Battle of Neighbourhoods

Vidya Sankar.S

Jul 31, 2019

1. Introduction

1.1. Background

Immigration to US and Canada is happening at a very great rate these days. Both the countries support the immigrants in many ways. For anyone from a developing country, settling off in one among these countries is a big dream come true. In today's world, leaving one's homeland and settling off in any of the foreign countries is a very common thing.

As part of this project, I am trying to identify the best suited place among two famous cities such as New York and Toronto for a common man to move into. Both the cities have their own pros and cons.

When New York city is well known for its diversity, energy, landmarks, excitement, opportunities and convenience, Toronto is well suited for a much more peaceful life in terms of cost of living, commutation, travel and connectivity to nature.

The purpose of moving into these cities will vary for each and every person. The decision of which city to chose when it comes to a living completely depends upon the purpose.

1.2. Problem

The purpose of this project is to help people understand various factors and identify which city would be preferable and what factors to consider when planning to move in.

The internet has tons of raw data already about these two famous cities. However, it would be best when a statistical analysis is performed with the data available. It would be much more better when the findings are portrayed through user friendly charts and plots, so that one can easily understand the facts.

In this project, let us consider a restauranteur from a developing country planning to move into one among these two cities. The factors considered would be one that would help them establish their business over these places.

1.3. Interest

Obviously, this would help anyone in future who has such plans of starting off in either of the places.

2. Data Acquisition and Cleaning

2.1. Data sources

I needed data corresponding to the New York city and Toronto to proceed further. If had to collect the Postcodes along with the Borough and Neighbourhood information, for these corresponding cities. I was able to get the data for US from the Kaggle dataset here.

For Toronto, I had to scrape the data from the webpage https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M.

2.2. Data Cleaning

The NY data downloaded was perfect and required no cleaning.

However, the Toronto data scraped had to be cleaned in order to ensure that only correct data is being used for analysis.

First, there were Postcodes which did not have Borough information. I had to drop those columns as further processing would be possible only if the borough information is available.

Second, few Boroughs did not have neighbourhood values populated. For those cases, I copied the corresponding Borough as the neighbourhood as so would be the case most of the times.

Third, the Toronto data did not have the Latitude and Longitude co-ordinates for the respective Postal codes. The co-ordinates where available in the dataset here. This data was fetched and merged to the scraped data.

The data for the cities where stored in the below format in respective dataframes.

2.3. Feature selection

Once the data cleaning was done for the Toronto data, I observed that there were multiple neighbourhoods for the same Post codes. I grouped all those postcodes and joined the corresponding neighbourhoods in the same column. This reduced the size of data from (288,3) to (103,3).

Once the feature selection was done, the data looked in the below format:

Toronto Data

	Postcode	Borough	Neighbourhood	Latitude	Longitude
0	M1B	Scarborough	Rouge,Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek,Rouge Hill,Port Union	43.784535	-79.160497
2	M1E	Scarborough	Guildwood,Morningside,West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476

New York Data

]:

	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

Once data was formatted in the above way for both the cities, next task was to consider the best suited Borough among both the cities which can be used for further analysis.

Factors considered to decide the same were as follows:

Lifestyle
Proximity to tourist attraction
Weather
Value of money
Cost of Living
Processes involved in starting a new restaurant
Education

Various web pages were browsed (links can be found in **Appendix I**)to decide on the Boroughs and the following boroughs were considered for the respective cities.

Borough	**City**
Manhattan	New York
Scarborough	Toronto

Once the borough was decided, Foursquare API was used to obtain furthermore information of the two boroughs considered.

Explore API was used to collect information and land on the top 10 venues for each borough. The result was optimised using K mean clustering.

Manhattan Data:

