

# CCT COLLEGE

**Data Visualisation**

**Assessment**

# Dublin 2022

**Nuno Alfredo Ribeiro Teixeira de Almeida - 2021310**

# HOUSE PRICING IRELAND

# ANALYSIS

Work developed in the analysis of a Dataset and Elaborate a report to obtain the first general grade of the Data Visualisation discipline of HDip Data Analytics - FT - Sept 2021 at CCT course.

Lectures: Marina Lantorno

# Dublin

# 2022

ILLUSTRATION LIST

Figure 1 – Size about Data Set;.……………………………………………………..……….…………...…05

Figure 2 – Data Set.…….………………………………………………………………………………….….06

Figure 3 – Analyzing variable “Year”;...…………………………………………………………………......07

Figure 4 – Using graphs to show analysis Year X Price by House Pricing Ireland Data set.….….…....07

Figure 5 – External Research about migration process in Ireland..………….………………….…….…08

Figure 6 – Frequency of Values ….…………………………………………………………………………09

Figure 7 –  Price Analyses by County..……………………………………………………….…….………10

Figure 8 – Density of Values..…………………………………………..……………………………...…....10

Figure 9 – Analysis about Price X Área X Price Group.…………………………………………….….…11

Figure 10 – Data Set Cleaned ….……………………………………..……………………………….……11

Figure 11 –  Analysis between Price X Class X Price Group..…………………………..………….……12

Figure 12 – Analysis between Price X Class X Price.x Area.………………………………………….....13

Figure 13 – Analysis Price X Area X Year..……………………………..……………………………….…13

Figure 14 – News by Journals ….……………………………………………..………………………….…15

**SUMMARY**

1. **INTRODUCTION …………………………………………………………………..03**
2. **SECTIONS……...…………………………………………………………………..07**

**2.1 GENERAL AUDIENCE…..…………………………………………..……….........07**

**2.2 MANAGEMENT TEAM ………………….…………………..……………………09**

**2.3 BUSINESS INTELLIGENCE TEAM …..……………………………………………11**

1. **CONCLUSION………………….…………………………………………………. 15**

**REFERENCES…………………………..…………………………………………16**

**01. INTRODUCTION**

For the complementary study of the Data Visualization discipline, it was necessary to analyse two data sets. In this report, we will work with Data analysis in one of them with the original name “House Pricing Ireland.” The archive carries information about the price of houses in Dublin and the Irish sales market. We will delve into further reading and further development of the data.

Despite being an essential deepening of the data, our goal is to meet a demand for three groups of readers, which will be highlighted in the following sections.

1. General Audience using Matplotlib;

2. Management team using Seaborn

3. Business Intelligence team using another library.

**Data Set**

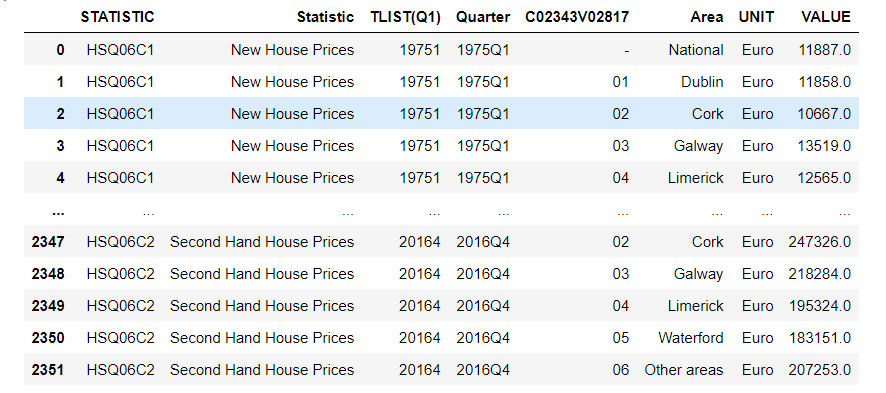
-9 Columns originals( Id; Statistic; Statistic; TList (Q1); Quarter; C02343V02817; Area; UNIT; VALUE )

-2352 Entries

Figure 1 – Size about Data Set;

**Data Dictionary**

The first scenario to understand the Dataset would be to interpret the values ​​in each column, because as we check the column names they do not give us clear information about what each of them brings us as results.

