Using Machine Learning to find out best restaurant to Lviv(Ukraine) (IBM Capstone Project)

Introduction and Business Problem

Introduction: The city of Lviv, Ukraine is the largest city in western Ukraine and the seventh-largest city in the country overall, with a population of 724,241 as of 2020. Lviv is one of the main cultural centers of Ukraine. Mysterious and architecturally lovely, this Unesco-listed city is the country's least Soviet and exudes the same authentic Central European charm as Krakow once did. It's also a place where the candle of Ukrainian national identity burns brightest. Simply put Lviv is just one of the most beautiful and tourist city in Europe yet underrated. For people that are new to Lviv, due its geographic size, it can be daunting to figure out what restaurants are worth going to based on their current location. For people living in Lviv or are visiting Lviv, how do you know what are the best places to get delicious meal or drinks?

Business Problem: For this project, I am going to create a simple guide to help users to choose which restaurant fit the best their needs based on Foursquare data: *number of likes, restaurant category and geographic location data for restaurants in Lviv*.

Data Requirements and Methodology

Data Requirements

For this project, I will be utilizing the Foursquare API to pull the following location data on restaurants in Lviv, Ukraine:

- Venue Name
- Venue ID
- Venue Location
- Venue Category
- Count of Likes

Data Acquisition Approach

To acquire the data mentioned above, I will need to do the following:

- Get lat and long coordinates for Lviv via Geocoder package
- Use Foursquare API to get a list of all venues in Lviv
- Get venue name, venue ID, location, category, and count of likes

Methodology

The approach I use in this project is to consider the **count of likes** as restaurant receive on the Foursquare platform as an indicator of **quality**. So the higher number of likes a restaurant/bar got, the better is the quality of the restaurant. This might be very simplistic way of evaluating the quality of a restaurant but due to limited amount of calls we can make to the api(as we use the free account) we will not dig deeper into details to get more specific information (prices, ratings, reviews).

After retrieving all restaurants with their location, category and count of likes from Foursquare API, I will categorize the quality of the restaurant (*Poor*, *Average*, *Good*, *Excellent*), so I can cluster appropriately later on.

Then I will categorize all the restaurants based on type of cuisine (Ukrainian, Italian, American...). This will allow the user to easily find the best restaurant based on type of cuisine they like. Here is how restaurant were categorized:

bars = ['Wine Bar', 'Hookah Bar','Beer Bar','Cocktail Bar', 'Brewery', 'Irish Pub', 'Gastropub']

others = ['Coffee Shop', 'Creperie','Chocolate Shop', 'Café', 'Sandwich Place','Breakfast Spot']

ukrainian_food = ['West-Ukrainian Restaurant', 'Tatar Restaurant', 'Restaurant']

italian_food = ['Pizza Place','Italian Restaurant']

asian_food = ['Vietnamese Restaurant', 'Sushi Restaurant', 'Asian Restaurant']

Finally, Using k-means clustering algorithm, I will group restaurants into clusters so that people looking to eat in Lviv can easily see which restaurants are the best fit?, what cuisine is available?, and where the restaurant is located in Lviv.

