Determining optimum location of an Argentinian Restaurant in Amsterdam

Casali Nicolas

July 3rd, 2019

Introduction

As an **Argentinian**, I was surprised when I first arrived to Amsterdam. It was neither the red-light district nor the great amount of coffee shops out there. The thing that caught my attention was the number or **Argentinian restaurants** in there.

It is interesting to look for the ways this restaurants are located to see if there is an optimum "stereotype" of neighbourhood where this kind of venues can maximize the success likelihood.

Motivation

There are a number of reasons to support this analysis, namely:

- Opening a restaurant abroad does not require optimum cooking skills. People from other countries usually are not as critic with foreign food as they are with they own (If you don't believe me, ask a Japanese what do they think about sushi restaurants outside japan). This makes the endeavour a little bit easier in terms of the skilled workers needed.
- Argentina is going through a severe economic crisis, which means that more people are considering going to live abroad and, as a matter of fact, lots of them are sons and daughters of European immigrants (which means they have the legal right to move to European countries and live there).
- Global trends suggest people spend a considerable amount of their income in social events like eating outside. Furthermore, they tend to look for new experiences, like foreign food, fusion cuisine, new flavours, etc. Have you notice more food venues in your neighbourhood recently?
- New trade deal between EU and some European countries (including Argentina) could potentially decrease price of raw materials (like Argentinian meat) which in turn could make the situation even more attractive.

Data description

Data sources

I used **demographic data of Amsterdam**, which I found in : claircitydata.cbs.nl. The dataset contains a lot of information like number of immigrants, population density, married habitants, etc. The initial DataFrame was composed by 579 rows and 37 columns.

Additionally, I used **Forsquare API** to get the most common venues of given neighborhoods of Amsterdam.

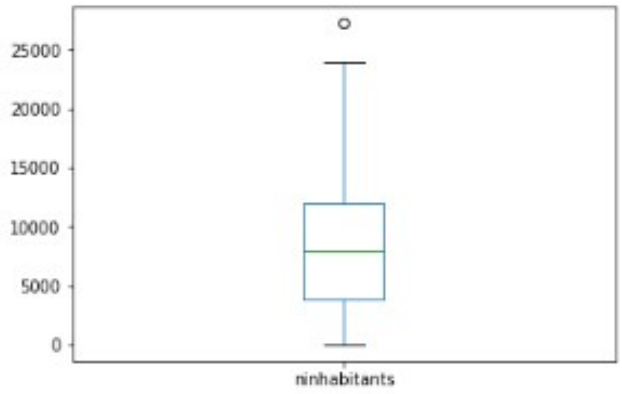
Data cleaning

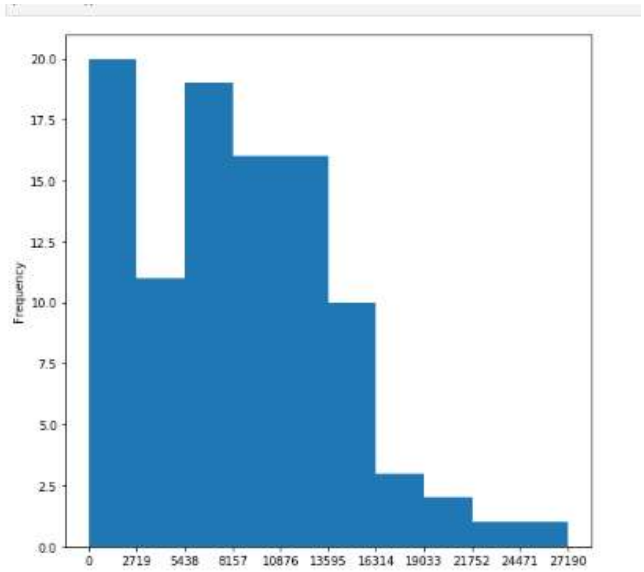
First of all, I filtered the DataFrame to only have data of **neighborhoods,** which led me to only 99 rows.

```
[31]: amsterdam = amsterdam[amsterdam['regio_type']=='Wijk']
      amsterdam.shape

[31]: (99, 37)
```

After that, I took a look at the summary statistics provided by the **describe()** method. Box plots and bar charts helped me to identify a couple of neighborhoods with zero habitants, which I consequently removed from the analysis.

