# BATTLE OF THE NEIGHBORHOODS

## Manhattan vs Toronto

By Asavari Paluskar

# Battle of the Neighborhoods

Contents:

# 1. Introduction :

**Objective:**

- Analysing and exploring the data of **Manhattan and Toronto** to find the the **most favoured type of activity** by the customers and find the venue which has the maximum number of visitors (customer foot-fall).
- **Select the best type of business venture in each of the geographical regions.**

**Client:** ABC & Co.

**Success Criteria:** Finding the type of business with maximum customer foot-fall in Manhattan and Toronto.

**Manhattan**

**Toronto**

By Asavari Paluskar

# Neighborhoods Analysed:

## (A) New York:

- Most populous city in the United States
- Home to a wide variety of businesses and financial capital of the United States
- Major centre for banking and finance, retailing, world trade, transportation, tourism, real estate, new media, traditional media, advertising, legal services, accountancy, insurance, theatre, fashion, and the arts .
- Wide customer base and ever increasing population. High competition.
- Five boroughs: Brooklyn, Queens, Manhattan, Bronx and Staten Island.

## (B) Manhattan:

- Urban core of New York
- City's economic and administrative centre, cultural identifier and historical birthplace.
- Hosts the United Nations Headquarters.

By Asavari Paluskar

# Neighborhoods Analysed:

## (C) Toronto:

- Most populous city in Canada
- International centre of business, finance, arts, and culture
- Recognized as one of the most multicultural and cosmopolitan cities in the world
- Diverse population – important destination for immigrants.
- Prominent centre for music, theatre, motion picture production, and television production.
- Ever increasing population – ideal for starting a new business.

By Asavari Paluskar

# Problem Description:

- Starting a new business venture in Manhattan or Toronto can be very risky.
- Ever increasing competition and high cost of establishment.
- Various types of businesses operational in both geographical areas like restaurants offering various cuisines, cafes, coffee shops, amusement parks, gyms, bars, bakeries etc.
- The demand structure in Manhattan and Toronto is very different.
- Necessary for an aspiring entrepreneur to study the customer base thoroughly and select an area of business that is most favoured by the target audience
- We need to analyse data of both geographical regions separately, visualize it and **find the most profitable business venture in each geographical area, in terms of customer foot-fall.**

By Asavari Paluskar

# 2. Data:

**(A) Libraries and dependancies used:**

Numpy, Pandas, json, geopy, matplotlib, kmeans clustering, folium and BeautifulSoup

**(B) Datasets for Geographical Regions Analysed:**

**New York**: "https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs/newyork_data.json"

**Manhattan**: Data for New York is much bigger than the data for Toronto, hence, New York data is sliced and only the data pertaining to Manhattan is selected for further analysis, for ease of comparability with Toronto data.

**Toronto**: Data not readily available.
Webpage:
https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
Data extracted using BeautifulSoup.

By Asavari Paluskar

# 3. Methodology:

## Exploratory Data Analysis:

### (A) New York:

- Load data from newyork_data.json file.
- Transform it in a pandas dataframe and loop the data to fill the dataframe containing geographical co-ordinates of neighborhoods in NYC.
- Data used to get details of venues in NYC from Foursquare API.

**Dataframe for New York**

| | Borough | Neighborhood | Latitude | Longitude |
|---|---|---|---|---|
| 0 | Bronx | Wakefield | 40.894705 | -73.847201 |
| 1 | Bronx | Co-op City | 40.874294 | -73.829939 |
| 2 | Bronx | Eastchester | 40.887556 | -73.827806 |
| 3 | Bronx | Fieldston | 40.895437 | -73.905643 |
| 4 | Bronx | Riverdale | 40.890834 | -73.912585 |

