**Guidelines for Data Visualization and Analysis Project**

**About the Project:**

In this project, you will be working with a dataset from the Superstore, aiming to answer 30 scenario-based questions through data visualisation and analysis. Your objective is to select the best chart for each question, explain your choice. This project will showcase your proficiency in data visualisation, critical thinking, and effective communication.

**Skills Required:**

* Proficiency in data visualisation concepts and techniques.
* Familiarity with Tableau or a similar data visualisation tool.
* Strong analytical and problem-solving skills.
* Ability to choose appropriate charts based on data characteristics and question requirements.
* Clear and concise communication skills.

**Deliverables:**

* A Google document containing solutions to the scenario based questions including the screenshot of relevant chart picked for each scenario, presented in a concise and well-structured format. Make sure to provide explanations that highlight your problem-solving skills.

**Rubrics for Assessment:**

Question Responses:

* Accuracy and completeness of answers for all 30 questions.
* Clear and concise explanations that address the question's context.

Chart Selection and Explanation:

* Thoughtful rationale for choosing specific chart types.
* Justification based on data characteristics, context, and communication goals.

Creative Enhancements:

* Effective use of creative elements to enhance visualisation quality.
* Enhancements that contribute to better understanding or engagement.

**Note**:

* Duplicate this document and proceed to write your solutions.
* For each scenario and question, provide a justification for the choice of chart type. Explain why it is the best option to visualise the data effectively.
* Attach screenshots of the charts you have created in Tableau for each scenario and question using the Superstore dataset. Label them clearly to match the corresponding questions in the Google Document.
* Submit the duplicated google doc file after completion.

Use these guidelines to structure your data visualisation and analysis project. Remember to maintain consistency in your responses, explanations, and visualisation styles. This project will not only demonstrate your skills but also your ability to effectively communicate complex information through visualisations. Good luck!

**Problem Statement: Choose the Best chart for any 30 scenario based questions from Superstore Dataset.**

Imagine you are a data enthusiast aiming to excel in data visualisation and analysis. In this task, you have been given any 30 scenario-based questions derived from the Superstore dataset, and your objective is to provide insightful answers using appropriate charts. For each question, you need to select a chart that best represents the data, explain why you chose that specific chart, and then proceed to build the chosen chart using Tableau.

Your responses should be succinct, organised, and illustrative of your problem-solving capabilities.

**Dataset Link:**

<https://community.tableau.com/s/question/0D54T00000CWeX8SAL/sample-superstore-sales-excelxls>

**Please keep in mind:**

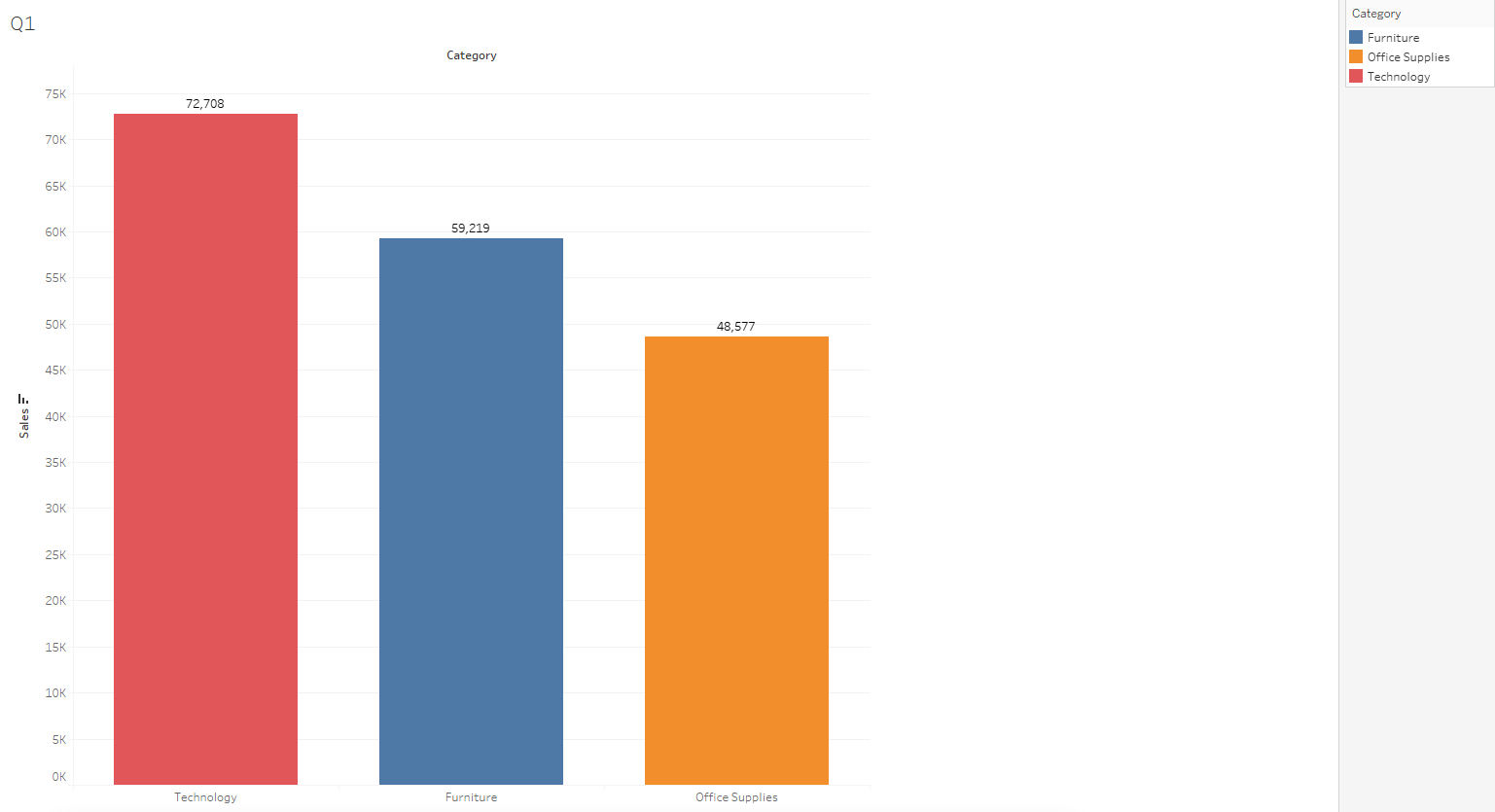
1. **Answer Completion**: Ensure that you furnish answers for all any 30 questions and build charts for them.
2. **Encouraged Creativity**: Don't hesitate to employ visuals, creative elements, or any other innovative approaches to enhance the quality of your responses.

By completing this task effectively, you'll not only demonstrate your proficiency in data visualisation and analysis but also showcase your ability to effectively communicate complex concepts through both text and charts.

**Good luck!**

**Questions:**

1. Which product categories have the highest total sales in the "Superstore" dataset?



We can see from the graph that the category ‘technology’ has the highest total sales followed by ‘furniture’ sales.

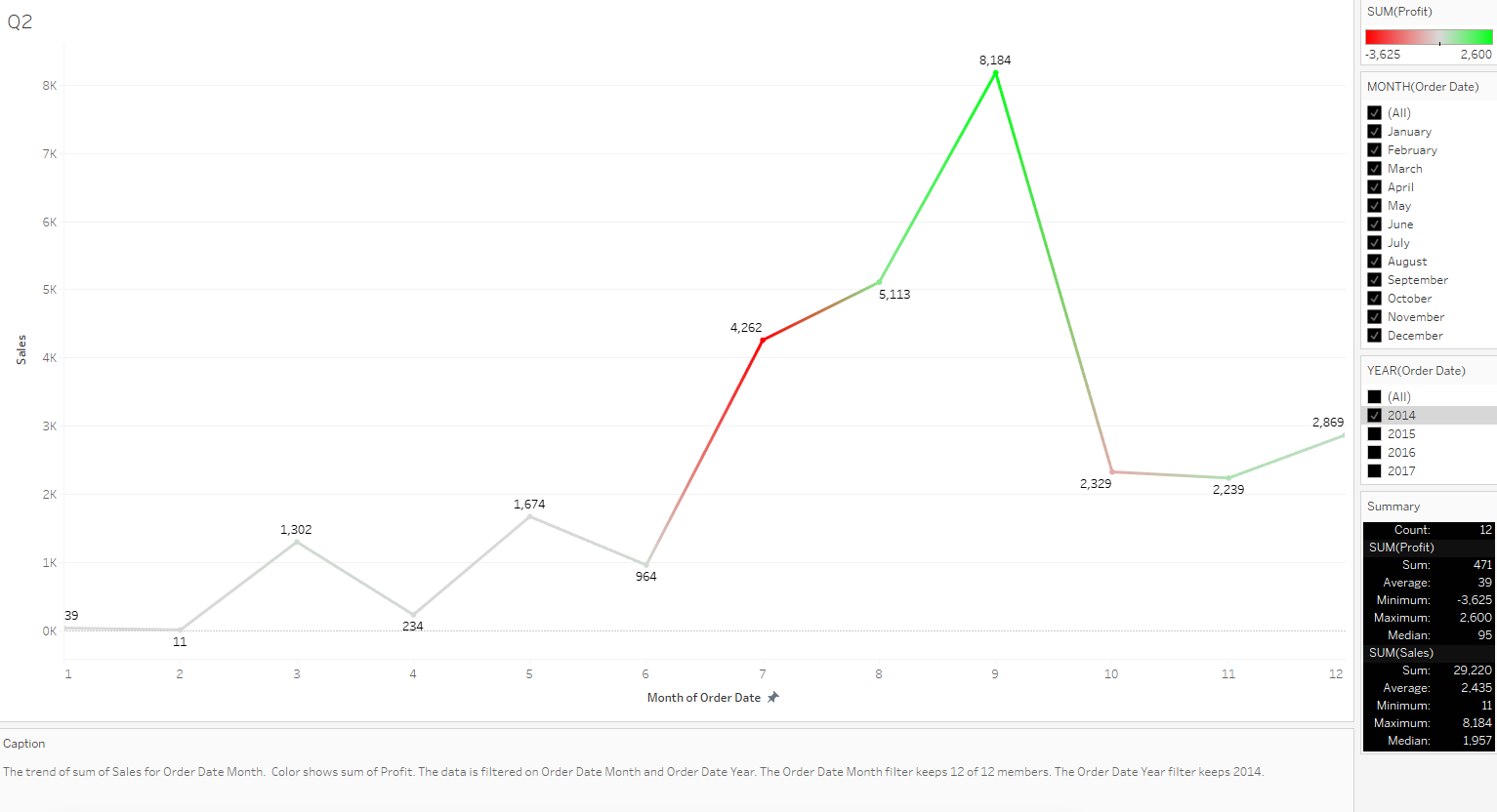
The lowest sales among the three categories was from the category ‘Office Supplies’.

The graph used here is a Bar Graph as it is one of the easiest and the most common graphs used to show the difference in values pertaining to categories.

We have also added Legends on the right side to make it easier to understand the names of different categories which are also differentiated with separate colours for distinguishing easily.

1. How do the monthly sales amounts change over the course of a year?

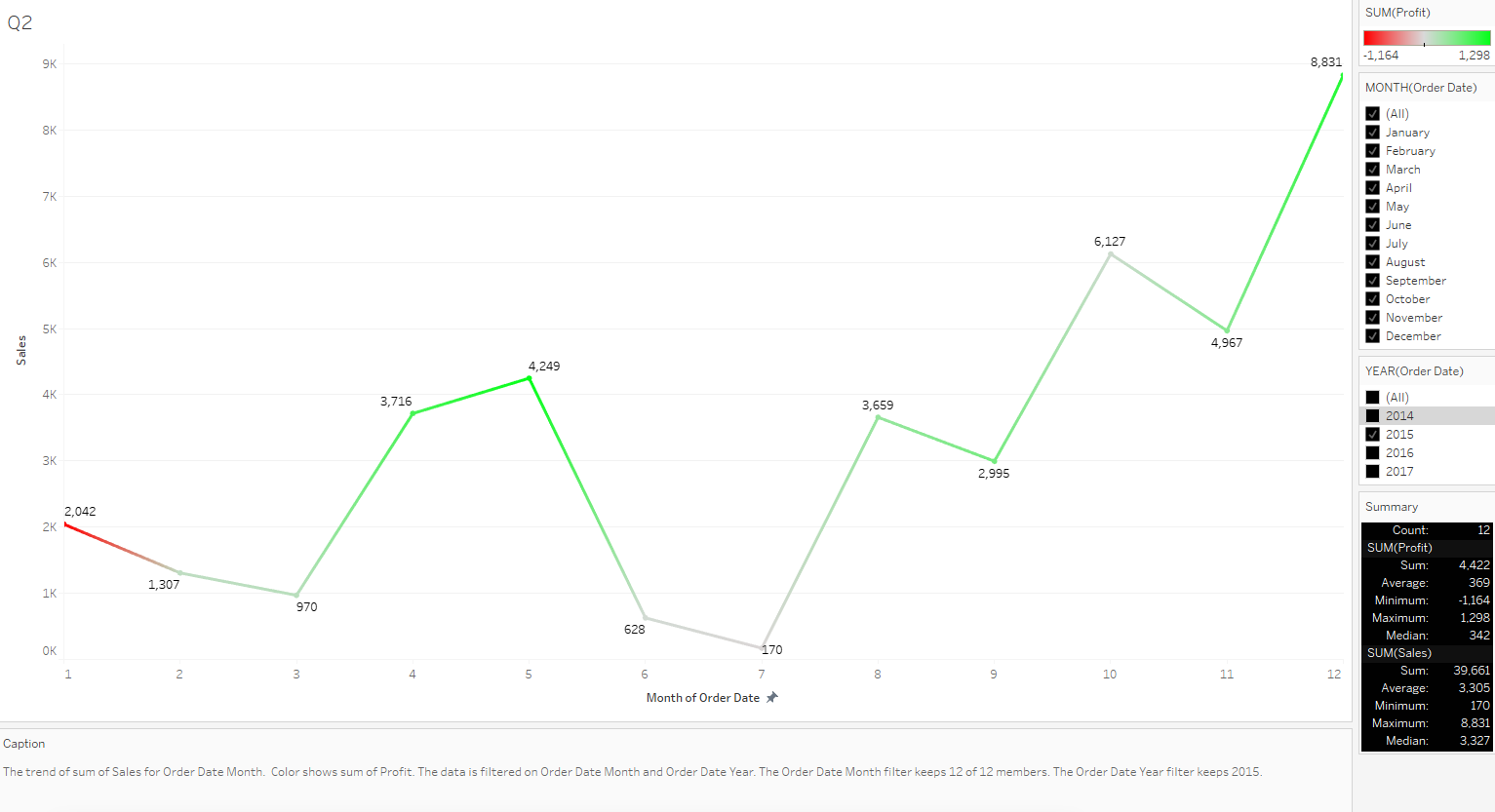
Year:- 2014



The monthly sales amount for the year 2014 show a maximum profit of $2,600 in the month of September and a loss of $3,625 incurred in the month of July. The average remains at 39$, which is not a bad sign considering it's not a loss but should be improved in the coming years as it remains close to the break-even-point.

It also shows a maximum of 8,184 units sold in the month of September and a minimum of 11 units sold in the month of July. The average remains at 2,435 units sold for the year 2014.

Year:- 2015

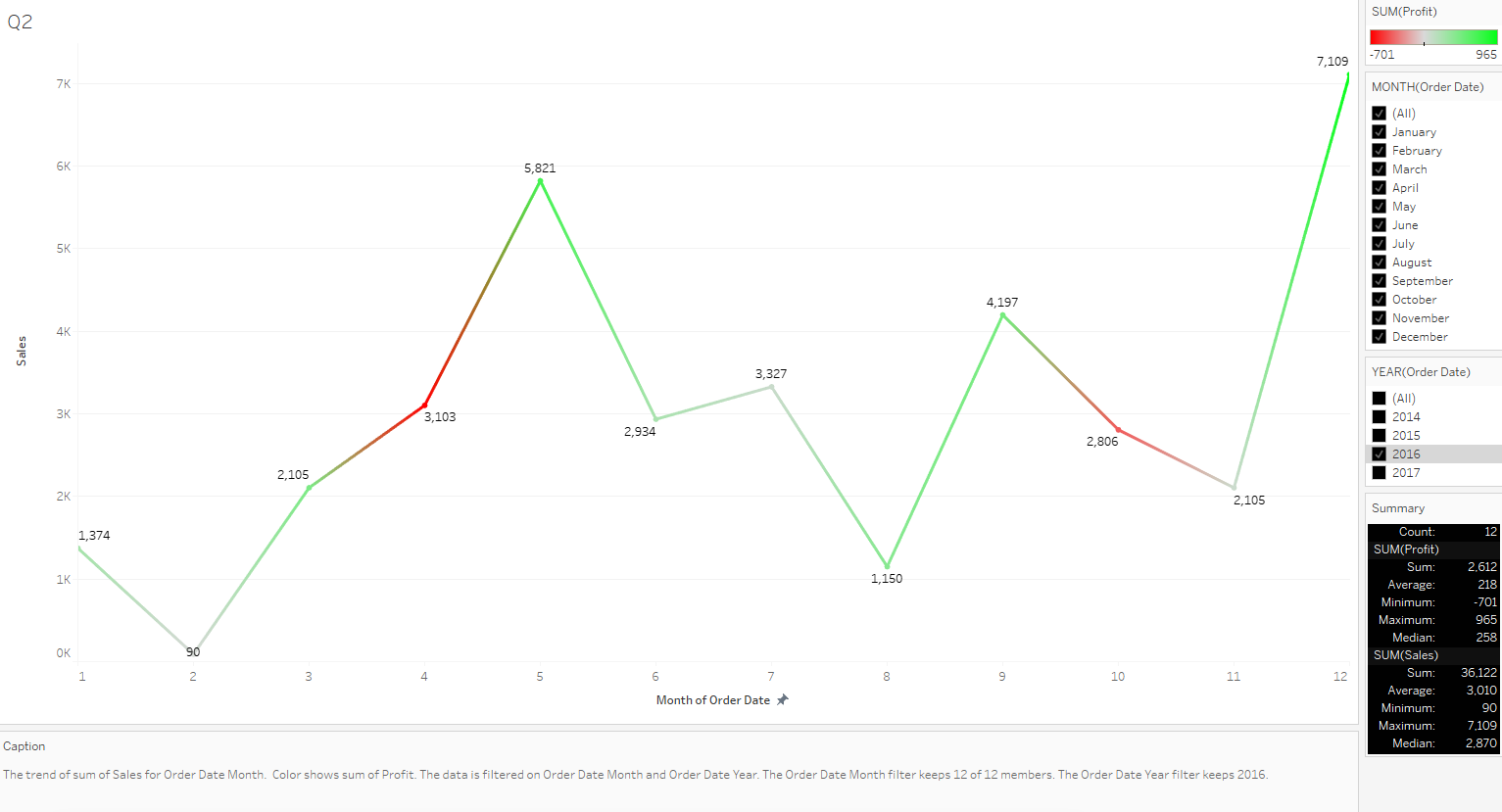


The monthly sales amount for the year 2015 show a maximum profit of $1,298 in the month of May and a loss of $1,164 incurred at the very start of the year in the month of January. The peak profit is less than the previous year but the highest loss incurred is also less indicating a steady rate of increase. The average remains at 369$, which is a good sign considering it's better than the previous year but once again should be improved in the coming years as it remains close to the break-even-point.

It also shows a maximum of 8,831 units sold in the month of December and a minimum of 170 units sold in the month of July. The average remains at 2,435 units sold for the year 2015.

The overall status looks good for the next year as the loss incurred was in the beginning of the year while the most units sold was at the end of the year indicating a steady growth and ending the year with the most units sold.

