



TUGAS AKHIR – TI 141501

**ANALISA PERBANDINGAN DARI PERILAKU HARGA
TIKET GARUDA AIRLINES ANTARA WAKTU
KEBERANGKATAN PEAK DAN NON-PEAK BERDASARKAN
HARI PEMBELIAN**

ADELIA STEPHANIE ZERLINDA
NRP 2512 100 087

Pembimbing
Dr. Eng. Ir. Ahmad Rusdiansyah, M.Eng, CSCP
NIP. 19681109 199503 1 003

JURUSAN TEKNIK INDUSTRI
FAKULTAS TEKNOLOGI INDUSTRI
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2016



FINAL PROJECT – TI 141501

**COMPARISON ANALYSIS OF PRICING BEHAVIOUR OF
GARUDA AIRLINES TICKET PRICES BETWEEN PEAK AND
NON-PEAK DEPARTURE TIME BY DAYS OF PURCHASE**

ADELIA STEPHANIE ZERLINDA
NRP 2512 100 087

Supervisor
Dr. Eng. Ir. Ahmad Rusdiansyah, M.Eng, CSCP
NIP. 19681109 199503 1 003

DEPARTMENT OF INDUSTRIAL ENGINEERING
FACULTY OF INDUSTRIAL TECHNOLOGY
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2016

APPROVAL SHEET

COMPARISON ANALYSIS OF PRICING BEHAVIOUR OF GARUDA AIRLINES TICKET PRICES BETWEEN PEAK AND NON-PEAK DEPARTURE TIME BY DAYS OF PURCHASE

FINAL PROJECT

Proposed to fulfill the requirement to get Industrial Engineering Bachelor Degree
Industrial Engineering Department
Faculty of Industrial Technology
Institut Teknologi Sepuluh Nopember

By :

ADELIA STEPHANIE ZERLINDA

NRP 2512 100 087

Approved by Supervisor :



Dr. Eng. Ir. Ahmad Rusdiansyah, M.Eng, CSCP

NIP. 19681109 199503 1 003



**ANALISA PERBANDINGAN DARI PERILAKU HARGA TIKET
GARUDA AIRLINES ANTARA WAKTU KEBERANGKATAN PEAK
DAN NON-PEAK BERDASARKAN HARI PEMBELIAN**

Penulis : Adelia Stephanie Zerlinda
NRP : 2512.100.087
Jurusan : Department of Industrial Engineering, ITS
Pembimbing : Dr. Eng. Ir. Ahmad Rusdiansyah, M.Eng, CSCP

ABSTRAK

Globalisasi dan pertumbuhan teknologi yang cepat membawa banyak dampak positif pada kehidupan manusia. Salah satu contoh dampak positifnya adalah perkembangan teknologi transportasi. Perkembangan transportasi membuat manusia semakin mudah untuk berpergian. Salah satu bentuk perkembangan transportasi adalah adanya transportasi udara yang mampu menempuh jarak jauh dengan cepat dan nyaman. Menurut catatan dari Badan Pusat Statistic, penggunaan transportasi udara terus meningkat dari tahun 1999 sampai 2014, baik penerbangan dalam maupun luar negeri. Kenyamanan menggunakan transportasi udara juga didukung dengan mudahnya mendapatkan tiket maskapai. Saat ini, calon penumpang dapat dengan mudah membeli tiket penerbangan melalui biro perjalanan secara (online) melalui website mereka. Bahkan saat ini, tidak tiket penerbangan yang bisa didapat dari biro perjalanan, melainkan juga berbagai hal lain. Penelitian ini berfokus pada data tiket penerbangan maskapai dari biro perjalanan (online) untuk mendapatkan informasi lain yang berguna untuk pengguna jasa penerbangan. Informasi ini didapat dengan 2 cara yaitu descriptive analysis dan multiple regression. Descriptive analysis digunakan untuk mengetahui dan membandingkan perilaku/karakteristik data pada hari Senin hingga Kamis untuk maskapai Garuda Indonesia. Software R-Project digunakan dalam menyelesaikan multiple regression untuk mengetahui pengaruh dari perubahan harga suatu waktu penerbangan ke waktu penerbangan yang lain. Penelitian ini dapat mengidentifikasi pola dan pengaruh dari penetapan harga tiket untuk setiap waktu keberangkatan ke waktu keberangkatan lain. Harga tiket pada waktu keberangkatan non-peak berubah mengikuti harga tiket pada waktu keberangkatan peak namun harga tiket pada waktu keberangkatan peak berubah tidak mengikuti harga tiket pada waktu keberangkatan non-peak.

Keywords: Airline Pricing Behaviour, Airfare Mining, Descriptive Analysis, Multiple Regression, R-Project.

**COMPARISON ANALYSIS OF PRICING BEHAVIOUR OF GARUDA
AIRLINE TICKET PRICES BETWEEN PEAK AND NON-PEAK
DEPARTURE TIME BY DAYS OF PURCHASE**

Author : Adelia Stephanie Zerlinda
Student ID : 2512.100.087
Department : Department of Industrial Engineering, ITS
Supervisor : Dr. Eng. Ir. Ahmad Rusdiansyah, M.Eng, CSCP

ABSTRACT

The globalization and rapid development of technology bring many positive impacts to the human life. One of the examples is the development of transportation facilities. The development of transportation facilities makes people easy to travel. One of the example is the develepment of air transportation facilities as a transportation mode that can facilitates people with fast moving and comfortable transportation. The statistics recorded by Badan Pusat Statistic shows that the flight demand trend from year 1999 to 2014 is increasing for both domestic and international flights. Furthermore, the development of technology also makes it easier to obtain the airline tickets. Nowadays, people can easily purchase the airlines tickets trough online travel agent websites which is very helpful for human life. However, there are many things that could be obtained from the online travel agent websites, not only the airline tickets. This research focused on mining the airline ticket prices data from online travel agent website to find some useful information for the customer. The useful information is obtained through two ways, which are descriptive analysis and multiple regression. The descriptive analysis is done to know the behaviour of the data and compare the data behaviour of Monday and Thursday for Garuda Indonesia Airlines. The multiple regression is done by using R-Project software to know whether there is influence of each departure time's price changes to other departure time's price changes. A departure time as the dependent variable of the multiple regression is regressed to the lag departure times as the independent variables. From this research, the behaviour of Monday and Thursday flights by days of purchase and the influence of each departure time's price changes to other departure time's price changes are identified. The non-peak departure time price changes correspond to the peak departure time prices but the peak departure time price changes incorrespond to the non-peak departure prices.

Keywords: Airline Pricing Behaviour, Airfare Mining, Descriptive Analysis, Multiple Regression, R-Project.

TABLE OF CONTENTS

APPROVAL SHEET	i
DISCLAIMER.....	ii
ABSTRACT	iii
ABSTRAK	iv
ACKNOWLEDGEMENT	v
TABLE OF CONTENTS.....	vii
LIST OF TABLES.....	ix
LIST OF FIGURES	xi
CHAPTER 1 INTRODUCTION.....	1
1.1 Background.....	1
1.2 Problem Formulation.....	6
1.3 Research Objectives	6
1.4 Research Contribution.....	6
1.5 Scope of Research	7
1.5.1 Boundaries	7
1.5.2 Assumptions.....	7
1.6 Research Outline	7
CHAPTER 2 LITERATURE REVIEW	9
2.1 Airlines Business Model.....	9
2.1.1 Full Service Business Model.....	9
2.1.2 Low Cost Carrier Business Model	11
2.1.3 Charter Carrier Business Model.....	14
2.2 Revenue Management	14
2.2.1 Tactics for Price Differentiation.....	16
2.2.2 Airlines Revenue Management Strategy	20
2.3 Business Statistics	22
2.3.1 Regression Analysis	23
2.3.2 Multiple Regression Analysis	24
CHAPTER 3 RESEARCH METHODOLOGY.....	33

3.1	Research Methodology Flowchart	33
3.2	Explanation of Research Methodology Flowchart.....	34
3.2.1	Data Collection.....	35
3.2.2	Model Development	35
3.2.3	Numerical Experiment.....	35
3.2.4	Analysis.....	35
3.2.5	Conclusions and Recommendations	36
CHAPTER 4 DATA COLLECTION AND NUMERICAL EXPERIMENT		37
4.1	Existing condition	37
4.2	Data Collection	38
4.3	Numerical Experiment.....	45
4.3.1	Descriptive Analysis	45
4.3.2	Multiple Regression.....	50
CHAPTER 5 INTERPRETATION AND ANALYSIS		57
5.1	Descriptive Analysis Interpretation.....	57
5.2	Multiple Regression Interpretation and Analysis.....	58
5.2.1	Garuda Indonesia based on Timely Flight Groups.....	58
5.2.2	Garuda Indonesia (Morning Flights) based on Lag Period.....	76
CHAPTER 6 CONCLUSIONS AND RECOMMENDATIONS		81
6.1	Conclusions.....	81
6.2	Recommendations	82
REFERENCES		83

LIST OF TABLES

Table 1.1 Number of Passenger Departures from Indonesia Airports.....	1
Table 2.1 Cost Gap Between LCC and FSC.....	13
Table 2.2 The Three Levels of Revenue Management Decisions.....	15
Table 2.3 Characteristics of Leisure vs. Business Customers.....	20
Table 2.4 Typical Airline Market Segmentation.....	21
Table 4.1 Classification of Garuda Indonesia Airlines on Departure Date December 14th, 2015.....	50
Table 4.2 Classification of Garuda Indonesia Airlines on Departure Date December 17th, 2015.....	52
Table 4.3 Clustering Result on Departure Date December 14th, 2015.....	56
Table 4.4 Clustering Result on Departure Date December 17th, 2015.....	56
Table 5.1 Regression Summary for Morning Flight Group.....	59
Table 5.2 Significance Code and Multiple R-Squared Regression Summary for Morning Flight Group.....	61
Table 5.3 Regression Summary for Noon Flight Group.....	63
Table 5.4 Significance Code and Multiple R-Squared Regression Summary for Noon Flight Group.....	64
Table 5.5 Regression Summary for Afternoon Flight Group.....	66
Table 5.6 Significance Code and Multiple R-Squared Regression Summary for Afternoon Flight Group.....	67
Table 5.7 Regression Summary for Evening Flight Group.....	69
Table 5.8 Significance Code and Multiple R-Squared Regression Summary for Evening Flight Group.....	70
Table 5.9 Regression Summary for Monday All Flight Groups with Lag t-1.....	75
Table 5.10 Regression Summary for Thursday All Flight Groups with Lag t-1...76	
Table 5.11 Regression Summary for Monday Morning with Lag t-1, t-2, t-3, t-4 and t-5.....	77
Table 5.12 Regression Summary for Thursday Morning with Lag t-1, t-2, t-3, t-4 and t-5.....	79

LIST OF FIGURES

Figure 1.1 Common Purchased Online Products by Indonesian People.....	4
Figure 1.2 Change Over Time for United Airlines Roundtrip Flight #168:169 LAX-BOS Departing on January 12.....	5
Figure 2.1 The Role of Business Statistics.....	22
Figure 2.2 Multiple Regression Hyperplane for Population.....	26
Figure 3.1 Research Methodology Flowchart.....	34
Figure 4.1 Garuda Indonesia Airlines Ticket Price Changes on December 14th, 2015.....	39
Figure 4.2 Garuda Indonesia Airlines Ticket Price Changes on December 17th, 2015.....	40
Figure 4.3 Garuda Indonesia Airlines Ticket Price Changes of Monday and Thursday Departure Day for Morning Flights.....	41
Figure 4.4 Garuda Indonesia Airlines Ticket Price Changes of Monday and Thursday Departure Day for Noon Flights.....	42
Figure 4.5 Garuda Indonesia Airlines Ticket Price Changes of Monday and Thursday Departure Day for Afternoon Flights.....	43
Figure 4.6 Garuda Indonesia Airlines Ticket Price Changes of Monday and Thursday Departure Day for Evening Flights.....	44
Figure 4.7 Garuda Indonesia Airlines Ticket Price Regression Result of Class 1 for Departure Date December 14th, 2015.....	47
Figure 4.8 Garuda Indonesia Airlines Ticket Price Regression Result of Class 2 for Departure Date December 14th, 2015.....	47
Figure 4.9 Garuda Indonesia Airlines Ticket Price Regression Result of Class 1 for Departure Date December 14th, 2015.....	48
Figure 4.10 Garuda Indonesia Airlines Ticket Price Regression Result of Class 1 for Departure Date December 17th, 2015.....	49
Figure 4.11 Garuda Indonesia Airlines Ticket Price Regression Result of Class 2 for Departure Date December 17th, 2015.....	49

Figure 4.12 Garuda Indonesia Airlines Ticket Price Regression Result of Class 3 for Departure Date December 17th, 2015.....	50
Figure 4.14 R-Project Coding for Monday Flight.....	53
Figure 4.15 R-Project Regression Result for Monday Morning Flight.....	54
Figure 5.1 Garuda Indonesia Airlines Ticket Price Regression Result Summary.....	57
Figure 5.2 Graphical Mapping of Monday Regression Result.....	72
Figure 5.3 Graphical Mapping of Thursday Regression Result.....	73

CHAPTER 1

INTRODUCTION

The primary aspects of this research will be explained in this chapter through several subchapters. The primary aspects of this research are background, problem formulation, research objectives, research benefits, scope of research and research outline.

1.1 Background

The globalization and rapid development of technology bring many positive impacts to the human life. One of the examples is the development of transportation facilities. The development of transportation facilities makes people easy to travel. Nowadays, people not only talking about domestic destinations, but also international destinations. It is not only for business but also for study purpose or even for a holiday. The fastest and comfortable way to travel around the world is by using air transportation mode.



Figure 1.1 Number of Passenger Departures from Indonesia Airports (source: BPS, 2016)

It can be seen in figure 1.1 above that the flight demand trend from year 1999 to 2014 is increasing for both domestic and international flights. This means

that the people need to travel fast are increased. The increasing of flight demand means that people are willing to change their transportation mode preference into air transportation mode, which is believed to be fastest and comfortable to travel.

The reason behind the increasing of flight demand is not only because of the fast moving and comfortable aspects of air transportation mode, the safety aspect is also one of the reason. In case of air transportation mode's safety, Indonesia has a very good aviation and flight safety. Based on the article released by Indonesian Ministry of Transportation, the latest score of Indonesia aviation safety by International Civil Aviation Organization (ICAO) is 93.6 which is considered as high aviation safety (hubud.dephub.go.id, 2016).

The other reason that might influence the increasing of flight demand is the airfare. There are many Low Cost Carrier (LCC) airlines operating in Indonesia. Currently, the listed operating airlines that departure from Juanda International Airport (SUB) to Soekarno-Hatta International Airport (CGK) are AirAsia Indonesia, Sriwijaya Air, Lion Air, Citilink, Batik Air, and Garuda Indonesia Airlines (traveloka.com, 2016). There are four LCC airlines, which are AirAsia Indonesia, Sriwijaya Air, Lion Air, and Citilink, more than the Full Service (FSC) airlines, which are Batik Air and Garuda Indonesia Airlines. The number of flights by LCC airlines is also more than the number of flights by FS airlines. Based on one of the online travel agent, for the flight from SUB to CGK, the LCC airlines have total flights around forty eight flights a day while the FSC airlines have total flights around thirty four flights a day, lower than the LCC airlines (traveloka.com, 2015). This statistic has given more opportunity for the passengers to travel less expensive by using LCC airlines, even though the FSC airlines has given opportunity to travel more comfortable. Based on the aviation news, it is stated that currently the economic flight airfares has decreased about five percents due to the decreasing price of avtur and the strengthening of the exchange rate of Rupiah to US Dollar (infopenerbangan.com, 2016).

Airfare is one of the most interesting topics to be discussed in aviation. Airlines might have many fare-classes for the seats in a flight (economy, business, and first class) and the airfare frequently vary over time based on some factors including seasonal time (high season, shoulder season, and low season), the

availability of seats, competitive strategy from other airlines, and more. The complex policies are often used by the airlines to vary the airfare over the time. One of the most sophisticated strategy that is used by the airlines to maximize its revenue is dynamic pricing strategy (Etzioni et al., 2003). The airlines might also use different sales channels (e.g. consolidators, travel agents) to increase the sales of the flight tickets.

As the development of technology, the use of internet is also increased. This phenomenon has influenced the global economic condition. Almost everyone in the world has known how to use and utilize internet to help human activity. There are many activities that can be done by using internet, such as using social media, browsing website, selling and buying products, and more. Based on the survey done by Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) on 2014, the products that are usually bought using internet. The products can be seen in figure 1.2 below.



Figure 1.2 Common Purchased Online Products by Indonesian People (source: Asosiasi Penyelenggara Jasa Internet Indonesia, 2015)

By knowing the commonly purchased products online, as showed in figure 1.2, the business runner can enhance its promotion strategy to be aligned with the development of technology. Many companies has changed their sales promotion strategy into an online system, which is using the internet as the transaction tool. Nowadays, the flight tickets can be easily booked and purchased through the online travel agents. This is one of the strategies that is used by airlines to increase their sales. This strategy can be beneficial for the airlines but also dangerous if the airlines can not manage their pricing strategy well. The consumer can easily look for the best airline ticket prices by comparing the price offered by all the airlines. This strategy is also advantageous for the consumer because they can purchase the tickets easily everywhere and every time.

Even though the airline tickets become increasingly available online, the consumers have to be smart so they can purchase the airline tickets efficiently. The consumer has an opportunity to be more sophisticated shoppers, understanding how the airlines vary the process over time. The consumers can become smart shoppers by tracking price over time, identifying airlines pricing patterns so they can make a decision to rush or delay purchasing the tickets based on anticipated price changes. However, proprietary algorithms and hidden variables (e.g. the number of unsold seats on a flight, overbooking strategy, and more) are definitely used by the airlines in changing ticket prices (Etzioni et al., 2003). The example of the pricing behavior can be seen in figure 1.3 below.

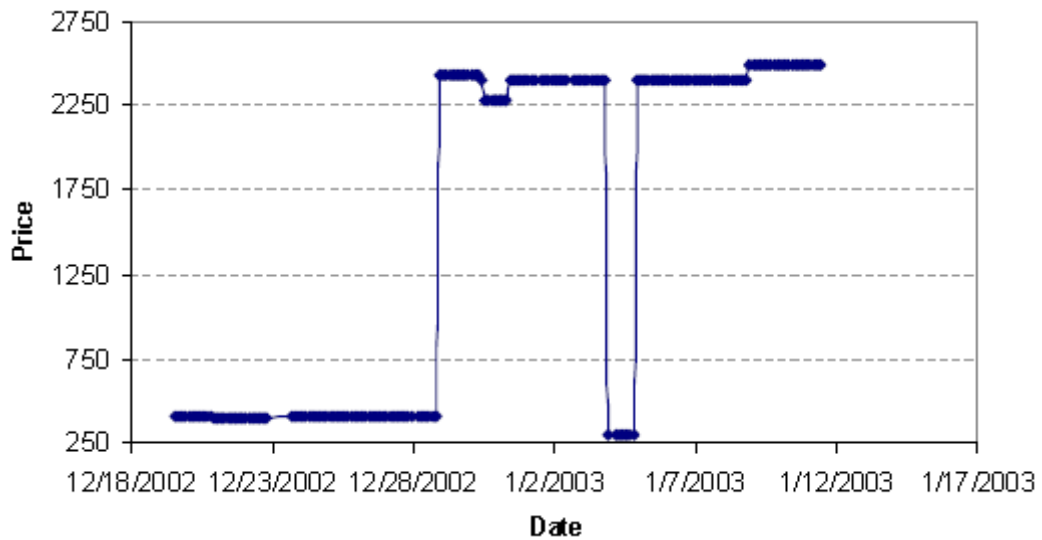


Figure 1.3 Price Change Over Time for United Airlines Roundtrip Flight #168:169 LAX-BOS Departing on January 12 (Etzioni et al., 2003).

The pricing behaviour that is called two price tiers is showed in figure 1.3 above. Two price tiers means that there are two price range over time that is considered stable separated by a rapid price drop in between. The number of tiers typically varies from two to four, depending on the airline and the particular flight (Etzioni et al., 2003).

Statistics may be defined as the procedures and techniques that comprise business statistics include those specially designed to describe data, such as charts, graphs, and numerical measures, and also includes inferential procedures that help decision makers draw inferences from a set of data (Groebner et al., 2010). Regression is one of the statistics tools that can be used in processing the data, in this case, multiple regression can be used to analyze the airline ticket pricing patterns based on the purchasing date. Multiple regression is used because this research has multiple independent variables that are need to be processed. Those multiple independent variables are including departure date, carrier, and purchasing date.

