

Applied Data Science Capstone Project

Recommending location for a new Restaurant through Data Analysis

Nithin John Jacob

26 April 2020

This report presents the details of recommending an ideal location for an Indian restaurant in Toronto, Canada by analyzing the demographics and neighborhood of various boroughs of Toronto

Contents

1.	Introduction	3
1.1.	Business Problem.....	3
1.2.	Target stakeholders	3
2.	Literature Review	3
3.	Data.....	4
3.1.	Toronto Demographics Data.....	4
3.2.	Geographic data of Toronto neighbourhoods.....	4
3.3.	Foursquare data.....	4
4.	Methodology	5
4.1.	Exploratory Data Analysis	5
4.1.1.	Statistics	5
4.1.2.	Correlation of features	8
4.1.3.	Data in Maps	11
4.1.	Machine Learning Algorithm	12
5.	Results.....	16
6.	Discussion	17
7.	Conclusion	17

Table of Tables

Table 1: Correlation of Ethnic group population with number of Indian Restaurants.....	9
Table 2: Results of K Means Clustering.....	13
Table 3: Cluster containing Woburn	14
Table 4: Recommended places for starting Indian Restaurant	16

Table of Figures

Figure 1: Population by ethnic background.....	6
Figure 2: Number of Restaurants of different cuisines	6
Figure 3: Population Distribution of Top 10 ethnic groups	7
Figure 4: Population of Indians and No. of Indian Restaurants against Neighborhoods	7
Figure 5: Correlation of top 20 restaurants and ethnic groups.....	8
Figure 6: Correlation between ethnic groups.....	10
Figure 7: Correlation between Features.....	10
Figure 8: Heat map of Population of Indians in Toronto	11
Figure 9: Restaurants of Toronto	11
Figure 10: Indian Restaurants in Toronto	12
Figure 11: Groups clustered as per demographics on [Left] vs Groups clustered as per Restaurant Profile [Right]	13
Figure 12: Mesh points in selected neighbourhoods, clustered to 20 groups	15
Figure 13: Centres of Groups, recommended locations.....	15

1. Introduction

Setting up a new restaurant is a huge investment. If you're thinking about opening a restaurant, location is everything. Experts often argue a bad location is one of the biggest, if not THE biggest, reason restaurants fail. Accessibility, Parking facility, Transportation, Size of restaurant, Competition, Labor Costs, the Minimum Wage and target customers are some of the parameters to be checked.

We will be focusing on the “target customers “ part, i.e. the demographics of various neighbourhoods of Toronto for this. We will be using various machine learning tools to analyse the data to identify a location. This report maybe useful for all stakeholders who would like to open a restaurant serving Indian cuisine in Toronto.

1.1. Business Problem

This project aims to find an ideal location for a new Indian restaurant in Toronto, Canada by interpreting the data acquired from various sources.

1.2. Target stakeholders

All restaurateurs interested in opening an Indian restaurant in Toronto. The data acquisition and research done in this project may provide useful insights in determining a location for their new restaurant.

2. Literature Review

Various researches have emphasized the importance of location over and again.

Some of the studies can be found at

<https://study.com/academy/lesson/location-analysis-for-the-restaurant-industry.html>

<https://www.thebalancesmb.com/choosing-restaurant-location-2888543>

<https://www.restaurantindia.in/article/5-smart-ways-to-choose-best-location-for-your-restaurant.9444>

https://www.researchgate.net/publication/301578175_Restaurant_Location_and_Price_Fairness_as_Key_Determinants_of_Brand_Equity_A_Study_on_Fast_Food_Restaurant_Industry

<https://www.webstaurantstore.com/article/81/restaurant-environmental-analysis.html>

3. Data

Different data sets are used in this project

3.1. Toronto Demographics Data

This data is obtained from the open Toronto portal. The csv file tabulates various statistics of all 140 neighbourhoods. Like, ethnicity, population, area , average income, etc. we use the average income ethnicity distribution and area of all the 140 neighbourhoods. This is downloadable from this link. <https://open.toronto.ca/dataset/neighbourhood-profiles/>.

3.2. Geographic data of Toronto neighbourhoods

This file was made from the Neighbourhood information from Namara.io and Toronto's Open Data portal. This Geojson file contains boundaries of City of Toronto Neighbourhoods in latitude and longitudes.. The Centroid of the neighbourhoods and distance of centroid to each of the boundary points of the neighbourhood is calculated and taulated

This data is available at <https://github.com/jasonicarter/toronto-geojson> .

3.3. Foursquare data

The information regarding the restaurants in the neighbourhood is obtained using the foursquare venue search API. Since Foursquare AP returns only 50 venues at a time , in order to collect maximum data , the request is made three times with varying(near, average, far) search radius for each neighbourhood. The results are combined and cleaned, ensuring that there is no repetition of venues and that each venue belonged to their corresponding neighbourhood. This data is made to a data frame containing, location information like

latitude, longitude, address, name and category of each venue and their corresponding neighbourhood details.

4. Methodology

In this project, the relationship between various features of data set and the number of Indian restaurants in neighbourhood is explored. Pearson correlation is used to evaluate the relationships

The optimum neighbourhoods to consider is identified using clustering technique. The neighbourhoods are grouped as per their demographics and also as per their restaurant profile i.e. the number of restaurants of different kinds in each neighbourhood. The ideal neighbourhoods are selected from this groups using the population of Indians data.

Once the candidate neighborhoods are selected , a mesh of points 200*200m in these neighbourhoods is created. All points that are close to an Indian restaurant are removed from the mesh. The remaining points of the mesh are divided to 20 groups using k means cluster and get the address of the centres of these groups are obtained by using geopy. This will be the starting points given to stake holders for further exploration.

4.1. Exploratory Data Analysis

4.1.1. Statistics

The population of ethnic groups in all neighbourhoods are summed and sorted in descending order to get the bar chart as shown in Figure 1. The largest ethnic group is the Chinese closely followed by English and Canadian. Indians occupy the sixth position. Italians occupy the seventh position but their population just about half of the Chinese population.

