# HR Analytics using MySQL & Tableau

**PROJECT REPORT**

**18CSE415J – FOUNDATION OF ANALYTICS**

**(2018 Regulation)**

**III Year/ V Semester**

**Academic Year: 2023 -2024**

By

**DEVANANDA REDDY (RA2111004010086)**

**MASKU RAHUL (RA2111004010087)**

Under the guidance of

**Dr. Subha T**

**Assistant Professor**

**Department of Cloud Computing, IoT, Machine Learning**



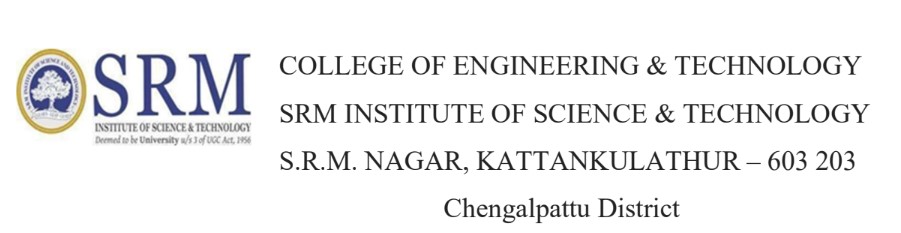
**FACULTY OF ENGINEERING AND TECHNOLOGY**

**DEPARTMENT OF ELECTRONICS AND COMMUNICATION**

**SRM INSTITUTE OF SCIENCE AND TECHNOLOGY**

**Kattankulathur, Kancheepuram**

**NOVEMBER 2023**



**BONAFIDE**

This is to certify that **18CSE415J – FOUNDATION OF ANALYTICS project report** titled “**HR Analytics using MySQL & Tableau”** is the bonafide work of **DEVANANDA REDDY RA2111004010086**

who undertook the task of completing the project within the allotted time.

|  |  |
| --- | --- |
| **SIGNATURE**  Dr. Subha T  **Course Faculty**  Assistant Professor  Department of Cloud Computing, IoT, Machine Learning  SRM Institute of Science and Technology  Kattankulathur | **SIGNATURE** |

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **S.NO** | **CONTENTS** | **PAGE NO** |
| **1.** | **AIM & ABSTRACT** | **3** |
| **2.** | **DATASET DESCRIPTION** | **5** |
| **3.** | **TOOLS** | **6** |
| **4.** | **PROJECT OVERVIEW** | **7** |
| **5.** | **UML DIAGRAMS** | **9** |
| **6.** | **RESULTS & OUTPUT** | **16** |
| **7.** | **DATASET TABLES** | **22** |
| **8.** | **CONCLUSION** | **23** |
| **9.** | **REFERENCES** | **23** |

# 

# HR Analytics using MySQL & Tableau

# AIM:-

The aim of the "HR Analytics using MySQL & Tableau Software" project is to revolutionize human resource management through the integration of robust data storage and advanced visualization tools. The primary objective is to establish a comprehensive and scalable solution for HR analytics, leveraging MySQL for efficient data management and Tableau for dynamic and interactive visualizations. This project aims to empower HR professionals with data-driven insights, facilitating informed decision-making, strategic workforce planning, and the optimization of human resource practices within the organization. Ultimately, the project seeks to elevate the role of analytics in HR, fostering a culture where organizational decisions are grounded in a thorough understanding of workforce dynamics achieved through the synergy of MySQL and Tableau.

# ABSTRACT:-

This project explores the integration of HR analytics using MySQL and Tableau software to enhance human resource management. In the contemporary corporate landscape, data-driven decision-making has become pivotal for effective HR strategies. The project aims to leverage the power of MySQL for data storage, management, and retrieval, coupled with Tableau's advanced visualization capabilities, to derive meaningful insights from HR data.

The process begins with the collection and organization of diverse HR datasets encompassing employee demographics, performance metrics, and other relevant information. MySQL is employed to ensure a robust and scalable database infrastructure, allowing for efficient storage and retrieval of HR data. The structured data is then subjected to exploratory data analysis, uncovering patterns, trends, and correlations crucial for understanding workforce dynamics.

Furthermore, the project delves into hypothesis testing, allowing for the validation or refutation of formulated hypotheses related to key HR metrics such as employee satisfaction and turnover rates. MySQL facilitates the statistical computations required for hypothesis testing, ensuring the accuracy and reliability of the results.

The heart of the project lies in the integration of Tableau for data visualization. Interactive dashboards are created to present HR insights in a visually appealing and comprehensible manner. Tableau's user-friendly interface enables HR professionals to explore data, identify trends, and make informed decisions seamlessly. The dashboard includes visual representations of employee performance, demographic distributions, and other critical HR metrics.

The synergy between MySQL and Tableau in HR analytics empowers organizations to move beyond conventional HR practices. The project demonstrates the effectiveness of this integrated approach in fostering a data-driven HR culture, where strategic decisions are grounded in comprehensive analytics. As organizations strive for enhanced workforce management, the amalgamation of MySQL and Tableau emerges as a potent tool for transforming HR practices, ensuring they align with the evolving demands of the modern business landscape

**DATASET DESCRIPTION:**

**Here the dataset represents: -**

1. Attrition Using Count-plot
2. The count Sum of Employee, Attrition, Attrition Rate, Active Employees, Average age
3. Total Count of Gender.
4. Attrition by Gender by bar plot.
5. Department-wise Attrition using pie chart.
6. Number of Employees by Age Group using bins.
7. Job Satisfaction Rating using Crosstab.
8. Education Field Wise Attrition Counts using Descending Order bar plot.
9. Attrition Rate by Gender for Different Age Groups using Lollipop chart.

**TOOLS:-**

For the project "HR Analytics using MySQL & Tableau software," the following tools are utilized to achieve comprehensive data management, analysis, and visualization:

**MySQL:**

* **Purpose:** MySQL serves as the relational database management system (RDBMS) for storing, managing, and retrieving HR data.
* **Functionality:** MySQL is employed to create and maintain a structured database that houses diverse HR datasets, ensuring data integrity and efficient query processing.

**Key Features:**

* **Data Storage:** Efficiently stores and organizes HR-related information.
* **Query Language:** Utilizes SQL for querying and managing the database.
* **Scalability:** Provides a scalable solution for handling growing volumes of HR data.

**Tableau:**

* **Purpose:** Tableau is used for data visualization and interactive dashboard creation.
* **Functionality:** Tableau facilitates the exploration of HR data through visually appealing and interactive dashboards, enabling stakeholders to gain insights quickly.
* **Key Features:**
* **Data Connectivity:** Connects to MySQL databases to fetch real-time data.
* **Visualization:** Offers a wide array of visualization options for HR metrics.
* **Interactivity:** Enables users to interact with and explore data dynamically.

**Excel:**

* **Purpose:** Excel is employed for data preprocessing and analysis before importing data into MySQL.
* **Functionality:** Excel is used for tasks such as cleaning, formatting, and organizing raw HR data before integration into the MySQL database.
* **Key Features:**
* **Data Preparation:** Assists in preparing datasets for MySQL ingestion.
* **Analysis:** Provides basic analysis and validation of HR data.

**Python (Optional):**

* **Purpose:** Python may be used for advanced data preprocessing or statistical analysis tasks.
* **Functionality**: Python scripts can automate data cleansing, transformation, and preliminary analysis, enhancing the efficiency of the overall HR analytics process.
* **Key Features:**
* **Data Manipulation:** Leverages Python libraries for efficient data manipulation.
* **Statistical Analysis:** Supports advanced statistical analysis if needed.

