**Finding the best location for the new pick-up point in Berlin**

Introduction/Business Problem:

Postal companies are very interested in providing the best service to the customers, when they are ordering their parcels to the pick-up points.

Its very crucial for every delivery service to find the best location for those type of delivery, which could serve as many customers as possible,

and be open/available for the customers for as many hours, as possible.

Pick-up points normally include different types of the location - post offices, parcel lockers, small convenient shops, gas stations etc.

One of the local Berlin express couriers has decided to place a new parcel locker in Treptow-Koepenick - one of the largest districts with the low density. Therefore bz placing a parcel locker, the courier can secure 24x7 access to their facility and very important that the location won't be man-powered which is even more secure during the pandemic.

As an analyst, I have been asked to find the best location for this new parcel locker in Treptow-Koepenick. As the area is quite large, wise idea would be to choose a location not far from one of the supermarkets at this area, as supermarkets are not always one of the main "attention" attractors in any districts, but also could guarantee an additional security to our future parcel locker, as in this case there would be always people around. Ideally and optionally would be great to have a look at the supermarket's working hours as this could help us to plan the customers visiting time to the parcel locker.

Data Collection and Usage:

I am planning to utilize the Foursquare API to get the location data on the supermarkets located in Treptow-Koepenick, after that i would analize and cluster this data in order to find out the optimal location to the parcel locker, which supposed to be located very close to the supermarket, also taking into account the working hours of the supermarket. After that I would be using the K-means algorithm to classify the data.

I would only be looking at the supermarkets located in specific areas and within the working hours after 8pm, because I am interested only in the supermarkets which are open late. Also, I have decided to take the center point as Rathaus Koepenick - this is the area which is not the geographical center, but the cultural center of the district. I would set up the large radius from it in order to use data at Foursquare.

Methodology and Analysis:

Let's get started with the fetching the data, which we would need before we start with analysis

*Data collection*

Using Foursquare account we can pull the data which would be used later for the data analysis - i am interested in the supermarkets (“categories”), located around Rathaus Koepenick ( i have found the latitude and longitude for this location - 52.4454164 and 13.5752346. Iam considering the distance around this central spot 2km (2 000 m).

After we have pulled the data, we should drop all the irrelevant columns, just not to get a confusing table at the end.

the table before the analysis would look like below:



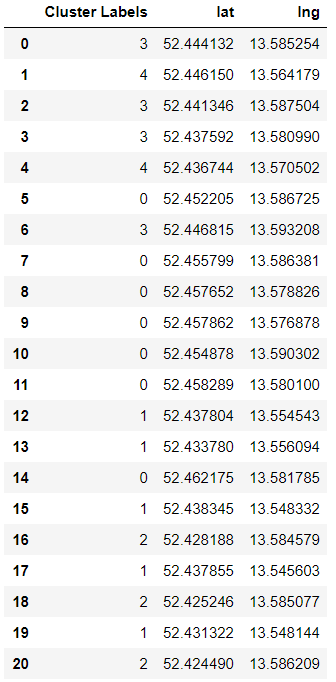
*Cleaning*

So we got a list with 21 different supermarkets surrounding Rathaus Koepenick within 2 km - some of them could be the supermarket from the same chain, but it doesn't matter for this case, as we ignore this factor for this analysis, and are only concerned about the distance. Therefore, lets go to the next step - clustering.

*Clustering*

At this step we would like to use K-means in order to cluster those supermarkets. As I am interested to see the groups of the supermarkets grouped by distance, I have chosen to implement ks=5.

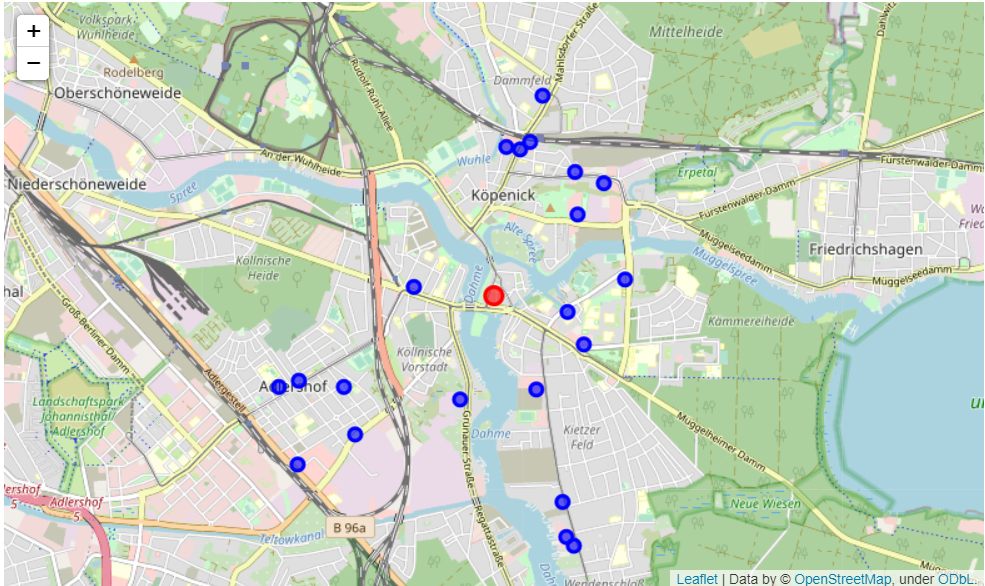
So, after that we would receive four clusters (from 0 to 4), below.



