NEXUS PROJECT - 1

NAME : BANDREDDY NITHIN

Project Title: Employee Analysis (Entry Level)

Objective: Conduct basic data analytics and visualization on the Employee Data Set to gain insights into the workforce. No machine learning is required for this entry-level project, and tools such as PowerBI, Tableau, or any preferred tool can be used for analysis and visualization.

Tools used : PowerBI and GoogleColab

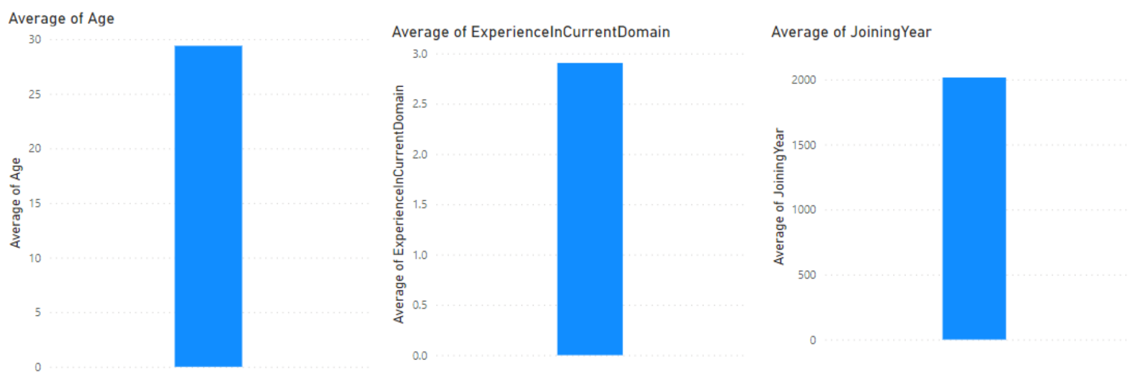
Summary of CSV :

- **Education:** Most of the individuals have a Bachelor's degree (3601 out of 4653).
- **JoiningYear:** The individuals joined between the years 2012 and 2018. The average joining year is around 2015.
- **City:** Most of the individuals are from Bangalore (2228 out of 4653).
- **PaymentTier:** The payment tier ranges from 1 to 3, with an average of approximately 2.7.
- **Age:** The age of the individuals ranges from 22 to 41, with an average age of approximately 29.4.
- **Gender:** Most of the individuals are male (2778 out of 4653).
- **EverBenched:** Most of the individuals have never been benched (4175 out of 4653).
- **ExperienceInCurrentDomain:** The experience in the current domain ranges from 0 to 7, with an average of approximately 2.9.
- **LeaveOrNot:** The average of this column is approximately 0.34, indicating that about 34% of the individuals are likely to leave.

Task-1 :

Using PowerBi :

TASK - 1
Descriptive Statistics: Calculate and present the average age, average experience in the current domain, and average years since joining for the entire workforce.



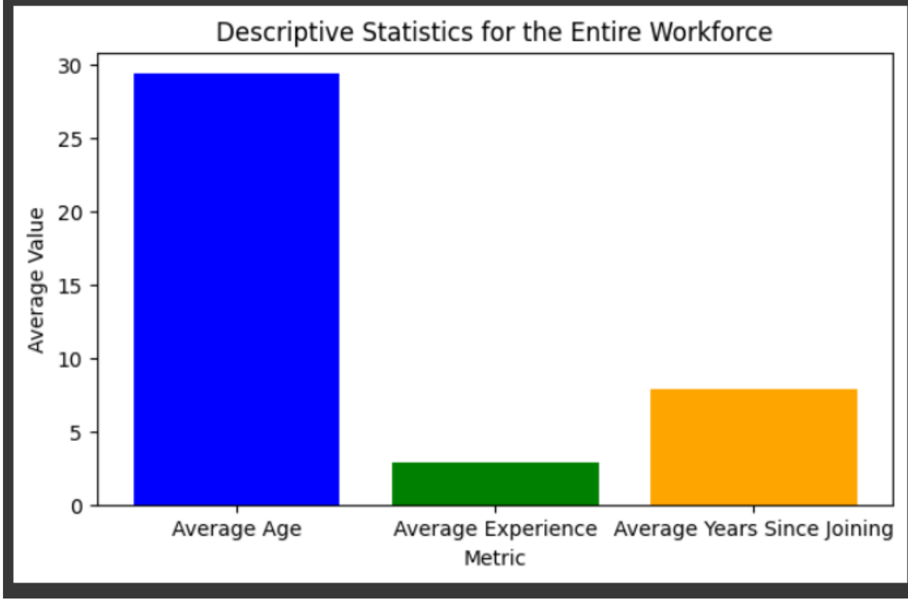
Average of Age	Average of ExperienceInCurrentDomain	Average of JoiningYear
29.39	2.91	2015.06

