Name- Vedant Roy

Registered Email ID- vedantroy3@gmail.com

Report: Analysis of Employee Data

Objective

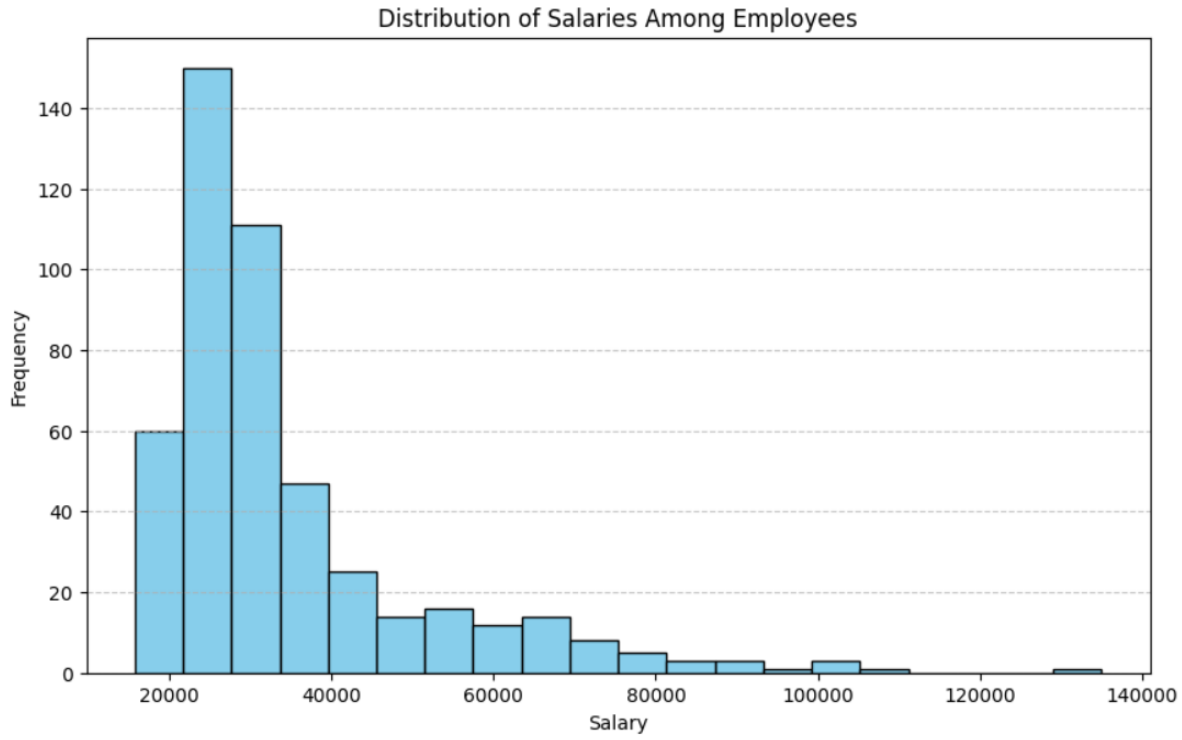
The objective of this analysis is to gain insights into the employee dataset and present a comprehensive overview of various aspects such as salary distribution, gender-based salary comparison, relationship between previous work experience and current salary, and the distribution of educational backgrounds among employees.

