Coursera Capstone Project Hamburg Neighbourhoods

Hendrik Welsch, January 2021

Introduction

- **Problem**: finding suitable neighbourhood in Hamburg, Germany, for living or opening business

- **Aim**: group neighbourhoods using
  - rent data
  - popular venues

- **Stakeholder**:
  - house/apartment-hunters & realtors
  - entrepreneurs looking to open a business, e.g. nightlife location

Data Aqcuisition (1)

- Neighbourhoods of Hamburg
  - Neighbourhood names
  - Mean rent/m2 [€] (from 2017)

Source: https://mietspiegeltabelle.de/mietspiegel-hamburg/

Mietpreis Hamburg

Durchschnittliche Miete in Euro je Quadratmeter des Jahres 2017.

sehr niedrig | niedrig | durchschnittlich | sehr hoch

Stadtteil	Miete pro m^2
Allermöhe	8,70 Euro
Alsterdorf	10,72 Euro
Altengamme	8,00 Euro
Altenwerder	9,08 Euro
Altona-Altstadt	10,60 Euro

Data Aqcuisition (2)

- Venue data
  - Retrieved from Foursquare
  - Limit: Top 100 venues
  - Radius: 600 m
  - From this: selected venue type categories at three different levels of detail


```
'venue': {'id': '54200eb6498e5af295bdd77c',
          'name': 'cantinetta ristorante & bar',
          'location': {'address': 'Pickhuben 3',
           'lat': 53.54411350571698,
           'lng': 9.994533061981201,
           'labeledLatLngs': [{'label': 'display',
             'lat': 53.54411350571698,
             'lng': 9.994533061981201}],
           'distance': 158,
           'postalCode': '20457',
           'cc': 'DE',
           'city': 'Hamburg',
           'state': 'Hamburg',
           'country': 'Deutschland',
           'formattedAddress': ['Pickhuben 3', '20457 Hamburg', 'Deutschland']},
          'categories': [{'id': '4bf58dd8d48988d110941735',
           'name': 'Italian Restaurant',
           'pluralName': 'Italian Restaurants',
           'shortName': 'Italian',
           'icon': {'prefix': 'https://ss3.4sqi.net/img/categories_v2/food/italian_'
            'suffix': '.png'},
           'primary': True}],
          'photos': {'count': 0, 'groups': []}},
         'referralId': 'e-0-54200eb6498e5af295bdd77c-2'},
        {'reasons': {'count': 0,
          'items': [{'summary': 'This spot is popular',
            'type': 'general',
            'reasonName': 'globalInteractionReason'}]}
```


Venue category

Highest-level category

Higher-level category

Data Cleaning & Preparation (1)

- Neighbourhood & Rent Data
  - Translation to English
  - Rent conversion to float
  - Removing neighbourhood outside of city area
  - Retrieving coordinates (lat, long) for each neighbourhood using geopy

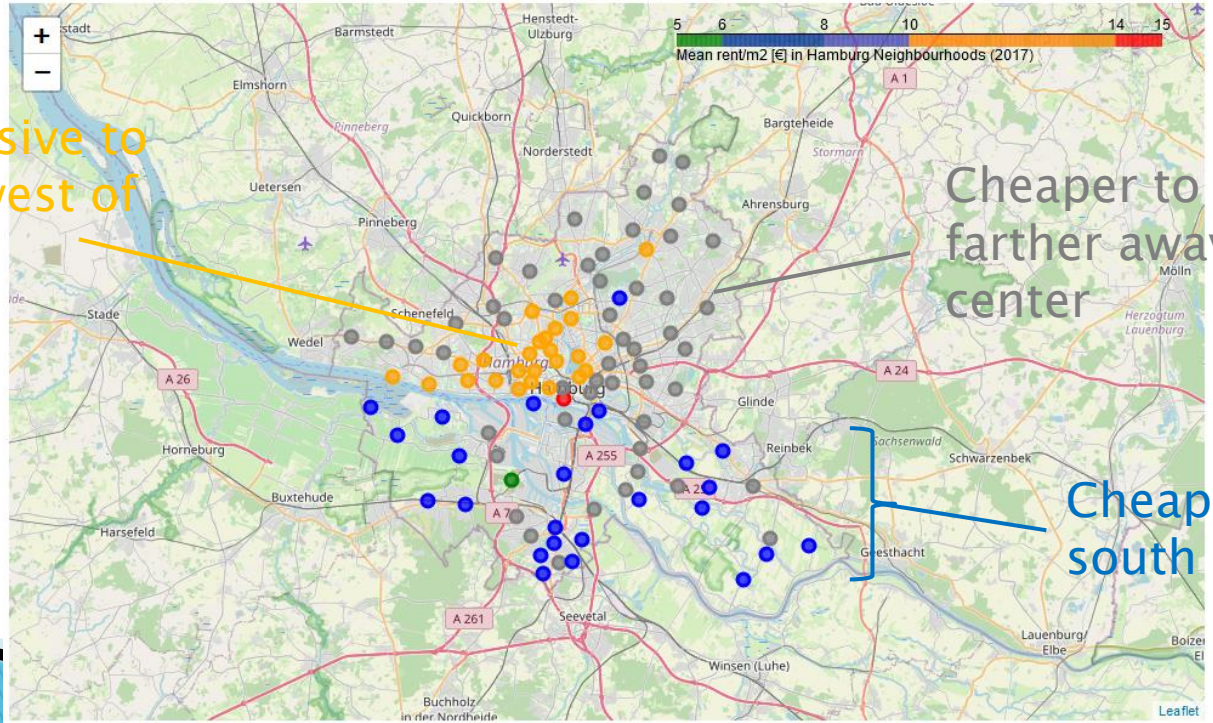
Data Cleaning & Preparation (2)