Using Google Colab :

```python
# Calculate averages
avg_age = df['Age'].mean()
avg_experience = df['ExperienceInCurrentDomain'].mean()
avg_years_since_joining = 2023 - df['JoiningYear'].mean()

# Display results
print(f"Average Age: {avg_age:.2f} years")
print(f"Average Experience in Current Domain: {avg_experience:.2f} years")
print(f"Average Years Since Joining: {avg_years_since_joining:.2f} years")
```

```
Average Age: 29.39 years
Average Experience in Current Domain: 2.91 years
Average Years Since Joining: 7.94 years
```

```python
# Plot bar graph
plt.figure(figsize=(7, 4))
plt.bar(['Average Age', 'Average Experience', 'Average Years Since Joining'],
        [avg_age, avg_experience, avg_years_since_joining], color=['blue', 'green', 'orange'])
plt.title('Descriptive Statistics for the Entire Workforce')
plt.xlabel('Metric')
plt.ylabel('Average Value')
plt.show()
```

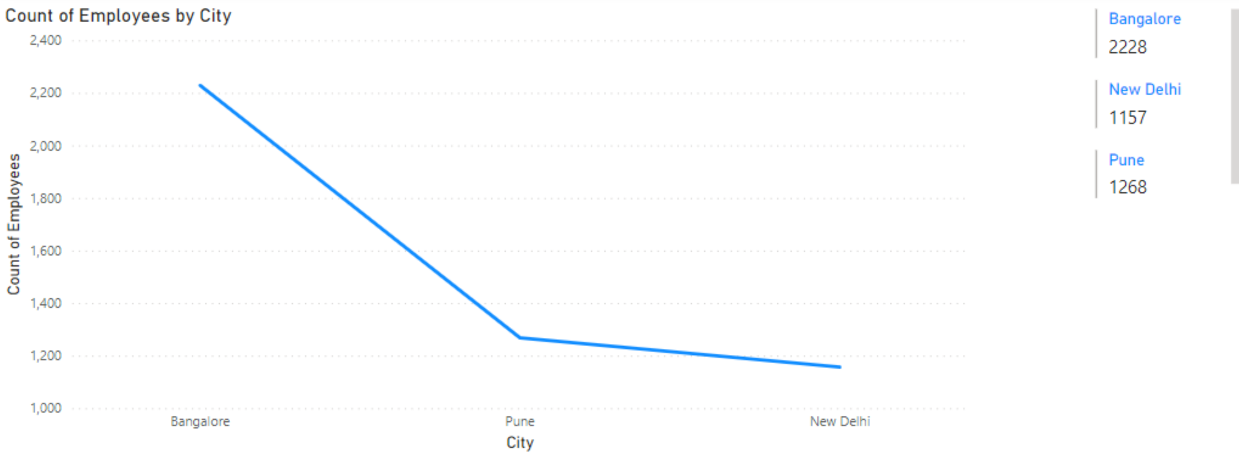


Task – 2 :

Using PowerBi :

TASK-2

City-wise Analysis: Explore the distribution of employees across different cities using charts or maps to identify significant locations.



At 2,228, Bangalore had the highest Count of City and was 92.57% higher than New Delhi, which had the lowest Count of City at 1,157.

Bangalore had the highest Count of City at 2,228, followed by Pune at 1,268 and New Delhi at 1,157.

Bangalore accounted for 47.88% of Count of City.

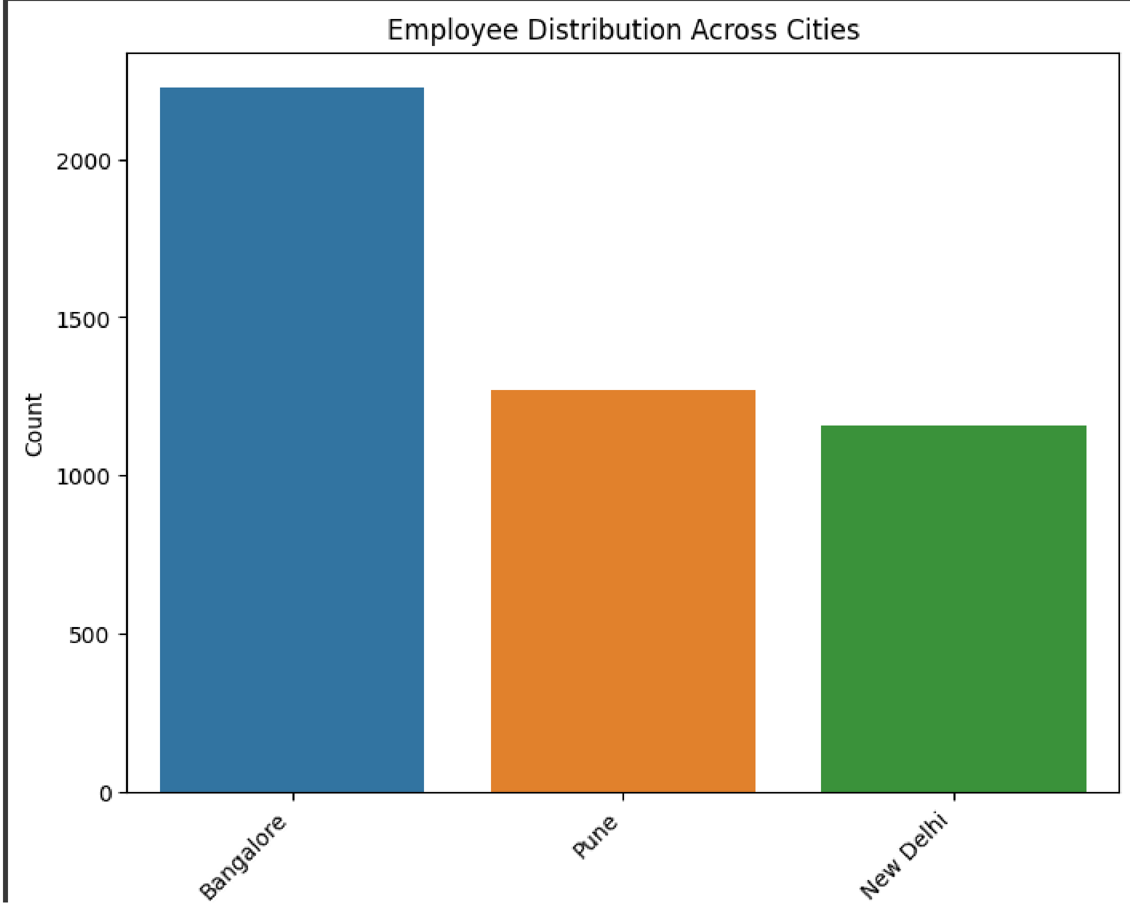
Bangalore had 2,228 Count of City, Pune had 1,268, and New Delhi had 1,157.

