Trainity Data Analytics Training

Project 4

Hiring Process Analytics

Date: 7/12/24 **Name-Sudipta Samanta**

Project Description:

The project involves analyzing a hiring process dataset to derive insights that can help the company improve its hiring practices. Tasks include handling missing data, detecting and managing outliers, performing statistical analyses, and creating visualizations for better understanding. The findings will be compiled into a report to present to stakeholders.

Approach:

1. Handling Missing Data

Checked for missing values in the dataset.

Execution:

Used Excel's Filter or Conditional Formatting to identify blank cells.

Applied appropriate strategies:

Removed rows with critical missing data if their absence doesn’t skew analysis.

Replaced missing values using averages/medians for numerical data or the most common value for categorical data.

2. Clubbing Columns

Simplified categories by combining them where applicable.

Execution:

Analyzed categorical columns with multiple categories (e.g., "Job Type").

Used Excel's Find & Replace or formulas to merge similar categories.

3. Outlier Detection and Management

Detection and handled outliers.

Execution:

Used the Boxplot feature in Excel or calculate IQR (Interquartile Range).

Removed or adjusted outliers based on their impact on analysis.

4. Hiring Analysis

Determining gender distribution.

Execution:

Used COUNTIF function or a Pivot Table to calculate the number of males and females.

5. Salary Analysis

Calculating the average salary.

Execution:

Used AVERAGE function in Excel

6. Salary Distribution

Created class intervals for salary data.

Execution:

Used Excel's Data Analysis Toolpak or create bins manually.

Visualized using a Histogram.

7. Departmental Analysis

Visualized the proportion of employees in each department.

Execution:

Created a Pie Chart or Bar Graph using the department data.

8. Position Tier Analysis

Analyzed distribution across different position tiers.

Execution:

Used a Bar Chart or Stacked Column Chart.

Tech-Stack Used:

Software: Microsoft Excel 2022

Purpose: Data cleaning, statistical analysis, and visualization.

Insights:

1. Gender Distribution: Analyzed the male-to-female hiring ratio to identify any hiring biases.

2. Salary Trends: Determined the average salary and visualized its distribution, which could help in setting competitive salaries.

3. Departmental Insights: Highlighted which departments have the highest and lowest hiring rates, aiding in resource allocation.

4. Position Tier Insights: Analyzed tier distributions to optimize the organizational structure

Results:

A. **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 females have been hired by the company?

Query: (In Excel)

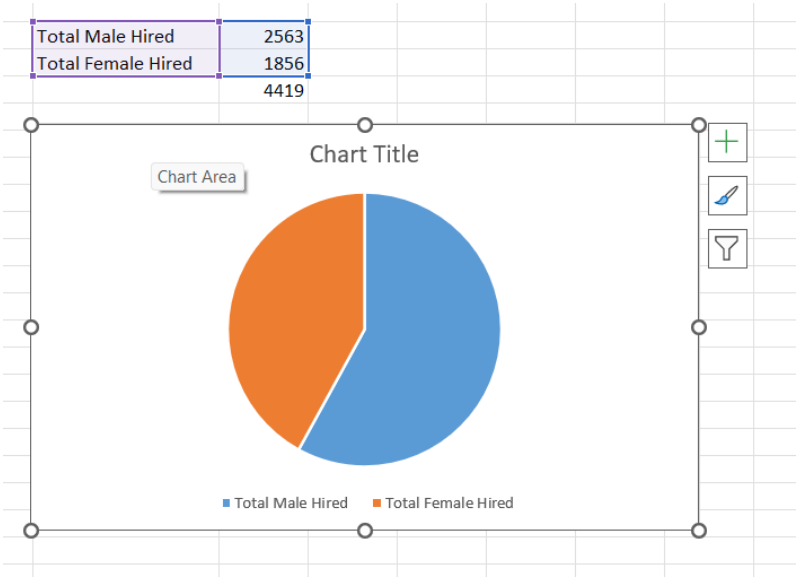
For Males

=COUNTIFS(Sheet1!D:D,"Male",Sheet1!C:C,"Hired")

For Females

=COUNTIFS(Sheet1!D:D,"Female",Sheet1!C:C,"Hired")

Output:



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.

