**CCT College Dublin**

|  |  |
| --- | --- |
| **Module Title(s):** | *Programming for DA*  ***Statistics for Data Analytics***  *Machine Learning for Data Analysis*  *Data Preparation & Visualisation* |
| **Assessment Title:** | *MSC\_DA\_Integr\_Repeat\_Sem1* |
| **Lecturer Name(s):** | ***Marina Iantorno/John O’Sullivan***  *Sam Weiss*  *Muhammad Iqbal*  *David McQuaid* |
| **Student Full Name:** | Candan Bayar |
| **Student Number:** | 2022352 |
| **Assessment Due Date:** | *26th May 2023* |
| **Date of Submission:** | *26th May 2023* |

****

**Declaration**

By submitting this assessment, I confirm that I have read the CCT policy on Academic Misconduct and understand the implications of submitting work that is not my own or does not appropriately reference material taken from a third party or other source. I declare it to be my own work and that all material from third parties has been appropriately referenced. I further confirm that this work has not previously been submitted for assessment by myself or someone else in CCT College Dublin or any other higher education institution.

**TABLE OF CONTENTS**

TABLE OF CONTENTS2

LIST OF FIGURES2-3

LIST OF TABLES 2-3

Introduction 4

Data Preparation and Visualisation5

**Analytical Framework** 5

**Problem Identification** 6

**Data Wrangling**

Statistics for Data Analytics10

Descriptive Statistics11-15

Statistical Tests16

Challenges20

Machine Learning Data Analytics20

Clustering21-22

Lineaar Regression 23-24

Discussion and Conclusion25

Dashboard Overview25-28

References28-30

# **LIST OF FIGURES**

[Figure 1: Geo plot of Construction Cost in Europe in 2021 4](#_Toc123881225)

[Figure 2: Heat map of missing observations in each of the data frames 4](#_Toc123881226)

[Figure 3: Ireland barplot 5](#_Toc123881227)

[Figure 4: Austria barplot 6](#_Toc123881228)

[Figure 5: Ireland histplot 7](#_Toc123881229)

[Figure 6: Austria hisplot 9](#_Toc123881230)

[Figure 7: Ireland Mean Visualisation 13](#_Toc123881231)

[Figure 8: Ireland Mean and Standard Deviation Visualisation 14](#_Toc123881232)

[Figure 9: Austria Mean Visiualisation 15](#_Toc123881233)

[Figure 10: Austria Mean and Standard Deviation Visiualisation 19](#_Toc123881234)

[Figure 11: Ireland Variance analysis Visualisation 19](#_Toc123881235)

[Figure 12: Austria Variance analysis Visualisation 20](#_Toc123881236)

[Figure 13: Ireland Heatmap Correlation Visualisation 22](#_Toc123881237)

[Figure 14: Austria Heatmap Correlation Visualisationl 22](#_Toc123881238)

[Figure 15: Dashboard Overview 1 25](#_Toc123881239)

[Figure 16: Dashboard Overview 2 26](#_Toc123881240)

[Figure 17: Dashboard Overview 3 27](#_Toc123881241)

# **LIST OF TABLES**

[Table 1: Null Values 9](#_Toc123881245)

[Table 2: Ireland and Austria Mean, Median, Standard Deviation table 10](#_Toc123881246)

[Table 3: Ireland and Austria T-statistics table 11](#_Toc123881247)

[Table 4: Ireland and Austria Wilcoxon statistics table 11](#_Toc123881248)

[Table 5: Ireland and Austria Chi-square test table 12](#_Toc123881249)

[Table 6: Ireland and Austria Variance Test table 12](#_Toc123881250)

[Table 7: Ireland and Austria Cluster (2014-2015) 16](#_Toc123881251)

[Table 8: Ireland Cluster 0,1,2 17](#_Toc123881252)

[Table 9: Austria Cluster 0,1,2 17](#_Toc123881253)

**INTRODUCTION**

Today, National House Construction Costs vary from country to country. In some countries these costs are quite high, in some countries the costs are low. One of the reasons for this is that each country's house building model is different, some countries use materials such as iron, cement, stone, and some countries use more flexible and less costly materials. Countries that use durable materials are generally located in the earthquake zone and they build durable houses accordingly.

In this study, the data of Ireland and Austria between 1994 and 2016 were analyzed and the costs of the two countries in the same years were compared and Machine Learning models were applied. In addition, data from countries in Europe between 2005 and 2021 were analyzed and geographically visualized, calculations were made with various statistical data and compared with Ireland and Austria.

A large part of the data belonging to Ireland, Austria and Europe consists of numerical data, so data analysis methods suitable for numerical variables were applied and analyzed. There were no null values in the datasets, but there were columns with a value of 0, which shows that there is a year whose value was not entered in the dataset, which was evaluated in these cases when performing data analysis.

In the analysis stages, the following stages were applied respectively; Programming for Data Analysis, Data Preparation and Visualisation, Statistics for Data Analysis, Machine Learning. These stages were implemented with the Python programming language and written in the report.

