



# Airline Satisfaction Investigation

Team: Graff, Ruiz, Foltz, Sheahan

# Airline Data Conceptual Map

129,000 rows of data regarding individual flights...

## Customer

- Satisfaction
- Age
- Gender
- Price Sensitivity
- Year of First Flight
- Number of Flights
- Type of Travel

## Loyalty

- Airline Status
- % Flights w/Other Airlines
- # of Other Loyalty Cards

## Airline

- Airline Code
- Airline Name

Origin City

Origin State

Destination City

Destination State

Class

Flight Date

Spend at Airport

Scheduled Departure

Flight Time

Flight Distance

## Flight

Flight Cancelled

Arrival Delay in Mins

Arrival Delay > 5 Mins

Departure Delay in Mins

Departure Delay > 5 Mins

## Flight Performance

## **Satisfaction Questions**

---

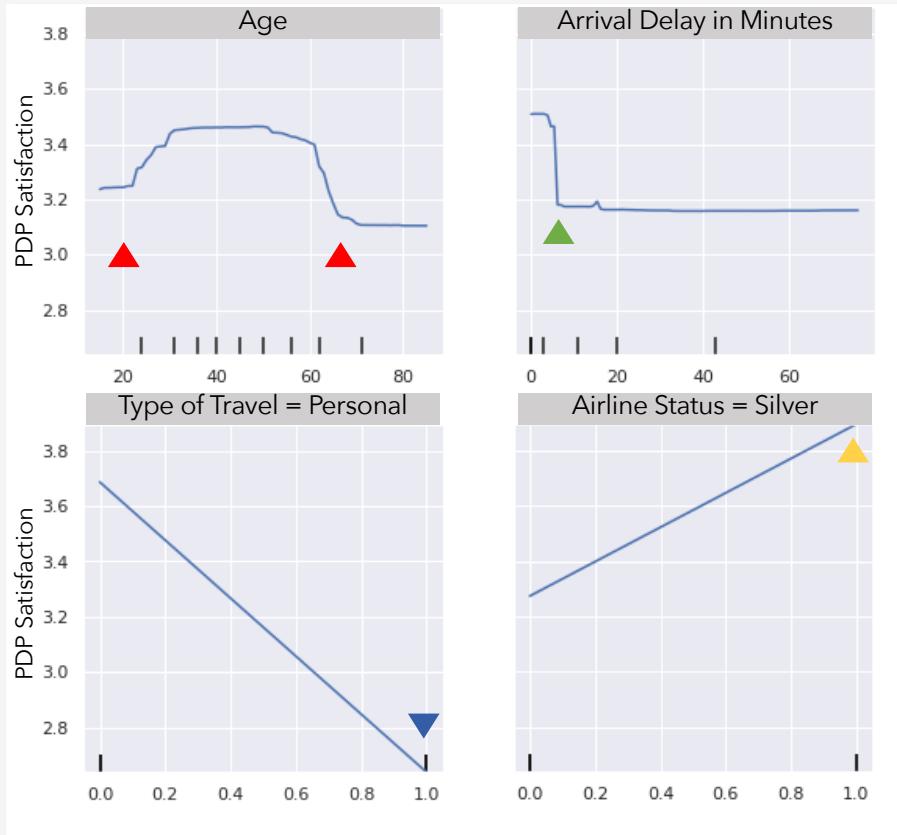
- 1) Where does consumer satisfaction vary and why?**
- 2) Is there a case for investing in satisfaction?**
- 3) Are there gender differences in satisfaction?**

