Hiring Process Analytics

ABHISHEK BHAKAT

Agenda

Project Description

Approach

Tech-Stack Used

Insights

Results

Project Description

The aim of this project is to analyze the hiring trend of a MNC using a dataset provided by the company. As a lead Data Analyst in Google the task is to extract the meaningful insight from the data provided the detail report to the hiring department. The dataset include about people who registered for a particular post in a department of this company.

Approach

- Understanding Data Columns and their content.
- Checking for missing data.
- Clubbing columns with multiple category.
- Identifying and handling outliers.
- Create data summary.

Tech-Stack Used

Microsoft Excel 365

Excel was chosen for its extensive data analysis functionalities, including pivot tables, charts and statistical function. It facilitated in-depth exploration and visualization of the hiring data. Enabling a through understanding of the underlying trends and patterns.

Data Cleaning

Handling Missing Values

1. 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 projects i.e. gender of the candidate in not Known.

2. Column **offered salary** has 1 rows with null values. 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** Posts Name. The median comes out to **45400.**

3. Column **Post Name** has 1 rows with "**-**" as its values. These can be termed as Null values. 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 **85000** and **96000** which is "**c9**"

Data Cleaning

Error Rectification

1. Column **Post Name** has a category **“c-10”** which seems to be a typo and the correct category should be “**c10**” which we rectified.

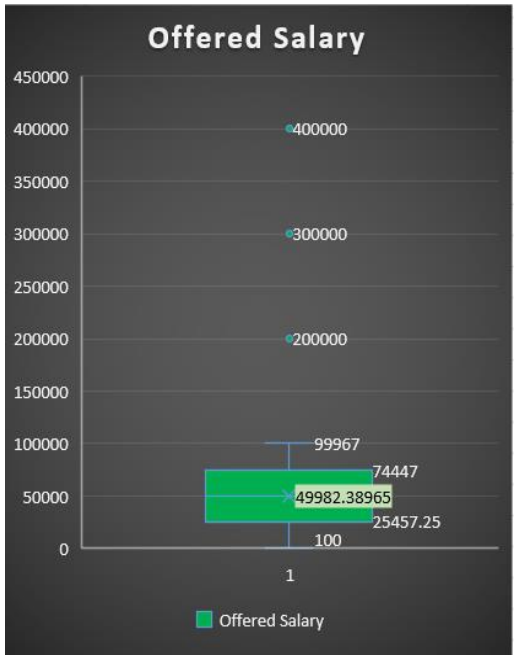
Handling Duplicate Values

Column **application_id** 54 rows with duplicate values. They should be either be removed or replaced with correct value.

Data Cleaning

Handling Outliers

From the **Box plot** of column **Offered salary**, we can see that there are three rows whose columns values are **200000, 300000, 400000**. We replaced them with median value of **Offered Salary** for corresponding Department and **Post Name.**




INSIGHTS

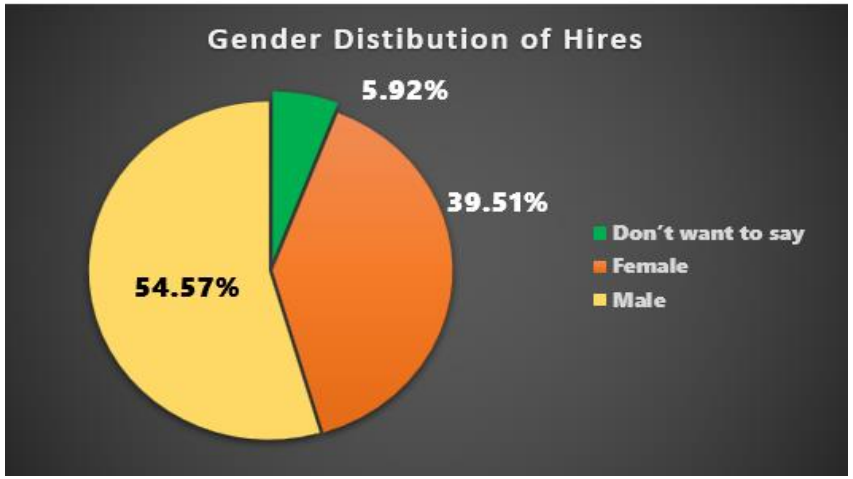
Hiring Analysis

The Hiring Process involves bringing new individuals into the organization for various roles.

