**MSc in Data Analytics – Integrated continuous Assessment 2**

Valentina Martucci SBS23006

[sbs23006@student.cct.ie](mailto:sbs23006@student.cct.ie)

26th May 2023

**Abstract**

*This work seeks to provide research about classification and prediction methods in regard to the Irish construction sector between 1997 and 2016. The key area is the labour cost represented mainly by construction employees’ wages, in relation to additional factors such as number of enterprises, construction types and production volume. Additionally, an analysis has been performed to compare the Irish labour cost with the other European countries, and an executive dashboard has been produced. Data from the Irish department of housing as well as Central Statistics Office were used in a selection of machine learning algorithms to produce these forecasts. Sentiment analysis has been performed in relation to the Ireland housing cost, the results of which are presented and discussed in Section 3 below. Finally plans for possible future extensions to this work are outlined in the conclusion.*

# Introduction

# Materials and Methods

## Methods and Tools

## Data

# Results and Discussion

## 3.1 Ireland construction sector

3.1.2 Inferential Statistic

3.1.3 Comparison within EU

## 3.2 Sentiment Analysis

## 3.3 Models

# Conclusion

# Introduction

The construction sector plays a significant role in the Irish economy, contributing approximately 7% to the country's Gross Domestic Product (GDP) and employing over 140,000 people as of 2021 (Central Statistics Office, 2021).

Data analytics has become an increasingly important tool in the construction sector, allowing stakeholders to gather and analyze data to make informed decisions and improve processes. This research paper aims to explore the use of data analytics in the Irish construction sector, examining its current state, identifying key challenges, and comparing it to other European countries.

Specifically, this research will focus on the labour cost and wages of construction sector employees over time, as well as the construction of residential buildings and the people’s sentiment on social media in regard to it. The insights gained from this research may be useful for policymakers, academics, and the general public in developing effective strategies to address the challenges facing the Irish construction sector.

This document is structured as follows. Section 2 describes the tools used during this research also providing information about methodologies approach. Section 3 provides details coming for the analysis of relevant datasets as well as data scraped from social media. Conclusion for this work are then summarised in Section 4.

# Materials and Methods

## 2.1 Methods and tools

For this work the phases defined by the Cross-Industry Standard Process for Data Mining, as known as CRISP-DM (Smart Vision Europe, 2017), were implemented in the following manner:

**Business & Data Understanding**: the starting point was analysing the brief, clarifying the objective of the work. The relevant data source has been identified and available datasets have been reviewed and selected. This work is detailed in section 1.

**Data Preparation:** All the data underwent a through E.D.A. and cleaning to prepare it for analysis. This work is detailed in section 2.

**Modelling & Evaluation:** the data were fit with a variety of machine learning modules depending on the desired outcome or result. Results were evaluated and new iterations have been performed as needed to make predictions. This work is detailed in section 3.

**Deployment:** This document and its supporting documents (data and code files) represent the deployment of the work, along with the creation of an executive dashboard.

Python notebooks have been primarily used to handle datasets. The open-source project called Jupyter (Jupyter.org) has been used to execute the supporting python code for this analysis.

Report’s notebooks, along with datasets and report can be found at <https://github.com/sbs23006/MsC_DA_CA2>

The report wordcount (including titles, references, and all sections) is xxx.

## Data

The following datasets were analysed for this research:

BAA12.20230506T200513.csv (data.cso.ie, 2023)

BBA02.json (data.gov.ie, n.d.)

BEA04.20230506T200502.csv (data.cso.ie, 2023)

estat\_lc\_lci\_lev\_en.csv (Europa.eu, 2021)

While the following one has been produced during the analysis:

data.csv

Data sourced from Eurostat has been used under their free re-use of data policy (ec.europa.eu, n.d).

Data sourced from CSO has been used under the CSO data policy for researchers ([www.cso.ie](http://www.cso.ie), n.d).

Data sourced from Ireland Department of Construction has been sourced under the Open Data Directive (data.gov.ie, n.d.).

# Results and Discussion

The goal of this section is to provide an overview of the conducted analysis along with the main highlights.

## 3.1 Ireland construction sector

The first step of this analysis has been to explore the datasets to get some initial insights into the Irish construction sector, along with the identification of a suitable dataset for further work in relation to machine learning models. The code for this work can be found in the accompanying Jupyter notebooks *Analysis.ipynb* and *Ireland Analysis + Eu comparison.ipynb*.

Visualization tools like matplotlib and seaborn were useful to plot data in regard to the type of building and construction in Ireland over the years (Figure 1), along with the average index of employment in building and construction industry (Figure 2).

