Coursera Capstone Project

IBM Data Science Professional Certificate

Strategic Locations for Supermarket-Chain in Berlin, Germany

Helal Chowhdury

April 2020

Business Problem

- Background: Population and visitors in Berlin are on the rise, flood of asylums lead to increasing demand for new supermarket

- Business question: In Berlin, if a local or foreign supermarket chain looks for opening few new branches, which strategic locations should be preferred considering business potential?

- Objective: Analyze and recommend the set of 5 best neighborhoods for new branch of a supermarket

- Challenge: Selecting strategic location considering high business potential and less risk

Data

- **Required Data :**
  - Neighborhoods, population, density
  - Latitude, longitude or each neighborhood
  - Venue data, especially accurate supermarket data

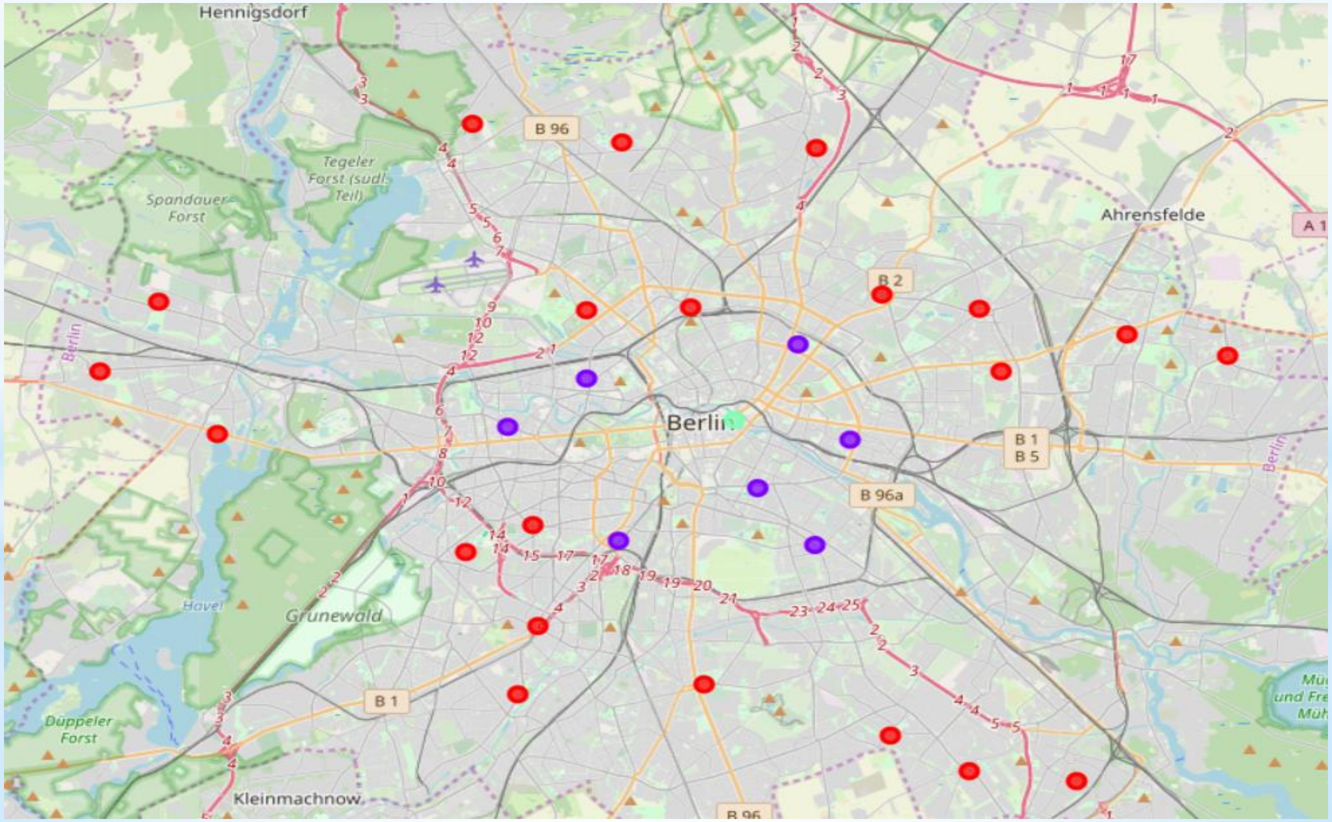
- **Sources of Data:**
  - Wikipedia: en.wikipedia.org/wiki/Boroughs_and_neighborhoods_of_Berlin
  - Geocoder for latitude, longitude data
  - Forsquare APIs for Venue data