	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	Marble Hill	Sandwich Place	Coffee Shop	Discount Store	Yoga Studio	Supplement Shop	Steakhouse	Spa	Shopping Mall	Seafood Restaurant	Clothing Store
1	Chinatown	Chinese Restaurant	Cocktail Bar	Salon / Barbershop	American Restaurant	Vietnamese Restaurant	Dim Sum Restaurant	Dumpling Restaurant	Spa	Bakery	Bubble Tea Shop
6	Central Harlem	African Restaurant	Chinese Restaurant	French Restaurant	Public Art	American Restaurant	Bar	Cosmetics Shop	Art Gallery	Seafood Restaurant	Cycle Studio
9	Yorkville	Coffee Shop	Italian Restaurant	Gym	Bar	Sushi Restaurant	Pizza Place	Deli / Bodega	Wine Shop	Japanese Restaurant	Mexican Restaurant
10	Lenox Hill	Coffee Shop	Italian Restaurant	Sushi Restaurant	Pizza Place	Sporting Goods Shop	Burger Joint	Cosmetics Shop	Gym	Gym / Fitness Center	Thai Restaurant
11	Roosevelt Island	Park	Sandwich Place	Coffee Shop	Greek Restaurant	Farmers Market	Outdoors & Recreation	Supermarket	School	Scenic Lookout	Liquor Store
12	Upper West Side	Italian Restaurant	Wine Bar	Bar	Coffee Shop	Vegetarian / Vegan Restaurant	Mediterranean Restaurant	Bakery	Cosmetics Shop	Indian Restaurant	Seafood Restaurant
16	Murray Hill	Coffee Shop	Sandwich Place	Japanese Restaurant	Hotel	French Restaurant	Bar	Gym / Fitness Center	Italian Restaurant	Gym	Mediterranean Restaurant
19	East Village	Bar	Wine Bar	Chinese Restaurant	Pizza Place	Ice Cream Shop	Mexican Restaurant	Vegetarian / Vegan Restaurant	Ramen Restaurant	Coffee Shop	Cocktail Bar
20	Lower East Side	Coffee Shop	Café	Pizza Place	Bakery	Sandwich Place	Ramen Restaurant	Art Gallery	Chinese Restaurant	Japanese Restaurant	Cocktail Bar
27	Gramercy	Bar	Italian Restaurant	Bagel Shop	American Restaurant	Pizza Place	Hotel	Thai Restaurant	Thrift / Vintage Store	Mexican Restaurant	Coffee Shop
28	Battery Park City	Park	Coffee Shop	Hotel	Memorial Site	Gym	Clothing Store	Wine Shop	Italian Restaurant	Shopping Mall	Sushi Restaurant
29	Financial District	Coffee Shop	Steakhouse	Wine Shop	Gym	Hotel	Event Space	Cocktail Bar	Gym / Fitness Center	American Restaurant	Pizza Place
30	Carnegie Hill	Coffee Shop	Pizza Place	Café	Cosmetics Shop	Yoga Studio	Japanese Restaurant	Bookstore	Bakery	French Restaurant	Grocery Store
33	Midtown South	Korean Restaurant	Hotel	Japanese Restaurant	Dessert Shop	Cosmetics Shop	Hotel Bar	American Restaurant	Coffee Shop	Italian Restaurant	Gym / Fitness Center
34	Sutton Place	Gym / Fitness Center	Italian Restaurant	Indian Restaurant	Furniture / Home Store	American Restaurant	Gym	Juice Bar	Dessert Shop	Pizza Place	French Restaurant
35	Turtle Bay	Italian Restaurant	Sushi Restaurant	Coffee Shop	Steakhouse	Wine Bar	Ramen Restaurant	Café	French Restaurant	Indian Restaurant	Park

Scarborough data:

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 M Comm Ven
Scarborough	1	Intersection	Electronics Store	Rental Car Location	Breakfast Spot	Pizza Place	Medical Center	Mexican Restaurant	Vietnamese Restaurant	Coffee Shop	Fried Chicken Joint
Scarborough	1	Coffee Shop	Korean Restaurant	Pharmacy	Vietnamese Restaurant	Clothing Store	General Entertainment	Fried Chicken Joint	Fast Food Restaurant	Electronics Store	Departm Store
Scarborough	1	Thai Restaurant	Athletics & Sports	Bakery	Bank	Fried Chicken Joint	Lounge	Caribbean Restaurant	Hakka Restaurant	Convenience Store	Grocery Store
Scarborough	1	Chinese Restaurant	Department Store	Coffee Shop	Hobby Shop	Clothing Store	Grocery Store	General Entertainment	Fried Chicken Joint	Fast Food Restaurant	Electroni Store
Scarborough	1	Bus Line	Soccer Field	Bakery	Intersection	Fast Food Restaurant	Bus Station	Park	Metro Station	Vietnamese Restaurant	Coffee Shop
Scarborough	1	General Entertainment	Skating Rink	Café	College Stadium	Vietnamese Restaurant	Clothing Store	Grocery Store	Fried Chicken Joint	Fast Food Restaurant	Electroni Store
Scarborough	1	Indian Restaurant	Pet Store	Chinese Restaurant	Latin American Restaurant	Vietnamese Restaurant	Sandwich Place	Soccer Field	Fried Chicken Joint	Fast Food Restaurant	Electroni Store
Scarborough	1	Smoke Shop	Bakery	Breakfast Spot	Middle Eastern Restaurant	Vietnamese Restaurant	Coffee Shop	General Entertainment	Fried Chicken Joint	Fast Food Restaurant	Electroni Store
Scarborough	1	Clothing Store	Skating Rink	Breakfast Spot	Lounge	Vietnamese Restaurant	General Entertainment	Fried Chicken Joint	Fast Food Restaurant	Electronics Store	Departm Store
Scarborough	1	Pizza Place	Fried Chicken Joint	Thai Restaurant	Bank	Fast Food Restaurant	Italian Restaurant	Pharmacy	Noodle House	Chinese Restaurant	Vietname Restaura
Scarborough	1	Fast Food Restaurant	Chinese Restaurant	Indian Restaurant	Pharmacy	Cosmetics Shop	Coffee Shop	Grocery Store	Bubble Tea Shop	Nail Salon	Pizza Pla

