CAPSTONE PROJECT

IBM DATA SCIENCE PROFESSIONAL CERTIFICATE

IDENTIFYING AN IDEAL LOCATION FOR A SUPPLEMENT SHOP IN MANHATTAN

AGENDA

- Introduction
- Business Problem/Problem Statement
- Data
- Methodology
- Results
- Discussion
- Conclusion

INTRODUCTION

- An individual or a partnership is interested to setup a Supplement Shop in the neighborhoods of Manhattan.

- Manhattan is often described as the cultural, financial, media, and entertainment capital of the world.

- To setup and run a business entity in such location requires a great deal of consideration and deliberation.

- One way to identify an ideal location is by analyzing the presence and popularity of such business ventures in the proposed area versus the demand.

Contoso

BUSINESS PROBLEM/PROBLEM STATEMENT

Identify a location for setting up a new Supplement Store in one of the neighborhoods in Manhattan, which may help to maximize the sales.

Contoso

DATA

To address the business problem we need the following data:

• Name of all the neighborhoods in Manhattan along with the geographical coordinates (latitude and longitude). (**https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkills Network-DS0701EN-SkillsNetwork/labs/newyork_data.json**)

• List of popular venues in each of these neighborhoods obtained using Foursquare API.


Contoso

NEIGHBOURHOODS OF MANHATTAN
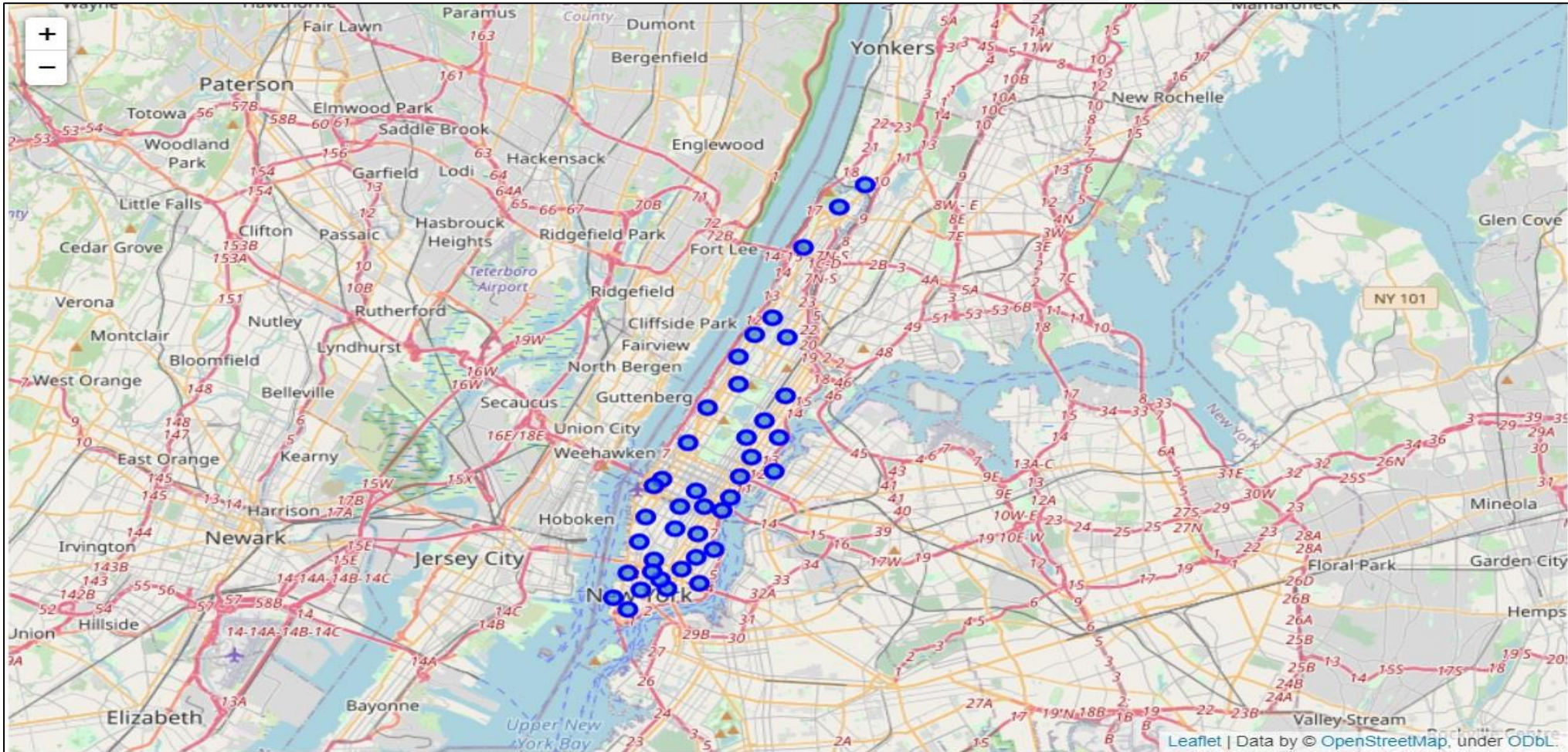
METHODOLOGY

- Identify the popular venues in each neighborhood using Foursquare API

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
1	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
2	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	Diner
3	Marble Hill	40.876551	-73.91066	Starbucks	40.877531	-73.905582	Coffee Shop
4	Marble Hill	40.876551	-73.91066	Dunkin'	40.877136	-73.906666	Donut Shop

- Consolidate the data on the basis of neighborhood. Based on the frequency of occurrence, top 10 venues were selected for each neighborhood.
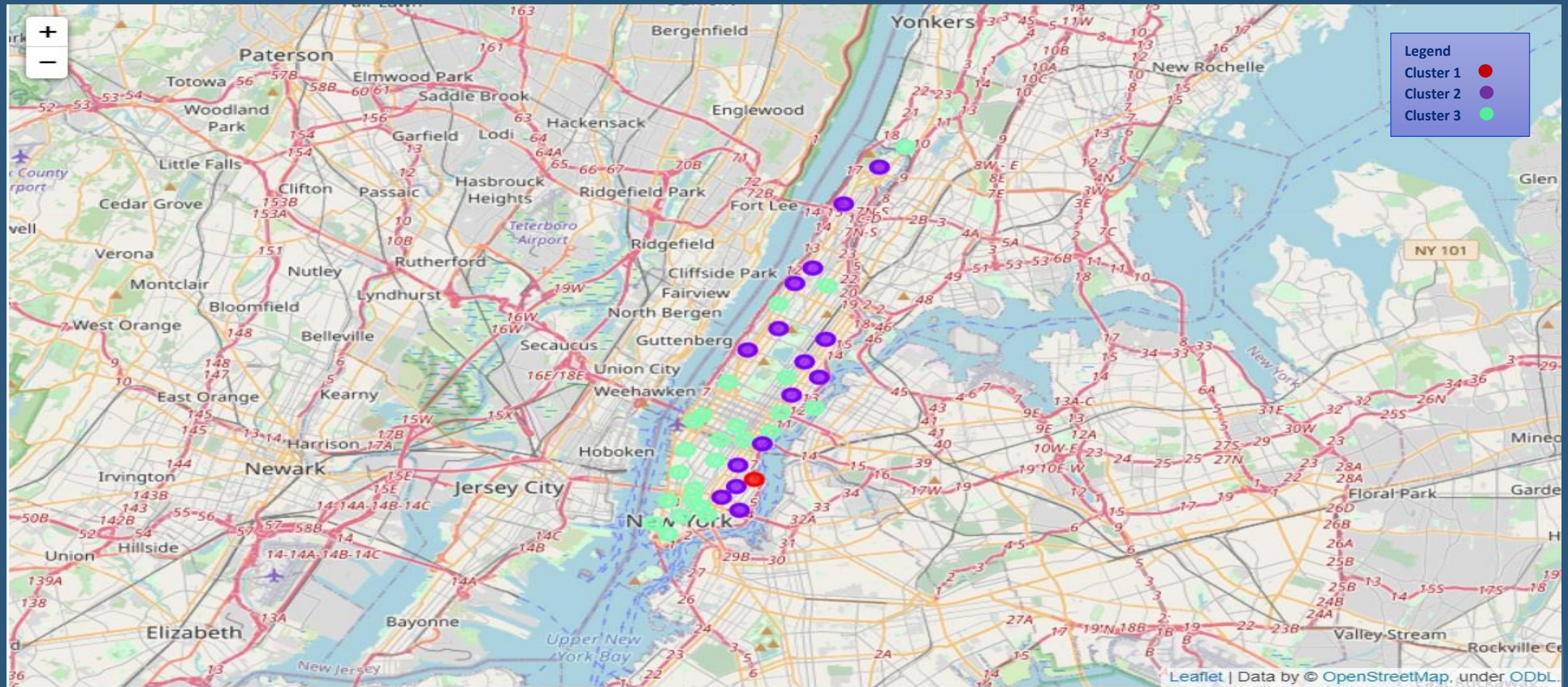
Contoso

