Project Flow Summary

Project Title: HR Analytics – Employee Attrition Prediction

Objective:

To analyze employee data and predict attrition using machine learning, providing insights into factors influencing employee turnover.

Workflow:

1. **Data Collection:** IBM HR Attrition dataset loaded into Kaggle Notebook.

2. **Data Exploration & Cleaning:** Checked data structure, handled IDs/constants, converted target variable.

3. **EDA (Exploratory Data Analysis):** Visualized attrition by department, age, job role, overtime, and income using Seaborn/Matplotlib.

4. **Feature Engineering:** Created Attrition_Flag (Yes=1, No=0) and performed one-hot encoding for categorical variables.

5. **Model Building:** Logistic Regression model trained using Scikit-learn with balanced class weights.

6. **Model Evaluation:** Achieved 71.7% accuracy, 61.7% recall, ROC-AUC of 0.78. Generated confusion matrix and classification report.

7. **Feature Importance:** Identified key drivers such as Overtime, Marital Status, Environment Satisfaction, Business Travel, Stock Options, and Job Role.

8. **Result Export:** Scored dataset prepared with predictions and probabilities, ready for visualization.

Tools & Libraries Used:

- Kaggle Notebook (development platform)

- Python (programming language)

- Pandas, NumPy (data manipulation)

- Matplotlib, Seaborn (visualization)

- Scikit-learn (machine learning)

Status: Project 1 successfully completed ■