Data Visualization

1. Histogram: Distribution of Salaries Among Employees

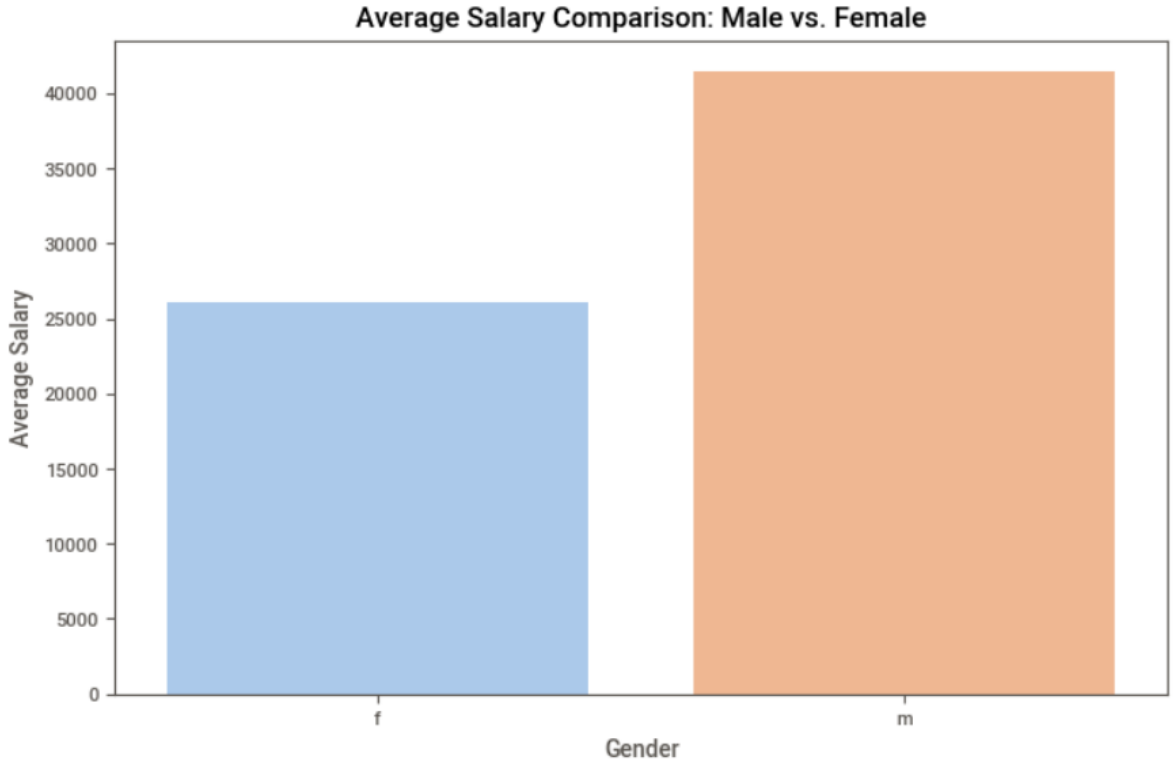
The histogram provides a visual representation of the distribution of salaries among employees. The x-axis represents salary ranges, while the y-axis represents the frequency of employees within each range.

```python
import pandas as pd
import matplotlib.pyplot as plt

# Load the CSV file into a DataFrame
employee_data = pd.read_csv('/content/Employee data.csv')

# Assuming the column containing salaries is named 'salary'
salary_column = 'salary'

# Plotting the histogram
plt.figure(figsize=(10, 6))
plt.hist(employee_data[salary_column], bins=20, color='skyblue', edgecolor='black')
plt.title('Distribution of Salaries Among Employees')
plt.xlabel('Salary')
plt.ylabel('Frequency')
plt.grid(axis='y', linestyle='--', alpha=0.7)

# Show the plot
plt.show()
```



2. Bar Plot: Average Salary Comparison - Male vs. Female

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

employee_data = pd.read_csv('/content/Employee data.csv')

# Assuming the columns containing gender and salary are named 'gender' and 'salary'
gender_column = 'gender'
salary_column = 'salary'

# Group the data by gender and calculate the average salary
average_salary_by_gender = employee_data.groupby(gender_column)[salary_column].mean().reset_index()

