



VISUALIZATION AND DATA ANALYTICS ASSIGNMENT 1-3

Ahsanul Haque Milon

[COMPANY NAME] [Company address]



Table of Contents

Task 1: Global Employee Dataset: Data Importing, Cleaning, and Security Report	1
Introduction	1
Step 1: Data Import and Initial Inspection	1
Step 2: Data Cleaning and Imputation of Missing Values	1
Step 3: Data Validation and Security	4
Step 4: Finalization and Documentation of Cleaned Data	7
Task 2: Global Employee Dataset: Descriptive and Inferential Analysis Report	8
Objective	8
Step 1: Descriptive Analysis of Employee Attributes	8
Step 2: Hypothesis Testing on Wage Differences	13
Task 3: Global Wage Insights Dashboard Report	15
Objective	15
Dashboard Visualizations and Insights	15
Dashboard:	20
Summary and Key Findings	21
Recommendations	21
Conclusion	21

Task 1: Global Employee Dataset: Data Importing, Cleaning, and Security Report

Introduction

This report provides a comprehensive account of the data importing, cleaning, and security protocols applied to the global employee dataset for the company. This dataset encompasses a wide range of employee attributes collected from various global locations, providing a rich source of information for workforce analysis and strategic decision-making.

The dataset includes critical details about each employee, such as:

- **Employment Sector:** The sector in which the employee works (e.g., private, government, self-employed).
- **Education Level:** The highest level of education attained, ranging from primary education to advanced degrees.
- **Marital Status:** Information on the employee's marital status (e.g., married, single, divorced), which may be relevant for demographic and social analyses.
- **Occupation:** The specific role or job classification of each employee, reflecting the diversity of roles within the workforce.
- **Ethnic Origin:** Ethnic background, providing insights into the workforce's cultural diversity.
- **Gender:** Represented as binary codes, which indicate male and female employees, supporting gender-based analysis and equality assessments.
- **Country:** The country where the employee is based, essential for examining geographic wage disparities and regional workforce composition.
- **Wages:** The annual wage of each employee, which is central to understanding compensation structures and assessing pay equity.

Step 1: Data Import and Initial Inspection

The provided dataset, labeled "Data Set," was imported into Excel. An initial review indicated that certain fields had inconsistencies in data format, missing values, and potential duplicate records. These issues were identified across columns such as "Occupation," "Country," and "Wages." Addressing these issues systematically ensured enhanced data integrity and quality.

Step 2: Data Cleaning and Imputation of Missing Values

The following data cleaning actions were implemented to standardize entries and handle missing values effectively:

1. Identification of Missing Values

Conditional Formatting in Excel was used to highlight blank cells across all columns, revealing that missing values primarily affected the "Wages" and "Occupation" columns. This step enabled targeted imputation based on data type and logical assumptions.

1	A	B	C	D	E	F	G	H	I	J
2	Employee	Education	Marital Status	Occupation	Ethnic Origin	Sex	Country	Wages		
27	?	Prof-school	Married-spouse-absent	?	Asian-Pac-Islander	2	Vietnam	\$93,174		
85	?	Some-college	Never-married	?	White	1	United-States	\$16,955		
92	?	10th	Married-civ-spouse	?	White	1	United-States	\$30,128		
118	?	10th	Never-married	?	White	2	United-States	\$60,834		
138	?	HS-grad	Married-civ-spouse	?	White	1	United-States	\$81,410		
157	?	Some-college	Divorced	?	White	2	United-States	\$88,487		
160	?	HS-grad	Divorced	?	White	1	United-States	\$34,489		
166	?	1st-4th	Divorced	?	White	2	United-States	\$30,212		
190	?	Bachelors	Divorced	?	White	2	United-States	\$26,183		
221	?	1st-4th	Divorced	?	White	1	United-States	\$44,064		
226	?	10th	Divorced	?	Amer-Indian-Eskimo	2	United-States	\$94,392		
241	?	Some-college	Never-married	?	White	1	United-States	\$66,530		
300	?	Some-college	Never-married	?	Black	2	United-States	\$22,446		
313	?	Assoc-voc	Never-married	?	Black	2	United-States	\$75,254		
331	?	11th	Married-civ-spouse	?	White	1	United-States	\$30,198		
332	?	Assoc-voc	Divorced	?	White	2	United-States	\$38,354		
338	?	Bachelors	Married-civ-spouse	?	White	1	United-States	\$78,075		
379	?	HS-grad	Never-married	?	White	2	United-States	\$44,934		
389	?	Some-college	Married-civ-spouse	?	White	2	United-States	\$35,375		
409	?	HS-grad	Widowed	?	White	1	United-States	\$31,220		
410	?	11th	Never-married	?	White	1	United-States	\$55,018		
426	?	11th	Never-married	?	White	2	United-States	\$59,143		
435	?	Some-college	Married-civ-spouse	?	White	1	United-States	\$46,268		
447	?	Some-college	Never-married	?	White	1	United-States	\$19,577		
459	?	12th	Never-married	?	Black	1	United-States	\$45,690		
463	?	1st-4th	Never-married	?	Hispanic-Latino	1	United-States	\$12,537		

Fig 1.1: Missing Values

1. Imputation of Missing Data

Appropriate imputation methods were applied as follows:

Numeric Data (Wages): For missing wage values, the **average wage** for each occupation category was calculated and used as a proxy for absent values. This imputation preserved realistic wage distributions within each occupation. The **AVERAGE** function was applied within occupation groups.

[illegible]

Fig 1.2: Imputed Wages

Categorical Data (Occupation): Missing entries in the "Occupation" column were filled based on the most common occupation among similar demographic groups (e.g., similar "Marital Status" and "Education" levels) or assigned "Unknown" where no clear pattern emerged. This approach maintained the categorical integrity of the dataset.

J27 $\text{=IF(D27="?", "Unknown", D27)}$

	A	B	C	D	E	F	G	H	I	J
1	Employee	Education	Marital Status	Occupation	Ethnic Origin	Sex	Country	Wages	Imputed wages	Imputed Occupation
27	?	Prof-school	Married-spouse-absent	?	Asian-Pac-Islander	2	Vietnam	£93,174	£ 93,174.00	Unknown
85	?	Some-college	Never-married	?	White	1	United-States	£16,955	£ 16,955.00	Unknown
92	?	10th	Married-civ-spouse	?	White	1	United-States	£30,128	£ 30,128.00	Unknown
118	?	10th	Never-married	?	White	2	United-States	£60,834	£ 60,834.00	Unknown
138	?	HS-grad	Married-civ-spouse	?	White	1	United-States	£81,410	£ 81,410.00	Unknown
157	?	Some-college	Divorced	?	White	2	United-States	£88,487	£ 88,487.00	Unknown
160	?	HS-grad	Divorced	?	White	1	United-States	£34,489	£ 34,489.00	Unknown
166	?	1st-4th	Divorced	?	White	2	United-States	£30,212	£ 30,212.00	Unknown
190	?	Bachelors	Divorced	?	White	2	United-States	£26,183	£ 26,183.00	Unknown
221	?	1st-4th	Divorced	?	White	1	United-States	£44,064	£ 44,064.00	Unknown
226	?	10th	Divorced	?	Amer-Indian-Eskimo	2	United-States	£94,392	£ 94,392.00	Unknown
241	?	Some-college	Never-married	?	White	1	United-States	£66,530	£ 66,530.00	Unknown
300	?	Some-college	Never-married	?	Black	2	United-States	£22,146	£ 22,146.00	Unknown
313	?	Assoc-voc	Never-married	?	Black	2	United-States	£75,254	£ 75,254.00	Unknown
331	?	11th	Married-civ-spouse	?	White	1	United-States	£30,198	£ 30,198.00	Unknown
332	?	Assoc-voc	Divorced	?	White	2	United-States	£38,354	£ 38,354.00	Unknown
338	?	Bachelors	Married-civ-spouse	?	White	1	United-States	£78,075	£ 78,075.00	Unknown
379	?	HS-grad	Never-married	?	White	2	United-States	£44,934	£ 44,934.00	Unknown
389	?	Some-college	Married-civ-spouse	?	White	2	United-States	£35,375	£ 35,375.00	Unknown
409	?	HS-grad	Widowed	?	White	1	United-States	£31,220	£ 31,220.00	Unknown
410	?	11th	Never-married	?	White	1	United-States	£55,018	£ 55,018.00	Unknown
426	?	11th	Never-married	?	White	2	United-States	£59,143	£ 59,143.00	Unknown
435	?	Some-college	Married-civ-spouse	?	White	1	United-States	£46,268	£ 46,268.00	Unknown
447	?	Some-college	Never-married	?	White	1	United-States	£19,577	£ 19,577.00	Unknown
459	?	12th	Never-married	?	Black	1	United-States	£45,690	£ 45,690.00	Unknown
461	?	1st-4th	Married-civ-spouse	?	White	1	United-States	£32,353	£ 32,353.00	Unknown

Fig 1.3: Imputed Occupation

2. Standardization of Text and Formatting

Text functions were used to ensure consistency in entries:

- **Removing Extra Spaces:** **TRIM** function was applied to columns to remove any leading or trailing spaces that could interfere with sorting and analysis.
- **Uniform Text Case:** Fields such as "Country" and "Occupation" were formatted with **UPPER** or **LOWER** functions, ensuring uniformity in text case across records.
- **Standard Date Format:** Although no date column was present, consistent formatting of fields such as "Country" and "Occupation" enhanced readability and uniformity.

