PROJECT 4

HIRING PROCESS ANALYTICS

Dataset: Statistics

application_id	Interview Taken on	Status	event_name	Department	Post Name	Offered Salary
383422	01-05-2014 11:40	Hired	Male	Service Department	c8	56553
907518	06-05-2014 08:08	Hired	Female	Service Department	c5	22075
176719	06-05-2014 08:08	Rejected	Male	Service Department	c5	70069
429799	02-05-2014 16:28	Rejected	Female	Operations Department	i4	3207
253651	02-05-2014 16:32	Hired	Male	Operations Department	i4	29668
289907	01-05-2014 07:44	Hired	Male	Sales Department	-	85914
959124	06-05-2014 16:27	Rejected	Male	Sales Department	i7	69904
86642	09-05-2014 13:17	Rejected	Male	Sales Department	i7	11758
751029	02-05-2014 13:09	Hired	Female	Service Department	i4	15156
434547	02-05-2014 13:11	Rejected	Female	Service Department	i4	49515
518854	01-05-2014 09:00	Rejected	Male	Service Department	n10	26990
649039	07-05-2014 10:48	Hired	Female	Service Department	b9	200000
199526	07-05-2014 10:50	Hired	Male	Service Department	b9	86787
539803	15-05-2014 09:31	Hired	Male	Finance Department	b9	2308
191009	09-05-2014 12:48	Hired	Female	Service Department	i7	56688
195323	09-05-2014 12:48	Hired	-	Service Department	i7	81757
51318	02-05-2014 08:07	Hired	Male	Service Department	i5	15134
742283	02-05-2014 08:11	Rejected	-	Service Department	i5	100
513166	01-05-2014 22:53	Hired	Female	Operations Department	i1	73579
791372	01-05-2014 22:54	Rejected	Male	Operations Department	i1	50351
47857	01-05-2014 22:55	Rejected	Female	Operations Department	i1	38462
834101	01-05-2014 22:53	Rejected	Don't want to say	Operations Department	i1	82510
985008	01-05-2014 09:41	Rejected	Male	Service Department	i6	52554
891568	01-05-2014 16:28	Hired	Female	Operations Department	i7	3423
935899	10-05-2014 14:17	Rejected	Male	Service Department	i1	88744

Dataset link:
https://drive.google.com/file/d/1rmenuYAKlXHm4ugL8qJuE6TShYOhGuzR/view?usp=drivesdk

PROJECT DESCRIPTION:

As a data analyst at a multinational company like Google, the objective is to analyze the company's hiring process data to derive meaningful insights. Understanding trends such as the number of rejections, interviews, job types, and vacancies can provide valuable insights for the hiring department. The dataset contains records of previous hires, and the task is to analyze this data to answer specific questions that can help improve the hiring process.

APPROACH:

The approach involves:

1. Handling missing data by deciding on the best strategy.

2. Clubbing columns with multiple categories to simplify analysis.

3. Detecting and deciding on how to handle outliers.

4. Summarizing findings through statistical measures and visualizations.

TECH STACK USED:

Microsoft PowerPoint 2010

Purpose: Data analysis, statistical calculations, and visualization.

INSIGHTS:

1. Gender Distribution of Hires: Analyzed the dataset to determine the number of males and females hired by the company.

2. Average Salary: Calculated the average salary offered by the company using Excel functions.

3. Salary Distribution: Created class intervals to understand the salary distribution within the company.

4. Departmental Analysis: Visualized the proportion of people working in different departments using suitable charts or graphs.

5. Position Tier Analysis: Represented the distribution of positions across different tiers using charts or graphs.

RESULTS:

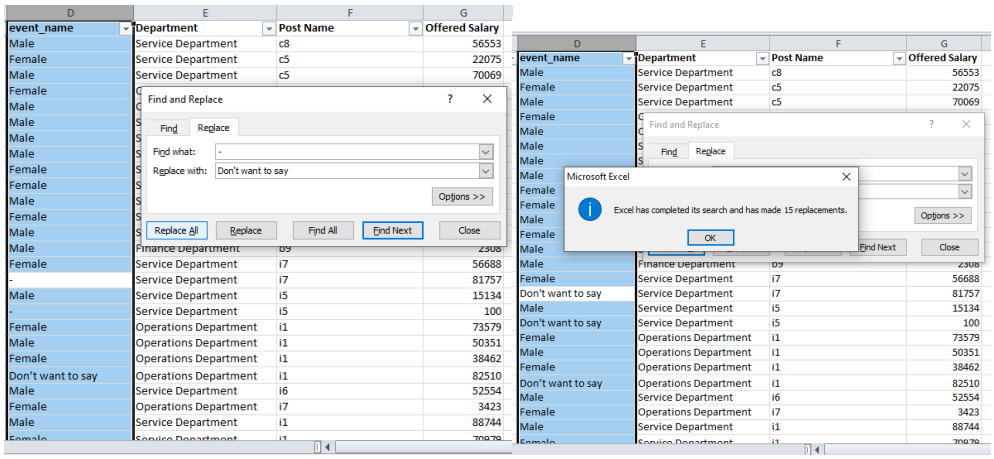
Through the project, insights were gained into the hiring process, gender distribution, salary trends, departmental composition, and position tier distribution. The analysis contributes to a better understanding of the hiring process analytics and provides actionable insights for improving the company's hiring strategies.

Handling Missing Data: Check if there are any missing values in the dataset. If there are, decide on the best strategy to handle them.

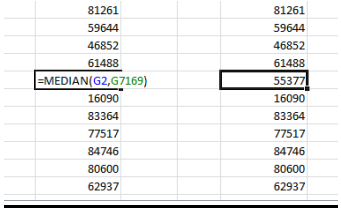
Insights:

Select the column in which you want to replace the value. Ctrl+H, find and replace dialog box will appear enter the values in 'find' and 'replace' column and select replace all.

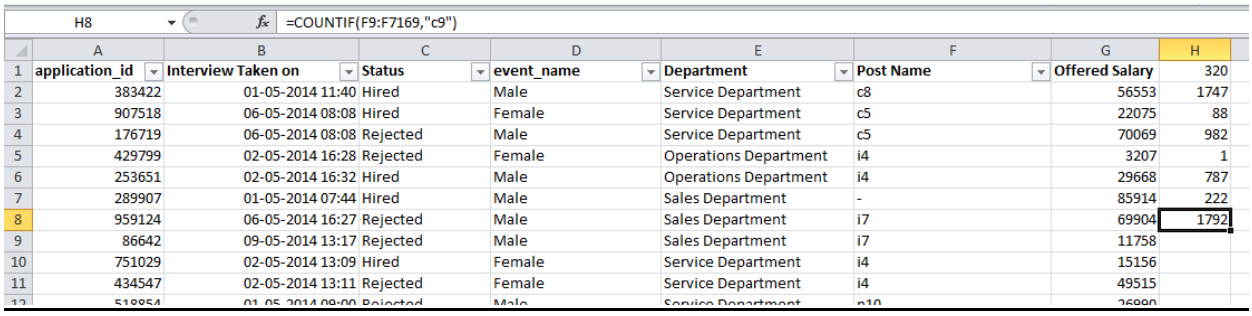
In column **event_name,** there are cells filled with '-' which have been replaced by "Don't want to say".



In column **Offered Salary** has 1 cell which is null. The corresponding value in **Department** column is **Sales Department** and **Post Name** is **i7.** Take median of all the values in **Offered Salary** column and the value came out is **55377**.



Column **Post Name** has a NULL value where **Department** is **Sales Department** and **Offered Salary** is 85914. Replacing it with majority count of Posts for candidates in **Sales Department** and whose **Offered Salary** is between 85,000 and 96,000, which is "**c9**".



	A	B	C	D	E	F	G	H
1	application_id	Interview Taken on	Status	event_name	Department	Post Name	Offered Salary	320
2	383422	01-05-2014 11:40	Hired	Male	Service Department	c8	56553	1747
3	907518	06-05-2014 08:08	Hired	Female	Service Department	c5	22075	88
4	176719	06-05-2014 08:08	Rejected	Male	Service Department	c5	70069	982
5	429799	02-05-2014 16:28	Rejected	Female	Operations Department	i4	3207	1
6	253651	02-05-2014 16:32	Hired	Male	Operations Department	i4	29668	787
7	289907	01-05-2014 07:44	Hired	Male	Sales Department	-	85914	222
8	959124	06-05-2014 16:27	Rejected	Male	Sales Department	i7	69904	1792
9	86642	09-05-2014 13:17	Rejected	Male	Sales Department	i7	11758	
10	751029	02-05-2014 13:09	Hired	Female	Service Department	i4	15156	
11	434547	02-05-2014 13:11	Rejected	Female	Service Department	i4	49515	

Clubbing Columns: If there are columns with multiple categories that can be combined, do so to simplify your analysis.

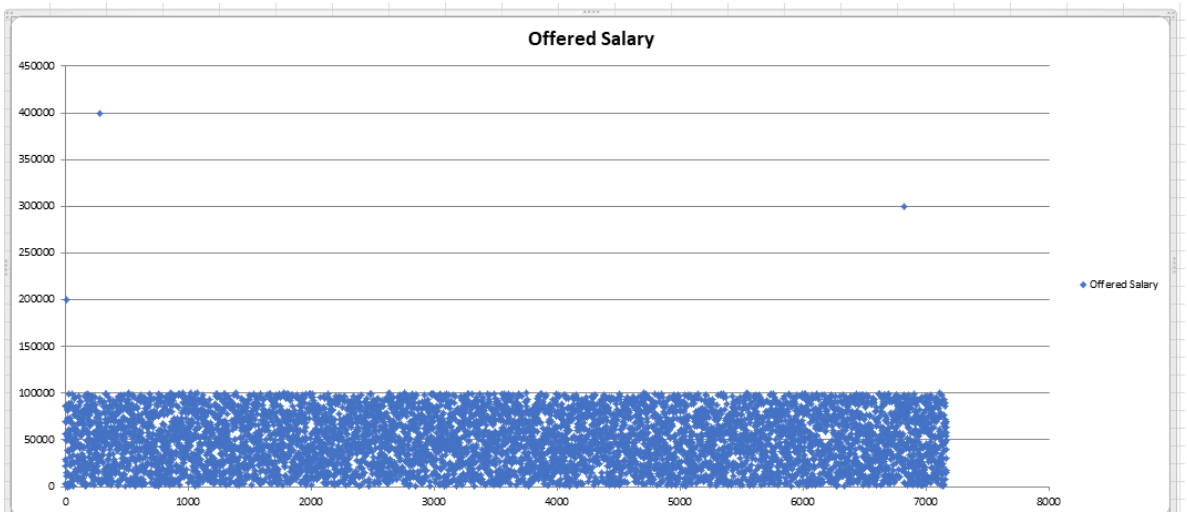
Insights:

Two columns can be clubbed here i.e., **Department** and **Post Name.**

=CONCATENATE(E2,"-",F2)

C	D	E	F	G	H	I
tus	event_name	Department	Post Name	Offered Salary	320	Department and Post
ed	Male	Service Department	c8	56553	1747	Service Department-c8
ed	Female	Service Department	c5	22075	88	Service Department-c5
ected	Male	Service Department	c5	70069	982	Service Department-c5
ected	Female	Operations Department	i4	3207	1	Operations Department-i4
ed	Male	Operations Department	i4	29668	787	Operations Department-i4
ed	Male	Sales Department	c9	85914	222	Sales Department-c9
ected	Male	Sales Department	i7	69904	1792	Sales Department-i7
ected	Male	Sales Department	i7	11758		Sales Department-i7

Outlier Detection: Check for outliers in the dataset that may skew your analysis.



Insights:

From the below **Scatter Plot** of Column **Offered Salary**, there are three outliers and the values are **200000, 300000, 400000**.

Removing Outliers: Decide on the best strategy to handle outliers. This could be removing them, replacing them, or leaving them as is, depending on the situation.

After removing outliers from the **Offered Salary** column

Insights:

Replacing the outliers with median value of **Offered Salary** i.e., 55377.

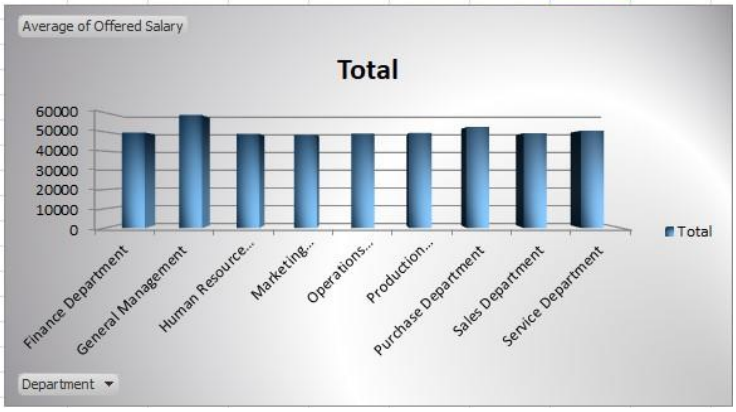
Data Summary: After cleaning and preparing your data, summarize your findings. This could involve calculating averages, medians, or other statistical measures. It could also involve creating visualizations to better understand the data.

Insights:

The chart below shows the status of **Hired** and **Rejected** people.

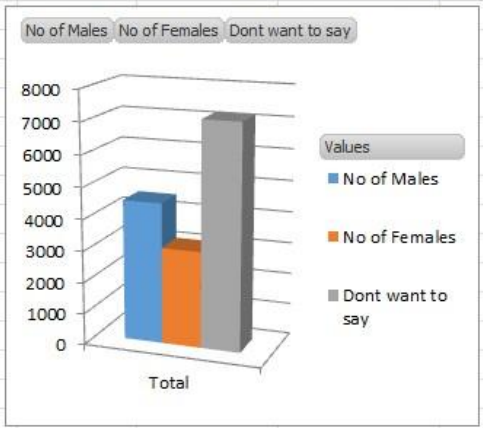
The chart below shows the **Average salary offered** according to department .

Row Labels	Average of Offered Salary
Finance Department	49628.00694
General Management	58722.09302
Human Resource Department	49002.27835
Marketing Department	48489.93538
Operations Department	49151.35438
Production Department	49448.48421
Purchase Department	52564.77477
Sales Department	49318.50201
Service Department	50629.88418
Grand Total	**49983.78153**



The chart below shows no of males, females and don't want to say values.

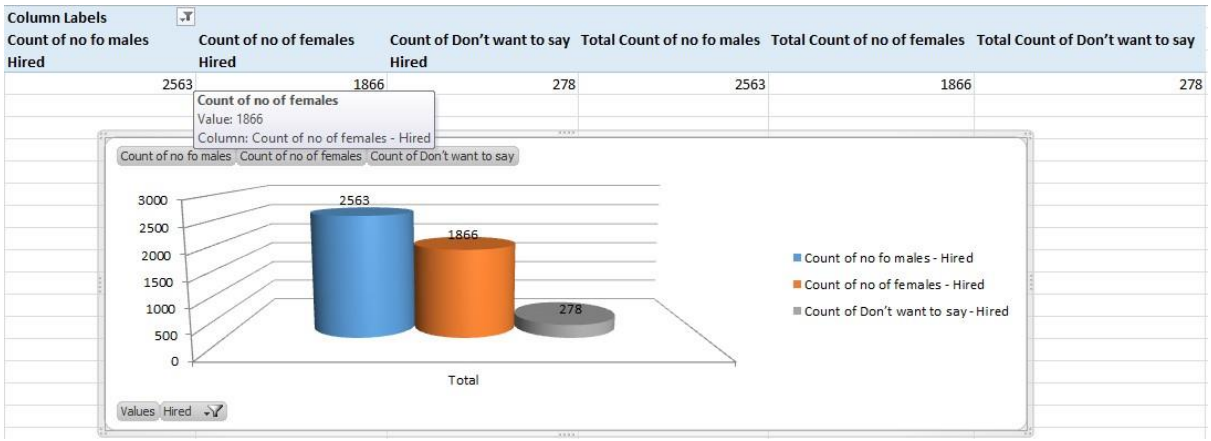
No of Males	No of Females	Dont want to say
4493	3083	7153



DATA ANALYTICS TASK:

A. Hiring Analysis: The hiring process involves bringing new individuals into the organization for various roles.

Your Task: Determine the gender distribution of hires. How many males and females have been hired by the company?



Insights:

More than half of the hired candidates are Males (2563). Hired Females are 1866. The rest haven't disclosed yet.

B. <u>Salary Analysis:</u> 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.

Your Task: What is the average salary offered by this company? Use Excel functions to calculate this.

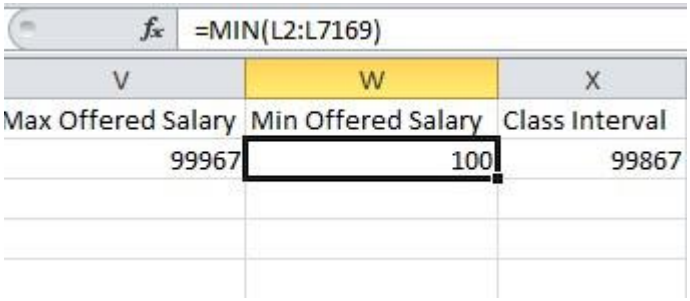
Average Salary	Average Slary of Hired candidates
49881.4	49596.65

<u>Insights:</u>

Average Offered Salary is **49881.4** and **Average Offered Salary** of **Hired Candidates** is **49596.65**. This shows that the hiring team is recruiting candidates as per the pre-determined salary bands of the organization.

C. <u>Salary Distribution:</u> 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.

Your Task: Create class intervals for the salaries in the company. This will help you understand the salary distribution.



Row Labels	Count of Candidates Offered	Count of Candidates Hired
(blank)		
100-10099	444	242
10100-20099	487	241
20100-30099	457	254
30100-40099	488	225
40100-50099	523	253
50100-60099	496	259
60100-70099	450	248
70100-80099	479	254
80100-90099	462	254
90100-100099	408	241
190100-200099	1	
290100-300099	1	
390100-400099	1	
Grand Total	**4697**	**2471**

<u>Insights:</u>

Class Interval is 99867.

Formula: Max Offered Salary – Min Offered Salary

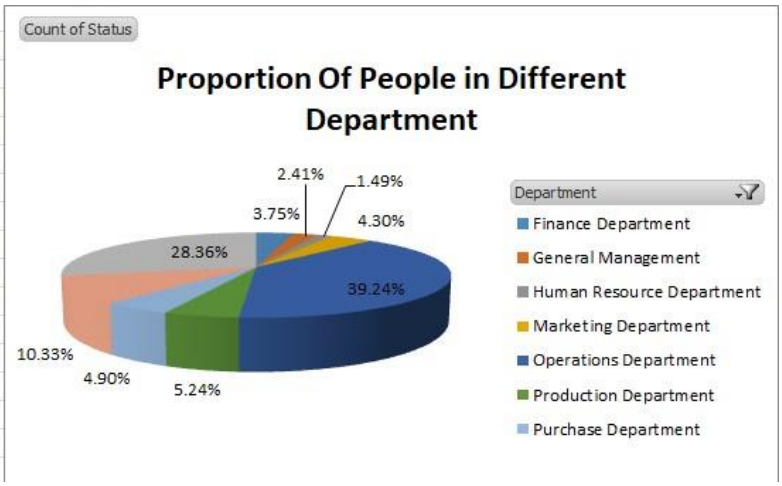
Maximum offered salary is in the interval of 40100–50099 while minimum offered salary are in intervals of 190100–200099, 290100-300099 and 390100-400099.

Maximum Hired salary is in the interval of 50100-60099 while minimum offered salary are in interval of 30100-40099.

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

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

Row Labels	Count of Status
Finance Department	3.75%
General Management	2.41%
Human Resource Department	1.49%
Marketing Department	4.30%
Operations Department	39.24%
Production Department	5.24%
Purchase Department	4.90%
Sales Department	10.33%
Service Department	28.36%
Grand Total	**100.00%**



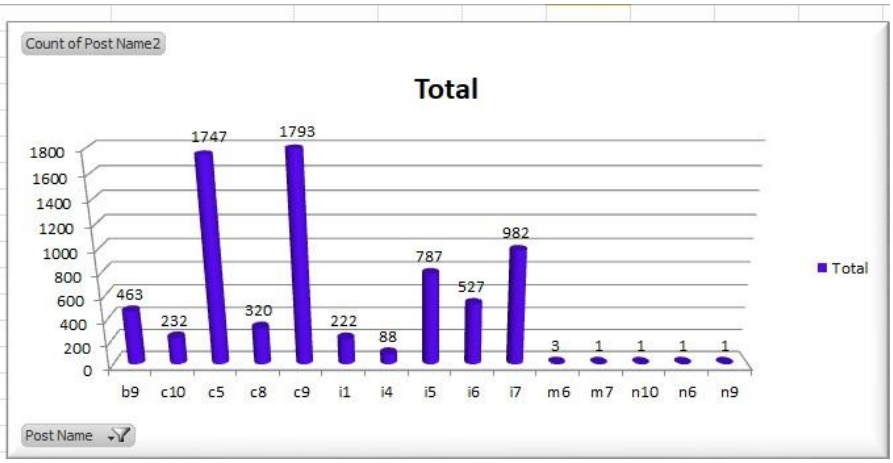
<u>Insights:</u>

Maximum candidates hired in **Operations Department** and Minimum candidates are hired in **Human Resource Department**.

E. <u>Position Tier Analysis:</u> Different positions within a company often have different tiers or levels.

Your 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.

Row Labels	Count of Post Name2
b9	463
c10	232
c5	1747
c8	320
c9	1793
i1	222
i4	88
i5	787
i6	527
i7	982
m6	3
m7	1
n10	1
n6	1
n9	1
Grand Total	**7168**



<u>Insights:</u>

Most candidates are hired in c9 post and least candidates are hired in m7, n10, n6, n9 posts.