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

Status	Hired

Row Labels	Count of event_name
Don't want to say	278
Female	1856
Male	2563
Grand Total	**4697**



Hiring Analysis

INSIGHTS

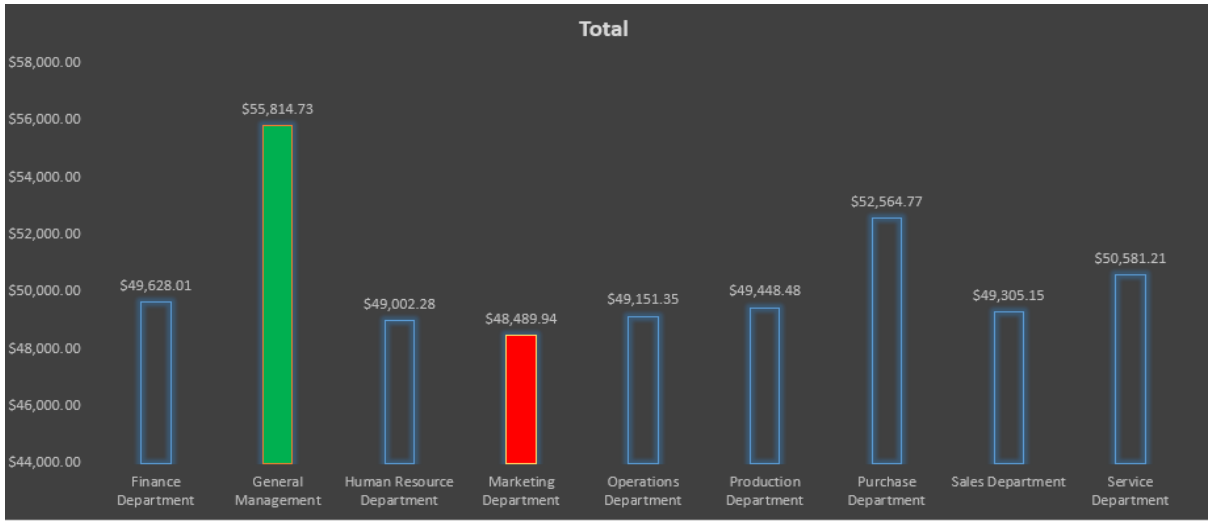
More than half of the hired candidate are male and only 39.51% are female. The rest haven’t disclose there gender. High Gender Ration (Ratio of Male to Female) may negatively impact the Organization’s image in public domain. The organization should focus on decreasing the Gender Ratio bringing close to 1.

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 function to calculate this.

Row Labels	Average of Offered Salary
Finance Department	$49,628.01
General Management	$55,814.73
Human Resource Department	$49,002.28
Marketing Department	$48,489.94
Operations Department	$49,151.35
Production Department	$49,448.48
Purchase Department	$52,564.77
Sales Department	$49,305.15
Service Department	$50,581.21
Grand Total	**$49,898.67**



Salary Analysis

INSIGHTS

The Average Offered Salary is **$49898.67**

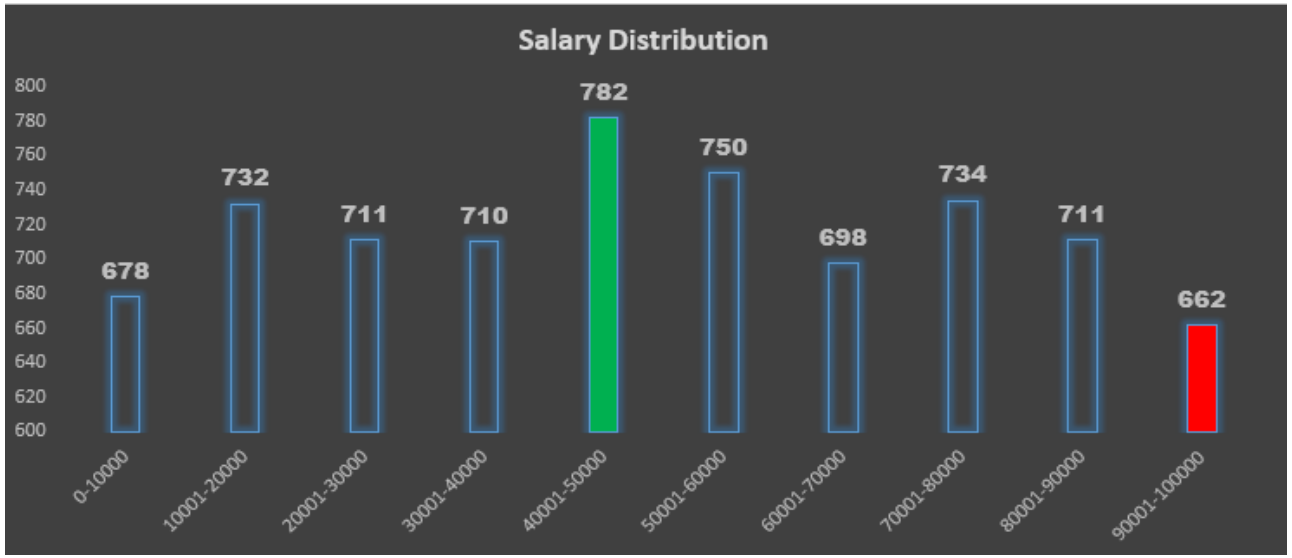
The Average Offered Salary form the **General Management** Department has the highest Salary Offered and the **Marketing Department** has the lowest Salary offered among all.

