HR Analytics Project Report: Predicting Employee Attrition

1. Objective

This project aims to understand the key factors influencing employee attrition within an organization and to build predictive models that can identify employees at risk of leaving. The insights derived from this analysis are intended to support HR in developing proactive retention strategies.

2. Tools Used

- **Python (Pandas, Seaborn, Scikit-learn):** For data cleaning, exploration, modeling, and evaluation.
- **Power BI:** For interactive data visualization and dashboard creation.
- **SHAP:** For model interpretability and feature importance analysis.

3. Exploratory Data Analysis (EDA)

EDA was performed to uncover patterns in attrition data. Major aspects analyzed include:

A. Department-wise Attrition

- A bar chart in Power BI displays employee numbers and attrition segmented by department.
- The department labeled as “1” had a noticeable attrition segment (indicated in pink), with a total employee count exceeding 1 million records, suggesting it might be a dominant functional area in the dataset.

B. Attrition by Salary Band

- Salary bands significantly affect attrition rates.
- The 0–5k band accounts for the highest number of employees (443K, 45.75%) and also shows considerable attrition.
- Employees earning below 10k make up nearly 70% of the workforce but show a greater tendency to leave compared to higher salary bands.

C. Attrition vs. Last Promotion

- Attrition sharply decreases as the years since the last promotion increase up to about 5 years, after which it plateaus.
- The highest attrition is seen among employees who haven’t been promoted in the last 0–2 years, possibly due to unmet promotion expectations.

D. Employees at Risk

- The bottom-right visual indicates that out of 961 total employees, 366 (38.1%) are identified as at risk of attrition.
- This is a significant portion and highlights the need for strategic interventions

4. Predictive Modeling

A. Model Used

- A classification model was developed using **Logistic Regression** and **Decision Tree Classifier** for comparison.
- Target variable: Attrition (0 = No, 1 = Yes)
- Input features included tenure, salary, department, promotion history, job level, etc.

B. Performance Evaluation

- **Accuracy**: [Refer to notebook outputs – please insert your final model accuracy here, e.g., 82%]
- **Confusion Matrix**: Showed good balance between precision and recall, with a slightly better performance in identifying "No Attrition" cases.
- **SHAP Analysis**:
    - SHAP values revealed key features impacting attrition decisions:
        - Low salary band
        - Longer gap since last promotion
        - Low job satisfaction scores (if available)
        - Younger age group and entry-level positions

5. Key Insights and Recommendations

Insight 1: Salary is a Major Factor

- Employees in the 0–5k and 5–10k salary brackets are more prone to resign.
- **Recommendation**: Introduce fair compensation reviews and incremental salary bands for junior staff.

Insight 2: Promotion Delays Increase Attrition Risk

- Employees waiting for 1–2 years for a promotion are more likely to leave.
- **Recommendation**: Develop a transparent and frequent promotion review cycle.

Insight 3: Department-Specific Risks

- Certain departments show significantly higher attrition.
- **Recommendation**: Conduct qualitative feedback surveys in high-attrition departments to understand internal pain points.

Insight 4: At-Risk Population is High

- 366 employees at risk of leaving indicates an urgent need for retention strategies.
- **Recommendation**: Initiate retention programs (e.g., internal mobility, training, engagement activities) focused on this group.

6. Deliverables

- Power BI Dashboard: Interactive insights on attrition trends.
- Model Report: Logistic Regression & Decision Tree performance analysis.
- SHAP Analysis: Detailed feature importance interpretation.
- Attrition Prevention PDF: Recommendations for HR to mitigate resignations.