Employee Data Analysis Report

Prepared by: Bara’a Al-Yateem

Date: 7/20/2025

python-docx

Company x /  Data Cleaning and Exploratory Data Analysis

# Executive Summary

This report provides an overview of employee data after a thorough cleaning process and exploratory data analysis. Various demographic and departmental attributes have been visualized to uncover patterns and relationships that can support strategic decisions. This exploratory data analysis showcases different types of connections and their relevance with the aim of improving the company’s employee knowledge.

# Table of Contents

[Executive Summary 1](#_Toc203823236)

[Table of Contents 2](#_Toc203823237)

[Table of figures: 3](#_Toc203823238)

[Table of tables: 3](#_Toc203823239)

[1. Introduction 4](#_Toc203823240)

[2. Data Understanding 4](#_Toc203823241)

[2.1 Data Cleaning 4](#_Toc203823242)

[2.2 Data Description 4](#_Toc203823243)

[2.3 Data Visualization 4](#_Toc203823244)

[3. Univariate Analysis 5](#_Toc203823245)

[4. Bivariate Analysis 7](#_Toc203823246)

[5. Key Findings and Insights 12](#_Toc203823247)

[6. Recommendations 13](#_Toc203823248)

[Appendix 13](#_Toc203823249)

[I-Cleaned Data 13](#_Toc203823250)

[II-Data Dictionary Table for the dataset 14](#_Toc203823251)

[III-Summary Statistics Table 15](#_Toc203823252)

# Table of figures:

[Figure 1: Ethnicity Distribution Pie Chart 5](#_Toc203820556)

[Figure 2:Country Diversity Pie Chart 5](#_Toc203820557)

[Figure 3:Gender Balance Pie Chart 6](#_Toc203820558)

[Figure 4:Annual Salary per Job Title 7](#_Toc203820559)

[Figure 5:Age per Job Title 8](#_Toc203820560)

[Figure 6: Unique Business Units per Department 9](#_Toc203820561)

[Figure 7:Bonus % vs Salary 10](#_Toc203820562)

[Figure 8: Salary Comparison by Gender 11](#_Toc203820563)

# Table of tables:

[Table 1:Dataset Data dictionary 14](#_Toc203823127)

[Table 2:Summary Statistics Table 15](#_Toc203823128)

# Introduction

This report presents the results of an exploratory data analysis (EDA) on the latest employee dataset (Employee Sample Data - A). With the aim of understanding workforce characteristics and compensation trends to work on improving work conditions and therefore performance in the long term.

# Data Understanding

## 2.1 Data Cleaning

- Filled missing non-critical fields with defaults.  
- Dropped rows with critical missing values.  
- Reformatted date fields for uniform structure.  
- Added fictional test records to test the accuracy of graph representation.

## 2.2 Data Description

- Numerical fields such as Age and Salary were described using summary statistics (Min, Max, Median).  
- Categorical fields such as Gender and Department were analyzed by frequency of occurrence throughout the dataset.

## 2.3 Data Visualization

- Various charts were created to explore distributions and relationships. -This report showcases analysis across both one variable and two. -Each graph will

# 3. Univariate Analysis

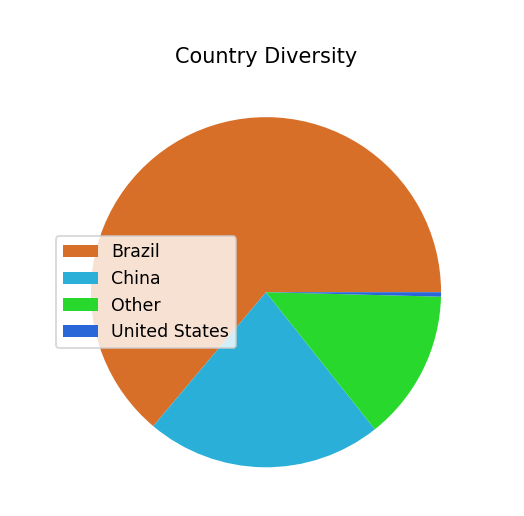


Figure 2:Country Diversity Pie Chart

Figure one displays the ethnic distribution of employees. Asians form the majority, indicating a concentration in that demographic, while Black and Caucasian groups are moderately represented. Latino and Other categories make up a small portion of the workforce, highlighting limited ethnic diversity in those groups.

