Excel Sheet Hyperlink:

https://docs.google.com/spreadsheets/d/18OzWYra9nHRP3arQSAstHsilYk0IjqTC/edit?usp=sharing&ouid=117438581175883142153&rtpof=true&sd=true

Some of the features of Microsoft excel used for this project are not available on Google sheets, when previewing. Hence, please open the file on Microsoft Excel.

Project Description:

Hiring Process Analytics is crucial for several reasons: it helps identify bottlenecks and inefficiencies, leading to faster recruitment, helps optimize budgets by understanding resource allocation, improves the quality of hires by revealing effective sources and methods, highlights gaps in diversity and inclusion for a balanced workforce, provides predictive insights for proactive talent management and enhances the candidate experience. Overall, it offers actionable insights for data-driven decisions, resulting in a more effective and efficient hiring process.

Approach:

The dataset consists of 7 columns and 7168 rows, with each column representing a unique attribute. The initial approach involves cleaning and pre-processing the data to ensure it is well-structured and usable. This process includes handling missing data points, correcting errors, and removing or replacing duplicate values. The modifications made during data cleaning and pre-processing is mentioned below. After data cleaning and preprocessing in Microsoft Excel, the next steps to derive insights typically include detailed data analysis using pivot tables and functions, creating visualizations like charts and graphs to identify trends and outliers, segmenting the data into meaningful groups, analyzing trends over time, examining correlations to understand relationships between variables, and compiling the findings into comprehensive reports or dashboards. These steps help transform raw data into valuable insights for strategic decision-making and improving the hiring process.

Handling Missing Data:

- The ‘event_name’ column had 15 rows with “-” values, which were treated as null values. These were replaced with “Don’t want to say” to indicate unknown gender.
- The ‘Offered Salary’ column had one null value (empty space). For the corresponding “Sales Department” and “i7” Post Name, the median salary of 45,400 was used as a replacement.
- The ‘Post Name’ column had one row with a “-” value, treated as null value. For the corresponding “Sales Department” and an Offered Salary of 85,914, the most common Post Name for similar salaries, “c9”, was used.
- The ‘event_name’ column, which actually represented gender, was renamed to ‘Gender’ for clarity.

Error Handling:

- The ‘Post Name’ column contained a category “c-10,” which appeared to be a typo. This was corrected to “c10.”

Handling Duplicate Values:

- Since application_id is a unique value for each candidate, there should be no duplicates. But, the ‘application_id’ column had 54 rows with duplicate values. These duplicates should either be removed or replaced with the correct values. For the purposes of this project, these

duplicate application_id values have been highlighted in red since the values in the corresponding columns vary and are not quite repetitive.

Tech-Stack Used:

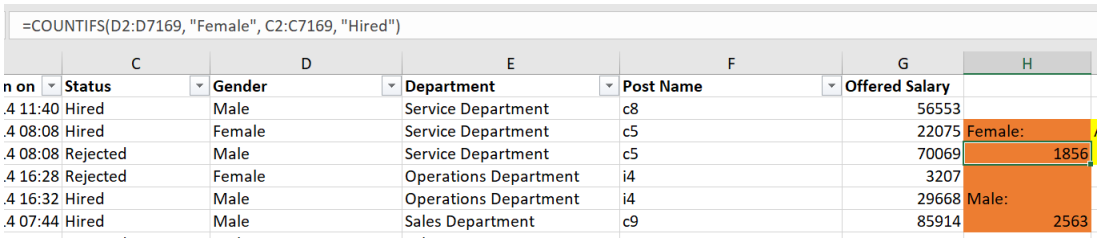
- Microsoft Excel 2019

Several features of Microsoft Excel were particularly useful for this project: data filtering for sorting and identifying patterns, pivot tables for summarizing large datasets, graphical tools for visualizing trends, formulas and functions for complex calculations, data cleaning tools for preprocessing accuracy, and conditional formatting for highlighting key insights. These collectively made Excel a powerful tool for analyzing and visualizing the hiring process data.

