

Battle Of Neighborhood

Coursera Capstone Project (IBM)

Shivam Pandey

11th June 2020

Introduction:

BUSINESS PROBLEM

A Person , currently lives in west side of Toronto , And he loves his Neighborhood very much. But , he got a great job offer from a great company on some other side of city . He cannot reject such an great opportunity .

So , he decided to move there , we need to help him find a place closest to its job , which has most similar neighborhood as its previous location .

So , if we think about solving , first we need to understand what are the factors because of which the man loves his neighborhood .

it may be :

1. Supermarket/ grocery store
2. Parks
3. Transport (metro + buses)
4. School
5. Rent/Property Rate

There can also be some factors like , bond with neighbors ,etc. We will ignore such factors here.

So , We will use Foursquare API to extract the location of our above required parameter for our data . So that we can analyze for our result .

We can simply collect geographical info's , and find closest location to office which have similar to initial location .

In our research we will collect all cities/town in Toronto and define parameter counts ,to compare cities and will cluster them using k means based on report .

We will use python in Jupyter Lab to solve this problem .

DATA COLLECTION

Neighborhood in Toronto

we will use week 3 data for our project To list cities .

Link – https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

We will use this list collect data from Foursquare API , such as Latitude , Longitude , various venues as defines in our Business statement .

After finding the list of neighborhoods, we then connect to the Foursquare API to gather information about venues inside each and every neighborhood. For each neighborhood, we have chosen the radius to be 3000 meter.

1. Neighborhood
2. Neighborhood Latitude
3. Neighborhood Longitude
4. Venue

we will generate a equation representing scores (dependency on parameter) .

Later we will use this and group them in cluster (best possible)
based on Location ans scores . (may use SVM to fit even 3d plot if needed).

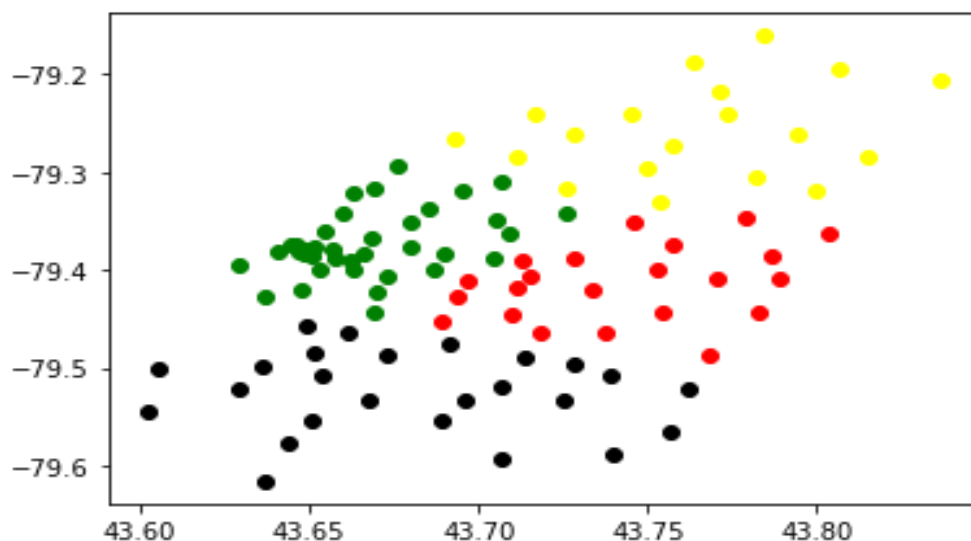
METHODOLOGY:

Here after cleaning data , we will use **FOURSQUARE API** to extract Latitude and Longitude of Postal Code .

After That we will also extract Counts of various basic amenities from foursquare .
And Store all this in a Data Frame . We can assume all amenities such as school ,
transport ,hospitals, parks are equally important .

After that normalize each column and calculate a new column score i.e. sum of normalized points of our parameters (amenities) and store it in a separate column.

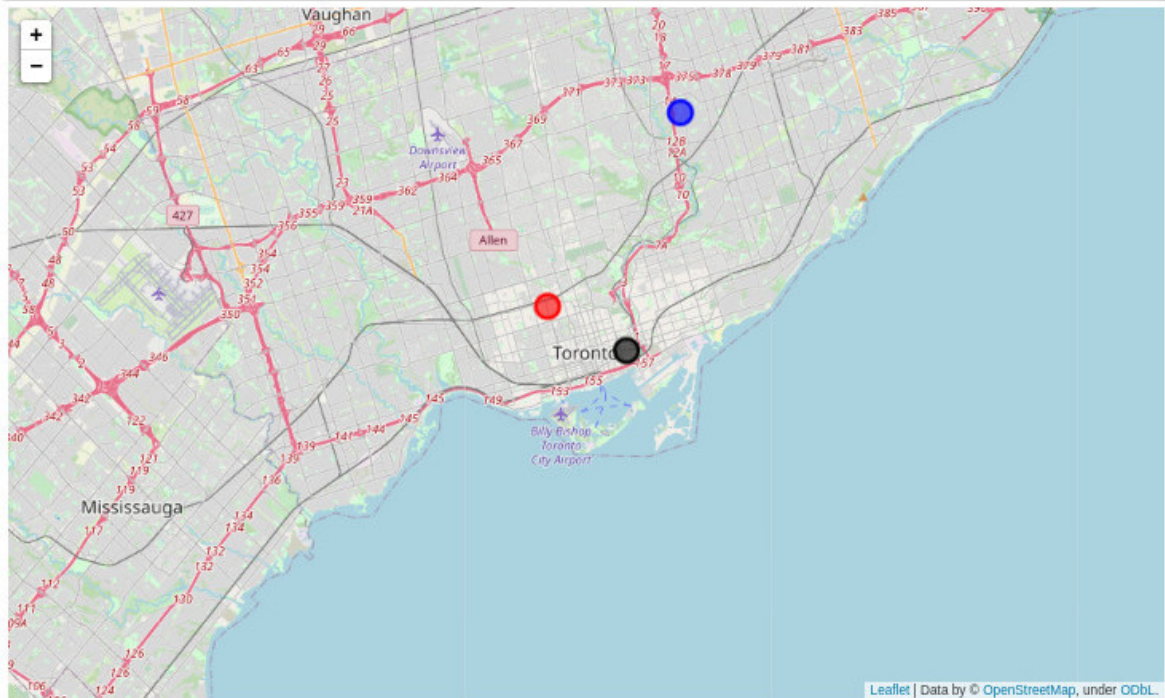
After that Apply K – Means Clustering , and provide best possible values of k .



After that we can input our current_city and target_city , and will find city in cluster of target_city with ,minimum difference between score column with current_city .
And That's The answer (Prediction) .

Drawback= if facilities are unevenly distributed in a country then this can give a wrong result , but this is not case with Canada .

Out[119]:



RESULT:

Finally we stored information and used that in a useful form which can be used for our prediction .
We allotted scores to each Parameter and finally compare them on basis of that after Clustering.
It works best with Postal code info's of out current and final city .

For Code you can refer

https://github.com/shivam-pandey9/Coursera_Capstone/blob/master/BattleOfneighborhood.ipynb