By Asavari Paluskar

# Exploratory Data Analysis (continued):

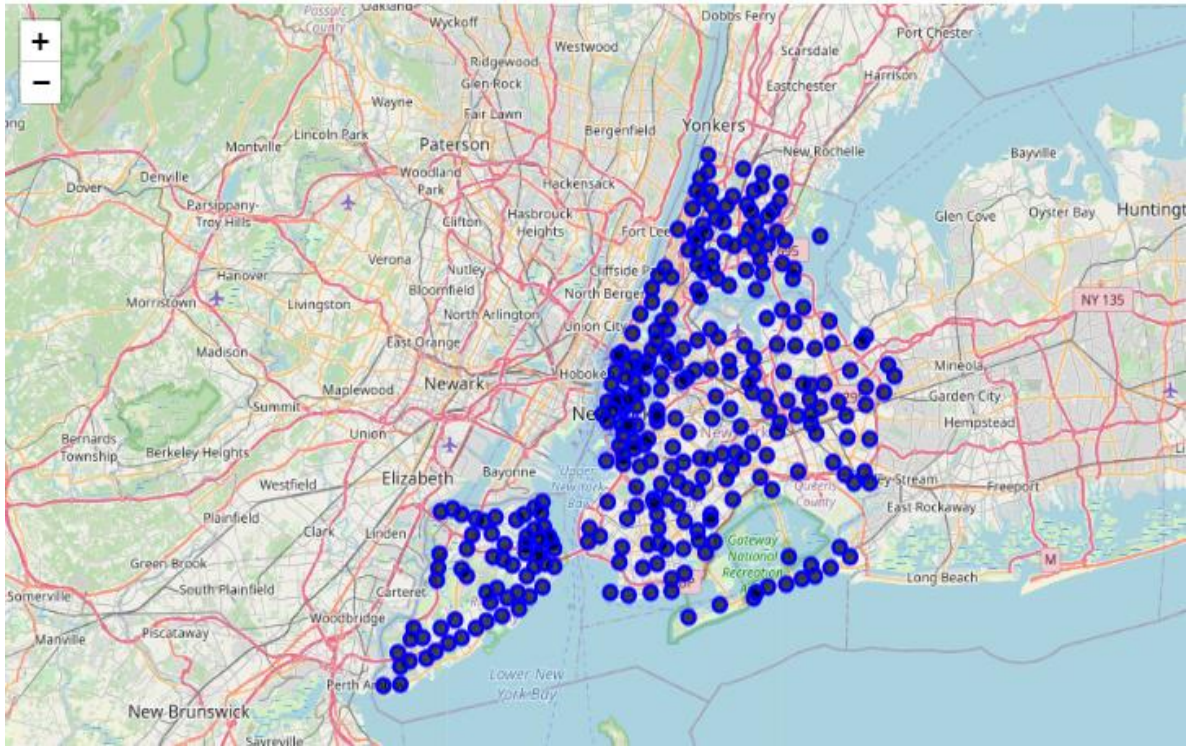
## (A) New York (continued):

- Geopy and Folium used to get the geographic co-ordinates and create a map of New York with all the neighborhoods superimposed on top.

- Foursquare API is used to get details of venues in New York.

- New York dataset much larger than Toronto dataset, so the dataset is sliced to select data for only Manhattan, which is comparable to Toronto in terms of size.

**Map of New York using Folium**




By Asavari Paluskar

# Exploratory Data Analysis (continued):

## (B) Manhattan:

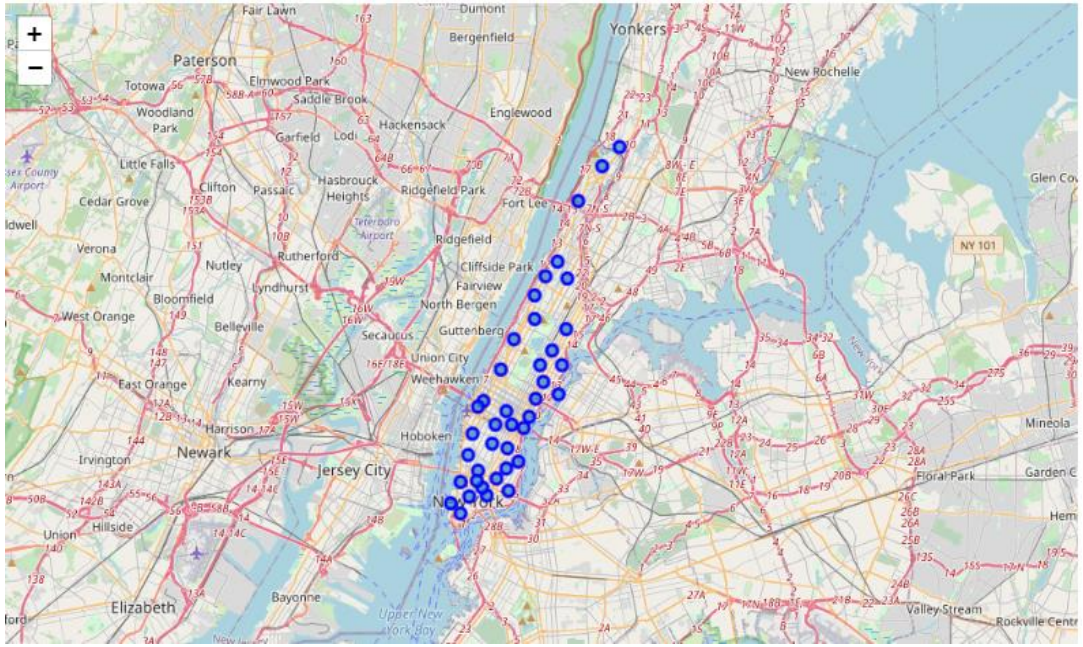
- New York dataset is sliced to select data for Manhattan.

