THE BATTLE OF NEIGHBORHOOD

AUGUST 20, 2020

APOORVA SAXENA

Introduction

1.1 Background

The average American moves about eleven times in their lifetime. This brings us to the question: Do people move until they find a place to settle down where they truly feel happy, or do our wants and needs change over time, prompting us to eventually leave a town we once called home for a new area that will bring us satisfaction? Or, do we too often move to a new area without knowing exactly what we are getting into, forcing us to turn tail and run at the first sign of discomfort?

To minimize the chances of this happening, we should always do proper research when planning our next move in life. Consider the following factors when picking a new place to live so you do not end up wasting your valuable time and money making a move, you will end up regretting. Safety is a top concern when moving to a new area. If you do not feel safe in your own home, you are not going to be able to enjoy living there.

1.2 Problem

The crime statistics dataset of Berlin found on Kaggle has crimes in each Boroughs of Berlin from 2012 to 2019. The year 2019 being the latest we will be considering the data of that year which is old information as of now. The crime rates in each borough may have changed over time.

This project aims to select the safest borough in Berlin based on the total crimes, explore the neighborhoods of that borough to find the 10 most common venues in each neighborhood and finally cluster the neighborhoods using k-mean clustering.

1.3 Interest

Expats who are considering relocating to Berlin will be interested to identify the safest borough in Berlin and explore its neighborhoods and common venues around each neighborhood.

2. Data Acquisition and Cleaning

2.1 Data Acquisition

The data acquired for this project is a combination of data from three sources. The first data source of the project uses a Berlin crime data that shows the crime per borough in Berlin. The dataset contains the following columns:

- Year: Year of reported counts, 2012-2019
- District: Common name for Berlin borough
- Code: code for Lower Super Output Area in Greater Berlin
- Location: Neighborhood
- Robbery: Robbery incidents
- Street_robbery: Robbery incidents happened at the streets
- Injury: Crime which resulted in injury
- Agg_assault: Crile relating to aggravated assault
- Threat: Crile relating to issuing threats
- Theft: Crile relating to thefts
- Car: Crile relating to stealing of cars
- From_car: Crile relating to breaking into cars
- Bike: Crile relating to stealing of bikes

- Burglary: Crile relating to burglary
- Fire: Crile relating to causing fire
- Arson: Crile relating to arson
- Damage: Crile relating to causing damage
- Graffiti: Crile relating to graffiti
- Drugs: Crile relating to drugs

2.2 Data Cleaning

The data preparation for each of the three sources of data is done separately. From the Berlin crime data, the crimes during the most recent year (2019) are only selected (see *fig 2.1*).

	Borough	Neighborhood	Robbery	Street_robbery	Injury	Agg_assault	Threat	Theft	Car	From_car	Bike	Burglary	Fire	Arson	Damage	Graffiti	Drugs
0	Mitte	Tiergarten Süd	60	35	365	92	128	2271	15	198	296	55	13	6	347	77	231
1	Mitte	Regierungsviertel	42	20	554	136	152	3692	13	172	352	22	19	4	497	162	170
2	Mitte	Alexanderplatz	173	102	1966	500	420	11233	63	587	940	137	43	12	1307	381	1133
3	Mitte	Brunnenstraße Süd	40	29	268	64	79	1859	39	182	361	64	18	7	424	172	86
4	Mitte	Moabit West	66	29	685	210	202	2107	47	322	326	93	28	15	641	91	618

Fig 2.1 Berlin crime data after preprocessing

3. <u>Methodology</u>

3.1 Exploratory Data Analysis

3.1.1 Statistical summary of crimes

The describe function in python is used to get statistics of the Berlin crime data, this returns the mean, standard deviation, minimum, maximum, 1st quartile (25%), 2nd quartile (50%), and the 3rd quartile (75%) for each of the major categories of crime (See *fig 3.1.1*).