These tools collectively form a robust and integrated environment for HR analytics, allowing for efficient data storage, analysis, and visualization. The synergy between MySQL and Tableau provides a seamless workflow from data storage to insightful visualizations, empowering HR professionals to make informed decisions and derive meaningful insights from the organization's workforce data.

## **PROJECT OVERVIEW:-**

The HR Analytics project using MySQL & Tableau software aims to revolutionize human resource management within the organization by adopting a data-driven approach. Leveraging MySQL as the database management system and Tableau for advanced data visualization, the project integrates these powerful tools to provide a comprehensive solution for HR analytics. The overarching goal is to empower HR professionals with actionable insights into employee performance, satisfaction, and overall workforce dynamics.

**Objectives:**

**Data Integration and Storage:** Utilize MySQL to establish a robust database infrastructure for efficient storage, retrieval, and management of diverse HR datasets.

**Exploratory Data Analysis (EDA):** Conduct thorough EDA to uncover patterns, trends, and correlations within the HR dataset, providing a deeper understanding of workforce dynamics.

**Hypothesis Testing:** Formulate and test hypotheses related to key HR metrics, including employee satisfaction, turnover rates, and performance indicators, using MySQL for statistical computations.

**Data Visualization:** Leverage Tableau's advanced visualization capabilities to create interactive and visually appealing dashboards. These dashboards will offer a user-friendly interface for HR professionals to explore and interpret HR data seamlessly.

**Dashboard Interactivity:** Design Tableau dashboards that allow HR professionals to interactively explore HR metrics, enabling them to make informed decisions quickly.

**Key Components:**

**MySQL Database:**

* Establishing a relational database for the storage and organization of HR datasets.
* Ensuring data integrity and scalability to handle evolving HR data requirements.

**Exploratory Data Analysis (EDA):**

* Analyzing employee demographics, performance metrics, and other relevant factors to derive meaningful insights.
* Identifying trends and correlations through statistical analysis.

**Hypothesis Testing:**

* Formulating hypotheses related to HR metrics and conducting statistical tests using MySQL.
* Validating or refuting hypotheses to contribute to evidence-based decision-making.

**Tableau Visualization:**

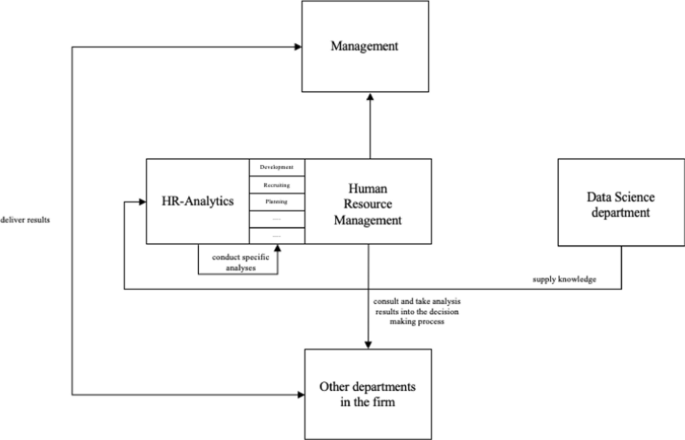
* Creating visually compelling and interactive dashboards using Tableau.
* Visualizing key HR metrics, performance indicators, and demographic distributions.

**Expected Outcomes:**

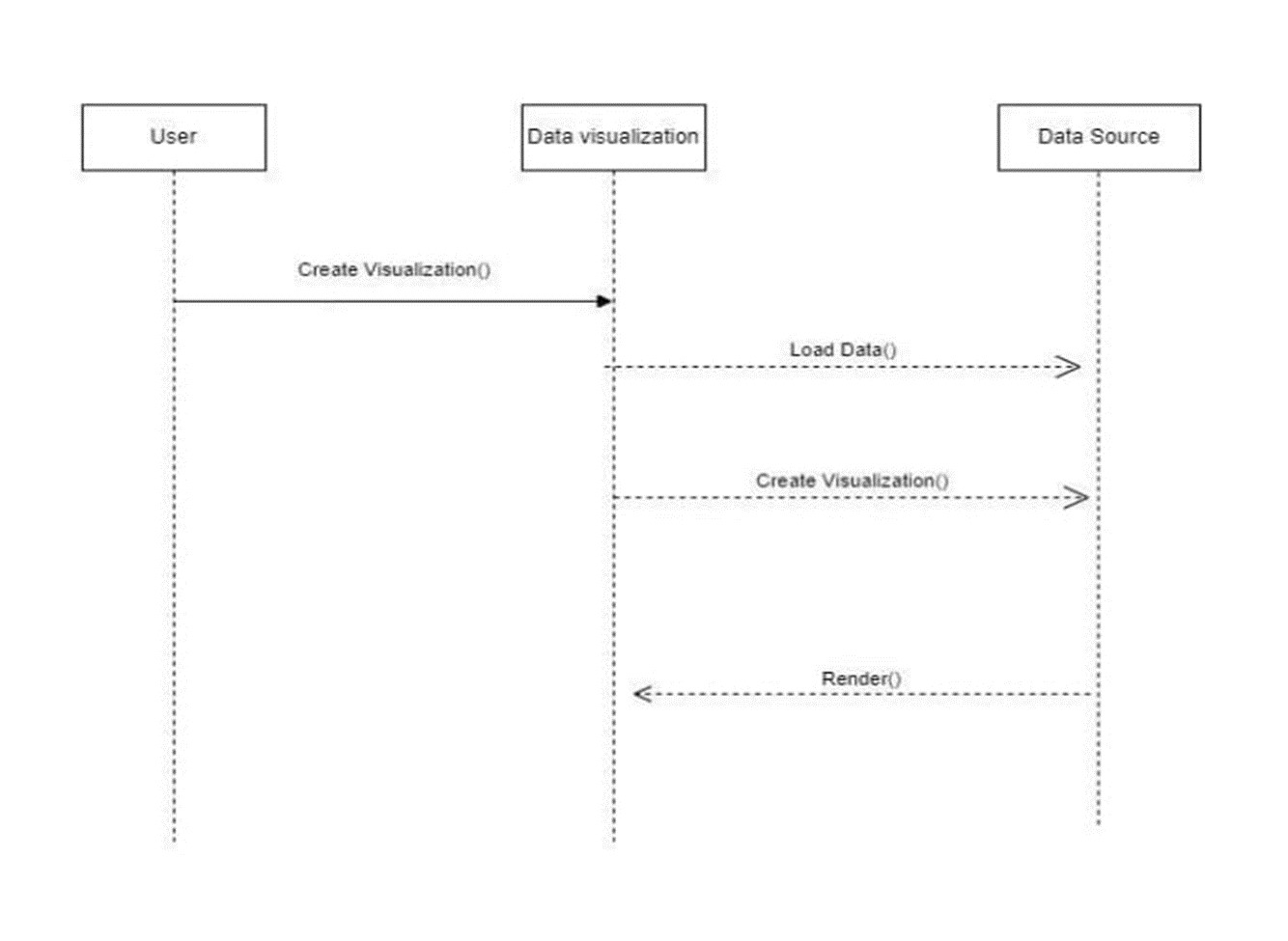
* **Informed Decision-Making:** Empower HR professionals with data-driven insights to make informed decisions related to workforce management, employee engagement, and performance optimization.
* **Enhanced Visualization:** Develop interactive Tableau dashboards that provide a user-friendly interface for exploring HR data, fostering a culture of data-driven decision-making.
* **Strategic HR Planning:** Enable the organization to implement strategic HR initiatives based on a comprehensive understanding of workforce dynamics.
* **Efficient Data Management:** Implement a robust MySQL database infrastructure for efficient storage, retrieval, and management of HR datasets.

