Project Report

Customer Satisfaction Score Analysis for Airline Passengers

Summary:

The dataset "Satisfaction Survey" originates from the airline industry. Our primary goal is to analyze which of the 129889 variables in 30 different attributes influence the "Customer Satisfaction" score for passengers traveling through flights. Based on the analysis we would be able to understand the pain points of this industry and in turn, provide actionable recommendations for improving the customer experience.

Our team focused on analyzing the data through visualization to understand if any trends exist. Using the concepts, we learned during IST 687; we will test the dataset through linear regression, a statistical technique to find which variables contribute towards customer satisfaction. We also calculate the association rules, a data mining process (confidence, support and lift) as there are 30 different attributes which can coincide and in turn affect the score. It will also help determine some unique pattern among the variables. Understanding these combinations could boost or hamper the industry.

Content

1. Introduction

It was almost 115 years ago that the Wright brothers took “the first flight” and after 11 years, the first commercial airline flight took place. The airline industry has never looked back since then. There have been several technology advancements with “Passengers” as the pivotal point. However, after 115 years, getting people from one destination to another in the shortest time is not the only goal. Today, personalizing a passenger’s trip, the best price that can be offered and retaining passengers are some of the goals that this industry strives for.

The dataset contains 129889 variables distributed across 30 attributes which affect the Customer Satisfaction Score. Keeping “South East” airlines as our airline of interest and the goals of airline industry, we will be analyzing the dataset through linear regression, association rules and build a Support Vector Machine model which can predict the customer satisfaction level in future passengers. Along with these tests and model, data visualization will help us understand the strength and weakness of the industry,mainly our targeted airline. This report data awill deliver conclusions and actionable insights based on these results.

2. Business Questions Addressed

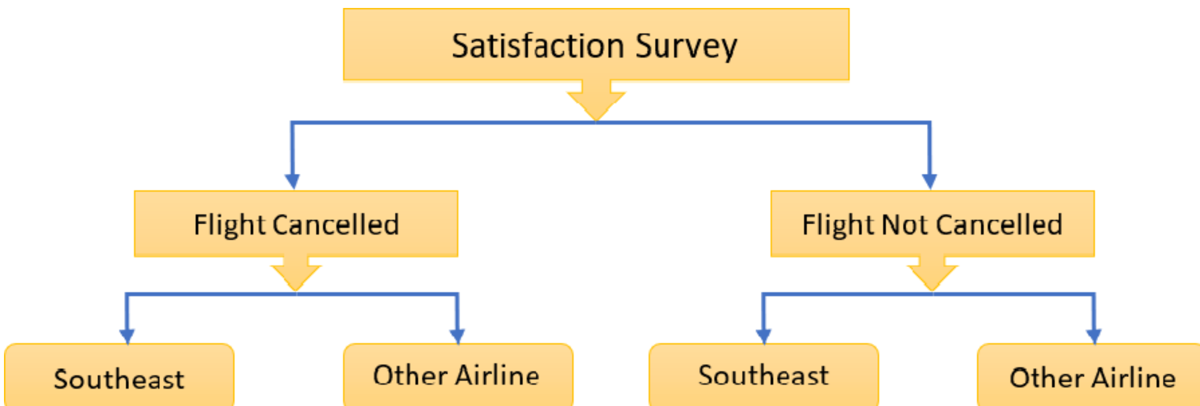
- Attributes affecting the customer satisfaction level the most.
- Difference in satisfaction level between customers whose flights were canceled or not canceled.
- Difference in satisfaction level between “South East” and other airlines.
- Category of customers dictates the greatest numbers of “low satisfaction score”.
- Area of the U.S. having the highest chance for flight cancellation.

3. Data Acquisition, Cleansing, Transformation, Munging

3.1 General data munging:

For achieving better results through our tests and models, we divided the dataset, cleaned it of irrelevant values, converted one data type to another to obtain a more structured dataset.

- Divided the data frame into two different subsets: Canceled flights, and Not Cancelled.
- Removed NAs for both subsets separately. Reason being, for passengers that have their flight, canceled they would have NA values for variables such as ***Departure.Delay.in.Minutes*** and ***Arrival.Delay.in.Minutes.*** If we were to remove NAs without dividing the dataset, we would risk deleting all passengers that had their flight canceled. Therefore, for the flight canceled subset we deleted the NA columns instead of deleting rows with NA.
- The dataset was further segregated into "Southeast" and "Other Airline" for analysis.



- Transforming ***Satisfaction*** and ***Scheduled.Departure.Hour*** into categorical variables for cross- group comparison for satisfied and unsatisfied customers based on their flight schedules.

```
Satisfaction$SatiClass<-replicate(dim(Satisfaction)[1],"Average")
Satisfaction$SatiClass[Satisfaction$Satisfaction<4]<-"Low"
Satisfaction$SatiClass[Satisfaction$Satisfaction>4]<-"High"
Satisfaction$SatiClass<-as.factor(Satisfaction$SatiClass)
```

```
Satisfaction$DeparturetimeClass<-replicate(dim(Satisfaction)[1],"Normal")
Satisfaction$DeparturetimeClass[Satisfaction$Scheduled.Departure.Hour<9]<-"Early"
Satisfaction$DeparturetimeClass[Satisfaction$Scheduled.Departure.Hour>18]<-"Late"
Satisfaction$DeparturetimeClass<-as.factor(Satisfaction$DeparturetimeClass)
```

- We also created new columns ***departState*** and ***arriveState*** from ***Origin.State*** and ***Destination.State*** for converting the state names into lower case character for ***ggmap***

- Transforming categorical variables ***Class, Airline.Status*** into factors for data visualization and ordering them (level) as the following: ***Blue, Silver, Gold, Platinum*** and ***Eco, Eco Plus, Business.***

```
Satisfaction$Airline.Status<-factor(Satisfaction$Airline.Status,
                                    levels = c("Blue","Silver","Gold","Platinum"))
Satisfaction$Class<-factor(Satisfaction$Class,levels = c("Eco","Eco Plus","Business"))
```

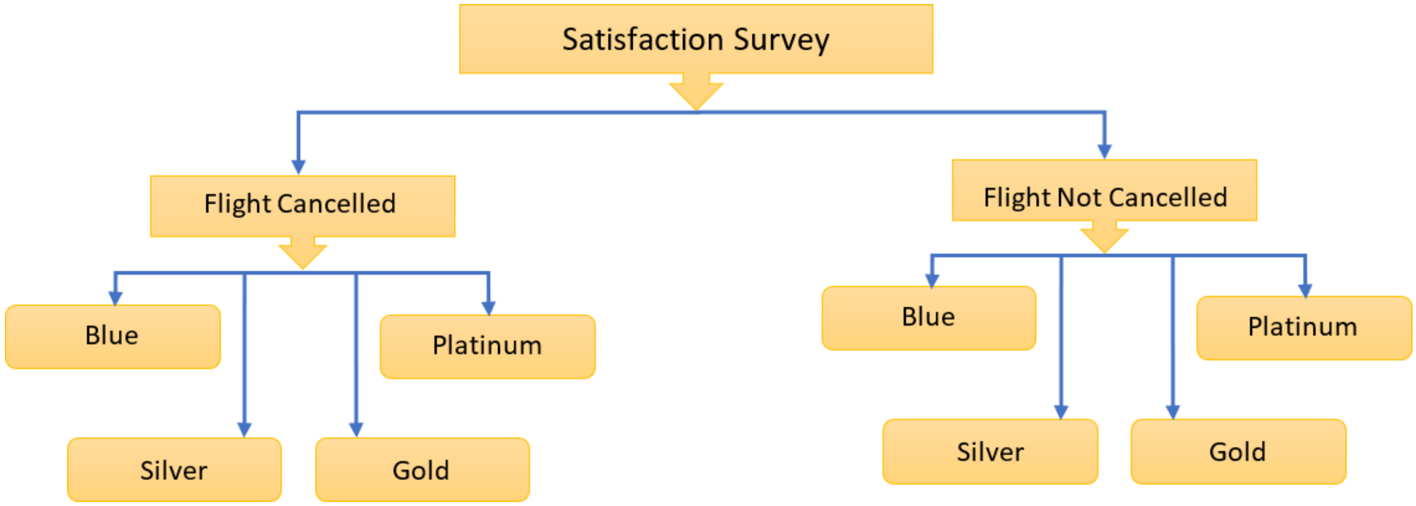
- Classified age into different groups of “0-15”,”16-25” till “76-85” for data visualization.
- Transformed the ***Flight.Date*** variables into standard date format using “as.Date”.

3.2 For running linear regression:

- Ensuring that numbers remain as numeric values.

```
Satisfaction<-SatisfactionSurvey
Satisfaction$Satisfaction<-as.numeric(Satisfaction$Satisfaction)
```

- We divided the dataset into “Southeast” and “Other Airline”.
- Based on the results, we decided to divide it further into the different ***Airline.Status:***



3.3 For Association Rules:

- We used the same dataset as shown in the figures above for our analysis.

- Mapped numerical attributes to categorical values, such as “High,” “Average,” “Low” for

  1. Age
  2. Airline.Status
  3. Gender
  4. Price.Sensitivity
  5. Year.of.First.Flight
  6. No.of.Flights.p.a.
  7. No.of.Flight.with.other.Airlines
  8. Type.of.Travel
  9. No.of.Other.Loyalty.Cards
  10. Shopping.Amount.at.Airport
  11. Eating.and.Drinking.at.Airport
  12. Class
  13. Airline.Name
  14. DeparturetimeClass
  15. Flight.Distance

```
fuc = function(vec){
  q<-quantile(vec,c(0.4,0.6))
  vBuckets = replicate(length(vec), "Average")
  vBuckets[vec > q[2]] = "High"
  vBuckets[vec < q[1]] = "Low"
  vBuckets=as.factor(vBuckets)
  return(vBuckets)
}
```

```
newdf_1$Price.Sensitivity<-replicate(dim(newdf)[1],"Average")
newdf_1$Price.Sensitivity[newdf$Price.Sensitivity<3]<-"Low"
newdf_1$Price.Sensitivity[newdf$Price.Sensitivity>3]<-"High"
newdf_1$Price.Sensitivity<-as.factor(newdf_1$Price.Sensitivity)
```

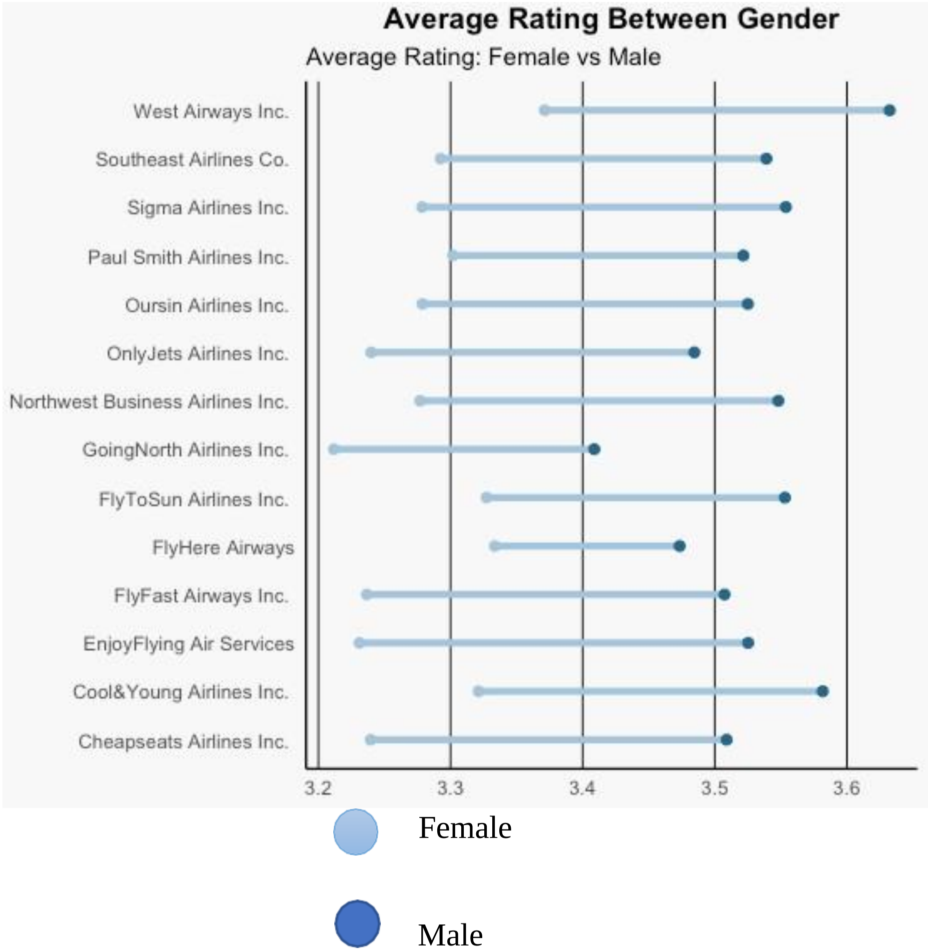
- Transformed all numerical columns to “factors” to run the arules.

3.4 For SVM model:

- Created a new column that categorizes the customers into Happy or Unhappy customers in the original cleaned dataset.
- Convert this new column into factors.
- Created training and test dataset.

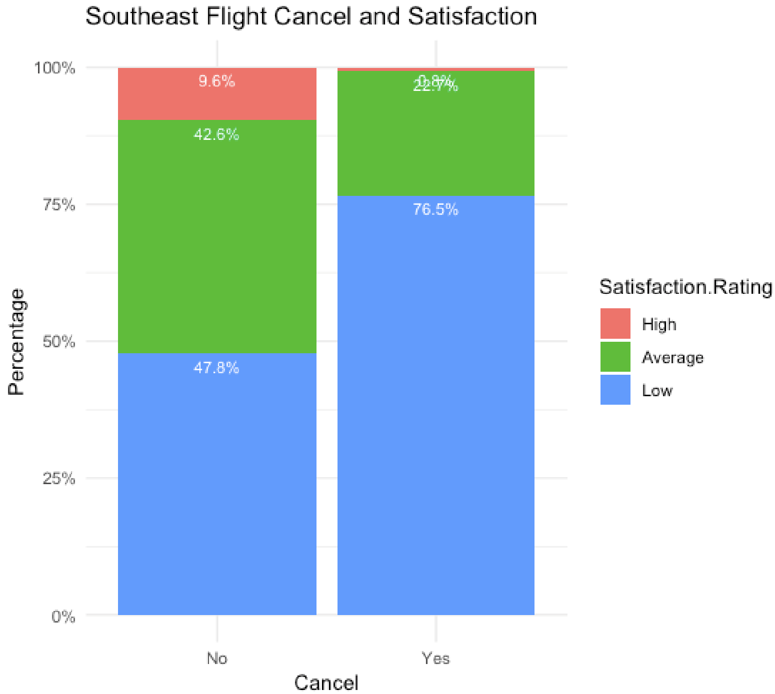
4. Descriptive statistics & Visualization

4.1 Average rating for each airline based on gender

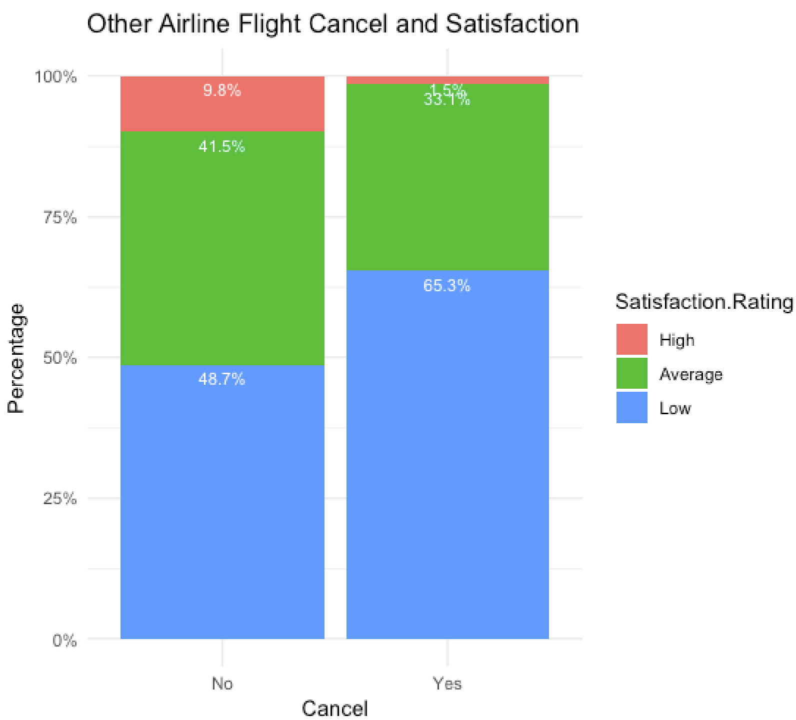


4.2 Customer Satisfaction for passengers when their flight is cancelled

a. For South East Airlines (%)



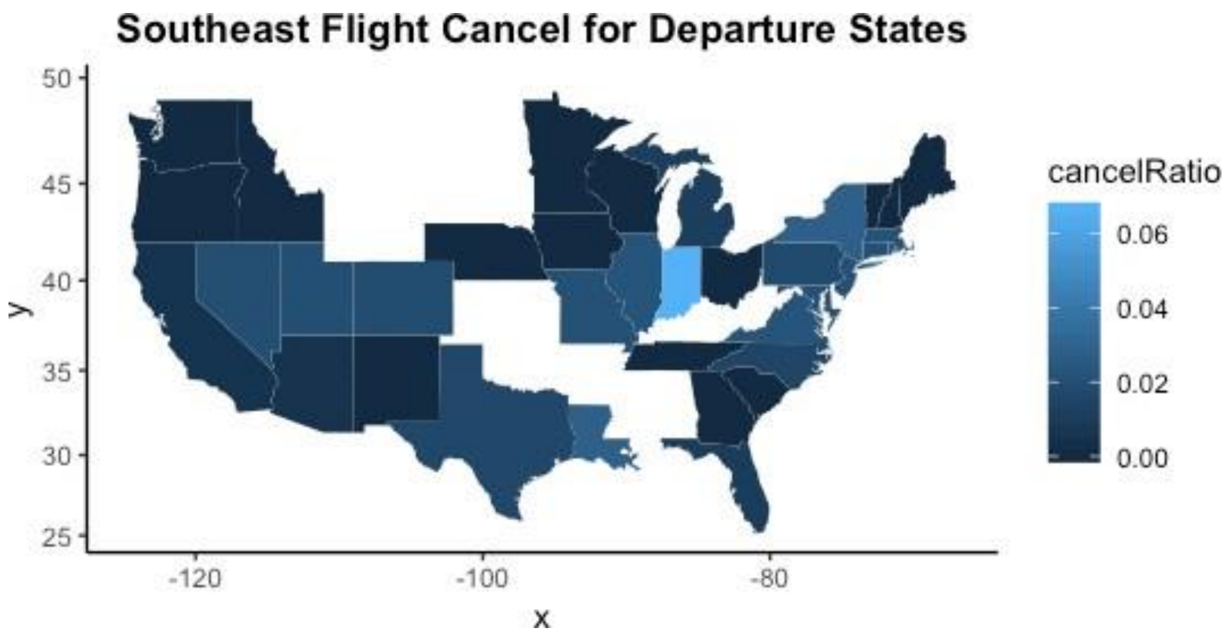
b. For other airlines (%)



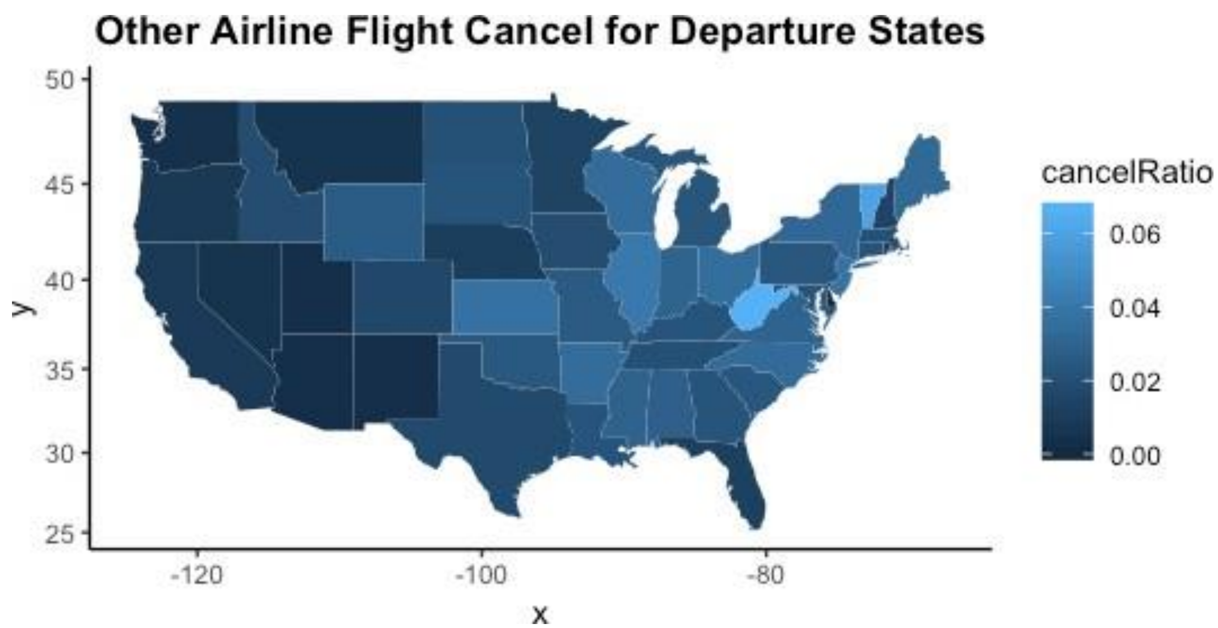
Customer Satisfaction proportion for cancelled flights for South east and other airlines

Customer Satisfaction Rating	Not canceled (NC)	Canceled (C.)	NC ratio	C ratio
Average	52979	782	41.56%	32.57%
High	12518	36	9.82%	1.50%
Low	61991	1583	48.62%	65.93%

4.3 Delay in departure for U.S. States



Flights originating from Indiana have a higher chance of getting cancelled.