# Plotting the bar plot
plt.figure(figsize=(8, 5))
sns.barplot(x=gender_column, y=salary_column, data=average_salary_by_gender, palette='pastel')
plt.title('Average Salary Comparison: Male vs. Female')
plt.xlabel('Gender')
plt.ylabel('Average Salary')
plt.show()
```



This bar plot compares the average salary between male and female employees. The x-axis represents gender, and the y-axis represents the average salary. It indicates any disparities in average salary based on gender.

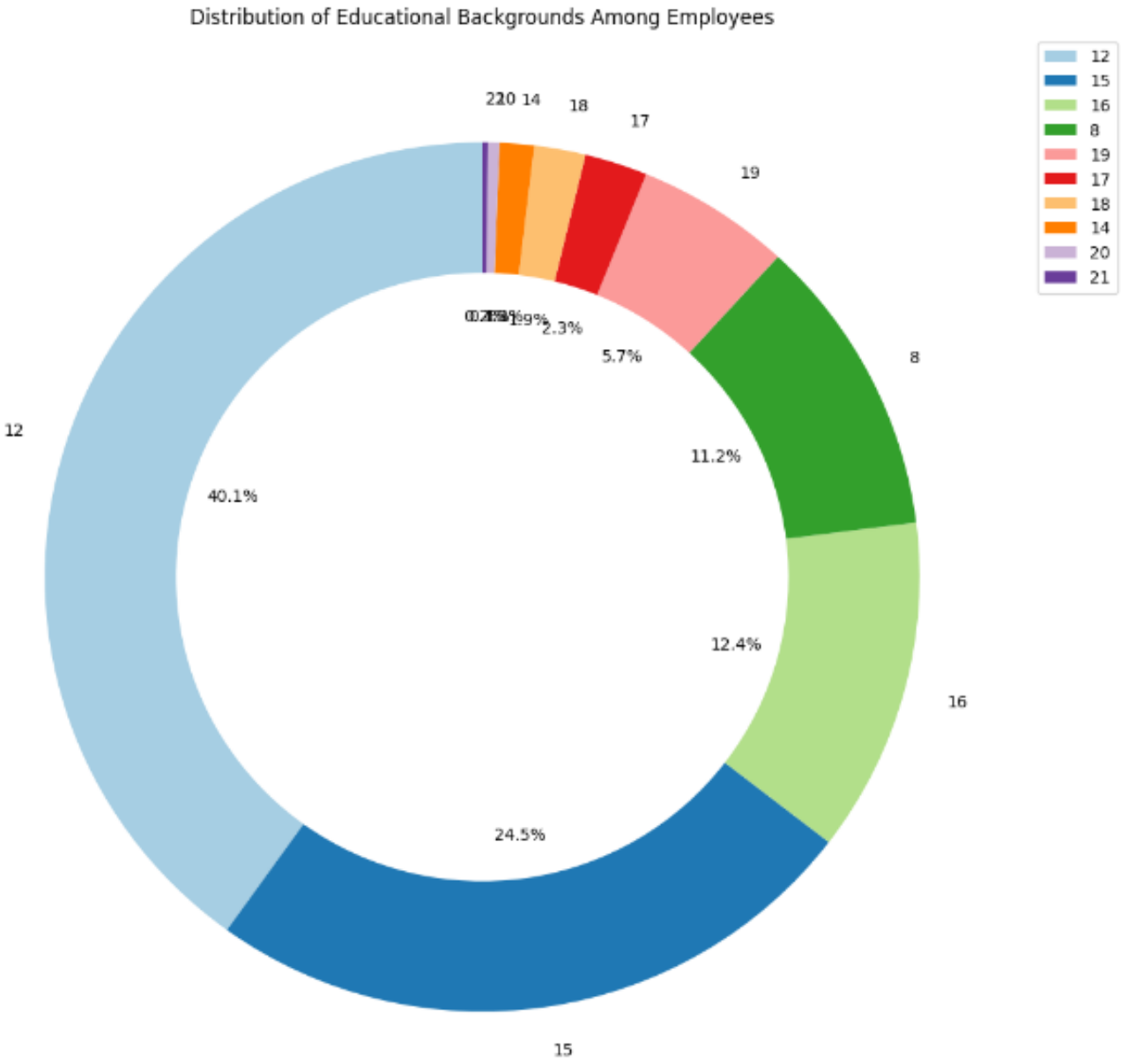
3. Scatter Plot: Relationship between Previous Work Experience and Salary

```python
import pandas as pd
import matplotlib.pyplot as plt

# Load the CSV file into a DataFrame
employee_data = pd.read_csv('/content/Employee data.csv')

# Assuming the columns containing previous work experience and salary are named 'prevexp' and 'salary'
prevexp_column = 'prevexp'
salary_column = 'salary'