Data Preparation and Visualisation

**Analytical Framework**

To this end, a modified Data Science Method (DSM) framework (Johnson, 2018; Martinez, et al., 2021) was adopted which follows processes such as:

**Problem Identification**

The problem identification step involves specifying and establishing the research problem that the project seeks to address. This step allows the formation of the problem and the establishment of a solution scope. For instance, the current study seeks to propose a machine-learning model for predicting houses cost the future year in Ireland and Austria .

**Data Wrangling**

During data wrangling, various processes are undertaken including the collection of data, organization, and cleaning.

**Exploratory Data Analytics (EDA)**

Exploratory data analytics allows the building of data profile tables and plots to determine issues that are present in the data. This step also enables the detection of underlying relationships between various attributes as well as the distribution of the variables of interest. With respect to the current study, both the statistics for the data analytics and visualization tasks will be conducted during the Exploratory data analytics step in DSM.

During EDA, a dashboard was built to facilitate the visualization of the relationships and distributions of various attributes.

**Construction costs Europe in 2021**

harita içeren bir resim

Açıklama otomatik olarak oluşturuldu

Figure 1: Geo plot of Construction Cost in Europe in 2021

All Europe, National Home Construction Costs are visualized. Compared to 2021, the country with the highest costs is Turkey.

### 

### Checking missing observations

Null values in all data were checked and visualized with heatmap. No null values were found in the entire dataset.

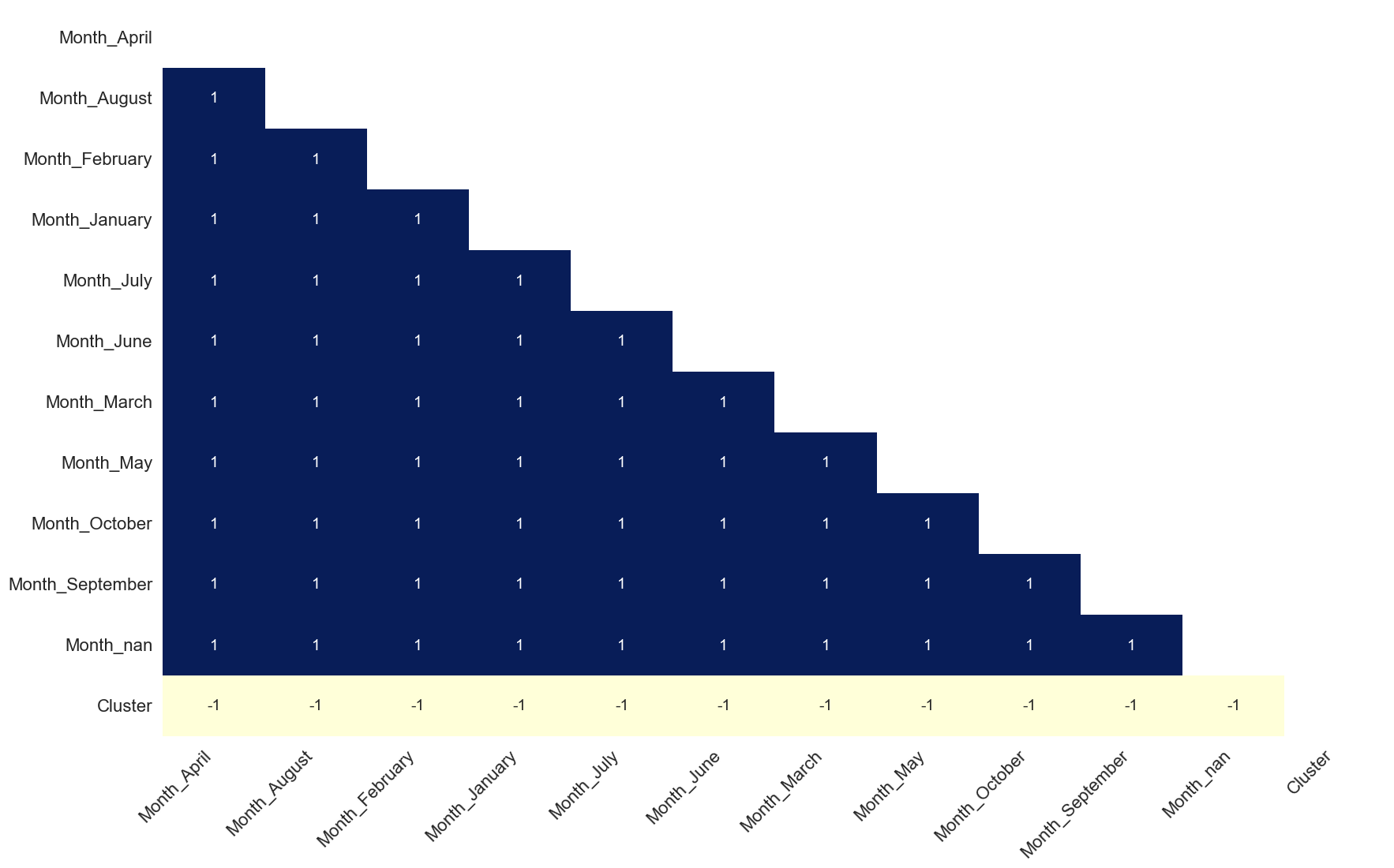
****

Figure 2: Heat map of missing observations in each of the data frames

The data for Ireland and Austria were visualized with barplot. According to the barplot visualization, the rate of Ireland is higher than that of Austria, which shows that the costs are more affordable in Austria.

