# AN APPLICATION OF BUSINESS INTELLIGENCE TECHNIQUES IN THE CIVIL AVIATION CARGO PLANNING

**A THESIS SUBMITTED TO**
**THE INSTITUTE OF THE SOCIAL SCIENCE**
**OF ANKARA YILDIRIM BEYAZIT UNIVERSITY**

**IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR**
**THE DEGREE OF MASTER OF SCIENCE IN**
**MANAGEMENT INFORMATION SYSTEMS**

**By**
**Musab Talha AKPINAR**
**DECEMBER 2017**

**Ankara Yıldırım Beyazıt University,**

**The Institute of Social Sciences**

**Department of Management Information Systems**

# AN APPLICATION OF BUSINESS INTELLIGENCE TECHNIQUES IN THE CIVIL AVIATION CARGO PLANNING

**Musab Talha AKPINAR**

**Master's Thesis**

**Abdulkadir HIZIROĞLU Ph.D.**

**Ankara, 2017**

**T.C.**
**YILDIRIM BEYAZIT ÜNİVERSİTESİ**
**SOSYAL BİLİMLER ENSTİTÜSÜ BİLİMLER ENSTİTÜSÜ**

Tez Adı
AN APPLICATION OF BUSINESS INTELLIGENCE TECHNIQUES IN THE CIVIL AVIATION CARGO PLANNING

Tez Yazarının Adı
MUSAB TALHA AKPINAR

YÜKSEK LİSANS Tezi

Tez Savunma Sınav Tarihi

29.12.2017

Tez Danışmanı Adı
DOÇ. DR. ABDULKADİR HIZIROĞLU

Jüri Üyelerinin Adı
DOÇ. DR. ABDULKADİR HIZIROĞLU
DOÇ. DR. H. KEMAL İLTER
YRD.DOÇ. DR. HALİL İBRAHİM CEBECİ

Okuduğumuz ve Savunmasını dinlediğimiz bu tezin bir Yüksek Lisans derecesi için gereken tüm kapsam ve kalite şartlarını sağladığını beyan ederiz.

Tezi Onaylayanın Adı

Enstitü Müdürü

Bu tezin Yüksek Lisans derecesi için gereken tüm şartları sağladığını tasdik ederim.

Tez 22.27.2017

ORIJINALLIK RAPORU

| %6 | %8 | %4 | %2 |
|---|---|---|---|
| BENZERLIK ENDEKSI | İNTERNET KAYNAKLARI | YAYINLAR | ÖĞRENCI ÖDEVLERI |

BIRINCIL KAYNAKLAR

| | | |
|---|---|---|
| 1 | www.tachowdhury.com<br>İnternet Kaynağı | %1 |
| 2 | public.dhe.ibm.com<br>İnternet Kaynağı | %1 |
| 3 | ijbds.usb.ac.ir<br>İnternet Kaynağı | %1 |
| 4 | docplayer.net<br>İnternet Kaynağı | %1 |
| 5 | www.thaitransport.org<br>İnternet Kaynağı | %1 |
| 6 | www.futuretravelexperience.com<br>İnternet Kaynağı | %1 |
| 7 | documents.pageflip-flap.com<br>İnternet Kaynağı | %1 |
| 8 | digitalcommons.morris.umn.edu<br>İnternet Kaynağı | %1 |
| 9 | www.pomsmeetings.org<br>İnternet Kaynağı | %1 |

ABSTRACT

**AN APPLICATION OF BUSINESS INTELLIGENCE TECHNIQUES IN THE CIVIL AVIATION CARGO PLANNING**

**AKPINAR, Musab Talha**
**Master Program, Department of Management Information Systems**
**Supervisor: Assoc. Prof. Abdulkadir HIZIROĞLU**
**December, 2017**

Civil aviation has been growing 5 % per year since 1980 all over the world while in Turkey there has been 15% of growth over the last decade. Moreover, airlines in developed countries have adapted several business intelligence and analytics implementations in order to support their decision-making activities. However, this adoption has not satisfactorily achieved in developing countries such as Turkey. Especially, applied business intelligence studies as well as the academic research in this field have remained insufficient in those countries. The aim of this study is to put forth the different types of BI applications in civil aviation sector. Therefore, the data have been obtained from the airline company was analyzed in order to demonstrate different levels of business analytics. Analyses of the thesis include different analytics from descriptive to prescriptive in order to profile the data of international passenger's baggage. After the completion of data preprocessing steps, multidimensional analysis has been made and special outcomes were put forward. Following that, selected data mining techniques have implemented like clustering and association rules in predictive context. Lastly, using the outcome which is acquired from predictive analytics was supported by DEMATEL method that is one of the noteworthy multi-criteria decision-making methods and prescriptive analytics techniques. Consequently, effective regional differences were tried to obtain from passenger baggage profiles. This study also included with some real-world application examples, contributes both flight planning and cargo departments in the air transportation sector by setting an example for different levels of analytics in the academic sense.

**Keywords: Aviation Analytics, Air Transportation, Business Intelligence, Clustering, Data Mining, Multidimensional Analysis**

ÖZET

## SİVİL HAVACILIK KARGO PLANLAMA İŞ ZEKASI TEKNİKLERİNİN UYGULAMASI

**AKPINAR, Musab Talha**
**Yüksek Lisans Programı, Yönetim Bilişim Sistemleri Bölümü**
**Danışman: Doç. Dr. Abdulkadir HIZIROĞLU**
**Aralık, 2017**

Sivil havacılık dünyada 1980 yılından bu yana yılda %5 büyürken, Türkiye'de ise son on yılda yüzde 15'lik bir büyüme kat etmiştir. Üstelik, gelişmiş ülkelerdeki havayolları, karar verme faaliyetlerini desteklemek için çeşitli iş zekâsı ve analiz uygulamalarını uyarlamışlardır. Bununla birlikte, bu gelişmeler Türkiye gibi gelişmekte olan ülkelerde tatminkâr bir şekilde başarılmamıştır; bu nedenle, uygulamalı iş zekâsı (İZ) çalışmaları ve bu alandaki akademik araştırmalar yetersiz kalmıştır. Bu çalışmanın amacı, sivil havacılık sektöründe farklı seviyelerde İZ uygulamalarının ortaya konmasıdır. Bu amaca ulaşmak için havayolu şirketinden alınan veriler farklı seviyelerde çözümleme teknikleri ile analiz edilmiştir. Tezin analizleri, uluslararası yolcu bagaj verilerinin profillerini çıkarmak için tanımlayıcı analizden kuralcı analize kadar farklı boyutlar içermektedir. Ön hazırlık aşamalarının eksiksiz olarak tamamlanmasının ardından çok boyutlu analiz yapılmış ve belirli sonuçlar ortaya konulmuştur. Ardından, seçilen veri madenciliği teknikleri, kümeleme ve birliktelik kuralı ile tahminsel analizler yapılmıştır. Son olarak, bu analitikten elde edilen bulgular, çok kriterli karar verme yöntemi ve kural destekleme tekniklerinden olan DEMATEL yöntemleriyle desteklenmiştir. Özellikle yolcu bagaj profillerinde etkin, bölgesel farklılıklar elde edilmiştir. Gerçek uygulama örneklerini de içeren bu çalışma, akademik anlamda farklı analitik seviyeleri bir arada sunarken, hava taşımacılık sektörünün uçuş planlama ve kargo bölümlerine de katkı sağlamaktadır.

**Anahtar Kelimeler:** Sivil Havacılık Veri Analizi, Veri Madenciliği, İş Zekâsı, Kümeleme, Sınıflandırma, Hava Taşımacılığı

**ACKNOWLEDGEMENTS**

The author wishes to express his deepest gratitude to his supervisor Abdulkadir Hızıroğlu Phd. for his guidance, advice, criticism, encouragements and insight throughout the research. Without his contributions, this thesis would not be complete.

The author wishes also to express his special gratitude for his lecturer H. Kemal İlter Phd. H. İbrahim Cebeci PHd. and Keziban Seçkin Phd., for their self-sacrificing and very important contributions, advices, encouragements, inspirations and valuable efforts throughout the research.

The author would also like to thank, Muhammed Emin KARABACAK, Judge Abdurrahman Remzi AKPINAR and Semih Ceyhan for their valuable suggestions and comments throughout the master program.

The author wishes to express his special thanks to Kadir ÇOŞKUN for his valuable contributions and support. And also, the technical assistance from Turkish Airlines is also gratefully acknowledged.

## TABLE OF CONTENTS

## LIST OF TABLES



## LIST OF FIGURES & GRAPHS

### FIGURES

## GRAPHS

**LIST OF ABBREVIATIONS**

| | |
|---|---|
| Asymp. Sig. | Asymptotic Significance |
| BA | Business Analytics |
| BI | Business Intelligence |
| CRM | Customer Relationship Management |
| df | Degree of Freedom |
| DM | Data Mining |
| EIS | Executive Information Systems |
| ERP | Enterprises Resources Planning |
| ETL | Extract-Transform-Load |
| GRI | Generalized Rule Induction |
| IT | Information Technologies |
| OLAP | Online Analytical Processing |
| SCM | Supply Chain Management |
| SQL | Structured Query Language |
| Std. D. | Standard Deviation |
| THY | Turkish Airlines |
| USA | United States of America |

# INTRODUCTION

In this section, general expressions and concepts related to the thesis will be given. Motivation of the thesis and the aims of the study will also be mentioned in the subtitles. Furthermore, the flow of the study will be revealed and finally, the expected contributions and values will be highlighted and the organizational scheme will be shown. The scope of the basic concepts is presented firstly.

## 1.1. Background and Motivation

Information systems and technologies have begun to spread after World War II. Until the 1960s, the role of most information systems was simple: transaction processing, record keeping, accounting, and other electronic data processing (EDP) applications. Then another role was added, namely, the processing of all these data into useful, informative reports and the concept of management information systems (MIS) was born (O'Brien and Marakas, 2005). By the 1970s, it was evident that the specified information products produced by such management information systems were not adequately meeting the decision-making needs of management, so the concept of decision support systems (DSS) was born. The new role for information systems was to provide managerial end users with ad hoc, interactive support of their decision-making processes. Decision support systems combine data and mathematical models to help decision makers in their work. In the 1970s there began to arise within enterprises increasingly complex needs to devise software applications, called management information systems (MIS), in order to ease access to useful and timely information for decision makers

In the 1980s, several new roles for information systems appeared such as microcomputer processing power, application software packages, telecommunications networks, management information systems or the analytical modeling capabilities of decision support systems and also executive information systems (EIS). Finally, the rapid growth of the Internet, intranets, extranets, and other inter- connected global networks in the 1990s dramatically changed the capabilities of information systems in business at the beginning of the 21st century.

BI refers to all applications and technologies in the organization that are focused on the gathering and analysis of data and information that can be used to drive strategic business decisions. Through the use of BI technologies and processes, organizations can gain valuable insight into the key elements and factors—both internal and external—that affect their business and competitiveness in the marketplace. BI relies on sophisticated metrics and analytics to "see into the data" and find relation- ships and opportunities that can be turned into profits.

Briefly, management information systems provide information in the form of specified reports and displays to support business decision making for instance, sales analysis, production performance, and cost trend reporting systems. Decision support systems provide interactive ad hoc support for the decision-making processes of managers and other business professionals for example, product pricing, profitability forecasting, and risk analysis systems. Executive information systems provide critical information from MIS, DSS, and other sources tailored to the information needs of executives for instance, systems for easy access to analyses of business performance, actions of competitors, and economic developments to support strategic planning. Executive information systems, then began to offer additional visualization, alerts, and performance measurement capabilities. By 2006, the major commercial products and services appeared under the umbrella term business intelligence (BI) which combines architectures, tools, databases, analytical tools, applications, and methodologies. It is a content-free expression, so it means different things to different people. By analyzing historical and current data, situations, and performances, decision makers get valuable insights that enable them to make more informed and better decisions. The process of BI is based on the transformation of data to information, then to decisions, and finally to actions.

### 1.1.1. Business Intelligence: An Emerging Area

Data, or pieces of information, have been collected and used throughout history. However, in the today's world, advances in digital technology have considerably boosted the ability to find, collect, store, analyze and interpret data. However, the information is valuable when they are used on the true state and right format. Every type of data is precious data merely, when they are analyzed and used to inform decision-making (Akerkar, 2014). In recent years, technological developments have increased importance of data analytics. Moreover, technology-driven process for analyzing data and creating some applications in order to support executives, business managers and other decision-makers is called business intelligence (BI). BI can be shown as a result of the natural evolution of information technology.

The term Bl was coined by the Gartner Group in the mid-1990s. However, the concept is much older; it has its roots in the Management Information Systems (MIS) reporting systems of the 1970s. During that period, reporting systems were static and two- dimensional and had no analytical capabilities. In the early 1980s, the concept of executive information systems (EIS) emerged. This concept expanded the computerized support to top-level managers and executives. Some of the capabilities introduced were dynamic multidimensional (ad hoc or on-demand) reporting, forecasting and prediction, trend analysis, drill down to details, status access, and critical success factors (CSFs). These features appeared in dozens of commercial products until the mid-1990s. Then the same capabilities and some new ones appeared under the name BI. Today, a good BI-based enterprise information system contains all the information executives need. So, the original concept of EIS was transformed into BI. By 2005, BI systems started to include artificial intelligence capabilities as well as powerful analytical capabilities.

Business intelligence represents analytical tools and systems that allow a company to gather, store, access and analyze corporate data in order to aid in decision-making models and strategic planning process. BI is the system and a collection of integrated operational as well as decision-support applications and databases that provide the business world easy access to data (Moss and Atre, 2003).

Lexical meaning of intelligence is the capacity for learning, reasoning, understanding, and similar forms of mental activity, relationships with knowledge, the gathering or distribution of information, interchange of information. Before elaborating the business intelligence (BI), there is a need for mentioning business analytics (BA) in order to settle the conflict. Business intelligence and analytics are actually two prominent terms that include different tools and serve different purposes. BI is the main topic which cover main infrastructure of the systems, analytics, reporting, and visualization. In deed BI offer the opportunity to work with large data sets in a data warehouse environment and will learn the use of Online Analytical Processing (OLAP) and Visualization capabilities to create visualizations and dashboards. Moreover, BI have had a profound impact on corporate strategy, performance, and competitiveness and broadly encompass decision support systems, business intelligence systems, and visual analytics. At that point, BA is one of the stage the business intelligence process (Baars and Kemper, 2008). In fact, BI includes all components of the data operation, from when data is collected to when it is accessed. Analytics, on the other hand, is the process performed on data that has been delivered by BI for the purpose of generating insights to drive decisions, actions, and eventually, revenue or other impacts. Data is converted to insights using analytics tools.

In other words, data analysis lies at the heart of decision-making in all business applications (Cadez, and Guilding, 2008). Most of the analysis is made for explain or put forth “what is the matter or what is the situation”. All those studies can be counted in business intelligence. As it is clearly explained above, BI is a broad concept which contains all following operations and processes: collect, filter, combine, aggregate, visualize, interpret, internalize, revise, check, enrich, share, remember, decide, distribute anticipate. On the other side BA also covers a lot of technologies and tools which can explore system information, develop business processes and support the forthcoming plan and performance. However, BA include only analytical items like statistical analysis, predictive modeling, various kinds of data processing platforms and data driven decision. In other words, Business Analytics relates to the historical data from a lot of different source systems through statistical and quantitative analysis, data mining, predictive modelling and other technologies and techniques (Acito and Khatri, 2014).

Business intelligence, a term that encompasses all the capabilities required to turn data into intelligence, has emboldened companies to strive for the ultimate goal getting the right information to the right people at the right time through the right channel (Bitzan and Peoples, 2016). Grujar and Rathore (2013) diverse description for business intelligence is an overarching term which includes the applications, infrastructure and tools, and best practices that enable access to and analysis of information to improve and optimize decisions and performances.

Historically, BI was originally devised by Richard Millar Devens' in the 'Cyclopedia of Commercial and Business Anecdotes' in 1865, more than 150 years ago. He used the BI term in following sense, gathering and analyzing process include whole business information (Okkonen and et. al., 2002).

### 1.1.2. Business Intelligence in Service Context

Businesses and systems are overwhelmed with data, making the ability to store, process, analyze, interpret, consume, and act upon that data a primary concern (Tansley and Tolle, 2009). As a result of these rising trends, there is a widespread realization that an enormous volume of data can be captured, stored, and processed. After that the knowledge gleaned from such data has advantages for everyone:

business sectors, governments issues, academic disciplines, engineering working, communities', and individuals' studies (Akerkar, 2014).

Last decade of rapidly improving technology has given rise to powerful business intelligence tools. Those tools are handling spread several service industries; such as "Finance and Insurance, Health, Tourism, Telecom, Transportation" in order to improve operational efficiencies, reduce expenses, predict risk and increase customer retention and also develop Customer Relationship Management (CRM).

**Finance and Insurance**

Thanks to the internet and the proliferation of using mobile devices and their applications, today's financial institutions face competition, an ever-changing client demands and require for strict control and risk management in an extremely dynamic market (Claessens and Laeven, 2004). Tools that the banking, finance and insurance industry can use to leverage customer data for insights that can lead to smarter management practices and better business decisions. While insurance institutions are commonly using BI solutions to predict risk, banks are using BI to improve operational efficiencies for competitive advantages (Wu et al, 2014).

**Health**

Health sector, which is also known as healthcare business line, has used business intelligence tools after the evaluation of the computers (Olszak and Batko, 2012). In particular, the healthcare industry was on the brink of transformation after the millennium. Health sector, pharma and health care industries, advertising, retailer business management and many more service sector have been using business intelligence applications, and clearly, they will continue to use them increasingly (Bandyopadhyay and Sen, 2011). For better understanding, BI application on health sector can be subcategorized under the three main headings, healthcare decision supports systems, analytical hospital process management and analytics on medical diagnostic and disease/treatment monitoring (Cebeci and Hiziroglu, 2016). The new wave of technologies, lower prices, and greater availability of patient data have created a useful opportunity for health sector. Clinical analytics refers to a broad category of information technologies in order to gathering data and analyze that clinical inputs to help managers and clinicians make better decision making (Dowie, 1993). Clinical Business intelligence is the use of data analysis to improve care delivery regarding to improve operational efficiency and developing CRM. Not only descriptive, but also predictive analytics tools can assess patient risk of diseases, potential costs, analyze clinical data, and guide providers for billing process.

**Tourism**

Tourism is one of the significant industries which is used BI (Fuchs, 2013). The sector is one of the important participants of the e-commerce sector. Most of the tourism transactions can be done over the internet. Because of the fact that, reservations, purchasing, registration, enrollment, transcription can work through from the Internet. Tourism companies would gain a lot of advantages whole those data gathering from the online shopping, customer feedback and comment if they will use BI tools for understanding customer requirement increase customer loyalty and optimize services.

**Telecommunications**

Telecommunication is also one of the common usage areas or BI. Telecom companies are using business intelligence in order to measure roaming customer value and client activity, their margins and profitability by type of client (Nagash, 2004). Technology is transforming all service industries. Its advancement has not been done yet in fact, it can be said that it is at infancy age (Barrett et al, 2015). In addition, those institutions that adopt and fully utilize BI solutions to manage risk, increase operational efficiency, and provide their services that meet real customer needs will be better positioned to enjoy sustained growth, profitability and a competitive edge for years to come.

**Transportation**

BI technologies are capable of handling large amounts of structured and unstructured data to help identify, develop and create new strategic business and transportation opportunities. In transportation industry BI is very helpful not only for reducing labor or operation cost, but also for satisfy the clients and gives ability to look at data quickly to help assist with decisions (Akpinar and Hiziroglu, 2016). Identifying new opportunities and implementing an effective strategy based on insights can provide transportation companies with a competitive market advantage and not only short term, but also long-term stability. Associated with transportation as well as tourism and telecom can have lots of advantages for the BI applications and tools. For instance, improving capacity management by applying BI to better understand, customers' behavior and developing CRM using BI tools (Gangadharan and Swami, 2004).

Nowadays, increasing of the globalization and integration of the world, one of the most notable part of the transportation sector is civil aviation which is the continuing to grow at an average annual 5% since 1980 (IATA, 2016). Within that increase, automatically and manually accumulating data in airline company systems is impracticable due to the mass amount of data produced every day. The goal of BI is to allow easy interpretation of these large volumes of data. Because of the huge amount of the data in air transportation, airlines corporations are obligated to utilize the business intelligence tools.

The development of computerized reservations systems (CRSs) began in the 1950s, when American Airlines partnered with IBM in order to that purpose. (McKenney et al, 1995). To illustrate, "Semi Automated Business Research Environment" which is called as SABRE supplied airline reservations agents manage the distribution process, both as well as centralized reservations office and other ticketing offices which can be city, airport or another country. SABRE was "the first real-time business application of computer technology, an automated system with complete passenger records available electronically to any agent connected to the SABRE system" (Smith et al, 2001). The airline companies both operate predictions future customer behavior using historical data and improve a scalable analytics service based on their current requirements (Ayhan et. al., 2013). From these standpoints, BI can be very useful in involving several decision-makers with multiple criteria to arrive at optimal operational and financial solution.

### 1.2. Research Problems

Civil aviation industry, in developed countries have adapted several Business intelligence and analytics implementations in order to support their decision-making activities. However, when checking of the academic literature, this adoption has not satisfactorily achieved in developing countries such as Turkey. Especially, applied Business intelligence studies as well as the academic research in this field have remained insufficient in those countries. The research problem which are also motivation of this thesis is the gap in academic literature and deficiency of regarding studies in this field. Moreover, this study will also meet the needs of the aviation sector and will be the unique example with real world data.

Aviation sector has so much data which cannot be manually controlled. Aircraft always record data down in their black box. Planes equipped with flight recording data typically record up to 500 variables of data - described in these flight data recordings are time, altitude, vertical acceleration, and heading - per second for the duration the plane is being operated (Das and et. al., 2011). Furthermore, airlines have not only flights data, but also their own different type and size data like passenger, cargo, luggage, sales, etc. As it is mentioned before, not only military aeronautics, but also civil aviation is one of the best able to adapt to technological developments and business intelligence applications. Airline companies are using some of business intelligence techniques in order to reduce their cost, optimize flights and crew schedules. Airlines should use BI for rising operational efficiencies, in order to ensure permanence in a competitive market. Moreover, they ought to improve CRM in order to increase customer loyalty and manage external risk with BI techniques.

This study will focus on the passenger baggage analytics because it is a common knowledge that every item on board makes a plane heavier therefore burns more fuel. An airliner's cost of operating rises with every laptop (33 cents per flight), pillow (6 cents), or magazine (5 cents) you bring along (Stone, 2017). The most important factor of affecting the fuel cost is depending aircraft weights which are include net weight of the aircraft, the amount of fuel and the weight of the passenger baggage. (Abdelghany, Abdelghany and Raina, 2005). Thus, only can optimize and reduce 1 kg of each baggage, the airline can save about 30 million US dollar. There are some studies concern to decrease the aircraft weight and flights fuel optimization, however they didn't focus passenger baggage optimizations. Generally, luggage optimization studies are about draft and design aircraft part production like containers design (Yan et al, 2008) or some air cargo network design (O'Kelly et al, 1994; Kuby and Gray, 1993).

The luggage values in the different flights have difference between them based on distance and countries. To better understand this state of affairs, it is necessary to make the country and the distance-based profiling. Operational planning can be either tactic centric or budget centric (Axson, 2007). With this thesis results, airline companies can use the Business intelligence tools for some operational management for both tactic and budget centric. However, there is not enough a comprehensive and a noteworthy study in academic field. Some studies in this context that can be count (Gurbuz et al, 2011; Akerkar, 2014; Akpinar and Hiziroglu, 2016) do not include all phase of the BA. Especially studies in Turkey about this field have been very primitive, general and superficial. In this study, planned to use all three kinds of analytics techniques which are descriptive, predictive and prescriptive with real-world data. In order to meet the needs of the aviation sector and to fill the gap in the academic literature the focus of the study is international flights passenger baggage data which is will be mentioned later detailed.

As a result, there are two different sides of problem; first one is sectoral issue and the second one is gap in this field on academic literature. It has been thought that to describe and predict passenger baggage values are the problems for airline companies. If airline companies can profile passenger baggage amount, they will be able to optimize their cargo transfer or they can reduce oil expense. More importantly, there is no remarkable study regarding civil aviation passenger baggage optimization using business analytics techniques especially covering all such as, descriptive, predictive and prescriptive in academic literature.

### 1.2.1. Purpose

This thesis aims to illustrate civil aviation passenger baggage analytics using three different types of business analytics techniques in order to sort out research problem that are mentioned above. Furthermore, study reveals lucrative result regarding luggage compliance analytics and it will demonstrate novel applications with real data. More specifically secondary data were obtained from Turkish Airlines and essential data preprocessing was performed. Afterwards, processed data was implemented with descriptive, predictive and prescriptive methods. Owing to descriptive analytics, multidimensional reports and data cubes will be derived in order to facilitate managers' decision that working at strategic level. Predictive analytics will draw passenger baggage values structure and put forth meaningful cluster. Thanks to those clusters, extracting significant rules can be derived for forming marketing strategies and reducing expenses. Lastly, with the prescriptive analytics, effects of flights attributes will be recognized in order to support operational and financial decision. While, the results increase the profitability in the air transportation sector, on the other hand, study, which covers all phases of the analytics, will fill the gap in the literature, and that will be original and exemplary work in this field.

### 1.2.2. Objectives

To fulfill that purpose, analytics are organized under the three-main headings. First one is descriptive analytics, which includes data cubes and multidimensional report. Multidimensional reporting would be a crucial part of reportage, especially if the data has big amount and also is disorganized. Airlines have a tough information and long-term passenger and operations data. Moreover, their operations involve a very different variety of data field. To demonstrate, they have passenger behavior, selling fraud detection, flight safety control, airport security operations, financial forecasting, flights planning and design, targeted marketing, weather and external factors predictions and more. Furthermore, looking at all those main headings more detail, it can be discovered more subdivisions. To illustrate, cargo or passenger baggage can detail elaborately. Passenger flight preference can be analyzed by time, by place, by type, by frequency and also by attitude. The study will represent convertible tables and multidimensional reporting with real airline flights data.

Moreover, descriptive analytics are past data analysis using data aggregation and data mining. Descriptive analysis or statistics describes and explains in brief the raw data in an intelligible form. However, among the three types of analytics, 80% of organizations use descriptive analytics (Zhou, 2014), which are the most elementary form of analytics.

Secondly, predictive analytics will be shown which cover couple of data mining tools and algorithms. After the completed preprocessing steps, the data has been analyzed with predictive methods such as, data mining. Predictive analytics will be profile passenger cargo value in aviation sector. The reason for the focusing on passenger baggage is that, cargo is the one of the most profitable parts of air transportation, however, it is also significant expense item. Because, airlines fuel expense increases, depending on the factors that affect weight, like cargo. On the other hand, airlines can carry also cargo on passenger flights. Therefore, if passenger baggage profile is understandable, a profitability enhancer cargo optimization can be also made.

Lastly using prescriptive analytics, some multi criteria decision-making can be done. Furthermore, future prediction can be analyzed depending on past term luggage and cargo data. For example, depending on passenger baggage profiling and other attitudes, flight scheduling and aircraft planning can arrange. Especially airlines which flight to wide variety of point or countries, need to take into account past records and should profile flights in order to determine a requirement.

This study tries to present two main contributions in different fields. The first one is trying to support the air transportation sector with business intelligence applications by making support decision-makers with analytics results. In fact, the study aims at accomplishing all different phases of analytics from descriptive to prescriptive and demonstrate empiric application related to passenger luggage. The second one is an academic contribution which is picking up the gap in BI studies on the civil aviation sector. Aviation industry has grown rapidly over the last decade as Turkey's GDP. On the other hand, scientific studies in this field can count something on the fingers of one hand. In this study, all phases of the analysis will be revealed step by step and exemplariness applications and meaningful results would be demonstrated. Namely, the thesis, sets the pace revealed applications and results, furthermore, will fill in the gap in Business Analytics practice in air transportation.

Finally, from these applications and analysis, significant results and commercial-grade rules would be obtained for sector. Moreover, these rules may initially seem meaningful just for airline companies, but then they may turn into the entire service sector at a later stage. At the same time, this study will try to fill the gaps in academic field.

### 1.3. Methodology

The aim of this study is to follow from descriptive to prescriptive analytical process in order to improve usable result examples in air transportation. In this thesis, data has taken from Turkish Airlines which is one of the biggest 10 airlines in terms of the passenger number. Although this airline has not the most points in the globe, but it is the airline that flies to most different countries. Because of this, they have very different passenger and baggage profiles. As of January 1, 2016, the company has flown to 284 destinations in 113 different countries in 4 different continents, 43 domestic and 235 internationals. The data include 229.465 flights which are from

Istanbul to international point in during in 2014 and 2015. This flight has different types of attributes such as; time, location, number of the total passenger, number of the luggage, business and infant passenger rates and the distance of arrival airports.

