Coursera Capstone – Battle of Neighborhoods, New York – Final Presentation

Opening a Pattiserie/Bakery in New York

By

Preetha Venkatasamy

January, 2021

Business Problem

This project aims to find the best locations in New York City to open a Bakery/Patisserie that would attract the culturally diverse population. It should have

1) great visibility,
2) should be easy to find and
3) should attract enough initial customer interest

With the help of Data Science methodology and tools, this project aims at addressing the business problem by helping the investor find the best locations in New York City, to start the business keeping in mind all the essential factors.

Data

Data Required

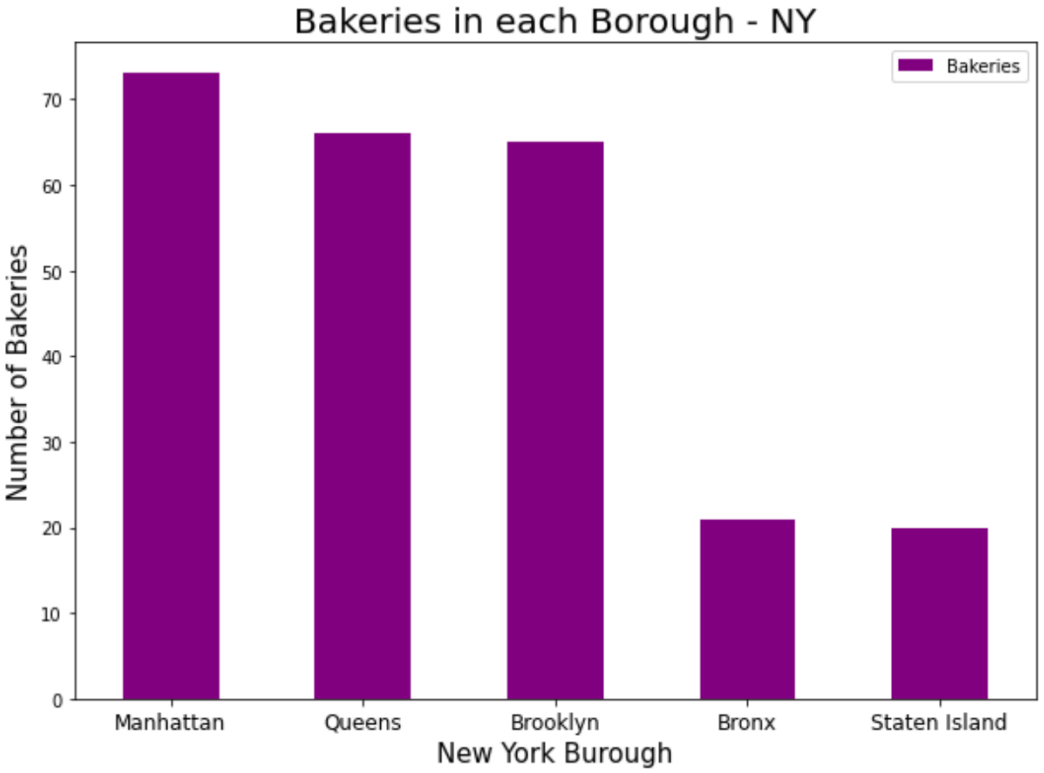
1) List of Boroughs and Neighborhoods on New York
2) Latitude and longitude coordinates for each neighborhood
3) Venue data for the preferred Neighborhoods or Boroughs

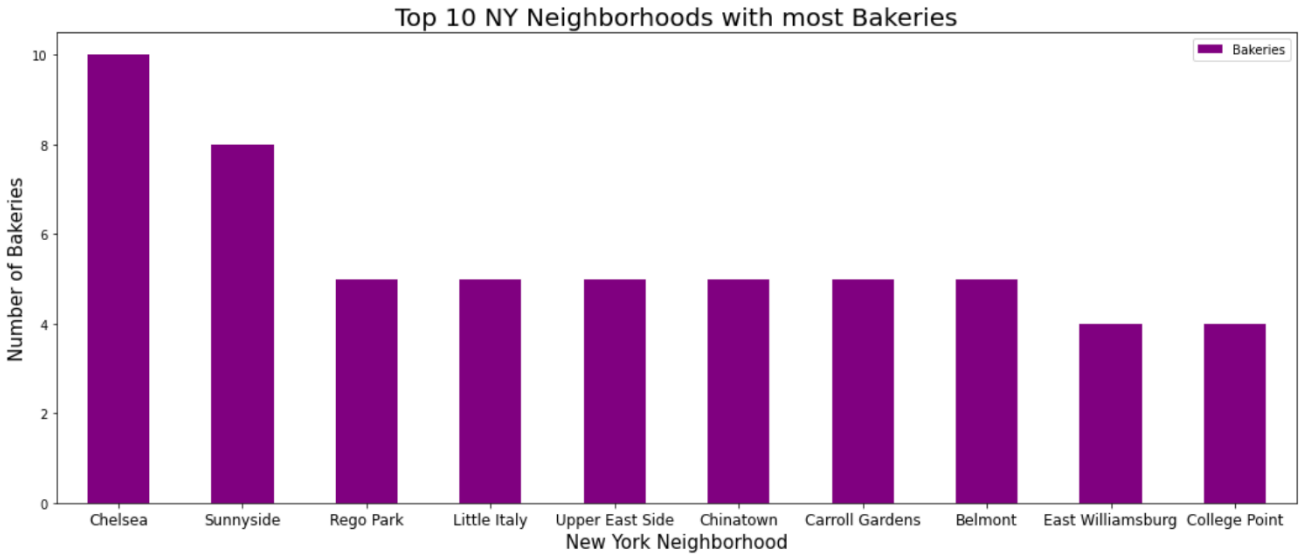
Data Source and Extraction

1) New York city data containing neighborhoods and boroughs from open data source: https://cocl.us/new_york_dataset
2) Python Geocoder package for getting latitude and longitude co-ordinates
3) Foursquare API to extract venues for preferred Neighborhoods
4) Matplot and Folium libraries for data analysis and Visualization

