Coursera Capstone

Find the best district to open a mall in Warsaw

Introduction

Malls has become a place to meet people and talk about their experiences. That's why malls are highly valuated in the city planning and having a great place to build one is hard, because the property developers needs to look to every aspect related to location

Business Problem

- The objective of this capstone project is to analyze and select the best locations in the city of Warsaw, Poland to open a new shopping mall.
- This project aims to provide solutions to answer the business question: In the city of Warsaw, Poland, if a property developer is looking to open a new shopping mall, where would you recommend that they open it?

Data

- List of districts in Warsaw. To define the scope for this project which is confined to the city of Warsaw.
- Latitude and longitude coordinates of those Districts. This is required in order to plot the map and also to get the venue data.
- Data of the venues in the city, this data should correspond to shopping malls. This data is used to perform clustering on the districts.

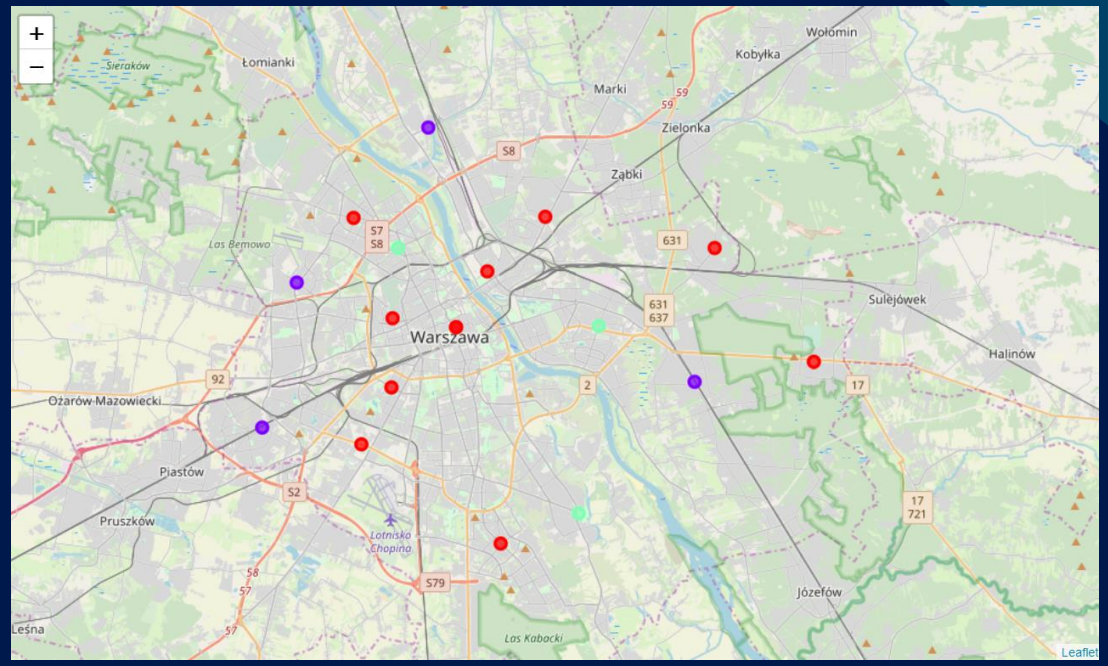
Methodology

- Web scraping Wikipedia page for districts list
- Get latitude and longitude coordinates using Geocoder
- Use Foursquare API to get venue data
- Group data by districts
- Filter venue category by Shopping Mall
- Perform clustering on the data by using k-means clustering
- Visualize the clusters in a map using Folium

Results

- Districts in 3 clusters:

- Cluster 0. Districts with low number or don’t have malls
- Cluster 1. Districts with the more concentration of malls
- Cluster 2. Districts with a moderate number of malls



Discussion

- Some more experienced people maybe can use better or more advanced web scrapping techniques or other APIs instead of Foursquare which could lead in a different or more precise result.

Conclusion

- With the results obtained we can see that most of the malls area located in cluster 1, while cluster 2 has less malls and cluster 0 has no shopping malls. This means that the districts located in cluster 1 have more competition than in other clusters