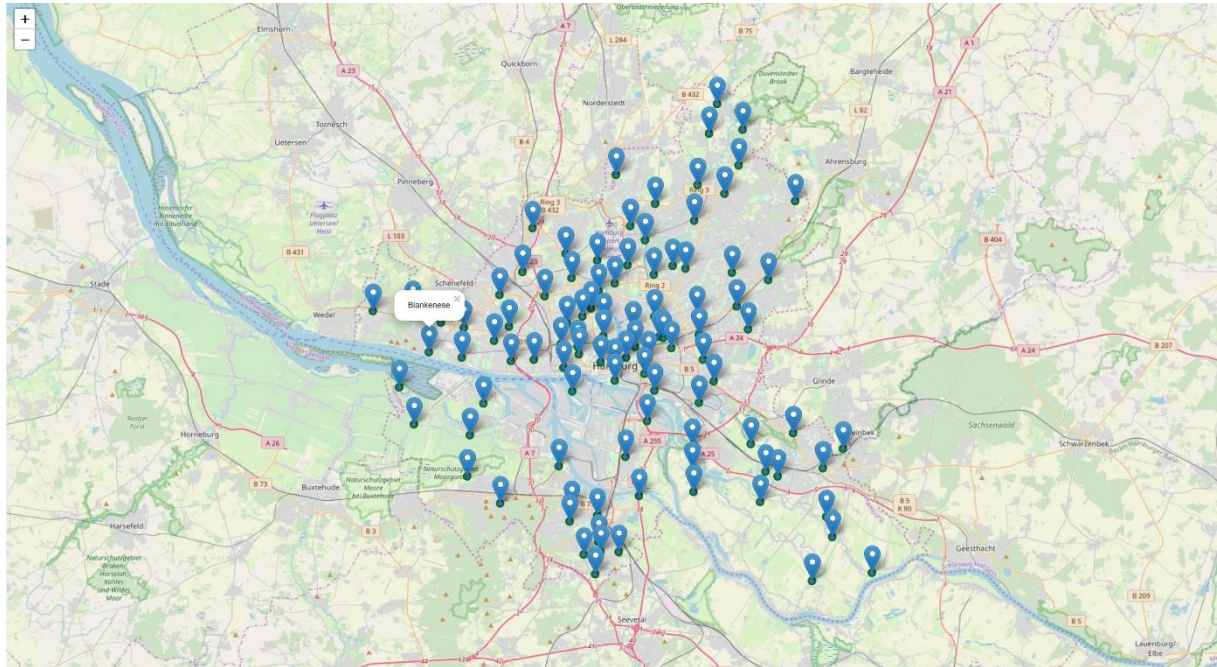# Finding the best districts of Hamburg to establish a new book-publishing house.

1. Introduction/Business Problem

As an aspiring entrepreneur and founder of a new book-publishing house, selecting the best location for the company headquarters is an important task. There are a lot of different criteria to keep in mind, if this endeavor will highly likely result in a successful business. For example, is the location conveniently situated to be easily reachable by train, car or other means for aspiring authors to visit the location or attractive enough for renown editors to relocate their main place of residence into this neighborhood? But of course typical rental prices in the different boroughs and spatial proximity to the center of a city are also pretty relevant. And having a lot of bookstores around could give lots of opportunities for additional marketing opportunities with selected bookstores and easily accessible readers frequenting the bookstores regularly and are potentially recruitable as voluntary book testers or even as long reaching influencers over the different channels of the internet. So in summary, considering all of the aspects and variables above, can we get insight by the means of a data analysis to make a better informed decision of where to establish a new book-publishing house in the city of Hamburg?

2. Discussion of the Data used in the Analysis

The analysis has been undertaken by combining data from three different sources. The first source is data about demographic, traffic and other urban related information concerning the 104 different districts of Hamburg. It can be found at

https://www.statistik-nord.de/fileadmin/Dokumente/Datenbanken_und_Karten/Stadtteilprofile/StadtteilprofileBerichtsjahr2017.xlsx

This data set contains variables and descriptions in German, but the necessary translations in English will be performed during this data analysis. We are especially interested in prices for real estates in the different districts.

