

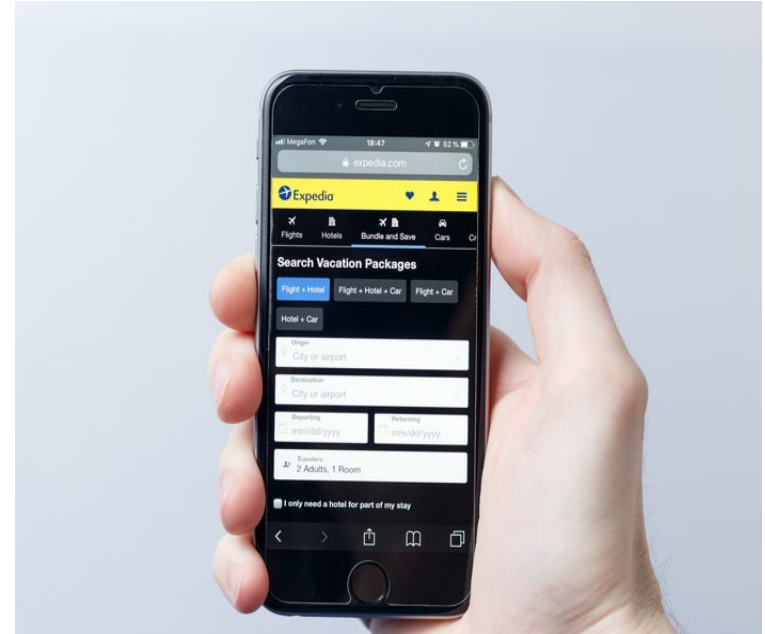
# Expedia Hotel Recommendations

By Lais Amorim Menezes



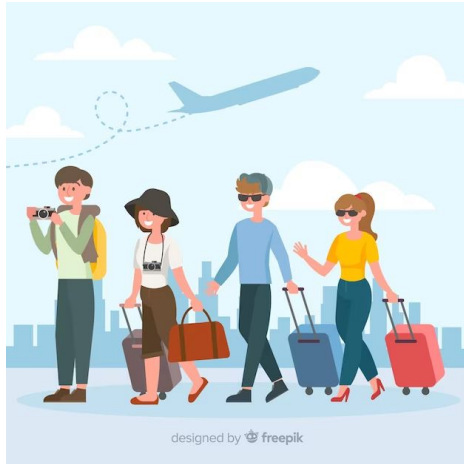
# Problem statement

- The website displays random Hotels to all the customer.
- Customer spend too much time on the website trying to find a hotel that matches their preferences.
- Expedia loses customer to other websites that are quick and easier to navigate.



