FINDING SUITABLE NEIGHBOURHOODS FOR OPENING UP A RESTAURANT IN HAMBURG, GERMANY

1. INTRODUCTION

1.1 PROBLEM DESCRIPTION

The pizza company "Fast Pizza" is successfully operating a restaurant in the neighbourhood of Eimsbüttel in Hamburg, Germany. They would like to expand and open up a new branch in a different neighbourhood. The objective of the project is to find a neighbourhood in Hamburg that is similar to Eimsbüttel in terms of demographics and food preferences.

1.2 INTEREST

The target audience for the analysis is the owner of "Fast Pizza". The analysis could also be useful for any other restaurant owner wishing to expand in Hamburg.

2. DATA ACQUISITION AND CLEANING

2.1 DATA SOURCES

- Demographic data for Hamburg neighbourhoods: https://www.statistik-nord.de/fileadmin/Dokumente/Datenbanken_und_Karten/Stadtteilprofile/StadtteilprofileBerichtsjahr2017.xlsx
- Locations of Hamburg Neighbourhoods: https://de.wikipedia.org/wiki/Liste_der_Bezirke_und_Stadtteile_Hamburgs
- Foursquare data on restaurants in Hamburg

2.2 USAGE OF DATA

Demographic data is used to cluster neighbourhoods by similarity in terms of population. Demographic data includes a variety of features such as age profile, average income, and unemployment rate. K means algorithm will be used to perform the clustering.
Locations data (latitude and longitude of neighbourhoods) is used to be able to plot neighbourhoods on a map and to query Foursquare API.
Data from Foursquare is used to cluster neighbourhoods based on restaurant preferences. Foursquare API is able to pull nearby popular venues for a given location. Restaurants are categorized by type of cuisine which allows us to understand food tastes in a given area. K means algorithm is used to perform clustering.
Finally, the results of demographic clustering and restaurant preference clustering is combined to give a recommendation as to where to open a new branch of the pizza restaurant.

2.3 DATA EXPLORATION

2.3.1 Demographic data set

Extract showing statistics for demographic dataset:

	cluster	minors %	elderly %	foreigners %	migration background %	people per household	single households %	single parent households %	population density	regular employment %	unemployment %	social security receivers %
count	99.000000	99.000000	99.000000	99.000000	99.000000	99.000000	99.000000	99.000000	99.000000	99.000000	99.000000	99.000000
mean	1.838384	16.847481	17.891426	17.542484	33.119345	1.883936	50.774932	23.624371	4298.901178	56.186869	4.631396	9.667386
std	1.952011	3.942186	5.511490	11.791703	15.209527	0.265083	12.015794	6.122686	4312.936187	6.157203	2.113647	8.048910
min	0.000000	5.631537	3.376491	3.112840	6.862745	1.304214	28.506098	9.523810	44.675489	29.700000	1.008471	0.697961
25%	0.000000	14.442278	14.304833	10.813988	22.272739	1.658282	40.879690	19.642945	1304.682186	52.550000	2.958236	3.639424
50%	1.000000	16.943765	18.874326	14.666219	31.189781	1.887123	48.589146	24.351747	2956.084047	57.000000	4.637484	8.155533
75%	3.000000	19.376766	20.775180	20.940954	41.032284	2.086830	61.440202	27.323775	5922.853761	60.300000	5.747283	12.307405
max	6.000000	30.996662	33.067010	77.968526	85.188679	2.733994	78.006166	44.736842	18536.000973	66.900000	11.686494	49.418605

Standard deviation for some features such as population density was very high. For the other features as well, discrepancies in between neighbourhoods are significant.