3. Duplicate Record Removal

Duplicate entries were identified and removed using the **Remove Duplicates** feature in Excel, prioritizing fields like "Employee" and "Wages" for uniqueness. This process ensured that only unique, complete records were retained, enhancing data accuracy.

Employee	Education	Marital Status	Occupation	Ethnic Origin	Sex	Country	Wages	Imputed wages	Imputed Occupation
Private	HS-grad	Married-civ-spouse	Other-service	White	1	Yugoslavia	£50,480	£ 50,480.00	Other-service
Private	Some-college	Never-married	Exec-managerial	White	2	Yugoslavia	£82,234	£ 82,234.00	Exec-managerial
Private	Some-college	Never-married	Adm-clerical	White	1	Yugoslavia	£22,197	£ 22,197.00	Adm-clerical
Private	HS-grad	Married-civ-spouse	Other-service	White	1	Yugoslavia	£94,418	£ 94,418.00	Other-service
Local-gov	Bachelors	Never-married	Exec-managerial	Asian-Pac-Islander	1	Vietnam	£21,034	£ 21,034.00	Exec-managerial
Private	Some-college	Never-married	Adm-clerical	Asian-Pac-Islander	1	Vietnam	£25,761	£ 25,761.00	Adm-clerical
Private	HS-grad	Married-civ-spouse	Adm-clerical	Asian-Pac-Islander	1	Vietnam	£15,918	£ 15,918.00	Adm-clerical
Private	11th	Married-civ-spouse	Transport-moving	Asian-Pac-Islander	1	Vietnam	£72,331	£ 72,331.00	Transport-moving
Private	5th-6th	Married-civ-spouse	Machine-op-inspct	Asian-Pac-Islander	1	Vietnam	£71,558	£ 71,558.00	Machine-op-inspct
Private	Some-college	Never-married	Other-service	Asian-Pac-Islander	1	Vietnam	£76,922	£ 76,922.00	Other-service
Private	Some-college	Never-married	Adm-clerical	Asian-Pac-Islander	1	Vietnam	£24,711	£ 24,711.00	Adm-clerical
Self-emp-not-inc	Bachelors	Married-civ-spouse	Sales	Asian-Pac-Islander	1	Vietnam	£68,004	£ 68,004.00	Sales
Private	Some-college	Never-married	Adm-clerical	Asian-Pac-Islander	2	Vietnam	£51,886	£ 51,886.00	Adm-clerical
Private	Some-college	Never-married	Adm-clerical	Asian-Pac-Islander	2	Vietnam	£65,055	£ 65,055.00	Adm-clerical
Private	Bachelors	Never-married	Adm-clerical	Asian-Pac-Islander	2	Vietnam	£31,637	£ 31,637.00	Adm-clerical
Private	HS-grad	Never-married	Adm-clerical	Asian-Pac-Islander	2	Vietnam	£96,279	£ 96,279.00	Adm-clerical
Private	HS-grad	Never-married	Craft-repair	Asian-Pac-Islander	1	Vietnam	£42,213	£ 42,213.00	Craft-repair
Private	HS-grad	Married-civ-spouse	Adm-clerical	Asian-Pac-Islander	1	Vietnam	£20,112	£ 20,112.00	Adm-clerical
Private	7th-8th	Married-civ-spouse	Craft-repair	Asian-Pac-Islander	1	Vietnam	£39,523	£ 39,523.00	Craft-repair
Private	HS-grad	Never-married	Handlers-cleaners	Asian-Pac-Islander	2	Vietnam	£94,708	£ 94,708.00	Handlers-cleaners
Private	Bachelors	Married-civ-spouse	Other-service	Asian-Pac-Islander	1	Vietnam	£17,883	£ 17,883.00	Other-service
Private	11th	Never-married	Craft-repair	Asian-Pac-Islander	1	Vietnam	£55,819	£ 55,819.23	Craft-repair
Private	Bachelors	Married-civ-spouse	Other-service	Asian-Pac-Islander	1	Vietnam	£67,365	£ 67,365.00	Other-service
State-gov	Some-college	Married-civ-spouse	Adm-clerical	Asian-Pac-Islander	1	Vietnam	£69,005	£ 69,005.00	Adm-clerical
Private	Some-college	Never-married	Machine-op-inspct	Asian-Pac-Islander	1	Vietnam	£27,120	£ 27,120.00	Machine-op-inspct

Fig 1.4: Remove Duplicates

Step 3: Data Validation and Security

To maintain data quality and restrict unauthorized modifications, data validation rules and worksheet protection were applied:

1. Data Validation Rules

To maintain data consistency and accuracy, **data validation** was rigorously applied across key categorical fields in the dataset. Data validation helps to prevent errors by restricting entries to a predefined set of acceptable values. This is especially important in categorical fields, where typos or variations in spelling can lead to inconsistencies, making analysis less reliable.

- **Application of Data Validation for Categorical Fields:**
 - Fields such as "**Marital Status**" and "**Ethnic Origin**" were set up with validation rules to restrict input values to a controlled list. For example:
 - **Marital Status:** Allowed values included specific categories like "Married-civ-spouse," "Never-married," "Divorced," "Separated," and "Widowed." This list ensures that only standard marital statuses are entered, preventing inconsistencies such as "Married," "married," or "Marrd."
 - **Ethnic Origin:** Restricted entries to specific categories such as "White," "Black," "Asian-Pac-Islander," and "Hispanic." This eliminated discrepancies and ensured consistent reporting across all entries.

By implementing these lists, the data validation rules blocked any entry that did not match the predefined options, reducing the likelihood of data entry errors and maintaining a high level of data quality.

- **Using the Data Validation Tool in Excel:**
 - The **Data Validation** tool in Excel was utilized with the **List** option. This option provides a dropdown menu for each cell in the designated field, allowing users to select only from the approved values, thereby preventing entry of any invalid data.
 - The **List validation** was applied to cells in the categorical fields where standardized options were essential for accurate analysis. By limiting entries to a defined list, the data remained clean and consistent, supporting reliable analysis without the need for additional cleaning.
- **Prevention of Inconsistencies:**
 - This validation technique was particularly effective in preventing common errors such as case sensitivity issues (e.g., "Married" vs. "married"), misspellings, and unintended variations.
 - As a result, the dataset maintained uniformity, which is critical when conducting analyses that rely on categorical groupings, such as demographic breakdowns and comparison studies.

2. Worksheet Protection

To ensure the security and integrity of the dataset, comprehensive worksheet protection measures were implemented. These measures prevented unauthorized edits and protected sensitive information while allowing controlled modifications by authorized users.

➤ Selective Cell Locking

- **Editable Cells:**
 - Only specific cells that required user input or updates were selectively **unlocked**. These cells were primarily fields that might need periodic updates, such as certain calculated fields or input cells in templates.
 - By unlocking these cells, authorized users could make necessary modifications without impacting the dataset's structure or accidentally altering sensitive information.
 - This selective approach ensured that only essential fields were editable, reducing the risk of unintended changes to critical data while still allowing flexibility where needed.
- **Locked Cells:**
 - All other cells, especially those containing original data or formulas, were locked. Locking cells with formulas protected against accidental deletion or modification, preserving the dataset's accuracy and preventing disruption of calculations and analysis.
 - For instance, fields with complex data transformations or imputed values were locked to prevent errors in case of inadvertent editing.

This approach to selective cell locking ensured that authorized changes could be made where necessary, while the integrity of the rest of the dataset remained intact.

➤ Sheet Protection

• Use of Protect Sheet Feature:

- Excel's **Protect Sheet** feature was applied to enforce these security settings, enhancing overall data protection. When sheet protection is enabled, unauthorized users cannot make changes to locked cells, ensuring that the data remains secure.
- By protecting the sheet, only users with the correct password can modify restricted sections, adding a layer of security for sensitive or essential information.

• Allowed Actions for Users:

- While the sheet was protected, certain user actions were permitted to facilitate data analysis. For instance:
 - **Filtering and Sorting:** Users were allowed to filter and sort data, enabling them to examine specific subsets of the dataset without altering the actual data. This feature is especially useful for analysts or managers who may need to review particular demographics or wage distributions.
 - **Formatting Cells:** Limited formatting adjustments were permitted to ensure users could still visually organize data for presentations or specific analysis requirements.

