MSc in Data Analytics – Integrated continuous Assessment 2

Valentina Martucci SBS23006
sbs23006@student.cct.ie

26th May 2023

Abstract

This work seeks to provide research about the Irish construction sector between 1997 and 2016, using techniques such as data analysis, statistics, and machine learning. 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. Sentiment analysis has been performed in relation to the Ireland housing cost, the results of which are presented and discussed in Section 3 below. Regression models have been used for forecasting of wages, and classification models have been developed to classify social media posts. Finally, plans for possible future extensions to this work are outlined in the conclusion.

Table of Contents

1. 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 Irish construction sector using data analytics, 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. Conclusions for this work are then summarised in Section 4.

2. 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 sections 2.1 and 3.

Data Preparation: All the data underwent a through E.D.A. and cleaning to prepare it for analysis. Initial insights have been visualized. This work is detailed in section 3.

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.

Four Jupyter notebooks have been produced for this study and, as this is a data analysis project, modules like *pandas, matplotlib, sns* and other have been used for data manipulation and visualization. To maximise code reutilization for this work, several functions have been defined in a file called *utils.py* that to allow code reutilization in all the notebooks.

As every data project, big part of the work has focused on data manipulation, in this case sourced datasets were in different formats (csv and json). Aggregation methods in Python are essential for processing and manipulating data from multiple data structures. Some of the commonly used aggregation methods include Pandas (reading, cleaning, and analyzing structured tabular data), NumPy (useful for numerical computing), and others like Apache Spark and Dask that use parallel computing and are better suited for large datasets (not the case for this project). Another option that has been explored for this work has been Polars (www.pola.rs), a python library to explore DataSets, and some performance tests have been ran to compare it with Pandas. Results are summarized in the table in Figure 1 below. Pandas performed better in data manipulation, while Polars is slightly better when loading the csv file. Considering the minimum difference and the lack of Polars developer communities, Pandas has been preferred for this work.

Operation	Library	Result	Winner
Load csv	Pandas	0.0040	Polars
	Polars	0.0020	
Print dataframe	Pandas	0.0000	Equal
	Polars	0.0000	
Aggregate	Pandas	0.0010	Polars
	Polars	0.0000	
Sort	Pandas	0.0000	Pandas
	Polars	0.0010	
Filter	Pandas	0.0000	Pandas
	Polars	0.0010	

Figure 1 – comparison between Pandas and Polars

The first step has been storing the sourced data in a Pandas Dataframe object: for the csv datasets, pandas built-in *read_csv* functionality has been used to do so, while for json datasets the approach used has been downloading the json file using requests module, and then open the file within the code and store its content in a pandas dataframe; *json.loads()* method has been used as it allows easy parsing of JSON strings and convert them into a Python Dictionary. (GeeksforGeeks, 2020) Pandas *read_json* function has been also evaluated, but as the downloaded dataset didn't have a nested, tabular structure, *json.loads()* provided a more lightweight approach and with few lines of code the dataframe was ready to be used.

Most of this work can be observed in the accompanying Jupyter notebook *Analysis.ipynb*.

Another important aspect when dealing with code is making sure that the code does what expected. There are several testing techniques with the main ones being unit testing (individual units of code), integration testing (interaction between multiple components), functional testing (testing a full functionality) and code review (ask peers to review the code to identify errors and improve quality). Integration testing is better suited for projects with a large code base, for example a distributed system where several microservices interact between themselves. As this is a data analysis project the scripting aspect of Python has been used (no OOP components or server applications have been created), as the goal is to use coding to support the analysis, not to create web services.

Because of all those reasons, unit testing has been used (constant use of *df.head()* or *print*), as well as functional when possible (for example, in the file *Sentiment Analysis.ipybn* tests have been ran to make sure the functions returned the expected output).

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 excluding Appendix) is 3480.

2.2 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, n.d).

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

3. 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*(section 1) and *Ireland Analysis + Eu comparison.ipynb* (section 2).

Visualization tools like matplotlib and seaborn have been used to plot data about 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).

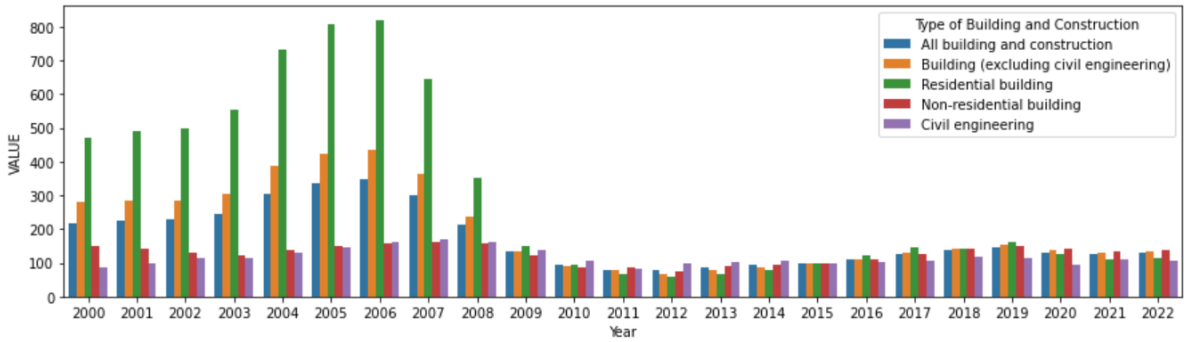


Figure 2 – Building and Construction types between 2000 and 2022

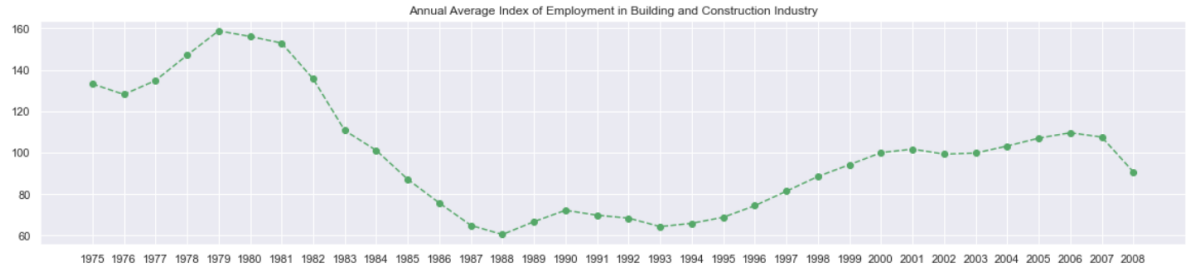


Figure 3 - Average employment index in building and construction between 1975 and 2008

It's noticeable from the two plots above the decreased values for building and construction from 2008, specifically for residential buildings. In the same year a drop in the index of employment is observed. This can be explained as a consequence of the 2008 great recession and the property bubble that saw a combination of increased speculative construction and rapidly rising prices (Malzubris, 2008).

