IBM Applied Data Science Capstone :

‘The Battle of the Neighbourhoods’

Analysis using location data via Foursquare API

Mohammad Saman

**Table of contents:**

1. **Introduction**
2. **Data Section**
3. **Methodology**
4. **Results**
5. **Discussion**
6. **Conclusion**
7. **Introduction :**

As the commercial capital of the Netherlands and one of the top financial centres of Europe, *Amsterdam* is known for its international atmosphere. Amsterdam is also one of the world's most multicultural cities, with at least 177 nationalities. It is home to close to half a million expatriates, with the majority of them from other parts of Europe. Due to its vibrant and diverse nightlife among other things, Amsterdam hosts close to 16.5 million visitors every year from all around the world.

Now, having stayed in Amsterdam for about a year and a half, I found quite a lot of different ethnic restaurants. Also adding to the fact that Amsterdam is a major multicultural city. As an Indian with a fondness for Italian food, I have explored the streets of Amsterdam looking for Italian restaurants and there are only a few to choose from. The restaurants can range from being luxurious and expensive to being ordinary and affordable.

The purpose of this project is to use Foursquare location data and regional clustering of venue information to determine what might be the best neighbourhood to open a restaurant. When looking to open a business of your own, there are many variables to consider such as Location, competition, Rental costs etc. As part of this project, we are going to do an analysis of the neighbourhoods of Amsterdam for the consideration of opening an Italian restaurant. This project would be of interest to people who are looking to find suitable neighbourhoods closer to high footfall areas with reasonable rent options to open a restaurant.

1. **Data Section :**

For this project we need data about :

* Neighbourhood boundaries of Amsterdam
* Venues/count of venues in different neighbourhoods
* Commercial Rented space for restaurants

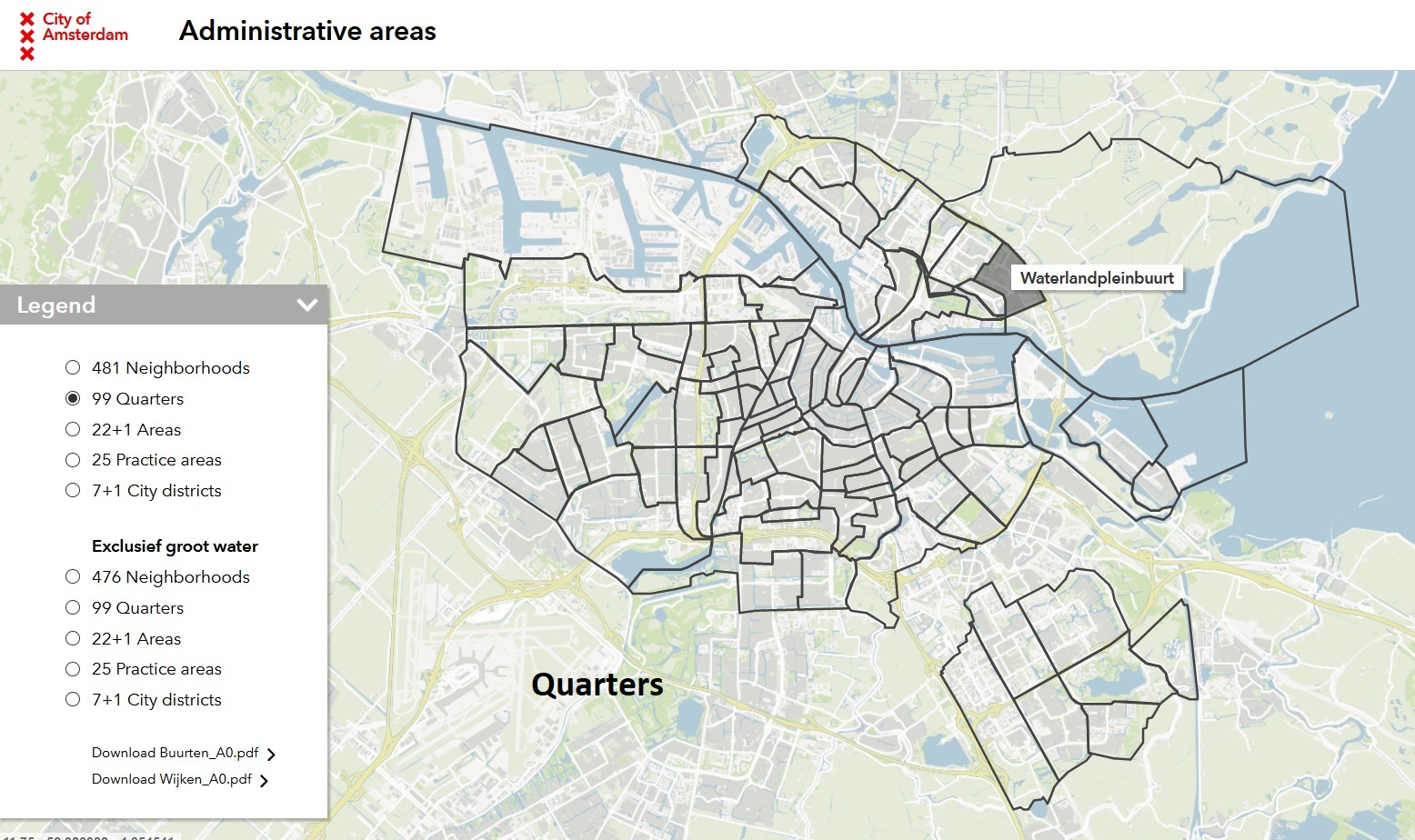
The data that we are going to use as part of this analysis has been taken from multiple locations. For instance, the data concerning the different neighbourhoods of Amsterdam was taken from the ‘clair city data portal’.

