# Analysis of Factors Affecting House Prices: Case of Dublin, Ireland

Akhileshkumar M Subbarao 19210140
School of Computing, Dublin City University, Ireland
akhileshkumar.mutuguppesubbarao2@mail.dcu.ie

Kadesh Basavaraj Huddar 19210323
School of Computing, Dublin City University, Ireland
kadesh.basavarajhuddar3@mail.dcu.ie

***Abstract*—In the age of rapid motorisation, walking, as both a green mode of travel and the most healthy method has recently been promoted as a specific form of everyday physical activity in Ireland. The term "walkability" explains to what degree an area or a neighbourhood is walking-friendly. Recently, rising inflation of home purchases and rentals have been seen as a possible catalyst for more greener ways of commuting. In this paper, we examine the factors that influence house prices in the city of Dublin. We propose a method to define a consistent spatial trend of spatial heterogeneity within the city in terms of walkability between amenities. Our experimental findings indicate the division of the rating for residential walkability in Dublin varied greatly. We can see that areas with more amenities tend to have greater infrastructure and that there are clusters with high walkability scores in the city centre. We can conclude from our study that walkability societies have typically concentrated rich amenities, lower crime rates and less population and thus will have higher values for home properties.The study also includes model developed for determining optimal store placement which helps to find the best place to open a new retail store.**

***Keywords*—*Walkability, Amenities, Clustering, Crimes, KNN, Haversine, Distance Matrix***

## I. Introduction

Dublin City's housing market has hit high rates in the last few years. The average house price for sale, for example, is €388,000 and for rent is €2,044 in the Dublin area, which is obviously beyond affordability for most families. In addition to rapid urbanisation and significant population migration that triggered the property bubble, determinants that led to spatial variance in home prices could be influenced by the range and number of amenities in the vicinity of the neighbourhoods, such as connections to bus stops, train stations, fitness centres, hospitals and schools. This is based on an article published in the Irish Times in June 2019 [1].

The advantages of ease and accessibility to various kinds of amenities affect the house prices accordingly. More price would be paid for properties either to rent or purchase in areas with amenities by those who wish to seek a high-quality living environment. Whilst housing prices differ based on convenience proximity, the values of those amenities can not be quantified easily. Implementing and developing the Green transit project is an important way to address these issues for it helps to give us a green and low-carbon travel environment.

[1]https://www.thejournal.ie/house-prices-increase-dublin-nationwide-\ /4696800-Jun2019/

Walking plays a significant role in supporting the greening of our environment. Furthermore, walking as the primary form of physical activity has many health benefits for people of all ages. It is considered the most common type of day-to-day physical activity and considered safe [3].

This paper considers a large number of properties in the Dublin city area, examines a range of amenities surrounding the properties, includes the crimes reported in the local Garda districts, the distance of each property from city centre, and we combine all this information to provide a score for each property within the 24 Dublin postal districts. Based on the resulting property scores and the property values in each area, we analyse how property pricing varies with these external factors. The paper also includes data-driven analysis of where to optimally place the next retail store in such a way so as to maximise a reduction in walkability for the greatest number of properties. In present Ireland situation, sustainability and social issues like walkability are the mainstream agenda items for government to move society towards a greener lifestyle and the work reported here could help inform how that greener lifestyle can be delivered.

## II. Literature Review

Local environment elements that correlate with walkability include neighbourhoods amenities. Public facilities are the services or amenities that local people use for various uses including food retailing, transport, education, leisure, social and cultural events and health [2].

In 2013 Emily Washington conducted a study of walkability in Washington D.C. [13]. It looks at the relationship between a city's Walk Score and its median housing prices when adjusting for economic opportunities, and it offers a first look at the relationship between home prices and Walk Scores. The author also says that residents in U.S. cities are willing to pay more in order to live in areas with better walkable amenities. House values are generally expensive in highly walkable areas when compared to less walkable neighbourhoods [13].

Azmi and Karim have published a study which suggests that the availability of community amenities impacts a neighbourhood's walking behaviour [1]. Their analysis shows that factors influencing walking behaviours are the size of area, catchment area, a pattern of distance, location, accessibility, density and land use [1].

Feuillet *et al.* talk about the socio-ecological model for health-related behaviours [3], and they discuss how physical activity is affected by environmental factors. Their paper describes the relationship between active commuting like walking and cycling to/from work and residential environments.Their method involved 4,164 adults who are residents of Paris, France and who were studied using a weighted Poisson regression model [3].

Research by Hao *et al.* [5] suggests that housing prices are strongly associated with transportation accessibility, which is more important while ignoring other variables, and improving the road network would increase various effects on transportation accessibility and housing prices. In order to encourage stability of housing prices, attention should also be paid to enhancing the closeness and straightness of residential land next to the road line.

Song and Knaap [10] have found that constant business activities of business circles will promote ease and quality of life for residents. Hence, they have a positive effect on the prices of neighbouring properties.

Erik *et al.* in [11] suggest that the spatial sensitivity of the geographical accessibility measures depends much on the nature and location of the region. The authors' findings were considerably affected by the existence of areas that provides several amenities. Their study uses the accessibility indicator to research the correlation between the price of housing and the area accessibility. Easy accessibility to public service plays a significant role in the quality of a land transaction.

In recent work, Xia and colleagues describe that on average, overall home values decreased by 0.5% if the distance from a convenience store increased by as little as 200m [14]. Hence with rising distance from the nearest grocery store, housing prices will decrease. Much-needed facilities like primary and elementary schools, hospitals, restaurants may have a bigger effect on the price of housing.

John *et al.* examined 170 neighbourhoods in a city to see how walkability impacts efficiency and sustainability in an area. They developed models that assessed the connection between the walkability in an area and three specific urban sustainability measures: assessment of public housing; foreclosures; and crime. Here their research shows a positive impact not only on the price of housing in the neighbourhood but also on crime [4].

Authors here have identified important factors for restaurants site selection. Visibility, return of investment and residential population are the three most important factors for selecting a place to open a new restaurant. Delphi technique was used to come up with the selection [7]. Feng *et al.* have considered online reviews to study about the optimal placement of restaurant. Features like a review-based market attractiveness features, market competitiveness features, and geographic features are researched to predict the number of visits for a restaurant in a particular location [12]. With the use of urban planning, authors have come up with a score impacting store placement. Scores are calculated through Foursquare metrics [9].

In summary, we see there is a good amount of published literature going back in once case, throughout the 2000s and 2010s. The cities examined are global in terms of geography, covering the U.S. to China, and the set of city amenities which were considered varied across each published case. In the work, we focus on Dublin city, Ireland, and we include a very wide range of amenities, wider than in all of the previously published examples. We measure the expense of housing properties covering both the purchase price as well as the cost of renting. Almost all of the data we will use in our analysis is scraped or automatically downloaded from publicly available resources on the internet.

In the next section, we present the data sources for amenities and properties and the pre-processing we performed.

## III. Data Sources and Pre-processing

In order to gather data for our work, we used properties for sale and properties for rent in the Dublin city area but our work could be applied to any city where the following data sources are available.

**Property Data (Rent and Sale):** The data on properties for rent and for sale was scraped from daft.ie (a rental market site) available from 12/06/2019 until 31/05/2020 using the python library 'beautiful soup'[2]. Daft.ie[3] is Ireland's No.1 property website and app which provides low-cost online advertising to Irish estate agencies, landlords, homeowners and tenants since 1997. It is a nationwide website and has property available in every county. Daft.ie delivers 220 million pages of properties to over 2.19 million visitors each month and every minute over 1,000 searches are carried out on the site.