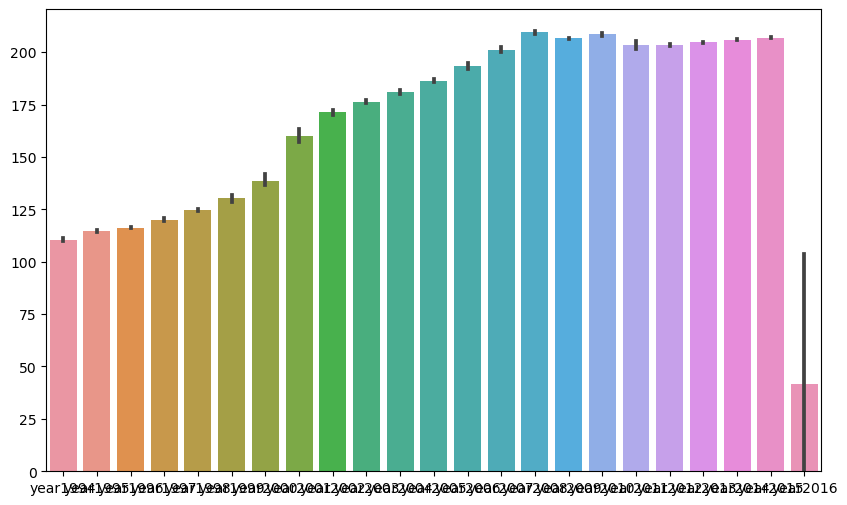
****

Figure 3: Ireland barplot

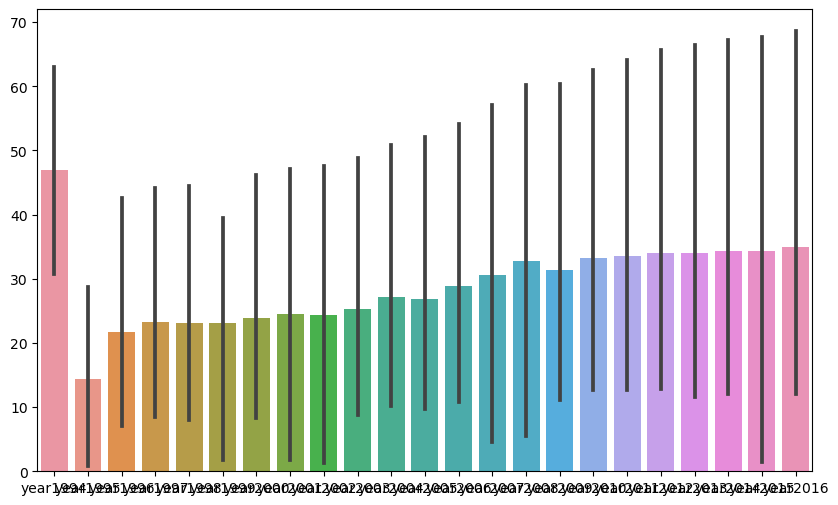
****

Figure 4: Austria barplot

To make a comparison according to the histogram graph, the costs are high in Ireland and are evenly distributed. In Austria, these rates are lower and evenly distributed. You can see it in figures 5 and 6.

