Report

Predicting Best place for Warehouse Store

By Praveen Kumar

(14th Jan 2020)

1. Introduction/Business Problem

1.1 Background

It's never been easier to go for Warehouse Store. Why leave the comfort of your warm bed when you can simply press a few buttons and have your full order (produce and all) delivered to your Place after sometime. Warehouse Contractor wanted to develop and open it at such a place where his team can deliver the product to their neighbours in that Borough. The Warehouse should be the most benefited among all in that Borough. The least time contractor will take to deliver the products, more benefits they will be getting.So how this can be possible??

In order to minimize the chance of getting late they should plan and do research in a way to get the least delay for Customers. Satisfaction of Customers need no delay, good quality, optimum price e.t.c…

1.2 Problems

The daily work for Warehouse's Contractor to deliver products to the local Customer in the least time. The place of Delivery may be far or near. And there are many regular Customer who are doing their Business great and their demands are very high. These are most valuable Customers whom they won't want to have delay. How they will manage to deliver products in minimum time??

We are taking the data of Toronto City in which many Borough are includes. We are manipulating the data of Toronto which is taken from wikipedia page.
Link: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

1.3 Interest

We are fond of finding the best location in Etobicoke Borough in Toronto City such that the nearest places of Delivery should be the most valuable Customers from whom Warehouse are getting the most benefits. Neighbourhood places of the Borough Etobicoke should have more number of Customers.

2. Data Acquisition and Cleaning

2.1 Data Acquisition

The data acquired for this project is a combination of data from three sources. The first data source of the project uses a List of postal codes of Canada: M that shows the neighbours per borough in Toronto. The dataset contains the following columns:

- **Post_Code** : Post Code for all regions in Toronto.
- **Borough** : Common name for London borough.
- **Neighbourhood** : All the neighbours in that Borough.

The second source of dataset is created from scratch using the list of neighbourhood available on the site Latitudes and Longitudes . This page contains additional information about the boroughs, the following are the columns:

- **Post_Code** : Post Code for all regions in Toronto.
- **Latitudes** : Latitudes of all regions of each Borough in Toronto.
- **Longitudes** : Longitudes of all regions of each Borough in Toronto.
- **Neighbourhood:** Name of the neighbourhood in the Borough.

The third data source is the Foursquare API as found on the given link. This dataset is responsible for information of all neighbours latitude and longitude by requesting url using Foursquare API. This contains:

- **CLIENT_ID** = # your Foursquare ID
- **CLIENT_SECRET** =# your Foursquare Secret
- **VERSION** = # Foursquare API version

2.1 Data Cleaning

The data preparation for each of the three sources of data is done separately. From the Toronto data, the Borough post_code and their neighbourhood are present in our datasets.

The part A data is scraped directly from wikipedia which had 'Not assigned' values. After cleaning data of part A to part B we can see good form of dataset having no such stuffs.

	Postcode	Borough	Neighborhood
0	M1A	Not assigned	Not assigned
1	M2A	Not assigned	Not assigned
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Harbourfront

part A

	Postcode	Borough	Neighborhood
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Harbourfront
5	M6A	North York	Lawrence Heights
6	M6A	North York	Lawrence Manor

part B

***Figure-2.1.1*:** Data from Wikipedia

Now we will use Geospatial_data to get Latitude and longitude of our neighbours on 'postcode' and finally we merge them to get a new dataframe as shown in below figure:

	Postcode	Borough	Neighborhood	Latitude	Longitude
161	M8V	Etobicoke	Humber Bay Shores	43.605647	-79.501321
162	M8V	Etobicoke	Mimico South	43.605647	-79.501321
163	M8V	Etobicoke	New Toronto	43.605647	-79.501321
164	M8W	Etobicoke	Alderwood	43.602414	-79.543484
165	M8W	Etobicoke	Long Branch	43.602414	-79.543484

We needed the venues and vanue_Category for manipulation to get the result and conclusion. For that we will use Foursquare API to collect all the relevant data to reach at conclusion of our problem'answer.

	Postal Code	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Summary	Venue Category	Distance
0	M8V	Humber Bay Shores	43.605647	-79.501321	LCBO	This spot is popular	Liquor Store	408
1	M8V	Humber Bay Shores	43.605647	-79.501321	Huevos Gourmet	This spot is popular	Mexican Restaurant	532
2	M8V	Humber Bay Shores	43.605647	-79.501321	Sweet Olenka's	This spot is popular	Dessert Shop	512
3	M8V	Humber Bay Shores	43.605647	-79.501321	Kitchen on 6th	This spot is popular	Breakfast Spot	540
4	M8V	Humber Bay Shores	43.605647	-79.501321	Cellar Door Restaurant	This spot is popular	Italian Restaurant	790

fig 2.1.2: Data using Foursquare API

Neighborhood	Automotive Shop	Bakery	Bank	Burrito Place	Bus Line	Donut Shop	Dessert Shop	Cupcake Shop	Cheese Shop	Café	Business Service	Grocery Store	Garden Center	Hardware Store	Hotel
Albion Gardens	0	0	0	0	1	0	0	0	0	0	0	3	0	1	0
Alderwood	0	0	0	0	0	1	0	0	0	0	0	1	1	0	0
Beaumond Heights	0	0	0	0	1	0	0	0	0	0	0	3	0	1	0
Bloordale Gardens	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0
Cloverdale	0	0	1	0	0	0	0	0	0	1	0	1	0	0	1

fig 2.1.2: Neighbour-category Data using Foursquare API

Now we will use **k-means** to cluster neighbours into k=5 clusters finally to get best result output for our result and conclusion. Here we are showing some of our best venue as output using this algorithm.