**Dataset for Manhattan**

| | Borough | Neighborhood | Latitude | Longitude |
|---|---|---|---|---|
| 0 | Manhattan | Marble Hill | 40.876551 | -73.910660 |
| 1 | Manhattan | Chinatown | 40.715618 | -73.994279 |
| 2 | Manhattan | Washington Heights | 40.851903 | -73.936900 |
| 3 | Manhattan | Inwood | 40.867684 | -73.921210 |
| 4 | Manhattan | Hamilton Heights | 40.823604 | -73.949688 |

- Get geographical co-ordinates using geopy and create a map using Folium

**Map of Manhattan using Folium**




By Asavari Paluskar

# Exploratory Data Analysis (continued):

## (B) Manhattan (continued):

- Foursquare API used to get details of venues in each neighbourhood and their co-ordinates.
- Onehot encoding used to transform the dataset based on venues and neighborhoods.
- Data grouped based on neighborhoods by taking mean of the frequency of occurrence of each.
- Data grouped based on the 10 most commonly visited venues by customers in different neighborhoods and transformed into a pandas dataframe.
  *(dataframe on next slide)*

By Asavari Paluskar

# Exploratory Data Analysis (continued):

**(B) Manhattan (continued):** Dataset of 10 most commonly visited venues by customers in different neighborhoods transformed into a pandas dataframe.

| | 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 | Clothing Store | Hotel | Park | Gym | Memorial Site | Gourmet Shop | Burger Joint | Beer Garden | Food Court |
| 1 | Carnegie Hill | Coffee Shop | Cosmetics Shop | Café | Yoga Studio | Bookstore | Gym | French Restaurant | Pizza Place | Wine Shop | Bakery |
| 2 | Central Harlem | African Restaurant | French Restaurant | Chinese Restaurant | American Restaurant | Cosmetics Shop | Seafood Restaurant | Bar | Southern / Soul Food Restaurant | Gym | Ethiopian Restaurant |
| 3 | Chelsea | Coffee Shop | Bakery | American Restaurant | Art Gallery | French Restaurant | Seafood Restaurant | Italian Restaurant | Ice Cream Shop | Hotel | Park |
| 4 | Chinatown | Bakery | Chinese Restaurant | Hotpot Restaurant | Cocktail Bar | Optical Shop | Spa | Dessert Shop | American Restaurant | Salon / Barbershop | Shanghai Restaurant |

