# PREDICTING HOTEL BOOKING CANCELLATION

Waad Alotaibi | DataScience Bootcamp

# Goal

- A significant number of hotel bookings are called-off due to cancellations or no-shows.

- The goal of this project is to use classification models to predict the hotel booking cancellation and help in formulating profitable policies for cancellations and refunds.
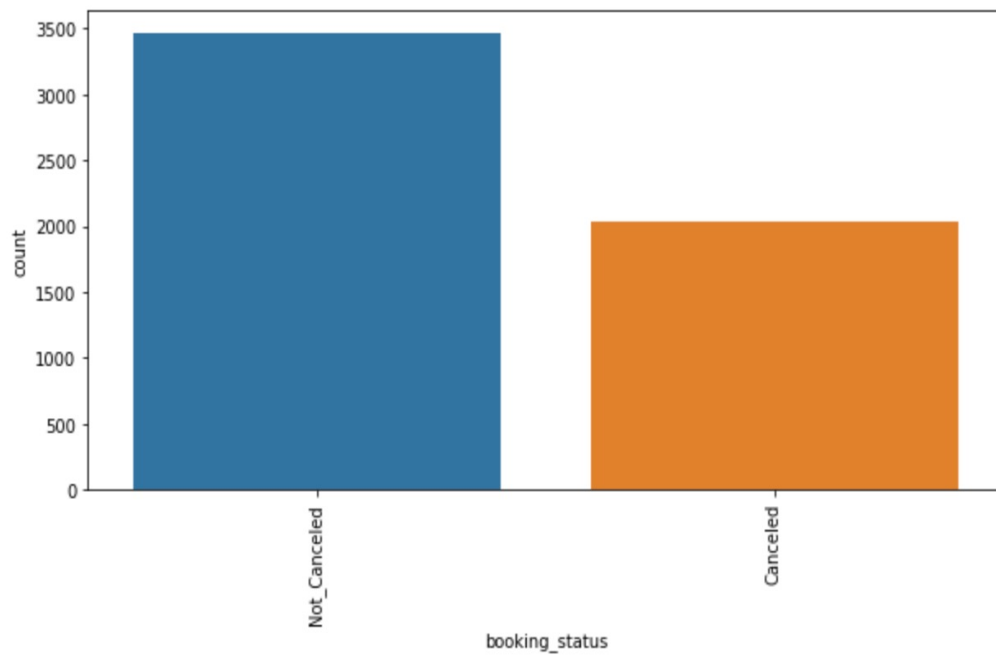
# Dataset

'Customers Booking Details' dataset. It used for supervised binary classification tasks.