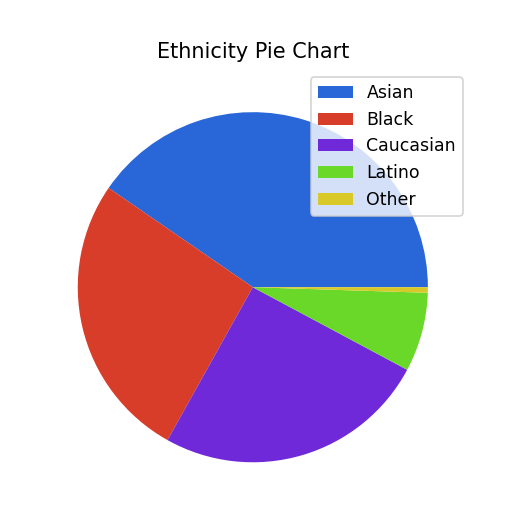


Figure 1: Ethnicity Distribution Pie Chart

Figure two visualizes the employee distribution by country, highlighting that the majority of employees are from Brazil, with smaller proportions from China, Other countries, and the United States. This indicates a significant concentration of the workforce in Brazil, with diverse representation from other regions.

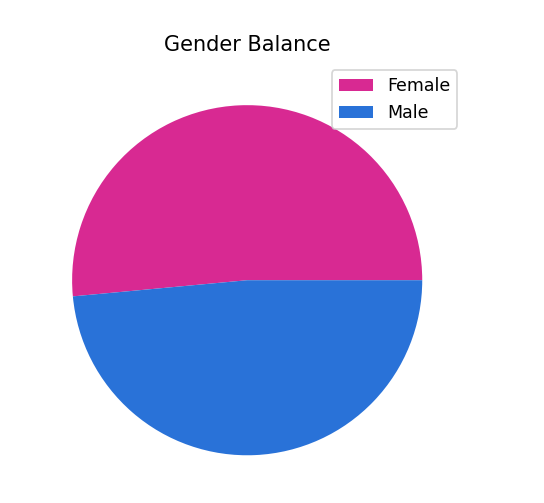


Figure 3:Gender Balance Pie Chart

Figure three depicts the gender balance within the organization, showing that the number of females and males is nearly equal. Slightly tipping toward the female percentile.

The graphs presented in the EDA report aim to highlight information, connect and form relations and infer characteristics relating each section of data to help shape future plans of improvement.

# Bivariate Analysis

The figure above displays the annual salary distribution across various job titles within the organization. Each bar represents the minimum, maximum and the middle salary for a specific position, illustrating significant variation in compensation levels. Notably, executive roles like Vice President and Director tend to have higher salary ranges, with some reaching above $200,000, while entry-level or technical roles such as Network Administrator and System Analyst exhibit lower salary ranges. This visualization helps in understanding the salary hierarchy and disparities among different job functions.