- Foursquare data
  - Removing transport-type venues (e.g. Bus Stop)
  - One-hot encoding: generating data frames with frequencies of each venue type category in each neighbourhood
  - Removing neighbourhoods with 0 popular venues

Exploratory Data Analysis (1)

- Visualization of mean rent

More expensive to live north-west of city center

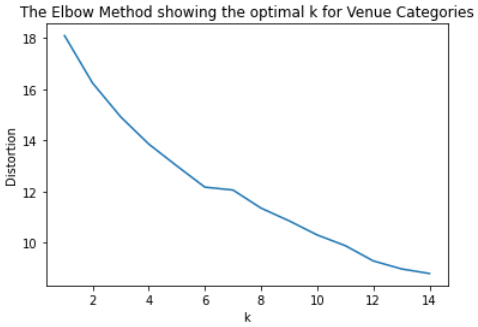


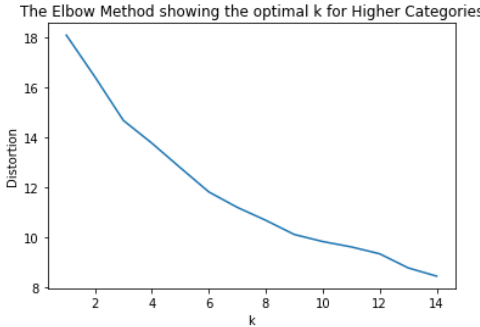
Cheaper to live farther away of city center

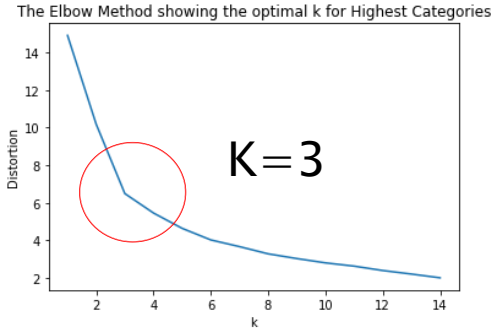
Cheaper to live south of River Elbe

Exploratory Data Analysis (2)

