



# HR Analytics and Employee Management

Data Analysis Report

## Contents

Introduction .....	2
Data Analysis Lifecycle Stages.....	2
1. Business Understanding.....	2
2. Data Understanding .....	2
2.1 Data Sources .....	2
2.2 Key Metrics .....	3
<b>3. Data Preparation (Cleaning)</b> .....	3
3.1 Data Cleaning Steps .....	3
<b>4. Data Analysis</b> .....	4
4.1 Descriptive Analysis .....	4
4.2 Diagnostic Analysis .....	4
<b>5. Data Visualization</b> .....	5
5.1 Gender and Ethnicity Breakdown.....	5
5.2 Salary and Bonus Distribution .....	5
5.3 Hiring and Exit Trends.....	7
5.4 Tenure and Retention Analysis.....	7
<b>5.5 Geographical Distribution of Employees</b> .....	7
<b>6. Data Interpretation and Insights</b> .....	7
6.1 Gender and Ethnic Diversity .....	8
6.2 Salary Inequity .....	8
6.3 Turnover and Retention .....	8
6.4 Age and Tenure.....	8
<b>7. Recommendations</b> .....	8
7.1 Strengthening Diversity and Inclusion.....	8
7.2 Addressing Salary Disparities.....	8
7.3 Enhancing Retention Strategies .....	8
7.4 Expanding Regional Workforce .....	9
<b>8. Conclusion</b> .....	9

## Introduction

This report provides an in-depth analysis of HR data focusing on employee demographics, diversity, salary distribution, and retention trends. The insights aim to support decision-making related to recruitment, employee development, diversity initiatives, and compensation strategies.

The report follows the Data Analysis Lifecycle, beginning from data collection, cleaning, and ending with actionable recommendations based on insights from data analysis.

## Data Analysis Lifecycle Stages

### 1. Business Understanding

The primary objective of this analysis was to gain insights into the following:

- The diversity and demographics of the workforce.
- Compensation patterns across various departments and job titles.
- Employee retention trends and areas with high turnover.
- The distribution of employees across different geographies.

The overall goal was to derive actionable recommendations for improving diversity, retention, compensation, and strategic hiring.

### 2. Data Understanding

#### 2.1 Data Sources

The data provided consisted of the following attributes:

- Employee ID: Unique identifier for each employee.
- First Name/Last Name: Employee's name.
- Gender: Gender of the employee.
- Ethnicity: Ethnic background of the employee.
- Age: Age of the employee.
- Job Title: Current role or position.

- Department: Department to which the employee belongs.
- Salary: Annual salary.
- Bonus %: Percentage of bonus relative to salary.
- Date of Hire: When the employee was hired.
- Country/City: Geographic location of the employee.
- Tenure: Number of years the employee has been with the company.

## 2.2 Key Metrics

- **Diversity Metrics:** Gender, ethnicity, and age distribution.
- **Compensation Metrics:** Salary and bonus percentages by job title and department.
- **Retention Metrics:** Hire and exit trends, turnover rates, and employee tenure.

## 3. Data Preparation (Cleaning)

The data cleaning process focused on ensuring consistency and accuracy in the dataset:

### 3.1 Data Cleaning Steps

#### 1. *Correcting Data Types:*

- Ensured that numeric fields like Salary, Bonus%, and Tenure were appropriately cast as numbers.
- Text fields such as Job Title, Department, and Country were treated as categorical data.
- Dates, such as Hire Date, were correctly formatted as date types.

#### 2. *Handling Missing Data:*

- Verified that all rows contained complete information for key metrics.
- Where possible, inferred missing values or flagged incomplete records for future data collection.

#### 3. *Removing Duplicates:*

- Ensured that there were no duplicate employee entries.

#### 4. *Standardizing Formats:*

- Standardized the format of Country and City names to maintain consistency across geographic data.

The data was now ready for analysis, with appropriate data types and cleaned fields.

## 4. Data Analysis

### 4.1 Descriptive Analysis

This phase focused on identifying general trends in the dataset using key metrics and visualizations.

