IBM Data Science Capstone Project

Picking right location for a new restaurant in berlin

Senura Fernando

November 2020

1. Introduction

The IBM Data Science Professional certificate course on Coursera concludes with a Capstone Project. This project is about using data science toolset on a real-life problem and demonstrating the creation of value by applying the learned skills. This report presents the capstone project which is done for picking right location for a new restaurant in Berlin.

2. Problem Definition

a. The Problem

Berlin is the capital and largest city of Germany by both area and population. In an competitive field, Berlin is making a strong claim to be the world's premier city. The German capital can lay claim to reasonable rent, an English-speaking population, thriving businesses, great food and nightlife, and a burgeoning startup scene.

In here, business problem is to picking up a best location to open a new restaurant in Berlin. Taking the population, price level at which the restaurant will operate. The intent is to find an optimal location in an area which is easily accessible for tourists and for wealthier local citizens as well.

b. Assumptions and business logic

The assumption behind the analysis is that , unsupervised machine learning can be used to create clusters of neighborhoods of Berlin that will give with a list of areas for consideration for the restaurant. The intent is that restaurant to be situated close to one of the gastronomical centers and touristic hotspots.

c. Audience

This analysis could be useful for group of market players and business people who have idea to open new restaurant in berlin and also tourists to take idea about restaurants around Berlin

3. Data

To perform this analysis , following data required:

1. List of the boroughs and neighborhoods of Berlin
2. Geo-coordinates of the boroughs in Berlin
3. Top venues of boroughs

List of boroughs will be obtained from Wikipedia.

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

Geo-coordinates of boroughs will be obtained with the help of the geocoder tool in the notebook.

Top venues data will be obtained from Foursquare through an API

4 . Methodology

a) Use of data and a high – level roadmap

After exploring the data , K – Means clustering will be applied for creating clusters of boroughs. Silhouette score can be used for selecting the optimal number of clusters.

b) Analysis

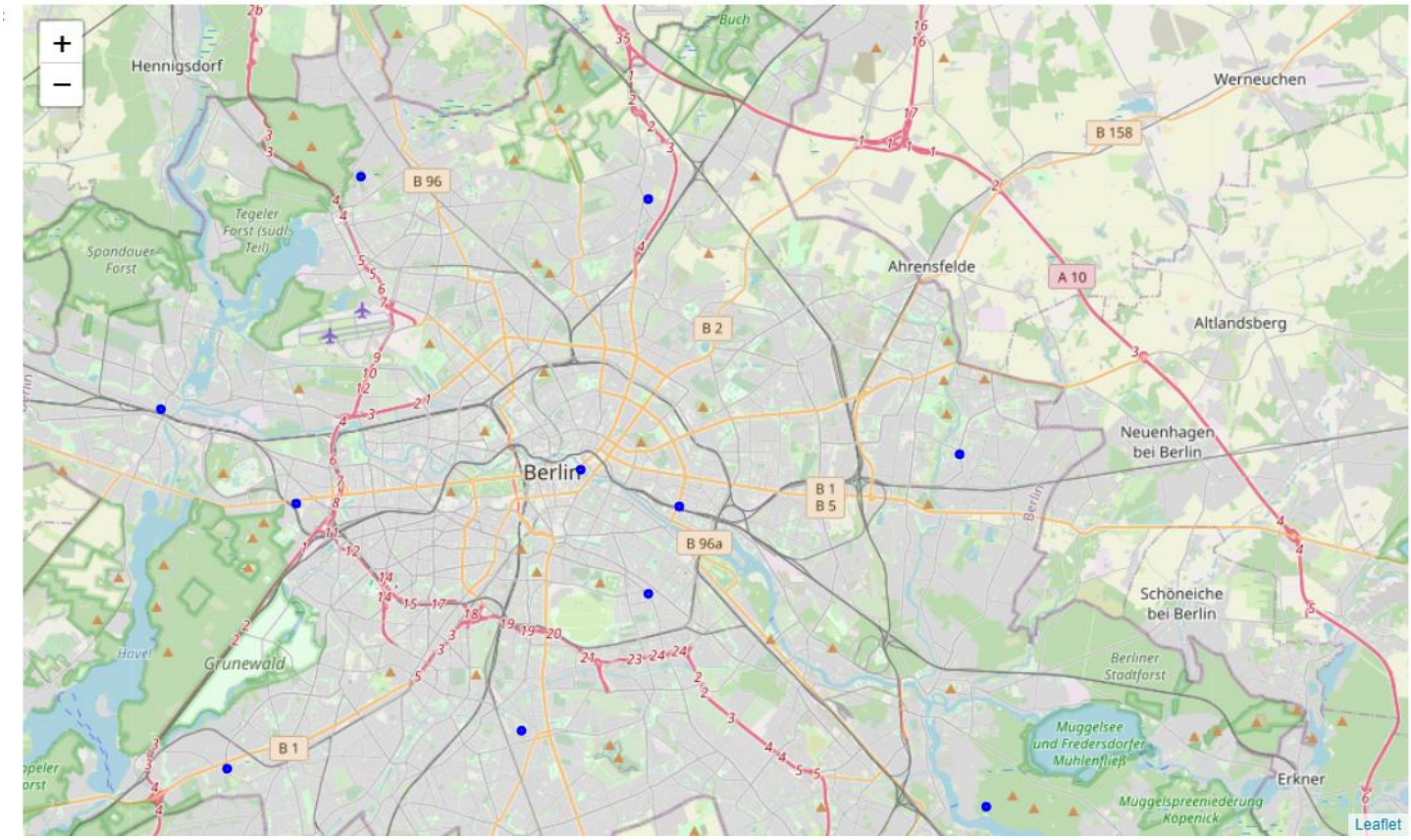
- **Data Preparation and exploration**

When preparing data , initially list of boroughs of berlin are created and add the geo-coordinates of each borough to this table. This is done by first importing list of boroughs and then using the list and geocode python library , latitude and longitude coordinates are added to each borough. After the resulting data frame is created as follows.

	Borough	Latitude	Longitude
0	Mitte	52.517690	13.402376
1	Friedrichshain-Kreuzberg	52.506862	13.450642
2	Pankow	52.597917	13.435316
3	Charlottenburg-Wilmersdorf	52.507856	13.263952
4	Spandau	52.535788	13.197792
5	Steglitz-Zehlendorf	52.429205	13.229974
6	Tempelhof-Schöneberg	52.440603	13.373703
7	Neukölln	52.481150	13.435350
8	Treptow-Köpenick	52.417893	13.600185
9	Marzahn-Hellersdorf	52.522523	13.587663
10	Lichtenberg	48.921296	7.481227
11	Reinickendorf	52.604763	13.295287

There are 12 boroughs in the Berlin and this table shows the corresponding longitude and latitude values for each borough.

In the next step , visual representation is used to show how these boroughs are situated in Berlin using folium library.



In the next step of analysis , the boroughs were explored in greater detail. It means venues were collected for each borough via Foursquare API. The data from Foursquare is received in JSON format. After arranging the data , 100 venues for each borough were captured. Venues are collected within a radius of 1000 meters from the point of borough coordinates. The collected and arranged data looks like this.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Mitte	52.51769	13.402376	Lustgarten	52.518469	13.399454	Garden
1	Mitte	52.51769	13.402376	Kuppelumgang Berliner Dom	52.518966	13.400981	Scenic Lookout
2	Mitte	52.51769	13.402376	Radisson Blu	52.519561	13.402857	Hotel
3	Mitte	52.51769	13.402376	Bronzestatue "Heiliger St. Georg im Kampf mit ...	52.516290	13.405558	Outdoor Sculpture
4	Mitte	52.51769	13.402376	Designpanoptikum - surreales Museum für indust...	52.516941	13.406072	Museum

Following table gives the summary for each neighborhood.

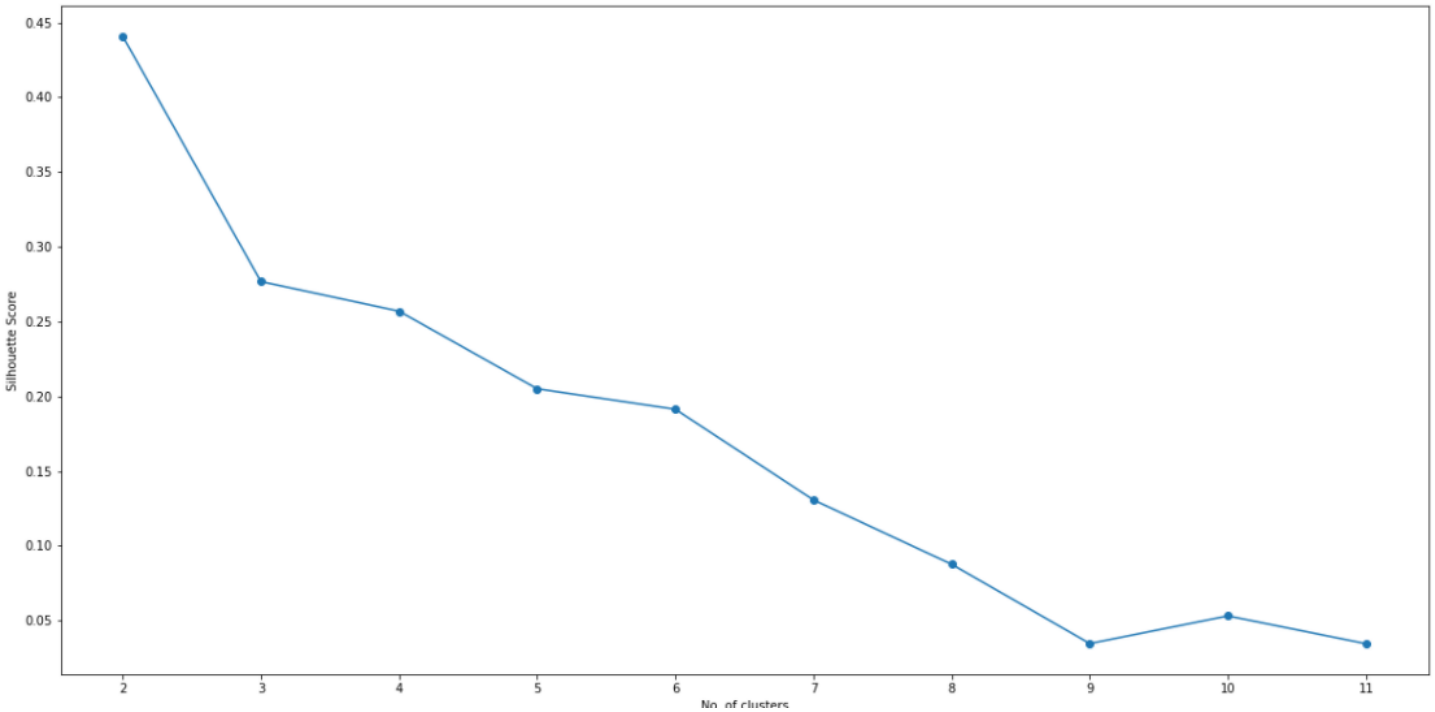
Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Charlottenburg-Wilmersdorf	45	45	45	45	45	45
Friedrichshain-Kreuzberg	100	100	100	100	100	100
Lichtenberg	2	2	2	2	2	2
Marzahn-Hellersdorf	15	15	15	15	15	15
Mitte	100	100	100	100	100	100
Neukölln	100	100	100	100	100	100
Pankow	8	8	8	8	8	8
Reinickendorf	7	7	7	7	7	7
Spandau	64	64	64	64	64	64
Steglitz-Zehlendorf	24	24	24	24	24	24
Tempelhof-Schöneberg	27	27	27	27	27	27
Treptow-Köpenick	7	7	7	7	7	7

For analysis the boroughs , only venue categories are focused. Following table shows the results of top ten most common venues for each borough after manipulations with the dataset.

	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	Café	Italian Restaurant	Hotel	German Restaurant	Bus Stop	Plaza	Pizza Place	Drugstore	Rooftop Bar	Bowling Alley
1	Friedrichshain-Kreuzberg	Vegetarian / Vegan Restaurant	Coffee Shop	Bar	Nightclub	Middle Eastern Restaurant	Falafel Restaurant	Café	Music Venue	Hotel	Thai Restaurant
2	Lichtenberg	Hostel	Historic Site	Dog Run	Falafel Restaurant	Exhibit	Escape Room	Dumpling Restaurant	Drugstore	Donut Shop	Doner Restaurant
3	Marzahn-Hellersdorf	Supermarket	Drugstore	Metro Station	Bowling Alley	Shopping Mall	Lake	Trail	Donut Shop	Falafel Restaurant	Exhibit
4	Mitte	Hotel	Art Gallery	History Museum	Plaza	Coffee Shop	Museum	Restaurant	Bookstore	Ice Cream Shop	Café

c) Clustering

Now clustering is performed. Here K - Means algorithm is mainly used which is unsupervised learning method. In order to select best number of cluster we can use the graph which shows the silhouette scores with different number of clusters.

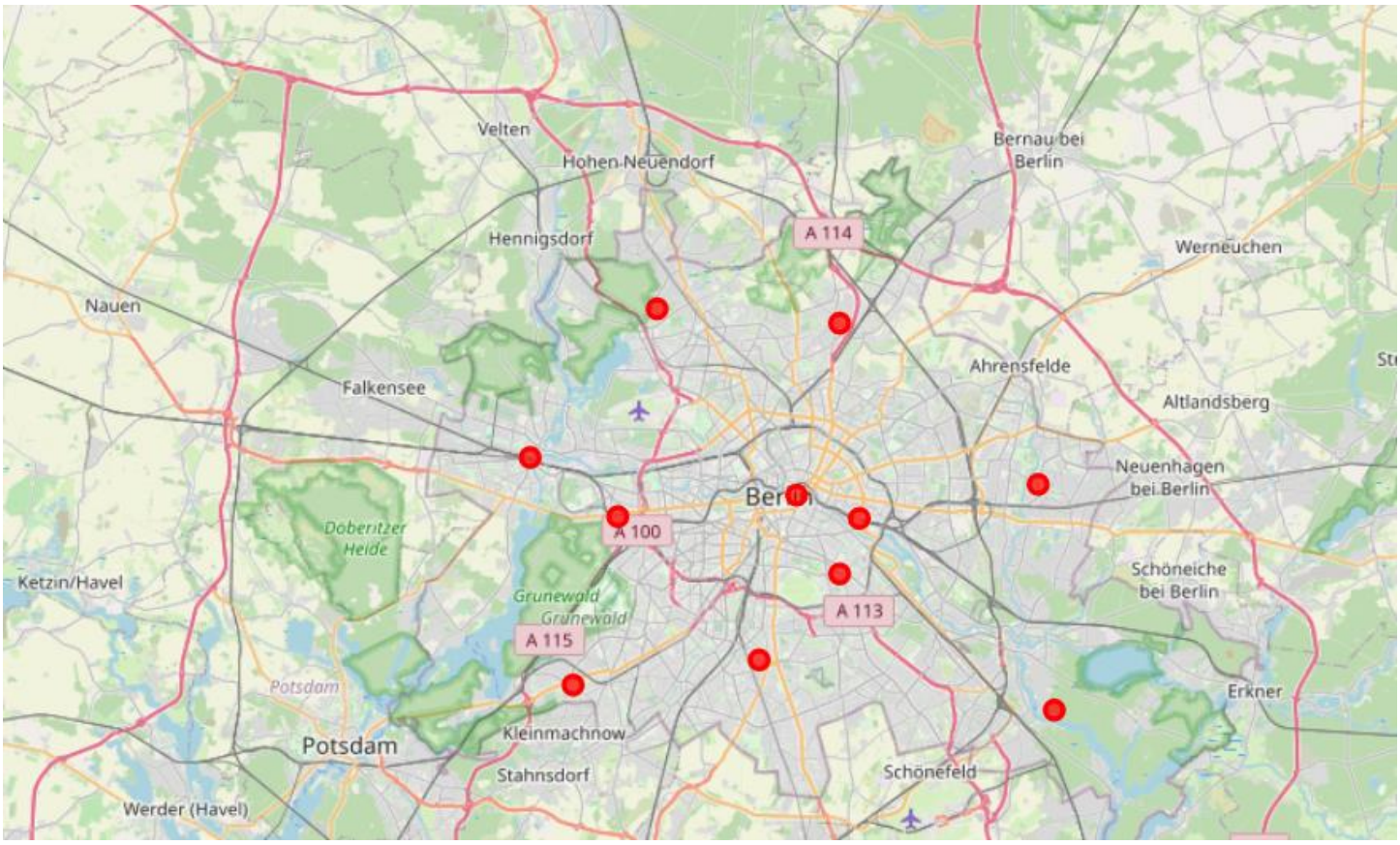


According to the graph , optimal number of cluster is 2 which gives the highest score.

Then K – Means clustering is performed in order to create 2 clusters. Following table shows the venues and corresponding cluster.

	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
0	Mitte	52.517690	13.402376	0	Hotel	Art Gallery	History Museum	Plaza	Coffee Shop	Museum	Restaurant	Bookstore	Ice Cream Shop
1	Friedrichshain-Kreuzberg	52.506862	13.450642	0	Vegetarian / Vegan Restaurant	Coffee Shop	Bar	Nightclub	Middle Eastern Restaurant	Falafel Restaurant	Café	Music Venue	Hotel
2	Pankow	52.597917	13.435316	0	Tram Station	Supermarket	Light Rail Station	Lake	Asian Restaurant	Bus Stop	Donut Shop	Falafel Restaurant	Exhibit
3	Charlottenburg-Wilmersdorf	52.507856	13.263952	0	Café	Italian Restaurant	Hotel	German Restaurant	Bus Stop	Plaza	Pizza Place	Drugstore	Rooftop Bar
4	Spandau	52.535788	13.197792	0	Bakery	Supermarket	Drugstore	Clothing Store	Hotel	Ice Cream Shop	Fast Food Restaurant	Italian Restaurant	Mobile Phone Shop
5	Steglitz-Zehlendorf	52.429205	13.229974	0	Italian Restaurant	Bus Stop	History Museum	Bar	Park	Plaza	Chocolate Shop	Pub	Café
6	Tempelhof-Schöneberg	52.440603	13.373703	0	Supermarket	Bakery	Steakhouse	Industrial Estate	Drugstore	Brewery	Farmers Market	Fast Food Restaurant	Restaurant
7	Neukölln	52.481150	13.435350	0	Coffee Shop	Bar	Café	Cocktail Bar	Bistro	German Restaurant	Italian Restaurant	Turkish Restaurant	Indie Movie Theater
8	Treptow-Köpenick	52.417893	13.600185	0	Gastropub	Plaza	Tram Station	Lake	River	Beach	Dive Bar	Escape Room	Dumpling Restaurant
9	Marzahn-Hellersdorf	52.522523	13.587663	0	Supermarket	Drugstore	Metro Station	Bowling Alley	Shopping Mall	Lake	Trail	Donut Shop	Falafel Restaurant
10	Lichtenberg	48.921296	7.481227	1	Hostel	Historic Site	Dog Run	Falafel Restaurant	Exhibit	Escape Room	Dumpling Restaurant	Drugstore	Donut Shop
11	Reinickendorf	52.604763	13.295287	0	Hostel	Italian Restaurant	Nature Preserve	Gastropub	Bakery	Liquor Store	Ice Cream Shop	Gym / Fitness Center	History Museum

Visualizing the clusters



d) Limitations

This analysis is done in borough level.

