Human Resources Data Set Analysis Report

Objective:
Analyze HR data to improve retention and performance.

Scope:
Data cleaning, EDA, forecasting, visualization.

Tools:
Excel, SQL, Power BI

Prepared by:
1. Marwa Lotfy Ismail
2. Donia Osama Mohamed
3. Mennat Allah Fikry Abdelkader
4. Mostafa Hassan Mohamed
5. Shahinda Adel Hussein

Data Schema:

1. **Employees Table**
    - Demographics
    - Salary
    - Roles
    - Attrition
2. **Performance Rating Table**
    - Satisfaction
    - Ratings
3. **Education Level Table**
    - Mapping of education levels
4. **Rating Level Table**
    - Performance Categories
5. **Satisfied Level Table**
    - Satisfaction Scale

Basic Descriptive Statistics

1. **Total Number of Employees**

```sql
SELECT COUNT(*) AS total_employees FROM employees;
```

Result :

- Total number of employees: **1470**

2. **Department Distribution**

```sql
SELECT Department, COUNT(*) AS Employee_Count
FROM employees
GROUP BY Department
ORDER BY Employee_Count DESC;
```

Result :

Department	Employee_Count
Technology	961
Sales	446
Human Resources	63

3. **Average Salary by Department**

```sql
UPDATE employees
SET Salary = REPLACE(Salary, '$', '');

UPDATE employees
SET Salary = REPLACE(Salary, ',', '');

SELECT Department,
       COUNT(*) AS Employee_Count,
       ROUND(AVG(Salary), 0) AS Avg_Salary,
       MIN(Salary) AS Min_Salary,
       MAX(Salary) AS Max_Salary
FROM employees
GROUP BY Department
ORDER BY Avg_Salary DESC;
```

Result :

Department	Employee_Count	Avg_Salary	Min_Salary	Max_Salary
Human Resources	63	119699	131047	99052
Sales	446	119118	100314	98435
Technology	961	109655	100576	99936

4. Attrition Rate

```sql
SELECT
    (COUNT(CASE WHEN Attrition = 'Yes' THEN 1 END) * 100.0 / COUNT(*)) AS Attrition_Rate
FROM employees
```

Result :

Attrition_Rate
16.12245

5. Attrition by Education Level

```sql
SELECT edu_level,
       COUNT(*) AS Total_Employees,
       COUNT(CASE WHEN Attrition = 'Yes' THEN 1 END) AS Employees_Left,
       (COUNT(CASE WHEN Attrition = 'Yes' THEN 1 END) * 100.0 / COUNT(*)) AS Attrition_Percentage
FROM employees
GROUP BY edu_level
ORDER BY Attrition_Percentage DESC;
```

Result:

edu_level	Total_Employees	Employees_Left	Attrition_Percentage
No Formal Qualifications	170	31	18.23529
Bachelors	572	99	17.30769
High School	282	44	15.60284
Masters	398	58	14.57286
Doctorate	48	5	10.41667

6. Attrition by Salary

```sql
SELECT
    CASE
        WHEN Salary < 30000 THEN 'Low (Below 30K)'
        WHEN Salary BETWEEN 30000 AND 60000 THEN 'Medium (30K-60K)'
        WHEN Salary BETWEEN 60000 AND 100000 THEN 'High (60K-100K)'
        ELSE 'Very High (100K+)'
    END AS Salary_Range,
    COUNT(*) AS Total_Employees,
    COUNT(CASE WHEN Attrition = 'Yes' THEN 1 END) AS Employees_Left,
    (COUNT(CASE WHEN Attrition = 'Yes' THEN 1 END) * 100.0 / COUNT(*)) AS Attrition_Percentage
FROM employees
GROUP BY Salary_Range
ORDER BY Attrition_Percentage DESC;
```

Result:

Salary_Range	Total_Employees	Employees_Left	Attrition_Percentage
Low (Below 30K)	157	45	28.66242
Medium (30K-60K)	449	93	20.71269
High (60K-100K)	333	46	13.81381
Very High (100K+)	531	53	9.98117

7. Attrition by Age Group

```sql
SELECT
    CASE
        WHEN Sum_of_Age < 25 THEN 'Under 25'
        WHEN Sum_of_Age BETWEEN 25 AND 34 THEN '25-34'
        WHEN Sum_of_Age BETWEEN 35 AND 44 THEN '35-44'
        WHEN Sum_of_Age BETWEEN 45 AND 54 THEN '45-54'
        ELSE '55+'
    END AS Age_Group,
    COUNT(*) AS Total_Employees,
    COUNT(CASE WHEN Attrition = 'Yes' THEN 1 END) AS Employees_Left,
    (COUNT(CASE WHEN Attrition = 'Yes' THEN 1 END) * 100.0 / COUNT(*)) AS Attrition_Percentage
FROM employees
GROUP BY Age_Group
ORDER BY Attrition_Percentage DESC;
```