The airlines departure time in this research is divided into peak and non-peak departure times by managerial assumption. The peak departure times are the departure time with the highest price of each flight groups, other flights are the non-peak departure times.

This research will provide the information of how the data is processed and the result of statistical analysis and the multiple regression process. The result of this research be used by the consumers to know the airlines ticket pricing behaviour based on purchasing date and departure time.

1.2 Problem Formulation

As stated in the background before, in order to become a smart shopper, the airlines pricing patterns should be analyzed by consumers before the airline ticket is decided to be purchased. In order for the airlines pricing patterns to be analyzed, understanding the airlines ticket pricing behaviour over time and using the multiple regression is necessary to detect the behaviour of airlines price data.

1.3 Research Objectives

In order for the problems formulated above to be answered, there are several research objectives in conducting this research, including:

1. To identify the behaviour of Garuda Indonesia Airlines ticket on departure date December 14th, 2015 and December 17th, 2015 by days of purchase
2. To identify the influence of each departure time's price change to other departure time's price change
3. To compare the behaviour of Garuda Indonesia Airlines ticket price changes behaviour on peak and non-peak departure time

1.4 Research Contribution

The contributions that could be obtained by the airlines ticket consumers through this research are as follows:

1. The consumers can know the airlines pricing behaviour over time for Garuda Indonesia Airlines on departure date December 14th, 2015 and December 17th, 2015 that can be seen in the online travel website
2. The consumer will know how to predict the change of airline prices by knowing the influence of each departure time's prices to other departure time's prices

1.5 Scope of Research

The scope of this research is divided into two parts, which are the research boundaries and research assumptions. The detailed information of the scope of research are as follows:

1.5.1 Boundaries

The boundaries that are used in conducting this research can be seen in the following details:

1. The research is focused on mining the airfare data from online travel agent to minimize ticket purchasing price
2. The data of airfare is obtained from www.traveloka.com
3. The data of airfare is limited only for the flight from Surabaya (SUB) to Jakarta (CGK)
4. The data of airfare is collected for two departure date, which are December 14th, 2015 (Monday) and December 17th, 2015 (Thursday)
5. The data of airfare is collected since 30 (thirty) days prior to the departure date
6. The data of airfare is collected 20 (twenty) times per day with the sampling time is in random

1.5.2 Assumptions

The assumptions that are used in conducting this research can be seen in the following details:

1. The airfare data collected is valid

1.6 Research Outline

The outline of this research report will be explained in the following details:

CHAPTER I: INTRODUCTION

This chapter explains the primary aspects of the research, including the background that motivates the researcher to conduct the research, followed by the

problem identified and formulated for the research, then the research objectives and benefits, the scope of research, and the research outline.

CHAPTER II: LITERATURE REVIEW

This chapter explains about the fundamental theory that is used in the research. The literature focuses on the multiple regression methods that is used to analyze the airlines pricing patterns, and other literature needed to support the method used in the research.

CHAPTER III: RESEARCH METHODOLOGY

This chapter explains the methodologies used to conduct the research that will be presented in flowchart followed with the detailed explanation about each phase of the research.

CHAPTER IV: DATA COLLECTION AND DATA PROCESSING

This chapter presents the airfare data collected from the online travel agent website needed to conduct the research and presents how the airfare data is processed by using multiple regression methods.

CHAPTER V: ANALYSIS

This chapter explains about the analysis of the result of the processed airfare data in the previous chapter. The analysis will mainly explain the airlines pricing behaviour and patterns, and also the impact or saving that can be obtained by the consumers while considering the research results when purchasing tickets.

CHAPTER VI: CONCLUSIONS AND RECOMMENDATIONS

This chapter explains about the conclusions of the research as the answer due to the research objectives and the recommendations for further research that can be done based on this initial research and the research topics.

CHAPTER 2

LITERATURE REVIEW

This chapter explain about the fundamental theory that is used in the research of mining the airfare data from online travel agent to minimize ticket purchasing price. The main contents of this chapter is data mining, and also other supporting literatures needed for the research.

2.1 Airlines Business Model

Airlines make strategic planning as a continual process, which take account to the challenging and ever-changing competitive environments as well as answering the consumer needs. For example, the business consumers tend to be sensitive to flight times and expect a high service level, the short-haul business consumers tend to be more sensitive to the ticket prices than long-haul business consumers, while the leisure consumers tend to be more sensitive to price but less demanding about service levels (boeing, 2016).

In order to overcome those problems, airlines vary their ticket prices into different classes, or even different business models. The airlines business model is mainly divided into three, they are full service carrier (FSC), low cost carrier (LCC) and charter carrier (CC). The further explanation about both airlines business model will be explained in the next section.

2.1.1 Full Service Business Model

A full-service carrier is an air carrier, typically a traditional or major carrier that operates on a relatively extensive route network (thus also referred to as a network carrier) and provides a full range of services including different seating classes, in-flight entertainment, meals and beverages, on-board store, and ground facilities such as waiting lounges for premium class passengers or frequent flyer programme members (ICAO, 2004).

In the case study US Airlines 2008, a full-service carrier (FSC) is defined as an airline company developed from the former state-owned flag carrier,

through the market deregulation process, into an airline company with the following elements describing its business model:

- *Core business*: Passenger, Cargo, Maintenance.
- *Hub-and-spoke network*: This has as its major objective the full coverage of as many demand categories as possible (in terms of city-pair) through the optimization of connectivity in the hub.
- *Global player*: Domestic, international and intercontinental markets are covered with short-, medium-, and long-haul flights from the hubs to almost every continent.
- *Alliance development*: No individual airline has developed a truly global network. Thus the network is virtually enlarged by interlining with partner carriers and become part of multi-HS systems.
- *Vertical product differentiation*: This is affected through in-flight and ground service, electronic services (internet check-in) and travel rules to cover all possible market segments.
- *Customer relationship management (CRM)*: Everu FSC has a loyalty program to retain the most frequent flyers. The frequent flyers program (FFP) have become part of a broader strategy called CRM. The general purpose of CRM is to enable carriers to better manage their customers through the introduction of reliable processes and procedures for interacting with those customers. The final aim of the CRM is to enhance the passenger's buying and travelling experience in order to personalize the carrier's services. In this perspective, the CRM is an extra tool to differentiate the airline product.
- *Yield management and pricing*: To support product differentiation, pricing and yield management is sophisticated, with the aim of minimizing the network revenues.
- *Multi-channel sales*: Sales channels are divided into indirect off-line (intermediate travel agencies) or indirect on-line (web intermediate electronic-agents); direct on-line: the passenger buys the tickets directly via airline's internet site; direct off-line: the passenger buys the tickets

directly via airline's call centre, the airlines city office (CTO), or the airline's airport office (ATO). The FSC cover all of these channels.

- *Distribution system*: The complexity of the distribution system described above is technologically supported by external companies called Global Distribution Systems (GDSs). Among the most diffused GDSs are: Galileo, Amadeus, WorldSpan, Sabre. (Cento, 2009).

2.1.2 *Low Cost Carrier Business Model*

A low-cost carrier refers to an air carrier that has a relatively low-cost structure in comparison with other comparable carriers and offers low fares or rates. Such a carrier may be independent, the division or subsidiary of a major carrier or, in some instances, the ex-charter arm of an airline group (ICAO, 2004).

The concept of 'low-cost carriers' or LCC originated in the United States with Southwest Airlines at the beginning of the 1970s. In Europe, the Southwest model was copied in 1991, when the Irish company Ryanair, previously a traditional carrier, transformed itself into an LCC and was followed by other LCCs in the UK (e.g. easyJet in 1995). In the literature, there are several similar definitions of an LCC, also known as a low fare or no-frills airline (Cento, 2009).

In the case study US Airlines 2008, LCC is considered as an airline company designed to have a competitive advantage in terms of costs over an FSC. In order to achieve this advantage, an LCC relies on a simplified business model (compared with the FSC), a model which is characterized by some or all of the following key elements:

- *Core business*: This is passenger air-service despite the ancillary offers are increasing and becoming part of the LCC core business.
- *Point-to-point network*: The network is developed from one or a few airports, called 'bases', from which the carrier starts operating routes to the main destinations. No connections are provided at the airport bases, which function as aircraft logistics and maintenance bases.
- *Secondary airports*: City-pairs are connected mainly from the secondary or even tertiary airports, such as London Luton, that are less expensive in

terms of landing tax and handling fee and experience less congestion than the larger ones, such as London Heathrow. Small airports will strive to gain the LCC's operation and the usual way is to reduce airport charges. Similarly, air transport activity generates welfare that is a multiple of the airports' activities, inducing regional economic and social development. Local authorities recognize that the LCC operation is a potential driver for social and economic developments, and are willing to provide financial help (for example: tax exemption, marketing support while LCCs start a new connection). The reduced airport fees can be understood as an incentive, as most of these secondary airports are public.

- *Single aircraft fleet:* In general, the LCC operates with one type of aircraft such as the Boeing 737 series with a configuration of 149 seats. The fleet composition also depends on the fact that they operate on only short- or medium-haul routes.
- *Aircraft utilization:* The aircraft is in the air, on average, more hours a day compared with FSC's that have to respect the connectivity schedule.
- *No frills service:* The product is not differentiated as they do not offer lounge services at airports, choice of seats, and in-flight service, and they do not have a frequent flyer program. Fare restrictions are removed so that the tickets are not refundable and there is no possibility to rebook with other airlines.
- *Minimized sales/reservation costs:* all tickets are electronic and the distribution system is implemented via the Internet or telephone sales centre (only direct channels). Passengers receive an e-mail containing their travel details and confirmation number, when they purchase. The LCC does not intermediate the sale with travel agents and nor does it outsource the distribution to GDS companies.
- *Ancillary services:* LCC increasingly have revenue sources other than ticket sales. Typical examples are commissions from hotels and car rental companies, credit card fees, (excess) luggage charges, in-flight food and

beverages, advertising space. The potential growth of this revenue comes from telephone operations and gambling on board (Cento, 2009).

Not every low-cost airline implements all the points mentioned above. The differences between the FSC and LCC business models are multifaceted. The significant structural cost gap between the two models results from these fundamental differences. The cost gap between those models can be seen below.

Table 2.1 Cost Gap Between LCC and FSC

	Cost Reduction	Cost per Seat
<i>Full-service carrier</i>		100%
<i>Low-service carrier</i>		
Operating advantages		
Higher seating density	-16	84
Higher aircraft utilization	-2	82
Lower flight and cabin crew cost	-3	79
Use cheaper secondary airports	-4	75
Outsourcing maintenance/single aircraft type	-2	73
Products/service features		
Minimal station costs and outsourced handling	-7	66
No free in-flight catering, fewer passenger services	-5	61
No differences in distribution		
No agents of GDS commissions	-6	55
Reduced sales/reservation costs	-3	52
Other advantages		
Smaller administration and fewer staff/offices	-3	49
<i>Low-cost carrier compared with a full-service carrier</i>		49%

(Source: Cento, 2009)

LCCs have successfully designed a focused, simple operating model around nonstop air travel to and from high-density markets. On the other hand, the FSC model is cost-penalized by the synchronized hub operations (e.g. long aircraft turns, slack built into schedules to increase connectivity) that implicitly accept the extra-time needed for passengers and baggage to make connections. In addition, the FSC business model relies upon highly sophisticated information systems and infrastructure to optimize its hubs (Cento, 2009).

2.1.3 *Charter Carrier Business Model*

A charter carrier is a non-scheduled air carrier that operates only charter flights (ICAO, 2004). In the case study US Airlines 2008, a charter carrier is defined as an airline company that operates flights outside normal schedules, by a hiring arrangement with a particular customer. Charter flights have acquired the more specific meaning of a flight whose only function is to transport holidaymakers to tourist destinations. However, tickets are not sold directly by the charter airline, but by tour operator companies who have chartered the flight.

Although charter airlines typically carry passengers who have booked, individually or as small groups to beach resorts, historic towns, or cities where a cruise ship is waiting for them, sometimes an aircraft is chartered by a single group, such as members of a company, a sports team, or the military. In general, charter flights are sold as part of a package holiday in which the price paid includes flights, accommodation and other services (Cento, 2009).

CCs were divided into the ones that form part of vertically-integrated tour operating groups and those that remain independent. The sources of cost advantage that the two types of charter airline have over the LCCs were analysed and identified as the following:

- Larger aircraft and longer-haul destinations;
- Higher load factor, aircraft utilization and labor productivity;
- Lower distribution costs, landing fees, aircraft leasing costs, and admin and finance costs (Cento, 2009).

2.2 **Revenue Management**

Revenue management refers to strategy and tactics used by a number of industries, notably the passenger airlines, to manage the allocation of their capacity to different fare classes over time in order to maximize revenue (Philips, 2005). Revenue management can be seen as a special case of pricing with constrained supply.

Revenue management is highly used in the airlines companies because the airfare has high complexity to manage. The revenue management is applicable

for the business, especially airlines or other transportation companies, under the following condition:

1. The seller is selling a fixed stock of perishable capacity;
2. The customers book capacity prior to departure;
3. The seller manages a set of fare classes, each of which has a fixed price (at least in the short run);
4. The seller can change the availability of fare classes over time (Philips, 2005).

Revenue management is based on setting and updating the availability of fare classes, where each fare class has an associated fare that remains constant through the booking period, not based on setting and updating prices. The term *revenue management industries* is used to refer the industries meeting the conditions listed above and *revenue management companies* is used to refer the companies that is using booking controls and fare classes.

Consistent execution of three levels of revenue management decisions is required to obtain successful revenue management. The three levels of revenue management can be seen in the table 2.2 below.

Table 2.2 The Three Levels of Revenue Management Decisions

Level	Description	Frequency
Strategic	Segment market and differentiate prices	Quarter or annually
Tactical	Calculate and update booking limits	Daily or weekly
Booking control	Determine which bookings to accept and which to reject	Real time

(source: Philips, 2005)

Revenue management strategy is the identification of customer segments and the establishment of products and prices targeted at those segments. Revenue management tactics require setting and updating limits on how much of a particular product can be sold at a particular fare to each segment for some period of time. Booking control is the moment-to-moment determination of which booking request should be accepted and which should be rejected (philips, 2005).

2.2.1 Tactics for Price Differentiation

Price differentiation is a powerful way for sellers to increase contribution but not all of the sellers implementing price differentiation. The reason is that there are powerful real-world limits to price differentiation. Those powerful real-world limits are as follows:

- *Imperfect segmentation.* The brain-scan technology required to determine the precise willingness to pay of each customer has not yet been developed. The best that can be done is to create market segments such that the *average* willingness to pay is different for each segment.
- *Cannibalization.* Under differential pricing, there is a powerful motivation for customers in high-price segment to find a way to pay the lower price.
- *Arbitrage.* Price differentials create a strong incentive for third-party *arbitrageurs* to find a way to buy the product at the low price and resell to high-willingness-to-pay customers below the market price and keeping the difference for themselves (Philips, 2005).

The presence of any of these factors can eliminate the benefits of price differentiation. The properties of price differentiation are including:

- Price differentiation allows sellers to increase profitability by changing different prices to customers with different willingness to pay;
- Cannibalization, imperfect segmentation, or arbitrage can destroy, or even reverse, the benefits of price differentiation (Philips, 2005).

The first property explain that sellers have tremendous incentive to find creative way to set the position and price of their products differently to the different market segments. The second property explain that the price differentiation strategy needs to be carefully planned and managed to be successfully conducted. The most common and effective approaches to price differentiation used in different market will be explained further below.

2.2.1.1 Group Pricing

The tactic of offering different prices to different groups of customers for exactly same product is considered as *group pricing*. The idea is to offer a lower price to customers with a low willingness to pay and a higher price to those with a high willingness to pay (Philips, 2005). Examples of pure group pricing are student discounts, senior citizen discounts, family specials, and more. In order to conduct the group pricing successfully, four criteria that must be hold are as follows:

1. There must be an unambiguous indicator of group membership. Furthermore, it must be difficult or impossible for members of one group to masquerade as member of another.
2. Group membership must strongly correlate with price sensitivity.
3. The product or service should not be easily traded or exchanged among purchasers. This is necessary to avoid arbitrage.
4. The segmentation must be both culturally and legally accepted. Differentiating prices on the basis of other than common characteristics, such as race and gender, are controversial or illegal (Philips, 2005).

This pure group pricing is relatively rare to be used in direct customer sales, but it is common to be used in services. Pure group pricing is also common in business-to-business sales.

2.2.1.2 Channel Pricing

The practice of selling the same product for different prices through different distribution channels is called *channel pricing* (Philips, 2005). The example of channel pricing are as follows:

- Special “web-only” fares for airline tickets are available through the internet but not through travel agencies.
- Many fashion and home furnishing merchants offer lower prices for merchandise through mail-order catalogs than for the same merchandise sold through retail outlets (Philips, 2005).

There might be more than one reason for the sellers to charge different price through different channels. The one and the most common is cost. For many

companies, especially small business, selling through the internet is cheaper than selling through traditional channels. However, it is also the case that customers arriving via different channels have different price sensitivities. Thus differential willingness to pay is also a motivation for channel pricing.

2.2.1.3 Regional Pricing

Regional pricing is differentiating prices based on region. Regional pricing is an extremely common price differentiation technique. Examples of regional pricing are as follows:

- In Latin America, McDonalds sells hamburgers for higher prices in wealthy neighborhood than in poorer ones.
- A round trip New York-Tokyo ticket purchased in Japan will usually cost more for many airlines than the same ticket purchased in the United States.
- A glass of beer costs more at an airport bar than at the corner bar (Philips, 2005).

The price difference is based on the supplier's desire to exploit differences in price sensitivity between locations (Philips, 2005).

2.2.1.4 Couponing and Self-Selection

In a *self-selection* approach, both the list price and a discounted price are available to all customers, but it takes additional time, effort, or flexibility to obtain the discounted price (Philip, 2005). The idea is that those willing to make the additional effort to get the discount are generally more price sensitive than those who are not. Examples of coupon and self-selection approach are as follows:

- Retailers commonly offer discount coupon through newspaper, direct mail, and magazines.
- Retailers often offer mail-in rebates for purchasers of a good.
- Movie theaters charge lower prices for a weekday matinee than for a Saturday night show.

- Brand-name retailers such as Ralph Lauren, the Gap, and Liz Claiborne operate outlet stores in somewhat out-of-the-way locations in which merchandise is available for a substantial discount (Philips, 2005).

2.2.1.5 Product Versioning

When pure group pricing is not feasible, companies use other strategies to differentiate prices. The most notable of these is designing or developing products (either virtual or real) that may have only minor differences but enable the seller to exploit differences in price sensitivity among customer segments. This can involve developing an “inferior” variant and/or a “superior” variant of an existing product (Philips, 2005).

In developing an “inferior” product, the motivation of the the seller is a desire to sell a product cheaply to customers with lower willingnesses to pay without cannibalizing sales of the full-price product. This is achieved in each case by creating an inferior version of the “standard product” (Philips, 2005).

While in developing “superior” goods, the sellers think in creating a superior good in order to extract a higher price from less price-sensitive customers. In some ways, a superior-good strategy is safer than an inferior-good strategy because it does not threaten cannibalization of existing sales. Of course, it presumes an ability to create and establish a product that themarket perceives as truly superior to the existing product and that there is a customer segment willing to pay a premium for the superior product (Philips, 2005).

Establishing a *product line* is the natural extension of creating inferior or superior products. A product line is a series of similar products serving the same general market but sold at different prices. For our purposes, we will consider vertical product lines, where almost all customers would agree that a higher-priced product is superior to a lowerpriced one.

2.2.1.6 Time-Based Differentiation

Time-based differentiation is a common form of product versioning. Examples of the usage of time-based differentiation are as follows:

- Amazon offers 5-to-9-day “super-saver” shipping free while charging \$3.97 for “standard shipping”.
- Passenger airlines offer discount rates to customers who book a week or more prior to departure.
- Software and hardware support contracts charge more for “two-hour response” than for “two-day response”
- Fashion goods cost most during the beginning of the season and are marked down toward the end of the season.

