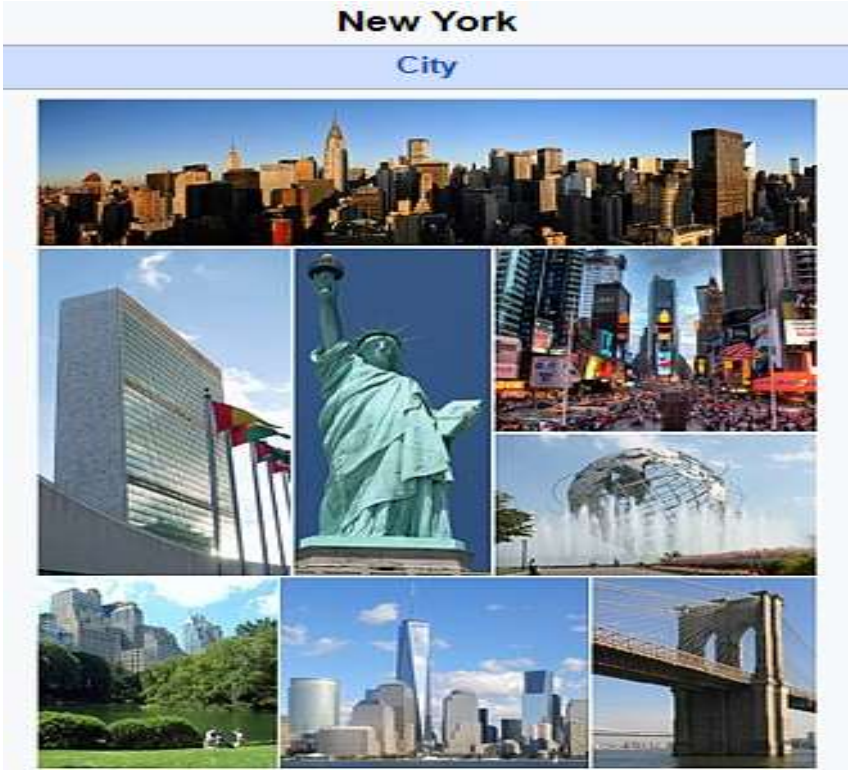
APPLIED DATA SCIENCE CAPSTONE PROJECT

The Battle of the Neighborhoods

New York, United States



BUSINESS PROBLEM DEFENTION

AND

SOLUTION DESIGN & DATA REFERENCES

Week1 - Submission

By SAJITH M P

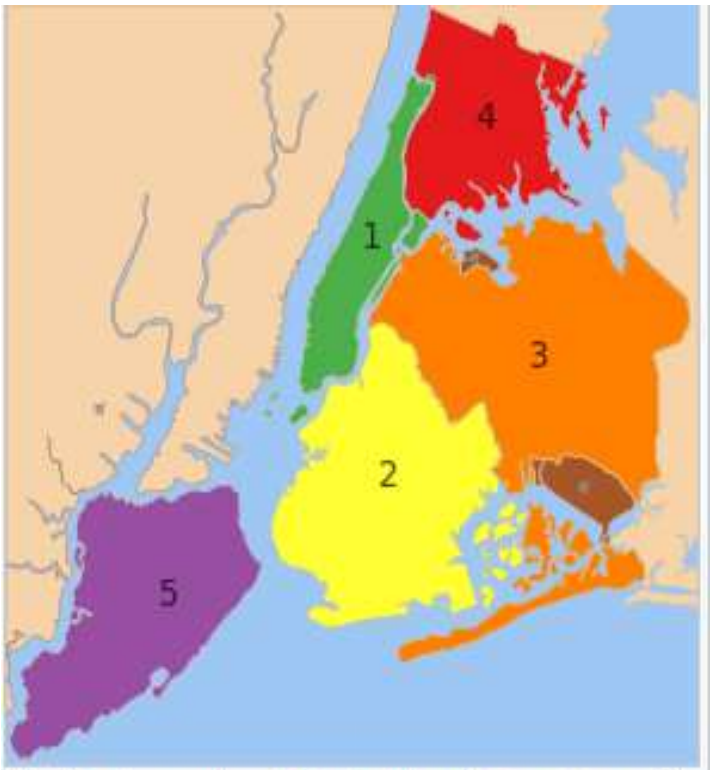
Introduction

The City of New York, usually called either New York City (NYC) or simply New York (NY), is the most populous city in the United States. With an estimated 2018 population of 8,398,748 distributed over a land area of about 302.6 square miles (784 km^2),

