Airline Passenger Satisfaction

Lauren Sweeney

Springboard – DSC

Capstone Project #2

Problem Statement

- Our client, Super Important Airline ("SIA"), wants to increase revenue
- SIA's board of directors has requested that we analyze survey data from their airline passengers. Their question:
    - What separates SIA's "satisfied" airline passengers from their "neutral or dissatisfied" passengers?
- Dataset was procured from SIA's passenger survey, which included data from over 120,000 airline passengers
- We determined which variables were most important for passenger satisfaction by calculating the highest Accuracy, Precision, & Recall scores by building machine learning models

Dataset Variables – Qualitative & Continuous

Variable Name	Information re Variable
Gender	Male | Female
Customer Type	First-Time | Returning
Type of Travel	Business | Personal
Class	Business | Economy | Economy Plus
Age	Numerical ranging from 7 to 85
Flight Distance	Numerical
Departure Delay	Numerical
Arrival Delay	Numerical

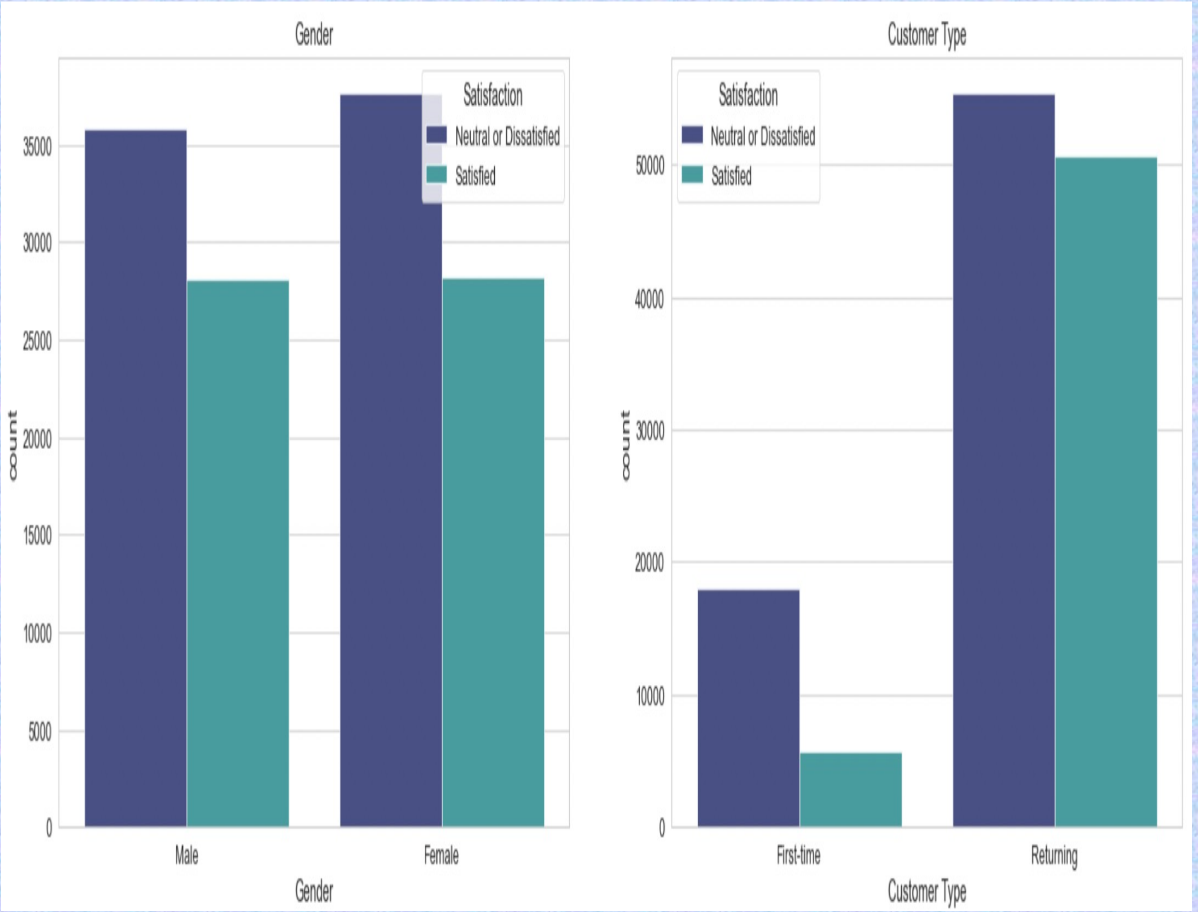
Dataset Variables – Passenger Survey Questions

