Capstone Project: The Battle of German Cities
Finding the best city to live in Germany

Introduction

Nowadays, recommender systems play an essential role in our daily lives for making decisions. In the business context, entrepreneurs seek the most suitable place to open a new business. For tourists, they look for a better place to stay for their next travelling destination. The business problem of this project is to help someone looking for a German city to live in. To better describe this problem, let's set a scenario: Lisa, a freelance writer, is looking for a city in Germany for living. She is about to relocate to Germany, and she would like to live in a town having a good environment to optimize her work-life balance, such as jogging in the park.
Also, in a city that offers opportunities for leisure activities, she enjoys visiting museums and theatres in her free time.

Data

In order to find a suitable German city for Lisa to live in, several data is required; firstly, a list of the city is needed, which define the scope of this project. Secondly, these cities' Latitude and longitude coordinates are required to identify them in the map and gain the venue information. This data can be found on a website: Simple Maps, which supporting cities databases and interactive maps in the world. Further, venue data is also needed; this data helps in understanding which venues are particular in a specific area. In order to get the venue information, Foursquare API is implemented. Foursquare is the location technology platform providing users to explore the world. For example, it enables app developers to create customized functionalities for their future uses by understanding human movement in the world map. Finally, packages and libraries applied in this project for handling/analyzing data are: Packages- geopy, geocoder, folium and Libraries- pandas, numPy, random, requests, json_normalize, matplotlib, Kmeans.
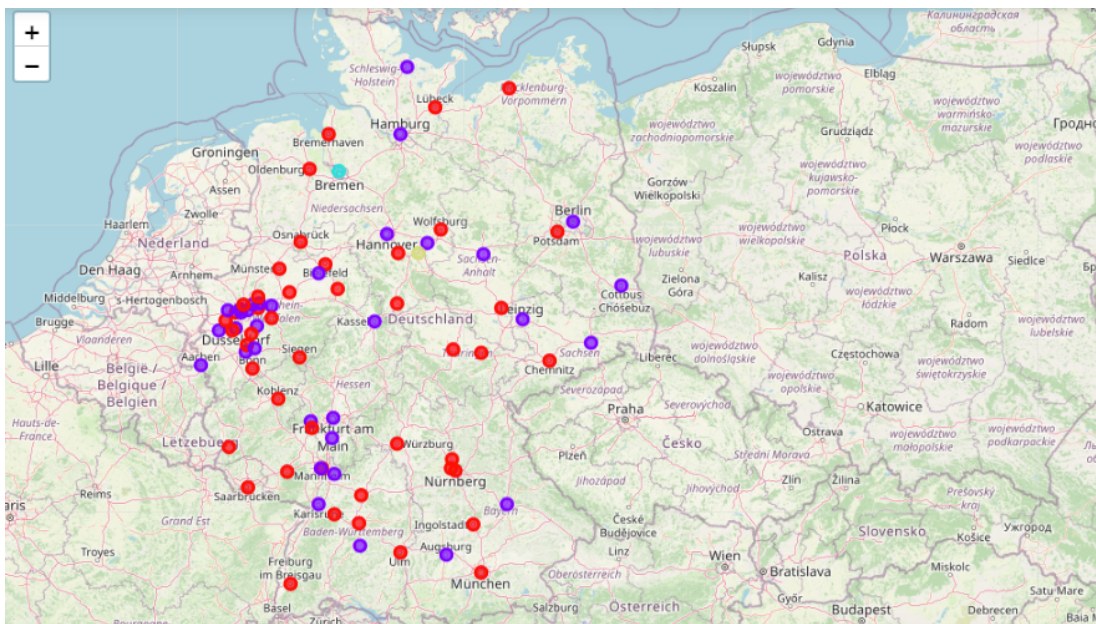
Methodology

The first step, acquiring data by downloading German cities data (CSV format) from "Simple Maps" and convert it into a data frame. Next, pre-process the data by deleting rows having missing data (e.g., rows have no populating data), removing unnecessary columns and modify column names to make the data frame better for analysing. The next step, using Foursquare API to retrieve venue data, and the data is gathered in JSON format (JavaScript Object Notation), which is a lightweight format for storing and transporting data. The fourth step, explore the data, including grouping the cities to discover how many venues in each city and the number of unique venues each city has. The fifth step, implementing the One-hot encoding approach to understand the frequency of each venue category's occurrence. In turn, the top 10 most common venues for each city are found.

