

Customer Satisfaction In The Airline Industry

*SouthEast Airlines*

Data Analysis Report

* Hope Puzzanghera
* Shwet Jain
* Osama Junaid
* Pat Carlin

CONTENTS

1. **INTRODUCTION** *[pg. 3]*
   1. Introduction
   2. Background
   3. Context
   4. Scope
2. **BUSINESS QUESTIONS ADDRESSED** *[pg. 5]*
3. **DATA EXPLORATION** *[pg. 6]*
   1. Acquisition
   2. Cleansing
   3. Transformation
   4. Munging
4. **DESCRIPTIVE STATISTICS AND VISUALIZATION** *[pg. 9]*
5. **USE OF MODELING TECHNIQUES** *[pg. 26]*
6. **ACTIONABLE INSIGHTS** *[pg.*
7. **CONCLUSION**
8. **APPENDIX**

**Chapter 1 -** INTRODUCTION

Introduction:

In the last 25 years, the aviation industry has been growing rapidly. In addition to its technological developments, the growing of airline industry is due to its role as supporting the world trade, international investment, and tourism activities. Because of these roles, it is often said that the aviation industry is the center of globalization for other industries. The growing of the airline industry provides opportunities as well as challenges to the business entities in this industry. The opportunities arise due to the increasing demand for the airline services. While the challenges arise not only because of the high level of competition between the airlines, but also due to growing consumer demands for better service. Growing of this industry has led to the governments around the world to deregulate the industry by enabling companies to form private airlines. Further, the stringent rules on safety flight is currently applied in almost all countries. These conditions cause all players in the airline industry require to continuously innovate in terms of both services and technology used to deliver services and better safety to the consumers. This report provides analysis of SouthEast Airlines on different factors that influence consumer satisfaction which can then be used to drive customers satisfaction and thus provide a competitive edge.

Background

The Dataset contains about 129889 responses (rows) from airline customers survey throughout 3 months, and contains data from 14 airlines. The Dataset has 28 columns, which consist of data obtained from surveys submitted by its airline customers. The columns broadly focus on several categories, including customer’s gender, age, number of flights, shopping amount at the airport, type of travel, etc.

Objective

We are working as a consulting company for Southeast Airlines. The focus of our survey analysis project is to provide some useful insights to Southeast Airlines so that they can improve customer satisfaction which will drive business and in turn profits.



Context:

The Airlines data is a dataset collected from information about the customers taking various airline flights and giving their satisfaction ratings about the overall experience they had with the flight.

Scope :

The study will benefit the airline company ‘South Asia’ in evaluating their performance. It could have a competitive advantage if they could brand themselves appropriately. This study will help Airline ‘South Asia’ and airport authority in developing an effective service. The results of the study will be significant to the airline ‘South Asia’ in which the majority passengers belong to and also it enables other airlines to offer better service. It will serve as a guide as how passengers prefer airline and the satisfaction level of passengers will guide the airline ‘SouthAsia’’ for their improvement.

**Chapter 2 -**  BUSINESS QUESTIONS ADDRESSED

Every research or analysis needs to have a business reason.

The two key reasons are:

**Improved customer service** such as those below. There can be more than one reason :

* Increasing take up of services among key groups to achieve targets
* Making it easier to access services
* Giving a better service
* Giving a service targeted to individual needs
* Giving access to a broader range of services

**Improving efficiency** by one or more of:

* Increasing take up among key groups to increase income
* Increasing early take up and reducing more expensive interventions later
* Improving processes to streamline services and reduce costs (including one touch contact)
* Switching customers to more cost efficient channels

1. Does Gender play an important role wrt. Satisfaction?
2. What are the important attributes that drive Satisfaction ?
3. Does Origin City and Destination City affect Satisfaction?
4. Does No.of Other Flights taken by the customer affect Satisfaction?
5. Discuss the correlation between Distance Covered and Arrival Delay?
6. Does the relationship between the Origin City and the Destination City together have something to do with Customer Satisfaction.
7. Does the relationship between Price Sensitivity, Type Of Travel affect the corresponding Satisfaction?
8. Relationship between the Airline Status and Satisfaction.

**Chapter 3 -**  DATA EXPLORATION

The process of amending or removing data in a database that is incorrect, incomplete, improperly formatted, or duplicated.

Data Acquisition

1) Approximately 130,000 survey responses

2) 25 fields in the Survey

3) Some entries in the data-set are blank (NA)

File Used: “**Satisfaction Survey.csv**”

Pre-Cleansing

Incorrect, Incomplete, Improperly formatted and duplicated



Also, handling the NA’ s in the data-set.

Cleansing, Transformation & Munging

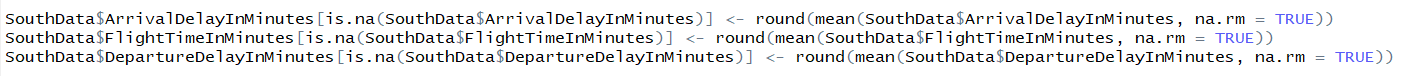
1. Dealing with Column Names:



1. Dealing with ‘ . ’ s in the Column Names:



1. Dealing with NA’s: NA's will be replaced by means of their respective columns



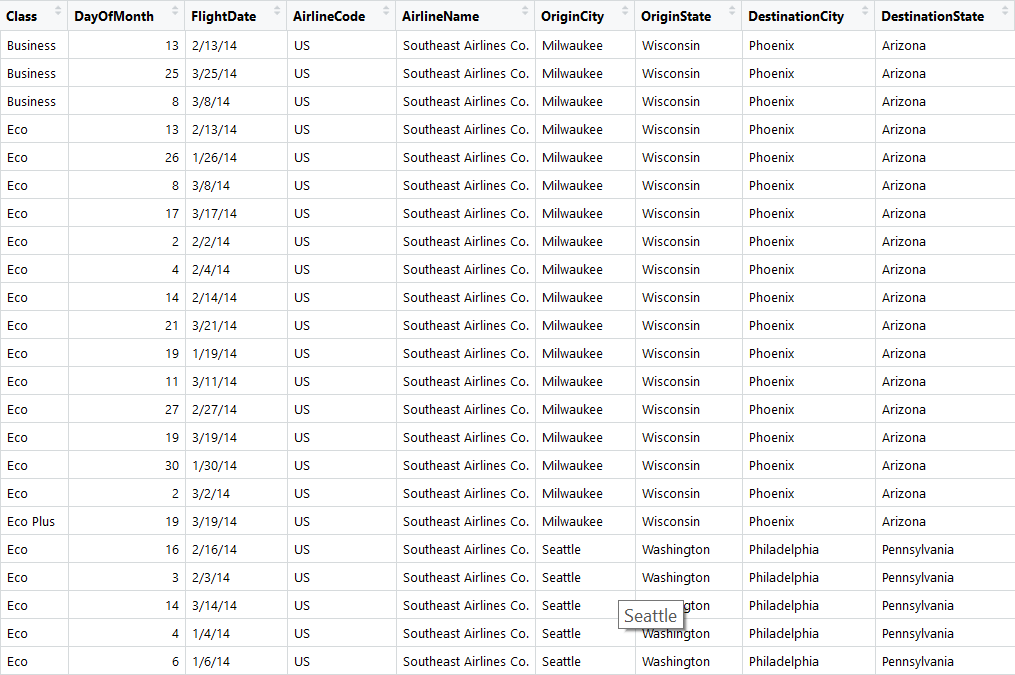
1. Dealing with City, State in the Same Column:

****

1. Extracting Data (SouthEast Airlines)

****

Data Set After Performing Cleansing, Transformation & Munging Operations

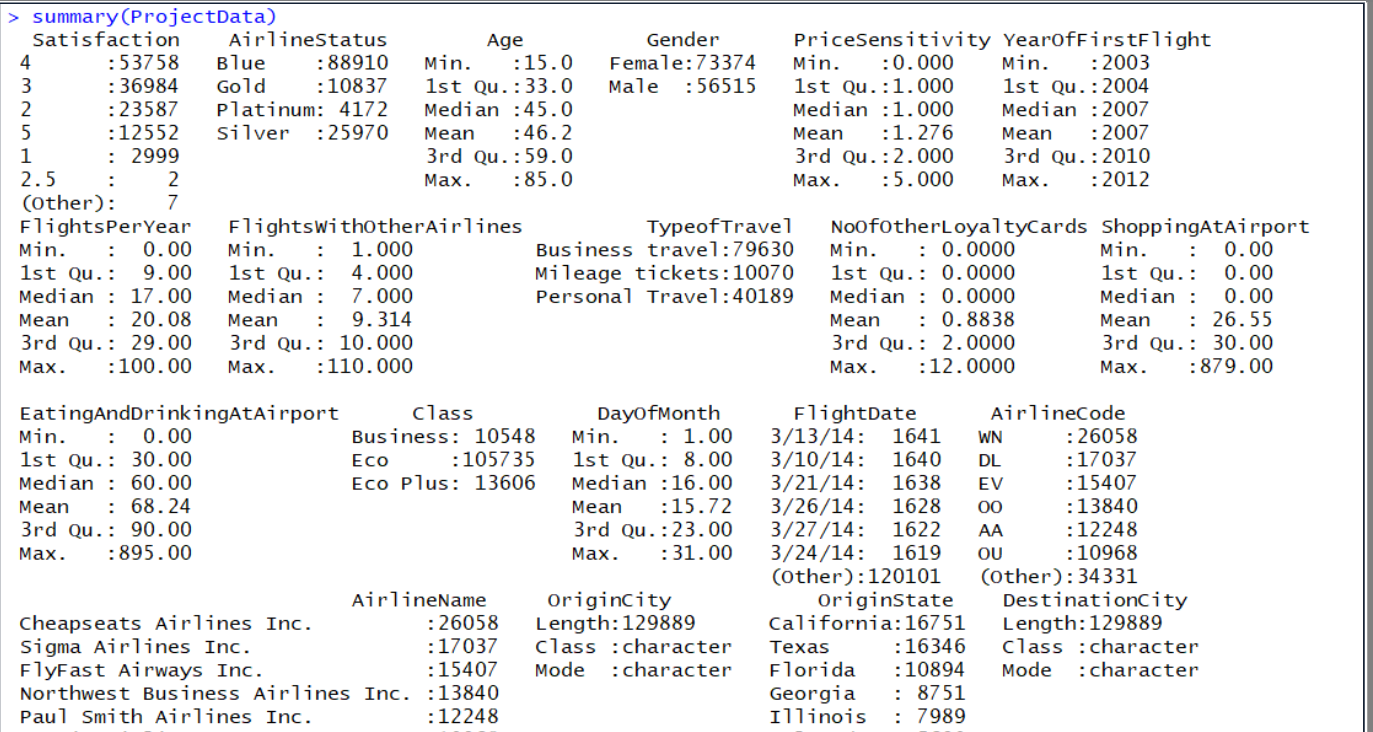


**Chapter 4 -**  DESCRIPTIVE STATISTICS AND VISUALIZATION

Descriptive Statistics

* Used to describe the basic features of the data in a study.
* Provide simple summaries about the sample and the measures.

Screenshots: Using the summary(), range(), and sd() functions we get

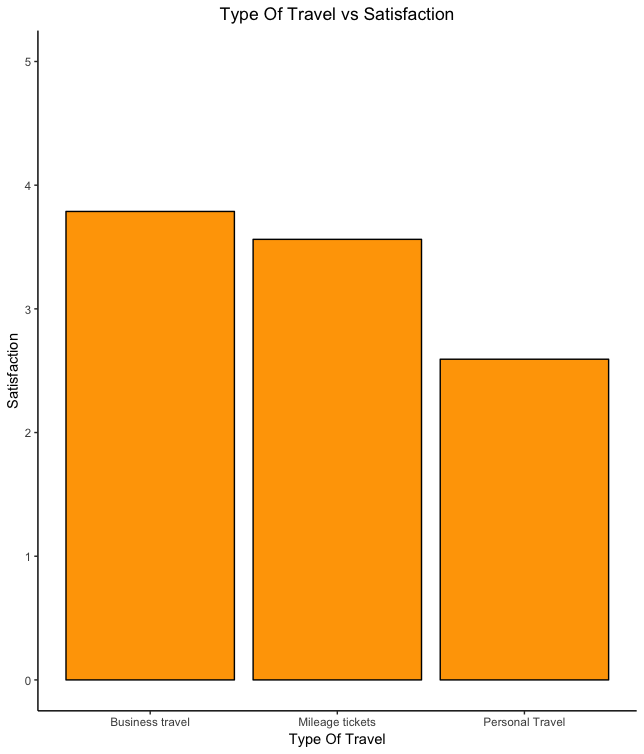


This helped us get a basic overview of the entire data-set

Visualization

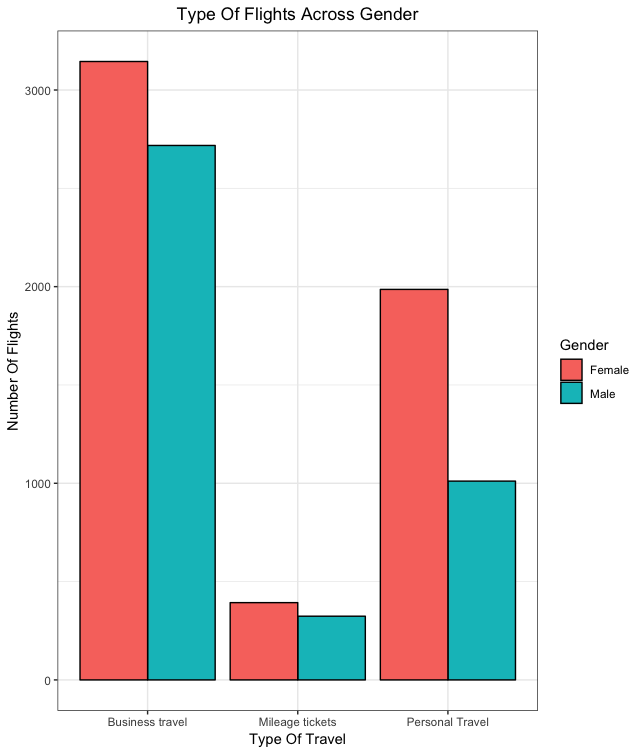
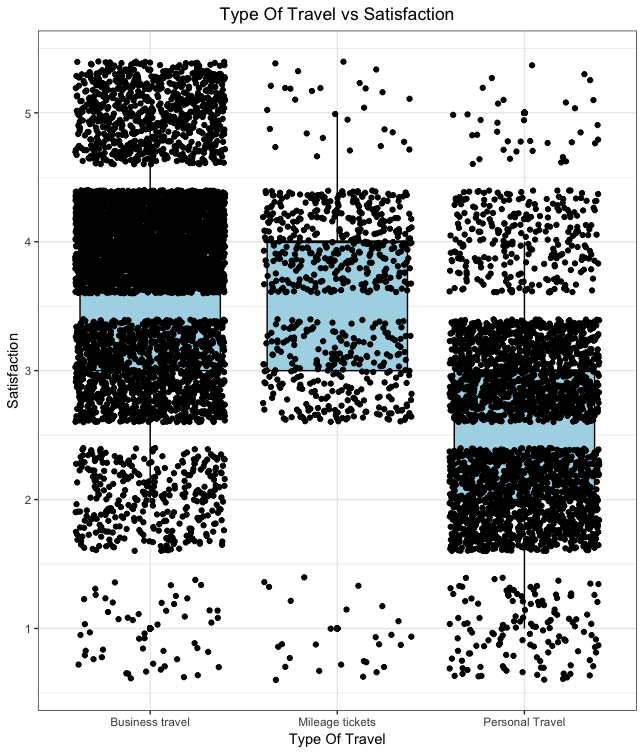
Visualization is one of the most important tasks in data analysis. Being able to visually represent an issue and a solution within a company will allow the technical analyst to touch all employees of a company. Aimed to appeal to everyone from upper management to the most technical employee at Southeast Airlines Co., well developed and thought provoking insights into the happenings and issues that have arisen. Our goal is to visualize our interpretation of the business questions previously discussed with sound graphical analysis.

1. *Viewing the Overall Satisfaction within separate Flights and Ages [Graph 1.1]*

****

This visual representation graphs the different types of travel compared to the averaged overall customer satisfaction. Notice that Business travel has the highest satisfaction with around 3.8, second is the mileage tickets at 3.5, and last would be the personal travel at well below 3. This brings up an apparent question when we decide on where to go next with our analysis; why, out of all the types of travel, is the personal (leisure) travel have the lowest satisfaction?