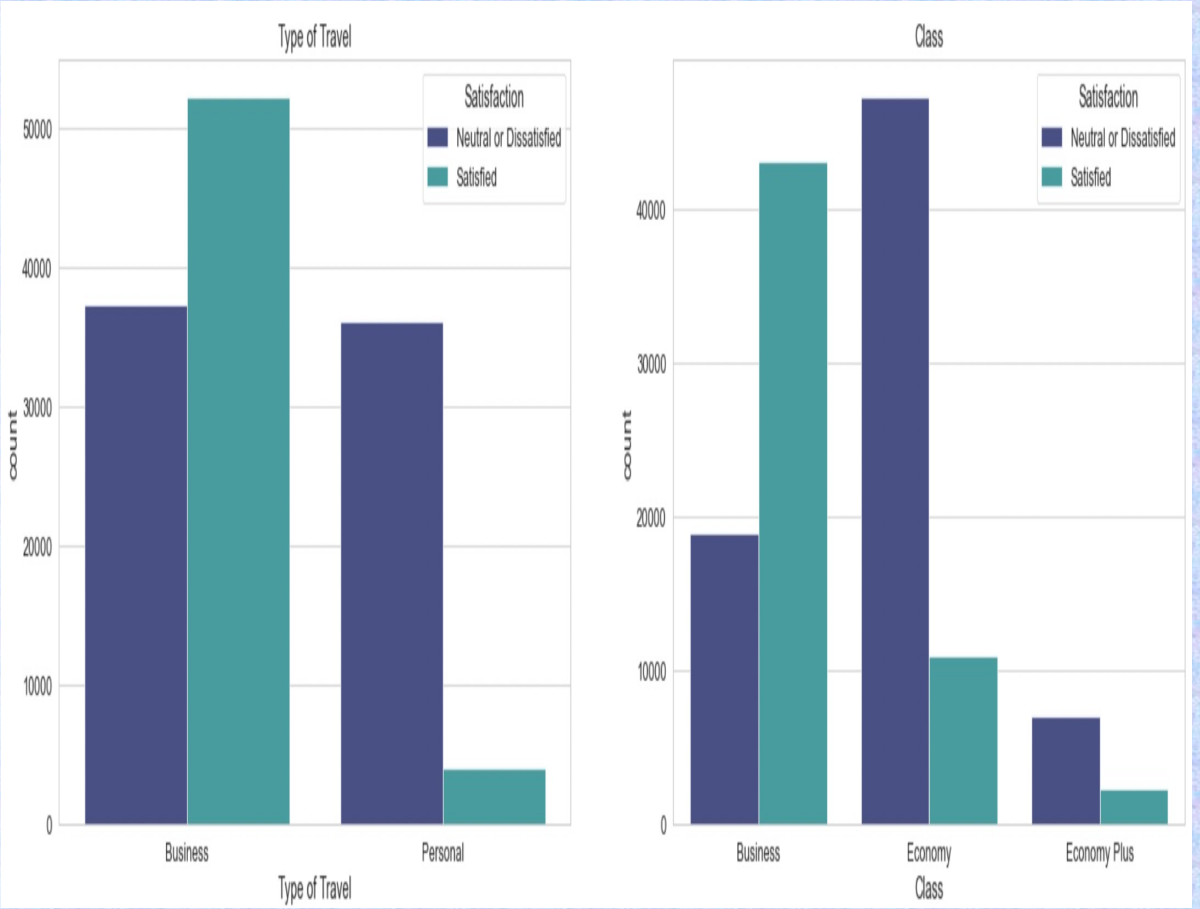
Variable Name	Information re Variable
Departure & Arrival Time Convenience	Numerical ranging between 0 & 5
Ease of Online Booking	Numerical ranging between 0 & 5
Check-In Service	Numerical ranging between 0 & 5
Online Boarding	Numerical ranging between 0 & 5
Gate Location	Numerical ranging between 0 & 5
On-Board Service	Numerical ranging between 0 & 5
Seat Comfort	Numerical ranging between 0 & 5
Leg Room Service	Numerical ranging between 0 & 5
Cleanliness	Numerical ranging between 0 & 5
Food & Drink	Numerical ranging between 0 & 5
In-Flight Service	Numerical ranging between 0 & 5
In-Flight WIFI Service	Numerical ranging between 0 & 5
In-Flight Entertainment	Numerical ranging between 0 & 5
Baggage Handling	Numerical ranging between 0 & 5

