

CLUSTERING NEIGHBOURHOODS IN BERLIN  
CITY (GERMANY) TO FIND OUT WHICH  
NEIGHBOURHOODS REQUIRE MORE PUBLIC  
PARKS, BASED ON PUBLIC PARKS AVAILABLE  
TO RESIDENTS

CAPSTONE PROJECT



# BACKGROUND

1. City parks play a vital role in the **social**, **economic**, and **physical well-being** of cities and their residents.
2. City parks provide access to recreational opportunities, increase property values, spur local economies, combat crime, and protect cities from environmental impact.
3. City parks encourage **active lifestyles** and **reduce health costs** also parks help clean the air and improve public health.



# PROBLEM

1. Parks can only reach their peak potential if **every household can access such green space** and that access shouldn't be tied to a private vehicle or an unreasonable walking distance.
2. Unequal park access is creating another form of environmental injustice.
3. There is an urgency to understand park access at the metropolitan scale. In response, metropolitan areas **need regional plans** to bring **parks closer to all residents**.
4. **Local planners** and real **estate developers** **lack data-driven insights** which will help them to create a plan which will reduce spatial and economic barriers tied to park access.



# REQUIRED SOLUTION

1. A **summary generation of parks distribution** is required to reduce spatial and economic barriers tied to park access.
2. This can be done by **clustering the localities** into various segments and figure out the localities which require immediate attention and the ones which have optimum park access.
3. This summary generation will help to **build more parks with data driven insights** and prioritize the localities for the same.



# DATA SOURCES

- **Berlin Neighbourhood Data:**

[https://en.wikipedia.org/wiki/Boroughs\\_and\\_neighborhoods\\_of\\_Berlin](https://en.wikipedia.org/wiki/Boroughs_and_neighborhoods_of_Berlin)

The data includes 96 localities in Berlin city.

- **Location Coordinates for each Neighbourhood:**

The Geocoder Python package will be used to get latitude and longitude of each Neighbourhood.

- **Berlin Venue Recommendations from FourSquare API**

(FourSquare website: [www.foursquare.com](http://www.foursquare.com))



# EXPLORATORY DATA ANALYSIS

Total Localities : 96

Total Parks found : 66

Total Localities with parks : 40

Total number of localities with no parks : 56

Hence 66 parks were found to be distributed over 40 localities with the number of parks ranging from 1 to 5 per locality.

