



Applied Data Science Capstone by IBM/Coursera

“The Battle of Hamburg’s Districts for a new Study Café”

Capstone project by Karima Chakroun

March 9, 2019

Overview

1. Introduction

2. Data

3. Methodology

3.1. Data exploration

3.2. Clustering analysis

3.3. Data visualization using Folium

4. Results

5. Discussion

6. Conclusion

1. Introduction



Background:

- Hamburg is a major science and education metropole with several universities and over 100 000 students ^[1-3]
- however, university buildings and campuses only offer limited space and are often overcrowded, old, and decrepit ^[4]
- clear need for more off-campus student venues, e.g. study cafés

Business problem: find the optimal location for opening a new study café

- near one or more universities
- in a lively district with other food & drink venues, but few cafés
- in a district with low to moderate rental prices

Interested parties: any private and public investors (e.g. universities)

2. Data

Data from different sources were collected and cleaned:

- **University names:** scraped from <https://www.4icu.org/de/hamburg/>
- **University locations:** from geopy client (Nominatim) & Google maps
- **District names and geographical borders:** from geojson file on github ^[5]
- **District center locations:** from geopy client (Nominatim)
- **District venues:** from the Foursquare API (explore?, limit=100, radius=500m)
- **Rental prices:** scraped from <https://www.4icu.org/de/hamburg/>,
district names were adjusted to match the names in the geojson file

2. Data

Resulting dataframes:

- **Universities:** (18 rows)

	University	Lat	Lon
0	Universität Hamburg	53.480616	10.240777
1	Technische Universität Hamburg	53.461007	9.969227
2	Hochschule für Angewandte Wissenschaften Hamburg	53.493382	10.200562

- **District locations & prices:** (104 rows)

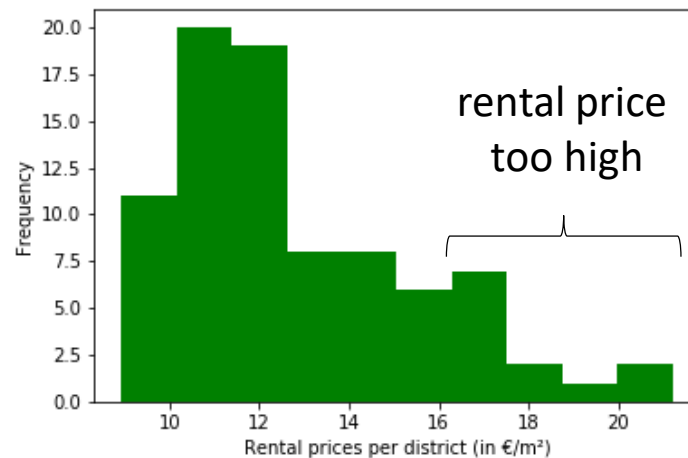
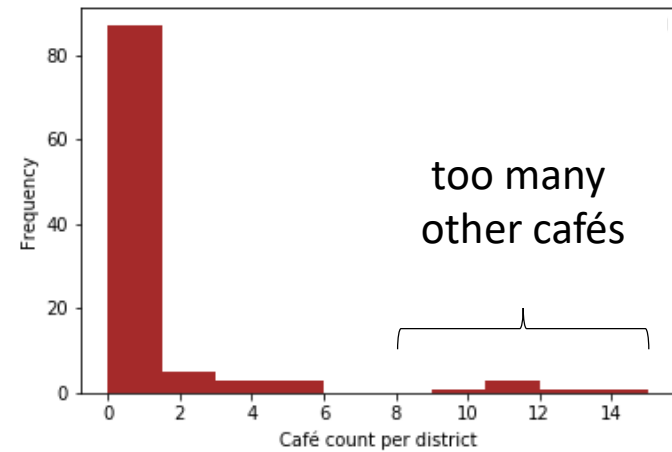
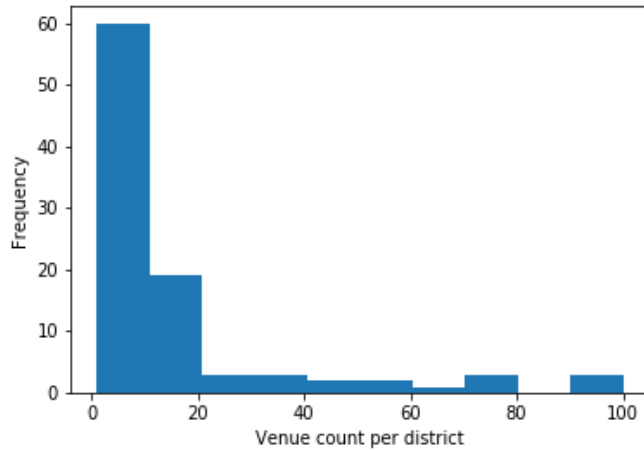
	District	Lat	Lon	Price
0	Allermöhe	53.483600	10.125000	11.21
1	Alsterdorf	53.610541	10.003889	14.04
2	Altengamme	53.429725	10.272787	NaN

