Increasing Airline Passenger Satisfaction

Classification Module Project

Abidemi Adeboje, January 2022

Introduction

- QUESTION: How can **airlines** increase satisfaction among passengers during the pandemic?

- OBJECTIVE: Can we build a classification model from an airline passenger survey with several features (i.e. *gate location, food and drink, online boarding)* in order **identify a satisfied airline passenger**?

Data

- Kaggle dataset with over **100k unique records** of an **airline passenger satisfaction survey** with over 20 features was used.

Algorithms

- Data was split into training and validation sets. A test set was included.

- Sklearn Python Library Models
  - Decision Tree
  - Logistic Regression
  - Random Forest

- Seaborn library for data visualization

Results

- **ROC curve** was used to rate models with a high true positive rate (satisfied passenger) and low false positive rate (dissatisfied passenger).

- **Random Forest** model was the best performing and selected for tuning.
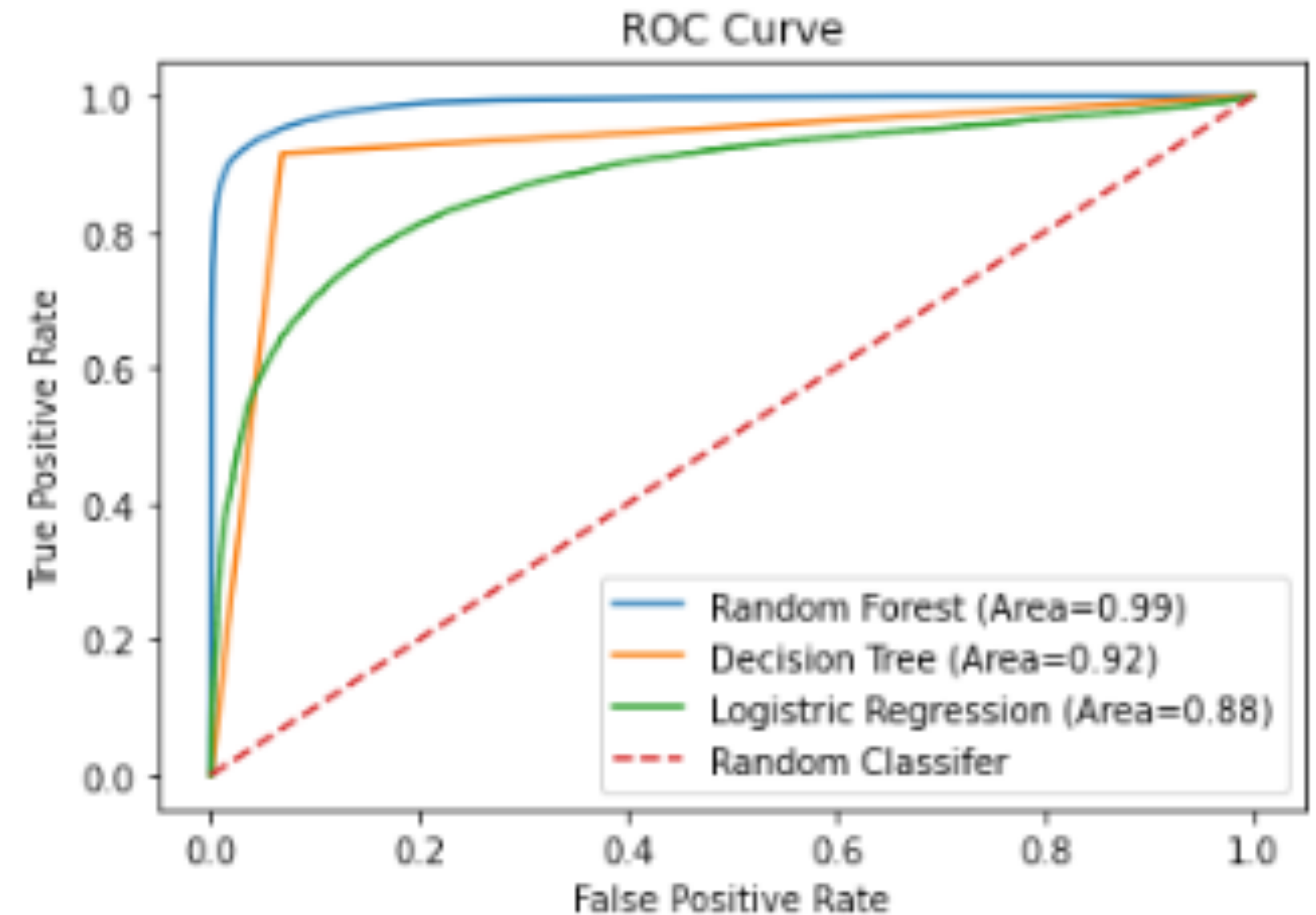
Results

- **Hyperparameters:** max_depth (level of the tree) and n_estimators (number of trees)

- **RMSE and ROC AUC Score** were the metrics used to evaluate the models.