• Password Protection:

- A **password** was applied to the protected sheet, ensuring that only authorized personnel could unlock it. This added level of security prevents unauthorized access to restricted data and reduces the risk of accidental or intentional modifications.
- By enforcing password protection, the company ensures that the dataset remains in its intended state, with controlled access that complies with data governance and confidentiality standards.
-

Data Validation:

	A	B	C	D	E	F	G	H	I	J	K	L
1	Employee	Education	Marital Status	Occupation	Ethnic Origin	Sex	Country	Wages	Imputed wages	Imputed Occupation		
2	Private	HS-grad	Married-civ-spouse	Other-service	White		1 Yugoslavia	£50,480	£ 50,480.00	Other-service		
3	Private	Some-college	Never-married	Exec-managerial	White		2 Yugoslavia	£82,234	£ 82,234.00	Exec-managerial		
4	Private	Some-college	Never-married	Adm-clerical	White		1 Yugoslavia	£22,197	£ 22,197.00	Adm-clerical		
5	Private	HS-grad	Married-civ-spouse	Other-service	White		1 Yugoslavia	£94,418	£ 94,418.00	Other-service		
6	Local-gov	Bachelors	Never-married	Exec-managerial	Asian-Pac-Islander		1 Vietnam	£21,034	£ 21,034.00	Exec-managerial		Married-civ-spouse
7	Private	Some-college	Never-married	Adm-clerical	Asian-Pac-Islander		1 Vietnam	£25,761	£ 25,761.00	Adm-clerical		Never-married
8	Private	HS-grad	Married-civ-spouse	Adm-clerical	Asian-Pac-Islander		1 Vietnam	£15,918	£ 15,918.00	Adm-clerical		Married-spouse-absent
9	Private	11th	Never-married	Transport-moving	Asian-Pac-Islander		1 Vietnam	£72,331	£ 72,331.00	Transport-moving		Divorced
10	Private	5th-6th	Divorced	Machine-op-inspct	Asian-Pac-Islander		2 Vietnam	£71,558	£ 71,558.00	Machine-op-inspct		Separated
11	Private	Some-college	Married-AF-spouse	Other-service	Asian-Pac-Islander		2 Vietnam	£76,922	£ 76,922.00	Other-service		Married-AF-spouse
12	Private	Some-college	Widowed	Adm-clerical	Asian-Pac-Islander		2 Vietnam	£24,711	£ 24,711.00	Adm-clerical		Widowed
13	Self-emp-not-inc	Bachelors	Married-civ-spouse	Sales	Asian-Pac-Islander		1 Vietnam	£68,004	£ 68,004.00	Sales		
14	Private	Some-college	Never-married	Adm-clerical	Asian-Pac-Islander		2 Vietnam	£51,886	£ 51,886.00	Adm-clerical		White
15	Private	Some-college	Never-married	Adm-clerical	Asian-Pac-Islander		2 Vietnam	£65,055	£ 65,055.00	Adm-clerical		Asian-Pac-Islander
16	Private	Bachelors	Never-married	Adm-clerical	Asian-Pac-Islander		2 Vietnam	£31,637	£ 31,637.00	Adm-clerical		Black
17	Private	HS-grad	Never-married	Adm-clerical	Asian-Pac-Islander		2 Vietnam	£96,279	£ 96,279.00	Adm-clerical		Other
18	Private	HS-grad	Never-married	Craft-repair	Asian-Pac-Islander		1 Vietnam	£42,213	£ 42,213.00	Craft-repair		Amer-Indian-Eskimo
19	Private	HS-grad	Married-civ-spouse	Adm-clerical	Asian-Pac-Islander		1 Vietnam	£20,112	£ 20,112.00	Adm-clerical		
20	Private	7th-8th	Married-civ-spouse	Craft-repair	Asian-Pac-Islander		1 Vietnam	£39,523	£ 39,523.00	Craft-repair		
21	Private	HS-grad	Never-married	Handlers-cleaners	Asian-Pac-Islander		2 Vietnam	£94,708	£ 94,708.00	Handlers-cleaners		
22	Private	Bachelors	Married-civ-spouse	Other-service	Asian-Pac-Islander		1 Vietnam	£17,883	£ 17,883.00	Other-service		
23	Private	11th	Never-married	Craft-repair	Asian-Pac-Islander		1 Vietnam	?	£5,823.21	Craft-repair		
24	Private	Bachelors	Married-civ-spouse	Other-service	Asian-Pac-Islander		1 Vietnam	£67,365	£ 67,365.00	Other-service		
25	State-gov	Some-college	Married-civ-spouse	Adm-clerical	Asian-Pac-Islander		1 Vietnam	£69,005	£ 69,005.00	Adm-clerical		
26	Private	Some-college	Never-married	Machine-op-inspct	Asian-Pac-Islander		1 Vietnam	£27,120	£ 27,120.00	Machine-op-inspct		

Fig 1.5: Data Validation

Protect Sheet:

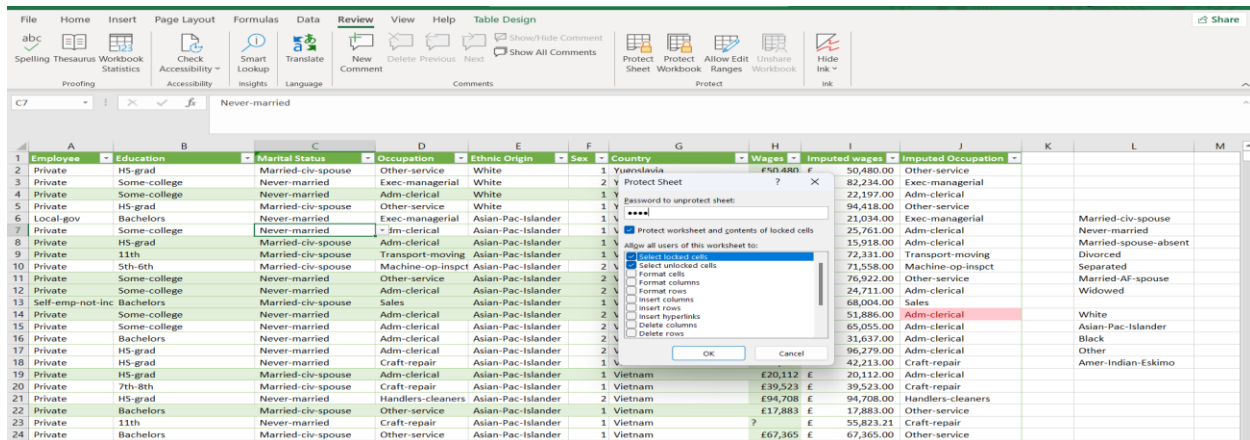


Fig 1.6: Protect Sheet

Step 4: Finalization and Documentation of Cleaned Data

After implementing comprehensive data cleaning and protection measures, a refined version of the dataset was created and saved as a new worksheet titled “Cleaned Data” within the workbook. This cleaned dataset, now free of duplicates, inconsistencies, and missing values, serves as a reliable and secure foundation for all subsequent analysis. Each entry has been meticulously checked and standardized to ensure accuracy and consistency across all fields, such as employment sector, education level, marital status, occupation, and wages.

