IBM Capstone Project

USING DATA SCIENCE TO DETERMINE BEST PLACE TO STAY IN NEW YORK

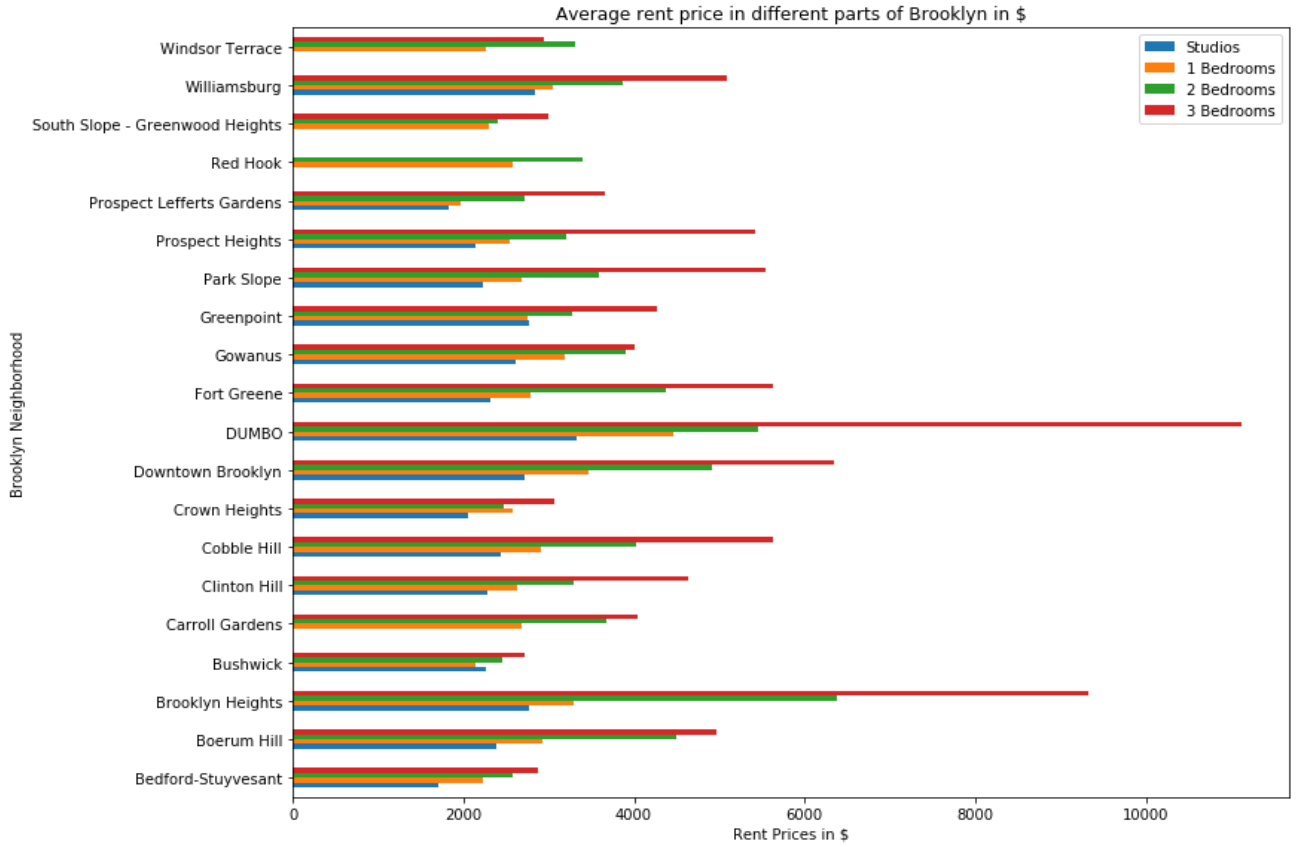
Introduction

- In the globalized world today, it is exceptionally valuable to have international work experience. Most young adults like myself in Singapore would have considered moving overseas for opportunities to major cities such as Toronto or New York. However, such cities can have a high cost of living. For the purpose of this project, we will be picking New York City to determine its quality of life there should we decide to work there.
- My target audience will be anyone interested to work in New York City and rent a place there.The information is useful for anyone who wish to work in New York City even though I talked about young adults.
- There are 2 factors that we should consider for working in New York City.
- The cost of living (essentially average apartment rent in Brooklyn & Manhattan)
- Attractiveness of the neighborhood (top venues analysis)
- We will use unsupervised machine learning (k-means clustering) to answer this question