Using Google Colab :

```
# Calculate city-wise distribution
city_distribution = df['City'].value_counts()
print("City-wise Distribution:")
print(city_distribution)
```

```
City-wise Distribution:
Bangalore    2228
Pune         1268
New Delhi    1157
Name: City, dtype: int64
```

```
# Plot city-wise distribution
plt.figure(figsize=(8, 6))
sns.countplot(x='City', data=df, order=city_distribution.index)
plt.title('Employee Distribution Across Cities')
plt.xlabel('City')
plt.ylabel('Count')
plt.xticks(rotation=45, ha='right')  # Rotate x-axis labels for better readability
plt.show()
```

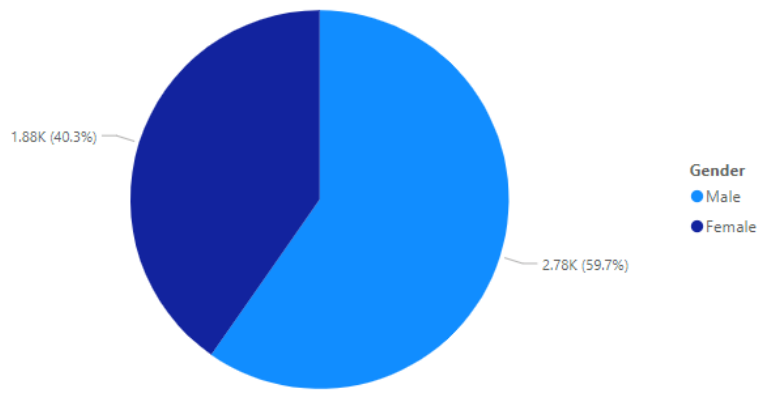


Task – 3 :

Using PowerBI :

Task - 3
Gender Diversity: Visualize the gender distribution within the company to understand the level of gender diversity

Diversity by Gender



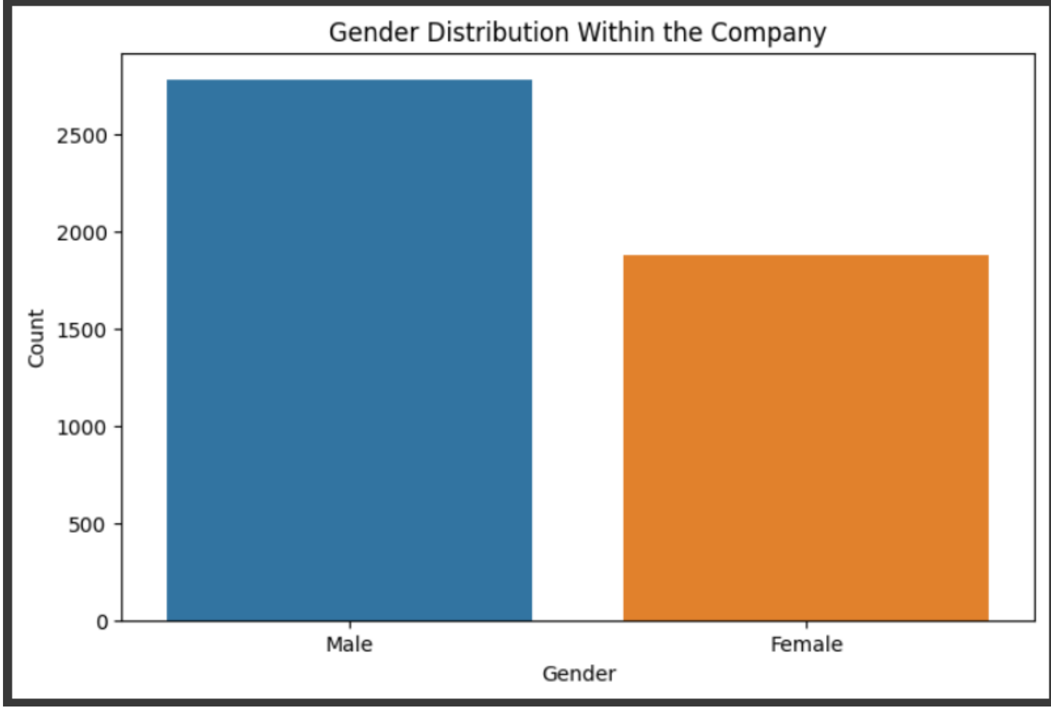
Gender	Count of Employees
Male	2778
Female	1875
Total	**4653**