Insights:

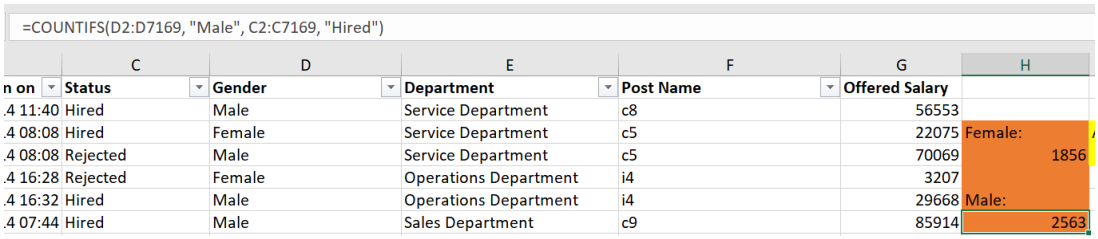
All the insights observed and outputs obtained are highlighted in light blue colour.

A. Hiring Analysis: Determine the gender distribution of hires. How many males and females have been hired by the company?



=COUNTIFS(D2:D7169, "Female", C2:C7169, "Hired")

	C	D	E	F	G	H
n on	Status	Gender	Department	Post Name	Offered Salary	
.4 11:40	Hired	Male	Service Department	c8	56553	
.4 08:08	Hired	Female	Service Department	c5	22075	Female:
.4 08:08	Rejected	Male	Service Department	c5	70069	1856
.4 16:28	Rejected	Female	Operations Department	i4	3207	
.4 16:32	Hired	Male	Operations Department	i4	29668	Male:
.4 07:44	Hired	Male	Sales Department	c9	85914	2563

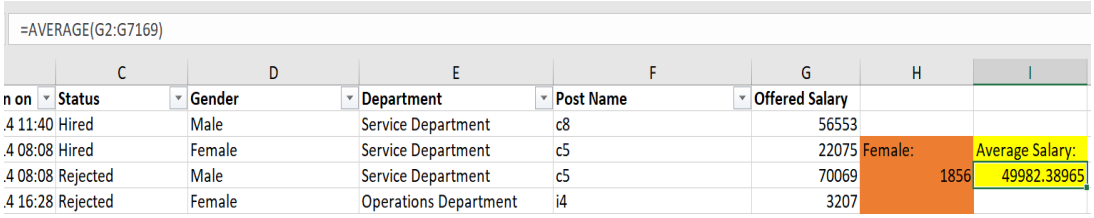


=COUNTIFS(D2:D7169, "Male", C2:C7169, "Hired")

	C	D	E	F	G	H
n on	Status	Gender	Department	Post Name	Offered Salary	
.4 11:40	Hired	Male	Service Department	c8	56553	
.4 08:08	Hired	Female	Service Department	c5	22075	Female:
.4 08:08	Rejected	Male	Service Department	c5	70069	1856
.4 16:28	Rejected	Female	Operations Department	i4	3207	
.4 16:32	Hired	Male	Operations Department	i4	29668	Male:
.4 07:44	Hired	Male	Sales Department	c9	85914	2563

As per the output, number of female hires is 1856 and number of male hires is 2563. More than half of the hired candidates are male, while only 39.51% are female. The remaining candidates have not disclosed their gender. A high gender ratio (male to female) could negatively impact the organization's public image. Therefore, the organization should aim to reduce this ratio, bringing it closer