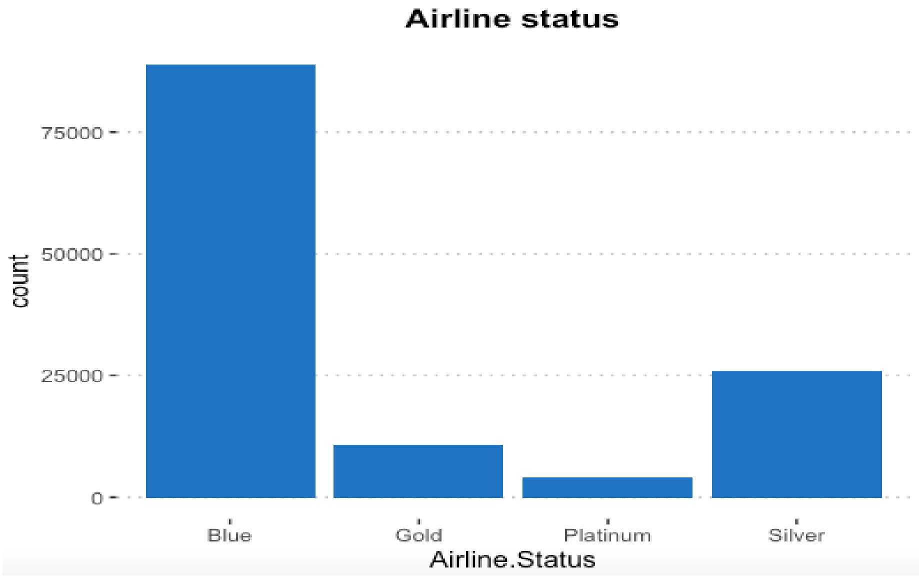
Flights originating from west Virginia have a higher chance of getting cancelled

4.4 Delay in arrival for U.S. States

All South East airlines headed towards Georgia have the highest chance of getting cancelled

All other airlines heading towards West Virginia and New Hampshire on east coast are equally likely to be cancelled

4.5 Distribution of passengers in different airline status

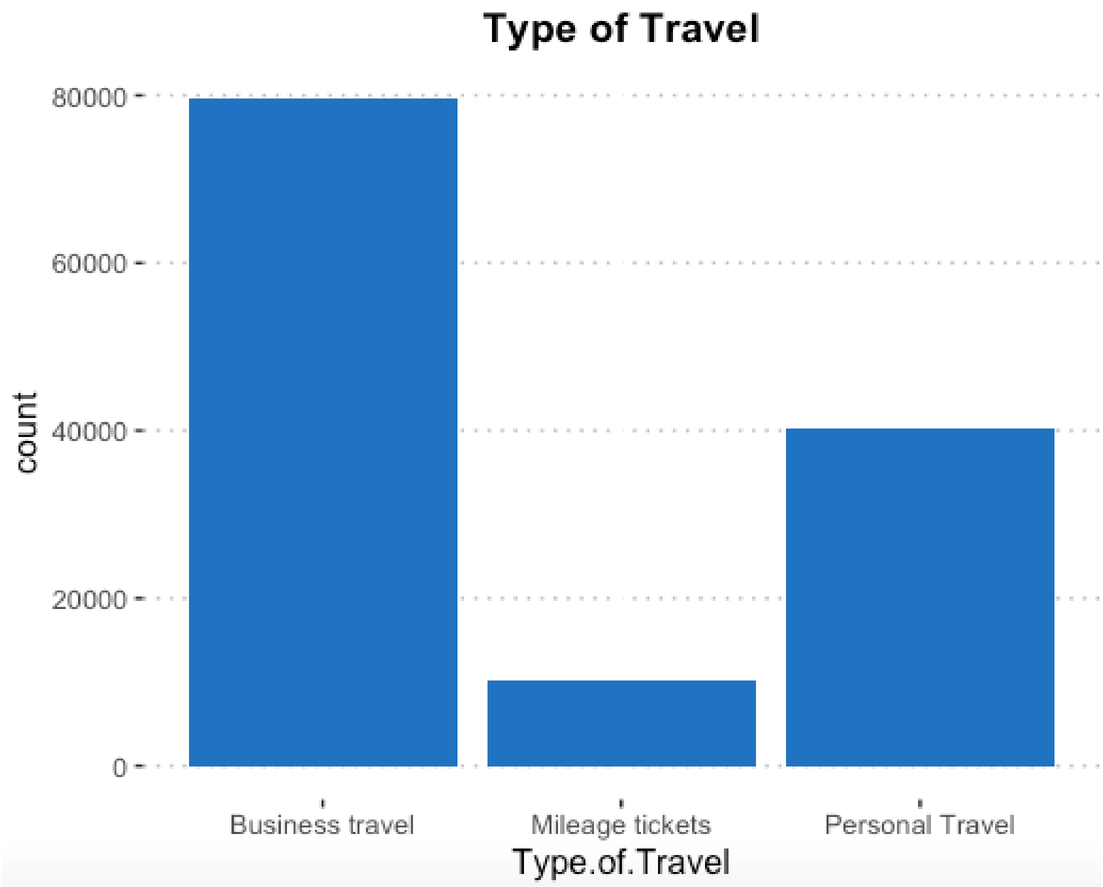


Count of passengers in different airline classes with "Blue" passengers with an overwhelming majority compared to other classes.

Customer Satisfaction Rating	Blue	Gold	Platinum	Silver
Average	38.09%	37.53%	23.75%	57.14%
High	3.60%	27.02%	38.49%	18.55%
Low	58.31%	35.45%	37.75%	24.31%

A table containing Customer Satisfaction Rating (%) for different airline statuses.

4.6 Passengers distribution in different types of travel

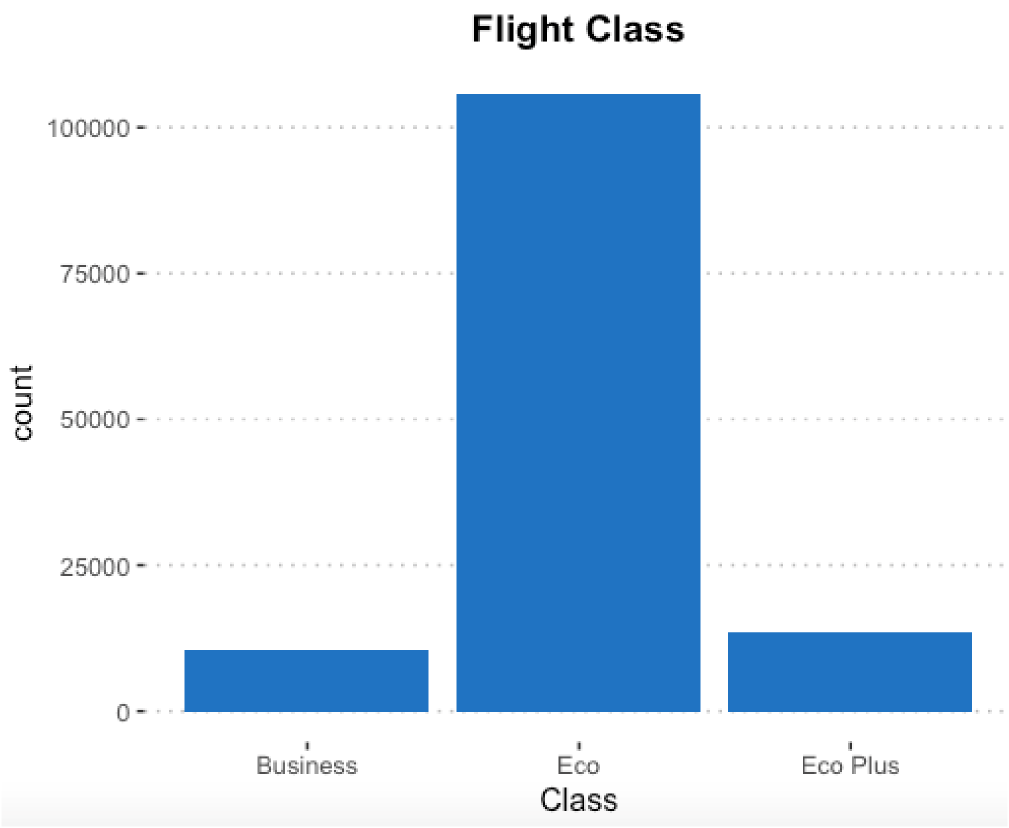


We observe that frequent flyers are mainly traveling for Business purpose, followed by personal and mileage-based tickets.

4.7 Price sensitivity distribution among different status

Price Sensitivity	Blue	Silver	Gold	Platinum
0	3005	720	220	139
1	57185	19262	8372	3260
2	27070	5790	2173	745
3	1485	179	65	25
4	165	18	7	3
5	0	1	0	0

4.8 Distribution of passengers in different airline classes

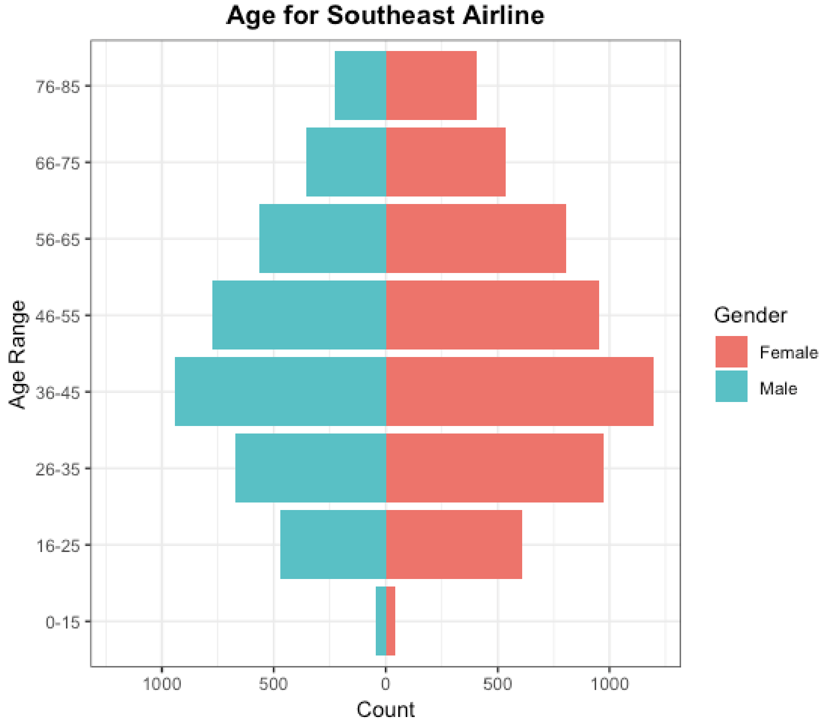


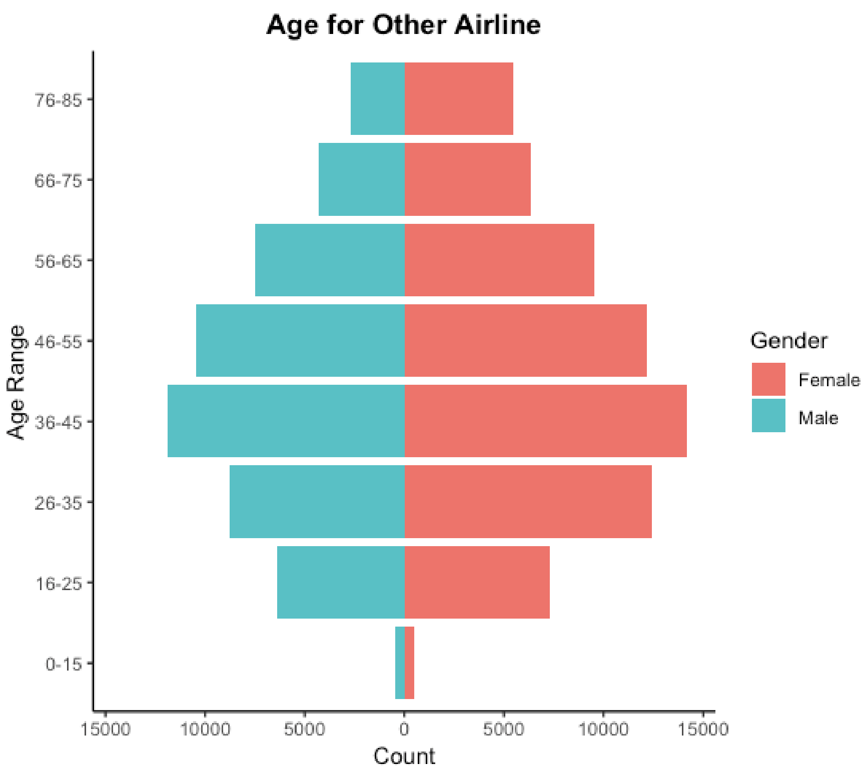
Majority of passengers choose economy class while travelling

Customer Satisfaction Rating	Business	Eco	Eco Plus
Average	4855	43266	5640
High	1380	10267	907
Low	4313	52202	7059

4.9 Age distribution

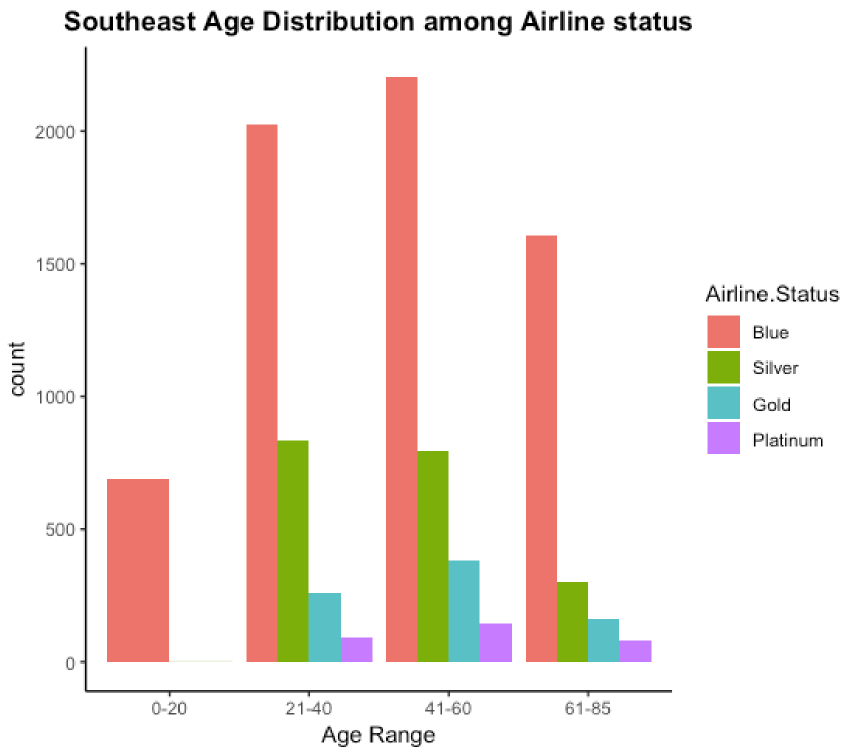
a. Age distribution based on Gender

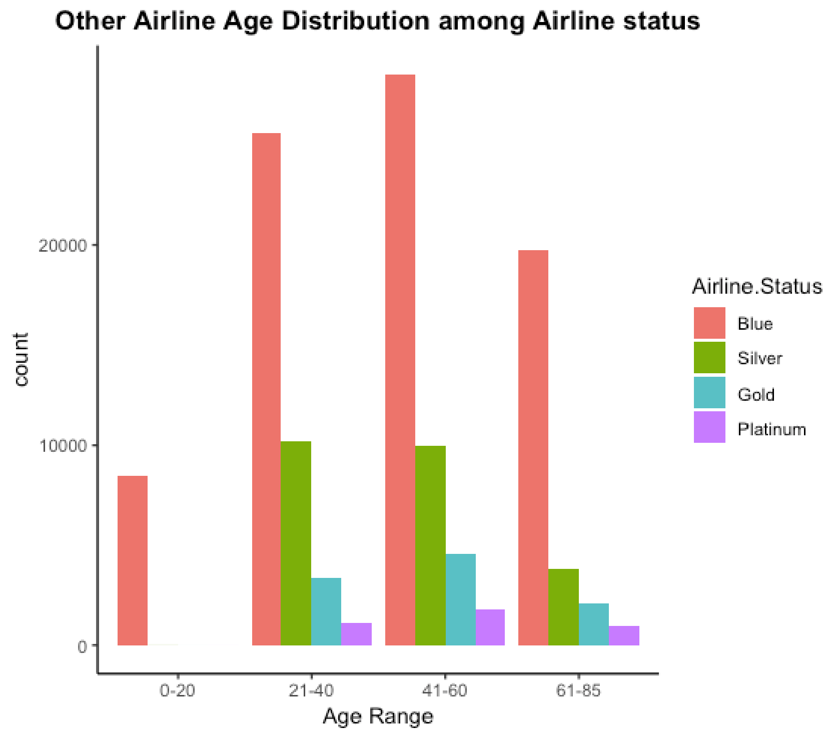




Overall distributions show that there are more female passengers aged between 25-55

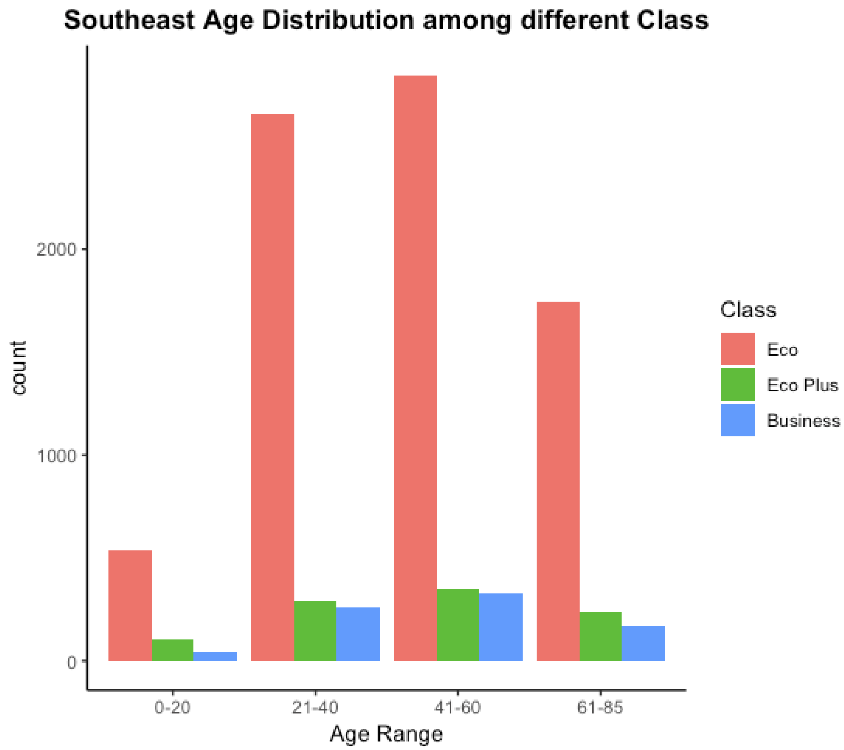
b. Age distribution based on airline status

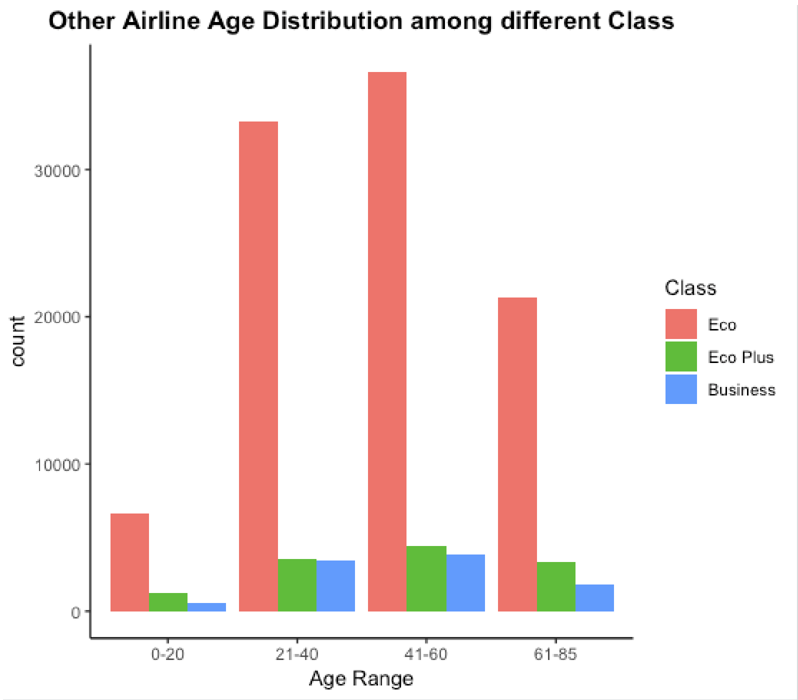




Passenger composition to Airline Status are similar between South East and Other Airlines with Blue passengers occupying an overwhelming majority.

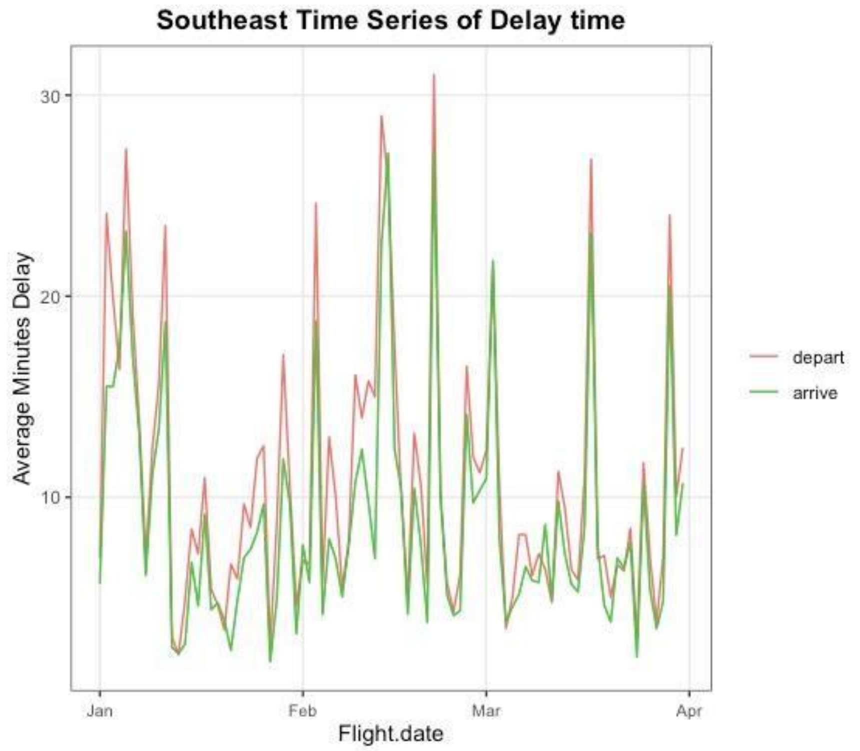
c. Age distribution based on different class

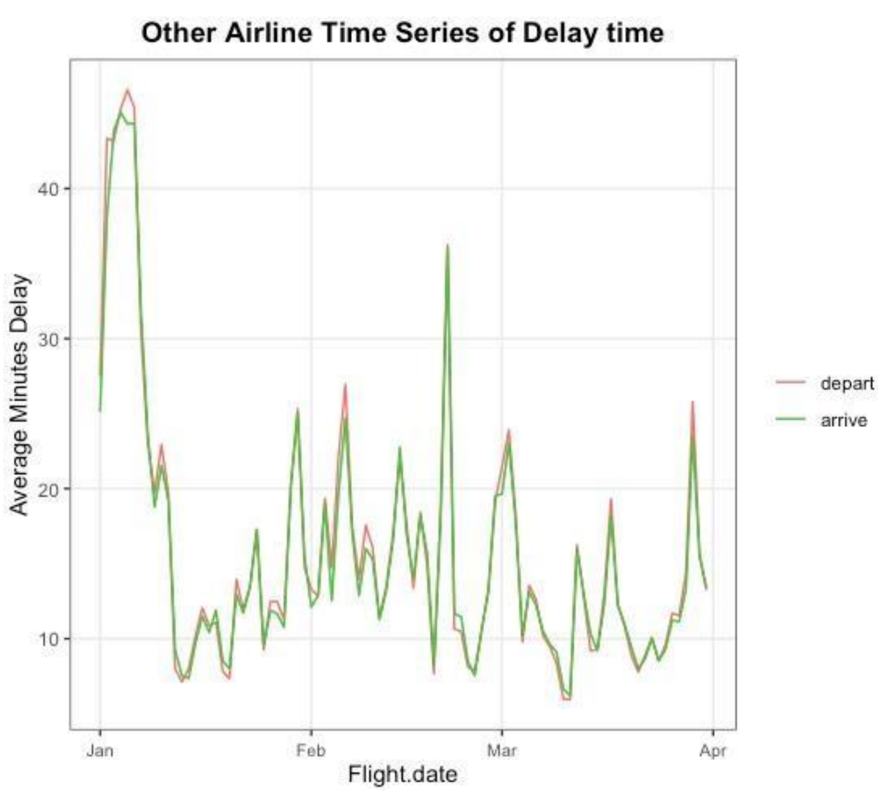




Passenger composition to Classes are similar between South East and Other Airlines with Economy passengers occupying an overwhelming majority.

