Atharva More

HR Analytics – Predicting Employee Attrition

1. Introduction

Employee attrition is a significant challenge for companies. It impacts productivity, morale, and hiring costs. The goal of this project is to identify the key factors that influence employee attrition using HR data and build a machine learning model to predict whether an employee is at risk of leaving.

2. Abstract

I used ***IBM's HR Analytics dataset*** from Kaggle and performed a complete data science pipeline—from data cleaning and exploratory data analysis (EDA), to predictive modelling and SHAP-based explainability. I built a logistic regression model with 84% accuracy and visualized our insights through Power BI. SHAP analysis provided transparency into which features influenced the model's predictions the most. This report highlights both technical steps and business recommendations to prevent attrition.

Kaggle Dataset Link:
https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset

3. Tools & Technologies Used

- Python (Pandas, Seaborn, Scikit-learn, SHAP)
- Google Colab (for model development)
- Power BI (for interactive dashboard)
- Excel / CSV (for dataset management)

4. Methodology

a. Data Preparation

- Removed non-informative columns (e.g., EmployeeCount, Over18)
- Converted categorical features using one-hot encoding
- Mapped target variable Attrition from Yes/No → 1/0

b. Exploratory Data Analysis (EDA)

- Found high attrition in employees under 30 years old
- Noted strong attrition correlation with overtime and low salary
- Visualized trends by age group, department, and income

c. Model Building

- Used Logistic Regression as a baseline classifier
- Achieved ~84% accuracy
- Evaluated using confusion matrix and classification report

d. SHAP Analysis

- Used SHAP values to interpret model predictions
- Found OverTime, MonthlyIncome, Age, and DistanceFromHome as top influencing factors

e. Power BI Dashboard

- Built visualizations for attrition trends by department, age, income, and feature importance
- Added interactive slicers for filtering by gender, job role, etc.

