Best value affluent living in Dublin

IBM DATA SCIENCE PROFESSIONAL CERTIFICATION

COURSERA CAPSTONE ASSIGNMENT

Objective

- Determine the **characteristics** of **high-value** locations in Dublin
- Determine locations that share **high value characteristics** but are in **low-cost** areas.

Data Sources

- Property Price Register
  - 408,000 records in over 10 yrs
- Townlands.ie
  - Parish data for Ireland
  - by Irish OpenStreetMap community
- Foursquare API
  - Location-based service
  - venue recommendations

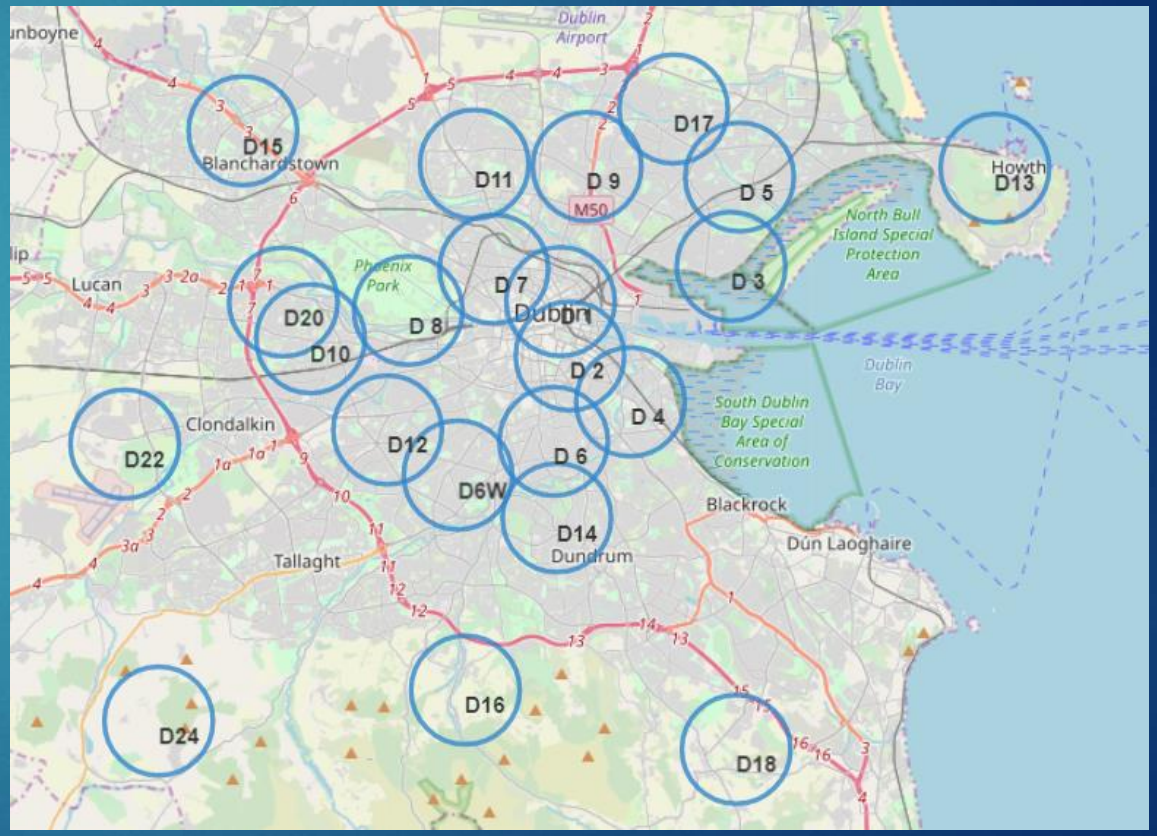
Methodology – get the data and cleanse

- **House price** data (PPR dataset)
  - Pandas DataFrame
  - Filter for Dublin
  - Cleanse

