



HR Project

CA12_DAT1_G1

HR Project

CA12_DAT1_G1

| Team Members | ID |
|-----------------------------|-----------|
| Eslam Gamal Abdelrahman | 21093859 |
| Salma Ayman Mohamed | 21083449 |
| Reham Ehab Hassan | 21089282 |
| Fatma ELzahraa Ahmed Salama | 21060326 |
| Naglaa Mohamed Mahmoud | 21079509 |

Supervised By:

Dr: Anas Isamil

Table of Contents:

1- Introduction

| | |
|--|--|
| 1.0 Project Introduction and overview..... | |
| 1.1 Problem Statement..... | |
| 1.2 Project Objectives..... | |
| 1.3 Questions Statement..... | |

2- Data Sources & Description

| | |
|--------------------------------------|--|
| 2.0 Overview of the dataset..... | |
| 2.1 Columns and Data Attributes..... | |
| 2.2 Tools & Technologies Used..... | |

3- Data cleaning & pre-processing

| | |
|---------------------------------------|--|
| 3.0 Introduction..... | |
| 3.1 Handling Missing Values..... | |
| 3.2 Removing Duplicates..... | |
| 3.3 Check wrong spellings..... | |
| 3.4 Data Preparation & Modelling..... | |

4- Analysis Summary

| | |
|-------------------------------|--|
| 4.0 Attrition Analysis..... | |
| 4.1 Salary Analysis..... | |
| 4.2 Performance analysis..... | |

INTRODUCTION & OVERVIEW:

1- Introduction

HR Analysis is the process of using employee data to gain insights that help organizations make better decisions about their people. It involves collecting, organizing, and analyzing data related to employee **performance, satisfaction, hiring, training, and turnover** to understand what drives success within the workforce.

Rather than relying on assumptions or general observations, HR Analysis allows HR teams to use real data to answer important questions such as:

- Why are employees leaving?
- What factors contribute to high performance?
- How satisfied are employees with their jobs?
- Where should we invest in training and development?

By turning workforce data into meaningful insights, HR Analysis helps create a stronger, more productive, and more engaged team.

2- Problem Statement

Every organization has its stars - those dedicated, talented individuals who drive the company forward. But what happens when those stars fade away?

Employee turnover is like a missing piece in a puzzle that can disrupt the whole picture. While companies pour energy into making the workplace a great place to be, the “why” behind employees leaving often remains a mystery. This project is all about solving that mystery. By examining the story behind past employees’ journeys, we’ll uncover the hidden factors that lead to departure. Through this analysis, we can predict which employees are at risk of leaving and understand the reasons that might drive them away. With these insights, companies can not only hold on to their talented workforce but also create a culture where employees feel valued, supported, and excited to stay for the long run.

3- The Goal of this Project:

The goal of this project is to analyze historical employee data to identify the key factors that contribute to employee attrition, and to develop predictive models that can help organizations understand who is at risk of leaving. By uncovering patterns and trends, the project aims to provide actionable insights that can help companies:

1. Identify Causes of Attrition: Discover the underlying reasons behind why employees leave, whether it's related to job satisfaction, work environment, compensation, or other factors.
2. Predict Future Attrition: Use data-driven models to forecast which employees might be at risk of leaving, allowing companies to take preventive measures.
3. Improve Employee Retention: Offer recommendations on how to create a more supportive work environment, improve employee satisfaction, and reduce turnover rates.
4. Optimize HR Strategies: Enable HR departments to make more informed, data-driven decisions when it comes to employee engagement, retention, and recruitment.

Ultimately, the goal is to help organizations retain valuable employees, reduce the costs and disruptions of turnover, and improve overall employee satisfaction and engagement.

Questions

- What is the total number of employees in the dataset?
- How many employees have left the company (attrition rate)?
- What is the distribution of attrition across different departments?
- What is the average length of service for employees who left versus those who stayed?
- Is there a relationship between attrition and the number of years since the last promotion?
- Does a higher number of training opportunities correlate with lower attrition?
- What is the attrition rate by education level or field of study?
- Which job roles have the highest attrition rates?
- Is there a relationship between job satisfaction and attrition?
- Do employees with higher performance ratings have a lower attrition rate?
- What is the attrition rate based on years in the current role?
- Does age correlate with attrition? Are younger employees more likely to leave?
- Is there a difference in attrition rates between males and females?
- What is the attrition rate by marital status (e.g., single, married, divorced)?
- How does attrition change based on different salary levels?
- Is there a correlation between attrition and percentage of salary increase?

