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

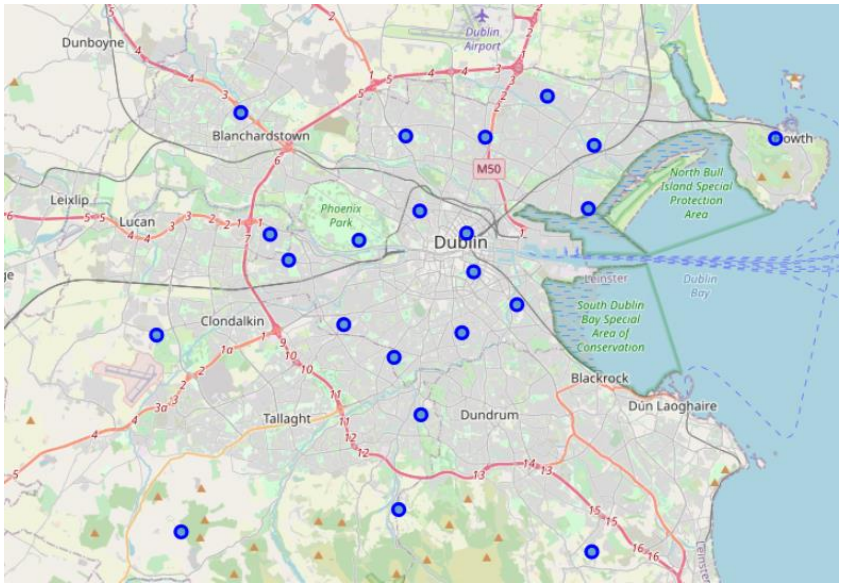
- Imported the required dependencies.
- Extracted Dublin Postal District Codes from Wikipedia
  https://en.wikipedia.org/wiki/List_of_Dublin_postal_districts
  Beautiful Soup package was used to extract the data from the Wikipedia and created a pandas dataframe which contains the list of the postal districts

	Neighorhood
0	Dublin 1
1	Dublin 2
2	Dublin 3
3	Dublin 4
4	Dublin 5
5	Dublin 6
6	Dublin 7
7	Dublin 6W

- 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

	Neighorhood	latitude	longitude
0	Dublin 1	53.352488	-6.256646
1	Dublin 2	53.338971	-6.252679
2	Dublin 3	53.361223	-6.185467
3	Dublin 4	53.327507	-6.227486
4	Dublin 5	53.383454	-6.181923
5	Dublin 6	53.317698	-6.259525
6	Dublin 7	53.360551	-6.284470
7	Dublin 6W	53.309282	-6.299435
8	Dublin 9	53.386050	-6.245577
9	Dublin 8	53.350263	-6.320213

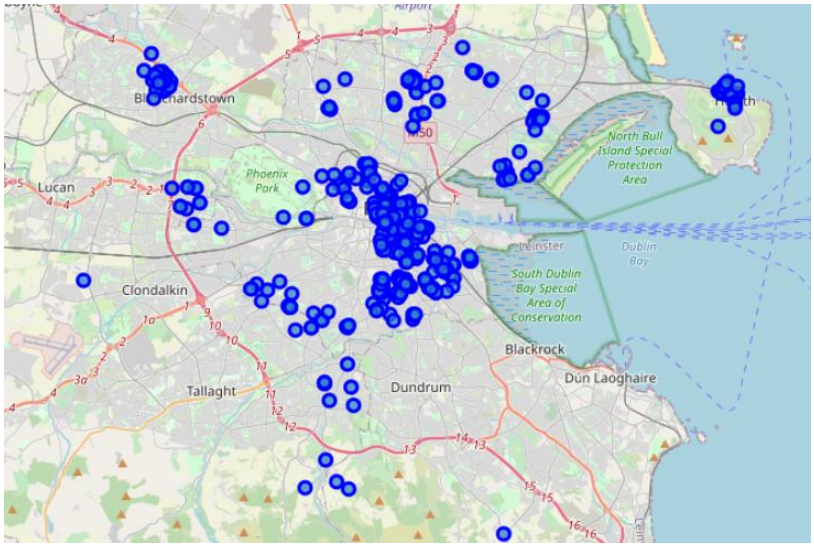
- To check the coordinates, these are displayed on the map of Dublin using Folium package



- 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.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Dublin 1	53.352488	-6.256646	147 Deli	53.353410	-6.259807	Deli / Bodega
1	Dublin 1	53.352488	-6.256646	Holiday Inn Express Dublin City Centre Hotel	53.352306	-6.260955	Hotel
2	Dublin 1	53.352488	-6.256646	The Celt	53.350442	-6.255071	Pub
3	Dublin 1	53.352488	-6.256646	The Hop House (Kim Chi)	53.352981	-6.260772	Korean Restaurant
4	Dublin 1	53.352488	-6.256646	Gate Theatre	53.353113	-6.261997	Theater

- 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.

