Project Description:

- Hiring process is very important for a company and understanding trends such as the number of rejections, interviews, job types, and vacancies can provide valuable insights for the hiring department.
- So, the task is to analyse the given data so that we could provide better insights and make this process easier.

Approach:

- For this we will clean the data, by clearing the missing data, detecting the outliers.
- We create charts and calculate few values for this purpose.

Tech-Stack Used:

- MS Excel 2019 has been used for this task.

Insights:

The following are found using the excel functions:

- Hiring analysis based on gender.
- Average Salary Analysis
- Distribution of salaries
- Department Analysis
- Position tier analysis

Understanding the **Data**

Application ID – INT datatype and is unique.

Interview taken on – Date type.

Status – String

Event_name – Gender: string

Department – String datatype

Post_name – both text and numbers

Offered salary – int datatype.

Cleaning Data----------

Application id – no missing data

Interview taken on – no missing data

Status – no missing data

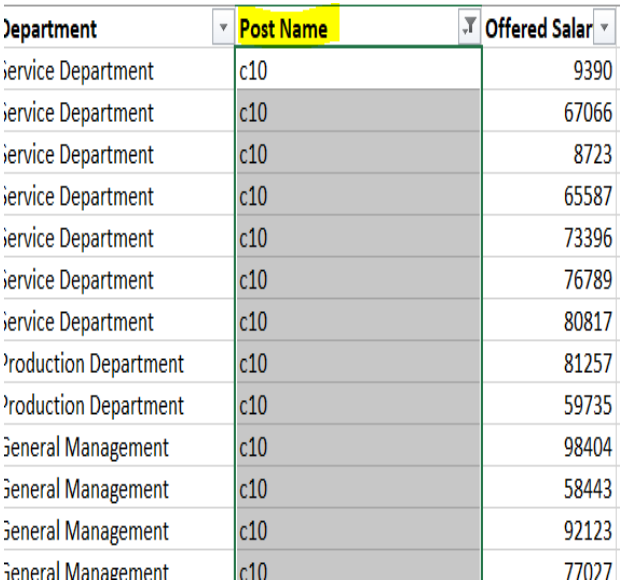
Event_name – There are rows where the data is missing so we delete them