Eimsbüttel population includes less minors and less elderly people than the average Hamburg neighbourhood. It is mainly populated by students and professionals. Unemployment is low, but income is slightly below average. Population density is very high.

neighbourhood	minors %	elderly %	foreigners %	migration background %	people per household	single households %	single parent households %	population density	regular employment %	unemployment %	social security receivers %
Eimsbüttel	12.832589	12.606663	12.02447	23.501207	1.527631	67.354742	25.409674	17839.135373	62.1	3.773973	4.480284

2.4 DATA CLEANING

2.4.1 Location data:

Location data was parsed from Wikipedia using BeautifulSoup:

	neighbourhood	quarter	borough	area	inhabitants	population density	coordinates	map
0	Hamburg-Altstadt!Hamburg-Altstadt	101!101–102	Hamburg-Mitte	02,4!2,4	2305	960	53° 33′ 0″ N, 10° 0′ 0″ O	
1	HafenCity!HafenCity	103!103–104	Hamburg-Mitte	02,2!2,2	3627	1649	53° 32′ 28″ N, 10° 0′ 1″ O	
2	Neustadt!Neustadt	105!105–108	Hamburg-Mitte	02,3!2,3	12.719	5530	53° 33′ 7″ N, 9° 59′ 8″ O	
3	St.Pauli!St. Pauli	109!109–112	Hamburg-Mitte	02,5!2,5	22.501	9000	53° 33′ 25″ N, 9° 57′ 50″ O	
4	St.Georg!St. Georg	113!113–114	Hamburg-Mitte	02,4!2,4	11.055	4606	53° 33′ 18″ N, 10° 0′ 44″ O	

Final dataset: The neighbourhood name and coordinates had to be transformed to a usable format, some of the other information was dropped.

	neighbourhood	borough	latitude	longitude
0	Hamburg-Altstadt	Hamburg-Mitte	53.550000	10.000000
1	HafenCity	Hamburg-Mitte	53.541111	10.000278
2	Neustadt	Hamburg-Mitte	53.551944	9.985556
3	St. Pauli	Hamburg-Mitte	53.556944	9.963889
4	St. Georg	Hamburg-Mitte	53.555000	10.012222

In the demographic set some sparsely populated neighbourhoods were combined (e.g. Waltershof and Finkenwerder). In order to assign a location to these neighbouhoods, mean values for latitude and longitude were calculated.

2.4.2 Demographic data:

Originally there were 66 features for each of the 99 neighbourhoods. However, there was some redundancy in the features. Duplicate and very similar features were dropped from the start. Absolute values were transferred into percentage values (e.g. number of doctors per inhabitant instead of total number of doctors. The data set had German labels, so they needed to be translated to English.