When it comes about the labour data, analysing the values from the Enterprise dataset, a drop of 60.2% is observed between 2008 and 2013 in the number of employees, demonstrating the impact on the labour market.

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).

The discussed labour cost trends can also be quickly spotted in Figure 4.

Figure 4 - Representation of Irish labour cost structure

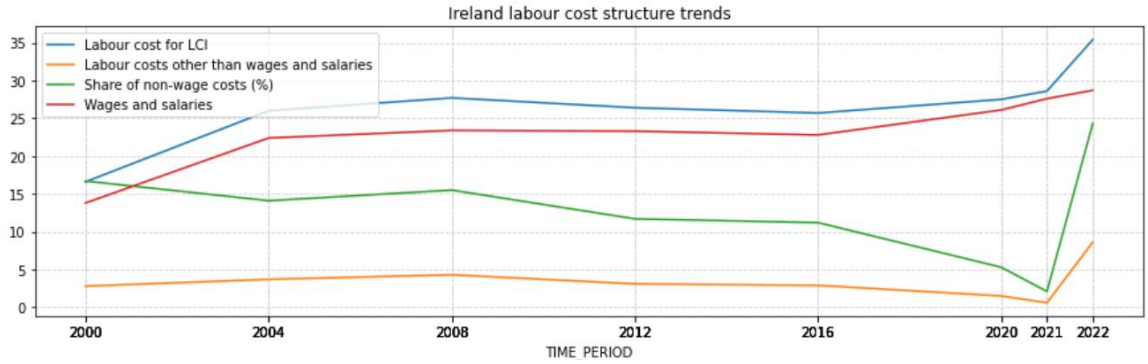


Figure 5

An executive dashboard has been created using Dash, an open-source Python framework used for building analytical web applications (dash.plotly.com). As Jupyter has been used to run the Python code for this project, *jupyter-dash* module has been installed and used for this project (dash.plotly.com). The reason behind the choice of dash and plotly is the dashboard scalability, being easily accessible on different screens as well as the possibility for the user to interact with the data. The first step has been the actual definition and creation of the single figures, and then combine them in the app layout along with callback definitions. Full code for this can be found in the file *Analysis.ipynb* (section 1.2). The Figure 5 below shows a quick look of the dashboard, a user guide with additional details is available in the Appendix.

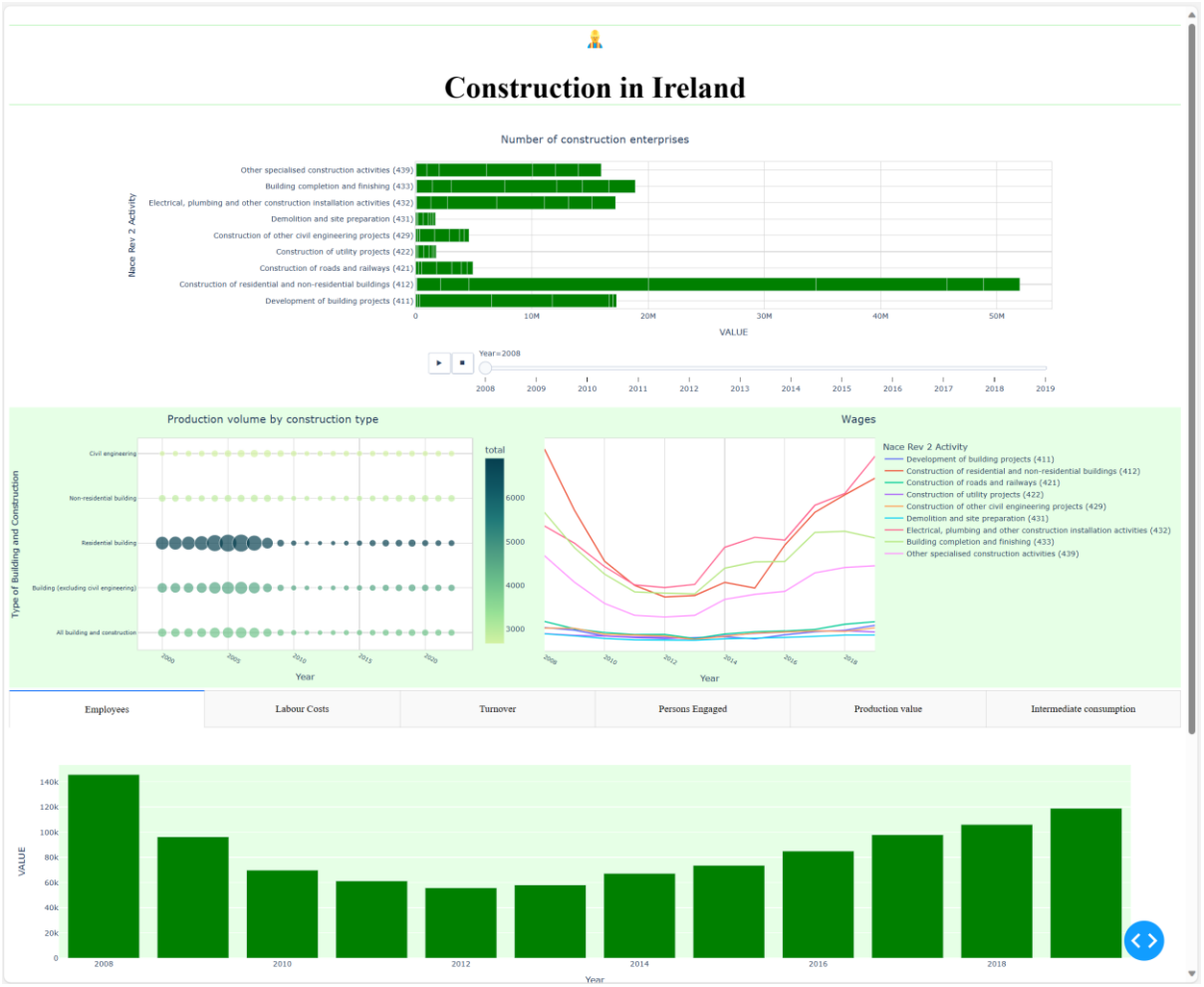


Figure 6 - Ireland construction report

Both for the dashboard and any visualization of this work, an attempt was made to follow Tufte's guiding principles of data visualization (Tufte, 2001). These can be summarized as follows:

- Above all else show the data
- Maximise the data ink ratio
- Erase non-data ink
- Erase redundant data ink
- Revise and edit

In some cases, such as Figure 4, an editorial decision was made to include elements which Tufte would recommend against. In this example, grid lines have been included in the plot to provide a clearer understanding of the value represented for each year. A choice was made tough to change the default grid colour to grey in an attempt to erase non-data ink (same as Figure 6, where the grid colour is light grey).

