

Buying a home in Germany - Categorical geospatial study

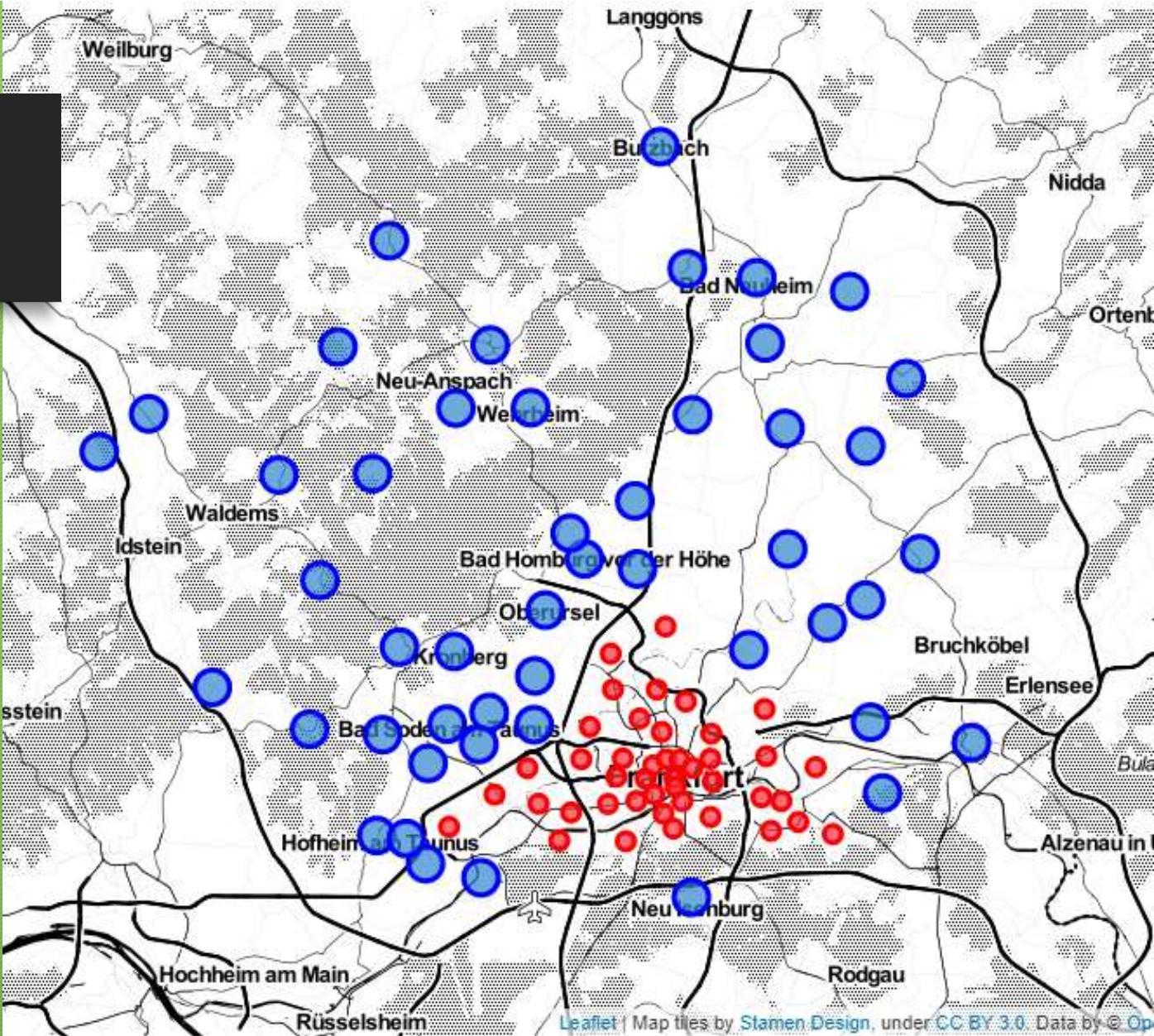
Applied Data Science Capstone by IBM/Coursera

Buying a home in Germany - Categorical geospatial study in Bad Homburg vor der Höhe

