IST 687 FINAL PRESENTATION

Shwet jain

Hope Puzzanghera

Osama Junaid

Pat Carlin

OVERVIEW

Objective: Analyze and manipulate the dataset of customers flying with Southeast Airlines

Goal: Provide actionable insights into who, and what, was affecting their overall customer satisfaction. Before coming up with specific business questions, our first step was to clean, manipulate, and transform our data into a workable set.

Basic Understanding of the data Set: Approximately 130,000 survey responses, with 25 fields, and some entries in the data-set are blank (NA)

Content

1. Business Questions
2. Descriptive Statistics
3. Visualization
4. Modeling
5. Key Insights

Data Cleansing Involved in the Data-Set

1. Dealing with Column Names (Incorrect, Incomplete, Improperly formatted)
2. Dealing with ‘ . ’ s in the Column Names:
3. Dealing with NA’s: NA's will be replaced by means of their respective columns
4. Dealing with City,State in the Same Column:
5. Extracting Data (SouthEast Airlines)

Data-Set After Cleansing, Transformation & Munging

Class	DayOfMonth	FlightDate	AirlineCode	AirlineName	OriginCity	OriginState	DestinationCity	DestinationState
Business	13	2/13/14	US	Southeast Airlines Co.	Milwaukee	Wisconsin	Phoenix	Arizona
Business	25	3/25/14	US	Southeast Airlines Co.	Milwaukee	Wisconsin	Phoenix	Arizona
Business	8	3/8/14	US	Southeast Airlines Co.	Milwaukee	Wisconsin	Phoenix	Arizona
Eco	13	2/13/14	US	Southeast Airlines Co.	Milwaukee	Wisconsin	Phoenix	Arizona
Eco	26	1/26/14	US	Southeast Airlines Co.	Milwaukee	Wisconsin	Phoenix	Arizona
Eco	8	3/8/14	US	Southeast Airlines Co.	Milwaukee	Wisconsin	Phoenix	Arizona
Eco	17	3/17/14	US	Southeast Airlines Co.	Milwaukee	Wisconsin	Phoenix	Arizona
Eco	2	2/2/14	US	Southeast Airlines Co.	Milwaukee	Wisconsin	Phoenix	Arizona
Eco	4	2/4/14	US	Southeast Airlines Co.	Milwaukee	Wisconsin	Phoenix	Arizona
Eco	14	2/14/14	US	Southeast Airlines Co.	Milwaukee	Wisconsin	Phoenix	Arizona
Eco	21	3/21/14	US	Southeast Airlines Co.	Milwaukee	Wisconsin	Phoenix	Arizona
Eco	19	1/19/14	US	Southeast Airlines Co.	Milwaukee	Wisconsin	Phoenix	Arizona
Eco	11	3/11/14	US	Southeast Airlines Co.	Milwaukee	Wisconsin	Phoenix	Arizona
Eco	27	2/27/14	US	Southeast Airlines Co.	Milwaukee	Wisconsin	Phoenix	Arizona
Eco	19	3/19/14	US	Southeast Airlines Co.	Milwaukee	Wisconsin	Phoenix	Arizona
Eco	30	1/30/14	US	Southeast Airlines Co.	Milwaukee	Wisconsin	Phoenix	Arizona
Eco	2	3/2/14	US	Southeast Airlines Co.	Milwaukee	Wisconsin	Phoenix	Arizona
Eco Plus	19	3/19/14	US	Southeast Airlines Co.	Milwaukee	Wisconsin	Phoenix	Arizona
Eco	16	2/16/14	US	Southeast Airlines Co.	Seattle	Washington	Philadelphia	Pennsylvania

Descriptive Statistics

- Provide simple summaries about the sample and the measures.
- Overview of what we want to cover and form business questions

Used the following inbuilt R functions:

- summary()
- range()
- sd()

