

IST687 M009- DATA ANALYSIS FOR CUSTOMER SATISFACTION INDEX

GROUP 3-XUEHAN CHEN
YANQI YAO
ADVAIT RAMESH IYER
WEICHENG WU



Objective:

To identify various trends in the customer data of flights in the US to formulate various business strategies for our client: Southeast Airlines Co.

Methodology:

- Used multiple linear regression analysis and association rules to develop hypothesis
- Used Support Vector Machine algorithm to validate the hypothesis



CONTENT



BUSINESS QUESTIONS

DESCRIPTIVE STATISTICS

DATA ANALYSIS

RECOMMENDATIONS



VARIABLES IN THE MODEL

- Satisfaction (Range: 0-5)
- Airline Status (Range: Blue, Gold, Silver, Platinum)
- Age (Range: 15-85 years)
- Gender
- Price Sensitivity (Range: 0-5)
- Year of First Flight (Range: 2003-2012)
- Number of flights per annum (Range: 0-100)
- Percentage of flights with other airlines (Range: 1-100)
- Type of travel (Range: Business travel, Mileage travel, Personal travel)
- Number of other loyalty cards (Range: 0-12)
- Shopping amount at airport (Range: \$0-\$879)
- Eating and drinking at airport (Range: \$0-\$895)
- Class (Range: Economy, Economy Plus, Business)
- Day of month (Range: 1-31)
- Flight date (Range: 1/1/14 to 3/9/14)
- Airline name: 14 different airlines
- Origin city
- Origin state
- Destination city
- Destination state
- Scheduled departure hour
- Departure delay in minutes (Range: 0-1592 min)
- Arrival delay in minutes (Range: 0-1584 min)
- Flight cancelled (Yes/No)
- Flight time in minutes (Range: 8-669 min)
- Flight distance (Range: 31-4983 miles)
- Arrival delay greater than 5 minutes (Yes/No)



BUSINESS QUESTIONS





BUSINESS QUESTIONS

- A. What are the key reasons for customer satisfaction or dissatisfaction?
- B. Which clusters generally come off as happy customers? Which of them come off as unhappy customers?
- C. What should be the pricing strategy of airline packages which lead to a higher satisfaction level among the unhappy customers?
- D. What can be any strategic improvements which can improve customer satisfaction index?



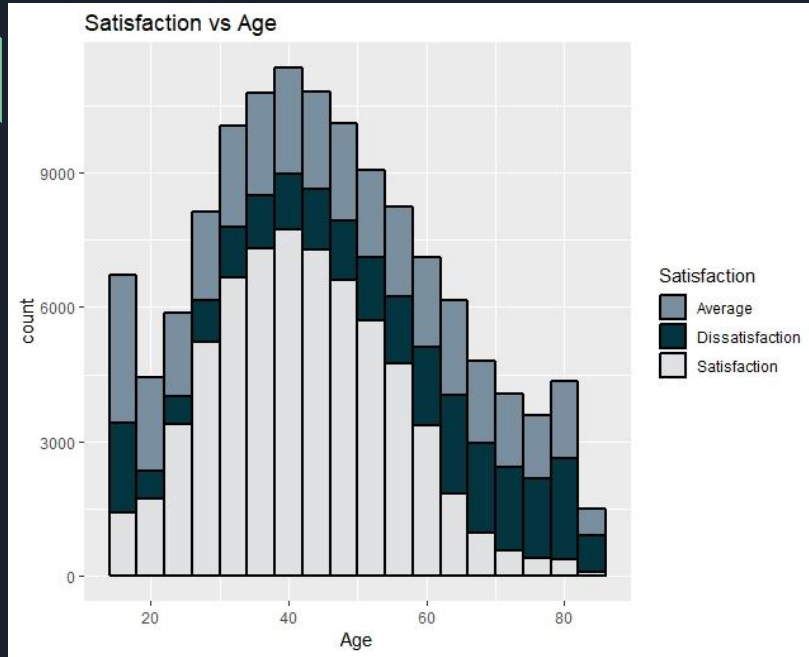
DESCRIPTIVE STATISTICS



```
> summary(C)
```

