Project 1 - Diplodocus Team

*KC Oh, Alfonso Camero, Marie Prosper*

Project Description

Since its inception in 2008, Airbnb has disrupted the traditional hospitality industry as more travelers decide to use Airbnb as their primary means of accommodation. Airbnb offers travelers a more unique and personalized way of accommodation and experience.

This dataset has one file- AB\_US\_2020.csv which has columns describing features such as host id, hostname, listing id, listing name, latitude and longitude of listing, the neighborhood, price, room type, minimum number of nights, number of reviews, last review date, reviews per month, availability, host listings and city.

Inspiration

* Can we predict the price of each house in different regions?
* Can we describe a region using the names of listings in that region?
* What can we learn about different regions from the data?
* Based on different factors is it possible to recommend a title to the host for his/her listing?
* Can we estimate the popularity of a listing based on given features?

What are we trying to solve? Independent/Dependent variables

* + What cities offer the most availability?
  + What areas offer the best prices?
  + How does population affect price and availability?
  + What are the most established market for Airbnb based on customer review?

Datasets

* + Csv file: <https://www.kaggle.com/kritikseth/us-airbnb-open-data>
  + Citipy, US census

Tasks Assignments (draft)

* + Marie Prosper: Project manager, Histogram
  + KC Oh: Bar Graphs
  + Alfonso Camero: Maps