	Date of Sale (dd/mm/yyyy)	Address	Postal Code	Description of Property	Price
0	04/01/2010	The Mews, 11A Wyckham Park Road, Dundrum	Dublin 14	Second-Hand Dwelling house /Apartment	345000.00
1	05/01/2010	134 longboat quay north, sir john rogersons quay	Dublin 2	New Dwelling house /Apartment	225000.00
2	05/01/2010	3 Myrtle Square, The Coast, Baldoyle	Dublin 13	New Dwelling house /Apartment	265198.00
3	05/01/2010	87 MANGERTON RAOD, DRIMNAGH	Dublin 12	Second-Hand Dwelling house /Apartment	182556.12
4	05/01/2010	9 Temple Manor Drive, Walkinstown, Dublin 12	Dublin 12	Second-Hand Dwelling house /Apartment	300000.00

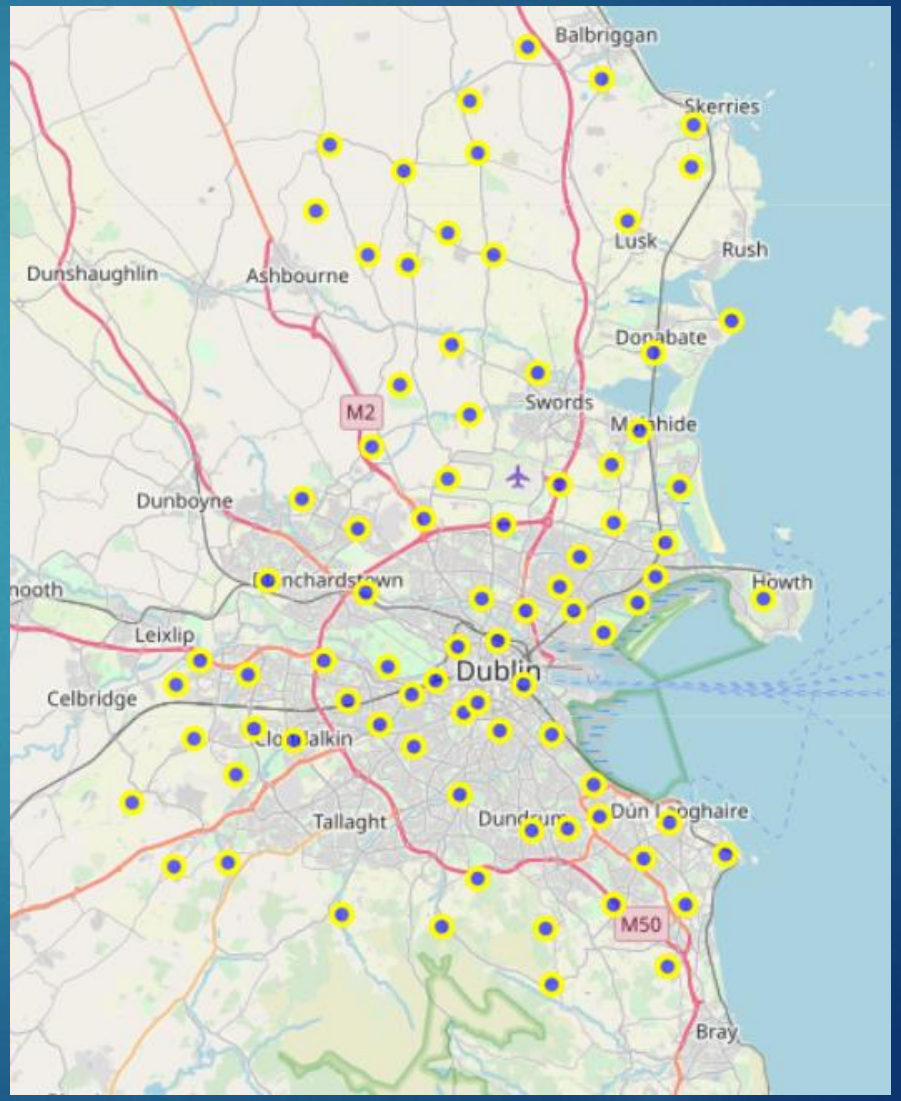
Methodology – get the data and cleanse

- **Postal code** data
  - Scrape Wikipedia w/Beautiful Soup
  - Use OpenCage API to get lat&long
  - 22 postal codes



Methodology – get the data and cleanse

- **Parish** data (Townlands.ie)
  - 83 parishes
  - better spread of location using parishes!

