**Data Analytics CA2**

**Introduction**

We have been asked to calculate data on house prices from the year 2010 to 2017 in Ireland. We collected the data from [www.propertypriceregister.ie](http://www.propertypriceregister.ie) and used it to conclude our findings. We compiled our data using excel and the presenting it via PowerPoint.

**Database Table**

This is the sql query that we used to describe the database table – DESCRIBE `sales20102017`. This is what the query shows in the database. We set all names to varchars, the date to date and the price to text. We used the following fields listed below. We set the varchars to various numbers biased on the sizes of the text in the csv file. We did not set the date or text to numbers as they included dates and prices.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Field | Type | Null | Key | Default | Extra |
| Date of Sale | date | NO |  | *NULL* |  |
| Address | varchar (500) | NO |  | *NULL* |  |
| Postal Code | varchar (20) | NO |  | *NULL* |  |
| County | varchar (40) | NO |  | *NULL* |  |
| Price (€) | text | NO |  | *NULL* |  |
| Not Full Market Price | varchar (5) | NO |  | *NULL* |  |
| VAT Exclusive | varchar (5) | NO |  | *NULL* |  |
| Description of Property | varchar (255) | NO |  | *NULL* |  |
| Property Size Description | varchar (255) | NO |  | *NULL* |  |

**Data Summary**

We had to find the max, min, mean (average) and standard deviation of the house prices for the years 2016 and 2017. For both years we found the max, min, mean and standard deviation for each county. Dublin was the county that had the highest max value, the highest average value, and the highest standard deviation value for 2016. Mayo was the county that had the lowest min value for 2016.

**This is year 2016**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| County | Max | Min | Mean | Standard Deviation |
| Carlow | 470000 | 10000 | 138195 | 70864 |
| Cavan | 2104000 | 10000 | 107133 | 103405 |
| Clare | 33180000 | 6000 | 167364 | 967681 |
| Cork | 5950000 | 7500 | 205745 | 172029 |
| Donegal | 941000 | 10000 | 108079 | 78070 |
| Dublin | 69208163 | 5400 | 404824 | 1028178 |
| Galway | 6161485 | 6667 | 189581 | 221545 |
| Kerry | 2700000 | 10000 | 147766 | 123927 |
| Kildare | 12073334 | 9000 | 260119 | 330669 |
| Kilkenny | 1940000 | 10000 | 179133 | 126295 |
| Laois | 20800000 | 6941 | 155865 | 743815 |
| Leitrim | 605000 | 7000 | 88997 | 65642 |
| Limerick | 1775000 | 7000 | 144321 | 111361 |
| Longford | 550000 | 9000 | 83641 | 58062 |
| Louth | 2250000 | 10000 | 166638 | 109968 |
| Mayo | 585000 | 5260 | 112899 | 81121 |
| Meath | 2120000 | 6000 | 225419 | 138790 |
| Monaghan | 508000 | 10000 | 118357 | 75096 |
| Offaly | 1145000 | 11000 | 116903 | 84116 |
| Roscommon | 960000 | 7500 | 92291 | 73973 |
| Sligo | 2393000 | 6000 | 116005 | 132184 |
| Tipperary | 1250000 | 7000 | 126124 | 101997 |
| Waterford | 2000000 | 11668 | 139335 | 108587 |
| Westmeath | 1218000 | 15000 | 132292 | 86517 |
| Wexford | 1566100 | 5500 | 146778 | 101230 |
| Wicklow | 5000000 | 10000 | 309963 | 226061 |

Dublin was the county that had the highest max value for 2017. Kerry was the county that had the lowest min value and the highest average value for 2017. Clare was the county that had the highest standard deviation value for 2017.

