

IBM Data Science  
Capstone Project: The Battle of Neighborhoods

## **Apartment hunting in emerging Neighborhoods Of Utrecht, NL**

Asif Anwar

Linkedin: <https://www.linkedin.com/in/anwarasif/>

## **Table of Contents**

<b>1. Introduction:</b>	<b>2</b>
<b>2. Data Acquisition:</b>	<b>3</b>
2.1: Rental Data Collection:	3
2.2 Location Data Collection:	3
2.3. Venue Data Collection:	5
<b>3. Data Exploration and Clean-Up:</b>	<b>5</b>
3.1 Neighbourhood data:	5
3.2. Venue Data:	7
3.3. Rental Data Exploration:	8
<b>4. K-means Clustering:</b>	<b>10</b>
Cluster 0 (RED):	11
Cluster 1 (PURPLE):	12
Cluster 2 (BLUE):	13
Cluster 3 (GREEN):	13
Cluster 4 (ORANGE):	14
<b>5. Selecting the apartments</b>	<b>15</b>
5.1 Apartment Clusters	15
<b>5.2 Finalizing Apartment</b>	<b>16</b>
<b>6. Conclusion</b>	<b>18</b>

# 1. Introduction:

I have recently moved to Utrecht, The Netherlands for my higher masters. Utrecht is a very beautiful and nice city in Netherland which is only 20 minutes intercity train ride away from Amsterdam. As Amsterdam is one of the most visited cities of the world by the tourists, the inhabitant of the Netherlands likes to stay away from the busy city centre and live in a quieter and peaceful place like Utecht. Not only the commute to Utrecht from most of the major cities is very good, but the city is also very popular for different cultural activates, events, universities, parks, bars and restaurants. As a result, the demand for an apartment rental is very high in Utrecht.

A large number of expat and locals are always looking for suitable housing in Utrecht. The target of this project will help all the apartment hunters to find a suitable apartment or house based on the budget in the emerging neighbourhoods of Utecht by data-driven analysis.

At a glance a target audience will able to get benefited from the below feature of this project:

- Finding the emerging neighbourhoods in Utecht, NL
- Compare between these new neighbourhoods
- Rent compare the apartments based on their size, rooms and the location in the new neighbourhood clusters
- Finally selecting the top apartments based on budget, location preference etc.

Personally, I was very lucky to find an apartment in the heart of Utrecht city centre. But it took me around three months to find a decent place for living. Although I really like the location of my apartment, the price I am currently paying feels really high considering the size of the place. So, this data-driven analysis to help me find a suitable rental apartment in the city so that I can compare with the current location and decide if I should move to a different location or not.

## 2. Data Acquisition:

To develop the recommendation engine for this data-driven project, I needed different sets of data on locations, venues, rentals, apartment price, size, zip code etc. I have selected below sources to get all required data for this project.

### 2.1: Rental Data Collection:

I used the most popular rental website [www.pararius.com](http://www.pararius.com) to collect the available rental locations in Utrecht. Rental data has been collected from this website by using BeautifulSoup Library in python. The rental information was saturated in a total of 14 pages of the website. By crawling through each page below data fields were collected for analysis.

	Title	Type	Postcode	Neighborhood	Rent	Size	Rooms	Furnished
0	Van Heesstraat	Apartment	3555GH	Noordwest	1250	55	1	upholstered
1	Abstederdijk	House	3582BE	Oost	1250	75	2	furnished
2	Hildebranddreef	Room	3561VD	Overvecht	49	27	1	furnished
3	Nieuwekade	Apartment	3511RW	Binnenstad	1590	72	2	furnished
4	Adelaarstraat	Apartment	3514CH	Noordoost	1495	75	2	furnished

### 2.2 Location Data Collection:

#### Neighbourhood Data:

To complete the analysis I needed the Neighbourhood data of Utrecht. Unfortunately, I could not find any source where all the data was readily available. Hence I had to crawl through multiple pages of

**Postcodebijadres.nl** website to collect Neighbourhood data. Below steps were taken to collect district, neighbourhood, neighbourhood latitude and neighbourhood longitude.

Step 1: I had to parse [www.postcodebijadres.nl](http://www.postcodebijadres.nl) using BeautifulSoup library to get JSON file of the Utrecht page.

Step 2: From the Utecht JSON file I had to collect the list of district available in the Utrecht area. After pursuing I found that there are a total of **11 districts** are available in the Utrecht area.

Step 3: After collecting the district information I crawled through 11 district pages to collect geolocation of each district and list of Neighbourhoods in each district.

Step 4: A total of 113 neighbourhoods are found in 11 districts of Utrecht. I have traversed through each neighbourhood page to collect neighbourhood latitude and longitude.

Below fields were collected by using the above method:

	<b>District</b>	<b>DLatitude</b>	<b>DLongitude</b>	<b>Neighborhood</b>	<b>Latitude</b>	<b>Longitude</b>
0	Wijk 00 Bunnik	52.061679	5.185625	Bunnik	52.068070	5.201628
1	Wijk 00 Bunnik	52.061679	5.185625	Vechten	52.064471	5.172355
2	Wijk 00 Bunnik	52.061679	5.185625	Verspreide huizen in het noorden	52.071684	5.171363
3	Wijk 00 Bunnik	52.061679	5.185625	Verspreide huizen in het zuiden	52.054542	5.183541
4	Wijk 01 West	52.105072	5.071596	Welgelegen I Den Hommel	52.082837	5.089841

Postcode data: The site **Postcodebijadres.nl** also had geolocation for each postcode which was very useful to find the geolocation of each rental places. So instead of using google geo API to collect latitude and longitude of each apartment, I have collected geolocation for each postcode situated in Utrecht, NL by using the similar method mentioned in the neighbourhood data collection. below are the main fields which were collected.