*Visualization*

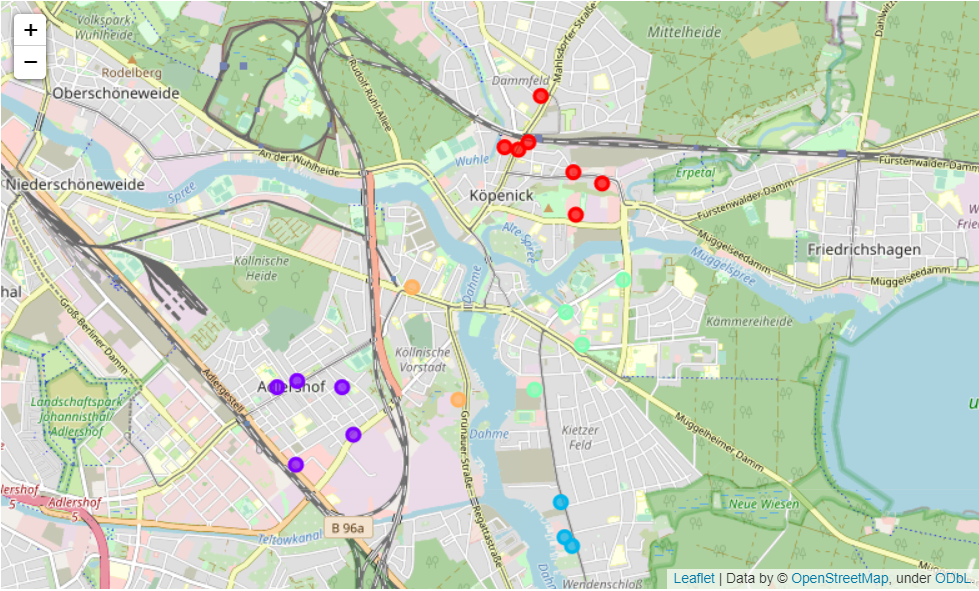
Based on the information we got at the previous step, i would like to visualize, firstly the supermarkets at the area, and their relation to the central point and secondly, the visualization of the clustering which we got on the table as well already.

At the picture below you could see the red spot in the middle - this is the central spot which we identified earlier. Blue points around are the supermarkets which are located within 2 km distance.



After clustering, the supermarket's map looks like the following (below). As you can see there are 5 different clusters, marked with different colours.

As we have chosen the distance, it's easy to see that the supermarkets which are located close to each other and at the same time located at the similar distance from the central point. Interesting to see (also logical) that there are more supermarkets are located closer to the center, and less at the suburbs of the Koepenick. This is related to the density of the district, and more people are living in the center of Koepenick, and way less (mostly one-storey building area). As a result, this visualization map below helps me to understand that at least one of the parcel lockers of the courier should be located in each cluster.



Results and recommendations:

It's important to mention that the map we got above is the first important step for finding the best location for the parcel locker in Koepenick. As we got 5 different clusters, based on the distance from the central spot, we could understand that it totally makes sense to have a parcel locker in each of the clusters, as it's visible that we could see the pattern of the location of the shops.

For the courier its important that their locker would be easily accessible for the customers, that means that the areas full of supermarkets would suit the best.

We already have the location of each of the supermarket, which would be very helpful to proceed with the next steps, which will help us to make the final decision.

Conclusion:

As mentioned earlier, this analysis is the first step to the complete project of finding the best spot for the parcel lockers in Berlin.

Due to this data research and clustering, we have found the potential spots, which is suitable for the placing parcel locker, however, there are more points, that need to be taken into consideration:

- Competitor analysis: there are several curriers, postal and parcel delivery companies in Germany, and specifically in Berlin, which have their lockers around the city, and this is very important, that our locker would be located strategically

- Cost calculation was not taken into account for this decision yet, so its important to analyse the pricing of placing the parcel locker in the potential place which have been chosen as the result of the analysis as well as the costs for the utilization this locker at the metined location

- Customers satisfaction - how customers would rate the chosen location, as locals opinion is also important

-In order to provide the best experience for the customers, as a next step it would be great to investigate the locations close to the supermarkets which are open late. Not all supermarkets in Berlin are open late (after 8pm) therefore, having a pick-up point at the location which will have customers traffic later in the evening, would at the same time will help the courier to plan the capacity (and as a result plan the delivery/collection to/from the locker)