- **Computational cost** was also used to select the final random forest model.

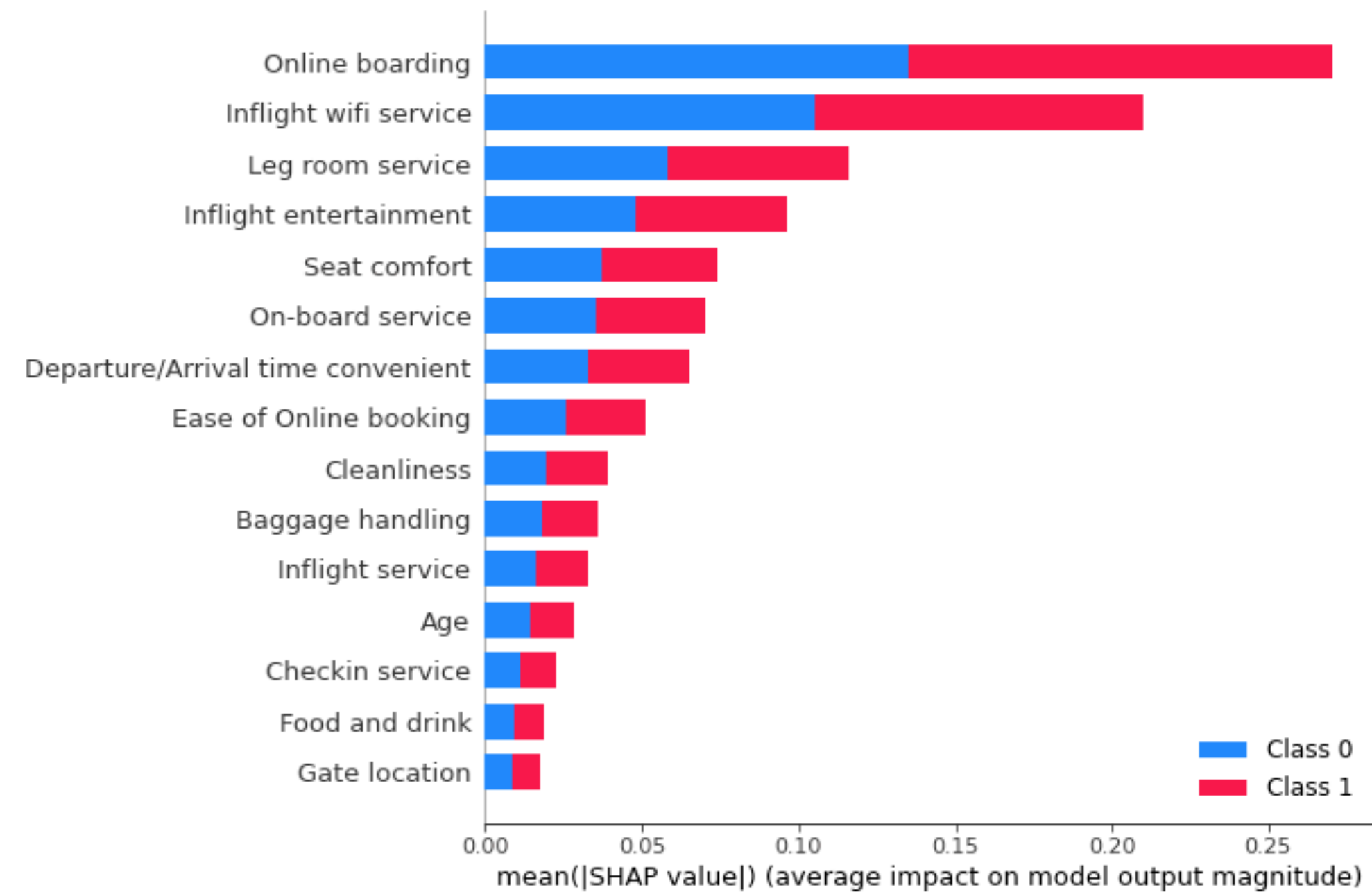 - **Random Forest** model with max_depth = 9 and n_estimators= 100.
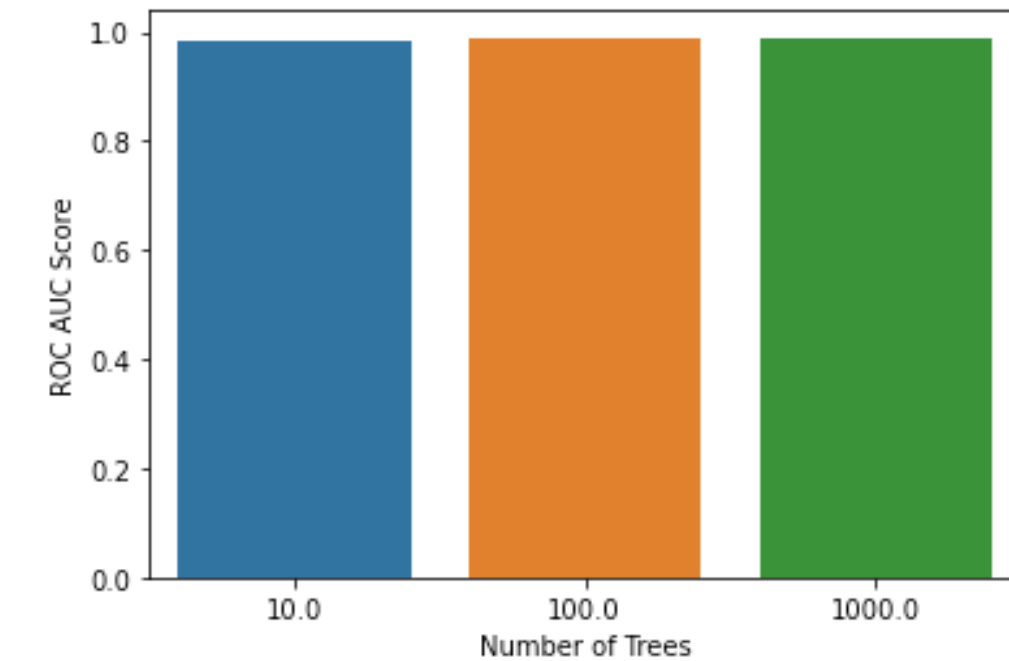