The HR Analytics project using MySQL & Tableau software represents a strategic initiative to align HR practices with modern data analytics, ultimately contributing to the overall success and efficiency of the organization's human resource management

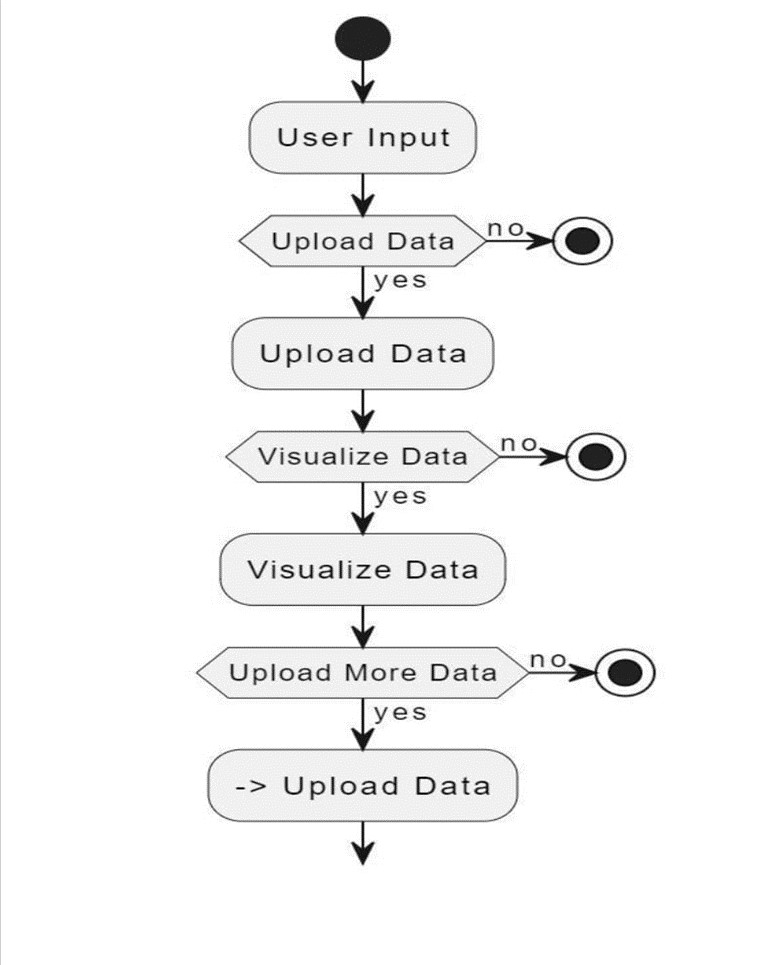
**CLASS DIAGRAM:-**



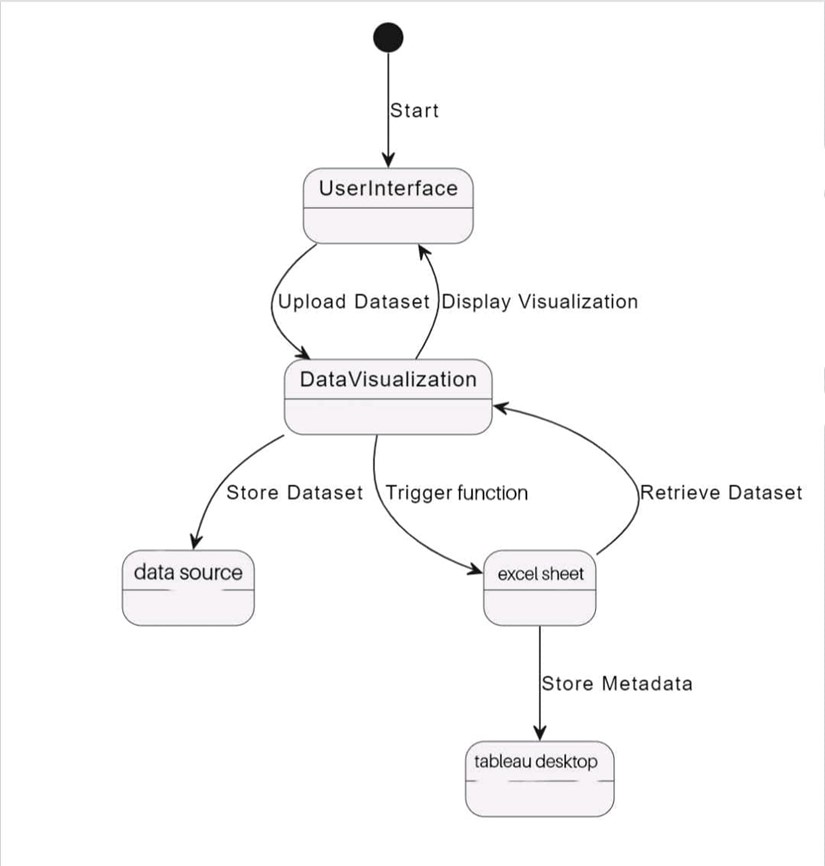
SEQUENCE DIAGRAM



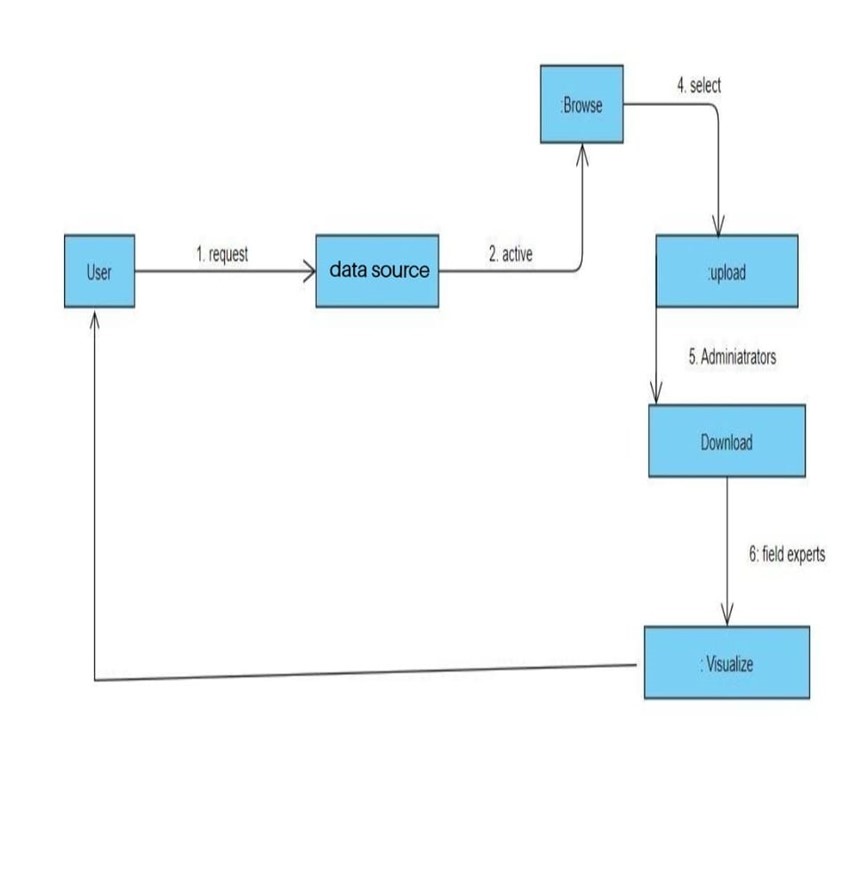
## ACTIVITY DIAGRAM



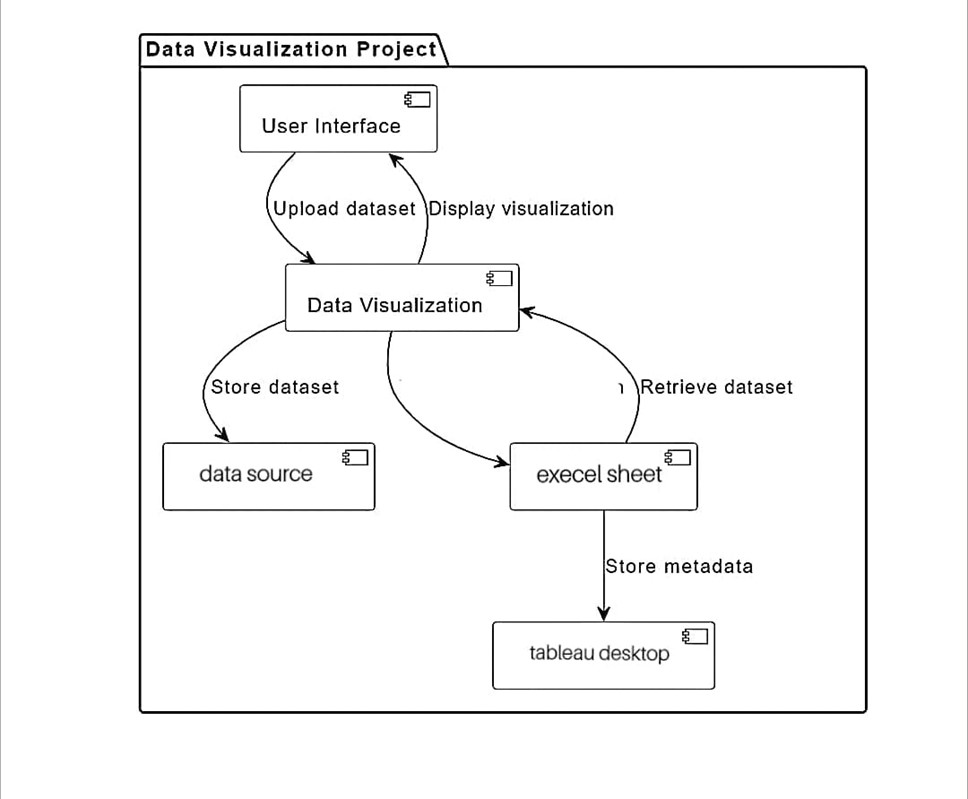
## STATE DIAGRAM



## COLLEBORATION DIAGRAM



## COMPONENT DIAGRAM



**EDA CODES WITH PROPER OUTPUT:-**

1. **COUNTS OF EMPLOYEE, ATTRITION, ATTRITION RATE, ACTIVE EMPLOYEES, AVG.AGE:-**

**EDA CODE:-**

# Assuming 'Employee Count' is a column in your DataFrame

# Check if 'Employee Count' column is present in the DataFrame

if 'Employee Count' in df.columns:

# Calculate the sum of 'Employee Count'

total\_employee\_sum = df['Employee Count'].sum()

# Display the sum of 'Employee Count'

print(f'Sum of Employee Count: {total\_employee\_sum}')

else:

print("Error: 'Employee Count' column not found in the DataFrame.")

#--------------------------------------------------------------------------------------------Sum of Employee Count

# Assuming 'Attrition' is a column in your DataFrame

# Check if 'Attrition' column is present in the DataFrame

if 'Attrition' in df.columns:

# Calculate the attrition rate

attrition\_count = df['Attrition'].value\_counts().get('Yes', 0)

total\_employees = len(df)

attrition\_rate = (attrition\_count / total\_employees) \* 100 if total\_employees > 0 else 0

# Display the attrition rate

print(f'Attrition Rate: {attrition\_rate:.2f}%')

else:

print("Error: 'Attrition' column not found in the DataFrame.")

# Assuming 'Attrition' is a column in your DataFrame

#--------------------------------------------------------------------------------------------Attrition Rate

# Check if 'Attrition' column is present in the DataFrame

if 'Attrition' in df.columns:

# Calculate the count of active employees

attrition\_count = df['Attrition'].value\_counts().get('Yes', 0)