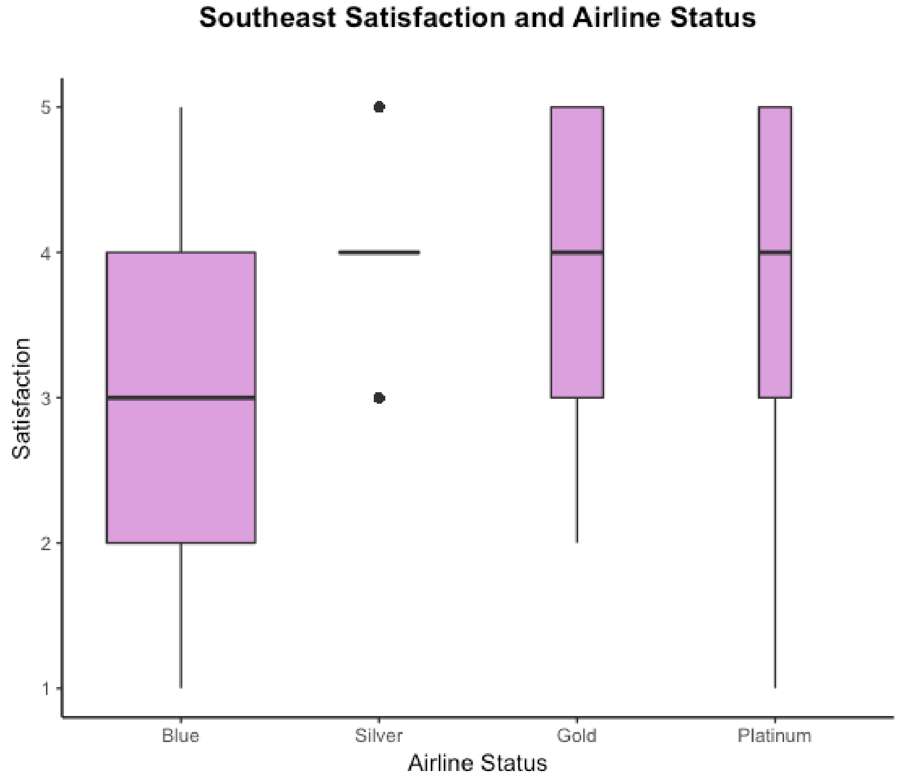
4.10 Time Series for delay in departure and arrival

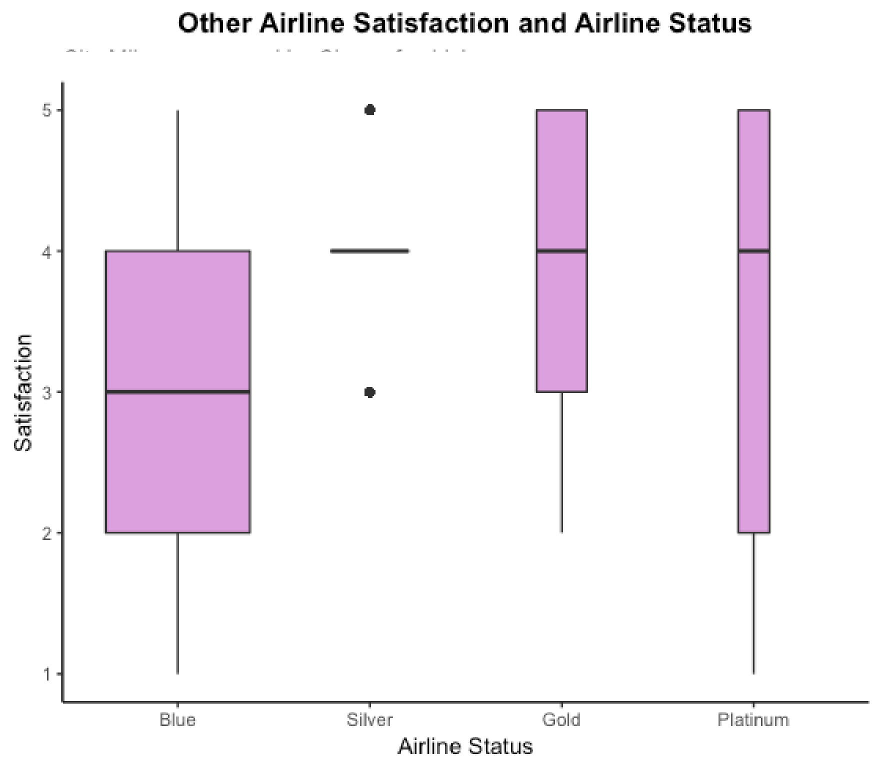




South East Airline has a lesser delay in their arrivals compared to their departures. Compared to industry average South East excels in the Arrival department.

4.11 Customer Satisfaction based on different airline status





South East Airline's customer satisfaction distribution are similar to Other Airline. With some distinctions where the Blue passengers are visibly less satisfied than other classes. In addition, South East performed better for their platinum customers as their 25th percentile score is a 3 compared to Others' score of 2.

5 Data Analysis

5.1 Multiple linear regression

As per the results we get from running a multiple linear regression we notice that for the flights that were "Not Cancelled", customer satisfaction is dependent on a myriad of factors, namely airline status ("Gold, Silver, Platinum"), Delay time exceeding 9 hours, delay of more than 5 minutes on arrival and copious amount of airport shopping and surprisingly the male population. Some of significant variables that we've found interesting for example included age and price sensitivity. Our interpretation was that as passengers gets older they tend to be more sensitive in a negative way with any changes in their flight schedule or price fluctuation. Another observation we had was that passengers sitting in Economy /Plus Class of the flights have more dissatisfied customers due to various factors such as less leg room, crammed up seats etc. Other significant factors are Mileage tickets based on a loyalty card, personal travel, year of first flight, flights per annum and eating and drinking at the airport.

Coefficients:	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-5.89E+00	1.36E+00	-4.323	1.54E-05	***
x$Age	-2.25E-03	1.28E-04	-17.523	< 2e-16	***
x$Airline.StatusGold	4.42E-01	7.47E-03	59.083	< 2e-16	***
x$Airline.StatusPlatinum	2.66E-01	1.16E-02	22.866	< 2e-16	***
x$Airline.StatusSilver	6.20E-01	5.21E-03	119.003	< 2e-16	***
x$GenderMale	1.32E-01	4.20E-03	31.464	< 2e-16	***
x$Price.Sensitivity	-4.04E-02	3.75E-03	-10.769	< 2e-16	***
x$Year.of.First.Flight	4.86E-03	6.79E-04	7.159	8.18E-13	***
x$No.of.Flights.p.a.	-3.27E-03	1.53E-04	-21.332	< 2e-16	***
x$Type.of.TravelMileage tickets	-1.47E-01	7.78E-03	-18.865	< 2e-16	***
x$Type.of.TravelPersonal Travel	-1.08E+00	4.98E-03	-216.15	< 2e-16	***
x$Shopping.Amount.at.Airport	1.63E-04	3.83E-05	4.258	2.06E-05	***
x$Eating.and.Drinking.at.Airport	-8.59E-05	3.96E-05	-2.168	0.0301	*
x$ClassEco	-7.72E-02	7.38E-03	-10.459	< 2e-16	***
x$ClassEco Plus	-6.99E-02	9.47E-03	-7.378	1.61E-13	***
x$DeparturetimeClassLate	5.70E-02	6.83E-03	8.338	< 2e-16	***
x$DeparturetimeClassNormal	5.69E-02	5.01E-03	11.366	< 2e-16	***
x$Arrival.Delay.greater.5.Minsyes	-3.42E-01	4.25E-03	-80.434	< 2e-16	***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

For the flight "Cancelled" category we observe that "Silver" airline status, customers traveling on mileage tickets based on loyalty cards, Eating and drinking options at the airport significantly affect customer satisfaction. On the contrary Age, "Platinum" airline status, customers that booked personal travel tickets or are in economy/plus have a significant negative affect on the customer satisfaction score. These factors upon increasing will further reduce the customer satisfaction score.

Interestingly, the customers traveling in Eco/Plus are dissatisfied in both the cases but a little less for “Not Cancelled.”.

Coefficients:	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	3.6762343	0.0932835	39.409	< 2e-16	***
x1$Airline.StatusGold	0.0667407	0.0596011	1.12	0.2629	
x1$Airline.StatusPlatinum	-0.3663962	0.0829498	-4.417	1.04E-05	***
x1$Airline.StatusSilver	0.7260149	0.0378798	19.166	< 2e-16	***
x1$Age	-0.0034322	0.0007265	-4.724	2.45E-06	***
x1$Type.of.TravelMileage tickets	0.1933178	0.0485971	3.978	7.16E-05	***
x1$Type.of.TravelPersonal Travel	-0.6755999	0.030207	-22.366	< 2e-16	***
x1$ClassEco	-0.2178703	0.0876747	-2.485	0.013	*
x1$ClassEco Plus	-0.1915649	0.0958823	-1.998	0.0458	*
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1					

Comparison between flights cancelled for South East Airline and Other:

The variables that are in common for both categories were Age, Type of Travel: mileage tickets, Type of Travel: personal travel, passengers travelling in economy, and economy plus class. With mileage tickets and personal travel being the most significant.
The only differences in variables that occurred between these two categories were Flights P.A. for Other Airlines and shopping amount at airport for Southeast Airlines. Implying that Southeast Air passenger’s satisfaction level for canceled flights are more influenced by airport shopping than the number of flights they take per year.

Comparison between flights not cancelled for South East Airline and Other:

Every single variable that is not Airline Name are significant for both categories with minimum differences between the variables. Implying that customer satisfaction would be more dependent on the different type of airline they fly with.

Other Airline Filght Cancelled

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.7770133	0.1010766	37.368	< 2e-16 ***
x1$Age	-0.0023299	0.0007974	-2.922	0.00351 **
x1$No.of.Flights.p.a.	-0.0021935	0.0010433	-2.103	0.03561 *
x1$Type.of.TravelMileage.tickets	0.20707	0.0525756	3.939	8.43e-05 ***
x1$Type.of.TravelPersonal Travel	-0.6816508	0.0329593	-20.682	< 2e-16 ***
x1$ClassEco	-0.2110297	0.0947023	-2.228	0.02595 *
x1$ClassEco Plus	-0.2153135	0.103558	-2.079	0.03771 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7169 on 2394 degrees of freedom
Multiple R-squared: 0.2314, Adjusted R-squared: 0.2295
F-statistic: 120.1 on 6 and 2394 DF, p-value: < 2.2e-16

Other Airline Filght Not Cancelled

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-1.19E+01	1.68E+00	-7.115	1.13e-12 ***
x$Age	-1.45E-03	1.73E-04	-8.387	< 2e-16 ***
x$GenderMale	1.37E-01	5.19E-03	26.434	< 2e-16 ***
x$Price.Sensitivity	-6.95E-02	4.63E-03	-15.01	< 2e-16 ***
x$Year.of.First.Flight	7.92E-03	8.36E-04	9.472	< 2e-16 ***
x$No.of.Flights.p.a.	-5.09E-03	1.91E-04	-26.704	< 2e-16 ***
x$X..of.Flight.with.other.Airlines	2.79E-03	3.18E-04	8.782	< 2e-16 ***
x$Type.of.TravelMileage tickets	-2.18E-01	9.57E-03	-22.737	< 2e-16 ***
x$Type.of.TravelPersonal Travel	-1.17E+00	6.11E-03	-191.541	< 2e-16 ***
x$No..of.other.Loyalty.Cards	-1.08E-02	2.64E-03	-4.096	4.22e-05 ***
x$Shopping.Amount.at.Airport	2.38E-04	4.71E-05	5.05	4.44e-07 ***
x$Eating.and.Drinking.at.Airport	3.33E-04	4.86E-05	6.845	7.67e-12 ***
x$ClassEco	-8.18E-02	9.08E-03	-9.01	< 2e-16 ***
x$ClassEco Plus	-1.11E-01	1.17E-02	-9.498	< 2e-16 ***
x$Airline.NameCool&Young Airlines Inc.	8.04E-02	2.51E-02	3.199	0.00138 **
x$Airline.NameEnjoyFlying Air Services	2.81E-02	1.10E-02	2.558	0.01054 *
x$Airline.NameFlyFast Airways Inc.	1.20E-02	9.06E-03	1.326	0.185
x$Airline.NameFlyHere Airways	3.76E-02	1.85E-02	2.029	0.04241 *
x$Airline.NameFlyToSun Airlines Inc.	8.65E-02	1.63E-02	5.302	1.15e-07 ***
x$Airline.NameGoingNorth Airlines Inc.	-4.62E-02	2.29E-02	-2.019	0.04353 *
x$Airline.NameNorthwest Business Airlines Inc.	5.54E-02	9.38E-03	5.907	3.50e-09 ***
x$Airline.NameOnlyJets Airlines Inc.	1.63E-03	1.35E-02	0.121	0.90395
x$Airline.NameOursin Airlines Inc.	4.34E-02	1.03E-02	4.215	2.50e-05 ***
x$Airline.NamePaul Smith Airlines Inc.	5.79E-02	9.79E-03	5.918	3.27e-09 ***
x$Airline.NameSigma Airlines Inc.	5.09E-02	8.79E-03	5.789	7.12e-09 ***
x$Airline.NameSoutheast Airlines Co.	7.25E-02	1.07E-02	6.786	1.16e-11 ***
x$Airline.NameWest Airways Inc.	1.38E-01	2.19E-02	6.306	2.88e-10 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7812 on 99000 degrees of freedom
Multiple R-squared: 0.3592, Adjusted R-squared: 0.3591
F-statistic: 2135 on 26 and 99000 DF, p-value: < 2.2e-16

Southeast Flight Cancelled

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.7428954	0.1008078	37.129	< 2e-16 ***
x1$Age	-0.0026753	0.0007754	-3.45	0.000570 ***
x1$Type.of.TravelMileage tickets	0.1952111	0.0522875	3.733	0.000193 ***
x1$Type.of.TravelPersonal Travel	-0.698074	0.0323112	-21.605	< 2e-16 ***
x1$Shopping.Amount.at.Airport	0.0005321	0.0002706	1.966	0.049369 *
x1$ClassEco	-0.2156724	0.0947161	-2.277	0.022872 *
x1$ClassEco Plus	-0.2116514	0.1035422	-2.044	0.041052 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.717 on 2394 degrees of freedom
Multiple R-squared: 0.2312, Adjusted R-squared: 0.2293
F-statistic: 120 on 6 and 2394 DF, p-value: < 2.2e-16

Southeast Flight Not Cancelled

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-1.19E+01	1.68E+00	-7.106	1.21e-12 ***
x$Age	-1.45E-03	1.73E-04	-8.381	< 2e-16 ***
x$GenderMale	1.37E-01	5.19E-03	26.367	< 2e-16 ***
x$Price.Sensitivity	-6.94E-02	4.63E-03	-14.978	< 2e-16 ***
x$Year.of.First.Flight	7.93E-03	8.36E-04	9.482	< 2e-16 ***
x$No.of.Flights.p.a.	-5.08E-03	1.91E-04	-26.613	< 2e-16 ***
x$X..of.Flight.with.other.Airlines	2.80E-03	3.18E-04	8.792	< 2e-16 ***
x$Type.of.TravelMileage tickets	-2.17E-01	9.58E-03	-22.684	< 2e-16 ***
x$Type.of.TravelPersonal Travel	-1.17E+00	6.12E-03	-191.472	< 2e-16 ***
x$No..of.other.Loyalty.Cards	-1.09E-02	2.64E-03	-4.116	3.87e-05 ***
x$Shopping.Amount.at.Airport	2.39E-04	4.71E-05	5.079	3.81e-07 ***
x$Eating.and.Drinking.at.Airport	3.36E-04	4.86E-05	6.91	4.88e-12 ***
x$ClassEco	-8.17E-02	9.08E-03	-8.995	< 2e-16 ***
x$ClassEco Plus	-1.11E-01	1.17E-02	-9.51	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7818 on 99013 degrees of freedom
Multiple R-squared: 0.3583, Adjusted R-squared: 0.3582
F-statistic: 4252 on 13 and 99013 DF, p-value: < 2.2e-16

Comparison between airline status for flights cancelled:

One interesting observation we had was that passenger satisfaction with different airline status react differently when their flights are cancelled: with Gold and Platinum members being more appeased by their shopping experiences in airports, while Silver male members care a little less than Silver female members, and then finally Blue members who are just universally upset regardless the other variables.

Comparison between airline status for flights not cancelled:

Again, customer satisfaction level between passengers of all airline status displayed minimum differences between each other, the only variable that can significantly affect the satisfaction level being again, the different types of airline they ride with.

Flight Cancelled: Blue

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.693144	0.097067	38.047	< 2e-16 ***
x1$No.of.Flights.p.a.	-0.002913	0.001015	-2.869	0.00416 **
x1$Type.of.TravelMileage tickets	0.219325	0.05249	4.178	3.04e-05 ***
x1$Type.of.TravelPersonal Travel	-0.704326	0.032083	-21.953	< 2e-16 ***
x1$ClassEco	-0.212269	0.09485	-2.238	0.02532 *
x1$ClassEco Plus	-0.227413	0.103638	-2.194	0.02831 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.718 on 2395 degrees of freedom
Multiple R-squared: 0.2287, Adjusted R-squared: 0.227
F-statistic: 142 on 5 and 2395 DF, p-value: < 2.2e-16

Flight Not Cancelled: Blue

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-1.19E+01	1.68E+00	-7.115	1.13e-12 ***
x$Age	-1.45E-03	1.73E-04	-8.387	< 2e-16 ***
x$GenderMale	1.37E-01	5.19E-03	26.434	< 2e-16 ***
x$Price.Sensitivity	-6.95E-02	4.63E-03	-15.01	< 2e-16 ***
x$Year.of.First.Flight	7.92E-03	8.36E-04	9.472	< 2e-16 ***
x$No.of.Flights.p.a.	-5.09E-03	1.91E-04	-26.704	< 2e-16 ***
x$X..of.Flight.with.other.Airlines	2.79E-03	3.18E-04	8.782	< 2e-16 ***
x$Type.of.TravelMileage tickets	-2.18E-01	9.57E-03	-22.737	< 2e-16 ***
x$Type.of.TravelPersonal Travel	-1.17E+00	6.11E-03	-191.541	< 2e-16 ***
x$No..of.other.Loyalty.Cards	-1.08E-02	2.64E-03	-4.096	4.22e-05 ***
x$Shopping.Amount.at.Airport	2.38E-04	4.71E-05	5.05	4.44e-07 ***
x$Eating.and.Drinking.at.Airport	3.33E-04	4.86E-05	6.845	7.67e-12 ***
x$ClassEco	-8.18E-02	9.08E-03	-9.01	< 2e-16 ***
x$ClassEco Plus	-1.11E-01	1.17E-02	-9.498	< 2e-16 ***
x$Airline.NameCool&Young Airlines Inc.	8.04E-02	2.51E-02	3.199	0.00138 **
x$Airline.NameEnjoyFlying Air Services	2.81E-02	1.10E-02	2.558	0.01054 *
x$Airline.NameFlyFast Airways Inc.	1.20E-02	9.06E-03	1.326	0.185
x$Airline.NameFlyHere Airways	3.76E-02	1.85E-02	2.029	0.04241 *
x$Airline.NameFlyToSun Airlines Inc.	8.65E-02	1.63E-02	5.302	1.15e-07 ***
x$Airline.NameGoingNorth Airlines Inc.	-4.62E-02	2.29E-02	-2.019	0.04353 *
x$Airline.NameNorthwest Business Airlines Inc.	5.54E-02	9.38E-03	5.907	3.50e-09 ***
x$Airline.NameOnlyJets Airlines Inc.	1.63E-03	1.35E-02	0.121	0.90395
x$Airline.NameOursin Airlines Inc.	4.34E-02	1.03E-02	4.215	2.50e-05 ***
x$Airline.NamePaul Smith Airlines Inc.	5.79E-02	9.79E-03	5.918	3.27e-09 ***
x$Airline.NameSigma Airlines Inc.	5.09E-02	8.79E-03	5.789	7.12e-09 ***
x$Airline.NameSoutheast Airlines Co.	7.25E-02	1.07E-02	6.786	1.16e-11 ***
x$Airline.NameWest Airways Inc.	1.38E-01	2.19E-02	6.306	2.88e-10 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7812 on 99000 degrees of freedom
Multiple R-squared: 0.3592, Adjusted R-squared: 0.3591
F-statistic: 2135 on 26 and 99000 DF, p-value: < 2.2e-16

Flight Cancelled: Silver

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.701	0.097737	37.867	< 2e-16 ***
x1$GenderMale	-0.021522	0.031033	-0.694	0.48805
x1$No.of.Flights.p.a.	-0.002808	0.001027	-2.734	0.00629 **
x1$Type.of.TravelMileage tickets	0.219402	0.052496	4.179	3.03e-05 ***
x1$Type.of.TravelPersonal Travel	-0.707083	0.032332	-21.87	< 2e-16 ***
x1$ClassEco	-0.212752	0.094863	-2.243	0.02501 *
x1$ClassEco Plus	-0.231746	0.103837	-2.232	0.02572 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7181 on 2394 degrees of freedom
Multiple R-squared: 0.2288, Adjusted R-squared: 0.2269
F-statistic: 118.4 on 6 and 2394 DF, p-value: < 2.2e-16

Flight Not Cancelled: Silver

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-1.19E+01	1.68E+00	-7.115	1.13e-12 ***
x$Age	-1.45E-03	1.73E-04	-8.387	< 2e-16 ***
x$GenderMale	1.37E-01	5.19E-03	26.434	< 2e-16 ***
x$Price.Sensitivity	-6.95E-02	4.63E-03	-15.01	< 2e-16 ***
x$Year.of.First.Flight	7.92E-03	8.36E-04	9.472	< 2e-16 ***
x$No.of.Flights.p.a.	-5.09E-03	1.91E-04	-26.704	< 2e-16 ***
x$X..of.Flight.with.other.Airlines	2.79E-03	3.18E-04	8.782	< 2e-16 ***
x$Type.of.TravelMileage tickets	-2.18E-01	9.57E-03	-22.737	< 2e-16 ***
x$Type.of.TravelPersonal Travel	-1.17E+00	6.11E-03	-191.541	< 2e-16 ***
x$No..of.other.Loyalty.Cards	-1.08E-02	2.64E-03	-4.096	4.22e-05 ***
x$Shopping.Amount.at.Airport	2.38E-04	4.71E-05	5.05	4.44e-07 ***
x$Eating.and.Drinking.at.Airport	3.33E-04	4.86E-05	6.845	7.67e-12 ***
x$ClassEco	-8.18E-02	9.08E-03	-9.01	< 2e-16 ***
x$ClassEco Plus	-1.11E-01	1.17E-02	-9.498	< 2e-16 ***
x$Airline.NameCool&Young Airlines Inc.	8.04E-02	2.51E-02	3.199	0.00138 **
x$Airline.NameEnjoyFlying Air Services	2.81E-02	1.10E-02	2.558	0.01054 *
x$Airline.NameFlyFast Airways Inc.	1.20E-02	9.06E-03	1.326	0.185
x$Airline.NameFlyHere Airways	3.76E-02	1.85E-02	2.029	0.04241 *
x$Airline.NameFlyToSun Airlines Inc.	8.65E-02	1.63E-02	5.302	1.15e-07 ***
x$Airline.NameGoingNorth Airlines Inc.	-4.62E-02	2.29E-02	-2.019	0.04353 *
x$Airline.NameNorthwest Business Airlines Inc.	5.54E-02	9.38E-03	5.907	3.50e-09 ***
x$Airline.NameOnlyJets Airlines Inc.	1.63E-03	1.35E-02	0.121	0.90395
x$Airline.NameOursin Airlines Inc.	4.34E-02	1.03E-02	4.215	2.50e-05 ***
x$Airline.NamePaul Smith Airlines Inc.	5.79E-02	9.79E-03	5.918	3.27e-09 ***
x$Airline.NameSigma Airlines Inc.	5.09E-02	8.79E-03	5.789	7.12e-09 ***
x$Airline.NameSoutheast Airlines Co.	7.25E-02	1.07E-02	6.786	1.16e-11 ***
x$Airline.NameWest Airways Inc.	1.38E-01	2.19E-02	6.306	2.88e-10 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7812 on 99000 degrees of freedom

Multiple R-squared: 0.3592, Adjusted R-squared: 0.3591
F-statistic: 2135 on 26 and 99000 DF, p-value: < 2.2e-16

Flight Cancelled: Platinum

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.7428954	0.1008078	37.129	< 2e-16 ***
x1$Age	-0.0026753	0.0007754	-3.45	0.000570 ***
x1$Type.of.TravelMileage tickets	0.1952111	0.0522875	3.733	0.000193 ***
x1$Type.of.TravelPersonal Travel	-0.698074	0.0323112	-21.605	< 2e-16 ***
x1$Shopping.Amount.at.Airport	0.0005321	0.0002706	1.966	0.049369 *
x1$ClassEco	-0.2156724	0.0947161	-2.277	0.022872 *
x1$ClassEco Plus	-0.2116514	0.1035422	-2.044	0.041052 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.717 on 2394 degrees of freedom
Multiple R-squared: 0.2312, Adjusted R-squared: 0.2293
F-statistic: 120 on 6 and 2394 DF, p-value: < 2.2e-16