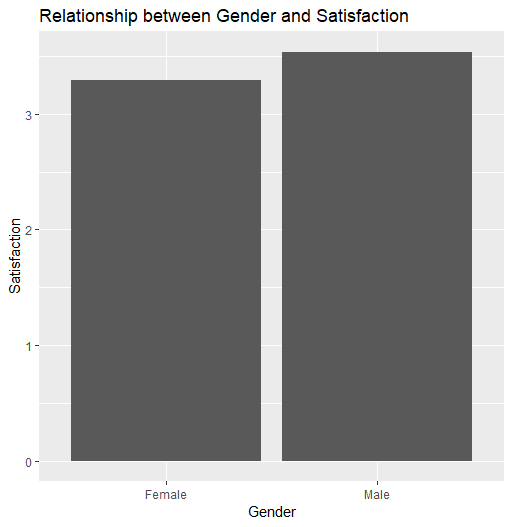
1. *Finding the frequencies of the Types of Travel [Graph 1.2:1.3]*



Furthering into our analysis of the overall customer satisfaction within the different types of travel leads us to the boxplot in ‘Type of Travel vs Satisfaction’ boxplot, by utilizing the geom\_jitter() function within our plot, we can see how the frequencies within each type are dispersed. Notice that within ‘Business travel’, having the highest frequency, tends to be skewed above the 3.5 median. Having a majority of the frequencies above the median in this case will ultimately lead us to a higher mean. transition to the travel type with the lowest mean in graph 1.1, ‘Personal Travel’ had a fairy even distribution around the median of 2.5, giving us validation of why we see the low average. With the least amount of visual frequencies ‘Mileage Tickets’ showed a fairly true distribution, around the median of 3.5.

Graph 1.3 is a visual comparing the frequencies of each Type of Travel, with gender taken into account for a deeper understanding of the data. We see that Graph 1.3 is clearly a better visual of how the frequencies of each travel type are allocated, again validating the reasoning behind Graph 1.1. But what is the reasoning for the ‘Personal Travel’, which has the second highest frequency, have the lowest average satisfaction? Notice the gender frequency for females within ‘Personal travel’ is double the frequency of males.

1. *Gender effects on Overall Satisfaction [Graph 1.4:1.5]*

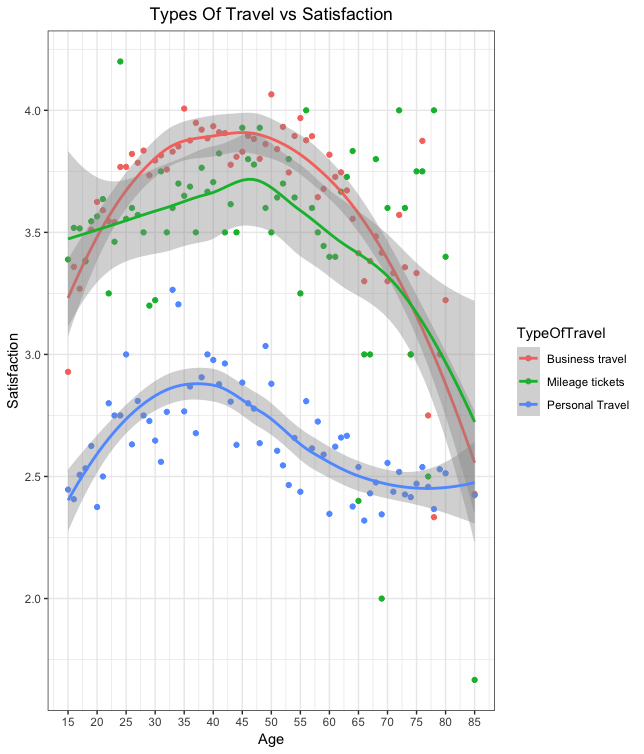
****

While discussing the impacts of types of travel on overall satisfaction, we notices that there was an unequal dispersion within the frequencies of male and female in personal travel. To take a step back and understand the reasoning behind the disparity, let’s take a look at overall averaged satisfactions of males and females across the age groups given.

Two interesting values arise from what is visible within Graph 1.4. Up until the age of around 21, men and women seem to give back the same average satisfaction. From there up until about age 55 do the two lines meet again. Why would that be, we know that women on average have a higher frequency across all categories of travel type. Would it be a causal effect of traveling more? Or an exogenous variable acting on the subset?

Another interesting point to take away from Graph 1.4 would be the downward trend from age 60-70. We notice that the satisfaction drops lower than at any other point in the graph within this group. Does retirement play a role in the the overall flight satisfaction. If does the type of travel change with the age?

1. *Age Impacts across Satisfaction and Type of Travel [Graph 1.6]*

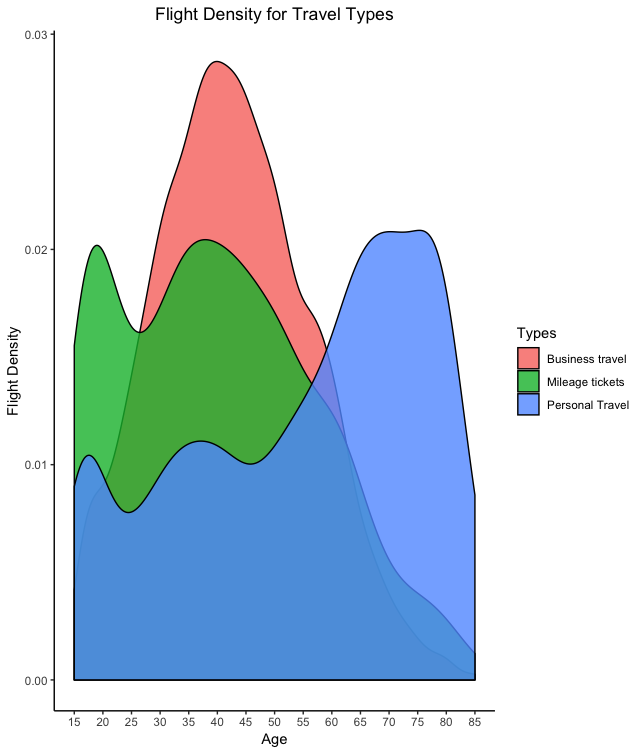
****

What we are attempting to view within this graph is how each type of travel changes across surveyed ages. For ‘Types of Travel vs Satisfaction’ we see a scatter plot, coupled with a smoothed average across ages. Visually we notice that as age increases, the overall satisfaction of the lines increases across all types of travel. But once we get passed 50 years of age, Business travel and Mileage tickets both fall exponentially. But the average satisfaction of Personal Travel seems to remain the, around 2.5 at age 85, while Business Travel and Mileage Tickets drop a full point of satisfaction across the same span.

But the smoothed line of Personal travel switched concavity at around 60 years of age. Just as we noticed before, something happens around age of 55-60 that spikes our attention. Graph 1.6 shows that something within Personal Travel will change as age increases.

Giving our insight into the data set, you need to understand the mindset of each flight, business travel is usually stipend while earning points towards other flights. On a business trip, being more of a duty than leisurely activity, amenities are not as big a factor as they would be on a personal travel where someone is paying for their flight.

1. *Visualizing the Density of Type Of Travel across Ages [Graph 1.7]*

****

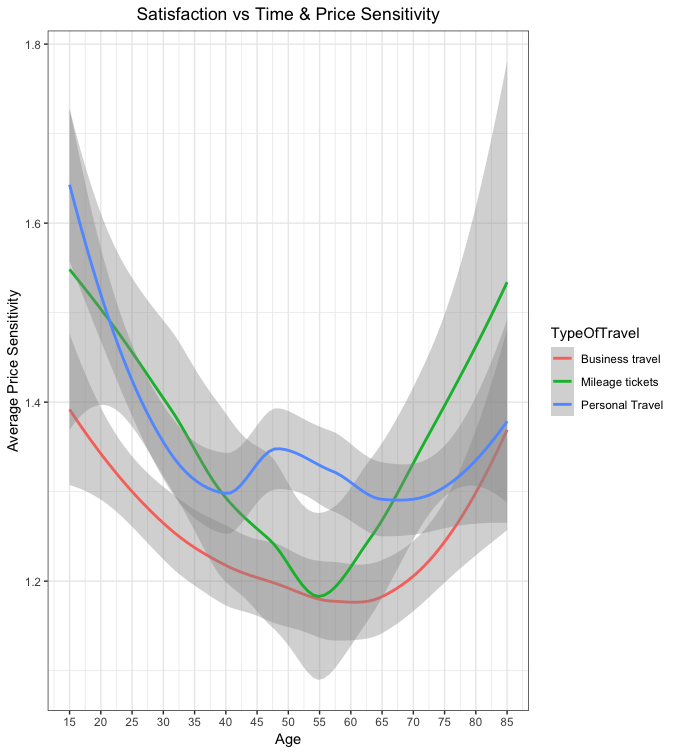
We are left wondering about how age impacted the specific types of travel, we needed to visualize how each type of travel frequency changes with age. Graph 1.7 depicts each type density, across ages 15 through 85.

We can notice the business travel has a parabolic path with a maximum value around the age of 45. Inferring that business travel is paid for by the employees company we will not see any fluctuation within the frequency of business travel. While travelling for business, employees usually earn mileage points which can be put towards mileage tickets.

We see that mileage tickets has an initial peak in between age 15 to 25. How I would translate this initial peak, would be that mileage tickets earned from a parents business travel are being used for their children and family. Families that use mileage tickets earned from business travel for their vacations (personal travel) would explain the spike in personal travel between age 15 to 25.

Based of of the visual in the “Flight Density for Travel Types” we can notice that at age 45 people decrease both Business travel and Mileage tickets, while Personal travel skyrockets until the age of 80 then begins to fall. As people begin to move out of the workforce they do not receive the same benefits from Business travel, therefore the type of travel turns into personal travel as one nears retirement.

1. *Price Sensitivity Graph against Age [Graph 1.8:1.9]*

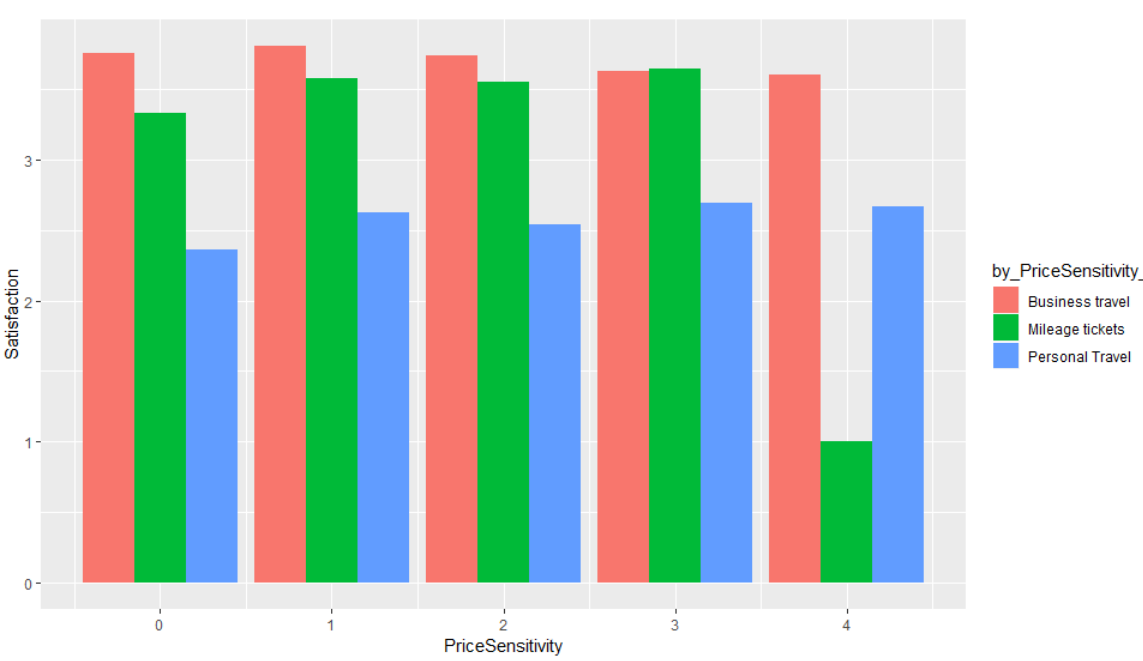
****

What we noticed in Graph 1.8 was how frequency of types of travel changed throughout the age scale given in the survey. Graphing the averaged price sensitivity across ages will show us how and why people decide on making their decisions for travel. Again the age 55 becomes an interesting value to look at when discussing the graph because it is the critical value of all the types of travel. At age 55, Business travel and mileage tickets reach their lowest price sensitivity. As viewed in graph 1.8, from age 55 to 65 is when we see the frequency of personal travel surpass that of the business travel and mileage tickets.

If we take a closer look at the personal travel. we notice that from age 15 to 40 the price sensitivity drops .3 of a sensitivity point, but from age 65 to 85 it remains unaffected. We can see this as after retirement people remain unaffected by the price of the tickets, unlike earlier ‘leisure’ travel.

This Visualization helps us understand the relationship between the Price Sensitivity and it’s respective mean Satisfaction. As we can see from the plot, Customer Satisfaction decreases as the Price Sensitivity increases.

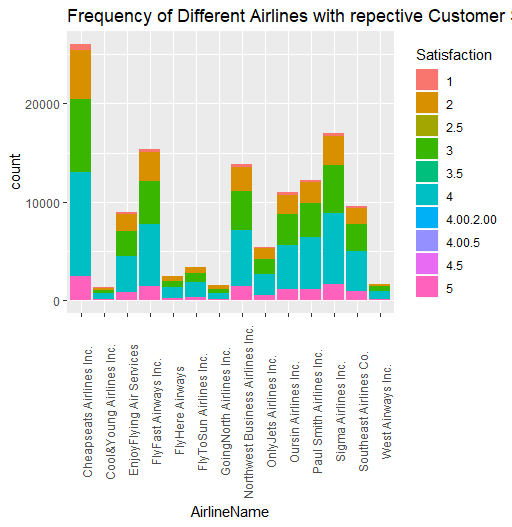
1. *Satisfaction and Price Sensitivity of Types of Travel [Graph 2.0]*

****

This Visualization describes the relationship between PriceSensitivity, TypeOfTravel and corresponding Satisfaction.

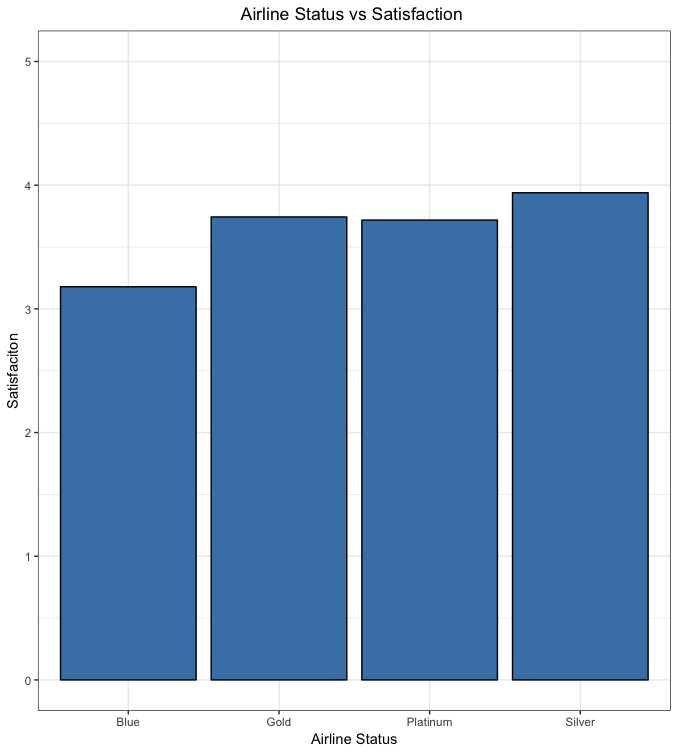
As we can see from the Plot, Price Sensitivity doesn’t really affect the Business Travel but in someway or another does affect the other Types of Travel.

1. *Satisfaction Overview of all Airlines [Graph 2.1]*

****

This Visualization helps us understand how Customer Satisfaction is distributed over all the other Airlines in the data-set along with the frequency.

