Project Title : Analysing key cities to set up a vegetable/fruit distribution centre that will cater mainly to food outlets

1. Introduction :

A1. Problem Statement

A big corporate house that is well established as a multibrand apparel retailer, is interested in diversifying its business. Its business has currently been hit by the pandemic and it is eager to diversify its business. It desires to carry out retailing of an essential commodity and has narrowed down on fruits and vegetables.

It has approached us(consultancy firm) for future market prospects, strategy and scope. It has a two year horizon(maximum) to start its new operations.

One key point highlighted by our team was that setting up retail sites for a new product line in a prime city will be expensive and highly unpredictable. There will be greater additional expenditure on advertising the brand to consumers as well as setting up multiple retail sites. Hence, at this initial stage of the venture, it is prudent to focus on business consumers. Rather than sell vegetables and fruits to the final customer, It is more feasible and viable to sell vegetables and fruits to the following set of customers in the following priority order.

1. Restaurants
2. Hotels/Motels
3. Fast Food Outlets
4. Bars and Pubs
5. Malls - Retail Stores or supermarkets

The client has corporate centres in the following cities.

1. Toronto
2. New York
3. Paris

Hence we will analyse these cities as the client can leverage their business networks wherein integration with their existing business will be easier. **To cater to these consumers the client can set up a distribution warehouse centre in one of the cities.**

A2. Data Description

To analyse the problem we will be using the following data as below:

1. City location Data:
    - Toronto – **Wikipedia data** on the postal codes and neighbourhoods in Toronto.
    - New York – **Json file from 'https://cocl.us/new_york_dataset'** which consists location data related to the neighbourhoods.
    - Paris – **Location data** for the administration zones of Paris which was sourced from www.aggdata.com
2. **Foresquare API** to capture venue details for each location coordinate in each city neighbourhood.

2. Methodology :

B1. Data Wrangling and Cleaning:

The data was obtained from the above sources, cleaned and tabulated. The following table is a summary of the data:

City	Boroughs	Neighbourhoods
Toronto	10	103
New York	5	306
Paris	NA	20*

*Data with respect to locations in the neighbourhood of Paris was not available. However data is available for 20 administration zones in Paris. We will select these zones and increase our search radius for each point in comparison to that of New York and Toronto. Another point to consider is that the area of Paris is **105 sqkm**. While that of Toronto and New York are **630 sqkm and 783 sqkm** respectively.

	Postal Code	Place Name	Latitude	Longitude
0	75001	Paris 01 Louvre	48.8592	2.3417
1	75002	Paris 02 Bourse	48.8655	2.3426
2	75003	Paris 03 Temple	48.8637	2.3615
3	75004	Paris 04 HÙtel-de-Ville	48.8601	2.3507
4	75005	Paris 05 PanthÈon	48.8448	2.3471

Paris Location Data

B2. Exploratory Analysis:

To further analyse the data we made API calls for each location coordinate and explored the neighbourhood. Different radius were considered for each city based on the number of neighbourhoods and total venue categorises.

One-Hot encoding was carried out for each city wherein the venue category was the dummy variable. **We grouped the data as per the neighbourhood and applied the mean operation across the neighbourhoods.**

Certain venue categories were eliminated based on the priority consumers as mentioned in the introduction section.

