

Applied Data Science Capstone

By Markus Engler

Contend

- Main Idea
- Used Data
- Used Methods
- Results
- Conclusion

Main Idear

- Main idea of the project is to compute a tool to search for the optimal geographical location for opening a specific new venue in a German city.
- The analysis should be valuable to all business owners or public organizations looking for a new location for their projects.
- The result can provide a first indication of optimal placement or area for optimal placement of a new venue. In the next steps the result should be enriched with several other datasets to improve information content and should be checked for accuracy for final decision.

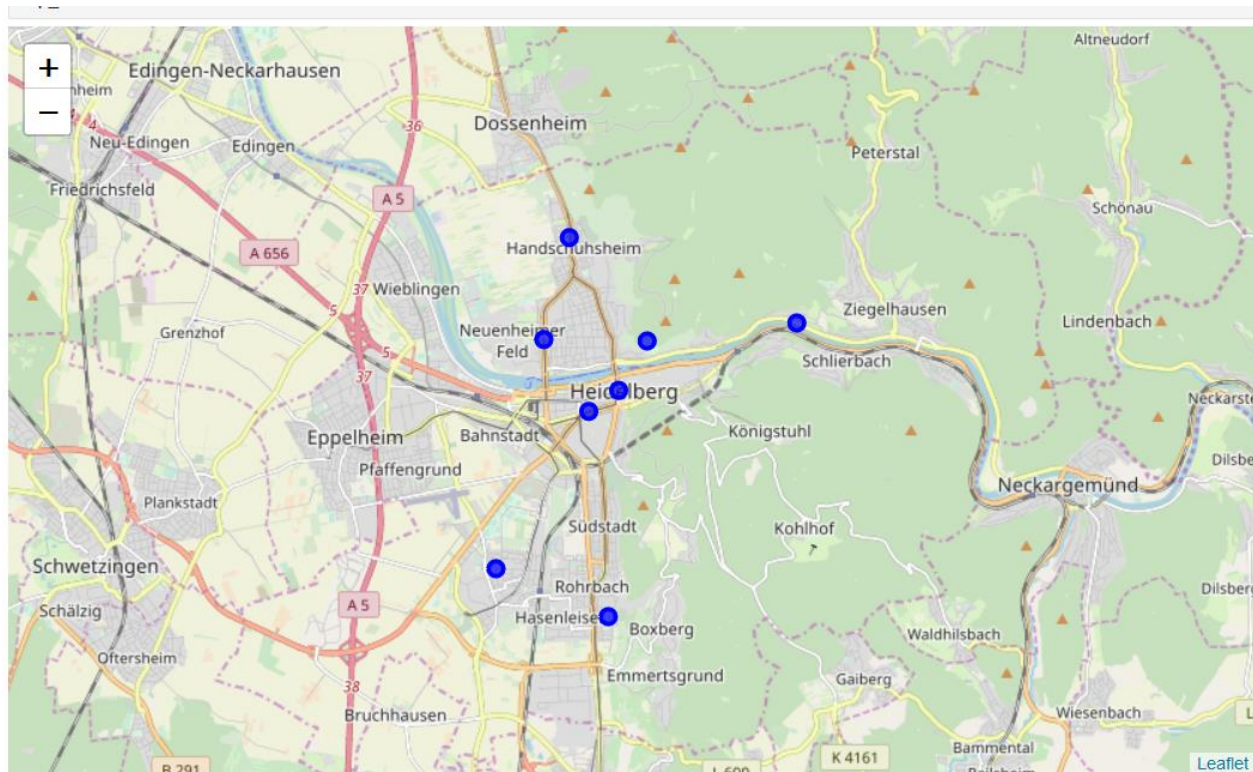
Used Data 1/3

- As first data source I will use information on German city location, which can be found on following page:
<https://raw.githubusercontent.com/TrustChainEG/postal-codes-json-xml-csv/master/data/DE/zipcodes.de.csv>
- This data includes the zip code of every city/region of Germany, which are used as identifier for the area of interest in this project.
- Example of German zip code data:

```
country_code,zipcode,place,state,state_code,province,province_code,community,community_code,latitude,longitude
DE,01945,Grünewald,Brandenburg,BB,"",00,Landkreis Oberspreewald-Lausitz,12066,51.4,14
DE,01945,Lindenau,Brandenburg,BB,"",00,Landkreis Oberspreewald-Lausitz,12066,51.4,13.7333
DE,01945,Hohenbocka,Brandenburg,BB,"",00,Landkreis Oberspreewald-Lausitz,12066,51.431,14.0098
DE,01945,Schwarzbach,Brandenburg,BB,"",00,Landkreis Oberspreewald-Lausitz,12066,51.45,13.9333
DE,01945,Guteborn,Brandenburg,BB,"",00,Landkreis Oberspreewald-Lausitz,12066,51.4167,13.9333
DE,01945,Hermsdorf,Brandenburg,BB,"",00,Landkreis Oberspreewald-Lausitz,12066,51.4055,13.8937
DE,01945,Ruhland,Brandenburg,BB,"",00,Landkreis Oberspreewald-Lausitz,12066,51.4576,13.8664
```