Determining Project Success & Constraints

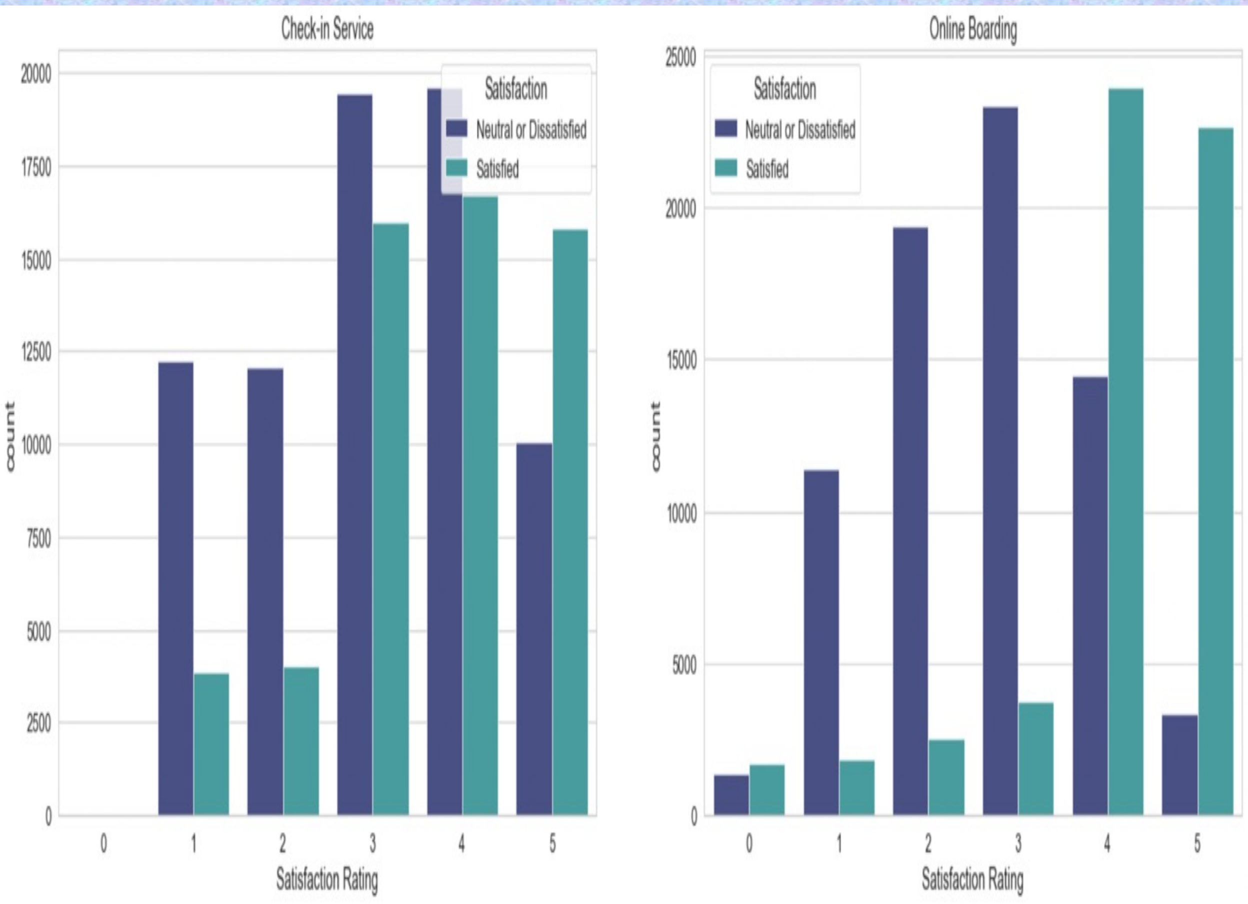
- Success:
  - Determining which variables impact the satisfaction of passengers the most
  - SIA can best tolerate False Negatives than False Positives, therefore, we needed to determine which model had the highest precision
- Constraints:
  - SIA's competitors were not surveyed
  - Costs of implementation not available

