IBM DATA SCIENCE CAPSTONE PROJECT

LOCATION SELECTION FOR A NEW RESTAURANT IN BERLIN

LOCATION SELECTION PROBLEM

- Berlin is one of Europe's most cosmopolitan cities. Its best restaurants are spanning the globe in their offering – from Turkish Döner Kebap to Japanese Sushi.

- A businessman who has fast-food restaurants in the US wants to open his new fast-food branch in Berlin, but he is unsure where he should open his new restaurant branch in the town.

- We will help him make a great decision to make his fast-food restaurant one of the tops visited venues in its borough.

DATASETS

We need a couple of Datasets to solve this problem

1. Boroughs and neighborhoods of Berlin

2. Geo-coordinates of boroughs in Berlin

3. Top venues in each borough from the Foursquare API

DATASETS

- Berlin is both a city and one of Germany's federated states. As of 2012, the 12 boroughs are made up of a total of 96 officially recognized localities. The dataset will be scraped from the Wikipedia Page: https://en.wikipedia.org/wiki/Category:Localities_of_Berlin

- Then, Geo-coordinates of Boroughs will be obtained with the Google Maps Geocoding API.

- the Foursquare API to obtain the most visited restaurant categories and the number of restaurants belonging to different types in each borough of Berlin.

METHODOLOGY

- After scraping the data and analyzing it, we will K- Means Clustering algorithm used as an unsupervised machine learning method to cluster boroughs based on the most visited venues.

- K value of the K-Means algorithm will be selected to plot the sum squared error graph and find its elbow point.

- According to clustering results, the most promising clusters for a new fast-food restaurant will be detected,

- and neighborhoods located in those clusters would be considered potential restaurant regions.
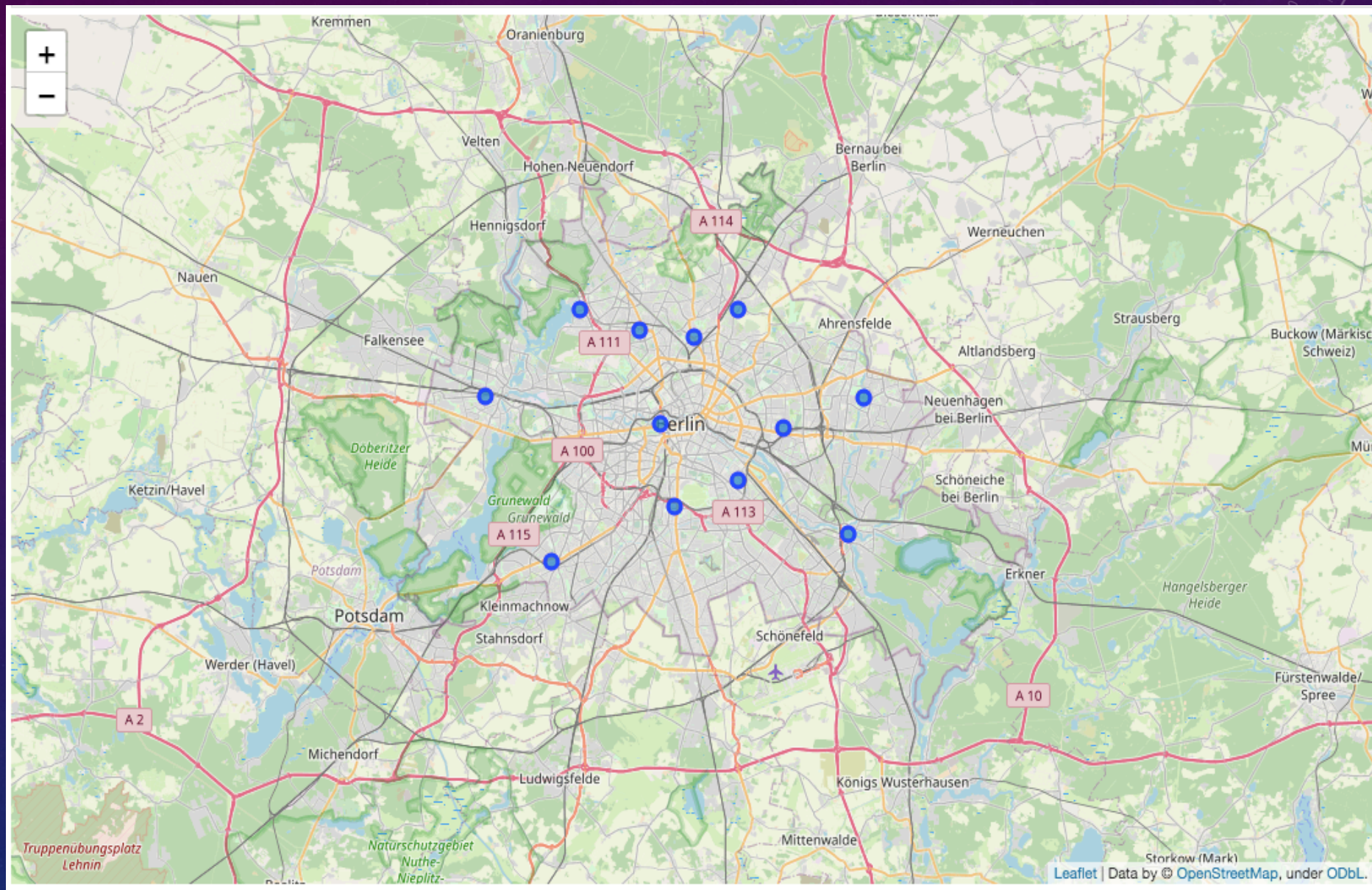
DATAFRAME (RAW AFTER SCRAPING)

	Borough	Neighborhood
0	Charlottenburg-Wilmersdorf	Charlottenburg,Charlottenburg-Nord,Grunewald,H...
1	Friedrichshain-Kreuzberg	Friedrichshain,Kreuzberg
2	Lichtenberg	Alt-Hohenschönhausen,Falkenberg,Fennpfuhl,Frie...
3	Marzahn-Hellersdorf	Biesdorf,Hellersdorf,Kaulsdorf,Mahlsdorf,Marzahn
4	Mitte	Gesundbrunnen,Hansaviertel,Mitte,Moabit,Tierga...
5	Neukölln	Britz,Buckow,Gropiusstadt,Neukölln,Rudow
6	Pankow	Blankenburg,Blankenfelde,Buch,Französisch,Buch...
7	Reinickendorf	Borsigwalde,Frohnau,Heiligensee,Hermsdorf,Konr...
8	Spandau	Falkenhagener,Feld,Gatow,Hakenfelde,Haselhorst...
9	Steglitz-Zehlendorf	Dahlem,Lankwitz,Lichterfelde,Nikolassee,Stegli...
10	Tempelhof-Schöneberg	Friedenau,Lichtenrade,Mariendorf,Marienfelde,S...
11	Treptow-Köpenick	Adlershof,Alt-Treptow,Altglienicke,Baumschulen...

DATAFRAME (AFTER GEOCODING)

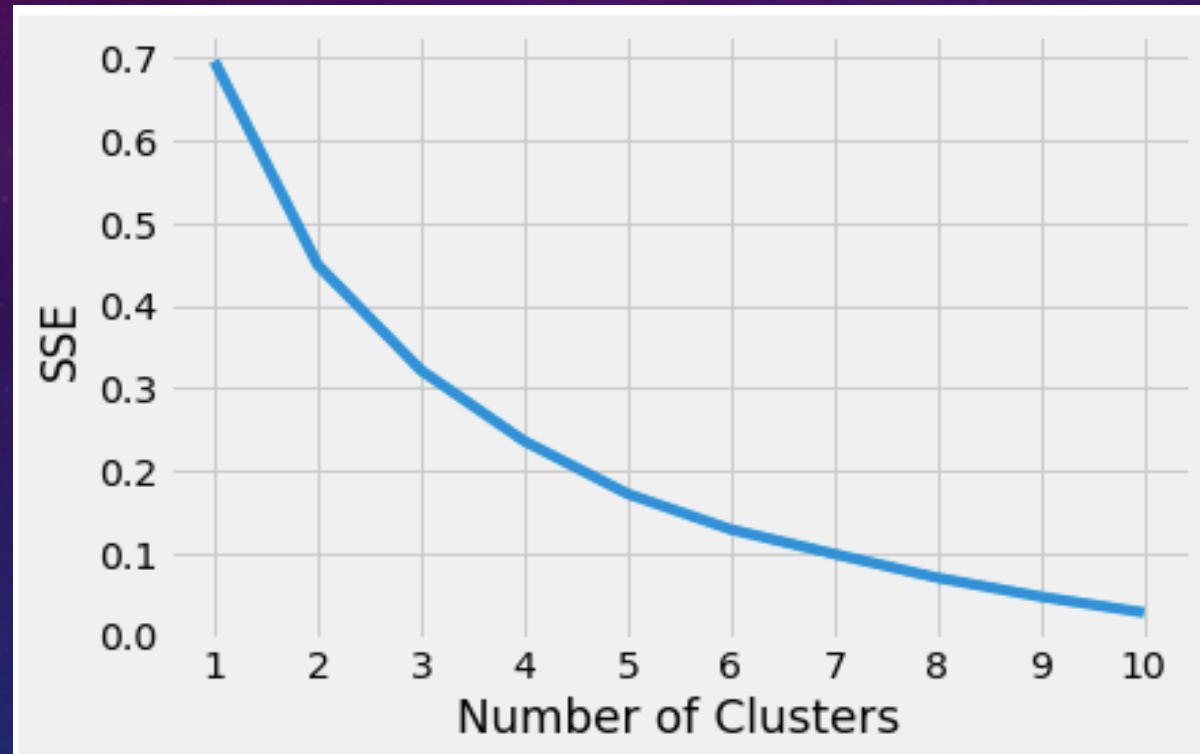|	Borough	Neighborhood	Latitude	Longitude
0	Charlottenburg-Wilmersdorf	Charlottenburg,Charlottenburg-Nord,Grunewald,H...	52.586016	13.283000
1	Friedrichshain-Kreuzberg	Friedrichshain,Kreuzberg	52.586016	13.450000
2	Lichtenberg	Alt-Hohenschönhausen,Falkenberg,Fennpfuhl,Frie...	52.514557	13.498307
3	Marzahn-Hellersdorf	Biesdorf,Hellersdorf,Kaulsdorf,Mahlsdorf,Marzahn	52.533001	13.583000
4	Mitte	Gesundbrunnen,Hansaviertel,Mitte,Moabit,Tierga...	52.516998	13.367000
5	Neukölln	Britz,Buckow,Gropiusstadt,Neukölln,Rudow	52.483002	13.450000
6	Pankow	Blankenburg,Blankenfelde,Buch,Französisch,Buch...	52.569595	13.403235
7	Reinickendorf	Borsigwalde,Frohnau,Heiligensee,Hermsdorf,Konr...	52.574093	13.345394
8	Spandau	Falkenhagener,Feld,Gatow,Hakenfelde,Haselhorst...	52.534073	13.181689
9	Steglitz-Zehlendorf	Dahlem,Lankwitz,Lichterfelde,Nikolassee,Stegli...	52.432999	13.252171
10	Tempelhof-Schöneberg	Friedenau,Lichtenrade,Mariendorf,Marienfelde,S...	52.466999	13.383000
11	Treptow-Köpenick	Adlershof,Alt-Treptow,Altglienicke,Baumschulen...	52.450001	13.567000

FOURSQUARE DATA

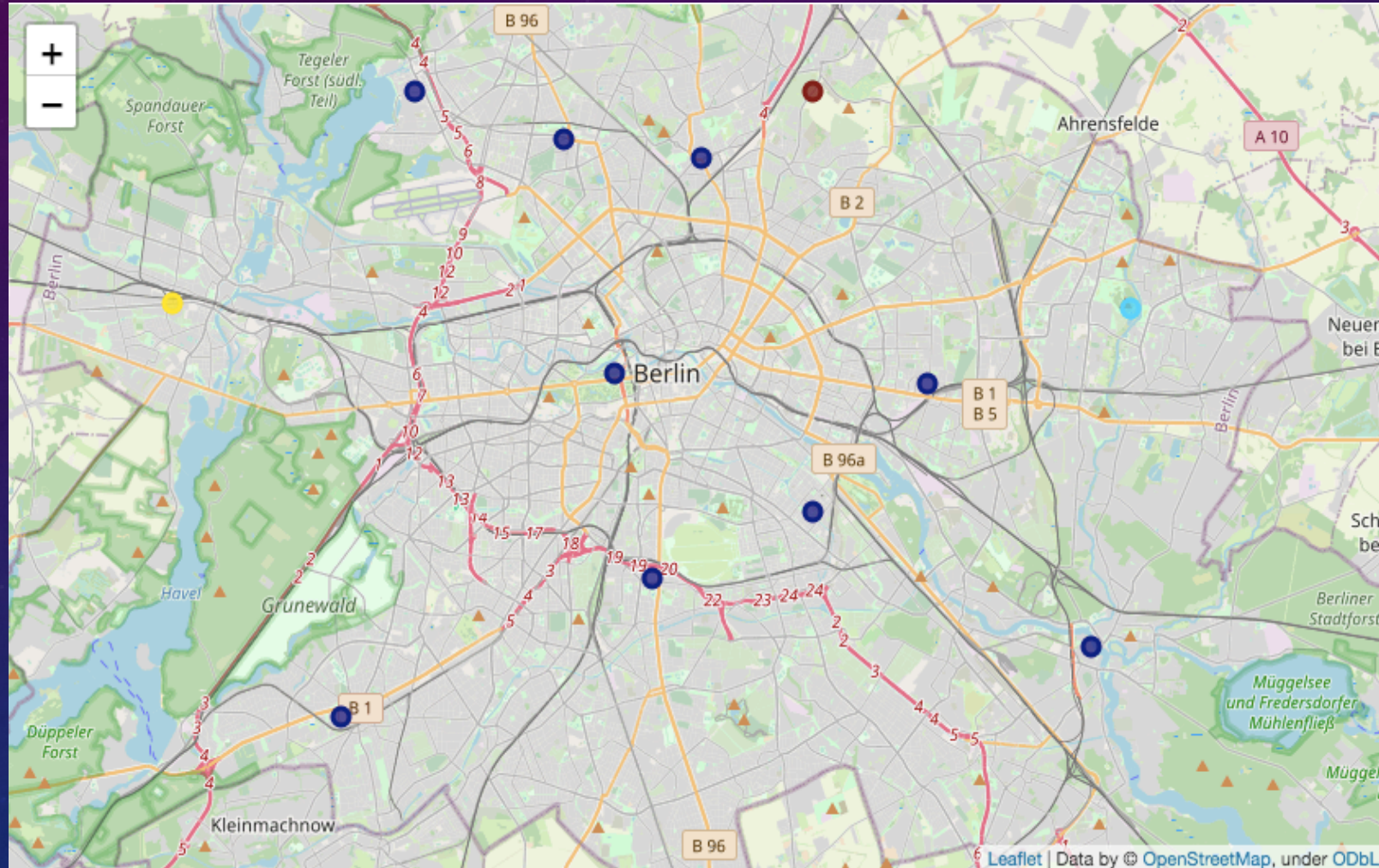|	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Charlottenburg-Wilmersdorf	Café	Clothing Store	Drugstore	Supermarket	Restaurant	Italian Restaurant	Shopping Mall	Ice Cream Shop	Indian Restaurant	Sandwich Place
1	Friedrichshain-Kreuzberg	Bakery	Light Rail Station	Greek Restaurant	Bus Stop	Electronics Store	Café	Farm	Forest	Food & Drink Shop	Flower Shop
2	Lichtenberg	Bakery	Supermarket	Park	Coffee Shop	ATM	Soccer Stadium	Bookstore	Mexican Restaurant	Pharmacy	Breakfast Spot
3	Marzahn-Hellersdorf	Garden	Supermarket	Scenic Lookout	Drugstore	Tea Room	Mountain	Theme Park Ride / Attraction	Yoga Studio	Falafel Restaurant	Food & Drink Shop
4	Mitte	Hotel	Plaza	Hotel Bar	Museum	Spa	Concert Hall	Coffee Shop	Italian Restaurant	Monument / Landmark	Park
5	Neukölln	Café	Bar	Supermarket	Italian Restaurant	Coffee Shop	Park	Vegetarian / Vegan Restaurant	Bistro	German Restaurant	Cocktail Bar
6	Pankow	Café	Supermarket	Bakery	Drugstore	Park	Dance Studio	Organic Grocery	Sushi Restaurant	Burger Joint	Gym / Fitness Center
7	Reinickendorf	Soccer Field	Supermarket	Metro Station	Plaza	Light Rail Station	Pool	Bus Stop	German Restaurant	Dry Cleaner	Park

