

Northeastern University

EECE5642: Data Visualization Midterm

**Rudra Patel
Feb 22, 2022**

Answer 1 (Graph Critique)

Part A.

In this question we need to list all the problems associated with the graph. The graph shows the sales revenue for 12 products across four quarters in the state of Kansas for a particular year.

There are several problems which makes this a poor visualization work which are discussed as follows:

1. The first thing that comes to the mind is about the 3D nature of the graph. It makes it very hard to read and interpret the data for such a dataset.
2. Here a line chart is used which is usually used to plot a time series data with time instances on x-axis to show trend between a current, previous and next instance of time (years, months, days of week etc.). However here the lines are used connect values which have no direct connections.
3. The 3D values which are encoded in graph have no values and causes distraction to the viewer.
4. There are too many tick marks and label on the x and y axis. Like on the vertical axis anything beyond 800,000 is never used and can be eliminated.
5. The graph is poorly labelled with no indication of title or currency type or year for which this data is plotted.
6. The grid lines for each axis are unnecessary and add to decreasing data-ink ratio.
7. Decimals are not needed while indicating revenue on y-axis.
8. The labels for the 12 different products are difficult to read easily at first glance.
9. There is a poor data-ink ratio because of the unnecessary color used on the x-axis and y-axis planes respectively.
10. There is no strong reason to highlight the labels for Q1, Q2, Q3 and Q4 in bold.

Overall, all the above stated reasons make this a weak chart. It is quite difficult to show all the categories of products for each quarter (Q1, Q2, Q3 and Q4) on a single plot.

Part B.

In this part we need to provide a solution / suggestion to each of the problem listed above.

Here is the suggestion for each of the 10 problems in the same order as listed in the previous part.

1. A 3D plot can easily be eliminated here. We can use a simple bar plot or a line chart which shows the variation in each quarter.
2. A suggestion here would be we can plot quarters Q1, Q2, Q3 and Q4 on the x-axis if we want to use a line chart to indicate some changing trends in the figure. We can also make use of bar plots or scatter chart as we will see in further discussion.
3. The 3D bars are not adding any value to this visualization, so we strictly need to eliminate it since we are not able to estimate values in this case like for the Q2 or Q3 value for the second product is very poorly shown. We need to show is more discretely.
4. We can set a limit on the y-axis range to limit it to 800,000 as there is not value going beyond it to reduce the number of ticks and labels.
5. A title can be put in the chart, so it is easy for a reader without a dataset to visualize what they're seeing. Also, we can indicate any arbitrary year and a type of currency.
6. We can remove unnecessary grid lines in order to improve and maximize the data-ink ratio.
7. Decimals are not adding any value so when defining the labels, we can eliminate them and simply use whole numbers.
8. We can try to accommodate the x-axis labels horizontally or at a rotated angle making them much easier to read as compared to vertical labels.
9. The background colors on the planes are not needed and can be removed to depict plane white background.
10. There is no need to emphasize the labels showing different quarters as they not adding any prominent value to the chart.

A Note here would be even after taking care of all the problems using the above listed suggestions the chart would still be a little cluttered because of the presence of too many categories for each of the four quarters.

Part C.

For this last section of the question, I have plotted five different plots each with its own advantages.