total\_employees = len(df)

active\_employees = total\_employees - attrition\_count

# Display the count of active employees

print(f'Active Employees: {active\_employees}')

else:

print("Error: 'Attrition' column not found in the DataFrame.")

#--------------------------------------------------------------------------------------------Active Employees

# Assuming 'Age' is a column in your DataFrame

# Check if 'Age' column is present in the DataFrame

if 'Age' in df.columns:

# Calculate the average age

average\_age = df['Age'].mean()

# Display the average age

print(f'Average Age: {average\_age:.2f}')

else:

print("Error: 'Age' column not found in the DataFrame.")

#--------------------------------------------------------------------------------------------Average Age

# Assuming 'Attrition' is a column in your DataFrame

# Check if 'Attrition' column is present in the DataFrame

if 'Attrition' in df.columns:

# Calculate the count of employees with attrition

attrition\_count = df['Attrition'].value\_counts().get('Yes', 0)

# Display the count of attrition

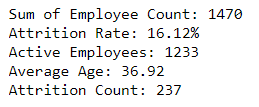
print(f'Attrition Count: {attrition\_count}')

else:

print("Error: 'Attrition' column not found in the DataFrame.")

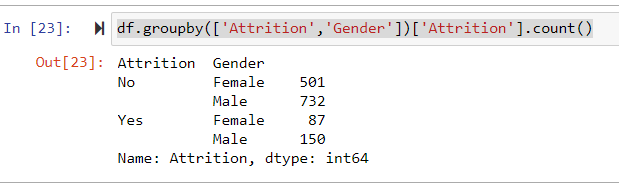
#------------------------------------------------------------------------------------------ Attrition Count

**OUTPUT OBTAINED:-**



1. **ATTRITION BY GENDER:-**

**CODE AND COUNT OUTPUT:-**



**CODE FOR GRAPH:-**

df.groupby(['Attrition','Gender'])['Attrition'].count()

import seaborn as sns

import matplotlib.pyplot as plt

import pandas as pd

# Assuming df is your DataFrame

# Group by 'Attrition' and 'Gender' and count occurrences

count\_series = df.groupby(['Attrition', 'Gender'])['Attrition'].count()

# Convert the MultiIndex Series to DataFrame for easy plotting

count\_df = count\_series.reset\_index(name='Count')

# Create a countplot

plt.figure(figsize=(8, 6))

sns.set(style="whitegrid")

sns.barplot(x='Attrition', y='Count', hue='Gender', data=count\_df, palette='viridis')