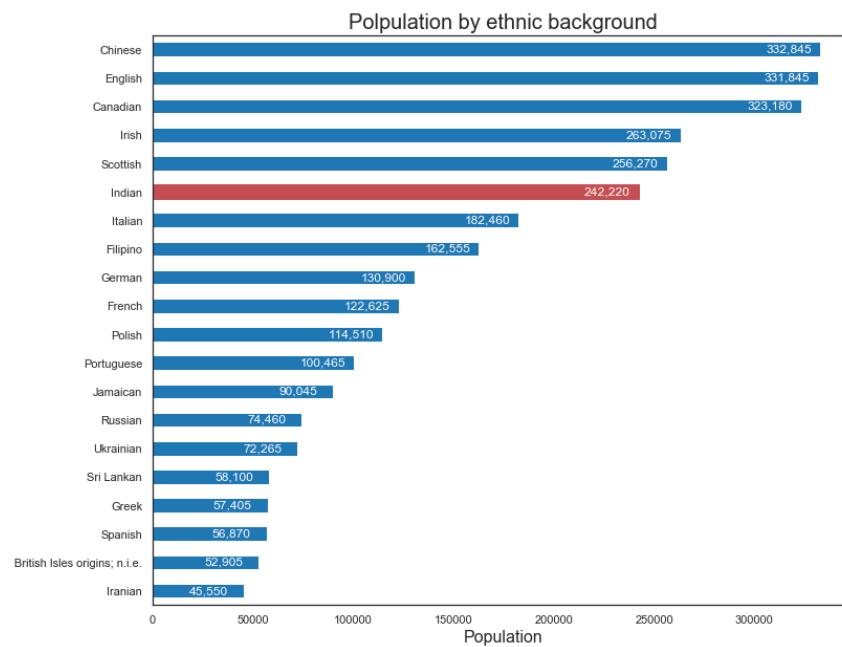


Figure 1: Population by ethnic background

The same process is done for the number of restaurants of different cuisines data set and the graph is shown in Figure 2

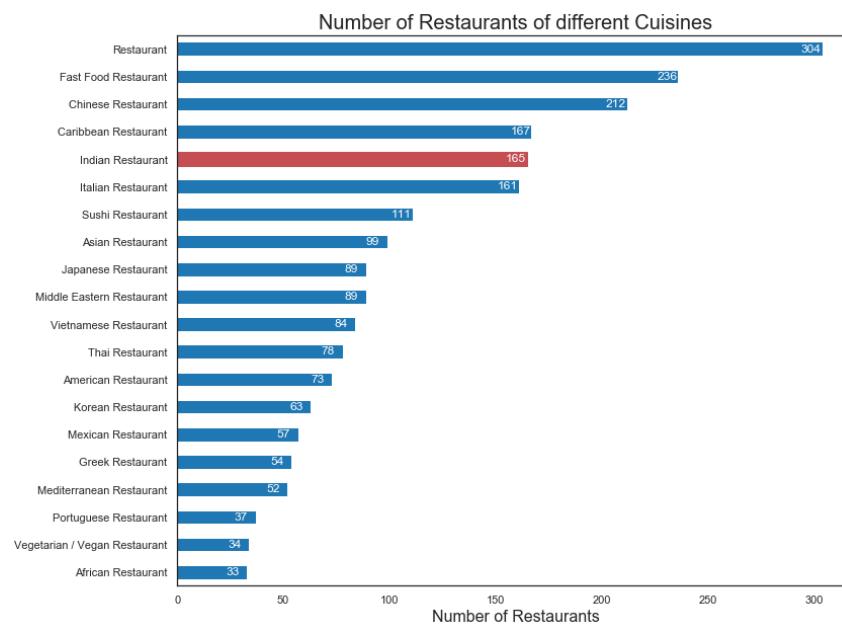


Figure 2: Number of Restaurants of different cuisines

The box plot shown in Figure 3 gives insights about the distribution of population of top 10 ethnic groups. Most groups have a lot of outliers.