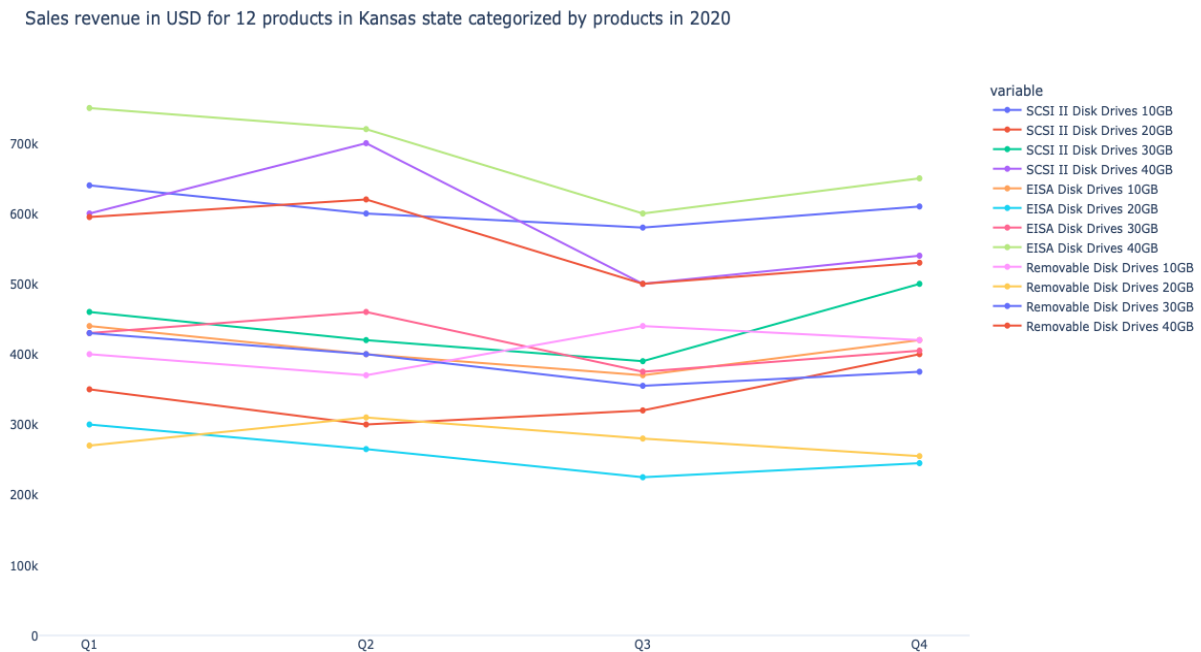


Figure1. Plot-ly All Data

- This plot is generated using the plotly library in python.
- Here we have the 4 quarters on the x-axis which depicts a time series of the data for all the 12 products.
- We see that here since there are high number of variables the plot looks a little clustered.
- A nice feature we can extract by using plotly is that we have an advantage of isolating all those data that we wish not to review. This can be done by clicking on that variable in the legend shown in the upper right corner.
- In the next chart we will see how the plot changes when we select only a few selected variables.

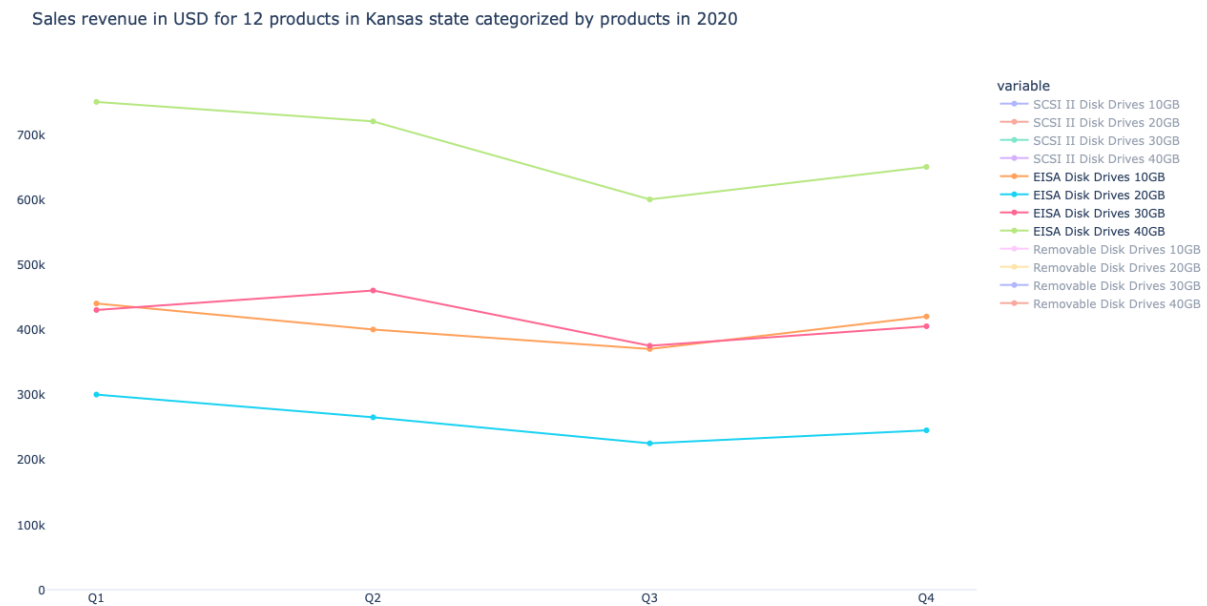


Figure 2. Plot-ly Selected Data

- As discussed above here we have selected only one group of disk drives. This removes all the other lines from the chart and aid us to focus on the data that is being currently reviewed.
- The variables that have been isolated have been blurred out in the legend as seen on upper right corner in the plot.
- We also have a feature wherein if we double click a particular variable, we will see only that time series thereby diverting the focus of the viewer on that series.
- The disadvantage in this type of chart is that we are not able to simulate it when we add in on a report or take a print of the graph.

Next, I have added a few different charts that are useful and can be used to plot the given data set.

Note: The values are assumed here since it is not possible to know the exact value from the given plot which is itself a disadvantage. Also, it is assumed that this data is from year 2020 just to be clear with the title.

