District Segmentation and Clustering

Cities and municipalities segmentation using k-means clustering of a district in southern Germany

Maik Sattelmaier

August 14, 2020

1 Introduciton

The *Rems-Murr-Kreis* is a district in the federal state of Baden-Württermberg in southern Germany. It is located in the northeast of the Stuttgart metropolitan region, which is known for its engineering and automotive industry which is home to the headquarters of Mercedes and Porsche and other world market leaders. The spatial proximity to this economic hub makes the district an attractive location for workers in the region. Here workers can take advantage of both the cultural offerings of a large city (Stuttgart) and the peace and nature of a rural region.

As a result of the popularity of the location and the welfare of the workforce, which benefits from a strong economic sector, rents, land and house prices are relatively high compared to the rest of Germany. However, prices vary greatly between the towns and communities within the district, although they are only a few minutes' drive apart. The aim of this project is to find out which places in the district are overpriced and underpriced in order to support the decision for future investors, shop owners and residents.

2 Business Problem

This section specifies the business problem and names the potential target groups.

2.1 The *Rems-Murr-Kreis*

Figure 1 shows the geographic location of municipalities (blue markers) and cities (red markers) wihtin the disctrict *Rems-Murr-Kreis*. This classification was made by the local administraion and can be freely accessible. The cities are located mainly close the Stuttgart and along two major roads that run through the area. The municipalities are typically located away from these two main roads, but often only a few minutes' drive from the cities.

Obviously this results in differences for commuters, but as mentioned above,

these are very small and do not justify the general price difference between the cities and the more rural communities. The objective of this project is to analyze which municipalities are underpriced and are similar to the cities wihtin the district in terms of popular nearby places and venues. A municipality which is officially classified as a municipality but is in the same cluster as a city is potentially underpriced and can be considered for settlement.

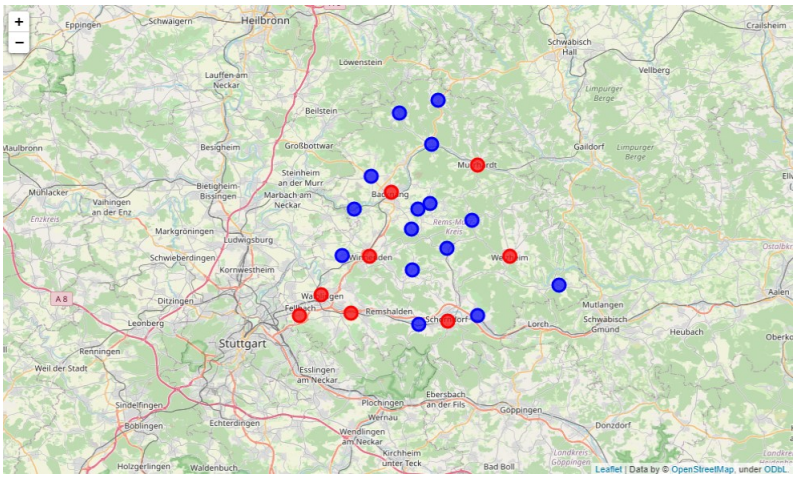

Figure 1: The Rems-Murr-Kreis: Blue markings indicate municipalites. Red markings indicate cities. Stuttgart is located in the southwest.

2.2 Potential Target Group

This project is addressed to a potentially broad audience taking advantage of underpriced settlements. Specifically, this includes all agents who intend to move near the economically strong city of Stuttgart: future workers, companies, restaurants, profit-oriented investors, shops and shopping centres.

3 Data and Methodology

This section describes the data used as well as the methodology for achieving results.

3.1 Data Set

I webscraped a list of cities and municipalites in the *Rems-Murr-Kreis* and merged it with their respective geographic coordinates, using Python's request and geocoder packages. I searched for the most common venues within 2000 meters of these coordinates using the 'Places API' of the Foursquare developer database, which completes my data set.

3.2 Methodology

Each location has been assigned its most popular venues which are proxies for their local amenities. I use k-means clustering to identify which locations are similar according to the most popular venues within the radius. As we only deal with binary qualitative responses (cities and municipalities) k is set to two. Finally I compare the labels of k-means clustering with the labels of the administration to identify which municipalities are wrongly classified as municipalities and are very similar to a city. In these municipalities the probability of underpriced housing is higher than in the correctly classified municipalities.