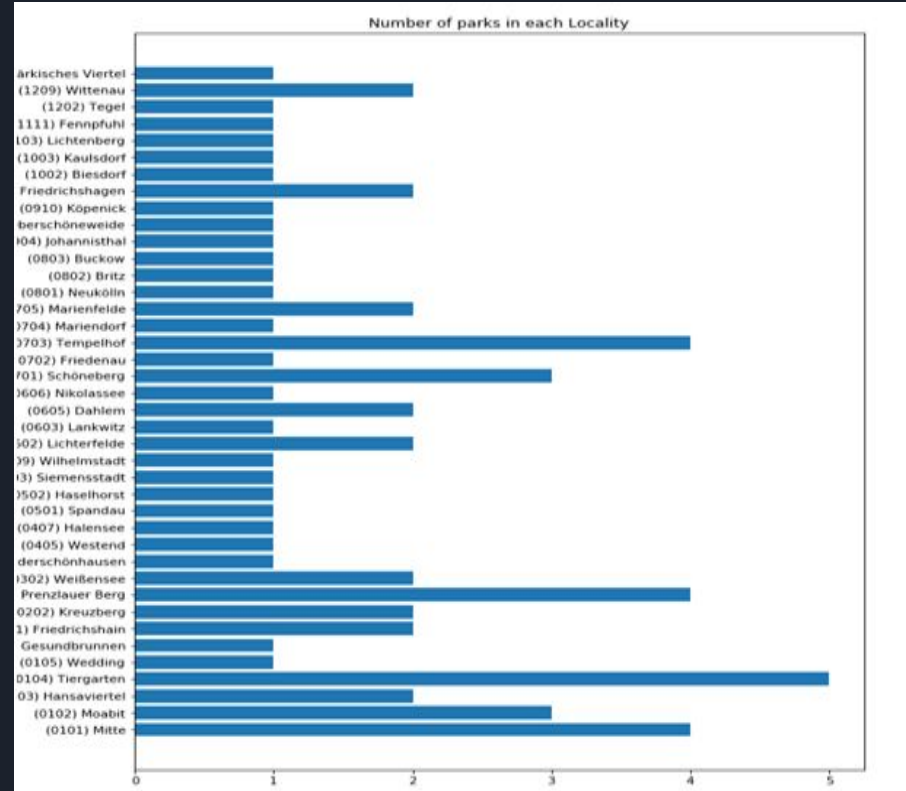
# EXPLORATORY DATA ANALYSIS

Distribution of parks in the 40 localities is as shown in the graph.(only those localities are considered that those have parks)

X-axis represents number of parks and Y-axis represents locality.

Highest number of parks are present in Tiergarten locality the number of parks is 5.

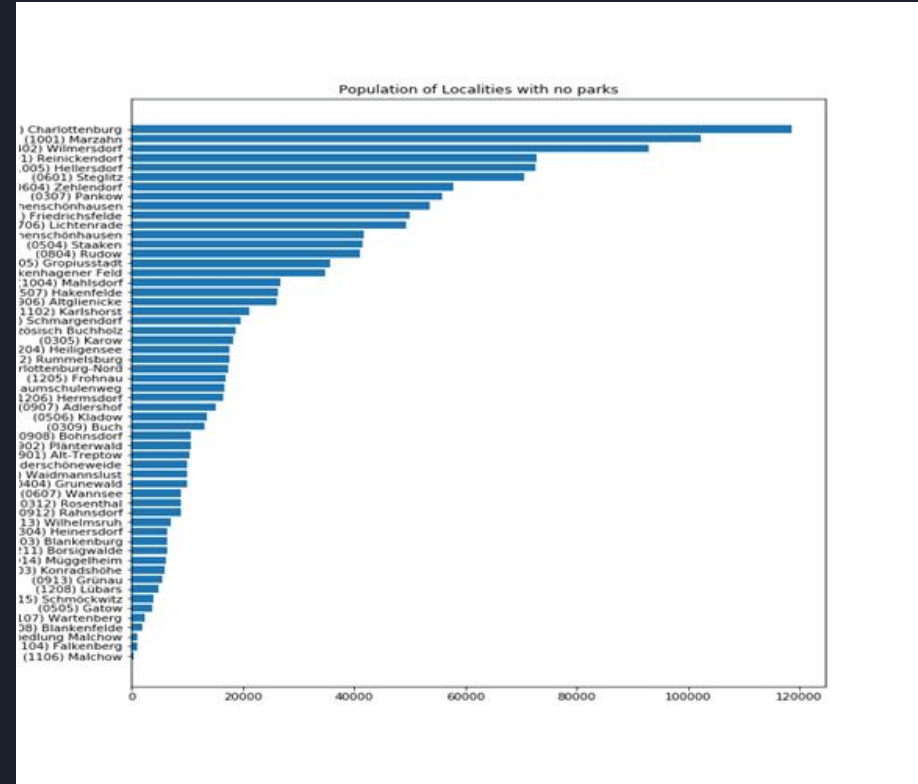
Maximum localities consist 1 park but is 1 park optimum or not will depend on the area and population of the locality.



# EXPLORATORY DATA ANALYSIS

The following graph shows the localities with no parks and corresponding population.

Top 5 localities in this category have population more than 70,000.





# CLUSTERING

K-means clustering resulted in the following clusters.

