HIRING PROCESS ANALYTICS

Excel file link

Project by Mayur Rajput

Content

- Insights & Results
- Project description
- Data Analysis
- Approach and Tech-Stack Used
- Handling Outliers
- Handling Missing Data

Project Description

In this project, the goal was to use our knowledge of statistics and Excel to to analyze this data and answer certain questions that can help the company improve its hiring process and draw meaningful conclusions and insights that could potentially help the company improve its hiring process and make better hiring decisions in the future.

Approach & Tech-Stack Used

Approach :

- Handling missing data and outliers in given data
- To analyze the provided company's hiring process data and draw meaningful insights from it to understand trends such as the number of rejections, interviews, job types, and vacancies can provide valuable insights for the hiring department.

Tech-Stack Used :

Microsoft Excel is used to handle outliers, missing values and also used to perform analysis and answer certain questions that can help the company improve its hiring process.

Handling Missing Data

a) Finding missing values in dataset:

Finding number of missing values in each column						
7168	7168	7168	7153	7168	7167	7167
Finding % of missing values						
	0.00%	0.00%	0.21%	0.00%	0.01%	0.01%
application_id	Interview Taken on	Status	event_name	Department	Post Name	Offered Salar
383422	2014-05-01 11:40	Hired	Male	Service Department	c8	56553
907518	2014-05-06 08:08	Hired	Female	Service Department	c5	22075
176719	2014-05-06 08:08	Rejected	Male	Service Department	c5	70069
429799	2014-05-02 16:28	Rejected	Female	Operations Department	i4	3207
253651	2014-05-02 16:32	Hired	Male	Operations Department	i4	29668
289907	2014-05-01 07:44	Hired	Male	Sales Department		85914
959124	2014-05-06 16:27	Rejected	Male	Sales Department	i7	69904
86642	2014-05-09 13:17	Rejected	Male	Sales Department	i7	11758
751029	2014-05-02 13:09	Hired	Female	Service Department	i4	15156
434547	2014-05-02 13:11	Rejected	Female	Service Department	i4	49515
518854	2014-05-01 09:00	Rejected	Male	Service Department	n10	26990
649039	2014-05-07 10:48	Hired	Female	Service Department	b9	200000
199526	2014-05-07 10:50	Hired	Male	Service Department	b9	86787
539803	2014-05-15 09:31	Hired	Male	Finance Department	b9	2308
191009	2014-05-09 12:48	Hired	Female	Service Department	i7	56688
195323	2014-05-09 12:48	Hired		Service Department	i7	81757
51318	2014-05-02 08:07	Hired	Male	Service Department	i5	15134
742283	2014-05-02 08:11	Rejected		Service Department	i5	100
513166	2014-05-01 22:53	Hired	Female	Operations Department	i1	73579
791372	2014-05-01 22:54	Rejected	Male	Operations Department	i1	50351
47857	2014-05-01 22:55	Rejected	Female	Operations Department	i1	38462
834101	2014-05-01 22:53	Rejected	Don't want to say	Operations Department	i1	82510

Missing values were found in columns named event_name, Post name and offered salary.

Handling Missing Data

b) Imputing missing values in dataset:

- Event_name column had 15 missing values which are imputed with "**Don't want to say** ".
- Post name column had 1 missing value which imputed with "**c9**" and it is calculated using Mode() function.
- Salary column had 1 missing value which imputed with "**49983**" and it is calculated using Mean() function.

