Using Machine Learning to find locations to open a Spanish Restaurant

INTRODUCTION

Background

For this Capstone project, I am creating a hypothetical scenario for an entrepreneur who wish to open a Spanish restaurant in Toronto where there may not be enough Spanish restaurants. This is the main idea behind this. This would be a great opportunity for this entrepreneur who is based on Canada. As Spanish food Ingredients are similar to Italian. This entrepreneur wants to open a restaurant in location where the Italian foods are popular. Finding a location for a Spanish restaurant is the most important decision for this entrepreneur and I am designing this project to help him find the most suitable location

Business Problem

The main aim of this project is to find the most suitable place for opening a Spanish restaurant .By using a various data science methods ,libraries, and machine learning methods such as clustering to find the neighourhoods ,This projects aims to provide solutions to answer the business question: In Toronto ,If an entrepreneur wants to open a Spanish restaurant,where should they consider opening it?

Target Audience

The entrepreneur who wants to find the location to open an authentic Spanish restaurant

Data

To solve this problem, I will need below data:

- ● List of Neighbourhoods in Toronto, Canada.
- ● Latitude and Longitude of these Neighbourhoods.
- ● Venue data related to European restaurants. This will help us find the Neighbourhoods that are

most suitable to open a Spanish Restaurant.

Extracting Data

- ● Scraping of Toronto Neighbourhoods via Wikipedia
- ● Getting Latitude and Longitude data of these Neighbourhoods via Geocoder package
- ● Using Foursquare API to get venue data related to these Neighbourhoods

Methodology :

The very first thing I need is to get the list of Neighborhoods In Toronto, Canada .This can be collected from the Wikipedia page. (“ https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M ”).This is done by web scraping using pandas html tabe scraping method as it is lot more easier to extract the table in a webpage and make them a Dataframe We have only the neigborhood names and postal codes .We also have to get the co-ordinates of the venues near these neighborhoods . This can be done by using FourSquare . I used the csv package which was already given by the IBM team.Using this I tried to match the corodinates of the Toronto Neighorhoods using postal codes.After all these are done,I visualize them in the world map using foilum and labeled the neigborhoods to check these are correct coordinates

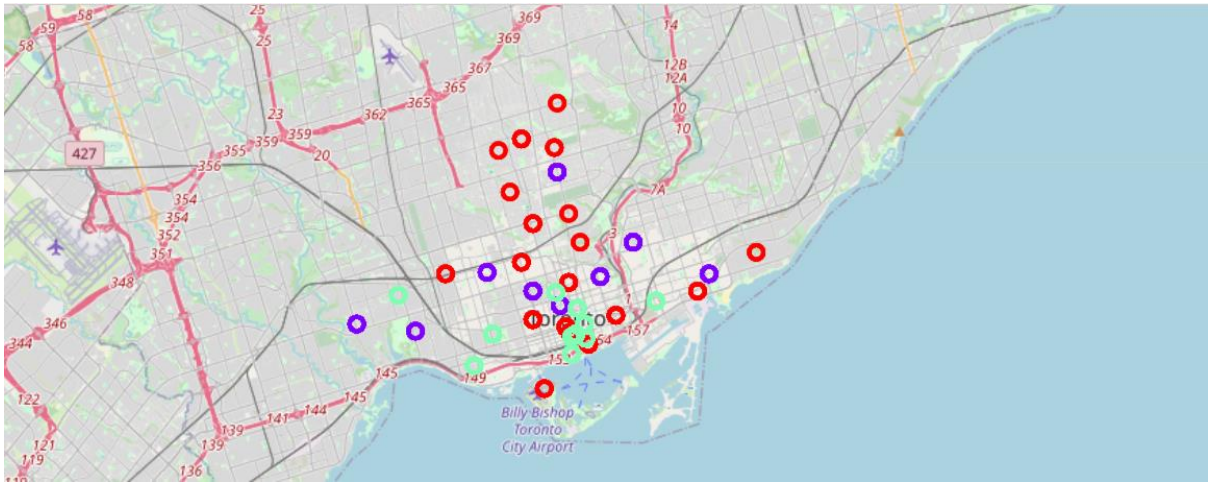
I used Foursquare API to get the list of top 100 venues within 500 meters radius.I have used my FourSquare account ID and API key to extract the data.Using FourSquare I got the latitudes and longitudes of each venue.I analysed each neigbourhood by grouping the rows by neighborhood and taking the mean on the frequency of each types and we can check the uniques categories also

Here we are looking for Italian Restaurant. Previously, so I looked for the restaurants closest to Spanish cuisine taste. Spanish food and Italian food are very similar in ingredients and taste, so my justification is that if there are people who enjoyed Italian food, they likely are going to enjoy Spansish food too. Lastly, I performed the clustering method by using k-means clustering. K-means clustering algorithm identifies k number of centeriods, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible. It is one of the simplest and popular unsupervised machine learning algorithms and it is highly suited for this project as well. Clustering is done I have clustered the neighborhoods in Toronto into 3 clusters based on their frequency of occurrence for “Italian food”. Based on the results (the concentration of clusters), I will

be able to recommend the ideal location to open the restaurant.

Results :

Clustering is shown below:



The results from k-means clustering show that we can categorize Toronto neighborhoods into 3 clusters based on how many Italian restaurants are in each neighborhood:

- Cluster 0: Neighborhoods with no Italian restaurants
- Cluster 1: Neighborhoods with high number of Italian restaurants bluish
- Cluster 2: Neighborhoods with a number less than cluster 1

The results are visualized in the above map

Cluster 0: Red

Cluster 1: Violet

Cluster 2: Bluish Green

Recommendation:

From the above, It is evident that most of the Italian Restaurants are found more in areas like Central Bay Street in cluster 1 and Brockton, Parkdale Village in cluster 2 and close to zero in cluster 0 .It is also possible to have a Spanish Restaurant in cluster 2 rather than cluster 1 as there are less Italian Restaurant.It is better to open a italian restaurant in cluster 0.If the food provided in that hotel is satisfactory,affordable and authethic then the restaurant will have great following.

Limitations and Suggestions for Future Research :

In this project, I only take into consideration of one factor: the occurrence / existence of Italian restaurants in each neighborhood. There are many factors that can be taken into consideration such as population density, income of residents, rent that could influence the decision to open a new restaurant. However, to put all these data into this project is not possible to do within a short time frame for this capstone project. Future research can take into consideration of these factors. In addition, I am relying on the existence of Italian restaurants only for this project but future research can take into consideration of other variables such as existence of European restaurants, European population level in each neighborhood etc.

Conclusion :

In this project, I have collected data required for the recommendation .After all the preprocessing done in the dataset and applying the right machine learning method ,In our case we have used K-means clustering which allowed me to give the right recommendation to this entrepreneur.

References :

List of neighborhoods in Toronto:

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

Foursquare Developer Documentation: https://developer.foursquare.com/docs