Using Google Colab :

```
# Calculate gender distribution
gender_distribution = df['Gender'].value_counts()
print("Gender Distribution:")
print(gender_distribution)
```

```
Gender Distribution:
Male      2778
Female    1875
Name: Gender, dtype: int64
```

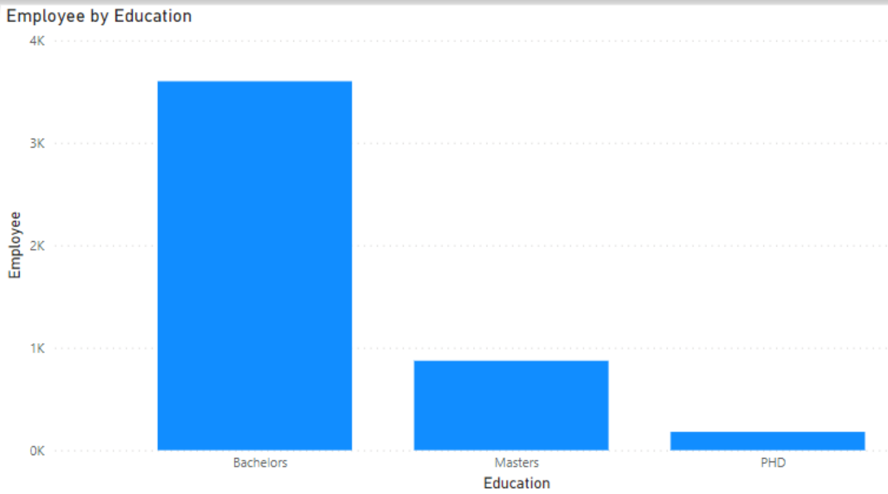
```python
# Plot gender distribution
plt.figure(figsize=(8, 5))
sns.countplot(x='Gender', data=df)
plt.title('Gender Distribution Within the Company')
plt.xlabel('Gender')
plt.ylabel('Count')
plt.show()
```



Task – 4 :

Using PowerBI :

Task - 4
Educational Background: Analyze and present the most common degrees and institutions among employees.



Employee	Education
3601	Bachelors
873	Masters
179	PHD
4653	

At 3601, Bachelors had the highest Employee and was 1,911.73% higher than PHD, which had the lowest Employee at 179.

Bachelors had the highest Employee at 3601, followed by Masters at 873 and PHD at 179.

Bachelors accounted for 77.39% of Employee.

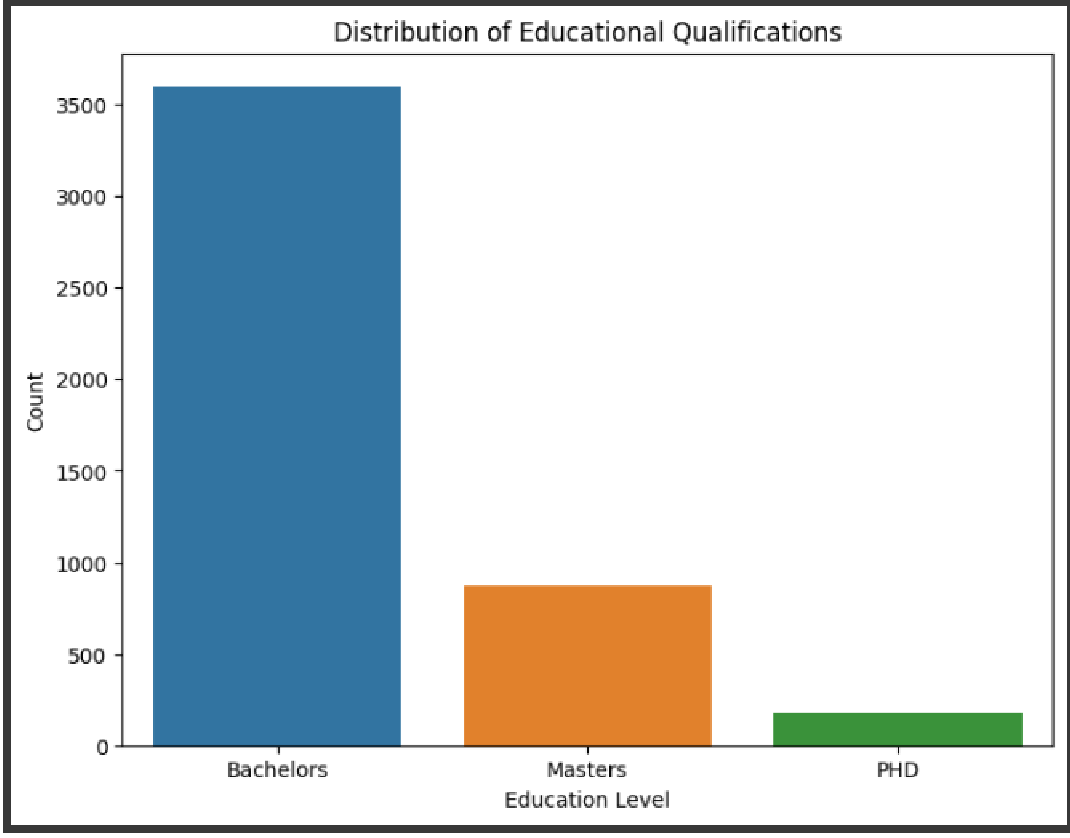
Bachelors had 3601 Employee, Masters had 873, and PHD had 179.