Figure 2 – Data Set;

***Id Column :*** mentions the unit identification, it is our index.

***STATISTICS:*** Brings nomenclature in the form of text; the two values ​​are HSQ06C1/ HSQ06C2;

***Statistic:*** It brings nomenclatures of the type of house that was analysed in the database, which are classified as New House Prices or Second Hand House Prices, which highlights whether the corresponding values ​​are for a new or second-hand house.

***TLIST(Q1) :*** Brings values ​​from 19751 to 20164.

***Quarter:*** Brings values 1975 + Quarter

***C02343V02817 :*** Text bringing the following values '-', '01', '02', '03', '04', '05', '06'

***Area:*** Names of cities for each unit surveyed. Being them: ‘National', 'Dublin', 'Cork', 'Galway', 'Limerick', 'Waterford', 'Other areas'.

***Unit :*** Currency of Exchange used. In this case the Euro.

***Value:*** Values ​​or Purchase price of the displayed unit. Minimum value of 10667E and Maximum value of 549330E

According to the last considerations, some steps are necessary for the preparation of the Dataset and external research to understand the subject and influence the relationship with external factors. The Data Preparation steps are cleaning columns and removing data.

We continue to analyse the data set from the values ​​presented, and we can observe that it is about sales comparisons, year, type of house, and location as the main variables. That report will make it clear that our object of study will be developed to present the results obtained using the existing metrics.

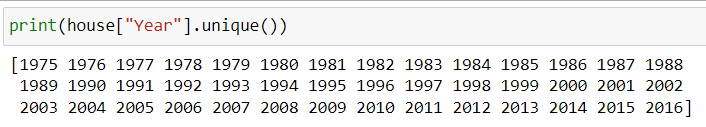
**02. SECTIONS**

2.1 GENERAL AUDIENCE

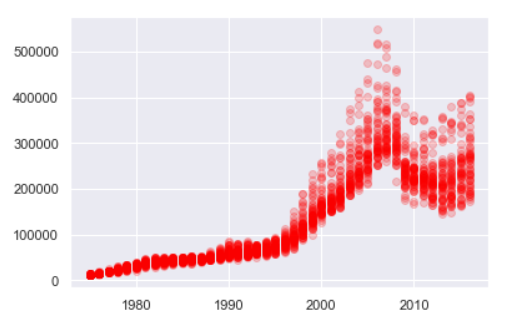
Understanding the data set used, we can be aware of what information was taken and used for study. Knowing the analysis for a General Audience, we can mention that the data set brings important and essential information for general analysis.

As we have home sales figures in Ireland and years, we can perform an annual analysis to understand the variation and compare it if their behaviour goes beyond 10% and house prices are now seven times average incomes as currently happens according to Whelan, 2022

After cleaning the Dataset, we noticed 58 houses per year from 1975 to 2016, with their variations in values ​​and type of house, whether new or second-hand.

Figure 3 – Analyzing variable “Year”;

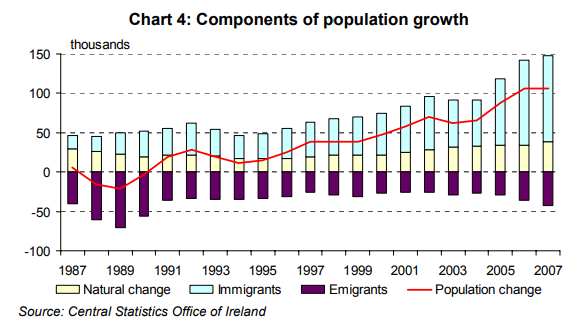
Looking at each year and seeking to understand the variation of houses, we can observe the following graph where we can analyse and find results on unit sales values ​​and the frequency of houses per year.

Figure 4 – Using graphs to show analysis Year X Price by House Pricing Ireland Data set;

The results obtained with this initial analysis were observed through this graph and we will explain that until the 90's we did not have houses with values ​​above 100k. These values ​​have been increasing in a non-gradual way in the mid-2000s. With this information, we can deepen our studies even more and understand what influenced this expressive increase.