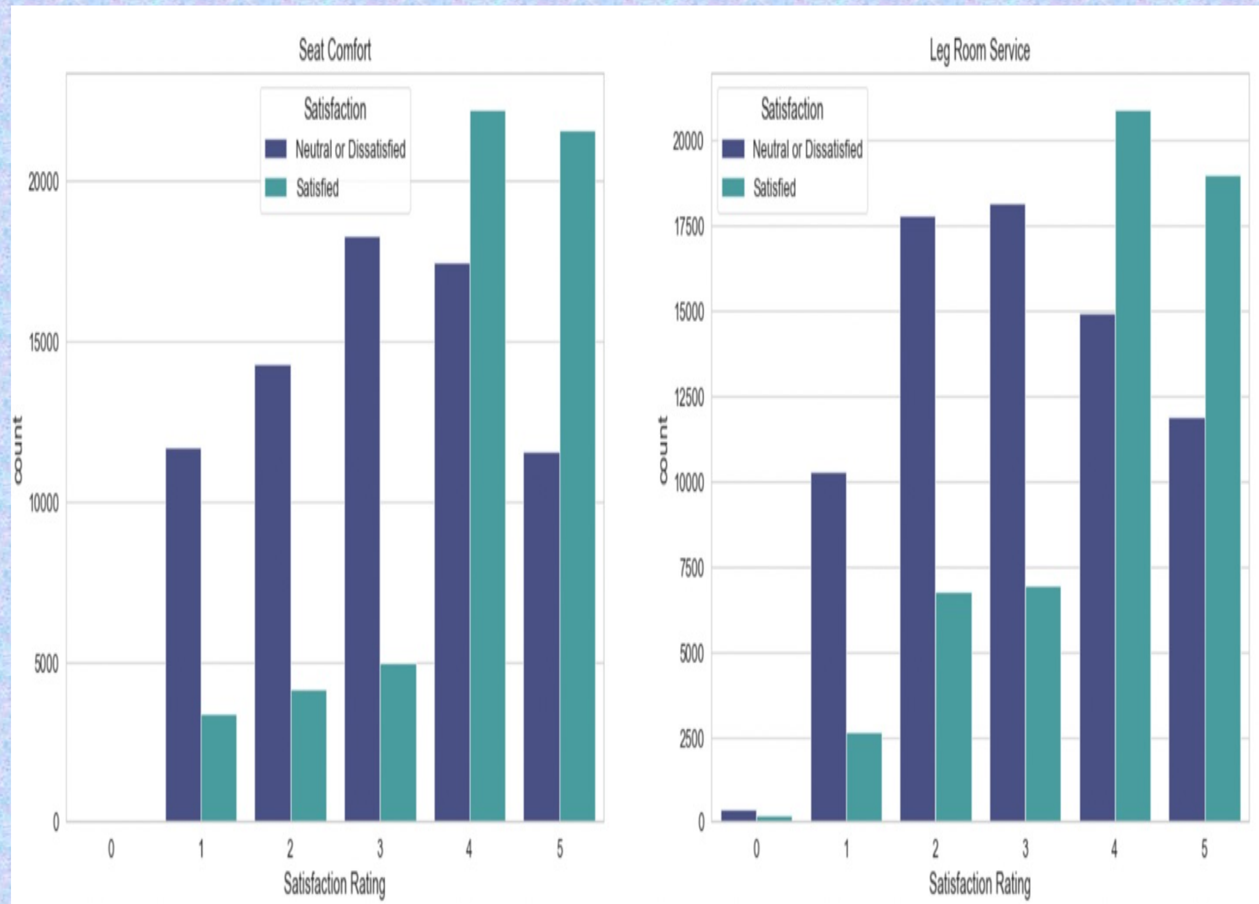
Exploratory Data Analysis



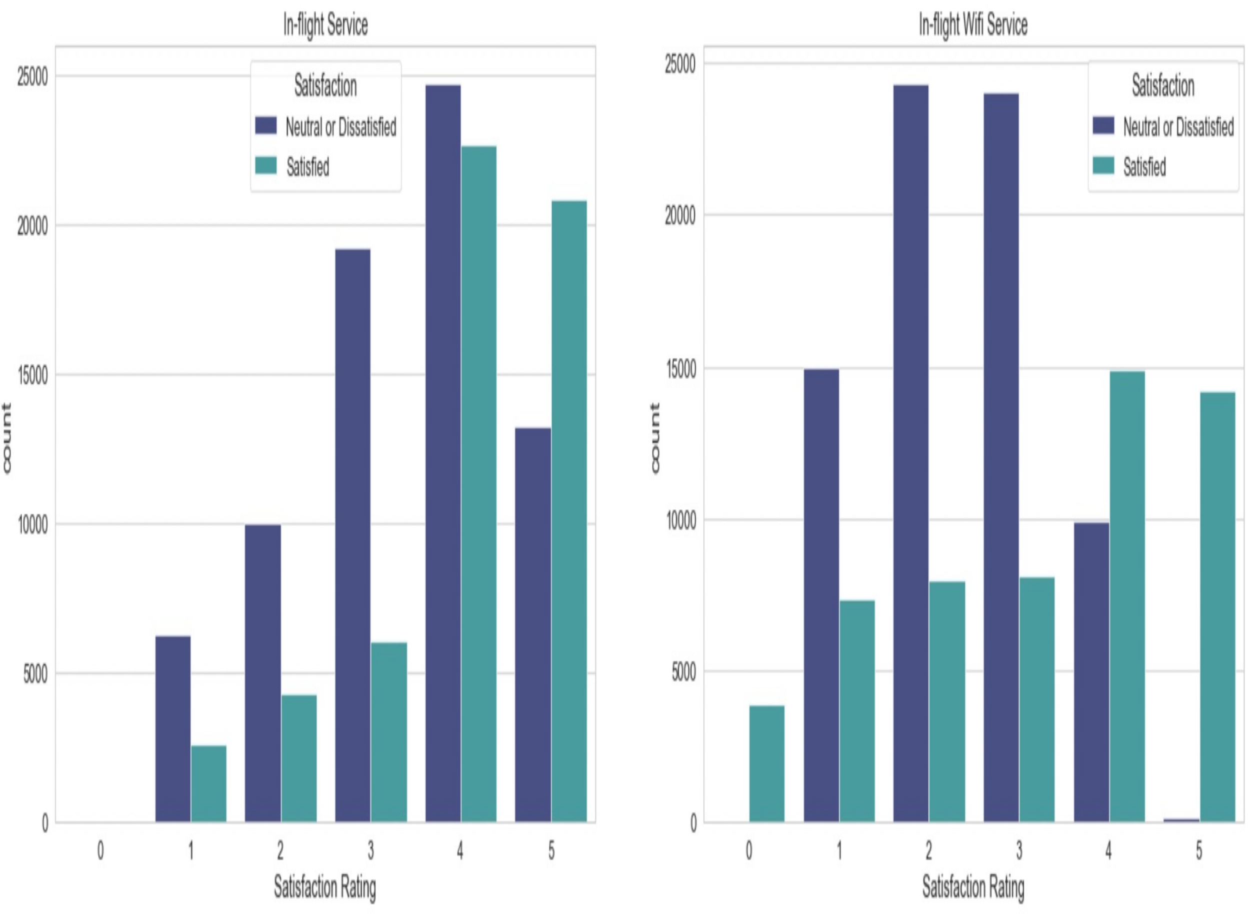