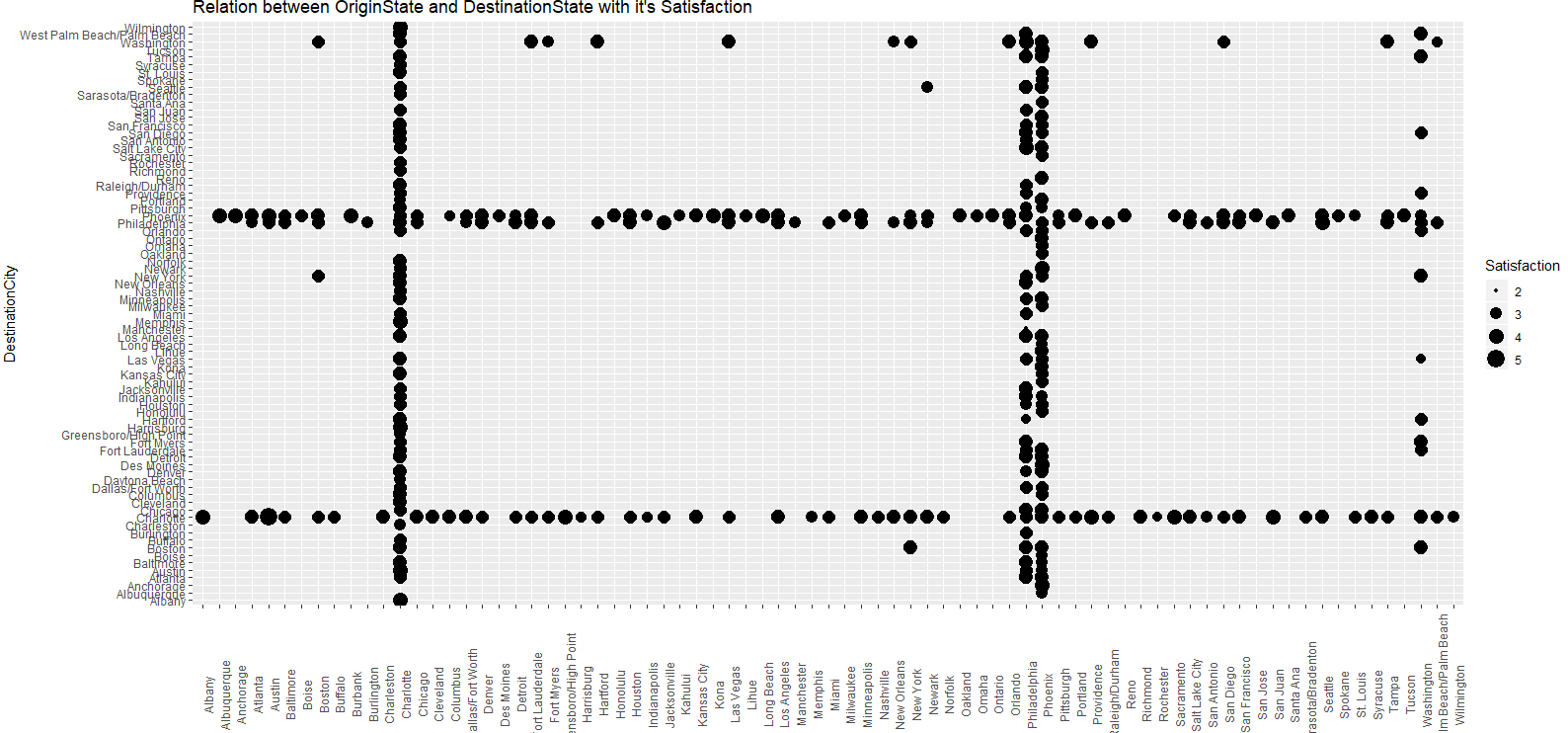
1. *Airline Status and Averaged Satisfaction [Graph 2.2]*

****

This Visualization helps us understand the relationship between the Airline Status and Satisfaction.

As we can see from the plot, Airline\_Status=Blue has a lower customer Satisfaction when compared to other Airline\_Status.

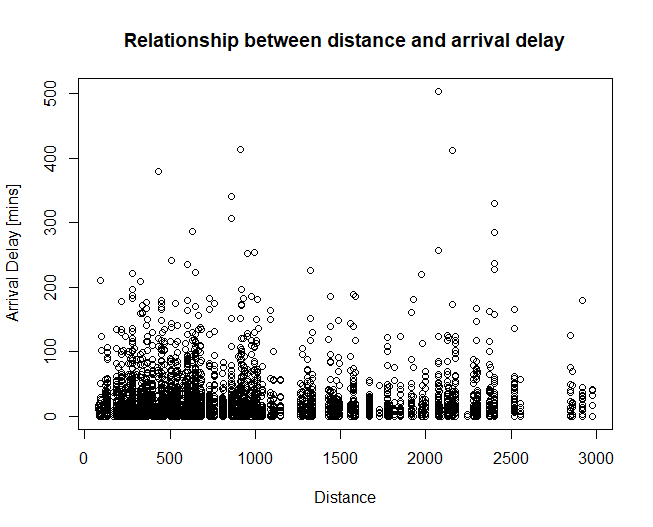
1. *Origin and Destination State with Satisfaction [Graph 2.3]*

****

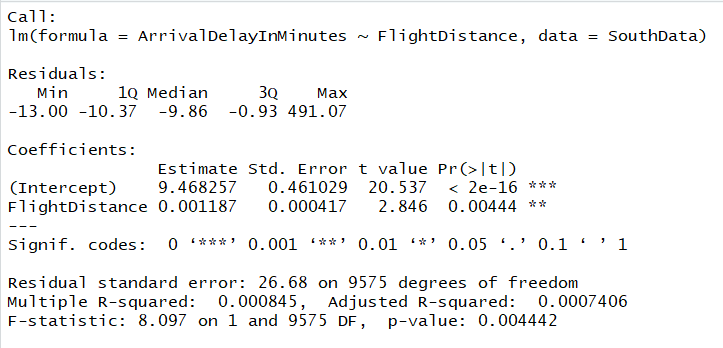
This visualization shows the relationship between the Origin City and the Destination City along with it’s mean Satisfaction.

This Visualization could be helpful for quick action. For example, as we can see “Washington” has a lower mean Satisfaction which is an Origin City. We could dive deep into the problem and necessary actions could be taken to solve the problem.

1. *Distance against Arrival Delay [Graph 2.4]*

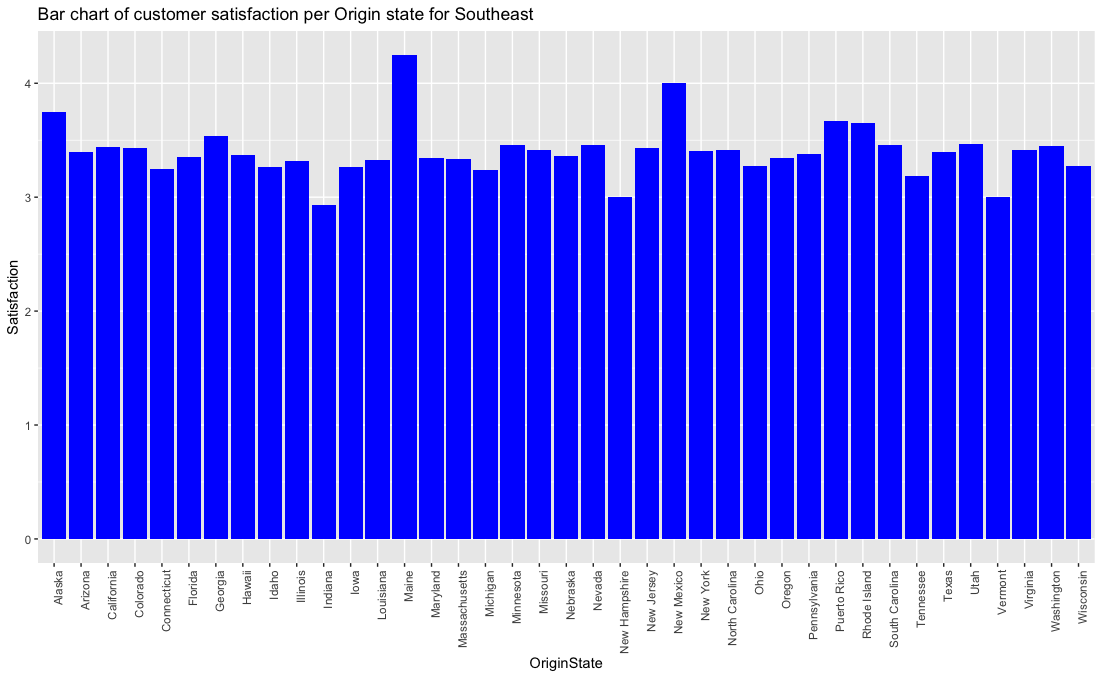
****

This visualization was created to find out the correlation between the Distance Travelled and the Arrival Delay.

****

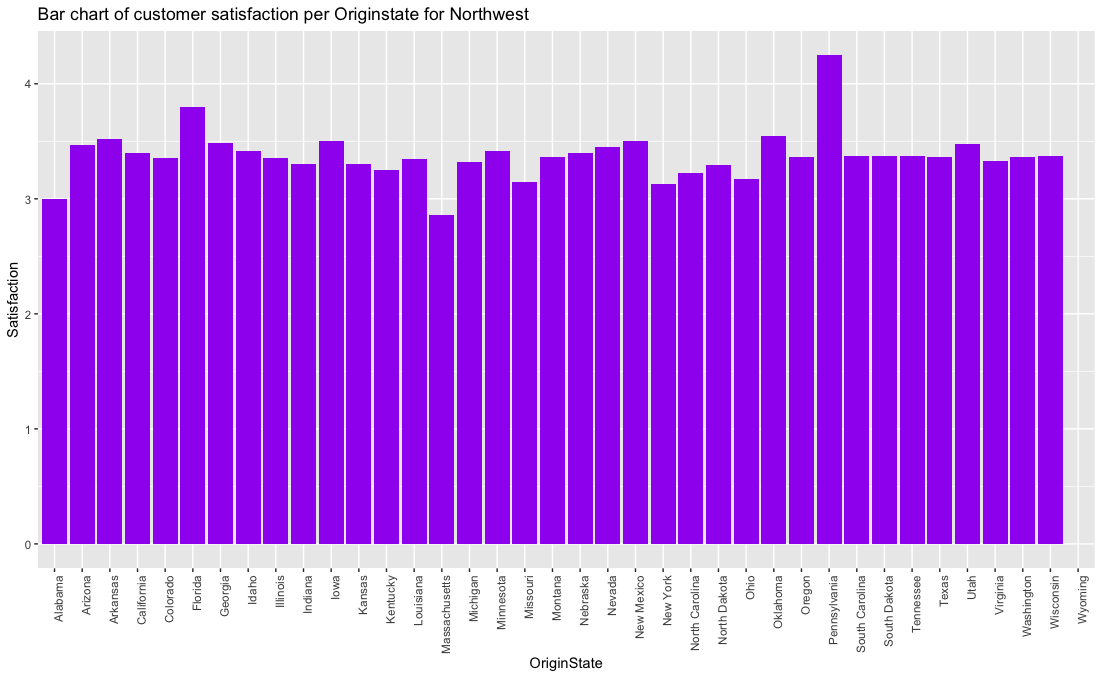
We carried out linear modelling tests for the same. And there was no significant relationship between the two variables.

1. *Customer Satisfaction against Origin State [Graph 2.5]*

****

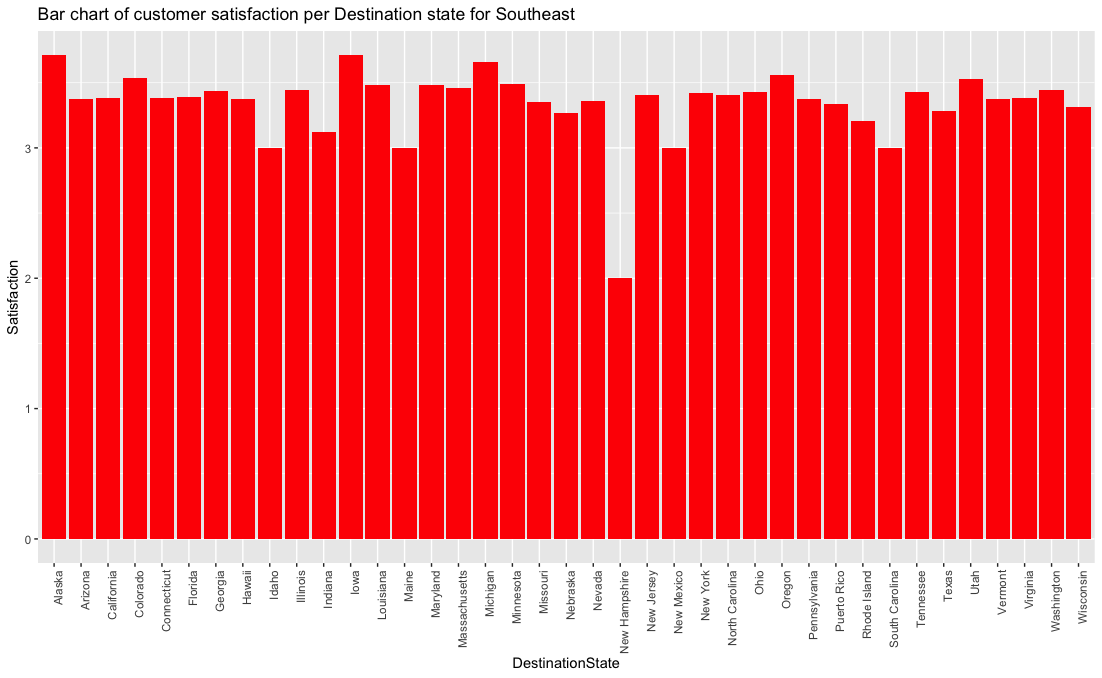
As you can see above, I constructed a Bar chart for the overall customer satisfaction per state for Origin of state. You can see that our Airline, Southeast Airlines Co flies out of 39 states instead of 50 or 51, which means that it doesn't fly out from the rest of the missing states which can be a problem because they are not maximizing profit. If you pay attention to the chart, Maine had the highest overall customer satisfaction (4.250000) and Indiana had the lowest (2.933333) and Vermont and New hampshire being the second lowest tied at (3.0000).

1. *Customer Satisfaction against Origin State for Northwest Business Airlines Inc. [Graph 2.6]*



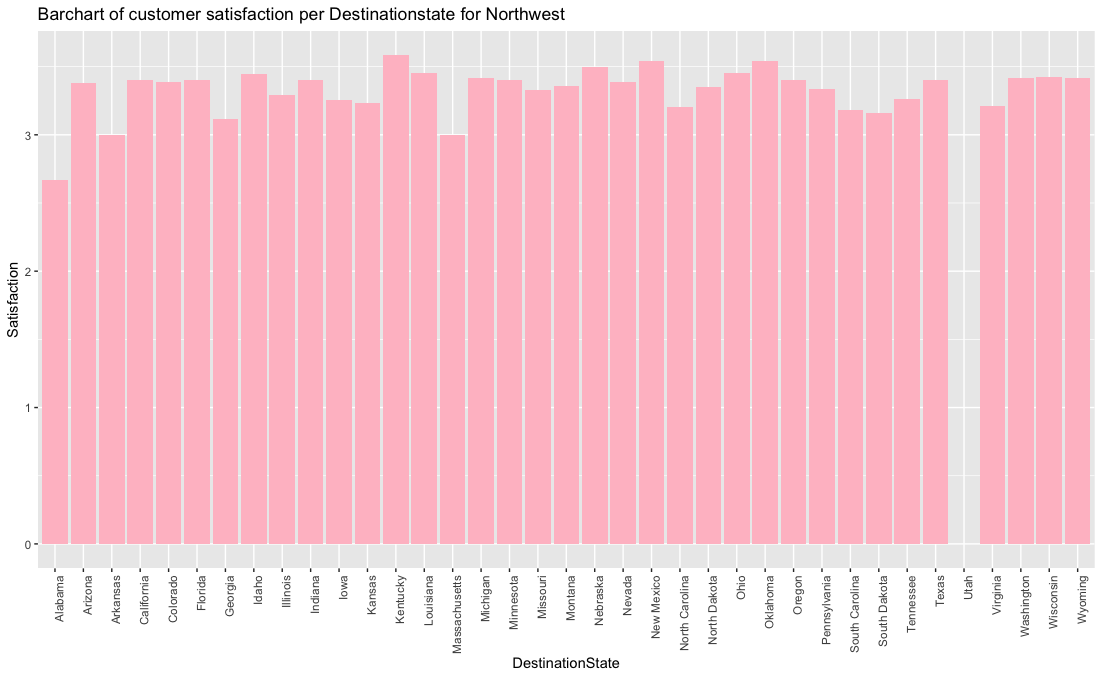
Constructed a different barchart for *Northwest Business Airlines Inc.* to compare it to *Southeast Airlines Co*. You can notice the difference and see that this airline only flies out of 38 states which is less than our airline. on average, there's a higher satisfaction rate compared to Southeast. Pennsylvania has the highest overall satisfaction rating.

