# Clustering Dublin neighborhoods

Marcos Gabriel Braz de Lima

March 13, 2020

## Introduction

Many countries around the world receive much immigrants all years. Ireland is one of them that has open much tecnologies companies few years ago and growth the demand of IT professionals. After that, Ireland has received much immigrants to work.

Therefor of job opportunities, Ireland have much english schools. Much other peoples go to Ireland for study english. Countries like Brazil, has Ireland as a most comun destiny to english exchange. The Irish are very welcoming and the study exchange can work during period of study, and pay your spending. Because of that, Ireland are chosen for much people to live or to have a experience as a exchange.

Dublin is the capital of Ireland and is the city that receive more immigrants. This city has about 500.000 of people. Has famous regions with many pubs and bar to enjoy the night. Other neighborhoods are more residential and have less trendy bars.

For new foreign residents, whatever for work or exchange, this study aims to group similar neighborhoods to map regions for different interests. This study will help people to find residential areas to live more peaceful or neighborhoods with more bars for those who like to enjoy the night. For people who want to open an establishment, this study can help to choose a location with few competitors. to help exchange students, this work can help them recommend neighborhoods with a higher number of venues to be able to look for a job. So, there are many situation that this study can help.

## Data acquisition and cleaning

The necessary data was a list of neighborhoods of Dublin with coordinate and the venues of each neighborhood. Using BeautifulSoup library, the list of neighborhood was obtained from wikipedia website. The neighborhoods are assumed as areas of the Northside and Southside listed in these wikipedia links below:

https://en.wikipedia.org/wiki/Southside,_Dublin

https://en.wikipedia.org/wiki/Northside,_Dublin

After that, setting the address as "neighborhood, Dublin", we use the Nominatim function from geopy library to find the latitude and longitude data.

The next step was map all venues in each neighborhood. For that, was used the Foursquare API to explore nightlife venues. I set a radius of 1500 meters from each neighborhood. Was used the venue categorie "4d4b7105d754a06376d81259" to filter just nightlife venues, like bars, pub, lounges and similars.

Some locations have been mapped by more than one neighborhood. Some locations appeared in up to nine different neighborhoods. In this case, the venues were assigned to the neighborhood with the shortest distance.

## Exploratory data analysis

In total 615 venues were found distributed in 116 neighborhoods (figure 1).
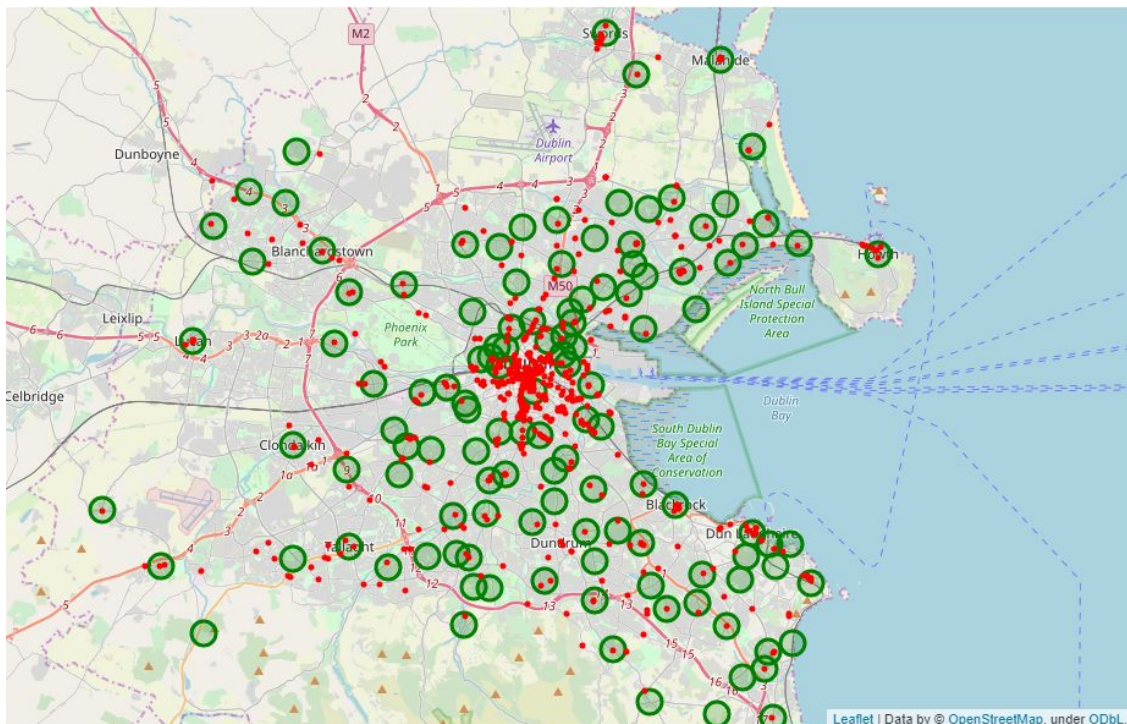