Flight Cancelled: Platinum

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-1.19E+01	1.68E+00	-7.115	1.13e-12 ***
x$Age	-1.45E-03	1.73E-04	-8.387	< 2e-16 ***
x$GenderMale	1.37E-01	5.19E-03	26.434	< 2e-16 ***
x$Price.Sensitivity	-6.95E-02	4.63E-03	-15.01	< 2e-16 ***
x$Year.of.First.Flight	7.92E-03	8.36E-04	9.472	< 2e-16 ***
x$No.of.Flights.p.a.	-5.09E-03	1.91E-04	-26.704	< 2e-16 ***
x$X..of.Flight.with.other.Airlines	2.79E-03	3.18E-04	8.782	< 2e-16 ***
x$Type.of.TravelMileage tickets	-2.18E-01	9.57E-03	-22.737	< 2e-16 ***
x$Type.of.TravelPersonal Travel	-1.17E+00	6.11E-03	-191.541	< 2e-16 ***
x$No..of.other.Loyalty.Cards	-1.08E-02	2.64E-03	-4.096	4.22e-05 ***
x$Shopping.Amount.at.Airport	2.38E-04	4.71E-05	5.05	4.44e-07 ***
x$Eating.and.Drinking.at.Airport	3.33E-04	4.86E-05	6.845	7.67e-12 ***
x$ClassEco	-8.18E-02	9.08E-03	-9.01	< 2e-16 ***
x$ClassEco Plus	-1.11E-01	1.17E-02	-9.498	< 2e-16 ***
x$Airline.NameCool&Young Airlines Inc.	8.04E-02	2.51E-02	3.199	0.00138 **
x$Airline.NameEnjoyFlying Air Services	2.81E-02	1.10E-02	2.558	0.01054 *
x$Airline.NameFlyFast Airways Inc.	1.20E-02	9.06E-03	1.326	0.185
x$Airline.NameFlyHere Airways	3.76E-02	1.85E-02	2.029	0.04241 *
x$Airline.NameFlyToSun Airlines Inc.	8.65E-02	1.63E-02	5.302	1.15e-07 ***
x$Airline.NameGoingNorth Airlines Inc.	-4.62E-02	2.29E-02	-2.019	0.04353 *
x$Airline.NameNorthwest Business Airlines Inc.	5.54E-02	9.38E-03	5.907	3.50e-09 ***
x$Airline.NameOnlyJets Airlines Inc.	1.63E-03	1.35E-02	0.121	0.90395
x$Airline.NameOursin Airlines Inc.	4.34E-02	1.03E-02	4.215	2.50e-05 ***
x$Airline.NamePaul Smith Airlines Inc.	5.79E-02	9.79E-03	5.918	3.27e-09 ***
x$Airline.NameSigma Airlines Inc.	5.09E-02	8.79E-03	5.789	7.12e-09 ***
x$Airline.NameSoutheast Airlines Co.	7.25E-02	1.07E-02	6.786	1.16e-11 ***
x$Airline.NameWest Airways Inc.	1.38E-01	2.19E-02	6.306	2.88e-10 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7812 on 99000 degrees of freedom
Multiple R-squared: 0.3592, Adjusted R-squared: 0.3591
F-statistic: 2135 on 26 and 99000 DF, p-value: < 2.2e-16

Flight Cancelled: Gold

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.7428954	0.1008078	37.129	< 2e-16 ***
x1$Age	-0.0026753	0.0007754	-3.45	0.00057***
x1$Type.of.Tr	0.1952111	0.0522875	3.733	0.000193***
x1$Type.of.Tr	-0.698074	0.0323112	-21.605	< 2e-16***
x1$Shopping.	0.0005321	0.0002706	1.966	0.049369*
x1$ClassEco	-0.2156724	0.0947161	-2.277	0.022872*
x1$ClassEco P	-0.2116514	0.1035422	-2.044	0.041052*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

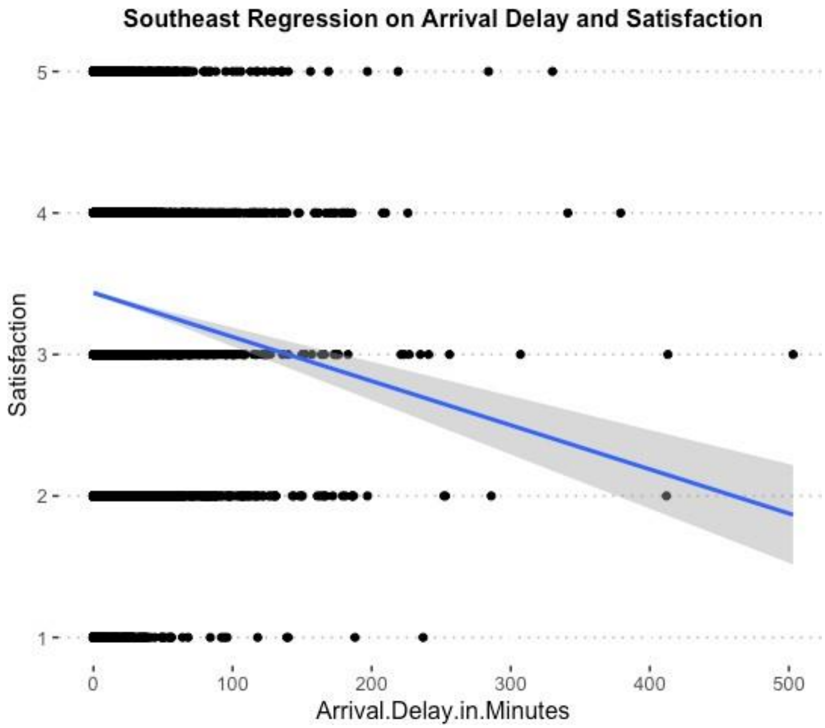
Residual standard error: 0.717 on 2394 degrees of freedom
Multiple R-squared: 0.2312, Adjusted R-squared: 0.2293
F-statistic: 120 on 6 and 2394 DF, p-value: < 2.2e-16

Flight Cancelled: Gold

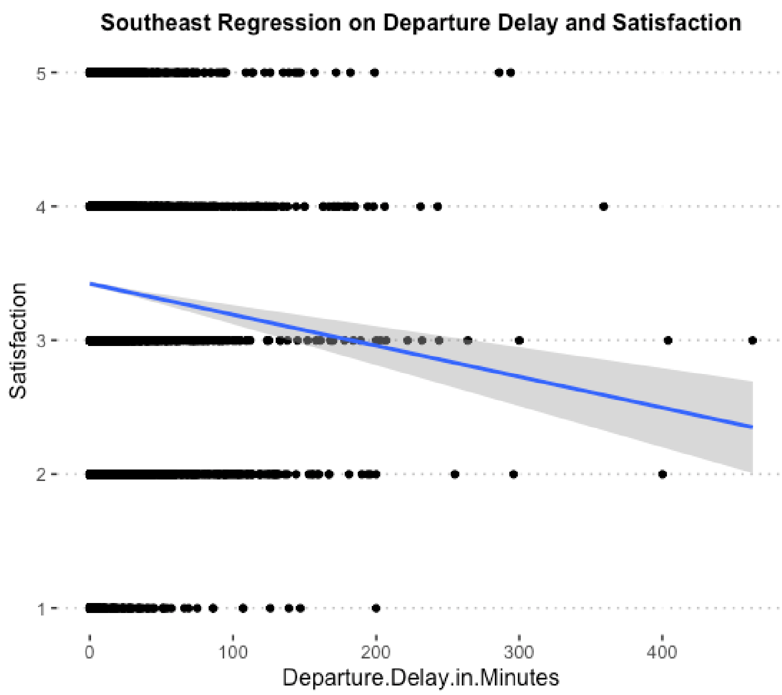
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-1.193e+01	1.677e+00	-7.115	1.13e-12 ***
x$Age	-1.45E-03	1.73E-04	-8.387	< 2e-16 ***
x$GenderMale	1.37E-01	5.19E-03	26.434	< 2e-16 ***
x$Price.Sensitivity	-6.95E-02	4.63E-03	-15.01	< 2e-16 ***
x$Year.of.First.Flight	7.92E-03	8.36E-04	9.472	< 2e-16 ***
x$No.of.Flights.p.a.	-5.09E-03	1.91E-04	-26.704	< 2e-16 ***
x$X..of.Flight.with.other.Airlines	2.79E-03	3.18E-04	8.782	< 2e-16 ***
x$Type.of.TravelMileage tickets	-2.18E-01	9.57E-03	-22.737	< 2e-16 ***
x$Type.of.TravelPersonal Travel	-1.17E+00	6.11E-03	-191.541	< 2e-16 ***
x$No..of.other.Loyalty.Cards	-1.08E-02	2.64E-03	-4.096	4.22e-05 ***
x$Shopping.Amount.at.Airport	2.38E-04	4.71E-05	5.05	4.44e-07 ***
x$Eating.and.Drinking.at.Airport	3.33E-04	4.86E-05	6.845	7.67e-12 ***
x$ClassEco	-8.18E-02	9.08E-03	-9.01	< 2e-16 ***
x$ClassEco Plus	-1.11E-01	1.17E-02	-9.498	< 2e-16 ***
x$Airline.NameCool&Young Airlines Inc.	8.04E-02	2.51E-02	3.199	0.00138 **
x$Airline.NameEnjoyFlying Air Services	2.81E-02	1.10E-02	2.558	0.01054 *
x$Airline.NameFlyFast Airways Inc.	1.20E-02	9.06E-03	1.326	0.185
x$Airline.NameFlyHere Airways	3.76E-02	1.85E-02	2.029	0.04241 *
x$Airline.NameFlyToSun Airlines Inc.	8.65E-02	1.63E-02	5.302	1.15e-07 ***
x$Airline.NameGoingNorth Airlines Inc.	-4.62E-02	2.29E-02	-2.019	0.04353 *
x$Airline.NameNorthwest Business Airlines Inc.	5.54E-02	9.38E-03	5.907	3.50e-09 ***
x$Airline.NameOnlyJets Airlines Inc.	1.63E-03	1.35E-02	0.121	0.90395
x$Airline.NameOursin Airlines Inc.	4.34E-02	1.03E-02	4.215	2.50e-05 ***
x$Airline.NamePaul Smith Airlines Inc.	5.79E-02	9.79E-03	5.918	3.27e-09 ***
x$Airline.NameSigma Airlines Inc.	5.09E-02	8.79E-03	5.789	7.12e-09 ***
x$Airline.NameSoutheast Airlines Co.	7.25E-02	1.07E-02	6.786	1.16e-11 ***
x$Airline.NameWest Airways Inc.	1.38E-01	2.19E-02	6.306	2.88e-10 ***

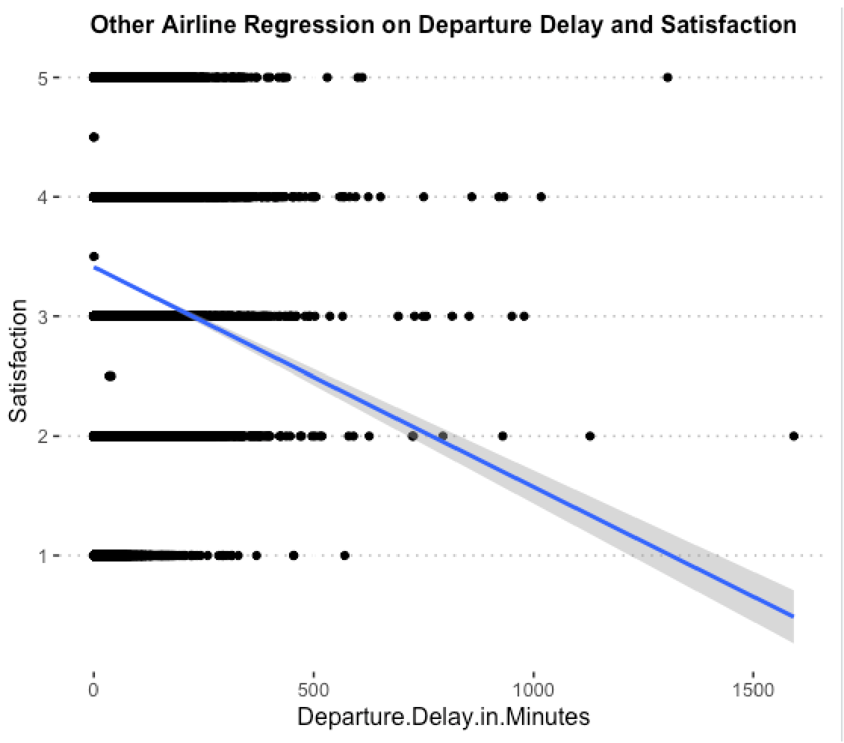
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7812 on 99000 degrees of freedom
Multiple R-squared: 0.3592, Adjusted R-squared: 0.3591
F-statistic: 2135 on 26 and 99000 DF, p-value: < 2.2e-16



For all airlines, we observe that a delay at arrival causes customer dissatisfaction. This can be concluded through the steep decrease in slope for both the regression lines.





A similar trend in customer satisfaction is observed when there is departure delay. The steep linear regression line suggests that as the delay increases, the satisfaction decreases

5.2 Association rules

In real-life, a single factor does not affect customer satisfaction. Various factors lead to customer satisfaction and dissatisfaction. We are implementing association rules on all airlines and south east to better understand which combinations of the 28 attributes boost or hamper the rating.

For implementing the rules:

a. We divided the datasets of Cancelled and Not Cancelled into Happy and Unhappy
b. With the association rules we observe the following results:

1. **All Airlines**

   Not Cancelled:

- **Happy Customers**: Based on our dataset, we observe that "Silver" status customers who are traveling for Business purpose or for whom the arrival delay (in minutes) is "average" are generally happier.

	lhs		rhs	support	confidence	lift	lift count
[1]	{Airline.Status=Silver, Arrival.Delay.in.Minutes=Average}	=>	{happy=Happy}	0.1037059	0.8327649	1.621308	13186
[2]	{Airline.Status=Silver, Type.of.Travel=Business travel}	=>	{happy=Happy}	0.1205996	0.8346851	1.625047	15334
[3]	{Airline.Status=Silver, Arrival.Delay.greater.5.Mins=no}	=>	{happy=Happy}	0.1095023	0.8327153	1.621212	13923
[4]	{Airline.Status=Silver, Arrival.Delay.in.Minutes=Average, Arrival.Delay.greater.5.Mins=no}	=>	{happy=Happy}	0.1037059	0.8327649	1.621308	13186

- **Unhappy Customers**: Customers in our dataset who are in the default "Blue" airline status traveling for personal reasons which happen to be older are on an average unhappy.

	lhs		rhs	support	confidence	lift	count
[1]	{Airline.Status=Blue, Type.of.Travel=Personal Travel}	=>	{happy=Unhappy}	0.228293	0.9624337	1.978841	29027
[2]	{Airline.Status=Blue,Age=High, Type.of.Travel=Personal Travel}	=>	{happy=Unhappy}	0.144147	0.9835784	2.022316	18328
[3]	{Airline.Status=Blue, No.of.Flights.p.a.=High, Type.of.Travel=PersonalTravel}	=>	{happy=Unhappy}	0.141733	0.9818033	2.018666	18021
[4]	{Airline.Status=Blue, Gender=Female, Type.of.Travel=Personal Travel}	=>	{happy=Unhappy}	0.152893	0.9728756	2.00031	19440
[5]	{Airline.Status=Blue, Type.of.Travel=Personal Travel} Departure.Delay.in.Minutes=Average}	=>	{happy=Unhappy}	0.138154	0.9592617	1.972319	17566

Cancelled:

- **Happy Customers:** Based on our dataset, we observe that "Silver" status customers who are traveling for "Business purpose" in the economy are happier. Interestingly, the younger age group who did travel in the "Business" category is happier even when their flights get canceled. One reason we could come up based on these results are younger passengers would prefer an all paid expense trip and hence are happy.

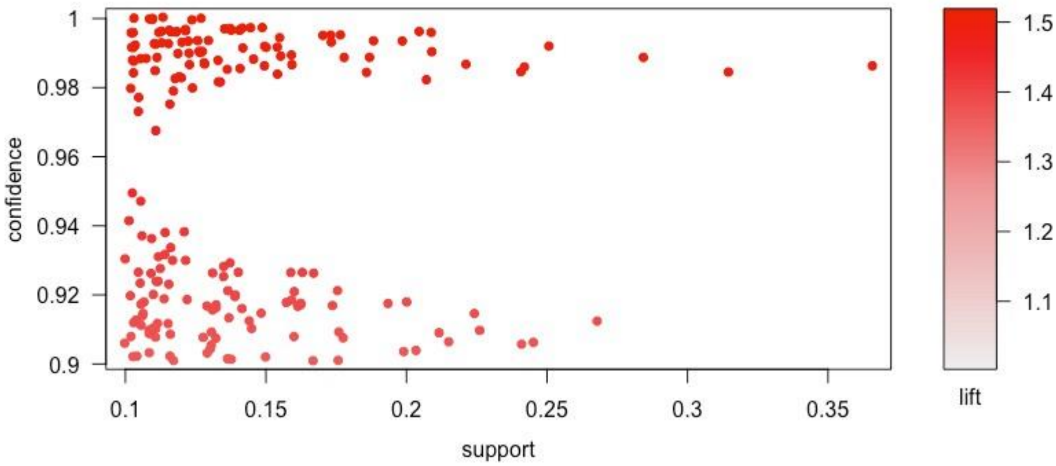
	lhs		rhs	support	confidence	lift	count
[1]	{Airline.Status=Silver}	=>	{happy=Happy}	0.11162	0.716578	2.103304	268
[2]	{Airline.Status=Silver,Class=Eco}	=>	{happy=Happy}	0.100375	0.728097	2.137115	241
[3]	{Age=Low, Type.of.Travel=Business travel}	=>	{happy=Happy}	0.117451	0.5875	1.724435	282
[4]	{Gender=Female, Type.of.Travel=Business travel}	=>	{happy=Happy}	0.141608	0.555556	1.630671	340
[5]	{Age=Low, Type.of.Travel=Business travel,Class=Eco}	=>	{happy=Happy}	0.102874	0.578454	1.697884	247
[6]	{Gender=Female, Type.of.Travel=Business travel,Class=Eco}	=>	{happy=Happy}	0.118284	0.551456	1.618639	284

- **Unhappy Customers:** Based on our results, we observe that passengers traveling for "Personal Purpose" mostly in "Blue" airline status are unhappier if their flight gets canceled. To add, the elder age group travelling under same circumstances are unhappier. One of the reasons could be that they could miss their connecting/subsequent flights which sometimes do not accommodate passengers.

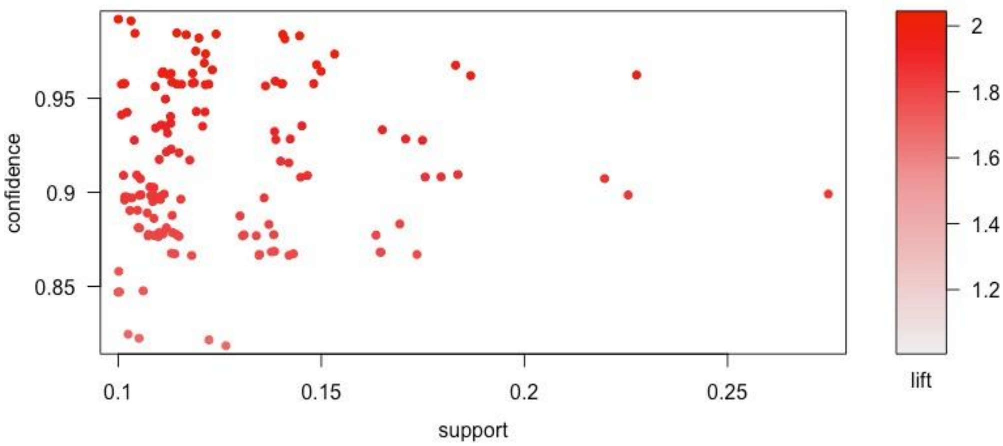
	lhs		rhs	support	confidence	lift	count
[1]	{Age=High, Type.of.Travel=Personal Travel}	=>	{happy=Unhappy}	0.245315	0.9061538	1.3744	589
[2]	{No.of.Flights.p.a.=High, Type.of.Travel=Personal Travel}	=>	{happy=Unhappy}	0.22324	0.9146758	1.387326	536
[3]	{Type.of.Travel=Personal Travel} Shopping.Amount.at.Airport=Average}	=>	{happy=Unhappy}	0.268222	0.9121813	1.383542	644
[2]	{Airline.Status=Blue, Type.of.Travel=Personal Travel}	=>	{happy=Unhappy}	0.366514	0.9865471	1.496336	880
[5]	{Age=High, Type.of.Travel=Personal Travel} No..of.other.Loyalty.Cards=Average}	=>	{happy=Unhappy}	0.240317	0.9058085	1.373876	577
[6]	{Airline.Status=Blue, Age=High, Type.of.Travel=Personal Travel}	=>	{happy=Unhappy}	0.208247	0.9960159	1.510698	500

Plots for association rules for all airlines:

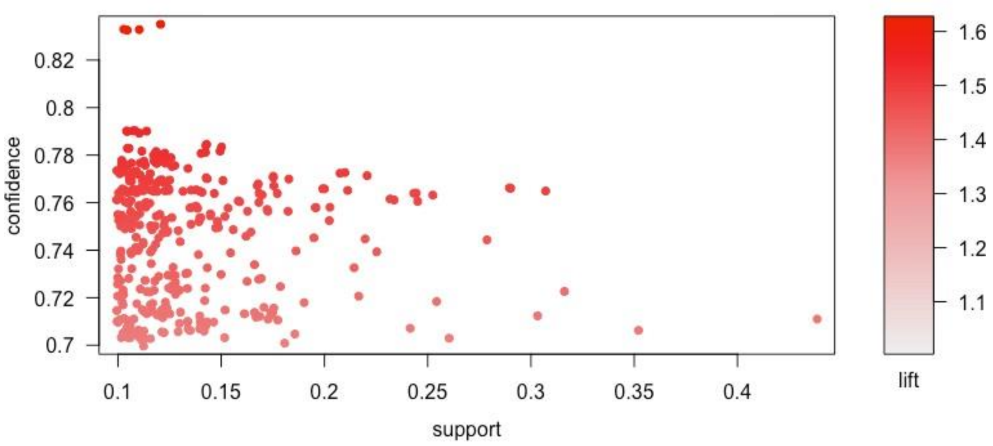
Unhappy customer for Cancelled



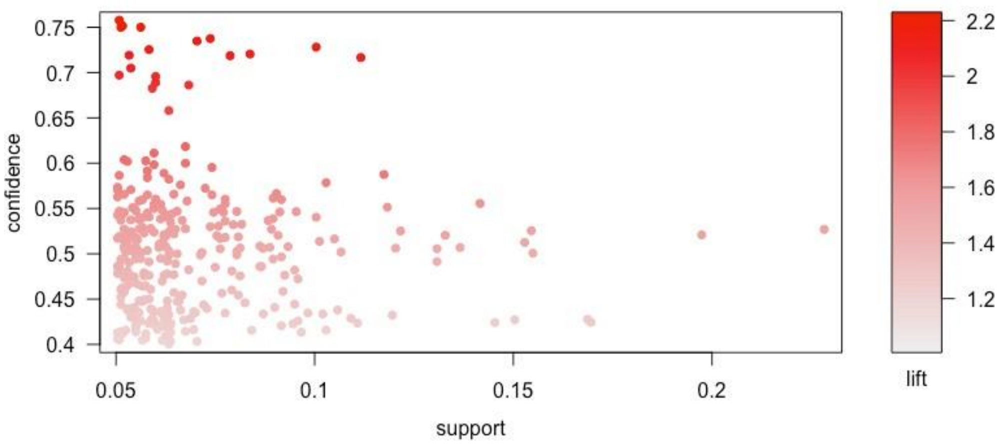
Unhappy customer for Not Cancelled



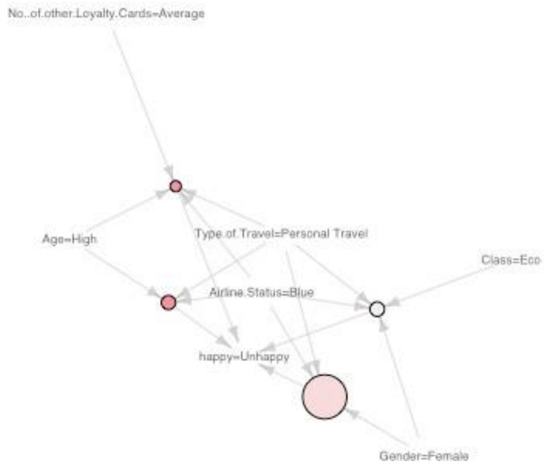
Happy customer for Not



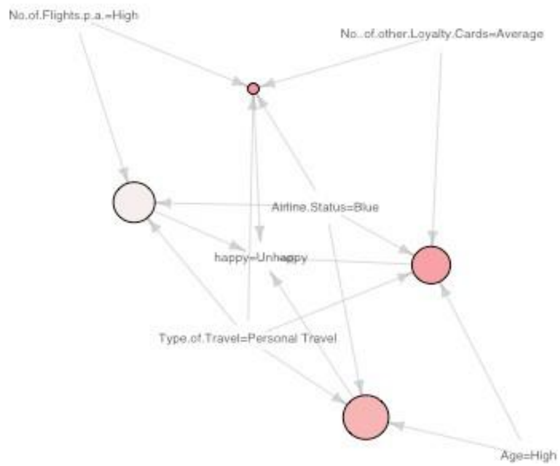
Happy customer for Cancelled



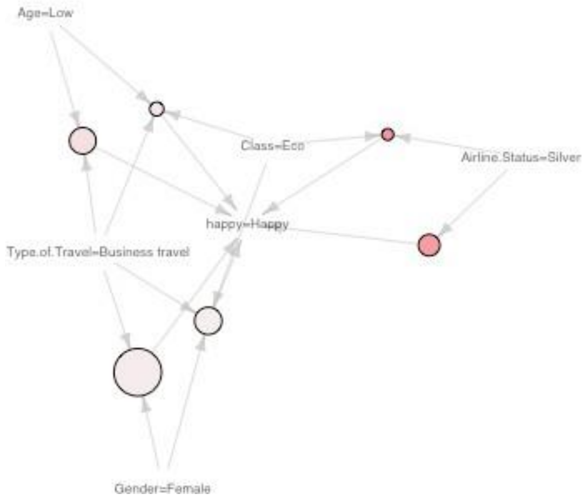
Unhappy customer for Cancelled flight



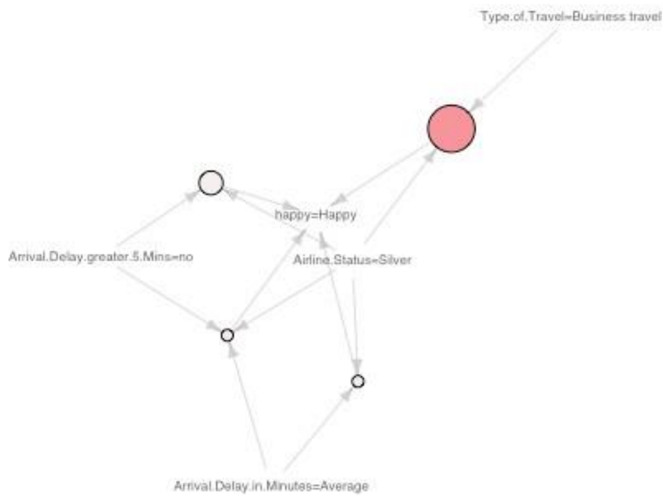
Happy customer for Cancelled



Unhappy customer for Not Cancelled



Happy customer for Not



2. **South East Airlines**

Not Cancelled:

- **Happy Customers**: "Silver" passengers who experience less than average delay (minutes) during departure and arrival are generally happier while travelling with South east airline.

	lhs		rhs	support	confidence	lift	count
[1]	{Airline.Status=Silver, Arrival.Delay.in.Minutes=Average}	=>	{happy= Happy}	0.10635 96	0.8508772	1.603306	194
[2]	{Airline.Status=Silver ,Arrival.Delay.greater.5.Mins=no}	=>	{happy= Happy}	0.11239 04	0.8541667	1.609504	205
[3]	{Airline.Status=Silver, Departure.Delay.in.Minutes=Average}	=>	{happy= Happy}	0.11458 33	0.8393574	1.581599	209
[4]	{Airline.Status=Silver, Arrival.Delay.in.Minutes=Average, Arrival.Delay.greater.5.Mins=no}	=>	{happy= Happy}	0.10635 96	0.8508772	1.603306	194
[5]	{Airline.Status=Silver, Departure.Delay.in.Minutes=Average, Arrival.Delay.greater.5.Mins=no}	=>	{happy= Happy}	0.10307 02	0.8584475	1.61757	188

- **Unhappy Customers**: Based on our dataset, we observe that customers who are traveling for personal reasons on a South east flight that is not cancelled are unhappy while travelling under "Blue" status or in economy. Surprisingly customer's whose origin state is Texas are also unhappy.

	lhs		rhs	support	confidence	lift	count
[1]	{Type.of.Travel=Personal Travel}	=>	{happy= Unhappy}	0.25987	0.90114	1.920188	474
[2]	{Airline.Status=Blue,Type.of.Travel=Personal Travel}	=>	{happy= Unhappy}	0.21382	0.96535	2.057	390
[3]	{Type.of.Travel=Personal Travel,No..of.other.Loyalty.Cards=Average}	=>	{happy= Unhappy}	0.20450	0.91198	1.943285	373
[4]	{Type.of.Travel=Personal Travel,Class=Eco}	=>	{happy= Unhappy}	0.20779	0.89176	1.900209	379
[5]	{Type.of.Travel=Personal Travel,Origin.State=Texas}	=>	{happy= Unhappy}	0.22917	0.89892	1.915466	418

Cancelled:

- **Happy Customers:** Young people mainly females who were travelling with South east airlines for business reasons are generally a little less upset upon their flight being cancelled.

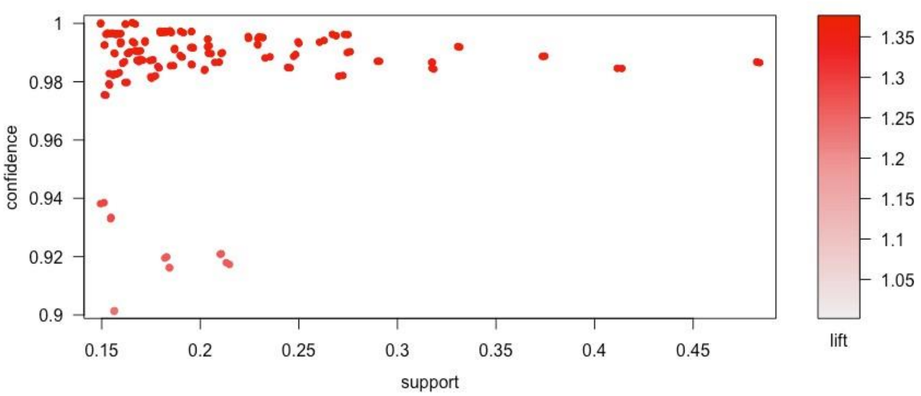
	lhs		rhs	support	confidence	lift	count
[1]	{Age=Low,Type.of.Travel=Business travel}	=>	{happy= Happy}	0.11897	0.56658	2.07936	217
[2]	{Gender=Female,Type.of.Travel=Business travel}	=>	{happy= Happy}	0.13213	0.54036	1.98313	241
[3]	{Age=Low,Type.of.Travel=Business travel,Class=Eco}	=>	{happy= Happy}	0.10197	0.55522	2.03768	186
[4]	{Airline.Status=Blue,Age=Low,Type.of.Travel=Busines s travel}	=>	{happy= Happy}	0.11897	0.56658	2.07936	217
[5]	{Airline.Status=Blue,Gender=Female,Type.of.Travel=B usiness travel}	=>	{happy= Happy}	0.13213	0.54036	1.98313	241
[6]	{Airline.Status=Blue,Age=Low,Type.of.Travel=Busines s travel,Class=Eco}	=>	{happy= Happy}	0.10197	0.55522	2.03768	186

- **Unhappy Customers:** Passengers who are travelling for personal reasons in economy class are more upset when their south east airline flight is cancelled.

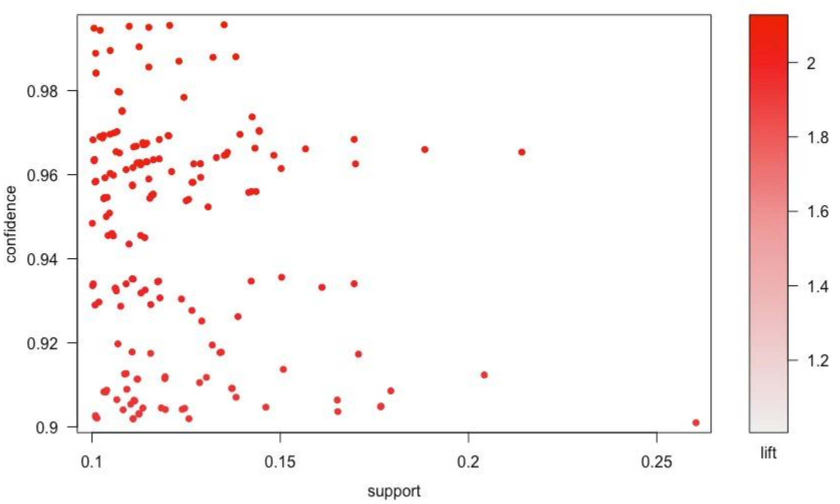
	lhs		rhs	support	confidence	lift	count
[1]	{Type.of.Travel=Personal Travel}	=>	{happy= Unhappy}	0.48246	0.98655	1.35604	880
[2]	{No..of.other.Loyalty.Cards=Average}	=>	{happy= Unhappy}	0.54167	0.74398	1.02262	988
[3]	{Type.of.Travel=Personal Travel,Class=Eco}	=>	{happy= Unhappy}	0.41283	0.98431	1.35297	753
[4]	{Airline.Status=Blue,Type.of.Travel=Personal Travel}	=>	{happy= Unhappy}	0.48246	0.98655	1.35604	880
[5]	{Airline.Status=Blue,No..of.other.Loyalty.Cards=Avera ge}	=>	{happy= Unhappy}	0.54167	0.74398	1.02262	988
[6]	{Airline.Status=Blue,Type.of.Travel=Personal Travel,Class=Eco}	=>	{happy= Unhappy}	0.41283	0.98431	1.35297	753

Plots for association rules for South East airline:

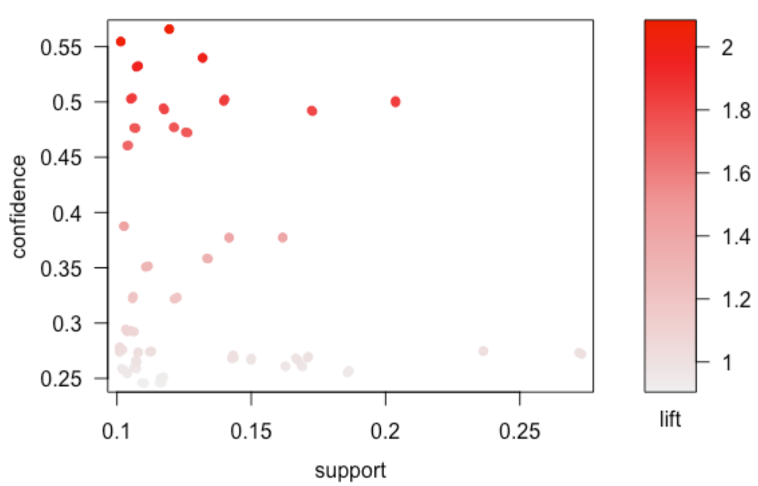
Unhappy customer for Cancelled



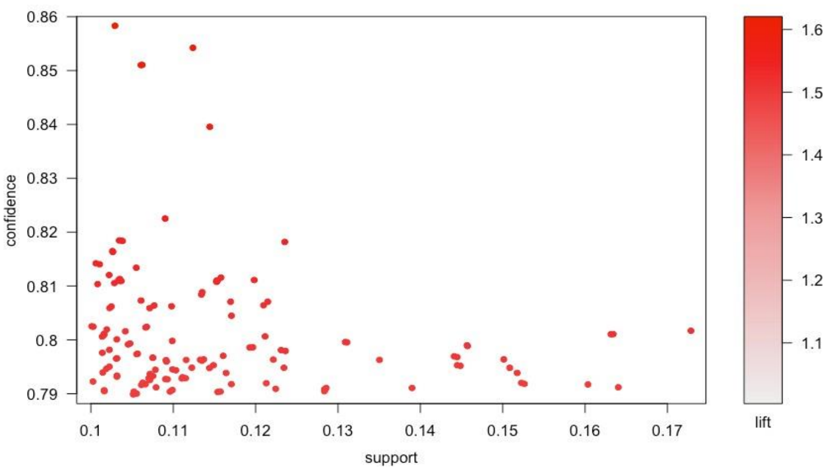
unHappy for no cancel southeast airline



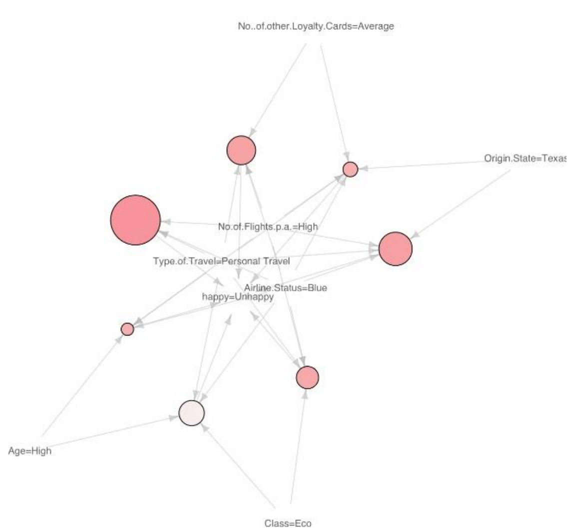
Happy for cancel southeast airline



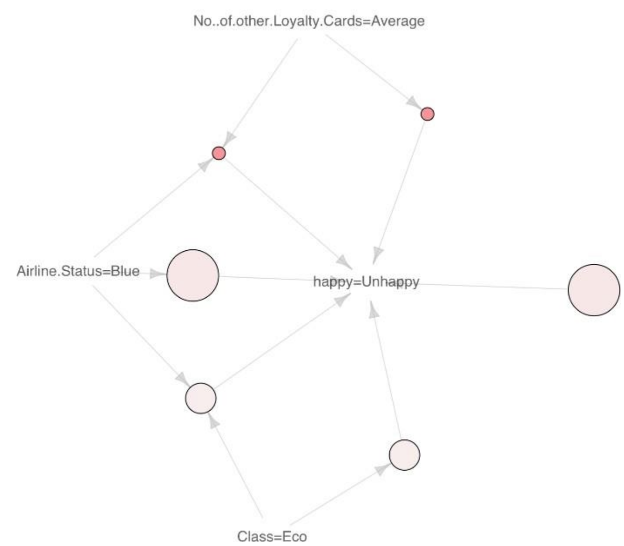
Happy for no cancel southeast airline



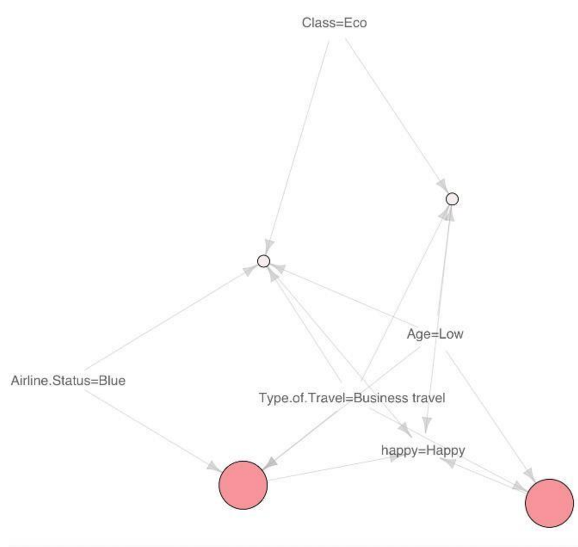
unHappy for no cancel southeast airline



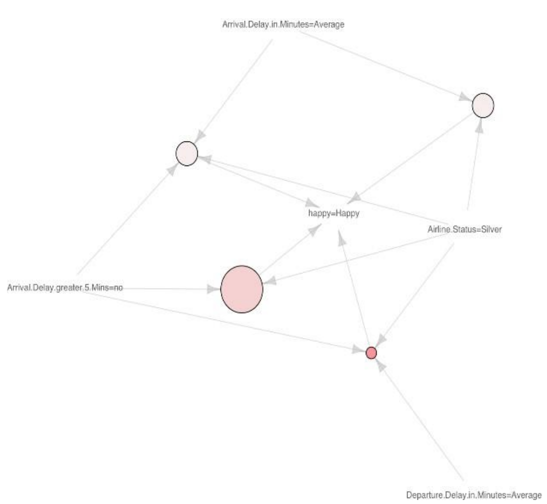
unHappy for cancel southeast airline



Happy for cancel southeast airline



Happy for no cancel southeast airline



5.3 SVM Models

Our model took into consideration all the variables in the dataset to predict if a passenger will be happy or not with various combinations of these variables.

We took 30000 samples from the not canceled flight subset and 2400 samples from the cancelled flights subset. Each sample was divided into 66% and 34% portions that was used as training and test data respectively.

For Not Cancelled Flights:

test_NC.happy	0	1
Happy	678	4425
Unhappy	3331	1566

From the confusion matrix for not cancelled flights, we achieved a good prediction of the test data with an error rate of 22.44%.

$$\frac{678 + 1566}{678 + 1566 + 4425 + 3331} = 0.22$$

For Cancelled Flights:

test_NC1.happy	0	1
Happy	111	158
Unhappy	431	100

For the flight canceled confusion matrix the error rate was slightly higher at 26.4%. The increase in error rate was due to the relatively small number of the sample size, which means less training and test data that can cause the Support Vector Machine analysis results to be more volatile. The calculations are below:

$$\frac{111 + 100}{111 + 100 + 158 + 431} = 0.2638$$

In addition, we also ran another Support Vector Machine model where instead of incorporating all the variables from the data frame, we only included the variables we used for our linear model and in our linear model the variables are more statistically significant than others, this is for Not Cancelled Flights.

Linear Model for

test_NC1.happy	0	1
Happy	328	4872
Unhappy	3099	1701

We discovered that the new SVM model experienced a drop-in error rate from 22.4% to 20.3%. This is significant due to the usage of variables that are statistically significant. Therefore, we would recommend incorporating this prediction model to analyse future test datasets, with additional refinements as we collect more and more training data over time.

We took a step even further by breaking up the customer segment we have for the Cancelled and None Cancelled subset into Airline Status (Platinum, Gold, Silver, and Blue). Results are show below:

Platinum:

Cancelled:

	0	1
Happy	3	1
Unhappy	15	5

Error Rate: 33.3%

Not Cancelled:

	0	1
Happy	13	189
Unhappy	54	60

Error Rate: 23.1%

Gold

Cancelled:

	0	1
Happy	8	6
Unhappy	26	6

Error Rate: 30.4%

Not Cancelled:

	0	1
Happy	22	512
Unhappy	175	125

Error Rate: 17.6%

Silver

Cancelled:

	0	1
Happy	5	79
Unhappy	4	32

Error Rate: 30.8%

Not Cancelled:

	0	1
Happy	36	1515
Unhappy	145	312

Error Rate: 17.3%

Blue

Cancelled:

	0	1
Happy	72	95
Unhappy	379	64

Error Rate: 22.3%

Not Cancelled:

	0	1
Happy	322	2591
Unhappy	2793	1136

Error Rate: 21.3%

In conclusion, we can observe that by further breaking down the Cancelled and Not Cancelled subsets into Airline Status, we get a significantly high and fluctuating error rates ranging from 22.3% to 33.3% for Cancelled passengers of all Status due to the small number of data available for analysis. On the other hand, Not Cancelled passengers exhibited a lower error rate with the range between 17.3% to 23.1%. The Law of Large numbers played a part in stabilizing the numbers for the Not Cancelled numbers.

We would recommend executives to adopt our Not Cancelled Airline Status support vector machine model as the sample sizes are large enough to make an accurate prediction for future passengers. Whereas for Cancelled passengers we would not recommend implementations until sufficient sample size is collected for analysis.

7. Conclusions:

- Upon comparing the customer satisfaction between South East and other airlines, we observe that south east is performing a little better as others when compensating the passengers for their cancelled flights. (figure 4.2 a)

- Demographically, South East airlines have a higher chance of getting cancelled on Arrival and Departure in Georgia and Indiana respectively. The gender distribution suggest that female proportions is more to male and are unhappier as well. The female passengers might encounter inconveniences due to security reasons, physiological needs or even when travelling during pregnancy or with their infants.

- Based on the analysis, we observed that the South East airline customers are unsatisfied with their trips mainly due on these three factors:

  1) Travelling under airline status “Blue”

     Passengers who might booking tickets not based on airline preferences might experience restriction in “seat upgrade” or “seat booking”, “leg room”. These passengers could also miss out on seats due to overbooking of the flight.

  2) Travelling for personal reasons

     Passengers traveling for personal reasons might include families in majority. Cancellation of flights can prove to be inconvenient to their successive plans.

  3) Travelling in economy class

     Passengers booking economical tickets experience similar restriction to “Blue” status passengers. Furthermore, these tickets are mostly non-refundable.