1. *Customer Satisfaction against Destination State [Graph 2.7]*



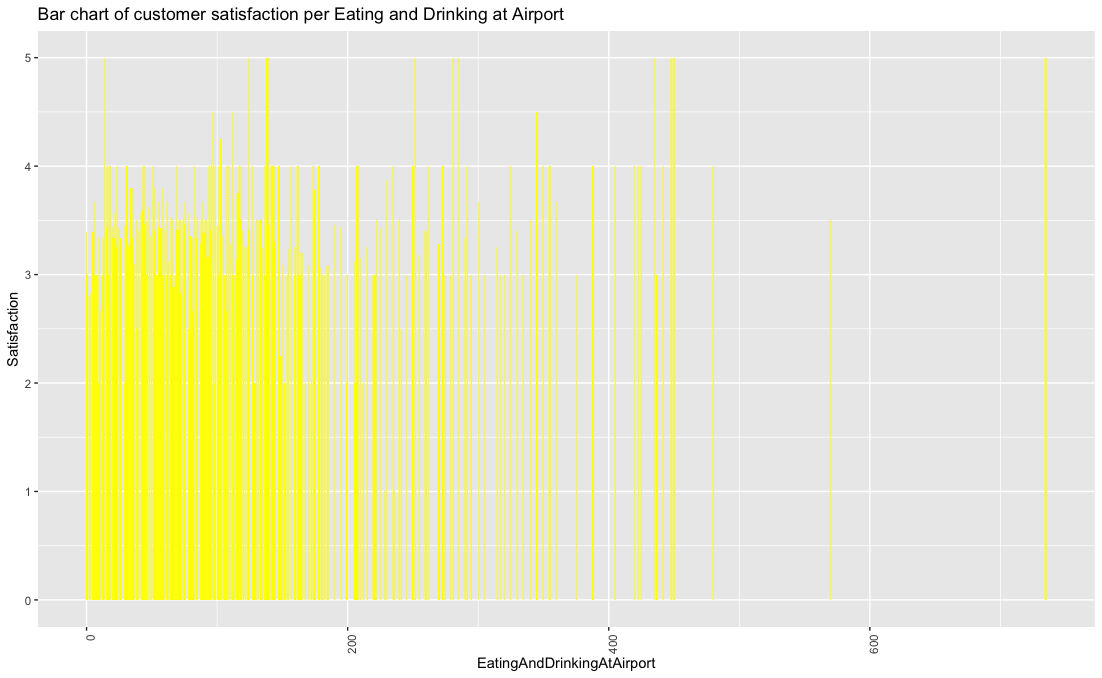
I constructed another Bar chart for the overall customer satisfaction per state for Destination of state. You can see again that Southeast Airlines Co. only flies to 39 states instead of 50 or 51, which means that it doesn't fly to the rest of the missing states which can be another problem because they are not maximizing profit for all the State. If you pay attention to the chart, Alaska has the highest overall customer satisfaction (3.714286) and New hampshire has the lowest (2.000000). South Carolina, New Mexico, Maine, and Idaho are the second lowest tied at (3.0000). From these 2 bar charts you can see that People like flying out of Maine but don’t like flying to Maine.

1. *Customer Satisfaction against Destination State for Northwest Business Inc. [Graph 2.8]*

****

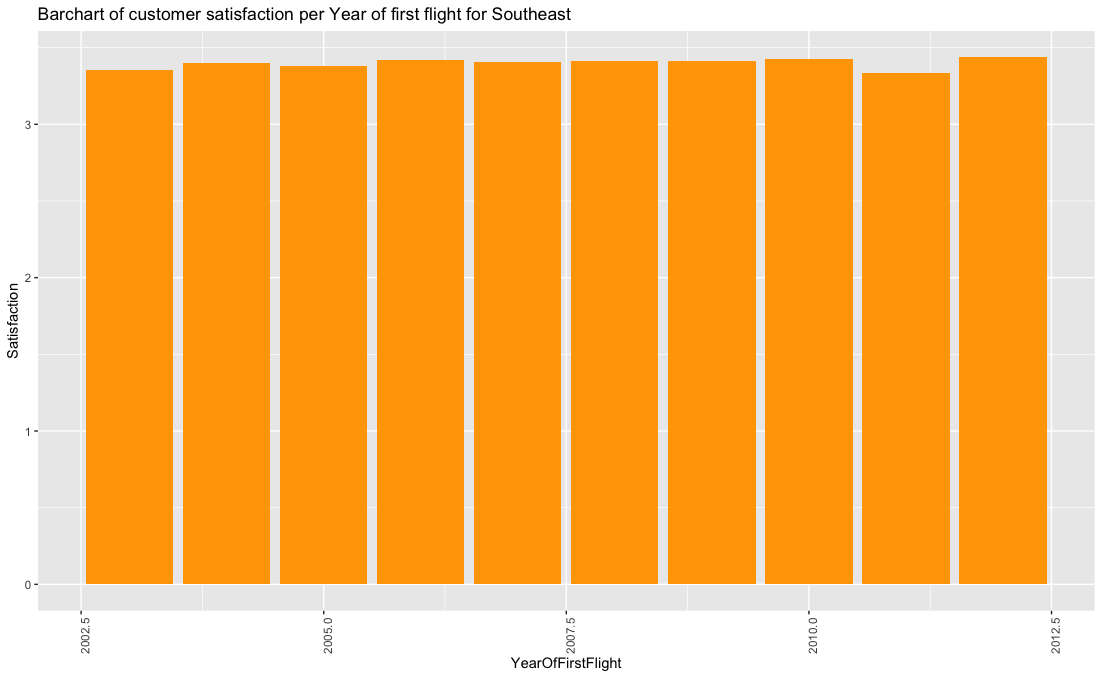
Overall customer satisfaction per state for Destination of state for Northwest Business Airlines Inc. you can see that this airlines only flies to 37 states which less than Southeast Airlines. For Utah there's no Satisfaction. Alabama has the lowest satisfaction rate.

1. *Satisfaction against Eating and Drinking at the Airport [Graph 2.9]*

****

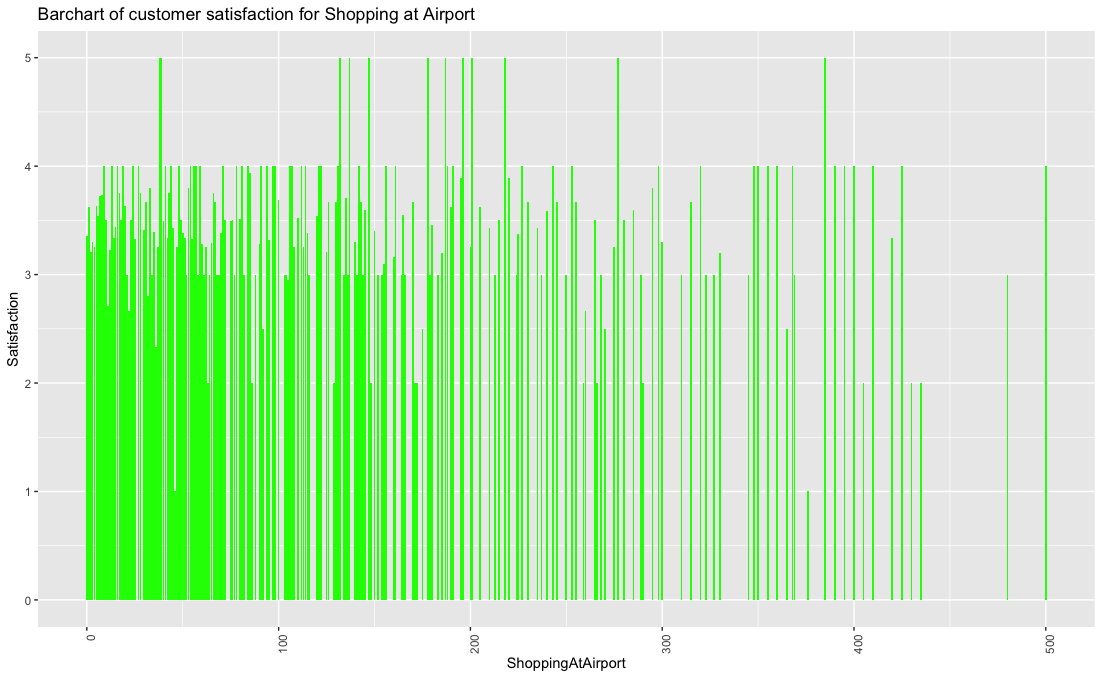
This barchart shows the overall customer satisfaction of eating and drinking at a Airport. The X-axis represents the amount of money people spent. There’s a outlier at 735 with a satisfaction of (5.000000) which can mean that someone with a lot of money spent a lot and enjoyed the food at the Airport which shows great customer service. You can say that the data is right skewed which means the mean is greater than the median.

1. *Satisfaction and Year of First Flight [Graph 3.0]*

****

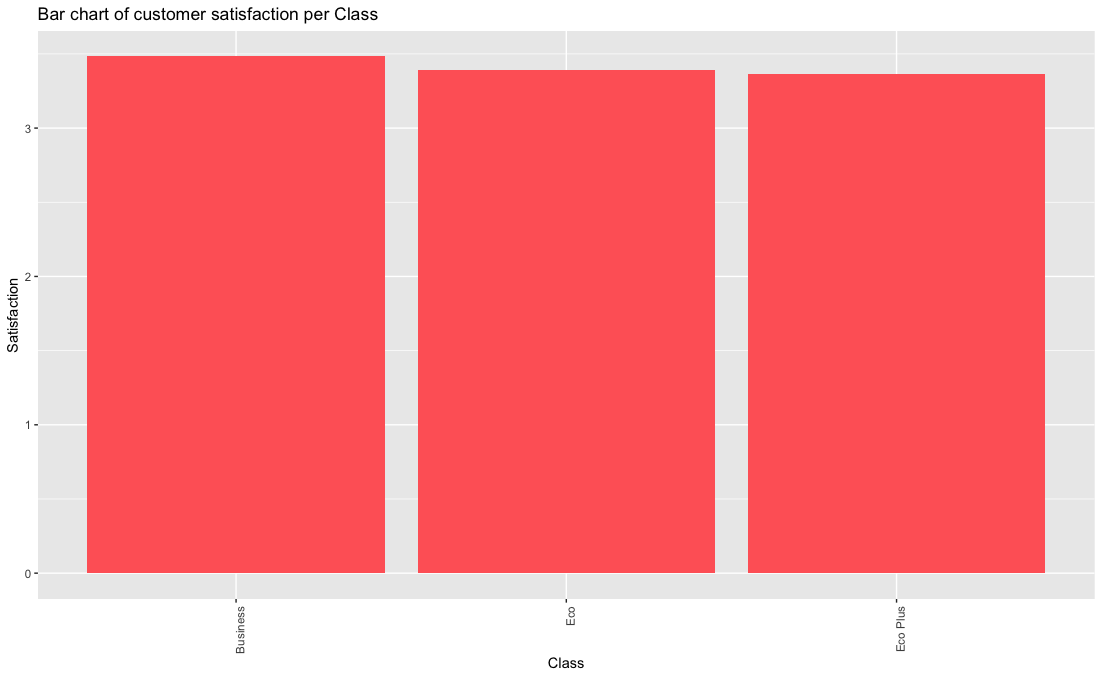
This Barchart was constructed to find out the overall customer Satisfaction per Year of first flight for Southeast. It proved that the most current “first flight” would have the highest satisfaction because it is new and understands what the customers want due to past research.

1. *Average Satisfaction Against Shopping at the Airport [Graph 3.1]*



Overall customer satisfaction for shopping at the Airport. The X-axis represents the money spent. A lot of people were in the range of $0 - $150. This shows that there are people that enjoy shopping at the Airport and are willing to spend money.

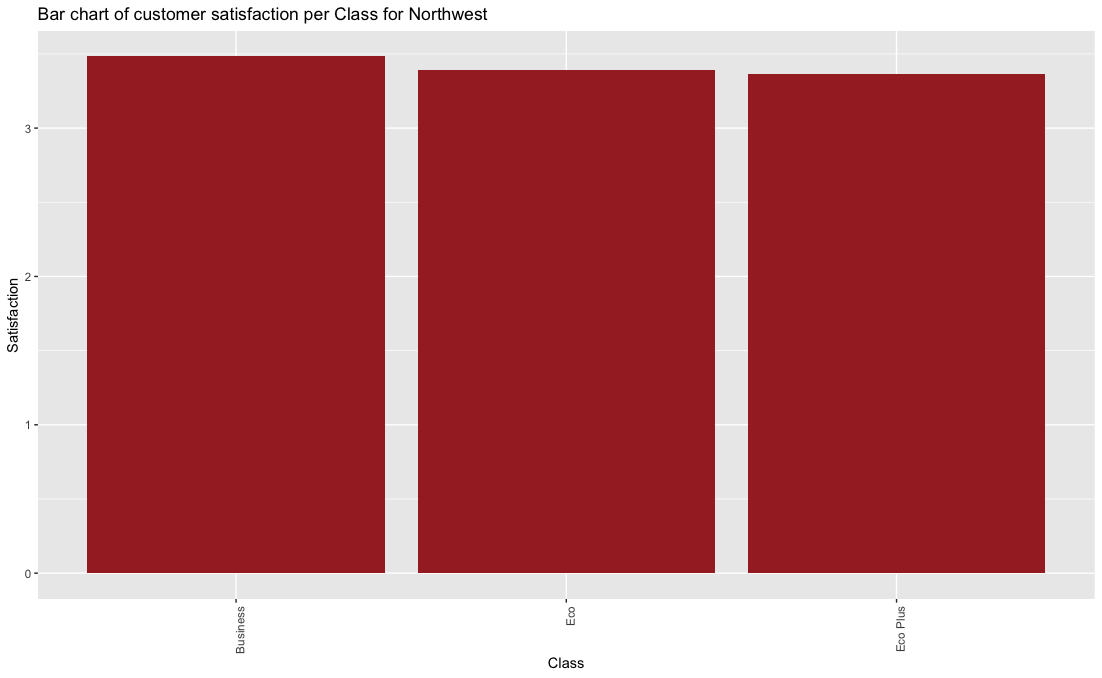
1. *Customer Satisfaction per Class in Southeast Airlines [Graph 3.2]*



This visual representation graphs the different types of classes, Business, Economy and Economy Plus. Business has the highest Satisfaction maybe due to the fact that you get what you pay for. The difference between the classes isn't great.

Class Satisfaction

1. Business 3.483146
2. Eco 3.392472
3. Eco Plus 3.361895
4. *Customer Satisfaction per Class in Northeast Airlines [Graph 3.3]*



This visual representation graphs the different types of classes for Northwest Airlines; Business, Economy and Economy Plus. It looks very similar to Southeast's classes

Class Satisfaction

1. Business 3.580292
2. Eco 3.310452
3. Eco Plus 3.279054

**Chapter 5 -** USE OF MODELLING TECHNIQUES

* Apriori Algorithm Model
* Linear Model
* Support Vector Machine Model

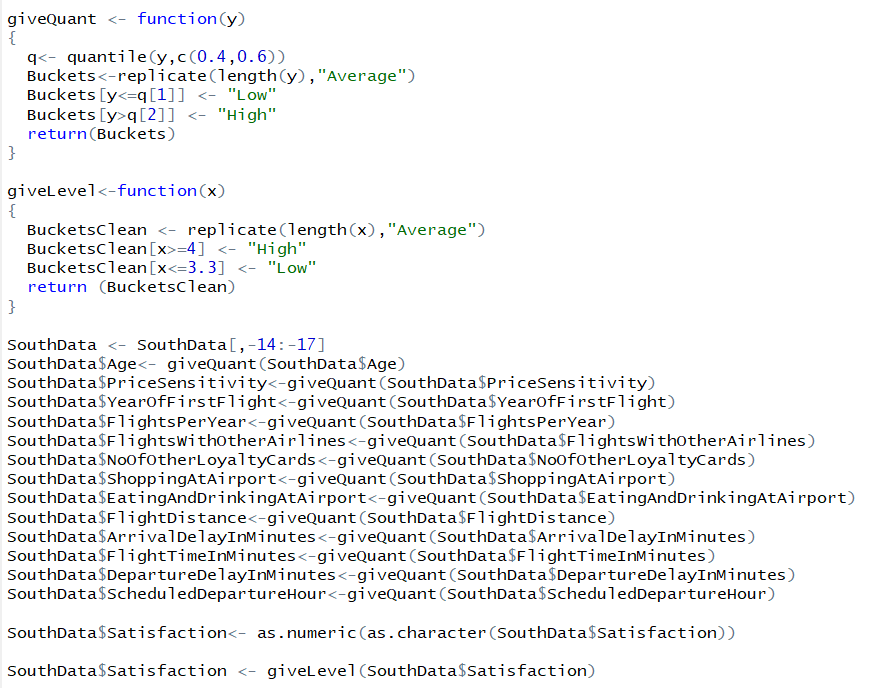
1. Apriori Algorithm (Association Rules)

Apriori is an algorithm for frequent item set mining and association rule learning over transactional databases.