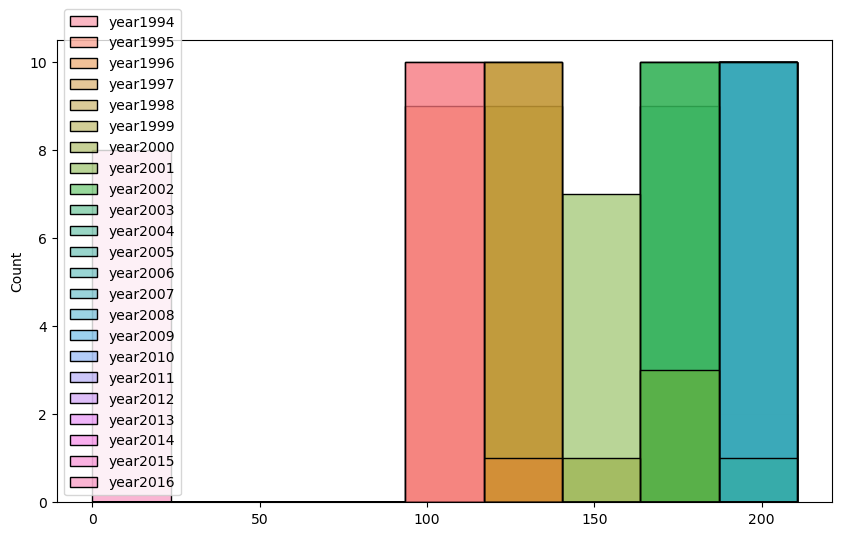
****

Figure 5: Ireland histplot

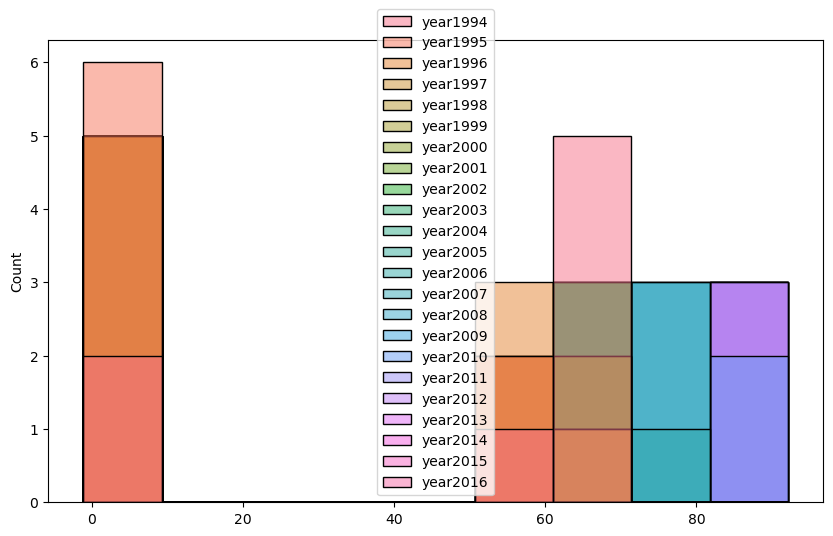
****

Figure 6: Austria hisplot

**STATISTICS FOR DATA ANALYTICS**

**Data Wrangling**

To achieve the research objective, data on National House Construction Costs in Europe were collected. The data included information on National House Construction Costs for the European country. In addition, Ireland and Austria were preferred from European countries and I collected data for the same years. Data were collected from <https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Construction_producer_price_and_construction_cost_indices_overview> , <https://www.oenb.at/isaweb/report.do?lang=EN&report=6.5> and <https://data.gov.ie/dataset/national-house-construction-cost-index/resource/188485c4-e5d7-4406-996b-1ff12a9a045a> . Prior to further analysis, data was filtered to include information for European countries only and restructured to convert indicators to column names and annual observations as row samples for relevant indicators. Indicators with missing observations were checked but not found to ensure that the research data had sufficient information to inform the purposeful findings.

|  |  |  |
| --- | --- | --- |
| **Year** | **Ireland** | **Austria** |
| **1994** | 0 | 0 |
| **1995** | 0 | 0 |
| **1996** | 0 | 0 |
| **1997** | 0 | 0 |
| **1998** | 0 | 0 |
| **2000** | 0 | 0 |
| **2001** | 0 | 0 |
| **2002** | 0 | 0 |
| **2003** | 0 | 0 |
| **2004** | 0 | 0 |
| **2005** | 0 | 0 |
| **2006** | 0 | 0 |
| **2007** | 0 | 0 |
| **2008** | 0 | 0 |
| **2009** | 0 | 0 |
| **2010** | 0 | 0 |
| **2011** | 0 | 0 |
| **2012** | 0 | 0 |
| **2013** | 0 | 0 |
| **2014** | 0 | 0 |
| **2015** | 0 | 0 |
| **2016** | 0 | 0 |

Table 1: Null Values

**Descriptive Statistics**

The Table below provides an overview of the mean, standard deviation, and median of the National Home Construction Costs index indicators for Ireland, Austria, respectively, from 1994 to 2016.

Instead of explaining all of them one by one, the years 2014-2015 were taken as an example and compared. The Irelanda mean for 2014 is less than 2015, so the Irelanda mean in 2015 is higher. In addition, when compared to Austria, the means for 2014-2015 were observed to be higher Ireland than for Austria. The standard deviation values of Ireland are not much different between 2014-2015, but the standard deviation of Ireland is lower in 2015. In addition, the standard deviation value of Austria is lower than that of Ireland in 2014-2015.

