Determining the best neighborhood to set up a indian restaurant.

Anish S Ghiya

Problem Statement

Looking at he background of toronto we observe that the there are a lot of indians in the city of toronto.So the more the indian folk the more would be the buisiness for the indian restaurants. But there are few areas where there are indian restaurants which are very popular. So the new setup must be in the areas where there is no competition or at places where the population is mostly indian with high population density and the average income is also decently high so there is high chances of profitable buisiness , the other acse would be to set up in a place where the competition exists but is not that strong but that would not be advised as there might be some other reasons also why the business might not be running properly in that area.

Data acquisition

- The main data was obtained from wikipedia ie the postal codes of Canada.But this was not sufficient to get the required attributes that were required. So for the remaining data ie the latitudes and longitudes were taken from the geospatial data that was provided in the previous modules.The remainong columns were from wikipedia ie Demographics of toronto by neighborhood.

- The fousquare api was alo used for this to attain the datasets for the venues in the particular neighborhood.

- The links to each of theses are :

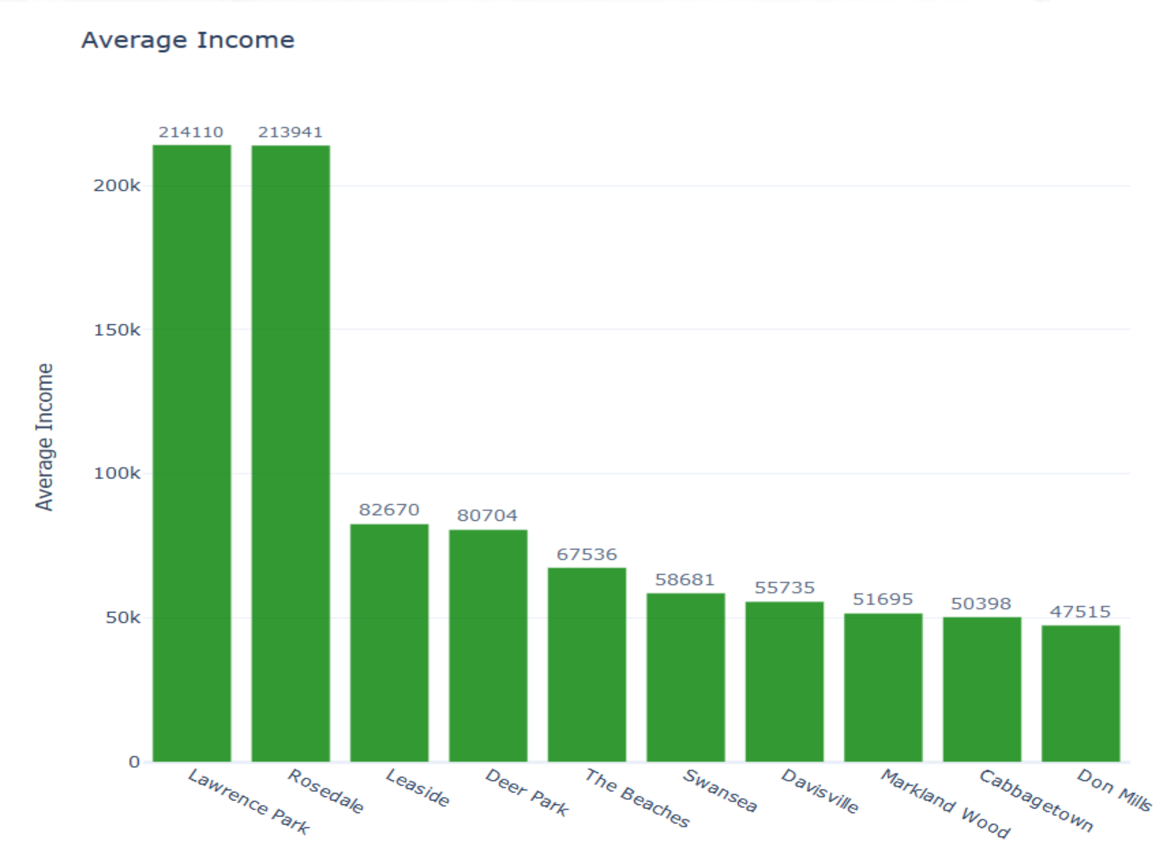
- i) https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