---

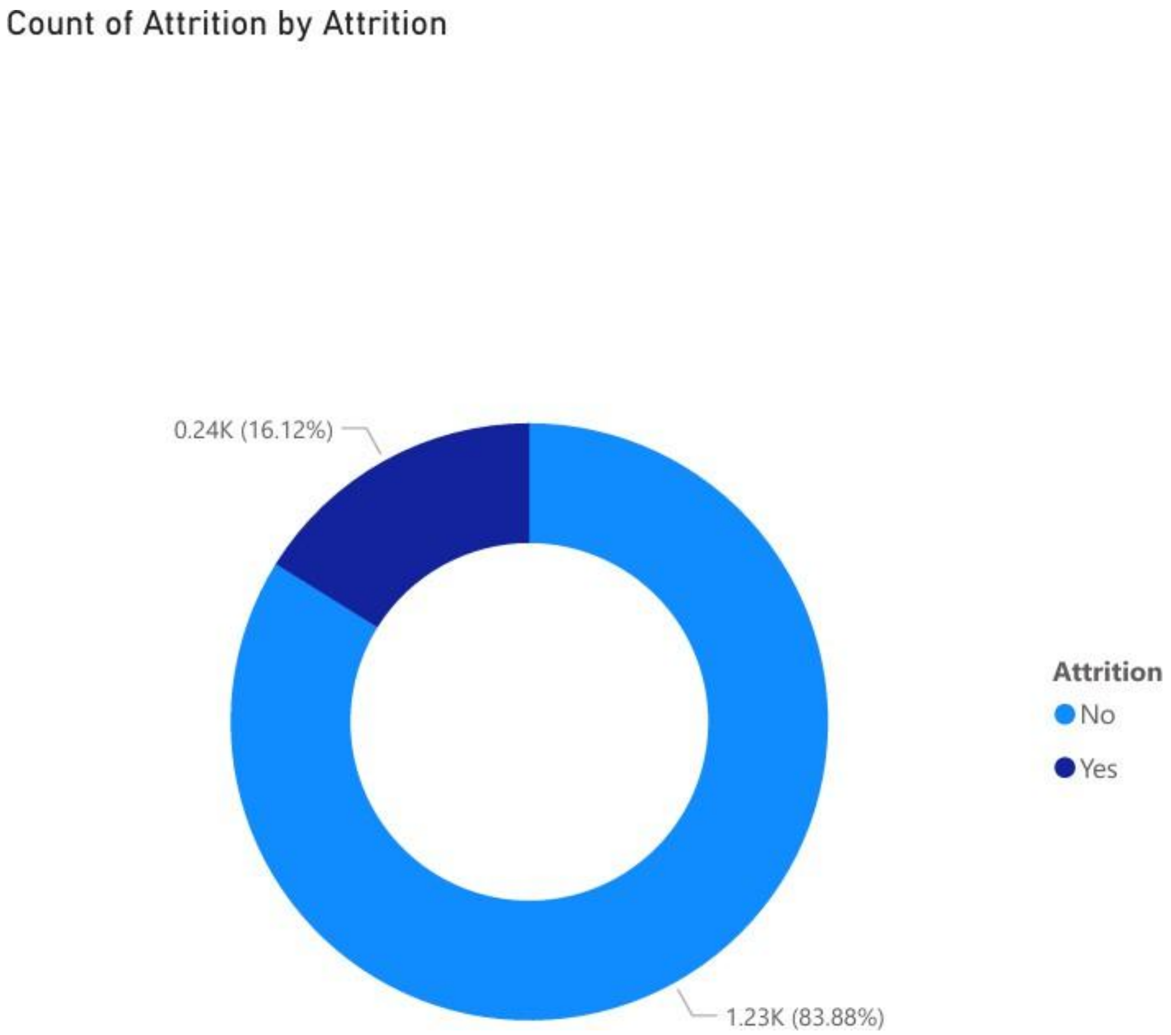
5. Key Findings

- Employees aged 18–30 had the highest attrition rate.
- OverTime was the most critical factor influencing attrition.
- Employees with lower monthly income showed higher resignation rates.
- Departments like Sales and Research & Development experienced more turnover.
- SHAP confirmed that the model made meaningful predictions based on real-world drivers.