	Robbery	Street_robbery	Injury	Agg_assault	Threat	Theft	Car	From_car	Bike	Burglary	Fire	Arson	Damage	Graffiti	Drugs
count	12.000000	12.000000	12.000000	12.000000	12.000000	12.000000	12.000000	12.000000	12.000000	12.000000	12.000000	12.000000	12.000000	12.000000	12.000000
mean	33.250000	20.000000	315.333333	80.000000	106.750000	1465.916667	40.416667	187.250000	203.583333	57.083333	19.000000	8.000000	313.833333	70.750000	143.333333
std	23.775943	16.431677	167.226756	54.883347	45.104172	866.710968	10.689488	78.008304	119.960947	24.489175	6.980492	2.558409	147.709071	46.912152	170.997253
min	8.000000	5.000000	117.000000	27.000000	41.000000	529.000000	26.000000	83.000000	78.000000	22.000000	10.000000	4.000000	135.000000	24.000000	30.000000
25%	19.750000	10.750000	215.000000	46.000000	78.500000	853.000000	33.500000	130.500000	99.000000	38.250000	12.000000	5.750000	216.000000	36.750000	46.000000
50%	23.500000	13.000000	252.000000	56.500000	93.500000	1133.000000	38.500000	164.000000	188.500000	54.500000	19.500000	8.500000	258.000000	48.500000	63.000000
75%	44.000000	25.000000	382.250000	99.000000	134.000000	1866.000000	45.250000	241.500000	257.000000	78.500000	23.750000	10.250000	384.750000	92.250000	161.500000
max	91.000000	64.000000	690.000000	195.000000	196.000000	3236.000000	60.000000	319.000000	455.000000	97.000000	31.000000	11.000000	594.000000	162.000000	581.000000

Fig 3.1.1 Statistical description of the Berlin crimes

The count for each of the major categories of crime returns the value 12 which is the number of Berlin boroughs. ‘Theft’ is the highest reported crime during the year 2019 followed by ‘Damage’ and ‘Drugs’.

3.1.2 Boroughs with the highest crime rates

Comparing five boroughs with the highest crime rate during the year 2019 it is evident that Mitte has the highest crimes recorded followed by Friedrichshain-Kreuzberg, Tempelhof-Schöneberg, Neuköl and Steglitz-Zehlendorf. Mitte has a significantly higher crime rate than the other 4 boroughs (*see fig 3.1.2*).

