LOCATION FOR OPENING A NEW SHOPPING MALL IN BERLIN

IBM Applied Data Science Capstone

PROBLEM

- Shopping malls have become more and more popular for people to visit

- The problem of selecting the location for opening a new shopping mall is always difficult

- This project aims at solving the problem of choosing locations for opening a new shopping mall in Berlin

DATA

- The neighborhoods of Berlin

- The corresponding venue data of these neighborhoods.

METHODOLOGY

- Scrape the data of neighborhoods of Berlin from Wikipedia by Beautifulsoup

- Ger the geographical coordinates by using Geocoder package

- Apply Foursquare API to get venue data

- Use Folium to visualize the data

- Analyze the data by using k-means clusters

RESULTS

- Cluster 0 (red) : The neighborhoods with no shopping malls

- Cluster 1(green) : the neighborhoods with high number
- of shopping malls

- Cluster 2 (purple): the neighborhoods with medium number
- of shopping malls
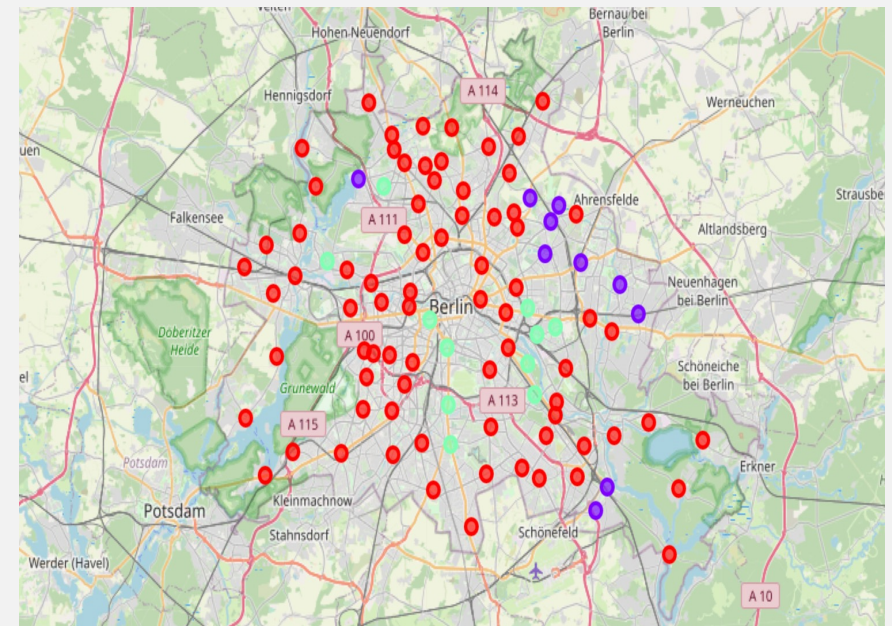
DISCUSSION

- Most of shopping malls are located in the city center

- Some ones are located outside the city center which are in the area of cluster 2

- There are still no shopping malls in the area of cluster 0.

CONCLUSIONS

- There will be a hard competition for the shopping malls in the areas of cluster 1

- The areas of cluster 0 may suffer the problem of lack of visitors

- It may be a good choice for a new shopping mall in the area of cluster 2 which is a balance of the avoiding hard competition and the lack of visitors

FUTURE WORK

- The problem can be further explored by considering more factors such as population distribution, investment const and so on, to build a more accurate model to help make better choices