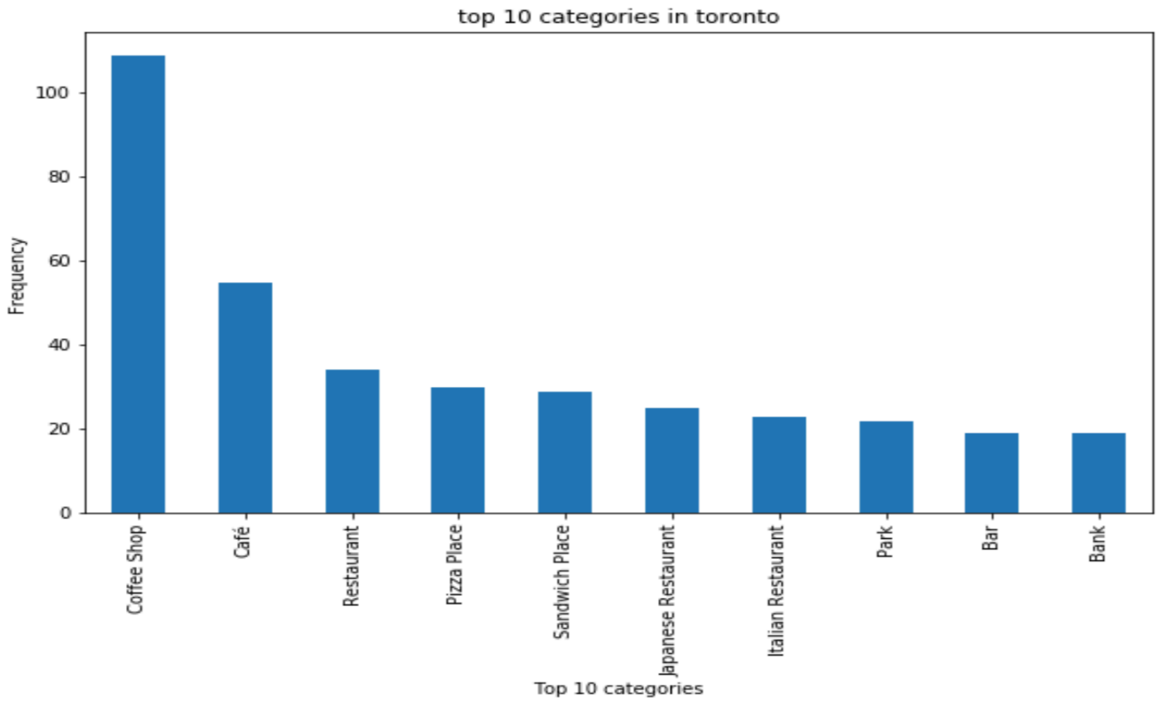
This data was tabulated for each city. The summary is as below:

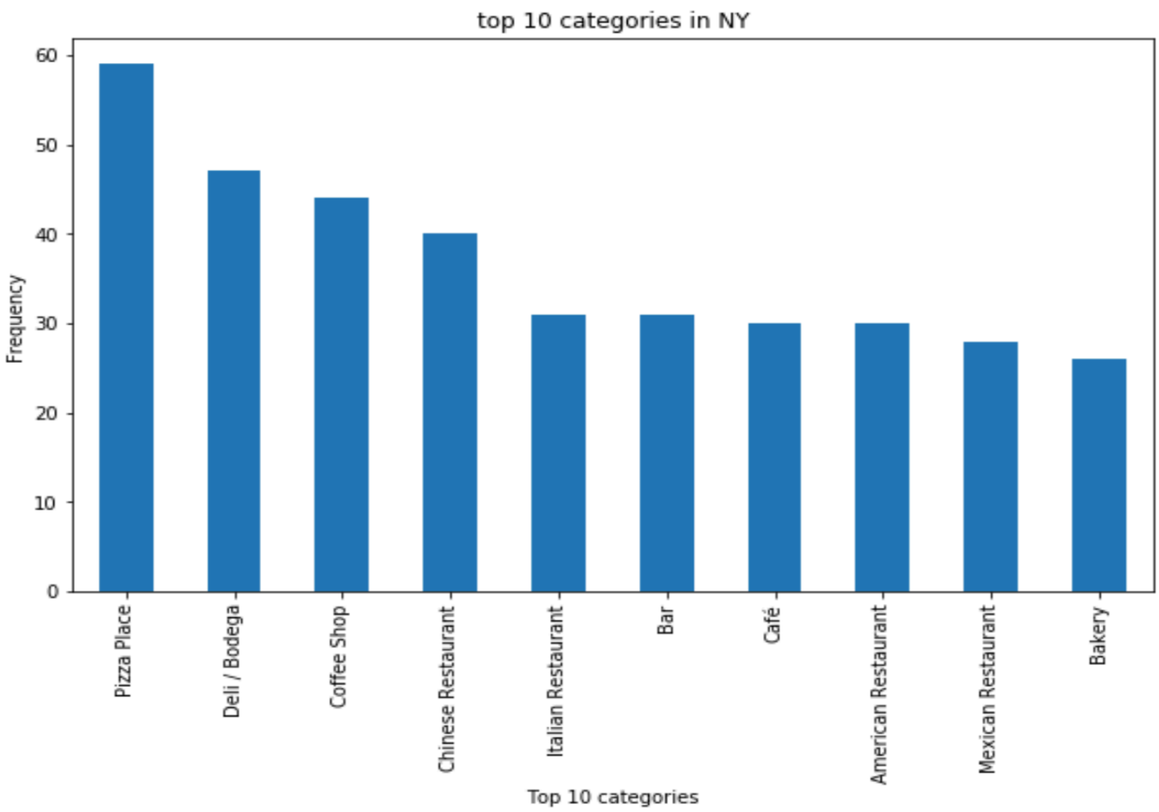
	Unique Categories	Neighborhoods/Zone	Radius for explore	Venues explored
Toronto	115	79	300	1151
New York	151	222	150	1467
Paris	134	20	500	1254

