Best value affluent living in Dublin

1. INTRODUCTION

In recent years, the housing market in Dublin has gradually stabilized [1] following a decade of economic recovery in Ireland. Now that house prices have stabilized, it is interesting to look at the characteristics of the parts of Dublin that make them valuable in terms of house price. In other words, what features of a region in Dublin make it valuable; and then, what low-cost regions share those high-value features? If we answer this question, we will have discovered the best value location for affluent living in Dublin. The following are two examples of stakeholders that would benefit from knowing the answer to the question posed above: (i) home buyers looking to buy a low-cost house in an affluent area, and (ii) developers looking to purchase low-cost sites in 'affluent' type areas to develop housing for sale.

The Central Bank of Ireland published an economic letter on Population Change and Housing Demand in Ireland [2] late in 2019. Population growth has significantly exceeded the rate of housing supply in Ireland. The Central Bank forecast that about 34,000 houses per annum are required to meet the demand in the next decade. Sustained high levels of in net-inward migration and a shortfall in housing supply between 2011 and 2019 are determined to be driving the demand. In summary, there is a large number of people looking to buy property at present. All of whom want value for money.

Dublin is Ireland's capital and has a population of 1.36 million, 28% of the country's population (4.83 million). House hunters who live (and want to live) in Dublin will be aware of the areas that are more expensive than others are. For example, properties on the South side of the River Liffey cost more, on average, than those on the North side of the river. There are other examples of this that are not as straightforward to define. This study sets out to identify and characterize emerging affluent areas (not as easy to define) that are good value for money.

2. DATA DESCRIPTION

The Property Services Regulation Authority (PSRA) publishes the date of sale, price, and address of every purchased residential property in Ireland via the Residential Property Price Register (PPR) [3]. This has been the case since 1st January 2010 by Irish law under the Property Services (Regulation) Act 2011. Ireland was the last country in the OECD to create a national postcode system; in July 2015 'Eircode' was introduced. Unlike other countries, where groups of addresses are identified using a postcode, Eircode identifies unique addresses. The system was heavily criticized, and adoption has been slow. Dublin postal addresses, however, have a district code attached (since 1961). Dublin has 22 districts. For Dublin only addresses, these codes are easily extractable in the PPR dataset. A more diverse representation of Dublin is possible by parish (a parish being place with a church); Dublin has 83 parishes. This information is not easily extractable from addresses in the PPR dataset. There is a rich dataset for parishes available at www.townlands.ie.

Foursquare is a location-based service that captures users' recommendations about places and venues they have been ('checked in' to), and makes recommendations to users based on their profile match to other similar users. Foursquare [4] offer an application-programming interface ('API', or 'RESTful API') that provides developers access to the location data and recommendations within the Foursquare data set. The data is provided in JavaScript Serialized Object Notation (JSON) format. For example, using the 'explore' endpoint of their API, for a given latitude and longitude and search radius the API can produce a list of venue recommendations about the coordinates given.

3. Methodology

3.1 House price data

3.1.1 Dublin postcodes data

The postal codes for Dublin, Ireland were scraped from Wikipedia using the Beautiful Soup [5] html parser for Python. A list was populated with postal codes from the Wikipedia source data table by searching for <*td*> HTML tags. The list was cast as a Pandas DataFrame. The DataFrame was stripped to contain only the 'Dublin' part of the postal code, followed by the one or two digit postal code identifier number. (One of the postal codes follows a different format to the other 24 codes; Dublin 6 West , Dublin 6w.) The DataFrame rows were passed to the OpenCage geocode API [6] to return the longitude and latitude for the center of each postal district. The result was appended to the DataFrame. A map of Dublin was generated using the Folium mapping library [7] and the location of the postal districts was marked, thereon, using DivIcon [8].

3.1.2 Dublin house price data

The Dublin house price dataset was downloaded as a csv file and converted to a DataFrame object. The dataset was checked for missing values. There were no missing house prices in a total of 408,419 rows. There were 331,998 missing postcodes. (The data set is for the country of Ireland and only the county of Dublin uses postal district – this account for the large number of missing postal codes.) The DataFrame was filtered by county (keeping only Dublin country entries); unwanted columns were dropped. Houses not having 'full market value' were dropped. The final DataFrame had 72,362 rows. The price column included a special character for the Euro (€) currency symbol; this was stripped from the column.