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Battery Park City	Coffee Shop	Hotel	Park	Memorial Site	Gym	Clothing Store	Burger Joint	Gourmet Shop	BBQ Joint	Plaza
1	Carnegie Hill	Coffee Shop	Café	Yoga Studio	Wine Shop	Bookstore	Pizza Place	Gym	Cosmetics Shop	Japanese Restaurant	Cocktail Bar
2	Central Harlem	African Restaurant	Seafood Restaurant	Art Gallery	Bar	French Restaurant	Chinese Restaurant	Cosmetics Shop	Gym / Fitness Center	American Restaurant	Park
3	Chelsea	Coffee Shop	Bakery	American Restaurant	French Restaurant	Art Gallery	Ice Cream Shop	Wine Shop	Seafood Restaurant	Cupcake Shop	Bookstore
4	Chinatown	Chinese Restaurant	Bakery	Cocktail Bar	Hotpot Restaurant	Salon / Barbershop	Spa	American Restaurant	Optical Shop	Dessert Shop	Ice Cream Shop

- The above dataset was used for segmenting and clustering the neighborhoods of Manhattan into three groups using k-means clustering algorithm.

- The value of k was determined by optimizing the value of Squared Error or Cost function.
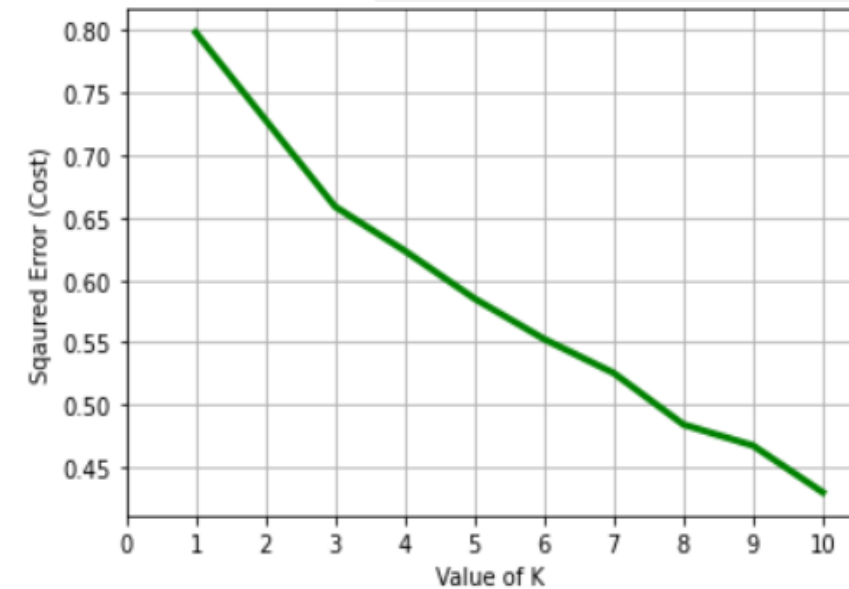
RESULTS

Clusters obtained after performing K-means Clustering

DISCUSSION

- Cluster 1
  - contains only one neighborhood namely Stuyvesant Town.

