

2. Data Sources and Description

2.1. Data Sources

2.1.1. <https://www.cybo.com> is a popular Global Business Directory which covers a vast majority of countries of the world. The intention is to scrape the information available about Cologne, Germany. <https://postal-codes.cybo.com/germany/cologne/> Information like, postal codes, Neighborhood, Population, Area, Male Population, Female Population, Median Age, Latitude, Longitude information can be sourced.

Information	
Timezone	Central European Time
Area	281.7 km ²
Population	977,691 (More Details)
Male Population	473,826 (48.5%)
Female Population	503,864 (51.5%)
Median Age	41.3
Postal Codes	50667, 50668, 50670 (43 more)
Area Codes	2203, 221

Figure 1 Generic Information on Cologne City

Information	
Population	977,691
Population Density	3,470 / km ²
Male Population	473,826 (48.5%)
Female Population	503,864 (51.5%)
Median Age	41.3
Male Median Age	40.5
Female Median Age	42
Businesses in Cologne	85,523
Population (1975)	702,101
Population (2000)	892,156
Population change from 1975 to 2015	+39.3%
Population change from 2000 to 2015	+9.6%

Figure 2 Further General Information with past Population Data

Postal Code	City	Administrative Region	Population	Area	Area Codes	Neighborhoods	Latitude	Longitude	Male Population	Female Population
50667	Cologne	North Rhine-Westphalia	4,512	1.031 km²	2203, 221	Altstadt-Nord, Innenstadt	50.940088623517646°	6.957684918323491°	2,187 (48.5%)	2,3 (51.5)
50668	Cologne	North Rhine-Westphalia	7,717	1.692 km²	2203, 221	Altstadt-Nord, Innenstadt, Kuniibertsviertel	50.950695435874°	6.965450051394539°	3,740 (48.5%)	3,9 (51.5)
50670	Cologne	North Rhine-Westphalia	8,612	1.9 km²	2203, 221	Agnesviertel, Innenstadt	50.95117220931733°	6.950957439590489°	4,174 (48.5%)	4,4 (51.5)
50672	Cologne	North Rhine-Westphalia	4,657	1.027 km²	2203, 221	Friesenviertel, Innenstadt	50.943440989539745°	6.938476189303402°	2,257 (48.5%)	2,4 (51.5)
50674	Cologne	North Rhine-Westphalia	6,643	1.389 km²	2203, 221	Innenstadt, Kuniibertsviertel	50.93358384190404°	6.937040819338811°	3,219 (48.5%)	3,4 (51.5)

Figure 3 DataFrame of Scraped Table from source

2.1.2. Cologne City Administration publishes many of the public data <https://www.offenedaten-koeln.de> of which a GeoJSON file for the city boundaries and boundaries of postal codes are used to identify the correct location of neighborhoods. <https://www.offenedaten-koeln.de/dataset/postleitzahlgebiete-koeln>

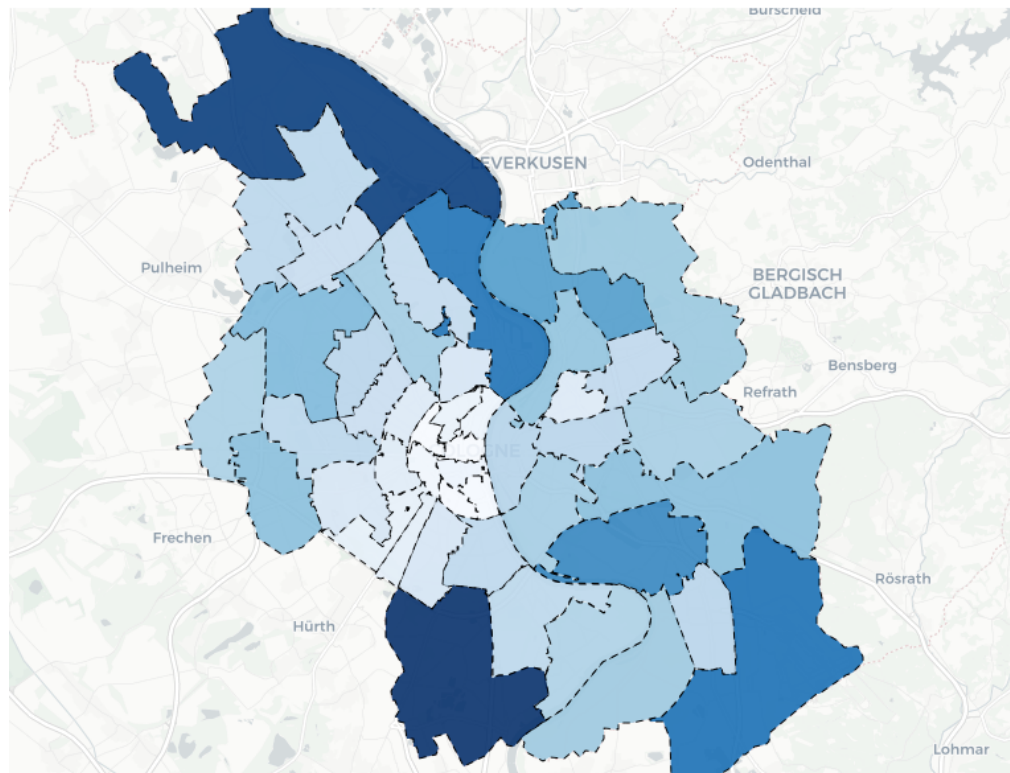


Figure 4 Cologne Postal Codes Boundaries showing Population in regions

2.1.3. Foursquare API will be used to fetch the venue information, venue categories by passing in the information obtained from above sources.