The thesis methodology will also contain three main analytics heading. First one is descriptive analytics stage, which is employed to describe the basic features of the data in a study and help the systems to estimate what happened in the past (Trochim and Donnelly, 2001). Mainly, first stage has exploratory data analysis.

Second stage is predictive analytics, which tries to give a recommendation for key decisions based on future outcomes and focuses on answering the question: “What is probably going to happen in the future?”. Predictive analytics provides organizations with actionable insights based on data and estimation regarding the likelihood of a future outcome via a variety of techniques, such as machine learning, data mining, modelling (Dhar, 2013). Predictive stage includes clustering and depending on those clusters association rules has been derived. These rules can be used both in marketing and production/planning departments, in the meantime it is expected that fetch a new flow of view for the industry.

Lastly, prescriptive analytics have a huge effect on all business processes and they can show that how determinations are made. It can impact any industry, organization and systems and help them become more effective and efficient (Delen and Demirkan, 2013). To demonstrate, prescriptive analytics can optimize scheduling, production, inventory and supply chain design to deliver the right products in the right amount in the most optimized way for the right customers on time. Taking the results from the descriptive and prescriptive analysis throughputs, attributes have been assessed in order to interview with sectoral experts. For the prescriptive analytics stage, DEMATEL method has been developed with interview questionnaire. Thus, the results will support and also upgrade them with prescriptive interpretation.

In detail, for the analytics, first TwoStep clustering technique is used to the determined the cluster number. Secondly, K means is one of the simplest algorithm which uses unsupervised learning method in order to solve all known clustering issues (Krishna and Murty, 1999; Jain, 2010). It is appropriate for large datasets and it has strong sensitivity to any outliers (Huang, 1998; Zhao, Ma and He 2009). Lastly, about the association rules, the more then 75 % of the results have taken notice. For the prescriptive analytics phase, DEMATEL method has been developed interview questionnaire and also results reliability and their validity. To conclude, workflow of the thesis methodology will be shown in Figure 1.

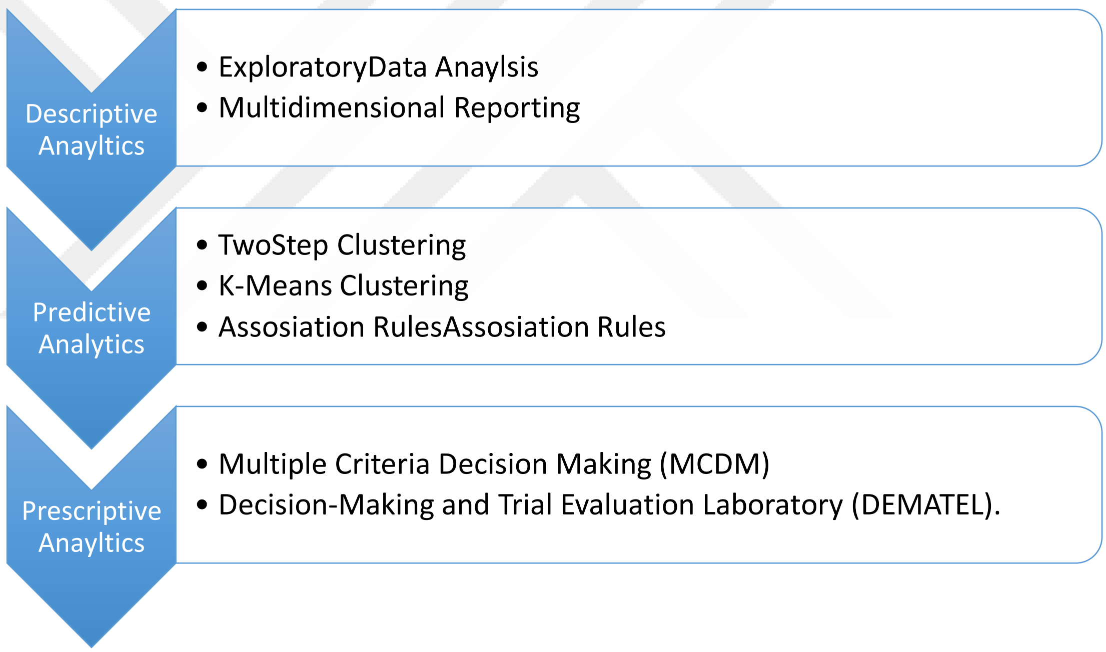



Figure 1: Workflow of the Thesis Methodology

## 1.4. Organization of the Study

The structure of the thesis is as follows. The second chapter is associated with literature which contains BI in civil aviation. The short history and future of the aviation is included, outline of the world civil aviation and general overview of air transportation in Turkey. Critical definition and description associated with business analytics and a brief introduction general aviation circumstance are presented. And also, in that section some BI studies and BA techniques are illustrated, regarding to civil aviation sector with the intent of providing readers unfamiliar with techniques some basic concepts in airline to understand the rest of the paper. Third Chapter includes methodology of the thesis. Methodology chapter will show the analytics process step by step which includes the method followed in accomplishing the critical analysis of the empirical studies. Empirical results and assessment are shown in the Chapter 4. Lastly, the Chapter 5 summarize by providing conclusion, discussion and also limitations and further studies of civil aviation business analytics research.

# LITERATURE

The Literature section is organized under five main topics: **(1)** the evaluation and development of civil aviation sector, **(2,3)** the literature of BI and demonstration of selected algorithms, **(4)** the literature of three different types of analytics with empirical studies, lastly **(5)** gaps in this field and focus on the thesis. The reason for this categorization is an opinion in business analytics including three main titles, and those will be shown separately in detail. Before analyze with real-world data in air transportation sector, it is important to comprehend the sectoral evolution and also to see the academic background of these concepts, especially with empiric studies in literature. It is also important to recognize the line requirements and realize the gaps in this field. The study plan of the literature section will also be shown in Figure 2.

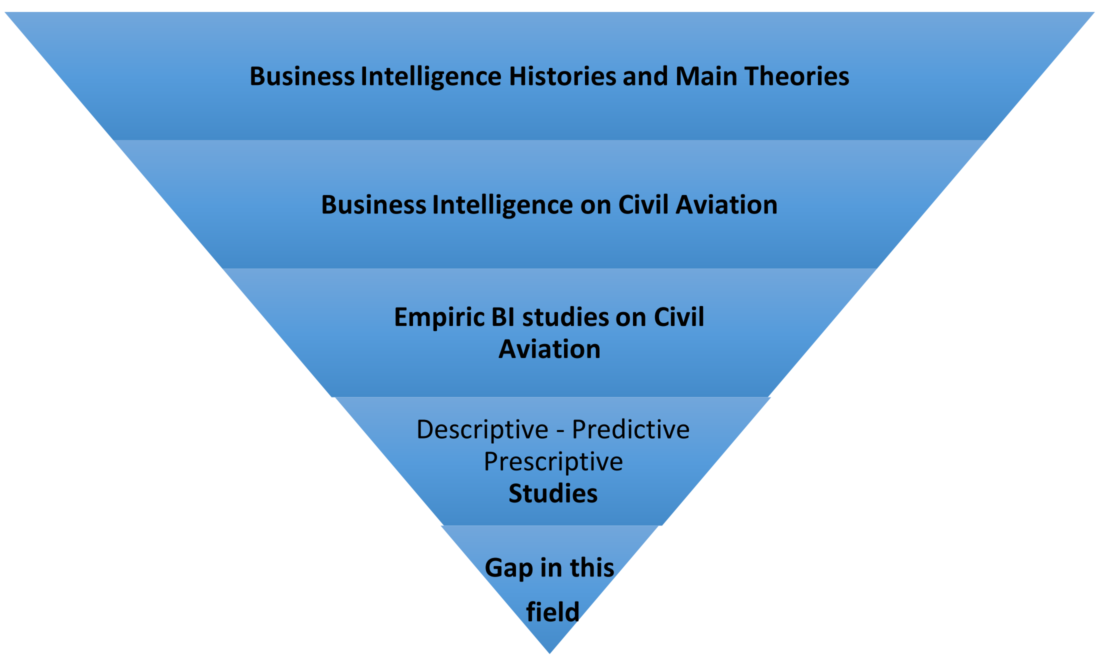



Figure 2: Study Plan of Literature Section

## 1.1.Evaluation of Civil Aviation and Future

Before the Wrights brothers first powered flight in a heavier-than-air machine in 1903, human being had always dreamed of flying throughout history and its performed experiments in different ways since 6th century. On the other hand, originate with the dawn of commercial aviation in 1914, aviation step into new age and become widespread. Thanks to the air transport industry, world has been globalized and removed the invisible bounders between not only countries, but also continents.

Regularly operating airlines in Europe, Africa, Australia, and South America in 1917, however people viewed air travel as a dangerous sport, not a safe mean of transportation on that time (Cleare, 2016). In 1920s governments begin to found their national airlines, for instance, British government who has formed Imperial Airways which is one of the first flag carrier. However, the aircrafts had to be larger, faster and more above all safer. Hence, lots of developments took place made, such as cockpit instruments, altimeters, airspeed indicators and they were installed in aircrafts in the 1930s that many believe it was the most innovative period in aviation history (Corn, 2002). One of the significant milestone was second world war, using air transport prevalence increased. In 1970's, air transfers carried about 300 thousand passengers and it has reached approximately 3.45 billion passengers in 2015 (ICAO, 2016).

On the economy side, there are significant crossroads for aeronautics as debut of first wide bodies aircraft produced by Boeing in 1969 (Irwin and Pavcnik, 2004). Moreover, economic crisis depending on oil prices, was also affected in 1970s like Arabs state embargos, Iran cut production, and started Iran Iraq war end of the 70s. Legitimate improvements were also experienced at the same time. First airline deregulations have started in order of cargo and passengers' liberalism had followed. After the millennium, most of airlines started to privatize and civil aviation authorities have completed liberalization. Unfortunately, the last milestone is 11/9 issues, it has changed many rules especially inflight and cabin safety and security.

According to International Civil Aviation Organization (ICAO) the aviation industry, composed of more than 1,400 commercial airlines, and about 4,130 airports continued to play a critical role over 2015 in fostering the growth of world tourism and trade (ICAO, 2016).

The IATA expectation is that air travel will reach more than double flights. And also, annual growth in air transportation passenger will be average % 4 during the next 20 years. However, according to the estimation, growth differs widely by market, e.g. in china civil aviation will increase more than 8. Moreover, according to those prediction, developing countries show fastest the growth like India, Indonesia, China and Brazil. And also, Turkey may be 7th most growing country and should be 10th air passenger ranking after 2030 (IATA, 2015).

On the other hand, according to the Boing long term reports, over the next 20 years World general GDP will increase % 3 while total airplane fleet will increase % 3.6, number of passenger will growth % 4 and also airline traffic will go up approximately 5 percent (Boing, 2015). Middle east long term estimation regarding to the Boing reports, airplane fleet will rise 5.2 percent and traffic growth will reach %6.2. Additionally, about 80 percent of the world's population lives within an eight-hour flight of the Middle East. That reminds of the words of Napoleon, "If the world was only one country, Istanbul would be its capital".

Turkey has fascinating growth in civil aviation sector after legal regulations and privatization of Flag carrier in 2003. To illustrate, the aircraft number has increased from 162 to 540 in 2016. According to IATA reports, one of the most significant impact of this situation demographic improvement. For instance, international flight number was approximately 218.000 in 2013 has increased 563.000 in 2016. Over and above domestic flights increased about 570 percent from 2003 to 2016 and its shows Turkey more developing within country (DGCA, 2017). The ICAO reports also show us this advancement hasn't finished for Turkey. By 2035, it is estimated for Turkey that the total number of passengers will be ranked 10th, even the domestic passenger number will rise 8th placed on the world (ICAO, 2015).

### 1.1.1. Civil Aviation in the World

Air transportation is a major industry in its own right and it also provides important inputs into wider economic, political, and social processes. The aviation industry supports $2.7 trillion (3.5%) of the world's gross domestic product (GDP). The world's airlines carry over three billion passengers a year and 50 million tons of freight. Providing these services generate 9.9 million direct jobs within the air transport industry and contributes $664.4 billion to global GDP. Compared with the GDP contribution of other sectors, the global air transport industry is larger than the automotive industry, which accounts for 1.2% of global GDP and chemicals manufacturing (2.1%). Tourism is fast becoming the world's largest industry and air transport plays a very important role in supporting this sector. Conservative analysis suggest that aviation supports $892.4 billion in economic activity within the tourism industry.

Nowadays, increasing of the globalization and integration of the world, one of the most notable part of the transportation sector is civil aviation sector, has grown at an average annual 5% since 1980 (IATA, 2016). Over and above civil aviation, passenger number has increased continuously, last 20 years and also in during the financial crisis in 2008 the only exception was 2001-02 because of the 11/9, it has really influenced. Another enchanting statistic about air passenger numbers, since 2009, the total number of the air transport carried passenger (for scheduled flights) increased approximately 54% and reached the 3.44 billion passengers in 2015 (ICAO, 2016). Moreover, airlines registered carrier departures were about 34 million worldwide also in 2016. Moreover, not only regular airlines but also Low-Cost Carriers (LCCs) almost reached 1 billion passengers in 2015, approximately 28 per cent of total scheduled passengers (ICAO, 2016).

On the other hand, scheduled air freight traffic which means air cargo grew by % 2.2 in 2015 – it was % 4.9 growth rate registered in 2014 because of the stagnating general economic situation worldwide in the last year. At the same time, air cargo load factor declined from %50 in 2014 to %47 in 2015, also mirroring generally weak world trade trends (ICAO, 2016).

Checking the airlines marshaling, according to the revenues, Forbes' world's biggest airlines are American, Delta and Continental airlines which has about $40 billion revenue each. American Airlines also is the first airlines regarding to the passenger carried sorting (Forbes, 2017). The world's busiest airports by passenger traffic according to each airport or country's civil aviation authority, was Atlanta, Beijing and Dubai in 2016. Hartsfield–Jackson Atlanta International Airport is the first airport which reached 100 million people in a year.

On the cargo side, according to the freight tonne-kilometres (FTK), world's most important companies are FedEx, Emirates and UPS airlines (IATA, 2016). About the fleet size, also American companies like American, Delta, United, Southwest and FedEx has been ranked among the top five. Furthermore, Turkish Airlines, Lufthansa and Air France serve the largest number of countries and Turkish Airlines flied 120 different countries in 2016.

### 1.1.2. Civil Aviation in Turkey

Air transport has always been seen as having an inherently strategic role. It has obvious direct military applications, which it is also highly visible, and for some period, and in some countries still, has been seen as a "flag carrier", a symbol of international commercial presence. Turkey has effective geopolitics location and also it is natural bridge between three continents and an emerging aviation hub at the intersection of sophisticated Europe with emerging markets in the Middle East and rest of Asia as well as in Africa.

Turkish aviation story is also one of the oldest aviation histories in the World. Ottoman government the first forays into aviation at an 'Aeroplane Station' set up at Ayastefanos (today's Yeşilköy), where the foundations of the history of aviation in Turkey were laid in 1911 (Skylife, 2013). The flag carrier of Turkey is Turkish

Airlines which began its journey in 1933 with just 5 airplanes and it was a monopoly not only passenger cargo flying but also airports handling until 1987 which has started civil aeronautics deregulation in Turkey (Cetin and Benk, 2011). New firms entered the market in 90s but they were not very competitor for THY, which was dominant, because it had not been privatized yet. After the 2003, new vision adopted for THY is still in progress. Even tough, the globally destructive effects of recent economic difficulties the company has achieved establishment and continued expansion of the world's most comprehensive route network.

From 2003 to 2014, Turkish Civil Aviation sector grew on average 13.7% each year, comparing to a growth rate of 5.7% per year for the aviation industry globally and Turkey got 9th highest aviation traffic in the world with approximately 97 million people in 2015 (ICAO, 2016). It is interesting that even during the crisis of 2009, while Turkey's GDP decreased 5%, air passenger numbers still increased by 6%. In 2015 Turkey had more than 743 thousand registered carrier departures, and grew more than five-fold in last decade (ICAO, 2016).

For the future, worsening security circumcision risks damaging Turkey's tourism and appeals to international business, which may in turn damage the civil aviation sector. However, Istanbul's third airport is now under construction, and is planned to be one of the biggest in the world. When it opens in 2018, it will certainly redraw the lines on the world aviation generic map and will likely cement Istanbul's position as a regional hub (Worldfolio, 2016).

## 1.2. An Overview of Business Intelligence

Business intelligence (BI) is a concept that managerial philosophy and applications are using in order to support work and organization to manage and clarify data and to make more useful for business decisions (Ghoshal and Kim, 1986). Business intelligence covers the whole systems and transforms the pure data into advantageous information in order to the decision-making process for the firms. According to Pirttimäki et al, BI term has 2 main meanings; first one is regarding to data in the business which includes organization, market, customer, competitors etc. and the second one is an intelligence producing process (Pirttimäki et al, 2006).

Let's think of a mom-and-pop store in the 1950s. He had 20 customers and he sold only bread, egg, milk and 1 newspaper. He kept an account for his revenue, cost of sales and his inventory account. He would have 240 transactions for every month and would need 2880 different cells in order to jot note for whole his small business. At that time, he only needed 60 sheets graph notebook for 5 years his accounting, inventory, business, customer and supplier management. Let's look at the present and think about mini supermarket which are in the streets of all the home. There are about 1000 different product and they have more than 100 suppliers for those and also minimum 5 shift workers. About the sales, they facing 2000 customer each day and they have to keep that account. Moreover, those stores are not independent, actually they have headquartered, warehouse and maybe about 3000 different stores. They really need to check their customer relationships, manage the human resources, conduct supplier supply chains, accounting, implement laws and regulate their own rules and many more issues. In this position, they would have more than 10 billion transactions to look at daily. And also, they should care their advertising, market future, competitors, macroeconomic assessment and strategic decision. Therefore, maybe 70 years ago one graph notebook was adequate however nowadays computers are not satisfactory without Business intelligence tools.

Subsequently, the 1970, some analytical software packages have begun showing up in the market. But, deficiency of computing power, difficulties with using and low user-friendliness, and bulky systems and manual integration with the transaction systems becloud BI tools from widespread use. The spreadsheet is still widely used in this area today, software like Lotus and Excel has become popular in the 1980s and they enabled end users creating own data models in order to business analysis (Rasmussen et al, 2002). Especially end of the 90s colorful software screens were started to become user-friendly with big buttons and easy to use, sheets reduce, the need for secretaries and assistants to write and print reports for them.

After the millennium the usage of SQL databases; data warehouse technologies; extraction, transformation, and loading (ETL) tools were progressing rapidly and carried a step further Business intelligence systems to last decade. With the advent and prevalent of the internet BI software vendors have released cloud systems and not only big companies, but also small firms can reach and use these kinds of systems with really low cost and hands down. Thanks to the intranet and internet, every employee can work own job and analyze data wherever his need from home, while traveling, or from any other location at wherever they have chance to be.

When looking at the economic side, worldwide analytic and performance management, in other words business applications software market, reach 15 billion dollar an increasing rise in 2015 (IDC, 2016). The market has been developing all along, even the last decade the average of the entire expansion is more than % 10 and the bossy companies of IS technologies which are SAP, Oracle and IBM continue to lead the market with more than % 40 of total shares (IDC, 2016).

Nowadays, Business intelligence cover data about historical information, diversified source generated and also using strategic, operational and tactical decision making. Before the millennium, BI application was only used for the information technology professionals, computer engineer or special analysts. However, last decade, especially improvement of the enterprise resources planning (ERP) systems worker of all levels can reach the all necessary data and they can also filter or pick essential ones and they may use their daily routine work. In fact, Business intelligence comprehends the whole data analysis tools, querying systems and ad-hoc networks, ERP, online analytical processing (OLAP), online and real-time data processing, operational data, cloud and software as a service Business intelligence, open source systems, geographic information system (GIS) and also data screening. The data visualization is also incorporating designing charts and infographics which are very popular in last years and also dashboards, scorecards that display visualizes data on business metrics and key performance indicators (Stacey et al, 2013).

With descriptive, predictive and prescriptive analytics, understanding your business will become easier and better-informed decisions can be made which take into account future outcomes. Entire content is summarized and shown below in the Figure 3.

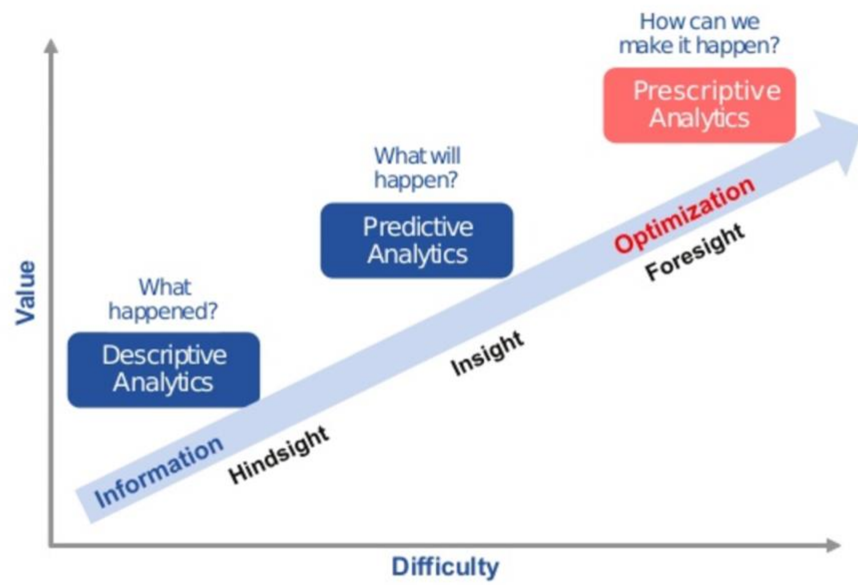



Figure 3: Types of Analytics

Source (Gartner and Ortag, 2012)

In this paper, there are also same processes just like historical development. It means that, the descriptive analysis will be shown about civil aviation with Multidimensional Reporting. Predictive analyses will be exhibited with some Data mining tools, such as clustering and association rules. Prescriptive analytics will be tried to be put down to fact with some Multiple Criteria Decision Making (MCDM) process and Decision-Making and Trial Evaluation Laboratory (DEMATEL).

### 1.3. Selected Business Intelligence Tools

As the history data analytics show that, last century people try to understand how the organizations or the world around them behaves by analyzing the available data. In the past, this used to be merely descriptive sciences. Descriptive statistics are employed to describe the basic features of the data in a study and help the systems what happen in the past (Trochim and Donnelly, 2001). These analyses provide simple summaries about the sample and the criteria, but approach to take in the future: learn from past behavior to influence future outcomes. Together with simple graphical analysis, they form the basis of virtually every quantitative analysis of data. In addition, currently estimated that more than 80% of Business Analytics are descriptive (Wu, 2014).

After the descriptive analyzing studies begin the new area of predictive analytics, which tries to give a recommendation for key decisions based on future outcomes and focuses on answering the question: “What is probably going to happen in the future?”. Predictive analytics provides organizations with actionable insights based on data and estimation regarding the likelihood of a future outcome via a variety of techniques, such as machine learning, data mining, and some modelling (Dhar, 2013). For example, predictive analytics can help to identify any risks or opportunities in the future and also can be used in all departments, from predicting customer behavior in sales and marketing, to forecasting demand for operations or determining risk profiles for finance. Because of the more data means better predictions, so it is important to have as much data as possible in predictive analyses.

Prescriptive analytics is the last step up in data analytics and is the final stage in understanding business, but it can be said, it is still in its infancy. Actually, prescriptive analytics not only foresees what will occur and when it will happen, but also why it will happen and provides recommendations how to act upon it in order to take advantage of the previsions. It practices a combination of many different optimization techniques and instruments such as managerial sciences, business rule algorithms, machine learning and computational modelling techniques as well as many different data sets ranging from historical and transactional data to public and social data sets. Prescriptive analytics try to determine what the effect of future decisions will be in order to adjust the decisions before they are actually created.

This will improve decision-making a lot as future outcomes are taken into consideration in the anticipation.

Prescriptive analytics could have a huge effect on all business and how determinations are made and it can impact any industry, organization and systems and help them become more effective and efficient (Delen and Demirkan, 2013). To demonstrate, prescriptive analytics can optimize your scheduling, production, inventory and supply chain design to deliver the right products in the right amount in the most optimized way for the right customers on time.

### 1.3.1. Descriptive Analytics: Multidimensional Reporting

Before the starting to the making all kind of analyses, data preprocessing was the significant necessity. It involves checking or logging all data in; checking the data for accuracy; entering the data into your structure; transforming the diverse data; and progressing and documenting a database structure that integrates the various measures (Kimball and Caserta, 2014). OLAP (online analytical processing) is computer processing that enables a user to easily and selectively extract and view information from different points of perspective. Generally, OLAP data is stored in the multidimensional database. Whereas those relational databases can be thought of as two-dimensional or multidimensional database considers each data attribute (such as distance, geographic location, region, and time) as a separate "dimension."

OLAP is the crucial technology behind lots of BI applications. OLAP is a robust technology in order to data discovery, and also capabilities for unlimited report viewing, complicated analytical calculations, and predictive “what if” scenario (prediction, propose, forecast) planning. Moreover, OLAP can be used also data mining or the discovery of previously undiscerned relationships between data items. In fact, the outcomes from the OLAP cubes or applications can be used in order to prescriptive analytics. OLAP is establishment for many types of business applications for Performance Management, Planning and Forecasting, Budgeting, Diverse and Multidimensional Reporting, Analysis, Simulation, Knowledge Discovery, and Data Warehouse Reporting (Negash and Gray, 2008).

In order to Multi-Dimensional Reporting, make useful and accomplished there are new techniques to specify how to arrange and organize complex crosstab definitions that will help organize reports and make them more easily understood by report consumers (Vogt and Johnson, 2011). To analyze and report on the health of a business and plan for future activity, many variable parameters must be tracked on a continuous. These variable groups or parameters are called dimensions in the OLAP environment.

On the other side, descriptive statistics is the term given to the analysis of data that helps describe, exhibit or summarize data in a meaningful way. For instance, with Descriptive Analyses, patterns may emerge from the data. However, descriptive analytics do not allow to make conclusions beyond the data we have analyzed regarding any hypotheses may have made. Descriptive statistics therefore enables us to present the data in a more meaningful way, which allows simpler interpretation of the data. (Trochim and Donnelly, 2001).

OLAP allows managers, and analysts to get an insight of the information through fast, consistent, and interactive access to information. Generally, OLAP classified three main types Relational OLAP (ROLAP), Multidimensional OLAP (MOLAP), Hybrid OLAP (HOLAP) (Bédard et al, 2001).

Firstly, ROLAP servers are placed between relational back-end server and client front-end tools. To store and manage warehouse data, ROLAP uses relational or extended-relational DBMS. ROLAP includes the following; Implementation of aggregation navigation logic, Optimization for each DBMS back end, Additional tools and helps. Secondly, MOLAP uses array-based multidimensional storage engines for multidimensional views of data. With multidimensional data stores, the storage utilization may be low if the data set is sparse. Therefore, many MOLAP server use two levels of data storage representation to handle dense and sparse data sets. Thirdly, Hybrid OLAP is a combination of both ROLAP and MOLAP. It offers higher scalability of ROLAP and faster computation of MOLAP. HOLAP servers allows to store the large data volumes of detailed information. The aggregations are stored separately in MOLAP store (Deshpande et al, 1999).

Since OLAP servers are based on multidimensional view of data, we will discuss OLAP operations in multidimensional data. OLAP operations has also four different varieties, Roll-up, Drill-down, Slice and dice Pivot. Roll-up performs aggregation on a data cube in any of the following ways; by climbing up a concept hierarchy for a dimension and by dimension reduction. The Figure 4 illustrates how roll-up works.