Each property that was scraped contains attributes like location, area code, bedroom, bathroom, longitude, latitude, etc. Coordinates are fetched from the script tag and are used to display the map location of the property. The data includes different types of properties like apartments, houses, flats and studio apartments. The number of property records for rent that we gathered is 1,834 and for sale is 2,370.

Data collected is drawn from Dublin postal codes D01 to D24. Postal codes for each area are extracted to place the properties accordingly. Attributes of sample properties are shown in Figure 1.

**Amenities Data:** Amenities for postal districts D01 to D24 were extracted from management-ware Google Maps Contact Extractor V2.0 using the Google API. The software is used to scrape Google maps website contacts. A total of 6,782 unique amenities spanning over 18 categories were scraped. These amenities were chosen based on the survey conducted on 2,000 people [14]. It includes important amenities for people on a day to day basis. Amenities tend to be more around the city centre and decrease gradually towards the city's outskirts. This shows that area tends to be popular with better infrastructure.

The 18 categories of amenities extracted along with the number of each amenity that we scraped are as follows :

- Banks : 160
- Bus Stops : 2,383

[2]https://pypi.org/project/beautifulsoup4/
[3]https://www.daft.ie/

| Location | Area code | Price(in Euros) | House_type | Beds | Bathroom | latitude | longitude |
|---|---|---|---|---|---|---|---|
| 1 Kirkpatrick House\|\| Spencer House\|\| IFSC\|\| Dublin 1 | D01 | 285000 | Apartment for sale | 1 | 1 | 53.35008 | -6.23869 |
| 17 Albert Place East\|\| Grand Canal Street Lower\|\| Dublin 2 | D02 | 850000 | Terraced House | 3 | 3 | 53.33944 | -6.24276 |
| 105 Shanard Road\|\| Santry\|\| Dublin 9 | D09 | 325000 | Semi-Detached House | 3 | 1 | 53.39128 | -6.26083 |
| 12a Maplewood Park\|\| Tallaght\|\| Dublin 24 | D24 | 295000 | Terraced House | 3 | 3 | 53.2859 | -6.38336 |
| Parnell Square West\|\| Dublin 1 | D01 | 2350 | Apartment to Rent | 2 | 2 | 53.35207 | -6.26435 |
| bantry square\|\| waterville\|\| Blanchardstown\|\| Dublin 15 | D15 | 1800 | Apartment to Rent | 2 | 1 | 53.39323 | -6.37147 |

Figure 1: Attributes of sample properties

- Cafes and Pubs : 616
- Colleges and Universities : 158
- Fire Stations : 14
- Garda Stations : 31
- General Practitioners : 167
- Gym, Swimming pool, Yoga : 380
- Hospitals : 255
- Hotels and Restaurants : 632
- Libraries : 87
- Pharmacies : 331
- Post Offices : 135
- Primary and Secondary Schools : 481
- Religious Activities : 85
- Retail Stores : 464
- Sports Facilities : 311
- Train and Tram Stations : 92

Importance of weights are calculated from the survey conducted [14]. These weights reflect the relative importance of each amenity. Weights are assigned in the range from 0 to 1, as shown in Figure 2.

| Amenity Type | Weight |
|---|---|
| Gym, Swimming pool, Yoga | 0.15 |
| Library, Pharmacy | 0.2 |
| General Practitioner | 0.35 |
| Fire Station, Garda, Religious Activity, Sports Facility | 0.4 |
| Bank, Hospital, Post Office | 0.5 |
| Retail Store | 0.75 |
| Cafe and Pub, Hotel and Restaurant | 0.8 |
| Bus Stop, Train and Tram Station, College and University, Primary and Secondary School | 1 |

Figure 2: Amenities and their weights for walkability assessment in Dublin

**Crime Data:** This data covers crimes reported at all Garda (police) stations from 2007 to 2018. There are a total of 12 categories of crimes reported which are further merged into three types based on severity as severe, moderate and minor. Each crime reported over the 12 years is added to compute a single entity. We use this historical archive of crime data to indicate the safety of each area broken down as follows:

- Severe
  * Attempts or threats to murder, assaults, harassment and related offences
  * Kidnapping and related offences
  * Robbery, extortion and hijacking offences
- Moderate
  * Burglary and related offences
  * Fraud, deception and related offences
  * Dangerous or negligent acts
  * Damage to property and to the environment
  * Weapons and Explosives Offences
- Minor
  * Theft and related offences
  * Controlled drug offences
  * Public order and other social code offences
  * Offences against government, justice procedures and organisation of crime

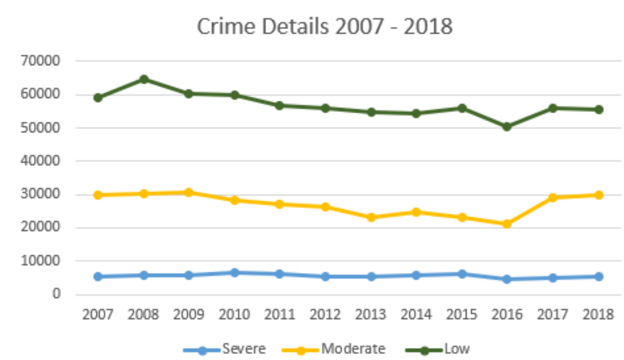



Figure 3: Crime details 2007 - 2018

The crime details spanning over a period of 12 years is shown in Figure 3.

**Population and Density Data:** The population of each Dublin postal district was collected from www.citypopulation.de/en/ireland/dublin/ which was gathered from the 2016 census data. This website had the population and density data for each electoral division. There are a total of 259 electoral divisions in Dublin. Since we are interested in population and density of postal districts, each electoral divisions are carefully mapped manually using the postal boundaries provided by Dublin Housing Observatory.

The electoral divisions shown in Figure 4 and 5 are within the boundary of D07 postal district. Hence overall density and population of an area will be the summation of all the electoral division in that area. We chose to calculate and use the density of people per $km^2$ and associate this with each area because the sizes of the postal districts are not uniform and thus population density is a better indication than the raw population count.

| Name | Status | County | Populati | Area_Co | Density kr |
|---|---|---|---|---|---|
| Arran Quay A | Electoral Division | Dublin City | 1,785 | D07 | 11,647 |
| Arran Quay B | Electoral Division | Dublin City | 4,166 | D07 | 5,486 |
| Arran Quay C | Electoral Division | Dublin City | 4,471 | D07 | 12,384 |
| Arran Quay D | Electoral Division | Dublin City | 3,109 | D07 | 9,004 |
| Arran Quay E | Electoral Division | Dublin City | 3,293 | D07 | 12,449 |
| Cabra East A | Electoral Division | Dublin City | 5,650 | D07 | 4,010 |
| Cabra East B | Electoral Division | Dublin City | 3,737 | D07 | 8,017 |
| Cabra East C | Electoral Division | Dublin City | 4,085 | D07 | 8,944 |
| Cabra West A | Electoral Division | Dublin City | 1,478 | D07 | 3,309 |
| Cabra West B | Electoral Division | Dublin City | 2,577 | D07 | 5,683 |
| Cabra West C | Electoral Division | Dublin City | 2,953 | D07 | 5,146 |
| Cabra West D | Electoral Division | Dublin City | 2,845 | D07 | 4,735 |
| Inns Quay A | Electoral Division | Dublin City | 3,919 | D07 | 12,346 |
| Inns Quay B | Electoral Division | Dublin City | 3,666 | D07 | 13,285 |
| Inns Quay C | Electoral Division | Dublin City | 2,757 | D07 | 9,991 |

Figure 4: Electoral divisions

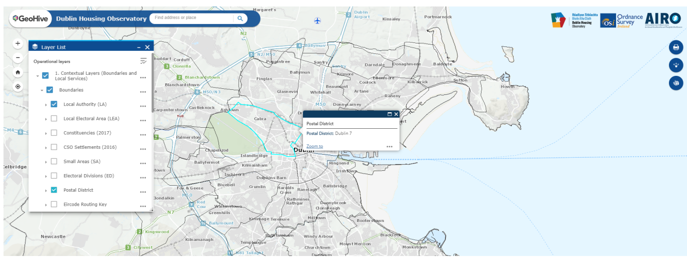



Figure 5: Postal code boundaries

## IV. Implementation and Analysis

Our data processing requires a number of steps, and we describe these steps in the following subsections.