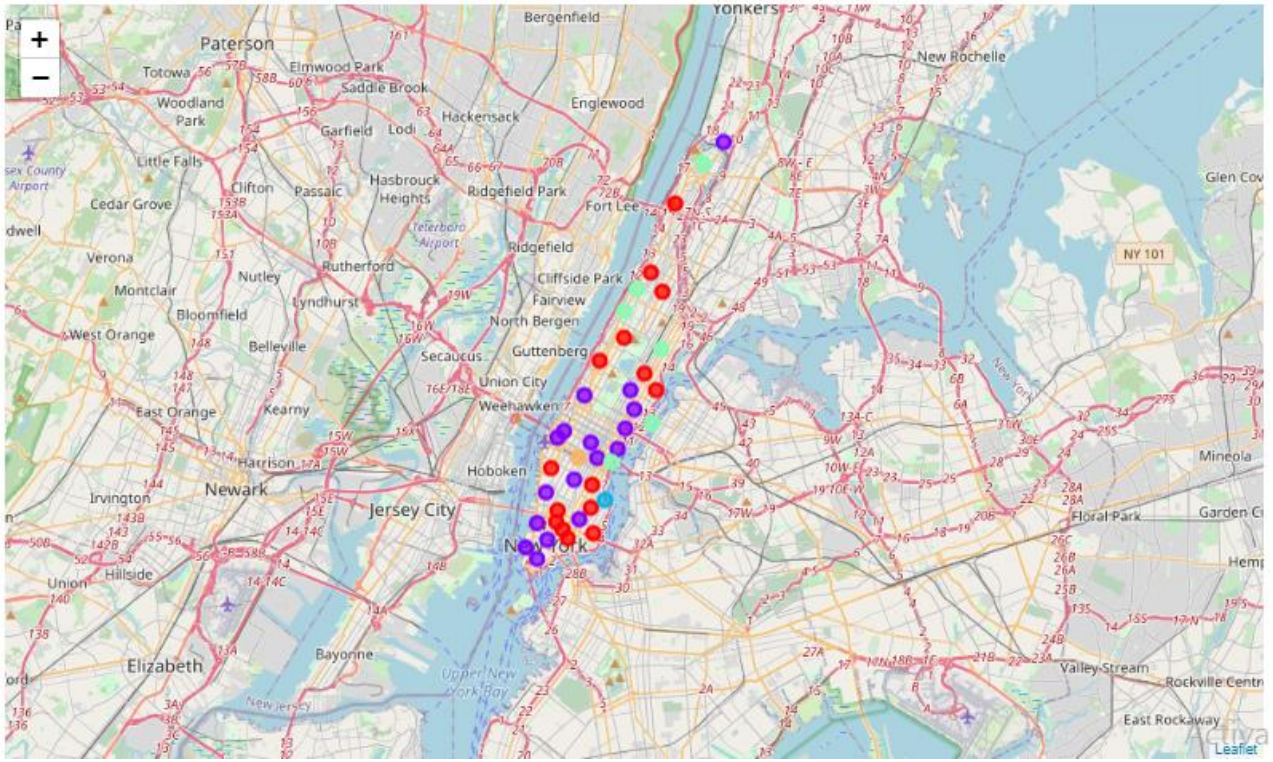
By Asavari Paluskar

# Exploratory Data Analysis (continued):

## (B) Manhattan (continued):

- Kmeans clustering used to divide the data in five clusters.
- Dataframe created containing details of all the neighborhoods, clusters and the top 10 venues for each neighbourhood.
- Cluster data visualised using Folium. Each cluster marked in a different colour on the map

**Map of Manhattan with Clusters using Folium**




By Asavari Paluskar

# Exploratory Data Analysis (continued):

**(B) Manhattan (continued):** Each cluster examined separately.

Extract of cluster one for example:

| | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Chinatown | Bakery | Chinese Restaurant | Hotpot Restaurant | Cocktail Bar | Optical Shop | Spa | Dessert Shop | American Restaurant | Salon / Barbershop | Shanghai Restaurant |
| 2 | Washington Heights | Café | Bakery | Mobile Phone Shop | Bank | Park | New American Restaurant | Coffee Shop | Grocery Store | Gym | Deli / Bodega |
| 4 | Hamilton Heights | Pizza Place | Café | Coffee Shop | Mexican Restaurant | Deli / Bodega | Yoga Studio | Caribbean Restaurant | Chinese Restaurant | School | Sandwich Place |
| 6 | Central Harlem | African Restaurant | French Restaurant | Chinese Restaurant | American Restaurant | Cosmetics Shop | Seafood Restaurant | Bar | Southern / Soul Food Restaurant | Gym | Ethiopian Restaurant |
| 9 | Yorkville | Italian Restaurant | Gym | Coffee Shop | Bar | Sushi Restaurant | Deli / Bodega | Mexican Restaurant | Wine Shop | Diner | Japanese Restaurant |

By Asavari Paluskar

# Exploratory Data Analysis (continued):

## (C) Toronto:

- Toronto data is not easily available. So the following webpage is used: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- BeautifulSoup used to get data from the webpage.
- Data transformed into a pandas dataframe and data looped in to fill the dataframe.
- Geocoder used to get the geographic co-ordinates for each postal code.
- The two tables are then merged to create one table containing details of neighborhoods and their co-ordinates.
- The dataset contains details for all the boroughs, so it is sliced to select data for Toronto. *(dataset in next slide)*

By Asavari Paluskar

# Exploratory Data Analysis (continued):

## (C) Toronto (continued):

- Foursquare API used to get details of venues in each neighbourhood and their co-ordinates.
- Onehot encoding used to transform the dataset based on venues and neighborhoods.
- Data grouped based on neighborhoods by taking mean of the frequency of occurrence of each.

**Extract of dataset for Toronto**

| | Postalcode | Borough | Neighborhood | Latitude | Longitude |
|---|---|---|---|---|---|
| 37 | M4E | East Toronto | The Beaches | 43.676357 | -79.293031 |
| 41 | M4K | East Toronto | The Danforth West, Riverdale | 43.679557 | -79.352188 |
| 42 | M4L | East Toronto | India Bazaar, The Beaches West | 43.668999 | -79.315572 |
| 43 | M4M | East Toronto | Studio District | 43.659526 | -79.340923 |
| 44 | M4N | Central Toronto | Lawrence Park | 43.728020 | -79.388790 |

By Asavari Paluskar

# Exploratory Data Analysis (continued):

## (C) Toronto (continued):

- Data grouped based on the 10 most commonly visited venues by customers in different neighborhoods and transformed into a pandas dataframe.