Result:

Age_Group	Total_Employees	Employees_Left	Attrition_Percentage
25-34	596	118	19.79866
Under 25	521	96	18.42610
35-44	263	18	6.84411
45-54	90	5	5.55556

8. Age Distribution

```sql
SELECT
    CASE
        WHEN Sum_of_Age < 25 THEN 'Under 25'
        WHEN Sum_of_Age BETWEEN 25 AND 34 THEN '25-34'
        WHEN Sum_of_Age BETWEEN 35 AND 44 THEN '35-44'
        WHEN Sum_of_Age BETWEEN 45 AND 54 THEN '45-54'
        ELSE '55+'
    END AS Age_Group,
    COUNT(*) AS Employee_Count
FROM employees
GROUP BY Age_Group
ORDER BY Age_Group
```

Result:

Age_Group	Employee_Count
25-34	596
35-44	263
45-54	90
Under 25	521

9. Marital Status Distribution

```sql
SELECT MaritalStatus, COUNT(*) AS Employee_Count
FROM employees
GROUP BY MaritalStatus
```

Result:

MaritalStatus	Employee_Count
Married	624
Single	549
Divorced	297

10. Ethnicity Distribution

```sql
SELECT Ethnicity, COUNT(*) AS Employee_Count
FROM employees
GROUP BY Ethnicity
```

Result:

Ethnicity	Employee_Count
White	860
Asian or Asian American	113
Black or African American	207
Mixed or multiple ethnic groups	198
Native Hawaiian	26
Other	16
American Indian or Alaska Native	50

11. Gender Distribution

```sql
SELECT Gender, COUNT(*) AS Employee_Count
FROM employees
GROUP BY Gender
```

Result:

Gender	Employee_Count
Female	675
Male	651
Other	144

12. Average Salary by Education

```sql
UPDATE employees
SET Salary = REPLACE(Salary, '$', '');

UPDATE employees
SET Salary = REPLACE(Salary, ',', '');

SELECT edu_level,
       COUNT(*) AS Employee_Count,
       ROUND(AVG(Salary), 2) AS Avg_Salary,
       MIN(Salary) AS Min_Salary,
       MAX(Salary) AS Max_Salary
FROM employees
GROUP BY edu_level
ORDER BY Avg_Salary DESC;
```

Result:

edu_level	Employee_Count	Avg_Salary	Min_Salary	Max_Salary
Doctorate	48	154268.79	102059	93181
Masters	398	117641.06	100576	99936
Bachelors	572	115405.43	100586	99203
High School	282	105180.54	100314	99602
No Formal Qualifications	170	94983.48	100818	98693

13. Overtime Distribution

```sql
SELECT OverTime, COUNT(*) AS Employee_Count
FROM employees
GROUP BY OverTime;
```

Result:

OverTime	Employee_Count
Yes	416
No	1054

Employees Demographics

1. Age Distribution

```sql
SELECT
    CASE
        WHEN Sum_of_Age < 25 THEN 'Under 25'
        WHEN Sum_of_Age BETWEEN 25 AND 34 THEN '25-34'
        WHEN Sum_of_Age BETWEEN 35 AND 44 THEN '35-44'
        WHEN Sum_of_Age BETWEEN 45 AND 54 THEN '45-54'
        ELSE '55+'
    END AS Age_Group,
    COUNT(*) AS Employee_Count
FROM employees
GROUP BY Age_Group
ORDER BY Age_Group
```

Result :

