Data Analysis Project

Chad Rice

X00171454  TA\_KACOM\_B Year 3

Avergage House Prices In Ireland throughout From 1997 t0 2015

# Motivation

Go on the Irish subreddit or read any newspaper and you’ll most likely hear about the prices of houses or the housing crisis within the first few minutes. It is probably the largest issue facing Irish citizens now especially with the rising cost of living. It is agreed by most people that this is a huge issue, but prices keep rising. This is why I have chosen to do my dataset on the housing prices of Ireland. My dataset looks at the average price of house in Dublin, Cork, Galway, Limerick, and Waterford from the years 1997 until 2015.

I will be primarily focusing on Dublin as Dublin seems to have the highest house prices; however, I will compare it against the other counties to show how much of a difference there is in the prices.

In 1995 the Celtic tiger started, and it ended in 2007 and so the dataset should show a rise in house prices followed by a drop in 2008 and onwards due to the recession and we should see if they had any increase in between 2008 and 2015 where housing prices should start increasing quickly again due to the end of the recession and due to an economic growth of 6.7% in 2015.

Before beginning I viewed my dataset and removed columns that I taught were unnecessary. I found two columns that I decided to remove which were the National average housing price as the averages for Dublin, Cork, Limerick, Galway, and Waterford would be included in this and the average housing prices for other areas as I decided it was better to compare Dublin against single counties rather than multiple counties’ averages. I also removed a header from the csv as it messed with the .shape() method, and I removed some whitespace at the end of Dublin which caused errors when I tried to call Dublin and forgot to add the space. I also changed the YEAR column heading to Year so it would fit with the other column headings.

***Link to Dataset:***

<https://data.gov.ie/en_GB/dataset/average-new-house-price>

***Dataset***

This dataset includes the data of house prices from 18 years across 5 counties.

Below you will see the first 5 columns and the last five columns. The names of the columns can also be seen blow.

Text, calendar

Description automatically generated with medium confidence

This is the dataset, and the features are as follows:

1. Year: The year in which the rows data belongs to(int).
2. Dublin: The average housing price of Dublin in the corresponding year (was considered an object however changed to int).
3. Cork: The average housing price of Cork in the corresponding year (was considered an object however changed to int).
4. Galway: The average housing price of Galway in the corresponding year (was considered an object however changed to int).
5. Limerick: The average housing price of Limerick in the corresponding year (was considered an object however changed to int).
6. Waterford: The average house price of Waterford in the corresponding year (was considered an object however changed to int).

There are no null values in this dataset, and I have not found any duplicate values.

# Exploratory Analysis

I created a subset with just my relevant rows (Year, Dublin, Cork, Galway, Limerick, and Waterford) and printed it out to make sure the dataset was read in correctly.

Graphical user interface, text

Description automatically generated

A black screen with white text

Description automatically generated with low confidence

I then created a line chart displaying house prices of my chosen counties from 1997 until 2015:

Chart, line chart

Description automatically generated

As you can see in my dataset Dublin has always had a higher house price average compared to the other counties. Waterford has had the lowest for the past few years which is something I wish to look into. All the counties’ average house prices declined in the early 2010 however Dublin’s began to increase sooner alongside Galway’s in 2011 before they both dropped again in 2012. The other counties average house price did not start recovering until later. Cork in 2013, Galway’s prices dropped in 2012 and didn’t start rising again until 2014 and so did Limerick’s. Waterford had the most significant drop, at the start of the recession in 2007 Waterford had the fourth highest of the five counties however the house prices in Waterford didn’t start increasing until 2014 and it now has the lowest of the five counties.

Shown below is a plot chart displaying housing prices in Dublin from 1997 to 2015.

Chart, scatter chart

Description automatically generated

Shown below is a scatter plot displaying housing prices in Dublin and Galway from 1997 to 2015.

Chart, scatter chart

Description automatically generated

Shown below is a bar plot displaying housing prices in Dublin (Blue) and Cork (Red) from 1997 to 2015.

Chart, bar chart

Description automatically generated

As shown in the bar plot Dublin has always had a higher house price average over cork. This is most likely due to the population size of Dublin being larger than Cork. In 1996 Dublin had a population around 954,000 while Cork had a population around 127,000. This continues up to 2015 as well. I divided the columns for Dublin and cork by 100 and then computed their difference in prices. After this I used describe to show the description of the data, such as the average, standard deviation, and the number of non-empty values.

Text

Description automatically generated

Scatter plot displaying housing prices in Dublin and Waterford from 1997 to 2015.

Chart, scatter chart

Description automatically generated

Chart, bar chart, histogram

Description automatically generated

In the charts above you can see that Waterford average house price began to drop in 2008 which makes sense as that was the year after the recession, however in the chart we can see that Dublin’s average new house price increased to over 300,000 again unlike Waterford which continued to decline until 2014.

I decided to drop the national and other areas columns from my data set rather than having to use the subset as it easier and I believe neater.

Text

Description automatically generated

Text, calendar

Description automatically generated

After this I added a function to allow me to generate the ecdf (estimated cumulative distribution function). This allows me to plot some of the data from my dataset from the least to the greatest and see it as if it was dispersed across my dataset.

Text

Description automatically generated

After this I used the Dublin column to compute and plot the ECDF.

Text

Description automatically generated

I also designed a pie chart to compare the sum of all the house prices of Dublin from 1997 to 2015 against the all the prices of Waterford from 1997 to 2015.

I also printed the difference between the two.

Graphical user interface, text

Description automatically generated

Chart, pie chart

Description automatically generated

I also created a boxplot to display the range of the average house price in each county throughout the years:

Chart, box and whisker chart

Description automatically generated

# Measures of Relationships

For my Null and alternative hypothesis, I will be considering the following:

**H0:** The year the house is built in results in lower house price.  
**H1:**The year the house is built in does not result in lower house price.

The level of significance I have chosen will be 0.05.

I have chosen to use the Pearson correlation coefficient and p-value for testing my correlation.

I then ran the Pearson test on Dublin and the years columns and another one for Waterford and the years to get my p-value.

Text

Description automatically generated



From the above results I have decided to reject the null hypothesis and so I believe the year the house is built does not affect the average house price in both Dublin and Waterford.