These data were further used to perform the ERD and come up with a solution.

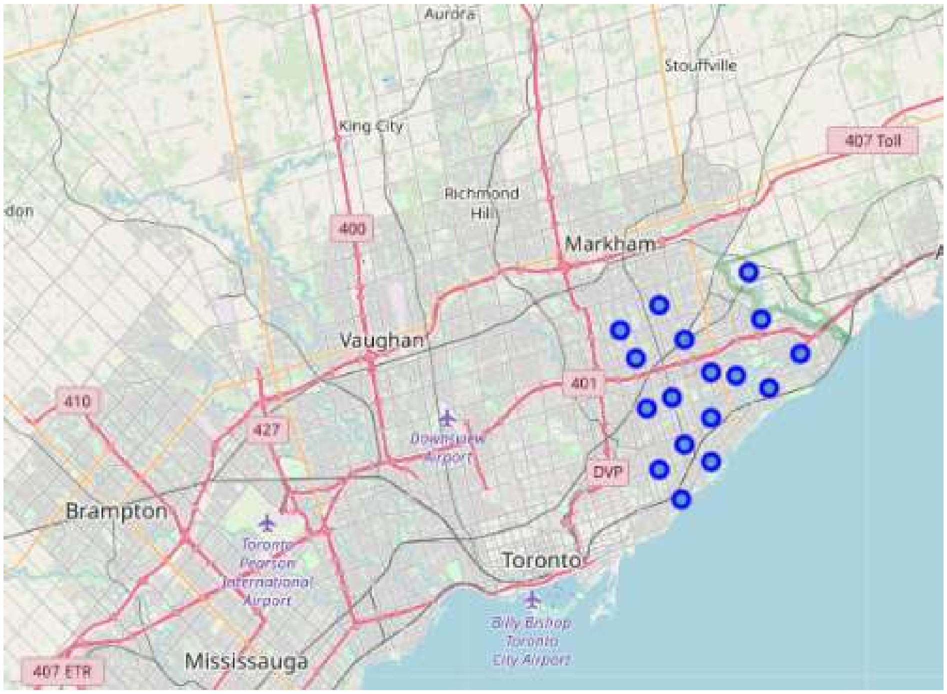
3. Exploratory Data Analysis:

3.1. Identification of neighbourhood within the boroughs:

For further analysis, it was required to identify the nearby venues in each of the borough. The Foursquare API was used to identify the nearby venues for each of the borough considered.