In each of these cases, companies have created differentiated products that allow customers to self-select. Time-based differentiation plays a very important role at passenger airlines, hotels, and rental car companies, in which time of booking and other factors, such as willingness to accept a Saturday night stayover, are used as indicators of whether or not a potential customer is traveling for leisure purposes or business purposes. Those traveling for leisure purposes are presumed to be more price sensitive than those traveling for business (Philips, 2005).

2.2.2 *Airlines Revenue Management Strategy*

Revenue management strategy consisted of identification of customer segments and the establishment of products targeted at those segments. A fundamental element of revenue management strategy at many hotels, rental car companies, and airlines is the distinction between leisure customers and business customers first recognized by American Airlines (Philip, 2005).

Table 2.3 Characteristics of Leisure vs. Business Customers

Leisure	Business
Highly price sensitive	Less price sensitive
Book earlier	Book later
More flexible to departure and arrival times	Less flexible
More accepting of restrictions such as Saturday night stayovers	Less accepting of restrictions

(source: Philip, 2005)

Table 2.3 showed the different characteristics of leisure and business customers. These different characteristics are used by the airlines to segment their market and create virtual products oriented toward the different segments. The typical approach of airline market segments can be seen in the table 2.4 below.

Table 2.4 Typical Airline Market Segmentation

	SEGMENT				
	BUSINESS ORIENTED			LEISURE ORIENTED	
Price sensitivity	Low	Moderate	Sensitive	Moderate	Very high
Schedule flexibility	Need	Important	Somewhat	Prefers	Very flexible
Inventory	Last seat	Somewhat restricted	Restricted	Restricted	Very limited
Product	Unrestricted Business	Corporate Discount	Discount Business	Regular Leisure	Sale fares Web only priceline.com

(source: Philip, 2005)

Based on the table 2.4, an airline has identified five customer segments, which are three business and two leisure, with at least one product targeted toward each segment. The leisure products have various restrictions (early purchase, Saturday night stay) that make them unattractive to many business travelers. Airlines used artificial restrictions to create an inferior product that they could sell at a lower price to a more price-sensitive segment of the market (Philip, 2005).

Product versioning is not the only price differentiation tactic used by the airlines. They also use most of the other tactics described in the tactics for price differentiation section in some form or other. Airlines sell products targeted to many segments such as government, senior citizens, groups, tour operators, and cruise lines. International airlines are great believers in regional pricing. They will sell the same tickets for different prices in different countries (even after adjusting for exchange rates) in order to exploit differences in price sensitivity. Airlines are also enthusiastic channel pricers; ticket prices on the Internet are often cheaper than through travel agents.

2.3 Business Statistics

Statistics may be defined as the procedures and techniques that comprise business statistics include those specially designed to describe data, such as charts, graphs, and numerical measures, and also includes inferential procedures that help decision makers draw inferences from a set of data. Then, business statistics is a collection of procedures and techniques that are used to convert data into meaningful information in a business environment (Groebner et al., 2010). The role of business statistics can be seen in the illustration below.

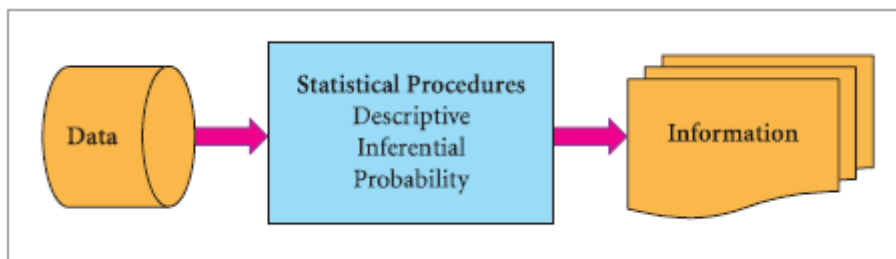


Figure 2.1 The Role of Business Statistics (Groebner et al., 2010)

Statistical inference procedures are the procedures that allow a decision maker to reach a conclusion about a set of data based on a subset of that data (Groebner et al., 2010). Estimation and hypothesis testing are two primary categories of statistical inference procedure.

Estimation is used if we are in situations in which we would like to know about all the data in a large data set but it is impractical to work with all the data, then decision makers can use techniques to estimate what the larger data set looks like. The estimates are formed by looking closely at subset of the larger data set.

Hypothesis testing is the procedure that is done by making an initial hypothesis then testing it to obtain the information to prove whether the initial hypothesis is accepted or rejected.

Populations and samples are two of the most important terms in statistics. Population is the set of all objects or individuals of interest or the measurement obtained from all objects or individuals interest. On the other hand, sample is a subset of the population. The entire set of measurement taken from the whole population is called census.

Descriptive numerical measures, such as an average or a proportion, that are computed from an entire population are called parameters. Corresponding measures for a sample are called statistics.

There are two categories of sampling techniques, which are statistical and nonstatistical. Statistical sampling techniques are those sampling methods that use selection techniques based on chance selection. On the other hand, nonstatistical sampling techniques are those methods of selecting samples using convenience, judgement, or other nonchance process. Regardless of which technique is used, the decision maker has the same objective, which is obtaining a sample that is close representative of the population.

Statistical sampling methods (also called probability sampling) allow every item in the population to have a known or calculable chance of being included in the sample. The fundamental statistical sample is called simple random sample. Simple random sampling is a method of selecting items from a population in such that every possible sample of a specified size has an equal chance of being selected (Groebner et al., 2010).

There are some techniques that can be used in statistics. In this research, the statistical technique that is used is multiple regression. The further explanation about statistical technique that is used in this research can be seen in the following section.

2.3.1 Regression Analysis

Often, you will want to use sample data to investigate the relationship among a group of variables, ultimately to create a model for some variable that can be used to predict its value in the future. The process of finding a mathematical model (an equation) that best fits the data is part of a statistical technique known as regression analysis (Mendenhall & Sincich, 2011).

In regression, the variable y to be modeled is called dependent (or response) variable and its true mean (or expected value) is denoted $E(y)$. The general form of probabilistic model in regression is as follows.

$$y = E(y) + \varepsilon$$

Where y = Dependent variable
 $E(y)$ = Mean (or expected) value of y
 ε = Unexplainable (or random) error

The process of finding the mathematical model that relates y to a set of independent variables and best fits the data is part of the process known as regression analysis. Independent variables are the variables used to predict (or model) y and are denoted by the symbols x_1, x_2, x_3 , etc. The goal of regression analysis is to build a good model, a prediction equation relating y to the independent variables, that will enable us to predict y for given values of x_1, x_2, \dots, x_k , and to do so with a small error of prediction. There are six steps to conduct regression modelling, including:

1. Hypothesize the form of the model for $E(y)$;
2. Collect the sample data;
3. Use the sample data to estimate unknown parameters in the model;
4. Specify the probability distribution of the random error term, and estimate any unknown parameters of this distribution;
5. Statistically check the usefulness of the model;
6. When satisfied that the model is useful, use it for prediction, estimation, and so on (Mendenhall & Sincich, 2011).

Regression analysis of data is a very powerful statistical tool. It provides a technique for building a statistical predictor of a response and enables you to place a bound (an approximate upper limit) on your error of prediction.

2.3.2 *Multiple Regression Analysis*

In multiple regression analysis, additional independent variables are added to the regression model to clear up some of the as yet unexplained variation in the dependent variable. Multiple regression is merely an extension of simple regression analysis; however, as we expand the model for the population from one independent variable to two or more, there are some new considerations

(Groebner et al., 2010). The general model of population multiple regression model is as follows.

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \cdots + \beta_kx_k + \varepsilon$$

Where β_0 = Population's regression constant

β_j = Population's regression coefficient for each variable $x_j = 1, 2, \dots, k$

k = Number of independent variables

ε = Model error

There are four assumptions that should be applied to multiple regression model, including:

1. Individual model errors, ε , are statistically independent of one another, and these values represent a random sample from the population of possible errors at each level of x .
2. For a given value of x there can exist many values of y , and therefore many possible values for ε . Further, the distribution of possible model errors for any level of x is normally distributed.
3. The distribution of possible ε -values have equal variances at each level of x .
4. The means of the dependent variable, y , for all specified values of x can be connected with a line called the population regression model (Groebner et al., 2010).

The estimated multiple regression model is an extension of an estimated simple regression model. The principal difference is that whereas the estimated simple regression model is the equation of a straight line in a two-dimensional space, the estimated multiple regression model forms a hyperplane (or response surface) through multidimensional space. Each regression coefficient represents a different slope. The equation of estimated multiple regression model can be seen as follows.

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + \cdots + b_kx_k$$

The regression hyperplane represents the relationship between the dependent variable and the k independent variable. The regression equation forms a slice (hyperplane) through the data such that $\sum(y - \hat{y})^2$ is minimized. This is the same least squares criterion that is used with simple linear regression. The illustration of hyperplane can be seen in figure 2.2 below.

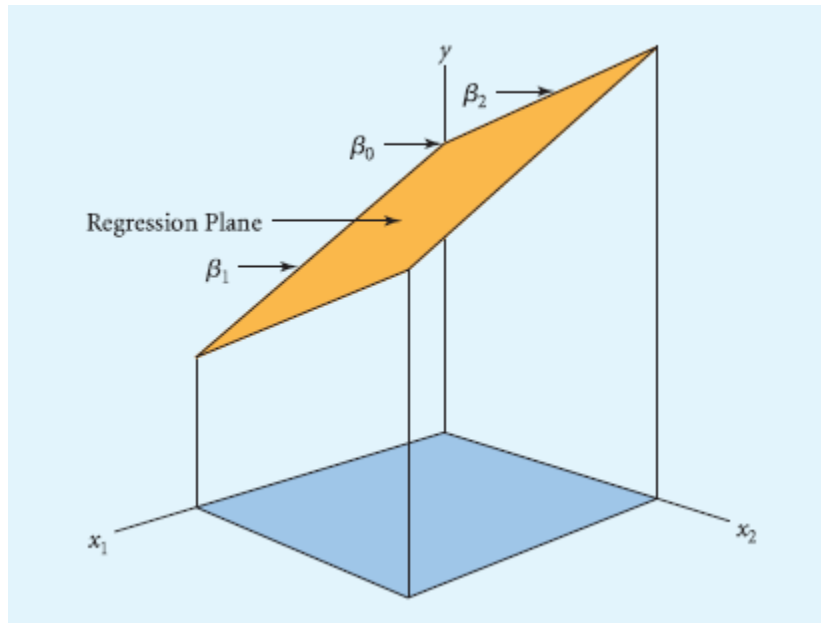


Figure 2.2 Multiple Regression Hyperplane for Population (source: Groebner et al, 2010)

An important activity in business decision making is referred to as model building. Model is a representation of an actual system using either a physical or a mathematical portrayal. Models are often used to test changes in a system without actually having to change the real system. Models are also used to help describe a system or to predict the output of a system based on certain inputs.

Model specification, or model identification, is the process of determining the dependent variable, deciding which independent variables should be included in the model and obtaining the sample data for all variables. As with any statistical procedure, the larger the sample size is the better, because the potential for extreme sampling error is reduced when the sample size is large. However, at a minimum, the sample size required to compute a regression model

must be at least one greater than the number of independent variables. There are three steps involved in modeling specification, including:

1. Decide what question you want to ask. The question being asked usually indicates the dependent variable.
2. List the potential independent variables for the model. Here, your knowledge of the situation you are modeling guides you in identifying potential independent variables.
3. Gather the sample data (observations) for all variables (Groebner et al., 2010).

Model building is the process of actually constructing a mathematical equation in which some or all of the independent variables are used in an attempt to explain the variation in the dependent variable.

Model diagnosis is the process of analyzing the quality of the model you have constructed by determining how well a specified model fits the data you just gathered. You will examine output values such as R-squared and the standard error of the model. At this stage, you will also assess the extent to which the model's assumptions appear to be satisfied. The following steps are employed in developing a multiple regression model:

1. Specify the model by determining the dependent variable and potential independent variables, and select the sample data.
2. Formulate the model. This is done by computing the correlation coefficients for the dependent variable and each independent variable, and for each independent variable with all other independent variables. The multiple regression equation is also computed.
3. Perform diagnostic checks on the model to determine how well the specified model fits the data and how well the model appears to meet the multiple regression assumptions (Groebner et al., 2010).

There is no way to determine whether an independent variable will be a good predictor variable by analyzing the individual variable's descriptive statistics, such as the mean and standard deviation. Instead, we need to look at the correlation between the independent variables and dependent variable, which is measured by the correlation coefficient. Correlation coefficient is a quantitative

measure of the strength of the linear relationship between two variables. The correlation coefficient, r , ranges from -1.0 to +1.0.

When we have multiple independent variables and one dependent variable, we can look at the correlation between all pairs of variables by developing a correlation matrix. Correlation matrix is a table showing the pairwise correlations between all variables (dependent and independent). Each correlation is computed using one of the equations as follows:

$$r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2 \Sigma(y-\bar{y})^2}} \quad \text{or} \quad r = \frac{\Sigma(x_i-\bar{x}_i)(x_j-\bar{x}_j)}{\sqrt{\Sigma(x_i-\bar{x}_i)^2 \Sigma(x_j-\bar{x}_j)^2}}$$

The appropriate formula is determined by whether the correlation is being calculated for an independent variable and the dependent variable or for two independent variables.

Coefficient of determination, R^2 , measures the proportion of variation in the dependent variable that can be explained by the dependent variable's relationship to a single independent variable. When there are multiple independent variable is a model, R^2 is called the multiple coefficient of determination. Multiple coefficient of determination is the proportion of the total variation of the dependent variable in a multiple regression model that is explained by its relationship to the independent variables. The equation of multiple coefficient of determination is as follows.

$$R^2 = \frac{\text{Sum of squares regression}}{\text{Total sum of squares}} = \frac{SSR}{SST}$$

There are several questions that should be answered to conduct the model diagnosis, including:

- Is the overall model significant?
- Are the individual variables significant?
- Is the standard deviation of the model error too large or to provide meaningful results?

- Is multicollinearity a problem?
- Have the regression analysis assumptions been satisfied?

Because the regression model constructed is based on a sample of data from the population and is subject to sampling error, we need to test the statistical significance of the overall regression model. If the null hypothesis is true and all the slope coefficients are simultaneously equal to zero, the overall regression model is not useful for predictive or descriptive purposes (Groebner et al., 2010).

The F -test is a method for testing whether the regression model explains a significant proportion of the variation in the dependent variable (and whether the overall model is significant). The F -test statistics for multiple regression model is shown as follows:

$$F = \frac{\frac{SSR}{k}}{\frac{SSE}{n-k-1}}$$

Where: SSR = Sum of squares regression
 SSE = Sum of squares error
 n = Sample size
 k = Number of independent variables

$$\text{Degrees of freedom} = D_1 = k \quad \text{and} \quad D_2 = (n - k - 1)$$

The equation of t -test for significance of each regression coefficient can be seen as follows:

$$t = \frac{b_j - 0}{s_{b_j}} \quad \text{with} \quad df = n - k - 1$$

Where: b_j = Sample slope coefficient for the j -th independent variable

s_{b_j} = Estimate of the standard error for the j -th sample slope coefficient

The t -tests are considered as conditional tests. This means that the null hypothesis is the value of each slope coefficient is 0, given that the other independent variables are already in the model.

An indication of how good the regression model is can be found by looking at the relationship between the measured values of the dependent variable and those values that would be predicted by the regression model. The standard deviation of the regression model (also called the standard error of the estimate), measures the dispersion of observed values, y , around the values predicted by the regression model (Groebner et al., 2010). The standard error of estimate can be seen as follows.

$$s_{\varepsilon} = \sqrt{\frac{SSE}{n - k - 1}} = \sqrt{MSE}$$

Where: SSE = Sum of squares error (residual)
 n = Sample size
 k = Number of independent variables

Sometimes, even though a model has a high R^2 , the standard error of the estimate will be too large to provide adequate precision for confidence and prediction intervals. A rule of thumb that we have found useful is to examine the range $\pm 2s_{\varepsilon}$. Taking into account the mean value of the dependent variable, if this range is acceptable from a practical viewpoint, the standard error of the estimate might be considered acceptable.

Even if the overall regression model is significant and each independent variable is significant, decision makers should still examine the regression model to determine whether it appears reasonable. This is referred to as checking for face validity. Specially, you should check to see that the signs on the regression coefficients are consistent with the signs on the correlation coefficients between the independent variables and the dependent variable.

Multicollinearity occurs when independent variables are correlated with each other and therefore overlap with respect to the information they provide in explaining the variation in the dependent variable. From a decision maker's viewpoint, you should be aware that multicollinearity can (and often does) exist

and recognize the basic problems it can cause. The following are some of the most obvious problems and indications of severe multicollinearity:

- Unexpected, therefore potentially incorrect, signs on the coefficients.
- A sizeable change in the values of the previously estimated coefficients when a new variable is added to the model.
- A variable that was previously significant in the regression model becomes insignificant when a new independent variable is added.
- The estimate of the standard deviation of the model error increases when a variable is added to the model (Groebner et al., 2010).

If the independent variables in a regression model are correlated and multicollinearity is present, another potential problem is that the t -tests for the significance of the individual independent variables may be misleading. That is, a t -test may indicate that the variable is not statistically significant when in fact it is. One method of measuring multicollinearity is known as the variance inflation factor (VIF). The equation of variance inflation factor can be seen below.

$$VIF = \frac{1}{(1 - R_j^2)}$$

Where: R_j^2 = Coefficient of determination when the j -th independent variable is regressed against the remaining $k-1$ independent variables

The effect of multicollinearity is to decrease the test statistic, thus reducing the probability that the variable will be declared significant. A related impact is to increase the width of the confidence interval estimate of the slope coefficient in the regression model. Generally, if the $VIF < 5$ for a particular independent variable, multicollinearity is not considered a problem for that variable. Values of $VIF \geq 5$ imply that the correlation between the independent variables is too extreme and should be dealt with by dropping variables from the model (Groebner et al., 2010).

The equation of confidence interval estimate for the regression slope can be seen below.

$$b_j \pm ts_{b_j}$$

Where: b_j = Point estimate for the regression coefficient for x_j

t = Critical t -value for the specified confidence level

s_{b_j} = The standard error of the j -th regression coefficient

CHAPTER 3

RESEARCH METHODOLOGY

This chapter explains about the methodologies used to conduct the research that will be presented in flowchart followed with the detailed explanation about each phase of the research.

3.1 Research Methodology Flowchart

This research will be done through several processes in collecting and mining the airfare data to obtain the desired output. The research methodology flowchart can be seen in figure 3.1 below.

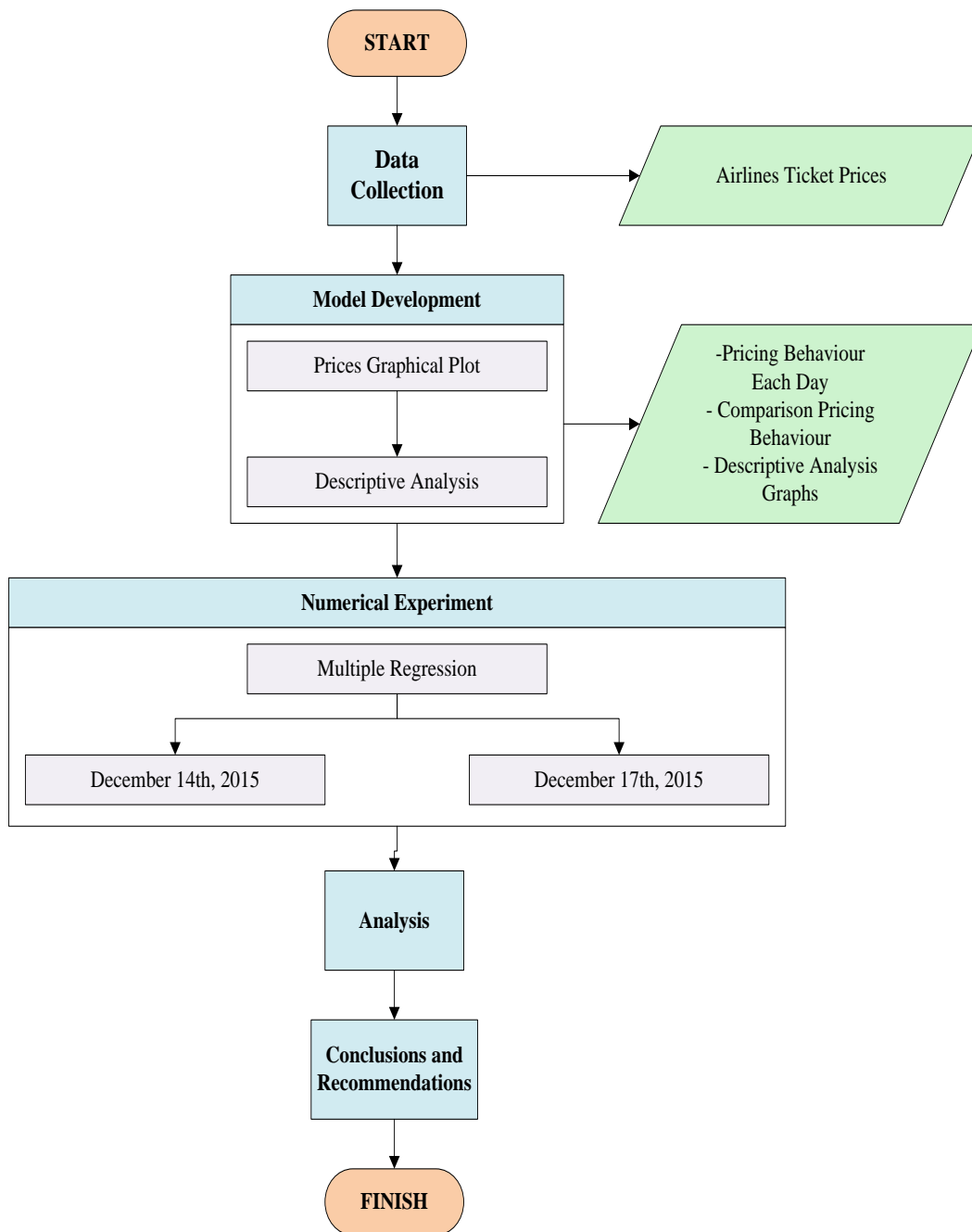


Figure 3.1 Research Methodology Flowchart

3.2 Explanation of Research Methodology Flowchart

The illustration of the research methodology is shown in figure 3.1. The further explanation about the research methodology can be seen in the next section.