Satisfaction	Airline Status	Age	Gender	Price Sensitivity
Min. :1.000	Length:127148	Min. :15.00	Length:127148	Min. :0.000
1st Qu.:3.000	Class :character	1st Qu.:33.00	Class :character	1st Qu.:1.000
Median :4.000	Mode :character	Median :45.00	Mode :character	Median :1.000
Mean :3.384		Mean :46.16		Mean :1.275
3rd Qu.:4.000		3rd Qu.:58.00		3rd Qu.:2.000
Max. :5.000		Max. :85.00		Max. :5.000
% of Flight with other Airlines	Type of Travel	Shopping Amount at Airport	Class	
Min. : 1.000	Length:127148	Min. : 0.00	Length:127148	
1st Qu.: 4.000	Class :character	1st Qu.: 0.00	Class :character	
Median : 7.000	Mode :character	Median : 0.00	Mode :character	
Mean : 8.324		Mean :16.23		
3rd Qu.:10.000		3rd Qu.:30.00		
Max. :21.000		Max. :60.00		
Arrival Delay in Minutes	Flight Distance			
Min. : 0.00	Min. : 31.0			
1st Qu.: 0.00	1st Qu.: 363.0			
Median : 0.00	Median : 631.0			
Mean : 15.29	Mean : 795.8			
3rd Qu.: 13.00	3rd Qu.:1027.0			
Max. :500.00	Max. :4983.0			

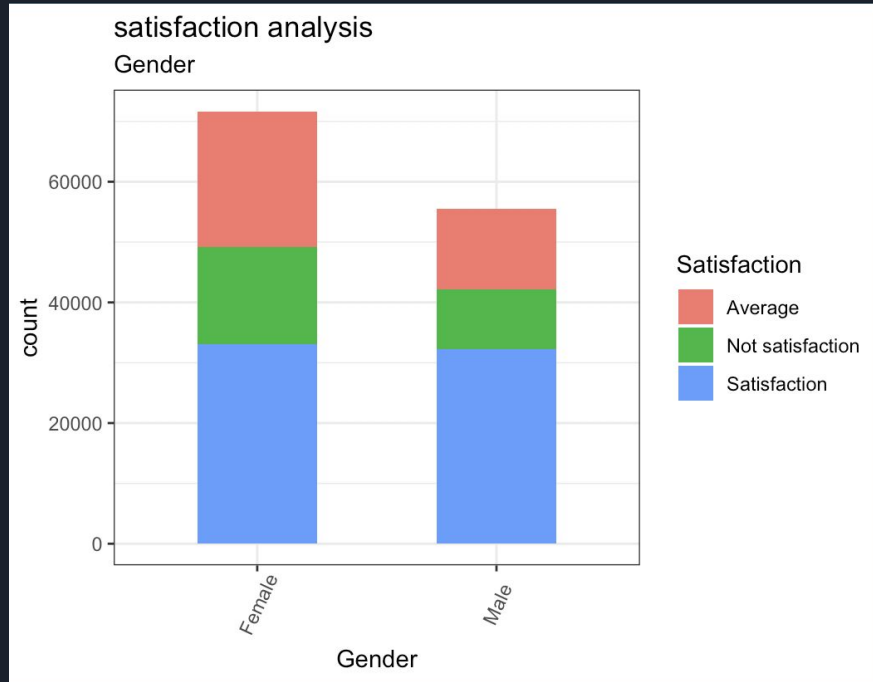
SATISFACTION VS AGE



Count of Satisfaction		Column Labels							
Row Labels		15-24	25-34	35-44	45-54	55-64	65-74	75-85	Grand Total
1		55	380	519	518	586	528	413	2999
2		2910	2075	2677	2955	3873	4551	4546	23587
2.5		1		1					2
3		6524	5333	5979	5311	5183	4766	3888	36984
3.5		1		1					2
4		4653	12166	14483	12169	7337	2090	860	53758
4.5			1	1					2
5		70	1850	4555	3845	1948	246	38	12552
Grand Total		14214	21805	28216	24798	18927	12181	9745	129886

Judging by the distribution of the histogram, the dissatisfaction and average ratings increase with increase in age, and the age group 60-80 are the most dissatisfied.

