

Project Report -Fintan M

GitHub URL

<https://github.com/Phinnnty/UCD-Data-Analytics-Course-Project>

Abstract

This Data Analytics project investigates the relationship between house prices in Ireland and the earnings of Irish workers. The project examines this relationship from 2010 to 2022, using two real-world datasets to determine whether the wage increase of Irish workers has kept pace with changing house prices.

The project aimed to showcase skills developed over the UCD Professional Academy course in Introduction to Data Analytics, which includes:

- Data Preparation
- Data Analysis
- Data Visualisation
- Deriving Insights

Overall, the project aims to leverage these skills to provide an insightful look at the relationship between housing prices and wages in Ireland over a 12 year period.

Introduction

There is a consensus in Ireland that the country is in the midst of a housing crisis. News articles, social media, and everyday discussion revolve around this issue. The Taoiseach Leo Varadkar went on record to admit that the country is short 250,000 houses that it desperately needs (Bowers, 2023). There are multiple contributing factors to this; for example, a short supply of materials and labour, and an increasing population.

One of the major symptoms of this issue is the perceived notion that buying a home in Ireland has become a far-off dream for many¹. Over the last 12 years people have begun to believe the cost of buying a home has become disproportionate to their earnings. This calls into question how long someone must save for to afford a mortgage, and whether their savings may be enough if house prices do go up. How much savings can be contributed towards a mortgage deposit is dependent on a buyer's wages, which begs the question – are Irish wages as competitive as the Irish housing market?

Thus, this project seeks to determine the average house price in each county in Ireland between the years 2010 and 2022, and explore its relationship to the change in earnings of Irish workers in different sectors over the same timeframe.

¹ While the Irish Times article states that house prices have gone down in 2023 by 0.6% on a monthly basis, there has still been a 12-year period of overall increase.

Dataset

The project uses two real-world datasets, including the Residential Property Price Register and the Weekly Average Earnings dataset from the Central Statistics Office (CSO), to investigate the relationship between house prices and wages in Ireland.

The datasets consist of the following information:

- The Residential Property Price Register includes Date of Sale, Price and Address of all residential properties purchased in Ireland since the 1st January 2010, as declared to the Revenue Commissioners for stamp duty purposes. (Authority, 2023)
- The data from the CSO included Weekly Average Earnings for Irish workers, across various positions, and Economic Sectors between 2008 and 2022. (Central Statistics Office, 2023)

The CSO is the national statistical office of Ireland, created with the mandate under the Statistics Act 1993 is "*The collection, compilation, extraction and dissemination for statistical purposes of information relating to economic, social and general activities and conditions in the State*". (Central Statistics Office, 2023)

The PSRA is the statutory body with responsibility for licensing and regulating the property services sector (Auctioneers, Estate Agents, Letting Agents and Property Management Agents) in Ireland which was established under the Property Services (Regulation) Act of 2011. (Authority, 2023)

Both sets of data can be considered to be reliably sourced by official government bodies of Ireland. Both datasets included in this project are available on the websites of the CSO and the PSRA.

Implementation Process

The implementation process began by importing the necessary libraries needed to carry out an analysis of this data. Key libraries include pandas, numpy, matplotlib, seaborn.

Data sets were then imported as csv files and read using pandas. Both datasets were been uploaded to github for reference purposes.

Dataset 1 Property Price Register:

Using the `tail()` function, 589000 rows of data were available for analysis; each row indicating a sale of a property including the date, price, address, and description of the property.

A `sns.countplot` was used to visualize the number of houses sold in each county over the specified period (Figure 1)

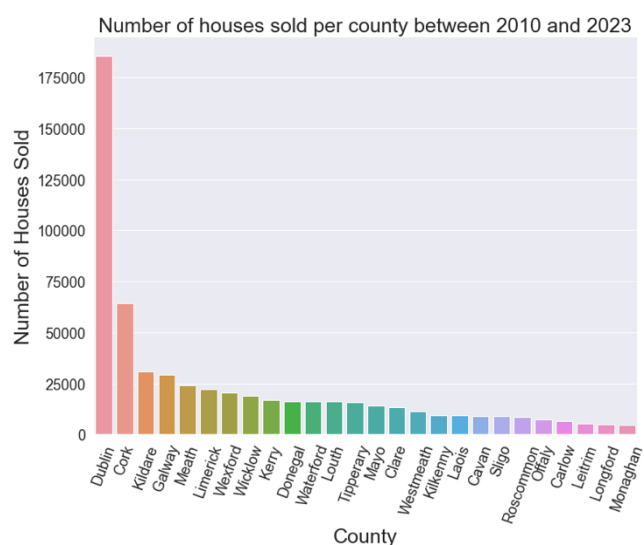


Figure 1

Further action was required to prepare the data for analysis:

1. Sum of .isnull carried out to determine which columns had missing values.
 - a. Eircode and Property Size description were missing 500,000 values.
2. Ensure the datatypes are workable (changed objects to numeric / date dtypes.)
3. Added in a “Year” column to refer to throughout the analysis.
4. Used group.by to filter the DF to only include County / Year / Price in Euro, and averaged the Price in Euro, per year for each county.
5. Using .groupby allowed removal of the null values in addition to paring down essential information.

The fourth step reduced the number of rows of data from 589,000 to 338.

6. Evaluate the change in house price during this period with . a catplot, which can be zoomed into in Fig 2:

The catplot demonstrated that there was an upward trend in the price of houses being bought in Ireland over the period 2010 – 2022.

A table ([Appendix 1](#)) was created to show that Average Price in 2010 and 2022, and calculate the % increase in property value over this period.

Seaborn was used to more effectively visualise the results of this tabulated data, the results of which will be covered in the [results](#) section of this report.

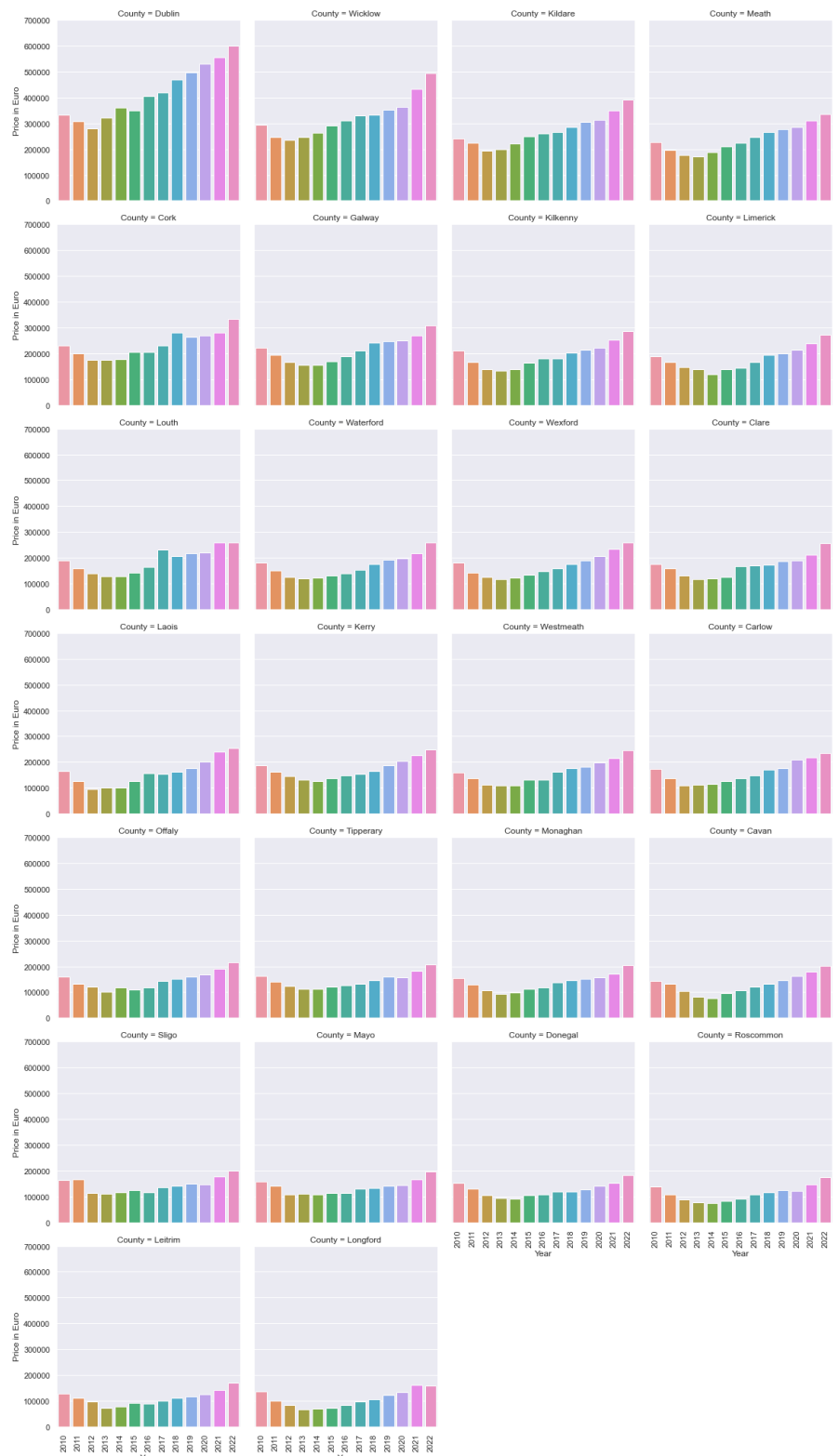


Figure 2

Dataset 2 CSO Average Weekly Earnings:

A data snapshot with `tail()` function, established that this dataset included 5039 rows of data including Average Weekly Earnings per Quarter, per Year, for each Economic Sector in Ireland across 4 types of Employees from 2008 until 2022.

The following steps were carried out to prepare the dataset.