	<b>PostCode</b>	<b>Neighborhood</b>	<b>Street</b>	<b>SLatitude</b>	<b>SLongitude</b>
0	3981AA	Bunnik	Stationsweg	52.064300	5.205860
1	3981AB	Bunnik	Stationsweg	52.066078	5.202329
2	3981AC	Bunnik	Stationsweg	52.064279	5.206843
3	3981AD	Bunnik	Stationsweg	52.065852	5.204077
4	3981AE	Bunnik	Schoudermantel	52.062812	5.209969

## 2.3. Venue Data Collection:

To find the emerging neighbourhoods in Utrecht, I need to find the popular venue information in the Utrecht area. For this, I have used the FourSquare API to collect the required information. First, I have selected the required districts. Then based on each district I have used the neighbourhood geolocation to find the most popular venues in each neighbourhood. Below data fields were collected by using this API.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Vogelenbuurt	52.098712	5.11959	Griftpark	52.098044	5.121264	Park
1	Vogelenbuurt	52.098712	5.11959	Garagebedrijf Autoweerd	52.097323	5.120107	Auto Garage
2	Vogelenbuurt	52.098712	5.11959	Guusjes Eten & Drinken	52.100043	5.121283	Bistro
3	Vogelenbuurt	52.098712	5.11959	EKKO	52.097525	5.114554	Music Venue
4	Vogelenbuurt	52.098712	5.11959	t Zusje	52.096298	5.117788	Tapas Restaurant

## 3. Data Exploration and Clean-Up:

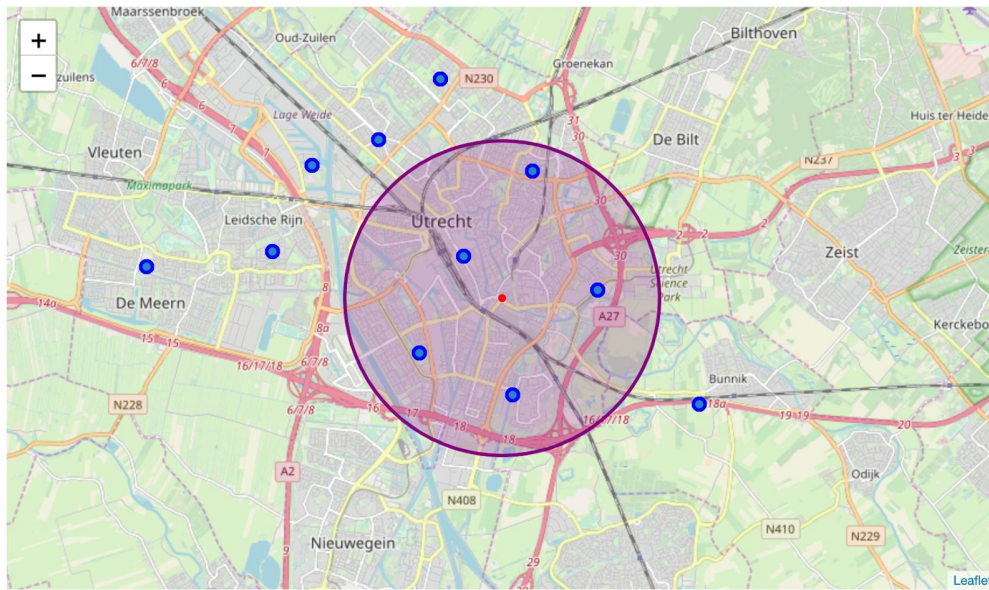
To come up with a good machine learning model, it is crucial to clean up the data and explore all the dimensions. Data clean up makes the dataset suitable for good analysis and outcome and on the other hand, data exploration gives us a good idea on the dataset so that we can prepare ideal model by understanding inside out of the information in hand. For each dataset, I had to use different methods to clear, select and explore. Full data handling process is describer below.

### 3.1 Neighbourhood data:

From the collected dataset on the Utrecht neighbourhoods, we can find that there are a total of 11 districts situated in Utecht. And these 11 districts have a total of 113 neighbourhoods.

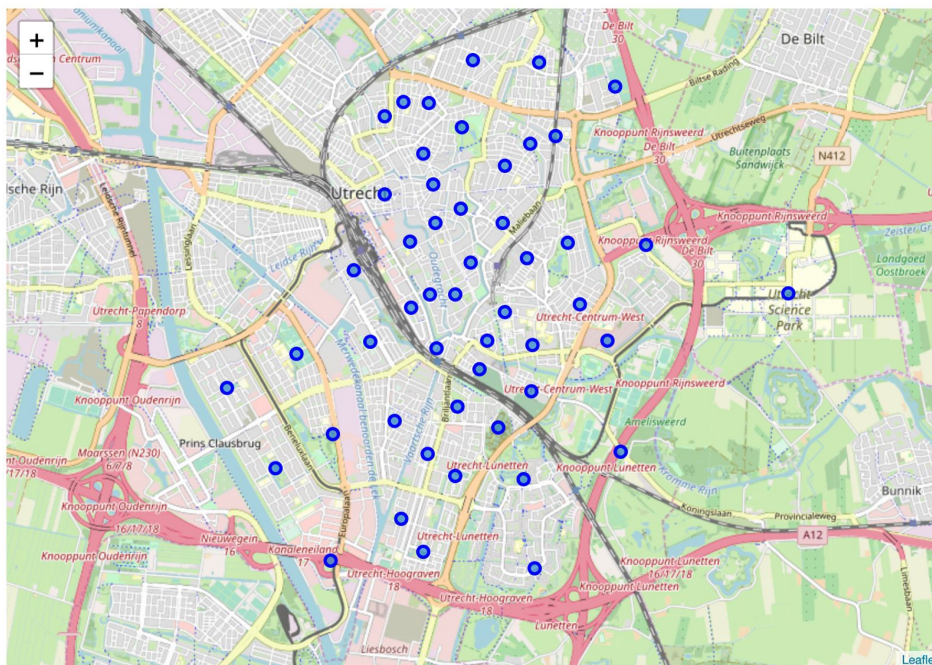
Utrecht is a comparatively large city in The Netherlands (approximately 100 square kilometres). Hence, it is essential to living close to the central station, so that commute to other towns becomes easy. Based on these, I have selected below Districts for my data exploration.