Salary Distribution

Class interval represent ranges of values, in this case, salary ranges. The class interval is the difference between the upper and the lower limits of a class.

TASK : Create a class intervals for the salaries in the company. This will help you to understand the salary distribution.

Row Labels	Count of application_id
0-10000	678
10001-20000	732
20001-30000	711
30001-40000	710
40001-50000	782
50001-60000	750
60001-70000	698
70001-80000	734
80001-90000	711
90001-100000	662
Grand Total	7168



Salary Distribution

INSIGHTS

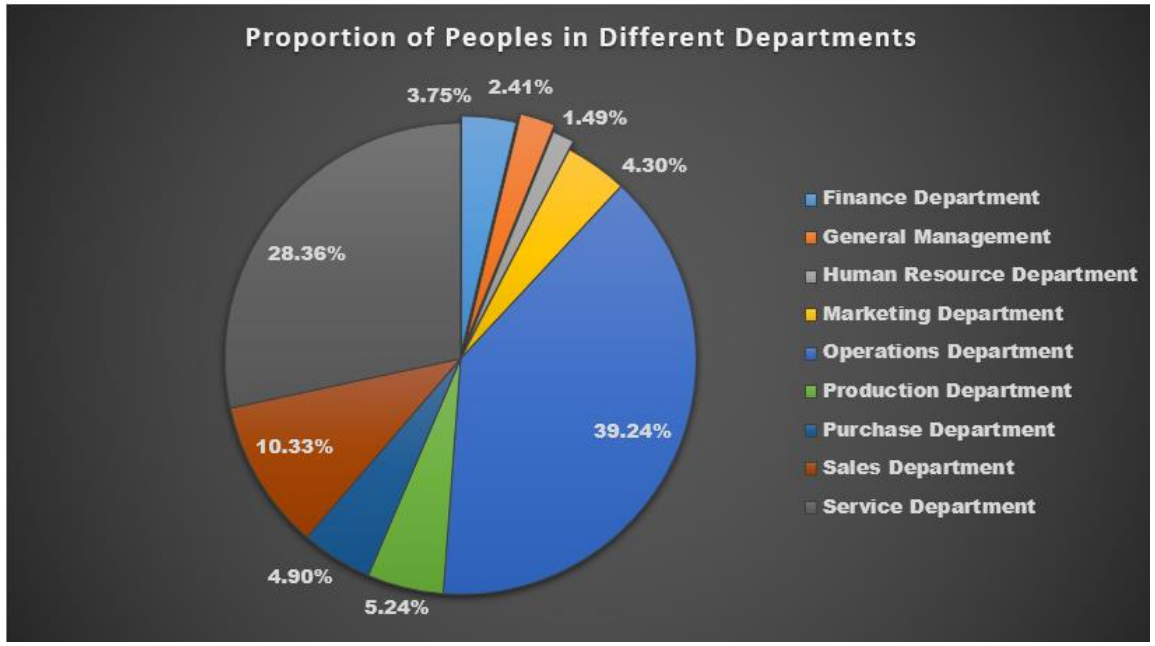
We can observe that the maximum offered salary is in the interval of $40001-50000. That is most of the job requirements was for the middle experience and least for senior most post and for freshers.

