## Airbnb – Capstone Project Proposal

## Problem

## Trying to predict in which country a new user on Airbnb, will make his or her first booking. 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.

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

## Datasets

## A list of users along with their demographics, web session records, and some summary statistics are provided by Airbnb. The challenge is to predict which country a new user's first booking destination will be.

## All the users in this dataset are from the USA.

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

## 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.

## File descriptions

## train\_users.csv - the training set of users

## test\_users.csv - the test set of users

## id: user id

## date\_account\_created: the date of account creation

## timestamp\_first\_active: timestamp of the first activity, note that it can be earlier than date\_account\_created or date\_first\_booking because a user can search before signing up

## date\_first\_booking: date of first booking

## gender

## age

## signup\_method

## signup\_flow: the page a user came to signup up from

## language: international language preference

## affiliate\_channel: what kind of paid marketing

## affiliate\_provider: where the marketing is e.g. google, craigslist, other

## first\_affiliate\_tracked: what is the first marketing the user interacted with before signing up

## signup\_app

## first\_device\_type

## first\_browser

## country\_destination: this is the target variable you are to predict

## sessions.csv – has details on the web sessions log depicting the browsing behavior of users

## user\_id: to be joined with the column 'id' in users table

## action

## action\_type

## action\_detail

## device\_type

## secs\_elapsed

## 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

## The training and test sets are split by dates. In the test set, you 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.

## Deliverable

## For every user in the dataset, submission files should contain two columns: id and country. The destination country predictions must be ordered such that the most probable destination country goes first.

## The file should contain a header and have the following format:

id,country  
000am9932b,NDF  
000am9932b,US  
000am9932b,IT  
01wi37r0hw,FR  
etc.