



# Expanding to Hamburg

Data Science Project for the Planned Expansion to Hamburg, Germany

# Overview

- Business Understanding
- Data Understanding
- Methodology
- Analysis
- Results
- Discussion and recommendation
- Conclusion

# 1. Business Understanding

## Target

- Entering the German market successfully with the PersonalCoffeeLovers Cafe.

## Background

- Germany is known as a core market in Europe and it is vital for the European success for our cafe brand. A preliminary analysis already found Hamburg as a suitable start. Hamburg offers a vivid coffee culture rooted back to the trade of the Hansa. Hamburg the second biggest city in Germany and is considered very livable. The people in Hamburg are known for appreciating quality. With our brand we offer the best and highest quality to a reasonable price. Hamburg is considered wealthy.

## Target Audience

- To solve this problem a Data Science team has to be formed. The team is expected to report to the company's management and to the shareholders to make a well informed and smart decision. The management expects rational problem solving for this project because it's crucial for entering the German market.

## Success Criteria

- The project is successful if a rational recommendation of a quarter of Hamburg is given. It should be considered to look for quarters with an already existing cafe culture and furthermore it should be affordable to start a business there (low price per m<sup>2</sup>) but reach a big customership (high demand).



## 2. Data

### Necessary Variables

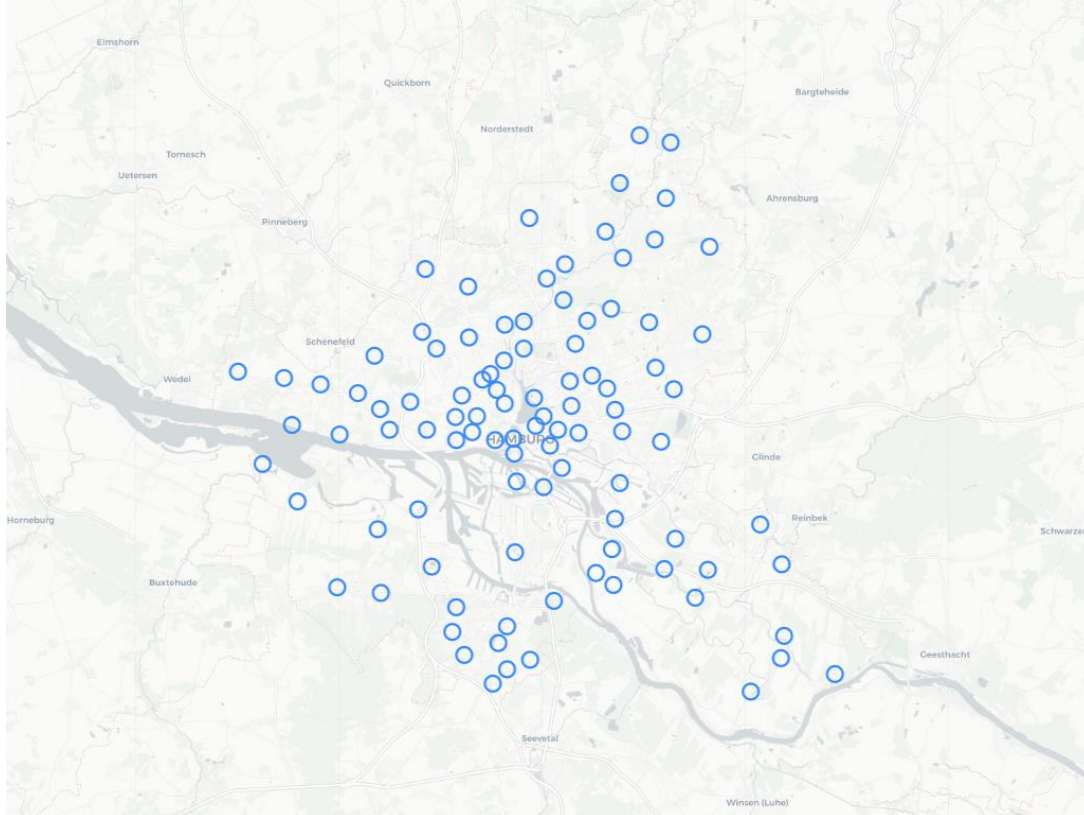
- Quarter
- Latitude
- Longitude
- Population: This could be an indicator for demand
- Price per  $m^2$  : This could be an indicator for the affordability
- Population under 18 in %
- Population over 64 in %
- Venues in the Quarter: This is needed to find the candidates for opening a cafe