Methodology

- Use web scrapping technique for wiki page
- Get lat/log data, then get venue data, 96 neighborhoods
- Preprocess data, group by high population, reduced to 50
- Group by higher density, reduced to 40
- Check supermarket category from all 1964 venues
- Find number of supermarkets for each neighborhood
- Find population per supermarket, take top 30
- Consider top 10 as expected locations
- Apply K-means to find less competitive neighborhoods
- Select the set of best 5

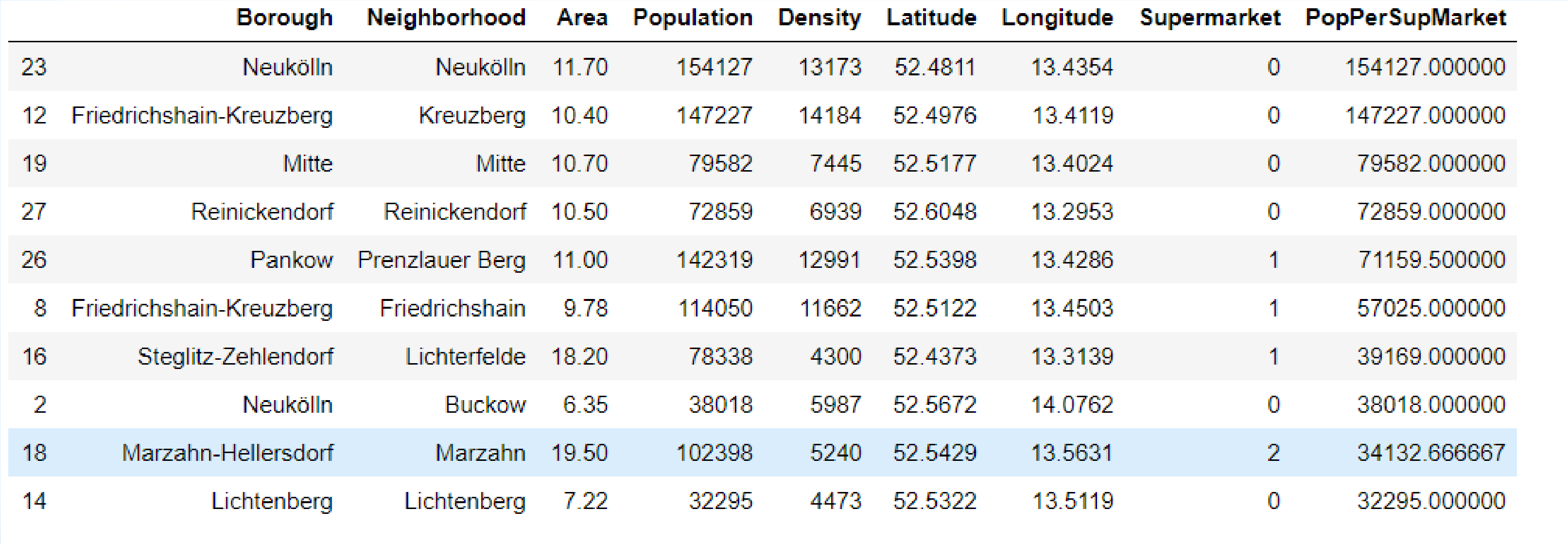
Result

- Separate locations where supermarket is not/less common: found 8 from K-means
- Compare with top 10 of PopPerSupMarket and pick the set of best 5