Final step, perform K-Means Clustering; the purpose of this algorithm is to segment similar data points into the same group. Before clustering, the optimal cluster number (k) has to be found. Then cluster the data points into k number of clusters. Moreover, using Folium to plot cities and clusters in the map, different clusters are shown with a different colours on the map. As a result, several data frames representing different clusters are created. These data frames include different city names, these cities' geo data (latitude and longitude coordinates) as well as the top 10 most common venues of the corresponding cities. Hence, this data allows us to identify which cluster contains the most suitable cities for Lisa to live in.

Discussions and results

According to K-Means Clustering, four is decided as the optimal number of clusters. Further, plotting these German cities on the map in clusters (showing in distinct colours in the following map).

Finally, examine different clusters:

-Cluster 1: Café, bars, restaurants, shopping

	city	population	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
2	Munich	1471508.0	0.0	Café	Hotel	Plaza	Clothing Store	Bavarian Restaurant	Coffee Shop	German Restaurant	Boutique	Shopping Mall	Cocktail Bar
6	Stuttgart	634830.0	0.0	Café	Bar	German Restaurant	Nightclub	Clothing Store	Plaza	Coffee Shop	Cocktail Bar	Wine Bar	Italian Restaurant
13	Nuremberg	518365.0	0.0	Café	German Restaurant	Coffee Shop	Bar	Plaza	Italian Restaurant	Hotel	Franconian Restaurant	Drugstore	Modern European Restaurant
14	Duisburg	498590.0	0.0	Café	Bar	Italian Restaurant	Drugstore	Music Venue	Greek Restaurant	Bakery	Gift Shop	Sushi Restaurant	Restaurant
15	Bochum	364628.0	0.0	Ice Cream Shop	Café	Bakery	Drugstore	German Restaurant	Toy / Game Store	Pedestrian Plaza	Beer Bar	Plaza	Coffee Shop
17	Bielefeld	333786.0	0.0	Café	Mediterranean Restaurant	Bar	Coffee Shop	Ice Cream Shop	Light Rail Station	Steakhouse	Mexican Restaurant	Spanish Restaurant	Castle
18	Bonn	327258.0	0.0	Café	Bar	Drugstore	Italian Restaurant	Plaza	Pub	Nightclub	Ice Cream Shop	Hotel	German Restaurant

