The Battle of the Neighborhoods in Middelburg,

The Netherlands

Teun Huibregtse

Table of contents

1. Preface

**In Dutch:

Dit bestand (notebook) en bijbehorende bijdragen is het resultaat van mijn laatste IBM Course voor Datascience, in dit notebook mag ik mijn eigen probleemdefinitie schrijven en hier tevens een oplossing voor bieden, het is echter noodzakelijk om bij het verzamelen van de data gebruik te maken van de Foursquare API, de API kan gebruikt worden voor het ophalen van interesses en beoordelingen van bijvoorbeeld restaurants of de lokale sportschool binnen een eigen gedefinieerde radius of het ophalen van gebruikersdata van Foursquare profielen, het uiteindelijke doel van de course is de mogelijkheid om problemen te beantwoorden van stakeholders met vraagstellingen als:

- Als ik zou verhuizen naar een plaats dichter bij mijn werk, naar welke wijk zou ik dan het beste naar toe kunnen verhuizen op basis van mijn intresses en behoeften?

* The rest of the document will be in English as this is the language that will help me pass the assignment, please note that my English language and grammar usage isn't the best, so please don't consider this as a professional document.

2. Introduction

In this chapter I will give you a description of the problem and a discussion of the background:

2.1. Background

This document will be the final report and end result of the capstone project in the IBM Data Science course. In this project I will use the data Foursquare API and Datascience practices to explore neighborhoods in Middelburg as this is my favorite city in Zeeland, The Netherlands.

2.2. Problem

I will explore the neighborhoods of Middelburg using the available data to solve and answer the following question:

- What neighborhood is the best appropriate place to open a new restaurant in Middelburg based on the already existing restaurants in the neighborhood?

2.3. Why Middelburg?

Middelburg is my favorite city because it's considered as a student's city in the province of Zeeland and I'm a student myself, the city has lots of bars and the people are rather young compared to neighboring cities and towns/villages, and I would like to live in Middelburg one day.

Also restaurants in The Netherlands have a hard time lately coping with government tax and salary of the workers and even finding the right amount of people willing to help running the business, so this tool would help owners getting the needed insights to start as efficiently and effectively as possible by getting better revenue streams based on the already available data.

2.4. Target audience and stakeholders

The target audience (and stakeholder) will be restaurant owners interested in opening a new restaurant in Middelburg or existing restaurants who are considering moving to a different location inside the City of Middelburg.

Also the municipality of Middelburg could play a role as a permit mediator to allow or dismiss a permit to a restaurant owner based on the needs of that particular neighborhood when issuing a new permit to deploy a restaurant.

3. Data

In this chapter I will give you a description if the data and how it will be used to solve the problem:

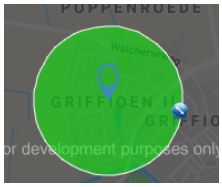
3.1. Available Data Sources

The following types of data will be used*:

Source	Link	Problem solver
Wikipedia	https://nl.wikipedia.org/wiki/Wijken_en_buurten_in_Middelburg	List of all neighborhoods in Middelburg according to CBS (Central Bureau of Statistics)
Maps.ie	https://www.maps.ie/coordinates.html	Finding the right coordinates of the neighborhoods in Middelburg
Foursquare API	https://developer.foursquare.com/	Scraping and gathering relevant information such as neighboring restaurants and it's menu/type of food based on its coordinates and radius

* It would be more accurate to get open data from the municipality Middelburg as the coordinates would be a lot more correct, however this would take a lot more time as I have to file a report and wait a couple of months. Neighborhoods with an industrial character will be ignored as seen in maps.ie as well as neighborhoods with less than 1000 residents.

3.2. Foursquare API and data gathering



Using the Foursquare API and the coordinates of the Neighborhoods inside Middelburg, I will be extracting and calculating the top 10 most common venues already available in the neighborhoods of Middelburg inside a 800 meters radius, from here we could calculate if the restaurant is really needed in the neighborhood, maybe there are already five Italian restaurant's so another Italian restaurant wouldn't necessarily be needed.

If a restaurant type such as sushi isn't available in the neighborhood, then the recommendation would be to place a new sushi restaurant there, as there will be more need and interests from the community or putting it in another direction there will be no direct competition.