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

Query: (In Excel)

=AVERAGE(SUMIFS(Sheet1!G:G,Sheet1!C:C,"Hired")/COUNTIF(Sheet1!C:C,"Hired"))

Output:

D
Avg salary of employees
49752.90

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.

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

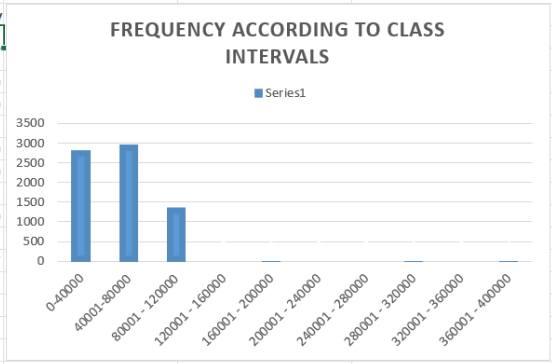
Query: (In Excel)

Class Intervals
0-40000 =COUNTIFS(Sheet1!G:G,">=0",Sheet1!G:G,"<=40000")
40001-80000
=COUNTIFS(Sheet1!G:G,">=40001",Sheet1!G:G,"<=80000")
80001 – 120000
=COUNTIFS(Sheet1!G:G,">=80001",Sheet1!G:G,"<=120000")
120001 – 160000
=COUNTIFS(Sheet1!G:G,">120001",Sheet1!G:G,"<=160000")
160001 – 200000
=COUNTIFS(Sheet1!G:G,">=160001",Sheet1!G:G,"<=200000")
200001 – 240000
=COUNTIFS(Sheet1!G:G,">=200001",Sheet1!G:G,"<=240000")
240001 – 280000
=COUNTIFS(Sheet1!G:G,">=240001",Sheet1!G:G,"<=280000")
280001 – 320000
=COUNTIFS(Sheet1!G:G,">=280001",Sheet1!G:G,"<=320000")

320001 – 360000
=COUNTIFS(Sheet1!G:G,">=320001",Sheet1!G:G,"<=360000")
360001 – 400000
=COUNTIFS(Sheet1!G:G,">=360001",Sheet1!G:G,"<=400000")

Output:

Class Intervals	Frequency
0-40000	2831
40001-80000	2963
80001 - 120000	1370
120001 - 160000	0
160001 - 200000	1
200001 - 240000	0
240001 - 280000	0
280001 - 320000	1
320001 - 360000	0
360001 - 400000	1
	7167



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

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

Query: (In Excel)

Proportion of people working in different deptartments
Finance Dept =COUNTIFS(Sheet1!E:E,"Finance Department",Sheet1!C:C,"Hired")

General Management
=COUNTIFS(Sheet1!E:E,"General Management",Sheet1!C:C,"Hired")

Human Resource Dept
=COUNTIFS(Sheet1!E:E,"Human Resource Department",Sheet1!C:C,"Hired")

Marketing Dept =COUNTIFS(Sheet1!E:E,"Marketing Department",Sheet1!C:C,"Hired")

Operations dept =COUNTIFS(Sheet1!E:E,"Operations Department",Sheet1!C:C,"Hired")

Production Dept =COUNTIFS(Sheet1!E:E,"Production Department",Sheet1!C:C,"Hired")