- ii) https://en.wikipedia.org/wiki/Demographics_of_Toronto_neighbourhoods

Exploratory Data Analysis

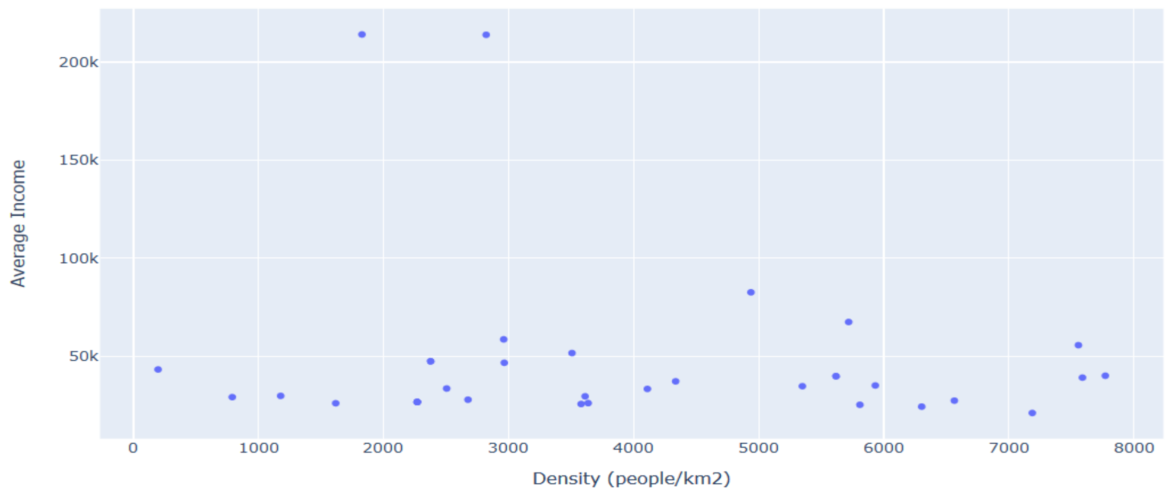
From the above graph we understand that the 2 neighborhood Lawrence Park and Rosedale have much higher average income when compared to the other neighborhoods

Following neighborhood come under the middle class income group and so this are the areas we should focus when it comes for the restaurant business



Density versus average income

The average income as we have seen plays a major role in the location picking as well as the population density but now let us see the relation between the population density and the average income which helps in understand the population of that neighborhood.



From the scatter plot we understand that the average income based on the density of population is almost a straight line ie it is almost constant when it comes to the regression between them . Later i have plotted the heatmap of correlation which will explain it better that the density has a descnet influnce on the average income. But the main influencer of the average income is the change in population from the year 2001.