<https://claircitydata.cbs.nl/organization/cbs>

This dataset consisted of the region codes, type of region (quarter/neighbourhood), number of inhabitants, Latitude, Longitude etc. among other things. Since some of the columns and the data was in Dutch, I had to translate and research a bit about the neighbourhoods. I found a website of the Dutch government which gave me a pretty good understanding of the distribution of the landscape of Amsterdam.

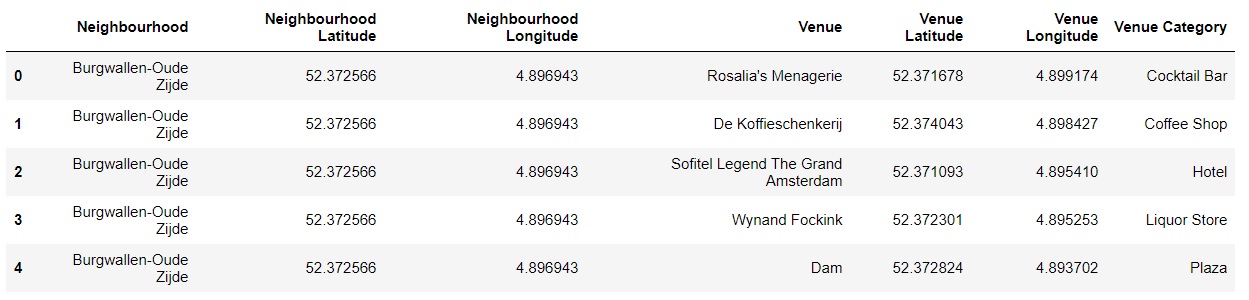
<https://maps.amsterdam.nl/gebiedsindeling/?LANG=en>

Upon examination of the dataset and with the help of the above site, I found that *Amsterdam* has 99 quarters and 481 neighbourhoods in total. Basically a quarter (‘Wijk) is an area consisting of 2 or more neighbourhoods(‘Buurten’). So using this information I have extracted the data by filtering on quarters (‘Wijk’) as we would have a few neighbourhoods(‘Buurten’) to choose from at the end of our analysis.