It is diverse and is the financial capital of USA. It is multicultural. It provides lot of business opportunities and business friendly environment. It has attracted many different players into the market. It is a global hub of business and commerce. The city is a major center for banking and finance, retailing, world trade, transportation, tourism, real estate, new media, traditional media, advertising, legal services, accountancy, insurance, theater, fashion, and the arts in the United States. This also means that the market is highly competitive. As it is highly developed city so cost of doing business is also one of the highest. Thus, any new business venture or expansion needs to be analysed carefully. The insights derived from analysis will give good understanding of the business environment, which help in strategically targeting the market. This will help in reduction of risk and better control on the Return on Investment

New York is also the most densely populated major city in the United States. Located at the southern tip of the state of New York.A global power city, New York City has been described as the cultural, financial, and media capital of the world, and exerts a significant impact upon commerce, entertainment, research, technology, education, politics, tourism, art, fashion, and sports.

NY is split up into five boroughs: the Bronx, Brooklyn, Manhattan, Queens, and Staten Island. Each borough has the same boundaries as a county of the state.



The five boroughs: 1: Manhattan, 2: Brooklyn, 3: Queens, 4: The Bronx, 5: Staten Island

BUSINESS PROBLEM

The City of New York is famous for it's excellent cuisine. It's food culture includes an array of international cuisines influenced by the city's immigrant history. Italian & Indian restaurants have become so popular in the United States now it seems that there is one on every corner, not only in major cities but also in smaller cities. One of my friends who is thinking of starting a restaurant in the NY neighbourhood, consulted with me to get some analysis done with the all-possible data available. Manhattan being the costliest place, it was decided to compare rest of the boroughs and pick one of the most suitable neighborhoods with in the shortlisted boroughs. Based on the data analysis, it is expected to logically conclude which restaurant type (Italian Or Indian) and its recommended location. All the choices to be rationalized with the data analysis & it helps to distinguish the selections, securing long-term success.

Overall Problem Statement can be broken into the following

- Exploring the Boroughs in NY and narrow down to one.
- Explore the Venues in the neighborhoods across that specific Borough
- Narrow down to handful of neighborhoods and then deep dive into the current Restaurants & Hotels landscape across those.
- Venue clustering by filtered neighborhoods and analyze the best choice of the restaurant and the best fit location.

SOLUTION DESIGN APPROACH & DATA REFERENCES

Solution is approached in seven steps as listed below

STEP 1: Pull all the boroughs & the respective neighborhood details of the New York data using newyork_data.json.['newyork_data.json' - https://cocl.us/new_york_dataset]

STEP 2: Narrowing down to one of the Boroughs – Basis of Population/Density analysis-

on the data available in Web.

https://en.wikipedia.org/wiki/Demographics_of_New_York_City"

STEP 3: Deep Dive into the shortlisted Borough from Step 2 Using **FourSquare APIs**

STEP 4: Explore Venues across the neighborhoods in that Borough & Narrow down to handful of it based on larger number of Venues Vs less number of Restaurants +Hotels

STEP 5: Deep Dive into the shortlisted neighborhoods using, **Word Cloud**, **Means of frequency** of each category of Restaurants & identifying the **Top5 Common** Restaurants/Hotels

STEP 6: Clustering the neighborhood using **K-means** & identifying the locations on the Map.

STEP 7: Concluding the Choices of Restaurants & Locations basis of the data analysis in Step

Conclusion

It's an attempt to explore the different possible analysis we could do in the available data and rationalize the decision. Although all of the goals of this project were met there is definitely room for further improvement by analyzing few more supplementary data points like demographic information, Average Spent of the population, Proximity of other crowd pulling venues like Malls, shopping complex etc. However, this project could definitely be handy to narrow down a Neighborhood and a type of Restaurant as a first step.