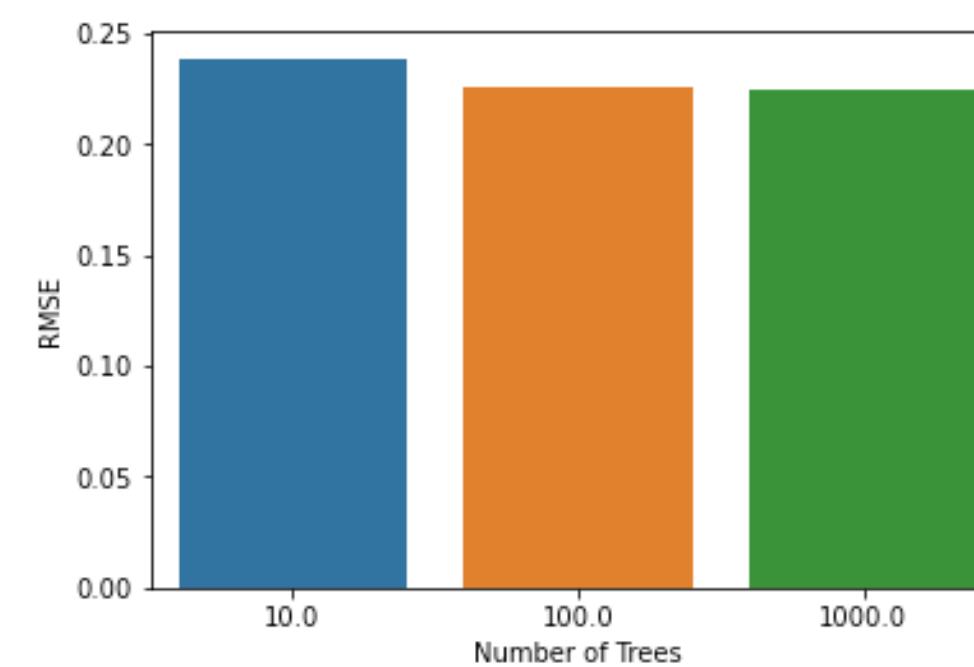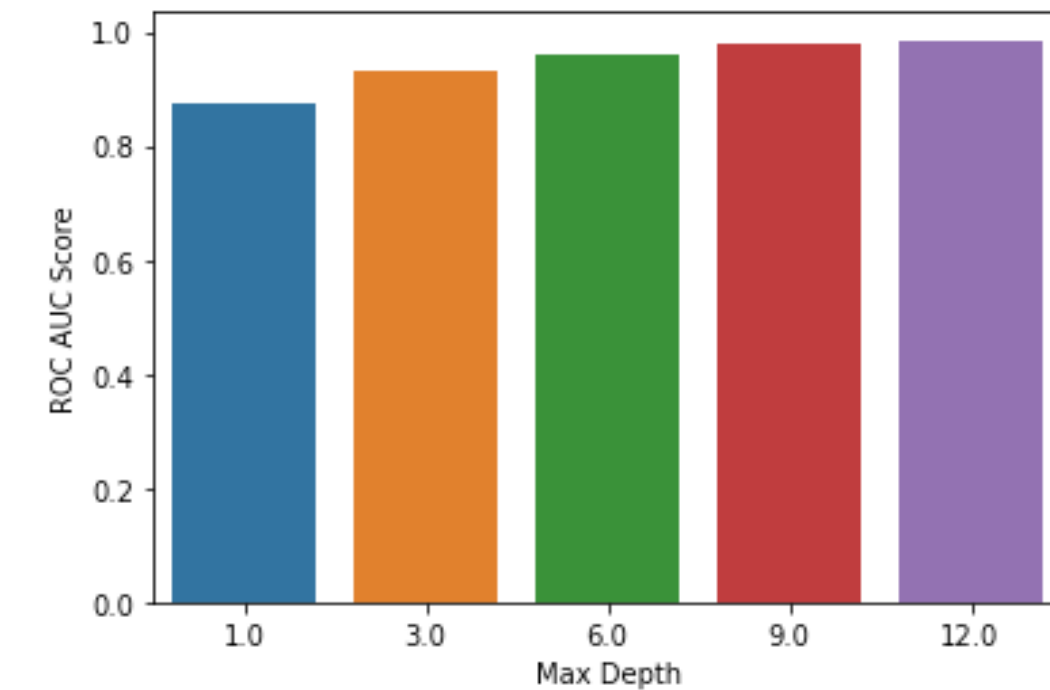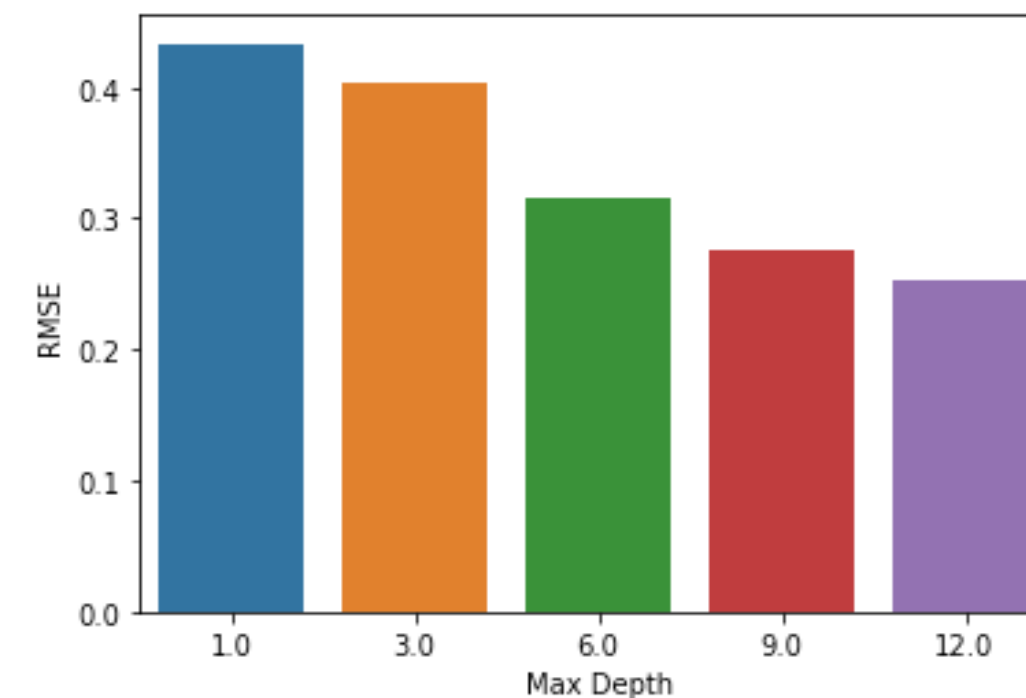
Results

