

NORTHEASTERN UNIVERSITY



INFO 6105 - Spring 2019

Data Science Engineering Methods

Topic : Airbnb

Under the guidance of

Prof. Ramkumar Hariharan

TA , Rohan Mandhanya

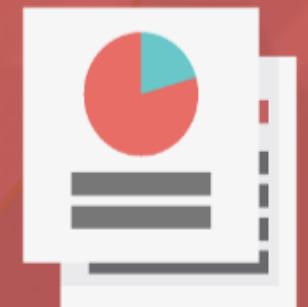
Team: FOURce

- **Huixiang Liu**
- **Meng Zhou**
- **Soumya M**
- **Sneha Shree**

kaggle Destination Prediction for New User Bookings

The project is about the analysis of the Airbnb dataset and the model we built to do the prediction of New user's country destination

The dataset comes from a kaggle competition supported by Airbnb.



Background and Motivation

- How to effectively enhance user experience and increase total bookings by utilizing huge dataset remains a question to answer
- Airbnb opens this challenge to its potential employees
- From their point of view, by accurately predicting the destination a new user might travel to for the first time, Airbnb can share more personalized content with their community, decrease the average time to first booking, and better forecast demand

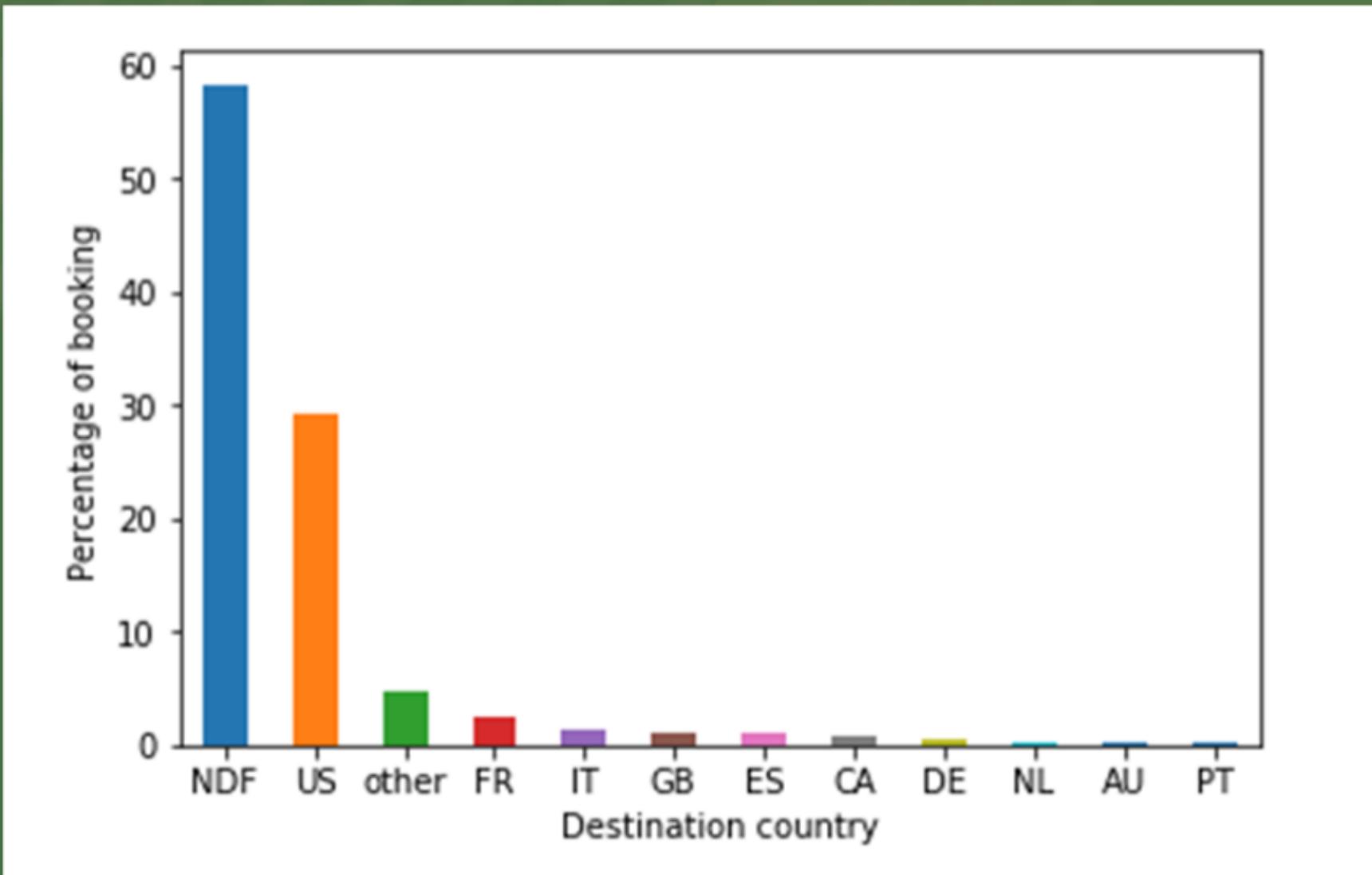
Problem Definition

- 1 As participants, it is required to predict the new users' first booking destination at the granularity of country.
- 2 There are 11 destinations to choose from, including Australia, Canada, Germany, Spain, France, Great Britain, Italy, Netherlands, Portugal, the U.S., and all the rest are labeled as "others".
- 3 Users who haven't made a booking are categorized with the "NDF" label in the destination field.

Idea behind the Project



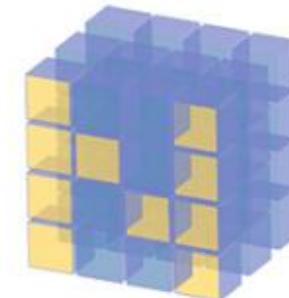
The idea here is how to accurately predict categorical variable by applying different machine learning algorithm.



Libraries



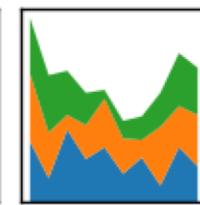
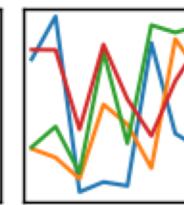
matplotlib



