**Capstone Project Proposal:** Predicting the first travel destination of a new user on AirBNB

Tracy Burns

**Project Goal**

With properties in 34,000+ destinations across 190+ countries, is it possible to personalise the content presented to new users on AirBNB in order to better forecast their needs and reduce the average time to first booking?

In this project, I aim to predict the destination country a new AirBNB user will choose for his/her first booking. The primary datasets will be acquired from the Kaggle competition, ‘AirBNB New User Bookings.’ In the dataset, AirBNB has provided a training dataset of approximately 180,000 entries from the USA, with 12 possible first destination countries: United States, Canada, Great Britain, France, Germany, Spain, Portugal, Italy, Netherlands, Australia, and other (or no booking made). A test set of approximately 40,000 entries has also been provided to test predictions against.

**Who might care?**

While this analysis focuses on AirBNB data, there is wide applicability for this style of predictive analysis across the online travel industry. Online travel agencies accumulate an enormous wealth of user data which could be used to enhance user experience and increase total bookings, as well as better forecasting demand.

**The Dataset**

The primary datasets will be acquired from the Kaggle competition, ‘AirBNB New User Bookings.’ The dataset reveals the basic information of new users over the time period of January 2010 – June 2014.

The AirBNB Kaggle dataset consists of:

* User information: Unique ID, age, gender, web browser, device, country destination, timestamp information, affiliate information, etc
* Session information: Web session log for users
* Countries information: Summary statistics for destinations in the dataset, with particular reference to their location in relation to the USA.
* Age & Gender: Summary statistics of users’ age group, gender and country of destination.

**Approach:**

* Gain understanding of the data and show findings on the characteristics of new users.
* Develop a feature selection strategy based on these observations.
* Create a prediction model, and test/optimise the model against the provided test dataset. Predicted results will be evaluated according to the NDCG (Normalized discounted cumulative gain) metric.
* The final deliverable will be a model that predicts five country destinations for each new user

**Possible limitations:**

The dataset provided by AirBNB only contains new users located in the US, so any model developed may be limited by this data. Whether factual or anecdotal, US travels are travel internationally less frequently than citizens of other nations.

**Deliverables** (available publicly on my Github account):

* All code (in Python) used to conduct the analysis, in the form of Jupyter notebooks.
* A final report detailing the process, outcomes and limitations (if any) of the analysis
* Presentation on the capstone project