Introduction

<u>Problem Statement</u>

In this project, we are required to pick the best location to open a new coffee shop among all the communities in both New York and Paris. We determine the best location using the following criteria -

1. The coffee shop needs to operate in a location without too much competition.
2. The owner of the shop needs to maximize profit and hence wants places with maximum possible population of the results from condition 1.

To solve this problem, we would need to first get information about the neighborhoods, populations, and locations (longitude and latitude) of both New York and Paris.

We would then collect information about the number of coffee shops in each of the neighborhoods using the foursquare API and the coordinates collected initially, while also calculating the population of the neighborhood is being served by each coffee shop.

Finally, we compare the maximum possible profit for each neighborhood and decide to open the coffee shop in the one promising the maximum profit.

For this project, we assume both neighborhoods are similar in terms of spending power and other necessary demographics. We also assume the population listed in the table all get the same amount of coffee daily.

This project would prove useful in choosing the area to situate a business venture.

Data Description

The data sources used in implementing this project are:

1. Neighborhood data for New York scraped from Wikipedia
2. Neighborhood data for Paris scraped from Wikipedia

The data scraped from Wikipedia on the New York neighborhoods contain columns describing the following.

1. The name of the neighborhood
2. The population of the neighborhood
3. The area in Km2 of the neighborhood
4. The community board to which the neighborhood belongs

The data scraped from Wikipedia on the Paris neighborhoods contain columns describing the following.

1. The name of the neighborhood
2. The population of the neighborhood
3. The area in Km2 of the neighborhood

4. The Arrondisement to which the neighborhood belongs

Other columns which are then added to the collected data for each of the neighborhoods include

1. The latitude and longitude of each neighborhood
2. The number of coffee shops open in each neighborhood
3. The number of people served per coffee shop in each neighborhood assuming each coffee shop has an equal chance of serving the same number of people.

A visualization of the number of people potentially served by each coffeeshop shows the potential for profit assuming the coffee shops all sell coffee at the same price.

A higher number of people served per coffeeshop means higher earning potential if we situate our coffeeshop in that neighborhood, hence encouraging the investor to open the coffeeshop in that neighborhood.

Methodology

The task at hand is one of choosing the best location to situate a coffee shop given the data regarding all the neighborhoods in New York and Paris, herein referred to as locations.

We pull data regarding the name, population, and area of the neighborhoods in both locations by scraping publicly available data from Wikipedia. We accomplish this by parsing the data available on the page using

- Pandas library
- BeautifulSoup from the bs4 library.

Next, we collect information about the longitude and latitude of each of the neighborhoods using

- Nominatim from the geopy library

We view the location of each of the neighborhoods on a map to give us initial information about each of the neighborhoods using

- Folium library

After visualizing the neighborhoods, we also gather information about the coffeeshops present in each of those locations. To achieve this, we use

- The Foursquare API

Finally, we calculate the number of people served by each coffeeshop and use this in our analysis to determine the location with the highest ceiling for profit.

We visualize the top five(5) results after sorting the neighborhoods by the possible sales of coffee to be made in descending order using

- The Matplotlib library

Results

Our results show that the top five locations to situate a coffee shop in Paris are

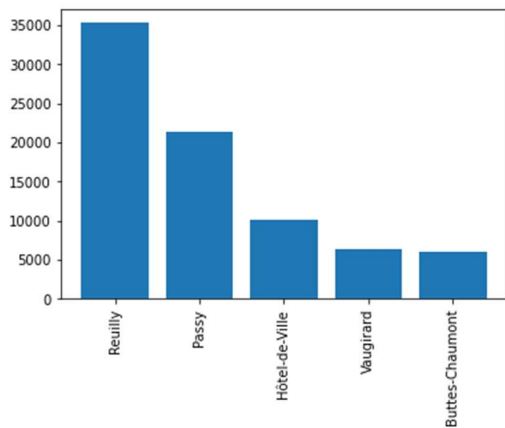| Neighborhood/Arrondisement	Population per coffee shop
Reuilly	35321
Passy	21357
Hôtel-de-Ville	10019
Vaugirard	6356
Buttes-Chaumont	6066

Figure 1: Bar chart of Population served per coffee shop in each neighborhood in Paris

Our results also show that the top five locations to situate a coffee shop in New York are

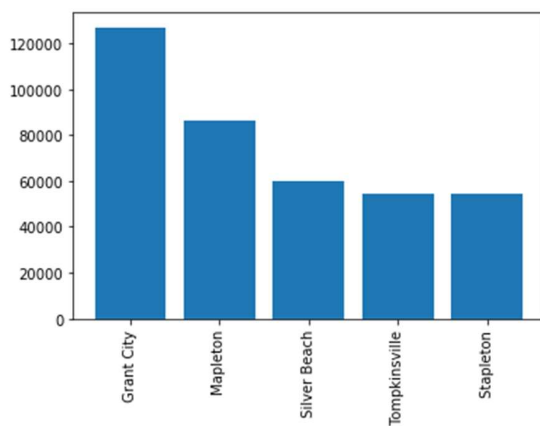| Neighborhood/Arrondisement	Population per coffee shop
Grant City	127071
Mapleton	86064
Silver Beach	60196
Tompkinsville	54203
Stapleton	54203

Figure 2: Bar chart of Population served per coffee shop in each neighborhood in New York

Discussion

The results we got in the previous section show that there is a far higher ceiling for sales in New York with the best neighborhood being Grant City with a population per coffee shop of 120, 000 compared to the maximum population per coffee shop of 35,000 in Reuilly, Paris.

Conclusion

The results of this analysis clearly point towards situating the coffee shop in Grant City, New York to maximize the potential profits.