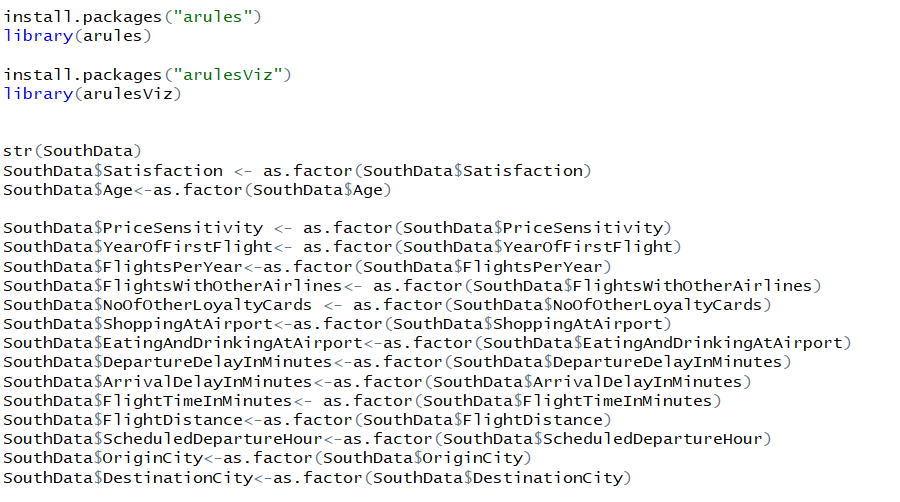
It proceeds by identifying the frequent individual items in the database and extending them to larger and larger item sets as long as those item sets appear sufficiently often in the database.

**Code Snippet**:

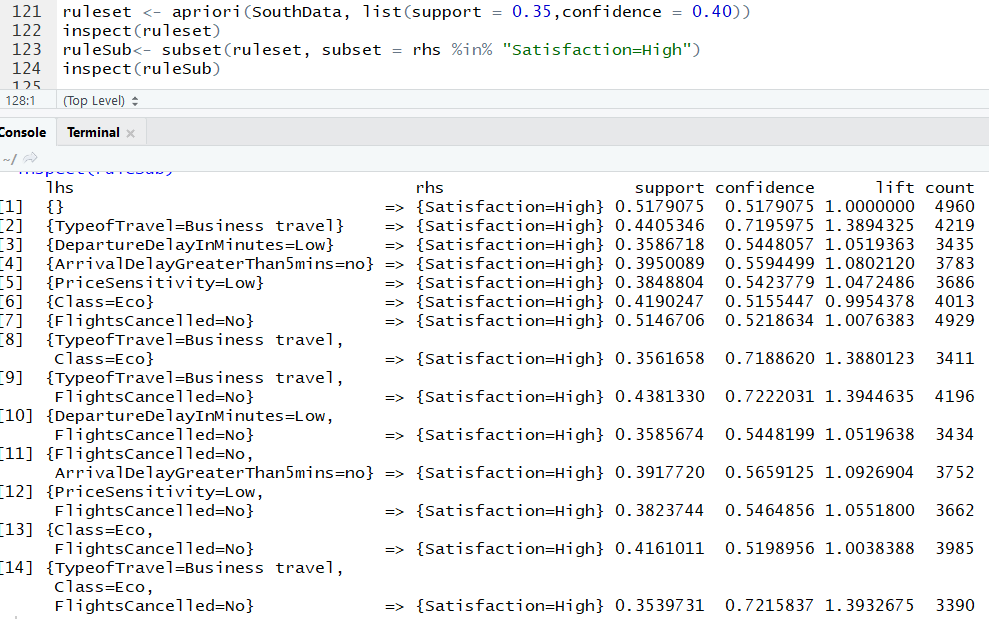
Made functions and passed variables in order to later convert them in factors (High, Low, Average)



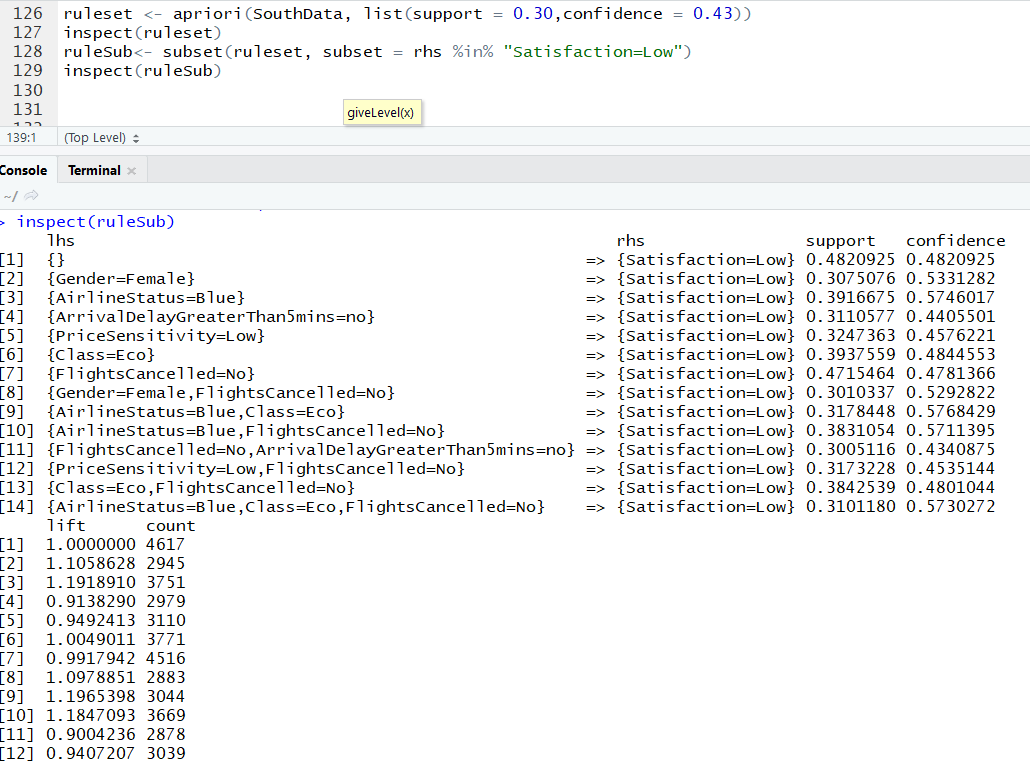
Install necessary packages for Apriori and convert them to factors.



Run the code in order to achieve the Association Rules for Satisfaction=High on RHS.

****

Run the code in order to achieve the Association Rules for Satisfaction=Low on RHS.



**Meaningful Insights (Apriori Algorithm):**

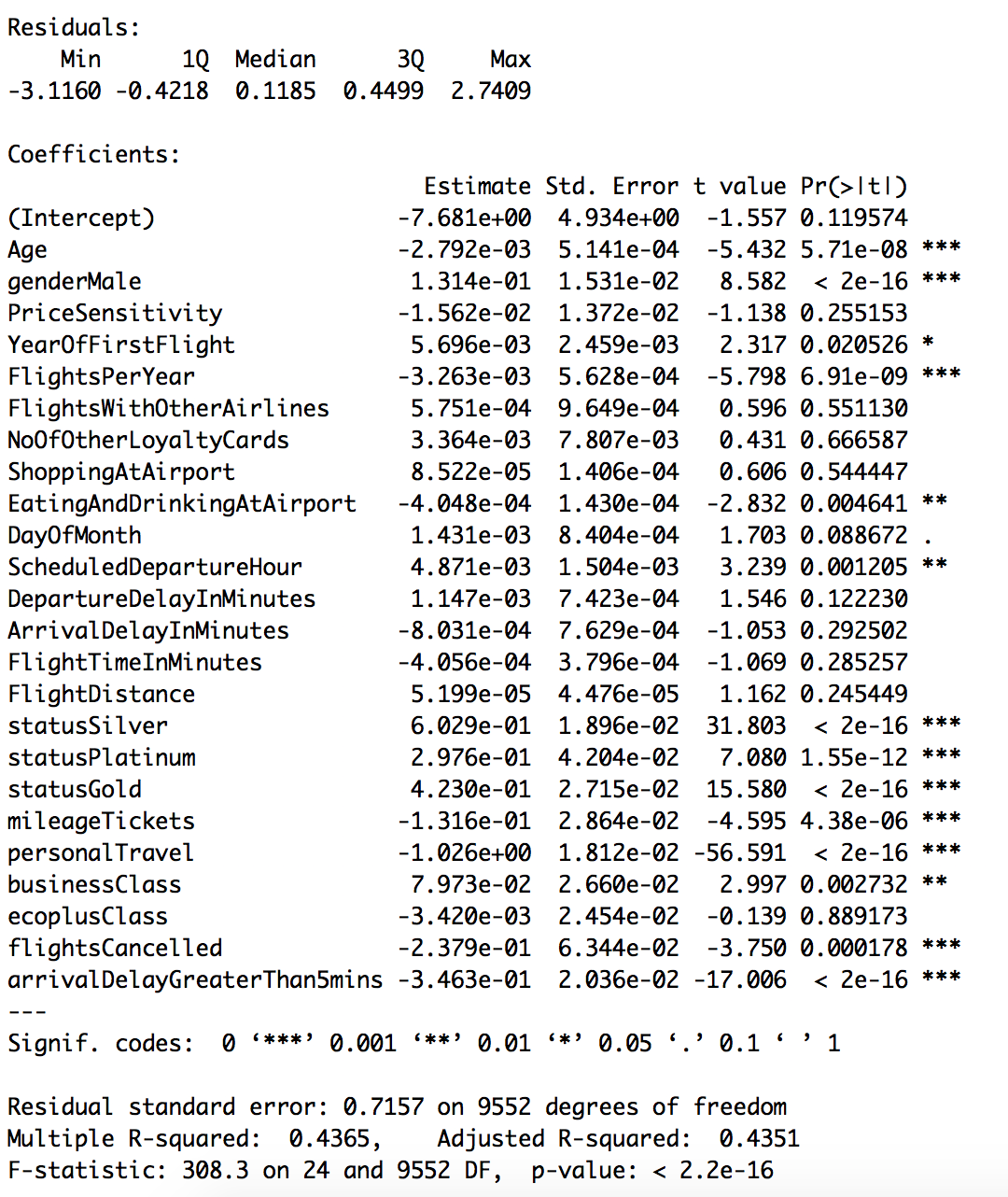
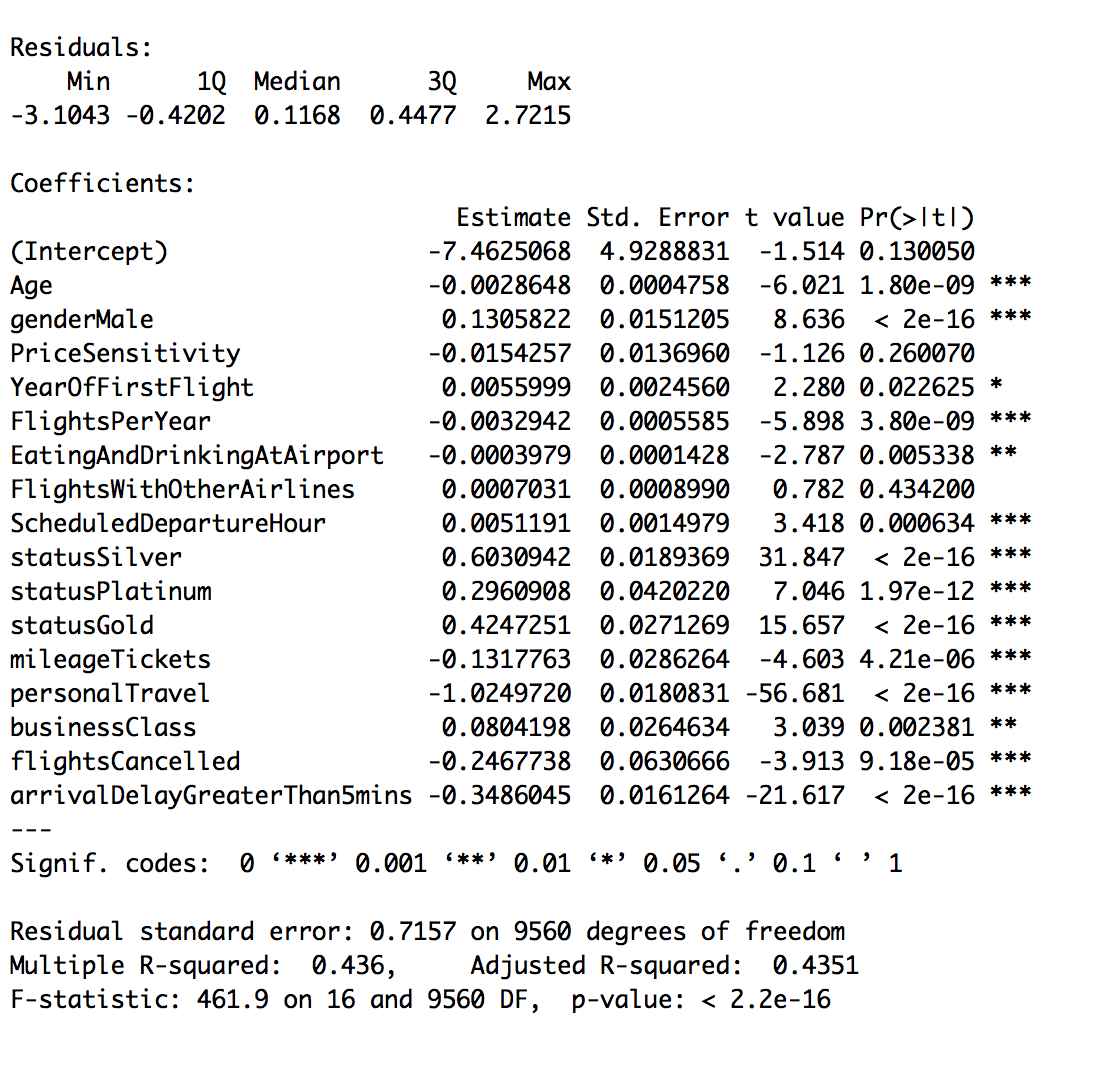
1. **Airline Status =** BLUE **→ Satisfaction =** LOW
2. **Gender =** Female **→ Satisfaction =** LOW
3. **Type Of Travel =** Business Travel **→ Satisfaction =** HIGH
4. **Price Sensitivity =** LOW **→ Satisfaction =** HIGH
5. **Departure Delay in Minutes =** LOW **→ Satisfaction =** HIGH
6. **Flights Cancelled =** NO **→ Satisfaction =** HIGH

2) Linear Model

* Linear models describe a continuous response variable as a function of one or more predictor variables.
* They can help you understand and predict the behavior of complex systems or analyze experimental, financial, and biological data.
* The variable that we're trying to model or predict is known as the dependent variable, and the variables that we use to make predictions are known as independent variables, or covariates

**Code Snippet**:





**Meaningful Insights**:

Positive Impact on Satisfaction:

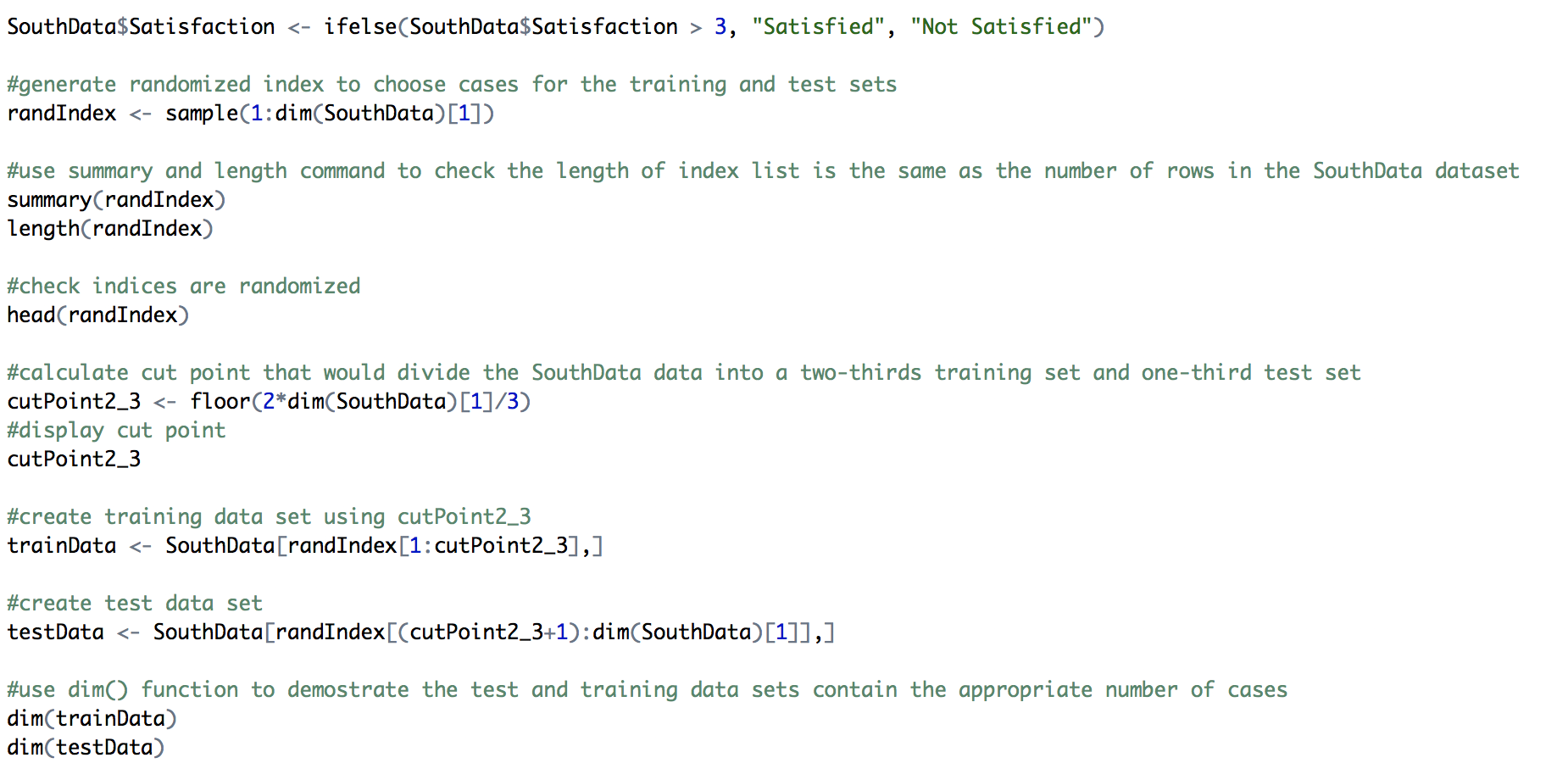
* + Gender
  + Year of First Flight
  + Scheduled Departure Hour
  + Airline Status - Silver, Gold, Platinum
  + Traveling Business Class

