Project 4

Hiring Process Analytics

Project Description: The hiring process is a crucial function of any company, and understanding trends such as the number of rejections, interviews, job types, and vacancies can provide valuable insights for the hiring department. This project aims to analyze the company's hiring process data and draw meaningful insights from it.

Approach: This project is executed using excel. It involves identifying the relation between the columns of the data from the excel worksheet, to analyze the insights from that data.

Tech-Stack Used: The tech-stack used for this project is EXCEL 2019 version.

Insights:

Case 1: Data cleaning

1) **Handling Missing Data:**

Descriptive Statistics of Offered Salary is analysed.

Descriptive Statistics of Offered Salary	
Mean	49983.02902
Standard Error	340.8317054
Median	49625
Mode	72843
Standard Deviation	28854.17689
Sample Variance	832563524
Kurtosis	2.610052003
Skewness	0.361578537
Range	399900
Minimum	100
Maximum	400000
Sum	358228369
Count	7167
Count of application_id	7168
Count of Offered Salary	7167
Missing row in Offered Salary	1

Mean, standard error, median, mode, standard deviation, sample variance, kurtosis, skewness, range, minimum value, maximum value, sum of values and count of data is calculated. It was found that there were 7167 rows of offered salary but the application_id had 7168 rows there was a missing value in offered salary. Therefore, that row was deleted as it was affecting the calculation of Descriptive Statistics of the data.

2) **Outlier Detection:**

First quartile	25452.75
Third quartile	74396.75
Interquartile range (IQR)	48944
Upper bound	147812.8
Lower bound	-47963.3
Outliers Found	3
Outlier values found	200000
Outlier values found	300000
Outlier values found	400000

The outliers were identified by calculating upper bound and lower bound of box plot.

The first quartile is calculated using the function =QUARTILE(G2:G7165,1),

Third quartile is calculated using the function =QUARTILE(G2:G7165,3),

Interquartile range (IQR) is calculated using the function =K4-K3, by subtracting the third quartile from first quartile.

The upper bound is calculated using the function =K4+(1.5*K5), and

The lower bound is calculated using the function =K3-(1.5*K5).

Three outliers were identified.

3) **Removing Outliers:**

200000, 300000 and 400000 are the three outliers identified from the data.

The three outliers were deleted from the data, as they affected in identifying the insights from the data.

4) Data Summary:

After cleaning the data, the descriptive statistics of offered salary is

Descriptive Statistics of Offered Salary	
Mean	49878.3318
Standard Error	334.989477
Median	49614.5
Mode	72843
Standard Deviation	28353.6486
Sample Variance	803929391
Kurtosis	-1.1794931
Skewness	0.01317731
Range	99867
Minimum	100
Maximum	99967
Sum	357328369
Count	7164
Count of application_id	7164
Count of Offered Salary	7164
Outliers Rows removed	3
Blank Rows removed	1
Total Rows removed	4

Case 2: Data Analytics Tasks

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

Hired Males	2562
Hired Females	1854

The number of male candidates hired by the company is calculated using the function **=COUNTIFS(Clean_Data_Sheet!C2:C7165,"Hired",Clean_Data_Sheet!D2:D7165,"Male")**

The number of male candidates hired by the company is calculated using the function **=COUNTIFS(Clean_Data_Sheet!C2:C7165,"Hired",Clean_Data_Sheet!D2:D7165,"Female")**

The company has hired 2562 male candidates and 1854 female candidates for different roles.

2) 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.