Year:- 2016

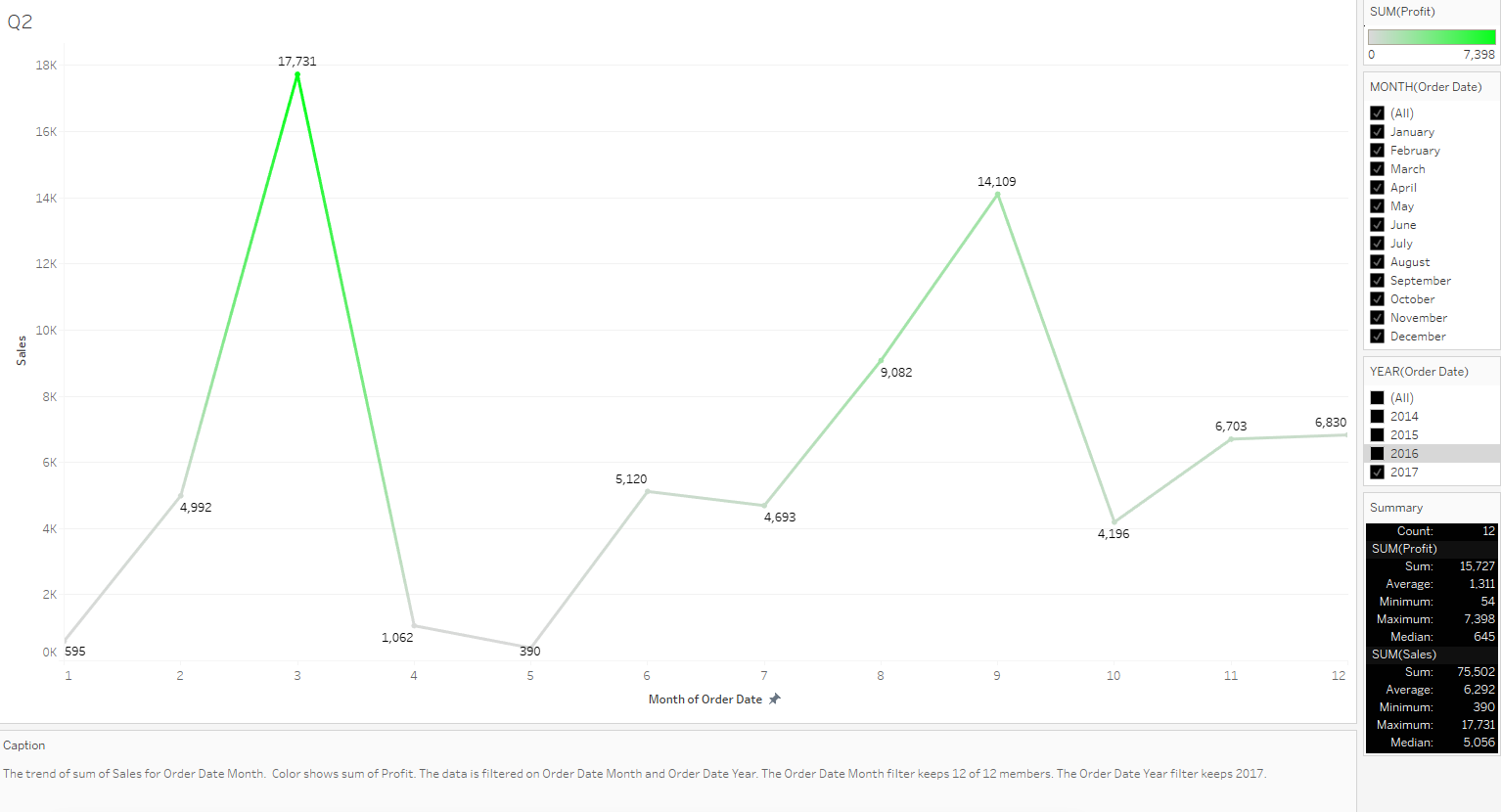


The monthly sales amount for the year 2016 show a maximum profit of $965 in the month of December and a loss of $701 incurred in the month of April. The peak profit is less than the previous year but the highest loss incurred is also less indicating improvements. The average is 218$, which is less than last year which should be improved in the coming years.

It also shows a maximum of 7,109 units sold in the month of December and a minimum of 90 units sold in the month of February. This indicates that generally the sales tend to rise by a great margin at the end of the year. The average is 3,010 units sold for the year 2016, which is higher than last year.

The overall status looks okay but still should be improved when considering a year-on-year improvement.

Year:- 2017

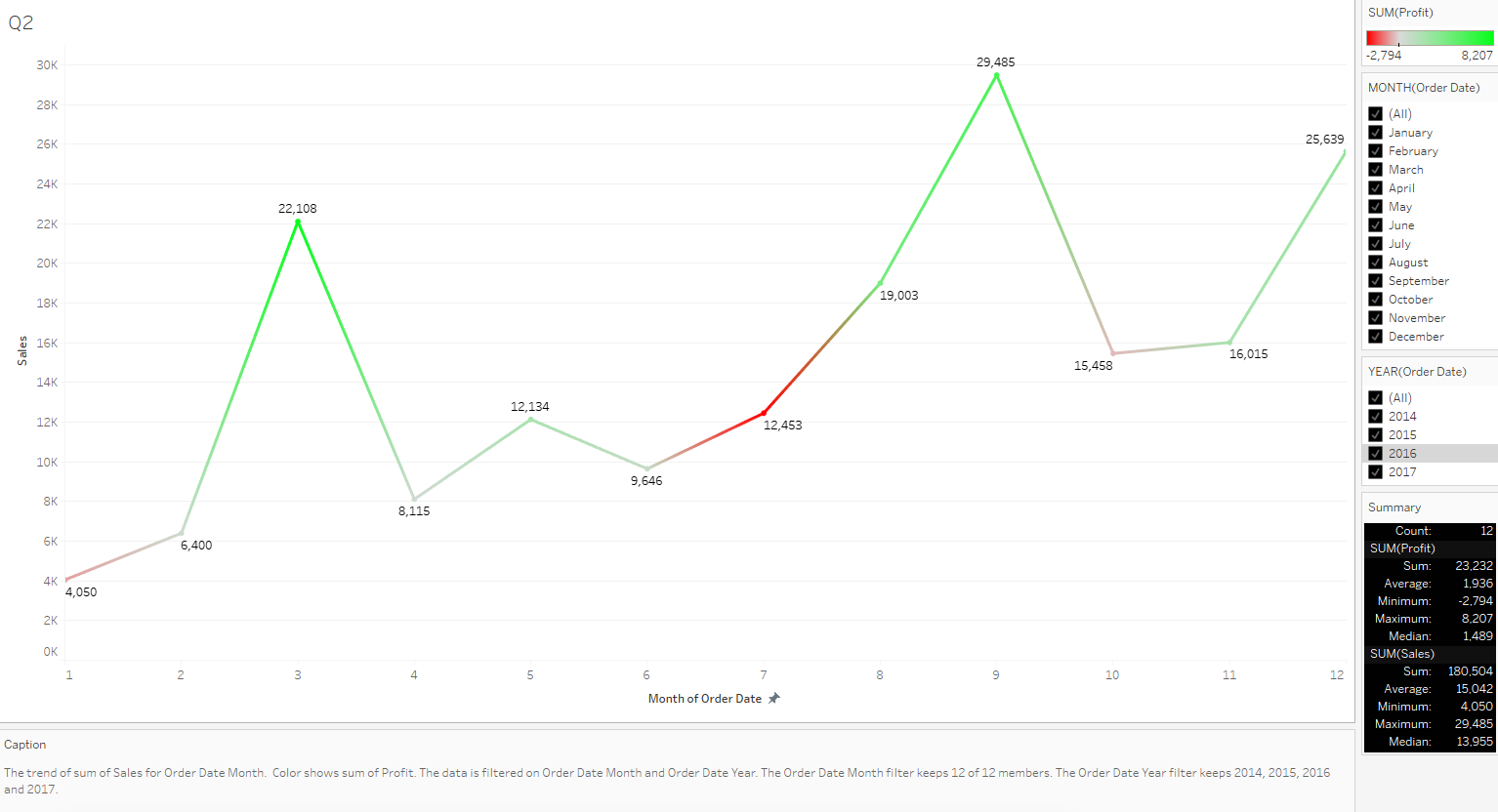


The monthly sales amount for the year 2017 show a maximum profit of $7,398 in the month of March and NO LOSS during the entire year which is surprising and reassuring for the superstore. The peak profit has improved by more than 7x when compared to the previous year and no loss incurred indicates massive improvements. The average is 1,311$, which is 6x more than last year indicating the beginning of great profits.

It also shows a maximum of 17,731 units sold in the month of March and a minimum of 390 units sold in the month of May. The average is 6,292 units sold for the year 2017, which is almost 2x than last year.

The overall status looks excellent for 2017 and indicates an outstanding year-on-year improvement.

All the years included:- 2014, 2015, 2016, 2017

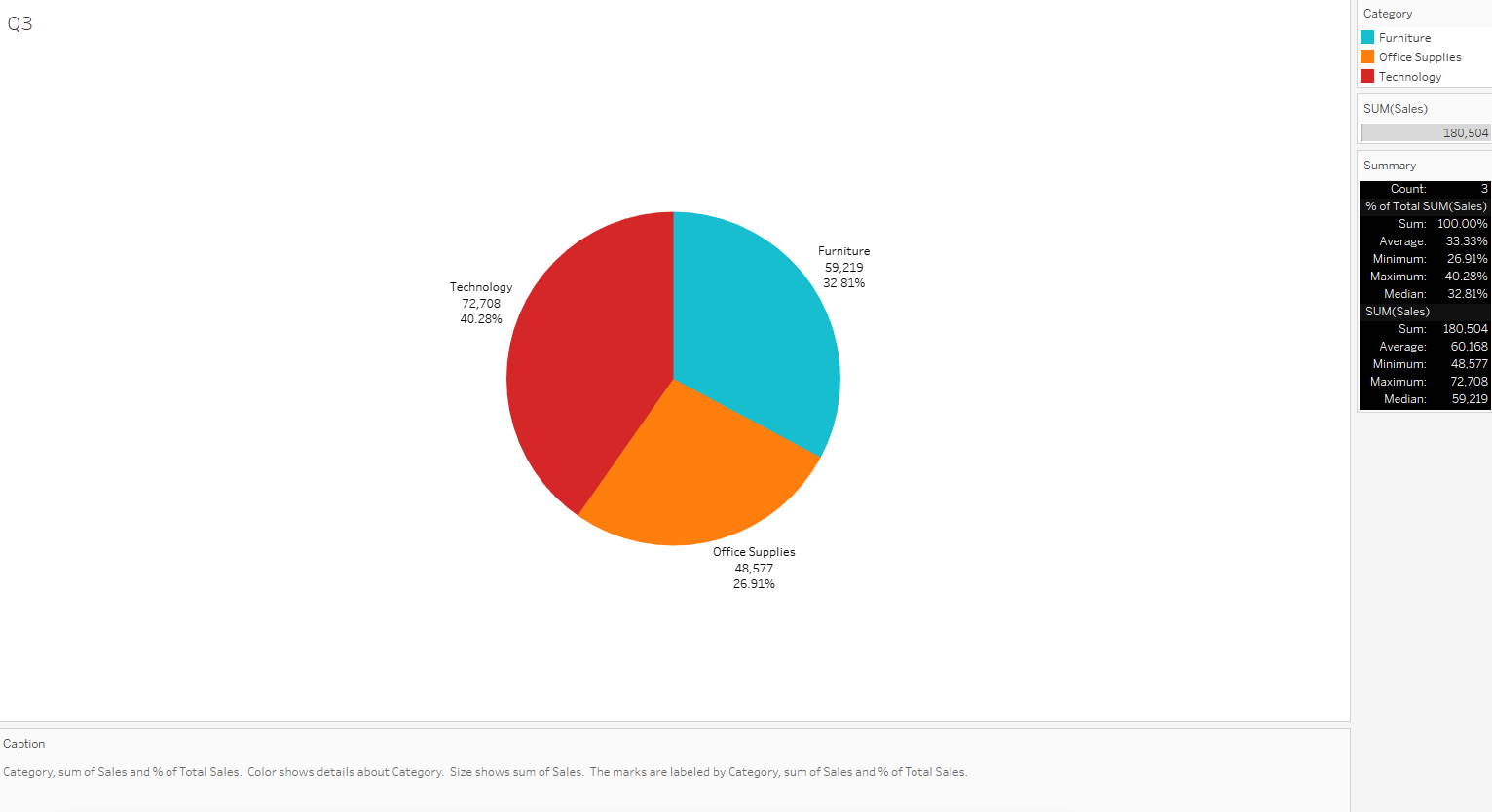


When accounting for all the 4 years, we find the total profit to be $1,396 with an average of 15,042 units sold.

We used Line Charts for visualisation as it is one of the best charts for showing trends over time. They help identify patterns, cycles, and overall directional changes in the data. It is useful for time-series analysis which is the use-case here.

We have added:

1. Legends on the right side to make it easier to understand the names of different categories which are also differentiated with separate colours for distinguishing easily.
2. Caption which shows brief interpretation and logic of all the components used and what it signifies.
3. Summary which gives a very important overview to the graph highlighting the maximum, minimum, average, etc. indicating the key pointers necessary for quick and effective data interpretation.
4. Filters for both years and months to give the utmost flexibility and ease in interpreting and understanding the data.
5. Colour Mark to the line chart and in legends highlighted with red and green gradient indicating profit or loss for the given month and year.
6. Labels indicating sales for the given month.
7. How is the total sales amount distributed among different product categories?



This Pie Chart shows the total sales amount and how it is distributed amongst different product categories.

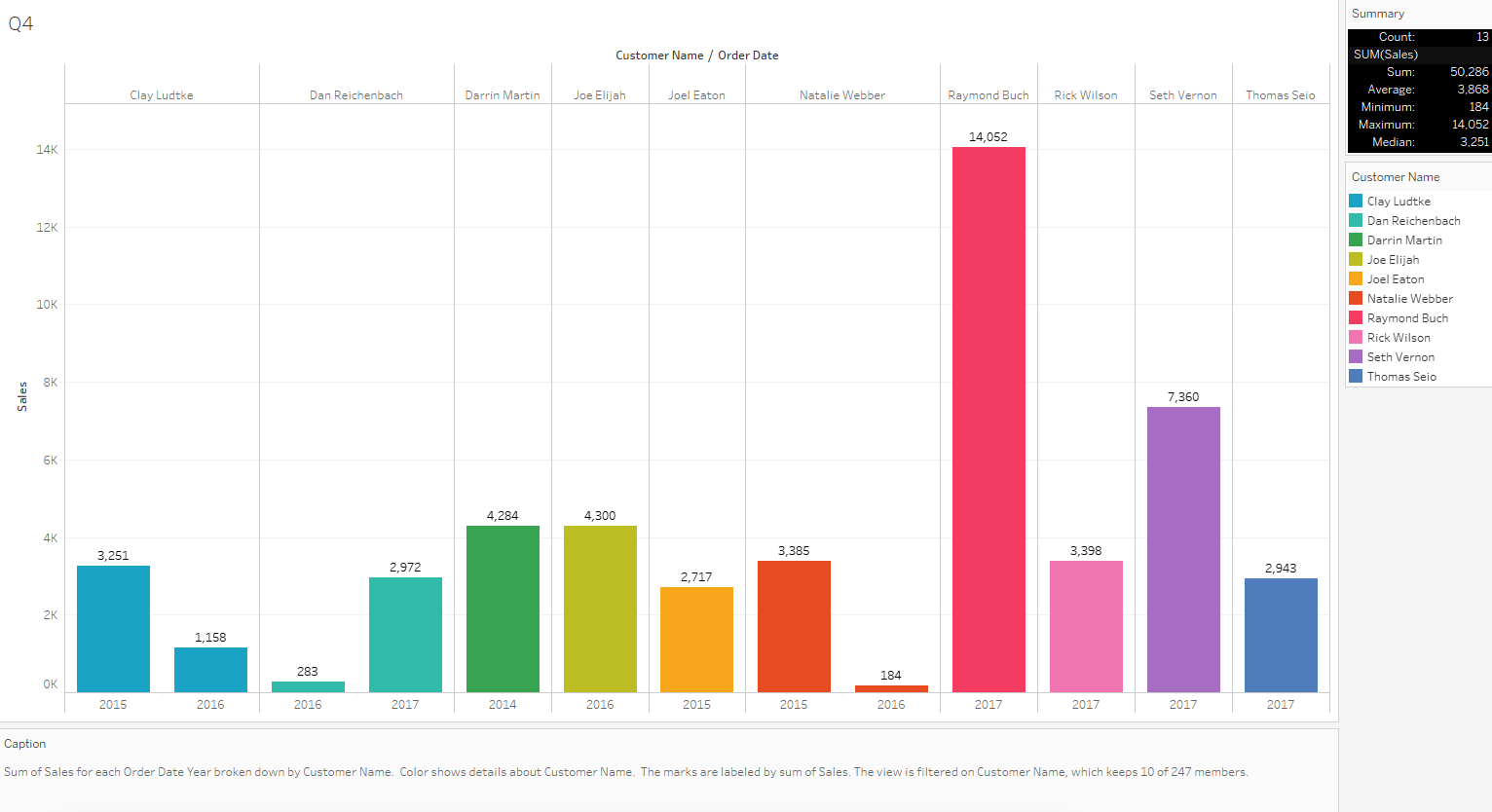
We can see that the category ‘Technology’ has the highest sales among the 3 categories with sales of 72,708 units and a share of 40.28% followed by the category ‘Furniture’ with sales of 59,219 units and a share of 32.81%.

The category with the least sales is ‘Office Supplies’ with a sale of 48,577 units and a share of 26.91%.

Total sales amount to 180,504 units.

We have added:

1. Legends on the right side to make it easier to understand the names of different categories which are also differentiated with separate colours for distinguishing easily.
2. Caption which shows brief interpretation and logic of all the components used and what it signifies.
3. Summary which gives a very important overview to the graph highlighting the maximum, minimum, average, etc. indicating the key pointers necessary for quick and effective data interpretation.
4. Separate contrast colours for easy differentiation.
5. All labels including category name, sales figure and share percentage for quick, easy understanding and relation.
6. Can we analyse the sales performance of individual customers over time?



In the above graph, we can see the sales performance for the top 10 individual customers over the period of time. This shows that all customers have a very fluctuating sales performance over time. Exceptions are there like in the case of ‘Raymond Buch’ who had the highest sales performance but only for the year 2017, whereas we can see ‘Natalie Webber’ whose sales performance reduced by a huge margin when going from the year 2105 to 2016. On a positive note, Dan Reichenbach’s sales performance improved by more than 8x going from the year 2016 to 2017.

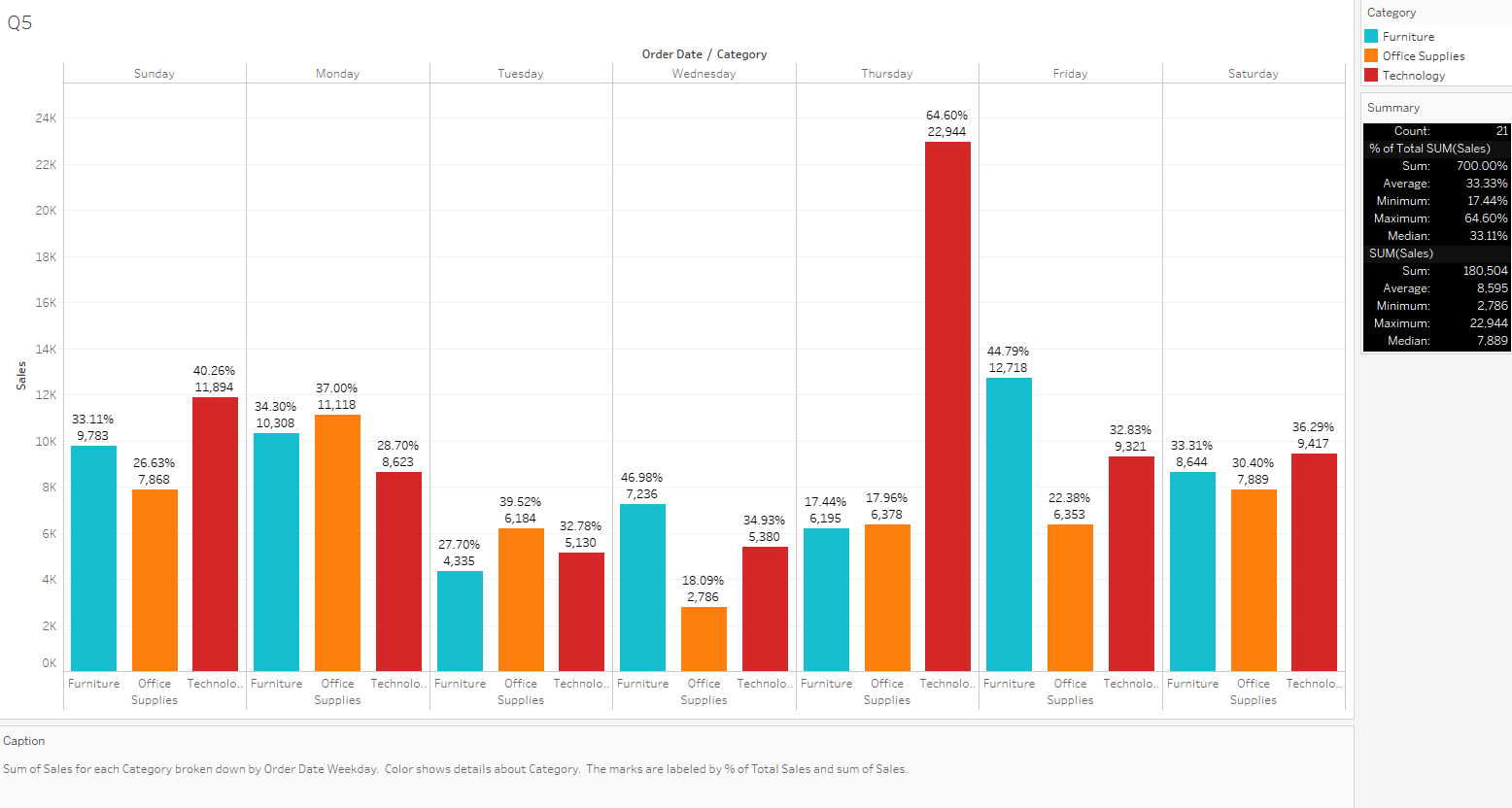
In this particular scenario, it is quite difficult to analyse the sales performance of individual customers as individual customer sales are not present for every year, therefore showing a trend using a line chart would provide inconsistent and incomplete information.