3. METHODOLOGY

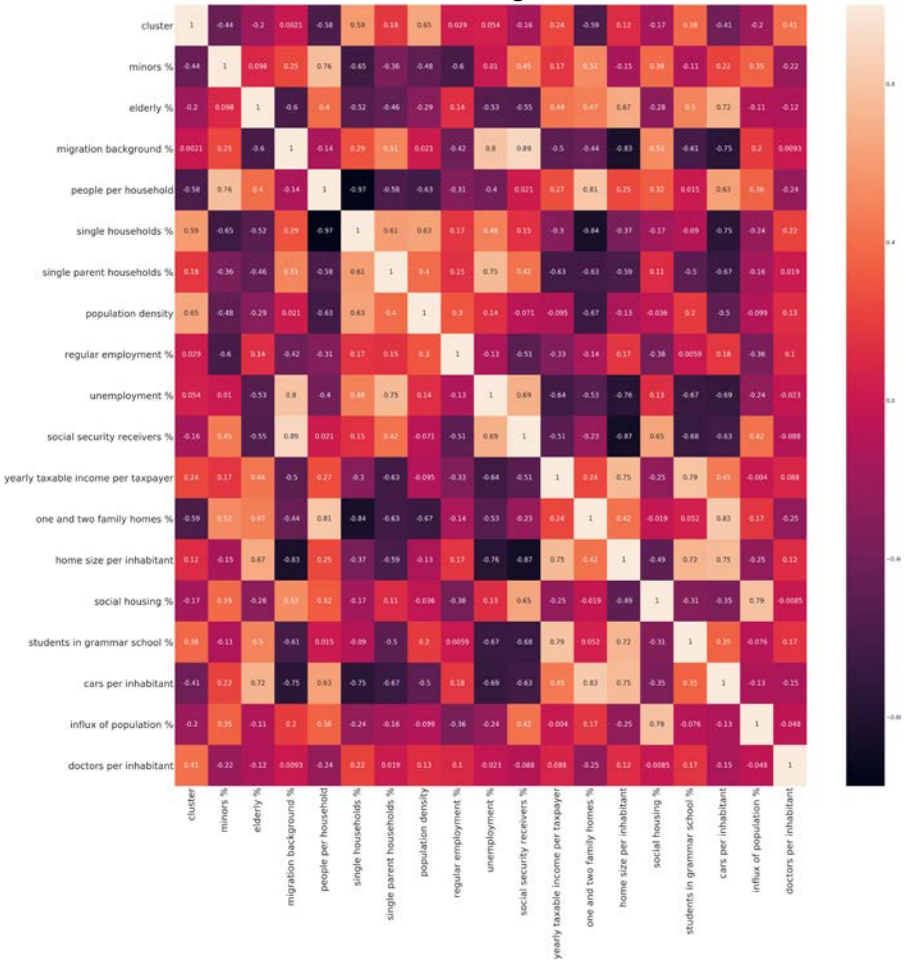
3.1 FEATURE SELECTION

After discarding redundant features in the demographic data set, the correlation of independent variables was inspected. Several pairs were highly correlated (Pearson correlation coefficient > 0.9).
From these highly correlated features, only one was kept, others were dropped from the dataset.

In the end, 19 features were selected

- minors %
- elderly %
- migration background %
- people per household
- single households %
- single parent households %
- population density
- influx of population %
- regular employment %
- unemployment %
- social security receivers %
- yearly taxable income per taxpayer
- one and two family homes %
- home size per inhabitant
- social housing %
- students in grammar school %
- doctors per inhabitant

Correlation matrix of remaining features:



3.2 CLUSTERING ALGORITHM

3.2.1 Normalization

Before performing the clustering, features were normalized: StandardScaler was used to transform the data such that its distribution has a mean value 0 and standard deviation of 1. This procedure makes sure that all features have a common scale.

3.2.2 K means

K-means algorithm was chosen to perform a clustering of neighbourhoods. The k-means algorithm searches for a pre-determined number of clusters within an unlabeled dataset. The "cluster center" is the arithmetic mean of all the points belonging to the cluster. Each point is closer to its own cluster center than to other cluster centers. The k-means algorithm was run 10 times with different centroid seeds. The final results is the best output of the 10 consecutive runs in terms of inertia.