# Add labels and a title

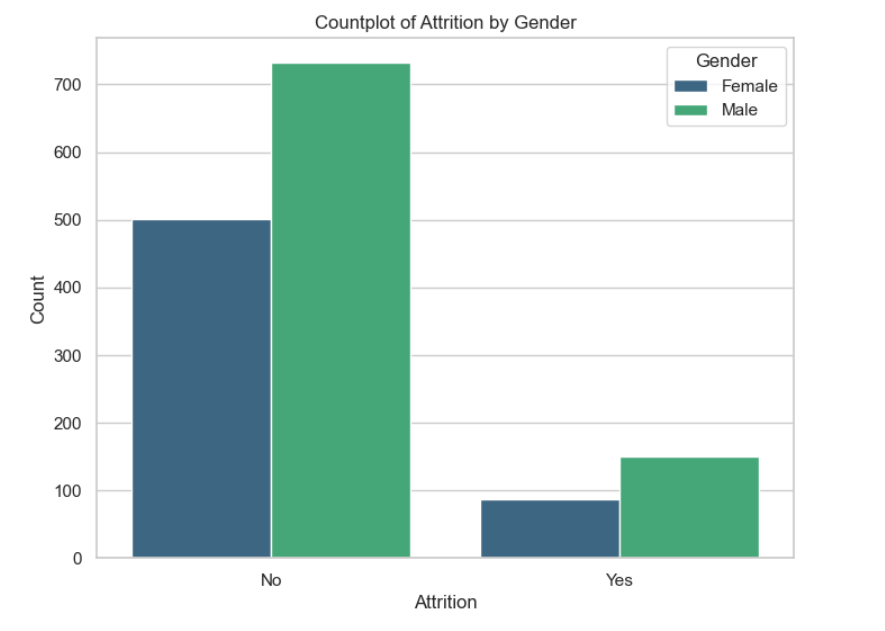
plt.xlabel('Attrition')

plt.ylabel('Count')

plt.title('Countplot of Attrition by Gender')

# Show the plot

plt.show()



1. **DEPARTMENT WISE ATTRTIION:-**

**CODE:-**

import seaborn as sns

import matplotlib.pyplot as plt

# Assuming 'Department' and 'Attrition' are columns in your DataFrame

# Assuming 'Attrition' column has values like 'Yes' or 'No'

# Create a DataFrame with department-wise attrition counts

attrition\_by\_department = df.groupby(['Department', 'Attrition']).size().reset\_index(name='Count')

# Filter for only the 'Yes' (attrition) values

attrition\_yes = attrition\_by\_department[attrition\_by\_department['Attrition'] == 'Yes']

# Plotting a pie chart

plt.figure(figsize=(8, 8))

plt.pie(attrition\_yes['Count'], labels=attrition\_yes['Department'], autopct='%1.1f%%', startangle=140, colors=sns.color\_palette('pastel'))

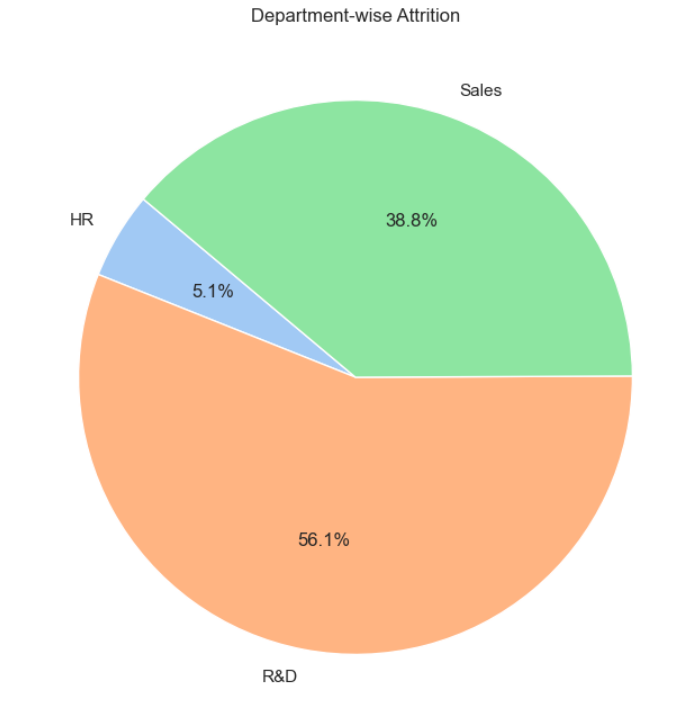
# Add a title

plt.title('Department-wise Attrition')

# Show the plot

plt.show()

**GRAPH OUTPUT:-**



1. **NO OF EMPLOYEE BY AGE GROUP:-**

**CODE:-**

import matplotlib.pyplot as plt

# Assuming 'Age' is a column in your DataFrame

# Assuming 'Age' is a numeric column

# Define the age bins and labels

age\_bins = [18, 25, 35, 45, 55, 65, 100] # You can adjust these bins as needed

age\_labels = ['18-25', '26-35', '36-45', '46-55', '56-65', '66+']

# Create a new column with age groups

df['AgeGroup'] = pd.cut(df['Age'], bins=age\_bins, labels=age\_labels, right=False)

# Plotting a bar chart

plt.figure(figsize=(10, 6))

df['AgeGroup'].value\_counts().sort\_index().plot(kind='bar', color='skyblue')

# Add labels and a title

plt.xlabel('Age Group')

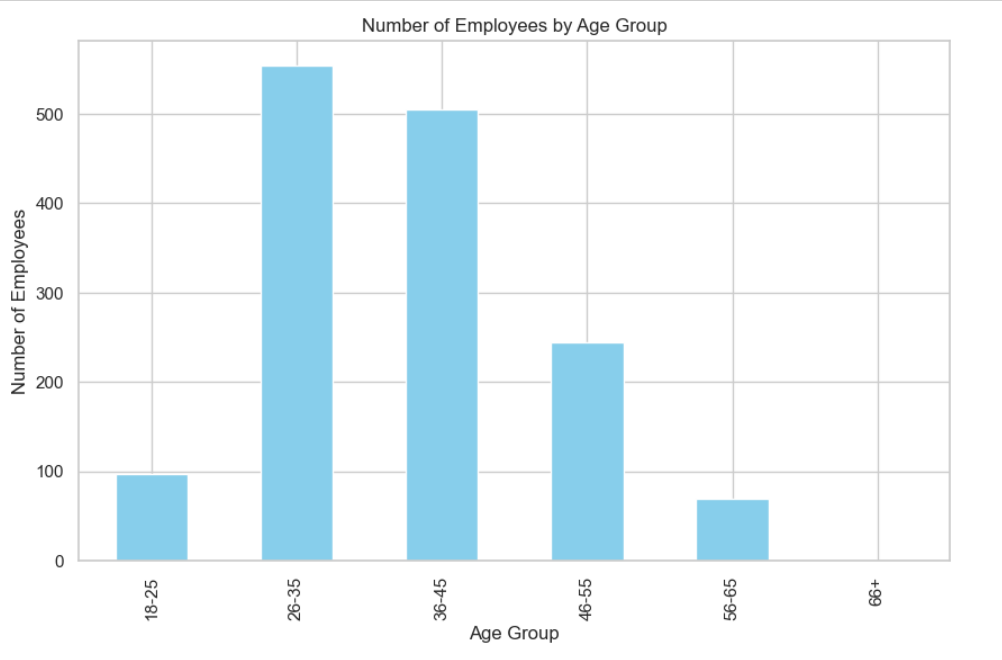
plt.ylabel('Number of Employees')

plt.title('Number of Employees by Age Group')

# Show the plot

plt.show()