**figure 1. Map with distribution of neighborhoods (green circles) and venues (red point)**

Dublin city is commonly divided into northern and southern parts. Figure 2 shows that the south side has more venues (336) than the north side (284).
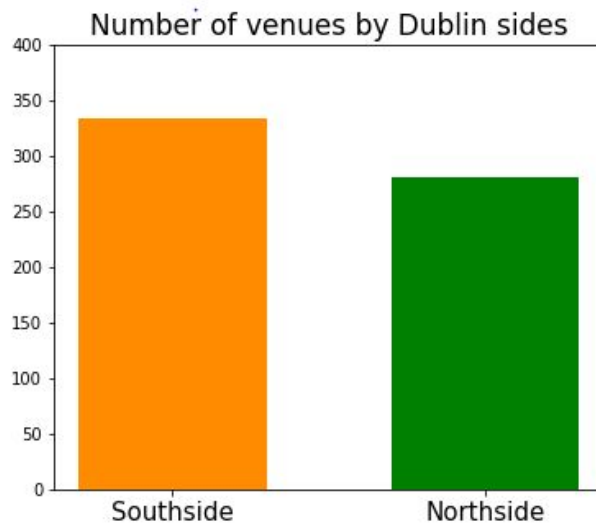
**figure 2. Number of venues by Dublin side**

