Correlation of venue composition and political affiliations of districts in Vienna

Zi-Jing Cai

May 06, 2020

Table of Contents

1. Introduction

1.1. Background

When we think about cities like Paris we have certain images in our heads like fashion boutiques, cafes on every corner, wineries and many museums. However, every city has unique areas that are very different from other parts of that same city. Because of that, some districts or neighbourhoods are more popular than others and thus generate more money because of a higher consumption of services and goods. That is why it would be useful to find some similarities of these more successful districts so it might be applied to the less popular districts to make them more prosperous.

1.2. Problem

For this project, we will look at the districts of Vienna. We know that some districts generate more revenue from the service industry because people from other districts or even tourists prefer these districts over others.

The main questions for this project are whether these successful districts have something in common and which factors are influencing their positive development. For the latter part, we will look at the leading political party of every district to determine if they have an influence regarding the development of their respective districts. The reason is that every district partially managed by a political party and they have some power over the development of their districts. Of course they are boasting of their great management and that is why we want to see if there are any correlation between the leading party of a district and their composition of their venues.

1.3. Interest

The results of this analysis might be of interest to political parties if we do find any positive correlations between the success of a district and its affiliation to a certain political party. Real estate developers might also be interested to find out about emerging districts that are still cheap compared to the most popular ones. Various types of businesses such as restaurant operators might also be interested for the same reason as real estate operators.

2. Data

2.1. Data Sources

The main data source will be Foursquare's database and the coordinates of Vienna's districts. Since I could not find any online sources with the coordinates of all 23 districts of Vienna, I have decided to get the data myself. For that I have created a dataframe containing all district names of Vienna and 2 extra columns for the latitude and longitude. The district names were filled manually and for the other data I have used the geocode function to fetch the longitude and latitude information. After retrieving this data it was possible to call the "venues/explore" API-Endpoint of Foursquare for every district and to get information about the surrounding venues. The only relevant information from Foursquare for this project is the "Venue Category". With this information it is then possible to generate a list of venues for every district and therefore a venue profile based on the venue composition. With this venue profile it is then possible to cluster similar ones together and link them to their districts.

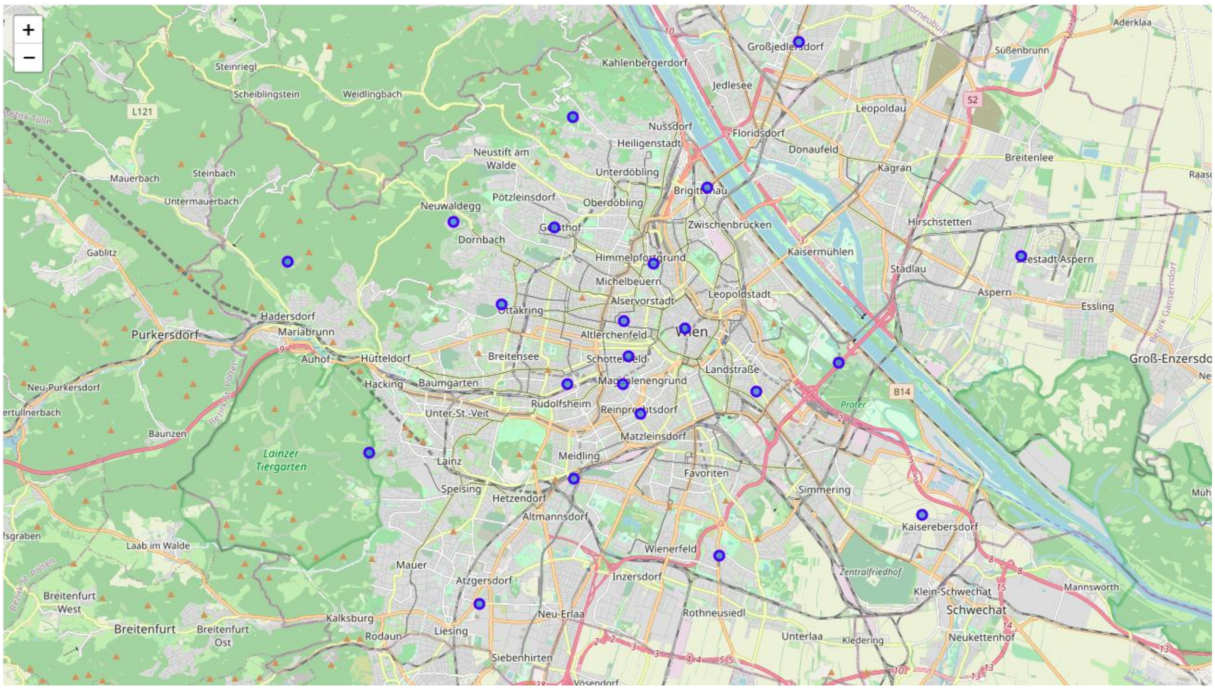
2.2. Data cleaning

There is not much work to do regarding cleaning because the location data for the districts are created by me with the help of geocode's function. I have verified the correctness of geocode's results, though the returned locations were sometimes not exactly in the middle of the district but still acceptable as we will see in later chapters. This means we are creating a dataframe that is already in the form we want it to be. As for the venue data from Foursquare we just select the venue type information for further processing.