	A	B	C	D	E	F	G	H	I	J
1	Employee	Education	Marital Status	Occupation	Ethnic Origin	Sex	Country	Wages	Imputed wages	Imputed Occupation
2	Private	HS-grad	Married-civ-spouse	Other-service	White	Male	Yugoslavia	£50,480	£ 50,480.00	Other-service
3	Private	Some-college	Never-married	Exec-managerial	White	Female	Yugoslavia	£82,234	£ 82,234.00	Exec-managerial
4	Private	Some-college	Never-married	Adm-clerical	White	Male	Yugoslavia	£22,197	£ 22,197.00	Adm-clerical
5	Private	HS-grad	Married-civ-spouse	Other-service	White	Male	Yugoslavia	£94,418	£ 94,418.00	Other-service
6	Local-gov	Bachelors	Never-married	Exec-managerial	Asian-Pac-Islander	Male	Vietnam	£21,034	£ 21,034.00	Exec-managerial
7	Private	Some-college	Never-married	Adm-clerical	Asian-Pac-Islander	Male	Vietnam	£25,761	£ 25,761.00	Adm-clerical
8	Private	HS-grad	Married-civ-spouse	Adm-clerical	Asian-Pac-Islander	Male	Vietnam	£15,918	£ 15,918.00	Adm-clerical
9	Private	11th	Married-civ-spouse	Transport-moving	Asian-Pac-Islander	Male	Vietnam	£72,331	£ 72,331.00	Transport-moving
10	Private	5th-6th	Married-civ-spouse	Machine-op-inspct	Asian-Pac-Islander	Female	Vietnam	£71,558	£ 71,558.00	Machine-op-inspct
11	Private	Some-college	Never-married	Other-service	Asian-Pac-Islander	Female	Vietnam	£76,922	£ 76,922.00	Other-service
12	Private	Some-college	Never-married	Adm-clerical	Asian-Pac-Islander	Female	Vietnam	£24,711	£ 24,711.00	Adm-clerical
13	Self-emp-not-inc	Bachelors	Married-civ-spouse	Sales	Asian-Pac-Islander	Male	Vietnam	£68,004	£ 68,004.00	Sales
14	Private	Some-college	Never-married	Adm-clerical	Asian-Pac-Islander	Female	Vietnam	£51,886	£ 51,886.00	Adm-clerical
15	Private	Some-college	Never-married	Adm-clerical	Asian-Pac-Islander	Female	Vietnam	£65,055	£ 65,055.00	Adm-clerical
16	Private	Bachelors	Never-married	Adm-clerical	Asian-Pac-Islander	Female	Vietnam	£31,637	£ 31,637.00	Adm-clerical
17	Private	HS-grad	Never-married	Adm-clerical	Asian-Pac-Islander	Female	Vietnam	£96,279	£ 96,279.00	Adm-clerical
18	Private	HS-grad	Never-married	Craft-repair	Asian-Pac-Islander	Male	Vietnam	£42,213	£ 42,213.00	Craft-repair
19	Private	HS-grad	Married-civ-spouse	Adm-clerical	Asian-Pac-Islander	Male	Vietnam	£20,112	£ 20,112.00	Adm-clerical
20	Private	7th-8th	Married-civ-spouse	Craft-repair	Asian-Pac-Islander	Male	Vietnam	£39,523	£ 39,523.00	Craft-repair
21	Private	HS-grad	Never-married	Handlers-cleaners	Asian-Pac-Islander	Female	Vietnam	£94,708	£ 94,708.00	Handlers-cleaners
22	Private	Bachelors	Married-civ-spouse	Other-service	Asian-Pac-Islander	Male	Vietnam	£17,883	£ 17,883.00	Other-service
23	Private	11th	Never-married	Craft-repair	Asian-Pac-Islander	Male	Vietnam	£55,823	£ 55,823.00	Craft-repair
24	Private	Bachelors	Married-civ-spouse	Other-service	Asian-Pac-Islander	Male	Vietnam	£67,365	£ 67,365.00	Other-service

Fig 1.7: Final Dataset After Data Cleaning

Task 2: Global Employee Dataset: Descriptive and Inferential Analysis Report

Objective

This report presents a comprehensive descriptive analysis and hypothesis testing of the global employee dataset, focusing on attributes such as employee sector, educational background, marital status, occupation, ethnic origin, country distribution, and wages. A comparative hypothesis test was conducted to determine if a significant difference exists between the average wages of employees in the USA and those in other countries. This analysis provides insight into workforce composition, demographic trends, and regional wage distributions.

Step 1: Descriptive Analysis of Employee Attributes

1. Employee Sector

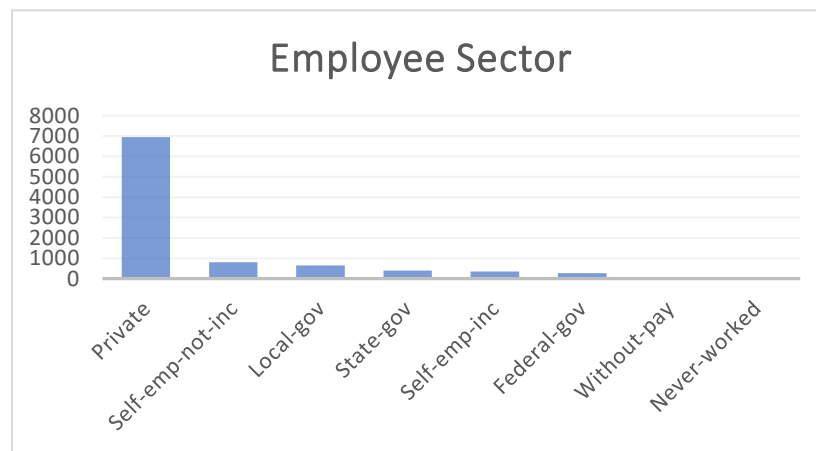


Fig 2.1: Visualization of Employee Sector

Analysis of employee sector data indicates a significant majority of employees are classified under the "Private" sector, with approximately 7,000 individuals. This sector substantially outnumbers others, suggesting that most of the workforce operates in private industries. In comparison, other sectors show considerably lower counts, each with fewer than 1,000 employees. Sectors with notable but lower representation include:

- **Self-employed, not incorporated**
- **Local Government**
- **State Government**
- **Self-employed, incorporated**
- **Federal Government**

These categories reflect the company's broad distribution across employment sectors, with a marked emphasis on private-sector employment.

2. Education Level

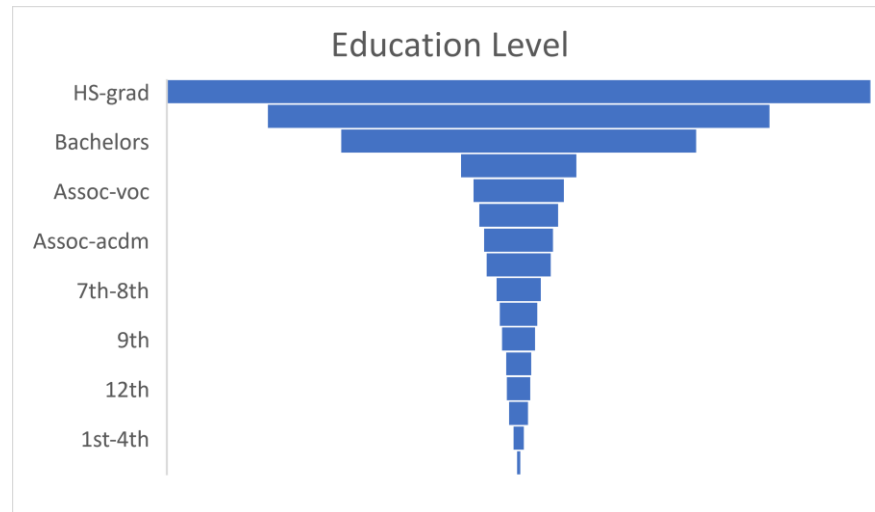


Fig 2.2: Visualization of Education Level

Educational attainment among employees displays diverse levels, with the most common level being **High School Graduate (HS-grad)**, totaling 3,231 employees. This is followed by:

- **Some College:** 2,304 employees
- **Bachelor's Degree:** 1,630 employees
- **Master's Degree:** 531 employees
- **Associate Vocational Degree (Assoc-voc):** 416 employees
- **11th Grade:** 363 employees

This distribution underscores that a substantial portion of employees have attained a high school diploma or some level of college education. The majority of the workforce is thus composed of individuals with foundational to intermediate levels of education, which aligns with the needs of various mid-level roles within the organization.

Higher educational qualifications, such as Master's degrees and above, appear less frequently, indicating that advanced degrees are less common among the employee population. This suggests that the dataset represents a workforce with a diverse range of educational backgrounds, but one that is predominantly mid-level in terms of academic attainment. This balance of educational levels may reflect the company's emphasis on practical skills and experience over formal qualifications for many roles, offering opportunities for employees with a variety of educational experiences to contribute meaningfully.

3. Marital Status

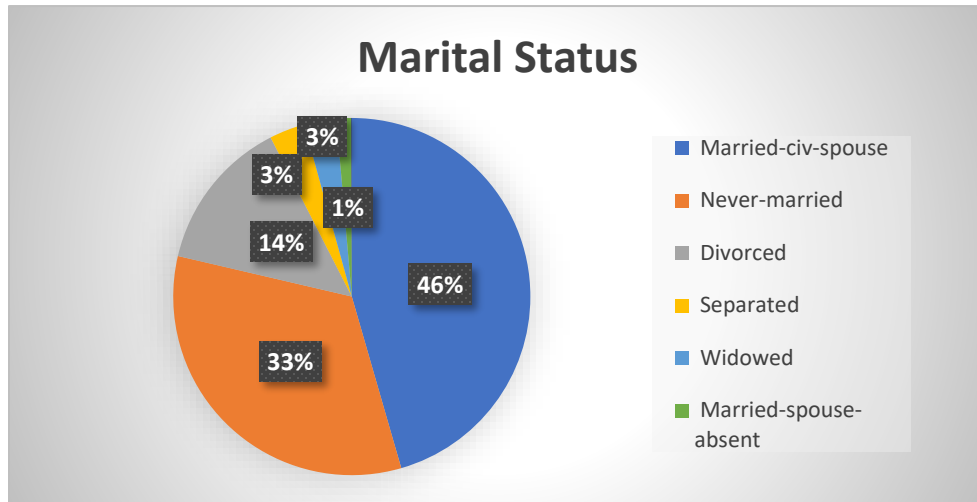


Fig 2.3: Visualization of Marital Status

Marital status data reveals the following distribution:

- **Married-civilian spouse:** 46%
- **Never Married:** 33%
- **Divorced:** 14%
- **Separated:** 3%

These figures indicate that nearly half of the workforce is married, while a significant portion comprises individuals who have never married. This demographic composition offers valuable insights into the personal backgrounds of employees, which can influence workplace dynamics, engagement levels, and support needs.