To combat this issue, we used a bar graph as it is particularly useful to see trends in discrete intervals. We showcased the sales performance of top 10 customers in various years as showing the data for all 240+ customer records would make the graph look very cluttered and difficult to understand.

We have added:

1. Legends on the right side to make it easier to understand the names of different categories which are also differentiated with separate colours for distinguishing easily.
2. Caption which shows brief interpretation and logic of all the components used and what it signifies.
3. Summary which gives a very important overview to the graph highlighting the maximum, minimum, average, etc. indicating the key pointers necessary for quick and effective data interpretation.
4. Hue colour palette contrast colours for easy differentiation.
5. Labels including category name, sales figure and year for quick, easy understanding and relation.

1. How do sales vary based on different days of the week and product categories?

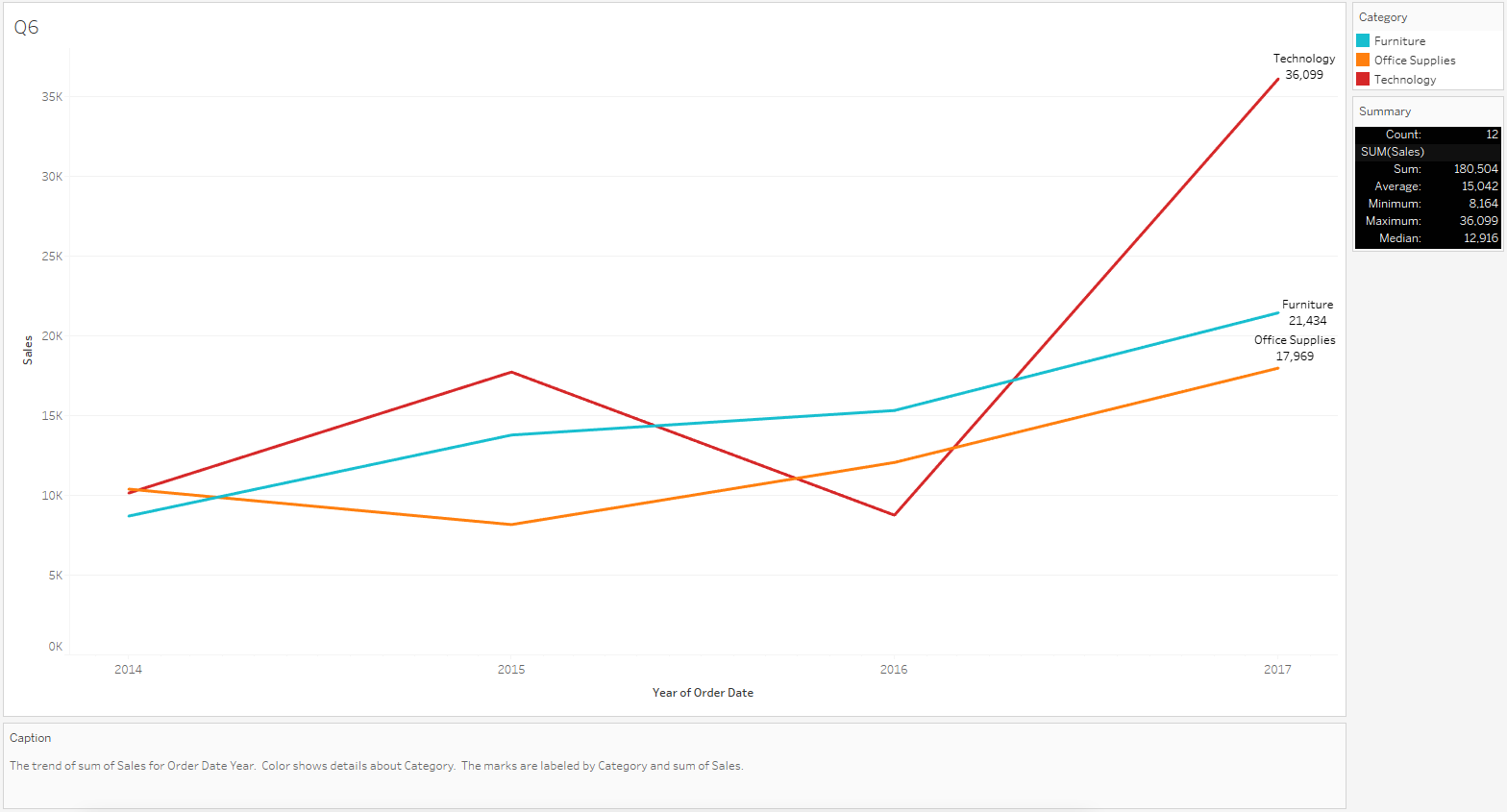


We can see from the graph how the sales vary based on different days of the week among product categories. Products from the category ‘Technology’ have the highest sales throughout the week, especially on ‘Thursday’ with a massive 64.6% boom while recording the highest sales of 22,944 units.

The graph used here is a Bar Graph as it is easy to to understand and used to show the difference in values pertaining to categories. It is particularly useful to see trends in discrete intervals.

We have added:

1. Legends on the right side to make it easier to understand the names of different categories which are also differentiated with separate colours for distinguishing easily.
2. Caption which shows brief interpretation and logic of all the components used and what it signifies.
3. Summary which gives a very important overview to the graph highlighting the maximum, minimum, average, etc. indicating the key pointers necessary for quick and effective data interpretation.
4. High contrast colour palette contrast colours for easy differentiation.
5. Labels including category name, sales figure and weekday for quick, easy understanding and relation.
6. Can we visualise the sales growth of different product categories over time?

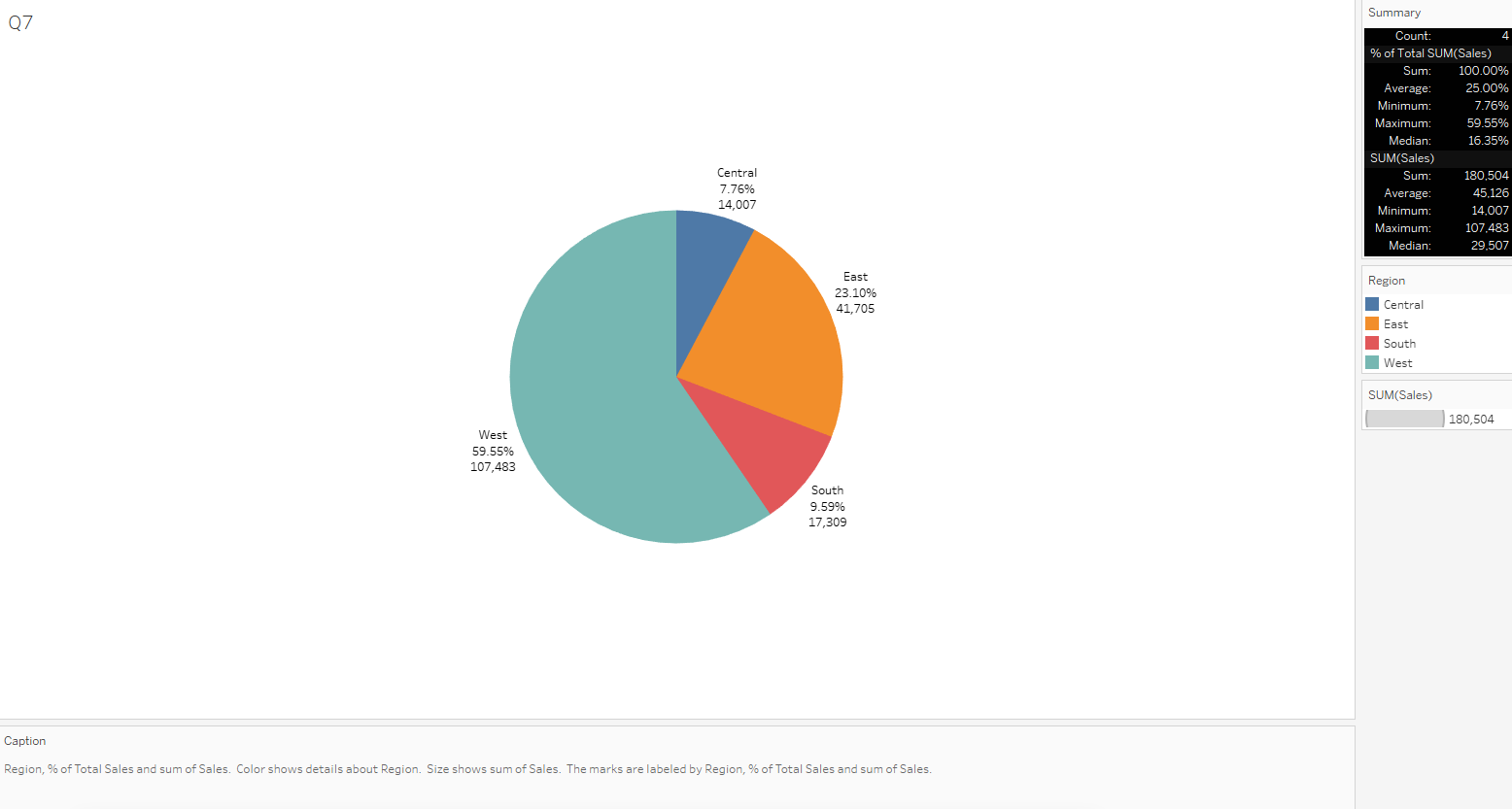


We can see from the graph how the sales vary during different years amongst product categories. The product category ‘Technology’ has seen a fluctuating decrease and increase during the years with a great increase in 2015 and the worst decrease in the year 2016 while the product categories ‘Furniture’ and ‘Office Supplies’ has seen a constant growth over the years. While the product categories ‘Furniture’ and ‘Office Supplies’ have seen a constant growth, ‘Technology’ makes the highest sales from the lowest dip to the highest reach in the year 2017.

The graph used here is a Line Chart as it is one of the best charts for showing trends over time. They help identify patterns, cycles, and overall directional changes in the data.

We have added:

1. Legends on the right side to make it easier to understand the names of different categories which are also differentiated with separate colours for distinguishing easily.
2. Caption which shows brief interpretation and logic of all the components used and what it signifies.
3. Summary which gives a very important overview to the graph highlighting the maximum, minimum, average, etc. indicating the key pointers necessary for quick and effective data interpretation.
4. High contrast colour palette contrast colours for easy differentiation.
5. Labels including category name and sales figure for quick, easy understanding and relation.
6. How does the sales distribution vary across different regions in the ‘Superstore’ dataset?



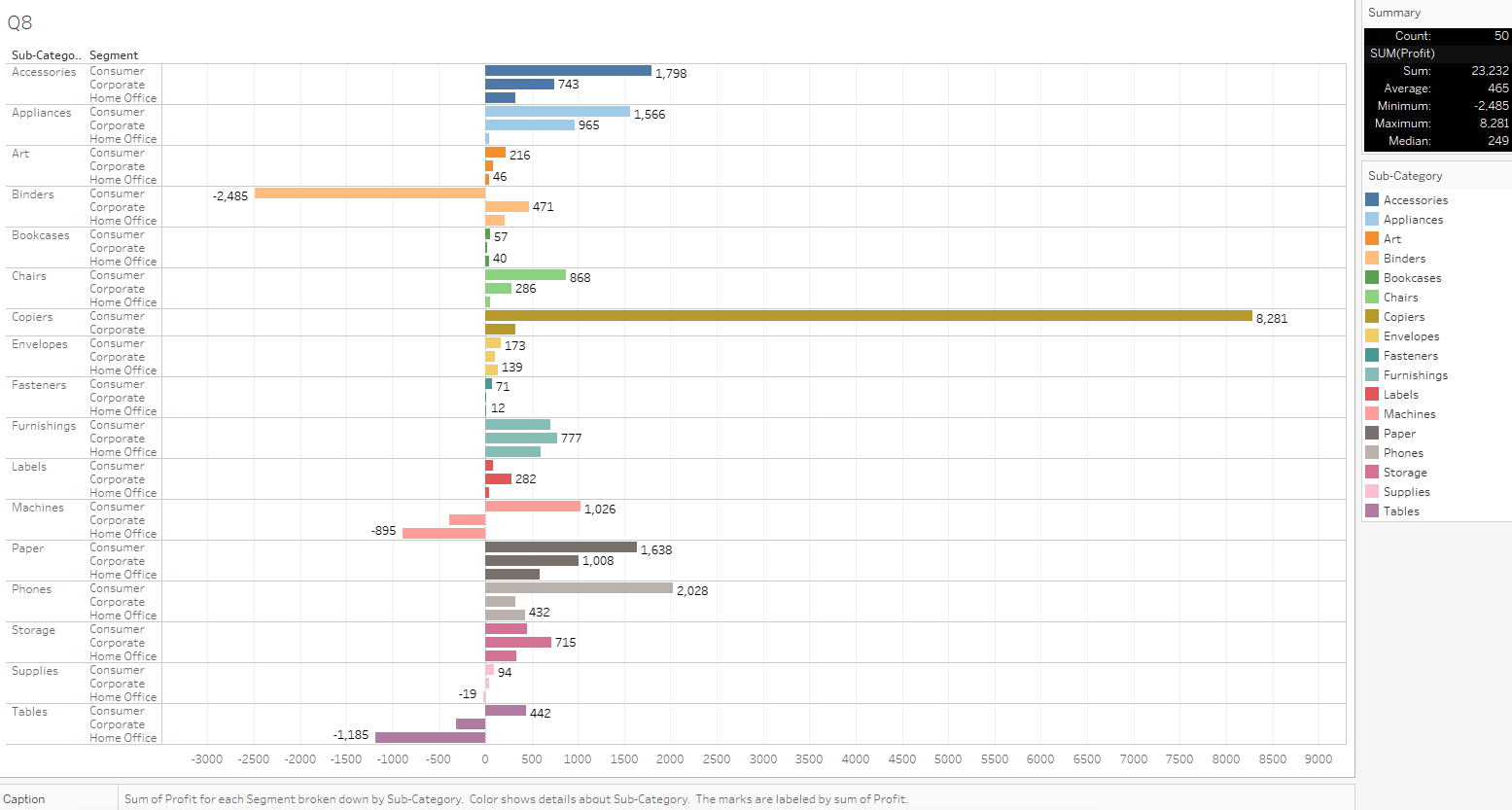
We can see from the graph the sales distribution across different regions in the ‘Superstore’ dataset. The region ‘West’ has the highest sales distribution with almost 60% of share and 107,483 units sold when compared to the other 3 regions. The lowest sales distribution is for the ‘Central’ region with only 14,007 units.

We used a pie chart for visualisation as it’s effective when we want to represent parts of a whole and emphasise the proportion of each component relative to the total.

We have added:

1. Legends on the right side to make it easier to understand the names of different categories which are also differentiated with separate colours for distinguishing easily.
2. Caption which shows brief interpretation and logic of all the components used and what it signifies.
3. Summary which gives a very important overview to the graph highlighting the maximum, minimum, average, etc. indicating the key pointers necessary for quick and effective data interpretation.
4. Different colours for each region for easy differentiation.
5. Labels including region name, sales figure and distribution percentage for quick, easy understanding and relation.

1. Can we visualise the composition of profits across various subcategories within different customer segments?

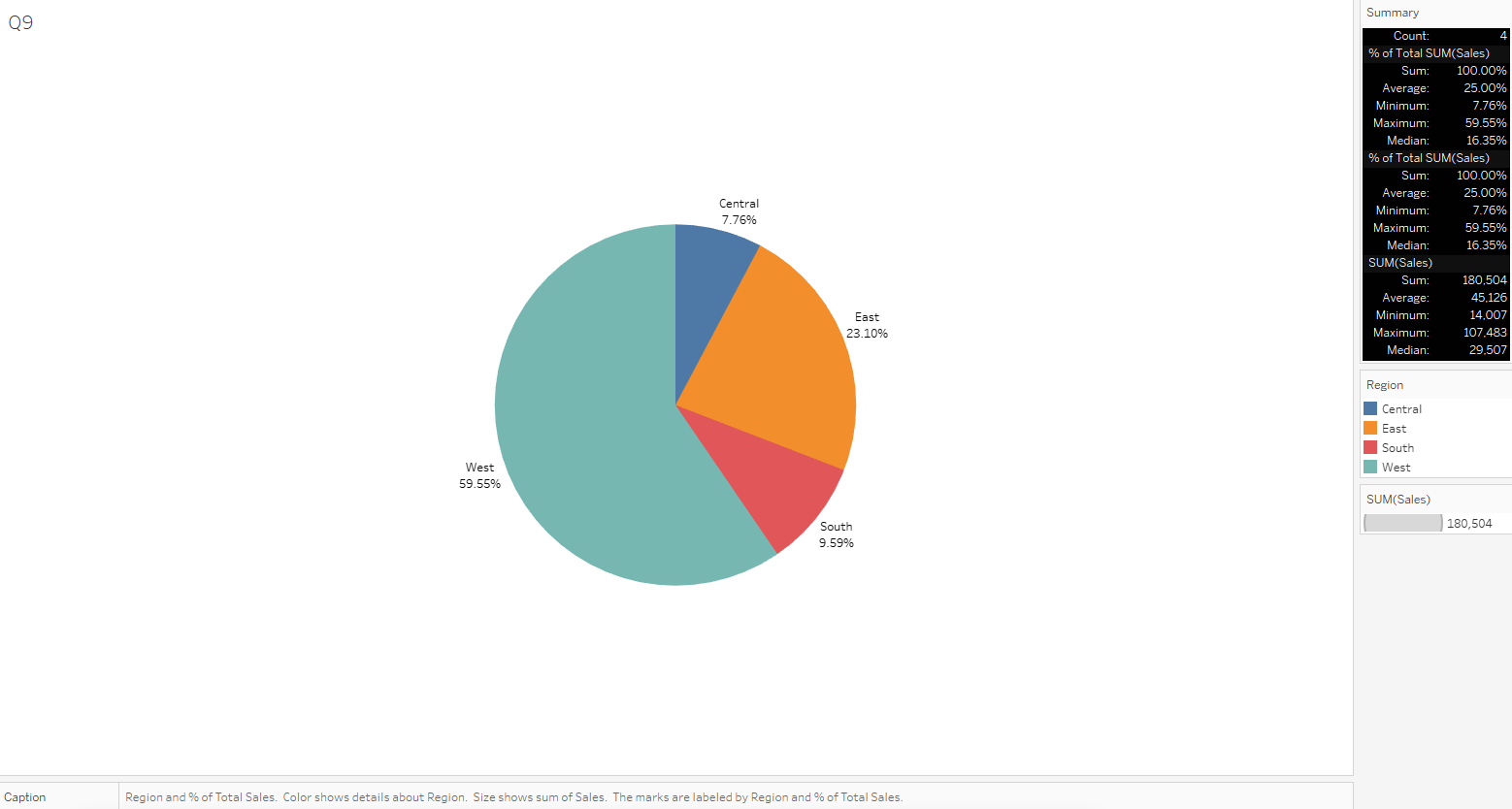


From the above graph, we can see the composition of profits across various subcategories within different customer segments. The highest profit was for ‘Copiers’ sub-category for ‘Consumers’ customer segment of $8,281 while the highest loss was for ‘Binders’ sub-category for ‘Consumers’ customer segment of $2,485.

We used a horizontal bar graph for visualisation as it’s excellent for comparing values of different categories and displaying the distribution of data effectively with horizontal bar graphs.

We have added:

1. Legends on the right side to make it easier to understand the names of different categories which are also differentiated with separate colours for distinguishing easily.
2. Caption which shows brief interpretation and logic of all the components used and what it signifies.
3. Summary which gives a very important overview to the graph highlighting the maximum, minimum, average, etc. indicating the key pointers necessary for quick and effective data interpretation.
4. Different colours for each sub-category for easy differentiation.
5. Labels including sub-category name, sales figure and consumer category for quick, easy understanding and relation.
6. What is the percentage contribution of each region to the overall sales?