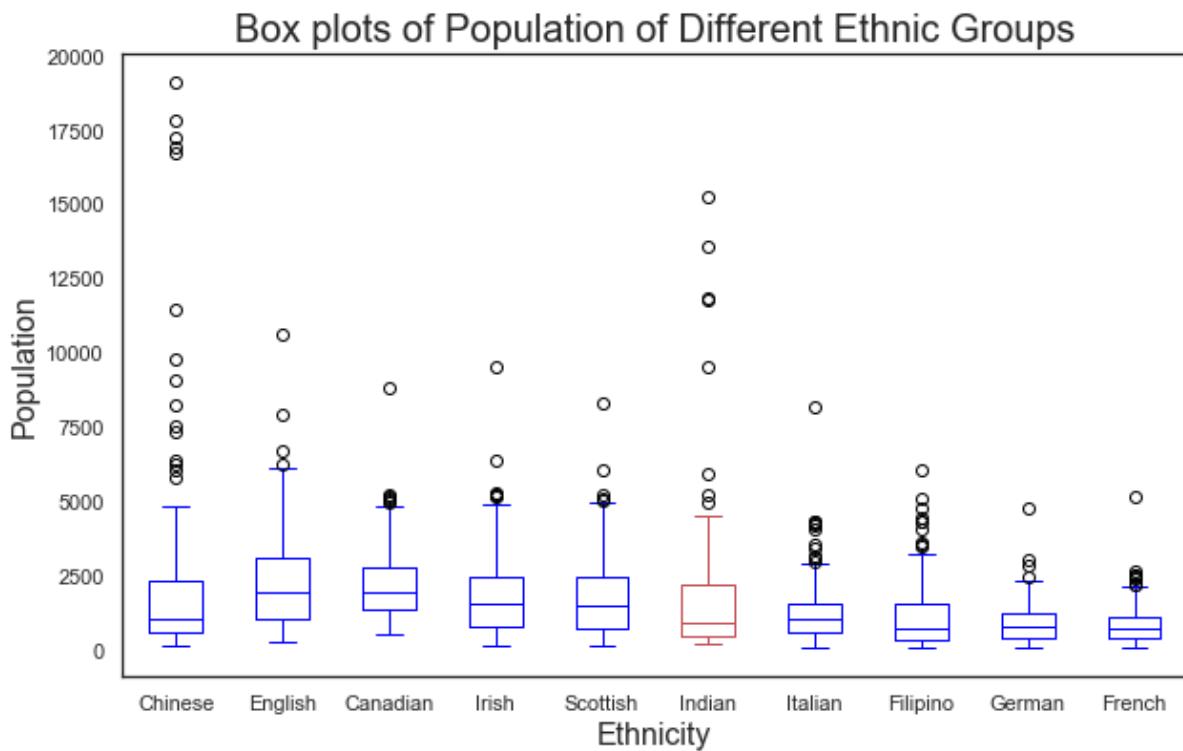


Figure 3: Population Distribution of Top 10 ethnic groups

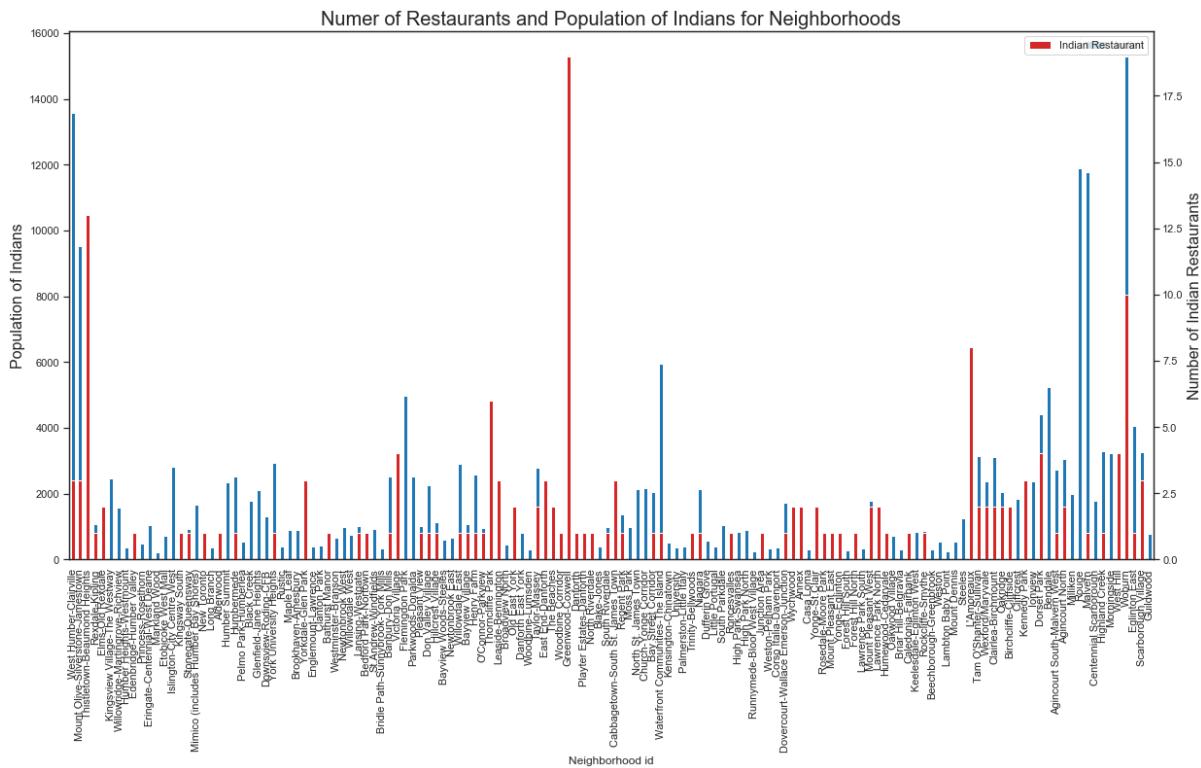


Figure 4: Population of Indians and No. of Indian Restaurants against Neighborhoods

The population of Indians and the number of Indian restaurants of neighbourhoods are plotted as bar chart and shown in Figure 4

4.1.2. Correlation of features

The Pearson correlation between top 20 types of restaurants and top 20 ethnic groups are calculated and shown in Figure 5.



Figure 5: Correlation of top 20 restaurants and ethnic groups

The correlation of Population of various ethnic groups with Number of Indian restaurants is calculated and sorted in descending order and top 20 are tabulated in Table 1.

Table 1: Correlation of Ethnic group population with number of Indian Restaurants

SINo	Feature	Correlation with Indian Restaurants
1	Nova Scotian	0.355933
2	Other Asian origins	0.332107
3	Pakistani	0.290206
4	Fijian	0.289514
5	Indian	0.284673
6	Bavarian	0.280163
7	Afghan	0.269483
8	Kittitian/Nevisian	0.244602
9	Guinean	0.243400
10	Macedonian	0.227512
11	Sicilian	0.225071
12	Arab; n.o.s	0.224620
13	Armenian	0.220627
14	Berber	0.215484
15	Congolese	0.210837
16	Arawak	0.209361
17	Syrian	0.208957
18	Togolese	0.208182
19	Newfoundland	0.205722
20	Pashtun	0.199409

The Correlation values are low. This shows poor relationship between the features.

The correlation between these 20 ethnic groups is shown in Figure 6. The population of Afghans and Pakistanis are correlated. There is also a small relationship between population of Indians and Other Asian origins group n.i.e. There is also a small relationship between population of Pakistanis and Other Asian origins group n.i.e

The Average income, Total population, Area and total number of restaurants show poor correlation. The calculated results are shown in Figure 7.