3.3. Data Cleansing and Formatted data

The data will be focused on the neighborhoods of the City of Middelburg in The Netherlands, de following rules apply when extracting and selecting the data:

Data source or method	Rule
Wikipedia Neighborhoods of Middelburg	Ignore neighborhoods with less than 1000 residents
Maps.ie	Ignore neighborhoods with an industrial character
Foursquare API	Set a radius of 800 meters (average radius of each neighborhood) when exploring common venues
Pd.read_csv()	Csv file should be named 'neighborhoods_middelburg.csv'

After formatting and creating/extracting the data of the panda data frame 'neighborhoods_middelburg.csv' should look like this:

	Neighborhood	Latitude	Longitude
0	Binnenstad	51.499732	3.613752
1	Griffioen	51.506202	3.595598
2	Klarenbeek	51.508051	3.610783
3	Nieuw Middelburg	51.506233	3.625254
4	Veersepoort	51.512197	3.622476
5	Stromenwijk 't Zand	51.491164	3.599305
6	Middelburg-Zuid	51.484007	3.614082
7	Dauwendaele	51.492077	3.625996
8	Mortiere	51.483405	3.635964
9	Nieuw - en Sint Joosland	51.481970	3.656884
10	Arnemuiden	51.500001	3.676246

3.4. Feature Selection

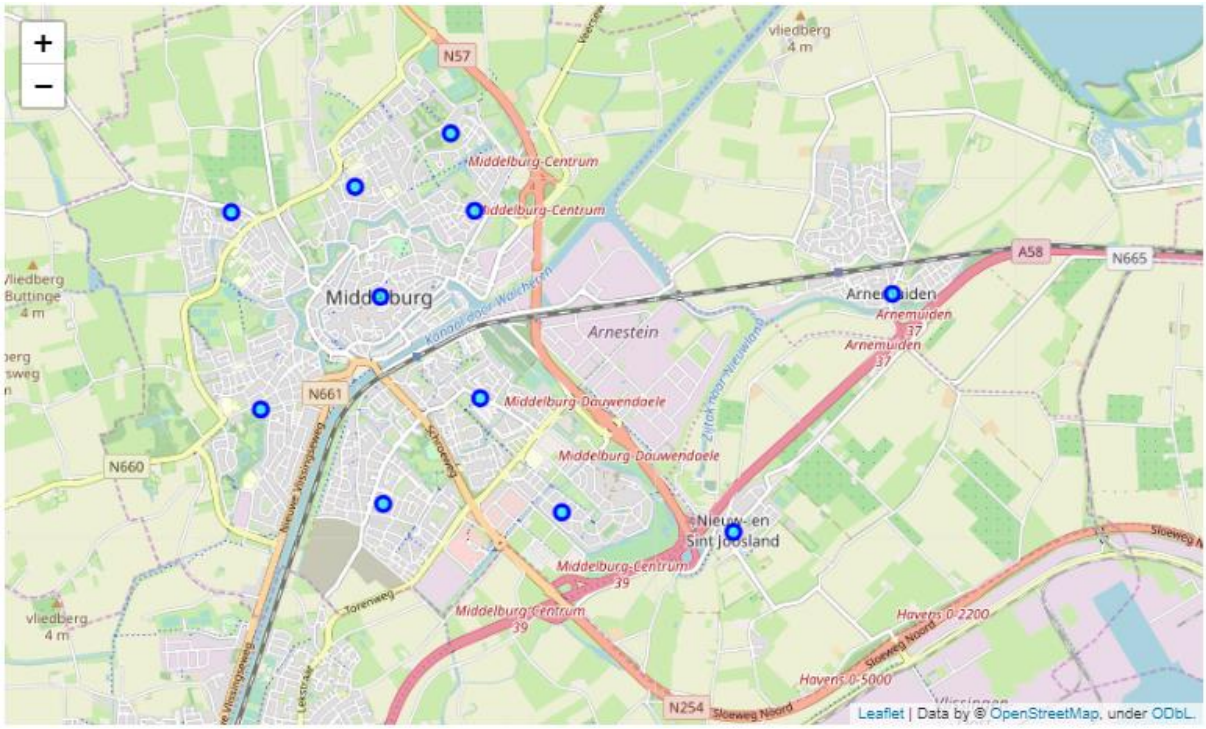
When getting the data of common venues inside a 800 meters radius of every neighborhood, calculate the top 10 most common venues based on its category as a feature.