[25]:

	Stadtteile	Bevölkerung	Unter 18-Jährige	65-Jährige und Ältere	Ausländerinnen und Ausländer	Bevölkerung mit Migrations-hintergrund	Unter 18-Jährige mit Migrations-hintergrund	Haushalte	Einpersonen-haushalte	Haushalte mit Kindern	Alleinerziehende	Bevölkerungs-dichte	Geburten	Sterbefälle	Zuzüge	Fortzüge
0	Hamburg-Altstadt	2 305	277	256	506	990	182	1 508	939	187	47	1 773	46	8	344	343
1	HafenCity	3 627	756	333	1 168	1 718	512	1 745	707	416	62	1 495	68	9	903	665
2	Neustadt	12 719	1 456	1 836	2 580	4 670	732	8 755	6 042	1 006	287	5 623	152	70	1 819	1 813
3	St. Pauli	22 501	2 991	2 150	4 880	8 309	1 435	15 163	10 462	1 974	638	10 034	239	150	2 933	2 823
4	St. Georg	11 055	1 108	1 397	2 552	4 274	588	7 207	4 892	732	179	6 065	116	90	2 005	1 770
5	Hammerbrook	3 939	448	133	1 812	2 470	379	2 009	1 314	201	55	1 578	61	8	1 356	1 615
6	Borgfelde	7 635	834	923	2 201	3 403	588	4 462	3 065	441	125	9 290	72	54	1 631	1 530
7	Hamm	38 716	4 453	5 398	7 037	13 792	2 795	26 171	17 808	2 871	888	9 956	487	331	5 488	5 066
8	Horn	38 569	6 001	6 039	9 906	19 729	4 474	22 763	13 535	3 596	1 188	6 540	439	417	4 622	4 587
9	Billstedt	70 539	13 653	12 646	17 768	40 525	10 337	35 135	16 421	7 814	2 378	4 168	873	694	5 825	5 470
10	Billbrook	2 097	650	116	1 635	1 806	632	628	464	68	18	333	40	10	645	743
11	Rothenburgsort	9 225	1 581	1 344	2 593	5 147	1 242	5 275	3 150	892	295	1 283	103	101	1 107	1 033
12	Veddel	4 667	925	388	2 106	3 374	848	2 510	1 433	520	146	1 068	52	15	609	583
13	Wilhelmsburg	54 169	11 236	7 294	17 593	32 447	8 767	26 250	12 879	6 068	1 604	1 531	743	475	5 229	6 176
14	Kleiner Grasbrook und Steinwerder	1 243	70	67	464	615	61	973	759	38	17	104	7	7	383	360
15	Waltershof und Finkenwerder	11 765	1 896	2 316	1 985	3 511	892	6 341	3 222	1 168	316	400	121	119	870	784
16	Altona-Altstadt	29 408	4 486	3 783	5 757	11 150	2 346	18 059	11 314	2 928	910	10 820	390	211	3 453	3 339
17	Sternschanze	8 080	1 146	828	1 612	2 962	587	5 250	3 542	759	232	14 732	97	110	1 254	1 221
18	Altona-Nord	22 946	3 647	2 297	4 387	8 365	1 875	13 954	8 752	2 299	651	10 346	331	113	3 371	2 901
19	Ottensen	35 480	5 585	4 891	4 577	9 342	1 933	21 616	13 264	3 745	1 026	12 693	457	214	3 801	3 916

The second data set is from Wikipedia.
https://de.wikipedia.org/wiki/Liste_der_Bezirke_und_Stadtteile_Hamburgs

