

Roland Teschke

Target areas to launch new gyms in Frankfurt, Germany

Capstone project, IBM Data Science Professional Certificate

1.2.2021

Table of contents

1) Introduction and business problem	2
a) Frankfurt, Germany: A brief summary about the city	2
b) Growing fitness market in Germany	2
c) Target audience, business problem and benefits provided by this report	2
2) Data	3
a) Overview of the data sources.....	3
b) GeoJSON data	3
c) Location data for gyms in and around Frankfurt.....	5
d) Statistical data	7
e) Other computed data, ratios and indices.....	13
3) Methodology	17
4) Results	20
a) Summary of results.....	20
i. General information about gyms in Frankfurt and Frankfurt boroughs	20
ii. Boroughs to consider for launching new gyms respectively expanding existing gym franchises.....	20
b) Detailed results.....	22
5) Discussion	40
6) Conclusion	40

1) Introduction and business problem

a) Frankfurt, Germany: A brief summary about the city¹

Frankfurt, officially Frankfurt am Main, is the most populous city in the German state of Hesse. Its 763,380 inhabitants as of December 31, 2019 make it the fifth-most populous city in Germany. The city is at the centre of the larger Rhine-Main Metropolitan Region, which is Germany's second-largest metropolitan region after the Rhine-Ruhr Region. Frankfurt is a global hub for commerce, culture, education, tourism and transportation. Frankfurt Airport is Germany's busiest. Frankfurt is one of the major financial centres of the European continent, with the headquarters of the European Central Bank, Deutsche Bundesbank, Frankfurt Stock Exchange, Deutsche Bank, DZ Bank, KfW, Commerzbank, to name just a few.

b) Growing fitness market in Germany²

Fitness market revenues and number of fitness-related venues in Germany have been increasing significantly in the previous years. Due to this major trend, it may be assumed that additional facilities are planned to be established in the near future.

As we will discover within this research, Frankfurt already has many gyms as of January 2021.

c) Target audience, business problem and benefits provided by this report

Target audience:

This report is targeted at

- owners and managers of already established gyms, gym franchises who intend to expand their business and set a new gyms in Frankfurt am Main, and
- people interested at setting up a new gym (i.e. who have previously not owned and/or managed a gym).

Business problem:

Given the vast existing competition and the potential desire to establish new gym, insights are desired as to which areas / boroughs might be most suitable in promising business success when setting up new gyms in Frankfurt am Main, Germany.

Benefits of this report:

As Frankfurt is a large and diversity city, this report will

- provide selected information about Frankfurt boroughs with regards to existing gyms, alternative facilities, average age, population, income data, and
- gain insights which factors might also be considered to establish a gym in Frankfurt (focusing on selected data, not being fully exhaustive), and
- provide a proposal which areas might be most interesting to launch new gyms in Frankfurt.

¹ Source: Abstract, based on <https://en.wikipedia.org/wiki/Frankfurt> - slightly amended for this report

² Source: <https://de.statista.com/statistik/daten/studie/6228/umfrage/umsatz-der-fitness-branche-in-deutschland/#statisticContainer>

2) Data

a) Source code and interactive outputs

- **NBViewer (recommended for rendering maps):**
https://nbviewer.jupyter.org/github/rolandteschke/Coursera_Capstone/blob/main/Target%20areas%20for%20opening%20a%20gym%20in%20Frankfurt.ipynb
- **Github:**
https://github.com/rolandteschke/Coursera_Capstone/blob/main/Target%20areas%20for%20Opening%20a%20gym%20in%20Frankfurt.ipynb

b) Overview of the data sources

This report is based on several data sources as follows:

- GeoJSON data for Frankfurt boroughs
- Gym locations in Frankfurt and surrounding areas, as offered by Foursquare
- Publicly available statistical data for boroughs of Frankfurt, Germany
- Self-computed data which is derived from processing the above mentioned data sources further

The following sub-chapters are going to provide more detailed information for each of these data sources.

c) GeoJSON data

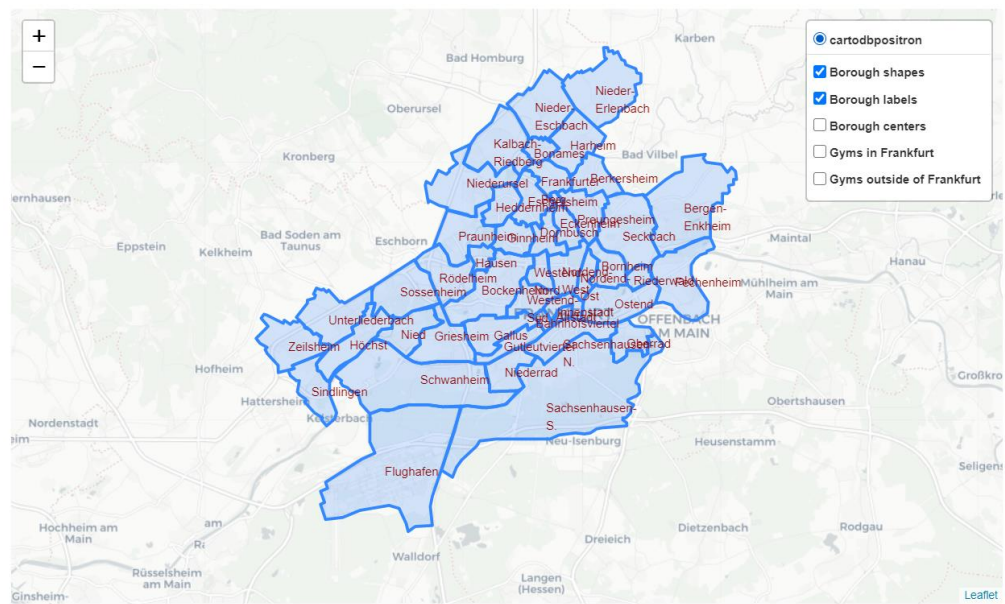
Source:

https://opendata-esri-de.opendata.arcgis.com/datasets/ca64da7abad04c0eb8717ca3ec486cae_0?geometry=7.594%2C49.967%2C9.680%2C50.275

Publically and officially available GeoJSON data has been retrieved for Frankfurt am Main in order to be able to properly visualize selected information in maps and choropleth maps for each of it's 46 boroughs.

Each borough center's coordinates have been computed in order to request gym information from Foursquare subsequently.

Example:



d) Location data for gyms in and around Frankfurt

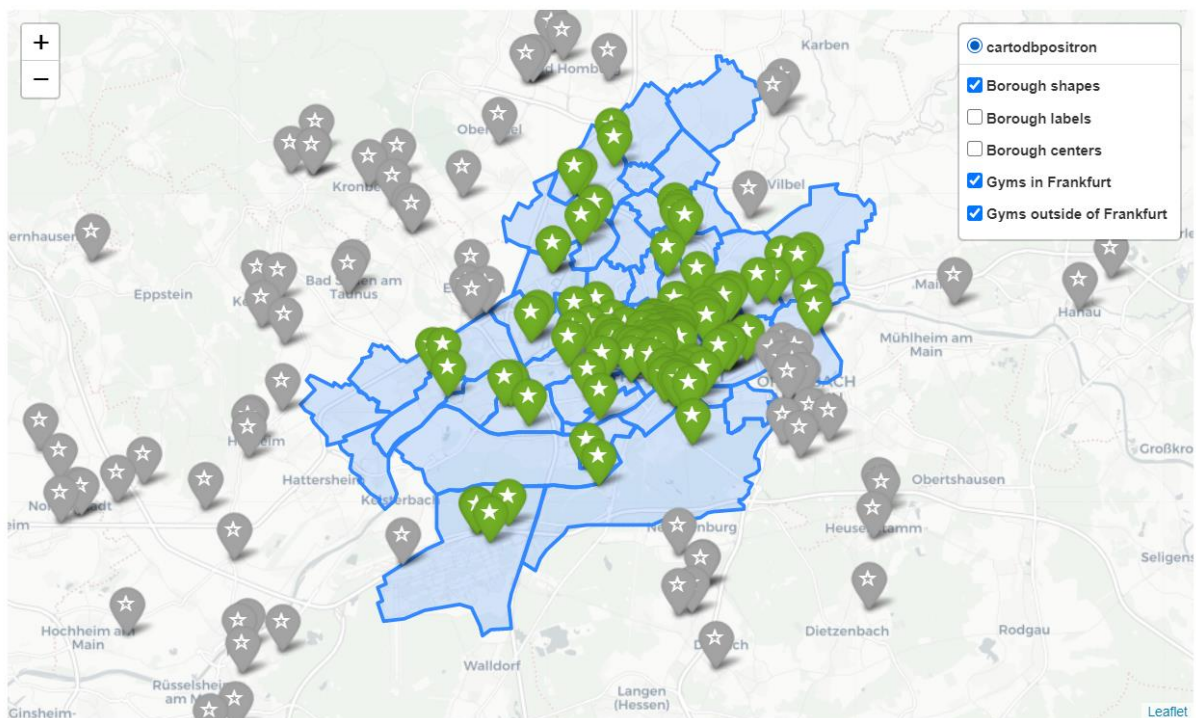
Source: <https://developer.foursquare.com/> , <https://api.foursquare.com>

The Foursquare API has been used in order to retrieve all gyms in an around Frankfurt for each of the 46 boroughs.

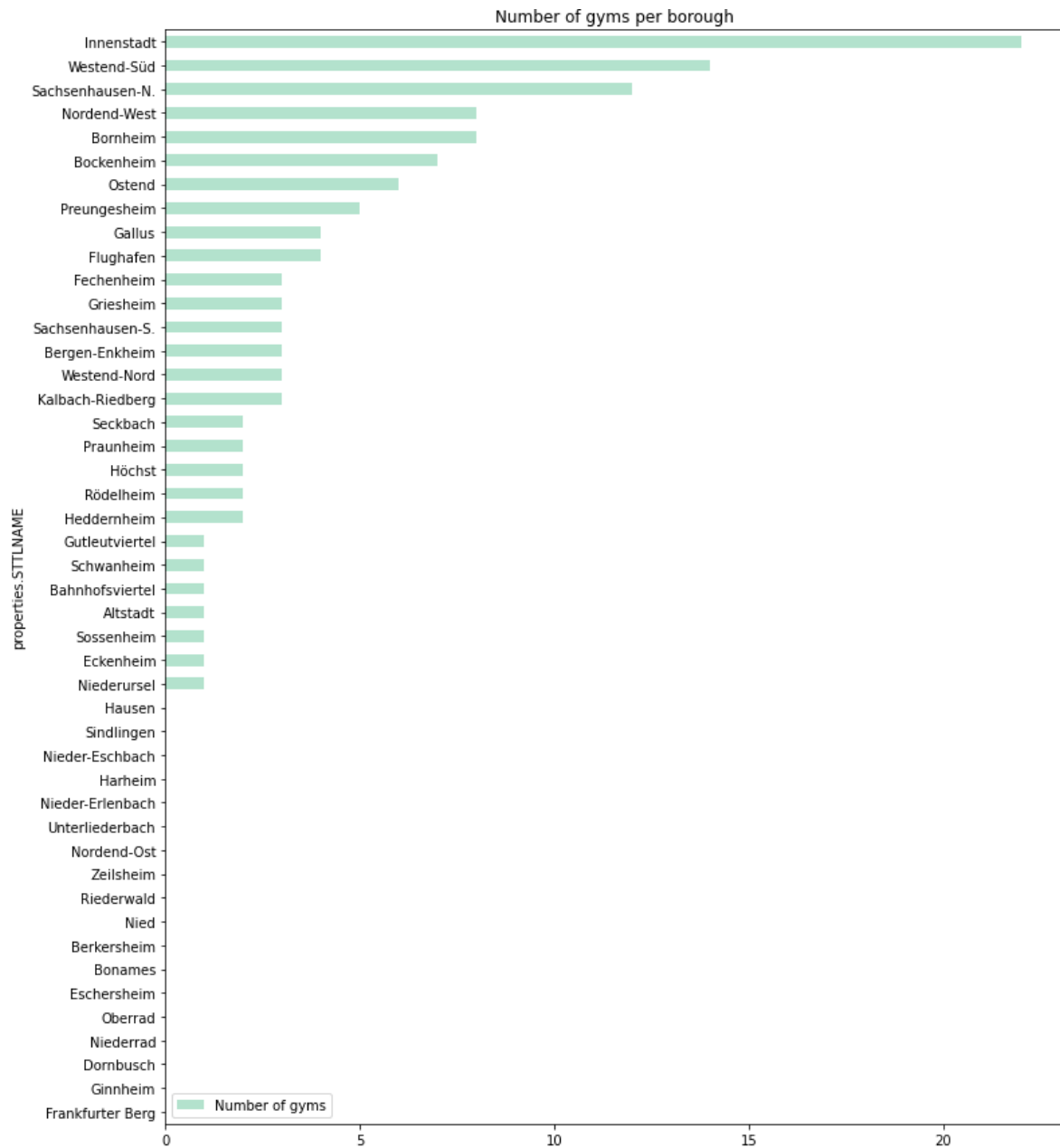
The data was processed further as follows:

- Location retrieval was consciously limited to include gyms only for each borough (based on the borough center locations)
- Duplicates had to be removed since the borough area sizes (i.e. also borough center distances) vary significantly. Approximately 200 gyms have been retrieved in total³.
- Gyms were classified as either to be within or outside of Frankfurt (**entire city**)
- Gyms were classified as either to be within the given **boroughs**

Examples:



³ very slightly varying depend on when the real-time request was performed



e) Statistical data

Source: https://statistik.stadt-frankfurt.de/strukturdatenatlas/downloads/Stadtteile_FFM.xlsx