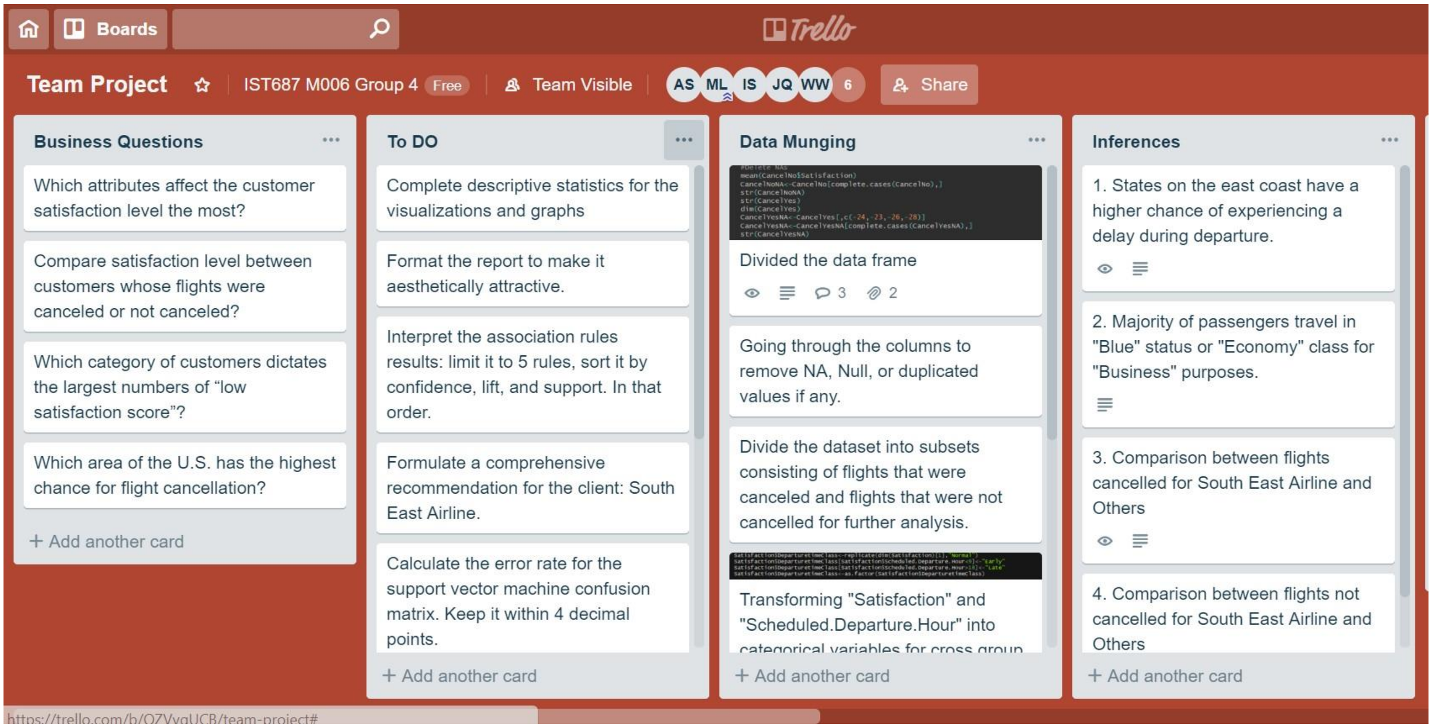
8. Recommendation

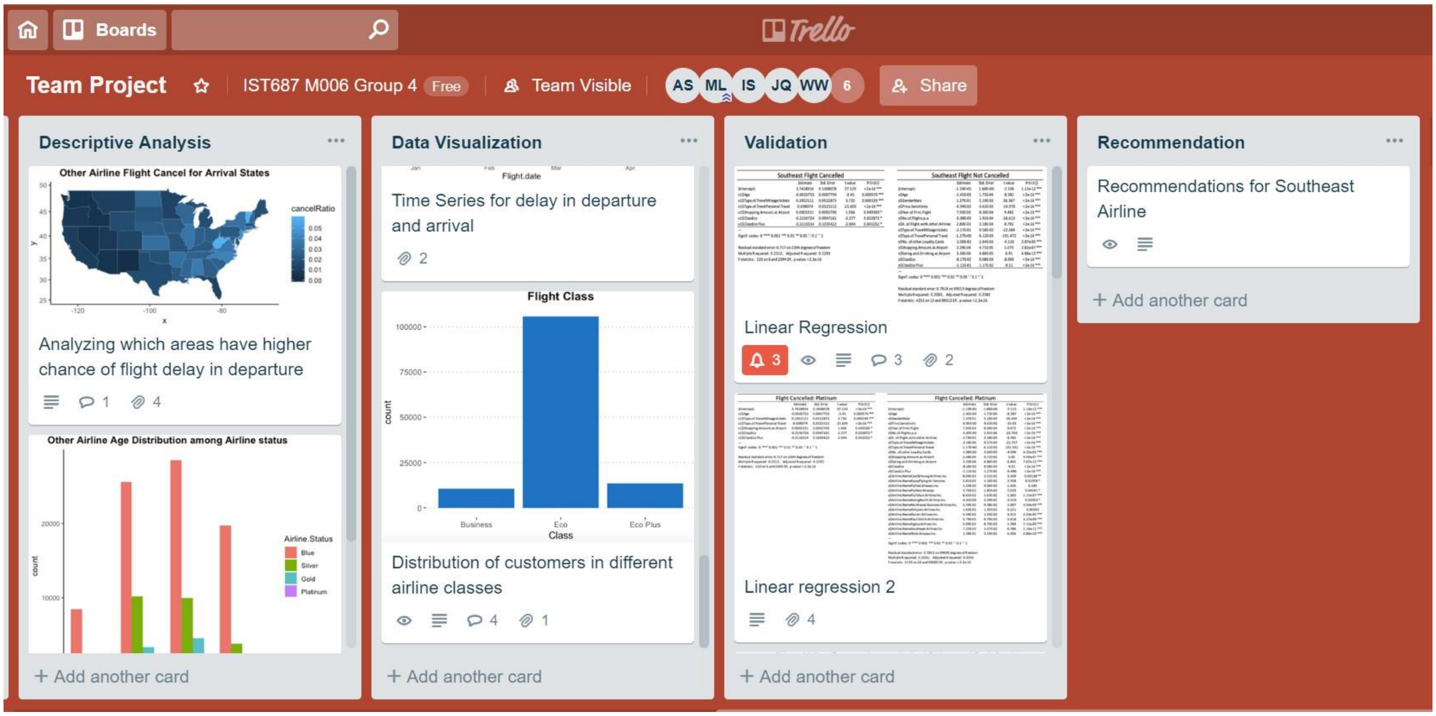
We would recommend South East airlines to implement the following for enhancing the customer satisfaction:

- Investment in technologies which take into consideration unforeseen circumstances such as weather, ground conditions, traffic etc. for improving flight schedules and avoid cancellation.
- Investigate Georgia, Indiana and West Virginia to improve the delays.
- Escort and staff services for women and senior citizens to travel on the airport.
- Advertising promotional offers to retain existing customers of different status and in turn attract new ones.
- More entertainment options (Radio, TV etc.) for every class passengers.
- Better Rewards for frequent flyers based on their age, frequency of travel and feedback.
- Seat Booking option should be made open to all passengers.