| | 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 | Berczy Park | Coffee Shop | Bakery | Cocktail Bar | Farmers Market | Seafood Restaurant | Restaurant | Pharmacy | Cheese Shop | Beer Bar | Café |
| 1 | Brockton, Parkdale Village, Exhibition Place | Café | Breakfast Spot | Nightclub | Coffee Shop | Yoga Studio | Performing Arts Venue | Burrito Place | Restaurant | Climbing Gym | Convenience Store |
| 2 | Business reply mail Processing Centre, South C... | Yoga Studio | Smoke Shop | Auto Workshop | Brewery | Burrito Place | Butcher | Comic Shop | Farmers Market | Fast Food Restaurant | Garden |
| 3 | CN Tower, King and Spadina, Railway Lands, Har... | Airport Lounge | Airport Service | Airport Terminal | Airport | Bar | Coffee Shop | Rental Car Location | Sculpture Garden | Boat or Ferry | Boutique |
| 4 | Central Bay Street | Coffee Shop | Sandwich Place | Café | Italian Restaurant | Thai Restaurant | Japanese Restaurant | Burger Joint | Bubble Tea Shop | Salad Place | Portuguese Restaurant |

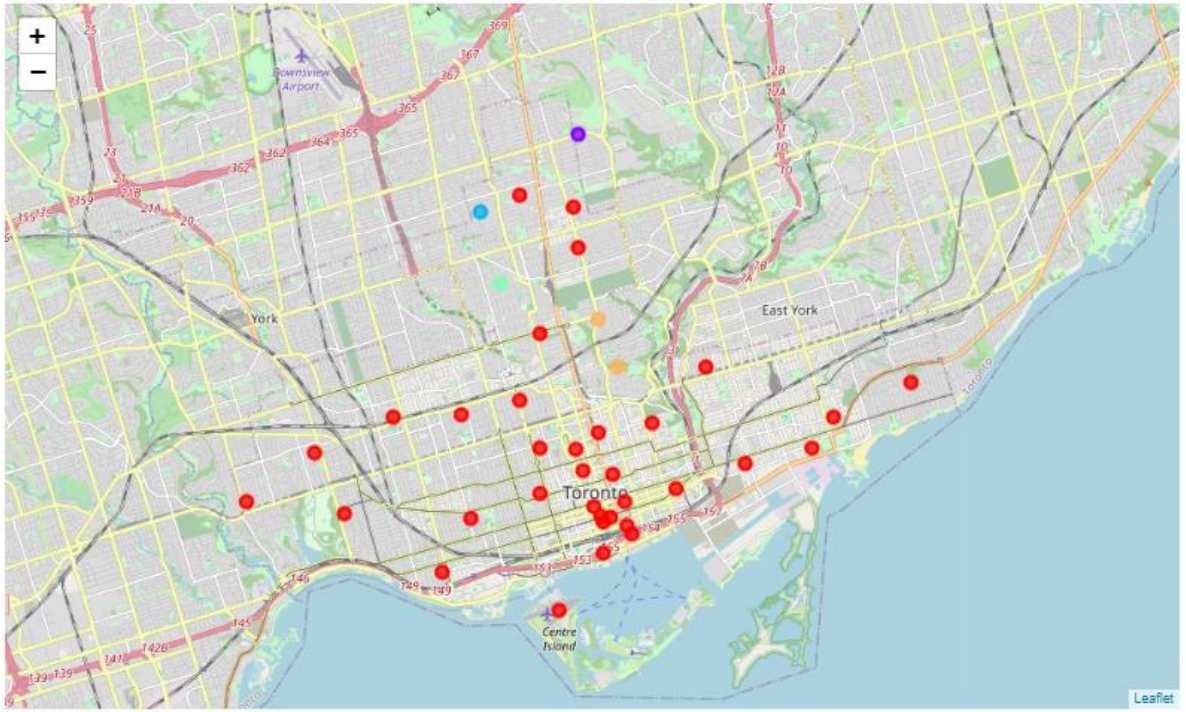
By Asavari Paluskar

# Exploratory Data Analysis (continued):

## (C) Toronto (continued):

- Kmeans clustering used to divide the data in five clusters.
- Dataframe created containing details of all the neighborhoods, clusters and the top 10 venues for each neighbourhood.
- Cluster data visualised using Folium. Each cluster marked in a different colour on the map