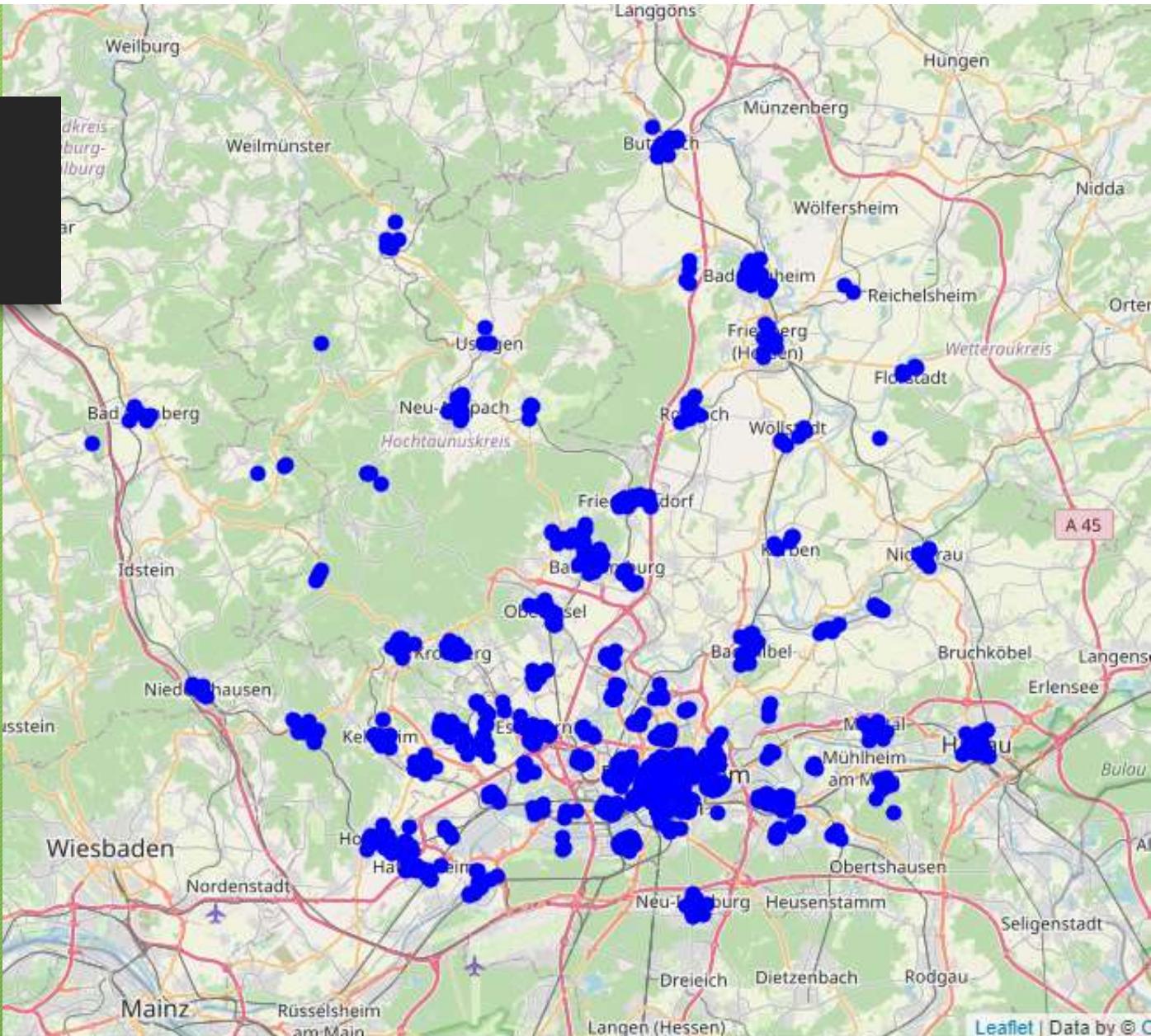
- Buying a home can be a good investment, especially if you are planning to take advantage of good market conditions. One of the most important and impactful decisions to make, when choosing to buy a home, is the location. The key factors of a great location are accessibility, appearance, and amenities available in a neighborhood.
- Considering the amenities available in a neighborhood, these translate directly to a better living and quality of life, for example: consider not having to travel far to go to a doctor's appointment, doing your shopping or dining or getting food in a good restaurant near your house, these are some of the hallmarks of a great neighborhood.
- Our final objective is clustering and comparing areas to try to find and group the best places to live.

Locations to study

- In the map we can see all areas involved in this study. Here you can see blue and red circles, these circles represent the individual areas where we will search for amenities. A blue circle area is 1000 meters, and the red circle area is 500 meters.
- Why did we choose to make this distinction? Because comparing a small city to a big city it's not feasible, the difference in amenities density would be too big.
- We will tolerate some overlapping because we want to find the places with the most amenities available and if it is within the radius of the area, we will keep it.



Amenities dataset



PLZ	Ort	Ort_Lat	Ort_Lon	Amenity	Am_Lat	Am_Lon	Am_Category
65520	Bad Camberg	50.299603	8.267312	REWE	50.301920	8.260702	Supermarket
65520	Bad Camberg	50.299603	8.267312	India Haus	50.298974	8.264521	Indian Restaurant
65520	Bad Camberg	50.299603	8.267312	Kurpark Bad Camberg	50.297865	8.272318	Park
65520	Bad Camberg	50.299603	8.267312	Banyan Garden	50.295113	8.269464	Chinese Restaurant
65520	Bad Camberg	50.299603	8.267312	Cafe Schäfer	50.302721	8.260552	Bakery

Amenities categories analysis

Am_Category	Amenity (Desc)	Am_Category	Amenity (Asc)
Supermarket	132	Accessories Store	1
Italian Restaurant	98	Medical Center	1
Café	92	Malga	1
Hotel	65	Malay Restaurant	1
German Restaurant	63	Locksmith	1
Bakery	51	Liquor Store	1
Drugstore	42	Laser Tag	1
Ice Cream Shop	41	Jazz Club	1
Pizza Place	39	Israeli Restaurant	1
Restaurant	31	Hockey Rink	1

We can't explore 234 categories, so let's categorize the categories into a sort of macro categories, converting 234 into 10 manageable categories.

The new categories will be:

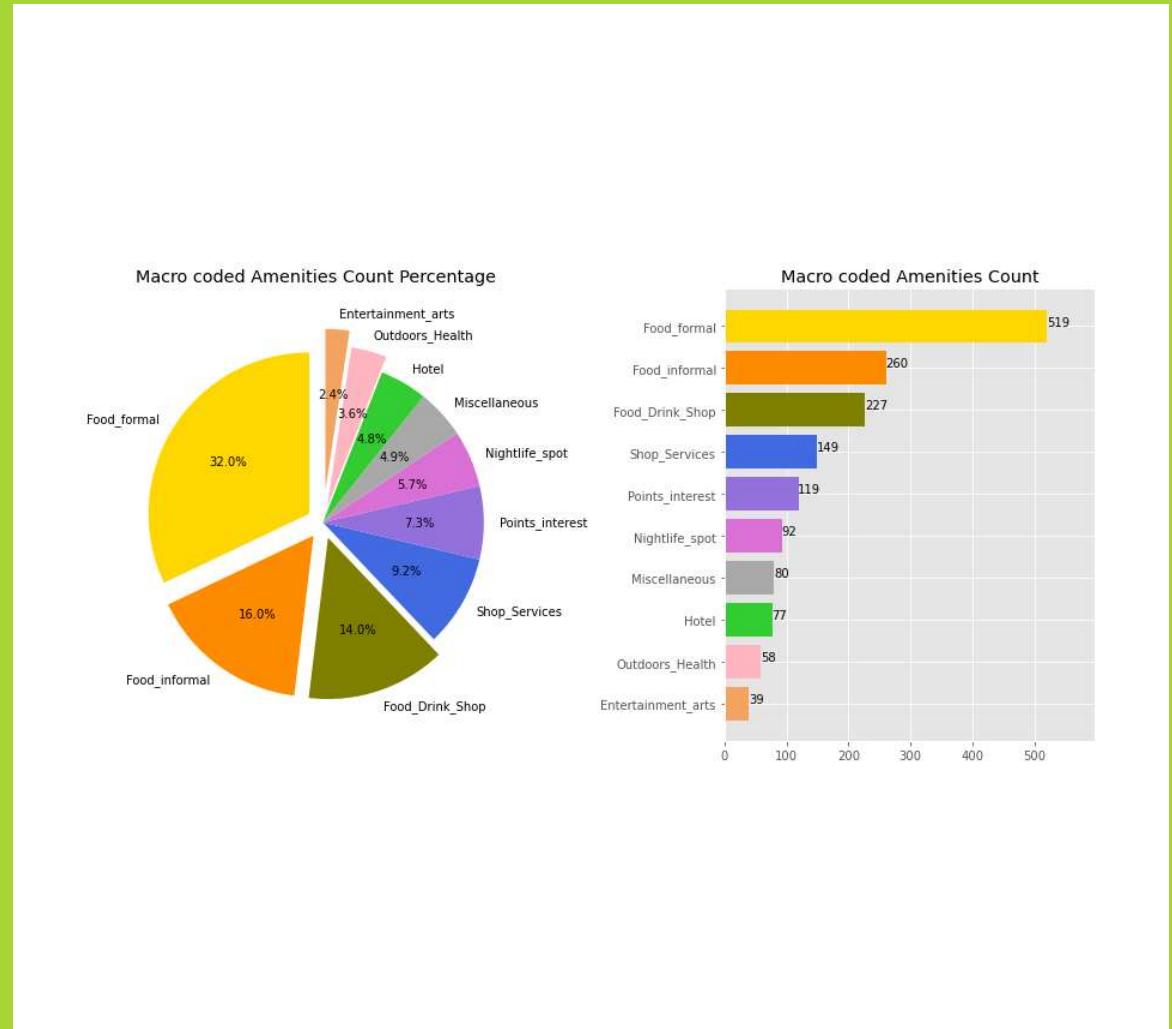
1. Food_formal - For this category we will consider pretty much all the categories that have the word restaurant in the name
2. Food_informal - This category will have all the places where you can get quick food
3. Food_Drink_Shop - Here we will have groceries and drink shops
4. Hotel - Accommodation spots
5. Nightlife_spot - Drinking spots, bars, etc.
6. Entertainment_arts - Clubs, Museums, Theater, etc.
7. Outdoors_Health - Sports and Health category
8. Shop_Services - Shops and Services
9. Points_interest - Landmarks
10. Miscellaneous - All the other stuff that can't quite fit in any other category

Recoded dataset

df_am.head(10)									
	PLZ	Ort	Ort_Lat	Ort_Lon	Amenity	Am_Lat	Am_Lon	Am_Category	Coded_Cat
0	65520	Bad Camberg	50.299603	8.267312	REWE	50.301920	8.260702	Supermarket	Food_Drink_Shop
1	65520	Bad Camberg	50.299603	8.267312	India Haus	50.298974	8.264521	Indian Restaurant	Food_formal
2	65520	Bad Camberg	50.299603	8.267312	Kurpark Bad Camberg	50.297865	8.272318	Park	Points_interest
3	65520	Bad Camberg	50.299603	8.267312	Banyan Garden	50.295113	8.269464	Chinese Restaurant	Food_formal
4	65520	Bad Camberg	50.299603	8.267312	Cafe Schäfer	50.302721	8.260552	Bakery	Food_Drink_Shop
5	65520	Bad Camberg	50.299603	8.267312	Bahnhof Bad Camberg	50.295386	8.256534	Train Station	Points_interest
6	61348	Bad Homburg vor der Höhe	50.226423	8.611368	Schloss Bad Homburg	50.227895	8.610084	Palace	Points_interest
7	61348	Bad Homburg vor der Höhe	50.226423	8.611368	Schreinerei Pfeiffer	50.228371	8.615125	German Restaurant	Food_formal
8	61348	Bad Homburg vor der Höhe	50.226423	8.611368	Bella	50.226868	8.616227	Ice Cream Shop	Food_informal
9	61348	Bad Homburg vor der Höhe	50.226423	8.611368	Taverne Kavala	50.231005	8.610217	Taverna	Food_formal