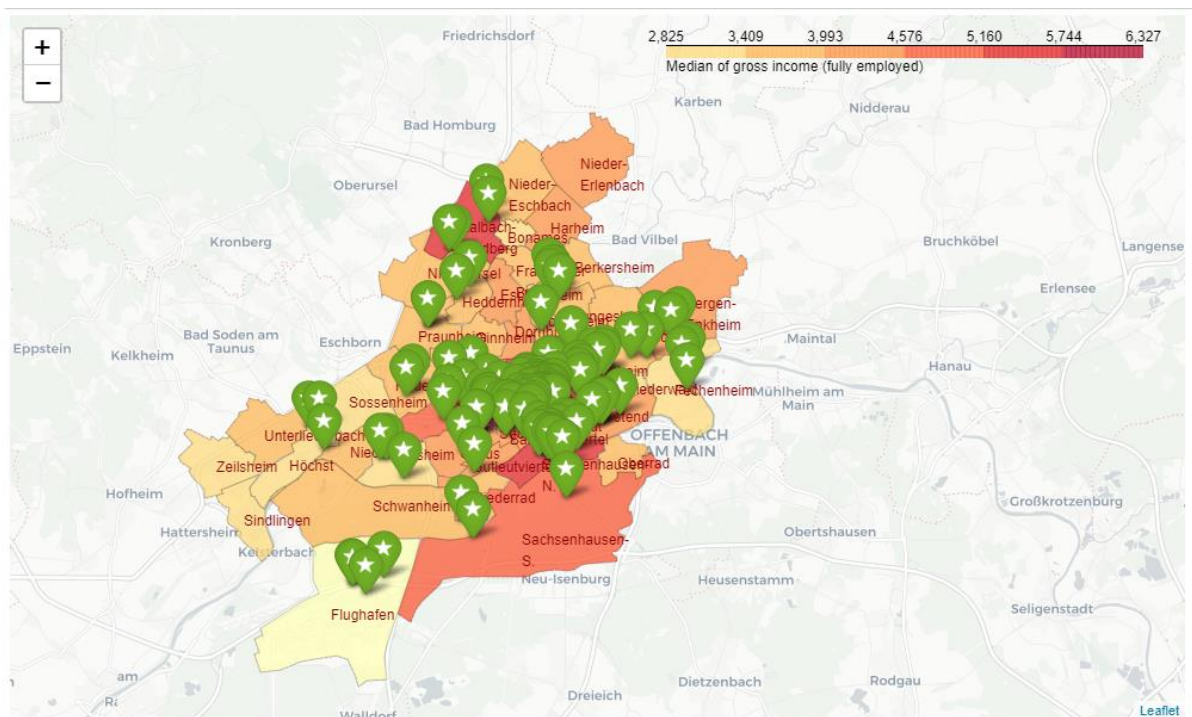
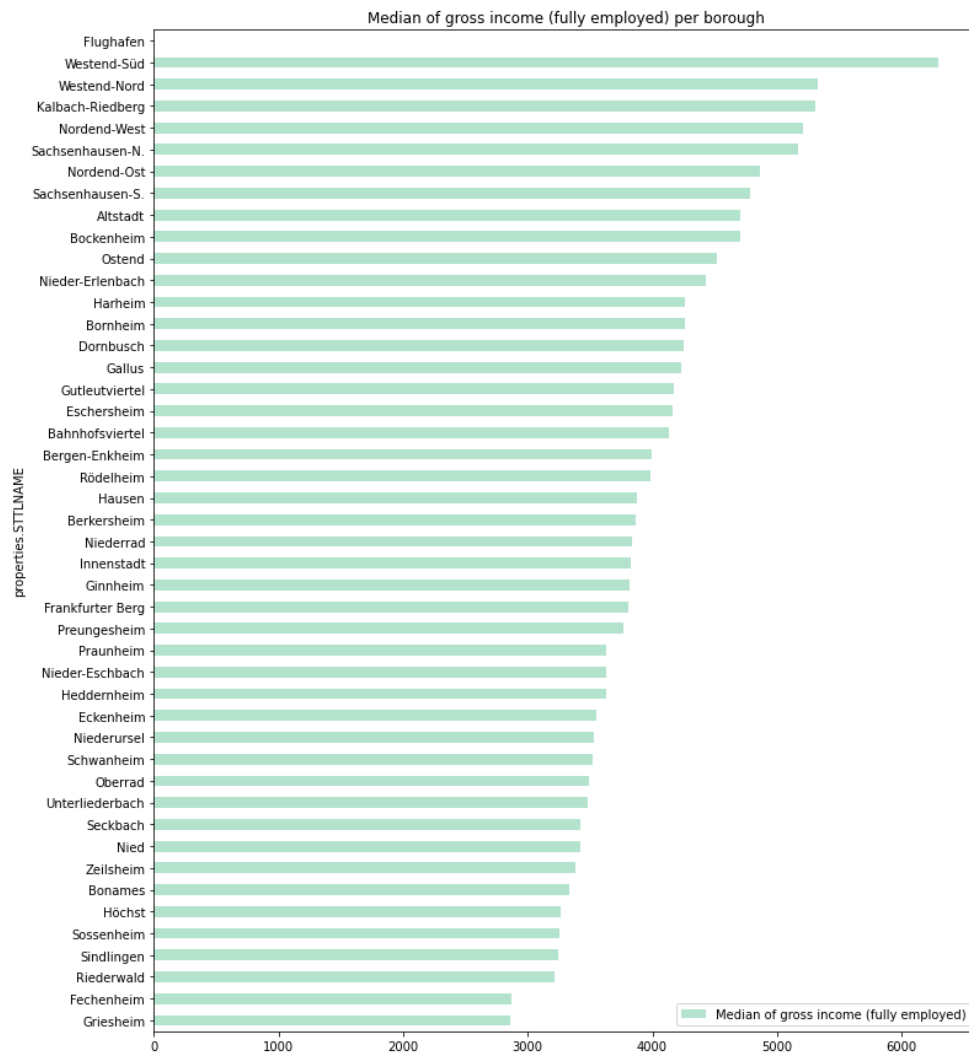
The city of Frankfurt offers a vast array of statistical data. The data selected for this research is the most up-to-date publicly available - that is, data from end of Dec 2019.

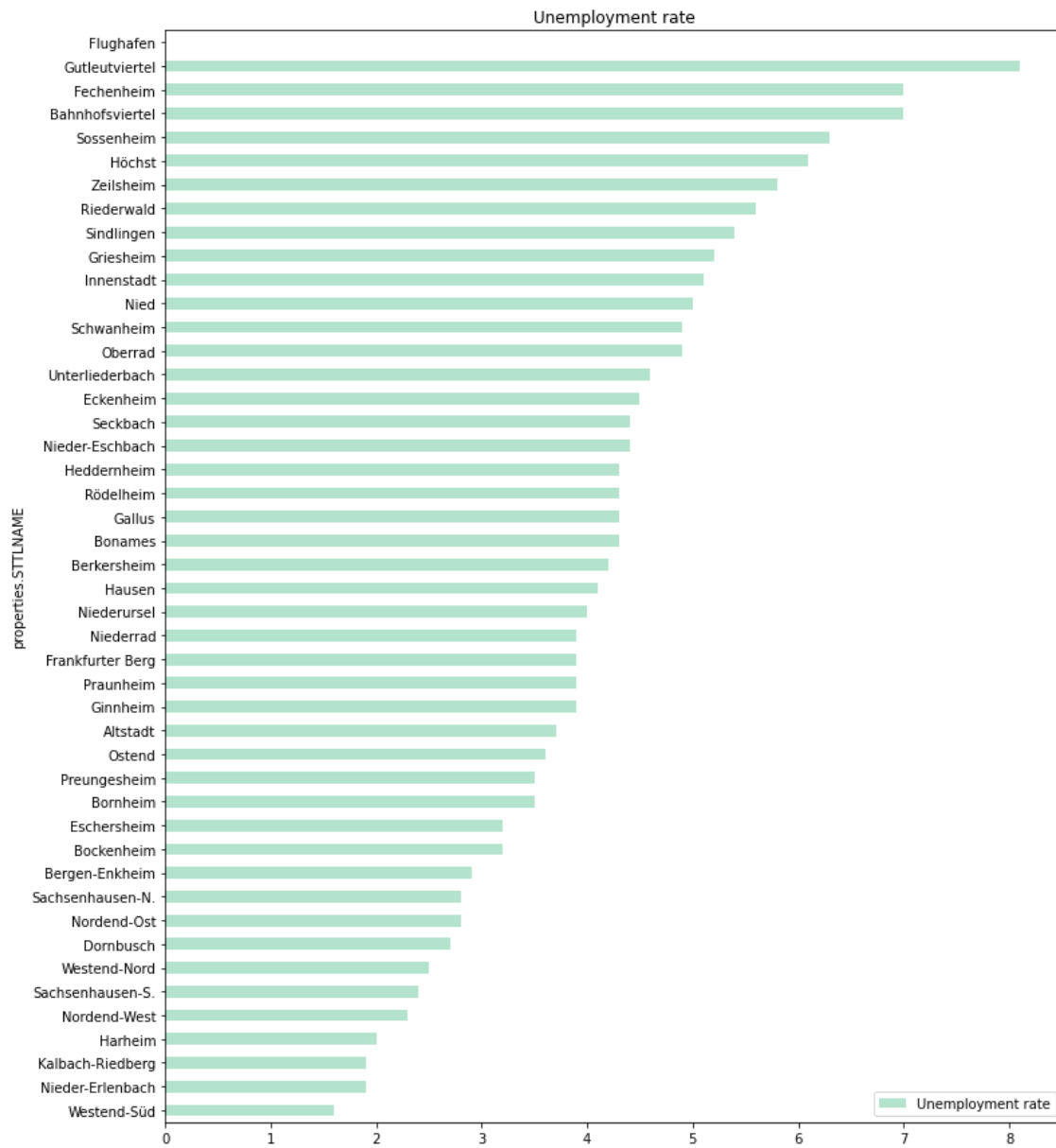
- The raw data was downloaded as an Excel file
- Potentially relevant metrics have been selected in the Excel file
- The excel cells did not contain clean header data as it contained several grouping mechanisms, so the data was slightly preprocessed in Excel in order to support a clean import to the Python notebook
- Subsequently been imported into Python notebook for further data processing as a Pandas dataframe

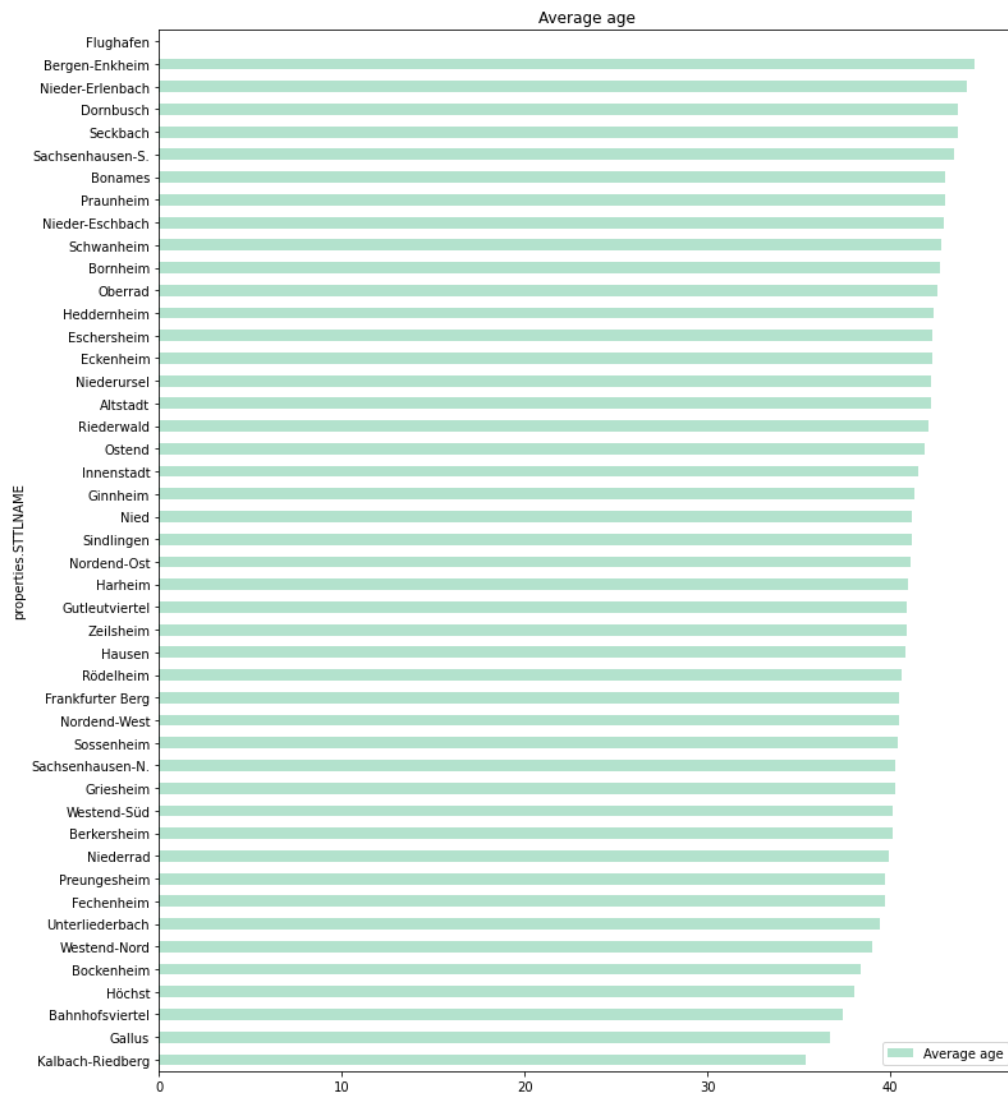
Several statistics have been selected for further research as follows:

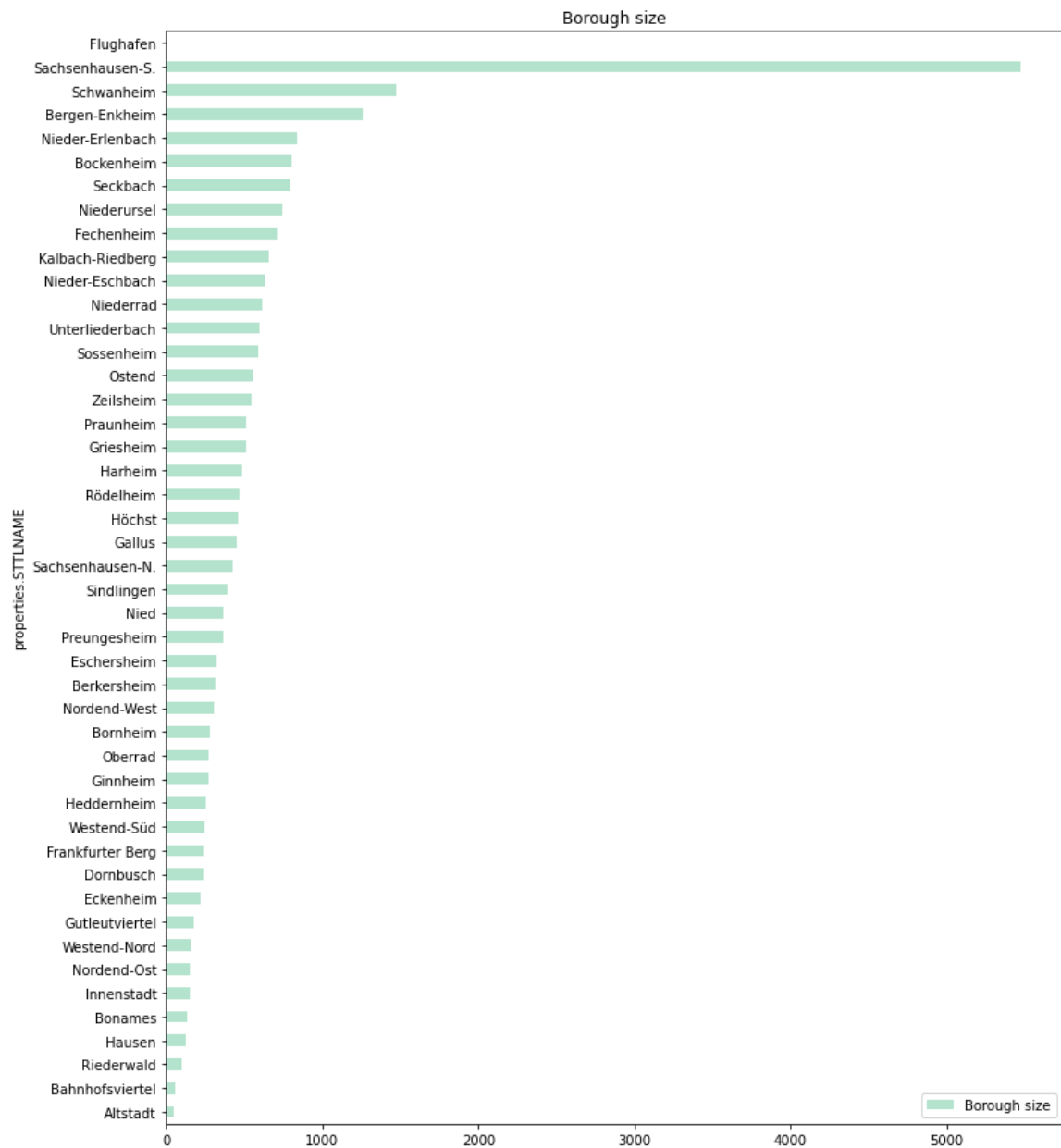
- Area information
 - Borough size
 - Size of sports, leisure and recreation areas
- Population details
 - Households
 - Inhabitants
 - People per household
 - Population density per ha
- Socio-demographic
 - Average age
 - Gender (female, male)
- Employment related
 - Employment total
 - Employment rate
 - Unemployment rate
- Financial
 - Median of gross income (fully employed)