It gives access to the geographical coordinates of the different districts of Hamburg. This information can be combined with the aforementioned data set to get quantitative information about how close a neighborhood is located in relation to the center of the city. This information also builds the backbone for the access of our third data set.
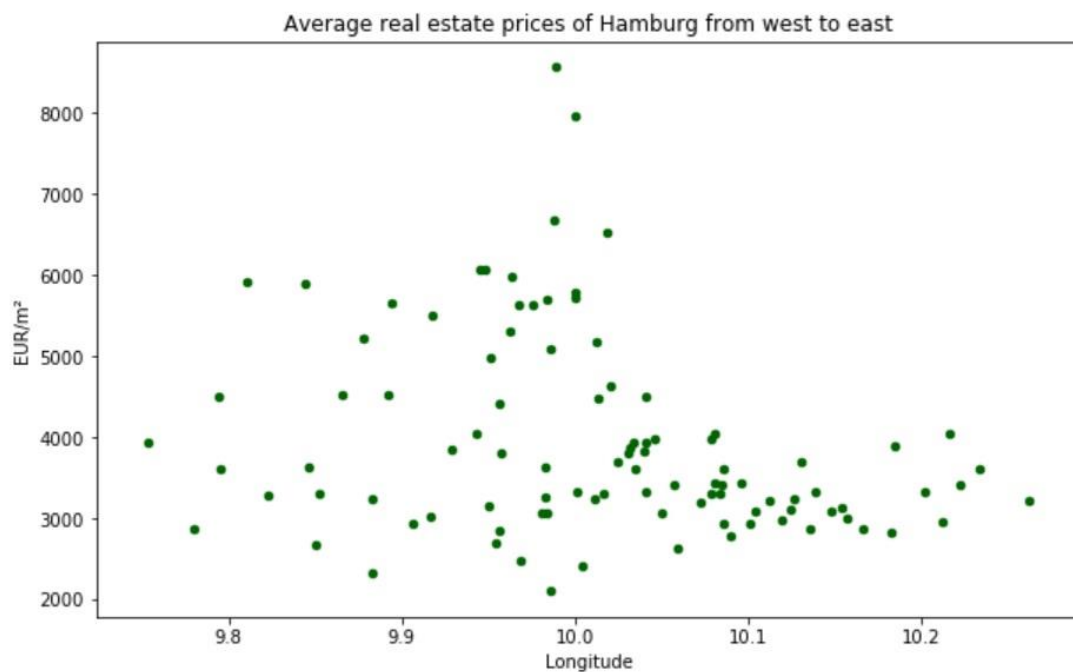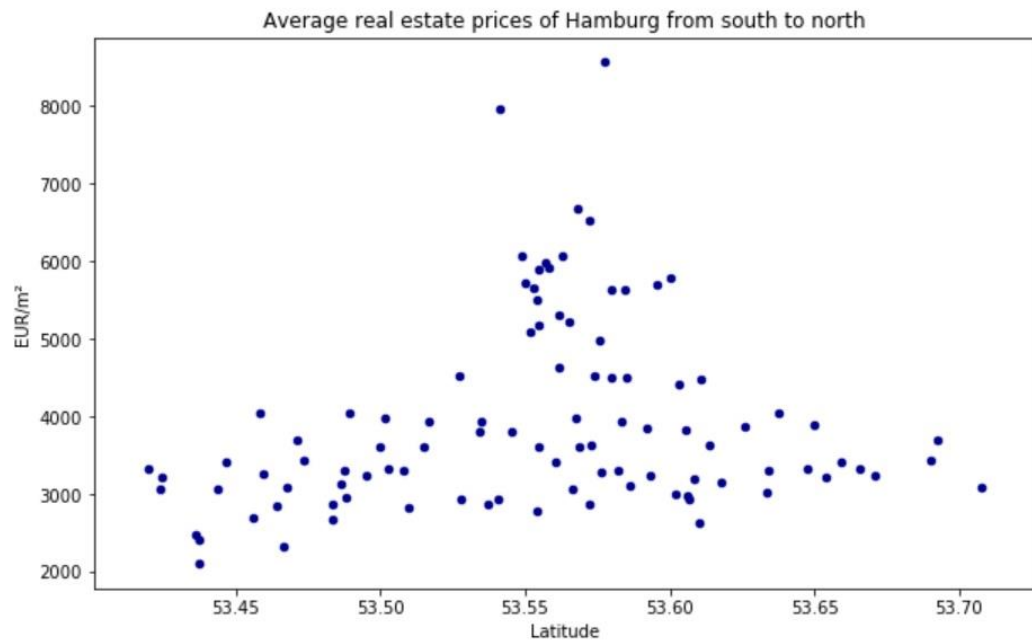
	Stadtteil	Ortsteile	Bezirk	Fläche(km²)	Einwohner	Bevölkerungsdichte(Einwohner/km²)	Koordinaten	Karte
0	Hamburg-Altstadt	NaN	Hamburg-Mitte	NaN	2272.000	947	53° 33′ 0″ N, 10° 0′ 0″ O	NaN
1	HafenCity	NaN	Hamburg-Mitte	NaN	4592.000	2087	53° 32′ 28″ N, 10° 0′ 1″ O	NaN
2	Neustadt	NaN	Hamburg-Mitte	NaN	12.920	5617	53° 33′ 7″ N, 9° 59′ 8″ O	NaN
3	St. Pauli	NaN	Hamburg-Mitte	NaN	22.436	8974	53° 33′ 25″ N, 9° 57′ 50″ O	NaN
4	St. Georg	NaN	Hamburg-Mitte	NaN	11.384	4743	53° 33′ 18″ N, 10° 0′ 44″ O	NaN
5	Hammerbrook	NaN	Hamburg-Mitte	NaN	4323.000	1441	53° 32′ 43″ N, 10° 1′ 50″ O	NaN
6	Borgfelde	NaN	Hamburg-Mitte	NaN	7696.000	9620	53° 33′ 17″ N, 10° 2′ 4″ O	NaN
7	Hamm	NaN	Hamburg-Mitte	NaN	38.773	9693	53° 33′ 39″ N, 10° 3′ 28″ O	NaN
8	Horn	NaN	Hamburg-Mitte	NaN	38.799	6576	53° 33′ 14″ N, 10° 5′ 24″ O	NaN
9	Billstedt	NaN	Hamburg-Mitte	NaN	70.355	4139	53° 32′ 26″ N, 10° 6′ 4″ O	NaN