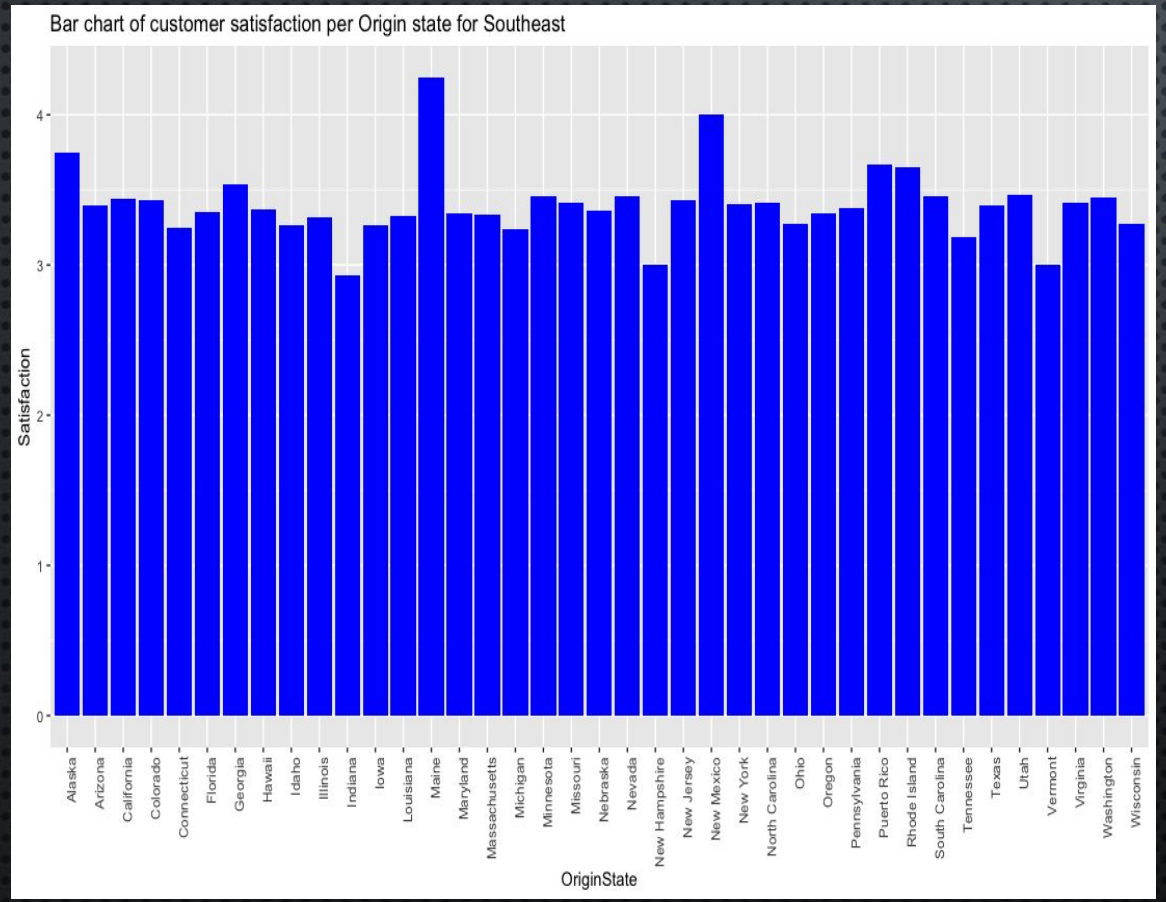
```
> summary(ProjectData)
 Satisfaction    AirlineStatus       Age          Gender     PriceSensitivity YearOfFirstFlight
4     :53758   Blue    :88910   Min.   :15.0   Female:73374   Min.   :0.000    Min.   :2003
3     :36984   Gold    :10837   1st Qu.:33.0   Male  :56515   1st Qu.:1.000    1st Qu.:2004
2     :23587   Platinum: 4172   Median :45.0                  Median :1.000    Median :2007
5     :12552   Silver  :25970   Mean   :46.2                  Mean   :1.276    Mean   :2007
1     : 2999                    3rd Qu.:59.0                  3rd Qu.:2.000    3rd Qu.:2010
2.5   :    2                    Max.   :85.0                  Max.   :5.000    Max.   :2012
(Other):    7
 FlightsPerYear   FlightsWithOtherAirlines        TypeofTravel   NoOfOtherLoyaltyCards ShoppingAtAirport
Min.   :  0.00   Min.   :  1.000        Business travel:79630   Min.   : 0.0000       Min.   :  0.00
1st Qu.:  9.00   1st Qu.:  4.000        Mileage tickets:10070   1st Qu.: 0.0000       1st Qu.:  0.00
Median : 17.00   Median :  7.000        Personal Travel:40189   Median : 0.0000       Median :  0.00
Mean   : 20.08   Mean   :  9.314                                Mean   : 0.8838       Mean   : 26.55
3rd Qu.: 29.00   3rd Qu.: 10.000                                3rd Qu.: 2.0000       3rd Qu.: 30.00
Max.   :100.00   Max.   :110.000                                Max.   :12.0000       Max.   :879.00

 EatingAndDrinkingAtAirport      Class        DayOfMonth      FlightDate      AirlineCode
Min.   :  0.00            Business: 10548   Min.   : 1.00   3/13/14:  1641   WN     :26058
1st Qu.: 30.00            Eco     :105735   1st Qu.: 8.00   3/10/14:  1640   DL     :17037
Median : 60.00            Eco Plus: 13606   Median :16.00   3/21/14:  1638   EV     :15407
Mean   : 68.24                              Mean   :15.72   3/26/14:  1628   OO     :13840
3rd Qu.: 90.00                              3rd Qu.:23.00   3/27/14:  1622   AA     :12248
Max.   :895.00                              Max.   :31.00   3/24/14:  1619   OU     :10968
                                                            (Other):120101   (Other):34331
                      AirlineName    OriginCity        OriginState   DestinationCity
Cheapseats Airlines Inc.       :26058   Length:129889   California:16751   Length:129889
Sigma Airlines Inc.            :17037   Class :character   Texas     :16346   Class :character
FlyFast Airways Inc.           :15407   Mode  :character   Florida   :10894   Mode  :character
Northwest Business Airlines Inc. :13840                    Georgia   : 8751
```