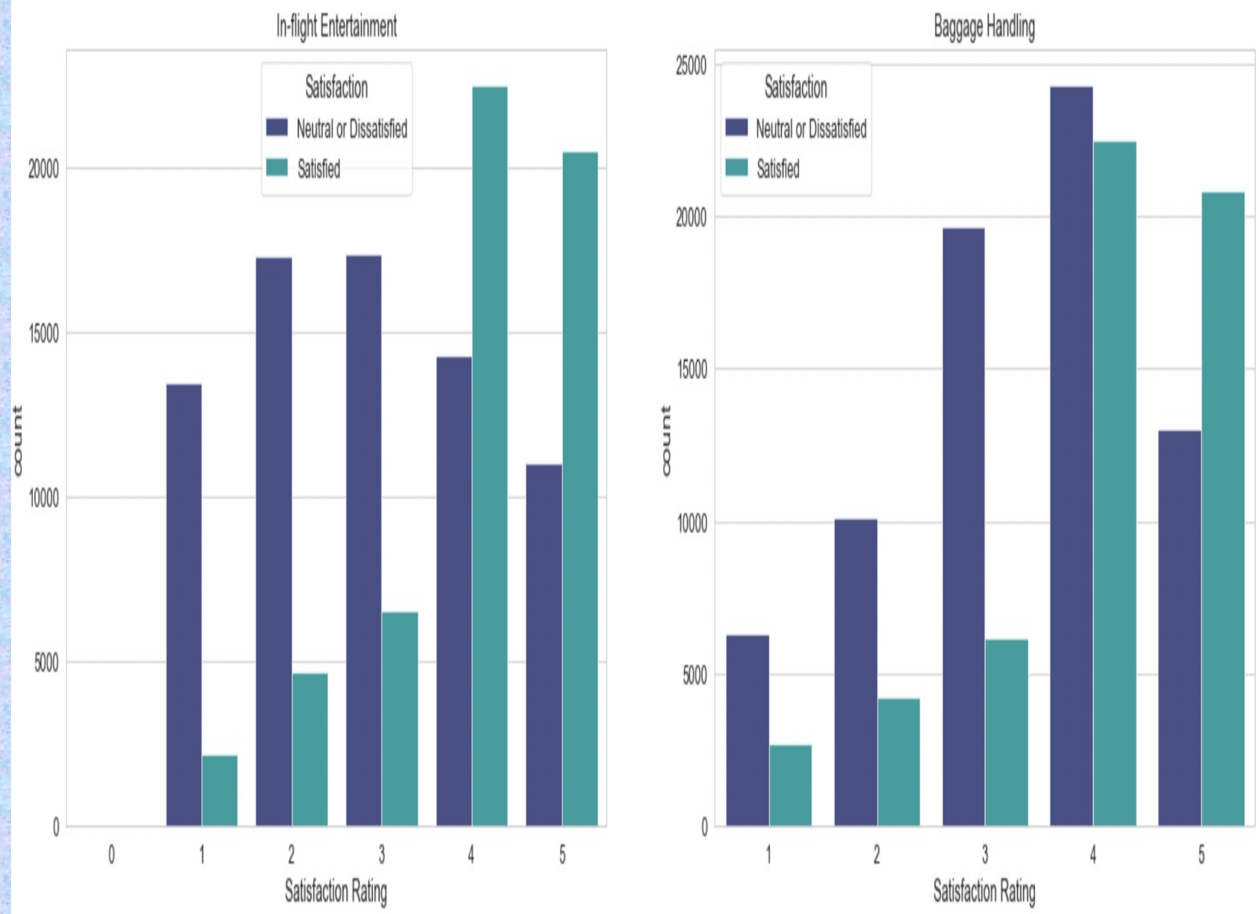
Exploratory Data Analysis



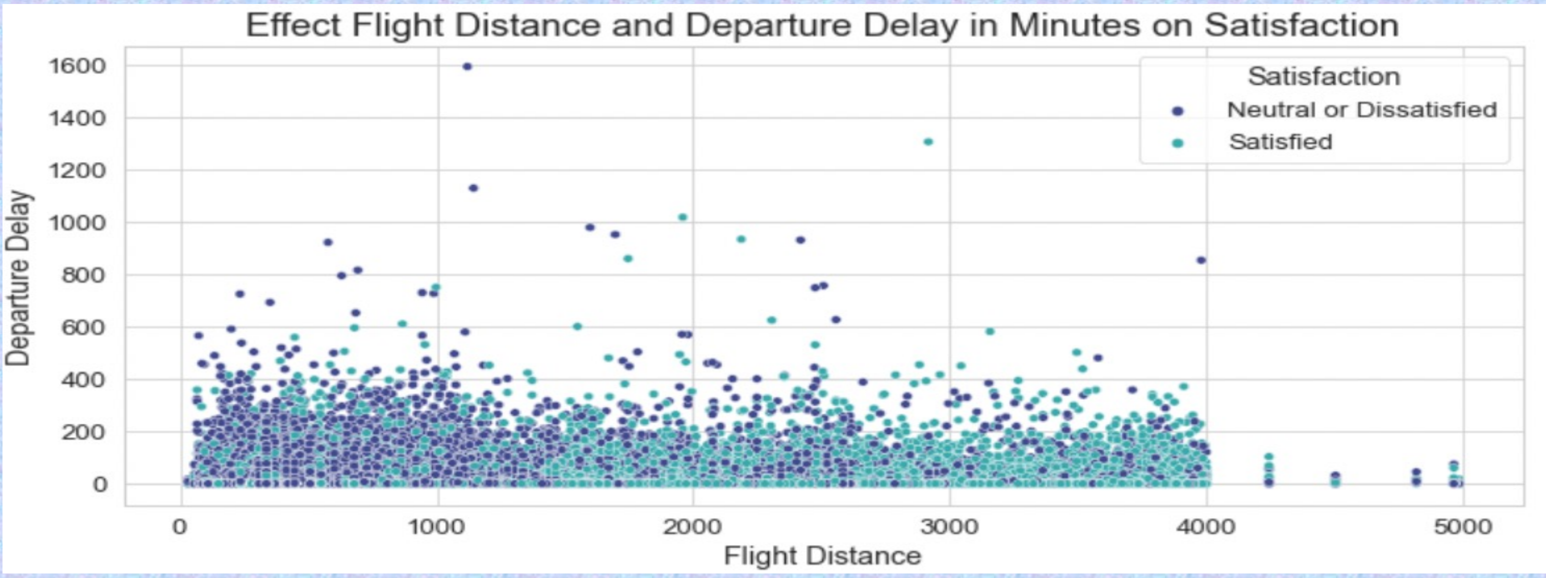


Exploratory Data Analysis





Exploratory Data Analysis



Model Selection

Model	Test Precision	Test Recall	Test F1-Score
Logistic Regression – RUS	0.68	0.58	0.63
Logistic Regression – SMOTE	0.68	0.59	0.63
Random Forest – RUS	0.97	0.96	0.97
Random Forest – SMOTE	0.96	0.98	0.97