3.1.3 Exploratory data analysis of Dublin house price data

A simple statistical analysis was applied to give the count, mean (μ), standard deviation (σ), minimum and maximum of the Dublin house price data. The lower (25), median (50), and upper (75) percentiles were also described. Some outliers in the data were removed by dropping prices outside the range: $\mu \pm 2\sigma$. The data was then analyzed by postal code. See the results section below (Section 4).

3.2 Location data by parish

The local parish data was downloaded as a 'geojson' file. The file was loaded and parsed to extract the title of each Dublin parish and its' geographical coordinates. The results were loaded to a DataFrame. The Foursquare API was used next. Each parish was passed to Foursquare to get venue recommendations within a radius of 1500 m. The result was normalized from json format and the venue categories added. This was performed for each parish result. All recommendations were then gathered in a single DataFrame. The first five rows of that the DataFrame is shown in Figure 1. (The Dublin Parishes are shown as 'Neighborhoods'.)

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Saggart	53.2712997522548	-6.44658461901325	Anvil Restaurant	53.280977	-6.443511	Restaurant
1	Saggart	53.2712997522548	-6.44658461901325	Dunnes Stores	53.279119	-6.446794	Supermarket
2	Saggart	53.2712997522548	-6.44658461901325	Centra	53.279113	-6.446824	Grocery Store
3	Saggart	53.2712997522548	-6.44658461901325	Saggart Pharmacy	53.279424	-6.445194	Pharmacy
4	Saggart	53.2712997522548	-6.44658461901325	China Garden Restaurant	53.280329	-6.444726	Chinese Restaurant

Figure 1 Venue recommendation table, first five rows.

Next, each parish was analyzed using this table. The frequency of occurrence of venue type was determined for each parish. Figure 2 shows an example of the top 10 most common venue types for the Booterstown parish. The frequency of occurrence of each venue type and for each parish, was gathered in a single DataFrame (this is the input data for the machine-learning algorithm, introduced in the next section.)

```
----Booterstown----
              venue  freq
0               Pub  0.13
1              Café  0.11
2       Coffee Shop  0.06
3 Convenience Store  0.06
4     Shopping Mall  0.04
5              Park  0.04
6             Hotel  0.04
7             Diner  0.04
8        Restaurant  0.04
9       Supermarket  0.04
```

Figure 2 Top 10 most common venues in Booterstown parish.

	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
0	Aderrig	Airport	Train Station	Gastropub	Hotel	History Museum	English Restaurant	Food Court	Food & Drink Shop	Flea Market
1	Artane	Supermarket	Bus Stop	Fast Food Restaurant	Grocery Store	Chinese Restaurant	Café	Pizza Place	Park	Pub
2	Baldongan	Racetrack	Cricket Ground	Golf Course	Zoo Exhibit	Falafel Restaurant	Food Service	Food Court	Food & Drink Shop	Flea Market
3	Baldoyle	Train Station	Supermarket	Bar	Badminton Court	Park	Gastropub	Fast Food Restaurant	Pub	Convenience Store
4	Balgriffin	Coffee Shop	Pet Store	Hotel Bar	Fast Food Restaurant	Shopping Mall	Mobile Phone Shop	Supermarket	Garden Center	Bakery

Figure 3 DataFrame with the most common venue type recommendations from Foursquare, first five rows.

Finally, a single DataFrame stored the results for all parishes, as shown in Figure 3 shows the first five rows of this DataFrame.

3.3 Machine learning model

The k-means clustering machine-learning algorithm is suitable for partitioning data according to similarity. It is an unsupervised algorithm. The algorithm objective is to minimize the *intra*-cluster distances (distance between data points within a cluster) and maximize the *inter*-cluster (distance between clusters). This algorithm is suitable for the problem described here, since we are attempting group together locations that are similar with respect to popular venue categories.

The k-means algorithm was applied so that parishes with similar results for frequency off occurrence of venue types were groups together. The cluster results, crosschecked with the house price data, gives an insight into best-value affluent living in Dublin.