Comparing the number of venues by neighborhood, it is clear that some neighborhoods are much more concentrated than others. Citywest, Smithfield, Harold's Cross, Rathmines regions have more than 20 venues (figure 3).
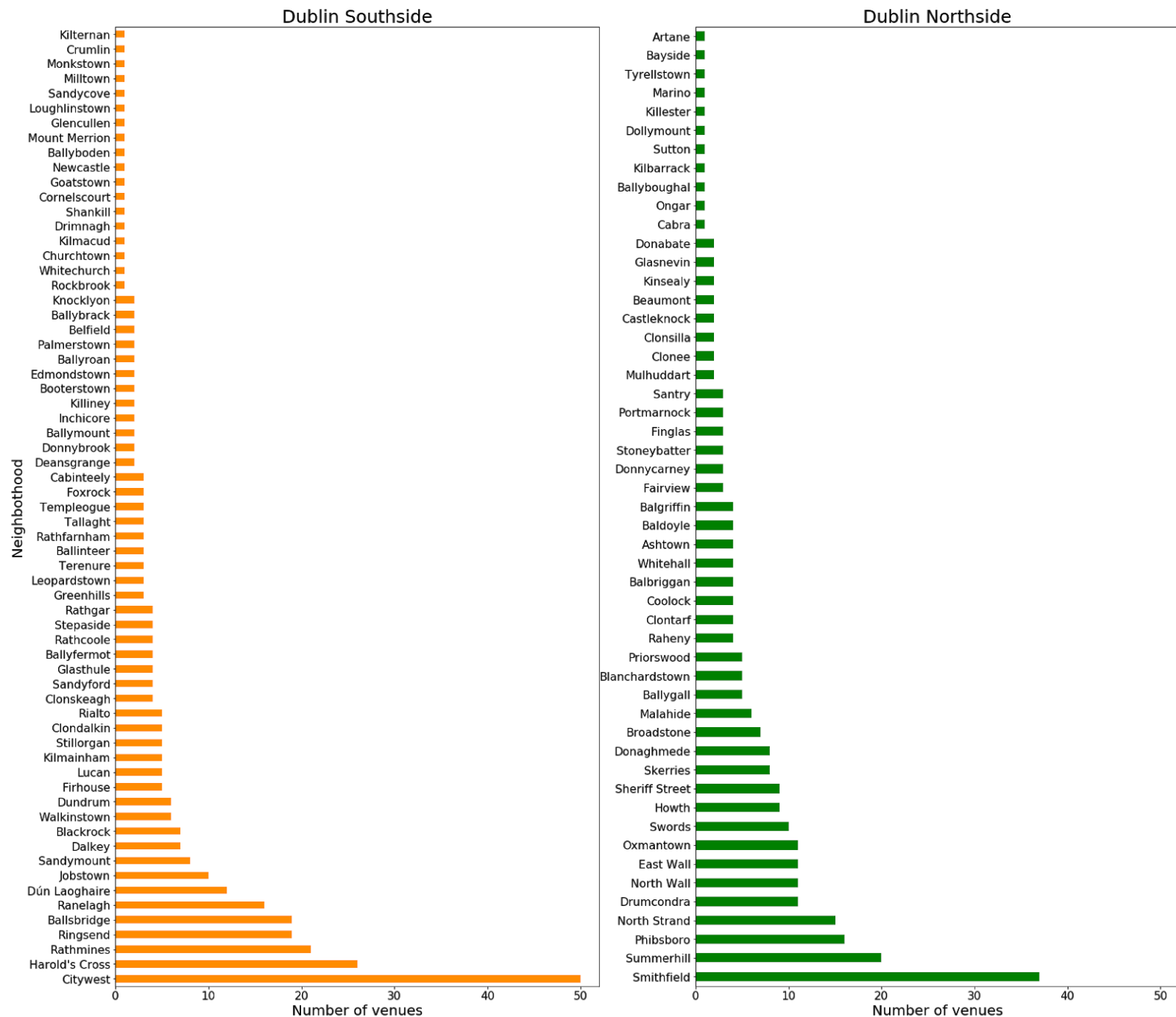
## Number of venues by neighborhood