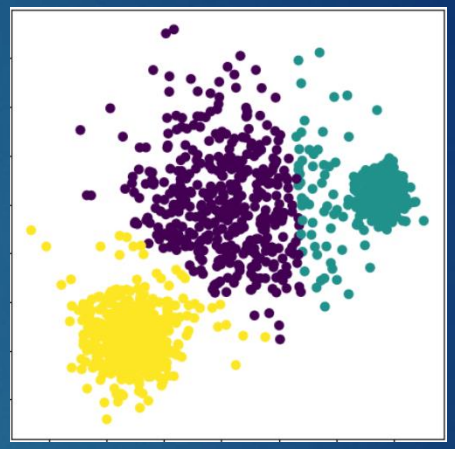


Methodology – ‘k-means’ algorithm

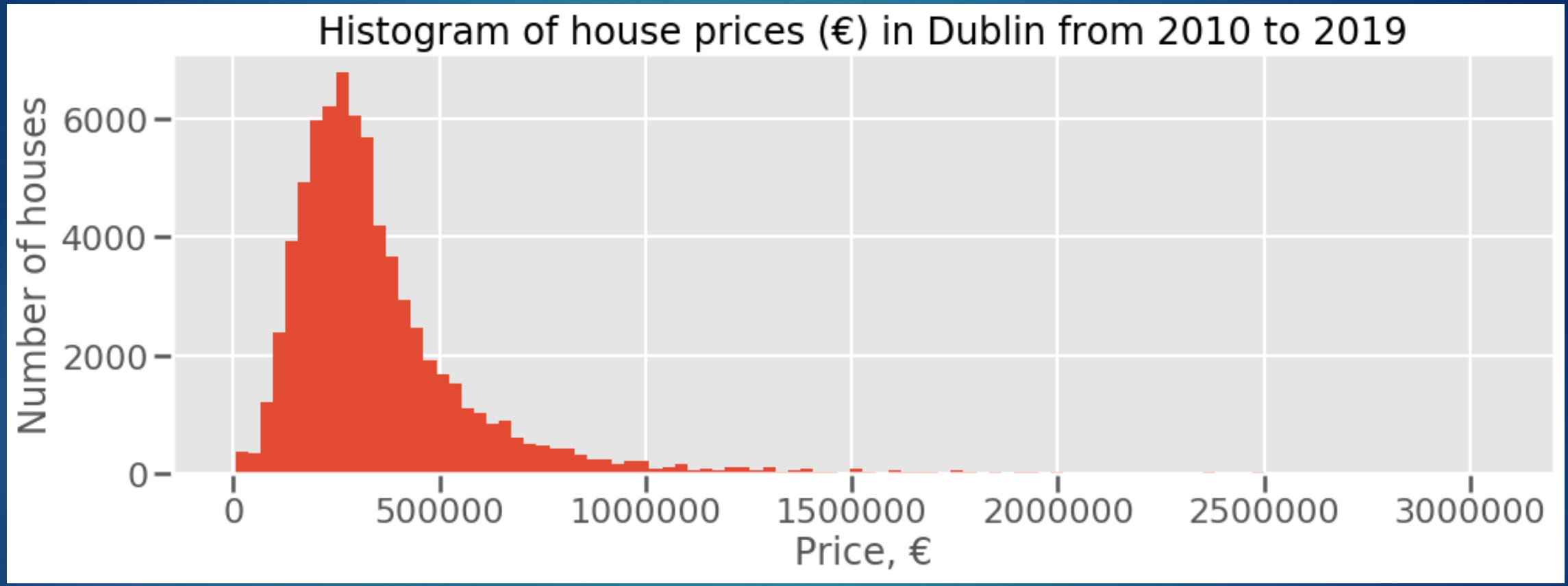
- **Clustering**
  - by most frequently occurring venue type
  - for each parish
  - using ScikitLearn

