**REPORT ON EMPLOYEE ATTRITION ANALYSIS**

**Introduction**

The primary objective of this project was to explore and analyze employee attrition data, identify key factors contributing to attrition, and develop predictive models to better understand the causes behind employee turnover. The dataset provided a variety of employee-related features such as demographic information, job role, satisfaction, income, and many other variables. This report will provide a detailed overview of the data analysis process, key findings, statistical tests, and the predictive model developed.

**Data Overview**

The dataset comprises 1,470 employees with 35 features, including both categorical and continuous variables. The key attributes in the dataset are:

* **Demographics**: Age, Gender, Marital Status, and Education.
* **Job-related variables**: Job Role, Department, Job Level, Monthly Income, Years at Company, Job Satisfaction, etc.
* **Performance indicators**: Attrition, OverTime, Environment Satisfaction, and Job Satisfaction.

The analysis aimed to understand what factors influence employee attrition and to predict whether an employee is likely to leave the company.

**Step 1: Data Cleaning and Preprocessing**

To begin the analysis, the dataset was preprocessed by handling missing data, converting categorical variables to factors, and ensuring data integrity:

* **Handling Missing Data**: There were no missing values found in the dataset.
* **Categorical Variables**: Several variables like Attrition, Gender, Department, and JobRole were converted to factor variables for proper analysis and modeling.

**Step 2: Exploratory Data Analysis (EDA)**

The exploratory data analysis (EDA) provided valuable insights into the dataset’s characteristics.

* **Age Distribution**: The age distribution showed that most employees were between the ages of 30 and 50, with a few younger and older employees. A histogram was plotted to visualize the distribution.
* **Attrition vs. Monthly Income**: A boxplot revealed that employees who left the company (Attrition = "Yes") generally had lower monthly incomes compared to those who stayed. This indicates a potential correlation between low income and higher attrition rates.
* **Attrition by Job Role**: A bar plot revealed that attrition rates varied by job role, with higher turnover in roles such as "Laboratory Technician" and "Sales Executive."
* **Attrition by Department**: A similar analysis showed that attrition rates were higher in the "Sales" department compared to "Research & Development."
* **Job Satisfaction and Attrition**: A bar plot showed that employees with lower job satisfaction were more likely to leave the company.

**Step 3: Statistical Testing**

To gain a deeper understanding of the factors influencing attrition, several statistical tests were conducted:

1. **Chi-Square Test for Job Role and Attrition**:  
   A chi-square test was performed to examine the relationship between JobRole and Attrition. The results showed a significant association (p-value = 2.75e-15), suggesting that certain job roles are more associated with higher attrition rates.
2. **T-Test for Monthly Income and Attrition**:  
   A t-test was used to compare the average monthly income between employees who stayed (Attrition = "No") and those who left (Attrition = "Yes"). The results (p-value = 4.43e-13) confirmed that employees with lower monthly incomes were significantly more likely to leave.

**Step 4: Correlation Analysis**

A correlation matrix was computed between continuous variables such as Age, MonthlyIncome, JobLevel, NumCompaniesWorked, and TotalWorkingYears. The results highlighted moderate correlations between MonthlyIncome and JobLevel, suggesting that higher job levels are associated with higher income. However, other correlations were relatively weak.

**Step 5: Predictive Modeling**

A logistic regression model was developed to predict the likelihood of an employee leaving the company (Attrition). The model considered the following predictor variables:

* **Age**
* **MonthlyIncome**
* **JobRole**
* **Department**
* **MaritalStatus**

The logistic regression model was trained on 80% of the data, and its summary revealed the following:

* **Age**: There is a significant negative relationship between age and attrition (p-value = 0.00112), suggesting that younger employees are more likely to leave.
* **MonthlyIncome**: Although monthly income was not significant (p-value = 0.47422), it is still important for further model improvement.
* **Job Role**: Some job roles, particularly "Laboratory Technician" and "Sales Executive", have significant relationships with attrition, indicating that employees in these roles are more likely to leave.
* **Department**: Employees in the "Sales" and "Research & Development" departments showed similar attrition patterns, with neither department standing out as significantly different from the other.
* **Marital Status**: There was no significant difference in attrition rates based on marital status.

**Step 6: Model Evaluation**

The logistic regression model's performance was evaluated using accuracy, precision, recall, and the area under the ROC curve (AUC). The model was able to identify the most critical features that drive attrition, but further tuning of the model (e.g., using different machine learning algorithms or adding more features) could improve predictive accuracy.

**Key Findings**

* **Income and Attrition**: Employees with lower incomes tend to have higher attrition rates.
* **Job Roles with High Attrition**: Employees in "Laboratory Technician" and "Sales Executive" roles are more likely to leave.
* **Younger Employees**: Younger employees (below 30) have higher attrition rates compared to older employees.
* **Departmental Differences**: There are no significant differences in attrition rates between departments, but the "Sales" department exhibited higher attrition compared to others.

**Recommendations**

Based on the analysis, the following recommendations can help reduce employee attrition:

1. **Increase Salary for High-Risk Employees**: Focus on improving the compensation for employees in lower-income brackets to reduce attrition.
2. **Focus on Job Satisfaction**: Conduct surveys and improve engagement initiatives to raise job satisfaction, particularly in roles with high turnover.
3. **Tailor Retention Strategies by Job Role**: Employees in roles such as "Laboratory Technician" and "Sales Executive" should be specifically targeted with retention strategies.
4. **Invest in Employee Development**: Employees in lower job levels (entry-level) should be given clear career development paths and growth opportunities.