**This is year 2017**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| County | Max | Min | Mean | Standard Deviation |
| Carlow | 550000 | 15000 | 148490 | 75262 |
| Cavan | 880000 | 7600 | 121714 | 86863 |
| Clare | 23070000 | 7000 | 168312 | 648697 |
| Cork | 18300000 | 5715 | 231225 | 364929 |
| Donegal | 2700000 | 5900 | 120029 | 105733 |
| Dublin | 33538000 | 7000 | 420039 | 558085 |
| Galway | 4750000 | 5864 | 211092 | 206020 |
| Kerry | 1495000 | 5250 | 155271 | 112913 |
| Kildare | 3065000 | 7700 | 264882 | 157543 |
| Kilkenny | 1196000 | 13333 | 179829 | 117371 |
| Laois | 1255000 | 19089 | 153946 | 95289 |
| Leitrim | 510000 | 6000 | 100636 | 64013 |
| Limerick | 2300000 | 7365 | 165357 | 137566 |
| Longford | 531200 | 9400 | 96346 | 58264 |
| Louth | 2300000 | 10000 | 231054 | 287723 |
| Mayo | 5854801 | 6000 | 131374 | 191060 |
| Meath | 2104463 | 10000 | 247077 | 126613 |
| Monaghan | 2114520 | 6500 | 139702 | 158840 |
| Offaly | 1460000 | 10000 | 142485 | 102292 |
| Roscommon | 1075000 | 9000 | 109233 | 90213 |
| Sligo | 875000 | 7500 | 136295 | 97183 |
| Tipperary | 1612000 | 7000 | 133018 | 107530 |
| Waterford | 970000 | 10000 | 153649 | 106454 |
| Westmeath | 1859358 | 10000 | 161546 | 132920 |
| Wexford | 590000 | 12570 | 157336 | 80754 |
| Wicklow | 5000000 | 12745 | 331495 | 231609 |

We then had to find the max, min, mean (average) and standard deviation of the house prices each province in the years 2016 and 2017. While looking at the data we found that Leinster was the province that had the max price in both years. See the data below.

**This is year 2016**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Province | Max Price (€) | Min Price (€) | Mean Price (€) | Standard Deviation Price (€) |
| Connacht | €6,161,484.58 | €5,260.00 | €143,062.93 | €168,908.01 |
| Leinster | €69,208,163.00 | €5,400.00 | €312,894.63 | €795,442.43 |
| Munster | €33,180,000.00 | €6,000.00 | €170,823.44 | €322,444.63 |
| Ulster | €2,104,000.00 | €10,000.00 | €109,371.48 | €85,926.67 |

**This is year 2017**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Province | Max Price (€) | Min Price (€) | Mean Price (€) | Standard Deviation Price (€) |
| Connacht | €5,854,801.00 | €5,864.00 | €163,358.37 | €178,140.92 |
| Leinster | €33,538,000.00 | €7,000.00 | €331,198.35 | €443,398.81 |
| Munster | €23,070,000.00 | €5,250.00 | €187,580.14 | €322,739.96 |
| Ulster | €2,700,000.00 | €5,900.00 | €123,928.26 | €111,546.30 |

We then had to rank each county by the mean of house prices for the years 2105, 2016 and 2017. We ranked the mean from the highest to lowest in each year and to two decimal places. We then added this information to a graph to display the data better.

**This is year 2015**

|  |  |
| --- | --- |
| County | Mean of Price 2015 |
| Dublin | 351106.62 |
| Wicklow | 290302.65 |
| Kildare | 248505.71 |
| Meath | 211103.53 |
| Cork | 206439.61 |
| Galway | 168680.23 |
| Kilkenny | 164461.91 |
| Louth | 142568.88 |
| Limerick | 138480.53 |
| Kerry | 137423.07 |
| Clare | 133654.28 |
| Wexford | 132006.98 |
| Waterford | 130103.14 |
| Westmeath | 126795.18 |
| Carlow | 126740.73 |
| Laois | 124418.05 |
| Tipperary | 124186.27 |
| Sligo | 122243.82 |
| Offaly | 114021.64 |
| Mayo | 112363.98 |
| Monaghan | 110912.93 |
| Donegal | 103253.83 |
| Cavan | 94981.93 |
| Roscommon | 90156.54 |
| Leitrim | 81478.01 |
| Longford | 72283.57 |

**This is year 2016**

|  |  |
| --- | --- |
| **County** | **Mean of Price 2016** |
| Dublin | 404823.50 |
| Wicklow | 309962.62 |
| Kildare | 260118.52 |
| Meath | 225418.86 |
| Cork | 205745.20 |
| Galway | 189581.21 |
| Kilkenny | 179133.06 |
| Clare | 167364.29 |
| Louth | 166638.27 |
| Laois | 155865.18 |
| Kerry | 147766.22 |
| Wexford | 146777.60 |
| Limerick | 144320.54 |
| Waterford | 139335.07 |
| Carlow | 138194.73 |
| Westmeath | 132291.58 |
| Tipperary | 126123.91 |
| Monaghan | 118357.44 |
| Offaly | 116903.40 |
| Sligo | 116005.27 |
| Mayo | 112898.57 |
| Donegal | 108078.64 |
| Cavan | 107133.12 |
| Roscommon | 92291.24 |
| Leitrim | 88996.87 |
| Longford | 83640.80 |