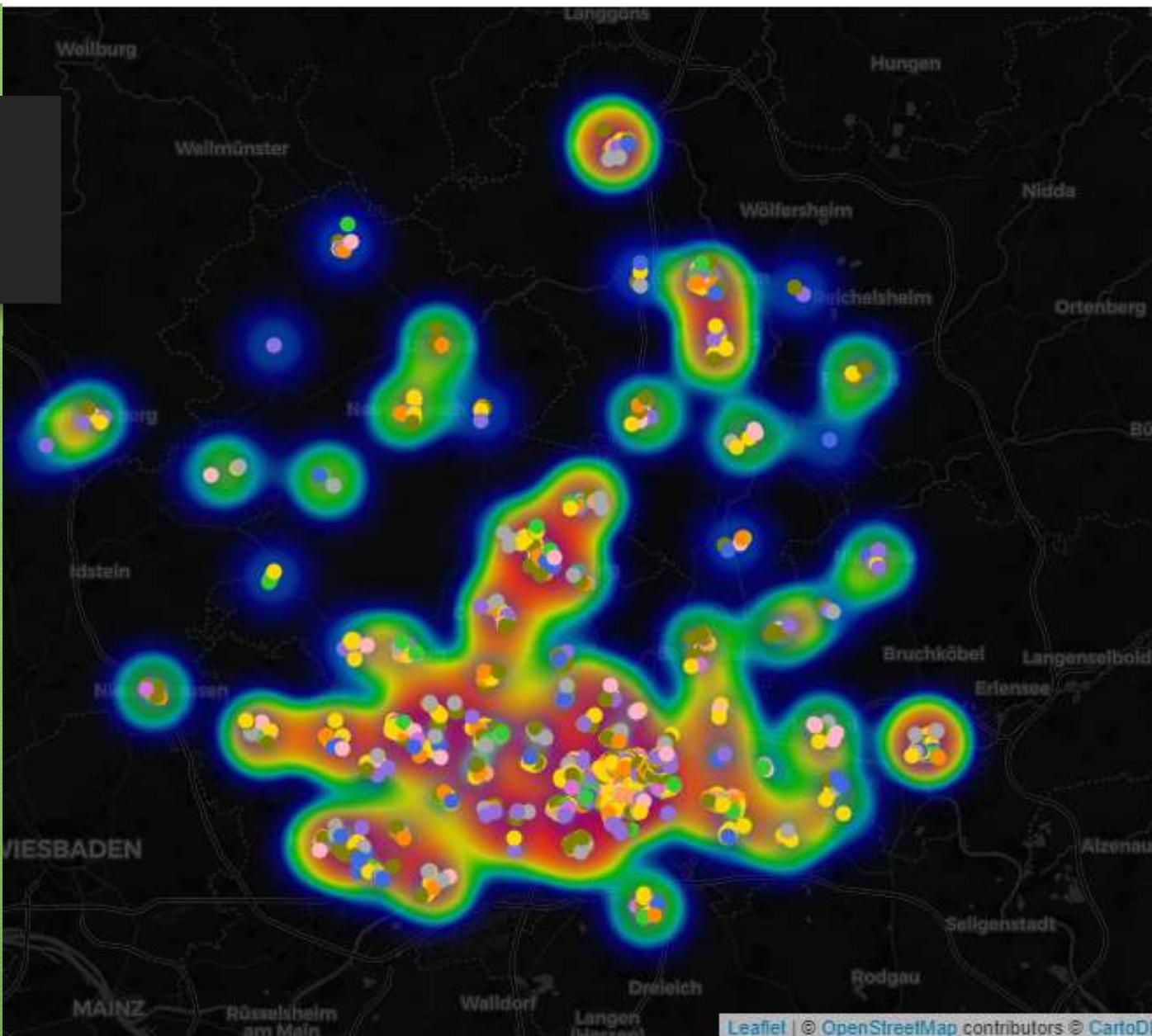
Amenities percentages and count

- We have many Food related amenities.
- We are poor in Entertainment and Arts amenities.
- The food category (formal plus informal) makes up almost 50% of all amenities.
- For fun the entertainment & arts, points of interest, outdoors & health and Nightlife Spots make up 19% of all amenities.



Mapping amenities with heatmap

- In this map you see the amenities density. The redder the area the more amenities available. As we can see the areas previously identified appear in bright red.



Preparing for K Means Clustering

1. Category onehot encoding
2. Row grouping by Area and by taking the mean of the frequency of occurrence of each category

Area	Accessories Store	African Restaurant	American Restaurant	Amphitheater	Apple Wine Pub	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Australian Restaurant	Austrian Restaurant
65520 - Bad Camberg	0	0	0	0	0	0	0	0	0	0	0	0	0
65520 - Bad Camberg	0	0	0	0	0	0	0	0	0	0	0	0	0
65520 - Bad Camberg	0	0	0	0	0	0	0	0	0	0	0	0	0
65520 - Bad Camberg	0	0	0	0	0	0	0	0	0	0	0	0	0
65520 - Bad Camberg	0	0	0	0	0	0	0	0	0	0	0	0	0