When the median values are compared, in 2014-2015, the value is higher in 2015. In addition, the values of Ireland are higher than that of Austria.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **IRELAND** | | | |  | **AUSTRIA** | | |
| **Year** | **Mean** | **Standard deviation** | | **Median** | | **Mean** | **Standard Deviation** | **Median** |
| 1995 | 114.450 | | 0.738 | 114.100 | | 14.325 | 24.539 | 1.600 |
| 1996 | 116.130 | | 0.469 | 115.950 | | 21.712 | 29.189 | 1.550 |
| 1997 | 120.010 | | 1.319 | 119.850 | | 23.312 | 29.037 | 2.900 |
| 1998 | 124.620 | | 1.219 | 124.500 | | 23.062 | 30.064 | 2.150 |
| 1999 | 130.250 | | 3.026 | 130.500 | | 23.175 | 30.710 | 2.300 |
| 2000 | 138.440 | | 5.426 | 136.850 | | 23.900 | 31.026 | 2.300 |
| 2001 | 159.960 | | 5.336 | 157.650 | | 24.438 | 31.598 | 2.500 |
| 2002 | 171.360 | | 2.171 | 169.950 | | 24.388 | 32.292 | 1.650 |
| 2003 | 176.390 | | 1.378 | 177.200 | | 25.275 | 32.721 | 2.850 |
| 2004 | 180.870 | | 1.447 | 180.100 | | 27.188 | 33.343 | 5.050 |
| 2005 | 186.410 | | 1.414 | 186.600 | | 26.875 | 34.647 | 2.350 |
| 2006 | 193.430 | | 2.470 | 194.250 | | 28.900 | 35.191 | 4.550 |
| 2007 | 201.220 | | 2.192 | 200.200 | | 30.587 | 36.471 | 4.800 |
| 2008 | 209.380 | | 1.069 | 209.650 | | 32.800 | 38.011 | 5.750 |
| 2009 | 206.540 | | 0.664 | 206.350 | | 31.362 | 40.167 | 3.950 |
| 2010 | 208.460 | | 1.060 | 208.650 | | 33.200 | 40.877 | 5.050 |
| 2011 | 203.250 | | 3.769 | 201.700 | | 33.500 | 42.375 | 3.300 |
| 2012 | 203.410 | | 0.601 | 203.600 | | 33.987 | 43.570 | 2.600 |
| 2013 | 204.670 | | 0.422 | 204.650 | | 33.950 | 44.795 | 2.200 |
| **2014** | **205.890** | | **0.576** | **205.950** | | **34.362** | **45.471** | **2.000** |
| **2015** | **206.960** | | **0.554** | **207.200** | | **34.375** | **46.266** | **1.650** |
| 2016 | 41.450 | | 87.384 | 0.000 | | 34.938 | 46.623 | 1.650 |

Table 2: Ireland and Austria Mean, Median, Standard Deviation table.

**Visualization (Mean,Median,Standard Deviation)**

**Ireland**

When the mean values of Ireland are analyzed, it was the highest year in 2010 and it continues to increase after 2015.

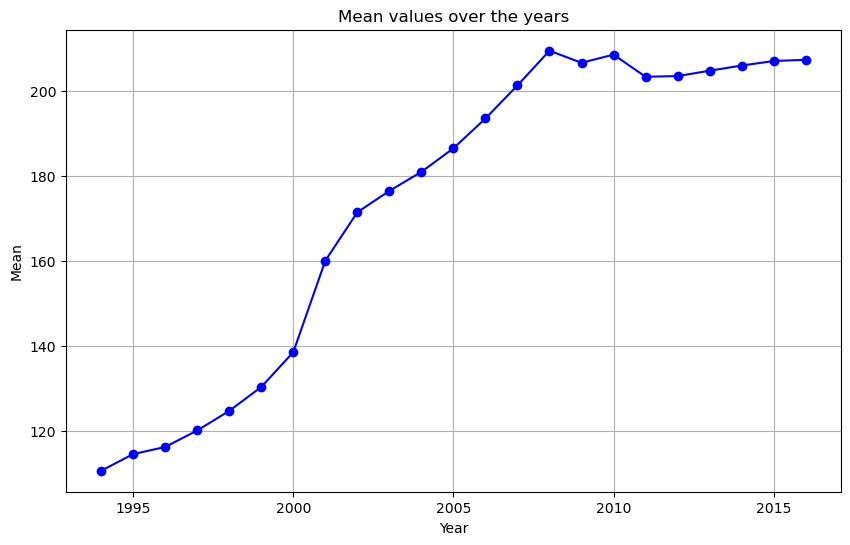


Figure 7: Ireland Mean Visualisation

When the standard deviations similar to the mean and comparison values of Ireland are examined, the highest year was 2010 and it continues to increase after 2015.

çizelge içeren bir resim

Açıklama otomatik olarak oluşturuldu

Figure 8: Ireland Mean and Standard Deviation Visualisation

**Austria**

When the mean values of Austria are analyzed, the costs are high in 1994 and before, but these mean values decreased afterwards. It increased slightly between 1996-1997, but then decreased. It has low mean values until 2015. It is expected to increase in 2015 and beyond.

çizelge içeren bir resim

Açıklama otomatik olarak oluşturuldu

Figure 9: Austria Mean Visiualisation

When comparing the standard deviation and mean, the deviation was not large in Austria and the mean has similar values.