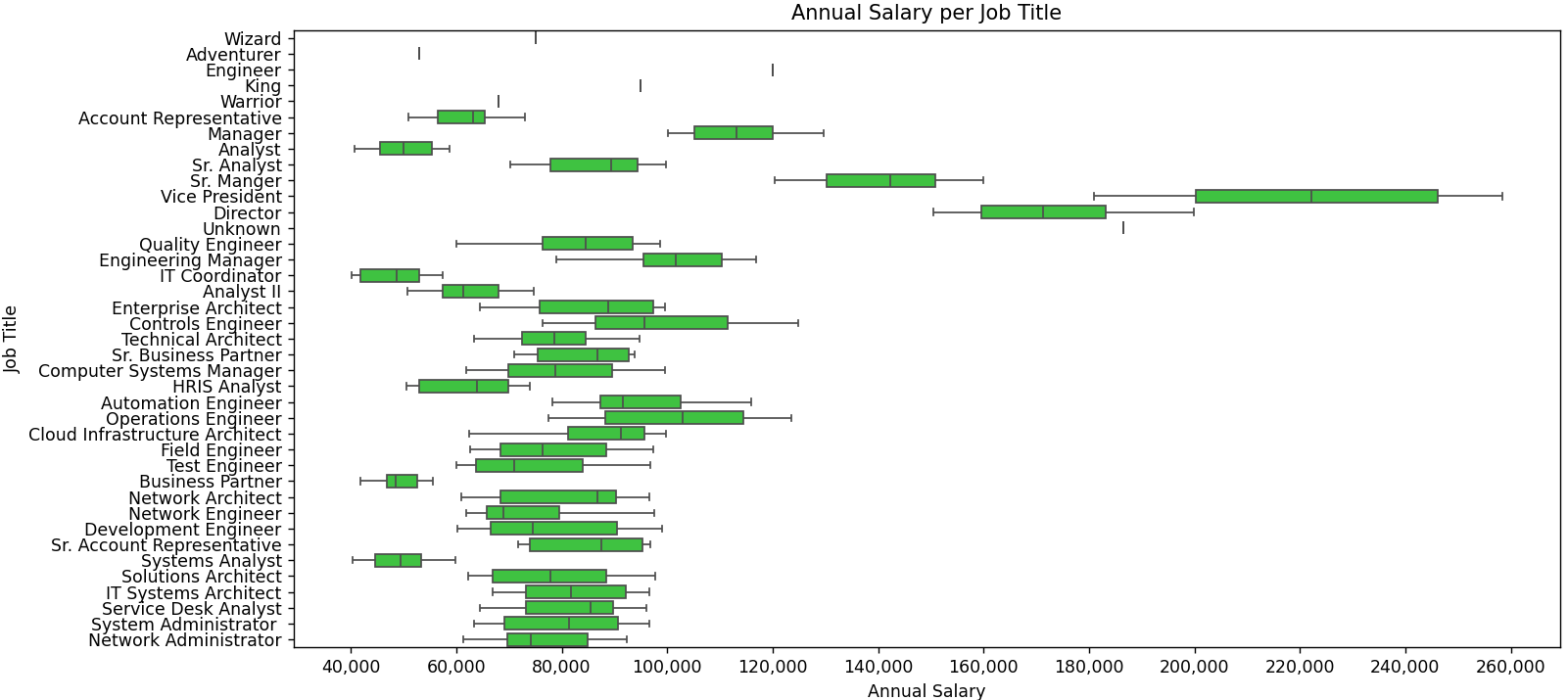


Figure 4:Annual Salary per Job Title

Figure five illustrates the age distribution across various job titles within the organization. It reveals that most employees are in the age range of approximately 30 to 50 years, indicating a predominantly experienced workforce. Certain roles, such as King, have a broader age range, extending above 60 years. The data highlights the diversity in age among employees, with a concentration of mid-career professionals in roles like Software Engineer, System Administrator, and Analyst. This information helps in understanding the age demographics across different job functions.