Data Sources & Description

The project is built on five primary datasets, each providing crucial information for HR Analysis:

- **Employees Dataset**

Employee 7.csv (Rows: 1470 & Columns: 23)

The dataset contains the following columns:

1. **EmployeeID**: Unique identifier for each employee.
2. **FirstName**: Employee's first name.
3. **LastName**: Employee's last name.
4. **Gender**: Gender of the employee (Male, Female, Non-Binary).
5. **Age**: Age of the employee.
6. **BusinessTravel**: Frequency of business travel (No Travel, Some Travel, Frequent Travel).
7. **Department**: Department where the employee works (Sales, Technology, Human Resources).
8. **DistanceFromHome (KM)**: Distance from home to the workplace in kilometers.
9. **State**: State where the employee resides.
10. **Ethnicity**: Ethnicity of the employee (White, Black or African American, Asian or Asian American, Mixed or Multiple Ethnic Groups, Other).
11. **Education**: Education level of the employee (High School, Technical Degree, Bachelor's Degree, Master's Degree, Doctorate).
12. **EducationField**: Field of education (e.g., Computer Science, Marketing, Information Systems).
13. **JobRole**: Job role of the employee (e.g., Sales Executive, Software Engineer, Data Scientist).
14. **MaritalStatus**: Marital status of the employee (Single, Married, Divorced).
15. **Salary**: Annual salary of the employee.
16. **StockOptionLevel**: Level of stock options granted to the employee.
17. **OverTime**: Whether the employee works overtime (Yes, No).
18. **HireDate**: Date when the employee was hired.
19. **Attrition**: Whether the employee has left the company (Yes, No).
20. **YearsAtCompany**: Number of years the employee has been with the company.
21. **YearsInMostRecentRole**: Number of years the employee has been in their most recent role.
22. **YearsSinceLastPromotion**: Number of years since the employee's last promotion.
23. **YearsWithCurrManager**: Number of years the employee has been with their current manager.

- **Education Level Dataset**

EducationLevel - Copy 5.csv(Rows: 5 & Columns: 2)

The dataset contains the following columns:

1. **EducationLevel ID:** Unique identifier for each education level.
2. **EducationLevel:** Description of the education level.

Education Levels

- **1:** No Formal Qualifications
- **2:** High School
- **3:** Bachelor's
- **4:** Master's
- **5:** Doctorate

This file provides a list of education levels with their corresponding IDs, which can be used for categorizing and analyzing employee education data.

- **PerformanceRating Dataset**

PerformanceRating 7.csv(Rows: 6709 & Columns: 11)

The dataset contains performance ratings of employees. The columns in the dataset are:

- **PerformanceID:** Unique identifier for each performance review.
- **EmployeeID:** Unique identifier for each employee.
- **ReviewDate:** Date of the performance review.
- **EnvironmentSatisfaction:** Rating of the employee's satisfaction with their work environment.
- **JobSatisfaction:** Rating of the employee's satisfaction with their job.
- **RelationshipSatisfaction:** Rating of the employee's satisfaction with their relationships at work.
- **TrainingOpportunitiesWithinYear:** Number of training opportunities provided to the employee within a year.
- **TrainingOpportunitiesTaken:** Number of training opportunities taken by the employee within a year.
- **WorkLifeBalance:** Rating of the employee's work-life balance.
- **SelfRating:** Employee's self-rating of their performance.
- **ManagerRating:** Manager's rating of the employee's performance.

- **SatisfiedLevel Dataset**

SatisfiedLevel 4.csv : (Rows: 5 & Columns: 2)

The dataset contains different levels of satisfaction with their corresponding IDs. The columns in the dataset are:

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

Satisfaction Levels

1. **Very Dissatisfied**
2. **Dissatisfied**
3. **Neutral**
4. **Satisfied**
5. **Very Satisfied**

These files provide a list of performance and satisfaction levels with their corresponding IDs, which can be used for categorizing and analyzing employee performance and satisfaction data.

Rating Level Dataset

Rating Level 3.csv : (Rows: 5 & Columns: 2)

The dataset contains different levels of performance ratings with their corresponding IDs. The columns in the dataset are:

- **RatingID:** Unique identifier for each rating level.
- **RatingLevel:** Description of the rating level.