----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



Results – house prices



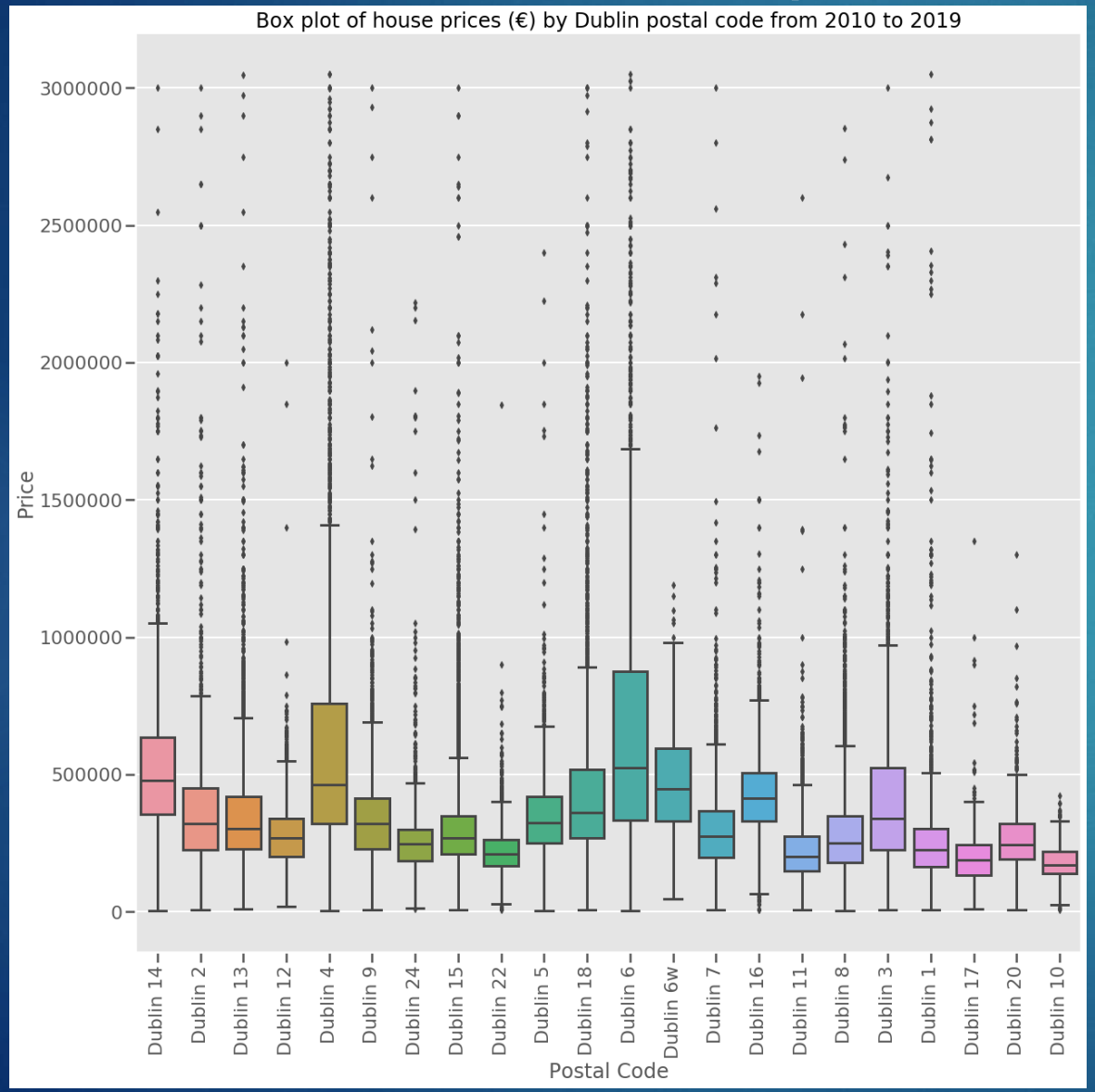
Mean, μ = € 407,902

Standard dev, σ = €1,326,543

Outliers excluded: $\mu \pm 2\sigma$

Positively skewed distribution

Results – house prices



- Lots of outliers
- Dublin 6 highest median
- Difference between odd (North)and even (South) districts:

μ_{odd} = €332,033

μ_{even} = €464,526

Results – clustering

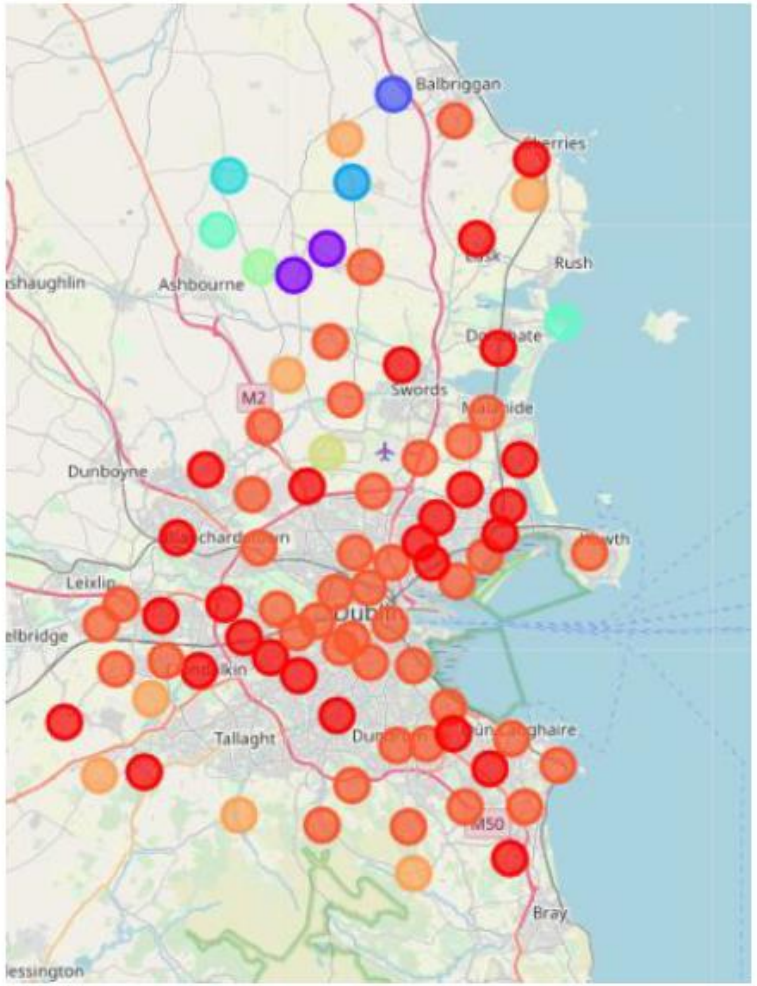


Figure 6 Clusters by color of all 83 no. Dublin parishes.