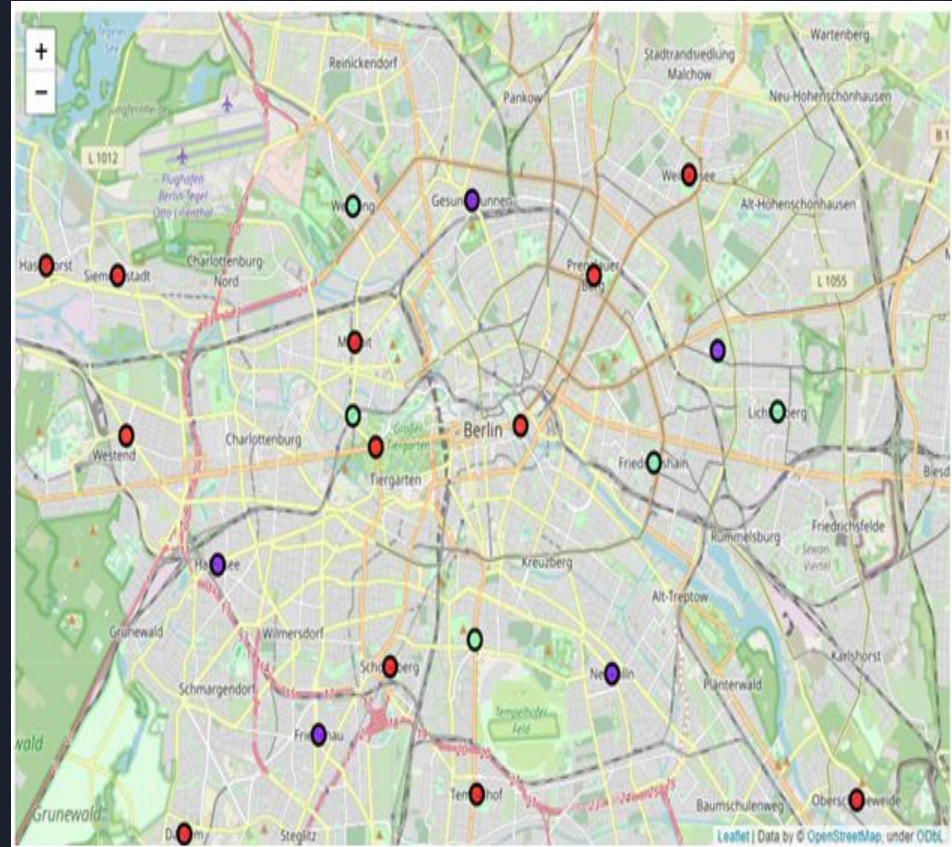
The partitioning was based on following attributes population, area of locality and number of parks.

Three clusters were created (only localities which had parks were chosen)

Cluster 0 - 24 localities

Cluster 1 - 6 Localities

Cluster 2 - 10 Localities



## CLUSTER 0

Localities can be categorized as **excellent** localities based on the number parks present for the residents in these localities i.e. according to the number of residents **sufficient** parks are present and do not require immediate attention as far as building more parks is concerned. But Maintaining these parks is a must.

## CLUSTER 1

Localities can be categorized as good localities. This is due to the small size of these localities. Parks are present in all of these but they are not sufficient for the residents. Also, these are densely populated regions with a smaller area and bigger population so due to smaller size many parks can't be built but a few more parks are a must.

## CLUSTER 2

Localities can be categorized as **average** localities based on the number parks present for the residents in these localities the area of these localities is sufficiently moderate and the number of parks is low compared to the number of residents.

There are 56 localities with No parks and out of these 5 localities have a population of more than 70,000 these need urgent efforts in planning and building of parks.



# CONCLUSION

Hence building more parks has the following priority according to the localities:

1. First, the localities with no parks at all also, among them the top 5 with a huge population.
2. Second priority is cluster 2 as the area is sufficient to build more parks and population to parks ration is moderately high.
3. Third will be cluster 1 due to population to parks ration is high but the area of these localities is small
4. Fourth are the 24 localities in cluster 0 having an optimum number of parks!!