**EMPLOYER’S**

**DATA**

Submitted by:

Student Name: Nandish P

Roll Number: 2461436

Course: Fundamentals of Data Analytics

Instructor: Dhiraj alate

Institution: Christ university

Date: 18/08/2025

**Abstract**

This study focuses on analysing employer data to evaluate salary distribution, employee experience, and gender disparities across various roles and education levels. The dataset, initially unstructured with missing values and redundant rows, was cleaned and prepared for meaningful analysis. Key objectives included comparing male and female salary allocations, examining the relationship between years of experience and compensation, and identifying roles with the highest contribution to overall salary costs. The analysis revealed that while males generally earn slightly higher than females across roles, significant gender pay gaps exist in managerial positions. Interns showed the lowest experience levels, whereas employees with advanced degrees (Master’s, PhD) accounted for the highest salary expenditures. Outcomes of the study highlight the importance of data-driven decision-making in human resource planning, salary structuring, and policy formulation. This work also enhanced skills in data cleaning, analytical reasoning, and effective communication of insights, supporting organizational goals of fairness, efficiency, and equity.

**Tools and Technologies used**

|  |  |
| --- | --- |
| Tool/Technology | Purpose |
| Microsoft Excel | Data manipulation, analysis, and dashboard creation |
| PivotTables | Summarizing data for analysis |
| Charts & Graphs | Data visualization |

**Scope of the project**

* Focused on data cleaning, analysis, and visualization only.
* No programming languages (like Python or R) or advanced statistical modelling used.
* All work is contained within a single Excel file.
* Analysis is limited to the provided dataset.

**Data Cleaning**

The initial inspection of the dataset, which contains 205 survey responses, revealed that it was generally well-structured. The primary cleaning task involved addressing missing data.

Initial State of the Data:

\* The dataset was complete, with all 205 rows having data for most of the 20 columns.

\* There were no duplicate rows found in the dataset.

\* Handling Missing Values:

\* The only column with missing data was 'What is your occupation? (optional)', which had 60 missing entries.

\* Since this field was optional, it's reasonable to assume that the missing entries indicate respondents who chose not to provide this information.

\* Action Taken: To ensure consistency for analysis, these missing values were filled with the placeholder text "Unknown". This retains all rows for analysis while clearly marking where occupation data was not provided.

Features Created for Analysis

The original dataset consists almost entirely of categorical survey responses. To facilitate more insightful analysis, the following new feature was engineered by transforming an existing column into a more structured format.

New Feature: Age Group

\* Rationale: The original column 'What is your age range?' contains text-based age brackets (e.g., "18-24", "25-34"). While useful, creating a standardized and ordered categorical feature makes it easier to segment and compare responses across different age demographics.

Process:

\* A new column named 'Age Group' was created.

\* The values from the original age range column were mapped directly into this new column, creating a clean, dedicated feature for age-based analysis. For this dataset, the values were already well-defined categories, so the new feature standardizes this into a distinct column for analysis.

**Dashboard Design Strategy**

A well-designed dashboard should guide the user's eye and present key information in a logical and easy-to-understand way. The goal is to move from a high-level overview to more specific details, allowing for quick insights and deeper analysis.

**Layout and Design**

A good dashboard layout follows a top-down, left-to-right hierarchy, mirroring how people naturally read.

* **Top Section:** This is the "at-a-glance" area. It should contain key performance indicators (KPIs) or single-number metrics that summarize the most critical information. For example, a card showing "Total Revenue”, or "Number of Active Users" would go here. This gives the user an immediate sense of the data's health.
* **Middle Section:** This is the primary content area where you present your main visualizations. It should be dedicated to the most important trends and comparisons. This is where you'll place charts that address the core questions of your analysis.
* **Bottom Section:** This area is for more detailed or secondary information, such as tables, detailed lists, or less critical visualizations that a user might want to explore after understanding the main points.

Consistency is key. Use a clean and consistent colour palette, maintain uniform spacing between elements, and use a simple, readable font to ensure the dashboard is not cluttered and is easy to scan.

**Choice of Visualizations**

The best visualizations are chosen to match the type of information you are trying to convey.

* **Line Chart:** A line chart is the ideal choice for **trends over time**. If you are showing how a value changes day-to-day, month-to-month, or year-to-year (for example, user sign-ups over a year), a line chart clearly illustrates the trend and rate of change.
* **Bar Chart:** A bar chart is perfect for **comparisons between categories**. Whether you are comparing sales by product, user demographics by region, or survey responses by age group, a bar chart makes it easy to see which category is highest or lowest. A horizontal bar chart can be particularly effective when category names are long.
* **Pie or Donut Chart:** Use a pie or donut chart to show **proportions of a whole**. This type of chart is best for a small number of categories (usually less than five) to show how each part contributes to the total, such as the market share of different product lines.
* **Key Performance Indicator (KPI) Cards:** These are single, large numbers with a label. They are great for a quick summary of a crucial metric like total revenue, average order value, or a conversion rate. Placing these at the top of the dashboard provides immediate context.

