



©arlan Santos

Airline Passenger Satisfaction: A Binary Classification Problem

Arlan Santos
Capstone Project
28/01/2022

BIO

- Former Systems Analyst at Marsh
- Bachelor's of Science Degree in Electronics and Communications Engineering
- Self-proclaimed AVGeek



OUTLINE

Introduction

- Problem Statement

Data Preparation

- Data Cleaning

EDA

- Feature Engineering, Distribution

Classification

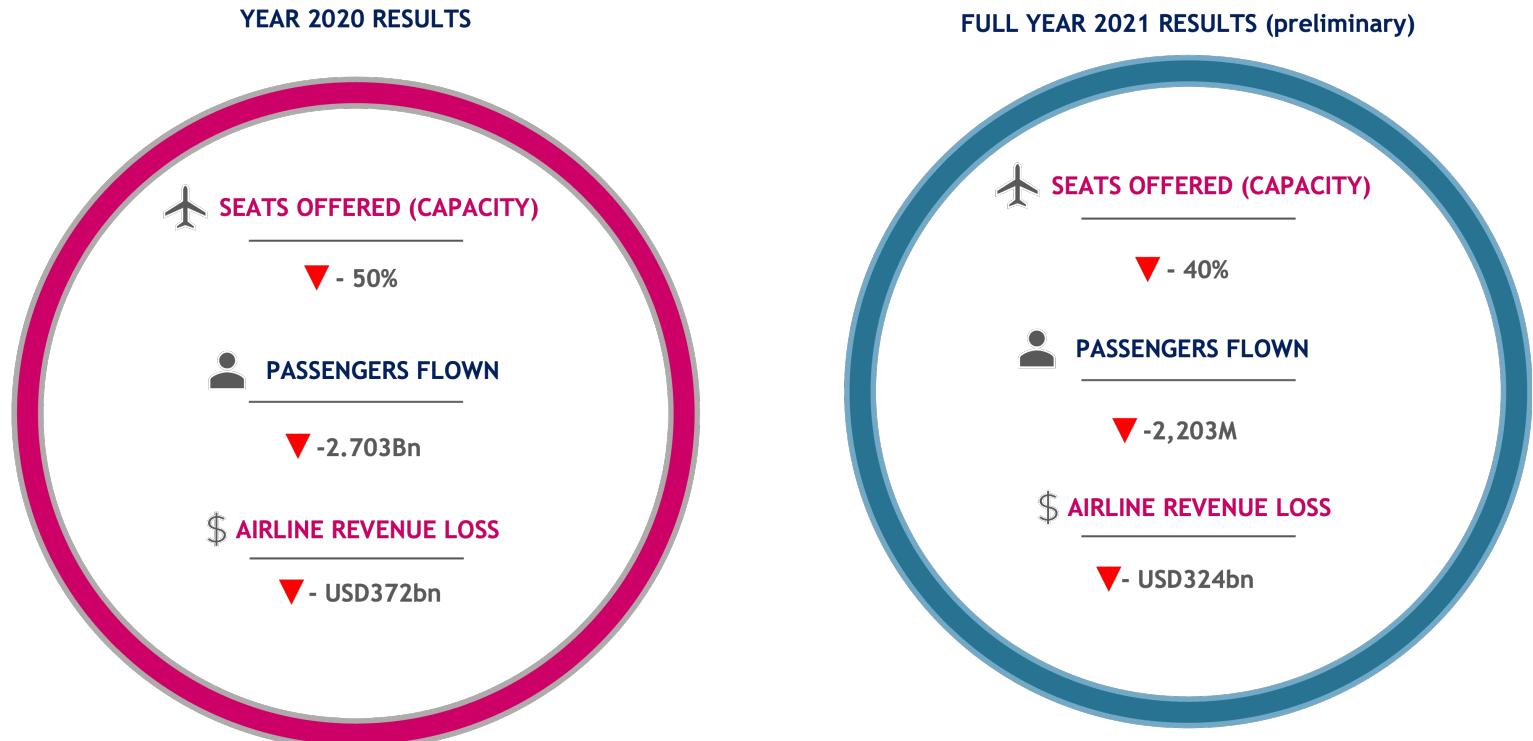
- Model Training and Evaluation

Conclusion

- Summary and Next Steps

CURRENT SITUATION - GLOBAL

- The **Civil Aviation industry** has suffered immensely due to the travel restrictions brought about by the pandemic
- UN ICAO: Border Closures & Restrictions:
 - ▼ Flights
 - ▼ Passenger
- **FINANCIAL LOSS!**



CURRENT SITUATION – QANTAS GROUP

Posted a **FULL YEAR LOSS** in 2021:

- ▼ \$4.18 Billion
- ▼ 80% reduction in seats offered
- ▼ 72% reduction in passengers flown

Source:

https://investor.qantas.com/FormBuilder/_Resource/_module/doLLG5ufYkCyEPjF1tpgyw/file/annual-reports/2021-Annual-Report.pdf

CURRENT SITUATION – QANTAS GROUP

However... Qantas Loyalty remained **STRONG**

- ▲ Maintained 91% profitability
- ▲ Generated \$1 Billion Gross Cash
(achieving record member satisfaction)
- ▲ 200,000 new Frequent Flyer members



Source:

[1] https://investor.qantas.com/FormBuilder/_Resource/_module/doLLG5ufYkCyEPjF1tpgyw/file/annual-reports/2021-Annual-Report.pdf

[2] <https://www.qantasnewsroom.com.au/media-releases/qantas-group-posts-significant-loss-from-full-year-of-covid/>

IMPORTANCE OF CUSTOMER SATISFACTION

Qantas introduced a Recovery Plan:

- Maintain **Customer Satisfaction**
- Return to double digit **Qantas Loyalty** earnings

- Satisfaction means...
- Increase Trust leads...
- Increase Profits



PROBLEM STATEMENT

- Which **flight scenarios** and **key areas of services**, both soft and hard products, should an airline allocate its budget and resources to in order to ensure passenger satisfaction and retention?
- To achieve the goal of returning to double-digit profitability, what are the **top important factors** that guarantee an enjoyable flight for the airline's **most loyal customers**?
- Given a collection of feedback from passengers, how can the airline carefully **assess whether the flight experience was positive or not**?