### A. Step 01. Distance calculation from the property:

• **Calculating straight line using Euclidean distance:** Haversine formula is used to compute the distance between properties and amenities [8]. Amenities address is for destination and we use the property as a origin. We measure the shortest path between each origin and destination as a Euclidean distance. Distance less than 1 km are considered as walkable, so amenities within this radius are used to calculate distance matrix . The euclidean distance on a sphere is computed using the haversine formula which can be written as the following trigonometric equation : haversine($\theta$) = sin2($\theta$=2)

$$dist[(x,y),(a,b)] = \sqrt{(x-a)^2+(y-b)^2} \text{ is}$$

$$2R\sin^{-1}\left(\sqrt{\sin^2(\frac{\phi_2-\phi_1}{2})+\cos(\phi_1)\cos(\phi_2)\sin^2(\frac{\lambda_2-\lambda_1}{2})}\right)$$

where R is radius of the earth (6371 km), dist[(x; y); (a; b)] is the distance between two points $\phi1$; $\phi2$ is the latitude of two points and $\lambda1$; $\lambda2$ is the longitude of two points respectively [6].

• **Calculating walkable distance using Google Map API:** Google Distance matrix is an efficient method to calculate the distance between two points. It returns actual traversing distance from source to destination.HTTP request is made to google maps using google map API keys. Parameters like source coordinates and destination coordinates along with mode, transit routing preference are sent to the distance matrix. The mode indicates the type of travel like walking, bicycling, etc. and transit routing preference is less walking as it indicates the google map distance matrix to return a shortest distance between the two points.

• **Amenities within 1 km footpath distance parsing:** Using Google Cloud Platform distance matrix API and haversine formula, a walkable distance less than 1km is computed between origin and destination. Since a large number of property and amenities are present in the dataset, API calls to each property to each amenity is not reliable, as API calls to be made will exceed more than 6 million requests. Hence haversine distance from each property to all the amenities are calculated first. If any amenities lies within 1km radius, then only that property coordinates are sent to distance matrix API. This drastically reduces the number of API calls needed and make calculations reliable. Once the distance matrix is computed, JSON files are returned as output, distance in JSON data is stripped from the output to get the distance between amenity and property. When calculating the distance to the nearest pubs and restaurants, these should not be very close to the property; hence only restaurants which are more than 500 meters and less than 1km are considered.

{'destination_addresses': ['76 White Oaks, Friarland, Dublin, Ireland'], 'origin_addresses': ['UCD Glenomena Residences, Belfie
ld, Dublin, Ireland'], 'rows': [{'elements': [{'distance': {'text': '1.9 km', 'value': 1947}, 'duration': {'text': '24 mins',
'value': 1449}, 'status': 'OK'}]}], 'status': 'OK'}
1.9 km

Figure 6: API output request from JSON

### B. Step 02. Score calculation for property:

• **Amenity Score**: Data collected from google map API for different categories are combined to assign score for each property. Based on total amenities surrounding property, final amenity score is formulated by adding all the corresponding weights of each amenities. For example : If a property has one bus stop and one hospital within 1 km walkable distance, the score is calculated by adding the weight 1 to the bus stop and 0.5 to hospital respectively. The total score for the property will be 1.5.

### C. Step 03. Visualisation of amenity spread:

• **Illustration in Tableau**: Tableau is an efficient visualisation tool, but it is very important that each data point and data components are defined clearly. Figures 7 and 8 has two layers:

1) **Dublin postal code boundaries**: It is important to define boundaries of the city so that it helps to identify which areas have more important amenities and walkability. This is a spatial dataset which is obtained from autoaddress [4]Irish grid website . It contains the boundary values of all postal code throughout Ireland.
2) **Property point and circle**: Each data point represents all the property coordinates. A circle represents the sum

[4]https://www.autoaddress.ie/support/developer-centre/resources/routing-key-boundaries

of weighted score of surrounding amenities. Circle reacts to the score quantification, and circle size varies as the amenity count increases.

- **Weighted amenity plot**: The main aim of this plot is to understand how different types of amenities are spread around Dublin, this helps us to address the area with better infrastructure which can affect the price of houses. Each point comprises of coordinates, activity label, score, importance weight. Amenities are broadly classified into eight categories, namely Transport, Social and Culture, Safety, Leisure, Education, Medical, Retail and Financial. Each category consists of subcategories as shown in Figures 7 and 8. The importance weight constitutes the total score for each property. Increased importance weighted amenity increases the total score for the property, whereas less important amenity score constitutes for less variation in property score.

From Figures 7 and 8 we can see that Retail and Transport have a better spread across Dublin city whereas Bank, Garda and Fire station are limited. It is evident that all the amenities are located at the heart of the city. Areas like Dublin 1, Dublin 2 which are close to the city centre have a high score as well as a large number of amenities. It clearly shows that urban developments are mainly concentrated at the core of the city. Infrastructure and walkability decrease as you move away from the heart of the city. Areas like Blanchardstown, Finglas, Tallaght have fewer numbers of amenities and lower walkability scores which suggest that the loosely defined city infrastructure in outskirts of Dublin. It shows the decrease in the property counts in outer parts of the city.

### *D. Step 04. Computing crime scores:*

Crimes in Dublin are categorised based on their severity. There are 31 Garda divisions which are examined for this study. We use the nearest Garda station present in the area in which the property is located. If there is more than one Garda station surrounding the property, then the Garda Jurisdiction under which the property falls is considered for calculating the property's score related to crimes.

Crime calculation is further carried out based on population density and level of severity committed. We have classified areas as low, medium and high density based on the Inter-Quartile Range. Less than 25% is low, more than 75% is high, and the rest is medium. The same is done to group the three (low, medium, high) subcategories of crimes based on crime rate quartiles, comparing along with severe, medium and minor crime categories. The overall score for crime listing equals 1.