- Wijk 04 Noordoost
- Wijk 05 Oost
- Wijk 06 Binnenstad
- Wijk 07 Zuid
- Wijk 08 Zuidwest



Map 1: Selected districts are highlighted in the circle

Now if we select only the above districts from Utrecht, we can find a total of 50 neighbourhoods. Our analysis will be done based on these locations. we can find all the selected neighbourhoods in the below map.



Map 2: Neighbourhoods around the Utrecht Central

During the exploration stage, we also noticed that there are no null or duplicated values in the neighbourhood dataset. This is very important for our further analysis.

## 3.2. Venue Data:

After collecting neighbourhood data, the next step of data exploration is to find the important values in all these selected neighbourhoods. As mentioned earlier, the foursquare API was used to explore the top venue data. Below steps were taken to collected venue data from the API source.

To get the nearby values in each neighbourhood, a function `getNearbyVenues` was designed. The function receives the neighbourhood name and their respective latitudes and longitudes and returns the nearby venues in a list.

To retrieve the venue list, the geocoordinate of neighbourhoods were sent along with foursquare API credentials. After API call with these data, we received all the venue locations of near the neighbourhoods in JSON format. Later JESON file was processed to retrieve all related venue information.

A category function was also defined to retrieve each venue got from the API sources. A total of 1334 venues were found nearby all the neighbourhood locations. And there are a total of 179 unique venues in all these neighbourhoods. Below are some sample information of the venues along with their categories.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Vogelenbuurt	52.098712	5.11959	Griftpark	52.098044	5.121264	Park
1	Vogelenbuurt	52.098712	5.11959	Garagebedrijf Autoweerd	52.097323	5.120107	Auto Garage
2	Vogelenbuurt	52.098712	5.11959	Guusjes Eten & Drinken	52.100043	5.121283	Bistro
3	Vogelenbuurt	52.098712	5.11959	EKKO	52.097525	5.114554	Music Venue
4	Vogelenbuurt	52.098712	5.11959	t Zusje	52.096298	5.117788	Tapas Restaurant

Table 3: Nearby venue data for all the neighbourhood (only showed five from the list)

To find the top venues of each location, first, the 'one-hot encoding' was done on each venue categories. Then the frequency data-frame was made to identify the rate of venues in each neighbourhood. And then function 'return most common venues' was created to get the top 10 venues in each neighbourhood. Below table shows the first five rows of the top venues.



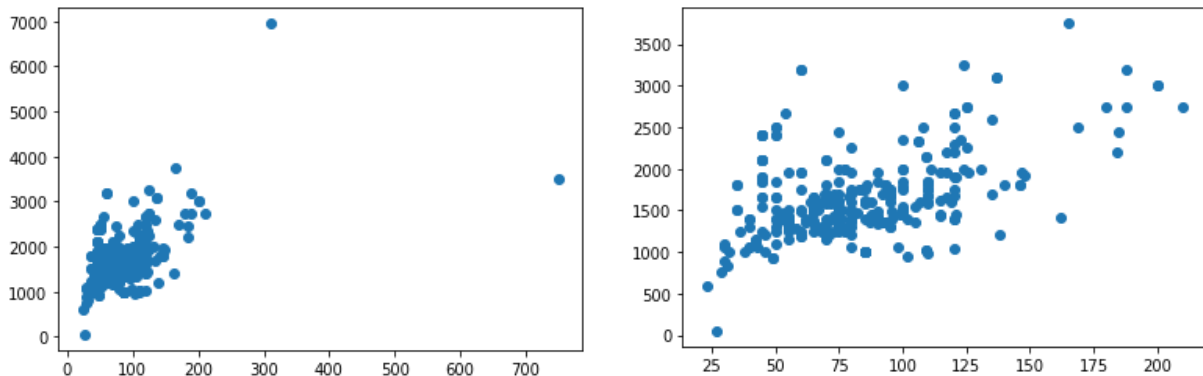
	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Absteede   Tolsteegsingel e.o.	Art Museum	Yoga Studio	Pharmacy	Fast Food Restaurant	Café	Garden	Supermarket	Planetarium	Theme Park Ride / Attraction	Science Museum
1	Bedrijfsgebied Kanaleneiland	Furniture / Home Store	Gym / Fitness Center	Bus Stop	Light Rail Station	Kids Store	Electronics Store	Breakfast Spot	Hotel	Lawyer	Martial Arts Dojo
2	Bleekstraat en omgeving	Bar	Coffee Shop	French Restaurant	Bakery	Park	Beer Bar	Café	Supermarket	Restaurant	Fish Market
3	Bokkenbuurt	Cosmetics Shop	Market	Metro Station	Fast Food Restaurant	Falafel Restaurant	Food	Flower Shop	Flea Market	Fish Market	Farmers Market
4	Breedstraat en Plompstorengracht en omgeving	Bar	Restaurant	Burger Joint	Coffee Shop	Sandwich Place	Italian Restaurant	Hotel	Café	Pizza Place	Hostel

### 3.3. Rental Data Exploration:

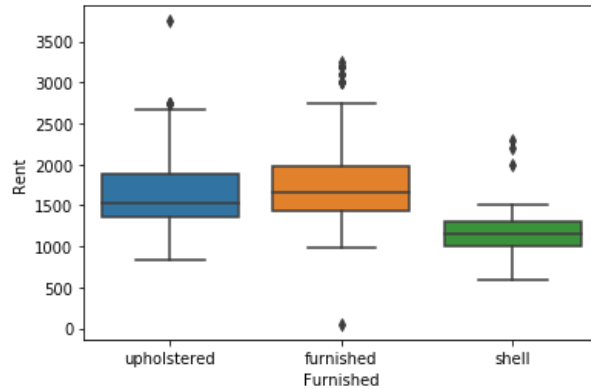
From the rental website, I have managed to scrap a total of 287 rental locations. After filtering the data based on our selected locations and removing duplicated and irrelevant data, we end up having 209 rental location in Utrecht. All the valid rental locations are plotted in the below map of Utrecht. Below steps were taken to prepare the dataset.