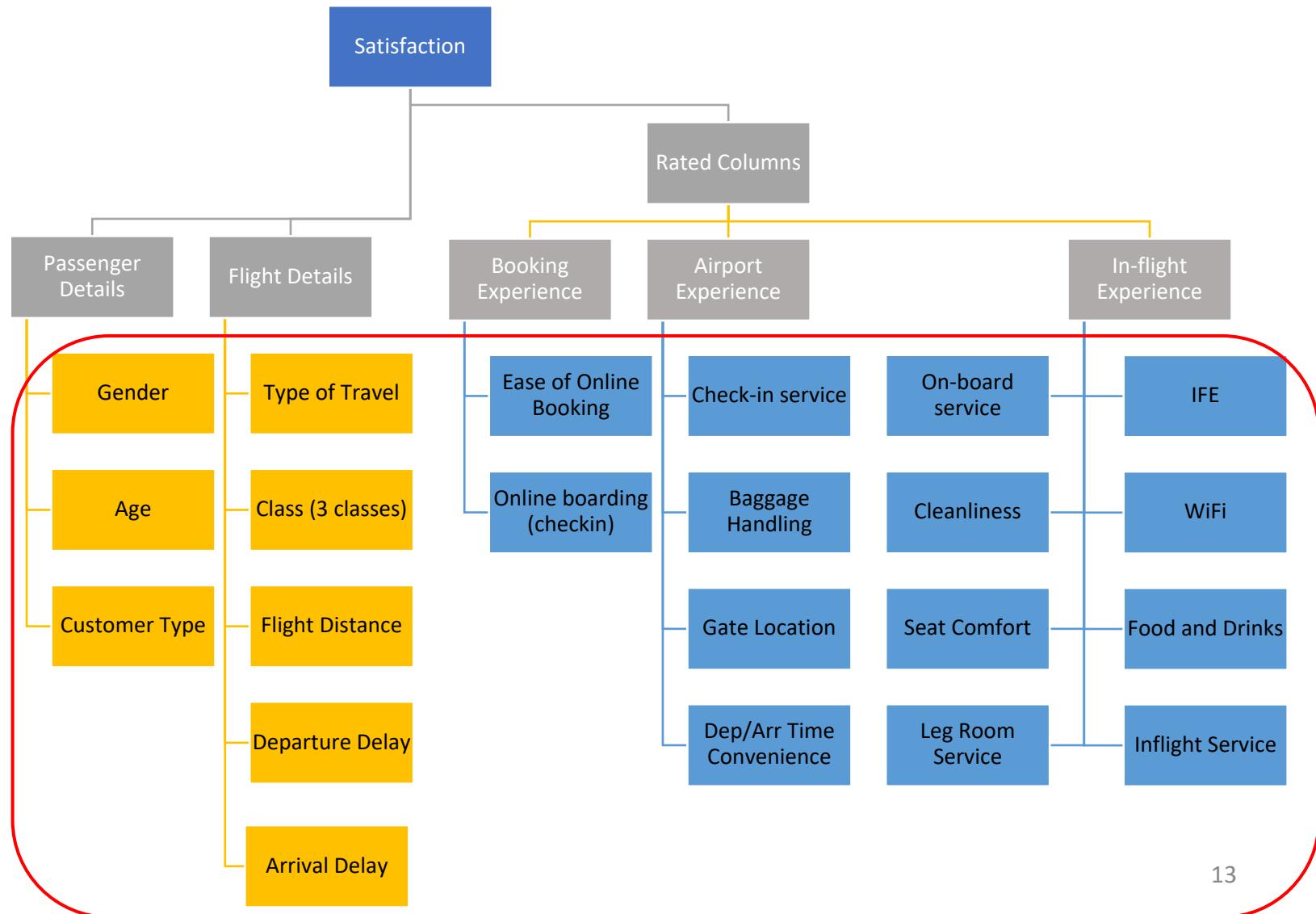


©carlansantos

DATA . EDA . CLASSIFICATION . CONCLUSION

DATA OVERVIEW

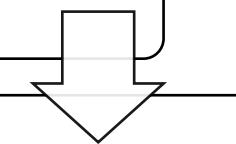
- Source: Kaggle (2018)
- 129,880 passengers
- 25 columns
- 0 duplicates
- Target: Satisfaction



DATA CLEANING

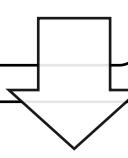
Removed Null Columns

- Arrival Delay in Minutes – 0.3% replaced NaN with 0s (assumption no delay)



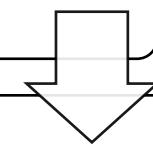
Removed Irrelevant Columns

- Unnamed Column



Reset DataFrame index

- ID Column has the same number of unique elements as the dataframe's row count.



Performed Feature Engineering

- Encoding of categorical variables
- Conversion of continuous variables to discrete variables



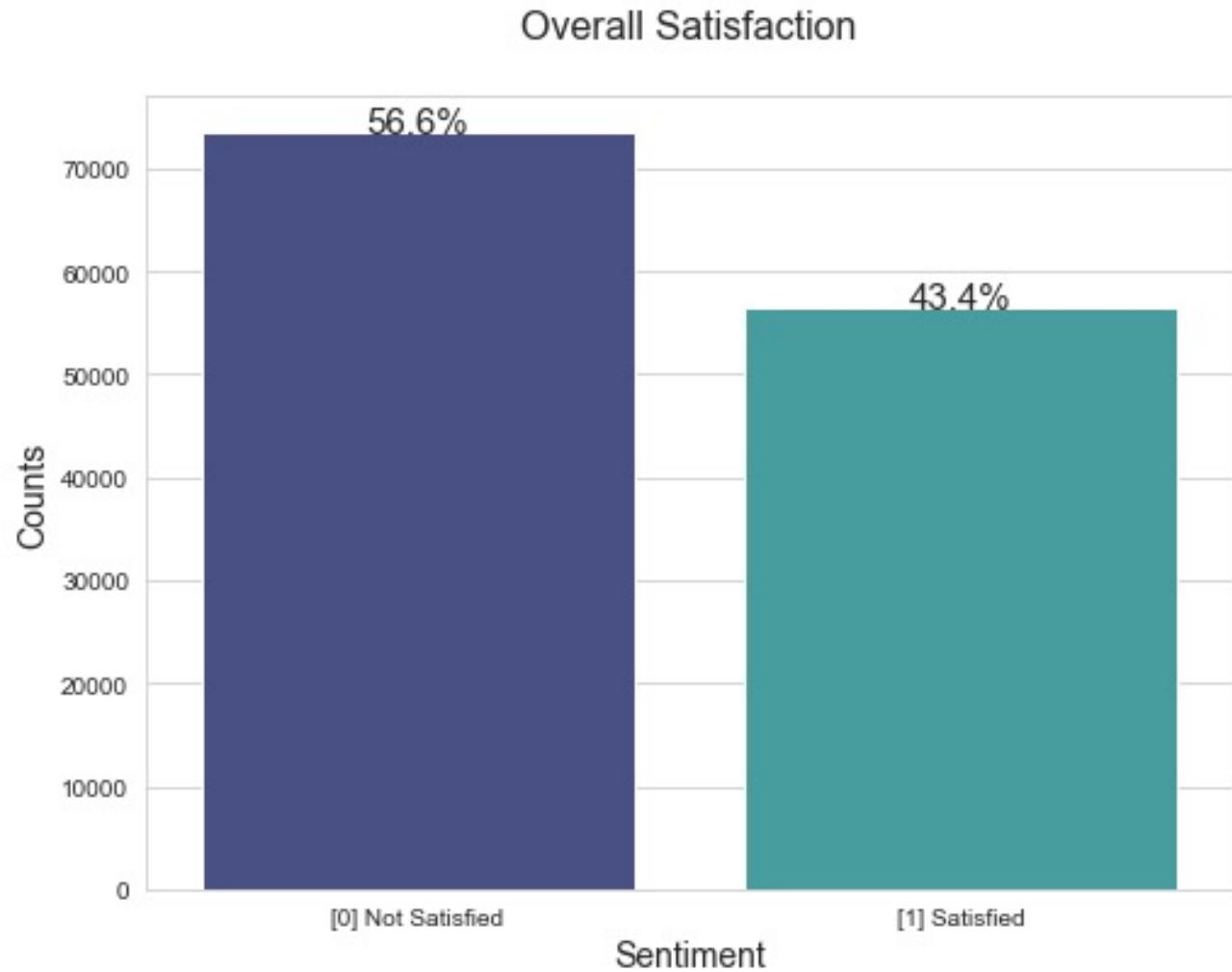
©carlonsantos

DATA . **EDA** . CLASSIFICATION . CONCLUSION

[Target] SATISFACTION DISTRIBUTION

SUMMARY:

- [0] Neutral/Not Satisfied – 73,452
- [1] Satisfied – 56,428



FEATURE SELECTION

Features Correlating with Satisfaction



SELECTION CRITERIA

- Correlation: $> +/- 0.1$
- Retain the feature engineered columns, but
- Avoid multi-collinearity:
 - Age vs Age Group
 - Flight Distance vs Flight Distance
 - Drop 1 out of 3 Flight Cabin Classes

DROPPED COLUMNS

- Flight Distance
- Arrival Delay in Minutes
- Departure Delay in Minutes
- Age
- class_Eco Plus
- Arrival Delay Group
- Departure Delay Group
- Departure/Arrival time convenience
- Gender
- Gate location