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Dublin 1	90	90	90	90	90	90
Dublin 10	2	2	2	2	2	2
Dublin 11	5	5	5	5	5	5
Dublin 12	13	13	13	13	13	13
Dublin 13	46	46	46	46	46	46
Dublin 14	7	7	7	7	7	7
Dublin 15	58	58	58	58	58	58
Dublin 16	4	4	4	4	4	4
Dublin 17	9	9	9	9	9	9
Dublin 18	1	1	1	1	1	1
Dublin 2	89	89	89	89	89	89
Dublin 20	11	11	11	11	11	11
Dublin 22	1	1	1	1	1	1
Dublin 3	15	15	15	15	15	15
Dublin 4	64	64	64	64	64	64
Dublin 5	14	14	14	14	14	14
Dublin 6	94	94	94	94	94	94
Dublin 6W	12	12	12	12	12	12
Dublin 7	29	29	29	29	29	29
Dublin 8	4	4	4	4	4	4
Dublin 9	33	33	33	33	33	33

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.

	Neighborhood	American Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Bagel Shop	Bakery	Bar	Bed & Breakfast	Beer Bar	Betting Shop	Bike Rental / Bike Share	Bistro	Bookstore
0	Dublin 1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	Dublin 1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	Dublin 1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	Dublin 1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	Dublin 1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

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

	Neighborhood	American Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Bagel Shop	Bakery	Bar	Bed & Breakfast	Beer Bar	Betting Shop	Bike Rental / Bike Share
0	Dublin 1	0.000000	0.022222	0.000000	0.000000	0.000000	0.000000	0.000000	0.022222	0.011111	0.011111	0.011111	0.000000	0.000000
1	Dublin 10	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2	Dublin 11	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
3	Dublin 12	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
4	Dublin 13	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.043478	0.000000	0.000000	0.000000	0.000000
5	Dublin 14	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
6	Dublin 15	0.034483	0.000000	0.000000	0.017241	0.051724	0.000000	0.017241	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
7	Dublin 16	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
8	Dublin 17	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
9	Dublin 18	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000	0.000000
10	Dublin 2	0.000000	0.000000	0.011236	0.011236	0.000000	0.000000	0.000000	0.022472	0.011236	0.000000	0.011236	0.000000	0.000000
11	Dublin 20	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
12	Dublin 22	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
13	Dublin 3	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
14	Dublin 4	0.000000	0.000000	0.000000	0.000000	0.031250	0.000000	0.015625	0.015625	0.015625	0.015625	0.000000	0.000000	0.000000
15	Dublin 5	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
16	Dublin 6	0.010638	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.010638	0.021277	0.000000	0.010638	0.000000	0.000000
17	Dublin 6W	0.000000	0.000000	0.000000	0.000000	0.000000	0.083333	0.000000	0.000000	0.083333	0.000000	0.000000	0.083333	0.000000
18	Dublin 7	0.000000	0.000000	0.000000	0.000000	0.034483	0.000000	0.000000	0.034483	0.068966	0.000000	0.000000	0.000000	0.034483
19	Dublin 8	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
20	Dublin 9	0.000000	0.000000	0.000000	0.000000	0.060606	0.000000	0.030303	0.000000	0.000000	0.000000	0.000000	0.060606	0.000000

