# The Battle of Neighbourhoods

## Dublin Postal Districts

## Introduction

Dublin is the capital of Ireland. It is situated on a bay on the east coast, at the mouth of the River Liffey.

Since 1836, the Carmelite Whitefriar Street Church has been the home of the relics of St Valentine – bones and a vial of his blood – they were given by Pope Gregory XVI to Dublin preacher Fr. John Spratt. A lot of young couples come to visit the relics of the Patron Saint of Engagement. With four Nobel prize winners (Yeats, Beckett, Shaw and Heaney), universities of global distinction in Trinity and UCD, numerous books festivals and a world-class new city library, it’s without a doubt that the capital has literature in its blood. With 1,252 acres Phoenix Park is the largest enclosed park in any capital city in Europe, it’s five times bigger than London’s Hyde Park. Dublin is an historical and contemporary centre for education, the arts, administration and industry.

Ireland's capital city attracts thousands of visitors every year. The city is divided into 22 postal districts. There are more than 8000 listings for Dublin City on Airbnb. Any first-time visitor should know which is the best district for short stay with all the amenities nearby. Further, any new entrepreneur who wants to venture into new cafe or restaurant, should be able to identify existing venues and areas already covers. With Dublin city becoming congested, multiple new residential and commercial developments have started outside the Dublin 1 and 2 districts. This analysis will provide ample information for starting a new venture where very less venues are available.

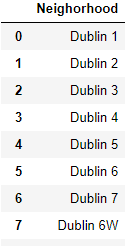
Data Description:

The data used to solve this problem is geolocation data collected from FourSquare. Wikipedia is used to extract the postal codes for the Dublin postal districts. Using the names of the Dublin postal districts, latitude and longitude are generated from Python Geopy (Nominatim). These coordinates are then used to extract nearby venues using FourSquare API. Top 10 common venue categories are then identified for each Postal District and then grouped using unsupervised learning Kmeans clustering. The clustering will provide information to tourists as per their taste to choose best place for short stay. The clusters based on categories will provide business opportunities for opening new restaurants or cafes.

## Methodology

* Imported the required dependencies.
* Extracted Dublin Postal District Codes from Wikipedia <https://en.wikipedia.org/wiki/List_of_Dublin_postal_districts>

Beautiful Soup package was used to extract the data from the Wikipedia and created a pandas dataframe which contains the list of the postal districts



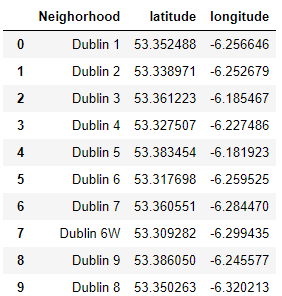
* The Wikipedia page did not have coordinates for Dublin postal districts. I used Nominatim Geocoding service, which is built on top of OpenStreetMap data. The method used to extract the coordinates is :

1. First delay the Geocoding 1 second between each address. This is convenient when you are Geocoding a large number of physical addresses as the Geocoding service provider can deny access to the service.

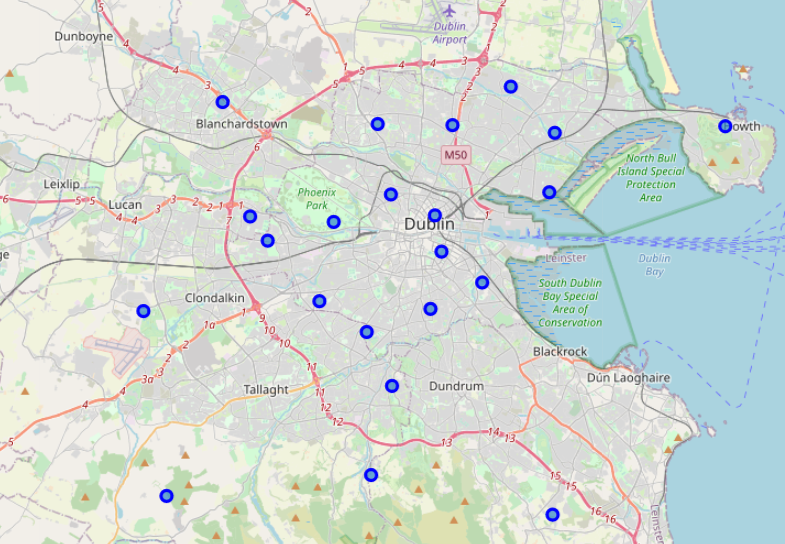
2. Create a df['location'] column by applying geocode we created.