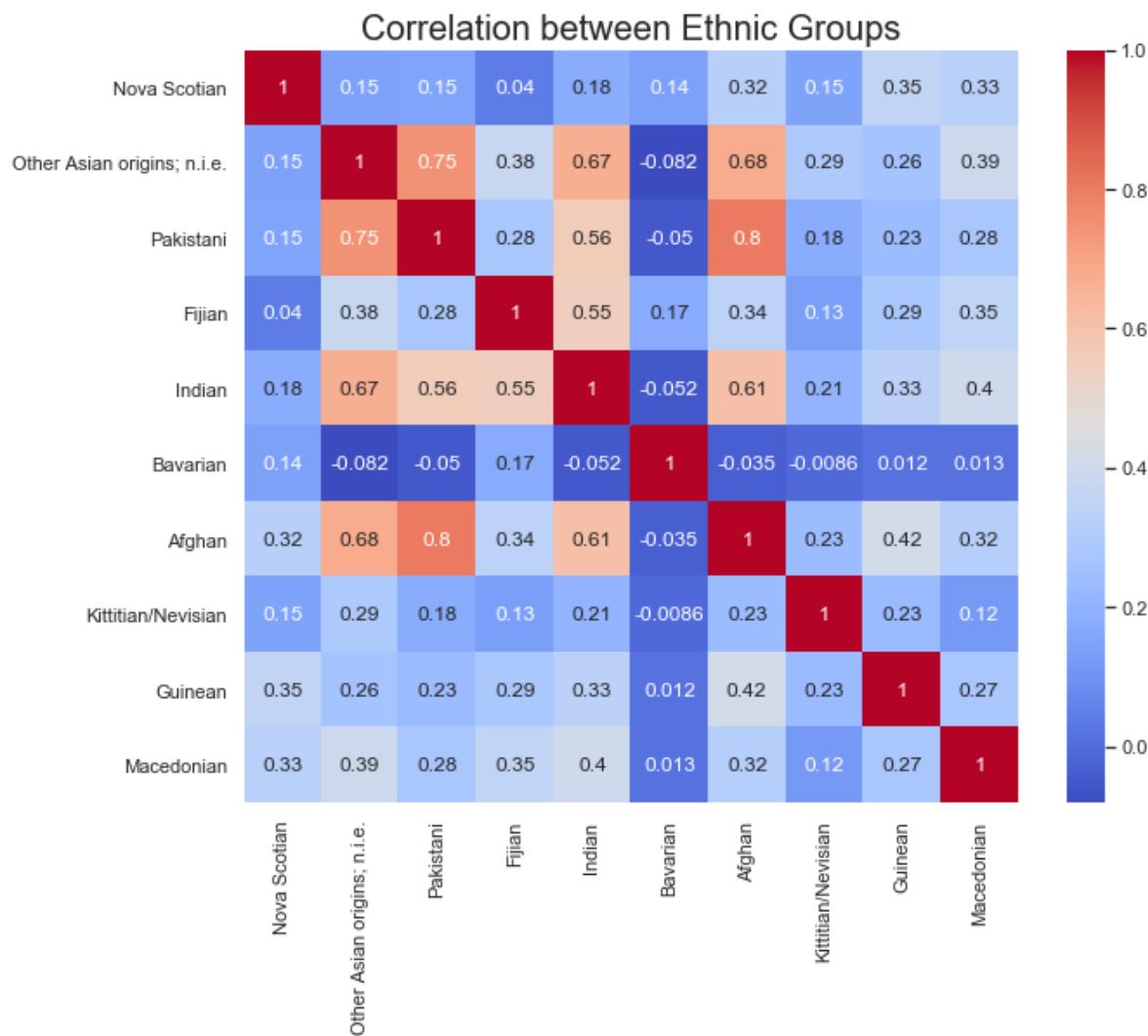


Figure 6: Correlation between ethnic groups

Correlation between Features

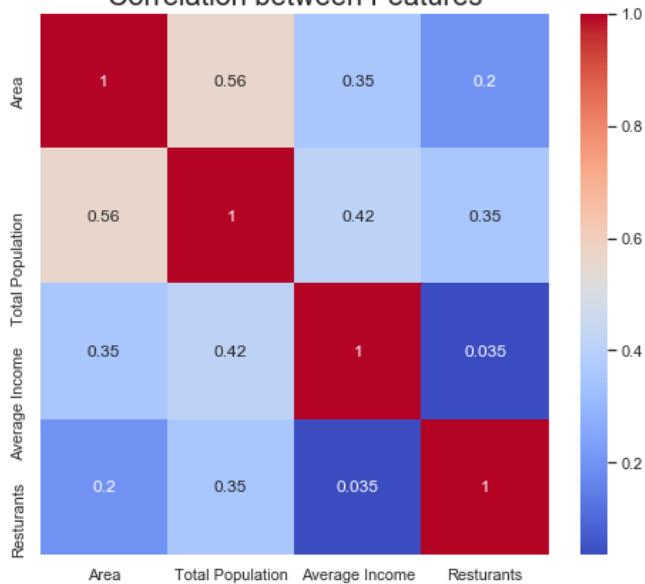


Figure 7: Correlation between Features

4.1.3. Data in Maps

The heat maps of population of Indians in Toronto is shown in Figure 8.

