Data Analyst Project Report

Sobia Anam Atif Ahmed

19 May 2025

Tools Used: Python (Jupyter Notebook), Power BI

Executive Summary

Description about the data

- Age: A period of employee life, measured by years from birth.
- Attrition: The departure of employees from the organization.
- BusinessTravel: Did the employee travel on a business trip or not.
- DailyRate: Employee salary for the period is divided by the amount of calendar days in the period.
- Department: In which department the Employee working.
- DistanceFromHome: How far the Employee live from the office location.
- Education: In education 1 means 'Below College', 2 means 'College', 3 means 'Bachelor', 4 means 'Master', 5 means 'Doctor'
- EducationField: In which field Employee complete his education.
- EmployeeCount: How many employee working in a department
- EmployeeNumber: An Employee Number is a unique number that has been assigned to each current and former State employee and elected official in the Position and Personnel DataBase (PPDB).
- Job involvement: Is the degree to which an employee identifies with their work and actively participates in it where 1 means 'Low', 2 means 'Medium', 3 means 'High', 4 means 'Very High'
- JobLevel: Job levels, also known as job grades and classifications, set the responsibility level and expectations of roles at your organization. They may be further defined by impact, seniority, knowledge, skills, or job title, and are often associated with a pay band. The way you structure your job levels should be dictated by the needs of your unique organization and teams.
- JobRole: What is the jobrole of an employee.

- JobSatisfaction: Employee job satisfaction rate where, 1 means 'Low', 2 means 'Medium', 3 means 'High', 4 means 'Very High'
- MaritalStatus: Marital status of the employee.
- MonthlyIncome: total monetary value paid by the organization to an employee.
- MonthlyRate: The per-day wage of the employee.
- NumCompaniesWorked: Before joining this organization how many organizations employee worked.
- Over18: Is the employee age over than 18 or not.
- OverTime: A Employee works more than 9 hours in any day or for more than 48 hours in any week.
- PercentSalaryHike:
- PerformanceRating 1 'Low' 2 'Good' 3 'Excellent' 4 'Outstanding'
- EnvironmentSatisfaction 1 'Low' 2 'Medium' 3 'High' 4 'Very High'
- RelationshipSatisfaction 1 'Low' 2 'Medium' 3 'High' 4 'Very High'
- StandardHours: Is the number of hours of production time that should have been used during an working period.
- StockOptionLevel: Employee stock options, also known as ESOs, are stock options in the company's stock granted by an employer to certain employees. Typically they are granted to those in management or officer-level positions. Stock options give the employee the right to buy a certain amount of stock at a specific price, during a specific period of time. Options typically have expiration dates as well, by which the options must have been exercised, otherwise they will become worthless.
- TotalWorkingYears: Total years the employee working in any organization
- TrainingTimesLastYear: Last year how many times employee took training session.

- WorkLifeBalance 1 'Bad' 2 'Good' 3 'Better' 4 'Best'
- YearsAtCompany: How many years the employee working in the current organization
- YearsInCurrentRole: How many years the employee working in the current position
- YearsSinceLastPromotion: How many years the employee working in the current position after promotion
- YearsWithCurrManager: How many years the employee working under the current manager

Introduction

Aim is to analyze the dataset completely with respect to each feature and find the reason behind Attrition of Employees and what are the top factors which lead to employee attrition

Data Cleaning & Preprocessing

1. Looked for missing values and duplicates
2. Checked shape of datasheet
3. Checked head, tail and sample of dataset
4. Checked Data Description
5. check the Data Types of the columns present in the data and changing object data type to int because object datatype won't work in Logistic Regression model

Exploratory Data Analysis (EDA)

Bar Chart was created for Attrition versus Department, EducationField, JobRole and Gender.

OBSERVATIONS

- Employees working in R&D department are more, but employees from sales department or at position like sales executive, sale Representative leaves the job early.
- Males are more under Attrition then Females
-Attrition with respect to bachelor can be seem more because they have more and more expectation from companies and it will be interesting to see the reason behind this in this dataset.

Analysis / Modeling

The method of Resampling was used which is a nonparametric method of statistical inference. We are going to use Over Sampling given that the data entries are very less. We will not use Under Sampling to avoid data loss.

Data would be split into training and testing set in 70:30 ratio.
Logistic Regression model was used from Classification which was accessed through scikit learn

Insights

OBSERVATIONS

- Employees working in R&D department are more, but employees from sales department or at position like sales executive, sale Representative leaves the job early.
- Males are more under Attrition then Females
-Attrition with respect to bachelor can be seem more because they have more and more expectation from companies and it will be interesting to see the reason behind this in this dataset

Recommendations

For 1st observation: **Re-design incentives:** Offer clearer, more transparent commission structures and shorter payout cycles.
Career ladders: Publish a concrete path from Sales Rep → Sr. Rep → Account Manager → Sales Manager with required skills & timelines.
For 2nd observation: Pulse surveys segmented by gender: Identify specific dissatisfaction drivers for male employees (e.g., pay fairness, overtime, manager feedback).
For 3rd observation: Give Bachelor's employees a fixed annual budget for certifications (sales, analytics, product) and link completion to promotion readiness.