**OUTPUT GRAPH:-**



1. **JOB SATISFACTION RATING:-**

**CODE:-**

import pandas as pd

# Read the CSV file

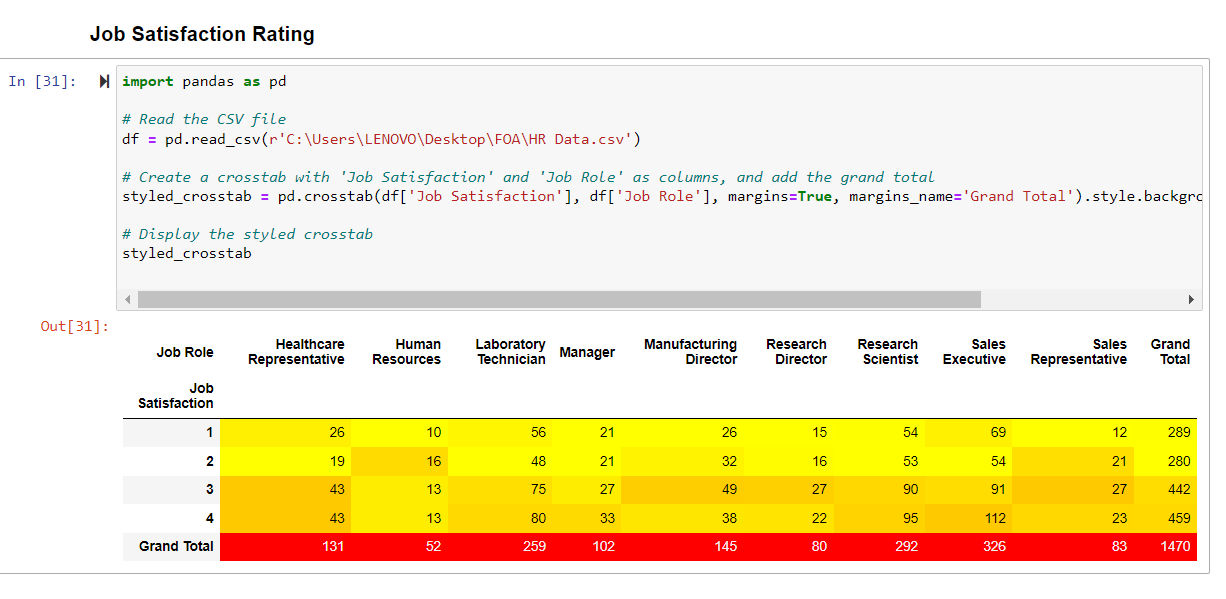
df = pd.read\_csv(r'C:\Users\LENOVO\Desktop\FOA\HR Data.csv')

# Create a crosstab with 'Job Satisfaction' and 'Job Role' as columns, and add the grand total

styled\_crosstab = pd.crosstab(df['Job Satisfaction'], df['Job Role'], margins=True, margins\_name='Grand Total').style.background\_gradient(cmap='autumn\_r')

# Display the styled crosstab

styled\_crosstab



1. **EDUCATION FIELD WISE ATTRITION:-**

**CODE:-**

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

# Read the CSV file

df = pd.read\_csv(r'C:\Users\LENOVO\Desktop\FOA\HR Data.csv')

# Create a crosstab to calculate the sum of attrition counts by education field

education\_field\_attrition\_sum = df.groupby('Education Field')['Attrition'].apply(lambda x: (x == 'Yes').sum()).reset\_index(name='AttritionSum')

# Sort by attrition count in descending order

education\_field\_attrition\_sum = education\_field\_attrition\_sum.sort\_values(by='AttritionSum', ascending=False)

# Plot the bar chart

plt.figure(figsize=(12, 8))

sns.set(style="whitegrid")

# Create a bar chart for education field-wise sum of attrition counts

sns.barplot(x='AttritionSum', y='Education Field', data=education\_field\_attrition\_sum, palette='viridis')

# Add labels and a title

plt.xlabel('Sum of Attrition Count')

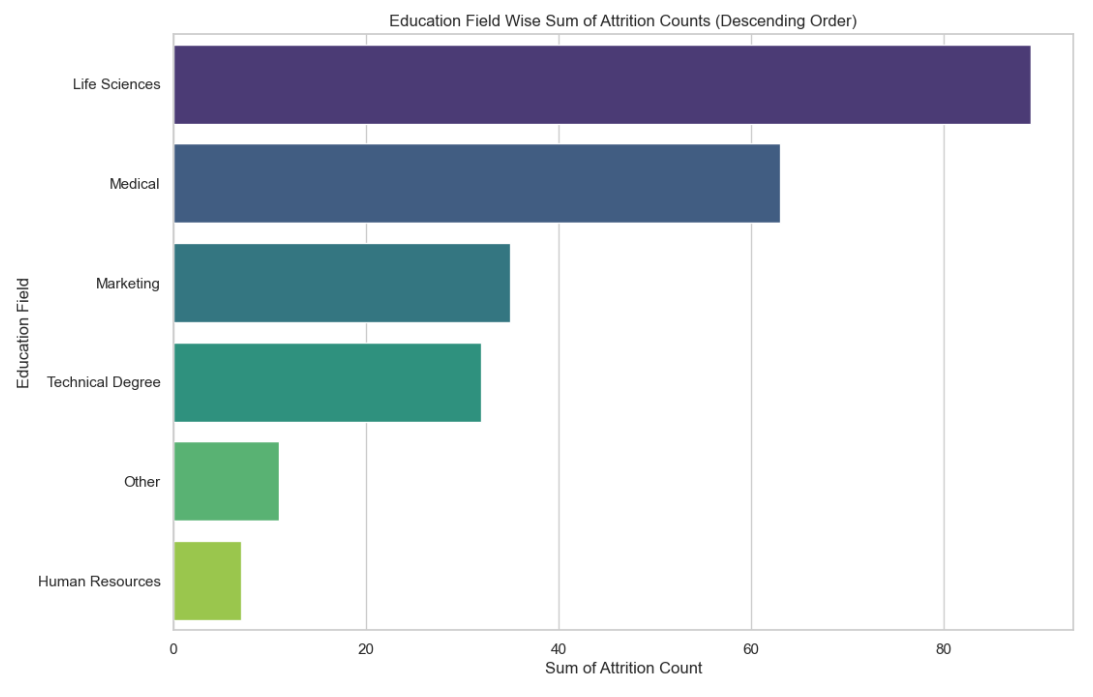
plt.ylabel('Education Field')

plt.title('Education Field Wise Sum of Attrition Counts (Descending Order)')

# Show the plot

plt.show()

**GRAPH OUTPUT:-**



ATTRITION RATE BY GENDER FOR DIFFERENT AGE GROUP:-

**CODE:-**

import seaborn as sns

import matplotlib.pyplot as plt

import pandas as pd

# Assuming 'Age', 'Gender', and 'Attrition' are columns in your DataFrame

# Check if all columns are present in the DataFrame

if all(col in df.columns for col in ['Age', 'Gender', 'Attrition']):

# Create a new column with age groups

age\_bins = [18, 25, 35, 45, 55, 65, 100]

age\_labels = ['18-25', '26-35', '36-45', '46-55', '56-65', '66+']

df['AgeGroup'] = pd.cut(df['Age'], bins=age\_bins, labels=age\_labels, right=False)

# Create a DataFrame with attrition counts by age group and gender

attrition\_by\_age\_gender = df.groupby(['AgeGroup', 'Gender', 'Attrition']).size().reset\_index(name='Count')

# Calculate attrition rates

attrition\_by\_age\_gender['AttritionRate'] = attrition\_by\_age\_gender.groupby(['AgeGroup', 'Gender'])['Count'].transform(lambda x: (x / x.sum()) \* 100)

# Create a lollipop chart

plt.figure(figsize=(12, 8))

sns.set(style="whitegrid")

# Lollipop chart for attrition rates by gender

sns.lineplot(x='AgeGroup', y='AttritionRate', hue='Gender', marker='o', data=attrition\_by\_age\_gender)