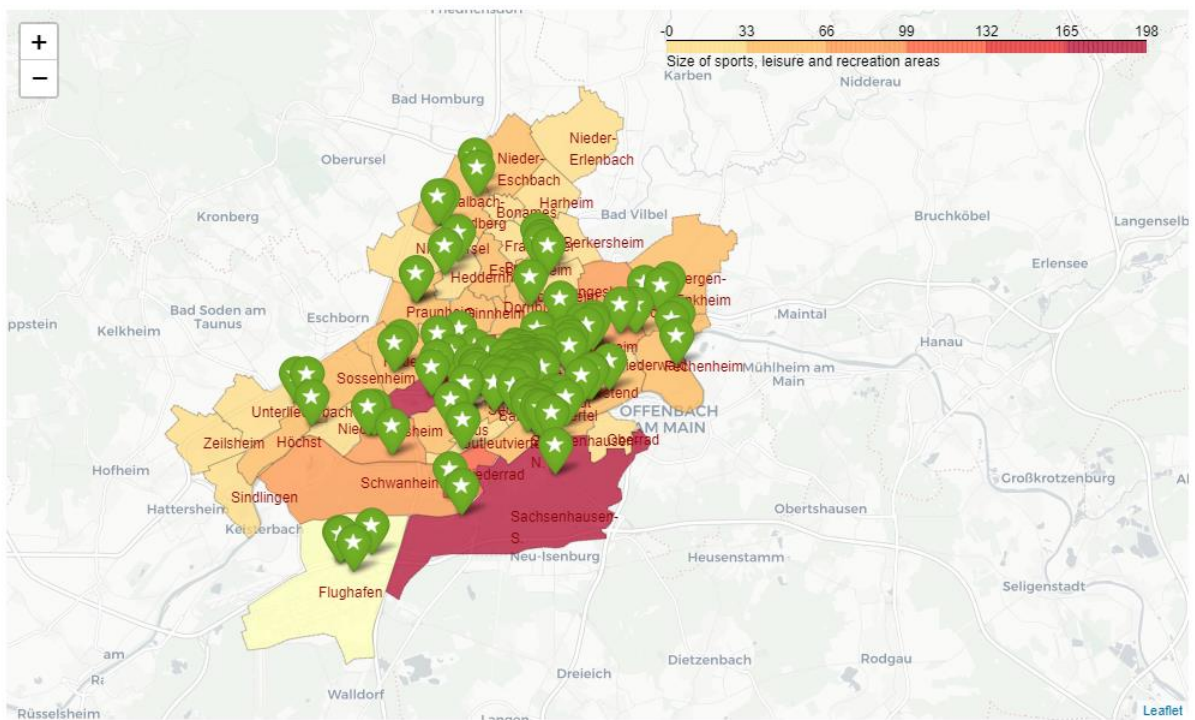
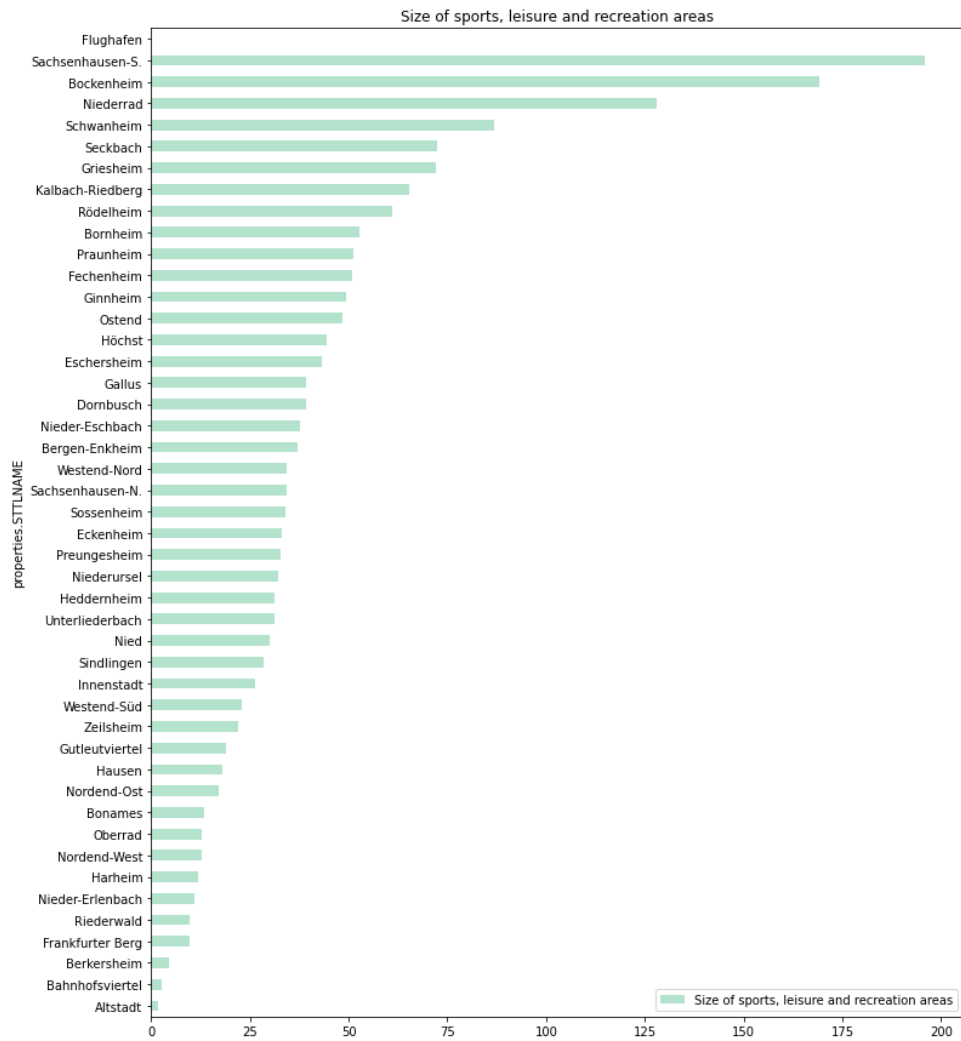
The following examples illustrate variances for the data retrieved for the various boroughs.











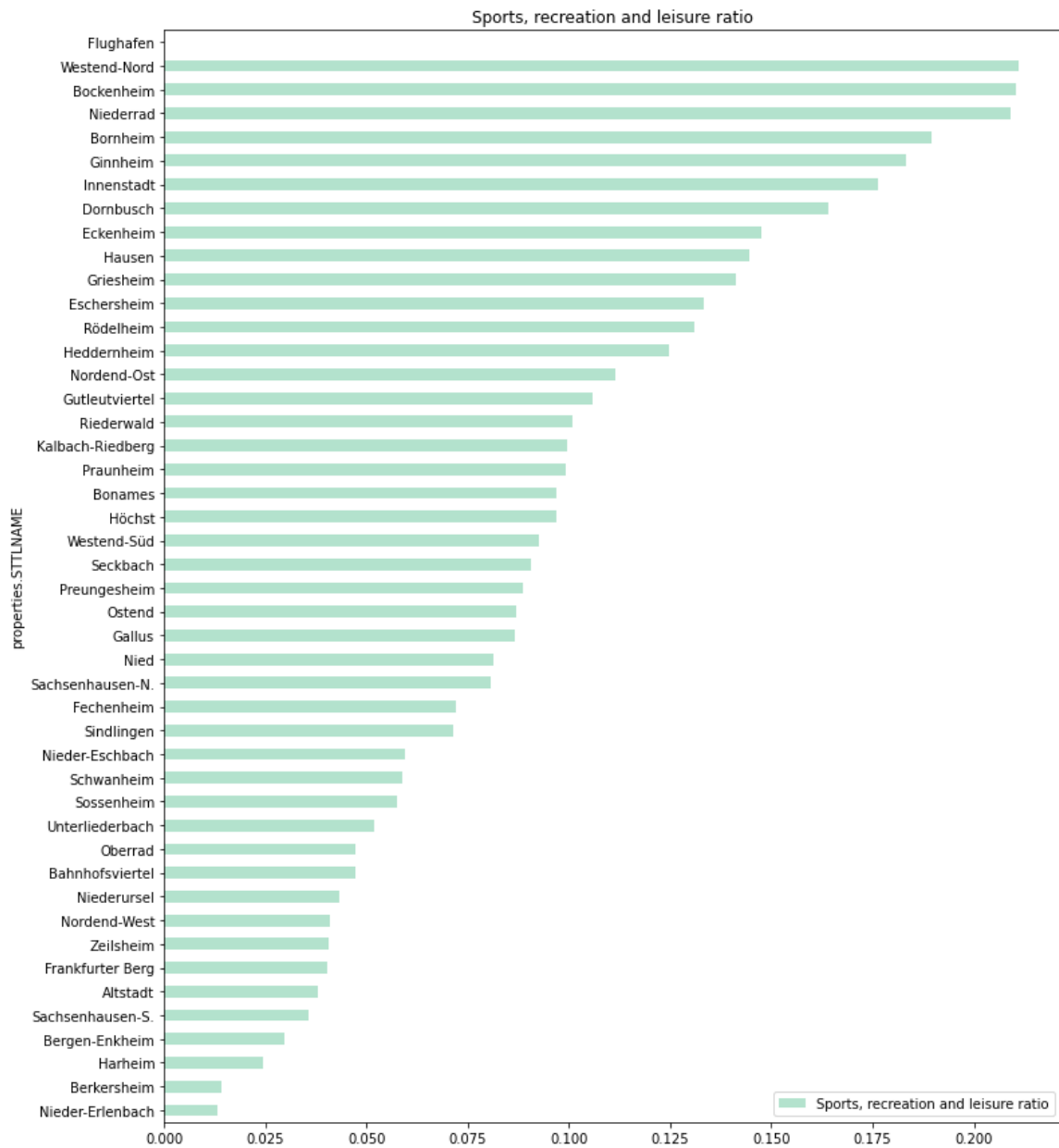
f) Other computed data, ratios and indices

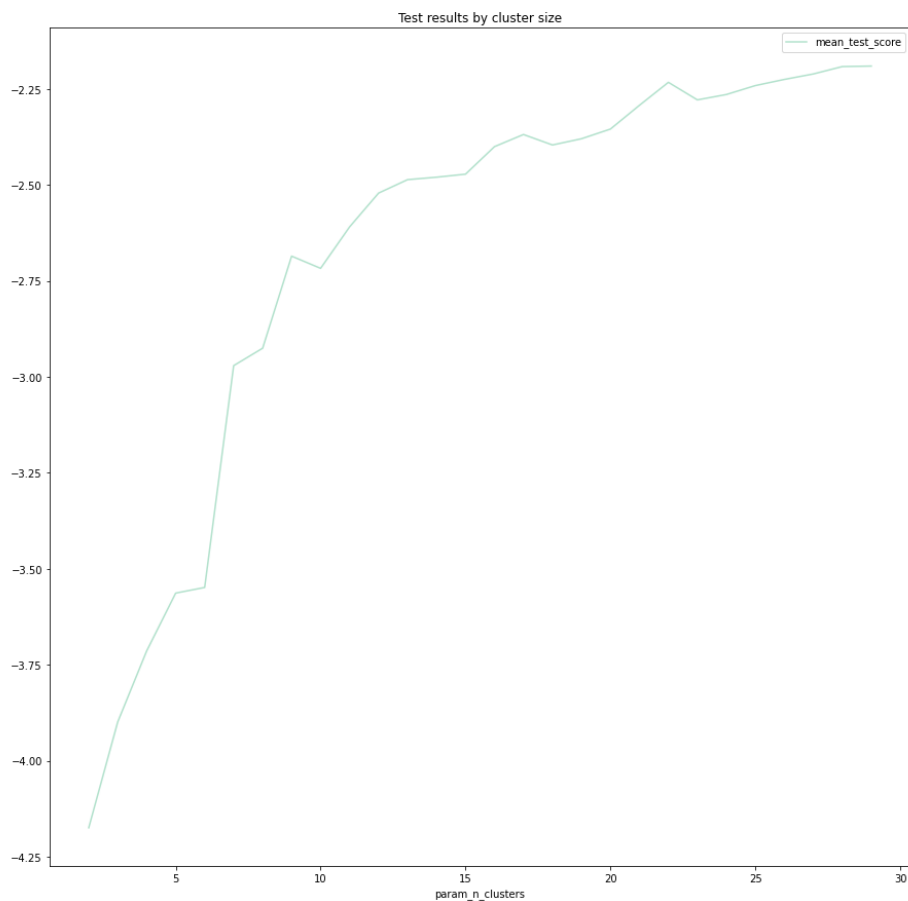
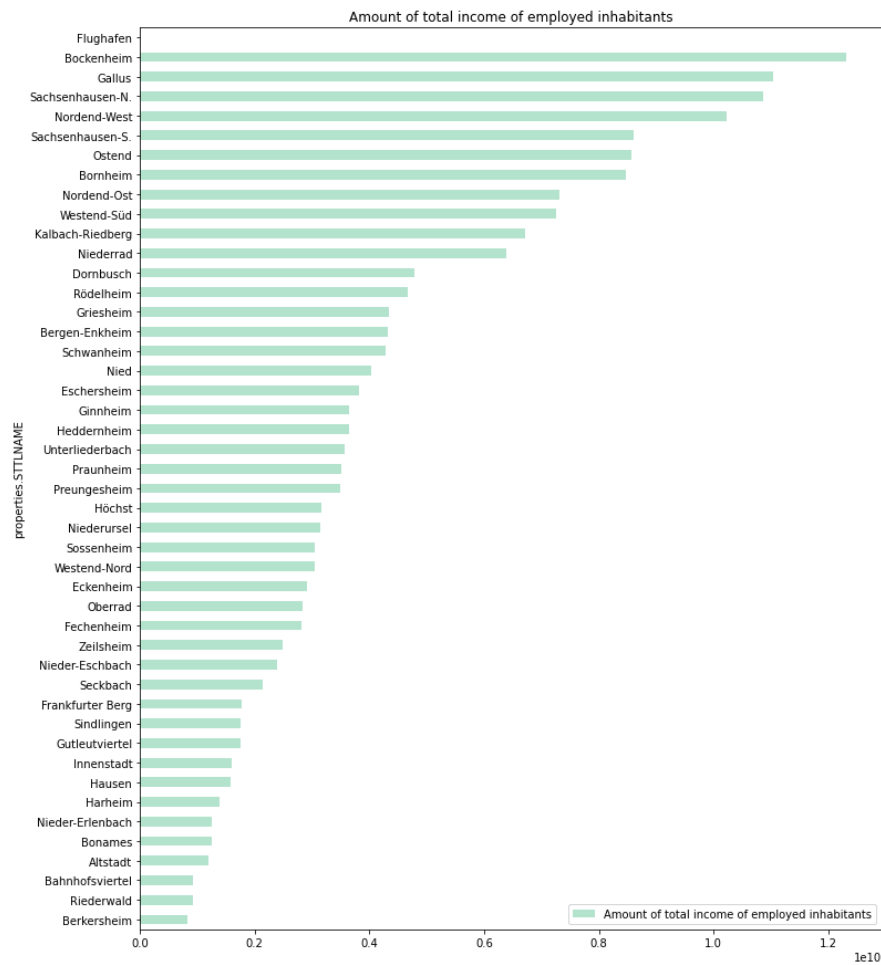
The following data was computed based on processing the various data sources and as either a prerequisite for further processing, or as a finding of some analyses described in more detail in the upcoming methodology and results sections:

- **Totals:**
 - o Amount of total income of employed inhabitants for each borough (= inhabitants * employment rate * median gross income)
 - o Number of gyms per borough
- **Boolean:**
 - o Determine whether gyms are inside or outside of Frankfurt (entire city)
 - o Determine whether gyms are inside specific boroughs of Frankfurt
- **Categories:**
 - o Gym competition level (based on number of gyms)
 - o Clusters of similarity based on a few selected variables (see cluster in analyses in subsequent report sections)
- **Continuous:**
 - o GridSearch results to help determine the number of clusters to choose for a cluster analysis
- **Gym expansion index**
 - o Gym expansion target index for each borough. It is important to note that the variables had to be scaled upfront to ensure the different ranges of the 3 variables can be used to create a meaningful target expansion index.
The index has been computed **from normalized values** as
 - + Amount of total income of employed inhabitants for each borough
 - 50% of the ratio of sports, leisure and recreation areas
 - 50% of the number of gyms in the borough

Here are some examples of the computed data:







3) Methodology

A vast set of methods were used to conduct this research. The following overview provides a concise summary.