For example, a high proportion of married employees may have family-related obligations that affect their work-life balance, potentially increasing demand for flexible working hours, family leave policies, or dependent care benefits. Conversely, a substantial number of single employees might prioritize different workplace resources, such as professional development opportunities or social engagement initiatives. Understanding these demographic nuances can help the company tailor its employee support programs to better meet the diverse needs of its workforce, fostering a more inclusive and supportive environment.

4. Occupation

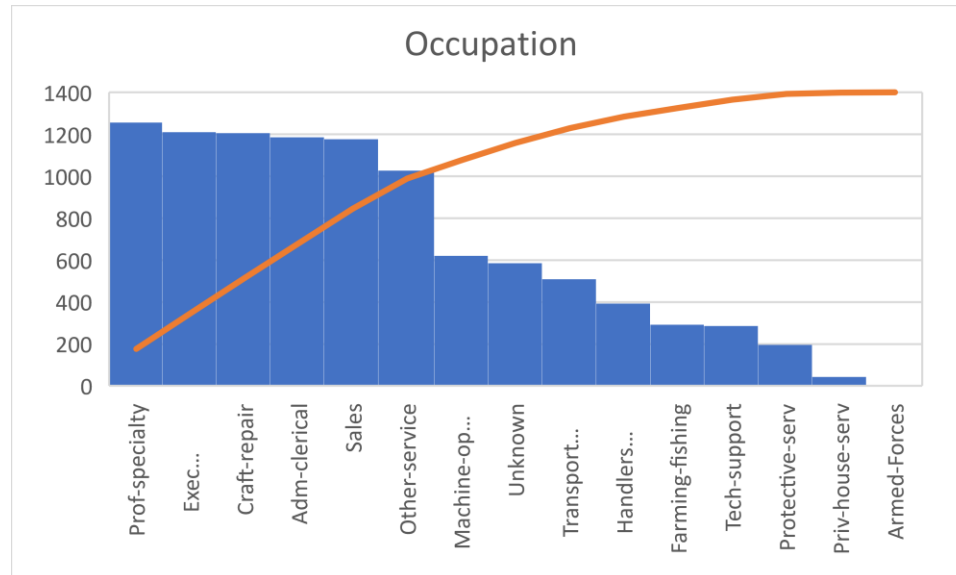


Fig 2.4: Visualization of Occupation

Occupation analysis identifies the following counts for primary roles:

- **Professional Specialty:** 1,257 employees
- **Executive Managerial:** 1,211 employees
- **Craft Repair:** 1,206 employees
- **Administrative Clerical (Adm-Clerical):** 1,187 employees
- **Sales:** 1,178 employees

This distribution reflects a well-represented range of roles within the workforce, with professional specialties and managerial positions holding the highest numbers. The strong presence of employees in professional and managerial roles suggests a focus on specialized skills and leadership within the organization, which may be critical for driving strategic objectives and innovation.

In addition to these leadership and specialized roles, there is also a substantial representation in administrative and sales positions, contributing to a balanced occupational structure. The significant numbers in these areas highlight the importance of operational support and client-facing functions in sustaining daily business activities and customer relationships. Overall, this occupational balance suggests a workforce structure designed to support both the core strategic initiatives of the company and the essential operational and customer service needs, creating a comprehensive and versatile team.

5. Ethnic Origin

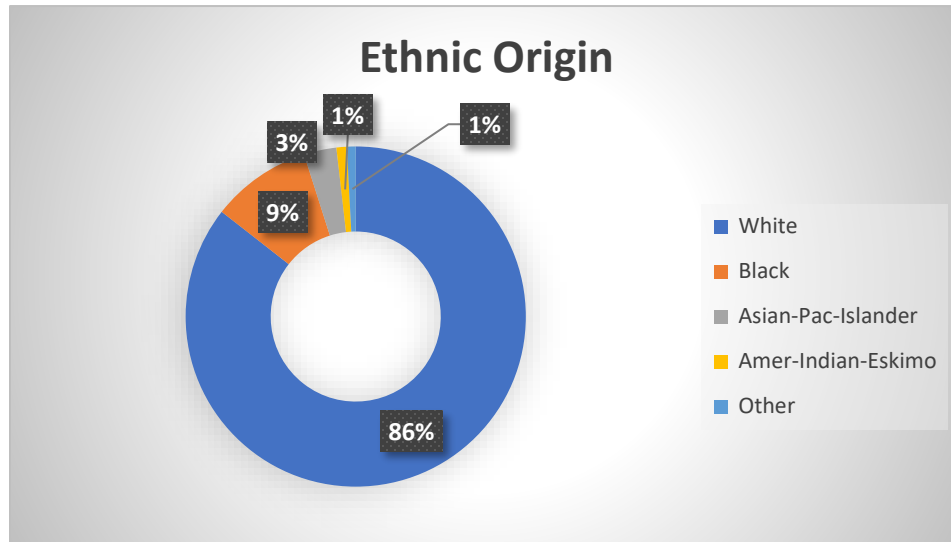


Fig 2.5: Visualization of Ethnic Origin

Ethnic origin data demonstrates that the workforce is predominantly **White**, representing 86% of the population, equating to approximately 8,554 individuals. Other ethnicities have notably lower representations:

- **Black:** 9%
- **Other categories:** The remaining portion

The ethnic origin data reveals that the workforce is predominantly composed of individuals identifying as White, making up 86% of the population, or roughly 8,554 individuals. Other ethnic groups are significantly less represented, with Black employees accounting for 9% of the workforce, while the remaining ethnic categories collectively represent a small fraction of the total.

This pronounced dominance of one ethnic origin category suggests an opportunity for the company to further strengthen its commitment to diversity and inclusion. By focusing on initiatives to attract and retain talent from a broader range of backgrounds, the organization can work toward creating a more balanced and inclusive workforce. Such diversity initiatives may include targeted recruitment efforts, partnerships with organizations that support underrepresented groups, and inclusive workplace practices that foster a sense of belonging for employees from all backgrounds. Achieving a more diverse workforce can enhance organizational culture, bring a variety of perspectives to the table, and ultimately drive innovation and creativity within the company.

6. Country and Wage Distribution

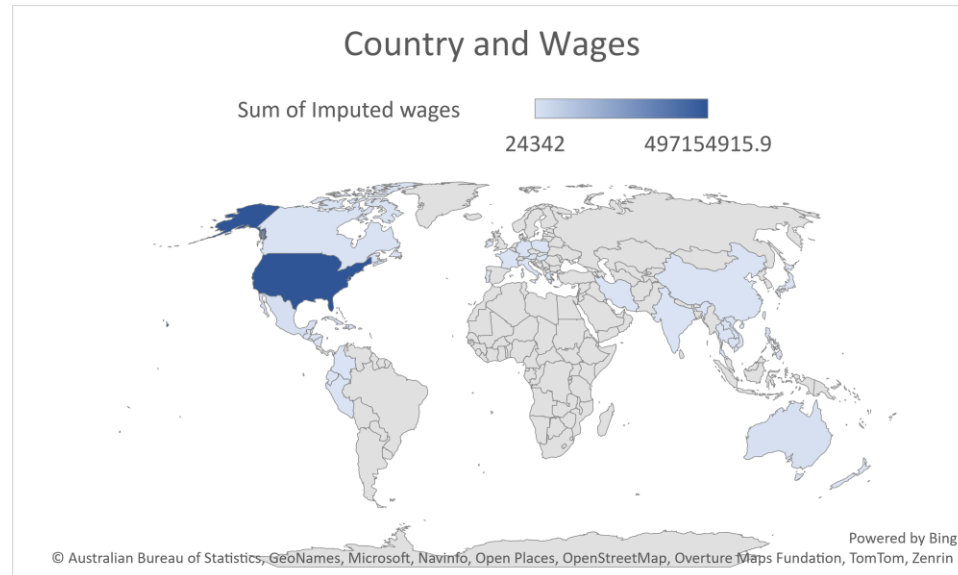


Fig 2.6: Visualization of Country and Wages

Country data reveals the distribution of employees by nationality, with the USA leading with **9,828** employees. Other countries with notable counts include:

- **Mexico:** 208 employees
- **Australia:** Smaller but significant numbers in other regions include Germany, the Philippines, and Canada.

