

EXPLORATORY DATA ANALYSIS REPORT

1. Data Overview

- Dimension: The dataset contains 1,999 rows and 20 columns.
- Columns: Employee_ID, Department, Gender, Age, Job_Title, Hire_Date, Years_At_Company, Education_Level, Performance_Score, Monthly_Salary, Work_Hours_Per_Week, Projects_Handled, Overtime_Hours, Sick_Days, Remote_Work_Frequency, Team_Size, Training_Hours, Promotions, Employee_Satisfaction_Score, Resigned
- Data types: bool(1: Resigned), float64(2: Monthly_Salary, Employee_Satisfaction_Score), int64(12), object(5)

2. Summary Statistics

- Age: The average employee age is around 41 years, ranging from 22 to 60.
- Years At Company: The average number of years spent at the company is 4.5, with a maximum of 9 years.
- Monthly Salary: The mean salary is \$6394, with values ranging from \$3850 to \$9000.
- Performance Score: Performance scores range from 1 to 5, with a mean of 2.96.
- Work Hours Per Week: Employees work an average of 45 hours, with a standard deviation of about 9 hours.
- Department: 9 unique departments ('IT', 'Finance', 'Customer Support', 'Engineering', 'Marketing', 'HR', 'Operations', 'Sales', 'Legal')
- Gender: 3 gender categories (likely including Male, Female, and perhaps a third option).
- Job Title: 7 unique job title ('Specialist' 'Developer' 'Analyst' 'Manager' 'Technician' 'Engineer' 'Consultant').
- Hire Date: Over 1553 unique dates (indicating a wide hiring span).
- Education Level: 4 education levels ('High School' 'Bachelor' 'Master' 'PhD').
- Resigned: Boolean (True/False) for whether the employee has resigned.

3. Visualization

- Monthly Salary: Salaries are relatively normally distributed, with a peak around \$6500.
- Years at Company: The majority of employees have been with the company for 6 to 10 years.

DATA QUALITY ANALYSIS REPORT

1. Missing Values: There are no missing values in the dataset, as indicated by the "Missing Data Report" with 0 missing values across all columns. The only missing values were detected in the new column we created to assign the different range of years in the company to each employee: from 0 to 2, from 3 to 5...
2. Duplicate Rows: There are 0 duplicate rows, meaning all rows in the dataset are unique.
3. Outlier Handling: No extreme outliers were detected.

Building Business Questions

Potential business questions based on the provided data description:

1. What is the average employee satisfaction score (0-5) by department?
2. Which departments have the highest number of employee resignations?
3. What is the average number of projects handled by employees in each department?
4. How does remote work frequency affect employee performance scores (1-5)?
5. How do training hours impact employee satisfaction and performance?
6. Does working a lot of overtime (>5 hours) make it more likely that an employee will get a promotion?
7. What is the age distribution of employees across different job titles?
8. Which job titles have the highest salaries and number of promotions