Line and scatter chart

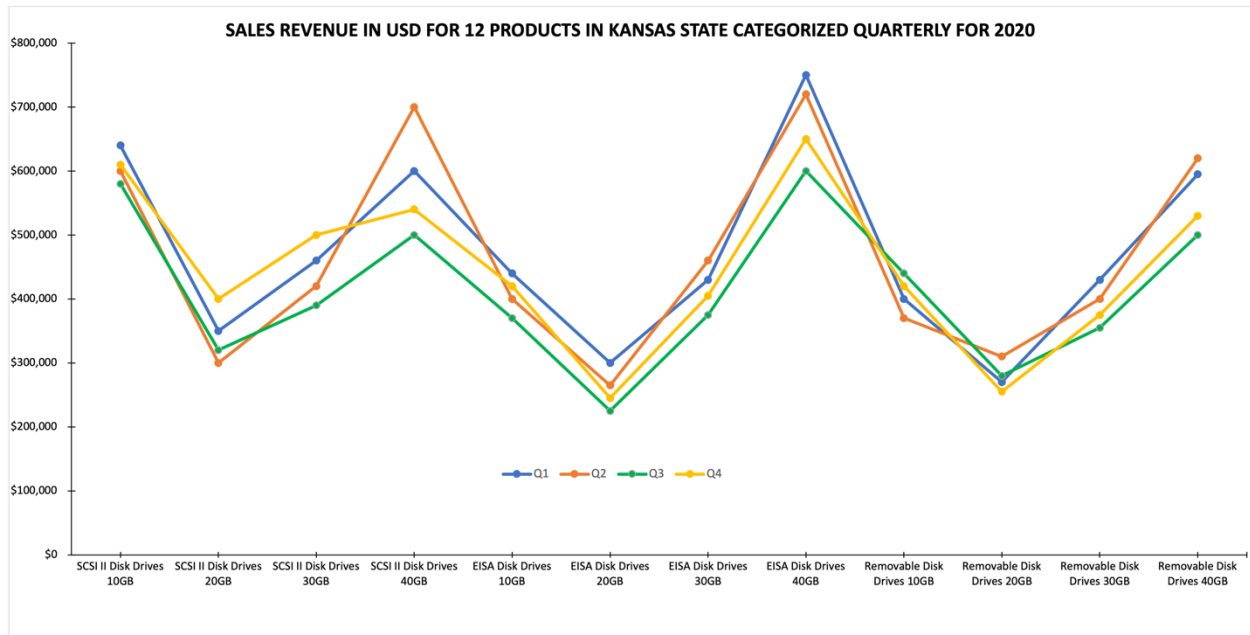


Figure 3. Line Chart categorized quarterly

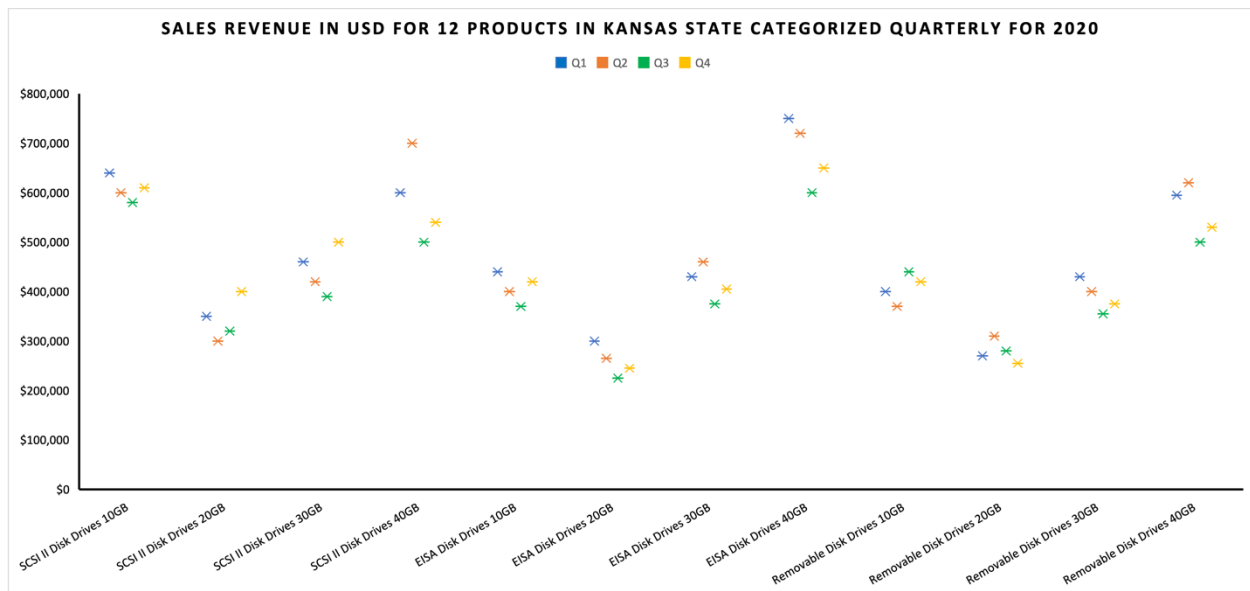


Figure 4. Scatter Plot categorized quarterly

- Both the above charts are categorized according to the quarterly data. On the x-axis we have listed the 12 different products. In the scatter plot we get rid of the lines to improve the data-ink ratio.

- Labels on the scatter chart are rotated to better fit in the x-axis and at the same time easy to read.
- With the scatter chart we can study a product more efficiently since we got rid of lines which sometimes comes in way when we need to analyze the data. It helps us understand how the trend is changing (increasing or decreasing) quarter after quarter for each product.