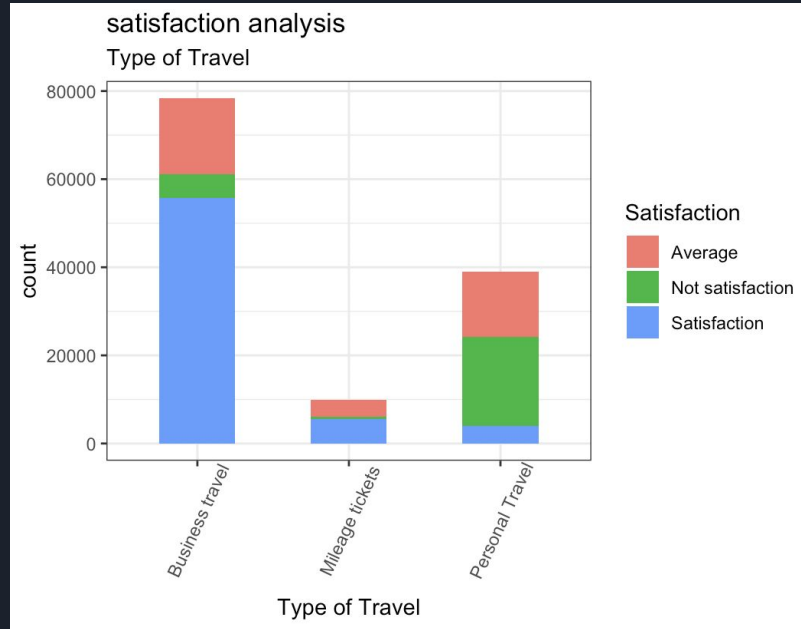
SATISFACTION VS GENDER





Count of Satisfaction		Column Labels		
Row Labels		Female	Male	Grand Total
1		1816	1183	2999
2		14646	8941	23587
2.5		2		2
3		23258	13726	36984
3.5			2	2
4		29429	24329	53758
4.5			2	2
5		4222	8330	12552
Grand Total		73373	56513	129886

There is a higher satisfaction level among the male customers as compared to female customers. There are more female customers who are either not satisfied or moderately satisfied.

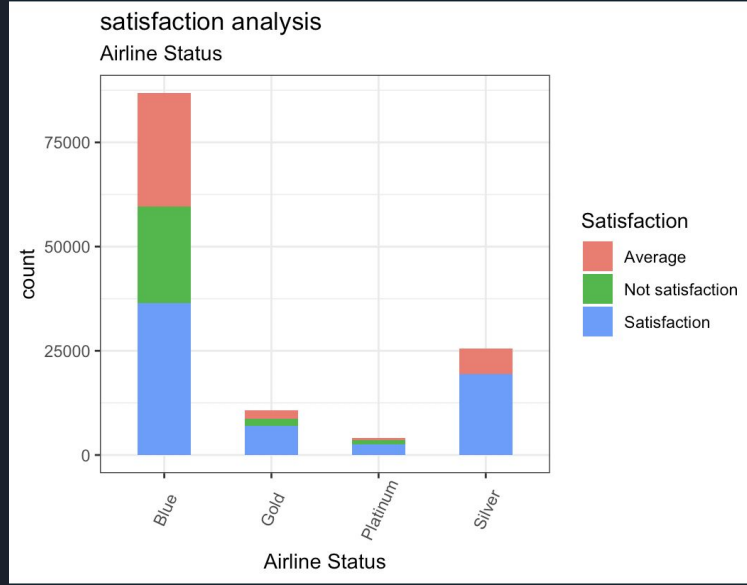
SATISFACTION VS TYPE OF TRAVEL



Count of Satisfaction		Column Labels 			
Row Labels		Business travel	Mileage tickets	Personal Travel	Grand Total
1		770	376	1853	2999
2		4746		18841	23587
2.5				2	2
3		17655	3908	15421	36984
3.5		2			2
4		44739	5351	3668	53758
4.5		2			2
5		11713	435	404	12552
Grand Total		79627	10070	40189	129886

- Business travel is a more satisfying experience than personal travel because of reimbursements and better services
- Personal travel involves customer's own money, which is why they might be left unsatisfied

