Trainity Assignment-4

Muthiah Sivavelan
Ph:8525021258

Excel document: Hiring Process Analysis

Project Description:

Imagine you're a data analyst at a multinational company like Google. Your task is to analyze the company's hiring process data and draw meaningful insights from it. 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.

As a data analyst, you'll be given a dataset containing records of previous hires. Your job is to analyze this data and answer certain questions that can help the company improve its hiring process.

Here's what you'll be doing:

1. Handling Missing Data: Check if there are any missing values in the dataset. If there are, decide on the best strategy to handle them.

2. Clubbing Columns: If there are columns with multiple categories that can be combined, do so to simplify your analysis.

3. Outlier Detection: Check for outliers in the dataset that may skew your analysis.

4. Removing Outliers: Decide on the best strategy to handle outliers. This could be removing them, replacing them, or leaving them as is, depending on the situation.

5. Data Summary: After cleaning and preparing your data, summarize your findings. This could involve calculating averages, medians, or other statistical measures. It could also involve creating visualizations to better understand the data.

The objective of this project is to use my knowledge of statistics and Excel to draw meaningful conclusions about the company's hiring process which could potentially help the company improve its hiring process and make better hiring decisions in the future.

Approach:
First, I found out that the missing values where stored as – and blanks in the given dataset. I replaced all the – with blanks and then checked the number of rows which has the blank values and since the number of rows are very less in comparison to the total number of rows, I decided to delete the rows which contain blank values. I have used box and whisker plot to find out that the salaries which are greater than 1 lakh are outliers and then I decided to remove the outliers(3 rows). I didn’t remove any rare occurrences in post Name column because there could be positions which are very less in the company like CEO which is not an outlier. Then, based on the question given I used various formulas(such as sumif, countif, etc.) and charts to get the solution for each problem.

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

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

Approach:

I used the countifs command to select the event_name column which contains the gender value and then provided the criteria as male and female respectively in separate cells. Along with this, I have also added the criteria where the status should be Hired.

Query:

=COUNTIFS(D:D,"Male",C:C,"Hired")
=COUNTIFS(D:D,"Female",C:C,"Hired")

Output:

# Qn		
1	**Hired**	
	Males	2561
	Females	1854

B) Salary Analysis:

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

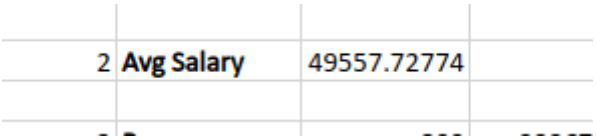
Approach:

The idea is to sum all the salaries where the status is Hired and then divide by the count of all rows where the status is Hired. This will give the average salary.

Query:

=SUMIF(C:C,"Hired",G2:G7149)/COUNTIF(C:C,"Hired")

Output:



C) Salary Distribution:

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

Approach:

First, I found out the range of the salary values using the min and max functions. Then, I found out their difference. Then, I decided to take the number of intervals as 10. I found out the interval width by dividing the Range and number of intervals. I know that the starting value of interval is the minimum salary value(in the first row). The starting value of the an interval is the sum of starting value of previous interval and the interval width. I used this formula to get the 10 rows(for 10 intervals). For the ending values, I used the formula starting value+interval width -1.

Thereby, I found the 10 intervals. To find the count in the interval, I used the countifs function and selected the salary column and used the criteria as “>=”& starting value. Similarly I added the less than condition for ending value.

Output:

Start Interval	End Interval	Count	Cumulative Count		
800	10716	480	480	0	
10717	20633	465	945	0	
20634	30550	458	1403	0	
30551	40467	486	1889	0	
40468	50384	517	2406	1	Median Interval
50385	60301	497	2903	1	
60302	70218	439	3342	1	
70219	80135	480	3822	1	
80136	90052	451	4273	1	
90053	99969	410	4683	1	

D) Departmental Analysis:

Objective: Visualizing data through charts and plots is a crucial part of data analysis. Use a pie chart, bar graph, or any other suitable visualization to show the proportion of people working in different departments.