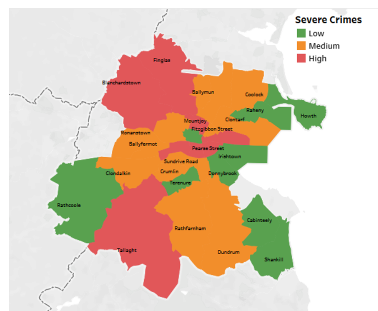
A decision matrix of 9*3 is prepared from below criteria and problem tables to get crime score for each property. Combining all the properties in a particular area gives crime score of that area. Low density, along with low crime is the best area to live as it is less populated and has lower crime rate, so this category gets the highest score. Low density and high crime areas are more dangerous to live in as there is more risk of things going wrong within a less populated zone. High density and high crime grouping gets an average score since the soaring crime tends to be greater in populated areas.

| **Criteria** | Low Crime | Moderate Crime | Severe Crime |
|---|---|---|---|
| | 0.2 | 0.3 | 0.5 |

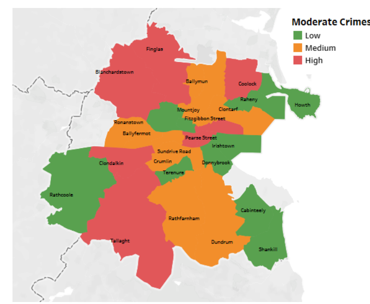
| **Problem** | Low Crime | Medium Crime | High Crime |
|---|---|---|---|
| Low Dense | 0.5 | 0.075 | -0.075 |
| Medium Dense | 0.3 | 0.045 | -0.045 |
| High Dense | 0.2 | 0.03 | -0.03 |

### *E. Step 05. Visualising crime scores:*

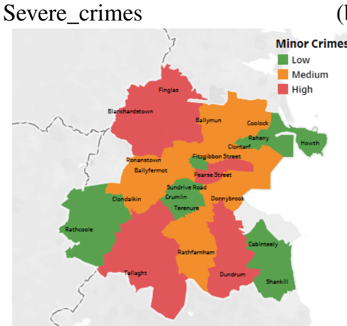
Garda jurisdiction boundaries are taken from CSO 2011 Garda sub-district data. Each of the sub-divisions is merged with jurisdiction boundaries. Severe, Moderate, Minor crime in each Garda sub-division are shown in Figure 9. Low crime is marked with green, followed by orange for medium and red for high crime areas.




(a) Severe_crimes




(b) Moderate_crimes



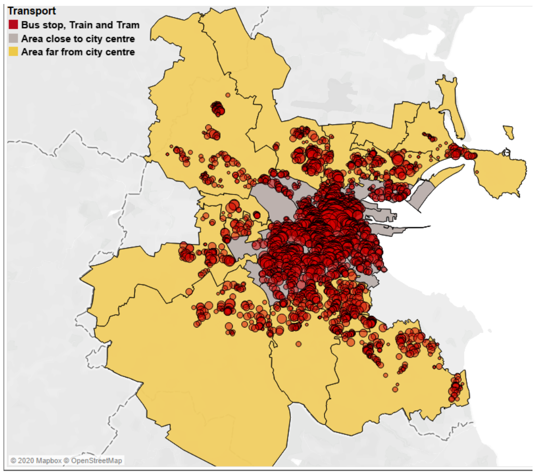

(c) minor_crimes

Figure 9: Crime in Dublin garda sub-divisions

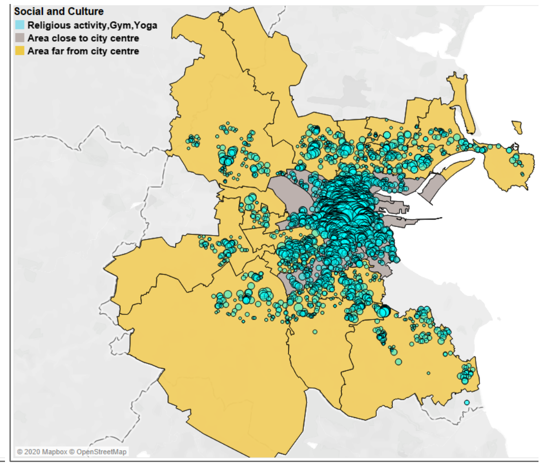
### *F. Step 06. Distance to city centre calculation:*

Each property distance from the Spire (public monument regarded by many as being the centre of Dublin) is computed to get the distance from the city centre. Properties near the city centre have an added advantage over those which are far away. For every km distance, 0.34 times the distance of the property from the city centre, is reduced [6]. Areas which are isolated tend to get lower scores when compared to those near the centre.

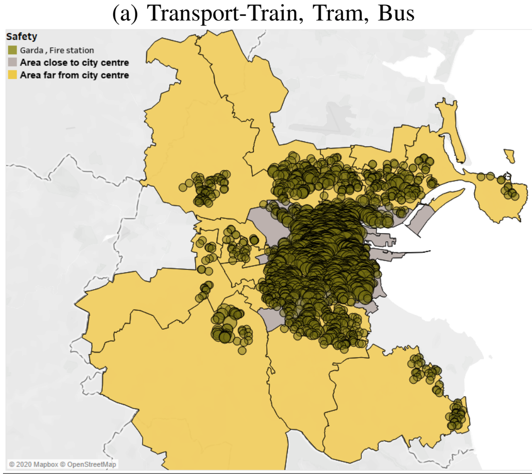
For example, if the score of property is 20 and its distance from the spire is 5 km, then 0.34*5 = 1.7 is subtracted from 20, which lowers the score to 18.3 .



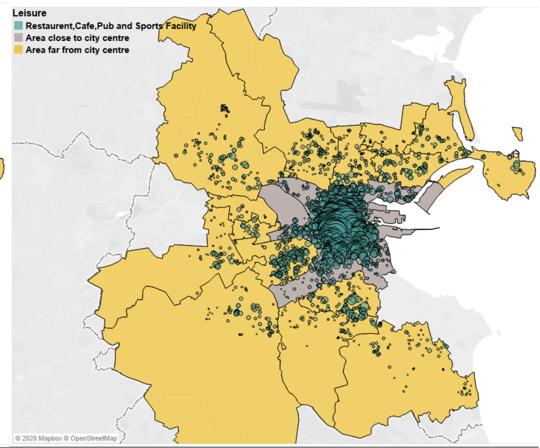

(a) Transport-Train, Tram, Bus




(b) Social-Religious, Gym, Yoga




(c) Safety-Garda, Fire station




(d) Leisure-Restaurant, Cafe, Pub, Sports facility

Figure 7: Amenity spread around properties for sale (part 1)

### *G. Step 07. Overall score computation:*

For each property taken into consideration, the overall score is computed by combining the scores obtained by amenity weight, the outcome of the crimes in the area and distance from the city centre. The experimental results we achieved are based on quantifying the relative importance of proximity factors which influence the price of a house or apartment in Dublin, for sale or rent.

