Airbnb New User Bookings

This is a summary document of the important steps taken to cleanup data from Kaggle Airbnb challenge.

Data used is titled “train\_user\_2.csv”. This file contains attributes of guests who have booked a reservation in a country on Airbnb. The end goal of analyzing this dataset is to be able to predict country of destination of users who haven’t booked a listing.

* Step 1: Cleanup age column. Consider numeric variables that looks like date of birth to age. Place different age to age groups.
* Step 2: Categorize first\_browser to five browser types
* Step 3: Categorize first\_device\_types to two devise types
* Step 4: Convert two-digit language to language code for the ease of reading
* Step 5: Add a column that indicates the primary language used for the destination\_country
* Step 6: Find the time difference between the date\_account\_created and date\_first\_booking