EDA Report: Employee Attrition Dataset

Author: Ahmed Mohammad Fayad
Date: 20/12/2024

---

1. Introduction

This report provides an overview of the exploratory data analysis (EDA) conducted on the Employee Attrition dataset. The primary objective was to identify factors contributing to attrition, extract meaningful insights, and prepare the data for further analysis and predictive modeling.

---

2. Data Overview

Initial Observations

The dataset includes demographic, job-related, and performance metrics along with attrition labels. Both numerical and categorical features were analyzed to assess data completeness and variability.

---

3. Data Cleaning

Steps Taken

- Handled missing values in both numerical and categorical features using imputation techniques.
- Addressed inconsistent data by standardizing categorical variables and resolving typographical errors.
- Removed rows with extensive missing data where imputation was infeasible.

---

4. Feature Engineering

Enhancements Made

- Created new features to capture tenure groups, attrition in binary form, and interaction effects.
- Transformed numerical features for standardization and enhanced interpretability.

---

5. Univariate Analysis

Process Overview

- Analyzed the distribution of numerical and categorical variables.
- Identified patterns in target variable distribution and its relationship with individual features.
- Used visualization techniques to highlight key trends in the dataset.

---

6. Bivariate and Multivariate Analysis

Process Overview

- Explored relationships between features and attrition through advanced visualization techniques.
- Conducted correlation analysis to understand dependencies between numerical features.
- Highlighted pairwise relationships between key variables and attrition patterns.

---

7. Outlier Detection and Handling

Approach

- Identified potential outliers using statistical techniques and visualizations.
- Retained valid data points representing realistic scenarios while addressing extreme outliers for consistency.

---

8. Advanced Insights

Findings

- Identified departments and tenure groups with higher attrition rates.
- Observed significant patterns in income distribution, job satisfaction, and work-life balance that correlated with attrition.

---

9. Key Visualizations

Examples

- Attrition rates across departments.

- Income distribution and its relationship with attrition.
- Work-life balance and job satisfaction trends.

---

10. Challenges and Solutions

Challenges

- Addressing missing and inconsistent data.
- Balancing analysis of numerical and categorical variables.
- Identifying actionable insights from multivariate patterns.

Solutions

- Applied advanced imputation, transformation, and visualization techniques.
- Focused on feature engineering to enhance interpretability and predictive value.

---

11. Conclusion and Recommendations

Conclusion

The EDA process successfully uncovered key patterns and prepared the dataset for machine learning applications. Insights from this analysis provide a solid foundation for addressing employee attrition challenges.

Recommendations

- Implement strategies to improve work-life balance and job satisfaction.
- Focus retention efforts on high-risk groups such as specific departments and tenure ranges.
- Address disparities in income distribution and other performance metrics.

---

12. Deliverables

- **Cleaned Dataset**: Ready for further analysis.
- **EDA Notebook**: Documenting all steps and insights.