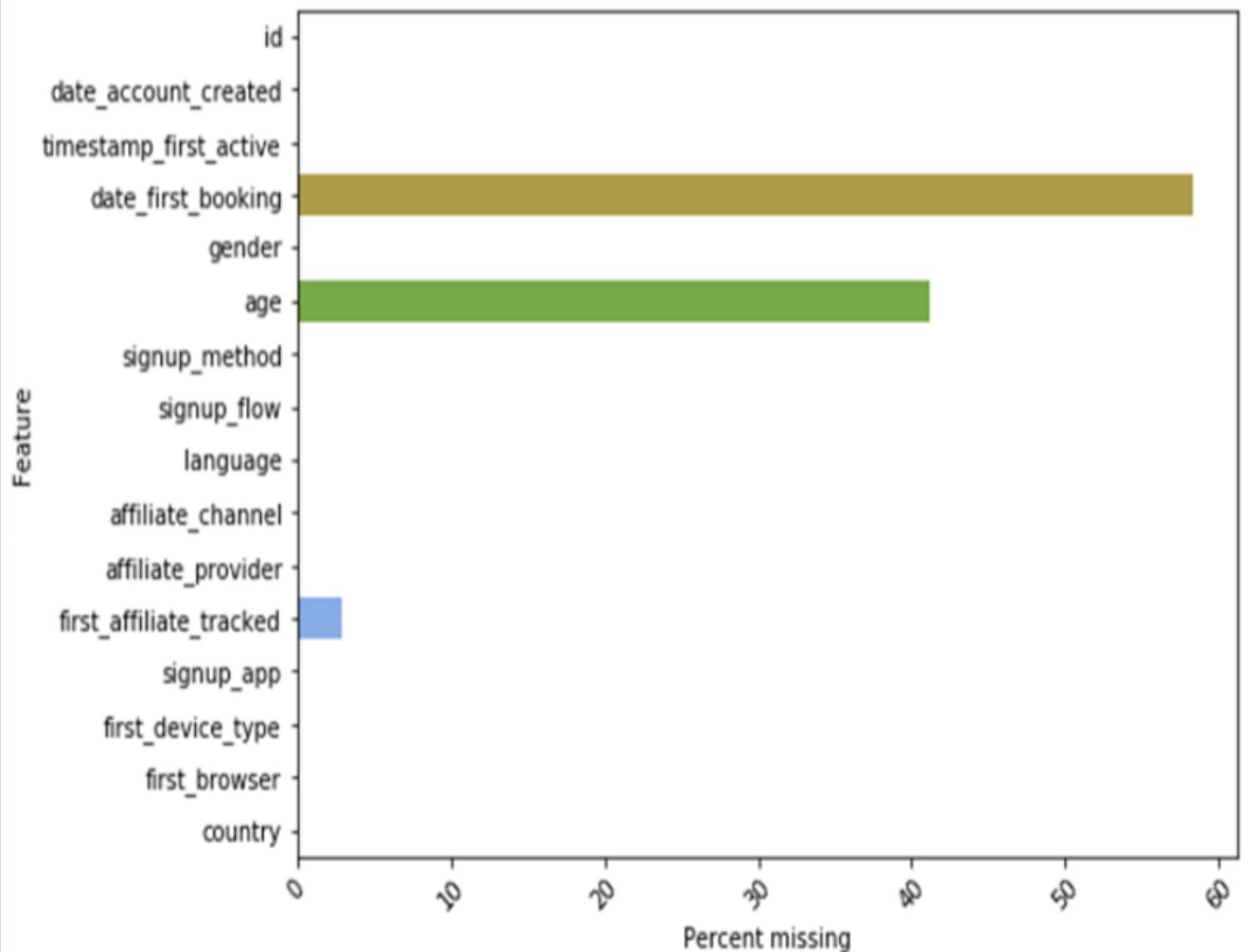
NumPy

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$

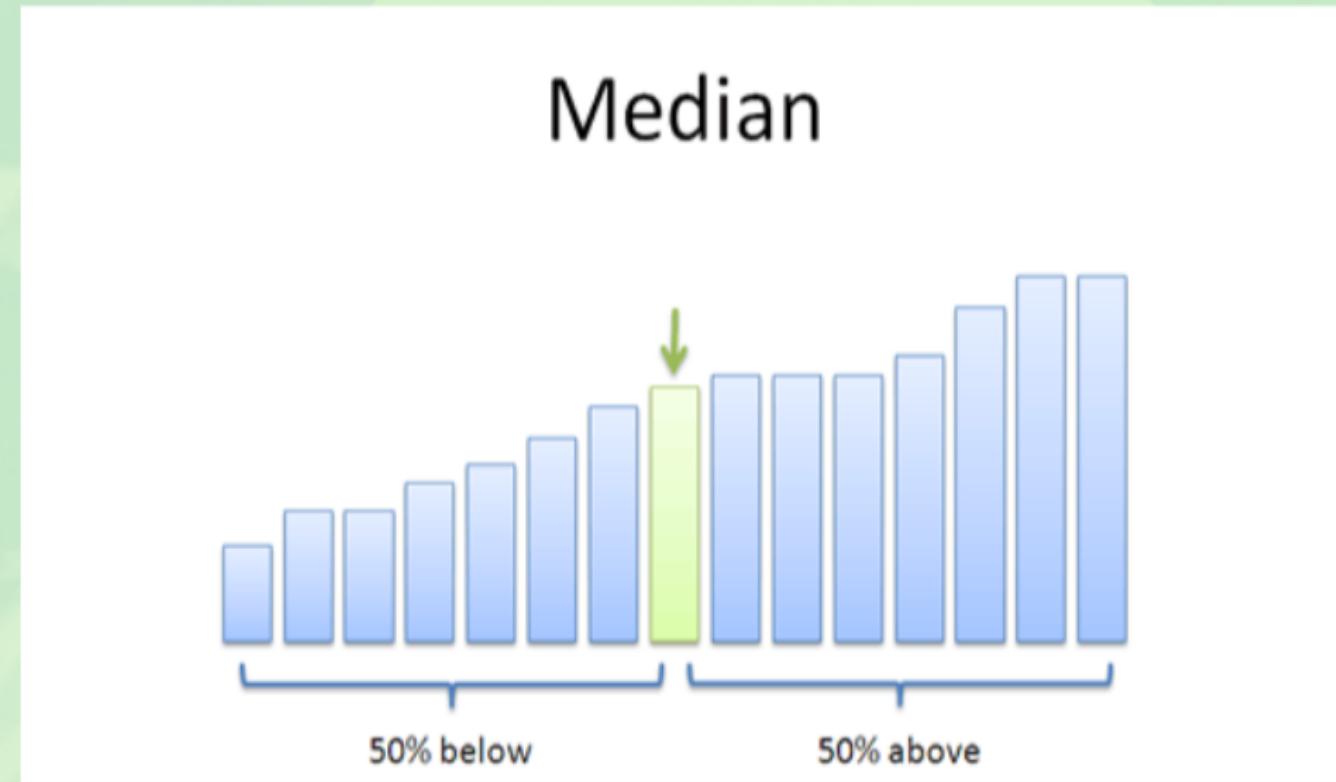


Data Preprocessing