3.1.2 Inferential Statistic

As discussed before, this study will mainly focus on labour data and constructions type. Inferential Statistic has been used to gain insights on both populations and find confidence intervals.

The dataset provides data between 2008 and 2019, so this can be considered our sample. When calculating the average of this sample, there's always some uncertainty as it might not represent perfectly the entire population. Confidence intervals are a way to measure that uncertainty, as they present the estimate average not as a single number but as a range (Bruce, Bruce and Gedeck, 2020).

To produce the confidence intervals for the plots of simulated IQ data above, *t*, *sem*, and *mean* functions available in the *scipy.stats* package have been used. Another option is to use the *tconfint_mean* function from the *statsmodels* package. Both methods give nearly identical results.

The table below summarizes the findings about both populations for the years between 2008 and 2019.

Population	**Calculated Mean**	**Interval (95% confidence)**
Wages and Salaries	4043069.16	(3049317.15, 5036821.18)
Employees	86280.66	(68739.28, 103822.05)

Table 1

Hypothesis testing techniques have been applied to get additional insights into the wages of employees in the construction industry. T-test has been used to validate the null hypothesis of Irish wages being equal to 50000, that has been rejected.

The construction type population has been also analyzed for inferential statistics, a sample of 10 items has been taken for each construction type. Unfortunately, Anova condition have not been met (distribution is normal and looks like there's some dependency between population).

Code for those calculations can be found in the file *Analysis.ipybn* (section 3).

3.1.3 Comparison within EU

Choropleth module from *Plotly Express* has been used to visualise Ireland values compared to other European countries, and statistical tests have been performed to compare it to specific countries.

Overall, Ireland hourly wages in construction have constantly been higher than the values in the Euro Area, with the difference being bigger from 2004 (Figure 6).

Figure 6

Figure 7 below shows that in 2022 Ireland is one of the countries with higher wages and salaries, and it's noticeable that the value decrease in southern Europe.

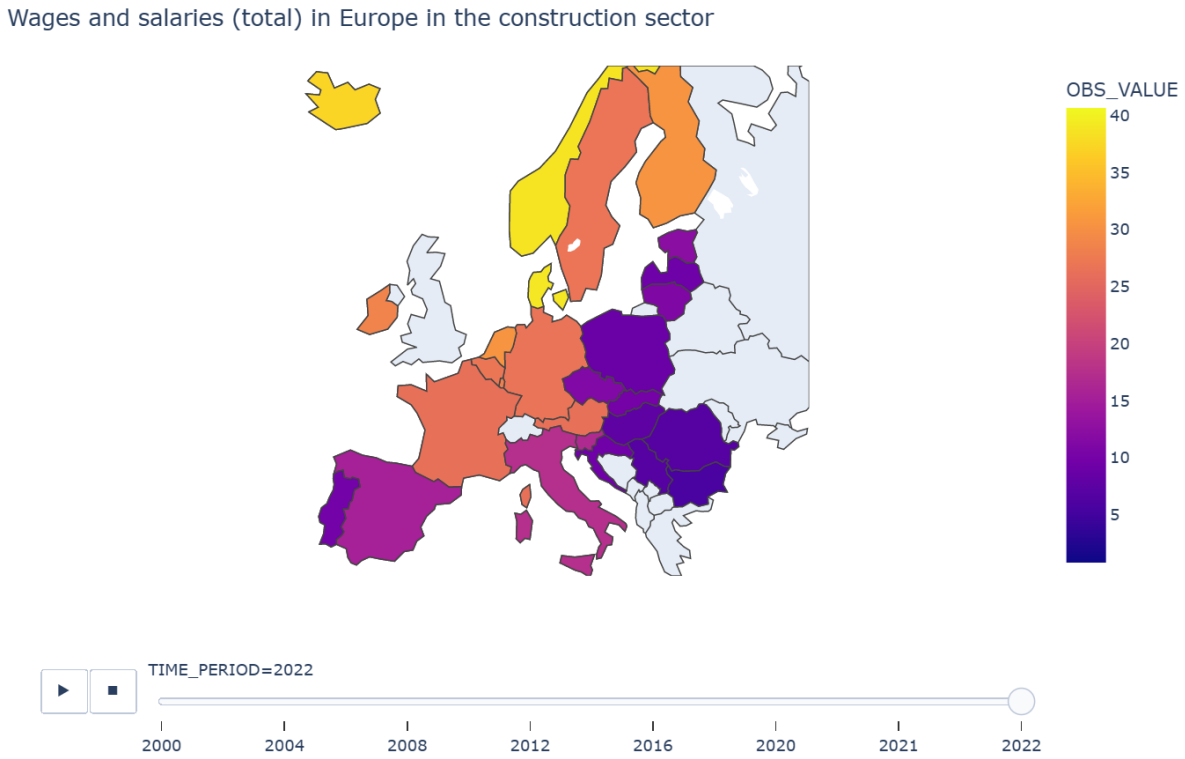


Figure 7 - Wages and salaries comparison in 2022

From a statistical perspective, Irish values seem to have similar distribution as Sweden, Germany, and Finland (Figure 7), so inferential statistic tests have been performed to compare those countries.