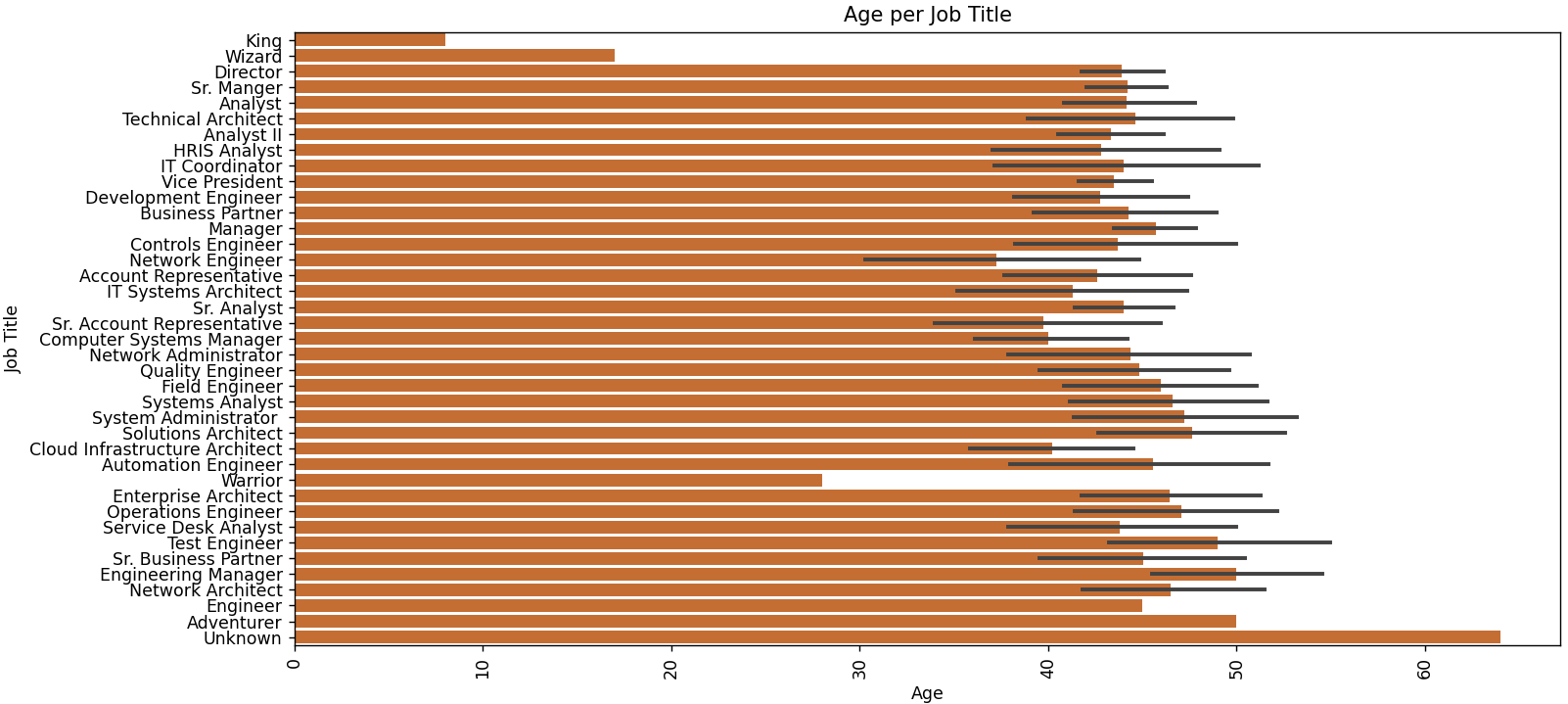


Figure 5:Age per Job Title

The bar chart in the figure above presents the number of unique business units within each department of the organization. It shows that most departments, including Accounting, Engineering, Finance, Human Resources, Marketing, IT, and Sales, have four units each. However, the Disney department stands out with the highest number of units, totaling five. This suggests a more complex or diversified structure within the Disney department compared to others. This visualization helps to understand the organizational complexity across different departments.

