

Opening a New Hotel in Manhattan

Peer-graded Assignment: Capstone Project - The Battle of
Neighborhoods (Week 2)

Table of contents

- Introduction
- About Data
- Business Problem
- Methodology
- Results and Discussion
- Conclusion
- References
- Appendix

Introduction

- For many tourists, traveling is a good way to relax. They can go shopping in local malls, admire cultural relics in museums, and enjoy the local cuisine. The real estate industry also uses this trend to build more hotels in the city to cater to the needs of travelers. As a result, a large number of hotels were built in Manhattan. Of course, there are many commercial factors to consider when building a hotel. One of the most important factors is the geographic location.

Bussiness Problem

- The objective of this capstone project is to analyze and select the best locations in Manhattan to open a new hotel mainly for travelers. The project aims to provide solutions to answer the business question: In New York City if a property developer wants to open a new hotel which mainly provides services to travelers, where would you recommend that they open it?

Interested Audience

- I think that property developers and tourists are interesting audiences. On the one hand, property developers can get more profits from where to build hotels. On the other hand, tourists can also know where to live when they visit Manhattan.

About Data

- To solve the problem, we will need the following data:
 - List of Boroughs and neighborhoods of Manhattan with their geodata (latitude and longitude)
 - List of Subway metro stations in Manhattan with their address location
 - Venues data, particularly data related to travelers, for each Manhattan neighborhood (that can be clustered)
- Data source:
 - https://geo.nyu.edu/catalog/nyu_2451_34572 Foursquare API

How data will be used to solve the problem

- 1. The link above contains a list of neighborhoods in Manhattan, with a total of 5 boroughs and 306 neighborhoods. We will download it from that link.
- 2. After that, we will use Foursquare API to get the venue data for those boroughs and neighborhoods.
- 3. We need to select our interested information from those data, because the API will provide a lot of categories of data.
- 4. Then, data cleaning, data wrangling will be done by using multiple machine learning skills in this process.
- 5. Lastly, we will use K-means clustering and data visualization skills to show the result.

Methodology

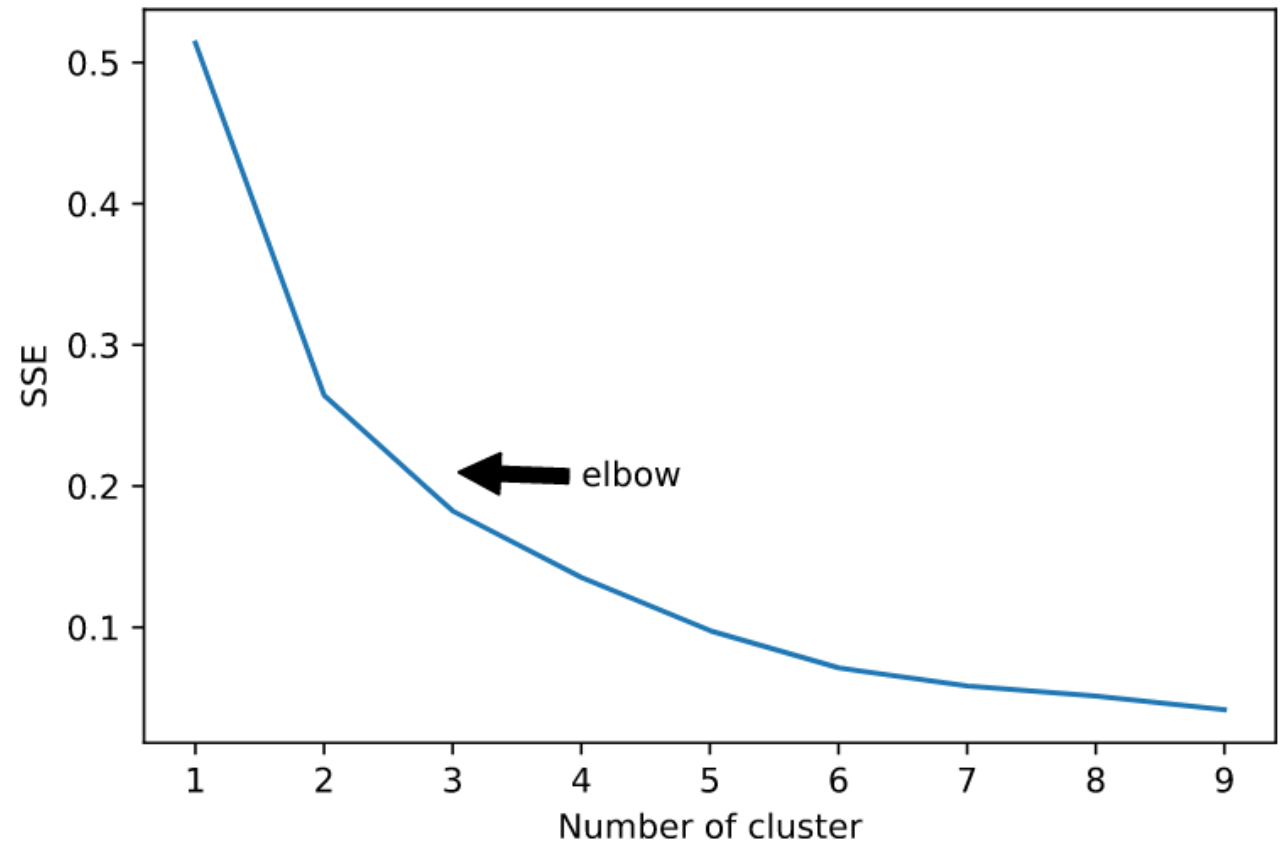
- Firstly, we need to get the list of neighborhoods in Manhattan. In this project, we use the data from this link (https://geo.nyu.edu/catalog/nyu_2451_34572). After gathering the data, we will populate the data into a pandas data frame and then visualize the neighborhoods in a map by using the Folium package. This allows us to perform a sanity check to make sure that the geographical coordinates data are correctly plotted in the Manhattan.

