Summary of all that I did in this project

Python

1. **Library Setup**:
    - Imported necessary libraries: pandas, numpy, matplotlib, seaborn, sklearn, and shap.
2. **Data Preprocessing**:
    - Loaded dataset (Dataset.csv) using pandas.
    - Checked for and removed any missing values.
    - Dropped irrelevant columns: 'EmployeeCount', 'Over18', 'StandardHours', 'EmployeeNumber'.
    - Converted the target column Attrition to binary (Yes = 1, No = 0).
    - Encoded categorical columns using LabelEncoder.
3. **Exploratory Data Analysis (EDA)**:
    - Plotted:
        - Countplot of **Attrition by Department**.
        - Boxplot of **MonthlyIncome vs Attrition**.
        - Histogram of **YearsSinceLastPromotion vs Attrition**.
4. **Modeling**:
    - Split data into training and test sets (80/20).
    - Trained a **Decision Tree Classifier**.

Power BI

1. **Visuals Created**:
    - **Stacked Column Chart**: Count of Attrition by **JobSatisfaction** and **Department**.
    - **Stacked Column Chart**: Count of EmployeeCount by **YearsSinceLastPromotion** and **Attrition**.
    - **Heatmap**:
        - Based on **JobRole** and shows **Count of Attrition**.
2. **Slicers Used**:
    - Gender, OverTime, and Department slicers for filtering the visuals.

In the Python portion of the project, the necessary libraries such as pandas, numpy, matplotlib, seaborn, sklearn, and shap were imported. The dataset, named "Dataset.csv," was loaded using pandas, after which any missing values were identified and removed. Irrelevant columns including 'EmployeeCount', 'Over18', 'StandardHours', and 'EmployeeNumber' were dropped to focus the analysis on meaningful features. The target variable, Attrition, was converted into a binary format where 'Yes' was encoded as 1 and 'No' as 0. Additionally, all categorical columns were encoded using LabelEncoder to transform them into numerical format suitable for machine learning algorithms.

For exploratory data analysis (EDA), several visualizations were created to better understand the relationships in the data. A countplot was generated to display Attrition distribution across different Departments, helping identify which departments experienced higher attrition rates. A boxplot was plotted to examine the distribution of MonthlyIncome with respect to Attrition, providing insights into income differences between employees who stayed versus those who left. Lastly, a histogram was used to visualize the YearsSinceLastPromotion variable segmented by Attrition status, revealing trends related to promotion timing and employee turnover.In the modeling phase, the dataset was split into training and testing subsets in an 80/20 ratio to build and evaluate the predictive model. A Decision Tree Classifier was trained on the training data to predict employee attrition.

For the Power BI protion of the project, a stacked column chart was designed to show the count of Attrition by JobSatisfaction and Department, with the legend indicating whether Attrition was 'Yes' or 'No'. Another stacked column chart displayed the count of employees by YearsSinceLastPromotion, further segmented by Attrition status. Additionally, a heatmap was developed based on JobRole, illustrating the count of Attrition across different roles . To enhance interactivity and filtering capabilities, slicers for Gender, OverTime, and Department were also incorporated.