Project Description

- Conduct an EDA on a dataset provided by a company.
- The dataset has information about people who registered for a post in a department.
- Use statistical knowledge to make sense of the data and draw conclusions.
- Understand the data columns, check for missing data, club similar columns.
- Check for outliers, draw data summary, use data visualization.
- Draw conclusions based on data analysis.
- Goal: Gain insights and inform decision-making.

Approach

- Data Understanding: Examine and understand that the data provided is of the hiring process

- Data Cleaning: observed one blank cell, some no value, and outlier data same has highlighted and removed

- Data Exploration: Used Pivot table and visualization techniques to explore the data.

- Data Analysis: Identify relationships, compare groups, and draw conclusions.

- Conclusions: Present findings and identify actionable insights.

- Additional Steps: Depending on project requirements, additional steps such as modeling or predictive analytics may be required.

- the approach involves understanding, cleaning, exploring, analyzing, and presenting the data to gain insights that can inform decision-making.

Tech-Stack Used

EXCEL

Microsoft 365

- Microsoft 365 Excel is a very powerful tool for DATA Analytics
- It offers new functions and formulas that make work hassle-free
- It also has good quality charts to prepare visualization
- Limited automation options compared to other tools that offer more advanced scripting and programming capabilities.
- Large file sizes can make it difficult to work with large datasets on older or less powerful devices.

Insights

A.) Hiring: Process of intaking of people into an organization for different kinds of positions.
Task: How many males and females are Hired ?

Hiring Status	Hired ▼
Count of Employees	
Gender ▼	**Total**
Don't want to say	268
Female	1854
Male	2561
Grand Total	**4683**

- After cleaning the data and converting it into a table, we inserted a Pivot Table. This approach ensures that any additions to the original table are automatically reflected in the Pivot Table after refreshing the data.
- Since we needed to calculate the count of hired employees, we placed the "Status" column in the filter area, and the "Event name Renamed as Gender" column in both the column and value areas of the Pivot Table. By doing this, the Pivot Table displayed the count of employees hired based on gender.
- Overall, using a Pivot Table in this manner helps in creating an interactive summary report that can be easily updated, filtered, and analyzed to provide valuable insights.

Insights

B.) Average Salary: Adding all the salaries for a select group of employees and then dividing the sum by the number of employees in the group.
Task: What is the average salary offered in this company ?

Department	(All)	
Average of Offered Salary		
Post Name		**Total**
b9		49281.957
c-10		51244.359
c5		50190.246
c8		50821.310
c9		50202.468
i1		49937.955
i4		44841.954
i5		49512.149
i6		48839.249
i7		49777.694
m6		34521.333
m7		41402.000
n10		26990.000
n6		44700.000
n9		46219.000
Grand Total		**49881.143**

- The Pivot Chart is a powerful tool that can be used to derive statistics from a large dataset. To find the average salary of employees in a selected group, we can use the Pivot Table functionality.

- First, we need to add the required data into the Pivot Table by selecting the data range and converting it into a table. Then, we need to select the "Department" column as the filter field to toggle various fields and derive the average salary of employees.

- We can group the employees by their "Post Name" as it impacts their salary. We add the "Salary" column into the "Values" area and use the "Average" option to summarize the values. This will calculate the average salary of employees in the selected group.

- By using Pivot Table and Pivot Chart, we can quickly and easily analyze the data and identify the average salary of employees in different departments or groups. This helps to make informed decisions and drive business growth.

Insights

C.) Class Intervals: The class interval is the difference between the upper class limit and the lower class limit.

Task: Draw the class intervals for salary in the company ?

Offered Salary
56553
22075
70069
3207
29668
69904
11758
15156
49515
26990
86787
2308
56688
15134
73579

Class Intervals for Salary

Salary Class Interval

Class intervals shown: (__, 26300], (26300, 31400], (31400, 36500], (36500, 41600], (41600, 46700], (46700, 51800], (51800, 56900], (56900, 62000], (62000, 67100], (67100, 72200], (72200, 77300], (77__

- In statistics, a class interval is a range of values that is used to group continuous data into discrete categories or classes. This allows for easier analysis and visualization of the data.
- Histograms are a useful tool in Microsoft Excel for creating class intervals.
- Next, you can customize the histogram by adjusting the bin size and number of bins, adding axis labels and titles, and changing the chart colors and styles. You can also use the "Chart Filters" option to include or exclude specific data points or categorie