Rating Levels

1. **Unacceptable**
2. **Needs Improvement**
3. **Meets Expectation**
4. **Exceeds Expectation**
5. **Above and Beyond**

Tools & Technologies Used

Python

Power bi

Data cleaning & Processing:

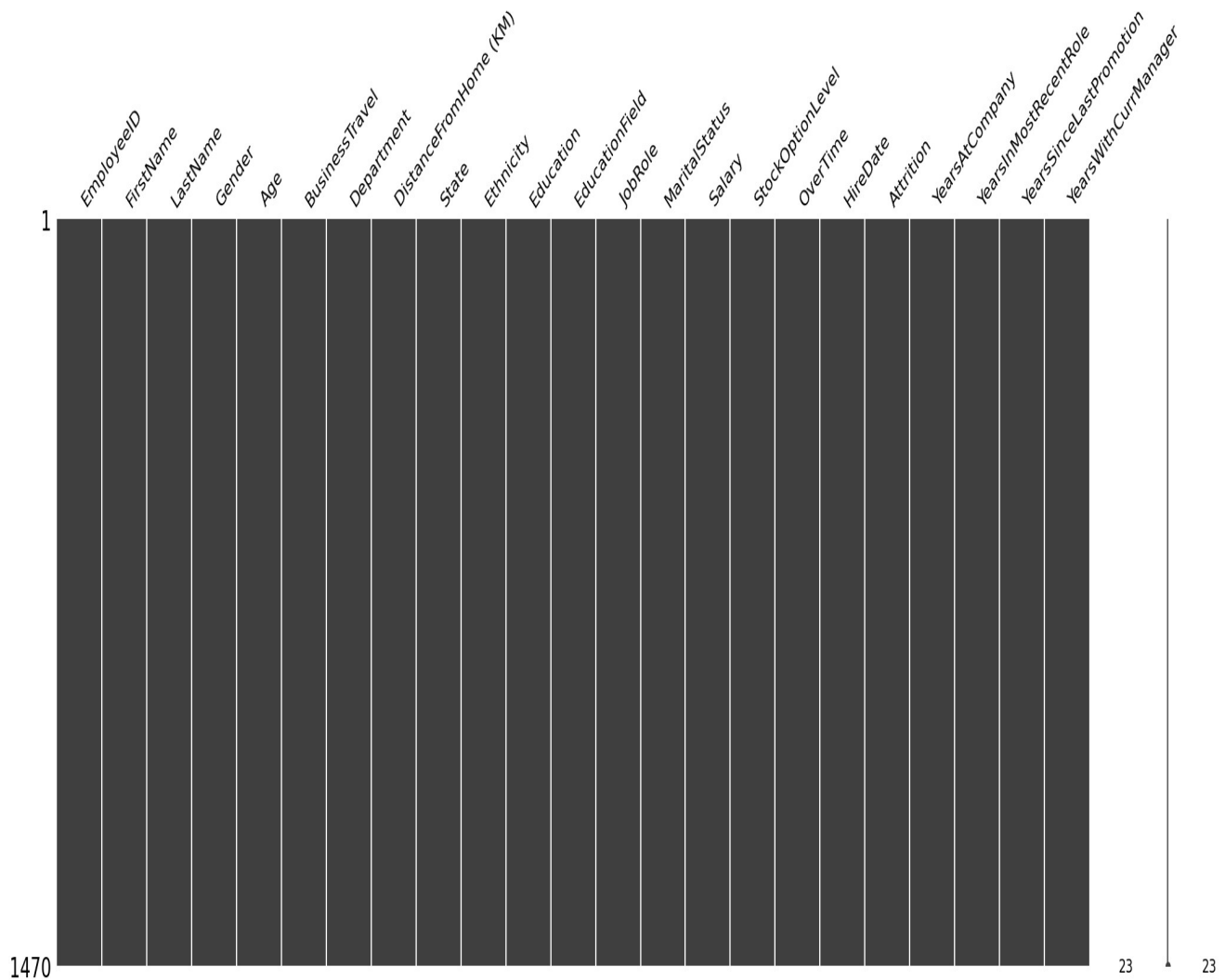
- **Introduction**

Data cleaning is the process of fixing or removing incorrect, messy, or incomplete data to make it accurate and reliable for analysis.

Data processing is the broader process of taking raw data and transforming it into meaningful information. This includes steps like collecting, cleaning, organizing, analyzing, and visualizing the data to make it useful for decision-making.