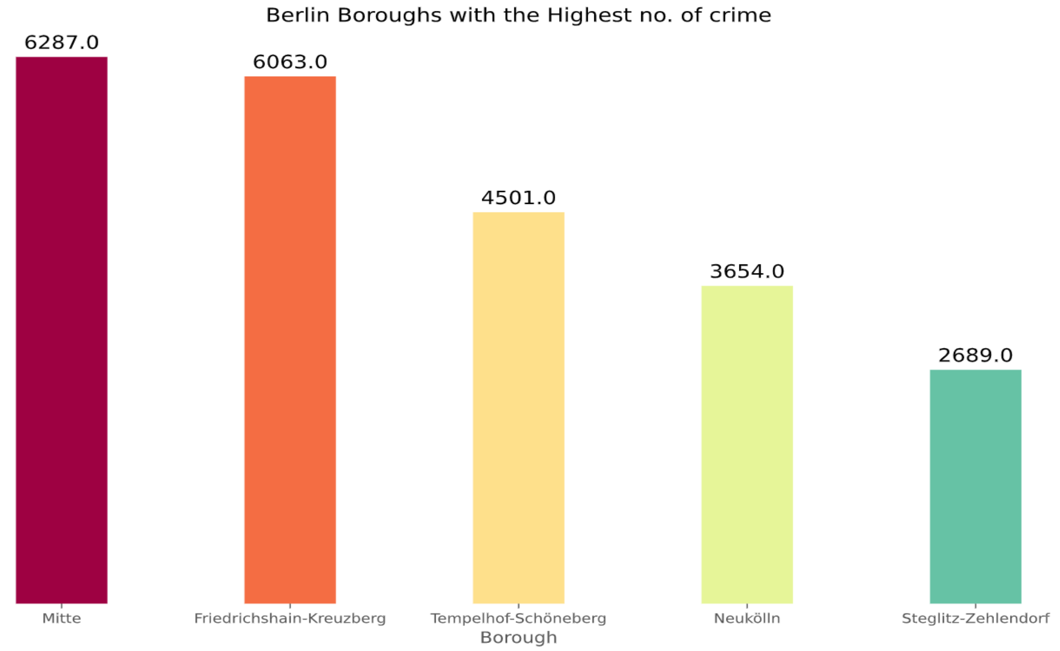

Fig 3.1.2 Boroughs with the highest crime rates

3.1.3 Boroughs with the lowest crime rates

Comparing five boroughs with the lowest crime rate during the year 2019, Treptow-Köpenick has the lowest recorded crimes followed by Lichtenberg, Reinickendorf, Marzahn-Hellersdorf and Spandau (*see fig 3.1.3*).

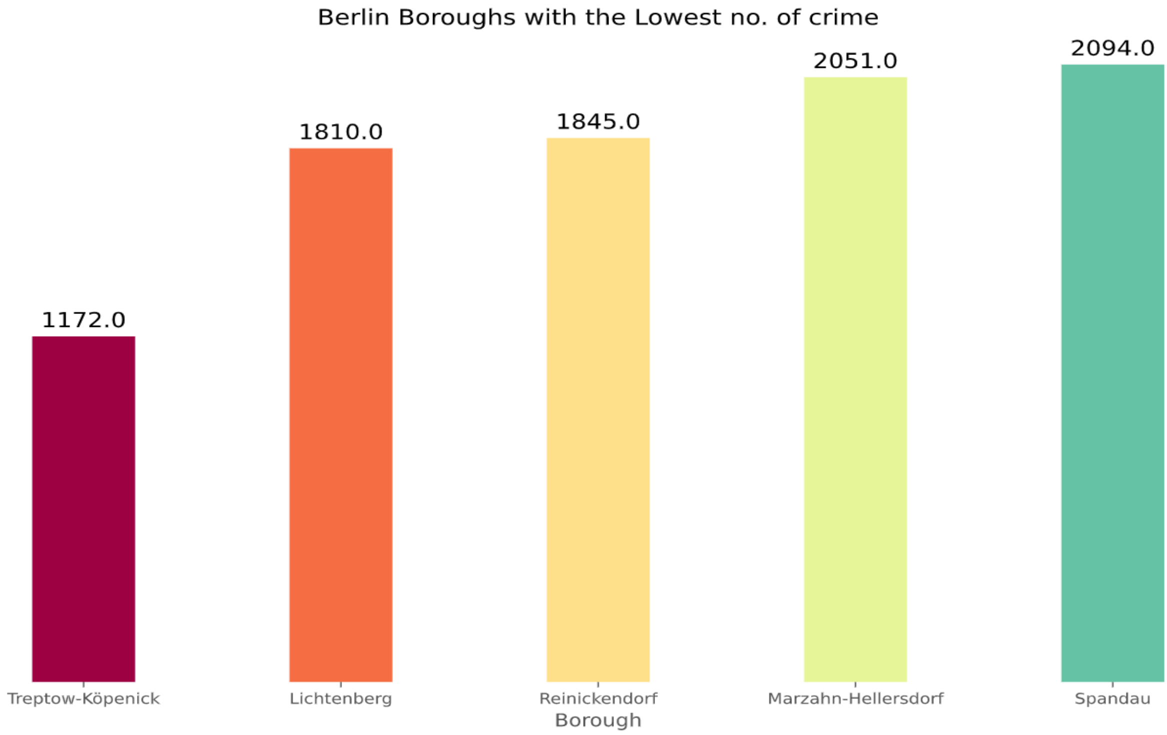

Fig 3.1.3 Boroughs with the lowest crime rates

Let us see the Treptow-Köpenick borough with the lowest crime rate as the safest borough in Berlin.