- Removing abnormal values for apartment size and the rent was removed from the dataset.

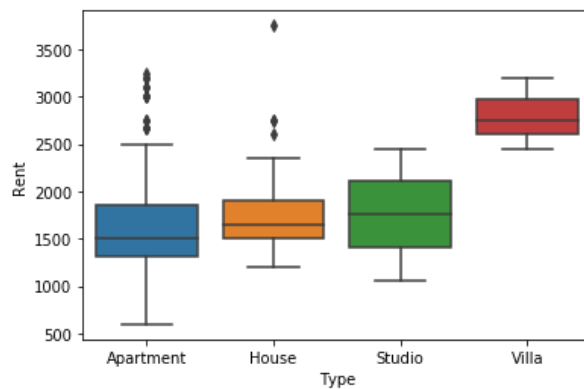
Below scatter plot demonstrate there were two anomalies in our data.



- In the dataset, there were a total of 19 unknown values for 'furnished' filed. By replacing unknown with the most common answer, we get below result, where it shows rent vs. apartment furnished or not.



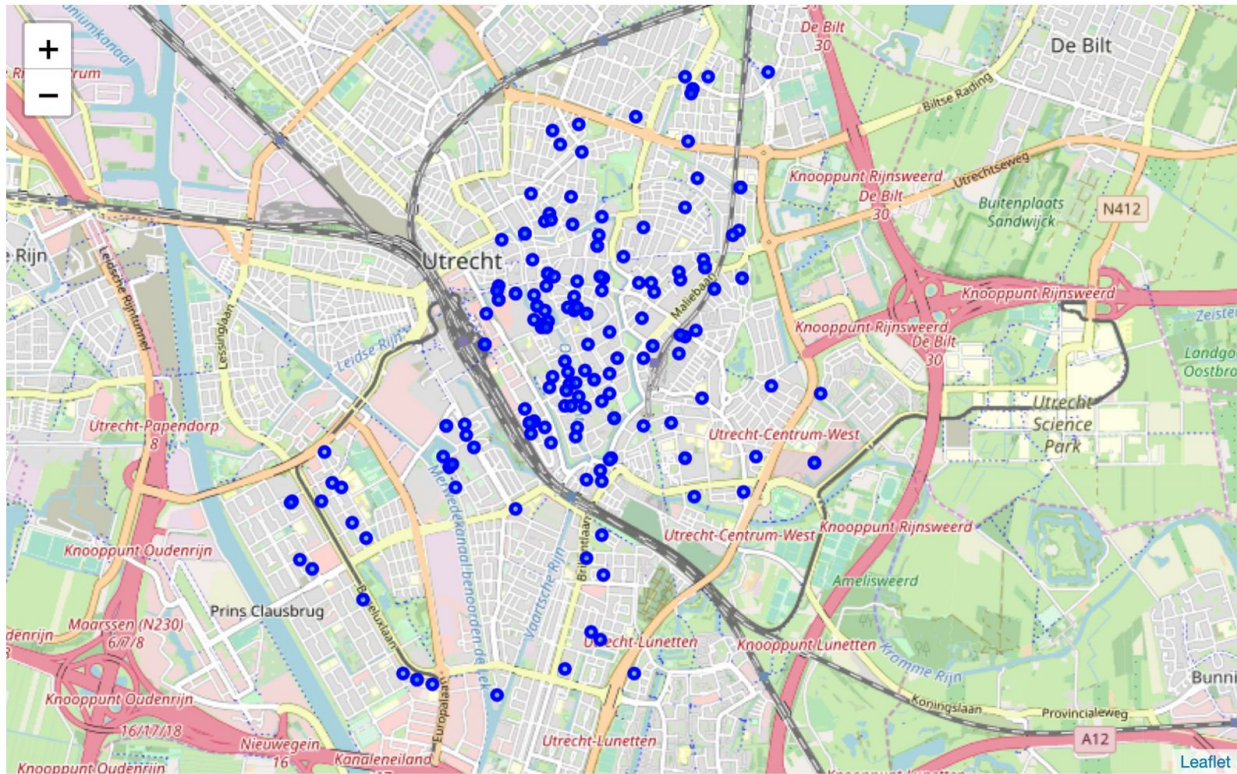
- For our analysis, we are not considering the rooms rents. After excluding the rooms the dataset looks like below.



- Last but not least, all the duplicate and null values were replaced to make the workable data. After preprocessing the rental dataset we get a total of 287 rental places in the Utrecht City. Now we can merge the rental dataset with Postcode dataset to get the geolocations of each rental locations. And by only selecting the rentals in your required districts we end up the data with 209 rows. Below are the first few rows of the rental dataset.

	Title	Postcode	District	Street	Type	Size	Rooms	Furnished	Neighborhoodr	SLatitude	SLongitude	Rent
0	Zakkendragershof	3511AE	Wijk 06 Binnenstad	Zakkendragershof	Apartment	90	2	furnished	Binnenstad	52.092719	5.115757	1495
1	Zakkendragershof	3511AE	Wijk 06 Binnenstad	Zakkendragershof	Apartment	73	3	furnished	Binnenstad	52.092719	5.115757	1695
2	Hekelsteeg	3511AK	Wijk 06 Binnenstad	Oudegracht aan de Werf	Apartment	148	2	upholstered	Binnenstad	52.091565	5.118934	1920
3	Oudegracht152-K	3511AZ	Wijk 06 Binnenstad	Oudegracht	Studio	75	1	furnished	Binnenstad	52.091868	5.118060	2450
4	Oudegracht	3511AZ	Wijk 06 Binnenstad	Oudegracht	Apartment	137	1	furnished	Binnenstad	52.091868	5.118060	3100

The rental locations are displayed in the Utrecht map below.