Age_Group	Employee_Count
25-34	596
35-44	263
45-54	90
Under 25	521

2. Marital Status Distribution

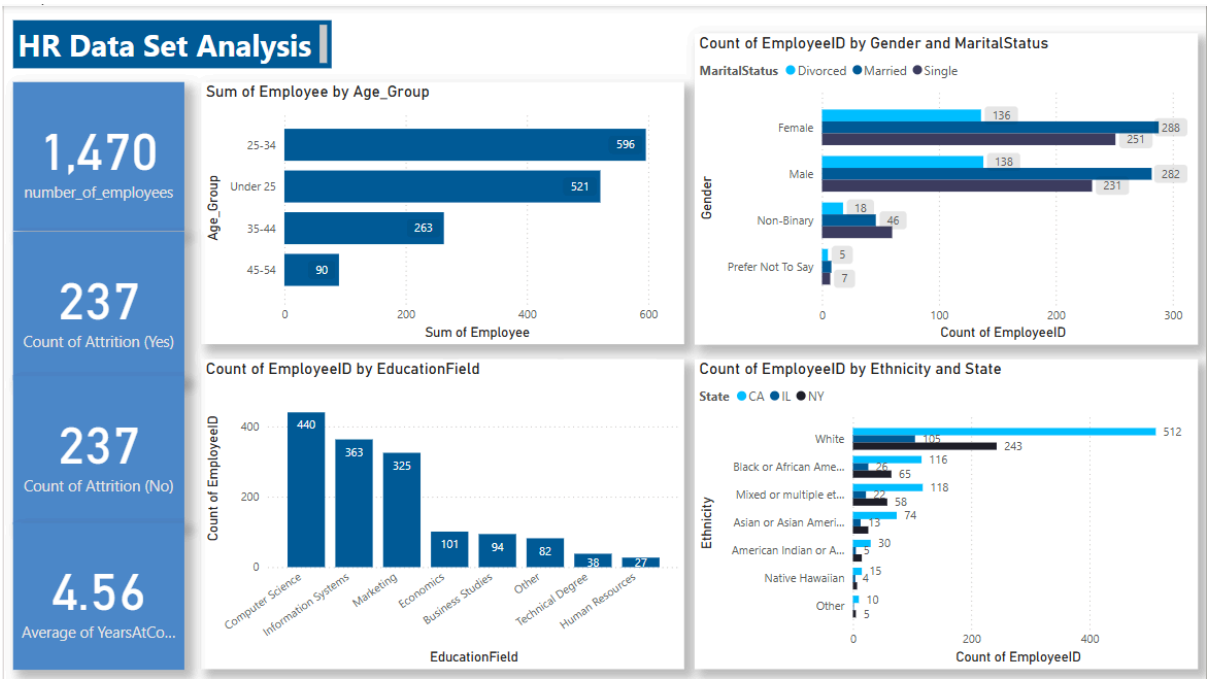
```
1 •   SELECT MaritalStatus, COUNT(*) AS Employee_Count
2     FROM employees
3     GROUP BY MaritalStatus
```

Result :