**This is year 2017**

|  |  |
| --- | --- |
| **County** | **Mean of Price 2017** |
| Dublin | 420039.17 |
| Wicklow | 331494.74 |
| Kildare | 264882.23 |
| Meath | 247077.49 |
| Cork | 231225.01 |
| Louth | 231053.61 |
| Galway | 211091.65 |
| Kilkenny | 179828.67 |
| Clare | 168312.12 |
| Limerick | 165357.27 |
| Westmeath | 161545.94 |
| Wexford | 157335.90 |
| Kerry | 155271.42 |
| Laois | 153945.91 |
| Waterford | 153648.82 |
| Carlow | 148490.22 |
| Offaly | 142484.66 |
| Monaghan | 139702.11 |
| Sligo | 136295.45 |
| Tipperary | 133018.00 |
| Mayo | 131373.58 |
| Cavan | 121713.51 |
| Donegal | 120029.47 |
| Roscommon | 109233.22 |
| Leitrim | 100636.28 |
| Longford | 96345.93 |

We then had to provide a histogram of house prices by province for the year 2017. We calculated the groups for each province and then used that data to graph the histograms. See all four below.

**Connacht**

|  |  |
| --- | --- |
| **Max** | 5854801 |
| **Min** | 5864 |
| **Difference** | 5848937 |
|  |  |
| **Lower Quartile** | 75000 |
| **Upper Quartile** | 210000 |
| **Interquartile Range** | 135000 |

|  |  |
| --- | --- |
| **Group** | **Count** |
| 5000-<100000 | 2219 |
| 100000-<200000 | 2152 |
| 200000-<300000 | 1101 |
| 300000-<400000 | 330 |
| 400000-<500000 | 135 |
| 500000-<600000 | 51 |
| 600000-<700000 | 22 |
| 700000-<800000 | 10 |
| 800000-<900000 | 11 |
| 900000-<1000000 | 7 |
| 1000000-<1100000 | 3 |
| 1100000-<1200000 | 2 |
| 1200000-<1300000 | 0 |
| 1300000-<1400000 | 0 |
| 1400000-<1500000 | 4 |
| 1500000-<1600000 | 2 |
| 1600000-<1700000 | 0 |
| 1700000-<1800000 | 1 |
| 1800000-<1900000 | 0 |
| 1900000-<2000000 | 0 |
| 2000000-<2100000 | 3 |
| 2100000-<2200000 | 1 |
| 2200000-<2300000 | 0 |
| 2300000-<2400000 | 0 |
| 2400000-<2500000 | 0 |
| 2500000-<2600000 | 0 |
| 2600000-<2700000 | 1 |
| 2700000-<2800000 | 0 |
| 2800000-<2900000 | 0 |
| 2900000-<3000000 | 1 |
| 3000000-<3100000 | 0 |
| 3100000-<3200000 | 0 |
| 3200000-<3300000 | 0 |
| 3300000-<3400000 | 0 |
| 3400000-<3500000 | 0 |
| 3500000-<3600000 | 0 |
| 3600000-<3700000 | 0 |
| 3700000-<3800000 | 0 |
| 3800000-<3900000 | 0 |
| 3900000-<4000000 | 1 |
| >4000000 | 2 |
|  | 6059 |
|  |  |

This is the Graph.

**Leinster**

|  |  |
| --- | --- |
| **Max** | 33538000 |
| **Min** | 7000 |
| **Difference** | 33531000 |
|  |  |
| **Lower Quartile** | 175000 |
| **Upper Quartile** | 375000 |
| **Interquartile Range** | 200000 |

|  |  |
| --- | --- |
| **Group** | **Count** |
| 5000-<100000 | 2782 |
| 100000-<200000 | 7198 |
| 200000-<300000 | 9261 |
| 300000-<400000 | 5764 |
| 400000-<500000 | 2722 |
| 500000-<600000 | 1466 |
| 600000-<700000 | 967 |
| 700000-<800000 | 580 |
| 800000-<900000 | 375 |
| 900000-<1000000 | 257 |
| 1000000-<1100000 | 111 |
| 1100000-<1200000 | 99 |
| 1200000-<1300000 | 101 |
| 1300000-<1400000 | 73 |
| 1400000-<1500000 | 49 |
| 1500000-<1600000 | 46 |
| 1600000-<1700000 | 29 |
| 1700000-<1800000 | 26 |
| 1800000-<1900000 | 23 |
| 1900000-<2000000 | 54 |
| 2000000-<2100000 | 25 |
| 2100000-<2200000 | 17 |
| 2200000-<2300000 | 9 |
| 2300000-<2400000 | 18 |
| 2400000-<2500000 | 5 |
| 2500000-<2600000 | 4 |
| 2600000-<2700000 | 4 |
| 2700000-<2800000 | 7 |
| 2800000-<2900000 | 8 |
| 2900000-<3000000 | 5 |
| >3000000 | 78 |
|  | 32163 |

This is the Graph.

