DEPI
Final project of data analysis
Human Resource

Supervised by: **Date:**

Eng/Osama Adel   **24/10/2024**

Agenda

About Data

This dataset provides comprehensive HR-related data for employees at Company X, including personal information, job roles, performance reviews, training records, and attrition details. It is structured into four interconnected tables. Below is a summary of each table and its columns:

1. Employees Table

This table stores the personal and job-related details of employees.

Columns:

- **EmployeeID**: Unique identifier for each employee.
- **full_name**: Employee's full name.
- **Gender**: Gender of the employee.
- **Age**: Employee's age.
- **TravelFrequency**: Frequency of work-related travel.
- **Department**: The department where the employee works.
- **DistanceFromHome**: Distance from the employee's home to the workplace.
- **State**: The state where the employee is located.
- **Ethnicity**: Employee's ethnicity.
- **EducationLevel**: Employee's highest education level.
- **EducationField**: Field of study.
- **JobRole**: Employee's job title.
- **Marital Status**: Marital status of the employee.
- **Salary**: Annual salary of the employee.
- **StockOptionLevel**: Level of stock options provided.
- **Overtime**: Indicates if the employee works overtime (Yes/No).
- **HireDate**: Date when the employee was hired.
- **Attrition**: Whether the employee left the company (Yes/No).
- **YearsAtCompany**: Total years the employee has worked at the company.
- **RecentRoleYears**: Years spent in the most recent role.
- **PromotionGapYears**: Years since the last promotion.
- **YearsWithCurrManager**: Years under the current manager.
- **Attrition Year**: The year when the employee left the company.

2. Performance Table

This table records details of employee performance reviews and training activities.

Columns:

- **PerformanceID**: Unique identifier for each performance review.
- **EmployeeID**: Foreign key linking to the Employees table.
- **ReviewDate**: Date of the performance review.
- **EnvironmentSatisfaction**: Rating of satisfaction with the work environment.
- **Job satisfaction**: Level of satisfaction with the job.
- **RelationshipSatisfaction**: Satisfaction with interpersonal relationships at work.
- **AnnualTrainingOpportunities**: Number of training opportunities provided annually.
- **TrainingTaken**: Number of training courses taken.
- **WorkLifeBalance**: Work-life balance rating.
- **Self-Rating**: Employee's self-assessment score.
- **Manager rating**: Manager's rating of the employee's performance.

3. Rating Table

This table provides details about performance rating levels.

Columns:

- **RatingID:** Unique identifier for the rating level.
- **RatingLevel:** Name or description of the rating level.

4. Satisfied Table

This table contains information about satisfaction levels and their descriptions.

Columns:

- **SatisfactionID:** Unique identifier for each satisfaction level.
- **SatisfactionLevel:** Description of the satisfaction level

Cleaning Data

Using Python

- Read Data

- merge the column of First & Last names as full name

- Rename columns

- Check Validation of data

  - Check Nulls

  - Check Missing in Data

  - Check Data Type

  - Check Duplicated

  - Delete rows that are illogical in performance Table

  - Check columns if containing zero or negative is valid or not valid

  - Check the value of the foreign key to see if it is true

- Merge the table of Employees and the table of Education

- Download the modified data

Using SQL Server

- Import the modified data to the SQL Server

- Edit data type in SQL Server

- Define Primary key

- Do a diagram of the data

- Create the relation between tables using a foreign key

- Create a column for attrition year in the Employees

- Overview data and understand the columns

Analysis Data &Prediction

- Check outliers in salary.
    - there are 124 outliers (99/168 managers - 25 Employees).
- Check the average Salary in each year.
    - The average salary for each job Role decreases over time
- Check the Hire and Fire of employees Each year.
    - The rate of hire increases over time
- The Attrition of the employees in each state in each job role.
- The relation between stock option level and Year at the company
- The Relationship between the year with the current manager and the attrition rate.
- The Relation Between Manager Rating and Attrition Rate
- The Relation between Relation Satisfaction and Attrition Rate
- The Relation between Life Balance and Attrition Rate
- The relation between Attrition rate and travel frequency
- The Relation between Overtime and Attrition Rate
- The Relation between overtime, travel frequency, and Attrition Rate.

Prediction

predict the Average salary for each job role in the next five years.

Data Visualization

Create a dashboard that shows an overview of the Capacity of the company (Tableau)

Create Dashboards to measure the reason for the Attrition rate in each state (Bower Bi)

Recommendations

- Create a system to control the salary to justify the qualifications of the employees and take into account their education level, job role, overtime, and Travel frequency.

- As a result of lowering the average salary over the last few years, the company can hire more Employees in each department.

- there is a strong relation between the existence of the employee in the company and his chance to have a stock options level.

- there are relations between overtime, travel frequency, and the attrition rate of the Employees.

- There is a relationship between the number of years with the current manager and the attrition rate of the employees.

- There is a relationship between the manager Rating and attrition

- There is a relation between the work-life balance and attrition rate

- There is a relationship between the relation job satisfaction and attrition Rate