Human Resources Dataset Analysis

Final Project

Period→ 4 WEEKS

Team Members

Omar Mohy El Din Hegazy

Abdelrhman Mohamed Yasser

Mostafa Mohamed

Hazem Ahmed Abd Elkader

- ## Week 1 → Build Data Model, Data Cleaning and Preprocessing

1. Activities and Tasks

In the first week, we focused on collecting relevant data .

After the data collection phase, we moved into the **Preprocessing** stage using Python. This involved cleaning the data and preparing it for further analysis.

2. Tools Used

- **Programming Language**: Python
- **Python Libraries**: We used libraries such as `pandas` and `numpy` to handle data cleaning and transformation.
- **Data Analysis Environment**: Jupyter Notebook was used to write and execute the code.

3. Challenges Encountered

During the preprocessing stage, we faced a few challenges related to the data types of certain columns:

- **Hire Date**: The date format was inconsistent, and we converted the data type to ensure uniformity across all records.
- **Salary**:
  - There were unnecessary spaces in the salary column, which were removed using the `strip()` function.
  - We also removed commas from salary values using the `replace()` function to enable numeric conversion.

- **Removing Duplicates**:
  - Some records were repeated in the dataset, so we used the `drop_duplicates()` function to eliminate duplicate rows, ensuring data integrity.

- **Standardization of Columns**:
  - To ensure consistency across various columns, we applied **standardization**. This included unifying text formats, such as making all entries in categorical columns Uppercase or capitalizing the first letter where necessary. This process helped reduce inconsistencies that could affect later analysis.

- The columns needed for the analysis have been identified, and a **calculated column** was created for the **Age** and **Salary** columns to categorize them into different groups.
- Create a calculated column for the state to categorize it as either HQ or branches .

Deliverables

Cleaned dataset ready for analysis.
Data preprocessing notebook.""

- **Week 2→ Analysis Questions Phase**

Task

Determine Data Analysis Questions:

Determine all possible analysis questions that can be deducted from the given dataset and would be of interest to the organization's decision makers.

1. Average Age, Highest, and Lowest Age
2. Highest Department Salary and Job Role
3. Average Salary by Ethnicity
4. State by Salary and Count of Employees
5. Correlation between Education Levels and Salary
6. Highest Paid Role and Department
7. Pie Chart of Marital Status and Correlation with Salaries
8. Correlation between Marital Status and Overtime
9. Total Salaries
10. Correlation between Distance from Home and Overtime
11. Hire Date by Count of Employees (Line Chart)
12. Count of Employees and Count of Terminated Employees
13. Most Turnover Year
14. Salaries by Gender
15. Average Years at the Company
16. Correlation between Years at the Company and Marital Status
17. Correlation between Years at the Company and Salaries
18. Department with Fastest Promotions
19. State with Age
20. Correlation between Work-Life Balance and Marital Status
21. Correlation between Manager Rating and Promotion
22. Difference between Self-Rating and Manager Rating
23. Training Opportunities Taken and Correlation with Salaries or Promotions

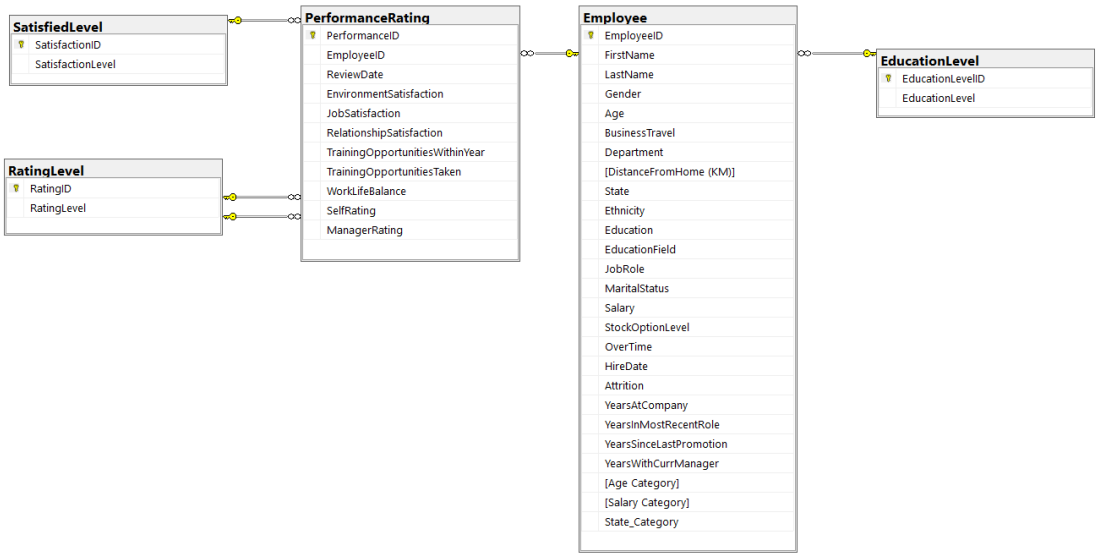
Deliverables:

Set of analysis questions that can be answered via the dataset.

Week 3→ Data Modeling and SQL Queries

Task

I focused on data modeling using SQL Server. I designed the database structure and established relationships between the tables to ensure efficient data flow.



After that, I wrote SQL queries to answer specific questions and derive insights that I had previously identified. This helped in understanding the data more clearly and provided actionable insights based on the queries

```sql
SElECT e.FirstName,l.EducationLevel
FROM Employee e
JOIN EducationLevel l ON l.EducationLevelID=e.Education
ORDER BY FirstName;
```

Week 4→ Visualization Dashboard and Forecasting

Following the data modeling and querying phase, I transitioned to data visualization and forecasting using Tableau. This stage involved creating interactive dashboards to display key insights and trends derived from the data. Additionally, I incorporated forecasting techniques to predict future outcomes based on historical data, providing valuable projections that support decision-making

- **Average Age, Highest, and Lowest Age**: Analyzes age demographics and diversity within the workforce.
- **Highest Department Salary and Job Role**: Identifies top-paying roles and departments for talent investment insights.
- **Average Salary by Ethnicity**: Reveals salary disparities across ethnic groups, promoting equity discussions.
- **State by Salary and Count of Employees**: Visualizes salary distribution and employee counts across states.
- **Correlation between Education Levels and Salary**: Examines how education impacts salary for recruitment strategies.
- **Highest Paid Role and Department**: Highlights the highest compensation role and department for benchmarking.
- **Pie Chart of Marital Status and Correlation with Salaries**: Shows marital status distribution and its relationship with salary.
- **Correlation between Marital Status and Overtime**: Analyzes overtime hours based on marital status for work-life balance insights.
- **Total Salaries**: Summarizes overall salary expenditure for budget planning.
- **Correlation between Distance from Home and Overtime**: Investigates if longer commutes lead to more overtime hours.
- **Hire Date by Count of Employees (Line Chart)**: Tracks hiring trends over time to inform recruitment strategies.
- **Count of Employees and Count of Terminated Employees**: Compares employee retention and turnover rates.
- **Most Turnover Year**: Identifies the year with the highest turnover for retention strategy insights.
- **Salaries by Gender**: Analyzes salary distribution by gender to detect pay gaps.
- **Average Years at the Company**: Provides insights into employee tenure and retention effectiveness.
- **Correlation between Years at the Company and Marital Status**: Investigates how tenure relates to marital status.
- **Correlation between Years at the Company and Salaries**: Assesses how salary increases correspond with tenure.
- **Department with Fastest Promotions**: Identifies departments with the highest promotion rates for career development insights.
- **State with Age**: Examines age distribution across states for regional HR strategies.

- **Correlation between Work-Life Balance and Marital Status**: Analyzes how marital status affects work-life balance perceptions.

- **Correlation between Manager Rating and Promotion**: Investigates the relationship between manager ratings and promotions.

- **Difference between Self-Rating and Manager Rating**: Compares self-assessments with manager evaluations for bias detection.

- **Training Opportunities Taken and Correlation with Salaries or Promotions**: Examines how training impacts salary and promotion outcomes.