Multiple Bar chart and Stacked Bar chart

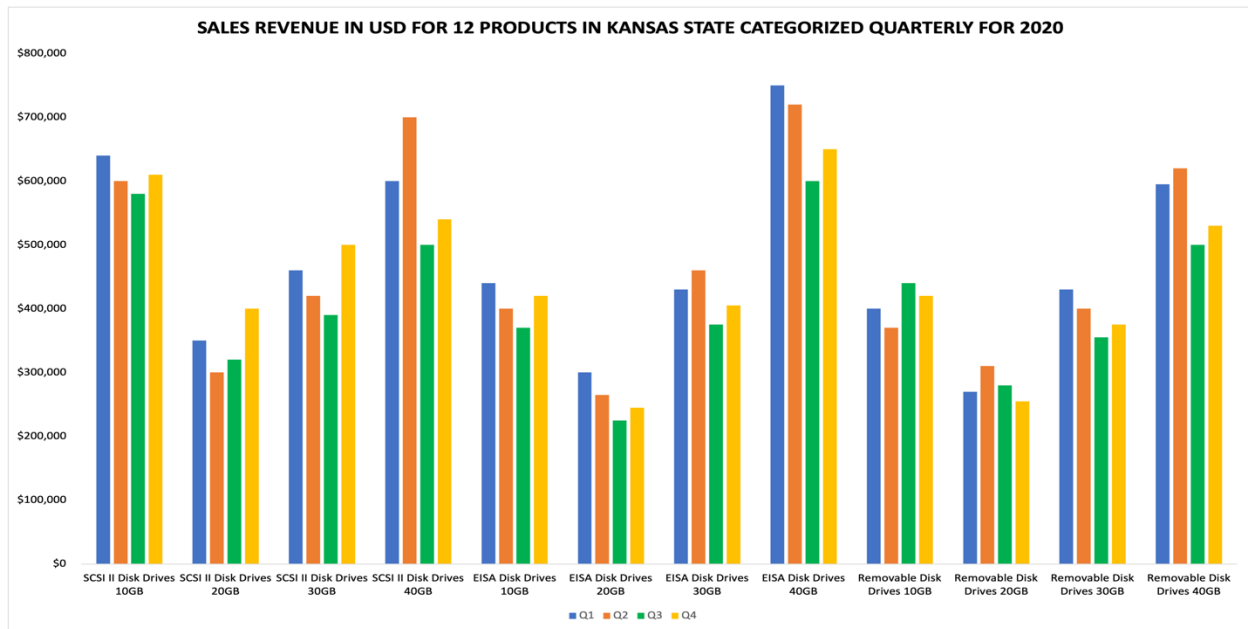


Figure 5. Multiple bar chart categorized quarterly

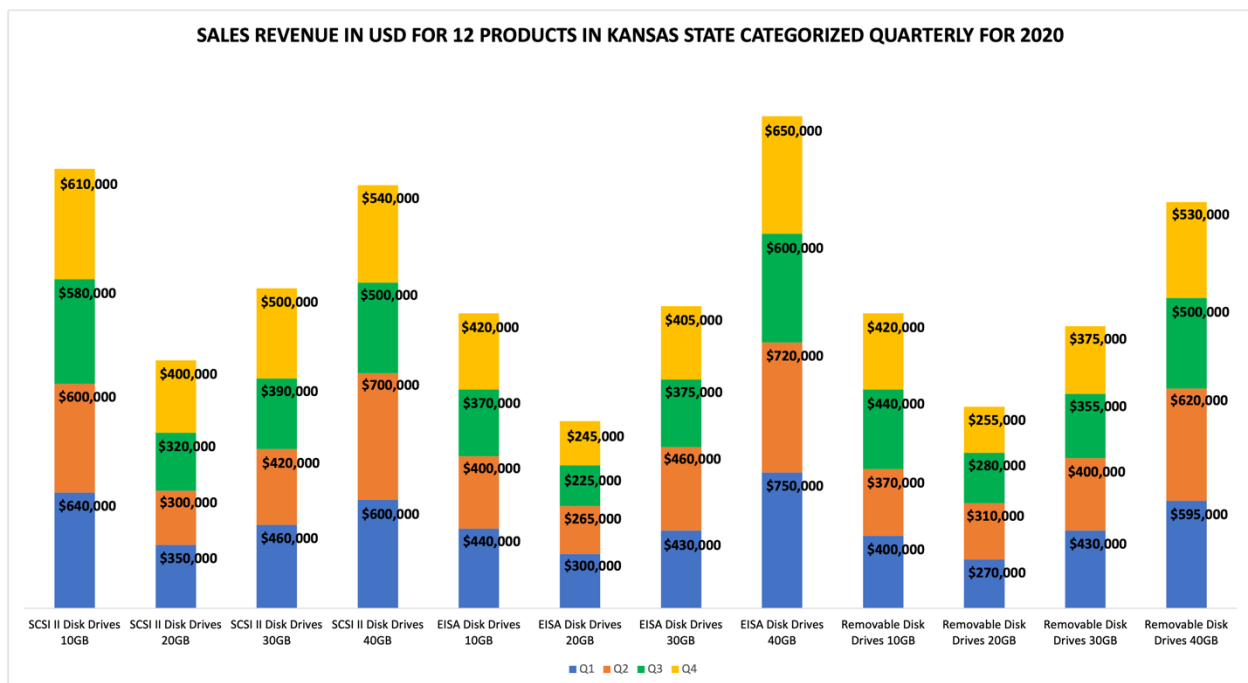


Figure 6. Stacked bar chart categorized quarterly

- Both above charts are special types of Bar plots.
- In each of them we have listed the different products on the x-axis. These are again categorized quarterly using values for each quarter of all the products.
- In the multiple bar chart, we have the sales revenue on the y-axis. This helps us in communicating well and interpret each product to its sales revenue easily and more efficiently as compared to line and scatter plots.
- In the second stacked bar chart I have gotten rid of the y-axis and instead added data values on the stacked chart. This is done in order to help the user know the exact value in each quarter for all the products which otherwise would have need to be calculated in a traditional stacked bar chart without data labels.
- The total height of the bar in stacked bar chart also helps us in determining which product has the maximum sales revenue and which has the least.

The complexity of the data set is quite more since it is difficult to plot all the 12 products for each quarter on a single chart. However, the best of an attempt has been made here in order to communicate with the viewer easily.

Answer 2 (Visualization Design)

In this question we are told that the VP is only interested in degree to which the actual expenses deviated from the budget and not the actual dollar amount. For this type of deviation elements, the best possible graph here would be a bar chart or a line graph with defined baseline showing the deviation of values. I would be discussing two major cases and two additional cases while answering the design criteria and advantage for each of them.

Case 1.