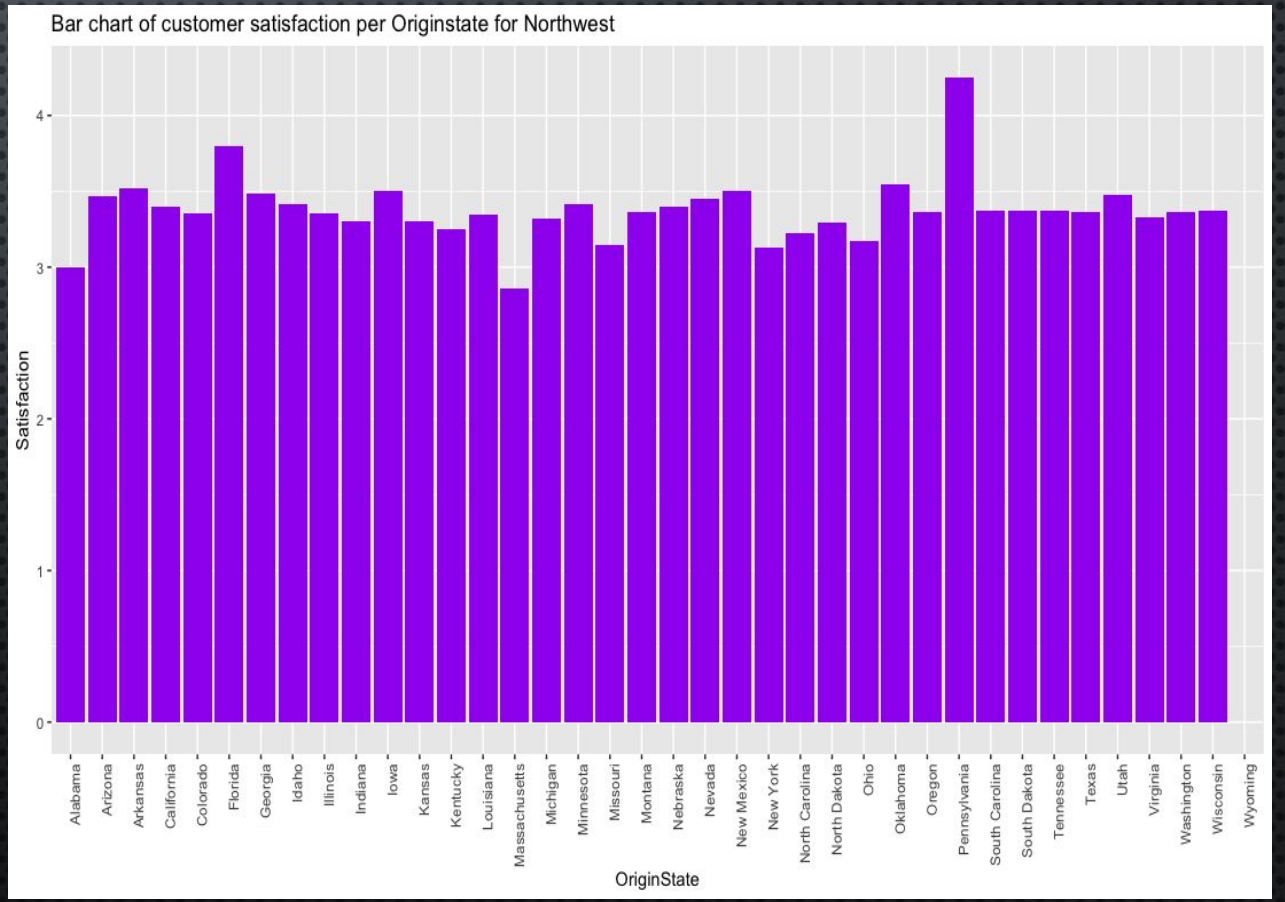
BUSINESS QUESTIONS

- Does Gender play an important role wrt. Satisfaction?
- What are the important attributes that drive Satisfaction ?
- Discuss the correlation between Distance Covered and Arrival Delay?
- Does the relationship between the Origin City and the Destination City together have something to do with Customer Satisfaction.
- Relationship between the Airline Status and Satisfaction.
- What is the relationship between Age, Price Sensitivity, and Satisfaction between types of travel?

VISUALIZATION

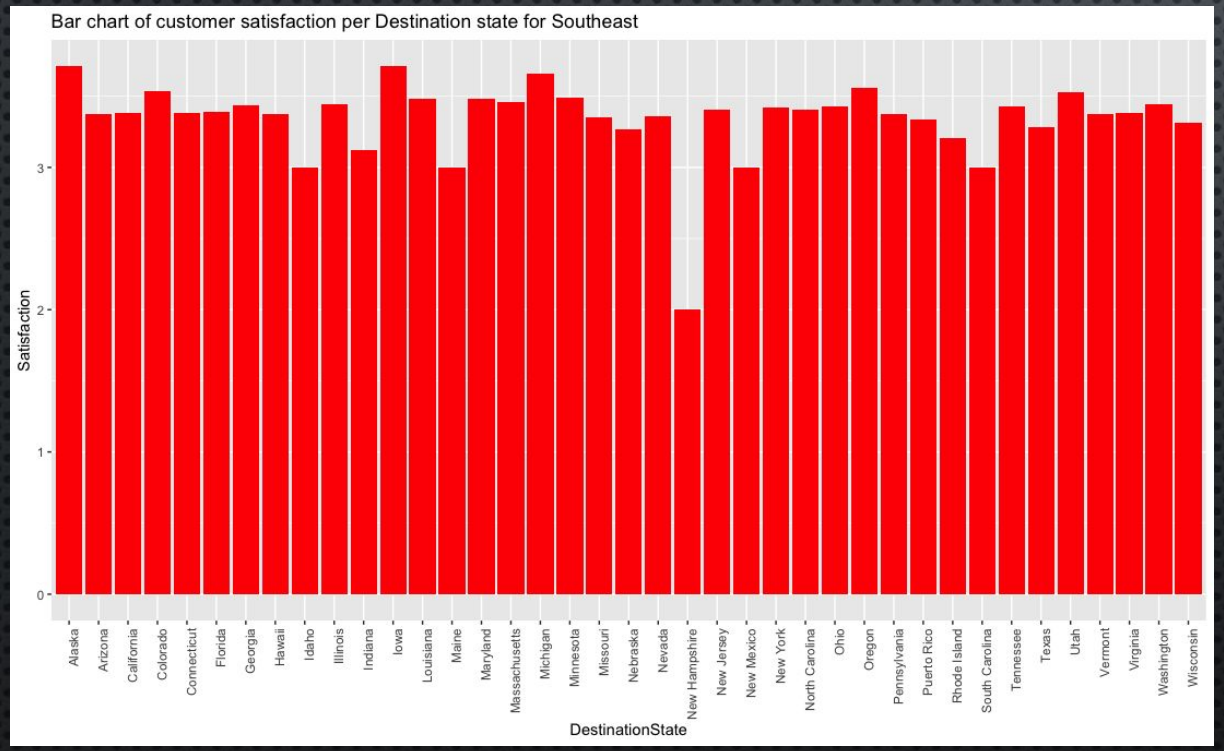
We compared the Origin State of SouthEast Airlines with NorthWest Airlines in order to find out the weaknesses of the competitor in a particular origin state and then in some way that could give a competitive edge over Northwest

