

Final Capstone Project

The Restaurant Battle of Neighborhoods in Berlin

1 - Introduction

Berlin is the capital and largest city of Germany by both area and population. Its population of around 3.8 million inhabitants makes it the most populous city of the European Union, according to population within city limits. One of Germany's 16 constituent states, Berlin is surrounded by the state of Brandenburg, and contiguous with Potsdam, Brandenburg's capital. Berlin's urban area has a population of around 4.5 million and is the second most populous urban area in Germany after the Ruhr. The Berlin-Brandenburg capital region has about six million inhabitants and is Germany's third-largest metropolitan region after the Rhine-Ruhr and Rhine-Main regions.

National and international migration into the city has a long history. Berlin is home to at least 180,000 Turkish and Turkish German residents, making it the largest Turkish community outside of Turkey. In December 2019, there were 777,345 registered residents of foreign nationality and another 542,975 German citizens with a "migration background". Foreign residents of Berlin originate from about 190 different countries. Berlin in 2009 was estimated to have 100,000 to 250,000 unregistered inhabitants. Boroughs of Berlin with a significant number of migrants or foreign born population are Mitte, Neukölln and Friedrichshain-Kreuzberg.

There are more than 20 non-indigenous communities with a population of at least 10,000 people, including Turkish, Polish, Russian, Lebanese, Palestinian, Serbian, Italian, Bosnian, Vietnamese, American, Romanian, Bulgarian, Croatian, Chinese, Austrian, Ukrainian, French, British, Spanish, Israeli, Thai, Iranian, Egyptian and Syrian communities.

Our objective is to make a summary of the type of restaurant that a tourist can find in each neighborhood of Berlin. Where to eat American food, German food,...

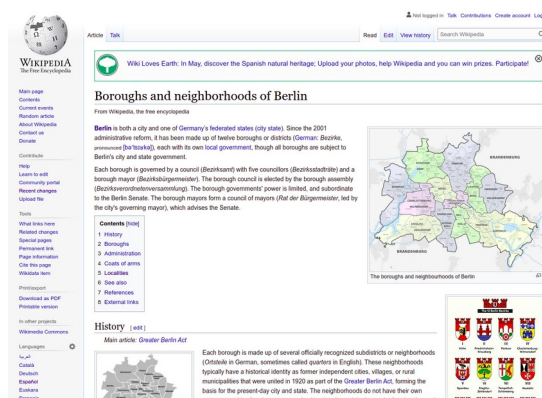
2 - Data Section

I will extract information Foursquare about Berlin sites. Foursquare Labs Inc., commonly known as Foursquare, is an American technology company. The company's location platform is the foundation of several business and consumer products, including the Foursquare City Guide and Foursquare Swarm apps. Here is an example of a asian restaurant in Berlin:

<https://de.foursquare.com/explore?mode=url&near=Berlin%2C%20Deutschland&nearGeoId=72057594040878095&q=Asian%20Restaurant>

I will use the overview of Boroughs of Berlin from Wikipedia:

https://en.wikipedia.org/wiki/Boroughs_and_neighborhoods_of_Berlin



3 - Methodology

In this section which represents the main component of the report where I discuss and describe any exploratory data analysis that I did, any inferential statistical testing that I performed, if any, and what machine learning were used and why.

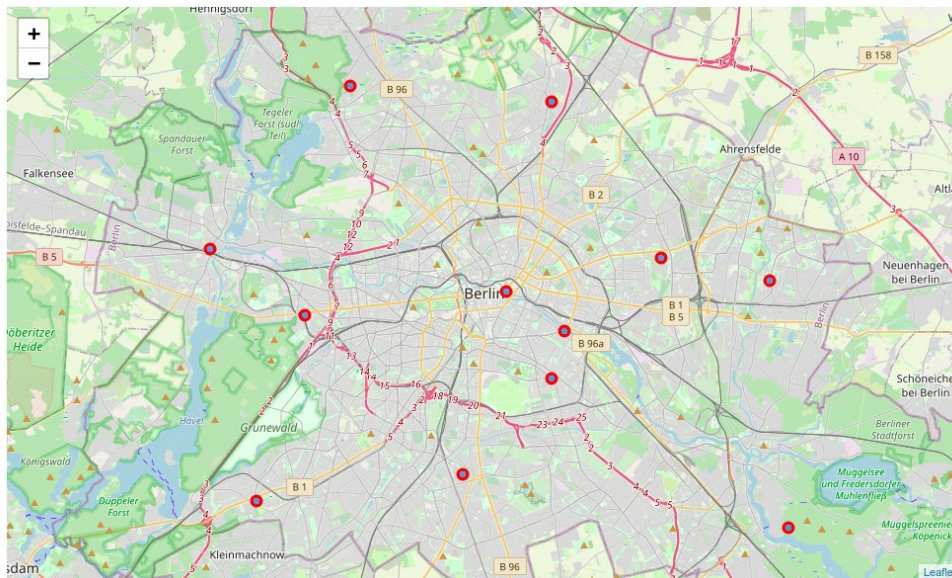
- Firstly, I scraped data from Wikipedia to create a dataframe with the Berlin Boroughs: https://en.wikipedia.org/wiki/Boroughs_and_neighborhoods_of_Berlin. I have used the pandas read function. I had to clean the resulting data frame in terms of unnecessary information or data that could not be handled in a data frame, such as I have changed the name of Lichtenberg's Borough because there is another village with the same name and it could be confusing in Foursquare. I have extracted the *borough*, the *population*, the *area* (in Km²) and the population *density* (people per Km²). Here the data frame:

	Borough	Population	Area	Density
0	Charlottenburg-Wilmersdorf	319,628	64.72	4,878
1	Friedrichshain-Kreuzberg	268,225	20.16	13,187
2	Lichtenberg Berlin	259,881	52.29	4,952
3	Marzahn-Hellersdorf	248,264	61.74	4,046
4	Mitte	332,919	39.47	8,272
5	Neukölln	310,283	44.93	6,804
6	Pankow	366,441	103.01	3,476
7	Reinickendorf	240,454	89.46	2,712
8	Spandau	223,962	91.91	2,441
9	Steglitz-Zehlendorf	293,989	102.50	2,818
10	Tempelhof-Schöneberg	335,060	53.09	6,256
11	Treptow-Köpenick	241,335	168.42	1,406

- I have used the *nominatim* function to add *geospatial* data. You can see the latitude and the longitude on the right side of the following data table:

	Borough	Population	Area	Density	Latitude	Longitude
0	Charlottenburg-Wilmersdorf	319,628	64.72	4,878	52.507856	13.263952
1	Friedrichshain-Kreuzberg	268,225	20.16	13,187	52.501115	13.444285
2	Lichtenberg Berlin	259,881	52.29	4,952	52.532161	13.511893
3	Marzahn-Hellersdorf	248,264	61.74	4,046	52.522523	13.587663
4	Mitte	332,919	39.47	8,272	52.517885	13.404060
5	Neukölln	310,283	44.93	6,804	52.481150	13.435350
6	Pankow	366,441	103.01	3,476	52.597917	13.435316
7	Reinickendorf	240,454	89.46	2,712	52.604763	13.295287
8	Spandau	223,962	91.91	2,441	52.535788	13.197792
9	Steglitz-Zehlendorf	293,989	102.50	2,818	52.429205	13.229974
10	Tempelhof-Schöneberg	335,060	53.09	6,256	52.440603	13.373703
11	Treptow-Köpenick	241,335	168.42	1,406	52.417893	13.600185

- Using the *folium* and the data frame, I have created a map with 12 boroughs on it. You can see the map here below. Each point marks the center of each Berlin borough:

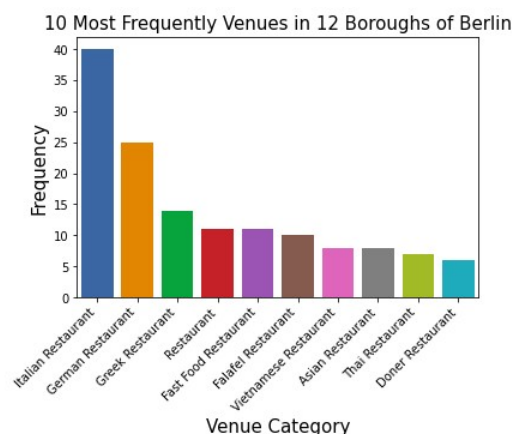


- Now I used foursquare data. Firstly, I did a quick view of the *Mitte* borough data and checked that the results match. We can see some venues in *Mitte* within a radius of 3500 meters:

	name	categories	lat	lng
0	Designpanoptikum - surreales Museum für indust...	Museum	52.516941	13.406072
1	Bronzestatue "Heiliger St. Georg im Kampf mit ...	Outdoor Sculpture	52.516290	13.405558
2	Kuppelgang Berliner Dom	Scenic Lookout	52.518966	13.400981
3	Nikolaiviertel	Neighborhood	52.516782	13.406453
4	Tigertörtchen	Cupcake Shop	52.517150	13.407926
5	Radisson Blu	Hotel	52.519561	13.402857
6	Lustgarten	Garden	52.518469	13.399454
7	Julchen Hoppe	German Restaurant	52.517397	13.404751
8	Nikolaikirchplatz	Plaza	52.516700	13.406839
9	Bonne Vie Café	Café	52.517287	13.407735
10	Altes Museum	History Museum	52.519537	13.398803
11	Buchhandlung Walther König	Bookstore	52.521301	13.400758
12	Neptunbrunnen	Fountain	52.519539	13.406925
13	The Greens	Café	52.515485	13.408987
14	Museum Ephraim-Palais	Museum	52.515856	13.407191
15	LUSH	Cosmetics Shop	52.519844	13.410409
16	Mutter Henne	German Restaurant	52.517614	13.405078

- Then, I retrieved the foursquare data for all venues on foursquare with a distance of less than 3500 meters from each center of each borough. The result was a list of 1081 venues all over Berlin. Out of these 1081 venues, 221 were restaurants. These 221 restaurants come from 34 unique restaurant categories, such as Spanish, African or German food.

- Late I plotted a bar chart with the frequency of the 10 most frequently restaurants in the city. We can see that Italian, German and Greek restaurants are the most frequently occurring restaurants in Berlin:



- Now we need to find clusters of restaurant types in the different boroughs, I had transformed the data frame with the restaurant venues, associated to boroughs, by one-hot encoding. The next step was to group the data by frequency of each category of restaurants in each borough:

	Neighborhood	African Restaurant	Argentinian Restaurant	Asian Restaurant	Caucasian Restaurant	Chinese Restaurant	Doner Restaurant	Dumpling Restaurant	Eastern European Restaurant	Falafel Restaurant	...	Ramen Restaurant	Restaurant
0	Charlottenburg-Wilmersdorf	0.000000	0.038462	0.076923	0.000000	0.038462	0.000000	0.000000	0.000000	0.038462	...	0.000000	0.000000
1	Friedrichshain-Kreuzberg	0.047619	0.000000	0.000000	0.000000	0.000000	0.000000	0.047619	0.000000	0.285714	...	0.000000	0.000000
2	Lichtenberg Berlin	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000
3	Marzahn-Hellersdorf	0.000000	0.000000	0.090909	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.090909
4	Mitte	0.000000	0.000000	0.000000	0.166667	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.166667	0.000000
5	Neukölln	0.095238	0.000000	0.000000	0.000000	0.000000	0.000000	0.095238	0.000000	0.095238	...	0.047619	0.095238
6	Pankow	0.000000	0.000000	0.083333	0.000000	0.083333	0.083333	0.000000	0.000000	0.000000	...	0.000000	0.083333
7	Reinickendorf	0.000000	0.060606	0.000000	0.000000	0.030303	0.000000	0.000000	0.060606	0.030303	...	0.000000	0.090909
8	Spandau	0.000000	0.083333	0.000000	0.000000	0.041667	0.041667	0.000000	0.000000	0.000000	...	0.000000	0.083333
9	Steglitz-Zehlendorf	0.000000	0.000000	0.086957	0.000000	0.000000	0.086957	0.000000	0.000000	0.000000	...	0.000000	0.043478
10	Tempelhof-Schöneberg	0.000000	0.000000	0.095238	0.000000	0.095238	0.095238	0.000000	0.047619	0.000000	...	0.000000	0.047619
11	Treptow-Köpenick	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000

- I created a data frame in which you can see the most common restaurant venue types per borough:

	Neighborhood	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	Charlottenburg-Wilmersdorf	Italian Restaurant	German Restaurant	Vietnamese Restaurant	Asian Restaurant	Greek Restaurant	Indian Restaurant	Mediterranean Restaurant	Falafel Restaurant	Argentinian Restaurant	French Restaurant
1	Friedrichshain-Kreuzberg	Falafel Restaurant	Middle Eastern Restaurant	Thai Restaurant	African Restaurant	Mediterranean Restaurant	Lebanese Restaurant	Italian Restaurant	Spanish Restaurant	German Restaurant	French Restaurant
2	Lichtenberg Berlin	Vietnamese Restaurant	Italian Restaurant	Greek Restaurant	German Restaurant	Syrian Restaurant	Indian Restaurant	Ramen Restaurant	Middle Eastern Restaurant	Moroccan Restaurant	New American Restaurant
3	Marzahn-Hellersdorf	Italian Restaurant	Fast Food Restaurant	Greek Restaurant	Restaurant	Mexican Restaurant	Asian Restaurant	Thai Restaurant	Syrian Restaurant	Sushi Restaurant	Spanish Restaurant
4	Mitte	Seafood Restaurant	Vegetarian / Vegan Restaurant	Caucasian Restaurant	Italian Restaurant	Middle Eastern Restaurant	Ramen Restaurant	Restaurant	Mexican Restaurant	Moroccan Restaurant	New American Restaurant
5	Neukölln	African Restaurant	Falafel Restaurant	Vegetarian / Vegan Restaurant	Turkish Restaurant	Spanish Restaurant	Restaurant	Korean Restaurant	Dumpling Restaurant	Vietnamese Restaurant	Sushi Restaurant
6	Pankow	Greek Restaurant	Italian Restaurant	Restaurant	Asian Restaurant	Chinese Restaurant	Doner Restaurant	Thai Restaurant	Mexican Restaurant	German Restaurant	African Restaurant
7	Reinickendorf	Italian Restaurant	German Restaurant	Restaurant	Indian Restaurant	Seafood Restaurant	Greek Restaurant	Eastern European Restaurant	Argentinian Restaurant	Sushi Restaurant	New American Restaurant
8	Spandau	Italian Restaurant	German Restaurant	Fast Food Restaurant	Argentinian Restaurant	Turkish Restaurant	Restaurant	Vietnamese Restaurant	Halal Restaurant	Greek Restaurant	Mexican Restaurant
9	Steglitz-Zehlendorf	Italian Restaurant	German Restaurant	Asian Restaurant	Doner Restaurant	French Restaurant	Fast Food Restaurant	Restaurant	Greek Restaurant	Mexican Restaurant	Sushi Restaurant
10	Tempelhof-Schöneberg	Italian Restaurant	Asian Restaurant	Thai Restaurant	Chinese Restaurant	Doner Restaurant	Fast Food Restaurant	Korean Restaurant	Greek Restaurant	Restaurant	Middle Eastern Restaurant
11	Treptow-Köpenick	German Restaurant	Fast Food Restaurant	Sushi Restaurant	Greek Restaurant	Seafood Restaurant	Restaurant	Middle Eastern Restaurant	Moroccan Restaurant	New American Restaurant	Ramen Restaurant

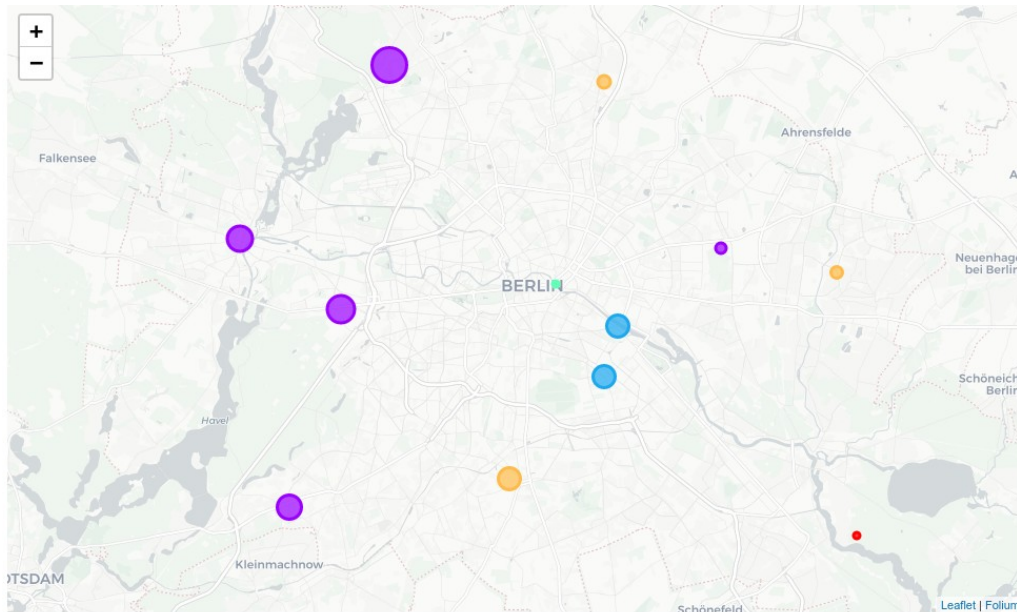
- I could finally run a k-means clustering algorithm. We could have used the elbow method to systematically define the k value, but in this case I have chosen a k = 5.

4 – Results

- Below we can see the table with the boroughs and their most common venues, and they now have been assigned different cluster labels from 0 to 4:

	N° Cluster	Neighborhood	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	1	Charlottenburg-Wilmersdorf	Italian Restaurant	German Restaurant	Vietnamese Restaurant	Asian Restaurant	Greek Restaurant	Indian Restaurant	Mediterranean Restaurant	Falafel Restaurant	Argentinian Restaurant	French Restaurant
1	2	Friedrichshain-Kreuzberg	Falafel Restaurant	Middle Eastern Restaurant	Thai Restaurant	African Restaurant	Mediterranean Restaurant	Lebanese Restaurant	Italian Restaurant	Spanish Restaurant	German Restaurant	French Restaurant
2	1	Lichtenberg Berlin	Vietnamese Restaurant	Italian Restaurant	Greek Restaurant	German Restaurant	Syrian Restaurant	Indian Restaurant	Ramen Restaurant	Middle Eastern Restaurant	Moroccan Restaurant	New American Restaurant
3	4	Marzahn-Hellersdorf	Italian Restaurant	Fast Food Restaurant	Greek Restaurant	Restaurant	Mexican Restaurant	Asian Restaurant	Thai Restaurant	Syrian Restaurant	Sushi Restaurant	Spanish Restaurant
4	3	Mitte	Seafood Restaurant	Vegetarian / Vegan Restaurant	Caucasian Restaurant	Italian Restaurant	Middle Eastern Restaurant	Ramen Restaurant	Restaurant	Mexican Restaurant	Moroccan Restaurant	New American Restaurant

- We can use this cluster labels to show the boroughs marked with a cluster-specific color on a map:



- We are going to analyze the different clusters we have (0-4):

- **Cluster 0:** It could be called the **German cluster**.