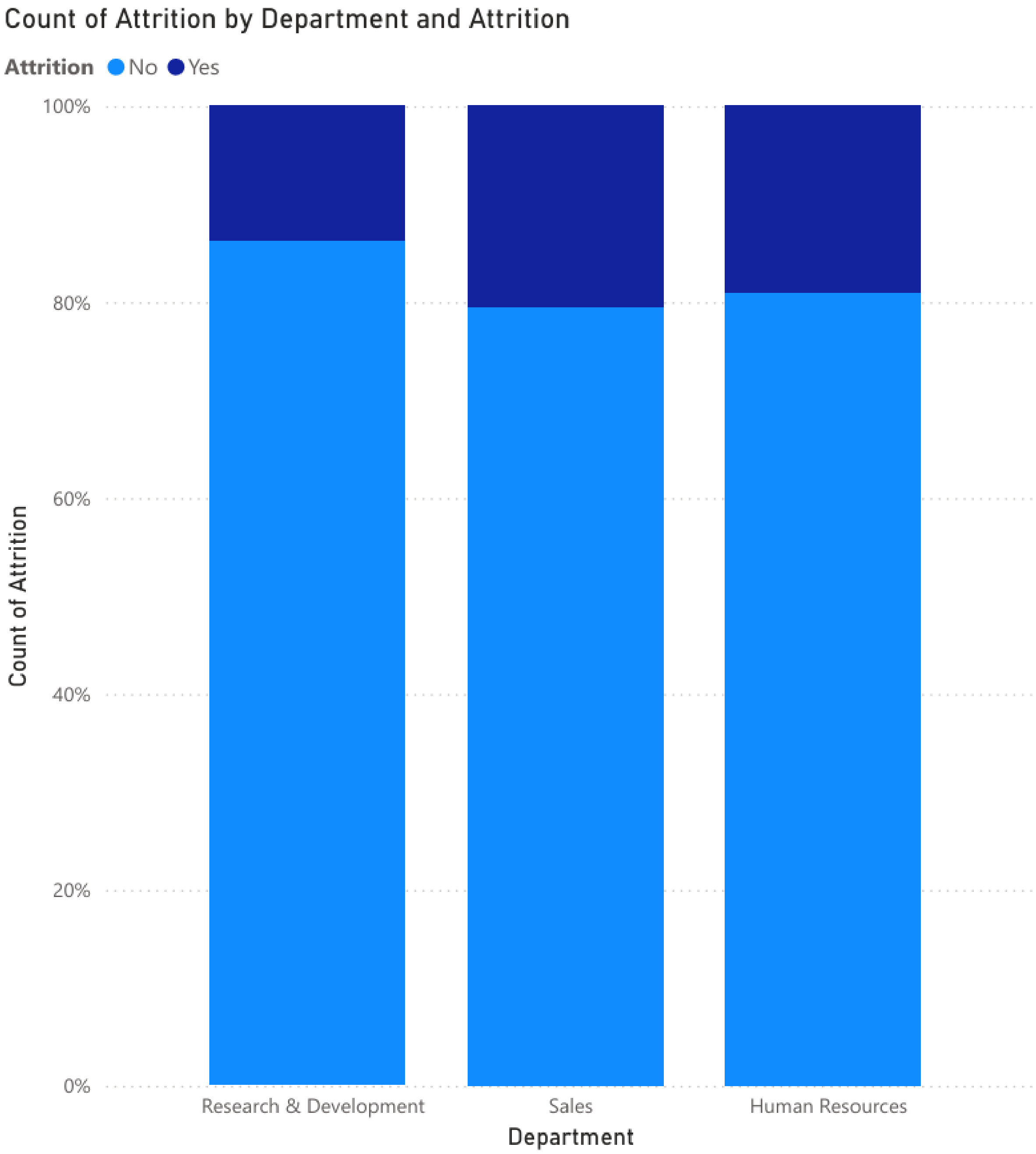
---

6. Visuals from Power BI dashboard

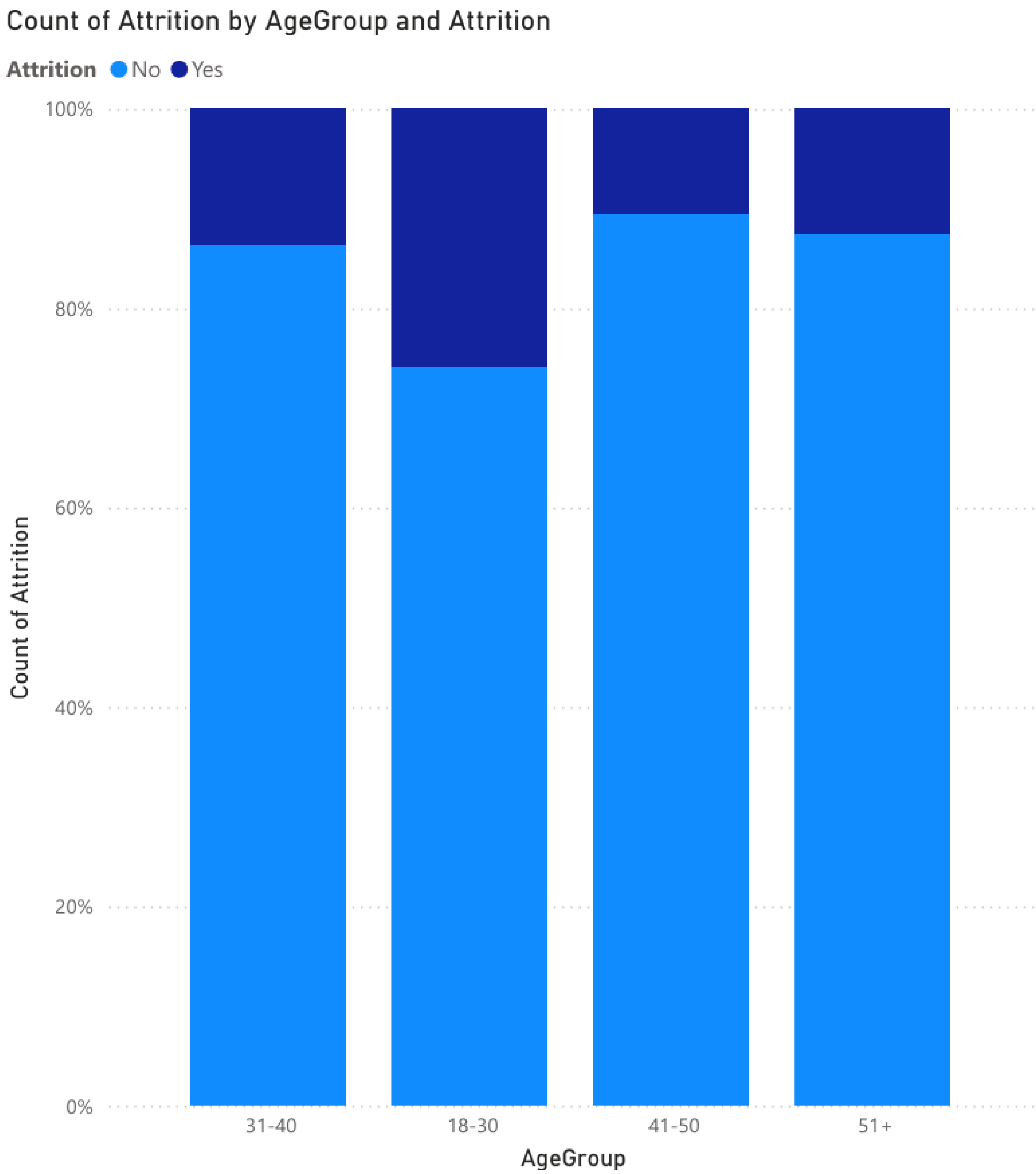
- **Attrition Count (Donut chart)**



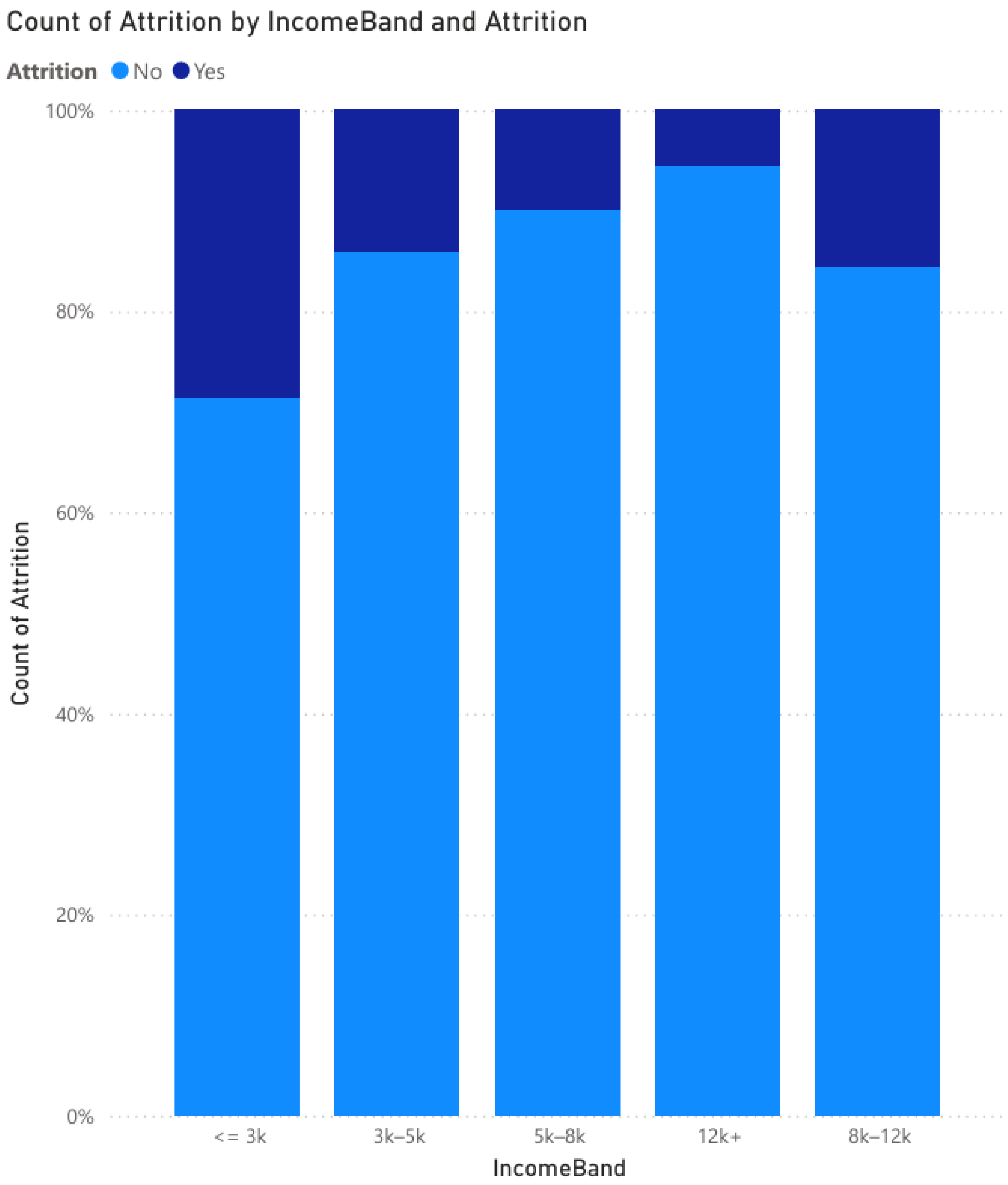
