



Get Started

ElementContent-
TitleEmployee Attrition Analysis & Prediction
Subtitle- A Python-Powered Deep Dive into HR Risk Factors
Code Focus- Pandas, Matplotlib, Seaborn, Plotly
Presented By- Group 02

The Business Problem & Data Prep

- **Problem:** Reducing preventable employee turnover is critical for cost savings and operational stability.
- **Goal:** Use data science to identify patterns and create actionable insights for HR intervention.
- **Data Setup (Pandas):**
 - Loaded employee_attrition.csv.
 - Handled missing values (df.dropna()).
 - Encoded the Target Variable: Converted Attrition from categorical ('Yes', 'No') to numerical (\$1, 0\$).
- **Initial Check:** Statistical summary provides quick insight into the distribution of numerical features (e.g., average Age, min/max Monthly Income).



Visualization Toolkit Implementation

Library/Tool	Code Section	Used Visualization	Created Value Added
Pandas	Section 2	Data Cleaning & Encoding	Prepares data for ML models
Matplotlib	Section 4 & 7	3D Scatter Plot, Pie Chart	Shows basic distribution and overall attrition rate.
Seaborn	Section 5 & 6	Correlation Heatmap, Countplot	Reveals complex relationships and departmental risk.
Plotly	Section 8 & 9	Interactive Scatter & Bar Charts	Enables dynamic exploration by end-users (HR).



Attrition Rate & 3D Exploration (Matplotlib)

1. Overall Attrition Percentage (Pie Chart)

- Code Reference: Matplotlib Pie Chart (Section 7)
- Insight: Clearly shows the overall percentage of employees who left vs. stayed (e.g., \$16.1\%\$ left).

2. The 3D Risk View

- Code Reference: Matplotlib 3D Scatter Plot (Section 4)
- Plot: Visualizes Attrition (Z-axis) across Age (X-axis) and Monthly Income (Y-axis).
- Observation: Employees who left (e.g., red points, $Z=1\$$) tend to cluster in areas of Lower Age and Lower Monthly Income, visually confirming the high-risk demographic.

Correlation Mapping (Seaborn Heatmap)

- **Code Reference:** Seaborn Heatmap (Section 5)
- **How it Works:** Visualizes the ρ (correlation coefficient) between every numerical feature pair.
- **Key Insight:** Allows for rapid identification of predictors:
- **Strong Predictors of Attrition (High ρ with 1/Yes):** Variables like OverTime, Job Level (Inverse).
- **Strong Retention Factors (High ρ with 0/No):** Variables like TotalWorkingYears, Age, MonthlyIncome.

Departmental Risk Analysis (Seaborn & Plotly Bar Charts)



1. Static Department Risk (Seaborn)

- Code Reference: Seaborn Countplot (Section 6)
- Insight: Highlights departments with the highest absolute number of leavers (e.g., Sales) and the departments with the highest attrition rate (e.g., Human Resources).



2. Interactive Department Risk (Plotly)

- Code Reference: Plotly Interactive Bar Chart (Section 9)
- Value: This visualization is interactive, allowing the user to hover over bars to see exact counts. It is ideal for inclusion in an HR dashboard.



The Interactive Risk Dashboard (Plotly)

- **Code Reference:** Plotly Interactive Scatter Plot (Section 8)
- **Visualization:** Scatter plot of Age vs. MonthlyIncome, where point color represents Attrition and size represents Job Satisfaction.
- **Interactivity:** The hover_data includes Department, Education, and Gender, allowing the user to click or hover to drill down into the specific profiles of high-risk employees.
- **Business Impact:** This dashboard is the ultimate tool for HR managers, enabling them to:
 1. Pinpoint specific individuals or small groups for intervention.
 2. Filter based on job satisfaction to understand burnout risk.



Summary of Impact

- **Data-Driven Decisions:** The analysis transforms the abstract problem of attrition into quantifiable metrics linked to salary, age, and job role.
- **Code Efficiency:** Python libraries provided the ability to perform complex analysis and generate interactive reports with minimal, concise code.
- **Value to IBM:** Supports the shift to a truly predictive HR model, maximizing retention ROI.



Thank You!