### Sources:

- Foursquare API
- Transparezportal Hamburg
- Geocoder



## 2. Data



- 99 Quarters of Hamburg for further analysis

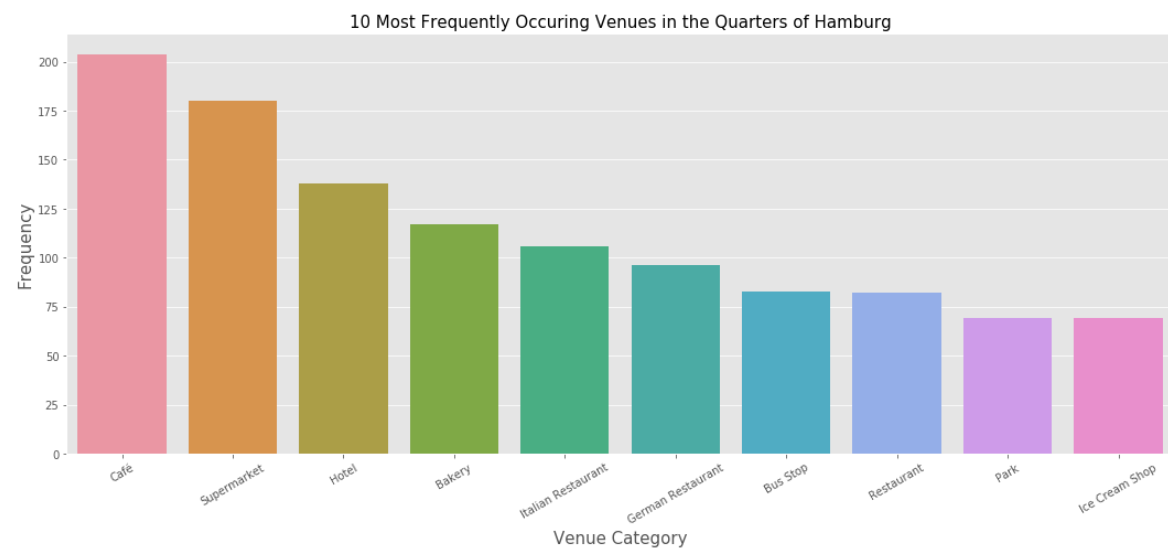
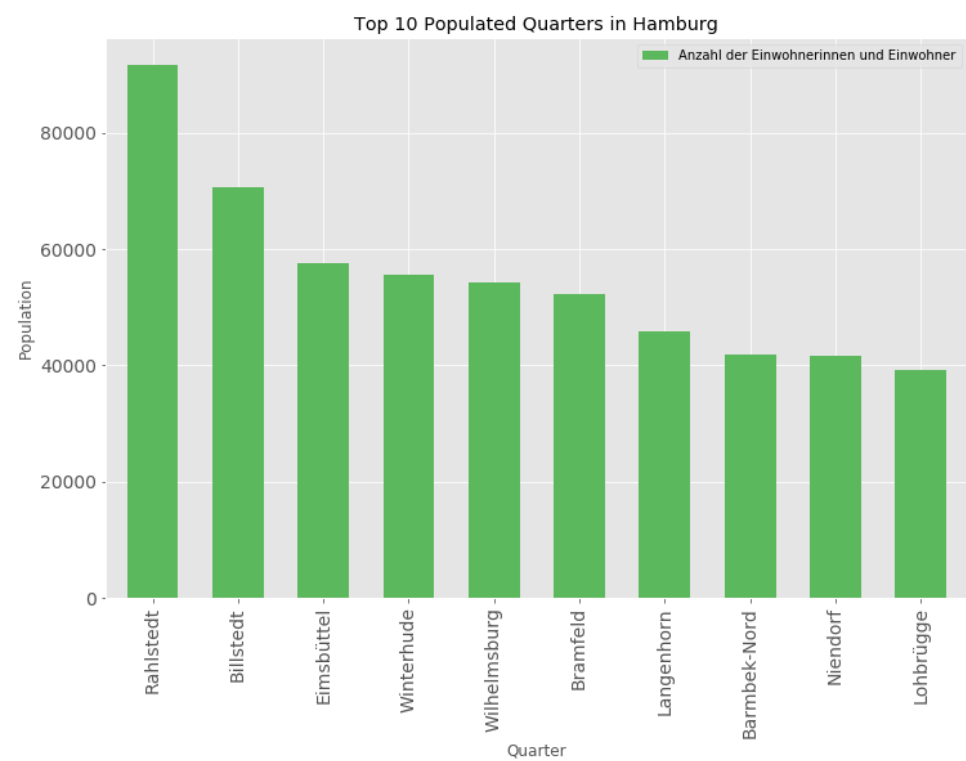
Quarter	Population in %	People under 18 years in %	People over 64 years in %	Price per m2 in EUR	latitude	longitude
Rahlstedt	4.895785	17.447630	22.276261	2990	53.603883	10.158155
Billstedt	3.765894	19.355250	17.927671	2932	53.548899	10.122625
Eimsbüttel	3.071965	12.832589	12.608663	4982	53.572483	9.950100
Winterhude	2.971062	12.529874	14.556791	5780	53.596390	10.003832
Wilhelmsburg	2.891942	20.742491	13.465266	3249	53.492292	9.998217

