**HR Analytics Dashboard**

### 1. Project Overview

**Project Title:** HR Analytics: Understanding Employee Attrition

**Authors:** Mohamed Fathy, Mohamed Haitham

**Date:** August 12, 2025

**Version:** v1.0

**Stakeholders:** Human Resources Department, Management, Data Science Team

This document provides a comprehensive technical guide for the HR Analytics project, including code snippets from the core scripts. Its purpose is to detail the methodology, code, and insights derived from an analysis of employee attrition data, ensuring the reproducibility and transparency of the findings.

### 2. Business Objective

The primary goal of this project is to analyze employee attrition within an organization. By identifying the key factors that contribute to employees leaving, this analysis provides the Human Resources and management teams with data-driven insights. These insights can be used to develop targeted strategies for improving employee retention, enhancing job satisfaction, and ultimately reducing operational costs associated with high turnover.

**3. Data Provenance & Description**

**3.1 Data Source**

The dataset for this project was obtained from a public repository on Kaggle. The raw data is stored in the file HR Analytics.csv.

**3.2 Dataset Details**

* **Original File:** HR Analytics.csv
* **Cleaned File:** HR\_Analytics\_Clean.csv
* **Initial Size:** The raw dataset contains 1,470 rows and 35 columns.
* **Key Columns:** The dataset includes a wide range of features, from employee demographics (Age, Gender, MaritalStatus), job-related information (Department, JobRole, MonthlyIncome, JobLevel), to satisfaction metrics (JobSatisfaction, WorkLifeBalance). The target variable is Attrition, a binary flag indicating whether an employee has left the company.

A screenshot of a computer

AI-generated content may be incorrect.

**4. Analysis Methodology**

The analysis pipeline was executed in a Jupyter Notebook (DA\_Project\_PRE.ipynb) and the results were deployed in a Streamlit application (app.py). Scripted in Python in VS Code IDE.

**4.1 Data Preprocessing**

The raw data required significant cleaning and transformation to prepare it for analysis and modeling.

1. **Non-Informative Column Removal:** Columns with a single unique value (EmployeeCount, Over18, StandardHours) were dropped as they provide no predictive power.
2. **Column Name Standardization:** Column names were cleaned by replacing spaces with underscores to improve code readability and access (e.g., Monthly Income became Monthly\_Income).



1. **Missing Value Imputation:** For numeric columns with missing values (NumCompaniesWorked, YearsInCurrentRole, etc.), the missing data was imputed with the mean of the respective column.

A computer screen shot of a computer code

AI-generated content may be incorrect.

1. **Feature Engineering:** A new categorical feature, Age\_Group, was created by binning the Age column into logical ranges (e.g., '20-29', '30-39', etc.).

A screenshot of a computer

AI-generated content may be incorrect.

1. **Categorical to Ordinal Mapping:** Several categorical features were mapped to more descriptive, human-readable labels:
   * Education, EnvironmentSatisfaction, JobInvolvement, JobSatisfaction, RelationshipSatisfaction, WorkLifeBalance were mapped from numerical ratings (1-4) to categories like "Low," "Medium," "High," and "Very High."

A close-up of a computer code

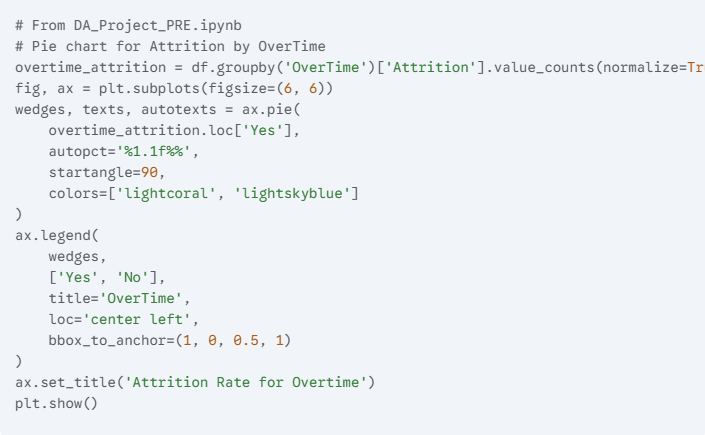
AI-generated content may be incorrect.

1. **Outlier Capping:** Outliers in continuous variables such as Age, Monthly Income, and Total Working Years were handled using an IQR-based method to cap values at the 1.5 IQR range. This prevents extreme values from skewing the analysis.

**4.2 Exploratory Data Analysis (EDA)**

EDA was performed to uncover initial insights and form hypotheses. Key findings were visualized to understand the distribution of attrition across different employee characteristics:

* **Overtime and Attrition:** The analysis revealed a strong correlation between working overtime and attrition.
* **Job Role & Department:** Certain job roles, particularly in the Sales department, showed higher attrition rates compared to roles in Research & Development or Human Resources.
* **Monthly Income:** A clear relationship was observed where employees with lower monthly incomes were more likely to leave.
* **Job & Environment Satisfaction:** Low scores on satisfaction metrics, such as JobSatisfaction and EnvironmentSatisfaction, were linked to increased attrition.



**4.3 Predictive Modeling**

A **Logistic Regression** model was chosen to predict attrition. This model is suitable for binary classification problems and provides a clear interpretation of feature importance.

* **Target Variable:** Attrition (Yes/No)
* **Features Used:** A set of key features including Age, JobSatisfaction, JobInvolvement, OverTime, YearsAtCompany, YearsInCurrentRole, YearsWithCurrentManager, and YearsSinceLastPromotion were selected for the model.
* **Procedure:** The dataset was split into training and testing sets to train and evaluate the model's performance on unseen data.

A screenshot of a computer program

AI-generated content may be incorrect.

### 5. Key Insights & Recommendations

The analysis revealed several actionable insights that can be used to improve employee retention:

* **Early Career Attrition:** Employees in the early stages of their careers tend to switch jobs more frequently.
* **Compensation is Key:** Higher salaries and stock options are significant motivators that lead to lower attrition rates.
* **Work-Life Balance:** While good work-life balance is a motivator, it is not a sufficient condition for retention, as employees may still leave for better opportunities.
* **Job Satisfaction:** High job satisfaction and environment satisfaction scores are correlated with greater employee loyalty.
* **Departmental Hotspots:** Sales and other performance-driven departments have higher attrition. This indicates a need for targeted retention programs in these areas.

### 6. Dashboard & Visualization

The project includes an interactive dashboard developed with Streamlit. The app.py script powers the dashboard, which contains:

* **Interactive Visualizations:** Users can filter data by department, gender, and job role to view visualizations of attrition rates.
* **Attrition by Factor:** Charts are provided to show the relationship between attrition and Gender, Department, MaritalStatus, Age\_Group, JobSatisfaction, OverTime, and BusinessTravel.
* **Attrition Prediction Tool:** An interactive widget allows users to input employee details and receive an attrition prediction from the trained Logistic Regression model.

### 7. Data Limitations & Assumptions

* **Limited Departmental Data:** The dataset is limited to only three departments (Sales, Research & Development, Human Resources). This may not be representative of the attrition patterns in other departments within the organization.
* **Generalizability:** The findings and the model are based on a specific dataset and may not be fully generalizable to all companies without further validation.
* **Causality:** While the analysis shows strong correlations, it does not prove causality. For example, while overtime is correlated with attrition, other unmeasured factors might also be at play.