## **Satisfaction Questions**

---

**1) Where does consumer satisfaction vary and why?**

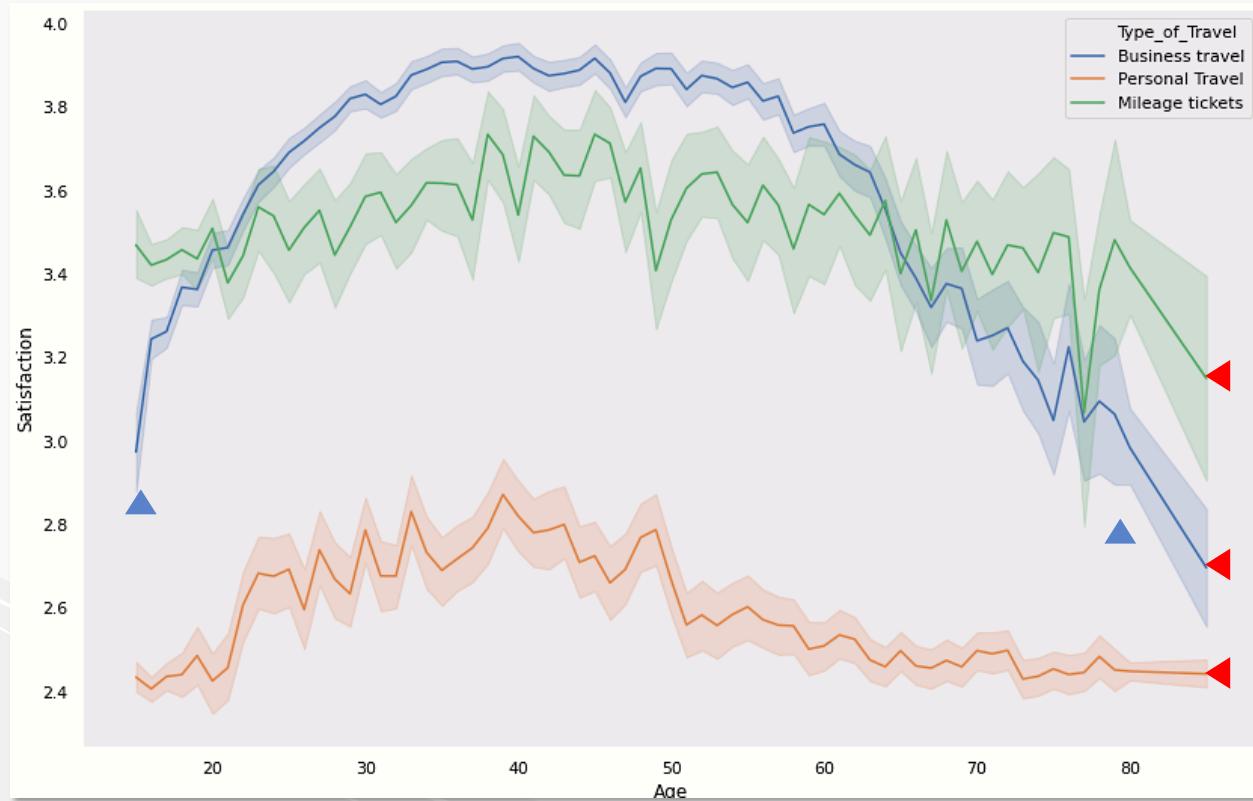
## Satisfaction: Partial Dependency Plots



At right we have the four most important features from our Satisfaction model:

- ▶ Age has low satisfaction in the young and old, something seen across travel type and status
- ▶ Delays become meaningful just before the 10-minute mark
- ▶ Personal travelers have significantly less satisfaction than business and mileage ticket peers
- ▶ Silver status is associated with a satisfaction benefit

## Satisfaction: Important Features



Here we can see two of the most important features for satisfaction

- ▶ There are meaningful differences in satisfaction by type of travel
- ▶ With business travel we see lower satisfaction in younger and older travelers – something seen to a lesser extent with personal and mileage ticket travel

## Satisfaction: Important Features



Silver has the highest satisfaction across most ages

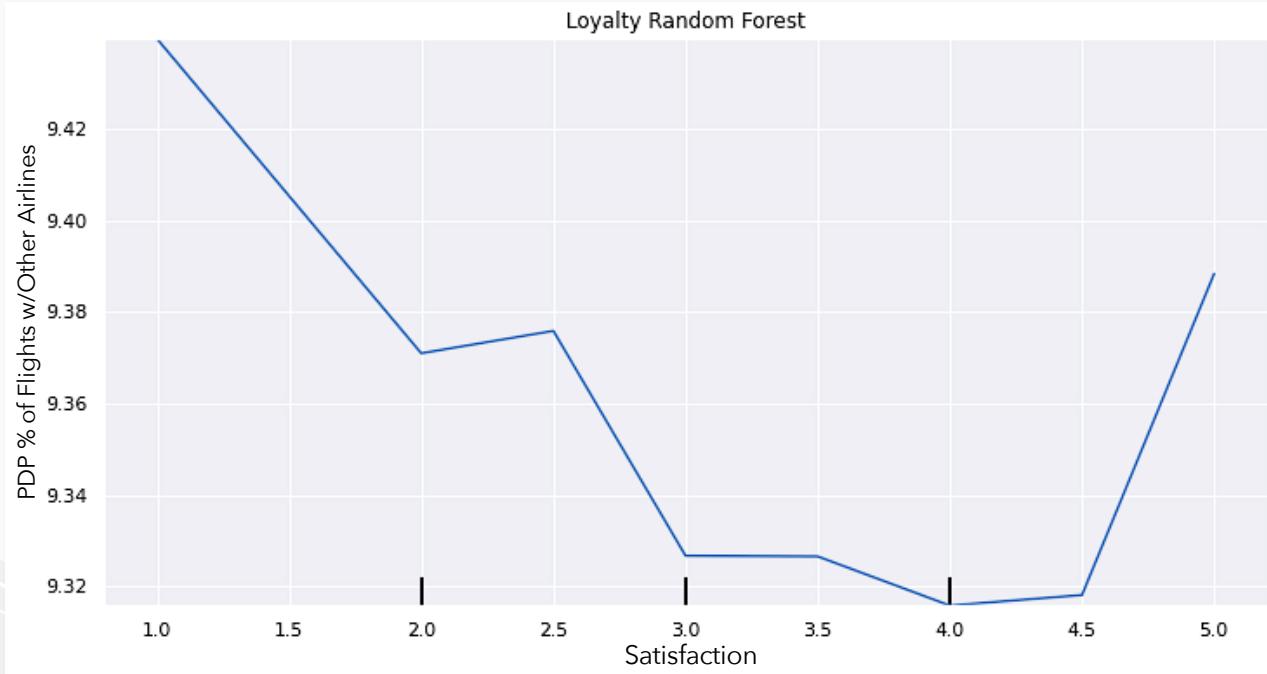
All tiers have lower satisfaction as age increases but the higher tiers (Gold and Platinum) have comparatively lower satisfaction, breaking from the Silver tier at around age 60