# Scatter plot to show individual data points

sns.scatterplot(x='AgeGroup', y='AttritionRate', hue='Gender', marker='o', data=attrition\_by\_age\_gender, legend=False)

# Add labels and a title

plt.xlabel('Age Group')

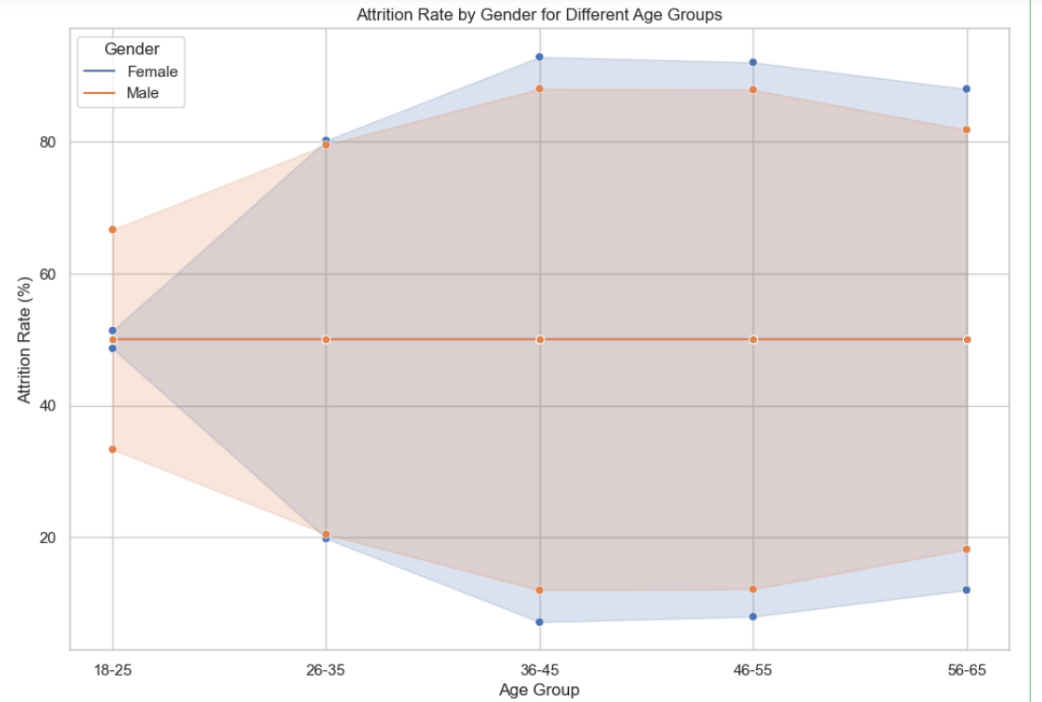
plt.ylabel('Attrition Rate (%)')

plt.title('Attrition Rate by Gender for Different Age Groups')

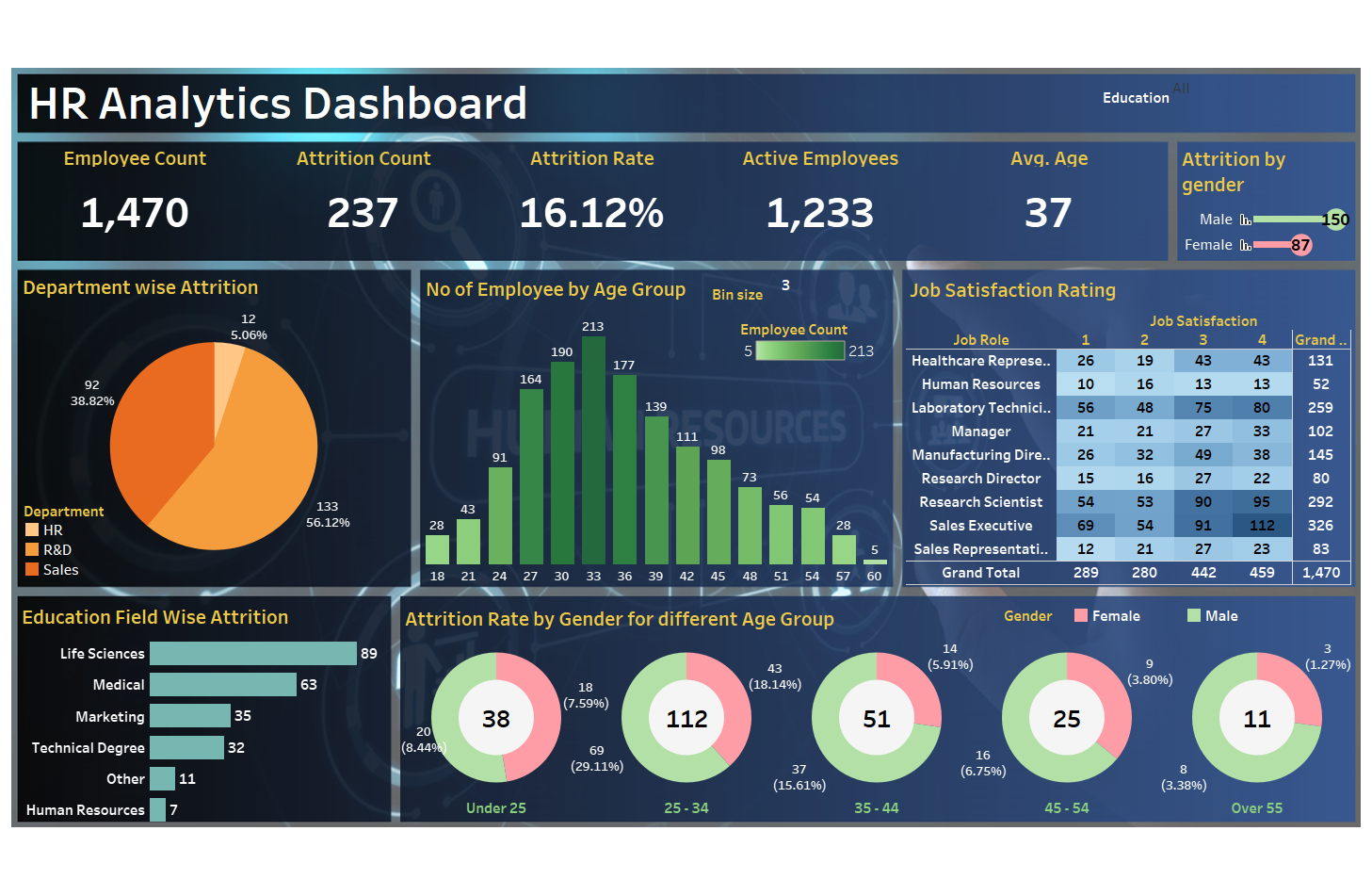
# Show the plot

plt.show()