Methodology

- Next, we will use the Foursquare API to get the top 200 venues that are within a radius of 1000 meters. We use the Foursquare Developer Account, which registered in the previous section, to obtain the Foursquare ID and secret key. We then make API calls to Foursquare passing the geographical coordinates of our Manhattan neighborhoods. Foursquare will return the venue data in JSON format and we will extract the venue name, venue category, venue latitude, and longitude. With the data, we can check how many venues were returned for each neighborhood and examine how many unique categories are needed. And we found that there are about 300 unique categories in those venue data. However, some of them are very similar, for example, "Chinese Restaurant and India Restaurant". So I combine those kinds of categories into one, like "Restaurant". Also, tourists may go to a lot of attractions, but these types are very scattered, I also put these into one category. Categories like "Art Gallery", "Theater", "Museum", "Park", "Historical Site", "Bar", "Stadium" are combined into one category names "Sites". Then, we will analyze each neighborhood by grouping the rows by neighborhood and taking the mean of the frequency of occurrence of each venue category. Since we are analyzing the "Hotel" for visitors, we will filter the "Restaurant", "Hotel", "Sites" as the venue category for the neighborhoods.

Methodology

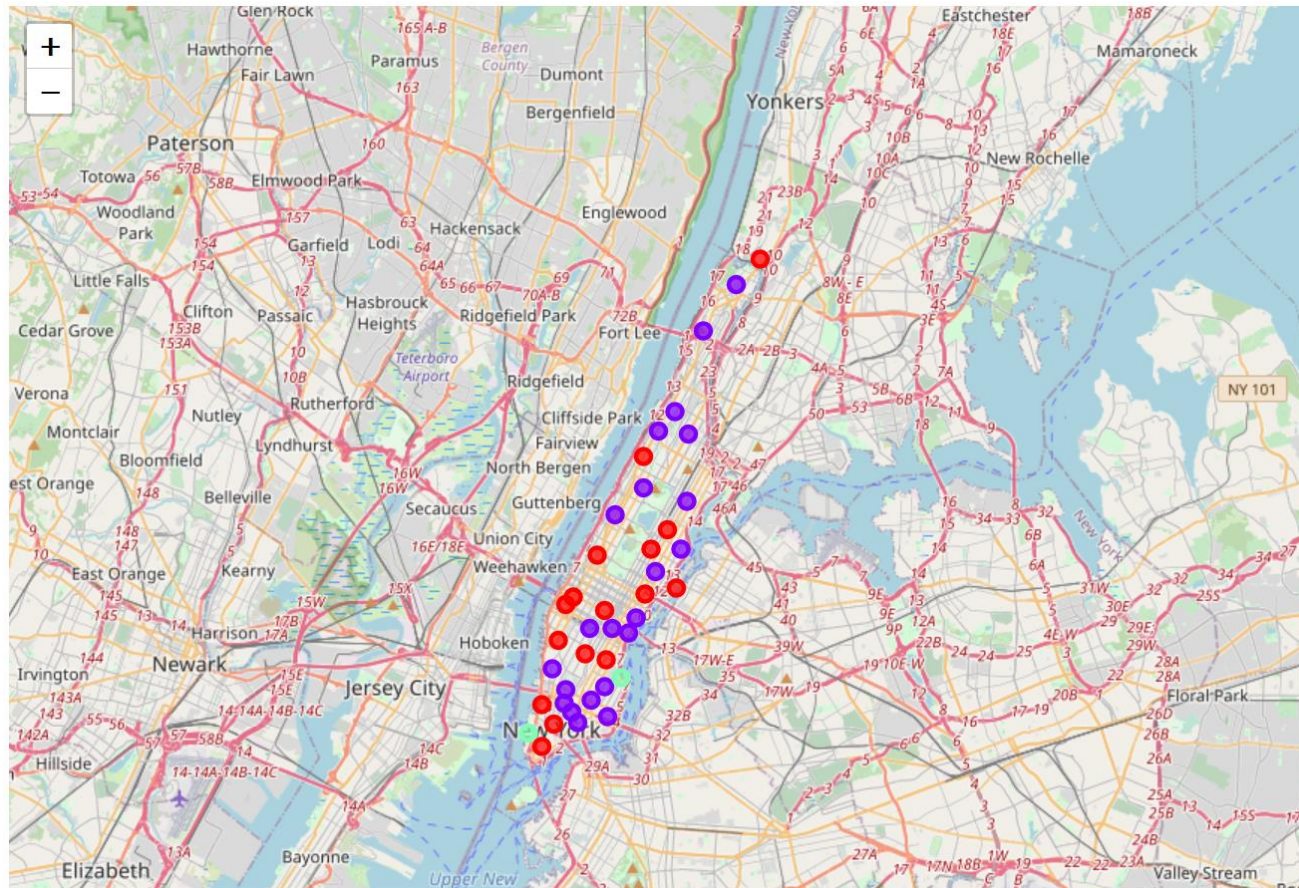