Results

After Running the K-mean clustering algorithm on the dataset, I was able to generate four clusters of restaurants. These are as follows:
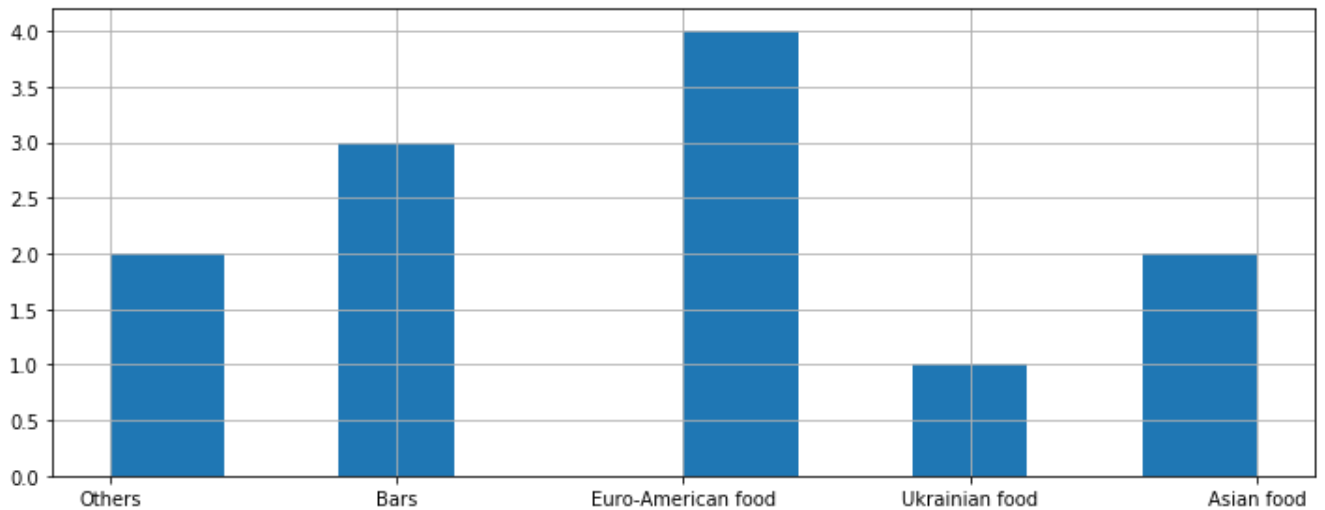
Cluster 1

- Characteristics
  - **Map of Clusters for Users**
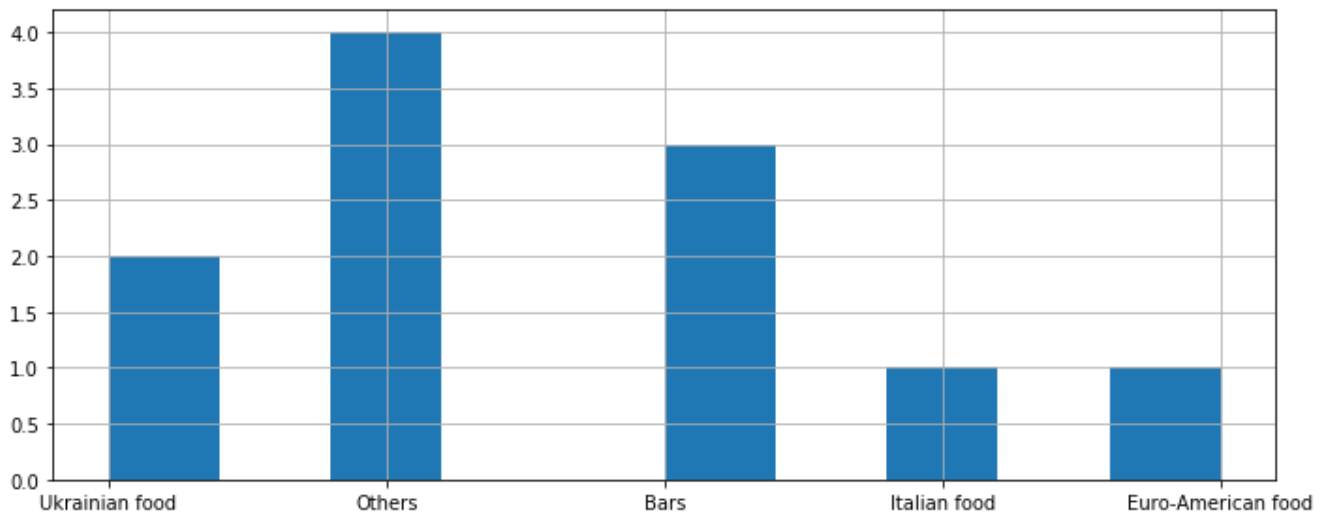 - Mostly **Bars and Others**
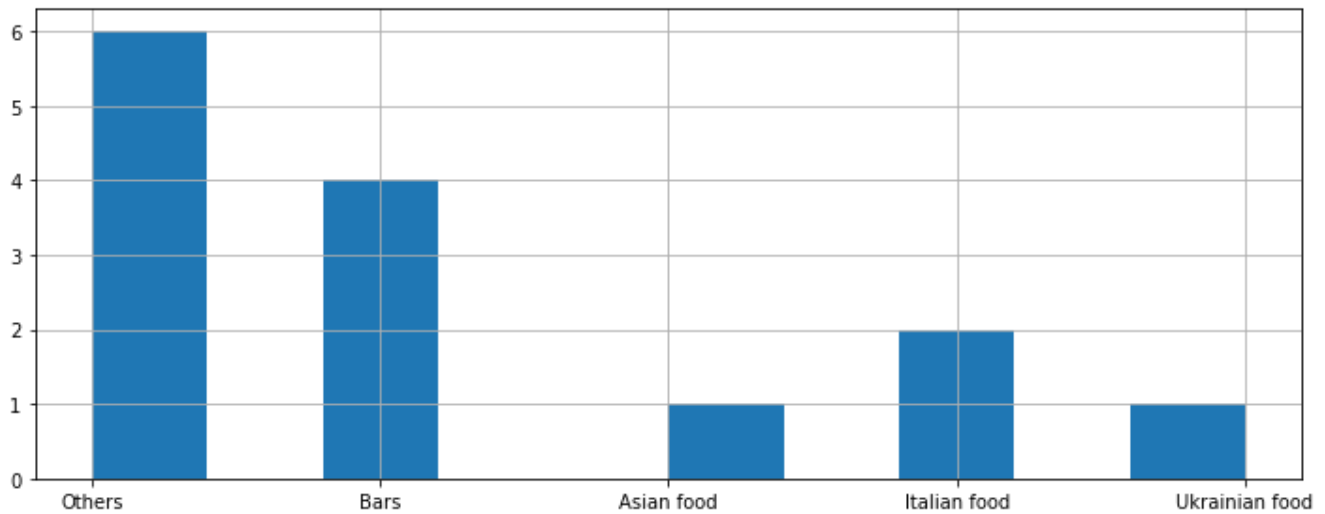
Cluster 2