Label	Type	Description
1st Most common Venue	String.	Venue category name
2nd Most common Venue	String.	Venue category name
3rd Most common Venue	String.	Venue category name
4th Most common Venue	String.	Venue category name
5th Most common Venue	String.	Venue category name
6th Most common Venue	String.	Venue category name
7th Most common Venue	String.	Venue category name
8th Most common Venue	String.	Venue category name
9th Most common Venue	String.	Venue category name
10th Most common Venue	String.	Venue category name

4. Methodology

4.1. Explore and Visualize a Dataset

To get a good view of the accuracy of the coordinates of the neighborhoods in de data frame I'm going to use the folium library (python-visualization.github, sd) inside the Jupyter notebook (jupyter, sd) using the Python 3.6 kernel of the used IBM Watson Studio (IBM Cloud watson-studio, sd) to visualize the coordinates:



It looks like all the markers are in the correct place on the map corresponding to the defined neighborhoods.

4.2. Extracting and examining Foursquare Location data

Using the following property's inside the Foursquare API (Foursquare Developers, sd) we should get up to 100 venues for each coordinate of a neighborhood in a 800 meters radius:

Foursquare property's	**Setting**
Type	'/venues/explore'
Limit	100
Radius	800
Version	20180605

Sample result:


```
{'meta': {'code': 200, 'requestId': '5dd5350c3907e7001be7f2f1'},
 'response': {'suggestedFilters': {'header': 'Tap to show:',
   'filters': [{'name': 'Open now', 'key': 'openNow'}]},
  'headerLocation': 'Middelburg',
  'headerFullLocation': 'Middelburg',
  'headerLocationGranularity': 'city',
  'totalResults': 71,
  'suggestedBounds': {'ne': {'lat': 51.5069319072, 'lng': 3.625296541714101},
   'sw': {'lat': 51.49253189279999, 'lng': 3.6022078582858996}},
  'groups': [{'type': 'Recommended Places',
    'name': 'recommended',
    'items': [{'reasons': {'count': 0,
       'items': [{'summary': 'This spot is popular',
         'type': 'general',
         'reasonName': 'globalInteractionReason'}]},
      'venue': {'id': '4bdaf85863c5c9b6dba92568',
       'name': "B'tje Anders",
       'location': {'address': 'Reigerstraat 3',
        'lat': 51.499650648136054,
        'lng': 3.6158709231274946,
```


4.3. Data pre-processing

After exploring all it's venues of each neighborhood using the Foursquare API we get the following output eventually (example Binnenstad Neighborhood):

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Category	Venue Latitude	Venue Longitude
0	Binnenstad	51.499732	3.613752	B'tje Anders	Fast Food Restaurant	51.499651	3.615871
1	Binnenstad	51.499732	3.613752	De Drukkerij	Bookstore	51.497921	3.611670
2	Binnenstad	51.499732	3.613752	Markt	Plaza	51.498415	3.611665
3	Binnenstad	51.499732	3.613752	St John	Café	51.497060	3.614056
4	Binnenstad	51.499732	3.613752	Biercafé De Vliegende Hollander	Bar	51.500203	3.617550

Then apply one-hot encoding to the venues categories to extract the mean of each category for every neighborhood:

	Neighborhood	Arcade	Asian Restaurant	Athletics & Sports	Bagel Shop	Bakery	Bar	Basketball Court	Beer Garden	Boat or Ferry	...
0	Arnemuiden	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	0.142857	0.000000	0.000000	...
1	Binnenstad	0.000000	0.0	0.000000	0.014085	0.028169	0.070423	0.000000	0.014085	0.014085	...
2	Dauwendaele	0.000000	0.0	0.000000	0.000000	0.066667	0.000000	0.000000	0.000000	0.000000	...
3	Griffioen	0.166667	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...
4	Klarenbeek	0.000000	0.0	0.125000	0.000000	0.125000	0.000000	0.000000	0.000000	0.000000	...
5	Middelburg-Zuid	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...
6	Mortiere	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...
7	Nieuw - en Sint Joosland	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...
8	Nieuw Middelburg	0.000000	0.0	0.071429	0.000000	0.071429	0.000000	0.000000	0.000000	0.000000	...
9	Stromenwijk 't Zand	0.000000	0.1	0.100000	0.000000	0.000000	0.000000	0.100000	0.000000	0.000000	...
10	Veersepoort	0.000000	0.0	0.166667	0.000000	0.083333	0.000000	0.000000	0.000000	0.000000	...

