**EDA Report: Employee Attrition Prediction**

**1. Dataset Overview:**

Number of records: **59130** rows  
Number of features: **24** columns

Target Variable: Attrition (Binary: **Stayed (1)** / **Left (0)**)

Objective: Predict whether an employee will leave the company based on personal, professional, and organizational features.

**2. Data Types:**

* **Numerical Features:**  
  Employee ID, Age, Years at Company, Monthly Income, Number of Promotions, Distance from Home, Number of Dependents, Company Tenure.
* **Categorical Features:**  
  Gender, Job Role, Work-Life Balance, Job Satisfaction, Performance Rating, Overtime, Education Level, Marital Status, Job Level, Company Size, Remote Work, Leadership Opportunities, Innovation Opportunities, Company Reputation, Employee Recognition, Attrition.

**3. Missing Values:**

Insight: No missing values detected.

**4. Statistical Summary:**

Insight:

* Monthly income ranges from **1316.0** to **16149.0** with a median of **7350.5**
* Number of promotions is **left-skewed**
* Distance from home **varies significantly** between employees

**5. Target Variable Distribution (Attrition):**

Insight:

* About **47.55%** of employees left the company (Attrition = Left)
* Class imbalance may require resampling later (SMOTE)

**7. Feature Engineering Performed:**

* Salary\_per\_Performance = Monthly Income / Performance Rating
* Tenure\_Group = binned version of Years at Company
* Distance\_Overtime = Distance × Overtime
* Performance\_Satisfaction = Job Satisfaction × Performance
* Encoded categorical features using:
  + Label Encoding

**8. Correlation Analysis:**

Created a heatmap to visualize correlations between features:

* Job Level, Marital Status and Remote Work show strong correlation with Attrition

**9. Recommendations:**

* Handle class imbalance during modeling
* Remove/reduce highly correlated features if using linear models
* Try models like Random Forest or Gradient Boosting
* Consider SMOTE or weighted loss functions
* Validate with stratified cross-validation