FEATURE SELECTION

Features Correlating with Satisfaction



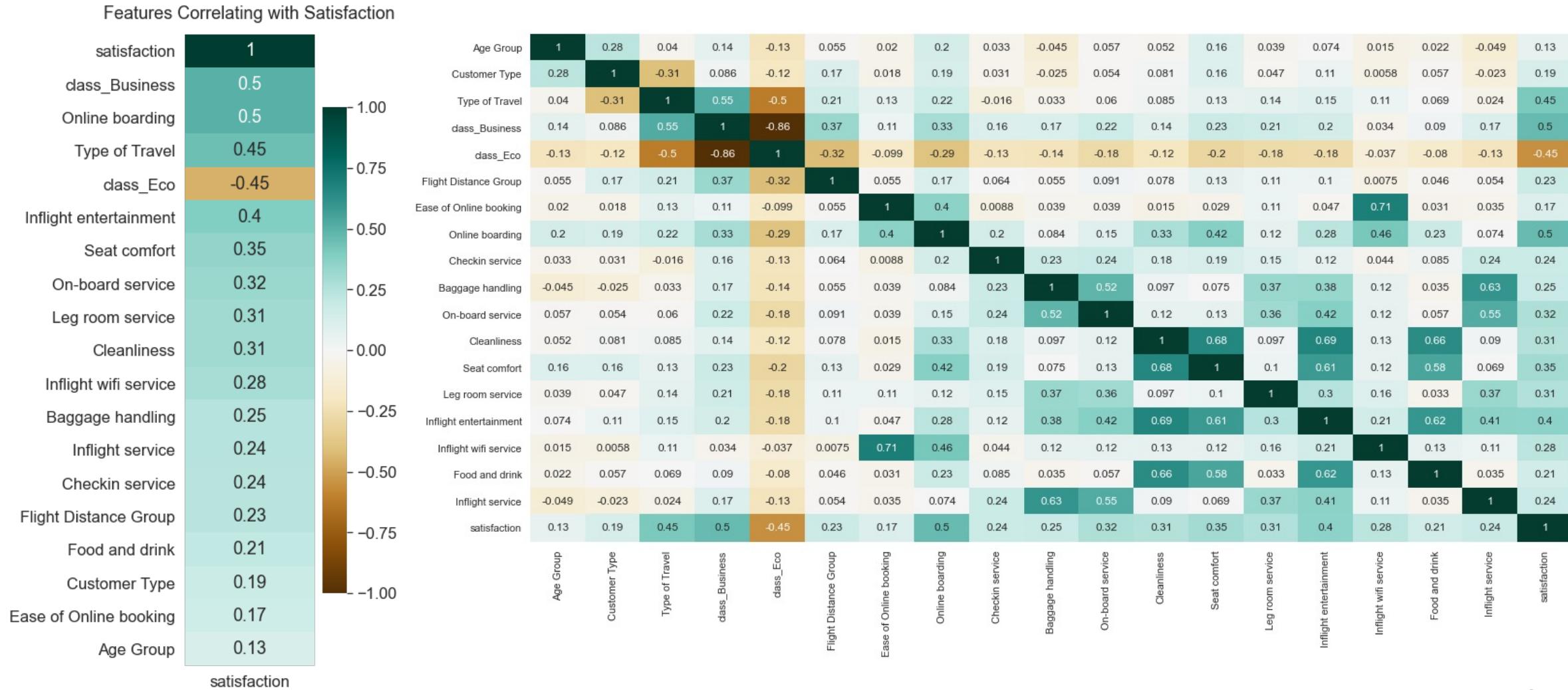
SELECTION CRITERIA

- Correlation: $> +/- 0.1$
- Retain the feature engineered columns, but
- Avoid multi-collinearity:
 - Age vs Age Group
 - Flight Distance vs Flight Distance
 - Drop 1 out of 3 Flight Cabin Classes

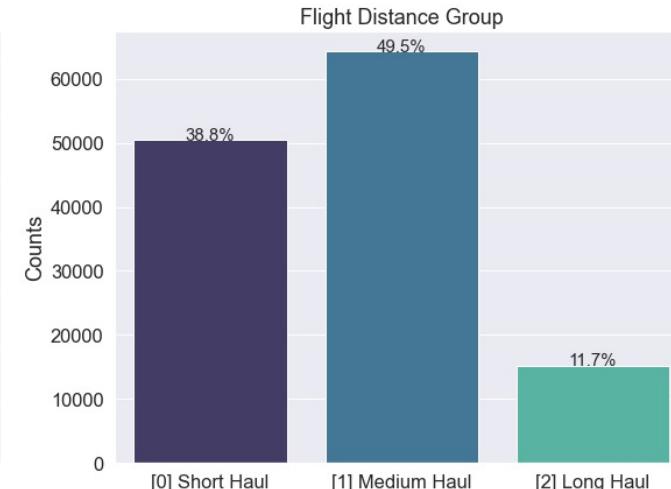
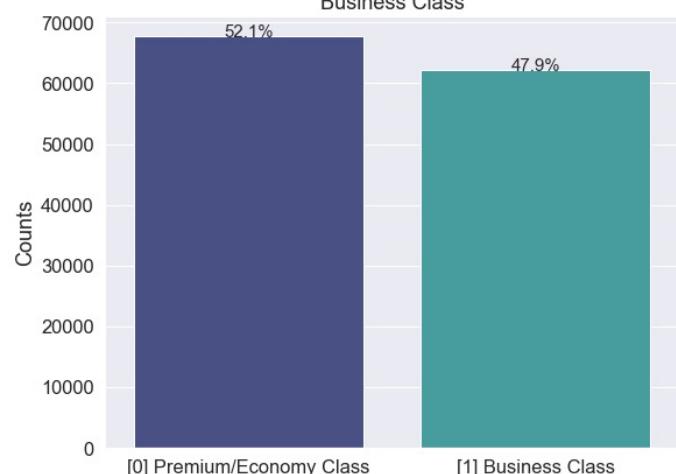
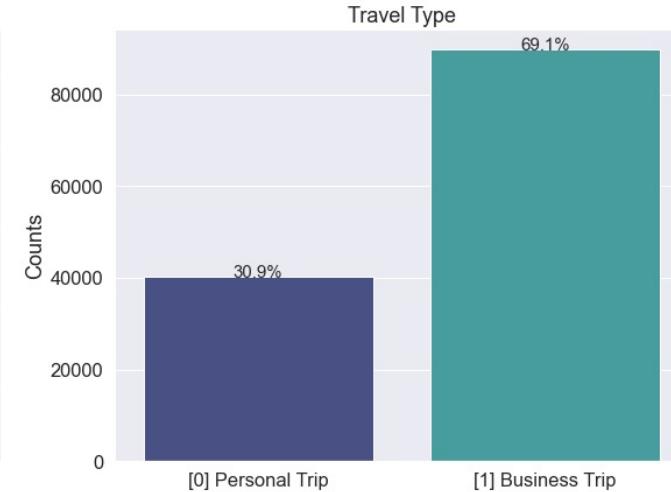
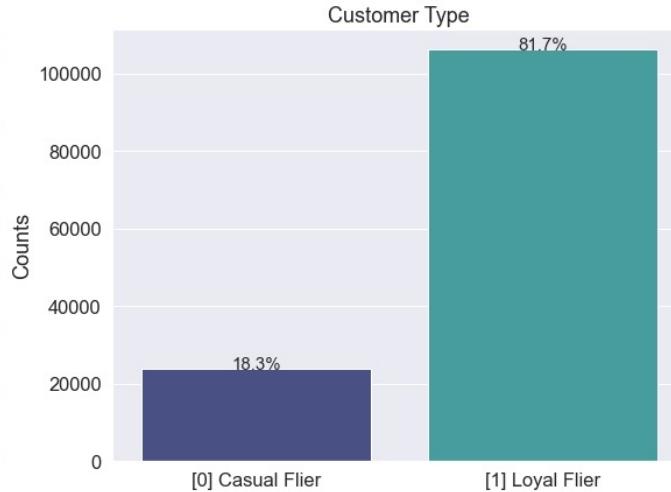
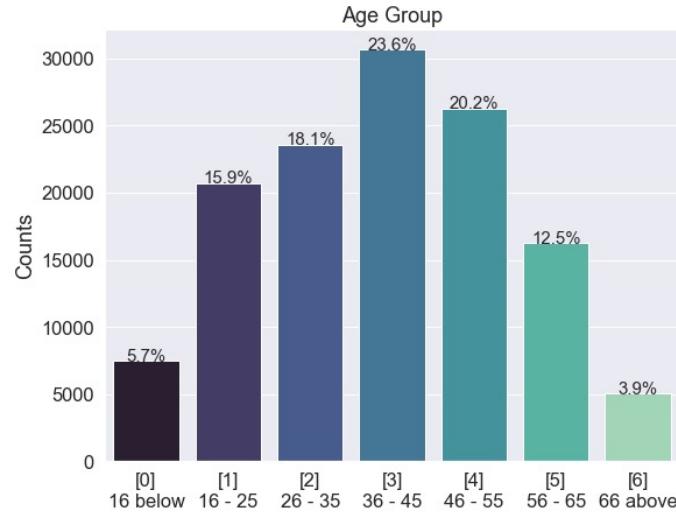
DROPPED COLUMNS