Using GoogleColab :

```
# Calculate and print most common degree with its count
most_common_degree_counts = df['Education'].value_counts()
most_common_degree = most_common_degree_counts.idxmax()
most_common_degree_count = most_common_degree_counts.max()

print(f"Most Common Degree: {most_common_degree}")
print(f"Count: {most_common_degree_count}")
```

```
Most Common Degree: Bachelors
Count: 3601
```

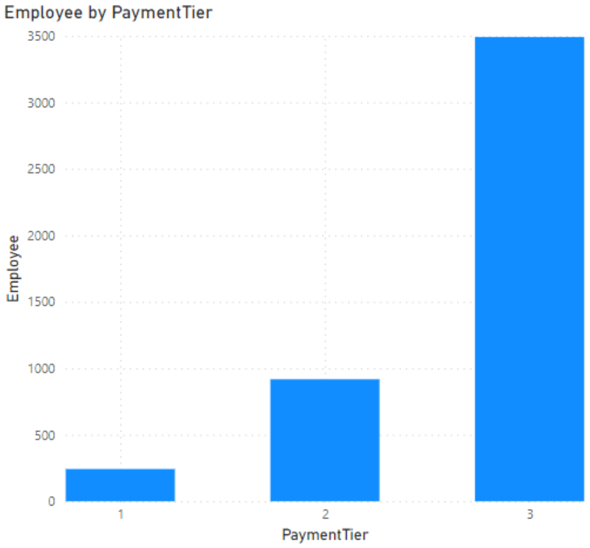
```
# Plot educational background distribution
plt.figure(figsize=(8, 6))
sns.countplot(x='Education', data=df, order=df['Education'].value_counts().index)
plt.title('Distribution of Educational Qualifications')
plt.xlabel('Education Level')
plt.ylabel('Count')
plt.show()
```



Task 5 :

Using PowerBI :

Task - 5
Salary Tier Analysis: Investigate the distribution of employees across different payment tiers and visualize the salary distribution to identify trends.



PaymentTier	Employee
1	243
2	918
3	3492
Total	**4653**

At 3492, 3 had the highest Employee and was 1,337.04% higher than 1, which had the lowest Employee at 243.

3 had the highest Employee at 3492, followed by 2 at 918 and 1 at 243.

3 accounted for 75.05% of Employee.

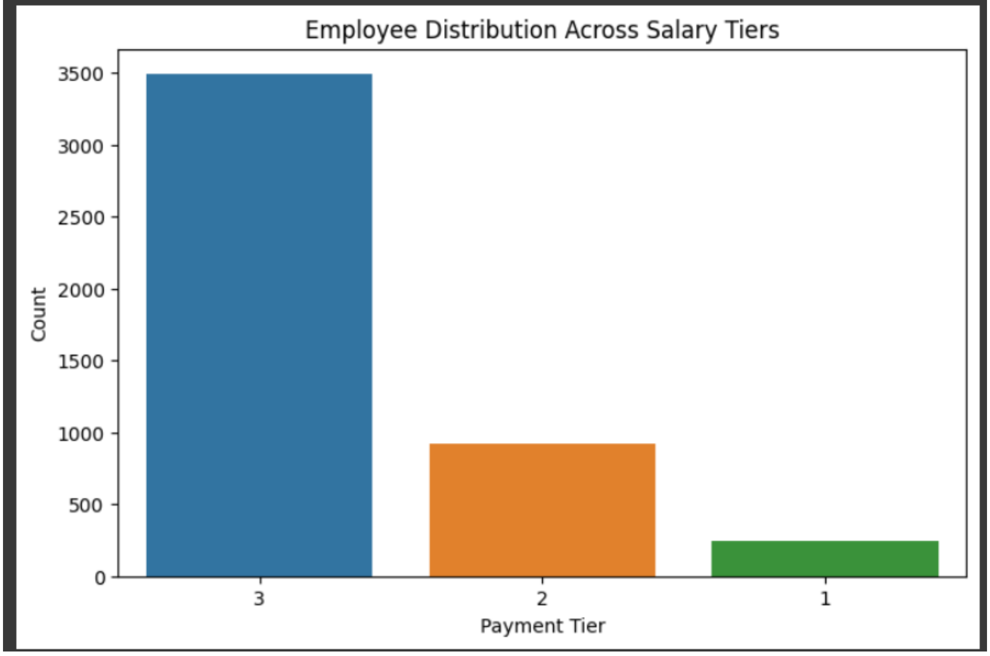
3 had 3492 Employee, 1 had 243, and 2 had 918.