**figure 3. Number of venues by neighborhood**

Analyzing the number of places by categories, the first three categories with the most occurrences are pub with 308, second is bar with 133, followed by gastropub with 28 occurrences. Pub and Bar are two categories with much higher occurrences compared to the others (figure 4).
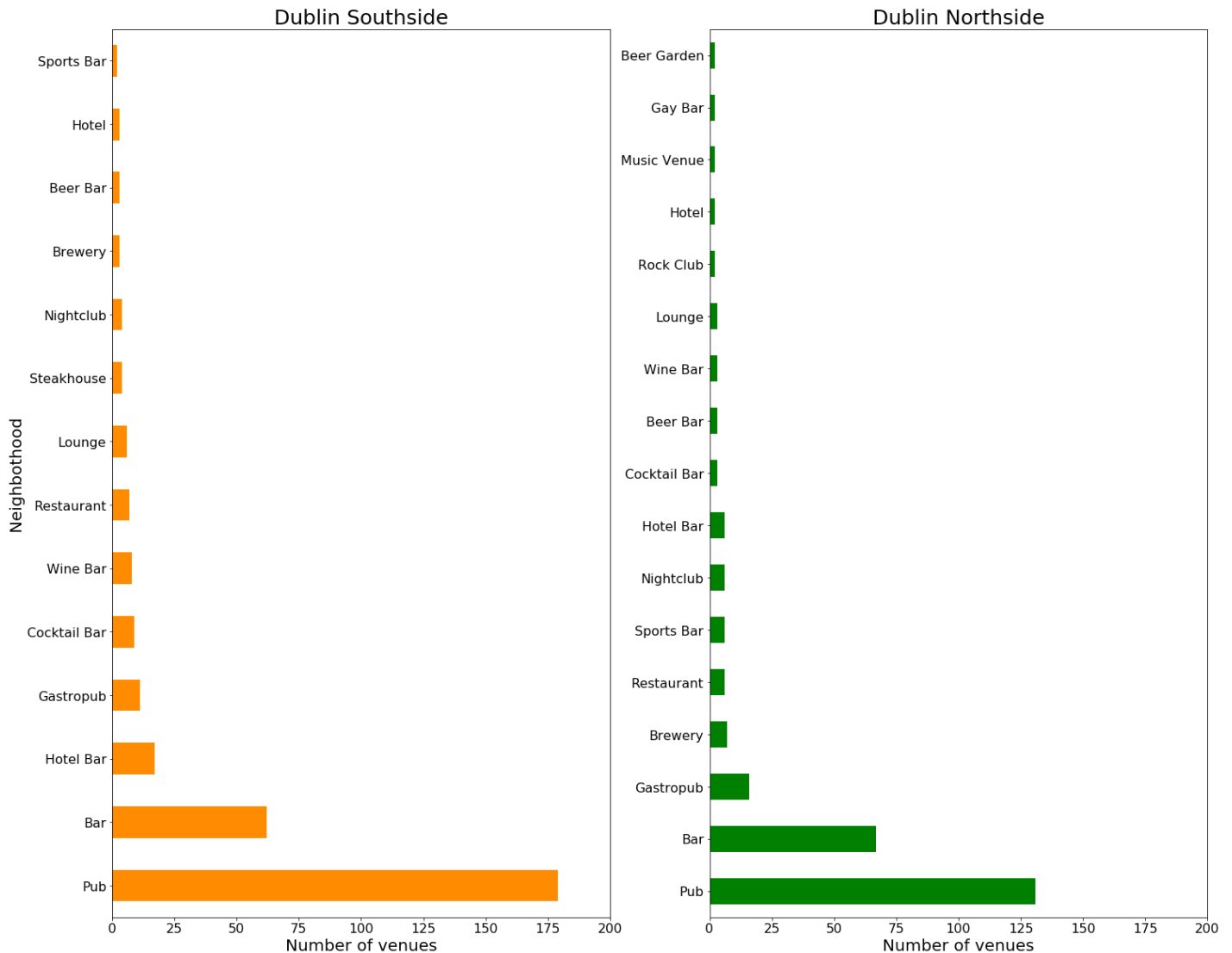
**figure 4. Number of venues by category.**

# Clustering model

Clustering are unsupervised models used to create groups within your database. So, I used the K-Mean clustering algorithm to group similar neighborhoods. To apply the model, I first transformed and standardized the entire database. Then, I used the elbow method to find the best k number, with k being the cluster number to be used. Figure 5 shows that the distortion after 3 decreases slowly. This decrease in distortion shows us the ideal value of k. Therefore, the number of k used was 3.
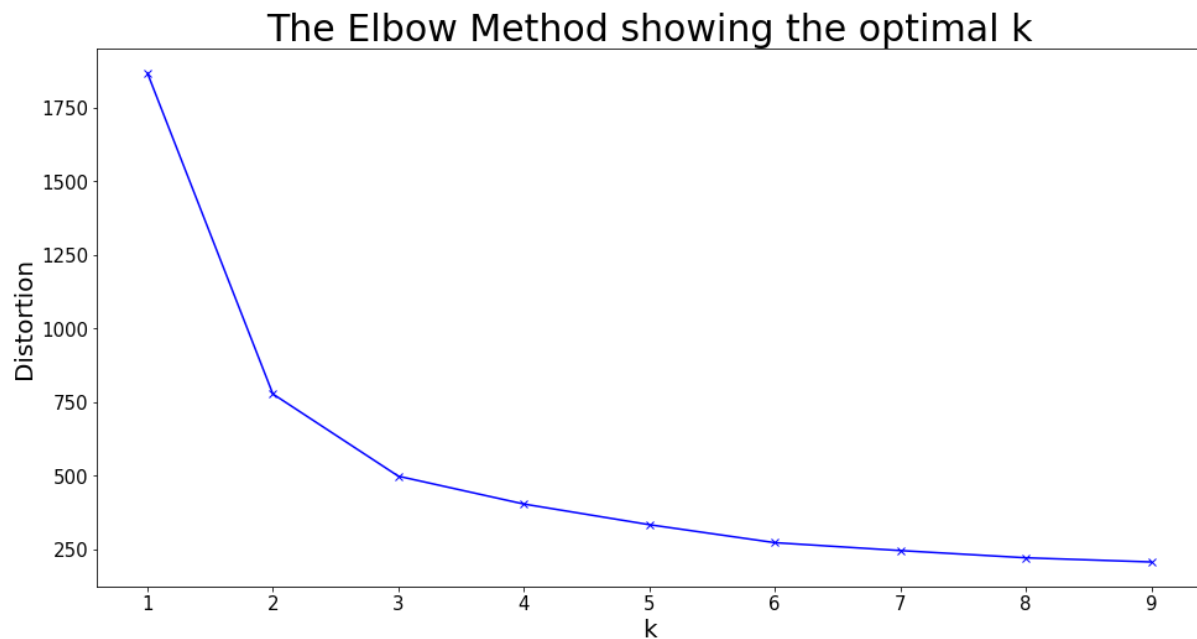
**figure 5. Elbow test**