- Flight Distance
- Arrival Delay in Minutes
- Departure Delay in Minutes
- Age
- class_Eco Plus
- Arrival Delay Group
- Departure Delay Group
- Departure/Arrival time convenience
- Gender
- Gate location

FEATURE SELECTION



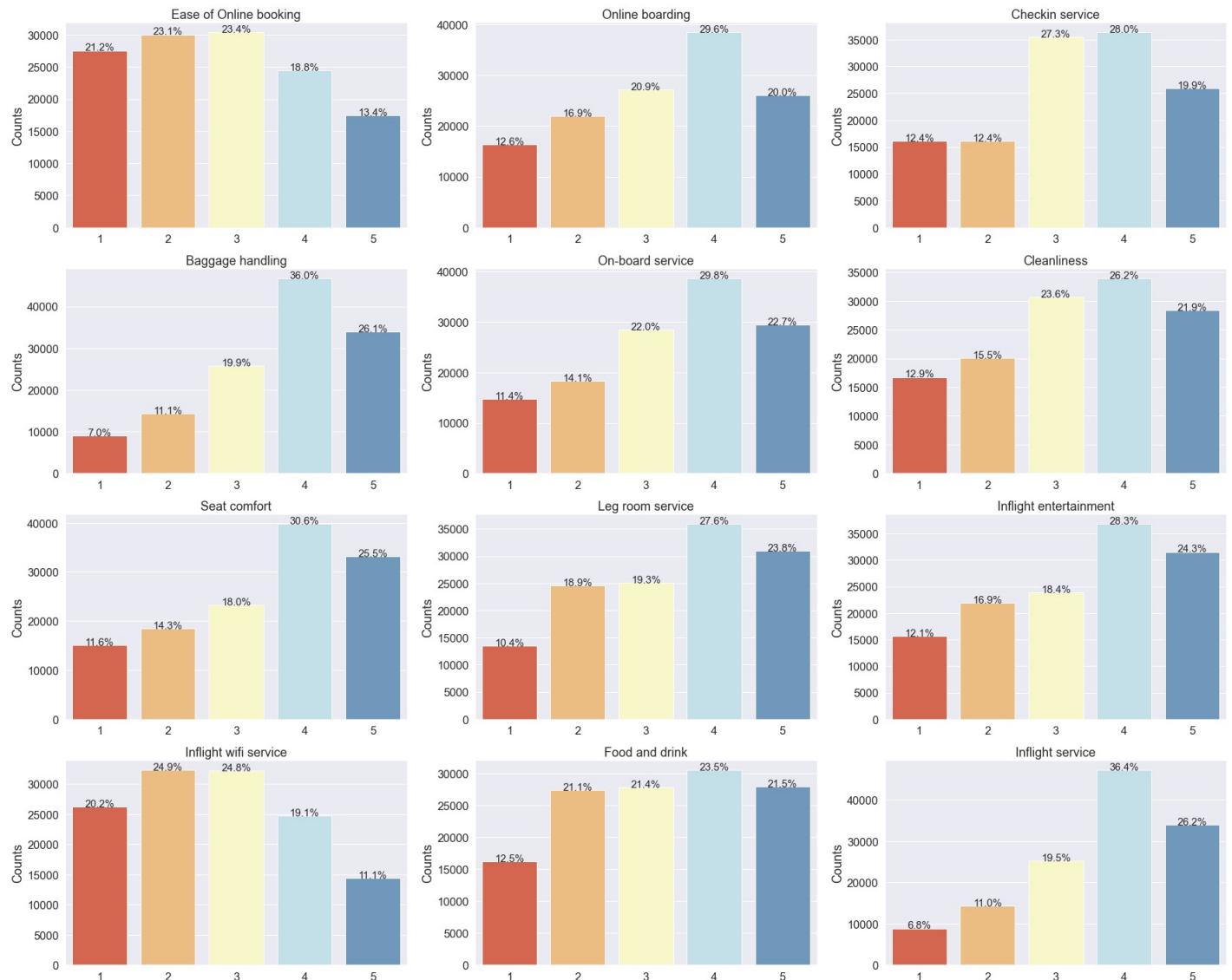
DEMOGRAPHICS DISTRIBUTION



RATED COLUMNS DISTRIBUTION

QUICK INSIGHTS

- **Lowest Ratings:**
 - Ease of Online Booking
- **Highest Ratings:**
 - Baggage Handling