Using GoogleColab :

```
# Calculate salary tier distribution
salary_tier_distribution = df['PaymentTier'].value_counts()
# Print salary tier distribution values
print("Salary Tier Distribution:")
print(salary_tier_distribution)
```

```
Salary Tier Distribution:
3    3492
2     918
1     243
Name: PaymentTier, dtype: int64
```

```python
# Plot salary tier distribution
plt.figure(figsize=(8, 5))
sns.countplot(x='PaymentTier', data=df, order=salary_tier_distribution.index)
plt.title('Employee Distribution Across Salary Tiers')
plt.xlabel('Payment Tier')
plt.ylabel('Count')
plt.show()
```

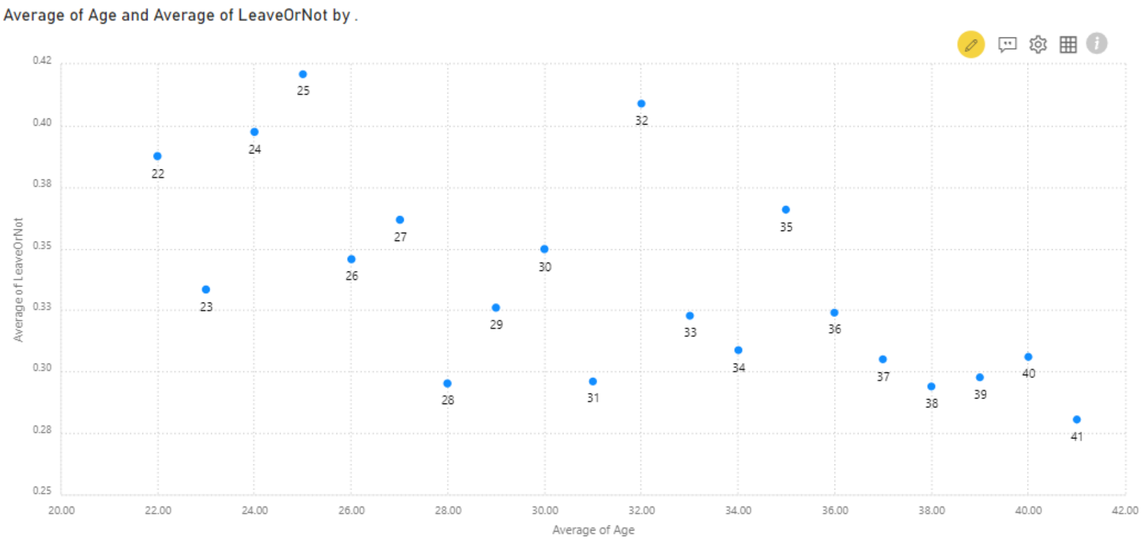


Task – 6 :

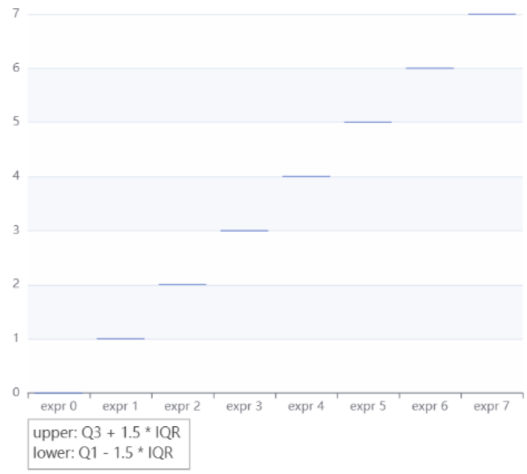
Using PowerBI :

Task - 6

Leave Analysis: Explore factors influencing leave decisions. Visualize the correlation between factors like age, experience, and the decision to take leave.