8. Trello Board View:

Source: https://trello.com/b/QZVyqUCB/team-project





Appendix:

```
#Question: what variables will effect customer satisfaction
#1. Linear regression, 2. arules, 3. svm model

SatisfactionSurvey=read.csv("SatisfactionSurvey.csv")
library(dplyr)
library(ggplot2)
library(ggpubr)
#1. Clean the data
  str(SatisfactionSurvey)
  Satisfaction<-SatisfactionSurvey
  Satisfaction$Satisfaction<-as.numeric(Satisfaction$Satisfaction)
  colnames(Satisfaction)
  Satisfaction[,c(2,4,9,13,15:21,25,28)]<-lapply(Satisfaction[,c(2,4,9,13,15:21,25,28)],as.factor)
  str(Satisfaction)
  mean(Satisfaction[which(Satisfaction$Flight.cancelled=="Yes"),1])
  #1.1 Add classification variables:
    #1.1.1Satisfaction
    hist(Satisfaction$Satisfaction)
    median(Satisfaction$Satisfaction,na.rm = TRUE)
    dim(Satisfaction)
    Satisfaction$SatiClass<-replicate(dim(Satisfaction)[1],"Average")
    Satisfaction$SatiClass[Satisfaction$Satisfaction<4]<-"Low"
    Satisfaction$SatiClass[Satisfaction$Satisfaction>4]<-"High"
    Satisfaction$SatiClass<-as.factor(Satisfaction$SatiClass)
    #1.1.2departure hours
    hist(Satisfaction$Scheduled.Departure.Hour)
    dim(Satisfaction)
    Satisfaction$DeparturetimeClass<-replicate(dim(Satisfaction)[1],"Normal")
    Satisfaction$DeparturetimeClass[Satisfaction$Scheduled.Departure.Hour<9]<-"Early"
    Satisfaction$DeparturetimeClass[Satisfaction$Scheduled.Departure.Hour>18]<-"Late"
    Satisfaction$DeparturetimeClass<-as.factor(Satisfaction$DeparturetimeClass)
    #1.1.4 add lower case state variables
    Satisfaction$departState<-tolower(Satisfaction$Origin.State)
    Satisfaction$arriveState<-tolower(Satisfaction$Destination.State)
    #1.1.5 add age classification
    Satisfaction$ageClass<-Satisfaction$Age
    Satisfaction$ageClass[which(Satisfaction$Age>15&Satisfaction$Age<=25)]<-"16-25"
    Satisfaction$ageClass[which(Satisfaction$Age<=15)]<-"0-15"
    Satisfaction$ageClass[which(Satisfaction$Age>25&Satisfaction$Age<=35)]<-"26-35"
    Satisfaction$ageClass[which(Satisfaction$Age>35&Satisfaction$Age<=45)]<-"36-45"
    Satisfaction$ageClass[which(Satisfaction$Age>45&Satisfaction$Age<=55)]<-"46-55"
    Satisfaction$ageClass[which(Satisfaction$Age>55&Satisfaction$Age<=65)]<-"56-65"
    Satisfaction$ageClass[which(Satisfaction$Age>65&Satisfaction$Age<=75)]<-"66-75"
    Satisfaction$ageClass[which(Satisfaction$Age>75&Satisfaction$Age<=85)]<-"76-85"
    #add level
    Satisfaction$Airline.Status<-factor(Satisfaction$Airline.Status,levels = c("Blue","Silver","Gold","Platinum"))
```

```
  Satisfaction$Class<-factor(Satisfaction$Class,levels = c("Eco","Eco Plus","Business"))

  SatisfactionAll<-Satisfaction

#Southeast
   Satisfaction<-SatisfactionAll[which(SatisfactionAll$Airline.Name == "Southeast Airlines Co. "),]
  #1.2 Splite the dataset based on flight cancelation status
  CancelYes<-filter(Satisfaction, Flight.cancelled=="Yes")
  CancelNo<-filter(Satisfaction, Flight.cancelled=="No")
  str(CancelYes)
  str(CancelNo)
  quantile(CancelNo$No..of.other.Loyalty.Cards,c(0.4,0.99))
  range(CancelNo$No..of.other.Loyalty.Cards)
  #Delete NAs
  mean(CancelNo$Satisfaction)
  CancelNoNA<-CancelNo[complete.cases(CancelNo),]
  str(CancelNoNA)
  str(CancelYes)
  dim(CancelYes)
  CancelYesNA<-CancelYes[,c(-24,-23,-26,-28)]
  CancelYesNA<-CancelYesNA[complete.cases(CancelYesNA),]
  str(CancelYesNA)


#2. Customer Image
  #2.0 satisfaction with cancellation
  a<-table(Satisfaction$SatiClass,Satisfaction$Flight.cancelled)
  a<-data.frame(a)
  name<-c("Satisfaction.Rating","Cancel","Count")
  colnames(a)<-name
  a$Satisfaction.Rating<-factor(a$Satisfaction.Rating,levels = c("High","Average","Low"))
  a<-a[c(3,1,2,6,4,5),]
  library("plyr")
  library(dplyr)
  a <- ddply(a, "Cancel",transform, accCount=cumsum(Count))

  ggplot(data=a, aes(x=Cancel, y=Count, fill=Satisfaction.Rating)) +
    geom_bar(stat="identity")+ggtitle("Southeast Flight Cancel and Satisfaction")+
    geom_text(aes(y=accCount, label=Count), vjust=1.6, color="white", size=3.5)+
    theme(plot.title = element_text(face = "bold", colour = "black",hjust = 0.5,size = 12))+
    theme_minimal()

  a$Percentage<-a$Count
  a[which(a$Cancel=="No"),5]<-a[which(a$Cancel=="No"),3]/a[3,4]
  a[which(a$Cancel=="Yes"),5]<-a[which(a$Cancel=="Yes"),3]/a[6,4]
  a <- ddply(a, "Cancel",transform, accPct=cumsum(Percentage))
  ggplot(data=a, aes(x=Cancel, y=Percentage, fill=Satisfaction.Rating)) +
    geom_bar(stat="identity")+ggtitle("Southeast Flight Cancel and Satisfaction")+
```

```
    geom_text(aes(y=accPct, label=percent(Percentage)), vjust=1.6, color="white", size=3)+
    scale_y_continuous(labels = scales::percent_format(accuracy = 1))+
    theme(plot.title = element_text(face = "bold", colour = "black",hjust = 0.5,size = 12))+
    theme_minimal()
  #why we split the table, because fight cancelled customers would more likely to give lower grade, and these two
sets of customers have different data
  #2.1histogram for Age
  hist(CancelNoNA$Age) #most customers are at [35-60]
  #2.2Hist for gender
  ggplot(data = Satisfaction,aes(x=Gender, na.rm=TRUE))+
  geom_bar(fill="#0073C2FF")+theme_pubclean()+ggtitle("Gender")+
  theme(plot.title = element_text(face = "bold", colour = "black",hjust = 0.5))
  #2.3 sensitivity
  plot(Satisfaction$Satisfaction,Satisfaction$Price.Sensitivity)# relation between satisfaction and sensitivity
  ggplot(data = Satisfaction,aes(x=Price.Sensitivity, na.rm=TRUE))+
  geom_bar(fill="#0073C2FF")+theme_pubclean()+ggtitle("Customer Sensitivity")+
    theme(plot.title = element_text(face = "bold", colour = "black",hjust = 0.5))#most customers are in 1
  #2.4 Type of travel
  ggplot(data = Satisfaction,aes(x=Type.of.Travel, na.rm=TRUE))+
    geom_bar(fill="#0073C2FF")+theme_pubclean()+
    ggtitle("Type of Travel")+
    theme(plot.title = element_text(face = "bold", colour = "black",hjust = 0.5))#majority are business travel
table(Satisfaction$SatiClass,Satisfaction$Type.of.Travel) #personal travel would more likely to give lower
ratinf, business would more likely to give high rating
prop.table(table(Satisfaction$SatiClass,Satisfaction$Type.of.Travel))
  #2.5 Airline status
  ggplot(data = Satisfaction,aes(x=Airline.Status, na.rm=TRUE))+geom_bar(fill="#0073C2FF")+
    theme_pubclean()+ggtitle("Airline status")+
    theme(plot.title = element_text(face = "bold", colour = "black",hjust = 0.5))#most are "blue"
  table(Satisfaction$SatiClass,Satisfaction$Airline.Status) #
  prop.table(table(Satisfaction$SatiClass,Satisfaction$Airline.Status))

  #2.5 Class
  ggplot(data = Satisfaction,aes(x=Class, na.rm=TRUE))+
    geom_bar(fill="#0073C2FF")+theme_pubclean()+
    ggtitle("Flight Class")+ theme(plot.title = element_text(face = "bold", colour = "black",hjust = 0.5))
  table(Satisfaction$SatiClass,Satisfaction$Class)
  #2.6 Shopping amount customers have flight cancelled and customers do not have flight cancelled
  ggplot(data = CancelNoNA,aes(x=Shopping.Amount.at.Airport, na.rm=TRUE))+
  geom_histogram(fill="#0073C2FF")+theme_pubclean()+xlim(0,1500)+ylim(0,4000)+
  ggtitle("Shopping Amout for Fight Cancelled Customers")+
    theme(plot.title = element_text(face = "bold", colour = "black",hjust = 0.5))

  ggplot(data = CancelYesNA,aes(x=Shopping.Amount.at.Airport, na.rm=TRUE))+
    geom_histogram(fill="#0073C2FF")+theme_pubclean()+
    ggtitle("Shopping Amout for Fight not Cancelled Customers")+
    theme(plot.title = element_text(face = "bold", colour = "black",hjust = 0.5,size = 12))
  #2.7 Eating
```

```
ggplot(data = CancelNoNA,aes(x=Eating.and.Drinking.at.Airport, na.rm=TRUE))+
  geom_histogram(bins=30,fill="#0073C2FF")+theme_pubclean()+
  ggtitle("Eating for Flight Cancelled Customers")+
  theme(plot.title = element_text(face = "bold", colour = "black",hjust = 0.5,size = 12))

ggplot(data = CancelYesNA,aes(x=Eating.and.Drinking.at.Airport, na.rm=TRUE))+
  geom_histogram(bins=30,fill="#0073C2FF")+theme_pubclean()+
  ggtitle("Eating for Fight not Cancelled Customers")+
  theme(plot.title = element_text(face = "bold", colour = "black",hjust = 0.5,size = 12))
#2.8 Number of fights per annual
ggplot(data = CancelYesNA,aes(x=No.of.Flights.p.a., na.rm=TRUE))+
  geom_histogram(bins=30,fill="#0073C2FF")+theme_pubclean()+
  ggtitle("Number of fights per annual")+xlab("Number of fights per annual")+
  theme(plot.title = element_text(face = "bold", colour = "black",hjust = 0.5,size = 12))
#2.9 Airname Name
ggplot(data = Satisfaction,aes(x=Airline.Name, na.rm=TRUE))+geom_bar(fill="#0073C2FF")+theme_pubclean()+
  theme(axis.text.x = element_text(angle = 90,hjust = 1))+
  ggtitle("Airline Name Ditribution")+
  theme(plot.title = element_text(face = "bold", colour = "black",hjust = 0.5,size = 12))

#2.10 Distance
ggplot(data = CancelNoNA,aes(x=CancelNoNA$Flight.Distance, na.rm=TRUE))+
  geom_histogram(bins=30,fill="#0073C2FF")+theme_pubclean()+
  ggtitle("Fight Distance")+xlab("Fight Distance")+
  theme(plot.title = element_text(face = "bold", colour = "black",hjust = 0.5,size = 12))

tapply(CancelNoNA$Flight.Distance,CancelNoNA$SatiClass,mean) #distance does not have much influence on the satisfaction
tapply(CancelYesNA$Flight.Distance,CancelYesNA$SatiClass,mean)

#2.11 Arrival.Delay.greater.5.Mins
tapply(CancelNoNA$Satisfaction,CancelNoNA$Arrival.Delay.greater.5.Mins , sum)
table(CancelNoNA$SatiClass, CancelNoNA$Arrival.Delay.greater.5.Mins)
#2.12 satisfaction and minutes of delay
ggplot(data = CancelNoNA,aes(x=Arrival.Delay.in.Minutes ,y=Satisfaction))+
  geom_point()+geom_smooth(method = "lm",formula = y~x)+theme_pubclean()+
  ggtitle("Southeast Regression on Arrival Delay and Satisfaction")+
  theme(plot.title = element_text(face = "bold", colour = "black",hjust = 0.5,size = 12))

ggplot(data = CancelNoNA,aes(x=Departure.Delay.in.Minutes ,y=Satisfaction))+
  geom_point()+geom_smooth(method = "lm",formula = y~x)+theme_pubclean()+
  ggtitle("Southeast Regression on Departure Delay and Satisfaction")+
  theme(plot.title = element_text(face = "bold", colour = "black",hjust = 0.5,size = 12))

#2.13 ggmap
library(ggmap)
us<-map_data("state")
```

```
 #calculate fight cancellation ratio for each departure state
 cancelYesDepartureStateCount<-CancelYesNA%>%
 group_by(departState)%>%
 summarise(Flight.cancelled=n())
 TotalFlightDepart<-Satisfaction%>%
  group_by(departState)%>%
  summarise(Flight.cancelled=n())
 bbDepart<-left_join(TotalFlightDepart,cancelYesDepartureStateCount,by='departState')
 bbDepart$Flight.cancelled.y[is.na(bbDepart$Flight.cancelled.y)]<-0
 bbDepart$cancelRatio<-bbDepart$Flight.cancelled.y/bbDepart$Flight.cancelled.x
 #calculate fight cancellation ratio for each destination state
 cancelYesArrivalStateCount<-CancelYesNA%>%
  group_by(arriveState)%>%
 summarise(Flight.cancelled=n())
 TotalFlightArrive<-Satisfaction%>%
 group_by(arriveState)%>%
 summarise(Flight.cancelled=n())
 bbArrive<-left_join(TotalFlightArrive,cancelYesArrivalStateCount,by='arriveState')
 bbArrive$Flight.cancelled.y[is.na(bbArrive$Flight.cancelled.y)]<-0
 bbArrive$cancelRatio<-bbArrive$Flight.cancelled.y/bbArrive$Flight.cancelled.x
 #ggplot for fight cancellation ratio for departure state
 ggplot(bbDepart,aes(map_id=departState))+
 geom_map(map=us,aes(fill=cancelRatio))+
 expand_limits(x=us$long,y=us$lat)+coord_map()+
  ggtitle("Southeast Flight Cancel for Departure States")+
  theme(plot.title = element_text(face = "bold", colour = "black",hjust = 0.5))

 #ggplot for flight cancellation ratio for arrival state
 ggplot(bbArrive,aes(map_id=arriveState))+
 geom_map(map=us,aes(fill=cancelRatio))+
 expand_limits(x=us$long,y=us$lat)+coord_map()+
 ggtitle("Southeast Flight Cancel for Arrival States")+
  theme(plot.title = element_text(face = "bold", colour = "black",hjust = 0.5))

#Other Airline
 Satisfaction<-SatisfactionAll[which(SatisfactionAll$Airline.Name != "Southeast Airlines Co. "),]
 #1.2 Splite the dataset based on flight cancelation status
 CancelYes<-filter(Satisfaction, Flight.cancelled=="Yes")
 CancelNo<-filter(Satisfaction, Flight.cancelled=="No")
 str(CancelYes)
 str(CancelNo)
 quantile(CancelNo$No..of.other.Loyalty.Cards,c(0.4,0.99))
 range(CancelNo$No..of.other.Loyalty.Cards)
 #Delete NAs
 mean(CancelNo$Satisfaction)
 CancelNoNA<-CancelNo[complete.cases(CancelNo),]
 str(CancelNoNA)
 str(CancelYes)
```

```
dim(CancelYes)
CancelYesNA<-CancelYes[,c(-24,-23,-26,-28)]
CancelYesNA<-CancelYesNA[complete.cases(CancelYesNA),]
str(CancelYesNA)
#2.0 satisfaction with cancellation
a<-table(Satisfaction$SatiClass,Satisfaction$Flight.cancelled)
a<-data.frame(a)
name<-c("Satisfaction.Rating","Cancel","Count")
colnames(a)<-name
a$Satisfaction.Rating<-factor(a$Satisfaction.Rating,levels = c("High","Average","Low"))
a<-a[c(3,1,2,6,4,5),]
library("plyr")
library(dplyr)
a <- ddply(a, "Cancel",transform, accCount=cumsum(Count))
a$Percentage<-a$Count
a[which(a$Cancel=="No"),5]<-a[which(a$Cancel=="No"),3]/a[3,4]
a[which(a$Cancel=="Yes"),5]<-a[which(a$Cancel=="Yes"),3]/a[6,4]
a <- ddply(a, "Cancel",transform, accPct=cumsum(Percentage))
ggplot(data=a, aes(x=Cancel, y=Percentage, fill=Satisfaction.Rating)) +
  geom_bar(stat="identity")+ggtitle("Other Airline Flight Cancel and Satisfaction")+
  geom_text(aes(y=accPct, label=percent(Percentage)), vjust=1.6, color="white", size=3)+
  scale_y_continuous(labels = scales::percent_format(accuracy = 1))+
  theme(plot.title = element_text(face = "bold", colour = "black",hjust = 0.5,size = 12))+
  theme_minimal()

#2.12 satisfaction and minutes of delay
install.packages("ggthemes")
library(ggthemes)
ggplot(data = CancelNoNA,aes(x=Arrival.Delay.in.Minutes ,y=Satisfaction))+
  geom_point()+geom_smooth(method = "lm",formula = y~x)+theme_pubclean()+
  ggtitle("Other Airline Regression on Arrival Delay and Satisfaction")+
  theme(plot.title = element_text(face = "bold", colour = "black",hjust = 0.5,size = 12))+


ggplot(data = CancelNoNA,aes(x=Departure.Delay.in.Minutes ,y=Satisfaction))+
  geom_point()+geom_smooth(method = "lm",formula = y~x)+theme_pubclean()+
  ggtitle("Other Airline Regression on Departure Delay and Satisfaction")+
  theme(plot.title = element_text(face = "bold", colour = "black",hjust = 0.5,size = 12))

#2.13 ggmap
library(ggmap)
us<-map_data("state")
#calculate fight cancellation ratio for each departure state
cancelYesDepartureStateCount<-CancelYesNA%>%
group_by(departState)%>%
summarise(Flight.cancelled=n())
TotalFlightDepart<-Satisfaction%>%
  group_by(departState)%>%
```

```
  summarise(Flight.cancelled=n())
bbDepart<-left_join(TotalFlightDepart,cancelYesDepartureStateCount,by='departState')
bbDepart$Flight.cancelled.y[is.na(bbDepart$Flight.cancelled.y)]<-0
bbDepart$cancelRatio<-bbDepart$Flight.cancelled.y/bbDepart$Flight.cancelled.x
#calculate fight cancellation ratio for each destination state
cancelYesArrivalStateCount<-CancelYesNA%>%
  group_by(arriveState)%>%
summarise(Flight.cancelled=n())
TotalFlightArrive<-Satisfaction%>%
group_by(arriveState)%>%
summarise(Flight.cancelled=n())
bbArrive<-left_join(TotalFlightArrive,cancelYesArrivalStateCount,by='arriveState')
bbArrive$Flight.cancelled.y[is.na(bbArrive$Flight.cancelled.y)]<-0
bbArrive$cancelRatio<-bbArrive$Flight.cancelled.y/bbArrive$Flight.cancelled.x
#ggplot for fight cancellation ratio for departure state
ggplot(bbDepart,aes(map_id=departState))+
geom_map(map=us,aes(fill=cancelRatio))+
expand_limits(x=us$long,y=us$lat)+coord_map()+
  ggtitle("Other Airline Flight Cancel for Departure States")+
  theme(plot.title = element_text(face = "bold", colour = "black",hjust = 0.5))

#ggplot for flight cancellation ratio for arrival state
ggplot(bbArrive,aes(map_id=arriveState))+
geom_map(map=us,aes(fill=cancelRatio))+
expand_limits(x=us$long,y=us$lat)+coord_map()+
ggtitle("Other Airline Flight Cancel for Arrival States")+
  theme(plot.title = element_text(face = "bold", colour = "black",hjust = 0.5))
#Southeast
#1. Plot for age and gender
str(Satisfaction)
Satisfaction<-SatisfactionAll[which(SatisfactionAll$Airline.Name == "Southeast Airlines Co. "),]
agePlot<-Satisfaction[,c(4,33,2,13,1)]
agePlot$Airline.Status<-factor(agePlot$Airline.Status,levels = c("Blue","Silver","Gold","Platinum"))
agePlot$ageClass<-as.factor(agePlot$ageClass)
agePlot$count<-1
#str(agePlot)
library(plyr)
library(dplyr)
AgeSum<-agePlot %>%
  group_by(ageClass,Gender) %>%
  summarise(count=n())

#ageClass4
agePlot$ageClass4<-Satisfaction$Age
agePlot$ageClass4[which(Satisfaction$Age<=20)]<-"0-20"
agePlot$ageClass4[which(Satisfaction$Age>20&Satisfaction$Age<=40)]<-"21-40"
agePlot$ageClass4[which(Satisfaction$Age>40&Satisfaction$Age<=60)]<-"41-60"
agePlot$ageClass4[which(Satisfaction$Age>60)]<-"61-85"
```

```
#age and status
#1.1.5 add age classification
agePlot$ageClass4<-Satisfaction$Age
agePlot$ageClass4[which(Satisfaction$Age<=20)]<-"0-20"
agePlot$ageClass4[which(Satisfaction$Age>20&Satisfaction$Age<=40)]<-"21-40"
agePlot$ageClass4[which(Satisfaction$Age>40&Satisfaction$Age<=60)]<-"41-60"
agePlot$ageClass4[which(Satisfaction$Age>60)]<-"61-85"

#ggplot
#1 age and gender
library(ggplot2)
ggplot(data = AgeSum,
       mapping = aes(x = ageClass, fill = Gender,
                     y = ifelse(test = Gender == "Male",
                                yes = -count, no = count))) +
  geom_bar(stat = "identity") +
  scale_y_continuous(labels = abs, limits = max(AgeSum$count) * c(-1,1)) +
  labs(y = "Count",x="Age Range") +ggtitle("Age for Southeast Airline")+
  theme(plot.title = element_text(face = "bold", colour = "black",hjust = 0.5))+
  coord_flip()

#2 age and airline status
AgeStatus<-agePlot %>%
  group_by(ageClass4,Airline.Status) %>%
  summarise(count=n())
AgeStatusSum.m<-melt(AgeStatusSum,id.vars = name)
ggplot(AgeStatus,aes(ageClass4,count))+
geom_bar(aes(fill=Airline.Status),position = "dodge",stat = "identity")+
ggtitle("Southeast Age Distribution among Airline status")+xlab("Age Range")+
theme(plot.title = element_text(face = "bold", colour = "black",hjust = 0.5))+
theme(plot.title = element_text(face = "bold", colour = "black",hjust = 0.5))

#3 age and class
AgeClass<-agePlot %>%
  group_by(ageClass4,Class) %>%
  summarise(count=n())

ggplot(AgeClass,aes(ageClass4,count))+
  geom_bar(aes(fill=Class),position = "dodge",stat = "identity")+
  ggtitle("Southeast Age Distribution among different Class")+xlab("Age Range")+
  theme(plot.title = element_text(face = "bold", colour = "black",hjust = 0.5))

#4 airline satisfaction and gender
#install.packages("ggalt")
library(ggalt)
theme_set(theme_classic())
```

```
Airline<-Satisfaction %>%
  group_by(Airline.Name,Gender) %>%
  summarise(mean=mean(Satisfaction,na.rm=TRUE))

Female<-Airline[which(Airline$Gender=="Female"),]
Male<-Airline[which(Airline$Gender=="Male"),]
AirlineGender<-data.frame(Airline=Female$Airline.Name,female=Female$mean,male=Male$mean)

ggplot(AirlineGender, aes(x=female,xend=male,y=Airline, group=Airline))+
  geom_dumbbell(colour="#a3c4dc",
                size=1.5,
                colour_xend="#0e668b") +
  scale_x_continuous()+
  labs(x=NULL,
       y=NULL,
       title="Average Rating Between Gender",
       subtitle="Average Rating: Female vs Male")+
  theme(plot.title = element_text(hjust = 0.5, face = "bold"),
        plot.background=element_rect(fill="#f7f7f7"),
        panel.background=element_rect(fill="#f7f7f7"),
        panel.grid.minor=element_blank(),
        panel.grid.major.y=element_blank(),
        panel.grid.major.x=element_line(),
        axis.ticks=element_blank(),
        legend.position="top",
        panel.border=element_blank())

#delay and time
install.packages("lubridate")
library(lubridate)
theme_set(theme_bw())

departDelay<-Satisfaction %>%
  group_by(Flight.date) %>%
  summarise(mean=mean(Departure.Delay.in.Minutes,na.rm=TRUE))
departDelay$variable<-"depart"
arriveDelay<-Satisfaction %>%
  group_by(Flight.date) %>%
  summarise(mean=mean(Arrival.Delay.in.Minutes,na.rm=TRUE))
arriveDelay$variable<-"arrive"
delay<-rbind(departDelay,arriveDelay)
delay$Flight.date<-as.character(delay$Flight.date)
str(delay)
#install.packages("xts")
library(xts)
delay$Flight.date<-as.Date(delay$Flight.date, '%m/%d/%y')
brks <- delay$Flight.date[seq(1, length(delay$Flight.date), 12)]
lbls <- lubridate::year(brks)
```

```
ggplot(delay,aes(x=Flight.date))+
  geom_line(aes(y=mean,col=variable))+
  labs(title="Southeast Time Series of Delay time",
       y="Average Minutes Delay",
       color=NULL)+
  #scale_x_date(lables=lbls ,breaks = brks)+
  scale_color_manual(labels = c("depart", "arrive"),
                     values = c("depart"="#00ba38", "arrive"="#f8766d")) +
  theme(axis.text.x = element_text( vjust=0.5, size = 8),
        panel.grid.minor = element_blank())+
  theme(plot.title = element_text(face = "bold", colour = "black",hjust = 0.5))

#box plot: satisfation and age
theme_set(theme_classic())
ggplot(agePlot, aes(Airline.Status, Satisfaction))+
  geom_boxplot(varwidth=T, fill="plum") +
  labs(title="Southeast Satisfaction and Airline Status",
       subtitle="City Mileage grouped by Class of vehicle",
       x="Airline Status",
       y="Satisfaction")+
  theme(plot.title = element_text(face = "bold", colour = "black",hjust = 0.5))

#Other
#1. Plot for age and gender
str(SatisfactionAll)
Satisfaction<-SatisfactionAll[which(SatisfactionAll$Airline.Name != "Southeast Airlines Co. "),]
str(Satisfaction)
agePlot<-Satisfaction[,c(4,33,2,13,1)]
agePlot$Airline.Status<-factor(agePlot$Airline.Status,levels = c("Blue","Silver","Gold","Platinum"))
agePlot$ageClass<-as.factor(agePlot$ageClass)
agePlot$count<-1
#str(agePlot)
library(dplyr)
AgeSum<-agePlot %>%
  group_by(ageClass,Gender) %>%
  summarise(count=n())

#ageClass4
agePlot$ageClass4<-Satisfaction$Age
agePlot$ageClass4[which(Satisfaction$Age<=20)]<-"0-20"
agePlot$ageClass4[which(Satisfaction$Age>20&Satisfaction$Age<=40)]<-"21-40"
agePlot$ageClass4[which(Satisfaction$Age>40&Satisfaction$Age<=60)]<-"41-60"
agePlot$ageClass4[which(Satisfaction$Age>60)]<-"61-85"


#age and status
#1.1.5 add age classification
```

```
agePlot$ageClass4<-Satisfaction$Age
agePlot$ageClass4[which(Satisfaction$Age<=20)]<-"0-20"
agePlot$ageClass4[which(Satisfaction$Age>20&Satisfaction$Age<=40)]<-"21-40"
agePlot$ageClass4[which(Satisfaction$Age>40&Satisfaction$Age<=60)]<-"41-60"
agePlot$ageClass4[which(Satisfaction$Age>60)]<-"61-85"

#ggplot
#1 age and gender
library(ggplot2)
ggplot(data = AgeSum,
       mapping = aes(x = ageClass, fill = Gender,
                     y = ifelse(test = Gender == "Male",
                                yes = -count, no = count))) +
  geom_bar(stat = "identity") +
  scale_y_continuous(labels = abs, limits = max(AgeSum$count) * c(-1,1)) +
  labs(y = "Count",x="Age Range") +ggtitle("Age for Other Airline")+
  theme(plot.title = element_text(face = "bold", colour = "black",hjust = 0.5))+
  coord_flip()

#2 age and airline status
AgeStatus<-agePlot %>%
  group_by(ageClass4,Airline.Status) %>%
  summarise(count=n())
AgeStatusSum.m<-melt(AgeStatusSum,id.vars = name)
ggplot(AgeStatus,aes(ageClass4,count))+
geom_bar(aes(fill=Airline.Status),position = "dodge",stat = "identity")+
  ggtitle("Other Airline Age Distribution among Airline status")+xlab("Age Range")+
  theme(plot.title = element_text(face = "bold", colour = "black",hjust = 0.5))+
  theme(plot.title = element_text(face = "bold", colour = "black",hjust = 0.5))

#3 age and class
AgeClass<-agePlot %>%
  group_by(ageClass4,Class) %>%
summarise(count=n())
ggplot(AgeClass,aes(ageClass4,count))+
  geom_bar(aes(fill=Class),position = "dodge",stat = "identity")+
  ggtitle("Other Airline Age Distribution among different Class")+xlab("Age Range")+
  theme(plot.title = element_text(face = "bold", colour = "black",hjust = 0.5))

#4 airline satisfaction and gender
#install.packages("ggalt")
library(ggalt)
theme_set(theme_classic())
Airline<-Satisfaction %>%
  group_by(Airline.Name,Gender) %>%
  summarise(mean=mean(Satisfaction,na.rm=TRUE))

Female<-Airline[which(Airline$Gender=="Female"),]
```

```
Male<-Airline[which(Airline$Gender=="Male"),]
AirlineGender<-data.frame(Airline=Female$Airline.Name,female=Female$mean,male=Male$mean)

ggplot(AirlineGender, aes(x=female,xend=male,y=Airline, group=Airline))+
  geom_dumbbell(colour="#a3c4dc",
                size=1.5,
                colour_xend="#0e668b") +
  scale_x_continuous()+
  labs(x=NULL,
       y=NULL,
       title="Average Rating Between Gender",
       subtitle="Average Rating: Female vs Male")+
  theme(plot.title = element_text(hjust = 0.5, face = "bold"),
        plot.background=element_rect(fill="#f7f7f7"),
        panel.background=element_rect(fill="#f7f7f7"),
        panel.grid.minor=element_blank(),
        panel.grid.major.y=element_blank(),
        panel.grid.major.x=element_line(),
        axis.ticks=element_blank(),
        legend.position="top",
        panel.border=element_blank())

#delay and time
#install.packages("lubridate")
library(lubridate)
theme_set(theme_bw())

departDelay<-Satisfaction %>%
  group_by(Flight.date) %>%
  summarise(mean=mean(Departure.Delay.in.Minutes,na.rm=TRUE))
departDelay$variable<-"depart"
arriveDelay<-Satisfaction %>%
  group_by(Flight.date) %>%
  summarise(mean=mean(Arrival.Delay.in.Minutes,na.rm=TRUE))
arriveDelay$variable<-"arrive"
delay<-rbind(departDelay,arriveDelay)
delay$Flight.date<-as.character(delay$Flight.date)
str(delay)
#install.packages("xts")
library(xts)
delay$Flight.date<-as.Date(delay$Flight.date, '%m/%d/%y')
brks <- delay$Flight.date[seq(1, length(delay$Flight.date), 12)]
lbls <- lubridate::year(brks)

ggplot(delay,aes(x=Flight.date))+
  geom_line(aes(y=mean,col=variable))+
  labs(title="Other Airline Time Series of Delay time",
       y="Average Minutes Delay",
```

```
     color=NULL)+
  #scale_x_date(lables=lbls ,breaks = brks)+
  scale_color_manual(labels = c("depart", "arrive"),
                     values = c("depart"="#00ba38", "arrive"="#f8766d")) +
  theme(axis.text.x = element_text( vjust=0.5, size = 8),
        panel.grid.minor = element_blank())+
  theme(plot.title = element_text(face = "bold", colour = "black",hjust = 0.5))

#box plot: satisfation and age
theme_set(theme_classic())
ggplot(agePlot, aes(Airline.Status, Satisfaction))+
  geom_boxplot(varwidth=T, fill="plum") +
  labs(title="Other Airline Satisfaction and Airline Status",
       subtitle="City Mileage grouped by Class of vehicle",
       x="Airline Status",
       y="Satisfaction")+
  theme(plot.title = element_text(face = "bold", colour = "black",hjust = 0.5))

Satisfaction<-SatisfactionAll
table(SatisfactionAll$Price.Sensitivity,SatisfactionAll$Airline.Status)


#3.linear regression
#3.1 linear regression for fights not cancelled
x<-CancelNoNA
str(x)

aa<-lm(data = x,formula = x$Satisfaction~
         x$Age+
         x$Airline.Status+
         x$Age+
         x$Gender+
         x$Price.Sensitivity+
         x$Year.of.First.Flight+
         x$No.of.Flights.p.a.+
         x$X..of.Flight.with.other.Airlines+
         x$Type.of.Travel+
         x$No..of.other.Loyalty.Cards+
         x$Shopping.Amount.at.Airport+
         x$Eating.and.Drinking.at.Airport+
         x$Class+
         x$Airline.Name+
         x$DeparturetimeClass+
         x$Departure.Delay.in.Minutes+
         x$Flight.time.in.minutes+
         x$Flight.Distance+
         x$Arrival.Delay.greater.5.Mins)
summary(aa)
```

```
#3.1.1 only with star
aa<-lm(data = x,formula = x$Satisfaction~
       x$Age+
       x$Airline.Status+
       x$Age+
       x$Gender+
       x$Price.Sensitivity+
       x$Year.of.First.Flight+
       x$No.of.Flights.p.a.+
       x$Type.of.Travel+
       x$Shopping.Amount.at.Airport+
       x$Eating.and.Drinking.at.Airport+
       x$Class+
       x$DeparturetimeClass+
       x$Arrival.Delay.greater.5.Mins)
summary(aa)
#3.2 linear regression for fight that cancelled
x1<-CancelYesNA
str(x1)
aa<-lm(data = x1,formula = x1$Satisfaction~
       x1$Age+
       x1$Airline.Status+
       x1$Age+
       x1$Gender+
       x1$Price.Sensitivity+