**Map of Toronto with Clusters using Folium**




By Asavari Paluskar

# Exploratory Data Analysis (continued):

**(C) Toronto (continued):** Each cluster examined separately.

Extract of cluster one for example:

| | Borough | Cluster Labels | 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 | East Toronto | 0 | Asian Restaurant | Health Food Store | Trail | Pub | Yoga Studio | Dumpling Restaurant | Dog Run | Doner Restaurant | Donut Shop | Electronics Store |
| 41 | East Toronto | 0 | Greek Restaurant | Coffee Shop | Italian Restaurant | Ice Cream Shop | Furniture / Home Store | Bookstore | Indian Restaurant | Spa | Pub | Japanese Restaurant |
| 42 | East Toronto | 0 | Sandwich Place | Fast Food Restaurant | Sushi Restaurant | Pub | Brewery | Board Shop | Restaurant | Italian Restaurant | Fish & Chips Shop | Steakhouse |
| 43 | East Toronto | 0 | Coffee Shop | American Restaurant | Bakery | Brewery | Café | Gastropub | Yoga Studio | Diner | Park | Middle Eastern Restaurant |
| 45 | Central Toronto | 0 | Gym / Fitness Center | Hotel | Pizza Place | Department Store | Sandwich Place | Breakfast Spot | Food & Drink Shop | Park | Gastropub | Gift Shop |

By Asavari Paluskar

# Exploratory Data Analysis (continued):

## (D) Analysing Manhattan and Toronto based on venues:

- Analysing both geographical regions to find the most favoured activity by the target audience in these regions.

- **Pie charts** used for data visualisation*.

- Using the base data for neighborhoods and venues that was used earlier, a pandas dataframe containing details of the most common venues in each neighbourhood in Manhattan and Toronto is created. *(dataframes in next slide)*

** Note: the preferred mode for visualising and presenting the results was wordcloud, but as there was some error while downloading the package, we had to use pie chart.*

By Asavari Paluskar

# Exploratory Data Analysis (continued):

## (D) Analysing Manhattan and Toronto based on venues (continued):

- Extract of dataframes created:

**Manhattan**

| | Neighborhood | Most Common Venue |
|---|---|---|
| 0 | Battery Park City | Coffee Shop |
| 1 | Carnegie Hill | Coffee Shop |
| 2 | Central Harlem | African Restaurant |
| 3 | Chelsea | Coffee Shop |
| 4 | Chinatown | Bakery |

**Toronto**

| | Neighborhood | Most Common Venue |
|---|---|---|
| 0 | Berczy Park | Coffee Shop |
| 1 | Brockton, Parkdale Village, Exhibition Place | Café |
| 2 | Business reply mail Processing Centre, South C... | Yoga Studio |
| 3 | CN Tower, King and Spadina, Railway Lands, Har... | Airport Lounge |
| 4 | Central Bay Street | Coffee Shop |

By Asavari Paluskar

# Exploratory Data Analysis (continued):

## (D) Analysing Manhattan and Toronto based on venues (continued):

- Proceed by taking a count of the neighborhoods for each of the 'Most Common Venues' for Manhattan and Toronto. This would help in creating pie charts. (Extract of data:)
- Matplotlib and ggplot are used to create pie-charts

**Manhattan**

| Most Common Venue | Neighborhood |
|---|---|
| African Restaurant | 1 |
| American Restaurant | 1 |
| Bakery | 2 |
| Bar | 2 |
| Boat or Ferry | 1 |