4. RESULTS

Figure 4 shows a histogram of house prices in Dublin from 2010 to 2019 after outlier outside the range: $\mu \pm 2\sigma$, were removed. Figure 5 shows a box plot of house prices by Dublin postal code from 2010 to 2019.

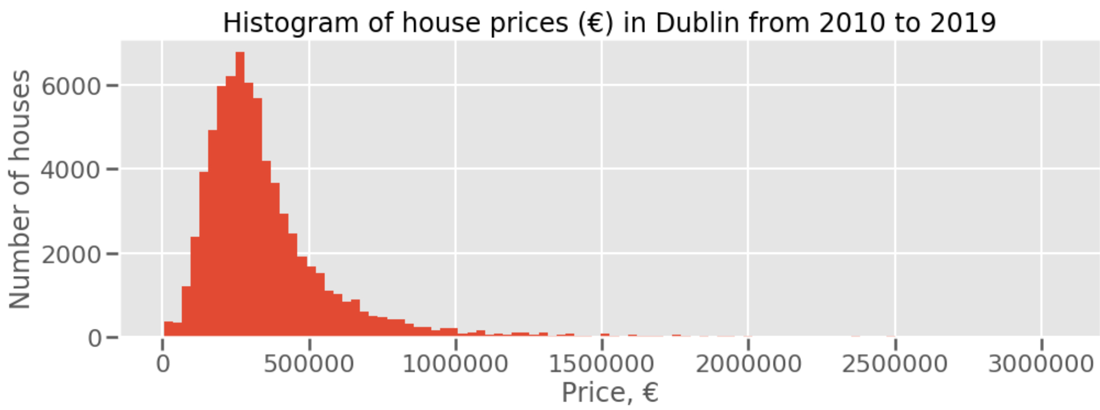

Figure 4 Histogram of house prices in Dublin from 2010 to 2019.