Finding number of missing values in each column						
7168	7168	7168	7168	7168	7168	7168
Finding % of missing values						
	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
application_id	Interview Taken on	Status	event_name	Department	Post Name	Offered Salar
383422	41760.48668	Hired	Male	Service Department	c8	56553
907518	41765.33926	Hired	Female	Service Department	c5	22075
176719	41765.33951	Rejected	Male	Service Department	c5	70069
429799	41761.68679	Rejected	Female	Operations Department	i4	3207
253651	41761.68919	Hired	Male	Operations Department	i4	29668
289907	41760.32243	Hired	Male	Sales Department	c9	85914
959124	41765.68605	Rejected	Male	Sales Department	i7	69904
86642	41768.55389	Rejected	Male	Sales Department	i7	11758
751029	41761.54858	Hired	Female	Service Department	i4	15156
434547	41761.54943	Rejected	Female	Service Department	i4	49515
518854	41760.37563	Rejected	Male	Service Department	n10	26990
649039	41766.45029	Hired	Female	Service Department	b9	200000
199526	41766.45145	Hired	Male	Service Department	b9	86787
539803	41774.39688	Hired	Male	Finance Department	b9	2308
191009	41768.53398	Hired	Female	Service Department	i7	56688
195323	41768.53373	Hired	Don't want to say	Service Department	i7	81757
51318	41761.33829	Hired	Male	Service Department	i5	15134
742283	41761.34105	Rejected	Don't want to say	Service Department	i5	100
513166	41760.95384	Hired	Female	Operations Department	i1	73579

Handling Outliers

a) Finding outliers in dataset:

Q1	25464
Median (Q2)	49628
Q3	74429
IQR	48965
Upper bound	147877
Lower bound	-47984

application_id	Interview Taken on	Status	event_name	Department	Post Name	Offered Salar	Outlier
383422	2014-05-01 11:40	Hired	Male	Service Department	c8	56553	FALSE
907518	2014-05-06 08:08	Hired	Female	Service Department	c5	22075	FALSE
176719	2014-05-06 08:08	Rejected	Male	Service Department	c5	70069	FALSE
429799	2014-05-02 16:28	Rejected	Female	Operations Department	i4	3207	FALSE
253651	2014-05-02 16:32	Hired	Male	Operations Department	i4	29668	FALSE
289907	2014-05-01 07:44	Hired	Male	Sales Department	c9	85914	FALSE
959124	2014-05-06 16:27	Rejected	Male	Sales Department	i7	69904	FALSE
86642	2014-05-09 13:17	Rejected	Male	Sales Department	i7	11758	FALSE
751029	2014-05-02 13:09	Hired	Female	Service Department	i4	15156	FALSE
434547	2014-05-02 13:11	Rejected	Female	Service Department	i4	49515	FALSE
518854	2014-05-01 09:00	Rejected	Male	Service Department	n10	26990	FALSE
649039	2014-05-07 10:48	Hired	Female	Service Department	b9	200000	TRUE
199526	2014-05-07 10:50	Hired	Male	Service Department	b9	86787	FALSE
539803	2014-05-15 09:31	Hired	Male	Finance Department	b9	2308	FALSE
191009	2014-05-09 12:48	Hired	Female	Service Department	i7	56688	FALSE
195323	2014-05-09 12:48	Hired	Don’t want to say	Service Department	i7	81757	FALSE
51318	2014-05-02 08:07	Hired	Male	Service Department	i5	15134	FALSE
742283	2014-05-02 08:11	Rejected	Don’t want to say	Service Department	i5	100	FALSE
513166	2014-05-01 22:53	Hired	Female	Operations Department	i1	73579	FALSE
791372	2014-05-01 22:54	Rejected	Male	Operations Department	i1	50351	FALSE

Handling Outliers

b) Removing outliers in dataset:

- 3 Outliers were found in Salary column which are then replaced with upper limit as outliers value was more than upper limit.

Offered Salar	New_Offered_salary	Outlier
200000	147877	TRUE
400000	147877	TRUE
300000	147877	TRUE

Offered Salar	New_Offered_salary	Outlier
56553	56553	FALSE
22075	22075	FALSE
70069	70069	FALSE
3207	3207	FALSE
29668	29668	FALSE
85914	85914	FALSE
69904	69904	FALSE
11758	11758	FALSE
15156	15156	FALSE
49515	49515	FALSE
26990	26990	FALSE
200000	147877	TRUE
86787	86787	FALSE
2308	2308	FALSE
56688	56688	FALSE
81757	81757	FALSE
15134	15134	FALSE
100	100	FALSE
73579	73579	FALSE
50351	50351	FALSE
38462	38462	FALSE