3. Third, we can create latitude, longitude, and altitude as a single tuple column.

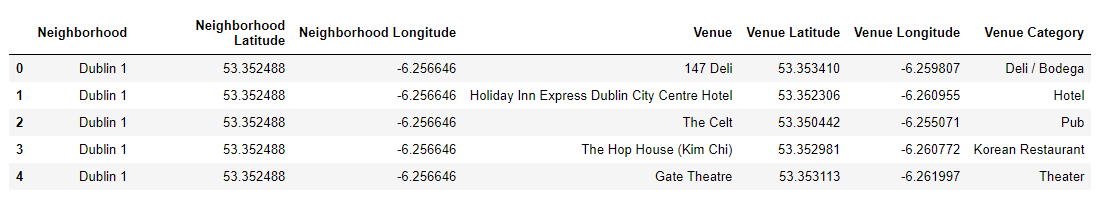
4. Finally, split latitude, longitude, and altitude columns into three separate columns



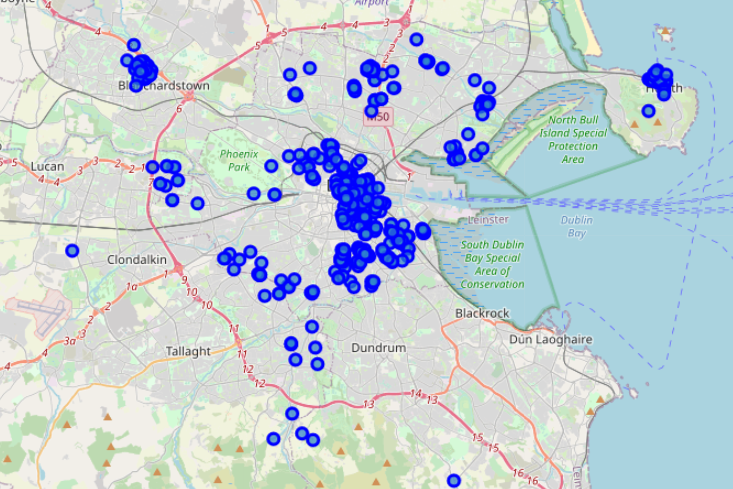
- To check the coordinates, these are displayed on the map of Dublin using Folium package



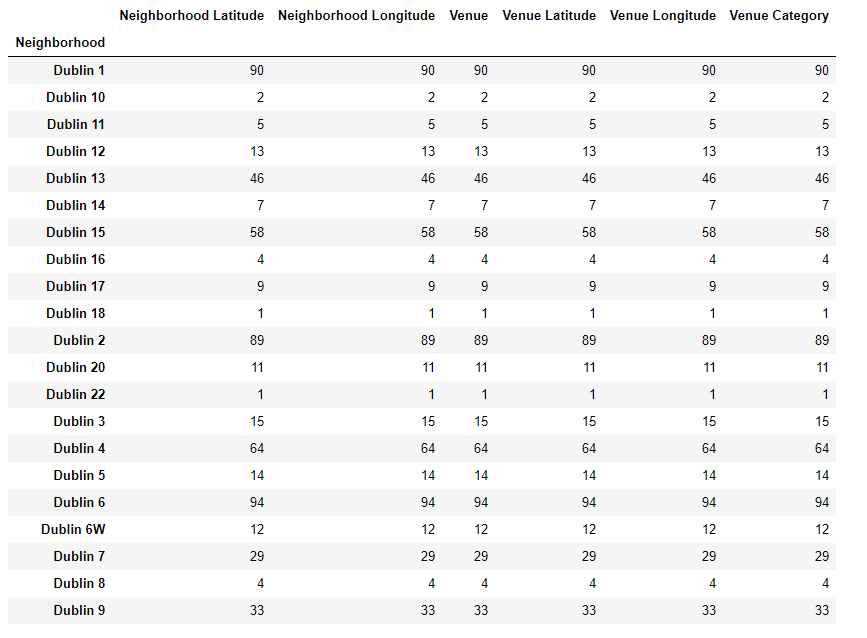
* Foursquare API is used to extract nearby venues within 10 km. Since only Postal districts are spread around in area, some districts outside the city centre are more than 10 km in area.