	City												
19	Münster	314319.0	0.0	Café	Italian Restaurant	Bar	Plaza	Gastropub	Bookstore	Ice Cream Shop	Pub	Bakery	German Restaurant
21	Mannheim	309370.0	0.0	Café	Plaza	Italian Restaurant	Hotel	Restaurant	Bar	German Restaurant	Turkish Restaurant	Sushi Restaurant	Farmers Market
29	Chemnitz	245403.0	0.0	Café	Ice Cream Shop	Plaza	Clothing Store	Pub	Italian Restaurant	Drugstore	Outdoor Sculpture	German Restaurant	Sushi Restaurant
30	Halle	239257.0	0.0	Café	German Restaurant	Bar	Supermarket	Drugstore	Clothing Store	Eastern European Restaurant	Plaza	Sandwich Place	Italian Restaurant
32	Freiburg im Breisgau	230241.0	0.0	Café	Italian Restaurant	German Restaurant	Plaza	Sushi Restaurant	Asian Restaurant	Hotel	Beer Garden	Historic Site	Bar
33	Krefeld	227020.0	0.0	Café	Clothing Store	Drugstore	Sushi Restaurant	Shopping Mall	Bar	Bank	Theater	Bakery	Fast Food Restaurant
34	Lübeck	217198.0	0.0	Café	German Restaurant	Coffee Shop	Candy Store	Clothing Store	Seafood Restaurant	Restaurant	Italian Restaurant	Plaza	Drugstore
36	Rostock	208886.0	0.0	Café	Hotel	Historic Site	Greek Restaurant	Rental Car Location	Nightclub	Tram Station	Event Space	Electronics Store	English Restaurant
37	Erfurt	203254.0	0.0	German Restaurant	Café	Drugstore	Ice Cream Shop	Italian Restaurant	Plaza	Bar	Clothing Store	Public Art	Restaurant
39	Hagen	188814.0	0.0	Clothing Store	Bakery	Pizza Place	Park	Shopping Mall	Café	Bookstore	Electronics Store	Optical Shop	Beer Bar
40	Mainz	184997.0	0.0	Café	German Restaurant	Plaza	Coffee Shop	Italian Restaurant	Irish Pub	African Restaurant	Bus Stop	Doner Restaurant	Supermarket
41	Saarbrücken	180741.0	0.0	Bar	Café	Ice Cream Shop	Italian Restaurant	Nightclub	German Restaurant	Hotel	Thai Restaurant	Clothing Store	Sushi Restaurant
42	Hamm	179111.0	0.0	Zoo	German Restaurant	Furniture / Home Store	Supermarket	Falafel Restaurant	Escape Room	Ethiopian Restaurant	Event Space	Exhibit	Fast Food Restaurant
43	Potsdam	178089.0	0.0	German Restaurant	Café	Restaurant	Science Museum	Drugstore	Organic Grocery	Market	Plaza	Exhibit	Electronics Store
46	Oldenburg	168210.0	0.0	Café	Bar	Nightclub	Clothing Store	Italian Restaurant	Drugstore	Coffee Shop	Plaza	Vietnamese Restaurant	French Restaurant
47	Osnabrück	164748.0	0.0	Café	German Restaurant	Spanish Restaurant	Ice Cream Shop	Burger Joint	Department Store	Bar	Pub	Museum	Hotel
48	Leverkusen	163838.0	0.0	Italian Restaurant	Clothing Store	Bakery	Hotel	Shopping Mall	Ice Cream Shop	Café	Electronics Store	German Restaurant	Restaurant
50	Solingen	159360.0	0.0	Light Rail Station	Cocktail Bar	Ice Cream Shop	Mongolian Restaurant	Bakery	Hockey Rink	Kitchen Supply Store	Furniture / Home Store	Clothing Store	Café
53	Neuss	153796.0	0.0	Café	Drugstore	Turkish Restaurant	Gym / Fitness Center	Tram Station	Bar	Lounge	Theater	Italian Restaurant	German Restaurant
55	Paderborn	150580.0	0.0	Furniture / Home Store	Pub	Restaurant	Asian Restaurant	Supermarket	Falafel Restaurant	English Restaurant	Escape Room	Ethiopian Restaurant	Event Space
56	Ingolstadt	136981.0	0.0	Café	German Restaurant	Greek Restaurant	Italian Restaurant	Bar	Steakhouse	Pedestrian Plaza	Drugstore	Plaza	Theater
57	Würzburg	127880.0	0.0	Café	German Restaurant	Plaza	Wine Bar	Bar	Sushi Restaurant	Steakhouse	Hotel	Beer Garden	Juice Bar
58	Fürth	127748.0	0.0	Café	Vietnamese Restaurant	Gastropub	German Restaurant	Coffee Shop	Theater	American Restaurant	Pub	Spanish Restaurant	Irish Pub
59	Ulm	126329.0	0.0	Café	Bar	German Restaurant	Pub	Italian Restaurant	Burger Joint	Asian Restaurant	Coffee Shop	Plaza	Drugstore
60	Heilbronn	125960.0	0.0	Italian Restaurant	German Restaurant	Burger Joint	Ice Cream Shop	Café	Greek Restaurant	Sushi Restaurant	Gastropub	Sports Bar	Bookstore
61	Pforzheim	125542.0	0.0	Café	Fast Food Restaurant	Supermarket	Asian Restaurant	Plaza	Thai Restaurant	Platform	Clothing Store	Nightclub	Chinese Restaurant

	city	population											
62	Wolfsburg	124151.0	0.0	Drugstore	Bistro	Italian Restaurant	Bar	Café	Beer Garden	Clothing Store	Korean Restaurant	Sushi Restaurant	Supermarket
63	Göttingen	119801.0	0.0	Café	Italian Restaurant	Supermarket	Drugstore	Pub	Bar	Coffee Shop	Historic Site	Outdoor Sculpture	Pastry Shop
64	Bottrop	117383.0	0.0	Café	Supermarket	Plaza	Sushi Restaurant	Pub	Clothing Store	Fast Food Restaurant	Steakhouse	Ice Cream Shop	Hotel
66	Koblenz	114024.0	0.0	Italian Restaurant	Café	Clothing Store	German Restaurant	Bar	Plaza	Ice Cream Shop	Drugstore	Shopping Mall	Bakery
67	Bremerhaven	113634.0	0.0	Hotel	Drugstore	Bus Stop	Irish Pub	Italian Restaurant	Café	Zoo	Falafel Restaurant	Escape Room	Ethiopian Restaurant
68	Recklinghausen	112267.0	0.0	Clothing Store	Italian Restaurant	Café	Drugstore	Ice Cream Shop	Shopping Mall	Shoe Store	Supermarket	German Restaurant	Hotel
70	Erlangen	111962.0	0.0	Café	Coffee Shop	Drugstore	Cocktail Bar	Greek Restaurant	Mediterranean Restaurant	Clothing Store	Sushi Restaurant	German Restaurant	Bakery
71	Jena	111407.0	0.0	Café	German Restaurant	Hotel	Pub	Drugstore	Bar	Plaza	Nightclub	Burger Joint	Fast Food Restaurant
73	Trier	110636.0	0.0	German Restaurant	Hotel	Drugstore	Café	Bar	Coffee Shop	Clothing Store	Pizza Place	Ice Cream Shop	Italian Restaurant
76	Siegen	102836.0	0.0	Clothing Store	Café	Drugstore	Supermarket	Fast Food Restaurant	Electronics Store	Nightclub	Frozen Yogurt Shop	Sushi Restaurant	Pizza Place
77	Hildesheim	101990.0	0.0	Seafood Restaurant	Café	Drugstore	Restaurant	German Restaurant	Hotel	Coffee Shop	Sandwich Place	Chinese Restaurant	Steakhouse
80	Kaiserslautern	100030.0	0.0	Bar	Café	Hotel	Clothing Store	German Restaurant	Ice Cream Shop	Drugstore	Pub	Sushi Restaurant	Supermarket

