Opportunities for Bakery shops in New York

Yongpeng Song

26 Jan 2021

Introduction

Background

As New York is one of the international metropolises located on the east coast of the United States, there are many ethnic groups and cultures gathered here, and the cultures of all ethnic groups in the world have been further developed and sublimated here. And because New York has a unique commercial and cultural atmosphere, the economy here has developed rapidly, and it has become one of the few important economic centres in the United States.

The bakery is an indispensable store in every community. People can buy their favourite pastries and order their desserts and drinks here every day and have a good time with their friends and family. And because the cost of the bakery shop is not very high, the return on investment can be realized in a short time, which makes investing in a bakery shop a good business. If you want to achieve rapid profitability at a low cost, bakery shops have become a good choice today.

Problem

The purpose of this project is to select the best location for opening a pastry shop in New York. For a retail store, the choice of location is inseparable from the success of the business. For this project, the purpose is to 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.

Interest

The most interested in this project will be small business owners who are interested in opening a dessert shop in New York. After they understand the project, they will have a better understanding and understanding of the location of the dessert shop. At the same time, this project can also be used by dessert lovers, who can learn where to find their favourite desserts.

Data acquisition and Cleaning

Data requirement

In this project, we need the following data

- We need a data set 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

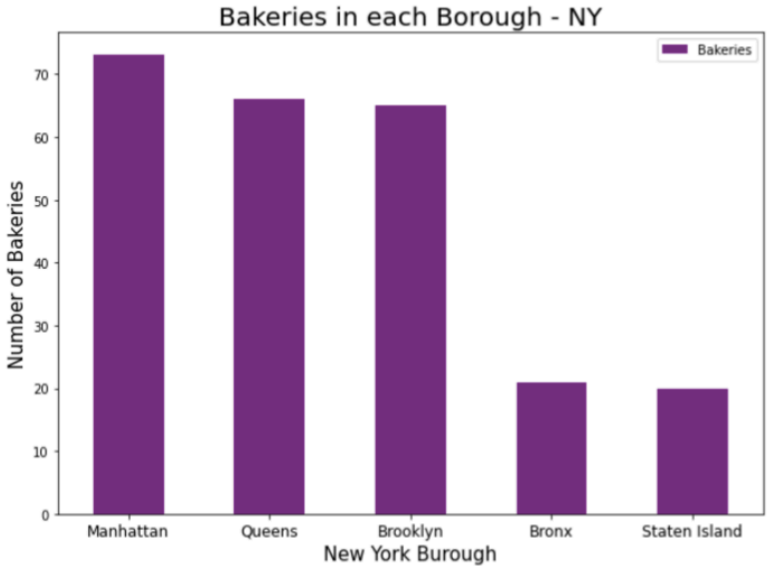
Get data on all boroughs and neighbourhoods in New York from https://cocl.us/new_york_dataset. Then use python Geocoder to get the corresponding latitude and longitude.
Then use Foursquare API to explore all neighbourhoods, to find the venues related to the bakeries in each neighbourhood, and then use these features to cluster neighbourhoods. In clustering, we use the K-mean method to operate.
Finally, we use the Folium library to visualize.