**Munster**

|  |  |
| --- | --- |
| **Max** | 23070000 |
| **Min** | 5250 |
| **Difference** | 23064750 |
|  |  |
| **Lower Quartile** | 92500 |
| **Upper Quartile** | 235000 |
| **Interquartile Range** | 142500 |

|  |  |
| --- | --- |
| **Group** | **Count** |
| 5000-<50000 | 1089 |
| 50000-<100000 | 2684 |
| 100000-<150000 | 2536 |
| 150000-<200000 | 2625 |
| 200000-<250000 | 1897 |
| 250000-<300000 | 1220 |
| 300000-<350000 | 681 |
| 350000-<400000 | 422 |
| 400000-<450000 | 237 |
| 450000-<500000 | 143 |
| 500000-<550000 | 70 |
| 550000-<600000 | 53 |
| 600000-<650000 | 39 |
| 650000-<700000 | 34 |
| 700000-<750000 | 13 |
| 750000-<800000 | 20 |
| 800000-<850000 | 9 |
| 850000-<900000 | 5 |
| 900000-<950000 | 6 |
| 950000-<1000000 | 14 |
| 1000000-<1050000 | 5 |
| 1050000-<1100000 | 2 |
| 1100000-<1150000 | 3 |
| 1150000-<1200000 | 3 |
| 1200000-<1250000 | 1 |
| 1250000-<1300000 | 3 |
| 1300000-<1350000 | 1 |
| 1350000-<1400000 | 1 |
| 1400000-<1450000 | 2 |
| 1450000-<1500000 | 2 |
| 1500000-<1550000 | 3 |
| 1550000-<1600000 | 0 |
| 1600000-<1650000 | 2 |
| 1650000-<1700000 | 1 |
| 1700000-<1750000 | 1 |
| 1750000-<1800000 | 0 |
| 1800000-<1850000 | 0 |
| 1850000-<1900000 | 0 |
| 1900000-<1950000 | 1 |
| 1950000-<2000000 | 1 |
| >2000000 | 12 |
|  | 13841 |

This is the Graph.

**Ulster**

|  |  |
| --- | --- |
| **Max** | 2700000 |
| **Min** | 5900 |
| **Difference** | 2694100 |
|  |  |
|  |  |
| **Lower Quartile** | 65000 |
| **Upper Quartile** | 158595 |
| **Interquartile Range** | 93595 |

|  |  |
| --- | --- |
| **Group** | **Count** |
| 0-<100000 | 1189 |
| 100000-<200000 | 1096 |
| 200000-<300000 | 250 |
| 300000-<400000 | 55 |
| 400000-<500000 | 12 |
| 500000-<600000 | 4 |
| 600000-<700000 | 1 |
| 700000-<800000 | 1 |
| 800000-<900000 | 2 |
| >1000000 | 5 |
|  | **2615** |

This is the Graph.

We then had to calculate the Mean and Standard Deviation for all the years.

|  |  |  |
| --- | --- | --- |
| **Calculations By Year** | **Average** | **Standard Deviation** |
| Year 2010 | 243025.228 | 239967.7616 |
| Year 2011 | 215063.078 | 239567.7091 |
| Year 2012 | 193594.861 | 294958.2722 |
| Year 2013 | 205056.963 | 658918.6664 |
| Year 2014 | 214709.273 | 648243.0034 |
| Year 2015 | 220350.199 | 428922.6053 |
| Year 2016 | 244997.873 | 629692.2064 |
| Year 2017 | 266331.799 | 390271.0665 |

**Hypothesis Tests**

1. We had to find if the average house prices differed between the provinces in 2017.

These are the tests we did for all six tests.