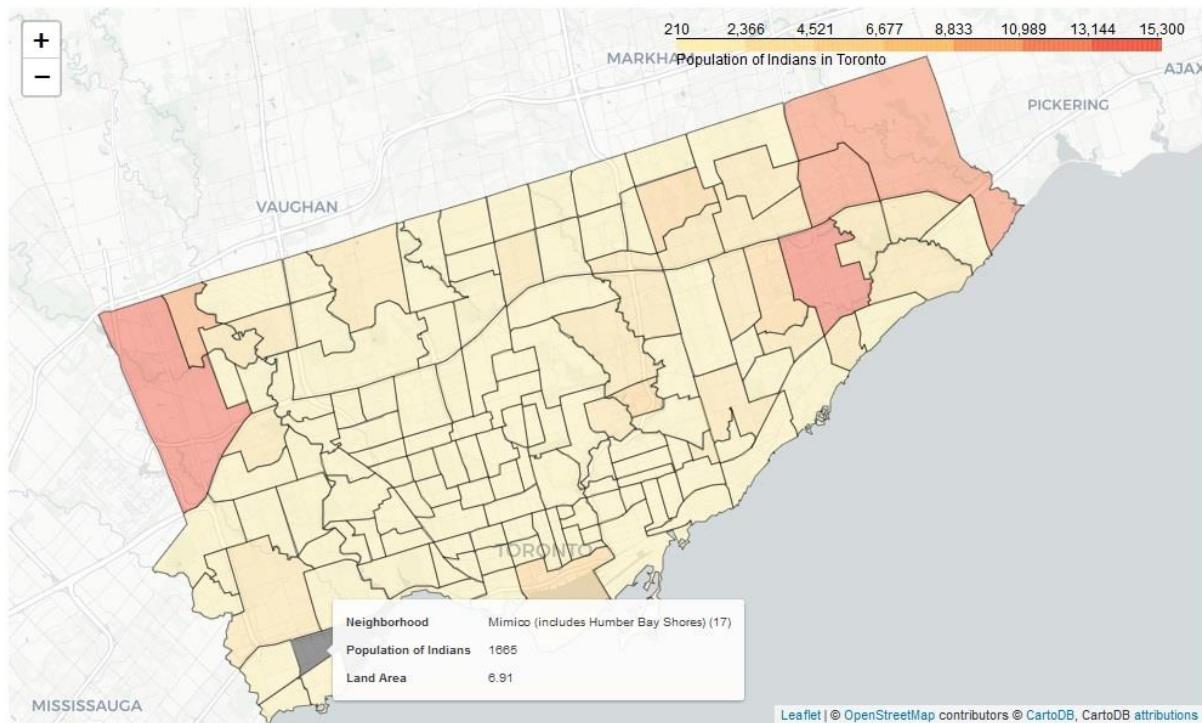


Figure 8: Heat map of Population of Indians in Toronto

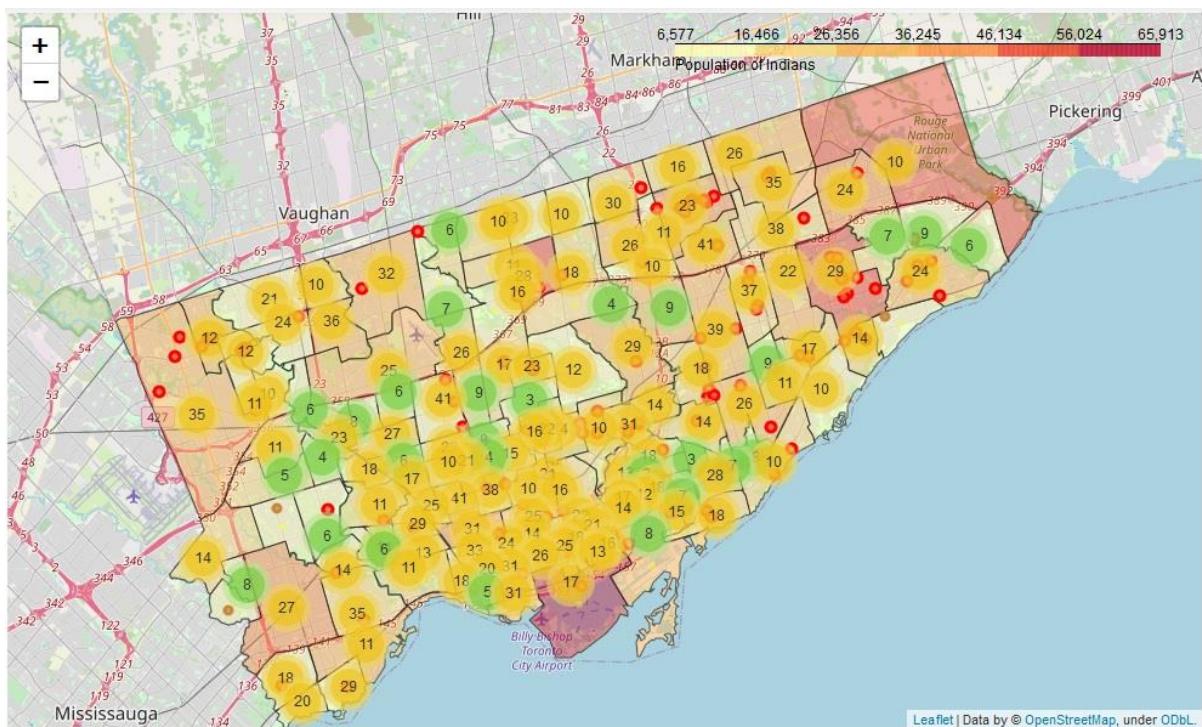


Figure 9: Restaurants of Toronto

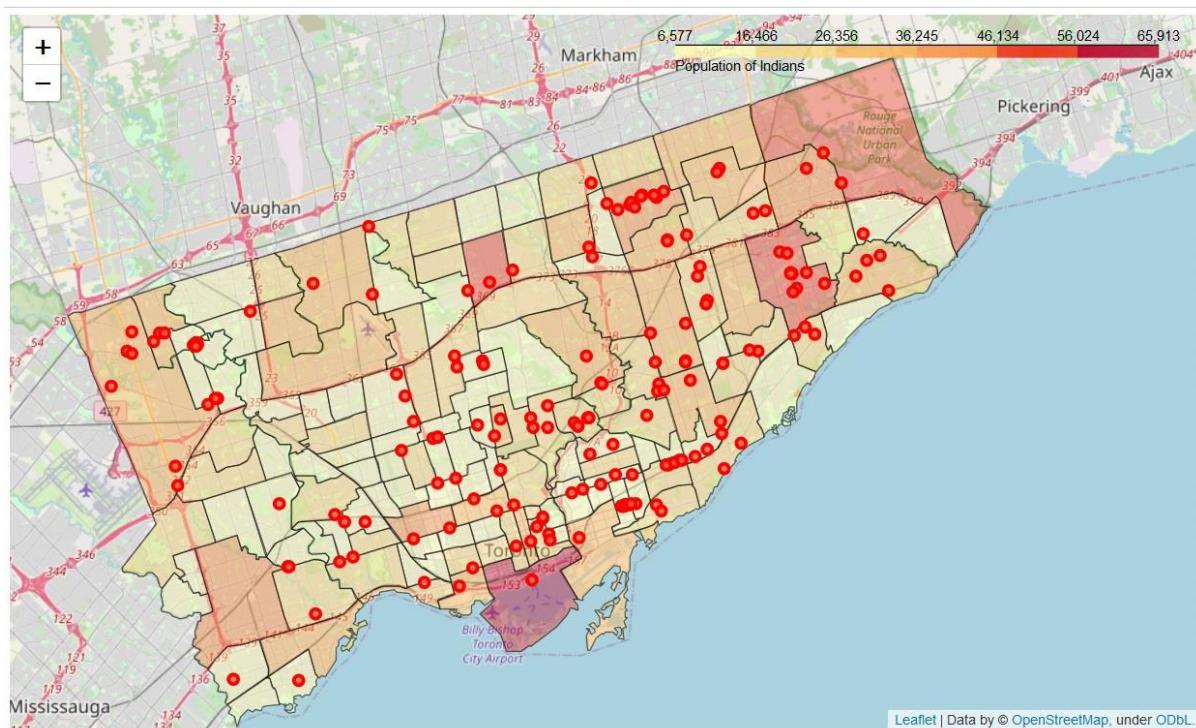


Figure 10: Indian Restaurants in Toronto

The Number of restaurants in Toronto are clustered into small groups to give better readability. The number shown on the icon indicates the number of restaurants in its vicinity. The Figure 9 shows the location of all 2605 restaurants given by foursquare API in the map.

The Indian restaurants are shown as red dots and shown in map in Figure 10. There are 165 restaurants in Toronto. The maximum number is nineteen in Greenwood-coxwell , Then thirteen in Thistletown-Beaumont Heights and ten in Woburn and eight in L'Amoreaux and six in Thorncliffe Park. The little India in Greenwood-Coxwell explains the large number of restaurants there.

4.2. Machine Learning Algorithm

We use the data to divide the neighbourhood into 10 groups. We use the features of demographics ,i.e the population different ethnic groups to do so. We use the standard scaler od preprocessing to fit and transform the data. We use n_init as 30 and init as k-means++ for KMeans algorithm. Once the groups are identified we plot it in map. The

neighbourhoods belonging to same groups are shaded with same color, Ten different groups can be seen on the map.

