Coursera Capstone

IBM Applied Data Science Capstone

Open Coffee Shop in Cracow, Poland

January 2021

1. Introduction

Visit in coffee shop is associated with relax and meeting with friends or family. In this time can rest from everyday life. People like places like that so there will be plenty of demand for such places. Living in Cracow is focus about central location of city. In some areas missing an atmospheric café (By writing atmospheric I do not think about chain of cafes.

Business Problem

The objective of this project is to analyze the best locations in the Cracow to open new coffee shop. Using data science and machine learning methodology like clustering this project try to answer the business problem: Where is the best area in Cracow to open new coffee shop?

Target Audience of project

This paper can be useful to everyone who thinking about open own business with coffee shop.

2. Data

Requirement of data:

- List of neighborhoods in Cracow,
- Latitude and longitude coordinates of these neighborhoods,
- Venue data

Source of data:

- The Wikipedia page contains a list of neighborhoods in Cracow,
- Foursquare API to get venue data for these neighborhoods

In this project will go through some data science abilities from web scraping (from Wikipedia), connect with API (Foursquare), data cleaning and wrangling, use machine learning algorithm and on the final visualize on map. In the next point, will be presented methodology.

3. Methodology

First point is list of neighborhoods. This list is available in the Wikipedia page. This data will be scrapping by Python requests. Next step what we need is geographical coordinates data. This data is available in Foursquare API. Geocoder package allow to convert address into format with latitude and longitude.

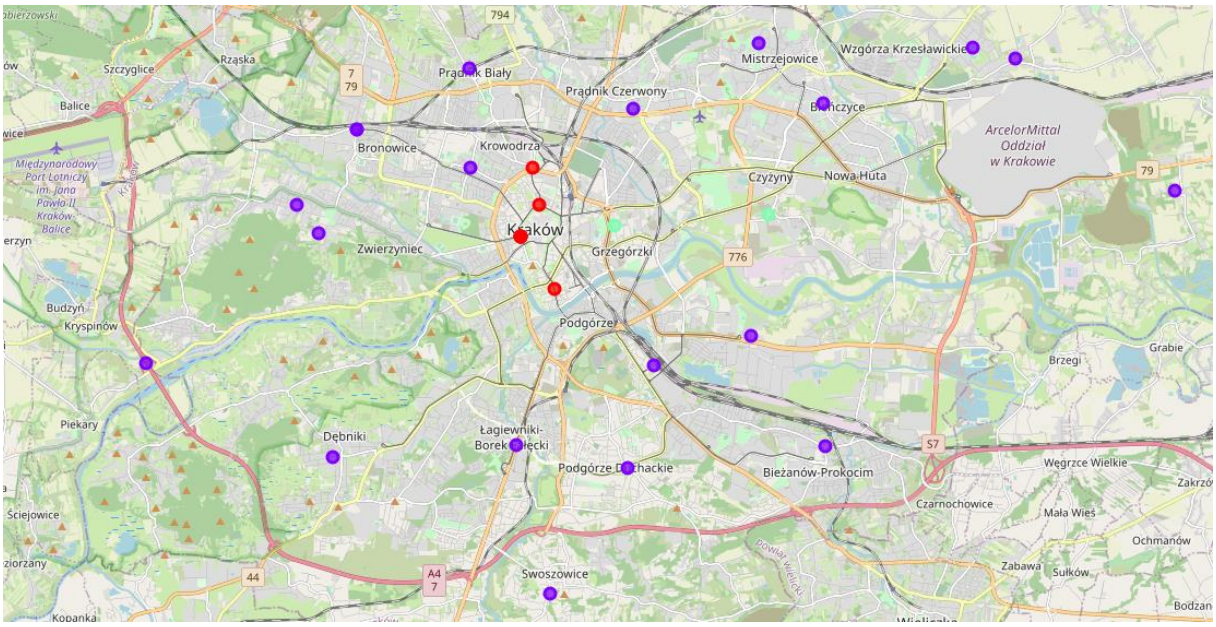
After that data will be use Foursquare API to get top 100 venues within a radius of 1000 meters. API return the venues data in JSON format, which can extract to venue name, latitude and longitude. Then will be analyze each neighborhood by grouping rows and taking the mean of frequency of occurrence of each category.

Final step is perform clusters on data using k means algorithm. This is very popular and easy to implement method for unsupervised data. This algorithm divided data into 3 clusters. This result will help to answer to business question.

4. Results

The results of clustering show that we can categorize the neighborhoods into 3 clusters:

- Cluster 0 (red color): Neighborhood of center of city with moderate number of coffee shops,
- Cluster 1 (purple color): Neighborhood with low number of coffee shops,
- Cluster 2 (mint color): Neighborhood with high number of coffee shops.



5. Discussions

By observation of map can be noticed that, most of coffee shops are concentrated in center of city. Cluster 1 has low number of these kind of places. Areas around center give a great opportunity to open new coffee shops, because of little to competition from others Coffee Shops. In other hand coffee shops in cluster 0 and 2 likely suffering from intense competition due to oversupply. Based on analysis above recommends open new coffee shops in around of cluster 1 and should avoids to neighborhoods in cluster 0.

6. Conclusions

In this project we have gone through the process of data scientist: identifying business problem/s, perform the data requirements, preparing data, clustering. To answer of the business question: The neighborhoods in cluster 1 is the most preferred locations to open new coffee shop.

Of course this is not the end of possibilities analyze the data. For example can be make more deeply analysis in the specific area.