# Impact

Customer

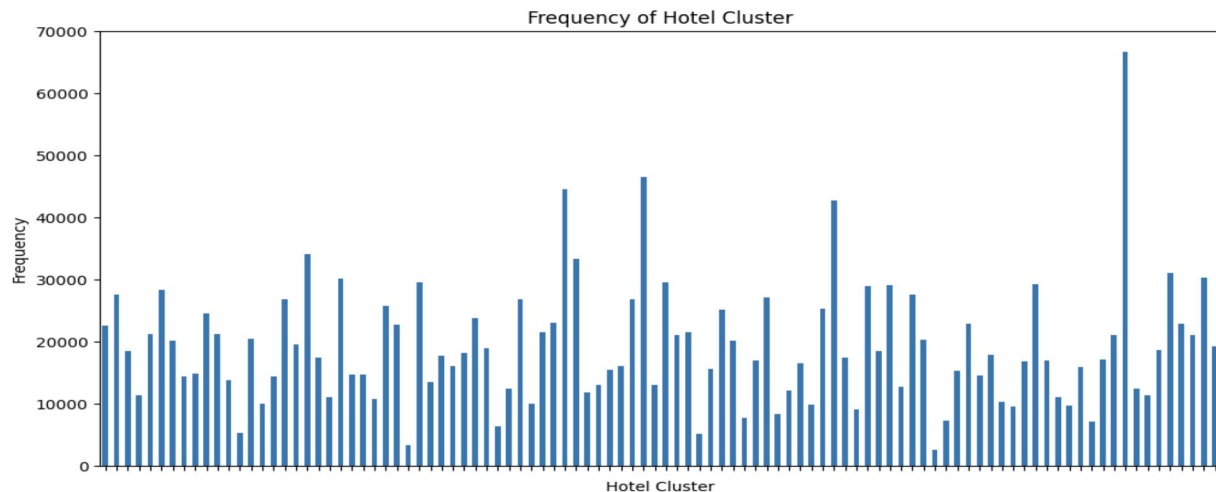


Company



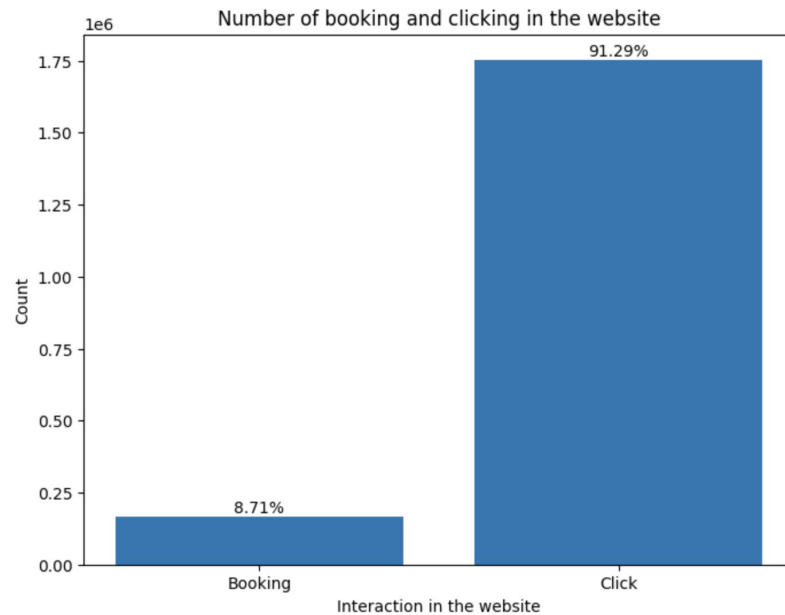
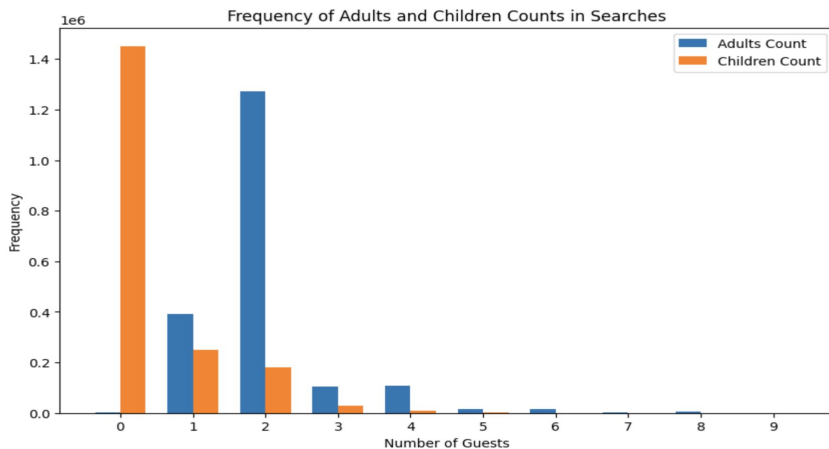
# Dataset overview

- The train data from 2013 to July 2014.
- The test data from August to December 2014.
- The dataset doesn't contain categorical columns, instead, all columns are stored numerically.
- The target column will be hotel cluster, the distribution is multimodal with several peaks.

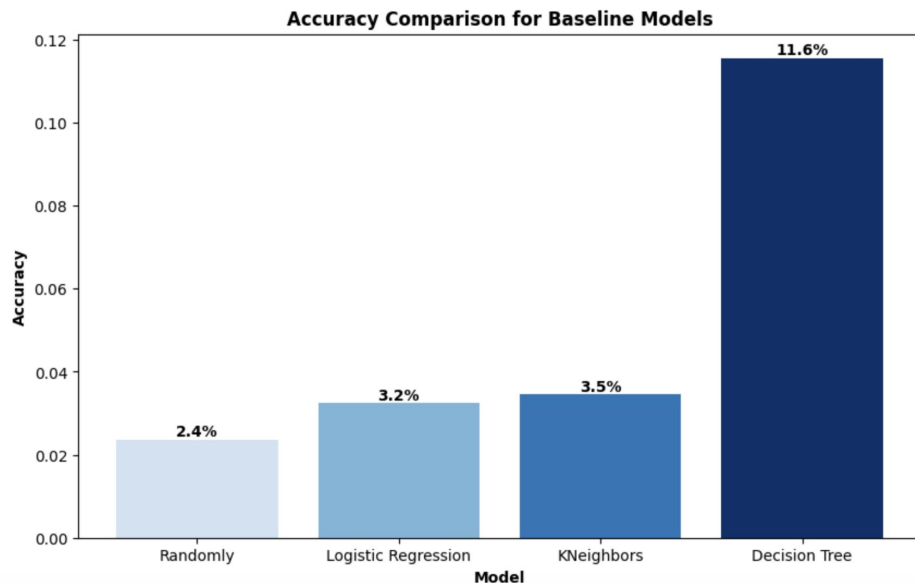


# Data overview

- Most of the customers book a hotel for 1 day and for 2 persons without childrens.
- 91.29% of the data is just customers chicking.



# Baseline Models and Evaluation Metrics



- Randomly
- Logistic Regression
- KNeighbors
- Decision Tree (The best one)

# Next Steps

- Implement advanced models
- Fine-tune hyperparameters
- Improve the model performance





Thank you!