Contents

[Introduction 2](#_Toc14962968)

[Objective 3](#_Toc14962969)

[Data 3](#_Toc14962970)

[Methodology 7](#_Toc14962971)

[Analyze Cities 7](#_Toc14962972)

[K-mean Cluster cities 9](#_Toc14962973)

[Result 13](#_Toc14962974)

[Discussion 21](#_Toc14962975)

[Conclusion 22](#_Toc14962976)

**Introduction**

The government wants to maintain the live ability of areas where the population is shrinking or where decline is forecast. Populations are declining in several parts of the Netherlands, especially northern and eastern Groningen and southern Limburg. This could mean more unoccupied homes and more schools having to close down. The provincial and municipal authorities hold primary responsibility for tackling the consequences of population decline and demographic ageing. In this, they receive support from central government.

Brief information about both cities:

**Heerlen** (Dutch pronunciation: [[ˈɦeːrlə(n)]](https://en.wikipedia.org/wiki/Help:IPA/Dutch) ([About this sound](https://en.wikipedia.org/wiki/File:135_Heerlen.ogg)[listen](https://upload.wikimedia.org/wikipedia/commons/0/09/135_Heerlen.ogg)); [Limburgish](https://en.wikipedia.org/wiki/Limburgish_language" \o "Limburgish language): *Heële*) is a [city](https://en.wikipedia.org/wiki/City) and a [municipality](https://en.wikipedia.org/wiki/Municipality) in the southeast of the [Netherlands](https://en.wikipedia.org/wiki/Netherlands). It is the third largest settlement proper in the province of [Limburg](https://en.wikipedia.org/wiki/Limburg_(Netherlands)). Measured as municipality, it is the fourth municipality in the province of [Limburg](https://en.wikipedia.org/wiki/Limburg_(Netherlands)).

Heerlen forms part of the city-region of [Parkstad Limburg](https://en.wikipedia.org/wiki/Parkstad_Limburg" \o "Parkstad Limburg), an agglomeration with about 250,000 inhabitants and encompassing 8 municipalities. It is to the east of [Maastricht](https://en.wikipedia.org/wiki/Maastricht) and north of the German city of [Aachen](https://en.wikipedia.org/wiki/Aachen).

After its early Roman beginnings and a modest medieval period, Heerlen became a centre for the coal mining industry in the Netherlands in the late 19th century. In the 20th century, architect [Frits Peutz](https://en.wikipedia.org/wiki/Frits_Peutz) played a major role in shaping the city as we know it today. His most famous design, and a distinctive building in the city centre, is the so-called [Glaspaleis](https://en.wikipedia.org/wiki/Glaspaleis" \o "Glaspaleis) (*Glass Palace*), listed as one of the world's thousand most architecturally important buildings of the 20th century.

(source: <https://en.wikipedia.org/wiki/Heerlen>)

**Kerkrade** ([Kerkrade dialect](https://en.wikipedia.org/wiki/Kerkrade_dialect): *Kirchroa*; [German](https://en.wikipedia.org/wiki/German_language): *Kerkrade* or *Kirchrath*)[[5]](https://en.wikipedia.org/wiki/Kerkrade#cite_note-5) is a [town](https://en.wikipedia.org/wiki/Town) and a [municipality](https://en.wikipedia.org/wiki/Municipality) in the southeast of [Limburg](https://en.wikipedia.org/wiki/Limburg_(Netherlands)), the southernmost province of the [Netherlands](https://en.wikipedia.org/wiki/Netherlands). It forms part of the [Parkstad Limburg](https://en.wikipedia.org/wiki/Parkstad_Limburg" \o "Parkstad Limburg)agglomeration.

Kerkrade is the western half of a [divided](https://en.wikipedia.org/wiki/Divided_region) city; it was part of the [German](https://en.wikipedia.org/wiki/Germany) town of [Herzogenrath](https://en.wikipedia.org/wiki/Herzogenrath" \o "Herzogenrath) until the [Congress of Vienna](https://en.wikipedia.org/wiki/Congress_of_Vienna) in 1815 drew the current Dutch-German border and separated the towns.[[6]](https://en.wikipedia.org/wiki/Kerkrade#cite_note-nijmegen-6)

The two towns, including outlying suburban settlements, have a population approaching 100,000, of which nearly 47,000 are in Kerkrade.

(source: <https://en.wikipedia.org/wiki/Kerkrade>)

**Objective**

In this project, we will study in details the area classification using Foursquare data and machine learning segmentation and clustering. The aim of this project is to segment areas of Kerkrade and Heerlen based on the most common places captured from Foursquare.