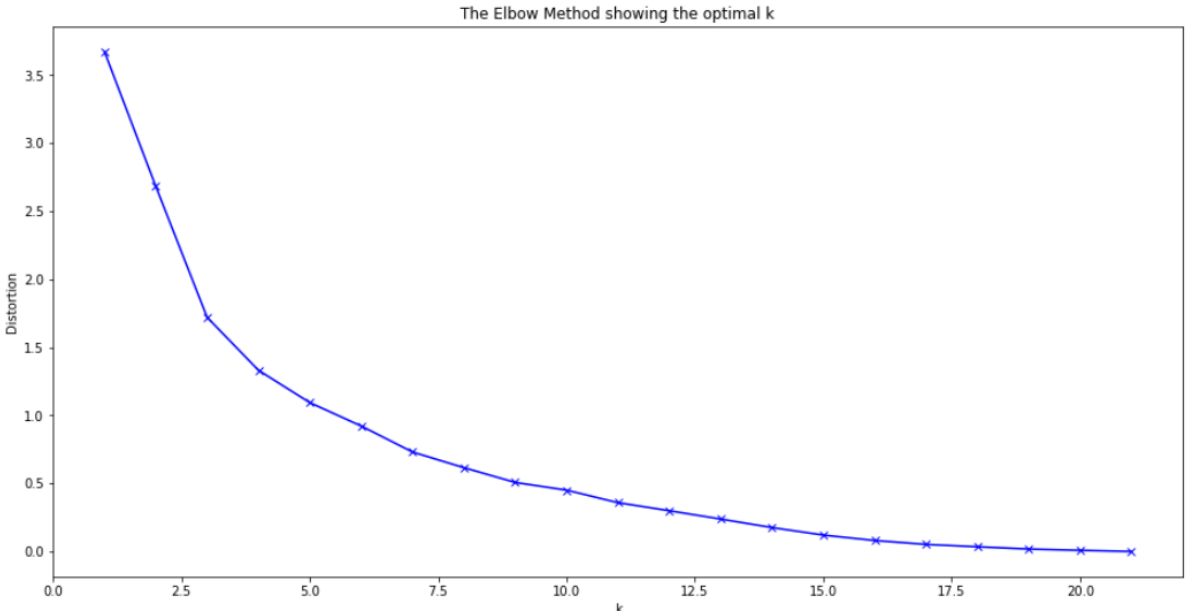
- top 5 venue categories were printed for each district. Dublin 1 and Dublin 2 has high frequency of coffee shops

```
----Dublin 1 ----
                venue  freq
0         Coffee Shop  0.13
1                 Pub  0.11
2               Café  0.09
3  Italian Restaurant  0.04
4             Theater  0.03


----Dublin 10 ----
                    venue  freq
0             Supermarket   0.5
1          Grocery Store   0.5
2            Pastry Shop   0.0
3  North Indian Restaurant   0.0
4           Optical Shop   0.0


----Dublin 11 ----
```

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

	Neighborhood	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	Dublin 1	Coffee Shop	Pub	Café	Italian Restaurant	Restaurant	Gastropub	Theater	Supermarket	Pizza Place	Sandwich Place
1	Dublin 10	Supermarket	Grocery Store	Wine Shop	Farmers Market	Flea Market	Fish Market	Fish & Chips Shop	Fast Food Restaurant	Falafel Restaurant	Food Court
2	Dublin 11	Hobby Shop	Supermarket	Pet Store	Coffee Shop	Fast Food Restaurant	Wine Shop	Farmers Market	Flea Market	Fish Market	Fish & Chips Shop
3	Dublin 12	Supermarket	Coffee Shop	Convenience Store	Café	Shopping Mall	Bistro	Fast Food Restaurant	Grocery Store	Pub	Diner
4	Dublin 13	Seafood Restaurant	Pub	Café	Harbor / Marina	Deli / Bodega	Ice Cream Shop	Trail	Coffee Shop	Fish & Chips Shop	Bar

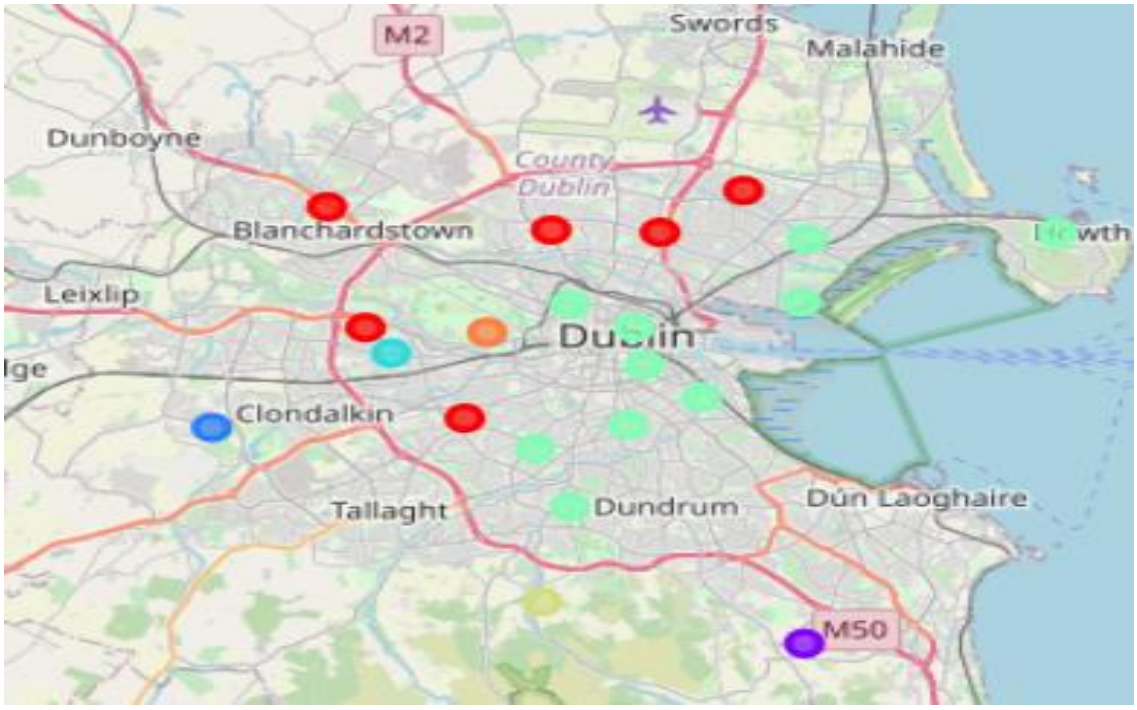
- Used KNN for clustering of these districts
- Initial tested with 5 nearest neighbours but that clustered most of the highly dense districts. So, used elbow method to find optimal number of clusters.