Purpose	Method(s)	Variables / Notes
Data import	Process GeoJSON	Borough names and geoshapes of Frankfurt boroughs
	Process Excel	<i>Many statistical variables, see data and results sections</i>
	Foursquare REST API and processing JSON	Gyms in and around Frankfurt am Main
Data (pre-)processing to enable more detailed research	<ul style="list-style-type: none"> - Compute borough centers - Classify whether gyms are inside or outside of Frankfurt - Classify in which borough a gym is located - Determine (sub-)type of gyms - Classify gym competition levels by boroughs - Ratio of each borough's sports, leisure and recreation areas compared to borough size 	(see methods)
Data exploration to investigate distributions for boroughs and computed classes and indices	Various descriptive statistics, predominantly: <ul style="list-style-type: none"> - Grouping - Counts - Means 	<i>Variables explored were visualized as seen in the subsequent rows of this table</i>
Data visualization of distributions for boroughs and computed classes and indices	Visualization: Bar charts (unstacked & stacked)	<ul style="list-style-type: none"> • Number of gyms per borough • Number of boroughs with regards to gym competition level • Number of gym (sub-)types by borough • Median of gross income (fully employed) per borough • Employment rate by borough • Unemployment rate by borough • Average age by borough • Borough sizes

		<ul style="list-style-type: none"> • Size of sports, leisure and recreation areas • Population density per ha • Sports, recreation and leisure ratio • Amount of total income of employed inhabitants • Gym expansion - target index
	Visualization: Maps	Interactive map showing <ul style="list-style-type: none"> • Gym locations in and outside of Frankfurt • Borough shapes • Borough labels • Borough centers
	Visualization: Choropleth maps highlighting borough differences	<ul style="list-style-type: none"> • Median of gross income (fully employed) • Inhabitants • Population density • Size of sports, leisure and recreation areas • Sports, recreation and leisure ratio • Gym expansion - target index
Descriptive analyses to determine potentially relevant variables for KMeans and creation of index for recommending target areas	Decision tree analysis for gym competition levels Note of caution: Based on the given data, the decision tree results vary across different runs. Hence results are just used as additional information, but feature selection should have some additional theoretic assumptions. It should be noted that causality might be questioned, but correlations might occur. For the decision tree analysis, the gym competition level is used as primary target as shown below. This ensures better legibility of the results, and it can be argued that number of gyms as continuous variable might be too granular and yields many more depth levels.	X = Borough size; Size of sports, leisure and recreation areas; Households; Inhabitants; female; male; People per household; Population density per ha; Average age; Employment total; Employment rate; Unemployment rate; Median of gross income (fully employed) Y = Gym competition level
Group clusters of similar types of boroughs	KMeans clustering analysis Several scaling methods were run to identify most suitable. As expected and according to SKLearn documentation,	X= Borough size; Sports; recreation and leisure ratio; Inhabitants; Average age; Employment total;

	<p>MaxAbsScaler returned the 2nd best and most meaningful results for this dataset since the dataset has only few observations since it is largely based on aggregate data by borough already. The Normalize scaler was discarded since it's intended for using in text mining analyses.</p> <p>A GridSearch was performed to help identify a suitable number of clusters for the KMeans clustering. Results have been visualized in a line plot.</p> <p>Since it is meaningful to balance number of clusters vs. test score and given the curve in the grid search results, a cluster size of 9 is chosen (to have a sufficient but not too finegrained clustering). Random state will be the same as in the gridsearch.</p>	<p>Employment rate; Unemployment rate; Median of gross income (fully employed); Number of gyms</p> <p>Results were visualized as:</p> <ul style="list-style-type: none"> • Choropleth map • Table of statistics for each cluster
Recommend target areas for opening new gyms	<p>Index creation</p> <p>After inspecting of both clustering and decision tree, it was deemed suitable to specifically create and visualize an index for recommending in which boroughs gyms might be opened.</p> <p>It is important to note that the variables had to be scaled upfront to ensure the different ranges of the 3 variables can be used to create a meaningful target expansion index. The same scaling method was used for the cluster analysis.</p> <p>The Gym expansion for each borough is computed as follows <u>from normalized values</u> as:</p> <p><i>(plus)</i> Amount of total income of employed inhabitants for each borough</p> <p><i>(less)</i> 50% of the ratio of sports, leisure and recreation areas</p> <p><i>(less)</i> 50% of the number of gyms in the borough</p>	<p>Gym expansion index</p> <p>Results were visualized as:</p> <ul style="list-style-type: none"> - Bar chart of the index by borough, sorted to visualize top target borough for gym expansion / openings - Choropleth map - Table of relevant statistics for top 10 boroughs

Further details are available in the notebook, data and results sections of this report.

4) Results

The following provides first a summary of the findings. Detailed charts, maps and tables are listed afterwards.

a) Summary of results

i. General information about gyms in Frankfurt and Frankfurt boroughs

- **Frankfurt has 46 boroughs**, of which “Airport” is often discarded in the enhanced statistics since the statistical data is not published for this area
- **206 gyms** were retrieved for Frankfurt and the surrounding areas.
 - There are **125 gyms in the city of Frankfurt**, and 81 outside of Frankfurt.
- There is a **huge variety of number of gyms depending on borough**.
 - 18 boroughs have no gyms
 - 20 boroughs have between 1 and 4 gyms
 - 5 boroughs have 5 and 10 gyms
 - 3 boroughs have more than 10 gyms
- **Most gyms are general gyms**. There are **only a few specialized sub-categories of gyms** (e.g. Yoga Studio, Martial Arts gym, boxing gym, climbing gym)
- **The median gross income of employed people vary a lot across the different boroughs** (under 3000 € to over 6000 € per borough). With the exception of Kalbach-Riedberg, most boroughs with high gross income are towards the inner center of the city.
- **Unemployment rates vary a lot across the different boroughs** (under 2% to over 6000 € per borough)
- In terms of borough sizes, there is Sachsenhausen-Süd as an outlier (much bigger). All other boroughs have a variance, but not as extreme.
- **Size of sports, leisure and recreation areas and the ratio compared to the borough size vary much across the boroughs**. In both metrics, Bockenheim ranks in the top2 section.
- Running a decision tree analysis, **the size of sports, leisure and recreation areas is the top variable to make a difference when trying to investigate the gym competition levels across different boroughs**.
- Population is more dense in the inner center boroughs compared to the outskirts.

ii. Boroughs to consider for launching new gyms respectively expanding existing gym franchises

- **KMeans clustering**: Following a grid search using different potential cluster sizes, **Frankfurt’s boroughs were grouped into 9 similar clusters**, with key findings as follows (more details in the notebook):
 - Cluster 4 and 5 show the highest mean index for targeting launch or expanding gyms.
 - Both clusters shows various overlaps also when plotting the target index for the top boroughs that launching new gyms should be considered.
 - Population is very dense, employment rates are high.
 - Median gross income of the 6 boroughs contained in both clusters are the highest compared to the other clusters.

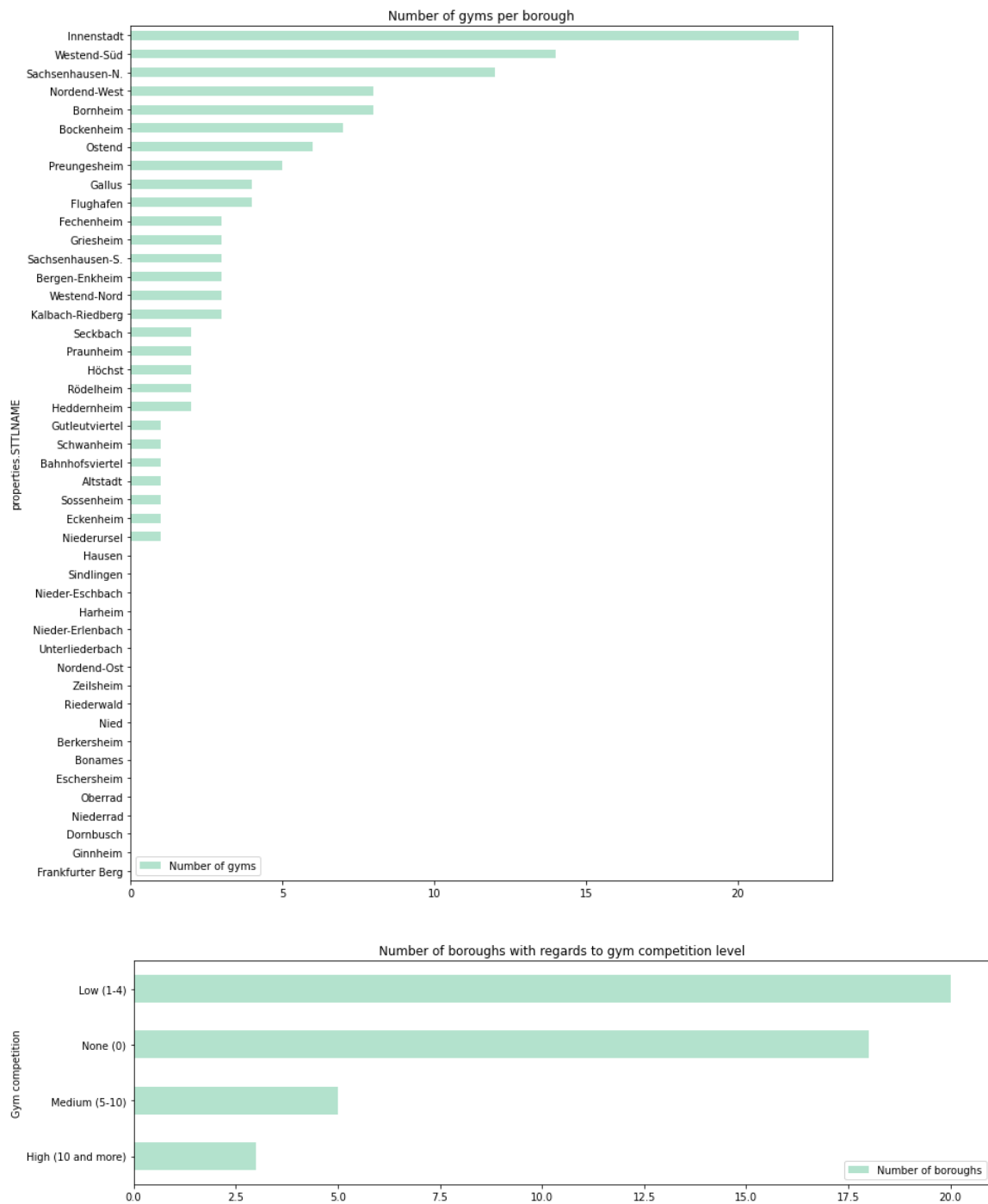
- **Gym expansion target index:**
 - The top 5 items created in fact are all in either cluster 4 or 5. So at least for top items, a combination of both approaches seems very reasonable.

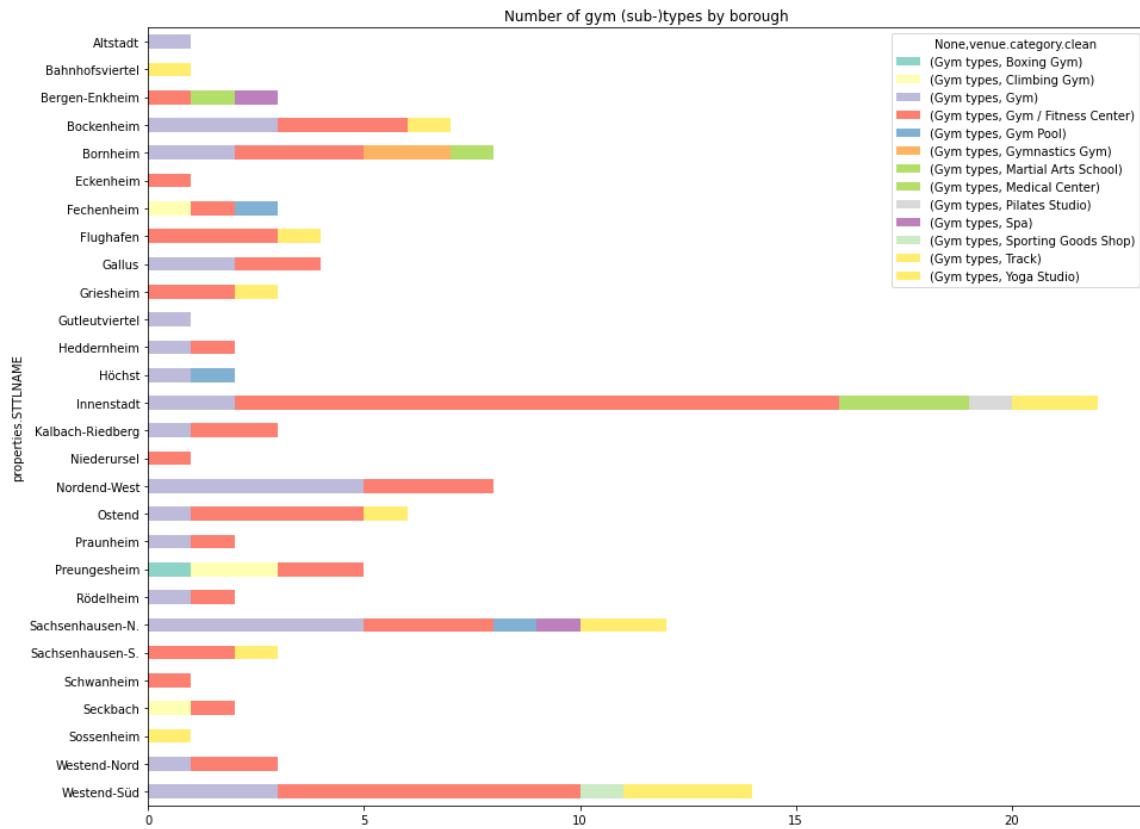
Hence the final recommendation for launching new gyms are the following boroughs:

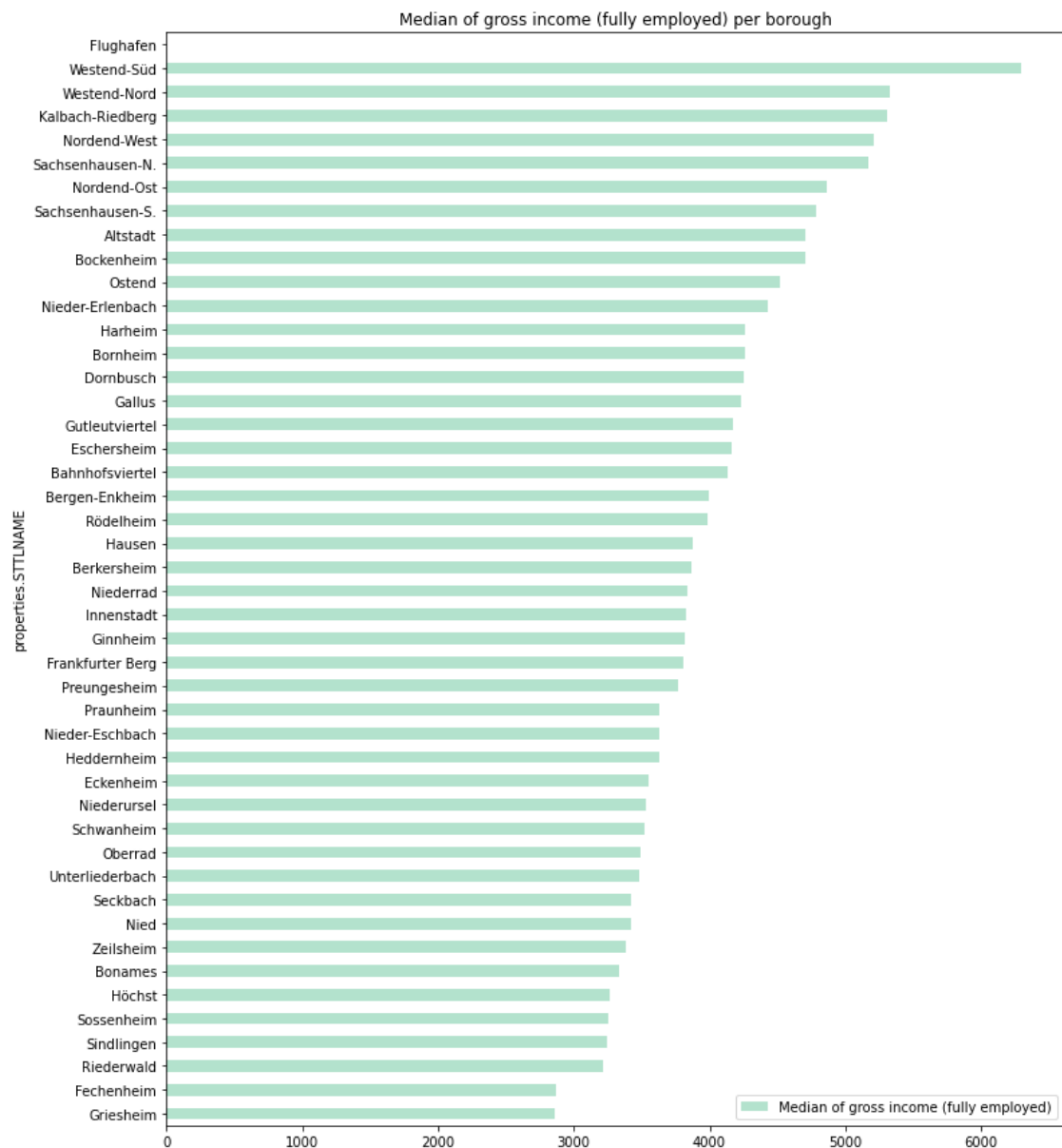
- Gallus
- Nordend-West
- Sachsenhausen Süd
- Sachsenhausen Nord
- Ostend