5.Results

Understanding the clusters.

Cluster 1

```
[101]: berlin_merged.loc[berlin_merged['Cluster Labels'] == 0, berlin_merged.columns[[0] + list(range(5, berlin_merged.shape[1]))]]
```

[101]:

	Neighborhood	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	Mitte	Art Gallery	History Museum	Plaza	Coffee Shop	Museum	Restaurant	Bookstore	Ice Cream Shop	Café
1	Friedrichshain-Kreuzberg	Coffee Shop	Bar	Nightclub	Middle Eastern Restaurant	Falafel Restaurant	Café	Music Venue	Hotel	Thai Restaurant
2	Pankow	Supermarket	Light Rail Station	Lake	Asian Restaurant	Bus Stop	Donut Shop	Falafel Restaurant	Exhibit	Escape Room
3	Charlottenburg-Wilmersdorf	Italian Restaurant	Hotel	German Restaurant	Bus Stop	Plaza	Pizza Place	Drugstore	Rooftop Bar	Bowling Alley
4	Spandau	Supermarket	Drugstore	Clothing Store	Hotel	Ice Cream Shop	Fast Food Restaurant	Italian Restaurant	Mobile Phone Shop	Pizza Place
5	Steglitz-Zehlendorf	Bus Stop	History Museum	Bar	Park	Plaza	Chocolate Shop	Pub	Café	Liquor Store
6	Tempelhof-Schöneberg	Bakery	Steakhouse	Industrial Estate	Drugstore	Brewery	Farmers Market	Fast Food Restaurant	Restaurant	Motorcycle Shop
7	Neukölln	Bar	Café	Cocktail Bar	Bistro	German Restaurant	Italian Restaurant	Turkish Restaurant	Indie Movie Theater	Dive Bar
8	Treptow-Köpenick	Plaza	Tram Station	Lake	River	Beach	Dive Bar	Escape Room	Dumpling Restaurant	Drugstore
9	Marzahn-Hellersdorf	Drugstore	Metro Station	Bowling Alley	Shopping Mall	Lake	Trail	Donut Shop	Falafel Restaurant	Exhibit
11	Reinickendorf	Italian Restaurant	Nature Preserve	Gastropub	Bakery	Liquor Store	Ice Cream Shop	Gym / Fitness Center	History Museum	Exhibit

This include lots of related places like Hotel , Asian Restaurant , Bakery

Cluster 2

```
berlin_merged.loc[berlin_merged['Cluster Labels'] == 1, berlin_merged.columns[[0] + list(range(5, berlin_merged.shape[1]))]]
```

	Neighborhood	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
10	Lichtenberg	Historic Site	Dog Run	Falafel Restaurant	Exhibit	Escape Room	Dumpling Restaurant	Drugstore	Donut Shop	Doner Restaurant

6.Discussion

Based on cluster details , we can advise businessman to open new restaurant in boroughs which are related to cluster 0. There are lots of similar places in this cluster and will have a potential location for a new restaurant.

7.Conclusion

This paper discuss the real world business related problem which can be easily solved using location data. The analysis was performed using python and related toolset for data analysis. In order to get the details about location , Foursquare API was used. The output provides better recommendation for the target audience in order to open new restaurant.

8.Reference

Note book of the code:

https://github.com/senura96/Coursera_Capstone/blob/main/IBM%20Coursera%20Capstone%20Final.ipynb