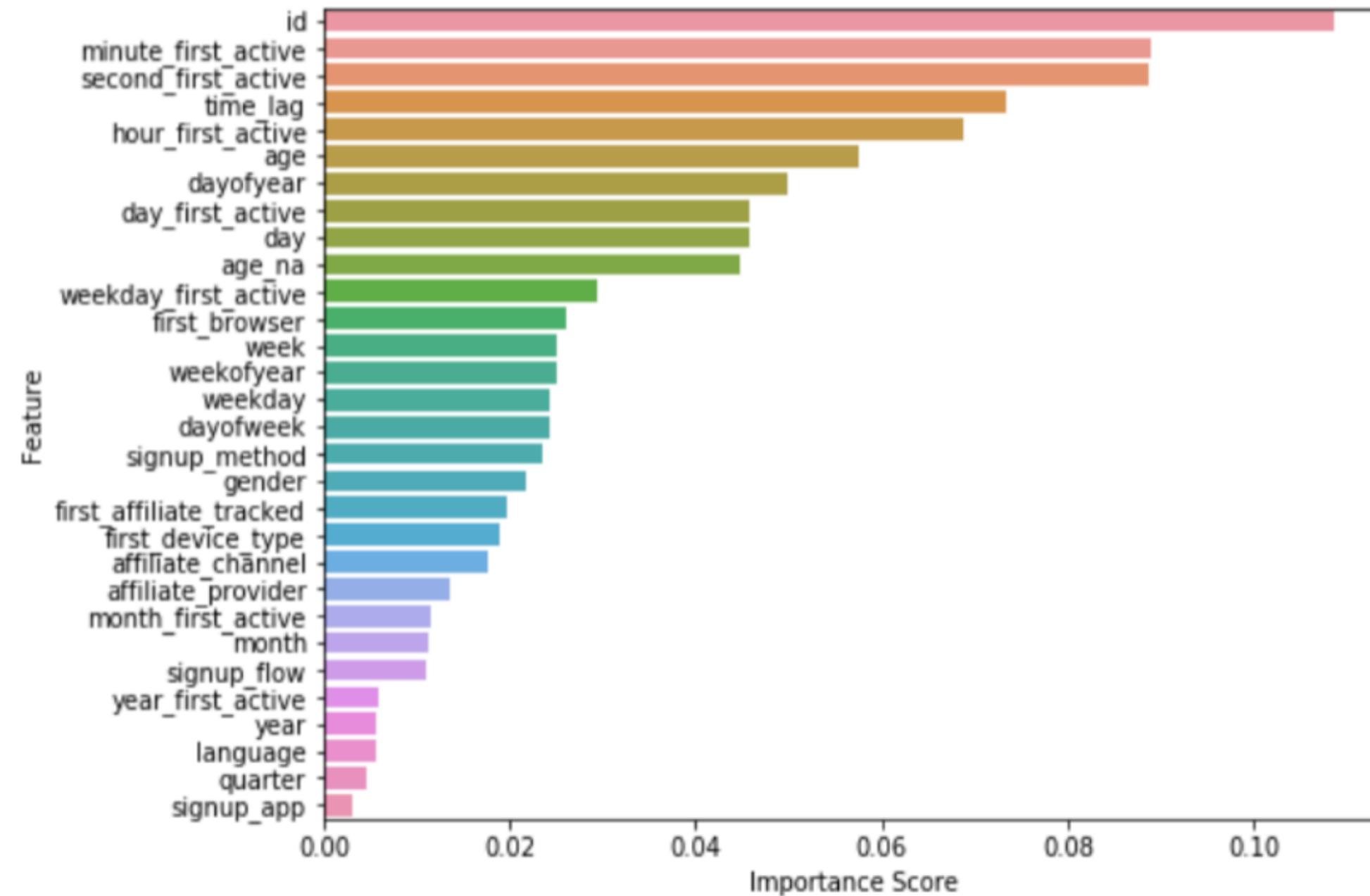


id	0
date_account_created	0
timestamp_first_active	0
date_first_booking	124543
gender	0
age	87990
signup_method	0
signup_flow	0
language	0
affiliate_channel	0
affiliate_provider	0
first_affiliate_tracked	6065
signup_app	0
first_device_type	0
first_browser	0
country	0

