📑 Documentation: Employee Attrition Analysis (Analytics Phase)

Project Title: Employee Attrition Analysis
Prepared By: Harsvardhan Rajgarhia
Date: September, 2025
Tools Used: Python (Pandas, Seaborn, Matplotlib), Jupyter Notebook

---

📌 1. Executive Summary

This project analyzes employee attrition patterns using the IBM HR Analytics dataset. The objective was to uncover key factors that contribute to employee turnover and provide actionable business insights for HR teams.

Attrition (employee turnover) represents the percentage of employees leaving the organization. In this dataset, approximately **16% of employees left the company**, while **84% stayed**.

The analysis reveals attrition is concentrated among **Sales employees, younger professionals, lower-income staff, and those working overtime**. Business recommendations have been provided to address these issues.

📂 2. Dataset Description

- **Source:** IBM HR Analytics Dataset (Kaggle)
- **Rows:** 1,470 employees
- **Columns:** 35 features (categorical & numerical)

Key Columns:

- **Attrition** → Target variable (Yes/No)
- **Age, Gender, Marital Status** → Demographic factors
- **Department, Job Role, Business Travel, Overtime** → Job-related factors
- **Monthly Income, Job Level** → Compensation details
- **Years at Company, Years in Current Role** → Experience and tenure
- **Satisfaction Ratings** → Work-life balance, job satisfaction

---

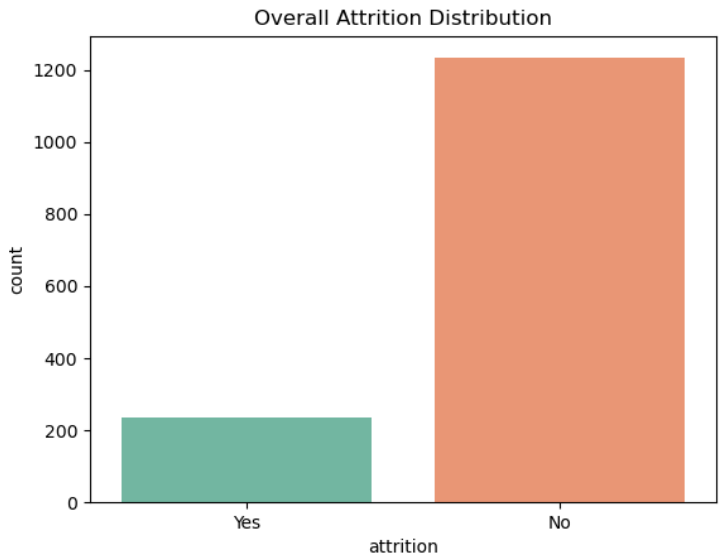
🔎 3. Analysis Approach

1. **Data Cleaning**
    - Standardized column names
    - Removed inconsistencies (e.g., spaces, case issues)
2. **Exploratory Data Analysis (EDA)**
    - Generated summary statistics
    - Created visualizations for attrition distribution across demographics, job roles, compensation, and workload factors
3. **Business Insights Extraction**
    - Interpreted each visualization
    - Linked findings to actionable HR strategies

📊 4. Key Visual Insights

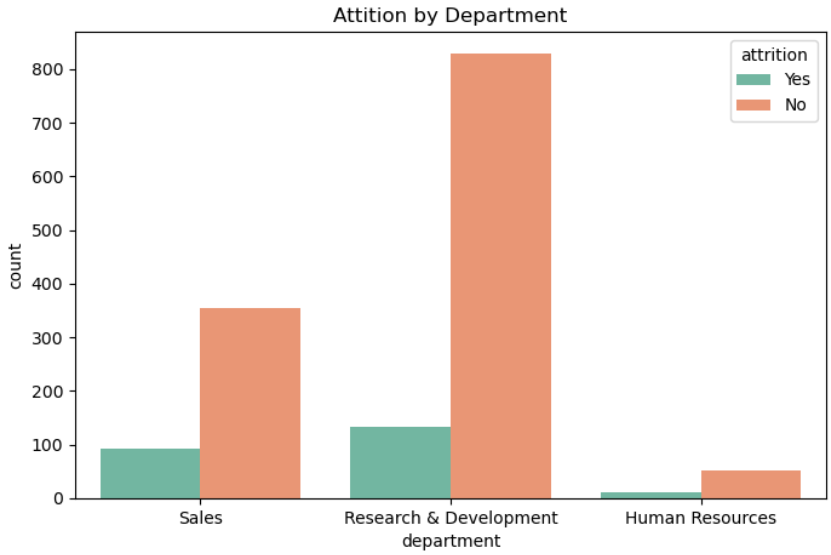
4.1 Overall Attrition

📌 About 16% of employees left the company.



