Opportunities for Bakery shops in New York

Yongpeng Song

27 Jan 2021

Problem

- The purpose of this project is to select the best location for opening a pastry shop in New York.
- Use data science methods and tools to help investors who are interested in opening a dessert shop in New York find the best location to open a shop, to maximize the interests of the owners.

Data

Data requirement

- Dataset containing all the data of 5 boroughs and 306 neighbourhoods
- The latitude and longitude of each neighbourhood
- All the venue categories in each neighbourhood
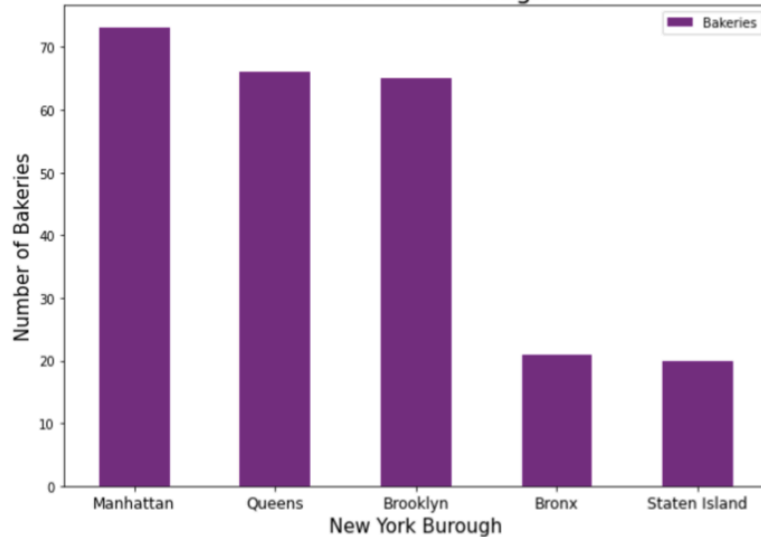
Data source and extraction

- Get data on all boroughs and neighbourhoods in New York from https://cocl.us/new_york_dataset
- Use python Geocoder to get the corresponding latitude and longitude
- Use Foursquare API and K-mean method to explore all neighbourhoods
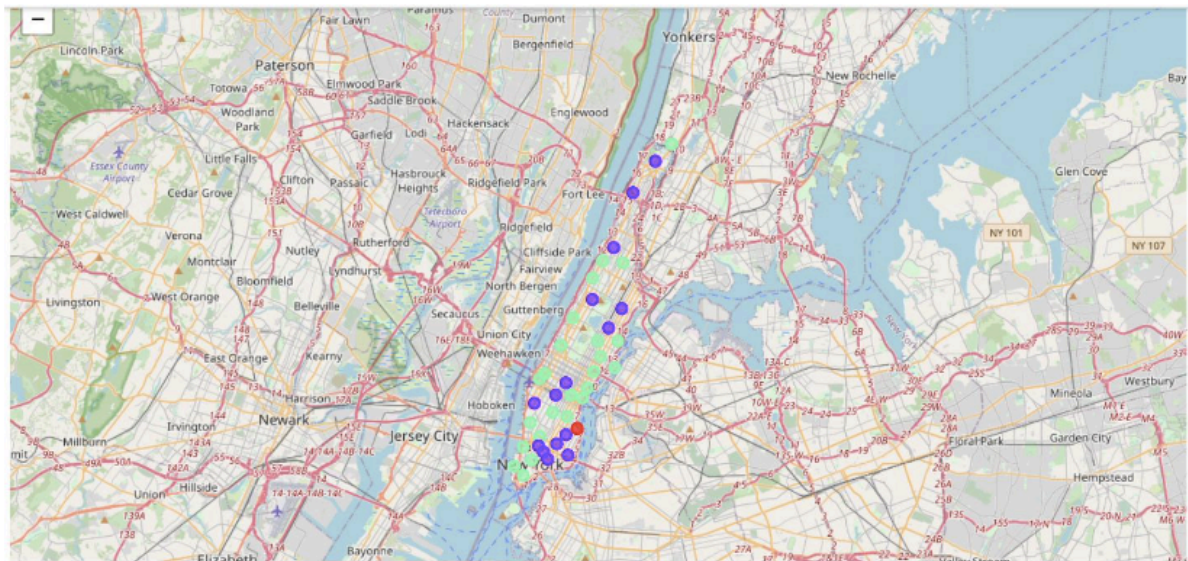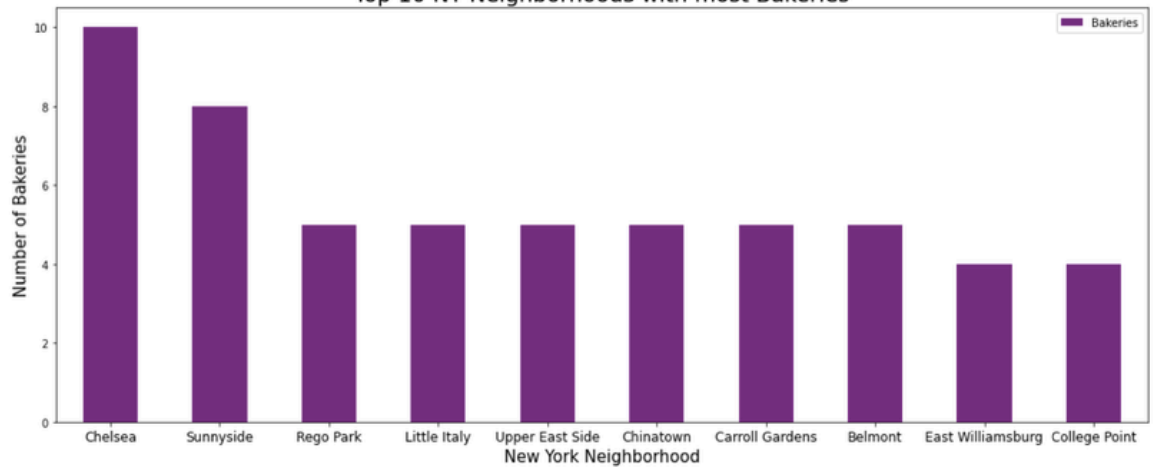- Use the Folium library to visualize

Methodology

- Get geographic information of all 5 boroughs and 306 neighbourhoods from https://cocl.us/new_york_dataset
- Obtain coordinates through the Geocoder package
- Clean datasets
- Get all the neighbourhoods' venues through Foursquare API, and then analyse the neighbourhoods with the most bakeries
- Select the highest ranked borough and use K-means for clustering
- Use one-hot encoding prepare data, and then group and sort the top 10 highest-ranked venues in each neighbourhood in the selected boroughs
- Use the K-means method to divide the prepared data into 3 clusters, and then join the resulting cluster labels with the neighbourhood's data table showing their top 10 venues
- Explore the resulting clusters with similar characteristics, and then select the most suitable neighbourhood
- Use map Visualize data

Result

Bakeries in each Borough - NY

Top 10 NY Neighborhoods with most Bakeries

- Cluster 1 (red) is not suitable for opening a baking shop
- Cluster 2 (green) is the most suitable for opening a bakery because there are many restaurants and bakery here
- Cluster 3 (purple) is not as good as cluster 2, and it is not recommended to open a bakery here

Conclusion

- In Manhattan, cluster 2 has the ideal neighbourhoods to open a bakery. Opening a bakery here can be relatively easy to obtain continuous benefits.