©arlansantos

DATA . EDA . **CLASSIFICATION** . CONCLUSION

CLASSIFICATION MODELS

Logistic
Regressor

SVC

Gaussian
NB

KNN
Classifier

Decision
Tree

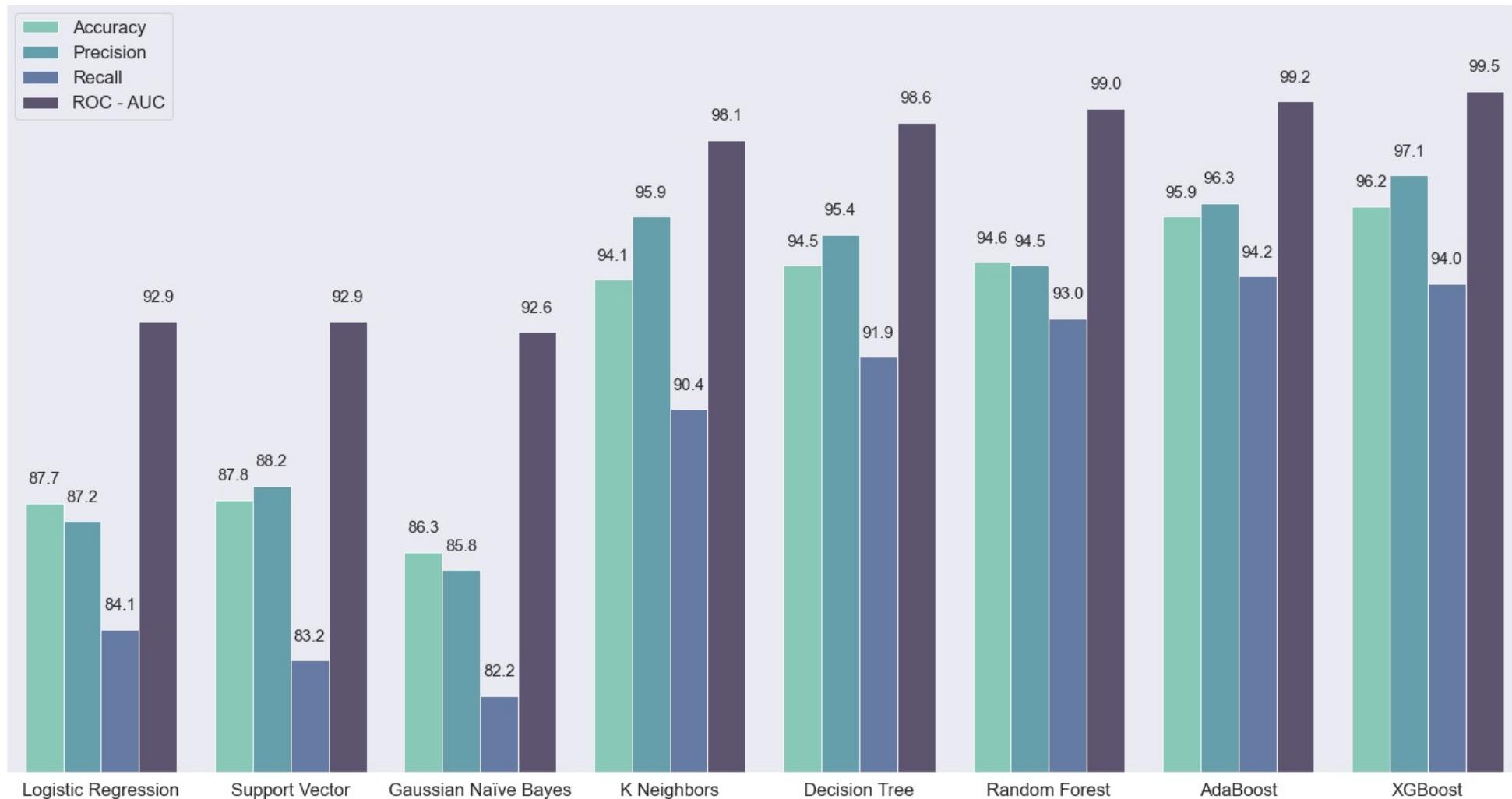
Random
Forest

AdaBoost

XGBoost

MODEL EVALUATION

Classification Model
Metric Comparison



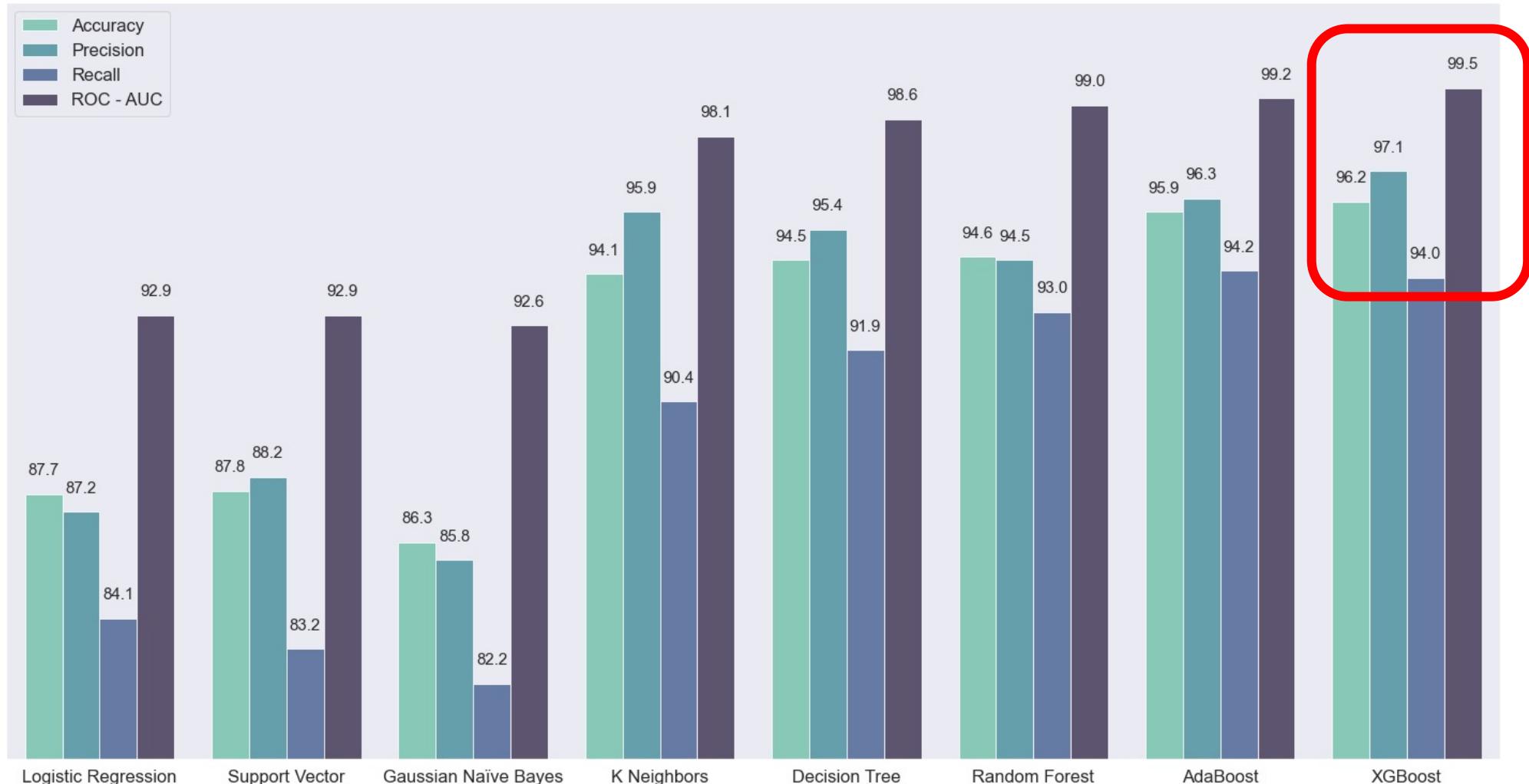
MODEL EVALUATION

WHICH METRIC/s TO CONSIDER?