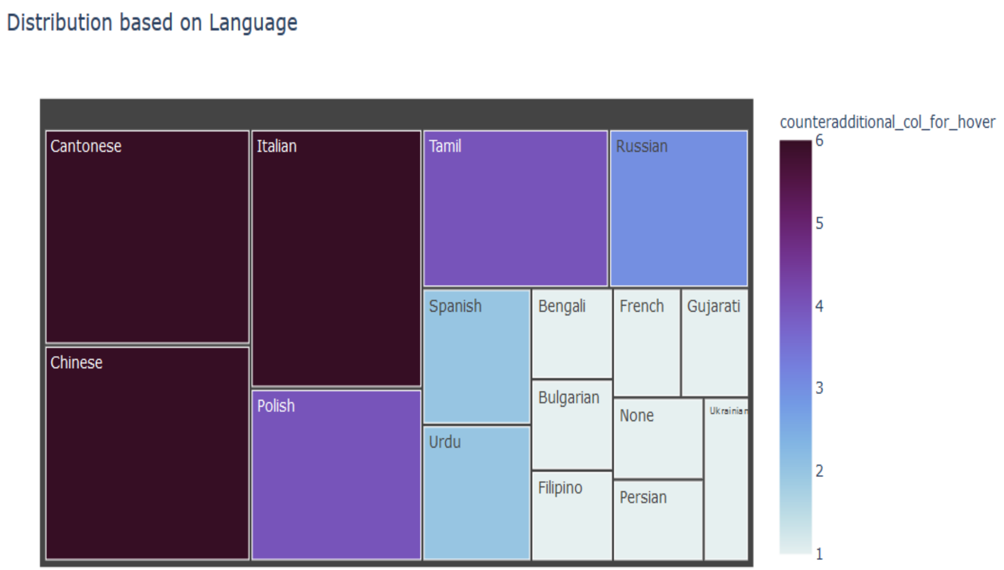
Language distribution

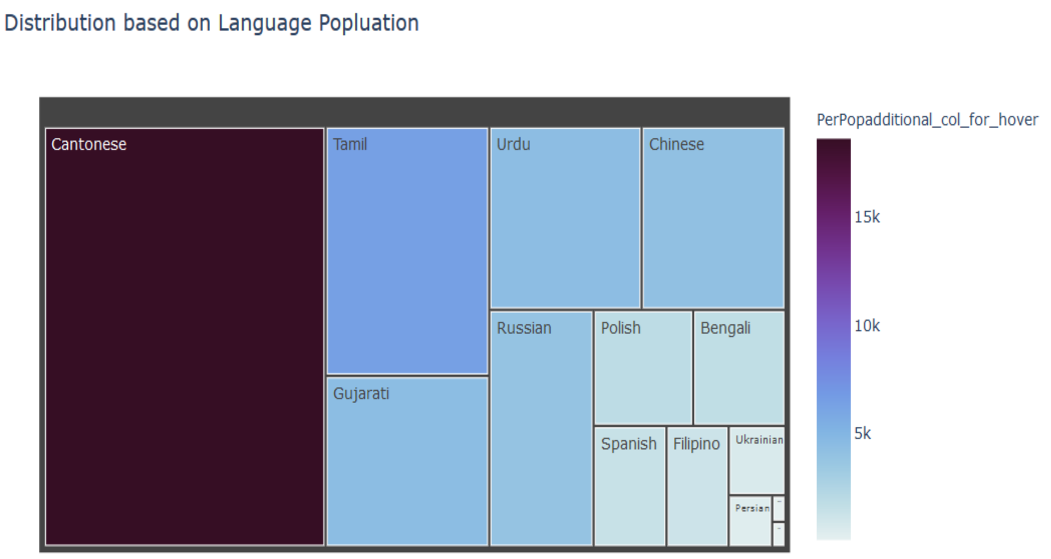
The language will play a major role in detemining the population type of the particular neighborhood and so there will be a lot of emphasis on this as the more the population of that language in the region they will like that type of cuisine more.

From the tree map we can see that cantonese,chinese,italian are the major second languages in the cty of toronto

Indians are not behind but as we speak different languages in different states so we are diversified by language but if we sum all of these we can get ahead of all the other languages.