- **Handling Missing Values**

Check Missing values: No missing values were found in all columns



- Removing Duplicates

```
[27] data.duplicated()
✓ 0.0s
... 0      False
     1      False
     2      False
     3      False
     4      False
     ...
    1465    False
    1466    False
    1467    False
    1468    False
    1469    False
    Length: 1470, dtype: bool
```

```
[28] data.duplicated().unique()
✓ 0.0s
... array([False])
```

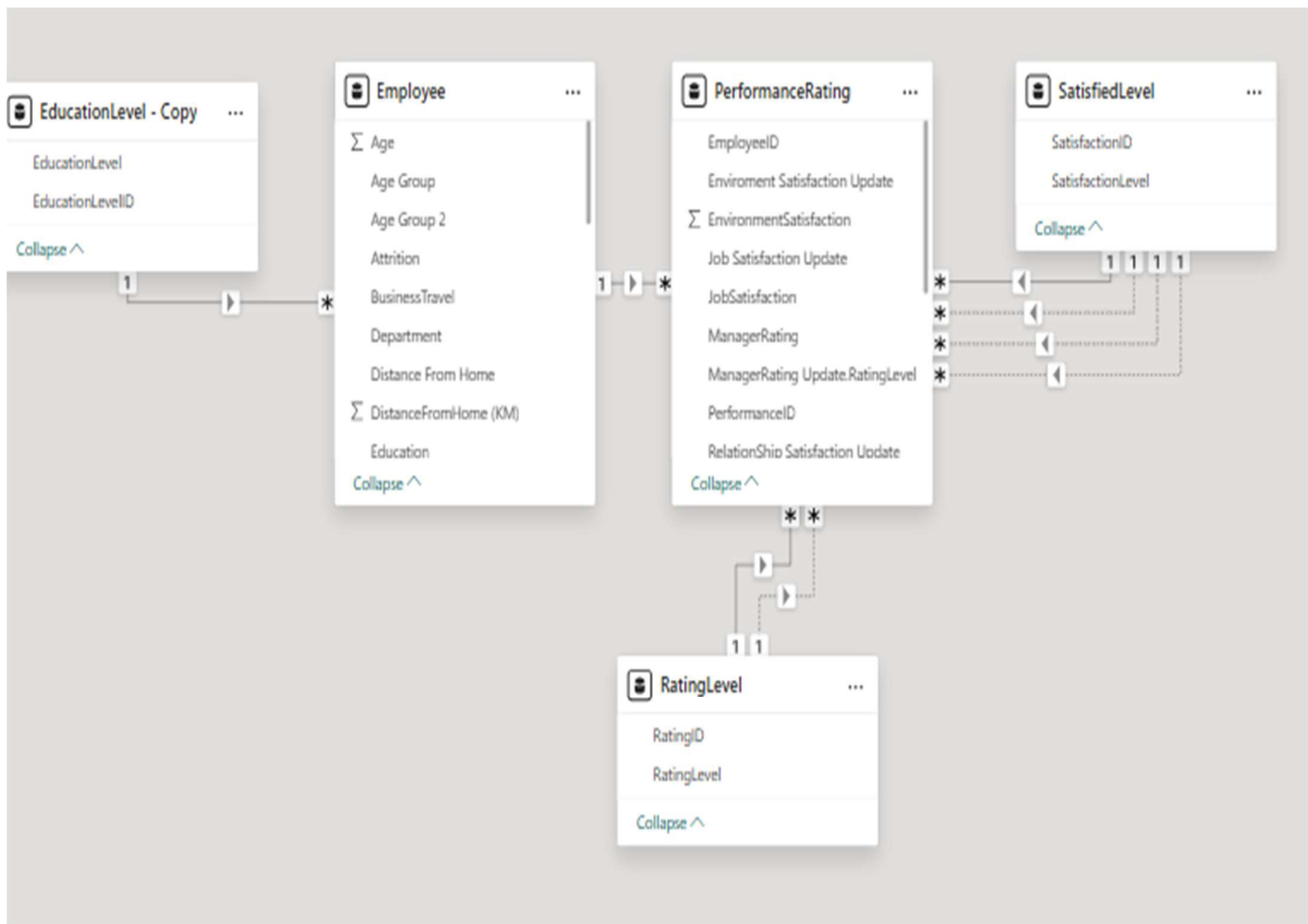
- Check wrong spellings
- Check wrong spellings: Found space after “No Travel “ in Business Travel column.
Found duplicated “Marketing” education field due to extra space.