Postal Code	PostCode	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
51149	50.906367	7.048909		Packstation 132	50.903793	7.058213	Shipping Store
51149	50.906367	7.048909		Rangierbahnhof Köln-Gremberg	50.899960	7.052129	Train Station
51149	50.906367	7.048909		A.T.U Köln - Gremberghoven	50.911754	7.056019	Automotive Shop
51149	50.906367	7.048909		Kienbaum Consultants International	50.913817	7.050480	Business Service
51149	50.906367	7.048909		CANCOM Pironet	50.914348	7.048726	Business Service

Figure 5 Venue Information from Foursquare API

2.1.4. Folium is a Data Visualisation library that lets us manipulate data in python and visualize the data in leaflet.js library. Folium will be used for Data Visualization.

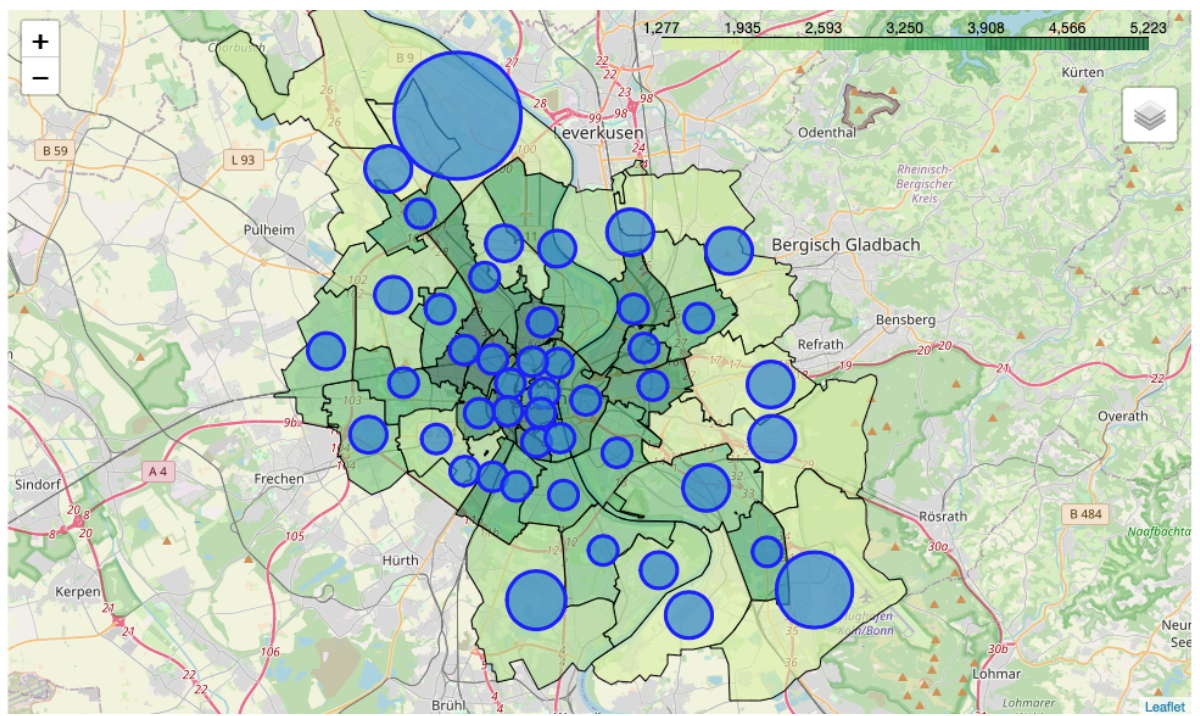


Figure 6 Folium Map showing used Radius

2.2. Data Explanation

- 2.2.1. The scraped data from cybo, will be used as all the information available about Cologne, primarily Postal Codes, Neighborhoods names, Population in each neighborhood and its Area along with Latitude and Longitude.
- 2.2.2. The boundaries GeoJSON file is obtained from Open Data of Cologne website which will help in visualization of region under lens and shadow regions.
- 2.2.3. Foursquare API is used to fetch the venue information in the region which is restricted to 100 venues in a query and a radius is chosen dynamically.
- 2.2.4. Folium is used to plot the map data available
- 2.2.5. Machine Learning techniques will be used to find the optimum location for Supermarket as part of solution to the business problem.