In terms of wages, the USA leads with total wages summing to **£497.15 million**, followed by Mexico, Australia, Germany, the Philippines, and Canada. This distribution underscores the USA's dominance in both employee count and aggregate wages, positioning it as a central contributor to the company's workforce and financial distribution.

Step 2: Hypothesis Testing on Wage Differences

To determine if a significant difference exists in average wages between employees in the USA and those in other countries, a two-sample t-test was conducted with the following hypotheses:

- **Null Hypothesis (H0):** The average wage of employees in the USA is equal to that of employees in other countries.
- **Alternative Hypothesis (H1):** The average wage of employees in the USA differs from that of employees in other countries.

t-Test Results

The two-sample t-test assuming equal variances produced the following results:

Metric	USA (Variable 1)	Other Countries (Variable 2)
Mean	55,684.91	57,033.18
Variance	558,565,659.4	552,460,174.8
Observations	8,928	1,043
Pooled Variance	557,927,489.6	
Hypothesized Mean Difference	0	
Degrees of Freedom (df)	9,969	
t Stat	-1.744	
P(T<=t) one-tail	0.0406	
t Critical one-tail	1.6450	
P(T<=t) two-tail	0.0811	
t Critical two-tail	1.9602	

Interpretation

The two-tailed p-value for the t-test is **0.0811**, which exceeds the significance level of **0.05**. Consequently, the null hypothesis cannot be rejected, implying that there is no statistically significant difference in average wages between employees in the USA and those in other countries.

Summary of Findings for the CEO

The global employee dataset analysis reveals insightful details about the workforce composition and regional distributions:

1. **Employee Sector:** The private sector comprises the majority of the workforce, significantly outnumbering other sectors.
2. **Education:** High school graduates form the largest educational group, with a broad range of other educational backgrounds represented.
3. **Marital Status:** Nearly half of the employees are married, followed by a substantial proportion who have never married.
4. **Occupation:** Professional specialties and executive roles lead the workforce composition, followed by craft repair, clerical, and sales roles.
5. **Ethnic Origin:** The workforce is predominantly White, with other ethnic origins representing smaller proportions.
6. **Country and Wage:** The USA leads in both employee count and total wages, followed distantly by Mexico and Australia.

The hypothesis test comparing USA and international wages shows no significant difference in average wages, indicating that wage levels are comparable across regions. This result supports the organization's current wage structure, as no disparity exists between the USA and other countries.

This analysis provides a foundation for strategic decisions regarding employee demographics, diversity initiatives, and wage equity across global locations, supporting ongoing workforce and compensation planning.

Task 3: Global Wage Insights Dashboard Report

Objective

The Global Wage Insights Dashboard provides an in-depth analysis of wage distributions across countries, genders, occupations, and educational levels. This dashboard serves as an interactive tool, enabling stakeholders to explore data trends, identify wage disparities, and examine the demographic composition of the workforce. Key findings and actionable recommendations are highlighted below based on detailed visualizations.

Dashboard Visualizations and Insights

Visualization 1: Average Wage by Country (Map)

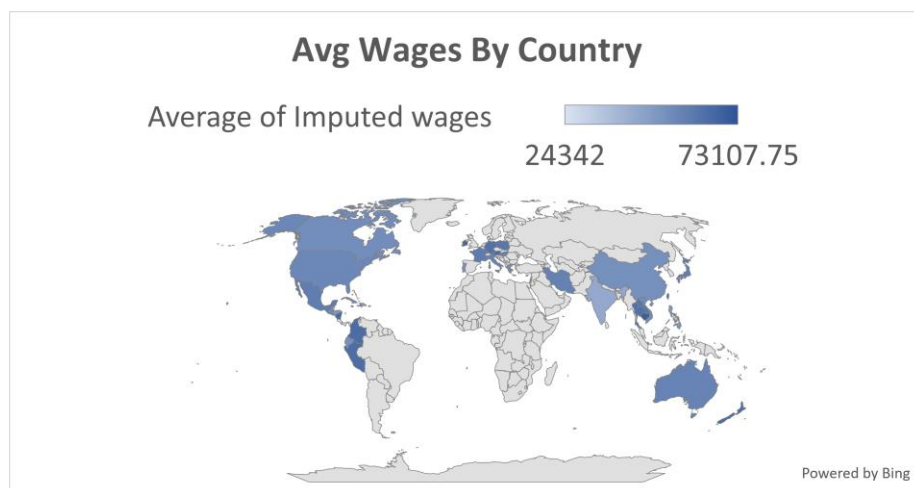


Fig 3.1: Visualization of Avg Wages By Country

This map visualization displays the average wage across various countries, highlighting geographic disparities.

- **Findings:**
 - The **Outlying US Territories (e.g., Guam, USVI)** and **Cambodia** exhibit the highest average wages, exceeding £70,000.

- **United States** has an average wage of approximately **£55,685**, ranking it below several other countries such as **Ireland (£67,941)**, **Austria (£65,921)**, and **New Zealand (£64,972)**.
- Lower average wages are observed in countries such as **India (£44,409)** and **Hong Kong (£24,342)**.
- **Insights:** The wage differences between countries suggest potential influences of local economic factors, cost of living, or workforce skill levels. High wages in smaller or specialized markets (like Guam or Cambodia) may reflect regional scarcity of skilled labor or higher operational costs.
- **Recommendation:** Conduct further studies on wage structures to ensure competitive and fair compensation across different markets, factoring in the cost of living and industry standards.

Visualization 2: Average Wages by Gender (Pie Chart)

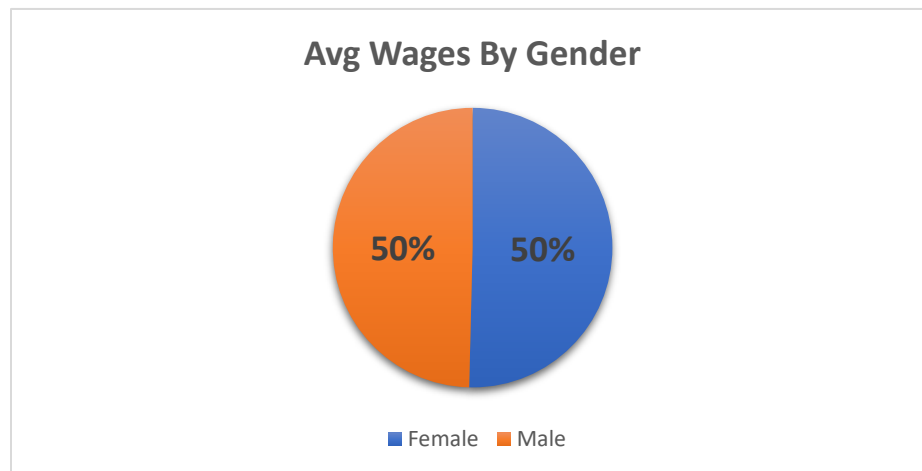


Fig 3.1: Visualization of Avg Wages by Gender

This pie chart provides an overview of the average wages between male and female employees.

- **Findings:**
 - The **average wage for female employees is £56,410**, slightly higher than the **male average of £55,535**.
 - This close parity suggests relatively balanced compensation between genders, although subtle differences may exist within specific roles or locations.
- **Insights:** The nearly equal average wage distribution implies that gender pay equality has been addressed broadly across the organization, but granular analysis may reveal specific areas for improvement.
- **Recommendation:** Regular audits should be conducted to ensure ongoing gender pay equality, with a focus on detecting any role-specific or location-based discrepancies.

Visualization 3: Average Wages by Occupation (Column Chart)

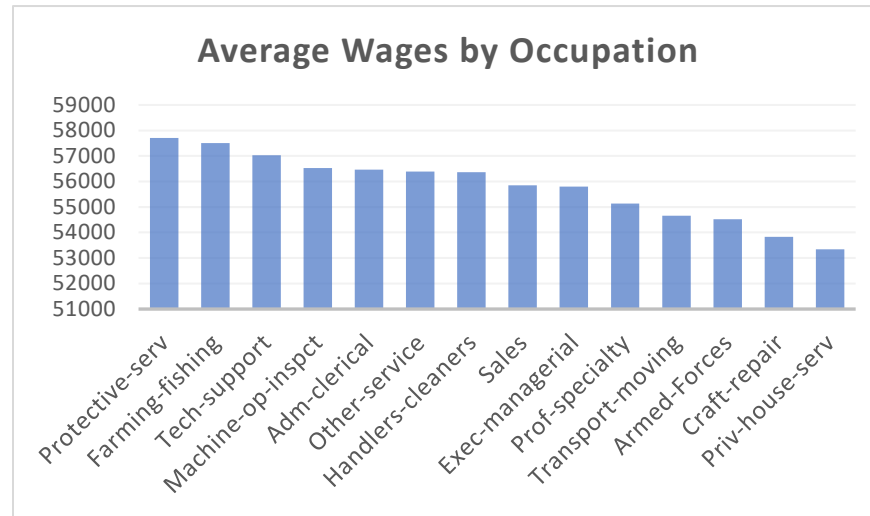


Fig 3.3: Visualization of Avg wages by Occupation

This column chart shows the average wage by occupation, providing insight into pay distribution across different roles.