Q1	25464
Median (Q2)	49628
Q3	74429
IQR	48965
Upper limit	147877
Lower limit	0

Cleaned Dataset

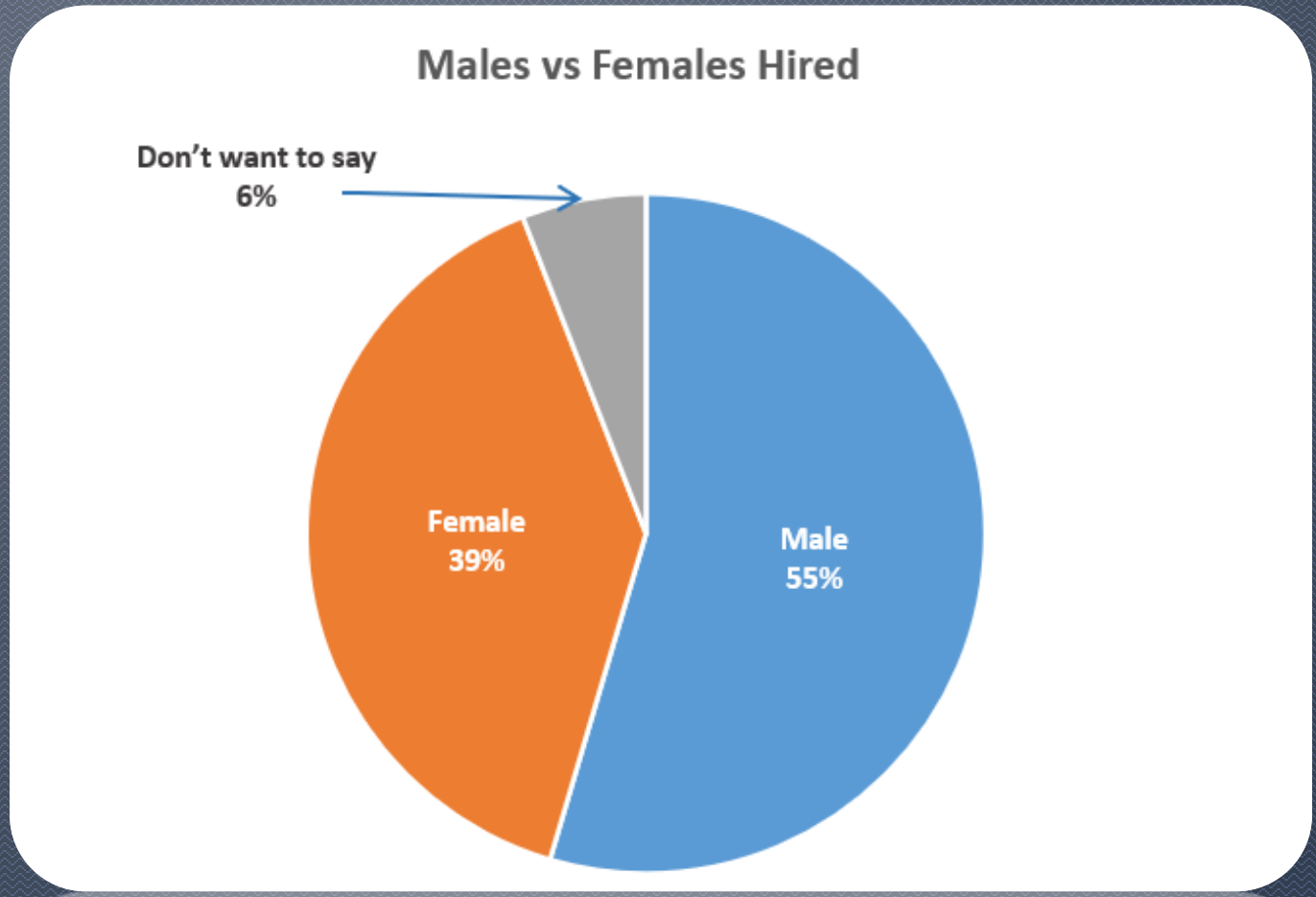
	A	B	C	D	E	F	G
1	application_id	Interview Taken on	Status	event_name	Department	Post Name	Offered Salary
2	383422	2014-05-01 11:40	Hired	Male	Service Department	c8	56553
3	907518	2014-05-06 08:08	Hired	Female	Service Department	c5	22075
4	176719	2014-05-06 08:08	Rejected	Male	Service Department	c5	70069
5	429799	2014-05-02 16:28	Rejected	Female	Operations Department	i4	3207
6	253651	2014-05-02 16:32	Hired	Male	Operations Department	i4	29668
7	289907	2014-05-01 07:44	Hired	Male	Sales Department	c9	85914
8	959124	2014-05-06 16:27	Rejected	Male	Sales Department	i7	69904
9	86642	2014-05-09 13:17	Rejected	Male	Sales Department	i7	11758
10	751029	2014-05-02 13:09	Hired	Female	Service Department	i4	15156
11	434547	2014-05-02 13:11	Rejected	Female	Service Department	i4	49515
12	518854	2014-05-01 09:00	Rejected	Male	Service Department	n10	26990
13	649039	2014-05-07 10:48	Hired	Female	Service Department	b9	147877
14	199526	2014-05-07 10:50	Hired	Male	Service Department	b9	86787
15	539803	2014-05-15 09:31	Hired	Male	Finance Department	b9	2308
16	191009	2014-05-09 12:48	Hired	Female	Service Department	i7	56688
17	195323	2014-05-09 12:48	Hired	Don’t want to say	Service Department	i7	81757
18	51318	2014-05-02 08:07	Hired	Male	Service Department	i5	15134
19	742283	2014-05-02 08:11	Rejected	Don’t want to say	Service Department	i5	100
20	513166	2014-05-01 22:53	Hired	Female	Operations Department	i1	73579
21	791372	2014-05-01 22:54	Rejected	Male	Operations Department	i1	50351
22	47857	2014-05-01 22:55	Rejected	Female	Operations Department	i1	38462
23	834101	2014-05-01 22:53	Rejected	Don’t want to say	Operations Department	i1	82510
24	985008	2014-05-01 09:41	Rejected	Male	Service Department	i6	52554
25	891568	2014-05-01 16:28	Hired	Female	Operations Department	i7	3423
26	935899	2014-05-10 14:17	Rejected	Male	Service Department	i1	88744
27	780839	2014-05-10 14:18	Hired	Female	Service Department	i1	70979

