Clustering Analysis Mannheim

1. Motivation for Clustering Analysis

- Clustering of the districts of Mannheim (Germany)
- Clustering with respect to demographic variables as well as venue data
- Main goal is to provide an overview of the similarity and the dissimilarity of the different districts

Target Audience and impact

New businesses

New openings

Facilitate the opening of new venues with respect to reduced uncertainty when it comes to acceptance of the new venue

Expanding businesses

Facilitate expansion based on data insights through clustering

Municipality

Identify opportunities

The municipality can identify business opportunities more effective and can set incentives for businesses to open in a particular district

Increase inhabitant satisfaction

Municipality can effectively incentivise new businesses in districts where none of such venues exists.

2. Data Overview

Demographic Variables (per district)

- **Total Population**
- **Population growth (CAGR)**
- **Age structure (six different age groups)**
- **Data from the last federal election to see party preferences**

Venue Data (per district)

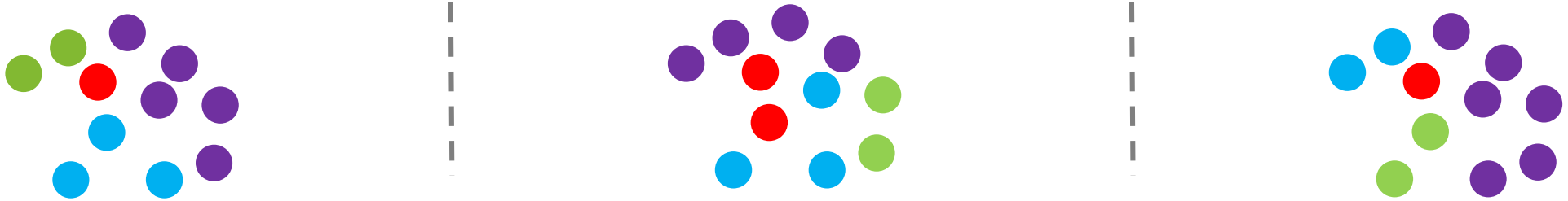
- **Multiple venue categories retrieved via the Foursquare API**

- Demographic data from the municipality's open data portal
- Venue data vie the Foursquare API
- Jointly the datasets can be used to get a holistic picture of a district
- Resulting differences in clustering emphasizes potential for new businesses

3. Methodology

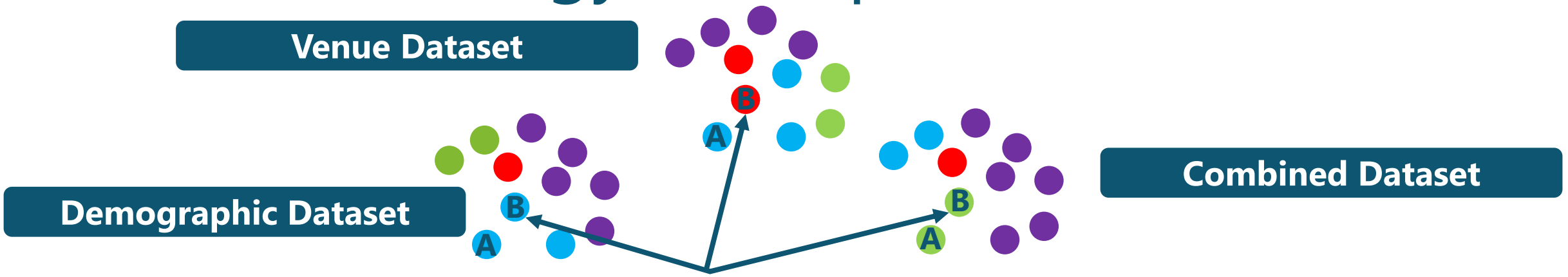
Demographic Dataset | **Venue Dataset** | **Combined Dataset**

K-means clustering



- Clustering is performed with 3 different datasets
- The result is then evaluated to see if there are crucial differences in clustering
- If there are crucial differences this indicates that demographic data and venue data do not produce the same clustering → in some districts that are similar (from demographic point of view) some venues are not established

