Trainity – Project 4

Mayank Kumar Gandharv

HIRING PROCESS ANALYTICS

Finding missing data if any:

There are no missing data in the provided dataset as such. Some of the data in “event_name” column is missing but it would be best practice if that particular row is not deleted because corresponding to that missing data there is other column data associated which could be used to make analysis better and get more insights.

Clubbing Columns:

I don’t think any column could be clubbed together because every column has its unique data.

Changing column name:

The given dataset has a column named “event_name”. this column name is changed to “Gender”.

Outlier detection:

There is no need to calculate outliers because the all the attributes in the table is needed for analysis.

DATA ANALYTICS TASKS:

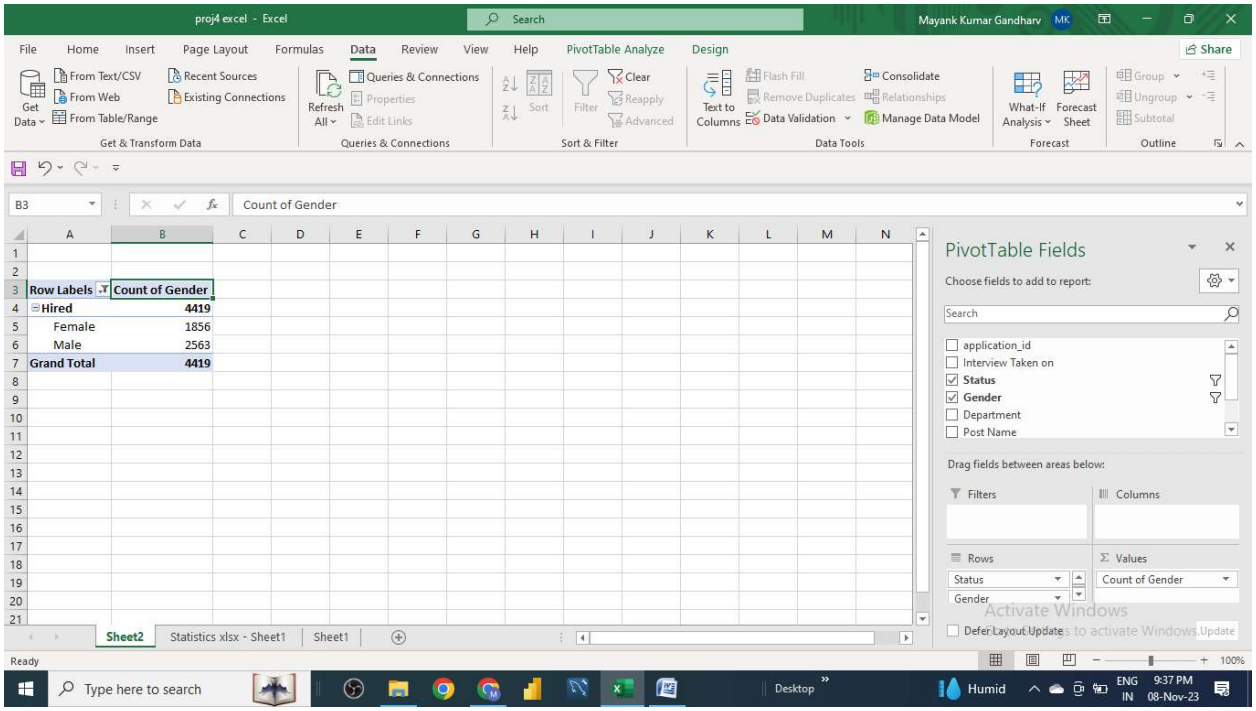
A. **Hiring Analysis:** The hiring process involves bringing new individuals into the organization for various roles.
Your Task: Determine the gender distribution of hires. How many males and females have been hired by the company?

SOLUTION:
To solve this i used “COUNTIFS” function in order to calculate the number of hired male and female candidates from the dataset. The syntax is given below:
For Male:
=COUNTIFS(D:D,"Male",C:C,"hired")
Output:
2563

For Female:
=COUNTIFS(D:D,"Female",C:C,"hired")
Output:
1856

OR

BY USING PIVOTR TABLE:



OUTPUT:

Row Labels	Count of Gender
Hired	**4419**
Female	1856
Male	2563
Grand Total	**4419**

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.