#### 4.1.1 Demographics Analysis

- **Gender:** A balanced gender distribution, with women occupying more senior roles than men.
- **Ethnicity:** Diverse ethnic representation, though Black employees constituted a smaller portion (10%) of the workforce, particularly at senior levels.
- **Age:** Majority of employees were in their 40s, with fewer in their 20s and 60s.

#### 4.1.2 Geographical Distribution

- The majority of employees were based in Seattle, with Brazil (Manaus) being the least represented region.

#### 4.1.3 Job Title and Department

- IT and Engineering had the highest employee numbers, while Support Roles were underrepresented.

#### 4.1.4 Salary Distribution

- Salaries showed disparities across locations, with Seattle employees earning significantly higher than those in Manaus.

### 4.2 Diagnostic Analysis

This phase focused on identifying reasons behind key trends and patterns.

#### 4.2.1 Gender and Leadership

- **Observation:** Women were more prevalent in senior roles than men.
- **Diagnosis:** This may be a result of promotion policies favoring gender diversity at leadership levels.

#### 4.2.2 Ethnic Disparity in Leadership

- **Observation:** Black employees were less represented in senior positions.
- **Diagnosis:** This may be due to barriers in career progression for minority employees or inadequate diversity-driven leadership programs.

### 4.2.3 Salary Disparities by Location

- **Observation:** Salaries were significantly higher in Seattle than in Manaus.
- **Diagnosis:** This may be due to geographic cost-of-living differences or unequal compensation structures between regions.

## 5. Data Visualization

To enhance understanding and provide actionable insights, various visualizations were created using Power BI, supplemented with Python scripts where necessary, to display trends, relationships, and distributions across different HR aspects:

### 5.1 Gender and Ethnicity Breakdown

- **Pie/Donut Charts:** Displayed the *gender distribution* within the company, helping to visualize the proportion of male and female employees across departments.
- **Stacked Bar Chart:** Used to show the *ethnic distribution* across different business units, highlighting areas with greater or lesser diversity.

### 5.2 Salary and Bonus Distribution

- **Python Correlation Heatmap:** Displayed the *relationship between job title, department, salary, and bonus percentage*, allowing for insights into whether certain job titles or departments were associated with higher compensation packages.

```
# The following code to create a dataframe and remove duplicated rows is
always executed and acts as a preamble for your script:

# dataset = pandas.DataFrame(Bonus %, Annual Salary, Job Title, Department)
# dataset = dataset.drop_duplicates()

# Paste or type your script code here:
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Drop duplicate rows
dataset = dataset.drop_duplicates()

# Convert categorical columns into numerical categories
dataset['Job Title'] = pd.factorize(dataset['Job Title'])[0]
dataset['Department'] = pd.factorize(dataset['Department'])[0]

# Select relevant columns
df = dataset[['Job Title', 'Department', 'Annual Salary', 'Bonus %']]
```

```

# Calculate the correlation matrix
corr_matrix = df.corr()

# Plot the heatmap of the correlation matrix
plt.figure(figsize=(12, 10))

# Create the heatmap
ax = sns.heatmap(corr_matrix, annot=True, cmap='tab20b', fmt='.2f',
                  annot_kws={"size": 32}, # Font size for annotations
                  cbar_kws={"shrink": .8}) # Adjust color bar size

# Adjust the size of the color bar legend text
colorbar = ax.collections[0].colorbar
colorbar.ax.tick_params(labelsize=20) # Font size for color bar ticks
colorbar.ax.set_ylabel(colorbar.ax.get_ylabel(), fontsize=20) # Font size
for color bar label

# Title and axis labels with increased font size
plt.title("Correlation Between Job Title, Department, Salary, and Bonus %",
          fontsize=22)
plt.xlabel('Variables', fontsize=32)
plt.ylabel('Variables', fontsize=32)

# Adjust tick label font sizes
plt.xticks(fontsize=18)
plt.yticks(fontsize=18)

# Display the plot
plt.show()

```

- *Histogram*: Visualized the *distribution of annual salaries*, showing the frequency of different salary ranges within the company.

```

# The following code to create a dataframe and remove duplicated rows is
always executed and acts as a preamble for your script:

# dataset = pandas.DataFrame(Annual Salary)
# dataset = dataset.drop_duplicates()

# Paste or type your script code here:
import matplotlib.pyplot as plt
import numpy as np
x = np.random.normal(dataset["Annual Salary"])