## Clusters

The three clusters can be seen in figure 6 below. The black colored cluster concentrates neighborhoods in the center of Dublin. The other orange and green clusters, on the other hand, have the neighborhoods scattered throughout the city.
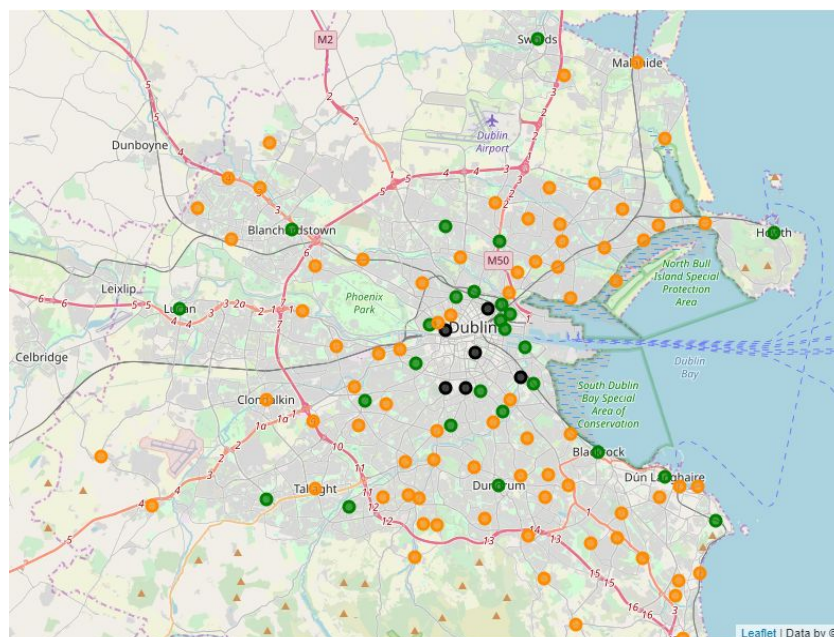