Average salary	49878.3318

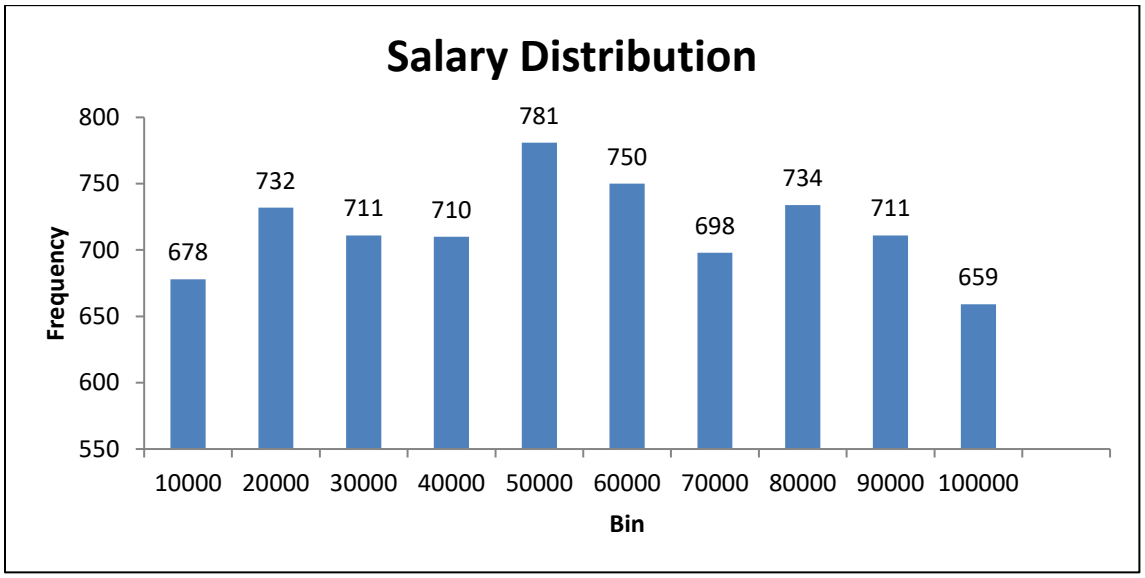
The average salary of the employees is calculated using the function **=AVERAGE(Clean_Data_Sheet!G2:G7165)**

The average salary of the employees in the organization is 49878.3318.

3) 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.

Maximum Offered Salary	99967
Minimum Offered Salary	100
Width	99867
Approximate width	10000

Bin	*Frequency*
10000	678
20000	732
30000	711
40000	710
50000	781
60000	750
70000	698
80000	734
90000	711
100000	659



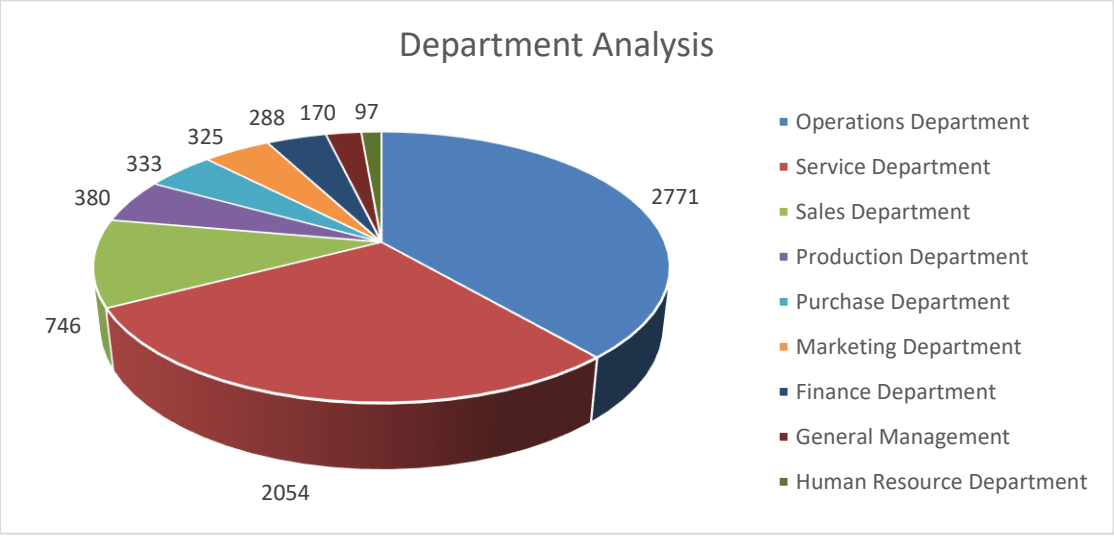
The above images show the salary distribution by creating class intervals for the salaries in the company. By approximating the salary width to 10000, 10 bins are created to analyze the salary distribution.

