

**HR Analytics Project**

**Supervised by**

**DR: Ahmed Samir**

**Egypt (2024- 2025)**

**Team Members:**

**Abeer Magdy Saad (Team Leader)**

**Malek Kamal El\_Sayed**

**Rahma Maher Sayed**

**Introduction**

Human Resource Analytics (**HR Analytics**) is a modern field that contributes to improving organizational performance and enabling data-driven decision-making. Today, companies rely on data analysis to gain valuable insights into employees, such as performance, turnover rates, and employee satisfaction, which helps enhance the work environment and increase productivity.

This project aims to study human resource data using various data analysis tools, including **Excel, Power BI, Tableau, SQL, and Python**, to extract valuable insights that enhance HR management efficiency. By utilizing these tools, we can effectively process data and conduct advanced analyses that help understand the factors influencing employee performance and enable smarter decision-making.

**Abstract**

This project focuses on analyzing human resource data to explore the factors that impact job performance and support data-driven decision-making. Data is collected, cleaned, and analyzed using a diverse set of specialized tools, namely **Excel, Power BI, Tableau, SQL, and Python**, each offering a unique aspect of analysis, such as data processing, interactive reporting, and extracting key insights.

Various data analysis techniques are applied to derive insights that assist in improving recruitment processes, reducing employee turnover rates, and enhancing workforce efficiency. By combining statistical analysis with interactive visualizations, we provide data-driven recommendations to optimize organizational performance and improve the overall work environment.

**Table of contents**

Table of contents:

**Chapter One (Introduction):**

Introduction................................................................................................ 4

Problem definition …….…….………………….............……………………. 4

Project Objectives …….…….………………….............………………...…... 5

Significance of the Project.......................................................................... 5

Tools & Technologies Used ....................................................................... 5

Dataset Overview ...................................................................................... 6

**Chapter Two (Technical):**

Introduction …………..………..…………..………………............…………..9

Handling Missing Values………….………..…..……………………….....… 9

Checking for missing values using SQL ……...……………….…………… 9

Checking for missing values using Pandas …………..…………..…….......10

Checking for full duplicate rows using Pandas ...………………….....…… 11

Data Type Corrections …………..………..……………………..………...… 13

Merging Datasets …………..………………..…..………………………….... 14

**Chapter Theree (Analysis Questions):**

Analysis Questions …………..……….…..………………............…………..16

**Chapter Four (Forcasting Analysis)**

Introduction …………..………..…………..………………............…………..52

**Chapter Five (Visualisation)**

Introduction …………..………..…………..………………............…………..63

HR Summary Dashboard ……….………..…..……………………….......… 63

Employees Analysis Dashboard ...........……...……………….…………… 64

Salary Analysis Dashboard …….......................……..…………..…….......65

Attrition Analysis Dashboard ...…………………................................…… 66

Employee Performance Analysis Dashboard …………..….....………...… 67

**Chapter Six (Recommendation)**

Recommendations ...................…………………................................…… 68

**Chapter 1**

**Introduction :-**

In today’s data-driven world, companies rely heavily on analytics to make informed decisions—especially in the Human Resources (HR) department, where understanding employee behavior, performance, and trends is essential.

This graduation project focuses on **HR Data Analysis**, with the aim of transforming raw HR data into meaningful insights that support effective decision-making processes.

The project utilizes **real-world HR data** and applies various data analysis techniques using a combination of professional tools including **Excel**, **Power BI**, **Tableau**, **SQL**, and **Python**. These tools were used collaboratively to explore workforce dynamics, clean and structure data, uncover hidden patterns, and visualize key HR metrics in an intuitive and professional way.

By leveraging these technologies, the project aims to deliver practical, interactive dashboards and evidence-based insights that can assist HR departments in improving employee management and organizational strategies.

**Problem Definition :-**

Human Resources departments often face challenges in making strategic decisions due to the lack of actionable insights from raw data. While HR departments usually collect vast amounts of information about employees, such as attendance, salaries, performance, and job roles, this data often remains underutilized.

Without proper analysis, organizations may:

* Fail to detect high turnover risks.
* Overlook patterns in absenteeism or overtime.
* Struggle to allocate salaries fairly or optimize employee performance.
* Miss opportunities to improve employee satisfaction or productivity.

This project aims to bridge this gap by analyzing HR data to uncover valuable insights that can support better decision-making and improve overall HR operations.

**Project Objectives :-**

This project aims to:

* Analyze employee data to detect trends in attendance, salary, and job satisfaction.
* Identify factors contributing to employee turnover.
* Discover patterns in overtime and absenteeism.
* Build interactive dashboards using Tableau and Power BI.
* Provide data-driven recommendations for HR decision-making.

**Significance of the Project :-**

Human Resources departments often rely on assumptions when managing employee affairs. By transforming HR data into meaningful visual insights, this project helps shift decision-making from intuition to data-driven strategies.

It supports better planning, fairness in compensation, and enhanced employee satisfaction.

**Tools & Technologies Used :-**

To ensure accurate, insightful, and professional HR data analysis, this project leverages a combination of industry-standard tools and technologies. Each tool played a key role in the data analysis pipeline, from data cleaning and querying to advanced visualization.

**🔹 Microsoft Excel**

Excel was used for initial data exploration, cleaning, and formatting. It played a vital role in identifying missing values, correcting data types, and performing preliminary calculations to prepare the dataset for deeper analysis.

**🔹 SQL (Structured Query Language)**

SQL was used to manage and manipulate structured datasets efficiently. It enabled the project team to extract specific records, filter large tables, perform joins between multiple sheets, and prepare meaningful subsets of data for analysis.

**🔹 Python**

Python was used for advanced data manipulation and analysis using popular libraries such as:

* pandas – for data wrangling and transformation
* numpy – for numerical operations
* matplotlib and seaborn – for data visualization (when needed)
* Custom scripts – for automating repetitive tasks and generating insights

Python helped in handling complex logic, creating derived fields, and validating trends statistically.

**🔹 Power BI**

Power BI was used to build dynamic and interactive dashboards. Its features allowed the creation of slicers, filters, and multi-page reports, enabling users to explore the data with ease and extract meaningful information on demand.

**🔹 Tableau**

Tableau was used to design visually compelling dashboards and data stories. Its drag-and-drop interface and advanced visualization capabilities made it possible to represent HR metrics in a clear, attractive, and decision-oriented manner.

**Employee Dataset Overview (23 columns|1470** **row)**

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Description** | **Sample Values / Notes** |
| EmployeeID | Unique identifier for each employee | E.g., 3012-1A41 |
| FirstName | First name of the employee | Alix, Annabela |
| LastName | Last name of the employee | Simco, Berrie |
| Gender | Employee gender | Male, Female |
| Age | Employee’s age | E.g., 25, 35, 42 |
| BusinessTravel | Frequency of business travel | Some Travel,Non-Travel |
| Department | Department the employee belongs to | HR, IT, Sales |
| DistanceFromHome | Distance from employee’s home to work (in KM) | E.g., 5, 15, 30 |
| State | State where the employee resides | CA, NY, IL |
| Ethnicity | Ethnic background of the employee | E.g., Arab, African, Asian |
| Education | Education level (coded as numbers or labels) | 1 to 5 or e.g., High School, Master’s |
| EducationField | The employee’s field of study | Marketing, Computer Science |
| JobRole | The position or role the employee holds | Data Scientist, Manager, Recruiter |
| MaritalStatus | Marital status of the employee | Single, Married, Divorced |
| Salary | Monthly salary of the employee | E.g., 8000, 15000 |
| StockOptionLevel | Level of stock options offered | 0, 1, 2, 3 |
| OverTime | Whether the employee works overtime | Yes, No |
| HireDate | Date the employee was hired | E.g., 2020-01-01 |
| Attrition | Whether the employee left the company | Yes, No |
| YearsAtCompany | Total number of years at the company | E.g., 3, 10, 15 |
| YearsInMostRecentRole | Years spent in the most recent job role | E.g., 1, 5, 7 |
| YearsSinceLastPromotion | Time since the last promotion | E.g., 0, 2, 6 |
| YearsWithCurrManager | Years under current manager | E.g., 2, 4, 8 |

**Performance Dataset Overview (11 columns|6710** **row)**

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Description** | **Sample Values / Notes** |
| PerformanceID | Unique identifier for each performance review | E.g., PR01, PR02 |
| EmployeeID | Unique identifier for the employee | E.g., B61E-0F26 |
| ReviewDate | Date when the performance review was conducted | E.g., 2020-05-01 |
| EnvironmentSatisfaction | Employee's satisfaction with the work environment | 1,2,3,4,5 |
| JobSatisfaction | Employee's satisfaction with the job role | 1,2,3,4,5 |
| RelationshipSatisfaction | Employee's satisfaction with work relationships | 1,2,3,4,5 |
| TrainingOpportunitiesWithinYear | Whether the employee had training opportunities within the year | Yes, No |
| TrainingOpportunitiesTaken | Whether the employee took the training opportunities | Yes, No |
| WorkLifeBalance | Employee's balance between work and personal life | 1,2,3,4,5 |
| SelfRating | Employee's self-rated performance | E.g., 1 to 5 |
| ManagerRating | Manager's rating of the employee's performance | E.g., 1 to 5 |

**Education Dataset Overview (2 columns|6** **row)**

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Description** | **Sample Values / Notes** |
| EducationLevelID | A unique identifier for each education level. This column contains numbers or unique values used to represent each education level individually. | E.g., 1,2,3,4,5 |
| EducationLevel | Represents the actual education level of the employee. This column contains textual descriptions specifying the type of education attained by the employee. | E.g., High School, Bachelor's Degree, Master's Degree |

**Satisfaction** **Dataset Overview (2 columns|6** **row)**

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Description** | **Sample Values / Notes** |
| SatisfactionID | A unique identifier for each satisfaction level. This column contains numeric or unique values used to represent each satisfaction level individually. | E.g., 1,2,3,4,5 |
| SatisfactionLevel | Represents the satisfaction level of the employee. This column contains a textual description specifying the employee's satisfaction level regarding their job, work environment, or other relevant factors. | E.g., Dissatisfied, Neutral, Satisfiedو Very Satisfied |

**RatingLevel Dataset Overview(2 columns|6** **row)**

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Description** | **Sample Values / Notes** |
| RatingID | A unique identifier for each rating level. This column contains numeric or unique values used to represent each rating level individually. | E.g., 1,2,3,4,5 |
| RatingLevel | Represents the rating level given to the employee. This column contains a textual description that specifies the level of the rating, such as performance rating or evaluation score. | E.g., Meets Expectation, Exceeds Expectation , Unacceptable |

**Chapter 2**

**Data Cleaning & Transformation**

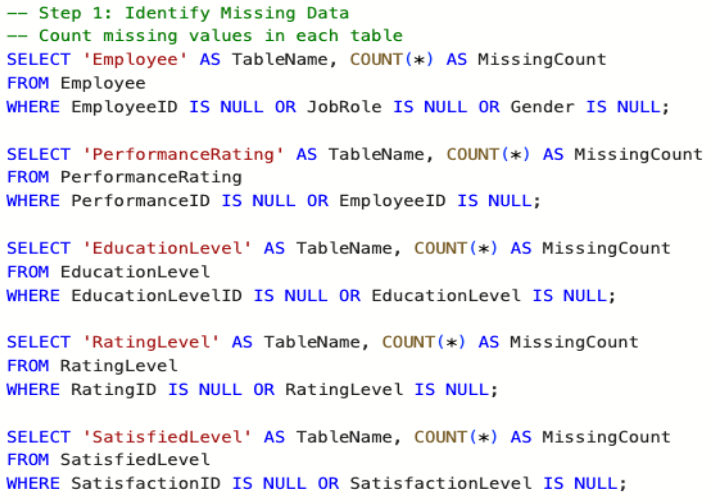
**Introduction :-**

In this chapter, the focus will be on the essential steps of Data Cleaning and Transformation. Data from various sources often contains inconsistencies, missing values, and irrelevant information. Cleaning and transforming this data ensures the integrity of the analysis and enables better insights. This chapter will detail the steps taken to clean and transform the dataset, addressing issues such as missing values, duplicates, and formatting errors.

**2.1. Handling Missing Values:-**

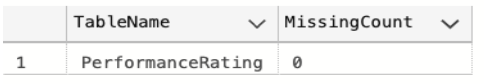
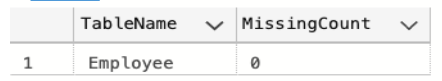
Before proceeding with any data analysis, it was essential to verify whether the dataset contained any missing values that could impact the accuracy of results.

After carefully inspecting the dataset using tools like Excel, SQL, and Python (Pandas), **no missing values were found** in any of the columns. This ensured that the data was complete and ready for further analysis without the need for imputation or deletion.

** Checking for missing values using SQL**

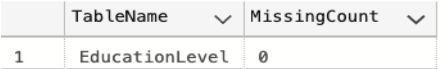
**Employee Table**

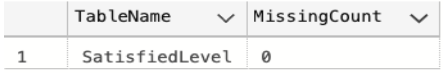
**PerformanceRating Table**



**EducationLevel Table**

**RatingLevel Table**

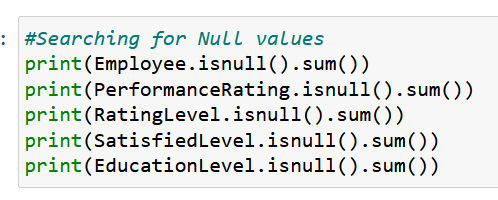




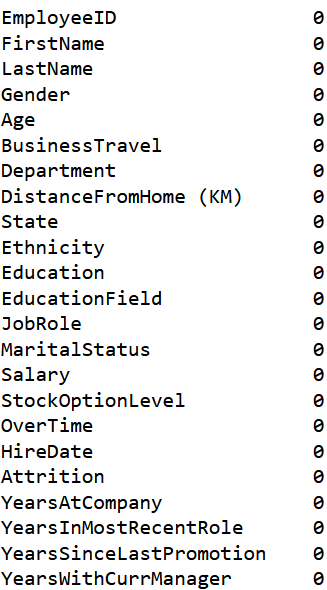
**SatisfiedLevel Table**

**The query shows that there are no missing values in the dataset.**

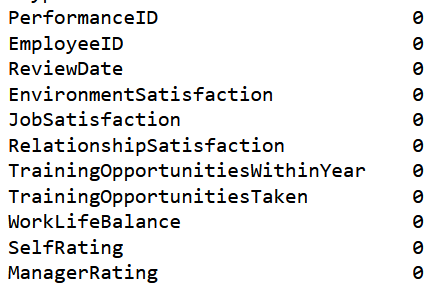
**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_----**

**Checking for missing values using Pandas**

**Employee Table**



**PerformanceRating Table**



**EducationLevel Table**

**RatingLevel Table**



**SatisfiedLevel Table**



**The output confirms that all columns have zero missing entries.**

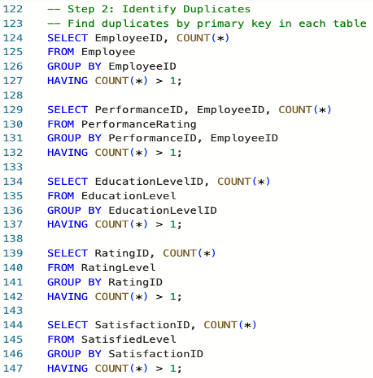
**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_----**

**2.2 Handling Duplicate Records:-**

Before proceeding with data analysis, it was essential to ensure that the dataset did not contain any duplicate records that might bias the results or lead to inaccurate conclusions.

Duplicate records occur when the same row appears more than once in the dataset, either entirely or based on specific key columns. These can result from repeated data entry, system issues, or merging datasets.

To handle this, I used both **SQL** and **Python (Pandas)** to check for and remove any duplicate records.