## Satisfaction Questions

---

**2) Is there a case for investing in satisfaction?**

## Loyalty: Important Features



We modeled the % of flights with other airlines to see if increasing satisfaction was associated with higher loyalty (lower %)

This is true up until satisfaction scores of 5.0, around 7% of the total traveling population

The movement from 1.0 to 4.5 in satisfaction could be used to help quantify investments in traveler satisfaction

## Understanding Customers

The factors collected allow for accurate understanding of whether a customer will be satisfied



	precision	recall	f1-score
1	0.92	0.99	0.95
2	0.91	0.92	0.92
3	0.89	0.80	0.84
4	0.86	0.93	0.90
5	0.94	0.83	0.88
accuracy			0.89
macro avg	0.91	0.89	0.90
weighted avg	0.89	0.89	0.88

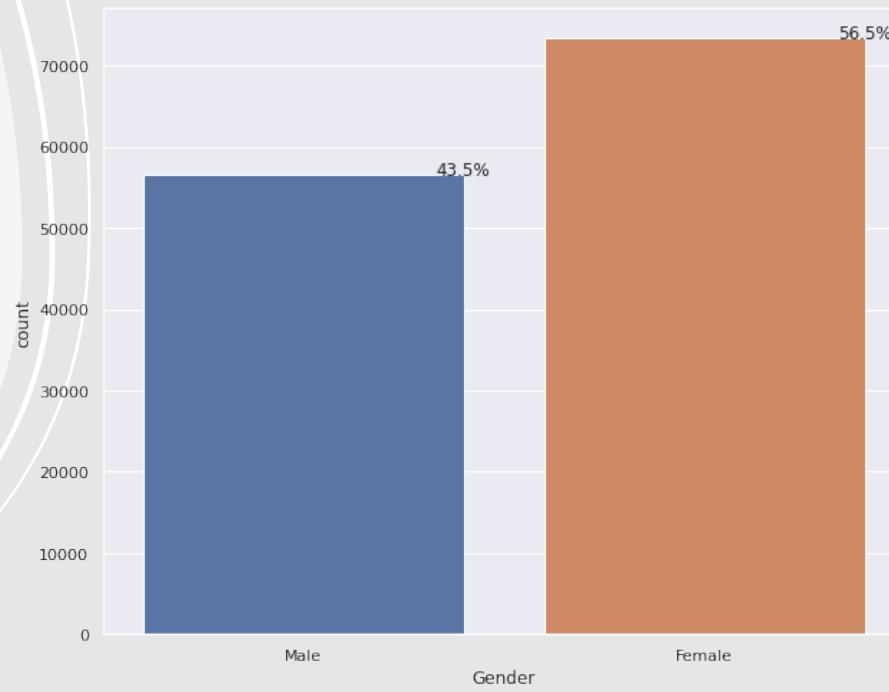
## **Satisfaction Questions**

---

**3) Are there gender differences in satisfaction?**

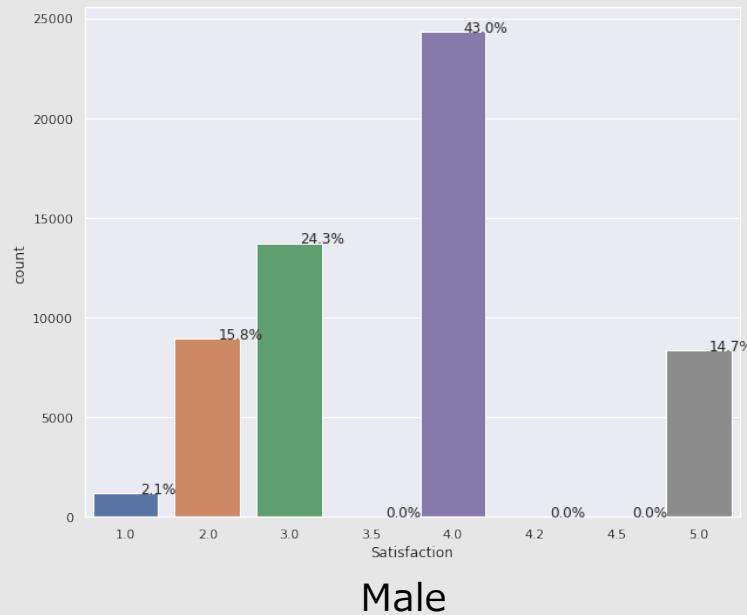
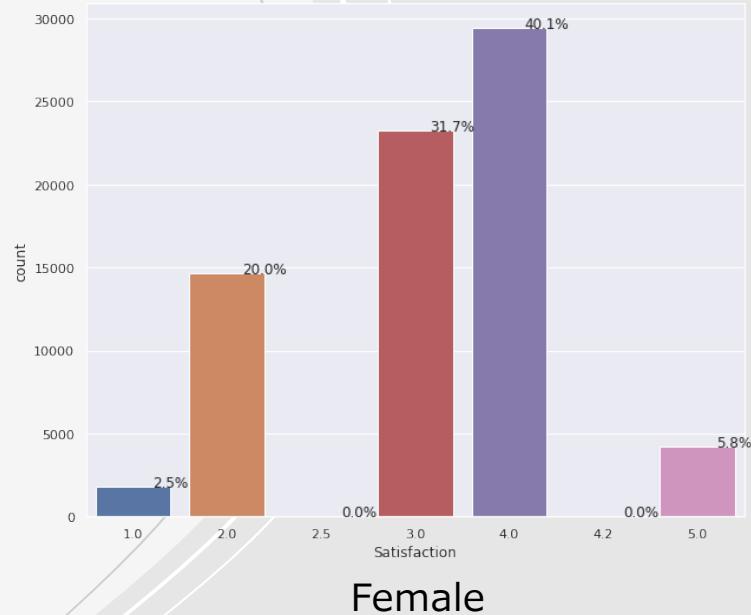
# Exploratory Data Analysis