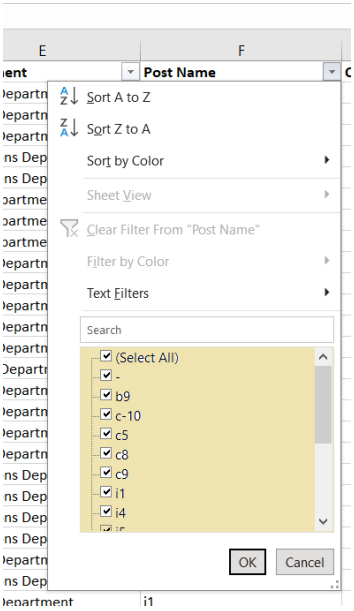
Department – No missing data

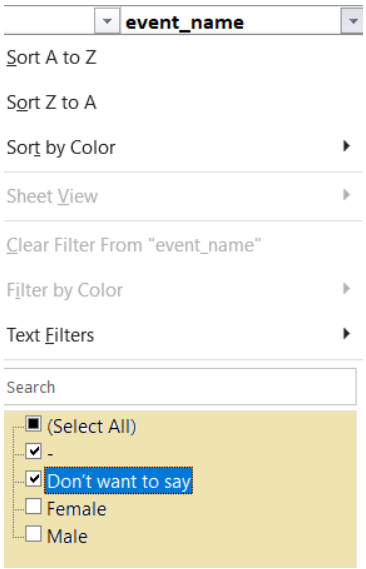
Post_name – 1 missing data

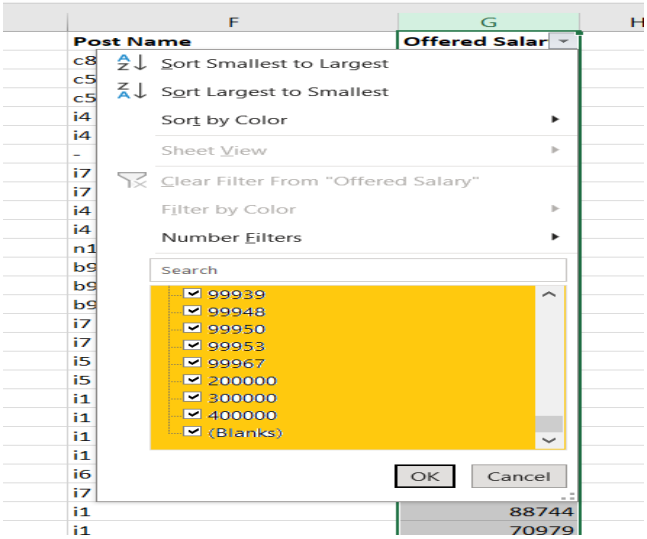
Offered salary – 1 blank row

- Clearing missing and formatting the dataset, helps us to analyse the data in a better way.
- Below we have removed the blanks from offered salary, changed “c-10” to “c10” as all other post_name includes a number and an alphabet.
- We have removed rows containing “-” in event_name column









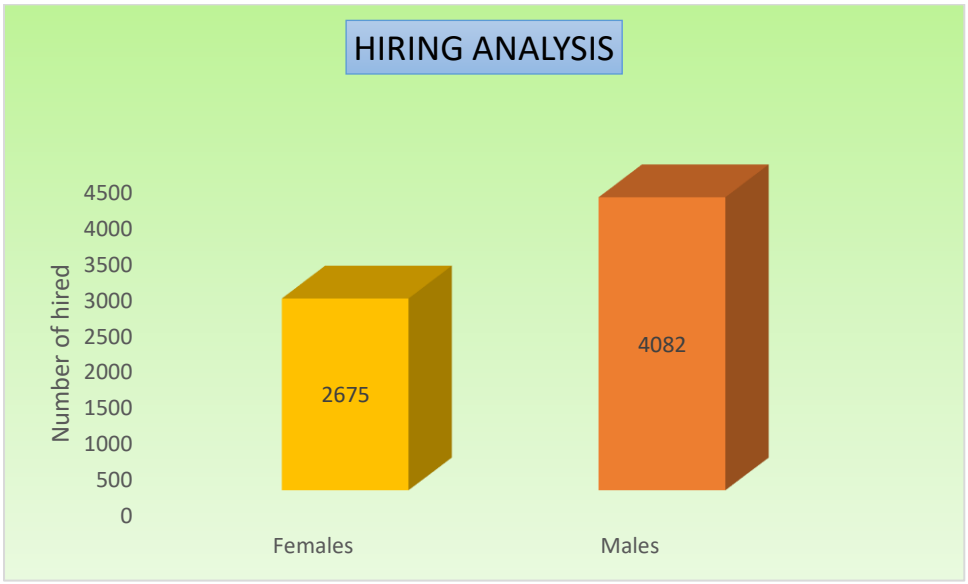
Finding outliers in the data----------

Values(SALARY)			
MIN	800		
q1	25516.5	48871	IQR
q2	49625		
q3	74387.5	147694	Upper Bound
MAX	400000	-47790	Lower Bound

- There are few outliers in the data which has been removed. Outliers can be misleading hence its important to remove them.
- Any offered salary below -47790 and above 147694, will be removed from the dataset.
- Now we have 7152 rows in the data.

