Analysing Neighbourhoods of Warsaw For Starting New Restaurant.

Krzysztof Kamiński

IBM Applied Data Science

Introduction

- Warsaw : The biggest city in Poland
- Most populous urban area in Poland, with population of 1.79 mln
- The population comprises of people of various ethnicities from all over the world

Business Problem

- Start a restaurant
- Neighbourhood that is most likely to give a good business

Data

- Neighbourhoods of Warsaw

– Neighbourhoods of Warsaw wikipedia page through data scraping.

- Geographical coordinates of the neighbourhoods

– Using GeoPy library.

- Venue data from FourSquare

– Using FourSquare API

Methodology

- Feature Extraction

  – One Hot Encoding

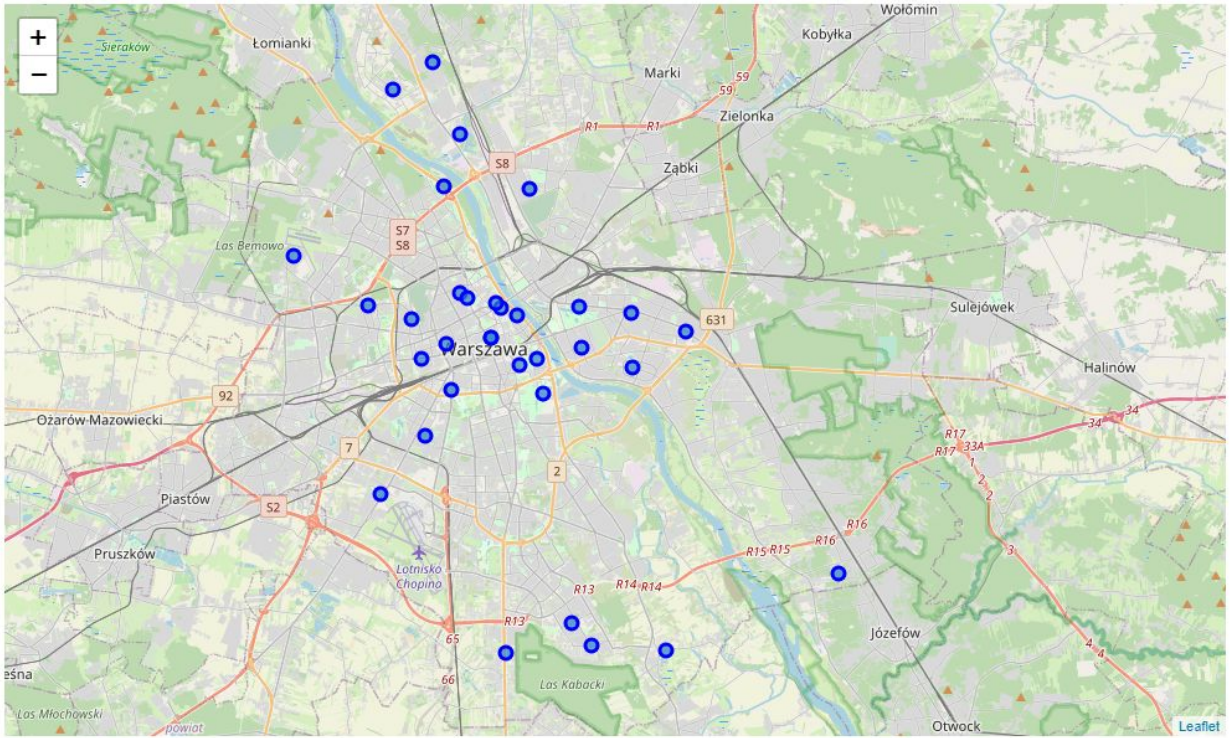
```
waw_1hot = pd.get_dummies(explore_waw[['Venue Category']], prefix="", prefix_sep="")


# Add neighbourhood column back to dataframe
waw_1hot['Neighbourhood'] = explore_waw['Neighbourhood']


# Move neighbourhood column to the first column
fixed_columns = [waw_1hot.columns[-1]] + waw_1hot.columns[:-1].values.tolist()
waw_1hot = waw_1hot[fixed_columns]

waw_1hot.head()
```