All of the tests we did showed that they all differed. The first three test scores are quite large compared to the other three as seen below.

|  |  |  |
| --- | --- | --- |
| **Province by Province** | **Differ** | **Test Score** |
| Leinster by Ulster | Yes | 20.16 |
| Leinster by Munster | Yes | 35.07 |
| Leinster by Connacht | Yes | 29.51 |
| Munster by Ulster | Yes | 8.59 |
| Munster by Connacht | Yes | 5.49 |
| Connacht by Ulster | Yes | 9.65 |

1. We had to see if the house prices increased from 2010 to 2017.

|  |  |
| --- | --- |
| **Leinster** | **Ulster** |
| Steps |  |
| 1 | Null Hypothesis: Mean A - Mean B= 0 |
| 2 | Alternative Hypothesis: Mean A - Mean B <>0 |
| 3 | Significance level 95% |
|  |  |
| 4 | Critical 5% region |
|  | T-Distribution |
|  | Use n1+n2-2 Degrees of Freedom |
|  | 71954.00 |
|  | Critical Value |
|  | 1.96 |
|  |  |
| 5 | Critical Region is Test Score >1.96 0r <-1.96 |
|  |  |
| 6 | Find Test Score |
|  |  |
|  | Numerator |
|  | 24950.67 |
|  | Denominator |
|  | 2957.970611 |
|  | Test Score |
|  | 8.44 |
| 7 | In critical region so reject the null hypothesis that Mean A=Mean B |
|  | Decide methods not the same |
|  | P Score |
|  | 0.000000000000003% |

Our test score was 8.44, which indicated to us that average house prices have increased over the years.

|  |  |
| --- | --- |
| **House Prices Increased from 2010 - 2017** | **Test Score** |
| Yes, Average House Prices have increased | 8.44 |

1. We had to find out if new houses cost more than second hand houses in the region of Dublin 10. Look below at our calculations. Our test score wasn’t positive.

|  |  |  |
| --- | --- | --- |
| **Houses in Dublin 10** | **Cost More** | **Test Score** |
| New vs Second Hand | Second hand houses cost more in Dublin 10 | -0.76 |





**Correlation**

We then had to find out if there was correlation between the house prices in Dublin and Louth throughout the years (2010 – 2017). We calculated the averages of the house prices from Dublin and Louth for each year.



**Prediction**

1. We had to produce a scatterplot showing the prices of all houses sold in Dublin 10 from January 1st 2010 to December 31st 2017. This is the scatterplot we got from our calculations:
2. Biased on the scatterplot we then had to extend the trend line to visually estimate what the mean of sale prices of houses would be by the end of June 2019.

The prediction of the mean of sale prices of houses would be by the end of June 2019 is 205,000.

|  |  |
| --- | --- |
| **Prediction** | **Value** |
| Mean Value June 2019 | 205,000 |

1. We then had to find the mean and standard deviation of house prices sold in 2017 in Dublin 10.

|  |  |
| --- | --- |
| **County** | **Dublin 10** |
| Mean | 168572.9013 |
| Standard Deviation | 88626.88139 |

1. Assuming that the price of houses sold in June 2019 in Dublin 10 was normally distributed, we then used the mean and standard deviation from 2017 to estimate the proportion of houses sold that will be between 200k and 300k euro. The table below shows the percentages we got between the range of 200K and 300K.

|  |  |  |
| --- | --- | --- |
| **Probability of Between 200 & 300k** | **Prob** | **Percentage** |
| x | Prob | Percentage |
| 200000 | 0.36144461 | 36.14% |
| 210000 | 0.320095181 | 32.01% |
| 220000 | 0.280867882 | 28.09% |
| 230000 | 0.244124081 | 24.41% |
| 240000 | 0.210141488 | 21.01% |
| 250000 | 0.179109777 | 17.91% |
| 260000 | 0.151130807 | 15.11% |
| 270000 | 0.126223058 | 12.62% |
| 280000 | 0.104329617 | 10.43% |
| 290000 | 0.085328885 | 8.53% |
| 300000 | 0.069047048 | 6.90% |

1. We then had to estimate the proportion of houses sold that will exceed 400k euro. The percentage we got was 0.45%. Look at our table below.

|  |  |  |
| --- | --- | --- |
| **Probability of >400k** | **Prob** | **Percentage** |
| x | Prob | Percentage |
| 400000 | 0.004510573 | 0.45% |

1. We then had to calculate the probability that 6 out of 10 houses will sell at a price of 300k or more. These are the results we got:

|  |  |  |
| --- | --- | --- |
| **x** | **Prob** | **Percentage** |
| 300000 | 0.069047048 | 6.90% |
|  |  |  |
| X | Prob |  |
| 6 | 0.0017% |  |
| 7 | 0.000072% |  |
| 8 | 0.0000020% |  |
| 9 | 0.000000033% |  |
| 10 | 0.00000000025% |  |
| Overall Prob | 0.0018% |  |

**Conclusion**

Overall, we found that Dublin was the county that had the highest max price value for both 2016 and 2017. This did not come as a surprise as Dublin is the capital of Ireland. It cost more to buy second hand houses in Dublin 10 biased off our calculations.