10	Billbrook	NaN	Hamburg-Mitte	NaN	1992.000	327	53° 31′ 40″ N, 10° 5′ 10″ O	NaN
11	Rothenburgsort	NaN	Hamburg-Mitte	NaN	9268.000	1252	53° 32′ 6″ N, 10° 2′ 27″ O	NaN
12	Veddel	NaN	Hamburg-Mitte	NaN	4632.000	1053	53° 31′ 0″ N, 10° 2′ 0″ O	NaN
13	Wilhelmsburg	NaN	Hamburg-Mitte	NaN	54.068	1527	53° 29′ 42″ N, 10° 0′ 40″ O	NaN
14	Kleiner Grasbrook	NaN	Hamburg-Mitte	NaN	1200.000	267	53° 31′ 52″ N, 9° 59′ 37″ O	NaN
15	Steinwerder	NaN	Hamburg-Mitte	NaN	39.000	5	53° 32′ 4″ N, 9° 57′ 26″ O	NaN
16	Waltershof	NaN	Hamburg-Mitte	NaN	0.000	0	53° 31′ 40″ N, 9° 54′ 2″ O	NaN
17	Finkenwerder	NaN	Hamburg-Mitte	NaN	11.808	599	53° 31′ 38″ N, 9° 51′ 58″ O	NaN
18	Neuwerk	NaN	Hamburg-Mitte	NaN	40.000	5	53° 55′ 16″ N, 8° 30′ 2″ O	NaN
19	Altona-Altstadt	NaN	Altona	NaN	29.305	10466	53° 32′ 56″ N, 9° 56′ 52″ O	NaN
20	Sternschanze	NaN	Altona	NaN	8095.000	16190	53° 33′ 42″ N, 9° 57′ 44″ O	NaN
21	Altona-Nord	NaN	Altona	NaN	24.153	10979	53° 33′ 47″ N, 9° 56′ 43″ O	NaN
22	Ottensen	NaN	Altona	NaN	35.585	12709	53° 33′ 14″ N, 9° 55′ 4″ O	NaN
23	Bahrenfeld	NaN	Altona	NaN	29.976	2828	53° 34′ 26″ N, 9° 53′ 33″ O	NaN
24	Groß Flottbek	NaN	Altona	NaN	11.086	4619	53° 33′ 55″ N, 9° 52′ 39″ O	NaN
25	Othmarschen	NaN	Altona	NaN	15.591	2599	53° 33′ 10″ N, 9° 53′ 40″ O	NaN