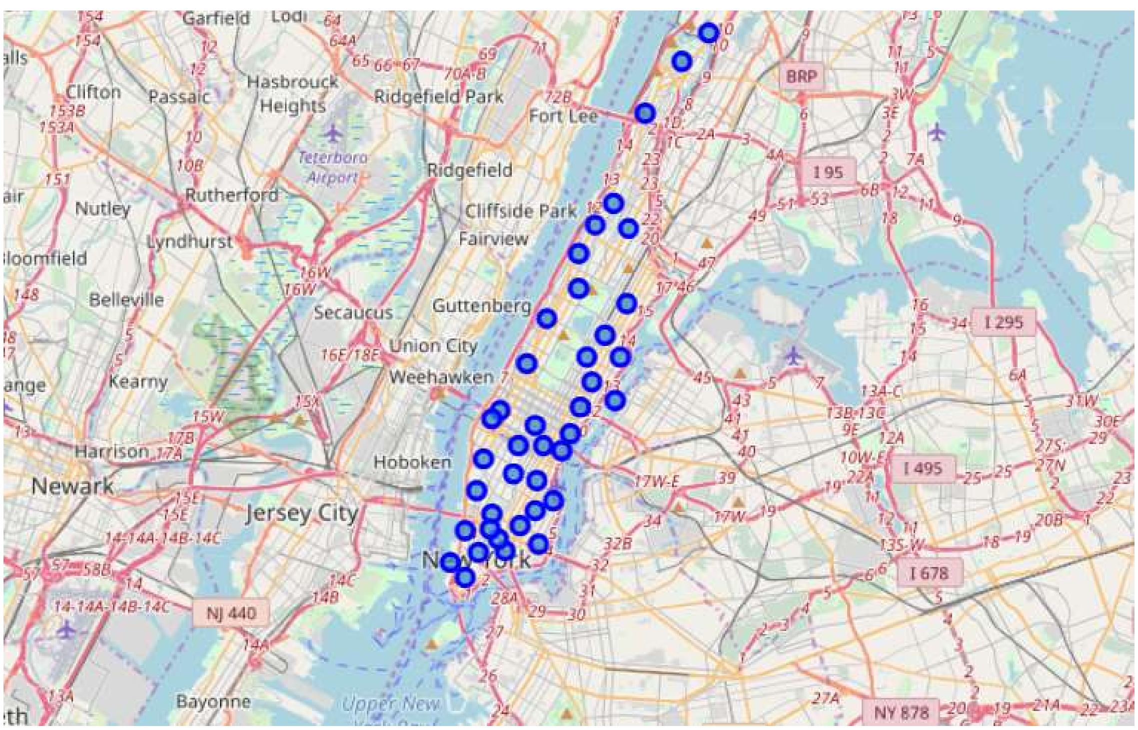
3.1.1 Neighbourhood of Scarborough

The localities identified for Scarborough can be found in **Appendix IV**



3.1.2 Neighbourhood of Manhattan

The places in the neighbourhood of Manhattan can be found in **Appendix V**

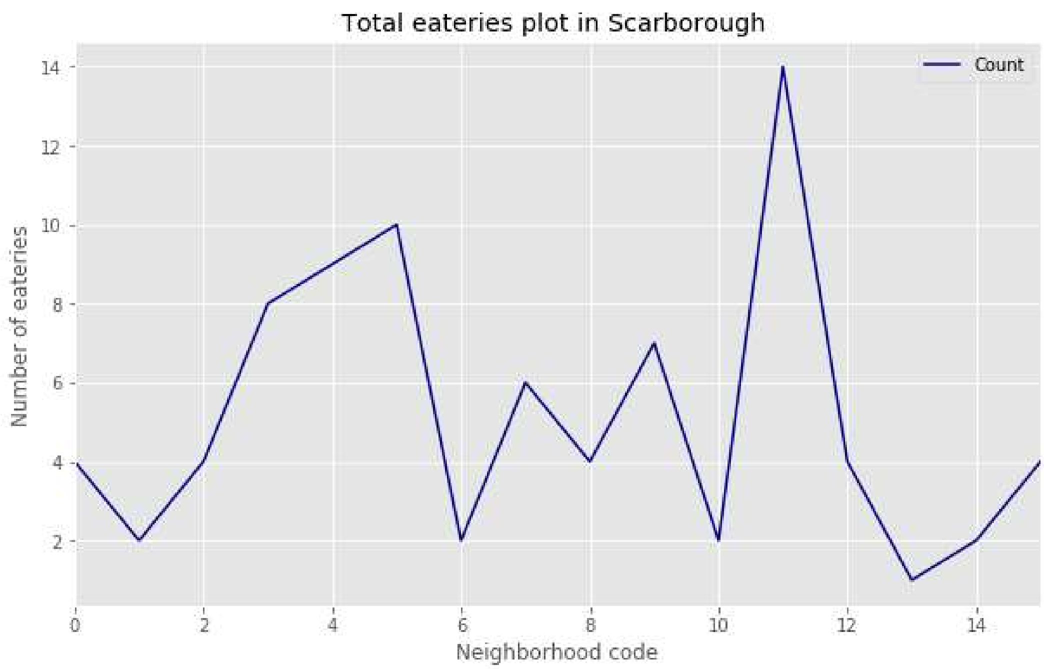


3.2. Identification of venues within neighbourhoods:

Once the neighbourhoods were identified, the next step was to identify the landmark venues within each neighbourhood. This was as well achieved using Foursquare API.