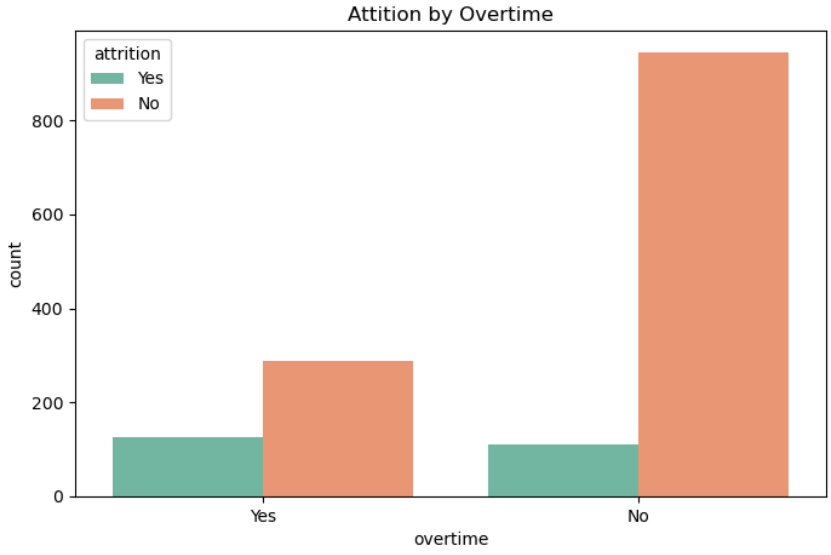
4.2 Attrition by Department

📊 **Insight:** Sales department has the highest attrition (~20%).
💡 **Recommendation:** Implement targeted retention and incentive programs in Sales roles.



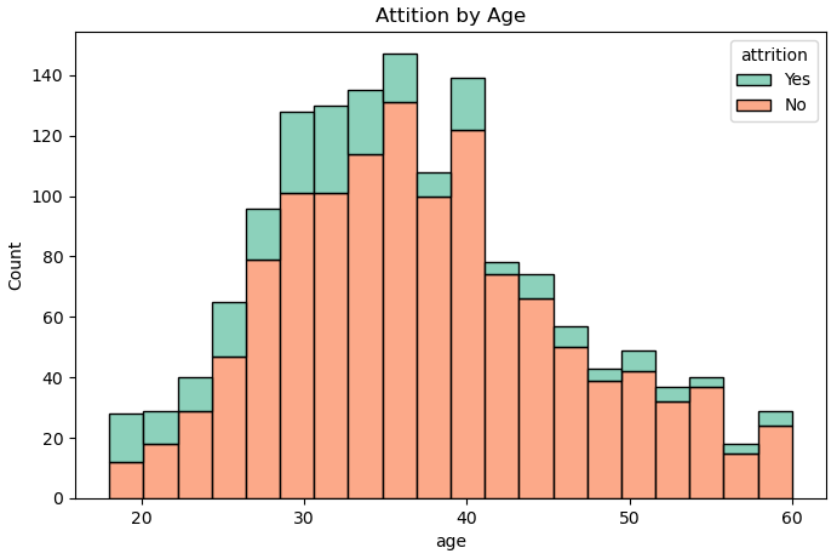
4.3 Attrition by Overtime

📊 **Insight:** Employees working overtime are ~3x more likely to quit.
💡 **Recommendation:** Balance workloads and introduce flexible scheduling or overtime pay policies.