-Cluster 2: Hotels and leisure (museums, theaters, parks)

	city	population	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	Berlin	3644826.0	1.0	Plaza	Gourmet Shop	Monument / Landmark	Café	Sandwich Place	Bookstore	Restaurant	Clothing Store	Cocktail Bar
1	Hamburg	1841179.0	1.0	Coffee Shop	Hotel	Vietnamese Restaurant	Café	Cosmetics Shop	German Restaurant	Restaurant	Burger Joint	Clothing Store
3	Cologne	1085664.0	1.0	Hotel	Italian Restaurant	Café	Coffee Shop	Bakery	German Restaurant	Plaza	Art Museum	Gift Shop
4	Frankfurt	753056.0	1.0	Café	Plaza	Coffee Shop	Bar	Nightclub	Pastry Shop	Italian Restaurant	Art Museum	Hotel
7	Düsseldorf	619294.0	1.0	Art Museum	Brewery	Nightclub	Bar	Plaza	Hotel	Park	Steakhouse	Beer Garden
8	Dortmund	587010.0	1.0	Café	Coffee Shop	Plaza	Bakery	Ice Cream Shop	Hotel	Pizza Place	Clothing Store	Japanese Restaurant
9	Essen	583109.0	1.0	Hotel	Plaza	Italian Restaurant	Bakery	Coffee Shop	Drugstore	Cocktail Bar	Seafood Restaurant	Furniture / Home Store
10	Dresden	555805.0	1.0	German Restaurant	Hotel	Burger Joint	Italian Restaurant	Museum	Café	Plaza	Art Museum	Clothing Store
11	Leipzig	542529.0	1.0	Hostel	Tram Station	Zoo	Falafel Restaurant	English Restaurant	Escape Room	Ethiopian Restaurant	Event Space	Exhibit
12	Hannover	538068.0	1.0	Coffee Shop	Café	Italian Restaurant	Burger Joint	Sushi Restaurant	Theater	Clothing Store	Plaza	Cocktail Bar
16	Wuppertal	354382.0	1.0	Theater	Cable Car	Grilled Meat Restaurant	Liquor Store	Trail	Supermarket	Electronics Store	English Restaurant	Escape Room
20	Karlsruhe	313092.0	1.0	Theme Park Ride / Attraction	Museum	Outdoor Sculpture	Palace	Park	Café	Restaurant	Botanical Garden	African Restaurant
22	Augsburg	295135.0	1.0	Italian Restaurant	Indie Movie Theater	Historic Site	Bar	Theater	Steakhouse	Bavarian Restaurant	Garden	Clothing Store
23	Wiesbaden	278342.0	1.0	Café	Plaza	Hotel	Thai Restaurant	Theater	Sushi Restaurant	Seafood Restaurant	Turkish Restaurant	Supermarket
24	Mönchengladbach	261454.0	1.0	Convenience Store	Hotel	Playground	Event Space	Zoo	Falafel Restaurant	English Restaurant	Escape Room	Ethiopian Restaurant