Approach:

I added a second sheet and copied the department column. Then I removed the duplicates using the data tab. I added a new column named Number of people and for each row in this column, I used the countifs function to count the number of rows in department column of the first spreadsheet which is equal to the distinct department value in the current spreadsheet. I also added the condition that the status should be Hired to include only the employees of the company. Then, I selected the two columns and clicked on pie chart and customized it a bit to look pleasing.

Query:

=COUNTIFS(Sheet1!E3:Sheet1!E7153,A5,Sheet1!C3:Sheet1!C7153,"Hired")

Output:

QN #4	
Department	**Number of people**
Service Department	1326
Operations Department	1840
Sales Department	483
Finance Department	176
Production Department	246
Purchase Department	230
Marketing Department	201
General Management	111
Human Resource Departmer	70

E) Position Tier Analysis:

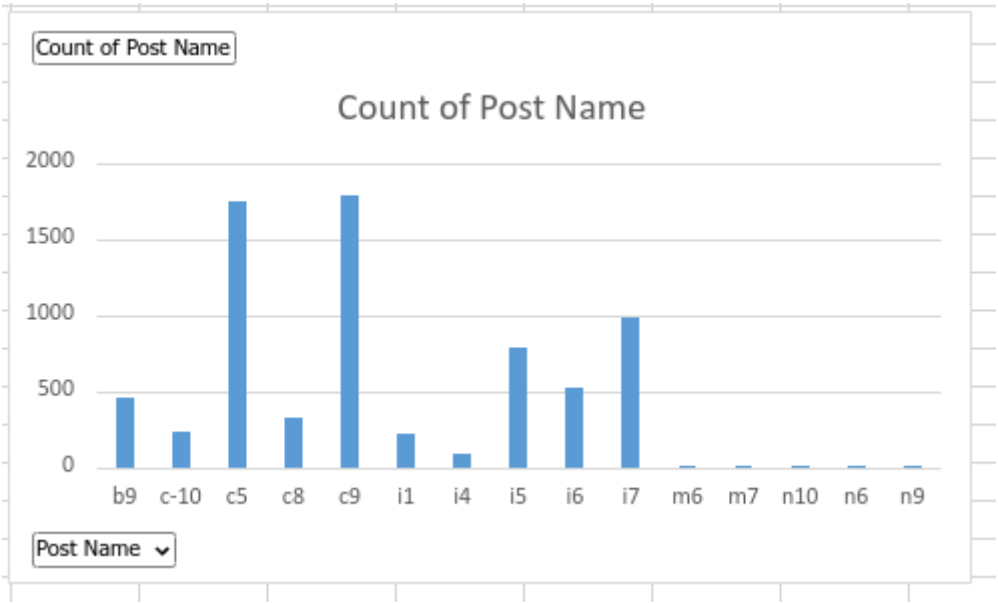
Objective: Different positions within a company often have different tiers or levels. Use a chart or graph to represent the different position tiers within the company. This will help you understand the distribution of positions across different tiers.

Approach:

I used the clustered bar chart from the recommended charts option after selecting the Post Name column. This opened a new sheet which contained the Post Name column values along with its count. I can also follow the procedure similar to the previous question but this saved me a bit of time. I got the bar chart for each Post Name showing its count.

Output:

Qn #5	
Post Name	Count of Post Name
b9	461
c-10	231
c5	1742
c8	319
c9	1790
i1	220
i4	87
i5	785
i6	527
i7	979
m6	3
m7	1
n10	1
n6	1
n9	1



Tech Stack Used:
I have used Microsoft Excel 2019, since that is the one that came preinstalled in my laptop. When some functions are missing I would upload the file in OneDrive and open the file using microsoft365.com to execute certain functions.

Insights:

- Males were hired in more numbers compared to females among those people who revealed their gender.
- The average salary of the employees of the company is 49558.

- The median salary lies in the range 40468 and 50384.
- The company hired the most number of people for operations department followed by service department.
- The post names which are least hired(<10) are m6, m7, n10, n6, n9 and depending on the type of tasks, these posts can be considered outliers or not. m6 has count of 3 while all others which are mentioned aboved has the count=1.

Result:

From this project, I have learnt to use my knowledge of excel and statistics for data analysis to form conclusions that could potentially improve the hiring process of the company and make better hiring decisions in the future.