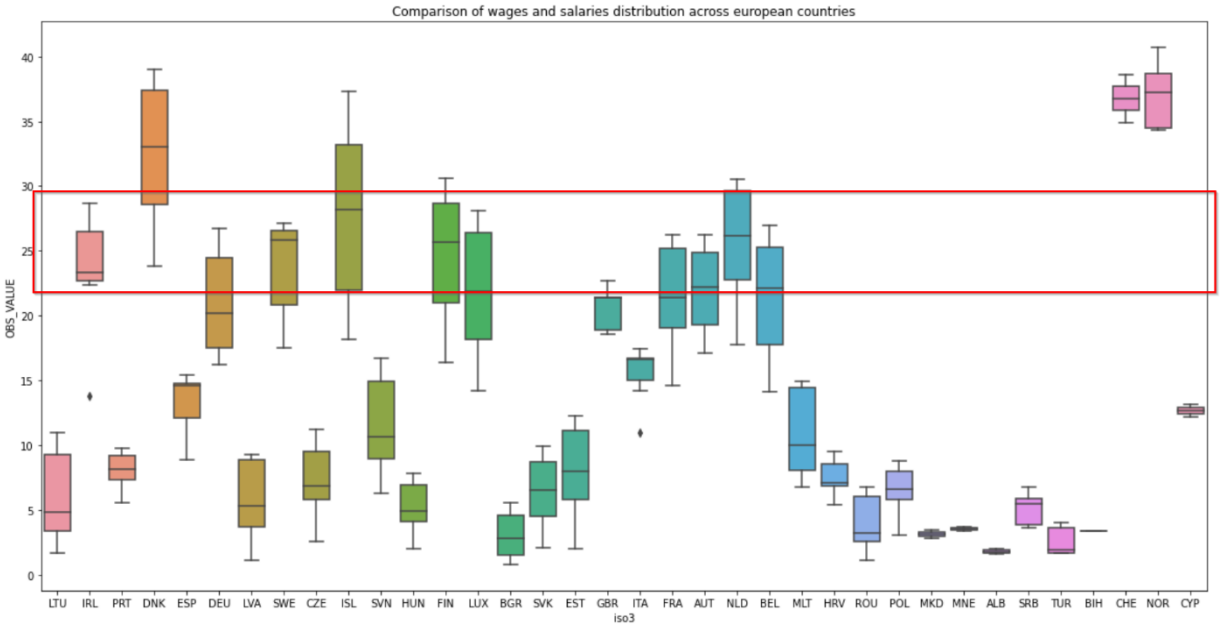


Figure 8 - Wages and salaries distribution across European countries

Two different approaches have been used to analyze and compare the data: parametric and non-parametric tests.

Parametric tests assume that data follow a normal distribution. Samples have been collected for wages and salaries of Ireland, Germany, Finland and Sweden, and Shapiro-Wilk test have been performed to check data normality. T-test confirmed that Sweden and Ireland's means are similar (confirmed also by Mann-Whitney), while ANOVA test failed to accept the null hypothesis for Ireland, Sweden, and Germany. *violinplot* was used to visualize the distribution of those three countries, confirming the ANOVA test's outcome (Figure 9).

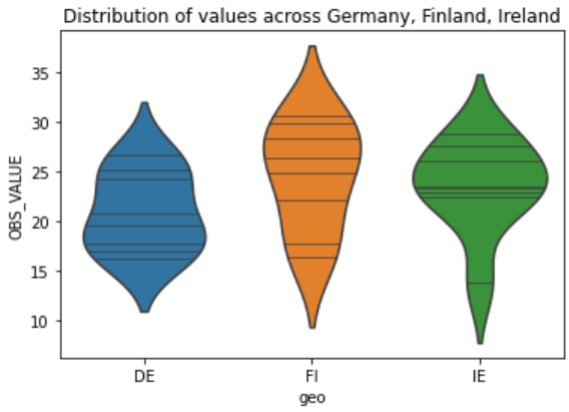


Figure 9 - Ireland vs Germany vs Finland

Wilcoxon test was used to compare years this time, across all countries, resulting in a difference between construction wages between those two years and indicating changes in the economics of the construction sector labour.

3.2 Sentiment Analysis

Ireland economy has drastically recovered since the 2008 crash but, as observed in Figure 1, the construction volume of residential building has not changed leading to a demand higher than supply. This topic is constantly being discussed in the dedicated Reddit channel '*r/ HousingIreland*' (reddit.com).

Python provides a powerful tool that allows quick interaction with Reddit APIs (praw.readthedocs.io). Reddit credentials are secret values, therefore their values are not hard-coded in the notebook but stored in a text file that won't be uploaded on Github using the *.gitignore* file (git-scm.com).

The first step of the analysis has been data gathering, a dataframe has then been built with the extracted reddit posts about housing in Ireland. Valencer Aware Dictionary and sentiment reasoner (VADER) provides a score given text determining if there is a positive, neutral or negative sentiment. A lambda function *analyze_sentiment* has been created to apply VADER classification to our dataset and add its value in a new variable called 'sentiment'. The distribution of sentiment after classification is as follows:

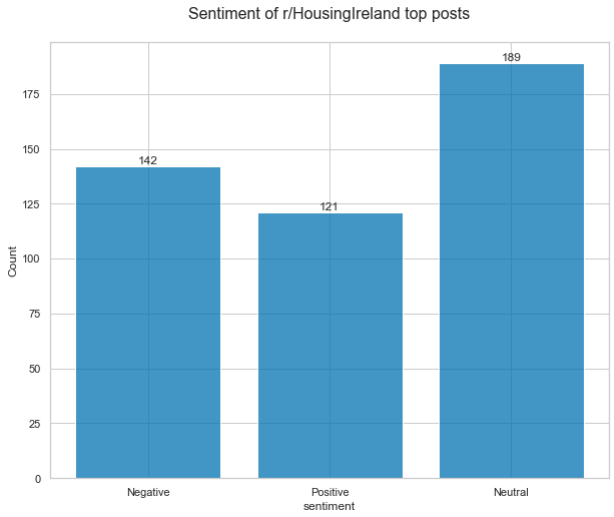


Figure 10

It's noticeable that there are more negative posts than positives, even if not too many, while 41.8% of the posts are classified as neutral. In addition to sentiment classification, Latent Dirichlet Allocation has been used to identify the top topics, along with a graphical representation using WordCloud.