Data Analysis

Task 1 : Hiring Analysis

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

Status	Hired
Gender	**Count of employees**
Male	2563
Female	1856
Don’t want to say	278

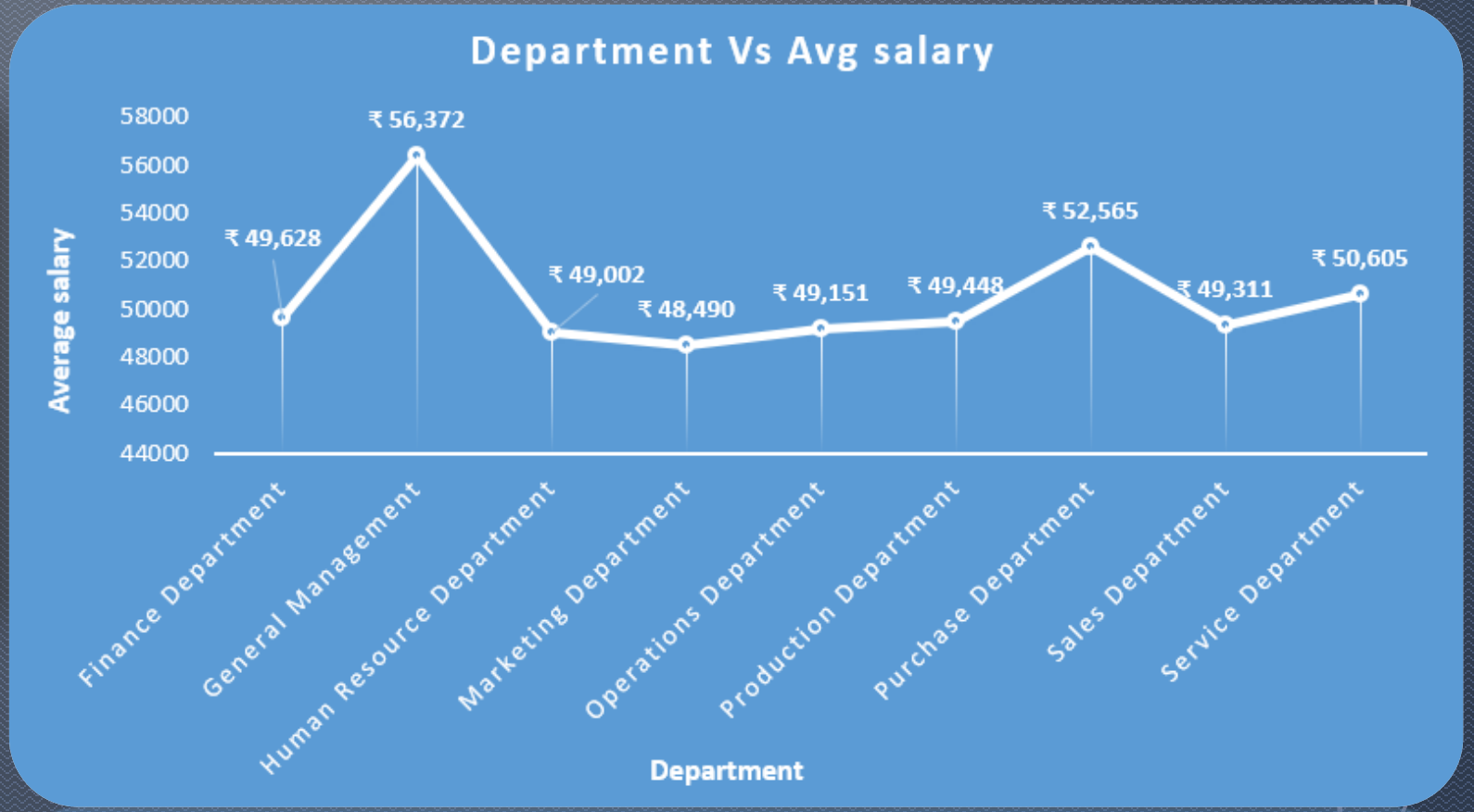


Data Analysis

Task 2 : Salary Analysis

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

Row Labels	Average Offered Salary
Finance Department	49628
General Management	56372
Human Resource Department	49002
Marketing Department	48490
Operations Department	49151
Production Department	49448
Purchase Department	52565
Sales Department	49311
Service Department	50605
Grand Total	**49919**