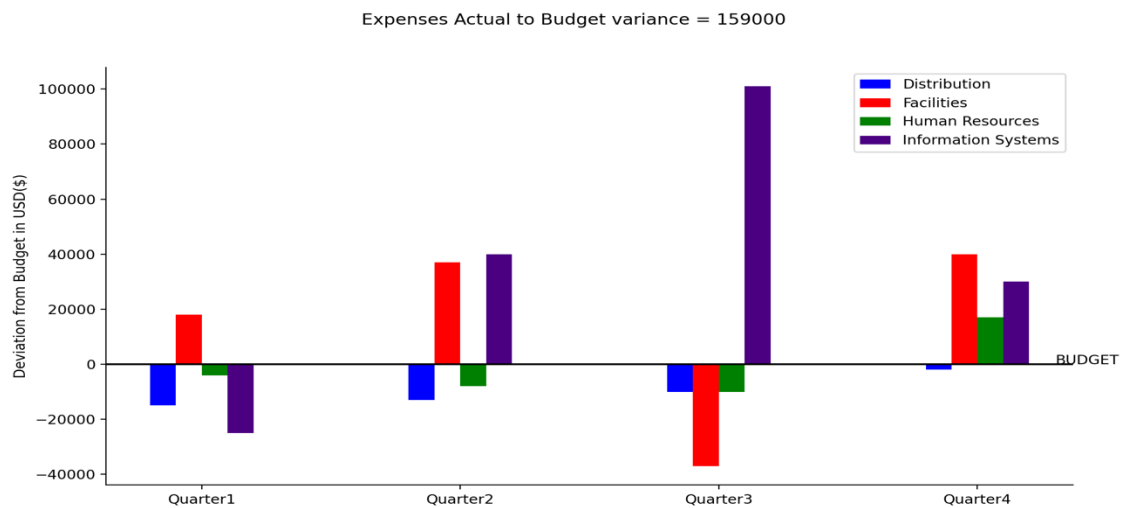


Figure 7. Variance Deviation from proposed budget

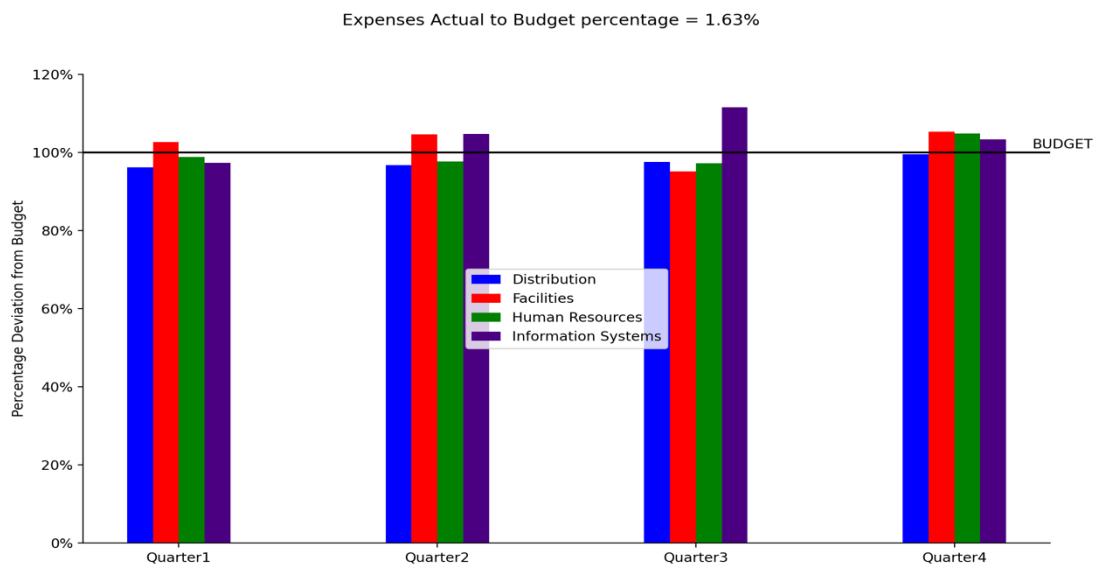


Figure 8. Percentage Deviation from proposed budget

The design criteria used for the above plots is as follows:

- For each case the value for (Actual – Budget) has been calculated which gives us the deviation while choosing budget as the baseline.
- In the first case we see the actual deviation in USD of different departments quarterly.
- In the second case we have the percentage deviation from the baseline for each value.
- The baseline is set to 0 when we see the deviation in values for USD while in the second case the baseline is set to 100% which allows us to see how over or under a department stands and by what percentage.
- The baseline labels have been made on both the graphs and legends are provided wherever necessary.
- The total USD deviation has been listed in the title in 1st chart whereas the overall percentage deviation has been listed in the title for 2nd chart.

Advantage:

- From both charts we can easily figure out how each department is performing when it comes to planning out the budget preparation.
- We can clearly analyze each quarter separately which helps us audit out the department that has gone over the budget proposal.
- The height of bars gives us the magnitude or degree of deviation, meaning larger/longer bars mean there is a high deviation and shorter bars mean that the proposed and actual expenditure were close.

Case 2.