When you have the mean of every value, you can now create a new data frame with the top 10 most common venues for each neighborhood that is ready to be used for clustering:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Arnemuiden	Train Station	Gas Station	Gym	Supermarket	Food Court	Basketball Court	Fast Food Restaurant	Yoga Studio	Electronics Store	Dance Studio
1	Binnenstad	Café	Restaurant	Bar	Hotel	Snack Place	Plaza	Concert Hall	Coffee Shop	Clothing Store	French Restaurant
2	Dauwendaele	Hotel	Fish Market	Chinese Restaurant	Bus Stop	Bus Station	Intersection	Bowling Alley	Farm	Playground	Bakery
3	Griffioen	Fast Food Restaurant	Arcade	Supermarket	Restaurant	Chinese Restaurant	Bakery	Farm	Dance Studio	Department Store	Diner
4	Klarenbeek	Park	Hotel	Athletics & Sports	Theater	Bakery	Supermarket	Food Court	Restaurant	Cosmetics Shop	Dance Studio
5	Middelburg-Zuid	Bus Stop	Supermarket	Shopping Mall	Snack Place	Drugstore	Furniture / Home Store	Cosmetics Shop	Dance Studio	Fast Food Restaurant	Intersection
6	Mortiere	Golf Course	Furniture / Home Store	Toy / Game Store	Fast Food Restaurant	Electronics Store	Buffet	Mini Golf	Mattress Store	Intersection	Restaurant
7	Nieuw - en Sint Joosland	Golf Course	Sporting Goods Shop	Intersection	Soccer Field	French Restaurant	Gas Station	Cosmetics Shop	Dance Studio	Department Store	Diner
8	Nieuw Middelburg	Yoga Studio	Sporting Goods Shop	French Restaurant	Fast Food Restaurant	Gym / Fitness Center	Harbor / Marina	Drugstore	Park	Pharmacy	Furniture / Home Store
9	Stromenwijk 't Zand	Sports Club	Snack Place	Asian Restaurant	Athletics & Sports	Grocery Store	Basketball Court	Discount Store	Café	Drugstore	Fast Food Restaurant
10	Veersepoort	Athletics & Sports	Supermarket	Fast Food Restaurant	Furniture / Home Store	Pharmacy	Bakery	Flower Shop	Plaza	Soccer Field	Drugstore

4.4. Clustering using k-means

In this project we are going to be using clustering using the k-means clustering algorithm because we have an unlabeled dataset, this is also known as an unsupervised learning project.

K-means aims to partition ‘n’ observations into ‘k clusters’ by which each observation of a cluster belongs with the nearest mean.

When clustering all the neighborhoods we can find out the pattern in them, identifying the identical neighborhoods and focusing on the target.

