

**Opening pizza restaurant in one of the districts of Toronto City**

**Report**

Introduction

Toronto is a big city with a lot of opportunities and possibilities. Citizens of Toronto like to eat at foreign food restaurants.  And especially Italian food.

The idea is to open a pizza restaurant at best most appropriate location and district of Toronto.

Problem and Background

The problem is that Toronto is a big diverse city with a lot of districts. It is vital to choose the ‘right’ place for pizzeria so that there will be a demand and not too many competitors around.

We need to collect and analyse data from different district of Toronto in order to make right decision about possible location of a new restaurant.

Data

We are going to use extract list of postal codes of Toronto from Wikipedia, geographical location imported from csv or Geocodes and Toronto Foursquare location data to explore different districts.

When postal codes Toronto info is imported from Wikipedia (BeautifulSoup) it has to be normalized first with help of pandas data frames.

Data for each Toronto district will be analysed separately and then we will explore the best location.

Wikipedia URL: <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>

Toronto Geo csv: <https://cocl.us/Geospatial_data/Geospatial_Coordinates.csv>

Note: Etobicoke and Scarborough districts analysis were out of scope for this project.

Methodology

In this section I will provide description of what machine learning techniques were used and why.

We have picked up location data provider, geolocation API – Foursquare, classification and clustering analysis methodology for nationhood analysis for respective district.

By using clustering, we got into the category of unsupervised Machine Learning because we have a goal to group/cluster observation/data that have similar parameters. We were using one of the most popular clustering method K-Means with a number of clusters equal to 5.

Results

As it is shown in TorontoFinal.ipynb notebook, I have made analysis of neighbourhoods and clusters of four major district of Toronto city: North York, Downtown Toronto, East York and York.

As we can see from Cluster analysis the biggest demand for new pizza restaurant is in North York and Downtown Toronto as well as those are two larges districts which is great for customer demand.

Another interesting thing that can be concluded from cluster analysis of Downtown Toronto Cluster 1 is that there are a lot of Coffee Shops so if a Pizza Restaurant will be opened there it should have a good coffee offers in menu.

From the Foursquare APIs and maps we can pick up the best place for the restaurant in North York.

Discussion

During the analysis I found out that same code is used multiple times when analysing four different districts of Toronto. I did some refactoring in order to create functions which contained login in duplicate code:

getDistrictGrouped

return\_most\_common\_venues

displayTopVenues

getDistrictOnehot

getVenues

I found out as well that good comments and notes in notebook help a lot during comparing results and analyzing data.

Conclusion

For this type of analysis, which involves geolocation data, venues, clusters it was appropriate to user Machine Learning: numpy, pandas, geocoder and especially Foursquare APIs.