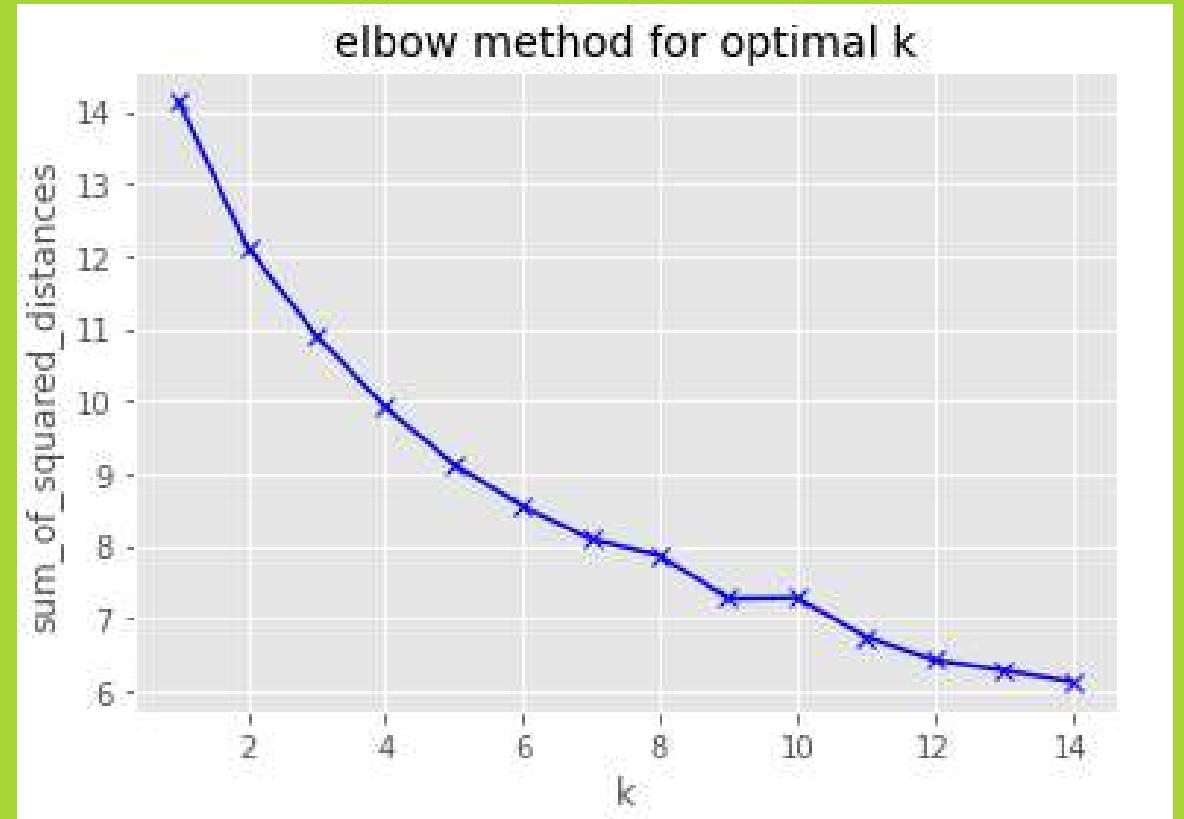
Area	Accessories Store	African Restaurant	American Restaurant	Amphitheater	Apple Wine Pub	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant
35510 - Butzbach	0.00	0.00	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000
60311 - Frankfurt am Main	0.00	0.00	0.012658	0.0	0.0	0.0	0.0	0.025316	0.0	0.000000
60313 - Frankfurt am Main	0.01	0.01	0.010000	0.0	0.0	0.0	0.0	0.000000	0.0	0.010000
60314 - Frankfurt am Main	0.00	0.00	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.024390
60316 - Frankfurt am Main	0.00	0.00	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.017857

Most common amenities in each area

Area	1st Most Common Amenity	2nd Most Common Amenity	3rd Most Common Amenity	4th Most Common Amenity	5th Most Common Amenity	6th Most Common Amenity	7th Most Common Amenity	8th Most Common Amenity	9th Most Common Amenity	10th Most Common Amenity
35510 - Butzbach	Supermarket	Drugstore	Construction & Landscaping	Shoe Store	Sandwich Place	Café	Turkish Restaurant	Greek Restaurant	Italian Restaurant	Indie Movie Theater
60311 - Frankfurt am Main	Café	German Restaurant	Plaza	Coffee Shop	Restaurant	Gym / Fitness Center	Scenic Lookout	Boutique	Italian Restaurant	Hotel
60313 - Frankfurt am Main	Café	Italian Restaurant	Restaurant	Bar	Thai Restaurant	Coffee Shop	Japanese Restaurant	Plaza	Sandwich Place	Doner Restaurant
60314 - Frankfurt am Main	Hotel	Gym / Fitness Center	Bakery	Escape Room	Supermarket	Italian Restaurant	Laser Tag	French Restaurant	Hawaiian Restaurant	Outdoor Supply Store
60316 - Frankfurt am Main	Café	Italian Restaurant	Burger Joint	Bakery	Thai Restaurant	Bar	Plaza	Modern European Restaurant	Supermarket	Japanese Restaurant
60318 - Frankfurt am Main	Café	Italian Restaurant	Supermarket	Ice Cream Shop	Pizza Place	Health Food Store	Bakery	Vietnamese Restaurant	Falafel Restaurant	Greek Restaurant
60320 - Frankfurt am Main	Pizza Place	Park	Bus Stop	Concert Hall	Metro Station	Market	Supermarket	Coffee Shop	Cafeteria	Drugstore
60322 - Frankfurt am Main	Italian Restaurant	Café	Bus Stop	Deli / Bodega	Falafel Restaurant	Thai Restaurant	Road	Coffee Shop	Supermarket	Lounge
60323 - Frankfurt am Main	Italian Restaurant	Bakery	Café	German Restaurant	Supermarket	Steakhouse	Thai Restaurant	Japanese Restaurant	Drugstore	Bistro

Evaluating optimal number of clusters - K

- From the graph we are going to choose 7 as the optimal number of clusters for our analysis. We had some doubts whether to choose 7 or 8, but the subsequent metrics analysis proved better for 7.