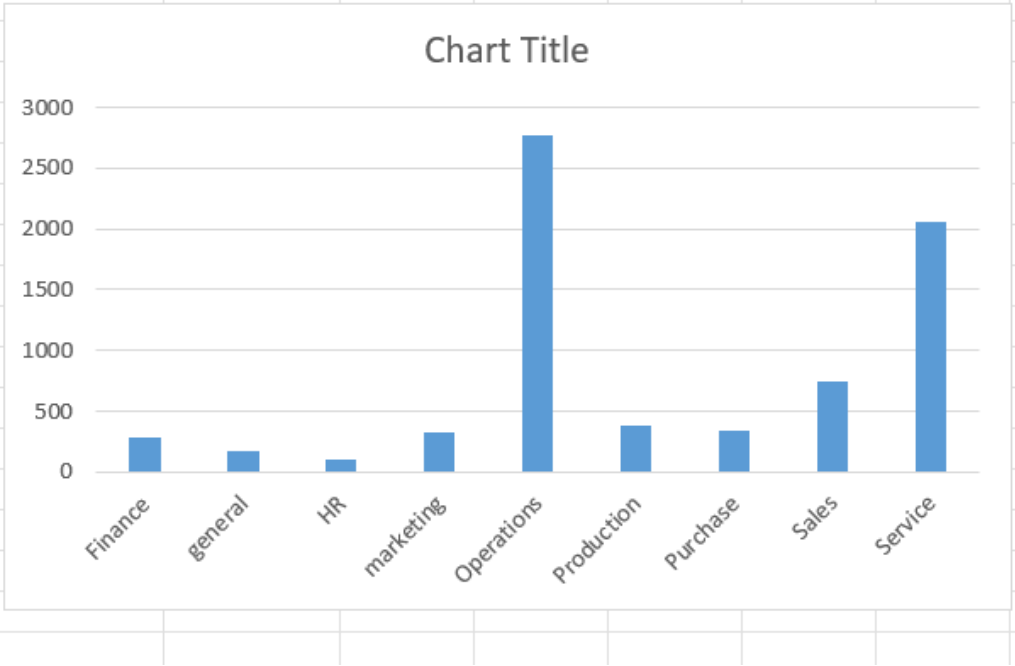
Purchase Dept =COUNTIFS(Sheet1!E:E,"Purchase Department",Sheet1!C:C,"Hired")

Sales Dept =COUNTIFS(Sheet1!E:E,"Sales Department",Sheet1!C:C,"Hired")

Service Dept =COUNTIFS(Sheet1!E:E,"Service Department",Sheet1!C:C,"Hired")

Output:

Finance	288
general	172
HR	97
marketing	325
Operations	2770
Production	380
Purchase	333
Sales	745
Service	2052



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

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.

Query: (In Excel)

b9
=COUNTIF(Sheet1!F:F,"b9")

c-10
=COUNTIF(Sheet1!F:F,"c-
10")

c5
=COUNTIF(Sheet1!F:F,"c5")

c8
=COUNTIF(Sheet1!F:F,"c8")

c9
=COUNTIF(Sheet1!F:F,"c9")

i1
=COUNTIF(Sheet1!F:F,"i1")

i4
=COUNTIF(Sheet1!F:F,"i4")

i5
=COUNTIF(Sheet1!F:F,"i5")

i6
=COUNTIF(Sheet1!F:F,"i6")

i7
=COUNTIF(Sheet1!F:F,"i7")

m6
=COUNTIF(Sheet1!F:F,"m6")

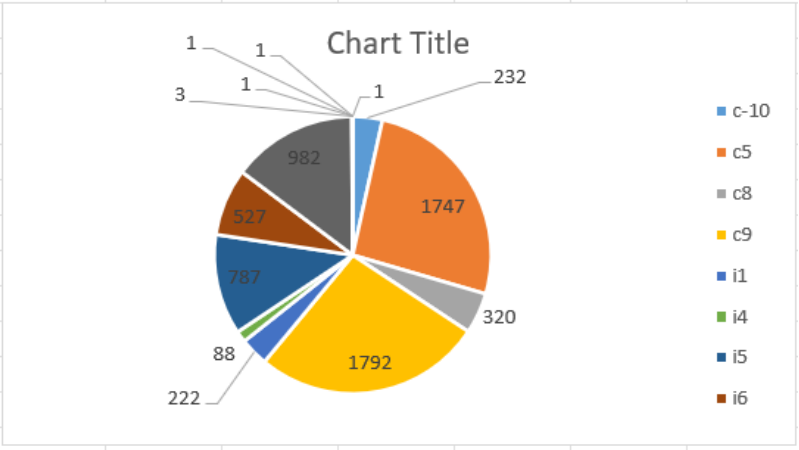
n10
=COUNTIF(Sheet1!F:F,"n10")

n6
=COUNTIF(Sheet1!F:F,"n6")

n9
=COUNTIF(Sheet1!F:F,"n9")

Output:

b9	463
c-10	232
c5	1747
c8	320
c9	1792
i1	222
i4	88
i5	787
i6	527
i7	982
m6	3
m7	1
n10	1
n6	1
n9	1



Access Detailed Insights via This Link (Hyperlink)

..\Downloads\Statistics.xlsxExcel Sheet