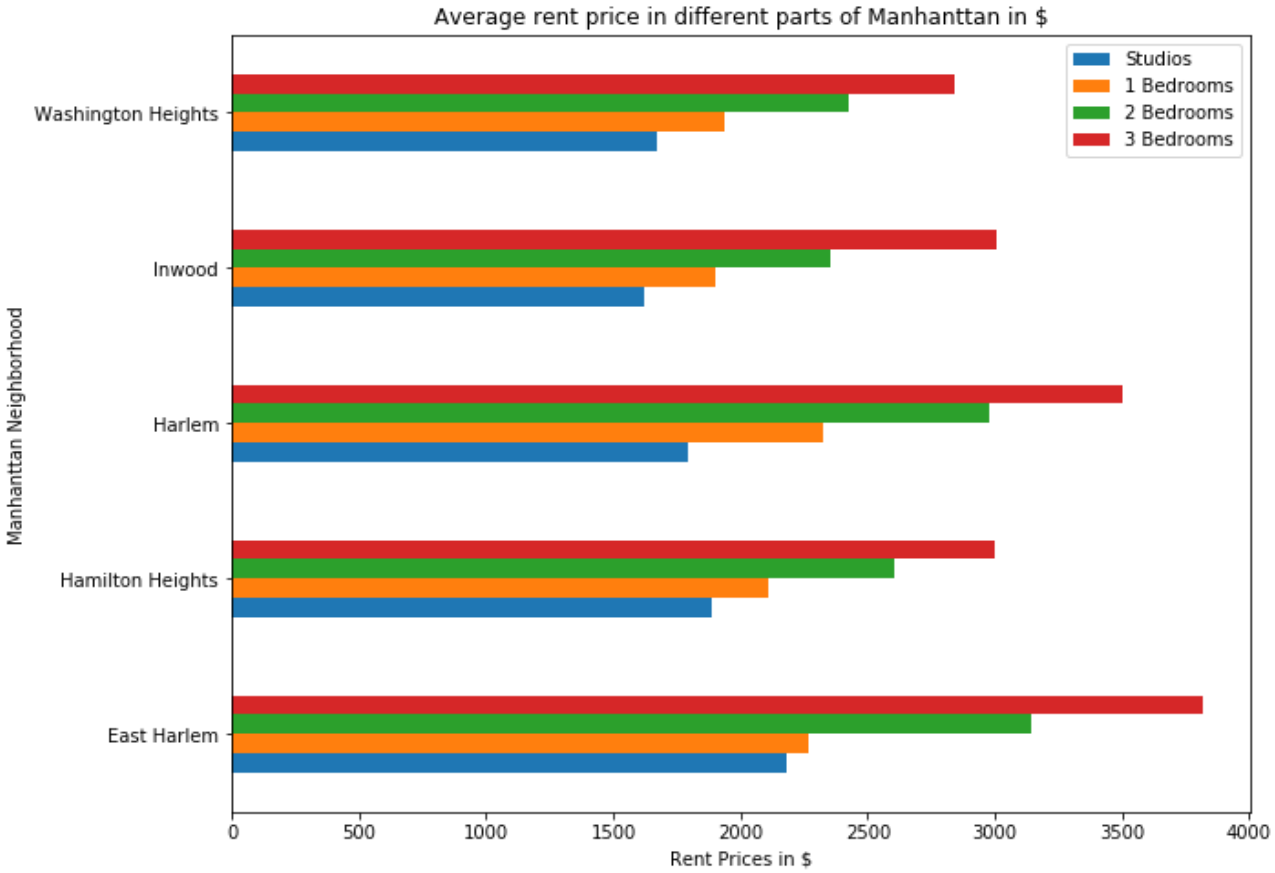
Type and sources of the datasets

- **Type and sources of the datasets**
- The datasets required for this project are:
- Data about Borough and Neighborhood in New York City
- Data about the location information (Latitude and Longitude of Borough and Neighborhood in NYC)
- Data about average apartment rent by Neighborhood
- With the longitude and latitude information, we can use Foursquare API to provide information on each NYC neighbourhood.
- 1) Data about Borough, Neighborhood, location information in NYC : Neighborhood has a total of 5 boroughs and 306 neighborhoods Wikipedia bourough and neighbourhood dataset
- Data Source : This dataset is available on the web in the following link : https://geo.nyu.edu/catalog/nyu_2451_34572 (https://geo.nyu.edu/catalog/nyu_2451_34572)
- This information can also be found in coursera lab
- 2) Data about average apartment rent by Neighborhood : NYC Average Neighborhood Rents in Feb 2019
- I have complied cleaned the data in excel before uploading it into this python notebook Data Source: This can be found in the url below https://www.cityrealty.com/nyc/market-insight/features/get-to-know/map-average-nyc-rent-prices-february-2019/26722

Graph 1


Average rent price in different parts of Brooklyn in $
Studios
1 Bedrooms
2 Bedrooms
3 Bedrooms
Windsor Terrace
Williamsburg
South Slope - Greenwood Heights
Red Hook
Prospect Lefferts Gardens
Prospect Heights
Park Slope
Greenpoint
Gowanus
Fort Greene
DUMBO
Downtown Brooklyn
Crown Heights
Cobble Hill
Clinton Hill
Carroll Gardens
Bushwick
Brooklyn Heights
Boerum Hill
Bedford-Stuyvesant
Brooklyn Neighborhood
0
2000
4000
6000
8000
10000
Rent Prices in $

Graph 2


Average rent price in different parts of Manhanttan in $
Studios
1 Bedrooms
2 Bedrooms
3 Bedrooms
Washington Heights
Inwood
Harlem
Hamilton Heights
East Harlem
Manhanttan Neighborhood
0
500
1000
1500
2000
2500
3000
3500
4000
Rent Prices in $

Summary