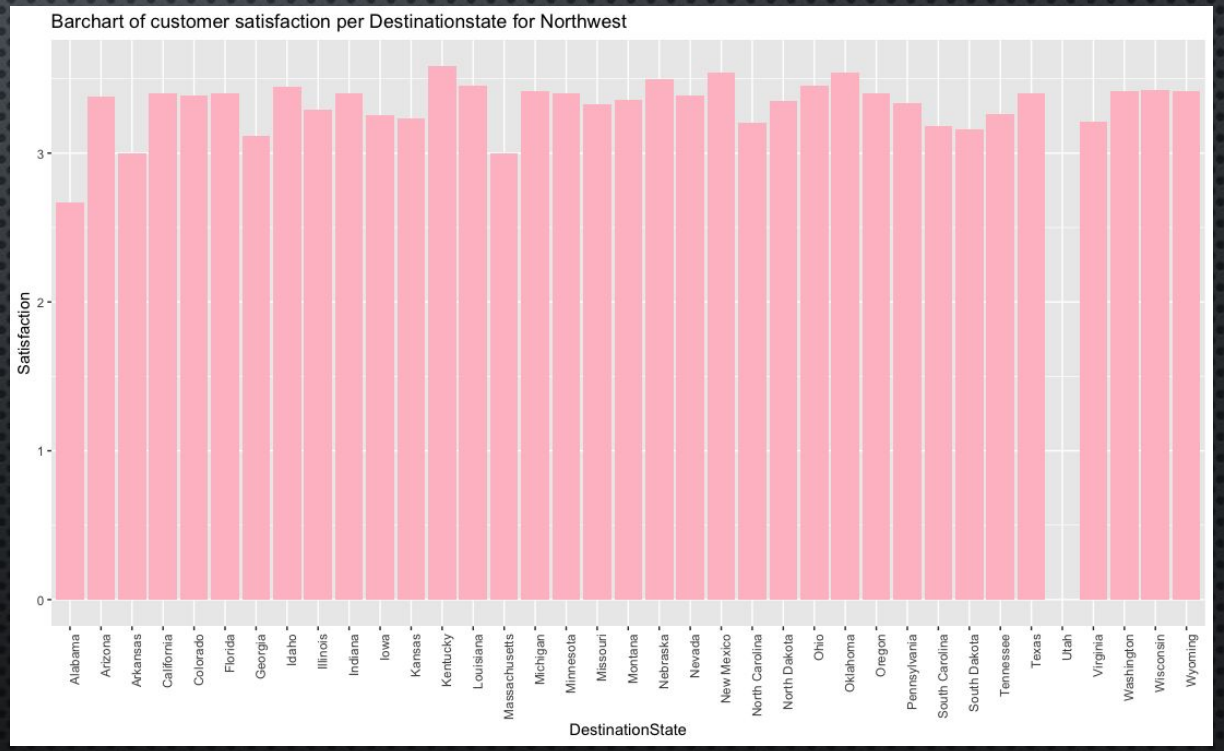




Visualization

We compared the Destination State of SouthEast Airlines with NorthWest Airlines in order to find out the weaknesses of the competitor in a particular origin state and then in some way that could give a competitive edge over Northwest





Key Insights

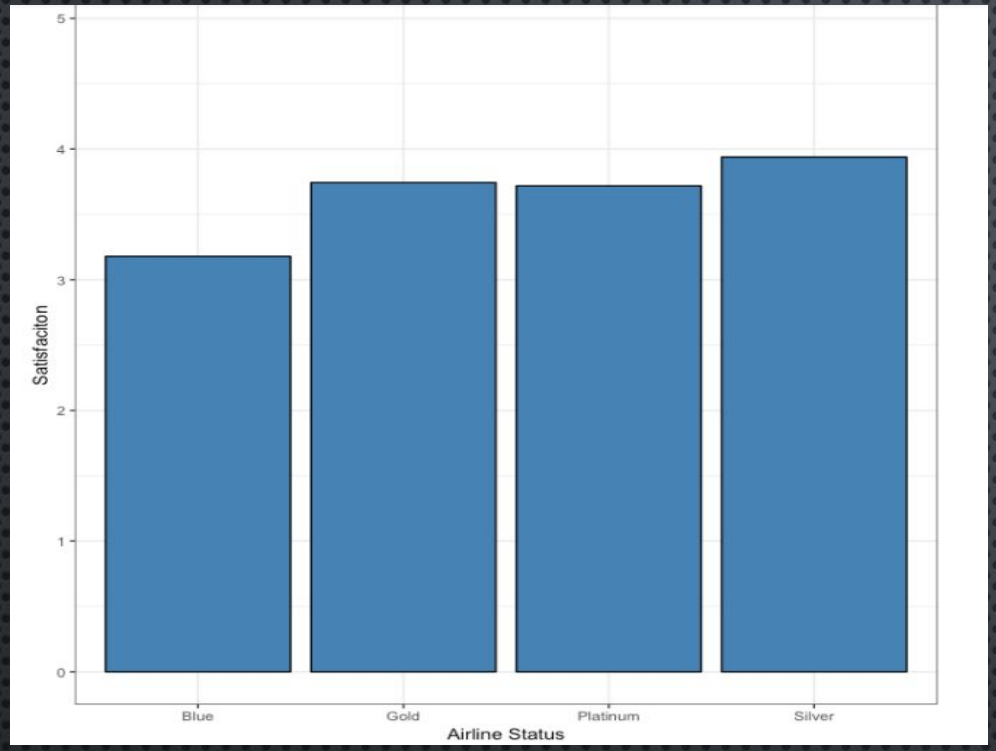
Southeast Airlines only flies out of and to 39 States.