**Checking for duplicate EmployeeID values using SQL**

****

**PerformanceRating Table**

**Employee Table**

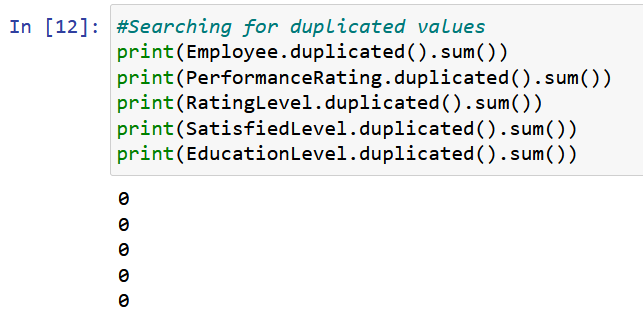
****

**SatisfiedLevel Table**

**RatingLevel Table**

**EducationLevel Table**

**The result showed that no duplicate records were found.**

**Checking for full duplicate rows using Pandas**

**The result indicated that no completely duplicated rows exist in the dataset.**

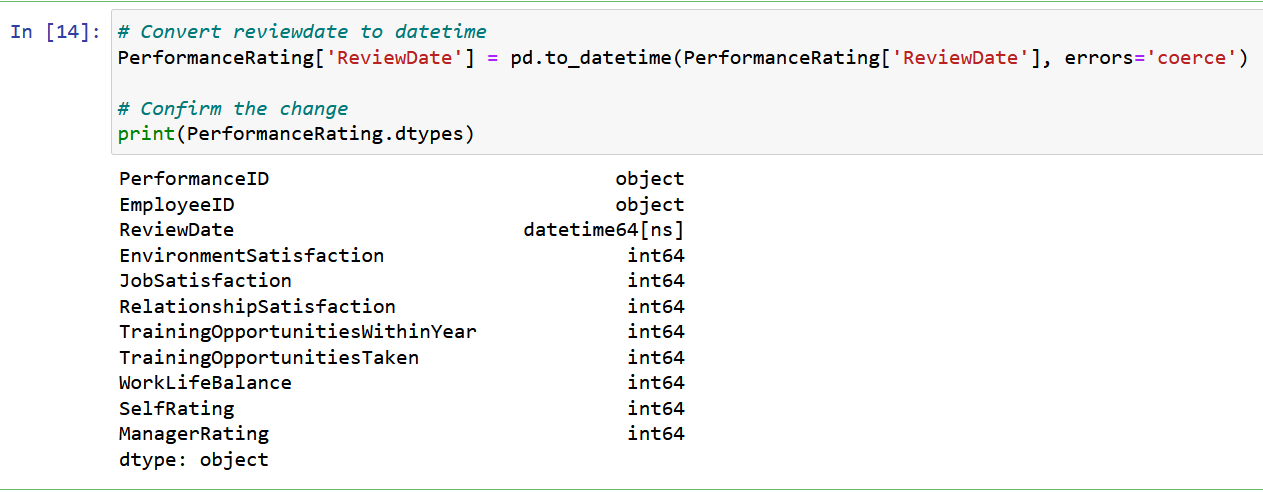
**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_----**

**2.3 Data Type Corrections:-**

As part of the data cleaning process, ensuring that each column has the correct data type is essential for accurate analysis and proper visualization.

In this project, all columns were reviewed to confirm their data types. Most columns had appropriate types such as integers for numerical values and strings for categorical data.

However, the **ReviewDate** column was initially in string (text) format, which would prevent any date-based calculations or filtering. Therefore, it was converted to a **Date/Datetime** format using Python and Excel to allow for accurate analysis, such as calculating the employee's tenure or filtering by hiring year.

**Converting ReviewDate** **to datetime using Python**

**This conversion allowed for accurate time-based analysis, such as determining the number of years an employee has been with the company.**

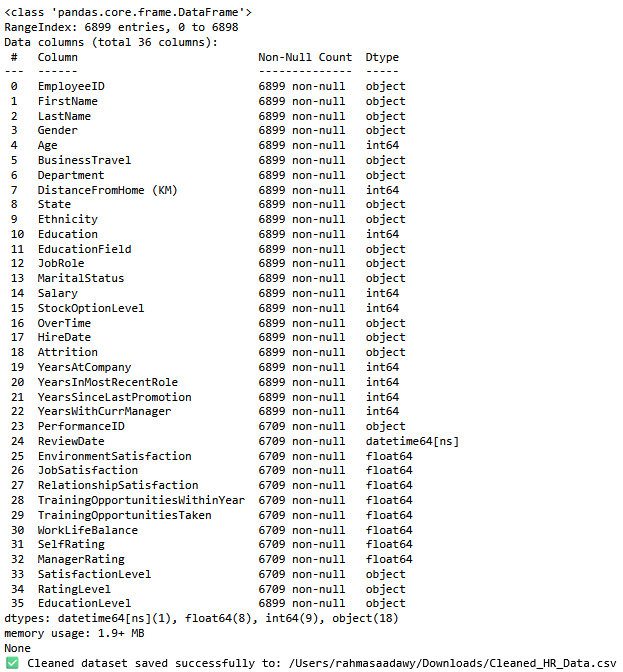
**2.4 Merging Datasets:-**

To perform comprehensive HR analysis, data from multiple sources needed to be combined into a single dataset. This merging process helps in creating a unified view of each employee by integrating various aspects such as personal details, performance reviews, satisfaction levels, and education background.

The following tables were merged:

* **Employee Table** – contains personal and job-related information.
* **Performance Table** – includes satisfaction scores and manager/self-ratings.
* **Education Level Table** – maps education level IDs to their respective labels.
* **Satisfaction Level Table** – provides descriptive levels for satisfaction scores.
* **Rating Table** – defines the meaning of different rating levels.

**Merging Employee and Performance Data using Python**

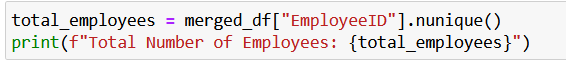


**Analysis Questions**

**Chapter 3**

**1:- Calculate and display the total number of employees**

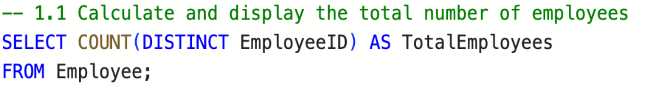
**Solution using Python:-**

****

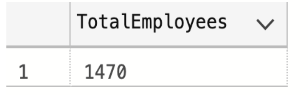
**Output**

****

**Solution using SQL:-**

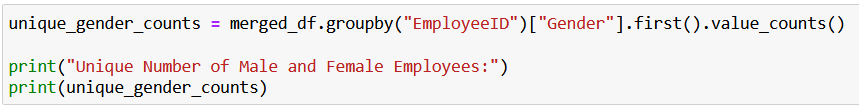


**Output**

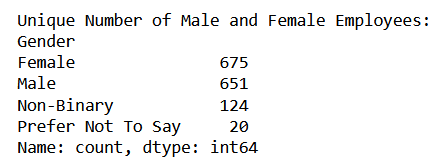


**2:- Count employees by gender**

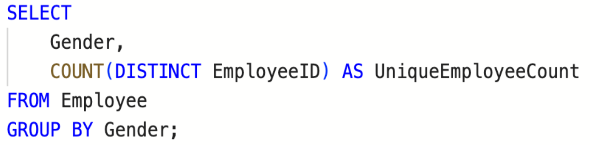
**Solution using Python:-**

****

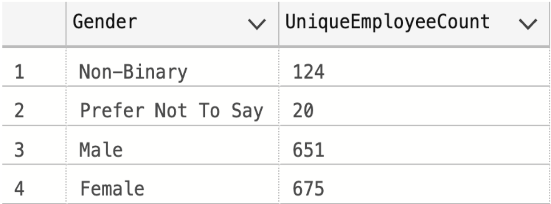
**Output**

****

**Solution using SQL:-**



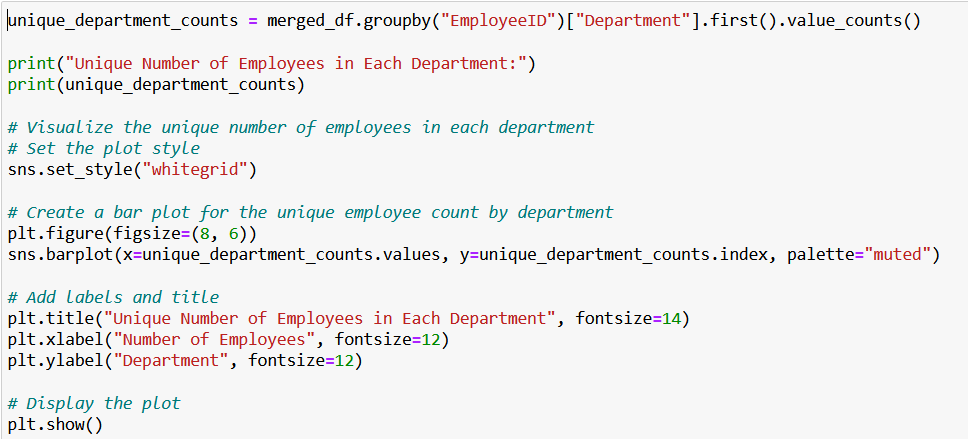
**Output**



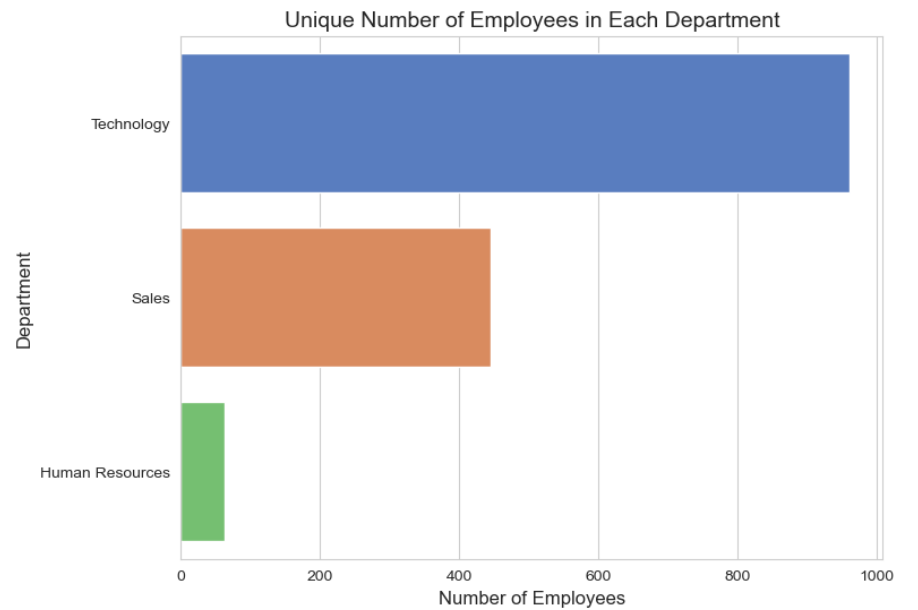
From the analysis, it appears that the number of **males** (651) is slightly lower than the number of **females** (675). As for **Non-Binary** individuals, there are 124, and 20 individuals prefer not to disclose their gender. This data may indicate greater diversity in gender identity within the organization, and it could be useful to explore this further in order to create a more inclusive work environment for everyone.

**3:- Count employees by department**

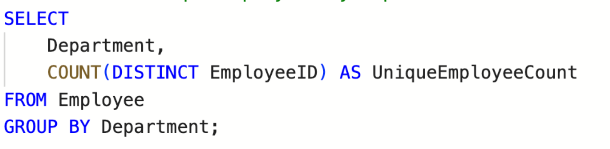
**Solution using Python:-**

****

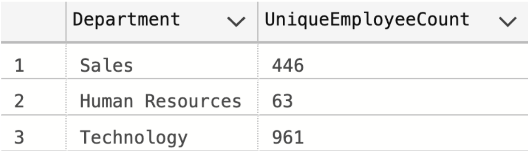
**Output**

****

**Solution using SQL:-**



**Output**



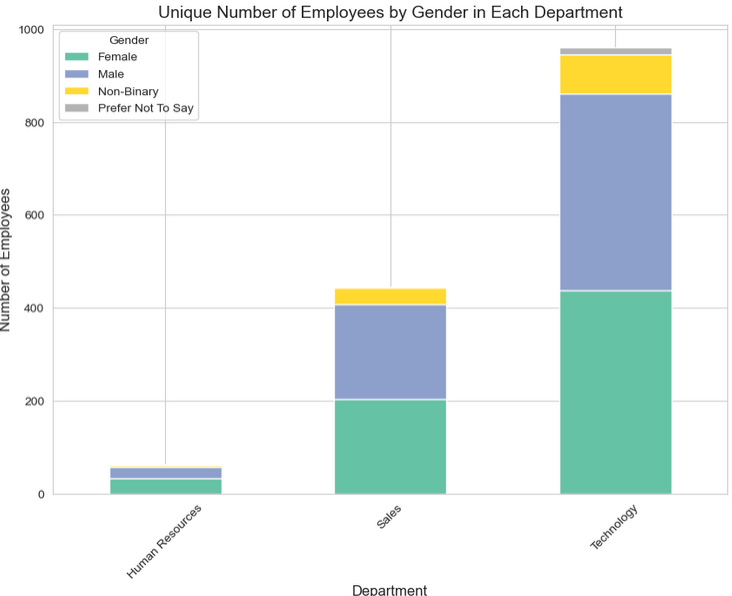
From the analysis, it is evident that the largest number of employees is in the **Technology** department (961 employees), followed by the **Sales** department with 464 employees, while the **HR** department has the smallest number of employees (63 employees). These distributions reflect a larger focus on resources in the technology and sales areas, which may indicate a need for larger teams in these departments to support growth and expansion.

**4:- Count employees by gender within each department**

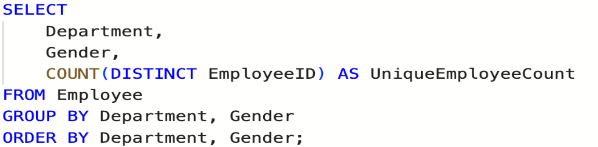
**Solution using Python:-**

****

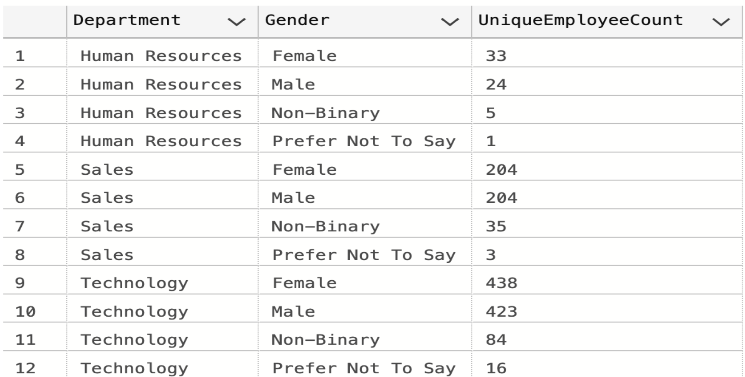
**Output**

****

**Solution using SQL:-**



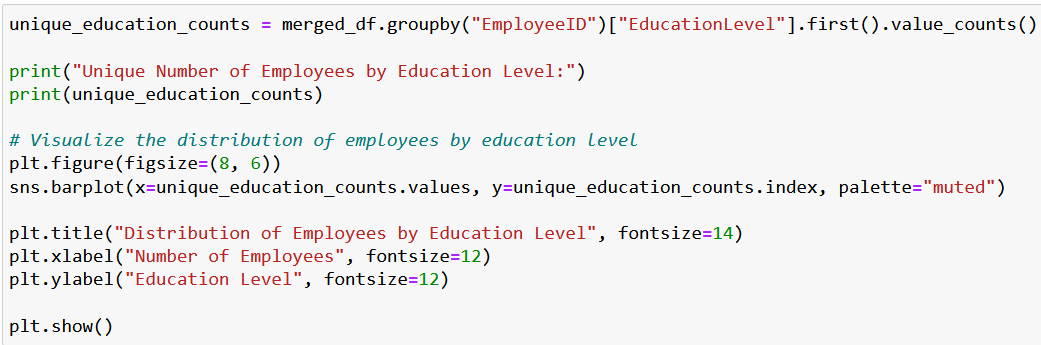
**Output**



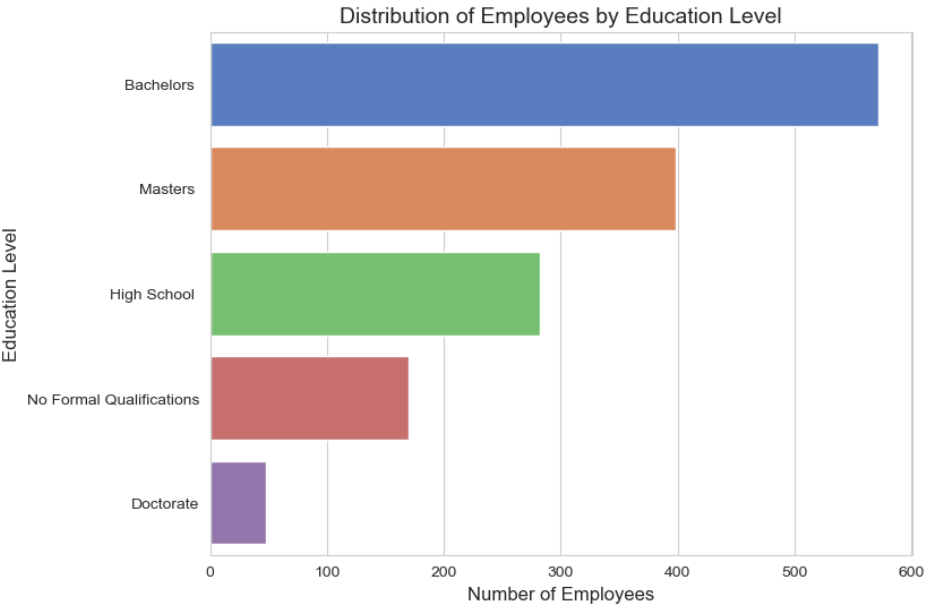
Through the analysis, the number of unique employees by gender within each department was calculated. The results show a diverse distribution of genders across different departments. For example, some departments may have a more balanced gender ratio, while others may be dominated by a specific gender. This data can provide insights into the diversity across departments and help in making strategic decisions to promote diversity and inclusion within the organization.