- **Findings:**
 - **Protective Services** employees have the highest average wage at **£57,706**, followed by **Farming-fishing** (£57,504) and **Tech-support** (£57,031).
 - Professions such as **Craft-repair** (£53,825) and **Private Household Services** (£53,340) rank at the lower end.
- **Insights:** The observed higher wages in roles such as protective services and technical support are likely due to the specialized skill sets, certifications, and potentially elevated risk factors associated with these positions. These roles may require additional training, expertise, or experience, justifying a premium in compensation. Conversely, lower wages in areas such as household and repair services may point to a potential gap in competitiveness, which could affect the company's ability to attract and retain skilled workers in these fields. If these lower-wage roles experience high turnover or if specialized skills are needed but not compensated accordingly, it may impact operational efficiency and service quality over time.
- **Recommendation:** Conduct a comprehensive review of wage structures in lower-paid occupations, with a focus on identifying roles with high turnover rates or where specialized skills are essential but currently undervalued. Adjusting compensation to be more competitive in these areas could enhance job satisfaction, reduce turnover, and strengthen the quality and continuity of work. This approach aligns with broader workforce retention strategies, ensuring that all roles are compensated fairly relative to their demands and contribution to the organization's overall success.

Visualization 4: Employee Count by Education Level (Waterfall Chart)

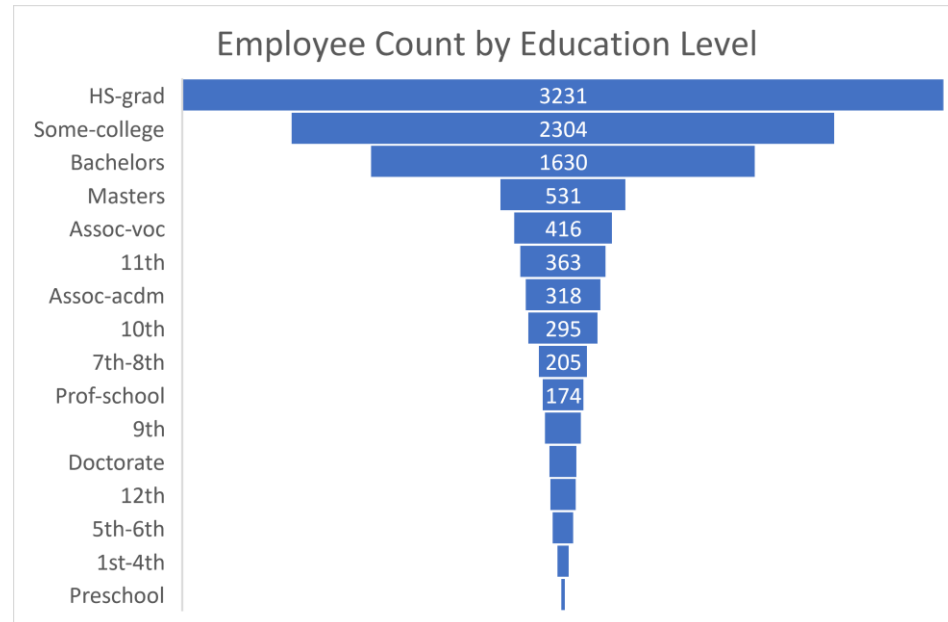


Fig 3.4: Visualization of Employee Count by Education Level

This waterfall chart represents the count of employees at each education level, illustrating the workforce's educational background.

- **Findings:**
 - The majority of employees are **high school graduates (3,231)**, followed by those with **some college education (2,304)**, and **Bachelor's degrees (1,630)**.
 - Advanced degrees such as **Doctorates (116)** and **Professional School (174)** are less common.
- **Insights:** The workforce displays a diverse range of educational backgrounds, with a notable concentration in high school and some college education levels. This distribution may suggest that many roles within the organization do not require advanced degrees, making them accessible to a wider talent pool. Additionally, this trend could reflect the company's commitment to inclusive hiring practices, providing opportunities to individuals from various educational backgrounds and fostering a workforce that values practical skills alongside formal qualifications.
- **Recommendation:** To support professional growth and encourage internal mobility, consider implementing career development programs and educational support initiatives for employees interested in advancing their qualifications. This can be particularly impactful for those in entry-level roles, providing them with pathways to gain new skills, pursue certifications, or complete higher education degrees. By investing in employee development, the company can strengthen its talent pipeline, retain motivated employees, and reduce turnover in key roles. Such initiatives align with the company's commitment

to building a skilled and adaptable workforce, prepared to meet the evolving demands of the organization.

Visualization 5: Gender Distribution Across Occupations (Stacked Column Chart)

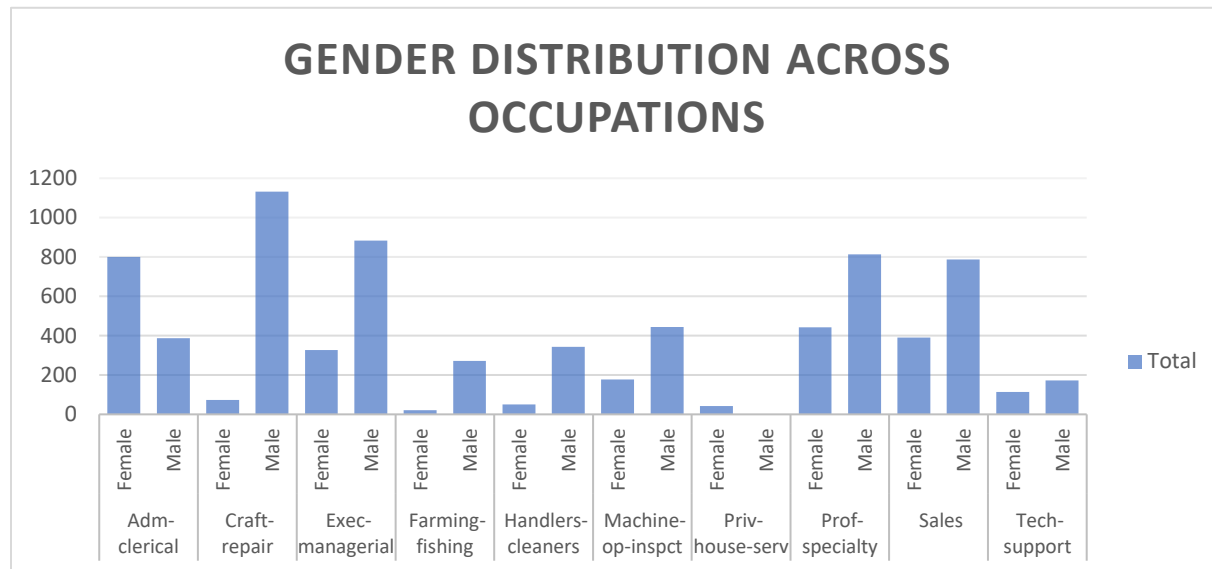


Fig 3.5: Visualization of Gender Distribution Across Occupation

This stacked column chart displays the distribution of male and female employees across various occupations.

- Findings:**
 - Female representation** is notably high in **Administrative Clerical (800)** roles but low in **Craft-repair (74)** and **Farming-fishing (21)**.
 - Male representation** dominates in roles such as **Craft-repair (1,132)** and **Farming-fishing (271)**.
- Insights:** The gender distribution across occupations reveals traditional role segmentation, with certain roles predominantly occupied by one gender. For example, technical and manual labor roles may have higher male representation, while administrative and clerical positions often see a higher proportion of female employees. This occupational segregation could inadvertently contribute to subtle wage disparities, as roles dominated by one gender may experience different compensation trends. Furthermore, this imbalance may limit diversity in specific job functions, potentially restricting the range of perspectives and experiences within these roles.
- Recommendation:** To promote a more balanced workforce, consider implementing targeted diversity initiatives across all roles, with a particular focus on positions showing significant gender imbalances. These initiatives could include:

- i. Develop outreach programs aimed at attracting underrepresented genders in specific roles, such as encouraging women to apply for technical roles or men to consider administrative positions.
- ii. Provide training and development opportunities that prepare employees to take on non-traditional roles, bridging the skill gap that might exist due to historic occupational segregation.
- iii. Establish mentorship or sponsorship programs where employees from underrepresented groups in certain roles receive guidance, support, and encouragement to progress in their careers.