* Some of the categories were removed. These categories are large buildings which are not in all the postal districts and also cannot be started new as an entrepreneur.
* Again, plotted these venues on the Dublin Map and initial analysis from the map shows clusters in some of the districts such Dublin 1, Dublin 2, Dublin 6 and Dublin 15. Dublin 1 and Dublin 2 are in city centre and highly populated with both business and residences. Dublin 15 and Dublin 6 has large shopping centres. Dublin 13 is a harbour point with many seafood restaurants and cafes.

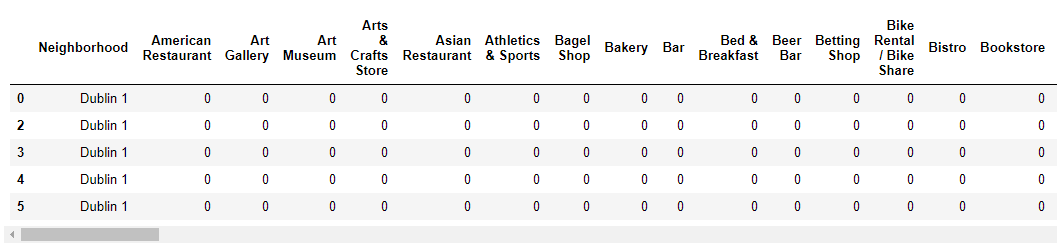


* Grouped these venues by districts in which they are located to count the venues in each district.