Similar process is done with features of restaurant profile data i.e. number of restaurants of different cuisines per neighbourhood. The groups are plotted like before. The two clustering results are plotted side by side and shown in Figure 11.

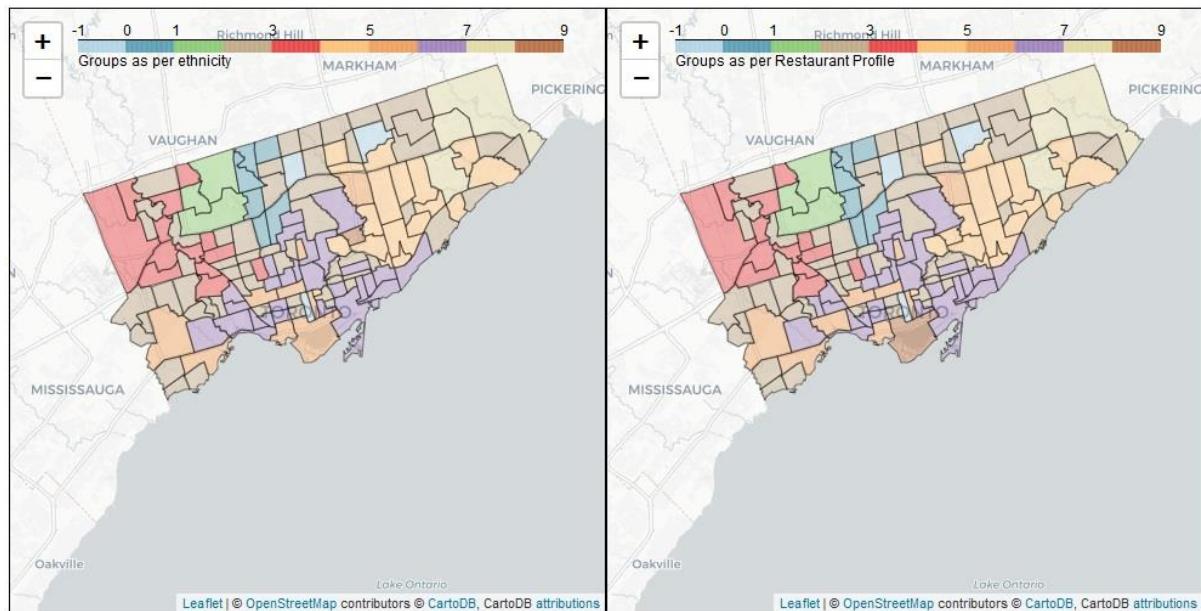


Figure 11: Groups clustered as per demographics on [Left] vs Groups clustered as per Restaurant Profile [Right]

Table 2: Results of K Means Clustering

Neighborhood id	Neighborhood	Indian Restaurant	Restaurant Cluster	Demographics Cluster	Indian Population
65	Greenwood-Coxwell	19	3	3	725
3	Thistletown-Beaumont Heights	13	3	1	2485
137	Woburn	10	9	9	16390
117	L'Amoreaux	8	1	0	4765
55	Thorncliffe Park	6	1	0	5020
136	West Hill	4	1	5	3480
126	Dorset Park	4	1	1	4800
43	Victoria Village	4	1	0	2230
1	West Humber-Clairville	3	5	5	14125
31	Yorkdale-Glen Park	3	3	1	820

The neighbourhood containing maximum population of Indians and their cluster groups are tabulated in Table 2.

The third row, neighbourhood Woburn is identified as a good candidate because it has large population of Indians. We add all neighborhoods in the same cluster as Woburn to candidate list. These are our neighbourhoods to be considered for further processing.

The neighbourhoods in same cluster as Woburn is shown in Table 3.

Table 3: Cluster containing Woburn

Neighborhood id	Neighborhood	Indian
131	Rouge	12580
132	Malvern	12640
137	Woburn	16390

The mesh points in the selected neighbourhoods are shown in Figure 12.