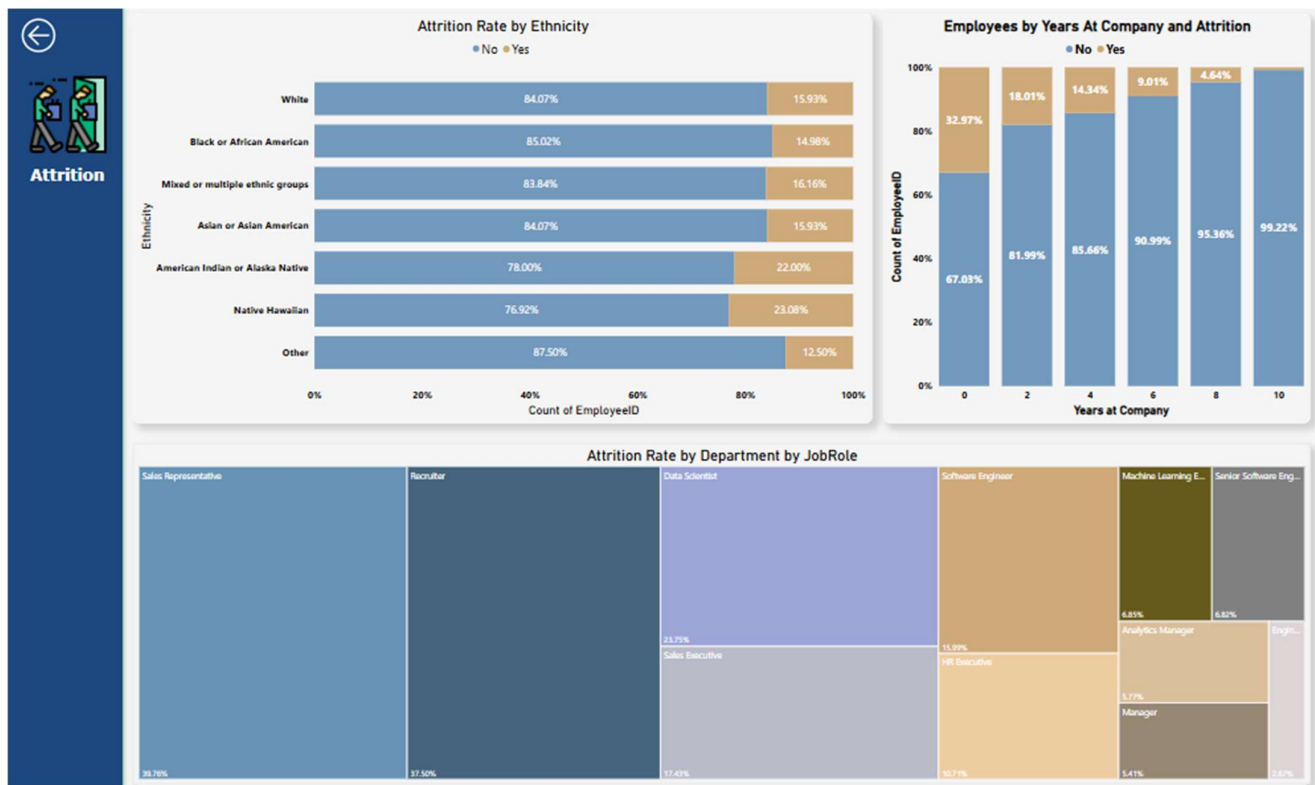
```
▶ data['EducationField'].unique()
[11] ✓ 0.0s
... array(['Marketing', 'Marketing ', 'Computer Science', 'Technical Degree',
          'Information Systems', 'Other', 'Economics', 'Human Resources',
          'Business Studies'], dtype=object)
```

```
data['EducationField'] = data['EducationField'].replace('Marketing ', 'Marketing')
✓ 0.0s

data['EducationField'].unique()
✓ 0.0s
array(['Marketing', 'Computer Science', 'Technical Degree',
      'Information Systems', 'Other', 'Economics', 'Human Resources',
      'Business Studies'], dtype=object)
```

- **Data Preparation & Modelling**





Attrition Analysis Summary

Key Influencers of Attrition

- Young employees with ≤ 1 year of experience are 3.11x more likely to leave.
- Working OverTime increases attrition likelihood by 2.5x.
- Lower salaries, being Single, and Frequent Business Travel also contribute significantly to attrition.

Attrition by Department

- The Technology department has the highest attrition rate (56.12% of total leavers).
- Sales comes next with 38.82%, and Human Resources the least (5.06%).
- Internal department attrition rates: Sales: 20.58%, HR: 19.05%, Technology: 13.85%.