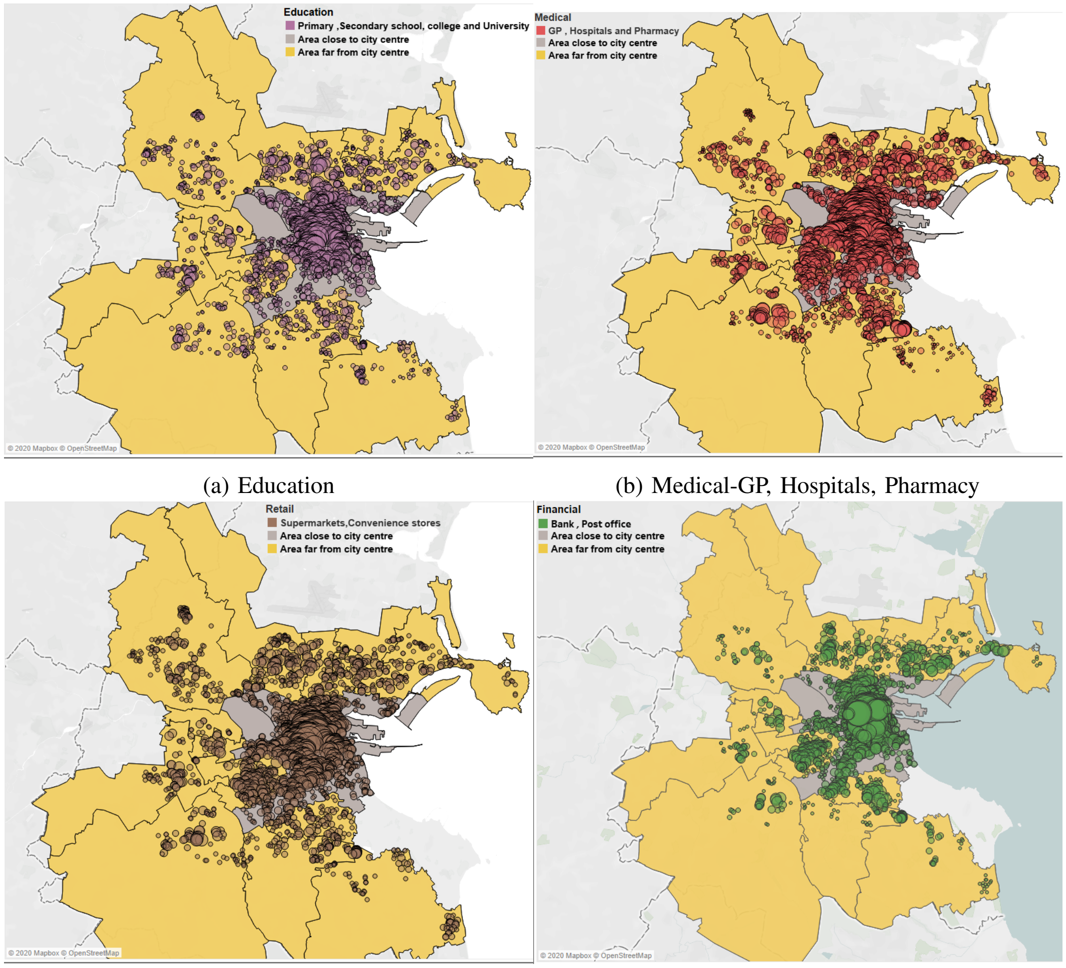
The main part of the analysis focused on how important is the distance to the nearest shop, school, bus top, DART station, etc. along with crimes in an area and then distance to the city centre in determining property value. House types include "Apartment to rent/sale" having two bedrooms and one bathroom. This combination is analysed because in every area, we can find these properties so as to get a granularity in property type. Our analysis would be hampered if were to consider a range of different house types like bungalows, studio apartments, etc. as we cannot find the same type in every area, which would limit our analysis.

## V. Initial Results

We now present the results of our rent and sale analysis before looking at the optimal placement of new amenities.

[•] **Rent Analysis** : From Figures 10 and 11, we can say that there is not much difference in rent price in all areas across Dublin. This is because landlords are less concerned about looking at property neighbourhoods when they fix a price for a property and a proprietor naturally tends to over-price a property. The score obtained in each area varies considerably according to the range of factors we have taken into account. Top 2 area with the highest rents are Dublin 02 and Dublin 01, and lowest are Dublin 11 and Dublin 22.

[•] **Sale Analysis** : Purchasing a house is one of the biggest financial decisions you'll make in your life. When people decide to buy a house, they should consider things that will affect their current and future financial state and style of living. A great location will remain an asset no matter how future fluctuations in the real estate market may occur. Thus it is always great idea to invest in a property located in good area.




(a) Education (b) Medical-GP, Hospitals, Pharmacy

(c) Retail (d) Financial-Bank, Post office

Figure 8: Amenity spread around the properties for sale (part 2)

In our research into sale prices, the price v/s score in all areas are correlated in most cases. If a property is advertised for sale, its price is affected by the factors surrounding it. From Figures 12 and 13, we can see that as the score for an area decreases, the price decreases as well. This tells us that vendors are more careful when tagging a price for their property. People look at the neighbourhood, the amenities as well as the crime in that particular area. Top 2 areas with highest relative sale prices are Dublin 04 and Dublin 02 whereas the bottom 2 are Dublin 24 and Dublin 22.

For renting, external factors are less when we move away from the city centre, but the disparities in rent is less. This results in random variations. In the case of sales, as we move far away from the city centre, the infrastructure decreases so does the price. External factors are more carefully taken into consideration in the sale of properties than renting to quote a price. However there are some exceptional cases like Dublin 18, where the infrastructure is less but the price is high, and such cases provide some insights into where many other external factors that affect the prices, but in this case, it might be the IT park in the location. Overall, we can clearly see variations of house prices with external factors.

## VI. Optimal amenity placement quantification

From our previous analysis of property sales data, there is careful consideration of amenities in tagging a price for a property. Hence we use our sales walkability data for further analysis. In the retail field, the selection of a suitable location for opening a new store is an important activity when entering a market that has not been explored previously. Launching of retail stores near the city centre or the outskirts of the city can contribute to a tremendous difference in business income, which is strongly linked to business performance.

In this paper, we have developed a system that helps to pick the best retail store location using the geographical data and predictive model. Based on the weights given to amenities in [14], retail stores are one of the most important amenities to have an impact on the overall population in a location on day to day basis.

### *A. Step 01. K-Means clustering of property sales data:*

K-means clustering is a popular unsupervised machine learning technique for cluster analysis. The method partitions

| Area_code | House_type | Price | score |
|---|---|---|---|
| D01 | Apartment to Rent | 2084.846154 | 30.637626 |
| D02 | Apartment to Rent | 2205.549296 | 37.547698 |
| D03 | Apartment to Rent | 1929.923077 | 21.765751 |
| D04 | Apartment to Rent | 2076.447368 | 25.479476 |
| D05 | Apartment to Rent | 1800.000000 | 17.206096 |
| D06 | Apartment to Rent | 1922.800000 | 21.059428 |
| D07 | Apartment to Rent | 1837.714286 | 25.351513 |
| D08 | Apartment to Rent | 1920.272727 | 26.400441 |
| D09 | Apartment to Rent | 1716.111111 | 17.676459 |
| D10 | Apartment to Rent | 1900.000000 | 19.904028 |
| D11 | Apartment to Rent | 1560.000000 | 13.362378 |
| D12 | Apartment to Rent | 1734.000000 | 18.003819 |
| D13 | Apartment to Rent | 1876.666667 | 16.508844 |
| D14 | Apartment to Rent | 1923.750000 | 18.725914 |
| D15 | Apartment to Rent | 1813.571429 | 16.109359 |
| D18 | Apartment to Rent | 2036.666667 | 19.527089 |
| D22 | Apartment to Rent | 1668.750000 | 14.685436 |
| D24 | Apartment to Rent | 1792.250000 | 14.753665 |

Figure 10: Rent price with area score

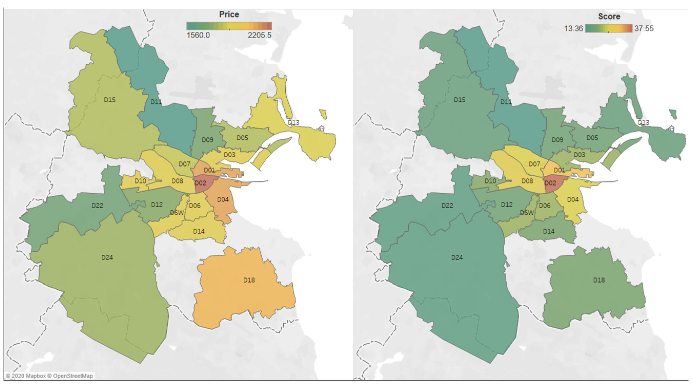


Figure 11: Rent price v/s Score comparison

observations into k clusters, where each observation is in a particular cluster based on the nearest mean. Properties with more than 100 surrounding amenities are excluded from our computation as we are interested in placing a new amenity in an area having no retail store surrounding the properties. In the next step, one-hot encoding is used to identify the total count of retail stores surrounding each property. Using the MinMaxscalar method data is normalised before applying clustering.

