

# Analysis of Düsseldorf, Germany

## A machine learning approach to neighbourhood clustering

M. Gehring

26th May 2020

# Why Analyse Düsseldorf?

## Background and Problem Statement

Analysis of  
Düsseldorf,  
Germany

M. Gehring

Introduction

Data

Results and  
Discussion

Conclusion

- ▶ Düsseldorf is a business hot-spot area with a large number of job opportunities
- ▶ International city with many influences by a large Japanese ex-pat community and close neighbours Belgium and The Netherlands
- ▶ Connected to the world by a long-distance train station and an international airport

### Problem statement

Which areas of Düsseldorf are characterised by a high living quality with affordable rental apartments?

## Data Sources

- ▶ Geodata of city and location data on train stations is provided by OpenData Düsseldorf
- ▶ Geolocation of boroughs and apartments is pulled using bing maps API
- ▶ Geodata of tram/bus stations is provided by the regional service provider *Verkehrsverbund Rhein-Ruhr*

## Data Cleaning

This data required minimal attention. Geocoordinates provided for tram/bus stops was turned from MRCV format to decimal coordinates.

## Data Sources

- ▶ Apartment data is scraped from rental agent website  
*Immobilien Scout24*

## Data Cleaning

- ▶ 1460 apartments were retrieved on 25 May 2020.
- ▶ Address data was separated to street address, borough and city information.
- ▶ 146 apartments with incomplete address data were removed from the data set.
- ▶ Price, size, and number of rooms information was turned from string format with German notation to float data.

# Data Retrieval and Cleaning

Borough venue data for clustering

Analysis of  
Düsseldorf,  
Germany

M. Gehring

Introduction

Data

Results and  
Discussion

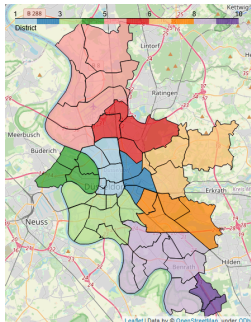
Conclusion

## Data Sources

- ▶ Venue data was pulled for the centre location of each borough using the Foursquare API

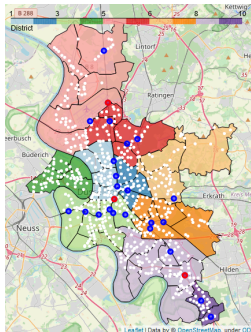
## Data Cleaning

- ▶ This data was used without modification.
- ▶ 808 venues were identified within 154 unique categories.



### Administrative structure

- ▶ 10 districts
- ▶ 54 boroughs
- ▶ 9 out of 10 districts are east of the Rhine river



### Public transportation infrastructure

- ▶ 3 long distance train stations
- ▶ 25 city train stations
- ▶ 1405 bus/tram stops
- ▶ trains follow mainly north-south or west-east axis with Düsseldorf main station at the centre

# Results

## Location of apartments in Düsseldorf

Analysis of  
Düsseldorf,  
Germany

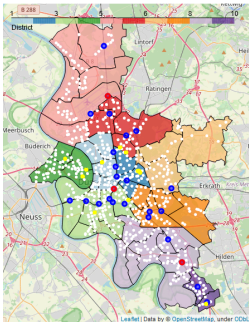
M. Gehring

Introduction

Data

Results and  
Discussion

Conclusion



### Apartment location

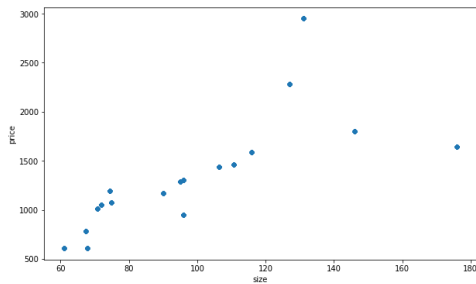
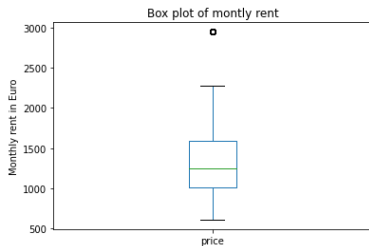
- ▶ sample of 200 apartments
- ▶ densely located in 8 boroughs/6 districts
- ▶ mostly in districts 2,3,4, and 6



# Results

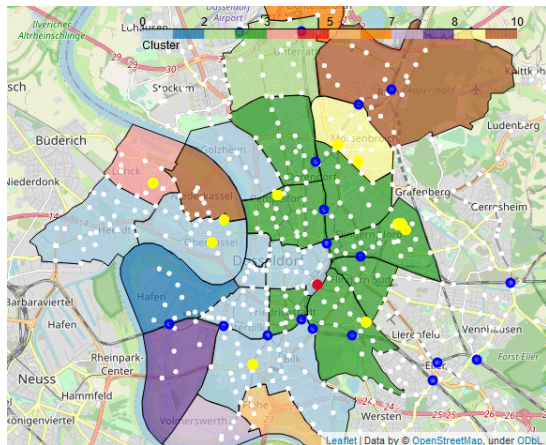
## Apartment statistics

|       | price       | size        | rooms       |
|-------|-------------|-------------|-------------|
| count | 1314.000000 | 1314.000000 | 1314.000000 |
| mean  | 1345.598333 | 98.816111   | 2.722222    |
| std   | 563.489658  | 30.426843   | 0.711729    |
| min   | 610.000000  | 61.000000   | 2.000000    |
| 25%   | 1011.710000 | 72.000000   | 2.000000    |
| 50%   | 1243.230000 | 95.480000   | 2.750000    |
| 75%   | 1588.170000 | 115.800000  | 3.000000    |
| max   | 2950.000000 | 176.000000  | 4.000000    |



- ▶ Linear relation of price and size in two sections with few outliers
- ▶ Steep increase between 60 m<sup>2</sup> and 80 m<sup>2</sup>
- ▶ Less steep increase above 80 m<sup>2</sup>

## Clustered boroughs of Düsseldorf



## Clusters

- ▶ Two large clusters (when number of clusters is 11)
- ▶ All clusters display similar characteristics

- ▶ Apartment prices are related more to their distance from the city centre than their respective clusters
- ▶ Cheapest apartments are found in district 6 and 10.
- ▶ Apartments in district 4 were more expensive than the others

# Conclusion

## Lessons learned

- ▶ Clustering the city of Düsseldorf based on its boroughs is difficult
- ▶ 808 venues in 25 boroughs is not very much
  - ▶ either Foursquare data is insufficient or
  - ▶ radius around borough centre was chosen too small (500 m)
- ▶ 2-3 feasible apartments could actually be identified

- ▶ To improve results apartments could also be clustered to identify common traits more easily
- ▶ the analysed radius could be increased
- ▶ other data sources such as median income, crime rate, or others could be added to the analysis to improve results
- ▶ another machine learning method could be employed to allow for asymmetric clusters