Mehreteab Kidane

2nd Capstone Project Proposal
Title: **Airbnb New User Bookings**
Springboard

Problem Statement

New users on Airbnb can book a place to stay in 34,000+ cities across 190+ countries. By accurately predicting where a new user will book their first travel experience, Airbnb can share more personalized content with their community, decrease the average time to first booking, and better forecast demand. Airbnb is the client of this project, the goal is to predict in which country a new user will make his or her first booking so to better forecast demand and therefore increase revenue.

Dataset

The data set is already available to in the form of Kaggle competition by Airbnb in 2015.
https://www.kaggle.com/rounakbanik/airbnb-new-user-bookings/data
The data provided by Airbnb is in the form CSV files and are listed below.

train_users.csv - the training set of users
test_users.csv - the test set of users. Contains User information such as gender, age, language, signup and device information
sessions.csv - web sessions log for users. Contains time, type and details of various user actions
countries.csv - summary statistics of destination countries in this dataset and their locations
age_gender_bkts.csv - summary statistics of users' age group, gender, country of destination
sample_submission.csv - correct format for submitting your predictions

Significance of the problem and my approaches

The approach to solving this problem is subject to change as I progress with career track and learn new concepts and approaches. Here are my approaches –

1. Data collecting: downloading all the data provided by Airbnb to local.
2. Data Wrangling: data cleaning, seek mistakes in data, look for peculiar behavior, fix missing data.
3. Data exploration: use of classification, inferential statistics and data visualization to find interesting trends and identify significant features in the data set.
4. Data Analysis: data manipulation and modeling.
5. Complete and submit final deliverables

Deliverables:

The deliverables will be the codes and visualization techniques on GitHub in the form of Jupyter Notebooks, and a slide desk. This will include a report and I intend to write a documentation explaining the code and the results.