	Population	Longitude	N° Cluster	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
11	241,335	13.600185	0	German Restaurant	Fast Food Restaurant	Sushi Restaurant	Greek Restaurant	Seafood Restaurant	Restaurant	Middle Eastern Restaurant	Moroccan Restaurant	New American Restaurant	Ramen Restaurant

- **Cluster 1:** It could be called the **Italian and Vietnamese cluster**.

	Population	Longitude	N° Cluster	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	319,628	13.263952	1	Italian Restaurant	German Restaurant	Vietnamese Restaurant	Asian Restaurant	Greek Restaurant	Indian Restaurant	Mediterranean Restaurant	Falafel Restaurant	Argentinian Restaurant	French Restaurant
2	259,881	13.511893	1	Vietnamese Restaurant	Italian Restaurant	Greek Restaurant	German Restaurant	Syrian Restaurant	Indian Restaurant	Ramen Restaurant	Middle Eastern Restaurant	Moroccan Restaurant	New American Restaurant
7	240,454	13.295287	1	Italian Restaurant	German Restaurant	Restaurant	Indian Restaurant	Seafood Restaurant	Greek Restaurant	Eastern European Restaurant	Argentinian Restaurant	Sushi Restaurant	New American Restaurant
8	223,962	13.197792	1	Italian Restaurant	German Restaurant	Fast Food Restaurant	Argentinian Restaurant	Turkish Restaurant	Restaurant	Vietnamese Restaurant	Halal Restaurant	Greek Restaurant	Mexican Restaurant
9	293,989	13.229974	1	Italian Restaurant	German Restaurant	Asian Restaurant	Doner Restaurant	French Restaurant	Fast Food Restaurant	Restaurant	Greek Restaurant	Mexican Restaurant	Sushi Restaurant

- **Cluster 2:** It could be called the **Middle Eastern food and African cluster**.

	Population	Longitude	N° Cluster	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	268,225	13.444285	2	Falafel Restaurant	Middle Eastern Restaurant	Thai Restaurant	African Restaurant	Mediterranean Restaurant	Lebanese Restaurant	Italian Restaurant	Spanish Restaurant	German Restaurant	French Restaurant
5	310,283	13.435350	2	African Restaurant	Falafel Restaurant	Vegetarian / Vegan Restaurant	Turkish Restaurant	Spanish Restaurant	Restaurant	Korean Restaurant	Dumpling Restaurant	Vietnamese Restaurant	Sushi Restaurant

- **Cluster 3:** It could be called the **Seafood cluster**.

	Population	Longitude	Nº Cluster	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
4	332,919	13.40406	3	Seafood Restaurant	Vegetarian / Vegan Restaurant	Caucasian Restaurant	Italian Restaurant	Middle Eastern Restaurant	Ramen Restaurant	Restaurant	Mexican Restaurant	Moroccan Restaurant	New American Restaurant

- **Cluster 4:** It could be called the **Italian and Greek cluster**.

	Population	Longitude	Nº Cluster	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
3	248,264	13.587663	4	Italian Restaurant	Fast Food Restaurant	Greek Restaurant	Restaurant	Mexican Restaurant	Asian Restaurant	Thai Restaurant	Syrian Restaurant	Sushi Restaurant	Spanish Restaurant
6	366,441	13.435316	4	Greek Restaurant	Italian Restaurant	Restaurant	Asian Restaurant	Chinese Restaurant	Doner Restaurant	Thai Restaurant	Mexican Restaurant	German Restaurant	African Restaurant
10	335,060	13.373703	4	Italian Restaurant	Asian Restaurant	Thai Restaurant	Chinese Restaurant	Doner Restaurant	Fast Food Restaurant	Korean Restaurant	Greek Restaurant	Restaurant	Middle Eastern Restaurant

5 – Discussion

As we have indicated above, we could also have used the elbow method to obtain the ideal number of clusters to obtain better precision. The K-means algorithm allows us to carry out these types of interesting researchs that can have multiple uses.

6 - Conclusion

We have obtained reasonable results, in this way anyone who wants to visit or lives in Berlin can know what kind of restaurants there are in the area, obviously the precision could be improved if instead of using the boroughs, we had done it with the neighborhoods that make up these boroughs.