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

SOLUTION:

Syntax to find average salary using sum and count:

=SUM(G:G)/COUNT(G:G)

USING AVERAGE:

=AVERAGE(G:G)

OUTPUT:

49983.03

USING AVERAGEIF:

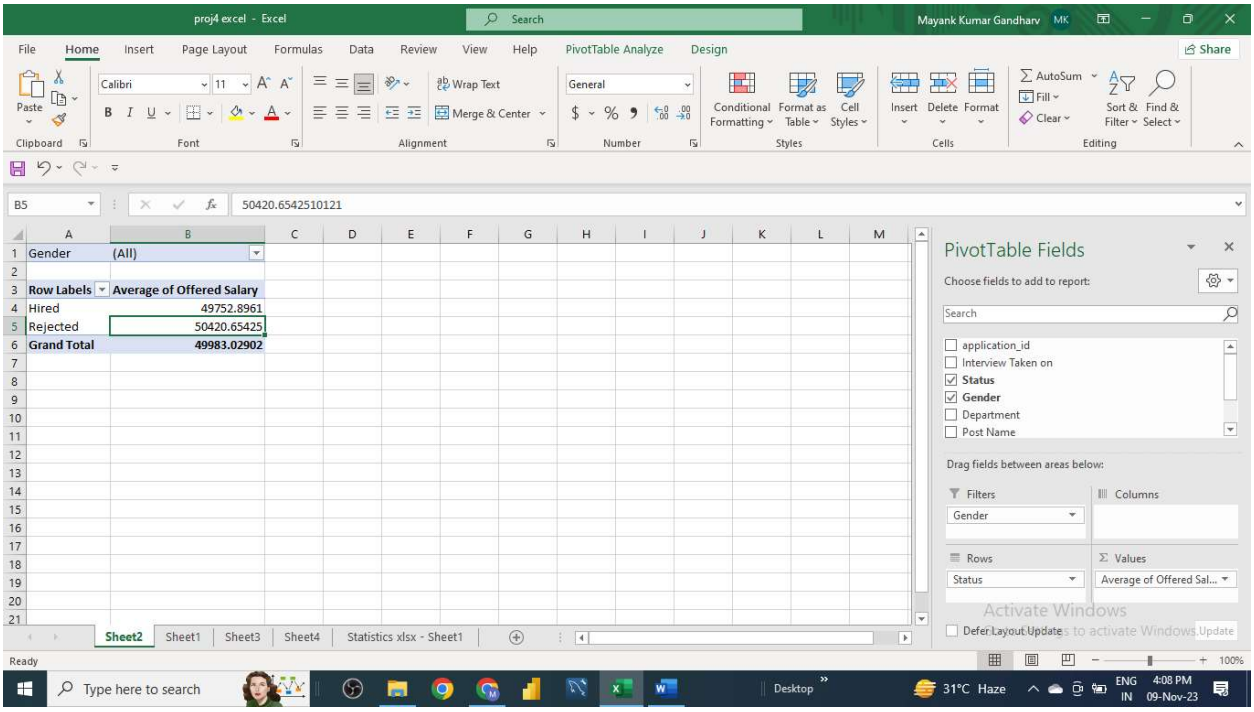
=AVERAGEIF(C:C,"Hired",G:G)

OUTPUT:

49752.9

OR

In order to calculate the average salary we need to calculate the avg salary for those who are already working in the organization (i.e Hired). Since there is more than 2 genders i.e "-" and "Donâ€™t want to say", I have included these genders as well because these gender entries are also been specified in other columns such as "Department" and "Offered Salary".



OUTPUT:

Row Labels	Average of Offered Salary
Hired	49752.8961
Rejected	50420.65425
Grand Total	**49983.02902**