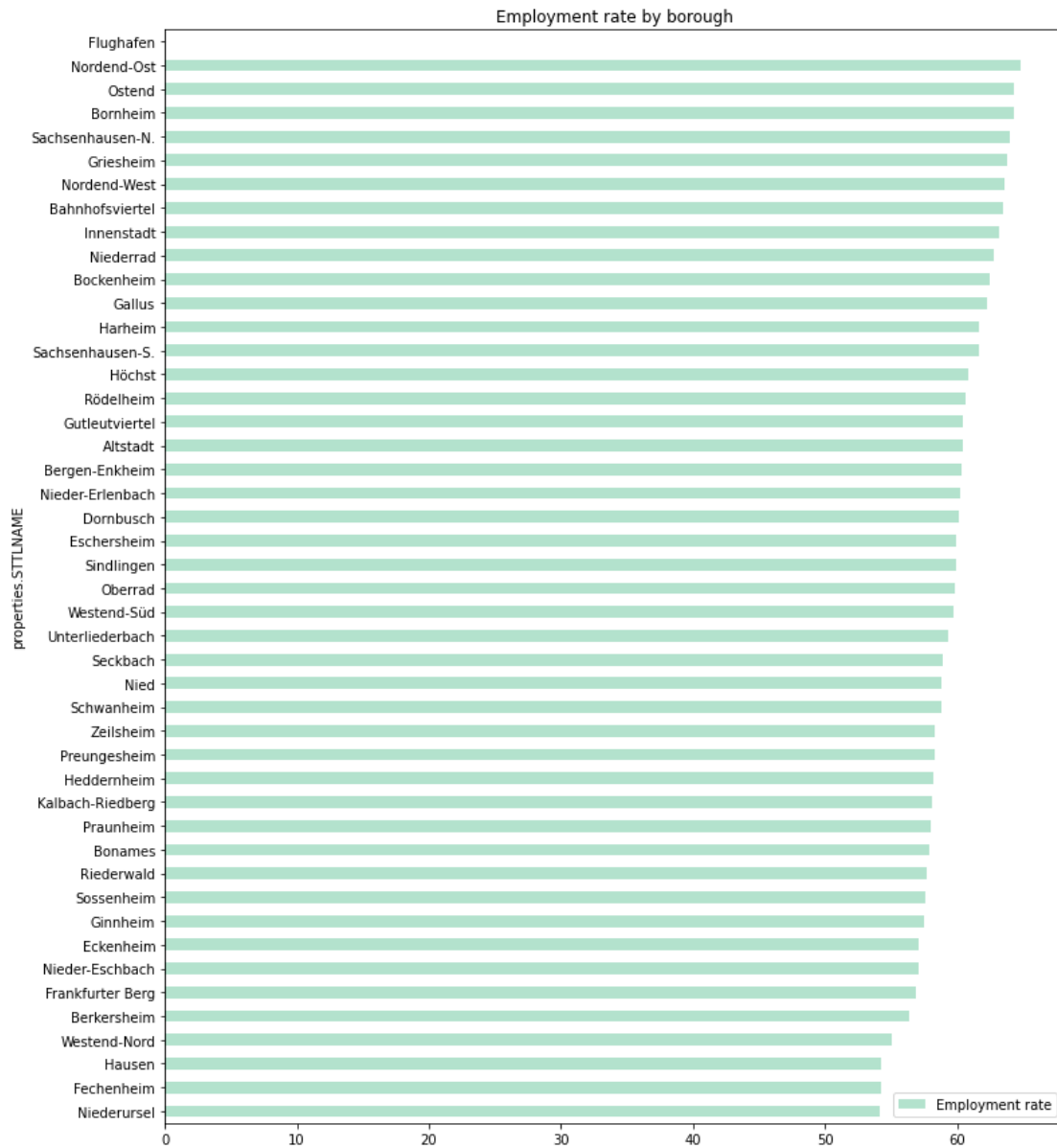


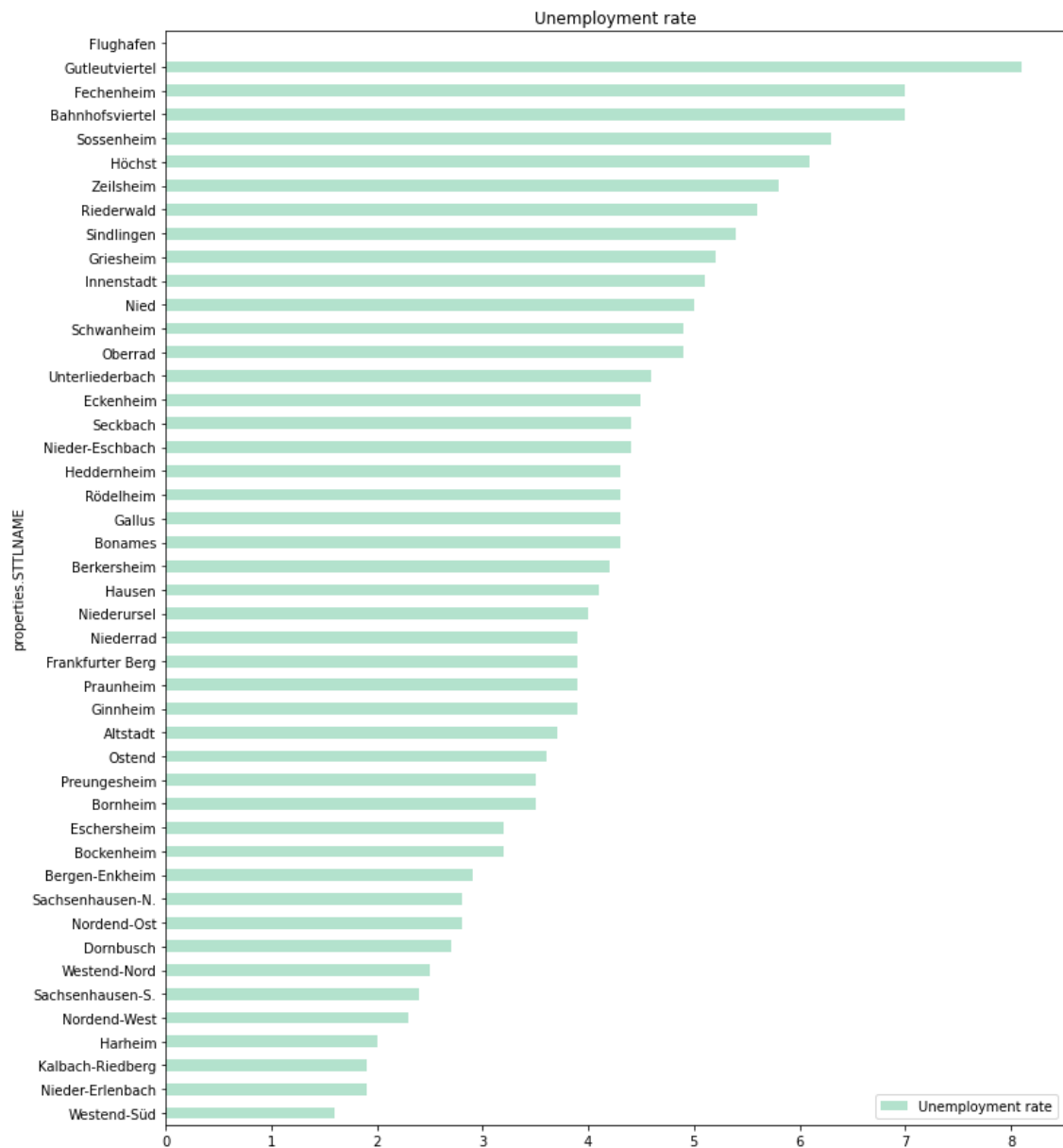
b) Detailed results



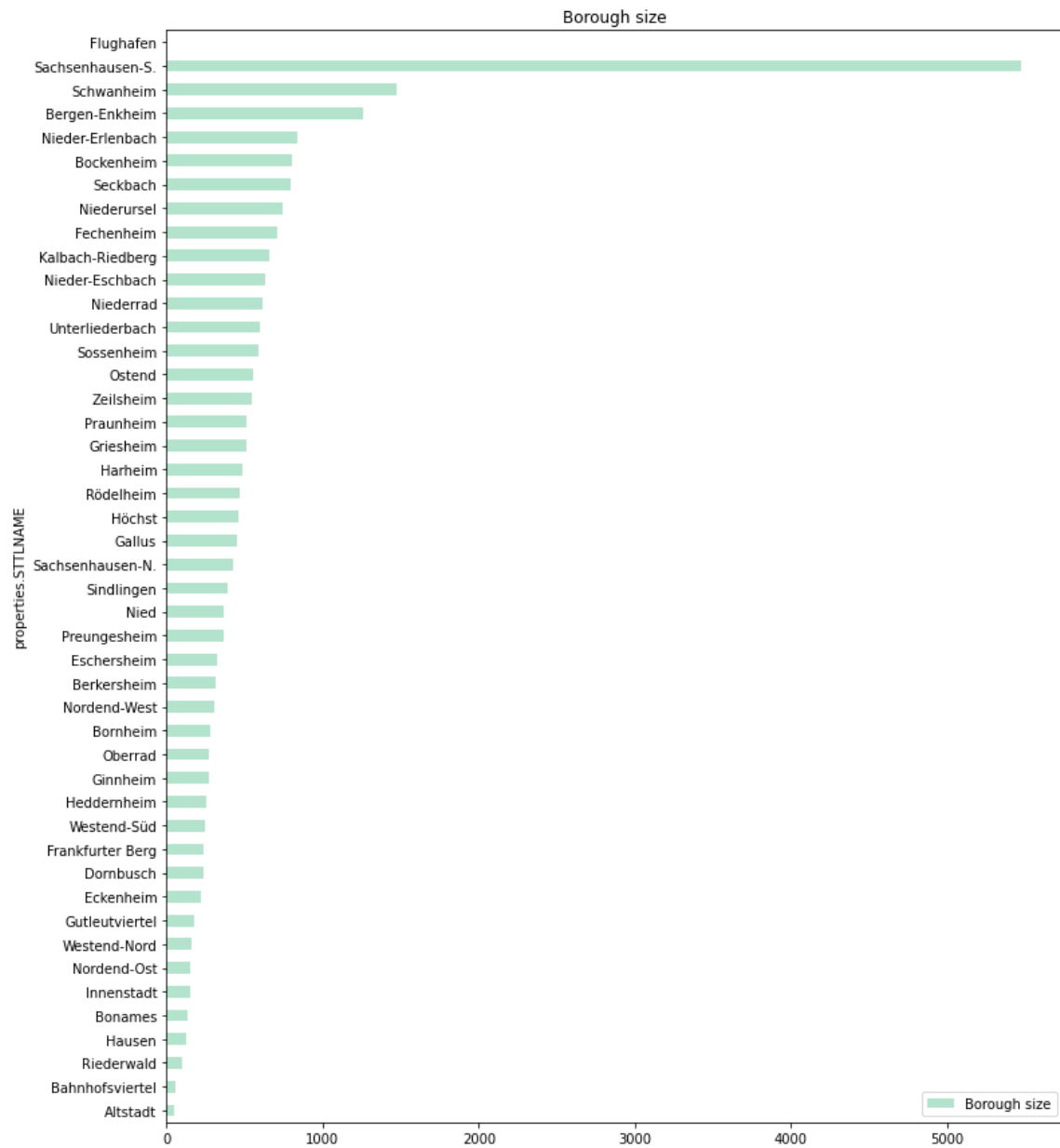


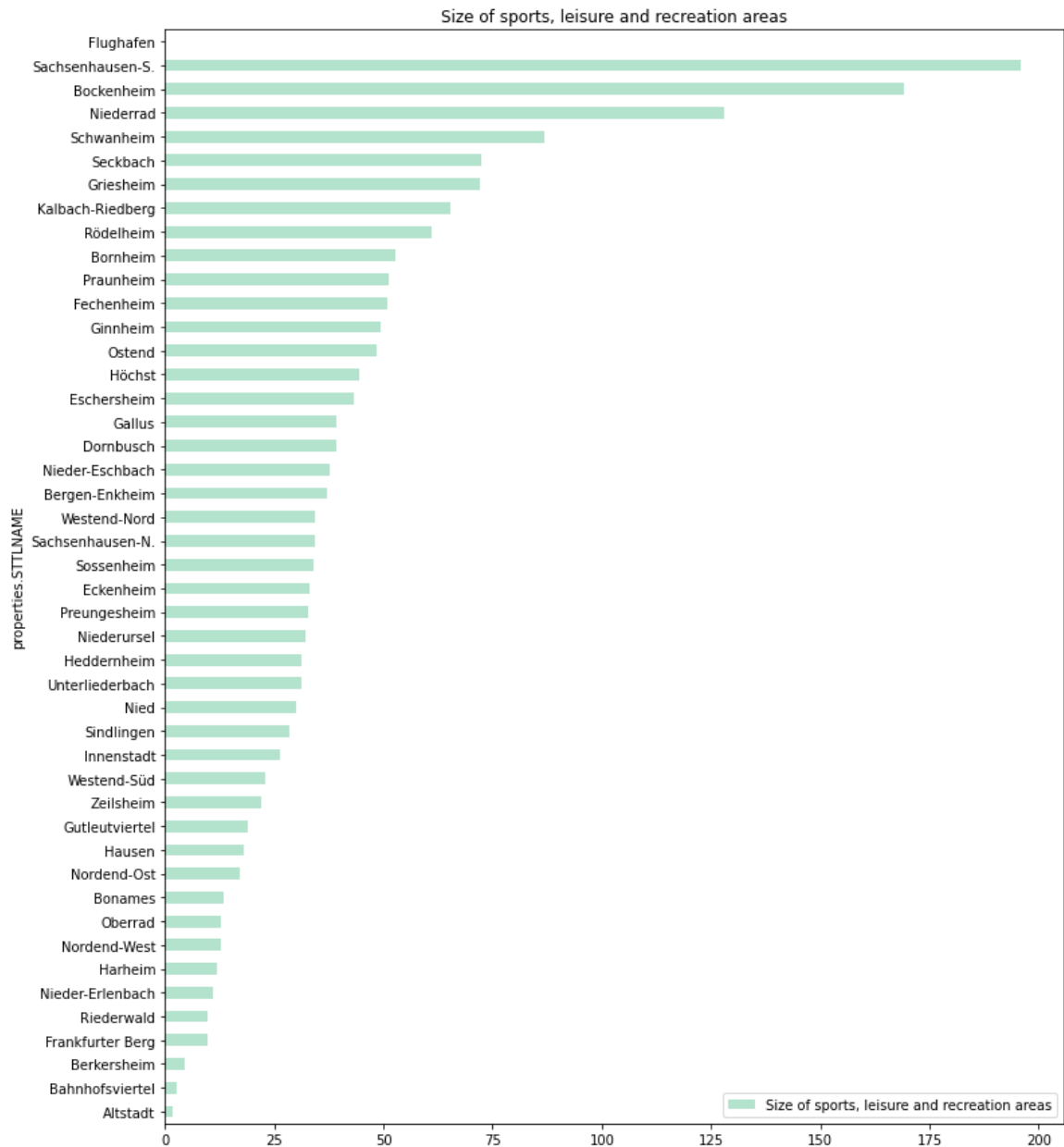


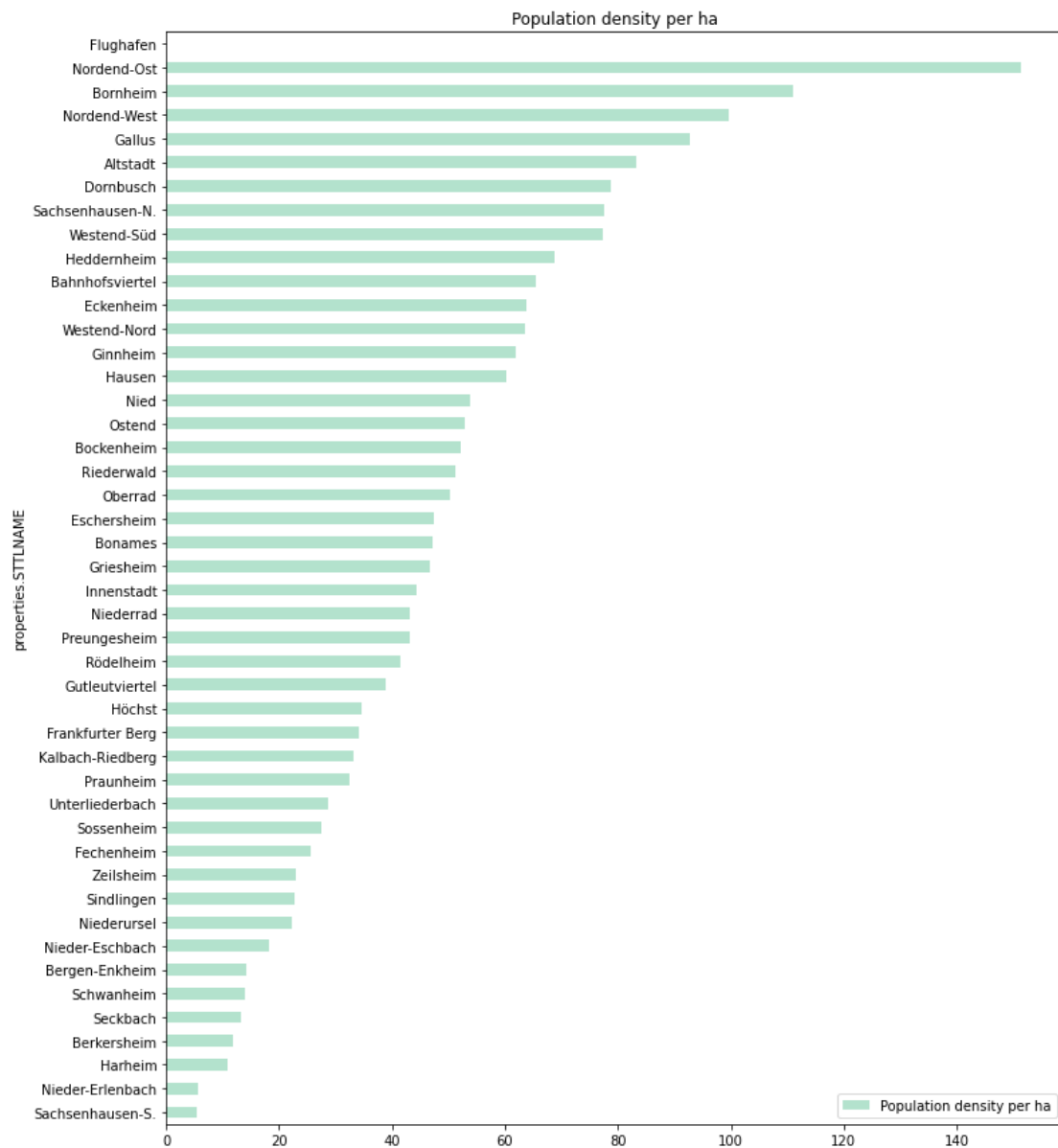


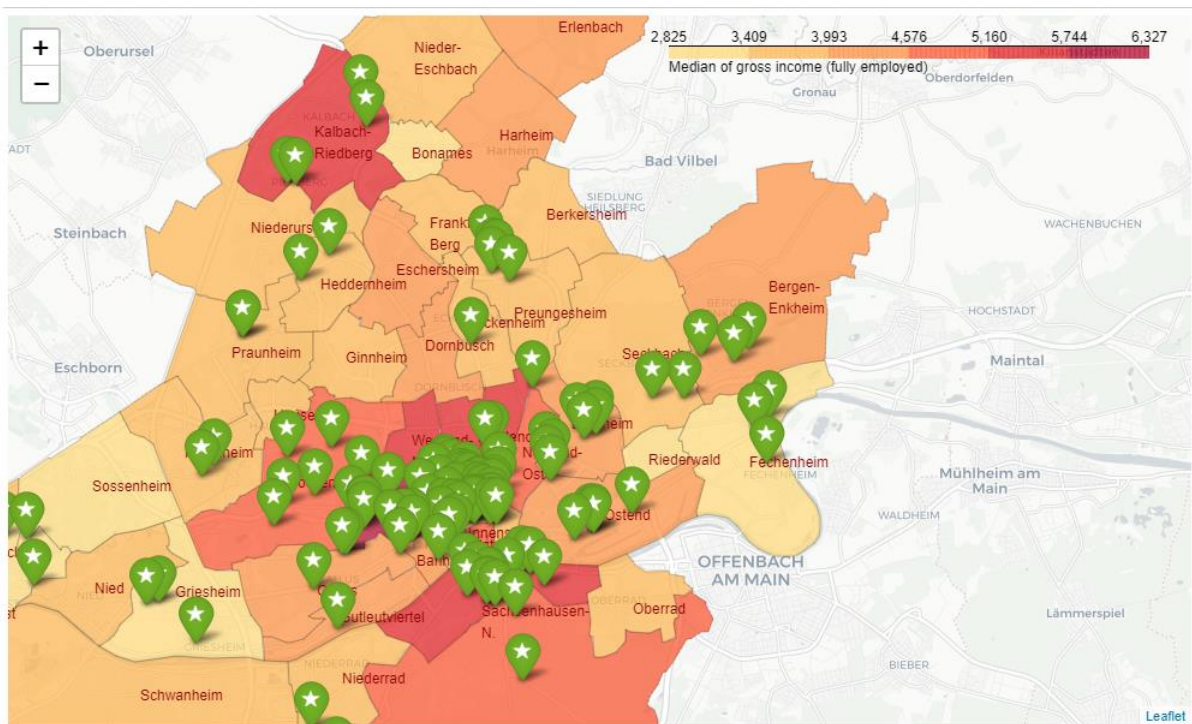
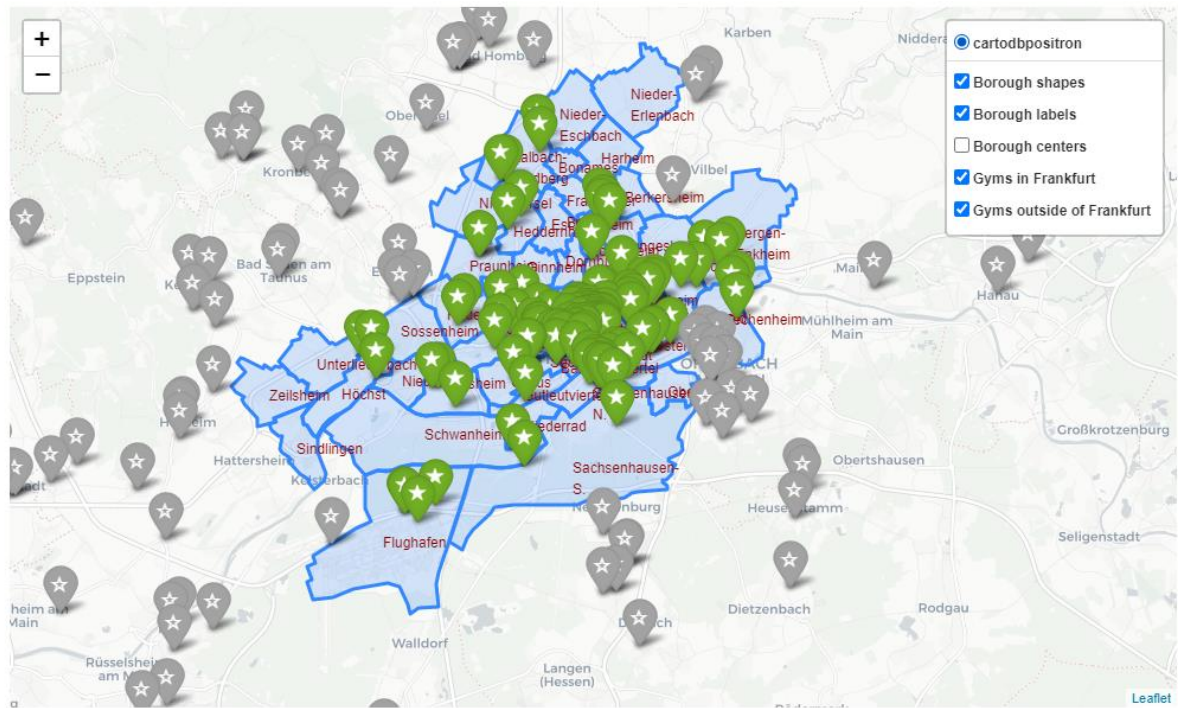


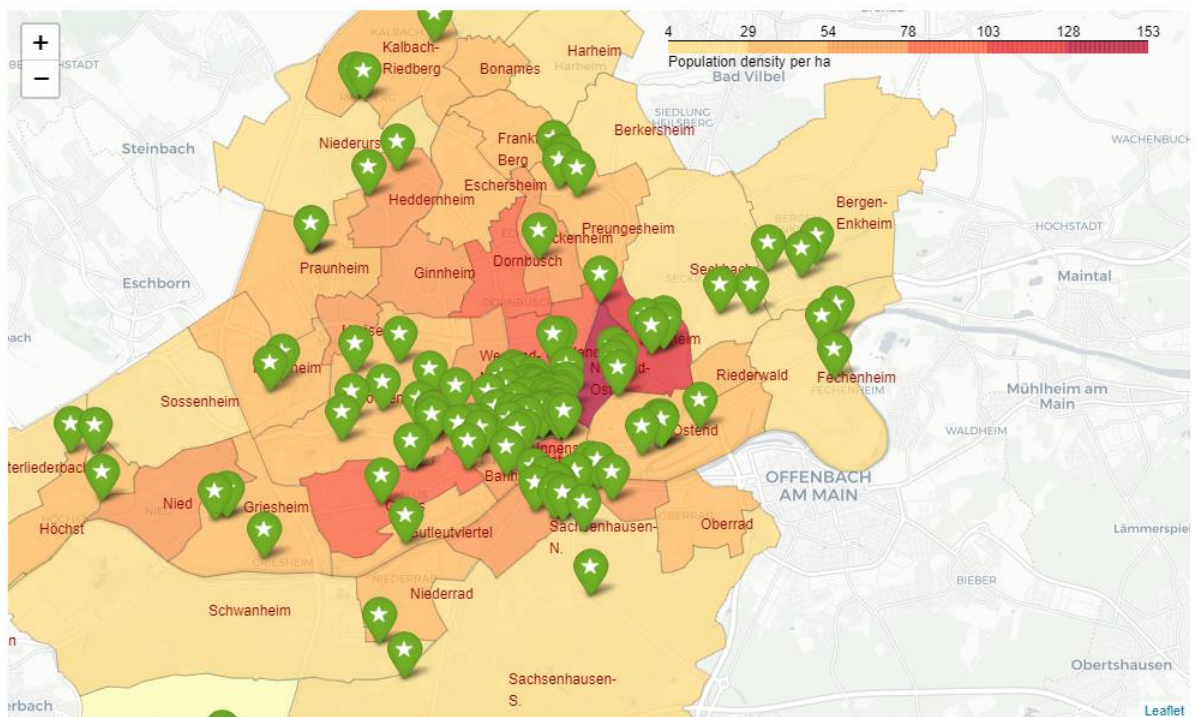
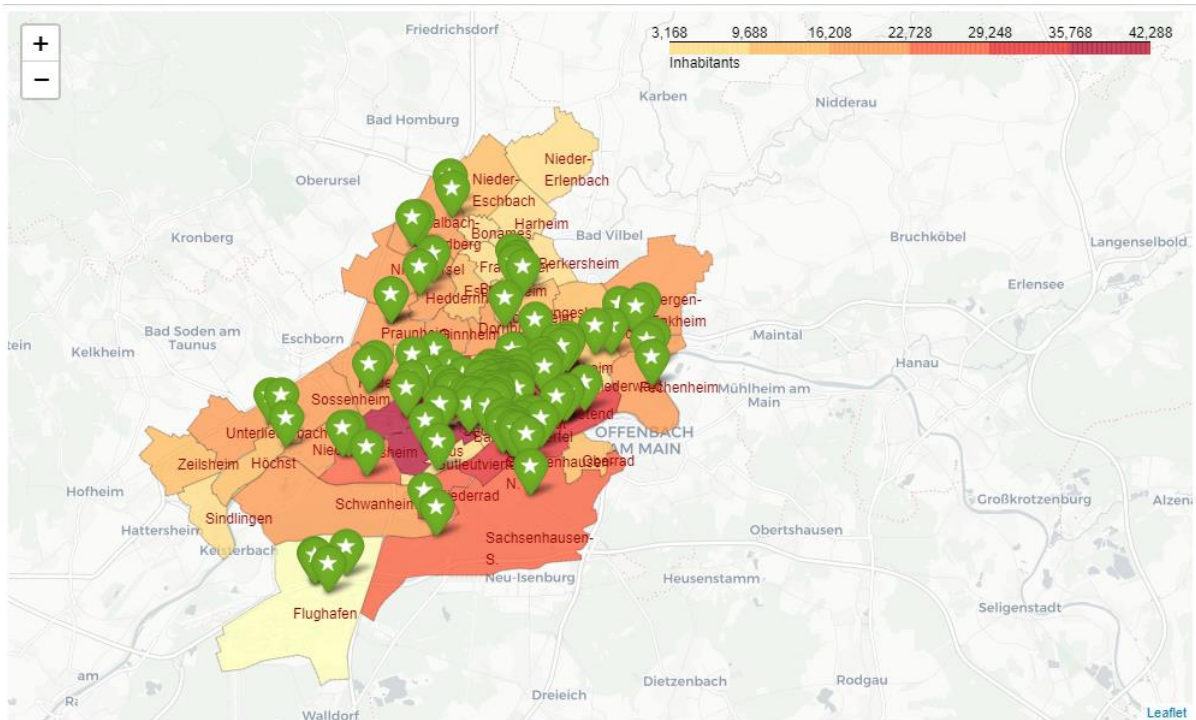