Used Data 2/3

- Coordinates for all districts of area of interest are displayed on map. As an example we choose the city of Heidelberg, Germany. (Blue)



Used Data 3/3

- In a second step the zip code data is enriched with venue information of the area using the Foursquare API.
- Example of resulting data frame including venue information:

```
[12]: frankfurt_venues.head(10)
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	60306	50.1159	8.6702	Opernplatz	50.115399	8.671772	Plaza
1	60306	50.1159	8.6702	Alte Oper	50.115350	8.671428	Opera House
2	60306	50.1159	8.6702	Manufactum brot&butter	50.115958	8.670772	Food & Drink Shop
3	60306	50.1159	8.6702	The Ivory Club	50.114309	8.669109	Indian Restaurant
4	60306	50.1159	8.6702	ZENZAKAN - Pan Asian Supperclub	50.114867	8.669458	Asian Restaurant
5	60306	50.1159	8.6702	Moriki	50.113863	8.669530	Japanese Restaurant
6	60306	50.1159	8.6702	Schneider's Café Snackbar	50.115685	8.669928	Café
7	60306	50.1159	8.6702	Kameha Suite	50.114732	8.670210	Mediterranean Restaurant
8	60306	50.1159	8.6702	MEYER Feinkost Frankfurt	50.114819	8.673398	Gourmet Shop
9	60306	50.1159	8.6702	Sofitel Frankfurt Opera	50.116294	8.673624	Hotel

Used Methods 1/5

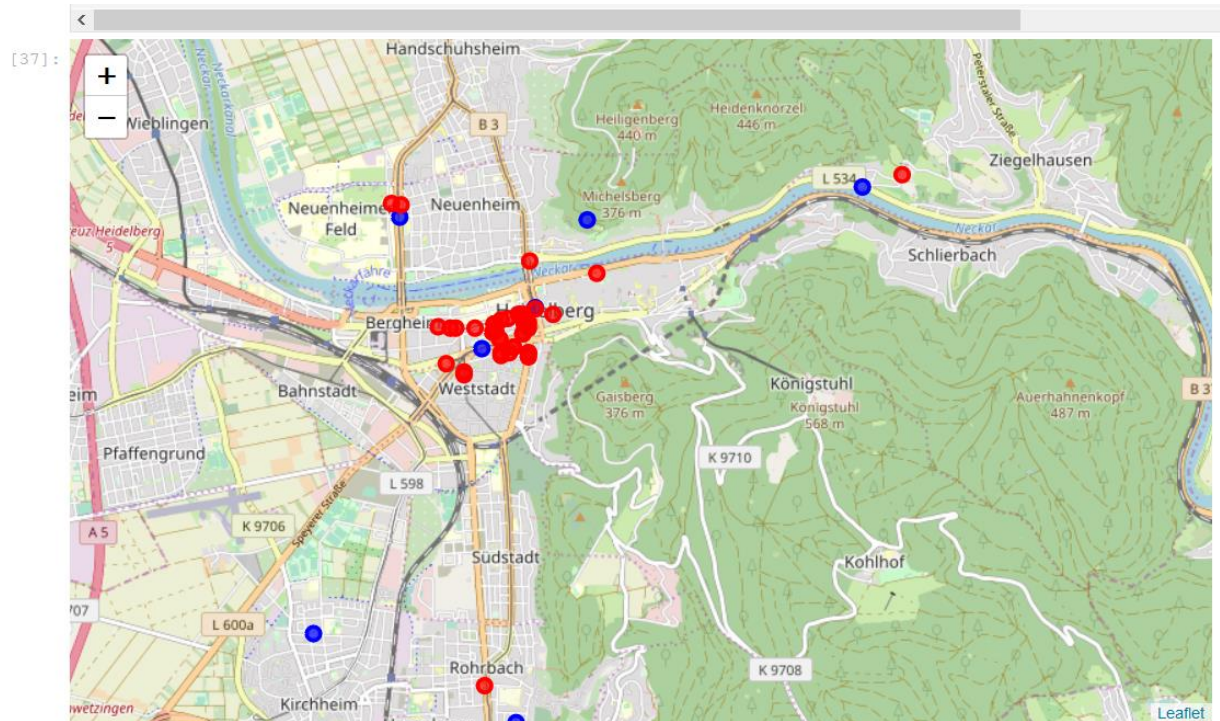
- The 10 most common venues for each district are displayed to enable the user to chose from existing venue categories.

[23]:

	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	69115	Hotel	Italian Restaurant	Restaurant	Coffee Shop	Café	Chinese Restaurant	Asian Restaurant	Supermarket	Vegetarian / Vegan Restaurant	Middle Eastern Restaurant
1	69117	Café	Hotel	Bakery	Cocktail Bar	Vegetarian / Vegan Restaurant	Plaza	Coffee Shop	Italian Restaurant	Chinese Restaurant	Clothing Store
2	69118	German Restaurant	Brewery	Wine Bar	Cupcake Shop	Halal Restaurant	Gym / Fitness Center	Grocery Store	Gastropub	Gas Station	Garden
3	69120	Tram Station	Café	Supermarket	Italian Restaurant	Bakery	Drugstore	German Restaurant	Gas Station	Falafel Restaurant	Fast Food Restaurant
4	69121	Hotel	Gas Station	Supermarket	Organic Grocery	Cupcake Shop	Gym / Fitness Center	Grocery Store	German Restaurant	Gastropub	Garden