B. Salary Analysis: What is the average salary offered by this company? Use Excel functions to calculate this.

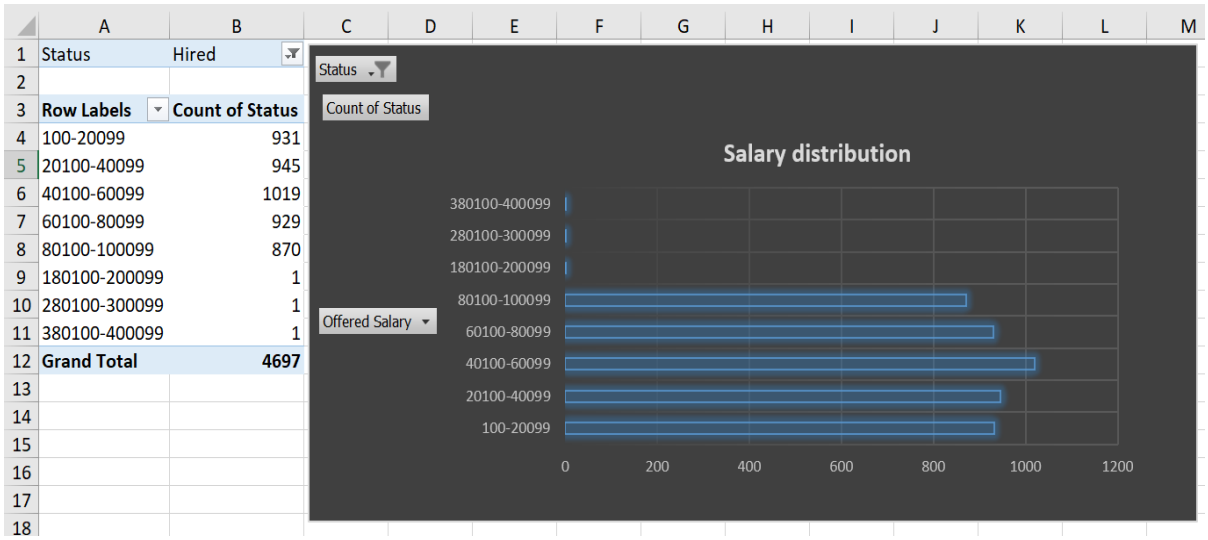


=AVERAGE(G2:G7169)

	C	D	E	F	G	H	I
n on	Status	Gender	Department	Post Name	Offered Salary		
.4 11:40	Hired	Male	Service Department	c8	56553		
.4 08:08	Hired	Female	Service Department	c5	22075	Female:	Average Salary:
.4 08:08	Rejected	Male	Service Department	c5	70069	1856	49982.38965
.4 16:28	Rejected	Female	Operations Department	i4	3207		

The average salary of hired candidates across domains is 49,982.39. This is nearly identical to the offered salary, indicating that the hiring team is recruiting candidates in line with the organization's predetermined salary bands.

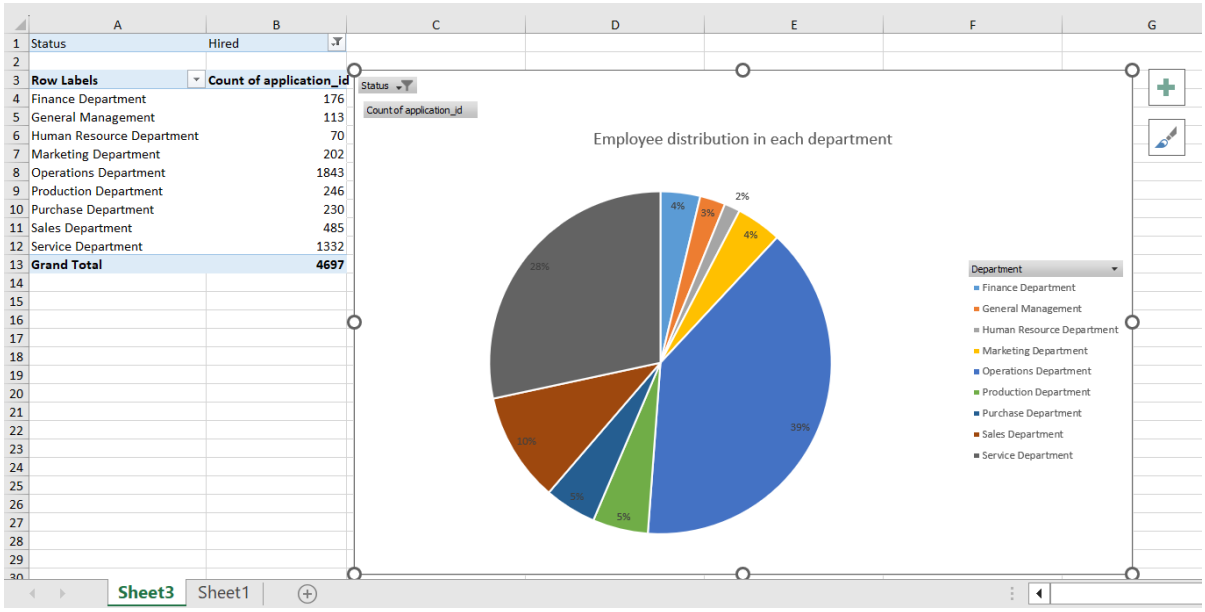
C. Salary Distribution: Create class intervals for the salaries in the company. This will help you understand the salary distribution.