3.2.1. Venues for Scarborough grouped by category.

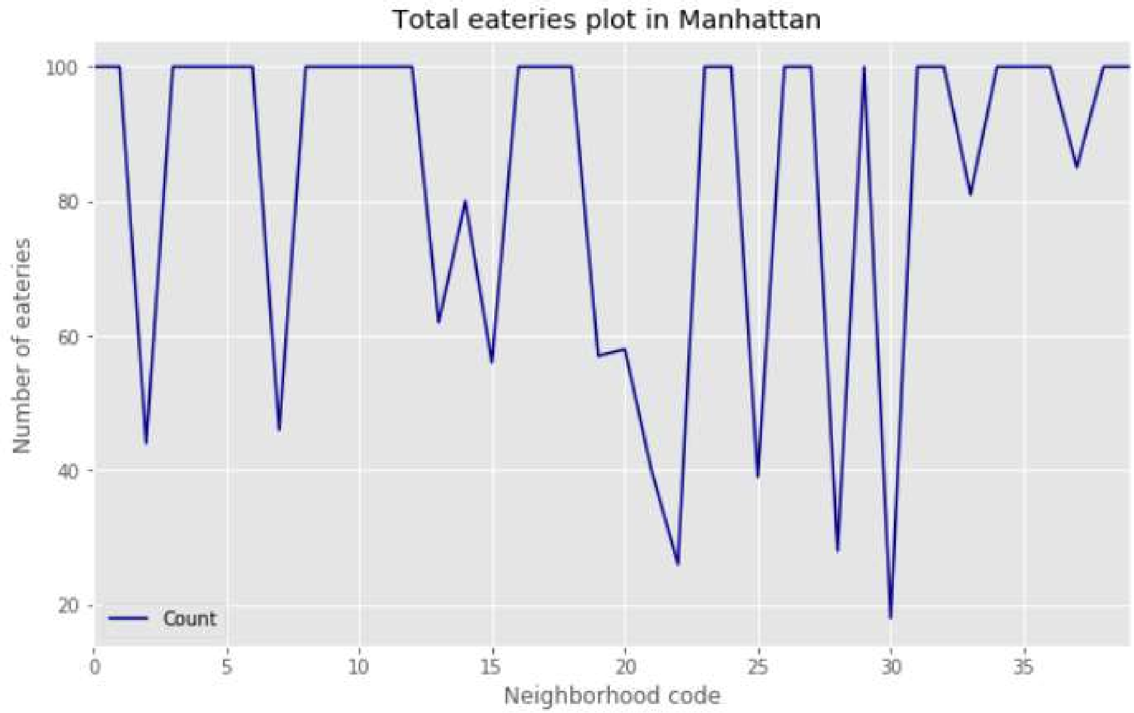
The food destination venues were identified for each of the neighbourhoods and are as follows:



The neighbourhood names, corresponding codes and number of eateries list can be found in **Appendix VI**

From the above plot, it is evident that L'Amoreaux West has the maximum number of eateries and Rouge, Malvern has the least number of eateries.

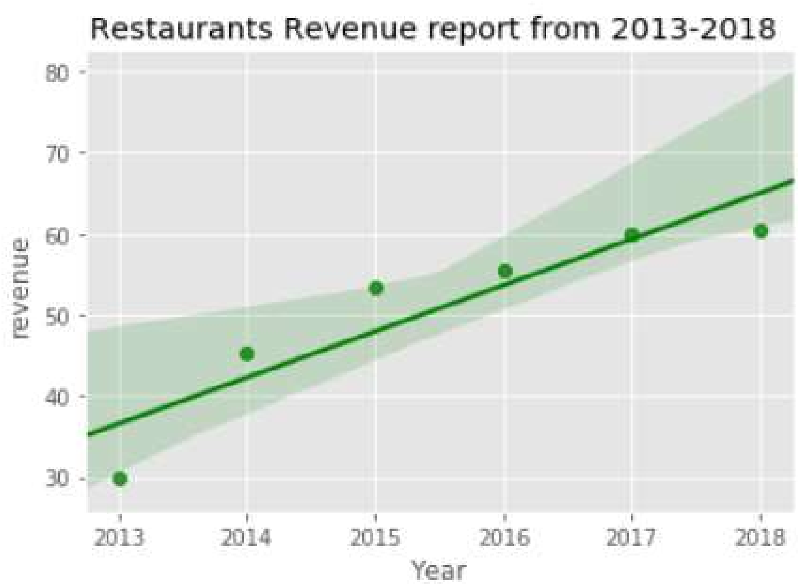
3.2.2. Venues for Manhattan grouped by category.



3.3. Revenue statistics from 2013 – 2018

3.3.1. Scarborough analysis:

Appendix II has the revenue details from the period 2013 – 2018 for in the Scarborough neighbourhood.



The report shows a positive linear relationship between the revenue and the year which is a positive sign for the business over Scarborough.

3.3.2. Manhattan analysis:

Appendix VIII has the revenue details from the period 2013 – 2018 for in the Manhattan neighbourhood.