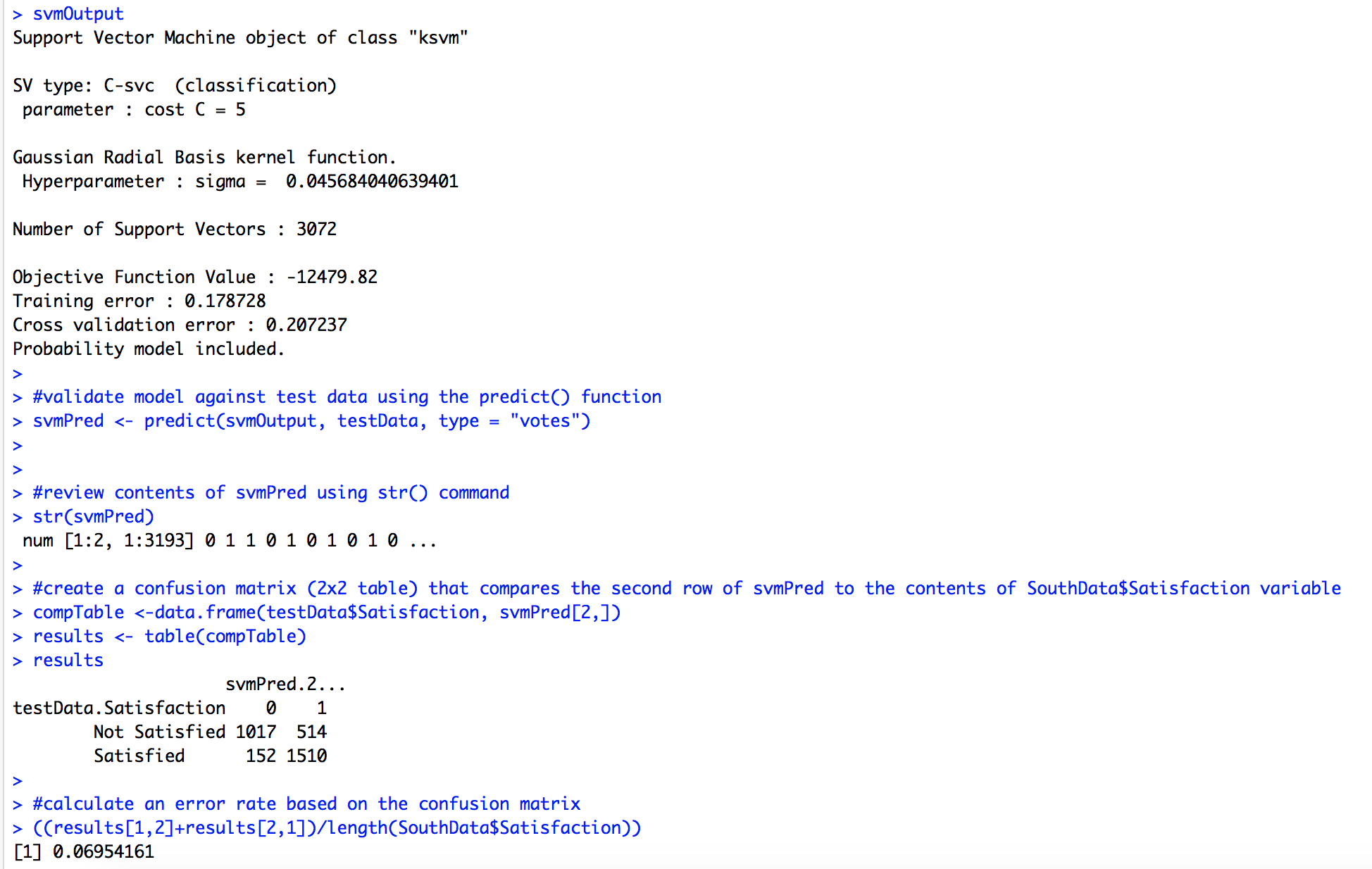
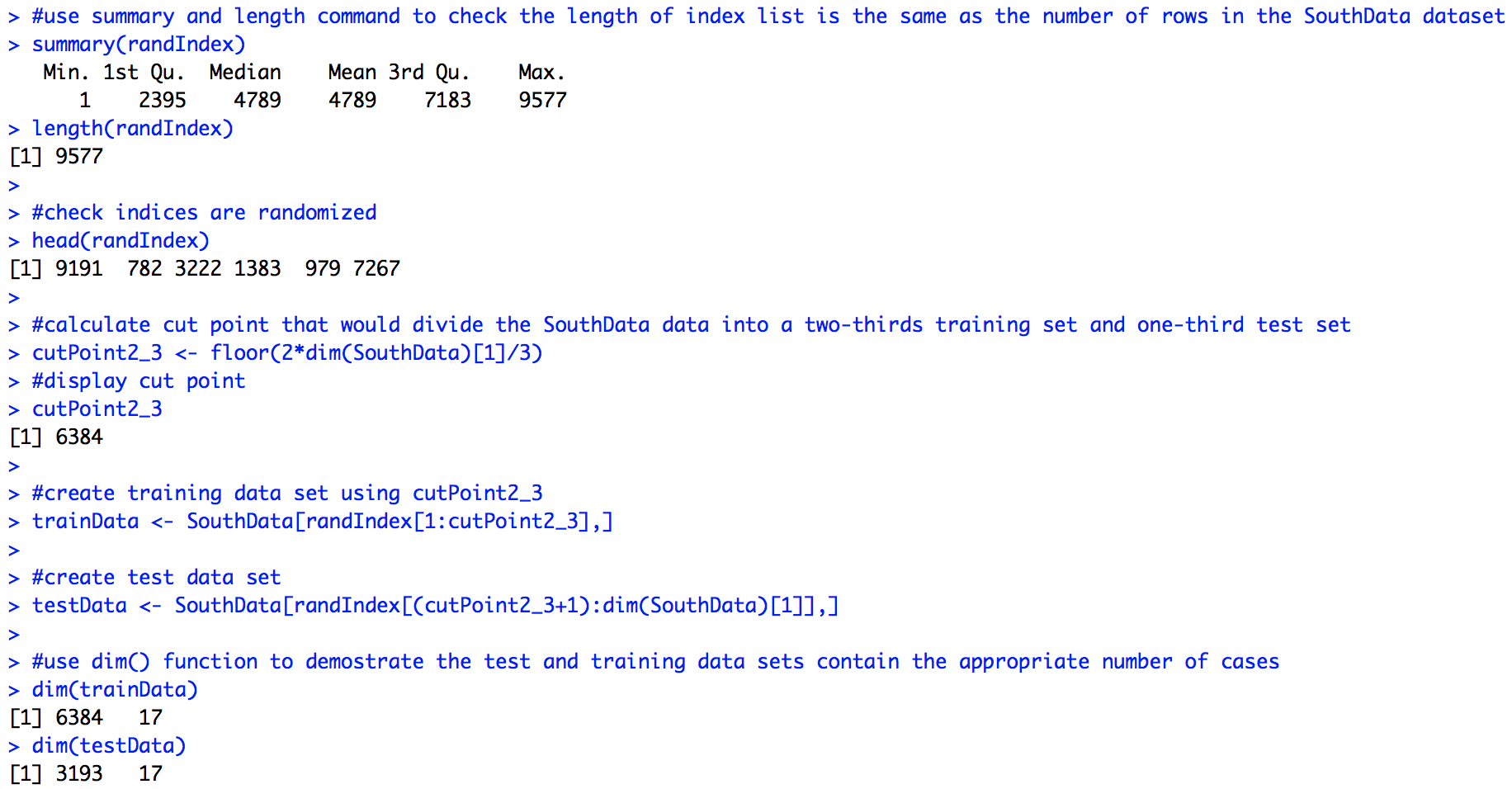
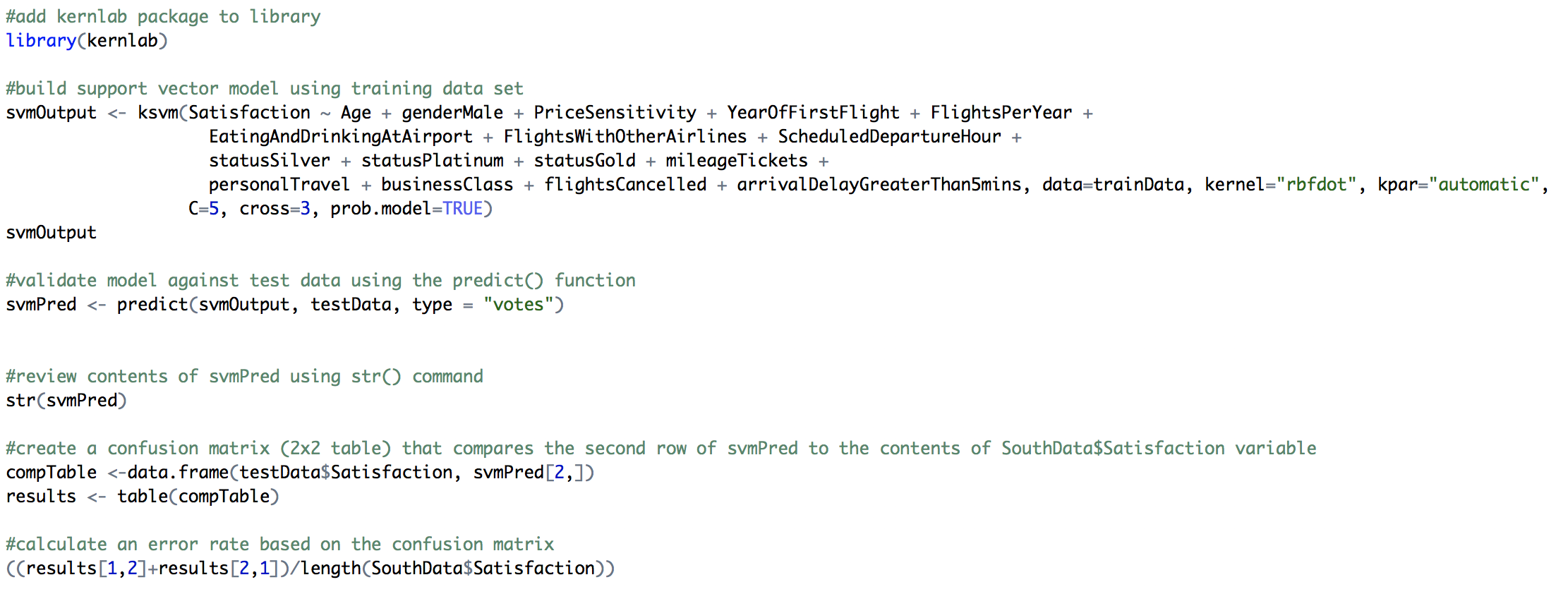
Negative Impact on Satisfaction:

* + Age
  + Number of Flights per Year
  + Eating and Drinking at Airport
  + Type of Travel - mileage
  + Flight Cancelled
  + Arrival Delay Greater Than 5 Minutes

3) Support Vector Machines

* A Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyperplane.
* In other words, given labeled training data (supervised learning), the algorithm outputs an optimal hyperplane which categorizes new examples.
* In two dimensional space this hyperplane is a line dividing a plane in two parts where in each class lay in either side.

**Code Snippet :**

****

**Meaningful Insights:**

1. Satisfaction was reported as greater than 3
2. Not Satisfied Customers, Satisfaction was reported as 3 or less
3. SVM Model Correctly Identified
   1. 1063 NotSatisfied Customers
   2. 1480 Satisfied Customers
4. 80% of the responses were predicted correctly from the svm model

**Chapter 6 -** ACTIONABLE INSIGHTS / OVERALL INTERPRETATION OF THE RESULTS

|  |  |
| --- | --- |
| OBSERVATION | RECOMMENDATION |
| As age increases, Personal travel surpasses both Business and Mileage ticket travel | Possible stipend for seniors, as their frequency of travel increase while price sensitivity remains unchanged, essentially allowing seniors to travel more aiding in higher satisfaction |
| Price Sensitivity for Business travel and Mileage has the least variation throughout age and frequency | To increase profits, increasing the price of business travel would have little no impact on the overall change in frequency. Coupled with an increase benefit |
| Mileage tickets and Personal travel are correlated from 15 - 25, while satisfaction is rather low | Promote a higher mileage point award system for business travel around age 45, where the highest frequency is, and where we believe people are using mileage tickets for both personal and mileage type of travel outside of work. This will raise the overall satisfaction as price sensitivity is correlated to satisfaction |
| Customers who travel by BLUE STATUS flights are unsatisfied | Improve amenities for blue status flights and work |
| Customers who are in the category of Personal travel have the lowest overall satisfaction | Provide family plans to reduce flight cost, this would increase overall satisfaction within the subcategory of Personal travel. |
| Customers in a higher age range are unsatisfied | Better healthcare facilities and proper care of their needs which will drive their Satisfaction. Discount Offers for Senior citizens. |
| Women have 1000 more flights in Personal travel, which has the lowest average satisfaction. | Women have 1000 more flights in Personal travel, which has the lowest average satisfaction. By alleviate low satisfaction for this group in Personal travel in any way would increase in the overall satisfaction of that type of travel. |

**Chapter 7 -** CONCLUSION

The data analyzed for Southeast Airlines provided useful insights that helped Southeast Airlines in order to improve customer satisfaction which will in turn gain profits giving a competitive edge over competitors.

**Chapter 8 -** APPENDIX

### Pat Carlin

### IST 687 FinalProject

#################### Section 0: Loading and Cleaning the Data

### Insalling Packages

install.packages("ggplot2")

install.packages("dplyr")

install.packages("ggmap")

install.packages("mapdata")

install.packages("wordcloud")

install.packages("tm")

library(mapdata)

library(ggmap)

library(ggplot2)

library(dplyr)

library(wordcloud)

library(tm)

### Reading the data into R

ProjectData <- read.csv("~/IST 687 Project/ProjectFullDataSet.csv")

str(ProjectData)

### Cleaning the names and punctuation of the dataset

colnames(ProjectData)

colnames(ProjectData)[6]<- "YearOfFirstFlight"

colnames(ProjectData)[8]<- "FlightsWithOtherAirlines"

colnames(ProjectData)[7]<- "FlightsPerYear"

colnames(ProjectData)[9]<- "TypeOfTravel"

colnames(ProjectData)[10]<- "NoOfOtherLoyaltyCards"

colnames(ProjectData)[11]<- "ShoppingAtAirport"

colnames(ProjectData)[12]<- "EatingAndDrinkingAtAirport"

colnames(ProjectData)[14]<- "DayOfMonth"

colnames(ProjectData)[15]<- "FlightDate"

colnames(ProjectData)[25]<- "FlightsCancelled"

colnames(ProjectData)[26]<- "FlightTimeInMinutes"

colnames(ProjectData)[28]<- "ArrivalDelayGreaterThan5mins"

colnames(ProjectData)[23]<- "DepartureDelayInMinutes"

colnames(ProjectData)[24]<- "ArrivalDelayInMinutes"

colnames(ProjectData)[18]<- "OriginCity"

ProjectData$DestinationCity<- gsub("(.\*),.\*", "\\1",ProjectData$DestinationCity)

ProjectData$OriginCity<- gsub("(.\*),.\*", "\\1",ProjectData$OriginCity)

names(ProjectData) <- gsub("\\.","", names(ProjectData))

names(ProjectData)

#################### Section 1: Focusing on our Southest Airlines Co.