3.2.1 Data Collection

Data collection process is done by collecting the airlines prices through online travel agent website (www.traveloka.com). The data collected are the airline prices for two departure dates, which are December 14th, 2015 and December 17th, 2015 with the route Surabaya (SUB) to Jakarta (CGK). The airline prices data is collected since 30 (thirty) days prior to the departure date. The data are collected 20 (twenty) times per day with the sampling time is in random. The data collected are covering all carriers and flight times on the departure date. The data collection is done by capturing the airline ticket prices at the random sampling time, then recording it into Microsoft Excel.

3.2.2 Model Development

In the model development, first, the airline ticket prices is plotted into a graph so the pricing behaviour can be seen through the graphs. After that, the descriptive analysis is done to mine more information from the data. Based on all of those information, the multiple regression model is made for numerical experiment.

3.2.3 Numerical Experiment

After the multiple regression model is made, then the model is run in two experiment. The first experiment is running the model on Monday, December 14th, 2015. The second experiment is running the model on Thursday, December 17th, 2015. The model will identify whether there is influence of each departure time's price change to other departure time's price change.

3.2.4 Analysis

After the numerical experiment is done, the analysis can be extracted from the experiment. The analysis will mainly discuss about the pricing behaviour of airline ticket pricing between peak and non-peak departure time based on the purchasing date (days before departure) and the influence of each departure time's price change to other departure time's price change.

3.2.5 *Conclusions and Recommendations*

The conclusion can be extracted from the analysis to answer the objectives of the research and the recommendations for further research that can be done based on this initial research and the research topics.

CHAPTER 4

DATA COLLECTION AND NUMERICAL EXPERIMENT

This chapter elaborates the data collection and data processing of the research. This chapter includes the existing condition of the data, the data collection, and the data processing.

4.1 Existing condition

This subchapter describes the existing condition of the data collected. The data is collected from the online travel website www.traveloka.com. The online travel website provides ticket prices from all airlines that is operating in the selected condition. The condition refers to the criteria sets by the online shopper, such as departure date, return date, origin, destination, number of person, and other filters.

In this case, the airline price data is collected by using several criteria. First, the origin is set from Surabaya (SUB) to Jakarta (CGK) as the destination with one number of person. Then, this research collect the data for two departure date, which are Monday, December 14th, 2015 and Thursday, December 17th, 2015. There is no other filter used beside the ones stated before.

The data is collected thirty days prior to the departure date. In this case, the data collection for the departure date December 14th, 2015 is collected since November 14th, 2015 and for the departure date December 17th, 2015, the data is collected since November 17th, 2015. The data is collected twenty times a day and the sampling time is in random.

There are seven airlines operating in this condition, which are Garuda Indonesia Airlines, Citilink, Lion Air, Batik Air, Sriwijaya Air, and AirAsia Indonesia. There are two full service carrier (FSC) between those airlines, which are Garuda Indonesia Airlines and Batik Air, and the others are low service carrier (LCC).

The data is collected for all departure time within the departure date of all airline carrier. The data is divided into four time range based on the departure

time, which are morning, noon, afternoon, and evening time based on the filter by Traveloka. The morning time range from 4.01 until 11.00, noon time range from 11.01 until 15.00, afternoon time range from 15.01 until 19.00, and evening time range from 19.01 until 4.00.

The peak and non-peak departure times are obtained from each flight groups. The peak departure times are the departure time with the highest price from each flight groups, other than that are the non-peak departure times. The peak departure times from both December 14th, 2015 and December 17th, 2015 are GA 8:45, GA 12:00, GA 18:45, and GA 19:30.

4.2 Data Collection

This subchapter describes about the airline price data collected. The airline price data will be described in graphical representation of each airline carrier per departure date. There will be also more specific graphical representation of the data that will be used in numerical experiment. The graphical representation of all collected data including some of the departure time on the departure date that will represent the ticket price for departure time in the morning, noon, afternoon and evening. The x axis of the graphical representation is the sampling period of the data. The sampling is done twenty (20) times a day so period 600 to 581 is the same as day 30 before departure, period 580 to 561 is day 29 before departure, period 560 to 541 is day 28 before departure, and so on.

The graphical representation of Garuda Indonesia Airline ticket price changes on December 14th, 2015 can be seen in the figure 4.1 below.

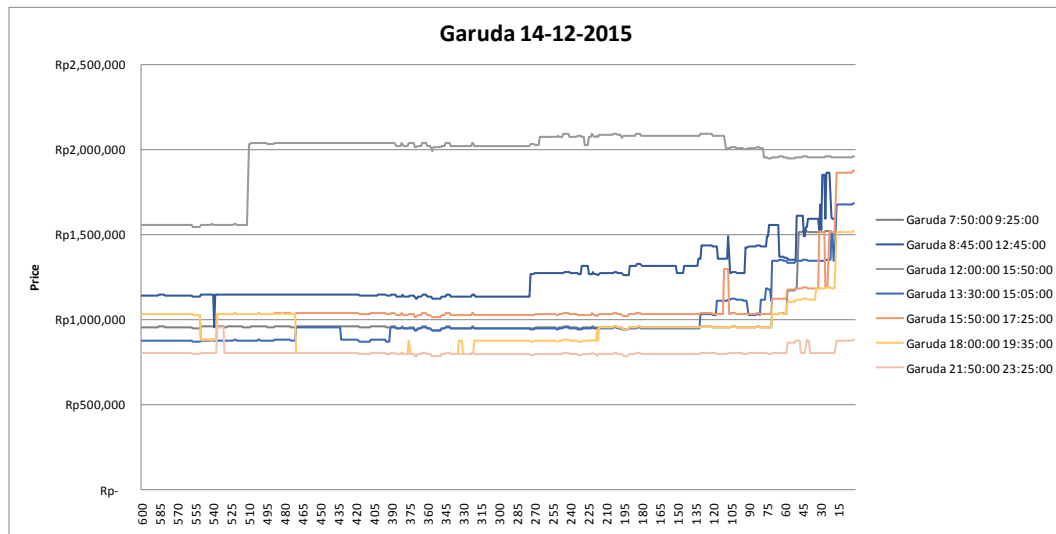


Figure 4.1 Garuda Indonesia Airlines Ticket Price Changes on December 14th, 2015

It can be seen in the figure that most of the schedule has similar trend, which is increasing as the time near to the scheduled departure except for the departure time 12:00 and 21:50. The price changes of departure time 12:00 is starting to increase around period 525 but decreasing near to the scheduled departure. Meanwhile, the price changes of departure time 21:55 tends to be stable and increasing a little near the departure schedule.

It can also be seen that the departure time 12:00 prices are the highest compared to other departure time and the departure time 21:55 prices are the lowest compared to other departure time. This condition might be caused by the different number of buyer. Garuda Indonesia Airlines is courageous to set a very high price for departure time 12:00 because more people might tend to like travelling in that time so even if they set a very high price, the tickets will still be purchased by the buyer. This condition is the opposite with the departure time 21:55. There might be not many people who like to travel around that time so they set the price lower than other departure time.

The graphical representation of Garuda Indonesia Airline ticket price changes on December 17th, 2015 can be seen in the figure 4.2 below.

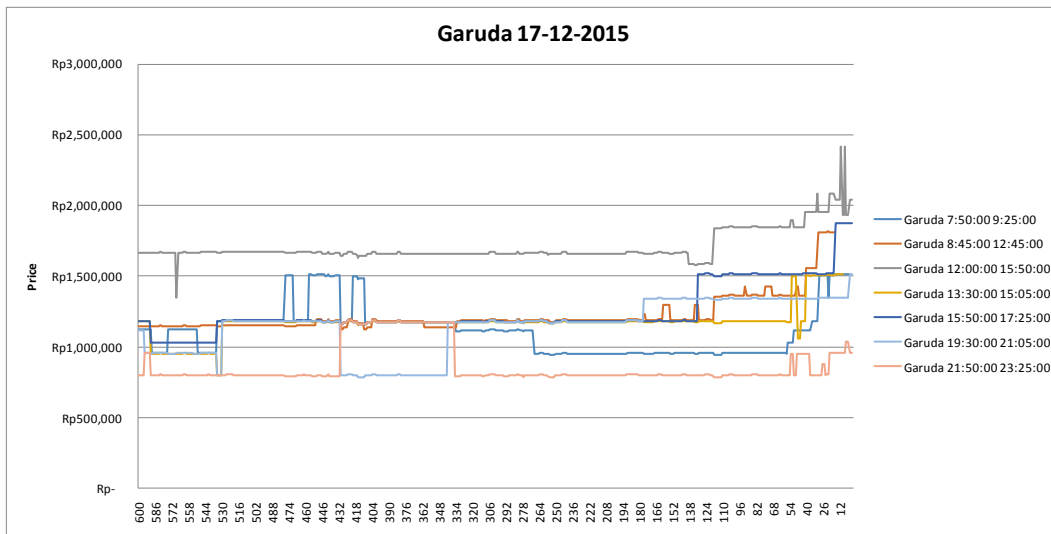


Figure 4.2 Garuda Indonesia Airlines Ticket Price Changes on December 17th, 2015

It can be seen in the figure that the Garuda Indonesia Airlines ticket price changes on departure date December 17th, 2015 is different with the ticket prices on departure date December 14th, 2015. The ticket price changes on departure date December 17th, 2015 varies more than the ticket prices on departure date December 14th, 2015 but the prices trend near the scheduled departure is also increasing similar to the ticket prices on departure date December 14th, 2015.

The prices on departure time 12:00 are also the highest and the prices on departure time 21:55 are the lowest, similar with the ticket prices on departure date December 14th, 2015. The different is that the prices for departure time 12:00 on December 17th, 2015 is lower than the prices on December 14th, 2015.

The data that will be used in numerical experiment is the data of Garuda Indonesia Airlines on departure date December 14th, 2015 and December 17th, 2015. The Garuda Indonesia Airlines data is chosen because it has the most departure time that can represent each flight time groups. The following figures are the graphical representation of the data that will be used in numerical experiment. The departure times that will be showed in the following figures are the highest and lowest price of each flight groups from Monday, December 14th, 2015 and Thursday, December 17th, 2015 departure date.

The following figure shows the Garuda Indonesia Airline ticket price changes for departure date December 14th, 2015 (Monday) and December 17th, 2015 (Thursday) only for morning flight representation.

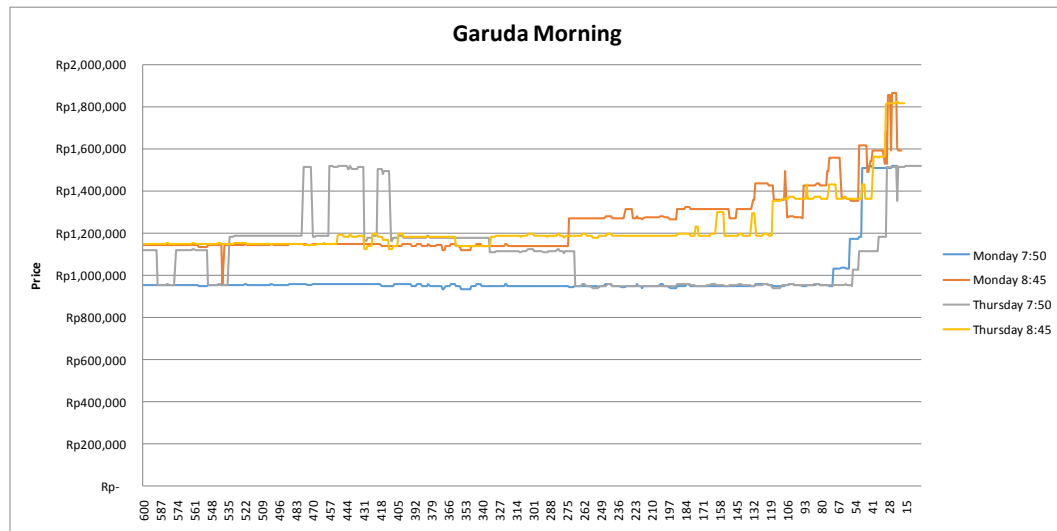


Figure 4.3 Garuda Indonesia Airlines Ticket Price Changes of Monday and Thursday Departure Day for Morning Flights

It can be seen in the figure that the prices on Monday 7:50 and Thursday 7:50 has different trend in the beginning but similar near to the scheduled departure. The prices on Thursday 7:50 tend to be increasing since the beginning period but decreasing since period 405 and then increasing again near the scheduled departure. Meanwhile, the prices on Monday 8:45 and Thursday 8:45 has the similar trend but the Monday 8:45 has higher prices than Thursday 8:45.

We can compare the behaviour by days of purchase by looking at the price changes from each week. We can divide the periods into four weeks, week 4 from period 600 to 451, week 3 from period 450 to 301, week 2 from period 300 to 151, and week 1 from period 150 to 1.

If we see in the figure 4.3 above, the prices on the week 4 are tend to be stable, except flight Thursday 7:50. The flight Thursday 7:50 has many ups and downs in the week 4. If we see the prices on the week 3, the other prices still tends to be stable, except flight Thursday 7:50. The flight Thursday 7:50 has many ups and downs and after the price back to the starting price, the price is decreasing in the week 3. If we see the prices on the week 2, the other prices still tends to be

stable, except flight Thursday 7:50 and Monday 8:45. The price of flight Thursday 7:50 is still decreasing, meanwhile the price of flight Monday 8:45 is starting to increase. If we see the prices on week 1, the prices of flight Monday 8:45 and Thursday 8:45 are increasing and the prices of flight Monday 7:50 and Thursday 7:50 are rapidly increasing around four days to the scheduled departure.

The following figure shows the Garuda Indonesia Airline ticket price changes for departure date December 14th, 2015 (Monday) and December 17th, 2015 (Thursday) only for noon flight representation.

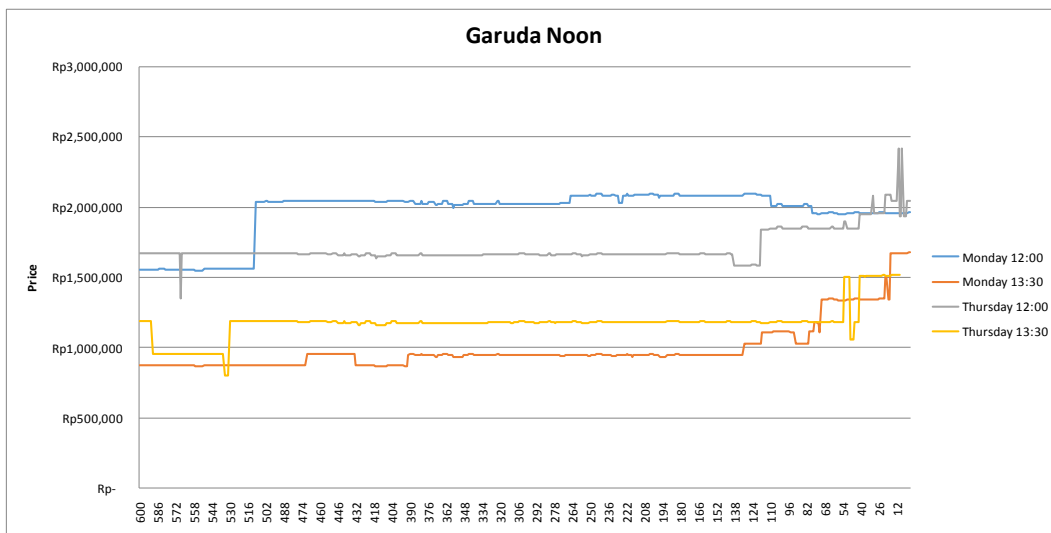


Figure 4.4 Garuda Indonesia Airlines Ticket Price Changes of Monday and Thursday Departure Day for Noon Flights

It can be seen in the figure that the Monday 12:00 has higher prices than Thursday 12:00 and the Monday 12:00 prices are increasing faster than Thursday 12:00, but the Thursday 12:00 prices are increasing more than Monday 12:00 near the scheduled departure. Meanwhile, the Thursday 13:30 has higher prices than Monday 13:30.

If we see in the figure 4.4 above, the prices on the week 4 are tend to be stable, except flight Monday 12:00 and Thursday 13:30. The flight Monday 12:00 is increasing once and then stable, while flight Thursday 13:30 is decreasing and then increasing again to the starting price. If we see the prices on the week 3, the all of the prices tend to be stable. If we see the prices on the week 2, all of the prices still tend to be stable. If we see the prices on week 1, all of the prices are

increasing towards the scheduled departure, except flight Monday 12:00. The price of flight Monday 12:00 is slightly decreasing towards the scheduled departure.

The following figure shows the Garuda Indonesia Airline ticket price changes for departure date December 14th, 2015 (Monday) and December 17th, 2015 (Thursday) only for afternoon flight representation.

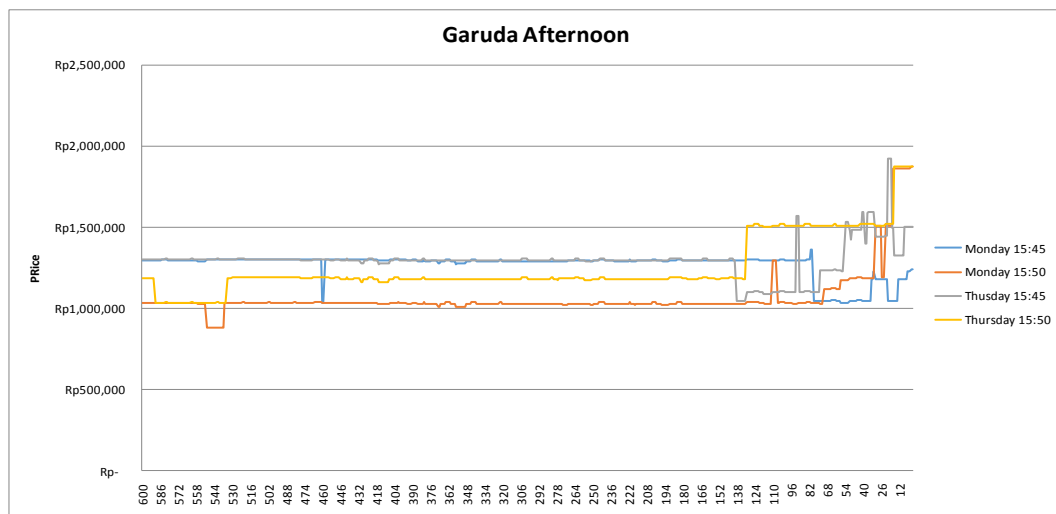


Figure 4.5 Garuda Indonesia Airlines Ticket Price Changes of Monday and Thursday Departure Day for Afternoon Flights

It can be seen in the figure that the Monday 15:45 prices and Thursday 15:45 prices have different trend. The Monday 15:45 prices tend to decrease near the scheduled departure while the Thursday 15:45 prices tend to increase near the scheduled departure. The Thursday 15:50 prices are also higher than the prices Monday 15:50. This means that the prices on Monday is not always higher than Thursday.