- **SHAP Values** used to interpret importance of feature to model.

- **Top 3 :**
  - Online Boarding
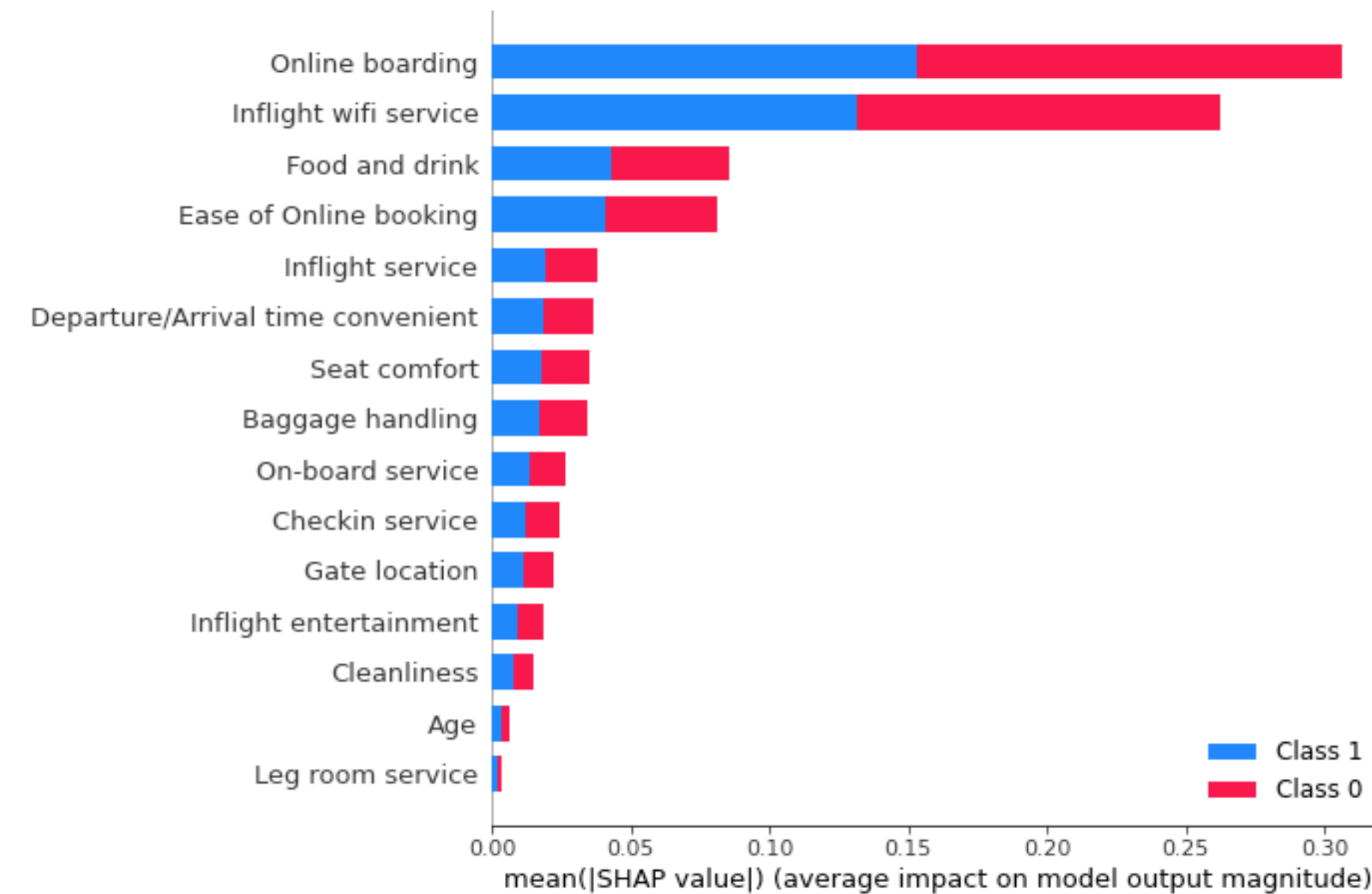 - In-flight Wi-Fi
  - Leg Room