### Isolating the Southeast Airlines Co.

SoutheastData <- ProjectData[which(ProjectData$AirlineName == "Southeast Airlines Co. " ),] #For Seperate DataFrame

View(SoutheastData)

### Transforming columns in SoutheastData to 'Factor' and 'Numeric'

SoutheastData$Satisfaction <- as.numeric(as.character(SoutheastData$Satisfaction))

SoutheastData$Age <- as.numeric(as.character(SoutheastData$Age))

SoutheastData$PriceSensitivity <- as.numeric(as.character(SoutheastData$PriceSensitivity))

SoutheastData$YearOfFirstFlight <- as.numeric(as.character(SoutheastData$YearOfFirstFlight))

SoutheastData$FlightsPerYear <- as.numeric(as.character(SoutheastData$FlightsPerYear))

SoutheastData$FlightsWithOtherAirlines <- as.numeric(as.character(SoutheastData$FlightsWithOtherAirlines))

#################### Section 3: Exploring how the TypeOfTravel effects the Satisfaction

### Plotting the average Satisfaction against TypeOfTravel

TypeOfTravSat <- SoutheastData %>% group\_by(TypeOfTravel) %>% summarise(TypeSat = mean(Satisfaction))

TypeOfTravPlot <- ggplot(TypeOfTravSat, aes(x = TypeOfTravel, y = TypeSat)) + geom\_bar(stat = "identity", color = "black", fill = "orange") + ylim(0,5)

TypeOfTravPlot <- TypeOfTravPlot + ggtitle("Type Of Travel vs Satisfaction") + xlab("Type Of Travel") + ylab("Satisfaction")

TypeOfTravPlot <- TypeOfTravPlot + theme\_classic() + theme(plot.title = element\_text(hjust = 0.5))

TypeOfTravPlot

### Creating a boxplot to visualize frequency of the TypeOfTravel

TypeSatBoxPlot <- ggplot(data = SoutheastData, aes(x = TypeOfTravel, y = Satisfaction)) + geom\_boxplot(fill = "lightblue", colour = "black")

TypeSatBoxPlot <- TypeSatBoxPlot + ggtitle("Type Of Travel vs Satisfaction") + xlab("Type Of Travel") + ylab("Satisfaction") + theme\_bw()

TypeSatBoxPlot <- TypeSatBoxPlot + theme(plot.title = element\_text(hjust = 0.5)) + geom\_jitter()

TypeSatBoxPlot

### Creating a smoothed scatter plot to see the effects of TypeOfTravel on Age and average Satisfaction

TypeAgeSatisfaction <- SoutheastData %>% group\_by(Age, TypeOfTravel) %>% summarise(TypeAgeSat = mean(Satisfaction))

AgeTypeLinePlot <- ggplot(data = TypeAgeSatisfaction, aes( x = Age, y = TypeAgeSat, group = TypeOfTravel, colour = TypeOfTravel)) + geom\_point() + geom\_smooth()

AgeTypeLinePlot <- AgeTypeLinePlot + xlab("Age") + ylab("Satisfaction") + ggtitle("Types Of Travel vs Satisfaction") + theme\_bw() + theme(plot.title = element\_text(hjust = 0.5))

AgeTypeLinePlot <- AgeTypeLinePlot + scale\_x\_continuous(breaks = seq(15, 85, 5))

AgeTypeLinePlot

### Plotting the AgeOfPlane against TypeOfTravel

AgeTypePlot <- ggplot(data = SoutheastData, aes(x = Age, fill = TypeOfTravel)) + geom\_histogram(binwidth = 2)

AgeTypePlot <- AgeTypePlot + xlab("Age") + ylab("Number Of Flights") + ggtitle("Types Of Travel Across Age") + theme\_classic() + theme(plot.title = element\_text(hjust = 0.5))

AgeTypePlot <- AgeTypePlot + scale\_x\_continuous(breaks = seq(15, 85, 5))

AgeTypePlot ### Need to find a better visual

### Finding the density plot for the TypeOfTravel and Age

TypeDensityPlot <- ggplot(SoutheastData, aes(Age)) + geom\_density(aes(fill = factor(TypeOfTravel)), alpha=0.8)

TypeDensityPlot <- TypeDensityPlot + xlab("Age") + ylab("Flight Density") + ggtitle("Flight Density for Travel Types") + theme\_classic() + theme(plot.title = element\_text(hjust = 0.5))

TypeDensityPlot <- TypeDensityPlot + scale\_x\_continuous(breaks = seq(15, 85, 5)) + guides(fill=guide\_legend(title="Types"))

TypeDensityPlot

### Seeing the effects of Age on average PriceSensitivity between TypeOfTravel

PriceAgeSat <- SoutheastData %>% group\_by(Age, TypeOfTravel) %>% summarise(PriceSat = mean(PriceSensitivity)

PriceAgePlot <- ggplot(data = PriceAgeSat, aes(x = Age, y = PriceSat, group = TypeOfTravel, colour = TypeOfTravel)) + geom\_smooth()

PriceAgePlot <- PriceAgePlot + ggtitle("Satisfaction vs Time & Price Sensitivity") + xlab("Age") + ylab("Average Price Sensitivity") + theme\_bw()

PriceAgePlot <- PriceAgePlot + theme\_bw() + theme(plot.title = element\_text(hjust = 0.5))

PriceAgePlot <- PriceAgePlot + scale\_x\_continuous(breaks = seq(15, 85, 5))

PriceAgePlot

############## Section 4: Exploring how Gender and Age affect Satisfaction within SoutheastData

### Showing variance in Gender and frequency of flight between the TypeOfTravel

TypeGenderPlot <- ggplot(data = SoutheastData, aes(x = TypeOfTravel, fill = Gender)) + geom\_bar(colour = "black", stat="count", position = position\_dodge(), size = .5)

TypeGenderPlot <- TypeGenderPlot + scale\_fill\_hue(name="Gender") + xlab("Type Of Travel") + ylab("Number Of Flights") + ggtitle("Type Of Flights Across Gender")

TypeGenderPlot <- TypeGenderPlot + theme\_bw() + theme(plot.title = element\_text(hjust = 0.5))

TypeGenderPlot

### Plotting average Satisfaction against Age for Gender

GenderAgeSat <- SoutheastData %>% group\_by(Age, Gender) %>% summarise(GenAgeSat = mean(Satisfaction))

GenderAgePlot <- ggplot(data = GenderAgeSat, aes(x = Age, y = GenAgeSat, group = Gender, colour = Gender)) + geom\_line()

GenderAgePlot <- GenderAgePlot + ggtitle("Satisfaction vs Time & Gender") + xlab("Age") + ylab("Satisfaction") + theme\_classic()

GenderAgePlot <- GenderAgePlot + theme\_bw() + theme(plot.title = element\_text(hjust = 0.5))

GenderAgePlot <- GenderAgePlot + scale\_x\_continuous(breaks = seq(15, 85, 5))

GenderAgePlot

##ShwetJain

##IST 687 Final Project

#CleanedData

#Different Plots

### Installing Packages

install.packages("ggplot2")

library(ggplot2)

### Reading the data into R

ProjectData <- read.csv("Satisfaction Survey.csv", header = TRUE)

str(ProjectData)

### Removing "." in ProjectData

names(ProjectData)

names(ProjectData) <- gsub("\\.","", names(ProjectData))

names(ProjectData)

View(ProjectData)

colnames(ProjectData)

colnames(ProjectData)[6]<- "YearOfFirstFlight"

colnames(ProjectData)[8]<- "FlightsWithOtherAirlines"

colnames(ProjectData)[7]<- "FlightsPerYear"

colnames(ProjectData)[10]<- "NoOfOtherLoyaltyCards"

colnames(ProjectData)[11]<- "ShoppingAtAirport"

colnames(ProjectData)[12]<- "EatingAndDrinkingAtAirport"

colnames(ProjectData)[14]<- "DayOfMonth"

colnames(ProjectData)[15]<- "FlightDate"

colnames(ProjectData)[25]<- "FlightsCancelled"

colnames(ProjectData)[26]<- "FlightTimeInMinutes"

colnames(ProjectData)[28]<- "ArrivalDelayGreaterThan5mins"

colnames(ProjectData)[23]<- "DepartureDelayInMinutes"

colnames(ProjectData)[24]<- "ArrivalDelayInMinutes"

colnames(ProjectData)[18]<- "OriginCity"

ProjectData$DestinationCity<- gsub("(.\*),.\*", "\\1",ProjectData$DestinationCity)

ProjectData$OriginCity<- gsub("(.\*),.\*", "\\1",ProjectData$OriginCity)

SouthData <- ProjectData[which(ProjectData$AirlineName == "Southeast Airlines Co. " ),] #For Seperate DataFrame

View(SouthData)

str(SouthData)

sum(is.na(SouthData$ArrivalDelayInMinutes)) #154

sum(is.na(SouthData$FlightTimeInMinutes)) #154

sum(is.na(SouthData$DepartureDelayInMinutes)) #129

#REPLACING THE NA’S

SouthData$ArrivalDelayInMinutes[is.na(SouthData$ArrivalDelayInMinutes)] <- round(mean(SouthData$ArrivalDelayInMinutes, na.rm = TRUE))

SouthData$FlightTimeInMinutes[is.na(SouthData$FlightTimeInMinutes)] <- round(mean(SouthData$FlightTimeInMinutes, na.rm = TRUE))

SouthData$DepartureDelayInMinutes[is.na(SouthData$DepartureDelayInMinutes)] <- round(mean(SouthData$DepartureDelayInMinutes, na.rm = TRUE))

install.packages("ggplot2")

library(ggplot2)

install.packages("dplyr")

library(dplyr)

SouthData$Satisfaction <- as.numeric(as.character(SouthData$Satisfaction))

str(SouthData)

by\_AirlineStatus <- group\_by(SouthData,AirlineStatus)

by\_AirlineStatus<-summarise(by\_day,Satisfaction=mean(Satisfaction))

Plot\_AirlineStatus <- ggplot(by\_day,aes(x=AirlineStatus,y=Satisfaction)) + geom\_col()+ theme(axis.text.x = element\_text(angle = 90))+ ggtitle("Relationship between AirlineStatus and Satisfaction") #Using the theme function and setting it to 90 will flip it and give a clear representation.

Age\_Gender<-ggplot(SouthData, aes(x=Age, fill=Gender, color=Gender)) + geom\_histogram(position="identity", alpha=0.5) + geom\_vline(data=mu, aes(xintercept=grp.mean, color=Gender),linetype="dashed")

Airlines\_Count <- ggplot(ProjectData, aes(AirlineName)) + geom\_bar() + theme(axis.text.x = element\_text(angle = 90)) + geom\_bar(aes(fill = Satisfaction)) + ggtitle("Frequency of Different Airlines with repective Customer Satisfaction")

ProjectData$Satisfaction<- as.factor(ProjectData$Satisfaction)

by\_Gender <- group\_by(SouthData,Gender)

by\_Gender<-summarise(by\_Gender,Satisfaction=mean(Satisfaction))

Gender\_Plot <-ggplot(by\_Gender, aes(x = Gender, y = Satisfaction)) + geom\_bar(stat = "identity") + ggtitle("Relationship between Gender and Satisfaction")

by\_PriceSensitivity <- group\_by(SouthData,PriceSensitivity)

by\_PriceSensitivity <-summarise(by\_PriceSensitivity,Satisfaction=mean(Satisfaction))

PriceS\_Plot<- ggplot(by\_PriceSensitivity, aes(x =PriceSensitivity, y = Satisfaction)) + geom\_bar(stat = "identity") + ggtitle("Relationship between Price Sensitivity and Satisfaction")

by\_Class <- group\_by(SouthData,Class)

by\_Class <-summarise(by\_Class,Satisfaction=mean(Satisfaction))

ClassPlot <-ggplot(by\_Class, aes(x=Class, y=Satisfaction)) + geom\_bar(stat="identity", fill="steelblue") + theme\_minimal()

by\_OriginDestination <- group\_by(SouthData,OriginCity,DestinationCity)

by\_OriginDestination <-summarise(by\_OriginDestination,Satisfaction=mean(Satisfaction),count=n())

OriginDestinationPlot<- ggplot(by\_OriginDestination, aes(x=OriginCity, y=DestinationCity)) + geom\_point(aes(size=Satisfaction)) + theme(axis.text.x = element\_text(angle = 90)) + ggtitle("Relation between OriginState and DestinationState with it's Satisfaction")

by\_NoOfOtherLoyaltyCards <- group\_by(SouthData,NoOfOtherLoyaltyCards)

by\_NoOfOtherLoyaltyCards <-summarise(by\_NoOfOtherLoyaltyCards,Satisfaction=mean(Satisfaction))

LoyaltyCardsPlot<- ggplot(by\_NoOfOtherLoyaltyCards, aes(x =NoOfOtherLoyaltyCards, y = Satisfaction)) + geom\_bar(stat = "identity")

by\_PriceSensitivity\_TypeOfTravel <- group\_by(SouthData,PriceSensitivity,TypeofTravel)

by\_PriceSensitivity\_TypeOfTravel<-summarise(by\_PriceSensitivity\_TypeOfTravel,Satisfaction=mean(Satisfaction),count=n())

plot(by\_PriceSensitivity\_TypeOfTravel)

ggplot(by\_PriceSensitivity\_TypeOfTravel, aes(fill=by\_PriceSensitivity\_TypeOfTravel$TypeofTravel, y=by\_PriceSensitivity\_TypeOfTravel$Satisfaction, x=by\_PriceSensitivity\_TypeOfTravel$PriceSensitivity)) + xlab("PriceSensitivity")+ ylab("Satisfaction")+geom\_bar(position="dodge", stat="identity")

#Association Model

### Insalling Packages

install.packages("ggplot2")

library(ggplot2)

### Reading the data into R

ProjectData <- read.csv("Satisfaction Survey.csv", header = TRUE)

str(ProjectData)

### Removing "." in ProjectData

names(ProjectData)

names(ProjectData) <- gsub("\\.","", names(ProjectData))

names(ProjectData)

View(ProjectData)

summary(ProjectData)

colnames(ProjectData)

colnames(ProjectData)[6]<- "YearOfFirstFlight"

colnames(ProjectData)[8]<- "FlightsWithOtherAirlines"

colnames(ProjectData)[7]<- "FlightsPerYear"

colnames(ProjectData)[10]<- "NoOfOtherLoyaltyCards"

colnames(ProjectData)[11]<- "ShoppingAtAirport"

colnames(ProjectData)[12]<- "EatingAndDrinkingAtAirport"

colnames(ProjectData)[14]<- "DayOfMonth"

colnames(ProjectData)[15]<- "FlightDate"

colnames(ProjectData)[25]<- "FlightsCancelled"

colnames(ProjectData)[26]<- "FlightTimeInMinutes"

colnames(ProjectData)[28]<- "ArrivalDelayGreaterThan5mins"

colnames(ProjectData)[23]<- "DepartureDelayInMinutes"

colnames(ProjectData)[24]<- "ArrivalDelayInMinutes"

colnames(ProjectData)[18]<- "OriginCity"

ProjectData$DestinationCity<- gsub("(.\*),.\*", "\\1",ProjectData$DestinationCity)

ProjectData$OriginCity<- gsub("(.\*),.\*", "\\1",ProjectData$OriginCity)

SouthData <- ProjectData[which(ProjectData$AirlineName == "Southeast Airlines Co. " ),] #For Seperate DataFrame

View(SouthData)

str(SouthData)

sum(is.na(SouthData$ArrivalDelayInMinutes)) #154

sum(is.na(SouthData$FlightTimeInMinutes)) #154

sum(is.na(SouthData$DepartureDelayInMinutes)) #129

SouthData$ArrivalDelayInMinutes[is.na(SouthData$ArrivalDelayInMinutes)] <- round(mean(SouthData$ArrivalDelayInMinutes, na.rm = TRUE))

SouthData$FlightTimeInMinutes[is.na(SouthData$FlightTimeInMinutes)] <- round(mean(SouthData$FlightTimeInMinutes, na.rm = TRUE))

SouthData$DepartureDelayInMinutes[is.na(SouthData$DepartureDelayInMinutes)] <- round(mean(SouthData$DepartureDelayInMinutes, na.rm = TRUE))

str(SouthData)

giveQuant <- function(y)

{

q<- quantile(y,c(0.4,0.6))

Buckets<-replicate(length(y),"Average")

Buckets[y<=q[1]] <- "Low"

Buckets[y>q[2]] <- "High"

return(Buckets)

}

giveLevel<-function(x)

{

BucketsClean <- replicate(length(x),"Average")

BucketsClean[x>=4] <- "High"

BucketsClean[x<=3.3] <- "Low"

return (BucketsClean)

