Coursera Capstone

IBM Applied Data Science Capstone

Opening a new Yoga Studio in New York City

Introduction

New York City(NYC), is the most populous city in the United States. With an estimated 2019 population of 8,336,817 distributed over about 302.6 square miles, New York City is also the most densely populated major city in the United States. Located at the southern tip of the U.S. State of New York, the city is the center of the New York metropolitan area, the largest metropolitan area in the world. New York City has been described as the cultural, financial, and media capital of the world, significantly influencing commerce, entertainment, research, technology, education, politics, tourism, art, fashion, and sports.

This final project explores the best locations for a Yoga Studio throughout the city of New York. However, as with any business, opening a new studio requires serious considerations and is more complicated than it seems from the first glance. In particular, the location of the studio is one of the most important factors that will affect whether it will have success or a failure. So our project will attempt to answer the questions “Where should the investor open a Yoga Studio?”

Business Problem

The main objective of this final project is to select and analyze the best locations in the city of New York to open a new Yoga Studio. Using Data Science methodology and instruments such as Data Analysis and Visualization, we want to help to provide solutions to answer the business question: Where in the city of New York, should be the best Neighborhood to open a Yoga Studio?

Target Audience of this project

This project is particularly useful to investors looking to open or invest in a Yoga Studio in the city of New York. New York is a great place to open a Yoga Studio, it is one of the most diverse city in the world, as every big city stress is a main topic and yoga is the perfect fighter against it. With its diverse culture, comes the aim to learn new things to improve overall quality of life.

Data

To solve the analysis, we will use the following data:

- New York City data containing the neighborhoods and boroughs.
- Latitude and longitude coordinates of those neighborhoods. The objective is to plot the map and get the venue data.
- Venue data, particularly data related to other yoga studios. We are going to use this data to perform further analysis of the neighborhoods.

Tools:

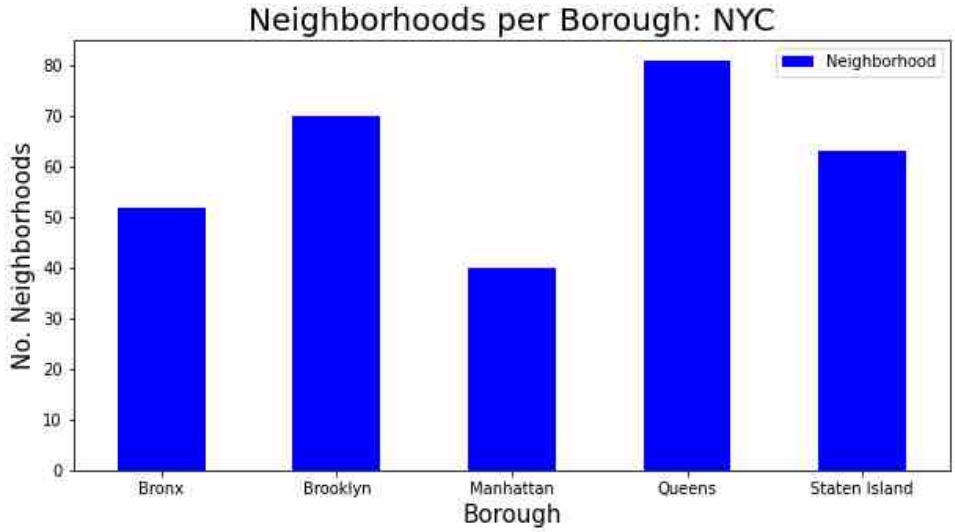
This project will require using of many data science skills, from web scrapping (open source dataset), working with API (Foursquare), data cleaning, data wrangling, to map visualization (Folium). In the next Methodology section, we will discuss and describe any exploratory data analysis that we did, any inferential statistical testing that we performed, and what machine learning techniques were used.

Methodology

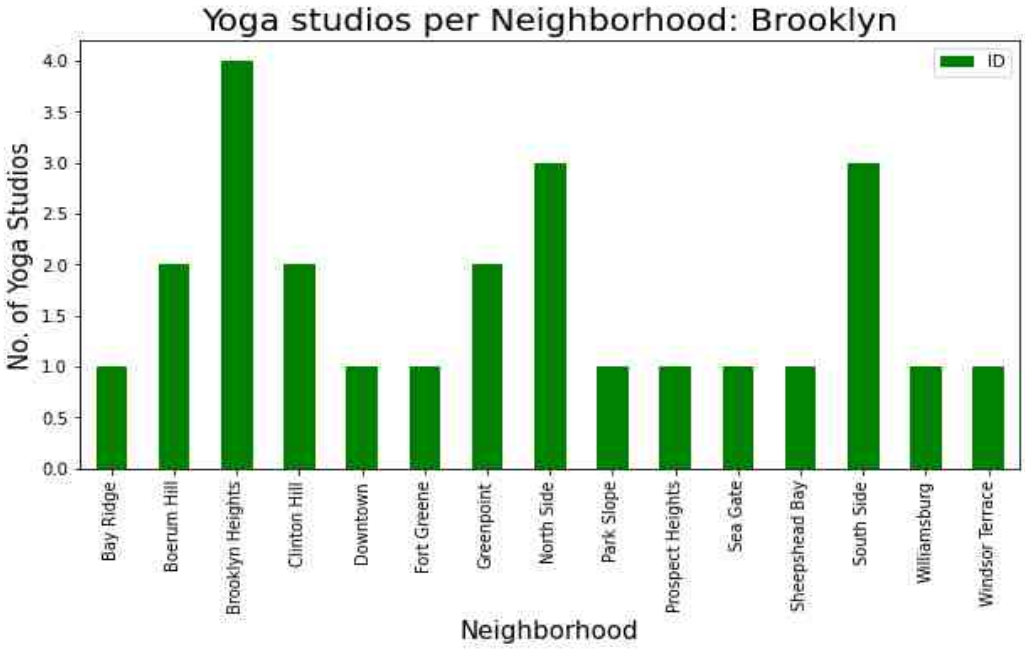
- Data will be collected from https://cocl.us/new_york_dataset and cleaned and processed into a dataframe.
- FourSquare be used to locate all venues and then filtered by Yoga Studios. Ratings, tips, and likes by users will be counted and added to the dataframe.
- Data will be sorted based on rankings.
- Finally, the data be will be visually assessed using graphing from Python libraries.

Results

We can see the Borough with their number of neighborhoods:



We will focus our research in the Borough of Brooklyn, as due his alternative style it could be a better fit for a new studio.



Downtown and Greenpoint have the best average rating Yoga Studios.

	Neighborhood	Average Rating
3	Downtown	9.000000
5	Greenpoint	9.000000
7	Park Slope	8.800000
11	Williamsburg	8.800000
0	Boerum Hill	8.750000
6	North Side	8.700000
4	Fort Greene	8.400000
10	South Side	8.366667
1	Brooklyn Heights	8.250000
8	Prospect Heights	8.200000
2	Clinton Hill	8.000000
9	Sheepshead Bay	7.400000

Discussion

Based in the results of our analysis we should check the neighborhoods with less number of studios and the ones with the lowest average ratings.

Conclusion

Following the discussion we will suggest to check neighborhoods as Sheepshead Bay, Clinton Hill, as the ratings of the yoga studios are not the highest and the quantity of yoga studios is low.

This studio could help investors as a first starting point, we could upgrade our analysis taking in account rents costs, type of people living in the neighborhoods and so on.

A more comprehensive analysis and future work would need to incorporate data from other external databases.