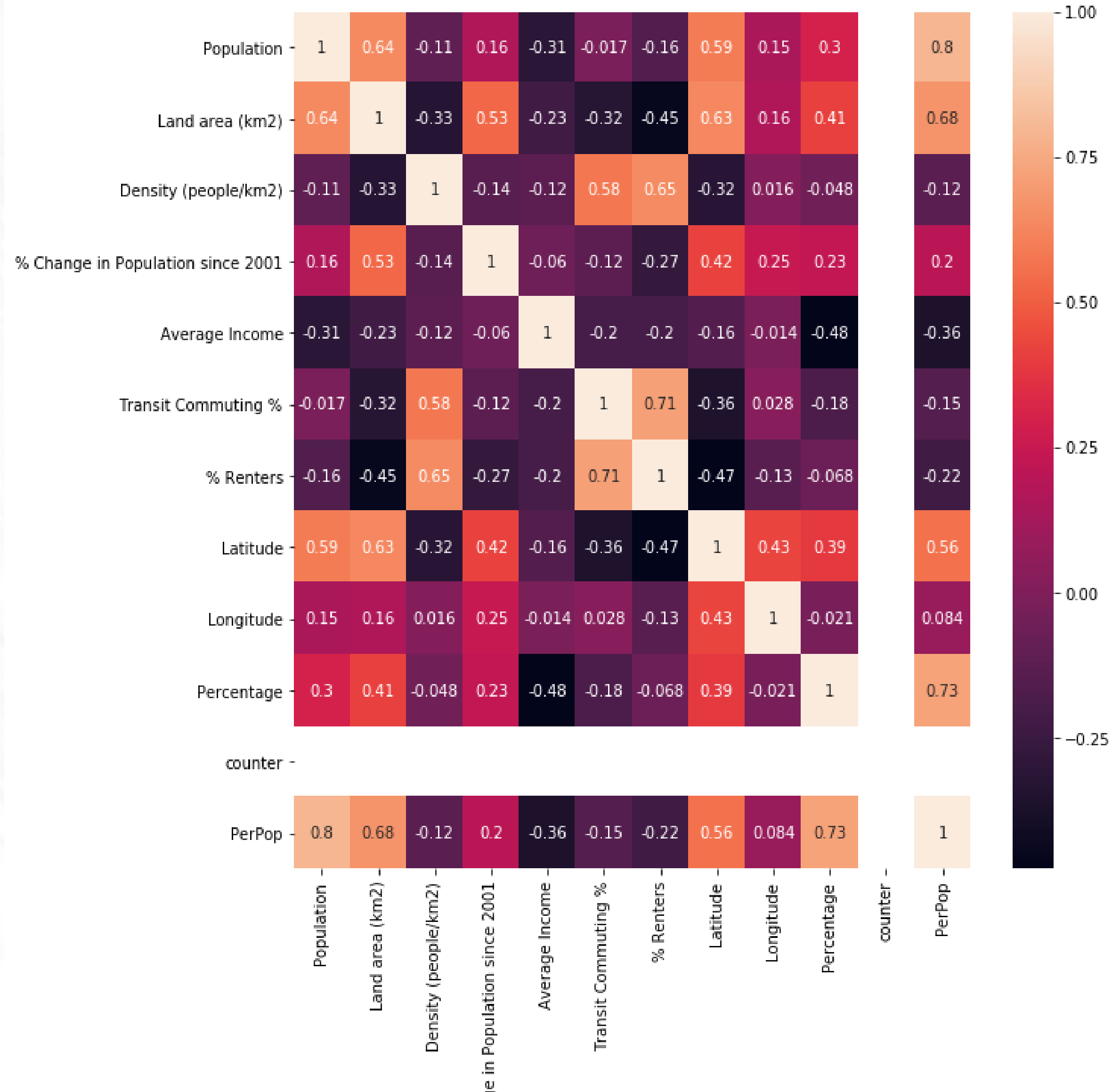
The above tree map is a descrption of how there a change when we see the population which speak the language amd here we can see that cantonese reamins at the top with the max population speaking the language and then is followed by tamil and Gujarati.