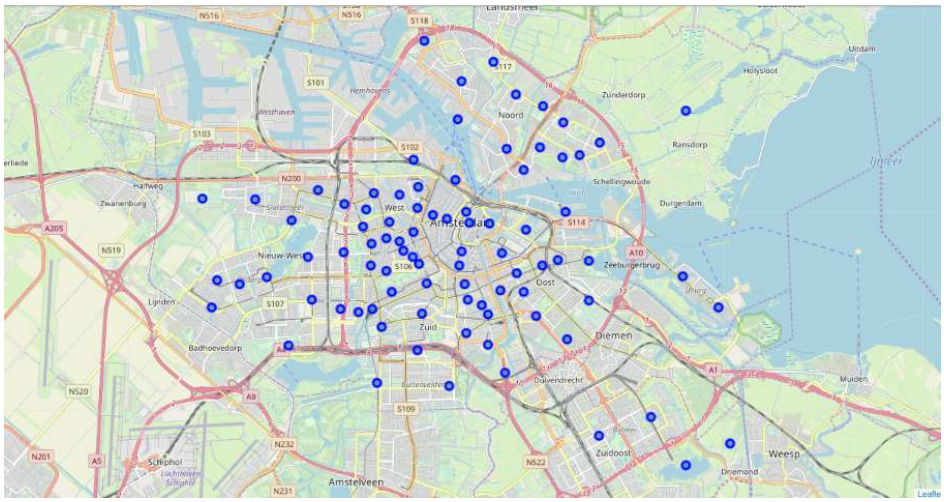




As we can see, there is an outlier. However, I did not remove it since I think it will add value to the analysis, since we can get some insights about why so many people live there.

Finally, I dropped the columns that where not going to be used for the analysis. The final output was a 89x11 DataFrame.

I used python **folium library** to visualize geographic details of Amsterdam and its neighborhoods and I created a map of Amsterdam with neighborhoods superimposed on top. I used latitude and longitude values to get the visual as below:

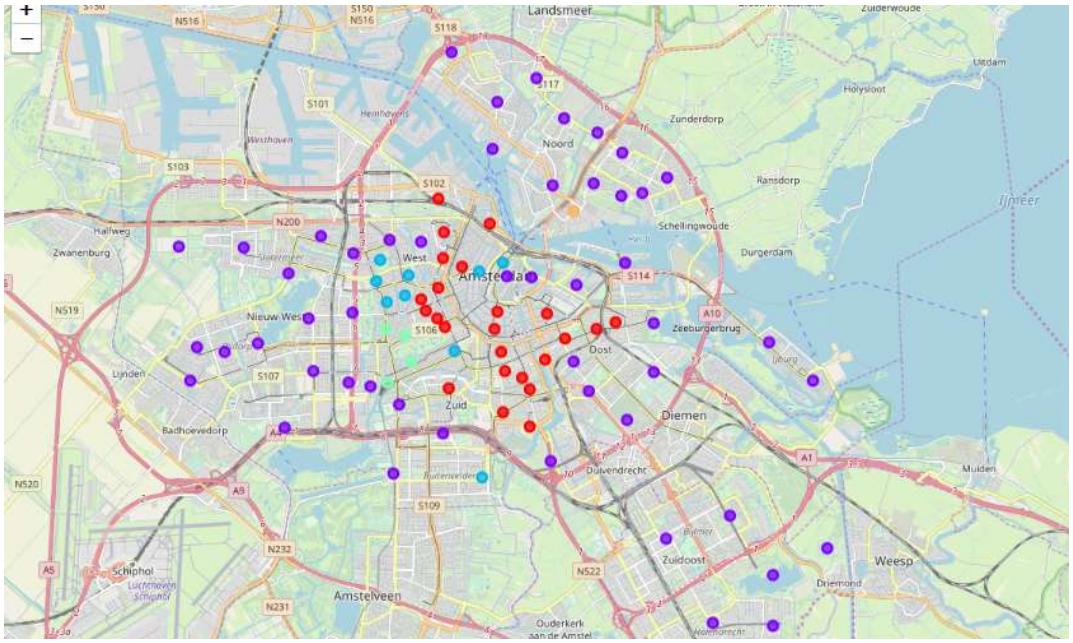


After that, I used Latitude and Longitude values of the neighborhoods to make exploratory requests via the Forsquare API, which I merged in a single DataFrame of 4700 rows and 7 columns. There were **246 different venue categories** in the aforementioned DataFrame.

Lastly, I merged the two DataFrames to get all the information in a master table, which would be used for the cluster analysis. Needless to say that I standardized the values before the analysis in order to avoid unwanted biases towards lower values.

Clustering

Elbow method suggested 5 as the optimum number of clusters, which I used to create the clustered map of Amsterdam neighborhoods displayed below:



Discussion

As I mentioned before, Amsterdam is a big city with a high population density in a narrow area. The kind of venues, as well as the demographic characteristics of the 89 neighborhoods under de analysis can be significant different. In such a complexity, very different approaches can be tried in

clustering and classification studies. Moreover, it is obvious that not every classification method can yield the same high-quality results for this city.

I used the Kmeans algorithm as part of this clustering study. When I tested the Elbow method, I set the optimum k value to 5. (with 89 neighborhood used). For more detailed and accurate guidance, the data set can be expanded and the details of the neighborhood or street can also be drilled.

I ended the study by visualizing the data and clustering information on the Amsterdam map. In future studies, web or telephone applications can be carried out to direct investors.

Conclusion

As a result, more Argentinians are turning to the old continent to start a new endeavour. For this reason, I think people can make informed decisions based on actual facts when considering moving to a different city. This study focused on Argentinian restaurants in the city of Amsterdam, but similar analysis can be made for different cities and venues.

To the future,

Casali Nicolas