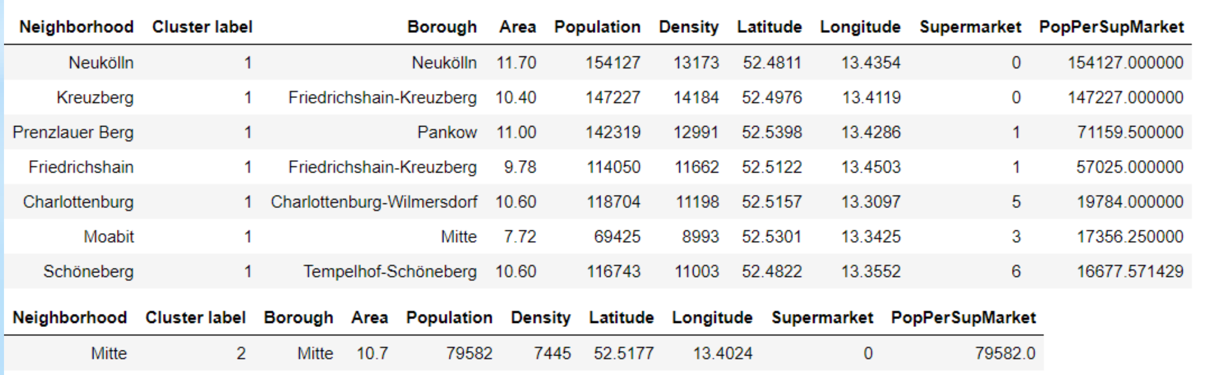


Result – Pick the best

- Top five: Neukolln, Kreuzberg, Mitte, Prenzlauerberg, and Friedrichshain



	Borough	Neighborhood	Area	Population	Density	Latitude	Longitude	Supermarket	PopPerSupMarket
23	Neukölln	Neukölln	11.70	154127	13173	52.4811	13.4354	0	154127.000000
12	Friedrichshain-Kreuzberg	Kreuzberg	10.40	147227	14184	52.4976	13.4119	0	147227.000000
19	Mitte	Mitte	10.70	79582	7445	52.5177	13.4024	0	79582.000000
27	Reinickendorf	Reinickendorf	10.50	72859	6939	52.6048	13.2953	0	72859.000000
26	Pankow	Prenzlauer Berg	11.00	142319	12991	52.5398	13.4286	1	71159.500000
8	Friedrichshain-Kreuzberg	Friedrichshain	9.78	114050	11662	52.5122	13.4503	1	57025.000000
16	Steglitz-Zehlendorf	Lichterfelde	18.20	78338	4300	52.4373	13.3139	1	39169.000000
2	Neukölln	Buckow	6.35	38018	5987	52.5672	14.0762	0	38018.000000
18	Marzahn-Hellersdorf	Marzahn	19.50	102398	5240	52.5429	13.5631	2	34132.666667
14	Lichtenberg	Lichtenberg	7.22	32295	4473	52.5322	13.5119	0	32295.000000



Neighborhood	Cluster label	Borough	Area	Population	Density	Latitude	Longitude	Supermarket	PopPerSupMarket
Neukölln	1	Neukölln	11.70	154127	13173	52.4811	13.4354	0	154127.000000
Kreuzberg	1	Friedrichshain-Kreuzberg	10.40	147227	14184	52.4976	13.4119	0	147227.000000
Prenzlauer Berg	1	Pankow	11.00	142319	12991	52.5398	13.4286	1	71159.500000
Friedrichshain	1	Friedrichshain-Kreuzberg	9.78	114050	11662	52.5122	13.4503	1	57025.000000
Charlottenburg	1	Charlottenburg-Wilmersdorf	10.60	118704	11198	52.5157	13.3097	5	19784.000000
Moabit	1	Mitte	7.72	69425	8993	52.5301	13.3425	3	17356.250000
Schöneberg	1	Tempelhof-Schöneberg	10.60	116743	11003	52.4822	13.3552	6	16677.571429

Neighborhood	Cluster label	Borough	Area	Population	Density	Latitude	Longitude	Supermarket	PopPerSupMarket
Mitte	2	Mitte	10.7	79582	7445	52.5177	13.4024	0	79582.0

Discussion

- Clusters of 22, 7 and 1 neighborhoods
- Top five are the best according higher population, higher density, lower competition, similarity and higher PopPerSupMarket
- Top 5 are in the vicinity of city center
- The 4th one Reinickendorf is not considered: low density, far from the city center

Conclusion and outlook

- Top 5 locations are selected based on exploratory data analysis and K-means clustering

- Findings gives a clear recommendation to stakeholders

- Limitations: Streets has not been suggested, household income data were not considered

- Outlook: Analyzing with more data and accurate data

THANK
YOU