Insights

D.) Charts and Plots: This is one of the most important part of analysis to visualize the data. Task Draw Pie Chart / Bar Graph (or any other graph) to show proportion of people working different department ?
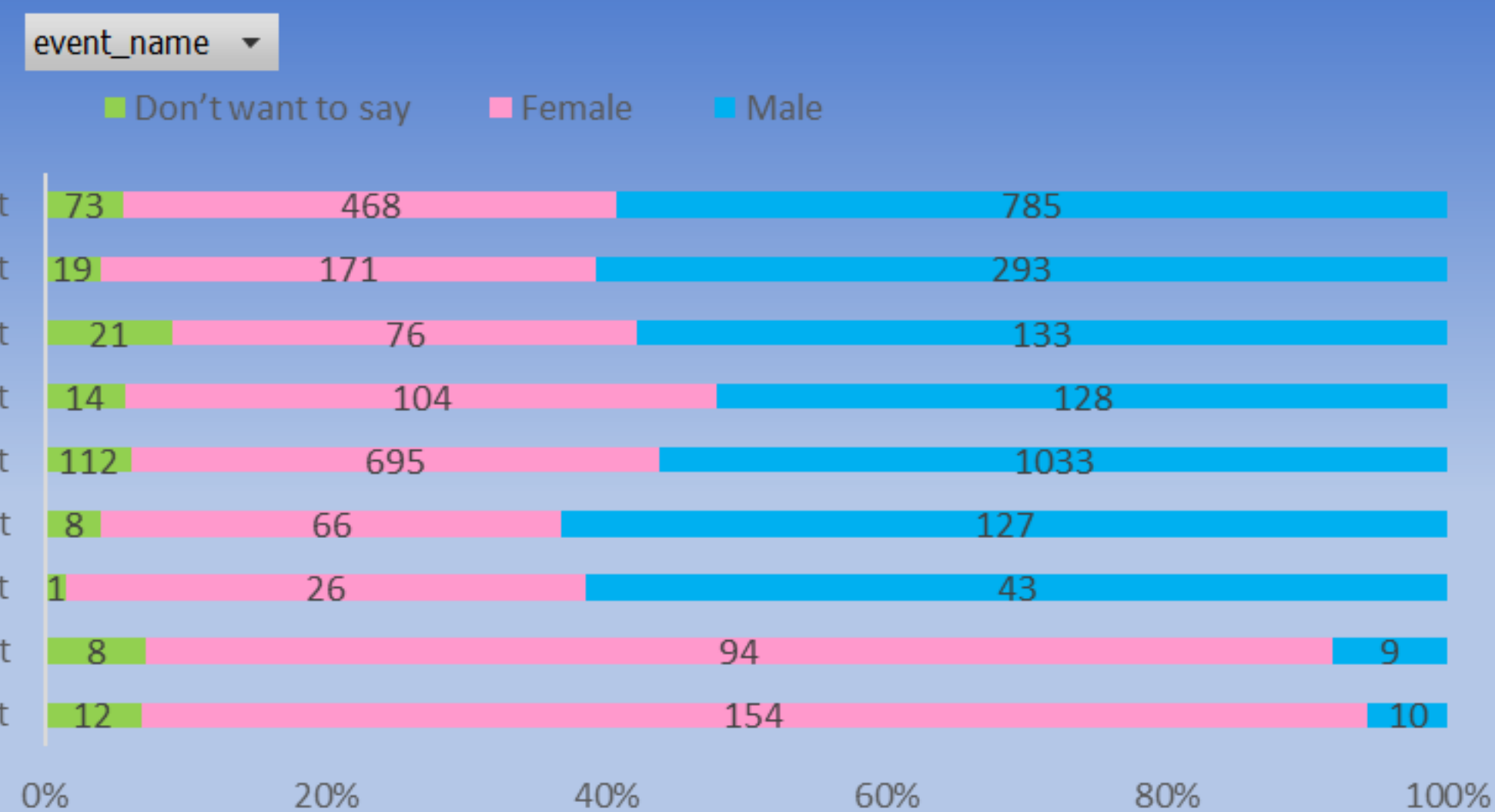

- Charts can be created using a Pivot table
- To show the proportion of people working in different departments, we used a stacked bar chart. This chart allows you to visualize the data for each department in terms of percentages.
- While using this chart, it is important to note that even if the value for a certain department or gender is small, it may still appear larger because it is only being compared to the data within its respective department. However, the stacked bar chart allows you to see the proportion of each group within the total dataset.

