The Battle of Students

A data analysis project for apartment-seeking students in Graz, Austria.

Bernhard Neumayer

October 11, 2019

Introduction

Background

Graz

- 300,000 residents
- 60,000 students
- Strong competition in apartment-hunting

Renting vs. Buying

- Many companies in and around Graz
- Students often move permanently after their studies
- Low interest rates

Problem

Renting

- Current situation in a borough of interest
- Cheap rent, close proximity to university

Buying

- Where ist it currently affordable?
- Will value increase/stay/decrease?

General

- Information source?
- Relying on assumptions and recommendations

Target Audience

Target Audience

- Students planning to study in Graz
- Looking for an apartment to rent
- Looking for property to buy

Data

Data Sources

Graz

- https://de.wikipedia.org/wiki/Liste_der_Stadtbezirke_von_Graz

Rent and Realty Prices

- https://www.immobilienscout24.at/
- https://www.immowert123.at/

Foursquare

- Vibrancy of borough based on availability of venues

Location Data

- geopy: Nominatim, ArcGIS

Data Acquisition

Pandas

- For data in tabular form
- Wikipedia
- `https://www.immobilienscout24.at/`

Beautiful Soup

- For less nicely provided data
- `https://www.immowert123.at/`

Data Cleaning, Data Preparation, Feature Selection

Graz

- Borough names
- Area
- Population Density

https://www.immobilienscout24.at/

- Use entire data

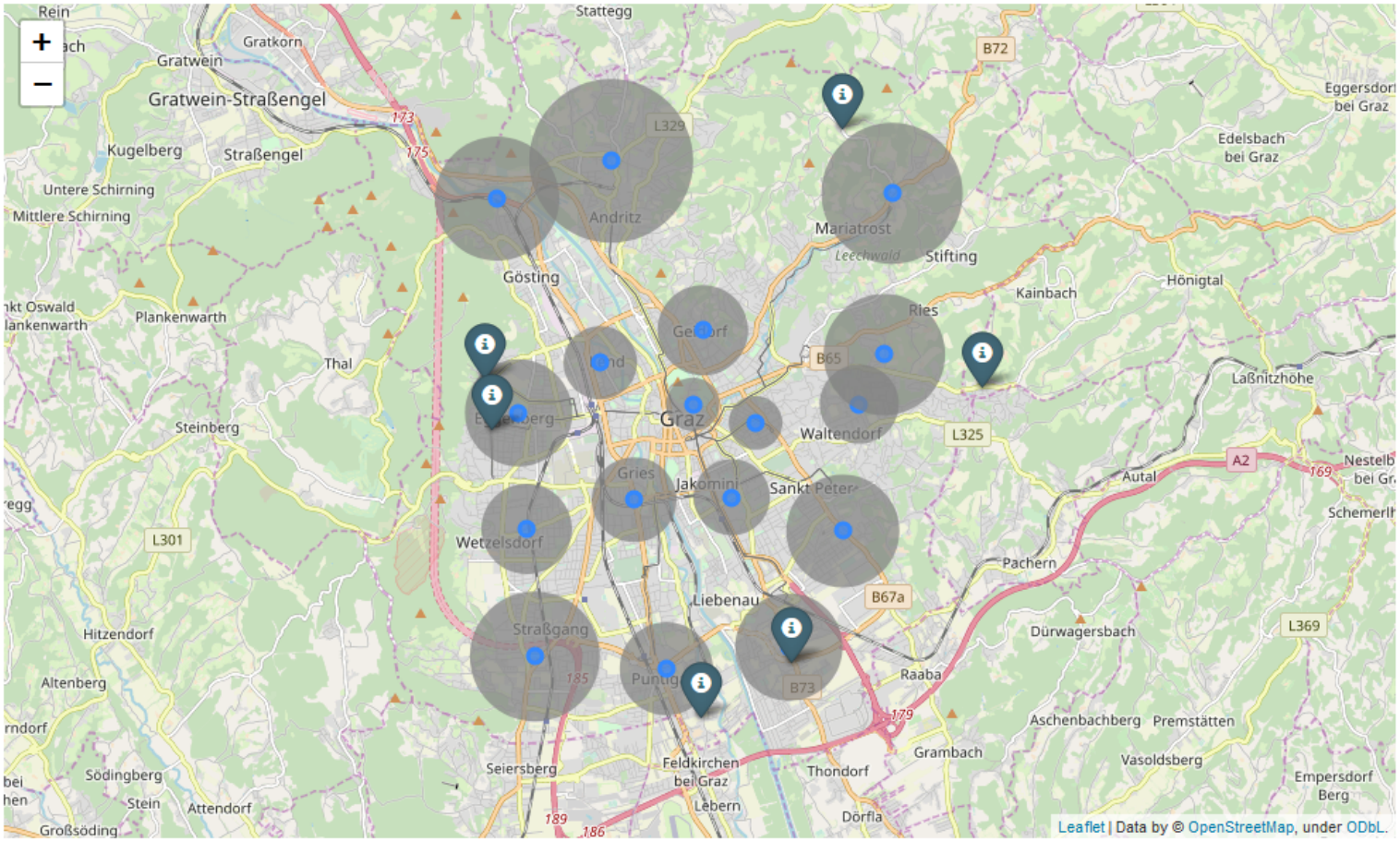
https://www.immowert123.at/

- Filter for entries with “wohnung” (apartment) in it

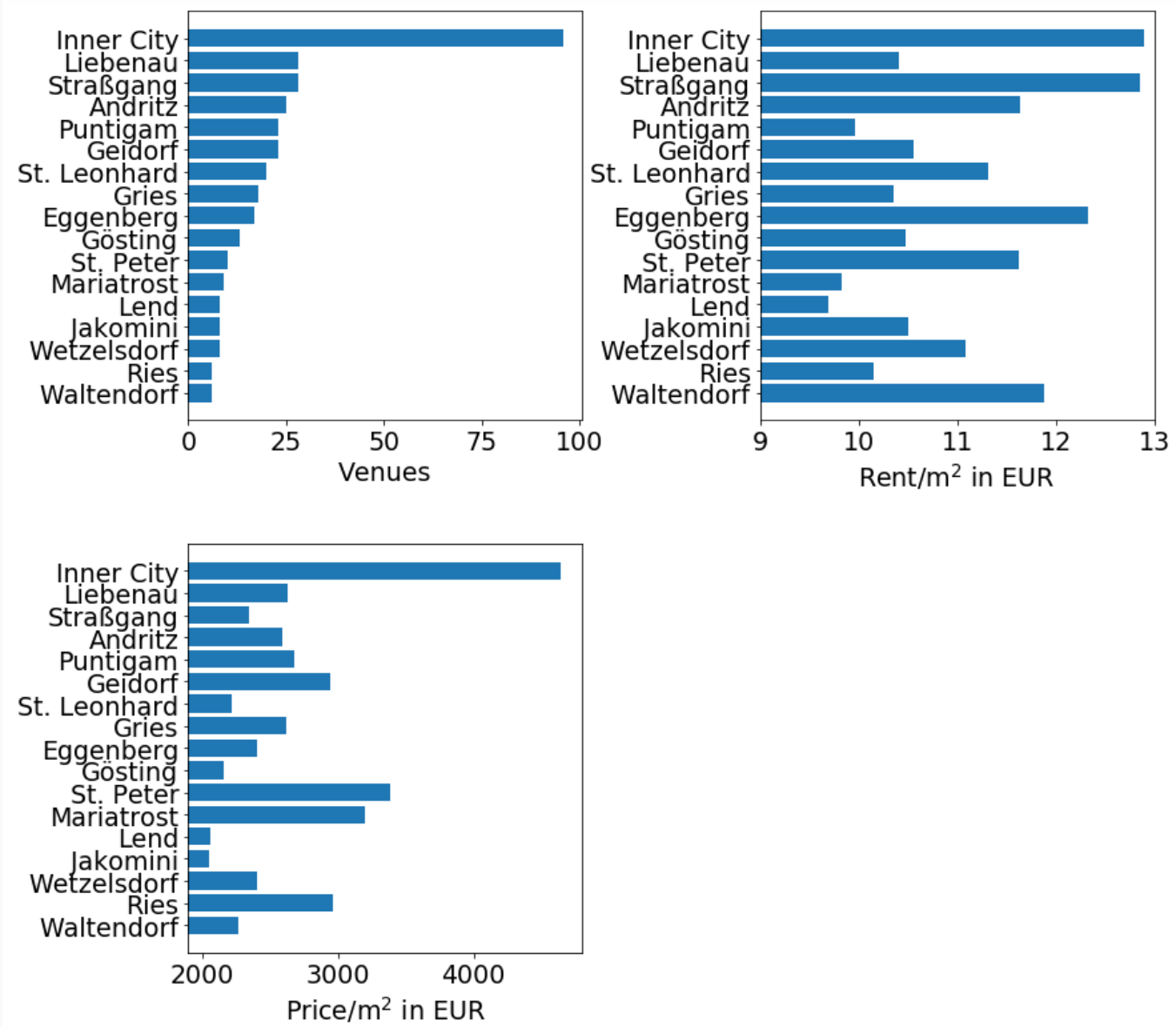
Methodology

Starting Point

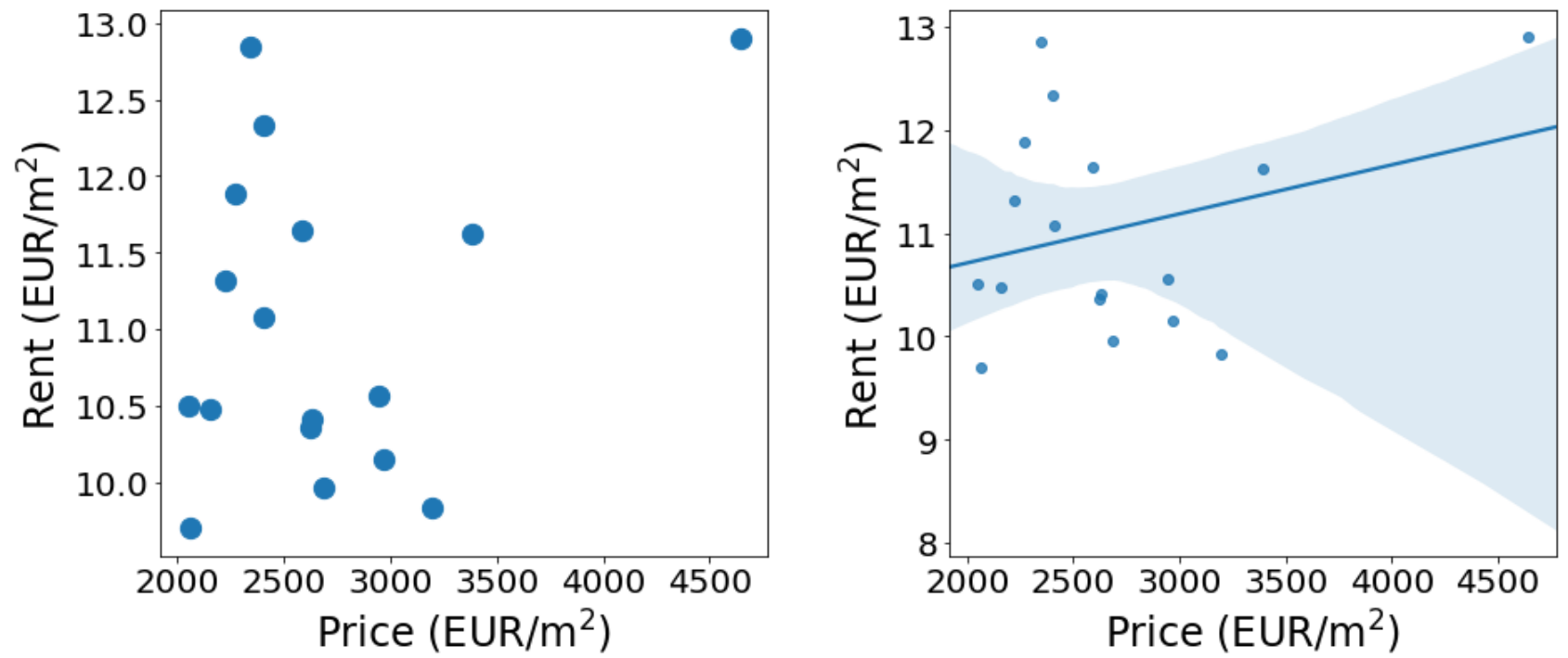
Boroughs, Districts, Area Size



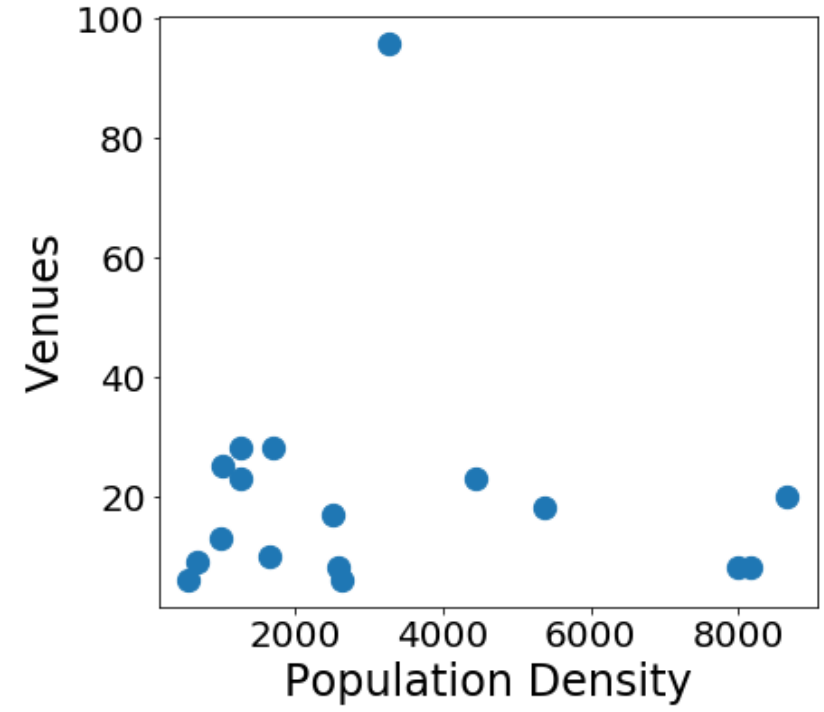
Exploratory Data Analysis: Vibrancy, Rent, Realty Price

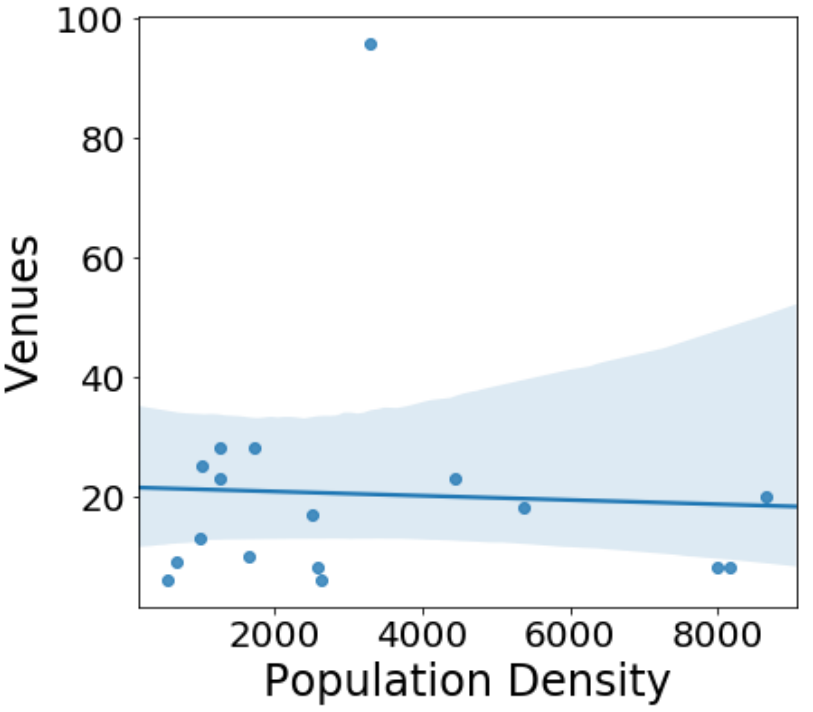


Exploratory Data Analysis: Realty Price vs Rent



Exploratory Data Analysis: Population Density vs Vibrancy





Conclusions from Data Analysis

- Variables are independent
- Small data set but maximum information
- Rent and prices may lead to different clusters

Clustering

Important Aspects

- Larger boroughs leave room for new buildings.
- Areas with high population density can be too crowded.
- Very few venues suggest only little entertainment.

Assumptions for Cluster Interpretation

- Buying: affordable price, moderate vibrancy
- Renting: affordable rent, high vibrancy
- Densely populated areas not preferred.
- Very little vibrancy is a drawback.
- Expensive areas not recommended