Using segmentation and clustering, we hope we can determine:

* the similarity or dissimilarity of both cities
* classification of area located inside the city whether it is residential, tourism places, or others

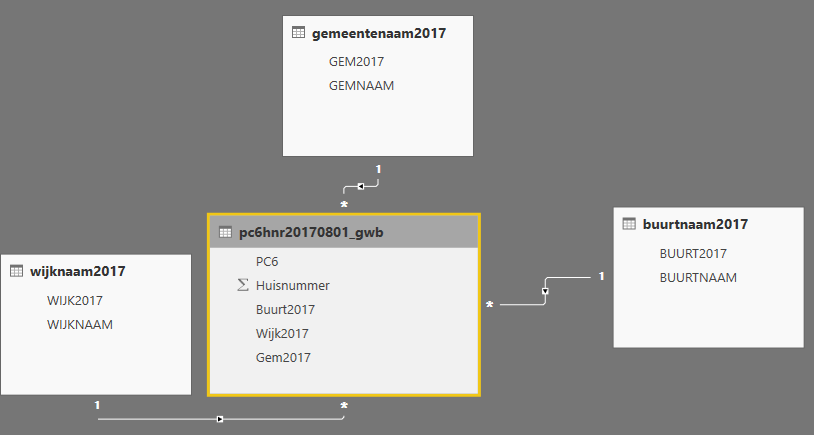
**Data**

The cities in the data set are divided in Neighborhoods the city of Kerkrade has a total of 10 boroughs and 34 neighborhoods, the city of Heerlen has a total of 14 boroughs and 82 neighborhoods . In order to segment the neighborhoods and explore them, we will essentially need a dataset that contains all the boroughs and the neighborhoods that exist in each borough as well as the latitude and longitude coordinates of each neighborhood.

Luckily, this dataset exists for free on the web. CBS. (2017b). Postalcode, house numbers per neighbourhood. Retrieved july 24, 2019, from <https://www.cbs.nl/-/media/_excel/2017/38/2017-cbs-pc6huisnr20170801_buurt.zip>

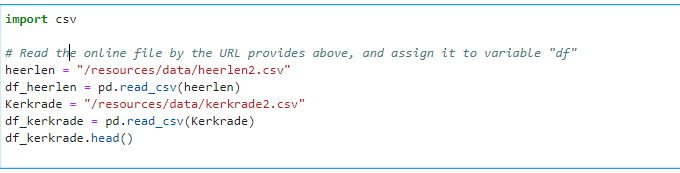
Another aspect to consider for this project is the Foursquare data. I believe that the data as good as provided, meaning although we are using Foursquare data for segmentation and clustering, the amount and accuracy of data captured can't 100% determine correct classification in real world.

The data was real big (+200MB of multible tables with links) I used PowerBI to extract the data for the use of the program. The following rwelations exist in the source data set

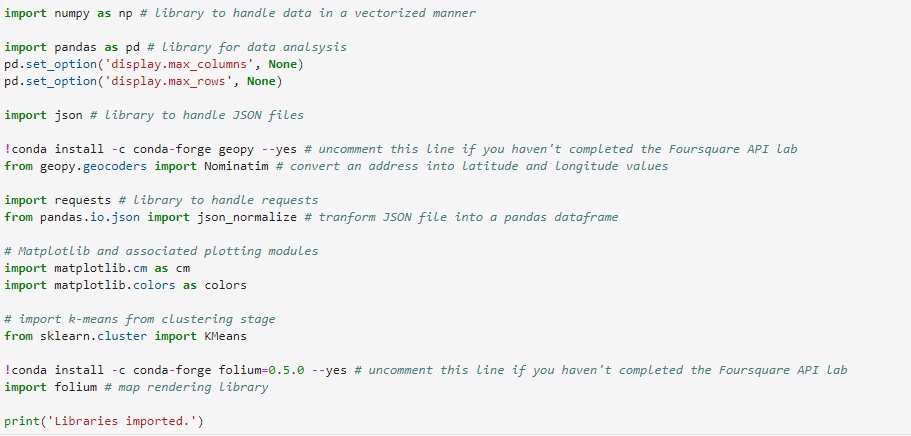


Power BI data cleansing/extraction from big files

I saved the data as a local file and imported the file as CSV



I imported the libraries for use



The next step is to identify the centre of the cities



**Methodology**

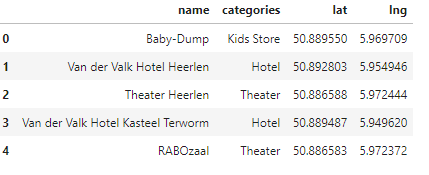
In this project, I will use the basic methodology as taught in Week 3 lab.

Above, we have done convert addresses into their equivalent latitude and longitude values. Then we will use the Foursquare API to explore neighborhoods in both cities, Kerkrade and Heerlen. After that, explore function to get the most common venue categories in each neighborhood, and then use this feature to group the neighborhoods into clusters K-means clustering algorithm will be use to complete this task. And also, the Folium library to visualize the neighborhoods in Kerkrade and Heerlen and their clusters.

