

IBM APPLIED DATA SCIENCE CAPSTONE

Co-working Facility Recommendation based on Attractiveness of Neighbourhood in Toronto

INTRODUCTION

Coworking and flexi work space concepts have gained a lot of traction in recent years. These are arrangements in which workers from different companies share an office space, allowing cost savings and convenience through the use of common infrastructure, such as equipment and utilities and in some cases includes meals, refreshments and postal services. It is attractive to independent contractors, independent scientists, telecommuting and work-at-home professionals, and people who travel frequently.

Problem Description

Coworking facilities have high maintenance costs and their profitability directly depends on the occupancy rates and overhead costs on account of rent, free services such as meals, refreshments, gyms etc. Therefore, location is very important for success of these facilities. So, identification of appropriate location for setting up such facility is a key success factor.

This project aims to aid the user with choosing an ideal location for such a facility in the Toronto area. It uses neighbourhood segmenting and clustering to group boroughs in the city and on this basis provides insights to 4 neighbourhood clusters to choose from based on 3 factors

- a. Surrounding facilities such as cafes, restaurants in the cluster (higher the better)
- b. Average real estate rental (lower the better)
- c. Cater to larger size of cluster in the cluster (larger the better)

The first 2 factors will enable lower overhead costs, whereas the last factor will increase the catchment areas increase potential customers. Finally, similarity of neighbourhood is very important for success of a coworking facility, so we also need to find the right cluster.

Target Audience

Companies and institutions which are into providing coworking or flexi workspace facilities and plans to move into Toronto as their next destination. This will also be helpful for any company who plans to relocate to Toronto.

DATA

Following data are chosen:

- a. Surrounding facilities such as cafes, restaurants in the cluster from Foursquare API
- b. Average real estate rental from Kaggle dataset on average rentals
- c. Cater to larger size of cluster in the cluster. Here postal number of postal codes is taken as surrogate of largeness of size. Postal codes are obtained from Wikipedia page.

THE FOURSQUARE API allows application developers to interact with the Foursquare platform. The API allows querying places and users, exploring popular places, and checking out reviews and photographs for these places. In this project we will be using Foursquare Places API to identify the 'most popular' venues

for each neighbourhood based on the locational coordinates. This is possible by the use of a explore call that returns a list of recommended locations in a specified area.

Toronto neighbourhoods data were scraped from a Wikipedia page titled 'List of Postal Codes of Canada: M' (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M). The dataset has list of postal codes along with name of borough and neighbourhoods. Geolocator was used the further enhance the dataset by latitudes and longitudes.

Real estate rentals in Toronto area is obtained from preformatted and precleaned csv file from Kaggle containing rentals and locational coordinates.

METHODOLOGY

Data Extraction and Cleaning

The data on postal codes was first scrapped and transformed into a pandas dataframe. A lot of missing values were identified where postal codes were not assigned to any borough or neighbourhood, and were dropped. Geolocator Nominatim from Geopy library was used to include locational coordinates (latitude and longitude) of each neighbourhood, as make a dataset which has postal codes, borough name, neighbourhood name, latitude and longitude. Post cleaning 4 boroughs and 20 neighbourhoods were identified as below. A snapshot of the 20 is provided below.

Table I

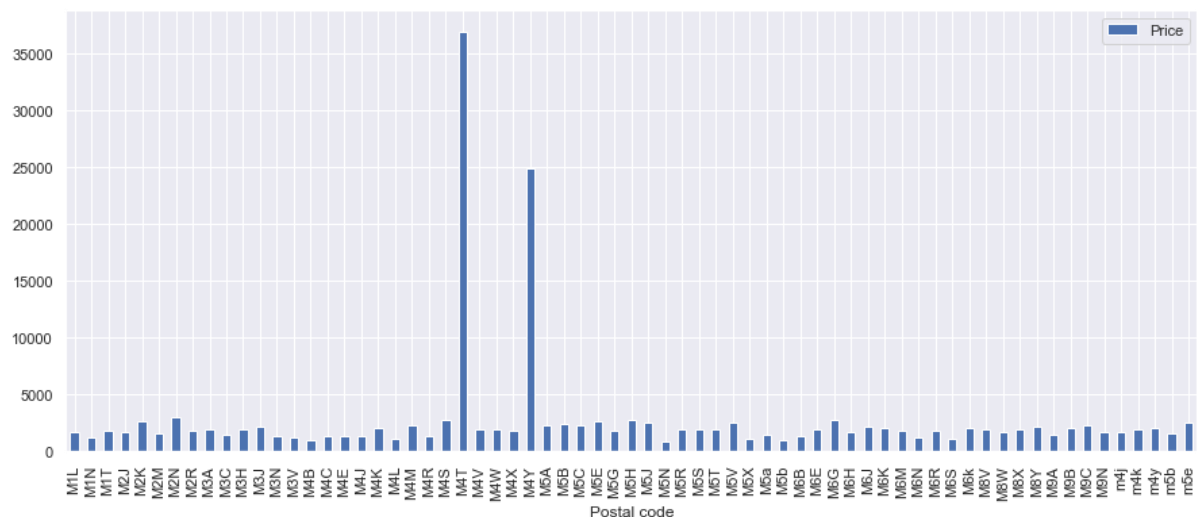
Postal code	Borough	Neighborhood	Latitude	Longitude
M5C	Downtown Toronto	St. James Town	43.669403	-79.372704
M4E	East Toronto	The Beaches	43.671024	-79.296712
M5E	Downtown Toronto	Berczy Park	43.647984	-79.375396
M5G	Downtown Toronto	Central Bay Street	43.660708	-79.385802
M6G	Downtown Toronto	Christie	43.664111	-79.418405
M6H	West Toronto	Dufferin / Dovercourt Village	43.660202	-79.435719
M5L	Downtown Toronto	Commerce Court / Victoria Hotel	43.648164	-79.377766
M4M	East Toronto	Studio District	43.649585	-79.390683
M4N	Central Toronto	Lawrence Park	43.729199	-79.403252
M5N	Central Toronto	Roselawn	43.699456	-79.454616
M4P	Central Toronto	Davisville North	43.697936	-79.397291
M5P	Central Toronto	Forest Hill North & West	43.701028	-79.425600
M4R	Central Toronto	North Toronto West	43.646547	-79.419526
M6R	West Toronto	Parkdale / Roncesvalles	43.641281	-79.451599
M4S	Central Toronto	Davisville	43.697936	-79.397291
M5S	Downtown Toronto	University of Toronto / Harbord	43.664096	-79.398669
M6S	West Toronto	Runnymede / Swansea	43.651778	-79.475923
M4W	Downtown Toronto	Rosedale	43.678356	-79.380746
M4X	Downtown Toronto	St. James Town / Cabbagetown	43.661517	-79.372140
M4Y	Downtown Toronto	Church and Wellesley	43.665524	-79.383801