Results

- **Subset - Age 20-29** Passengers

- **Top 3**:
  - Online Boarding
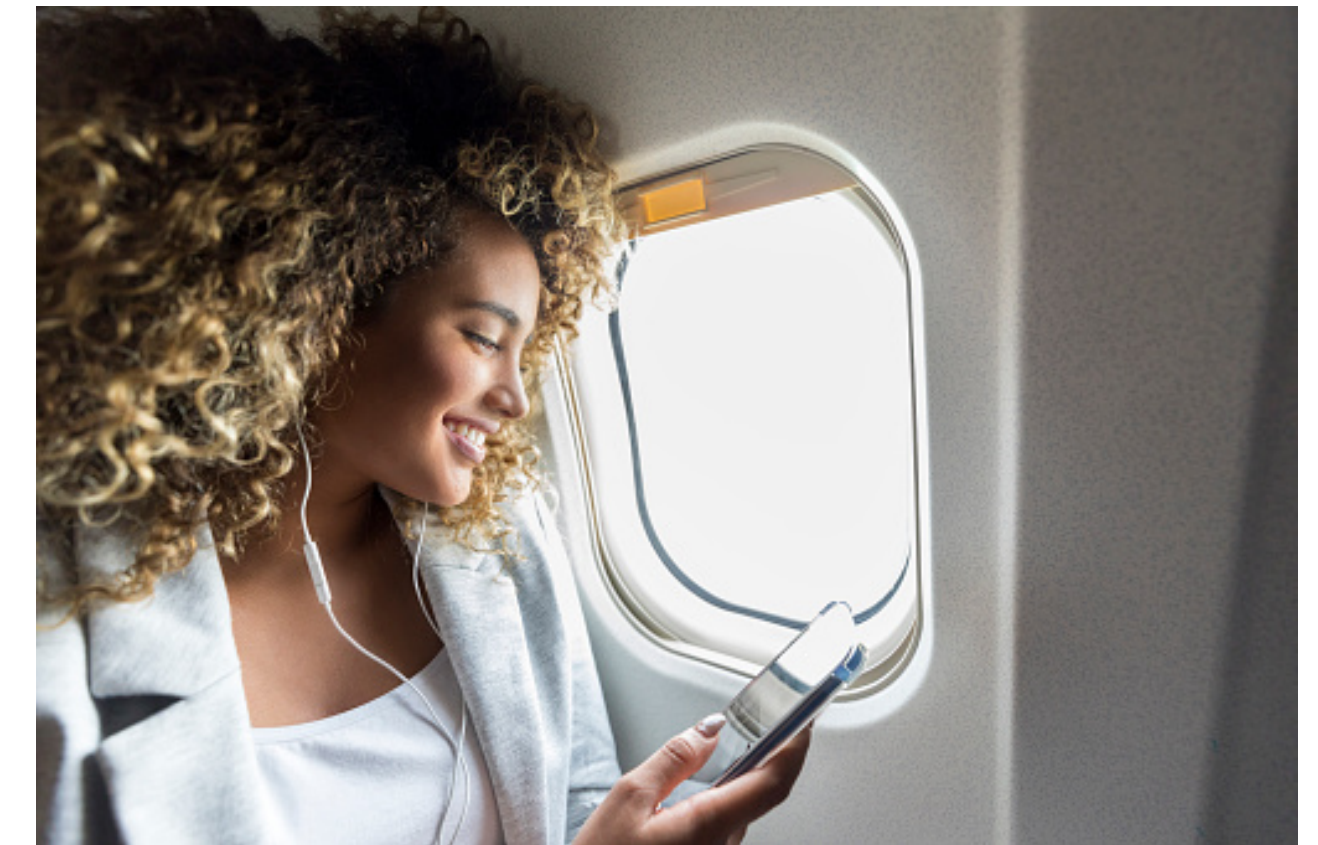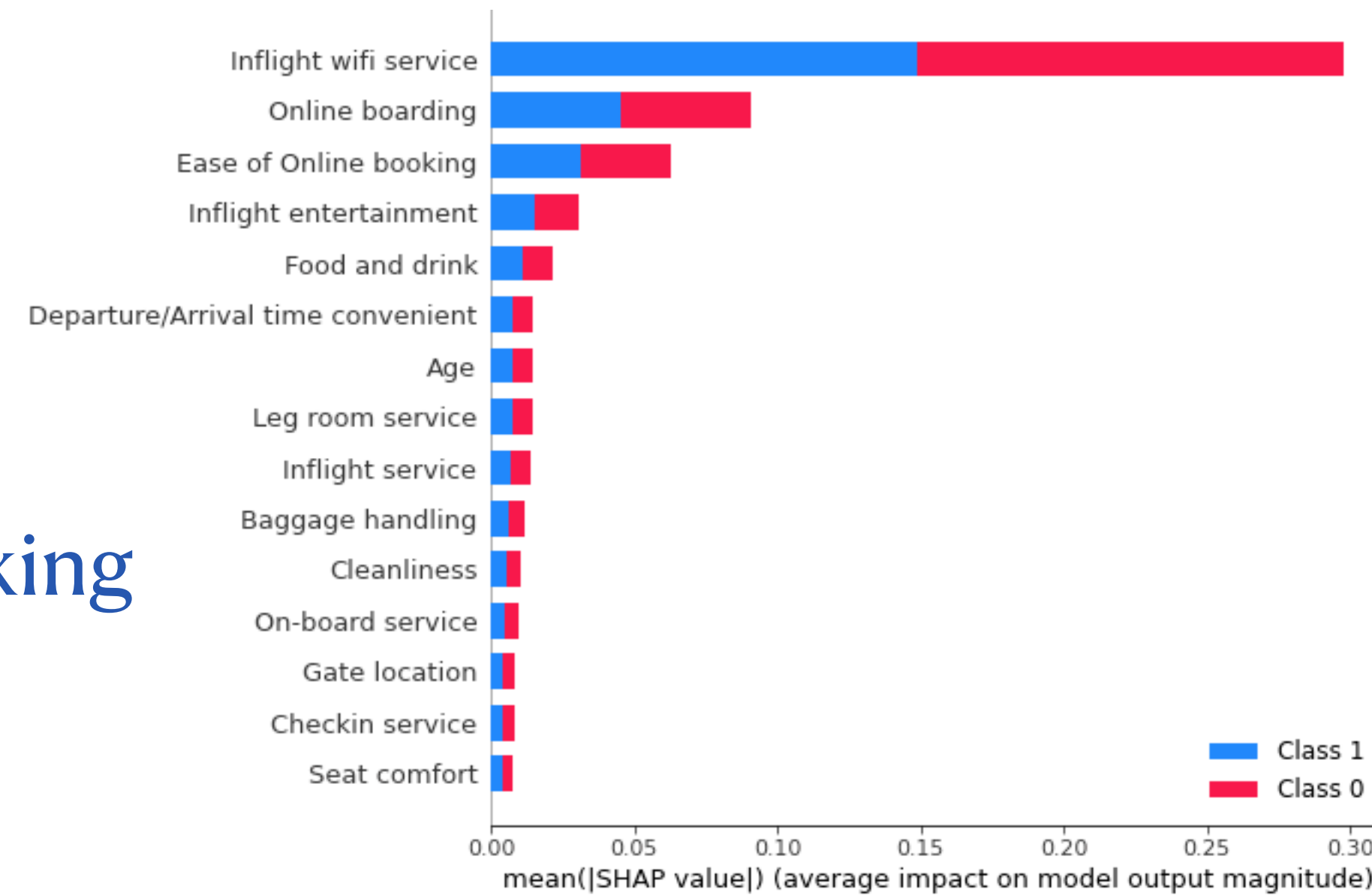 - In-flight Wi-Fi
  - Food and Drink

Results

- **Subset - Economy Class** Passengers

- **Top 3** :
  - In-flight Wi-Fi
  - Online Boarding
  - Ease of Online Booking

Conclusions

- Online Boarding and In-flight Wi-Fi were among the top 3 features impacting the model of all passenger groups studied.

- Airlines can study how to improve these features in order to increase passenger satisfaction.

Appendix

- Slideshow pictures from https://unsplash.com