If we see in the figure 4.5 above, the prices on the week 4 are tend to be stable, except flight Monday 15:50 and Thursday 15:50. The flight Monday 15:50 and Thursday 15:50 are decreasing and then increasing back to the starting price. If we see the prices on the week 3, the all of the prices tend to be stable. If we see the prices on the week 2, all of the prices still tend to be stable. If we see the prices on week 1, all of the prices has ups and downs while increasing towards the

scheduled departure, except flight Thursday 15:50. The flight Thursday 15:50 is constantly increasing towards the scheduled departure.

The following figure shows the Garuda Indonesia Airline ticket price changes for departure date December 14th, 2015 (Monday) and December 17th, 2015 (Thursday) only for evening flight representation.

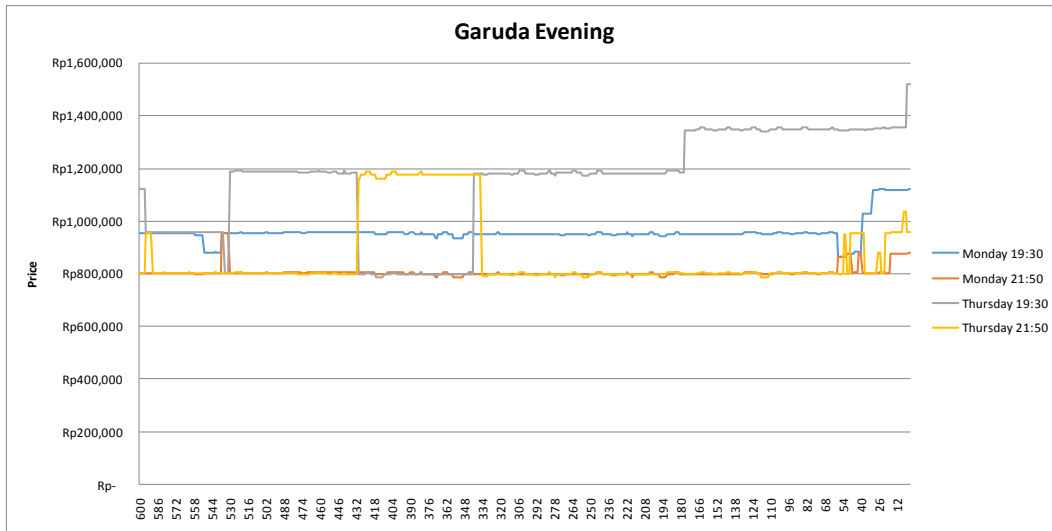


Figure 4.6 Garuda Indonesia Airlines Ticket Price Changes of Monday and Thursday Departure Day for Evening Flights

It can be seen that the prices on Thursday 19:30 are a lot higher than the prices on Monday 19:30. This phenomenon is also supporting the conclusion that the prices on Monday not always higher than Thursday. Meanwhile, the prices on Monday 21:50 and Thursday 21:50 has similar trend but the prices on Thursday 21:50 has more ups than the prices on Monday 21:50.

If we see in the figure 4.6 above, the prices on the week 4 are tend to be stable, except flight Thursday 19:30. The flight Thursday 19:30 is decreasing and increasing higher than the starting price. If we see the prices on the week 3, the other prices still tends to be stable, except flight Thursday 19:30 and Thursday 21:50. The flight Thursday 19:30 is decreasing again and back to the previous price, while flight Thursday 21:50 is increasing and decreasing back to the starting price. If we see the prices on the week 2, the other prices still tends to be stable, except flight Thursday 19:30. The price of flight Thursday 19:30 is increasing once then stable. If we see the prices on week 1, the price of flight Thursday 19:30

is increasing just a day before departure, while the other flights are increasing towards the scheduled departure with many ups and downs.

4.3 Numerical Experiment

This subchapter elaborates the numerical experiment done in this research. The numerical experiment will be done in two step, the first is descriptive analysis and then the multiple regression. The descriptive analysis is done to know the behaviour of the data and compare the data behaviour of Monday and Thursday for Garuda Indonesia Airlines. The multiple regression is done to know whether there is influence of each departure time's price change to other departure time's price change.

4.3.1 Descriptive Analysis

The descriptive analysis of the data is used to know behaviour of the data and compare the data behaviour of Monday and Thursday for each airline carrier. To know the behaviour of the data, the data of each airline carrier per departure date is classified into three class by using k-means so that the data is classified based on the similar price. The classification is done by using software Matlab.

The classification result for Garuda Indonesia Airlines on departure date December 14th, 2015 can be seen in the following table.

Table 4.1 Classification of Garuda Indonesia Airlines on Departure Date December 14th, 2015

Class 1	Class 2	Class 3
13:30:00	5:25:00	12:00:00
21:50:00	6:15:00	15:45:00
22:50:00	7:50:00	
	8:25:00	
	8:45:00	
	8:50:00	
	10:10:00	
	11:10:00	
	12:30:00	
	14:50:00	
	15:50:00	
	18:00:00	
	18:40:00	
	19:30:00	
	20:30:00	

After the data is classified, the data is regressed for each class to know the behaviour of each class. The regression result for the Garuda Indonesia Airlines on departure date December 14th, 2015 for class 1 can be seen in the figure 4.7 below.

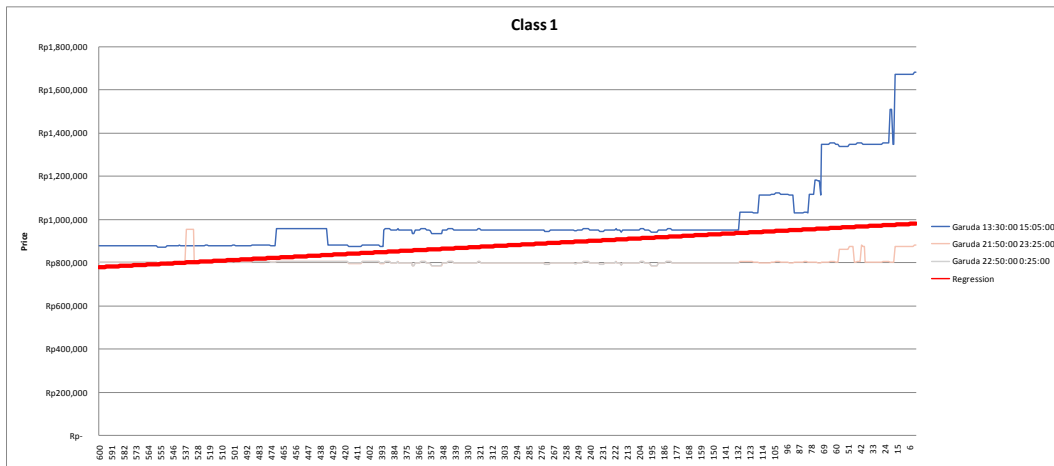


Figure 4.7 Garuda Indonesia Airlines Ticket Price Regression Result of Class 1 for Departure Date December 14th, 2015

The regression result for the Garuda Indonesia Airlines on departure date December 14th, 2015 for class 2 can be seen in the figure 4.8 below.

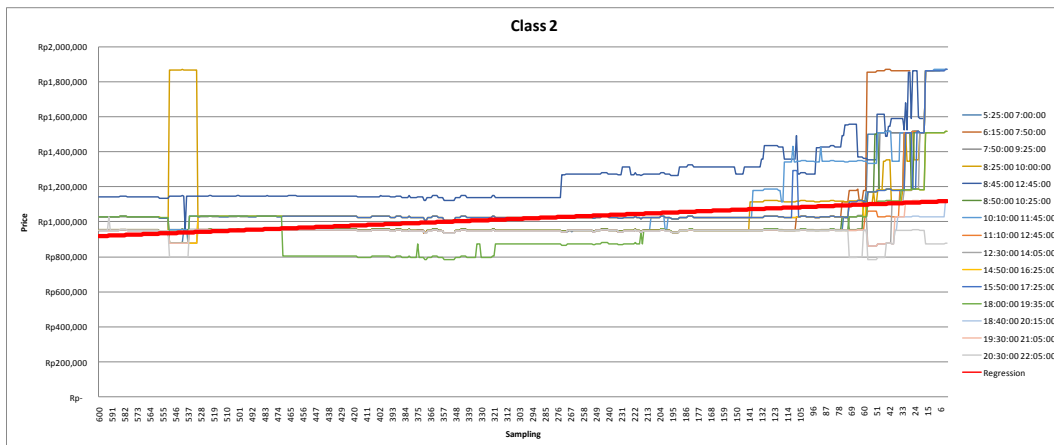


Figure 4.8 Garuda Indonesia Airlines Ticket Price Regression Result of Class 2 for Departure Date December 14th, 2015

The regression result for the Garuda Indonesia Airlines on departure date December 14th, 2015 for class 3 can be seen in the figure 4.9 below.

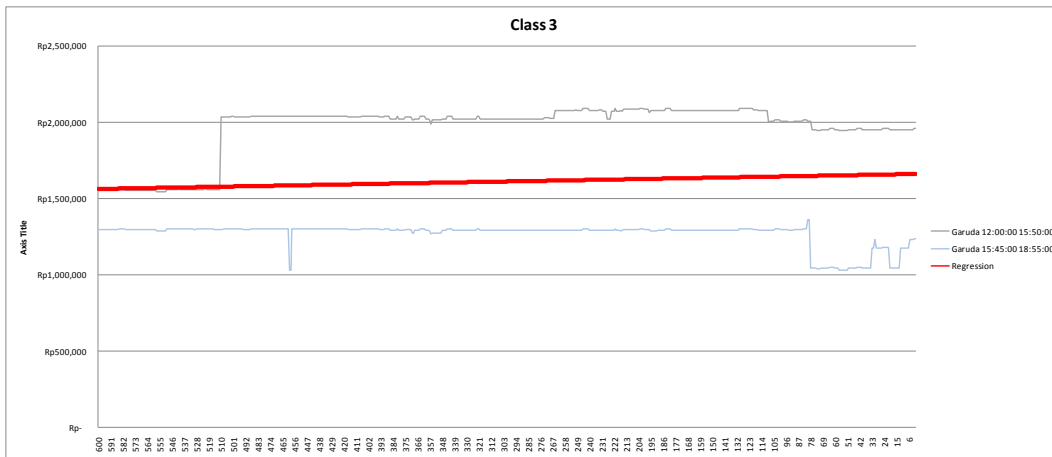


Figure 4.9 Garuda Indonesia Airlines Ticket Price Regression Result of Class 1 for Departure Date December 14th, 2015

The classification result for Garuda Indonesia Airlines on departure date December 17th, 2015 can be seen in the following table.

Table 4.2 Classification of Garuda Indonesia Airlines on Departure Date December 17th, 2015

Class 1	Class 2	Class 3
5:25:00	7:50:00	12:00:00
6:15:00	8:25:00	18:40:00
11:10:00	8:45:00	
20:30:00	8:50:00	
21:50:00	10:10:00	
22:50:00	12:30:00	
	13:30:00	
	14:50:00	
	15:45:00	
	15:50:00	
	19:30:00	

After the data is classified, the data is regressed for each class to know the behaviour of each class. The regression result for the Garuda Indonesia Airlines on departure date December 17th, 2015 for class 1 can be seen in the figure 4.10 below.

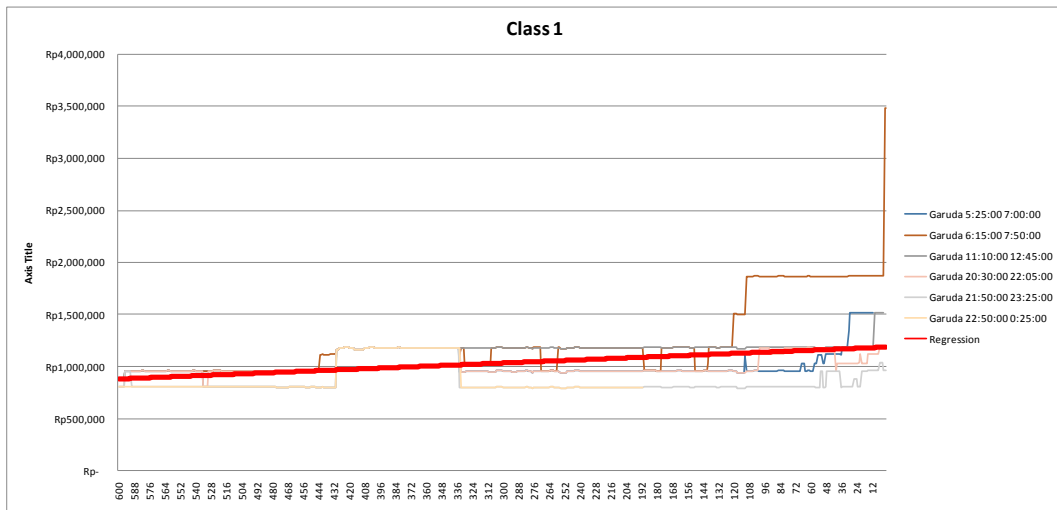


Figure 4.10 Garuda Indonesia Airlines Ticket Price Regression Result of Class 1 for Departure Date December 17th, 2015

The regression result for the Garuda Indonesia Airlines on departure date December 17th, 2015 for class 2 can be seen in the figure 4.11 below.

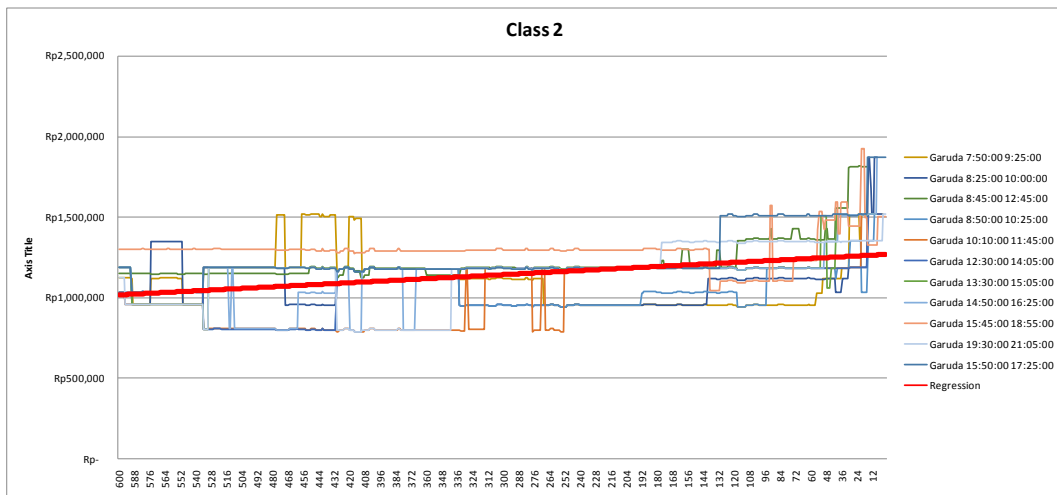


Figure 4.11 Garuda Indonesia Airlines Ticket Price Regression Result of Class 2 for Departure Date December 17th, 2015

The regression result for the Garuda Indonesia Airlines on departure date December 17th, 2015 for class 3 can be seen in the figure 4.12 below.

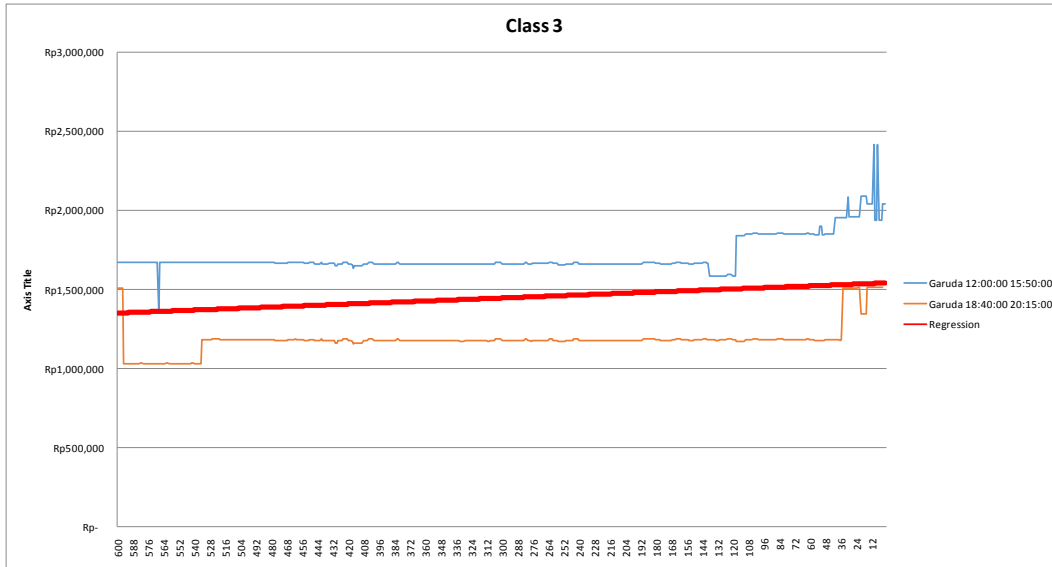


Figure 4.12 Garuda Indonesia Airlines Ticket Price Regression Result of Class 3 for Departure Date December 17th, 2015

4.3.2 Multiple Regression

The multiple regression is done to whether there is influence of each departure time's price change to other departure time's price change. The multiple regression will be done by using R-Project Software. The regression is done only for the Garuda Indonesia Airlines data for both December 14th, 2015 and December 17th, 2015 departure date.

The dependent variables of the model are the departure time of each airline carrier per departure date. The independent variables to look for the influence in price changes are the lag variable of the other departure time from the same departure date of the same airline carrier with the dependent variables.

The hypothesis of this experiment is that the fare of a departure time of an airline carrier react to their fares of other departure time within the same departure date in the previous period (lag) as well as its own fare in the preceding (lag) departure time. The model of the multiple regression are as follows:

$$P_t^{A_n} = \alpha + \beta_1 P_{t-m}^{A_1} + \beta_2 P_{t-m}^{A_2} + \beta_3 P_{t-m}^{A_3} + \dots + \beta_i P_{t-m}^{A_n} + Days + Days^2$$

With:

- $P_t^{A_n}$ = Price of Airline A on the departure time n in the period t
- Days = Number of period before departure
- m = Lag period

There are constant number of days before departure and the squared number of days before departure in the regression model. The constant number of days imply that the ticket price increases constantly (linear) while the squared number of days imply that the fare increases over time. By including those two variables, we can know whether a departure time price changes in linear or exponential behaviour by looking at the significant value.

The equation model contains an autoregressive component. A single equation with an autoregressive component may, under certain assumptions (stability and no serial correlation), be estimated using ordinary least squares. In order to account for correlation between disturbances for the same observation, the model is estimated using seemingly unrelated regression (SUR) (Pels and Rietveld, 2004).

Before the regression is done, the airline ticket price data are classified into several groups based on the departure time. The groups are morning flights, noon flights, afternoon flights, and evening flights. After the data is divided into these groups, the data similarities is checked by using cluster variables on Minitab software. After the data on each group is classified into several clusters, the data that are chosen to be regressed are the highest price of each cluster with the assumption that the schedules with the highest price are tigher than the other schedule. The detailed explanation can be seen in the next section.

4.3.2.1 Departure Time Clustering

The airline ticket price data is clustered by using Minitab software into three clusters. The clustering is done in order to simplify the data by grouping the that has similar price and behaviour into one cluster. The schedule that has the higher price in each cluster is chosen to be regressed. The clustering result of the Garuda Indonesia Airlines data on departure date December 14th, 2015 can be seen in the table 4.3 below.

Table 4.3 Clustering Result on Departure Date December 14th, 2015

Flight Groups	Clusters		
	1	2	3
Morning Flights	GA 5:25		GA 8:45
	GA 6:15		
	GA 7:50	GA 8:25	GA 10:10
	GA 8:50		
Noon Flights	GA 11:10		
	GA 12:30	GA 12:00	GA 13:30
	GA 14:50		
Afternoon Flights	GA 15:45	GA 15:50	GA 18:40
		GA 18:00	
Evening Flights	GA 19:30	GA 20:30	GA 22:50
	GA 21:50		

The clustering result of the Garuda Indonesia Airlines data on departure date December 17th, 2015 can be seen in the table 4.4 below.