Available rental locations in Utrecht

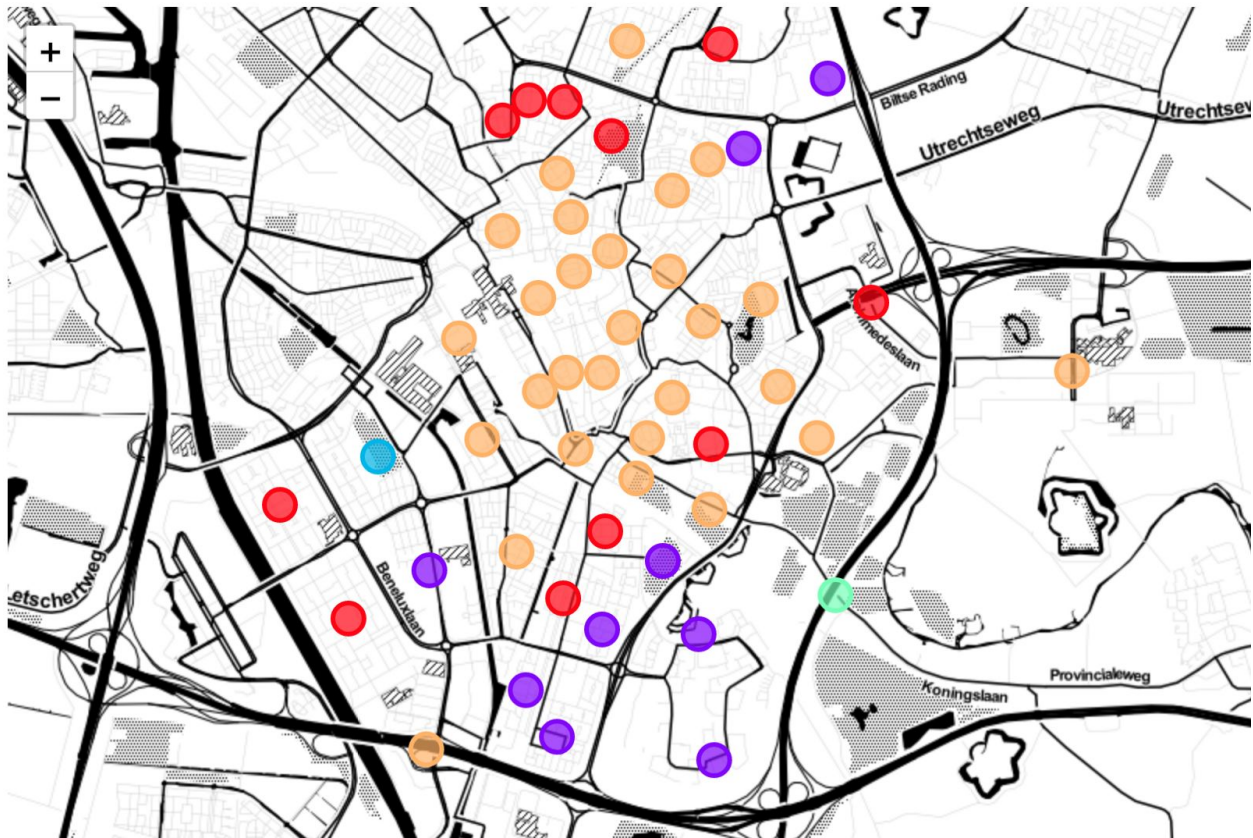
## 4. K-means Clustering:

k-means clustering is an unsupervised machine learning method which can be used to classify the unlabeled dataset. k-means clustering aims to partition the observations in the given number of clusters based on the features offered for each data.

Instead of looking for apartments in the big city without understanding the features of the neighbourhood is risky. For example, I can probably find a rental location based on price search from any rental websites. However, given that I don't know how the neighbourhood is can put me in a difficult situation, especially if I am new to the location.

To avoid such issue, I want to use k-means clustering method to identify each group of neighbourhoods based on the given features (the top venues) so that apartment hunter can understand in which area is suitable for based on his/her requirements.

After doing the k-means clustering method to our venue dataset, we find below clusters or in my term emerging neighbourhood based on the similarities. The clusters are shown in the figure.



Red: Cluster 0, Purple: Cluster 1, Blue: Cluster 2, Green: Cluster 3, Orange: Cluster 4

### Cluster 0 (RED):

This cluster has a total of 11 locations. The most common venues in this cluster are restaurants, supermarkets, Bus Stop, Farmers Markets. This cluster looks like a suitable location for living.