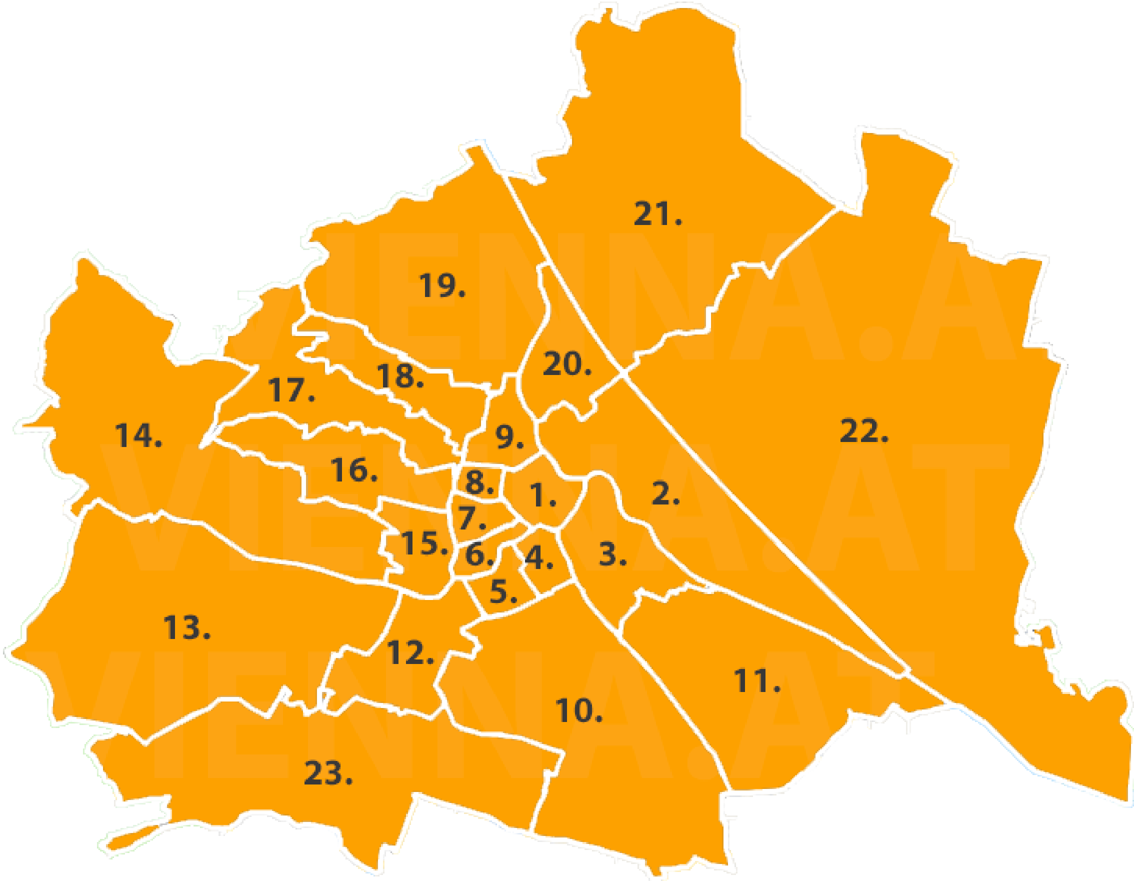
3. Methodology

3.1. Data processing

As described in the previous sections, the location data of the districts will be built in the way we need it so we can insert them into our calls to Foursquare's API. The following 2 compares the location data from geocode and the real districts of Vienna. The data seems accurate.



Map 1. Location generated based on geocode's data.



Map 2. The real districts of Vienna.

We will take the location data for each district and use the https://api.foursquare.com/v2/venues/explore endpoint to get the venues and assign them to the districts. I had to adapt the radius of the search based on the district size because the districts in the center are smaller than the outer districts. Initially the radius was set to 500 (meters) but that seems insufficient for the bigger districts as they have returned less than 10 venues, that is why for 9 districts I have increased the radius to 2km and it returned acceptable results.