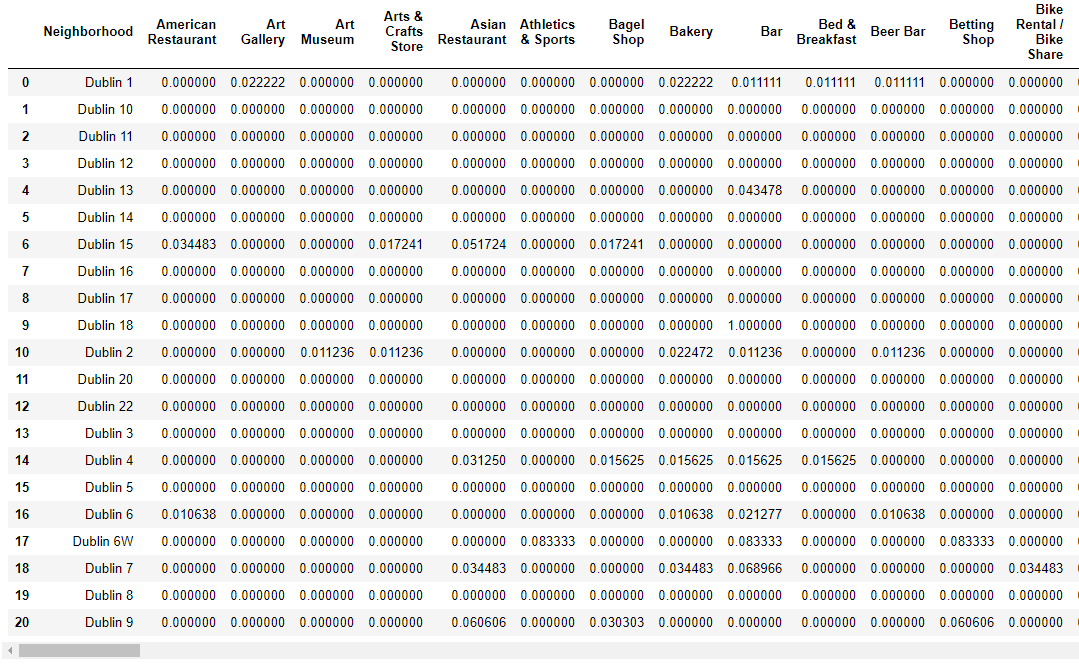


As per the map, clearly there are existing clusters around Dublin 1, 2, 6 and 15

* There are 143 unique categories of venues.
* Analysis of each postal district was done next
* One hot encoding was used for venue categories.



* grouped rows by postal district and by taking the mean of the frequency of occurrence of each category



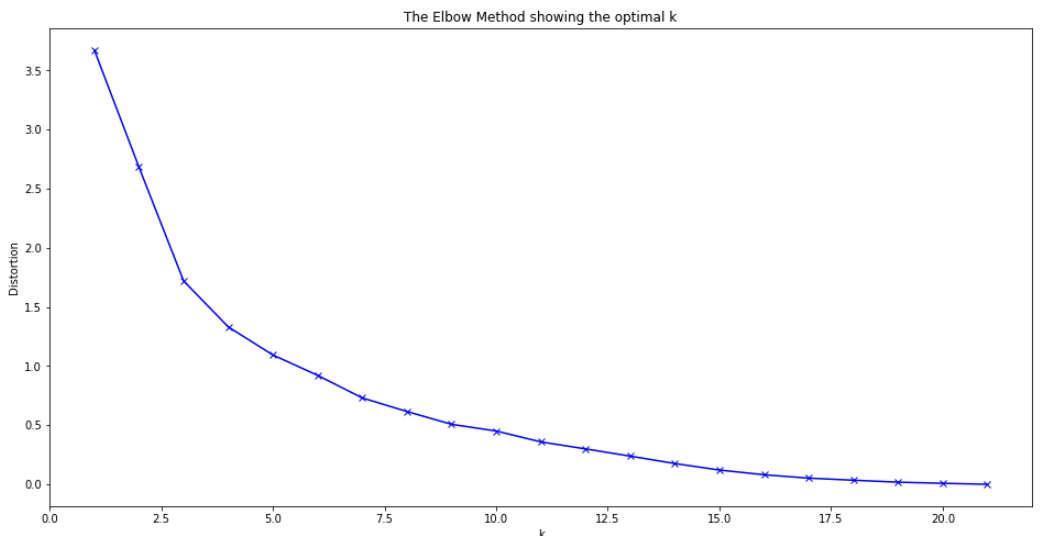
* top 5 venue categories were printed for each district. Dublin 1 and Dublin 2 has high frequency of coffee shops



