

[DATA-06P] Airbnb new user bookings

Miguel-Angel Canela, IESE Business School

October 7, 2016

Introduction

Instead of waking to overlooked “Do not disturb” signs, Airbnb travelers find themselves rising with the birds in a whimsical treehouse, having their morning coffee on the deck of a houseboat, or cooking a shared regional breakfast with their hosts.

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.

In a recruiting competition, Airbnb challenged the contestants to predict in which country a new user will make his or her first booking. Those who impress with their answer (and an explanation of how they got there) would be considered for an interview for the opportunity to join Airbnb’s Data Science and Analytics team.

The data set

In this challenge, you are given a list of users along with their demographics, web session records, and some summary statistics. You are then asked to predict which country a new user’s first booking destination will be. All the users in this data set 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 that there wasn’t a booking.

The file `users.csv` contains data about the users. The fields are:

- `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 sign 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 the signing up.
- `signup_app` .

- `first_device_type` .
- `first_browser` .
- `country_destination` : this is the target variable you are to predict.

The file `sessions.csv` contains the web sessions log for users. The fields are:

- `user_id` : to be joined with the column `id` in users table.
- `action` .
- `action_type` .
- `action_detail` .
- `device_type` .
- `secs_elapsed` .