

**»Data Mining course«**

**Project 1 - AirBnb**

**Dr.Farahani , Dr.Kheradpisheh**

**Ashkan Safavi Sohi**

**98422096**

**About dataset**

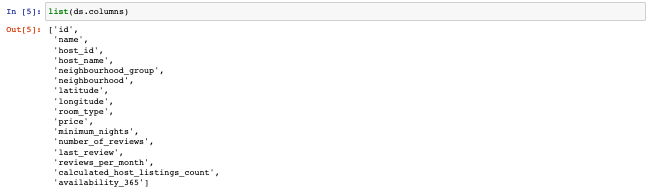
Since 2008, guests and hosts have used Airbnb to expand on traveling possibilities and present more unique, personalized way of experiencing the world. This dataset describes the listing activity and metrics in NYC, NY for 2019.This data file includes all needed information to find out more about hosts, geographical availability, necessary metrics to make predictions and draw conclusions.

**Analyze and Questions**

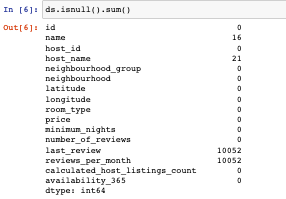
1. What can we learn about different hosts and areas?



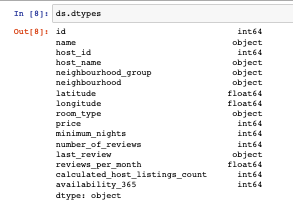
In the first step, we realize that our database consists of 16 columns and 48895 rows.



The information of the columns that are dataset feature can be seen in the list above.



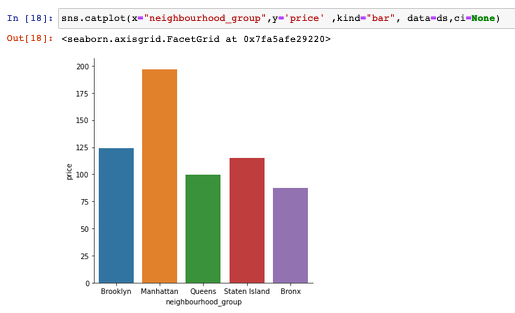
In the list above, it can be seen that this database also contains null data that can be deleted or replaced by the desired number if needed.



This database includes 10 numerical feature and 6 catgorical feature.

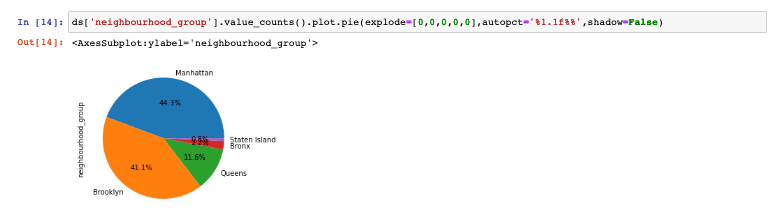


1. What can we learn from predictions? (ex: locations, prices, reviews, etc)



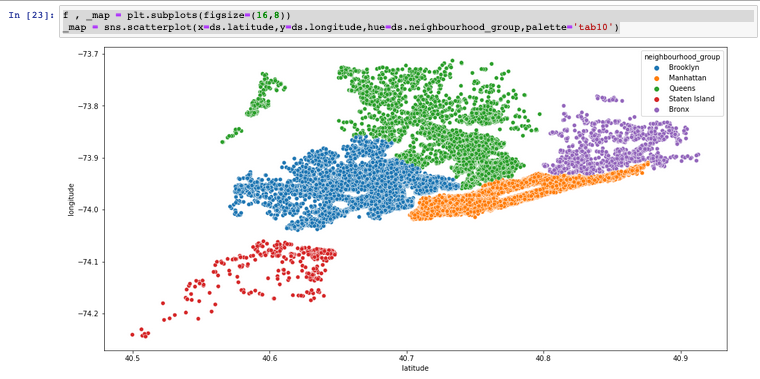
Manhattan has more expensive hotels than anywhere else

1. Which hosts are the busiest and why?



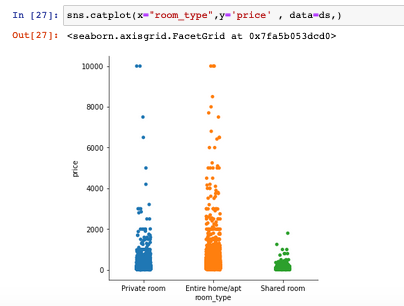
First Manhattan and then brooklyn have a larger share in the number of hotels

1. Is there any noticeable difference of traffic among different areas and what could be the reason for it?



The chart above shows the distribution of bookers by geographical location , Population density in the range of -74 to -79.3 longitude is quite clear .

1. Price dispersion based on types of places



As can be seen from the diagram above, the rooms have various prices.

1. What are the most popular neighborhoods?