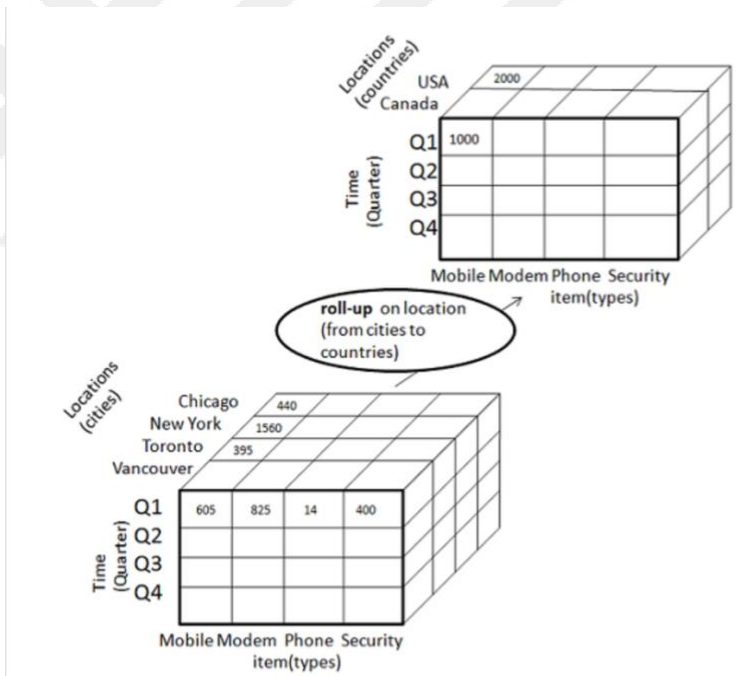



Figure 4: OLAP Roll-up Operations (Point, 2017)

Roll-up is performed by climbing up a concept hierarchy for the dimension location. Initially the concept hierarchy was "street < city < province < country". On rolling up, the data is aggregated by ascending the location hierarchy from the level of city to the level of country. The data is grouped rather into cities than countries. When roll-up is performed, one or more dimensions from the data cube are removed.

Moreover, drill-down is the reverse operation of roll-up. It is performed by either of the following ways; by stepping down a concept hierarchy for a dimension and by introducing a new dimension. The Figure 5 illustrates how drill-down works.

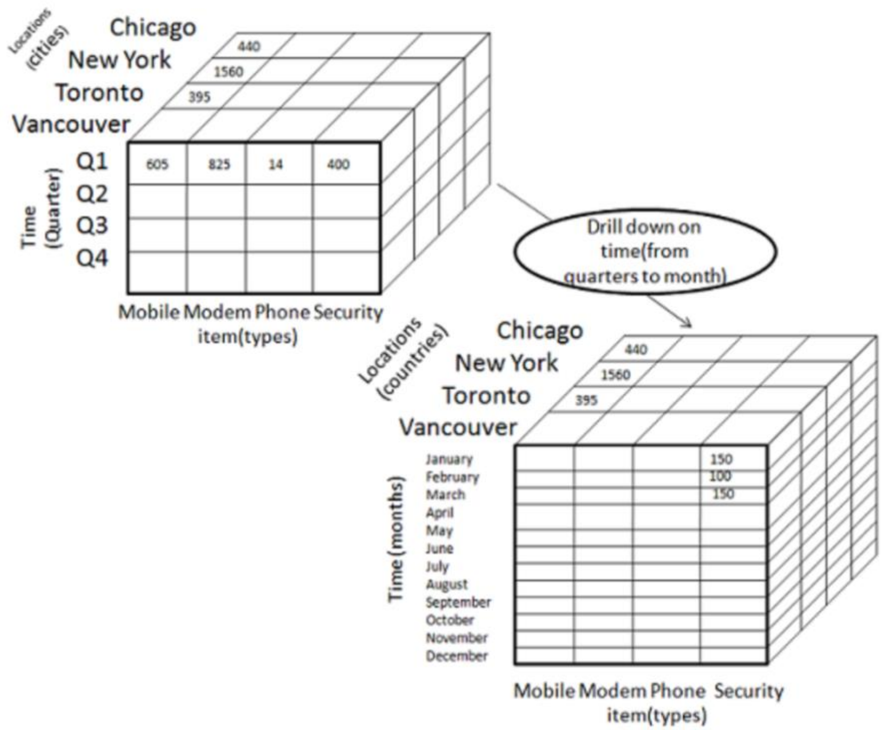



Figure 5: OLAP Drill-down Operations (Point, 2017)

Drill-down is performed by stepping down a concept hierarchy for the dimension time. Initially the concept hierarchy was "day < month < quarter < year." On drilling down, the time dimension is descended from the level of quarter to the level of month. When drill-down is performed, one or more dimensions from the data cube are added. It navigates the data from less detailed data to highly detailed data.

The slice operation selects one particular dimension from a given cube and provides a new sub-cube. Consider the Figure 6 that shows how slice works.

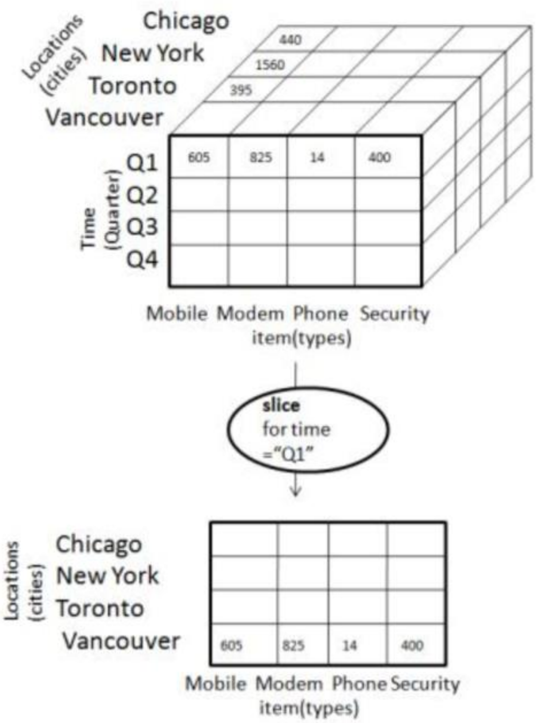



Figure 6: Slice Operations (Point, 2017)

Here Slice is performed for the dimension "time" using the criterion time = "Q1". It will form a new sub-cube by selecting one or more dimensions. Dice selects two or more dimensions from a given cube and provides a new sub-cube. Consider the Figure 7 that shows the dice operation.

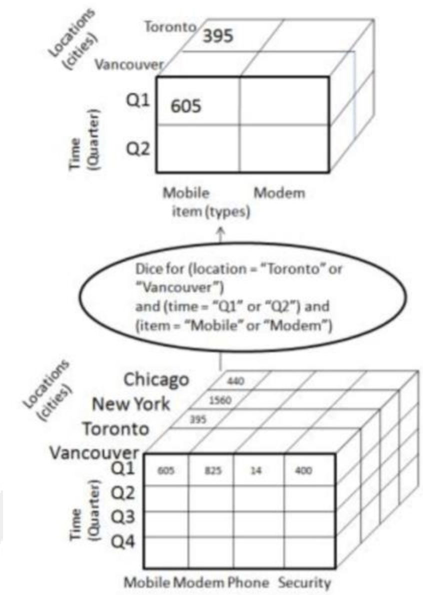



Figure 7: Dice Operations (Point, 2017)

The dice operation on the cube based on the following selection criteria involves three dimensions. (location = "Toronto" or "Vancouver") (time = "Q1" or "Q2") (item =" Mobile" or "Modem"). The pivot operation is also known as rotation. It rotates the data axes in view in order to provide an alternative presentation of data. Consider the Figure 8 that shows the pivot operation.

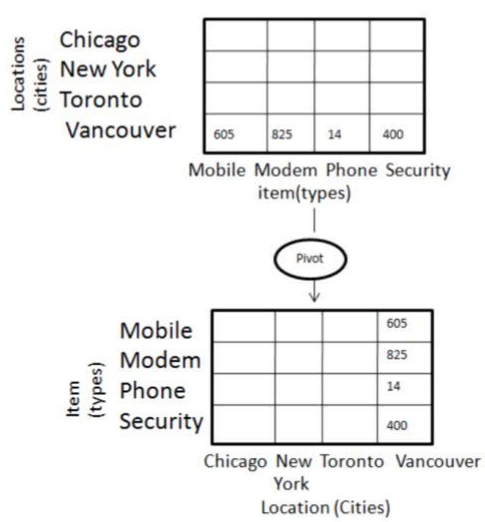



Figure 8: Pivot Operations (Point, 2017)

This study will focus on the MOLAP which is mentioned below. Multidimensional OLAP (MOLAP) uses array-based multidimensional storage engines for multidimensional views of data. With multidimensional data stores, the storage utilization may be low if the dataset is sparse. Therefore, many MOLAP servers use two levels of data storage representation to handle dense and sparse datasets. MOLAP tools process information with consistent response time regardless of level of summarizing or calculations selected. MOLAP tools need to avoid many of the complexities of creating a relational database to store data for analysis and also need faster possible performance. MOLAP server adopts two levels of storage representation to handle dense and sparse data sets. Sub-cubes are identified and stored as array structure. Sparse sub-cubes employ compression technology.

MOLAP has a lot of advantages when it compared with other OLAP tools. For instance, MOLAP allows the fastest indexing to the pre-computed summarized data. And also, it helps the users connected to a network who need to analyze larger, less-defined data. Moreover, using MOLAP is easier than others, thus MOLAP is suitable for inexperienced users.

### 1.3.2. Predictive Analytics: Data Mining

BI programs can also include advanced analytics forms, such as data mining, intelligent analytics, text analysis, statistical analysis, and large data analysis (Russom, P. 2011). However, in many cases, advanced analytical projects are conducted and managed by individual groups of scientists, statisticians, intelligent designers and other qualified analysts, and BI teams monitor simpler queries and analyze business data. Data mining describes from oxford dictionary very basic practice of examining large pre-existing databases in order to generate new information (Akpinar, and Karabacak 2017). It helps in extracting and refining useful knowledge from different types and size of datasets.

According to the description of Barai (2003) obtained and aggregated information can be used to form a prediction or classification model, identify trends and associations, refine an existing model, or provide a summary of the datasets which are mined. Definition of the Han and Kamber, (2000) from their books is that data mining is the task of discovering interesting patterns from large amounts of data, where the data can be stored in databases, data warehouses, or other information repositories.

Thomas Bayes's (1763) paper is published posthumously regarding a theorem for relating current probability to prior probability called the Bayes' theorem. From the second half of the 19th century data collection and database creation has thieved stepwise. Alan Turing (1936) introduced the idea of a Universal Machine capable of performing computations like our modern-day computers. The last quarter before the millennium database management systems flourished with hierarchical, network and relational database systems theories. After 80's advanced studies have been put forward about those fields. And then after the widespread use of the Internet in commercial area, XML and web based databased systems improved and data-information integration has spread (Vercellis, 2010). Data mining is a step in the knowledge discovery and process consisting of particular DM algorithms that, under some acceptable computational efficiency limitations, produces a particular enumeration of patterns (Piatetsky-Shapiro, 1996). Although the term data science has existed since 1960s, it wasn't until 2001 that William S. Cleveland introduced it as an independent discipline. The civil aviation industry is one of the earliest adopters of data science with amounts of data in their systems. Over the years, this industry has been making large investments to mine data and explore opportunities to improve operational efficiency and boost customer loyalty.

With one of the general definition; data mining (DM) is the analysis of huge amount of the data in order to discover classified information during the regular process of all business. According to description of Moss DM application touch all the bases use artificial intelligence, pattern recognition, databases, traditional statistics, and graphics to present hidden relationships and patterns found in the organization's data pool (Moss et al, 2003). However, DM is diverse from conventional statistical analysis and both of them have strengths and weaknesses. Statistical analysis and DM comparison are shown in Table 1.

Table 1: Statistical Analysis vs Data Mining

| **Statistical Analysis** | **Data Mining** |
|---|---|
| Statisticians usually start with a hypothesis. | Data Mining does not require a hypothesis. |
| Statisticians have to develop their own equations to match their hypothesis. | Data Mining algorithms can automatically develop the equations. |
| Statistical analysis uses only numerical data. | Data Mining can use different types of data (e.g., text, voice), not just numerical data. |
| Statisticians can find and filter dirty data during their analysis. | Data Mining depends on clean, well-documented data. |
| Statisticians interpret their own results and convey these results to the business managers and business executives. | Data Mining results are not easy to interpret. A statistician must still be involved in analyzing the Data Mining results and conveying the findings to the business managers and business executives. |

Source: (Moss and Atre, 2003)

Data mining tools need to be versatile for ability to apply a various model, scalable to work either a small data set or on larger data sets, capable of correctly assuming reactions between activities and results, and capable of automatic applications (Linoff and Berry, 2011). Even though, some analytic functions are often automated, human being should set up some implementing procedures. Moreover, analyst judgment is important for accomplished applications of data mining. Furthermore, before starting to make analyze, selection of data type and size is also critical. In addition, the transformation of data is often required (Olson and Delen, 2008).

On the other hand, data mining is the secrecy that surrounds its implementations. Moreover, they mentioned that some of the companies, that have implemented data mining hesitate to talk about their successes (Moss et al, 2003). Especially big firms in service sector which have vast quantity of customer and operations data.

There are many types of data mining applications. One of the important one is market management involving Cross-selling, Defecting customers, Promotions and campaigns, Prospecting, Market basket analysis. It is also used for Fraud detection; credit card or calling card fraud or insurance fraud. Moreover, risk management is one of the most significant applications of the data mining such as credit risk, quality control. Furthermore, financial service and distribution there are also prevalent applications of data mining. Data mining steps are shown in Figure 9

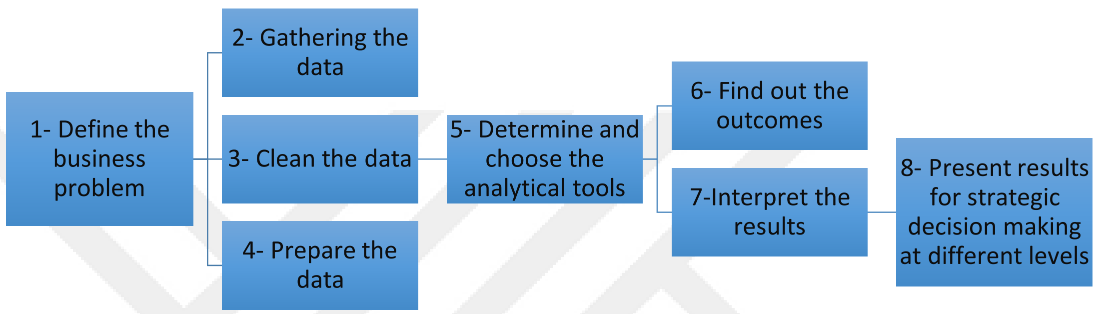



Figure 9: Data Mining Activities Process

**Functions of Data Mining**

The selection of segmentation techniques has become more important due to the fact that the developments in information and communication technologies, especially database management systems and data mining have changed (Hiziroglu, 2013). Data Mining can be achieved by, Classification, Clustering, Predictions, Sequential Patterns, Association and Similar Time Sequences (Chen et al, 1996; Olson, Shi and Shi, 2007; Han et al, 2011).

The term "Data Mining" was introduced in the 1990s, but data mining is the evolution of the field with a long history. Early methods of identifying patterns in data include Bayes' theorem (1700s) and regression analysis (1800s). The proliferation, ubiquity and increasing power of computer technology have increased data collection, storage and manipulations. As data sets have grown in size and complexity, direct hands-on data analysis has increasingly been augmented with indirect, automatic data processing. This has been aided by other discoveries in computer science, such as neural networks, clustering, genetic algorithms (1950s), decision trees (1960s), and support vector machines (1990s).

Classification and Prediction

There are two forms of data analysis that can be used for extracting models describing important classes or to predict future data trends which are classification and prediction. Briefly, classification algorithm predicts categorical class labels; and prediction algorithm predicts continuous valued functions. The major issue is preparing the data for Classification and Prediction. Preparing the data involves the following activities which are data cleaning, relevance analysis, normalization and generalization. The criteria for comparing the methods of classification and prediction are accuracy, speed, robustness, scalability and interpretability (Lim et al, 2000).

Classification is a data mining technique that assigns categories to a collection of data in order to aid in more exact predictions and analysis. Also, sometimes called a decision tree, classification is one of several methods intended to make the analysis of very large data sets more effective. If there is a class label associated with each data item, it should be a training set, and it can be used to build a model of the data which separates one class label from another. There are several classification algorithms, each of which builds a different type of model, and each with its pros and cons. If the label is continuous instead of discrete, the problem can be addressed by using regression techniques (Chandrika, 2009).

The ID3 algorithm was originally developed by J. Ross Quinlan at the University of Sydney, and he first presented it in 1975 in the book “Machine Learning”. ID3 identifies attributes that differentiate one class from another. All attributes must be known in advance, and must likewise be either continuous or selected from a set of known values. To find out which attributes are the most important, ID3 uses the statistical property of entropy. Entropy measures the amount of information in an attribute.

On the other hand, the C4.5 algorithm overcomes this problem by using other statistical property known as information gain. Information gain measures how good given attribute separates the training sets into the output classes.

One of the other approximation algorithm is that genetic programming (GP) has been vastly used in research in the past 10 years to solve data mining classification problems. The reason genetic programming is thus widely used is the fact that prediction rules are very naturally represented in GP. Additionally, GP has proven to produce dependable results with global search problems like classification. The search space for classification can be distinguished as having several ‘peaks’, this causes local search algorithms, such as simulated annealing, to perform badly.

The other diverse algorithm is neural networks which were modeled after the cognitive operations of the brain. They are capable of predicting new observations from existing observations. A neural network consists of interconnected processing elements also called neurons. The neurons within the network work together, in parallel, to produce an output function.

Although, predictive analytics is usually related to data mining to describe how information or data is processed, there are significant differences between these techniques. Predictive analytics and data mining use algorithms to discover knowledge and find the best solutions. Data mining is a process based on algorithms to analyze and extract useful information and automatically discover hidden patterns and relationships from data.

Data mining offers promising ways to reveal hidden patterns within large amounts of data. These hidden patterns can potentially be utilized to forecast future behavior. The availability of novel data mining algorithms, however, should be taken on with caution. Furthermore, in order to the best results, data analyst should compare or even combine available techniques in order to obtain the best possible results.

Clustering

Clustering is the process of grouping the data into clusters, so that objects within a cluster have high similarity in comparison one to another but are very dissimilar to objects in other clusters (Han and Kamber, 2006: 381). In literature, a lot of different clustering algorithms are existed. It is difficult to categorize the clustering methods because each category may overlap, so that a method may have features from several categories.

Clustering methods can be classified into the following categories; “Partitioning Method”, “Hierarchical Method”, “Density-based Method”, “Grid-Based Method”, “Model-Based Method” which are also shown in Figure 10.

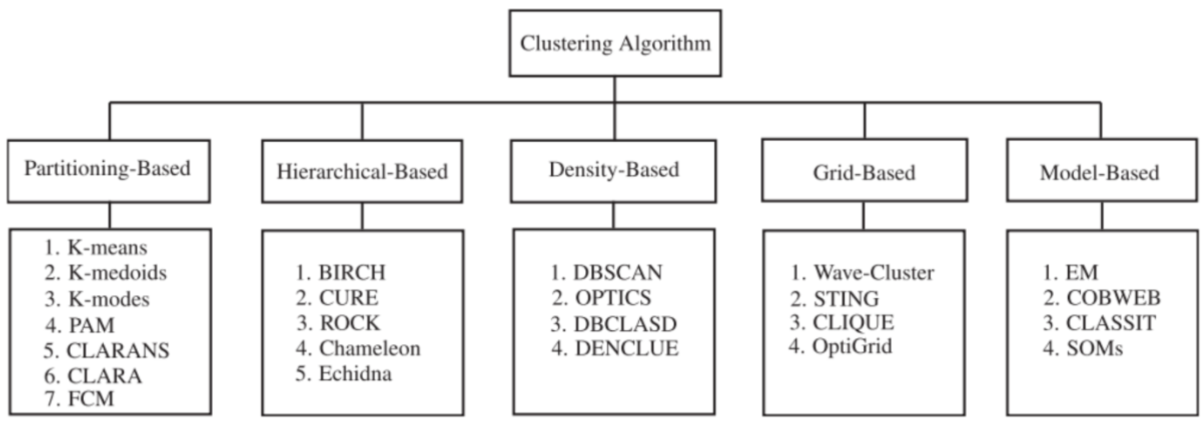



Figure 10: The basic Taxonomy of Clustering Algorithms

Source (Fahad et al, 2014)

The partitioning algorithms divide data objects into a number of partitions, where each partition represents a cluster. These clusters should fulfil the following requirements: (1) each group must contain at least one object, and (2) each object must belong to exactly one group.

Hierarchical-based Data are organized in a hierarchical manner depending on the medium of proximity. Proximities are obtained by the intermediate nodes. Hierarchical clustering methods can be bottom-up or top-down. The bottom-up should starts with one object for each cluster and recursively merges two or more of the most appropriate clusters. The top-down should starts with the dataset as one cluster and recursively splits the most appropriate cluster.

Moreover, in Density-based clustering, data objects are separated are based on their regions of density, connectivity and boundary. In other words, this method is based on the notion of density. The basic idea is to continue growing the given cluster as long as the density in the neighborhood exceeds some threshold.

In grid-based method, the objects together form a grid. The object space is quantized into finite number of cells that form a grid structure.

Lastly, in model-based methods, a model is hypothesized for each cluster to find the best fit of data for a given model. This method locates the clusters by clustering the density function. It reflects spatial distribution of the data points. This method also provides a way to automatically determine the number of clusters based on standard statistics, taking outlier or noise into account. It therefore yields robust clustering methods.

The pattern recognition techniques of clustering and also be applied to image segmentation tasks. In clustering, it's attempted to find natural groupings of the objects (in this case, pixels) in feature space such that pixels in a group are similar to each other but dissimilar to pixels in other groups (Johnson, 1967). By applying clustering techniques, pixels can be grouped from each object together, thus separating them from the background pixels

There are two key issues which must be addressed before applying a clustering algorithm (Mccallum, 2000). The first is the choice of a similarity metric. This essentially indicates how similar two objects are in the feature space. The most obvious choice of a similarity metric is the distance between the two data items. The Euclidean distance is often chosen, though another distance metrics are also possible. Alternatively, one can use a similarity matrix which directly provides the value of the "distance" between two objects. The second issue is the clustering criterion, which is used to evaluate a partitioning of the data items into groups. The task of clustering then becomes one of optimizing this function. In this study, hierarchical methods and partitioning methods are used to determine the clusters.

Selected Clustering Methods

K-Means Cluster

There are many algorithms for clustering. K-means clustering is an exclusive clustering algorithm. Each object is assigned precisely to one of a set of clusters. For this method of clustering, how many clusters would like be to from the data. It's called value k.

K-means (MacQueen, 1967) is one of the simplest unsupervised learning algorithms that solve the well-known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed a priori. The main idea is to define k centroids, one for each cluster. These centroids should be placed in a cunning way because of different location which causes different result. So, the better choice is to place them as much as possible far away from each other.

The next step is to take each point belonging to a given data set and associate it to the nearest centroid. When no point is pending, the first step is completed and an early groupage is done. At this point, there is need to re-calculate k new centroids as barycenter's of the clusters resulting from the previous step. After have these k new centroids, a new binding has to be done between the same data set points and the nearest new centroid. A loop has been generated. As a result of this loop it may be noticed that the k centroids change their location step by step until no more changes are done. In other words, centroids do not move any more. Finally, this algorithm aims at minimizing an objective function, in this case a squared error function. Although it can be proved that the procedure will always terminate, the k-means algorithm does not necessarily find the most optimal configuration, corresponding to the global objective function minimum. The algorithm is also significantly sensitive to the initial randomly selected cluster centers. The k-means algorithm can be run multiple times to reduce this effect (Stauffer and Grimson, 1999).

To shows all those steps item by item; (Bramer, 2007)

1. Choose a value of k.
2. Select k objects in an arbitrary fashion. Use these as the initial set of k centroids.
3. Assign each of the objects to the cluster for which it is nearest to the centroid.
4. Recalculate the centroids of the k clusters.
5. Repeat steps 3 and 4 until the centroids no longer move.

<u>TwoStep Cluster</u>

The TwoStep Cluster node provides a form of cluster analysis (Lin and Gerla, 1997). It can be used to cluster the dataset into distinct groups when you don't know what those groups are at the beginning. As with Kohonen nodes and K-Means nodes, TwoStep Cluster models do not use a target field. Instead of trying to predict an outcome, TwoStep Cluster tries to uncover patterns in the set of input fields. Records are grouped so that records within a group or cluster tend to be similar to each other, but records in different groups are dissimilar.

TwoStep Cluster has 2 diverse stage of clustering method. The first step makes a single pass through the data, during which it compresses the raw input data into a manageable set of subclusters. The second step uses a hierarchical clustering method to progressively merge the subclusters into larger and larger clusters, without requiring another pass through the data. Hierarchical clustering has the advantage of not requiring the number of clusters to be selected ahead of time. Many hierarchical clustering methods start with individual records as starting clusters and merge them recursively to produce ever larger clusters. Though such approaches often break down with large amounts of data, Twostep’s initial preclustering makes hierarchical clustering fast even for large datasets (Kašparová, et al, 2010).

Association Rules and Sequential Patterns

Association rules are created by analyzing data for frequent if/then patterns and using the criteria support and confidence to identify the most important relationships. Support is an indication of how frequently the items appear in the database. Confidence indicates the number of times the if/then statements have been found to be true.

In Association, the relation of a particular item in a data transaction on other items in the same transaction is used to predict patterns. In Classification, the methods are intended for learning different functions that map each item of the selected data into one of a predefined set of classes. Cluster analysis takes ungrouped data and uses automatic techniques to put this data into groups. Prediction analysis is related to regression techniques. The key idea of prediction analysis is to discover the relationship between the dependent and independent variables, the relationship between the independent variables (Hair et al, 1998).

In data mining, association rules are useful for analyzing and predicting customer behavior (Hsieh, 2004). They play an important part in shopping basket data analysis, product clustering, catalog design and store layout. Programmers use association rules to build programs capable of machine learning

In classic sequential pattern mining, no rules are generated. It only finds all frequent patterns. Generalized association rules, application-specific knowledge in the form of hierarchies over items are used to discover more interesting rules, whereas sequential pattern mining utilizes the time associated with the transaction data to find frequently occurring patterns.

Given a set of data-sequences each of which is a list of transactions ordered by the transaction time, the problem of mining sequential patterns is to discover all sequences with a user-specified minimum support. Each transaction contains a set of items. A sequential pattern is an ordered list (sequence) of item sets. The item sets that are contained in the sequence are called elements of the sequence. The support of a sequential pattern is the number of data sequences that contain the sequence (Thomas and Sawaragi, 1998).

Selected Association Methods

Rule induction is one of the most important techniques of machine learning. Since regularities hidden in data are frequently expressed in terms of rules, rule induction is one of the fundamental tools of data mining at the same time (Grzymala-Busse, 1998). The most frequent task of rule induction is to induce a rule set R that is consistent and complete and also R can be called discriminant. In global rule induction algorithms, the search space is the set of all attribute values, while in local rule induction algorithms the search space is the set of attribute-value pairs.

### 1.3.3. Prescriptive Analytics: DEMATEL

Prescriptive analytics is the area of Business Analytics dedicated to discovering the best course of action for a given situation. Prescriptive analytics are related to both descriptive and predictive analytics. While descriptive analytics aim to provide penetration into what has happened and predictive analytics helps model and forecast what might happen, prescriptive analytics seeks to determine the best solution or outcome among various choices, given the known parameters.

The relatively new field of prescriptive analytics allows users to “prescribe” a figure of different possible actions and guide them towards a solution. In a nut-shell, these analytics are all about offering advice. Prescriptive analytics attempt to measure the effect of future decisions in order to advise on possible outcomes before the decisions are actually made. At their best, prescriptive analytics predicts not just what will happen, but also why it will happen providing recommendations regarding actions that will take advantage of the predictions. These analytics go beyond descriptive and predictive analytics by recommending one or more possible courses of natural process. Essentially, they predict multiple futures and allow companies to assess a number of possible outcomes based upon their activities. Prescriptive analytics use a combination of techniques and tools such as business rules, algorithms, machine learning and computational modelling procedures. These techniques are applied against input from many different data sets including historical and transactional data, real-time data feeds, and big data. Most decision making is prescriptive or normative. It is aimed at making the best decision without uncertainties arising. Decision-makers should have the perfect insight and knowledge to take the most rational decision/solution in the end.

Namely, there are two universal types of analysis in Multi Criteria Decision Analysis (MCDA); prescriptive and evaluation of a past decision. Prescriptive analysis involves multi criteria scoring that can be separated by following set of options. The earliest known reference to Multiple Criteria Decision Making can be traced back to Benjamin Franklin (1706 - 1790). At that period, He had a simple paper system for deciding important issues (Pessoa et al, 2015). Since the 1960s, MCDM has been an active research area and produced many theoretical and applied articles and books. (Roy, 2015). The multi-criteria analysis has been an important tool that supports decisions and it is necessary to manage systems. An expressive number of multi-criteria methods and models have been developed, each one with 5specific characteristics and applicable to the various areas of production engineering (Mendoza and Martins, 2006).