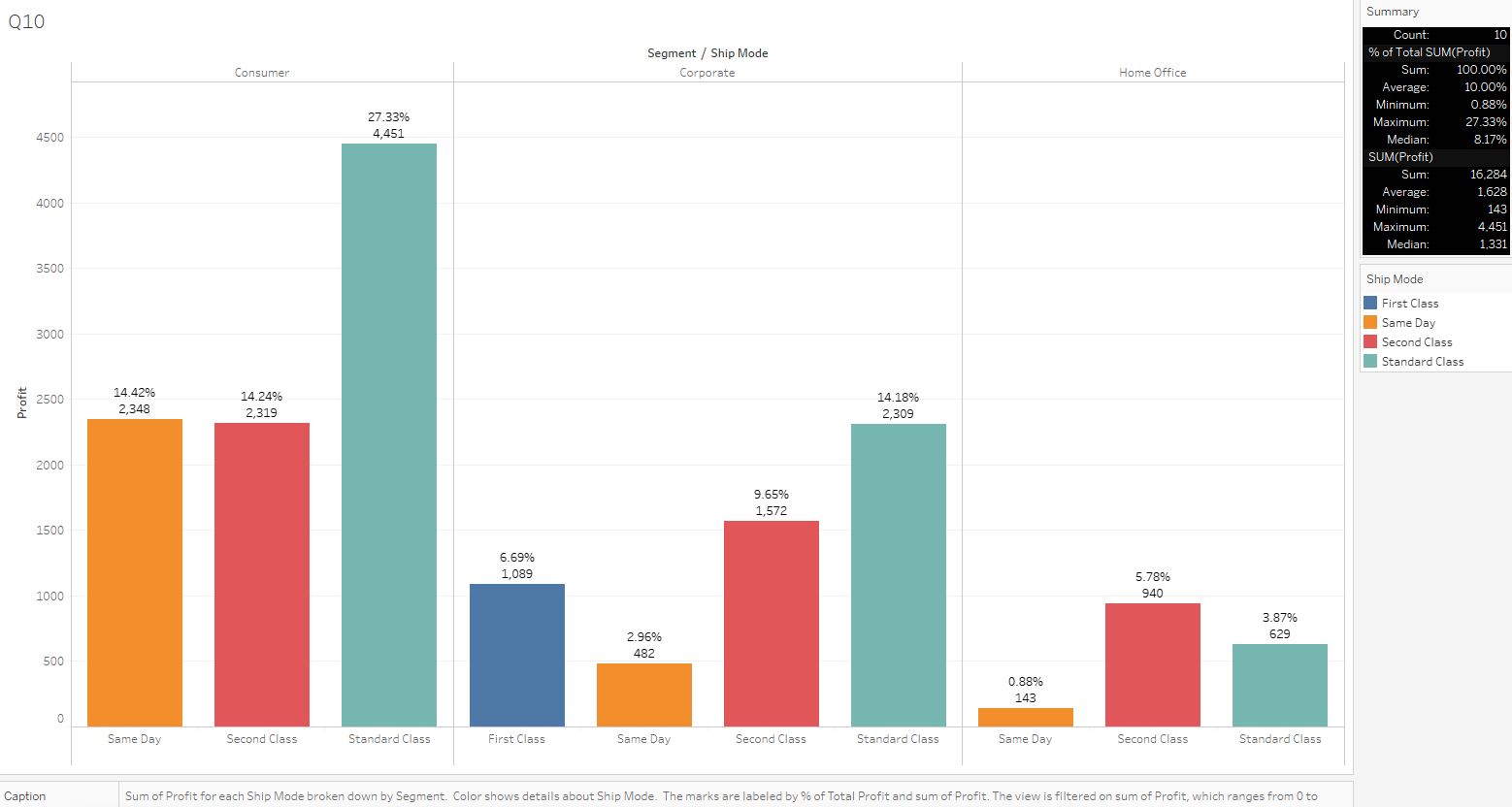


We can see from the graph the sales percentage contribution of each region to the overall sales in the ‘Superstore’ dataset. The region ‘West’ has the highest contribution to overall sales with almost 60% of share and 19,663 units sold when compared to the other 3 regions. The lowest sales contribution to overall sales is the ‘Central’ region with only 14,007 units.

We used a pie chart for visualisation as it’s effective when we want to represent parts of a whole and emphasise the proportion of each component relative to the total.

We have added:

1. Legends on the right side to make it easier to understand the names of different categories which are also differentiated with separate colours for distinguishing easily.
2. Caption which shows brief interpretation and logic of all the components used and what it signifies.
3. Summary which gives a very important overview to the graph highlighting the maximum, minimum, average, etc. indicating the key pointers necessary for quick and effective data interpretation.
4. Different colours for each region for easy differentiation.
5. Labels including region name and distribution percentage for quick, easy understanding and relation.
6. Can we visualise the profit margins associated with different shipping modes and customer segments?



From the above graph, we can see the profit margins associated with different shipping modes and customer segments. ‘Standard Class’ has the most profit margin of 27.3% associated with the ‘Consumer’ segment whereas ‘Second Class’ and ‘Same Day’ have similar profit margins.

‘Standard Class’ has the most profit margin of 14.1% associated with the ‘Corporate’ segment as well whereas ‘Same Day’ has the least profit margin of 2.96%.

‘Second Class’ has the most profit margin of 5.78% associated with the ‘Home Office’ segment whereas ‘Same Day’ has the least profit margin of less than 1%.

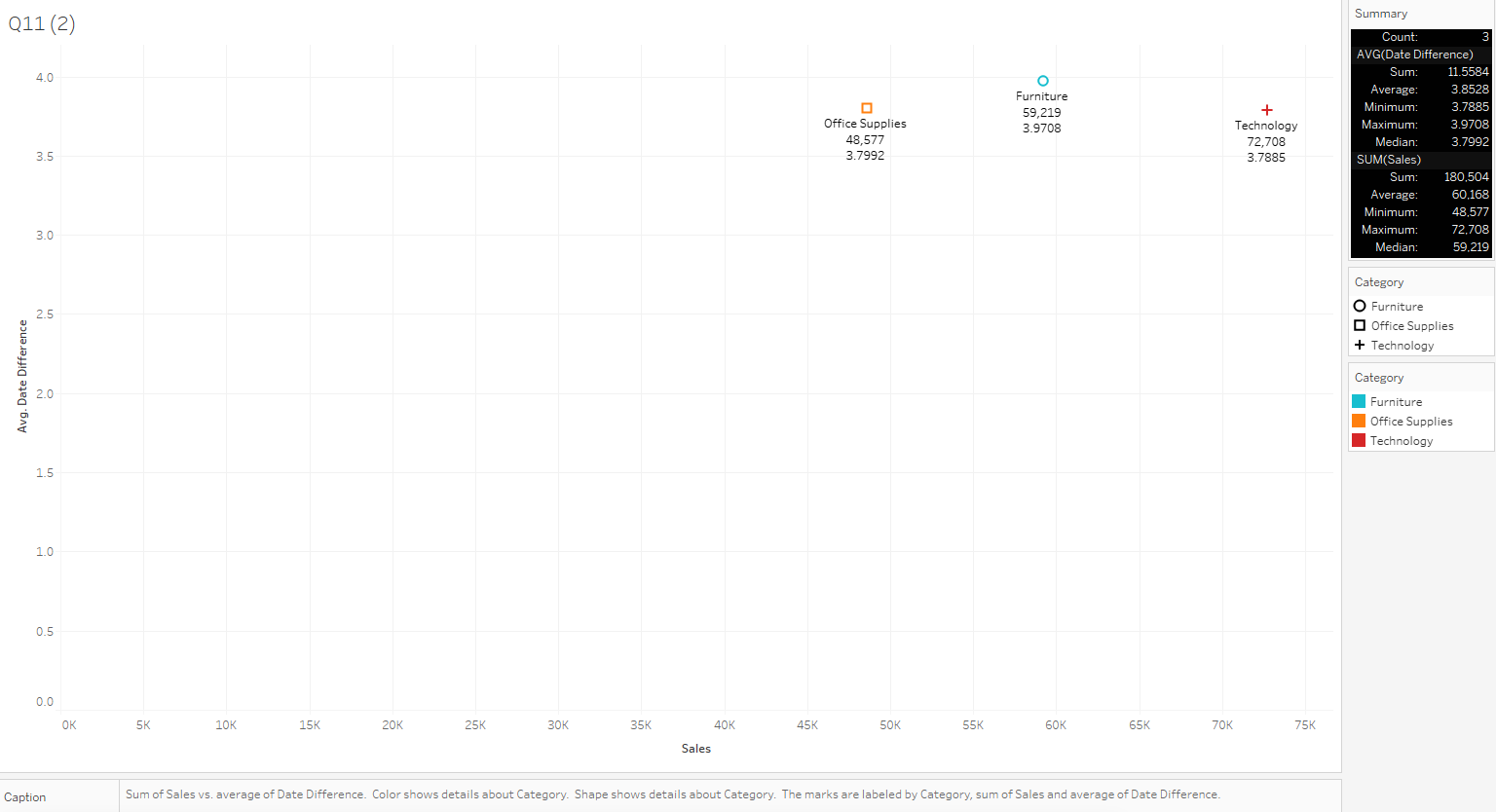
We can conclude that the ‘Consumer’ segment is the most profitable when compared to the rest and ‘Same Day’ shipping has the least profit margin in general.

We used a bar graph for visualisation as it’s excellent for comparing values of different categories and displaying the distribution of data effectively with vertical bar graphs.

We have added:

1. Legends on the right side to make it easier to understand the names of different categories which are also differentiated with separate colours for distinguishing easily.
2. Caption which shows brief interpretation and logic of all the components used and what it signifies.
3. Summary which gives a very important overview to the graph highlighting the maximum, minimum, average, etc. indicating the key pointers necessary for quick and effective data interpretation.
4. Different colours for each shipping mode for easy differentiation.
5. Labels including segment name, type of shipping mode, profit margin value and percentage for quick, easy understanding and relation.

1. How long does it take to process orders for different product categories?



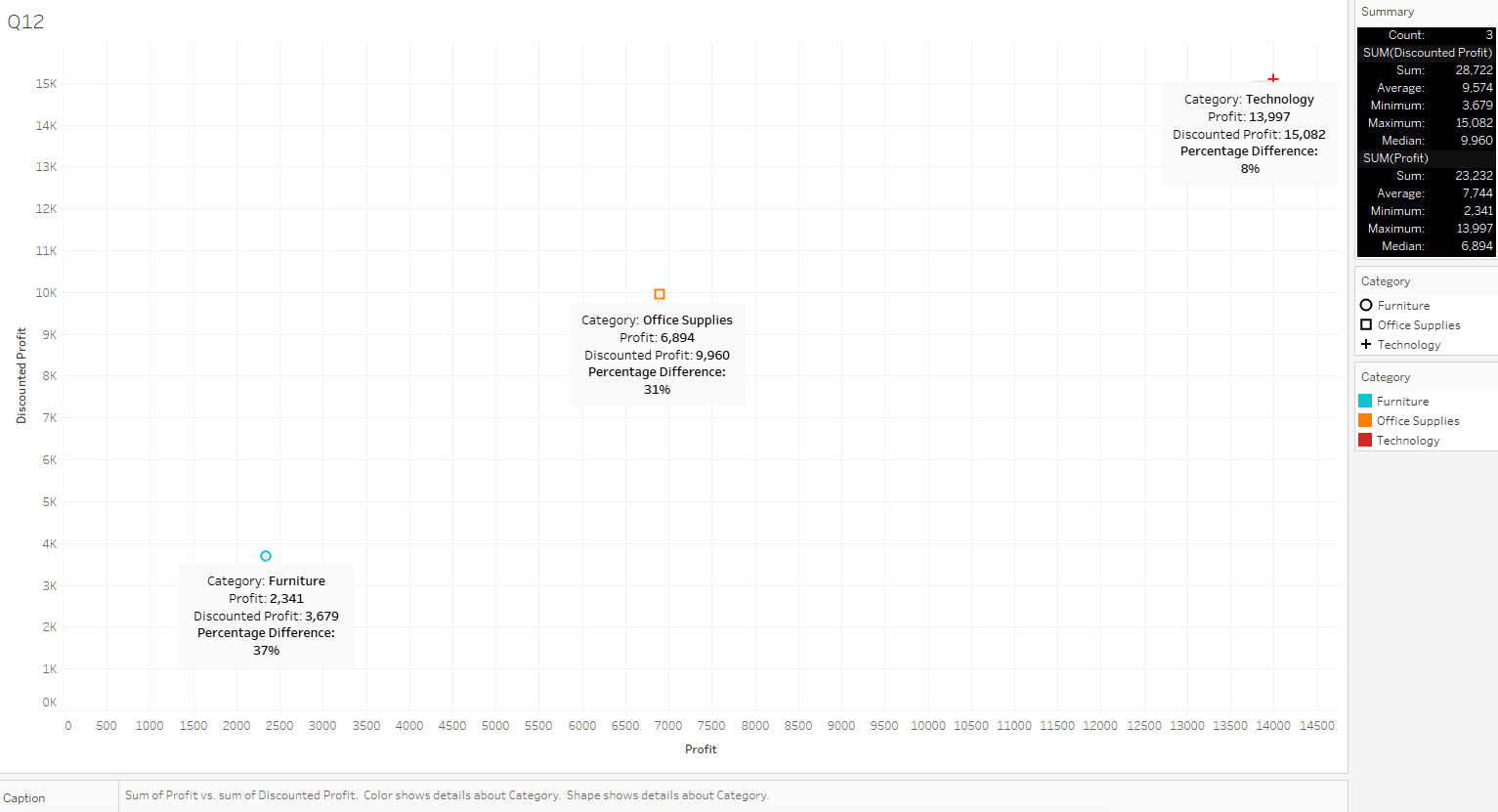
From the above graph, we can see the average time it takes to process orders for different product categories.

For obtaining this figure, we took values from two columns:- Order Date and Ship Date. We then created a calculated field by using the DATEDIFF function with the parameter as day. This gave us the difference between the two columns as integer values. We then took the average and from this, we can see that the average for all three categories are quite similar in the case of time it takes.

On average, we can see that all the categories take 3-4 days to process the orders but the difference comes in the number of orders shipped where the category ‘Furniture’ takes the win as it has the best ratio of shipping the maximum number of orders in the least amount of time whereas the rest 2 categories take more time to ship less number of orders.

We used a Scatter plot for visualisation as they provide a visual representation of how data points are distributed. Clustering or spread of points can give insights into the density and variability of the data. We have added:

1. Legends on the right side to make it easier to understand the names of different categories which are also differentiated with separate colours for distinguishing easily.
2. Caption which shows brief interpretation and logic of all the components used and what it signifies.
3. Summary which gives a very important overview to the graph highlighting the maximum, minimum, average, etc. indicating the key pointers necessary for quick and effective data interpretation.
4. Different colours for each category for easy differentiation.
5. Labels including category name, sales amount and average time in number of days for quick, easy understanding and relation.
6. How do discounts affect overall profit?



From the above graph, we can see how discounts affect the overall profit across different categories. We can the see two values for each category:

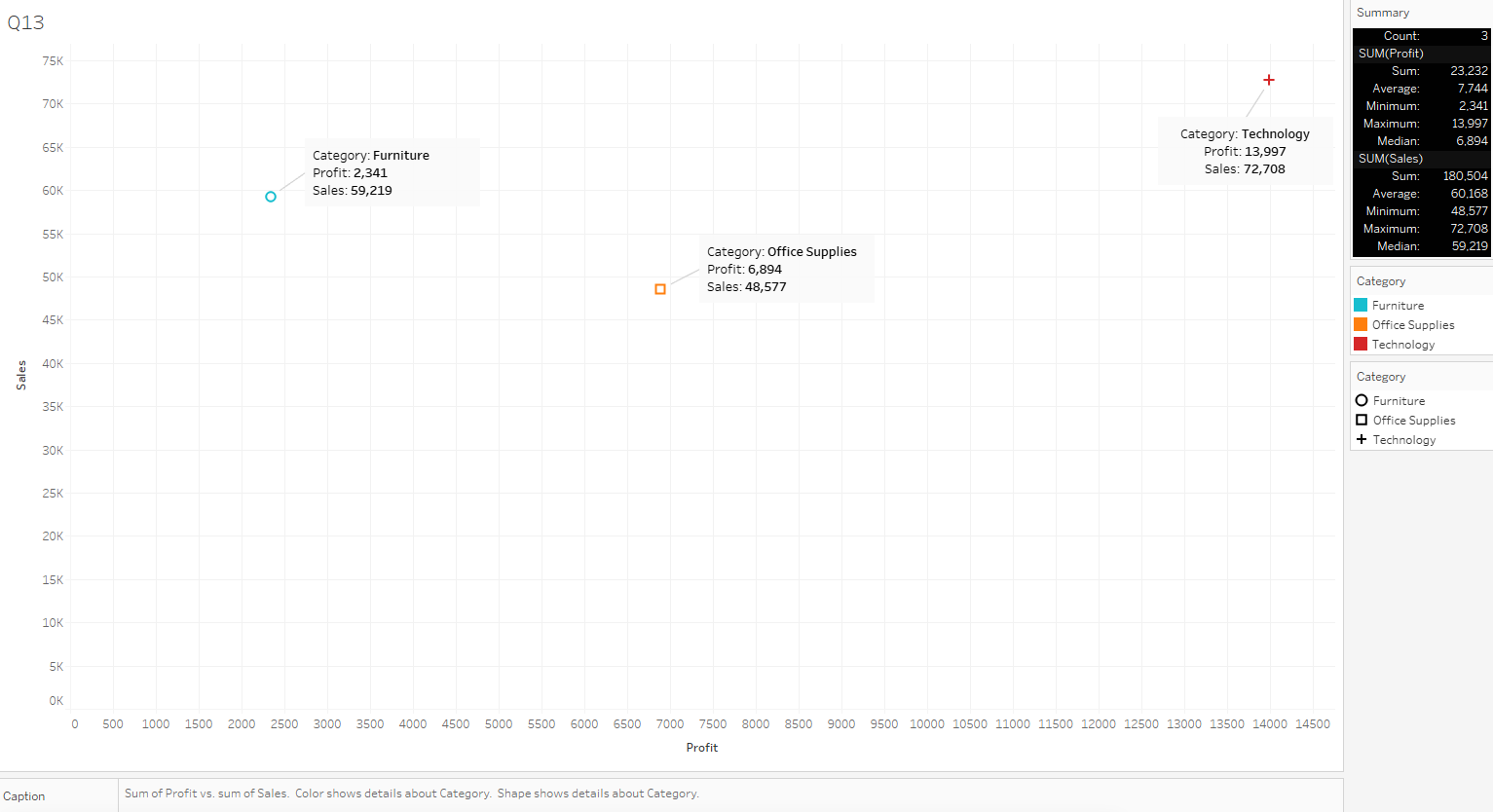
1. Profit which is before offering any discounts
2. Discounted Profit which is after offering discounts.
3. Percent difference between both.

From this, we can see how the overall profits increase when discounts are offered.

The category ‘Furniture’ has the highest gains of 37% by offering discounts while Technology has the least difference i.e of only 8%.

We used a scatter plot for plotting these points as they provide a spread visual representation of how data points are distributed. Clustering or spread of points can give insights into the density and variability of the data. We have added:

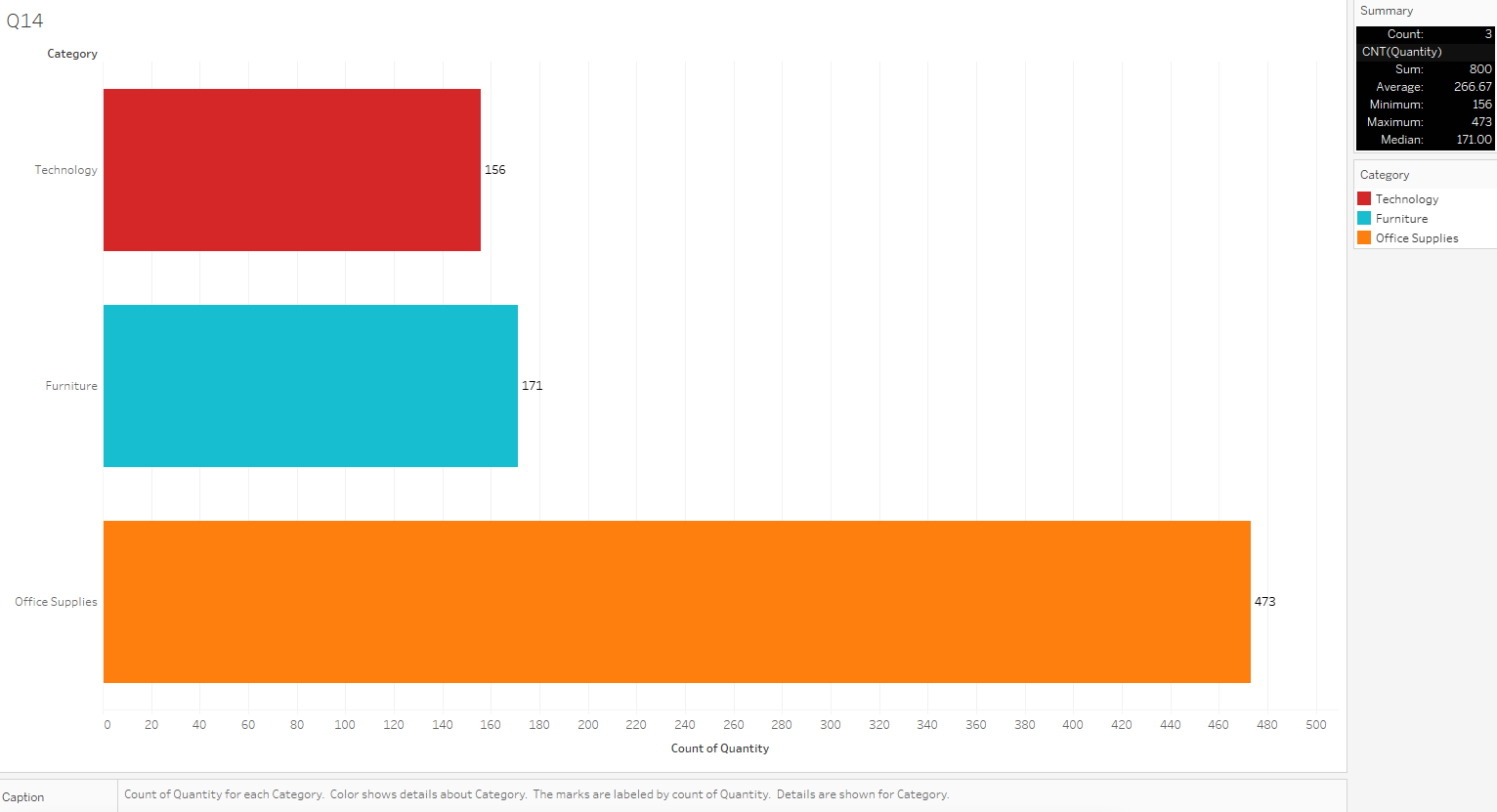
1. Annotations instead of Labels for a deeper insight, better understanding of all the data handles and for custom percentage difference.
2. Legends on the right side to make it easier to understand the names of different categories which are also differentiated with separate colours for distinguishing easily.
3. Caption which shows brief interpretation and logic of all the components used and what it signifies.
4. Summary which gives a very important overview to the graph highlighting the maximum, minimum, average, etc. indicating the key pointers necessary for quick and effective data interpretation.
5. Different colours for each category for easy differentiation.
6. Can we visualise the relationship between product sales and profitability for different product categories?