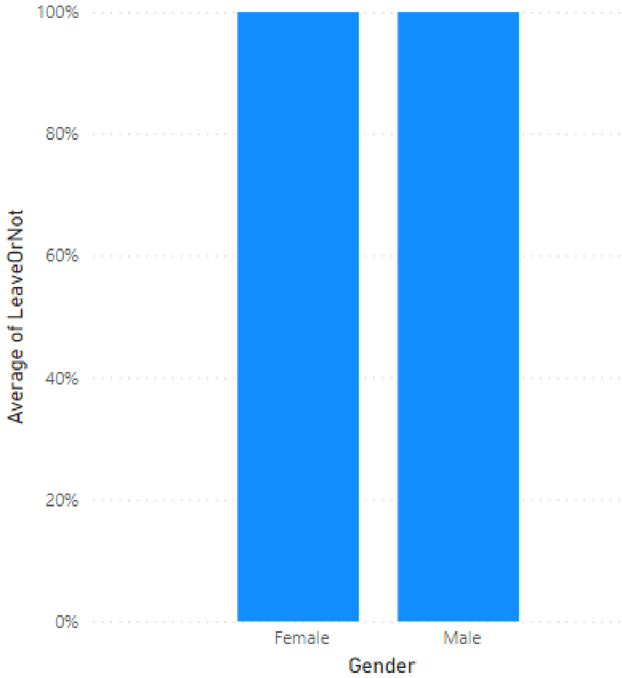


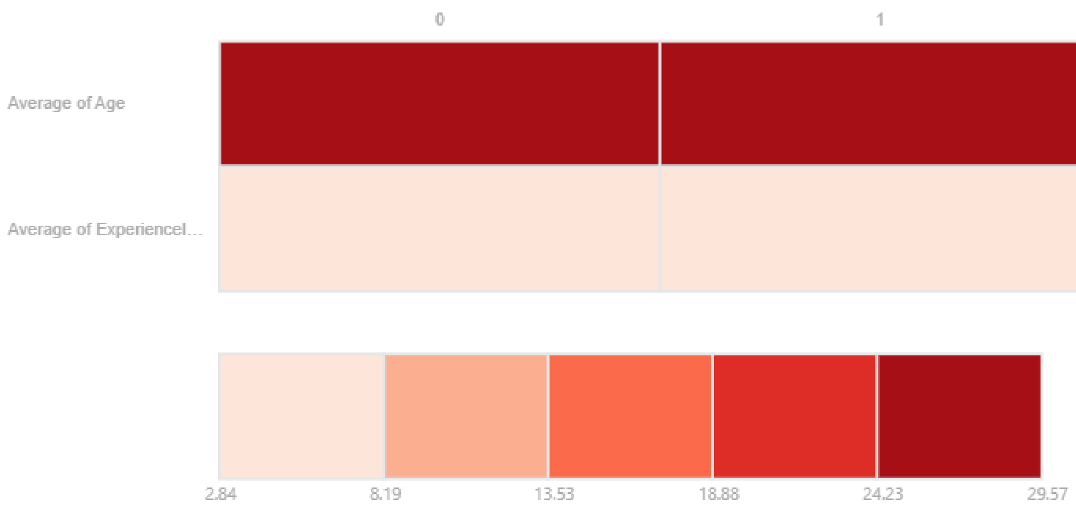
ExperienceInCurrentDomain



LeaveOrNot

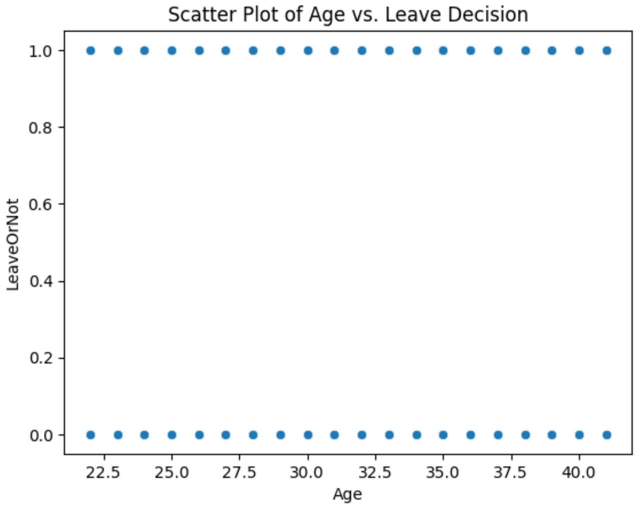
Average of LeaveOrNot by Gender

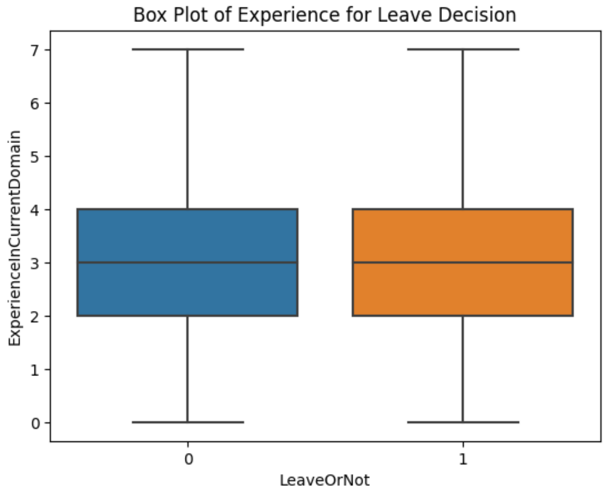


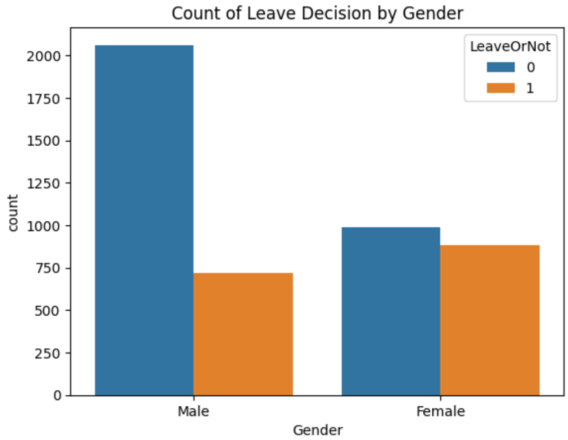


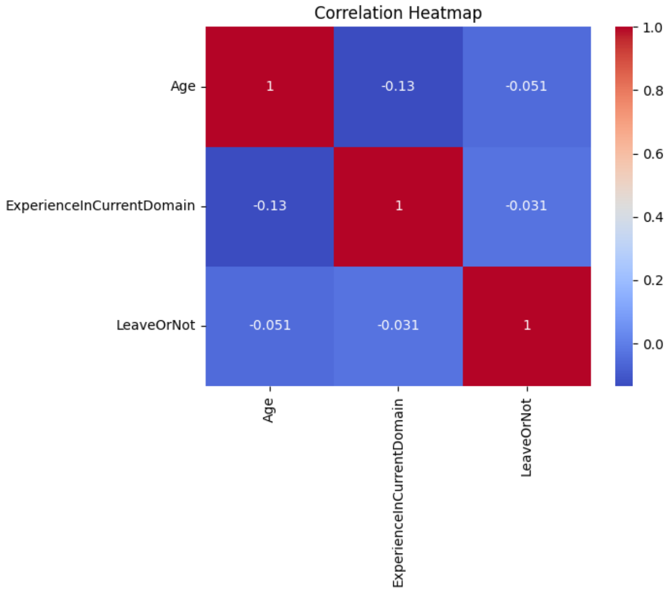
Using GoogleColab :

```
# Scatter plot
sns.scatterplot(x='Age', y='LeaveOrNot', data=df)
plt.title('Scatter Plot of Age vs. Leave Decision')
plt.show()

# Box plot
sns.boxplot(x='LeaveOrNot', y='ExperienceInCurrentDomain', data=df)
plt.title('Box Plot of Experience for Leave Decision')
plt.show()

# Grouped bar chart
sns.countplot(x='Gender', hue='LeaveOrNot', data=df)
plt.title('Count of Leave Decision by Gender')
plt.show()

# Heatmap for correlation
corr_matrix = df[['Age', 'ExperienceInCurrentDomain', 'LeaveOrNot']].corr()
sns.heatmap(corr_matrix, annot=True, cmap='coolwarm')
plt.title('Correlation Heatmap')
plt.show()
```


Scatter Plot of Age vs. Leave Decision
LeaveOrNot
1.0
0.8
0.6
0.4
0.2
0.0
22.5
25.0
27.5
30.0
32.5
35.0
37.5
40.0
Age


Box Plot of Experience for Leave Decision
ExperienceInCurrentDomain
7
6
5
4
3
2
1
0
0
1
LeaveOrNot


Count of Leave Decision by Gender
count
2000
1750
1500
1250
1000
750
500
250
0
LeaveOrNot
0
1
Male
Female
Gender


Correlation Heatmap
Age
ExperienceInCurrentDomain
LeaveOrNot
1
-0.13
-0.051
-0.13
1
-0.031
-0.051
-0.031
1
1.0
0.8
0.6
0.4
0.2
0.0
Age
ExperienceInCurrentDomain
LeaveOrNot

Summary of the Tasks :

1. **Task - 1: Descriptive Statistics**
    - Average Age: 29.39 years
    - Average Experience in Current Domain: 2.91 years