Restaurant	8.0
Supermarket	8.0
Bus Stop	6.0
Farmers Market	5.0
Fast Food Restaurant	4.0

Topmost common venues in the cluster 0



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
Coffee Shop	Restaurant	Bus Stop	Smoke Shop	Irish Pub	Pizza Place	Drugstore	Caribbean Restaurant	Café	Liquor Store
Restaurant	Supermarket	Irish Pub	Drugstore	Smoke Shop	Organic Grocery	Farmers Market	Middle Eastern Restaurant	Bus Stop	Liquor Store
Supermarket	Liquor Store	Coffee Shop	Farmers Market	Bistro	Restaurant	Middle Eastern Restaurant	Drugstore	Bus Stop	Pizza Place
Restaurant	Snack Place	Bar	Bus Stop	Skate Park	Cosmetics Shop	Farmers Market	Farm	Lake	Bistro
Zoo Exhibit	Shopping Mall	Playground	Supermarket	Garden	Electronics Store	Flea Market	Fish Market	Fast Food Restaurant	Farmers Market
Drugstore	Snack Place	Bookstore	Supermarket	Park	Ice Cream Shop	Restaurant	Turkish Restaurant	Grocery Store	Bakery
Coffee Shop	Park	Restaurant	Bus Stop	Office	Lawyer	Flea Market	Fish Market	Fast Food Restaurant	Farmers Market
Restaurant	Beer Bar	Cosmetics Shop	Wine Shop	Concert Hall	Mediterranean Restaurant	Fast Food Restaurant	Asian Restaurant	Supermarket	Farm
Bus Stop	Restaurant	Flower Shop	Fast Food Restaurant	Gym	Supermarket	Discount Store	Market	Farm	Playground
Park	Music Venue	Shopping Mall	Supermarket	Chinese Restaurant	Japanese Restaurant	Diner	Design Studio	Flea Market	Cosmetics Shop
Supermarket	Cosmetics Shop	Discount Store	Furniture / Home Store	Bagel Shop	Bakery	Department Store	Diner	Donut Shop	Food

## Cluster 1 (PURPLE):

Cluster 1 also has 9 locations in total. The topmost common venues in this cluster are Bust Station, Flea Market, Flower Shop, and Food places. This cluster has good commute facilities along with different types of shopping facilities.

Bus Stop	9.0
Flea Market	6.0
Flower Shop	5.0
Park	4.0
Food	4.0

Topmost common venues in the cluster 1

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
Soccer Field	Bus Stop	Athletics & Sports	Tennis Court	Farm	Food & Drink Shop	Food	Flower Shop	Flea Market	Fish Market
Bus Stop	Park	Soccer Field	French Restaurant	Hotel	Theater	Fast Food Restaurant	Restaurant	Food	Bus Station
Bus Stop	Park	Supermarket	Train Station	Electronics Store	Flower Shop	Flea Market	Fish Market	Fast Food Restaurant	Farmers Market
Bus Stop	Park	Playground	Farm	Snack Place	Design Studio	Deli / Bodega	Flower Shop	Cosmetics Shop	Flea Market
Bus Stop	Gym	Flea Market	Supermarket	Restaurant	Grocery Store	Drugstore	Harbor / Marina	Design Studio	Department Store
Cosmetics Shop	Bus Stop	Metro Station	Fast Food Restaurant	Convenience Store	Food	Flower Shop	Flea Market	Fish Market	Farmers Market
Bus Stop	Cosmetics Shop	Playground	Shopping Mall	Fast Food Restaurant	Supermarket	Chinese Restaurant	Flower Shop	Gym	Donut Shop
Bus Stop	Soccer Field	Intersection	Paper / Office Supplies Store	Restaurant	Flea Market	Park	Food	Pedestrian Plaza	Gym / Fitness Center
Bus Stop	Plaza	Business Service	Café	Farm	Bistro	Electronics Store	Pub	Skate Park	Japanese Restaurant

## Cluster 2 (BLUE):

This cluster has only one neighbourhood. And the main features of the cluster are fitness-centre, performing arts venue, casino and Electronic stores.

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
Gym / Fitness Center	Performing Arts Venue	Casino	Electronics Store	Food	Flower Shop	Flea Market	Fish Market	Fast Food Restaurant	Farmers Market

## Cluster 3 (GREEN):

This cluster also has just one neighbourhood and the most common venues in this cluster are jewellery stores, park, farm, soccer fields.

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
Jewelry Store	Park	Farm	Soccer Field	Zoo Exhibit	Food	Flower Shop	Flea Market	Fish Market	Fast Food Restaurant

## Cluster 4 (ORANGE):

Total 28 neighbourhoods are situated in this cluster. The main features of this cluster are bars, coffee shops, restaurants and cafes. The location is suitable for people who like to go out a lot.

Bar	16.
Coffee Shop	15.
Restaurant	15.
French Restaurant	14.
Café	13.

