**CCT College Dublin**

**Assessment Cover Page**

|  |  |
| --- | --- |
| **Module Title:** | *Statistics for Data Analytics*  *Data Preparation & Visualisation*  *Programming for Data Analytics*  *Machine Learning for Data Analysis* |
| **Assessment Title:** | *MSc\_Data\_Analytics\_CA2* |
| **Lecturer Name:** | *John O’Sullivan*  *David McQuaid*  *Sam Weiss*  *Muhammad Iqbal* |
| **Student Full Name:** | *Mahmut Zahid Malkoc* |
| **Student Number:** | *2023034* |
| **Assessment Due Date:** | *26/05/2023* |
| **Date of Submission:** | *26/05/2023* |

**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. |

**ABSTRACT:**

This study aims to examine the impact of employment number of persons and building permits on Construction Producer Prices (CPP). The analysis is based on the examples of the United Kingdom (UK), Turkey (TR), Ireland (IE), and the European Union (EU) countries. The objective of the study is to evaluate the relationship between Construction Producer Prices(CPP), Building Permits(BP), and Employment Number of Persons(ENOP) factors and analyze this relationship on a yearly basis.

In this study, macroeconomic indicators such as employment number of persons, as well as building permits which are crucial for the vitality of the construction sector, are taken into consideration. The data is obtained from the Eurostat and covers the period from 1996 to 2022.

The analysis results reveal the impact of employment and building permit numbers on CPP. An increase or decrease in employment and building permits are found to have a positive or negative effect on the demand in the construction sector and consequently influences CPP in a manner. Similarly, an increase in building permits encourages activities in the sector and contributes to the rise in CPP. However, downturns in employment and building permits due to economic fluctuations, policy changes, and other factors can adversely affect CPP.

In conclusion, this study emphasizes the dependence of CPP on employment and building permits and demonstrates how these factors affect prices in the construction sector. These findings can be taken into consideration by governments when formulating supportive policies for the construction sector and regulating the market. Additionally, this analysis can provide valuable insights to construction companies in understanding industry trends and determining their future strategies.

**INTRODUCTION: Construction Sector Regarding Ireland**

Analysis of the Increase or Decrease in the Construction Producer Prices According to the Number of Employment and Building Permits on a Yearly Basis considering Ireland (IE), United Kingdom (UK), European Union (EU), and Türkiye (TR).

Overall, this assessment was carried out on the construction sector in Ireland, the United Kingdom, European Union, and Türkiye provided that three different construction data were selected from Eurostat, the European Union open data portal.

The Building Permits dataset is based on 2015=100 from 1992 to 2022. The Employment Number of Persons dataset is based on 2015=100 from 1990 to 2020. The Construction Produce Prices dataset is based on 2015=100 from 1990 to 2022. In these three different datasets, there are data for Ireland, the United Kingdom, the European Union, and Türkiye.

**PROGRAMMING FOR DATA ANALYTICS:**

* **Programming:** Code and libraries suitable for the project are added to the top line of each cell where necessary.
* **Data Structures:** In this study, three different data distinct formats were used. These are .CSV file, .xlsx file and the web API in JSON format respectively.
* **Documentation:** The project documentation explains the rationale and explanation of the code choices where appropriate.
* **Testing&Optimisation:** In this study, code and programming are documented in appropriate sections where necessary.
* **Data Manipulation:** In this project, pandas and numpy libraries were actively used while extracting data from the data source in .csv and .xlsx formats. While extracting data in .xlsx format, the openpyxl library could be used, but with the pandas library, this process was handled with a much shorter process. Another data source is 'request' and 'pyjstat' libraries in JSON format via Web API. In the last stage, these three different datasets could be made with the 'concat' or 'merge' method in the pandas library. This was carried out simply with 'merge'.

**DATA PREPARATION AND PROGRAMMING:**

* As a result of the research, the data sets were obtained from Eurostat, the open data portal of the European Union. This data is gathered in three different formats as '.csv', '.xlsx' and 'JSON'. Since all of these datasets are received by the open data portal, they are licensed in accordance with open source licensing.

The open source license for all datasets used is below:

https://opendatacommons.org/licenses/odbl/odbl-10.txt

* According to the Exploratory Data Analysis stages, the datasets and the performed stages are as follows step by step.

**STATISTICS FOR DATA ANALYTICS:**

* Descriptive statistics are central predictive measures and dispersion measures with the 'description' and 'skimpy' methods as follows.
* Estimation is made by associating with Construction Produce Price in Building Permits and Employment Number of Persons datasets.
* Five different inferential statistical methods were used. These are t-test, , ANOVA, Wilcoxon Test, Chi-Squared Test, Corelation Test and OLS Regression Test respectively.
* During the dataset research process, there were difficulties in selecting the dataset in order to compare the correct attributes. Therefore, when these dataset selections are completed, the resulting dataframe is a numerical dataset. Some inferental statistical models cannot be applied because there is no categorical attribute in the data set. For this reason, the Chi-Squared Test could not be applied and it was explained as such.

**DATA VISUALIZATION AND PROGRAMMING:**

**MACHINE LEARNING FOR DATA ANALYTICS:**