OPTIMAL RESTAURANT LOCATION FINDER REPORT

1. Business Problem

With an increase in competition amongst food establishments, along with an increase in rent, it has become more difficult for smaller independant restaurants to survive. It is critical for restaurants to capitlize on the demand through optimal location in order to reach their niche consumers. Therefore predicting the location for a specific type of restaurant, will determine its success.

There is a large amount of public data available on how often people visit different venues. This project aims to use data on venue activity to determine which neighbourhoods would provide the best opportunities to open a restaurant. More specifically this project will give the user the freedom to input any type of restaurant eg: "Italian Restaurant, Thai Restaurant" and give back custom options of neighbourhoods based on the user's preference in restaurant.

This project would be of interest to anybody interested in opening up a restaurant, or for potential opportunity zones for new immigrants or business's who need a location to open up a restaurant.

2. Data

I used the Four Square API to obtain the data regarding all the different venues within a 10 kilometer raduis of each neighbourhood. More specifically the information obtained through the Four Square API was the list of venues in the GTA as well as their neighbourhoods.In order to get all of the different neighbourhoods that make up the GTA, I scraped the Wikipedia page containing all the postal codes starting with the letter "M". From Wikipedia I obtained a table with all of the "Borough","Neighbourhood", and "Postal Code" in the GTA.

Lastly I used a Geolocator to transfer the neighbourhood names into latitude and longitude values.

3.Methodology

3.1. Prepping data for Foursquare API

Initially I had a dataframe "df_Toronto" containing the scraped Wikipedia data contained, "Borough","Neighbourhood", and "Postal Code". To complete this dataframe and make it functional for the Foursesquare Api, I merged the coordinate values to "df_Toronto". I had obtained the Coordinates both using a CSV file, and simply with the Geopy geolocator. I now had a complete picture and visualized the neighbourhoods using folium.

3.2. Accessing Foursquare API

I then extracted the venue, venue frequency, and category using "getNearbyVenues" function as well as the "get_category_type"

3.3.Wrangling the Foursquare API Data

I had a dataframe with all the frequencies of visits for each venue category.Since my project was associated with restaurants I joined the data with neighbourhoods, and filtered the venues to only show restaurants.

There is also an input field, to let the user put in their preffered restaurant they want to open. The rest of the analysis will be based on this input. For this project I chose "Sushi Restaurant".

3.4.Data Overview

Now that all the data is cleaned and organized I began with looking at the top Neighbourhoods with the highest frequency of visits to Sushi Restaurants.

3.5. Machine Learning

I chose to use KMeans Algorithm for this Project. This is because I wanted an algorithm to cluster the neighbourhoods based on restaurant visits. KMeans was good clustering algorithm and, since it was unsupervised I just removed the neighbourhood labels, and let it cluster based on the existing data.

Optimal Paramters: However I needed to find the optimal Paramter for KMeans. I used the Elbow Method and found that a 6 clusters would be ideal.

To gather meaning out of the clusters I transposed Cluster Labels to the frequency and chose the Cluster with the highest mean frequency of Sushi Restaurant Visits. This being Cluster 0.

Cluster 0 had many neighbourhoods. To choose the otptimal neighbourhood , my key indicator was a neighbourhood that had the highest Sushi Restaurant visits, yet limited competition, that being the least amount of sushi restaurants.

A dataframe "Cluster_venues_type" was made which sorted the neighbourhood ascending based on "visits per neighbourhood" which means unique restaurant visits, and descending based on Sushi Restaurant Frequency.

I then visualized this on a bar plot however the y axis was based off Sushi Restaurant Frequency/ visits per neighbourhood, to find the nieghbourhood that had the most frequency and the least amount of restaurants.

<u>Results</u>

The resulting information is that out of the 22 neighbourhoods in cluster 0, the neighbourhoods associated with index 0-19 and neighbourhood 22 are all equally optimal for opening up a sushi restaurant.

Neighbourhoods 1-17 have the smallest mean frequency value of 0.01, yet only 1 unique restaurant that people are actively going to. The ratio between the supply and demand is 0.01

Neighbourhoods 18-20 have a larger mean frequency value of 0.02, but 2 unique restaurants people are actively going to. The demand is higher, but so is the supply therefore the ratio remains the same at 0.01

Neighbourhood 21, Downsview only has a mean frequency value of 0.01, whilst having 4 unique restaurants that people actively go to. Thus there is an over saturation of sushi restaurants whilst very little demand, compared to all of the other neighbourhoods. The ratio between the supply and demand is is a quarter of the rest of the neighbourhood at 0.0025.

Lastly Neighbourhood 22 has the highest mean frequency value of 0.05, but also the highest amount of unique restaurants being 5. Therefore the supply/demand ratio remains the same as Neighbourhoods 1-20 at 0.01.

Discussion:

Based on the results Neighbourhoods 1-20 & 22, are all good locations to open a sushi restaurant. However The compeition is already high in Neighbourhood 22, and success would have to see a even greater rise in demand which may plateau.

A good extension for this project in the future, is to find data that would able to create a timeseries plot of the frequency in visits, and use a regression to predict what the demand for such a restaurant is in the future.

Future potential: rent and property datasets, and API's are hard to obtain for toronto. With the availability of this data, another good project would be to find the most optimal places to rent/share commercial kitchens. That mixed with data on food delivery may be promising for business's to survive situations like covid where they can pay a fraction of the cost to share existing kitchens and have the bare minimum of staff. With chef's spread around neighbourhoods with the highest demand, not keeping all of the chefs "in the same basket" they would generate more revenue, and be more disaster resistant.

6. Conclusion

Overall by using Kmeans algorithm on Neighbourhood data, merged with foursquare data I was able to find the most optimal neighbourhoods in Toronto to open up a restaurant of a user's choice. This can be used as a tool for new entrepreneurs to open up their own restuarants, or for established business's to expand.