**5:- Count employees by education level**

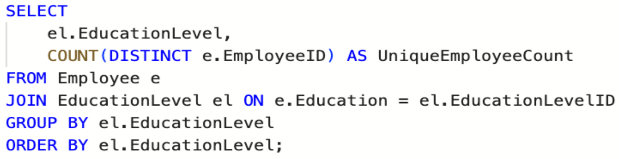
**Solution using Python:-**

****

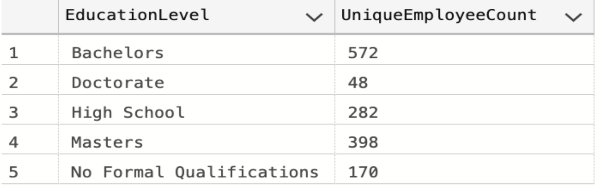
**Output**

****

**Solution using SQL:-**



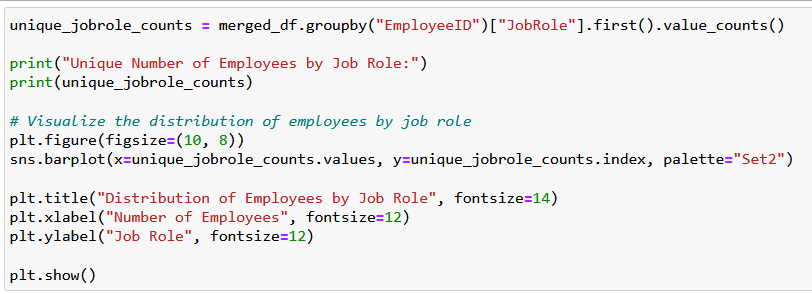
**Output**



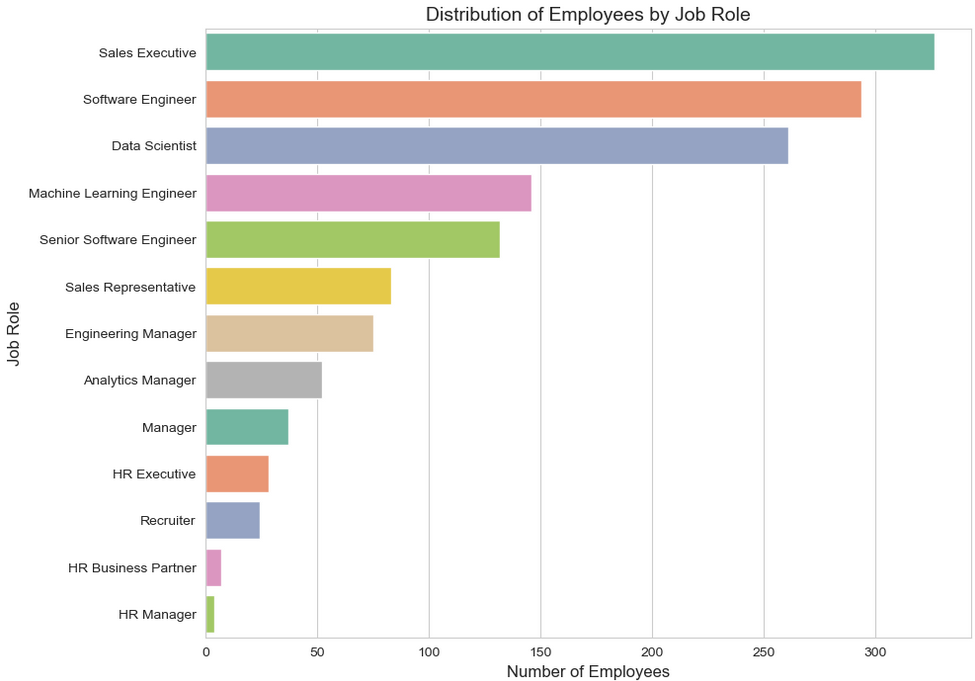
**Through the analysis, the number of unique employees was counted based on their education level. The results show that the largest group of employees holds a Bachelor's degree (572 employees), followed by those with a High School diploma (280 employees), then Master's degree holders (398 employees), and finally Doctorates (48 employees). Additionally, there are 170 employees with Non-formal education. This data indicates that most employees have a university education, while the smaller number with doctorates reflects a relatively low concentration of employees in advanced academic roles.**

**6:- Count employees by job role**

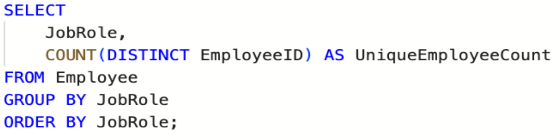
**Solution using Python:-**

****

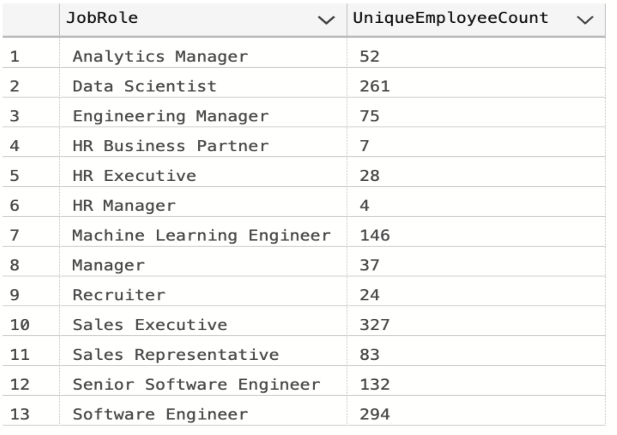
**Output**

****

**Solution using SQL:-**



**Output**



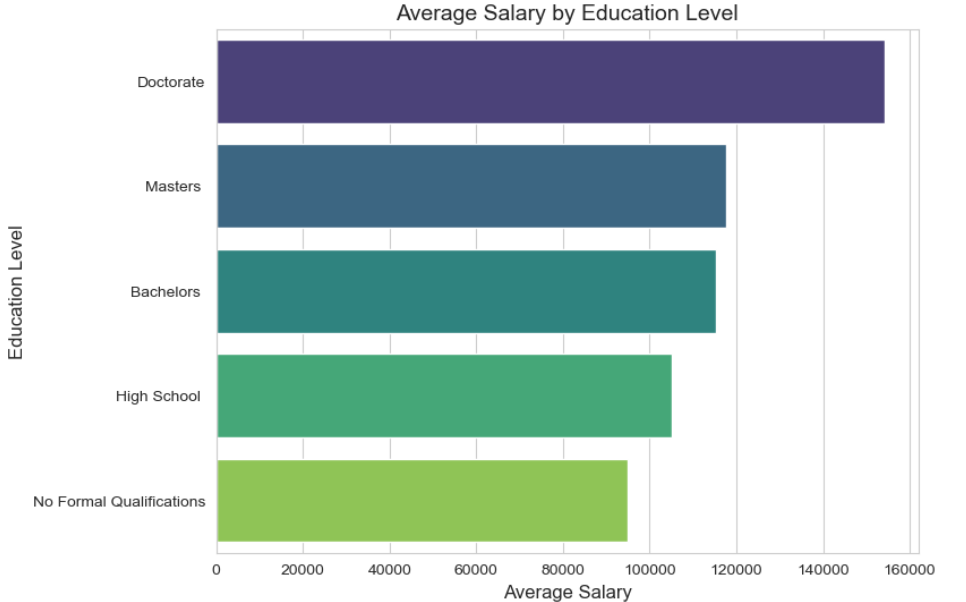
Through the analysis, the number of employees was counted based on their job role. The results show that some roles, such as **HR** (4 employees), have relatively small representation, while others, like **Data Scientist** (261 employees) and **Software Engineer** (294 employees), have a larger presence. The **Sales Executive** role had the highest number of employees (327 employees). This distribution reflects the significant diversity in job roles within the company and indicates a greater need for employees in departments like sales and software compared to other roles.

**7:- How does the average salary vary by education level?**

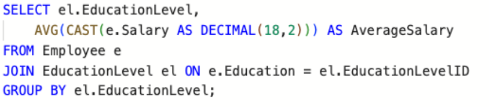
**Solution using Python:-**

****

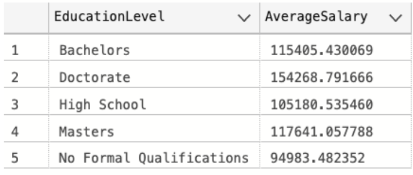
**Output**

****

**Solution using SQL:-**



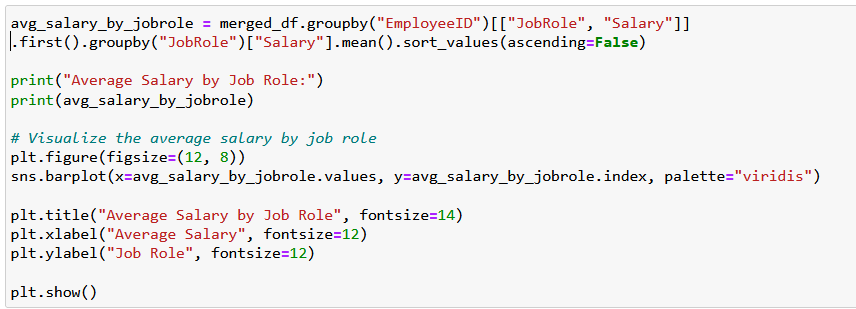
**Output**



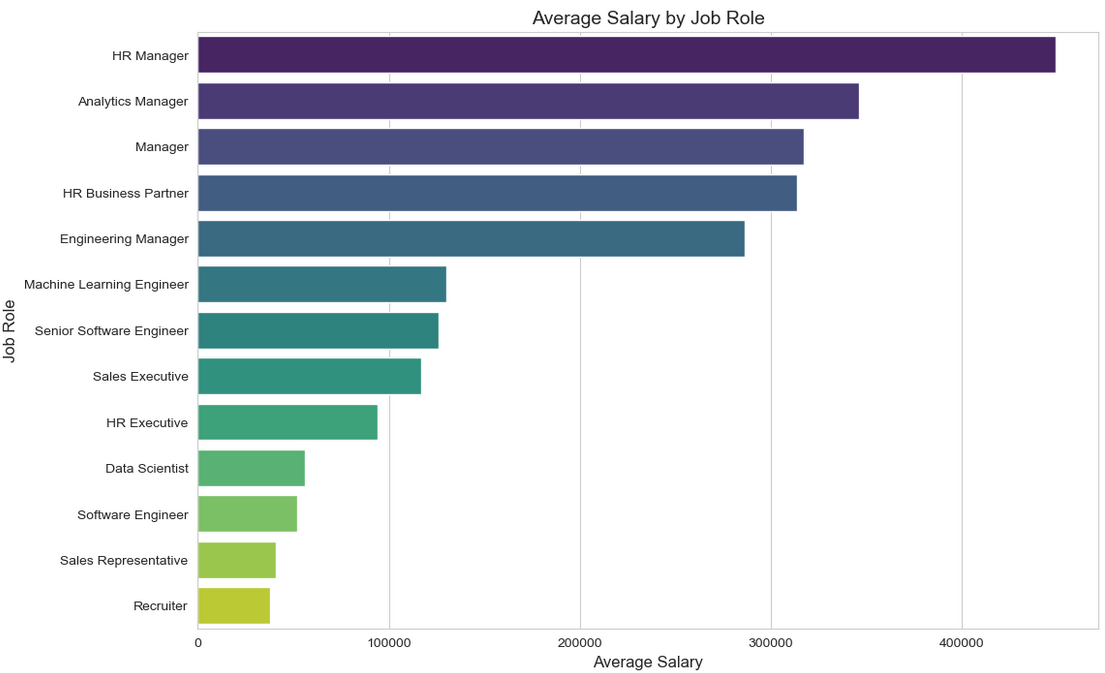
Through the analysis, it was found that **Doctorates** earn the highest average salary, followed by those with a **Master's degree**, then **Bachelor's degree** holders. **High School** employees had the lower average salary, and the **Non-formal** category had the lowest. This indicates that education level has a significant impact on salary, with higher academic qualifications generally leading to higher salaries.

**8:- Calculate the average salary by job role for employees**

**Solution using Python:-**

****

**Output**

****

**Solution using SQL:-**



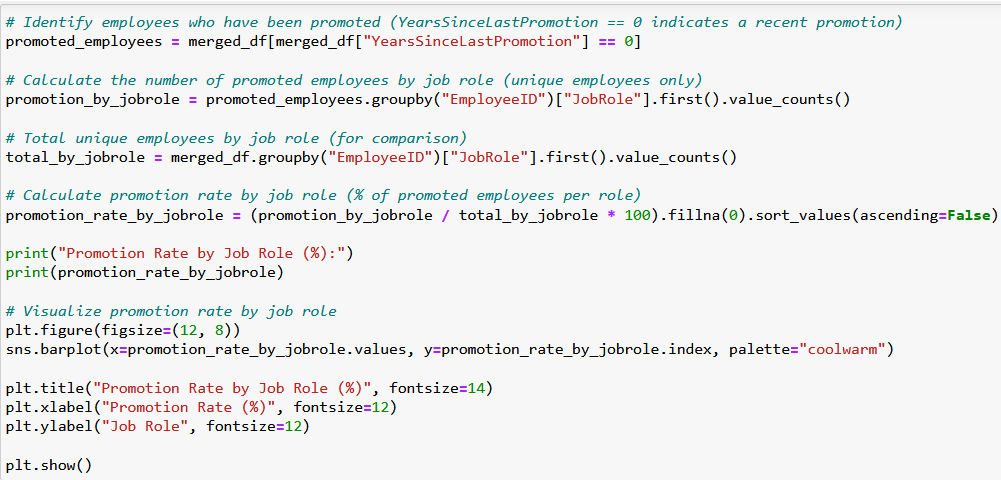
**Output**



Through the analysis, it was found that **HR** is the job role with the highest average salary, while other roles have more varied salary levels. Roles like **Sales Executive** and **Software Engineer** earned comparatively lower average salaries than **HR**, indicating a significant variation in salaries across different job roles within the company.

**9:- What is the salary distribution based on years of experience?**

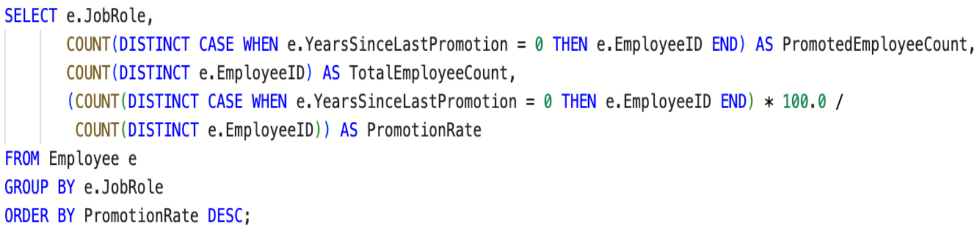
**Solution using Python:-**

****

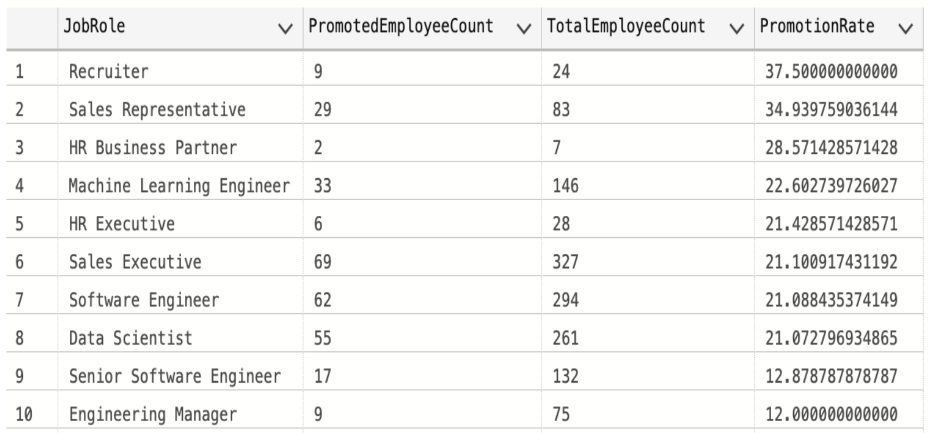
**Output**

****

**Solution using SQL:-**



**Output**

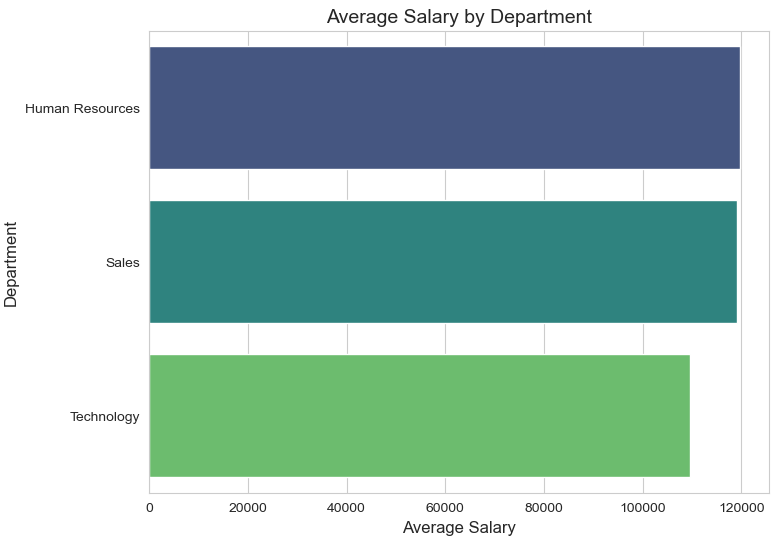


Through the analysis, it was found that salary distribution is highly dependent on **years of experience**. Employees with **less experience** tend to have lower salaries compared to those with **more experience**, with salaries increasing gradually as experience grows. This result confirms that experience is a key factor in determining salary levels within the company.