- Characteristics
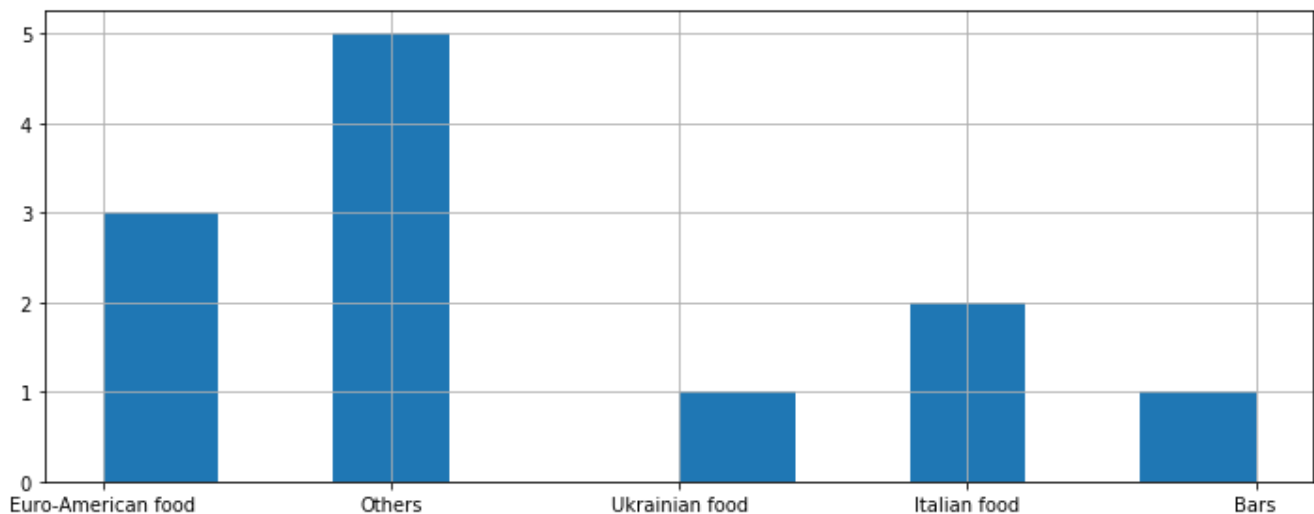 - **Average Quality**
  - Mostly **Euro-american food**

Cluster 3
- Characteristics
  - o **Excellent quality**
  - o Most **Ukrainian_food/others/bars**

Cluster 4
- Characteristics
  - o **Good Quality**
  - o Mostly **Others/ Euro-american food / italian food**

Based on our result we obtain 4 clusters, which will guide the user in the choice of the restaurant in Lviv. So for instance, a tourist who want to eat excellent food will not be disappointed if he choose an Ukrainian cuisine restaurant.