|	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
37	Stuyvesant Town	Park	Bar	Coffee Shop	Pet Service	Baseball Field	Gym / Fitness Center	Cocktail Bar	Heliport	Farmers Market	Harbor / Marina

- Cluster 2
  - Different kinds of restaurants and coffee shops are predominant in this neighborhood.
  - Fitness related or outdoor recreational venues are not that popular in this neighborhood.
  - Let's get the frequency distribution of this data to identify the most popular venues of this cluster

FREQUENCY DISTRIBUTION OF POPULAR VENUES IN CLUSTER 2
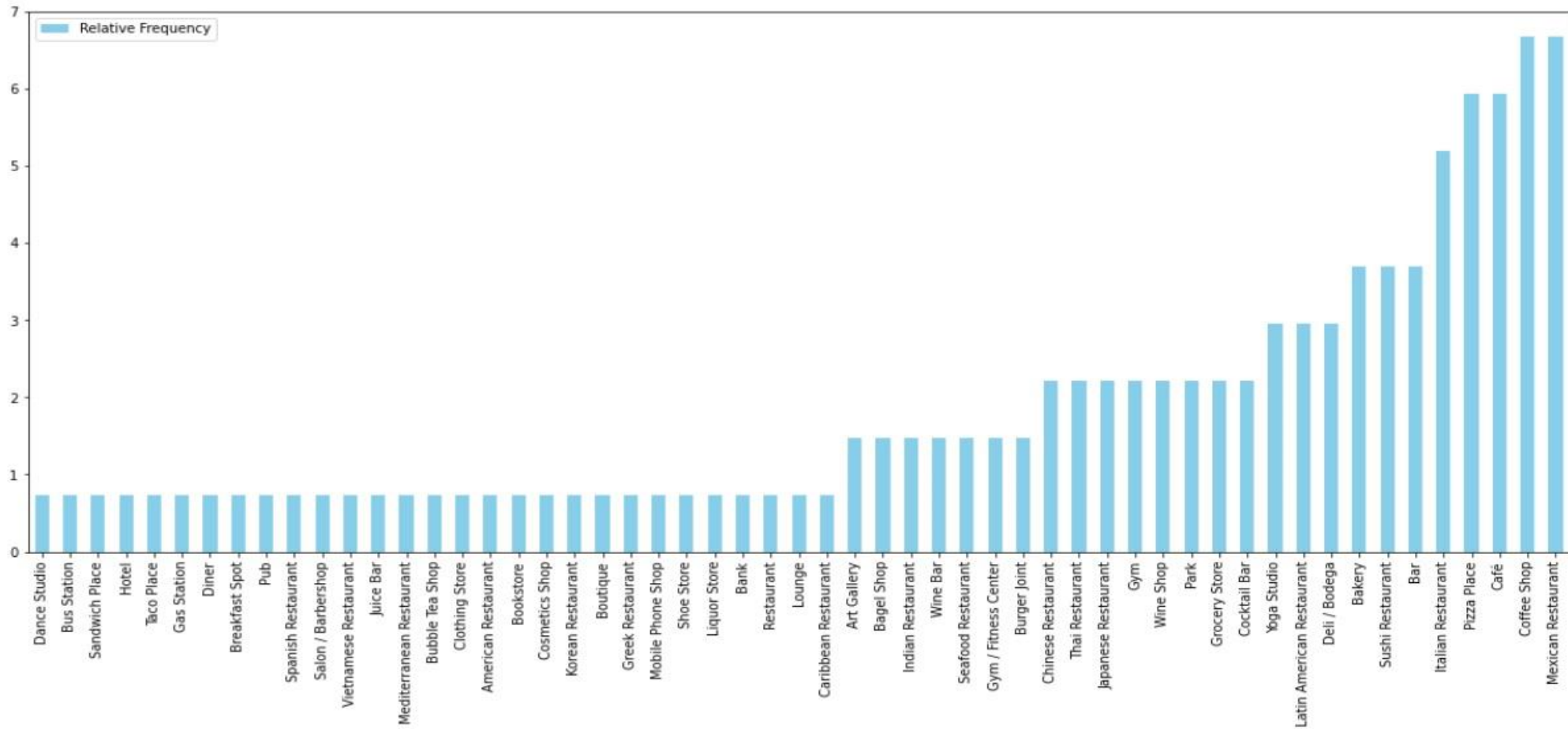
- Cluster 3
  - Fitness related or outdoor recreational venues are quite popular.
  - This cluster is a balanced mix of restaurants, fitness and recreational amenities as well commercial establishments.