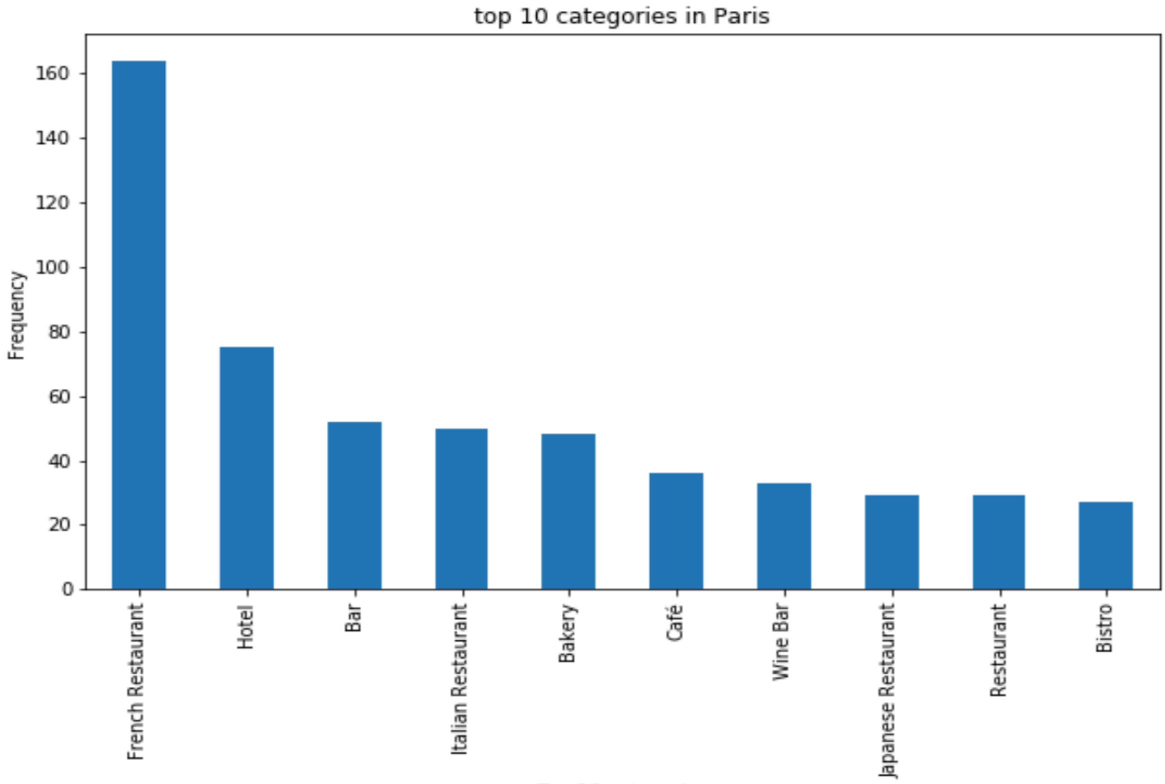
One-hot encoding summary table

As can be seen in the summary, the radius for exploring venues in each city has been selected to bring the number of unique venue categories and venues explored as close as possible for each city.

We then analysed the top 10 venue categories for each city. Here is the summary.







We can conclude the following points from the preliminary analysis of the following graph.

