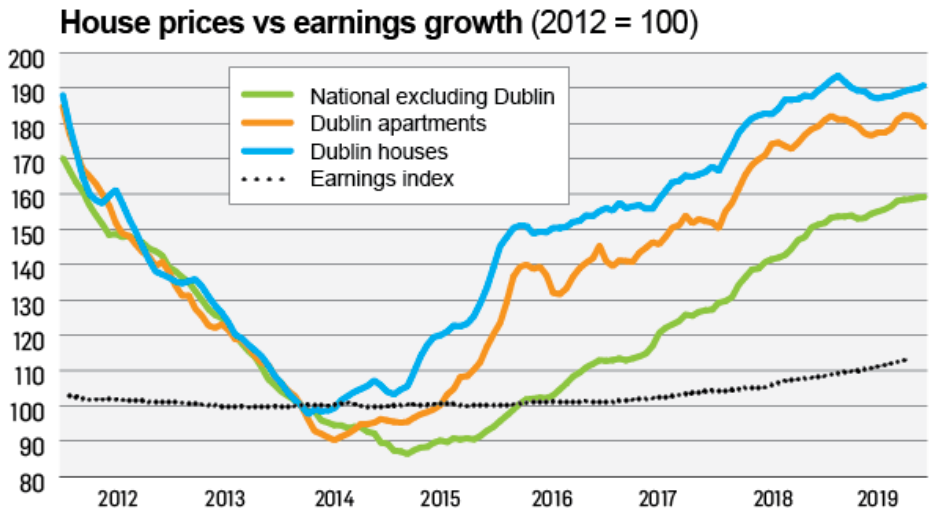
Battle of Neighbourhoods

Coursera IBM Applied Data Science Capstone

Yash Karle, March 2021

Background

- **Housing crisis in Dublin**
  - House prices in Dublin have risen by almost 90 percent, while wages have increased by only 18 percent since 2012
  - Ireland needs an estimated 30,000 new units built annually and it's been lagging behind for quite few years
- **15-minute city – impact of Covid-19**
  - Initiative which tries to mimic 15-minute city initiative pilots in major cities across the globe
  - Needs of the ever-increasing urban density, enhanced public transport and investment in public realm



Problem

- **The perfect neighbourhood - 5K lockdowns**
  - Level-5 restrictions in Ireland meant people had to restrict their movements within 5 km radius from their house.
  - A perfect neighbourhood in such scenarios would be able to satisfy all the “local” needs of the surrounding population in that neighbourhood.
  - Different neighbourhoods would have diverse demographic distribution amongst its populations.
- **Skewed distribution of amenities and opportunities**
  - Focus on smaller parts of a neighbourhood (2km, 5km radius) to see if the businesses’ and retailer’s setup as part of these local towns are enough to meet the majority needs of the immediate surrounding population.
  - Dublin (housing crisis) the construction of houses is skewed on top of this which magnifies the gap between the demand and the supply

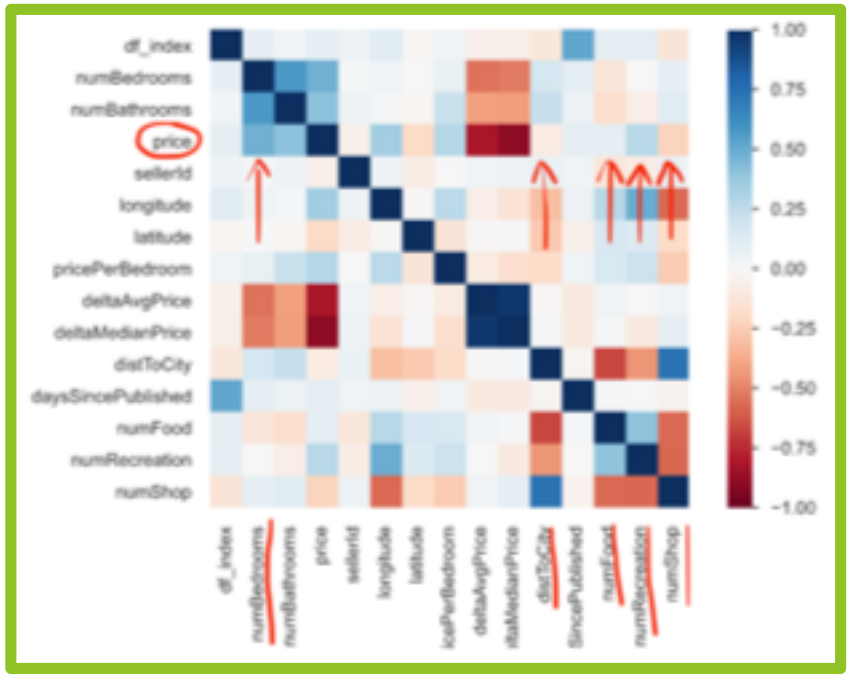
Data Sources

- **Daft-scraper API**
  - We fetch all such listings and build a dataframe containing all the useful features for each property which as seen below would consist of <price', 'facilities', 'address', 'num_bedrooms', 'num_bathrooms', 'latitude', 'longitude’>
  - This data would help us recommend properties to the targeted end-user as well as the geographical coordinates would help us visually analyse the data in question.