Female Passengers make up 56.5% of the dataset with +73k flights.



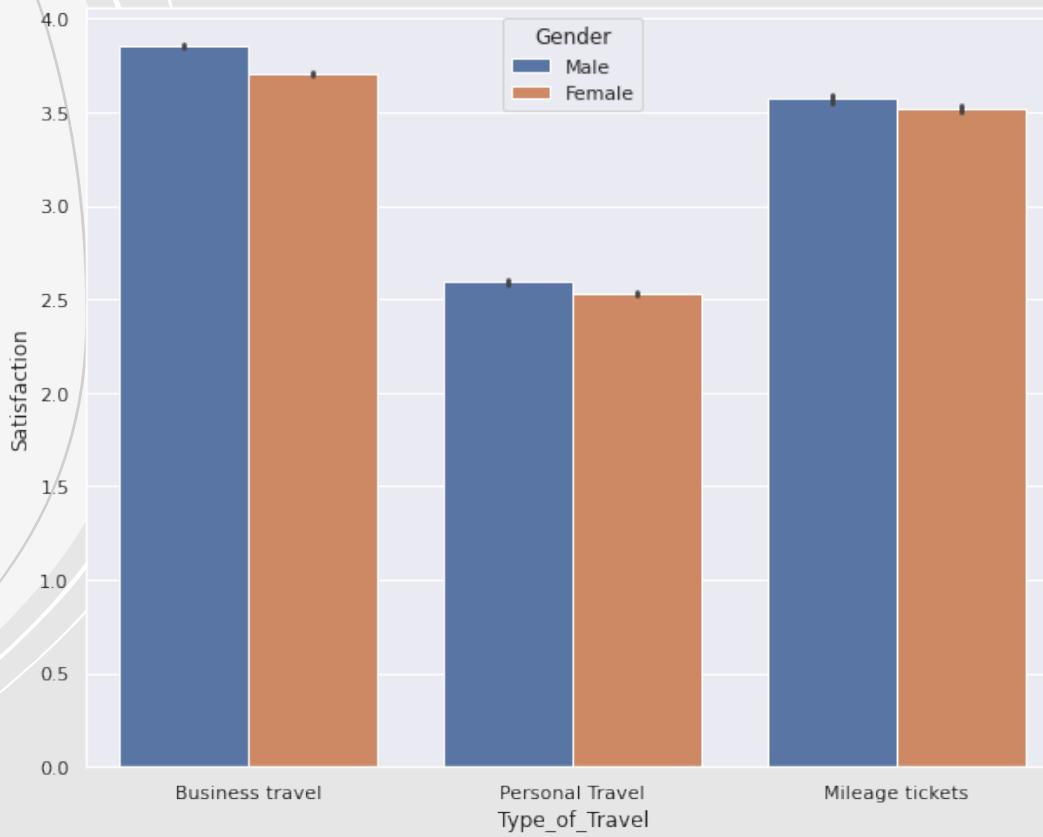
# Exploratory Data Analysis (cont)

Male passengers rated their flights as  $\geq 4$  following 58% of their flights, which is 12pts higher than the 46% for Female ratings.