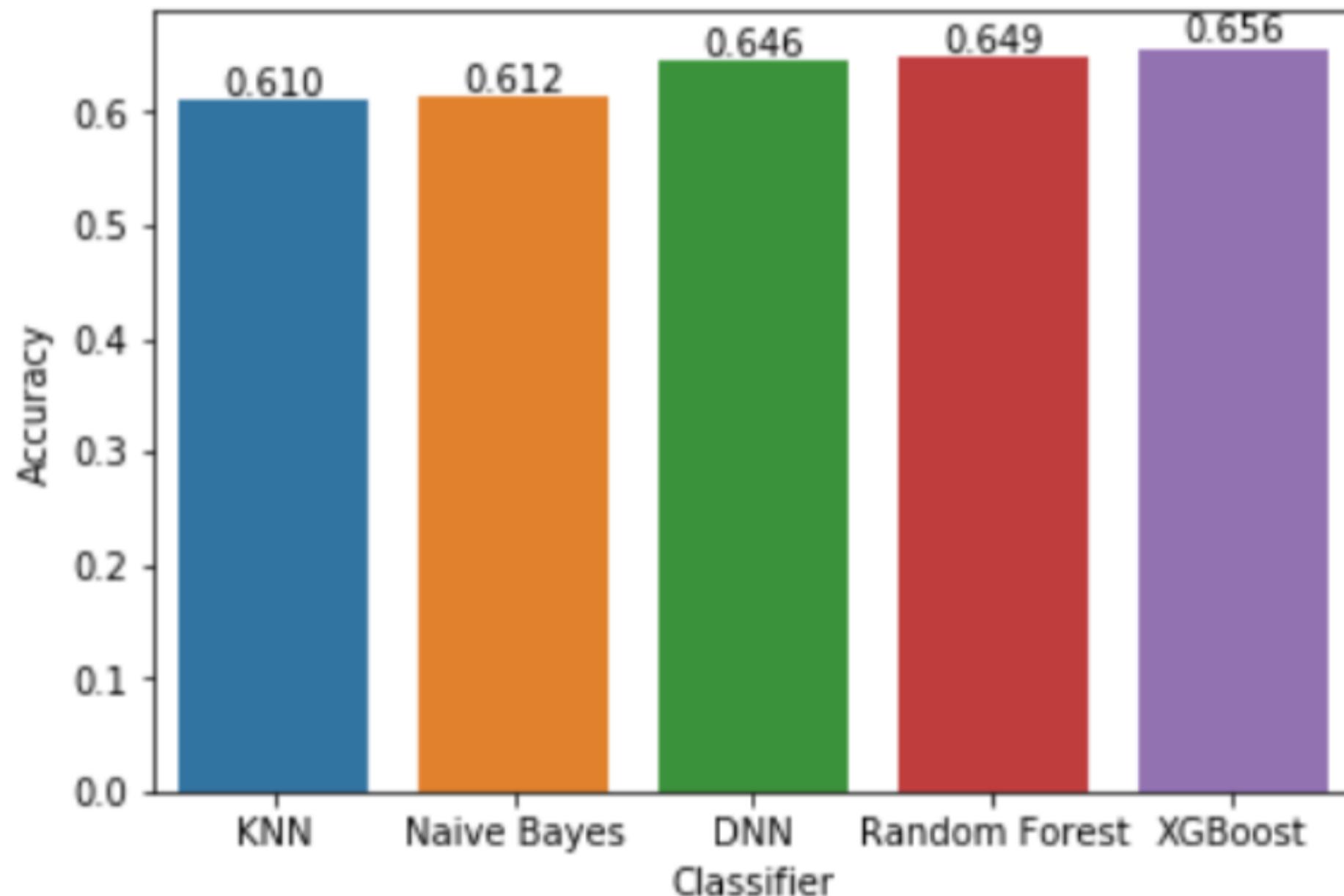
Data Preprocessing



Feature Importance



Model Classifiers



Kaggle Results

The evaluation metric for this competition is NDCG
(Normalized discounted cumulative gain)

Private Score

0.70467

Public Score

0.69824

References

1. Airbnb Recruiting: New User Bookings.

<https://www.kaggle.com/c/airbnb-recruiting-new-user-bookings/data>

2. Feature Selection in Python with Scikit-Learn.

<https://machinelearningmastery.com/feature-selection-in-python-with-scikit-learn/>

3. Feature Selection with Scikit-Learn

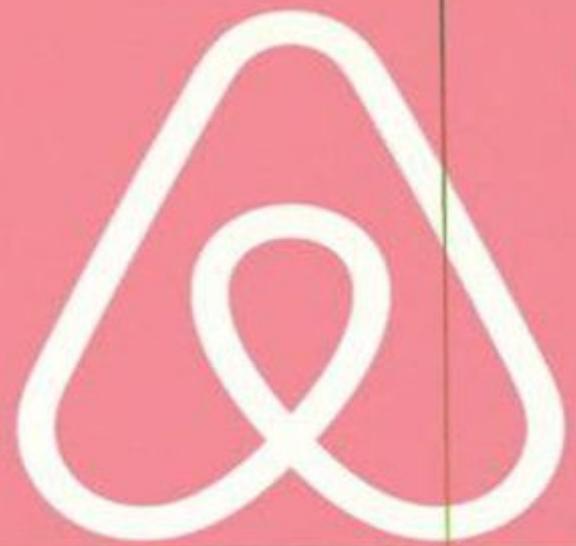
http://scikit-learn.org/stable/modules/feature_selection.html

4. Nearest Neighbors Classification

<http://scikit-learn.org/stable/modules/neighbors.html>



Thank You....



airbnb

