**Project Title:** Airbnb Usage in NYC vs. San Francisco  
  
**Team:** Amgad Nassif, Mikey Esteban, Benedict Waiharo, Christine Mazur  
  
**Project Description**: Tourist preference between NYC and SF based on Airbnb booking rate  
  
**Research Questions**:

1. Number of listings by price or range of prices
2. Density of listings by zip code?
3. Does review score affect price?
4. What are the peak rental times in NYC & SF?
5. Group by minimum nights allowed to determine which units are daily (1 day = hotel style), monthly (30 days = mo. to mo. apt. style), yearly (365 = std. apt. style).
6. Some hosts have more than 1 listing; group the listings by owners of 1, 2-5, 6-10, etc. to determine corporate vs individual ownership
7. What else is interesting to each of you?

**Datasets:** Airbnb .csv for NYC and SF (price, lat/lon, type, length, amenities).

Do we need data for actual listings rental rate? These datasets show only the listings themselves, not when they were actually booked during the year.

**Rough breakdown of tasks**:

* Create a new repository for this project called Project1
* Inside the Project1 git repository add: this file, a resources directory and a jupyter notebook
* Research and analyze datasets for New York and San Francisco Airbnb listings through Kaggle and Rapid .apis.
  + NYC\_Listings.csv & SF\_Listings.csv
  + <http://insideairbnb.com/index.html>
* Segment the NYC dataset due to large size (>180MB)
* Pandas
  + Load, read and combine datasets into Pandas
  + Determine merge key for datasets
* use matpltlib to visualize NYC and SF datasets
  + Total number of rentals by month
  + Average review score by month
  + Use latitude and longitude to plot rental locations on a SF map and a NYC map to show density comparisons visually
  + See review questions above
  + WHAT else?
* draw conclusions and summarize research results

**Contingency plan:** airline credit card company’s affiliation analysis.