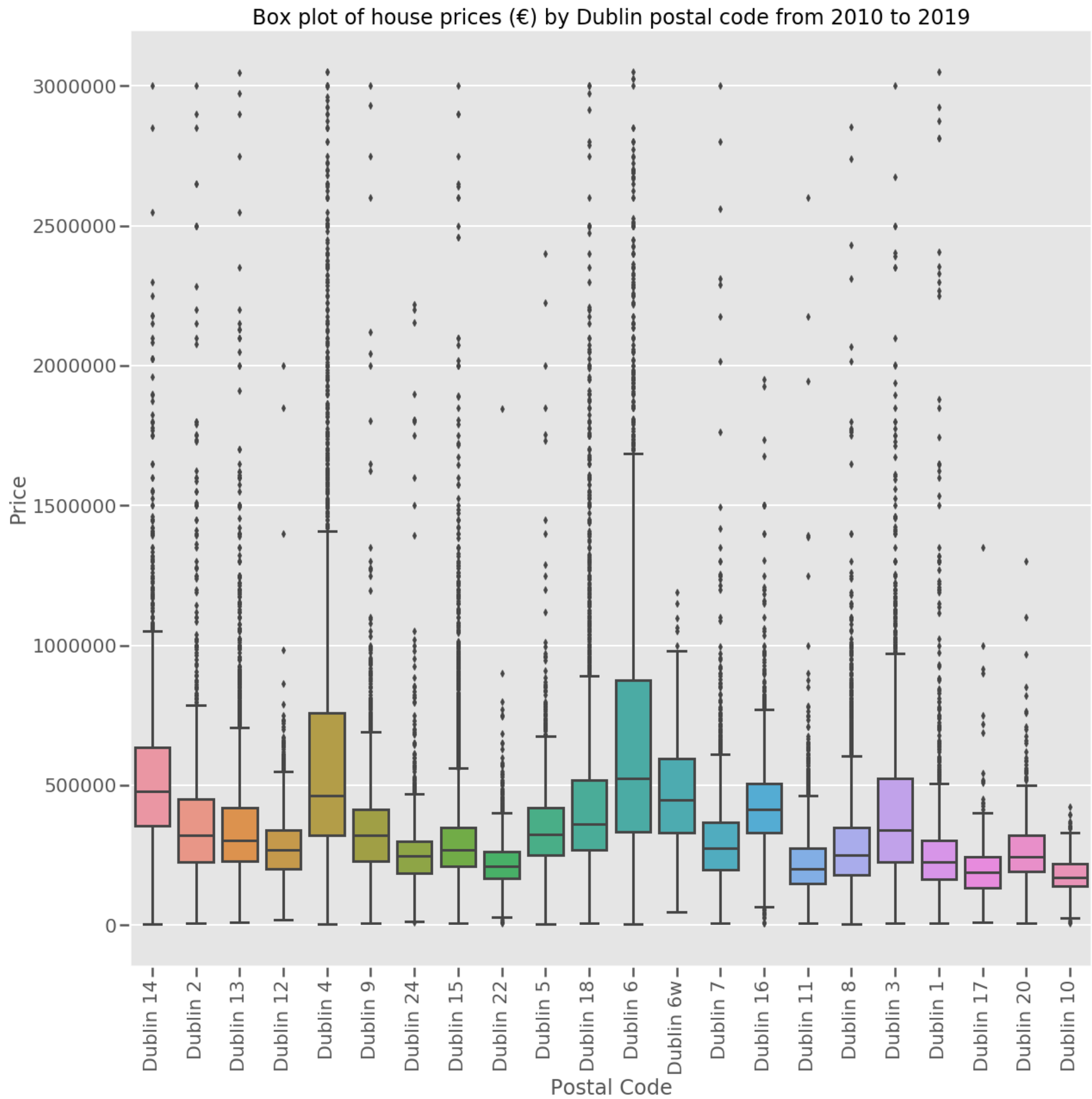

Figure 5 Box plot of house prices (€) by Dublin postal code from 2010 to 2019.

Figure 6 shows the clusters of Dublin parishes given filled circles of distinct colors.

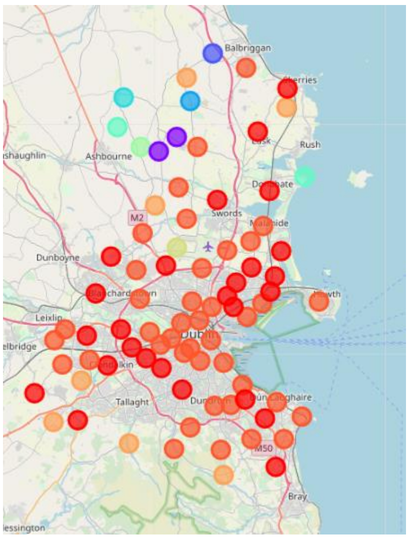

Figure 6 Clusters by color of all 83 no. Dublin parishes.

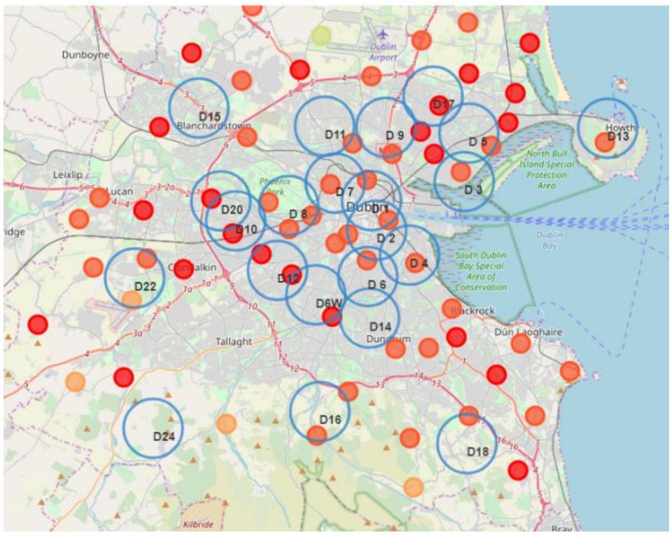

Figure 7 Clusters (filled circles) overlaid on Dublin postal code (open circles with postal code in annotation).

Figure 7 shows clusters of the Dublin parishes focused on the Dublin postal districts regions. Clusters are the filled circles of distinct colors, while the open circles are overlaid to represent the Dublin postal code regions centers. Each open circle has the corresponding postal code shown annotated.

