IST 687: INTRODUCTION TO DATA SCIENCE- FINAL PROJECT REPORT

IMPROVING THE CUSTOMER SATISFACTION OF AIRLINES IN THE USA

PROJECT BY:

AMBARISH PATANKAR

**Content**

**1)** [**Introduction**](file:///C:\Users\ambar\AppData\Local\Temp\Temp1_Project%20Docs%20(All).zip\WORD\SIYUN-687%20project_Word.docx#_heading=h.uivutar482ok)3

**2)** [**Business Questions**](file:///C:\Users\ambar\AppData\Local\Temp\Temp1_Project%20Docs%20(All).zip\WORD\SIYUN-687%20project_Word.docx#_heading=h.q4x8y26v2jr6)

**3)** [**Data Acquisition, Cleaning, Transformation, Munging**](file:///C:\Users\ambar\AppData\Local\Temp\Temp1_Project%20Docs%20(All).zip\WORD\SIYUN-687%20project_Word.docx#_heading=h.yr5ds6wvaty7)5

[3.1 Data Extraction](file:///C:\Users\ambar\AppData\Local\Temp\Temp1_Project%20Docs%20(All).zip\WORD\SIYUN-687%20project_Word.docx#_heading=h.e74zpxosiijt) 6

[3.2 Data](file:///C:\Users\ambar\AppData\Local\Temp\Temp1_Project%20Docs%20(All).zip\WORD\SIYUN-687%20project_Word.docx#_heading=h.h7y6kutjzre4) Cleaning 6

[3.3 Summary](file:///C:\Users\ambar\AppData\Local\Temp\Temp1_Project%20Docs%20(All).zip\WORD\SIYUN-687%20project_Word.docx#_heading=h.kyxbmr8qnhob)  7

**4)** **Descriptive Statistics and Visualization** 8

[4.1 Visualization](file:///C:\Users\ambar\AppData\Local\Temp\Temp1_Project%20Docs%20(All).zip\WORD\SIYUN-687%20project_Word.docx#_heading=h.h8pnbp37nblc)s 8

**5)** **Modeling Techniques and Visualization**

5.1 Linear Modelling

5.2 Insights generated from Linear Modeling

5.3 Association Rules Mining

5.4 Insights generated from Association Rules Mining

5.5 Support Vector Machine (SVM)

5.6 Insights generated from SVM

**6) Map Route Visualizations of Low Satisfaction Airlines**

6.1 Visualizations

6.2 Insights

**7)** **Recommendations to Improve Marketing Plans**

7.1 Marketing Segments and Actionable Insights

**8) References**

**1) Introduction:**

The United States of America is a huge country, and most people take flights to travel between states. Although driving from one state to another is feasible, people usually look to save time and take flights. Hence, many airlines go head-to-head in order to gain more customers and are always looking out to improve their products and services like loyalty, reliability, ease of use, etc. Southeast Airlines can be the top dogs in the market by improving the experience of their customers and by reducing customer churn.

In order to get an idea of experience of the customers, the loyalty program was introduced. However, the loyalty program did not work, as it was not enough in keeping low customer churn. In this project, I will extract attributes which are of use to me and generate models in order to generate insights and provide recommendations to the airlines.

**2) Business Questions:**

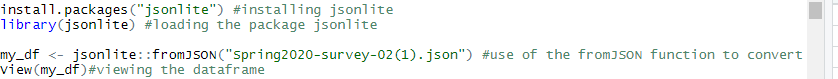
The business questions which I have answered in this project are as follows:

1. Which airlines have the lowest and highest levels of customer satisfaction?
2. Why I chose FlyFast Airways Inc. and Northwest Business Airlines Inc. for my analysis?
3. How old are people who have been taking the flights?
4. How is age affecting the likelihood of a customer recommending a specific airline to another customer?
5. Which age group is likely to recommend an airline?
6. Are males recommending the airlines more or are females?
7. How is price sensitivity affecting the satisfaction of a customer?
8. Is customer satisfaction affected by the type of travel of the customers?
9. Are customers with a certain airline status more likely to recommend an airline?
10. Are loyal customers most likely to recommend an airline?

**3)** [**Data Acquisition, Cleaning, Transformation, Munging**](file:///C:\Users\ambar\AppData\Local\Temp\Temp1_Project%20Docs%20(All).zip\WORD\SIYUN-687%20project_Word.docx#_heading=h.yr5ds6wvaty7)

3.1: Data Extraction

The first step of the project is to extract data from a JSON document. Initially, I downloaded the data and it consisted of 10282 different observations of 32 variables.



I ran the ‘jsonlite’ package and stored the data in a different dataframe ‘my\_d’, using the fromJSON. I went through all of the columns in the data and found out that the freeText column had a lot of NA values.

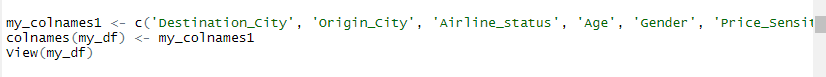
3.2: Data Cleaning



I checked the NA values for each column using the sapply() function. Further, I checked the values of the freeText column, in order to decide what I have to do with this column. I went ahead and deleted this column, as I think that this column would not help me in this analysis.



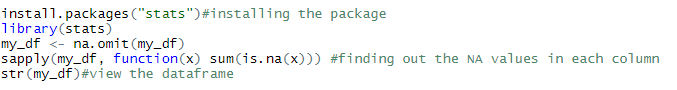
After deleting the freeText column, I decided to change the all the column names, for my own analysis



Next, I decided to eliminate all the NA values from the numeric variables present in the dataset. I did this by replacing the NAs with the mean values of each variable.



I used the sapply() function again to check the NA values in the numeric variables. There were no NA values in the numeric variables, but the non-numeric variables still had NA values. I used the na.omit function to remove these NA values, and then checked the NA values again.



3.3: Summary

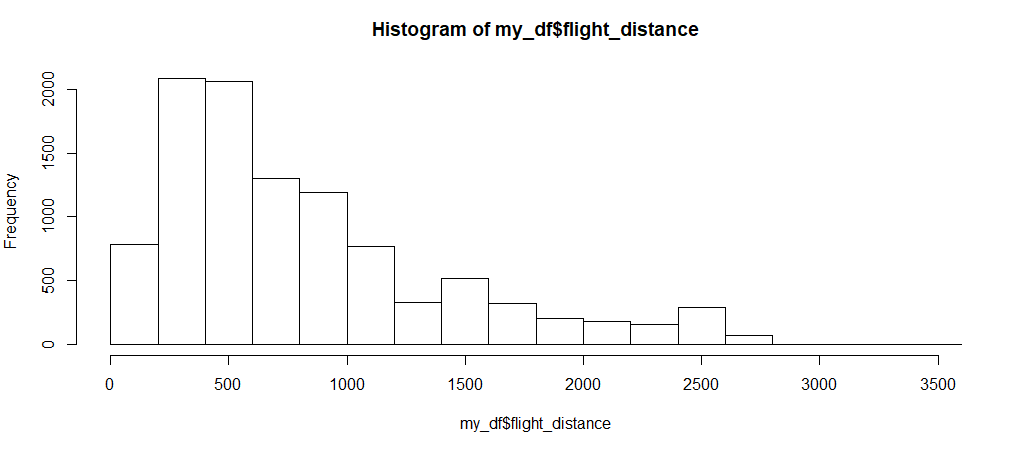
After understanding the data properly, I realized that the freeText column was not of any use for my analysis, so I deleted it. I changed the column names in order to get clarity and use these in my analysis. The ‘Likelihood to Recommend’ is the most important variable in this analysis. I ran the table() function for all the variables in order to understand the data better, and how I was going to go ahead with my analysis. Even though it would have been okay to keep some of the NA values, I decided to delete them in order to glean accurate insights.

