HR Data analysis

This project focuses on analysing HR data to identify key factors influencing employee retention, satisfaction, and overall workforce dynamics within the organization. Using a comprehensive dataset encompassing various attributes such as employee demographics, job roles, compensation, and satisfaction metrices, we performed extensive data analysis to uncover patterns and trends.

Utkarsh Sharma

sharma.utkarsh2402@gmail.com

**Project Overview**

The HR analytics project aimed to uncover key factors influencing employee retention, satisfaction, and overall workforce dynamics within the organization. By leveraging a comprehensive dataset that included various employee attributes, the project sought to identify patterns and trends that could inform strategic HR decisions.

1. **Dataset:** The dataset comprised columns such as 'EmpID', 'Age', 'AgeGroup', 'Attrition', 'BusinessTravel', 'DailyRate', 'Department', 'DistanceFromHome', 'Education', 'EducationField', 'EmployeeCount', 'EmployeeNumber', 'EnvironmentSatisfaction', 'Gender', 'HourlyRate', 'JobInvolvement', 'JobLevel', 'JobRole', 'JobSatisfaction', 'MaritalStatus', 'MonthlyIncome', 'SalarySlab', 'MonthlyRate', 'NumCompaniesWorked', 'Over18', 'OverTime', 'PercentSalaryHike', 'PerformanceRating', 'RelationshipSatisfaction', 'StandardHours', 'StockOptionLevel', 'TotalWorkingYears', 'TrainingTimesLastYear', 'WorkLifeBalance', 'YearsAtCompany', 'YearsInCurrentRole', 'YearsSinceLastPromotion', 'YearsWithCurrManager', 'BusinessTravel1', and 'Overtime'.
2. **Objectives:**

* Identify key factors contributing to employee attrition.
* Analyse the relationship between job satisfaction and employee retention.
* Explore the impact of compensation, benefits, and work-life balance on employee satisfaction.
* Develop targeted retention strategies based on demographic and job-related factors.

1. **Key Findings:**

* **Attrition Patterns:** This metric shows that how many employees leave the company at a specific period. And in this key metrics we found that the attrition rate of this dataset is 16.08%.
* **Impact of travelling -** In this metric we analyse that the job satisfaction of the employees according to their different business travel. And we found that those employees who are travel occasionally are highest satisfied with their jobs as the average rate is 2.89 compared to Frequent traveller, non-traveller and rare traveller. Rare travellers have the lowest satisfaction rate with the average rate is 2.70.
* **Compensation and Benefits-** Monthly income and salary slab data indicate that competitive compensation is crucial for retaining top talent. Employees in higher salary slabs are less likely to leave the company. Regular salary reviews and performance-based incentives can motivate employees to stay.
* **Demographic Factors -** Gender, marital status, and education field also play a role in employee satisfaction and retention. Tailored initiatives that address the unique needs of different demographic groups can foster a more inclusive and supportive work environment.

**Cleaning of the data**

In this phase we need to identify and remove inaccurate, duplicates, nulls, and irrelevant data, which affect our dataset when we analyse. It is a very crucial process for extracting the accurate insights from the data.

Here are some steps of how you clean this data: -

**Step1- First you need import all the library.**

A screen shot of a computer code

Description automatically generated

We use the variable called ‘reset’ to reset all the figure size used in this analysis.

**Step2- We need to read the database by using the panda’s library.**

A white rectangular object with red and blue text

Description automatically generated

**A screenshot of a computer

Description automatically generatedStep3- Now we need to see the top 5 rows from the dataset.**

**Step4 – Then we need to see the last 5 rows.**

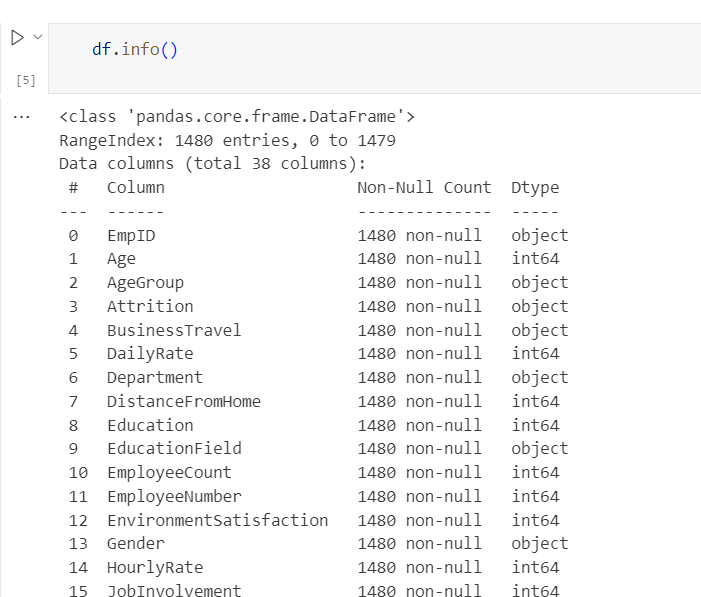
A screenshot of a computer

Description automatically generated

**A screenshot of a computer

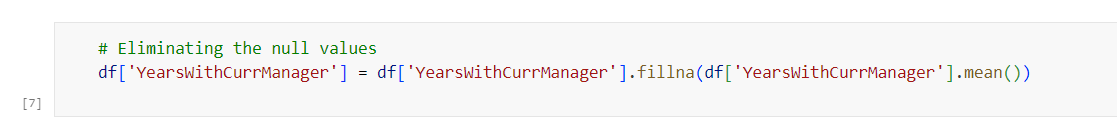
Description automatically generatedStep5- We need to see all the statistics values or 5 pointer theory.**

**Step6- We need to see the information related to the dataset.**

****

**A screenshot of a computer

Description automatically generatedStep7- We need to check all the null values in the dataset.**

**Step8 – After identifying all the null values we need to fill those null values with mean/ median/ mode according to the dataset requirement.**

**A white rectangular object with black text

Description automatically generatedStep9 – We need to remove all the null values (if present in the dataset).**

**Step10 – Shape function will show the count of rows and columns. Here 1473 are the total rows and 38 columns.**

**A white rectangular object with black text

Description automatically generated**