From the above graph, we can see the relationship between product sales and profitability for different product categories indicating the highest profit and sales made for the category ‘Technology’ with a profit of almost $14,000 and sales of 72,708 units. The category ‘Office supplies’ sits in the middle with a profit of almost $7,000 and sales of 48,577 units. The category ‘Furniture’ made the least profit of $2,341 despite its sales being more than ‘Office Supplies’ at 48,577 units.

Once again, we have used a scatter plot for plotting these points as plotting one variable on the x-axis and another on the y-axis helps visualise the trend and direction of the relationship. They also provide a spread visual representation of how data points are distributed. Clustering or spread of points can give insights into the density and variability of the data. We have added:

1. Annotations instead of Labels for a deeper insight, better understanding of all the data handles.
2. Legends on the right side to make it easier to understand the names of different categories which are also differentiated with separate colours for distinguishing easily.
3. Caption which shows brief interpretation and logic of all the components used and what it signifies.
4. Summary which gives a very important overview to the graph highlighting the maximum, minimum, average, etc. indicating the key pointers necessary for quick and effective data interpretation.
5. Different colours for each category for easy differentiation.
6. What is the distribution of order quantities for products in the dataset?

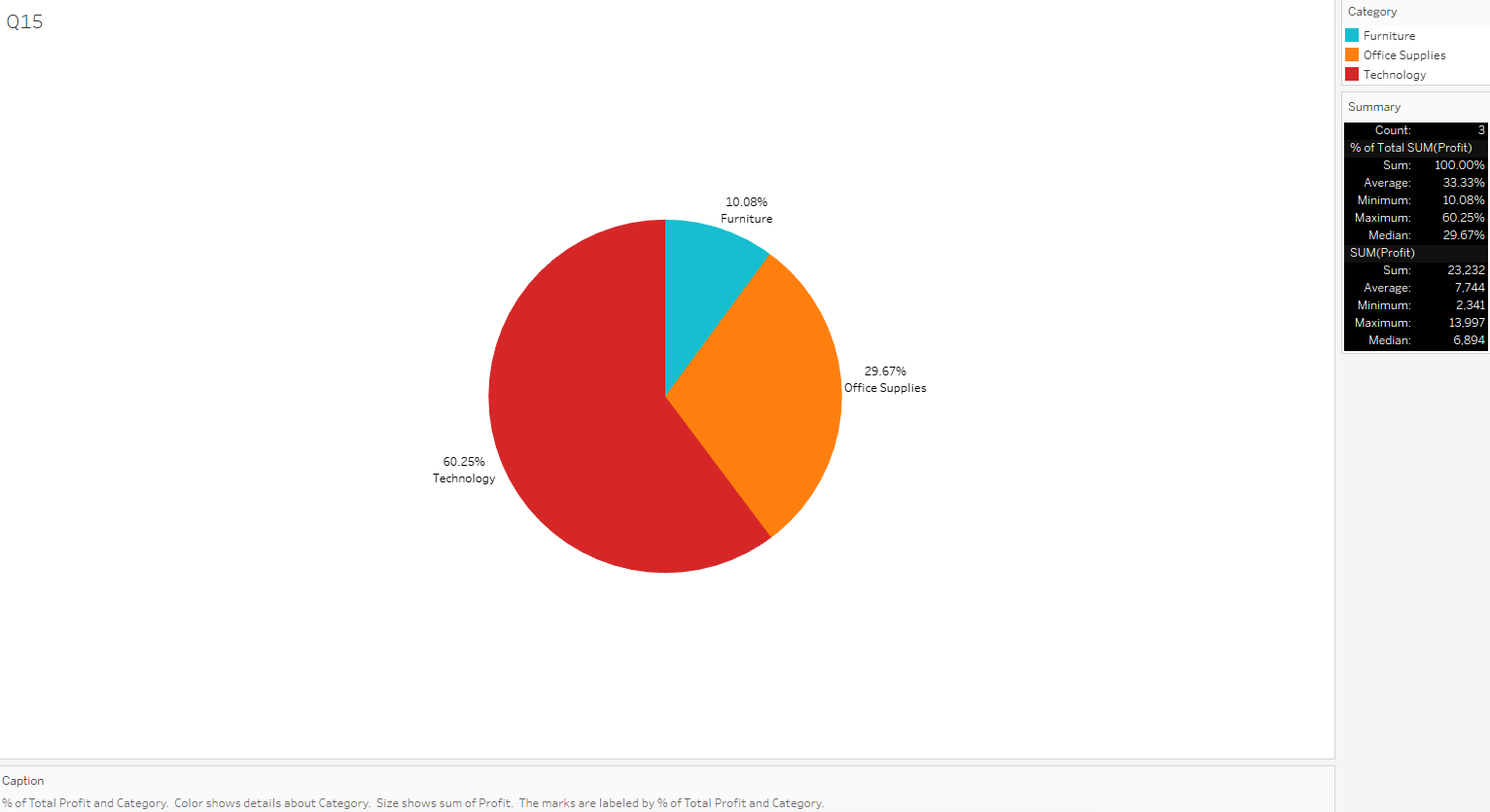


From the above graph, we can see the distribution of order quantities for the products in the dataset. We can clearly see the order quantities for the product ‘Office Supplies’ is the highest with a quantity of 473 orders whereas the rest of the two categories trail behind with less than half of the order quantities. When we compare this graph with the previous graph, we can see that even though ‘Office Supplies’ has the highest order quantities, it is not directly proportional to the number of sales and profit.

The graph used here is the horizontal bar graph for visualisation as it’s excellent for comparing values of different categories and displaying the distribution of data effectively with horizontal bar graphs.

We have added:

1. Legends on the right side to make it easier to understand the names of different categories which are also differentiated with separate colours for distinguishing easily.
2. Caption which shows brief interpretation and logic of all the components used and what it signifies.
3. Summary which gives a very important overview to the graph highlighting the maximum, minimum, average, etc. indicating the key pointers necessary for quick and effective data interpretation.
4. Different colours for each category for easy differentiation.
5. Labels including quantity orders for quick, easy understanding and relation.
6. How do the profit distributions vary across different product categories?



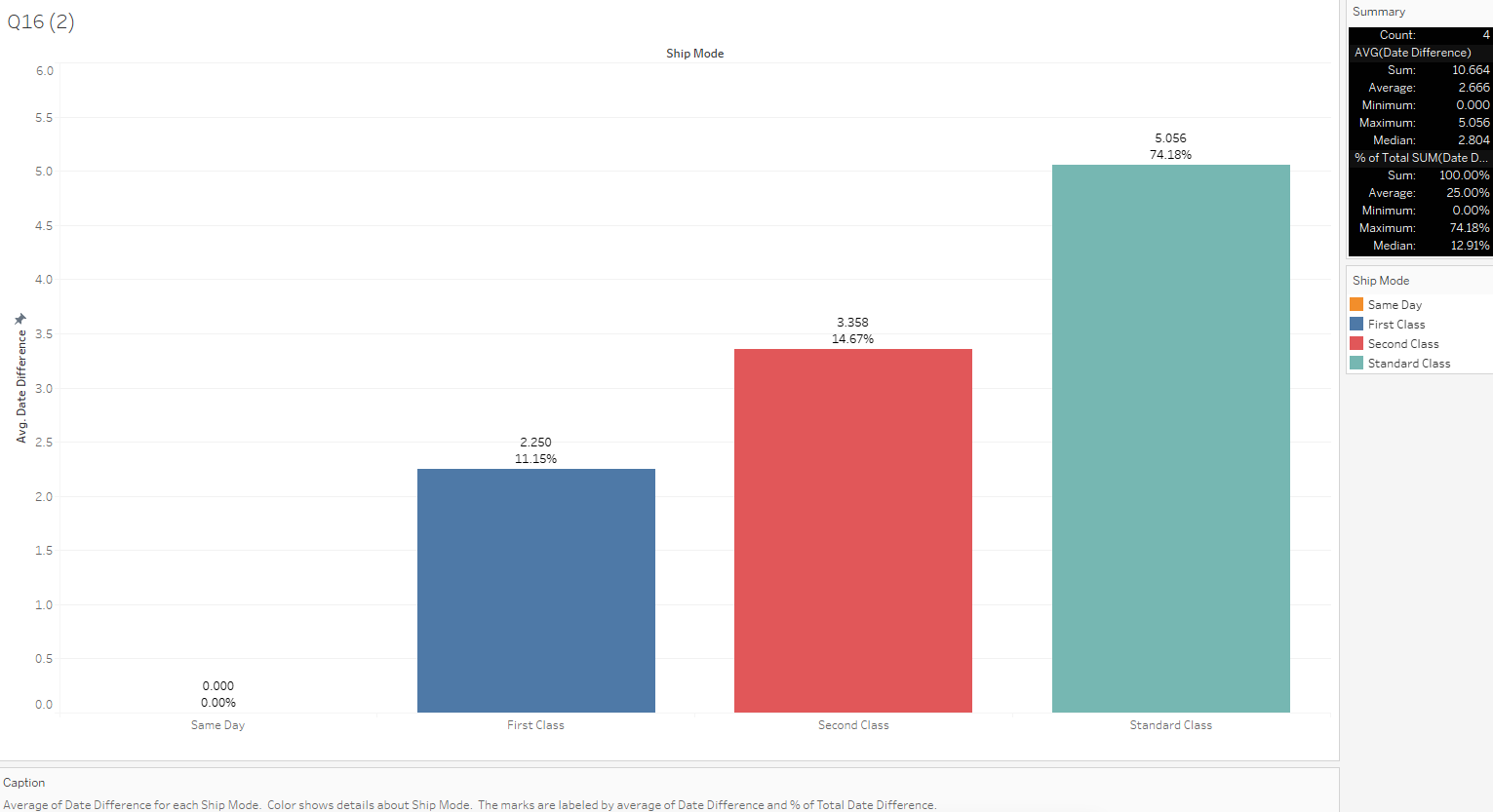
In the above graph, we can see how the profit distributions vary across different product categories. The highest profit is generated from the category ‘Technology’ with a share of 60.25% while the lowest is generated from the category ‘Furniture’ with only 10%. The category ‘Office Supplies’ sits in between the two with almost 30%.

The chart used here is a pie chart for visualisation as it’s effective when we want to represent parts of a whole and emphasise the proportion of each component relative to the total. It also allows us to get a good visual representation of distribution of percent as a whole.

We have added:

1. Legends on the right side to make it easier to understand the names of different categories which are also differentiated with separate colours for distinguishing easily.
2. Caption which shows brief interpretation and logic of all the components used and what it signifies.
3. Summary which gives a very important overview to the graph highlighting the maximum, minimum, average, etc. indicating the key pointers necessary for quick and effective data interpretation.
4. Different contrasty colours for each category for easy differentiation.
5. Labels including category name and profit percentage for quick, easy understanding and relation.

1. Can we compare the shipping time distributions for different shipping modes?



From the above graph, we can see the shopping time distributions for different shipping modes. We can see that the shipping mode ‘Same Day’ takes no time as it is being delivered on the same day but moving on to the other shipping modes, we see a difference. The shipping mode ‘First Class’ takes around 2 days on average to ship whereas the shipping mode ‘Second Class’ takes around 3 days. ‘Standard Class’ mode takes the most time of 5 days to ship covering 74.18 percent of the time as a whole whereas the shipping mode ‘Same Day’ covers nothing at 0%.

We used a bar graph for visualisation as it’s excellent for comparing values of different categories and displaying the distribution of data effectively with vertical bar graphs.

We have added:

1. Legends on the right side to make it easier to understand the names of different categories which are also differentiated with separate colours for distinguishing easily.
2. Caption which shows brief interpretation and logic of all the components used and what it signifies.
3. Summary which gives a very important overview to the graph highlighting the maximum, minimum, average, etc. indicating the key pointers necessary for quick and effective data interpretation.
4. Different colours for each shipping mode for easy differentiation.
5. Labels including type of shipping mode, percentage value and number of days for quick, easy understanding and relation.
6. What is the monthly trend in the number of orders shipped?

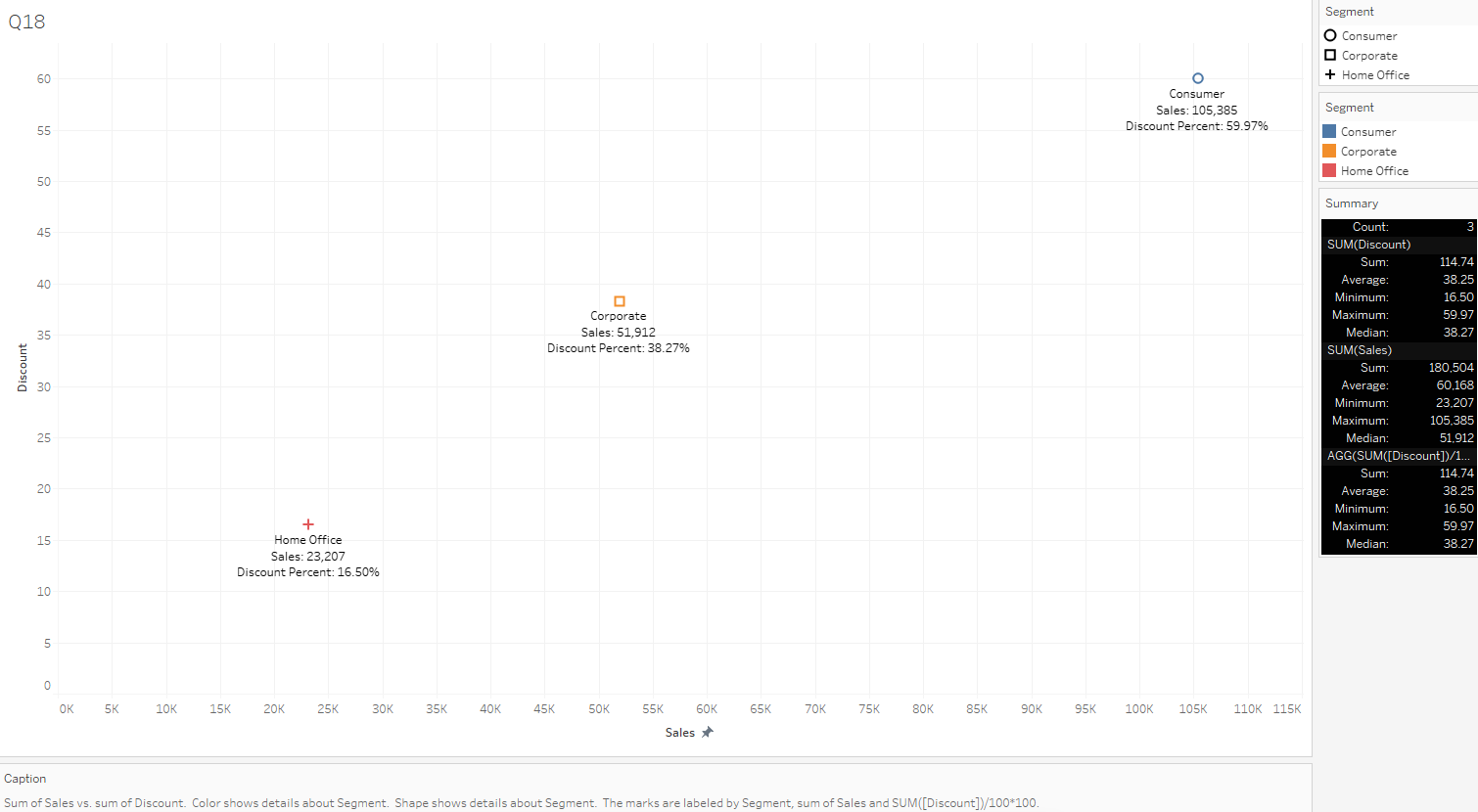


We can see from the above graph the monthly trend in the number of orders shipped. The highest number of orders shipped was in the month of September with a massive 29,485 orders shipped whereas the least was in the month of January with a number of only 4,050 orders shipped. The biggest drop which can be observed is when moving from the month of December to January i.e from 25,639 orders to 4,050 orders.

The graph used here is a Line Chart as it is one of the best charts for showing trends over time. They help identify patterns, cycles, and overall directional changes in the data.

We have added:

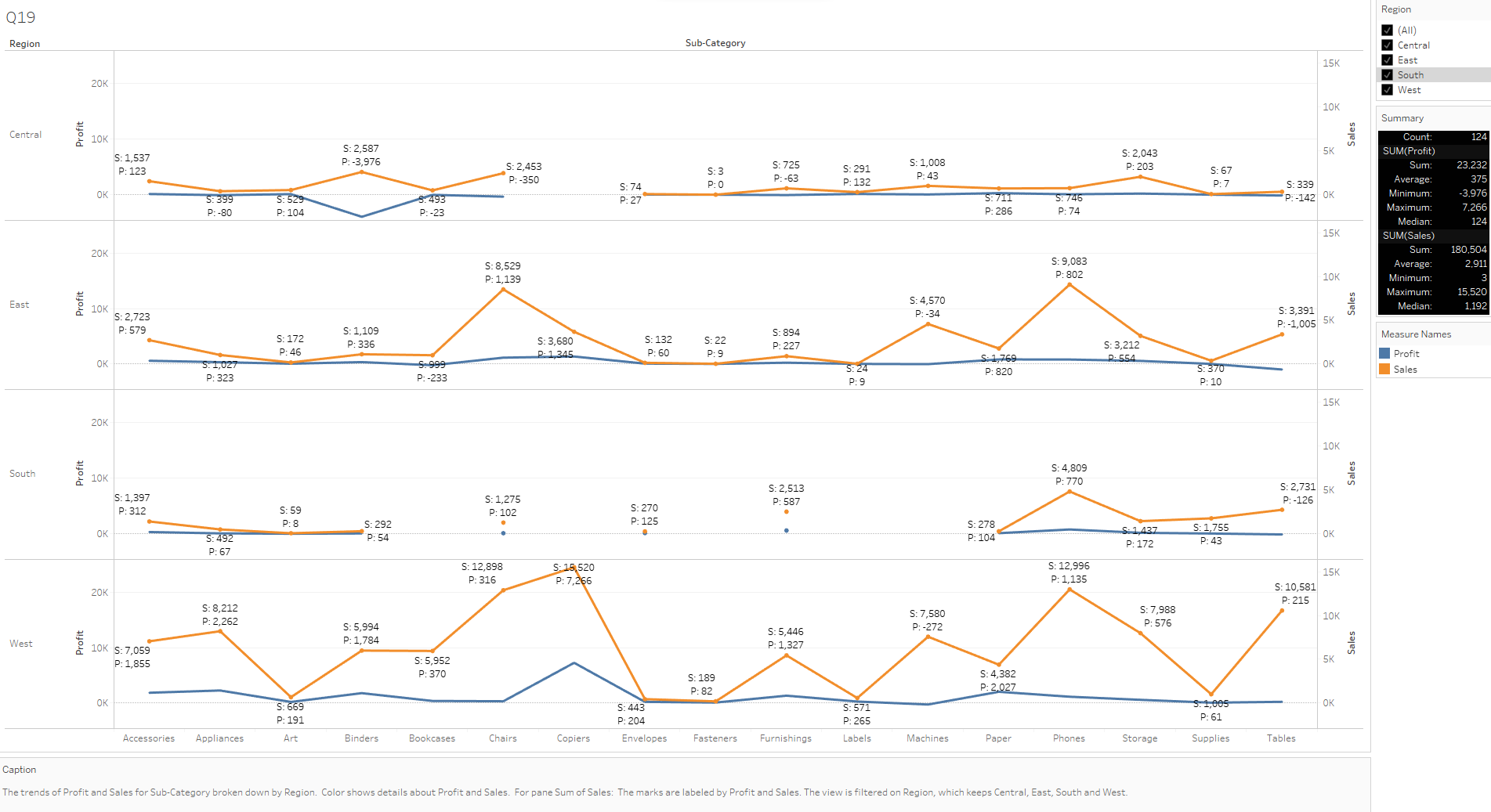
1. Caption which shows brief interpretation and logic of all the components used and what it signifies.
2. Summary which gives a very important overview to the graph highlighting the maximum, minimum, average, etc. indicating the key pointers necessary for quick and effective data interpretation.
3. High contrast colour palette contrast colours for easy differentiation.
4. Labels including sales figures for quick, easy understanding and relation.
5. How do different customer segments perform in terms of sales and discount rates?