The report shows a negative linear relationship across Revenue and Year over a period of time which is not a good sign to start off a new restaurant at this place

4. Predictive model:

There are two types of models, regression and classification, that can be used to determine the ideal location to start a new restaurant. Regression models can provide additional information on the amount of improvement while classification models focus on the probabilities that a new restaurant might thrive. The underlying algorithms are similar between regression and classification models, but different audience might prefer one over the other. For Eg: Regression model can be used to predict how the revenue would increase in the following years whereas classification model can be used to classify the factors within a neighbourhood which would impact the eat out proportion which in turn will increase the revenue.

4.1. Regression Model

I applied simple linear regression to predict how the outcast of revenue would be in each locality.

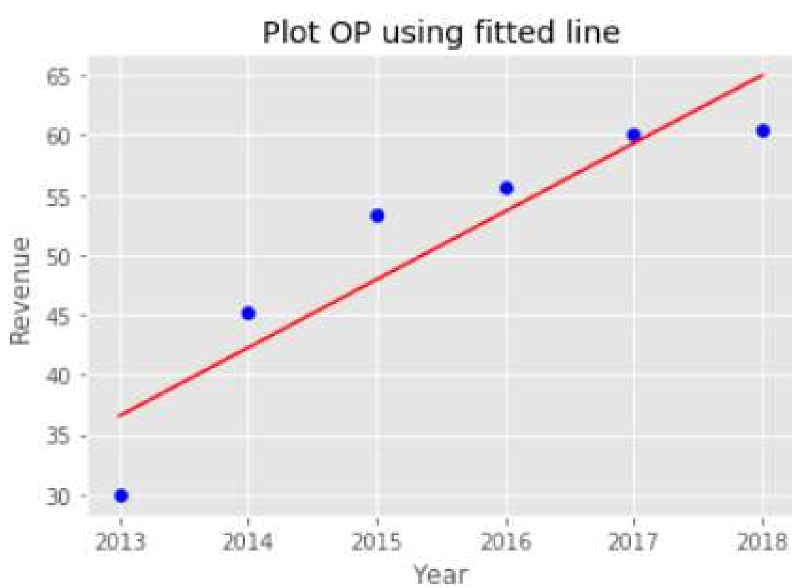
4.1.1. Scarborough Model

The linear formula derived for Scarborough was as follows:

5.68857143 -11414.53238095X

Where, X is the year.

The plot output using the fitted line is as below:



The values were predicted using the derived formula. The performance of the regression model was evaluated using the following metrics.

Mean Absolute Error	Residual Sum of Squares(MSE)	R2- score

3.69	17.71	0.81

The R-square being close to 1 shows that the predicted value is good.

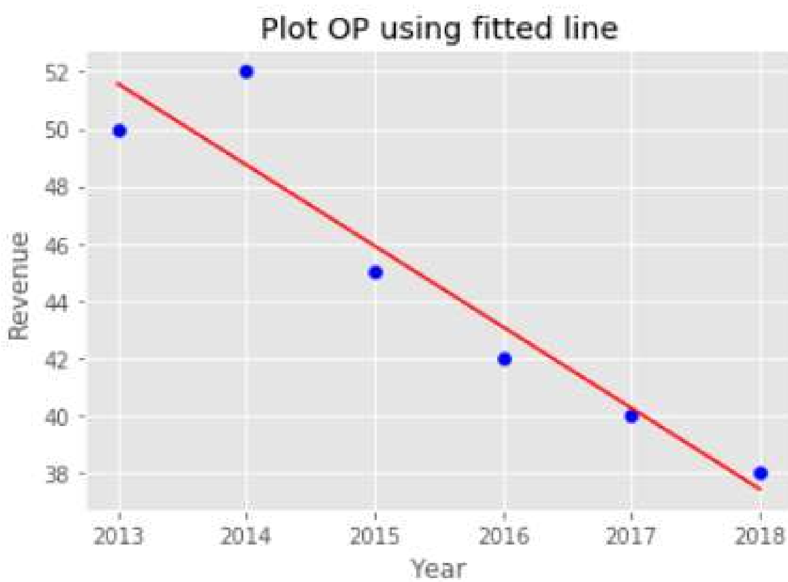
4.1.2. Manhattan Model

The linear formula derived for Scarborough was as follows:

5745.48571429 - 2.82857143X

Where, X is the year.

The plot output using the fitted line is as below:



The values were predicted using the derived formula. The performance of the regression model was evaluated using the following metrics.

Mean Absolute Error	Residual Sum of Squares(MSE)	R2- score
1.28	2.58	0.89

The R-square being close to 1 shows that the predicted value is good.