Departmental Analysis

Visualizing data through charts and plots is a crucial part of data analysis.

TASK : Use pie charts, bar graph or any others suitable visualization to show the proportion of people working in different departments.

Status	Hired	
Row Labels	**No. of Employees**	**Percentages**
Finance Departme	176	3.75%
General Managem	113	2.41%
Human Resource D	70	1.49%
Marketing Departn	202	4.30%
Operations Depart	1843	39.24%
Production Depart	246	5.24%
Purchase Departm	230	4.90%
Sales Department	485	10.33%
Service Departmer	1332	28.36%
Grand Total	**4697**	**100.00%**



Departmental Analysis

INSIGHTS

From the above pie chart we can observe that most candidates are hired in **Operations Department** followed by **Service Department** and **Sales Department** and least candidates are hired in **Human Resource Department.**

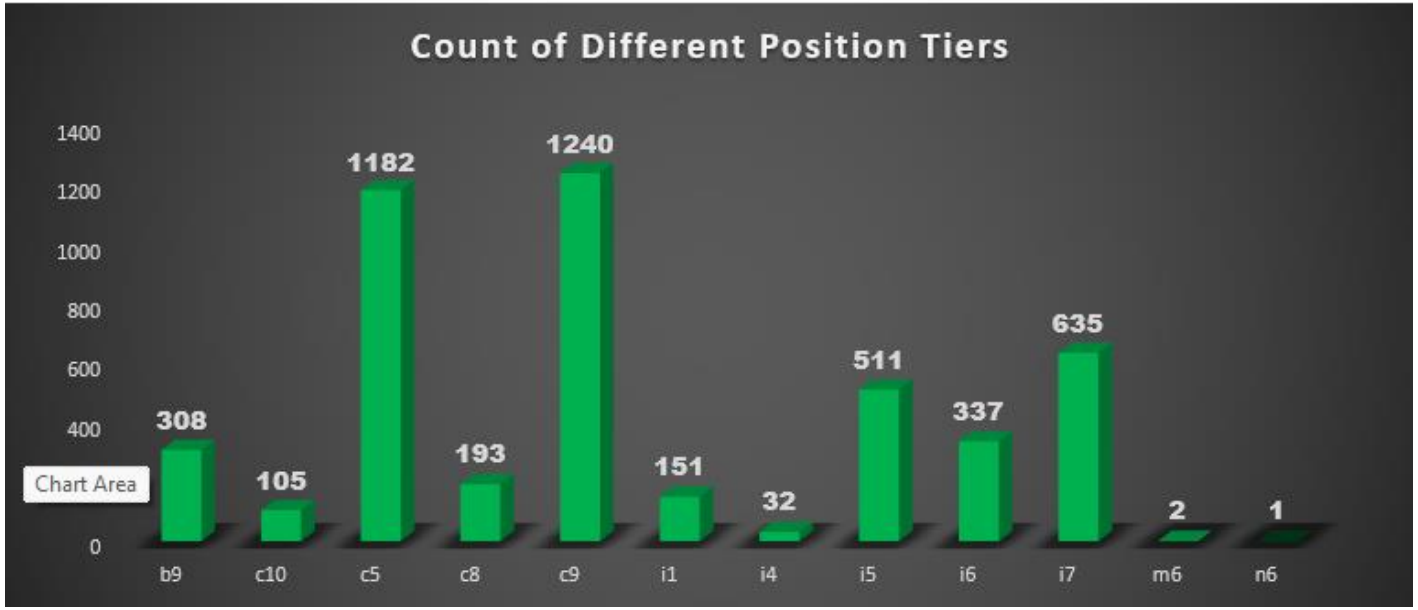
Position Tier Analysis

Different position within a company often have different tiers or levels.

TASK : Use charts or graph to represent the different position within the company. This will help you understand the distribution of position across different tiers.

Status	Hired

Row Labels	Count of Post Name
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**



Position Tier Analysis

INSIGHTS

Here, we can observe that the organization has hired most candidates for post tier **c9**
Followed by **c5** and then **i7** at distant third.

Summary

This project helped me in understanding how important Data Analytics is for Hiring Process of an organization as it provides valuable insights such as number of rejection, profile of applicants, vacancies etc. which helps the hiring department to take data Data-Driven Decisions.

Link to the dataset – Click Here

Thank you

Abhishek Bhakat

abhishekbhakat6@gmail.com