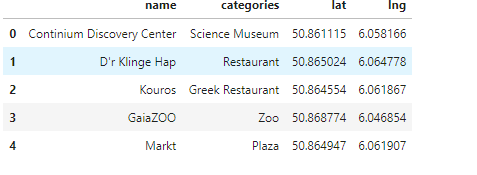
**Analyze Cities**

Based on data and venue extract we come to the most common venues in Heerlen and Kerkrade

Heerlen



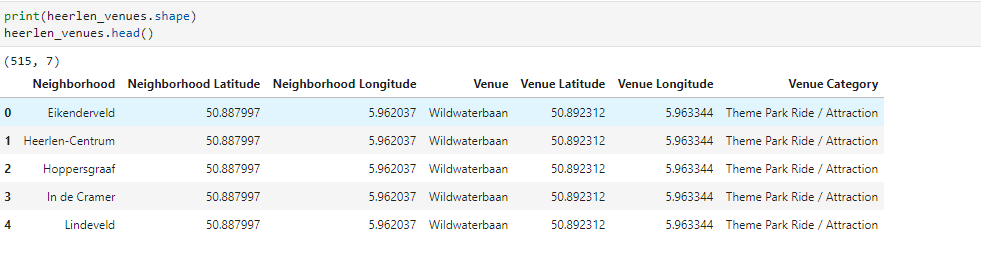
Kerkrade



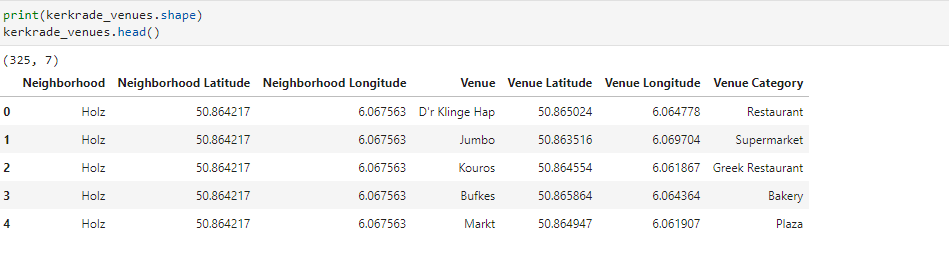
Based on this comparison we already see that Kerkrade has more the focus on the entertainment and Heerlen more on culture.

After further analysis we see the following split and venues for Heerlen and Kerkrade