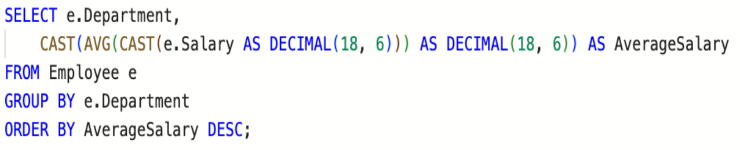
**10:- Calculate the average salary by department for employees Solution using Python:-**

****

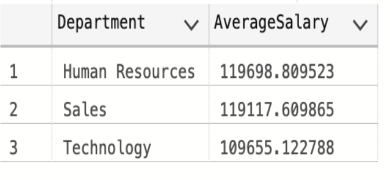
**Output**

****

**Solution using SQL:-**



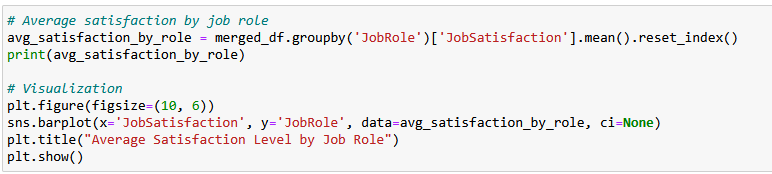
**Output**



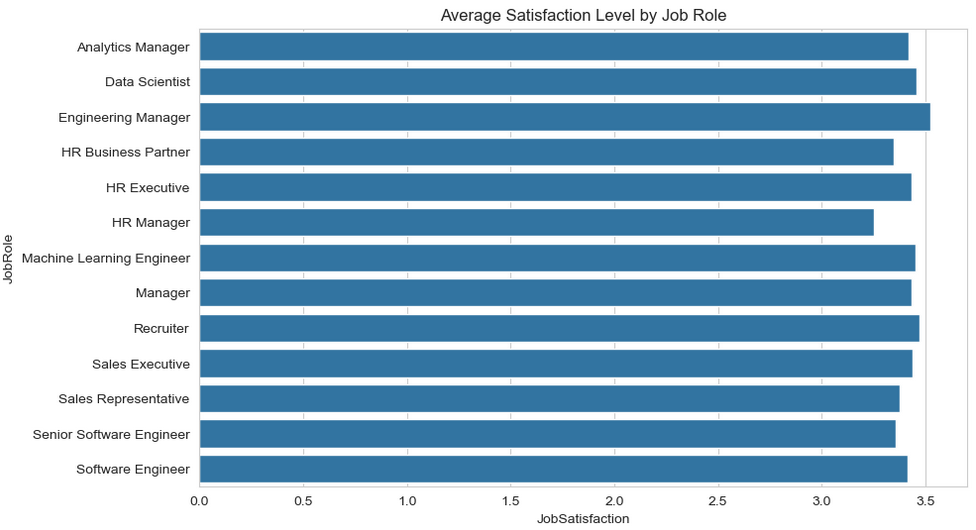
Through the analysis, it was found that salaries across departments are quite similar, with **HR** having the highest average salary, followed closely by **Sales** with only a small difference. After **Sales**, **Technology** followed with a minimal difference. This suggests that salaries are relatively close across most departments, with slight distinctions in average salaries for departments like HR and Sales.

**11:- What is the average satisfaction level across different job roles?**

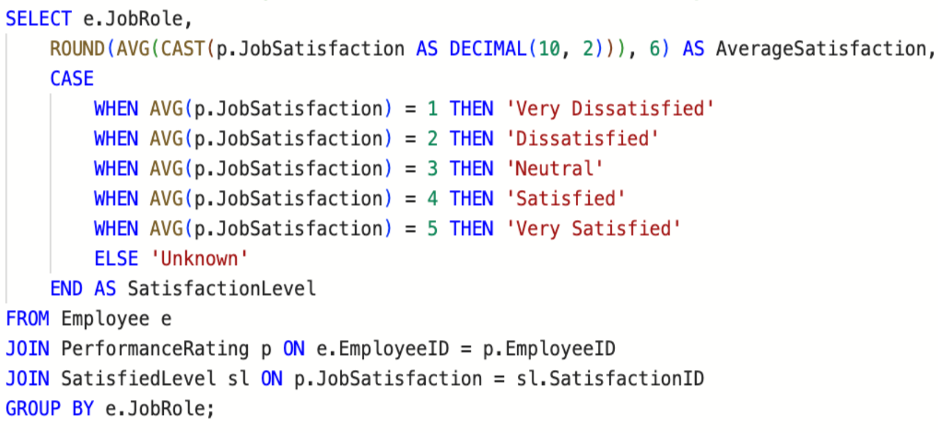
**Solution using Python:-**

****

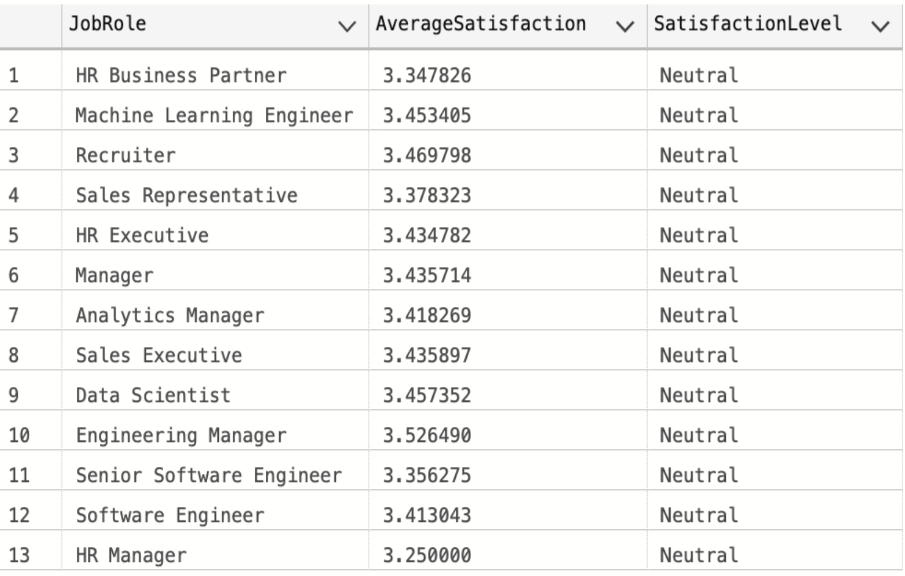
**Output**

****

**Solution using SQL:-**



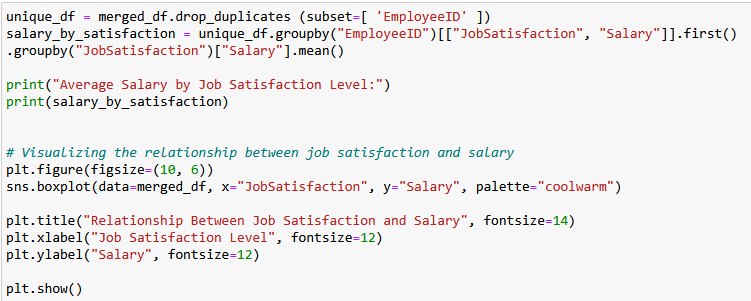
**Output**



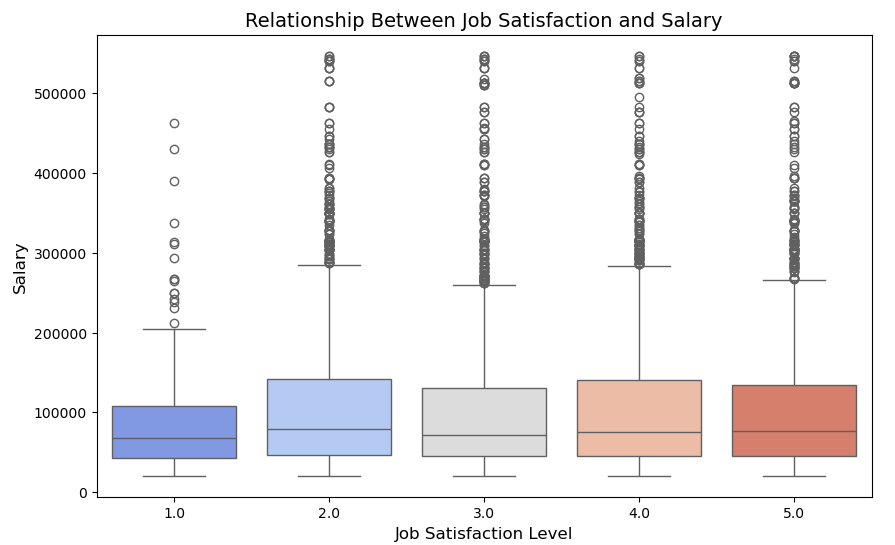
Through the analysis, it was found that the **satisfaction level** across different job roles is generally close. There is no significant variation between roles in terms of satisfaction level, indicating that employees in most roles have similar levels of satisfaction, with only minor differences in some roles.

**12:-Calculate AVG salary by satisfaction level for employees**

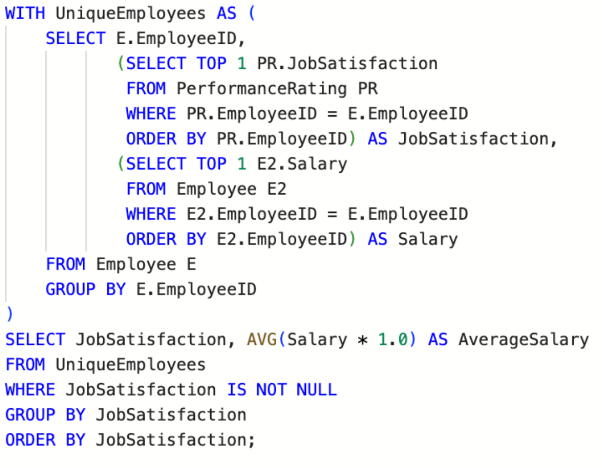
**Solution using Python:-**

****

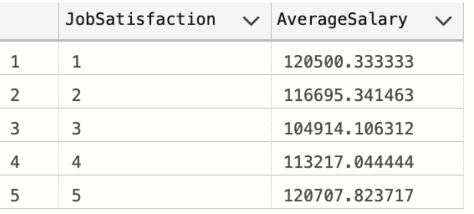
**Output**

****

**Solution using SQL:-**



**Output**



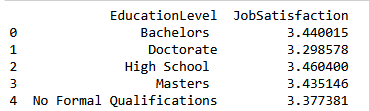
Through the analysis, it was found that the average **salary** across different **satisfaction levels** is very close. There are no significant differences between satisfaction levels in terms of salary, indicating that salaries are not strongly linked to the employees' satisfaction levels.

**13:-** **Do employees with higher education levels report higher satisfaction?**

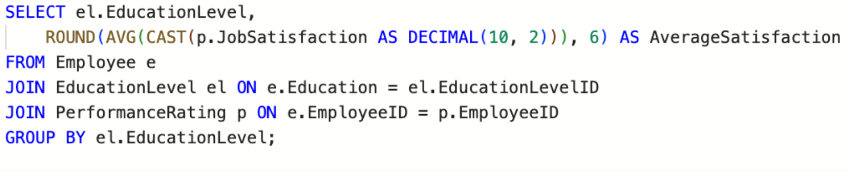
**Solution using Python:-**

****

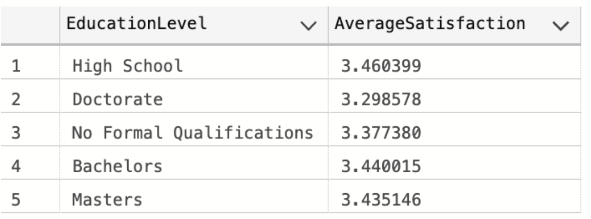
**Output**

****

**Solution using SQL:-**



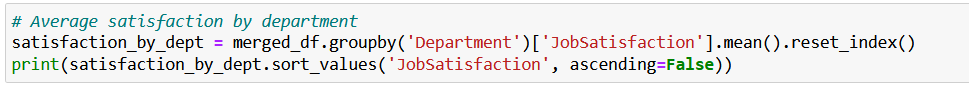
**Output**



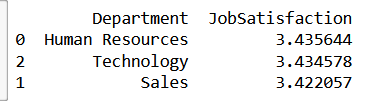
Through the analysis, it was found that **education level** does not have a significant impact on **satisfaction level**. All the results were quite similar, indicating that employees with different education levels report similar satisfaction levels, with no significant differences based on education.

**14:-** **Which departments have the most satisfied and least satisfied employees?**

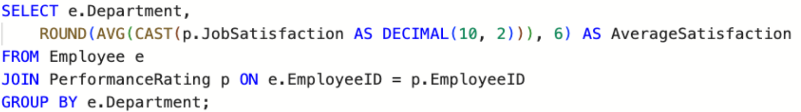
**Solution using Python:-**

****

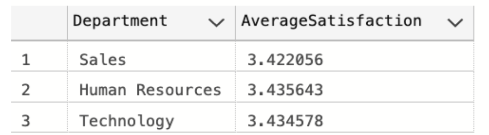
**Output**

****

**Solution using SQL:-**



**Output**



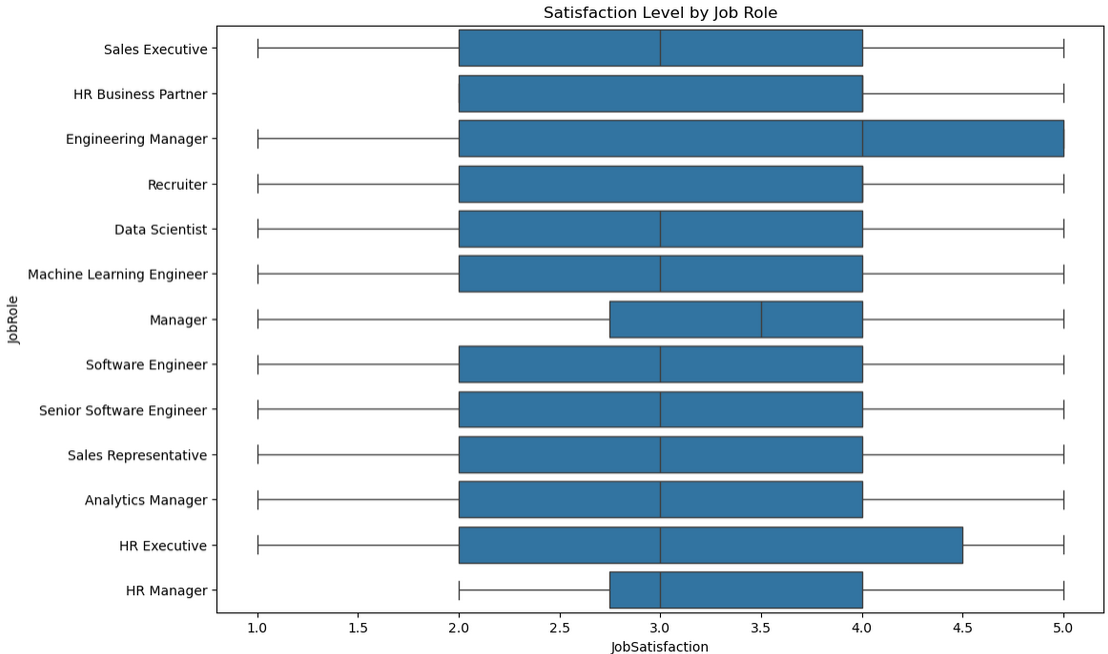
Through the analysis, it was found that the **satisfaction level** across departments varies slightly, but overall, the results were **quite similar**. No department stood out with a significant difference in satisfaction level compared to others, indicating that employees across most departments have similar levels of satisfaction.