The first was to test whether there were null values, resulting in the notice that 3325 values were missing from the VALUE column (Figure 3).

```
Presence of Null Values = True

Number of Null values = Statistic Label      0
Quarter                                     0
Economic Sector NACE Rev 2                 0
Type of Employee                           0
UNIT                                        0
VALUE                                     3325
Year                                        0
dtype: int64
```

Figure 3

A chart was plotted to determine where the missing values were in the VALUE column, it indicated that the VALUE column only had data on 'All Employees' post 2010 (Figure 4). The other employee data did not appear to be included post 2009, this was a non-issue as the scope of the project was only from 2010 onwards.

Using `.loc` function a new data frame was made with a Boolean condition that `Employee Type == All Employees`. The same chart was created with the new data frame and indicated that all values were included and there were no null values left. (Figure 5)

Before:

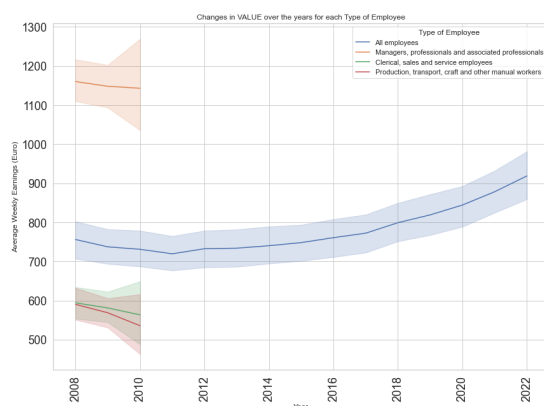


Figure 4

After:

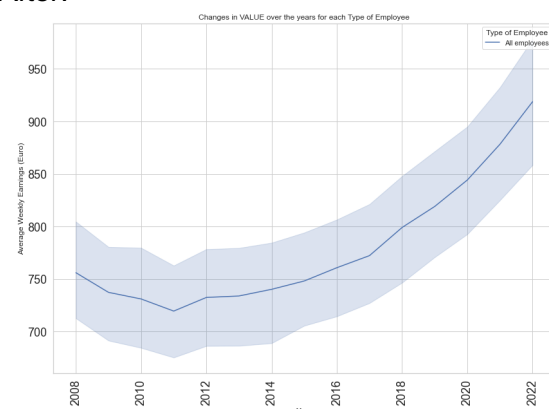


Figure 5

The next step carried out was to evaluate the dataset for irrelevant columns and duplicate info. Unique values of the names of Economic Sectors were printed and it was found that some sectors were actually combined sectors into one, and as such were removed i.e.

'Financial, insurance and real estate activities (K,L)' was a combination of 'Financial and insurance activities (K)' and 'Real estate activities (L)'.