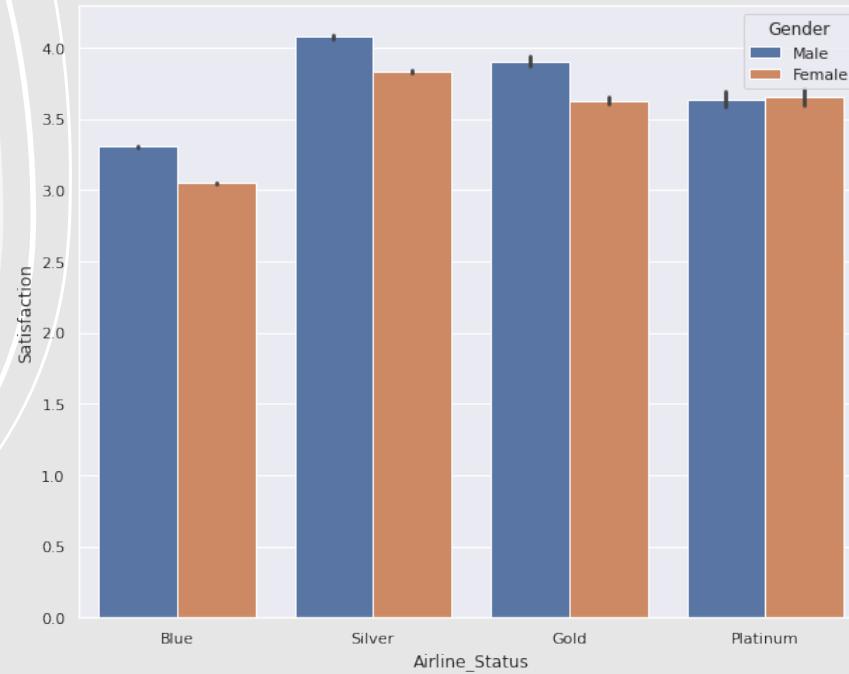
# Exploratory Data Analysis (cont)

Male passengers rated their flights higher, in aggregate, across all 3 Travel Types.



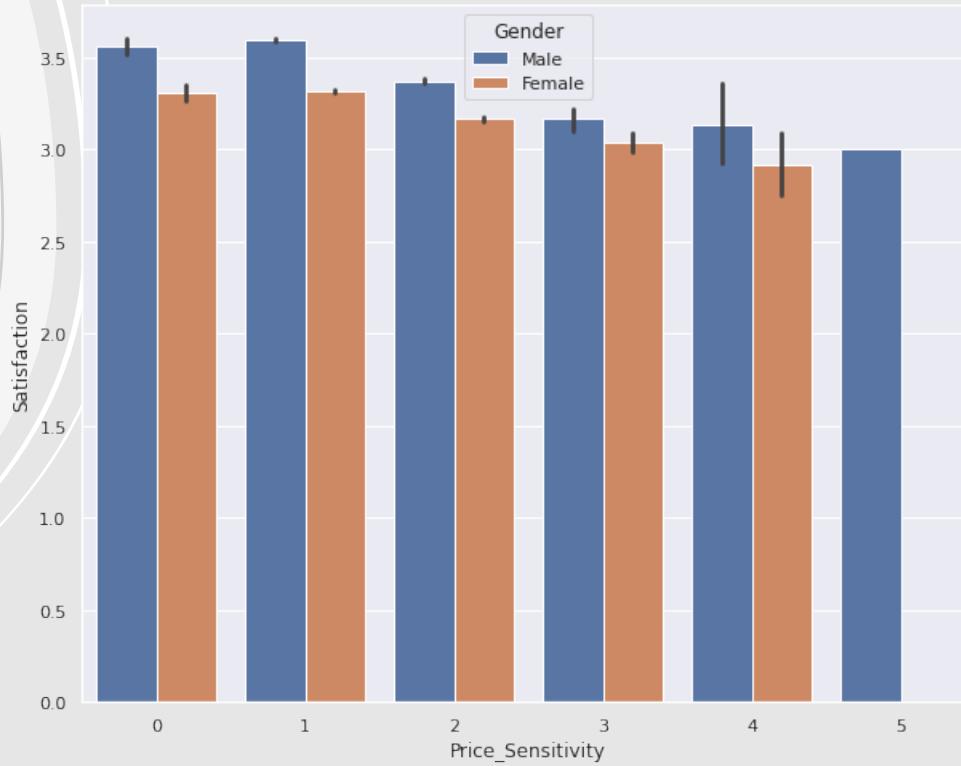
## Exploratory Data Analysis (cont)

Male passengers rated their flights higher, in aggregate, across 3 of 4 Airline Statuses. Females that flew Platinum had a better experience than the males. This is one of the few spots where Females had a higher Satisfaction score than Males.



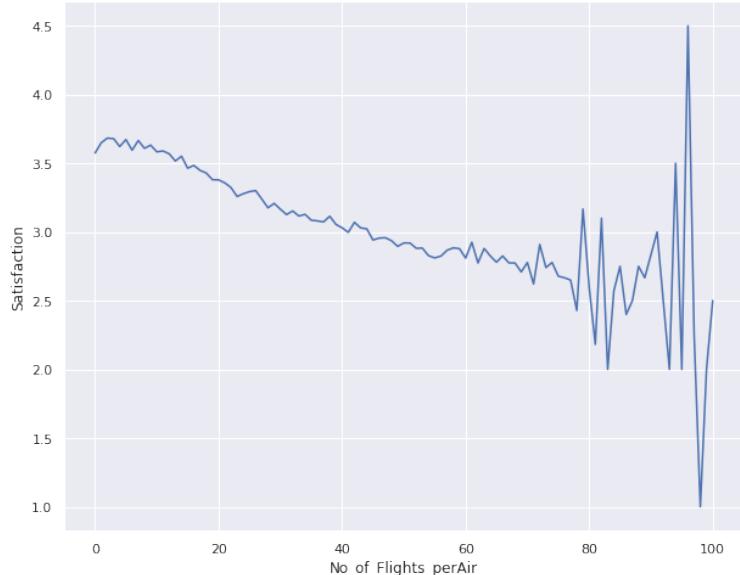
# Exploratory Data Analysis (cont)

Satisfaction scores decreased as flyers were more price sensitive.