}

SouthData <- SouthData[,-14:-17]

SouthData$Age<- giveQuant(SouthData$Age)

SouthData$PriceSensitivity<-giveQuant(SouthData$PriceSensitivity)

SouthData$YearOfFirstFlight<-giveQuant(SouthData$YearOfFirstFlight)

SouthData$FlightsPerYear<-giveQuant(SouthData$FlightsPerYear)

SouthData$FlightsWithOtherAirlines<-giveQuant(SouthData$FlightsWithOtherAirlines)

SouthData$NoOfOtherLoyaltyCards<-giveQuant(SouthData$NoOfOtherLoyaltyCards)

SouthData$ShoppingAtAirport<-giveQuant(SouthData$ShoppingAtAirport)

SouthData$EatingAndDrinkingAtAirport<-giveQuant(SouthData$EatingAndDrinkingAtAirport)

SouthData$FlightDistance<-giveQuant(SouthData$FlightDistance)

SouthData$ArrivalDelayInMinutes<-giveQuant(SouthData$ArrivalDelayInMinutes)

SouthData$FlightTimeInMinutes<-giveQuant(SouthData$FlightTimeInMinutes)

SouthData$DepartureDelayInMinutes<-giveQuant(SouthData$DepartureDelayInMinutes)

SouthData$ScheduledDepartureHour<-giveQuant(SouthData$ScheduledDepartureHour)

SouthData$Satisfaction<- as.numeric(as.character(SouthData$Satisfaction))

SouthData$Satisfaction <- giveLevel(SouthData$Satisfaction)

install.packages("arules")

library(arules)

install.packages("arulesViz")

library(arulesViz)

str(SouthData)

SouthData$Satisfaction <- as.factor(SouthData$Satisfaction)

SouthData$Age<-as.factor(SouthData$Age)

SouthData$PriceSensitivity <- as.factor(SouthData$PriceSensitivity)

SouthData$YearOfFirstFlight<- as.factor(SouthData$YearOfFirstFlight)

SouthData$FlightsPerYear<-as.factor(SouthData$FlightsPerYear)

SouthData$FlightsWithOtherAirlines<- as.factor(SouthData$FlightsWithOtherAirlines)

SouthData$NoOfOtherLoyaltyCards <- as.factor(SouthData$NoOfOtherLoyaltyCards)

SouthData$ShoppingAtAirport<-as.factor(SouthData$ShoppingAtAirport)

SouthData$EatingAndDrinkingAtAirport<-as.factor(SouthData$EatingAndDrinkingAtAirport)

SouthData$DepartureDelayInMinutes<-as.factor(SouthData$DepartureDelayInMinutes)

SouthData$ArrivalDelayInMinutes<-as.factor(SouthData$ArrivalDelayInMinutes)

SouthData$FlightTimeInMinutes<- as.factor(SouthData$FlightTimeInMinutes)

SouthData$FlightDistance<-as.factor(SouthData$FlightDistance)

SouthData$ScheduledDepartureHour<-as.factor(SouthData$ScheduledDepartureHour)

SouthData$OriginCity<-as.factor(SouthData$OriginCity)

SouthData$DestinationCity<-as.factor(SouthData$DestinationCity)

ruleset <- apriori(SouthData, list(support = 0.35,confidence = 0.40))

inspect(ruleset)

ruleSub<- subset(ruleset, subset = rhs %in% "Satisfaction=High")

inspect(ruleSub)

ruleset <- apriori(SouthData, list(support = 0.30,confidence = 0.43))

inspect(ruleset)

ruleSub<- subset(ruleset, subset = rhs %in% "Satisfaction=Low")

inspect(ruleSub)

#####Osama Junaid ##########

#final project

#install ggplot and ggmap

install.packages("ggplot2")

install.packages("mapdata")

install.packages("ggmap")

install.packages("dplyr")

library("ggplot2")

library(ggplot2)

library(ggmap)

library(maps)

library(mapdata)

library(dplyr)

install.packages("arules")

library(arules)

install.packages("arulesViz")

library(arulesViz)

install.packages("tmap")

install.packages("tmaptools")

install.packages("sf")

install.packages("leaflet")

library("tmap")

library("tmaptools")

library("sf")

library("leaflet")

ProjectData <- read.csv(file.choose(), header = TRUE)

View(ProjectData)

str(ProjectData)

### Removing "." in ProjectData

names(ProjectData)

names(ProjectData) <- gsub("\\.","", names(ProjectData))

names(ProjectData)

View(ProjectData)

colnames(ProjectData)

colnames(ProjectData)[6]<- "YearOfFirstFlight"

colnames(ProjectData)[8]<- "FlightsWithOtherAirlines"

colnames(ProjectData)[7]<- "FlightsPerYear"

colnames(ProjectData)[10]<- "NoOfOtherLoyaltyCards"

colnames(ProjectData)[11]<- "ShoppingAtAirport"

colnames(ProjectData)[12]<- "EatingAndDrinkingAtAirport"

colnames(ProjectData)[14]<- "DayOfMonth"

colnames(ProjectData)[15]<- "FlightDate"

colnames(ProjectData)[25]<- "FlightsCancelled"

colnames(ProjectData)[26]<- "FlightTimeInMinutes"

colnames(ProjectData)[28]<- "ArrivalDelayGreaterThan5mins"

colnames(ProjectData)[23]<- "DepartureDelayInMinutes"

colnames(ProjectData)[24]<- "ArrivalDelayInMinutes"

colnames(ProjectData)[18]<- "OriginCity"

ProjectData$DestinationCity<- gsub("(.\*),.\*", "\\1",ProjectData$DestinationCity)

ProjectData$OriginCity<- gsub("(.\*),.\*", "\\1",ProjectData$OriginCity)

names(ProjectData) <- gsub("\\.", "", names(ProjectData))

names(ProjectData)

View(ProjectData)

############################################################

#New data called Southeast

SoutheastData <- ProjectData[which(ProjectData$AirlineName == "Southeast Airlines Co. " ),] #For Seperate DataFrame

View(SoutheastData)

##############################################################################################################

#Calculating mean of satisfaction and grouping it by Origin state

originstatemean <- SoutheastData %>%

group\_by(OriginState) %>%

summarize(Satisfaction = mean(as.numeric(Satisfaction)))

#Plotting graph for satisfaction vs origin state

SoutheastData$Satisfaction <- as.numeric(as.character(SoutheastData$Satisfaction))

originstatemean<-as.data.frame(originstatemean)

originstatemean

originState <- ggplot(originstatemean, aes(x=OriginState, y=Satisfaction)) + geom\_col()

originState <- originState + ggtitle("Bar chart of customer satisfaction per Origin state for Southeast")

originState <- originState + theme(axis.text.x = element\_text(angle = 90, hjust = 1))

originState <- originState + geom\_bar(stat = "identity", fill ="blue")

originState

##################################################################################################################

#calculating mean of satisfaction and grouping by destination state

destinationstatemean <- SoutheastData %>%

group\_by(DestinationState) %>%

summarize(Satisfaction = mean(as.numeric(Satisfaction)))

#plotting graph of satisfaction vs destination state

destinationstatemean<-as.data.frame(destinationstatemean)

destinationstatemean

destinationState <- ggplot(destinationstatemean, aes(x=DestinationState, y=Satisfaction)) + geom\_col()

destinationState <- destinationState + ggtitle("Bar chart of customer satisfaction per Destination state for Southeast")

destinationState <- destinationState + theme(axis.text.x = element\_text(angle = 90, hjust = 1))

destinationState <- destinationState + geom\_bar(stat = "identity", fill ="Red")

destinationState

########################################################################################################################

#calculating mean of satisfaction and grouping by Eating and Drinking at Airport

EatandDrinkmean <- SoutheastData %>%

group\_by(EatingAndDrinkingAtAirport) %>%

summarize(Satisfaction = mean(as.numeric(Satisfaction)))

#plotting graph of satisfaction vs destination state

EatandDrinkmean<-as.data.frame(EatandDrinkmean)

EatandDrinkmean

EatingandDrinking <- ggplot(EatandDrinkmean, aes(x=EatingAndDrinkingAtAirport, y=Satisfaction)) + geom\_col()

EatingandDrinking <- EatingandDrinking + ggtitle("Bar chart of customer satisfaction per Eating and Drinking at Airport")

EatingandDrinking <- EatingandDrinking + theme(axis.text.x = element\_text(angle = 90, hjust = 1))

EatingandDrinking <- EatingandDrinking + geom\_bar(stat = "identity", fill ="yellow")

EatingandDrinking

########################################################################################################################

#culating mean of satisfaction and grouping by Year of first flight

yearofflightmean <- SoutheastData %>%

group\_by(YearOfFirstFlight) %>%

summarize(Satisfaction = mean(as.numeric(Satisfaction)))

#plotting graph of satisfaction vs destination state

yearofflightmean<-as.data.frame(yearofflightmean)

yearofflightmean

YearofFF <- ggplot(yearofflightmean, aes(x=YearOfFirstFlight, y=Satisfaction)) + geom\_col()

YearofFF <- YearofFF + ggtitle("Barchart of customer satisfaction per Year of first flight for Southeast")

YearofFF <- YearofFF + theme(axis.text.x = element\_text(angle = 90, hjust = 1))

YearofFF <- YearofFF + geom\_bar(stat = "identity", fill ="orange")

YearofFF

########################################################################################################################

#culating mean of satisfaction and grouping by Shopping at airport

shopatAirportmean <- SoutheastData %>%

group\_by(ShoppingAtAirport) %>%

summarize(Satisfaction = mean(as.numeric(Satisfaction)))

#plotting graph of satisfactionstate

shopatAirportmean<-as.data.frame(shopatAirportmean)

shopatAirportmean

shoppingatAirport <- ggplot(shopatAirportmean, aes(x=ShoppingAtAirport, y=Satisfaction)) + geom\_col()

shoppingatAirport <- shoppingatAirport + ggtitle("Barchart of customer satisfaction for Shopping at Airport")

shoppingatAirport <- shoppingatAirport + theme(axis.text.x = element\_text(angle = 90, hjust = 1))

shoppingatAirport <- shoppingatAirport + geom\_bar(stat = "identity", fill ="green")

shoppingatAirport

########################################################################################################################

#culating mean of satisfaction and grouping by class

classmean <- SoutheastData %>%

group\_by(Class) %>%

summarize(Satisfaction = mean(as.numeric(Satisfaction)))

#plotting graph of satisfaction vs destination state

classmean<-as.data.frame(classmean)

classmean

classes <- ggplot(classmean, aes(x=Class, y=Satisfaction)) + geom\_col()

classes <- classes + ggtitle("Bar chart of customer satisfaction per Class")

classes <- classes + theme(axis.text.x = element\_text(angle = 90, hjust = 1))

classes <- classes + geom\_bar(stat = "identity", fill ="#FF6666")

classes

############################################################################

#culating mean of satisfaction and grouping by Type of travel

typeoftravelmean <- SoutheastData %>%

group\_by(TypeofTravel) %>%

summarize(Satisfaction = mean(as.numeric(Satisfaction)))

#plotting graph of satisfaction vs destination state

typeoftravelmean<-as.data.frame(typeoftravelmean)

typeoftravelmean

typeoftravel <- ggplot(typeoftravelmean, aes(x=TypeofTravel, y=Satisfaction)) + geom\_col()

typeoftravel <- typeoftravel + ggtitle("Bar chart of customer satisfaction per Type of travel")

typeoftravel <- typeoftravel + theme(axis.text.x = element\_text(angle = 90, hjust = 1))

typeoftravel <- typeoftravel + geom\_bar(stat = "identity", fill ="green")

typeoftravel

########################################################################################################################

shoppingamountatAirport <- ggplot(SoutheastData, aes(x=shoppingatAirport, y=Satisfaction, colour="purple")) + geom\_point()

shoppingamountatAirport

##################################################################################################################

#created a new airlines

NorthwestData <- ProjectData[which(ProjectData$AirlineName == "Northwest Business Airlines Inc. " ),]

View(NorthwestData)

#type of travel for this airlines

#culating mean of satisfaction and grouping by Type of travel

typeoftravelmeann <- NorthwestData %>%

group\_by(TypeofTravel) %>%

summarize(Satisfaction = mean(as.numeric(Satisfaction)))

#plotting graph of satisfaction vs destination state

typeoftravelmeann<-as.data.frame(typeoftravelmean)

typeoftravelmeann

typeoftravell <- ggplot(typeoftravelmeann, aes(x=TypeofTravel, y=Satisfaction)) + geom\_col()

typeoftravell <- typeoftravell + ggtitle("Bar chart of customer satisfaction per Type of travel for Northwest")

typeoftravell <- typeoftravell + theme(axis.text.x = element\_text(angle = 90, hjust = 1))

typeoftravell <- typeoftravell + geom\_bar(stat = "identity", fill ="purple")

typeoftravell

####################################################################################################

originstatemeann <- NorthwestData %>%

group\_by(OriginState) %>%

summarize(Satisfaction = mean(as.numeric(Satisfaction)))

#Plotting graph for satisfaction vs origin state

NorthwestData$Satisfaction <- as.numeric(as.character(NorthwestData$Satisfaction))

originstatemeann<-as.data.frame(originstatemeann)

originstatemeann

originStatee <- ggplot(originstatemeann, aes(x=OriginState, y=Satisfaction)) + geom\_col()

originStatee <- originStatee + ggtitle("Bar chart of customer satisfaction per Originstate for Northwest")

originStatee <- originStatee + theme(axis.text.x = element\_text(angle = 90, hjust = 1))

originStatee <- originStatee + geom\_bar(stat = "identity", fill ="Purple")

originStatee

########################################################################################################################

EatandDrinkmeann <- NorthwestData %>%

group\_by(EatingAndDrinkingAtAirport) %>%

summarize(Satisfaction = mean(as.numeric(Satisfaction)))

#plotting graph of satisfaction vs destination state

EatandDrinkmeann <-as.data.frame(EatandDrinkmeann)

EatandDrinkmeann

EatingandDrinkingg <- ggplot(EatandDrinkmeann, aes(x=EatingAndDrinkingAtAirport, y=Satisfaction)) + geom\_col()

EatingandDrinkingg <- EatingandDrinkingg + ggtitle("Bar chart of customer satisfaction per Eating and Drinking at Airport for Northwest")

EatingandDrinkingg <- EatingandDrinkingg + theme(axis.text.x = element\_text(angle = 90, hjust = 1))

EatingandDrinkingg <- EatingandDrinkingg + geom\_bar(stat = "identity", fill ="green")

EatingandDrinkingg

########################################################################################################################

#culating mean of satisfaction and grouping by class

classmeann <- NorthwestData %>%

group\_by(Class) %>%

summarize(Satisfaction = mean(as.numeric(Satisfaction)))

#plotting graph of satisfaction vs destination state

classmeann<-as.data.frame(classmeann)

classmeann

classess <- ggplot(classmean, aes(x=Class, y=Satisfaction)) + geom\_col()

classess <- classess + ggtitle("Bar chart of customer satisfaction per Class for Northwest")

classess <- classess + theme(axis.text.x = element\_text(angle = 90, hjust = 1))

classess <- classess + geom\_bar(stat = "identity", fill ="brown")

classess

#######################################################################################################################

#culating mean of satisfaction and grouping by Shopping at airport for Northwest

shopatAirportmeann <- NorthwestData %>%

group\_by(ShoppingAtAirport) %>%

summarize(Satisfaction = mean(as.numeric(Satisfaction)))

#plotting graph of satisfactionstate

shopatAirportmeann <-as.data.frame(shopatAirportmeann)

shopatAirportmeann

shoppingatAirportt <- ggplot(shopatAirportmeann, aes(x=ShoppingAtAirport, y=Satisfaction)) + geom\_col()

shoppingatAirportt <- shoppingatAirport + ggtitle("Bar chart of customer satisfaction per Shopping at Airport for Northwest")

shoppingatAirportt <- shoppingatAirportt + theme(axis.text.x = element\_text(angle = 90, hjust = 1))

shoppingatAirportt <- shoppingatAirportt + geom\_bar(stat = "identity", fill ="red")

shoppingatAirportt

###################################################################################

destinationstatemeann <- NorthwestData %>%

group\_by(DestinationState) %>%

summarize(Satisfaction = mean(as.numeric(Satisfaction)))

#plotting graph of satisfaction vs destination state

destinationstatemeann<-as.data.frame(destinationstatemeann)

destinationstatemeann

destinationStatee <- ggplot(destinationstatemeann, aes(x=DestinationState, y=Satisfaction)) + geom\_col()

destinationStatee <- destinationStatee + ggtitle("Barchart of customer satisfaction per Destinationstate for Northwest")

destinationStatee <- destinationStatee + theme(axis.text.x = element\_text(angle = 90, hjust = 1))

destinationStatee <- destinationStatee + geom\_bar(stat = "identity", fill ="pink")

destinationStatee

Thank You