**Interactive Elements**

Adding interactive elements transforms a static report into a dynamic tool that users can explore on their own.

* **Slicers/Filters:** These are essential for allowing users to **filter data based on different dimensions**. You can add filters for time ranges, categories or any other attribute. This lets the user customize the view to answer specific questions.
* **Drill-Down Capability:** This feature allows a user to click on a high-level visualization (like a total sales number) and **see the underlying data in more detail** (like sales by individual store or salesperson). This provides a seamless way to move from a summary to the raw data without leaving the dashboard.
* **Tooltips:** When a user hovers over a data point on a chart, a tooltip can appear to show more precise details, such as the exact value and date. This adds a layer of useful information without cluttering the main visualization.

**Questions and Solutions**

### **Q1: Which role has the highest total salary?**

**Solution:** The **Master** category has the highest total salary of **661,775,000**.

### **Q2: Which role has the lowest total experience years?**

**Solution:** **Interns** have the lowest total experience of **301 years**.

### **Q3: On average, who earns more across all roles — Male or Female employees?**

**Solution:**

* Total Male salary = **593,350,000**
* Total Female salary = **560,465,000**

### **Q4: Which role has the largest gender salary gap?**

**Solution:** **Manager** role shows the largest difference:

* Male: **176,305,000**
* Female: **163,670,000**
* Gap: **12,635,000**

### **Q5: What is the average salary per year of experience for Executives?**

**Solution:**

* Total Salary (Executives) = **219,850,000**
* Total Experience (Executives) = **29,341 years**
* Average Salary per Year of Experience ≈ **7,490 per year**.

**Challenges faced and solutions**

|  |  |  |
| --- | --- | --- |
| **Challenge** | **Description** | **Solution** |
| Messy Headers | Dataset had extra rows (titles, labels) before actual data started. | Skipped unnecessary rows and redefined proper headers. |
| Missing Values | Many columns (e.g., *Unnamed: 7–20*) | Dropped irrelevant empty columns to simplify dataset. |
| Mixed Data Rows | Some rows were totals (e.g., *Grand Total*) instead of actual role data. | Separated totals from analysis to avoid double-counting. |
| Inconsistent Labels | “Row Labels” row created confusion in column naming. | Renamed columns to meaningful names (e.g., Role, Salary Female, Salary Male). |
| Gender Salary Comparison | Male and Female salaries were in different columns, making comparisons tricky. | Combined data logically and calculated differences (e.g., gender pay gap). |

**Outcomes**

Summary of Key Insights

 Clear **Understanding of Salary Distribution**

* A detailed comparison of male vs. female salaries across roles and education levels will show where disparities exist.

 Insight **into Experience-Salary Relationship**

* Identification of how years of experience correlate with salary, helping to determine whether higher experience justifies higher pay.

 Gender **Pay Gap Analysis**

* Specific roles with significant salary gaps between male and female employees will be revealed, supporting fairness and equality initiatives.

 Role**-Based Contribution to Salary Expenses**

* Recognition of which roles (e.g., Manager, Executive, Analyst) have the highest salary costs, helping in workforce budgeting and planning.

 Data**-Driven HR and Management Decisions**

* Actionable insights that guide recruitment strategies, promotion criteria, and salary structure improvements.

Reflection on Skills

 Data **Cleaning & Preparation**

* Learned how to handle raw datasets with missing values, extra headers, and redundant columns by restructuring them for analysis.

 Data **Analysis & Interpretation**

* Gained experience in extracting meaningful insights such as salary distribution, gender pay gaps, and experience-to-salary ratios.

 Critical **Thinking**

* Improved ability to ask the right questions (e.g., which role earns the most? Where is the largest gender gap?) and interpret data-driven answers.

 Visualization **& Communication**

* Developed skills in presenting data using tables, summaries, and charts that make complex information easier to understand.

 Decision**-Making Support**

* Understood how analytical results can inform HR and management decisions, bridging technical analysis with real-world business applications

**Objectives**

 Analyse **Salary Distribution**

* To compare salary allocation between male and female employees across different roles and education levels.

 Evaluate **Experience Contribution**

* To assess the relationship between years of experience and total salary across various roles.

 Identify **Gender Pay Gaps**

* To highlight positions with the largest salary differences between male and female employees.

 Determine **High-Value Roles**

* To identify which roles (e.g., Analyst, Engineer, Manager, Executive) contribute the most to overall salary expenses.

 Support **Decision-Making**

* To provide insights that can guide HR and management in policymaking regarding hiring, promotions, and salary structures.

**Screenshots of OutputsA close-up of a graph

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

**Conclusion**

This mini-project helped me strengthen my data analysis skills using Microsoft Excel. I gained practical insights into data cleaning, transformation, and creating effective visualizations to communicate findings. The hands-on analysis of a real-world dataset also enhanced my understanding of how data can be used to solve business problems and make informed decisions.

**References**

https://www.kaggle.com/datasets/gmudit/employer-data