1. The top 10 categories in Toronto have few restaurant categories as compared to New York and Paris. Toronto is more dominated by Cafes and Coffee Shops which are not large consumers of vegetables and fruits.
2. Hence we narrow our focus on New York and Paris and carry out cluster analysis using K-means clustering for both cities.
3. However it should noted that, at this stage, Paris seems more lucrative than New York to establish a vegetable and fruit distribution warehouse. This is because there are greater number of restaurants in Paris as compared to New York. Additionally, New York has a higher proportion of coffee shop and café as compared to Paris.

B2. K-means Analysis:

To carry out cluster analysis for Paris and New York we apply the **K-mean unsupervised machine learning algorithm**. The data is unlabelled with no target and independent variables hence we rely on unsupervised machine learning to find hidden patterns in our unlabelled data.

We selected the number of clusters for each city by iterating for each cluster value and selecting the value that resulted in the most uniform distribution of neighbourhood across clusters.
We then selected two clusters for each city having the highest neighbourhoods and further analysed the data.

City	Kmeans cluster
New York	7
Paris	2

B.2.2 New York Cluster Analysis

For each cluster we analysed the top three popular venues across the neighbourhoods. **There is a table for 'most common venue' or popular venues for each cluster of neighbourhoods for a city**. Following is the summary for New York City cluster one.

	1st Most Common Venue
Women's Store	43
Deli / Bodega	9
Ice Cream Shop	7
Bar	6
Liquor Store	6

CLUSTER ONE

	2nd Most Common Venue
Food	43
Women's Store	27
Coffee Shop	7
Grocery Store	6
Deli / Bodega	6

	3rd Most Common Venue
Flea Market	46
Falafel Restaurant	15
Electronics Store	12
Women's Store	12
Pizza Place	7

B.2.3 Paris Cluster Analysis

Following is the summary Paris City cluster zero.

	1st Most Common Venue
French Restaurant	9

CLUSTER ZERO

	2nd Most Common Venue
Hotel	3
Bar	3
Italian Restaurant	1
Café	1
Wine Bar	1

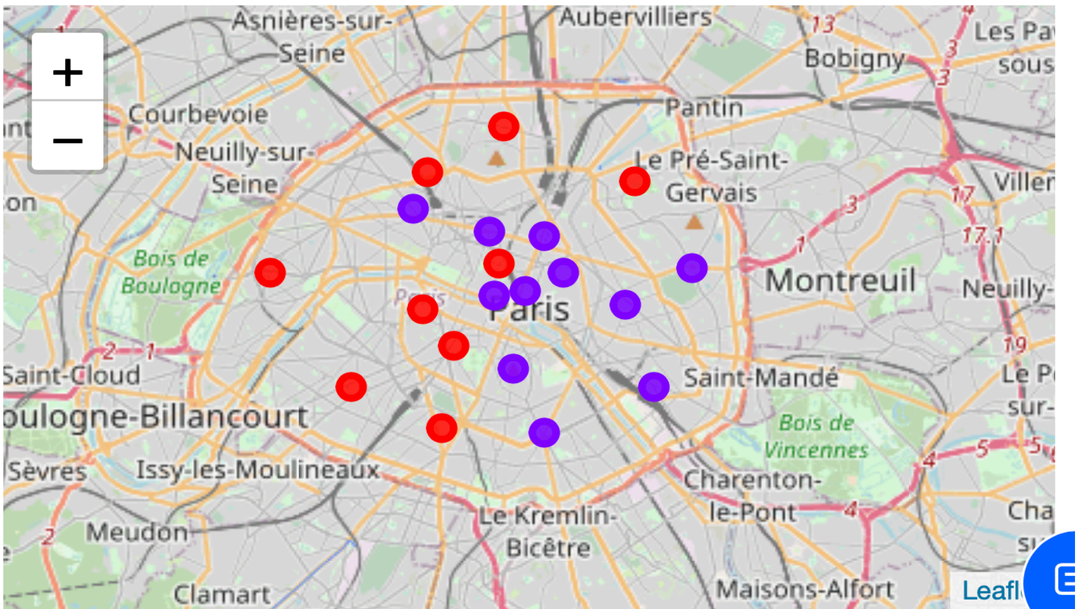
CLUSTER ZERO

	3rd Most Common Venue
Bar	2
Café	2
Chocolate Shop	1
Grocery Store	1
Pizza Place	1

3. Result :

After choosing the top two clusters and tabulating the top three common venues for each cluster, we can confidently conclude that Paris has more number of food outlets compared to New York. Even amongst the most common venues or popular venues, restaurants and food outlets have a higher proportion in Paris neighbourhood clusters as compared to that of New York.