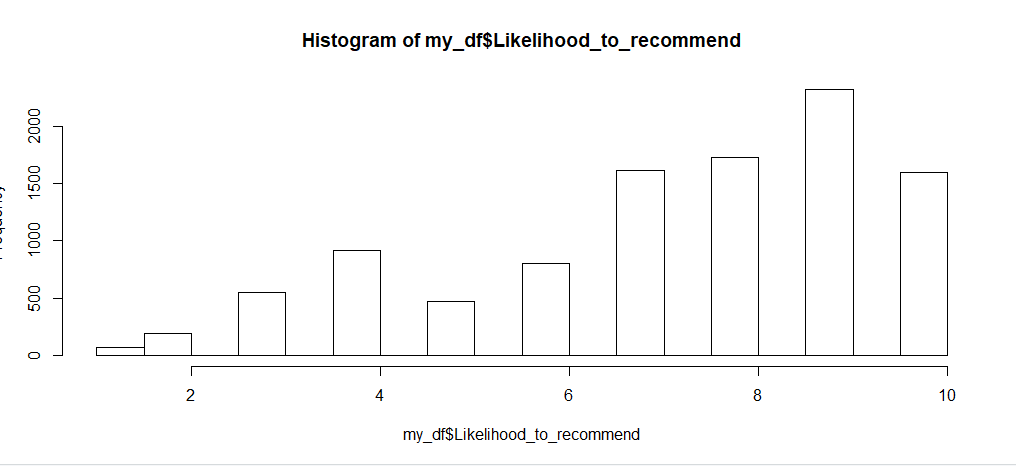
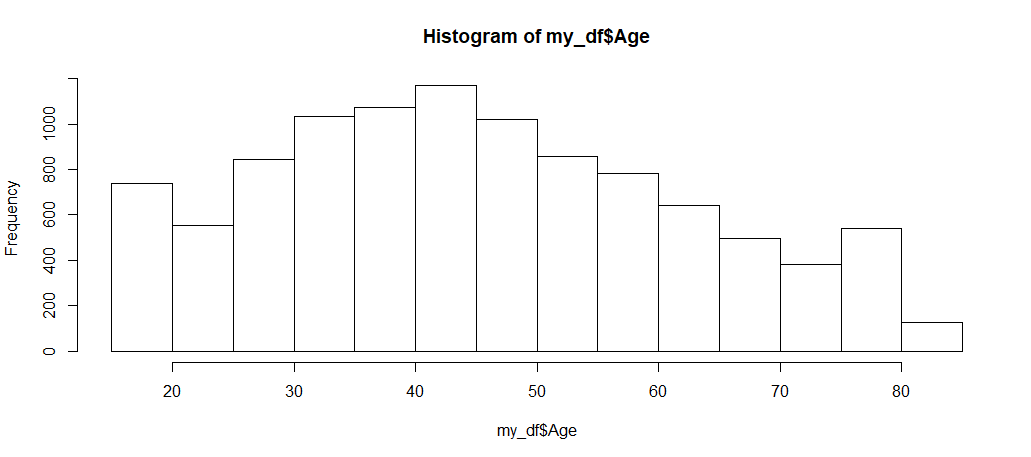
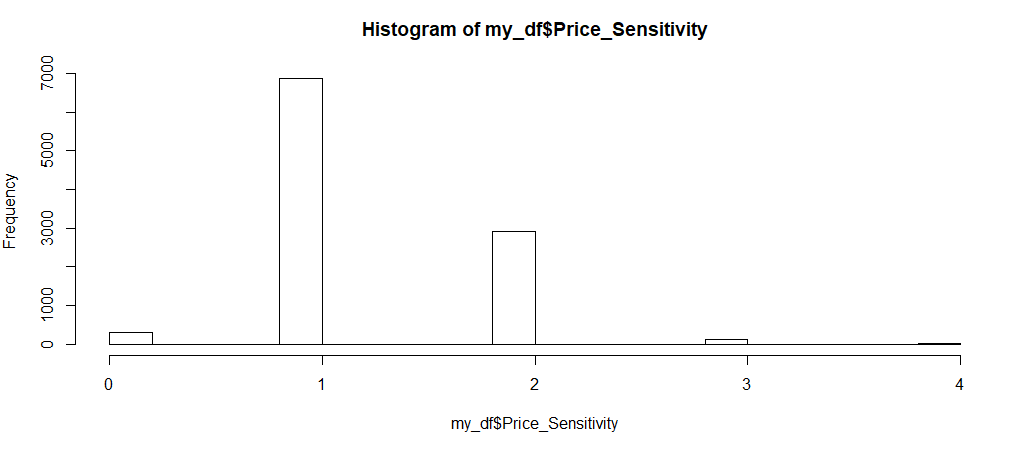
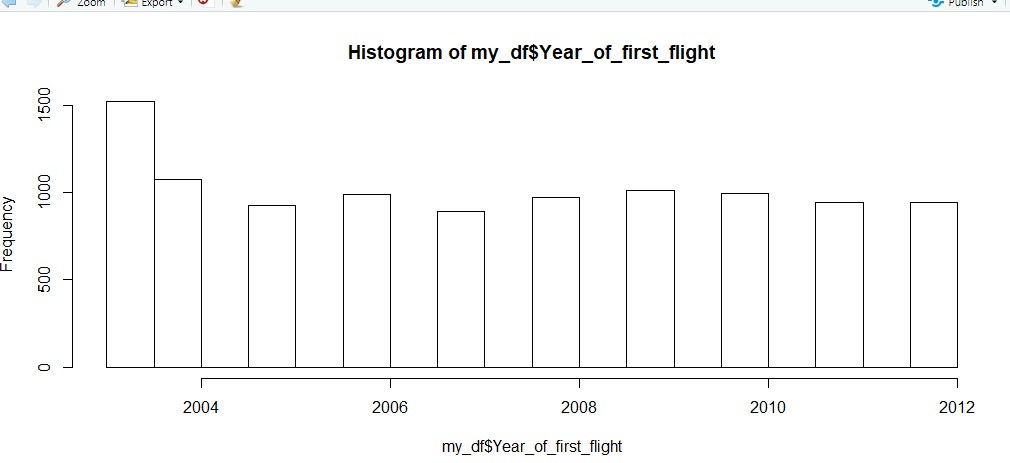
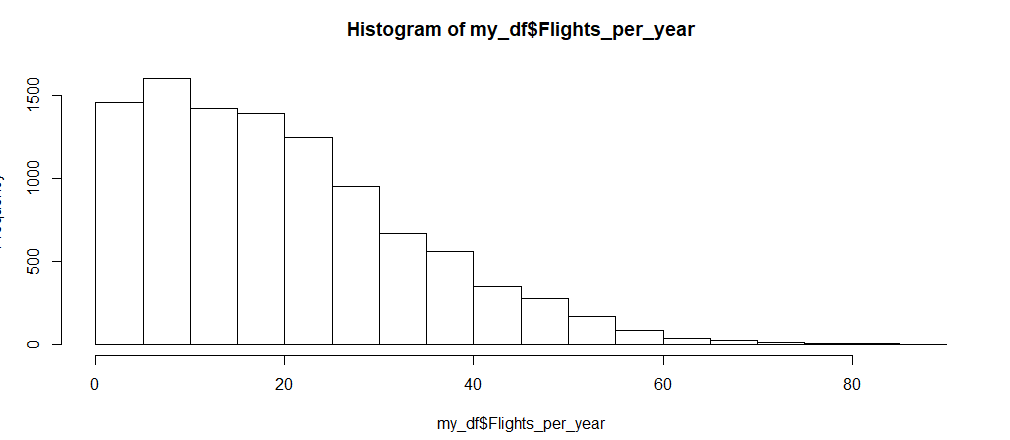
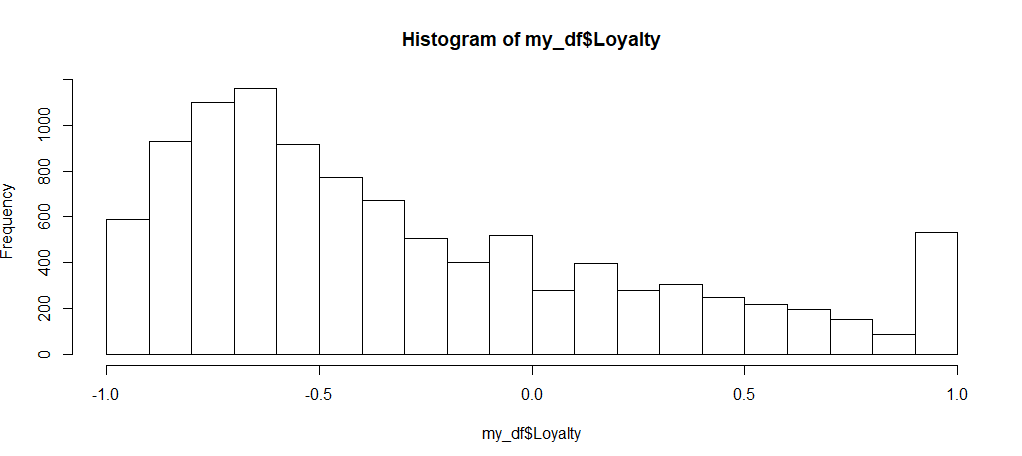
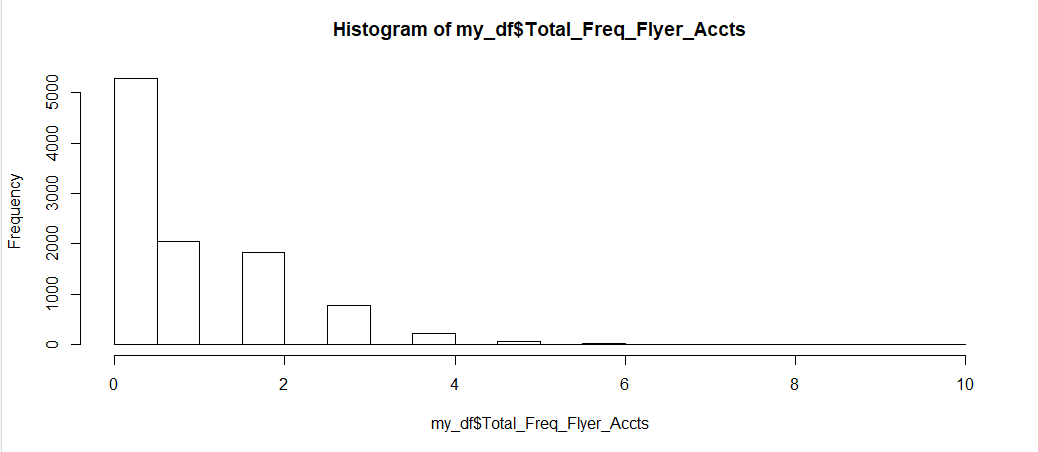
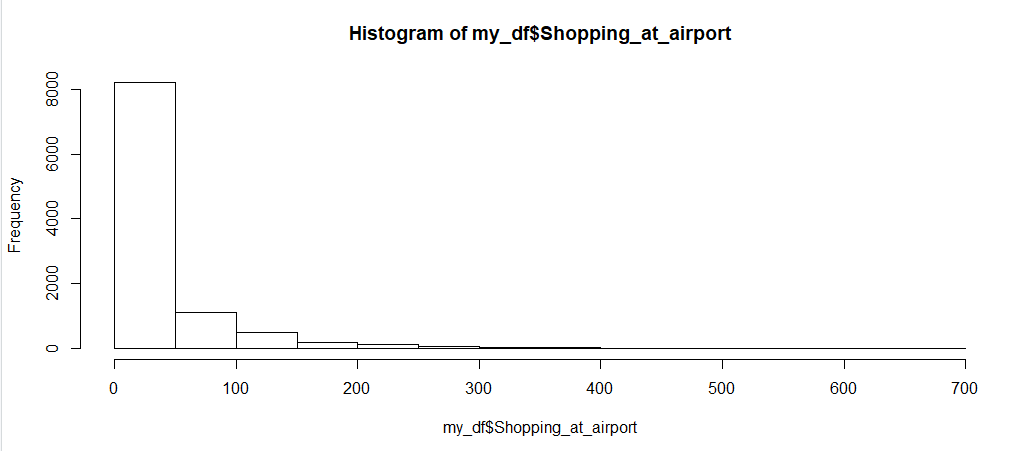
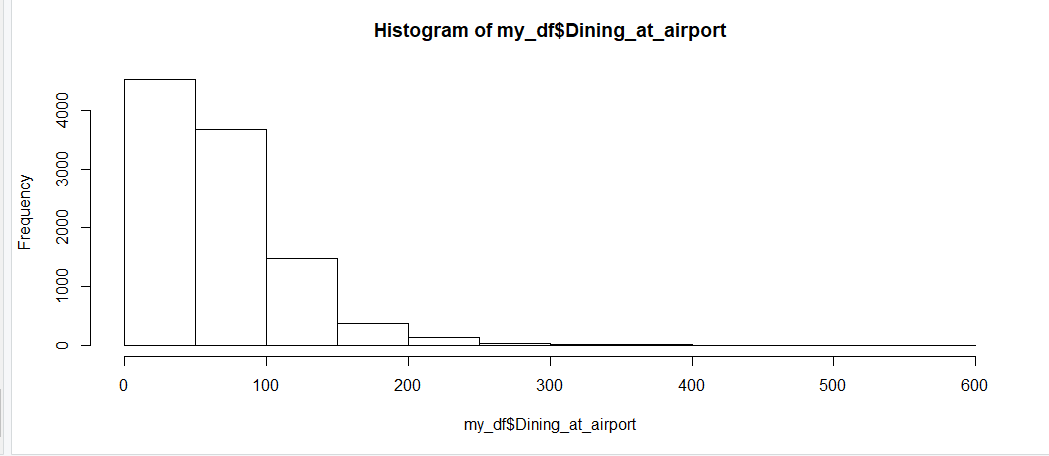
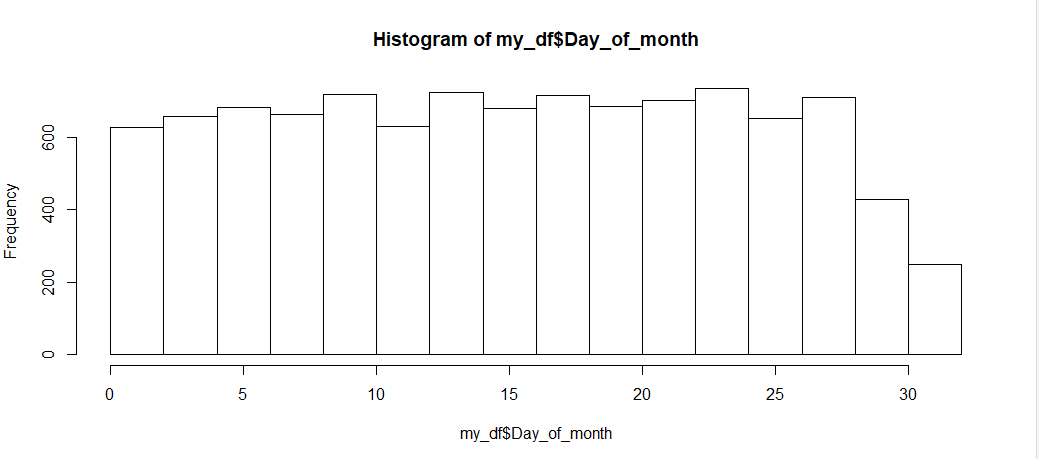
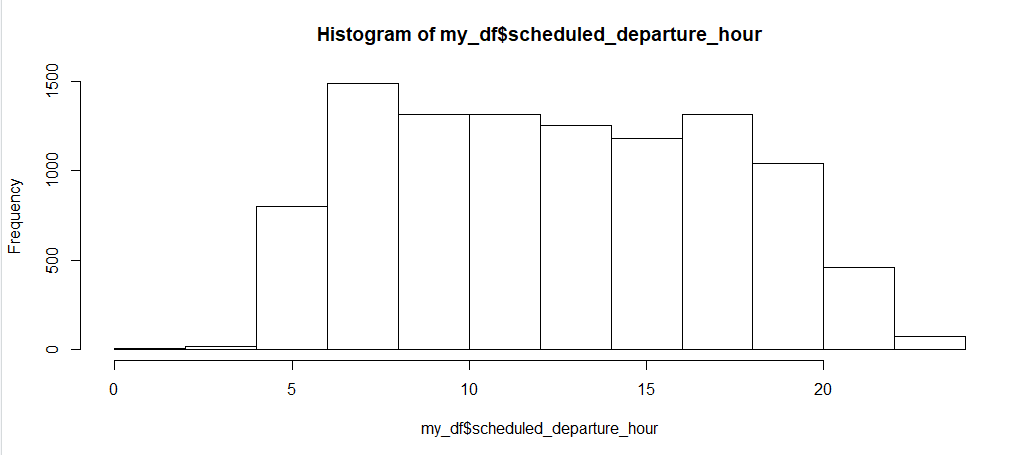
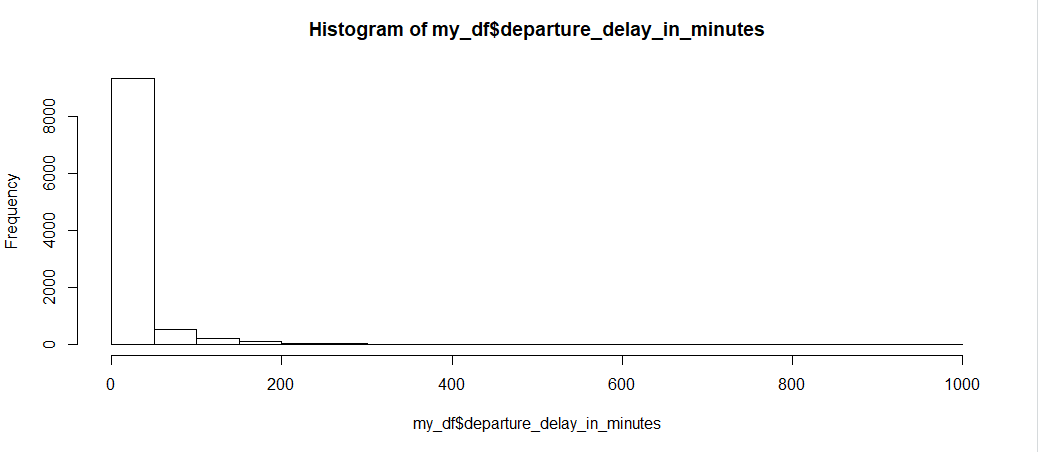
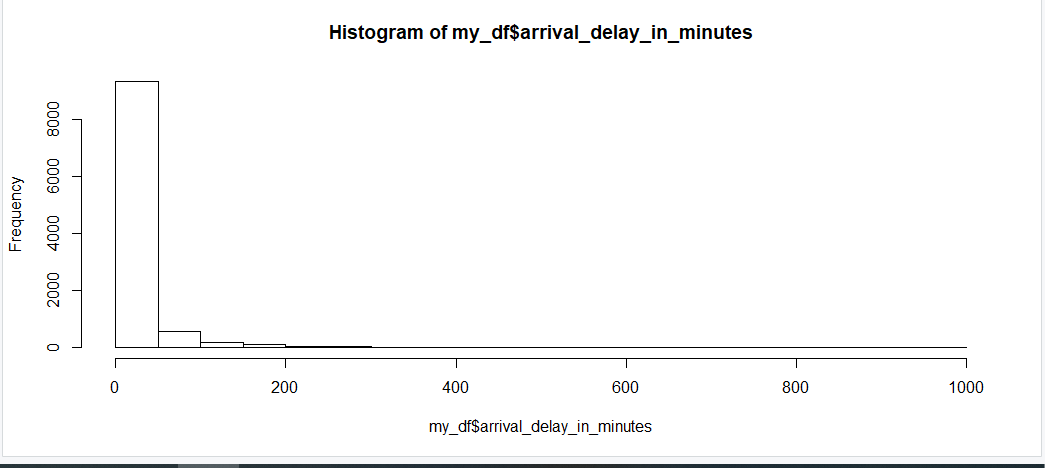
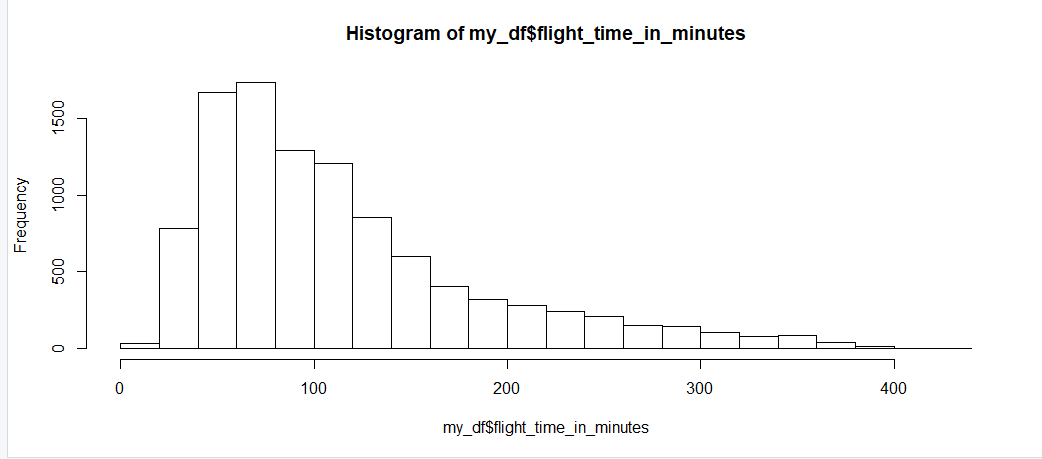
**4)** **Descriptive Statistics and Visualization**

4.1 Visualizations

As the customers did not value the Loyalty program much, it was not enough to keep the low customer churn. ‘Likelihood to Recommend’ is a good indicator to understand the customer churn, but the Net Promoter Score is three times more sensitive at predicting the customer churn. While conducting the descriptive analysis, I have considered multiple variables in comparison to the ‘Likelihood to Recommend’. I want to test the relationship between these variables and ‘Likelihood to recommend’.

In the beginning, I started of by simply creating histograms for each numeric variable. I realized that the histograms helped me in understanding the data better, in order to glean insights from it.





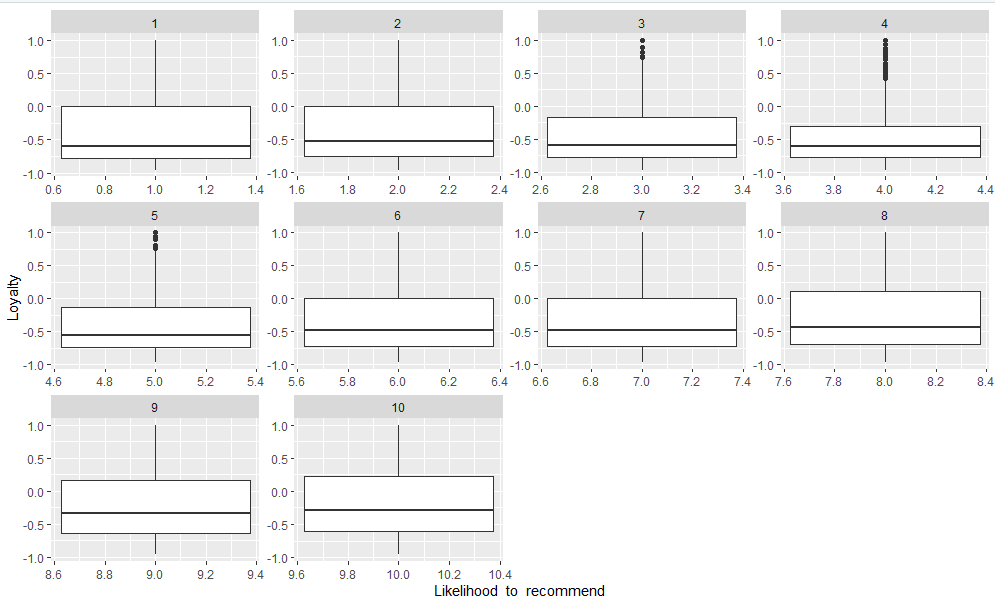
I have generated all of these visualizations in order to understand the shape of each histogram. Most of these histograms are symmetric or have a long right tail. As these histograms have a long right tail, they are positively skewed histogram. I gathered from these visualizations that the variables which have a positively skewed histogram have different mean, median and mode values.