# Plotting the scatter plot
plt.figure(figsize=(10, 6))
plt.scatter(employee_data[prevexp_column], employee_data[salary_column], color='coral', alpha=0.7)
plt.title('Relationship between Previous Work Experience and Salary')
plt.xlabel('Previous Work Experience (months)')
plt.ylabel('Salary')
plt.grid(True)
plt.show()
```

The scatter plot illustrates the relationship between an employee's previous work experience (in months) and their current salary. Each point represents an employee, and the x-axis corresponds to previous work experience, while the y-axis represents the current salary.

4. Pie Chart: Distribution of Educational Backgrounds

```
import pandas as pd
import matplotlib.pyplot as plt

# Load the CSV file into a DataFrame
employee_data = pd.read_csv('/content/Employee data.csv')

# Assuming the column containing educational backgrounds is named 'educ'
educ_column = 'educ'

# Count the occurrences of each educational background
educ_counts = employee_data[educ_column].value_counts()

# Plotting the pie chart with improved formatting
plt.figure(figsize=(12, 12))
plt.pie(educ_counts, labels=educ_counts.index, autopct='%1.1f%%', startangle=90, colors=plt.cm.Paired.colors, wedgeprops=dict(width=0.3))

# Move the legend outside the pie chart to avoid overlapping
plt.legend(bbox_to_anchor=(1, 1), loc='upper left')

plt.title('Distribution of Educational Backgrounds Among Employees')
plt.show()
```



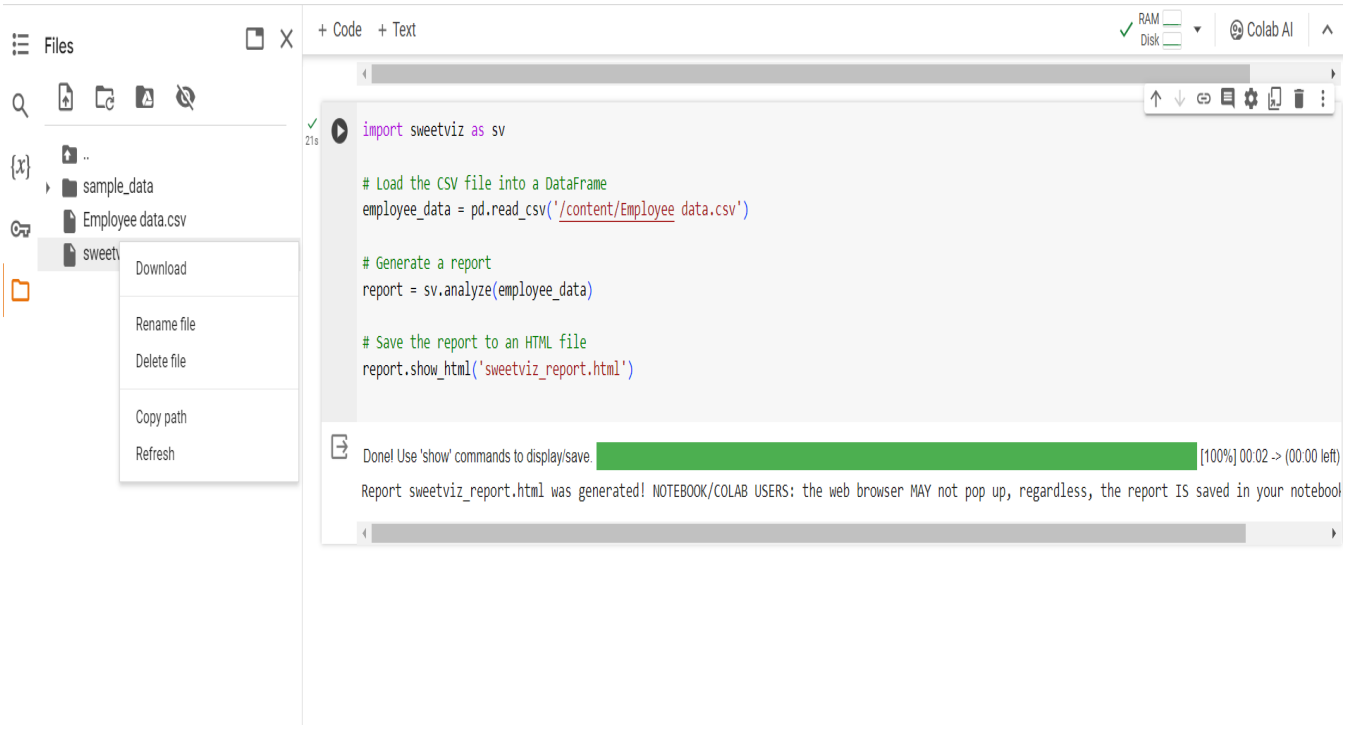
The pie chart visualizes the distribution of educational backgrounds among employees. Each slice represents a different level of education, and the percentages indicate the proportion of employees with each educational background.

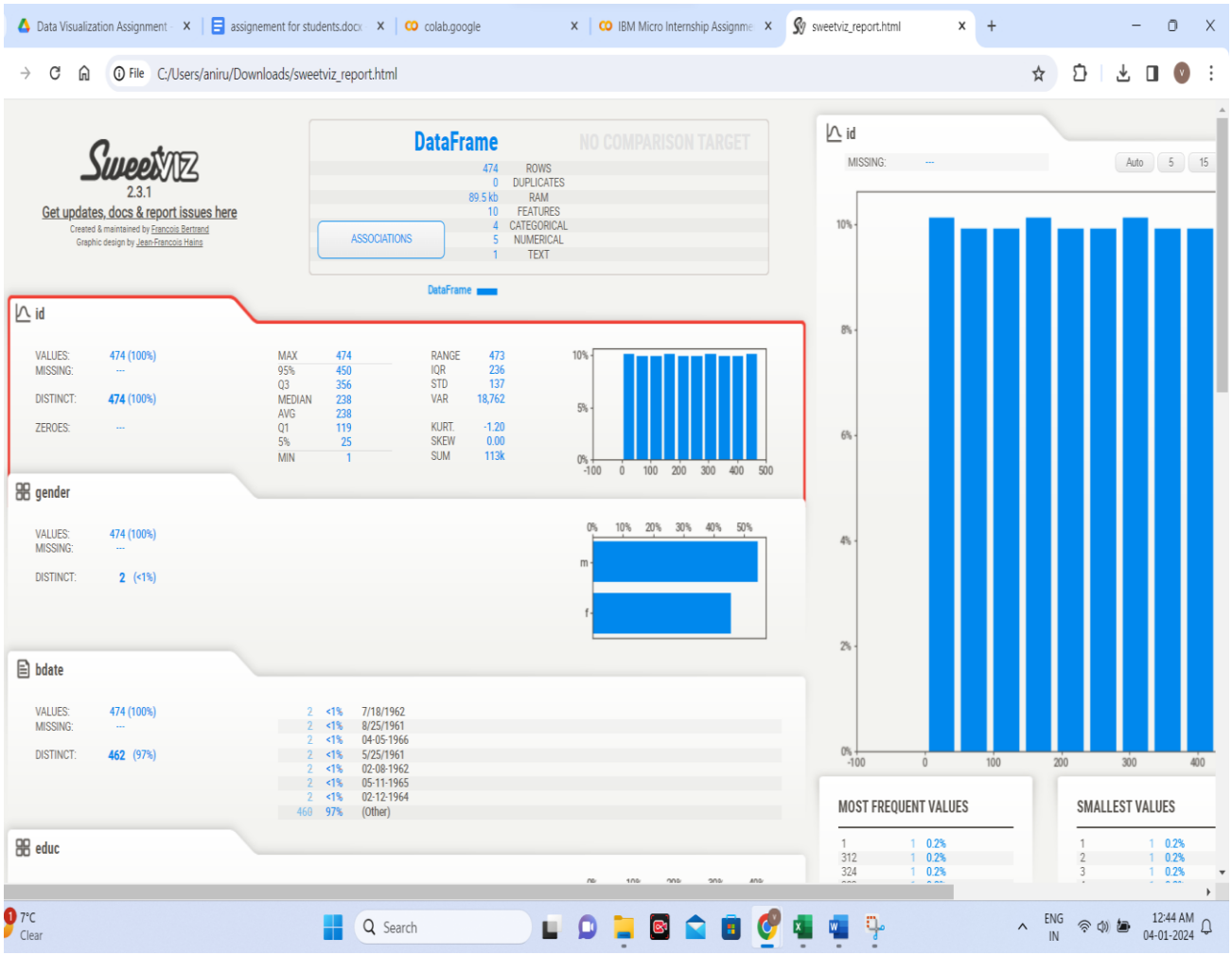
5. Summary Using Sweetviz

To provide a holistic understanding of the dataset, we employed the Sweetviz library to generate a detailed summary report. The Sweetviz summary report encompasses insights into data types, missing values, statistical metrics, and visual comparisons between different features.

