**CCT College Dublin**

**Assessment Cover Page**

*To be provided separately as a word doc for students to include with every submission*

|  |  |
| --- | --- |
| **Module Title:** | Data Preparation and Visualization |
| **Assessment Title:** | Public Sector employment and Earnings |
| **Lecturer Name:** | David MCQuaid |
| **Student Full Name:** | Farhad Khankishiyev |
| **Student Number:** | 2023068 |
| **Assessment Due Date:** | 06.08.2023 |
| **Date of Submission:** | 06.08.2023 |

**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. |

2023068 – MSc in Data Analytics

Author: Farhad Khankishiyev

e-mail: [2023068@student.cct.ie](mailto:20230%36%38@s%74u%64e%6e%74%2e%63%63%74%2e%69%65)

Student ID: 2023068

<https://github.com/FarhadKhankishiyev068/RepeatAssigment>

License

• The OECD makes data (the “Data”) available for use and consultation by the public. Data may be subject to restrictions beyond the scope of these Terms and Conditions, either because specific terms apply to those Data or because third parties may have ownership interests. It is the User’s responsibility to verify, either directly in the metadata or, if available, by clicking on the icon and then referring to the "source" tab, whether the Data is fully or partially owned by third parties and/or whether additional restrictions may apply, and to contact the owner of the Data before incorporating it in your work in order to secure the necessary permissions. The OECD in no way represents or warrants that it owns or controls all rights in all Data, and the OECD will not be liable to any User for any claims brought against the User by third parties in connection with the use of any Data.

• Permitted use

• Except where additional restrictions apply as stated above, You can extract from, download, copy, adapt, print, distribute, share and embed Data for any purpose, even for commercial use. You must give appropriate credit to the OECD by using the citation associated with the relevant Data, or, if no specific citation is available, You must cite the source information using the following format: OECD (year), (dataset name),(data source) DOI or URL (accessed on (date)). When sharing or licensing work created using the Data, You agree to include the same acknowledgment requirement in any sub-licenses that You grant, along with the requirement that any further sub-licensees do the same.

Statistics

Summary

In this assessment I examined the two different data about amount of employment and average annual wage of three countries with the same currency. In my data Germany, Italy and Ireland are researched. In this data we can reach information from 2000 to 2021 year.

To conduct this analysis, I meticulously gathered reliable data from trusted sources. By including data from various countries, I was able to provide a comprehensive evaluation. Analysing growth patterns over time helps identify emerging trends of Ireland's employment and how it compares to other nations.

About Dataset:

In my Dataset I have Columns with Country Code, Year (Time), Value of annual wage and Amount of employment.

Descriptive statistics:

I analysed the dataset for these countries. The descriptive statistics reveal key insights:

A table of numbers and a number

Description automatically generated with medium confidence

Choropleth:

A map of europe with different colored countries/regions

Description automatically generated

Value Index on a map, with each country's shading reflecting its respective index value. The animation feature enables viewers to observe changes in the index over time, identifying trends and patterns. The natural earth projection enhances geographic accuracy, while the concise title, "Value Index," ensures clarity.

Distribution graphs:

A diagram of a graph

Description automatically generated with medium confidence

A screen shot of a graph

Description automatically generated

For Wage Value distribution graph, we can see all countries maximum, medium and minimum wage distribution. For Germany Annual wage distribution is maximum 43.6 K, medium and minimum is 33.7k and 28 k euro. For Italy Annual wage distribution is maximum 29.9 K, medium and minimum is 27.5 k and 21.3 k euro. For Ireland Annual wage distribution is maximum 50.7 K, medium and minimum is 43.7 k and 27.3 k euro.

The wage distribution graph reveals distinct variations in maximum, median, and minimum annual wages for Germany, Italy, and Ireland. These differences highlight disparities in income levels and labour market conditions among the three countries.

Germany median is closer to minimum wage. On the other two countries median is closer to maximum wage.

This linked with labour force in these countries. In Germany labour force median is 38 million. In Italy this number is 22.5 million. In the other hand in Ireland this number is 2 million. This number allows us to say why median of wage in Germany is closer to minimum wage.

Line Graphs:

A graph with lines and numbers

Description automatically generated

A graph of a graph showing the growth of the company's stock market

Description automatically generated with medium confidence

In this Line graphs we can see how annual wage growing by the year. Between the years 2010 and 2014 we can see in Ireland there was a decrease in the wage. This can be linked to house crisis in Ireland.

In Germany we can see increase of wage by the year is constantly except the year of 2020. This can be related with pandemic.

In Italy we can see slow increase than the other two countries.

T-Test

Tests for Germany and Ireland

A screenshot of a computer

Description automatically generated

T-Test Results:

Labour Force: A highly significant difference was found (T=76.97, p<0.001) between the groups, indicating substantial variation in Labor Force.

Annual Wage: A significant difference was observed (T=-3.50, p=0.00115) between the groups, implying notable diversity in Annual Wage.

These t-test results demonstrate significant disparities in Labor Force and Annual Wage between the compared groups.

Test for Germany and Italy

A screenshot of a computer

Description automatically generated

T-Test Results for Germany and Italy:

Labour Force:

A highly significant difference was found (T=33.53, p<0.001) between Germany and Italy, indicating substantial variation in Labor Force between the two countries.