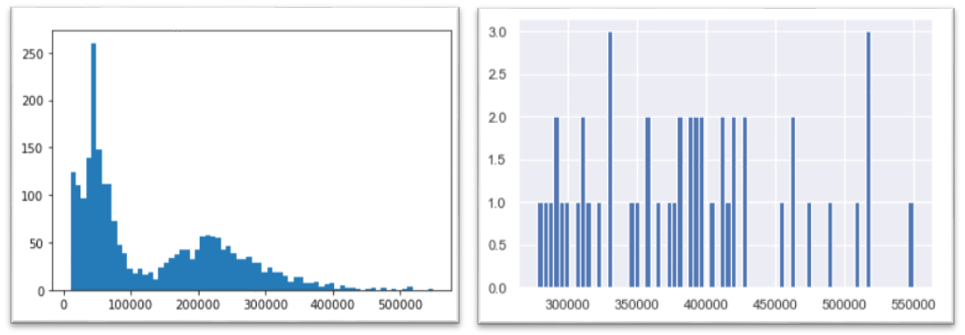
An external research where we seek to understand the changes that caused these increases to occur, in one of them they can be highlighted through the next figure, here one of the contributing factors may have been the increase in the population range of these years as we are analysing in parallel to the Dataset.

Figure 5 – External Research about migration process in Ireland;



The article in the figure above informs us about prices and population growth related to immigration and about Immigrants and Emigrants, showing the difference. “What can be one of the factors in the search for housing in Ireland?” (Malzubris, 2008 )

New house prices increased by 300 percent over the period and 340 percent in nominal ones between 1992 and 2006. At the same period, the housing stock increased by 150 percent. Much of the rise in real house prices can be attributed to strong housing demand fuelled by a young and growing population, rapid growth through real disposable income, low (and sometimes negative) real interest rates, and favourable tax treatment of residential property and house purchases for investment purposes.

Figure 6 – Frequency of Values; 

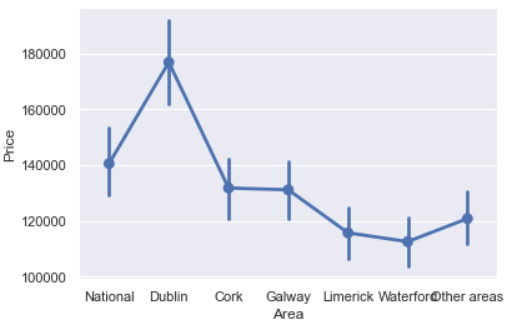
The frequency of values ​​analysed in the figure above shows us that most of the houses are with values ​​below 100k Euros. However, what we have to take into account about these values ​​are all the referring years. So as we previously noticed that the highest values ​​were around 2007, we performed a separation of the data in an earlier and a later year, so we obtained a result that the houses regardless of type were costing over 300k Euros.

2.2. MANAGEMENT TEAM

According to the analyses about the data set, we have the results obtained where the year with the highest peak is 2007, but other variables can be analysed so that decisions can be made and better deal with the attempts to understand which impacts were carried out by region with the values.

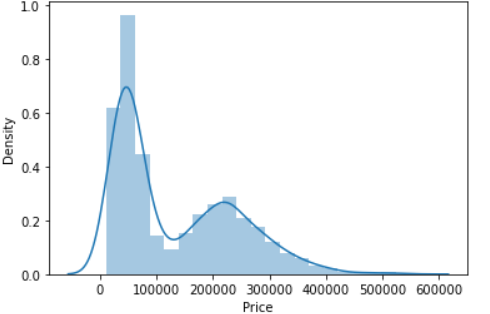
With this area analysis, we could also deal with the values ​​of square meters per region, but as the data set itself did not bring this type of information, the total value of a property may vary and be dependent on the size to perform this type of analysis. A very simple example, a house for 300k Euros analysed in Dublin can be equivalent to a house in Cork of the same value, but to deepen this analysis it is clear that we would need more data to work with and thus measure values ​​and more specific information.

Analysing the prices practiced by regions we can have the following result. Remembering that all results are accounting for every year.

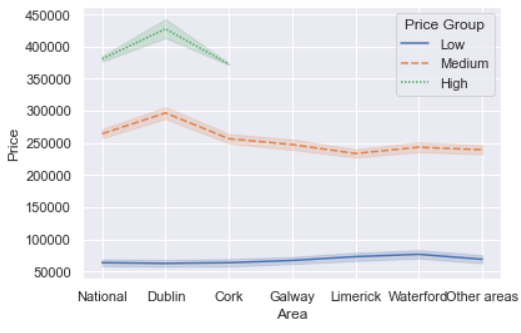
Figure 7 – Price Analyses by County;

The results obtained help us to understand the projection of values ​​in the Data set in a more specific way to the studies and characteristics; we can bring a more formal analysis of the data. Where we can use Seaborn Library to understand density through the same data.