- Unsupervised Learning

  – K-Means Clustering

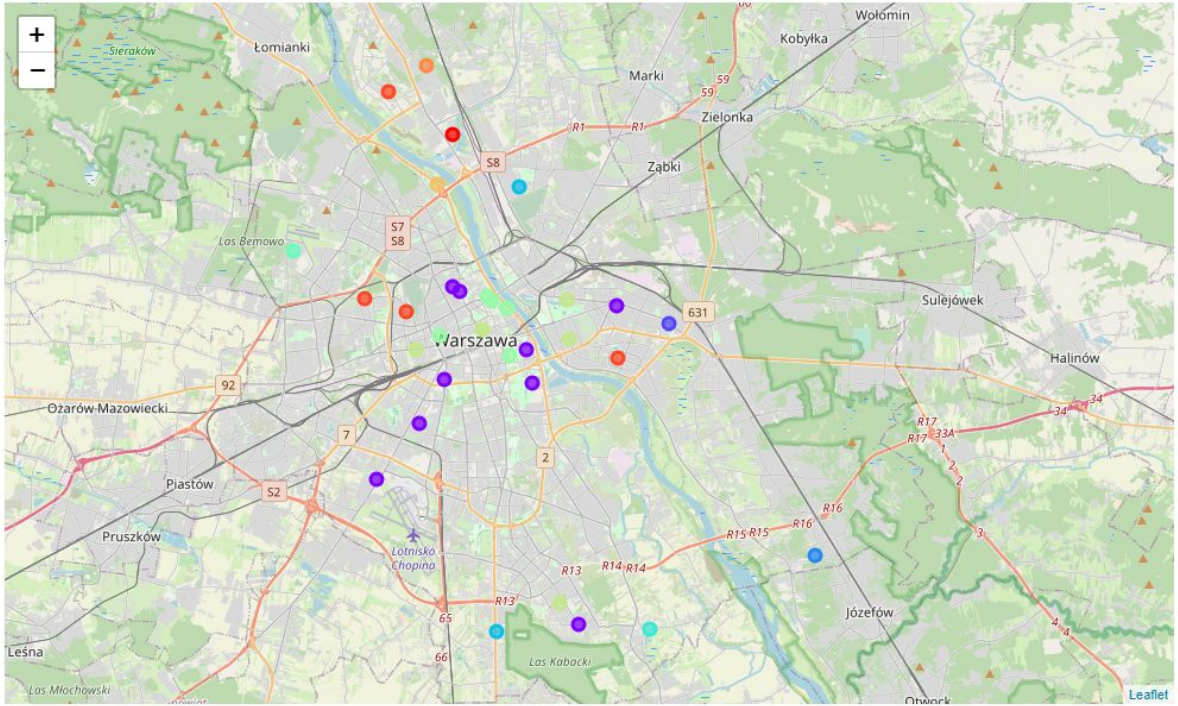
```
max_range = 15 #Max range 15 (number of clusters)

from sklearn.metrics import silhouette_samples, silhouette_score


indices = []
scores = []

for waw_clusters in range(2, max_range) :

    # Run k-means clustering
    waw_gc = waw_grouped_clustering
    kmeans = KMeans(n_clusters = waw_clusters, init = 'k-means++', random_state = 0).fit_predict(waw_gc)

    # Gets the score for the clustering operation performed
    score = silhouette_score(waw_gc, kmeans)

    # Appending the index and score to the respective lists
    indices.append(waw_clusters)
    scores.append(score)
```

- Plotting
  - Folium



Results

- Visualization of clusters



Discussion

- Most suitable neighbourhoods for starting the restaurant business are present in the cluster number 9.
- Our K-Means model worked perfectly and successfully clustered similar neighbourhoods together.
- After studying all four clusters, it is recommended to the client that neighbourhoods such as Czyste, Kamionek, Natolin, Saska Kępa that fall in cluster 9 look like good locations for starting their restaurant business.
- The client can go ahead and make a decision depending on other factors like availability and legal requirements that are out of scope of this project.

Conclusion

- Data analysis and machine learning techniques used in this project can be very helpful in determining solutions of certain business problems.
- Python’s inbuilt libraries such as GeoPy, Folium and BeautifulSoup make it very easy and effective to analyse a geographical location.
- In this project we studied the neighbourhoods of Warsaw city and came up with a recommendation of neighbourhoods where our client can start their restaurant business.