	Neighborhood	Group
6	Humber Bay	5
7	Humber Bay Shores	5
11	King's Mill Park	5
13	Kingsway Park South East	5

fig 2.1.2: k-means output

3. Methodology

In this project we will direct our efforts on detecting areas of Toronto that have low restaurant density, particularly those with low number of Warehouse's Team who are responsible for providing required stuffs for running small store and restaurant . We will limit our analysis to the particular Borough “Etobicoke". In first step we have collected the required **data: location and type (category) of every neighbours within from Borough centres** (Etobicoke). We have also **identified Categories** (according to Foursquare categorisation). Second step in our analysis will be calculation and exploration of '**Shop Centres density**' across different areas of Toronto(Etobicoke) - we will use **one hot coding** to make more readable of categorical data. In third and final step we will focus on most promising areas and within those create **clusters of locations that meet some basic requirements** established in discussion with stakeholders: we will take into consideration locations with **more closer shop centres and restaurants, and we want locations **without Italian restaurants in that Borough**. We will present map of all such locations but also create clusters (using **k-means clustering**) of those locations to identify general zones / neighbourhoods / addresses which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders. We are mainly focussing on top 3 Group of venues that will best fit for the Contractor of Warehouse Store.
The Concept here is more closer will be the number of customer(any kind of Shopkeepers, any Restaurant, Public Servicer and so all such..) from the ‘Warehouse’ centre more will be benefits for the Warehouse teams and can extend their business.

3.Analysis

We will perform some basic explanatory data analysis and derive some additional informations from our raw data. First we count the different neighborhoods inside “Etobicoke”. Pickling is a way to convert a python object (list, dict, etc.) into a character stream. The idea is that this character stream contains all the information necessary to reconstruct the object in another python script. In our project we are getting .txt file.

Again we will analyze how many venues for stores of shops and Restaurants which are crucial for Warehouse Contractor for each post_code. After analyzing the the total number of neighbours of Etobicoke, Toronto. Based on that we will be getting over on problem of how many shopping Centres and Restaurants in terms of total are there. Then We are looking for list of categories are there for doing the Business. More Categorical data we got more useful and benificials that Borough becomes for the Warehouse’s team..

We Will use Machine Learning Concept of k-means clustering by which we can cluster the topmost 3 valuable venues groups.

	Automotive Shop	Bakery	Bank	Burrito Place	Bus Line	Donut Shop	Dessert Shop	Cupcake Shop	Chees
G1	1.000000e+00	2.000000e+00	1.000000e+00	3.0	1.000000e+00	0.000000e+00	2.775558e-17	0.000000e+00	1.0000
G3	0.000000e+00	1.000000e+00	1.000000e+00	0.0	-5.551115e-17	6.938894e-18	2.000000e+00	1.000000e+00	0.0000
G4	0.000000e+00	-5.551115e-17	5.551115e-17	0.0	1.000000e+00	-6.938894e-18	0.000000e+00	0.000000e+00	0.0000
G2	4.163336e-17	5.551115e-17	5.294118e-01	0.0	2.352941e-01	1.176471e-01	0.000000e+00	2.775558e-17	4.163
G5	1.387779e-17	2.727273e-01	5.551115e-17	0.0	5.551115e-17	-6.938894e-18	2.727273e-01	1.387779e-17	1.387

fig 3.1: Best 5 groups of clustering

14	Kingsway Park South West
20	Mimico NW
32	Royal York South West
35	South of Bloor
40	The Queensway West

fig3.2:Second Best G1

22	Montgomery Road
27	Old Mill North
38	The Kingsway

fig 3.3:Second Best G3

4. Results and Discussion

Our analysis shows that although there is a great number of neighbours in Etobicoke Toronto. , there are pockets of low shopping centres and restaurant density fairly close to city center. Highest concentration of restaurants was detected Humber Bay, Humber Bay Shores, King's Mill Park, Kingsway Park South East, Mimico NE, Mimico South, New Toronto, Old Mill South Royal York South East, Sunnylea and The Queensway East.

so we focused our attention to these areas , corresponding to boroughs Etobicoke. So our attention was focused onthese centres which are more closeness to Warehouse center, strong economic point view.

Those location venues were then clustered to create zones of interest which contain greatest number of shopping centres locations. Addresses of centers of those zones were also generated using geocoding to be used as markers/starting points for more detailed local analysis based on other factors.Result of all this is 11 venues location containing largest number of potential for WArehouse's locations based on number of and distance to existing venues - both restaurants in general and all restaurants particularly. This, of course, does not imply that those zones are actually optimal locations for a new warehouse! Purpose of this analysis was to only provide info on areas close to Etobicoke Toronto.It is entirely possible that there is a very good reason for small number of Warehouse in any of those areas, reasons which would make them unsuitable for a new warehouses regardless of lack of competition in the area. Recommended zones should therefore be considered only as a starting point for more detailed analysis which could eventually result in location which has not only no nearby competition but also other factors taken into account and all other relevant conditions met.

5. Conclusion

Purpose of this project was to identify Etobicoke Toronto areas close to centre with higher number of shopping centres and restaurants in order to aid stakeholders for optimal location for a new Warehouse Centres. By calculating density of any types of shops and restaurant distribution from Foursquare data we have first identified general neighbourhoods of borough Etobicoke, Toronto that justify further analysis, and then generated extensive collection of locations which satisfy some basic requirements regarding existing nearby such location for opening new Warehouses. Clustering of those locations was then performed in order to create major zones of interest (containing greatest number of potential locations) and addresses of those zone centers were created to be used as starting points for final exploration by stakeholders.

Final decision on optimal Warehouse location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.