	Borough	Robbery	Street_robbery	Injury	Agg_assault	Threat	Theft	Car	From_car	Bike	Burglary	Fire	Arson	Damage	Graffiti	Drugs	Total
0	Treptow-Köpenick	8.0	5.0	117.0	27.0	41.0	529.0	28.0	83.0	102.0	22.0	10.0	5.0	135.0	30.0	30.0	1172.0

Fig 3.1.3.1 City of Berlin

3.1.4 Neighborhoods in Treptow-Köpenick

There are 17 neighborhoods in the royal borough of Treptow-Köpenick, they are visualized on a map using folium on python (*see fig 3.1.4*).

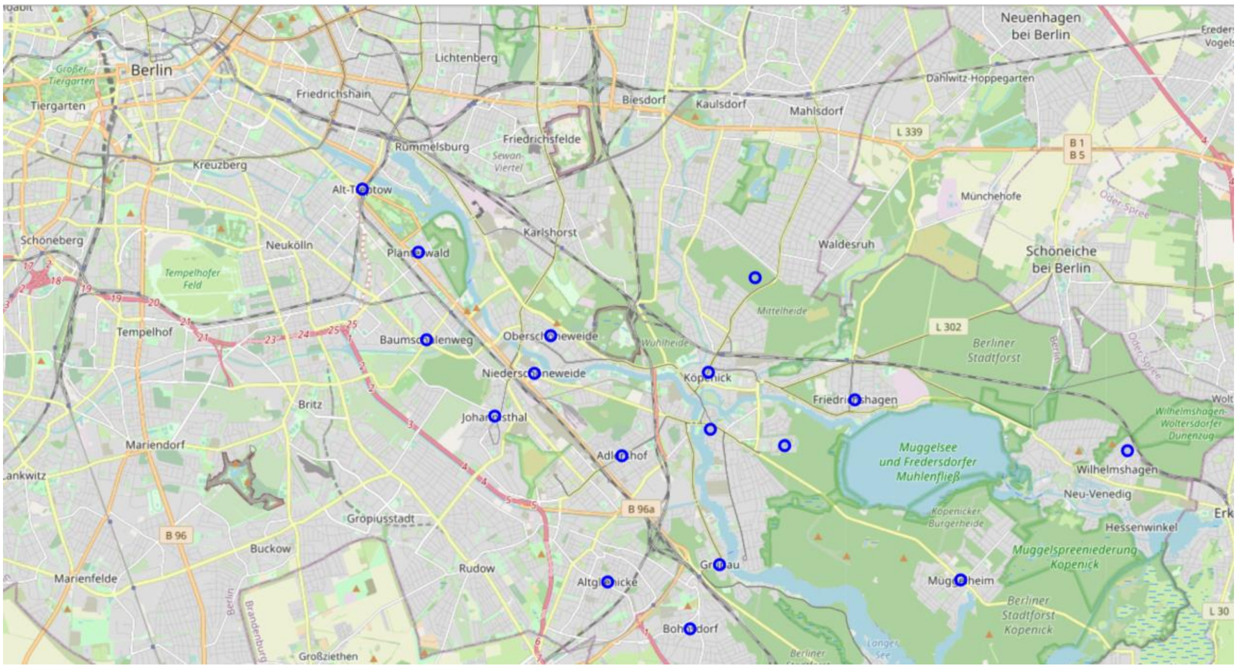

Fig 3.1.4 Neighborhoods in Treptow-Köpenick

3.2 Modelling

Using the final dataset containing the neighborhoods in Treptow-Köpenick along with the latitude and longitude, we can find all the venues within a 500-meter radius of each neighborhood by connecting to the Foursquare API. This returns a json file containing all the venues in each neighborhood which is converted to a pandas data frame. This data frame contains all the venues along with their coordinates and category (*see fig 3.2.1*).