From the above graph, we can see how different customer segments perform in terms of sales and discount rates. The highest sales and the highest discount percent can be seen in the ‘Consumer’ segment while the lowest sales and discount percent can be seen in the ‘Home Office’ segment. This shows how greatly discount affects the overall sales of each segment.

Once again, we have used a scatter plot for plotting these points as plotting one variable on the x-axis and another on the y-axis helps visualise the trend and direction of the relationship. They also provide a spread visual representation of how data points are distributed. Clustering or spread of points can give insights into the density and variability of the data. We have added:

1. Custom labels for a deeper insight, better understanding of all the data handles.
2. Legends on the right side to make it easier to understand the names of different categories which are also differentiated with separate colours for distinguishing easily.
3. Caption which shows brief interpretation and logic of all the components used and what it signifies.
4. Summary which gives a very important overview to the graph highlighting the maximum, minimum, average, etc. indicating the key pointers necessary for quick and effective data interpretation.
5. Different colours for each segment for easy differentiation.
6. What are the sales and profit trends across different product subcategories and regions in the Superstore dataset?



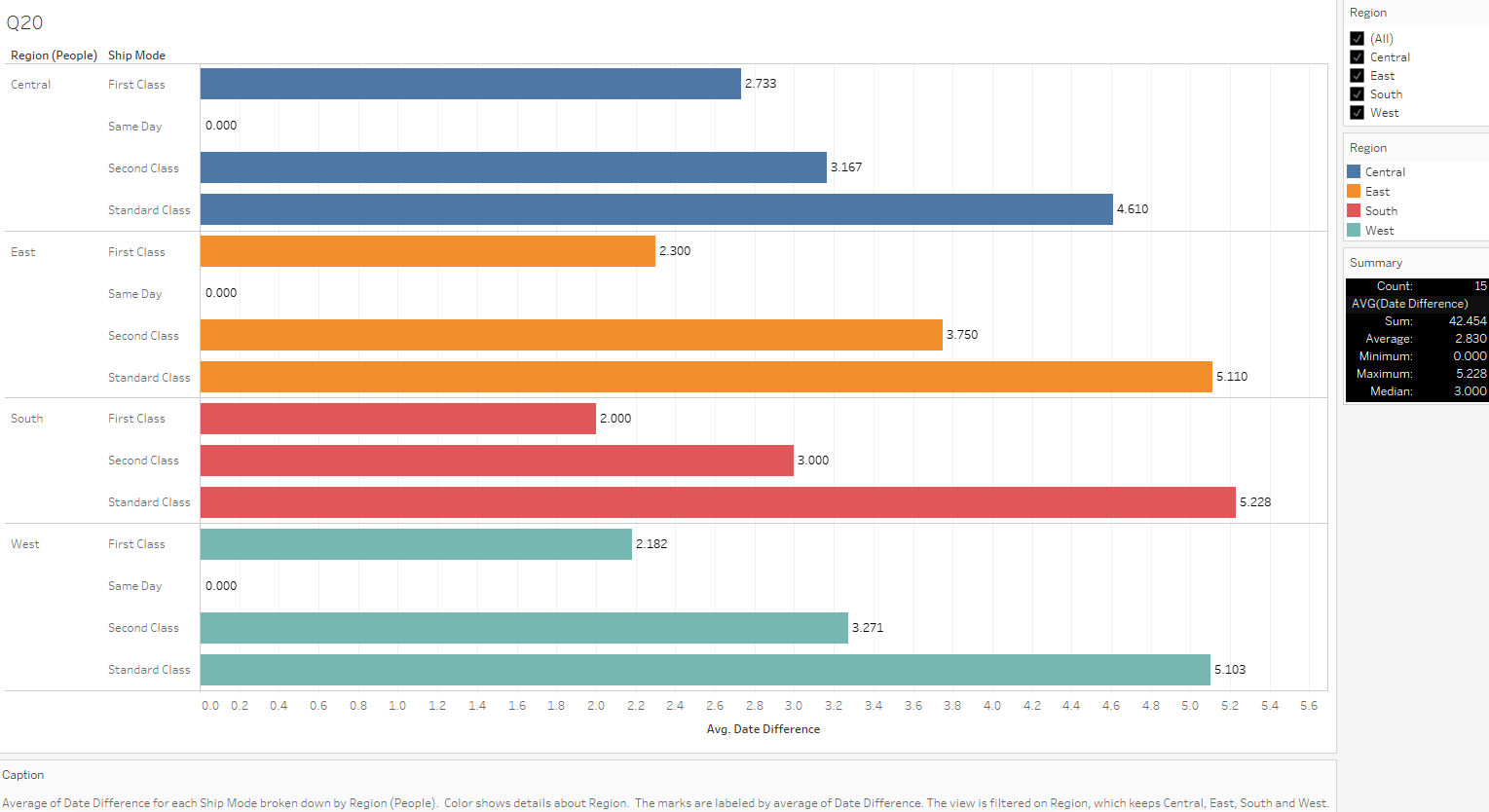
The above graph shows the sales and profit trends across different product subcategories and regions. The data with ‘S’ shows the sales figure while the data with ‘P’ shows the profit figure. The regions are shown in the y-axis whereas the sub-categories are shown in the x-axis. The profit is quite plain in the ‘Central’ region whereas the most increases in profit can be seen in the ‘West’ region. The same can be said for the sales as well as the highest sales can be seen in the ‘West’ region. The highest sales when including every region is from the ‘Phones’ sub-category.

The graph used here is a Multi-Line Chart as it allows us to display and compare the trends of multiple entities or categories simultaneously. This is particularly useful when we want to observe how different variables evolve over time or another sequential dimension. With multiple lines on the same chart, it becomes easier to identify patterns, trends, and anomalies in the data. We can also quickly spot similarities or differences in how various entities behave.

We have added:

1. Custom labels for a deeper insight, better understanding of all the data handles.
2. Filters on the right side to make it easier to filter the different regions individually. Profit and Sales are also differentiated with separate colours for distinguishing easily.
3. Caption which shows brief interpretation and logic of all the components used and what it signifies.
4. Summary which gives a very important overview to the graph highlighting the maximum, minimum, average, etc. indicating the key pointers necessary for quick and effective data interpretation.
5. Different colours for easy differentiation.

1. What is the average delivery duration for different regions and ship modes?

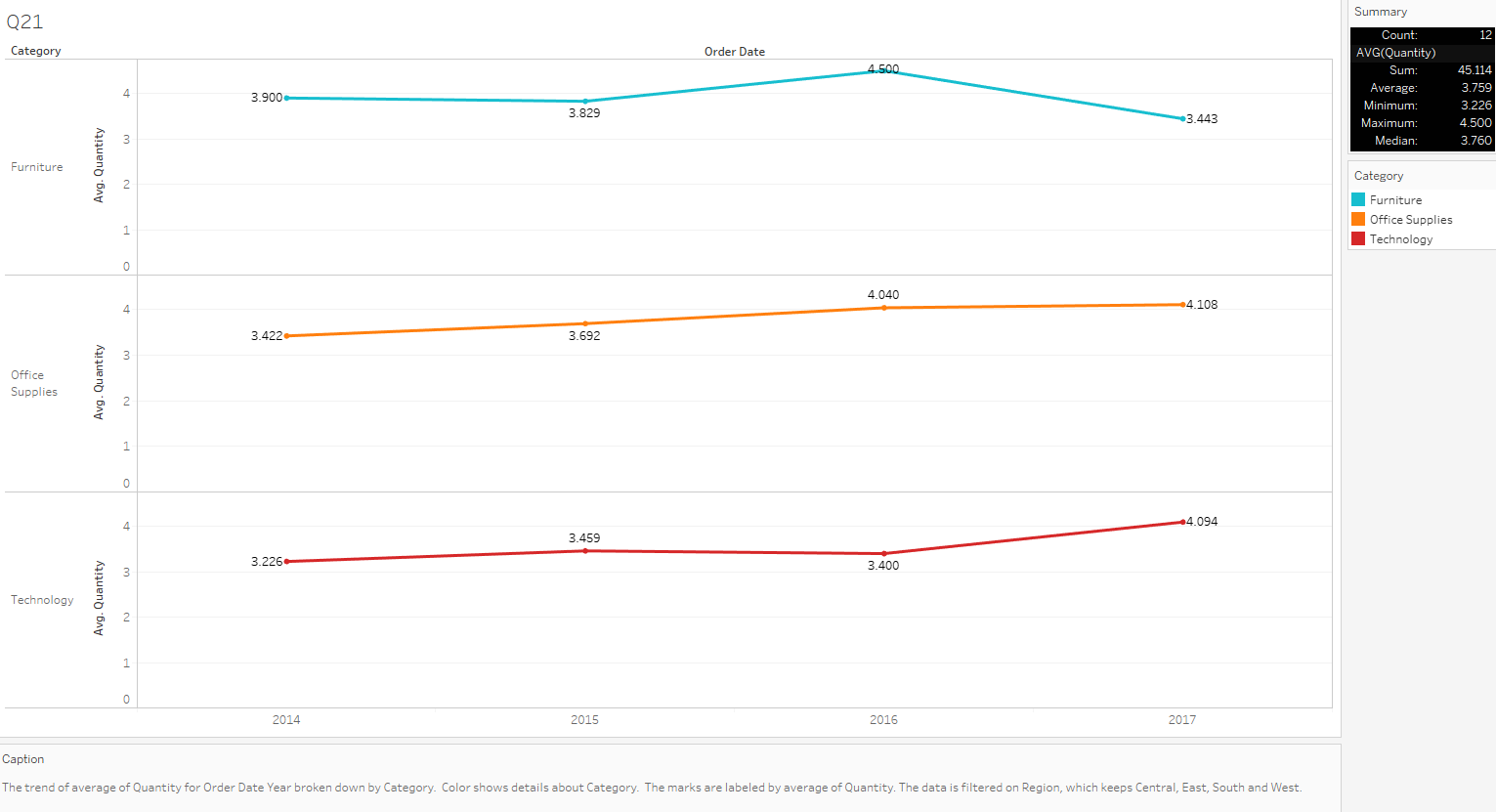


The above graph shows the average delivery duration for different regions and ship modes. We can see the most delivery time taken in every region is for the ‘Standard Class’ ship mode with the highest being recorded at 5.2 days average in the ‘South’ region. The lowest being for ‘Same Day’ ship mode with an average of 0 days in every region.

The graph used here is the multi-axis horizontal bar graph for visualisation as it’s excellent for comparing values of different regions and displaying the distribution of data effectively with horizontal bar graphs.

We have added:

1. Legends on the right side to make it easier to understand the names of different categories which are also differentiated with separate colours for distinguishing easily.
2. Caption which shows brief interpretation and logic of all the components used and what it signifies.
3. Summary which gives a very important overview to the graph highlighting the maximum, minimum, average, etc. indicating the key pointers necessary for quick and effective data interpretation.
4. Different colours for each region for easy differentiation.
5. Label showing day average for quick, easy understanding and relation.
6. How has the average order quantity changed over the years for various product categories?

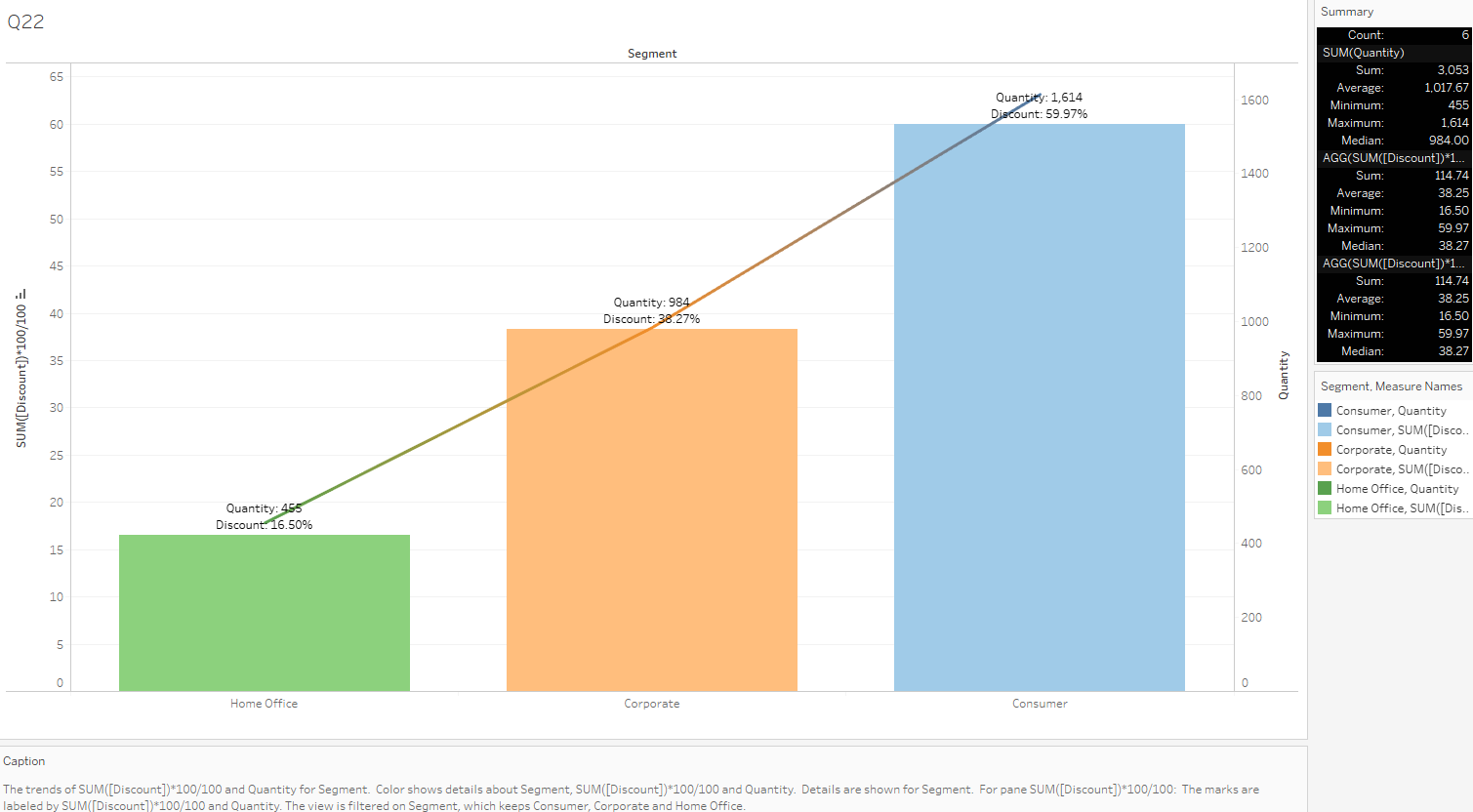


The above graph shows the average order quantity changed over the years for various product categories. The highest average quantity was in 2016 for the category ‘Furniture’ whereas for ‘Office Supplies’ it was in the year 2017. The category ‘Technology’ recorded its highest average in the year 2017 as well.

The graph used here is a Multi-Line Chart as it allows us to display and compare the trends of multiple categories simultaneously. This is particularly useful when we want to observe how different variables evolve over time or another sequential dimension. With multiple lines on the same chart, it becomes easier to identify patterns, trends, and anomalies in the data. We can also quickly spot similarities or differences in how various entities behave.

We have added:

1. Label for average quantities for a deeper insight, better understanding of all the data handles.
2. Caption which shows brief interpretation and logic of all the components used and what it signifies.
3. Summary which gives a very important overview to the graph highlighting the maximum, minimum, average, etc. indicating the key pointers necessary for quick and effective data interpretation.
4. Different colours for each category for easy differentiation.
5. Can we visualise the correlation between discount rates and order quantities for different customer segments?

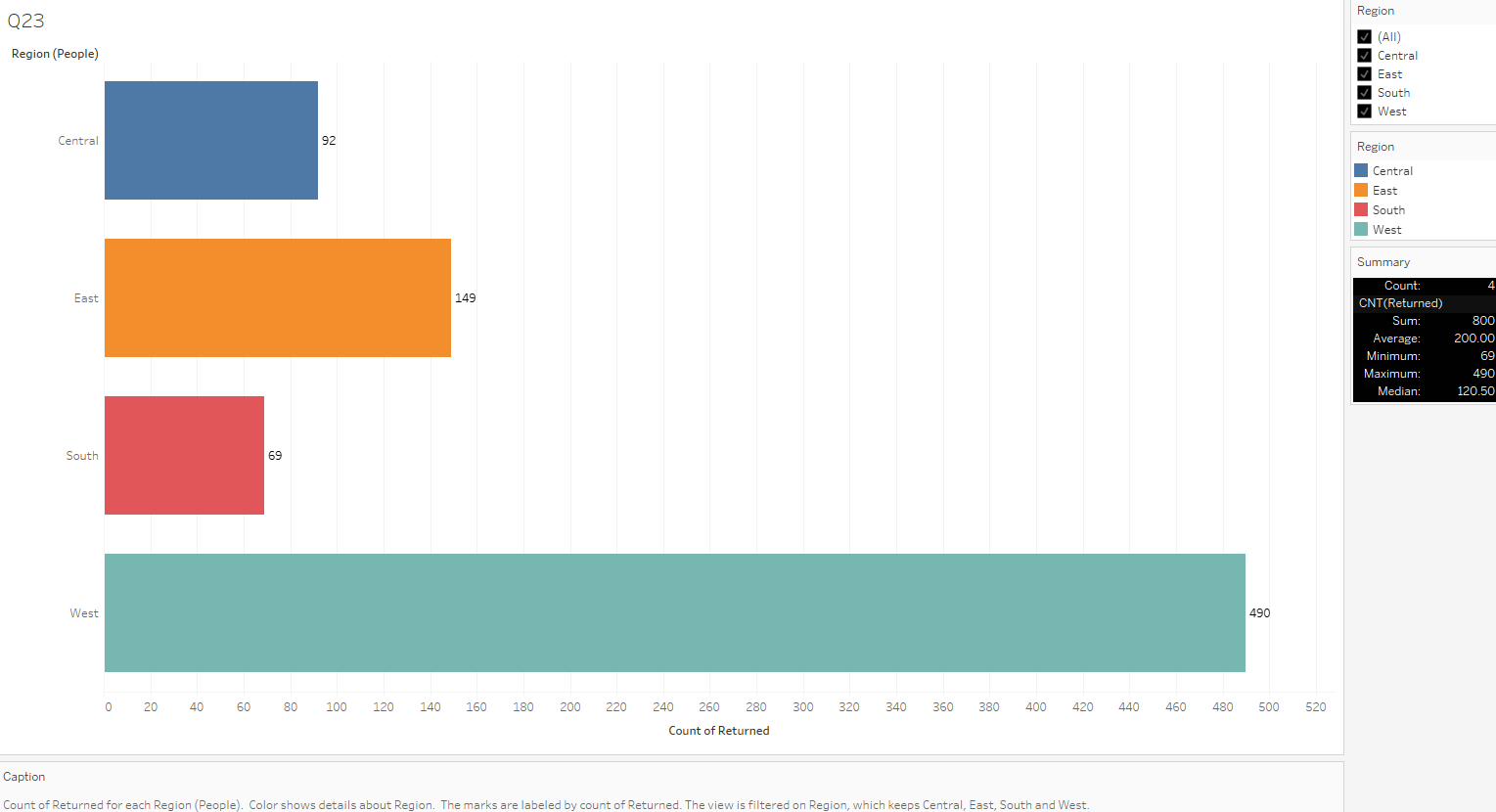


The above graph shows the correlation between discount rates and order quantities for different customer segments. The quantity is shown by the line chart whereas the discount is shown through the vertical bars. We can see an increasing ascent in the amount of quantities as the discount increases. The highest increase was seen in the ‘Consumer’ category with a discount of almost 60% and a quantity of 1,614 units sold whereas the least was for the ‘Home Office’ category with a discount of 16% and a quantity of 455 units sold.

The graph used here is a combination of bar chart and line chart for easy differentiation between quantities and discount as they are sharing the same axis.

We have added:

1. Custom labels for quantities and discount percentage for a deeper insight, better understanding of all the data handles.
2. Caption which shows brief interpretation and logic of all the components used and what it signifies.
3. Summary which gives a very important overview to the graph highlighting the maximum, minimum, average, etc. indicating the key pointers necessary for quick and effective data interpretation.
4. Different colours for each category for easy differentiation.
5. What is the proportion of orders returned in each region within the Superstore dataset?