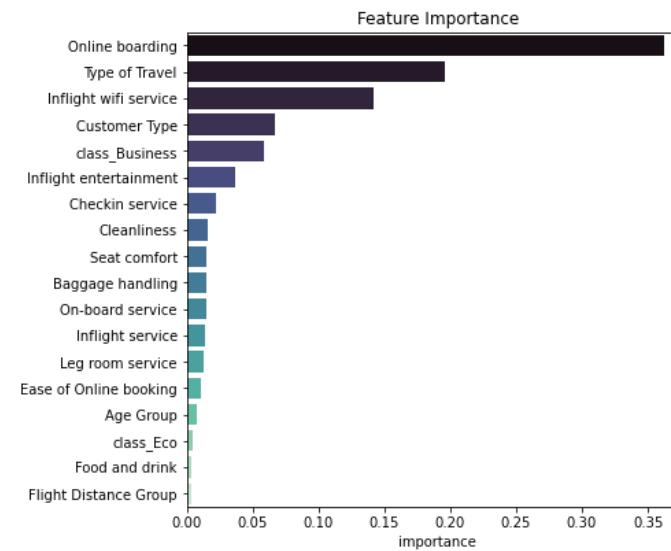
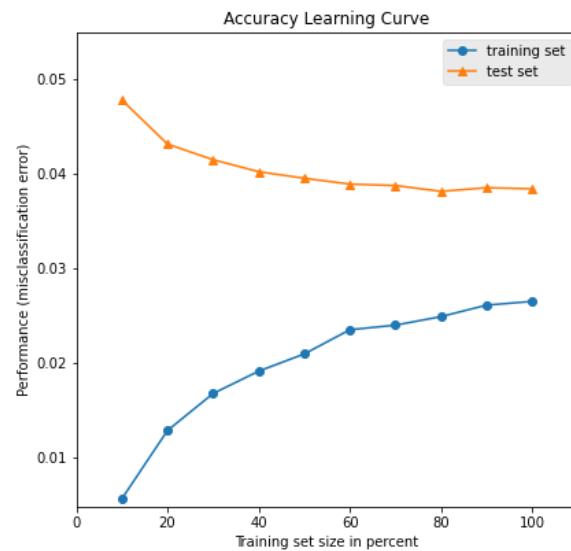
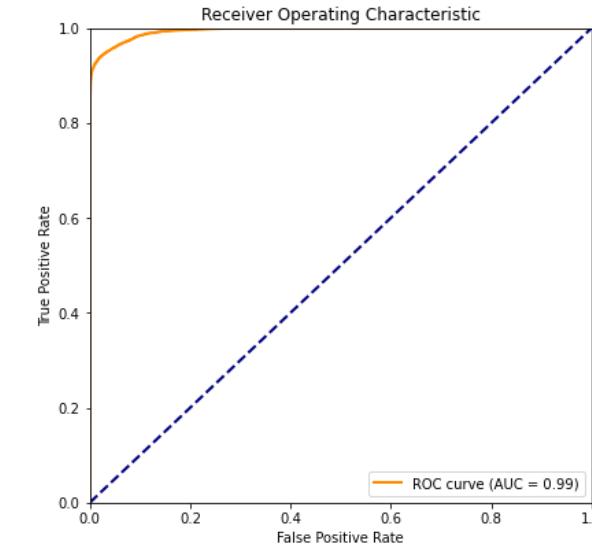
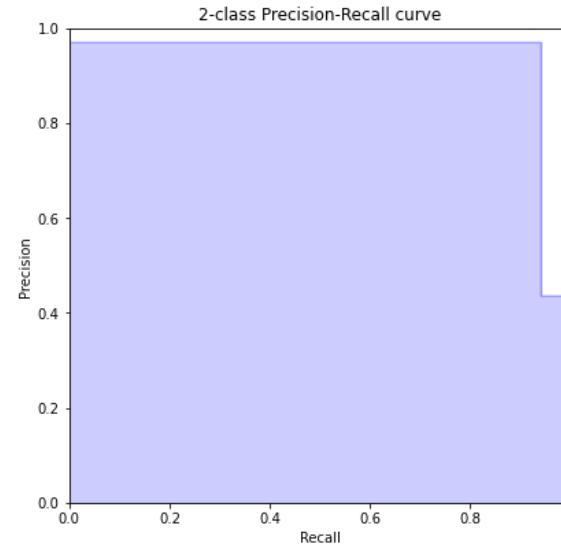
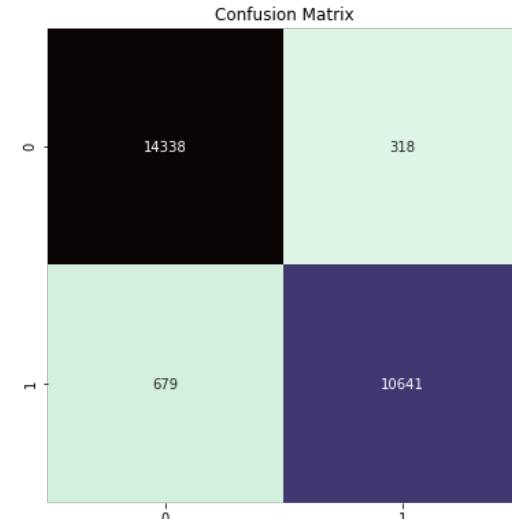
- Be able to capture both sentiments, satisfied and not satisfied/neutral - to know where you're good at and where there's room for improvement
- Maybe **Accuracy** is a good metric to evaluate?
- However it is critical not to misclassify a disgruntled passenger as someone who had an enjoyable flight – lead to loss of business
- Reduce our **False Positives** as much as possible:
Precision is the main metric of choice.