Insights
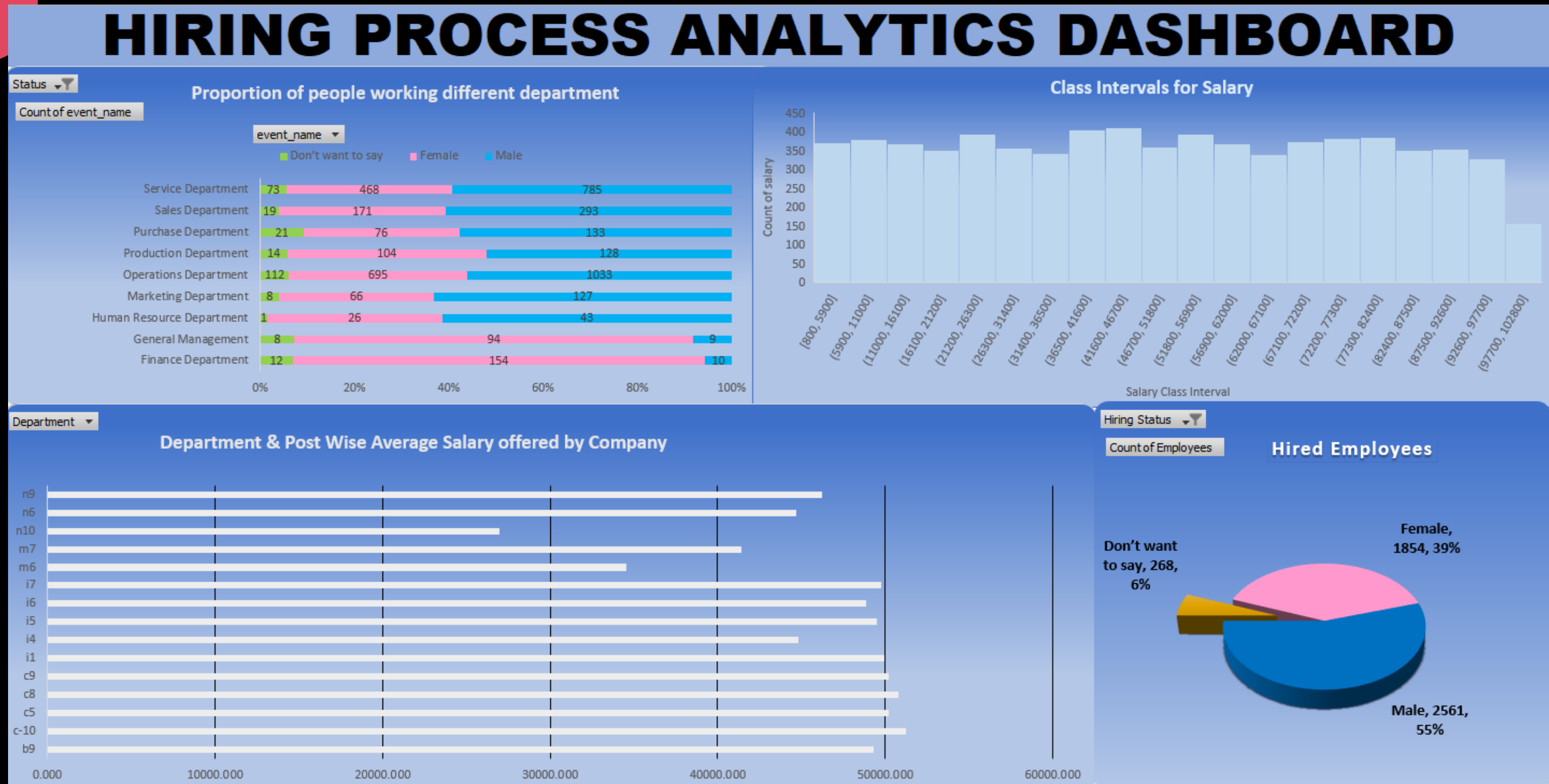
Result

- This project has helped me to analyze data and understand the importance of data types, rows, and columns in data analysis.
- During the data analysis process, I identified and dealt with blank cells and non-value data such as (-) that could skew the results in unexpected ways. I also learned about outliers and their impact on data visualization.
- Using this project, I was able to implement various features of Pivot tables and create multiple charts to visualize the data. I then grouped all the charts and created a dashboard.
- I understood what is EDA and how to perform it.
- Overall, colors played an important role in data visualization, making it easier to interpret and understand the data.

Attachment:- Link for project folder

THANK YOU