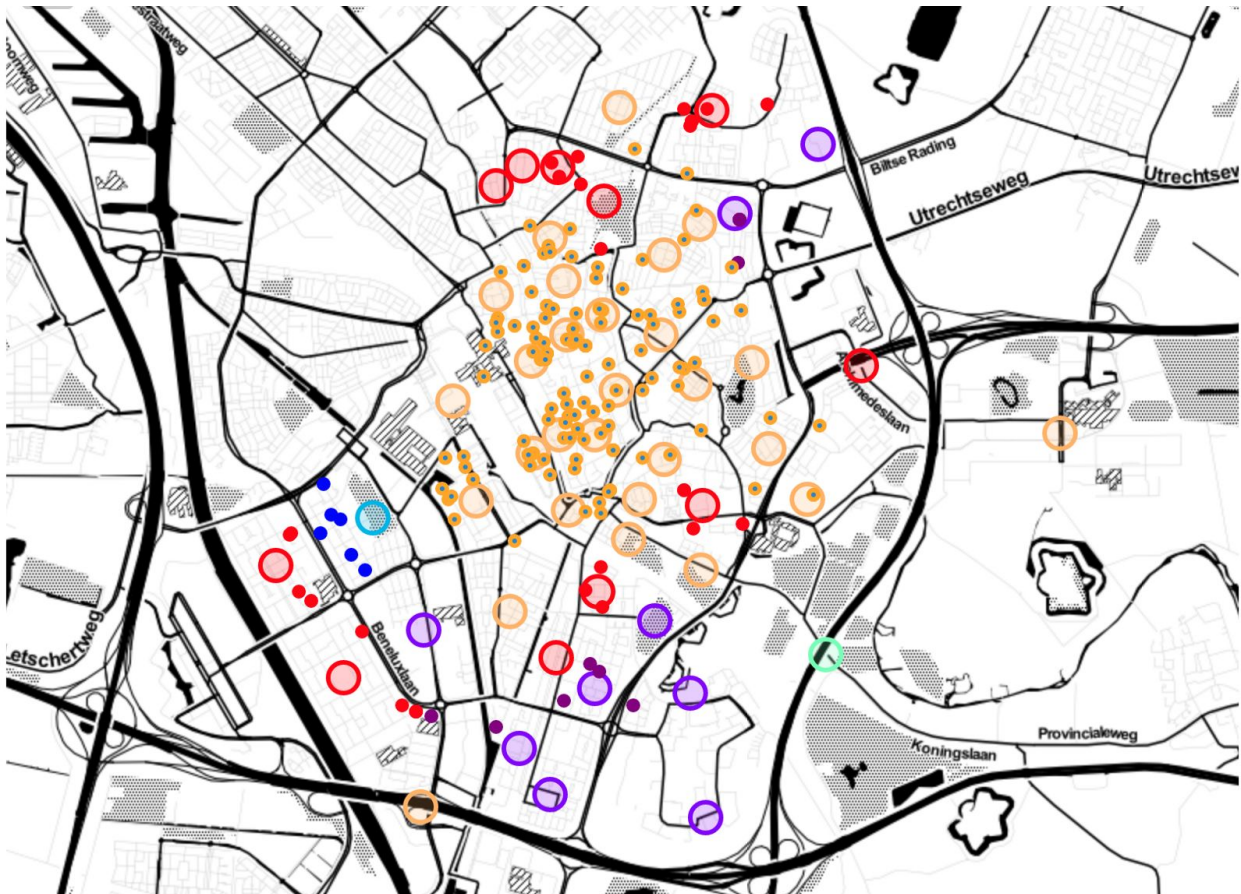
Topmost common venues in Cluster 4

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
Hotel	Vietnamese Restaurant	Tapas Restaurant	Bakery	Gay Bar	Ice Cream Shop	Irish Pub	Farmers Market	Cosmetics Shop	Bistro
Platform	Bakery	Bus Stop	Café	Train Station	Zoo Exhibit	Flower Shop	Flea Market	Fish Market	Fast Food Restaurant
Bus Stop	French Restaurant	Theater	Greek Restaurant	Hotel	Bar	Park	Fast Food Restaurant	Restaurant	Food
French Restaurant	Restaurant	Greek Restaurant	Theater	Burger Joint	Snack Place	Butcher	Laser Tag	Sandwich Place	Park
Coffee Shop	Restaurant	Deli / Bodega	French Restaurant	Pizza Place	Bus Stop	Indonesian Restaurant	Sandwich Place	Church	Butcher
Coffee Shop	Deli / Bodega	French Restaurant	Supermarket	Theme Park Ride / Attraction	Bus Stop	Breakfast Spot	Playground	Miscellaneous Shop	Pizza Place
Art Museum	Museum	Café	Planetarium	Theme Park Ride / Attraction	Train Station	Supermarket	Garden	Gym / Fitness Center	Yoga Studio
Café	Bar	Art Museum	Supermarket	Thai Restaurant	Fish Market	Japanese Restaurant	Falafel Restaurant	Beer Store	Gourmet Shop
Coffee Shop	French Restaurant	Train Station	Beer Bar	Japanese Restaurant	Bus Stop	BBQ Joint	Indie Movie Theater	Café	Vegetarian / Vegan Restaurant
Rental Car Location	Historic Site	Park	Furniture / Home Store	Garden	Garden Center	Eastern European Restaurant	Fish Market	Fast Food Restaurant	Farmers Market
Italian Restaurant	Ice Cream Shop	Gastropub	Sandwich Place	Restaurant	Breakfast Spot	French Restaurant	Supermarket	Bar	Bakery
French Restaurant	Bus Stop	Breakfast Spot	Lawyer	Coffee Shop	Bar	Bakery	Restaurant	Deli / Bodega	Garden

## 5. Selecting the apartments

### 5.1 Apartment Clusters

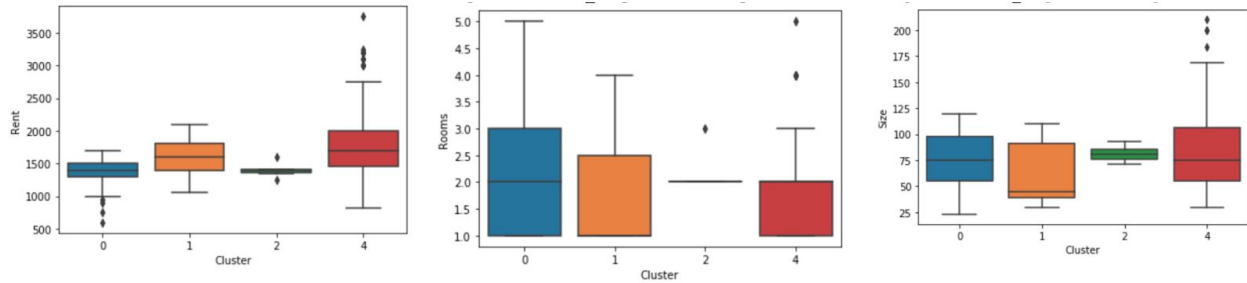
Now we have a clear idea about the neighbourhoods of Utrecht. From the clusters properties and our preferences, we can select our housing now. To do so first we need to assign cluster values to our rental dataset. We can do that by measuring the Euclidean distance method, the results are shown in the below map where each apartment are superimposed on top of the cluster map of Utrecht.



Map of available rentals based on the clusters

Now let's do some analysis to see the relation between apartment rent, size etc based on clusters in the below diagrams.





From the above analysis, we can understand that rent for cluster 4 is higher than the rest of the clusters, the number of rooms is higher in cluster 0 and floor place for cluster 1 is slightly higher than rest of the clusters.