Heerlen

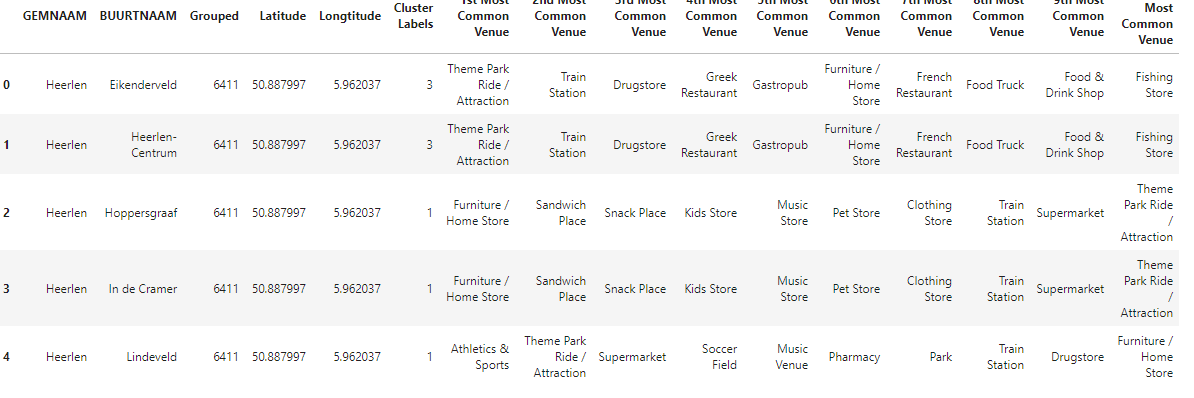


Kerkrade

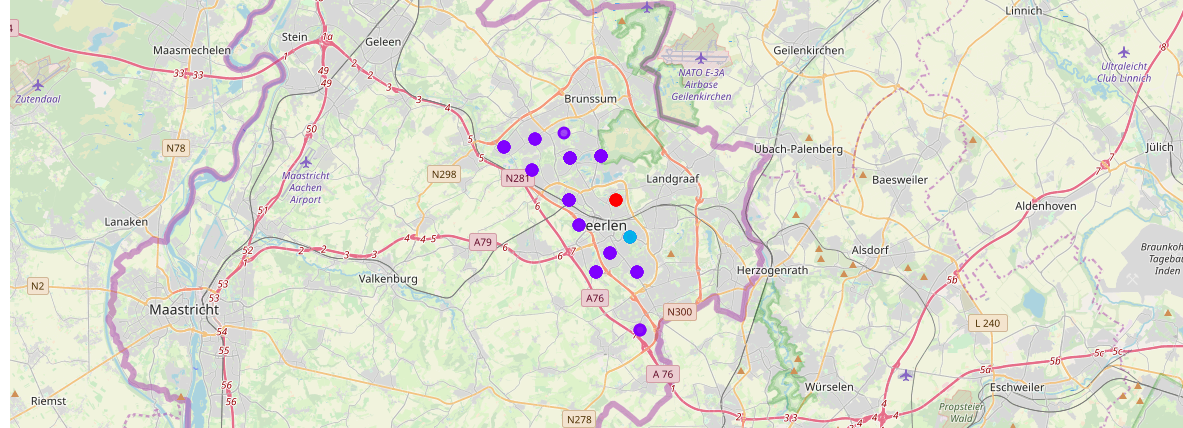


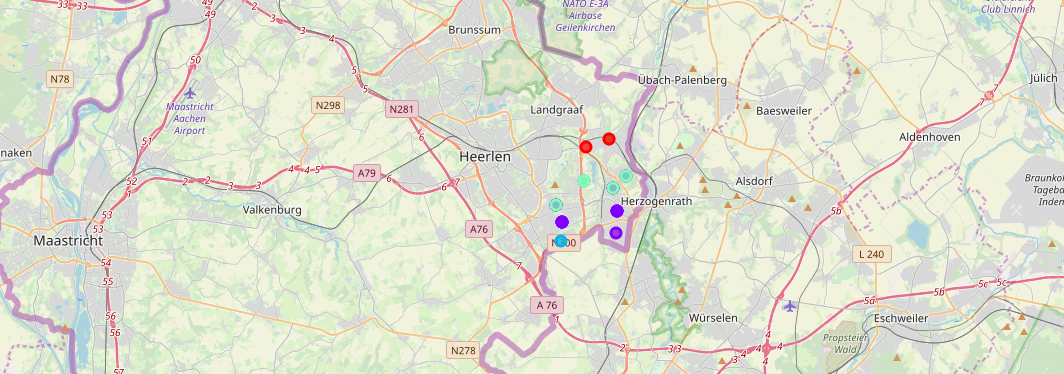
**K-mean Cluster cities**

Using K-mean to clustering data area with most common venue









Most common venues

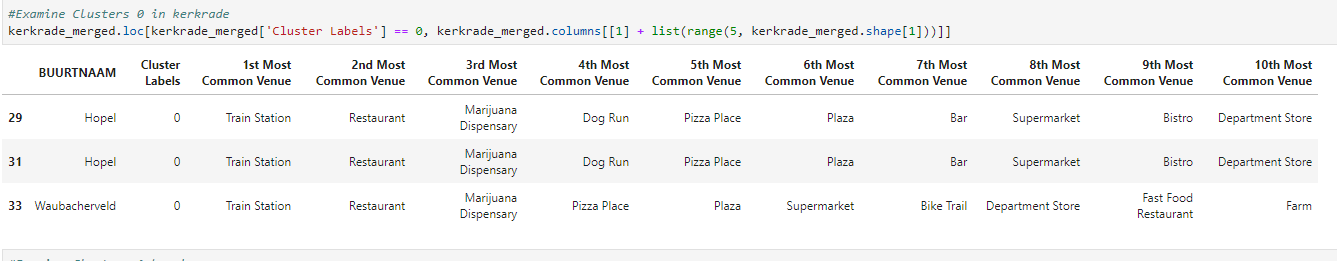
Heerlen



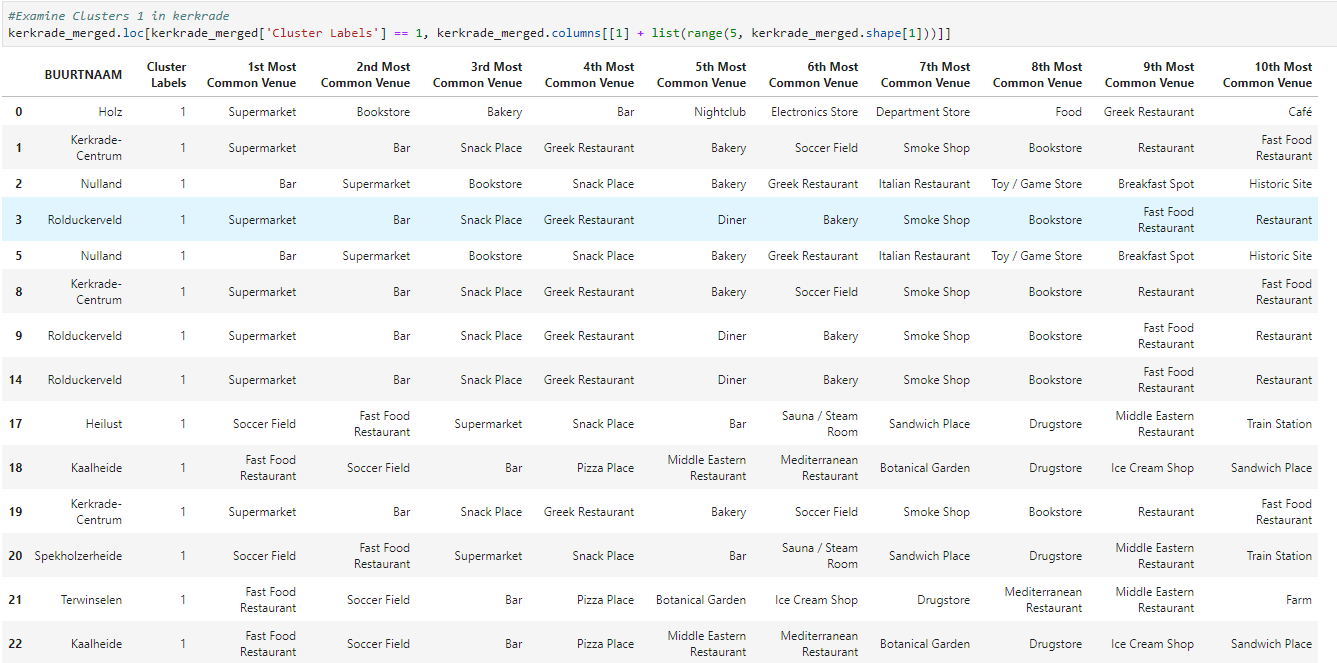
Kerkrade



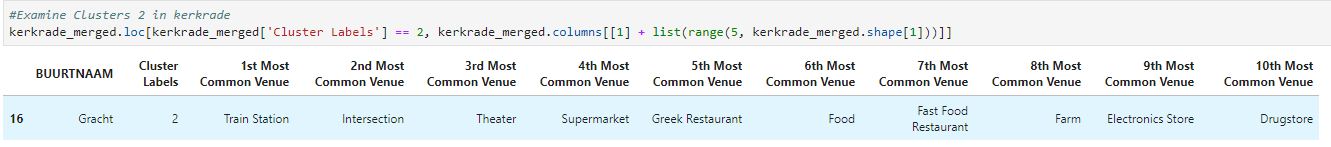
**Result**

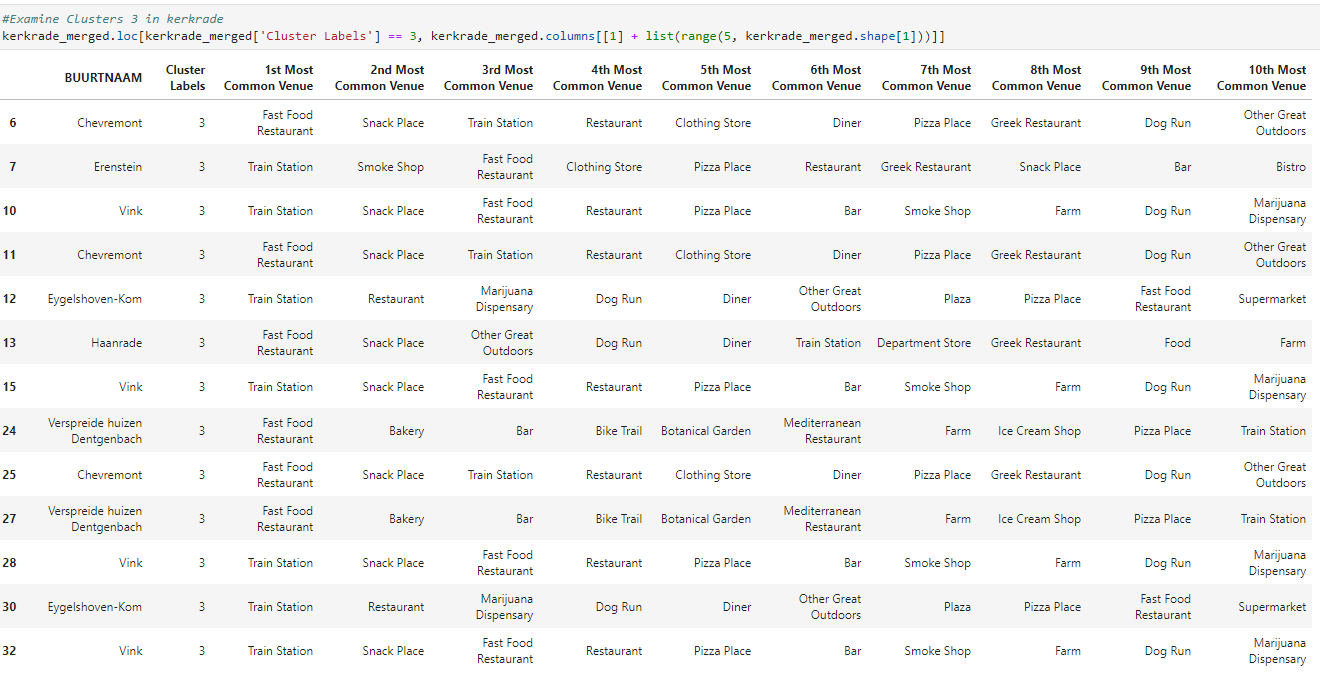
Cluster 0 for Kerkrade 

Cluster 1 for Kerkrade

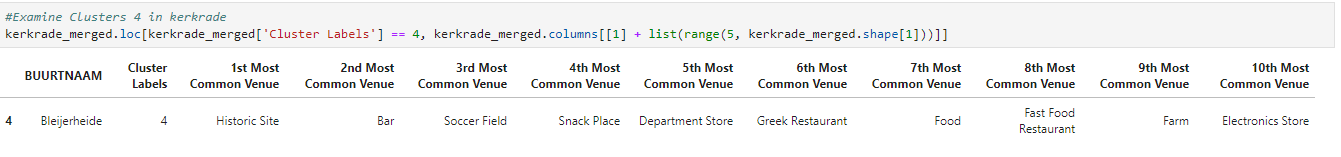


Cluster 2 for Kerkrade

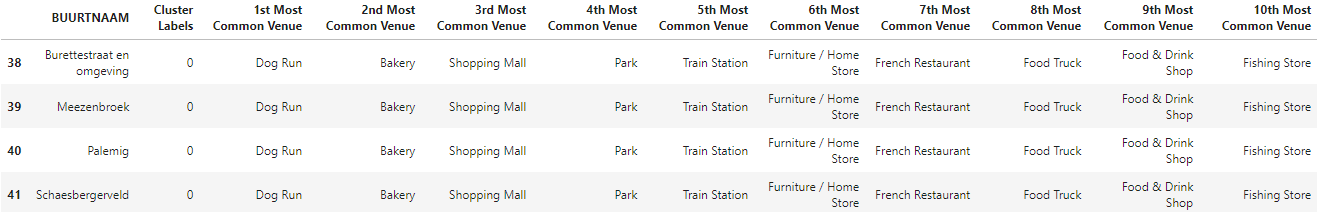


Cluster 3 for Kerkrade

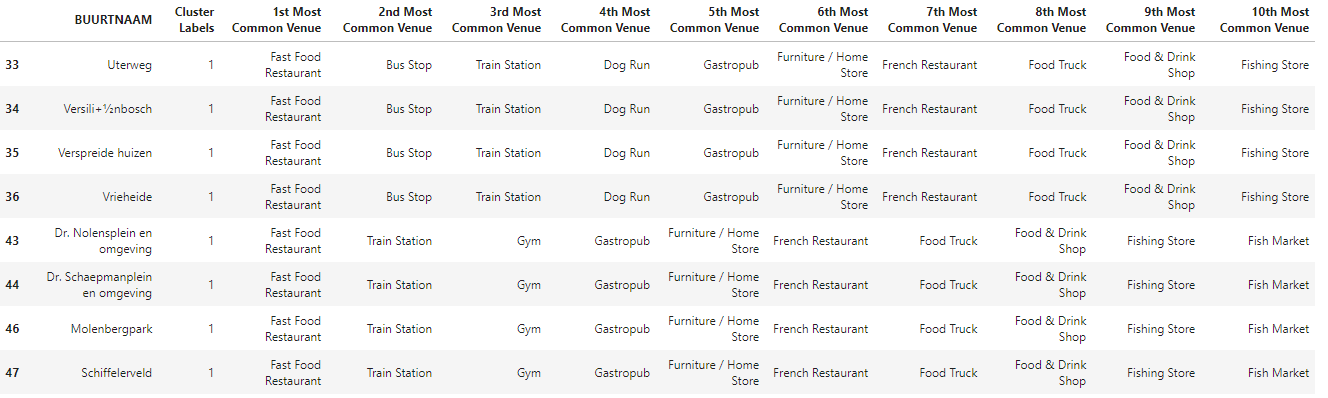