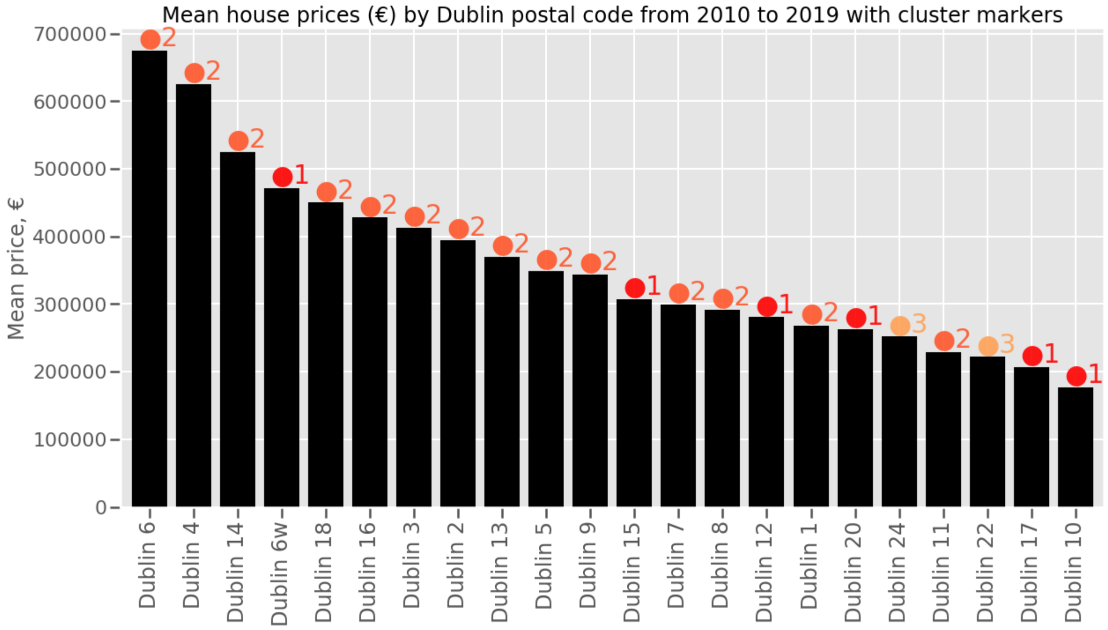

Figure 8 Mean house prices (€) by Dublin postal code from 2010 to 2019 with cluster markers; red (●) is cluster 1, orange (●) is cluster2, yellow (●) is cluster 3.

Figure 8 shows a bar plot of mean house price in the Dublin postal code districts from 2010 to 2019; the filled circles, at the top of each bar, denotes the dominant cluster for each postal code. Table 1 shows other metrics from the analysis.

Table 1 Other metrics from the analysis.

Number of Dublin postal districts	22
Number of Dublin parishes	83
Number of unique categories in Foursquare recommendations	217
Number of Foursquare recommendations for Dublin parishes	2151
Number of houses sold in Dublin from 2010 to 2019	72,362
Mean house price in Dublin from 2010 to 2019, μ	€407,902
Mean house price in Dublin from 2010 to 2019 of odd postal districts	€332,033
Mean house price in Dublin from 2010 to 2019 of even postal districts	€464,526
Standard deviation of house price in Dublin from 2010 to 2019, σ	€1,326,543

5. Discussion

The mean price of a house in a Dublin postal district is €407,902. The standard deviation is €1,326,543. The standard deviation is relatively large compared to the mean, so the spread or distribution of house prices about the mean value is quite large. Figure 4 shows a positively skewed distribution of house prices in Dublin. This result in further exposed in Figure 5 where a large number of outliers are present for each postal district. Dublin 6 has the highest median house price of all the districts and is the only district with a median above €500,000.

With reference to Table 1, it is interesting to look at the mean house price in districts in the north of Dublin (odd numbered) versus that of southern districts (even numbered). Houses in the North of Dublin cost about 71% of that on the South side.