5	69123	Garden	Concert Hall	Other Great Outdoors	Boat or Ferry	Japanese Restaurant	Scenic Lookout	Wine Bar	Drugstore	Falafel Restaurant	Fast Food Restaurant
6	69124	Bakery	Pizza Place	Trattoria/Osteria	Wine Bar	Gym / Fitness Center	Grocery Store	German Restaurant	Gastropub	Gas Station	Garden
7	69126	Ice Cream Shop	Rental Car Location	Bakery	Tram Station	Thai Restaurant	Pharmacy	Supermarket	Light Rail Station	Doner Restaurant	Drugstore

Used Methods 2/5

- Coordinates for all venues of specific kind are displayed on map, together with location of city districts. As an example we choose the venue category Restaurant. (Red)



Used Methods 3/5

- A pairwise calculation of distances between venues of the datasets is executed to calculate distance matrix.

	0	1	2	3	4	5	6	7	8	9	...	33
0	0.000000	147.937443	89.141217	146.835502	88.181823	439.327026	141.166470	293.309989	282.566147	222.865825	...	756.210220
1	147.937443	0.000000	60.033666	292.519395	82.186894	520.007059	271.368143	407.986347	234.164361	277.837925	...	647.375930
2	89.141217	60.033666	0.000000	232.529261	34.681095	476.182847	220.598247	364.860177	253.154670	255.281847	...	695.729779
3	146.835502	292.519395	232.529261	0.000000	219.871807	364.104077	132.048103	258.529088	404.178669	285.965055	...	891.896833
4	88.181823	82.186894	34.681095	219.871807	0.000000	443.054400	228.206837	377.787228	287.797664	281.325201	...	726.734144
5	439.327026	520.007059	476.182847	364.104077	443.054400	0.000000	495.175926	613.870691	719.848443	641.827116	...	1167.011007
6	141.166470	271.368143	220.598247	132.048103	228.206837	495.175926	0.000000	154.289153	301.831961	157.795407	...	795.202569
7	293.309989	407.986347	364.860177	258.529088	377.787228	613.870691	154.289153	0.000000	356.766410	178.867808	...	828.690002
8	282.566147	234.164361	253.154670	404.178669	287.797664	719.848443	301.831961	356.766410	0.000000	177.956331	...	493.387488
9	222.865825	277.837925	255.281847	285.965055	281.325201	641.827116	157.795407	178.867808	177.956331	0.000000	...	658.861430

Used Methods 4/5