The average salary provided to the employees is **49,753.**

And the average salary provided (regardless hired or rejected) is **49,983.**

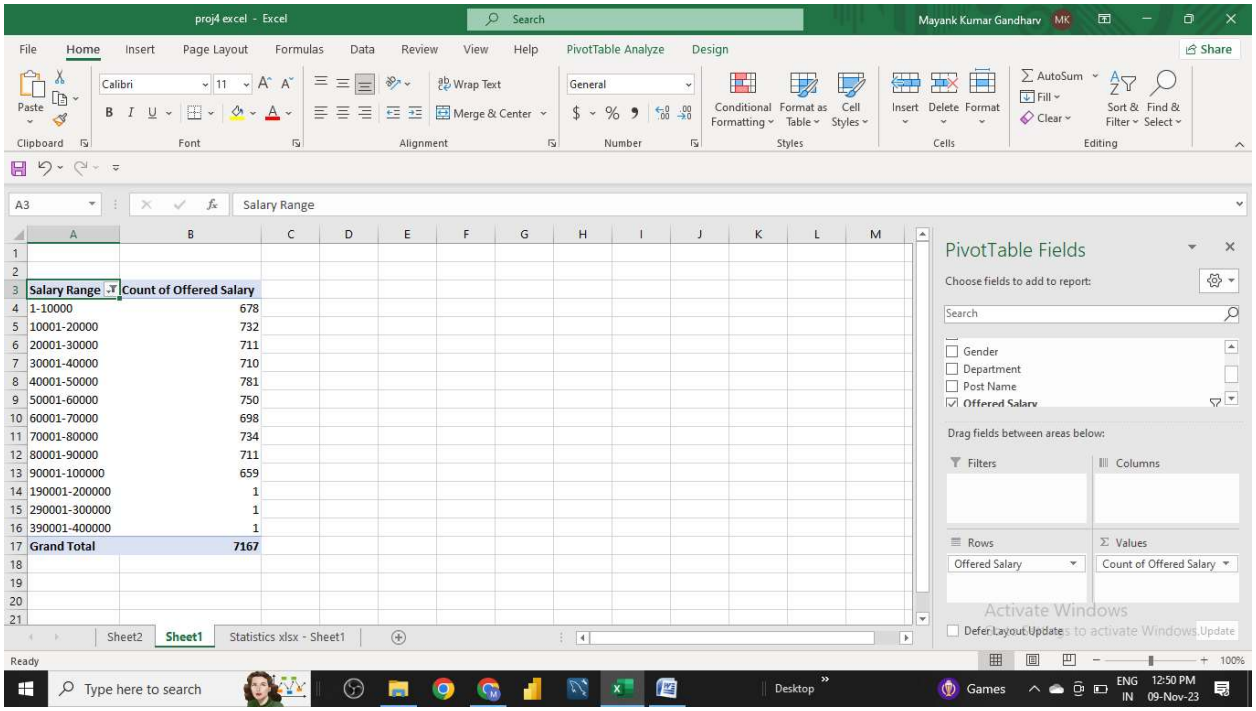
C. **Salary Distribution:** Class intervals represent ranges of values, in this case, salary ranges. The class interval is the difference between the upper and lower limits of a class.

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

SOLUTION:

In order to calculate the class interval of the salary first I used pivot table for all the candidates who were offered salaries irrespective of hired or rejected.

Then, grouped the offered salary with respect to the count of offered salary in 10,000 interval.



INSIGHT:

The maximum salary distribution is in the range of **40001-50000** and minimum is:

190001-200000
290001-300000
390001-400000

OUTPUT:

Salary Range	Count of Offered Salary
1-10000	678
10001-20000	732
20001-30000	711
30001-40000	710
40001-50000	781
50001-60000	750
60001-70000	698
70001-80000	734
80001-90000	711
90001-100000	659
190001-200000	1
290001-300000	1
390001-400000	1
Grand Total	**7167**