The next step was to manipulate the earnings data so it could be compared against house price data using annual average earnings.

To achieve this these steps were taken:

1. the 'Quarterly Earnings' column was created, which was equal to the 'VALUE' column * 13 – as there are 13 weeks in a quarter.
2. DF2_grouped for only 'Year', 'Economic Sector NACE Rev 2', 'UNIT' was created.
3. DF2_summed, summed the quarterly earnings per year, for each economic sector under DF2_grouped.
4. Quarterly Earnings was then renamed to Yearly Earnings.
5. `.head()` was carried out to check if it worked. (Figure 6)

	Year	Economic Sector NACE Rev 2	UNIT	Yearly Earnings
0	2008	Accommodation and food service activities (I)	Euro	18100.55
1	2008	Administrative and support service activities (N)	Euro	25683.97
2	2008	All NACE economic sectors	Euro	36762.44
3	2008	Education (P)	Euro	45131.06

Figure 6

These refinements allowed further analysis with focus on the primary objective. First, determining the average change in wages over a 12 year period was required. A table (Appendix 2) was created demonstrating the Avg Earnings in 2010 and 2022, and the % change between them. A bar chart was created to showcase this which can be found in the [Results](#) section.

Comparing the Data

At this stage both datasets had been prepared for comparison, and insights had been derived for each without investigating the relationship between them.

In order to compare the datasets, I merged the samples of data from both dataframes. Taking 6 of the counties including both the highest and lowest, as well as taking a sample of 6 of the economic sectors. These 12 datapoints were chosen to represent the range of data between highs, lows, and mid-range datapoints to represent the relationship effectively.

This new dataframe was used to create two line-charts to visualize the change in wages and house prices. Filters were then applied to the merged dataframe in order to calculate the slope of the line for each chart. The slope could then feed results indicating the rate of change in the increase of wages and house prices.

A user-defined function was created to reuse code so repetition could be reduced.

```
def compute_slope(df, x_col, y_col):
    x = df[x_col]
    y = df[y_col]
    slope, intercept = np.polyfit(x, y, 1)
    return slope
```

Results

Analysis of both datasets allowed investigation into the relationship between wage and housing price.

From Dataset 1 on the House Prices:

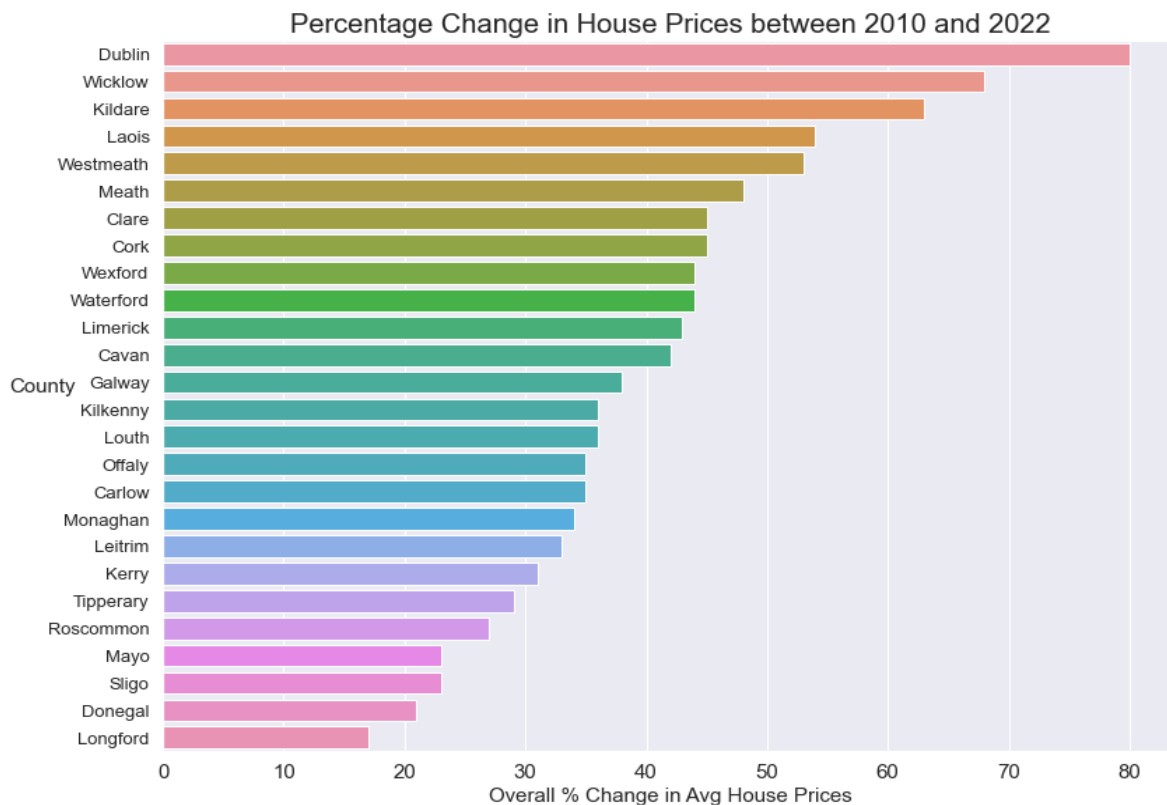


Figure 7

Overall change in house prices between 2010 and 2022 was illustrated in Figure 7. This bar chart was created based on the data manipulated to show the average house price per county per year. This bar chart shows that the price of houses has solidly increased over the time period with the highest being 80% while, the lowest is 17%.

Dataset 2 on Average Yearly Earnings:

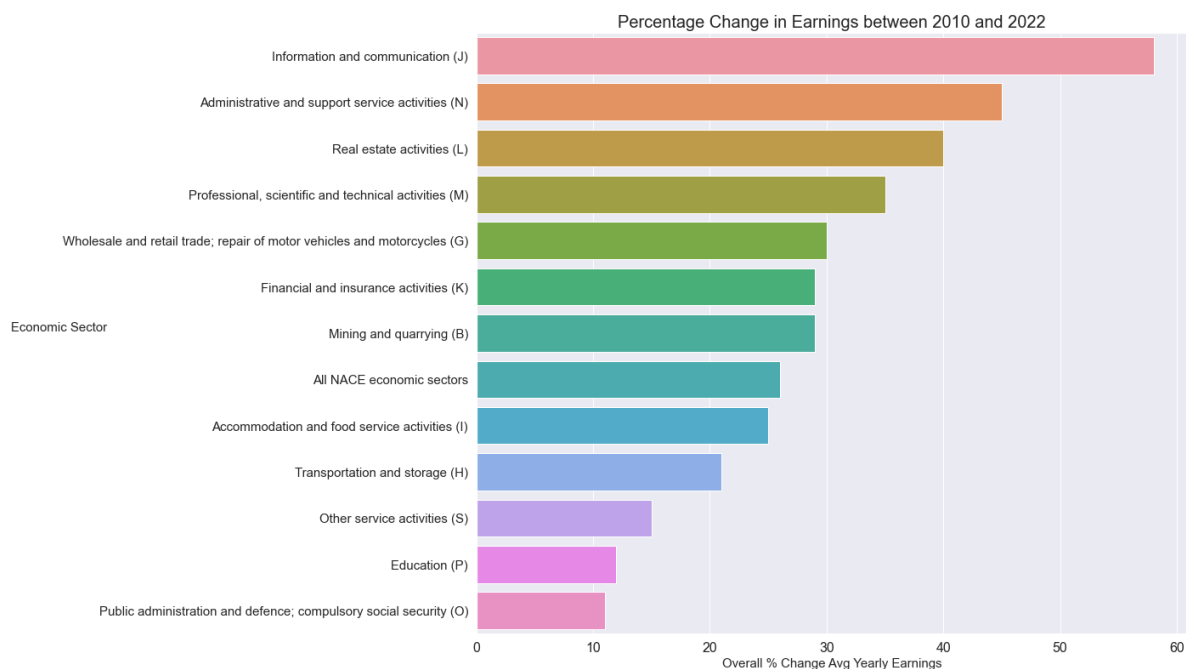


Figure 8

Figure 8 illustrates the overall increase in average yearly earnings for all Employees across Irish economic sectors, showing the largest increase is 58%, while the lowest is 11%.