#	Topic
1	irish market investment funds homes
2	crisis funds time ireland new
3	rent dublin ireland house new
4	house ireland protest prices government
5	homes new crisis government rent

Table 2 - Top 5 topics extracted using LDA

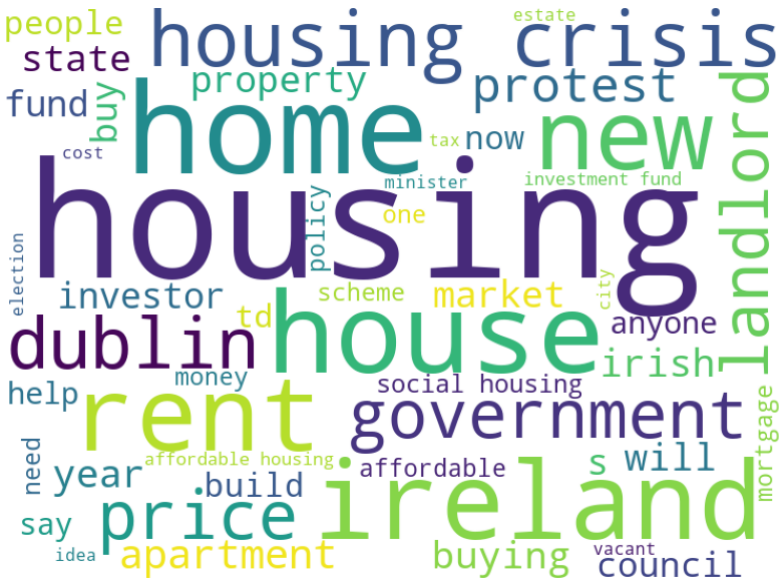


Figure 11 – WordCloud

It's evident that according to what users are debating on Reddit there is a housing crisis in Ireland, spanning across rent and house prices and with Dublin appearing in trending topics. Those insights were useful to scope the research further and analyze the polarity distribution for those two trending topics. Figure 12 below shows the visualizations where it is observed that the discussions about rent are more in comparison to purchase, with a normal distribution; on the other side, both Dublin and Ireland values seem lightly left skewed to the left, indicating most negative posts, with Dublin having more values on the negative side.

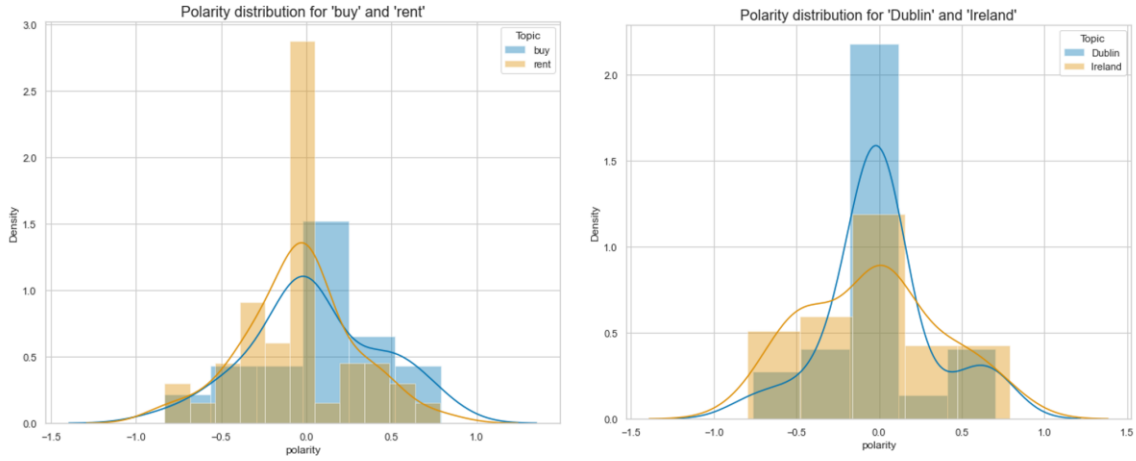


Figure 12 – polarity distributions

Text has been then processed to develop classification models to classify sentiment. Additional details can be found in the next section.

3.3 Models

This section will cover the classification models used on Reddit data, and regression models used to predict wages. Classification code can be found in the file *Sentiment Analysis.ipynb*, while steps to build regression model are in *Models.ipynb*.

3.3.1 Classification

As mentioned in the previous section, Latent Dirichlet Allocation (LDA) has been used to identify the top topics across the Reddit posts. LDA is an unsupervised machine learning model used for topic modelling; as such, it learns the topics without any labelled data and identifies them solely based on the patterns and relationships present in the data itself (Müller and Guido, 2017).

Before applying a supervised model, text data needs to be converted into numerical values and *CountVectorizer* has been applied. This technique, known also as 'bag of words' consists into converting text data into a matrix of token counts, the numerical representation captures the frequency of words in each document allowing the algorithm to learn from these and make classifications. The output of this operation resulted in a 452x1675 sparse matrix with 3114 stored elements, for high-dimensional, sparse data like this, the algorithm chosen is *LogisticRegression* that is known to work best with such data (Müller and Guido, 2017). The algorithm has been initially fit with processed and original text data and, surprisingly, the accuracy was better using original text. The second step

has been finding the best C parameter along with test split data to find the best parameter to build the model.

Another pre-processing technique has been used, called TF-ID (Term Frequency-Inverse Document Frequency) vectorizer, it considers the frequency of words and their importance across the entire document collection (Müller and Guido, 2017). TF-IDF values are non-negative and often skewed, they can be approximately treated as continuous and transformed into a more Gaussian-like distribution, therefore GaussianNB has been applied (Analytics Vidhya, 2021).

LogisticRegression resulted in being the best model, full results are available in the table below. Please refer to the accompanying Jupyter notebook for more information and visualizations.

Model	Accuracy
LogisticRegression with original text and CountVectorizer	0.58
LogisticRegression with processed text and CountVectorizer	0.55
GaussianNB using TfidVectorizer (original text)	0.38
LogisticRegression using TfidVectorizer (original text)	0.38

3.3.2 Regression

Before applying regression models, features have been extracted from the enterprise dataset analyzed in the section 3.1, null values removed, and columns renamed. Functions from the imported python file utils.py have been used to check outliers and features distribution, data has been scaled and visualizations techniques showed the linear relationship between target variable (wages and salaries) and the dependent variables. Additional details can be found in the accompanying Jupyter notebook Models.ipynb.

Several linear models have been explored and, from Figure below, Lasso seems the best.