- Lastly, we will perform a cluster on the data by using k-means clustering. We firstly for loop from 0 to 10 to find the proper number of clusters. And base the elbow point we found, we set the three to the cluster number. The results will allow us to identify which cluster has a high concentration of hotels while which neighborhoods have a fewer number of hotels with plenty of "sites". Based on the occurrence of hotels, restaurants, and sites, it will help us to answer the question as to which neighborhoods are most suitable to open new hotels.



Results and Discussion

- According to the results of clustering, we can see. First of all, the third category is definitely not our goal, because the number of restaurants and hotels is very small. Next, let's look at the first and second categories. In the first category, we find that restaurants, hotels, and tourist attractions are relatively evenly distributed, but this shows that this area has undergone some development, and competition may be fierce. The second type of area also has many equity and tourist attractions, but the number of hotels is relatively small. I think it is a good location to open a new hotel.

Results and Discussion



- Red :cluster 1
- Purple: cluster 2
- Green: cluster 3

Conclusion

- Therefore, this project recommends property developers to capitalize on these findings to open a new hotel in neighborhoods in cluster 1 with little to no competition. Property developers with unique selling propositions to stand out from the competition can also open new hotels in neighborhoods in cluster 0 with moderate competition. Lastly, property developers are advised to avoid neighborhoods in cluster 2 that lack of enough catering services.