- **Foursquare API**
  - The challenge here is to obtain different districts comprising within Dublin City and obtain their respective geographical coordinates using Nominatim geolocator.
  - Construct the final dataframe where each row would be an individual venue along-with the attributes of each of the venues including their geolocation coordinates.

Exploratory Data Analysis

- Typical (n x n) matrix where n is the number of features.
- The intersecting blocks/cells in the matrix are represented as heat map which gives us an indication about the correlation between each pair of features.
- Values of the correlation coefficient (as r approaches +1 the pair of features are positively correlated and likewise as r approaches -1 the pair are negatively correlated)



Clustering

- **Property characteristics**: <NumBedrooms, NumBathrooms, FloorArea> [Appendix I]
  - D15, D16, D22 share same sized representation within Cluster 0
  - D9, D13 are again very similar neighbourhoods looking at Cluster 2
  - D6, D11, D12, D24 represent similar sizes in Cluster 4
- **Location characteristics**: <Longitude, Latitude>
- **Neighbourhood characteristics**: <NumFood, NumRecreation, NumShop> [Appendix II]
  - D2, D16, D18 share the same neighbourhood characteristics as seen from Cluster 0 all of which are south Dublin neighbourhoods. Similarly, D1, D5 can be seen as similar neighbourhoods which in turn are north Dublin neighbourhoods.
  - As far as the remainder of Cluster 0 is concerned, D3, D4, D6, D7, D9, D12, D14 are all fairly equal sized and stacked together. That leaves D8, D11 as the largest sized neighbourhoods within Cluster 0.

Conclusion

- We have looked at fetching the housing data using the Daft-scraper API and joined the same with the Foursquare API
- To curate some useful features which allows us to understand the interactions with price in terms of price indicators (promoters and demoters)
- To know more about the different districts within Dublin city looking from multiple different perspectives – urban development, post-pandemic town planning and uplifting small businesses
- Help new buyers understand the market triggers and plan their move into a potentially new neighbourhood.

Future Directions

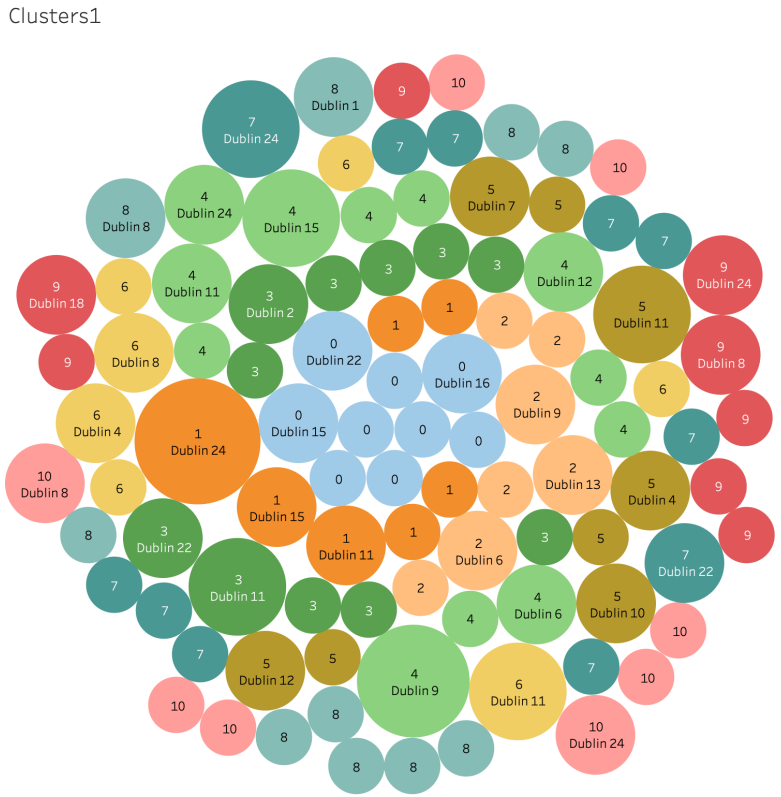
Some additional explorations possible beyond the current scope of this project:

- Fetch seller names from seller id and see if the data makes sense
- Most common & least common venues for each neighbourhood
- Removing outliers for price and then redo price deciles
- Popular transport routes, commute time to city center as an influencer on price
- Schools in the neighbourhood influencing house prices
- Crime rates in a neighbourhood and correlation with some of the other price indicators
- Pricing per sq. ft of area

References

- Irish Time Article: https://www.irishtimes.com/life-and-style/homes-and-property/ireland-s-housing-crisis-in-five-revealing-graphs-1.4150332
- Dublin Chamber of Commerce Blueprint Document: https://www.dublinchamber.ie/DublinChamberofCommerce/media/banners/Dublin_The-15-Minute-City.pdf
- Daft-scraper API: https://github.com/TheJokersThief/daft-scraper
- Dublin Postal Districts Wiki: https://en.wikipedia.org/wiki/List_of_Dublin_postal_districts

Appendix I
Clusters based on Property characteristics



Appendix II
Clusters based on Neighbourhood characteristics