- **District venues:** (1537 rows)

	District	District Lat	District Lon	Venue	Venue Lat	Venue Lon	Venue Category
0	Alsterdorf	53.610541	10.003889	Eppendorfer Moor	53.613315	10.002277	Nature Preserve
1	Alsterdorf	53.610541	10.003889	REWE	53.607687	10.005800	Supermarket
2	Alsterdorf	53.610541	10.003889	Best Western Premier Alsterkrug Hotel	53.613080	9.999037	Hotel

3. Methodology

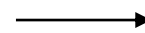
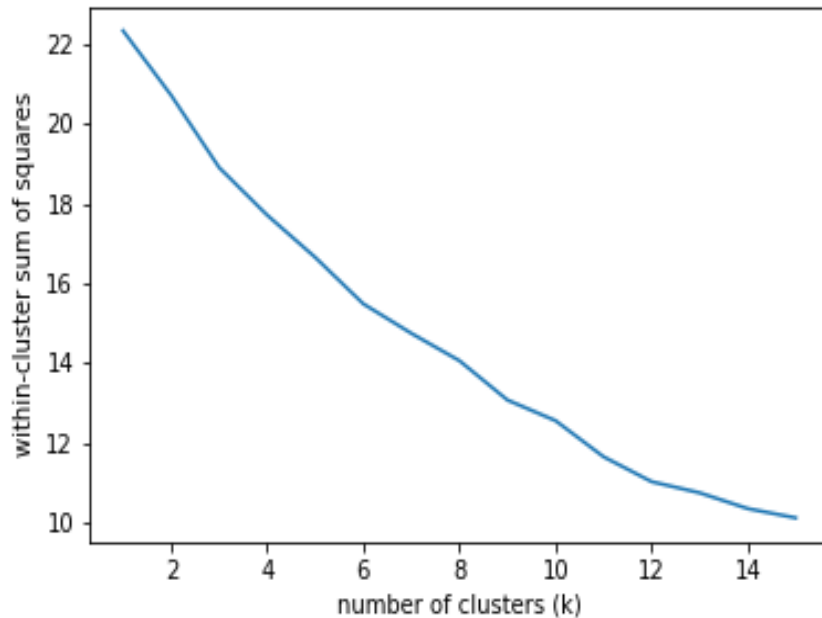
3.1. Data exploration (histograms)



3. Methodology

3.2. Clustering analysis

- use mean frequencies of venue categories per district for clustering
- finding the optimal k for k-means clustering (via sklearn^[6]) by elbow plot:



no clear elbow visible;
we use **k=5** based on the
rationale that we want enough
clusters to show the districts'
different venue characteristics,
but not too many clusters to
prevent overfitting

3. Methodology

3.2. Clustering analysis

- get cluster labels for each district and assign a name to each cluster (based on the 5 most common venue categories in that cluster):

Cluster 0: “Memorials” (1 district)

Cluster 1: “Supermarkets” (24 districts)

Cluster 2: “Outdoor & Recreation” (7 districts)

Cluster 3: “German Restaurants & Zoo” (4 districts)