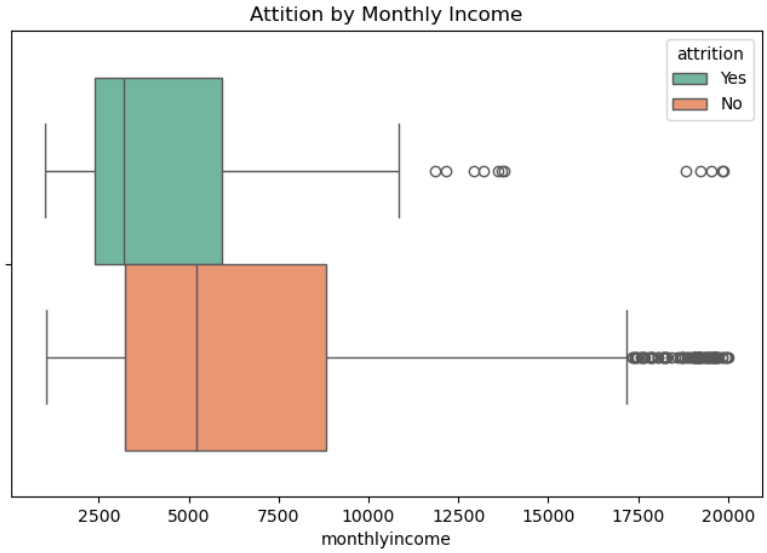
4.4 Attrition by Age

📊 **Insight:** Younger employees (25–35) face the highest attrition.
💡 **Recommendation:** Provide mentorship, learning opportunities, and career progression for early-career employees.



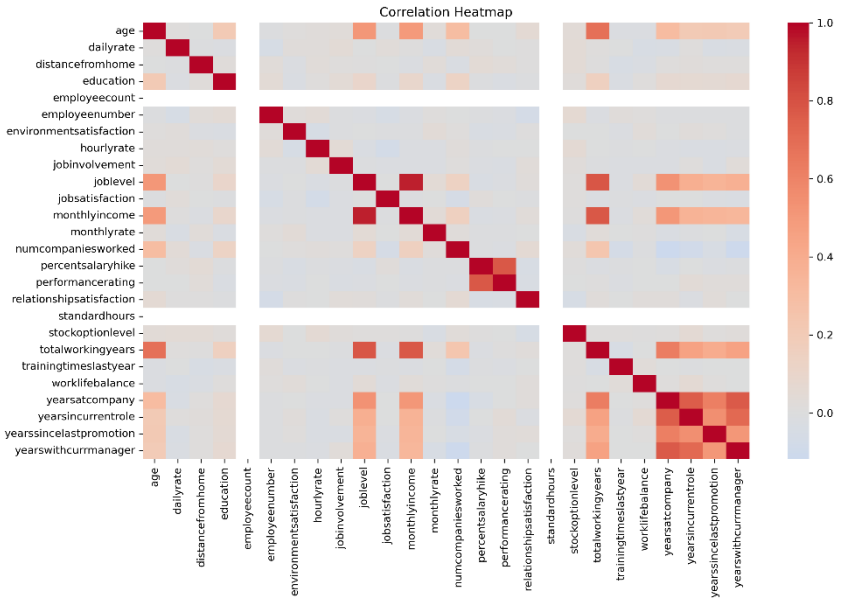
4.5 Attrition by Monthly Income

📊 **Insight:** Low-income employees (< Q2) have higher turnover.
💡 **Recommendation:** Revise compensation structures for lower income brackets.



4.6 Correlation Heatmap

📊 **Insight:** Strong correlations exist between Job Level ↔ Monthly Income, Age ↔ Total Working Years.
💡 **Recommendation:** Use these features in predictive modeling to identify at-risk employees.



📝 5. Business Recommendations (Summary)

- **Prioritize retention in Sales & entry-level employees** through mentorship and career growth paths.
- **Improve compensation** for low-income employees to reduce financial-driven attrition.
- **Address overtime workload** with flexible scheduling and wellness programs.
- **Engage younger employees (<35 years)** with mentorship and upskilling programs.

📈 6. Conclusion

The analytics phase provides clear visibility into attrition drivers within the company. Insights point toward workload management, fair compensation, and early-career employee engagement as the most impactful areas for retention improvement.

The next phase of this project will extend into **predictive modeling** (Machine Learning) to forecast which employees are at highest risk of leaving, enabling proactive HR interventions.

📎 7. Appendix

- **Dataset Link:** IBM HR Analytics Dataset (Kaggle)
- **Screenshots Folder:** Contains all visualizations (department, role, age, income, overtime, correlation heatmap).