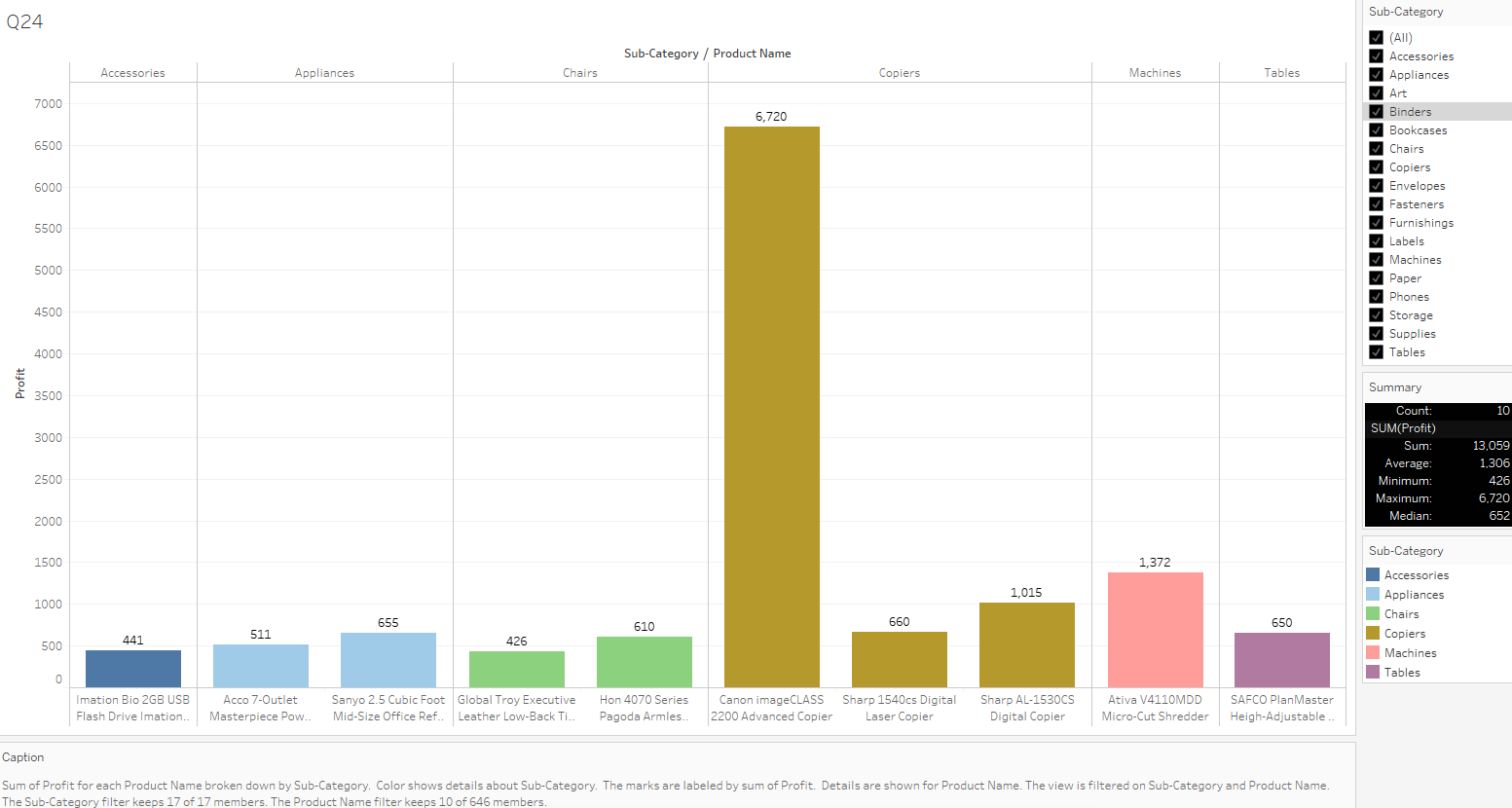
The above graph shows the proportion of orders returned in each region. We can see the highest number of orders returned is in the ‘West’ region at 490 units while the least number of orders returned is in the ‘South’ region at only 69 units.

We used a horizontal bar graph for visualisation as it’s excellent for comparing values of different regions and displaying the distribution of data effectively with vertical bar graphs.

We have added:

1. Legends on the right side to make it easier to understand the names of different categories which are also differentiated with separate colours for distinguishing easily.
2. Caption which shows brief interpretation and logic of all the components used and what it signifies.
3. Summary which gives a very important overview to the graph highlighting the maximum, minimum, average, etc. indicating the key pointers necessary for quick and effective data interpretation.
4. Different colours for each shipping mode for easy differentiation.
5. Label for number of returned orders for quick, easy understanding and relation.

1. Can you compare the profit of different products for different subcategories?



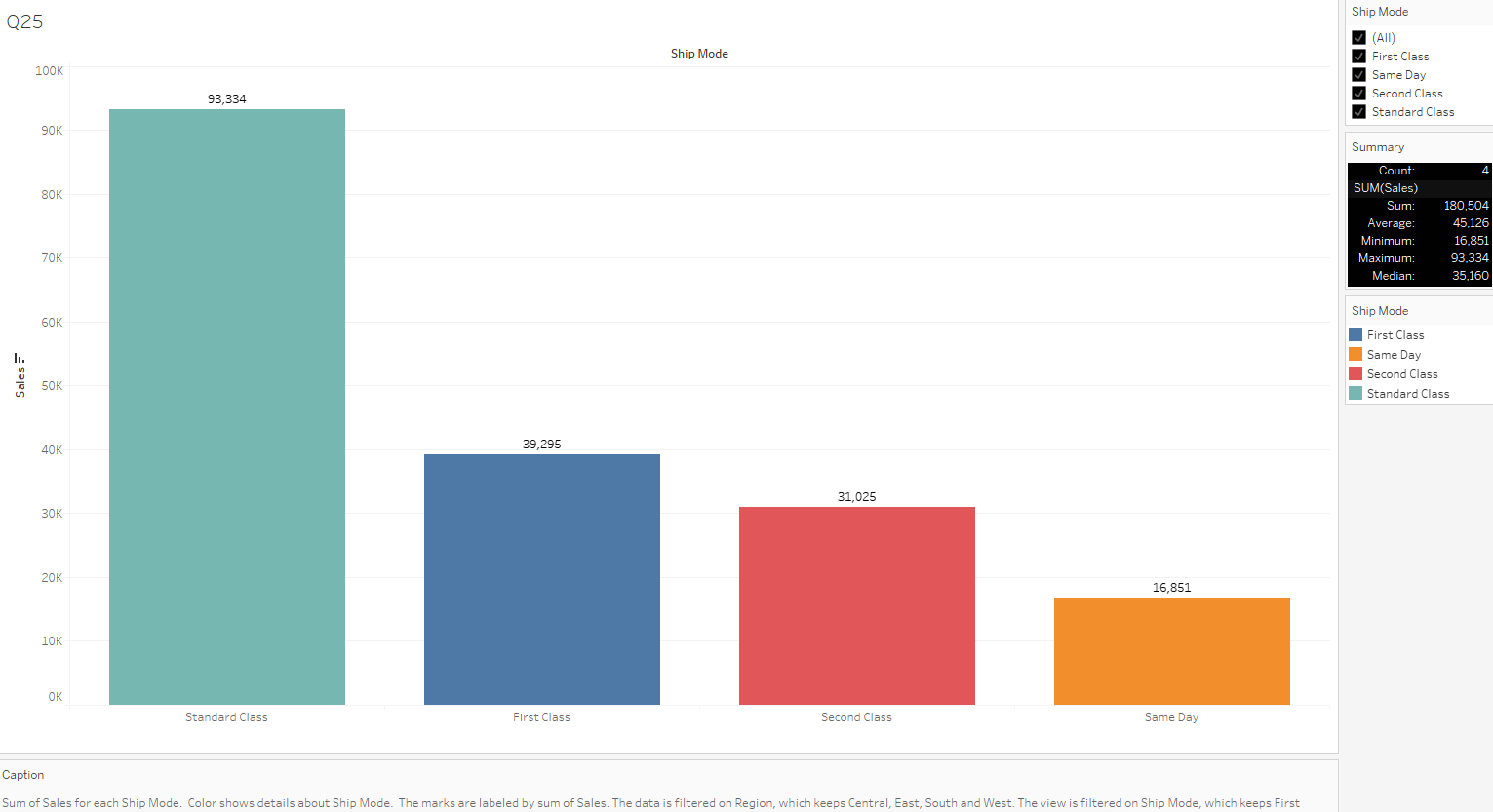
The above graph shows the profit of different products for different subcategories. We are showing only the top-ten profited products from each sub-category because there are a total of 646 products.

The highest of all is from the ‘Copiers’ sub-category at a profit of $6,720.

We used a bar graph for visualisation as it’s excellent for comparing values of different categories and displaying the distribution of data effectively with vertical bar graphs.

We have added:

1. Legends on the right side to make it easier to understand the names of different categories which are also differentiated with separate colours for distinguishing easily.
2. Filters for easily viewing each or a group of sub-categories.
3. Caption which shows brief interpretation and logic of all the components used and what it signifies.
4. Summary which gives a very important overview to the graph highlighting the maximum, minimum, average, etc. indicating the key pointers necessary for quick and effective data interpretation.
5. Different colours for each sub-category for easy differentiation.
6. Label showing the profit for quick, easy understanding and relation.
7. Which shipping mode is the most commonly used in the Sample Superstore dataset?

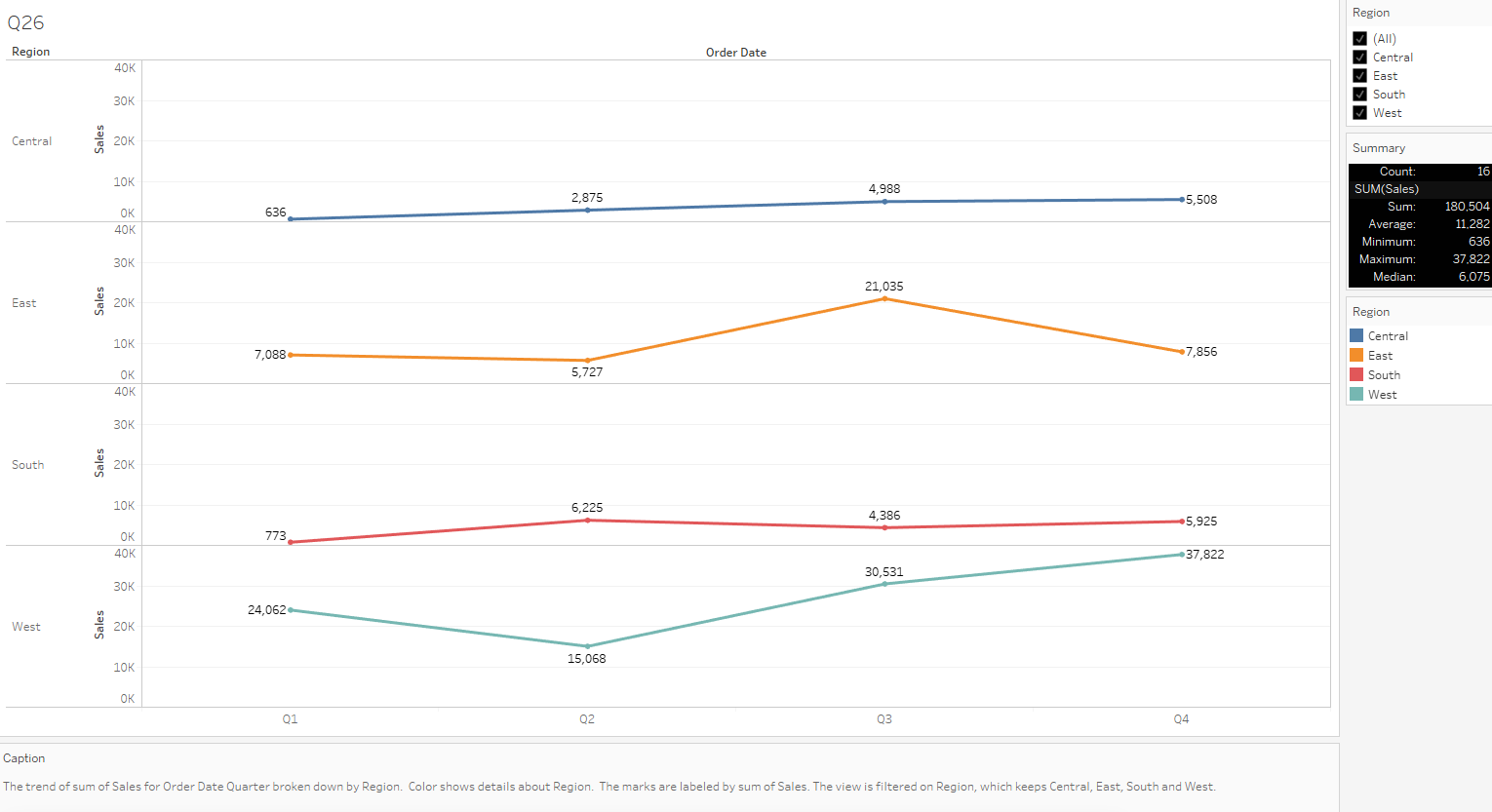


The above graph shows the most commonly used shipping mode. We can see the maximum sales happened for ‘Standard Class’ with over 93k sales whereas the lowest was for ‘Same Day’ shipping mode with only 16,851 sales while it takes the least amount of time.

The graph used here for visualisation is a Bar Graph as it’s excellent for comparing values of different categories and displaying the distribution of data effectively with vertical bar graphs.

We have added:

1. Legends on the right side to make it easier to understand the names of different categories which are also differentiated with separate colours for distinguishing easily.
2. Filters for easily viewing each or a group of shipping modes.
3. Caption which shows brief interpretation and logic of all the components used and what it signifies.
4. Summary which gives a very important overview to the graph highlighting the maximum, minimum, average, etc. indicating the key pointers necessary for quick and effective data interpretation.
5. Different colours for each shipping mode for easy differentiation.
6. Label showing the total sales for each shipping mode for quick, easy understanding and relation.
7. How does the sales performance of different regions evolve throughout the quarters of a year?

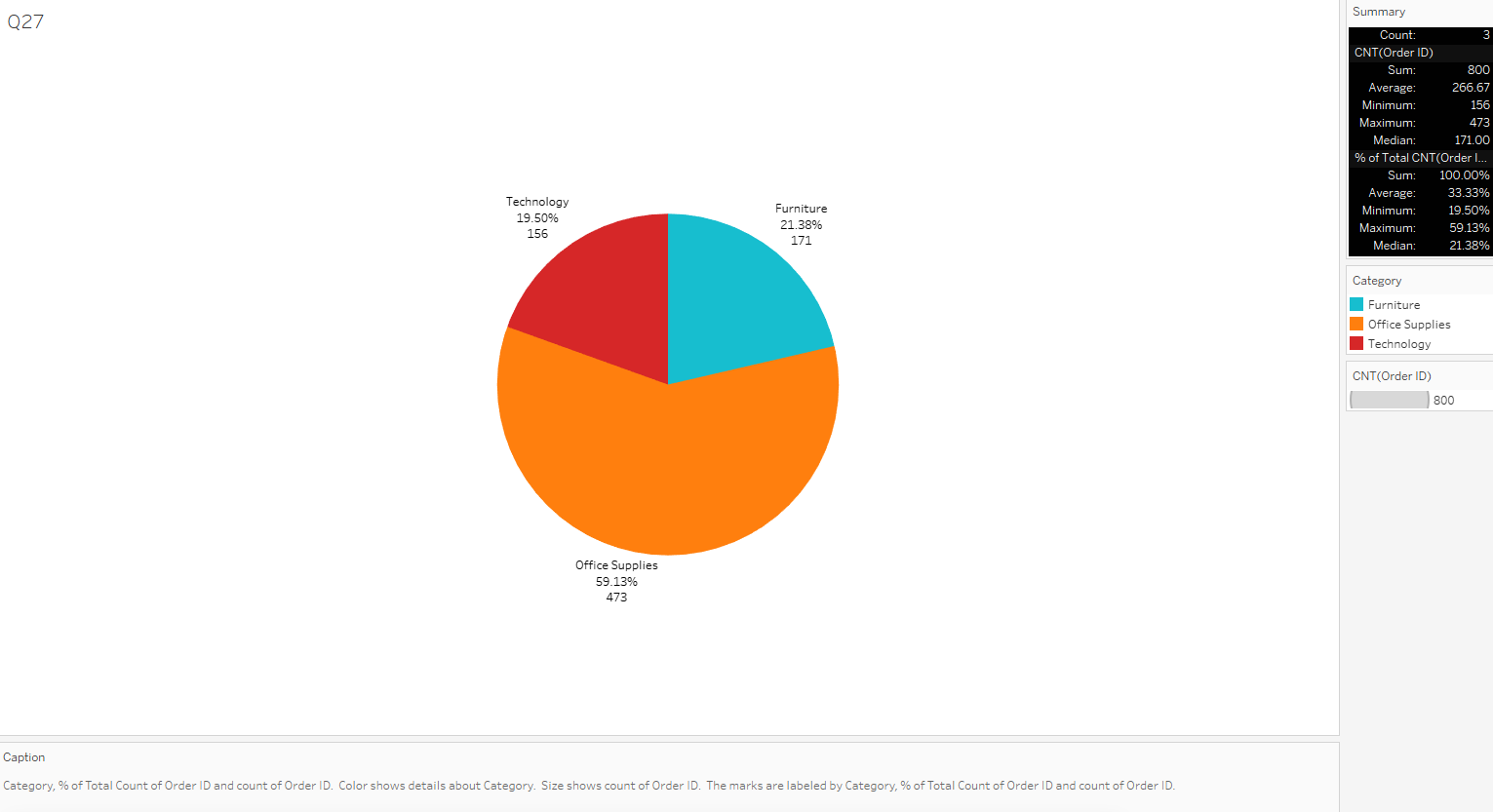


The above graph shows how the sales performance of different regions evolve throughout the quarters of a year. We can see that the ‘Central’ region has a steady and stable incline throughout all the quarters of the year whereas the ‘East’ region faces an incline in the third quarter and a decline afterwards in the 4th quarter. The ‘South’ region maintains a steady line while the ‘West’ region starts with a decline but covers it up with a steady and increasing incline in the 3rd and 4th quarter. The highest sales at the end of the year was in the ‘West’ region with over 37k sales while the lowest was in the ‘Central’ region at only 5,508 sales.

The graph used here is a Multi-Line Chart as it allows us to display and compare the trends of multiple regions simultaneously. This is particularly useful when we want to observe how different variables evolve over time or another sequential dimension. With multiple lines on the same chart, it becomes easier to identify patterns, trends, and anomalies in the data. We can also quickly spot similarities or differences in how various entities behave.

We have added:

1. Label for total sales in each quarter for a deeper insight, better understanding of all the data handles.
2. Filter for easily viewing each or a group of regions.
3. Caption which shows brief interpretation and logic of all the components used and what it signifies.
4. Summary which gives a very important overview to the graph highlighting the maximum, minimum, average, etc. indicating the key pointers necessary for quick and effective data interpretation.
5. Different colours for each region for easy differentiation.
6. What is the distribution of order priorities across different product categories?

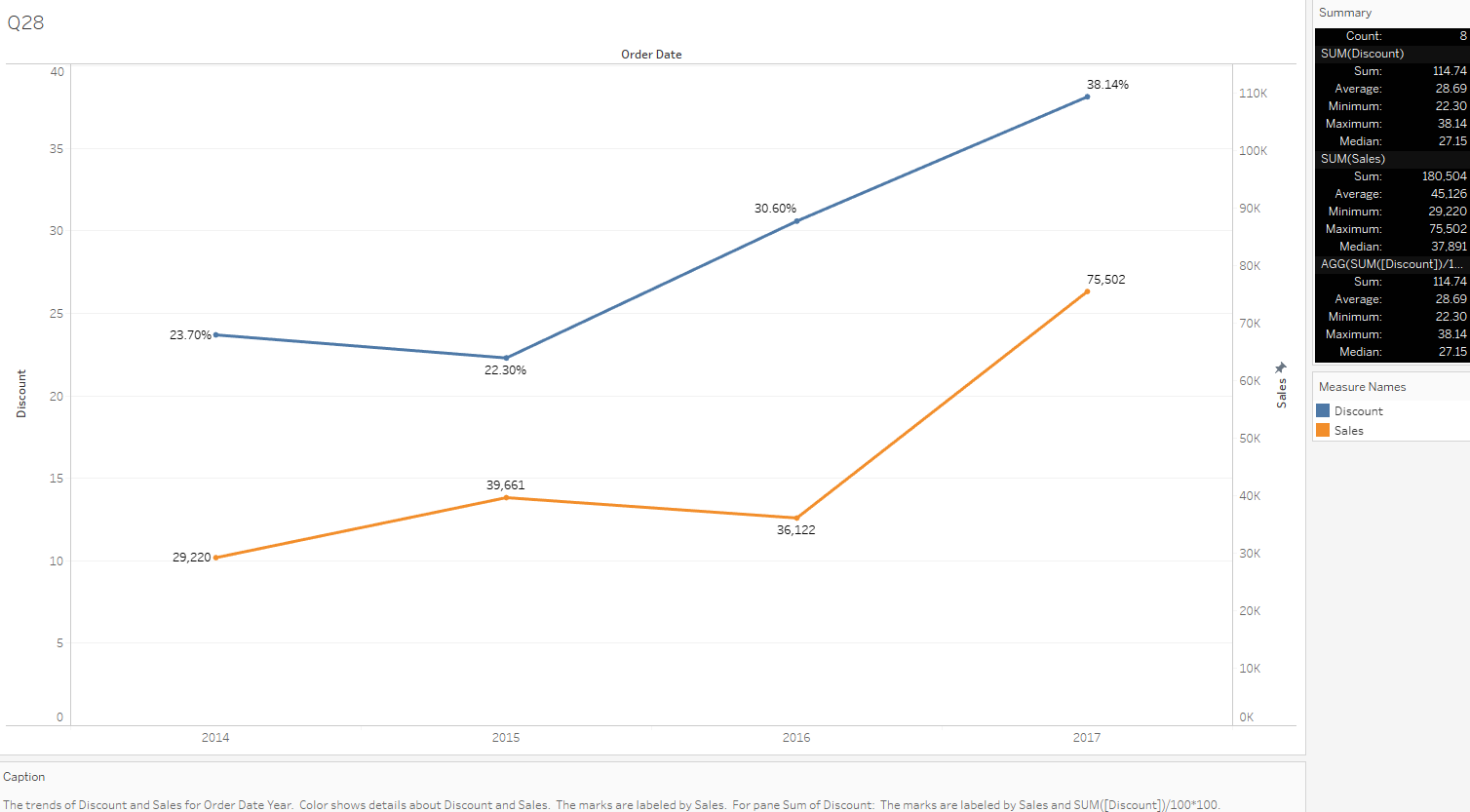


The above graph shows the distribution of order priorities across different product categories. We can see that the highest order and the biggest proportion is for ‘Office Supplies’ with 473 orders and over 59% share while the lowest proportion is for ‘Technology’ with 156 orders and over 19.50% share.