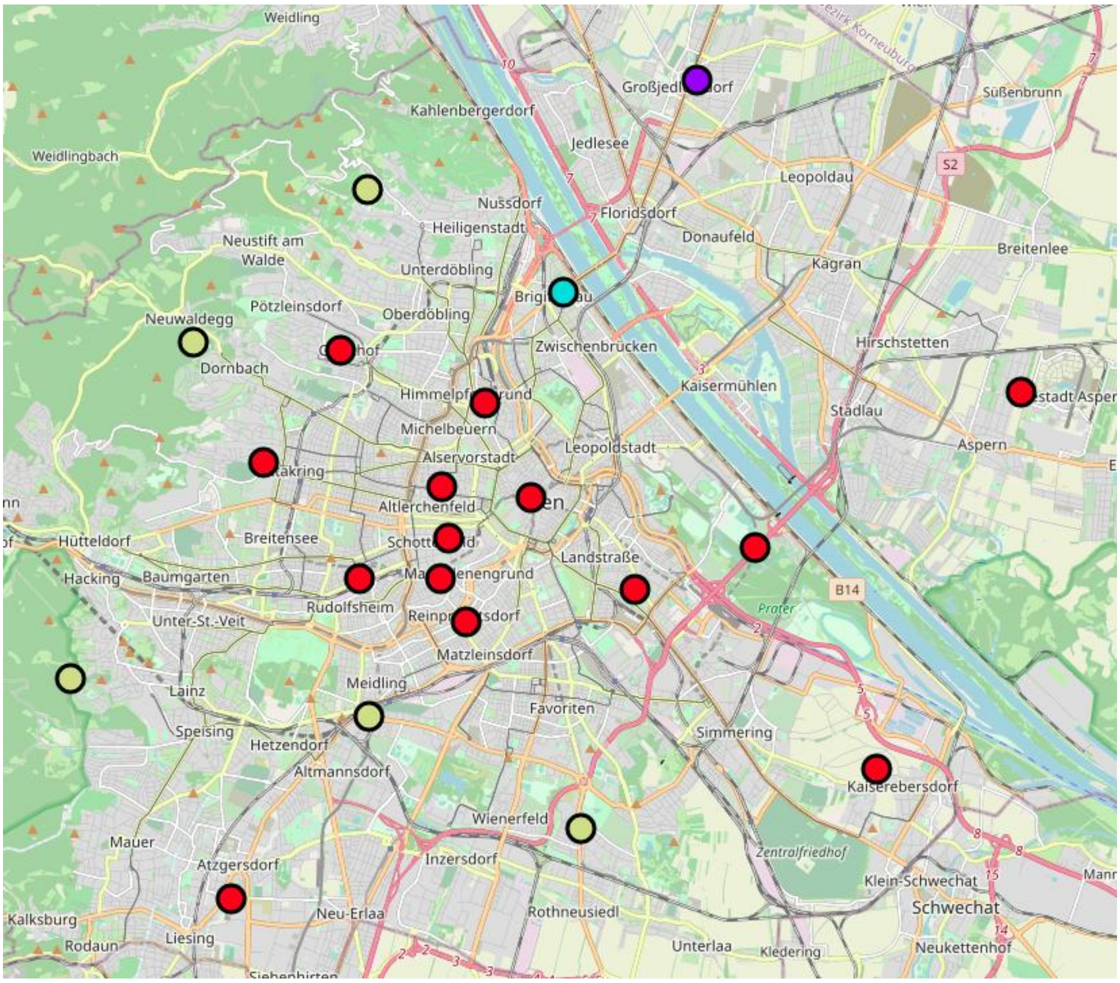
The venue category information was taken and assigned to each district. The one hot encoding method was applied and the mean of every venue category was calculated to create a venue profile for each district. With a profile for every district we are finally able to apply further machine learning methods.

3.2. Machine Learning Methods

Since we are interested in the similarities of the venue structure of each district, I have decided to apply clustering. For that I have set the number of clusters to 4, representing the 4 major political parties in Vienna. After feeding the model with our venue profile data, I was able to acquire some insightful data.

4. Results

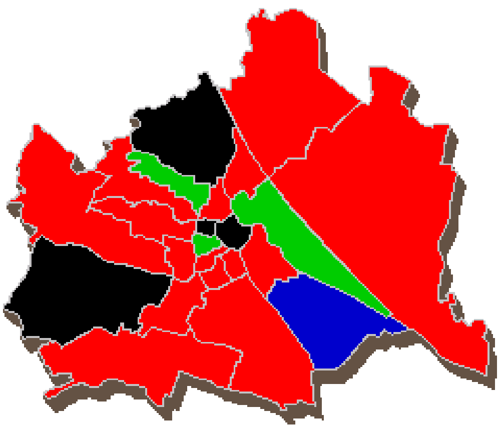
The following map shows the result of our clustering model.



Map 3. Result of clustering with 4 clusters.

As we can see the central districts (red) seem to be quite similar regarding their venue profile with some exceptions. The yellow clusters are all located on the outskirts. The Turquoise and purple clusters only have 1 district each so it is not possible to draw any connections with other districts.

If we compare these results with the map with its political affiliation, we cannot really draw any safe conclusions.



Map 4. Districts of Vienna with their political party affiliation.

5. Discussion

Our initial hypothesis that there might be a correlation between the venue structure and political affiliation of a district could not be confirmed. However, we were able to see a clear difference in the venue structure based on the location of the districts. Inner districts seem to be quite similar based on the kind of venues compared to the outer ones. The population density might be a reason why the mix of venues might be so different. One thing for sure is that there might be an underlying reason or factor for this difference and would be interesting for further study or exploration.

6. Conclusion

We can say that the political management of a district has little or no influence to the venue composition of a district. It makes sense, since these venues are opened by private companies and not by the government. However, the governing party still has some influence, especially in regards to developing the district such as parks, streets or buildings. This should have some influence but with the venue data from Foursquare we do not have sufficient data to further analyze it. That is why other data sources are required for further study.