Internship Project Report

Project Title: HR Analytics – Predict Employee Attrition

Intern Name: *BHAVESH SANTOSHKUMAR TUPE*

Duration: 2 Weeks

---

1. Introduction

Employee attrition is a critical concern for many organizations, as it affects productivity, team morale, and operational costs. This project investigates the key drivers of employee resignation and builds a predictive model to forecast attrition using HR data from IBM.

---

2. Abstract

The objective of this project is to use Python and Power BI for exploratory analysis and machine learning modeling on HR attrition data. The project highlights critical factors behind employee resignation, builds a logistic regression model to predict future leavers, and visualizes attrition trends through interactive dashboards.

---

3. Tools & Technologies Used

- **Python**: pandas, seaborn, matplotlib, scikit-learn
- **Power BI**: Data visualization and dashboards
- **Google Colab / Jupyter Notebook**: Development environment
- **DAX**: Measures and KPI creation in Power BI

---

4. Steps Involved

Data Cleaning (Python)

- Removed constant or irrelevant columns (EmployeeCount, Over18, StandardHours)
- Handled categorical encoding and binary mapping (e.g., Attrition, OverTime)
- Created AgeGroup bins for segmentation analysis

Exploratory Data Analysis (Python)

- Analyzed attrition by department, gender, age group, and overtime status
- Found most attrition in **Sales**, **OverTime workers**, and **26–35 age group**

Model Building (Python)

- Model used: **Logistic Regression** with class_weight='balanced'

- Applied StandardScaler for feature scaling
- Model trained and evaluated on a stratified 80-20 train-test split

Visualization (Power BI)

- Built KPIs: Total Employees, Attrition Rate, Avg Income of Leavers
- Used bar, pie, treemap, and column charts to show attrition breakdown
- Added slicers for interactive filtering by age, department, gender, etc.

---

5. Model Results & Performance

Metric	Value
Accuracy	**75.17%**
Precision (Class 1)	0.35
Recall (Class 1)	0.62
F1 Score (Class 1)	0.44
Overall Attrition Rate	16.1%
AUC Score	~0.74

Confusion Matrix:

[[192 55]

 [18 29]]

📌 Despite lower precision, the model prioritizes **recall** for resigning employees — which is ideal for HR intervention strategies.

---

6. Conclusion

This project effectively combines machine learning and visualization to uncover attrition patterns and predict resignations. Power BI offers clear insight into key HR indicators, while the logistic regression model provides actionable intelligence for retention planning.