- Expected Dataframe

### 3. Methodology

- **Webscraping:** We use public available data and bring it in a structured format.
- **Foursquare:** Retrieving venue information
- **k-Means Clustering:** Segement and cluster similar quarters and group them based on common venues.
- **Data Visualization:** For a better understanding we visualize the results.

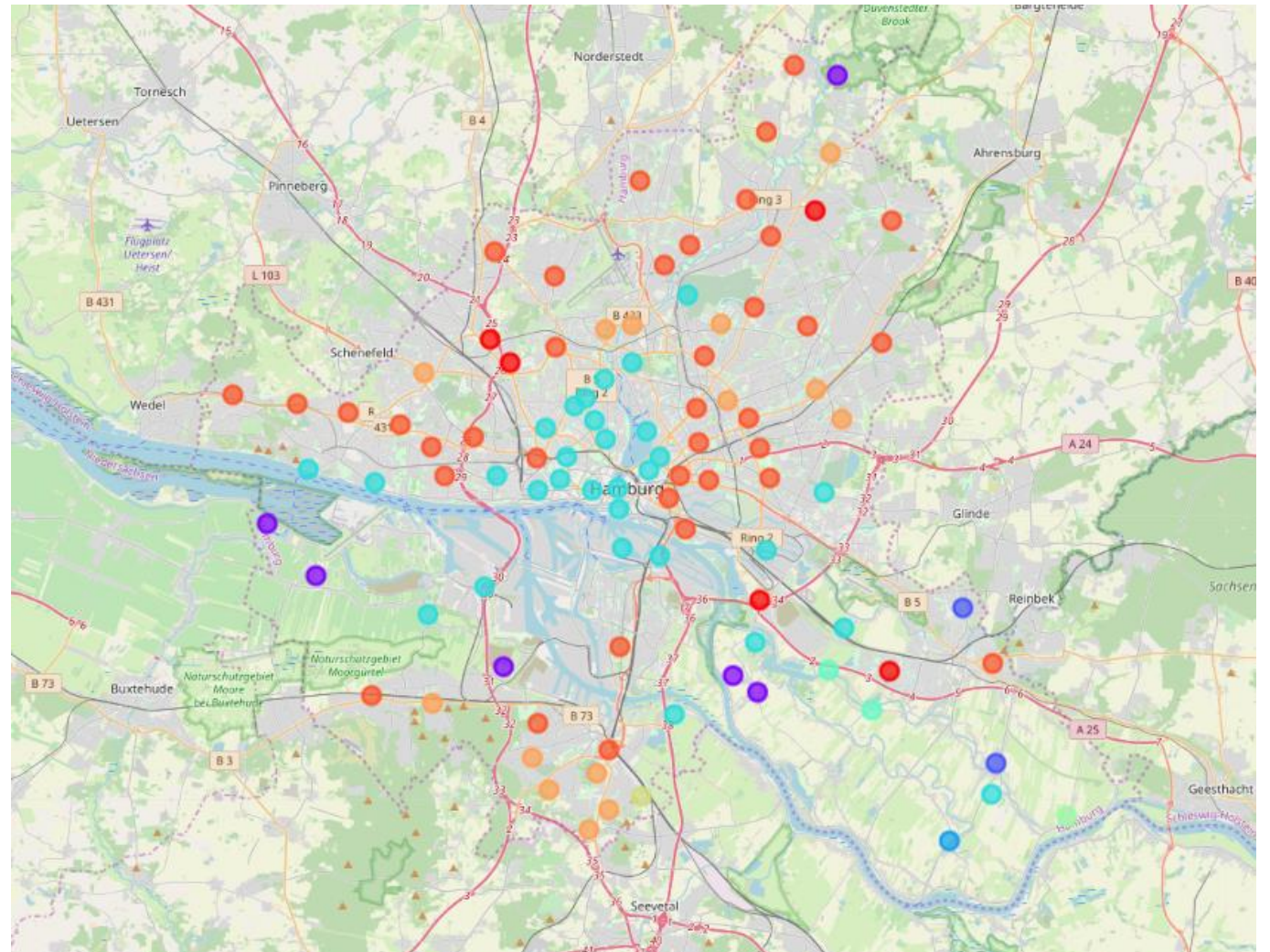
# 4. Analysis



# 5. Results

- Most interesting Cluster is number 5 (light blue).
- Contains mostly cafes and bars