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

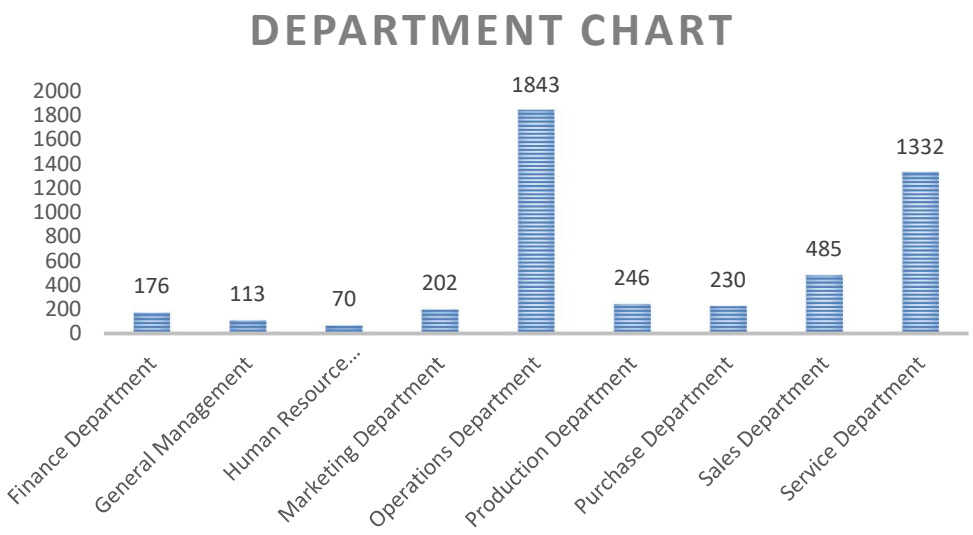
Your Task: Use a pie chart, bar graph, or any other suitable visualization to show the proportion of people working in different departments.

SOLUTION:

To solve this task I used “COUNTIFS” FUNCTION.

Finance Department: =COUNTIFS(E:E,"Finance Department",C:C,"Hired")
General Management: =COUNTIFS(E:E,"General Management",C:C,"Hired")
Human Resource Department: =COUNTIFS(E:E,"Human Resource Department",C:C,"Hired")

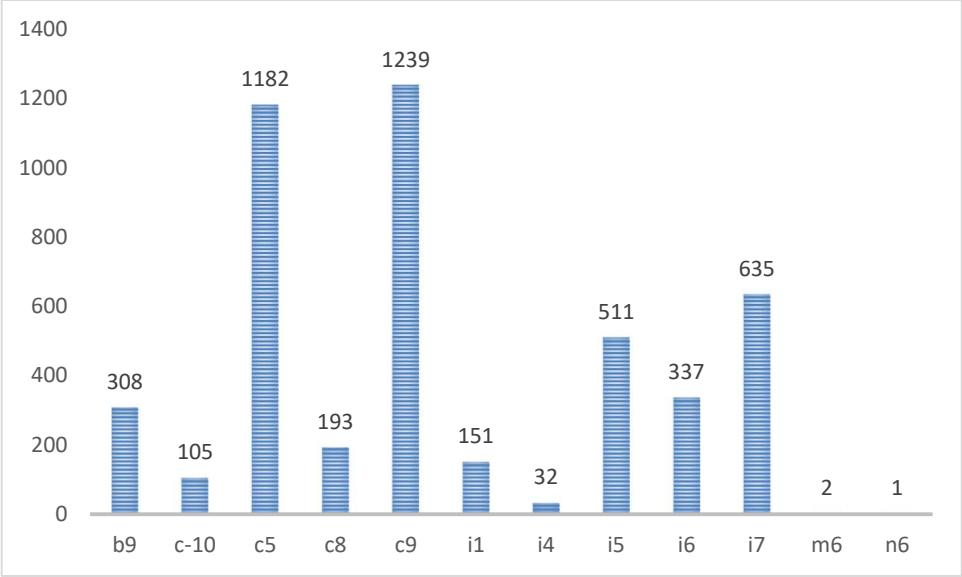
Marketing Department: =COUNTIFS(E:E,"Marketing Department",C:C,"Hired")
Operations Department: =COUNTIFS(E:E,"Operations Department",C:C,"Hired")
Production Department: =COUNTIFS(E:E,"Production Department",C:C,"Hired")
Purchase Department: =COUNTIFS(E:E,"Purchase Department",C:C,"Hired")
Sales Department: =COUNTIFS(E:E,"Sales Department",C:C,"Hired")
Service Department: =COUNTIFS(E:E,"Service Department",C:C,"Hired")



E. Position Tier Analysis: 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.

SOLUTION:
In order to do this task I used “COUNTIFS” Function and calculated the count of each tier and grouped by tiers itself.

SYNTAX:
=COUNTIFS(F:F,”Tier_Name",C:C,"Hired")


1400
1200
1000
800
600
400
200
0
308
105
1182
193
1239
151
32
511
337
635
2
1
b9
c-10
c5
c8
c9
i1
i4
i5
i6
i7
m6
n6