Finally, the slopes of each of the lines included in these graphs were calculated to compare the rate of change between House Prices and Irish Wages. (Figure 9, 10 & 11)

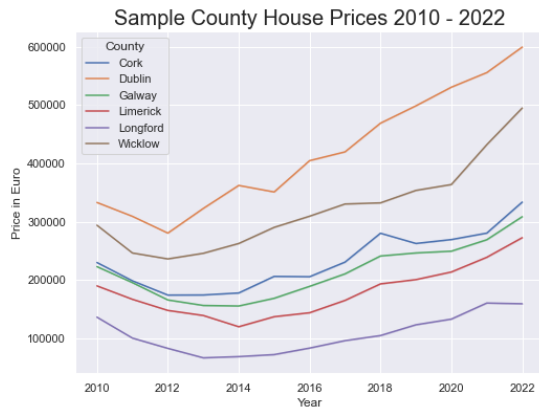


Figure 9

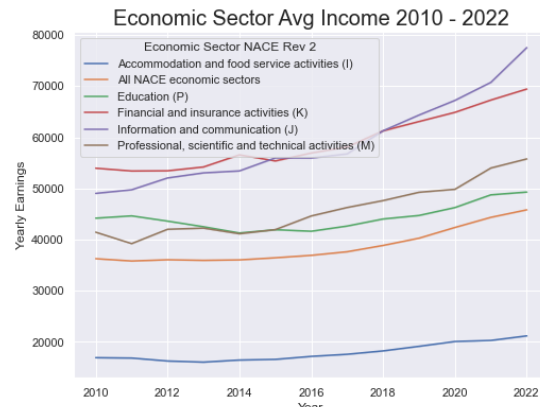


Figure 10

The slope of the All NACE economic sectors line is 796.9542857142621
The slope of the Information and communication (J) line is 2126.13214285709
The slope of the Financial and insurance activities (K) line is 1355.1014285713927
The slope of the Professional, scientific and technical activities (M) line is 1260.380714285675
The slope of the Accommodation and food service activities (I) line is 394.8864285714164
The slope of the Education (P) line is 409.53285714284357

The above slopes can be considered the rate of change in the earnings for Irish workers wages over 2010 - 2022, how does this compare against the house prices rate of change across Ireland

The slope of the Dublin House price line is 25487.56940431429
The slope of the Cork House price line is 10463.482036809117
The slope of the Longford House price line is 4967.512907285454
The slope of the Galway House price line is 9330.437763060181
The slope of the Wicklow House price line is 17281.33231324594
The slope of the Limerick House price line is 8119.849196820932

Figure 11

Insights

1. Based on the dataset from the Property Price Registry we can definitively state that there has been a consistent and even dramatic increase in the sale price of homes in Ireland.
2. The increase in house price appears to be focused in specific areas, with greatest effect seen in Dublin, Wicklow, and Kildare ([Figure 1](#)).
 - Dublin = 80% increase
 - Wicklow = 60% increase
 - Kildare = 63% increase

Determined the majority of residential property sales are in Dublin, Cork, and Kildare. Counties where the most homes are sold, are also the most expensive places to buy homes in Ireland. (Figure 7)

3. From the visualisation in Figure it is clear that the top three economic sectors with the highest increase in earnings are
 - Information and Communication = 58% increase
 - Administrative and Support Service Activities = 45% increase
 - Real Estate Activities = 40% increase
4. Having calculated the slope of the lines for our samples of each dataset it is clear that the rate of change in house prices far outstrips the rate of change of Irish earnings. The slope of each line represents the average increase in euro each year.
 - *The slope of the All NACE economic sectors line is 796.95*
 - *The slope of the Information and communication (J) line is 2126.13*
 - *The slope of the Education (P) line is 409.53*

If the average person's wage has gone up 796 euro, has it kept pace with the increase in house prices?

- *The slope of the Dublin House price line is 25487.56*
- *The slope of the Cork House price line is 10463.48*
- *The slope of the Longford House price line is 4967.51*

Dublin house prices appear to have gone up 25,487 euros a year, wages have not kept pace.