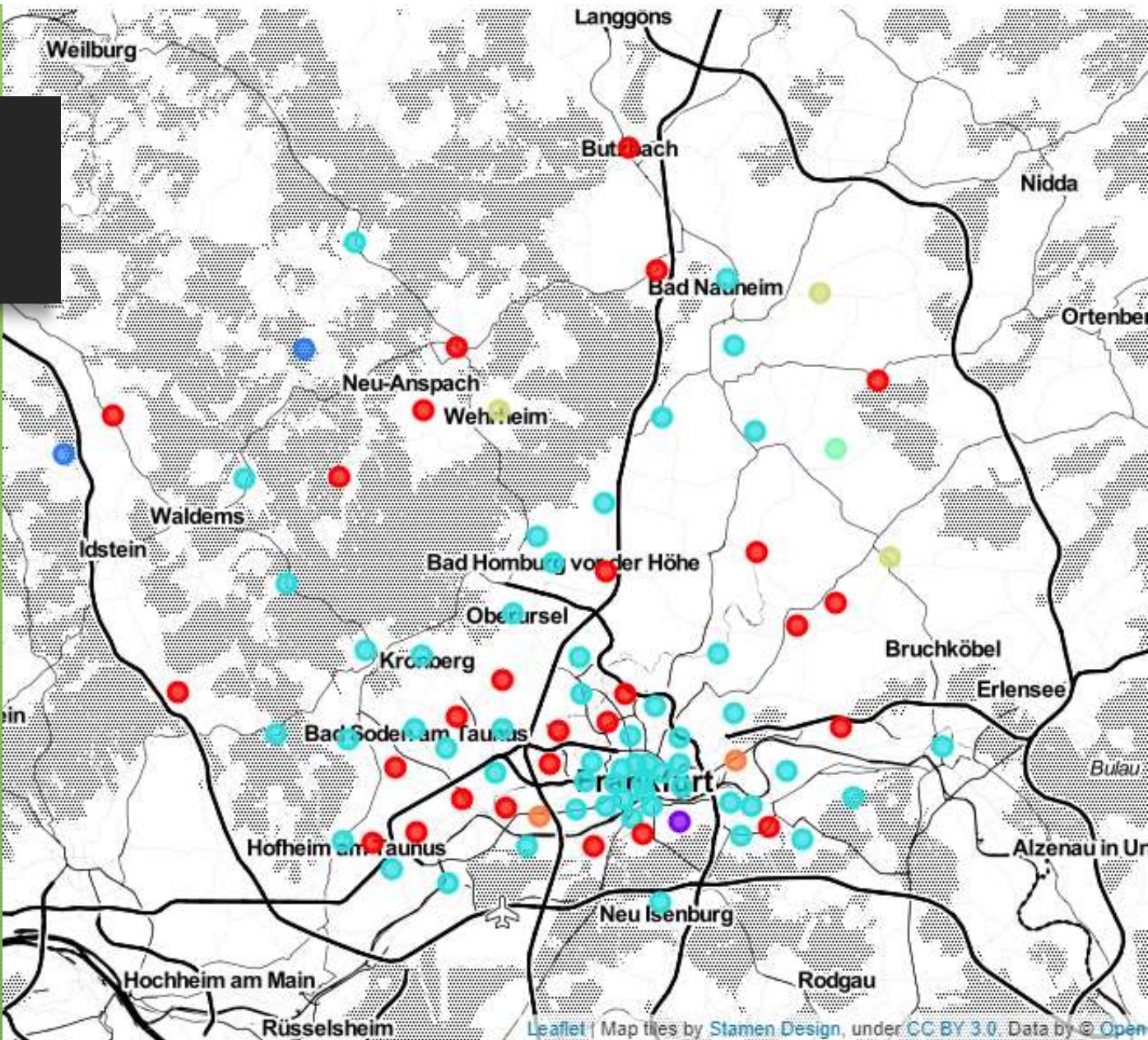


Areas - Cluster Labeled

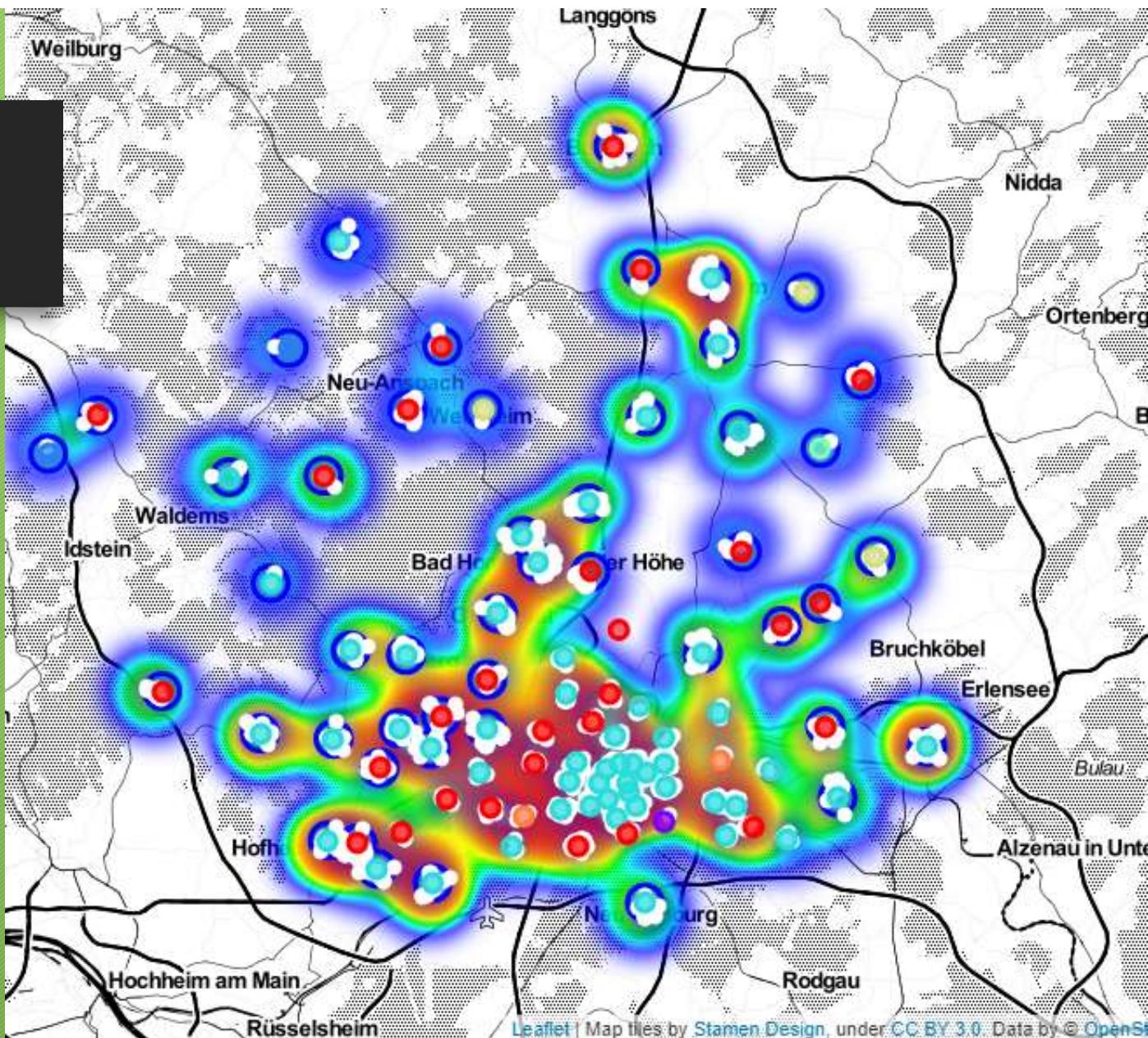
Area	Ortsteile	latitude	longitude	Cluster Labels	1st Most Common Amenity	2nd Most Common Amenity	3rd Most Common Amenity	4th Most Common Amenity	5th Most Common Amenity	6th Most Common Amenity	7th Most Common Amenity	8th Most Common Amenity
35510 - Butzbach	Butzbach	50.434042	8.671118	0	Supermarket	Drugstore	Construction & Landscaping	Shoe Store	Sandwich Place	Café	Turkish Restaurant	Greek Restaurant
60311 - Frankfurt am Main	Altstadt, Innenstadt	50.111320	8.683249	3	Café	German Restaurant	Plaza	Coffee Shop	Restaurant	Gym / Fitness Center	Scenic Lookout	Boutique
60313 - Frankfurt am Main	Altstadt, Innenstadt	50.115291	8.682501	3	Café	Italian Restaurant	Restaurant	Bar	Thai Restaurant	Coffee Shop	Japanese Restaurant	Plaza
60314 - Frankfurt am Main	Fechenheim, Ostend	50.113812	8.712702	3	Hotel	Gym / Fitness Center	Bakery	Escape Room	Supermarket	Italian Restaurant	Laser Tag	French Restaurant
60316 - Frankfurt am Main	Nordend-West, Nordend-Ost, Ostend	50.120735	8.696523	3	Café	Italian Restaurant	Burger Joint	Bakery	Thai Restaurant	Bar	Plaza	Modern European Restaurant

Cluster mapping

- Cluster 0 - red
- Cluster 1 - purple
- Cluster 2 - blue
- Cluster 3 - light blue
- Cluster 4 - green
- Cluster 5 - yellow
- Cluster 6 - orange



Putting it all together



Results

- From our analysis we can see that the areas from cluster 0 and 3 have the most amenities while the remaining clusters only have between 1 and 3 areas with very few amenities.
- Considering amenities density, cluster 0 areas, are near yellow zones and cluster 3 areas, are near red zones in the heatmap.
- Ranking the clusters by number of amenities available we get:
- **Cluster 3 with the most amenities of all and greater variety category wise. Food related amenities account for the most common amenities in all areas.**