Methodology

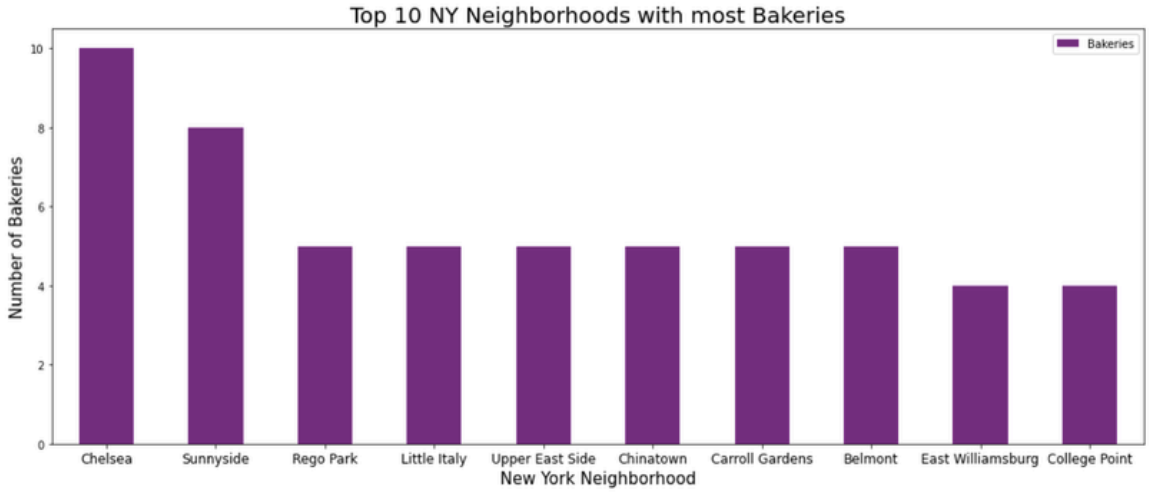
- Get geographic information of all 5 boroughs and 306 neighbourhoods from https://cocl.us/new_york_dataset
- Obtain coordinates through the Geocoder package
- Clean datasets
- Get all the neighbourhoods' venues through Foursquare API, and then analyse the neighbourhoods with the most bakeries
- Select 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

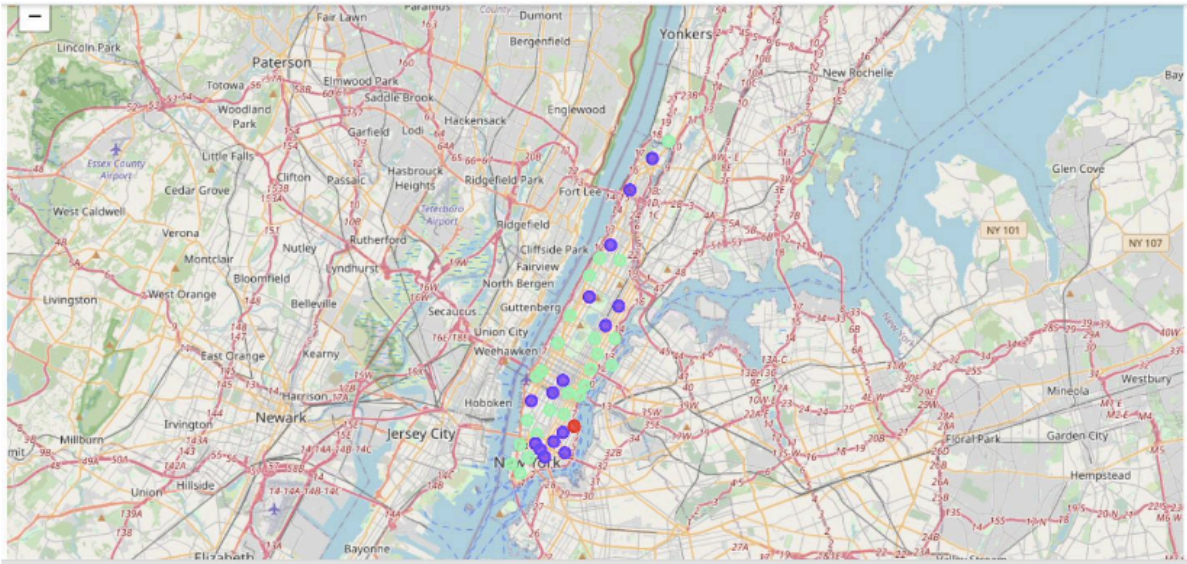
- According to the picture below, Manhattan has the most bakeries, so Manhattan is the most suitable borough



- Chelsea has the most bakery related venues



- Based on the above, it can be inferred that Manhattan is the most suitable for opening a bakery. Then use K-means clustering to find similar neighbourhoods to open a bakery shop. This neighbourhood must satisfy the interest of attracting enough customers and be easy to find
- Results of clusters include 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



Further Discussion

This model can be used as a basis. In the future, we will further analyse the neighbourhood that is most suitable for opening a bakery in Cluster 2 on this model.

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.