1. **Introduction:**
   1. **Background**

Two of the most important cities in the world are New York City and City of Toronto. Every year millions of tourists visit these cities for business and pleasure.

The main reasons for tourist to visit these cities are as follows:

* Discover different neighborhoods
* Explore similar and different venues
* Enjoying different international and experimental foods
* Visiting world renowned arts and galleries

People from different part of the world come to visit, they like to explore places, taste similar and different cuisines, enjoy popular sites and much more. When someone wants to visit and has to decide which city to choose, the visitor/visitors would like to compare two cities based on their likes and dislike.

Instead of the whole New York City, the borough of Manhattan is chosen to compare with Toronto.

* 1. **Problem**

Data from two cities can help decide tourists and tourism industry. The comparative analysis contributes to determining places to visit in these two cities. The annual or seasonal number of tourists and their interests influence the reshaping of these two cities.

* 1. **Interest**

A comparison of the venues between two cities will help people decide where to visit. A data analysis between New York city and Toronto which gives a picture of the sought-after venues will serve the purpose.

1. **Data collection and cleaning:**

**2.1 Data Sources**

In order to make a comparative analysis of venues of interest between two cities (New York and Toronto), we need effective datasets for both the two cities.

Following data sources will be needed to extract/generate the required information:

* New York City data will be obtained from a json file obtained from IBM Developer Skills Network
* City of Toronto postal codes are obtained from a Wikipedia page
* City of Toronto data will be obtained from a csv file that has the geographical coordinates of each postal code: <http://cocl.us/Geospatial_dataFollowing>
* Number of restaurants and their type and location in every neighborhood will be obtained using Foursquare API

* 1. **Data Cleaning**

City of Toronto postal codes are obtained from a Wikipedia page. Only the boroughs which are assigned were considered for the data. the boroughs that are not assigned were discarded. The geographical coordinates of each postal code were collected and mapped for each borough to create a single table which contains the postal code, boroughs, the neighborhoods and the coordinates.

For New York City data, the data were extracted from the json file. The data is extracted in a way to have boroughs, the neighborhoods and the coordinates to match the same features as the Toronto city data.

Communicating with the Foursquare database is done by their RESTful API. A uniform resource identifier or URI is created, and extra parameters are appended depending on the data that we are seeking from the database. Any call request you make is composed of, we can call this base URI, which is api.foursquare.com/v2, and you can request data about venues, users, or tips.

For New York City, we will be using the coordinates of Manhattan to conduct the search. For City of Toronto, the coordinates of Toronto city will be used.

1. **Data Analysis:**

The two cities chosen for the analysis are city of Toronto and the borough of Manhattan of New York city. Since Manhattan is the most visited borough among other boroughs of New York, we chose borough of Manhattan.

We are comparing the venues of city of Toronto and borough of Manhattan in New York city.

In the first step, we have gathered datasets for both the two cities and explored the neighborhoods based on number of venues.

In the second step, we have analyzed the neighborhoods with respect to top ten venues for each neighborhood.

In the final step, we have used K-means clustering to cluster venue categories. This means, it is an unsupervised algorithm. Objects within a cluster are similar, and objects across different clusters are different or dissimilar. For both the cities, we have taken venues within 500 meters from the location of Toronto and Manhattan. The clusters show which venue category dominate which cluster.

By analyzing the venue categories between the clusters of the two cities, we can compare the similarities or differences between two cities.

* 1. **Comparison of venues between and Toronto and Manhattan**:

The total number of venues calculated for Toronto is 2133 and 3252 for Manhattan.

3.2 **Comparison of unique venues between and Toronto and Manhattan**:

The total number of venues calculated for Toronto is 1391 and 2767 for Manhattan.

* 1. **Comparison of unique venue categories between and Toronto and Manhattan**:

The number of venues, unique venues and venue categories are more in Manhattan than in the city of Toronto.

|  |  |  |  |
| --- | --- | --- | --- |
| City | No. of Venues | No. of Unique Venues | No. of Unique Venue Category |
| Toronto | 2133 | 1391 | 268 |
| Manhattan | 3252 | 2767 | 330 |

A picture containing table

Description automatically generated

Chart

Description automatically generated

* 1. **Comparison of different restaurants between and Toronto and Manhattan**:

The number of unique restaurants is more in Manhattan than in Toronto.

|  |  |  |  |
| --- | --- | --- | --- |
| City | No. of Restaurants | No. of Restaurants found only in Manhattan | No. of Restaurants found only in Toronto |
| Toronto | 480 | -- | 9 |
| Manhattan | 923 | 35 | -- |

Restaurant business is one of the most important tourist attractions for any big city.

**List of the restaurants only found in Manhattan**:

'Tapas Restaurant', 'Spanish Restaurant', 'Turkish Restaurant', 'Australian Restaurant', 'Peruvian Restaurant', 'Argentinian Restaurant', 'Israeli Restaurant', 'Shanghai Restaurant', 'Hawaiian Restaurant', 'African Restaurant', 'Southern / Soul Food Restaurant', 'Dumpling Restaurant', 'Lebanese Restaurant', 'Udon Restaurant', 'Egyptian Restaurant', 'Kebab Restaurant', 'Japanese Curry Restaurant', 'Hotpot Restaurant', 'Jewish Restaurant', 'Austrian Restaurant', 'Cantonese Restaurant', 'Russian Restaurant', 'North Indian Restaurant', 'Tibetan Restaurant', 'Kosher Restaurant', 'South Indian Restaurant', 'South American Restaurant', 'Himalayan Restaurant', 'Scandinavian Restaurant', 'Czech Restaurant', 'Szechuan Restaurant', 'Swiss Restaurant', 'Empanada Restaurant', 'Paella Restaurant'

**List of the restaurants only found in Toronto**:

'Comfort Food Restaurant', 'Gluten-free Restaurant', 'Portuguese Restaurant', 'Belgian Restaurant', 'Cajun / Creole Restaurant', 'Theme Restaurant', 'Korean BBQ Restaurant', 'Colombian Restaurant', 'Doner Restaurant', 'Hakka Restaurant'

Chart

Description automatically generated

A picture containing chart

Description automatically generated

**3.5 Clustering Neighborhoods:**

We have used **K-means clustering** to cluster venue categories. This means, it's an unsupervised algorithm. Objects within a cluster are similar, and objects across different clusters are different or dissimilar. For both the cities, we have taken venues within 500 meters from the location of Toronto and Manhattan. The clusters show which venue category dominate which cluster.

By analyzing the venue categories between the clusters of the two cities, we can compare the similarities or differences between two cities.

Toronto’s cluster of neighborhoods:

Map

Description automatically generated

Manhattan’s cluster of neighborhoods:Map

Description automatically generated

1. **Conclusion**:

## The objective of this project was to compare the city of Toronto and Manhattan of New York city based on their venues. I compared the number of venues, number of unique venues, number of unique venue categories. I also compared the restaurants of these two cities. The number and types of restaurants offered by the two cities are compared. Both cities have common and unique restaurants. I have used K-means clustering to cluster venue categories in both cities. Three of the five clusters of both Toronto and Manhattan have Park, Coffee Shop and Restaurant. Toronto is more popular with Fast Food restaurants and Manhattan is more popular with Italian restaurants.

## This analysis will help the curious stakeholders to have a picture about venues in Toronto and Manhattan and help them make decisions.