- Neighbourhood Clustering: k-means
  - Determining optimal k with elbow method





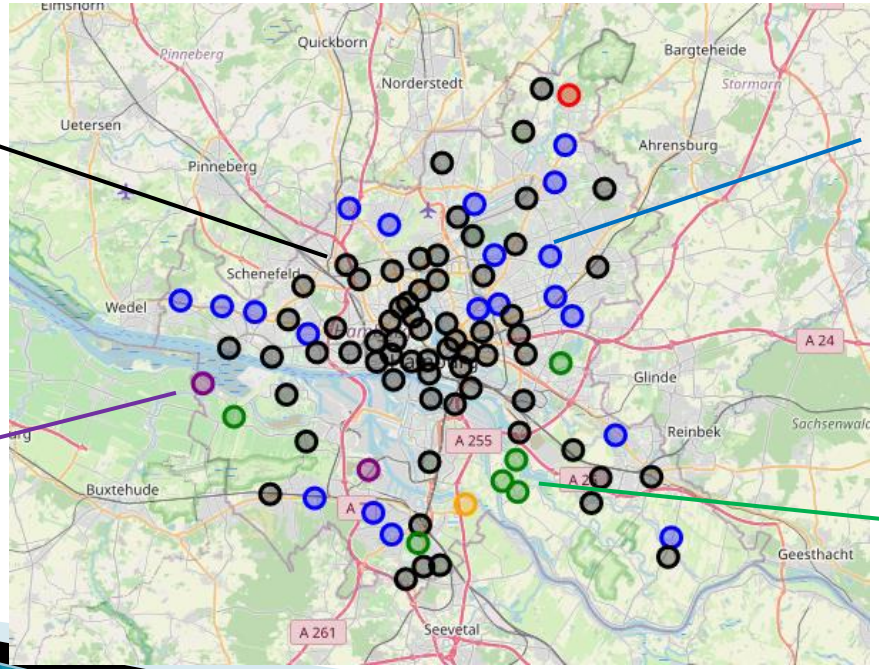


K=5 (Compromise of low distortion & low complexity)

Exploratory Data Analysis (3)

- Clustering based on Venue Category

„inner-city“-like
- restaurants, cafes
- hotels
- various shopping



„suburb“-like
- supermarkets
- bakeries
- groceries
- department stores

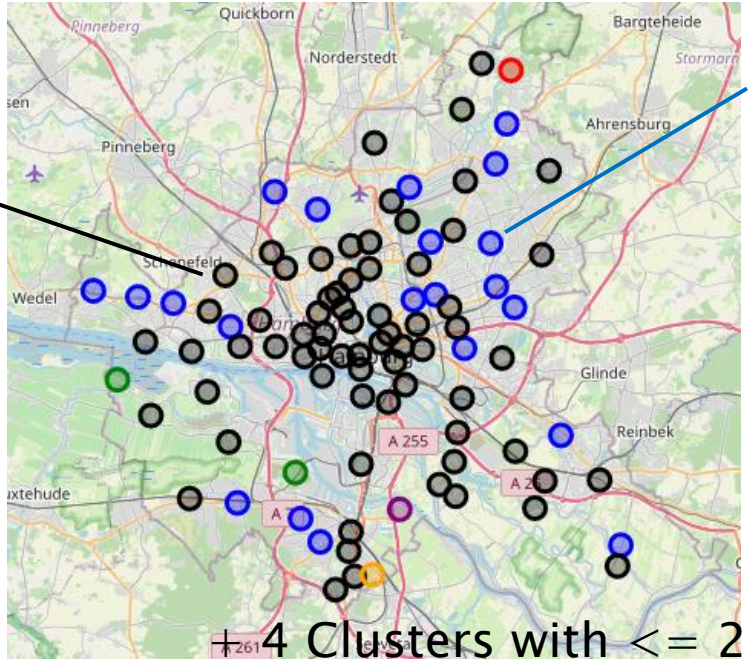
German restaurants
+ Zoo Exhibit
+French Restaurant

Defining:
German restaurants

Exploratory Data Analysis (4)

- Clustering based on Higher-Level Venue Category

„inner-city“-like
- cafes
- food
- bakeries
- hotels



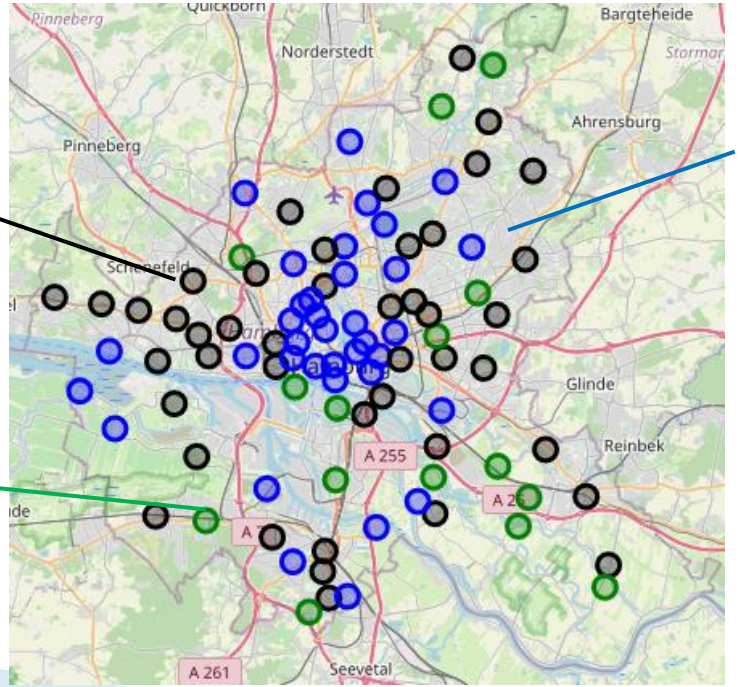
„suburb“-like
- Food/grocery stores

+4 Clusters with <= 2 neighbourhoods
→ no conclusion possible

Exploratory Data Analysis (5)