A picture containing text, plot, line, font

Description automatically generated

Figure 1 – Building and Construction types between 2000 and 2022

A picture containing line, plot, diagram, screenshot

Description automatically generated

Figure 2 – Average employment index in building and construction between 1975 and 2008

It’s noticeable from the two plots above the decreased values for building and construction from 2009, specifically for residential buildings. In the same year a drop in the index of employment is observed.

Eurostat data (Europa.eu, 2021) provided comprehensive data related to the labour cost structure where the values showed a peak in non-wages related costs in 2022. Figure 3 shows a representation of those values, with the dot size being related to the value; both the percentage of non-wage cost and costs other than salaries have their highest value in 2022. *Plotly Express* has been used to create this interactive dashboard, a high-level data visualization package that allow the creation of interactive plots (plotly.com).

Labour cost trends can also be quickly spotted in Figure 4.

A picture containing text, screenshot, diagram, plot

Description automatically generated

Figure 3 - Representation of Irish labour cost structure

A picture containing line, plot, diagram, slope

Description automatically generated

Figure 4

### 3.1.2 Inferential Statistic

Text text

### 3.1.3 Comparison within EU

## 3.2 Sentiment Analysis

## 3.3 Models

# Conclusion

The sector has undergone significant changes since the economic downturn in 2008, which led to a decrease in demand for construction services and a subsequent reduction in construction activity (Forde et al., 2020). However, recent years have seen a resurgence in the construction sector, with strong growth predicted in the coming years (Construction Industry Federation, 2021).

Construction Industry Federation. (2021). Construction industry forecasts. Retrieved from https://cif.ie/insight-and-analysis/cif-construction-industry-forecasts/

Forde, K., Lynskey, M., & Doherty, E. (2020). The impact of the global financial crisis on the Irish construction industry. Journal of Financial Management of Property and Construction, 25(1), 68-82. doi: 10.1108/JFMPC-02-2019-0011.

## References

Central Statistics Office. (2021). *Construction statistics*. Available at: <https://www.cso.ie/en/statistics/construction/> [Accessed ]

Europa.eu. (2021). Available at: <https://ec.europa.eu/eurostat/databrowser/view/lc_lci_lev/default/table?lang=en> [Accessed 14 May 2023].

data.gov.ie. (n.d.). BBA02 - *Annual Average Index of Employment (1975 - Date) in Building and Construction Industry - data.gov.ie*. [online] Available at: https://data.gov.ie/dataset/bba02-annual-average-index-of-employment-1975-date-in-building-and-construction-industry?package\_type=dataset [Accessed 14 May 2023].

data.cso.ie. (2023). BAA12 - *Construction Enterprises*. Available at: https://data.cso.ie/table/BAA12 [Accessed 14 May 2023].

data.cso.ie. (2023). BEA04 - *Indices of Total Production in Building and Construction Sector (Base 2015=100)*. Available at: https://data.cso.ie/table/BEA04 [Accessed 14 May 2023].

www.cso.ie. (n.d.). *Policies - CSO - Central Statistics Office*. [online] Available at: https://www.cso.ie/en/aboutus/lgdp/csodatapolicies/dataforresearchers/policies/#d.en.210341 [Accessed 14 May 2023].

ec.europa.eu. (n.d.). *Copyright notice and free re-use of data - Eurostat*. [online] Available at: https://ec.europa.eu/eurostat/en/web/main/about/policies/copyright [Accessed 14 May 2023].

data.gov.ie. (n.d.). *Open Data Directive - data.gov.ie*. [online] Available at: https://data.gov.ie/pages/open-data-directive [Accessed 14 May 2023].

Smart Vision Europe (2017). *What is the CRISP-DM methodology?* [online] Smart Vision - Europe. Available at: <https://www.sv-europe.com/crisp-dm-methodology/>.

Jupyter.org. (2019). *Project Jupyter*. [online] Available at: <https://jupyter.org>. [Accessed 14 Apr. 2023].

plotly.com. (n.d.). *Plotly Express*. [online] Available at: https://plotly.com/python/plotly-express/.

‌

‌

‌

‌

‌

‌

‌

Data Science Process Alliance. (n.d.). *Data Science Methodologies and Frameworks Guide*. [online] Available at: https://datascience-pm.com/data-science-methodologies [Accessed 14 Apr. 2023].

Mario Dobler, ‌Tim Großmann. (2020). *The Data Visualization Workshop: A Self-Paced, Practical Approach to Transforming Your Complex Data Into Compelling, Captivating Graphics*. Vol Third edition. Packt Publishing.