Multi-Criteria Decision-Making (MCDM) is for assessing trade-offs between multiple criteria (or objectives) in order to rank, prioritize or choose from among alternatives (Pohekar and Ramachandran, 2004). And also, MCDM can define that is a branch of operational research dealing with finding optimal results in complex scenarios including various indicators, conflicting objectives and criteria (Kumar et al, 2017).

Especially after the millennium, Multiple Criteria Decision-Making (MCDM) has grown as a part of operations research, concerned with designing computational and mathematical tools for supporting the subjective evaluation of performance criteria by decision-makers (Zavadskas et al, 2014). In this study, after the key criteria were found with MDCM, the second survey developed for applying Decision-Making Trial and Evaluation Laboratory (DEMATEL) method was issued to the determine the affected and affecting factors in evaluating the importance of criteria and constructing the causal relations among the criteria. Historical developments in multi-criteria analysis is shown in Table 2.

Table 2: Historical Developments in Multi-Criteria Analysis

| Authors | Period | Development |
|---|---|---|
| **Kuhn and Tucker** | 1951 | Formulated optimality conditions for nonlinear programming and considered problems with multiple objectives. |
| **Charnes, Cooper and Ferguson** | 1955 | Contained the essence of goal programming and 'goal programming' was first used. |
| **Ron Howard and G.E. Kimball** | 1959 | Used the term "decision analysis" for the first time. |
| **Bernard Roy** | 1960's | Developed the ELECTRE´s methods which is a family of Multi-Criteria Decision Analysis methods. |
| **Bruno Contini and Stan Zionts** | 1968 | A multiple-criteria negotiating model was developed |
| **Howard Raiffa** | 1968 | Involved in decision analysis early on, and work an important |
| **Thomas Saaty** | 1970's | Introduced the Analytic Hierarchy Process (AHP) and the Analytic Network Process (ANP). |
| **Gabus and Fontela** | 1972 | Decision-Making Trial and Evaluation Laboratory (DEMATEL) technique was employed |
| **Zionts and Jyrki Wallenius** | 1973 | Developed the Zionts-Wallenius interactive method for solving multiple-objective linear programming problems. |
| **Ralph Keeney and Howard Raiffa** | 1976 | Instrumental in establishing the theory of multi-attribute value theory (including utility theory) as a discipline. |

Source: (Pessoa et al. 2015)

Selected MDCM methods

The DEMATEL method is used to construct interrelations between criteria/factors (Fontela and Gabus, 1974, 1976) and to find the central criteria to represent the effectiveness of factors/aspects. It has been successfully applied in many areas, such as marketing strategies, control systems, safety problems, developing the competencies of global managers and group decision-making (Wu and Lee, 2007).

There are four steps in the DEMATEL method:

1. Calculate the average matrix.
2. Calculate the normalized initial direct-influence matrix
3. Derive the total relation matrix,
4. Set a threshold value and obtain the impact-relations map.

These steps completed as well as adequate information for further analysis and decision-making. The traditional method followed to set a threshold value is conducting discussions with experts. The results of the threshold values may differ among different researchers (Li and Tzeng, 2009).

At first the initial average matrix of pair-wise comparisons from experts is calculated. Firstly, respondents are asked to indicate the degree of direct influence each factor/element i exerts on each factor/element j, which is denoted by a(ij). We assume that the scales 0, 1, 2, 3 and 4 represent the range from "non-influence" to "very high influence". Each respondent would produce a direct matrix, and an average matrix "A" is then derived through the mean of the same factors/elements in the various direct matrices of the respondents. In the second step the initial influence matrix is calculated. The initial direct influence matrix X(X={x(ij)}) can be obtained by normalizing the average matrix A by following formulas X=λ×A. Then in the third step the full direct/indirect influence matrix T is derived using following formula: $T = X(1 - X)^{-1}$

Then each row sum and column sum of matrix T is calculated. Where r denotes the row sum of the i th row of matrix T and shows the sum of direct and indirect effects of factor/element i on the other factors/elements. Similarly, j c denotes the column sum of the j th column of matrix T and shows the sum of direct and indirect effects that factor/element j has received from the other factors/criteria. In addition, when i = j (i.e., the sum of the row and column aggregates) (r + c) provides an index of the strength of influences given and received, that is, (r + c) shows the degree of the central role that factor i plays in the problem. If ( r − c) is positive, then factor i is affecting other factors, and if (r − c) is negative, then factor is being influenced by other factors (Tzeng et al, 2007).

In this paper, passenger baggage values and flights distance has been chosen like a decision attribute. Additionally, Passenger Number, Baggage Number, Baggage Weight, Economy Class Rate, Infant-Child Passenger Number Rate, Flights Date (Time), Distance (Km), Region have also picked like a condition attributes. The detailed methodology and analysis process and result will be shown above in the Methodology and Results parts.

## 1.4. Business Intelligence in Civil Aviation: Empirical Studies

Nowadays, it is very hard to define the scope of the Business intelligence studies. Especially after the 2010 when is BI terms started to popular. Table 3 shows that the remarkable empirical papers about BI in civil aviation fields since 2001. They have categorized 6 different scope which are Efficiency, Airports, Airlines Market, Passenger, Cargo and Safety.

Table 3: Empirical Papers about BI in Civil Aviation Fields

| Topics | Titles | Number of Writer | P. Year | Used Technique | Analytics Type |
|---|---|---|---|---|---|
| Efficiency | A comparative performance analysis of airline strategic alliances using data envelopment analysis | 2 | 2016 | Data Envelopment Analysis | Predictive |
| Airports | A decision rules approach for improvement of airport service quality | 4 | 2011 | Data Driven Decision Making | Descriptive |
| Airlines Market | A factor-analytic generalized nested logit model for determining market position of airlines | 3 | 2014 | Factor Analysis | Descriptive |
| Passenger | A fast airplane boarding strategy using online seat assignment based on passenger classification | 6 | 2016 | Classification | Predictive |
| Cargo | A study of the competitiveness of airline cargo services departing from Korea: Focusing on the main export routes | 2 | 2015 | Decision Tree | Predictive |
| Efficiency | Aircraft taxi time prediction: Comparisons and insights | 5 | 2014 | Regression Analysis | Descriptive |

| | | | | | |
|---|---|---|---|---|---|
| Efficiency | Airline Route Profitability Analysis and Optimization Using BIG DATA Analytics on Aviation Data Sets under Heuristic Techniques | 4 | 2016 | Heuristic | Predictive |
| Airlines Market | An evaluation of the world's major airlines' technical and environmental performance | 2 | 2014 | Data Envelopment Analysis | Predictive |
| Cargo | Analyzing China's air cargo flows and data | 3 | 2004 | Factor Analysis | Descriptive |
| Cargo | Analyzing competition of international air cargo carriers in the Asian general air cargo markets | 2 | 2013 | Sensitivity Analysis | Predictive |
| Safety | Anomaly detection via a Gaussian Mixture Model for flight operation and safety monitoring | 4 | 2016 | Gaussian Mixture Model | Predictive |
| Airlines Market | Cities and air services: the influence of the airline industry | 2 | 2012 | Classification | Predictive |
| Cargo | Classification and competition analysis of air cargo logistics providers: The case of Taiwan's high-technology industry | 3 | 2011 | Classification | Predictive |
| Efficiency | Classification of jet fuels by fuzzy rule-building expert systems applied to three-way data by fast gas chromatography—fast scanning quadrupole ion trap mass spectrometry | 5 | 2011 | Classification | Predictive |
| Airlines Market | Classification rule discovery for the aviation incidents resulted in fatality | 3 | 2009 | Classification | Predictive |
| Airlines Market | Data mining and preprocessing application on component reports of an airline company in Turkey | 3 | 2011 | Regression Analysis | Predictive |
| Airports | Data mining modeling on the environmental impact of airport deicing activities | 7 | 2011 | Decision Tree | Predictive |
| Airlines Market | Data-driven Modeling of Airlines Pricing | 5 | 2015 | Data Driven Decision Making | Descriptive |
| Passenger | Entrepreneurial orientation as a basis for classification within a service industry: the case of retail pharmacy industry | 3 | 2005 | Classification | Predictive |
| Airlines | Evaluation of deregulated airline networks using data envelopment analysis combined with principal component analysis with an application to Western Europe | 2 | 2001 | Factor Analysis | Descriptive |
| Airlines | Improving the IATA delay data coding system for enhanced data analytics | 2 | 2014 | Data Envelopment Analysis | Predictive |
| Airlines Market | Market clustering and performance of U.S. OD markets | 4 | 2013 | Clustering | Predictive |
| Airports | Measuring airport quality from the airlines' viewpoint: an application of data envelopment analysis | 2 | 2001 | Data Envelopment Analysis | Predictive |
| Safety | Natural language processing for aviation safety reports: From classification to interactive analysis | 5 | 2016 | Classification | Predictive |
| Passenger | Optimizing airline passenger prescreening systems with Bayesian decision models | 2 | 2012 | Decision Tree | Predictive |
| Airports | Performance based clustering for benchmarking of US airports | 2 | 2004 | Clustering | Predictive |
| Efficiency | Prediction of Aircraft Flight Duration | 4 | 2006 | Clustering | Predictive |
| Efficiency | Route-based performance evaluation of Taiwanese domestic airlines using data envelopment analysis | 2 | 2006 | Data Envelopment Analysis | Predictive |
| Efficiency | Technical efficiency of mainstream airlines and low-cost carriers: New evidence using bootstrap data envelopment analysis truncated regression | 2 | 2014 | Data Envelopment Analysis | Descriptive |
| Passenger | Using decision rules to achieve mass customization of airline services | 3 | 2010 | Data Driven Decision | Descriptive |

Moreover, these studies and more will show and examine below titles in detail.

### 1.4.1. Descriptive Empirical Studies

Before start the BI in air transportation sector; first want to emphasis that, one of the most crucial sectors in 20th century was aviation sector in military and also commercial field. Total number of passengers carried on scheduled services reached 3.7 billion in 2016, a 6.0 per cent increase over last year (Philbin, 2016). Moreover, the air transportation industry is one of the first adopters of data science with amounts of huge data in their systems. Over the years, this industry has been making large investments to mine data and explore opportunities to improve operational efficiency and boost customer loyalty. In Obviously, the nature of data is critical to the success of data mining application. Therefore, the nature of the data is related to its source, utility, behavior and description. For the very reason aviation has most robust and pristine data. For why, aircraft always record data down in their black box. Finding patterns in aviation data manually is impracticable due to the mass amount of data produced every day (Pagels, 2015). Moreover, planes equipped with flight recording data typically record up to 500 variables of data -described in these flight data recordings are time, altitude, vertical acceleration, and heading- per second for the duration the plane is being operated (Das et al, 2010). Finding patterns in aviation data manually is impracticable due to the mass amount of data produced every day (Pagels, 2015). Airline companies' domain to better make sense of current and historical data, and make predictions using descriptive behavior, a scalable analytics service because of their real needed (Ayhan et al, 2013).

Civil aviation sector has large amounts of data are generated during studies on flights information, passengers' data and analysis, cost expenditure analysis mainly fuels prices and employee cost, airspace of countries and other political relations. And also, there are some other circumstances such us; measurements of wastes and damage to nature, technical affairs and mechanical matters like traffic signals and radar issues, aircraft interior design, advertising and other managerial issues and inventory etc. Based on all these data, decision-makers who are managers arrive at decision to solve a respective problem and optimize their resource. Managers should look out for ways to ease the pain in obtaining access to and applying different datasets.

About the baggage handling system, in order to increase security and fertility new baggage handling methods simulated and optimized (Rijsenbrij and Ottjes, 2008). On the passenger side, for minimizing total waiting time and better utilization of facilities, model was developed and optimized check-in time (Hsu et al, 2012). Moreover, there are also some lateral studies in aviation sector, like; airport security queue optimization (Lee and Jacobson, 2011) or other security operations (Nie, 2011).

On the other hand, Business intelligence is indispensable for the airline companies as well as airport, civil aviation institutes, flight producer and also other transportation and tourism directorates. Because BI tools provide data including agency sales volume, destination sales loyalty, ticket counts as well as the agency's contact information and helps tourism boards and destination marketing organizations identify the agencies most valuable to their destination. Moreover, with different tools passenger satisfaction benchmarking survey may design for airlines. Track and compare their customer satisfaction ratings with that of competitors. Those can cover all travel service aspects of the pre-flight, in-flight and post-flight passenger travel experience (Jacobson et al, 2003).

### 1.4.2. Predictive Empirical Studies

Nowadays, air transportation companies may allow operational teams to take action to minimize the impact about an approaching storm or even queues forming at the security point. The next footstep will be to utilize Business intelligence tools, analyze past event and combine live data feeds from multiple sources, both internal to the organization and external, to predict future events and take preventive action before they occur (Leitner et al, 2010).

Data can improve ground operations, supporting faster turnaround times and better airspace management solutions which are driving efficiencies. Also, it gives airlines a critical look at passengers to enable better and more personalized experience for each passenger, which in turn drives brand loyalty, increases customer satisfaction, enables stronger auxiliary revenue stream and finally supports scheduling/rebooking of passengers when delays happen. Due to both economic and social reasons, several journal articles were examined about aviation sector (Moorthy et al, 2005). However, it has focused Data Mining tools and applications related air transportation.

Actually, the lots of studies regarding aircraft production and some engineering issues about aviation in particular till 2000. Its inspired just with empirical studies and should be real-world data. And also, perusing articles related to the main purpose of using data mining techniques; the tools or the analysis should be included all echelon of data mining and analytics such as, integration, cleaning, transformation, pattern evaluation and knowledge (Delen, 2004). Moreover, distinguished articles considering air transformation issues except human automation interaction of pilots or stewards, employees survey works, accident and aircraft production. Exclusively data mining and analytics about airlines, airports, cargo, passenger, efficiency and safety. It can be said that the majority of the well-known science and social science journals were searched about civil aviation business analytics or data mining (Akpinar and Karabacak, 2017).

As it mentioned before, the earliest studies about data Mining and aviation industries begin after the world war two but most of ripe articles published after millennium. However, Gudmundsson wrote his paper before 2000, and it's refer the new entrant airlines life cycle analysis using with classification techniques (Gudmundsson, 1998). Thanks to some of data mining techniques, oil expenses can be reduced as much as possible. To illustrate the last example, aircrafts ordinarily do not fly with full tank, because it causes really high damage in terms of cost. However, in the event of using data mining techniques, fuel cost optimization precisely, taking into consideration weather conditions, passenger numbers, cargo weights, next flight, airport situation and whole lot more. One of the most studies in this field is about airlines marketing and performance analysis. Arjomandi and Seufert evaluate the major airlines performances based on technical and environmental used with data envelopment analysis (DEA) techniques. They also classify the different regions and types of airlines according to IATA and efficiencies of the airlines. Their results are empirical and they advocate that low-cost airlines more environmentally oriented (Seufert and Arjomandi, 2014). The other remarkable article was written by Cosmas et al. and aforementioned article is about clustering airlines performance in US market (Cosmas et al, 2013).

One of the widespread subject in this space is airport consideration. Adler and Joseph has measured airports using DEA and also Sarkis et al. clustered airports depends on the passenger performance (Adler and Joseph, 2001) (Sarkis et al, 2004). In addition, concerning air cargo the researchers studied air cargo analysis (Hui et al, 2004), cargo competition analysis (Wen et al, 2011) (Yoon and Park, 2015) (Shiao and Hwang, 2013) and estimation air cargo issues (Lo et al, 2015).

Eventually, one of the most crucial topics of these contexts is a client of the airlines are passengers (Park et al, 2004). Through a combination of the historically generated customer information, plus the current real-time information being gathered, specific and personalized advertising or programming can be offered from airline partners; thereby increasing ancillary revenues and improving customer satisfaction. And also, with historical customer knowledge and data regarding individual passengers' day-of-travel plans, an airline can offer customized price and once again improve customer satisfaction.

At the present time connected digital data is transforming the aviation industry including airlines and all other stakeholders in the aviation ecosystem. Regarding this matter, Liou has more than 5 different studies, different titles and several writers is about CRM of airlines (Liou, 2009), airline customer behavior (Liou and Tzeng,

2010) and variable consistency airline service strategies (Liou, 2011). Moreover, there are studies about passenger optimization (Majeske and Lauer, 2012) and about passenger classifications (Notomista et al, 2016).

### 1.4.3. Prescriptive Empirical Studies

In the future, the combination of Business intelligence, plus predictive analysis will help airlines better manage available resources to improve the passenger experience. It will also help airports to increase optimization of their infrastructure and space to improve services and revenues. Better still, it will allow the industry to add actual value for their passengers and their businesses by running scenarios and building the evidence, either to make process changes or introduce new products and services. Right now, forward-thinking airlines and airports are focused on becoming much more proactive and making interactions with passengers more relevant. Some of these initiatives offer a glimpse of the future prize ahead (SITA, 2014).Considering the complexities around the logistics of running an airline and the tough competition, the ones that exploit sophisticated analytical tools are the ones who would ultimately enjoy a competitive edge over the rest (Porter and Millar, 1985).

About the MCDM application in aviation, looking at the DEMATEL studies in air transportation sector, there are different kind of papers which they mentioned diverse points. To demonstrate, for building effective safety management systems, Liou has shown that fuzzy DEMATEL can be useful in visualizing the structural relations and identifying key factors in a complex system such as a safety management systems for airlines (Liou, Yen and Tzeng, 2008). And also about the evaluation of airline service quality, there are some survey studies include customer perceptions with DEMATEL techniques (Wang, Lin andTseng, 2011).

Using the DEMATEL it’s also possible to derive a hybrid multi-criteria model, with selection outsourcing providers criteria (Liou and Chuang, 2010) Liou also has another study about the develop an integrated model in order to the selection of strategic alliance partners in the civil aviation industry (Liou, 2012). Furthermore, DEMATEL techniques also cab merge marketing and other business models which is based on the perspectives of ‘‘marketing mix 4Ps’’ and ‘‘website quality’’. That can be used to analyze the relationship among criteria, and then the Analytic Network Process (ANP) is applied to compute the weight of each criterion, finally that can be modified VIse Kriterijumska Optimizacija i Kompromisno Resenje (VIKOR) method is used to rank the performance and they analyzed the websites of five air transportation companies. All those researches which are empirical studies about MCDM in civil aviation field and more is shown below in Table 4.

Table 4: Empirical Papers about MCDM in Civil Aviation Fields

| TOPICS | TITLES | NUMBER WRITERS | YEAR | USED TECHNIQUES |
|---|---|---|---|---|
| Airport | A hybrid fuzzy MCDM approach for mitigating airport congestion: A case in Ninoy Aquino International Airport | 2 | 2017 | DEMATEL<br>Analytic Network Process |
| Performance Evaluation | A hybrid type-2 fuzzy based supplier performance evaluation methodology: The Turkish Airlines technic case | 4 | 2017 | Analytic Hierarchy Process |
| Aircraft | A model for aircraft evaluation to support strategic decisions | 3 | 2015 | Decision Making Process<br>Analytic Hierarchy Process |
| Aircraft | A Multi-Criteria Decision-Making Scheme for Multi-Aircraft Conflict Resolution | 4 | 2017 | TOPSIS |
| Environmental Management | A multi-criteria decision support methodology for evaluating airport expansion plans | 3 | 2002 | Analytic Hierarchy Process |
| Performance Evaluation | A new hybrid simulation-based assignment approach for evaluating airlines with multiple service quality criteria | 6 | 2017 | TOPSIS |
| Environmental Management | A proposed Multi Criteria Analysis decision support tool for international environmental policy issues: a pilot application to emissions control in the international aviation sector | 3 | 2007 | Decision Support Systems |
| Route Analysis | Airline new route selection based on interval type-2 fuzzy MCDM: A case study of new route between Turkey- North American region destinations | 4 | 2017 | Type 2 Fuzzy |
| Performance Evaluation | An analysis of African airlines efficiency with two-stage TOPSIS and neural networks | 3 | 2015 | TOPSIS |
| Aircraft | An approach to operational aircraft maintenance planning | 5 | 2010 | Decision Support Systems |
| Environmental Management | An MCDM approach for selecting green aviation fleet program management strategies under multi-resource limitations | 5 | 2017 | DEMATEL<br>Analytic Hierarchy Process |
| Performance Evaluation | An outsourcing provider decision model for the airline industry | 3 | 2013 | DEMATEL<br>Analytic Hierarchy Process |
| Airport | Analytic Hierarchy Process assessment for potential multi-airport systems – The case of Cape Town | 2 | 2014 | Analytic Hierarchy Process |
| Performance Evaluation | Applying FMCDM to evaluate financial performance of domestic airlines in Taiwan | 1 | 2008 | Fuzzy Grey Relational Analysis |
| Performance Evaluation | Balanced scorecard based performance measurement of European airlines using a hybrid multi-criteria decision-making approach under the fuzzy environment | 3 | 2017 | DEMATEL |
| Performance Evaluation | Developing a hybrid multi-criteria model for selection of outsourcing providers | 3 | 2017 | DEMATEL<br>Analytic Hierarchy Process |
| Performance Evaluation | Development of a fuzzy ANP based SWOT analysis for the airline industry in Turkey | 7 | 2012 | Analytic Hierarchy Process |

| | | | | |
|---|---|---|---|---|
| Performance Evaluation | Evaluating airline service quality using a combined fuzzy decision-making approach | 1 | 2017 | DEMATEL |
| Airport | Evaluating the service quality of airports in Thailand using fuzzy multi-criteria decision-making method | 1 | 2016 | Decision Support Systems |
| Environmental Management | Low-carbon airline fleet assignment: A compromise approach | 3 | 2017 | Decision Making Process |
| Fuels | Multi-attribute sustainability evaluation of alternative aviation fuels based on fuzzy ANP and fuzzy grey relational analysis | 2 | 2015 | Fuzzy Grey Relational Analysis<br>Analytic Hierarchy Process |
| Cargo | Multi-criteria decision-making for complex bundling configurations in surface transportation of air freight | 3 | 2017 | Best Worst Method (BWM) |
| Performance Evaluation | Ranking of aircraft maintenance organization based on human factor performance | 2 | 2015 | Analytic Hierarchy Process |
| Performance Evaluation | Taking-off corporate social responsibility programs: An AHP application in airline industry | 2 | 2017 | Analytic Hierarchy Process |
| Fuels | Three-stage airline fleet planning model | 2 | 2015 | Decision Making Process |
| Performance Evaluation | Transshipment hub selection from a shipper's and freight forwarder's perspective | 4 | 2017 | Analytic Hierarchy Process |
| Airport | Unmanned Aerial Vehicle hub-location and routing for monitoring geographic borders | 2 | 2015 | Decision Support Systems |
| Performance Evaluation | Using decision rules to achieve mass customization of airline services | 4 | 2010 | Decision Support Systems |

DEMATEL also has used green aviation fleet management (Lee et al, 2017) mitigating airport congestion (Bongo and Ocampo, 2017), comparative scorecard based performance measurement of some airlines (Dinçer et al, 2017), evaluating airline service quality (Perçin, 2017) and also some aviation staff evaluating. For instance, they approach in order to air traffic controllers' workload stress problems in Philippines (Bongo et al, 2017).

## 1.5. Gaps in The Field and Focus of This Study

The civil aviation industry has a collective vision to make air travel a more personal and relevant experience for passengers in the near future. Business intelligence and data mining are at the forefront of data analysis work (Minelli, 2012). As in many sectors, it also increases productivity and profitability in the civil aviation sector. Large quantities of data, especially those held by commercial airlines, require the use of Business intelligence techniques (Gürbüz et al, 2011). To achieve the target aviation sector, pull out all the stop and they try to improve not only aircraft or airports, but also their systems, employee and all services. Air transport industry has tiger economy after second half the 20th century thanks to the keeping up pace with the technological advancements. And there are lots of studies regarding to the Business intelligence on civil aviation to illustrate efficiency, marketing analysis, airport or airlines flights optimizations, crew and employee preparing program, also passenger and cargo.

Regarding to civil aviation industry, there are studies about the service quality (Liou, Yen and Tzeng, 2010), reportage (Gürbüz et al, 2011), security (Tanguy et al, 2016) and delay researches (Wu and Truong, 2014) In addition, Hui and friends have already studied not about cargo planning but also air cargo analysis in china aviation sector. They mentioned network of the air cargo and mapped air cargo transformation data. Wu and Truong (2014) researched detailed flights delay and they focused delay time, delay reasons. Moreover, they classified them concerning the delaying code and put forward particular schemes. On the other study its analyzed text mining about air transportation security procedure and used phased, significant result has been demonstrated (Tanguy et al, 2016). Another noteworthy study is at the regard flight planning, and Sun (2015) made predictions and formulated flight schedule and clustered their result put out the terminations.

Under the cargo title, studies regarding air cargo analysis (Hui et al, 2004), optimization (Kasturi et al. 2016), competitiveness (Yoon and Park, 2015) classification (Wen et al. 2011) and assessment (Shiao and Hwang, 2013). In this headline, all of these articles relevant Asian Airlines which are China, Hong Kong, Taiwan and Korea.

When looking at the articles written in this field in Turkey, the data analysis and mining studies regarding to civil aviation has limited. In this context Peker and Baki (2009) analyzed airports efficiencies which has include Turkey airports passenger and flights number and they classificated that and remarked the situation. Another article about in this topic has written Avcı and Aktas (2015) mentioned airport performances with enveloping techniques.

Aslan and Yılmaz (2010) study about decision making systems and they touch on air transportation decision making systems. The other studies more about the industrial and aircraft parts issues, like Gürbüz et al (2009) mentioned data mining on aircraft disassembly reports and Yurtay et al (2014) made container optimizations using some techniques but not very about the airlines it's also related to sea freight.

As it found out, in the context of Business intelligence, there are only a little number of air transportation sector studies and the inadequacy of the work done in this area. In addition, although scientific research and studies related to reducing the weight of the aircraft or fuel planning have been carried out, studies on passenger baggage have been limited to the physical and design areas, there is not much evaluation has been made on the available real data and findings, and no estimates have been delivered. In order to fill in this gap, in this study its examined air transport passenger baggage and their profile classified some in meaningful rules. Consequences may irradiate marketing and sales strategies and make a huge contribution to improving all aviation sector.

# METHODOLOGY

The Figure 11 below is a brief representation of the research design used in this thesis. The analytics on application approach was followed in order to fulfill the purpose of exploring the innovation dynamics and gathering convenient and meaningful rules for civil aviation industry.

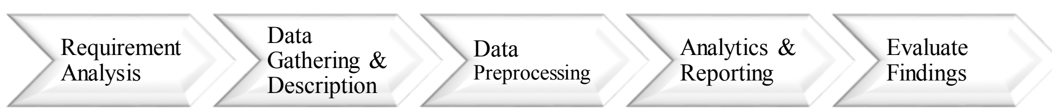



Figure 11: Methodology of the Study

Before starting to study on this, the situation assessment and needs investigation were carried with out in consultation with authorized directors. After that, data has taken from Turkish Airlines which is one of the biggest 10 airlines in terms of the number of passengers. Although this airline is not the most points in the globe, but it is the airline that flies to most different countries. Because of this, they have very different passenger and baggage profiles. Dataset is consisted of 855.250 different flights and has included 115.629.955 passengers from 2014 to 2015. The time phrase and domestic flights knowledge, international and local pass passenger information take part in dataset that picked out the flights which are departure from Istanbul.

Following, in data preparing phase, international flights were selected as the most likely to be able to explain the pattern of operation therefore only left about 230.000 rows. Lastly, some of them was deleted because of validation and incompletion of the data as it mentioned before and eventually has been left 229.465 flights which are from Istanbul to international point in during in 2014 and 2015. This flight has different types of attributes such as; time, location, number of the total passenger, number of the luggage, business and infant passenger rates and the distance of arrival airports. The dataset consists of 21 different attributes and more than two hundred flights were questioned on their operational feature.

The aim of this study is to follow from descriptive to prescriptive analytical process in order to improve usable result examples in air transportation. In descriptive analytics phase, descriptive statistics are employed to describe the basic features of the data in a study and help the systems what happen in the past (Trochim and Donnelly, 2001).

Second phase is predictive analytics, which tries to give a recommendation for key decisions based on future outcomes and focuses on answering the question: “What is probably going to happen in the future?”. Predictive analytics provides organizations with actionable insights based on data and estimation regarding the likelihood of a future outcome via a variety of techniques, such as machine learning, data mining, modelling and game theory (Dhar, 2013). Firstly, TwoStep clustering technique is used to the determined the cluster number. Secondly, K means is one of the simplest algorithm which uses unsupervised learning method to solve all known clustering issues (Krishna and Murty, 1999; Jain, 2010). It is appropriate for large datasets and it has strong sensitivity to any outliers (Huang, 1998; Zhao, Ma and He 2009). Lastly, about the association rules, the more then 75 % of the results have taken notice.