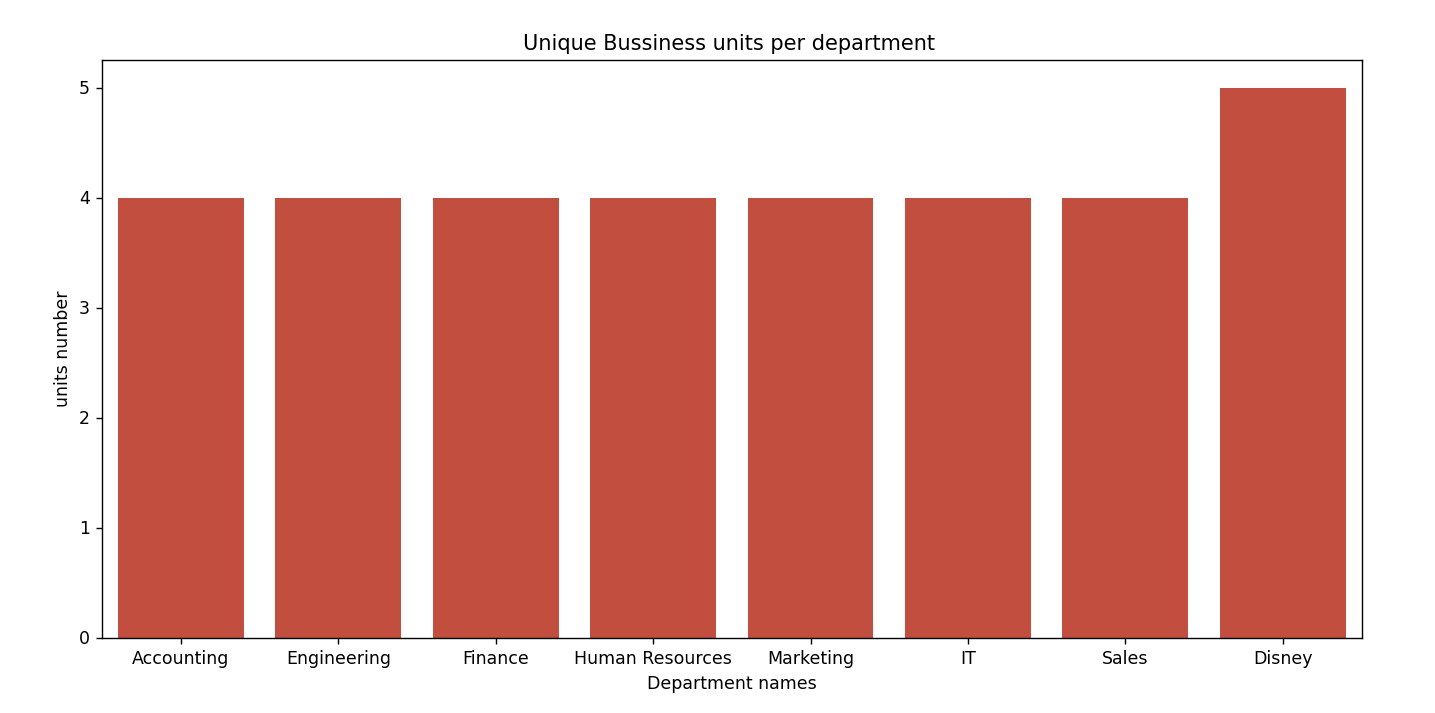


Figure 6: Unique Business Units per Department

The scatter plot in figure seven illustrates the relationship between annual salary and bonus, highlighting a positive correlation. As the annual salary increases, the bonus tends to increase as well, although with some variability. The data points show that higher salaries are generally associated with higher bonuses, suggesting a connection between compensation levels and bonus amounts within the organization. This visualization can provide insights into the compensation structure and how bonuses are distributed relative to salaries.