 Cluster 4 for Kerkrade



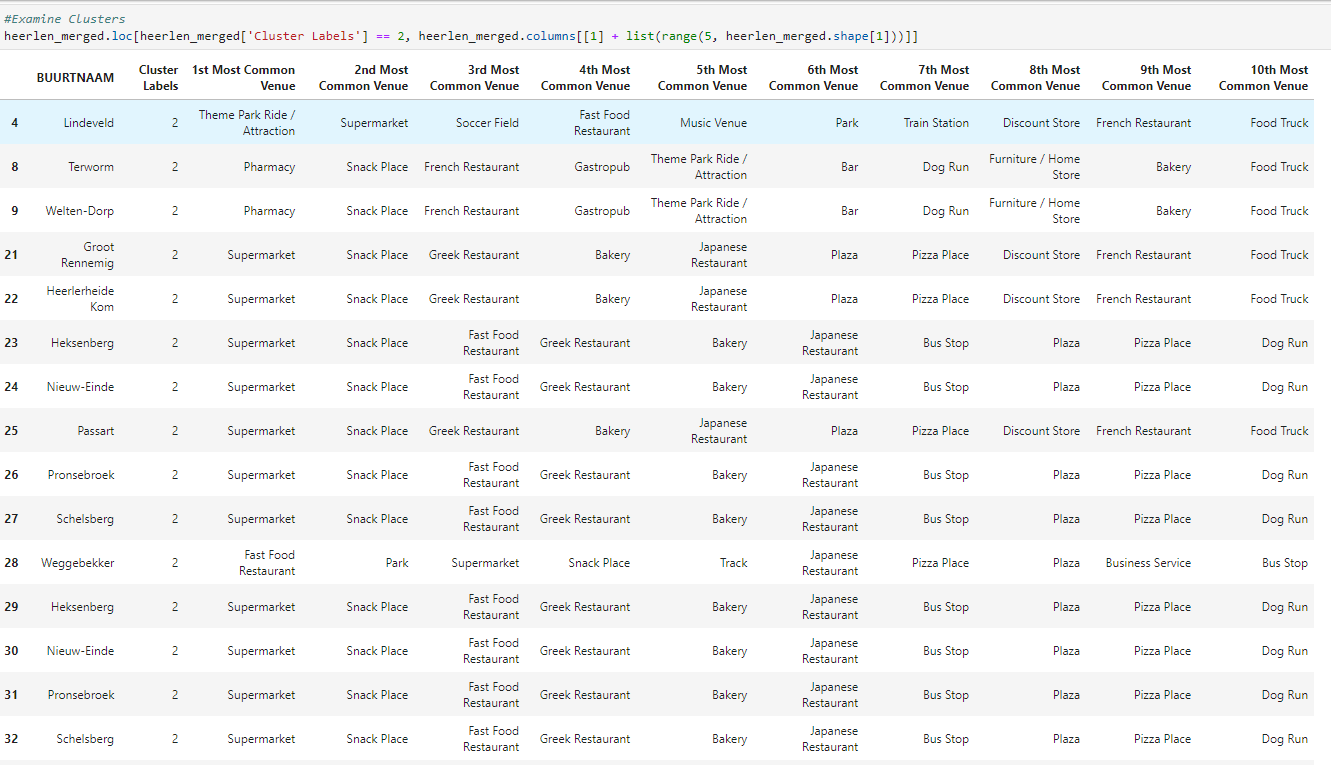
Cluster 0 Heerlen



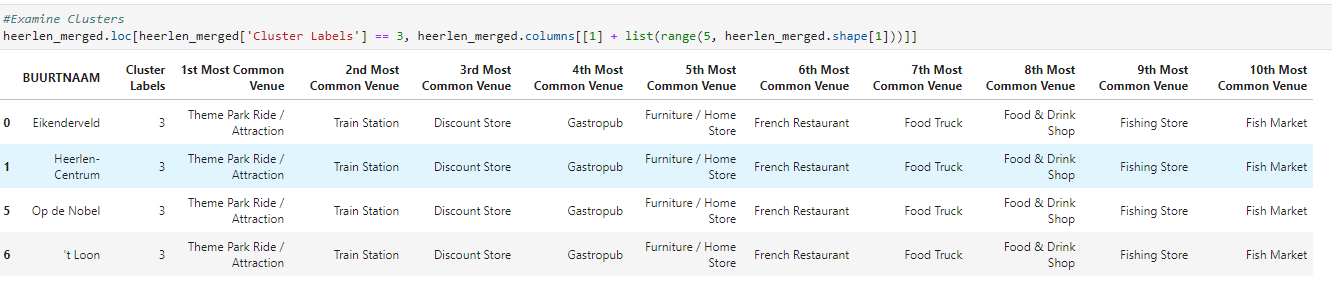
Cluster 1 for Heerlen



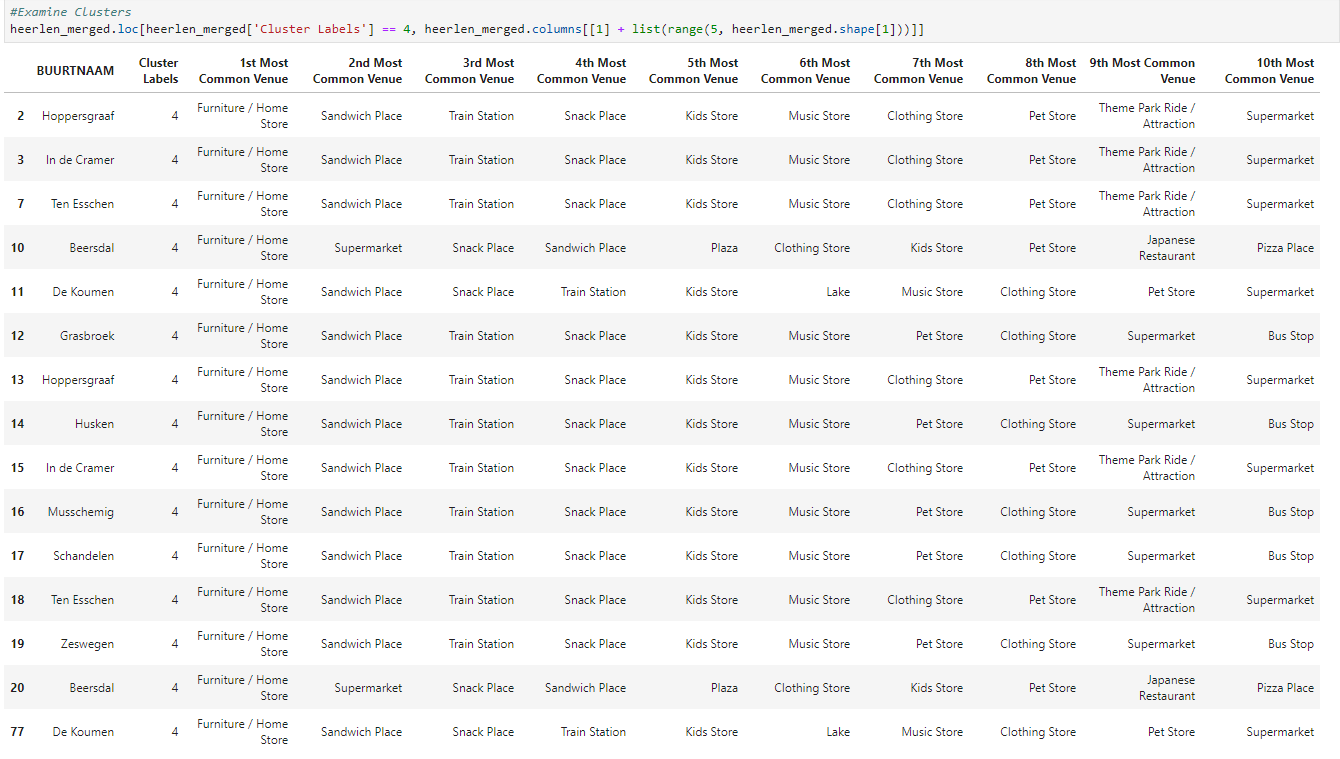
Cluster 2 for Heerlen



Cluser 3 Heerlen



Cluster 4 Heerlen



**Discussion**

Based on cluster for each cities above, we believe that classification for each cluster can be done better with calculation of venues categories (most common) in each cities. Refering to each clsuter, we can't deterimine clearly what represent in each cluster by using Foursquare - Most Common Venue data.

However, for the purpose of t project I assumed each cluster as follow:

|  |  |  |
| --- | --- | --- |
