Airbnb New User Bookings (Kaggle Competition)

I'm hoping to answer:

Where will a new guest book their first travel experience by Airbnb?

What data you're planning to use to answer the question?

Data provided by Airbnb for the purpose of the competition. Here're the summary of the data files that I will be using:

- **train_users.csv**
- **test_users.csv**
- **sessions.csv**
- **countries.csv**
- **age_gender_bkts.csv**
- **Sample_submission.csv**

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's 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

- **sessions.csv** - web sessions log for 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

What do you know about the data so far?

I chose this topic because, airbnb data always seemed interesting. Having used the service both as Guest and Host, I always wondered what one could predict from the airbnb data.

So far, I've explored the datasets and tried to learn what's available for me and what I could learn from it. I've also tried to think along the line of what type of models might help me predict the new users booking destination.

The available data is for:
- 10 different countries
  - 'AU', 'CA', 'DE', 'ES', 'FR', 'GB', 'IT', 'NL', 'PT', 'US'
- For the year 2015