Looking at the histogram of ‘Likelihood to Recommend’, it is negatively skewed, and many customers have given 9 as their ‘Likelihood to Recommend’ score. Similarly, the average of people taking these airlines is about 45. Considering the frequent flyer accounts, most of the customers don’t have a frequent flyer account, maybe this is because of the loyalty of the customers.

In order to understand the data better and make more comparisons, I have generated many more visualizations. Firstly, I have created a boxplot in order to understand loyalty and Likelihood to Recommend better.



I used the facet\_wrap function to generate some facet panels of different range in this distribution and present them in the same page with the order of ascending. The scales are set as free, which means that each small graph is free to adjust the coordinate scale range according to its own data range.

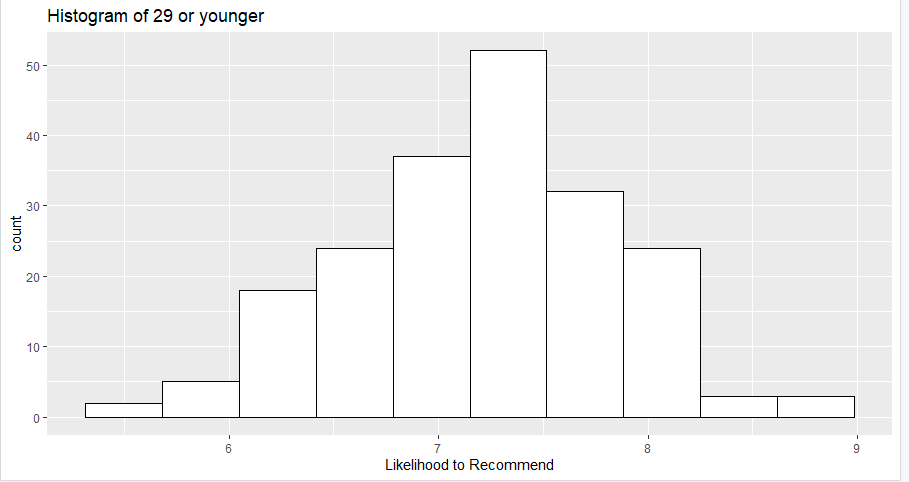


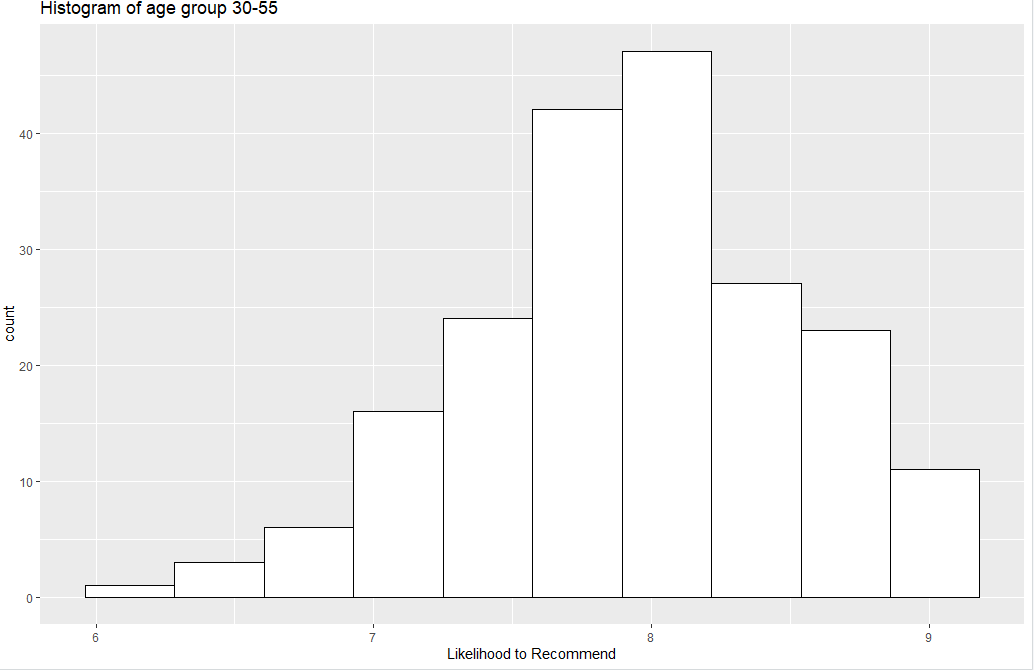
I have generated a boxplot for all values of Likelihood to Recommend score (1-10), in comparison to the loyalty. We see that a customer is unlikely to recommend an airline for a loyalty score which is below -0.5. As the loyalty score goes above -0.5 going more towards 0, the customer is likely to recommend an airline. We can also see that for values 3,4 and 5 of Likelihood to Recommend, there are some outliers.

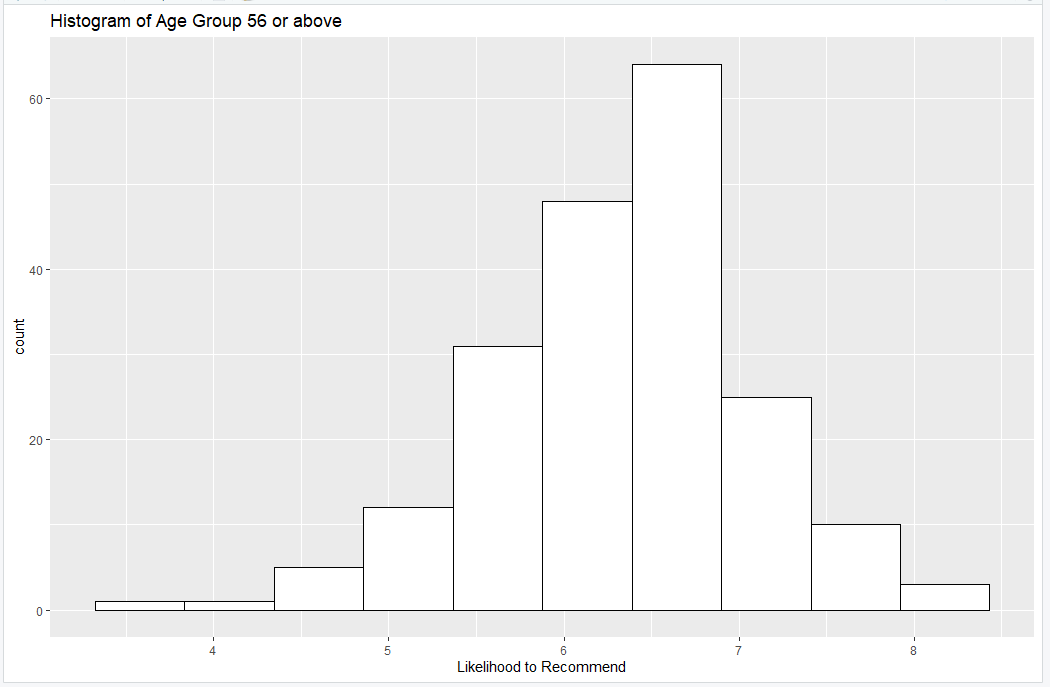
Further, let’s take a look at how age is affecting the likelihood to recommend of a customer. We already know that average age of recommending an airline is about 45.

I have created different age groups, which are

1. 29 or younger
2. 30-55
3. 56 and above

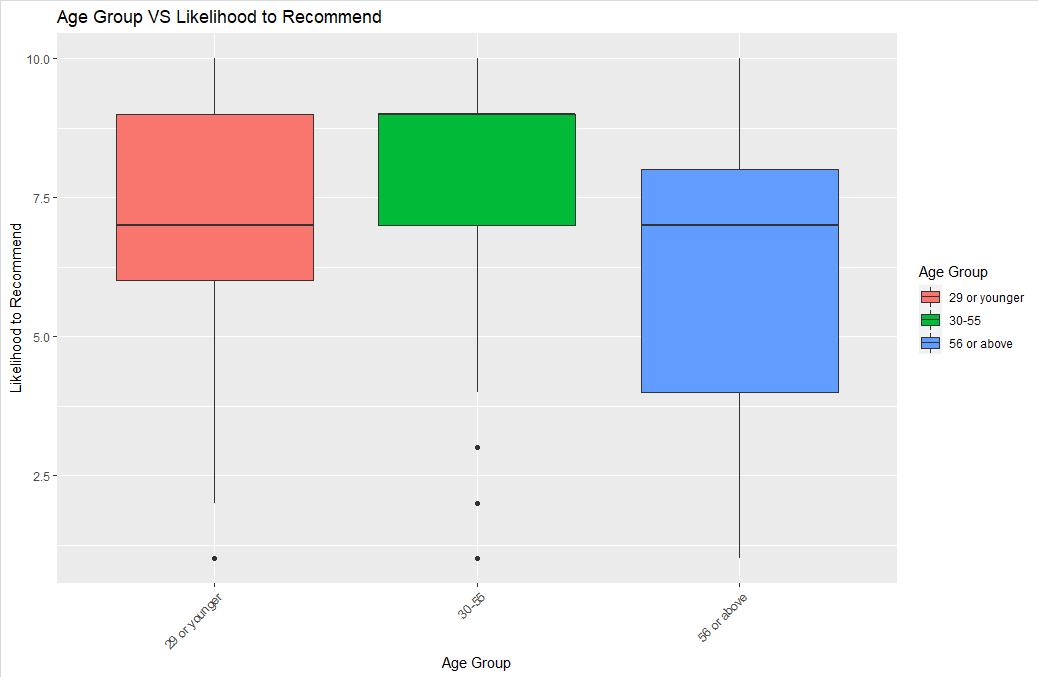






From the histograms above, we can see that the age group of 29 or younger is likely to recommend an airline between a value of 7 and 8, this shows that the first age group is less likely to recommend an airline as compared to the second group. The age group of 30-55 can be said to be ‘Promoters’ of airlines, as this age group is most likely to recommend an airline. The second age group has likelihood to recommend values of 7,8 or even 9 the most. The 3rd age group, doesn’t seem to recommend the airlines much either.

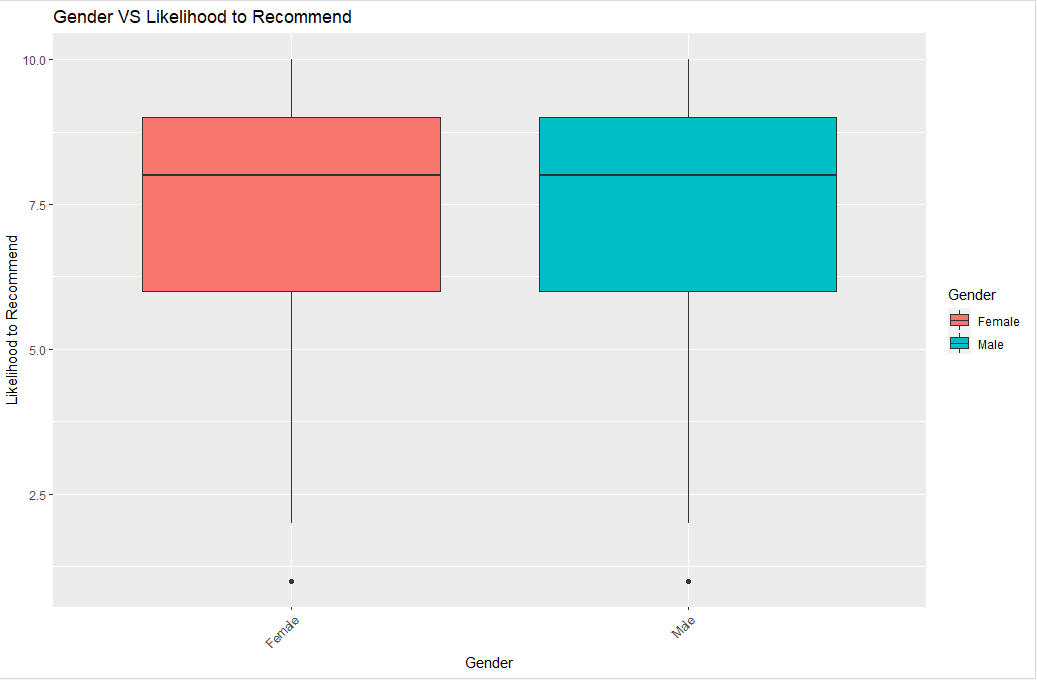
I have also created a boxplot visualization in order to understand the comparison better.



From this boxplot visualization, again we can gather that the 2nd age group is most likely to recommend an airline, and the 3rd age group is the least likely. There are some outliers in the 1st and 2nd age group.

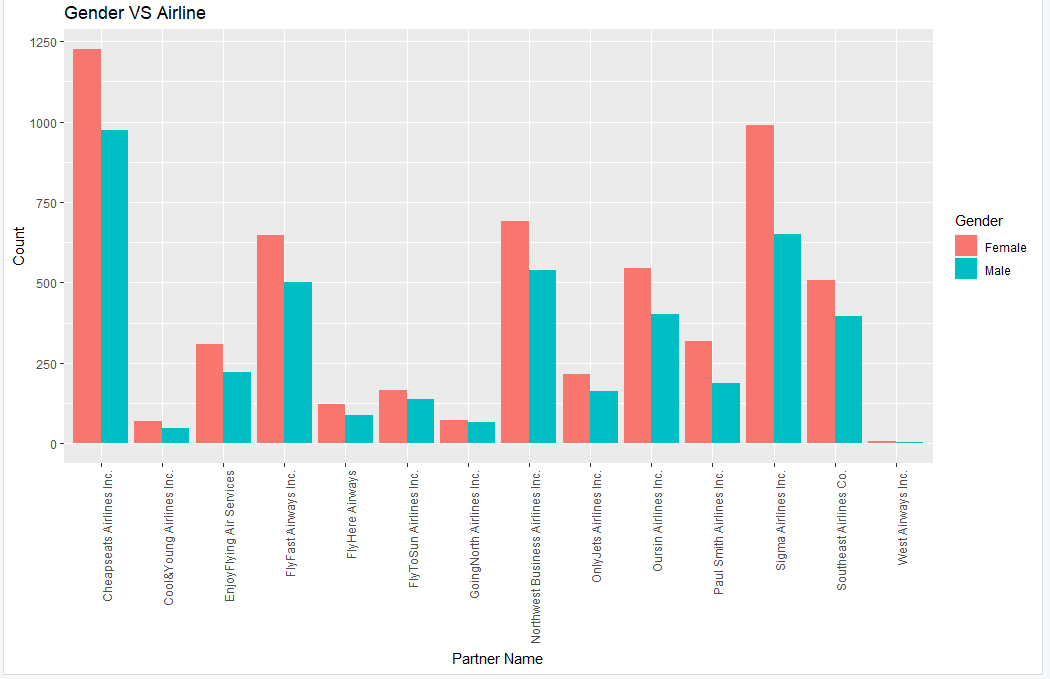
We can gather that the 2nd age group is pretty much satisfied by the service provided by the airlines, but the 3rd age group is not as satisfied. The airlines can look into this, by providing better service to the people who are above the age of 56.

In the next visualization, I have generated a ggplot which shows how gender is affecting the likelihood to recommend of an airline.

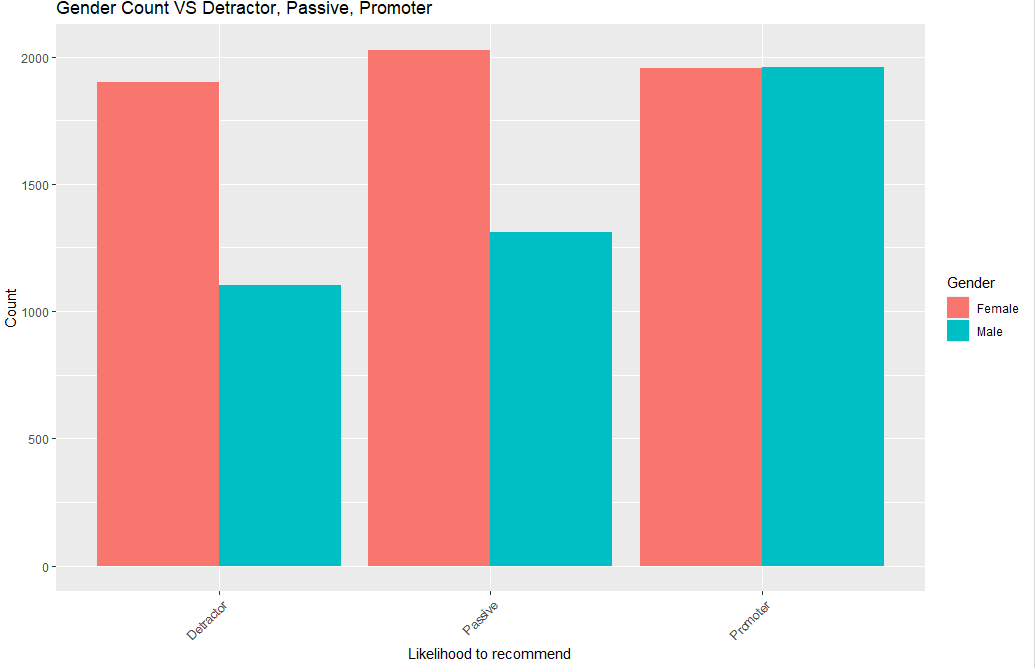


As we can see from this boxplot, there is not much difference between males and females, while recommending an airline. This boxplot gives us some type of results, but doesn’t really explain much taken into consideration that there are 14 different airlines.

In the next visualization, I have created a histogram which compares the gender with each airline.

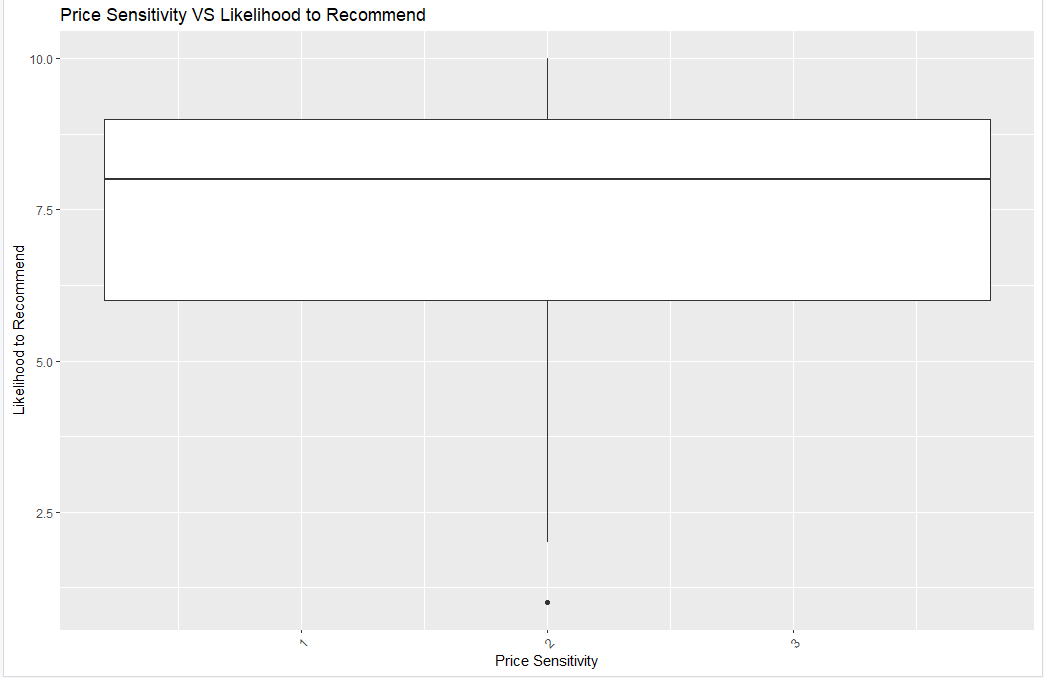


As we can see from the histogram above, ‘Cool & Young Airlines Inc.’ and ‘West Airways Inc.’ are the least recommended. ‘Cheapseats Airlines Inc.’ and ‘Sigma Airlines Inc.’ are highly recommended by Females, males prefer them too.



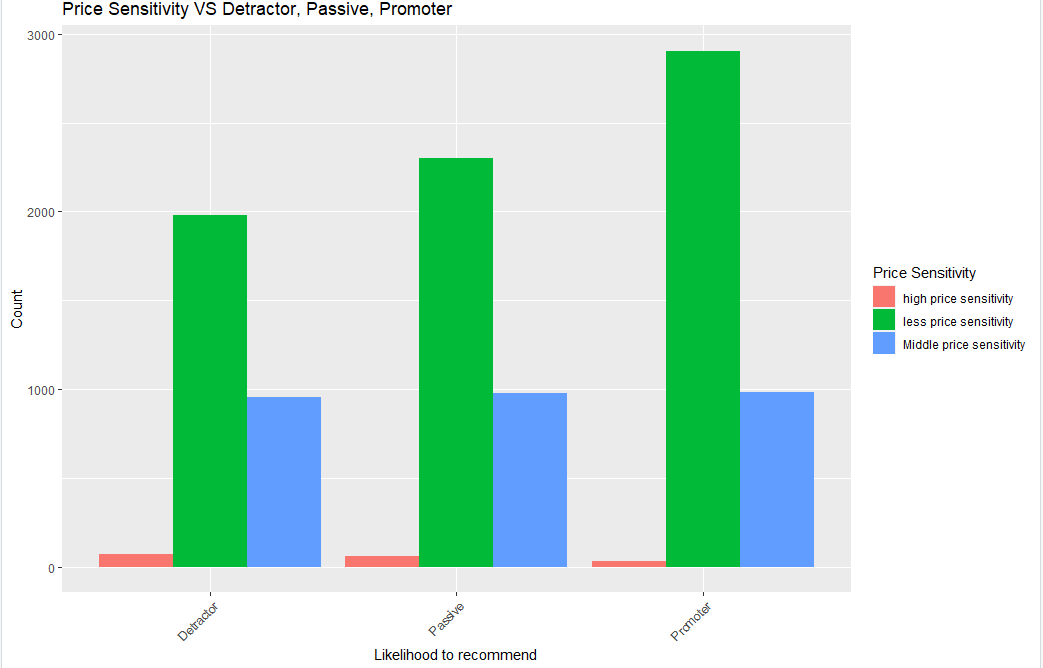
Looking at this visualization, we can say that more females are detractors as compared to the males. Considering for passive, again females are more as compared to the males. Looking at the promoters, both males and females equally promote a specific airline. Thus, from this analysis, we can say that females are not exactly promoting or demoting the airlines, they are passive. The males are actively promoting the airlines. In order to improve the customer experience of females, the airlines can provide extra services to them, like extra leg space on flights or convenient toilets.

The next visualizations display the effect of price sensitivity on likelihood to recommend



Price sensitivity is the grade to which the price affects the likelihood of a customer buying something. It has a range from 0 to 5.

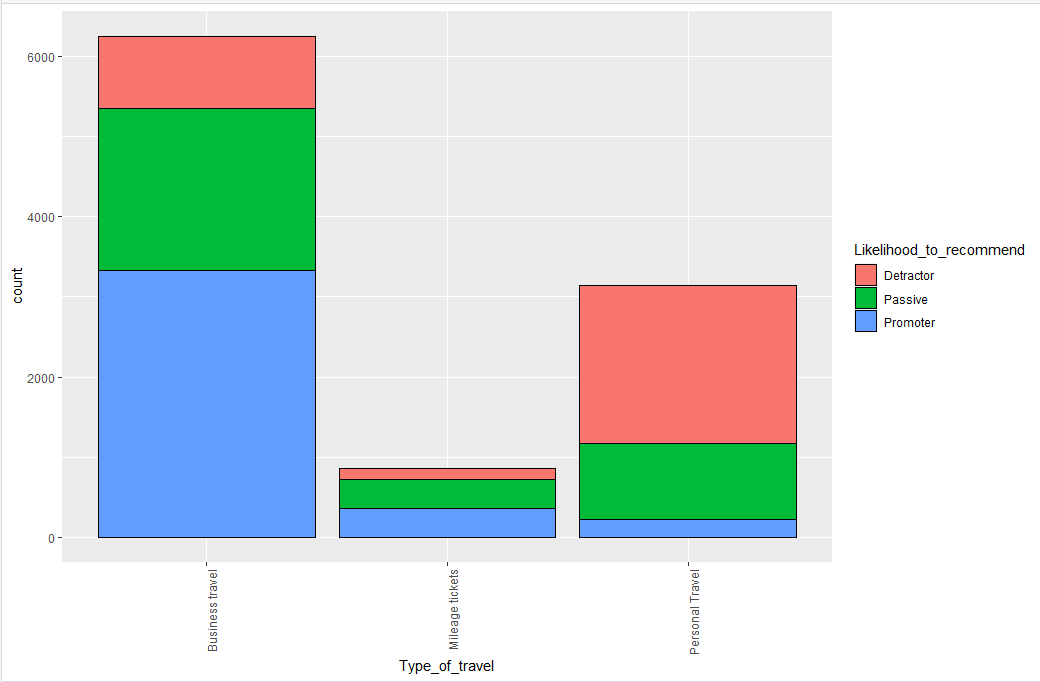
The boxplot above displays the likelihood to recommend at price sensitivity 2. We can see that the major values range from 6 to 9. The median in this case is around 8. There is one outlier in this case. Even though this boxplot gives us some insights, it is not enough to glean results from it. In the next visualization, I have created bins for the Likelihood to Recommend column, where if score is less than 6, then the customer is a detractor, if the score is 7 or 8, then the customer is passive and if the score is 9 or 10, then the customer is a promoter.



As we can see from the visualization above, for high price sensitivity, all three groups of customers are less likely to recommend an airline, whereas, for less price sensitivity, all three groups are more likely to recommend an airline.

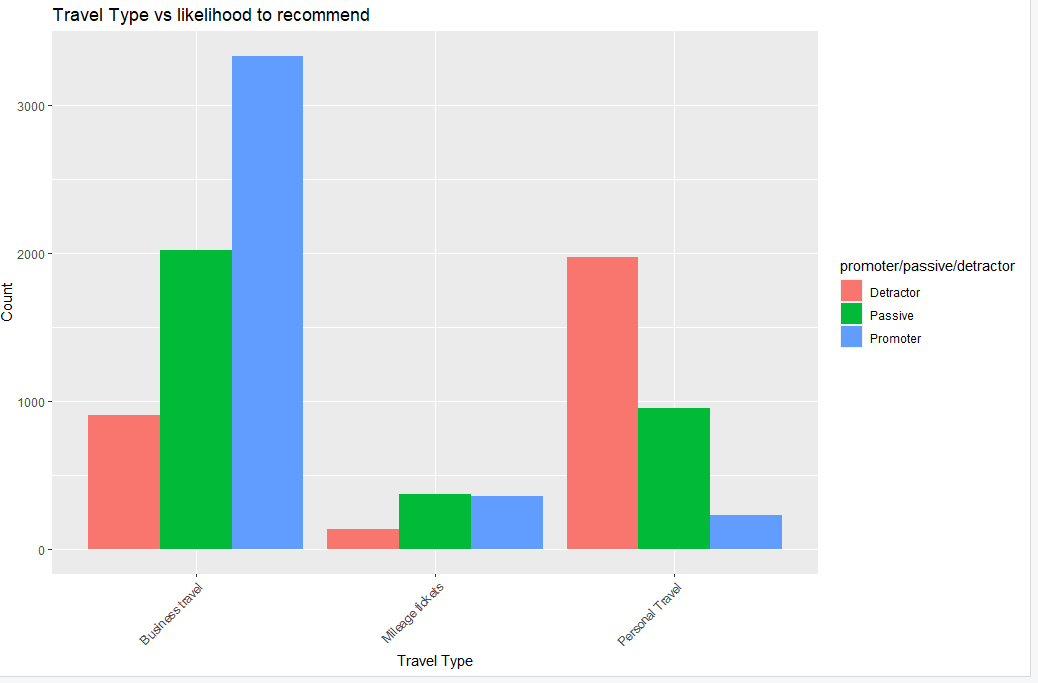
Thus, we can say that the airlines can make changes to their in-air prices, in order to improve their service and attract more customers.

The following visualization shows how the type of travel is affecting the satisfaction of a customer. There are three types of travelers here, Business Travel, Mileage ticket and Personal Travel. I have created bins for the Likelihood to Recommend column, where if score is less than 6, then the customer is a detractor, if the score is 7 or 8, then the customer is passive and if the score is 9 or 10, then the customer is a promoter.



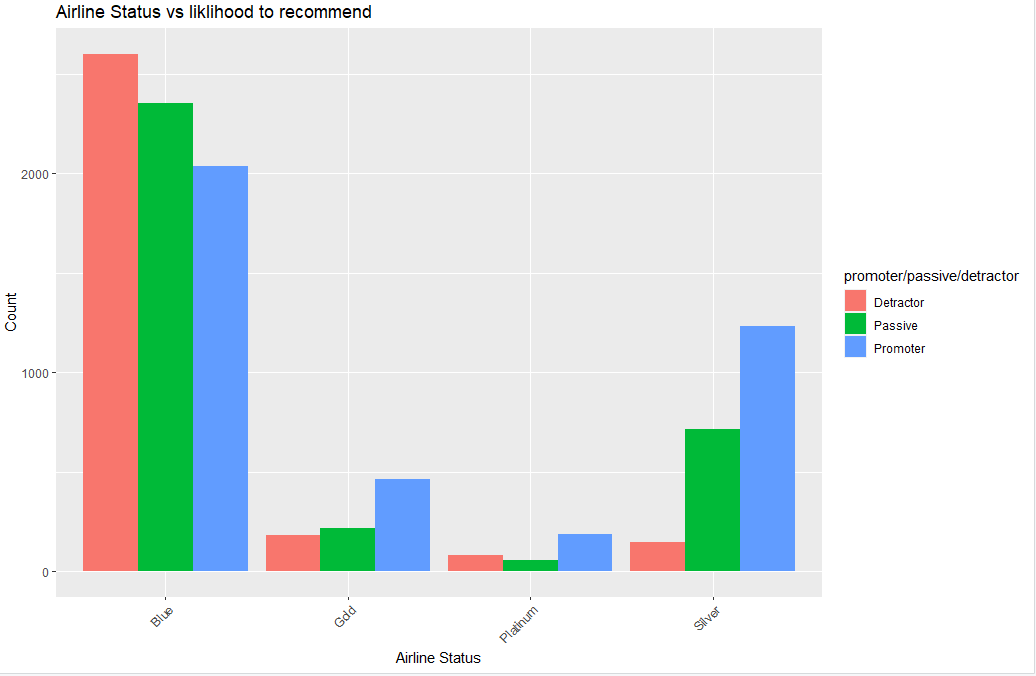
From the histogram above, we can see that the Business travelers are more in number as compared to the Mileage tickets and Personal travelers. Business travelers are promoting the airlines the most. Even though most of the business travelers are promoters, there are customers who are passive too. Most of the personal travelers are detractors, which shows that they are not happy with the customer experience.

I have created another visualization to understand the data better.



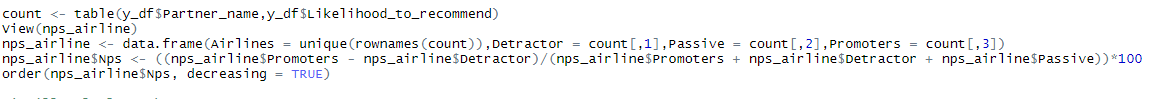
Thus, we can say that the personal travelers are not happy with their flight experience, as a result they are less likely to recommend to anyone else. In order to improve the customer experience for personal travelers, the airlines need to provide the same services as the business travelers to the personal travelers, this can be done by making the personal travelers feel comfortable during the flights. Business travelers are usually habituated to flying, whereas personal travelers are usually tourists.