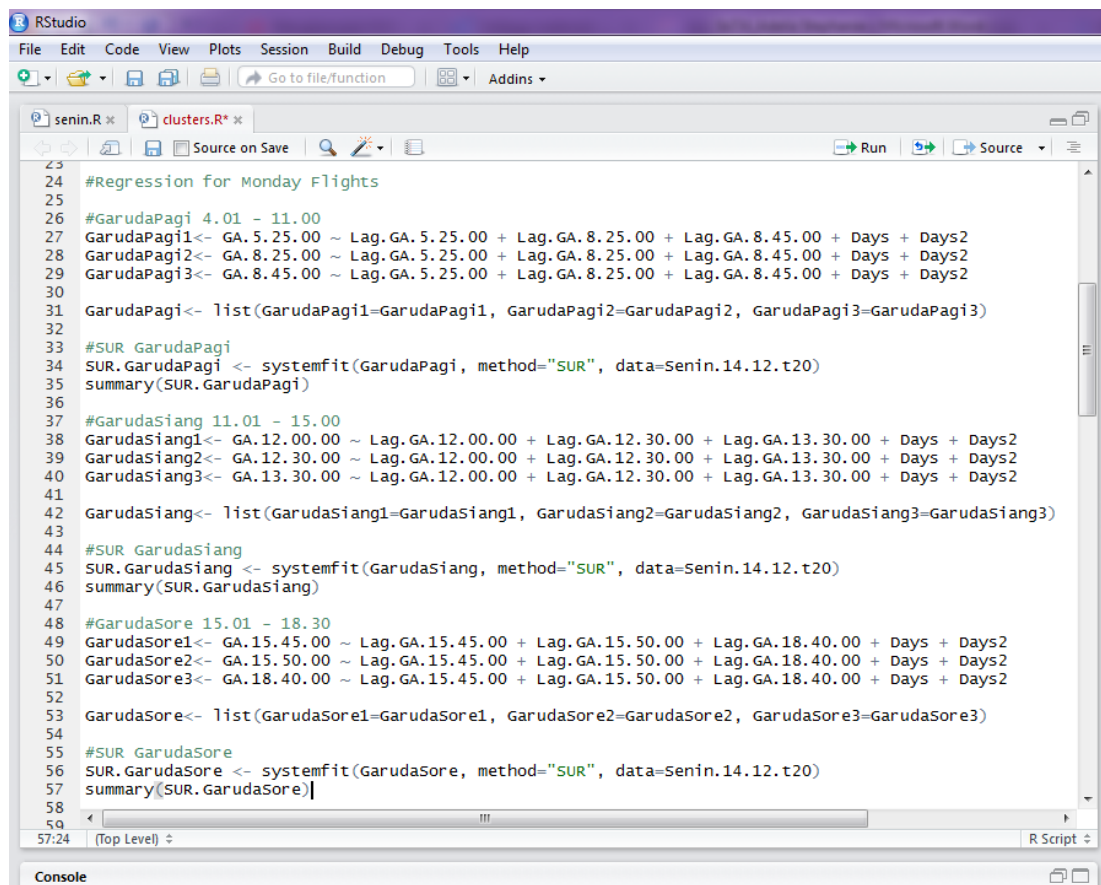
Table 4.4 Clustering Result on Departure Date December 17th, 2015

Flight Groups	Clusters		
	1	2	3
Morning Flights	GA 5:25	GA 6:15	
	GA 8:25	GA 8:45	GA 7:50
	GA 8:50	GA 10:10	
Noon Flights	GA 11:10		
	GA 12:30	GA 12:00	GA 13:30
	GA 14:50		
Afternoon Flights	GA 15:45	GA 18:40	GA 15:50
Evening Flights	GA 19:30	GA 20:30	GA 21:50
			GA 22:50

4.3.2.2 Multiple Regression for Garuda Indonesia Airlines

In this section, the regression analysis is done for all flight groups on both departure date December 14th, 2015 and December 17th, 2015. The regression is done by using the regression model in the previous section. The regression is done in several times with different lag. The lag period that are used in this experiment are $t-1$, $t-10$, and $t-20$. The purpose of using several lag period is to know the what model represents the actual behaviour of airline ticket price changes the best.

The regression is done by using R-Project software with SUR method. The example of the coding that is used to regress the Monday flight data with lag period $t-20$ can be in the figure 4.14 below.



```
23
24 #Regression for Monday Flights
25
26 #GarudaPagi 4.01 - 11.00
27 GarudaPagi1<- GA. 5.25.00 ~ Lag.GA.5.25.00 + Lag.GA.8.25.00 + Lag.GA.8.45.00 + Days + Days2
28 GarudaPagi2<- GA. 8.25.00 ~ Lag.GA.5.25.00 + Lag.GA.8.25.00 + Lag.GA.8.45.00 + Days + Days2
29 GarudaPagi3<- GA. 8.45.00 ~ Lag.GA.5.25.00 + Lag.GA.8.25.00 + Lag.GA.8.45.00 + Days + Days2
30
31 GarudaPagi<- list(GarudaPagi1=GarudaPagi1, GarudaPagi2=GarudaPagi2, GarudaPagi3=GarudaPagi3)
32
33 #SUR GarudaPagi
34 SUR.GarudaPagi <- systemfit(GarudaPagi, method="SUR", data=Senin.14.12.t20)
35 summary(SUR.GarudaPagi)
36
37 #GarudaSiang 11.01 - 15.00
38 GarudaSiang1<- GA.12.00.00 ~ Lag.GA.12.00.00 + Lag.GA.12.30.00 + Lag.GA.13.30.00 + Days + Days2
39 GarudaSiang2<- GA.12.30.00 ~ Lag.GA.12.00.00 + Lag.GA.12.30.00 + Lag.GA.13.30.00 + Days + Days2
40 GarudaSiang3<- GA.13.30.00 ~ Lag.GA.12.00.00 + Lag.GA.12.30.00 + Lag.GA.13.30.00 + Days + Days2
41
42 GarudaSiang<- list(GarudaSiang1=GarudaSiang1, GarudaSiang2=GarudaSiang2, GarudaSiang3=GarudaSiang3)
43
44 #SUR GarudaSiang
45 SUR.GarudaSiang <- systemfit(GarudaSiang, method="SUR", data=Senin.14.12.t20)
46 summary(SUR.GarudaSiang)
47
48 #GarudaSore 15.01 - 18.30
49 GarudaSore1<- GA.15.45.00 ~ Lag.GA.15.45.00 + Lag.GA.15.50.00 + Lag.GA.18.40.00 + Days + Days2
50 GarudaSore2<- GA.15.50.00 ~ Lag.GA.15.45.00 + Lag.GA.15.50.00 + Lag.GA.18.40.00 + Days + Days2
51 GarudaSore3<- GA.18.40.00 ~ Lag.GA.15.45.00 + Lag.GA.15.50.00 + Lag.GA.18.40.00 + Days + Days2
52
53 GarudaSore<- list(GarudaSore1=GarudaSore1, GarudaSore2=GarudaSore2, GarudaSore3=GarudaSore3)
54
55 #SUR GarudaSore
56 SUR.GarudaSore <- systemfit(GarudaSore, method="SUR", data=Senin.14.12.t20)
57 summary(SUR.GarudaSore)
58
59
```

Figure 4.14 R-Project Coding for Monday Flight

The example of the regression result for Monday morning flight can be seen in the figure 4.15 below.

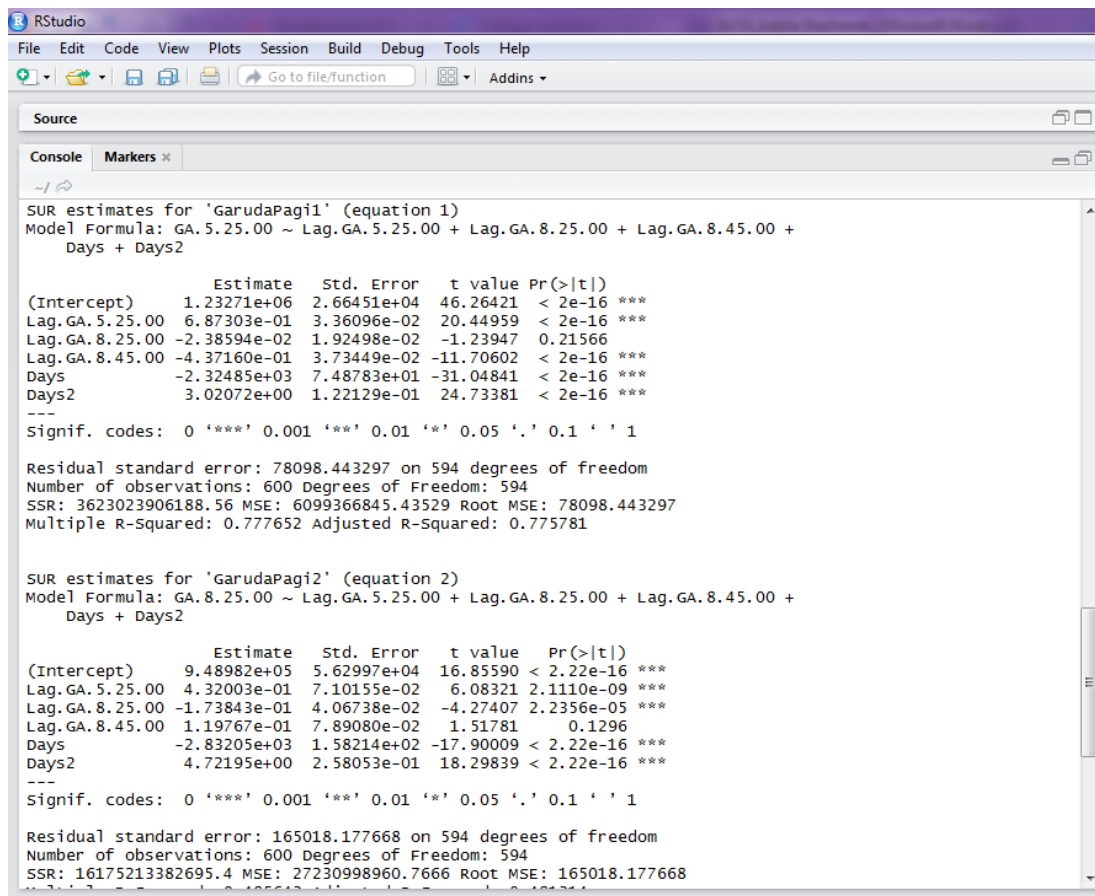


Figure 4.15 R-Project Regression Result for Monday Morning Flight

In order for the result to be easy to read and compare, the result is recorded into some result tables. The result tables and analysis can be seen in the next chapter.

4.3.2.3 Regression for Garuda Indonesia Airlines (Morning Flights)

In this section, the regression analysis more thorough for morning flights on both departure date December 14th, 2015 and December 17th, 2015. The regression is done by using the regression model in the previous section. The regression is done in several times with different lag than the previous experiment. The lag period that are used in this experiment are $t-1$, $t-2$, $t-3$, $t-4$ and $t-5$. The purpose of this experiment is to know how different the result will be between the lag periods with small difference one another. The researcher wants to know the pattern of the different result of the experiment by using small lag variable difference.

This experiment also done by using R-Project software with the same coding with the previous experiment. The result tables and analysis also can be seen in the next chapter.

(This page is intentionally left blank)

CHAPTER 5

INTERPRETATION AND ANALYSIS

In this chapter, the graphical representation, tables, and regression results will be interpreted and analyzed thoroughly. The interpretation analysis is done for all the data collection and numerical experiment.

5.1 Descriptive Analysis Interpretation

All the graphical representation of the descriptive analysis can be seen in the previous chapter. Based on those figures, it can be seen that all of the classes from both Monday and Thursday have increasing regression value. Even though all of the classes has the same trend, their price ranges are different. It can be seen in the graphical representation of regression summary.

The regression result summary for the Garuda Indonesia Airlines on both departure date and all classes can be seen in the figure 5.1 below.



Figure 5.1 Garuda Indonesia Airlines Ticket Price Regression Result Summary

The classification is done based on the average price similarity. Class 1 has the lowest level, class 2 is the middle level, and class 3 is the highest level. It can be seen in the figure 5.1 above that the regression result for Monday class 1

has lower prices than Thursday class 1. It means that the prices of departure times on Thursday class 1 has tendency to be higher than the prices of departure times on Monday class 1. The regression result for Monday class 2 also has lower prices than Thursday class 2. It means that the prices of departure times on Thursday class 2 has tendency to be higher than the prices of departure times on Monday class 2.

The Monday class 3 also has the different regression result, which is higher than Thursday class 3. It means that the prices of departure times on Monday class 3 has tendency to be higher than the prices of departure times on Thursday class 3.

It can also be seen that the Thursday class 1 and Monday class 2 has similar price range but the Thursday class 1's slope is slightly higher than Monday class 2.

Based on the regression result summary, it can be concluded that if the ticket prices data is classified by using k-means and then regressed for each classes, the prices on Thursday are higher than prices on Monday, except for class 3. The prices on Monday class 3 are higher than Thursday class 3.

5.2 Multiple Regression Interpretation and Analysis

The multiple regression coding example and result example from R-Project has been showed in the previous chapter. In this chapter, the results that will be showed are the result that has been interpreted into some result tables. The detailed analysis can be seen in the next sections.

5.2.1 Garuda Indonesia based on Timely Flight Groups

The summary of coefficient correlation regression result for Morning flight group by using R-Project software can be seen in the table 5.1 below.

Table 5.1 Regression Summary for Morning Flight Group

Lag Period	Monday				Thursday			
t-20	Coefficient	GA Morning			Coefficient	GA Morning		
	Lag Var.	GA 5:25	GA 8:25	GA 8:45	Lag Var.	GA 7:50	GA 8:25	GA 8:45
	GA 5:25	0.687	0.432	0.208	GA 7:50	0.984	0.168	0.314
	GA 8:25				GA 8:25	-0.653		-0.123
	GA 8:45	-0.437		-0.145	GA 8:45		0.148	
t-10	Coefficient	GA Morning			Coefficient	GA Morning		
	Lag Var.	GA 5:25	GA 8:25	GA 8:45	Lag Var.	GA 7:50	GA 8:25	GA 8:45
	GA 5:25	0.856	0.212	0.158	GA 7:50	0.910	0.154	0.254
	GA 8:25		0.431	-0.037	GA 8:25	-0.280	0.241	-0.173
	GA 8:45	-0.577	-0.293		GA 8:45	-0.211	-0.182	0.116
t-1	Coefficient	GA Morning			Coefficient	GA Morning		
	Lag Var.	GA 5:25	GA 8:25	GA 8:45	Lag Var.	GA 7:50	GA 8:25	GA 8:45
	GA 5:25	0.967			GA 7:50	1.066	0.074	0.081
	GA 8:25	-0.022	0.924	-0.025	GA 8:25		0.840	-0.077
	GA 8:45	-0.332	-0.331	0.466	GA 8:45	-0.217	-0.258	0.649

Based on the table 5.1 above, there are several lag variables that influence the price changes of a departure time prices and some of them did not have any influence to a departure time prices. The example of how to read the table above for Monday flight on departure time 5:25, if the price of lag 5:25 increased by 1 then the price of 5:25 is increased by 0.687 and if the price of lag 8:45 increased by 1 point then the price of 5:25 is decreased by 0.437. The columns that do not have any value means that there is no influence from the lag variable to the departure time.

It can be seen in the table that the results from Monday and Thursday are different. It also varies from different lag period. If we look at the Monday flights, the regression result with lag period $t-20$ has less correlations compared to the regression result with lag period $t-10$ and $t-1$. The regression result with lag period $t-20$ has 5/9 correlations, while the regression result with lag period $t-10$ and $t-1$ has 7/9 correlations. If we look at the Thursday flights, the regression result with

lag period $t-20$ also has less correlations compared to the regression result with lag period $t-10$ and $t-1$. The regression result with lag period $t-20$ has 6/9 correlations, the regression result with lag period $t-10$ has full correlations, and the regression result with lag period $t-1$ has 8/9 correlations.

If we compare the result from Monday and Thursday flights, the Thursday flights has more correlations than Monday flights. For lag period $t-20$, the Monday flights are 55% correlated to each other, while Thursday flights are 67% correlated to each other. For lag period $t-10$, the Monday flights are 78% correlated to each other, while Thursday flights are 100% correlated to each other. For lag period $t-1$, the Monday flights are 78% correlated to each other, while Thursday flights are 89% correlated to each other. It means that for Morning flight group, the Thursday flight has higher price competition than Monday flights.

We can also compare the result by looking at the significance of the correlation of each flights. Each of the coefficient result has different significance value. The ones with three stars (***) has significant value < 0.001 , the ones with two stars (**) has significant value $0.001 - < 0.01$, and the ones with one star (*) has significant value $0.01 - 0.05$. The significance code summary of those coefficient result and the multiple R-squared value of Morning flight group can be seen in the table 5.2 below.

Table 5.2 Significance Code and Multiple R-Squared Regression Summary for Morning Flight Group

Lag Period	Monday				Thursday			
	Sig. Code	GA Morning			Sig. Code	GA Morning		
t-20	Lag Var.	GA 5:25	GA 8:25	GA 8:45	Lag Var.	GA 7:50	GA 8:25	GA 8:45
	GA 5:25	***	***	***	GA 7:50	***	***	***
	GA 8:25				GA 8:25	***		***
	GA 8:45	***		***	GA 8:45		***	
	Mul. R Squared	0.778	0.486	0.891	Mul. R Squared	0.516	0.424	0.826
t-10	Sig. Code	GA Morning			Sig. Code	GA Morning		
	Lag Var.	GA 5:25	GA 8:25	GA 8:45	Lag Var.	GA 7:50	GA 8:25	GA 8:45
	GA 5:25	***	**	***	GA 7:50	***	***	***
	GA 8:25		***	**	GA 8:25	***	***	***
	GA 8:45	***	***		GA 8:45	*	**	***
	Mul. R Squared	0.875	0.525	0.887	Mul. R Squared	0.418	0.388	0.844
t-1	Sig. Code	GA Morning			Sig. Code	GA Morning		
	Lag Var.	GA 5:25	GA 8:25	GA 8:45	Lag Var.	GA 7:50	GA 8:25	GA 8:45
	GA 5:25	***			GA 7:50	***	***	***
	GA 8:25	**	***	*	GA 8:25		***	***
	GA 8:45	***	***	***	GA 8:45	***	***	***
	Mul. R Squared	0.964	0.921	0.922	Mul. R Squared	0.858	0.837	0.932

The ones with three stars (***) means that the lag variable has very significant influence to a departure date. The ones with two stars (**) means that the lag variable has significant influence to a departure date. The ones with one star (*) means that the lag variable has less significant influence to a departure date. How to read this table is the same with how to read table 5.1 but to know the positive or negative influence (correlation), it has to be read with the table 5.1. The example for departure time 5:25, the lag 5:25 has a very significant (***) positive (0.687) correlation to 5:25 prices and the lag 8:45 has a very significant (***) negative (-0.437) correlation to 5:25 prices.

If we want to compare the significance of the correlation of each flights, we can only read the table 5.2 above. We can see in this table that the significance of the correlation from Monday and Thursday flights are different. The

significance of the correlation is related to the previous correlation results. If we look at the Monday flights, the regression result with lag period $t-20$ has 5 very significant correlations. The regression result with lag period $t-10$ has 5 very significant correlations and 2 significant correlations. The regression result with lag period $t-1$ has 5 very significant correlations, 1 significant correlation, and 1 less significant correlation. If we look at the Thursday flights, the regression result with lag period $t-20$ has 6 very significant correlations. The regression result with lag period $t-10$ has 7 very significant correlations, 1 significant correlation, and 1 less significant correlation. The regression result with lag period $t-1$ has 8 very significant correlations.

If we compare the significance of the correlation from Monday and Thursday flights, the Thursday flights has more very significant correlation than Monday flights. For lag period $t-20$, the Monday flights are 100% very significantly correlated to each other, while Thursday flights are also 100% very significantly correlated to each other. For lag period $t-10$, the Monday flights are 71% very significantly correlated to each other, while Thursday flights are 78% very significantly correlated to each other. For lag period $t-1$, the Monday flights are 71% very significantly correlated to each other, while Thursday flights are 100% very significantly correlated to each other. There is no real competition for the correlation that does not have very significant correlation. This result affirms the previous conclusion that for Morning flight group, the Thursday flights have higher price competition than Monday flights.