- 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

	Neighorhood	latitude	longitude	Cluster Labels	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 M Comm Ve
0	Dublin 1	53.352488	-6.256646	4	Coffee Shop	Pub	Café	Italian Restaurant	Restaurant	Gastropub	Theater	Supermarket	Pizza Pl
1	Dublin 2	53.338971	-6.252679	4	Coffee Shop	Café	Pub	Restaurant	Plaza	Cocktail Bar	Wine Shop	Steakhouse	Ita Restau
2	Dublin 3	53.361223	-6.185467	4	Grocery Store	Modern European Restaurant	Italian Restaurant	Fish & Chips Shop	Indian Restaurant	Café	Sports Club	Supermarket	Restau
3	Dublin 4	53.327507	-6.227486	4	Pub	Coffee Shop	Café	Restaurant	Thai Restaurant	Indian Restaurant	Grocery Store	Italian Restaurant	Break
4	Dublin 5	53.383454	-6.181923	4	Convenience Store	Gym	Sandwich Place	Café	Pharmacy	Sports Club	Supermarket	Chinese Restaurant	Burger J
5	Dublin 6	53.317698	-6.259525	4	Pub	Restaurant	Coffee Shop	Supermarket	Grocery Store	Fast Food Restaurant	Café	Pizza Place	Ita Restau
6	Dublin 7	53.360551	-6.284470	4	Supermarket	Pub	Coffee Shop	Café	Bar	Gastropub	Soccer Stadium	Deli / Bodega	Pharm
7	Dublin 6W	53.309282	-6.299435	4	Indian Restaurant	Bar	Breakfast Spot	Sandwich Place	Restaurant	Betting Shop	Grocery Store	Gym	Athleti Sp
8	Dublin 9	53.386050	-6.245577	0	Fast Food Restaurant	Coffee Shop	Pub	Supermarket	Betting Shop	Clothing Store	Asian Restaurant	Grocery Store	Convenie S
9	Dublin 8	53.350263	-6.320213	6	Convenience Store	Café	Harbor / Marina	Other Great Outdoors	Wine Shop	Fast Food Restaurant	Flea Market	Fish Market	Fish & Cl S
10	Dublin 11	53.386614	-6.292627	0	Hobby Shop	Supermarket	Pet Store	Coffee Shop	Fast Food Restaurant	Wine Shop	Farmers Market	Flea Market	Fish Ma
11	Dublin 10	53.343217	-6.360964	3	Supermarket	Grocery Store	Wine Shop	Farmers Market	Flea Market	Fish Market	Fish & Chips Shop	Fast Food Restaurant	Fal Restau
12	Dublin 13	53.385651	-6.075383	4	Seafood Restaurant	Pub	Café	Harbor / Marina	Deli / Bodega	Ice Cream Shop	Trail	Coffee Shop	Fish & Cl S
13	Dublin 12	53.320529	-6.328824	0	Supermarket	Coffee Shop	Convenience Store	Café	Shopping Mall	Bistro	Fast Food Restaurant	Grocery Store	
14	Dublin 15	53.394935	-6.389209	0	Clothing Store	Supermarket	Fast Food Restaurant	Coffee Shop	Italian Restaurant	Asian Restaurant	Furniture / Home Store	Gym	Shoe S
15	Dublin 14	53.289133	-6.283512	4	Pizza Place	Pub	Supermarket	Shopping Mall	Grocery Store	History Museum	Food & Drink Shop	Department Store	Dim S Restau

- 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.