Based on the analysis above it is feasible to conduct a further analysis into the economics of setting up a distribution warehouse centre in Paris.



Cluster Map Paris

	Neighborhoods	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	Postal Code	Latitude	Longitude
1	Paris 02 Bourse	French Restaurant	Wine Bar	Pizza Place	Italian Restaurant	Cocktail Bar	Bistro	Cheese Shop	Pastry Shop	Restaurant	Hotel	75002	48.8655	2.3426
5	Paris 06 Luxembourg	French Restaurant	Hotel	Chocolate Shop	Wine Bar	Pastry Shop	Restaurant	Clothing Store	Cocktail Bar	Miscellaneous Shop	Coffee Shop	75006	48.8493	2.3300
6	Paris 07 Palais-Bourbon	French Restaurant	Hotel	Café	Italian Restaurant	Bistro	Coffee Shop	Chocolate Shop	Restaurant	Seafood Restaurant	Japanese Restaurant	75007	48.8565	2.3210
13	Paris 14 Observatoire	French Restaurant	Bar	Italian Restaurant	Hotel	Vietnamese Restaurant	Bistro	Ice Cream Shop	Fish Market	Southwestern French Restaurant	Restaurant	75014	48.8331	2.3264
14	Paris 15 Vaugirard	French Restaurant	Hotel	Bar	Coffee Shop	Italian Restaurant	Supermarket	Thai Restaurant	Japanese Restaurant	Korean Restaurant	Lebanese Restaurant	75015	48.8412	2.3003
15	Paris 16 Passy	French Restaurant	Café	Grocery Store	Chinese Restaurant	Supermarket	Sandwich Place	Restaurant	Pizza Place	Asian Restaurant	Italian Restaurant	75016	48.8637	2.2769
16	Paris 17 Batignolles-Monceau	French Restaurant	Bar	Wine Bar	Hotel	Italian Restaurant	Restaurant	Thai Restaurant	Asian Restaurant	Bookstore	Creperie	75017	48.8835	2.3219
17	Paris 18 Buttes-Montmartre	French Restaurant	Bar	Café	Pizza Place	Italian Restaurant	Restaurant	Gastropub	Deli / Bodega	Wine Bar	Coffee Shop	75018	48.8925	2.3444
18	Paris 19 Buttes-Chaumont	French Restaurant	Italian Restaurant	Bar	Restaurant	Supermarket	Seafood Restaurant	Lebanese Restaurant	Moroccan Restaurant	Hotel	Vietnamese Restaurant	75019	48.8817	2.3822

Cluster Zero Paris

Based on the above table one can observe that food outlets dominate the ten most common venue for category zero in Paris city.

4. Discussion :

We first carried out an exploratory or preliminary analysis based on which we eliminated Toronto city from further analysis. **Toronto city has a lesser proportion of top priority consumers i.e. restaurants and hotels - and a greater proportion of café and coffee houses**. We then carried out cluster analysis for New York and Paris.

From the cluster analysis we concluded that restaurants and hotels **do not dominate the 'most common venues' for New York City cluster**. However the case was opposite for Paris wherein for some neighbourhoods the **'top 10 common venues' were food outlets.**

The study was preliminary in nature and requires further analysis to take a firm investment decision. Factors such as ease of setting up a business and real estate expenses can be considered in further economic analysis.

One can also improve the current model by further streamlining the number of venue categories and increasing the number of neighbourhood points.

Additionally one could have also given additional weightage to certain food outlets that consume higher amount of vegetables and fruits based on the cuisine of the restaurant. In this model we gave equal weightage to each cuisine.

5. Conclusion:

Out of the three proposed cities we have narrowed down with a high degree of confidence on **Paris.** The city of Paris has a higher proportion of top priority consumers.

Considering the higher proportion of top priority consumers we can assume that there will be strong demand for the client's products.

However further economic analysis needs to be done before making an investment decision to set up a warehouse distribution centre.

6. References:

- https://developer.foursquare.com/
- www.aggdata.com
- https://cocl.us/new_york_dataset
- https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M