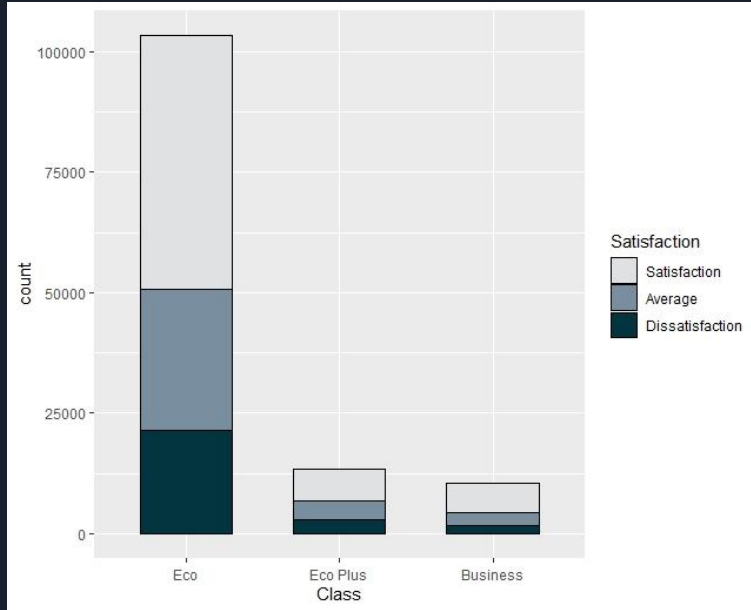
SATISFACTION VS AIRLINE STATUS



Count of Satisfaction		Column Labels			
Row Labels		Blue	Gold	Platinum	Silver
1		2597		402	2999
2		21012	1863	712	23587
2.5		2			2
3		28232	1978	461	6313
3.5		1	1		2
4		33863	4067	991	14837
4.5		2			2
5		3200	2928	1606	4818
Grand Total		88909	10837	4172	25968
					129886

- Silver is the most satisfaction-guaranteed airline status, followed by gold
- Platinum status has the least number of customers, whereas blue status have more than half the customers

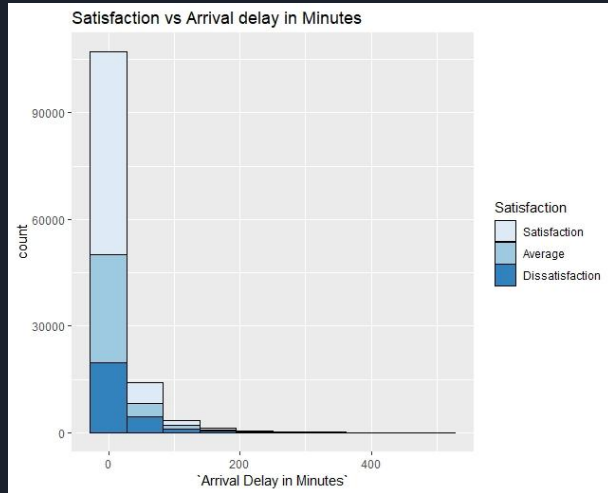
SATISFACTION VS CLASS



Count of Satisfaction		Column Labels		
Row Labels	Business	Eco	Eco Plus	Grand Total
1	209	2497	293	2999
2	1574	19460	2553	23587
2.5	1	1		2
3	2529	30242	4213	36984
3.5		2		2
4	4855	43263	5640	53758
4.5	1	1		2
5	1379	10266	907	12552
Grand Total	10548	105732	13606	129886

Economy class has the highest proportion of dissatisfied and moderately satisfied customers, followed by economy plus.