3. Methodology (Example)

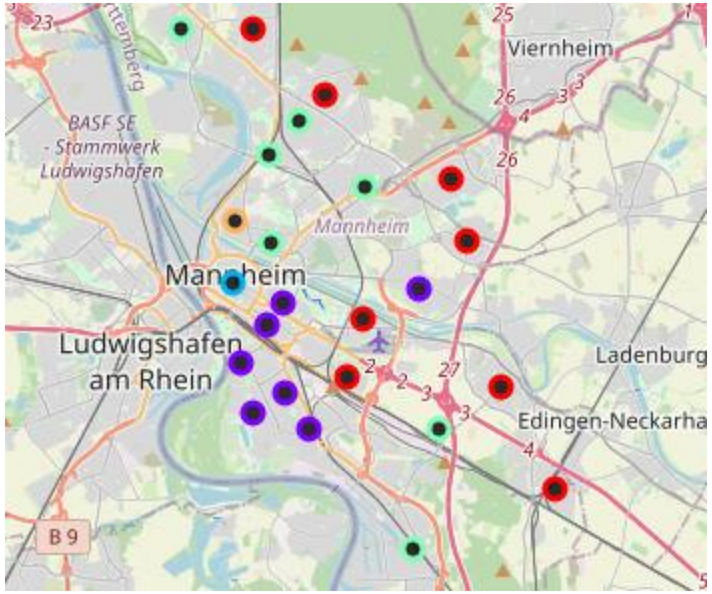


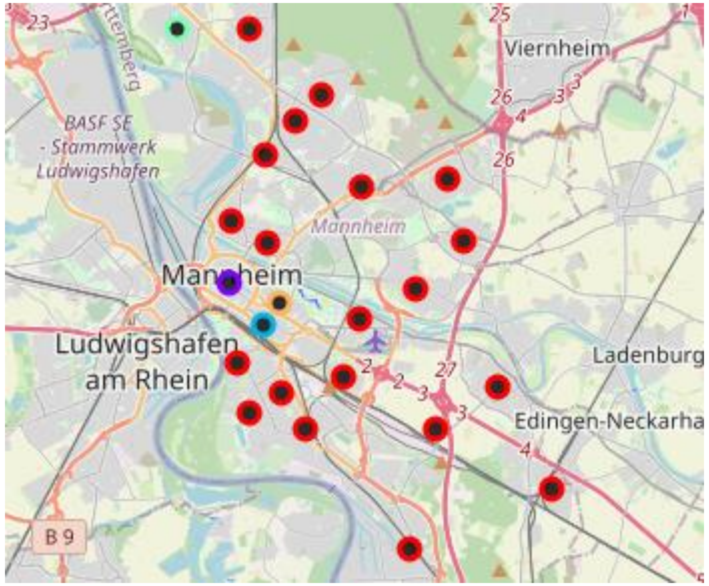
- District B is clustered differently in all clusters
- Clearly B is similar to A with respect to demographics but dissimilar with respect to venues in this district
- So for a business it would be a good idea to check what is a frequent venue in A and bring it to B
- The same demographics proxy for the acceptance of the new venue in the district