**LOLLIPOP GRAPH OUTPUT:-**

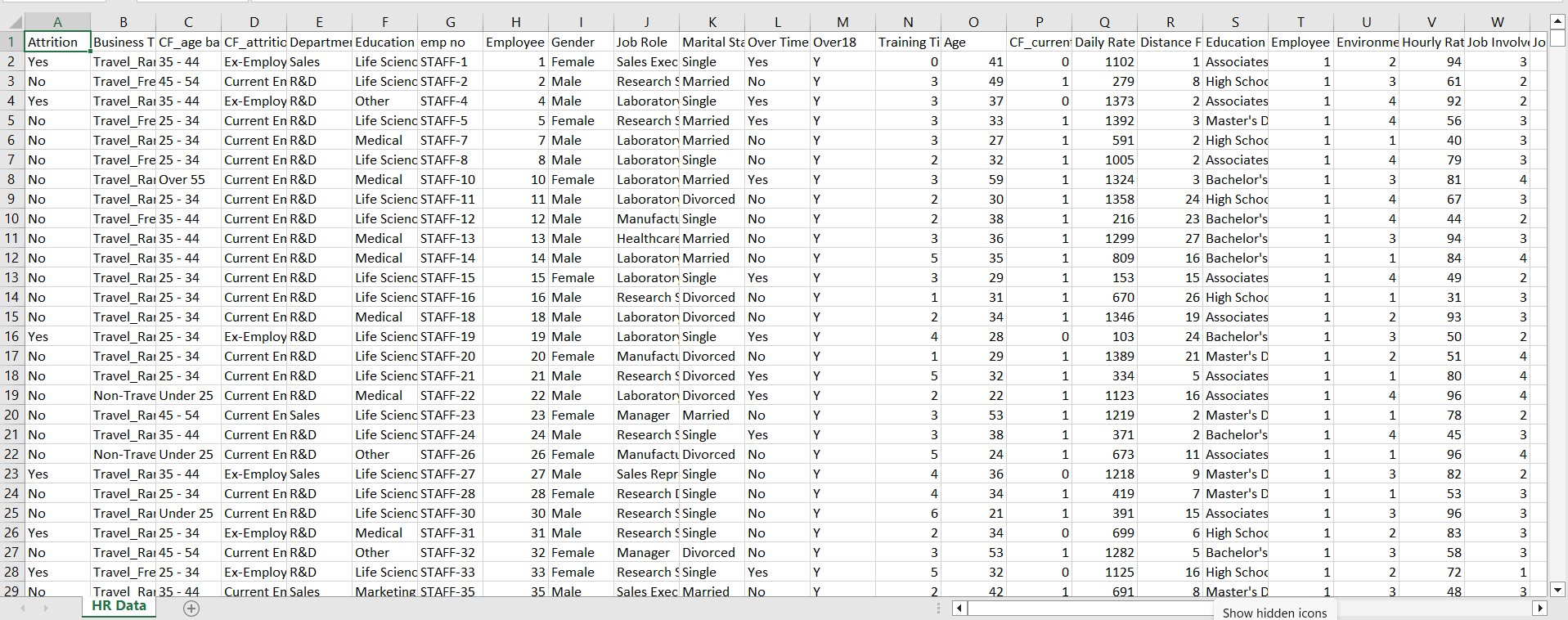


# TABLEAU DASHBOARD:-

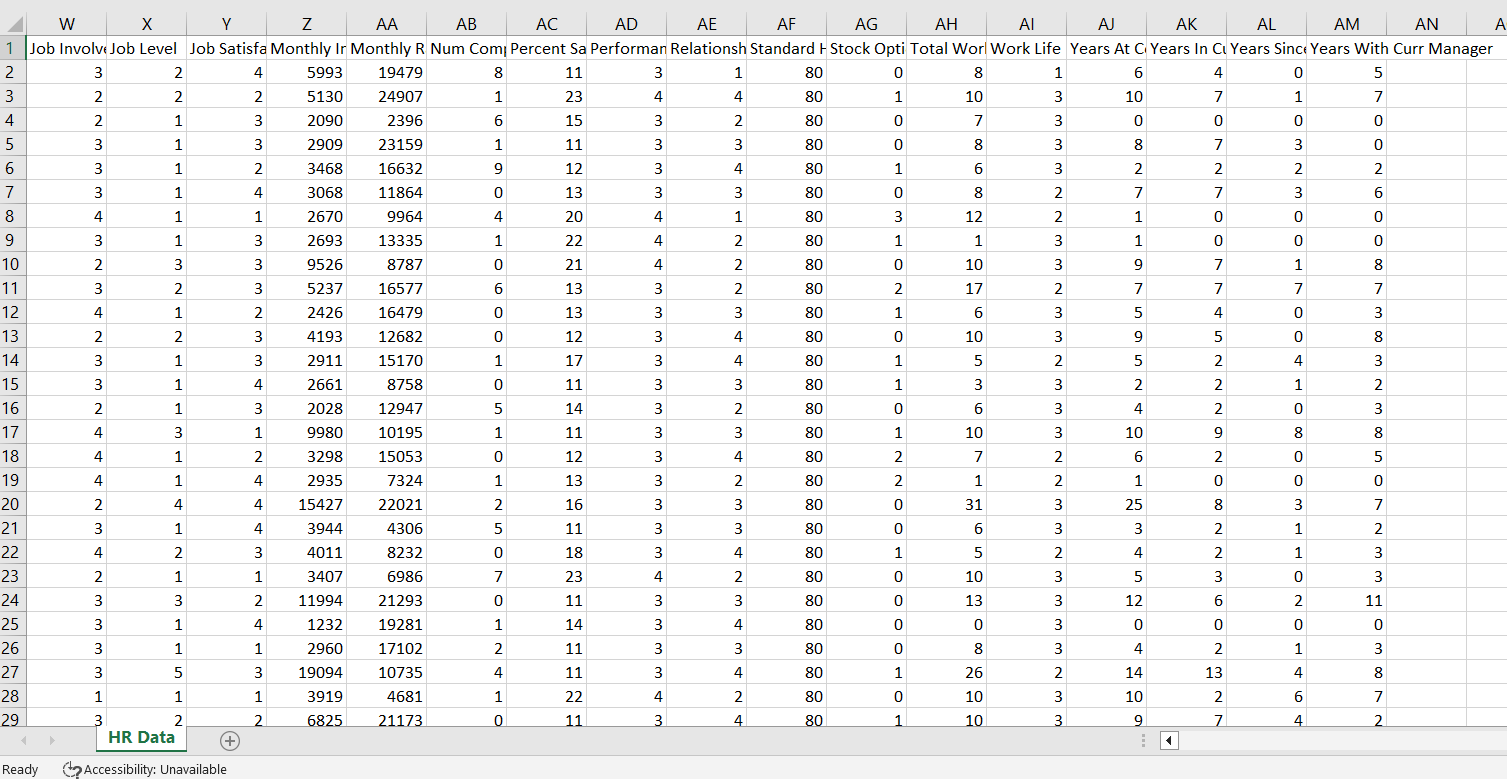


## **DATASET SCREENSHOTS:-**

SET 1 DATASET:-



SET 2 DATASET:-



## **CONCLUSION: -**

In conclusion, the integration of HR Analytics using MySQL and Tableau Software marks a pivotal advancement in human resource management. The project highlights the transformative impact of adopting a data-driven approach, providing comprehensive insights into workforce dynamics. MySQL's robust data infrastructure, coupled with Tableau's visualization capabilities, empowers HR professionals to make informed decisions based on a holistic understanding of employee demographics, performance metrics, and key HR indicators. The interactive dashboards created foster a collaborative and data-centric culture, enhancing accessibility and engagement. This strategic alignment contributes to efficient HR planning, optimization, and the establishment of a scalable data management system. In essence, the project sets the stage for a new era in HR Analytics, where the synergy of MySQL and Tableau propels organizations toward excellence in workforce management through precision and insight.

**Reference: -**

<https://www.geeksforgeeks.org/tableau-tutorial/>

<https://docs.google.com/spreadsheets/d/1-1Ldoe-DwZTL77tdMtRgZAIzeAzs0jh3/edit#gid=2089618187>

https://www.w3schools.com/mysql/