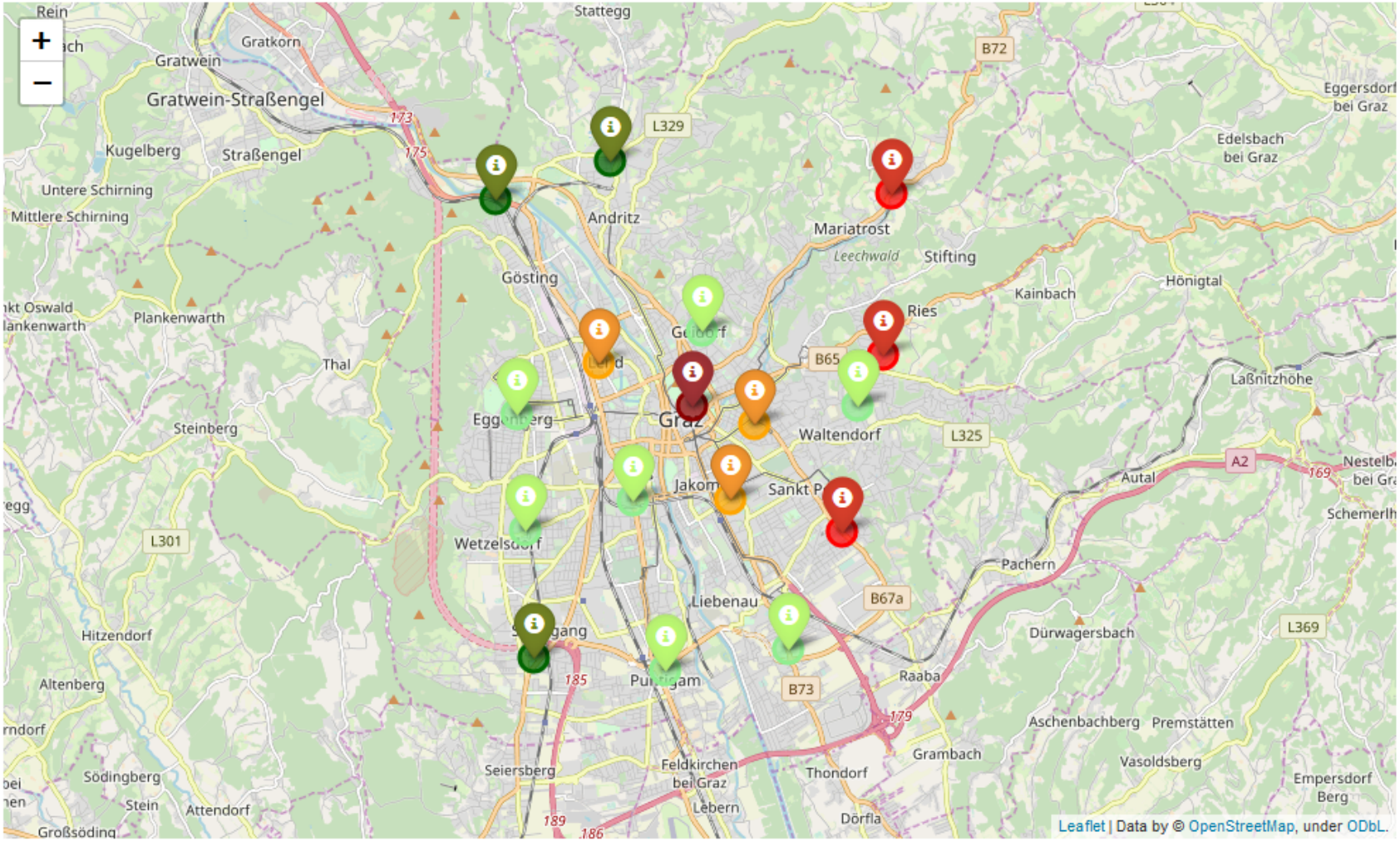
Results

Average Cluster Values

Cluster	Rent EUR/m^2	Average Price EUR/m^2	Area km^2	Population Density $1/km^2$	Venue Count
0	10.94	2566.15	6.11	2941.29	17.57
1	12.90	4642.89	1.16	3287.00	96.00
2	11.66	2363.28	13.68	1097.00	22.00
3	10.53	3184.44	11.00	968.67	8.33
4	10.51	2110.92	3.20	8263.67	12.00

- Largest cluster: cluster 0 (7 boroughs)
- Smallest cluster: cluster 1 (1 borough, inner city)
- Remaining clusters: 3 boroughs each
- Same clusters for rent and price

Map with Recommendations



Discussion

Discussion

Results

- Map provides overview and enables further filtering
- Largest cluster: renting
- Dense population: close to transport hubs
- Central borough most expensive
- Reasonable results

Method

- Using 4 clusters would combine clusters 2 and 4
- 5 clusters separate periphery and developing boroughs

Conclusions and Outlook

Conclusions and Outlook

Conclusions

- Objective method
- Approach is feasible, provides reasonable results
- Generally applicable to any city

Outlook

- More data will lead to better results
- Smaller disctricts to detect hotspots within boroughs
- More available venue data would allow target group-specific analysis