The histogram shows that the salary bin 40000 to 50000 has the largest count of employees of 781 employees and salary bin 90000 to 100000 has the least count of employees of 659 employees.

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

Departments	Employees
Operations Department	2771
Service Department	2054
Sales Department	746
Production Department	380
Purchase Department	333
Marketing Department	325
Finance Department	288
General Management	170
Human Resource Department	97
Grand Total	**7164**

The above image shows the distribution of employees through various departments. The operations department has the largest count of 2771 employees and the human resource department has the least count of 97 employees working in the organization.

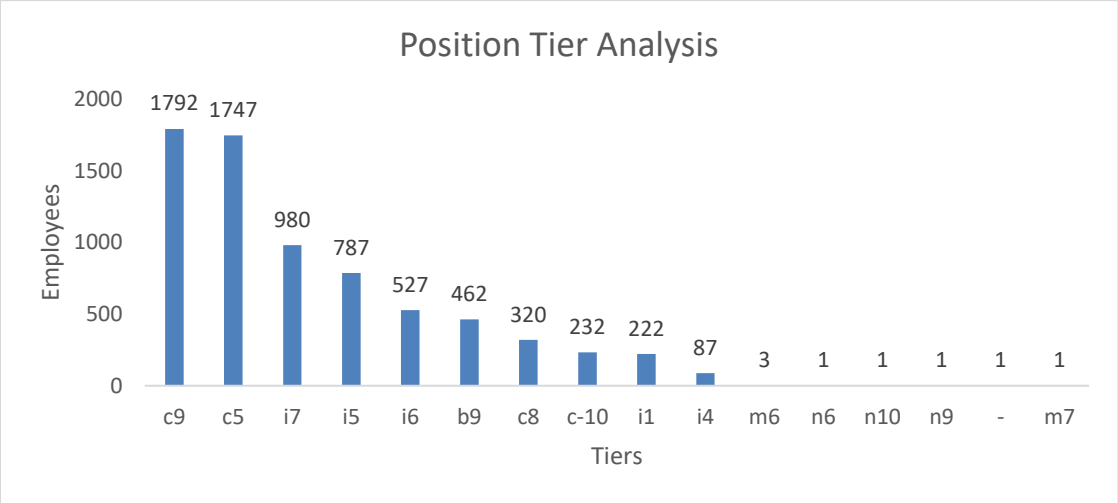


The above image shows the visualization of the distribution of employees in the organization in various department using pie chart.

5) Position Tier Analysis: Different positions within a company often have different tiers or levels.

Tiers	Employees
c9	1792
c5	1747
i7	980
i5	787
i6	527
b9	462
c8	320
c-10	232
i1	222
i4	87
m6	3
n6	1
n10	1
n9	1
-	1
m7	1
Grand Total	**7164**

The above image shows the distribution employees in various tiers. The c9 tier has the largest count of employees of 1792 employees and the n6, n10, n9 and m7 tiers has the least count of employee of 1 employee in the organization.



The above image shows the visualization of the distribution of employees in the organization in various tiers using cluster column chart.

Excel drive link:

The excel sheet has been saved and uploaded to google drive. To access the file following link can be used: https://docs.google.com/spreadsheets/d/1RqEVct_zma-PwxOZQc31halsuYlWjA8v/edit?usp=sharing&ouid=116970069599597283204&rtpof=true&sd=true.