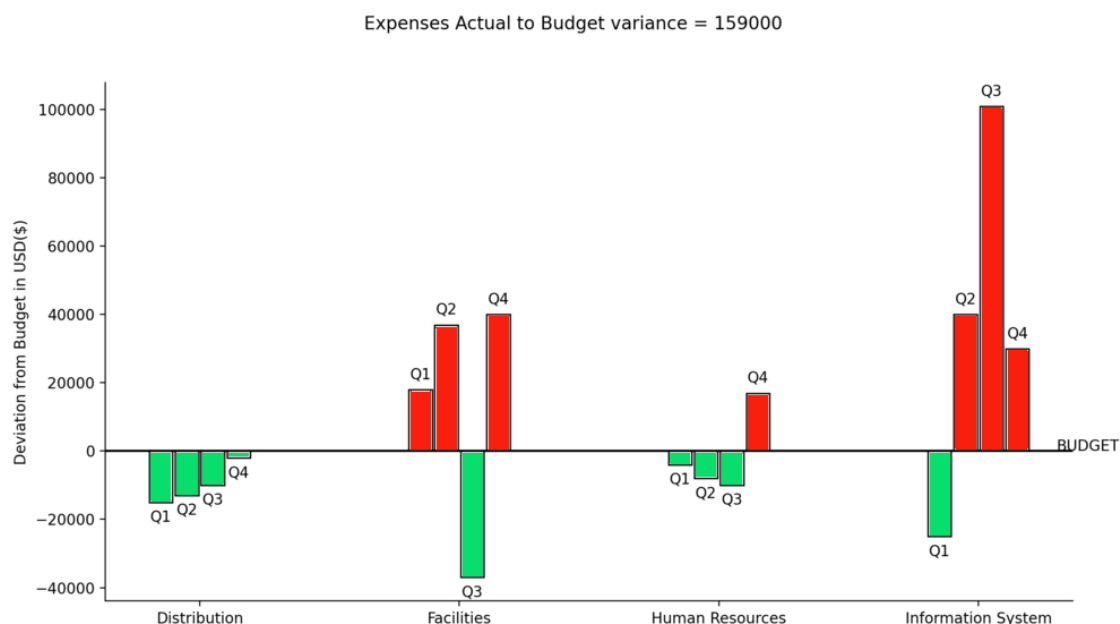


Figure 9. Variance deviation from proposed budget categorized by department

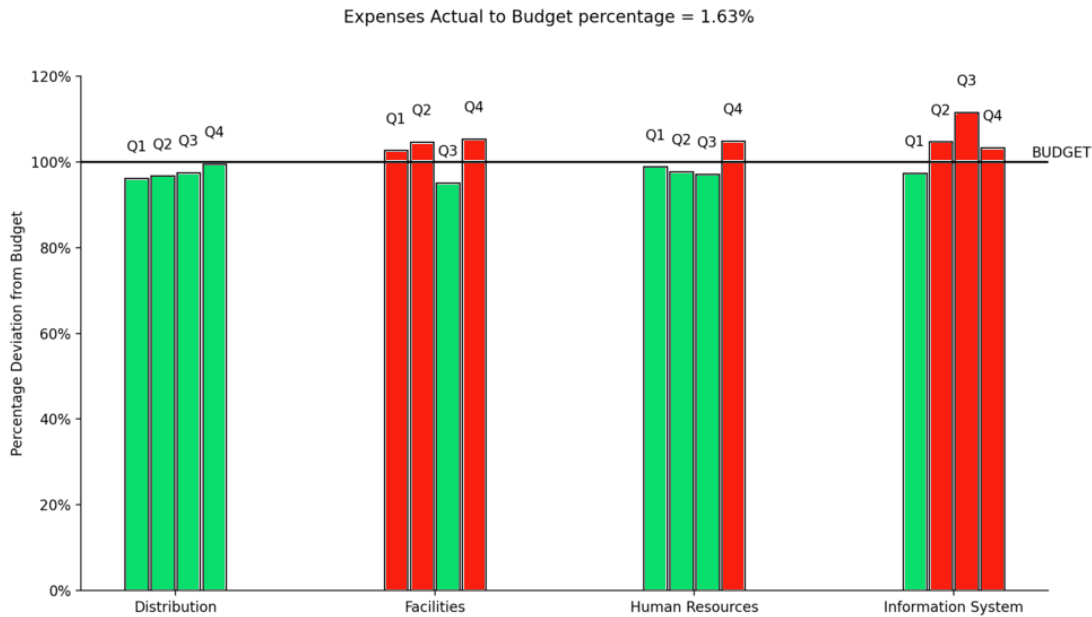


Figure 10. Variance deviation from proposed budget categorized by department

The design criteria used for the above plots is as follows:

- For each case the value for (Actual – Budget) has been calculated which gives us the deviation while choosing budget as the baseline.
- In the first case we see the actual deviation in USD for different quarters which are categorized using the departments.
- In the second case we have the percentage deviation from the baseline for each value.
- The baseline is set to 0 when we see the deviation in values for USD while in the second case the baseline is set to 100% which allows us to see how over or under a department stands and by what percentage.
- The baseline labels have been made on both the graphs and legends are provided wherever necessary.
- The total USD deviation has been listed in the title in 1st chart whereas the overall percentage deviation has been listed in the title for 2nd chart.
- I have used the analogy of green flag and red flags in these charts. A green flag which is often considered to be positive or good has been used to show all the instances when the proposed budget overpowered the actual expenditure.
- Red flags depict negative or bad instances which are reflected in the charts for all the instances where the actual expenditure went beyond the proposed budget.

Advantages:

- From both charts we can easily figure out how each department is performing when it comes to planning out the budget preparation.
- We can clearly analyze each department separately which helps us audit out the department that has gone over the budget proposal.
- The height of bars gives us the magnitude or degree of deviation, meaning larger/longer bars mean there is a high deviation and shorter bars mean that the proposed and actual expenditure were close.
- In this chart we have categorized it by departments which helps in a better visualization add since it makes us of only 2 colors and makes it look cleaner and easier to interpret.
- For instance, we can easily make out that the distribution department was highly efficient in planning the budget whereas the information systems department fared poorly as most of its proposals were inaccurate.
- One thing to notice here is that we have 9 bars of green color as compared to 7 of red color. This makes us think that overall, the entire company did a good job. However, a point to be taken care of is the magnitude of deviation which is higher on the negative instances which makes the overall budget underestimated.

Case 3.

