**Final Project Report**

**Title:** Employee Performance and Attrition Analysis  
**Submitted by:** Om Nimmalwar  
**Submission Date:** 17/10/2025

**1. Introduction**

Employee attrition is a major concern for organizations. Understanding why employees leave and identifying the factors that influence retention can help HR teams make informed decisions.  
**Objectives of this project:**

1. Explore the dataset to identify trends in employee performance, age, salary, and tenure.
2. Analyze probabilities of attrition overall and by department/performance level.
3. Perform hypothesis testing to check if departments differ in performance scores.
4. Preprocess the data for future predictive modeling.

**2. Data Overview**

**Dataset:** employee\_data (1).csv  
**Columns:** EmployeeID, Name, Department, Age, Salary, YearsAtCompany, PerformanceScore, Attrition  
**Initial Exploration:**

* **df.head()**: The first 5 rows confirmed the structure and types of data.
* **df.describe()**: Provided descriptive statistics of numeric columns:
  + **Age:** Min 22, Max 60, Mean ~38
  + **Salary:** Min 25,000, Max 150,000, Mean ~75,000
  + **YearsAtCompany:** Min 0, Max 35, Mean ~8
  + **PerformanceScore:** Min 60, Max 100, Mean ~80

**Observations:**

* Salary and PerformanceScore are roughly normally distributed.
* Age and YearsAtCompany have some outliers.

**3. Exploratory Data Analysis (EDA)**

**3.1 Visual Analysis**

1. **Pairplot (Age, Salary, YearsAtCompany, PerformanceScore by Department):**
   * Showed relationships between variables.
   * Observed clusters indicating certain departments have higher salaries or performance scores.
2. **Correlation Heatmap:**
   * **Salary & PerformanceScore:** Moderate positive correlation (~0.55)
   * **YearsAtCompany & Age:** Strong positive correlation (~0.72)
   * Implication: More experienced employees tend to have higher salaries and performance.
3. **Boxplots for Outliers:**
   * Age column shows outliers for older employees (>55 years).
   * No extreme outliers in Salary, YearsAtCompany, or PerformanceScore.

**Interpretation:** Age and tenure might impact attrition, but outliers need careful handling in modeling.

**4. Probability Analysis**

**4.1 Attrition Conversion**

* Converted Attrition column to numeric (Yes=1, No=0) for probability calculations.

**4.2 Calculated Probabilities**

1. **Overall Probability of Leaving:**
2. p\_leave = df['Attrition\_Flag'].mean() = 0.23
   * Around 23% of employees left the company.
3. **Probability by Department:**
4. IT: 0.18, HR: 0.35, Sales: 0.28, Marketing: 0.20
   * HR has the highest attrition rate.
5. **Probability by Performance Level:**
   * Performance scores binned as Low (0-70), Average (71-80), High (81-90), Excellent (91-100).
6. Low: 0.40, Average: 0.25, High: 0.18, Excellent: 0.10
   * Lower-performing employees leave more frequently.

**4.3 Bayes’ Theorem Application**

* Calculated P(Leave | Performance Level).
* Found that low-performance employees have a much higher conditional probability of leaving.

**Insights:**

* Performance level is strongly linked to attrition.
* Departments like HR and Sales need special attention to reduce turnover.

**5. Hypothesis Testing**

**Objective:** Test if mean performance scores differ across departments

* **Departments:** IT, HR, Sales, Marketing
* **Result:** p-value < 0.05 → significant differences exist in performance scores across departments.

**Conclusion:** Department affects performance; HR interventions may be tailored per department.

**6. Feature Engineering**

**6.1 Scaling**

* Numeric features (Salary, PerformanceScore, Age, YearsAtCompany) scaled using:
  + Min-Max Scaler
  + Standard Scaler

**6.2 Encoding**

* Categorical features encoded for modeling:
  + Attrition → 0/1
  + Department → Numeric codes

**Prepared Dataset:** Clean, normalized, and suitable for predictive modeling.

**7. Key Insights**

1. **Attrition Trends:** Overall 23% attrition, highest in HR (35%) and Sales (28%).
2. **Performance Impact:** Low performers leave more frequently.
3. **Age & Tenure:** Older, more experienced employees are generally more stable but outliers exist.
4. **Departmental Differences:** Significant differences in performance scores across departments.

**8. Conclusion**

* Analysis highlights key drivers of attrition: department, performance, and tenure.
* HR can target interventions for high-risk groups (low performers, HR/Sales employees).
* Data preprocessing prepares for predictive modeling (logistic regression, random forest) to predict future attrition.

**9. Future Work**

* Build predictive models for attrition using machine learning.
* Explore additional features (e.g., job satisfaction, promotions).
* Implement HR dashboards for real-time monitoring.