**15:-** **Does job role impact satisfaction level?**

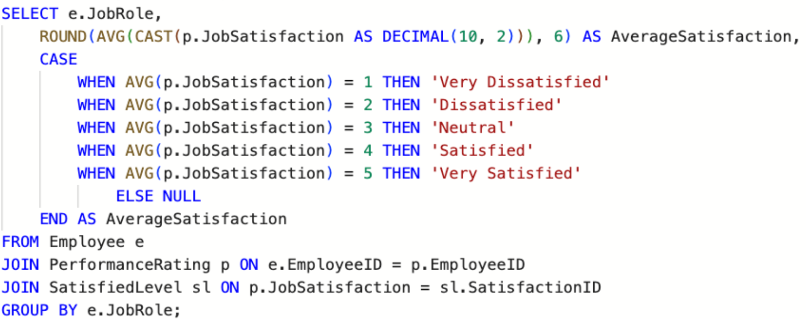
**Solution using Python:-**

****

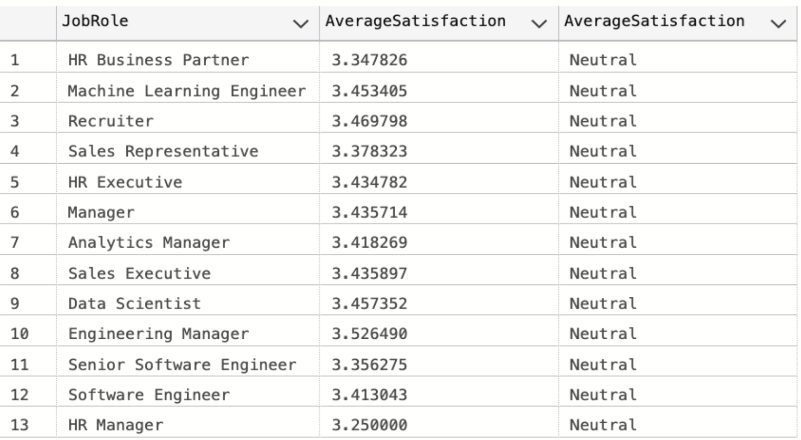
**Output**

****

**Solution using SQL:-**



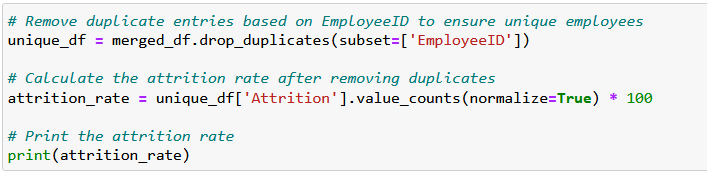
**Output**



Based on the data, the average satisfaction level across different job roles shows only **minor variations**. All job roles have an average satisfaction score around the neutral level (approximately between 3.25 and 3.52), indicating that **job role does not have a significant impact** on employee satisfaction. While roles like Engineering Manager and Recruiter show slightly higher averages, the differences are minimal and not enough to suggest a strong correlation.

**16:-** **What is the overall employee attrition rate?**

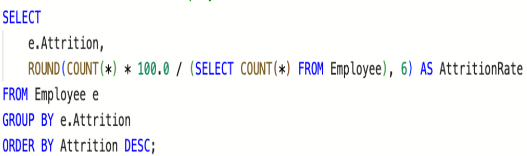
**Solution using Python:-**

****

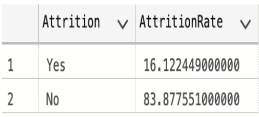
**Output**

****

**Solution using SQL:-**



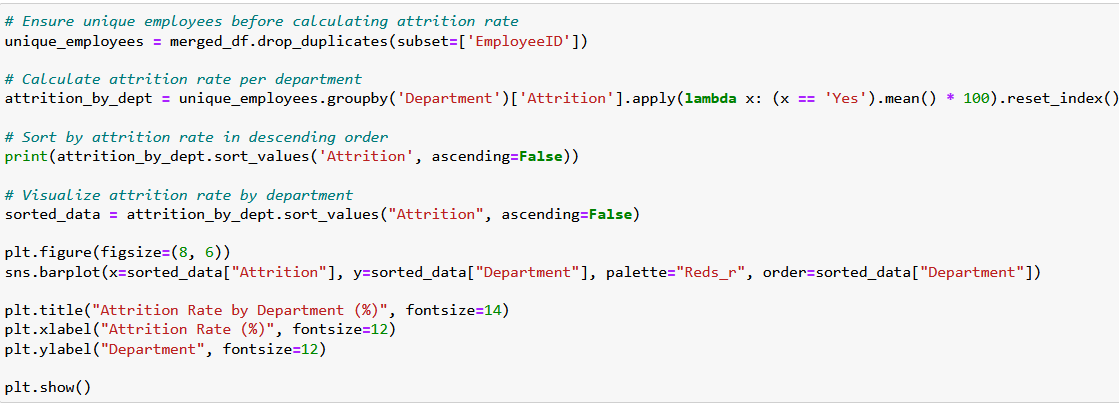
**Output**



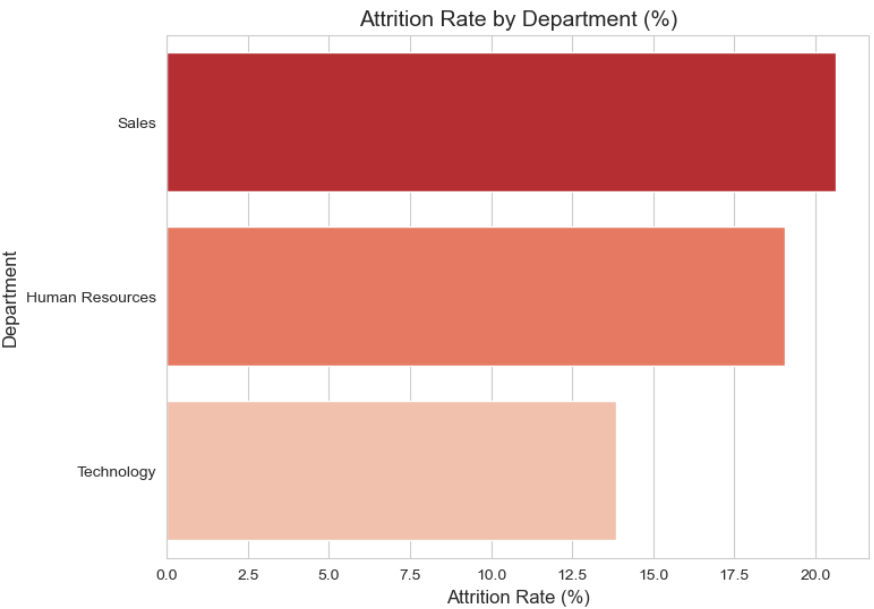
The overall employee attrition rate is approximately **16.12%**, indicating that a relatively small portion of the workforce has left the company. On the other hand, about **83.87%** of employees have remained. This suggests a **generally stable workforce**, with attrition not being a major issue at the organizational level.

**17:-** **Which department has the highest employee turnover?**

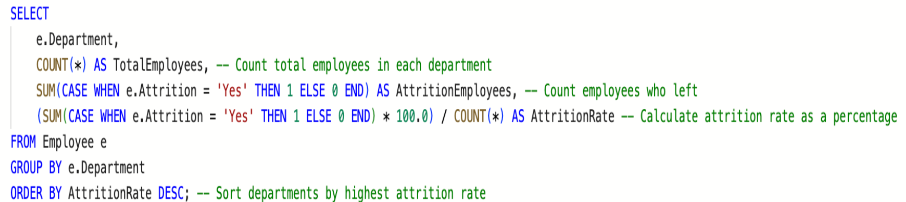
**Solution using Python:-**

****

**Output**

****

**Solution using SQL:-**



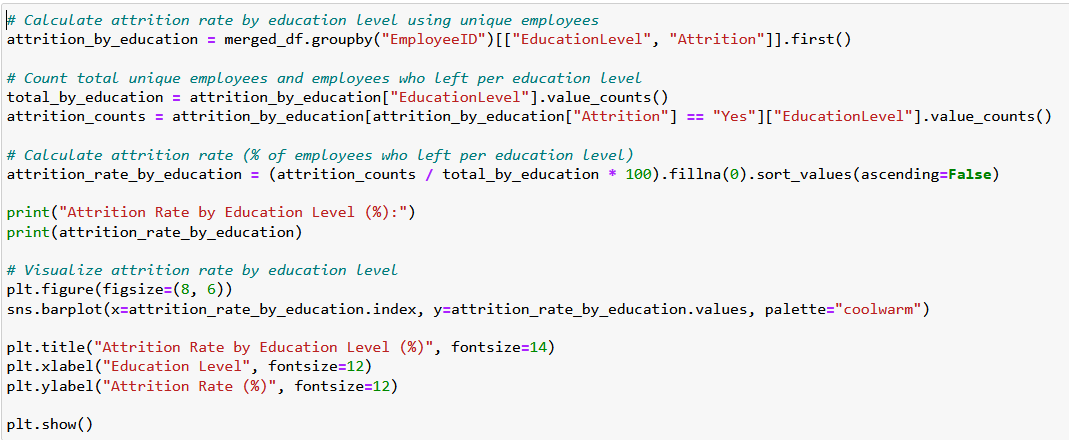
**Output**



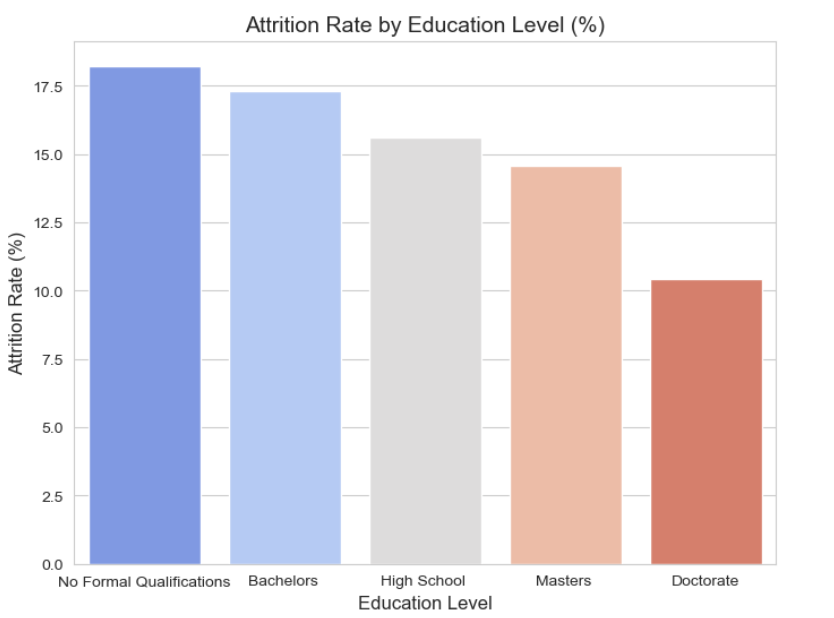
The **Sales** department has the highest employee turnover rate at **30%**, followed by **HR** at **19%**, and **Technology** at **13%**. This suggests that employees in the Sales department are more likely to leave compared to those in other departments, potentially indicating higher job pressure, dissatisfaction, or better external opportunities in that field.

**18:-** **Do employees with higher education levels have lower attrition rates?**

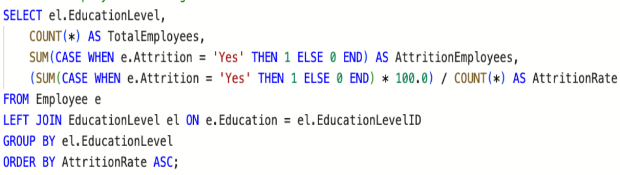
**Solution using Python:-**

****

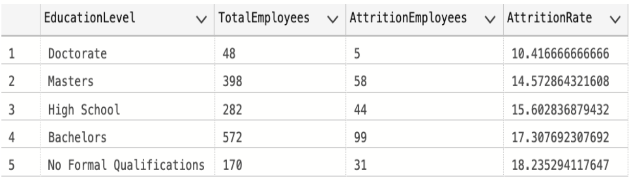
**Output**

****

**Solution using SQL:-**



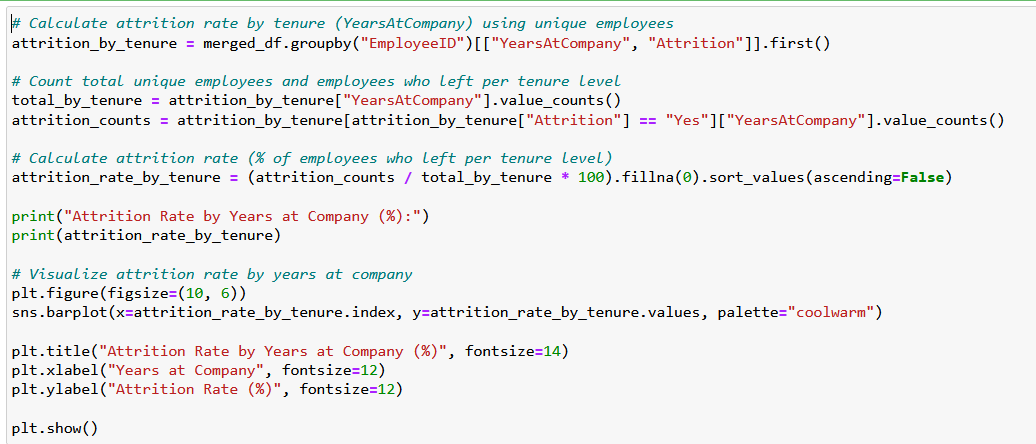
**Output**



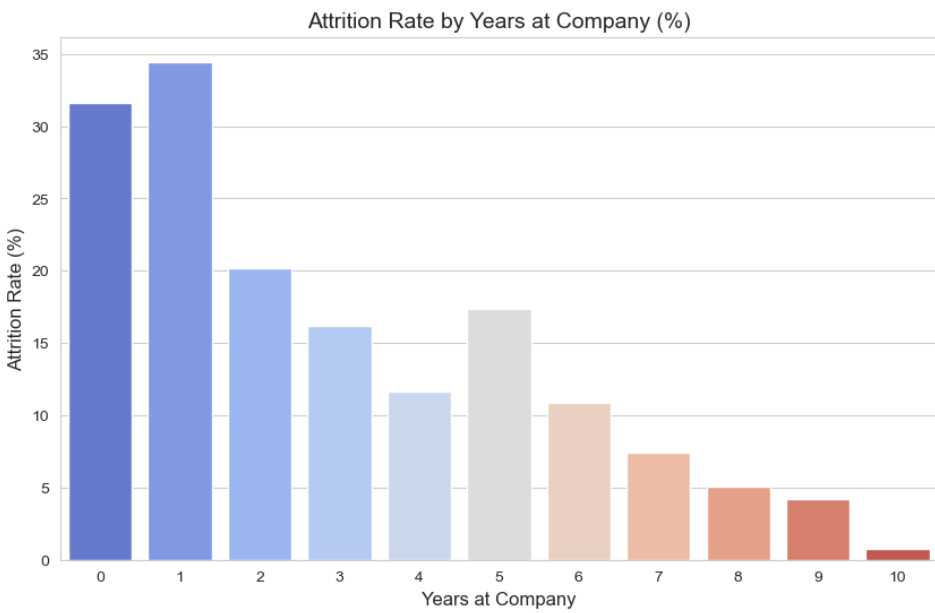
There is a noticeable trend indicating that employees with higher education levels tend to have lower attrition rates. Doctorate holders have the lowest attrition rate at **10.4%**, followed by Master's degree holders at **14.5%**. In contrast, employees with no formal education have the highest attrition rate at **18.2%**, suggesting that education may play a role in employee retention.