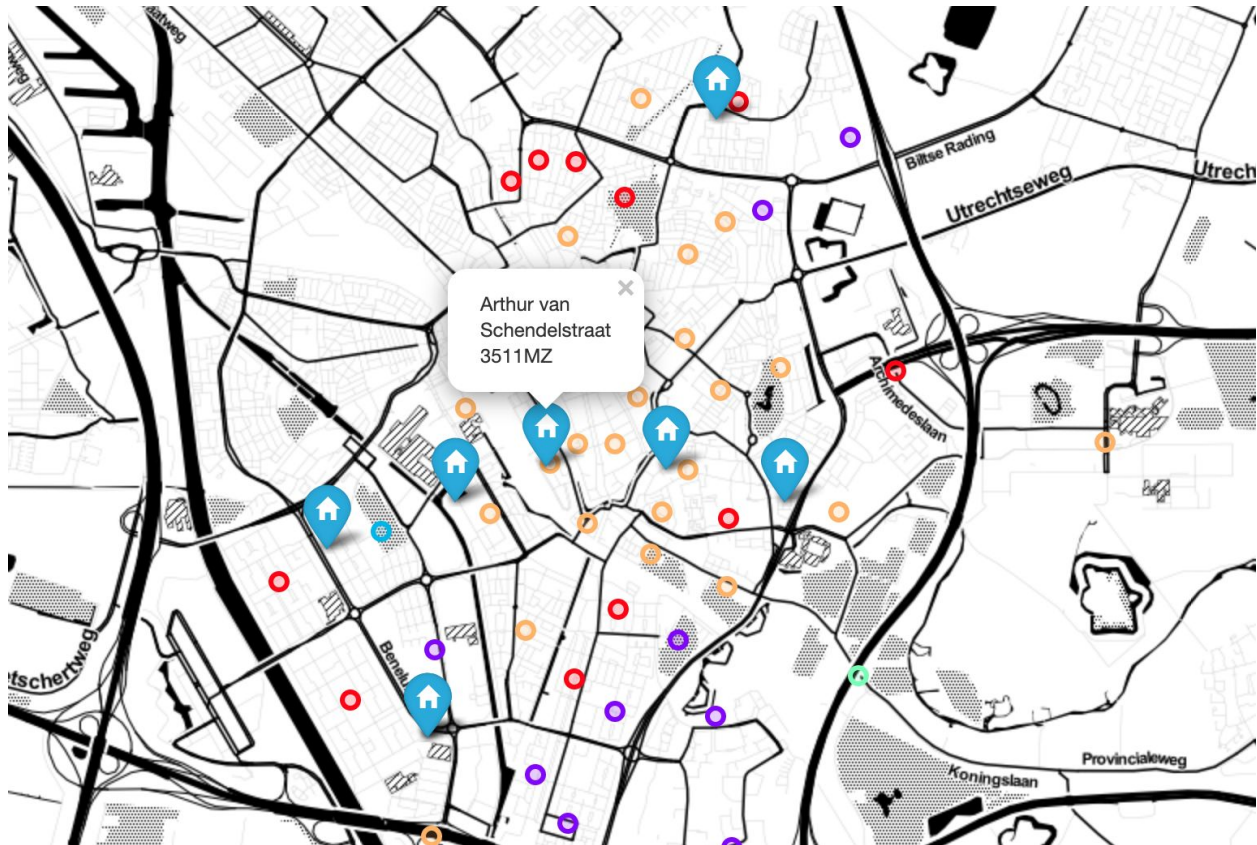
## 5.2 Finalizing Apartment

Now for me, the selecting criteria for an apartment I will be less than 1300 Euro per month, has more than one room and floor space of more than 60-meter squares. If I put the filter we get the below result.

	Title	Postcode	District	Street	Type	Size	Rooms	Furnished	Neighborhoodr	SLatitude	SLongitude	Rent	Distance	Cluster
35	Arthur van Schendelstraat	3511MZ	Wijk 06 Binnenstad	Carry van Bruggenstraat	Apartment	75	2	upholstered	Binnenstad	52.083785	5.117154	1295	0.0444714	4
118	Rolderdiephof	3521DA	Wijk 08 Zuidwest	Rolderdiephof	Apartment	80	2	furnished	Zuidwest	52.081562	5.107654	1215	0.275886	4
133	Aziëlaan	3526SJ	Wijk 08 Zuidwest	Aziëlaan	Apartment	73	2	upholstered	Zuidwest	52.066170	5.104754	1250	0.612787	0
134	Aziëlaan276	3526SJ	Wijk 08 Zuidwest	Aziëlaan	Apartment	102	3	shell	Zuidwest	52.066170	5.104754	950	0.612787	0
141	Van Vollenhovenlaan	3527JA	Wijk 08 Zuidwest	Van Vollenhovenlaan	Apartment	77	2	upholstered	Zuidwest	52.078419	5.094106	1250	0.40951	2

Apartment based on my requirement

I prefer to go out often along with my friends and family and I like to stay closer to the central hub. Hence, I prefer Cluster 4. So below Apartment is suitable for me.



By using k-means machine learning method, finally, I was able to under the neighbourhood of Utrecht city centre and did able to select a beautiful apartment in my preferred location. Below are the photos of the apartment I have selected.



## 6. Conclusion

The k-means cluster is a great method to explore and classify dataset similar to the location data I was using for this analysis. The source code in python, full report and datasets are shared in below locations. Anyone is welcome to use my codes to explore and select apartment in their preferred area.

**Python score code:**

[https://github.com/tamal2000/Capstone-Project/blob/master/Capstone\\_Project\\_Final.ipynb](https://github.com/tamal2000/Capstone-Project/blob/master/Capstone_Project_Final.ipynb)

**Datasets:** <https://www.kaggle.com/tamal2000/utrecht-geolocations>