Data Analysis

Task 3 : Salary Distribution

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

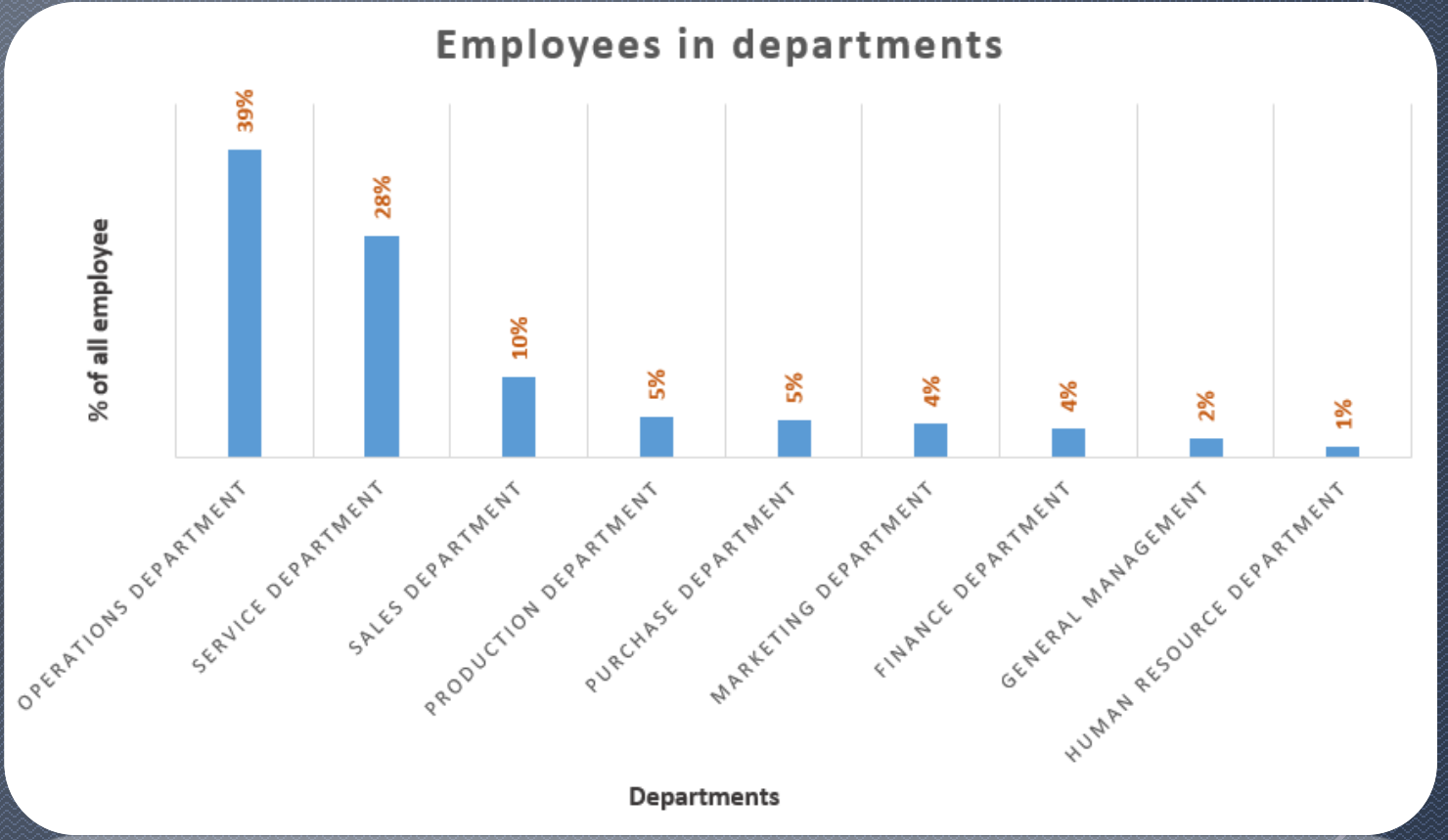
Status	Hired
Salary	**Count of employees**
100-10100	444
10100-20100	487
20100-30100	457
30100-40100	488
40100-50100	523
50100-60100	496
60100-70100	450
70100-80100	479
80100-90100	462
90100-100100	408
140100-150100	3
Grand Total	**4697**

Data Analysis

Task 4 : Departmental Analysis

Show the proportion of people working in different departments

Status	Hired	
Row Labels	**Count of working empl.**	**% of working empl.**
Operations Department	1843	39%
Service Department	1332	28%
Sales Department	485	10%
Production Department	246	5%
Purchase Department	230	5%
Marketing Department	202	4%
Finance Department	176	4%
General Management	113	2%
Human Resource Department	70	1%
Grand Total	**4697**	**100%**