```
pip install sweetviz
```

```
Collecting sweetviz
  Downloading sweetviz-2.3.1-py3-none-any.whl (15.1 MB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 15.1/15.1 MB 31.1 MB/s eta 0:00:00
Requirement already satisfied: pandas!=1.0.0,!=1.0.1,!=1.0.2,>=0.25.3 in /usr/local/lib/python3.10/dist-packages (from sweetviz) (1.5.3)
Requirement already satisfied: numpy>=1.16.0 in /usr/local/lib/python3.10/dist-packages (from sweetviz) (1.23.5)
Requirement already satisfied: matplotlib>=3.1.3 in /usr/local/lib/python3.10/dist-packages (from sweetviz) (3.7.1)
Requirement already satisfied: tqdm>=4.43.0 in /usr/local/lib/python3.10/dist-packages (from sweetviz) (4.66.1)
Requirement already satisfied: scipy>=1.3.2 in /usr/local/lib/python3.10/dist-packages (from sweetviz) (1.11.4)
Requirement already satisfied: jinja2>=2.11.1 in /usr/local/lib/python3.10/dist-packages (from sweetviz) (3.1.2)
Requirement already satisfied: importlib-resources>=1.2.0 in /usr/local/lib/python3.10/dist-packages (from sweetviz) (6.1.1)
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.10/dist-packages (from jinja2>=2.11.1->sweetviz) (2.1.3)
Requirement already satisfied: contourpy>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib>=3.1.3->sweetviz) (1.2.0)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.10/dist-packages (from matplotlib>=3.1.3->sweetviz) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib>=3.1.3->sweetviz) (4.47.0)
Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib>=3.1.3->sweetviz) (1.4.5)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib>=3.1.3->sweetviz) (23.2)
Requirement already satisfied: pillow>=6.2.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib>=3.1.3->sweetviz) (9.4.0)
Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib>=3.1.3->sweetviz) (3.1.1)
Requirement already satisfied: python-dateutil>=2.7 in /usr/local/lib/python3.10/dist-packages (from matplotlib>=3.1.3->sweetviz) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-packages (from pandas!=1.0.0,!=1.0.1,!=1.0.2,>=0.25.3->sweetviz) (
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.7->matplotlib>=3.1.3->sweetviz) (1.1
Installing collected packages: sweetviz
Successfully installed sweetviz-2.3.1
```





Python Scripting on Google Colab

For the analysis, we utilized Python language on google colab along with popular libraries such as pandas for data manipulation, matplotlib for basic plotting, seaborn for enhanced visualization, and sweetviz for generating a comprehensive data analysis report.

Conclusion

In conclusion, the analysis of the employee dataset has provided valuable insights into various aspects such as salary distribution, gender-based salary comparisons, the relationship between previous work experience and current salary, and the distribution of educational backgrounds. The visualizations and Sweetviz summary serve as powerful tools for understanding and communicating key patterns within the data.

The Python scripting, using libraries such as pandas, matplotlib, seaborn, and sweetviz, facilitated efficient data manipulation, visualization, and comprehensive analysis.

This report aims to assist in informed decision-making processes related to employee management and resource allocation within the organization.