**Toronto**

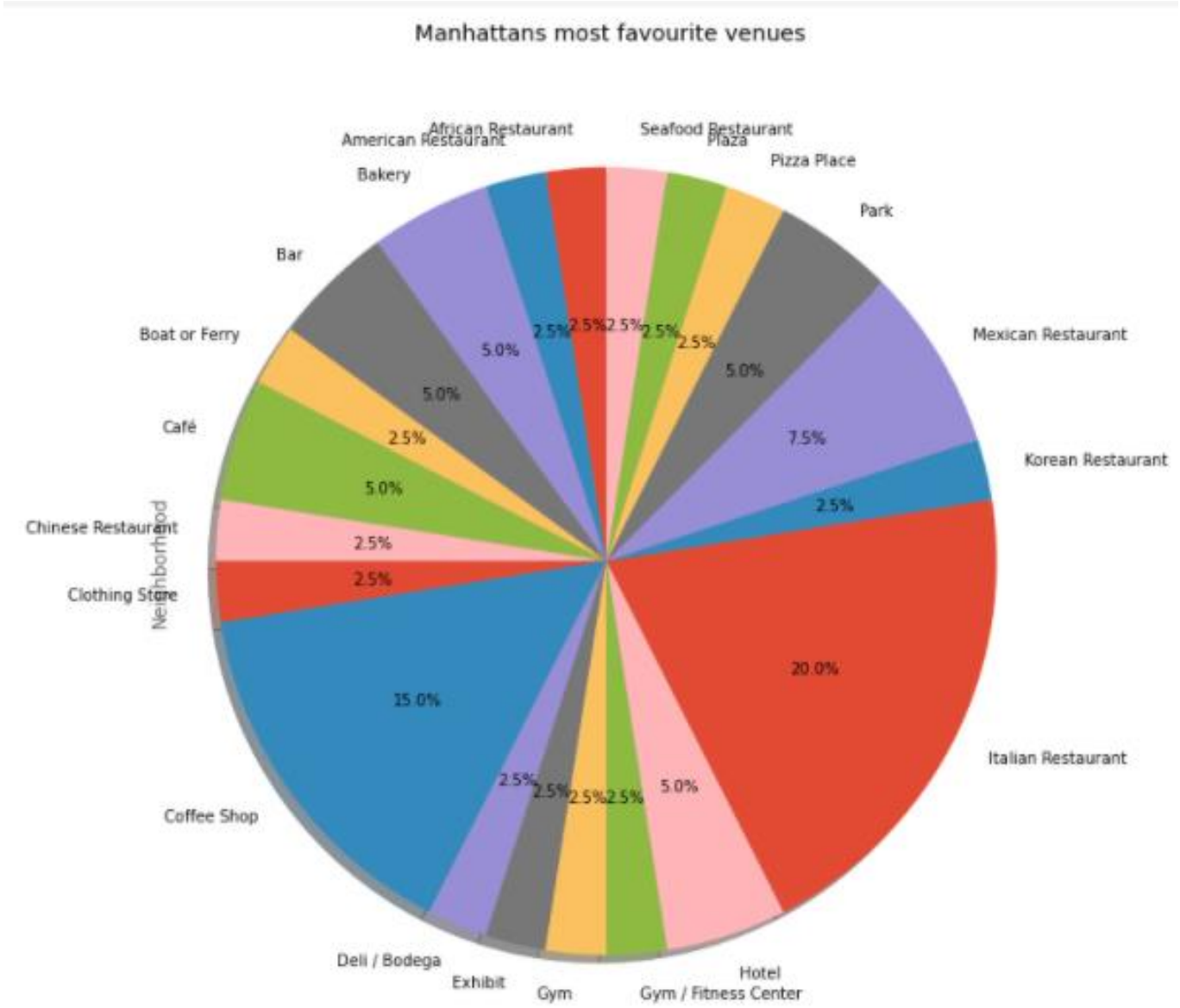
| Most Common Venue | Neighborhood |
|---|---|
| Airport Lounge | 1 |
| Asian Restaurant | 1 |
| Bakery | 1 |
| Bar | 1 |
| Breakfast Spot | 1 |