çizelge içeren bir resim

Açıklama otomatik olarak oluşturuldu

Figure 10: Austria Mean and Standard Deviation Visiualisation

**Statistical Tests**

**T – test (2014 – 2015)**

According to the Ireland T-statistic, a negative relationship was observed as a result of the analysis, rejecting the null hypothesis and showing that there is a significant relationship. According to Austria values, there is no relationship between t - statistic as there is no significant evidence against the null hypothesis. According to the European t-statistics, the hypothesis does not have significant relationships. At a 0.05 level of significance, it is observed that Ireland does not follow a normal distribution (p < 0.05) as opposed to samples from Austria and the rest of Europe with p-values greater than 0.05.

|  |  |  |
| --- | --- | --- |
|  | **T- statistics** | **p - value** |
| **IRELAND** | -4.232118752188478 | 0.0005011652982170444 |
| **AUSTRIA** | -0.0005450128213722044 | 0.9995728326467744 |
| **EUROPE** | 0.14006206685019307 | 0.8890508180047323 |

Table 3: Ireland and Austria T-statistics table

**Wilcoxon-statistics (2014 – 2015)**

Based on Irish data, the p-value is 0.001953125, which is approximately equal to 0.002. This p-value provides strong evidence against the null hypothesis. If the null hypothesis is true, we can reject the significance level because this statistic is below 0.05. According to the data in Austria, there is no significant hypothesis relationship. According to the data in Europe, it shows that there is a significant relationship between the hypotheses.

|  |  |  |
| --- | --- | --- |
|  | **Wilcoxon-statistics** | **p- value** |
| **IRELAND** | 0.0 | 0.001953125 |
| **AUSTRIA** | 18.0 | 1.0 |
| **EUROPE** | 261.5 | 0.9627007428430175 |

Table 4: Ireland and Austria Wilcoxon statistics table

**Chi-square test (2014 – 2015)**

According to the data in Ireland, there is no statistically significant relationship between the observations. According to the Austrian data, there is no significant relationship between the hypotheses. Moreover, more data is needed for these tests.

|  |  |  |
| --- | --- | --- |
|  | **Chi-square test** | **p value** |
| **IRELAND** | 0.026 | 1.000 |
| **AUSTRIA** | 0.638 | 0.999 |

Table 5: Ireland and Austria Chi-square test table

**Variance Test (2014 – 2015)**

According to data from Ireland, the F-statistic and the associated p-value were analyzed between 2014 and 2015. In addition Irelanda data to the F-statistic, a difference was observed in the null hypothesis. The P-value is below 0.05, which indicates no statistically significant difference. The Austria data for 2014-2015 are analyzed. The difference between the F-statistic and the p-value is small, and it is necessary to analyze other data for further evaluation.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **F-statistics** | **p value** | **Variance** |
| **IRELAND** | 14.440078585461814 | 0.0025297631984834944 | 2014 : 0.196  2015 : 0.289 |
| **AUSTRIA** | 6.70534194370195e-05 | 0.9935820456670429 | 2014 : 2067.646  2015 : 2126.771 |

Table 6: Ireland and Austria Variance Test table

**Ireland Variance Analysis Visualisation**

The variance rate in 2015 is higher than in 2014. It is shown in the image below.

metin, ekran görüntüsü, diyagram, dikdörtgen içeren bir resim

Açıklama otomatik olarak oluşturuldu

Figure 11: Ireland Variance analysis Visualisation

**Austria Variance Analysis Visualisation**

As a result of the data analysis of the Austria year 2014-2015, there is not much difference between the 2 years, but the variance rate is higher in 2015.

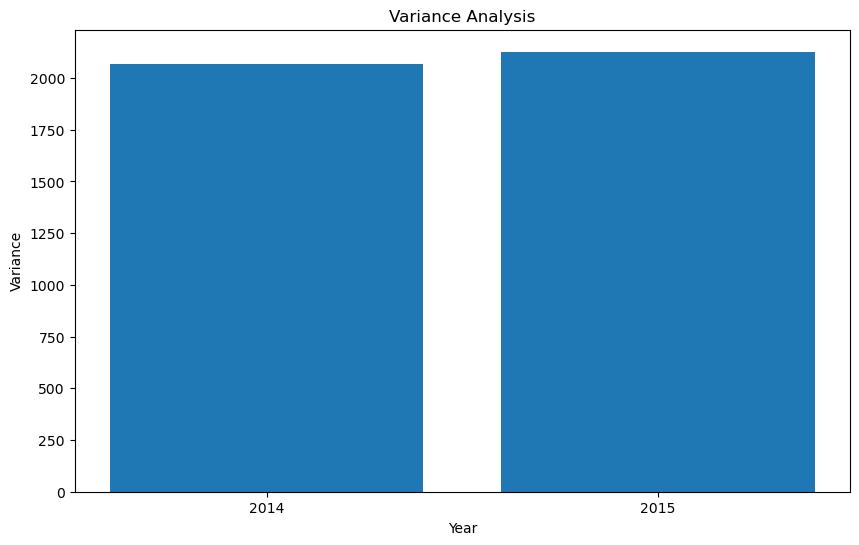


Figure 12: Austria Variance analysis Visualisation

**Ireland Heat Map and Correlation Analysis**

The Ireland correlation analysis is visualized below. Its correlation is between 0.75 and 1 and there is a positive correlation.

metin, fayans döşemeli, fayans, karo, çini, kiremit içeren bir resim

Açıklama otomatik olarak oluşturuldu

Figure 13: Ireland Heatmap Correlation Visualisation

**Austria Heat Map and Correlation Analysis**

According to the Austria data, there is not much correlation between each other. The correlation ratio is between 0.0 and 1 and it is a positive correlation.

