**Problem Description:**

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 challenges you to predict in which country a new user will make his or her first booking

In this challenge, you are given a list of users along with their demographics, web session records, and some summary statistics. You are asked to predict which country a new user's first booking destination will be. 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, you will predict all the new users with first activities after 7/1/2014 (note: this is updated on 12/5/15 when the competition restarted). In the sessions dataset, the data only dates back to 1/1/2014, while the users dataset dates back to 2010.

**Dataset details:**

* 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: whats the first marketing the user interacted with before the signing up
  + signup\_app
  + first\_device\_type
  + first\_browser
  + country\_destination: this is the target variable you are to predict

**Data Analysis steps followed:**

Data pre-processing and imputations:

* Finding missing values from provided test and train data.
* Data type of the object columns are converted in to the categorical. ‘str\_to\_cat’ function is used to convert columns with object data type to categorical data type
* Replacing Missing values: ‘find\_missing\_values’ function is used to find the number of missing values in the data set and we have replaced these blanks with median age value.
* Age outliers (Age Limit is 18-100) : Some age values are incorrect, like close to 2000 and less than 18. So we have cleaned such values.
* NDF (No destination found entries are removed): As there was no bookings the entries with NDF status are eliminated.
* Date column is converted in to date time format.

**Algorithms Used:**

1) Gradient boosting:

2) XG Boost:

3) Random Forest: