TRAINITY PROJECT 4: HIRING PROCESS ANALYTICS

PROJECT DESCRIPTION:

The objective of this project is to analyze the company's hiring process data and draw meaningful insights from it. The hiring process is a crucial function of any company, and understanding trends such as the number of rejections, interviews, job types, and vacancies can provide valuable insights for the hiring department. As a data analyst, we are given a dataset containing records of previous hires. Our job is to analyze this data and answer certain questions that can help the company improve its hiring process.

APPROACH:

The database file was provided in the attachment. The excel file was carefully observed for missing values, duplicate values and outliers for further analysis. With the help of Power Query editor these values were either altered or removed based on the context of analysis.

TECH- STACK USED:

The software I am using to do the analysis is Microsoft Excel It is a very powerful tool to filter data and plot various graphs to get insights about hiring process of the organization. The Power Query Editor is very helpfu to deect any anomalies in the data as well as correct them.

DATASET OVERVIEW:

The dataset contains 7,168 records of candidates who were interviewed previously and the columns includes the following:

application_id: ID of the applicant
Interview Taken on: Date and time of the interview
Status: Hired/ rejected
event_name: Gender of the applicant
Department: Name of the department for which interview was conducted
Post Name: Name of the post offered
Offered Salary: Salary offered for the job

DATA CLEANING:

Handling Missing Values:

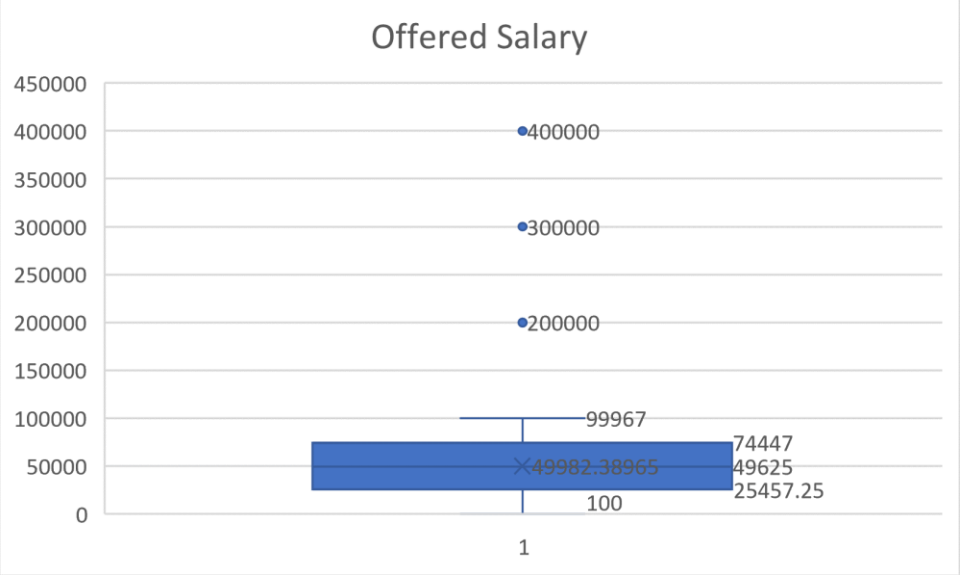
- Column **event_name** has **15** rows with '-' as its values. These can be termed as Null values. We replaced it with "**Don't want to say**" as they both implies the same thing in context of this project.
- Column Offered Salary has 1 row with null value. The corresponding value in Department column is "Sales Department" and Post Name is "**i7**". So, we replaced it

with median of Offered Salary for Sales Department and i7 Post Name. The median came out to **45400**.

- Column Post Name has 1 row with "**-**" as its value. It can be termed as Null value. The corresponding value in Department column is **"**Sales Department**"** and Offered Salary is "**85914**". So we replaced it with majority count of Posts for candidates in Sales Department and whose Offered Salary is between 80,000 and 90,000, which is "**c5**".

Handling Outliers:

I created a **Box plot** for column Offered Salary and found that three values can be termed as outliers. We replaced them with median value os Offered Salary for corresponding Department and Post Name.



Handling Duplicate Values:

Column **application_id** has **54** rows with **duplicate** values. As application_id is a unique id and can't be repeated, we removed them from the dataset for the correct analysis.

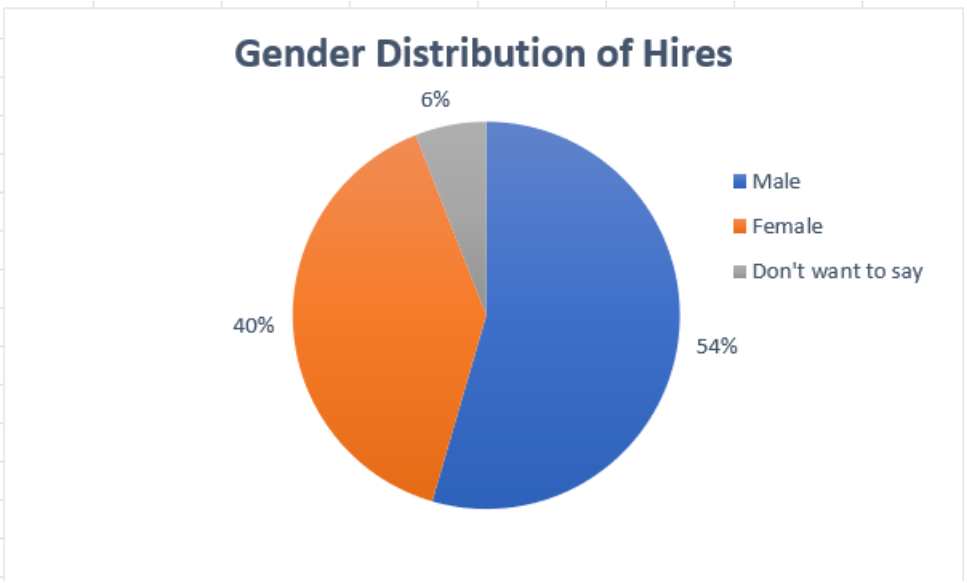
INSIGHTS:

A. Hiring Analysis: The hiring process involves bringing new individuals into the organization for various roles.
Task: Determine the gender distribution of hires. How many males and females have been hired by the company?

Hired		Count
Male		2551
Female		1851
Don't want to say		278
Total		4680

I used COUNT function of Excel to calculate the number of males and females hired in the company.

```
=COUNTIFS(D:D,"Male",C:C,"Hired")
```



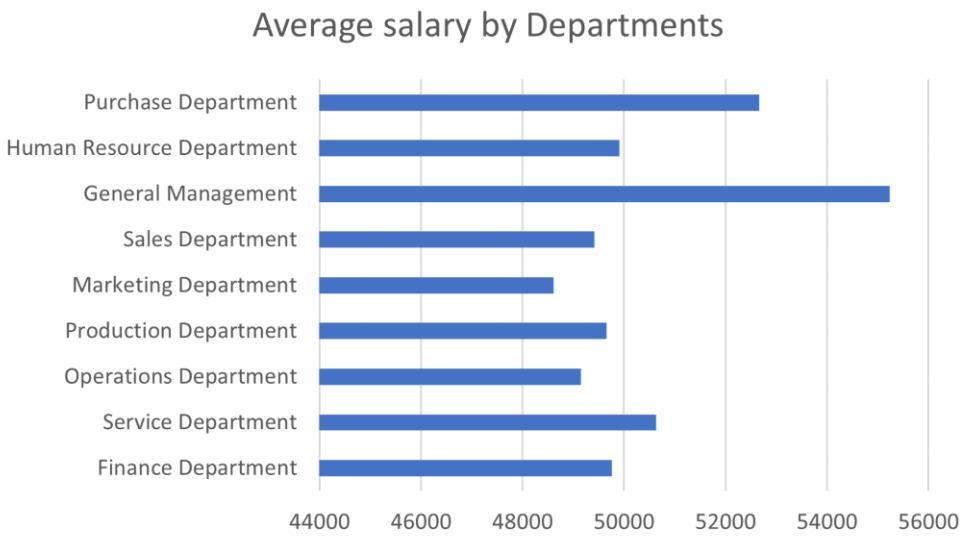
Only 40% females were hired as compared to 54% males which is an indication of skewed Gender ratio which the organisation needs to improve.

B. **Salary Analysis:** The average salary is calculated by adding up the salaries of a group of employees and then dividing the total by the number of employees.
Task: What is the average salary offered by this company? Use Excel functions to calculate this.

Department	Average salary
Finance Department	49759.2962
Service Department	50630.8185
Operations Department	49147.9515
Production Department	49653.443
Marketing Department	48607.8951
Sales Department	49415.3567
General Management	55233.5162
Human Resource Department	49903.7789
Purchase Department	52669.6386
Average offered salary	49948.1703

We use AVERAGEIF function to calculate the average salary.
=AVERAGE(Department column range, Name of the department, Salary column range)

The average salary of all the hired employees is $49948.17. The bar graph below shows the distribution of average salaries of different departments. Genera Management and Purchase Management Department have the highest average salary as compared to other departments.



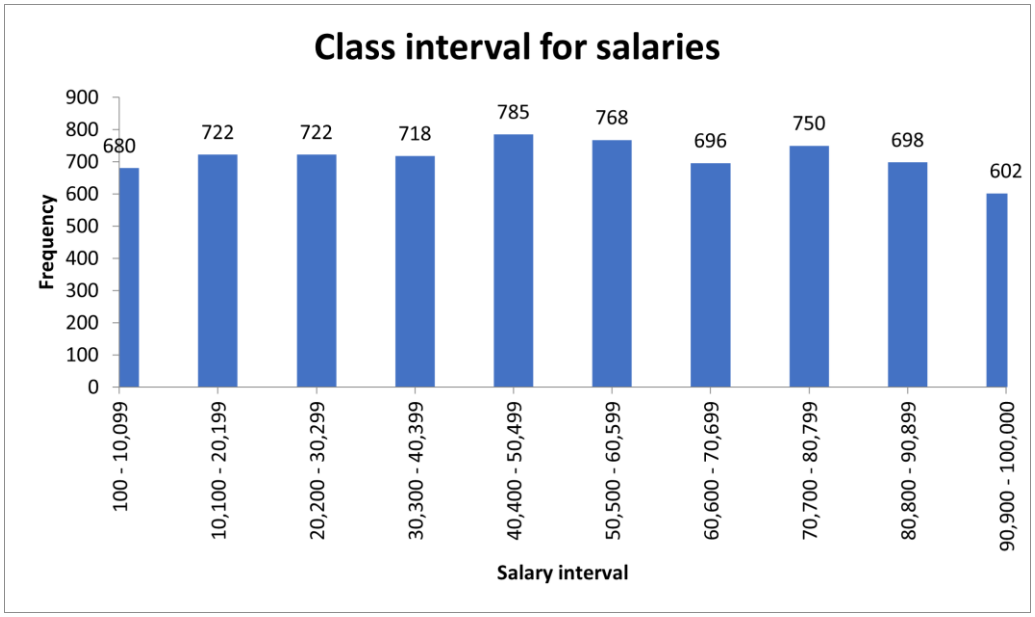
C. **Salary Distribution:** Class intervals represent ranges of values, in this case, salary ranges. The class interval is the difference between the upper and lower limits of a class. **Task:** Create class intervals for the salaries in the company. This will help you understand the salary distribution.

To divide the salary range into various class intervals, I found the maximum (=MAX(salary range)) and minimum (=MIN(salary range)) salary and then divide it into number of class intervals I wanted. Lets say number of class intervals= 10, so

Class width= (Max salary – Min salary)/ Number of intervals

Class interval	*bin*	*Frequency*
100 - 10,099	10099	680
10,100 - 20,199	20199	722
20,200 - 30,299	30299	722
30,300 - 40,399	40399	718
40,400 - 50,499	50499	785
50,500 - 60,599	60599	768
60,600 - 70,699	70699	696
70,700 - 80,799	80799	750
80,800 - 90,899	90899	698
90,900 - 100,000	100000	602

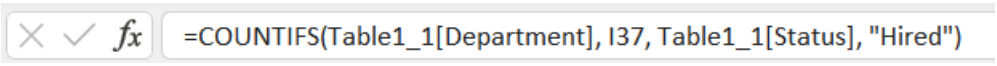
After writing class interval, just write down the upper boundary of every class interval. Histogram is a built in function in Excel which can be used to understand the salary distribution. In this Histogram, for most of the hired candidate's salary falls in the range of $40,400- $60,599.



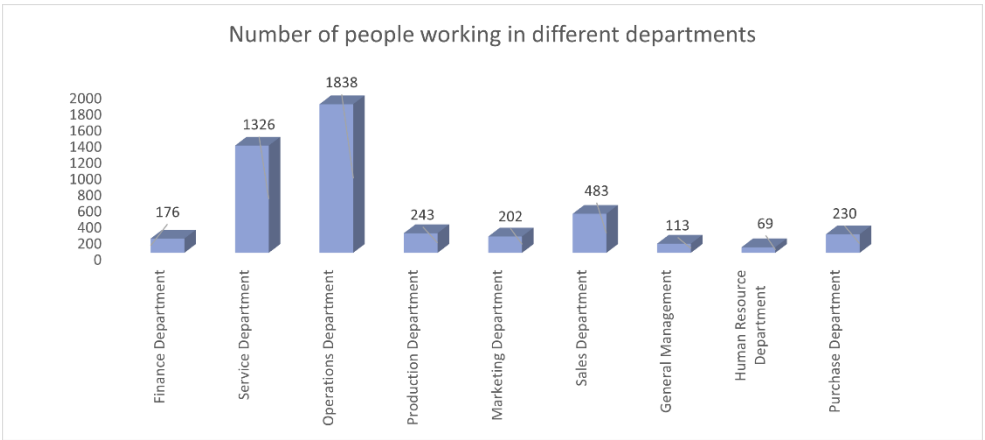
D. Departmental Analysis: Visualizing data through charts and plots is a crucial part of data analysis.

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

Finance Department	176
Service Department	1326
Operations Department	1838
Production Department	243
Marketing Department	202
Sales Department	483
General Management	113
Human Resource Department	69
Purchase Department	230



To count the number of people wrking in various department, we can use COUNTIFS function.
=COUNTIFS(Department column range, Department name, Salary column range,status=”Hired”)

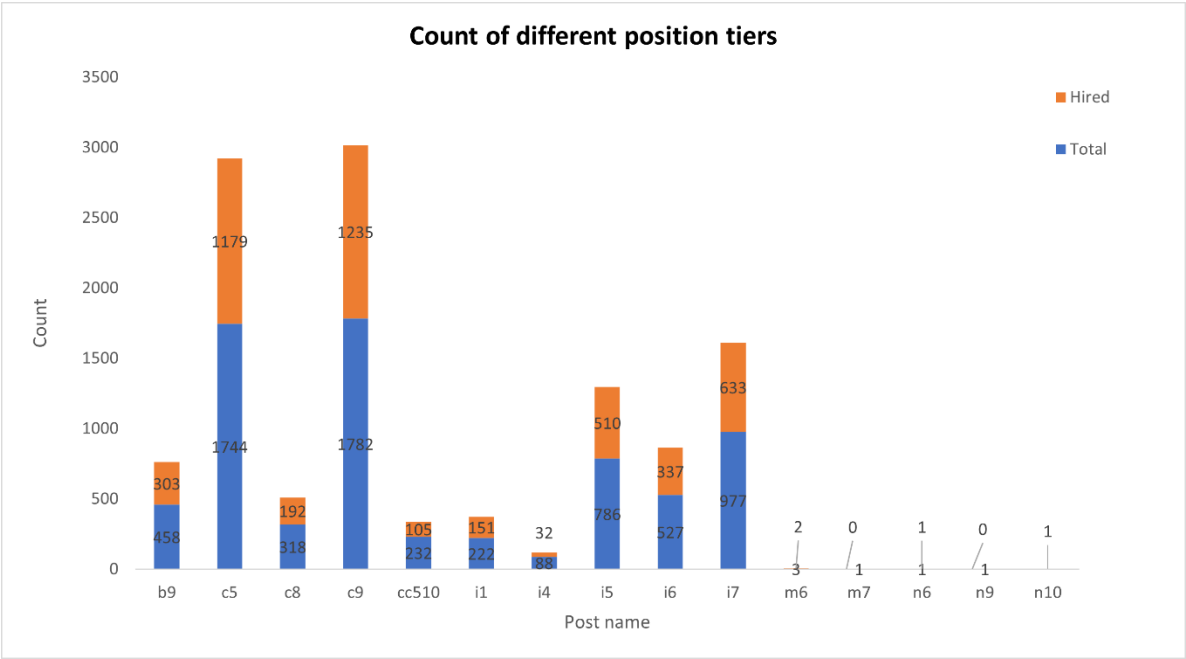


In this bar graph, we can see Operations Department and Service Department has the maimum number of employees and HR department needs to recruit more candidates.

E. **Position Tier Analysis:** Different positions within a company often have different tiers or levels.
Task: 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.

Post name	Total	Hired
b9	458	303
c5	1744	1179
c8	318	192
c9	1782	1235
cc510	232	105
i1	222	151
i4	88	32
i5	786	510
i6	527	337
i7	977	633
m6	3	2
m7	1	0
n6	1	1
n9	1	0
n10	1	0
	7141	4680

This requires us to find the number of candidates that are being hired in different position tiers. For this we can use COUNTIFS function.



The stacked bar graph represents that maximum number of candidates hired in c5 and c9 post whereas m7, n6, n9, n10 post tier have the minimum number of candidates.

RESULT:

The analysis of hiring data revealed key insights into department-wise distribution and salary trends. Recommendations include refining recruitment strategies and adjusting salary structures. Acknowledging limitations, future directions suggest **exploring additional data sources for deeper insights**. Overall, the project underscores the importance of data-driven decision-making in optimizing the hiring process for improved organizational outcomes.