The chart used here is a pie chart for visualisation as it’s effective when we want to represent parts of a whole and emphasise the proportion of each component relative to the total. It also allows us to get a good visual representation of the distribution of percent as a whole.

We have added:

1. Legends on the right side to make it easier to understand the names of different categories which are also differentiated with separate colours for distinguishing easily.
2. Caption which shows brief interpretation and logic of all the components used and what it signifies.
3. Summary which gives a very important overview to the graph highlighting the maximum, minimum, average, etc. indicating the key pointers necessary for quick and effective data interpretation.
4. Different contrasty colours for each category for easy differentiation.
5. Labels including category name, order count and profit percentage for quick, easy understanding and relation.
6. What is the relationship between discounts and sales?

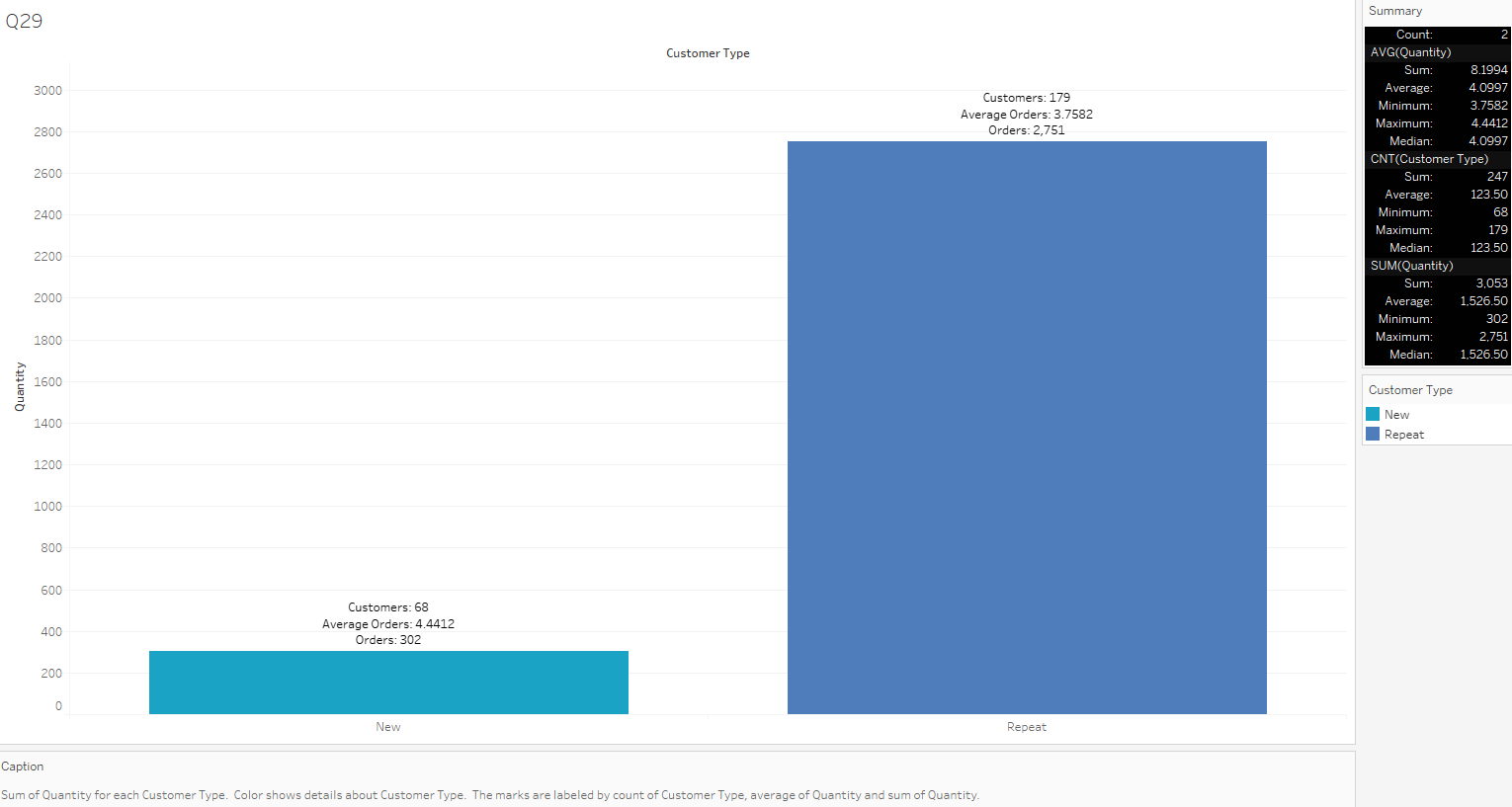


The above graph shows the relationship between discounts and sales. We can see that the common trend is:- the sales increase as the discounts increase meaning they have a proportional relationship in general if we exclude the year 2016. The sales reach its highest of over 75k when offered up-to 38% discount. The year 2016 seeing a downfall in sales shows the necessary increase in discount percent in the year 2017 which saw a staggering increase in sales.

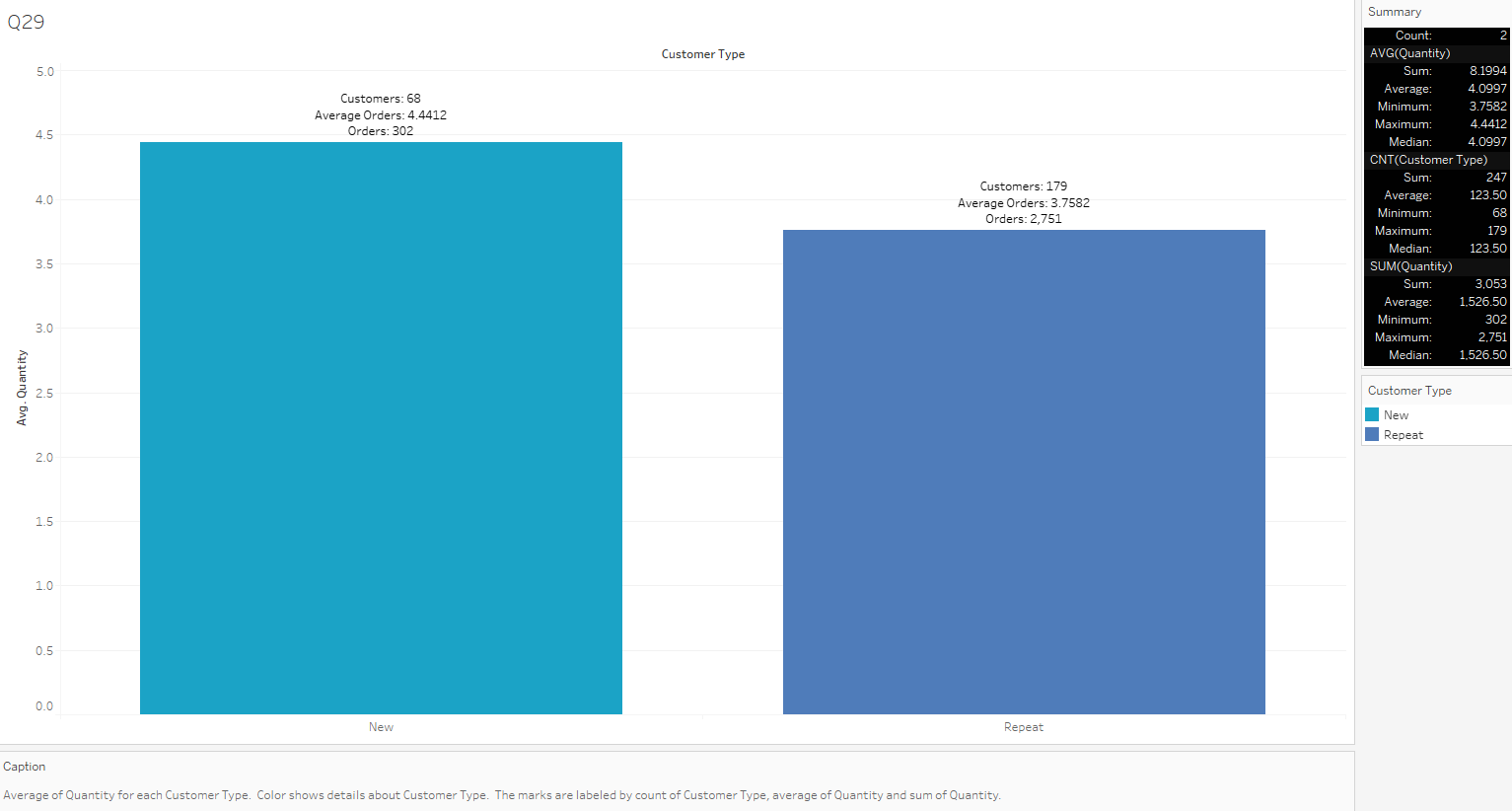
The graph used here is a Multi-Line Chart as it allows us to display and compare the trends of multiple regions simultaneously. This is particularly useful when we want to observe how different variables evolve over time or another sequential dimension. With multiple lines on the same chart, it becomes easier to identify patterns, trends, and anomalies in the data. We can also quickly spot similarities or differences in how various entities behave.

We have added:

1. Label for total sales and discount percent in each year for a deeper insight, better understanding of all the data handles.
2. Caption which shows brief interpretation and logic of all the components used and what it signifies.
3. Summary which gives a very important overview to the graph highlighting the maximum, minimum, average, etc. indicating the key pointers necessary for quick and effective data interpretation.
4. Different colours for both sales and discounts for easy differentiation.
5. How does the average order value differ between repeat customers and new customers?



The above graph shows how the average order value differs between repeat customers and new customers. We created a calculated field for differentiating new customers from repeat by checking whether they had orders on more than one date. If yes, then they are ‘Repeat Customers’ else they’re ‘New’. From the chart, we can see that ‘New’ customers account for only 68 while ‘Repeat’ customers account for 179 in quantity. Although the average order rate is higher for New customers when compared to the ratio of customers, the total quantity remains higher for ‘Repeat’ customers.



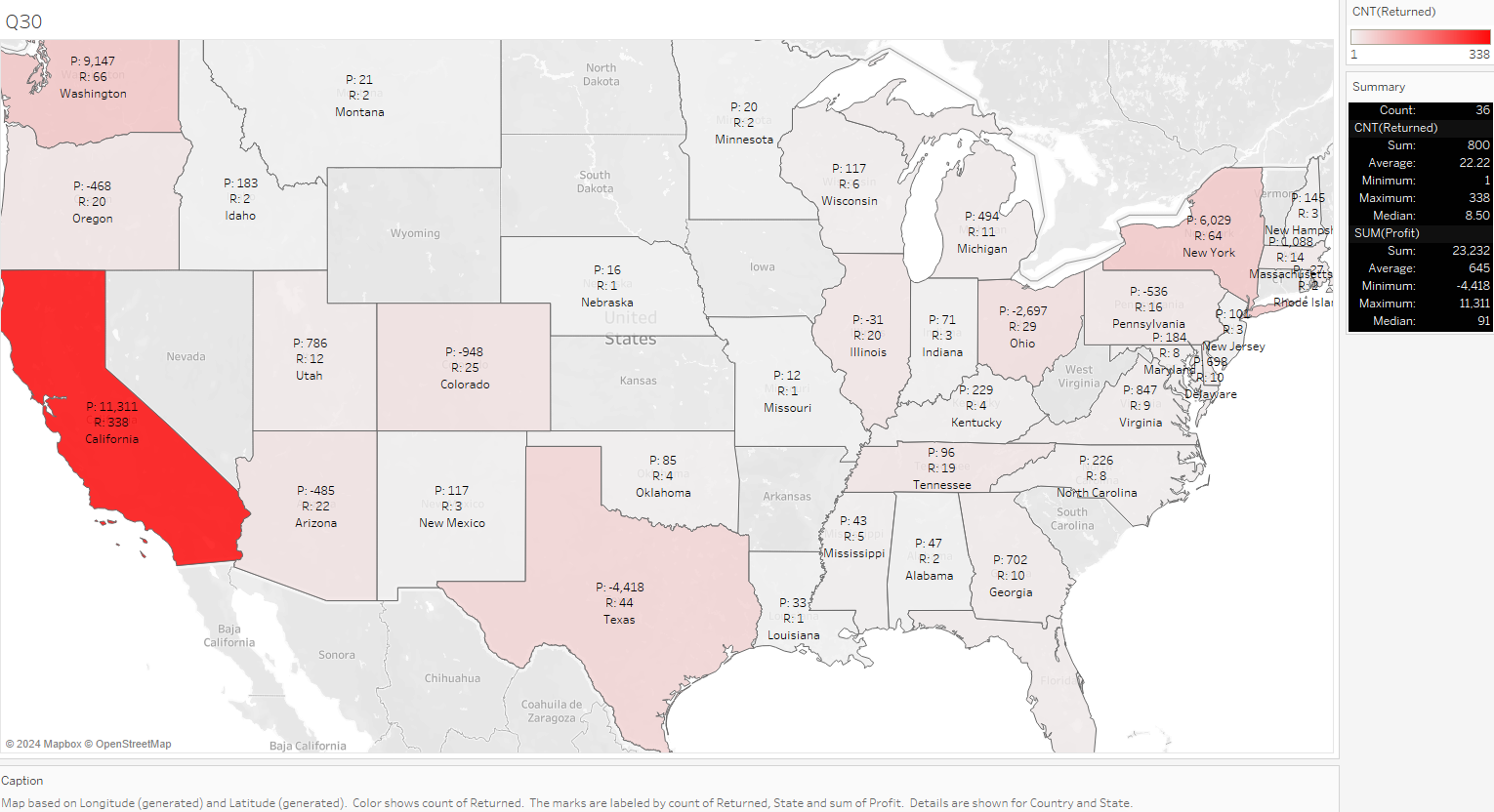
The above graph shows the ratio of orders to the number of customers where we can see the average order rate is higher for ‘New’ customers when compared to ‘Repeat’ customers even though their amount of orders is higher.

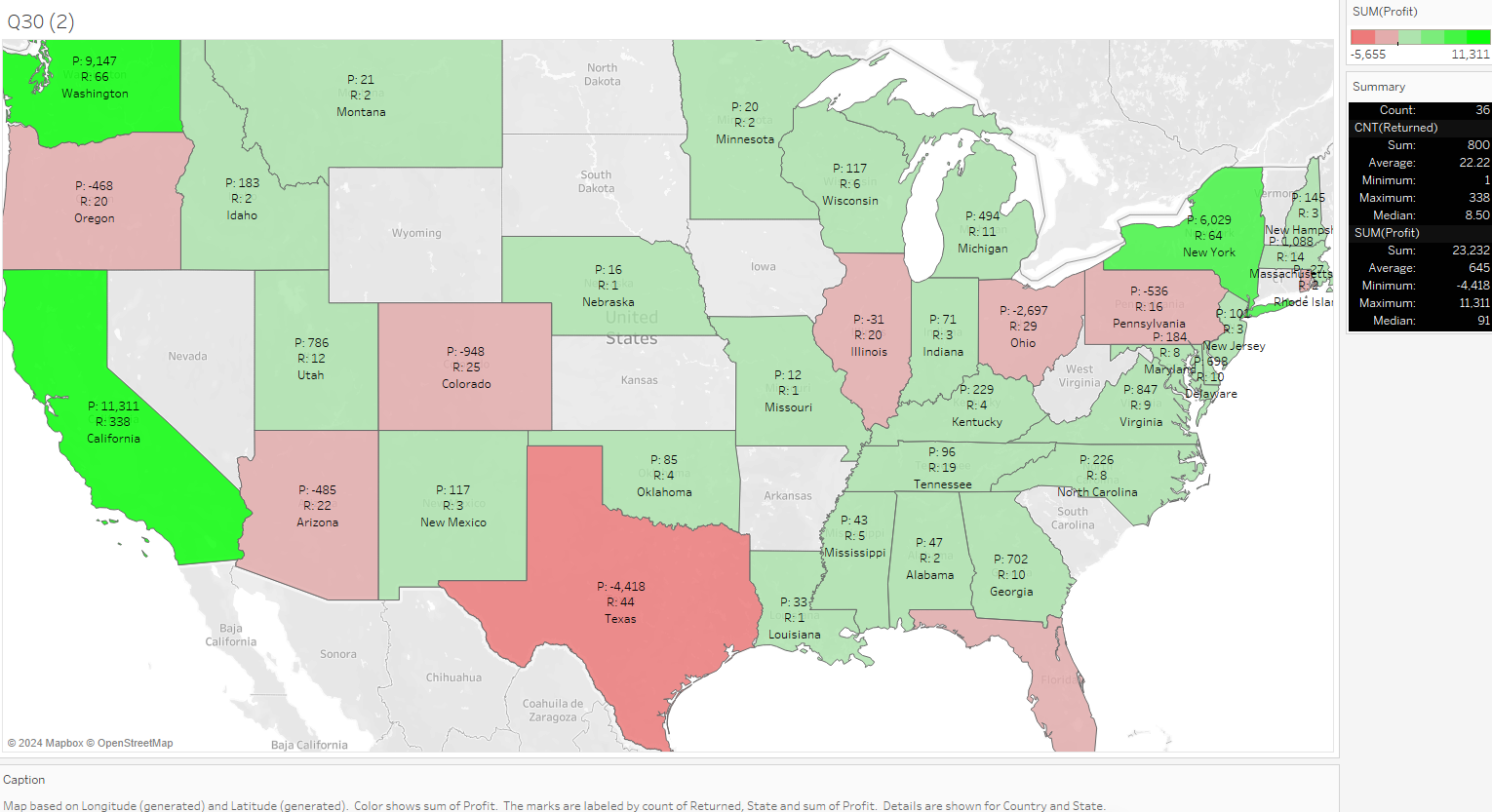
The graph used here for visualisation is a Bar Graph as it’s excellent for comparing values of different categories and displaying the distribution of data effectively with vertical bar graphs.

We have added:

1. Legends on the right side to make it easier to understand the names of different categories which are also differentiated with separate colours for distinguishing easily.
2. Caption which shows brief interpretation and logic of all the components used and what it signifies.
3. Summary which gives a very important overview to the graph highlighting the maximum, minimum, average, etc. indicating the key pointers necessary for quick and effective data interpretation.
4. Different colours for each customer type for easy differentiation.
5. Labels showing the number of customers, average and number of orders for quick, easy understanding and relation.

1. What is the geographical distribution of returns and its impact on overall profitability?





The above two graphs show the geographical distribution of returns and its overall impact on profitability. In the first graph, we can see which states had the most returns with the help of the red accent. It shows the highest returns being recorded in the state ‘California’ while the lowest being in the state ‘Missouri’. The second graph shows how it affects the overall profits with the help of red-green gradient. We observe that even though the state ‘California’ had the highest amount of returns, it is doing well in profit because of the high number of sales in that state. The rest of states having poor sales get affected with even the slightest of returns.

We have marked the amount of profit with the symbol ‘P’ and the number of returns with the symbol ‘R’. We have used short-forms to avoid clutter in data.

We used the symbol map graph here as it represents data points using custom symbols or markers on a geographic map. It is particularly useful for visualising and analysing location-based data.It also helps us visualise the geographic distribution of sales, profit, returns and other relevant data across different regions, countries and cities.

We have added:

1. Legends on the right side to make it easier to understand the amount of orders returned which are differentiated with separate colours for distinguishing easily.
2. Caption which shows brief interpretation and logic of all the components used and what it signifies.
3. Summary which gives a very important overview to the graph highlighting the maximum, minimum, average, etc. indicating the key pointers necessary for quick and effective data interpretation.
4. Gradient colour panel for easy differentiation between different states.
5. Custom labels showing the number of order returns, profit and state name for quick, easy understanding and relation.