Now, lets see how the airline flyer status affects the satisfaction of a customer



There are 4 types of airline status’, which are Blue, Gold, Platinum and Silver. The visualization above represents detractors, passive and promoters for each status type. As we can see from the visualization, Blue airline status has more customers as compared to Gold, Platinum and Silver. Even though the number of customers is more in Blue, it has the most detractors and the least number of promoters, which shows that the customers are not satisfied. In Gold, the number of promoters is the most and number of detractors is the least, which means that the customers are happier as compared to Blue. The same case is seen in Platinum and Silver, in fact, Silver seems to have a lot of promoters and very less detractors. In order to improve the experience of the customers in Blue, the airlines need to provide better service to customers, just like in Gold, Platinum and Silver.

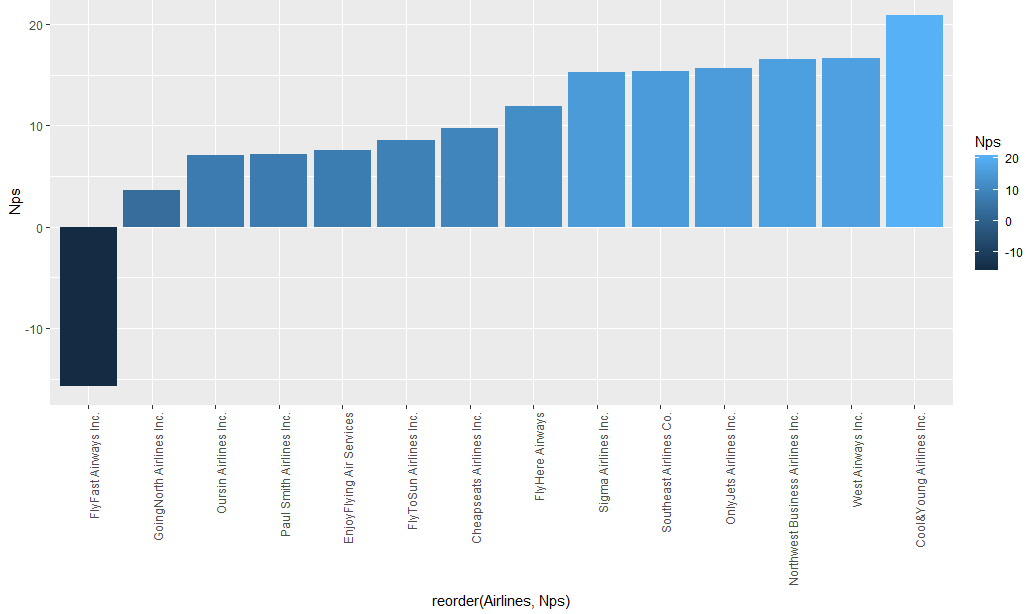
Further, I have calculated the Net Promoter Score for my analysis. , I have created bins for the Likelihood to Recommend column, where if score is less than 6, then the customer is a detractor, if the score is 7 or 8, then the customer is passive and if the score is 9 or 10, then the customer is a promoter. I used the following code to generate the Net Promoter Score.



The visualization is generated by running the following code



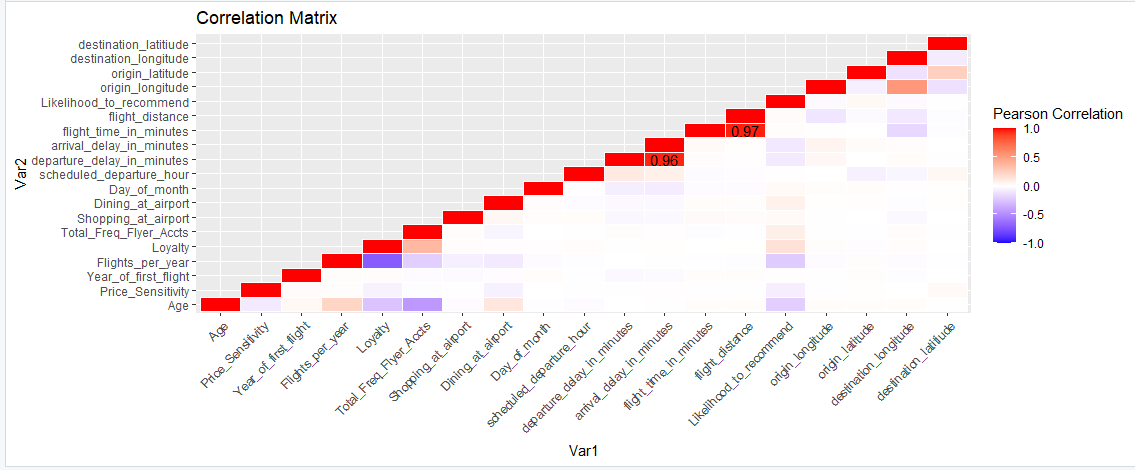
In the next visualization for the nps, I have used a bar chart to show the nps of each airline and used NPS as the color type to set as the range.



The visualization above shows us that ‘FlyFast Airways Inc.’, ‘GoingNorth Airlines Inc.’ and ‘Oursin Airlines Inc.’ have the worst Net Promoter Scores, whereas ‘Northwest Business Airlines Inc.’, ‘West Airways Inc.’ and ‘Cool & Young Airlines Inc.’ have the best NPS. For my map analysis, I have considered ‘FlyFast Airways Inc.’, ‘GoingNorth Airlines Inc.’ and ‘Oursin Airlines Inc.’, in order to understand the map routes of low-level satisfaction airlines. For analysis during modeling techniques, I have considered ‘FlyFast Airways Inc.’ and ‘Northwest Business Airlines Inc.’, as flyfast has the lowest level of satisfaction and Northwest is in the top three net promoter scores.

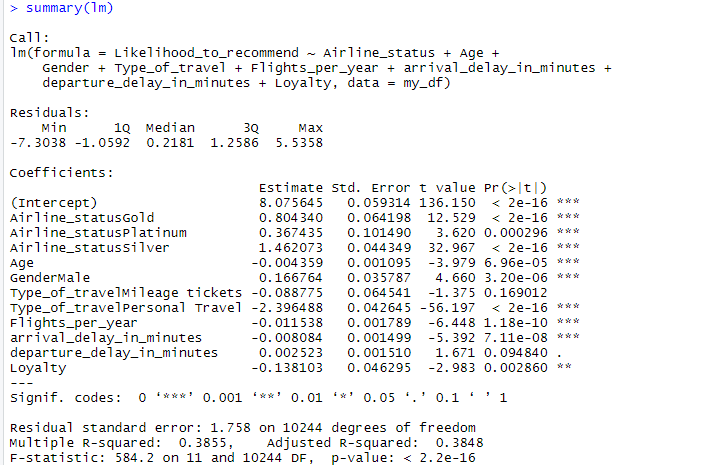
**5)** **Modeling Techniques and Visualization**

5.1: Linear Modeling

A work-horse method used by statisticians to interpret data is linear modeling, also known as linear regression, which is a term covering a wide variety of methods, from the relatively simple to the very sophisticated. The main motive of using linear models is for prediction. There are two types of variables in this modeling technique, independent and dependent variables. The dependent variables can be predicted using independent variables. In my analysis, I have applied linear modeling on the whole data first, and then on Flyfast and Northwest airlines, which I have chosen specifically for my analysis. In all cases, I have considered the Likelihood to recommend as the dependent variable. Firstly, I started by creating a correlation matrix, in order to find out which variables have the strongest correlation. 

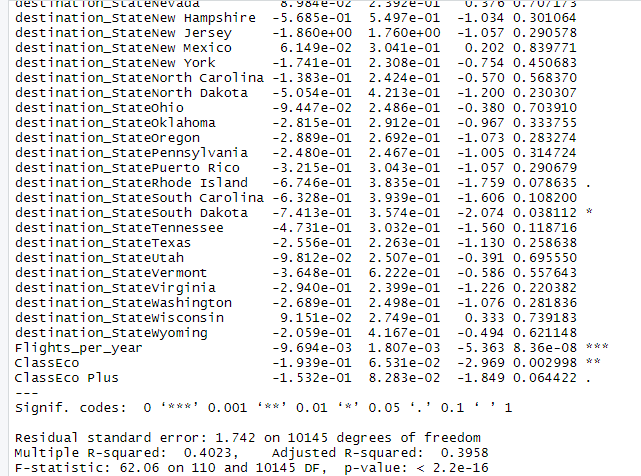
As we can see from the diagram, ‘arrival\_delay\_in\_minutes’ and ‘departure\_delay\_in\_minutes’ have the second strongest correlation, and ‘flight\_distance’ and ‘flight\_time\_in\_minutes’ have the strongest correlation. I chose the independent variables after analyzing the diagram.

After running the liner model for the whole dataset, this is what I got:

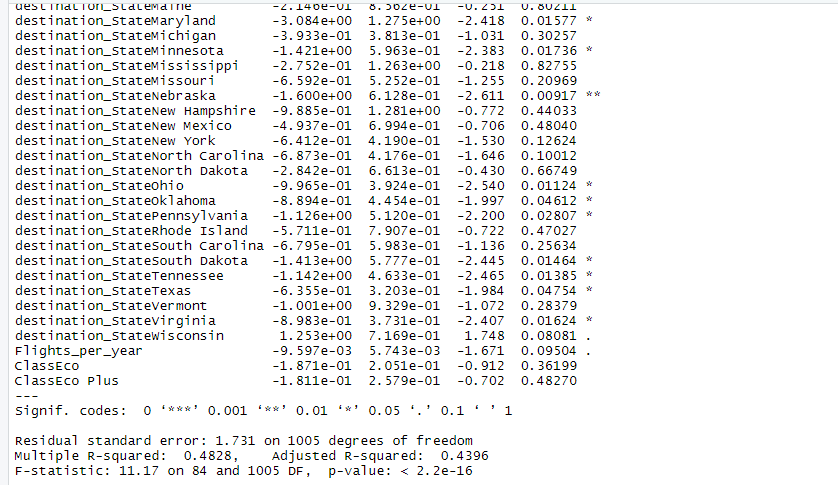


As we can see, we get the multiple r-squared value as 0.38555 and adjusted r-squared value as 0.3848. The adjusted R-squared is 0.3848, which means that the Likelihood to recommend variable accounts for about 38% of the variability. Here, we cannot determine whether a r-squared value is bad or good. We see that the p-value of the equation is less than 0.05, which means that equation is significant.

I have also applied the linear models to flyfast and northwest airlines.



This analysis is done for the northwest airlines. As we can see from the diagram above, the multiple r-squared value is 0.4023 and the adjusted r-squared value is 0.3958. The adjusted R-squared is 0.3958, which means that the Likelihood to recommend variable accounts for about 39.58% of the variability for northwest airlines. The p-value is less than 0.05, so the equation is significant. Now, let’s see what I got for flyfast airlines.



For flyfast airlines, the r-squared values have increased. The adjusted R-squared is 0.4396, which means that the Likelihood to recommend variable accounts for about 43.96% of the variability for flyfast airlines. The equation is significant in this case too

5.2: Insights generated from Linear Modeling:

I found out that the adjusted r-squared values for all of the models mentioned above between a range of 0.3 to 0.5, which can be good or bad, because the r-squared value doesn’t have good or bad values. This specifies that the likelihood to recommend can be predicted between a range of 30% to 50%, according to my models. All the equations are below the value of 0.05, which means all of them are significant. Even though I was able to glean some insights from linear modeling, I don’t think it was the best method to go ahead. Further in the project, I have generated better insights using Association rules mining and support vector machine.

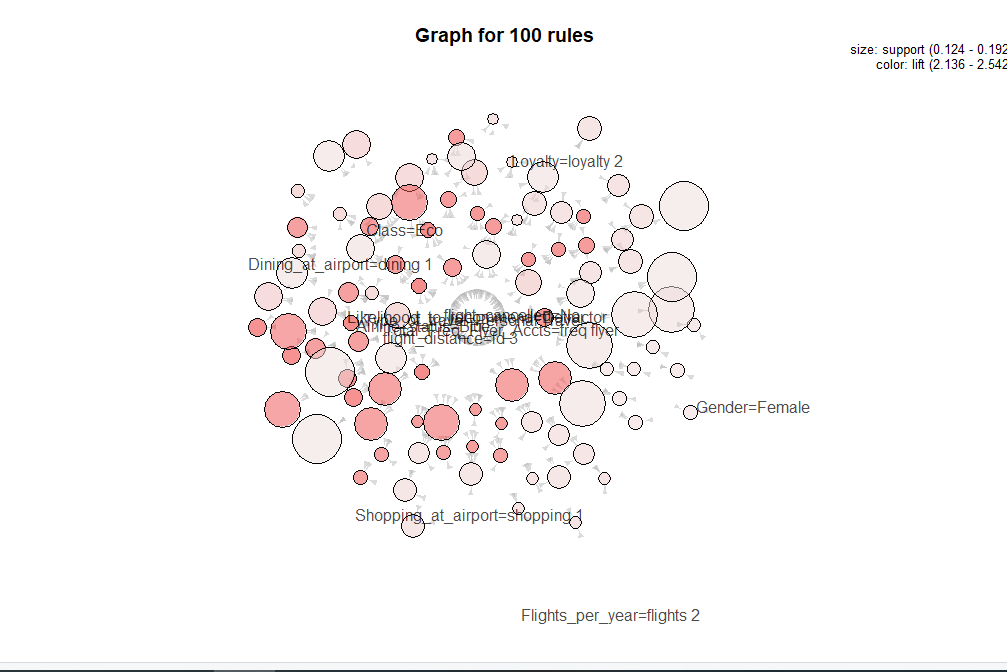
5.3: Association Rules Mining

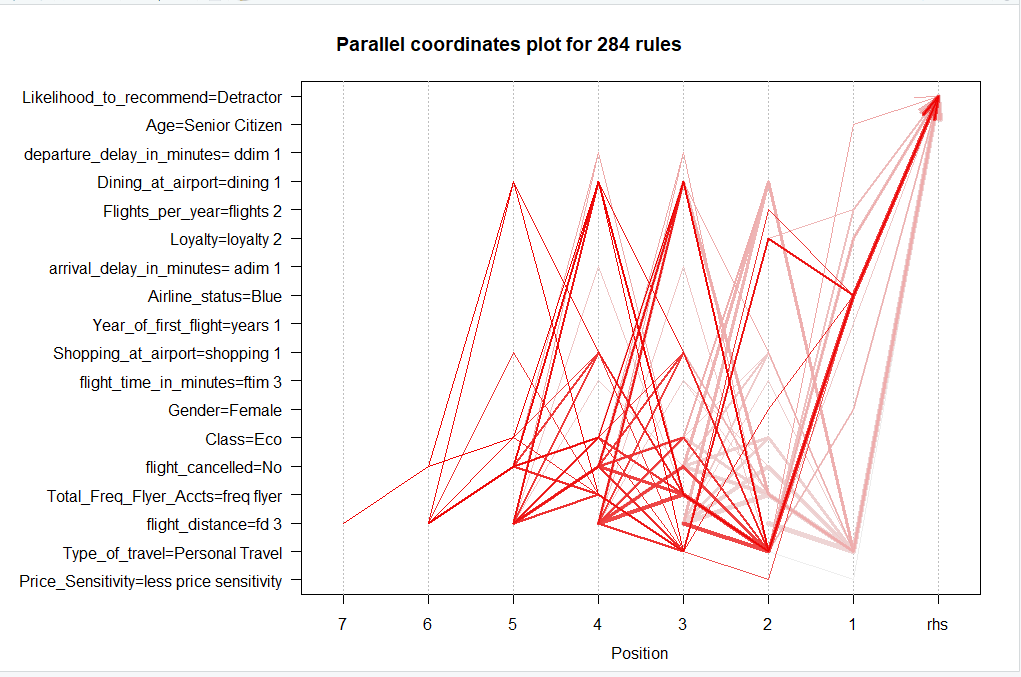
Association rule learning is a rule-based machine learning method for discovering interesting relations between variables in large databases. It is intended to identify strong rules discovered in databases using some measures of interestingness. As I mentioned before, I was not able to gain the best insights from linear modeling. For association rules mining, I have followed the same approach, where I have applied the association rules on the whole dataset first, and then on flyfast and northwest airlines. In order to apply association rules, I decided to divide all numeric variables into bins, except latitude and longitude.