Again Kmeans clustering technique is used to cluster these points into twenty groups. The centroids of these clusters are also obtained. The points are plotted in Figure 13.

The geopy module is used to get address from centroid points. These addresses and centroids are tabulated in result section in Table 4.

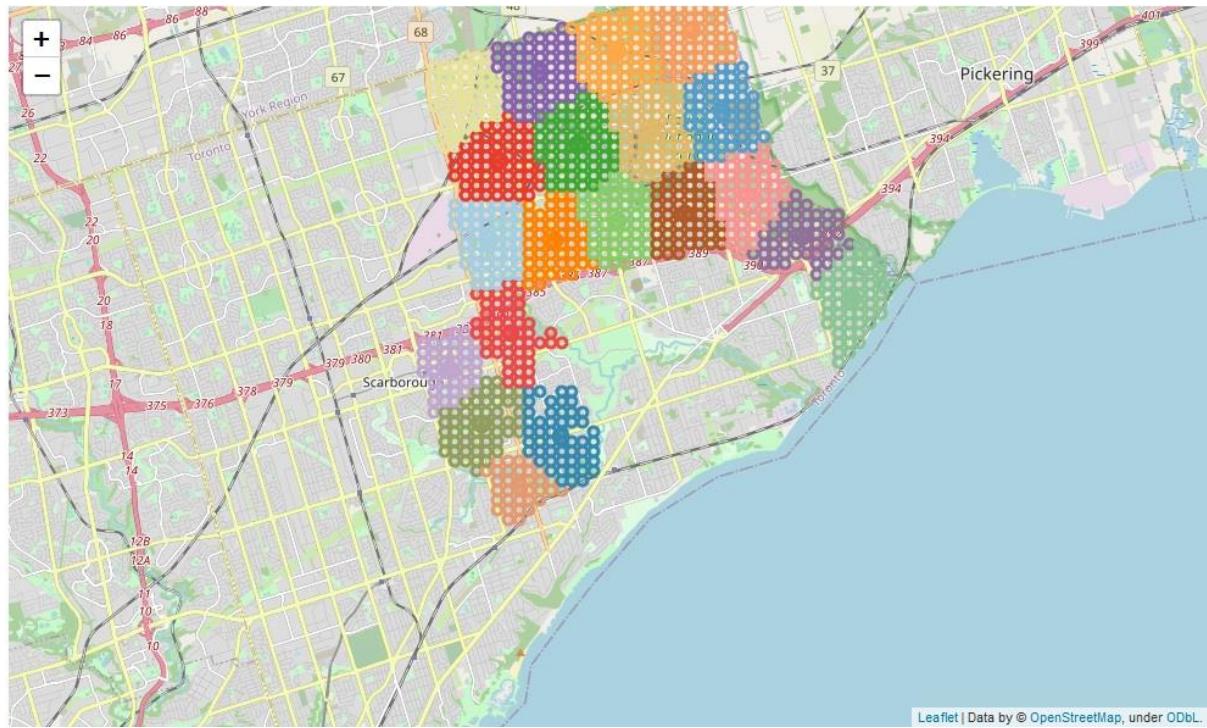


Figure 12: Mesh points in selected neighbourhoods, clustered to 20 groups

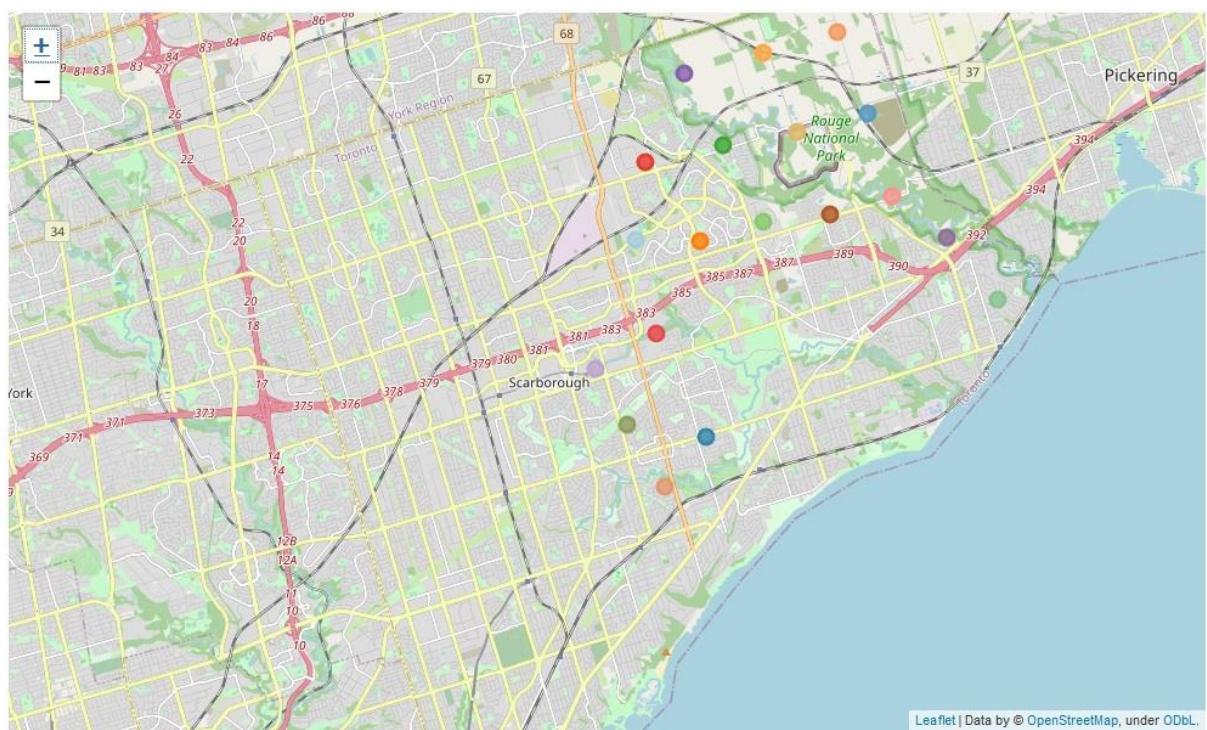


Figure 13: Centres of Groups, recommended locations

5. Results

The twenty recommended locations for further exploration are:

Table 4: Recommended places for starting Indian Restaurant

	Recommended Address	Location Coordinates
1	Rouge Park - Cedar Trail, Scarborough—Rouge Park, Scarborough, Toronto, Golden Horseshoe, Ontario, M1B 2X3, Canada	(43.827789098448534, -79.16790286803437)
2	3801, Lawrence Avenue East, Woburn, Scarborough—Guildwood, Scarborough, Toronto, Golden Horseshoe, Ontario, M1G 1R6, Canada	(43.761842148846334, -79.21343801392297)
3	31, Blueking Crescent, West Rouge, Scarborough—Rouge Park, Scarborough, Toronto, Golden Horseshoe, Ontario, M1C 2Z2, Canada	(43.79003170928593, -79.13119350842845)
4	Venture Drive, Scarborough—Rouge Park, Scarborough, Toronto, Golden Horseshoe, Ontario, M1B 3M5, Canada	(43.80586872227296, -79.19728311881808)
5	Heritage Park Public School, 80, Old Finch Avenue, Morningside Heights, Scarborough—Rouge Park, Scarborough, Toronto, Golden Horseshoe, Ontario, M1B 5J2, Canada	(43.821318575645414, -79.20866816280635)
6	Tredway Woodsworth Public School, 112, Sedgemount Drive, Woburn, Scarborough—Guildwood, Scarborough, Toronto, Golden Horseshoe, Ontario, M1H 1X9, Canada	(43.76435448556622, -79.23596305020745)
7	7, Twyn Rivers Drive, Rouge, Scarborough—Rouge Park, Scarborough, Toronto, Golden Horseshoe, Ontario, M1B 2K5, Canada	(43.81099599500024, -79.16087207985703)
8	Centennial College, 941, Progress Avenue, Scarborough—Guildwood, Scarborough, Toronto, Golden Horseshoe, Ontario, M1H 3B4, Canada	(43.7830759950002, -79.2276049369999)
9	Whitefield Christian Schools, Neilson Road, Scarborough North, Scarborough, Toronto, Golden Horseshoe, Ontario, M1X 1Z8, Canada	(43.81789599500025, -79.2305649369999)
10	Beare Road, Scarborough—Rouge Park, Scarborough, Toronto, Golden Horseshoe, Ontario, L1V 6M9, Canada	(43.84446474500028, -79.1765149369999)
11	14, Reesor Road, Scarborough—Rouge Park, Scarborough, Toronto, Golden Horseshoe, Ontario, M1B 6K4, Canada	(43.840189543387375, -79.19749074345151)
12	8, Greypoint Drive, Scarborough—Rouge Park, Scarborough, Toronto, Golden Horseshoe, Ontario, M1B 1M4, Canada	(43.80183599500023, -79.2150649369999)
13	14, Strandhill Road, Woburn, Scarborough—Guildwood, Scarborough, Toronto, Golden Horseshoe, Ontario, M1H 2A2, Canada	(43.7516430538237, -79.22508846641166)
14	Miracle Family Temple, 1100, Bellamy Road North, Scarborough—Guildwood, Scarborough, Toronto, Golden Horseshoe, Ontario, M1H 1H2, Canada	(43.775919071923276, -79.2446495523845)
15	Passmore Avenue, Morningside Heights, Scarborough—Rouge Park, Scarborough, Toronto, Golden Horseshoe, Ontario, M1X 2E6, Canada	(43.835945147542645, -79.21962086920328)
16	56, Royal Rouge Trail, Rouge Hill, Scarborough—Rouge Park, Scarborough, Toronto, Golden Horseshoe, Ontario, M1B 4X3, Canada	(43.80245432833356, -79.14563993699988)
17	Passmore Avenue, Armadale, Scarborough North, Scarborough, Toronto, Golden Horseshoe, Ontario, M1X 0A1, Canada	(43.83091907192334, -79.24055339853835)
18	Toronto Zoo, Old Finch Avenue, Scarborough—Rouge Park, Scarborough, Toronto, Golden Horseshoe, Ontario, M1B 5J2, Canada	(43.824143535983865, -79.18764198618022)
19	9390, Sheppard Avenue East, Rouge, Scarborough—Rouge Park, Scarborough, Toronto, Golden Horseshoe, Ontario, M1B 5X4, Canada	(43.807292291296534, -79.17832049255544)
20	103, Littleleaf Drive, Browns Corners, Scarborough North, Scarborough, Toronto, Golden Horseshoe, Ontario, M1B 1P6, Canada	(43.802070069074304, -79.23343160366656)

6. Discussion

The analysis shows that there are large number of restaurants in Toronto. The largest ethnic group is Chinese followed by English and Canadian. Indians occupy the sixth position.

The largest number of restaurants are of generic type, followed by Fast Food and Chinese. Indian Restaurant occupy the fifth position. There are 165 Indian Restaurants.

The 10 largest population groups are not distributed normally. there are lot of outliers. The population of indians and number of indian restuarants are not correlated very much. The chinese ,Portuguese and Jamaican population shows good correlation to their respective restaurants.

The features like Total Population, Area , Average income shows poor correlation.

Nova Scotian , other Asian orgins, Pakistani Fijian and Indian opulation show the highest correlation to indian restaurant. However the maximun correlation is only 0.356 which is not good.

We have grouped the neighborhoods using K means clustering using demographics data set and also with Restaurant profile data set. Both of these cluster groups shows excellent correlation. We have identified the cluster gruop containing neighborhood Woburn as an ideal candidate for our location. The neighborhoods in the cluster are Rouge, Malvern and Woburn. The polpulation of indians in the neighborhoods are relatively high.

We create a mesh of points with distance of 200x200m and remove the points near to an Indian restaurant. We club the remaining points to 20 groups. The centers of these groups are our recommended location to stakeholders These locations are to be further explored.

7. Conclusion

The Data was collected and analysed using statistical and machine learning techniques and twenty locations are identified for further exploration.