**19:-** **How does tenure (years at company) impact attrition?**

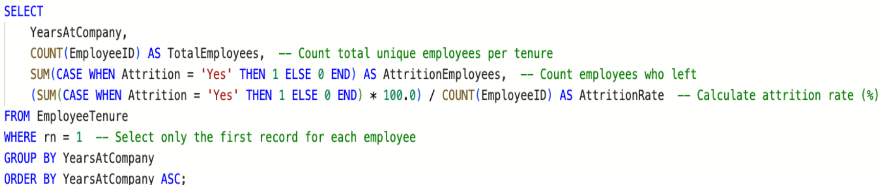
**Solution using Python:-**

****

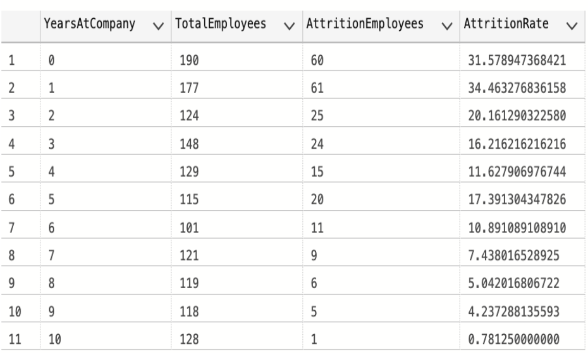
**Output**

****

**Solution using SQL:-**



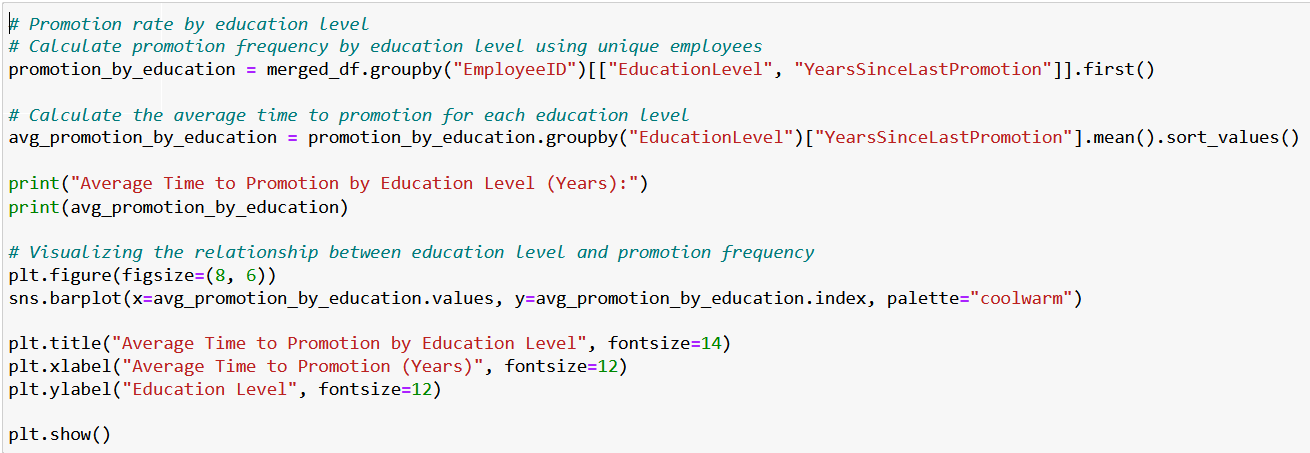
**Output**



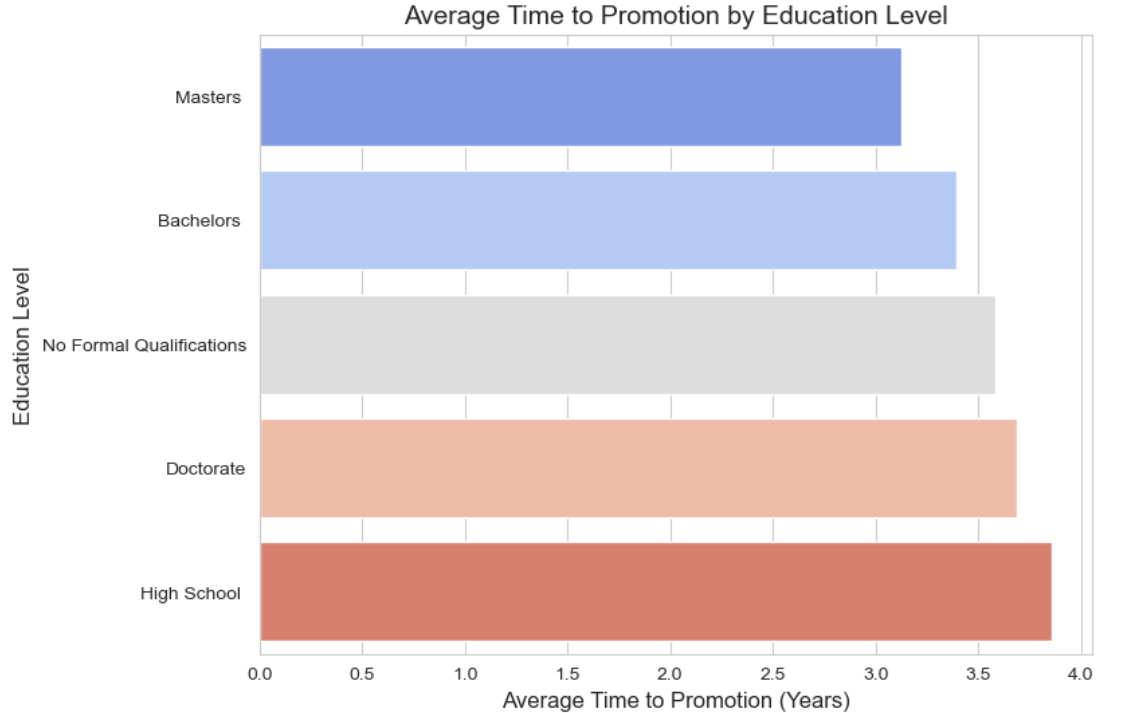
There is a clear trend showing that the longer employees stay at the company, the lower their attrition rate becomes. Employees with less than 2 years of tenure have the highest attrition rates—**31.5%** for those under 1 year and **34.6%** for those with exactly 1 year. However, as tenure increases, attrition significantly decreases. For example, employees with 5 years of service have a rate of **7.4%**, and those with 9 years have only **2%**, indicating that long-term employees are more likely to stay.

**20:-** **Is there a correlation between education level and promotion frequency?**

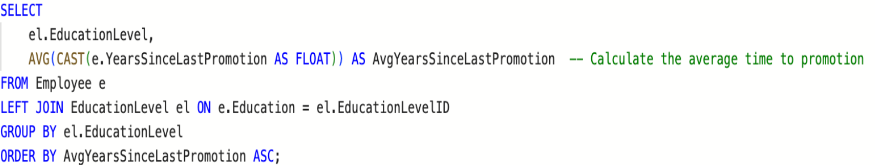
**Solution using Python:-**

****

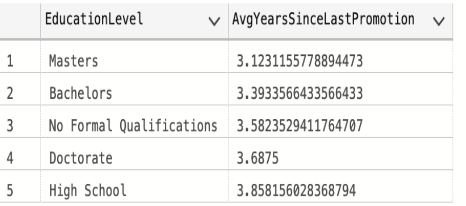
**Output**

****

**Solution using SQL:-**



**Output**



There appears to be a **slight correlation** between higher education levels and promotion frequency. Employees with **Doctorate degrees** had the **highest average promotions (3.6 times)**, followed closely by **non-formal education holders (3.5)** and **Master's degree holders (3.1)**. Meanwhile, **Bachelor’s degree** holders had an average of **3.3 promotions**, and **High School graduates** had **3.8**, which is unexpectedly higher. This indicates that while education level may play a role, it is **not the only factor** affecting promotion frequency.

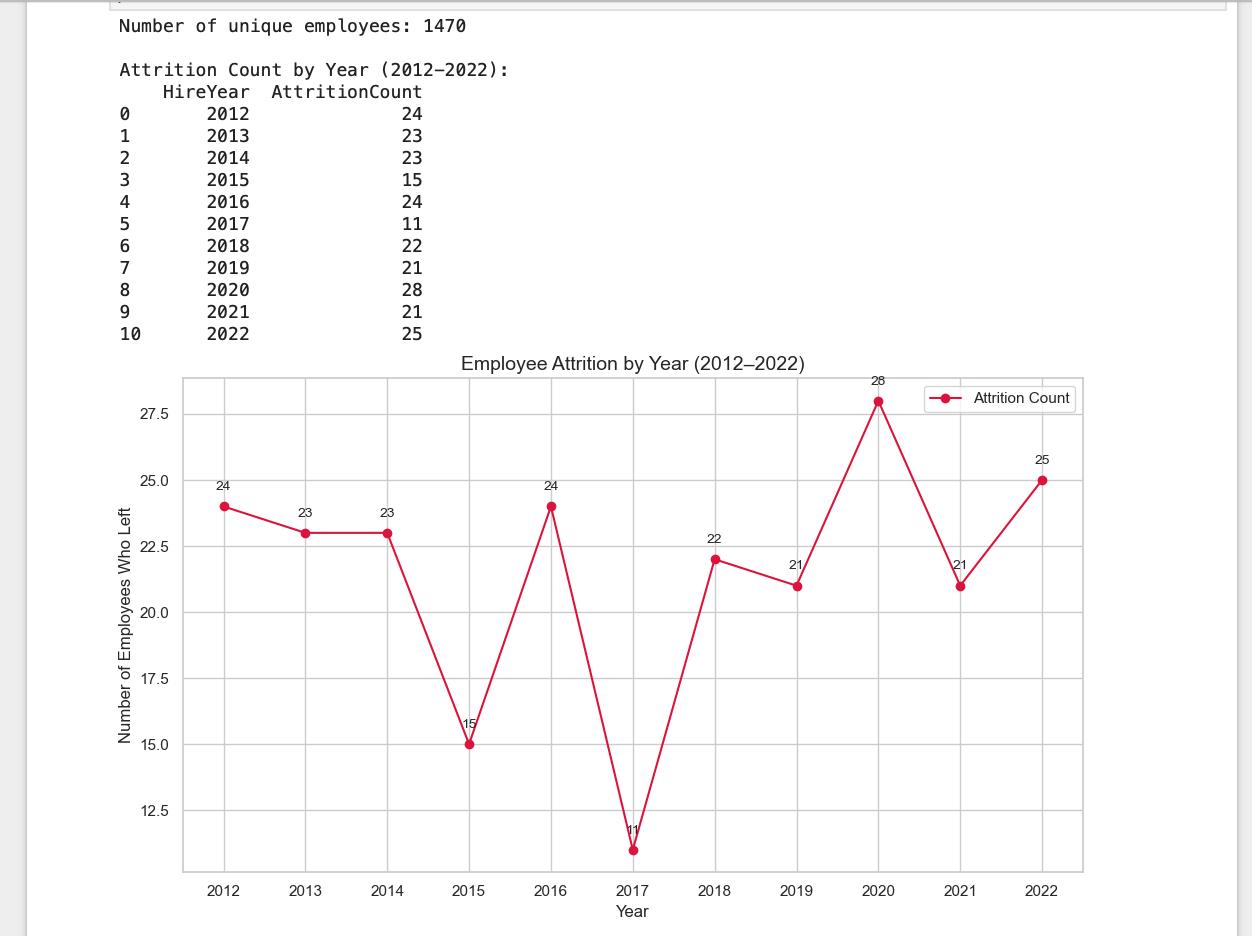
**Forcasting Analysis**

**Chapter 4**

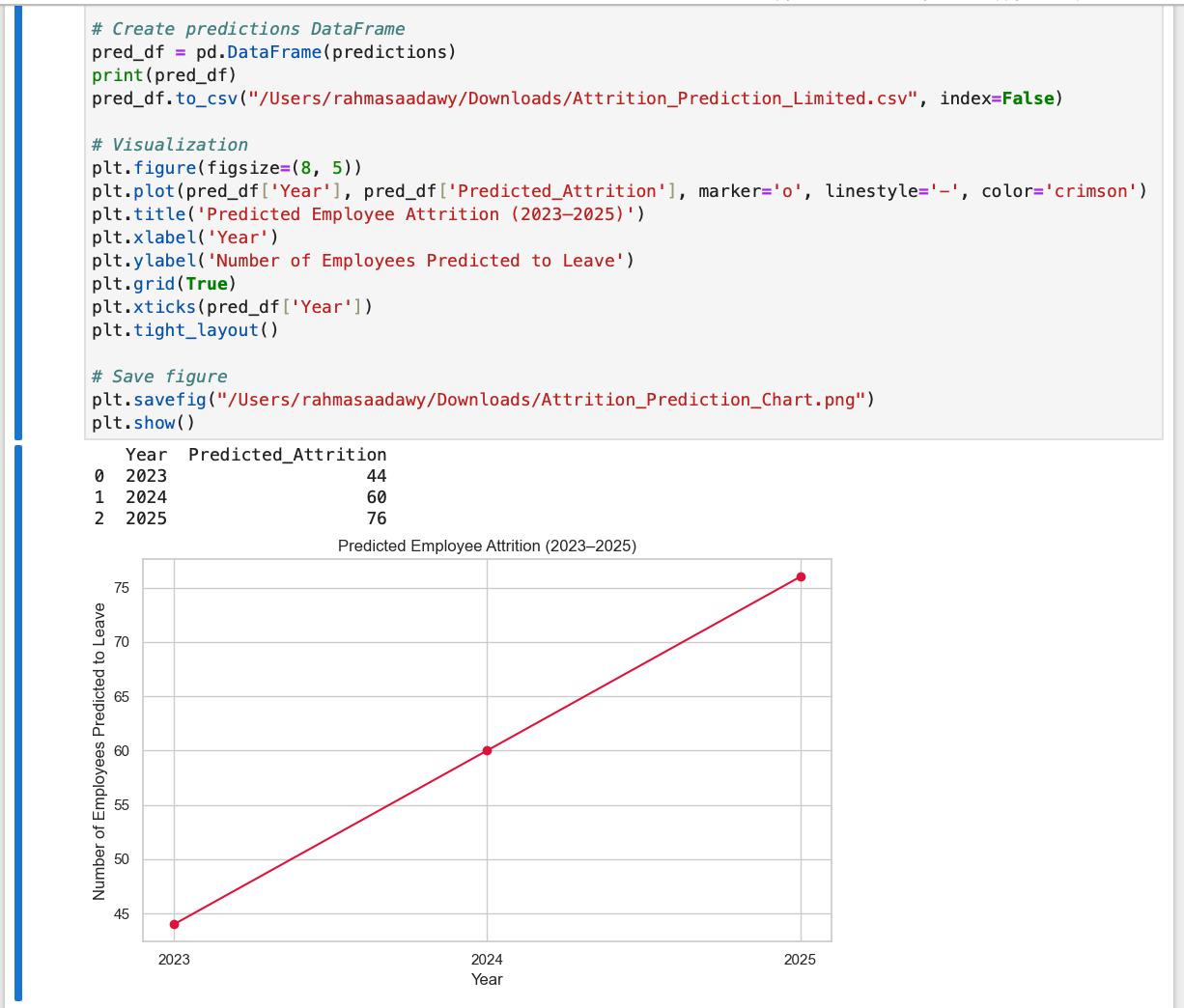
**Introduction :-**

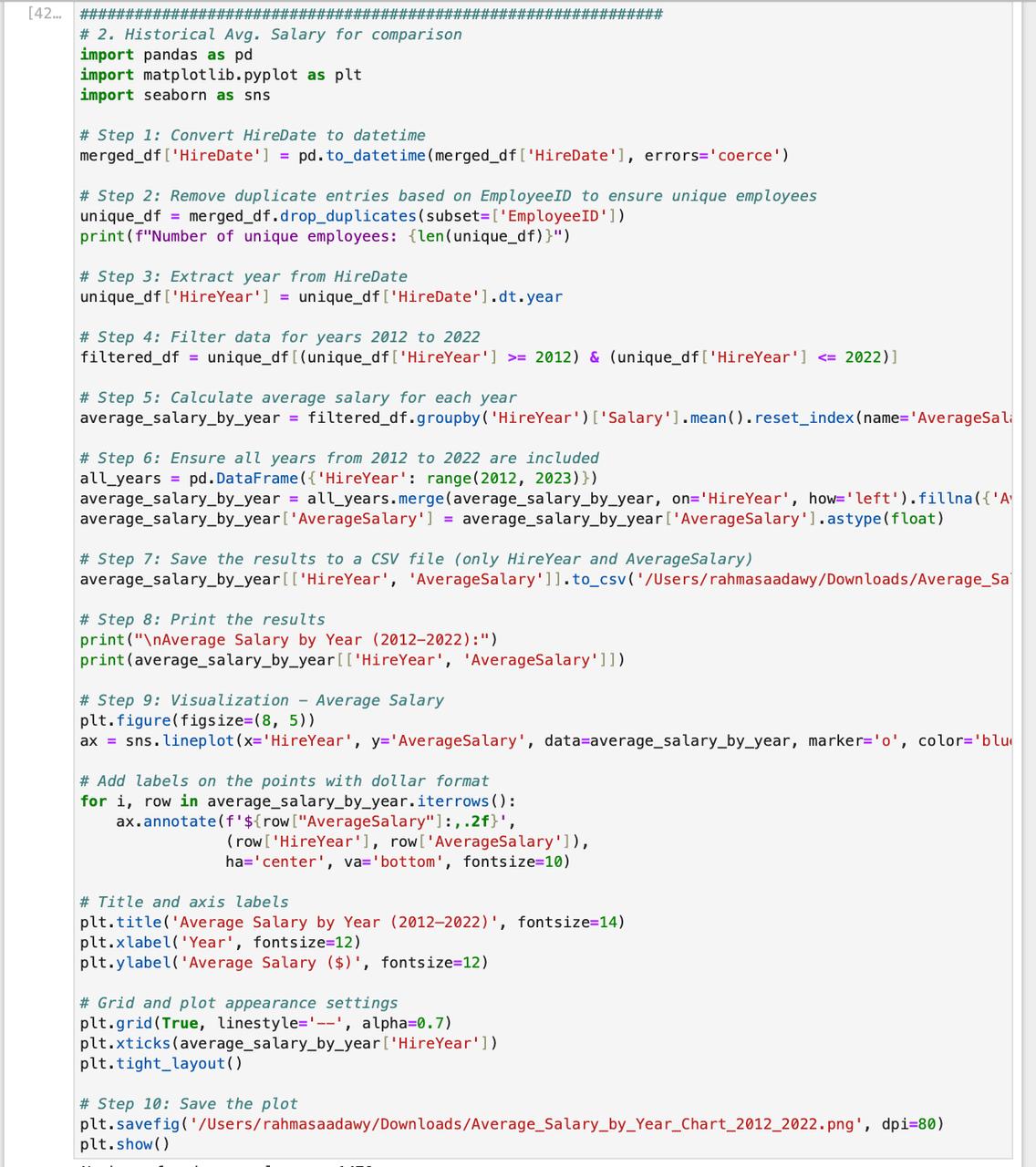
This chapter presents the forecasting analysis for employee attrition and average salary trends from **2023** to **2025**, building on historical data from **2012** to **2022**. By leveraging predictive models, we aim to identify potential challenges in workforce retention and compensation, offering insights into future trends. The forecasts highlight a significant rise in attrition and steady salary growth, prompting a deeper exploration of underlying factors and actionable strategies to enhance organizational stability and employee satisfaction.