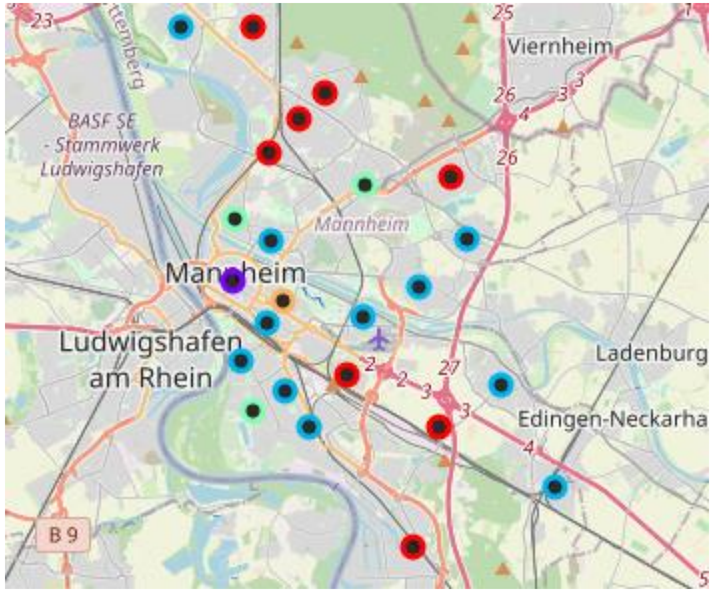
4. Results

Demographic Dataset | **Venue Dataset** | **Combined Dataset**

K-means clustering

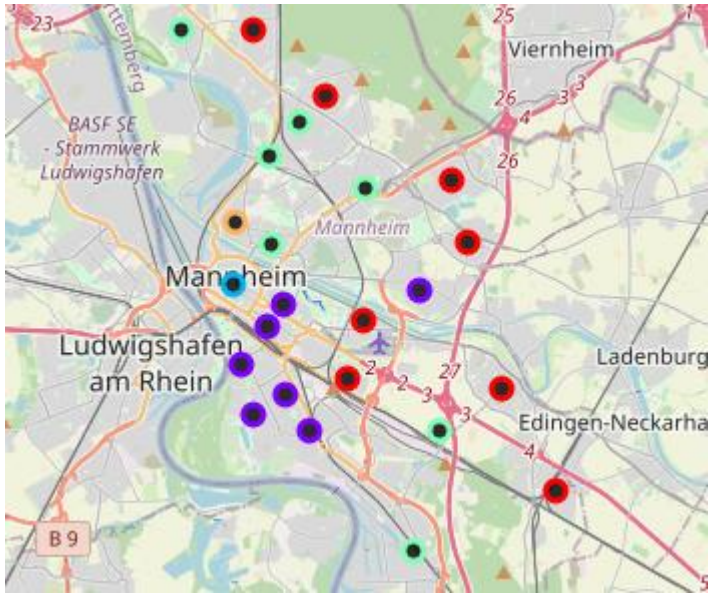




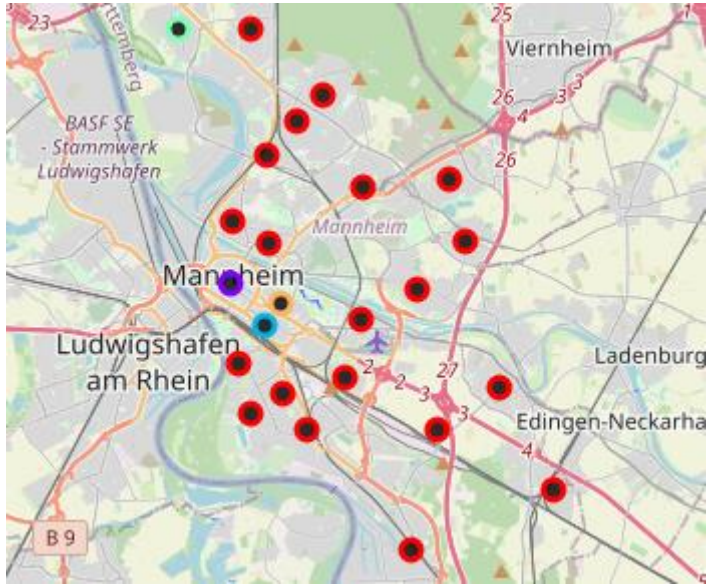


4. Results

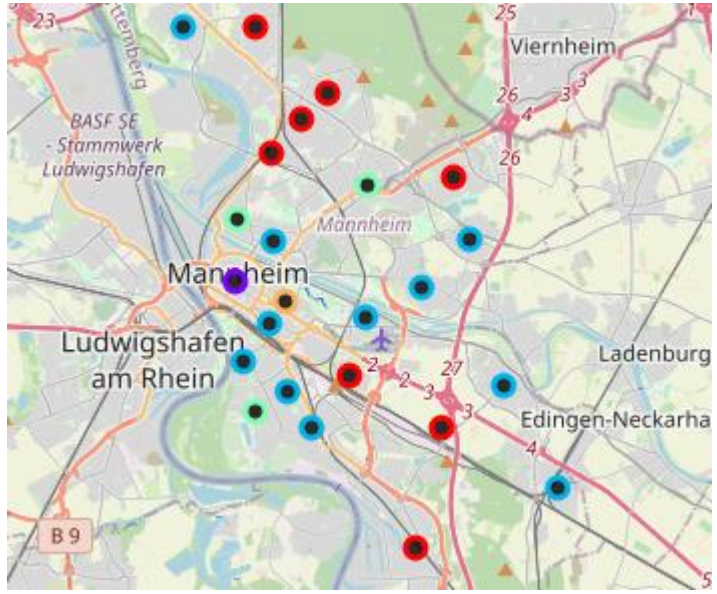
Demographic Dataset



Venue Dataset



Combined Dataset



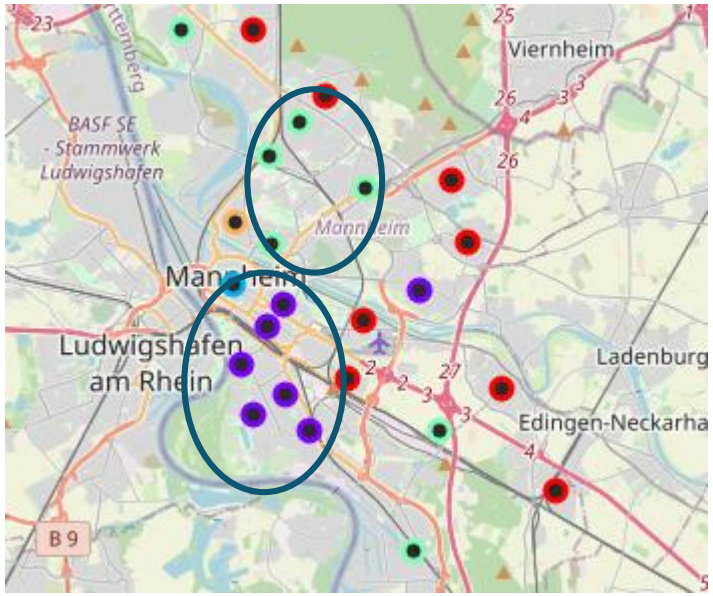
- As suspected the clustering with the different samples are quite different

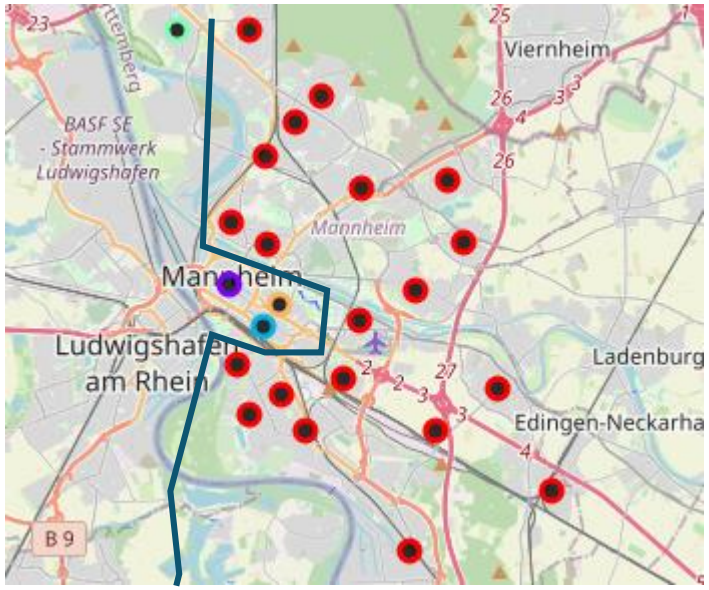
5. Discussion

Demographic Dataset

Venue Dataset

Clustering in geographic areas

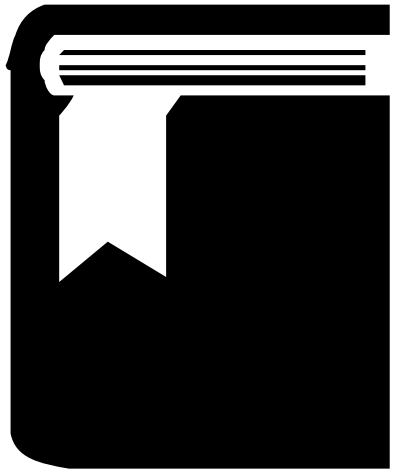




Clustering in distance from city center

- One can clearly see that some districts are differently clustered – here is some potential for new businesses

5. Conclusion



Demographic and location data result in different clustering of the districts

Demographic clustering is more geographic while venue clustering can be described as distance from the city center

Finally, there is potential to open up new businesses in Mannheim