The summary of coefficient correlation regression result for Noon flight group by using R-Project software can be seen in the table 5.3 below.

Table 5.3 Regression Summary for Noon Flight Group

Lag Period	Monday				Thursday			
t-20	Coefficient	GA Noon			Coefficient	GA Noon		
	Lag Var.	GA 12:00	GA 12:30	GA 13:30	Lag Var.	GA 12:00	GA 13:30	GA 14:50
	GA 12:00	0.322	-0.239	-0.106	GA 12:00	0.144	-0.423	-0.125
	GA 12:30	-0.226	0.299	-0.112	GA 13:30		0.794	-0.125
	GA 13:30	-0.193	0.211	0.465	GA 14:50	-0.120		0.250
t-10	Coefficient	GA Noon			Coefficient	GA Noon		
	Lag Var.	GA 12:00	GA 12:30	GA 13:30	Lag Var.	GA 12:00	GA 13:30	GA 14:50
	GA 12:00	0.307	-0.367	-0.256	GA 12:00	0.097	-0.477	-0.269
	GA 12:30		0.869		GA 13:30		0.815	
	GA 13:30	-0.274	-0.131	0.602	GA 14:50	-0.085		0.379
t-1	Coefficient	GA Noon			Coefficient	GA Noon		
	Lag Var.	GA 12:00	GA 12:30	GA 13:30	Lag Var.	GA 12:00	GA 13:30	GA 14:50
	GA 12:00	0.697	-0.180	-0.150	GA 12:00	0.520	-0.257	-0.289
	GA 12:30		1.007		GA 13:30	-0.128	0.822	-0.066
	GA 13:30	-0.162	-0.114	0.868	GA 14:50	-0.063		0.878

It can be seen in the table 5.3 above that the results from Monday and Thursday are different. It also varies from different lag period. If we look at the Monday flights, the regression result with lag period $t-20$ has more correlations compared to the regression result with lag period $t-10$ and $t-1$. The regression result with lag period $t-20$ has 9/9 correlations, while the regression result with lag period $t-10$ and $t-1$ has 7/9 correlations. If we look at the Thursday flights, the regression result with lag period $t-20$ also has more correlation compared to the regression result with lag period $t-10$ but less correlations compared to the regression result with lag period $t-1$. The regression result with lag period $t-20$ has 7/9 correlation, the regression result with lag period $t-10$ has 6/9 correlations, and the regression result with lag period $t-1$ has 8/9 correlations.

If we compare the result from Monday and Thursday flights, the Monday flights has more correlation than Thursday flights, except for lag period $t-1$. For lag period $t-20$, the Monday flights are 100% correlated to each other, while

Thursday flights are 78% correlated to each other. For lag period $t-10$, the Monday flights are 78% correlated to each other, while Thursday flights are 67% correlated to each other. For lag period $t-1$, the Monday flights are 78% correlated to each other, while Thursday flights are 89% correlated to each other. It means that for Morning flight group, the Monday flight has price higher competition than Thursday flights for lag period $t-20$ and $t-10$, and the Thursday flight has higher price competition than Monday flights for lag period $t-1$.

The significance code summary of those coefficient result and the multiple R-squared value of Noon flight group can be seen in the table 5.4 below.

Table 5.4 Significance Code and Multiple R-Squared Regression Summary for Noon Flight Group

Lag Period	Monday				Thursday			
	Sig. Code	GA Noon			Sig. Code	GA Noon		
t-20	Lag Var.	GA 12:00	GA 12:30	GA 13:30	Lag Var.	GA 12:00	GA 13:30	GA 14:50
	GA 12:00	***	***	**	GA 12:00	***	***	**
	GA 12:30	***	**	*	GA 13:30		***	*
	GA 13:30	***	**	***	GA 14:50	***		***
	Mul. R Squared	0.771	0.270	0.816	Mul. R Squared	0.473	0.503	0.525
t-10	Sig. Code	GA Noon			Sig. Code	GA Noon		
	Lag Var.	GA 12:00	GA 12:30	GA 13:30	Lag Var.	GA 12:00	GA 13:30	GA 14:50
	GA 12:00	***	***	***	GA 12:00	**	***	***
	GA 12:30		***		GA 13:30		***	
	GA 13:30	***	*	***	GA 14:50	***		***
	Mul. R Squared	0.787	0.547	0.864	Mul. R Squared	0.459	0.509	0.567
t-1	Sig. Code	GA Noon			Sig. Code	GA Noon		
	Lag Var.	GA 12:00	GA 12:30	GA 13:30	Lag Var.	GA 12:00	GA 13:30	GA 14:50
	GA 12:00	***	***	***	GA 12:00	***	***	***
	GA 12:30		***		GA 13:30	***	***	***
	GA 13:30	***	***	***	GA 14:50	***		***
	Mul. R Squared	0.909	0.922	0.960	Mul. R Squared	0.732	0.825	0.892

We can see in the table 5.4 above that the significance of the correlation from Monday and Thursday flights are different. The significance of the

correlation is related to the previous correlation results. If we look at the Monday flights, the regression result with lag period $t-20$ has 5 very significant correlations, 3 significant correlations and 1 less significant correlation. The regression result with lag period $t-10$ has 6 very significant correlations and 1 less significant correlations. The regression result with lag period $t-1$ has 7 very significant correlations. If we look at the Thursday flights, the regression result with lag period $t-20$ has 5 very significant correlations, 1 significant correlation and 1 less significant correlation. The regression result with lag period $t-10$ has 5 very significant correlations and 1 significant correlations. The regression result with lag period $t-1$ has 8 very significant correlations.

If we compare the significance of the correlation from Monday and Thursday flights, the Thursday flights has more very significant correlation than Monday flights. For lag period $t-20$, the Monday flights are 55% very significantly correlated to each other, while Thursday flights are 71% very significantly correlated to each other. For lag period $t-10$, the Monday flights are 86% very significantly correlated to each other, while Thursday flights are 83% very significantly correlated to each other. For lag period $t-1$, the Monday flights are 100% very significantly correlated to each other, while Thursday flights are also 100% very significantly correlated to each other. This result is different with the previous conclusion. There is no real competition for the correlation that does not have very significant correlation. The conclusion of this result is the Thursday flights have higher price competition than Monday flights for lag period $t-20$ and $t-1$ but has lower price competition than Monday flights for lag period $t-10$.

The summary of coefficient correlation regression result for afternoon flight group by using R-Project software can be seen in the table 5.5 below.

Table 5.5 Regression Summary for Afternoon Flight Group

Lag Period	Monday				Thursday			
t-20	Coefficient	GA Afternoon			Coefficient	GA Afternoon		
	Lag Var.	GA 15:45	GA 15:50	GA 18:40	Lag Var.	GA 15:45	GA 15:50	GA 18:40
	GA 15:45	0.485	-0.828	0.130	GA 15:45	0.411	-0.316	
	GA 15:50	0.225	0.773	0.308	GA 15:50	-0.082	0.219	-0.566
	GA 18:40	-0.924	0.360	-0.512	GA 18:40	-0.294	0.257	0.693
t-10	Coefficient	GA Afternoon			Coefficient	GA Afternoon		
	Lag Var.	GA 15:45	GA 15:50	GA 18:40	Lag Var.	GA 15:45	GA 15:50	GA 18:40
	GA 15:45	0.494	-0.583		GA 15:45	0.345	-0.347	
	GA 15:50	0.145	0.795	0.130	GA 15:50	-0.254	0.407	-0.188
	GA 18:40	-0.825		-0.131	GA 18:40			0.212
t-1	Coefficient	GA Afternoon			Coefficient	GA Afternoon		
	Lag Var.	GA 15:45	GA 15:50	GA 18:40	Lag Var.	GA 15:45	GA 15:50	GA 18:40
	GA 15:45	0.782	-0.164	-0.080	GA 15:45	0.678	-0.217	-0.229
	GA 15:50	0.069	1.004	0.064	GA 15:50	-0.239	0.698	-0.237
	GA 18:40	-0.804	-0.551	0.289	GA 18:40	0.077	0.088	1.004

It can be seen in the table 5.5 above that the results from Monday and Thursday are different. It also varies from different lag period. If we look at the Monday flights, the regression result with lag period $t-20$ and $t-1$ has more correlations compared to the regression result with lag period $t-10$. The regression result with lag period $t-20$ and $t-1$ has 9/9 correlations, while the regression result with lag period $t-10$ has 7/9 correlations. If we look at the Thursday flights, the regression result with lag period $t-20$ and $t-1$ also has more correlations compared to the regression result with lag period $t-10$. The regression result with lag period $t-20$ has 8/9 correlation, the regression result with lag period $t-10$ has 6/9 correlations, and the regression result with lag period $t-1$ has 9/9 correlations.

If we compare the result from Monday and Thursday flights, the Monday flights has more correlation than Thursday flights. For lag period $t-20$, the Monday flights are 100% correlated to each other, while Thursday flights are 89% correlated to each other. For lag period $t-10$, the Monday flights are 78%

correlated to each other, while Thursday flights are 67% correlated to each other. For lag period t-1, the Monday flights are 100% correlated to each other, while Thursday flights are also 100% correlated to each other. It means that for Afternoon flight group, the Monday flight has higher price competition than Thursday flights.

The significance code summary of those coefficient result and the multiple R-squared value of afternoon flight group can be seen in the table 5.6 below.

Table 5.6 Significance Code and Multiple R-Squared Regression Summary for Afternoon Flight Group

Lag Period	Monday				Thursday			
	Sig. Code	GA Afternoon			Sig. Code	GA Afternoon		
	Lag Var.	GA 15:45	GA 15:50	GA 18:40	Lag Var.	GA 15:45	GA 15:50	GA 18:40
t-20	GA 15:45	***	***	***	GA 15:45	***	***	
	GA 15:50	***	***	***	GA 15:50	*	***	***
	GA 18:40	***	*	***	GA 18:40	***	***	***
	Mul. R Squared	0.586	0.646	0.364	Mul. R Squared	0.239	0.801	0.393
	Sig. Code	GA Afternoon			Sig. Code	GA Afternoon		
	Lag Var.	GA 15:45	GA 15:50	GA 18:40	Lag Var.	GA 15:45	GA 15:50	GA 18:40
t-10	GA 15:45	***	***		GA 15:45	***	***	
	GA 15:50	***	***	***	GA 15:50	***	***	**
	GA 18:40	***		***	GA 18:40			*
	Mul. R Squared	0.590	0.744	0.230	Mul. R Squared	0.282	0.835	0.284
	Sig. Code	GA Afternoon			Sig. Code	GA Afternoon		
	Lag Var.	GA 15:45	GA 15:50	GA 18:40	Lag Var.	GA 15:45	GA 15:50	GA 18:40
t-1	GA 15:45	***	***	***	GA 15:45	***	***	***
	GA 15:50	***	***	***	GA 15:50	***	***	***
	GA 18:40	***	***	***	GA 18:40	***	***	***
	Mul. R Squared	0.790	0.933	0.446	Mul. R Squared	0.690	0.932	0.841

In the table 5.6 above, it can be seen that the significance of the correlation from Monday and Thursday flights are different. The significance of the correlation is related to the previous correlation results. If we look at the

Monday flights, the regression result with lag period $t-20$ has 8 very significant correlations and 1 less significant correlation. The regression result with lag period $t-10$ has 7 very significant correlations. The regression result with lag period $t-1$ has 9 very significant correlations. If we look at the Thursday flights, the regression result with lag period $t-20$ has 7 very significant correlations and 1 less significant correlation. The regression result with lag period $t-10$ has 4 very significant correlations, 1 significant correlation and 1 less significant correlation. The regression result with lag period $t-1$ has 9 very significant correlations.

If we compare the significance of the correlation from Monday and Thursday flights, the Thursday flights has more very significant correlation than Monday flights. For lag period $t-20$, the Monday flights are 89% very significantly correlated to each other, while Thursday flights are 87% very significantly correlated to each other. For lag period $t-10$, the Monday flights are 100% very significantly correlated to each other, while Thursday flights are 67% very significantly correlated to each other. For lag period $t-1$, the Monday flights are 100% very significantly correlated to each other, while Thursday flights are also 100% very significantly correlated to each other. There is no real competition for the correlation that does not have very significant correlation. This result affirms the previous conclusion that for afternoon flight group, the Monday flights have higher price competition than Thursday flights.

The summary of coefficient correlation regression result for evening flight group by using R-Project software can be seen in the table 5.7 below.

Table 5.7 Regression Summary for Evening Flight Group

Lag Period	Monday				Thursday			
t-20	Coefficient	GA Evening			Coefficient	GA Evening		
	Lag Var.	GA 19:30	GA 20:30	GA 22:50	Lag Var.	GA 19:30	GA 20:30	GA 21:50
	GA 19:30	0.444	-0.326		GA 19:30	0.368	-0.223	-0.301
	GA 20:30	-0.274			GA 20:30	-0.250	0.289	
	GA 22:50	-0.196	0.228		GA 21:50			0.502
t-10	Coefficient	GA Evening			Coefficient	GA Evening		
	Lag Var.	GA 19:30	GA 20:30	GA 22:50	Lag Var.	GA 19:30	GA 20:30	GA 21:50
	GA 19:30	0.747	-0.120		GA 19:30	0.454	-0.248	-0.363
	GA 20:30		0.337		GA 20:30	-0.321	0.292	-0.180
	GA 22:50	-0.822	-0.293		GA 21:50	-0.143	0.187	0.699
t-1	Coefficient	GA Evening			Coefficient	GA Evening		
	Lag Var.	GA 19:30	GA 20:30	GA 22:50	Lag Var.	GA 19:30	GA 20:30	GA 21:50
	GA 19:30	0.983			GA 19:30	0.778	-0.132	-0.151
	GA 20:30		0.893		GA 20:30	-0.170	0.793	-0.077
	GA 22:50	-1.133	-1.062	0.016	GA 21:50	-0.057		0.921

It can be seen in the table 5.7 above that the results from Monday and Thursday are different. It also varies from different lag period. If we look at the Monday flights, the regression result with lag period $t-20$, $t-10$ and $t-1$ has same correlations. The regression result with lag period $t-20$, $t-10$ and $t-1$ has 5/9 correlations. If we look at the Thursday flights, the regression result with lag period $t-20$ has less correlations compared to the regression result with lag period $t-10$ and $t-1$. The regression result with lag period $t-20$ has 6/9 correlation, the regression result with lag period $t-10$ has 9/9 correlations, and the regression result with lag period $t-1$ has 8/9 correlations.

If we compare the result from Monday and Thursday flights, the Monday flights has more correlation than Thursday flights. For lag period $t-20$, the Monday flights are 56% correlated to each other, while Thursday flights are 67% correlated to each other. For lag period $t-10$, the Monday flights are 56% correlated to each other, while Thursday flights are 100% correlated to each

other. For lag period $t-1$, the Monday flights are 56% correlated to each other, while Thursday flights are also 89% correlated to each other. It means that for evening flight group, the Thursday flight has higher price competition than Monday flights.

The significance code summary of those coefficient result and the multiple R-squared value of evening flight group can be seen in the table 5.8 below.

Table 5.8 Significance Code and Multiple R-Squared Regression Summary for Afternoon Flight Group

Lag Period	Monday				Thursday			
t-20	Sig. Code	GA Evening			Sig. Code	GA Evening		
	Lag Var.	GA 19:30	GA 20:30	GA 22:50	Lag Var.	GA 19:30	GA 20:30	GA 21:50
	GA 19:30	***	***		GA 19:30	***	***	***
	GA 20:30	***			GA 20:30	**	***	
	GA 22:50	**	***		GA 21:50			***
	Mul. R Squared	0.260	0.243	0.202	Mul. R Squared	0.712	0.424	0.397
t-10	Sig. Code	GA Evening			Sig. Code	GA Evening		
	Lag Var.	GA 19:30	GA 20:30	GA 22:50	Lag Var.	GA 19:30	GA 20:30	GA 21:50
	GA 19:30	***	*		GA 19:30	***	***	***
	GA 20:30		***		GA 20:30	***	***	**
	GA 22:50	***	***		GA 21:50	**	***	***
	Mul. R Squared	0.548	0.259	0.196	Mul. R Squared	0.811	0.587	0.617
t-1	Sig. Code	GA Evening			Sig. Code	GA Evening		
	Lag Var.	GA 19:30	GA 20:30	GA 22:50	Lag Var.	GA 19:30	GA 20:30	GA 21:50
	GA 19:30	***			GA 19:30	***	***	***
	GA 20:30		***		GA 20:30	***	***	*
	GA 22:50	***	***	*	GA 21:50	*		***
	Mul. R Squared	0.944	0.837	0.197	Mul. R Squared	0.944	0.906	0.929

In the table 5.8 above, it can be seen that the significance of the correlation from Monday and Thursday flights are different. The significance of the correlation is related to the previous correlation results. If we look at the Monday flights, the regression result with lag period $t-20$ has 4 very significant

correlations and 1 significant correlation. The regression result with lag period $t-10$ has 4 very significant correlations and 1 less significant correlation. The regression result with lag period $t-1$ also has 4 very significant correlations and 1 less significant correlation. If we look at the Thursday flights, the regression result with lag period $t-20$ has 5 very significant correlations and 1 significant correlation. The regression result with lag period $t-10$ has 7 very significant correlations and 2 significant correlations. The regression result with lag period $t-1$ has 6 very significant correlations and 2 less significant correlations.

If we compare the significance of the correlation from Monday and Thursday flights, the Thursday flights has more very significant correlation than Monday flights. For lag period $t-20$, the Monday flights are 80% very significantly correlated to each other, while Thursday flights are 83% very significantly correlated to each other. For lag period $t-10$, the Monday flights are 80% very significantly correlated to each other, while Thursday flights are 78% very significantly correlated to each other. For lag period $t-1$, the Monday flights are 80% very significantly correlated to each other, while Thursday flights are also 75% very significantly correlated to each other. This result is different with the previous conclusion. There is no real competition for the correlation that does not have very significant correlation. The conclusion of this result is the Thursday flights have higher price competition than Monday flights for lag period $t-20$ but has lower price competition than Monday flights for lag period $t-10$ and $t-1$.

If all the regression results based on lag period are compared, the multiple R-squared value is increasing from $t-20$ regression to $t-1$ regression. The multiple R-squared value of $t-10$ regression is higher than the multiple R-squared value of $t-20$ regression. Furthermore, the multiple R-squared value of $t-1$ regression is higher than the multiple R-squared value of $t-10$ regression. Then, it can be concluded that the regression by using lag variable $t-1$ is the most representative to model the behaviour of the airline ticket pricing data.

In order for the regression result per day to be seen clearly, the result is recaped into a graphical map. Each departure time that we can see in the tables before represent different clusters. The graphical mapping of the regression result for Monday based on clusters can be seen in the figure 5.2 below.