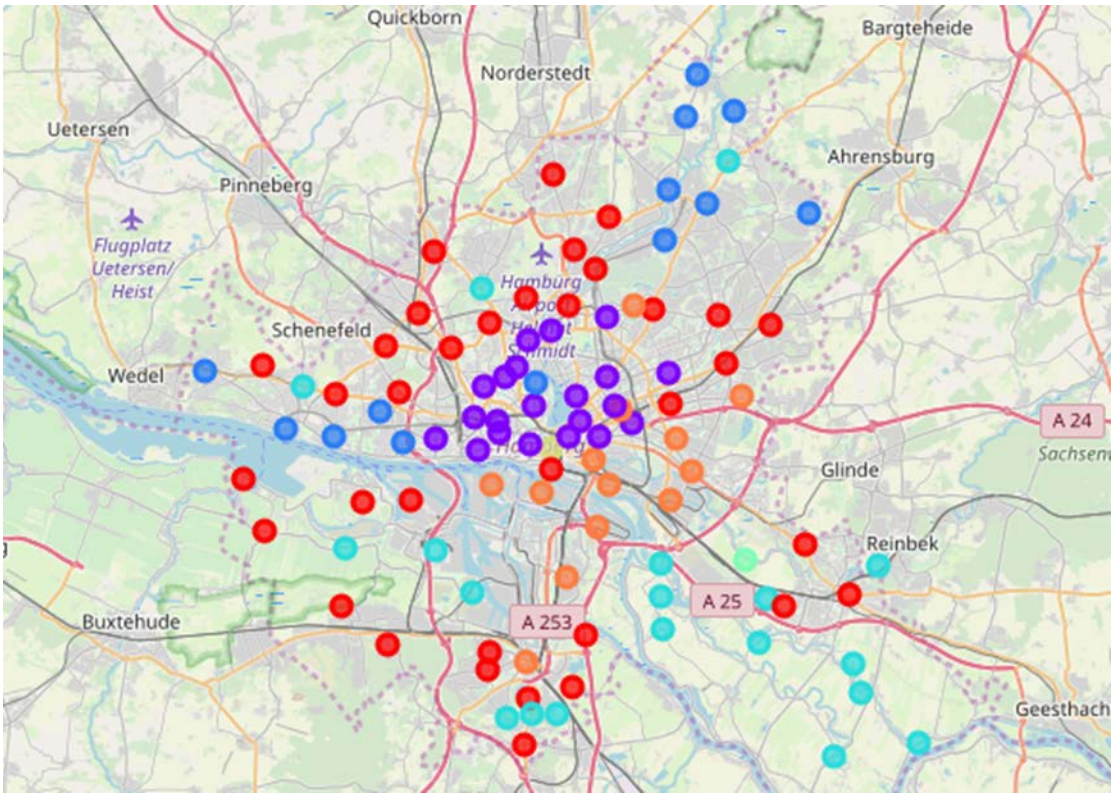
k-means cannot learn the number of clusters from the data, you must tell it how many clusters you expect. The best result was obtained for 7 clusters. Two of the clusters identified contained only one neighbourhood (outliers), five big clusters were identified by the algorithm.

K means algorithm was also used to cluster restaurant data retrieved via Foursquare API. Here as well, seven clusters were identified.

4. RESULTS

4.1 DEMOGRAPHIC CLUSTERS

Map of Hamburg with neighbourhoods color coded by cluster:



Clusters identified based on neighbourhood demographics:

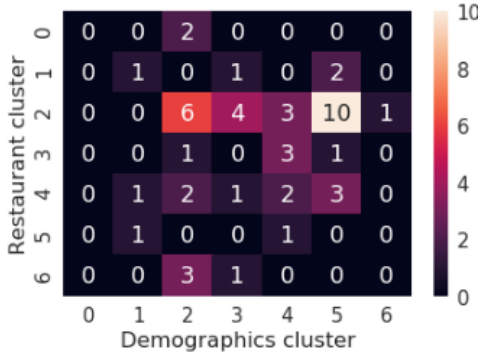
- **Neighbourhoods farther away from the centre, mainly families** e.g. Eidelstedt, Ohlsdorf and Bergedorf
- **Trendy neighbourhoods in central Hamburg** e.g. Eimsbüttel, Sternschanze and Ottensen
- **High income neighboorhoods not too close to the center** e.g. Nienstedten, Blankenese
- **Middle class neighboorhoods farther away from the center** e.g. Niendorf, Allermöhe
- **Poorer neighbourhoods** e.g. Wilhelsburg, Veddel
- Outliers: Hamburg-Altstadt (old town, low number of inhabitants), Billwerder (poor area with 50% of inhabitants living on social security)

4.2 RESTAURANT CLUSTERS

Italian and Asian restaurants are the most common restaurants in nearly all neighbourhoods of Hamburg. Most neighbourhoods fell into the same cluster, cluster 2, which showed a preference for Italian food and Fast food. In other clusters German (cluster 4) or Greek restaurants (cluster 1) were most popular.

4.3 RELATIONSHIP BETWEEN DEMOGRAPHIC CLUSTERS AND RESTAURANT PREFERENCES

There was no clear relationship between demographic clusters and restaurant clusters.