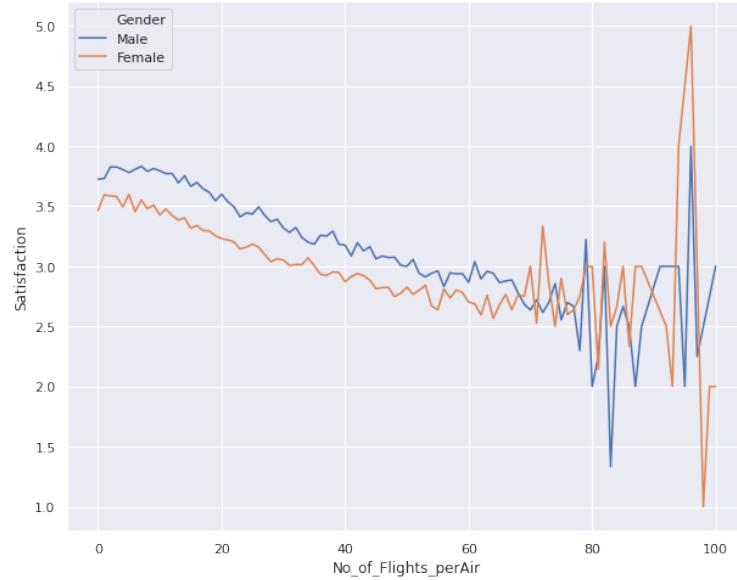


# Exploratory Data Analysis (cont)

Passenger Satisfaction decreases with the increase in # of flights. Female satisfaction has a ~1/2pt lower base of Satisfaction per flight vs Male satisfaction.



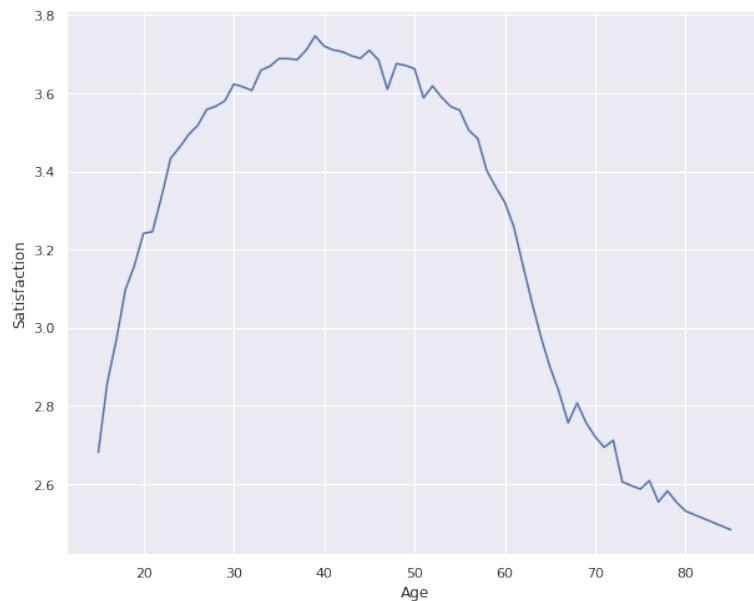
Aggregate



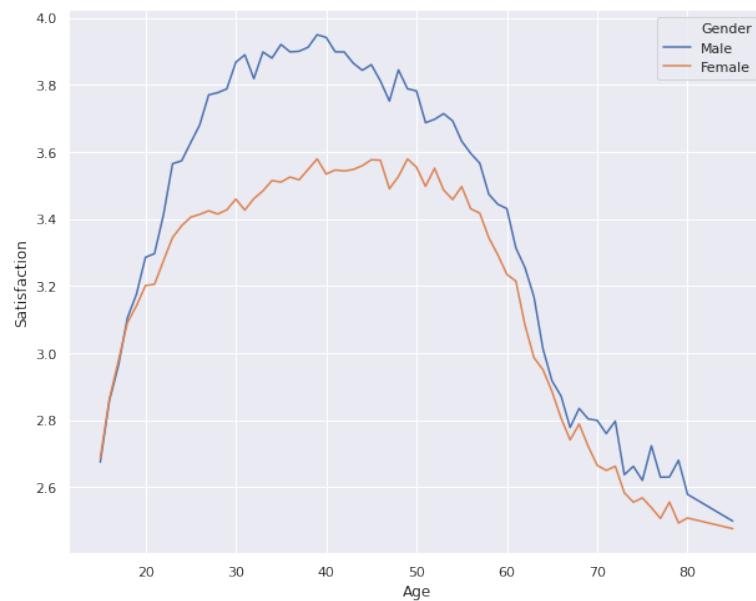
Male/Female Split

# Exploratory Data Analysis (cont)

Passenger Satisfaction and Age showed an interesting relationship. Notice how ages 30 - 55 have a score of  $\geq 3.6$ , then falls off the table on either side of 30 & 55. Looking at the Male/Female split, Females have a lower aggregate score around the age of 20 and beyond.