The well-known elbow method is used to identify the number of clusters based on the assumption that the optimal number of clusters must produce small inertia, or total intra-cluster variation. As such, there will be a trade-off between the inertia and the number of clusters. With this concept, we should choose a number of clusters so that adding another

| Area_code | House_type | Price | score |
|---|---|---|---|
| D01 | Apartment for sale | 302860.344828 | 42.130091 |
| D02 | Apartment for sale | 383636.363636 | 42.492887 |
| D03 | Apartment for sale | 309166.666667 | 31.656979 |
| D04 | Apartment for sale | 419197.500000 | 39.477391 |
| D05 | Apartment for sale | 273325.000000 | 26.738575 |
| D06 | Apartment for sale | 335000.000000 | 33.471093 |
| D07 | Apartment for sale | 282222.222222 | 33.708081 |
| D08 | Apartment for sale | 289045.454545 | 37.120041 |
| D09 | Apartment for sale | 243118.750000 | 22.407288 |
| D11 | Apartment for sale | 221264.285714 | 19.173052 |
| D12 | Apartment for sale | 215000.000000 | 19.587278 |
| D13 | Apartment for sale | 239987.500000 | 21.700781 |
| D14 | Apartment for sale | 328928.571429 | 31.420928 |
| D15 | Apartment for sale | 217111.111111 | 17.876998 |
| D16 | Apartment for sale | 288166.666667 | 26.741447 |
| D17 | Apartment for sale | 204983.333333 | 16.939889 |
| D18 | Apartment for sale | 306111.111111 | 29.313189 |
| D20 | Apartment for sale | 227500.000000 | 23.209242 |
| D22 | Apartment for sale | 186613.636364 | 15.080259 |
| D24 | Apartment for sale | 176666.666667 | 12.374896 |

Figure 12: Sale price with area score

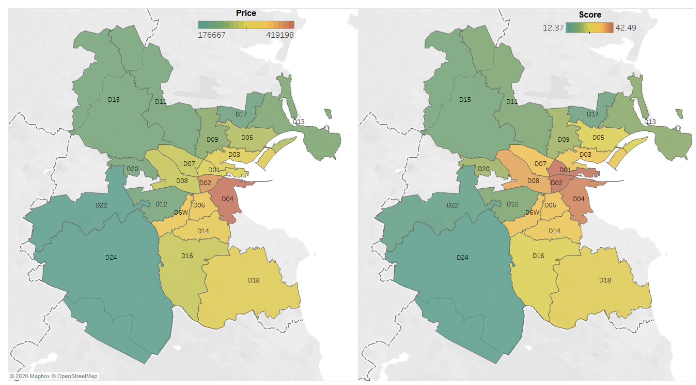


Figure 13: Sale price v/s Score comparison

cluster does not improve much.[5]

From Figure 14, we can clearly see that k = 10 is the optimal cluster with the lowest cost, after which model efficiency remains the same.

The clusters obtained are plotted using folium. Figure 15 shows total clusters, each represented with a different colour. Blue colour represents cluster 4, which has properties with zero neighbourhood retail stores. This will be used as the test set for our next predictions.

### *B. Step 02. Walkable distance prediction:*

Using property coordinates and walkable distance calculated earlier, we try to predict the optimal distance for the test set obtained from k-means clustering. Machine learning methods like KNN regression, random forest, decision trees

[5]https://medium.com/towards-artificial-intelligence/get-the-optimal-k-in-k-means-clustering-d45b5b8a4315

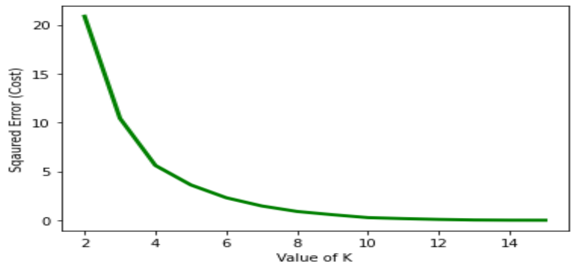



Figure 14: Variation of cost along with K

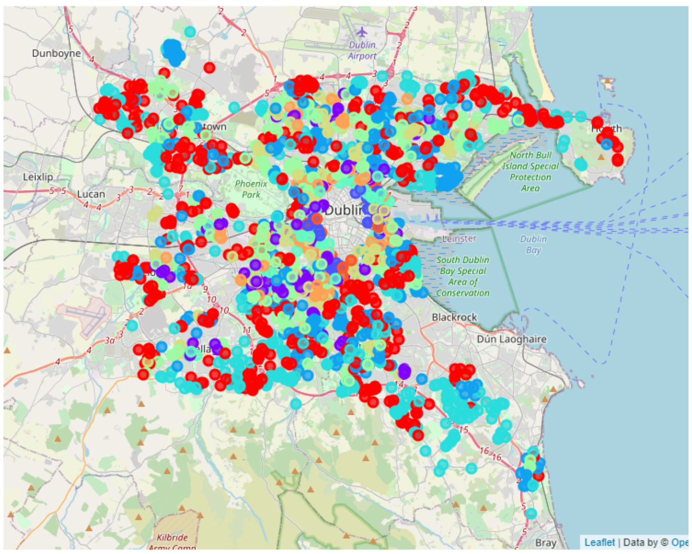



Figure 15: K-means cluster plot

are checked for efficiency. KNN regression provides less Root Mean Squared Error value of 0.2136 hence it is used for further prediction. The K-nearest neighbour algorithm is set

```
Mean Absolute Error: 0.17547510729613733
Mean Squared Error: 0.045626125188364335
Root Mean Squared Error: 0.21360272748343906
```

Figure 16: Efficiency of KNN

with n_neighbours to 15, the metric as haversine distance indicating that we use the haversine distance to calculate the new average value from each K neighbours. The aim of the prediction is to compute amenity placement distance which helps to find the optimal location for an amenity to be placed, from each property.

*C. Step 03. Retail store placement using predicted distance:*

Predicted walkable distance in the previous step is used as the radius to calculate the optimal location for a retail store. Rather than working with all properties, we create a subset where placing a retail store would affect a greater number of people. Population and density data collected in previous steps from www.citypopulation.de/en/ireland/dublin/ is used to determine the population spread across Dublin. Dublin 15 area which includes Castleknock, Clonsilla, Blanchardstown with more than 100,000 population will be used as placing a store there would influence a larger audience.