- **Cluster 0 contains mostly shopping related amenities. Supermarket is the most common amenity in all clustered areas.**
- **The remaining clusters 1,2,4,5 and 6 contain areas with low amenities density and little variety.**
- Cluster 3 contains 1436 amenities, while cluster 0 has 211 amenities and the remaining 26 belong to the combined clusters 1,2,4,5,6.

1 - Cluster 3 Areas

	Area	Ortsteile
14	60313 - Frankfurt am Main	Altstadt, Innenstadt
13	60329 - Frankfurt am Main	Bahnhofsviertel, Gallus, Gutleutviertel, Innen...
16	60311 - Frankfurt am Main	Altstadt, Innenstadt
26	60594 - Frankfurt am Main	Sachsenhausen-Nord
48	60385 - Frankfurt am Main	Bornheim, Nordend-Ost, Ostend
29	60318 - Frankfurt am Main	Innenstadt, Nordend-Ost, Nordend-West
38	60316 - Frankfurt am Main	Nordend-West, Nordend-Ost, Ostend
57	63454 - Hanau	Hanau, Kesselstadt, Mittelbuchen
1	61348 - Bad Homburg vor der Höhe	Bad Homburg vor der Höhe
18	60323 - Frankfurt am Main	Westend-Nord, Westend-Süd
20	60314 - Frankfurt am Main	Fechenheim, Ostend
4	61231 - Bad Nauheim	Bad Nauheim
80	63065 - Offenbach am Main	Offenbach am Main
5	65812 - Bad Soden am Taunus	Bad Soden am Taunus
9	65760 - Eschborn	Eschborn
59	65719 - Hofheim am Taunus	Hofheim am Taunus
24	60487 - Frankfurt am Main	Bockenheim, Hausen, Rödelheim, Westend-Süd
71	63263 - Neu-Isenburg	Neu-Isenburg
77	61440 - Oberursel	Oberursel (Taunus)
6	61118 - Bad Vilbel	Bad Vilbel
17	60325 - Frankfurt am Main	Bockenheim, Gallus, Westend-Süd
49	60439 - Frankfurt am Main	Heddernheim, Kalbach-Riedberg, Niederursel, Pr...
81	63067 - Offenbach am Main	Offenbach am Main, Kaiserlei
62	65779 - Kelkheim	Kelkheim
66	61476 - Kronberg	Kronberg im Taunus
22	60322 - Frankfurt am Main	Dornbusch, Innenstadt, Nordend-West, Westend-N...
54	61381 - Friedrichsdorf	Friedrichsdorf
32	60486 - Frankfurt am Main	Bockenheim, Gallus, Rödelheim, Westend-Süd