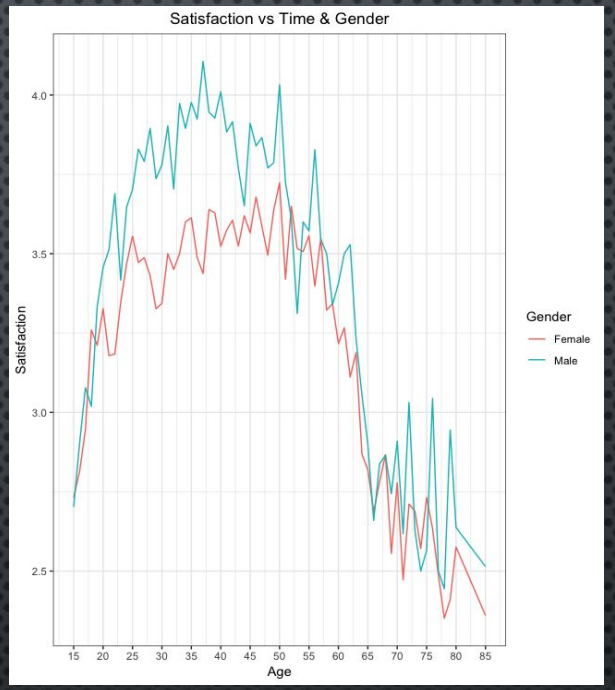
Maine has the highest Satisfaction for Origin State but the lowest for Destination State. Customers were satisfied flying out of Maine but were not Satisfied flying to Maine.

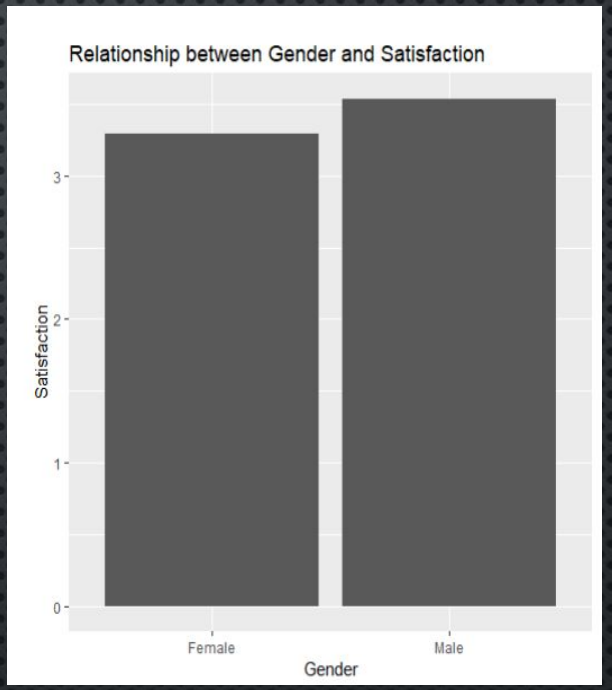
Compared to Southeast Airlines, Northwest Airlines only flies out of and flies to 38 States. Pennsylvania has the highest satisfaction for Origin State for Northwest with a satisfaction rate of 4.25 whereas compared to Southeast Airlines where Pennsylvania has a satisfaction rate of 3.3 . We can compare business models and see why Northwest did better compared to Southeast for Pennsylvania and adapt it.

Southeast Airlines doesn't fly to or out of Alabama but Northwest Airlines does. The satisfaction rate of Origin State and Destination State for Alabama is the lowest overall. We believe that if Southeast Airlines starts to fly to and out of Alabama, they can increase profits because people aren't happy with Northwest Airlines

Overall, we advice Southeast Airlines to fly out of and to more states if they want to increase profits.