MODEL SELECTION - XGBoost

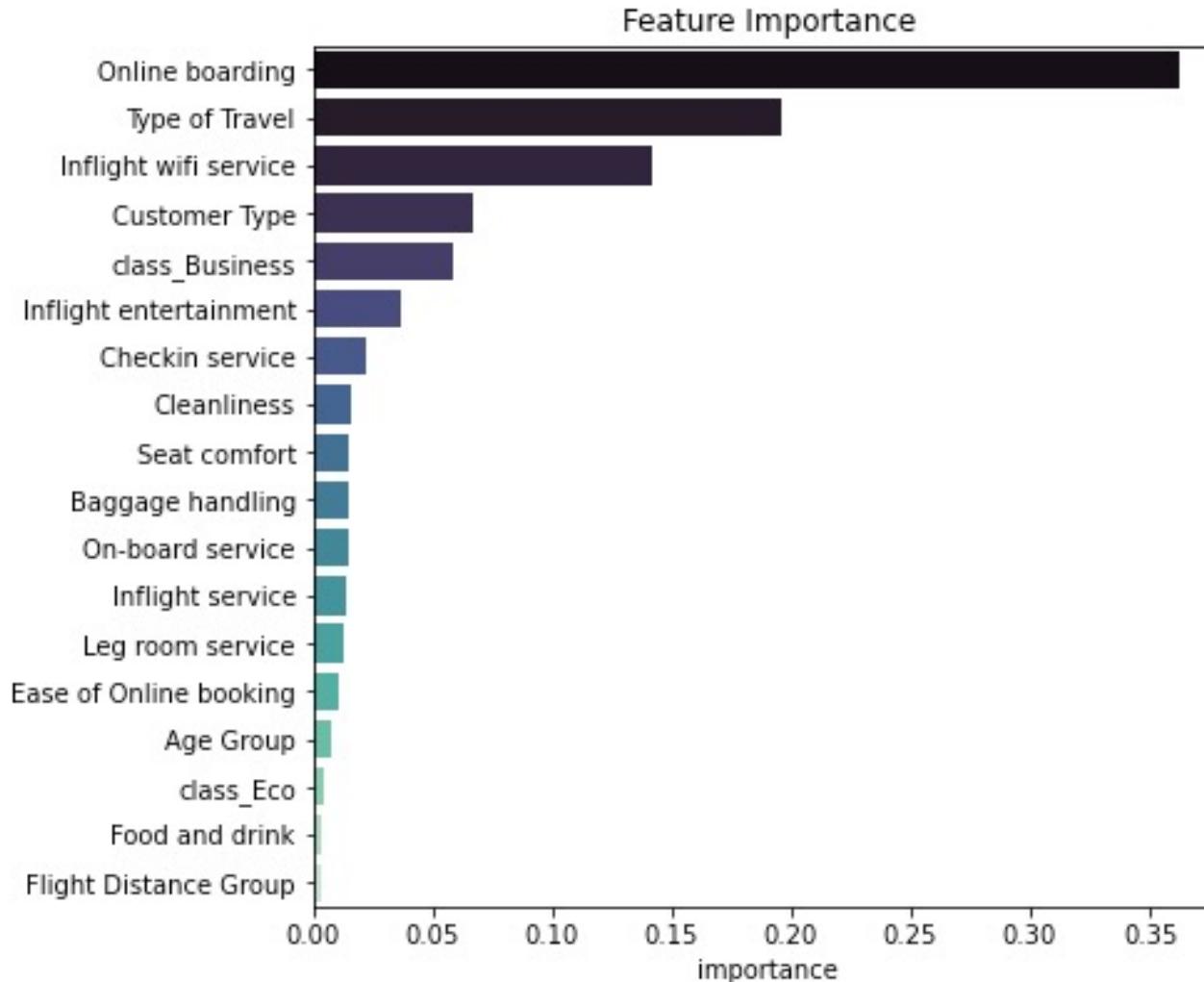
Classification Model
Metric Comparison



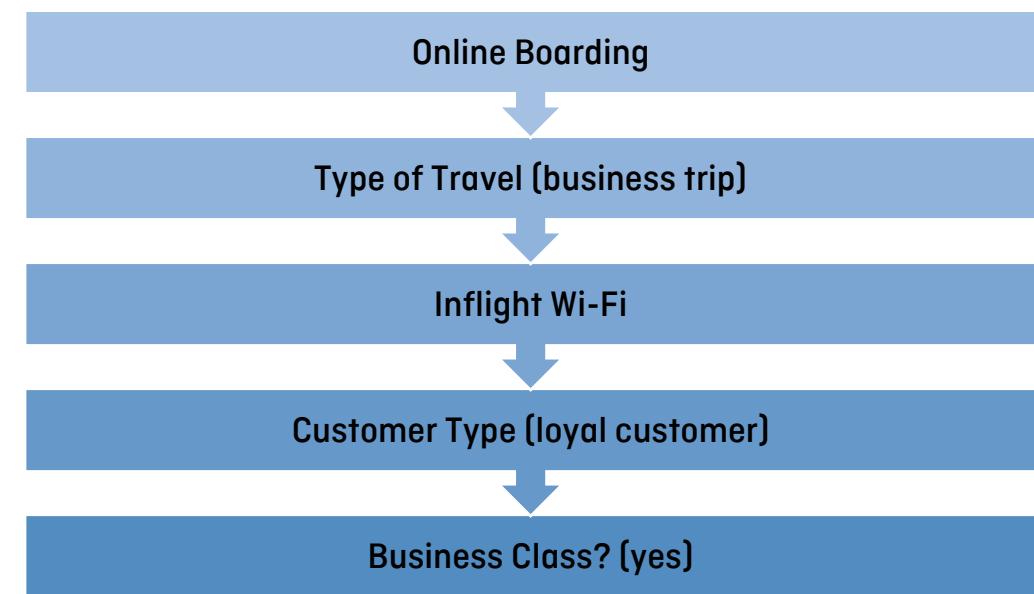
XGBOOST CLASSIFIER PERFORMANCE



FEATURE IMPORTANCE – TOP 5



Which **flight scenarios** and **key areas of services**, both soft and hard products, should an airline allocate its budget and resources in order to ensure passenger satisfaction and retention?



A FOCUS ON LOYAL MEMBERS

Prepare Data Set

- Customer Type is filtered to include Loyal (Frequent Flier) customers only

Perform EDA

- Verify if the target variable distribution is balanced: 52.2% vs 47.8% -> [0] vs [1]
- Select appropriate feature

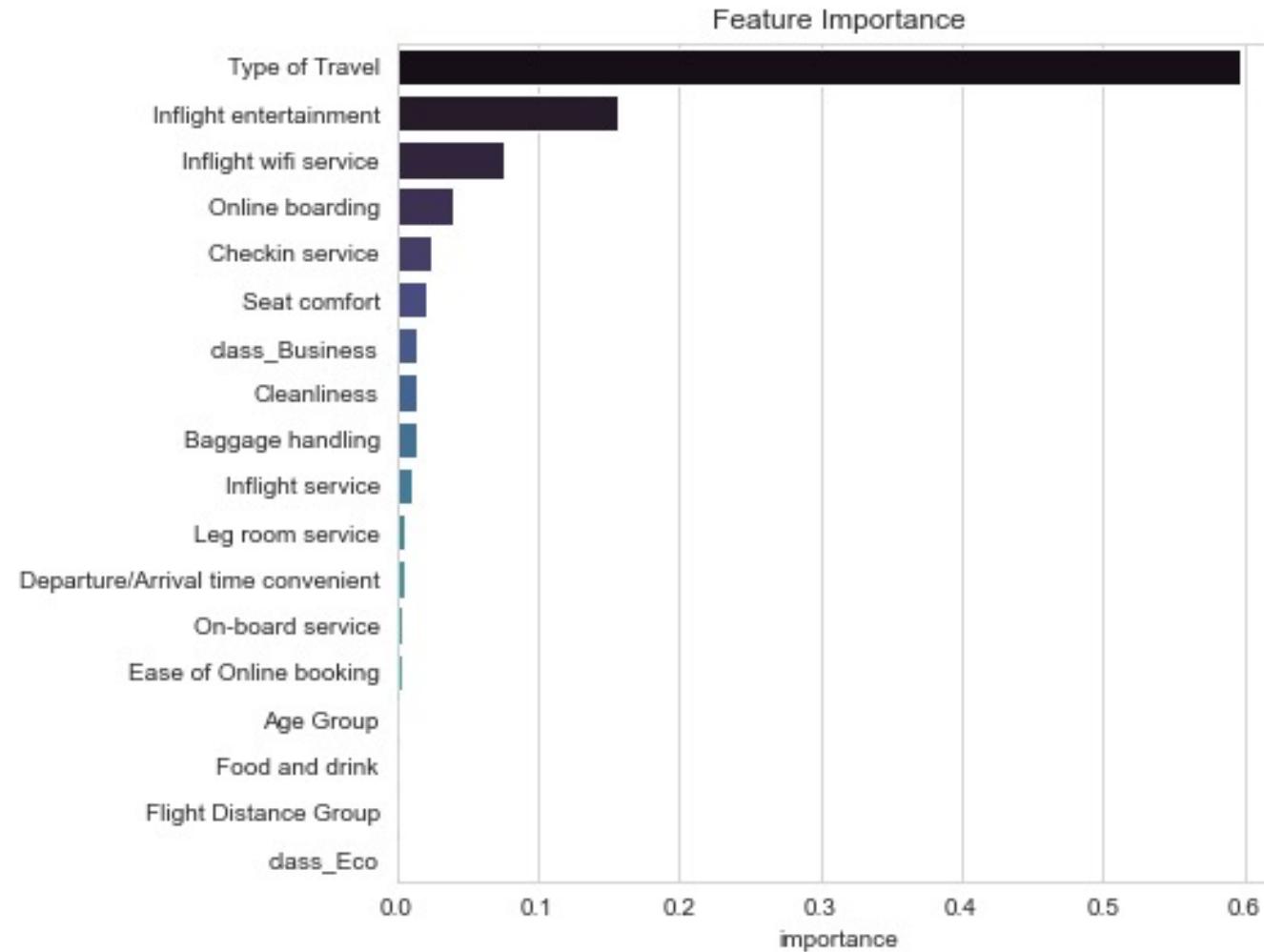
Perform Ensemble Modeling

- Same type of ensemble method used - XGBoost
- BUT... different parameters (`gridsearchCV`)