- Clustering based on **Highest**-Level Venue Category

„suburb“-like
- shops
- food

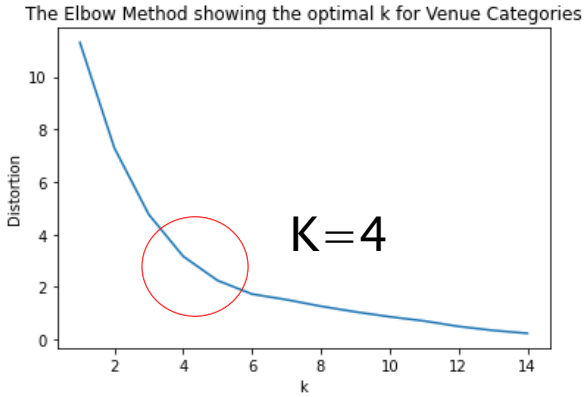


„inner-city“-like
- food (e.g. restaurants)
- shops
- **travel (e.g. hotels)**

Defining:
- parks
- outdoor activities

Exploratory Data Analysis (6)

- Clustering based on **Nightlife Venues**
- Data preparation
  - Filtering for highest category = „nightlife“
  - Removing problematic categories „Beer Store“ and „Other Nightlife“
- Determining optimal K
- Removing neighbourhoods with no nightlife locations

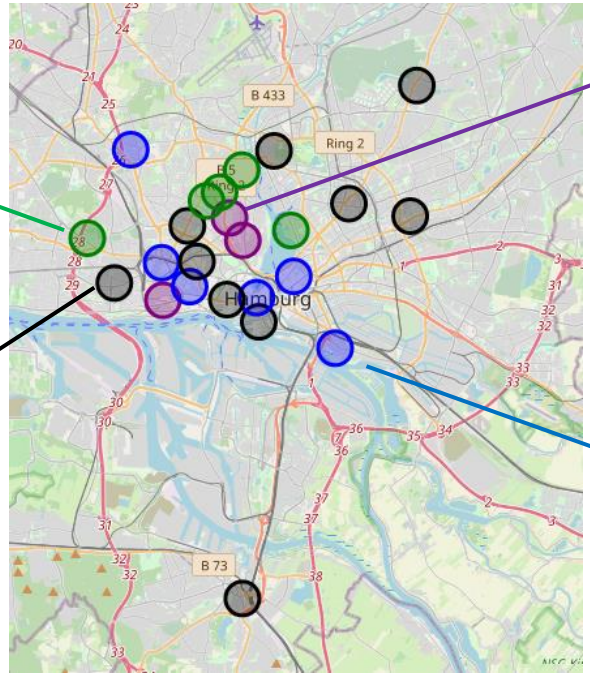


Exploratory Data Analysis (7)

- Clustering based on **Nightlife Venues**

Defining:
- cocktail bars

Mostly Bars
- normal
- whiskey
- hookah



Defining:
- Pubs

Defining:
- nightclubs

Results – Summary

- Higher rent (a) closer to city center, and (b) north of river Elbe
- Clustering approach able to separate neighbourhoods into
  - **Inner-city-like**: lively, many restaurants, cafes, bars, hotels
  - **Suburb-like**: dominated by supermarkets, other daily-life services
  - **Outdoor activity** supporting neighbourhoods
- Successful separation of neighbourhoods based on frequent nightlife venue types

Discussion

- Downsides of the approach:
  - Mean rent: no information whether neighbourhood has lots of industry or is living area
  - Limited meaningfulness of clusters, especially small ones
  - Problematic venue categories, especially „default“
  - Strong skewing potential of parameters LIMIT and RADIUS when retrieving Foursquare data

Discussion

- Possible future improvements:
  - Manually group venue categories into higher-level categories to achieve more meaningfull results
  - Include more neighbourhood data: mean age, mean income, distribution of education …
  - Include data about schools, medical facilities -> equally important for people looking for a place to live, but hard to define by „popularity“; other rating metrics needed

Conclusion

- Generated and visualized data set of Hamburg neighbourhoods
- Good starting point for someone
  - looking for a place to live
  - looking for a place to open a certain business
- Nightlife venues only one application, can be adjusted to other venue types