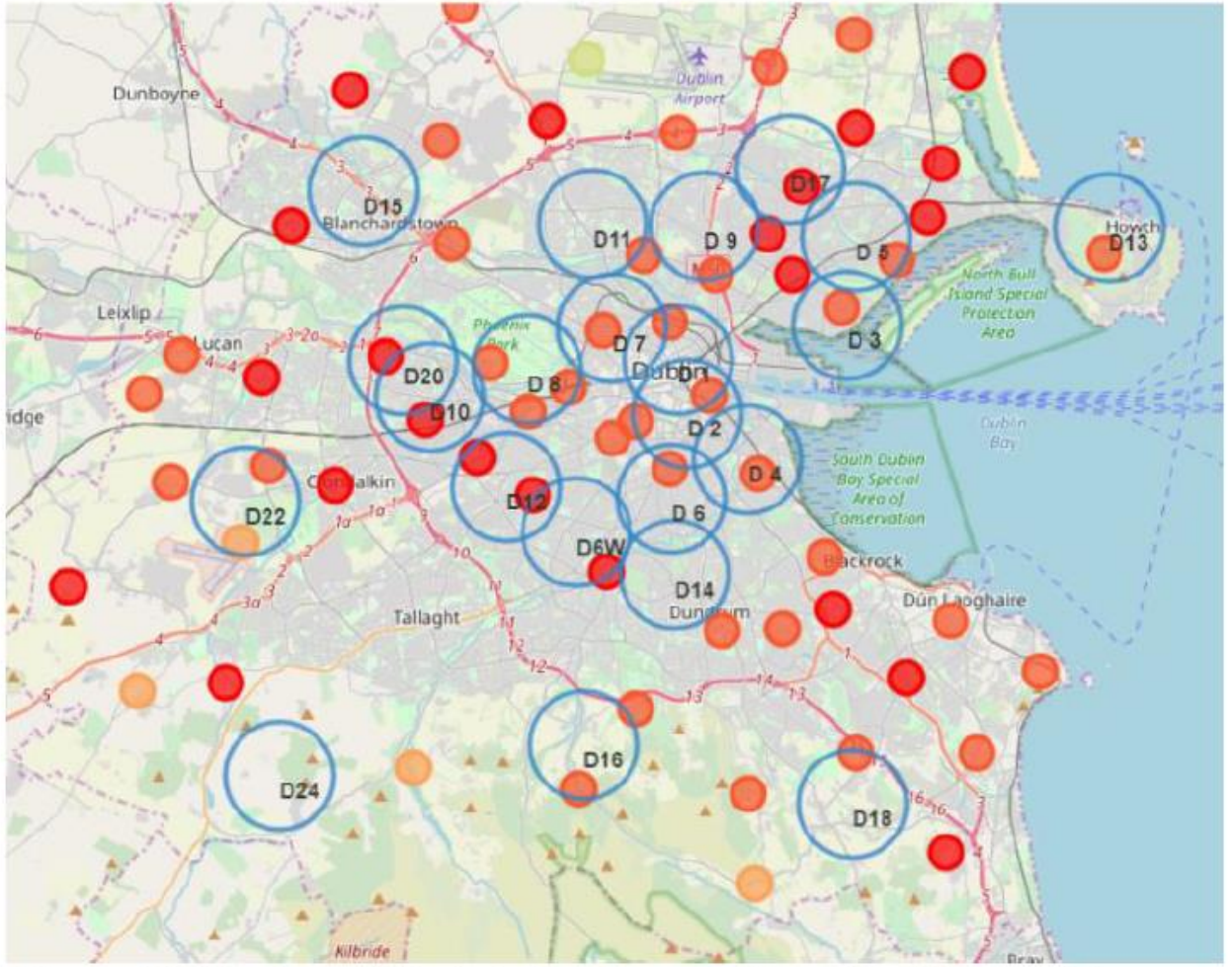
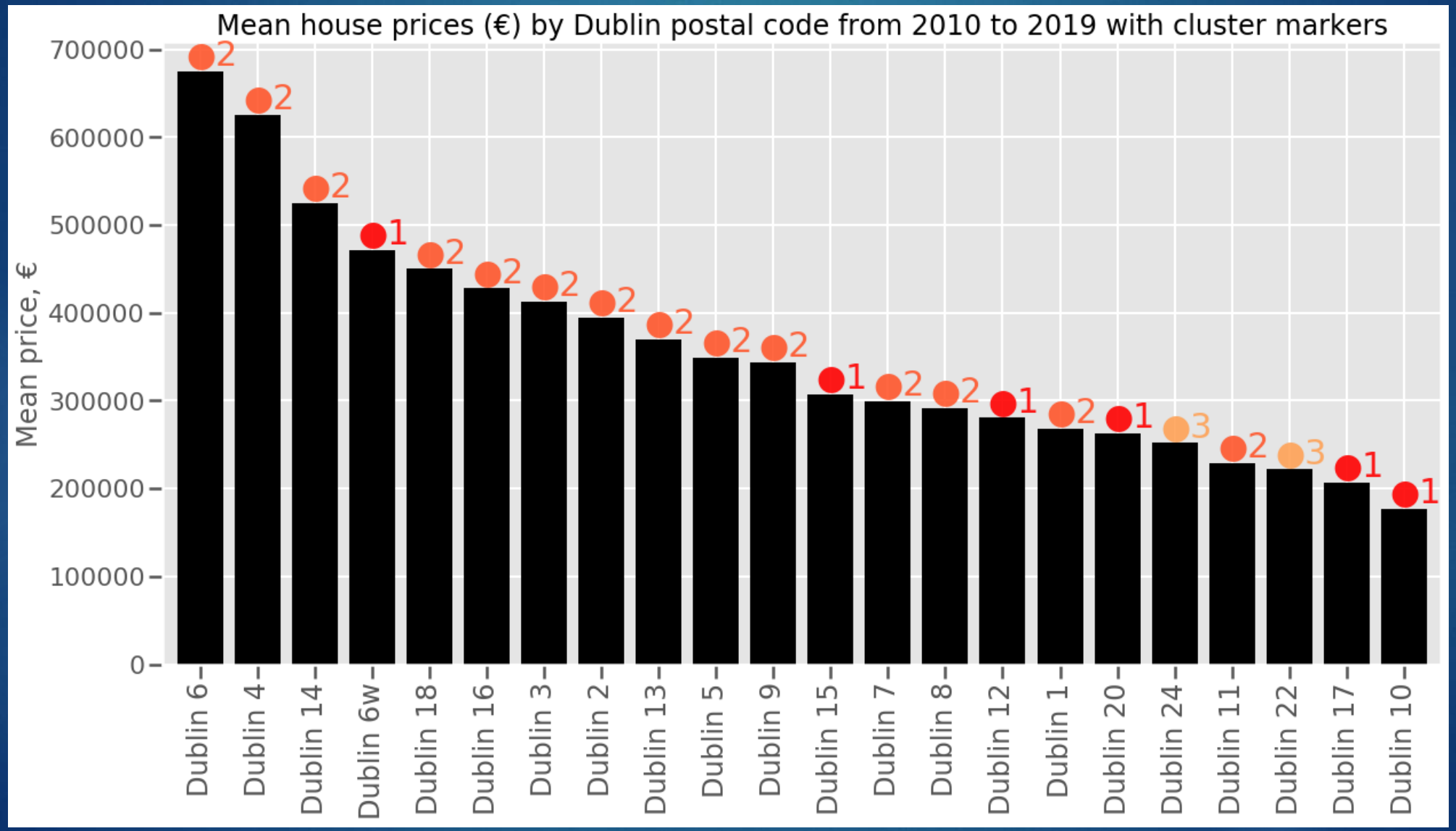


Figure 7 Clusters (filled circles) overlaid on Dublin postal code (open circles with postal code in annotation).

Results – clustering w/house prices



Discussion

- Wide spread house prices
- North better value than South
- Three clusters in postal districts
  - Red  **suburbs lower cost**, supermarkets, convenience grocery stores, pubs and restaurants
  - Orange  **urban expensive**, parks, pubs, restaurant and cafés
  - Yellow  **outdoor-type venues** near Dublin Mountains, golf courses, stables
- Value locations clear

Conclusions

- Dublin 6 is the most expensive district
- Dublin 10 is the least expensive district
- The North side is better value than South side.
- Dublin 1, Dublin 7, and Dublin 8 are good value in the cluster determined to have the highest valued houses.
- Dublin 17 and Dublin 20 are good value in the cluster determined to have the next highest valued houses.

Thank you.