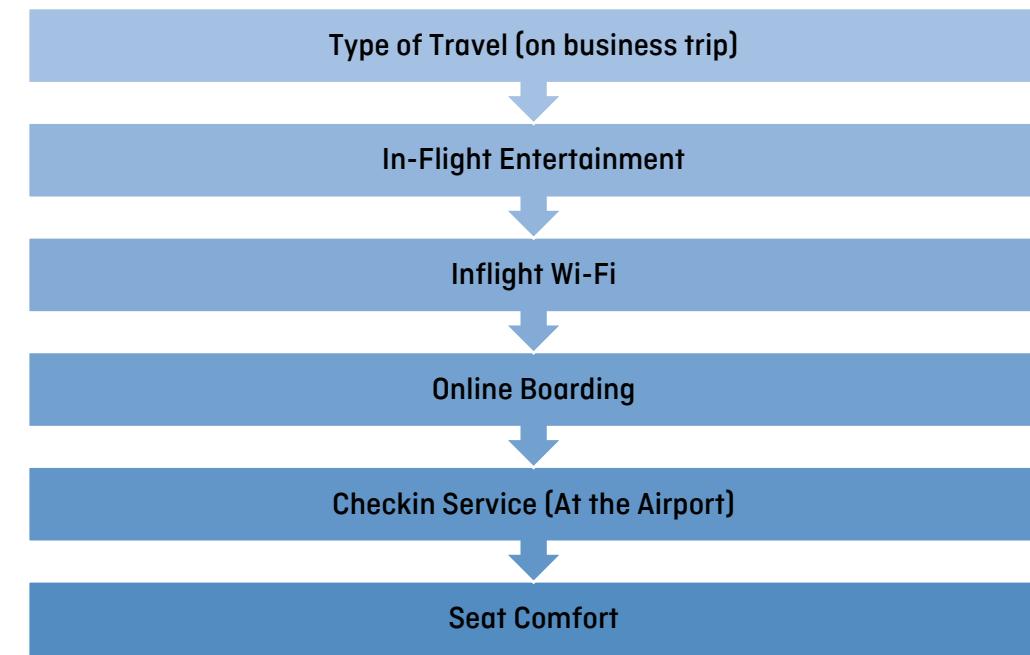
Evaluate Model

- Precision Score: **98.9**
- Feature Importance

A FOCUS ON LOYAL MEMBERS

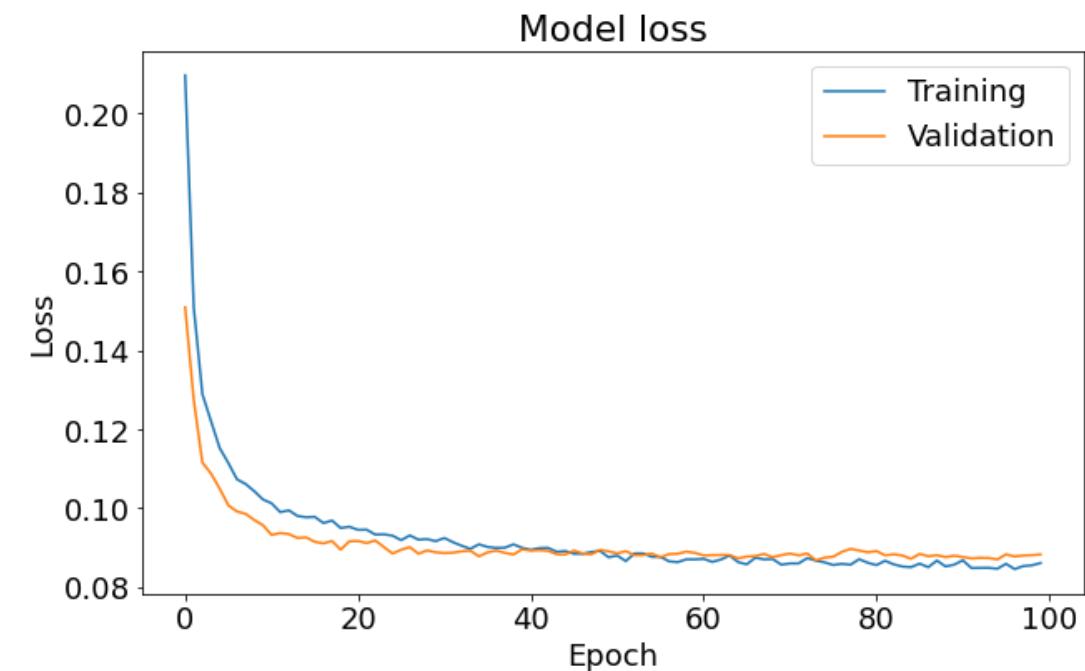
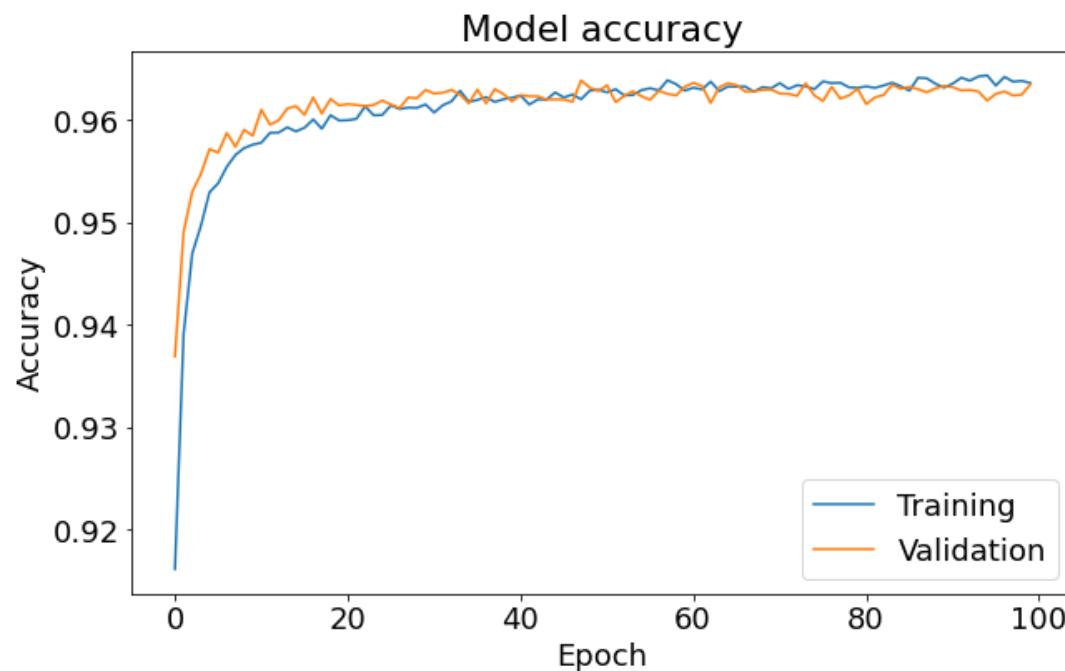


To achieve the goal of returning to double-digit profitability, what are the **top important factors** that guarantees an enjoyable flight for the airline's **most loyal customers**?



Neural Network Training Report

- Sequential with 2 hidden layers
- Implemented **Dropouts** in the hidden layers as the initial network was suffering **overfitting**
- Test loss: 88.35%, Test accuracy: 96.26%





©Carlonsantos

DATA . EDA . CLASSIFICATION . CONCLUSION

CONCLUSION

1. The airline can focus on guaranteeing the satisfaction of **a particular set of customers** in order to build a strong brand image, maintain their loyalty, and achieve profitability:

Loyal customers on a business trip in the business cabin – PREMIUM section

2. Focus areas to properly allocate the airline's budget and resources, particularly:

Check-in Technology (Online and Airport)

Cabin Connectivity

Business Class Seat

In-Flight Entertainment

3. Given a collection of feedback from passengers, how can the airline **carefully assess the flight experience** as something positive or not?

By deploying a machine learning model with a focus on HIGH Precision

NEXT STEPS

1. Create a dashboard on Tableau to visualise the overall and granular picture of passenger's satisfaction better.
2. Perform analysis on other types of customers (eg. Non-frequent fliers) to explore the potential of creating new customer base and increasing the airline's market reach.
3. Create an app to get feedback from customers and integrate the response to the trained model, and classify the passenger's sentiment. Simulate the process of full data modeling lifecycle.

References

- [1] <https://www.icao.int/sustainability/Pages/Economic-Impacts-of-COVID-19.aspx>
- [2] https://investor.qantas.com/FormBuilder/_Resource/_module/doLLG5ufYkCyEPjF1tpgyw/file/annual-reports/2021-Annual-Report.pdf
- [3] <https://www.qantasnewsroom.com.au/media-releases/qantas-group-posts-significant-loss-from-full-year-of-covid/>
- [4] <https://www.qantas.com/au/en/qantas-group/delivering-today/performance-against-our-strategic-pillars.html>
- [5] <https://www.kaggle.com/teejmahal20/airline-passenger-satisfaction>

*Note: All aviation images used are personal photos of the author

A large white Qantas Boeing 747 aircraft is positioned in the foreground, facing towards the right. The tail features the iconic red kangaroo logo. The fuselage is white with "Spirit of Australia" and "QANTAS" written on it, along with "Longreach". In the background, the dense urban skyline of Sydney is visible, featuring numerous skyscrapers and the distinctive spire of the Sydney Tower Eye. The sky is clear and blue.

**THANK
YOU**