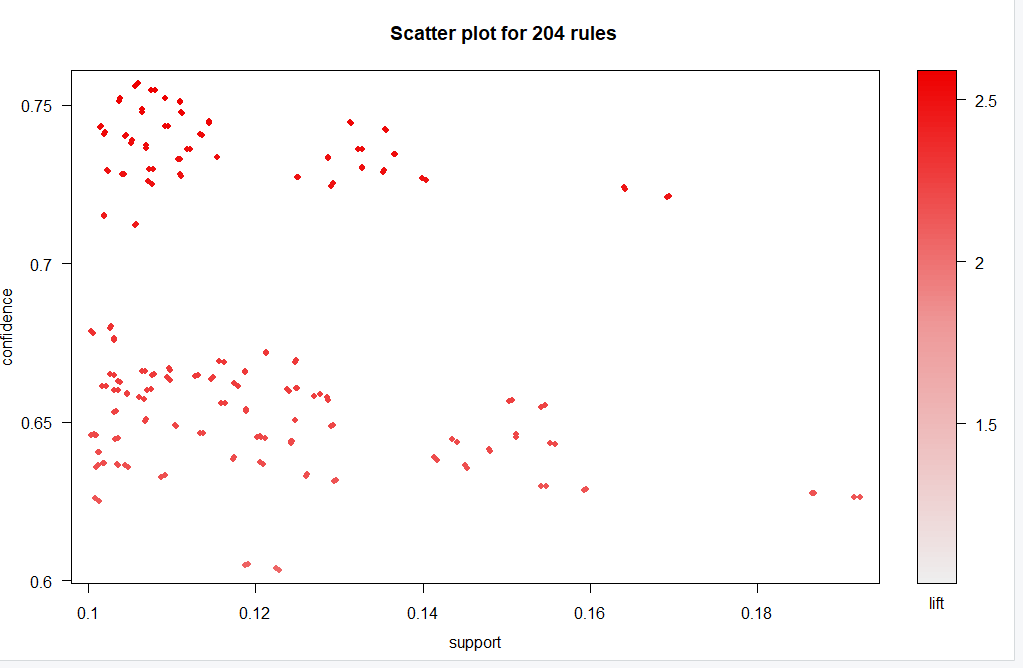
I have used the following association rules for the whole dataser.



The next few visualizations represent how the association rules have been applied for the first rule. I have generated a graph, scatterplot and parallel coordinates.







In the first visualization, the graph is generated for 100 association rules.

In the second visualization, the parallel coordinates generate 284 association rules.

In the third visualization, the scatter plot generates 204 rules, with support on the x-axis and confidence on the y-axis, having a range of 0 to 2.5.

I have generated such plots for all of my association rules.

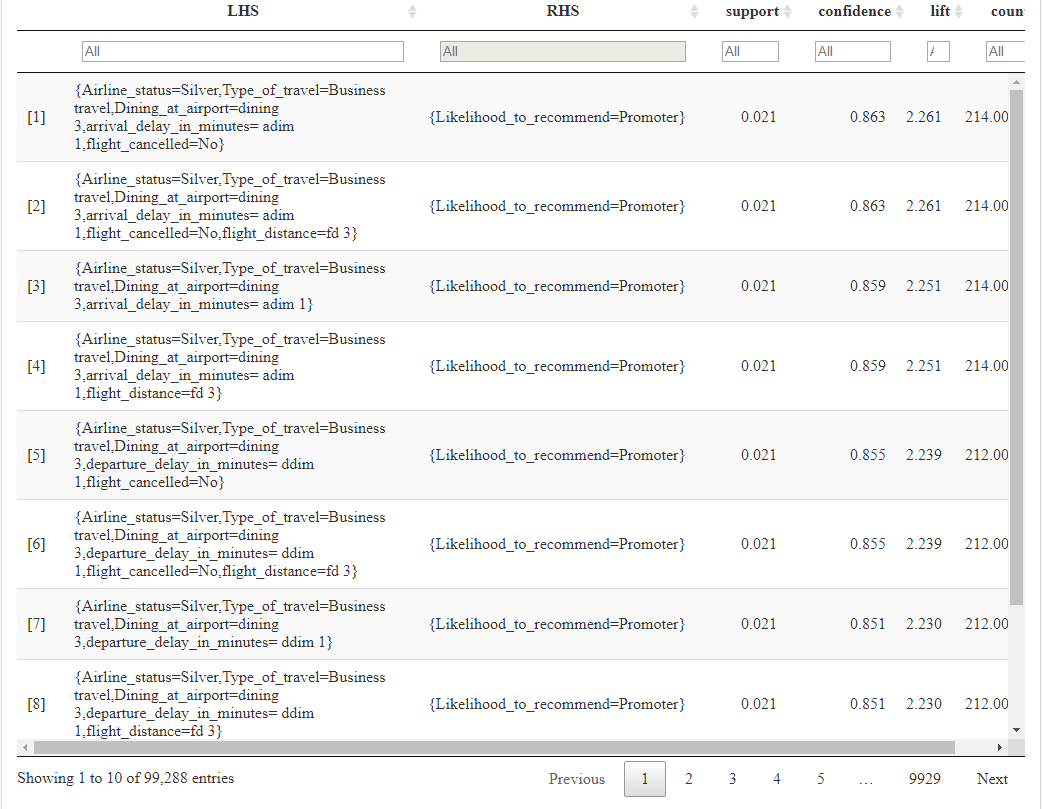
In the next visualization, I have run the InspectDT() function to glean analysis from the detractors.



The output shows us that I have decided to display the first 10 rules generated, I have sorted them as well. The result shows that a customer having a Blue airline flyer status, who is a personal traveler, who isn’t a frequent flyer and doesn’t dine at the airport, is not satisfied with the flight and is less likely to recommend the specific airline.



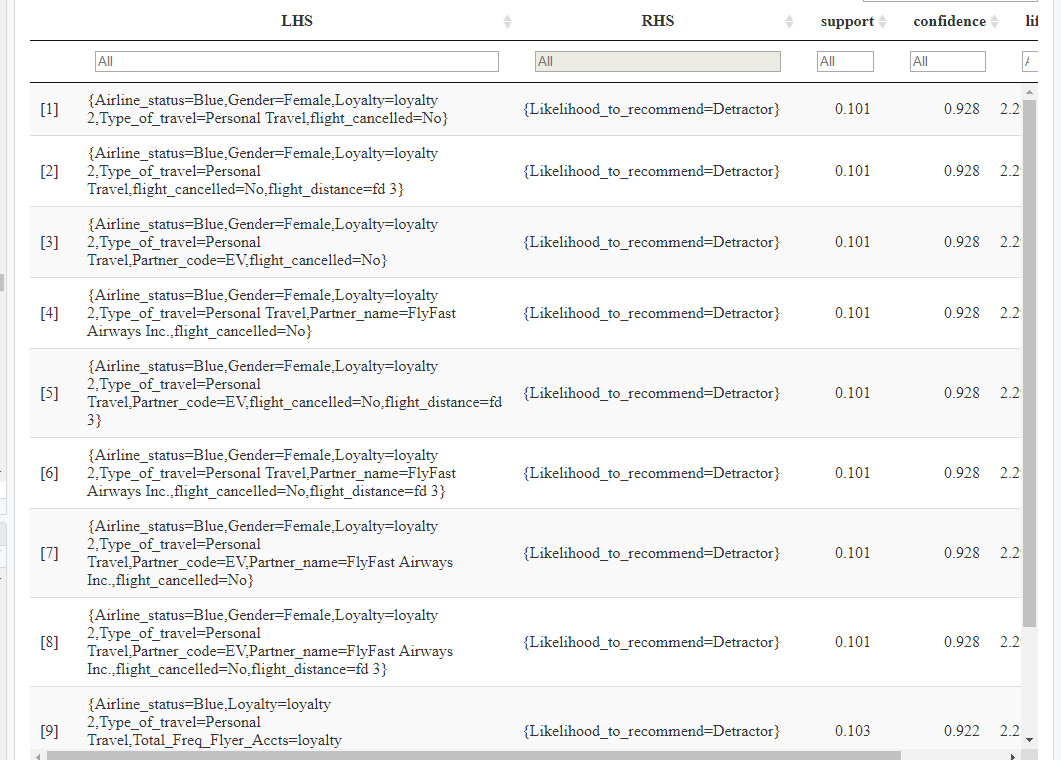
In this diagram, a customer who has a silver airline status, is a personal traveler, flies economy class, is considered to be a passive. Looking at the second customer, they are actually traveling for a long distance, which might be the reason that they are not promoting or demoting the airline.

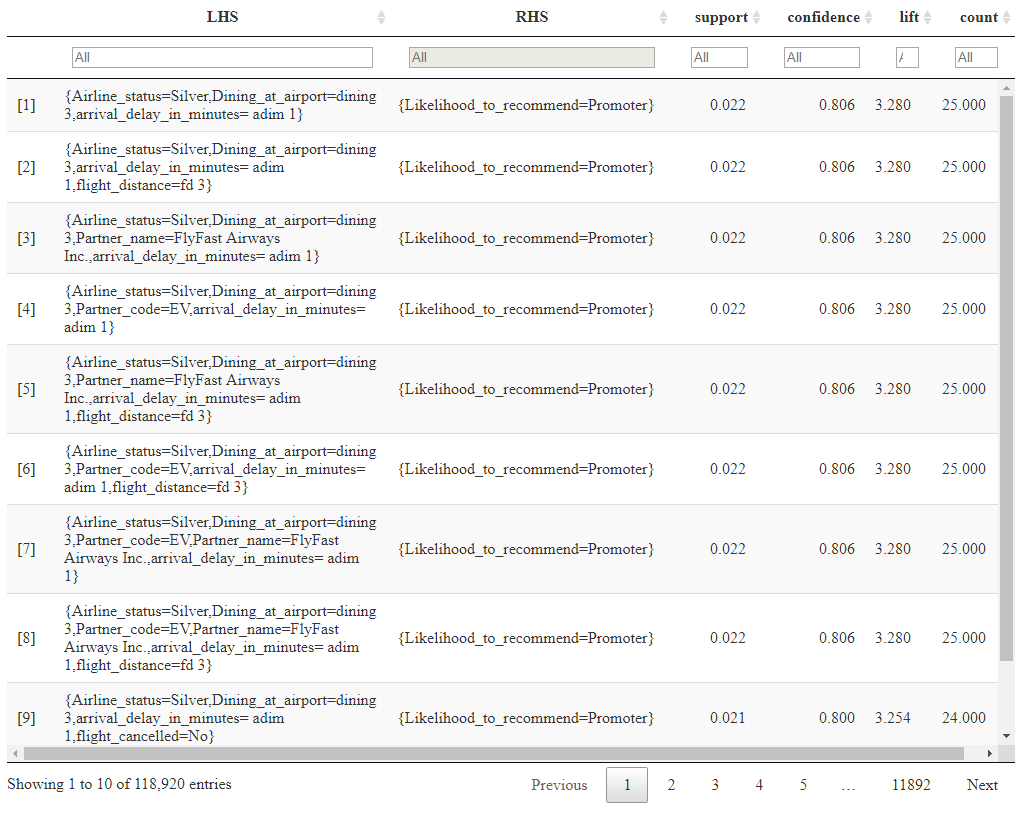
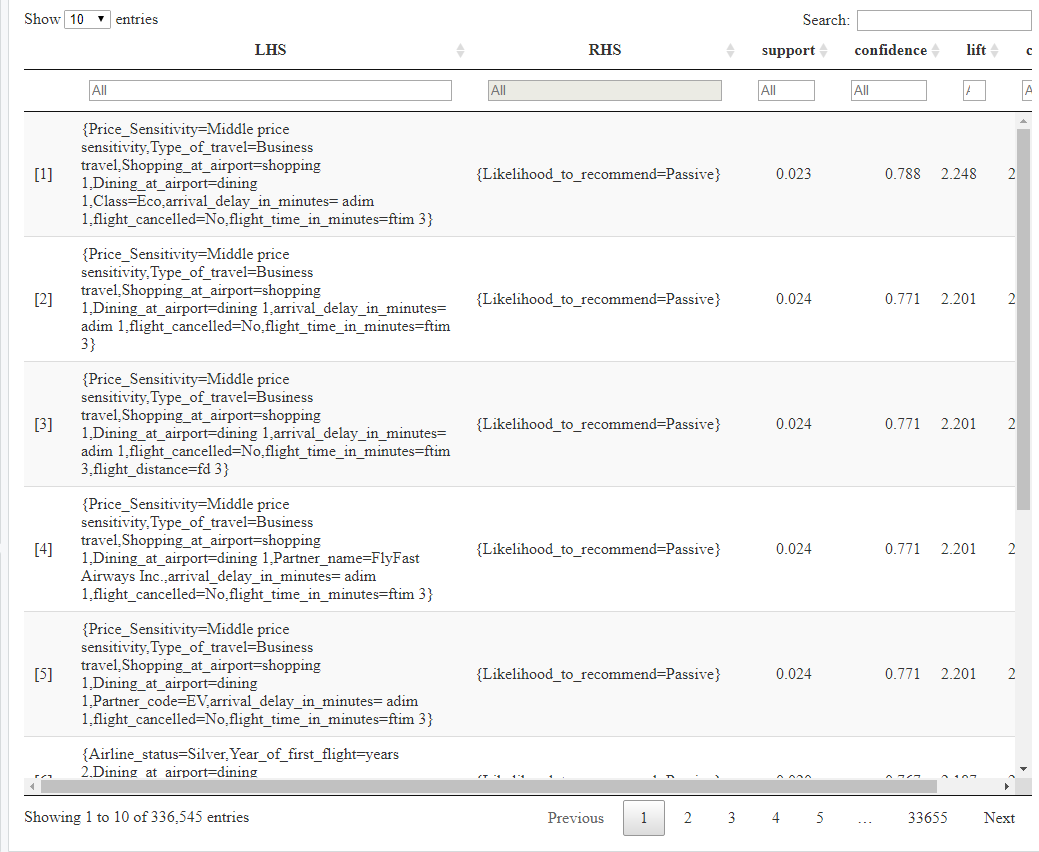


Finally, looking at the promoters, a customer having a silver airline flyer status, is a Business traveler, dines at the airport often and whose flight has very less arrival delay, is much more satisfied with the airline and actively promotes the airline.

From the information above, we can gather that the business travelers are much more satisfied as compared to the personal travelers. Also the customers in the economy class don’t seem to promote the airline much. The arrival delay was also very less for the promoters, which can also be considered as another factor for the satisfaction of the customers. Usually, the promoters are dining the most at the airports, this should also be considered.

The next few visualizations show the association rules for flyfast airlines, which has the least nps.





