Airbnb Case Study

By

*Ramya D & Jaydeep Nawagekar, DSC 27*

**Data Methodology**

**Note: Please kindly review the python and tableau files which are there along with the other files.**

* We conducted a thorough analysis of the Airbnb dataset. The process included:
  + Dropped columns such as name and last review which are not useful for analysis
  + Performed basic data checks to identify constant columns, missing values, duplicate rows etc
    - No constant columns or duplicate rows/values found in the dataset
    - After dropping last\_review and name columns, only reviews\_per\_month has missing values (20.5%) which are missing at random
    - Imputed the missing values in the column “reviews per month” with a constant value of 0
  + Used scatter plots to identify abnormal data points and perform sanity checks
  + Sanity checks included to find any abnormal values outside the logical range such as
    - Availability\_365 should be between 0 and 365
    - Price should be greater than 0 etc
  + Although we have identified a few outliers in the price column, where the listings have a price of 0$ per night, we’ve decided to let them be part of analysis because of two reasons
    - They are very few (11 out of 40k+)
    - Important information in the remaining columns need to be retained for analysis
  + We’ve converted numerical variables into categorical bins to qualitatively analyse relationships using quantiles.
  + Later we created various pivot tables, univariate and bivariate graphs, statistical analysis on important columns to identify key relationships and customer preferences