**Employee Data Summary Report**

**1. Dataset Overview**

Total Employees: 1470

Columns: 23

**Data Columns and Missing Values:**

|  |  |  |
| --- | --- | --- |
| **COLUMN** | **NON-NULL COUNT** | **MISSING VALUES** |
| EmployeeID | 1470 | 0 |
| FirstName | 1470 | 0 |
| LastName | 1470 | 0 |
| Gender | 1469 | 1 |
| Age | 1470 | 0 |
| BusinessTravel | 1469 | 1 |
| Department | 1469 | 1 |
| DistanceFromHome (KM) | 1470 | 0 |
| State | 1468 | 2 |
| Ethnicity | 1467 | 3 |
| Education | 1470 | 0 |
| EducationField | 1469 | 1 |
| JobRole | 1470 | 0 |
| MaritalStatus | 1468 | 2 |
| Salary | 1470 | 0 |
| StockOptionLevel | 1468 | 2 |
| OverTime | 1470 | 0 |
| HireDate | 1470 | 0 |
| Attrition | 1470 | 0 |
| YearsAtCompany | 1470 | 0 |
| YearsInMostRecentRole | 1470 | 0 |
| YearsSinceLastPromotion | 1470 | 0 |
| YearsWithCurManager | 1470 | 0 |

**Data Types:**

**Integer Columns:** Age, DistanceFromHome (KM), Education, Salary, YearsAtCompany, YearsInMostRecentRole, YearsSinceLastPromotion, YearsWithCurrManager

**Float Column:** StockOptionLevel

**Object Columns:** EmployeeID, FirstName, LastName, Gender, BusinessTravel, Department, State, Ethnicity, EducationField, JobRole, MaritalStatus, OverTime, HireDate, Attrition

**2. Key Statistics**

**Average Age:** 28.99 years

**Average Salary:** $112,956.50

**Attrition Rate:** 16.12%

**3. Distribution Analysis**

**3.1 Age Distribution**

The age distribution reveals that the average age of employees is approximately 29 years. This suggests a predominantly younger workforce. It is useful to analyze the age distribution further to understand the spread and identify any potential age-related trends or issues.

**3.2 Salary Distribution**

The average salary stands at about $112,956.50. A detailed salary distribution analysis would include examining the range of salaries, identifying any outliers, and understanding the distribution pattern, which could help benchmark against industry standards.

**3.3 Years at Company Distribution**

Analyzing the distribution of years at the company helps to understand employee tenure. This includes examining the average number of years employees have spent with the company and the range and distribution to identify patterns in employee retention.

**4. Detailed Breakdown**

**4.1 Salary by Department**

A breakdown is essential to understand how salaries vary across different departments. This will involve calculating average salaries for each department and analyzing any significant differences. It can also highlight departments with higher compensation levels, which may be aligned with job roles and responsibilities.

**4.2 Years at Company by Job Role**

Analyzing years at the company based on job roles can reveal insights into employee retention within different roles. It helps to identify which roles have higher or lower average tenure and can provide valuable information for improving retention strategies.

**5. Attrition Analysis**

With an attrition rate of 16.12%, it's crucial to explore factors contributing to employee turnover. This includes:

**Departmental Analysis:** Examining which departments have higher attrition rates.

**Job Role Analysis:** Identifying job roles with higher turnover.

**Distance from Home and Other Factors:** Exploring if distance from home or other variables are correlated with higher attrition.

**6. Steps involved**

**1. Handle Missing Values:**

Impute missing values in categorical columns using the most frequent category or consider data removal where applicable.

**2. Visualizations:**

Create histograms and boxplots to visualize age and salary distributions.

Use bar charts to compare salaries by department and tenure by job role.

**3. Further Analysis:**

Conduct correlation analysis between salary, age, and tenure.

Analyze the impact of business travel and distance from home on attrition rates.

**4. Data Preprocessing:**

Convert HireDate to datetime format for time-based analysis.

Encode categorical variables for predictive modeling if needed.

**7. Next Steps**

**1. Data Cleaning:** Address missing values and ensure data consistency.

**2. Exploratory Data Analysis (EDA):** Generate detailed visualizations and summaries.

**3.Predictive Modeling:** If applicable, prepare data for modeling and conduct predictive analysis to understand factors influencing attrition or other outcomes.