Total number of features in the dataset is 5499 distributing 18 attributes as some features input to the model and single attribute as output label.
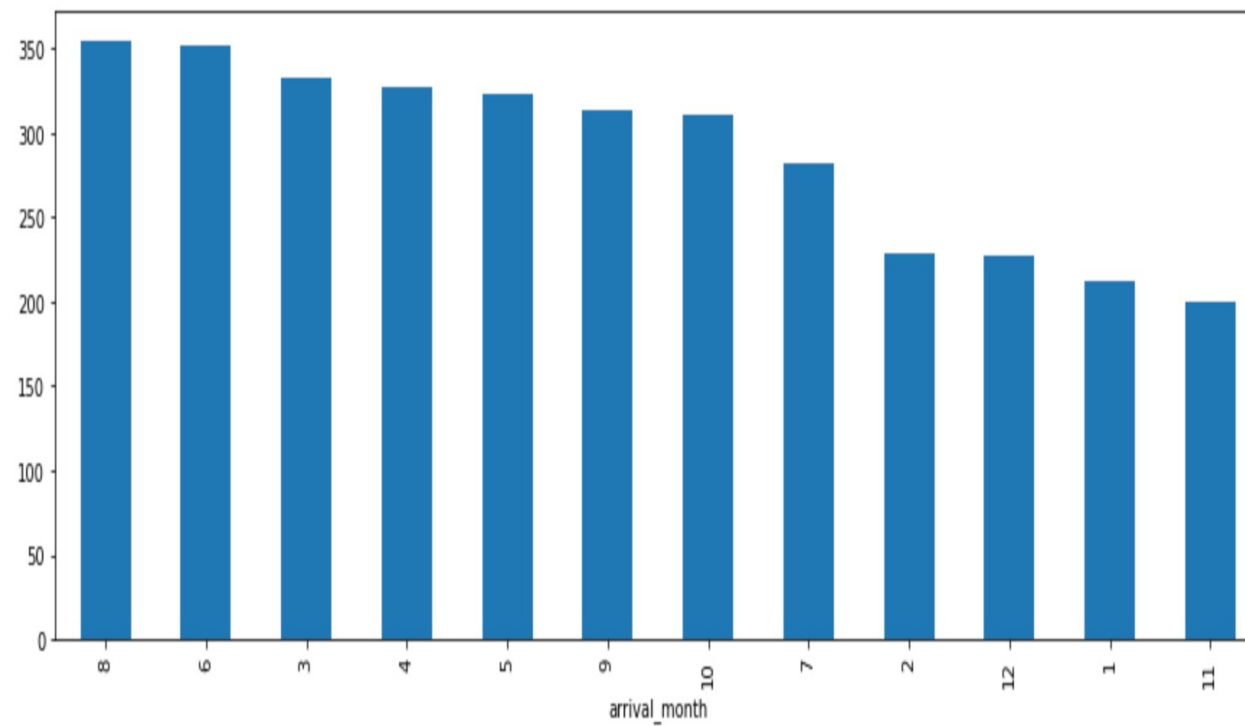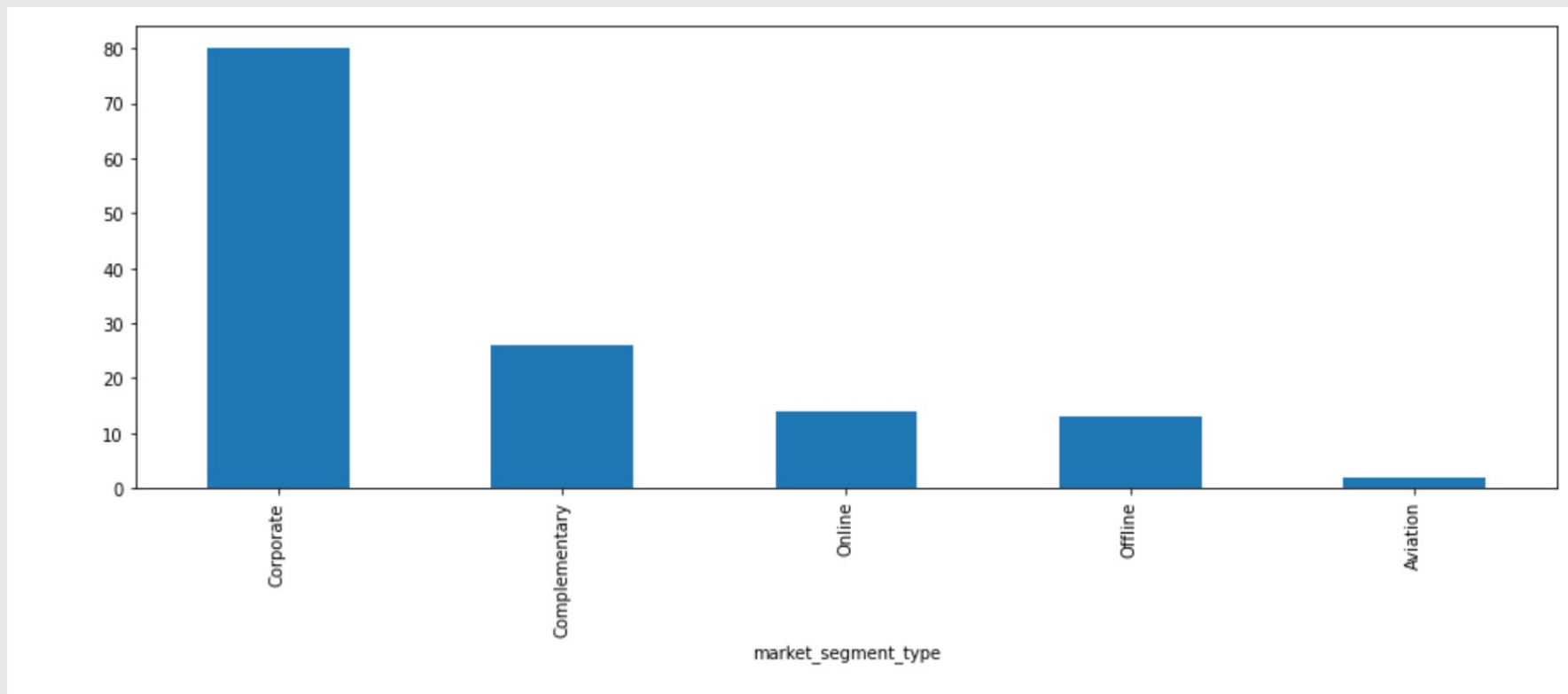
# EDA

## MARKET SEGMENT MOST OF THE GUESTS COME FROM

# MODELS AND EVALUATION

- After I checked the insgihts of dataset by perform the Exploratory Data Analysis (EDA).

- I performed the preprocessing to get the most relevant features for prediction.