Prescriptive analytics could have a huge effect on all business and how determinations are made and it can impact any industry, organization and systems and help them become more effective and efficient (Delen and Demirkan, 2013). To demonstrate, prescriptive analytics can optimize your scheduling, production, inventory and supply chain design to deliver the right products in the right amount in the most optimized way for the right customers on time. Taking the results from the descriptive and prescriptive analytics throughputs, attributes have been assessed in order to interview with sectoral experts. For the prescriptive analytics phase, DEMATEL method has been developed interview questionnaire and also results reliability and their validity.

## 3.1. Data Source

In this paper, for the analysis and result, different kind of data was used. Firstly, the data was received secondarily from THY. That data has been already collected by and readily available from other sources. Namely, secondary data is research data that has previously been gathered and can be accessed by researchers. The term contrasts with primary data, which is data collected directly from its source. That data cover from 2014 to 2015 all passenger information.

Before the making any analysis and transactions the data consist of 855.250 different flights which has include 115.629.955 passengers, 3.661 different flight routes, 294 different departure airports, 104.419.219 pieces and approximately 1,555 billion kg luggage. Moreover, it has derived new columns and new knowledge with different method from main data. More detailed information about the data will give on the next part and also first part of the results.

The summary data view is shown in the Table 5 below which has include 8 flights and their 24 different attributes. To explain some of abbreviations in the table that, Bag meaning is passenger baggage, Av meaning is average, BagW is baggage weight, Pax meaning is passenger, Y-Cls meaning economy class passenger, D is departure and Arr is arrival, lastly Int meaning is international.

Table 5:Summary of Data View

| No | Bag | Av Bag | Bag W | Av BagW | Y-Cls Pax | Y rate | C-Cls Pax | Infant Pax | Infant rate | Total Pax | Chld Pax | Chld Rate |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 8 | 0.888 | 139 | 15.4 | 9.0 | 1.0 | 0.0 | 0.0 | 0.0 | 9.0 | 0.0 | 0.0 |
| **2** | 82 | 1.07 | 1575 | 20.7 | 72.0 | 0.9 | 4.0 | 3.0 | 0.0 | 76.0 | 3.0 | 0.0 |
| **3** | 92 | 1.31 | 1674 | 23.9 | 65.0 | 0.9 | 5.0 | 3.0 | 0.0 | 70.0 | 2.0 | 0.0 |
| **4** | 107 | 1.32 | 1648 | 20.3 | 80.0 | 1.0 | 1.0 | 3.0 | 0.0 | 81.0 | 3.0 | 0.0 |
| **5** | 19 | 0.95 | 275 | 13.8 | 20.0 | 1.0 | 0.0 | 0.0 | 0.0 | 20.0 | 0.0 | 0.0 |
| **6** | 21 | 1.167 | 306 | 17.0 | 18.0 | 1.0 | 0.0 | 1.0 | 0.1 | 18.0 | 1.0 | 0.1 |
| **7** | 72 | 1.125 | 1126 | 17.6 | 60.0 | 0.9 | 4.0 | 1.0 | 0.0 | 64.0 | 7.0 | 0.1 |
| **8** | 43 | 1.23 | 795 | 22.7 | 29.0 | 0.8 | 6.0 | 0.0 | 0.0 | 35.0 | 2.0 | 0.1 |
| | **Date** | **Flight no** | **Target** | **D.Code** | **D.Country** | **D.Region** | **D.Continent** | **Arr Code** | **Arr.Country** | **Arr.Region** | **Arr.Continent** | **Int/Dom** |
| **1** | 01.01.2015 | TK0475 | SVX | IST | TR | TUR | TURKEY | SVX | RUSSIAN FEDERATION | North Europe | Europe | Int |
| **2** | 01.01.2015 | TK0386 | TBS | IST | TR | TUR | TURKEY | TBS | GEORGIA | Eastern Europe | Europe | Int |
| **3** | 01.01.2015 | TK0806 | ISU | IST | TR | TUR | TURKEY | ISU | IRAQ | Middle East | Asia | int |
| **4** | 01.01.2015 | TK0882 | TBZ | IST | TR | TUR | TURKEY | TBZ | IRAN | Middle East | Asia | Int |
| **5** | 01.01.2015 | TK0338 | GYD | IST | TR | TUR | TURKEY | GYD | AZERBAIJAN | Eastern Europe | Europe | imt |
| **6** | 01.01.2015 | TK0446 | UFA | IST | TR | TUR | TURKEY | UFA | RUSSIAN FEDERATION | North Europe | Europe | int |
| **7** | 01.01.2015 | TK0804 | EBL | IST | TR | TUR | TURKEY | EBL | IRAQ | Middle East | Asia | int |
| **8** | 01.01.2015 | TK0067 | IST | SIN | SINGAPORE | far east Asia | Asia | IST | TR | TUR | TURKEY | int |

There are also time phrase and domestic flights knowledge, international and local pass passenger information which are not shown on this table. In addition, data covered passenger classes like Y means the economy and C meaning is the business. Furthermore, another value are average passenger baggage numbers and theirs weight those are derived by own calculations. In addition, it is calculated Y and C class passenger rate on each flight and infant and child rate also. Moreover, it is assigned IATA airport code to each flight for arrival airport and it also defines their country, region, continent and distance from the departure airport. There is flag input, which is connected or direct separation distinction.

## 3.2. Data Preprocessing for Analytics

Basically, data preprocessing is the process include; collecting, cleaning, and consolidating data, primarily for use in different types of analytics (Azoff, 1993; Famili et al, 1997; Ye, 2003; Bazeley and Jackson, 2013). There are most of techniques in order to prepare to data. For validate data, necessary to identify and remove anomalies and inconsistencies. And also, data integration and transformation, to improve the accuracy and efficiency of learning algorithms. Lastly data size reduction to get the dataset with a lower number of attributes and records but which is as informative as the original dataset. In this study, only the techniques used will be shown below and it can be started with the validation of the data on next title.

### 3.2.1. Data Validation

In this study, it has been received the data from airline companies, so it's the one of the most reliable data because airlines enterprise software systems integrated international aviation institutes for safety and security. Therefore, there aren't any data were not recorded at the source in a systematic way and they were whole available when the transactions associated with a record took place. Moreover, there is no malfunctioning recording devices so it's not possible to some data were deliberately removed. Attention was paid in order to avoid mistake that transferring the data from the operational databases to the platform used for the analyzes.

**Noise**

Data may contain abnormal values known to be outliers. Actually, the data has some outlier because of the compressive all flights but they were removed after the checking flights codes because it's taking into account only scheduled flights. It is also erased heterogeneous flights which are include less than 10 passengers, because most of them were not the normal flights. (Bratu and Barnhart, 2005)

**Inconsistency**

Inconsistency meaning is that data contain discrepancies due because of changes in the coding system used for their representation. In the data involve IATA codes and international standard naming. It's also used in the analysis the same codes to reduce possibility of the making mistake.

### 3.2.2. Complete Data Integration

Data integration includes inspection, elimination identification, and substitution. However, inspection suffers from a high degree of arbitrariness, and is rather burdensome and time-consuming for large datasets. On the other hand, used data for analytics in this thesis has taken from reliable airlines company and they had already had an inspection by some regulatory institutions. Because of that inspected techniques didn’t use for the data preprocessing.

**Elimination**

Systematic elimination of records could be ineffective when the distribution of missing values varies in an irregular way across the different attributes. However, it can the risk of incurring a substantial loss of information. Before analyze the data, its eliminated some flights which has some missing attributes, however may count them on the fingers of one hand. To illustrate, they were 41 different flights has no passenger and 112 flights have not baggage information, which is less than 0.001 per cent of whole flights.

**Identification**

The conventional value might be used to encode and identify missing values in order to remove entire records from the given dataset. To demonstrate there is no information for connection flights but it has been identified for non-direct flights.

**Substitution**

Some criteria exist for the automatic replacement of missing data, although most of them appear somehow arbitrary. In the study, just adding some missing distance values and some region values which are identified company own website.

### 3.2.3. Data Reduction

When dealing with a small data set, the transformations described above are usually adequate to prepare input data for analytics. However, when confronting a large data set it is also appropriate to reduce its size, in order to make learning algorithms more efficient, without sacrificing the quality of the results obtained. There are three primary criteria to determine whether a data reduction technique should be used: efficiency, accuracy and simplicity of the models generated. However, only efficiency and accuracy was being used for this study. Since it is difficult to develop a data reduction technique that represents the optimal solution for all the criteria described (Jain et al, 1999). Lastly, because of the political and terror issues at 2015, it did not used high or low season as the analyses were performed monthly basis on the time phase.

**Efficiency**

Especially the predictive analytics process it is customary to run several alternative learning algorithms in order to distinguish the most accurate model. Thus, a reduction in processing times allows the analyses to be carried out more rapidly. Because of that, it has been reduced the data and it comprehend only two-year data which are 2014 and 2015. Because the firm is very fast-growing company and they added new flight point and also aircrafts mount by mount especially in last decade. Therefore, it was just chosen that years which are the one of the best stationary year for most of the analysis, because these years were the constant time, and there were no extreme events.

**Accuracy**

The accuracy of the models generated represents a critical success factor, and it is so the main criterion followed in order to select one class of learning methods over another. As a result, reduction techniques shouldn't importantly compromise the accuracy of whole model generated. The data taken from some special flights tracking systems which is also responsible for IATA and checked by them. Therefore, really few wrong faulty entries obtained which is about two dozen and it's less than 0,001 % of all the rows.

On the other hand, 279 flights which are have less then 10 passengers and 239 flights which are have less then 10 luggage number deleted, because those are very outlier for the study. And also regarding to the twin engine jets aircraft especially which are on that airlines fleet have minimum 124 passenger seats. When checking the occupancy, rate is about 70 percent for general. Namely less than 10 passenger number is not normal for scheduled flights. In addition, 35 flights were removed also they have more than 70 kg average luggage per passenger. It's also not very usual, because passengers can take maximum 70 kg extra baggage with themselves. Those flights data weren't incorrect but most probably those flights had their own special permission. All those removed flight number is 553 and its approximately %0.24 of all data. As a consequence, it's not going to be effect to the analysis and results.

### 3.2.4. Data Discretization

Data discretization is also one of the primary reduction method. However, discretization reduces continuous attributes to categorical ones characterized by a finite number of discrete values. On the other hand, data discretization aim is to significantly decrease the number of discrete values artificial by the categorical attributes. In this study, distance was categorized into 3 different hauls; short, medium and long. The characterization of Eurocontrol classification was taken into consideration, short haul is less than 1500 km, middle ones between 1500-4000 km and the long haul is more than 4000 km for flight routes. While the region designated it was defined according to corporate headquarters.

Moreover, the outcomes from predictive analytics have has been taken such as cluster numbers. Following all those steps, the prescriptive analytics have been done and the results has been shown in part 4 which is empirical results and assessment.

Lastly the letter codes were derived to conditional attributes in order to the better understanding and demonstrating in Table 6.

Table 6: Codes for Attributes

| Conditional Attributes | Codes |
|---|---|
| **Average Passenger (Number)** | Av.Pax |
| **Baggage Quantity (Number)** | Bag.Q. |
| **Baggage Weight (Kg)** | Bag.W |
| **Y (Economy) Class Rate (%)** | YRate |
| **Infant + Child Passenger Rate (%)** | CIRate |
| **Number of Flights (Time)** | N.Fl |
| **Distance (Km)** | Dis. |
| **Region** | Reg. |

## 3.3.Summary of the Analytics

The summary of analytics will be shown under this title. The three research methodology paradigms and using Business intelligence tools are shown on Table 7 detailed with the form of the headings.

Table 7: Summary of the Method

| Types of Analytics | The Main Purpose | Questions Answers | BI Methods / Techniques and Tools |
|---|---|---|---|
| **Descriptive** | To describe the data for state main data structure | What is the data pattern of passenger luggage based upon the number and weight? | Exploratory Data Analysis and Multidimensional Reporting / OLAP Cubes |
| **Predictive** | To determine customer profile and discovering insights about the future | What will be happened on the future behavior that relates to the profile of passenger baggage for different country and flight type? | Data Mining / Twostep Clustering & K-Means Clustering and Association Rules |
| **Prescriptive** | To support the predictive results, define the attributes' degrees of impact and feeding advice for decision-makers with meaningful results | What should be done to give an advice on possible outcomes using optimization and simulation algorithms, such as put forth some association rules depending on all those analytics and their results? | Multiple Criteria Decision Making / DEMATEL |

In addition to our primary questions above, we may also be looking for answers of the following question: Can be used an online analytical processing and multidimensional cubes in civil aviation sector and is it possible to promotes and support predictive analytics outputs such as association rules, with Decision-Making

Trial and Evaluation Laboratory (DEMATEL) method for making futuristic analytics?

For the prescriptive analytics, DEMATEL method has been developed interview questionnaire and also results reliability and their validity. Proposed criterion set will use factor analysis to clarify the structure's validity and reliability. Furthermore, the proposed set will show the origins and the initial assessment of the model's aspects and criteria. Transforming the qualitative and quantitative information into a comparable scale, evaluate the incomplete information and the DEMATEL in measuring the interrelationships among the criteria.

**Descriptive Analytics**

As it mentioned before, airlines and aviation sector have really much and different types of input and data from diverse sources. Air transportation sector as well as growth rapidly, their data growing increasingly. They have very much continuous information such as passenger, cargo, transfer, etc. The managers need instant or daily reports as they need long term reports in order to get strategic decisions. Reports will look like long term but the model may repeatable and quotable for each year because it’s used comprehensive data which include more than 90 million passengers and it covers 2 years 2014 and 2015.

After the gathering data with SAP HANA platform, Microsoft Excel 2016 and SPSS version 23.01 were utilized on the preprocessing stage. Data was loaded on SPSS in order to define the titles, derived new column, which are mentioned above and indicate data specified. And then, clean data upload the excel, normal and pivot tabulated respectively. To better understand this state of affairs, it is necessary to make the country and the distance-based profiling. After than some graphs coined and interpreted, the results were revealed. After whole data preprocessing and cleaning, according to dataset, there are about 31 million passengers on 229,465 flights in 2014 & 2015 from Istanbul to 247 different airports. All these flights took place to 4 continents, 11 regions and 112 different countries. Their average luggage quantity is 1.12 per person and its weight is 19.21 kg. The multi-dimensional cubes created which include three main layers; region, time and distance, shows that company's main target, high seasons, compact region and more.

**Predictive Analytics**

Predictive analysis was made which include clustering, profiling, classification and lastly rule extraction. After the counseling experts, it has been decided focus luggage optimization problem who is going to be very beneficial for ground service department and also it can be cash cow for cargo department. Furthermore, the predictive results going to be very helpful for marketing, flight and aircraft planning chairmanships.

When finished the descriptive analysis, data which has completed preprocessing process; uploaded whole data in SPSS and describe all variables and their features. Predictive analytics strategy is to first apply a hierarchical agglomeration algorithm, which determines the number of clusters and finds an initial clustering, and then use iterative relocation to improve the clustering. Initial cluster number was determined by using two-step technique as three clusters. Lastly, its derived some association rules with generalized rule induction (GRI) algorithms with the cluster results. Association rules meaning is, created by analyzed data for frequent if/then and using some criteria in order to identify support and confidence for most important relationships. The rules are related baggage weight and number, child passenger rate, business and economy passenger rate, total passenger number and also concerning location like country or region.

**Prescriptive Analytics**

In this study, the factors to be evaluated are considered based upon expert interviews. Propose of this study is an integrated model that combines descriptive analytics (multidimensional analytics) predictive analytics (data mining and more) and prescriptive analytics (MCDM and DEMATEL) in order to extract the critical factors for the improvement of airline baggage optimizations. We apply the dominance-based rough set approach to extract the essential factors. Taking the results from the descriptive and prescriptive analysis throughputs, attributes have been assessed in order to interview with sectoral experts. After that, taking feedback from them, DEMATEL method had been run. Namely, the decision-making trial and evaluation laboratory method (DEMATEL) with the clustering results are then used to construct the new evaluation and assessment system.

At first step of DEMATEL, attributes impact values have already asked to 8 different experts on working this field. The impact value which is from 0 to 4 (0 means that there are not any affect and 4 means that there are relations with those attributes a very dramatically effective) has collected each different specialist. 8 experts are asked to identify the degree of influence between the factors or elements (criteria) to calculate the average matrix of influence matrix.

Following the step of DEMATEL method is created the total influence matrix. The total influence matrix is defined the sum of the rows and the sum of the columns separately which can be denoted as vector r and c. There shows the sum of the direct and indirect effects that factor has received from the other factors. Let $i = j$ and $i, j \in \{1,2,..., n\}$; the horizontal axis vector ($r_i + s_i$) is then made by adding $r_i$ to $s_i$, which illustrates the importance of the criterion. Similarly, the vertical axis vector ($r_i - s_i$) is made by deducting $r_i$ from $s_i$, which may separate criteria into a cause group and an affected group. In general, when ($r_i - s_i$) is positive, the criterion is part of the cause group. On the contrary, if ($r_i - s_i$) is negative, the criterion is part of the affected group. Therefore, a causal graph can be achieved by mapping the dataset of ($r_i + s_i$, $r_i - s_i$), providing a valuable approach for decision-making

### 3.4. Reliability and Validity

In order for research data to be of value and of use, they must be both reliable and valid. Data reliability is a state that exists when data is sufficiently complete and error free to be convincing for its purpose and context. And also, it refers to the accuracy and completeness of computer-processed data, given the intended purposes for use. It has mentioned above the data accuracy and others preprocessing stages. There are several types of validity that contribute to the overall validity of a study.

The two main dimensions are internal and external validity, and further sub-types can be added under these headings. About internal validity, descriptive validity is that which is concerned with the initial stage of research, usually involving data gathering. In this study, data was taken from airline company secondarily, because of that data acquired reliably. Applying a statistical tool to examine validity is entirely anchored in what it is you think is being measured. For descriptive part, multidimensional tables have valid impression. About the predictive section, in order to the determined the cluster number, two-step tool was chosen which has using hierarchical techniques. Moreover, about the association rules, the more then 75 % of the results have taken notice. To conclusion, all analysis validity and reliability result will be given after their outcomes separately. Furthermore, predictive results valid with ANOVA tests and other tools which are shows below. For the prescriptive analytics, DEMATEL method has been developed interview questionnaire and also results reliability and their validity. Proposed criterion set will use factor analysis to clarify the structure's validity and reliability.

For the external validity, is concerned with the degree to which research findings can be applied to the real world, beyond the controlled setting of the research. This is the issue of generalizability and attempts to increase internal validity are likely to reduce external validity as the study is conducted in a manner that is increasingly unlike the real world. In this study, data has taken from only one company which is the one of the most lack sides. However, that company is one of the biggest company on the world and also flag carrier of Turkey. On the other side, data has only included 2 years, 2014-2015. It can be also increased with previous years data for better understanding.

# EMPIRICAL RESULTS AND ASSESSMENT

## 1.6.Results of Descriptive Analytics

### 1.6.1. Exploratory Data Analysis

Normally the data consist of 855.250 different flights which has include 115.629.955 passengers, 3.661 different flight routes, 294 different departure airports, 104.419.219 pieces and approximately 1,555 billion kg luggage. As it mentioned before, after finished completing the data, picked out the flights which are departure from Istanbul. The number was approximately 300.000 different flights. Then, the domestics flight has been removed and only left about 230.000 rows. Lastly, some of them was deleted because of validation and incompletion of the data as it mentioned before and eventually has been left 229.465 flights which are from Istanbul to international point in during in 2014 and 2015.

The airline flight more than 115 countries and 285 different points, hence it has very various passenger and luggage profiles and company that need the most these serviceable and functional reports. On the other hand, this study focuses international flights because of to reducing the effect of external factors. Therefore, only international flights which are from Istanbul to other destinations were selected. To summarize of the data from January 2014 to December 2015 with 229,465 different flights shown in Table 8.

Table 8: Summary of the Flights Data in 2014 & 2015

| Items | Min | Max | Total | Mean | S. D. |
|---|---|---|---|---|---|
| **Total Pax** | 10 | 349 | 30840110 | 134 | 60 |
| **Av. Bgg. Qtt** | 0.34 | 5 | 256316 | 1 | 0.2705486 |
| **Bgg. Wg** | 114 | 13602 | 591867425 | 2579 | 1478 |
| **Av. Bgg.Wg** | 5 | 70 | 4407949 | 19 | 6 |
| **Yclass Pax** | 3 | 349 | 28447258 | 124 | 52 |
| **Cclass Pax** | 0 | 49 | 2118925 | 9 | 8 |
| **CC Pax Rate** | 0 | 3 | - | 0.072759 | 0.0556143 |
| **CI.Pax** | 0 | 103 | 1313976 | 6 | 7 |
| **CI. Pax Rate** | 0 | 0.5085 | - | 0.041056 | 0.0405032 |
| **Distance** | 264 | 11057 | 594072990 | 2,589 | 2,069 |

The minimum passenger number of flight was 10 because lesser ones were removed as it mentioned methodology section. The maximum number of passenger is 349 which is the seat number of biggest aircraft in company and also the maximum C class number is 49 because of the design of the aircrafts. Lastly, distance of the arrival points from Istanbul airport change from 264 to 11057 km.

Moreover, as it shown above, for about 230,000 flights, the average number of passengers per flight is 134,4. The maximum flight number concerns a route from Istanbul to Tel Aviv Airport and there were more than 4849 different flights in 2 years. The maximum passenger number is characteristic for the HKG (Hong Kong) flights with average of about 305 passengers per flight. However, the NDJ (N'Djamena Airport, Chad) flights average number of passengers is approximately 35 and that is the lowest quantity by route during examined year. The average baggage weight of 19.21 kilograms is the lowest at 12.44 BUS (Alexander Kartveli Batumi International Airport, Georgia) flights and more then 45 kg at the highest on FIH (Kinshasa, the Democratic Republic of the Congo) flights. The highest Y rate is KSH (Kermanshah, Iran) airport with 99 per cent and the highest C rate is about 20 per cent on MRU (Sir Seewoosagur Ramgoolam International Airport, Mauritius) flights. The lowest child and infant passengers rate is only 1 per cent on PEK (BEIJING, China) flights and the highest is 13 per cent on ELQ (Prince Nayef bin Abdulaziz Regional Airport, Saudi Arabia) flights.

### 1.6.2. Multidimensional Analysis Results

It is a common knowledge that every item on board makes a plane heavier therefore burns more fuel. An airliner's cost of operating rises with every laptop (33 cents per flight), pillow (6 cents), or magazine (5 cents) you bring along (Stone, 2017). The luggage values in the different flights have difference between them based on distance and countries. To better understand this state of affairs, it is necessary to make the country and the distance-based profiling.

According to dataset, there are about 31 million passengers on 229,465 flights in 2014 & 2015. Their average luggage quantity is 1.12 per person and its weight is 19.21 kg. The economic class passenger rate for THY flights which are depart from Istanbul is 93 percent and depending on this business class passenger rate is 7 %. The total of the child and infant passengers rate is 4.11 %, but it is hard to remark the relation with a Child-Infant rate between continent, therefore it should be check also regions more detailed. The multi-dimensional cubes such as Figure 12, which include three main layers; region, time and distance, shows that company's main target is Middle Europe, Middle East and South Europe which have more number of passengers and flights. However, these are more than half of the company flights the business rate of the America flights has reached about 11 percent therefore the firm can increase the business class seat number or they may focus for the business passenger

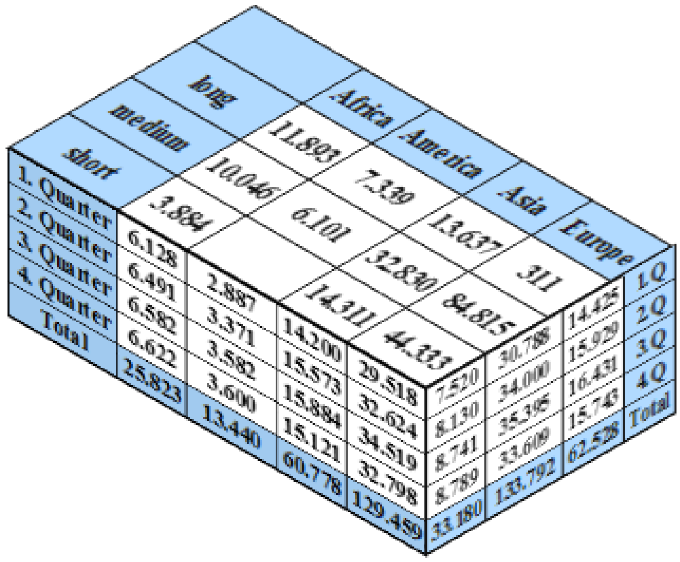



Figure 12: Descriptive Analytics Cube by Flights

There is relationship between distance and luggage weights. When people fly long haul, they need more luggage. To explain the data pattern of luggage, the relationship between the distance and luggage have not enough the power of knowledge. For this reason, predictive analysis is used to discover of data pattern.

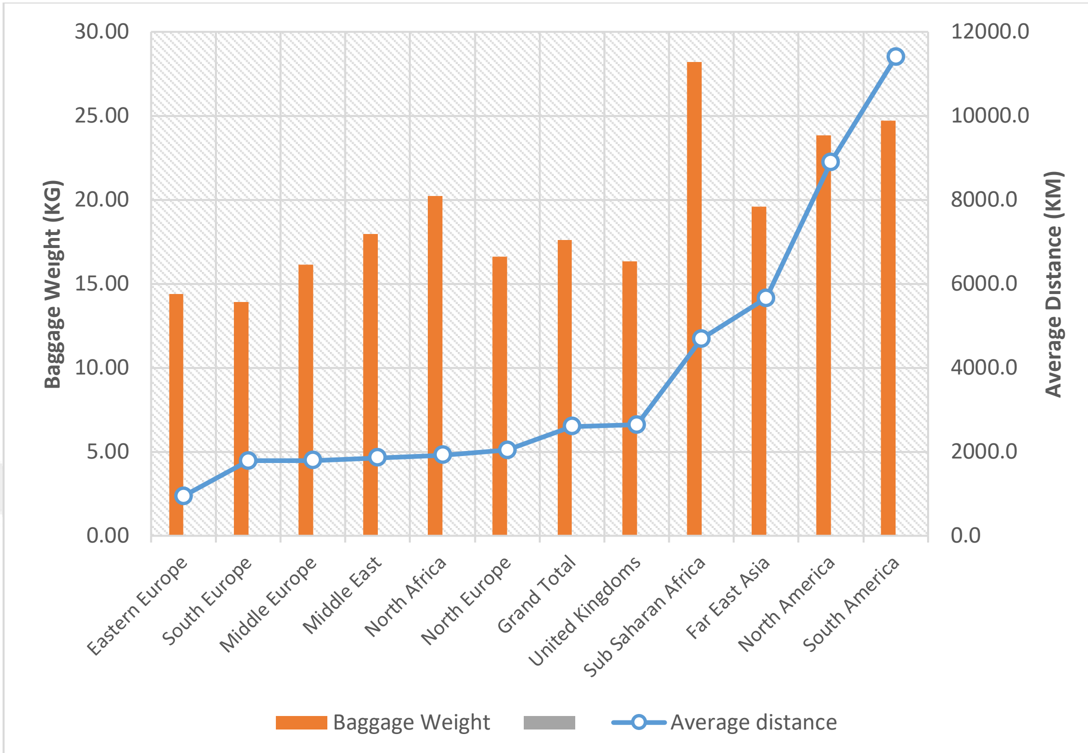
Moreover, Figure 12 shows that, one of the most important target for that company is Middle Europe, Middle East and South Europe which have most passenger and flights and also, they are more than half of the company flights. One of the most salient touch is CI rate in United Kingdom, it is about twofold the average rate, however, it is not easy to tell something just with two-years data.