4.2. Classification Model

I applied K-nearest neighbour to predict the proportion of people who prefer to eat out.

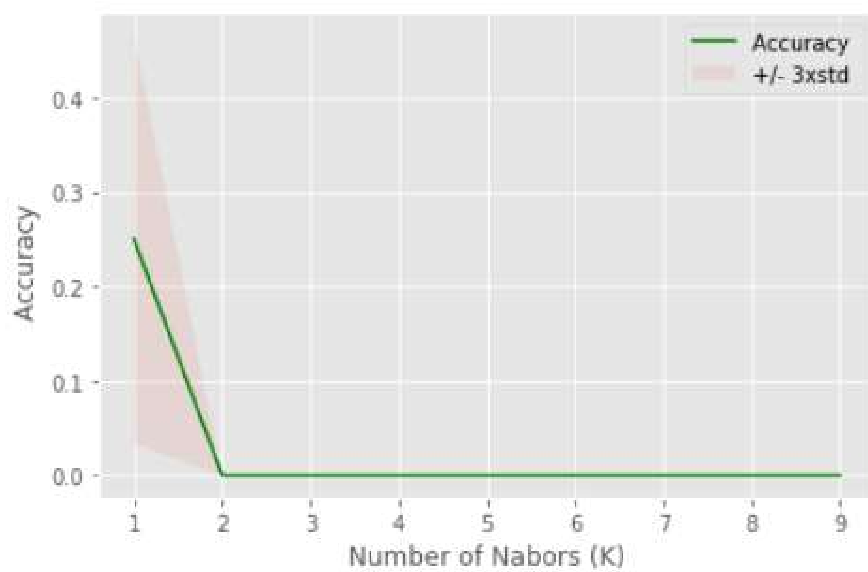
The data consisted of independent variables such as Neighborhood name, Neighborhood code, availability of organic products, Percentage of people working.

The dependent variable was category of the customer which might be one among the 3 values

CUSTCAT	Legend
1	People rarely eat out
2	People occasionally eat out
3	People eat out frequently

4.2.1. Scarborough Data

The best accuracy was 0.25 with K = 1.

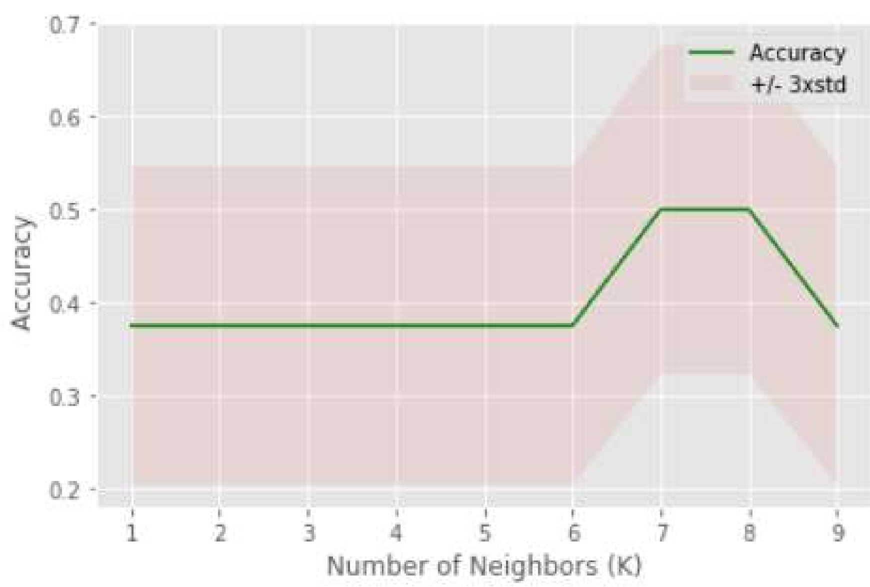


The values were predicted with K = 1 and the result was evaluated is using following metrics

Jaccard index	F1 score
1.0	1.0

4.2.2. Manhattan Data

Appendix IX has the document with data for classification analysis.
The best accuracy was 0.5 with K = 7.



The values were predicted with K = 7 and the result was evaluated is using following metrics

Jaccard index	F1 score
0.625	0.5892857142857143

Using the above prediction method, it was observed that more the availability of organic produce and more the percentage of working people, the tendency to eat out was more which was a positive sign for starting a new restaurant.

Looking at the data, the ideal neighbourhood in the respective boroughs are as follows.