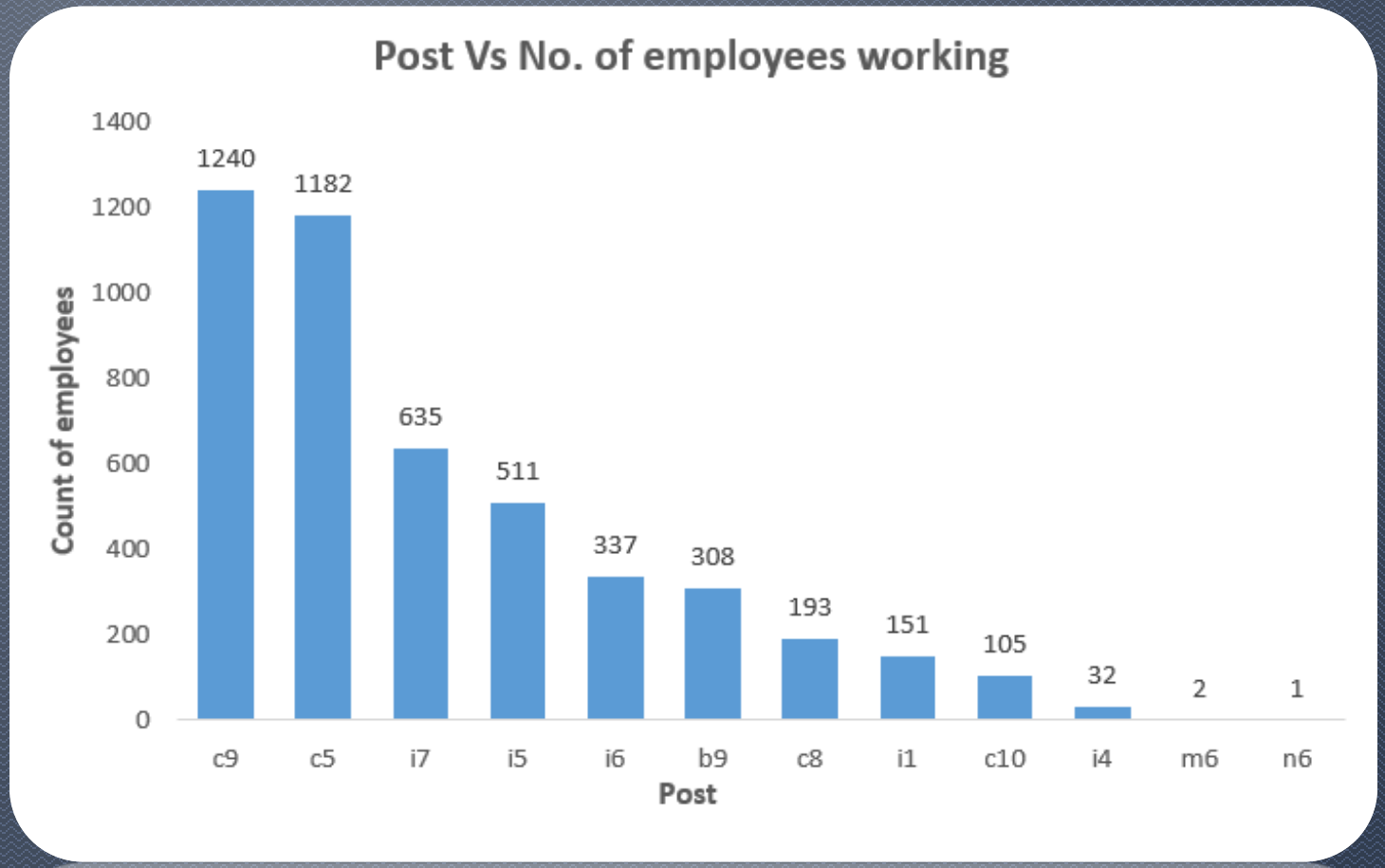


Data Analysis

Task 5 : Position Tier Analysis

Represent the different position tiers within the company.

Status	Hired
Post name	**Count of employees**
c9	1240
c5	1182
i7	635
i5	511
i6	337
b9	308
c8	193
i1	151
c10	105
i4	32
m6	2
n6	1
Grand Total	**4697**



Insights :

- Out of 4697 employees, 55 percent of employees are Male and 39 percent are Female.
- General management department has highest average offered salary (₹56,372) and Marketing department has lowest average offered salary (₹48,490) .
- Most of the employees are getting paid in range from ₹40100 – ₹50100.
- Out of 4697 employees, 39 percent of employees are working in operations department and 28 percent in service department.
- Out of 4697 employees, 1240 employees are working under c9 post name and 1182 employees are working under c5 post name.

Results :

This project helped me to advance Excel skills and problem solving ability. Through this project I learned how to handle missing data and outliers in data based on situation. Extensively worked on pivot table and charts which enabled me to give better representation of output in form of charts.

THANK YOU