Table 9: Statistics by Continents & Regions in 2014 - 2015

| Regions | Number of Flights | Total PAX | Baggage Quantity | Baggage Weight | Y rate | CI rate |
|---|---|---|---|---|---|---|
| Africa | **25,789** | **2,522,452** | **1.45** | **26.87** | **0.08** | **0.04** |
| **North Africa** | 10,575 | 1,354,035 | 1.29 | 21.76 | 0.07 | 0.04 |
| **Sub Saharan Africa** | 15,214 | 1,168,417 | 1.57 | 30.42 | 0.09 | 0.04 |
| America | **8,057** | **1,942,162** | **1.39** | **26.40** | **0.11** | **0.04** |
| **North America** | 6,620 | 1,755,088 | 1.40 | 26.28 | 0.11 | 0.05 |
| **South America** | 1,437 | 187,074 | 1.37 | 26.96 | 0.11 | 0.02 |
| Asia | **60,777** | **9,335,663** | **1.17** | **20.33** | **0.09** | **0.04** |
| **Far East Asia** | 23,353 | 4,247,418 | 1.22 | 21.37 | 0.10 | 0.03 |
| **Middle East** | 37,424 | 5,088,245 | 1.14 | 19.68 | 0.08 | 0.05 |
| Europe | **134,842** | **17,039,833** | **1.01** | **16.81** | **0.06** | **0.04** |
| **Eastern Europe** | 26,294 | 2,860,901 | 0.97 | 15.85 | 0.05 | 0.03 |
| **Middle Europe** | 46,367 | 6,110,226 | 1.06 | 17.72 | 0.06 | 0.04 |
| **North Europe** | 19,333 | 2,393,456 | 1.07 | 18.25 | 0.05 | 0.05 |
| **South Europe** | 32,599 | 4,192,277 | 0.93 | 15.13 | 0.07 | 0.03 |
| **United Kingdom** | 10,249 | 1,482,973 | 1.04 | 17.79 | 0.09 | 0.08 |
| Grand Total | **229,465** | **30,840,110** | **1.12** | **19.21** | **0.07** | **0.04** |

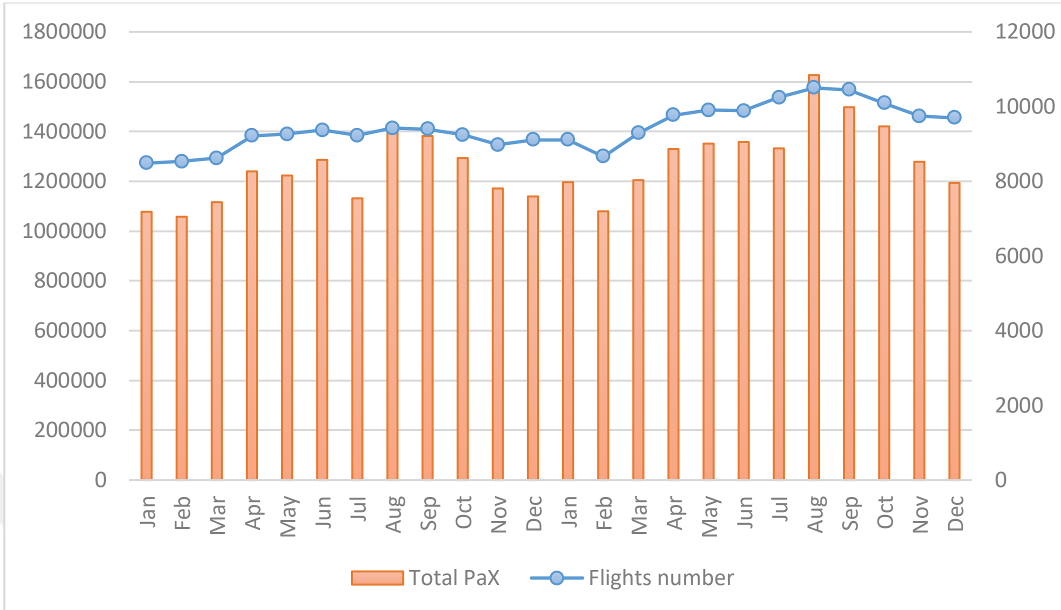
When Table 9 is checked it is obvious that there is relationship between distance and luggage weights judiciously. When people fly long haul, they need more luggage. On the other hand, if look at the Table 9 again detailed, it is not only about the distance, because South Americas flights are the farthest but they have not got the highest baggage weight and also numbers.

Another example is long haul Far East Asia flights; they have less baggage weight compared with medium haul North African flights. This should be regarding to people behavior and their culture. In addition, Far East Asia and North Europe flights in long haul, Eastern and South Europe in medium haul flights has number of baggage below the average remarkably. That results will be significant for cargo department and they can use their passenger flights for carrying cargo more.




Graph 1: Regions Distance & Baggage Statistics

When looking at the time phase, normally it looks like high and low season, however it's look unsteady in Graph 1. Normally high season start from July and finish October because of the summer season (Dax, 1975). For Turkish Airlines flights, high and low seasons can see also in Graph 2. Moreover, December and January can call second high season because they cover touristic holidays (Roy et al, 2001). Nevertheless, because of the political issues and election on July and terror attracts the high/low seasons didn't appear obviously. It can be proved that with July data, the firm has increased the number of the flights but the passenger number didn't raise.




Graph 2: Passenger and Flights Numbers in 2014 & 2015 month by month

Graph 2 have two years' data to improve the time phrase. Although, the rise of the fleet and flight number in 2015, passenger number didn't increase insomuch, because of the political and other external factor. The effect of the political issues is more apparent when checking distance features with month by month. For instance, check the month of Junes, there is no increasing in second one, because of the 7 June election in 2015. Meanwhile, some regions' weight and number of baggage decreased on high season such as; Middle East, Far East Asia and North Africa. On the other hand, some regions like Middle and North Europe, North America and United Kingdom have increased their average passenger baggage number and weight. If it focuses on this point, it can figure out the passenger's profile elaborately but it can be said only thing airlines should care differences of the expat people and tourists.

To summarize, when looking at the airline data with multidimensional and descriptive analysis, it can raise lots of rules and different outcomes. Some of them can be very important for the crew, aircraft or schedule planning departments and it may be also beneficial also in order to advertising and marketing departments.

Moreover, cargo and ground operation chairmanship can also get some prediction based on those results.

## 1.7. Results of Predictive Analytics

### 1.7.1. Cluster Analysis Results

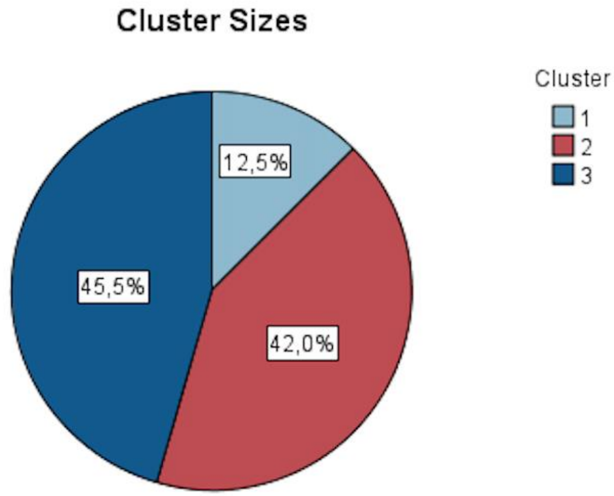
The data have 229,465 different flights, and clustering has made in order to profiling those flights and finding some significant rules. After defining the data and preliminary studies, clustered all the flights depending on variables which are the average number of baggage weight, baggage number and Child-Infant passenger rates.

In the predictive section, strategy that is to first apply a hierarchical agglomeration algorithm, which determines the number of clusters and finds an initial clustering, and then use iterative relocation to improve the clustering. Initial cluster number was determined by using two-step technique as three clusters. The number of clusters and their features which are derived by Two-Step technique is shown in Table 10.

Table 10: Clusters derived by Two-Step Technique

| **Clusters** | **Number** | **% of Combined** | **Mean** | |
|---|---|---|---|---|
| | | | **Av. Bgg. Qtt** | **Av. Bgg. Wg** |
| 1 | 28678 | 12.5% | 1.65 | 30.99 |
| 2 | 96425 | 42.0% | 1.18 | 20.45 |
| 3 | 104362 | 45.5% | 0.92 | 14.83 |
| Total | 229465 | 100.0% | 1.12 | 19.21 |

The cluster numbers and their percentages shown below in Graph 3. Moreover, their pie graphs shown in Graph 3 and ratio of biggest cluster is 45.5 % (104,362) and smallest size of cluster ratio is 12.5 % (28,678).




Graph 3: Two-Step Clusters on Pie Graph

Depending on that, ratio of sizes largest to smallest cluster is 3.64. The cluster quality has approximately 0.8 cohesion and separation measure shown below in Figure 13.

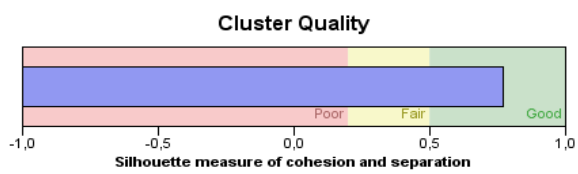
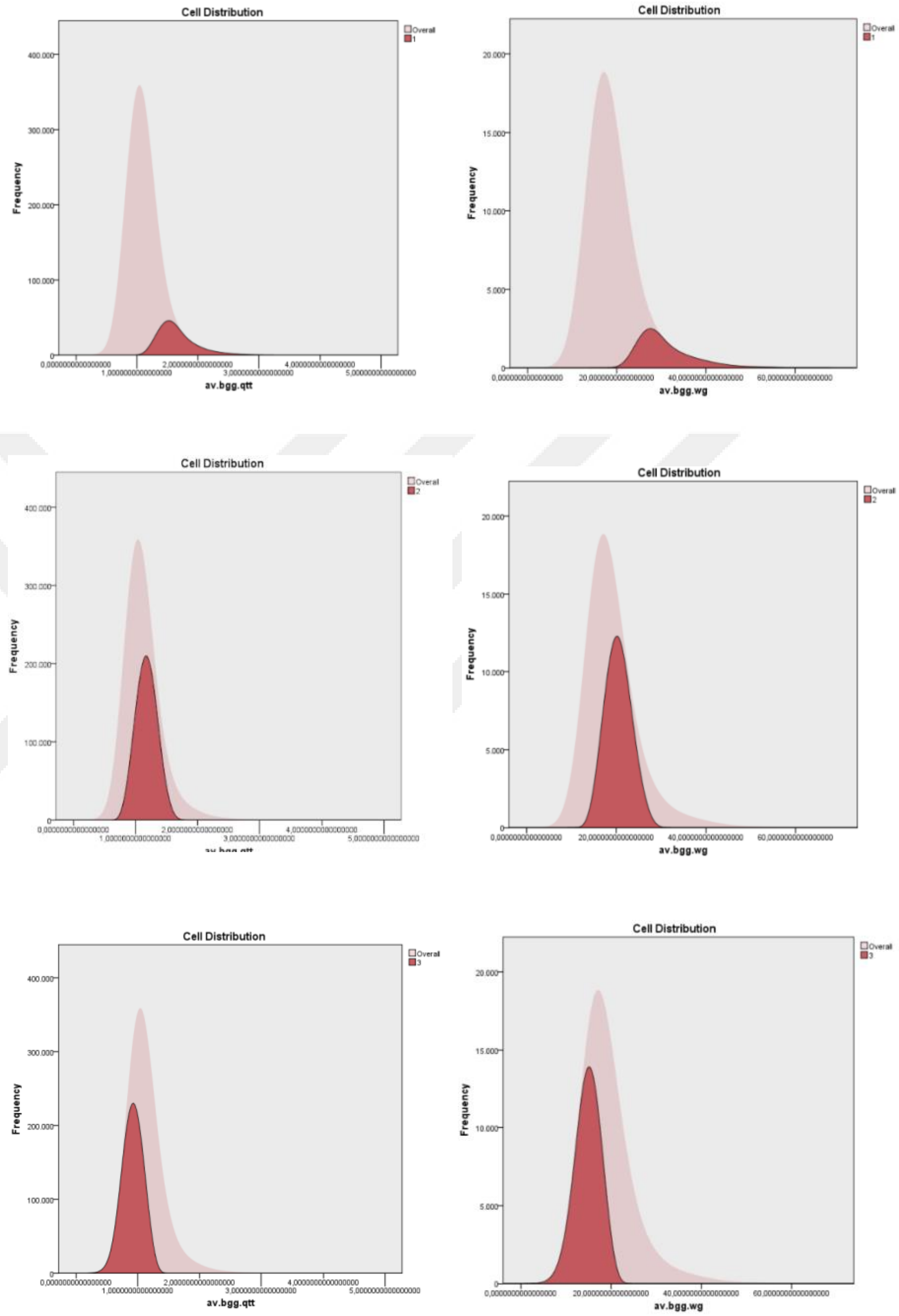



Figure 13: Cluster Summary and Quality

The cell distribution of average baggage quantity and weight also shown below in cluster by cluster.




It can be understanding from those graphs, depending on the average baggage weight and average baggage number, the distribution of the clusters is conspicuously apparent. Moreover, the cluster 2 is closer to the average and also include average of the all data, and cluster 3 contains low baggage weight and number and cluster 1 is above the mean.

On the other hand, TwoStep technique applied only for detect number of cluster. After defined optimum number of cluster is 3, K-Means classify technique applied. The cluster numbers picked 3, and iteration numbers picked 100. But only 34 iterations took places. History of iterations is shown in Annex 1. Convergence achieved due to no or small change in cluster centers. The maximum absolute coordinate change for any center is ,001 when iteration was 34. The minimum distance between initial centers is 32,262. Table 11 shows that the number of cases in each cluster derived from K-means technique.

Table 11: Number of Cases in Each Cluster

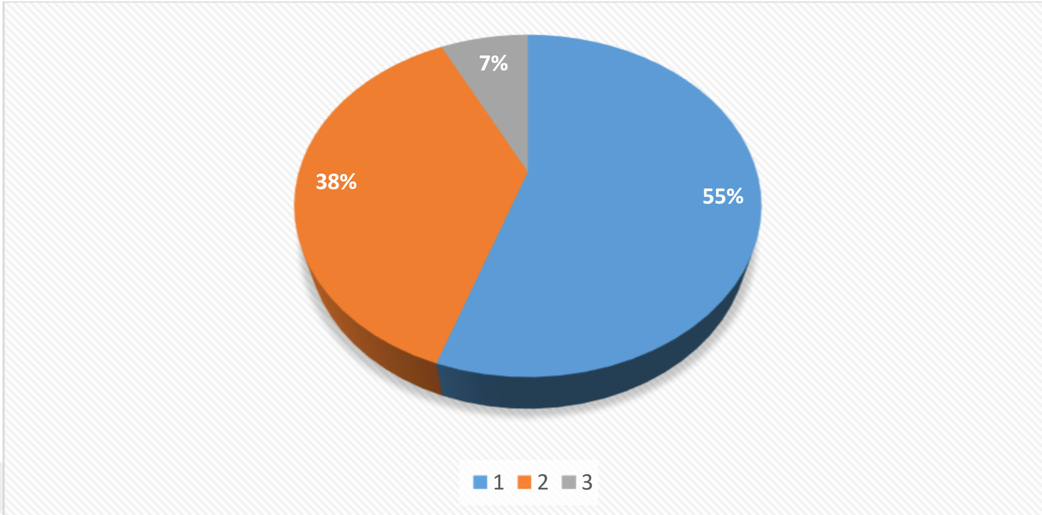
| | | |
|---|---|---|
| **Cluster** | 1 | 126795.0 |
| | 2 | 85879.0 |
| | 3 | 16791.0 |
| **Valid** | | 229465.0 |

In addition, K-means cluster centers are shown in Table 12.

Table 12: Clusters derived by K-means Technique

| | **Cluster** | | |
|---|---|---|---|
| | **1** | **2** | **3** |
| **Average Baggage Quantity** | .9475 | 1.2386 | 1.7746 |
| **Average Baggage Weight** | 15.3741 | 21.8942 | 34.4425 |

It can be understood from the table 12 and 13 that, cluster 1 is the biggest and dominant one and it also include low values of passenger baggage. On the other hand, cluster 3 has high baggage values and it really has values above the mean. Distribution of the clusters can be more visible in the Graph 4.




Graph 4: K-Means Cluster on Pie Graph

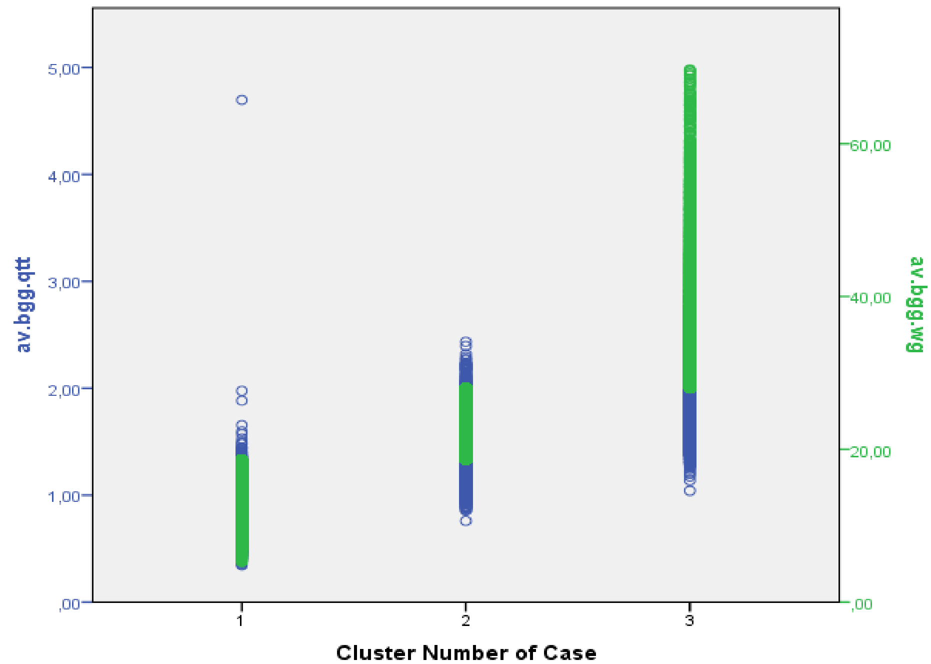
Consequently, clustering results were subjected to ANOVA test is below shown in Table 13. Therefore, one can assume, there is a statistically significant difference.

Table 13: Detailed ANOVA Analysis Results

| | Sum of Squares | df | Mean Square | F | Sig. | | |
|---|---|---|---|---|---|---|---|
| Av.Bgg.Qtt | Between Groups | (Combined) | 12175.368 | 2 | 6087.684 | 302318.184 | 0.000 |
| | | Linear Term | Unweighted | 10144.113 | 1 | 10144.113 | 503762.972 |
| | | | Weighted | 11649.610 | 1 | 11649.610 | 578526.903 |
| | | | Deviation | 525.758 | 1 | 525.758 | 26109.465 |
| | Within Groups | 4620.602 | 229462 | .020 | | | |
| | Total | 16795.970 | 229464 | | | | |
| Av.Bgg.Wg | Between Groups | (Combined) | 6380426.237 | 2 | 3190213.1 | 414264.187 | 0.000 |
| | | Linear Term | Unweighted | 5391311.060 | 1 | 5391311.060 | 700087.113 |
| | | | Weighted | 6061719.817 | 1 | 6061719.817 | 787142.845 |
| | | | Deviation | 318706.420 | 1 | 318706.420 | 41385.529 |
| | Within Groups | 1767067.263 | 229462 | 7.701 | | | |
| | Total | 8147493.500 | 229464 | | | | |

ANOVA analysis of the emerging clusters was performed, distance between cluster centers and centers was determined. Moreover, distance of each cluster, calculated according to data which is shown in annex 2.

The Graph 5 shows that the average baggage weights on the right and average baggage quantities on the left for each flights cluster with a plot diagram.




Graph 5: THY Flights Average Baggage Weights by Clusters in 2014 & 2015

### 1.7.2. Association Rules Results

Lastly, its derived some association rules with generalized rule induction (GRI) algorithms with the cluster results. Association rules meaning is, created by analyzed data for frequent if/then and using some criteria in order to identify support and confidence for most important relationships. In descriptive analytics, association rules are useful for analyzing and predicting customer behavior in general.

For association rules, it has derived 41 different rules with GRI algorithms. The number of valued transaction was 229.495 and the minimum support is 2 and maximum is about 57 percent. That means the rule includes 57 percent of the entire data. In general terms, if the confidence percent more than 70, it can be said foolproof, so in this study minimum confidence selected 75 percent.  Other features show below in Table 14.

Table 14: GRI Rules Features

| |
|---|
| **Analysis** |
| Number of Rules: 41 |
| Number of Valid Transactions: 229.497 |
| Minimum Support: 2 % |
| Maximum Support: 57.21% |
| Minimum Confidence: 75% |
| Maximum Confidence: 100,0% |
| Minimum Lift: 1,464% |
| Maximum Lift: 11,906% |
| **Consequents** |
| ClusterNumber |
| **Antecedents:** Ttl.Pax, Bgg.Qtt, Bgg.Wg, Yclass, CI.Pax, Arregion, Arrcontinent, Distype |
| **Build Settings** |
| Maximum number of antecedents: 3 |
| Minimum antecedent support (%): 0,0 |
| Minimum rule confidence (%): 80,0 |
| Algorithm: GRI |
| Model type: Association |

As can be concluded from the figure, the values within the cluster are very close to each other and the distance between the clusters is about the same as the distance. This allows the authors to observe the clustering value visually. Table 15 shows the accuracy rates and the amount of support of the rule occurs.

Table 15: Rules Derived from Generalized Rule Induction (GRI) Algorithms

| Number | Cluster Number | Rules | Support % | Confidence % |
|---|---|---|---|---|
| **1** | CN = 1 | CI.pax < 4,5 | 57.2 | 78.79 |
| **2** | CN = 1 | bgg.qtt < 141,5 | 54.22 | 76.09 |
| **3** | CN = 1 | av.CI.rate < 0,0272 | 45.89 | 77.12 |
| **4** | CN = 1 | CI.pax < 4,5 and ttl.pax > 73,5 | 45 | 78.12 |
| **5** | CN = 1 | bgg.qtt < 141,5 and arrcontinent = europe | 36.61 | 86.5 |
| **6** | CN = 1 | arrcontinent = europe and distance < 1826,3 | 31.82 | 81.98 |
| **7** | CN = 1 | CI.pax < 4,5 and ttl.pax > 73,5 and arrcontinent = europe | 29 | 84.21 |
| **8** | CN = 1 | av.CI.rate < 0,02725 and arrcontinent = europe | 27.3 | 85.59 |
| **9** | CN = 1 | CI.pax < 4,5 and CI.pax < 2,5 and arrcontinent = europe | 22.75 | 86.2 |

| | | | | |
|---|---|---|---|---|
| **10** | CN = 1 | distance < 1519,6 and bgg.qtt < 141,5 | 20.28 | 86.17 |
| **11** | CN = 1 | distance < 1519,6 and ttl.pax < 149,5 and bgg.qtt < 130,5 | 17.61 | 87.06 |
| **12** | CN = 1 | distance < 1519,6 and bgg.qtt < 141,5 and arrcontinent = europe | 16.35 | 90.74 |
| **13** | CN = 2 | CI.pax > 10,5 | 15.21 | 76.21 |
| **14** | CN = 1 | arrregion = south europe | 14.21 | 89.78 |
| **15** | CN = 2 | av.CI.rate > 0,0819 | 12.85 | 77.18 |
| **16** | CN = 1 | arrregion = eastern europe | 11.46 | 83.46 |
| **18** | CN = 2 | av.CI.rate > 0,08981 | 10.51 | 75.03 |
| **19** | CN = 1 | CI.pax < 4,5 and arrregion = south europe | 9.95 | 93.27 |
| **20** | CN = 1 | av.CI.rate < 0,0272 and arrregion = south europe | 8.3 | 94.12 |
| **21** | CN = 2 | av.CI.rate > 0,1087 | 7.5 | 78.21 |
| **22** | CN = 2 | CI.pax > 18,5 | 7.3 | 76.22 |
| **23** | CN = 2 | av.CI.rate > 0,13569 | 6.7 | 75.19 |
| **24** | CN = 1 | ttl.pax > 237,0 and bgg.qtt < 273,5 | 6.2 | 95.21 |
| **25** | CN = 3 | bgg.qtt > 405,5 | 5.3 | 80.46 |
| **26** | CN = 2 | arrregion = north africa and Distype = short | 5.2 | 85.26 |
| **27** | CN = 1 | ttl.pax > 237,0 and av.CI.rate < 0,00947 and arrcontinent = asia | 4.7 | 75.44 |
| **28** | CN = 1 | ttl.pax > 237,0 and av.CI.rate < 0,00947 and arrcontinent = asia | 4.4 | 77.81 |
| **29** | CN = 1 | ttl.pax > 237,0 and arrregion = south europe | 4.2 | 81.8 |
| **30** | CN = 3 | arrregion = sub saharan africa and Distype = short | 4.1 | 78.41 |
| **31** | CN = 1 | ttl.pax > 237,0 and Cclassrate < 0,0035 | 3.1 | 91.79 |
| **32** | CN = 2 | CI.pax > 44,5 and distance < 3359,5 | 3.1 | 90.7 |
| **33** | CN = 1 | ttl.pax > 237,0 and Cclassrate < 0,0035 and arrcontinent = asia | 2.9 | 93.85 |
| **34** | CN = 1 | ttl.pax > 237,0 and Cclassrate < 0,0035 and arrregion = middle east | 2.8 | 93.8 |
| **35** | CN = 1 | ttl.pax > 345,5 | 2.7 | 75.94 |
| **36** | CN = 1 | bgg.qtt < 12,5 | 2.5 | 81.97 |
| **37** | CN = 2 | CI.pax > 44,5 | 2.5 | 81.02 |
| **38** | CN = 1 | ttl.pax > 345,5 | 2.5 | 76.01 |
| **39** | CN = 2 | yclass < 8,5 | 2.4 | 75.04 |
| **40** | CN = 2 | Cclassrate > 0,8236 | 2.1 | 76.12 |

To illustrate the rules, they are related baggage weight and number, child passenger rate, business and economy passenger rate, total passenger number and also concerning location like country or region. Those rules have really significant knowledge for marketing and planning department. Moreover, they will help to solve some optimization problems in air transportation sector. For example, in cluster number 1, total number of child and infant passenger is less than 4.5 what confidence rate is approximately 79. This is also important for airlines since they can arrange their flights take into consideration that., The cluster number 2 is included the more than 10,5 child and infant passenger per flight (Confidence rate is 76.21). While cluster number have less than 141 baggage per flight at %77 confidence rate, cluster number 3 have more than 405.5 baggage quantity in each flight at %81 confidence rate. Additionally, there are also lots of different types of rules, such that business class passenger number is less than 8.5 per flight in cluster number 2.

Eventually, about thousands of rules derived from association rules GRI algorithms however those which are valuable in terms of strategic and application were chosen.

The above data contain statistical information particularly for flight planning, cargo and marketing departments. Especially the baggage planning and cargo departments can easily use these results with their operations. They can make decisions on the basis of the meaningful clusters and their outstanding information. Moreover, these results show that, distance is also a significant factor for passenger profiling, not only from potential perspective. Likewise, it affects profiles from different countries, different regions and different time. Furthermore, it would be an important factor in determining the aircraft type.

### 1.8. Results of Prescriptive Analytics

In this study, the factors to be evaluated are considered based upon expert interviews which is mentioned in Methodology. At first step of DEMATEL, attributes impact values have already asked to 8 different experts on working this field. The impact value which is from 0 to 4 (0 means that there are not any affect and 4 means that there are relations with those attributes a very dramatically effective) has collected each different specialist. 8 experts are asked to identify the degree of influence between the factors or elements (criteria) to calculate the average matrix of influence matrix.

The DEMATEL method has been applied to build a relation structure among the criteria for airline affecting factors. The five phases of the method have demonstrated below with consequence tables.

Phase 1:

Gathering data (evaluation and impact value) from sectoral and academic experts. On this phase, attributes impact values have already asked to 8 different experts on working this field. The impact value which is from 0 to 4 (0 means that there are not any affect and 4 means that there are relations with those attributes a very dramatically effective) has collected each different specialist.

Phase 2:

On this phase, the calculation of direct influence matrix (it can be called M matrix) 8 experts are asked to identify the degree of influence between the factors or elements (criteria). Based on those numbers calculated average matrix is possible. An integer scale is applied to rank the influence between the elements ranging from 0 to 4 as it mentioned before. It is possible to derive an average M matrix from any group of direct matrices that proceed from the responses of the 8 experts. Every element of the average matrix shows the mean of the answers for the same attributes provided by expert. All those average numbers are show in Table 16 with the derived codes which is mentioned on Data Preprocessing part.