K- MEANS K VALUE SELECTION


```python
from kneed import KneeLocator

kl = KneeLocator(range(1, 11), sse, curve="convex", direction="decreasing")
kl.elbow
```

```
4
```

CLUSTERING RESULTS

CLUSTER 1

	Borough	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	10th Most Common Venue
0	Charlottenburg-Wilmersdorf	13.283000	1	Café	Clothing Store	Drugstore	Supermarket	Restaurant	Italian Restaurant	Shopping Mall	Ice Cream Shop	Indian Restaurant	Sandwich Place
2	Lichtenberg	13.498307	1	Bakery	Supermarket	Park	Coffee Shop	ATM	Soccer Stadium	Bookstore	Mexican Restaurant	Pharmacy	Breakfast Spot
4	Mitte	13.367000	1	Hotel	Plaza	Hotel Bar	Museum	Spa	Concert Hall	Coffee Shop	Italian Restaurant	Monument / Landmark	Park
5	Neukölln	13.450000	1	Café	Bar	Supermarket	Italian Restaurant	Coffee Shop	Park	Vegetarian / Vegan Restaurant	Bistro	German Restaurant	Cocktail Bar
6	Pankow	13.403235	1	Café	Supermarket	Bakery	Drugstore	Park	Dance Studio	Organic Grocery	Sushi Restaurant	Burger Joint	Gym / Fitness Center
7	Reinickendorf	13.345394	1	Soccer Field	Supermarket	Metro Station	Plaza	Light Rail Station	Pool	Bus Stop	German Restaurant	Dry Cleaner	Park
9	Steglitz-Zehlendorf	13.252171	1	Café	Italian Restaurant	Supermarket	Doner Restaurant	Drugstore	Pizza Place	Yoga Studio	Fast Food Restaurant	Mobile Phone Shop	Gas Station
10	Tempelhof-Schöneberg	13.383000	1	Supermarket	Café	Bakery	Park	Doner Restaurant	Bus Stop	Italian Restaurant	Fried Chicken Joint	Drugstore	Plaza
11	Treptow-Köpenick	13.567000	1	Café	German Restaurant	Tram Station	River	Hotel	Soccer Stadium	Gas Station	Fast Food Restaurant	Boat or Ferry	Plaza

CLUSTER 0,2 AND 3

	Borough	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	10th Most Common Venue
1	Friedrichshain-Kreuzberg	13.45	0	Bakery	Light Rail Station	Greek Restaurant	Bus Stop	Electronics Store	Café	Farm	Forest	Food & Drink Shop	Flower Shop

	Borough	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	10th Most Common Venue
3	Marzahn-Hellersdorf	13.583	2	Garden	Supermarket	Scenic Lookout	Drugstore	Tea Room	Mountain	Theme Park Ride / Attraction	Yoga Studio	Falafel Restaurant	Food & Drink Shop

	Borough	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	10th Most Common Venue
8	Spandau	13.181689	3	Supermarket	Restaurant	Burger Joint	Mobile Phone Shop	Bus Stop	Automotive Shop	Clothing Store	Bakery	Big Box Store	Farmers Market

DISCUSSION

- Based on the clustering results, we recommend that cluster 3, which included the Spandau region, might be a suitable region for the new fast-food restaurant.

- Its most common venue types are supermarket, restaurant, and burger joint.

- It shows that fast-food costumers already visit this region to have a meal.

- Additionally, it is far from the city center, and this can allow becoming an unrivaled fast-food restaurant around the area.