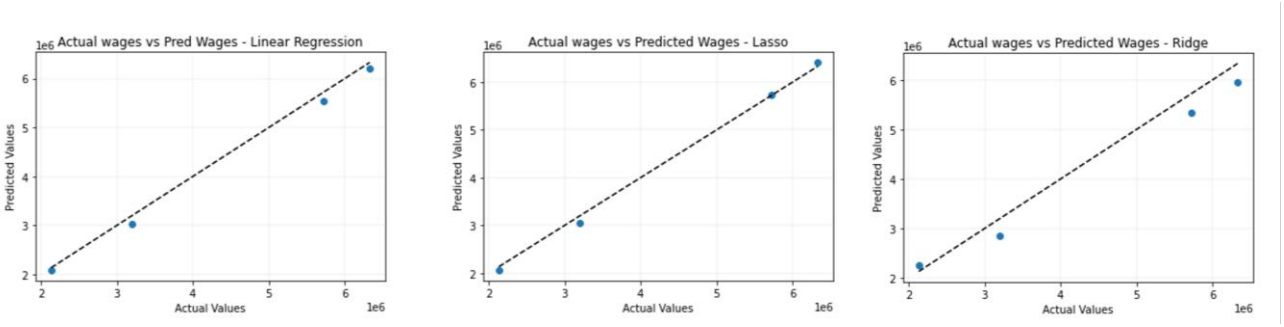


Figure 13 – algorithm predictions comparison

The table below shows the train and test scores for the applied models.

	Model	Train Precision	Test Precision
0	LinearRegression	1.000000	0.993536
1	Lasso	0.999696	0.997336
2	Ridge	0.971540	0.964517

Values are homogeneous with Ridge regression being slightly lower in test. Both training and testing accuracy are high and almost similar, meaning that there is no over-fitting or under-fitting. Increasing the dataset size could potentially impact those values considering the small shape of the dataframe used.

GridSearchCV has been used to identify the best algorithm and parameters to use for this scenario, with both original and scaled data. Ridge algorithm with alpha value equal to 5, has been selected as the best one to train on not scaled data.

4. Conclusion

In this work, valuable insights have been collected in regard to the Irish construction sector.

This sector in Ireland has undergone significant changes since the economic downturn in 2008, which led to a decrease in demand for construction services and a, as seen in Figure 1, 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). As the number of residential buildings production did not increase proportionally with the disposable income, a new situation of demand-supply imbalance rose over the last years in Ireland.

This topic is widely discussed on social media, and sentiment analysis of the Reddit channel *r/HousingIreland* showed a predominance of negative sentiment, with top topics being rent prices, housing crisis as well as discussions about Dublin. A classification model has been developed to further classify text posts coming from social media, reaching the best accuracy with LogisticRegression.

This work has then analyzed another aspect of the construction industry such as wages and salaries. Irish wages in this sector have constantly been higher than the Euro Area average between 2000 and 2020, with Irish hourly wages being equal to a total of 28.7€, with only Finland, Denmark, Norway and Netherlands having higher values according to Eurostat data (Figure 7). Coming back to Ireland data only, another interesting finding is the spike of labour costs other than wages and salaries between 2021 and 2022 (Figure 4); this can be explained with the rising of material costs that has been a challenge for many construction companies: 96% of companies reported a rise in the cost of building materials over the summer of 2022 (The Irish Times).

Features from Irish enterprise data have been selected, data has been analyzed and then prepared to fit a regression model to predict wages and salaries in Ireland. Due to the nature of the data, three algorithms have been applied resulting all in high test and train scores. Despite the good results, the small size of the data leaves the way open for further research into other forecasting methods using bigger samples, in an effort to achieve the same accuracy.

References

Central Statistics Office. (2021). Construction statistics. Available at: https://www.cso.ie/en/statistics/construction/ [Accessed 6 May 2023]

Europa.eu. (2021). Available at: https://ec.europa.eu/eurostat/databrowser/view/lc_lci_lev/default/table?lang=en [Accessed 6 May 2023].

data.gov.ie. (n.d.). BBA02 - Annual Average Index of Employment in Building and Construction Industry [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 6 May 2023].

data.cso.ie. (2023). BAA12 - Construction Enterprises. Available at: https://data.cso.ie/table/BAA12 [Accessed 6 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 6 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 6 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 6 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 6 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/. [Accessed 7 May 2023].

www.pola.rs. (n.d.). Polars, lightning-fast DataFrame library. [online] Available at: https://www.pola.rs/. [Accessed 7 May 2023].

Jupyter.org. (2019). Project Jupyter. [online] Available at: https://jupyter.org. [Accessed 7 May 2023].

plotly.com. (n.d.). Plotly Express. [online] Available at: https://plotly.com/python/plotly-express/. [Accessed 7 May 2023].

dash.plotly.com. (n.d.). Dash Documentation & User Guide | Plotly. [online] Available at: https://dash.plotly.com/ [Accessed 7 May 2023].

dash.plotly.com. (n.d.). Using Dash in Jupyter and Workspaces | Dash for Python Documentation | Plotly. [online] Available at: https://dash.plotly.com/workspaces/using-dash-in-jupyter-and-workspaces [Accessed 7 May 2023].

Tufte, E. (2001). The visual display of quantitative information, Cheshire: Graphic Press. – 2001

Bruce, P.C., Bruce, A. and Gedeck, P. (2020). Practical statistics for data scientists: 50+ essential concepts using R and Python. Sebastopol, Ca: O'reilly Media, Inc.

reddit. (n.d.). r/HousingIreland. [online] Available at: https://www.reddit.com/r/HousingIreland/ [Accessed 11 May 2023].

praw.readthedocs.io. (n.d.). PRAW: The Python Reddit API Wrapper — PRAW 7.4.0 documentation. [online] Available at: https://praw.readthedocs.io/en/stable/. [Accessed 11 May 2023].

git-scm.com. (n.d.). Git - gitignore Documentation. [online] Available at: https://git-scm.com/docs/gitignore. [Accessed 6 May 2023].

Muller, Guido. (2017) Introduction to Machine Learning with Python. O'Reilly Media,Inc

Analytics Vidhya. (2021). Sentiment classification using NLP With Text Analytics. [online] Available at: https://www.analyticsvidhya.com/blog/2021/09/sentiment-classification-using-nlp-with-text-analytics/. [Accessed 13 May 2023].

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.

Malzubris J. (2008). Ireland's housing market: bubble trouble. [online] Available at: https://ec.europa.eu/economy_finance/publications/pages/publication13187_en.pdf [Accessed 22 May 2023].