The clustering analysis was carried out using the parish data for Dublin (83 parishes), as opposed using to the postal code districts (22 districts). The parish data gives a wider geographical spread for recommendations from the Foursquare API. Figure 6 and Figure 7 illustrate the geographical differences of parishes and postal districts. Figure 7 shows the postal districts overlaid on the map of clusters given by the machine-learning k-means algorithm. Three clusters dominate the postal districts region of the map. Cluster 1 (the red-filled circles) are characterized by supermarkets, convenience and grocery stores, some pubs and restaurants. Cluster 2 (orange-filled circles) are characterized by parks, pubs, restaurant and cafés, and predominantly urban locations. Cluster 3 (yellow-filled circles) are characterized by outdoors-oriented venues like golf courses, cricket grounds and zoo exhibits, and are predominantly outside the major motorway (C-shape ring-road in Figure 7).

Figure 8 shows the mean price of houses by postal district in descending order. The color and number on top of each bar shows the associated cluster for each postal district. Returning to the objective set out in the introduction, we want to identify good value in affluent areas. The Cluster 2 (orange) has nine of the top ten most expensive house prices. However, there is a large difference between the most (€674,867) and least (€228,471) expensive district for this cluster. Postal districts with a mean house price at the lower end of the orange cluster therefore represent good value for money. Examples being, Dublin 1, Dublin 7, and Dublin 8. Cluster 1 (red) has a similar story, where the most expensive postal district is Dublin 6w (€471,329) and least expensive Dublin 10 (€176,284). Examples of good value in this cluster are Dublin 20 and Dublin 17. Cluster 3 appears to be the most geographically constrained of the clusters in that they are predominantly located outside the main ring road and characterized by outdoors or activity-related venues (for example, Gold courses). This is likely because of the proximity of these locations to the Dublin Mountains. Without prior knowledge of the area, this characteristic is not obvious; rather, it revealed itself through the methodology applied here.

There are a number of way to improve the analysis presented. The following are two suggestions to improve the analysis. The number of recommendations for rural parishes was smaller than that in urban areas. For example, some areas had only three recommendations while others had more than the Foursquare API limit of 100. More recommendations for rural areas would achieve a fairer comparison of parishes within the k-means algorithm. During the period 2010 to 2019, the housing market in Ireland experienced significant change. This was due to a large adjustment occurring after the economic crash of 2008/2009. It would be interesting to apply consumer price indices to adjust for inflation during this time. The result would very likely effect the analysis here in some way.

6. CONCLUSION

The conclusions of the analysis presented here are as follows:

1. Dublin 6 is the most expensive district of Dublin in terms of house price.
2. Dublin 10 is the least expensive district of Dublin in terms of house price.
3. The North side of Dublin is better value for money that the South side.
4. Dublin 1, Dublin 7, and Dublin 8 are good value in the cluster determined to have the highest valued houses.
5. Dublin 17 and Dublin 20 are good value in the cluster determined to have the next highest valued houses.

7. References

[1] "Irish Property Prices Analysis – Q3 2019." [Online]. Available: https://www.auctioneera.ie/irish-property-prices. [Accessed: 02-Feb-2020].
[2] "Economic Letter: Population Change and Housing Demand in Ireland." [Online]. Available: https://centralbank.ie/news-media/press-releases/press-release-economic-letter-population-change-and-housing-demand-in-ireland-10-december-2019. [Accessed: 02-Feb-2020].
[3] "Residential Property Price Register - home page." [Online]. Available: https://propertypriceregister.ie/. [Accessed: 02-Feb-2020].
[4] "Places API - Foursquare Developer." [Online]. Available: https://developer.foursquare.com/docs/api. [Accessed: 02-Feb-2020].
[5] "Beautiful Soup Documentation." [Online]. Available: https://www.crummy.com/software/BeautifulSoup/bs4/doc/index.html.
[6] "OpenCage API." [Online]. Available: https://opencagedata.com/api.
[7] "Folium." [Online]. Available: https://github.com/python-visualization/folium/blob/master/README.rst#python-data-leafletjs-maps.
[8] "Leaflet, DivIcon." [Online]. Available: https://leafletjs.com/reference-1.6.0.html#divicon.