[24]:

	Neighborhood	Neighborhood_Latitude	Neighborhood_Longitude	Venue	Venue_Latitude	Venue_Longitude	Venue_Category
0	Alt-Treptow	52.492563	13.459874	Der Holländer	52.491981	13.459416	Garden Center
1	Alt-Treptow	52.492563	13.459874	ELSE	52.495205	13.462637	Nightclub
2	Alt-Treptow	52.492563	13.459874	Tapas No. 6	52.490133	13.456732	Tapas Restaurant
3	Alt-Treptow	52.492563	13.459874	Molecule Man	52.497000	13.459093	Outdoor Sculpture
4	Alt-Treptow	52.492563	13.459874	Hafenräucherei Treptow "Löcknitz"	52.493991	13.463445	Seafood Restaurant

Fig 3.2.1 Venue details of each Neighborhood

One hot encoding is done on the venues data. (One hot encoding is a process by which categorical variables are converted into a form that could be provided to ML algorithms to do a better job in prediction). The Venues data is then grouped by the Neighborhood and the mean of the venues are calculated, finally the 10 common venues are calculated for each of the neighborhoods.

To help people find similar neighborhoods in the safest borough we will be clustering similar neighborhoods using K - means clustering which is a form of unsupervised machine learning algorithm that clusters data based on predefined cluster size. We will use a cluster size of 3 for this project that will cluster the 16 neighborhoods into 3 clusters. The reason to conduct a K- means clustering is to cluster neighborhoods with similar venues together so that people can shortlist the area of their interests based on the venues/amenities around each neighborhood.

4. <u>Results</u>

After running the K-means clustering we can access each cluster created to see which neighborhoods were assigned to each of the five clusters. Looking into the neighborhoods in the first cluster (*see fig 4.1*)

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Treptow-Köpenick	Alt-Treptow	52.492563	13.459874	0	Bakery	Café	Warehouse Store	Platform	Gym / Fitness Center	Vietnamese Restaurant	Electronics Store	Drugstore	Italian Restaurant	Light Rail Station
2	Treptow-Köpenick	Baumschulenweg	52.461694	13.481548	0	Asian Restaurant	Supermarket	Ice Cream Shop	Drugstore	Warehouse Store	Flower Shop	Doner Restaurant	Electronics Store	Falafel Restaurant	Farmers Market
3	Treptow-Köpenick	Johannisthal	52.445911	13.504547	0	Park	Tram Station	Taverna	Pizza Place	Sushi Restaurant	Movie Theater	Dessert Shop	Pub	Electronics Store	Convenience Store
4	Treptow-Köpenick	Oberschöneweide	52.462456	13.523476	0	Supermarket	Café	Burger Joint	Pizza Place	German Restaurant	Falafel Restaurant	History Museum	Asian Restaurant	Tram Station	Bank
5	Treptow-Köpenick	Niederschöneweide	52.454820	13.517877	0	Beach	Motel	Restaurant	Fast Food Restaurant	Italian Restaurant	Indian Restaurant	Ice Cream Shop	Miscellaneous Shop	Greek Restaurant	Convenience Store
6	Treptow-Köpenick	Adlershof	52.437893	13.547550	0	Trattoria/Osteria	Greek Restaurant	Bank	Supermarket	Steakhouse	Drugstore	Plaza	Italian Restaurant	Warehouse Store	Farmers Market
8	Treptow-Köpenick	Bohnsdorf	52.402243	13.570665	0	Ice Cream Shop	Electronics Store	Insurance Office	Italian Restaurant	Warehouse Store	Flower Shop	Doner Restaurant	Drugstore	Falafel Restaurant	Farmers Market
9	Treptow-Köpenick	Grünau	52.415427	13.580563	0	Dessert Shop	Boat or Ferry	Tram Station	Historic Site	Café	Restaurant	Doner Restaurant	Drugstore	Electronics Store	Falafel Restaurant
11	Treptow-Köpenick	Allende-Viertel	52.439869	13.602320	0	ATM	Café	Bus Stop	Food Court	Fast Food Restaurant	Doner Restaurant	Drugstore	Electronics Store	Falafel Restaurant	Farmers Market
12	Treptow-Köpenick	Altstadt-Kietz	52.443296	13.577312	0	Café	German Restaurant	Plaza	Greek Restaurant	Cocktail Bar	Palace	Ice Cream Shop	Sushi Restaurant	Bar	Park
13	Treptow-Köpenick	Müggelheim	52.412165	13.661954	0	American Restaurant	IT Services	Harbor / Marina	German Restaurant	Café	Warehouse Store	Fast Food Restaurant	Drugstore	Electronics Store	Falafel Restaurant
14	Treptow-Köpenick	Friedrichshagen	52.449380	13.626272	0	Café	Ice Cream Shop	Park	Brewery	German Restaurant	Flower Shop	Italian Restaurant	Drugstore	Organic Grocery	Cajun / Creole Restaurant
15	Treptow-Köpenick	Rahnsdorf/Hessenwinkel	52.438882	13.718101	0	Post Office	Café	Light Rail Station	Warehouse Store	Fast Food Restaurant	Doner Restaurant	Drugstore	Electronics Store	Falafel Restaurant	Farmers Market
16	Treptow-Köpenick	Dammvorstadt	52.454932	13.576645	0	Clothing Store	Electronics Store	Drugstore	Café	Shoe Store	Gym / Fitness Center	German Restaurant	Indian Restaurant	Doner Restaurant	Burger Joint