Fifty-one houses from the dataset which belong to the Dublin 15 area are used for further processing. The reverse haversine method in Figure 17 is used to compute the destination Longitude and Latitude, given the distance, source Longitude and Latitude and bearing angle.[6]

*dist = d/ R  --- arc distance*

$$(\sin^{-1}(\sin(\lambda 1) * \cos(dist) + \cos(\lambda 1) * \sin(dist) * \cos(\theta)) * 180)/\pi$$

$$((\Phi 1 + \mathrm{ATAN2}\left(\left(sin(\theta) * sin(dist) * cos(\lambda 1)\right), \left(cos(dist) - \left(sin(\lambda 1)\right) * \ sin(\lambda 1)\right)\right) + \pi)\%(2 * \pi)) - \pi$$

Figure 17: Reverse Haversine formula

where $\lambda 1$ is source latitude, $\phi 1$ is source longitude, sdist = predicted distance R where R is the radius of the earth 6371KM. $\theta$ is the bearing distance. $\theta$ is used to point the direction in which destination latitude and longitude to retrieve. $\theta$ = 0 represents North, $\theta = \pi$ is South, $\theta = \pi/2$ is East, and $\theta = 3*\pi/2$ is West.

Reverse haversine is calculated for all properties in Dublin 15 based on the predicted distance for each property. To illustrate this, the predicted distance for property 11 Inglewood Road, Clonsilla, Dublin 15 is 0.64667km. So reverse haversine is calculated in all four directions from this house to calculate the probable longitude and latitude in each direction. The resulting output for this property is [53.39657,-6.40906] [53.39076, -6.39931] [53.38494,-6.40906] [53.39076,-6.41881] in North, East, South and West directions respectively. This provides a probable location to place a retail store in all four directions within a predicted distance radius.

Once coordinates in each direction are known, we compute how many properties fall within 1 km of the generated coordinates in all directions. Figure 18 shows the coverage of properties from each predicted longitude and latitude.

There is a clear inference that the Northside of the properties like 34 Portersgate Heights, Clonsilla, Dublin 15 and 44 Portersgate Court, Clonsilla, Dublin 15 which covers 12 properties out of 18 would be the prime location to place a retail store. Figure 19 illustrates the impact of placing the amenity. The predicted walkable distance for the property 34 Portersgate Heights, Clonsilla, Dublin 15 is 0.734Km. A store is placed in all the four directions from the property marked in yellow, and the circular ring shows the 1 km radius coverage from the store location. The Northside with the red circle covers 12 houses in Clonsilla. It is evident that placing a retail store to the North of this property near coordinates [53.39359, -6.42039] and [53.39146, -6.42163] would cover most of the properties, in turn helping a large population.

[6]http://mathforum.org/library/drmath/view/51816.html

<table>
<tr><th>Location</th><th>Area</th><th>property_latitude</th><th>property_longitude</th><th>Predicted_distance</th><th>North_side_count</th><th>East_side_count</th><th>South_side_count</th><th>West_side_count</th><th>North_side_coordinates</th><th>East_side_coordinates</th><th>South_side_coordinates</th><th>West_side_coordinates</th></tr>
<tr><td>11 Inglewood Road|| Clonsilla|| Dublin 15</td><td>Dublin 15</td><td>53.390757</td><td>-6.40906</td><td>0.64666667</td><td>2</td><td>2</td><td>7</td><td>8</td><td>[53.39657, -6.40906]</td><td>[53.39076, -6.39931]</td><td>[53.38494, -6.40906]</td><td>[53.39076, -6.41881]</td></tr>
<tr><td>132 Lohunda Downs|| Clonsilla|| Blanchardstown|</td><td>Dublin 15</td><td>53.388462</td><td>-6.403396</td><td>0.57333333</td><td>2</td><td>2</td><td>11</td><td>6</td><td>[53.39362, -6.4034]</td><td>[53.38846, -6.39475]</td><td>[53.38331, -6.4034]</td><td>[53.38846, -6.41204]</td></tr>
<tr><td>126 Hazelbury Green|| Clonee|| Clonee|| Dublin 1</td><td>Dublin 15</td><td>53.401931</td><td>-6.433323</td><td>0.49066667</td><td>7</td><td>7</td><td>9</td><td>7</td><td>[53.40634, -6.43332]</td><td>[53.40193, -6.42592]</td><td>[53.39752, -6.43332]</td><td>[53.40193, -6.44072]</td></tr>
<tr><td>12 Portersgate Drive|| Clonsilla|| Clonsilla|| Dubli</td><td>Dublin 15</td><td>53.386412</td><td>-6.423903</td><td>0.72</td><td>11</td><td>7</td><td>4</td><td>8</td><td>[53.39289, -6.4239]</td><td>[53.38641, -6.41305]</td><td>[53.37994, -6.4239]</td><td>[53.38641, -6.43476]</td></tr>
<tr><td>101 Allendale Square|| Clonsilla|| Dublin 15</td><td>Dublin 15</td><td>53.38982567</td><td>-6.427479189</td><td>0.71333333</td><td>8</td><td>9</td><td>7</td><td>5</td><td>[53.39624, -6.42748]</td><td>[53.38983, -6.41672]</td><td>[53.38341, -6.42748]</td><td>[53.38983, -6.43824]</td></tr>
<tr><td>11 Aldemere Close|| Clonsilla|| Dublin 15</td><td>Dublin 15</td><td>53.386874</td><td>-6.426989</td><td>0.69333333</td><td>11</td><td>8</td><td>5</td><td>6</td><td>[53.39311, -6.42699]</td><td>[53.38687, -6.41653]</td><td>[53.38064, -6.42699]</td><td>[53.38687, -6.43744]</td></tr>
<tr><td>18 Castlegrange Dale|| Clonsilla|| Dublin 15</td><td>Dublin 15</td><td>53.3993555</td><td>-6.4283489</td><td>0.74666667</td><td>7</td><td>4</td><td>11</td><td>7</td><td>[53.40607, -6.42835]</td><td>[53.39935, -6.41709]</td><td>[53.39264, -6.42835]</td><td>[53.39935, -6.43961]</td></tr>
<tr><td>32 Barnwell Grove|| Hansfield|| Clonsilla|| Dublir</td><td>Dublin 15</td><td>53.3897649</td><td>-6.4436579</td><td>0.45066667</td><td>3</td><td>6</td><td>1</td><td>1</td><td>[53.39382, -6.44366]</td><td>[53.38976, -6.43686]</td><td>[53.38571, -6.44366]</td><td>[53.38976, -6.45045]</td></tr>
<tr><td>34 Portersgate Heights|| Clonsilla|| Clonsilla|| Du</td><td>Dublin 15</td><td>53.38699</td><td>-6.420389</td><td>0.73333333</td><td>12</td><td>5</td><td>4</td><td>8</td><td>[53.39359, -6.42039]</td><td>[53.38699, -6.40933]</td><td>[53.38039, -6.42039]</td><td>[53.38699, -6.43145]</td></tr>
<tr><td>23 Castlegrange Green|| Castaheany|| Clonsilla||</td><td>Dublin 15</td><td>53.398598</td><td>-6.429061</td><td>0.68666667</td><td>7</td><td>4</td><td>11</td><td>8</td><td>[53.40477, -6.42906]</td><td>[53.3986, -6.4187]</td><td>[53.39242, -6.42906]</td><td>[53.3986, -6.43942]</td></tr>
<tr><td>54 Allendale Square|| Clonsilla|| Clonsilla|| Dubli</td><td>Dublin 15</td><td>53.388853</td><td>-6.426407</td><td>0.69333333</td><td>11</td><td>9</td><td>6</td><td>6</td><td>[53.39509, -6.42641]</td><td>[53.38885, -6.41595]</td><td>[53.38262, -6.42641]</td><td>[53.38885, -6.43686]</td></tr>
<tr><td>4 Castlegrange Gardens|| Castaheany.|| D15RH93|</td><td>Dublin 15</td><td>53.398656</td><td>-6.430342</td><td>0.62666667</td><td>7</td><td>6</td><td>11</td><td>7</td><td>[53.40429, -6.43034]</td><td>[53.39866, -6.42089]</td><td>[53.39302, -6.43034]</td><td>[53.39866, -6.43979]</td></tr>
<tr><td>44 Portersgate Court|| Clonsilla|| Dublin 15</td><td>Dublin 15</td><td>53.384989</td><td>-6.421631</td><td>0.72</td><td>12</td><td>5</td><td>4</td><td>8</td><td>[53.39146, -6.42163]</td><td>[53.38499, -6.41077]</td><td>[53.37851, -6.42163]</td><td>[53.38499, -6.43249]</td></tr>
<tr><td>41 Beechfield Close|| Clonee|| Dublin 15</td><td>Dublin 15</td><td>53.402404</td><td>-6.439278</td><td>0.45066667</td><td>3</td><td>7</td><td>9</td><td>3</td><td>[53.40646, -6.43928]</td><td>[53.4024, -6.43248]</td><td>[53.39835, -6.43928]</td><td>[53.4024, -6.44608]</td></tr>
<tr><td>59 Allendale Drive|| Clonsilla|| Dublin 15</td><td>Dublin 15</td><td>53.392039</td><td>-6.429491</td><td>0.67333333</td><td>9</td><td>10</td><td>7</td><td>5</td><td>[53.39809, -6.42949]</td><td>[53.39204, -6.41934]</td><td>[53.38598, -6.42949]</td><td>[53.39204, -6.43965]</td></tr>
<tr><td>7 Castlegrange Dale|| Castaheany|| Clonsilla|| Du</td><td>Dublin 15</td><td>53.398682</td><td>-6.428067</td><td>0.71333333</td><td>7</td><td>4</td><td>11</td><td>8</td><td>[53.4051, -6.42807]</td><td>[53.39868, -6.41731]</td><td>[53.39227, -6.42807]</td><td>[53.39868, -6.43883]</td></tr>
<tr><td>8 Beechfield Green|| Clonee|| Dublin 15</td><td>Dublin 15</td><td>53.403266</td><td>-6.43702</td><td>0.45066667</td><td>3</td><td>7</td><td>8</td><td>3</td><td>[53.40732, -6.43702]</td><td>[53.40327, -6.43022]</td><td>[53.39921, -6.43702]</td><td>[53.40327, -6.44382]</td></tr>
</table>