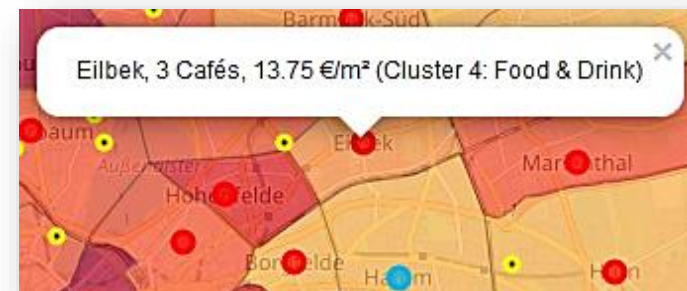
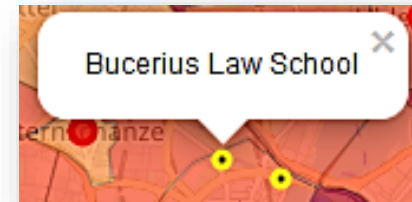
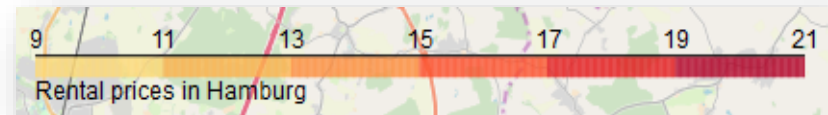
Cluster 4: “Food & Drink” (60 districts)

3. Methodology

3.3. Data visualization using Folium

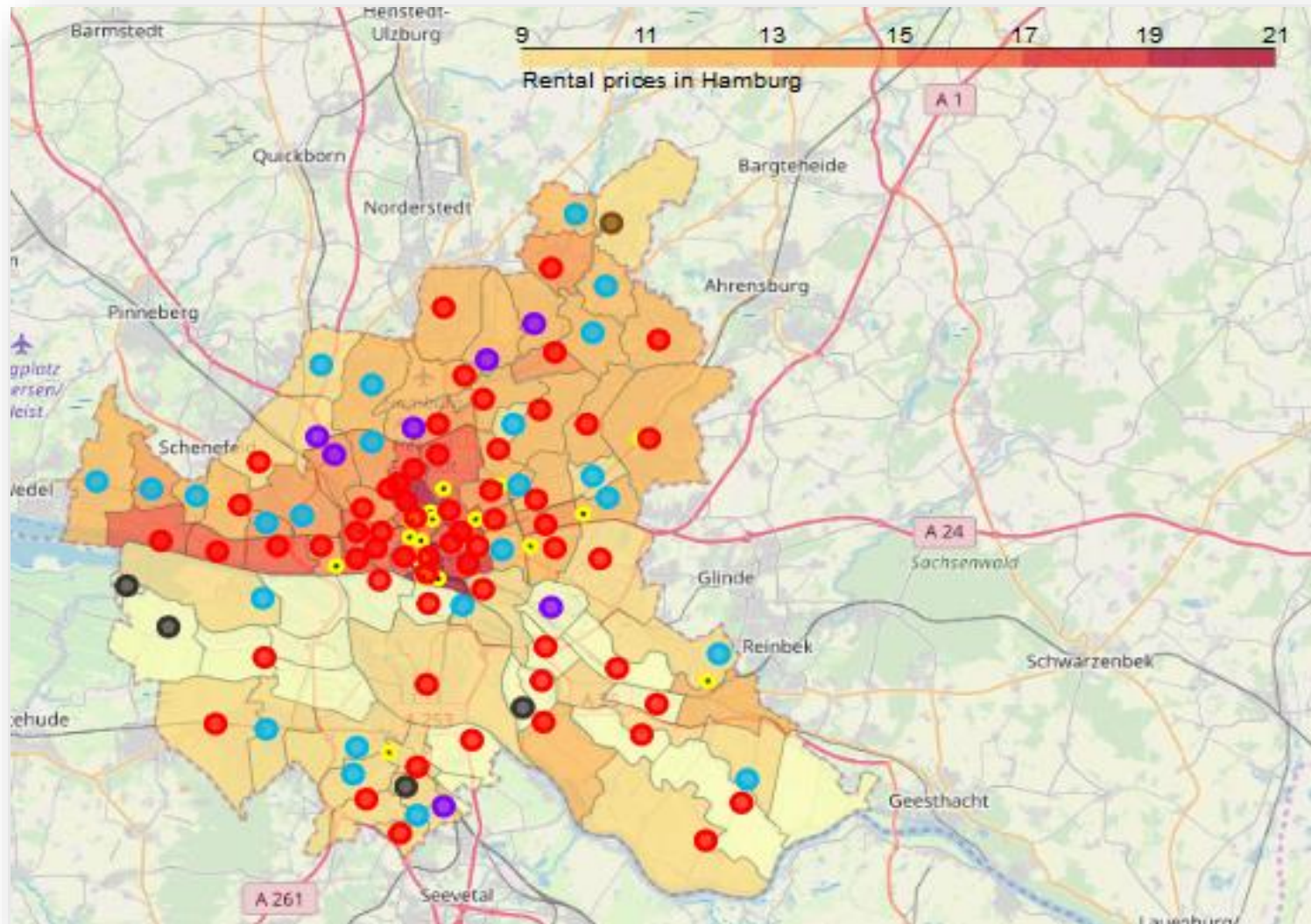
All data visualized in an interactive Folium^[7] map showing:

- rental prices of Hamburgs districts as a **choropleth map**
- locations of Hamburg's universities **marked by a small yellow circle** with a popup label showing its name
- locations of Hamburg's districts **marked by a larger circle in cluster color** with a popup label showing all relevant information



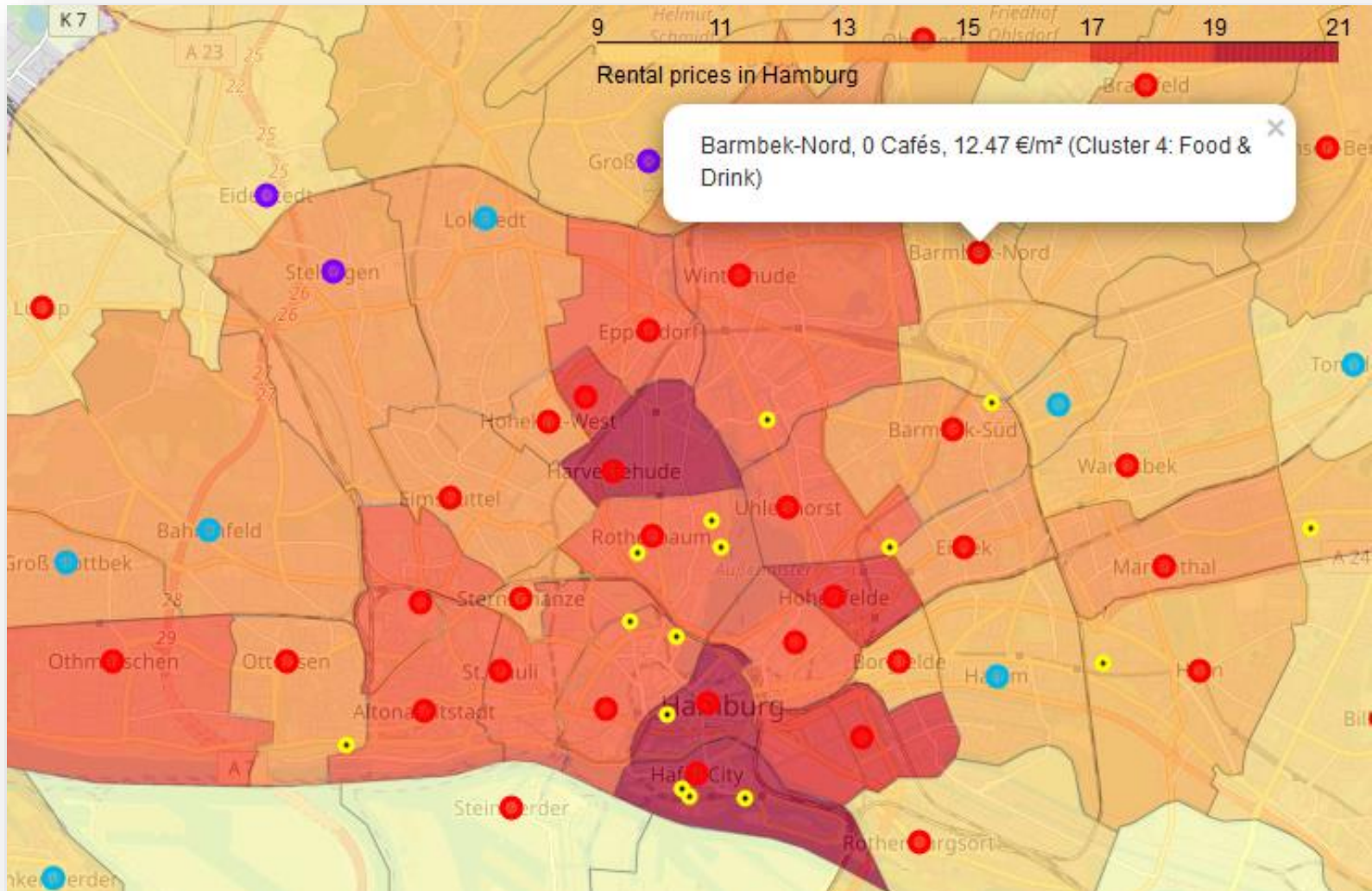
4. Results

Folium map of Hamburg (overview)