MaritalStatus	Employee_Count
Married	624
Single	549
Divorced	297

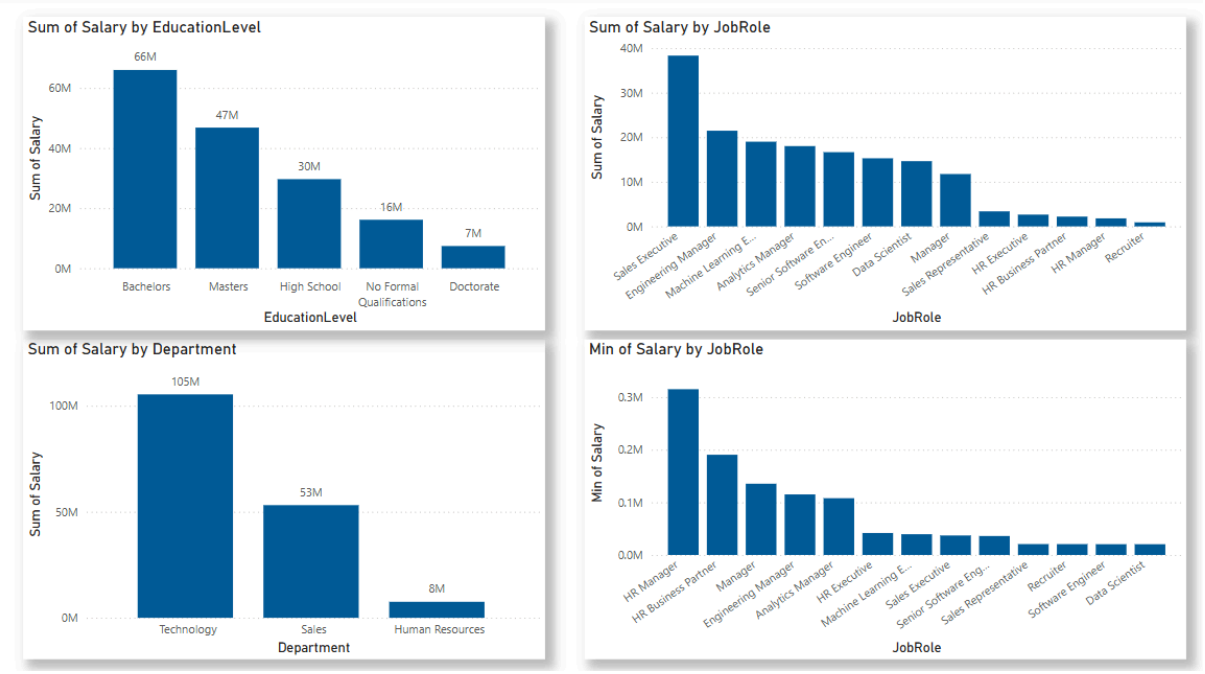
Power BI Analysis

1. Demographics Insights

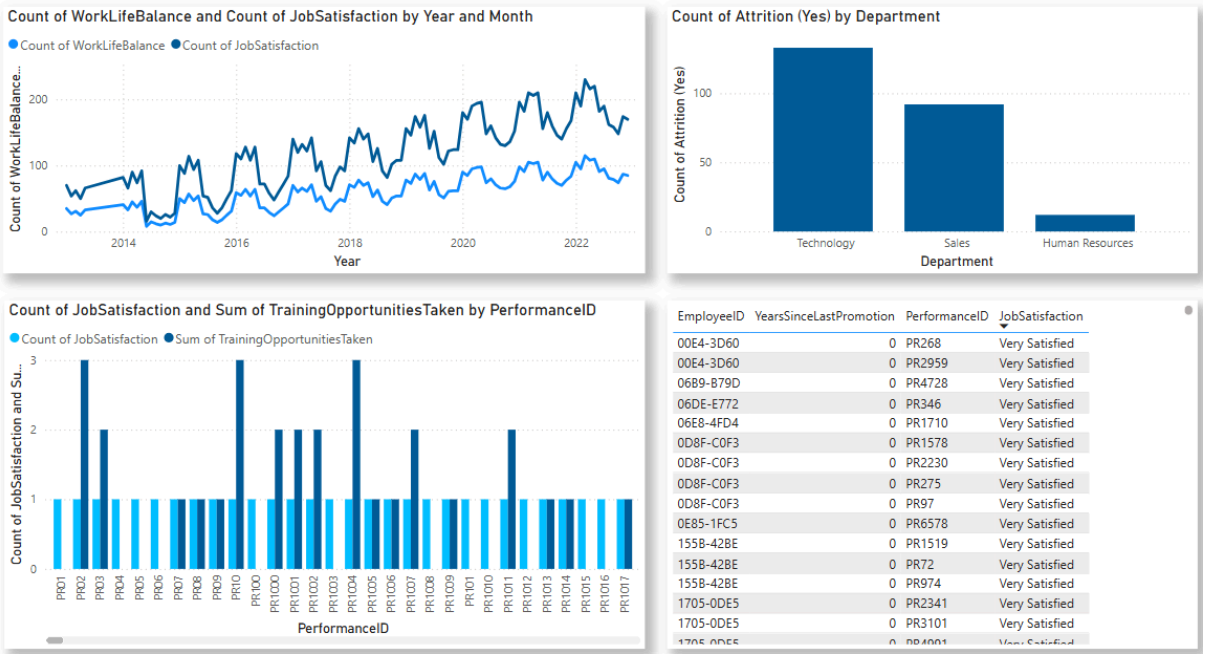


2. Education and Department Insights

3. Salary Insights



4. Job Satisfaction and Work Life Balance



EmployeeID	YearsSinceLastPromotion	PerformanceID	JobSatisfaction
00E4-3D60	0	PR268	Very Satisfied
00E4-3D60	0	PR2959	Very Satisfied
06B9-B79D	0	PR4728	Very Satisfied
06DE-E772	0	PR346	Very Satisfied
06E8-4FD4	0	PR1710	Very Satisfied
0D8F-C0F3	0	PR1578	Very Satisfied
0D8F-C0F3	0	PR2230	Very Satisfied
0D8F-C0F3	0	PR275	Very Satisfied
0D8F-C0F3	0	PR97	Very Satisfied
0E85-1FC5	0	PR6578	Very Satisfied
155B-42BE	0	PR1519	Very Satisfied
155B-42BE	0	PR72	Very Satisfied
155B-42BE	0	PR974	Very Satisfied
1705-0DE5	0	PR2341	Very Satisfied
1705-0DE5	0	PR3101	Very Satisfied