

# Coursera Capstone IBM Data Science Capstone

*OPENING A NEW SHOPPING CENTER IN GERMANY*

BY HENRIK VÉROLET

# Business Question

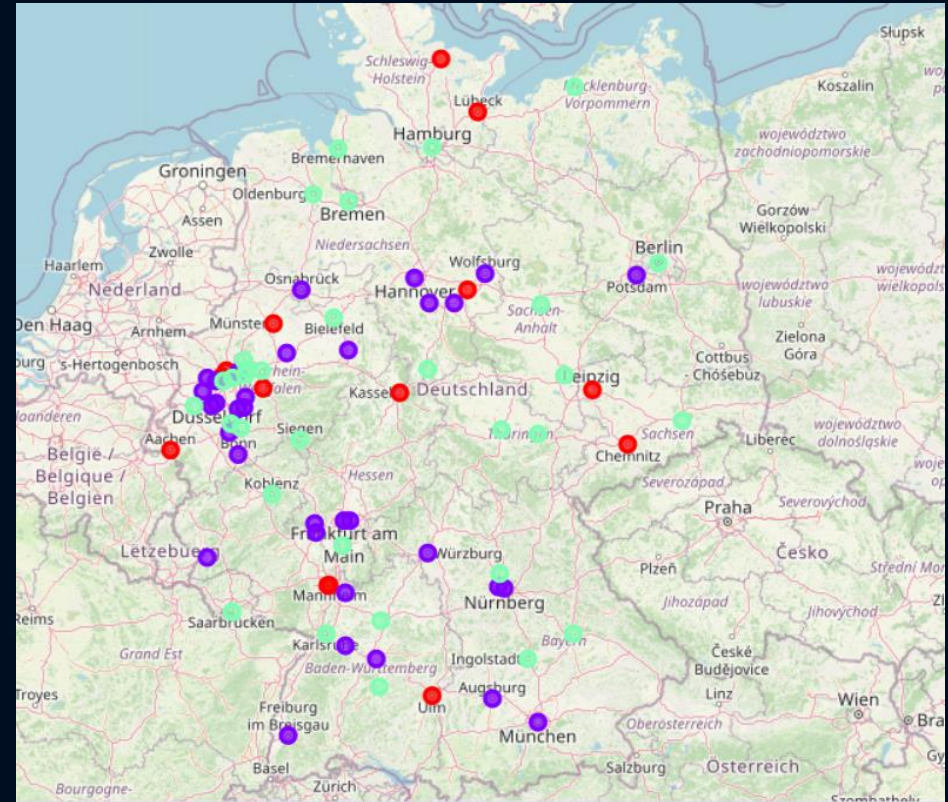
- Location of the shopping mall is one of the most important decisions that will determine whether the mall will be a success or a failure
- Objective: To analyse and select the best locations in Germany to open a new shopping mall
- This project is timely as the city is currently suffering from oversupply of shopping malls

# Methodology

- Web scraping Wikipedia page for cities list
- Get latitude and longitude coordinates using Geocoder
- Use Foursquare API to get venue data
- Group data by City and taking the mean of the frequency of occurrence of each venue category
- 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

- Categorized the cities into 3 clusters :
  - Cluster 0: cities with moderate number of shopping malls
  - Cluster 1: cities with low number to no existence of shopping malls
  - Cluster 2: cities with high concentration of shopping malls





# Discussion

- Highest number in cluster 2 and moderate number in cluster 0
- Cluster 1 has very low number to no shopping mall in the Cities

# Recommendation

- Open new shopping malls in Cities in cluster 1 with little to no competition
- Can also open in Cities in cluster 0 with moderate competition if have unique selling propositions to stand out from the competition
- Avoid Cities in cluster 2, already high concentration of shopping malls and intense competition

Thank You!