These efforts can contribute to a balanced workforce where employees of all genders feel empowered to pursue diverse career paths and contribute fully to the organization's success.

Dashboard:

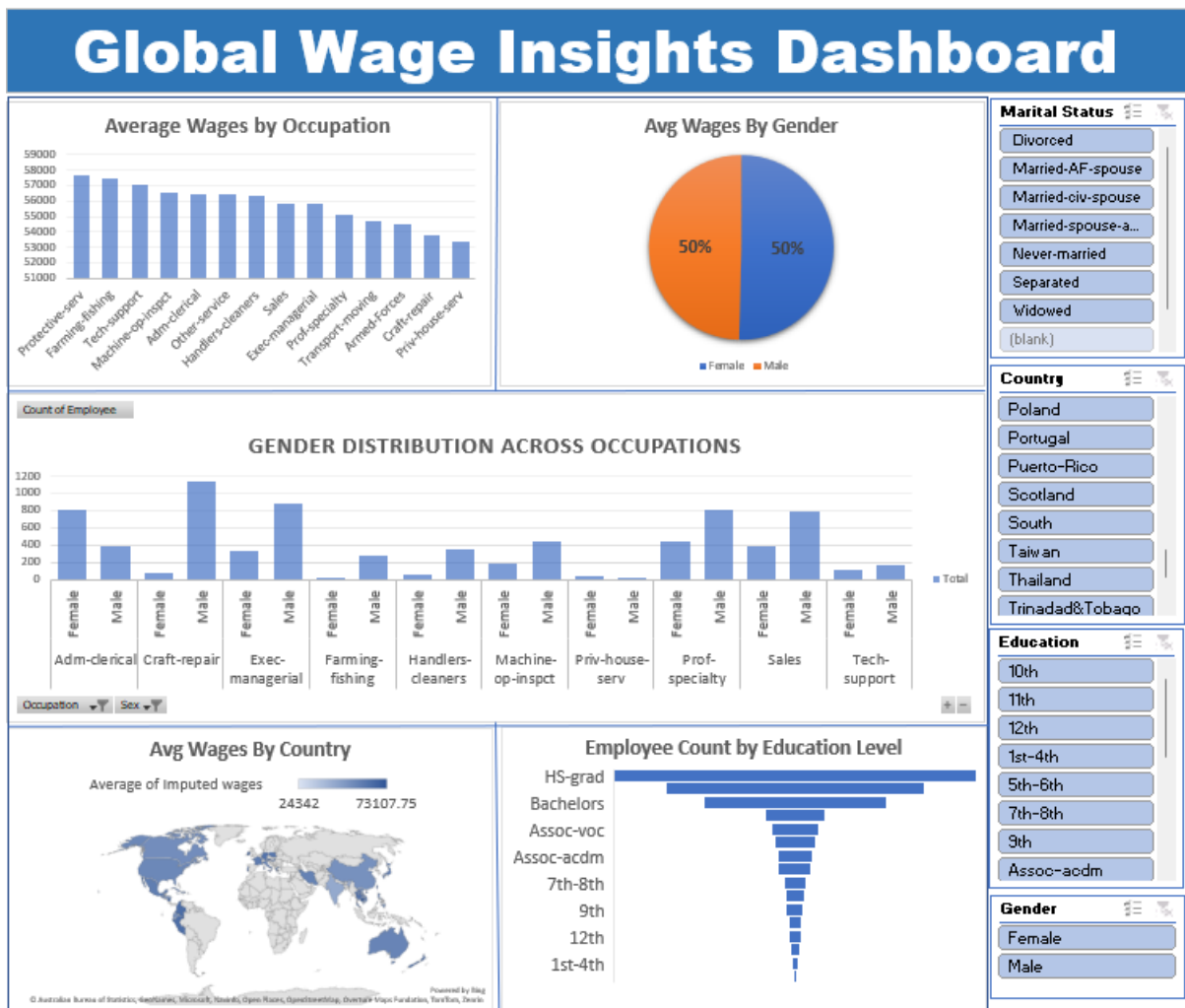


Fig 3.6: Final Interactive Dashboard in Excel

Summary and Key Findings

The dashboard's interactive components allow for granular insights into wage and demographic patterns. Major findings include:

1. **Geographic Wage Disparities:** Significant wage variations exist across countries, with some regions offering substantially higher wages than others. This disparity warrants further investigation to ensure equitable pay structures and competitiveness in each market.
2. **Gender Pay Parity:** Overall, average wages between male and female employees are nearly equal, indicating broad success in achieving gender pay equality. However, role-specific or location-based analysis could help identify any isolated discrepancies.
3. **Occupational Wage Trends:** Higher wages are observed in protective, technical support, and farming-fishing roles, reflecting skill-specific or risk-based compensation. Lower wages in household and repair services highlight opportunities for wage adjustments to attract skilled labor.
4. **Educational Composition:** The workforce has a diverse educational background, with most employees having high school or some college education. Advanced degrees are less common, suggesting that many roles do not require higher education or that opportunities exist to support further employee development.
5. **Gender Distribution in Occupations:** Occupational gender distribution highlights traditional role segmentation, with higher female representation in clerical roles and male dominance in technical and manual roles. This imbalance may limit diversity and contribute to implicit wage gaps.

Recommendations

1. **Wage Adjustments and Competitive Analysis:** Regularly review and adjust wages to align with market standards, especially in countries and occupations with lower average wages. This will help ensure fair compensation and attract top talent globally.
2. **Promote Gender Diversity in Roles:** Encourage more balanced gender representation across occupations, particularly in roles traditionally dominated by one gender. This can be achieved through targeted recruitment, training, and mentorship programs.
3. **Educational Development Programs:** Invest in career development and educational support for employees, particularly those at entry-level or in positions with low wage mobility. This will foster internal growth and support workforce stability.
4. **Ongoing Monitoring for Pay Equity:** Conduct periodic gender pay audits to maintain pay equity and detect any emerging disparities. This ensures ongoing commitment to fair compensation across all demographic groups.

Conclusion

The **Global Wage Insights Dashboard** is a strategic tool that brings clarity and depth to the analysis of wage distribution, demographic trends, and workforce composition across the company. Through its intuitive design and interactive elements, this dashboard transforms complex data into accessible insights, enabling stakeholders to make informed, data-driven

decisions that support equitable compensation practices and foster a diverse, inclusive workplace.

The dashboard's interactive components—such as filters and slicers—empower users to explore data dynamically, providing the flexibility to analyze specific workforce segments by factors like gender, occupation, education, and geographic location. This adaptability enhances the utility of the dashboard for diverse audiences within the organization, from executives seeking a high-level overview to HR and finance teams focused on specific compensation or demographic metrics.

The comprehensive visualizations within the dashboard—maps, pie charts, column charts, and stacked bar charts—illustrate critical workforce patterns, such as gender representation across roles, average wages by country, and education levels. By visualizing these trends, the dashboard enables quick identification of potential areas for improvement, such as regions with wage disparities, occupations with gender imbalances, and roles where wage adjustments may be needed to remain competitive.

The insights gathered from these visualizations underscore the importance of fair and consistent compensation structures across all demographics and geographies. The dashboard's design also supports the company's commitment to diversity and inclusion by highlighting areas where greater representation or equity may be beneficial. For example, the ability to compare average wages between male and female employees within the same occupation and location allows for a more nuanced analysis of gender pay equity, while visualizations of ethnic and educational diversity provide a holistic view of the workforce composition.

The recommendations derived from the dashboard aim to strengthen employee satisfaction and retention by addressing the following key areas:

1. **Equitable Compensation:** By identifying and addressing any wage disparities across gender, geography, and job roles, the company can foster a fair and supportive work environment, ensuring that all employees are compensated fairly for their contributions.
2. **Enhanced Diversity and Inclusion:** The dashboard highlights areas where representation may be imbalanced, such as occupations dominated by one gender. Encouraging diversity in these roles can create a more inclusive workforce and bring diverse perspectives to all levels of the organization.
3. **Career Development and Retention:** Insights into the educational background and wage distribution among employees can guide initiatives to support career progression, skill development, and internal mobility, especially in roles with lower educational attainment or compensation levels.
4. **Global Competitiveness:** By comparing wages across different countries and regions, the company can ensure that compensation remains competitive in each market. This global view enables the company to adjust pay structures based on local economic conditions, enhancing its ability to attract and retain top talent internationally.

In summary, the Global Wage Insights Dashboard is more than just a data visualization tool; it is a comprehensive resource for guiding strategic workforce planning. It empowers decision-

makers to prioritize equity, support diversity, and ensure competitive compensation, ultimately contributing to a balanced, satisfied, and high-performing workforce. By implementing the recommendations identified through the dashboard, the company positions itself to enhance overall organizational competitiveness, foster a positive work culture, and retain its most valuable asset—its people.