**figure 6. Map with neighborhoods by clusters.**

To better understand the clusters, figure 7 shows the average quantity per neighborhood of the 5 most common categories in each cluster.
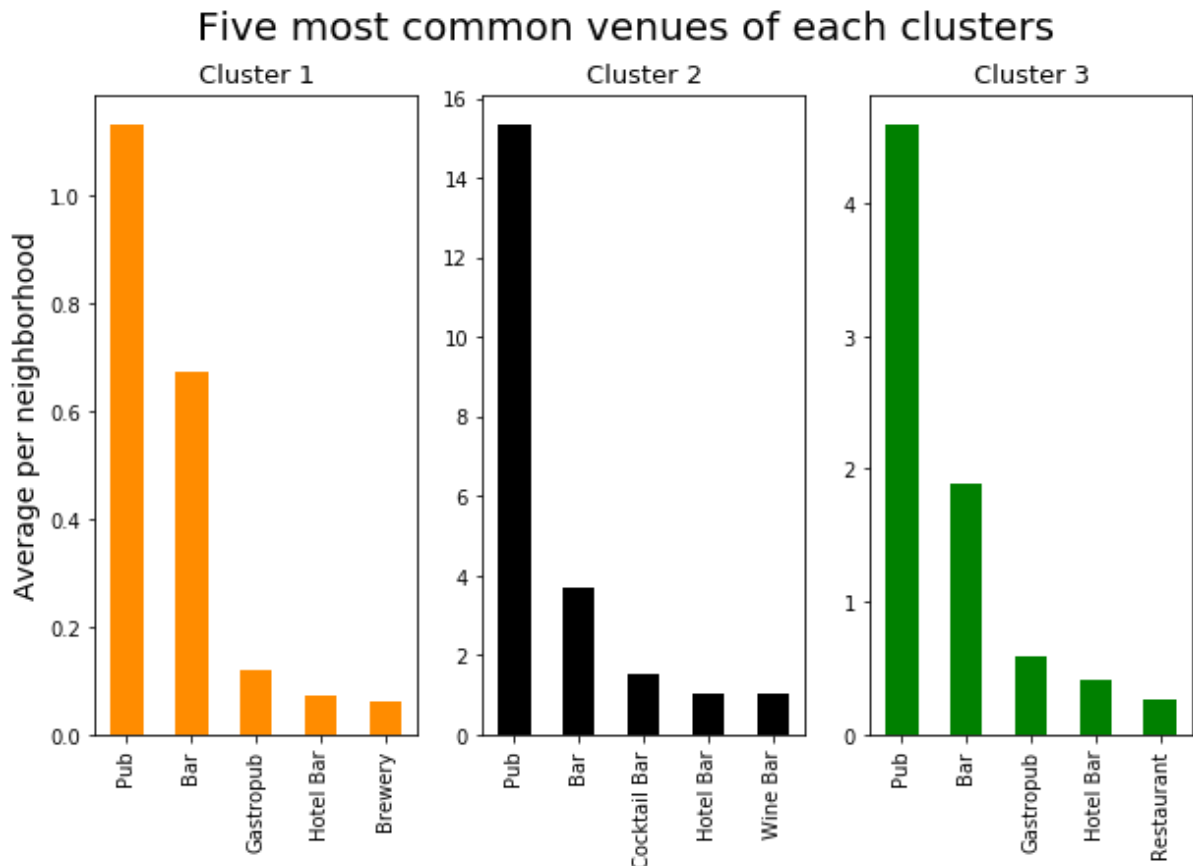


**figura 7. Average of number of venues per neighborhood of five most common venues.**

The central region of Dublin concentrates high commercial activity. Temple Bar is the most famous central region that concentrates traditional bars. The clustering model was able to identify these central neighborhoods with more nightlife. This central region is represented by the black cluster. Comparing the number of bars per neighborhood, the black cluster has about 15 bars per neighborhood, the green cluster has 4.5 and the orange cluster has 1.1.

The other two groups, orange and green, are scattered throughout the city. The orange cluster has a lower rate of venues per neighborhood. The first most common site has about 1.1 per neighborhood and the second about 0.7. On the other hand, the two most common locations in the green cluster are about 4.5 and 1.9 respectively.

From these clusters, we can make different recommendations. Given the large number of immigrants in Dublin, whether for work or for english study exchange, we can recommend neighborhoods to live in for example. For those who like quieter and more isolated places, the Orange Cluster neighborhoods are ideal. Who prefer to live where there are several places to enjoy the night, the central neighborhoods of the

black Cluster will be the best option. For those who like tranquility so much and also like to enjoy the night, the neighborhoods of Cluster Verde will be the best to recommend.

In a different situation, thinking about the best places to undertake, the neighborhoods of orange cluster are more interesting to open a hotel bar, for example. The lower rate of hotel bars per neighborhood indicates less competition.

In a specific context, exchange students can reconcile their homes near workplaces. For this, you can recommend looking for quieter neighborhoods in the orange cluster closer to the neighborhoods with potential workplaces in the black cluster.

In general, there are several situations in which this study can be used. Think of yourself in some situation!

## Conclusion

In this study, I mapped the venues in Dublin neighborhoods. I used the K-Means cluster model to group similar neighborhoods. Three clusters were created. The main difference between them is the average number of venues per neighborhood. With that, it was possible to recommend, for example, quieter neighborhoods to live in, as well as neighborhoods with more places to enjoy the night.