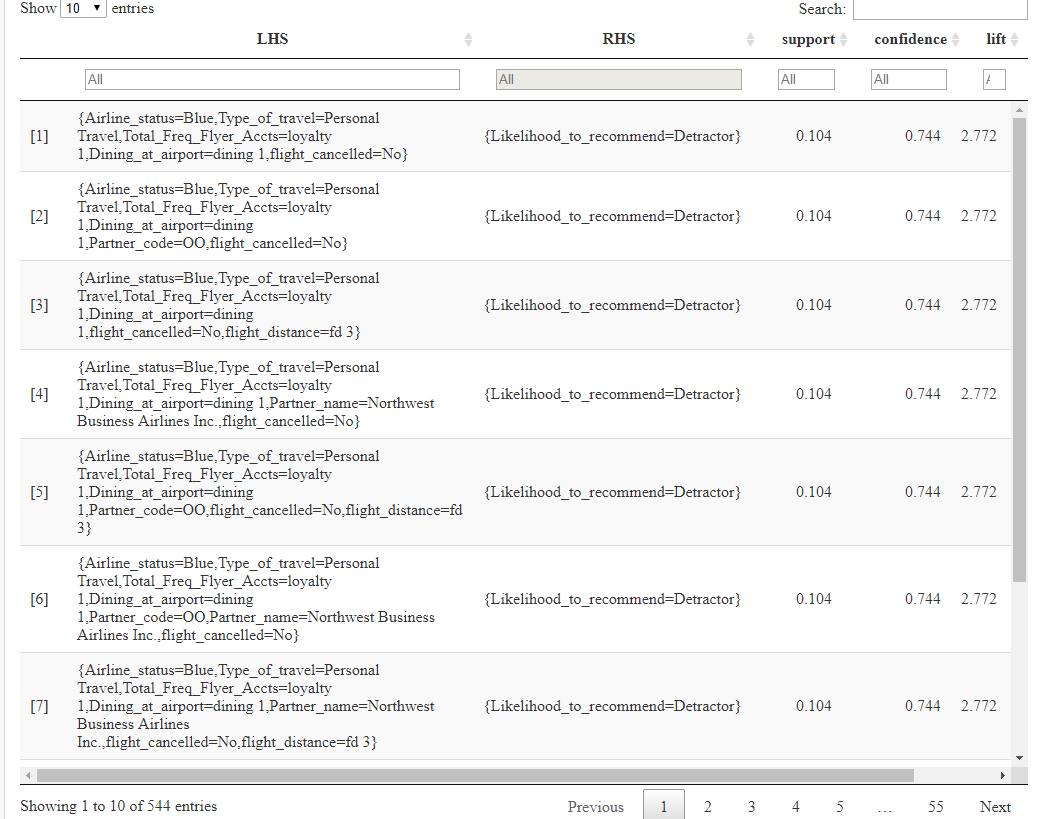
For flyfast airlines, we can see that females, travelling with a Blue airline flyer status and travelling for personal reasons, are detractors.

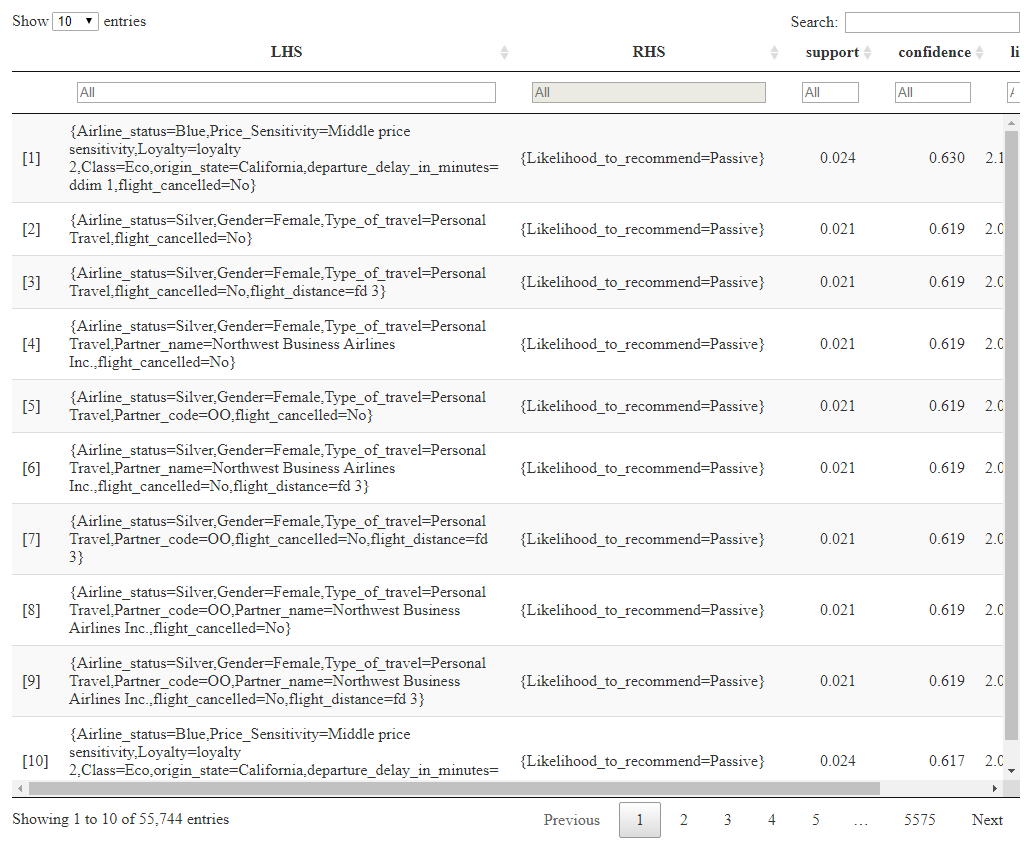
The customers whose purchasing is not affected by the prices, travelling for business purposes, shopping and dining the least at the airport are passive.

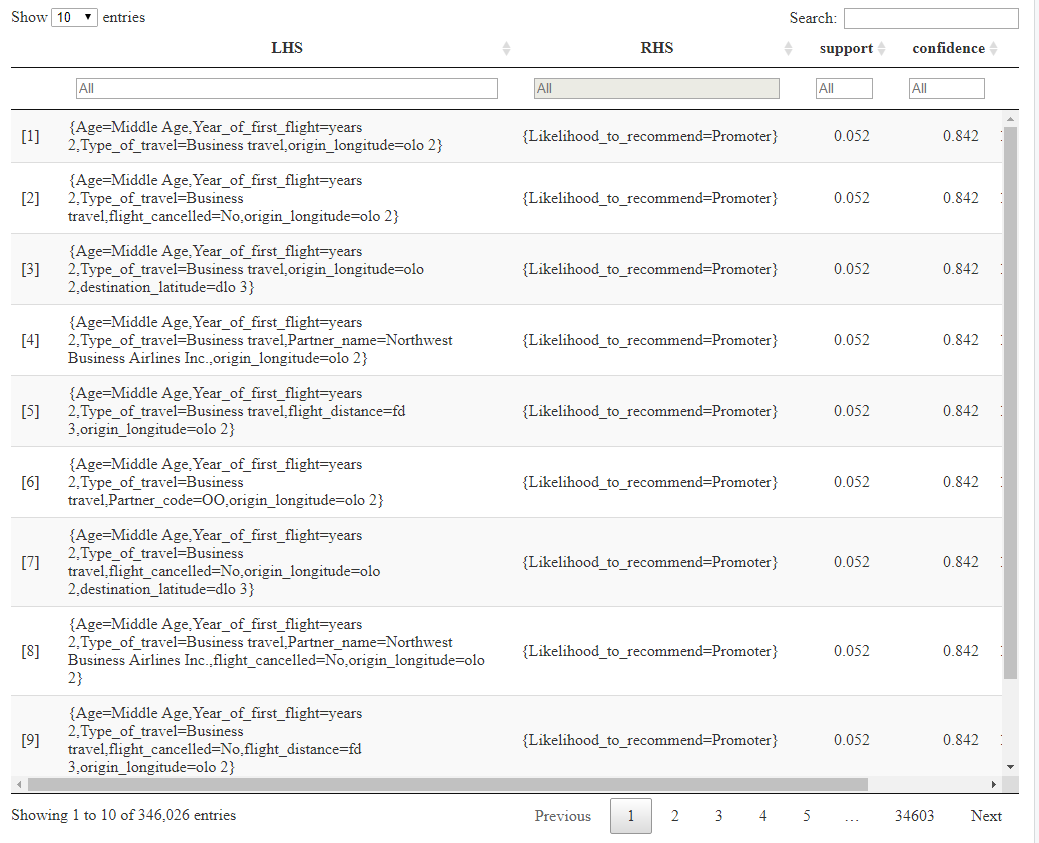
Customers with silver airline flyer status, who dine the most at the airport, whose flight arrival delay time is very less and the distance is long, are the promoters for flyfast airlines.

Again, we see that silver status customers, who dine at the airport are much more happier, as compared to Blue members, who are traveling for personal reasons.

The next few visualizations show the association rules for northwest airlines, which has one of the best nps.







For northwest airlines, the customers with Blue airline flyer status, traveling for personal reasons and who don’t dine much at the airport, are the detractors

Females who have a silver airline flyer status, who are traveling for personal reasons for a long distance are passive.

Customers aged between 30-55, traveling for business reasons, whose flight doesn’t get cancelled are the promoters.

From the given information, we can gather that Blue customers traveling for personal reasons are not satisfied, whereas, people aged between 30-55, travelling for business reasons are usually the promoters. Again , we can gather that the Silver customers are much more happier, the airlines needs to look into this and see how they can improve the customer experience of Silver customers.

5.4 Insights generated from Association Rules Mining

1) For the overall dataset: I have gathered that the business travelers are much more satisfied as compared to the personal travelers. Also the customers in the economy class don’t seem to promote the airline much. The arrival delay was also very less for the promoters, which can also be considered as another factor for the satisfaction of the customers. Usually, the promoters are dining the most at the airports, this should also be considered.

2) For FlyFast airlines: I saw that the silver status customers, who dine at the airport are much happier, as compared to Blue members, who are traveling for personal reasons.

3) For Northwest airlines: I have gathered that Blue customers traveling for personal reasons are not satisfied, whereas, people aged between 30-55, travelling for business reasons are usually the promoters. Again , we can gather that the Silver customers are much more happier, the airlines needs to look into this and see how they can improve the customer experience of Silver customers.

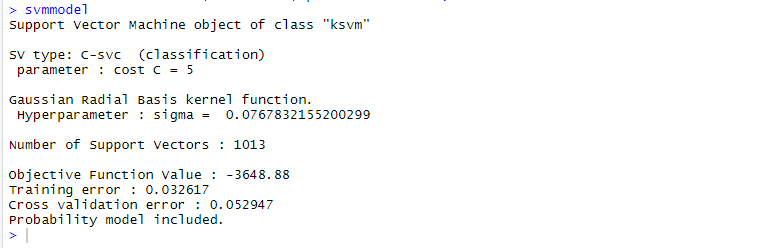
5.5 Support Vector Machine (SVM)

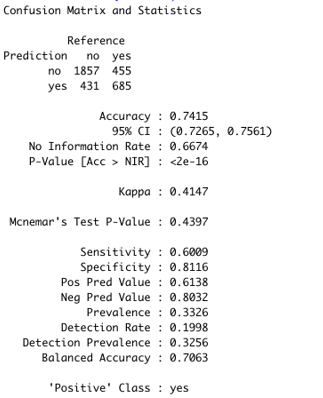
SVM tries to create a “hyperplane” to divide the data. It’s a linear model for classification and regression problems, which can solve linear and non-linear problems and work well for many practical problems. If we keep the discussion to a two-dimensional piece of paper, we can think of the hyperplane as a line dividing the two categories of data. The goal is to choose a hyperplane (the line in two dimensions) with the greatest possible margin between the hyperplane and any point within the training set, giving a greater chance of new data beingclassified correctly. SVM tries to make a decision boundary in a way that the separation between the two classes is as wide as possible.

In order to perform the analysis better, I have created bins again, dividing the likelihood to recommend into detractors, passive and promoters, just like before. I have also created an age group column for this analysis, where, age group 1 is between 15 and 29, age group 2 is in between 30 and 54 and age group 3 is above 54. I have also removed some unwanted columns from my data frame. I am attaching a snippet of my code below for SVM of the whole dataset.



Firstly, I have removed the unwanted columns from the data frame. Further, I have generated random indices and then took 2/3 of the values. I have also created two data frames for the training data and testing data. Further, I have installed the ‘caret’ package and then run the support vector machine in R. Finally, I have performed the prediction and generated the accuracy.



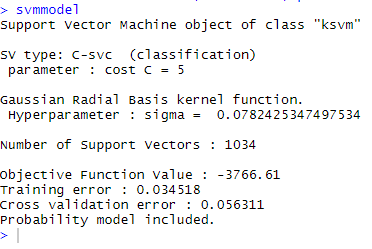


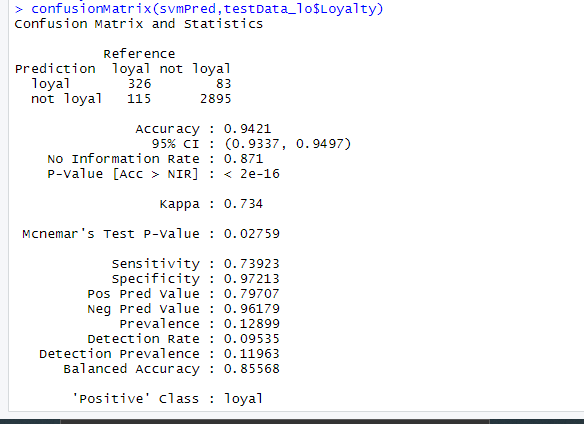
As we can see, the accuracy of the model is 0.7415, which means that the model works well in prediction. Besides, the sensitivity and specificity are considerable, which means that the model is reliable when predicting. The p-value is also less than 0.05, which means the equation is significant.

Just like for the whole dataset, I have also applied the support vector machine in order to predict the loyalty of a customer.



The following output was generated from this code:





The accuracy of the model is 0.9431, which means that the model works very well in prediction. Besides, the sensitivity and specificity are high, which means that the model is reliable when predicting. The confusion matrix shows that 2895 customers are not loyal. Thus, Southeast Airlines could use the model to predict the loyalty of a customer.

5.6 Insights generated from SVM

1) SVM for whole dataset: The accuracy of the model was 0.74, which wasn’t very high, but the sensitivity and specificity are considerable, which means that the model is reliable when predicting. The p-value is also less than 0.05, which means the equation is significant. Thus, support vector machines can used to find out the likelihood to recommend of a customer.

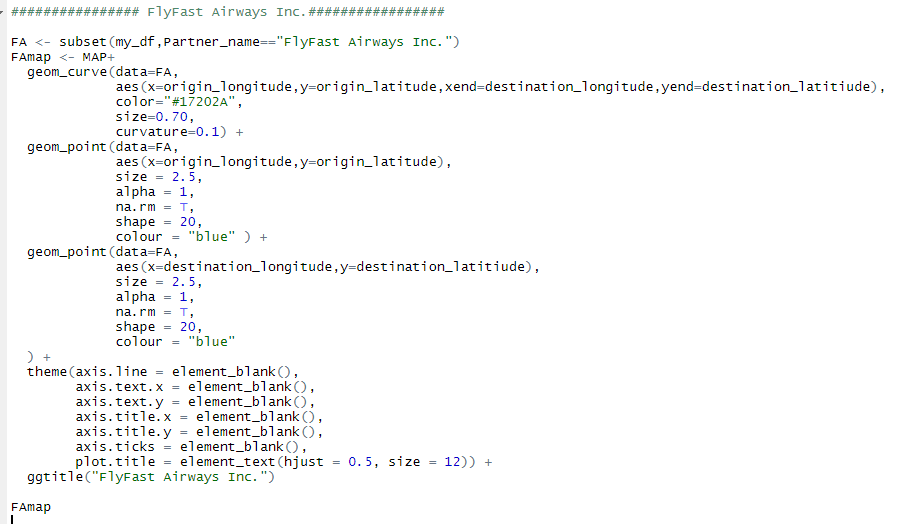
2) SVM for loyalty: In this case, the accuracy of the model is 0.9431, which means this model can be used for prediction. The sensitivity and specificity are also high, so the model is reliable for prediction. The confusion matrix shows that 2895 customers are not loyal..