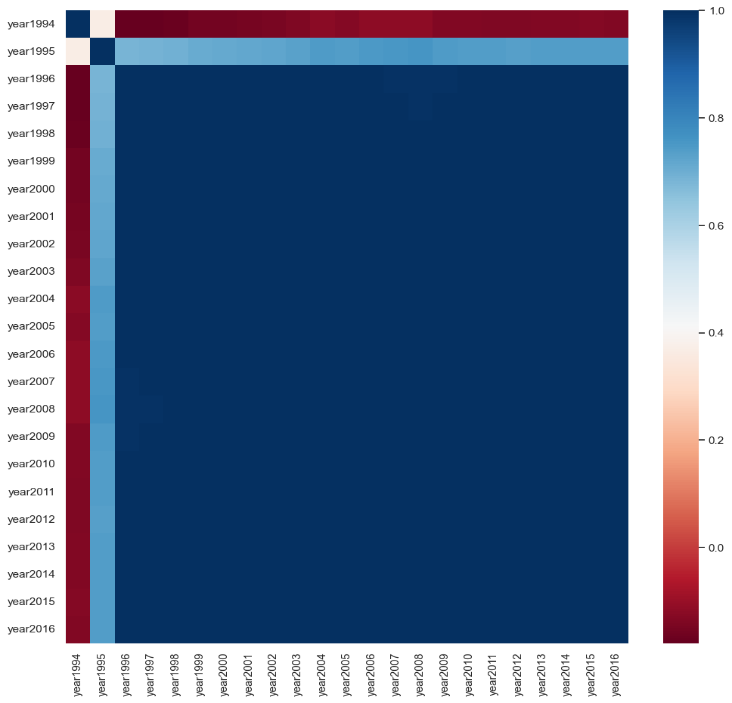


Figure 14: Austria Heatmap Correlation Visualisation

**Challenges**

Some problems were encountered while performing statistical calculations. One of these difficulties was the absence of numerical and categorical variables of the same type, and similar numerical variables were compared to eliminate this problem. Comparing the rates over the years gave more accurate results.

**Machine Learning for Data Analytics**

In order to apply Machine Learning models, first of all, the categorical variables in the data set must be converted to numerical variables. This is necessary for the dataset to model the data. Machine learning is perhaps the most important stage of a project because we can interpret the data set and make predictions about the future by making a conclusion with the models made in return for the visualizations and data preparation processes.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **IRELAND** | | | | **AUSTRIA** | | | |
| **1994** | **1995** | **1996** | **Cluster** | **1994** | **1995** | **1996** | **Cluster** |
| 109.200 | 113.500 | 115.900 | 2 | 0.000 | 0.000 | 61.000 | 2 |
| 109.300 | 114.100 | 115.700 | 2 | 0.000 | 0.000 | 0.000 | 1 |
| 109.400 | 114.100 | 115.900 | 2 | 61.000 | 0.000 | 0.000 | 0 |
| 110.300 | 114.100 | 115.900 | 0 | 62.100 | 0.000 | 0.000 | 0 |
| 110.100 | 114.100 | 115.900 | 0 |  |  |  |  |
| 110.600 | 114.100 | 116.000 | 0 |  |  |  |  |
| 110.700 | 114.500 | 116.000 | 0 |  |  |  |  |
| 110.900 | 114.500 | 116.000 | 0 |  |  |  |  |
| 112.200 | 115.700 | 116.900 | 1 |  |  |  |  |
| 112.700 | 115.800 | 117.100 | 1 |  |  |  |  |

Table 7: Ireland and Austria Cluster (2014-2015)

**Clustering**

Ideally, clustering is used to detect structures in a data . Since clustering is unsupervised, it works on data with no outcome (target) variable as well as that which no relationship between the observations is known. Clustering is useful in creating generalizations about groups in the data. During model selection, a silhouette score was used to determine the model with the optimal separation degree among the clusters.

According to the data, there are 3 clusters (0,1,2) and their corresponding values for each year are as follows:

**Ireland**

It is more important to know the specified criteria or algorithms in order to better evaluate the clustering output. According to the given data, it shows that each year has different values in itself.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Year** | **Cluster 0** | | **Cluster 1** | **Cluster 2** |
| **1994** | 110.300 | 112.200 | | 109.200 |
| **1995** | 114.100 | 115.700 | | 113.500 |
| **1996** | 115.900 | 116.900 | | 115.900 |

Table 8: Ireland Cluster 0,1,2

**Austria**

According to this information, each cluster can be evaluated separately for each year. Cluster 0 has values that are 0 in 1994, there are also non-zero values. Cluster 1 has a value of 0 in all of the years 1994-1995-1996. There is also a value of 0 in 1996 as well.

|  |  |  |  |
| --- | --- | --- | --- |
| **Year** | **Cluster 0** | **Cluster 1** | **Cluster 2** |
| **1994** | 61.000 | 0.000 | 0.000 |
| **1995** | 0.000 | 0.000 | 0.000 |
| **1996** | 0.000 | 0.000 | 61.000 |

Table 9: Austria Cluster 0,1,2

**Linear Regression**

**Linear regression:** This is a simple model that assumes a linear relationship between the input features and the output. It is often used as a baseline model for comparison with more complex models.