Construction Industry Federation. (2021). Construction industry forecasts. [online] Available at: https://cif.ie/insight-and-analysis/cif-construction-industry-forecasts/ [Accessed 24 May 2023].

The Irish Times. (n.d.). Construction sector warns rising costs will hit housing supply next year. [online] Available at: https://www.irishtimes.com/business/2022/10/05/construction-sector-warns-rising-costs-will-hit-housing-supply-next-year/. [Accessed 24 May 2023].

GeeksforGeeks. (2020). json.loads() in Python. [online] Available at: https://www.geeksforgeeks.org/json-loads-in-python/ [Accessed 20 May 2023].

Appendix – Dashboard user guide

This dashboard wants to serve as a single-entry point to quickly get insights on Irish construction trends over the years. Colour palette chosen is green, to symbolize Irish national colour. This section will provide a description of each graph along with interaction instructions.

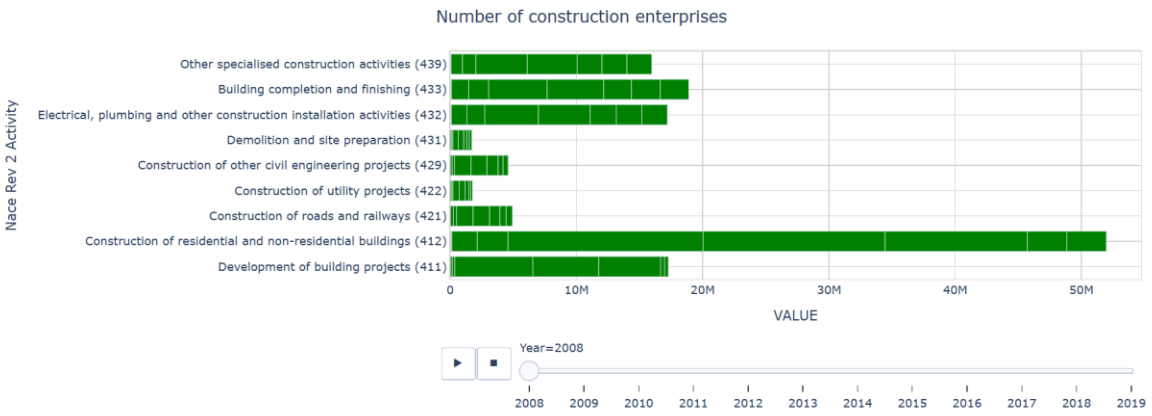


Figure 14

The first graph (Figure 14) provides an overview about the number of construction enterprises in Ireland, grouped by NACE code, a pan-European classification system that groups organisations according to their business activities (gov.ie).

To visualize data for a specific year, the bottom slider can be used to select it, or simply using the play button will create an animation across the years. Thanks to Plotly Express, it's also possible to hover on the bar to see the actual value.

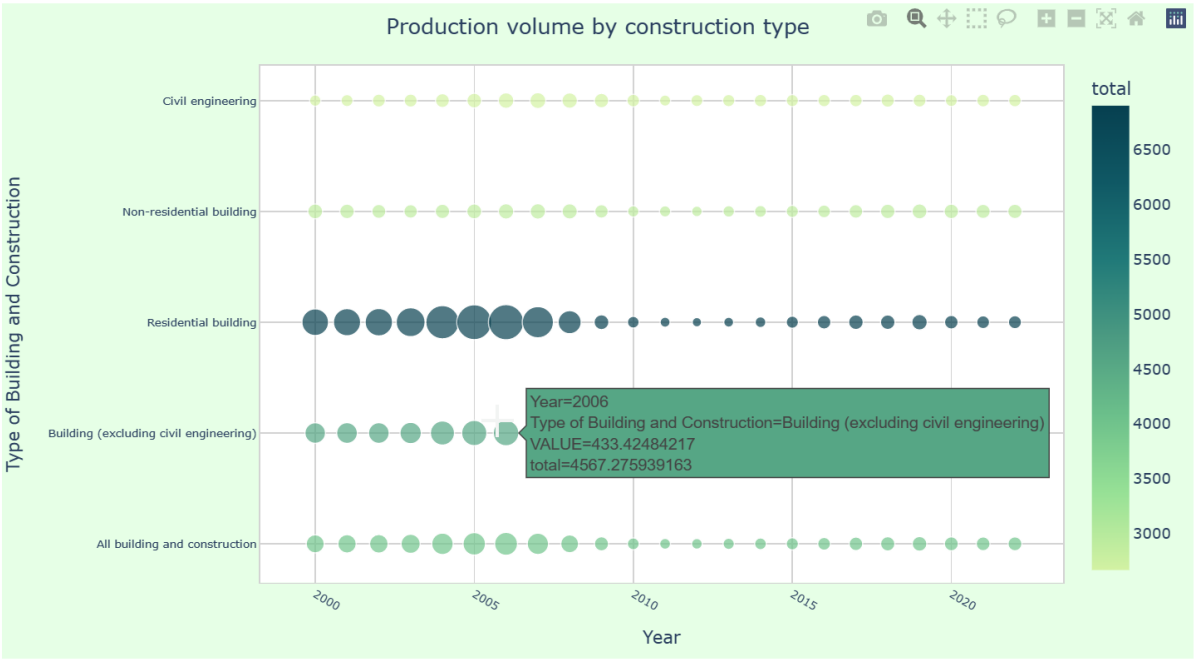


Figure 15

Scatterplot has been used to visualize the production value by construction type over years, this has been a natural choice as two variables are being compared. The dot size is correlated to the actual value, as well as the colour intensity. As shown in Figure 15, hover functionalities are provided here as well.

