MID COURSE ASSESSMENT

DATA VISUALIZATION AND ANALYSIS PROJECT

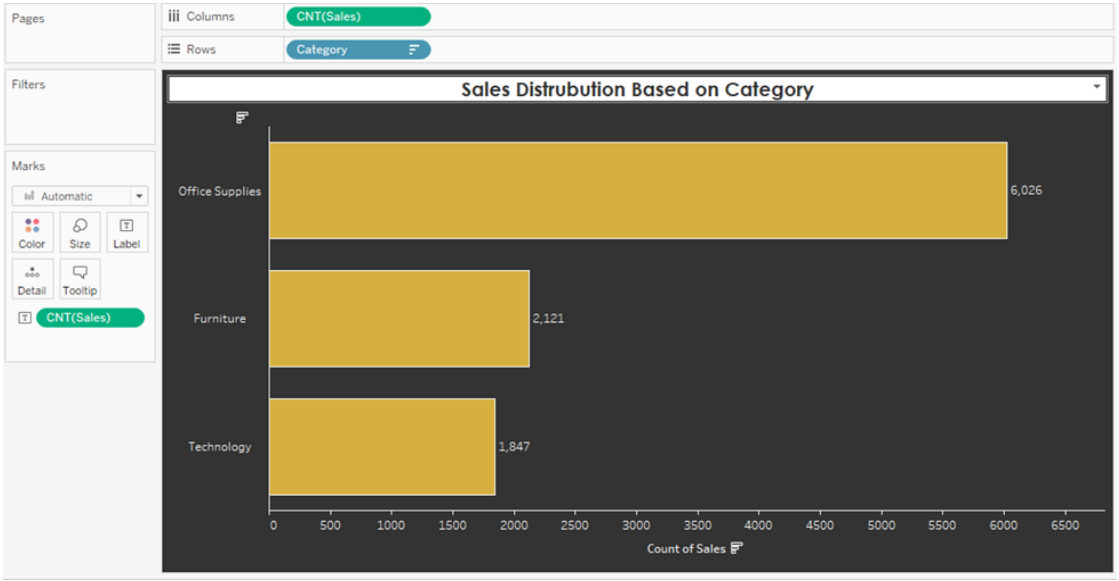
DATASET - SUPERSTORE

*NOTE: Based on certain questions criteria, have added a few columns to justify the ask of few of the questions like:

- Order Process Days
- Repeat/New Customer
- Avg Order Value

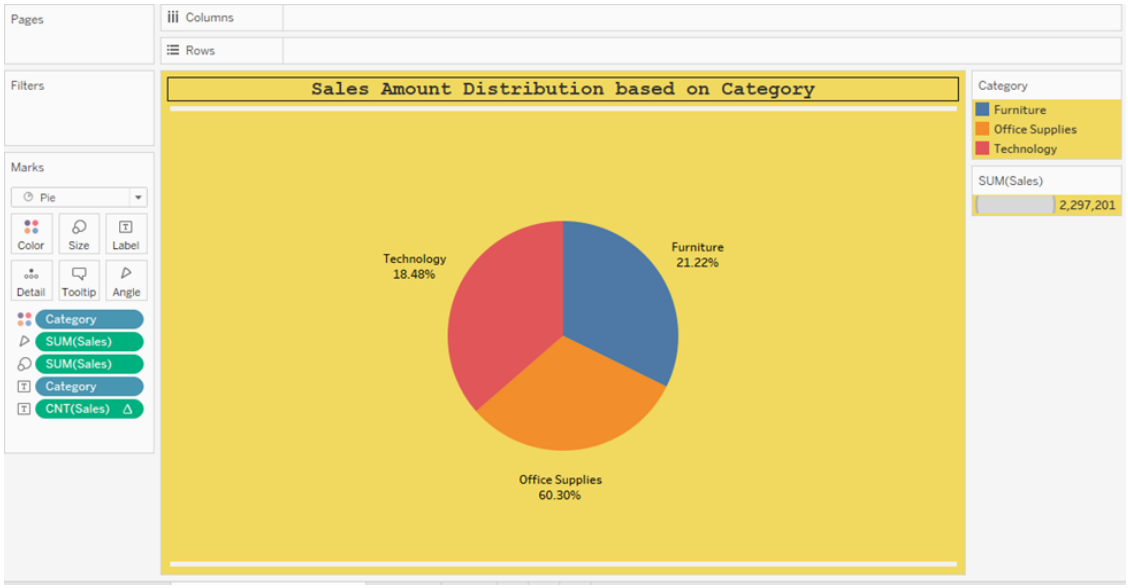
1. Which product categories have the highest total sales in the "Superstore" dataset?

The best graph that justifies the ask of the question is **Bar Chart** as it shows the clear total sales count of each Category. With sales of 6026 out of total sales 9994, office supplies constitute for the most as compared to furniture which adds 2121 and technology 1847 of total sales.