Figure 8 – Density of Values;



We can use the graph below to show us a deeper analysis where they showed values. For this, a division of values ​​by class was performed in the Dataset, where we have them defined in Low, Medium and High. As reference, values ​​so that we could sort our data in order to specify and use as a reference which of these values ​​really are high and accessible according to the region.

Figure 9 – Analysis about Price X Área X Price Group ;

Then Areas such as National, Dublin and Cork are represented by having values ​​above 300k in a single property more often than in other regions.

2.3 BUSINESS INTELLIGENCE TEAM

According with audience in question, we can considerate data using more one library that can bring a result with more deeper. After cleaning all the data and bring important information.

Figure 10 – Data Set Cleaned;

Knowing that the data has been treated and we are speaking to an audience with more excellent knowledge of data, we use the Plotly Library, which brings more dynamic and interactive graphs and a zoom projection as an example to follow.

In the following chart, we can divide three features where we have the accumulated value of each. Determined by its price division, the “Price Group” was defined where average values ​​were taken from the division in three parts of the importance ​​of the original dataset.

Figure 11 – Analysis between Price X Class X Price Group;

According to the Price Group, the values ​​obtained by area were found for analysis of the regions later: Low 1587, Medium 716, and High 49 houses. It can also be analysed by year by reading the data set through other analyses.

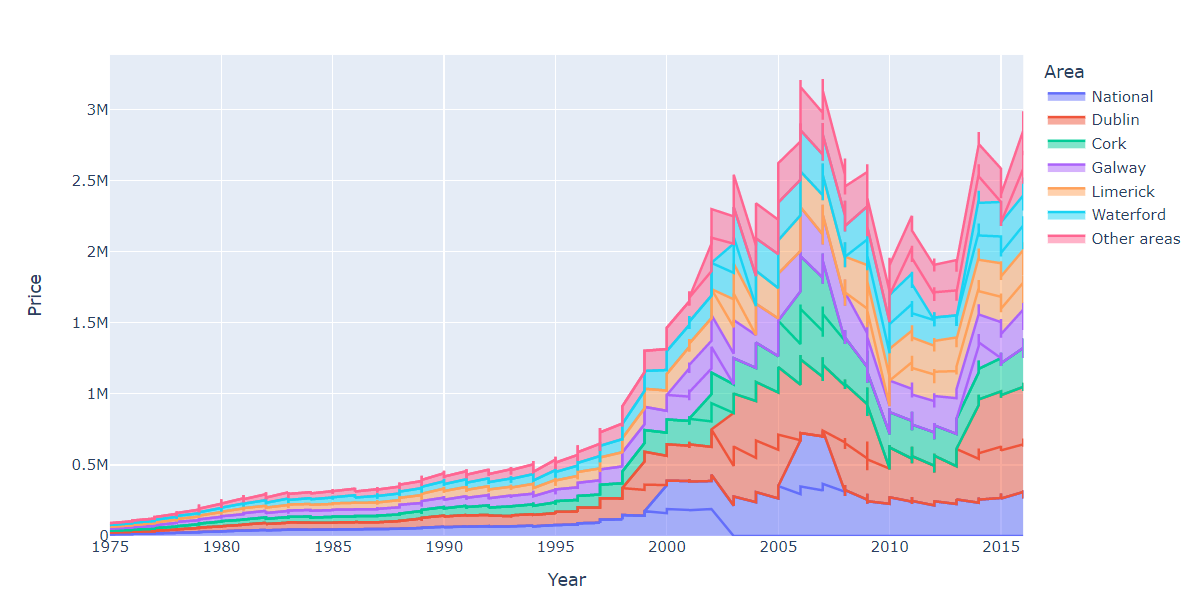
According to the more in-depth results, we can also analyse the chart below, which brings us the division by Price Group, Class (New or Second House), Prices, and Areas. They compare four variables that can bring us a deeper comparison of the impact of regions according to all these variables. Thus, we can extract several results according to the graph below, being aware that it is also interactive.

Figure 12 – Analysis between Price X Class X Price Group X Area;

The visible differences are in High Price in Galway, Limerick, Waterford and Other Areas; these do not have houses with high values ​​close to 400k. We can also see variations between new houses and Old House when we talk about the National and Cork area that in turn they only have used houses, where this negotiation can show that over the years there were few investments in construction in these areas until the year 2016, which brings the dataset. Therefore, a new concept for analysis would be to look for updated data to know the growth of values ​​in these regions with new houses.