Status	Hired
Row Labels	**Count of Status**
100-20099	931
20100-40099	945
40100-60099	1019
60100-80099	929
80100-100099	870
180100-200099	1
280100-300099	1
380100-400099	1
Grand Total	**4697**

We can see that the highest number of offered salaries fall within the range of 40,100-60,099, with 1,019 candidates in this bracket. Conversely, the lowest offered salary ranges are 180,100-200,099, 280,100-300,099, and 380,100-400,099, each with only one candidate in these intervals

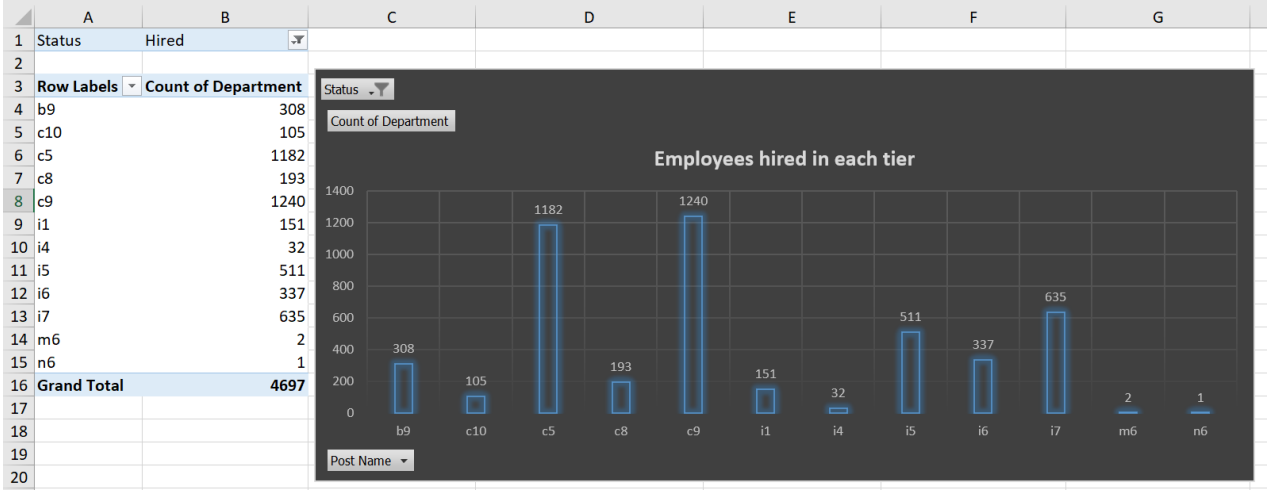
D. Departmental Analysis: Use a pie chart, bar graph, or any other suitable visualization to show the proportion of people working in different departments.



Status	Hired
Row Labels	**Count of application_id**
Finance Department	176
General Management	113
Human Resource Department	70
Marketing Department	202
Operations Department	1843
Production Department	246
Purchase Department	230
Sales Department	485
Service Department	1332
Grand Total	**4697**

There are a total of 4697 employees, out of which, majority of them work in Operations department (39%). Human Resource Department needs lesser manpower compared to the other departments of the organization (2%). The pie chart shows the percentage distribution of employee distribution across various departments. These figures may reflect the relative size and importance of these departments within the organization.

E. Position Tier Analysis: Use a chart or graph to represent the different position tiers within the company. This will help you understand the distribution of positions across different tiers.



Status	Hired
Row Labels	**Count of Department**
b9	308
c10	105
c5	1182
c8	193
c9	1240
i1	151
i4	32
i5	511
i6	337
i7	635
m6	2
n6	1
Grand Total	**4697**

The bar chart above illustrates the distribution of position tiers within the organization. It is evident that the majority of candidates have been hired for position tier c9 while the fewest have been hired for position tier n6.

Result

Completing this project yields several valuable outcomes. Handling missing data ensures the dataset is accurate and reliable, while clubbing columns simplifies analysis. Detecting and managing outliers prevents skewed results, and summarizing cleaned data provides a clear overview of key metrics. Understanding the gender distribution of hires highlights imbalances and informs diversity initiatives. Calculating the average salary assesses compensation practices, and creating salary intervals reveals pay distribution. Visualizing departmental proportions shows the organizational structure, and representing position tiers helps understand the hierarchy. Overall, this project enabled meaningful conclusions about the hiring process, potentially leading to improved strategies and better decision-making.