The map of Toronto with the neighbourhoods are shown below:



Foursquare API's explore option was utilised to identify list of venues and its categories across each of the neighbourhoods. In total 992 such venues were identified across 192 categories for the 39 neighbourhoods. OneHotEncoder was used to encode the dataset of venues and was grouped based on neighbourhoods for usage in clustering.

Dataset on average rentals in Toronto area is obtained from preformatted csv file from Kaggle containing rentals and locational coordinates was pre-cleaned. The dataset on rental was loaded and the average rental based on postal codes was identified using groupby and mean function to identify the average rental. The below graph shows postal code wise average rentals in CAD. It was observed that there were certain neighbourhoods where rentals were steep, therefore not most ideal for a coworking facility from cost standpoint.



Data Analysis

Based on the Foursquare dataset 5 most common venues for each neighbourhood were identified and merged with the above dataset shown in table I. A glimpse of the first 5 rows given below.

Table II

Postal code	Borough	Neighborhood	Latitude	Longitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
M5C	Downtown Toronto	St. James Town	43.669403	-79.372704	Coffee Shop	Pizza Place	Café	Grocery Store	Playground
M4E	East Toronto	The Beaches	43.671024	-79.296712	Beach	Japanese Restaurant	Breakfast Spot	Bar	Tea Room
M5E	Downtown Toronto	Berczy Park	43.647984	-79.375396	Coffee Shop	Italian Restaurant	Café	Restaurant	Japanese Restaurant
M5G	Downtown Toronto	Central Bay Street	43.660708	-79.385802	Coffee Shop	Sandwich Place	Japanese Restaurant	Middle Eastern Restaurant	Diner
M6G	Downtown Toronto	Christie	43.664111	-79.418405	Korean Restaurant	Coffee Shop	Indian Restaurant	Mexican Restaurant	Café

Further, since similarity of neighbourhood is very important for success of a coworking facility, K-Means clustering is done, K-means clustering from scikit learn library was used on dataset as depicted in table II (which was name of the borough, name of the neighbourhood and top five venues), with 4 target clusters using kmeans ++ as groups which have not been explicitly labelled in the data. Average rentals for the 4 clusters was the separately computed and appended to the cluster results.

Table III

Postal code	Borough	Neighborhood	Latitude	Longitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	Cluster Labels	Price
M5C	Downtown Toronto	St. James Town	43.669403	-79.372704	Coffee Shop	Pizza Place	Café	Grocery Store	Playground	3	2279.166667
M4E	East Toronto	The Beaches	43.671024	-79.296712	Beach	Japanese Restaurant	Breakfast Spot	Bar	Tea Room	3	1300.000000

Postal code	Borough	Neighborhood	Latitude	Longitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	Cluster Labels	Price
M5E	Downtown Toronto	Berczy Park	43.647984	-79.375396	Coffee Shop	Italian Restaurant	Café	Restaurant	Japanese Restaurant	1	2705.405405
M5G	Downtown Toronto	Central Bay Street	43.660708	-79.385802	Coffee Shop	Sandwich Place	Japanese Restaurant	Middle Eastern Restaurant	Diner	1	1813.461538
M6G	Downtown Toronto	Christie	43.664111	-79.418405	Korean Restaurant	Coffee Shop	Indian Restaurant	Mexican Restaurant	Café	3	2841.666667

RESULTS

As per the analysis above neighbourhood cluster2 (cls_2) is the preferred choice as check all the criteria/factors i.e. larger areas as highest number of postal codes (8), thereby highest surrounding facilities such as cafes, restaurants in the cluster and lowest average real estate rental (\$1677.25).

Cluster name	Average Rental	No of postal codes/ Area size	Number of venues	Preferred (Yes/No)
Cluster 1 (cls_1)	\$6163.31	6	High	No
Cluster 2 (cls_2)	\$1677.25	8	High	Yes
Cluster 3 (cls_3)	\$1300.00	1	Very Low	No
Cluster 3 (cls_3)	\$1825.00	1	Very Low	No