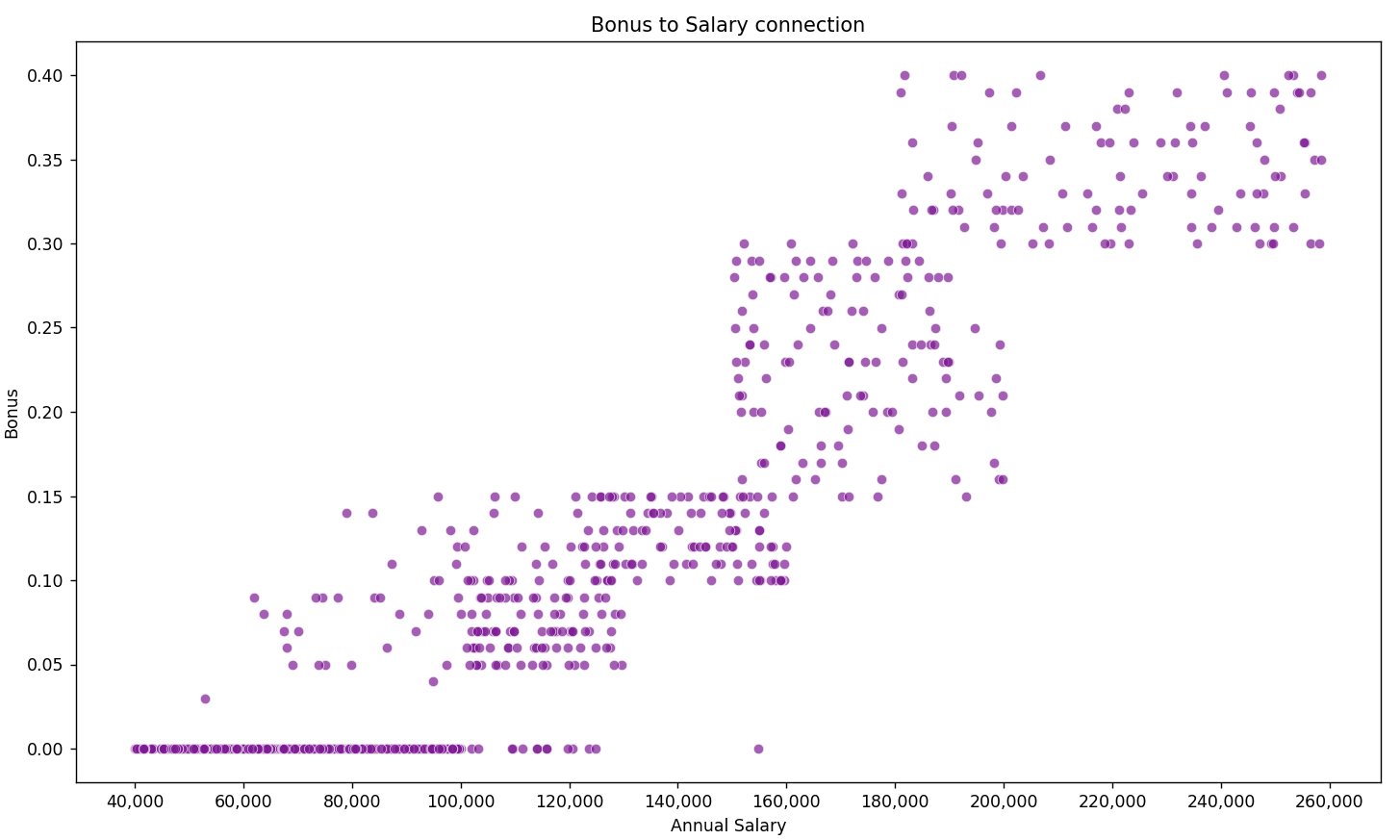


Figure 7:Bonus % vs Salary

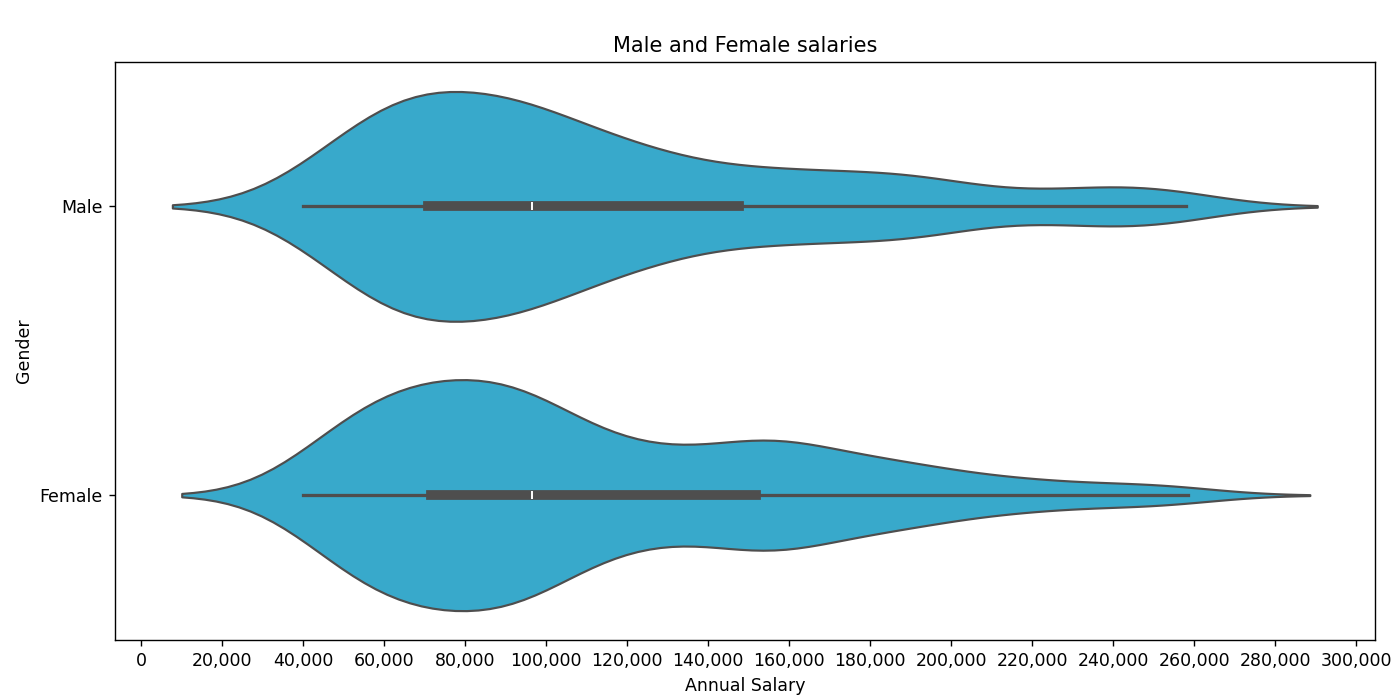


Figure 8: Salary Comparison by Gender

The violin plot compares the annual salaries of male and female employees. It reveals that the salary distributions for both genders are quite similar, with overlapping ranges. The median salaries appear to be nearly equal, indicating gender parity in compensation within the organization. The plot also shows a broader range of salaries for males, suggesting some higher earners, but overall, there is no significant disparity evident in the salary distribution between males and females.

# Key Findings and Insights

* **Workforce Distribution by Country:** The organization has a significant concentration of employees in Brazil, with smaller but diverse representations from China, the United States, and other countries. This indicates a predominantly regional workforce with some international diversity.
* **Gender Balance:** The gender distribution is relatively balanced, with nearly equal numbers of males and females across the organization, suggesting an inclusive environment regarding gender representation.
* **Salary and Bonus Relationship:** There is a positive correlation between annual salary and bonus amounts, indicating that higher earners tend to receive larger bonuses. This reflects a structured compensation approach where bonuses are aligned with salary levels.
* **Age Demographics:** The majority of employees are in the age range of 30 to 50 years, highlighting a workforce that is primarily mid-career professionals with considerable experience. Some roles, such as King, include employees over 60, indicating longevity and experience in certain positions.
* **Salary Variability by Job Title:** Salaries vary significantly across roles, with executive positions like Vice President and Director earning notably higher salaries. Entry-level roles and technical positions tend to have lower salary ranges, reflecting typical organizational salary hierarchies.
* **Organizational Structure:** The number of unique business units per department varies, with the Disney department having the most units, suggesting a more complex organizational structure within that division.