       x1$Year.of.First.Flight+
       x1$No.of.Flights.p.a.+
       x1$X..of.Flight.with.other.Airlines+
       x1$Type.of.Travel+
       x1$No..of.other.Loyalty.Cards+
       x1$Shopping.Amount.at.Airport+
       x1$Eating.and.Drinking.at.Airport+
       x1$Class+
       x1$Airline.Name+
       x1$DeparturetimeClass+
       x1$Flight.Distance)
summary(aa)
#3.2.1 cancel only with star
#Linear model
silvercancelyes<-CancelYesNA[which(x1$Airline.Status=="Silver"),]
summary(silvercancelno)
goldcancelyes<-CancelYesNA[which(x1$Airline.Status=="Gold"),]
bluecancelyes<-CancelYesNA[which(x1$Airline.Status=="Blue"),]
platinumcancelyes<-CancelYesNA[which(x1$Airline.Status=="Platinum"),]
#linear model based on different Airline status
#flight cancelled data
#1
lm_silvercancelyes<-lm(data = silvercancelyes,formula = x1$Satisfaction~
```

```
                    x1$Age+
                    x1$Gender+
                    x1$Price.Sensitivity+
                    x1$Year.of.First.Flight+
                    x1$No.of.Flights.p.a.+
                    x1$X..of.Flight.with.other.Airlines+
                    x1$Type.of.Travel+
                    x1$No..of.other.Loyalty.Cards+
                    x1$Shopping.Amount.at.Airport+
                    x1$Eating.and.Drinking.at.Airport+
                    x1$Class+
                    x1$Airline.Name+
                    x1$DeparturetimeClass+
                    x1$Flight.Distance)
#2
lm_goldcancelyes<-lm(data = goldcancelyes,formula = x1$Satisfaction~
                  x1$Age+
                  x1$Gender+
                  x1$Price.Sensitivity+
                  x1$Year.of.First.Flight+
                  x1$No.of.Flights.p.a.+
                  x1$X..of.Flight.with.other.Airlines+
                  x1$Type.of.Travel+
                  x1$No..of.other.Loyalty.Cards+
                  x1$Shopping.Amount.at.Airport+
                  x1$Eating.and.Drinking.at.Airport+
                  x1$Class+
                  x1$Airline.Name+
                  x1$DeparturetimeClass+
                  x1$Flight.Distance)


#3
lm_platinumcancelyes<-lm(data = platinumcancelyes,formula = x1$Satisfaction~
                      x1$Age+
                      x1$Gender+
                      x1$Price.Sensitivity+
                      x1$Year.of.First.Flight+
                      x1$No.of.Flights.p.a.+
                      x1$X..of.Flight.with.other.Airlines+
                      x1$Type.of.Travel+
                      x1$No..of.other.Loyalty.Cards+
                      x1$Shopping.Amount.at.Airport+
                      x1$Eating.and.Drinking.at.Airport+
                      x1$Class+
                      x1$Airline.Name+
                      x1$DeparturetimeClass+
                      x1$Flight.Distance)
```

```
#4
lm_bluecancelyes<-lm(data = bluecancelyes,formula = x1$Satisfaction~
                    x1$Age+
                    x1$Gender+
                    x1$Price.Sensitivity+
                    x1$Year.of.First.Flight+
                    x1$No.of.Flights.p.a.+
                    x1$X..of.Flight.with.other.Airlines+
                    x1$Type.of.Travel+
                    x1$No..of.other.Loyalty.Cards+
                    x1$Shopping.Amount.at.Airport+
                    x1$Eating.and.Drinking.at.Airport+
                    x1$Class+
                    x1$Airline.Name+
                    x1$DeparturetimeClass+
                    x1$Flight.Distance)


#not cancelled data
silvercancelno<-CancelNoNA[which(x$Airline.Status=="Silver"),]
goldcancelno<-CancelNoNA[which(x$Airline.Status=="Gold"),]
bluecancelno<-CancelNoNA[which(x$Airline.Status=="Blue"),]
platinumcancelno<-CancelNoNA[which(x$Airline.Status=="Platinum"),]
#1
lm_silvercancelno<-lm(data = silvercancelno,formula = x$Satisfaction~
                    x$Age+
                    x$Gender+
                    x$Price.Sensitivity+
                    x$Year.of.First.Flight+
                    x$No.of.Flights.p.a.+
                    x$X..of.Flight.with.other.Airlines+
                    x$Type.of.Travel+
                    x$No..of.other.Loyalty.Cards+
                    x$Shopping.Amount.at.Airport+
                    x$Eating.and.Drinking.at.Airport+
                    x$Class+
                    x$Airline.Name+
                    x$DeparturetimeClass+
                    x$Flight.Distance)
#2
lm_goldcancelno<-lm(data = goldcancelno,formula = x$Satisfaction~
                  x$Age+
                  x$Gender+
                  x$Price.Sensitivity+
                  x$Year.of.First.Flight+
                  x$No.of.Flights.p.a.+
                  x$X..of.Flight.with.other.Airlines+
```

```
                x$Type.of.Travel+
                x$No..of.other.Loyalty.Cards+
                x$Shopping.Amount.at.Airport+
                x$Eating.and.Drinking.at.Airport+
                x$Class+
                x$Airline.Name+
                x$DeparturetimeClass+
                x$Flight.Distance)

#3
lm_bluecancelno<-lm(data = bluecancelno,formula = x$Satisfaction~
                x$Age+
                x$Gender+
                x$Price.Sensitivity+
                x$Year.of.First.Flight+
                x$No.of.Flights.p.a.+
                x$X..of.Flight.with.other.Airlines+
                x$Type.of.Travel+
                x$No..of.other.Loyalty.Cards+
                x$Shopping.Amount.at.Airport+
                x$Eating.and.Drinking.at.Airport+
                x$Class+
                x$Airline.Name+
                x$DeparturetimeClass+
                x$Flight.Distance)
#4
lm_platinumcancelno<-lm(data = platinumcancelno,formula = x$Satisfaction~
                   x$Age+
                   x$Gender+
                   x$Price.Sensitivity+
                   x$Year.of.First.Flight+
                   x$No.of.Flights.p.a.+
                   x$X..of.Flight.with.other.Airlines+
                   x$Type.of.Travel+
                   x$No..of.other.Loyalty.Cards+
                   x$Shopping.Amount.at.Airport+
                   x$Eating.and.Drinking.at.Airport+
                   x$Class+
                   x$Airline.Name+
                   x$DeparturetimeClass+
                   x$Flight.Distance)

summary(lm_bluecancelno)
summary(lm_bluecancelyes)
summary(lm_goldcancelno)
summary(lm_goldcancelyes)
summary(lm_platinumcancelno)
summary(lm_platinumcancelyes)
```

```
summary(lm_silvercancelno)
summary(lm_silvercancelyes)


lm_silvercancelyes1<-lm(data = silvercancelyes,formula = x1$Satisfaction~
                          x1$No.of.Flights.p.a.+
                          x1$Type.of.Travel+
                          x1$Class)
summary(lm_silvercancelyes1)

lm_silvercancelno1<-lm(data = silvercancelno,formula = x$Satisfaction~
                         x$Age+
                         x$Gender+
                         x$Price.Sensitivity+
                         x$Year.of.First.Flight+
                         x$No.of.Flights.p.a.+
                         x$X..of.Flight.with.other.Airlines+
                         x$Type.of.Travel+
                         x$No..of.other.Loyalty.Cards+
                         x$Shopping.Amount.at.Airport+
                         x$Eating.and.Drinking.at.Airport+
                         x$Class+
                         x$Airline.Name)
summary(lm_silvercancelno1)

lm_platinumcancelno1<-lm(data = platinumcancelno,formula = x$Satisfaction~
                           x$Age+
                           x$Gender+
                           x$Price.Sensitivity+
                           x$Year.of.First.Flight+
                           x$No.of.Flights.p.a.+
                           x$X..of.Flight.with.other.Airlines+
                           x$Type.of.Travel+
                           x$No..of.other.Loyalty.Cards+
                           x$Shopping.Amount.at.Airport+
                           x$Eating.and.Drinking.at.Airport+
                           x$Class+
                           x$Airline.Name)
summary(lm_platinumcancelno1)

lm_platinumcancelyes1<-lm(data = platinumcancelyes,formula = x1$Satisfaction~
                            x1$Age+
                            x1$Type.of.Travel+
                            x1$Shopping.Amount.at.Airport+
                            x1$Class)
summary(lm_platinumcancelyes1)
```

```
lm_bluecancelno1<-lm(data = bluecancelno,formula = x$Satisfaction~
                    x$Age+
                    x$Gender+
                    x$Price.Sensitivity+
                    x$Year.of.First.Flight+
                    x$No.of.Flights.p.a.+
                    x$X..of.Flight.with.other.Airlines+
                    x$Type.of.Travel+
                    x$No..of.other.Loyalty.Cards+
                    x$Shopping.Amount.at.Airport+
                    x$Eating.and.Drinking.at.Airport+
                    x$Class+
                    x$Airline.Name)
summary(lm_bluecancelno1)

lm_bluecancelyes1<-lm(data = bluecancelyes,formula = x1$Satisfaction~
                     x1$No.of.Flights.p.a.+
                     x1$Type.of.Travel+
                     x1$Class)
summary(lm_bluecancelyes1)

lm_goldcancelno1<-lm(data = goldcancelno,formula = x$Satisfaction~
                    x$Age+
                    x$Gender+
                    x$Price.Sensitivity+
                    x$Year.of.First.Flight+
                    x$No.of.Flights.p.a.+
                    x$X..of.Flight.with.other.Airlines+
                    x$Type.of.Travel+
                    x$No..of.other.Loyalty.Cards+
                    x$Shopping.Amount.at.Airport+
                    x$Eating.and.Drinking.at.Airport+
                    x$Class+
                    x$Airline.Name)
summary(lm_goldcancelno1)

lm_goldcancelyes<-lm(data = goldcancelyes,formula=x1$Satisfaction~
                    x1$Age+
                    x1$Type.of.Travel+
                    x1$Shopping.Amount.at.Airport+
                    x1$Class)
summary(lm_goldcancelyes)

southeastcancelyes<-CancelYesNA[which(x1$Airline.Names=="Southeast Airlines Co. "),]
southeastcancelno<-CancelNoNA[which(x$Airline.Names=="Southeast Airlines Co. "),]
without_secancelyes<-CancelYesNA[which(x1$Airline.Names!="Southeast Airlines Co. "),]
without_secancelno<-CancelNoNA[which(x$Airline.Names!="Southeast Airlines Co. "),]
southeastcancelyes<-lm(data=southeastcancelyes,formula = x1$Satisfaction~
```

```
                  x1$Age+
                  x1$Type.of.Travel+
                  x1$Shopping.Amount.at.Airport+
                  x1$Class)

summary(southeastcancelyes)
southeastcancelno<-lm(data=southeastcancelno,formula = x$Satisfaction~
                 x$Age+
                 x$Gender+
                 x$Price.Sensitivity+
                 x$Year.of.First.Flight+
                 x$No.of.Flights.p.a.+
                 x$X..of.Flight.with.other.Airlines+
                 x$Type.of.Travel+
                 x$No..of.other.Loyalty.Cards+
                 x$Shopping.Amount.at.Airport+
                 x$Eating.and.Drinking.at.Airport+
                 x$Class)
summary(southeastcancelno)

without_secancelyes1<-lm(data = without_secancelyes,formula = x$Satisfaction~
                   x$Age+
                   x$Gender+
                   x$Price.Sensitivity+
                   x$Year.of.First.Flight+
                   x$No.of.Flights.p.a.+
                   x$X..of.Flight.with.other.Airlines+
                   x$Type.of.Travel+
                   x$No..of.other.Loyalty.Cards+
                   x$Shopping.Amount.at.Airport+
                   x$Eating.and.Drinking.at.Airport+
                   x$Class+
                   x$Airline.Name)
summary(without_secancelyes1)
without_secancelno1<-lm(data =without_secancelno,formula = x1$Satisfaction~
                  x1$Age+
                  x1$No.of.Flights.p.a.+
                  x1$Type.of.Travel+
                  x1$Class)
summary(without_secancelno1)

#4 arules
#4.1 arules for not cancel
#4.1.1
#based on the linear regression result, we choose
#Age, Airline.Status, Gender, Price.Sensitivity, Year.of.First.Flight, No.of.Flights.p.a.,
#Type.of.Travel, Shopping.Amount.at.Airport ,Eating.and.Drinking.at.Airport,
#Class, DeparturetimeClass, Arrival.Delay.greater.5.Minsyes
```

```
#4.1.2 convert into classification variables:
#4.1.2.1 Age
yy=step(aa,direction="both")
#fuction
fuc = function(vec){
  q<-quantile(vec,c(0.4,0.6))
  vBuckets = replicate(length(vec), "Average")
  vBuckets[vec > q[2]] = "High"
  vBuckets[vec < q[1]] = "Low"
  vBuckets=as.factor(vBuckets)
  return(vBuckets)
}
#delet variable
CancelNoNA=CancelNoNA[,-30]
str(CancelNoNA)
newdf=CancelNoNA[,c(-8,-14,-15,-16,-25)]
str(newdf)
newdf_1=newdf



newdf_1$Price.Sensitivity<-replicate(dim(newdf)[1],"Average")

newdf_1$Price.Sensitivity[newdf$Price.Sensitivity<3]<-"Low"
newdf_1$Price.Sensitivity[newdf$Price.Sensitivity>3]<-"High"
newdf_1$Price.Sensitivity<-as.factor(newdf_1$Price.Sensitivity)
#find int variable and transfer
str(newdf_1)
df_index=rep(NA,24)
for( i in 1:24){
  df_index[i]=is.integer(newdf[,i])
}

int_index=which(df_index=="TRUE")
length(int_index)
for (i in int_index){
  newdf_1[,i]=fuc(newdf[,i])
}
newdf_1$happy=replicate(dim(newdf_1)[1],"Happy")
newdf_1$happy[newdf_1$Satisfaction<4]="Unhappy"
newdf_1$happy=as.factor(newdf_1$happy)
newdf_1=newdf_1[,c(-1,-24)]
str(newdf_1)
library(arules)
library(arulesViz)
#happy
rules_df=as(newdf_1,"transactions")
```

```
  ruleset_high=apriori(newdf_1,parameter=list(support=0.1,confidence=0.7),appearance =
list(rhs=("happy=Happy")))
  plot(ruleset_high,main="Happy consumer for no cancel flight")
  summary(ruleset_high)
  inspect(ruleset_high)
  #unhappy
  ruleset_low=apriori(newdf_1,parameter=list(support=0.1,confidence=0.8),appearance =
list(rhs=("happy=Unhappy")))
  plot(ruleset_low,,main="Unhappy consumer for no cancel flight")
  summary(ruleset_low)
  inspect(ruleset_low)
  ########### by status
  levels(newdf_1$Airline.Status)
  status_index_bl=which(newdf_1$Airline.Status=="Blue")
  status_index_gold=which(newdf_1$Airline.Status=="Gold")
  status_index_Pla=which(newdf_1$Airline.Status=="Platinum")
  status_index_sil=which(newdf_1$Airline.Status=="Silver")
  str(status_index_bl)
  rules_df_bl=as(newdf_1[status_index_bl,-1],"transactions")
  rules_df_gol=as(newdf_1[status_index_gold,-1],"transactions")
  rules_df_pla=as(newdf_1[status_index_Pla,-1],"transactions")
  rules_df_sil=as(newdf_1[status_index_sil,-1],"transactions")


  #cancel no happy
ruleset_high_bl=apriori(rules_df_bl,parameter=list(support=0.1,confidence=0.73),appearance =
list(rhs=("happy=Happy")))
  plot(ruleset_high_bl)
  inspect(ruleset_high_bl)

  ruleset_high_gol=apriori(rules_df_gol,parameter=list(support=0.1,confidence=0.88),appearance =
list(rhs=("happy=Happy")))
  plot(ruleset_high_gol)
  inspect(ruleset_high_gol)

  ruleset_high_pla=apriori(rules_df_pla,parameter=list(support=0.1,confidence=0.85),appearance =
list(rhs=("happy=Happy")))
  plot(ruleset_high_pla)
  inspect(ruleset_high_pla)

  ruleset_high_sli=apriori(rules_df_sil,parameter=list(support=0.1,confidence=0.9),appearance =
list(rhs=("happy=Happy")))
  plot(ruleset_high_sli)
  inspect(ruleset_high_sli)
  #cancel no Unhappy
```

```
  ruleset_high_bl=apriori(rules_df_bl,parameter=list(support=0.1,confidence=0.985),appearance =
list(rhs=("happy=Unhappy")))
  plot(ruleset_high_bl)
  inspect(ruleset_high_bl)

  ruleset_high_gol=apriori(rules_df_gol,parameter=list(support=0.1,confidence=0.93),appearance =
list(rhs=("happy=Unhappy")))
  plot(ruleset_high_gol)
  inspect(ruleset_high_gol)

  ruleset_high_pla=apriori(rules_df_pla,parameter=list(support=0.1,confidence=0.85),appearance =
list(rhs=("happy=Unhappy")))
  plot(ruleset_high_pla)
  inspect(ruleset_high_pla)

  ruleset_high_sli=apriori(rules_df_sil,parameter=list(support=0.05,confidence=0.7),appearance =
list(rhs=("happy=Unhappy")))
  plot(ruleset_high_sli)
  inspect(ruleset_high_sli)
  ##########southeast
  status_index_se=which(newdf_1$Airline.Name=="Southeast Airlines Co. ")
  str(newdf_1)
  rules_df_se=as(newdf_1[status_index_bl,-12],"transactions")
ruleset_high_se=apriori(rules_df_se,parameter=list(support=0.2,confidence=0.8),appearance =
list(rhs=("happy=Unhappy")))
  plot(ruleset_high_se,method="graph",main="unHappy for no cancel southeast airline",cex=0.7)
inspect(ruleset_high_se)
ruleset_high_se=apriori(rules_df_se,parameter=list(support=0.1,confidence=0.83),appearance =
list(rhs=("happy=Happy")))
  plot(ruleset_high_se,method="graph",main="Happy for no cancel southeast airline",cex=0.7)
  inspect(ruleset_high_se)
  ##############################
  #cancel Yes
  CancelYesNA=CancelYesNA[,-26]
  str(CancelYesNA)
  newdf1=CancelYesNA[,c(-8,-14,-15,-16,-23)]
  str(newdf1)
  newdf_2=newdf1
  newdf_2$Price.Sensitivity<-replicate(dim(newdf1)[1],"Average")
  newdf_2$Price.Sensitivity[newdf1$Price.Sensitivity<3]<-"Low"
  newdf_2$Price.Sensitivity[newdf1$Price.Sensitivity>3]<-"High"
  newdf_2$Price.Sensitivity<-as.factor(newdf_2$Price.Sensitivity)

  str(newdf_2)
  df_index1=rep(NA,20)
  for( i in 1:20){
    df_index1[i]=is.integer(newdf1[,i])
  }
```

```
  int_index1=which(df_index1=="TRUE")
  length(int_index1)
  for (i in int_index1){
    newdf_2[,i]=fuc(newdf1[,i])
  }
  str(newdf_2)

  newdf_2$happy=replicate(dim(newdf_2)[1],"Happy")
  newdf_2$happy[newdf_2$Satisfaction<4]="Unhappy"
  newdf_2$happy=as.factor(newdf_2$happy)

  newdf_2=newdf_2[,c(-1,-20)]

  rules_df1=as(newdf_2,"transactions")
  ruleset_high1=apriori(newdf_2,parameter=list(support=0.05,confidence=0.4),appearance =
list(rhs=("happy=Happy")))
  summary(ruleset_high1)
  inspect(ruleset_high1)
  plot(ruleset_high1,main="Happy consumer for cancel flight")
ruleset_low1=apriori(newdf_2,parameter=list(support=0.1,confidence=0.9),appearance =
list(rhs=("happy=Unhappy")))
  summary(ruleset_low1)
  inspect(ruleset_low1)
  plot(ruleset_low1,main="Unhappy consumer for cancel flight")
  #cancel yes by status
  levels(newdf_1$Airline.Status)
  status_index_bl=which(newdf_2$Airline.Status=="Blue")
  status_index_gold=which(newdf_2$Airline.Status=="Gold")
  status_index_Pla=which(newdf_2$Airline.Status=="Platinum")
  status_index_sil=which(newdf_2$Airline.Status=="Silver")
  str(status_index_bl)
  rules_df_bl=as(newdf_2[status_index_bl,-1],"transactions")
  rules_df_gol=as(newdf_2[status_index_gold,-1],"transactions")
  rules_df_pla=as(newdf_2[status_index_Pla,-1],"transactions")
  rules_df_sil=as(newdf_2[status_index_sil,-1],"transactions")

  ruleset_high_bl=apriori(rules_df_bl,parameter=list(support=0.1,confidence=0.4),appearance =
list(rhs=("happy=Happy")))
  plot(ruleset_high_bl)
  inspect(ruleset_high_bl)

  ruleset_high_gol=apriori(rules_df_gol,parameter=list(support=0.1,confidence=0.5),appearance =
list(rhs=("happy=Happy")))
  plot(ruleset_high_gol)
  inspect(ruleset_high_gol)
```

```
  ruleset_high_pla=apriori(rules_df_pla,parameter=list(support=0.1,confidence=0.4),appearance =
list(rhs=("happy=Happy")))
  plot(ruleset_high_pla)
  inspect(ruleset_high_pla)

  ruleset_high_sli=apriori(rules_df_sil,parameter=list(support=0.1,confidence=0.85),appearance =
list(rhs=("happy=Happy")))
  plot(ruleset_high_sli)
  inspect(ruleset_high_sli)


  #unhappy
  ruleset_high_bl=apriori(rules_df_bl,parameter=list(support=0.1,confidence=0.985),appearance =
list(rhs=("happy=Unhappy")))
  plot(ruleset_high_bl)
  inspect(ruleset_high_bl)

  ruleset_high_gol=apriori(rules_df_gol,parameter=list(support=0.1,confidence=0.93),appearance =
list(rhs=("happy=Unhappy")))
  plot(ruleset_high_gol)
  inspect(ruleset_high_gol)

  ruleset_high_pla=apriori(rules_df_pla,parameter=list(support=0.3,confidence=0.7),appearance =
list(rhs=("happy=Unhappy")))
  plot(ruleset_high_pla)
  inspect(ruleset_high_pla)

  ruleset_high_sli=apriori(rules_df_sil,parameter=list(support=0.05,confidence=0.45),appearance =
list(rhs=("happy=Unhappy")))
  plot(ruleset_high_sli,method="graph")
  plot(inspect(ruleset_high_sli))




  ##################
  status_index_se=which(newdf_2$Airline.Name=="Southeast Airlines Co. ")
  str(newdf_2)
  rules_df_se=as(newdf_2[status_index_bl,-12],"transactions")
ruleset_high_se=apriori(rules_df_se,parameter=list(support=0.4,confidence=0.74),appearance =
list(rhs=("happy=Unhappy")))
  plot(ruleset_high_se,method="graph",main="unHappy for cancel southeast airline")
  inspect(ruleset_high_se)
  ruleset_high_se=apriori(rules_df_se,parameter=list(support=0.1,confidence=0.54),appearance =
list(rhs=("happy=Happy")))
  plot(ruleset_high_se,method="graph",main="Happy for cancel southeast airline")
```

```
  inspect(ruleset_high_se)

  ###############################
  CancelNoNA$happy=replicate(dim(CancelNoNA)[1],"Happy")
  CancelNoNA$happy[CancelNoNA$Satisfaction<4]="Unhappy"
  CancelNoNA$happy=as.factor(CancelNoNA$happy)



  lg=length(CancelNoNA$SatiClass)
  index=sample(1:lg)
  cp=floor(lg*2/3)
  train=CancelNoNA[index[1:cp],]
  train_NC=train[sample(nrow(train),20000),]
  test=CancelNoNA[index[(cp+1):lg],]
  test_NC=test[sample(nrow(test),10000),]
  str(train_NC)
  str(test)
  library(kernlab)
  train_NC=train_NC[,c(-1,-29)]
  test_NC=test_NC[,c(-1,-29)]
  str(test_NC)
  svm=ksvm(happy~.,kernel="rbfdot",data=train_NC,kpar="automatic",C=5,cross=3,prob.model=T)
svm2=ksvm(happy~+Age+Airline.Status+                Gender+Price.Sensitivity+               Year.of.First.Flight+
No.of.Flights.p.a.+Type.of.Travel+ Shopping.Amount.at.Airport+Eating.and.Drinking.at.Airport+

Class+DeparturetimeClass+Arrival.Delay.greater.5.Mins,kernel="rbfdot",data=train_NC,kpar="automatic",C=5
,cross=3,prob.model=T)
  table(predict(svm,test_NC))
  table(test_NC$happy)
  svm_pr=predict(svm,test_NC,type="vote")
  c_m=data.frame(test_NC$happy,svm_pr[1,])
  table(c_m)
  ##svm status
  svm_b=ksvm(happy~Age+                         Gender+Price.Sensitivity+                         Year.of.First.Flight+
No.of.Flights.p.a.+Type.of.Travel+ Shopping.Amount.at.Airport+Eating.and.Drinking.at.Airport+

Class+DeparturetimeClass+Arrival.Delay.greater.5.Mins,kernel="rbfdot",data=train_NC[which(train_NC$Airli
ne.Status=="Blue"),-1],kpar="automatic",C=5,cross=3,prob.model=T)
svm_pr_b=predict(svm_b,test_NC[which(test_NC$Airline.Status=="Blue"),-1],type="vote")
c_m_b=data.frame(test_NC[which(test_NC$Airline.Status=="Blue"),-1]$happy,svm_pr_b[1,])
  table(c_m_b)

  svm_g=ksvm(happy~Age+                         Gender+Price.Sensitivity+                         Year.of.First.Flight+
No.of.Flights.p.a.+Type.of.Travel+ Shopping.Amount.at.Airport+Eating.and.Drinking.at.Airport+

Class+DeparturetimeClass+Arrival.Delay.greater.5.Mins,kernel="rbfdot",data=train_NC[which(train_NC$Airli
ne.Status=="Gold"),-1],kpar="automatic",C=5,cross=3,prob.model=T)
```

```
svm_pr_g=predict(svm_g,test_NC[which(test_NC$Airline.Status=="Gold"),-1],type="vote")
c_m_g=data.frame(test_NC[which(test_NC$Airline.Status=="Gold"),-1]$happy,svm_pr_g[1,])
table(c_m_g)
svm_p=ksvm(happy~Age+ Gender+Price.Sensitivity+ Year.of.First.Flight+
No.of.Flights.p.a.+Type.of.Travel+ Shopping.Amount.at.Airport+Eating.and.Drinking.at.Airport+

Class+DeparturetimeClass+Arrival.Delay.greater.5.Mins,kernel="rbfdot",data=train_NC[which(train_NC$Airline.Status=="Platinum"),-1],kpar="automatic",C=5,cross=3,prob.model=T)
svm_pr_p=predict(svm_p,test_NC[which(test_NC$Airline.Status=="Platinum"),-1],type="vote")
c_m_p=data.frame(test_NC[which(test_NC$Airline.Status=="Platinum"),-1]$happy,svm_pr_p[1,])
table(c_m_p)
svm_s=ksvm(happy~Age+ Gender+Price.Sensitivity+ Year.of.First.Flight+
No.of.Flights.p.a.+Type.of.Travel+ Shopping.Amount.at.Airport+Eating.and.Drinking.at.Airport+

Class+DeparturetimeClass+Arrival.Delay.greater.5.Mins,kernel="rbfdot",data=train_NC[which(train_NC$Airline.Status=="Silver"),-1],kpar="automatic",C=5,cross=3,prob.model=T)
svm_pr_s=predict(svm_s,test_NC[which(test_NC$Airline.Status=="Silver"),-1],type="vote")
c_m_s=data.frame(test_NC[which(test_NC$Airline.Status=="Silver"),-1],svm_pr_s[1,])
table(c_m_s)


##################
svm_b=ksvm(happy~Age+Airline.Status+ Gender+Price.Sensitivity+ Year.of.First.Flight+
No.of.Flights.p.a.+Type.of.Travel+ Shopping.Amount.at.Airport+Eating.and.Drinking.at.Airport+

Class+Arrival.Delay.greater.5.Mins,kernel="rbfdot",data=train_NC[which(train_NC$Airline.Name=="Southeast Airlines Co. "),],kpar="automatic",C=5,cross=3,prob.model=T)
svm_pr_b=predict(svm_b,test_NC[which(test_NC$Airline.Name=="Southeast Airlines Co. "),],type="vote")
c_m_b=data.frame(test_NC[which(test_NC$Airline.Name=="Southeast Airlines Co. "),-1]$happy,svm_pr_b[1,])
table(c_m_b)



##########################################
CancelYesNA$happy=replicate(dim(CancelYesNA)[1],"Happy")
CancelYesNA$happy[CancelYesNA$Satisfaction<4]="Unhappy"
CancelYesNA$happy=as.factor(CancelYesNA$happy)
lg_1=length(CancelYesNA[which(CancelYesNA$Airline.Status=="Silver")])
index1=sample(1:lg_1)
cp1=floor(lg_1*2/3)
str(CancelYesNA)
train1=CancelYesNA[index1[1:cp1],]
str(train1)
train_NC1=train1[sample(nrow(train1),1500),]
test1=CancelYesNA[index1[(cp1+1):lg_1],]
```

```
test_NC1=test1[sample(nrow(test1),800),]
str(test1)
train_NC1=train_NC1[,c(-1,-25)]
test_NC1=test_NC1[,c(-1,-25)]
str(train_NC1)
svm1=ksvm(happy~Age+ Airline.Status+Gender+Price.Sensitivity+ Year.of.First.Flight+
No.of.Flights.p.a.+Type.of.Travel+
Shopping.Amount.at.Airport+Eating.and.Drinking.at.Airport+Class,,kernel="rbfdot",data=train_NC1,kpar="automatic",C=5,cross=3,prob.model=T)

table(predict(svm1,test_NC1))
table(test_NC1$happy)

svm_pr1=predict(svm1,test_NC1,type="vote")
c_m1=data.frame(test_NC1$happy,svm_pr1[1,])
table(c_m1)
#cancel yes status
svm_b1=ksvm(happy~+Age+ Gender+Price.Sensitivity+ Year.of.First.Flight+
No.of.Flights.p.a.+Type.of.Travel+
Shopping.Amount.at.Airport+Eating.and.Drinking.at.Airport+Class,kernel="rbfdot",data=train_NC1[which(train_NC1$Airline.Status=="Blue"),-1],kpar="automatic",C=5,cross=3,prob.model=T)
svm_pr_b1=predict(svm_b1,test_NC1[which(test_NC1$Airline.Status=="Blue"),-1],type="vote")
c_m_b1=data.frame(test_NC1[which(test_NC1$Airline.Status=="Blue"),]$happy,svm_pr_b1[1,])
table(c_m_b1)


svm_g1=ksvm(happy~+Age+ Gender+Price.Sensitivity+ Year.of.First.Flight+
No.of.Flights.p.a.+Type.of.Travel+ Shopping.Amount.at.Airport+Eating.and.Drinking.at.Airport+
        Class,kernel="rbfdot",data=train_NC1[which(train_NC1$Airline.Status=="Gold"),-
1],kpar="automatic",C=5,cross=3,prob.model=T)
svm_pr_g1=predict(svm_g1,test_NC1[which(test_NC1$Airline.Status=="Gold"),-1],type="vote")
c_m_g1=data.frame(test_NC1[which(test_NC1$Airline.Status=="Gold"),-1]$happy,svm_pr_g1[1,])
table(c_m_g1)


svm_p1=ksvm(happy~+Age+ Gender+Price.Sensitivity+ Year.of.First.Flight+
No.of.Flights.p.a.+Type.of.Travel+ Shopping.Amount.at.Airport+Eating.and.Drinking.at.Airport+
        Class,kernel="rbfdot",data=train_NC1[which(train_NC1$Airline.Status=="Platinum"),-
1],kpar="automatic",C=5,cross=3,prob.model=T)
svm_pr_p1=predict(svm_p1,test_NC1[which(test_NC1$Airline.Status=="Platinum"),-1],type="vote")
c_m_p1=data.frame(test_NC1[which(test_NC1$Airline.Status=="Platinum"),-1]$happy,svm_pr_p1[1,])
table(c_m_p1)


svm_s1=ksvm(happy~+Age+ Gender+Price.Sensitivity+ Year.of.First.Flight+
No.of.Flights.p.a.+Type.of.Travel+ Shopping.Amount.at.Airport+Eating.and.Drinking.at.Airport+
        Class,kernel="rbfdot",data=train_NC1[which(train_NC1$Airline.Status=="Silver"),-
1],kpar="automatic",C=5,cross=3,prob.model=T)
```

```
svm_pr_s1=predict(svm_s1,test_NC1[which(test_NC1$Airline.Status=="Silver"),-1],type="vote")
c_m_s1=data.frame(test_NC1[which(test_NC1$Airline.Status=="Silver"),-1]$happy,svm_pr_s1[1,])
table(c_m_s1)


svm_b=ksvm(happy~.
,kernel="rbfdot",data=train_NC1[which(train_NC1$Airline.Name=="Southeast Airlines Co. "),],kpar="automatic",C=5,cross=3,prob.model=T)
svm_pr_b=predict(svm_b,test_NC1[which(test_NC1$Airline.Name=="Southeast Airlines Co. "),],type="vote")
c_m_b=data.frame(test_NC1[which(test_NC1$Airline.Name=="Southeast Airlines Co. "),-1]$happy,svm_pr_b[1,])
table(c_m_b)
```