  - Let's get the frequency distribution of this data to identify the most popular venues of this cluster.
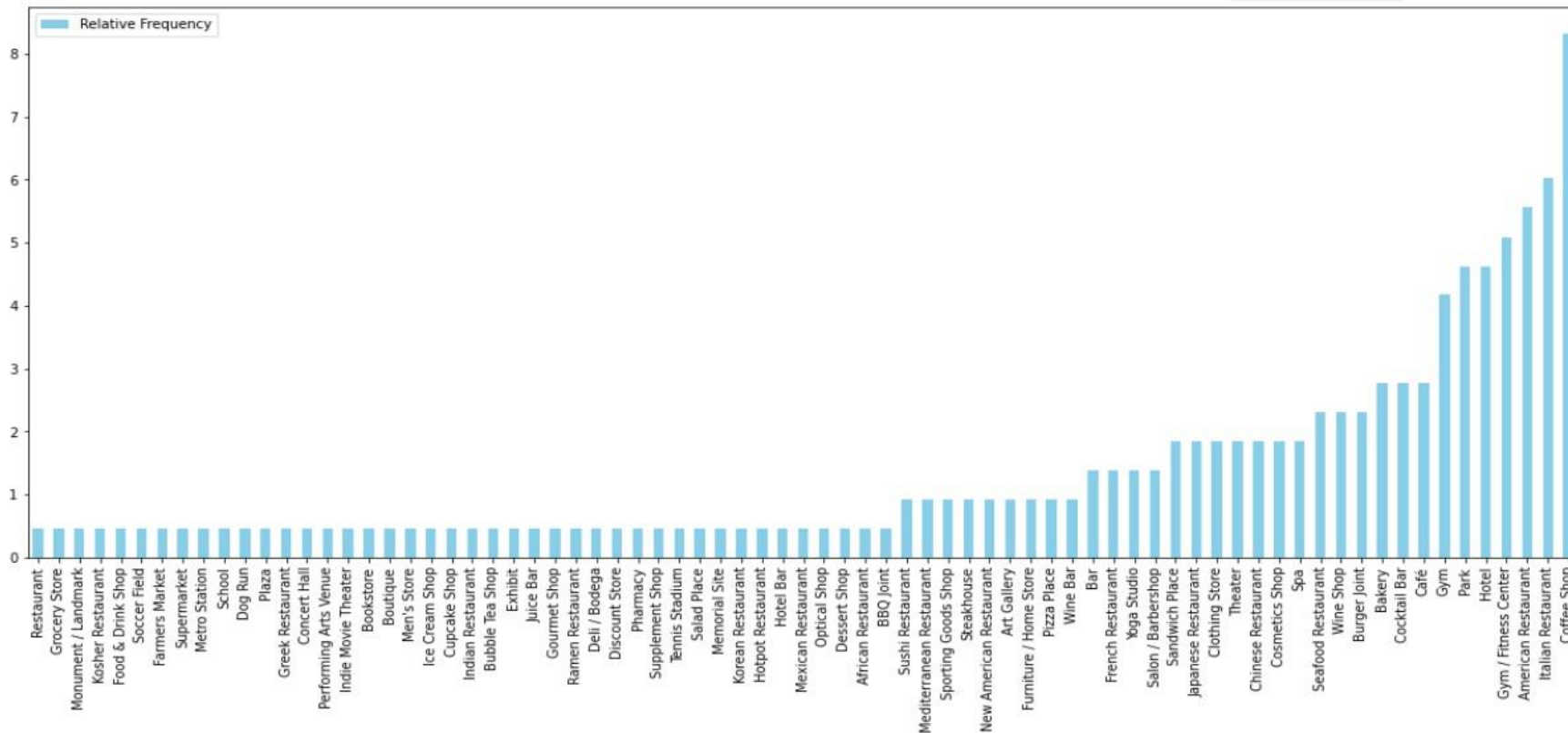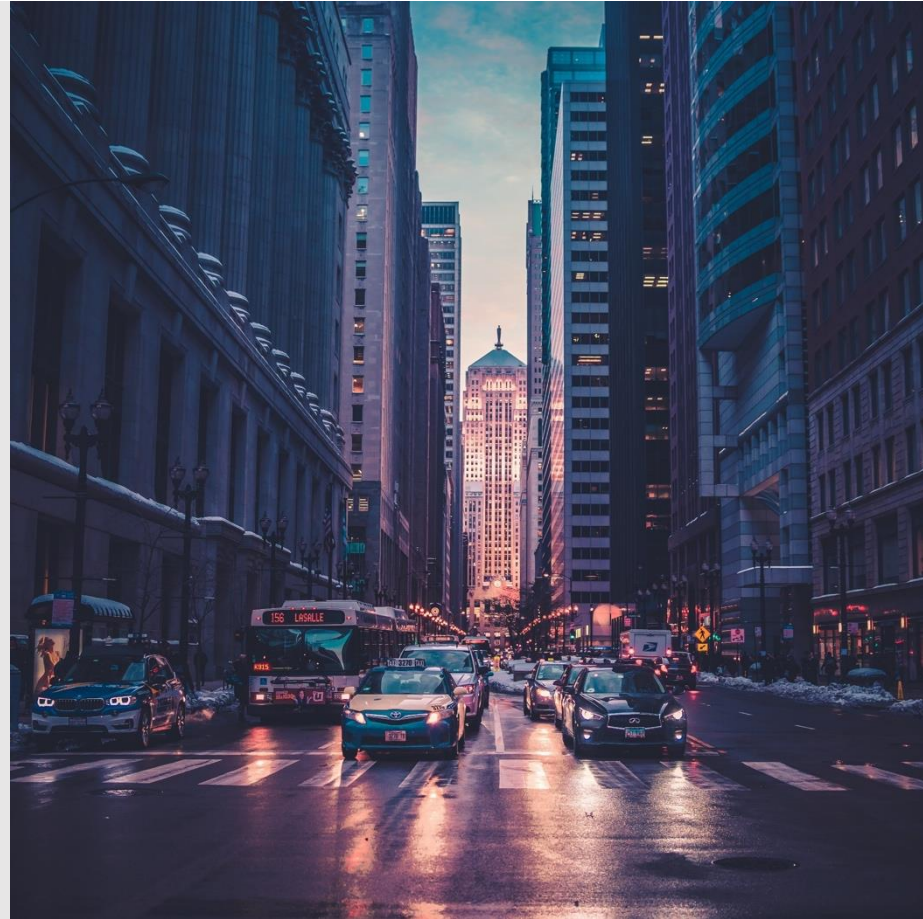
COMPARISON OF CLUSTER1 AND CLUSTER 2

CLUSTER 1

- This neighborhood has almost all sports grounds such as baseball field, tennis field etc. It also has a lot of recreational venues like parks, health and fitness related establishments such as gyms, hospitals etc.

- There are no supplement shops in this neighborhood.

CLUSTER 2

- Since fitness related venues are quite popular, we can consider the neighborhoods in this cluster as well, as a promising location.

- Please note that the neighborhood of Marble Hill already has a Supplement Store.

- However, the remaining neighborhoods may prove to be promising, but further analysis is required to pin point the neighborhoods.

CONCLUSION

- The neighborhoods of Manhattan were studied to identify a promising location for a Supplement Store.

- The defining feature data obtained from the Foursquare API was used for segmenting and clustering the neighborhoods into three groups using k-means clustering.

- Further analysis of the clusters suggested cluster 1 and cluster 3 as promising clusters.

- Among these neighborhoods Stuyvesant Town was identified to be the best fit.

THANK YOU!

✉ snehadevadas9@gmail.com

🔗 https://www.linkedin.com/in/sneha-devadas-56310b156/

Contoso