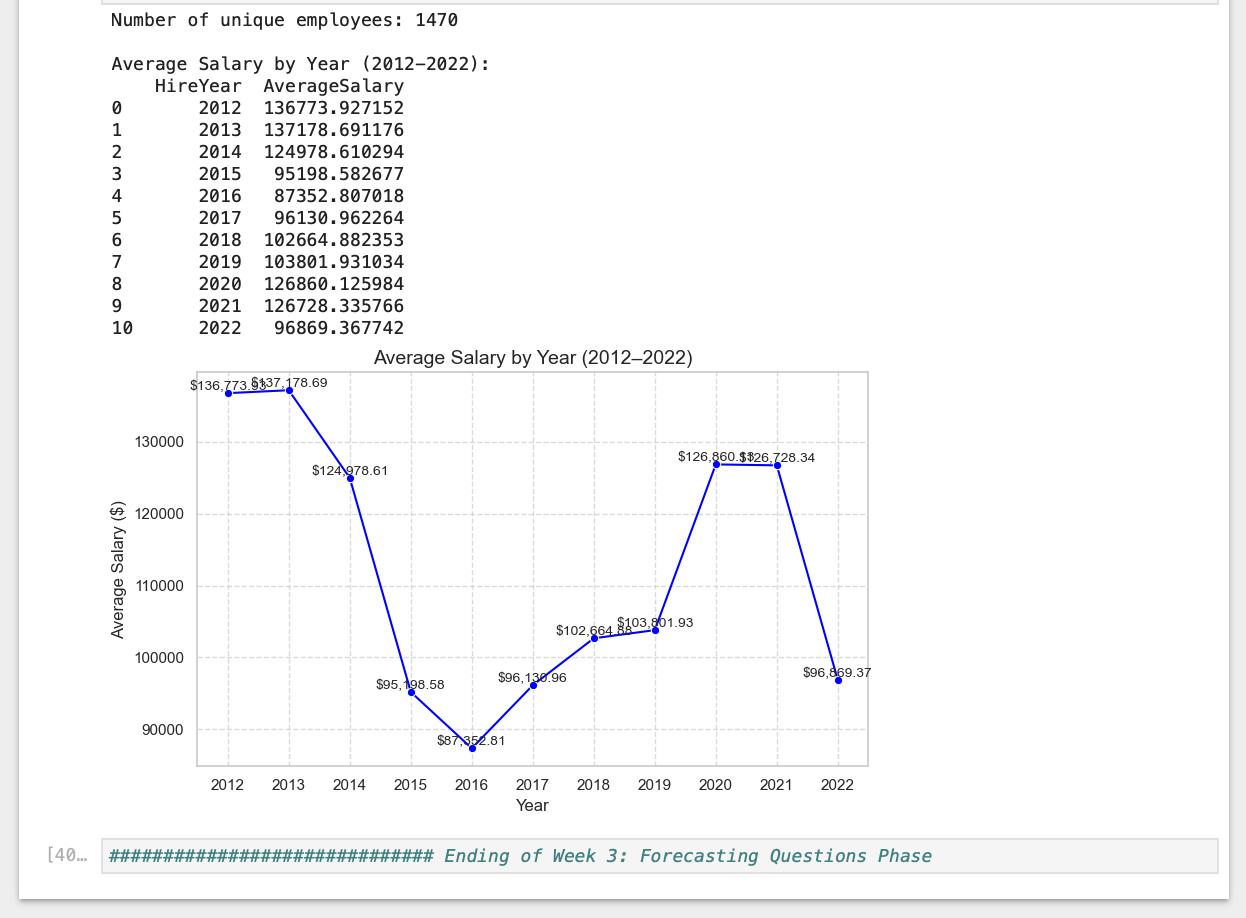
**Question 1** **Python code**

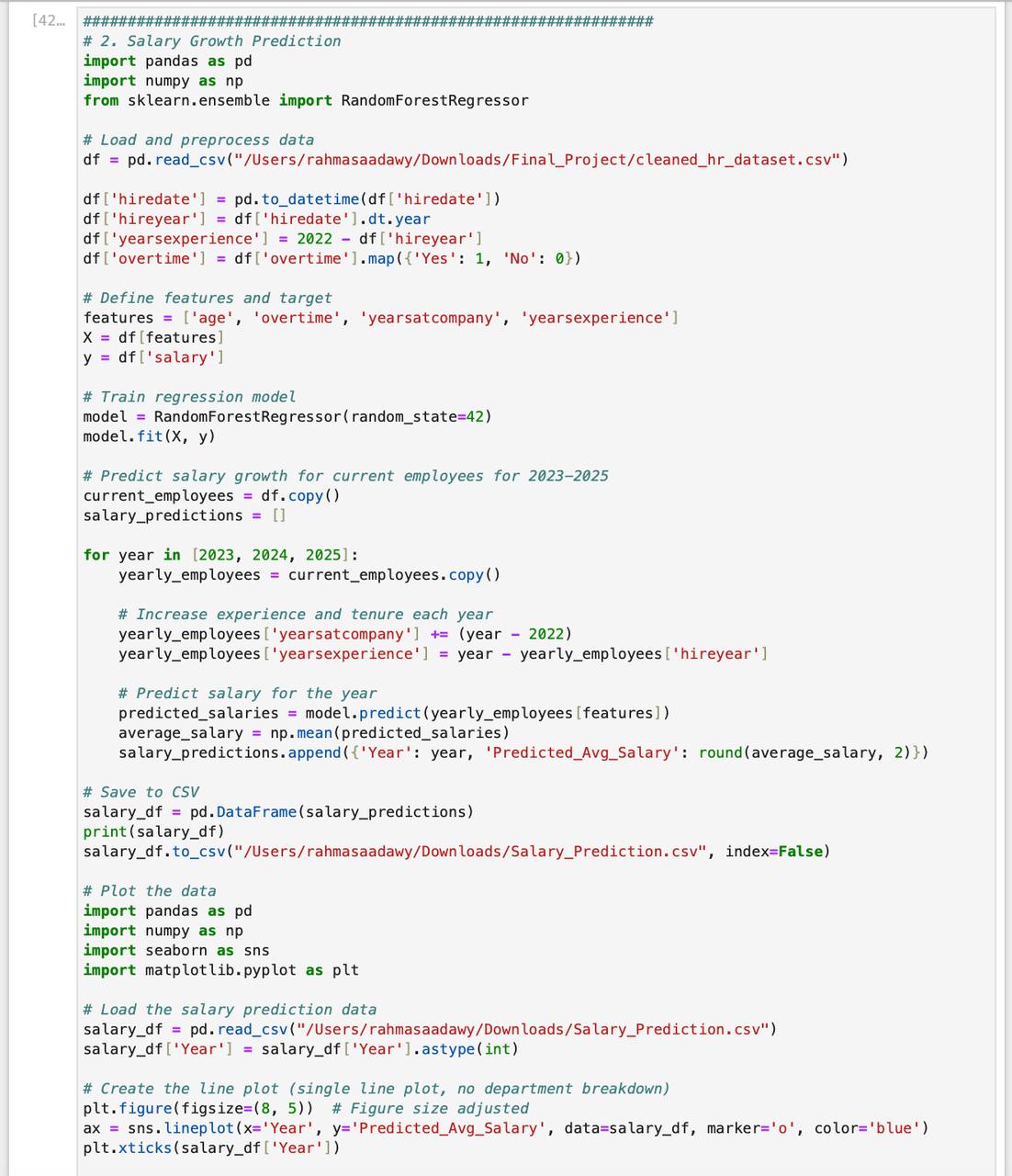
****

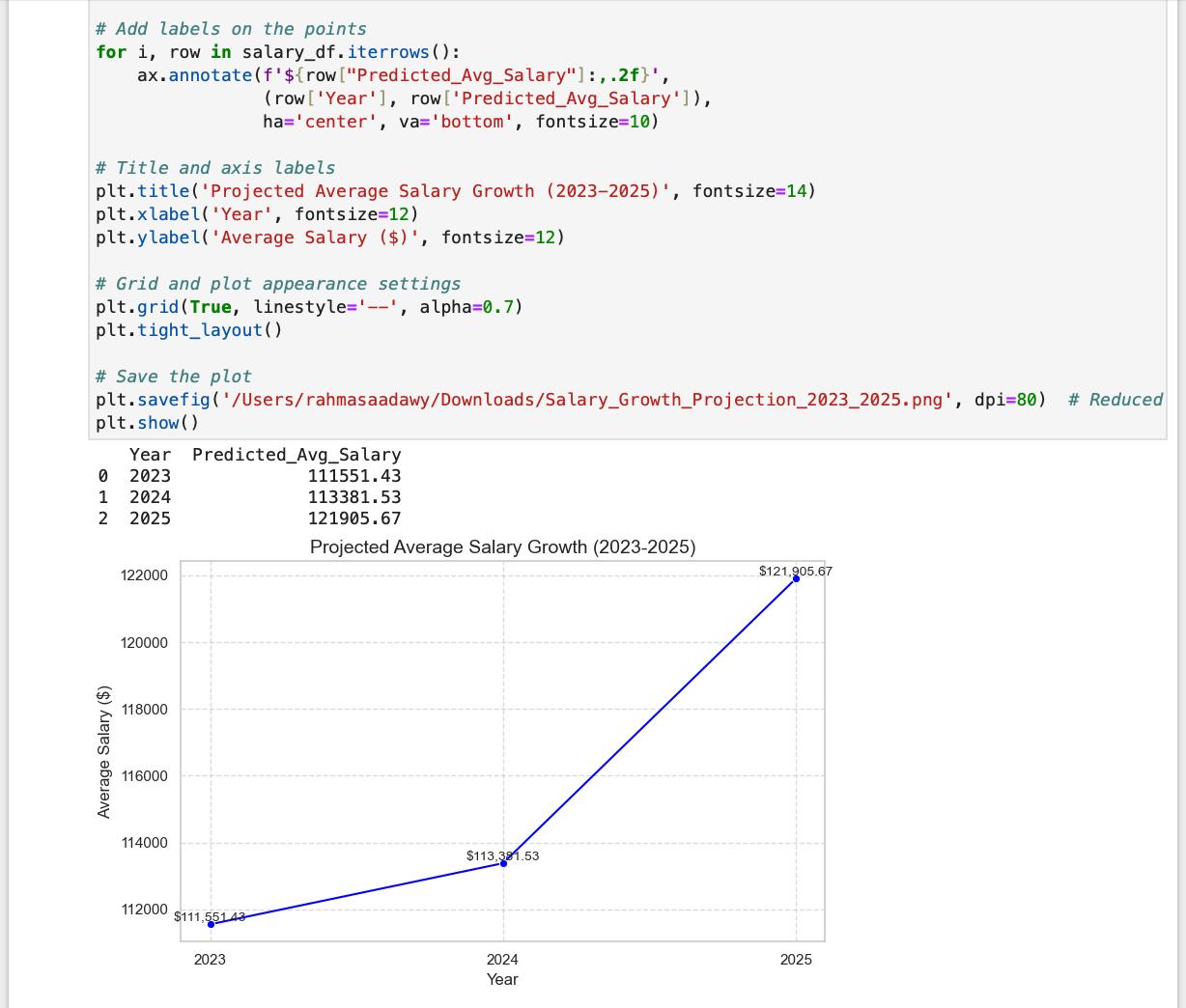
**Question 2 Python code**



**Question 3 Python code**



**Question 4 Python code**



# **Employee Attrition and Salary Analysis Report (2012–2025)**

## **Summary**

This report analyzes historical employee attrition and average salary trends from 2012 to 2022, alongside predictions for 2023 to 2025. The goal is to understand the forecasted increase in attrition and salary growth, compare them with past trends, and suggest improvements for better workforce management.

## **Historical Trends (2012–2022)**

Attrition: Attrition fluctuated significantly, dropping to a low of 11 employees in 2017 and peaking at 28 in 2020. The average attrition rate hovered around 21 employees per year, with notable volatility likely tied to economic or organizational changes.

Average Salary: Salaries saw a sharp decline from $136,773 in 2012 to $87,352 in 2016, possibly due to market adjustments or hiring lower-cost talent. A recovery followed, peaking at $126,860 in 2020, before dropping to $96,869 in 2022, reflecting economic instability or restructuring.

## **Predictions (2023–2025)**

Attrition Forecast: Predicted attrition rises sharply from 44 employees in 2023 to 76 in 2025—a 73% increase over three years. This steep upward trend contrasts with the historical average of 21, suggesting factors like declining job satisfaction, competitive job markets, or insufficient retention strategies may be at play.

Salary Forecast: Average salaries are projected to grow steadily from $111,551 in 2023 to $121,905 in 2025—a 9.3% increase. While this growth is positive, it’s slower than historical peaks (e.g., 2020’s $126,860), indicating cautious compensation adjustments despite rising attrition.

## **Analysis of Predictions**

The sharp rise in predicted attrition could stem from unaddressed employee dissatisfaction, possibly linked to the salary stagnation seen in 2022 ($96,869), which is well below the 2020 peak. Employees may be seeking better opportunities elsewhere, especially if competitors offer higher pay or better benefits. The modest salary growth forecast may not be enough to retain talent in a competitive market, further driving attrition. Historically, low salaries in 2016 ($87,352) coincided with higher attrition (24 employees), supporting this correlation.

## **Recommendations for Improvement**

1. Enhance Retention Strategies: Address potential dissatisfaction by improving work-life balance, offering career development, or increasing non-monetary benefits like flexible work options.

2. Adjust Compensation: Accelerate salary growth to align with or exceed market rates, especially since historical data shows a link between low salaries and higher attrition.

3. Conduct Exit Surveys: Gather data on why employees leave to refine the predictive model and address root causes.

4. Monitor Market Trends: Benchmark salaries and benefits against competitors to ensure the company remains attractive to talent.