Borough	Neighborhood
Manhattan	Marble Hill
Scarborough	Rouge, Malvern

5. Conclusions

In this study, I analysed various data pertaining to Scarborough and Manhattan. It was identified that in terms of revenue Scarborough is the ideal place to start off with. Considering other factors, Rouge, Malvern is the ideal neighbourhood in Scarborough to set up the restaurant.

6. Future Directions

The analysis I performed considered few factors like how much people are habituated to eat out, availability of organic produce etc. Further more analysis can be considered to optimise the results much more. Some of the factors worth considering when moving to a new city are cost of living, education, type of people etc. Similar analysis can be performed with such data as well.

Appendix I

https://www.timeout.com/newyork/things-to-do/reasons-manhattan-is-the-best-borough

https://www.quora.com/What-is-Manhattan-NYC-borough-known-for

https://en.wikipedia.org/wiki/Scarborough,_Toronto

https://moving2canada.com/where-to-live-in-toronto-neighbourhoods/

Appendix II

Scarborough restaurant revenue report from 2013 - 2018

scarborough_revenue
.csv

Appendix III

Eat-out preferences of Scarborough people

Scarboroug_class.csv

Appendix IV

Rouge,Malvern
Highland Creek,Rouge Hill,Port Union
Guildwood,Morningside,West Hill
Woburn

Cedarbrae
Scarborough Village
East Birchmount Park,Ionview,Kennedy Park
Clairlea,Golden Mile,Oakridge
Cliffcrest,Cliffside,Scarborough Village West
Birch Cliff,Cliffside West
Dorset Park,Scarborough Town Centre,Wexford Heights
Maryvale,Wexford
Agincourt
Clarks Corners,Sullivan,Tam O'Shanter
Agincourt North,L'Amoreaux East,Milliken,Steeles East
L'Amoreaux West
Upper Rouge

Appendix V

Marble Hill
Chinatown
Washington Heights
Inwood
Hamilton Heights
Manhattanville
Central Harlem
East Harlem
Upper East Side
Yorkville
Lenox Hill
Roosevelt Island
Upper West Side
Lincoln Square
Clinton
Midtown
Murray Hill
Chelsea
Greenwich Village
East Village
Lower East Side
Tribeca
Little Italy
Soho
West Village
Manhattan Valley

Morningside Heights
Gramercy
Battery Park City
Financial District
Carnegie Hill
Noho
Civic Center
Midtown South
Sutton Place
Turtle Bay
Tudor City
Stuyvesant Town
Flatiron
Hudson Yards

Appendix VI

Neighborhood	Count	ncode
Agincourt	4	0
Agincourt North,L'Amoreaux East,Milliken,Steeles East	2	1
Birch Cliff,Cliffside West	4	2
Cedarbrae	8	3
Clairlea,Golden Mile,Oakridge	9	4
Clarks Corners,Sullivan,Tam O'Shanter	10	5
Cliffcrest,Cliffside,Scarborough Village West	2	6
Dorset Park,Scarborough Town Centre,Wexford Heights	6	7
East Birchmount Park,Ionview,Kennedy Park	4	8
Guildwood,Morningside,West Hill	7	9
Highland Creek,Rouge Hill,Port Union	2	10
L'Amoreaux West	14	11
Maryvale,Wexford	4	12
Rouge,Malvern	1	13
Scarborough Village	2	14
Woburn	4	15

Appendix VII

Neighborhood	Count	ncode
Battery Park City	100	0
Carnegie Hill	100	1
Central Harlem	44	2
Chelsea	100	3
Chinatown	100	4
Civic Center	100	5
Clinton	100	6
East Harlem	46	7
East Village	100	8
Financial District	100	9
Flatiron	100	10
Gramercy	100	11
Greenwich Village	100	12
Hamilton Heights	62	13
Hudson Yards	80	14
Inwood	56	15
Lenox Hill	100	16
Lincoln Square	100	17
Little Italy	100	18
Lower East Side	57	19
Manhattan Valley	58	20
Manhattanville	40	21
Marble Hill	26	22
Midtown	100	23
Midtown South	100	24

Morningside Heights	39	25
Murray Hill	100	26
Noho	100	27
Roosevelt Island	28	28
Soho	100	29
Stuyvesant Town	18	30
Sutton Place	100	31
Tribeca	100	32
Tudor City	81	33
Turtle Bay	100	34
Upper East Side	100	35
Upper West Side	100	36
Washington Heights	85	37
West Village	100	38
Yorkville	100	39

Appendix VIII

Manhattan restaurant revenue report from 2013 - 2018

Manhattan_revenue.csv

Appendix IX

Manhattan eat-out preferences of people

Manhattan_class.csv