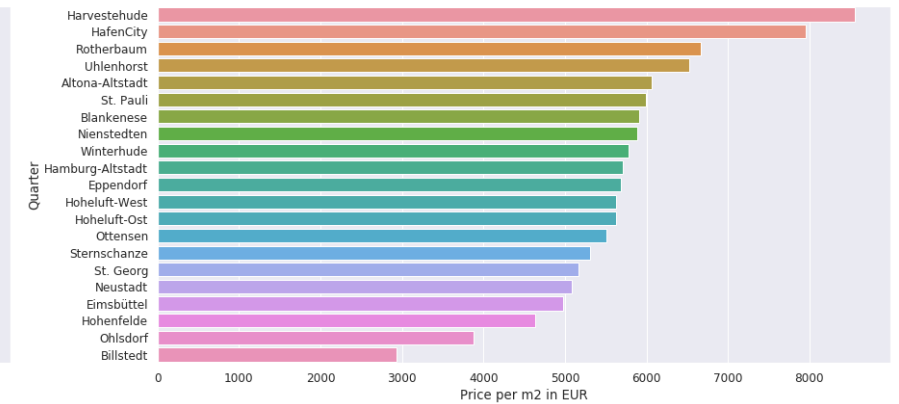
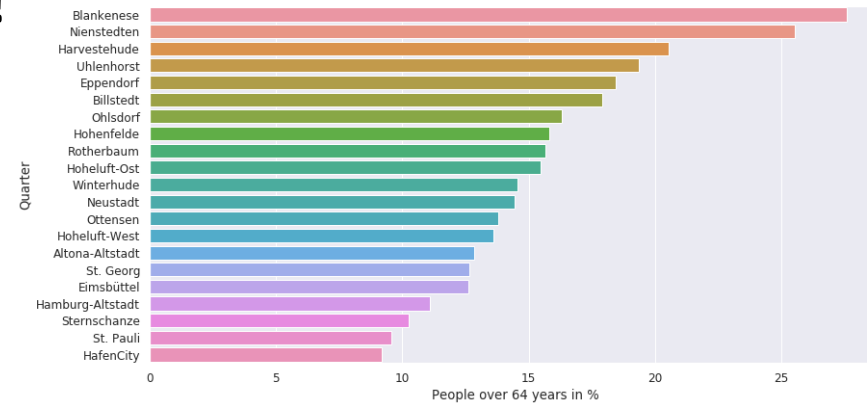
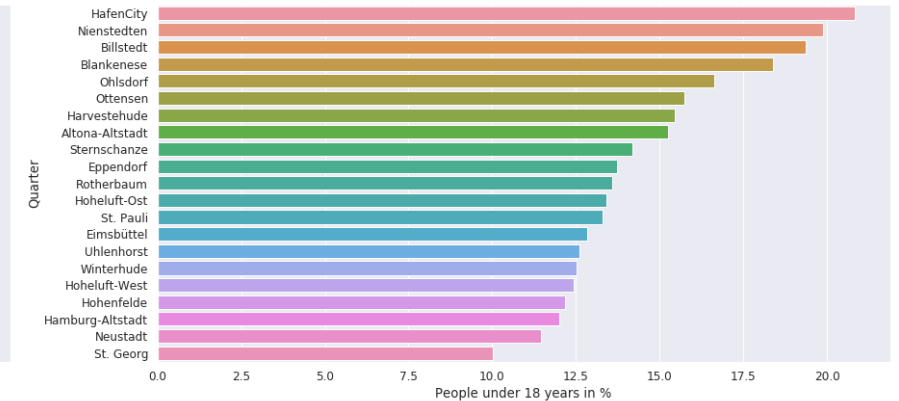
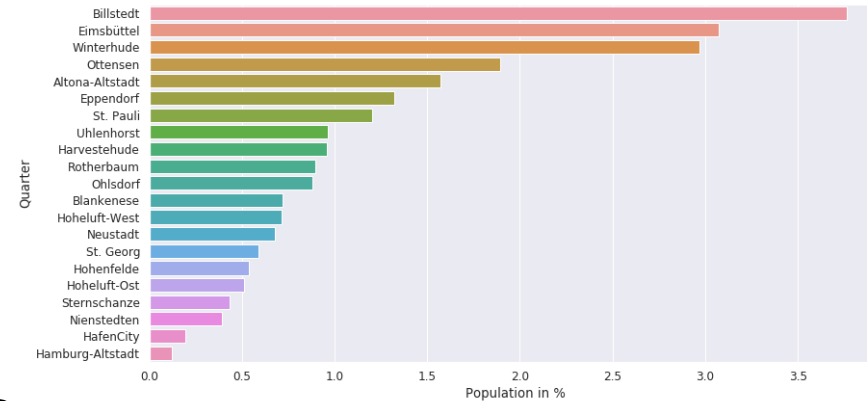
## Clustering with k-Means





# 5. Results – Further Investigation of Cluster 5

- Billstedt as a suitable candidate?



## 6. Discussion and recommendation

- As we can see, **Billstedt** is the quarter with the **highest population** in cluster 5. It contributes with nearly 3.8% to the total population of Hamburg. Billstedt is by far the **cheapest quarter** with an average price per m2 of ca. 3.000 EUR. This price is considered for buying a flat per m2 in EUR but could also reflect how expensive a quarter is in general.
- Furthermore, this quarter has with ca. 19% the **third highest rate of young people** (under 18) and has **fairly a lot people over the age of 64** (ca. 18%). So it could be interesting to consider this in the business strategie of the cafe.
- In general, this could be a **good quarter for setting up our cafe**.
- It has to be further researched where the optimal location for this cafe is in this quarter. Nevertheless we should not forget that other clustering algorithms than k-Means could deliver different outcomes. This has to be evaluated as well as other factors that could be vital for the success of our cafe.



## 7. Conclusion

- Finally to conclude this project, we got a small glimpse how a data science project could look like. We used the most used Python libraries, the Foursquare API, Geocoding and webscraped data from the internet. We did this to analyse the quarters of Hamburg and saw the results on a Folium map in order to choose a quarter for a cafe. This kind of analysis can be **suited as a first search for possible locations but cannot replace further research** with taking more factors into considerations.
- This project **demonstrated and enhanced the data mining and data science skill** as part of the IBM Data Science Capstone Project.
- You are invited to learn from it and copy it. If you copy, think about what happens in the code.