I can observe that the categories "Others and bars" is almost strongly represented in all the clusters, so we can probably break down those categories into more specific categories for more accuracy.
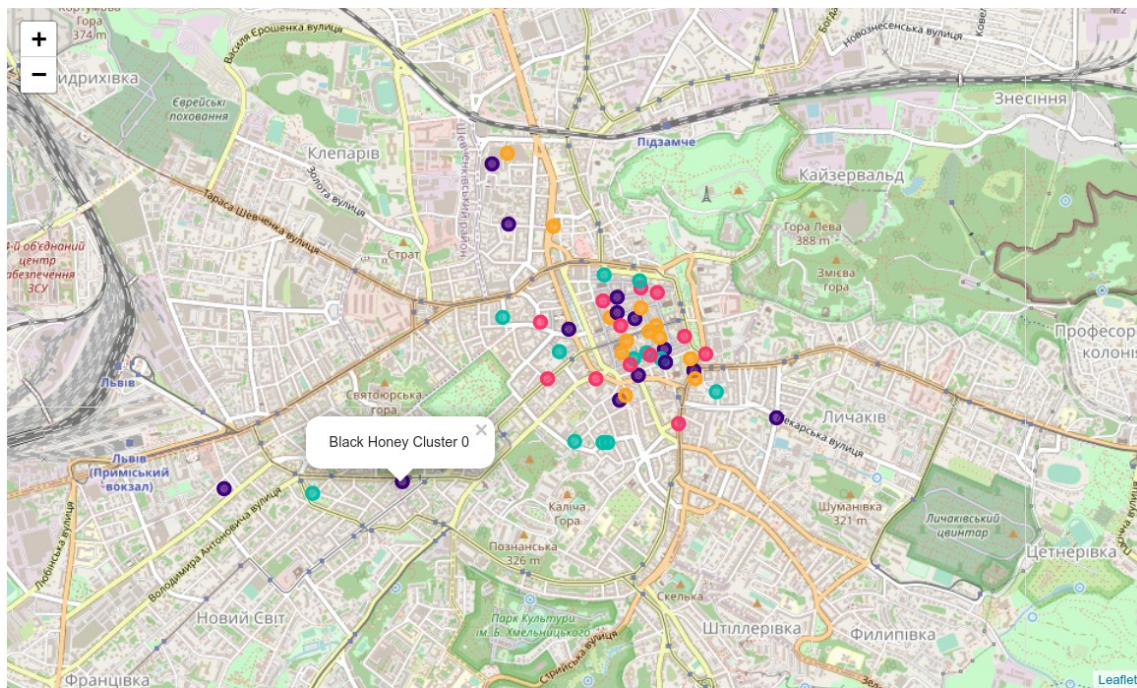
Fig: Map of Clusters for Users

Discussion and observations

A couple of observations:

- There are not enough data related to the city of Lviv on Foursquare. Most of the restaurants found are those are Ploshcha Rynok (Market Square) in the center of the city. Foursquare might not be very popular in Ukraine

- A lot of restaurant in our data set have a category **"Restaurant"**, which doesn't give us much information about the type of meal (ukrainian, italian, ...) the restaurant serves.

- For more accuracy of our classifier, I will recommend getting more relevant data about indicator as the **"price or average price"** of the meal in the restaurant, and We should dig deeper into the analysis of the **quality, quantity, and recency** of the reviews of each restaurant.

Conclusions

The aim of this project was to create a guide that can help user to find the best restaurant that fit their needs in the city of Lviv (Ukraine) using Foursquare API data (location of restaurants, categories, count of likes) and machine learning more precisely K-Means clustering. Based on the Results we got four clusters of restaurants. I have to emphasize that the approach I took in this study was very simplistic because the main

indicator of quality of the restaurant is the count of likes, which is not very realistic because we should dig deeper into the meaning of likes, does the user liked the quality of the food, the quality of the service or the quality of furniture in the restaurant **?**