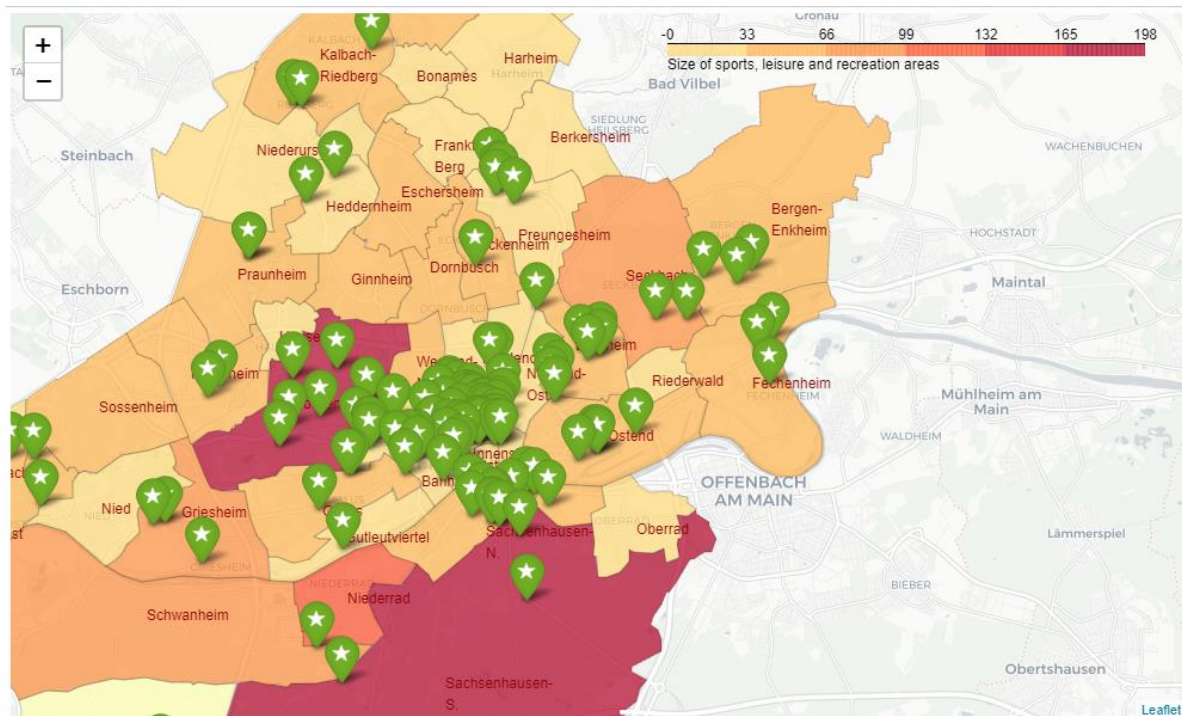












```

decision_tree_text = tree.export_text(decision_tree,
    feature_names=[
        'Borough size',
        'Size of sports, leisure and recreation areas',
        'Households',
        'Inhabitants',
        'female',
        'male',
        'People per household',
        'Population density per ha',
        'Average age',
        'Employment total',
        'Employment rate',
        'Unemployment rate',
        'Median of gross income (fully employed)'
    ],
    max_depth=10
)
print(decision_tree_text)

```

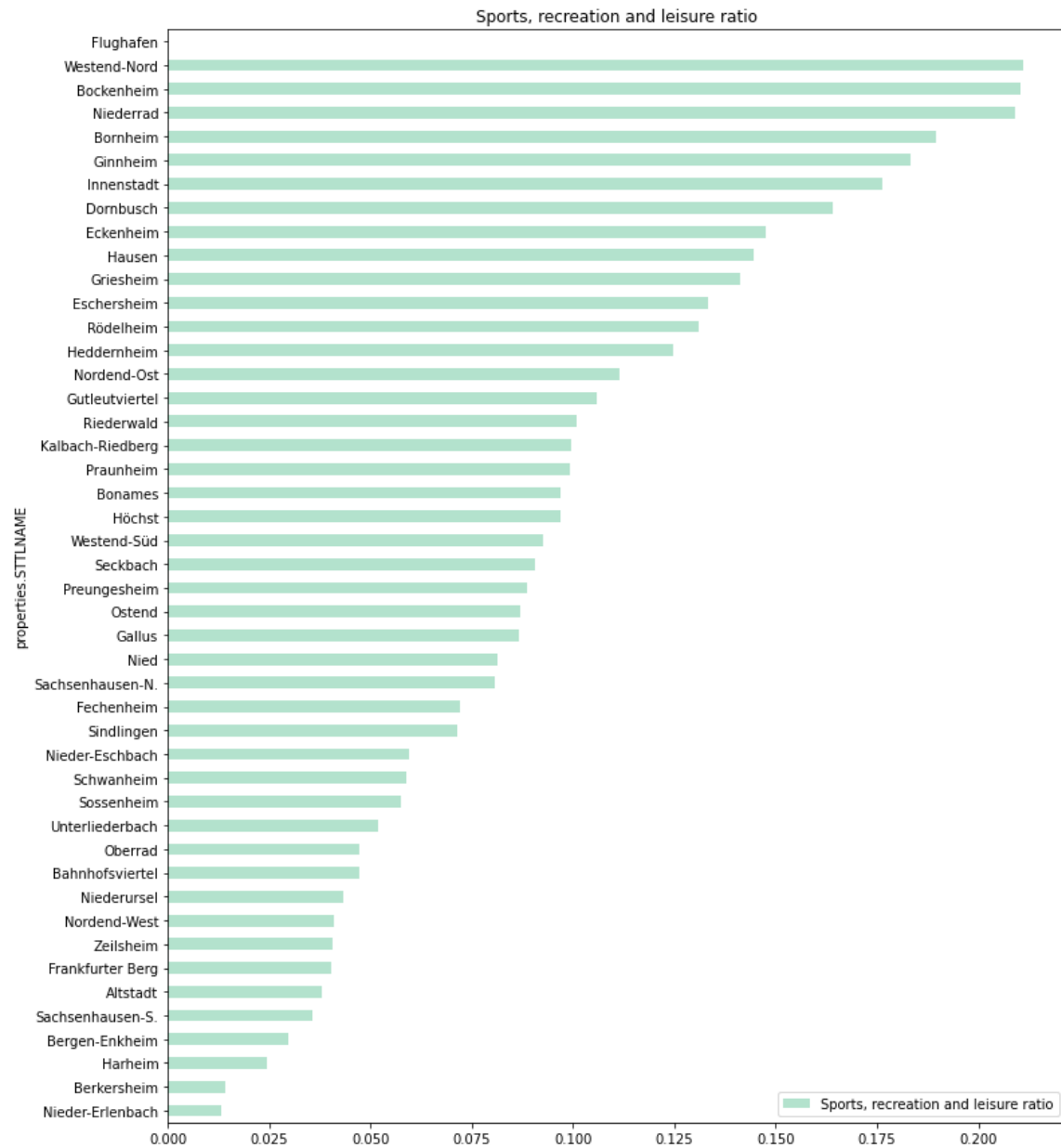
```

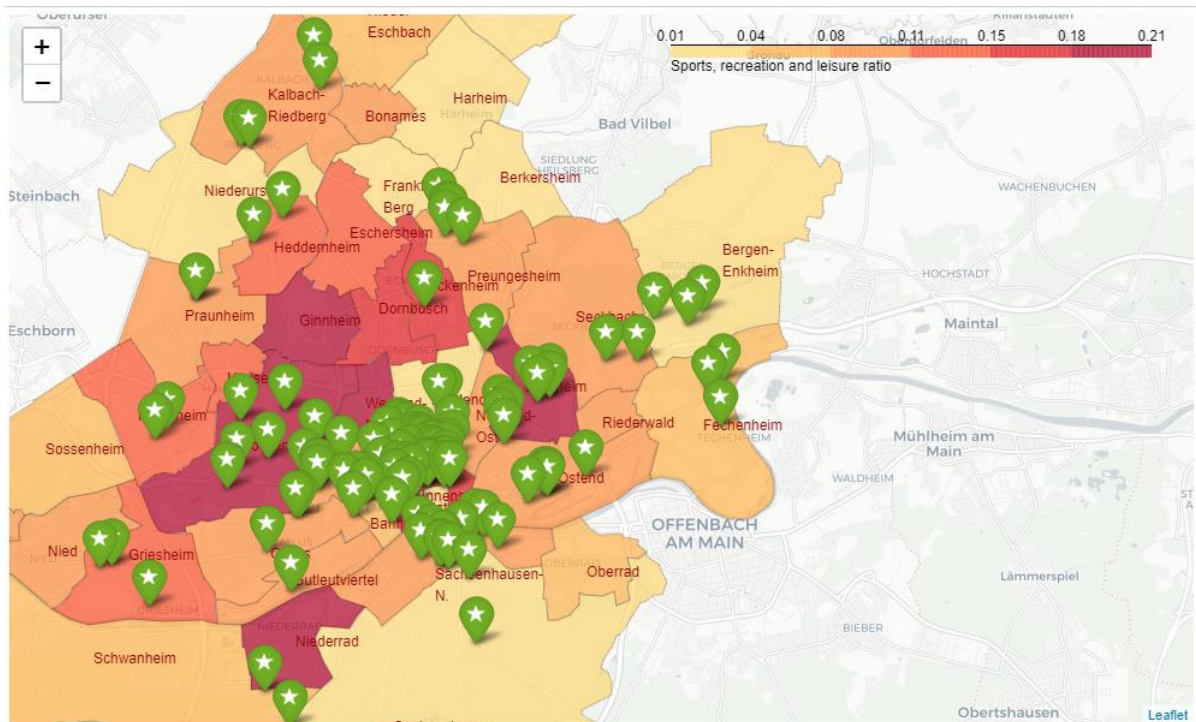
|--- Size of sports, leisure and recreation areas <= 31.28
|   |--- People per household <= 1.79
|   |   |--- female <= 3079.50
|   |   |   |--- class: Low (1-4)
|   |   |--- female > 3079.50
|   |   |   |--- Employment total <= 15546.00
|   |   |   |   |--- class: High (10 and more)
|   |   |   |--- Employment total > 15546.00
|   |   |   |   |--- Employment rate <= 64.20
|   |   |   |   |   |--- class: Medium (5-10)
|   |   |   |   |--- Employment rate > 64.20
|   |   |   |   |   |--- class: None (0)
|   |   |--- People per household > 1.79
|   |   |   |--- class: None (0)
|--- Size of sports, leisure and recreation areas > 31.28
|   |--- People per household <= 1.77
|   |   |--- Median of gross income (fully employed) <= 4936.13
|   |   |   |--- class: Medium (5-10)
|   |   |   |--- Median of gross income (fully employed) > 4936.13
|   |   |   |   |--- class: High (10 and more)
|   |   |--- People per household > 1.77
|   |   |   |--- Median of gross income (fully employed) <= 3625.26
|   |   |   |   |--- class: Low (1-4)
|   |   |   |--- Median of gross income (fully employed) > 3625.26
|   |   |   |   |--- Population density per ha <= 42.33
|   |   |   |   |   |--- Median of gross income (fully employed) <= 3626.22
|   |   |   |   |   |   |--- class: None (0)
|   |   |   |   |   |--- Median of gross income (fully employed) > 3626.22
|   |   |   |   |   |   |--- class: Low (1-4)
|   |   |   |--- Population density per ha > 42.33
|   |   |   |   |--- Average age <= 39.80
|   |   |   |   |   |--- Average age <= 39.35
|   |   |   |   |   |   |--- class: Low (1-4)
|   |   |   |   |   |--- Average age > 39.35

```

As a result, the **size of a boroughs' sports, leisure and recreation areas** seems to impact the competition level significantly when using 'Gym competition' as target variable. Hence a ratio of those areas compared to the overall borough size is going to be computed for subsequent analyses.
 Besides, as might expected, some additional variables seem to be relevant:

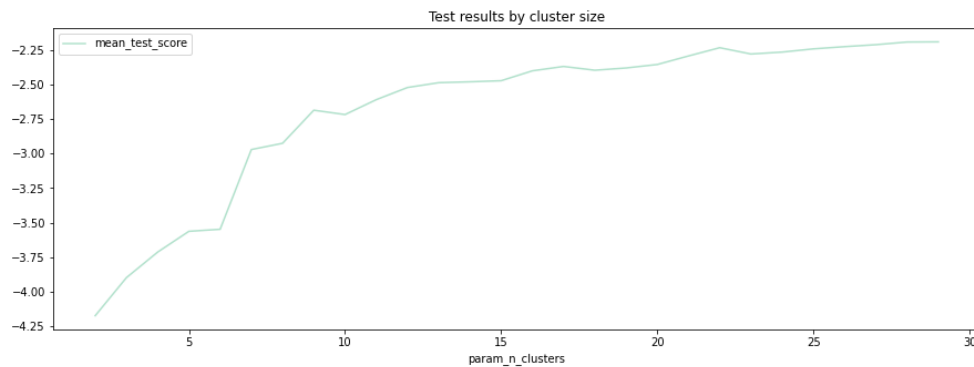
- households
- people per household
- employment
- median gross income
- population density
- average age