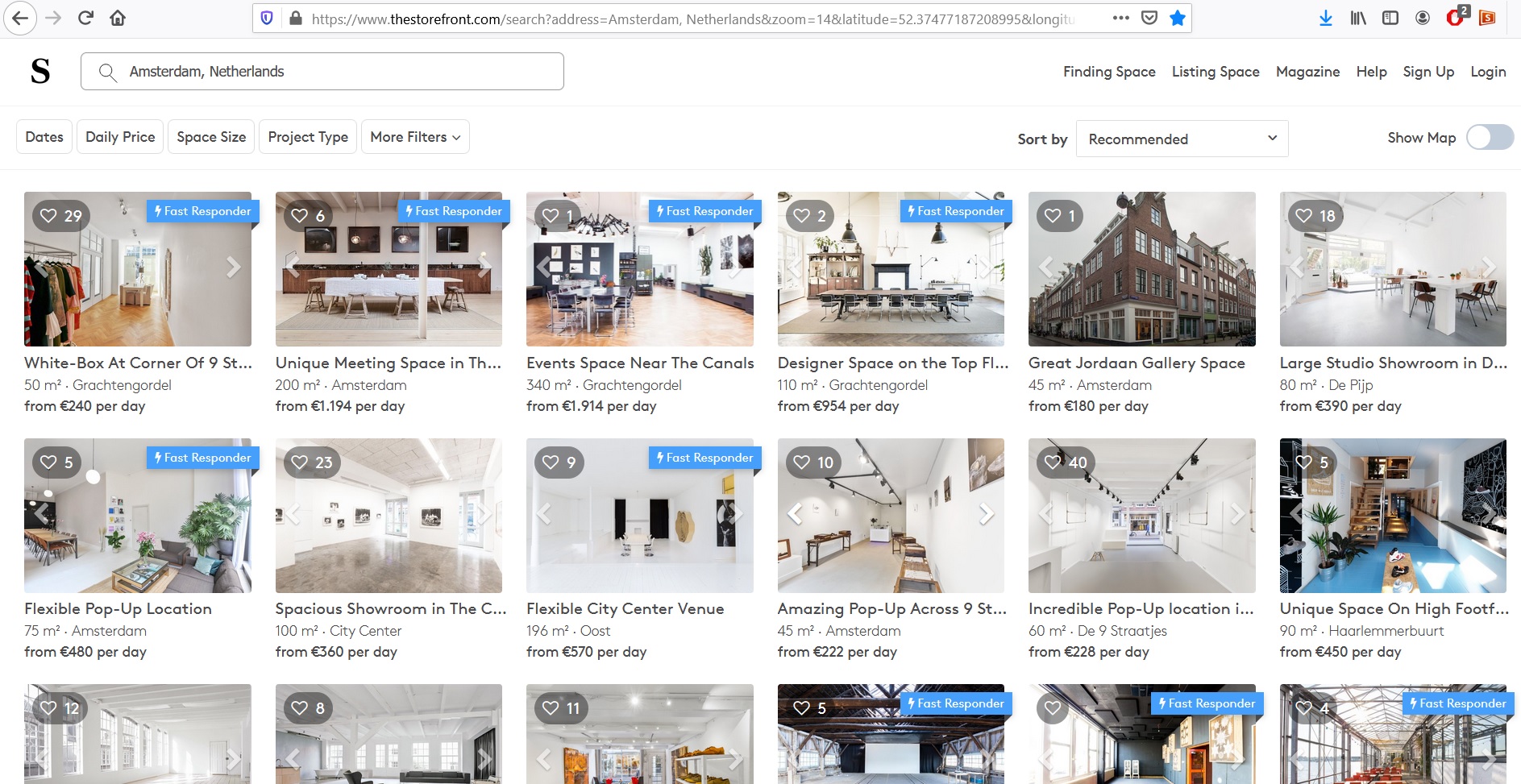




For the information regarding the venues with respect to the locations, we used the Foursquare API. We were able to get a list of venues for each neighbourhood corresponding to their latitude and longitude we extracted from the neighbourhoods dataset. For example,



Apart from the above, we also needed data for commercial rented space for restaurants. Obtaining data for this part was a little bit tricky. However, I found a website which lets its users browse thousands of pop-up retail spaces, showrooms and event venues of a certain location.



**Link** : <https://www.thestorefront.com/>

I web scraped the data from the above site for all retail spaces in Amsterdam by adding a filter for ‘Restaurants’. I extracted the location data (Latitude & Longitude) and the rental costs from the webpage into a pandas dataframe. Some of the spaces listed on the site had a ‘Price on request’ tag for which it was not possible to extract the rental cost data. For the purpose of our analysis all such listings had been removed from our dataset.