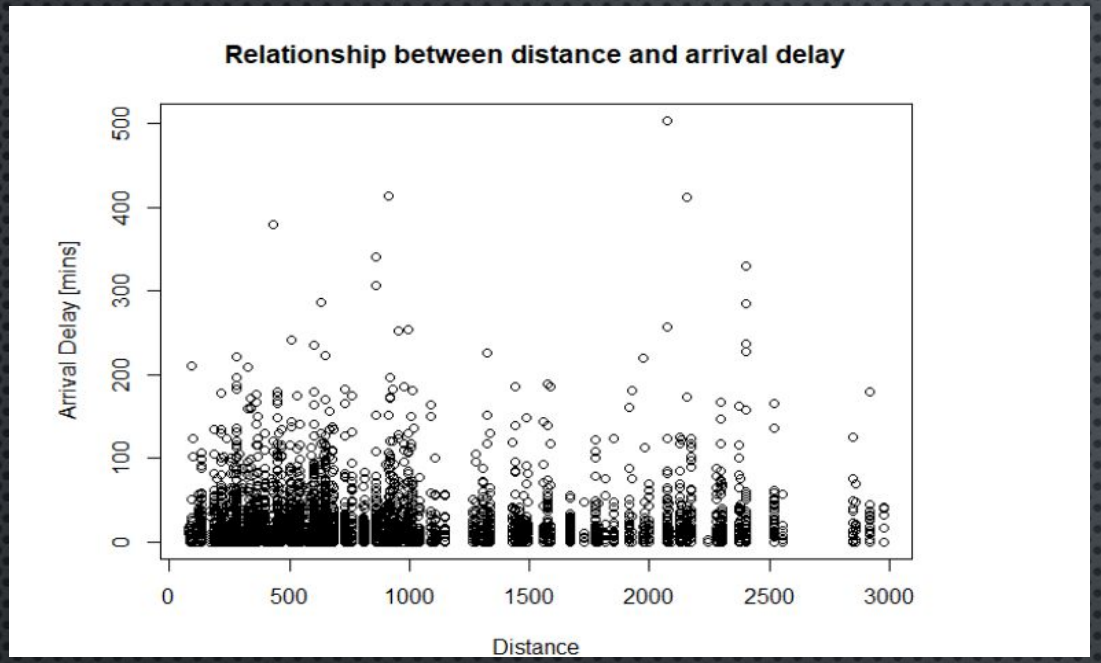




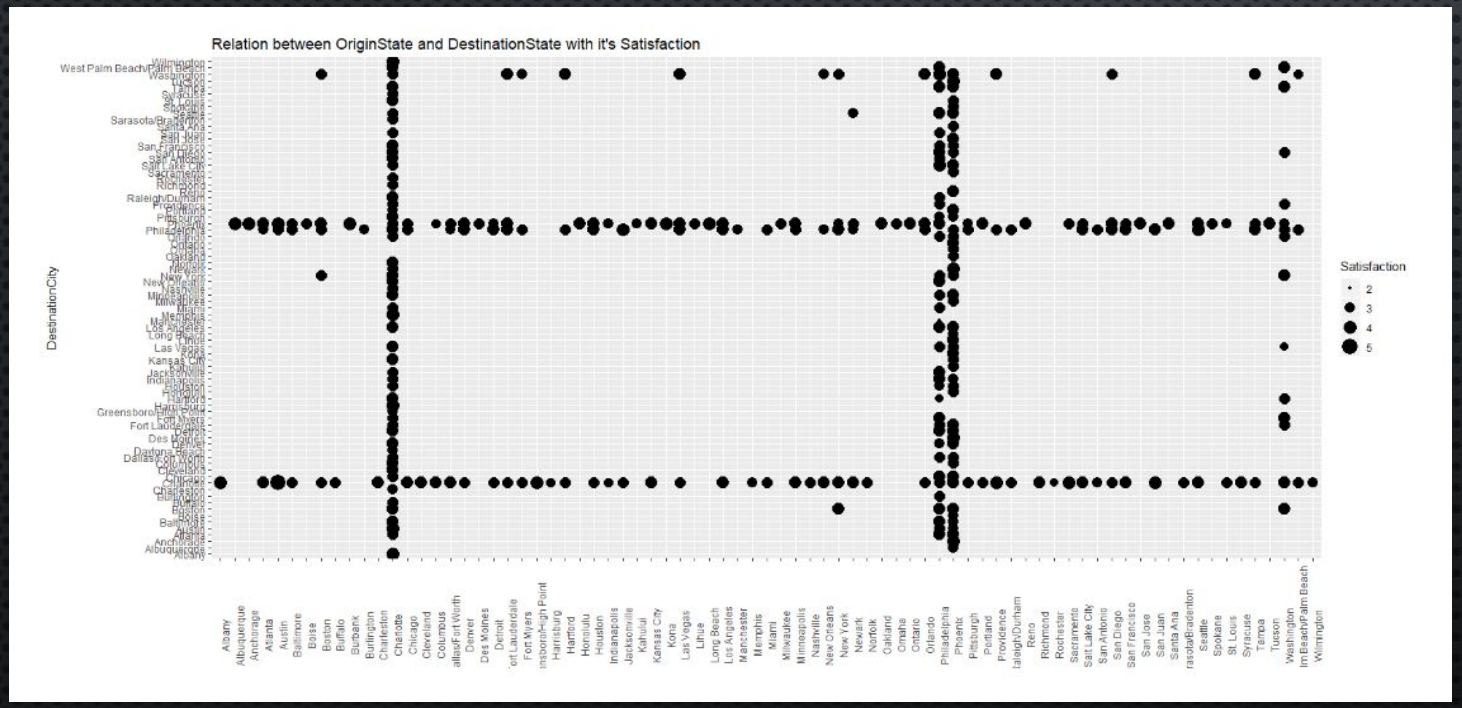


As we can see from the plot, Airline_Status=Blue has a lower customer Satisfaction when compared to other Airline_Status.

As we can see, Customers in a higher age range are unsatisfied.
Male(Satisfaction) is higher as compared to Female(Satisfaction)



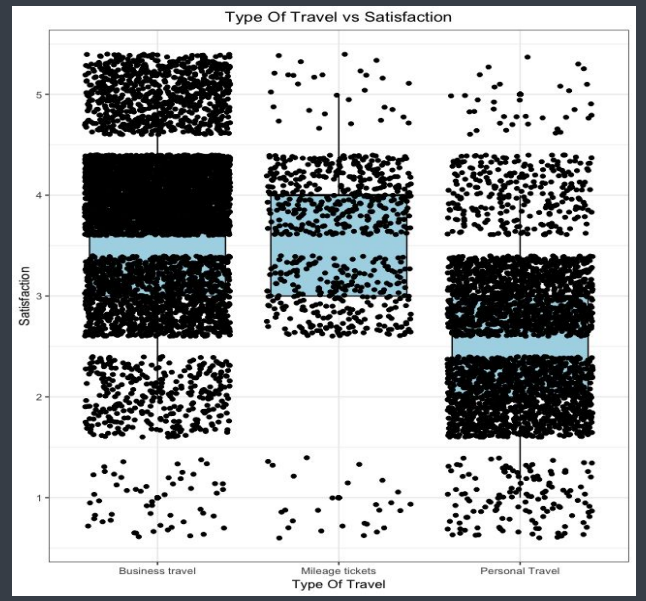
We carried out linear modelling tests for the same. And there was no significant relationship between the two variables.



