Capstone Project Proposal

Airbnb New User Bookings

Sandrine Estelle Meyantchop

INTRODUCTION

Airbnb is an online marketplace and hospitality service, enabling people to list or rent short-term lodging including vacation rentals, apartment rentals, homestays, hostel beds, or hotel rooms. Airbnb has gained a lot of momentum over the past two years and is the most popular vacation rental stay due to its affordability. Many people have started listing their places to be rented on airbnb's website. New users on Airbnb can book a place to stay in 34,000+ cities across 190+ countries. In particular the objective of this project ist to predict where a new user will book their first travel.

INTENTED CLIENT

With the result of the analysis Airbnb could share more personalized content with their community, decrease the average time to first booking, and better forecast demand.

REFERENCE DATA SOURCE

The dataset give a list of users along with their demographics, web session records, and some summary statistics. All the users in this dataset are from the USA.

There are 12 possible outcomes of the destination country: 'US', 'FR', 'CA', 'GB', 'ES', 'IT', 'PT', 'NL','DE', 'AU', 'NDF' (no destination found), and 'other'. Please note that 'NDF' is different from 'other' because 'other' means there was a booking, but is to a country not included in the list, while 'NDF' means there wasn't a booking.

The training and test sets are split by dates. In the test set, I will predict all the new users with first activities after **7/1/2014**. In the sessions dataset, the data only dates back to 1/1/2014, while the users dataset dates back to 2010.

Dataset on Kaggle: https://www.kaggle.com/c/airbnb-recruiting-new-user-bookings/data

SOLUTION APPROACHES

The intended solution approaches will include the following:

- Make the data exploration to be able to find the correlation between the data
- Use better machine learning model to predict which country a user's first booking destination will be

DELIVERABLES

The deliverables will include the following:

- Report paper for the project
- Code for data analysis and machine learning