Attrition by Marital Status

- 54.01% of those who left were Single.
- Married employees are less likely to leave (33.33%).
- Divorced employees represent 12.66% of attrition cases.

Attrition by Gender

- Females represent the highest proportion of leavers (48.10%).
- Males: 43.88%, Non-Binary: 8.02%.

Attrition by Age Group

- Young employees (64.57%) dominate the workforce and represent the highest attrition rates.
- Mid-Aged: 34.56%, Senior: Only 0.87%, with 0% attrition.

Attrition by Tenure (Years at Company)

- Highest attrition occurs during the first year (18.18%).
- Attrition sharply decreases with time and is nearly 0% after year 6.

Business Travel Impact

- 65.82% of frequent travelers left the company.
- Those who traveled 'rarely' or 'never' had lower attrition rates.

Attrition by Ethnicity

- Small differences across ethnic groups.
- Higher attrition seen among: Native Hawaiian: 22.03%, American Indian or Alaska Native: 23.08%, Multiple Ethnicities: 18.18%.

Attrition by Job Role

- Highest attrition observed among: Sales Representatives, Laboratory Technicians, Machine Learning Engineers.

Attrition in HR department:

- The distribution of attrition rate by job role in young employees was 47% for HR recruiters and 9.52% for HR executives, while in the mid-aged employees, it was 14.29% for HR recruiters versus 14.29% for HR executives.

Young:

- Young HR salary is the domain cause of their leave, as when it equal or less than \$49.606K with increase the attrition rate 7.60 times.
- The 47.06% of the leaved HR recruiter worked over time with average salary \$33.11K.
- On the other hand the leaved young HR executives were mixed groups with average salary \$44.63K or native Hawaiian with average salary \$179.56K
- The young leaved HR executives native Hawaiian worked less than 1 year and some travel.

Mid_Aged:

- Mid_Aged HR also salary and over time work the main reasons for attrition.
- The whole leaved mid_aged HR recruiters were white with average salary \$48.74K.
- The whole leaved mid_aged HR executives were white with average salary \$201.56K.

Attrition in Sales department:

Young:

- The main factors affect attrition rate between young in sales department are over time work, salary, single, Asian or American Asian, and frequent travel.
- The average years between young employees is equal or less than 2 years.
- Over time affect increase the attrition rate to 2.38x, as 51.35% of the total sales attrition rate work over time.
- Most leaved young employees are sales representative with average salary \$36.16K and they do some or frequent travel. 51.61% of them work over time.
- 50% of the total left young sales executive employees worked over time.
- Almost all left young sales executive employees do some or frequent travel.

Mid_Aged:

- The attrition rate of mid-aged sales were 12.98% for sales executive, 10.53% for sales representative, then 3.85% for manager.
- The most left mid_aged employees were sales executive because 66.67% of them worked over time and do some business travels. on the other hand 26.47% of total left ones frequent travel for business.
- The second left job role was sales representative, as the all of them frequent travel with average salary \$40.88K.
- The main reason for mid-aged sales manager to leave is business travels .
- The average years between mid_aged employees before left is equal or less than 2 years.