- After zooming into 2 cheapest area of NYC, we shortlisted Brooklyn and Upper Manhattan.
- Based on the needs required for myself, I would think that a studio apartment is most relevant.
- If I were to stay in Upper Manhanttan,Harlem will be my choice based on the cheapest price.
- While if I were to stay in Brooklyn, Prospect Lefferts Gardens or Bedford-Stuyvesant would be my obvious choice.
- Here are the prices summarized below
- Washington Heights - $1623
- Inwood - $1669
- Prospect Lefferts Gardens - $1827
- Bedford-Stuyvesan - $1700
- From here, I believe the price difference is not as important as other factors such as the venue analysis factors and the location of your job.
- However, based on the available dataset I have, I have used Brooklyn to do a further depth analysis.
- K-means will partition Bedford Stuyvesant,Brooklyn into mutually exclusive groups, in our example, into 4 clusters. The venues in each cluster are similar to each other. Now we can create a profile for each group, considering the common characteristics of each cluster. For example, the 4 clusters can be:
- Cluster 1 : Mostly food and drinks places
  Cluster 2 : Historic Site
  Cluster 3 : Dog Run
  Cluster 4 : Boutique Shop

Analysis of Results

- We have used data science to determine which neighbourhoods in NYC is good to move into. In terms of prices, we have considered Upper Manhattan and Brooklyn.
- However, based on the most accurate dataset we have, we have decided to explore Brooklyn further.
- From Brooklyn, we went deeper to find how that Bedford Stuyvesant and Prospect Lefferts Gardens are relatively cheaper than the rest.

- We went on to do a venue analysis using foursquare API.
- Bedford Stuyvesant has good clusters of food and drinks which is quite ideal for us and the low rental price makes it as appealing choice for me.

What can we do to improve further?

- Diversity of people (Better culture acceptance and friendship)
- Schools (For kids)
- Proximity to good transport (Reduce travelling time)