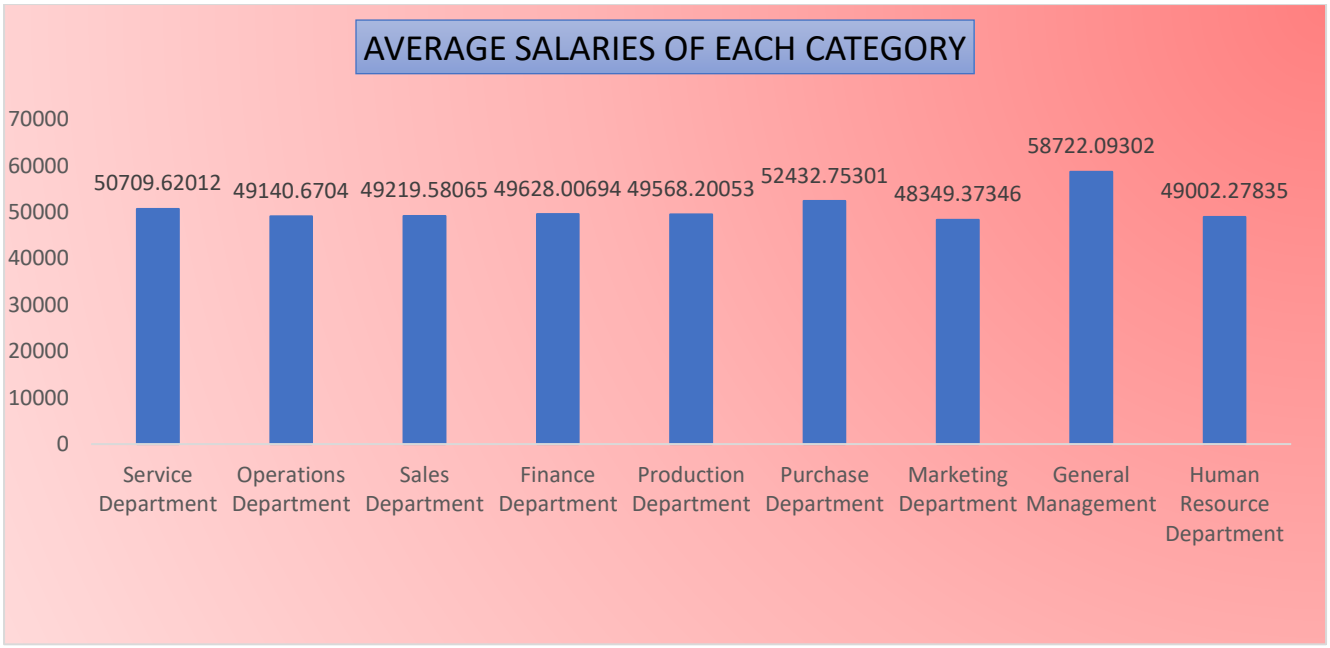
Tasks----------

Hiring Analysis based on gender:



Males are hired more than women. We use the CountIF formula for this task. (The big picture is shown in the excel sheet that is attached)

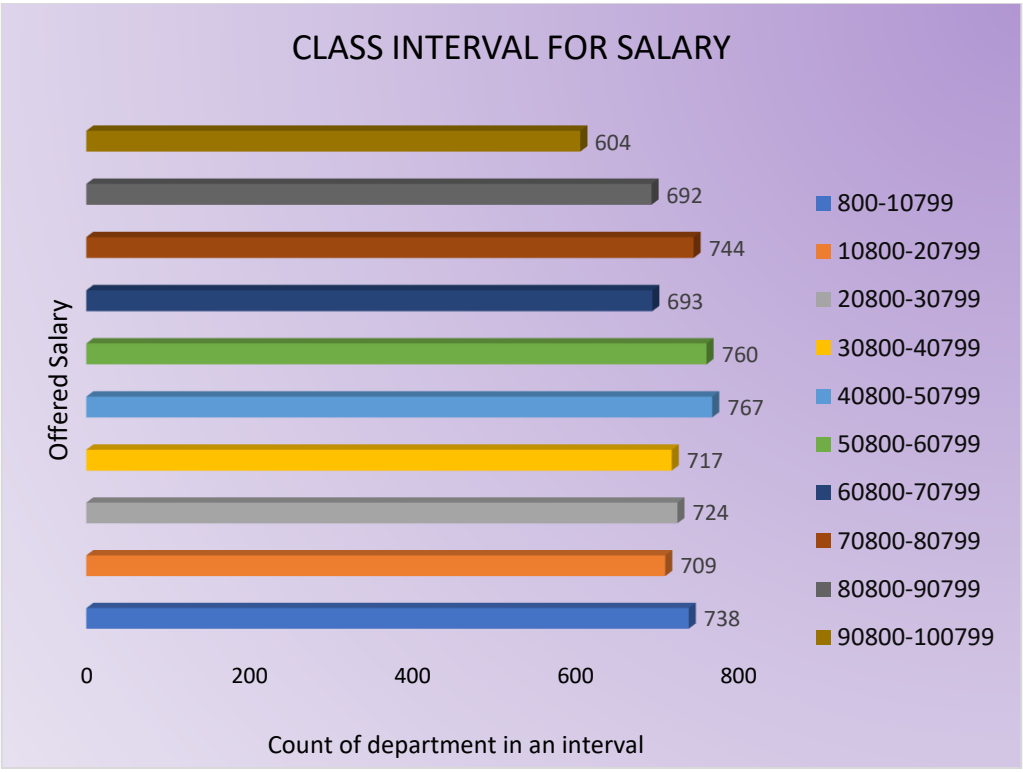
Salary Analysis:



- The average salaries of each department is calculated using AverageIF() along with round() for rounding it of to 5 places after the decimal.
- The least average salary is for Marketing Department i.e. 48349.37346

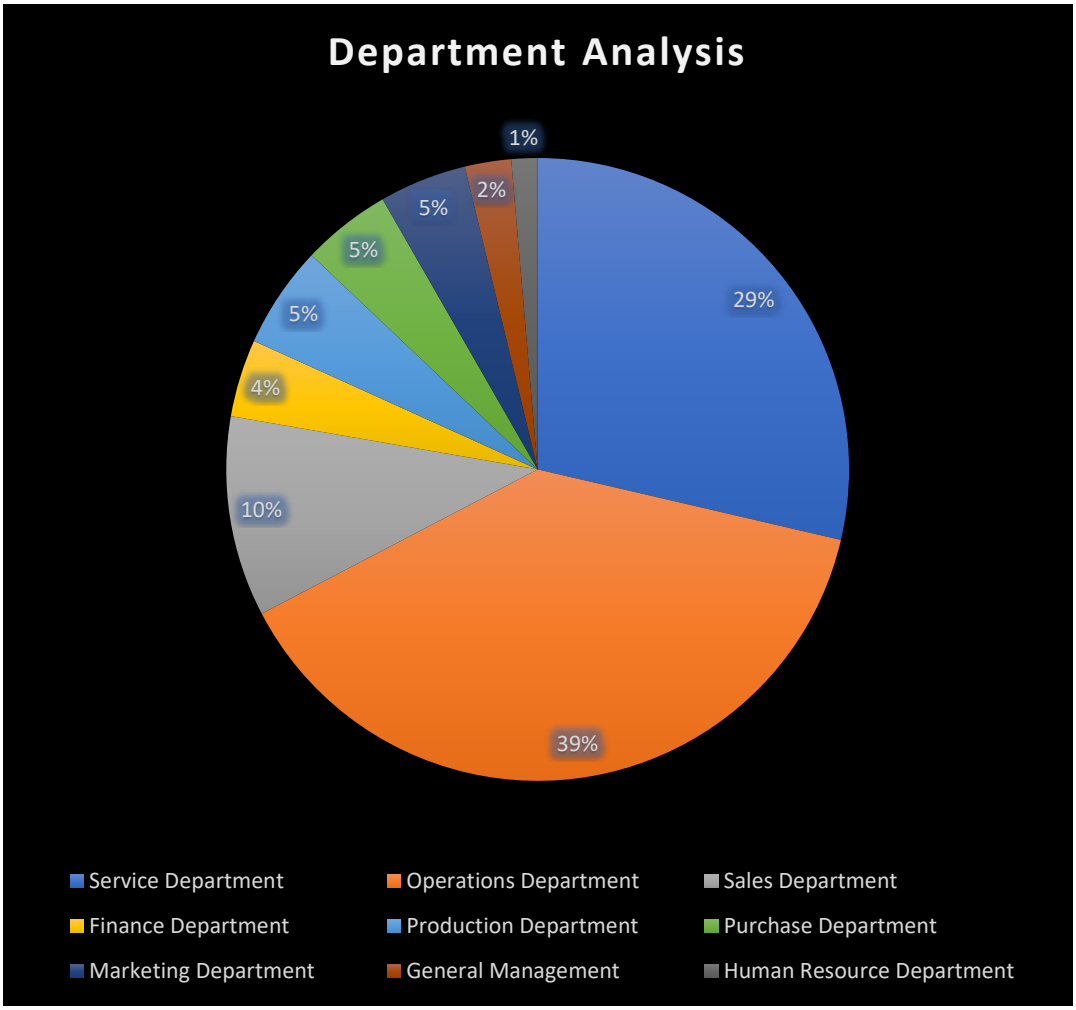
Salary Distribution:

- **Here we find the maximum number of departments fall under which interval.**
- **Clearly the maximum number is 767 which is under 40800-50799 interval.**



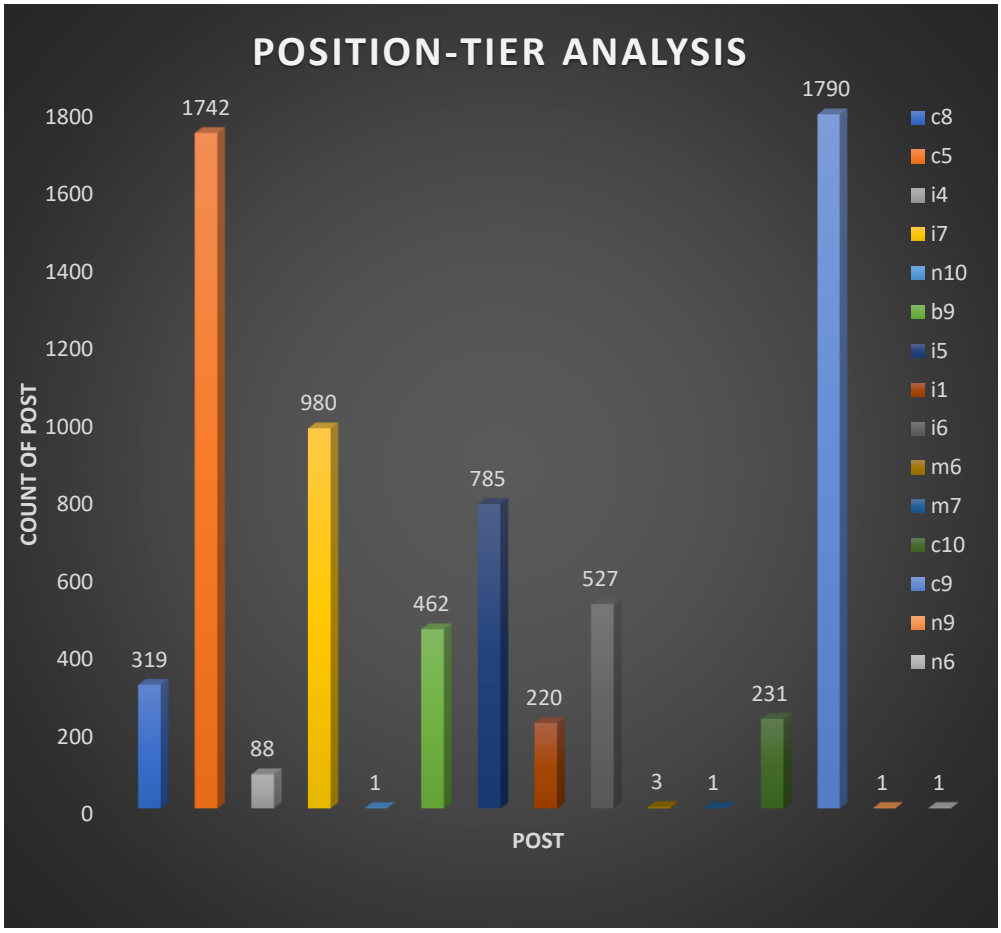
Departmental Analysis:

- **To show what proportion of people work under each category.**
- **We have used the pie chart. Maximum % of people are in the operations department.**
- **Production, purchase and marketing have 5% people working**



Position Tier Analysis:

- **<u>This is the distibution of post_tier.</u>**
- **<u>n10,n6,n7 and M7 have only 1 person each.</u>**
- **<u>C5 is the second highest and C9 is the highest. (The complete analysis is on excel workbook).</u>**



Results:

- **We get familiar with excel functions and learn to create charts.**
- **We learn to clean the data which is a cruical role for data analysis**
- **We learn about the quartiles and IQR and outliers.**