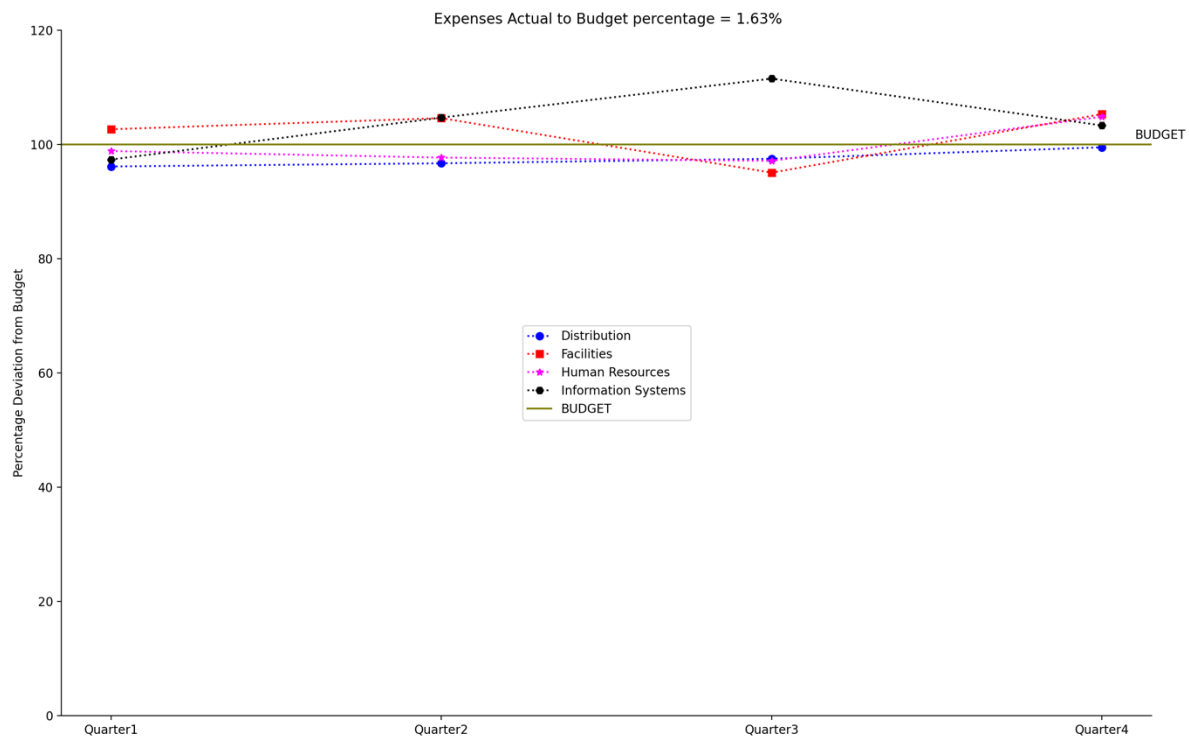


Figure 11. Line plot showing deviation from budget

- The above chart shows a line graph of the percentage deviation from the budget for all the departments categorized by different quarters.
- The design criteria for this chart is same as both the above cases.
- The difference here is that we instead of bars we use scatter points which are joined using lines to show the trending trends for each department quarterly.
- The baseline is again at 100% which is depicted in the graph and labelled accordingly.
- The advantage here is that it maximizes data-ink ratio by eliminating the big colored bars from the chart.
- Although the scale on the y-axis is quite large, but as a good visualization practice, we need to start the axis from 0 which still makes this graph readable and easy to follow.

Case 4.

