Employee Attrition Analysis & Prediction

By: Adapa Mahesh

Date: July 11, 2025

1. Introduction

This project aims to analyze and predict employee attrition using Python and Power BI. By studying various HR parameters, we explore the underlying reasons why employees leave and provide insights to help HR departments improve retention strategies.

2. Dataset Overview

The dataset used is the HR Employee Attrition dataset from Kaggle. It includes variables such as Age, Department, Monthly Income, Education, Work-Life Balance, and more. Each row represents an employee, with the target variable being 'Attrition'.

3. Data Cleaning & Feature Engineering

- Removed irrelevant columns like EmployeeCount, StandardHours, etc.
- Encoded the target column (Attrition) as binary (Yes=1, No=0).
- Created new features such as AgeGroup and IncomeLevel to group employees meaningfully.
- Checked and confirmed there were no missing/null values.

4. Exploratory Data Analysis (EDA)

Visualizations were created using Seaborn and Matplotlib to analyze trends:

- Countplot of Attrition distribution
- Attrition vs Department
- Attrition vs Age Group
- Monthly Income vs Attrition (Boxplot)

These helped uncover patterns and high-risk groups.

5. Power BI Dashboard

A professional dashboard was created in Power BI showcasing:

- Attrition by Department

- Attrition by Age Group

- Monthly Income vs Attrition

- Dynamic filtering by EducationField and Gender

This dashboard makes it easy for HR to explore employee data and attrition patterns.

6. Key Insights

- Employees aged 31-40 showed the highest attrition rate.

- The Sales department had more exits than others.

- Low monthly income and poor work-life balance correlated with higher attrition.

- Employees with fewer training opportunities were more likely to leave.

7. Conclusion & Future Work

This project successfully identified key drivers of employee attrition and visualized them through an interactive Power BI dashboard. In the future, predictive models such as Logistic Regression and Random Forest can be applied to forecast attrition with accuracy.