Picking Right Location For a New Restaurant in Berlin

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.

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.

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

METHODOLOGY

- 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)

	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.

- 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
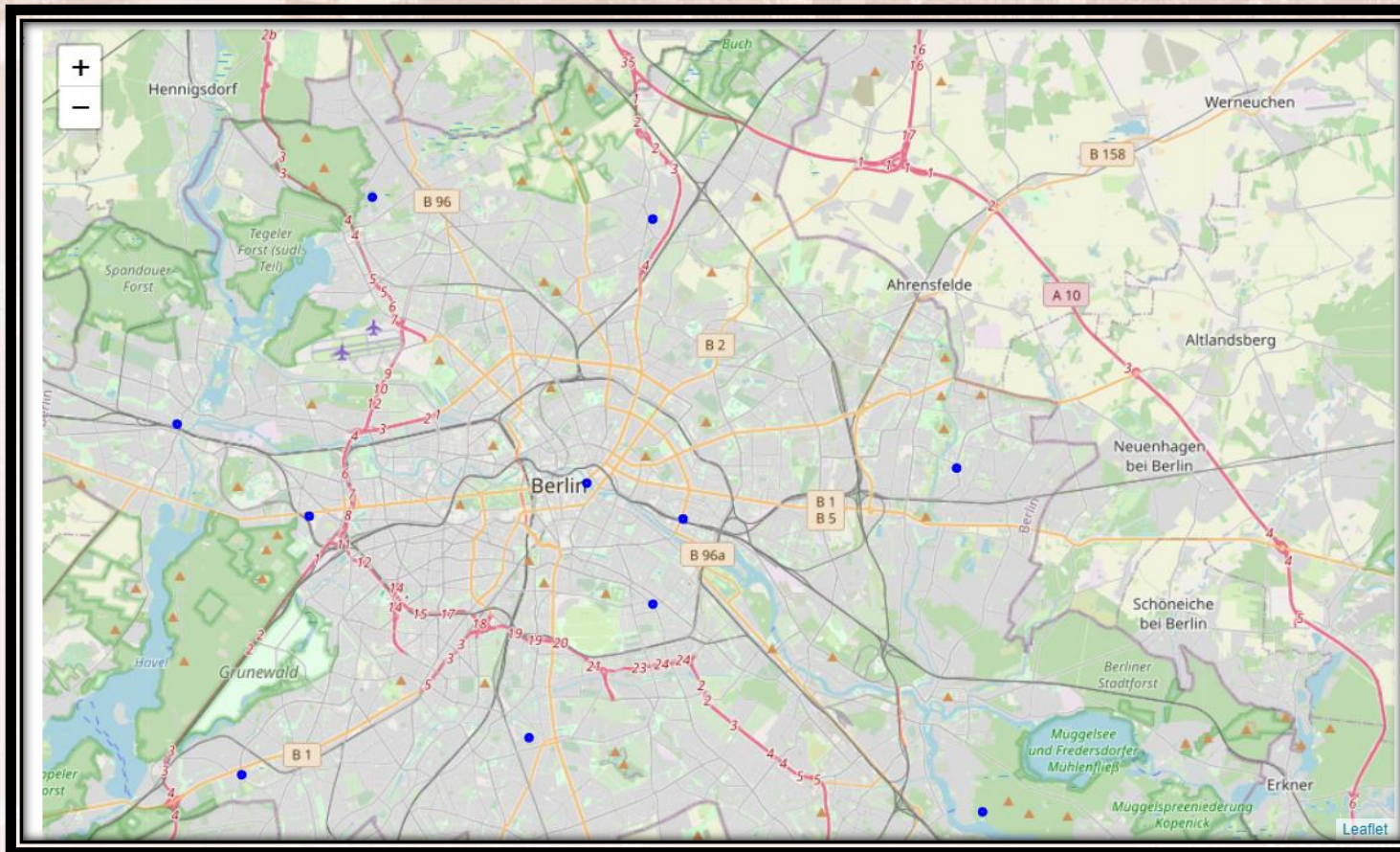
Boroughs of Berlin

- 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.

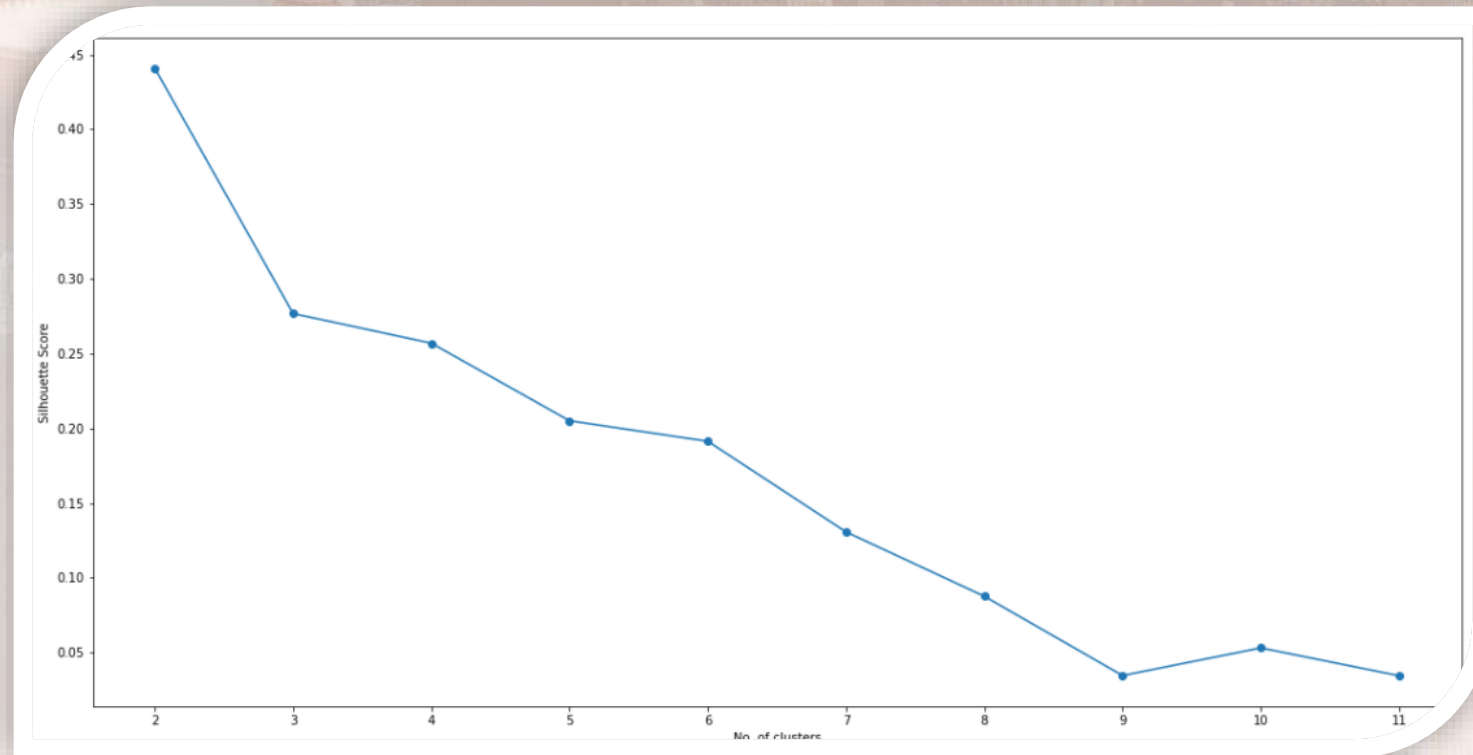|	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Mitte	52.51769	13.402376	Lustgarten	52.518469	13.399454	Garden
1	Mitte	52.51769	13.402376	Kuppelumgang Berliner Dom	52.518966	13.400981	Scenic Lookout
2	Mitte	52.51769	13.402376	Radisson 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

- 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

RESULTS

Cluster 0

```
[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

Cluster 1

```
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

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