

## **Table of Contents**

### **Contents**

1. PROJECT TITLE .....	2
2. PROJECT IDEA .....	2
3. PROJECT OBJECTIVES .....	2
4. TOOLS & TECHNOLOGIES USED .....	2
5. DATASET DESCRIPTION .....	3
6. PROJECT WORKFLOW .....	4
7. PROJECT FEATURES.....	5
8. EXPECTED OUTCOMES.....	6
9. TEAM MEMBERS & ROLES .....	6
10. CONCLUSION .....	6

# **HR DATA ANALYSIS PROJECT**

## **1. PROJECT TITLE**

HR Data Analysis for Employee Performance and Attrition

## **2. PROJECT IDEA**

This project focuses on analyzing Human Resources (HR) data to identify the key factors that influence employee attrition and performance.

The analysis explores relationships between variables such as job satisfaction, work-life balance, salary, training, promotions, business travel, and stock options. The main goal is to use data-driven insights to help organizations understand employee behavior and improve retention strategies.

## **3. PROJECT OBJECTIVES**

The main objectives of this project are:

- To analyze employee performance across different departments and job roles
- To study employee attrition and identify its main causes
- To analyze salary distribution and its impact on performance and satisfaction
- To measure job satisfaction levels
- To provide insights that support better human resource decision-making

## **4. TOOLS USED**

- Python (for data cleaning, analysis)
- Jupyter Notebook
- Tableau (Data Visualization)
- GitHub for version control and project hosting

## 5. DATASET DESCRIPTION

This project uses two main datasets related to Human Resources (HR) employee records and performance evaluations.

The first dataset, Employee.csv, contains **1,470** employee records with **23** columns.

The columns in this dataset are:

Column Name	Description
<b>EmployeeID</b>	Unique identifier for each employee
<b>FirstName</b>	Employee's first name
<b>LastName</b>	Employee's last name
<b>Gender</b>	Gender of the employee
<b>Age</b>	Age of the employee in years
<b>BusinessTravel</b>	Frequency of business travel
<b>Department</b>	Department where the employee works
<b>DistanceFromHome (KM)</b>	Distance between home and workplace
<b>State</b>	State of residence
<b>Ethnicity</b>	Employee's ethnic background
<b>Education</b>	Education level
<b>EducationField</b>	Field of education
<b>JobRole</b>	Current job position
<b>MaritalStatus</b>	Single, married, divorced
<b>Salary</b>	Monthly salary
<b>StockOptionLevel</b>	Stock options level
<b>OverTime</b>	Works overtime (Yes/No)
<b>HireDate</b>	Date of joining the company
<b>Attrition</b>	Whether employee left (Yes/No)
<b>YearsAtCompany</b>	Total years in the company
<b>YearsInMostRecentRole</b>	Years in current role
<b>YearsSinceLastPromotion</b>	Years since last promotion
<b>YearsWithCurrManager</b>	Years with current manager

The second dataset, PerformanceRating.csv, contains 6,709 records with 11 columns. This dataset stores multiple performance evaluations for employees across different review periods. The columns in this dataset are:

Column Name	Description
PerformanceID	Unique ID for each review
EmployeeID	Links to Employee.csv
ReviewDate	Date of evaluation
EnvironmentSatisfaction	Satisfaction with work environment
JobSatisfaction	Satisfaction with job
RelationshipSatisfaction	Satisfaction with colleagues
TrainingOpportunitiesWithinYear	Training offered in the year
TrainingOpportunitiesTaken	Training attended
WorkLifeBalance	Work-life balance score
SelfRating	Rating given by employee
ManagerRating	Rating given by manager

The two datasets are linked using the EmployeeID column, which allows combining employee personal information with their performance and satisfaction evaluations.

## 6. PROJECT WORKFLOW

### 1. Data Collection

The HR datasets (Employee.csv and PerformanceRating.csv) were collected and reviewed to understand the structure, size, and type of information available.

### 2. Data Cleaning

The raw datasets were cleaned to ensure high data quality. This stage included:

- Handling missing values
- Removing duplicate records
- Correcting inconsistent data formats
- Preparing the datasets for analysis

### **3. Data Analysis**

After cleaning, the analysis team member explored the data to identify patterns, relationships, and key HR insights such as attrition trends, performance distribution, and salary comparisons.

### **4. Data Visualization & Dashboard**

The visualization team member created charts and visual dashboards to present the analysis results in a clear and interactive manner for better understanding by decision-makers.

### **5. Documentation & Reporting**

The documentation stage included writing the full project report, dataset description, data dictionary, and the README file for GitHub to clearly explain the project structure and results.

### **6. Presentation Preparation**

Finally, the presentation team member prepared the PowerPoint slides to summarize the project idea, workflow, analysis results, team roles, and future improvements for the final project defense.

## **7. PROJECT FEATURES**

- Employee attrition analysis
- Performance analysis by department and job role
- Salary and income distribution
- Job satisfaction evaluation
- HR decision-support insights

## 8. EXPECTED OUTCOMES

Through this project, organizations can:

- Reduce employee turnover
- Improve employee performance
- Identify high-risk employees for attrition
- Enhance job satisfaction levels
- Optimize HR policies using data

## 9. TEAM MEMBERS & ROLES

Mostafa Bakry	Esraa Ayman	Menna Allah taha	Rahma Hassan
Data Cleaning, project Documentation	Data processing, Data Analysis	Data Visualization, dashboard	Reporting & Presentation

## 10. CONCLUSION

This HR Data Analysis Project demonstrates how data analytics can be applied to human resource management. By analyzing real employee data, the project provides meaningful insights that help organizations enhance workforce performance and reduce attrition using evidence based decisions.