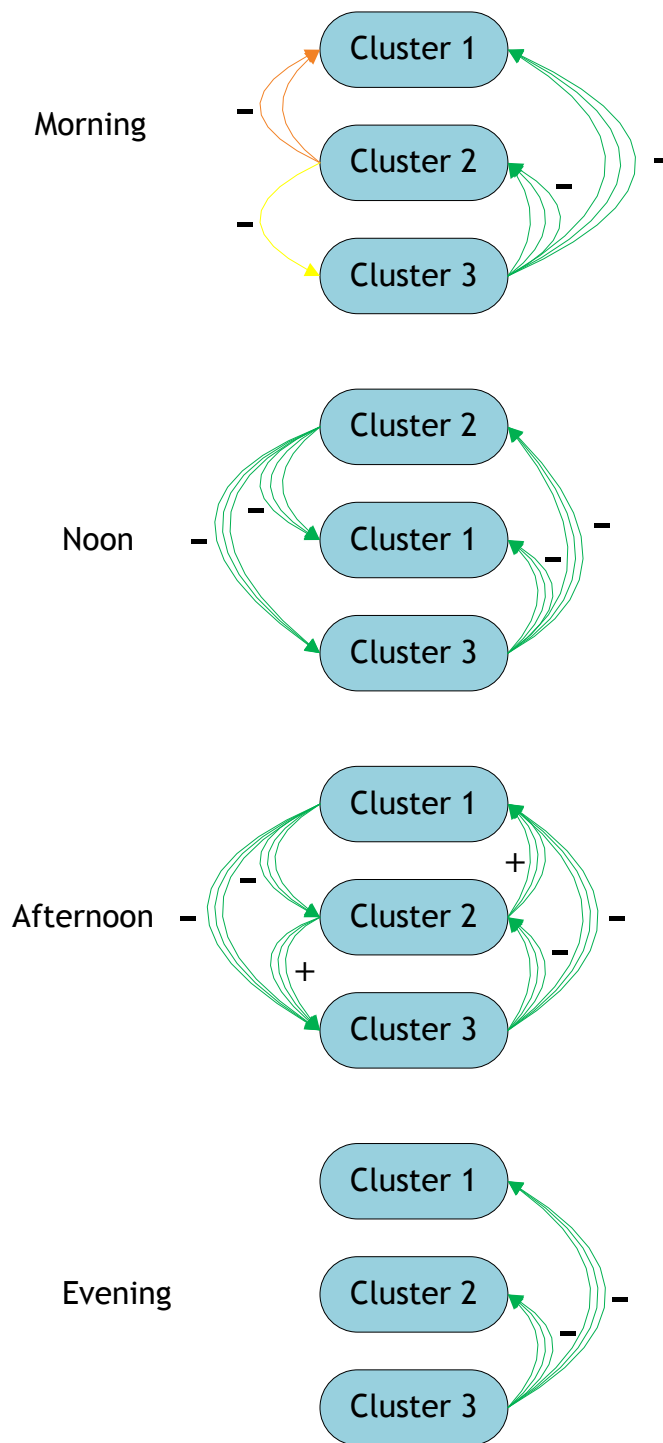


Figure 5.2 Graphical Mapping of Monday Regression Result

The graphical mapping of the regression result for Monday based on clusters can be seen in the figure 5.3 below.

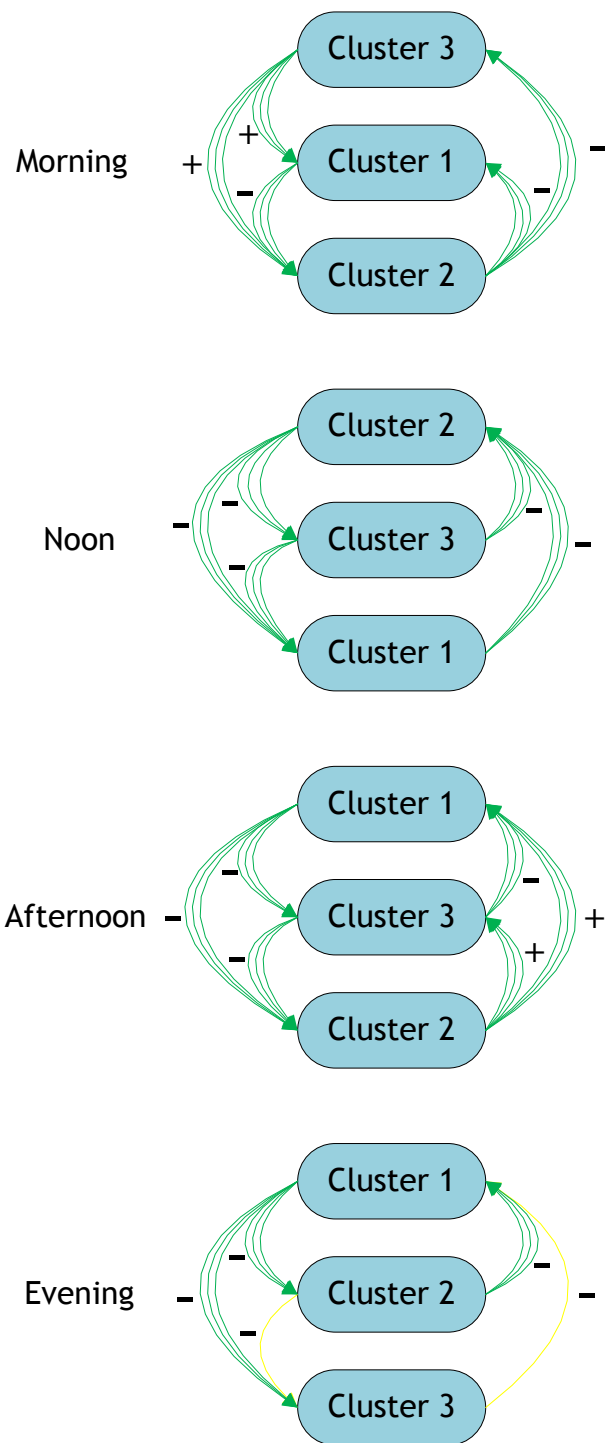


Figure 5.3 Graphical Mapping of Thursday Regression Result

If we compare the peak and non-peak departure time behaviour for Monday, almost all the peak departure time influence the non-peak departure time, except the lag 19:30. Meanwhile, not all the non-peak departure time influence the peak departure time. This means that the non-peak departure time

price always changes correspond to the peak departure time price changes. On the other side, most of the peak departure time price changes independently without really considering the non-peak departure time price changes.

If we see in the result table, the flight group with the most tight correlation is the afternoon group. Meanwhile, the flight group with the most untight correlation is the evening group. This means that the afternoon group has the most competitive prices, on the other side, the evening group has the less competitive prices.

For Thursday, almost all the peak departure time influence the non-peak departure time. Meanwhile, not all the non-peak departure time influence the peak departure time. This means that the non-peak departure time price always changes correspond to the peak departure time price changes. On the other side, most of the peak departure time price changes independently without really considering the non-peak departure time price changes.

If we see in the result table, the flight group with the most tight correlation is the afternoon group. Meanwhile, the other flight groups have similar tightness in correlation. This means that the afternoon group has the most competitive prices, on the other side, the other flight groups has the less competitive prices than the afternoon groups.

The regression is also done for all flight groups on Monday. The highest price of each flight group is chosen to be regressed. The regression is done with lag period $t-1$ because based on the previous result, the $t-1$ regression results have the highest multiple R-squared value. The regression result for Monday all flight groups with lag $t-1$ by using R-Project software can be seen in the table 5.9 below.

Table 5.9 Regression Summary for Monday All Flight Groups with Lag $t-1$

Coefficient	GA All			
Lag Variables	GA 8:45:00	GA 12:00:00	GA 15:45:00	GA 19:30:00
GA 8:45:00	0.6375	-0.2201	-0.1910	-0.1248
GA 12:00:00	-0.0447	0.8738	-0.0813	-0.0576
GA 15:45:00	-0.2204	-0.3198	0.6567	-0.1914
GA 19:30:00	-0.0977	-0.2463	-0.1596	0.8003
Sig. Code	GA All			
Lag Variables	GA 8:45:00	GA 12:00:00	GA 15:45:00	GA 19:30:00
GA 8:45:00	***	***	***	***
GA 12:00:00	*	***	***	***
GA 15:45:00	***	***	***	***
GA 19:30:00	*	***	***	***
Mul. R Squared	0.935771	0.948284	0.783603	0.703167

It can be seen in the table 5.9 above that all of the lag variables have influence to the original departure date prices whether it is negative influence or positive influence. The lag 8:45 and lag 15:45 has very significant influence to all the original departure time, while lag 12:00 and 19:30 has less significant influence to the departure time 8:45 and very significant influence to other departure times.

This table can also be used to see the competitiveness of all the peak departure times. If we see in the table above, all of the peak departure time has influence to each other. This means that all of the peak departure time has tight price competition.

The regression result for Thursday all flight groups with lag $t-1$ by using R-Project software can be seen in the table 5.10 below.

Table 5.10 Regression Summary for Thursday All Flight Groups with Lag $t-1$

Coefficient	GA All			
Lag Variables	GA 8:45:00	GA 12:00:00	GA 15:45:00	GA 19:30:00
GA 8:45:00	0.9005	-0.1295		-0.0646
GA 12:00:00	-0.2335	0.6043	-0.2720	-0.2318
GA 15:45:00	-0.0593	-0.0785	0.7904	-0.0739
GA 19:30:00	-0.0854	-0.1133	-0.0918	0.8954
Sig. Code	GA All			
Lag Variables	GA 8:45:00	GA 12:00:00	GA 15:45:00	GA 19:30:00
GA 8:45:00	***	***		**
GA 12:00:00	***	***	***	***
GA 15:45:00	***	***	***	***
GA 19:30:00	***	***	***	***
Mul. R Squared	0.950983	0.737452	0.729248	0.954296

It can be seen in the table 5.10 above that all of the lag variables have influence to the original departure date prices whether it is negative influence or positive influence, except the lag 8:45. The lag 12:00, lag 15:45, and lag 19:30 has very significant influence to all the original departure time, while lag 8:45 has very significant influence to departure time 8:45 and 12:00, significant influence to departure time 19:30, and no influence to departure time 15:45.

This table can also be used to see the competitiveness of all the peak departure times. If we see in the table above, all of the peak departure time has influence to each other, except lag 8:45 to departure time 15:45. This means that all of the peak departure time has tight price competition, except flight 15:45 because the price changes of flight 15:45 did not influenced by prices of flight 8:45 but the flight 8:45 price changes are influenced by prices of flight 15:45.

5.2.2 Garuda Indonesia (Morning Flights) based on Lag Period

. The regression result for Monday morning with lag $t-1$, $t-2$, $t-3$, $t-4$ and $t-5$ by using R-Project software can be seen in the table 5.11 below.

Table 5.11 Regression Summary for Monday Morning with Lag $t-1$, $t-2$, $t-3$, $t-4$ and $t-5$

Sig. Code	GA Morning (t-5)		
Lag Variables	GA 5:25:00	GA 8:25:00	GA 8:45:00
GA 5:25:00	*** (+)	* (+)	*** (+)
GA 8:25:00	** (-)	*** (+)	** (-)
GA 8:45:00	***(-)	*** (-)	*** (+)
Mul. R Squared	0.919359	0.703666	0.886994
Sig. Code	GA Morning (t-4)		
Lag Variables	GA 5:25:00	GA 8:25:00	GA 8:45:00
GA 5:25:00	*** (+)		*** (+)
GA 8:25:00	** (-)	*** (+)	** (-)
GA 8:45:00	***(-)	*** (-)	*** (+)
Mul. R Squared	0.927002	0.751828	0.888354
Sig. Code	GA Morning (t-4)		
Lag Variables	GA 5:25:00	GA 8:25:00	GA 8:45:00
GA 5:25:00	*** (+)		* (+)
GA 8:25:00	** (-)	*** (+)	** (-)
GA 8:45:00	***(-)	*** (-)	*** (+)
Mul. R Squared	0.935442	0.801106	0.89483
Sig. Code	GA Morning (t-2)		
Lag Variables	GA 5:25:00	GA 8:25:00	GA 8:45:00
GA 5:25:00	*** (+)		
GA 8:25:00	*** (+)	*** (+)	** (-)
GA 8:45:00	*** (+)	*** (-)	*** (+)
Mul. R Squared	0.946916	0.856238	0.903182
Sig. Code	GA Morning (t-1)		
Lag Variables	GA 5:25:00	GA 8:25:00	GA 8:45:00
GA 5:25:00	*** (+)		
GA 8:25:00	** (-)	*** (+)	* (-)
GA 8:45:00	***(-)	*** (-)	*** (+)
Mul. R Squared	0.964243	0.920564	0.92153

- Very Significant Correlation
- Significant Correlation
- Less Dignificant Correlation
- Multiple R-Squared value > 0.9
- Multiple R-Squared value > 0.8
- Multiple R-Squared value > 0.7
- Multiple R-Squared value > 0.6
- Multiple R-Squared value > 0.5

Based on the table 5.11 above, it can be seen that the multiple R-squared value is increasing from $t-5$ regression to $t-1$ regression. If the regression result of

the previous section also be considered, it can be concluded that the closer the lag variable to the original departure time, the bigger multiple R-squared value resulted. It means that the closer the lag variable to the original departure time, the more representative the regression model to the actual behaviour of the airline ticket price changes.

It also can be seen from the table 5.11 above that the correlation type (positive or negative) of each lag variable to the original departure time did not change based on the different lag period.

The regression result for Thursday morning with lag $t-1$, $t-2$, $t-3$, $t-4$ and $t-5$ by using R-Project software can be seen in the table 5.12 below.

Table 5.12 Regression Summary for Thursday Morning with Lag $t-1$, $t-2$, $t-3$, $t-4$ and $t-5$

Sig. Code	GA Morning (t-5)		
Lag Variables	GA 7:50:00	GA 8:25:00	GA 8:45:00
GA 7:50:00	*** (+)	*** (+)	*** (+)
GA 8:25:00	** (-)	*** (+)	*** (-)
GA 8:45:00	***(-)	*** (-)	*** (+)
Mul. R Squared	0.61223	0.540127	0.861281
Sig. Code	GA Morning (t-4)		
Lag Variables	GA 7:50:00	GA 8:25:00	GA 8:45:00
GA 7:50:00	*** (+)	*** (+)	*** (+)
GA 8:25:00	** (-)	*** (+)	*** (-)
GA 8:45:00	***(-)	*** (-)	*** (+)
Mul. R Squared	0.649508	0.587807	0.870101
Sig. Code	GA Morning (t-4)		
Lag Variables	GA 7:50:00	GA 8:25:00	GA 8:45:00
GA 7:50:00	*** (+)	*** (+)	*** (+)
GA 8:25:00	** (-)	*** (+)	*** (-)
GA 8:45:00	***(-)	*** (-)	*** (+)
Mul. R Squared	0.689982	0.642799	0.88246
Sig. Code	GA Morning (t-2)		
Lag Variables	GA 7:50:00	GA 8:25:00	GA 8:45:00
GA 7:50:00	*** (+)	*** (+)	*** (+)
GA 8:25:00	* (-)	*** (+)	*** (-)
GA 8:45:00	*** (+)	*** (-)	*** (+)
Mul. R Squared	0.734346	0.721306	0.901317
Sig. Code	GA Morning (t-1)		
Lag Variables	GA 7:50:00	GA 8:25:00	GA 8:45:00
GA 7:50:00	*** (+)	*** (+)	*** (+)
GA 8:25:00		*** (+)	*** (-)
GA 8:45:00	*** (+)	*** (-)	*** (+)
Mul. R Squared	0.857611	0.8371	0.932409

- Very Significant Correlation
- Significant Correlation
- Less Dignificant Correlation
- Multiple R-Squared value > 0.9
- Multiple R-Squared value > 0.8
- Multiple R-Squared value > 0.7
- Multiple R-Squared value > 0.6
- Multiple R-Squared value > 0.5

It can be seen in the table 5.12 above that the regression comparison result is the same with the one on Monday morning. The multiple R-squared value

is increasing from $t-5$ regression to $t-1$ regression. If the regression result of the previous section also be considered, it can be concluded that the closer the lag variable to the original departure time, the bigger multiple R-squared value resulted. It means that the closer the lag variable to the original departure time, the more representative the regression model to the actual behaviour of the airline ticket price changes.

It also can be seen from the table 5.12 above that the correlation type (positive or negative) of each lag variable to the original departure time did not change based on the different lag period, same with the regression result on Monday morning.

CHAPTER 6

CONCLUSIONS AND RECOMMENDATIONS

In this chapter, the conclusions of the research is presented and the recommendations from the research is offered. The recommendations offered for the further research with the similar topic with this research.

6.1 Conclusions

The conclusions of this research are as follows:

1. There are several price changing pattern that is followed by the Garuda Indonesia Airlines by days of purchase for departure date December 14th, 2015 and December 17th, 2015. First, the starting price is tend to be stable since period 600 (day 30), then the price is starting to increase from around period 300 (day 15) to the scheduled departure. Second, the starting price is tend to be stable since period 600 (day 30), then the price is rapidly increasing near the scheduled departure, around period 60 (day 3). Third, the starting price is stable since period 600 (day 30), then it is increasing in the middle period and decreasing lower than the starting price and rapidly increasing near the scheduled departure. Fifth, the price changes is tend to be stable all the time since period 600 until the scheduled departure.
2. In a flight time group, there are lag variables that gave influence to the original departure time prices, especially peak departure time, and there are also lag variables that did not influence to the original departure time prices. Every lag variable gave influence to at least one original departure time prices within the same flight time group. The flight group that has tightest price competition is afternoon flight group. Thursday flights have more tight price competition than Monday flights. For the regression between flight groups, on Monday, all lag variable gave influence to all the original departure time. Meanwhile, on Thursday, all

lag variable gave influence to all the original departure time, except lag 8:45.

3. The difference between changing price behaviour of peak departure time and non-peak departure time is the non-peak departure time price always changes correspond to the peak departure time price changes. On the other side, most of the peak departure time price changes independently without really considering the non-peak departure time price changes. It means that the peak departure times has thighter price competition than non-peak departure times.

6.2 Recommendations

The recommendations that is offered for the next research with the similar topic of this research are:

1. The data collected should be more than two departure dates so the data can be used to predict the future price changes. By having more than two departure dates data, e.g. a week or a month of departure date, the data can be used for research with the ojective to predict the future price changes or pricing patterns.
2. The data collected can also be expanded for two airport bases. For example, the data collected are flight from SUB to CGK and CGK to SUB so that the cause and effect correlation of both flights can be analyzed.
3. The result of this research can also be the input of the next research from the company's point of view because this research is from the customer's point of view. From the multiple regression result in this research, we can research why there is lag variable that did not influence the original departure time prices from the company's point of view.

REFERENCES

- Asosiasi Penyelenggara Jasa Internet Indonesia 2015, *Profil Pengguna Internet Indonesia 2014*, Asosiasi Penyelenggara Jasa Internet Indonesia, Jakarta.
- Badan Pusat Statistik 2016, Badan Pusat Statistik, accessed 13 April 2016, <<https://www.bps.go.id/>>
- Boeing 2016, *Airline Strategies and Business Models*, Boeing, United States, accessed 19 April 2016, <<http://www.boeing.com/commercial/market/long-term-market/airline-strategies-and-business-models/>>
- Cento, A 2009, *The Airline Industry Challenges in the 21st Century*, Physica-Verlag Haidelberg, Germany.
- Direktorat Jenderal Perhubungan Udara 2016, *Nilai Keamanan Penerbangan Indonesia Mencapai 93.6*, Media Release, Kementerian Perhubungan Republik Indonesia, accessed 13 April 2016, <<http://hubud.dephub.go.id/?id/news/detail/2845>>
- Etzioni, O, Tuchinda, R, Knoblock, CA, Yates, A 2003, 'To Buy or Not to Buy: Mining Airfare Data to Minimize Ticket Purchase Price', *KDD '03 Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 119-128.
- Groebner, DF, Shannon, PW, Fry, PC, Smith, KD 2010, *Business Statistics: A Decision-Making Approach*, Prentice Hall, New Jersey.
- InfoPenerbangan 2016, 'Tarif Penerbangan Kelas Ekonomi Turun 5 Persen', *InfoPenerbangan*, 13 April, accessed 13 April 2016, <<http://infopenerbangan.com/tarif-penerbangan-kelas-ekonomi-turun-5-persen/>>
- International Civil Aviation Organization 2004, *Manual on the Regulation of International Air Transport*, 2nd edn, International Civil Aviation Organization.
- Mendenhall, W & Sinich, T 2011, *A Second Course in Statistics: Regression Analysis*, 7th edn, Prentice Hall, United States.

Pels, Eric & Rietveld, Piet 2004, 'Airline Pricing Behaviour in The London Paris Market', *Journal of Air Transport Management*10, pp. 279-283.

Philips, R 2005, *Pricing and Revenue Optimization*, Stanford University Press, Stanford.

Traveloka 2015, Traveloka, accessed 14 November 2015, <<http://www.traveloka.com/>>

BIOGRAPHY



Author was born in Surabaya, September 30th, 1995. Author's name is Adelia Stephanie Zerlinda. Author has finished her study at SD Negeri Pucang 3 Sidoarjo on 2007, finished her study at SMP Negeri 1 Sidoarjo on 2010, finished her study at SMA Negeri 1 Sidoarjo on 2012, and continued her study at Industrial Engineering Department of Institut Teknologi Sepuluh Nopember (ITS) Surabaya. She joined some student activities during her college study, including Industrial Engineering Youth Club (IEYC), joined as committee of Industrial Engineering Games (IE Games), Industrial Challenge (INCHALL), fresh student orientation as IC, and some other student activities. Author can be reached through email to adelia.zerlinda@gmail.com.