Heatmap of the correlation between the features

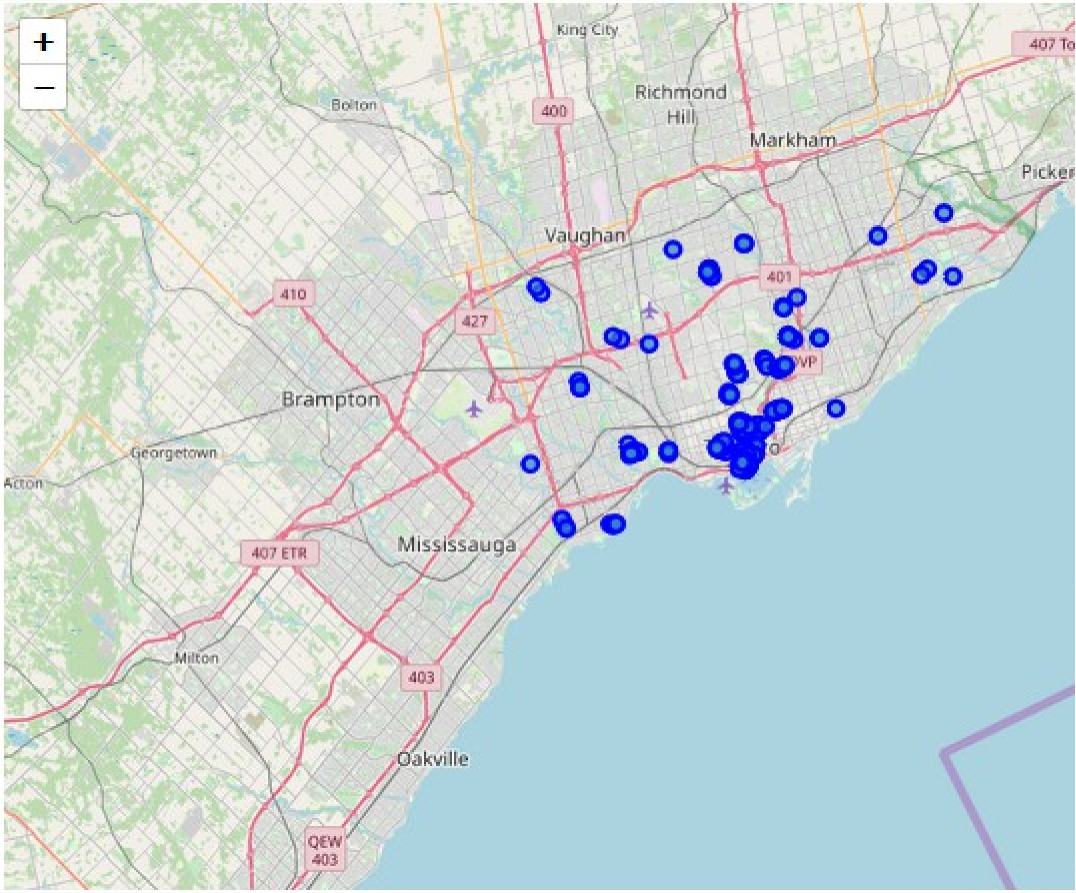
It is essential to understand the correlation between various data features so as to get a better understanding of the dataframe and with this correlation we can understand whatdetrmines the paarticular feature in the best possible way.



