EMPLOYEE ATTRITION ANALYSIS REPORT

1. OVERVIEW OF DATASET

The HR dataset contains 35 features and 1470 records. It includes employee demographic information, work experience, job role, salary, and whether the employee has left the company (Attrition).
Key variables: Age, Department, Monthly Income, Job Role, Overtime, Years at Company, etc.

2. KEY STATISTICS AND VISUAL INSIGHTS

- Employees aged 30–40 had the highest attrition.
- Departments with more attrition: Sales and R&D.
- High attrition seen among employees with overtime and low income.
- Attrition is higher for employees with less than 3 years at the company.

3. ATTRITION CORRELATIONS

Attrition shows significant correlation with:

- Overtime (strong positive)
- Years at Company (negative)
- Monthly Income (negative)
- Job Satisfaction (negative)

These were validated through a heatmap and pair-plots.

4. MACHINE LEARNING PIPELINE

A Random Forest Classifier was trained using:

- Label Encoding for categorical variables
- SMOTE for handling class imbalance

- Feature scaling via Standard-Scaler
- 5-fold Cross-validation for model reliability

---

5. MODEL PERFORMANCE

- **Accuracy:** ~86%
- **ROC AUC Score:** ~77%
- The classification report indicates strong precision and recall for both classes (Attrition = Yes/No).

---

6. SHAP INTERPRETABILITY SUMMARY

SHAP summary plot identified:

- Overtime, Age, Monthly Income, and Job Level as key drivers of attrition.
- Employees with high Overtime and low income are at highest risk.
- SHAP visualizations reinforced the importance of reducing workload and increasing engagement.

---

7. BUSINESS RECOMMENDATIONS

- Monitor and reduce excessive overtime.
- Improve compensation packages for junior roles.
- Focus retention efforts on employees within the first 3 years.
- Enhance job satisfaction through internal feedback and growth opportunities.