```

```
plt.hist(x, color = "purple", edgecolor = "white")
plt.title("Distribution of Annual Salary", fontsize = 15)
plt.xlabel("Annual Salary", fontsize = 15)
plt.ylabel("Frequency", fontsize = 15)

plt.show()
```

### 5.3 Hiring and Exit Trends

- *Line Charts:* Depicted *hiring trends over time*, showing the number of employees hired during different time periods. Another line chart was created for *exit trends*, revealing which timeframes saw higher employee turnover.
- *Gauge Chart:* Calculated and displayed the *overall turnover rate*, providing a clear view of how many employees exited the company relative to the total number of employees.

### 5.4 Tenure and Retention Analysis

- *Bar Chart:* Showcased *average tenure* by department and job title, helping to identify which areas of the business had the longest or shortest employee tenures.
- *Stacked Column Chart:* Analyzed *tenure by demographics* (e.g., gender, age, ethnicity), revealing patterns in retention across different demographic groups. This provided insight into whether certain groups had a higher or lower average tenure.

### 5.5 Geographical Distribution of Employees

- *Scatter Map:* Displayed the *distribution of employees across countries and cities*, providing a clear view of employee concentrations and helping identify regions with the largest employee presence.
- *Tree Map:* Used to visualize employee counts by *city and country*, emphasizing areas with higher concentrations of staff.

Each of these visualizations played a key role in uncovering actionable insights related to diversity, retention, and compensation, providing the company with a data-driven foundation for optimizing HR practices.

## 6. Data Interpretation and Insights

Based on the analysis and visualizations, the following insights were derived:



## 6.1 Gender and Ethnic Diversity

- The company has a fairly balanced gender distribution, but there is room for improvement in promoting Black and other underrepresented ethnicities to leadership positions.

## 6.2 Salary Inequity

- There are noticeable discrepancies in salary distribution, especially across different geographical locations. Employees in Manaus are undercompensated compared to those in Seattle, despite similar roles.

## 6.3 Turnover and Retention

- Departments like Finance and Operations exhibit higher turnover rates, indicating dissatisfaction or leadership challenges in those areas.

## 6.4 Age and Tenure

- Employees in their 40s are most likely to stay longer with the company, while younger employees (20s) tend to have shorter tenures.

# 7. Recommendations

Based on the insights from the analysis, the following recommendations are proposed:

## 7.1 Strengthening Diversity and Inclusion

- *Mentorship Programs:* Develop mentorship initiatives for minority employees, especially Black and underrepresented groups, to foster career growth and enhance representation in leadership roles.
- *Diverse Recruitment:* Implement recruitment strategies aimed at increasing diversity across departments, ensuring a broader, more inclusive candidate pool.

## 7.2 Addressing Salary Disparities

- *Salary Audit:* Conduct a comprehensive salary audit to identify and address pay discrepancies across geographic locations. Ensure compensation aligns with market standards for each role, particularly in regions like Brazil where gaps may exist.

## 7.3 Enhancing Retention Strategies

- *Turnover Reduction:* Implement regular exit interviews in high-turnover departments such as Finance and Operations to understand employee concerns. Use this feedback to create targeted engagement and retention programs.

- *Career Development for Young Employees:* Develop structured career development programs for employees in their 20s, offering clearer career progression opportunities and professional growth paths to improve retention.

#### 7.4 Expanding Regional Workforce

- *Global Expansion:* Expand the company's workforce into underrepresented regions, such as Brazil and Southeast Asia, to diversify talent pools and reduce over-reliance on high-cost locations like Seattle. Promote global hiring to create a more regionally balanced workforce.

### 8. Conclusion

This report outlined the process of analyzing HR data based on the Data Analysis Lifecycle. After cleaning and preparing the data, the analysis provided valuable insights into the company's workforce demographics, salary structures, and retention trends. Based on these findings, targeted recommendations were made to improve diversity, address salary inequities, and reduce turnover in key departments. Implementing these changes can result in a more equitable, diverse, and high-performing organization.