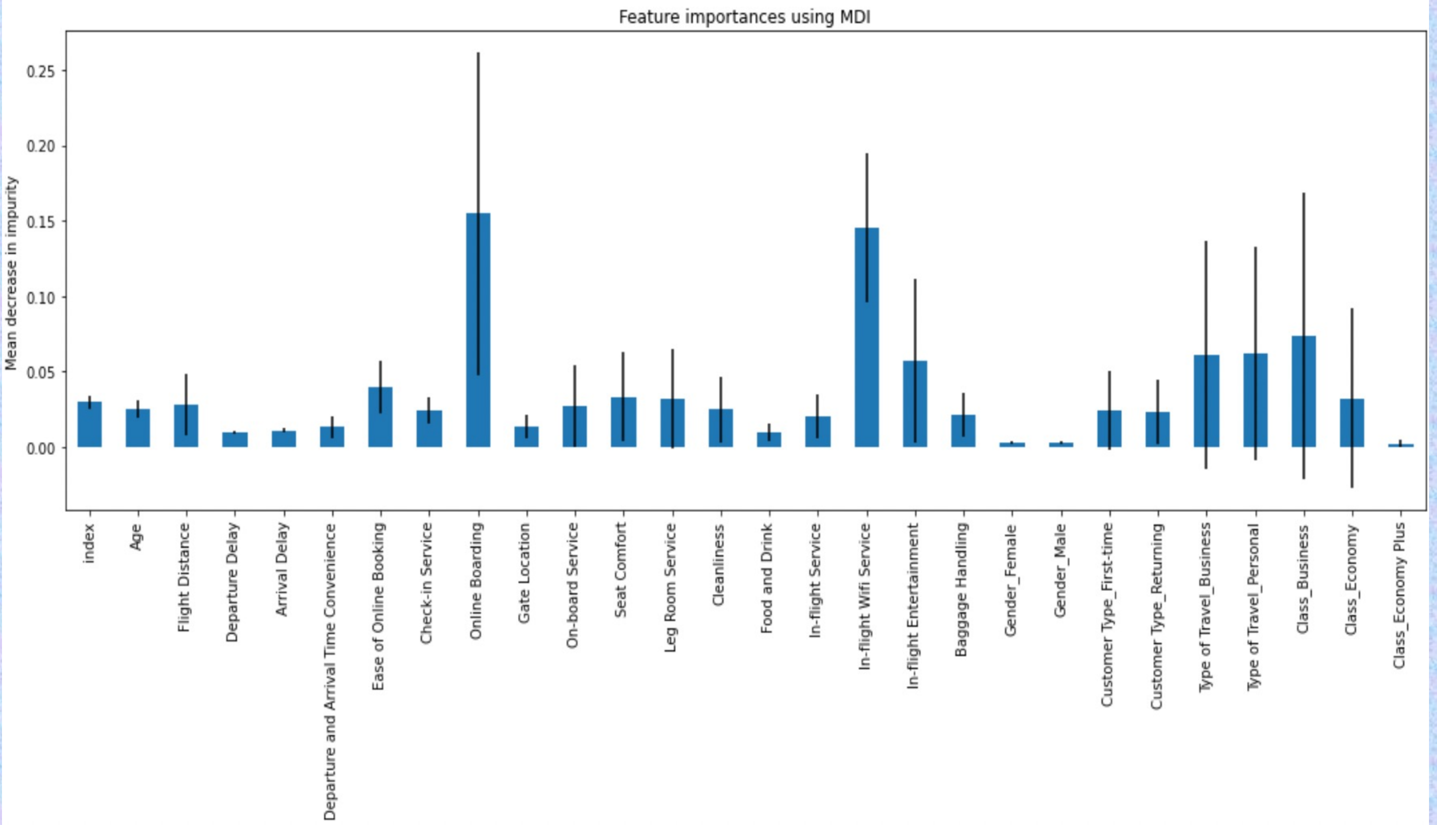
In total, we created 4 models:

- Logistic Regression using random under sampling
- Logistic Regression using SMOTE
- Random Forest using random under sampling
- Random Forest using SMOTE

- We performed hyperparameter tuning on the Random Forest models, which did not change their performance metrics.
- The Random Forest model using random under sampling still slightly outperformed the Random Forest model using SMOTE
- As a result, we determined that the best model for the needs of our client was the Random Forest model using random under sampling, due to its test precision of 0.97

Feature Importances

- Since we identified the best model for our client, we were then able to use that model to determine which features were the most important to the overall satisfaction of passengers
- The most important features:
  - Online Boarding
  - In-Flight WIFI Service
  - Class_Business
  - Type of Travel_Personal
  - Type of Travel_Business
  - In-Flight Entertainment
  - Ease of Online Booking
  - Seat Comfort
  - Class_Economy
  - Leg Room Service



Conclusions & Recommendations

- What did we learn during exploratory data analysis?
- What did we learn after building machine learning models?
- Recommendations for the future:
  - Perform similar analysis on competitors to determine whether there are differences in passenger satisfaction
  - After enacting changes, perform analysis on a future passenger survey to determine whether those changes increased overall passenger satisfaction
  - SIA should consider enacting changes to features that performed poorly

Thank You!