Internee Name: Saad Tariq

Company Name: Robx.ai

Final Report: Salary Analysis and Dashboard Creation

---

1. Introduction

This report documents a comprehensive analysis of job salaries, from data collection and preprocessing to visual representation through a Power BI dashboard. The primary objective was to analyze salary data in relation to various factors, such as years of experience, age group, and gender, to uncover patterns and insights that could benefit HR decision-making and industry analysis.

2. Objective

The goal of this project was to perform data preprocessing, analyze a dataset related to job salaries, and create an insightful and visually appealing dashboard in Power BI. The steps involved included:

1. Sourcing data from Kaggle.
2. Cleaning and preprocessing the data using Python.
3. Designing an interactive dashboard in Power BI to visualize key salary-related metrics.

3. Data Collection

I sourced the dataset from Kaggle's repository: [Salary Prediction for Beginners](). This dataset contained information about job salaries based on various factors, such as experience level, job title, age, and gender.

4. Data Preprocessing and Cleaning in Python

I used **Python** for the data preprocessing and cleaning phase. Python libraries, including **Pandas**, **NumPy**, and **Matplotlib**, facilitated this process. The following steps were performed to ensure data quality and integrity:

*   **Handling Missing Values**: I detected missing data points and dealt with them by either filling in plausible values or removing records where data was incomplete.
*   **Outlier Detection and Treatment**: Outliers were identified using statistical methods. Where appropriate, they were handled by either removing or adjusting values based on domain knowledge and statistical analysis.
*   **Data Inconsistencies**: I identified any inconsistencies, such as incorrect data types or mislabeled columns, and corrected them to ensure data coherence.
*   **Feature Engineering**: Additional columns, such as **Age Group** and **Experience Level**, were created to enhance analysis. These engineered features provided further segmentation for visual analysis in Power BI.

After these preprocessing steps, I saved the cleaned data as cleaned_data.csv, which I later imported into Power BI.

5. Power BI Dashboard Creation

Using the cleaned data, I developed an interactive and informative dashboard in **Power BI**. The dashboard includes a range of visualizations that provide insights into salary trends across various factors.

6. Dashboard Components and Insights

Below is a breakdown of the key components and insights derived from the Power BI dashboard:

1. **Total Employees and Employee Countries**:
   o The dashboard highlights 5,148 employees across 5 countries in the dataset. These metrics provide a snapshot of the dataset's size and geographic distribution.
2. **Average and Max Salary Metrics**:
   o The Average Salary metric is 114.27K, and the Max Years of Experience in the dataset is 34 years. These values give a quick overview of the dataset's salary structure and employee tenure.
3. **Number of Employees by Age Group**:
   o This bar chart illustrates the distribution of employees across different age groups. The 20-30 age group has the highest representation, indicating that the majority of employees in the dataset are early-career professionals.
4. **Employees by Experience Level**:
   o Another bar chart shows the distribution of employees by experience level: Mid-level, Junior, and Senior. This helps us understand the workforce composition in terms of experience, with a higher concentration in the mid-level category.
5. **Gender Distribution**:
   o The gender distribution, represented in a donut chart, is relatively balanced, with 54.55% Female and 45.45% Male employees.
6. **Yearly Data**:
   o The years represented in the dataset range from 2019 to 2024. This time-based segmentation allows for year-over-year trend analysis.
7. **Salary by Years of Experience**:
   o A line chart depicting median salary by years of experience shows an upward trend, indicating that salaries generally increase with experience. This trend is expected in most industries as experience is a key factor in salary determination.
8. **Salary Variation by Gender Across Countries**:
   o A line chart compares salary variation between genders across different countries. This visualization highlights salary disparities across countries and genders, providing insights into how location and gender can influence salary.
9. **Country Distribution**:
   o The geographical distribution of employees across countries is represented in both a map and a tree map. Countries such as the USA, Australia, China, UK, and Canada have the most representation, allowing for cross-country comparison in salary trends and employee demographics.

7. Insights and Recommendations

Insights:

- **Age and Experience Influence**: The dataset shows that younger age groups have higher representation, particularly in mid-level and junior positions. Senior roles are typically occupied by those with more years of experience.
- **Gender Salary Parity**: While the dataset shows a near balance in gender distribution, further analysis reveals gender-based salary disparities in some countries. This could inform policies aimed at achieving greater gender salary parity.
- **Regional Salary Trends**: Salaries vary significantly across different countries, with the USA and Australia exhibiting higher median salaries. This may reflect differences in cost of living, industry standards, and local economic conditions.
- **Experience-Driven Salary Growth**: A clear positive correlation exists between years of experience and salary. This suggests that investing in long-term employee development could yield returns in terms of higher salaries and retention.

---

8. Conclusion

This project provided hands-on experience in data preprocessing, visualization, and analysis. Starting with raw data from Kaggle, I used Python to clean and prepare the dataset, creating new features to enhance the analysis. Afterward, I used Power BI to create an interactive and insightful dashboard, showcasing key metrics and trends in salary data across various dimensions.

This report serves as a testament to the effectiveness of data-driven decision-making, providing actionable insights for understanding salary trends. By following this process, I developed valuable skills in data preparation, dashboard design, and data interpretation, which can be applied to real-world business scenarios.