## **Conclusion**

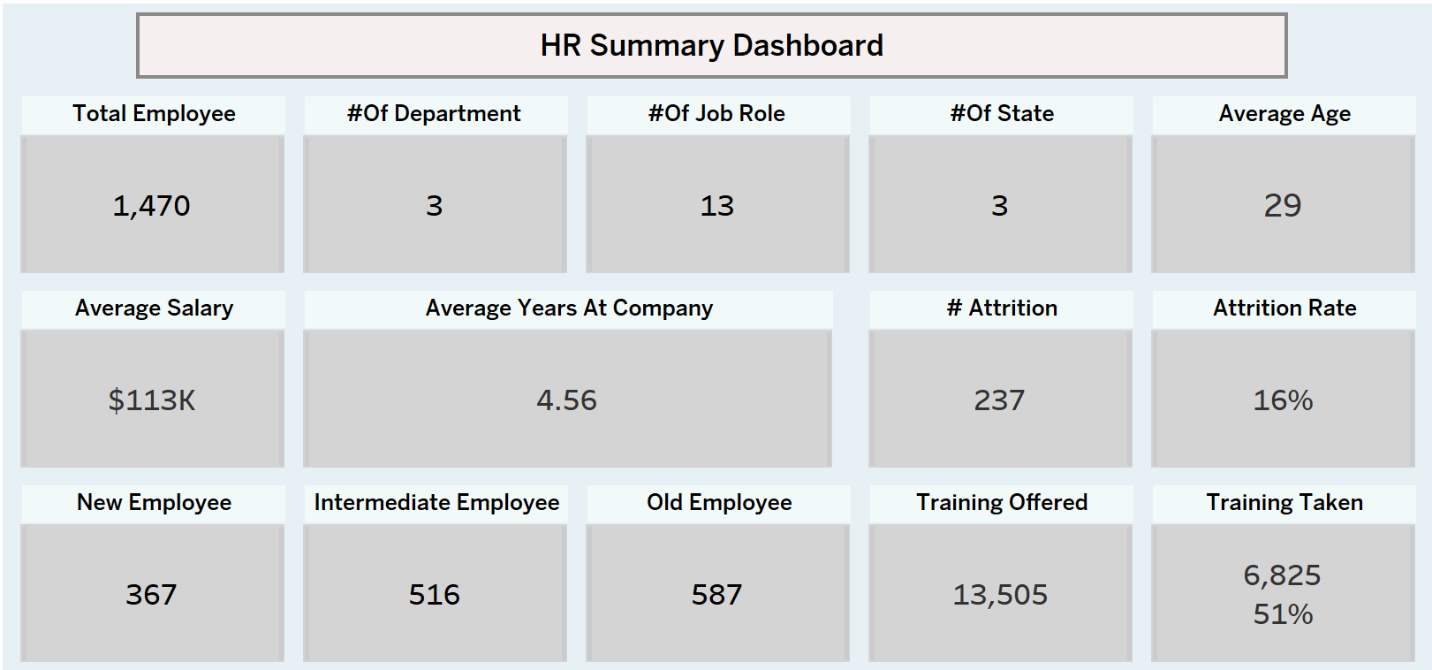
The predicted rise in attrition and modest salary growth highlight a critical need for proactive retention strategies. By addressing compensation and employee satisfaction, the company can mitigate the forecasted turnover spike and build a more stable workforce through 2025.

**Visualisation**

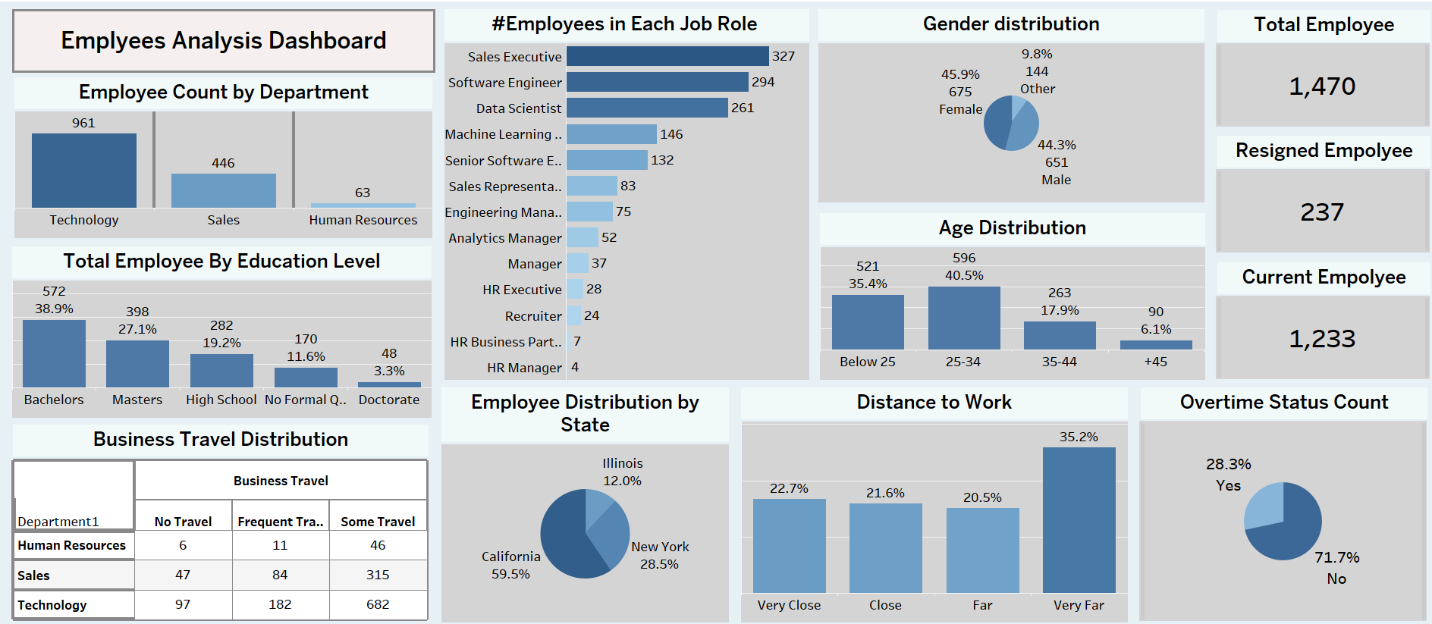
**Chapter 5**

**Introduction :-**

In this chapter, we review a set of HR-related dashboards that enable HR and management teams to track and analyze employee data in a visual and organized manner. The goal of these dashboards is to improve decision-making, understand job distribution, and track performance, attendance, hiring, and payroll.

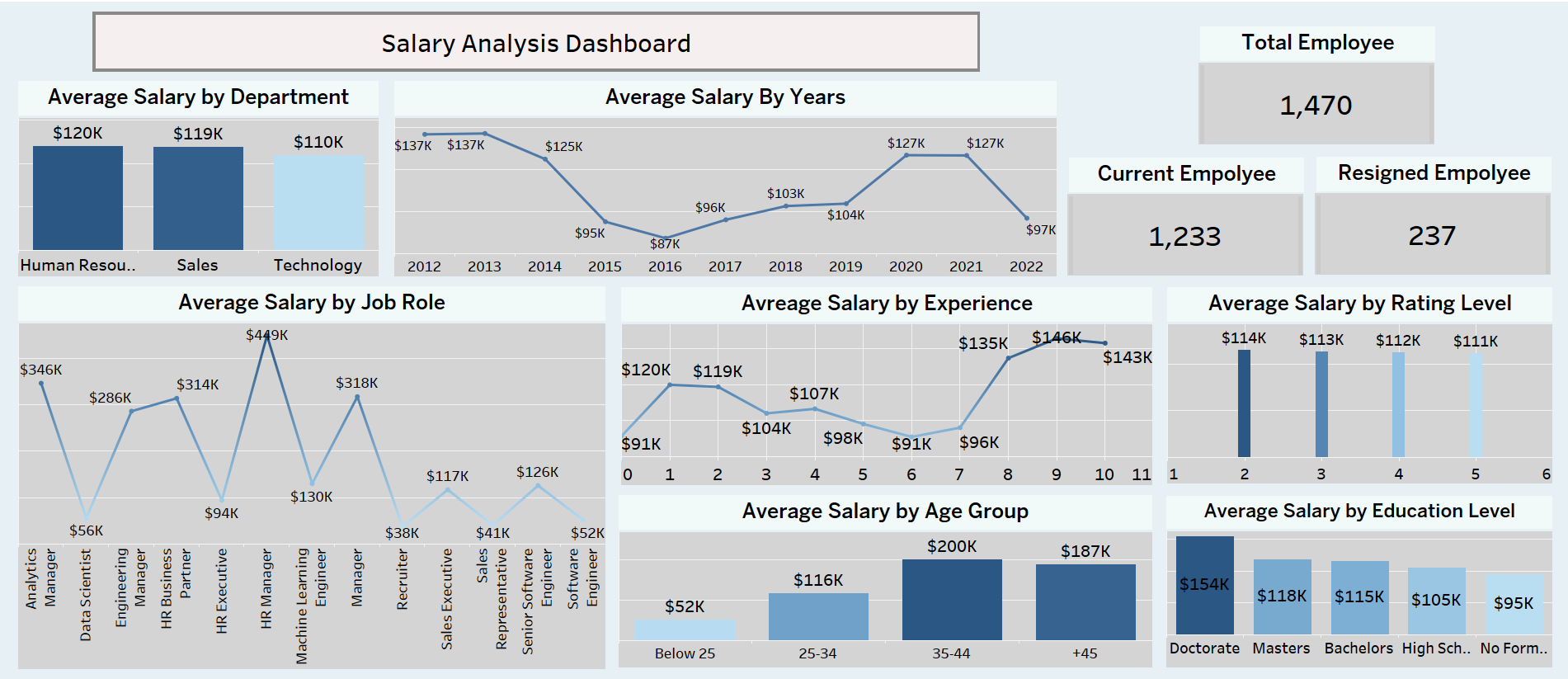
**HR Summary Dashboard**

The HR Summary Dashboard provides a high-level overview of key HR indicators such as total employees, job roles, salaries, attrition, and training. It's designed to help management and HR teams quickly understand workforce status and support strategic decisions.

**Employees Analysis Dashboard**

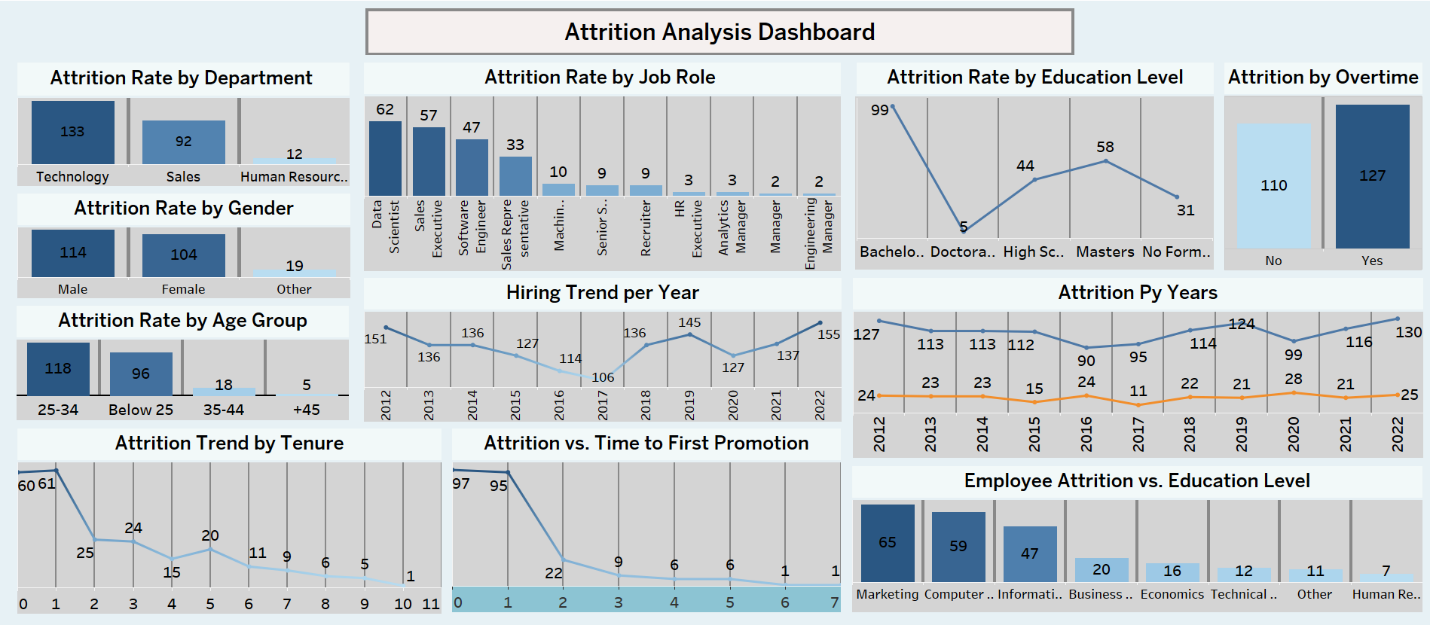
The Employees Analysis Dashboard breaks down the workforce by department, job roles, education level, gender, age group, and location. It helps management better understand workforce composition and supports data-driven decisions in hiring and resource allocation.

**Salary Analysis Dashboard**



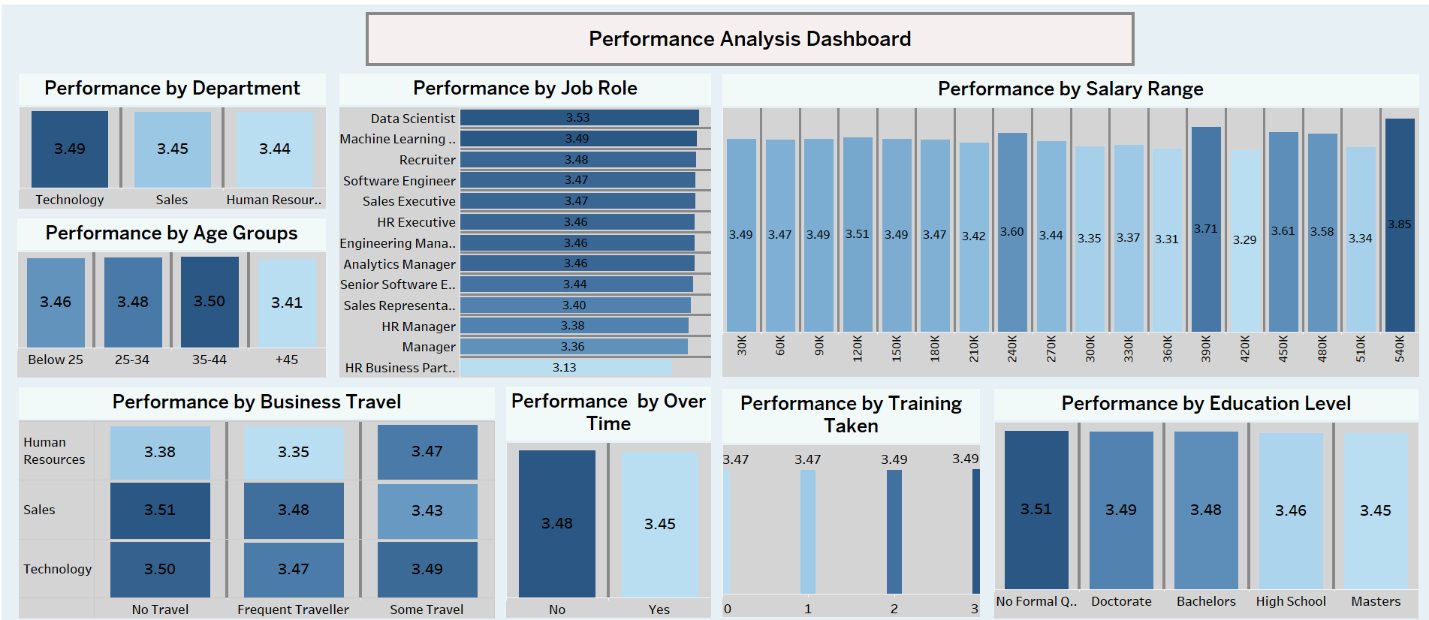
The Salary Analysis Dashboard highlights salary variations across the company based on department, years of experience, job role, rating level, age group, and education level. It supports fair and data-driven decisions regarding compensation and rewards.

**Attrition Analysis Dashboard**



This dashboard provides a comprehensive overview of employee attrition across various dimensions such as age, gender, department, job role, education level, and tenure. It also examines the impact of factors like overtime, time to first promotion, and hiring trends over the years. The visualizations help identify key patterns and potential causes of employee turnover, enabling better strategic decisions to improve employee retention and workplace satisfaction.

**Employee Performance Analysis Dashboard**



This dashboard provides an in-depth analysis of employee performance based on various factors such as department, job role, salary range, age group, business travel, overtime, training taken, and education level. It highlights performance patterns across different employee segments, helping organizations identify key strengths and areas for improvement to enhance overall productivity and workforce management.

**Recommendation**

**Chapter 6**

**Recommendations :**

1. Launch engagement programs for Sales & Tech departments.
2. Improve internal promotion cycles.
3. Offer hybrid work options for long-distance employees.
4. Increase training opportunities for mid-career employees.
5. Use forecasts for strategic workforce planning.