The final data source is the Foursquare database found at foursquare.com. Here we get information about the availability and quality of the bookstores in the vicinity and also about the quality of the transport connections via train and other kinds of traffic to this area of the city.
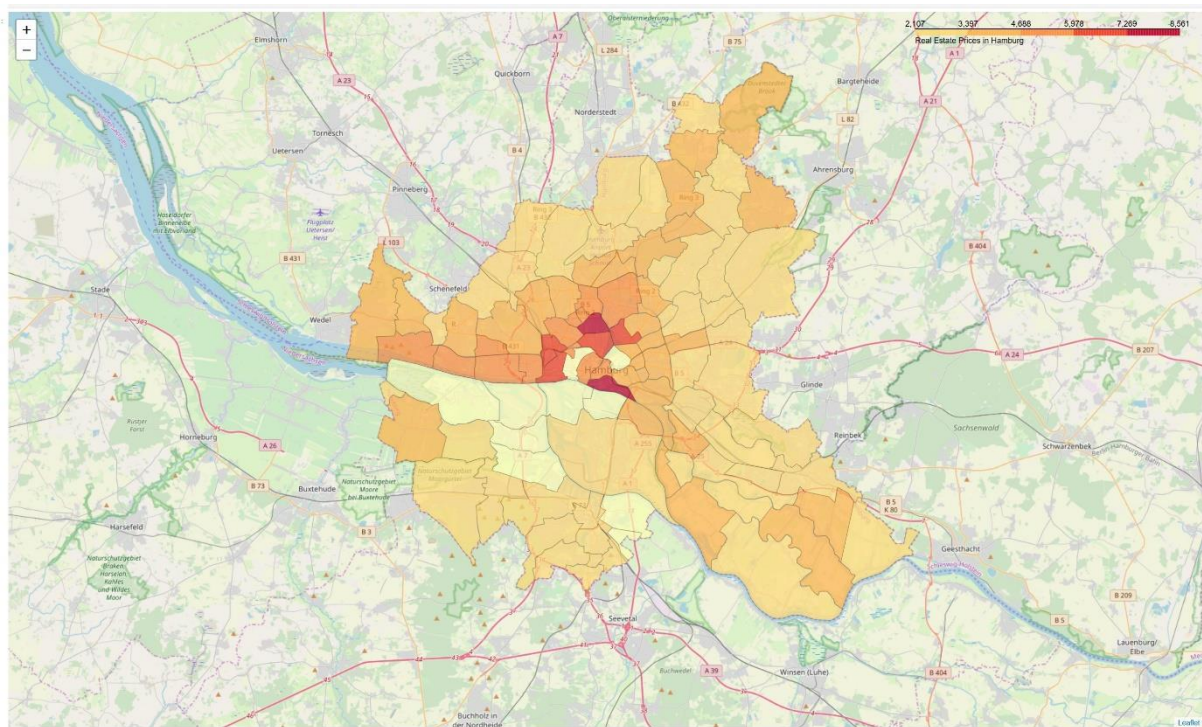
3. Methodology – Exploratory Data Analysis & Clustering