5. Further considerations and application to business settings include:
 - Increased potential for valuing tele-working or working from home if working outside of more expensive areas is available to employees. This may make the company a more prospective place to work for potential talent.
 - The government can evaluate this data to measure housing in Ireland. This data shows that the cost of buying a house in Ireland has increased at a far steeper pace than that of the average wage in Ireland. This can help inform policy decisions.
 - Banks can look at this may review this data when considering a mortgage application i.e. workers from "x" sector are more likely to be able to apply for mortgages in "y" counties.
 - Real estate developers can look at this data when considering new builds; i.e. building in Dublin may lead to a larger profit than other counties.

Additional Questions

1.
Machine Learning could be used to look at pattern recognition in data and make predictions about what this indicates. Machine learning could theoretically use the info and insights derived from the data used in this project to predict how housing prices may change in coming years. It could be used for predicting stock prices, looking at skin moles and detecting melanoma, creating tailor made ads for customers etc. There is a huge potential in machine learning.

2.
I would use regression methods with a dataset like this, the data revolves around money. Calculating the slope of the line as with this project touches on regression methods.

Appendix

	County	Avg Sale Price in Euro in 2010	Avg Sale Price in Euro 2022	Overall % Change in House Sale Price
0	Carlow	174221.0	235348.0	35.0
1	Cavan	142471.0	202309.0	42.0
2	Clare	175999.0	255190.0	45.0
3	Cork	230044.0	333625.0	45.0
4	Donegal	151829.0	183690.0	21.0
5	Dublin	332942.0	599245.0	80.0
6	Galway	223014.0	308575.0	38.0
7	Kerry	188045.0	247124.0	31.0
8	Kildare	240094.0	391494.0	63.0
9	Kilkenny	209911.0	284956.0	36.0
10	Laois	164331.0	253359.0	54.0
11	Leitrim	128098.0	170126.0	33.0
12	Limerick	190027.0	272606.0	43.0
13	Longford	136420.0	159377.0	17.0
14	Louth	189617.0	258321.0	36.0
15	Mayo	158670.0	195299.0	23.0
16	Meath	226326.0	334701.0	48.0
17	Monaghan	153118.0	204581.0	34.0
18	Offaly	159628.0	216138.0	35.0
19	Roscommon	137850.0	174710.0	27.0
20	Sligo	163200.0	200511.0	23.0
21	Tipperary	161320.0	207456.0	29.0
22	Waterford	181084.0	259923.0	44.0
23	Westmeath	160580.0	246390.0	53.0
24	Wexford	179564.0	257743.0	44.0
25	Wicklow	293963.0	494791.0	68.0

Appendix 1

	Economic Sector NACE Rev 2	Avg Yearly Earnings in Euro 2010	Avg Yearly Earnings in Euro 2022	Overall % Change Avg Yearly Earnings
0	Accommodation and food service activities (I)	16931.0	21194.0	25.0
1	Administrative and support service activities (N)	25286.0	36634.0	45.0
2	All NACE economic sectors	36277.0	45827.0	26.0
3	Education (P)	44196.0	49297.0	12.0
4	Financial and insurance activities (K)	53964.0	69438.0	29.0
5	Information and communication (J)	49039.0	77533.0	58.0
6	Mining and quarrying (B)	46172.0	59393.0	29.0
7	Other service activities (S)	24708.0	28448.0	15.0
8	Professional, scientific and technical activit...	41456.0	55777.0	35.0
9	Public administration and defence; compulsory ...	48735.0	53908.0	11.0
10	Real estate activities (L)	30597.0	42898.0	40.0
11	Transportation and storage (H)	36639.0	44294.0	21.0
12	Wholesale and retail trade; repair of motor ve...	26112.0	34026.0	30.0

Appendix 2

Works Cited

- Authority, P. S. R., 2023. *Residential Property Price Register*. [Online]
Available at: <https://www.propertypriceregister.ie/Website/NPSRA/pprweb.nsf/page/ppr-home-en>
[Accessed March].
- Bowers, S., 2023. *The Irish Times: Ireland's housing crisis facts and figures: All you need to know*. [Online]
Available at: <https://www.irishtimes.com/ireland/housing-planning/2023/03/23/irelands-housing-crisis-facts-and-figures-all-you-need-to-know/#:~:text=The%20State%20has%20too%20few,year%20from%202021%20to%202030.>
[Accessed April 2023].
- Central Statistics Office, 2023. <https://data.cso.ie/>. [Online]
Available at: <https://data.cso.ie/>
[Accessed March].