Table 16: Initial Influence Matrix M

| Conditional Attributes | Av.Pax | Bag.Q. | Bag.W | YRate | CIRate | N.Fl | Dis. | Reg. |
|---|---|---|---|---|---|---|---|---|
| **Av.Pax** | 0 | 2.63 | 2.88 | 3 | 2 | 1 | 2.88 | 1 |
| **Bag.Q.** | 1.88 | 0 | 3.13 | 1 | 1.25 | 3.13 | 3.33 | 3 |
| **Bag.W** | 1 | 2 | 0 | 2.25 | 2.88 | 2.88 | 3.13 | 2.88 |
| **YRate** | 3 | 2.75 | 2.88 | 0 | 1.75 | 0.5 | 3.33 | 3.1 |
| **CIRate** | 0.5 | 0.75 | 2.25 | 0.75 | 0 | 0 | 0.25 | 1.75 |
| **N.Fl** | 3.33 | 2.5 | 2.25 | 1 | 1 | 0 | 0.75 | 3 |
| **Dis.** | 3.75 | 2.88 | 2.75 | 3.33 | 0.5 | 1 | 0 | 1 |
| **Reg.** | 3.33 | 2.75 | 2.88 | 3.13 | 1.25 | 2.75 | 2 | 0 |

Phase 3:

After the defined matrix M, then matrix T which is normalized influence matrix will be calculate this formula; finding the maximum column and row total and chose the biggest ones. Then multiply the matrix M to resolve the matrix T which is shown in Table 17.

Normalize the direct-influence matrix based on the direct-influence matrix **D**, the normalized direct relation matrix **X** is acquired by using Equationss (1) and (2).

$$\boldsymbol{X} = s\boldsymbol{D} \qquad \text{(Equation 1)}$$

$$s = \max\left\{\max_{1\le i\le n}\sum_{j=1}^{n} d_{ij}, \max_{1\le j\le n}\sum_{i=1}^{n} d_{ij}\right\} \qquad \text{(Equation 2)}$$

Table 17: Normalized Influence Matrix T

| Conditional Attributes | Av.Pax | Bag.Q. | Bag.W | YRate | CIRate | N.Fl | Dis. | Reg. |
|---|---|---|---|---|---|---|---|---|
| **Av.Pax** | 0.000 | 0.138 | 0.151 | 0.158 | 0.105 | 0.053 | 0.151 | 0.053 |
| **Bag.Q.** | 0.099 | 0.000 | 0.165 | 0.053 | 0.066 | 0.165 | 0.175 | 0.158 |
| **Bag.W** | 0.053 | 0.105 | 0.000 | 0.118 | 0.151 | 0.151 | 0.165 | 0.151 |
| **YRate** | 0.158 | 0.145 | 0.151 | 0.000 | 0.092 | 0.026 | 0.175 | 0.163 |
| **CIRate** | 0.026 | 0.039 | 0.118 | 0.039 | 0.000 | 0.000 | 0.013 | 0.092 |
| **N.Fl** | 0.175 | 0.131 | 0.118 | 0.053 | 0.053 | 0.000 | 0.039 | 0.158 |
| **Dis.** | 0.197 | 0.151 | 0.145 | 0.175 | 0.026 | 0.053 | 0.000 | 0.053 |
| **Reg.** | 0.175 | 0.145 | 0.151 | 0.165 | 0.066 | 0.145 | 0.105 | 0.000 |

Phase 4:

After obtain the matrix T, the total influence matrix A can be defined the sum of the rows and the sum of the columns separately which can be denoted as vector r and c (comes from the DEMATEL calculation) within the total-influence matrix A. Finally, it has showed the sum of the direct and indirect effects that factor has received from the other factors.

Attaining the total-influence matrix **T**. Once the normalized direct-influence matrix **X** is obtained, the total-influence matrix **T** for NRM can be obtained through Equation (3), in which **I** denotes the identity matrix.

$$\boldsymbol{T} = \boldsymbol{X} + \boldsymbol{X}^2 + \boldsymbol{X}^3 + \ldots + \boldsymbol{X}^k$$
$$= \boldsymbol{X}(\boldsymbol{I} + \boldsymbol{X} + \boldsymbol{X}^2 + \ldots + \boldsymbol{X}^{k-1})(\boldsymbol{I} - \boldsymbol{X})(\boldsymbol{I} - \boldsymbol{X})^{-1} = \boldsymbol{X}(\boldsymbol{I} - \boldsymbol{X}^k)(\boldsymbol{I} - \boldsymbol{X})^{-1}$$
$$= \boldsymbol{X}(\boldsymbol{I} - \boldsymbol{X})^{-1}$$

when $k \to \infty$, $\boldsymbol{X}^k = [0]_{n \times n}$ ( Equation 3)

where $\boldsymbol{X} = [x_{ij}]_{n \times n}$, $0 \leq x_{ij} < 1$, $0 < \sum_{j=1}^{n} x_{ij} \leq 1$ and $0 < \sum_{i=1}^{n} x_{ij} \leq 1$. If at least one row or column of summation is equal to 1, but not all, then $\lim_{k \to \infty} \boldsymbol{X}^k = [0]_{n \times n}$.

Step 4: Analyze the results. In this stage, the sum of the rows and the sum of the columns are separately expressed as vector $\boldsymbol{r} = (r_1, \ldots, r_i, \ldots, r_n)'$ and vector $\boldsymbol{s} = (s_1, \ldots, s_j, \ldots, s_n)'$ by using Equations (4)-(6). Let $i = j$ and $i, j \in \{1, 2, \ldots, n\}$; the horizontal axis vector ($r_i + s_i$) is then made by adding $r_i$ to $s_i$, which illustrates the importance of the criterion. Similarly, the vertical axis vector ($r_i - s_i$) is made by deducting $r_i$ from $s_i$, which may separate criteria into a cause group and an affected group. In general, when ($r_i - s_i$) is positive, the criterion is part of the cause group. On the contrary, if ($r_i - s_i$) is negative, the criterion is part of the affected group. Therefore, a causal graph can be achieved by mapping the dataset of ($r_i + s_i$, $r_i - s_i$), providing a valuable approach for decision-making.

$\boldsymbol{T} = [t_{ij}]_{n \times n}$, $i, j = 1, 2, \ldots, n$ (Equation 4)

$$\boldsymbol{r} = \left[\sum_{j=1}^{n} t_{ij}\right]_{n \times 1} = [t_i]_{n \times 1} = (r_1, \ldots, r_i, \ldots, r_n)' \quad \text{(Equation 5)}$$

$$\boldsymbol{s} = \left[\sum_{i=1}^{n} t_{ij}\right]'_{1 \times n} = \left[t_j\right]_{n \times 1} = (s_1, \ldots, s_j, \ldots, s_n)' \quad \text{(Equation 6)}$$

where vector **r** and vector **s** express the sum of the rows and the sum of the columns from the total-influence matrix $\boldsymbol{T} = \left[t_{ij}\right]_{n \times n}$, respectively, and the superscript denotes the transpose (Chen et al, 2010a). Now we call the total-influence matrix $\boldsymbol{T}_C = \left[t_{ij}\right]_{n \times n}$ obtained by criteria and $\boldsymbol{T}_D = \left[t_{ij}^D\right]_{m \times m}$ obtained by dimensions (clusters) from experts' opinions. Then we normalize the ANP weights of dimensions (clusters) by using influence matrix $\boldsymbol{T}_D$.

Table 18: Total Influence Matrix A

| Conditional Attributes | Av.Pax | Bag.Q. | Bag.W | YRate | CIRate | N.Fl | Dis. | Reg. | Total (Y) |
|---|---|---|---|---|---|---|---|---|---|
| **Av.Pax** | 0.466 | 0.584 | 0.661 | 0.564 | 0.413 | 0.391 | 0.604 | 0.493 | **4.176** |
| **Bag.Q.** | 0.610 | 0.510 | 0.719 | 0.521 | 0.406 | 0.527 | 0.658 | 0.619 | **4.571** |
| **Bag.W** | 0.559 | 0.590 | 0.564 | 0.559 | 0.469 | 0.499 | 0.633 | 0.606 | **4.479** |
| **YRate** | 0.671 | 0.654 | 0.733 | 0.492 | 0.444 | 0.421 | 0.687 | 0.636 | **4.738** |
| **CIRate** | 0.229 | 0.241 | 0.338 | 0.228 | 0.145 | 0.161 | 0.223 | 0.281 | **1.846** |
| **N.Fl** | 0.588 | 0.549 | 0.597 | 0.449 | 0.348 | 0.328 | 0.477 | 0.548 | **3.883** |
| **Dis.** | 0.654 | 0.614 | 0.674 | 0.596 | 0.360 | 0.405 | 0.496 | 0.506 | **4.305** |
| **Reg.** | 0.709 | 0.677 | 0.757 | 0.647 | 0.438 | 0.535 | 0.651 | 0.523 | **4.937** |
| **Total (B)** | **4.486** | **4.420** | **5.043** | **4.056** | **3.023** | **3.267** | **4.428** | **4.211** | |

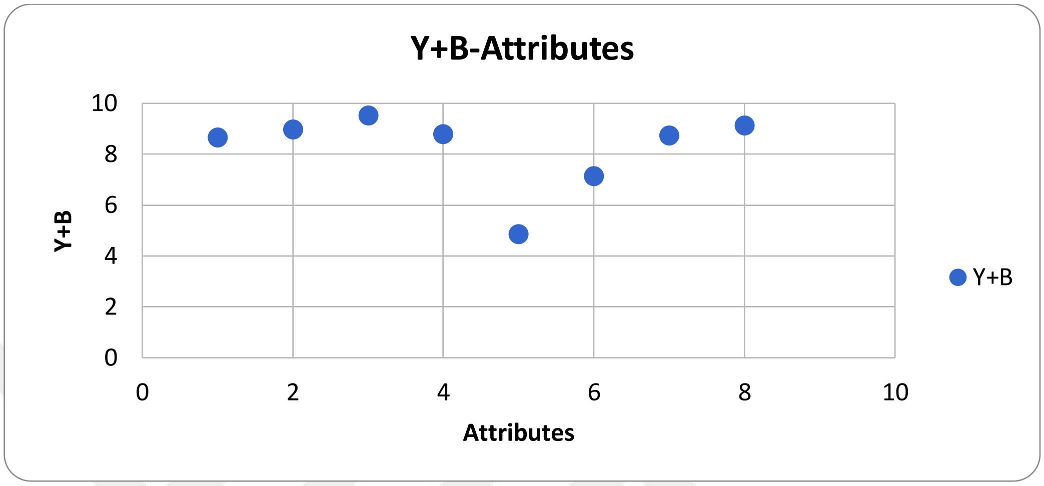
Phase 5:

Eventually, after achieved the matrix A, sum of influences given and received on criteria will be shown in Table 19.

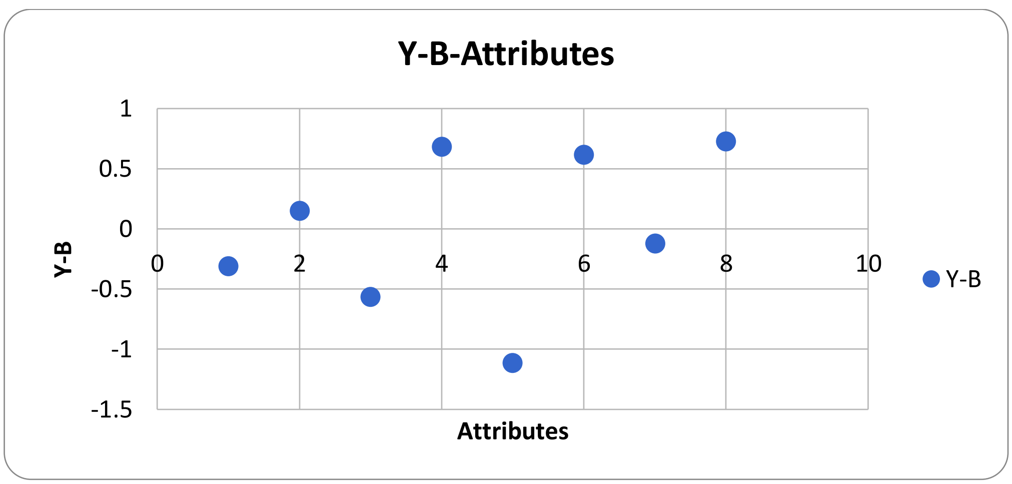
Table 19: Sum of Influences Given and Received on Criteria

| | Y | B | Y+B | Y-B |
|---|---|---|---|---|
| **Average Passenger (Number)** | 4.176 | 4.486 | 8.663 | -0.310 |
| **Baggage Quantity (Number)** | 4.571 | 4.420 | **8.991** | 0.151 |
| **Baggage Weight (Kg)** | 4.479 | 5.043 | **9.522** | -0.564 |
| **Y (Economy) Class Rate (%)** | 4.738 | 4.056 | 8.794 | **0.682** |
| **Infant + Child Passenger Rate (%)** | 1.846 | 3.023 | 4.869 | -1.177 |
| **Flights Number (Time)** | 3.883 | 3.267 | 7.151 | **0.616** |
| **Distance (Km)** | 4.305 | 4.428 | 8.733 | -0.124 |
| **Region** | 4.937 | 4.211 | **9.148** | **0.726** |

Moreover, Region factor also may be one of the most effective and also affected ones. The others comment derived from Graph 6 and 7.




Graph 6: Attributes Impulse Degree



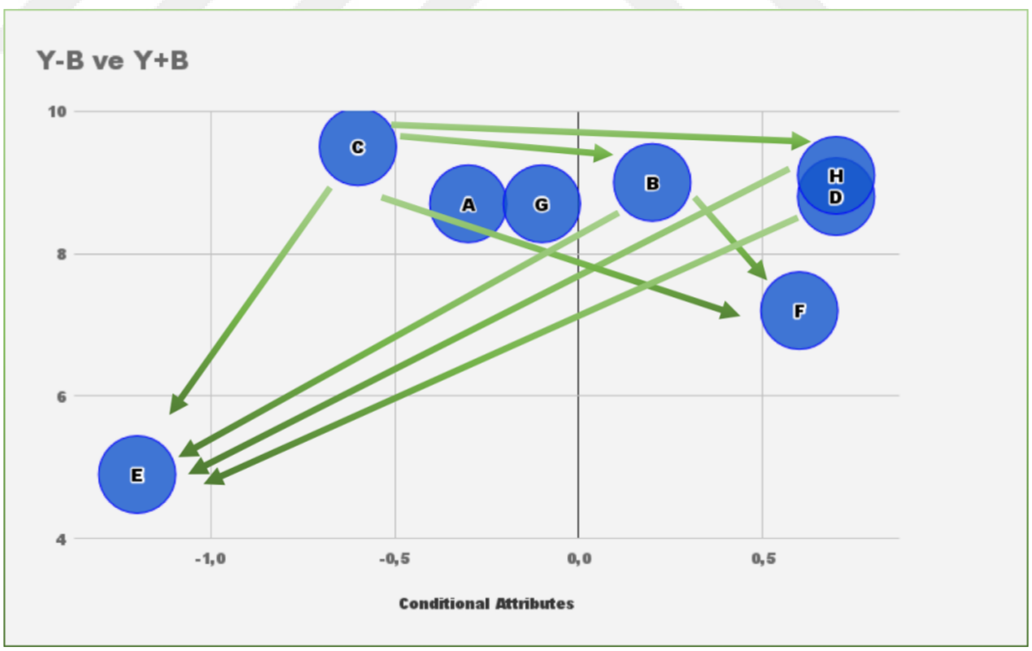

Graph 7: Attributes Affected Degree

The different demonstration of the Table 19 is influence network relation map (INRM) plotted using the total influence matrix A. In addition, with the choosing the bigger than the average method, the weighted super matrix and that significant values are shown below in Table 20. The weighted super matrix and that significant values are shown below in Table 20.

Table 20: Weighted Supermatrix with Average Method

| Conditional Attributes | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| **A** | | 0.584 | 0.661 | 0.564 | | | 0.604 | |
| **B** | 0.610 | | **0.719** | 0.521 | | 0.527 | 0.658 | 0.619 |
| **C** | 0.559 | 0.590 | 0.564 | 0.559 | | | 0.633 | 0.606 |
| **D** | 0.671 | 0.654 | **0.733** | | | | 0.687 | 0.636 |
| **E** | | | | | | | | |
| **F** | 0.588 | 0.549 | 0.597 | | | | | 0.548 |
| **G** | 0.654 | 0.614 | 0.674 | 0.596 | | | | |
| **H** | **0.709** | 0.677 | **0.757** | 0.647 | | 0.535 | 0.651 | 0.523 |

As a result of the prescriptive analytics, the conditional attributes' degrees of impacts have been identified. Moreover, predictive analytics results (cluster and rule outcomes) have been supported with DEMATEL technique. All those outcomes would give advices for decision-makers with meaningful results.




Graph 8: Evaluating Systems

Table 19 is illustrated in Graph 8. From the INRM, the direction of influence between dimensions and criteria can be visualized. The INRM indicates that Baggage Weight (C) is the most effective attributes can be understood from Graph 6.

### 1.9. Interpretation of Results

At this part, interpret study results will be shown and comment outcomes will be demonstrated. First, put forth different type of tables and multidimensional reports about civil aviation data. Diverse kind of tables can enhance and may increase in order to distinct situations. That graphs and tables can help the managers and decision-makers for seen together many different dimensions of data. On that tables showed that the company has different kind of passenger and flights profile. Depending on those outcome, decision-makers may resolve in order to operations. Furthermore, they can also see all different types of data together in Graph 2. To demonstrate, it can see easily all African flights in different hauls and also quarter by quarter.

Secondly, has tried to make classification and profiling for determine flights profile. Before that, it has done describe the data and inputs and then specified some data. For instance, passenger luggage number and weights has picked out and clustering done depending on that. After the profiling flights and determined cluster tried to associated rules regarding the cluster numbers. Those rules can also help decision-makers and especially strategy manager. To illustrate with numbers, Cluster 1 has less than 1 average baggage numbers, and most of the South Europe were in that cluster, so it can deduce that Cluster 1 flights are short generally.

Additionally, it can be understood different view of results from different tables. For example, if check the Graph 2, business rate of the America flights has reached about 11 percent, therefore the firm can increase the business class seat number or they may focus for the business passengers. Furthermore, one of the most salient touch is CI rate in United Kingdom, it is about twofold the average rate, however, it is not easy to tell something just with two-years data.

As a result, these outcomes can be an example of working in the field. Its improved the validity of results and showed all consequences frankly. Multidimensional reports can guidance not only civil aviation but also all service sector. Clustering and classification also can sample service sector Business intelligence applications. Association rules part on one hand can help the managers and on the other hand it can also be an example of predictive analytics. And also, DEMATEL has support the predictive results and put forth affect degree of the attributes. For example, depending derived Rules, Cluster 1 flights have less then 4.5 % child passenger rate in almost 79 % confidence. In addition, Cluster 1 has approximately 90 % confidentially eastern Europe flights. Hence, the flights planning or optimization departments can take into account those results. Consequently, owing to DEMATEL applications, it can understand that, Child rate is the one of the most affected attributes of that data.

# CONCLUSION & DISCUSSION

## 1.10. Conclusion and Discussion

Civil aviation is vital in many industries such as tourism. Regarding developing technology, business analytics process monitoring aims to forecast potential problems during process execution so that these problems can be handled proactively. On the other side, air transportation covers excessive different kinds of jobs and departments, like engineering issues, marketing, planning, micro/macro-economic issues and optimization. Several predictive monitoring techniques have been proposed in the past. However, those prediction techniques have been assessed independently from each other, making it hard to reliably compare their applicability and accuracy. Consequently, the analytics and the reporting have become crucial because aviation sector has developed rapidly via the use of information systems and they cannot be unconcerned with Business intelligence and all kinds of analytics.

One of the important Business intelligence phases is business analytics which includes descriptive, predictive and prescriptive. Airlines data analytics can improve ground operations, supporting faster turnaround times and better airspace management solutions which are driving efficiencies. They also can give airlines a critical profile of passengers to enable better and more personalized experiences for each passenger. Furthermore, the analytics foster brand loyalty, increase customer satisfaction, enable stronger auxiliary revenue stream and finally support scheduling/rebooking of passengers when delays occur. Propose of this study is an integrated model that combines descriptive analytics (multidimensional analytics), predictive analytics (data mining and more) and prescriptive analytics (MCDM and DEMATEL) in order to extract the critical factors for the improvement of airline baggage optimizations. It has applied for the dominance-based to extract the essential factors. To clarify, it has empirically analyzed and separately put forth that, the three different types of analytics with air transportation secondary data.

On this study, the data was taken from THY which is one of the 10 biggest airline companies in the world, and it has covered 2 years, 2014-2015. After the applied necessary preprocessing techniques for analytics, some meaningful results have been demonstrated. The results of the finding can be beneficial for aircraft and flight optimizations. The descriptive analytics consequences make an impression for profiling passengers and their behaviors. Especially regional analytics may serve for purposes of not only long-term but also short-term planning. In phase, the framework of the results obtained from this thesis should be use by airlines with current and difference analytics techniques and they should be adapted to evolve data technologies. These reports may have continuity and vitality; they can be evaluated weekly or monthly for efficient results. In addition, this study can expand by adding previous years data and in these way, it can be more effective for predicting next year situations so that projection of the past year may help next year flight planning.

Descriptive analytics results have set a precedent implication of multidimensional reports for service sector. In addition, rules that arise as outcomes of predictive analytics have really significant knowledge for marketing and planning department in civil aviation. For instance, the airline can make flight type arrangement depending on the rules 2, 5, 10 and more which are regarding to passenger baggage information. Because the airline has different types of aircraft which are also Furthermore, they will help to solve some optimization problems in air transportation sector. Owing to prescriptive analytics, displayed results supported by the MCDM and DEMATEL methods. Therefore, all stages of the analytics have been shown step by step on the real-world data implementations.

Briefly, for this study, it has been put forth big amount of data taken from the airline company. And necessary data preparing and cleaning process has completed. Then, multidimensional reports/outputs have produced which are can be useful for managerial and administrative stages. For instance, thanks to the reporting cubes, decision-makers can consider the different view of the data and may resolve and settle their operational issues. To demonstrate, they can easily see the different dimension of the data in several elements such as regions, time or distance. In addition, they can also focus on a range of units at any time or location. Predictive analytics has made with those data and meaningful rules were put forward with their outcomes so that they could be helpful for managers. Managers can use those rules for optimizing cargo, predicting passenger behaviors, checking the airports operations, making flights planning and designing projects for target marketing. Lastly, all those outcomes supported and the attributes evaluated with prescriptive analytics.

Emerging information systems and analytical research have increased productivity and profitability in the aviation sector. Large amounts of data about passenger and flight information accumulated over many years on commercial airlines which necessitate the use of Business analytics in different levels. Descriptive analytics studies for airlines with different kind of passenger or locations, will be beneficial for both marketing and flight planning departments. Furthermore, BI and analytics techniques are rapidly emerging, the use of data and data becomes even more crucial. Turkish Airlines has made great advancement in recent years, both economically and technologically; yet they should be more effective with analytics and Business intelligence which have been already used by major airlines.

Finally, there are some studies about Turkish Airlines in BI fields as it mentioned literature. However, those study can be regarding too with some analytics but they are not cover all the types which they are presented above. And also, THY is the best airline in Europe which flies to the most countries in the world. Because of the fact that, they have very different kind of passenger and flights thus they should profile them to better understand the situation and may give some right decisions. This study will support the airline with meaningful results and also it will fill in the gap in business analytics field academically. Consequently, civil aviation as a notable part of transportation is a growing and highly competitive sector. Moreover, automatically and manually accumulating information is impracticable due to the mass amount of data produced on each flight. Airlines have adapted several BI and analytics implementations in order to support their decision-making activities.

### 1.11. Limitations and Further Studies

This study has three main limitations and restriction which can be titling uncontrollable conditions data acquisition and data coverages. The external factors have changed the data on these years. Because of the THY is the flag carrier of the Turkey, they are most affected airline company depending on political and peripheral factors. According as geopolitical circumstance in middle east and north Africa, that data was so impressed. Especially civil war in Syria, Libya and Yemen, the company has closed more than 10 flight points which has include one of the busiest line like; Istanbul-Damascus.

Furthermore, terror attacks on different place and time in Turkey also affected air transportation in short term. That is also can be different study with these flights and passenger data. It didn't mention detailed because that topic also covers some political elements and public administration studies which are not close the work area. Furthermore, also some economic crisis and situation in neighborhood such as, Greece, also could affect the company flight slightly.

Moreover, the other nuisance of the study is changing company conditions and facilities. Because the airline company is one of the fastest growing airline not only in Europe but also all over the world. Therefore, their flights should normalize because they should add approximately 10 % new aircraft on those years. Hereby, number of flights has also increased depending on that number of passenger has risen. Therefore, depending those facilities, also checking these circumstances for not the misunderstood and do not make any calculation mistakes.

Lastly, some other external factors also affect always airlines and civil aviation sector such as weather condition and oil prices. Normally, they are looks very different fields but not for the airlines. Because weather is one of the crucial factor which has not affect only delay or cancelation also can cause different cost like de-icing or anti-icing which can use at the winter time for protect aircraft and ensure that completed the flights are safely. In addition, oil prices can manipulate by big company or some country, so it is also uncontrollable, but very affecting factor.

One of the most significant restraint is acquisition and attainment of the data. The aviation data gathering is not easy in Turkey or some developing country. In Turkey, it is very hard to attain this kind of secondary data which has gathered by company own system and acquire them. Especially it is difficult find data which has cover some financial record and passenger or employee personal data. Those compressive data which are include “data, flights code, arrival and departure airports, their cities countries and regions, number of passenger, different class passenger rate, rate of child and infant passenger, baggage numbers and weights, flights situation and capacities. In addition, because of the data has over 1 million rows, it’s difficult to handle and contemplate. The data are very crucial for the firms because of the security and privacy.

On the other hand, one of the lacks of this study was connected flights which include more than one leg flight. Because of the fact that, the data does not cover each passenger knowledge and it’s not possible to identify the passengers’ luggage for each individual. So, it could mislead the results and that is why the study didn’t include the connected flights.

In addition, in order to reduce the complexity of passenger variation in this study only international flights originating from Istanbul have been taken into consideration. If Istanbul flights originating from abroad were taken into consideration, the foreign passenger rate would be much higher because it would cover transit flights. Therefore, various evaluations can be made in different combinations of domestic and international lines.

As a result, depending on the data, very diverse study can be done. It can be examined monthly flights or passenger change or their other knowledge. Furthermore, it can also observe those situations effects which can be mentioned above. Increased of number of passenger can also investigate depending time series or taking into account seasonal changes. This study can be improved with the amount of data can expand like using 5 or 10 years data, but this state of affairs can cause with it some challenges.

# REFERENCES


Abdelghany, K., Abdelghany, A., ve Raina, S. (2005). A model for the airlines' fuel management strategies. Journal of Air Transport Management, 11(4), 199-206.

Acito, F., & Khatri, V. (2014). Business Analytics: Why now and what next?.

Adams, R. (2015). The ICAO Journal (3rd ed., Vol. 70, Ser. 2015, pp. 1-44, Rep.). Montreal: ICAO.

Adler, N., & Golany, B. (2001). Evaluation of deregulated airline networks using data envelopment analysis combined with principal component analysis with an application to Western Europe. European Journal of Operational Research, 132(2), 260-273.

Akerkar, R. (2014). Analytics on Big Aviation Data: Turning Data into Insights. IJCSA, 11(3), 116-127.

Akpinar, M. T., & Karabacak, M. E. (2017). Data mining applications in civil aviation sector: State-of-art review.

Akpınar, M. T., & Hızıroğlu, A. (2016). Sivil havacılık sektörü iş planlamasında bir iş zekası uygulaması. In V. Tecim (Ed.), Smart Technology and Smart Management (174-187). İzmir.

Anderberg, M. R. (2014). Cluster analysis for applications: probability and mathematical statistics: a series of monographs and textbooks (Vol. 19). Academic press.

Antonie, M. L., Zaiane, O. R., & Coman, A. (2001, August). Application of data mining techniques for medical image classification. In Proceedings of the Second International Conference on Multimedia Data Mining (pp. 94-101). Springer-Verlag.

Arslan, V. ve Yılmaz, G. (2010, July). Karar Destek Sistemlerinin Kullanımı İçin Uygun Bir Model Geliştirilmesi. Havacılık ve Uzay Teknolojileri Dergisi, 4(4), 75-82.

Avcı, T. ve Aktaş, M. (2015). Türkiye'de Faaliyet Gösteren Havalimanlarının Performanslarının Değerlendirilmesi. Uluslararası Alanya İşletme Fakültesi Dergisi, 7(3), 99-109.

Ayhan, S., Pesce, J., Comitz, P., Sweet, D., Bliesner, S., & Gerberick, G. (2013, April). Predictive analytics with aviation big data. In Integrated Communications, Navigation and Surveillance Conference (ICNS), 2013 (pp. 1-13). IEEE.

Baars, H., & Kemper, H. G. (2008). Management support with structured and unstructured data—an integrated business intelligence framework. Information Systems Management, 25(2), 132-148.

Babu, V.L.L., Batta, R., Lin, L., 2006. Passenger grouping under constant threat probability in an airport security system. Eur. J. Oper. Res. 168, 633e644.