64	61462 - Königstein	Königstein im Taunus
25	60320 - Frankfurt am Main	Dornbusch, Eckenheim, Nordend-West, Westend-No...
36	60326 - Frankfurt am Main	Gallus, Griesheim
58	65795 - Hattersheim	Hattersheim
19	60327 - Frankfurt am Main	Gallus, Gutleutviertel, Westend-Süd
21	60596 - Frankfurt am Main	Sachsenhausen-Nord
53	61169 - Friedberg (Hessen)	Friedberg (Hessen)
63	65451 - Kelsterbach	Kelsterbach
15	60438 - Frankfurt am Main	Kalbach-Riedberg
85	61191 - Rosbach v. d. Höhe	Rosbach v.d. Höhe
8	65817 - Eppstein	Eppstein
3	61350 - Bad Homburg vor der Höhe	Bad Homburg vor der Höhe
83	63069 - Offenbach am Main	Lauterborn, Offenbach am Main, Rosenhöhe
90	65843 - Sulzbach (Taunus)	Sulzbach (Taunus)
69	63165 - Mühlheim am Main	Mühlheim am Main
78	63073 - Offenbach am Main	Bieber
56	61279 - Grävenwiesbach	Grävenwiesbach
23	65936 - Frankfurt am Main	Nied, Sossenheim
95	61206 - Wöllstadt	Wöllstadt
43	60389 - Frankfurt am Main	Bergen-Enkheim, Bornheim, Nordend-Ost, Nordend...
46	60435 - Frankfurt am Main	Berkersheim, Bornheim, Dornbusch, Eckenheim, N...
55	61479 - Glashütten	Glashütten (Taunus)
79	63075 - Offenbach am Main	Bürgel, Offenbach am Main, Rumpenheim, Waldheim
92	65529 - Waldems	Waldems
34	60529 - Frankfurt am Main	Schwanheim
31	60388 - Frankfurt am Main	Bergen-Enkheim, Seckbach

2 - Cluster 0 Areas

	Area	Ortsteile
42	60528 - Frankfurt am Main	Niederrad, Sachsenhausen-Nord, Sachsenhausen-S...
7	35510 - Butzbach	Butzbach
65	65830 - Krifel	Krifel
70	61267 - Neu-Anspach	Neu-Anspach
27	60489 - Frankfurt am Main	Rödelheim
68	63477 - Maintal	Maintal
88	65824 - Schwalbach am Taunus	Schwalbach am Taunus
89	61449 - Steinbach	Steinbach (Taunus)
67	65835 - Liederbach am Taunus	Liederbach am Taunus
41	60431 - Frankfurt am Main	Bockenheim, Dornbusch, Eschersheim, Ginnheim, ...
2	61352 - Bad Homburg vor der Höhe	Bad Homburg vor der Höhe
45	65929 - Frankfurt am Main	Höchst, Unterliederbach
0	65520 - Bad Camberg	Bad Camberg
75	65527 - Niedernhausen	Niedernhausen
10	61197 - Florstadt	Florstadt
61	61184 - Karben	Karben
74	61138 - Niederdorfelden	Niederdorfelden
28	65934 - Frankfurt am Main	Höchst, Nied
35	65931 - Frankfurt am Main	Sindlingen, Zeilsheim
33	60488 - Frankfurt am Main	Hausesen, Praunheim, Rödelheim
87	61137 - Schöneck	Schöneck (Hessen)
86	61389 - Schmitten	Schmitten
44	60433 - Frankfurt am Main	Bonames, Dornbusch, Eschersheim, Frankfurter B...
30	60598 - Frankfurt am Main	Sachsenhausen-Nord, Sachsenhausen-Süd
76	61239 - Ober-Mörlen	Ober-Mörlen
82	63071 - Offenbach am Main	Offenbach am Main, Tempelsee
91	61250 - Usingen	Usingen

Clusters 1,2,4,5,6 Areas

	Area	Ortsteile
73	61130 - Nidderau	Nidderau
40	65933 - Frankfurt am Main	Griesheim
93	61273 - Wehrheim	Wehrheim
37	60386 - Frankfurt am Main	Bornheim, Fechenheim, Riederwald, Seckbach
84	61203 - Reichelsheim	Reichelsheim (Wetterau)
39	60599 - Frankfurt am Main	Oberrad, Sachsenhausen-Nord, Sachsenhausen-Süd
72	61194 - Niddatal	Niddatal
60	65510 - Idstein, Hünstetten	Idstein, Hünstetten
94	61276 - Weilrod	Weilrod

Conclusions

- This research aimed to provide information about areas to buy a house based on amenities availability. For that we used a machine learning algorithm - K means clustering - in a dataset comprised of the mean of the frequency of occurrence of each category in each area. As a result, we found the relevant clusters.
- For us it is clear, based on our study, that pertaining to amenities density **one should buy in one of the areas in cluster 3**.
- This, of course does not mean that you should be rushing to buy in any of those areas, further developing of this research is needed, this is only a starting point.
- These results matter as a start for a larger study, and also there is room for improvement in this study. We should try to expand our amenities dataset, by pinpointing optimal coordinates or getting other sources. We should also include new datasets like housing market, quality of living indexes, etc. to explore other avenues of research. We should also try other machine learning clustering algorithms and compare results.