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

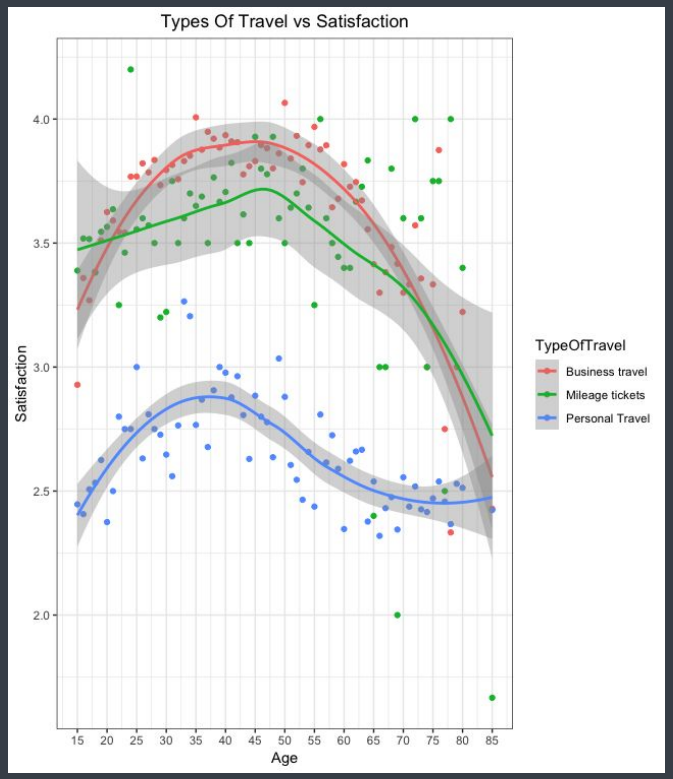
Visualizations - Type Of Travel (Actionable Insight)

Exploration



- Exploring the average satisfaction within the separate types of travel
- Median of each type would be the clear blue line

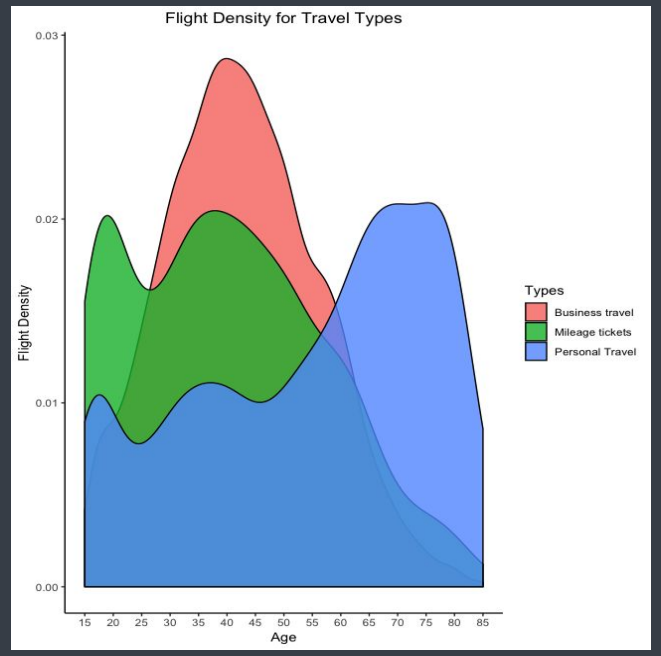
Refining our Scope



- Viewing the averaged satisfaction over age with respect to each type of travel

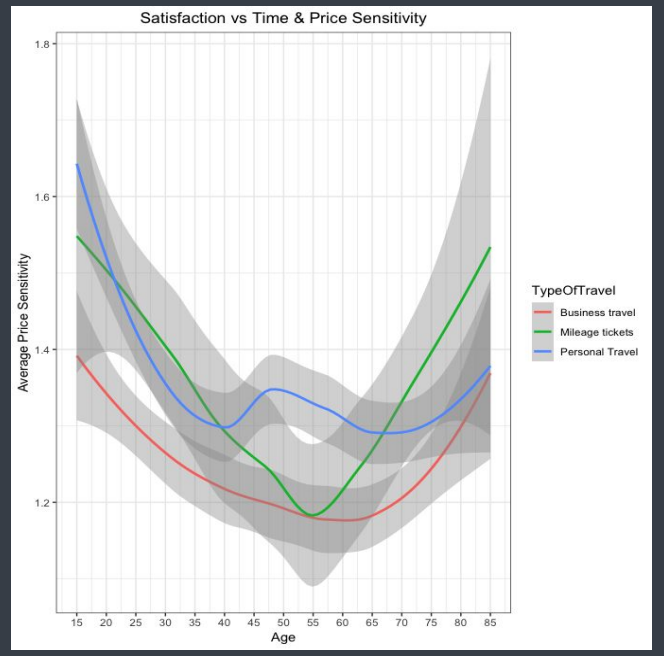
Visualization - Type Of Travel (Actionable Insight)

Diving Deeper



- This graph shows us the flight frequency density across the range of years
- Notice the peak in Personal travel from ages 55 - 65

Validation



- Depicting the Price sensitivity smoothed over age shows that sensitivity from age 50 - 85 essentially remains unchanged
- Notice how sensitive price was from ages 15 - 40 for personal travel

Different Modelling Techniques

1) APRIORI (Association Rules)
2) LINEAR MODEL
3) SUPPORT VECTOR MACHINE

APRIORI ALGORITHM/MODEL

```
121 ruleset <- apriori(SouthData, list(support = 0.35,confidence = 0.40))
122 inspect(ruleset)
123 ruleSub<- subset(ruleset, subset = rhs %in% "Satisfaction=High")
124 inspect(ruleSub)
125
```

128:1 (Top Level)