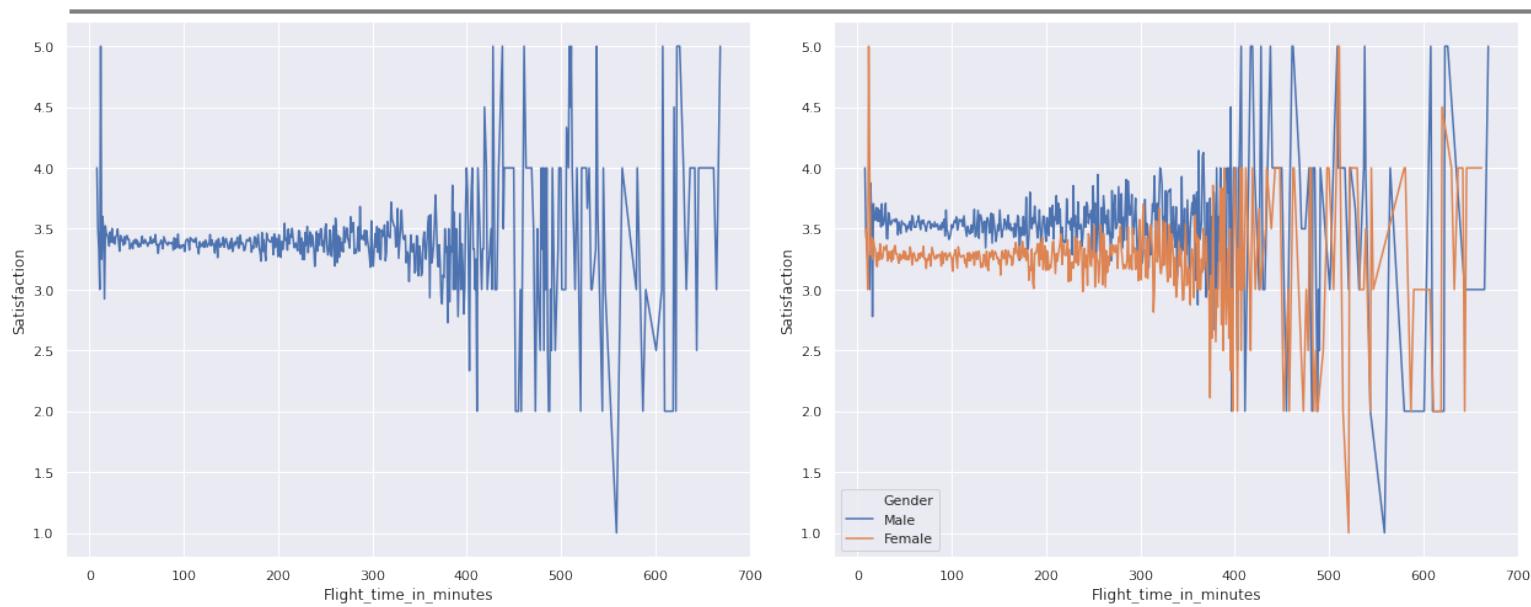
Aggregate



Male/Female Split

# Exploratory Data Analysis (cont)

Passenger Satisfaction and Flight Time increased in variance as flights approached the 6+ hour mark. Looking at the Male/Female split, Females have a lower aggregate score across the board.