Annual Wage: A significant difference was observed (T=6.79, p<0.001) between Germany and Italy, implying notable diversity in Annual Wage between the two countries.

These t-test results demonstrate significant disparities in Labor Force and Annual Wage between Germany and Italy.

Test for Italy and Ireland

A blue screen with white text

Description automatically generated

T-Test Results for Labour Force:

Italy and Ireland: A highly significant difference was found (T=210.91, p<0.001) in Labor Force between Italy and Ireland, indicating substantial variation in the labour force sizes of the two countries.

T-Test Results for Annual Wage:

Italy and Ireland: A significant difference was observed (T=-9.68, p<0.001) in Annual Wage between Italy and Ireland, implying notable diversity in annual wages between the two countries.

These t-test results demonstrate significant disparities in Labor Force and Annual Wage between Italy and Ireland, indicating differences in their workforce sizes and wage distributions.

ANOVA Test

A screenshot of a computer

Description automatically generated

ANOVA (Analysis of Variance) Test Results for Labour Force:

A highly significant difference was found (F=4204.84, p<0.001) in Labor Force among the groups being compared, indicating substantial variation in the labour force sizes.

ANOVA Test Results for Annual Wage:

A significant difference was observed (F=43.88, p<0.001) in Annual Wage among the groups being compared, implying notable diversity in annual wages.

These ANOVA test results suggest that there are significant differences in Labor Force and Annual Wage among the groups being compared, indicating variations in workforce sizes and wage distributions.

Chi-Squared Test

A screenshot of a computer

Description automatically generated

Chi-Squared Test Results:

Chi-Squared Statistic: 128.0

P-Value: 0.385

The Chi-Squared test was performed to determine if there is a significant association between the variables being compared. The obtained p-value (0.385) indicates that there is no significant association between the variables.

The degrees of freedom for the Chi-Squared test are 124, and the expected frequencies for the categories are shown in the matrix.

Paired T-test

A screenshot of a computer

Description automatically generated

Paired T-Test Results:

Germany (DEU): A significant difference was observed (T=6.66, p<0.001) before and after the treatment for Germany, indicating notable variation in the data.

Ireland (IRL): A highly significant difference was found (T=-29.71, p<0.001) before and after the treatment for Ireland, indicating substantial variation in the data.

Italy (ITA): A significant difference was observed (T=-6.97, p<0.001) before and after the treatment for Italy, implying notable diversity in the data.

These paired t-test results suggest that there are significant differences before and after the treatment for each country (Germany, Ireland, and Italy), indicating variations in the measured values for each country.

Wilcoxon Signed-Rank Test

A screenshot of a computer program

Description automatically generated

Wilcoxon Signed-Rank Test Results:

Germany (DEU): A significant difference was observed (Wilcoxon Statistic=8.0, p<0.001) before and after the treatment for Germany, indicating notable variation in the data.

Ireland (IRL): A highly significant difference was found (Wilcoxon Statistic=0.0, p<0.001) before and after the treatment for Ireland, indicating substantial variation in the data.

Italy (ITA): A significant difference was observed (Wilcoxon Statistic=5.0, p<0.001) before and after the treatment for Italy, implying notable diversity in the data.

These Wilcoxon Signed-Rank test results suggest that there are significant differences before and after the treatment for each country (Germany, Ireland, and Italy), indicating variations in the measured values for each country.

Correlation map

A graph of numbers and value

Description automatically generated with medium confidence

Correlation Map Results:

Year to Value: The correlation coefficient between the 'Year' and 'Value' variables is 0.64, indicating a moderate positive correlation. This suggests that there is some degree of association between the 'Year' and 'Value' columns, and as the 'Year' increases, the 'Value' tends to increase as well.

Year to Amount: The correlation coefficient between the 'Year' and 'Amount' variables is 0.051, which indicates a very weak positive correlation. This implies that there is little to no significant association between the 'Year' and 'Amount' columns.

Value to Amount: The correlation coefficient between the 'Value' and 'Amount' variables is -0.34, showing a moderate negative correlation. This suggests that as the 'Value' increases, the 'Amount' tends to decrease, and vice versa.

These correlation values provide insights into the relationships between the columns in the dataset, indicating the strength and direction of association between different variables.

Conclusion

This study was conducted with aim of analysing the Countries labour force and annual wages on the employment sector. Two datasets were compiled and processed in CSV format. Through data manipulation and cleaning, the data was brought into a form suitable for analysis.

In Visualization interactive map were used to compare countries annual wages. Interactive Line graphs and Distribution graphs were used to easily read and understand the differences or similarities between countries.

A Statistical Analysis was performed on the contribution of countries’ labour forces and annual wage. Our commentary focused on the contributions of the employment sector and annual wages of Ireland, Germany, Italy. These countries were selected because they represent different regions and economies within Europe, offering a diverse set of data for analysis.

Challenges I faced in the process.

It was difficult to find data about public sector employments for other countries. Most of the datasets are in different languages. Also, except Ireland most of the countries has not holding datasets about public sector employment and earnings or these datasets are private.