Console Terminal

~/

```
    lhs                                                rhs                        support confidence      lift count
[1]  {}                                             => {Satisfaction=High} 0.5179075  0.5179075 1.0000000  4960
[2]  {TypeofTravel=Business travel}                 => {Satisfaction=High} 0.4405346  0.7195975 1.3894325  4219
[3]  {DepartureDelayInMinutes=Low}                  => {Satisfaction=High} 0.3586718  0.5448057 1.0519363  3435
[4]  {ArrivalDelayGreaterThan5mins=no}              => {Satisfaction=High} 0.3950089  0.5594499 1.0802120  3783
[5]  {PriceSensitivity=Low}                         => {Satisfaction=High} 0.3848804  0.5423779 1.0472486  3686
[6]  {Class=Eco}                                    => {Satisfaction=High} 0.4190247  0.5155447 0.9954378  4013
[7]  {FlightsCancelled=No}                          => {Satisfaction=High} 0.5146706  0.5218634 1.0076383  4929
[8]  {TypeofTravel=Business travel,
      Class=Eco}                                    => {Satisfaction=High} 0.3561658  0.7188620 1.3880123  3411
[9]  {TypeofTravel=Business travel,
      FlightsCancelled=No}                          => {Satisfaction=High} 0.4381330  0.7222031 1.3944635  4196
[10] {DepartureDelayInMinutes=Low,
      FlightsCancelled=No}                          => {Satisfaction=High} 0.3585674  0.5448199 1.0519638  3434
[11] {FlightsCancelled=No,
      ArrivalDelayGreaterThan5mins=no}              => {Satisfaction=High} 0.3917720  0.5659125 1.0926904  3752
[12] {PriceSensitivity=Low,
      FlightsCancelled=No}                          => {Satisfaction=High} 0.3823744  0.5464856 1.0551800  3662
[13] {Class=Eco,
      FlightsCancelled=No}                          => {Satisfaction=High} 0.4161011  0.5198956 1.0038388  3985
[14] {TypeofTravel=Business travel,
      Class=Eco,
      FlightsCancelled=No}                          => {Satisfaction=High} 0.3539731  0.7215837 1.3932675  3390
```

Association Rules - Key Insights

1) **Airline Status** = BLUE → **Satisfaction** = LOW

2) **Gender** = Female → **Satisfaction** = LOW

3) **Type Of Travel** = Business Travel → **Satisfaction** = HIGH

4) **Price Sensitivity** = LOW → **Satisfaction** = HIGH

5) **Departure Delay in Minutes** = LOW → **Satisfaction** = HIGH

6) **Flights Cancelled** = NO → **Satisfaction** = HIGH

LINEAR MODEL

```
Residuals:
    Min      1Q  Median      3Q     Max
-3.1043 -0.4202  0.1168  0.4477  2.7215

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                   -7.4625068  4.9288831  -1.514 0.130050
Age                           -0.0028648  0.0004758  -6.021 1.80e-09 ***
genderMale                     0.1305822  0.0151205   8.636  < 2e-16 ***
PriceSensitivity              -0.0154257  0.0136960  -1.126 0.260070
YearOfFirstFlight              0.0055999  0.0024560   2.280 0.022625 *
FlightsPerYear                -0.0032942  0.0005585  -5.898 3.80e-09 ***
EatingAndDrinkingAtAirport    -0.0003979  0.0001428  -2.787 0.005338 **
FlightsWithOtherAirlines       0.0007031  0.0008990   0.782 0.434200
ScheduledDepartureHour         0.0051191  0.0014979   3.418 0.000634 ***
statusSilver                   0.6030942  0.0189369  31.847  < 2e-16 ***
statusPlatinum                 0.2960908  0.0420220   7.046 1.97e-12 ***
statusGold                     0.4247251  0.0271269  15.657  < 2e-16 ***
mileageTickets                -0.1317763  0.0286264  -4.603 4.21e-06 ***
personalTravel                -1.0249720  0.0180831 -56.681  < 2e-16 ***
businessClass                  0.0804198  0.0264634   3.039 0.002381 **
flightsCancelled              -0.2467738  0.0630666  -3.913 9.18e-05 ***
arrivalDelayGreaterThan5mins -0.3486045  0.0161264 -21.617  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7157 on 9560 degrees of freedom
Multiple R-squared:  0.436,     Adjusted R-squared:  0.4351
F-statistic: 461.9 on 16 and 9560 DF,  p-value: < 2.2e-16
```

Positive Impact on Satisfaction:

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

Negative Impact on Satisfaction:

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

SUPPORT VECTOR MACHINE

```
Support Vector Machine object of class "ksvm"

SV type: C-svc  (classification)
 parameter : cost C = 5

Gaussian Radial Basis kernel function.
 Hyperparameter : sigma =  0.0467305677792971

Number of Support Vectors : 3108

Objective Function Value : -12445.79
Training error : 0.178415
Cross validation error : 0.213346
Probability model included.
```

```
> results
                      svmPred.2...
testData.Satisfaction    0    1
        Not Satisfied 1063  513
        Satisfied      137 1480
>
> #calculate an error rate based on the confusion matrix
> ((results[1,2]+results[2,1])/length(testData$Satisfaction))
[1] 0.2035703
```

Satisfied Customers

- Satisfaction was reported as greater than 3

Not Satisfied Customers

- Satisfaction was reported as 3 or less

SVM Model Correctly Identified

- 1063 NotSatisfied Customers
- 1480 Satisfied Customers

~ 80% of the responses were predicted correctly from the svm model

Actionable Insights/ Overall Interpretation of the Result

FINDING	RECOMMENDATION
As age increases, Personel travel suprasses both Business and Mileage ticket travel	Possible stipend for seniors, as their frequency of travel increase while price sensitivity remains unchanged, essentially allowing seniors to travel more aiding in higher satisfaction
Price Sensitivity for Business travel and Mileage has the least variation throughout age and frequency	Could increase the price of business class as it will vary the most
Mileage tickets and Personal travel are correlated from 15 - 25, while satisfaction is rather low	Promote a higher mileage point award system for business travel around age 45, where the highest frequency is, and where we believe people are using mileage tickets for both personal and mileage type of travel outside of work. This will raise the overall satisfaction as price sensitivity is correlated to satisfaction
Customers who travel by BLUE STATUS flights are unsatisfied	Improve amenities for blue status flights and work
Customers who travel for PERSONAL reasons are unsatisfied	Provide family plans to reduce flight cost
Customers in a higher age range are unsatisfied	Better healthcare facilities and proper care of their needs which will drive their Satisfaction. Discount Offers for Senior citizens.
Females are less satisfied compared to Males	

THANK YOU FOR YOUR ATTENTION !

https://trello.com/b/KhK7E37T/ist-687-project-group-1