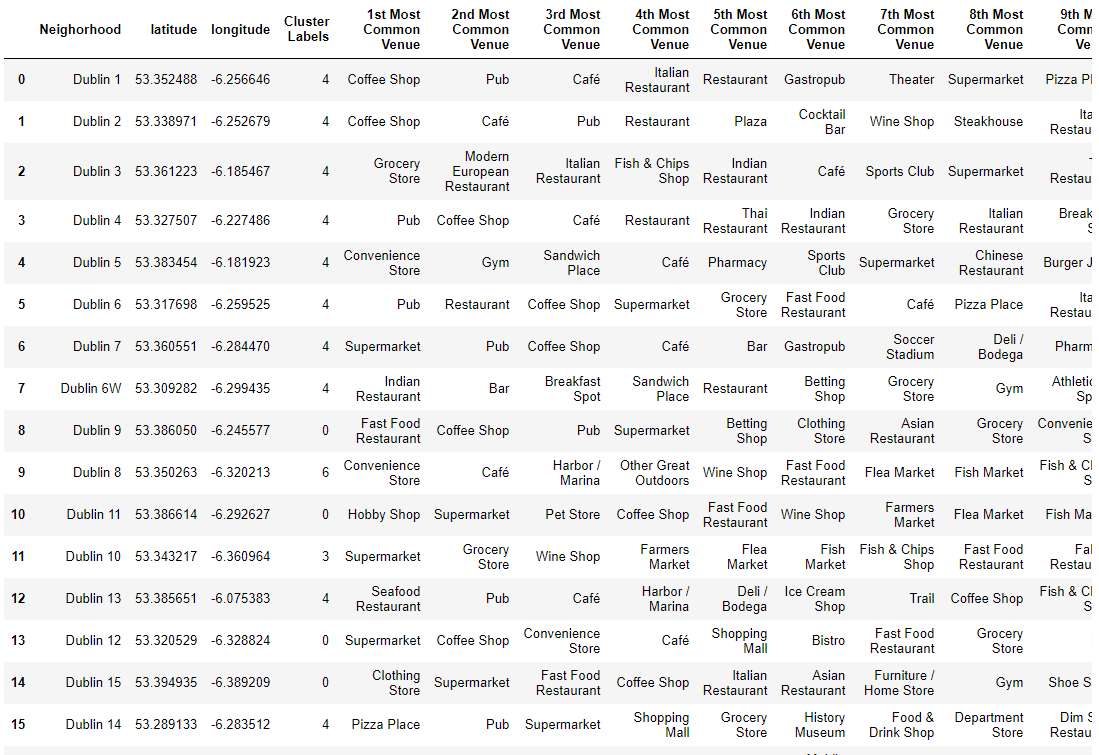
* A new dataframe is created which as top 10 venues for each district.



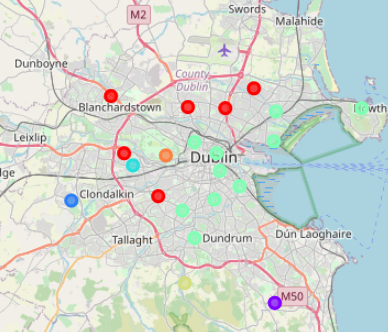
* Used KNN for clustering of these districts
* Initial tested with 5 nearest neighbours but that clustered most of the highly dense districts. So, used elbow method to find optimal number of clusters.



* From the plot it was clear, to use 8-10 clusters, so opted for 9 clusters.
* A new dataframe is created which included clusters as well as top 10 venues for each postal district



* Finally, the clusters are plotted on the map using Folium.



* Each cluster is examined to check which districts it contains.

## Results

* Cluster 1 have 6 postal districts, with top venues as supermarket, coffee shops. These districts are outside city centre limits and with good residential denseness and small and medium business
* Cluster 2 just have Dublin 18, which is also outside city centre limits. It has both residential and large business.
* Cluster 3 also have just one postal district, Dublin 22. This district has industrial estate with most of the car dealership.
* Cluster 4 near same as cluster 3, both with good residential areas, so venues are different to cluster 3
* Cluster 5, the biggest cluster that covers Dublin city centre. Similarities of the venues has clustered these districts together.
* Cluster 6 is on the footsteps of mountains, so different venues as compared to other clusters
* Cluster 7 just have one postal district, which have the Europe’s biggest Park inside the capital city.

## Discussion

The cluster analysis provided a good view of the postal districts in Dublin. Districts in dense, high population areas were clustered together. Clusters with just one districts are large areas which have industrial estates and less residentials, so they show opportunities for starting small to medium businesses. The large clusters are having opportunities for smaller businesses like café, coffee shops. Also, these larger clusters are tourist attractions. The analysis can provide information to first time visitors to city, where to book there Airbnb with all major venues nearby.

## Conclusion

This report can be advanced further by getting population and number of houses data in the postal districts and combined with venues can give a 360-degree view of the area. Also, adding crime data, it can help to find better district to buy a house for living, with all amenities nearby.

Dublin is a tourist destination. Adding tourist attractions to the maps along with other venues, can help tourist guides and tourists to plan the visits beforehand.

Due to restrictions on Four Square API, some of the areas in postal districts has been exempted from the analysis.