	city	population	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
25	Gelsenkirchen	260654.0	1.0	Spanish Restaurant	Cocktail Bar	Chinese Restaurant	Hotel	Ice Cream Shop	Soup Place	Hookah Bar	Business Service	Mexican Restaurant	
26	Braunschweig	248292.0	1.0	Mexican Restaurant	Café	Brewery	Italian Restaurant	Burger Joint	Breakfast Spot	Bookstore	Falafel Restaurant	Nightclub	
27	Kiel	247548.0	1.0	Coffee Shop	Plaza	Clothing Store	Hotel	Fast Food Restaurant	Cocktail Bar	Restaurant	Drugstore	Café	
28	Aachen	247380.0	1.0	Drugstore	Ice Cream Shop	Italian Restaurant	Plaza	Bakery	Hotel	Café	German Restaurant	Sushi Restaurant	
31	Magdeburg	238697.0	1.0	Hotel	Bakery	Platform	Bookstore	Ice Cream Shop	Italian Restaurant	Café	Asian Restaurant	Pizza Place	
35	Oberhausen	210829.0	1.0	Bus Station	Food & Drink Shop	Drugstore	Pub	Hotel	Sandwich Place	Steakhouse	Rock Club	Bookstore	
38	Kassel	201585.0	1.0	Clothing Store	Bakery	Italian Restaurant	Coffee Shop	Hotel	Drugstore	Plaza	Café	Restaurant	
44	Ludwigshafen	171061.0	1.0	Platform	Hotel Bar	Fast Food Restaurant	Park	Doner Restaurant	Metro Station	Rental Car Location	Tram Station	Bus Stop	
45	Mülheim	170880.0	1.0	Hotel	Ice Cream Shop	Bar	Restaurant	Italian Restaurant	Museum	Metro Station	Coffee Shop	Thai Restaurant	
49	Heidelberg	160355.0	1.0	Hotel	Plaza	Bar	German Restaurant	Cocktail Bar	Trail	Restaurant	Coffee Shop	Burger Joint	
51	Darmstadt	159207.0	1.0	Turkish Restaurant	Sushi Restaurant	Coffee Shop	Bar	Sporting Goods Shop	Gas Station	Church	Supermarket	Bookstore	
52	Herne	156374.0	1.0	Bakery	Drugstore	Platform	Supermarket	Bus Station	Fast Food Restaurant	Theater	Insurance Office	Metro Station	
54	Regensburg	152610.0	1.0	German Restaurant	Plaza	Art Gallery	Food & Drink Shop	Mexican Restaurant	Arts & Entertainment	Park	Hotel	Deli / Bodega	
65	Reutlingen	115966.0	1.0	Miscellaneous Shop	Hardware Store	Breakfast Spot	Hotel	Insurance Office	Gym / Fitness Center	Supermarket	Escape Room	Ethiopian Restaurant	
69	Bergisch Gladbach	111966.0	1.0	Art Museum	German Restaurant	Music Venue	Business Service	Restaurant	Mexican Restaurant	Farmers Market	Falafel Restaurant	Escape Room	
75	Moers	103725.0	1.0	Pool Hall	Gym	Insurance Office	Pet Store	Exhibit	Electronics Store	English Restaurant	Escape Room	Ethiopian Restaurant	
78	Cottbus	100219.0	1.0	German Restaurant	Hotel	Drugstore	Gym / Fitness Center	Café	Burger Joint	Bowling Alley	Market	Clothing Store	
79	Gütersloh	100194.0	1.0	Park	Rock Club	Liquor Store	Chinese Restaurant	Hotel	Food	Fish Market	Fast Food Restaurant	Farmers Market	

-Cluster 3: Event places, soccer fields

	city	population	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
5	Bremen	724909.0	2.0	Soccer Field	Zoo	Farmers Market	English Restaurant	Escape Room	Ethiopian Restaurant	Event Space	Exhibit	Falafel Restaurant	Fast Food Restaurant

-Cluster 4: Event places, intersections

	city	population	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
74	Salzgitter	104948.0	3.0	Intersection	Zoo	Farmers Market	English Restaurant	Escape Room	Ethiopian Restaurant	Event Space	Exhibit	Falafel Restaurant	Fast Food Restaurant

According to the top 10 most common venues of each city, it shows cluster 2 is recommended for Lisa to live because this cluster contains a city with more opportunities to do leisure activities..

Conclusion

To conclude, cluster 2 contains the most suitable cities for Lisa to move to; as mentioned in the introduction part, she wishes to move to a city to optimize her work-life balance and fulfil her interests in leisure activities. Therefore, cluster 2, which include cities having common venues of parks, museums and theatres, is recommended to Lisa. Cities that belong to cluster 2 are: Berlin, Hamburg, Cologne, Frankfurt, Düsseldorf, Dortmund, Essen, Dresden, Leipzig, Hannover,

Wuppertal, Karlsruhe, Augsburg, Wiesbaden, Mönchengladbach, Gelsenkirchen, Braunschweig, Kiel, Aachen, Magdeburg, Oberhausen, Kassel, Ludwigshafen, Mülheim, Heidelberg, Darmstadt, Herne, Regensburg, Reutlingen, Bergisch Gladbach, Moers, Cottbus, Gütersloh.