### 

**Capstone Statistical Analysis**

**By Sumita Soundararajan**

***Foundations of Data Science Workshop by Springboard***

**Airbnb Recruiting - New User Bookings**

## ***Where will a new guest book their first travel experience?***



## **Problem**

## Trying to predict in which country a new user on Airbnb, will make his or her first booking. There are 11 potential countries along with a 12th class - NDF (No Destination Found), indicating the user did not make any booking.

## 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.

## URL -> <https://www.kaggle.com/c/airbnb-recruiting-new-user-bookings>

**Statistical Analysis**

## **Training User data**

## Country\_destination - has 12 valid classes with NDF meaning no booking. About 123,489 users ie 58% of users have not made a booking.

## 87,590 users have made a booking, out of that 61,457 users have chosen US – That is among those that book, 70% chose to do it in US. The rather imbalanced classes will be addressed during training the model.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **PT** | **AU** | **DE** | **NL** | **CA** | **ES** | **GB** | **IT** | **FR** | **Other** | **US** | **NDF** |
| 83 | 152 | 250 | 247 | 440 | 707 | 731 | 979 | 1435 | 3655 | 20095 | 45041 |

## Age - there are many values in thousands and some in single digits. 41% of users have missing age. 23% of users with missing age have made a booking.

## Further analysis shows that the bulk of booking is done by users in age-group 30-60.

## The bookings drop significantly once the user’s age exceeds 60.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Age group** | **GB count** | **GB %** | **US count** | **US %** | **FR count** | **FR %** | **ES count** | **ES %** | **IT count** | **IT %** | **Other count** | **Other %** |
| 15-30 | 563 | 32 | 17,514 | 37 | 1,185 | 32 | 685 | 41 | 697 | 35 | 2,598 | 35 |
| 30-60 | 1,049 | 60 | 28,124 | 59 | 2,256 | 62 | 915 | 55 | 1,168 | 59 | 4,459 | 60 |
| >60 | 126 | 7 | 2,028 | 4 | 210 | 6 | 76 | 5 | 123 | 6 | 355 | 5 |
| **Total** | **1,738** |  | **47,666** |  | **3,651** |  | **1,676** |  | **1,988** |  | **7,412** |  |

## Language -The language spoken is distributed as shown below. It is not surprising that most users speak English since Airbnb is a company located in US and its customers are mostly Americans.

|  |  |  |
| --- | --- | --- |
| **Language** | **Count** | **Percent** |
| en | 206314 | 96.66% |
| zh | 1632 | 0.76% |
| fr | 1172 | 0.55% |
| es | 915 | 0.43% |
| ko | 747 | 0.35% |
| de | 732 | 0.34% |
| it | 514 | 0.24% |
| ru | 389 | 0.18% |

1. Gender –Users with 'unknown' gender book less frequently than those with a known one – either MALE or FEMALE. We can see that there are a lot of missing values for gender. Almost half of the users did not input their gender information.

|  |  |  |
| --- | --- | --- |
| **Gender** | **Count** | **Percent** |
| Unknown | 95,688 | 44.9% |
| Female | 63,041 | 29.6% |
| Male | 54,440 | 25.5% |
| Other | 12 | 0.0% |

## Below plot shows the booking done by each gender – Male or Female. There is no

## significant difference in the booking based on gender.