    - Average Years Since Joining: 7.94 years
2. **Task - 2: City-wise Distribution**
    - Bangalore: 2228
    - Pune: 1268
    - New Delhi: 1157
3. **Task - 3: Gender Distribution**
    - Male: 2778
    - Female: 1875
4. **Task - 4: Most Common Degree**
    - Most Common Degree: Bachelors
    - Count: 3601
5. **Task - 5: Salary Tier Distribution**
    - Tier 3: 3492
    - Tier 2: 918
    - Tier 1: 243
6. **Task - 6: Visualizations**
    - **Scatter Plot of Age vs. Leave Decision:**
        - X-Axis: Age (ranging from 22.5 to 40)
        - Y-Axis: Leave Decision (labeled as 0 and 1)
        - Scatter plot visually shows the relationship between age and the decision to take leave**.**
    - **Box Plot of Experience for Leave Decision:**
        - X-Axis: Leave Decision (categorized as 0 and 1)
        - Y-Axis: Experience (specifically for experience = 1 and leave_or_not = 0)
    - **Grouped Bar Chart of Leave Decision by Gender:**
        - X-Axis: Gender
        - Y-Axis: Count of Leave Decision
        - Visualizes the count of leave decisions for each gender, categorized by 0 and 1.

- **Correlation Heatmap:**
  - Heatmap displays correlation coefficients between age, experience, and leave_or_not.
- **Values:**
  - Age vs. Experience: -0.13
  - Age vs. Leave_or_Not: -0.051
  - Experience vs. Leave_or_Not: -0.031

The summary provides an overview of descriptive statistics, distributions, and visualizations related to age, experience, leave decisions, city, gender, education, and salary tiers. It gives a comprehensive view of employee characteristics and their relationships within the dataset.