SATISFACTION VS ARRIVAL DELAY IN MINUTES



Count of Satisfaction		Column Labels								
Row Labels	(blank)	0-199 min	200-399 min	400-599 min	600-799 min	800-999 min	1000-1199 min	1200-1399 min	1400-1599 min	Grand Total
1		64	2912	19	4					2999
2		527	22805	227	19	6	1	1	1	23587
2.5			2							2
3		1142	35550	260	24	4	4			36984
3.5		1	1							2
4		936	52491	292	31	4	3	1		53758
4.5			2							2
5		68	12384	89	10			1		12552
Grand Total		2738	126147	887	88	14	8	2	1	129886

Higher arrival delay leads to a lesser customer satisfaction

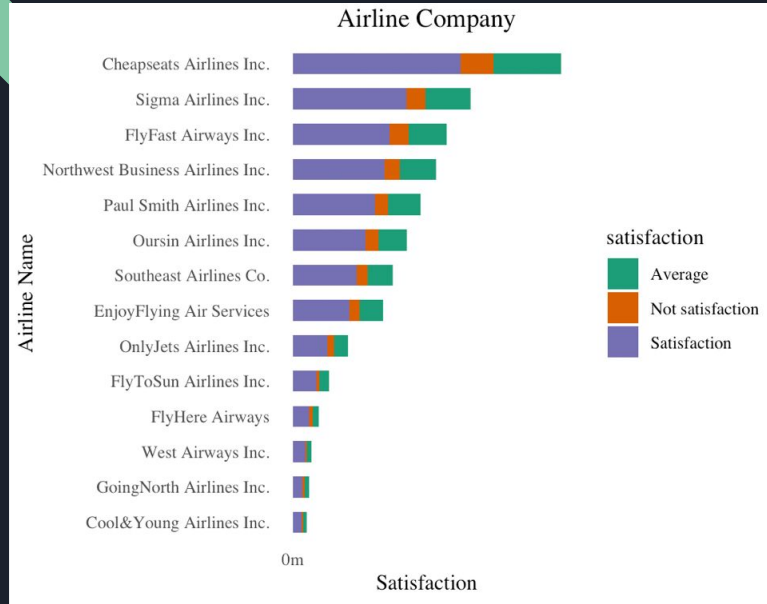
SATISFACTION VS SHOPPING AMOUNT



Count of Satisfaction		Column Labels									
Row Labels		0-99	100-199	200-299	300-399	400-499	500-599	600-699	700-799	800-899	Grand Total
1		2754	206	28	10	1					2999
2		21721	1450	304	79	22	10	1			23587
2.5		2									2
3		33859	2399	523	143	46	9	3		2	36984
3.5		2									2
4		48996	3629	822	242	49	15	2	3		53758
4.5		2									2
5		11453	858	176	52	11	1	1			12552
Grand Total		118789	8542	1853	526	129	35	7	3	2	129886

- With increase in shopping amount, there is an increase in the proportion of satisfied customers
- Most of the customers do not shop at airports

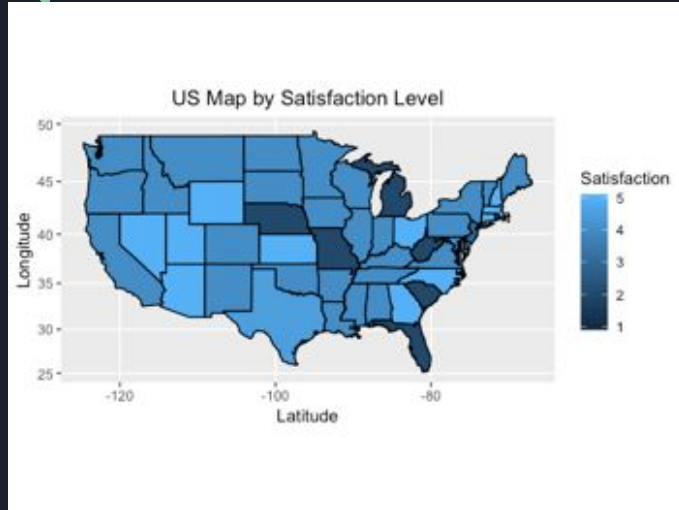
SATISFACTION VS AIRLINE COMPANY



Count of Satisfaction Level	Column Labels			
Row Labels	Average	Dissatisfied	Satisfied	Grand Total
Cheapseats Airlines Inc.	28%	22%	50%	100%
Cool&Young Airlines Inc.	28%	17%	55%	100%
EnjoyFlying Air Services	29%	21%	50%	100%
FlyFast Airways Inc.	28%	22%	50%	100%
FlyHere Airways	27%	21%	52%	100%
FlyToSun Airlines Inc.	29%	18%	53%	100%
GoingNorth Airlines Inc.	27%	25%	48%	100%
Northwest Business Airlines Inc.	29%	20%	52%	100%
OnlyJets Airlines Inc.	29%	22%	49%	100%
Oursin Airlines Inc.	28%	20%	52%	100%
Paul Smith Airlines Inc.	29%	19%	52%	100%
Sigma Airlines Inc.	29%	20%	52%	100%
Southeast Airlines Co.	29%	19%	52%	100%
West Airways Inc.	27%	16%	57%	100%
Grand Total	28%	20%	51%	100%