# 6. Recommendations

* **Enhance Diversity and Inclusion Efforts:** While gender balance appears satisfactory, further initiatives could be explored to promote broader diversity, including ethnic and regional representation, especially considering the workforce's geographic distribution.
* **Career Development Programs:** Given the age demographics, implementing targeted training and leadership development programs can support career progression for mid-career employees and prepare younger talent for senior roles.
* **Salary Transparency and Equity:** The salary distribution analysis indicates disparities across roles and experience levels. Conducting a comprehensive pay equity review can ensure fair compensation practices, fostering employee trust and satisfaction.
* **Additional Compensation Incentives:** To motivate high performers and retain top talent, consider expanding bonus programs or introducing other incentive schemes aligned with organizational goals.
* **Organizational Simplification:** The varied number of business units suggests complexity within some departments. Streamlining or consolidating units could improve operational efficiency and clarity in reporting structures.

# Appendix

## I-Cleaned Data

[Link to clean data.](https://github.com/BaraaAlyateem/Task3_DataCleaning/blob/master/Project_Results.xlsx)

## II-Data Dictionary Table for the dataset

|  |  |  |
| --- | --- | --- |
| Variable Name | Description | Definitions and Notes |
| Employee ID (EEID) | Unique identifier assigned to each employee within the organization | Used for record-keeping; not linked to personal identifiers publicly |
| Full Name | The complete name of the employee | Typically used internally; often omitted or anonymized in published reports |
| Job Title | The designated position or role held by the employee | Examples: Engineer, Manager, VP; indicates level and responsibilities |
| Department | The major organizational division the employee belongs to | Examples include R&D, HR, Finance, Operations |
| Business Unit | Subdivision within a department or organization representing a specific operational segment | Number of units within a department; indicates organizational complexity |
| Gender | The gender identity of the employee | Categories: Male, Female |
| Ethnicity | Ethnic or racial background of the employee | Examples: Asian, Black, Caucasian, Latino, Other |
| Age | The age of the employee at the time of data capture | Represents employee's age; range from early 20s to over 60 |
| Hire Date | The date when the employee first joined or was hired by the organization | Used to calculate tenure, experience, or seniority within the company |
| Annual Salary | The total gross salary earned annually, before taxes and deductions | Reflects compensation level; varies by role, experience, geography |
| Bonus Percentage | The annual bonus as a percentage of the base salary | Usually a positive value; can be zero; indicates performance-based incentives |
| Country | The country where the employee is based or recorded for employment purposes | Indicates geographic distribution; e.g., US, China, Brazil |
| City | The specific city within the country where the employee works or is stationed | Provides geographical granularity within the country |
| Exit Date | The date when the employee left or was terminated from the organization | If null, indicates the employee is currently employed |

Table 1:Dataset Data dictionary

## III-Summary Statistics Table

|  |  |  |
| --- | --- | --- |
| Variable | Statistic | Value/Range/Description |
| Age | Median | ~40-45 years |
|  | Range | Most employees between 30-50; |
| Salary | Minimum | ~$40,000 |
|  | Maximum | Over $260,000 |
|  | Median | Not specified precisely, but typical for organization |
| Gender | Female | ~50% |
|  | Male | ~50% |
| Ethnicity | Asian | Predominant ethnic group |
|  | Others | Black, Caucasian, Latino, etc. |
| Number of Departments | Varies; e.g., Disney has 3 units | Organization complexity varies |
| Bonus | Correlation with salary | Positive; higher pay, larger bonus |

Table 2:Summary Statistics Table