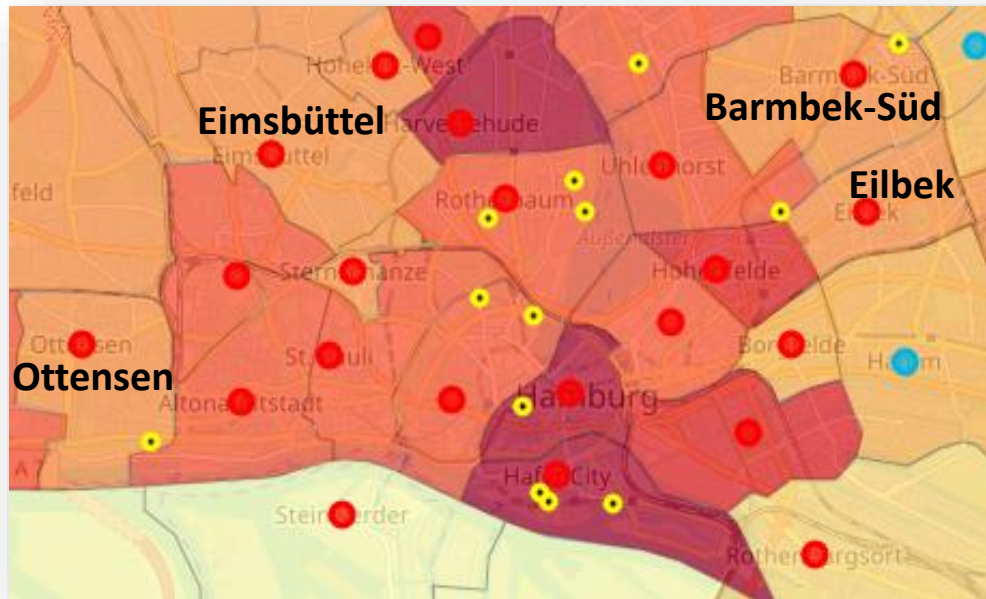
4. Results

Folium map of central Hamburg



5. Discussion

- 4 districts meet all the required criteria for the study café:



	District	Lat	Lon	Price	Cluster	ClusterName	Cafés
0	Barmbek-Süd	53.579885	10.043251	14.50	4	Food & Drink	0
1	Eilbek	53.567237	10.045241	13.75	4	Food & Drink	3
2	Eimsbüttel	53.572483	9.950100	14.12	4	Food & Drink	3
3	Ottensen	53.555066	9.919819	14.42	4	Food & Drink	2

6. Conclusion

Aim of this project: find the optimal location for opening a new study café by using an interactive Folium map of Hamburg showing several relevant data (rental prices, universities, districts and their venue characteristics)

4 districts recommended: [Barmbek-Süd](#), [Eimsbüttel](#), [Eilbek](#), [Ottensen](#)

All 4 districts meet criteria for opening a new study café:

- near one or more universities
- in a lively district with other food & drink venues, but few cafés
- in a district with low to moderate rental prices

Thank you!

References

- [1] <https://www.hamburg.de/buergermeisterreden-2017/9969574/hamburg-eine-metropole-der-wissenschaft-im-norden/>
- [2] <https://en.wikipedia.org/wiki/Hamburg>
- [3] <https://www.abendblatt.de/hamburg/article208791261/Hamburg-hat-so-viele-Studenten-wie-nie-zuvor.html>
- [4] <https://www.abendblatt.de/hamburg/article212674037/Rekordzahl-an-Studenten-in-Deutschland.html>
- [5] <https://gist.github.com/webtobesocial/935759ba975ffd9f6df6d1059fe5ad82/raw>
- [6] <https://scikit-learn.org/>
- [7] <https://pypi.org/project/folium/>