| Place | Cluster | Description |
| Kerkrade | 0 | Shopping |
| Kerkrade | 1 | Tourism |
| Kerkrade | 2 | Shopping\_Groceries |
| Kerkrade | 3 | Recreation |
| Kerkrade | 4 | Tourism\_historic |
| Heerlen | 0 | Park\_recreation |
| Heerlen | 1 | Eat places |
| Heerlen | 2 | Shopping |
| Heerlen | 3 | Tourism |
| Heerlen | 4 | Shopping |

What is lacking at this point is a systematic, quantitative way to identify and distinguish different district and to describe the correlation most common venues as recorded in Foursquare. The reality is however more complex: similar cities might have or might not have similar common venues. A further step in this classification would be to find a method to extract these common venues and integrate the spatial correlations between different of areas or district.

I believe that the classification we propose is an encouraging step towards a quantitative and systematic comparison of the different cities. Further studies are indeed needed in order to relate the data acquired, then observe it to more meaningful and objective results.

**Conclusion**

Using Foursquare API, we can captured data of common places all around the world. Using it, we refer back to our main objectives, which is to determine;

The similarity or dissimilarity of both cities classification of area located inside the city whether it is residential, tourism places, or others In conclusion, both cities Heerlen and Kerkrade are reinventing each other to be attractive in the region. People are not coming to a city because of shops anymore However, to declare both cities are similar or dissimilar base on common venues visited is quite difficult. Both cities is similar in some venues also dissimilar in certain venues. And for classification based on common venues, again we must have more systematic or quantitative way to identify and declare this. Comparison can be made, but no such method or quantitative data to determine this. We hope in the future, a method to determine it can be establish and explore for references.