Aggregate

Male/Female Split

# Model Performance

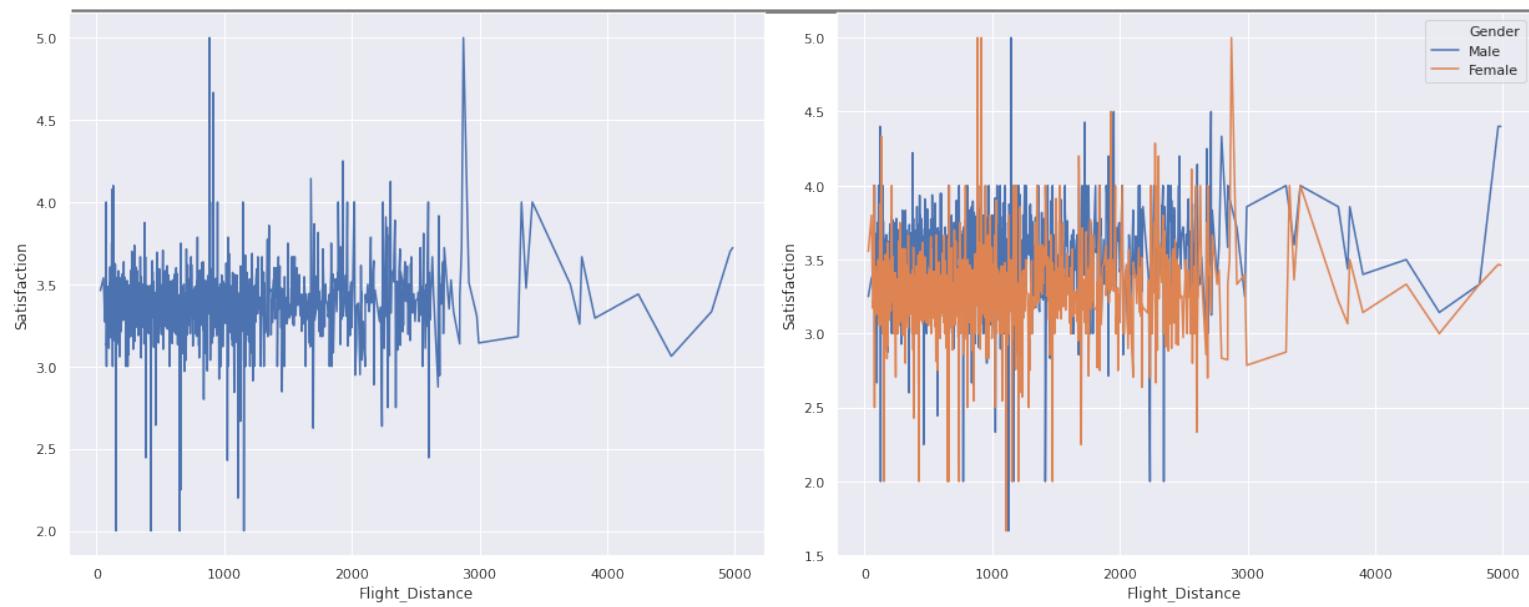
Model	Inputs	Accuracy Measure(s)	Time to Run (secs)	Time to Run (mins)	Time to Run (hrs)
OLS Regression #1	Price_Sensitivity + Age + Gender + Type_of_Travel + Airline_Status + Class + No_of_Flights_perAir + Destination_State + Origin_Status + Flight_cancelled + Flight_time_in_minutes	RMSE = 0.7393147791986532 Adjusted R^2 = 0.42	14.8	0.2	0.0
OLS Regression #2	Price_Sensitivity + Age + Gender + Type_of_Travel + Airline_Status + Class + No_of_Flights_perAir + Flight_cancelled + Flight_time_in_minutes	RMSE = 0.7400339204068892 Adjusted R^2 = 0.419	5.6	0.1	0.0
Random Forest #1	All Variables (dummy + numeric); n_estimators = 100	Accuracy = 80.6% Mean Absolute Error: 0.52 degrees	1,278.5	21.3	0.4
Random Forest #2	All Variables (dummy + numeric); n_estimators = 500	Accuracy = 80.7% Mean Absolute Error: 0.52 degrees	4,932.2	82.2	1.4
XGBoost #1	All Variables (dummy + numeric); n_estimators = 200	Accuracy = 80.7% Mean Absolute Error: 0.52 degrees	193.0	3.2	0.1
Random Forest #3 Sub	Top 13 Features (dummy + numeric); n_estimators = 100	Accuracy = 79.8% Mean Absolute Error: 0.54 degrees	360.0	6.0	0.1
Random Forest #4 Grid Optimized	Top 13 Features (dummy + numeric); n_estimators = 500	Accuracy = 80.4% Mean Absolute Error: 0.52 degrees	1,346.0	22.4	0.4
Random Forest Loyalty	All Variables (dummy + numeric); n_estimators = 100	Accuracy = 65.2% Mean Absolute Error: 2.4 degrees	189.0	3.2	0.1
SVC, kernal = linear	All variables, as dummy. Y input was Satisfaction of 5	f1 score = 1		420	7
SVC, kernal = poly, degree =8	All variables, as dummy. Y input was Satisfaction of 5	f1 score =1		402	6.7
SVC, kernal = rbf	All variables, as dummy. Y input was Satisfaction of 1	f1 score =1		378	6.3
SVC, kernal = linear	Dummy +Numeric. Y input was Satisfaction (Used subset of data only about 15%	f1 score = 1		120	2



# Appendix

# Exploratory Data Analysis (cont)

Passenger Satisfaction and Flight Distance showed no discernable trends. Looking at the Male/Female split, Females have a slightly lower aggregate score across the board.



Aggregate

Male/Female Split

## Sources

Repo:

<https://github.com/itsAmeMario0o/BDA718-PROJECT>

Images:

[https://www.freepik.com/free-vector/airplane-realistic-poster\\_3796059.htm#page=1&query=airline&position=0](https://www.freepik.com/free-vector/airplane-realistic-poster_3796059.htm#page=1&query=airline&position=0)

PPT Template:

<https://www.slidescarnival.com/category/free-templates>