A Tale of Two Cities

Introduction:

In the Capstone Project of Data Science Profession certificate we will compare two cities, New York City, USA and Mumbai City, India. Both the cities are financial and cultural centres of their respective countries. Also, Mumbai as well as New York is have one of the best educational institutes with respect to their countries. Lots of students look for a comparison between the cities all over the world when it comes to spending money on education and its overall worth after completion of it. They look for segmentation of neighborhoods, population, desirability rank, number of ranked universities, employer activity rank, affordability rank etc. In this project we will segment the neighborhoods as well as compare the cities on parameters mentioned above.
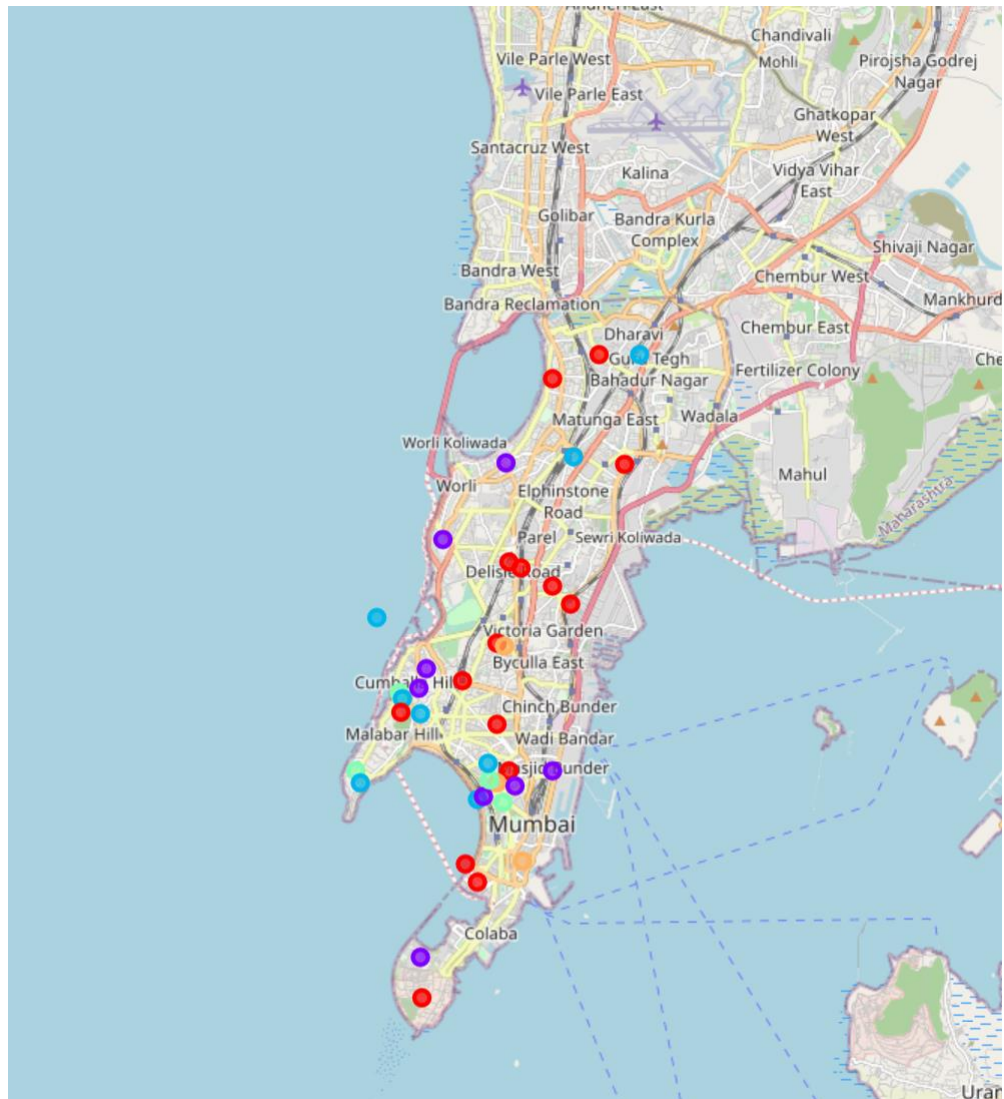
Data:

1. Forsquare APIs will be used for location of neighborhoods and getting the venues in radius of 1 Kms. The data for Mumbai is taken from the Wikipedia and link for the same is: https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Mumbai. The data is scrapped using "pandas" and taken into dataframe and analysed further.
2. The data for population, desirability rank, student affordability rank, number of ranked universities, employer activity rank etc. has been taken from QS best student cities page for Mumbai and Newyork. The link for the same is as follows:
   Mumbai: https://www.topuniversities.com/university-rankings-articles/qs-best-student-cities/mumbai
   New York: https://www.topuniversities.com/university-rankings-articles/qs-best-student-cities/new-york

Methodology:

- As we have already learnt the neighbourhood venue data preparation for New York City in week 2, here we will go through the data preparation mechanism for Mumbai City, India.
  1. A data is scrapped from the Wikipedia page which gave us all neighborhoods of Mumbai City including Suburbs.
  2. We will compare mainly the Manhattan region of New York City and South Mumbai region of Mumbai city which show certain resemblance as both are coastal region and financial centers of respective cities.
  3. The data for Mumbai suburbs should be dropped and a resulting dataframe will provide us the necessary fields i.e. borough, neighbourhood, latitude and longitude.
- A K-Means clustering technique is used to segment neighbourhood in South Mumbai and Manhattan Newyork and value for K is taken as 5.
- For the second part of comparison between two cities, data is scrapped from QS website (link for the same is provided in data section).
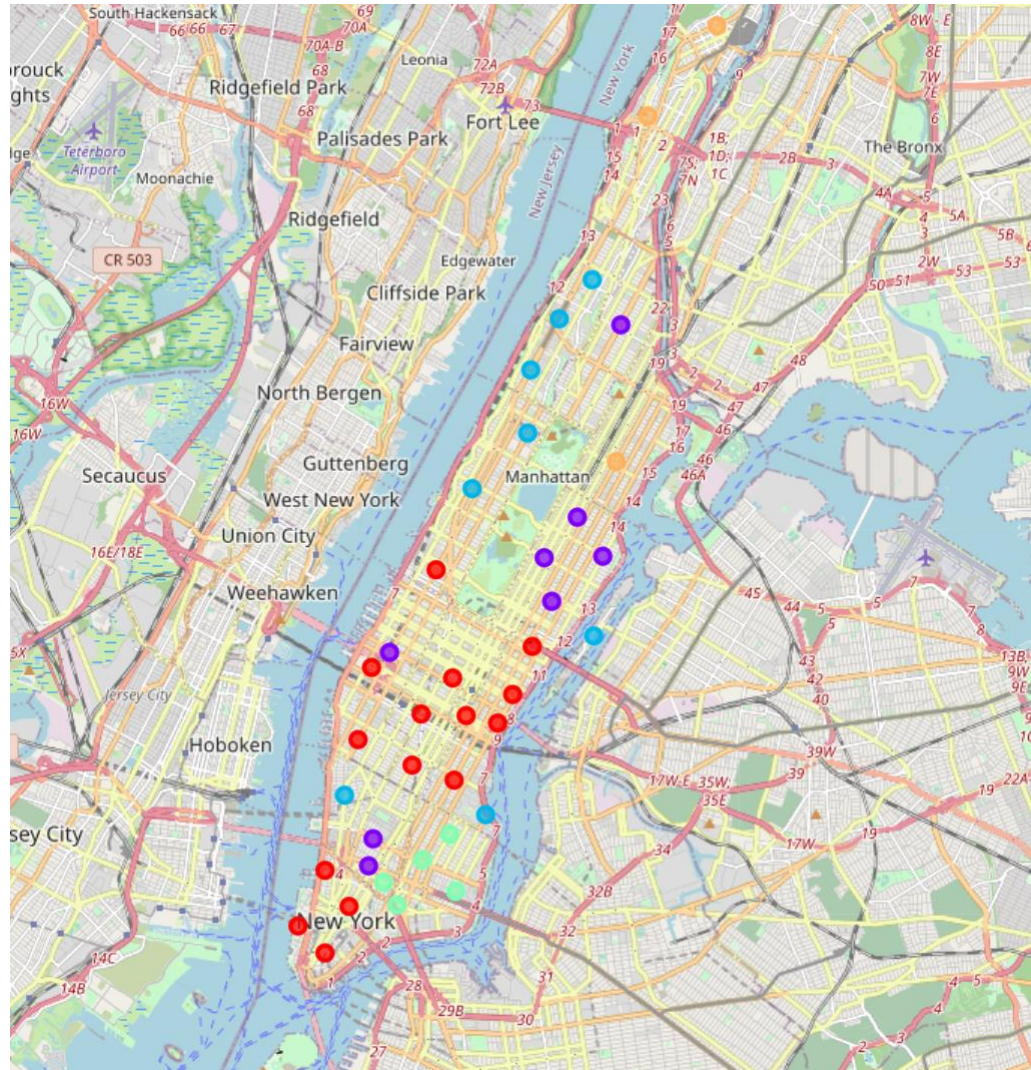- For comparative analysis a bar chart is created.
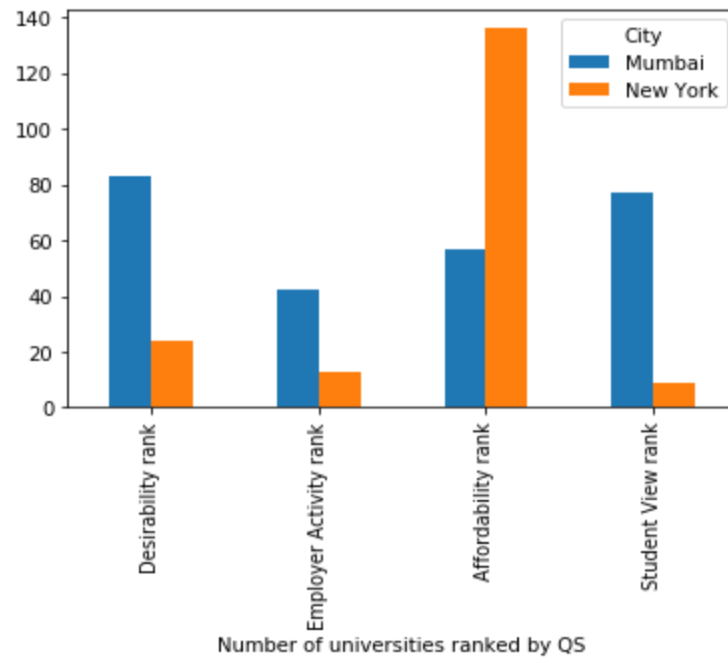
Results:

1. Mumbai City Clustering Map:

2. Manhattan, New York Clustering Map

3.

Referring to cluster of neighborhoods in Manhattan, NY and Mumbai, it is quite clear that similar kinds of venues are located nearby in Manhattan, NY while Mumbai City is quite diverse.

3. Education Related Parameters

Looking at above bar-chart it is quite sure that New York is way ahead of Mumbai in higher education while it lags in affordability index.

Discussion:

1. Mumbai being more diverse place than New York on base of venue segmentation.
2. Students prefer mostly New York when it comes to choosing which university they want to study in.
3. Education in Mumbai is way cheaper than New York.
4. While analysing the most common places in each clusters, it is found that in every cluster of Mumbai most common place is Indian Restaurant.

Conclusion:

On comparing both financial centres of respective countries, it is found that Mumbai is diverse than Manhattan region of New York. Also when it comes to choice of place of study, students choose New York over Mumbai. Mumbai is affordable than New York.