

# **Coursera IBM Data Science with Python Specialization – Course 9 Capstone Project**

## **Contents:**

- 1. Business Problem**
- 2. Data Description**
- 3. Methods used for data exploration**
- 4. Results**
- 5. Discussion**
- 6. Conclusion**

## **1. Business Problem**

In Munich, if someone is looking to open a restaurant, where is the best place to open and what kind of restaurant is missing there? Munich is a city with a lot of different types of restaurants (italian, bavarian, german, indian, asian, etc. ).

Alex our persona wants to open a restaurant in Munich, he is experienced in Indian, Asian and Italian food and can not decide which of these restaurants he should open, so he can make more profit. So he wants to analyze the district around the store and find the best possible solution to open his restaurant.

questions to be answered:

1. what kind of restaurants are the most common ones in each district around his store?
2. how do the districts differ from each other?

## **2. Data Description**

For analyzing the data described data from the foursquare API in the city of Munich is used. The API was connected to Python and called through it. As each regular call is limited to 50 calls. For each type of restaurant, a separate call has been used to make most of the data.

For extracting each type of restaurant key words for searching were used, trial and error shows that a search for 'indian' gives better results than 'indian restaurant' this was also true for 'asian' and 'italian'. Therefore these key words were used to get data through the foursquare API and the data was cleaned afterwards.

## **3. Methods used for data exploration**

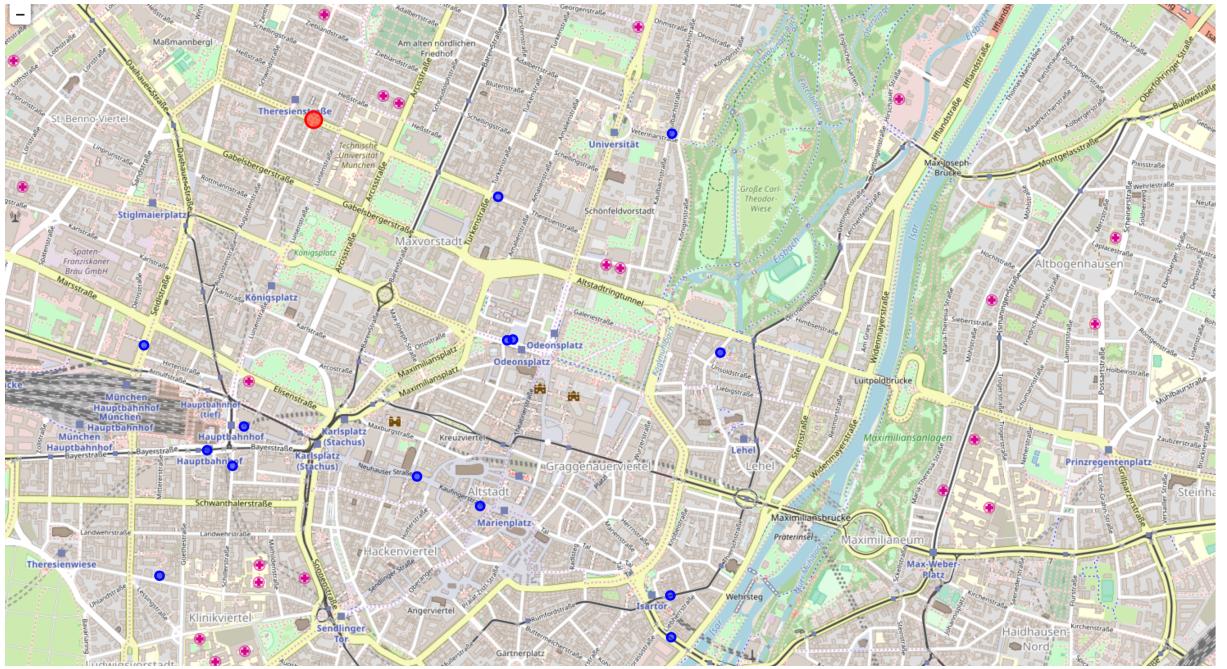
The general methodologies investigating the munich data include:

- - Data wrangling with Python
- - Pandas dataframe manipulation
- - Data visualization Folium
- - K-Means clustering with sklearn and KMeans

The imported data from foursquare was cleaned through additional functions in the panda library. Using functions like groupby, filtering columns etc. helped the cleaning of the data

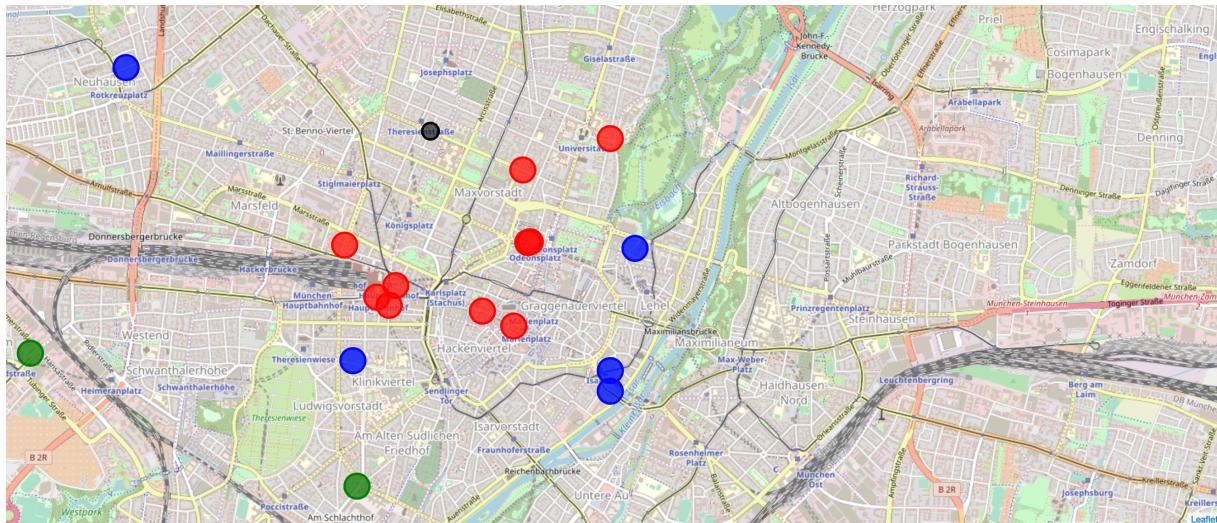
## **4. Results**

Using the Folium library, maps were created to initially analyze the restaurants around our store. (see figure below)



Our store is shown in red, other stores around it in blue

Using the k-Means method the restaurants around our store were clustered based on their category (Italian, indian, asian) the optimal number for k-means was obtained through trial and error and investigating the clusters. (see figure below)



Our store in black and three clustered restaurants around it

Further Analyzing the clusters show that in the red cluster there is only one indian restaurant, and that italian and asian food are the favorites around the store. The green and blue cluster have the most indian restaurant which have a larger distance A lack of indian restaurants around our store could be a good opportunity to open an indian restaurant

## 5. Discussion

Result from our investigations show tha data from foursquare in Munich is incomplete as the app is not used by many user. Most of the venues have no ratings or tips which could be analysed for this project to make better decision making.

Many of the restaurant around our store were not registered on foursquare which means we need more data for decision making. Many of the restaurants were also filtered out because of faulty information.

## **6. Conclusion**

Investigating data from Foursquare with data science tools shows first results that an Indian restaurant could be a good option for Alex, but the investigation also showed that data from Foursquare is faulty and incomplete and is not entirely trustable. Therefore, Alex should use more data resources to make a better decision. Many people in Munich use Google maps for rating etc. as Google maps API is not freely available, Alex should hire a data scientist and purchase the API to make the best out of his business.