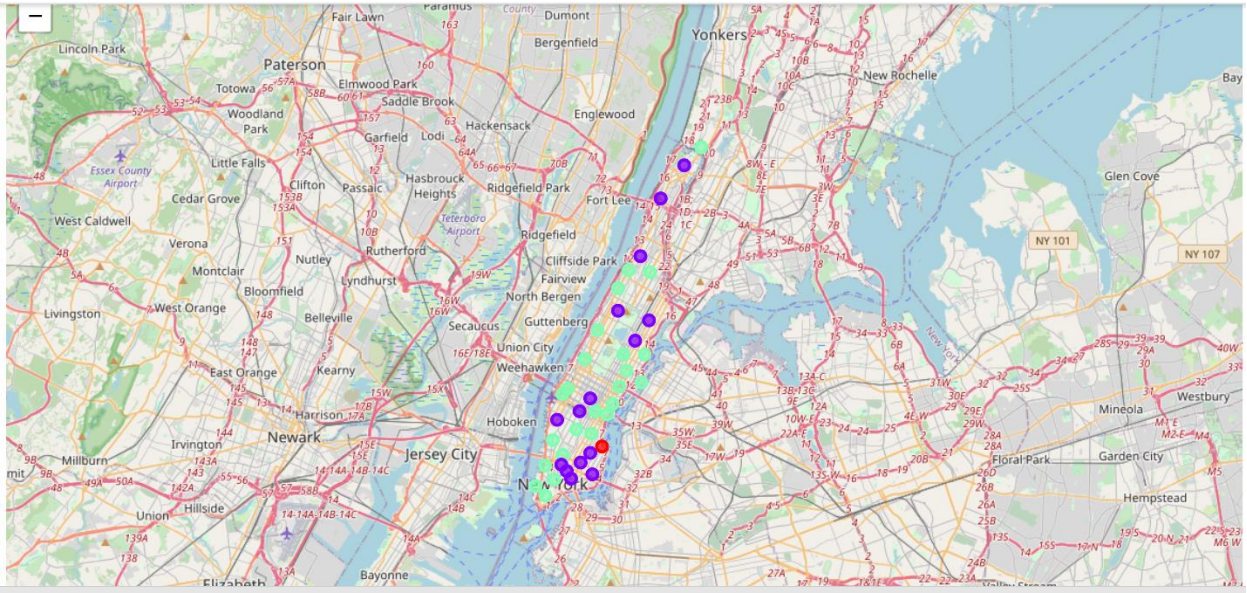
Methodology

1) Collect New York data from https://cocl.us/new_york_dataset that has all the 5 boroughs and 306 neighborhoods
2) Access any required co-ordinates using python Geocoder package
3) Once all the boroughs and neighborhoods and corresponding latitude and longitude coordinates are obtained, explore and clean the dataset.
4) Access Foursquare API to extract venues for all New York neighborhoods. Analyze the venues specific to Bakery and locate the most popular Boroughs and Neighborhoods with bakeries
5) Select the most popular Borough for further analysis and clustering using k-Means algorithm
6) Prepare data for clustering by one-hot encoding, grouping and sorting the 10 most popular venues from each neighborhood in the chosen Borough
7) Cluster the prepared data into 3 clusters and join the cluster labels the neighborhoods showing their top 10 venues
8) Explore the resulting clusters with similar characteristics and choose the neighborhoods that best answers the business questions
9) Visualize data using maps and plot as needed throughout the project


Bakeries in each Borough - NY
Bakeries
Number of Bakeries
0
10
20
30
40
50
60
70
Manhattan
Queens
Brooklyn
Bronx
Staten Island
New York Burough


Top 10 NY Neighborhoods with most Bakeries
Bakeries
Number of Bakeries
0
2
4
6
8
10
Chelsea
Sunnyside
Rego Park
Little Italy
Upper East Side
Chinatown
Carroll Gardens
Belmont
East Williamsburg
College Point
New York Neighborhood

Results



1) Cluster 1 - **NOT a preferred cluster** to open a new Bakery/Patisserie. Does not answer the business questions as compared to Cluster 2 and 3.
2) Cluster 2 -Seems like the **MOST preferred cluster** to start a new Bakery/Patisserie. It has a lot of restaurants and bakeries that are among the top most venues in these Neighborhoods.
3) Cluster 3 - Although this cluster has many restaurants among its top venues, it does not look as popular compared to cluster 2. As a result, this is **NOT a preferred cluster** to start a new Bakery/Patisserie.

Conclusion

Cluster 2 in Manhattan has the **MOST preferred neighborhoods** with many restaurants and bakeries among the top 10 venues. Starting a Patisserie here will yield great success and consistent profit due to high visibility, easy to locate and ability to attract enough initial customer interest.

As a final note, all of the above analysis is dependent on the accuracy of Foursquare data.

Future enhancements: This data and model can be used for more detailed and comprehensive analysis in the future to find **the BEST neighborhood in Cluster 2** to open a new Patisserie/Bakery.