**6) Map Route Visualizations of Low Satisfaction Airlines**

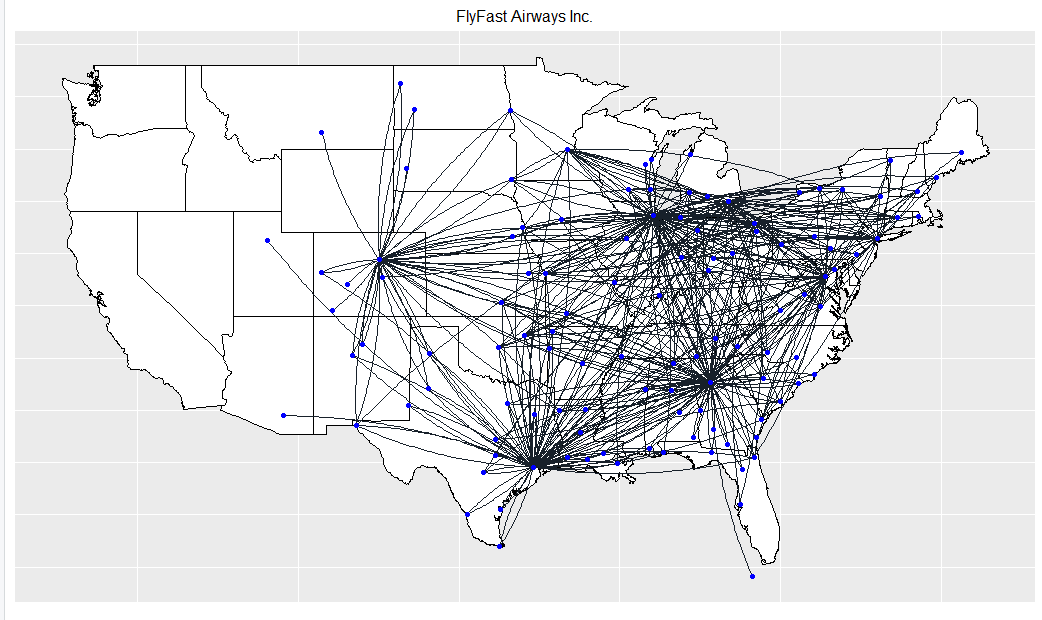
6.1 Visualizations

Earlier, when I found out the net promoter scores of each airline, I found out that Flyfast Airways Inc., Oursin airlines and GoingNorth airlines had the worst nps. In this visualization, I have considered these airlines and tried to glean why they might have low satisfaction routes. I have also created a visualization for Northwest airlines

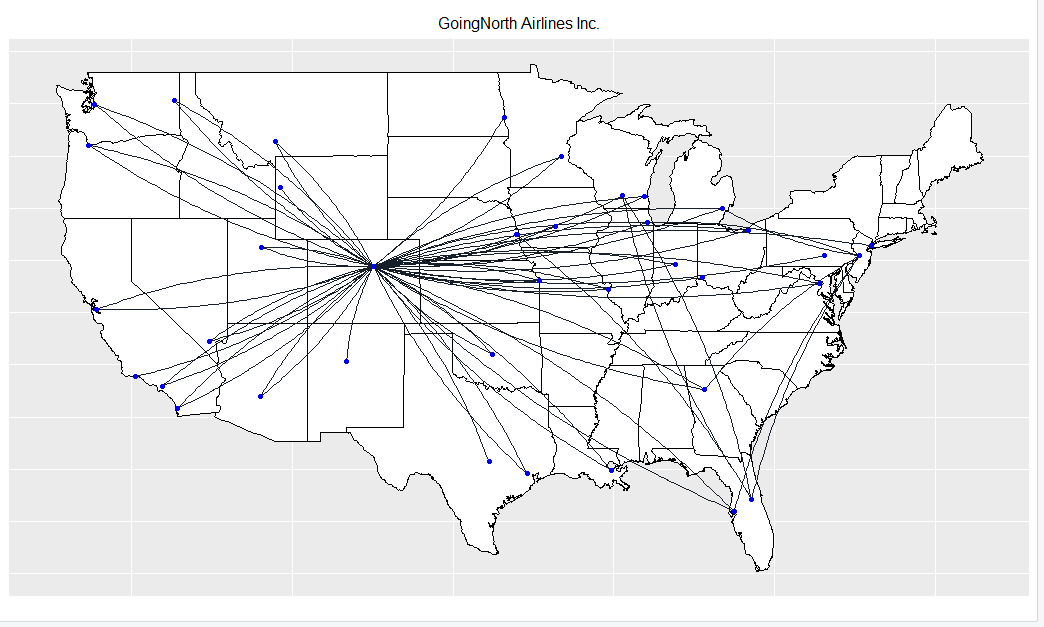
Firstly, I started with Flyfast airlines, which has the worst nps. I have attached the code below.



This code gave the following output:

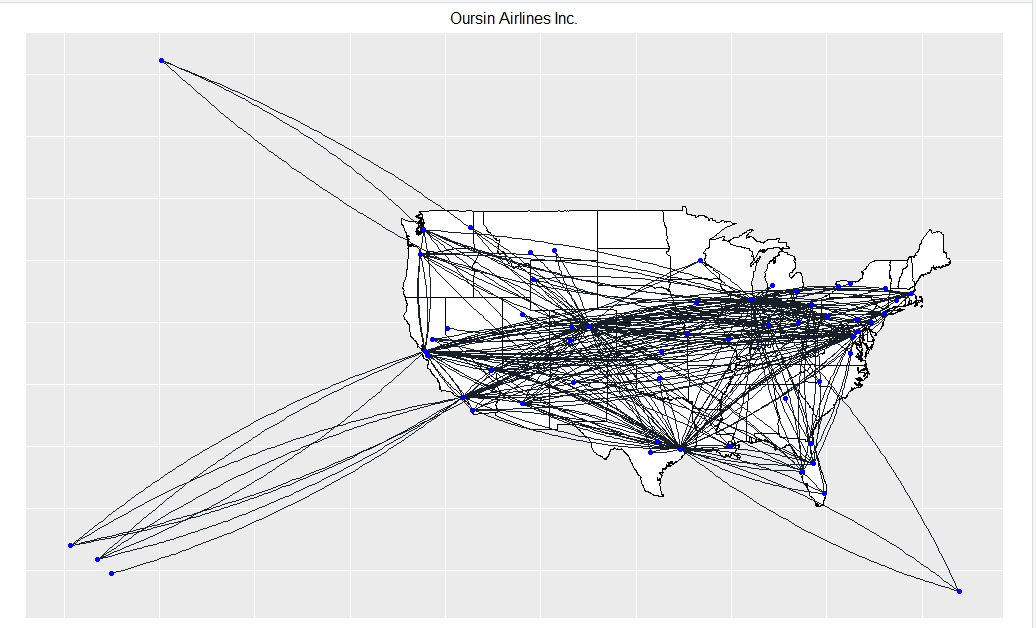


As we can see from the visualization, Flyfast airways has a lot of flights, most of which are flying in the east coast of the United States and some in the centre. Flyfast airways has absolutely no flights in the west coast, which could be one of the reasons of their low customer satisfaction, because most of the tourist cities like Los Angeles, San Diego, Las Vegas, etc. are on the west coast. Also, the flights in states like Colorado, Wyoming, Utah are very few. There are not many flights going back and forth from Florida as well, which is a huge tourist attraction in the USA. There is just one flight going to Havana, which is also a tourist attraction. There are no flights going to Hawaii or The Bahamas, which is surely another factor.

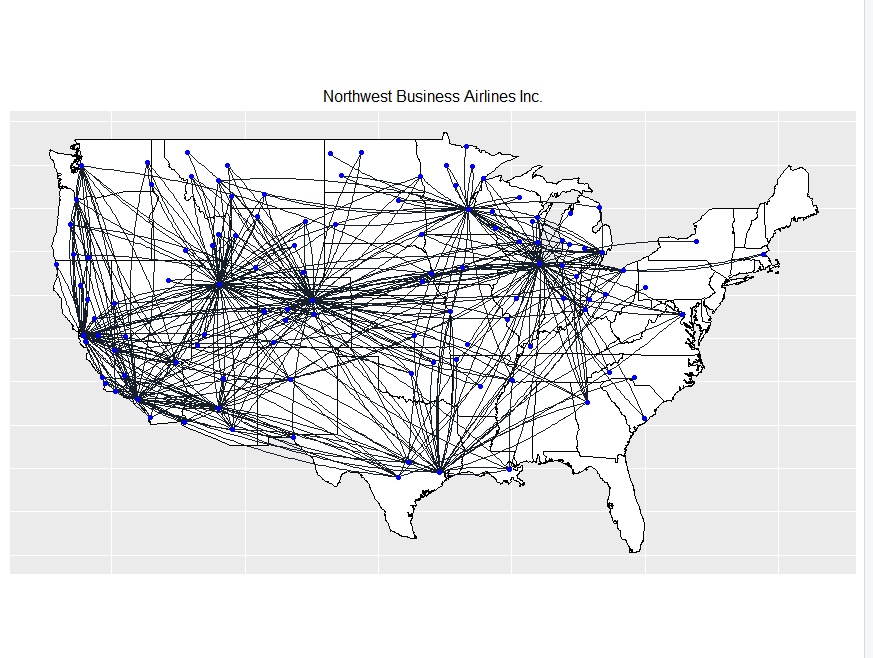
The next visualization shows the flight routes for GoingNorth airlines

From this visualization, we can see that GoingNorth doesn’t have many flights in the USA. As compared to flyfast, GoingNorth has more flights in the west coast, but still they are very less, which could be one of the reasons of their low customer satisfaction, because most of the tourist cities like Los Angeles, San Diego, Las Vegas, etc. are on the west coast. GoingNorth has a lot of flights going back and forth from Colorado. Maybe the airline has a hub there, and the service at Colorado airport is not the best. The airlines must improve the dining and shopping services at the airport, in order to improve the customer experience. Just like FlyFast airways, GoingNorth airlines doesn’t have any flights going to Hawaii, The Bahamas or Havana. There are hardly any flights going to Florida too.

The next visualization shows the flight routes for Oursin airlines



Oursin Airlines has a lot more flights, as compared to GoingNorth. As compared to flyfast and Goingnorth, Oursin has flights going back and forth in the east and west coast of USA, which is surely a good thing for Oursin airlines. Although Oursin airlines has flights going to Florida, there are not a lot of them, which can be a reason for the low customer satisfaction. As compared to flyfast and Goingnorth, Oursin has flights going back and forth from Hawaii, Alaska, Havana and Cuba.

The next visualization shows the flight routes for Oursin airlines

Northwest airlines has one of the best net promoter scores. As we can see from the visualization above, Northwest has a lot of flights going to the west coast of the USA, especially to places like Los Angeles, San Diego and San Francisco, which are tourist places. This might effect the customer satisfaction, since the customer is already in a good mood while flying. Even though Northwest has a lot of flights going to the west coast, it has very less flights going to the east coast, which should actually be looked at, because the customers travelling to the west coast seem to be much more satisfied. Surprisingly, Northwest has no flights going to Florida.

6.2 Insights

1) FlyFast Airways Inc: Flyfast airways has absolutely no flights in the west coast, which could be one of the reasons of their low customer satisfaction, because most of the tourist cities like Los Angeles, San Diego, Las Vegas, etc. are on the west coast. There are not many flights going back and forth from Florida as well and no flights going to Hawaii, The Bahamas or Havana. which are huge tourist attractions in the USA.

2) GoingNorth Airlines Inc.: As compared to flyfast, GoingNorth has more flights in the west coast, but still they are very less, which could be one of the reasons of their low customer satisfaction, because most of the tourist cities like Los Angeles, San Diego, Las Vegas, etc. are on the west coast. There are hardly any flights going to Florida too, or to Hawaii.

3) Oursin Airlines Inc.: As compared to flyfast and Goingnorth, Oursin has flights going back and forth in the east and west coast of USA. Oursin has flights going back and forth from Hawaii, Alaska, Havana and Cuba, which might result into a few happy customers, because the customer is already in a good mood while flying.

4) Northwest Business Airlines Inc.: Northwest airlines has one of the best net promoter scores. As we can see from the visualization above, Northwest has a lot of flights going to the west coast of the USA, especially to places like Los Angeles, San Diego and San Francisco, which are tourist places. This might effect the customer satisfaction, since the customer is already in a good mood while flying. Surprisingly, Northwest has no flights going to Florida, Hawaii, Havana, The Bahamas or Cuba.

**7) Recommendations to Improve Marketing Plans**

While conducting my analysis, I have generated some actionable insights. Based on these insights, I have developed the following marketing plans in order to improve the customer satisfaction.

7.1 Marketing segments and Actionable insights

1) **Age**: After conducting a thorough analysis, I have found out that the middle age group (between 30 and 54) are promoting the airlines, as compared to the third age group (above 55). Most of the customers between the age of 30 and 54 are travelling for business reasons, which is a major factor in them being the promoters.

**Actionable insights**: From my analysis, I have gathered that the third age group is not satisfied, this group mostly consists of senior citizens. One of the reasons can be that the senior citizens cannot navigate their way properly through the airport, which might eventually lead to missing their flight. The solution to this can be to assign special airlines helpers to the senior citizens, so that they can find the gate of their flights easily, which will lead to the customer being much happier. We also have to consider that most senior citizens are fatigued and finding your flights on the airport can be very challenging. In addition, the senior citizens might have to wait in long queues, which may irritate them. Express lines can be created for senior citizens, where they can get through the check point process easily. We also have to consider the in-air experience of these people. This can be improved by providing in-flight entertainment, for example, uploading some movies onto the in-flight entertainment system, which these people can enjoy. Other then that, comfortable seats can be allotted to citizens who are in the economy class. If a customer is a frequent flyer, then they can be given rewards or even bumped up to business class!

2) **Frequent Flyer Accounts:** From my analysis, I have gathered that most customers don’t have a frequent flyer account, which means that customers are changing airlines quite often while travelling. After applying association rules to Northwest Business Airlines Inc., I found out that the customers with low frequent flyer account were detractors.

**Actionable insights**: The airlines can urge the customers to create an account on the website, while booking their flights. This can be done by providing numerous awards to the customers. For example, for a customer’s first flight with an airline, the customer can be provided with a free meal on the airplane, but the customers can only register for a free meal if they create an account and register for it online. The customers can also be provided with promotional codes, by sending them these codes via e-mail. The most important thing would be to give miles to the customers whenever they complete a flight, which are stored on their personal frequent flyer accounts. The customers can also be rewarded with points, if they refer the airline to another customer, who creates an account too.

**3) Shopping and Dining at the airport:** From my analysis, I have gathered that the customers who do not shop or dine at the airport, or don’t shop or dine rarely, are not satisfied with the airlines. After applying association rules, I found out that the customers who dine very less or don’t dine at all at the airport are detractors. Also, the number of people dining at the airport is more than the people shopping.

**Actionable insights**: The best thing to improve the shopping at an airport would be to promote the airport as a shopping destination. Let’s take an example, the Dubai International Airport advertises itself as a place to have fun, even while travelling. The airport has a range of restaurants to dine at and even cinemas. The shopping at an airport can be improved by providing discounts to customers who have frequent flyer accounts. The duty-free shopping center can also offer sales at their stores, in order to attract customers. This method would surely work as millions of people travel daily, and a little advertising can help in getting more customers to shop at the airport. In order to improve the dining options, the airlines can provide lounge access to customers. The lounges can have their own food buffet system, so the customers can pay for the buffet and relax in the lounge while enjoying their food!

**8) References**

1. Saltz, J. S., & Stanton, J. M. (2018). An introduction to data science.
2. Skift (2014). *The Biggest Issues Faced by Elderly Travelers at American Airports.* Retrieved from <https://skift.com/2014/04/16/the-biggest-issues-faced-by-elderly-travelers-at-american-airports/>
3. Faraway, J. J. (2005). Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models. United Kingdom: CRC Press.
4. O’Donnell (2003). *How Frequent Flyer Programs Work*. Retrieved from <https://adventure.howstuffworks.com/destinations/travel-guide/tips/ff-programs.htm>
5. n.a. (2020). *What is Net Promoter Score?* Retrieved from <https://www.qualtrics.com/experience-management/customer/net-promoter-score/>