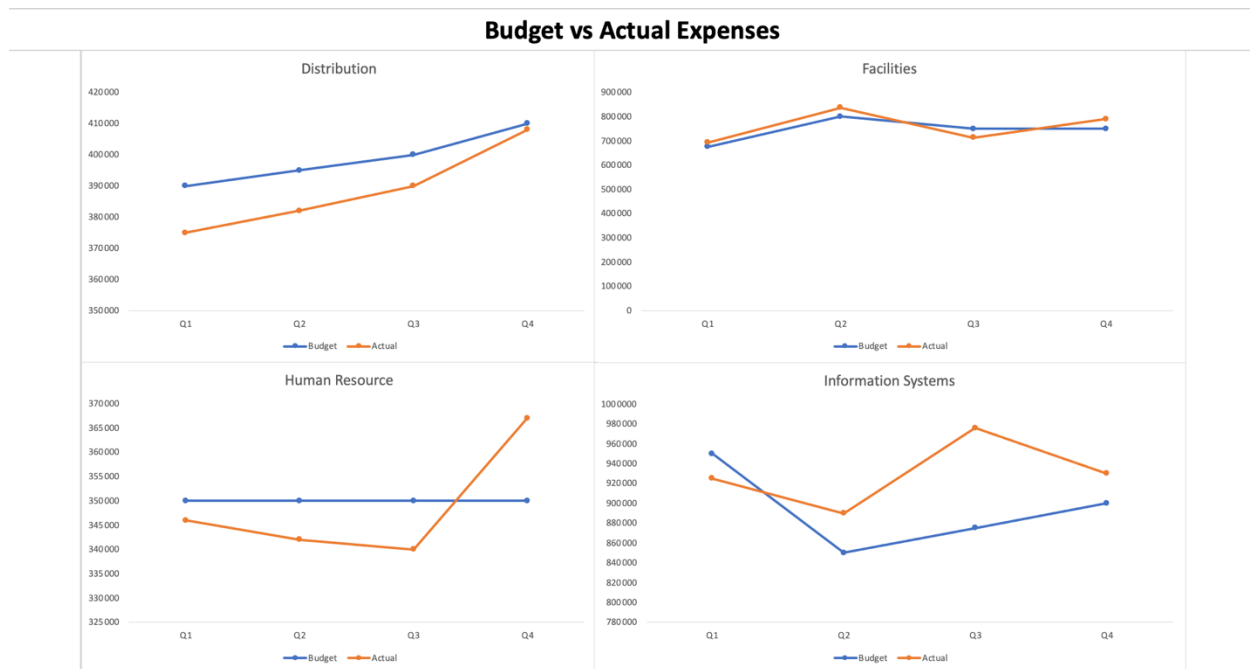


Figure 12. Budget vs Expenses

- This chart is a modification of line chart seen in case 3.
- The change here is that we have used 4 different subplots for each of the departments which helps us focus on each one of them separately.
- As compared to the previous line chart, here we have used the deviation in form of USD instead of percentage.
- Another difference here is that we have plotted the actual values here which is not strictly required but can still be used.
- The way we can interpret the chart here is that we can sense the gap between the markers for each data value quarter wise to see how big of a deviation in present for actual expenses as compared to proposed budget.
- The advantage here is that when we need to analyze the graph using actual values this type of chart makes it more clear and easier.