According The Irish Times (2021), Homeland Projects is introducing roughly 20 units in Limerick's Abbey Grove, Mungret Gate, with prices ranging from €190,000 for one-bed flats to €310,000 for semi-detached homes. Then Nowadays the houses market are increasing and the prices are incoming.

The following line graphs also provides explanatory information about the evolution of each region on price. The same graph generated for General Audience, but with “Area” Feature divisions. We can observe a value that has not been dealt with before, which is the National area with the highest presence and frequency in the analysis of the graph, followed by Dublin and Cork.

Figure 13 – Analysis Price X Area X Year;

After analysed the graphs before, that shows the average by County( “Area” ), we can bring a news that was founded in The Gardians (2011) about what was happening if the numbers after 2007 which addresses the topic of shopping for new homes. Just as at that time, the market was highly prepared for sales in 2007 and everything that came after it was not expected for the case there was no clear external to the factors that would help to research.

According [O'Carroll](https://www.theguardian.com/profile/lisaocarroll) (2012) the fall was unsurprising because just 13,000 mortgages were approved in 2011, compared to 200,000 in 2006. "Given that the property market in every industrialized country is tightly tied to the mortgage industry, it's no wonder that values are down 50% or more while lending is down 90%."

**3. CONCLUSION**

This critical audience works with data as a way of putting together plans. In this way, dynamic and interactive graphics were chosen where I work with the most significant number of available Features.

As an excellent conclusion factor, we can mention the difficulties of updating the Data Set itself, where it did not have the data. As mentioned before, Nomenclatures required a whole Data Cleaning and Feature Engineering to complete.

Some extra data could be added to be used for future analysis. They are the square meters of each unit to calculate the value by region and if it is following the metric that exists today in the market. The study was also in-depth in finding values ​​and prices. Still, even if it had these specific characteristics, they could be complemented and better developed to search for new analyses and decisions.

Because of the analysis, we can see that there was a considerable increase in the years 2006, 2007 and 2008, where 168 houses analysed, almost 48 had an increase greater than 10% in value over the years, if we consider the same characteristics and type of house.

In Dublin, houses are increasing more quickly in price, up by 13.5%, and while apartment prices increased by 9.7%. Nationally, the median cost of a dwelling is now €276,000, while in Dublin; the median price for a dwelling was €405,000 in the year to November. ( O'REGAN, Ellen )

Some research was based on these headlines where they bring us a relation of migratory processes AND the fall and increase of areas around Dublin. So all the negotiations carried out in this study are based on past data, external research for complementation. Where it can be further explored below under new dealings in a Data Analysis.

Figure 14 – News by Journals;

**REFERENCES**

Malzubris, Jānis (2009) Ireland's housing market: bubble trouble. Avaiable at: https://ec.europa.eu/economy\_finance/publications/pages/publication13187\_en.pdf (Accessed: 24 March 2022 )

[O'Carroll](file:///C:\Users\Nuno\Desktop\Data\2º%20Semester\CA1\Data%20Visualisation%20Techniques\O'Carroll), Lisa (2012) Ireland's house prices at lowest levels since 2000. Avaiable at : <https://www.theguardian.com/business/2012/jan/03/ireland-house-prices-2000-levels> (Accessed: 29 March 2022 )

O'REGAN, Ellen (2022) Regional property prices continue to rise at faster rate than capital Avaiable at: <https://www.irishexaminer.com/news/arid-40788540.html> (Accessed: 29 March 2022 )

The Irish Times (2021) New homes around Ireland Avaiable at: <https://www.irishtimes.com/life-and-style/homes-and-property/new-to-market/new-homes-around-ireland-1.4674800> (Accessed: 02 April 2022 )

Whelan, Zuzia (2022) House prices across Ireland are almost 10% higher than they were a year ago.

Avaiable at: https://www.thejournal.ie/myhome-ie-report-house-prices-2022-5644942-Jan2022/ (Accessed: 24 March 2022 )

Libraries

***Plotly (***Avaiable at: <https://plotly.com/python/plotly-express/>)

***Seaborn*** ( Avaiable at : <https://seaborn.pydata.org/> )

***Matplotlib*** ( Avaiable at : <https://matplotlib.org/> )