- **Attrition by Department**



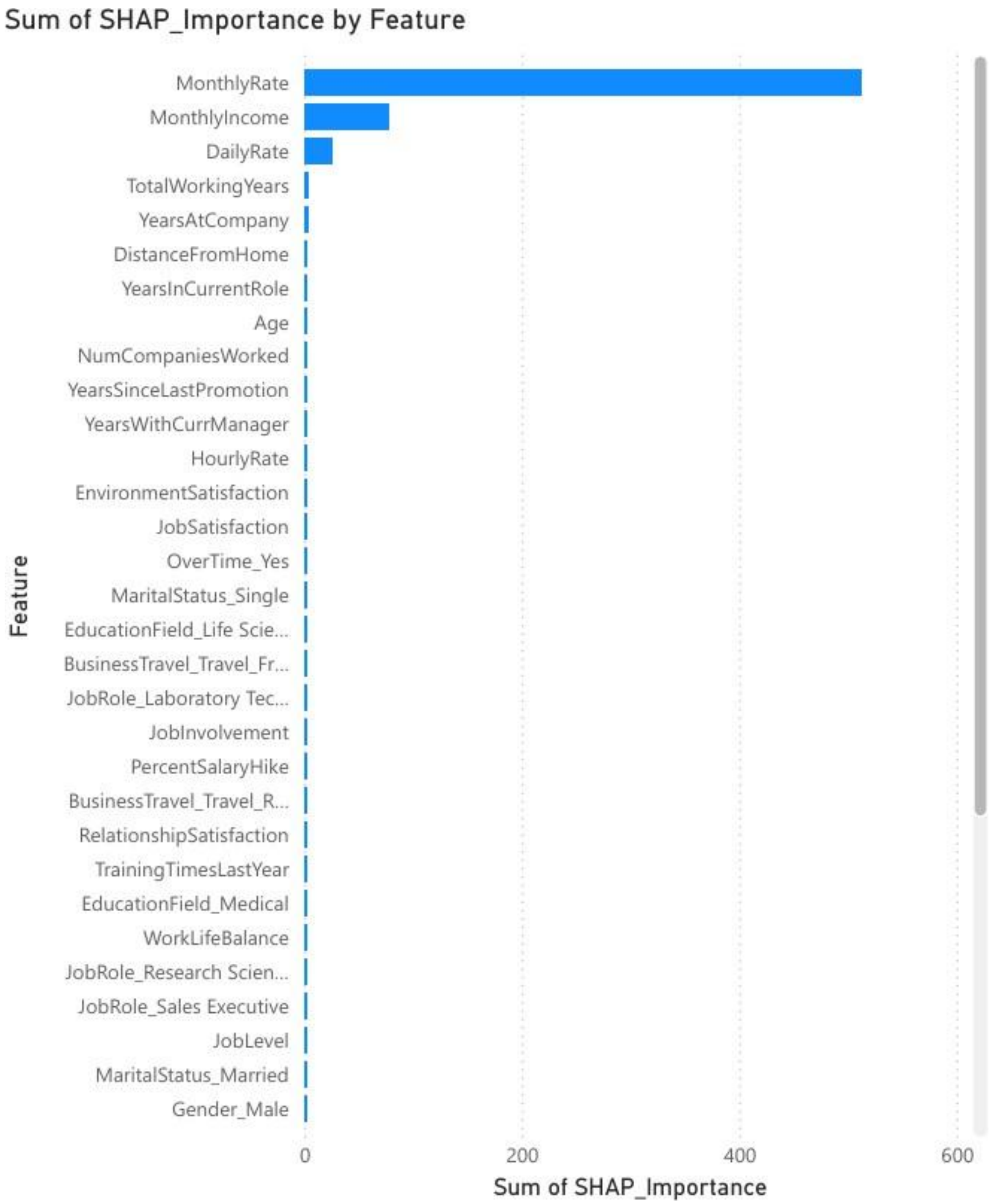
- **Attrition by Age Group**



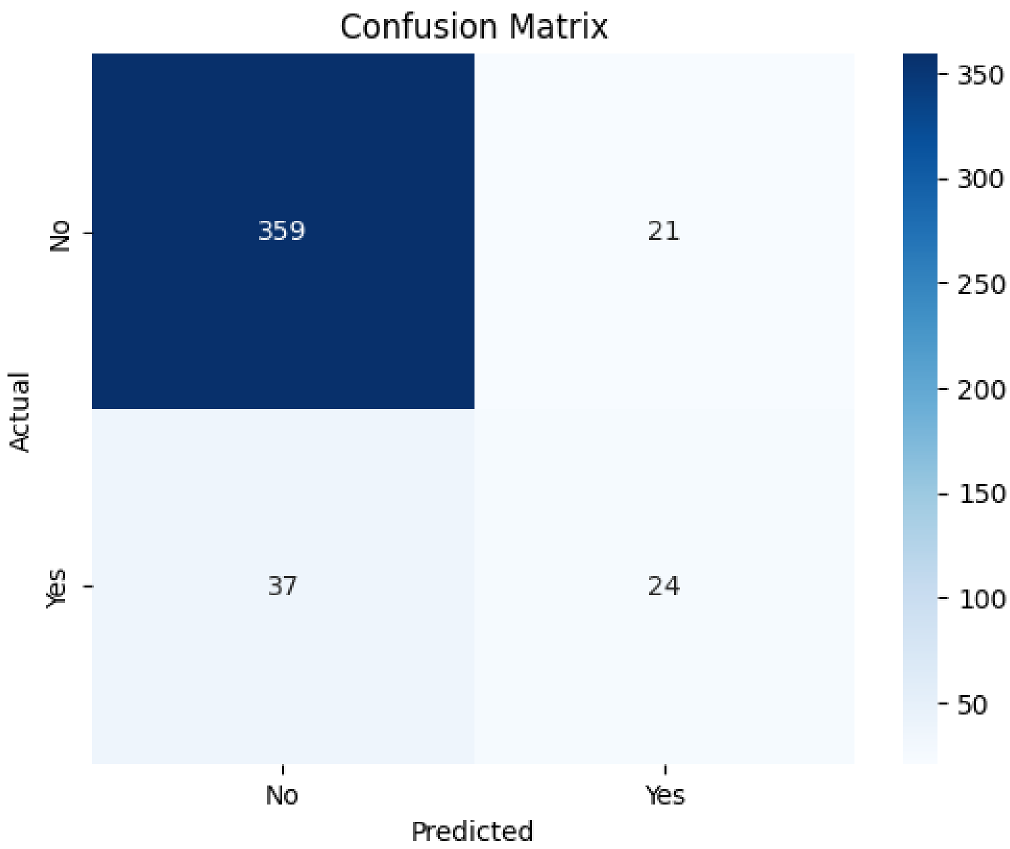
- **Monthly Income vs Attrition**



- **SHAP Feature Importance**



- **Confusion Matrix**



7. Conclusion

This HR analytics project delivered a predictive model with strong accuracy and explainability. By identifying key attrition triggers, HR teams can now intervene proactively. The combination of machine learning and visual dashboards enables both data-driven decisions and transparent insights into employee behaviour.