Food preferences in neighbourhoods demographically similar to Eimsbüttel:

neighbourhood	1st most common venue	2nd most common venue	3rd most common venue	4th most common venue	5th most common venue
Neustadt	French Restaurant	Italian Restaurant	Scandinavian Restaurant	Turkish Restaurant	Swabian Restaurant
St. Georg	Indian Restaurant	Italian Restaurant	Halal Restaurant	Asian Restaurant	Falafel Restaurant
Hamm	Greek Restaurant	Vietnamese Restaurant	Falafel Restaurant	Italian Restaurant	Indian Restaurant
Sternschanze	Vietnamese Restaurant	Vegetarian / Vegan Restaurant	Asian Restaurant	Korean Restaurant	Kumpir Restaurant
Altona-Nord	German Restaurant	French Restaurant	Vietnamese Restaurant	Falafel Restaurant	Italian Restaurant
Eimsbüttel	Italian Restaurant	Korean Restaurant	Asian Restaurant	Spanish Restaurant	Middle Eastern Restaurant
Rotherbaum	Falafel Restaurant	Italian Restaurant	German Restaurant	Vietnamese Restaurant	Indian Restaurant
Hoheluft-West	Italian Restaurant	Greek Restaurant	Chinese Restaurant	Spanish Restaurant	Japanese Restaurant
Hoheluft-Ost	Italian Restaurant	Chinese Restaurant	Mexican Restaurant	Paella Restaurant	Vietnamese Restaurant
Eppendorf	Argentinian Restaurant	German Restaurant	Vietnamese Restaurant	Falafel Restaurant	Italian Restaurant
Winterhude	Italian Restaurant	Thai Restaurant	Sushi Restaurant	Falafel Restaurant	Doner Restaurant
Uhlenhorst	Italian Restaurant	Asian Restaurant	Vietnamese Restaurant	Falafel Restaurant	Indian Restaurant
Barmbek-Süd	Doner Restaurant	Falafel Restaurant	German Restaurant	Italian Restaurant	Indian Restaurant
Barmbek-Nord	German Restaurant	Vietnamese Restaurant	Falafel Restaurant	Italian Restaurant	Indian Restaurant
Eilbek	Dim Sum Restaurant	Vietnamese Restaurant	Falafel Restaurant	Italian Restaurant	Indian Restaurant
Wandsbek	Greek Restaurant	Vietnamese Restaurant	Falafel Restaurant	Italian Restaurant	Indian Restaurant

4.4 RECOMMENDATION

The following recommendation could be given to a restaurant owner successfully operating in Eimsbüttel:

- Hoheluft of Winterhude have both similar demographics and a similar restaurant structure as Eimsbüttel. This could indicate that a restaurant currently operating in Eimsbüttel might successfully expand to these neighbourhoods.
- There could also be an opportunity in expanding to neighbourhoods that are demographically similar to Eimsbüttel, but do not have so many Italian Restaurants, for example Altona-Nord or Eppendorf. In these neighbourhoods competition is likely to be lower.

5. DISCUSSION

5.1 DEMOGRAPGIC DATASET

The clusters identified by k-means could be easily labeled using local knowledge. A Hamburg local might have performed a similar clustering which is a good sanity check. The result given by demographic clustering seems reliable.

5.2 FOURSQUARE DATA

Foursquare data for Germany does not seem to be as extensive as for other parts of the world. For some of the less popular/central neighbourhoods there was only a very limited number of restaurants found, or even none at all. Therefore, for a thorough analysis it would make sense to query a second data source for restaurant data.

6. CONCLUSION

Demographic data on Hamburg neighbourhoods as well as restaurant preferences obtained via Foursquare AOI have been used to identify similarities between different neighbourhoods. A recommendation was given to an Italian restaurant owner currently operating in Eimsbüttel as to which other neighborhoods she could expand to. The analysis could be useful for anybody seeing to open a restaurant in Hamburg.