Fig 4.1 Cluster 1

The cluster one is the biggest cluster with 12 of the 16 neighborhoods in the borough Treptow-Köpenick. Upon closely examining these neighborhoods we can see that the most common venues in these neighborhoods are Restaurants, and bus stop and train station.

Looking into the neighborhoods in the second and third clusters, we can see these clusters have two and one neighborhood, respectively. This is because of the unique venues in each of the neighborhoods, hence they could not be clustered into similar neighborhoods (*see figures 4.2, 4.3)*.

Cluster 2

```
cluster_2 = df_merge[df_merge['Cluster Labels']==1]
cluster_2
```

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	Treptow-Köpenick	Plänterwald	52.479544	13.478808	1	Bus Stop	Supermarket	Light Rail Station	Warehouse Store	Fast Food Restaurant	Doner Restaurant	Drugstore	Electronics Store	Falafel Restaurant	Farmers Market
10	Treptow-Köpenick	Köpenick-Süd	52.474385	13.592336	1	Tram Station	Bus Stop	Warehouse Store	Flower Shop	Doner Restaurant	Drugstore	Electronics Store	Falafel Restaurant	Farmers Market	Fast Food Restaurant

Fig 4.2 Cluster 2

Cluster 3

```
[48]: cluster_3 = df_merge[df_merge['Cluster Labels']==2]
cluster_3.head()
```

[48]:

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
7	Treptow-Köpenick	Altglienicke	52.411838	13.542646	2	Shipping Store	Warehouse Store	Fast Food Restaurant	Dessert Shop	Doner Restaurant	Drugstore	Electronics Store	Falafel Restaurant	Farmers Market	Flower Shop

Fig 4.3 Cluster 3

Visualizing the clustered neighborhoods on a map using the folium library (see fig 4.5).