Attrition in Technology department:**Young:**

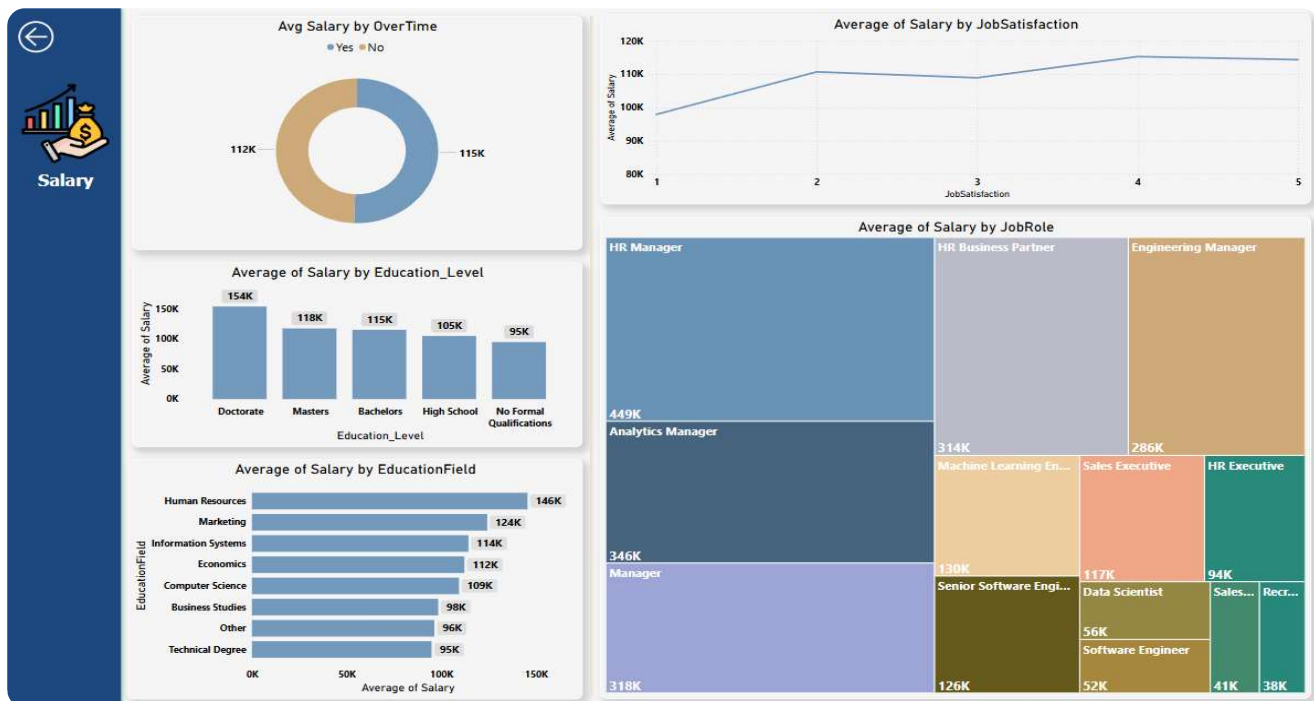
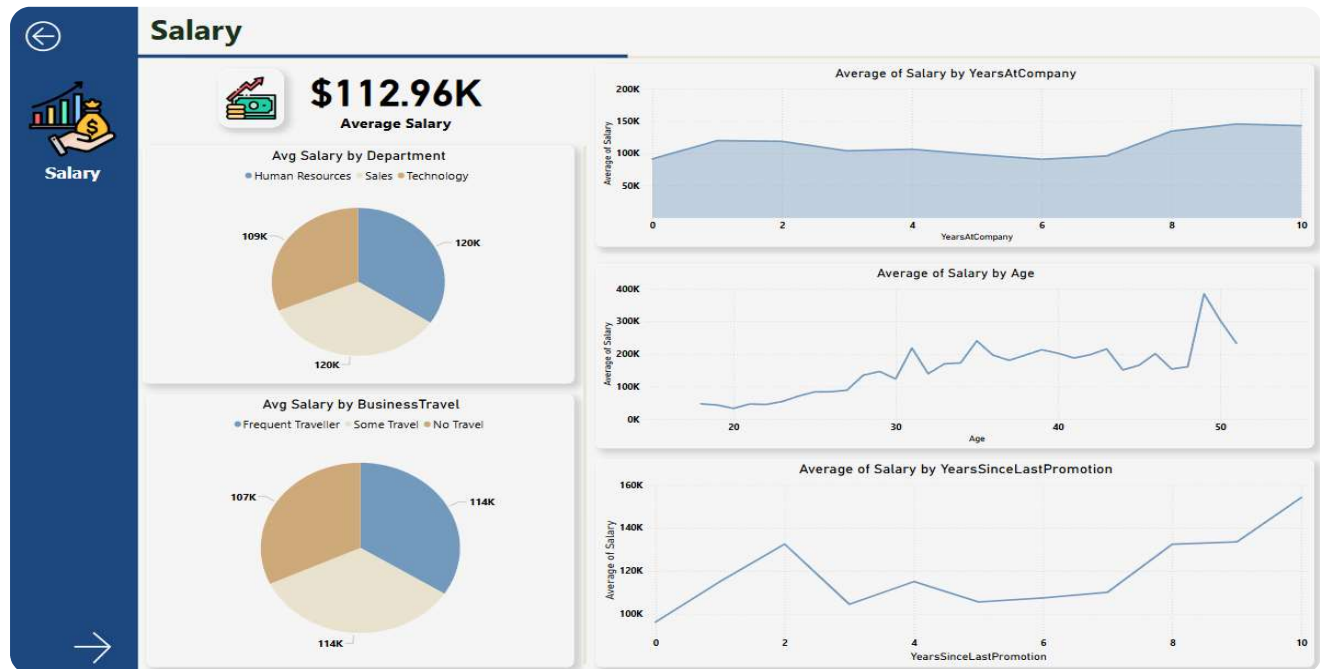
- The main factors affect attrition rate between young in technology department are over time work, salary, single, and frequent travel.
- The average years between young employees before left is equal or less than 1 years.
- Business travel, overtime, and a low average salary (\$48,580) are the factors that have caused a high attrition rate among young people in the data scientist field, as 49.06% of them work over time, 71.7% do some business travel and 24.53% of them travel frequently.
- 69.05% of the left-wing young software engineers worked overtime. Almost all of them had business travel with an average salary (\$37.75K), as 59.52% of them had some travel and 30.95% travel frequently.
- All young machine learning engineers left because they had business travel, as 77.78% of them had some travel and 22.22% travel frequently, and 44.44% of them worked overtime.
- Approximately 70% of the total left young senior software engineers had business travel, as 50% of them had some travel and 33.33% travel frequently. Working overtime also affect the attrition as 33.33% of them worked overtime.
- Business travel and working overtime resulted in the attrition of the young Analytics manager.