By Asavari Paluskar

# Exploratory Data Analysis (continued):

## (D) Analysing Manhattan and Toronto based on venues (continued):

**Pie-chart for Manhattan:**

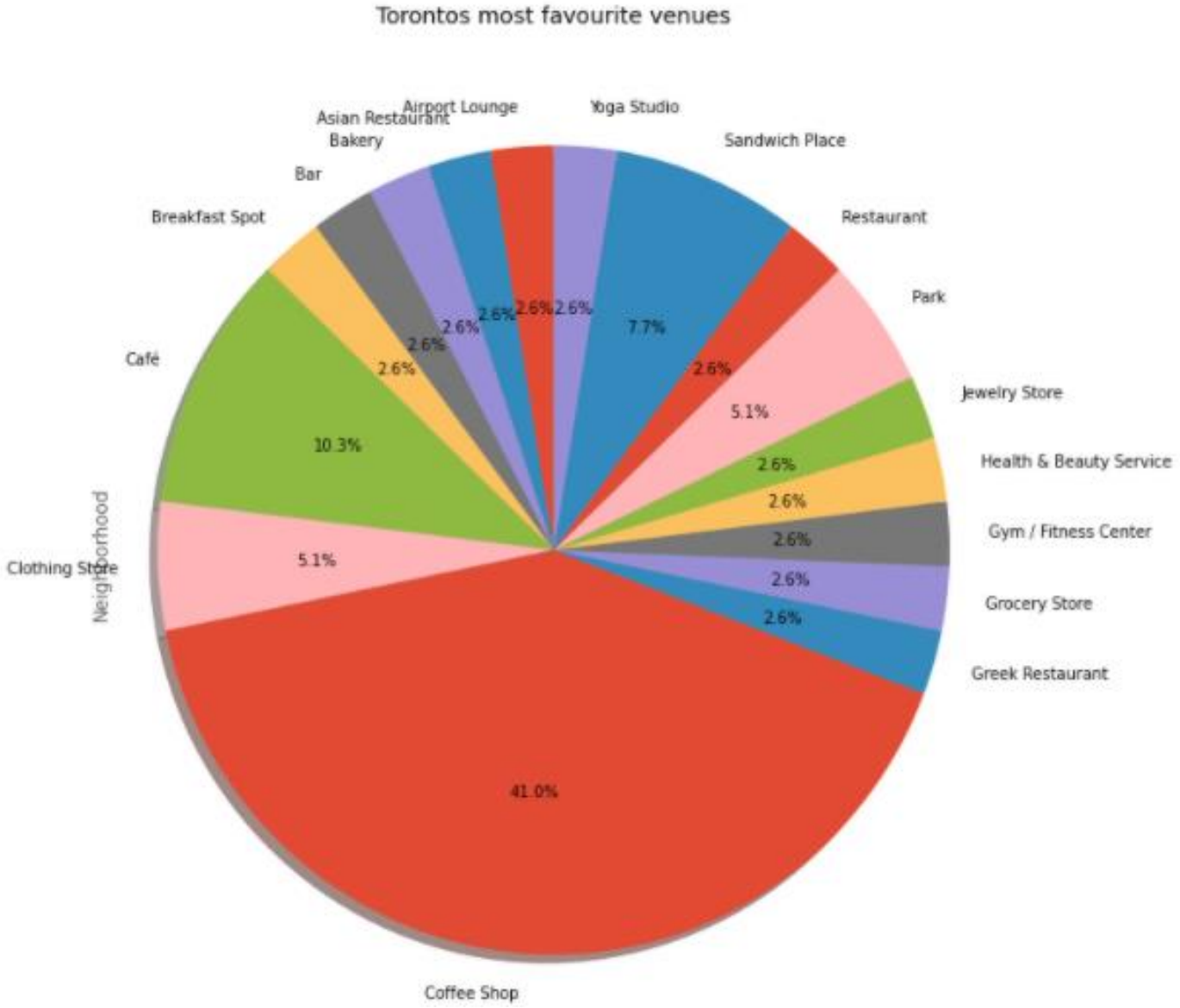



By Asavari Paluskar

# Exploratory Data Analysis (continued):

## (D) Analysing Manhattan and Toronto based on venues (continued):

**Pie-chart for Toronto:**




By Asavari Paluskar

# 4. Results:

**(a) Based on the analysis, we see that both Manhattan and Toronto are:**

- Similar in terms of being multi-cultural and cosmopolitan cities
- Major centres for various businesses
- Provide a large customer base
- Attract new businesses and are subject to every increasing competition in the business environment.

*(Analysis of Manhattan and Toronto on next slide)*

(b) An aspiring entrepreneur needs to carefully **analyse both the geographical regions** before deciding on the type of business venture.

(c) The type of venue that has the **most customer footfall could be considered as an ideal type of business** for an aspiring entrepreneur, in order to **ensure profit**.

By Asavari Paluskar

# 4. Results (continued):

## Manhattan:

- Based on the pie-chart, we see that the most commonly visited venues and their share in customer foot-fall are:
  - ***(a) Italian Restaurants – 20%***
  - ***(b) Coffee Shops – 15%***
  - ***(c) Mexican Restaurants – 7.5%***
- All three put together account for about **42.5%** of the total customer foot-fall.
- Following them are Parks, Bakeries, Bars, Cafes and Hotels, each with a share of 5%.

## Toronto:

- Based on the pie-chart, we see that the most commonly visited venues and their share in customer foot-fall are:
  - ***(a) Coffee Shops – 41%***
  - ***(b) Cafes – 10.3%***
- Both put together account for about **51.3%** of the total customer foot-fall, which is more than half the market size.
- Following them are Sandwich shops(7.3%), Clothing Stores (5%) and Parks (5%).

By Asavari Paluskar

# 5. Discussion:

**The analysis is based on various sources of information and assumptions:**

- The venue information extracted using **Foursquare API**, which may not account for all the customers in the geographical regions.

- **Manhattan region from New York** is used for the purpose of analysis and to **make it comparable to Toronto** data. However, the Manhattan data **may not be an exact representation** for the entire New York dataset. **Customer preferences may differ** depending on the geographical regions within New York.

- **High customer footfall** has been assumed to be the representative of **high sales and high profit**, however, it may not always be so.

- Information regarding the **profit margins and initial capital expenditure** for each of the businesses/venues analysed **could further help in deciding the most profitable type of business.**

By Asavari Paluskar

# 6. Conclusion:

- Manhattan and Toronto are very different in terms of activities that are most favoured by the customers. Hence, establishing the same type of business in both the regions may not yield the same level of returns.

- In **Manhattan**, an aspiring entrepreneur could expect a favourable response from the target customers if they were to open an Italian Restaurant or a Coffee Shop or a Mexican Restaurant.

- In **Toronto**, on the other hand, they could expect a favourable response from the target customers if they were to open a Coffee Shop or a Café.

*Note:The analysis is based on limited data and market segment information. Further information about customer preferences, profit margins and capital investments in various businesses in both the geographical regions could help refine the results.*

By Asavari Paluskar

# THANK YOU !

By Asavari Paluskar