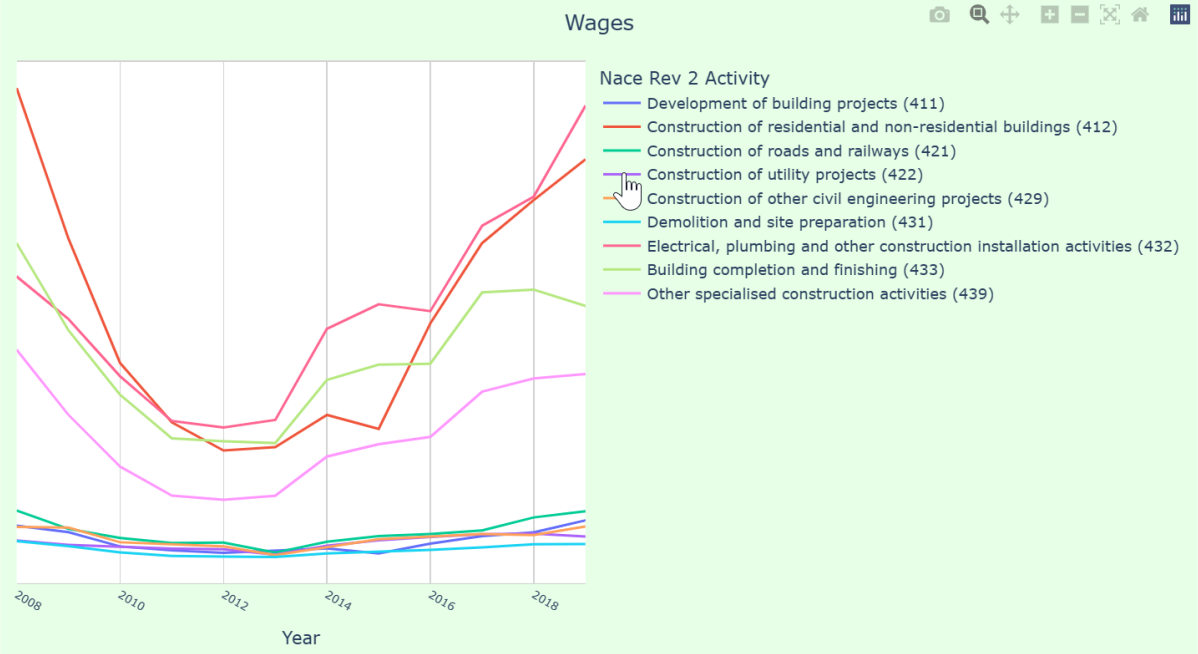


Figure 16

Plot in Figure 16 shows the trend of wages in the construction sector, line plot has been identified as the most appropriate as it gives an easily visualization of the change over the years, line plot has been identified as the most appropriate as it gives an easily visualization of the change. By default, all Nace categories are displayed, if the reader would like to focus on specific values only, it possible removing categories from the plot by clicking on its line colour on the right.

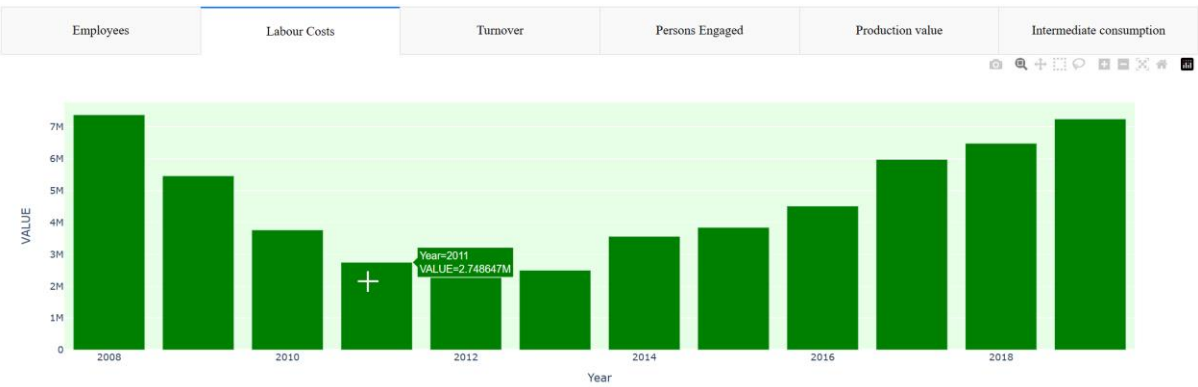
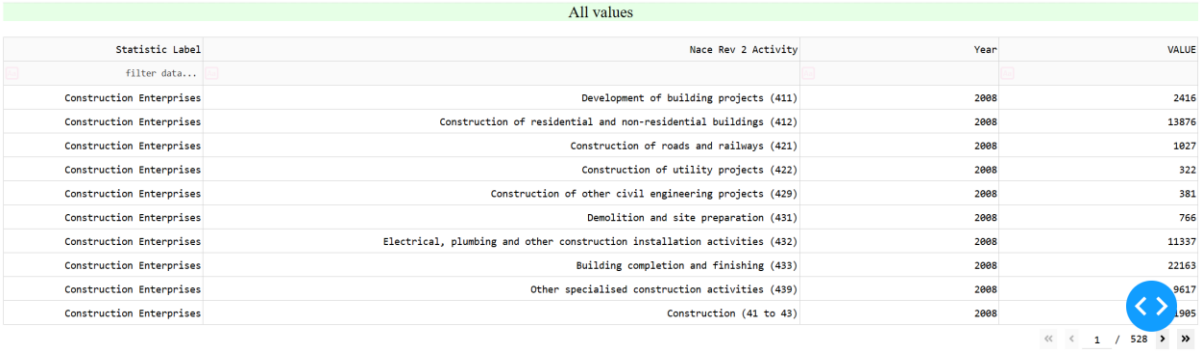


Figure 17

Bar plot in Figure 17 provides a quick and intuitive visualisation of the overall main data categories analyzed in this dashboard. The category can be selected in the bar on top and the

content will reflect the user's selection. Same as the previous graph, hovering capabilities are provided within this graph.



All values

Statistic Label	Nace Rev 2 Activity	Year	VALUE
filter data...			
Construction Enterprises	Development of building projects (411)	2008	2416
Construction Enterprises	Construction of residential and non-residential buildings (412)	2008	13876
Construction Enterprises	Construction of roads and railways (421)	2008	1027
Construction Enterprises	Construction of utility projects (422)	2008	322
Construction Enterprises	Construction of other civil engineering projects (429)	2008	381
Construction Enterprises	Demolition and site preparation (431)	2008	766
Construction Enterprises	Electrical, plumbing and other construction installation activities (432)	2008	11337
Construction Enterprises	Building completion and finishing (433)	2008	22163
Construction Enterprises	Other specialised construction activities (439)	2008	9617
Construction Enterprises	Construction (41 to 43)	2008	905

« ‹ 1 / 528 › »

Figure 18

Lastly, raw data is provided to the user for any additional query of interest. Filtering capabilities are in place to quickly identify values by typing the value in the first row, case sensitivity can be also enabled for strings, as shown in Figure 19 below.

Figure 19