Mid_Aged:

- The highest attrition rate among mid-aged people in the technology department was in the data science branch, with 11.84%, followed by software engineers at 7.14%, software engineers at 6.38%, analytics managers at 5.56%, engineer managers at 4.88%, and machine learning engineers at 1.79%.
- All left mid-aged data scientists were doing business travel, 55.56% of them working overtime, and having a low average salary (\$65,92K).
- 80% of the left mid-aged software engineers worked overtime. All of them had business travel with an average salary (\$56.59K), as 60% of them had some travel and 40% travel frequently.
- The main factor for mid-aged senior software engineers to leave is business travel. 66.67% had some travel meanwhile 33.33% frequent travel.
- Overtime working was the primary factor for the whole analytics manager to leave the company, then business travel, as all ones left had over time and half of them had business travels.
- The primary reason for engineer managers to quit is having business travel, and sometimes they had to work overtime. All of them had business travels and half of them worked overtime.

The total quit mid-aged machine learning engineers had to travel frequently.

Salary Dashboard



Salary Analysis

This report presents a detailed analysis of employee salaries across multiple factors such as department, job role, education, experience, business travel, and job satisfaction. The main goal is to find out what factors have the biggest impact on salary, and to help managers make fair and smart decisions about how much employees should be paid.

- The average salary across all employees is **\$112.96K**.
- By Department:
 - **Sales** and **Human Resources** have the highest average salaries (both **\$120K**) as these departments are often core to a company's growth.
 - **Technology** has a slightly lower average salary; this could be because it includes more junior or support roles that usually pay less.
- By Job Role:
 - HR Manager has the highest average salary at **\$449K**.
 - Analytics Manager: **\$346K**, Manager: **\$318K**.
 - Lowest paid roles: Sales Representative (**\$41K**) and Recruiter (**\$38K**) as these are often junior jobs.
- By Years Since Last Promotion:
 - Salaries increase with time since last promotion.
 - Peak average salary is at 10 years since last promotion (~**\$154K**).
- By Years at Company:
 - Salary rises with more years at the company.
 - Highest averages are seen between **8 to 10 years** of service.
- **By Age:**
 - Salaries increase with age; older usually means more experience
 - A noticeable peak is around **age 50**; Many employees at this age are likely in senior or leadership positions.
- **By Job Satisfaction:**

Higher satisfaction levels correlate with higher salaries; this might be because satisfied employees often perform better and stay longer so the company rewards them with higher pay.
- **By Overtime:**

Employees who work overtime earn a bit more (**\$115K**) than those who don't (**\$112K**).

- **By Business Travel:**
 - Employees who travel sometimes or often earn more (**\$114K**)
 - Employees who don't travel earn less (**\$107K**)
- **By Education Level:**
 - Doctorate holders earn the most (**\$150K**).
 - Followed by Masters and Bachelors (**both \$120K**).
 - Lowest: No Formal Qualification (**\$90K**).
- **By Education Field:**
 - Top earners studied Human Resources (**\$146K**), Marketing (**\$124K**), and Information Systems (**\$114K**).
 - Lower salaries in Business Studies, Technical Degrees, and Other fields (**~\$95K–\$98K**)