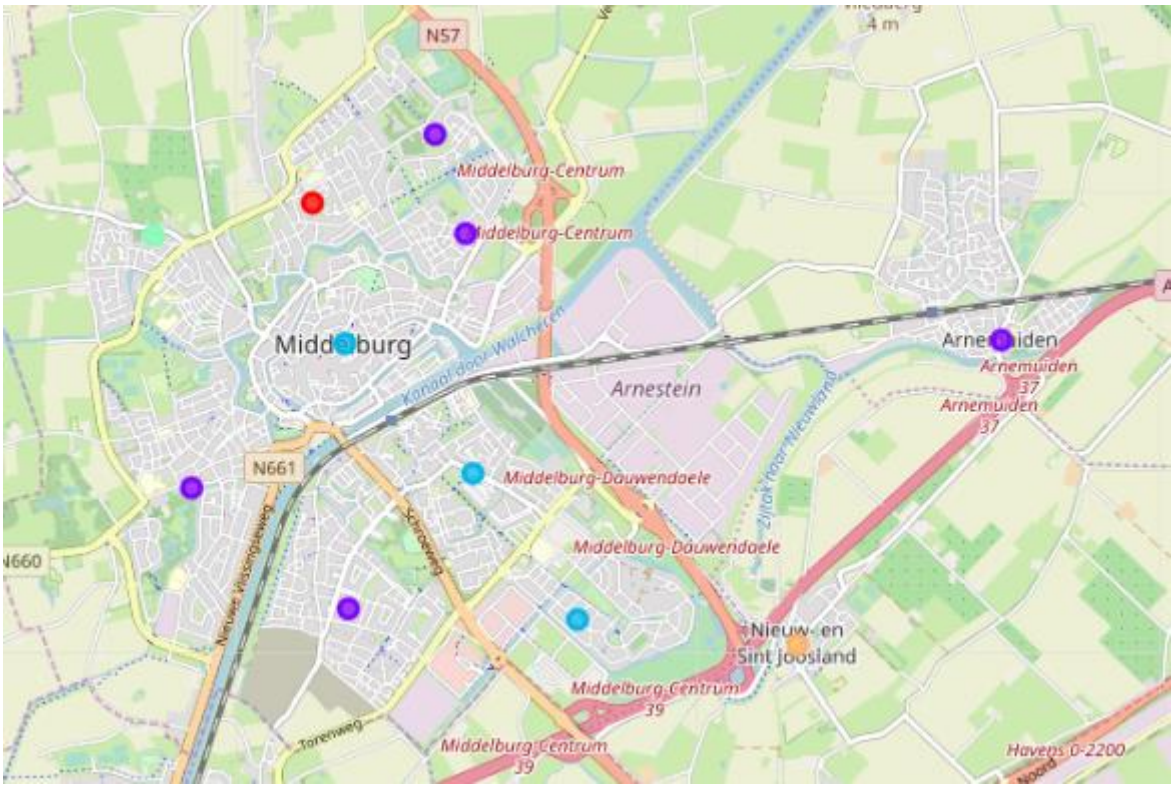
Examining and setting the right amount of k-means is very difficult to do and could be solved by using machine learning or other techniques when calculating the accuracy of the model, I had chosen ‘k=5’ as the experience hyperparameter for the model to cluster the neighborhoods inside five clusters.

Neighborhood	**Cluster**
Binnenstad	1
Griffioen	3
Klarenbeek	1
Nieuw Middelburg	1
Veersepoort	1
Stromenwijk 't Zand	0
Middelburg-Zuid	1
Dauwendaele	1
Mortiere	1
Nieuw - en Sint Joosland	2
Arnemuiden	4

5. Final Results

5.1. Visualizing the Clusters

Using the folium I generate the same map now with clustered markers:



We can see that there are three main clusters of cluster 3, outside of the city center you have cluster 2 who are very similar, and a few unique cluster with different characteristics.

5.2. Data frame of each cluster

5.2.1. Cluster 1

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
2	Klarenbeek	Hotel	Athletics & Sports	Bakery	Wine Shop	Theater	Supermarket	Snack Place	Skate Park	Yoga Studio	Cosmetics Shop

5.2.2. Cluster 2

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Nieuw Middelburg	French Restaurant	Snack Place	Furniture / Home Store	Flower Shop	Fast Food Restaurant	Wine Shop	Gym / Fitness Center	Harbor / Marina	Drugstore	Hotel
Veersepoort	Athletics & Sports	Supermarket	Plaza	Soccer Field	Flower Shop	Fast Food Restaurant	Drugstore	Pharmacy	Skate Park	Furniture / Home Store
Stromenwijk 't Zand	Chinese Restaurant	Athletics & Sports	Café	Basketball Court	Flower Shop	Supermarket	Sports Club	Discount Store	Snack Place	Drugstore
Middelburg-Zuid	Supermarket	Shopping Mall	Furniture / Home Store	Cosmetics Shop	Flower Shop	Fast Food Restaurant	Hotel	Sporting Goods Shop	Bus Stop	Drugstore
Arnemuiden	Soccer Field	Gas Station	Train Station	Gym	Food Court	Basketball Court	Supermarket	Construction & Landscaping	Jewelry Store	Drugstore

5.2.3. Cluster 3

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Binnenstad	Café	Restaurant	Bar	Hotel	Plaza	Park	Concert Hall	Coffee Shop	Clothing Store	Seafood Restaurant
Dauwendaele	Hotel	Bus Stop	Fast Food Restaurant	Playground	Café	Intersection	Bowling Alley	Convenience Store	Farm	Furniture / Home Store
Mortiere	Golf Course	Fast Food Restaurant	Toy / Game Store	Furniture / Home Store	Mini Golf	Restaurant	Bus Stop	Hotel Bar	Buffet	Paper / Office Supplies Store