Because we are mainly interested in the pricing components of our demographic data set we use this data to analyze, if there are any differences between different regions from south to north and from west to east?

Average real estate prices of Hamburg from south to north

Average real estate prices of Hamburg from west to east

There is evidently a correlation towards the western districts of Hamburg. And not surprising the central districts lead the ranking of most expensive real estate prices. This is especially good to see on a choropleth map with the real estate costs colored per district from yellow to red.

What we are now interested in is to find the bookstores per districts and to build a sum of how many there are to find per district. This shows what areas are especially useful for regional marketing campaigns.

	District	District Latitude	District Longitude	Bookstore	Bookstore Latitude	Bookstore Longitude
0	Hamburg-Altstadt	53.550000	10.000000	Bücher Koje	53.550427	10.001408
1	Hamburg-Altstadt	53.550000	10.000000	Murmann Publishers	53.548762	9.999167
2	Hamburg-Altstadt	53.550000	10.000000	Michelle Records	53.552378	10.000935
3	Hamburg-Altstadt	53.550000	10.000000	Thalia	53.552254	10.003478
4	Hamburg-Altstadt	53.550000	10.000000	Thalia	53.551386	9.996020
5	Hamburg-Altstadt	53.550000	10.000000	Boysen+Maasch	53.553431	10.000236
6	Hamburg-Altstadt	53.550000	10.000000	Dr. Götze Land & Karte	53.552708	9.997119
7	Hamburg-Altstadt	53.550000	10.000000	stories!	53.553359	9.988813
8	Hamburg-Altstadt	53.550000	10.000000	Lehmanns Fachbuchhandlung	53.552979	10.003275
9	Hamburg-Altstadt	53.550000	10.000000	Boysen & Mauke	53.549319	9.992606
10	Hamburg-Altstadt	53.550000	10.000000	Felix Jud & Co.KG Buchhandlung	53.552471	9.992102
11	Hamburg-Altstadt	53.550000	10.000000	Buchhandlung im Haus der Photographie	53.546719	10.006457
12	Hamburg-Altstadt	53.550000	10.000000	SPIEGEL-Shop	53.545706	10.002961
13	Hamburg-Altstadt	53.550000	10.000000	Fachbuchhandlung Wede	53.543310	9.999812
14	Hamburg-Altstadt	53.550000	10.000000	TASCHEN store Hamburg	53.551492	9.988654
15	Hamburg-Altstadt	53.550000	10.000000	Buchhandlung Walter König	53.551739	10.009727
16	Hamburg-Altstadt	53.550000	10.000000	P&B Press and Books	53.553272	10.005850
17	Hamburg-Altstadt	53.550000	10.000000	Buchhandlung und Antiquariat Dr. Robert Wohler...	53.556609	10.010148
18	Hamburg-Altstadt	53.550000	10.000000	Kunstantiquariat Joachim Lührs	53.549449	9.984963
19	Hamburg-Altstadt	53.550000	10.000000	st. Katharinen	53.545969	9.993911
20	Hamburg-Altstadt	53.550000	10.000000	Zeilenreich	53.550246	9.992986
21	Hamburg-Altstadt	53.550000	10.000000	k presse + buch	53.553243	10.005821
22	Hamburg-Altstadt	53.550000	10.000000	Zweitausendeins	53.555297	9.990689
23	Hamburg-Altstadt	53.550000	10.000000	Buchhandlung Antiquariat Bernhardt	53.549498	10.005159
24	Hamburg-Altstadt	53.550000	10.000000	Colón Bookshop	53.556744	9.989690
25	Hamburg-Altstadt	53.550000	10.000000	Wiedebusch	53.558020	9.988352
26	Hamburg-Altstadt	53.550000	10.000000	Buchhandlung Laatzen	53.558015	9.989727
27	Hamburg-Altstadt	53.550000	10.000000	Sautter + Lackmann	53.548779	9.985005
28	HafenCity	53.541111	10.000278	Fachbuchhandlung Wede	53.543310	9.999812
29	HafenCity	53.541111	10.000278	Boysen & Mauke	53.549319	9.992606
30	HafenCity	53.541111	10.000278	SPIEGEL-Shop	53.545706	10.002961
31	HafenCity	53.541111	10.000278	Buchhandlung im Haus der Photographie	53.546719	10.006457
32	HafenCity	53.541111	10.000278	Bücher Koje	53.550427	10.001408
33	HafenCity	53.541111	10.000278	Murmann Publishers	53.548762	9.999167

We can summarize this expansive data set via table of the total counts of book stores per district, which finally will become a part of our machine learning data set.
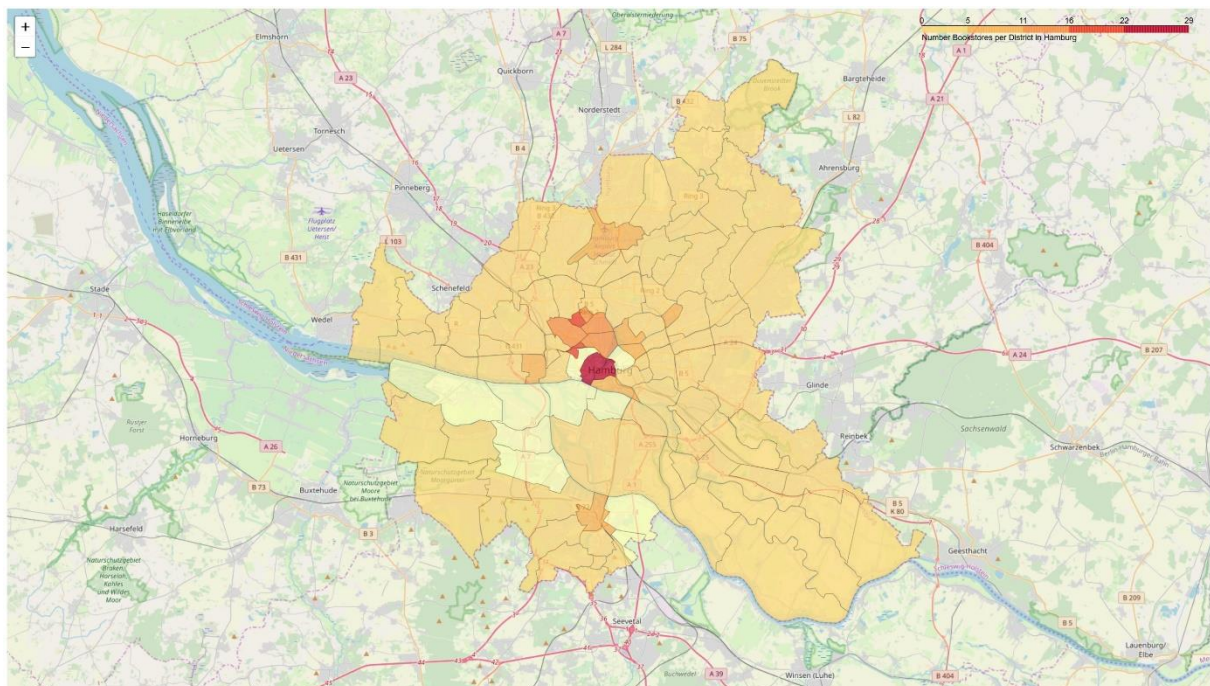
	District	Bookstores
0	Hamburg-Altstadt	28
1	HafenCity	10
2	Neustadt	28
3	St. Pauli	15
4	St. Georg	20
5	Hammerbrook	1
6	Borgfelde	3
7	Hamm	1
8	Billstedt	2
9	Waltershof und Finkenwerder	1
10	Altona-Altstadt	4
11	Sternschanze	18
12	Altona-Nord	4
13	Ottensen	8
14	Groß Flottbek	1
15	Othmarschen	1
16	Osdorf	3
17	Blankenese	4
18	Rissen	1
19	Eimsbüttel	11
20	Rotherbaum	15
21	Harvestehude	12
22	Hoheluft-West	16
23	Niendorf	1
24	Eidelstedt	3
25	Stellingen	1
26	Hoheluft-Ost	12
27	Eppendorf	4