- West Airways Inc. and Cool&Young Airlines Inc. have lesser dissatisfaction than Southeast Airlines Co.
- Overall, Southeast Airlines performs better than most of its competitors

Geographical service benchmarking



Comparing satisfaction levels with all flights considered and Southeast Airlines Co., it can be identified that service in the following states can be improved by the client:
Nevada, Utah, Arizona, New Hampshire, Virginia and Ohio



DATA ANALYSIS



Transformation for Linear Regression



Initial Correlation Matrix



Final Correlation Matrix

- The correlation matrix was plotted, and multicollinearity was removed
- The outliers were removed from the data
- All the numerical variables were Z-transformed to remove scaling issues

Multiple Linear Regression Analysis

Key Interpretations

Variable	Change in Variable	Effect on Customer Satisfaction Index	Coefficients:				
				Estimate	Std. Error	t value	Pr(> t)
Age	Increase	Decrease	(Intercept)	0.163645	0.009116	17.952	< 2e-16 ***
			Z\$Age	-0.040994	0.002336	-17.549	< 2e-16 ***
			Z\$'Price Sensitivity'	-0.021992	0.002322	-9.471	< 2e-16 ***
Price Sensitivity	Increase	Decrease	Z\$'Shopping Amount at Airport'	0.012298	0.002560	4.803	1.56e-06 ***
			Z\$'Arrival Delay in Minutes'	-0.152884	0.003064	-49.901	< 2e-16 ***
			Z\$'No of Flights p.a.'	-0.049351	0.002387	-20.674	< 2e-16 ***
			Z\$'Airline Status'Gold	0.454715	0.007834	58.042	< 2e-16 ***
Shopping amount at Airport	Increase	Increase	Z\$'Airline Status'Platinum	0.273630	0.012197	22.434	< 2e-16 ***
			Z\$'Airline Status'Silver	0.642075	0.005454	117.716	< 2e-16 ***
			Z\$GenderMale	0.135663	0.004401	30.826	< 2e-16 ***
			Z\$'Type of Travel'Mileage tickets	-0.150765	0.008169	-18.457	< 2e-16 ***
			Z\$'Type of Travel'Personal Travel	-1.115242	0.005218	-213.730	< 2e-16 ***
Number of flights per annum	Increase	Decrease	Z\$ClassEco	-0.079178	0.007751	-10.215	< 2e-16 ***
			Z\$ClassEco Plus	-0.072470	0.009945	-7.287	3.18e-13 ***
			Z\$'Airline Name'Cool&Young Airlines Inc.	0.063831	0.021617	2.953	0.00315 **
			Z\$'Airline Name'EnjoyFlying Air Services	0.011758	0.009408	1.250	0.21142
Airline Status	Change in status	Silver > Gold > Platinum > Blue	Z\$'Airline Name'FlyFast Airways Inc.	0.019886	0.007812	2.546	0.01091 *
			Z\$'Airline Name'FlyHere Airways	0.010605	0.016041	0.661	0.50854
			Z\$'Airline Name'FlyToSun Airlines Inc.	0.027316	0.013840	1.974	0.04842 *
			Z\$'Airline Name'GoingNorth Airlines Inc.	-0.052528	0.019668	-2.671	0.00757 **
			Z\$'Airline Name'Northwest Business Airlines Inc.	0.026189	0.008018	3.266	0.00109 **
Gender	Change in gender	Male > Female	Z\$'Airline Name'OnlyJets Airlines Inc.	0.007530	0.011424	0.659	0.50983
			Z\$'Airline Name'Oursin Airlines Inc.	0.023893	0.008656	2.760	0.00578 **
			Z\$'Airline Name'Paul Smith Airlines Inc.	0.024742	0.008338	2.967	0.00300 **
Type of travel	Change in type of travel	Business > Mileage > Personal	Z\$'Airline Name'Sigma Airlines Inc.	0.026212	0.007495	3.497	0.00047 ***
			Z\$'Airline Name'Southeast Airlines Co.	0.027690	0.009099	3.043	0.00234 **
			Z\$'Airline Name'West Airways Inc.	0.082946	0.018998	4.366	1.27e-05 ***
			Z\$Number_of_years	0.014636	0.002122	6.898	5.30e-12 ***
Class	Change in class	Business > Eco Plus > Eco	--- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
			Residual standard error: 0.7546 on 127120 degrees of freedom				
			Multiple R-squared: 0.4306, Adjusted R-squared: 0.4305				
Number of years	Increase	Increase	F-statistic: 3561 on 27 and 127120 DF, p-value: < 2.2e-16				