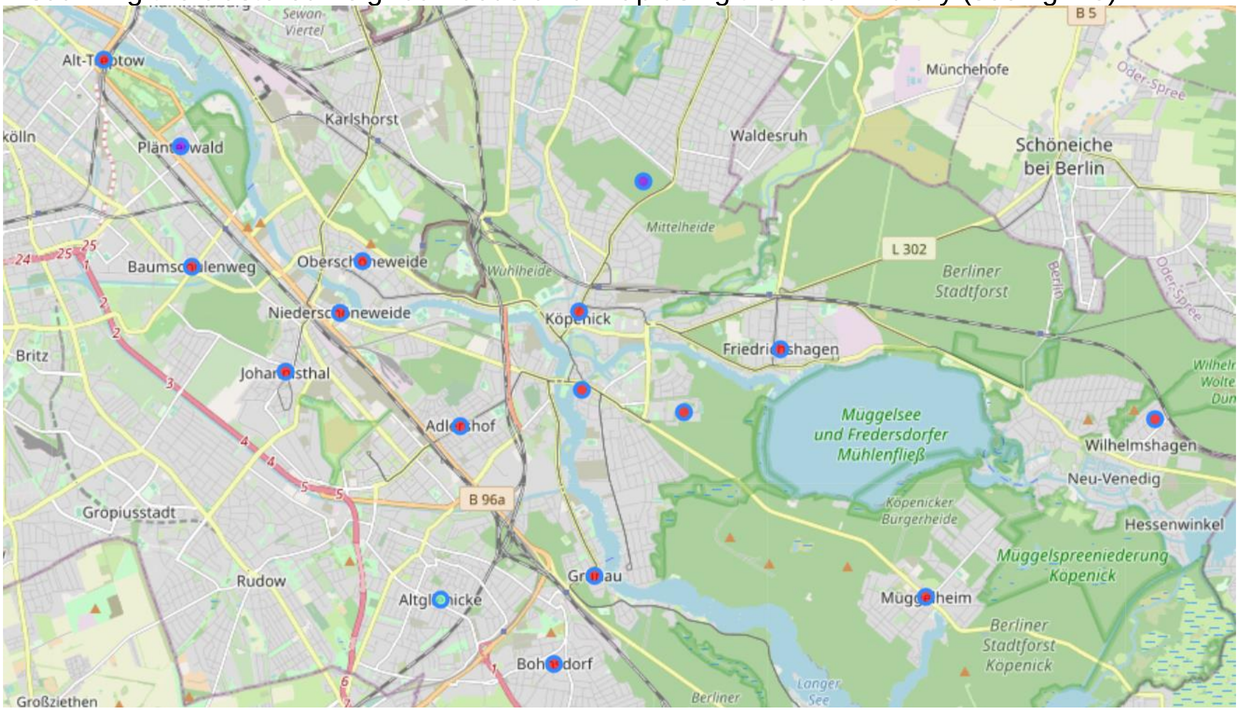

Fig 4.5 Clustered neighborhoods in the Borough of Treptow-Köpenick

Each cluster is color coded for the ease of presentation.

5. <u>Discussion</u>

The aim of this project is to help people who want to relocate to the safest borough in Berlin, expats can choose the neighborhoods to which they want to relocate based on the most common venues in it. For example, if a person is looking for a neighborhood with good connectivity and public transportation, we can see that Clusters 1 has Train stations and Bus stops as the most common venues. If a person is looking for a neighborhood with stores and restaurants in a proximity, then the neighborhoods in clusters 2 & 3 are suitable. For a family I feel that the neighborhoods in Cluster 2 are more suitable dues to the common venues in that cluster, these neighborhoods have common venues such as Parks, Gym/Fitness centers, Restaurants and Electronics Stores which is ideal for a family.

6.<u>Conclusion</u>

This project helps a person get a better understanding of the neighborhoods with respect to the most common venues in that neighborhood. It is always helpful to make use of technology to stay one step ahead i.e. finding out more about places before moving into a neighborhood. We have just taken safety as a primary concern to shortlist the borough of Berlin. The future of this project includes taking other factors such as cost of living in the areas into consideration to shortlist the borough based on safety and a predefined budget.