We also calculate the euclidean distance from each district in relation to the center to have a better grasp of how influential a more distant neighborhood from the center region could be for our future business in relation to the benefits of reduced costs for the office or maybe more office space for the same price.

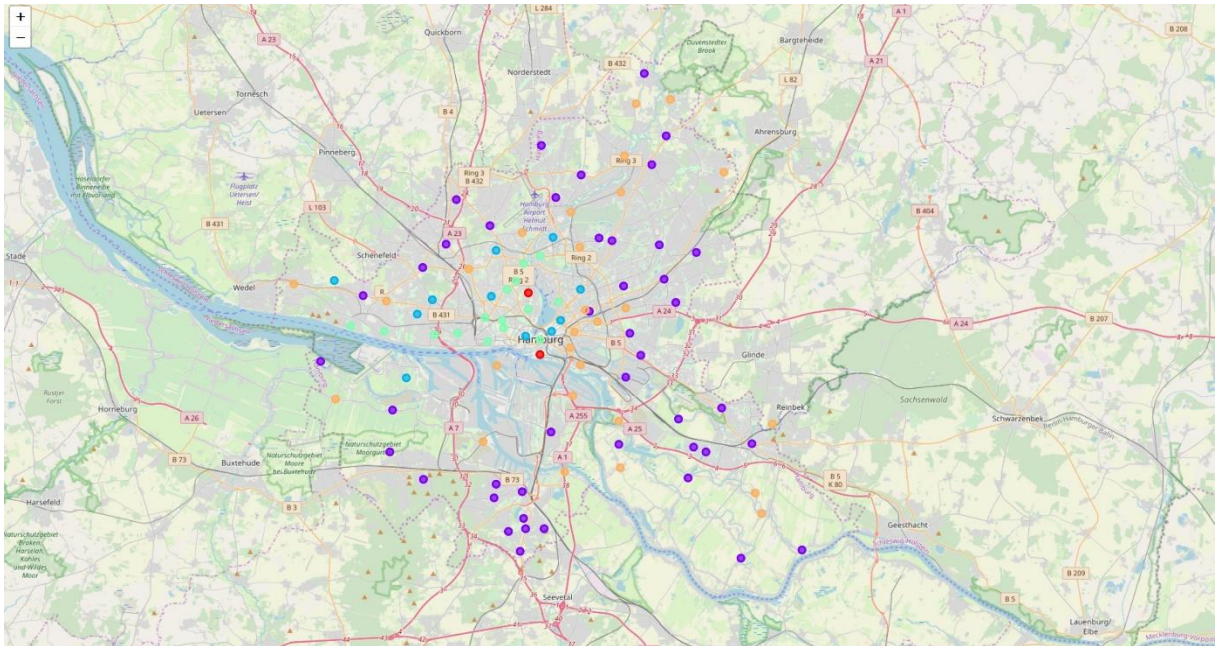|	District	Prices	Latitude	Longitude	Bookstores	Distance
0	Hamburg-Altstadt	5710	53.550000	10.000000	28.0	0.000000
1	HafenCity	7958	53.541111	10.000278	10.0	0.008893
2	Neustadt	5081	53.551944	9.985556	28.0	0.014574
3	St. Pauli	5991	53.556944	9.963889	15.0	0.036773
4	St. Georg	5169	53.555000	10.012222	20.0	0.013205
5	Hammerbrook	3814	53.545278	10.030556	1.0	0.030919
6	Borgfelde	3614	53.554722	10.034444	3.0	0.034766
7	Hamm	3414	53.560833	10.057778	1.0	0.058785
8	Horn	2789	53.553889	10.090000	0.0	0.090084
9	Billstedt	2932	53.540556	10.101111	2.0	0.101551
10	Billbrook	2932	53.527778	10.086111	0.0	0.088932
11	Rothenburgsort	3937	53.535000	10.040833	0.0	0.043501