5.2.4. Cluster 4

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Griffioen	Fast Food Restaurant	Arcade	Chinese Restaurant	Supermarket	Electronics Store	Convenience Store	Cosmetics Shop	Department Store	Diner	Discount Store

5.2.5. Cluster 5

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Nieuw - en Sint Joosland	Golf Course	Soccer Field	Intersection	Sporting Goods Shop	French Restaurant	Electronics Store	Convenience Store	Cosmetics Shop	Department Store	Diner

5.3. Final answer to the question

In the beginning the following question was defined:

- In a city of your choice, if someone is looking to open a restaurant, where would you recommend that they open it?

The city of my choice was obviously Middelburg, now the question depends on what type of restaurant there needs to be placed, such as a Chinese restaurant for example and if there's need for a particular restaurant in the neighborhood.

After examining the five clusters I have come to the following conclusion:

Cluster NO.	Neighborhood(s)	Answer
1	Klarenbeek	Cluster one has a snack place, an Asian restaurant, and a Fast food restaurant, looking at the most common venues I would consider that a different type than an Asian Restaurant or fast-food restaurant will be welcome in this neighborhood, however there doesn't seem to be a lot of restaurant, so it isn't a popular place. I would consider cluster one a residential neighborhood.
2	Nieuw Middelburg, Veersepoort, Stromenwijk 't Zand, Middelburg-Zuid, Arnemuiden	Cluster two has a lot more restaurants and café's than cluster one for example, I would consider this the outskirts neighborhoods of Middelburg because it got everything you need, depending on the type of restaurant there will be need. There's less need for a snack place, fast-food restaurant, and Chinese/French restaurant's.

3	Binnenstad, Dauwendaele, Mortiere	Cluster three seems to be the best place to start a new restaurant because of the large variety of restaurant's and it's localization as a city center type of neighborhoods, also more expensive restaurant's would be more suitable here because of the luxury activities such as golf for example.
4	Griffioen	Cluster four would also be a great place to start a new restaurant targeting younger people, there seem to be mostly younger people as there are dancing studio's and arcade game halls as well.
5	Nieuw - en Sint Joosland	Cluster five would be considered a place for the more quieter restaurants as it is (far) more outside of the center of Middelburg but also a lot less customers.

Conclusion:

Cluster three seems to have the best neighborhoods in Middelburg for a (new) Restaurant but a lot of needed variables depend on this conclusion as you can see in the above table.

6. Discussion

As previously answered (see 4.4. Clustering using k-means) the k-means algorithm works good for calculating the answer to the question of the project, although there are a few side notes:

- The used k-clustering of 5 clusters could be more accurately and precise when calculating the accuracy using different clusters numbering.
- Adding more features such as the accessibility of public transport, average income of residents in neighborhoods of Middelburg and population for better and more accurate calculations.
- The time to finish this project and assessment was minimal, as I had a very strict deadline from school and to myself to finish this course on time, so I had to make cuts along the way to finish on time such as calculating the best k-means clustering value.

7. Final conclusion

The answer to the **question** “What neighborhood is the best appropriate place to open a new restaurant in Middelburg based on the already existing restaurants in the neighborhood?”

Is: “Cluster three seems to have the best neighborhoods in Middelburg for a (new) Restaurant but a lot of needed variables depend on this conclusion as you can see in the above table.”

The neighborhood data is collect through Wikipedia using CBS (Central Bureau of Statistics) data and transformed using the k-means algorithm using five clusters and venue data of those neighborhoods using the Foursquare API with the most common top ten venue categories.

The models could be improved to get better results as discussed in the discussing chapter.

Reference list

(sd). Opgehaald van jupyter: https://jupyter.org/

Foursquare Developers. (sd). Opgehaald van Foursquare: https://developer.foursquare.com/

IBM Cloud watson-studio. (sd). Opgehaald van IBM : https://www.ibm.com/cloud/watson-studio

python-visualization.github. (sd). Opgehaald van Folium: https://python-visualization.github.io/folium/