- The resulting distance matrix is used to assign the two nearest venues to each datapoint (venue). Resulting data frame including two nearest venues of each venue:

[47]:

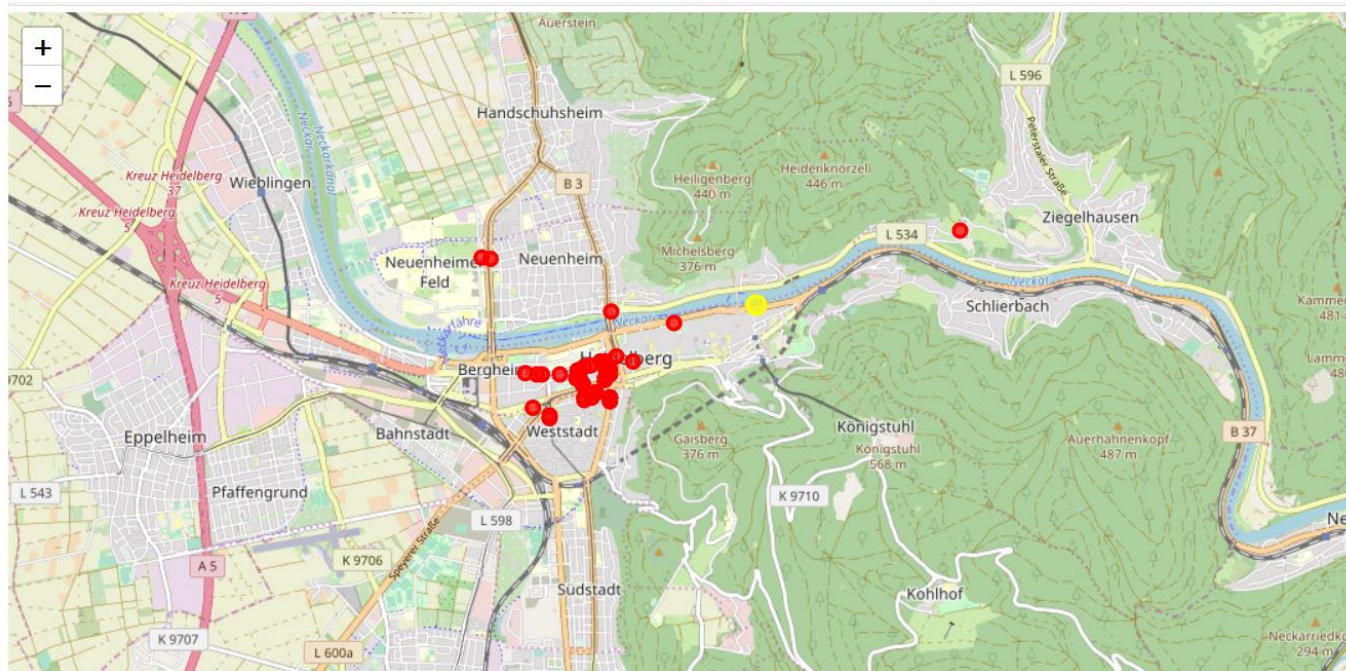
	First Venue	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue1 Latitude	Venue1 Longitude	Venue Category	Second Venue	Venue2 Latitude	Venue2 Longitude	Third Venue	Venue3 Latitude	Venue3 Longitude
0	Baan Thai	69126	49.3757	8.6911	49.378656	8.687056	Thai Restaurant	Same Same	49.404080	8.684501	Da Baggio	49.404234	8.684587
1	BräuStadel	69120	49.4169	8.6764	49.418056	8.675446	German Restaurant	Enoteca Cesarino	49.417925	8.676667	Qube Restaurant	49.407923	8.681275
2	Da Baggio	69115	49.4062	8.6868	49.404234	8.684587	Italian Restaurant	Same Same	49.404080	8.684501	Palmyra	49.404923	8.682225
3	Darwisch Persische	69115	49.4062	8.6868	49.405476	8.692586	Middle Eastern Restaurant	Hot Pot House	49.405851	8.692674	Hot Pot House	49.405851	8.692674
4	Enoteca Cesarino	69120	49.4169	8.6764	49.417925	8.676667	Italian Restaurant	BräuStadel	49.418056	8.675446	Qube Restaurant	49.407923	8.681275

Used Methods 5/5

- From these three venues a triangular is constructed to calculate the centron of the involved venues.
- Afterwards the distances of the centron to each of the venues is calculated.
- To find the coordinates for optimal venue placements, the algorithm looks for the maximum of the sum of distances between the centron and the venues.

Results

- This calculation provides the coordinates of the centron with the highest possible distance to all venues in the dataset, with the coordinates of latitude 49.41 and longitude 8.71. (Yellow)



Conclusion

- Presented tool gives an indication for optimal placement of new venue of specific kind in a city/region of interest.
- The user can choose from all cities/regions in Germany and is able to specify a common venue category based on existing top venues or own choice.
- The resulting coordinates give the optimal location based on trigonometrical calculations using existing venues of the specified type.
- Calculation is based only on available venues in the database and uses distances between venues as optimization criteria only.