We add another choropleth map with the number of book stores per district, not surprisingly centering around the central districts of Hamburg.

Finally we take our data set with all the real estate prices, number of book stores per district and distance to the center and create five clusters per machine learning. Because there is most likely not a single best solution for our problem of where to establish the main office space for our new book-publishing house, it makes sense to gather several comparatively similar solutions in the same clusters. It makes the most sense to compare the clusters visually on a map with an excerpt of the clustering result in our data set.

	District	Prices	Latitude	Longitude	Bookstores	Distance	Cluster Labels
0	Hamburg-Altstadt	5710	53.550000	10.000000	28.0	0.000000	3
1	HafenCity	7958	53.541111	10.000278	10.0	0.008893	0
2	Neustadt	5081	53.551944	9.985556	28.0	0.014574	2
3	St. Pauli	5991	53.556944	9.963889	15.0	0.036773	3
4	St. Georg	5169	53.555000	10.012222	20.0	0.013205	2
5	Hammerbrook	3814	53.545278	10.030556	1.0	0.030919	4
6	Borgfelde	3614	53.554722	10.034444	3.0	0.034766	4
7	Hamm	3414	53.560833	10.057778	1.0	0.058785	4
8	Horn	2789	53.553889	10.090000	0.0	0.090084	1
9	Billstedt	2932	53.540556	10.101111	2.0	0.101551	1
10	Billbrook	2932	53.527778	10.086111	0.0	0.088932	1
11	Rothenburgsort	3937	53.535000	10.040833	0.0	0.043501	4
12	Veddel	3937	53.516667	10.033333	0.0	0.047140	4
13	Wilhelmsburg	3249	53.495000	10.011111	0.0	0.056111	1
14	Kleiner Grasbrook und Steinwerder	3814	53.534444	9.957222	0.0	0.045519	4
15	Waltershof und Finkenwerder	4514	53.527222	9.866111	1.0	0.135813	2
16	Altona-Altstadt	6064	53.548889	9.947778	4.0	0.052234	3
17	Sternschanze	5302	53.561667	9.962222	18.0	0.039539	3
18	Altona-Nord	6064	53.563056	9.945278	4.0	0.056258	3
19	Ottensen	5512	53.553889	9.917778	8.0	0.082314	3
20	Bahrenfeld	4514	53.573889	9.892500	0.0	0.110122	2
21	Groß Flottbek	5211	53.565278	9.877500	1.0	0.123449	2
22	Othmarschen	5646	53.552778	9.894444	1.0	0.105593	3
23	Lurup	3239	53.593056	9.882778	0.0	0.124879	1
24	Osdorf	3621	53.572778	9.846667	3.0	0.155016	4
25	Nienstedten	5886	53.555000	9.844722	0.0	0.155358	3

4. Results

As we can see from this data analysis project, there are several interesting options to choose from for the new location of our publishing house. As is often the case in the realm of statistics and data analysis there are no best or worst solutions. It really depends on where the founders of the publishing house see the most benefit for themselves and their future plans. So the districts in one cluster may be the preferred solution in one scenario, and the districts in another cluster in the next. This is the beauty of clustering as one tool in the box of every data scientist.

5. Discussion

The blue, green and red cluster that we can see on the map of our clustering analysis are very interesting. There seems to be no clear path on where to steer with the results of this analysis. I see this as an opportunity to go more deeply in the cluster analysis, changing the number of clusters, the starting centroids and the strength of the variables that we use in our analysis.

6. Conclusion

This report was an interesting opportunity to work with a real data set and try to come up with a solution for a realistic business problem. Having done all these different aspects of a single comprehensive data analysis projects gives a good starting point for future data analysis projects. My main insight is, that there can be a lot of interesting aspects surrounding a maybe simple business question, that are worth to really dig deep into over the course of several days.