Bandyopadhyay, D., & Sen, J. (2011). Internet of things: Applications and challenges in technology and standardization. Wireless Personal Communications, 58(1), 49-69.

Barai, S. K. (2003). Data Mining applications in transportation engineering. Transport, 18(5), 216-223.

Bazeley, P., & Jackson, K. (Eds.). (2013). Qualitative data analysis with NVivo. Sage Publications Limited.

Bédard, Y., Merrett, T., & Han, J. (2001). Fundamentals of spatial data warehousing for geographic knowledge discovery. Geographic data mining and knowledge discovery, 2, 53-73.

Berkhin, P. (2006). A survey of clustering data mining techniques. Grouping multidimensional data, 25, 71.

Bitzan, J., & Peoples, J. (2016). A comparative analysis of cost change for low-cost, full-service, and other carriers in the US airline industry. Research in Transportation Economics, 56, 25-41.

Boeing. (2015). Current market outlook 2014-2034 (pp. 1-58, Rep. No. 1). Seattle.

Bongo, M. F., Alimpangog, K. M. S., Loar, J. F., Montefalcon, J. A., & Ocampo, L. A. (2017). An application of DEMATEL-ANP and PROMETHEE II approach for air traffic controllers' workload stress problem: A case of Mactan Civil Aviation Authority of the Philippines. Journal of Air Transport Management.

Bramer, M. (2007). Principles of Data Mining (Vol. 180). London.: Springer.

Bratu, S., & Barnhart, C. (2005). An analysis of passenger delays using flight operations and passenger booking data. Air Traffic Control Quarterly, 13(1), 1-28.

Cadez, S., & Guilding, C. (2008). An exploratory investigation of an integrated contingency model of strategic management accounting. Accounting, organizations and society, 33(7), 836-863.

Cebeci, H. I., & Hiziroglu, A. (2016). Review of Business Intelligence and Intelligent Systems in Healthcare Domain. In Intelligent Techniques for Data Analysis in Diverse Settings (pp. 192-206). IGI Global.

Çetin, T., & Benk, S. (2011). Regulation, Deregulation, and Competition in the Turkish Airline Industry. In the Political Economy of Regulation in Turkey (pp. 193-214). Springer New York.

Chandrika, K. (2009). Scientific Data Mining: a practical perspective.

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. MIS quarterly, 36(4).

Chung, W., Chen, H., & Nunamaker, J. F. (2003, January). Business Intelligence explorer: a knowledge map framework for discovering Business Intelligence on the Web. In System Sciences, 2003. Proceedings of the 36th Annual Hawaii International Conference on (pp. 10-pp). IEEE.

Claessens, S., & Laeven, L. (2004). What drives bank competition? Some international evidence. Journal of Money, credit, and Banking, 36(3), 563-583.

Cleare, A. B. (2016). Business of Travel and Tourism in the 21st Century: A Caribbean Approach. Bahamas: Lulu.

Corn, J. J. (2002). The Winged Gospel: America's Romance with Aviation. JHU Press.

Cosmas, A., Love, R., Rajiwade, S., & Linz, M. (2013). Market clustering and performance of US OD markets. Journal of Air Transport Management, 28, 20-25.

Countries with The Highest Number Of Airline Passengers. (2016, July 25). Retrieved April 06, 2017, from http://www.worldatlas.com/articles/countries-with-the-highest-number-of-airline-passengers.html

Das, S., Matthews, B. L., Srivastava, A. N., & Oza, N. C. (2010, July). Multiple kernel learning for heterogeneous anomaly detection: algorithm and aviation safety case study. In Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and Data Mining (pp. 47-56). ACM.

Dax, E. C. (1975). Australia and New Zealand. World history of psychiatry, 704-728.

Delen, D. (2014). Real-world data mining: applied business analytics and decision making. FT Press.

Delen, D., & Demirkan, H. (2013). Data, information and analytics as services.

DGCA. (2017). Directorate general of civil aviation. Annual Report (Ser. 2016, pp. 1-128, Rep.). Ankara.

Dhar, V. (2013). Data science and prediction. Communications of the ACM, 56(12), 64-73.

Faber, V. (1994). Clustering and the continuous k-means algorithm. Los Alamos Science, 22(138144.21).

Fayyad, U. M., Piatetsky-Shapiro, G., & Smyth, P. (1996, August). Knowledge Discovery and Data Mining: Towards a Unifying Framework. In KDD (Vol. 96, pp. 82-88).

Fontela, E. and GABUS, A. (1976). The DEMATEL observer. Battelle Institute, Geneva Research Center.

Fontela, E., & GABUS, A. (1974). DEMATEL, innovative methods, Report no. 2, Structural analysis of the world problematique. Battelle Geneva Research Institute.

Forbes. (2017). The World's Biggest Public Companies. Retrieved April 06, 2017, from https://www.forbes.com/global2000/list/#header:revenue_sortreverse:true_industry:Airline

Frey, B. J., & Dueck, D. (2007). Clustering by passing messages between data points. science, 315(5814), 972-976.

Fuchs, M., Abadzhiev, A., Svensson, B., Höpken, W., & Lexhagen, M. (2013). A knowledge destination framework for tourism sustainability: A business intelligence application from Sweden. Turizam: znanstveno-stručni časopis, 61(2), 121-148.

Gartner, G., & Ortag, F. (2012). Advances in location-based services. Springer.

Ghoshal, S., Kim, S. K. (1986) "Building Effective Intelligence Systems for Competitive Advantage", Sloan Management Review, Vol. 28, No. 1, pp49-58.

Grzymala-Busse, J. W. (1997). A new version of the rule induction system LERS. Fundamenta Informaticae, 31(1), 27-39.

Gudmundsson, S. V. (1998). New-entrant airlines' life-cycle analysis: growth, decline and collapse. Journal of Air Transport Management, 4(4), 217-228.

Gürbüz, F., Özbakir, L., ve Yapici, H. (2011). Data Mining and preprocessing application on component reports of an airline company in Turkey. Expert Systems with Applications, 38(6)

Gürbüz, F., Özbakır, L., ve Yapıcı, H. (2009). Türkiye'de Bir Havayolu İşletmesine Ait Parça Söküm Raporlarına İlişkin Veri Madenciliği Uygulaması. Gazi Üniversitesi Mühendislik Ve Mimarlık Fakültesi Dergisi, 24(1), 73-78.

Gurjar, Y. S., & Rathore, V. S. (2013). Cloud Business Intelligence—is what business need today. International Journal of Recent Technology and Engineering, 1(6), 81-86.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (1998). Multivariate data analysis (Vol. 5, No. 3, pp. 207-219). Upper Saddle River, NJ: Prentice hall.

Han, J., & Kamber, M. (2000). Data Mining: concepts and techniques (the Morgan Kaufmann Series in data management systems).

Han, J., Pei, J., & Kamber, M. (2011). Data mining: concepts and techniques. Elsevier.

Hiziroglu, A. (2013). Soft computing applications in customer segmentation: State-of-art review and critique. Expert Systems with Applications, 40(16), 6491-6507.

Howson, C. (2007). Successful Business Intelligence. Tata McGraw-Hill Education

Hsieh, N. C. (2004). An integrated data mining and behavioral scoring model for analyzing bank customers. Expert systems with applications, 27(4), 623-633.

Hsu, C. I., Chao, C. C., & Shih, K. Y. (2012). Dynamic allocation of check-in facilities and dynamic assignment of passengers at air terminals. Computers & Industrial Engineering, 63(2), 410-417.

Huang, J. Y., Kao, Y. C., Lu, W. C., & Shieh, J. C. P. (2014, December). An inductive literature review of REITs by Data Mining methods. In Applied Engineering Sciences: Proceedings of the 2014 AASRI International Conference on Applied Engineering Sciences, Hollywood, LA, USA (Vol. 1, p. 161). CRC Press.

Huang, Z. (1998). Extensions to the k-means algorithm for clustering large data sets with categorical values. Data Mining and knowledge discovery, 2(3), 283-304.

Hui, G. W., Van Hui, Y., & Zhang, A. (2004). Analyzing China's air cargo flows and data. Journal of Air Transport Management, 10(2), 125-135.

Hui, G. W., Van Hui, Y., & Zhang, A. (2004). Analyzing China's air cargo flows and data. Journal of Air Transport Management, 10(2), 125-135.

IATA. (2016, June). Annual Review 2016, 72nd Annual General Meeting. http://www.iata.org/about/Documents/iata-annual-review-2015.pdf, (06.06.2016).

ICAO. (2016). Air transport, passengers carried. Retrieved April 05, 2017, from http://data.worldbank.org/indicator/IS.AIR.PSGR?end=2015&start=1970&view=chart

IDC. (2016). Worldwide Semiannual Software Tracker (240th ed., Vol. 25, Ser. 716, Tech.). Framingham, MA: IDC.

Irwin, D. A., & Pavcnik, N. (2004). Airbus versus Boeing revisited: international competition in the aircraft market. Journal of international economics, 64(2), 223-245.

Jacobson, B. F., Munster, M., Smith, A., Burnand, K. G., Carter, A., Abdool-Carrim, A. T. O., ... & Calvert-Evers, J. L. (2003). The best study-a prospective study to compare business class versus economy class air travel as a cause of thrombosis. South African Medical Journal, 93(7), 522-528.

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. Pattern recognition letters, 31(8), 651-666.

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a review. ACM computing surveys (CSUR), 31(3), 264-323.

Johnson, S. C. (1967). Hierarchical clustering schemes. Psychometrika, 32(3), 241-254.

Kamath, C. (2009). Scientific data mining: a practical perspective. Society for Industrial and Applied Mathematics. (pp. 127-132)

Kašparová, M., Křupka, J., & Chýlková, J. (2010). Comparison of cluster analysis algorithms for heavy metals contamination in atmospheric precipitation. WSEAS Transactions on Information Science and Applications, 7(3), 341-350.

Kasturi, E., Devi, S. P., Kiran, S. V., & Manivannan, S. (2016). Airline Route Profitability Analysis and Optimization Using BIG DATA Analyticson Aviation Data Sets under Heuristic Techniques. Procedia Computer Science, 87, 86-92.

Krishna, K., & Murty, M. N. (1999). Genetic K-means algorithm. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 29(3), 433-439.

Kuby, M. J., & Gray, R. G. (1993). The hub network design problem with stopovers and feeders: The case of Federal Express. Transportation Research Part A: Policy and Practice, 27(1), 1-12.

Kumar, A., Sah, B., Singh, A. R., Deng, Y., He, X., Kumar, P., & Bansal, R. C. (2017). A review of multi criteria decision making (MCDM) towards sustainable renewable energy development. Renewable and Sustainable Energy Reviews, 69, 596-609.

Lee, A. J., & Jacobson, S. H. (2011). The impact of aviation checkpoint queues on optimizing security screening effectiveness. Reliability Engineering & System Safety, 96(8), 900-911.

Leitner, P., Michlmayr, A., Rosenberg, F., & Dustdar, S. (2010, July). Monitoring, prediction and prevention of sla violations in composite services. In Web Services (ICWS), 2010 IEEE International Conference on (pp. 369-376). IEEE.

Li, C. W., & Tzeng, G. H. (2009). Identification of a threshold value for the DEMATEL method using the maximum mean de-entropy algorithm to find critical services provided by a semiconductor intellectual property mall. Expert Systems with Applications, 36(6), 9891-9898.

Lin, C. R., & Gerla, M. (1997). Adaptive clustering for mobile wireless networks. IEEE Journal on Selected areas in Communications, 15(7), 1265-1275.

Linoff, G. S., & Berry, M. J. (2011). Data mining techniques: for marketing, sales, and customer relationship management. John Wiley & Sons.

Liou, J. J. (2009). A novel decision rules approach for customer relationship management of the airline market. Expert systems with Applications, 36(3), 4374-4381.

Liou, J. J. (2011). Variable Consistency Dominance-based Rough Set Approach to formulate airline service strategies. Applied Soft Computing, 11(5), 4011-4020.

Liou, J. J., & Tzeng, G. H. (2010). A dominance-based rough set approach to customer behavior in the airline market. Information Sciences, 180(11), 2230-2238.

Liou, J. J., Yen, L., & Tzeng, G. (2010). Using decision rules to achieve mass customization of airline services. European Journal of Operational Research, 205(3), 680-686.

Lo, W. W. L., Wan, Y., & Zhang, A. (2015). Empirical estimation of price and income elasticities of air cargo demand: The case of Hong Kong. Transportation Research Part A: Policy and Practice, 78, 309-324.

Majeske, K. D., & Lauer, T. W. (2012). Optimizing airline passenger prescreening systems with Bayesian decision models. Computers & Operations Research, 39(8), 1827-1836.

Majeske, K.D., Lauer, T.W., 2012. Optimizing airline passenger prescreening systems with Bayesian decision models. Comput. Operations Res. 39, 1827e1836.

Marjanovic, O., Dinter, B., & Ariyachandra, T. (2017, January). Introduction to Organizational Issues of Business Intelligence, Business Analytics and Big Data Minitrack. In Proceedings of the 50th Hawaii International Conference on System Sciences.

McCallum, A., Nigam, K., & Ungar, L. H. (2000, August). Efficient clustering of high-dimensional data sets with application to reference matching. In Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 169-178). ACM.

McKenney, J. L., Copeland, D. C., & Mason, R. O. (1995). Waves of change: Business evolution through information technology. Harvard Business Press.

Mendoza, G. A., & Martins, H. (2006). Multi-criteria decision analysis in natural resource management: a critical review of methods and new modelling paradigms. Forest ecology and management, 230(1), 1-22.

Minelli, M., Chambers, M., & Dhiraj, A. (2012). Big data, big analytics: emerging business intelligence and analytic trends for today's businesses. John Wiley & Sons.

Moss, L. T., & Atre, S. (2003). Business Intelligence roadmap: the complete project lifecycle for decision-support applications. Addison-Wesley Professional, 124-125.

Negash, S. (2004). Business intelligence. The communications of the Association for Information Systems, 13(1), 54.

Nie, X. (2011). Risk-based grouping for checked baggage screening systems. Reliability Engineering & System Safety, 96(11), 1499-1506.

Notomista, G., Selvaggio, M., Sbrizzi, F., Di Maio, G., Grazioso, S., & Botsch, M. (2016). A fast airplane boarding strategy using online seat assignment based on passenger classification. Journal of Air Transport Management, 53, 140-149.

O'Kelly, M. E., & Miller, H. J. (1994). The hub network design problem: a review and synthesis. Journal of Transport Geography, 2(1), 31-40.

Okkonen, J., Pirttimäki, V., Hannula, M., & Lönnqvist, A. (2002, May). Triangle of Business Intelligence, Performance Measurement and Knowledge Management. In IInd Annual Conference on Innovative Research in Management, May 9-11, Stockholm, Sweden.

Olson, D. L., & Delen, D. (2008). Advanced Data Mining techniques. Springer Science & Business Media.

Olson, D. L., Shi, Y., & Shi, Y. (2007). Introduction to business Data Mining (Vol. 10, pp. 2250-2254). Englewood Cliffs: McGraw-Hill/Irwin.

Olszak, C. M., & Batko, K. (2012, September). The use of business intelligence systems in healthcare organizations in Poland. In Computer Science and Information Systems (FedCSIS), 2012 Federated Conference on (pp. 969-976). IEEE. Chicago

Oxley, D. (2015). Air Passenger Forecasts Global Report (pp. 1-37, Rep.). Geneva: IATA.

Pagels, David A. (2015). "Aviation Data Mining," Scholarly Horizons: University of Minnesota, Morris Undergraduate Journal: Vol. 2: Iss. 1, Article 3

Peker, İ, & B. B. (2009). Veri Zarflama Analizi İle Türkiye Havalimanlarında Bir Etkinlik Ölçümü Uygulaması. Ç.Ü. Sosyal Bilimler Enstitüsü Dergisi, 18(2), 72-88.

Pessoa, I. C., Trojan, F., Oliveira, G. A., & Setti, D. (2015). The statistical sampling about levels of utilization of multi-criteria methods to solve problems in POM. Negash, S., & Gray, P. (2008). Business intelligence. Handbook on decision support systems 2, 175-193.

Philbin, A. (2016). Traffic growth and airline profitability were highlights of air transport in 2016 //. Retrieved March 09, 2017, from http://www.icao.int/Newsroom/Pages/traffic-growth-and-airline-profitability-were-highlights-of-air-transport-in-2016.aspx

Piatetsky-Shapiro, G. (1996). Advances in knowledge discovery and Data Mining (Vol. 21). U. M. Fayyad, P. Smyth, & R. Uthurusamy (Eds.). Menlo Park: AAAI press.

Pirttimäki, V., Lönnqvist, A., & Karjaluoto, A. (2006). Measurement of Business Intelligence in a Finnish telecommunications company. The Electronic Journal of Knowledge Management, 4(1), 83-90.

Pohekar, S. D., & Ramachandran, M. (2004). Application of multi-criteria decision making to sustainable energy planning—a review. Renewable and sustainable energy reviews, 8(4), 365-381.

Point, T. (2017, November 06). Data Warehousing OLAP. Retrieved November 23, 2017, from https://www.tutorialspoint.com/dwh/dwh_olap.htm

Punj, G., & Stewart, D. W. (1983). Cluster analysis in marketing research: Review and suggestions for application. Journal of marketing research, 134-148.

Rasmussen, N. H., Goldy, P. S., & Solli, P. O. (2002). Financial Business Intelligence: trends, technology, software selection, and implementation. John Wiley & Sons.

Rijsenbrij, J. C., & Ottjes, J. A. (2007). New developments in airport baggage handling systems. Transportation Planning and Technology, 30(4), 417-430.

Roy, B. (2005) Paradigms and challenges. In J. Figueira, S. Greco, & M. Ehrgott (Eds.), Multiple-criteria decision analysis. state of the art surveys (pp. 3–24). New York, NY: Springer.

Roy, C., Weeks, S., Rouault, M., Nelson, G., Barlow, R., & Van der Lingen, C. (2001). Extreme oceanographic events recorded in the Southern Benguela during the 1999-2000 summer season. South African Journal of Science, 97(11-12), 465-471.

Russom, P. (2011). Big data analytics. TDWI best practices report, fourth quarter, 19, 40.

Sarkis, J., & Talluri, S. (2004). Performance based clustering for benchmarking of US airports. Transportation Research Part A: Policy and Practice, 38(5), 329-346.

Seufert, J. H., & Arjomandi, A. (2013). An Evaluation of the World's Major Airlines' Technical and Environmental Performance.

Shiao, G. C., & Hwang, C. C. (2013). Analyzing competition of international air cargo carriers in the Asian general air cargo markets. Transport policy, 27, 164-170.

Shiao, G. C., & Hwang, C. C. (2013). Analyzing competition of international air cargo carriers in the Asian general air cargo markets. Transport Policy, 27, 164-170.

SITA (Ed.). (2014). Air Transport Industry Insights, Smart Thinking, Working towards better Business Intelligence for airlines and airports (pp. 1-12, Rep.). SITA. doi:http://www.futuretravelexperience.com/wp-content/uploads/2014/04/smart-thinking.pdf

Skylife. (2013, July). The history of aviation. Retrieved April 06, 2017, from http://www.turkishairlines.com/en-kr/skylife/makaleler/2008/january/ucaklarin-tarihi

Smith, B.C., Gunther, D.P., Venkateshwara Roa, B., and Ratliff, R.M. (2001) "E-Commerce and Operations Research in Airline Planning, Marketing and Distribution," Interfaces, Vol. 31, No. 2, pp. 37–55.

Stacey, M., Salvatore, J., & Jorgensen, A. (2013). Visual Intelligence: Microsoft Tools and Techniques for Visualizing Data. John Wiley & Sons.

Stauffer, C., & Grimson, W. E. L. (1999). Adaptive background mixture models for real-time tracking. In Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on. (Vol. 2, pp. 246-252). IEEE.

Sun, J. Y. (2015). Clustered airline flight scheduling: Evidence from airline deregulation in Korea. Journal of Air Transport Management, 42, 85-94.

Tanguy, L., Tulechki, N., Urieli, A., Hermann, E., & Raynal, C. (2016). Natural language processing for aviation safety reports: From classification to interactive analysis. Computers in Industry, 78, 80-95.

Tansley, S., & Tolle, K. M. (2009). The fourth paradigm: data-intensive scientific discovery (Vol. 1). T. Hey (Ed.). Redmond, WA: Microsoft research.

Thomas, A. (2016, December 28). Statistics of Lost Luggage. Retrieved April 03, 2017, from http://trackimo.com/statistics-of-lost-luggage/

Thomas, S., & Sarawagi, S. (1998, August). Mining Generalized Association Rules and Sequential Patterns Using SQL Queries. In KDD (pp. 344-348).

Trochim, W. M., & Donnelly, J. P. (2001). Research methods knowledge base.

TZENG, G.H., CHIANG, Ch.H. & LI, Ch.W. (2007). Evaluating intertwined effects in elearning programs: a novel hybrid MCDM model based on factor analysis and DEMATEL. Expert Systems with Applications, 32 (4), 1028–1044.

Van der Aalst, W. M. (2011). Data Mining. In Process Mining (pp. 59-91). Springer Berlin Heidelberg.

Vercellis, C. (2010). Business Intelligence: Data Mining and optimization for decision making. Chichester: John Wiley & Sons.

Vesset, D. (Ed.). (2016, July). Worldwide Business Analytics Software Market Shares, 2015: Healthy Demand Despite Currency Exchange Rate Headwinds (Rep.). Retrieved April 11, 2017, from IDC website: https://www.sas.com/content/dam/SAS/en_us/doc/analystreport/idc-business-analytics-software-market-shares-108014.pdf

Vogt, W. P., & Johnson, R. B. (2011). Dictionary of Statistics & Methodology: A Nontechnical Guide for the Social Sciences: A Nontechnical Guide for the Social Sciences. Sage.

Wang, R., Lin, Y. H., & Tseng, M. L. (2011). Evaluation of customer perceptions on airline service quality in uncertainty. Procedia-Social and Behavioral Sciences, 25, 419-437.

Wen, C. H., Chen, T. N., & Fu, C. (2014). A factor-analytic generalized nested logit model for determining market position of airlines. Transportation Research Part A: Policy and Practice, 62, 71-80.

Wen, C. H., Tsai, M. C., & Lin, C. H. (2011). Classification and competition analysis of air cargo logistics providers: The case of Taiwan's high-technology industry. Journal of Air Transport Management, 17(2), 106-109.

Worldfolio, T. (2016, April). World's biggest airport under construction in Istanbul. Retrieved April 06, 2017, from http://www.theworldfolio.com/interviews/worlds-biggest-airport-under-construction-in-istanbul/3796/

Wu, C., & Truong, T. (2014). Improving the IATA delay data coding system for enhanced data analytics. Journal of Air Transport Management, 40, 78-85.

Wu, D. D., Chen, S. H., & Olson, D. L. (2014). Business intelligence in risk management: Some recent progresses. Information Sciences, 256, 1-7.

Wu, M. (2014, February). Smarter data. Admap Magazine. Retrieved April 22, 2017, from https://www.warc.com/Topics/Admap.topic

WU, W.W., and LEE, Y.T. (2007). Developing global managers' competencies using the fuzzy DEMATEL method. Expert Systems with Applications, 32 (2), 499–507.

Yan, S., Shih, Y. L., & Shiao, F. Y. (2008). Optimal cargo container loading plans under stochastic demands for air express carriers. Transportation Research Part E: Logistics and Transportation Review, 44(3), 555-575.

Yoon, S. H., & Park, J. W. (2015). A study of the competitiveness of airline cargo services departing from Korea: Focusing on the main export routes. Journal of Air Transport Management, 42, 232-238.

Yoon, S. H., & Park, J. W. (2015). A study of the competitiveness of airline cargo services departing from Korea: Focusing on the main export routes. Journal of Air Transport Management, 42, 232-238.

Yurtay, Y. (2014). Konteynerde Yük Optimizasyonu: Örnek Uygulama. Jmer Journal of Management and Economics Research, (23), 228-228.

Zavadskas, E. K., Turskis, Z., & Kildienė, S. (2014). State of art surveys of overviews on MCDM/MADM methods. Technological and economic development of economy, 20(1), 165-179.

Zhao, W., Ma, H., & He, Q. (2009, December). Parallel k-means clustering based on mapreduce. In IEEE International Conference on Cloud Computing (pp. 674-679). Springer, Berlin, Heidelberg.

Zhou, Z. H., Chawla, N. V., Jin, Y., & Williams, G. J. (2014). Big data opportunities and challenges: Discussions from data analytics perspectives [discussion forum]. IEEE Computational Intelligence Magazine, 9(4), 62-74.

Azoff, E. M. (1994). Neural network time series forecasting of financial markets. John Wiley & Sons, Inc..

Famili, A., Shen, W. M., Weber, R., & Simoudis, E. (1997). Data preprocessing and intelligent data analysis. Intelligent data analysis, 1(1-4), 3-23.

Ye, N. (Ed.). (2003). The handbook of data mining (Vol. 24). Mahwah, NJ/London: Lawrence Erlbaum Associates, Publishers.

Dowie, J. (1993). Clinical decision analysis: background and introduction. Analyzing how we reach clinical decisions. RCP Publications, 7-25.

Barrett, M., Davidson, E., Prabhu, J., & Vargo, S. L. (2015). Service innovation in the digital age: key contributions and future directions. MIS quarterly, 39(1), 135-154.

Gangadharan, G. R., & Swami, S. N. (2004, June). Business intelligence systems: design and implementation strategies. In Information Technology Interfaces, 2004. 26th International Conference on (pp. 139-144). IEEE.

Kimball, R., & Caserta, J. (2011). The Data Warehouse? ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data. John Wiley & Sons.

Chen, M. S., Han, J., & Yu, P. S. (1996). Data mining: an overview from a database perspective. IEEE Transactions on Knowledge and data Engineering, 8(6), 866-883.

Han, J., Pei, J., & Kamber, M. (2011). Data mining: concepts and techniques. Elsevier.

Lim, T. S., Loh, W. Y., & Shih, Y. S. (2000). A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms. Machine learning, 40(3), 203-228.

Moorthy, K., Vincent, C., & Darzi, A. (2005). Simulation based training: is being extended from training individuals to teams. BMJ: British Medical Journal, 330(7490), 493.

Park, J. W., Robertson, R., & Wu, C. L. (2004). The effect of airline service quality on passengers' behavioural intentions: a Korean case study. Journal of Air Transport Management, 10(6), 435-439.

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a review. ACM computing surveys (CSUR), 31(3), 264-323.

Deshpande, P. M., Ramasamy, K., Shukla, A., & Naughton, J. F. (1998, June). Caching multidimensional queries using chunks. In ACM SIGMOD Record (Vol. 27, No. 2, pp. 259-270). ACM.

## Annex 1

Table 21: Cluster Iteration History from Analysis

**Iteration History**

| | Change in Cluster Centers | | |
|---|---|---|---|
| **Iteration** | 1 | 2 | 3 |
| **1** | 11.276 | 10.834 | 11.240 |
| **2** | .079 | .249 | 9.846 |
| **3** | .030 | .657 | 5.105 |
| **4** | .124 | .739 | 2.655 |
| **5** | .164 | .651 | 1.575 |
| **6** | .159 | .512 | 1.007 |
| **7** | .141 | .402 | .729 |
| **8** | .116 | .313 | .566 |
| **9** | .094 | .237 | .408 |
| **10** | .072 | .181 | .333 |
| **11** | .059 | .146 | .275 |
| **12** | .046 | .112 | .210 |
| **13** | .035 | .085 | .161 |
| **14** | .029 | .069 | .132 |
| **15** | .023 | .056 | .111 |
| **16** | .019 | .044 | .085 |
| **17** | .015 | .035 | .066 |
| **18** | .011 | .026 | .051 |
| **19** | .009 | .022 | .041 |
| **20** | .007 | .018 | .039 |
| **21** | .006 | .015 | .029 |
| **22** | .006 | .013 | .026 |
| **23** | .005 | .011 | .021 |
| **24** | .004 | .008 | .014 |
| **25** | .003 | .007 | .010 |
| **26** | .003 | .005 | .009 |
| **27** | .002 | .003 | .005 |
| **28** | .001 | .003 | .005 |
| **29** | .001 | .002 | .003 |
| **30** | .001 | .002 | .003 |
| **31** | .001 | .001 | .001 |
| **32** | .000 | .001 | .001 |

## Annex 2

Table 22: THY Flights and Passengers Data Analysis by Clusters Distance

**Distances between Final Cluster Centers**

| Cluster | 1 | 2 | 3 |
|---|---|---|---|
| **1** | | 6.527 | 19.086 |
| **2** | 6.527 | | 12.560 |
| **3** | 19.086 | 12.560 | |