Figure 18: Clonsilla properties

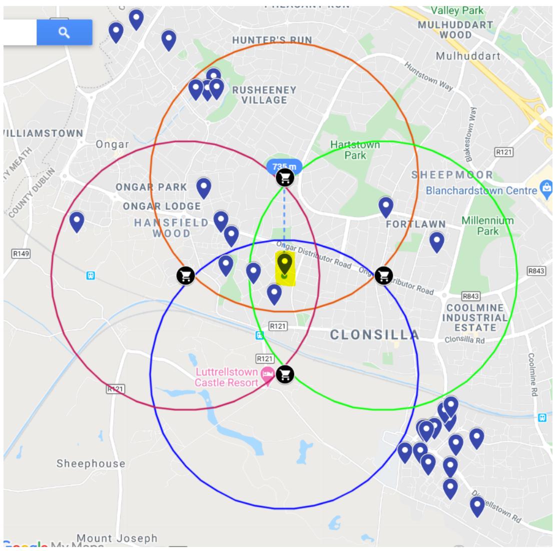



Figure 19: Placed amenity coverage

## VII. Conclusions and Future Work

This study explored how accommodation price varies in accordance with external factors like surrounding amenities, crime and distance to the city centre. This puts a perspective on the impact of walkability on property values. It is evident that sales properties have a variation very sensitive to their surrounding environment, safeness of the area and this changes when the location of the property moves away from the city. Though there are plenty of external factors that have an influence on property price, it is evident that our external factors do impact property price. Future work can include even more external factors.

A model was developed for determining optimal retail store placement which helps to find the best place to open a new retail store. This uses our existing walkability data to place a store and can be extended to place any amenity like a bus stop, Garda station, School etc. as principle remains the same. We realise our model is far from being perfect as there are many factors affecting opening a store and future work can include details like site information, the spending power of people in that area, income for that area, etc. Our model has a wide range of use cases as it helps to determine the optimal location so that opening a store impacts the largest number of citizens. In turn, this helps to inform the government of the areas that need better infrastructure for efficient urbanisation and betterment of Dublin city.


*Acknowledgement*

We thank our Practicum supervisor Alan Smeaton for his guidance and Minkun Kim for providing his research insights.


## References


[1] Diyanah Inani Azmi and Hafazah Abdul Karim. A comparative study of walking behaviour to community facilities in low-cost and medium cost housing. *Procedia-Social and Behavioral Sciences*, 35:619–628, 2012.

[2] Kylie Ball, Adrian Bauman, Eva Leslie, and Neville Owen. Perceived environmental aesthetics and convenience and company are associated with walking for exercise among Australian adults. *Preventive Medicine*, 33(5):434–440, 2001.

[3] Thierry Feuillet, Hélène Charreire, Mehdi Menai, Paul Salze, Chantal Simon, Julien Dugas, Serge Hercberg, Valentina A Andreeva, Christophe Enaux, Christiane Weber, et al. Spatial heterogeneity of the relationships between environmental characteristics and active commuting: towards a locally varying social ecological model. *International Journal of Health Geographics*, 14(1):1–14, 2015.

[4] John I Gilderbloom, William W Riggs, and Wesley L Meares. Does walkability matter? An examination of walkability's impact on housing values, foreclosures and crime. *Cities*, 42:13–24, 2015.

[5] L.X. Huang, L.J. Chen, J.M. Hao, D.C. Wang, and L.G Jin. Study on the relationship between housing price and transportation accessibility in urban district of Tianjin China. *Theor. Empir. Res. Urban Manag.*, 12:48–63, 2017.

[6] M. [Minkun] Kim. Walkability in Dublin City. Dublin City University M.Sc. Thesis, 2019.

[7] Kunsoon Park and Mahmood A Khan. An exploratory study to identify the site selection factors for us franchise restaurants. *Journal of Foodservice Business Research*, 8(1):97–114, 2006.

[8] C Carl Robusto. The cosine-haversine formula. *The American Mathematical Monthly*, 64(1):38–40, 1957.

[9] Jeroen Rook, Brent Verpaalen, Daniela Gawehns, and Mitra Baratchi. A multi-objective approach for optimal store placement.

[10] Yan Song and Gerrit-Jan Knaap. Quantitative classification of neighbourhoods: The neighbourhoods of new single-family homes in the Portland Metropolitan Area. *Journal of Urban Design*, 12(1):1–24, 2007.

[11] Taede Tillema, Erik Verhoef, Bert van Wee, and Dirk van Amelsfort. Evaluating the effects of urban congestion pricing: geographical accessibility versus social surplus. *Transportation Planning and Technology*, 34(7):669–689, 2011.

[12] Feng Wang, Li Chen, and Weike Pan. Where to place your next restaurant? optimal restaurant placement via leveraging user-generated reviews. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, pages 2371–2376, 2016.

[13] Emily Washington. Role of walkability in driving home values. *Leadership and Management in Engineering*, 13(3):123–130, 2013.

[14] Zelong Xia, Hao Li, and Yuehong Chen. Assessing neighborhood walkability based on usage characteristics of amenities under Chinese metropolises context. *Sustainability*, 10(11):3879, 2018.