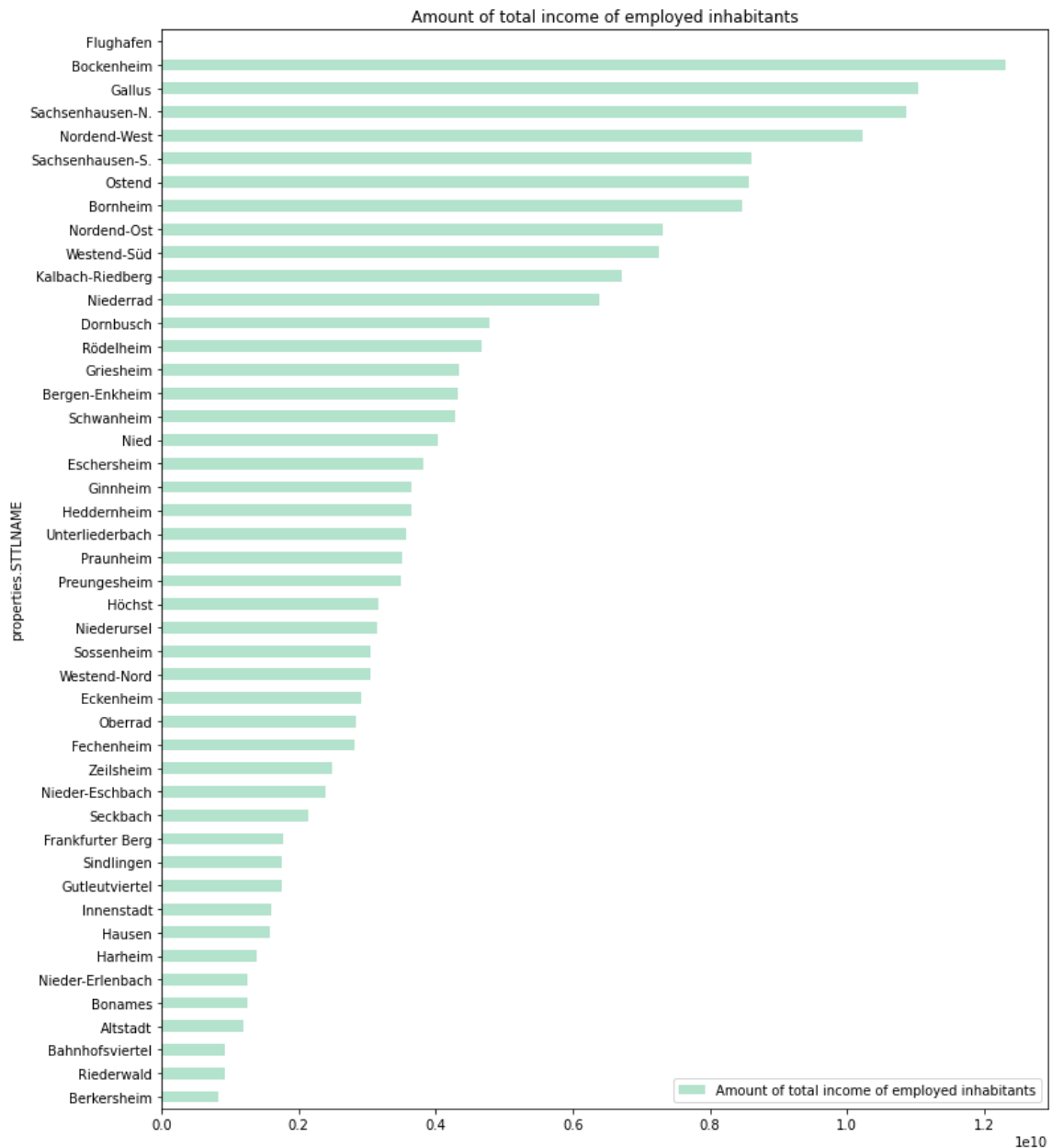




Visualize gridsearch results to determine how many clusters will be used in final KMeans algorithm

```
In [60]: line_chart(results_gridsearch,x_val='param_n_clusters',y_val=['mean_test_score'],title='Test results by cluster size',sortcol='param_n_clusters',figsize_val=(15,5))
```



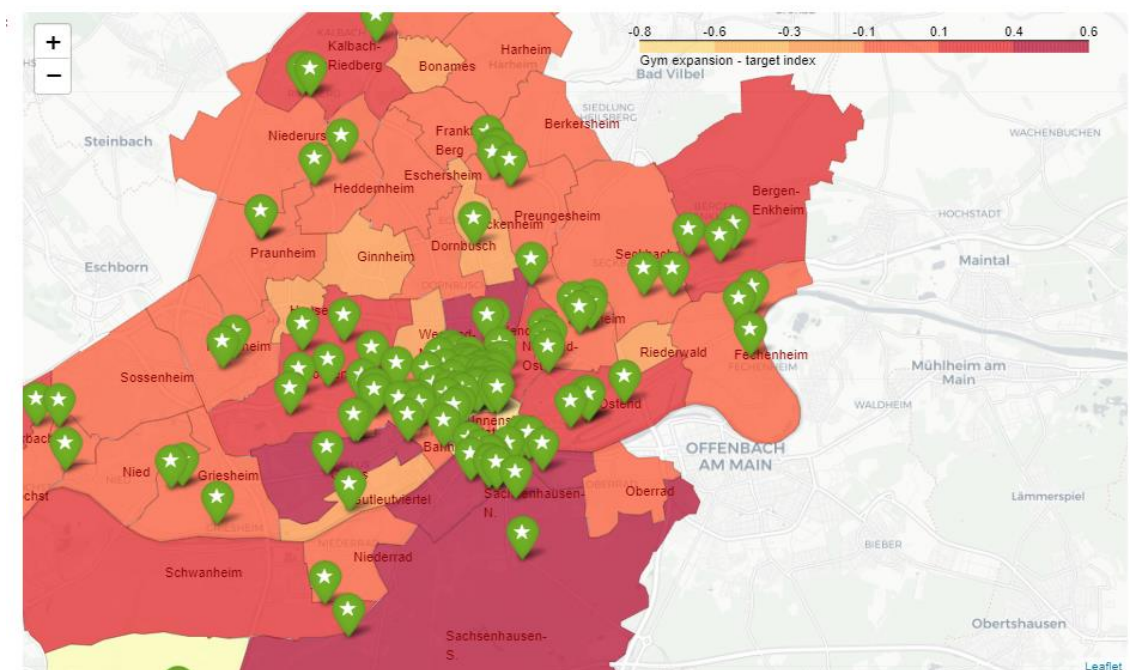




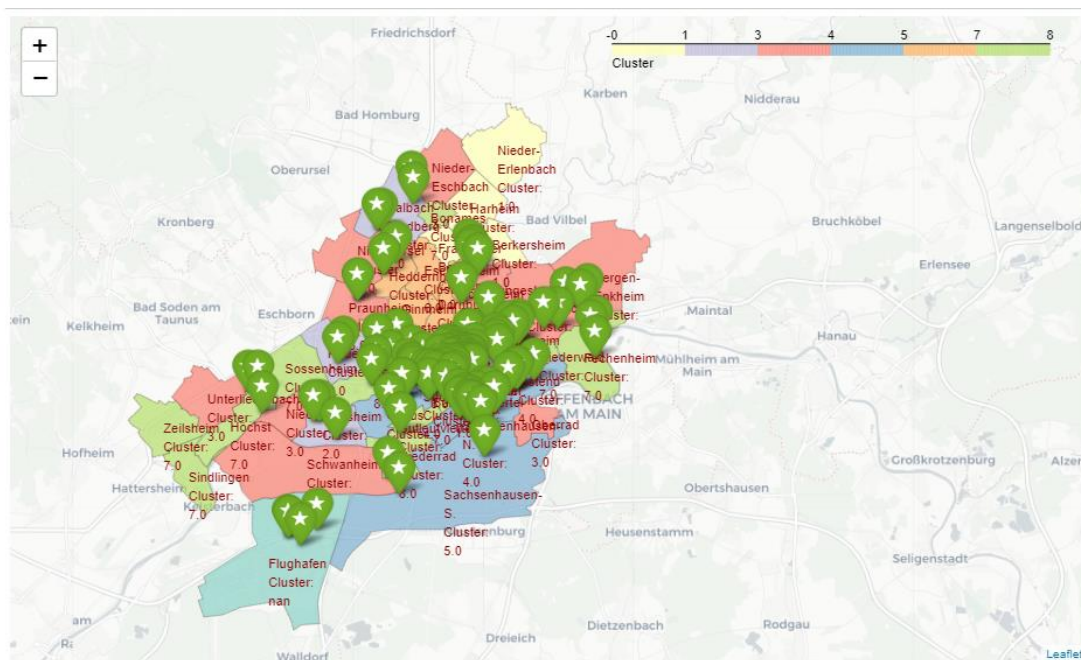
Statistics for the top10 target boroughs, here based on the target expansion index:

Out[74]:

	0	1	2	3	4	5	6
properties.STTLNAME	Gallus	Nordend-West	Sachsenhausen-S.	Sachsenhausen-N.	Ostend	Bockenheim	Nordend-Ost
Gym expansion - target index	0.600595	0.552259	0.546423	0.418095	0.352567	0.342258	0.32975
centroid.latitude	50.102319	50.130401	50.070682	50.098492	50.11622	50.122213	50.12757
centroid.longitude	8.638973	8.685188	8.674516	8.685992	8.721066	8.630767	8.697659
Number of gyms	4.0	8.0	3.0	12.0	6.0	7.0	0.0
Gym competition	Low (1-4)	Medium (5-10)	Low (1-4)	High (10 and more)	Medium (5-10)	Medium (5-10)	None (0)
Borough size	451.730563	310.017541	5471.102593	423.463536	556.403464	803.075235	153.174038
Size of sports, leisure and recreation areas	39.183601	12.689284	195.848676	34.230558	48.466343	169.016658	17.083692
Sports, recreation and leisure ratio	0.086741	0.040931	0.035797	0.080835	0.087106	0.210462	0.111531
Population density per ha	92.645934	99.662103	5.328176	77.496637	52.977744	52.17942	151.344186
Inhabitants	41851.0	30897.0	29151.0	32817.0	29477.0	41904.0	23182.0
female	19824.0	15845.0	15001.0	16845.0	15108.0	20948.0	12045.0
male	22027.0	15052.0	14150.0	15972.0	14369.0	20956.0	11137.0
Average age	36.7	40.5	43.5	40.3	41.9	38.4	41.1
Households	23069.0	18568.0	16506.0	19097.0	17619.0	24065.0	14631.0
People per household	1.85	1.69	1.79	1.75	1.67	1.76	1.61
Employment rate	62.3	63.6	61.7	64.0	64.3	62.5	64.8
Unemployment rate	4.3	2.3	2.4	2.8	3.6	3.2	2.8
Median of gross income (fully employed)	4235.34252	5206.382353	4787.166667	5171.768657	4516.125	4700.5	4867.905063
Amount of total income of employed inhabitants	11042881823.000393	10230797477.541176	8610277912.350002	10862203648.477613	8559732808.987499	12310609500.0	7312535831.483544
Cluster	4.0	4.0	5.0	4.0	4.0	8.0	2.0
(Gym types, Boxing Gym)	NaN	NaN	NaN	NaN	NaN	NaN	NaN
(Gym types, Climbing Gym)	NaN	NaN	NaN	NaN	NaN	NaN	NaN
(Gym types, Gym)	2.0	5.0	NaN	5.0	1.0	3.0	NaN
(Gym types, Gym / Fitness Center)	2.0	3.0	2.0	3.0	4.0	3.0	NaN
(Gym types, Gym Pool)	NaN	NaN	NaN	1.0	NaN	NaN	NaN
(Gym types, Gymnastics Gym)	NaN	NaN	NaN	NaN	NaN	NaN	NaN
(Gym types, Martial Arts School)	NaN	NaN	NaN	NaN	NaN	NaN	NaN
(Gym types, Medical Center)	NaN	NaN	NaN	NaN	NaN	NaN	NaN
(Gym types, Pilates Studio)	NaN	NaN	NaN	NaN	NaN	NaN	NaN
(Gym types, Spa)	NaN	NaN	NaN	1.0	NaN	NaN	NaN



Results of clustering similar boroughs based on KMeans algorithm:



Excerpt of statistics for the 9 clusters based on cluster analysis (full details in notebook):

	Cluster	0.0	1.0	2.0	3.0	4.0	5.0
properties.STTLNAME	count	1	5	4	10	5	1
	unique	1	5	4	10	5	1
	top	Innenstadt	Nieder-Erlenbach	Griesheim	Bergen-Enkheim	Ostend	Sachsenhausen-S.
	freq	1	1	1	1	1	1
	mean	NaN	NaN	NaN	NaN	NaN	NaN
	std	NaN	NaN	NaN	NaN	NaN	NaN
	min	NaN	NaN	NaN	NaN	NaN	NaN
	25%	NaN	NaN	NaN	NaN	NaN	NaN
	50%	NaN	NaN	NaN	NaN	NaN	NaN
	75%	NaN	NaN	NaN	NaN	NaN	NaN
Gym expansion - target index	count	1.0	5.0	4.0	10.0	5.0	1.0
	unique	NaN	NaN	NaN	NaN	NaN	NaN
	top	NaN	NaN	NaN	NaN	NaN	NaN
	freq	NaN	NaN	NaN	NaN	NaN	NaN
	mean	-0.788249	0.038061	0.136428	0.087894	0.39507	0.546423
	std	NaN	0.032531	0.178842	0.100688	0.216273	NaN
	min	-0.788249	-0.015227	-0.049723	-0.086009	0.051835	0.546423
	25%	-0.788249	0.03353	0.005211	0.016653	0.352567	0.546423
	50%	-0.788249	0.04823	0.132843	0.124304	0.418095	0.546423
	75%	-0.788249	0.053695	0.26406	0.158318	0.552259	0.546423
Number of gyms	count	1.0	5.0	4.0	10.0	5.0	1.0
	unique	NaN	NaN	NaN	NaN	NaN	NaN
	top	NaN	NaN	NaN	NaN	NaN	NaN
	freq	NaN	NaN	NaN	NaN	NaN	NaN
	mean	22.0	0.2	2.0	1.4	8.8	3.0
	std	NaN	0.447214	1.414214	1.846545	4.147288	NaN
	min	22.0	0.0	0.0	0.0	4.0	3.0
	25%	22.0	0.0	1.5	0.0	6.0	3.0
	50%	22.0	0.0	2.5	1.0	8.0	3.0
	75%	22.0	0.0	3.0	2.0	12.0	3.0
Gym competition	count	1	5	4	10	5	1
	unique	1	2	2	3	3	1
	top	High (10 and more)	None (0)	Low (1-4)	Low (1-4)	High (10 and more)	Low (1-4)
	freq	1	4	3	5	2	1
	mean	NaN	NaN	NaN	NaN	NaN	NaN
	std	NaN	NaN	NaN	NaN	NaN	NaN
	min	NaN	NaN	NaN	NaN	NaN	NaN

5) Discussion

Several points should be considered when reviewing the research, and/or adapting it for other locations:

- Removal of duplicates of gym locations was essential since the retrieval process using the Foursquare API has rate limits. Using the approach to get data for each borough ensures data completeness, but yields many duplicates which have to be cleaned.
- The decision tree in this research was only intended as descriptive measure to aid the feature selection. It is not intended to predict values.
- The dataset contains a lot of aggregate data on a per-borough base. Hence algorithms may not be suitable to predict new values necessarily if the borough shapes would be changed officially in future.
- The statistical data and the location data are about 1 year apart. This is due to the limited availability of real time information for the official statistical data.
- The gym expansion target index could be enhanced on a theoretical level if more detailed data would be available. However, given the constraints of the existing data and the sanity check using two different methods, it seems reasonable as an indicator for gym expansions / setups.
- The target index and clustering does not consider the cost of setting up new gyms. For example, the analyses could be improved in future if per-borough rental information would be available.
- It would be advisable to run additional research either on a qualitative and/or quantitative basis. One could image running surveys targeting gym managers and/or gym members. Such research would need to be thought through in detail and will likely have some methodological constraints itself as well.

6) Conclusion

Given the existing research findings, it is advisable to inspect the following boroughs in more detail when aiming to launch new gyms or expand gym franchises:

- Gallus
- Nordend-West
- Sachsenhausen Süd
- Sachsenhausen Nord
- Ostend