MULTIPLE LINEAR REGRESSION ANALYSIS

RECOMMENDATIONS

- Satisfaction level for the client is high among their passengers, so they should now focus on the segment of passengers who are price sensitive, and implement flexibility of pricing
- There must be more capital invested into improving the service in economy class and blue status carriers, which is targeted towards the personal travel segment
- There must be more marketing done for the silver and gold status carriers, as well as the business class seats; the target segment for this marketing must be mileage travel and business travel passengers



ASSOCIATION RULES

Category	Common traits	Rating
High	Silver status business travel Male middle aged no delay low price sensitivity	Greater than 4
Average	Personal travel Female economy class no delay high number of flights per annum	Between 3 and 4
Low	Blue status personal travel delayed	Below 3



SUPPORT VECTOR MACHINE

- To validate the precision of the SVM model, we calculated the error rate of this model. The error rate is 22.56%
- The SVM model accurately predicts whether a customer is satisfied or not 78% of the time

Classification	Satisfaction(predict)	Dissatisfaction(predict)
Dissatisfaction	1547 (Wrongly classified)	2767
Satisfaction	3859	384 (Wrongly classified)



RECOMMENDATIONS



RECOMMANDATIONS – DESCRIPTIVE STATISTICS

- There is scope for improvement for client in terms of overall service in the following states: Nevada, Utah, Arizona, New Hampshire, Virginia and Ohio
- There can be more capital expenditure directed to improve operations in these areas



RECOMMANDATIONS – LINEAR MODEL

- Satisfaction level for the client is high among their passengers, so they should now focus on the segment of passengers who are price sensitive, and implement flexibility of pricing
- There must be more capital invested into improving the service in economy class and blue status carriers, which is targeted towards the personal travel segment
- There must be more marketing done for the silver and gold status carriers, as well as the business class seats; the target segment for this marketing must be mileage travel and business travel passengers



RECOMMENDATIONS – ASSOCIATION RULES

High satisfaction:

- The client could focus on marketing the silver status package more, to improve sales. Even if they evolve the pricing strategy into a dynamic pricing, the cluster of passengers who are on business travel, or are not price sensitive, would be attracted due to the services that are available to them
- The airline routes which have lesser traffic should nest a silver or gold status airline, as the target customer demographics would comprise of older people, who highly value arrival-on-time and good hospitality
- The client should focus on creating partnership with any Multinational Companies, and developing bundle-packages for them in the business class segment. The target segment comprises of many middle-aged business class travelers, who would be highly satisfied



RECOMMANDATIONS – ASSOCIATION RULES

Moderate satisfaction:

- Low-priced business class plans must be created for this segment. This would result in a higher satisfaction level
- The client must focus on improving service in the blue status airlines, which should improve the satisfaction level in the passengers in this segment
- Females traveling through economy give a strong priority on the overall value associated with the airline's service. The client should consider creating value-oriented bundle plans, or lowering the ticket prices, specifically for women



RECOMMANDATIONS – ASSOCIATION RULES

Low satisfaction:

- The client must focus upon improving services for the blue airline status, as well as the economy class seats
- Lower price setting for business class seats in blue airline status might trigger a higher satisfaction level



THANKS!

Any
questions?