- Then, I checked  the prediction through four different classification models:

- **Random Forest, Decision Tree, Support Vector Machine (SVC) and k-nearest neighbors(KNN).**

- Then compare all the models through different performance metrics : **accuracy, precision, recall and f1-score.**

# RESULTS

This table shows the performance of the four machine learning models we used on our dataset.

```
+--------------+----------------+-------------------------+----------------------------+-----------------+
| Metrics      | decision tree  |  k-nearest neighbours   |  support vector machine    |  random forest  |
|--------------+----------------+-------------------------+----------------------------+-----------------|
| accuracy     |     0.81       |         0.817273        |          0.828182          |     0.847273    |
| precision    |   0.863235     |         0.853521        |          0.858939          |     0.881598    |
| recall       |   0.834993     |         0.86202         |          0.874822          |     0.87909     |
| f1-score     |   0.848879     |         0.857749        |          0.866808          |     0.880342    |
+--------------+----------------+-------------------------+----------------------------+-----------------+
```
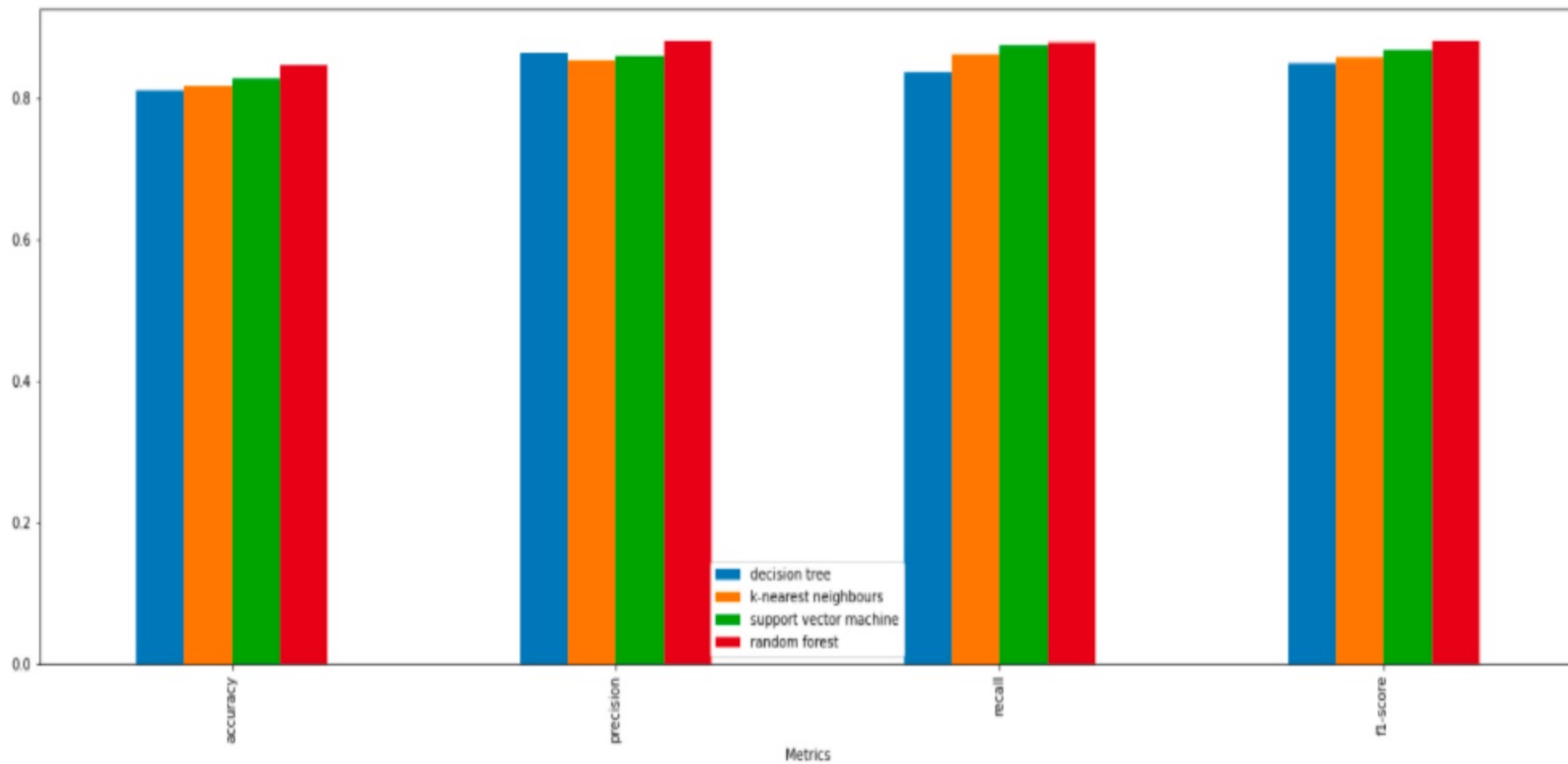
# CONCLUSION

It shows that **Random Forest** attains the highest Accuracy with 0.84% on testing data and high other metrics as compared to other Models. For this reason, we will use Random Forest model as our predictive model.

# THANK YOU