**Battle of the Dublin Districts**

Yash Karle

**1. Introduction**

**1.1 Background**

* Housing crisis in Dublin
* 15-minute cities – impact of Covid-19
* Reimaging neighbourhoods

**1.2 Problem**

* The perfect house
* 5K lockdowns
* Skewed distribution of amenities and opportunities

**1.3 Interest**

* Find the right house for me!
* Grow Remote
* Hyper-personalisation

**2. Data acquisition and cleaning**

**2.1 Data sources**

* Daft-scraper API
* Foursquare API

**2.2 Data cleaning**

* Feature Extraction (Neighbourhood, SellerId, FloorArea, Longitude, Latitude, BerRating)
* Missing data (Price, NumBedrooms, NumBathrooms, BerRating)

**2.3 Feature selection**

* Feature Synthesis (PricePerBedroom, DeltaAvgPrice, DeltaMedianPrice, NorthSouthFlag, DistanceToCity, DaysSincePublished, NumFood, NumRecreation, NumShop)
* Aggressive filtering (Zero Price, Nulls for newly added features)

**3. Exploratory Data Analysis**

**3.1 Relationship between Price and SellerId**

**3.2 Relationship between Price and NumBedrooms**

**3.3 Relationship between Price and FloorArea**

**3.4 Relationship between Price and DistanceToCity**

**3.5 Relationship between Price and NumFood**

**3.6 Relationship between Price and NumRecreation**

**3.7 Relationship between Price and NumShop**

**4 Clustering**

**4.1 DBSCAN clustering**

* Property characteristics [NumBedrooms, NumBathrooms, FloorArea, PropertyType, Rating]
* Location characteristics [Longitude, Latitude, DistanceToCity]
* Neighbourhood characteristics [NumFood, NumRecreation, NumShop]

**4.2 Solution to the problem**

**5. Conclusion**

**6. Future directions**