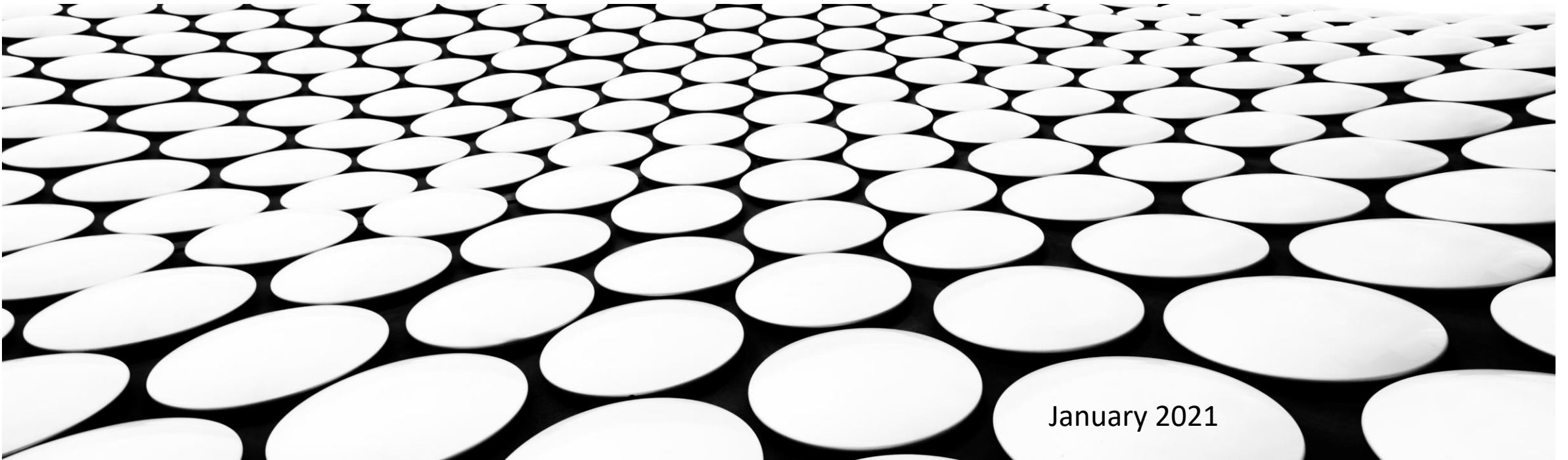

COURSERA CAPSTONE

IBM APPLIED DATA SCIENCE CAPSTONE

OPEN COFFEE SHOP IN CRACOW, POLAND



January 2021



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?

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

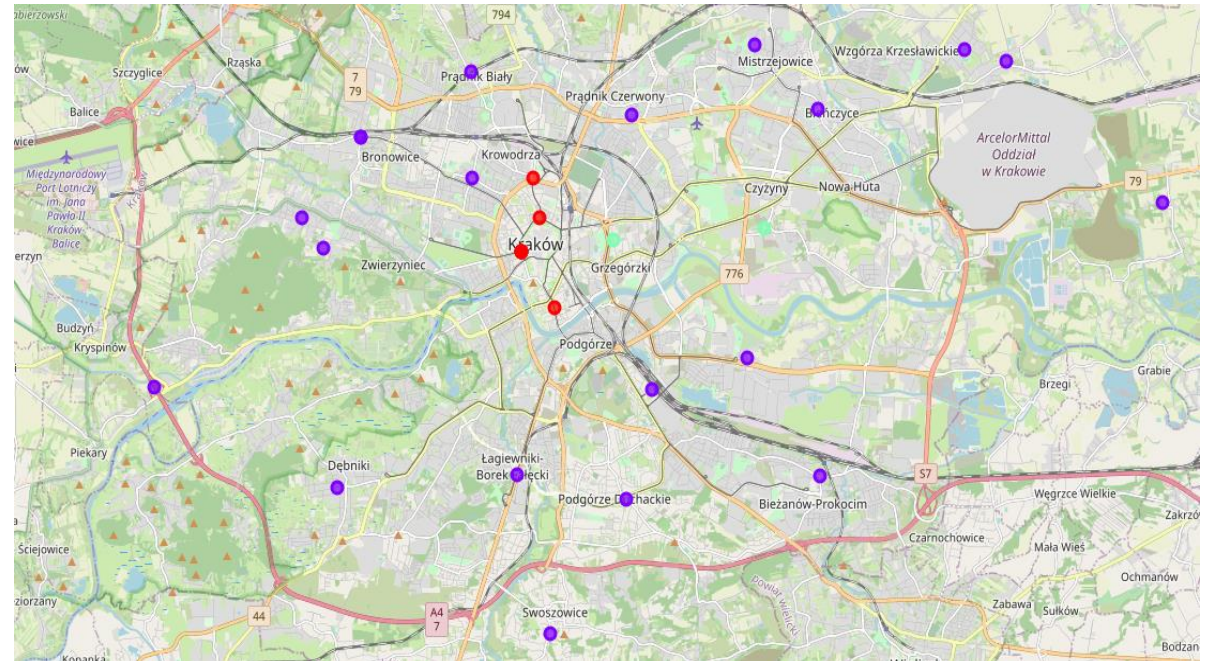
METHODOLOGY

- Web scraping Wikipedia page
- Get latitude and longitude
- Use Foursquare API
- Group data by neighborhood and taking the mean of the frequency of occurrence of each category
- Filter category by coffee shops
- Perform clusters
- Visualize the clusters on map

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.



DISCUSSION

- 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.



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.