Working with the restaurants

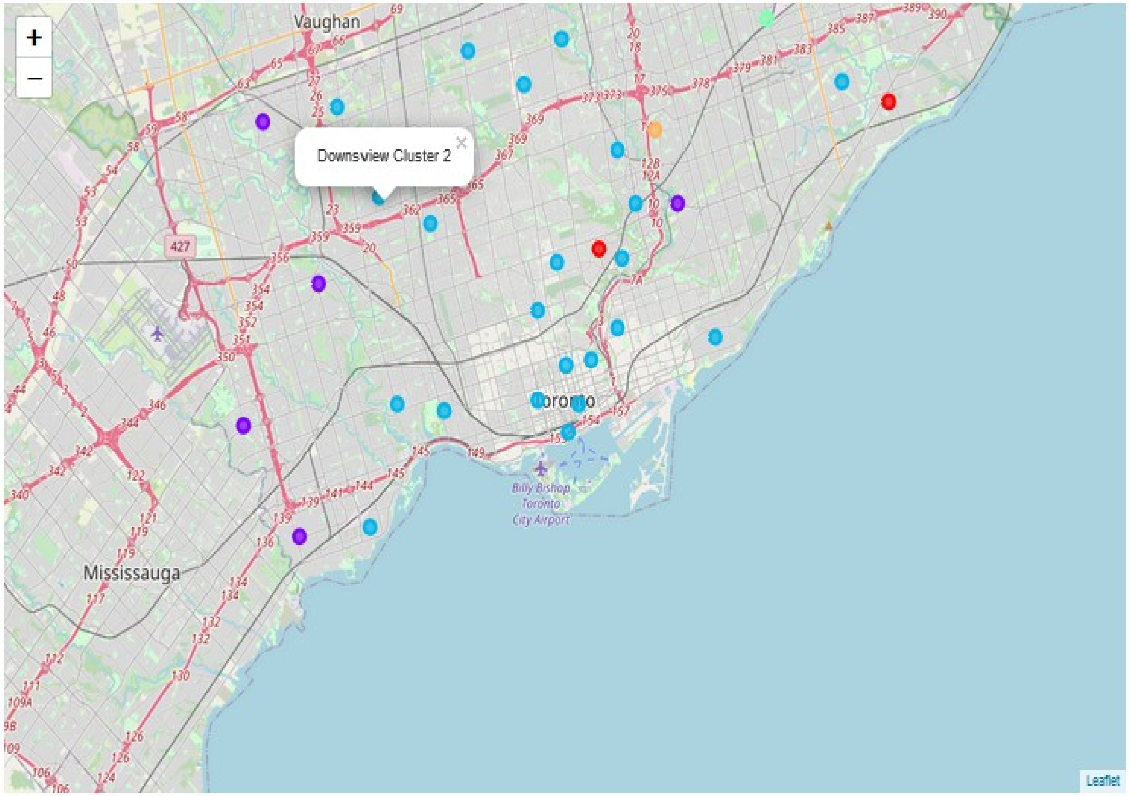
Now that we have the data set ready let us see the restaurants in the whole dataset

Now we need to this so as to understand the areas which have already got a lot of restaurants and wheter they hve got the data on the population or they have just opened with respect to their speciality and so which would open an window for us to set up our restaurant



Predictive Modelling :

Kmeans algorithm is an iterative algorithm that tries to partition the dataset into Kpre-defined distinct non-overlapping subgroups (clusters) where each data point belongs to only one group. It tries to make the inter-cluster data points as similar as possible while also keeping the clusters as different (far) as possible. It assigns data points to a cluster such that the sum of the squared distance between the data points and the cluster's centroid (arithmetic mean of all the data points that belong to that cluster) is at the minimum. The less variation we have within clusters, the more homogeneous (similar) the data points are within the same cluster.



Result

The Rouge neighborhood in the Scarborough would be best to set up a south indian restaurant in this cluster

```
1 toronto_clustered_india.iloc[0]
```

```
FM_x                                                          S
% Change in Population since 2001_x                         175
Average Income_x                                          29230
Transit Commuting %_x                                      12.1
% Renters_x                                                 2.7
Postal code_x                                               NaN
Borough_x                                           Scarborough
Latitude                                                43.8067
Longitude                                              -79.1944
Language_x                                                Tamil
Percentage_x                                               15.6
counter_x                                                     0
PerPop_x                                                3544.94
Transit Comm pop_x                                       2749.6
Rental pop_x                                            613.548
Cluster Labels                                                2
1st Most Common Venue                      Fast Food Restaurant
2nd Most Common Venue                     Vietnamese Restaurant
3rd Most Common Venue                Eastern European Restaurant
4th Most Common Venue                         German Restaurant
5th Most Common Venue                         Fried Chicken Joint
Name                                                      Rouge
FM_y                                                          S
Population                                                22724
Land area (km2)                                           28.72
Density (people/km2)                                        791
% Change in Population since 2001_y                         175
Average Income_y                                          29230
Transit Commuting %_y                                      12.1
% Renters_y                                                 2.7
Postal code_y                                               NaN
Borough_y                                           Scarborough
Language_y                                                Tamil
Percentage_y                                               15.6
counter_y                                                     0
PerPop_y                                                3544.94
Transit Comm pop_y                                       2749.6
Rental pop_y                                            613.548
Measure                                                 1.14142
Name: 16, dtype: object
```