**Linear Regression Ireland 2014-2015**

Size of X\_train : (8, 33)

Size of y\_train : (8, 1)

Size of X\_test : (2, 33)

Size of Y\_test : (2, 1)

To get rid of the negatives values, we take the square, we want to eliminate the measurement problem. As a result, the success rate of 0.99 is high, the model learned well.

**Mean Squared Error:** 0.09952840319283998

R-squared ranges from 0 to 1, where a higher value indicates a better fit. An R-squared value of 0 means that the model does not explain any of the variance in the dependent variable, whereas an R-squared value of 1 means that the model explains all of the variance in the dependent variable. R-squared rate is 0.60 and this success rate is low and it is necessary to increase this rate.

**R-squared:** 0.6018863872286401

**Linear Regression Austria 2014-2015**

Size of X\_train : (6, 22)

Size of y\_train : (6, 1)

Size of X\_test : (2, 22)

Size of Y\_test : (2, 1)

As a result, the success rate of 0.94 is high, the model learned well.

**Mean Squared Error:** 0.9491741076679562

R-squared ranges from 0 to 1, where a higher value indicates a better fit. An R-squared value of 0 means that the model does not explain any of the variance in the dependent variable, whereas an R-squared value of 1 means that the model explains all of the variance in the dependent variable. As a result, the R-squared value is 0 and the dataset model has not learned enough.

**R-squared:** 0.0

**Discussion and Conclusion**

As a result, many methods were used to analyze the dataset and these results were shared.

It was analyzed and written in the report to get the most accurate results using methods suitable for the dataset. National Housing Construction Costs of Ireland and Austria were compared between the two countries and estimates were made by applying Machine learning models. It was also analyzed with European data and visualized geographically and the results were shared in the report.

**Dashboard Overview**

2 separate dashboards were prepared for 2 countries and displayed as output in the images below.

**metin, ekran görüntüsü, renklilik, diyagram içeren bir resim

Açıklama otomatik olarak oluşturuldu**

Figure 15: Dashboard Overview 1

**metin, ekran görüntüsü, diyagram, ekran, görüntüleme içeren bir resim

Açıklama otomatik olarak oluşturuldu**

Figure 16: Dashboard Overview 2

**metin, ekran görüntüsü, çizgi, ekran, görüntüleme içeren bir resim

Açıklama otomatik olarak oluşturuldu**

Figure 17: Dashboard Overview 3

**Github link :** [**https://github.com/candanbayar/2022352\_ID\_MSC\_DA\_Integr\_Repeat\_Sem1\_Statistics\_**](https://github.com/candanbayar/2022352_ID_MSC_DA_Integr_Repeat_Sem1_Statistics_)

**References**

[1] <https://w3.unece.org/PXWeb/en/CountryRanking?IndicatorCode=8> <https://data.gov.ie/dataset/beq03-indices-of-total-production-in-building-and-construction-sector-base-2010100/resource/1d57e24f-b6d2-47f1-b557-04d09922068b>

[2] <https://www.oenb.at/en/Statistics/Standardized-Tables/Prices--Competitiveness/Sectoral-Price-Development/Construction-Price-Index-and-Construction-Cost-Index.html>

[4] <https://www.oenb.at/isaweb/report.do?lang=EN&report=6.5>

[5] <https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Construction_producer_price_and_construction_cost_indices_overview>

[6] <https://data.gov.ie/dataset/national-house-construction-cost-index/resource/188485c4-e5d7-4406-996b-1ff12a9a045a>

[7] <https://medium.com/analytics-vidhya/how-to-visualize-pandas-descriptive-statistics-functions-480c3f2ea87c>

[8] <https://medium.datadriveninvestor.com/visualising-geospatial-data-with-python-d3b1c519f31>

[9] <https://www.sfu.ca/~mjbrydon/tutorials/BAinPy/01_intro.html>

[10] <https://towardsdatascience.com/a-beginners-guide-to-data-analysis-in-python-188706df5447>

[11] <https://machinelearningmastery.com/machine-learning-in-python-step-by-step/>

[12] <https://pyimagesearch.com/2019/01/14/machine-learning-in-python/>

[13] <https://www.w3schools.com/python/python_ml_getting_started.asp>

[14] <https://www.freecodecamp.org/learn/machine-learning-with-python/#tensorflow>

[15] <https://data-science-blog.com/blog/2016/04/26/machine-learning-mit-python-tutorial-minimalbeispiel/>

[16] <https://link.springer.com/book/10.1007/978-1-4842-2866-1>

[17] <https://www.packtpub.com/product/python-machine-learning-third-edition/9781789955750>