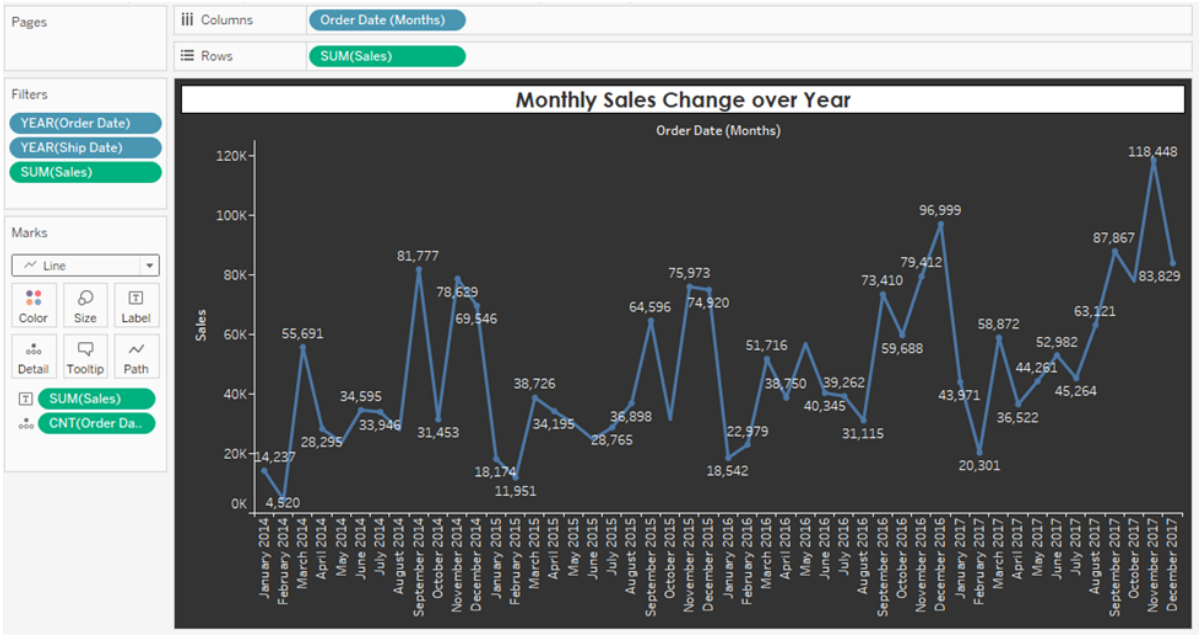
2. How is the total sales amount distributed among different product categories?

The best graph to showcase distribution of product categories is **Pie Chart**. The chart depicts clearly that office supplies have the largest share out of all the product categories. It occupies a percentage share of 60.30% followed by furniture which is 21.22% and technology 18.48%.



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

When it comes to time trend analysis then **Line Chart** is the best one to portray the case. We can see monthly sales change over the years. Wherein November 2017 marks the highest sales of 118,448.

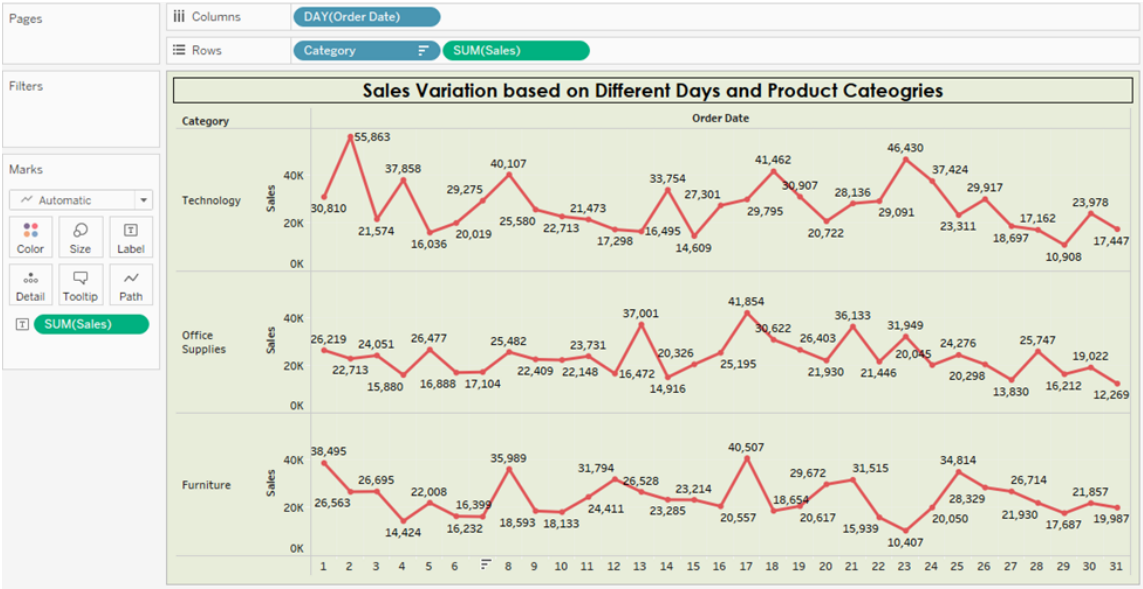


4. Can we analyze the sales performance of individual customers over time?

Based on the need and the parameters, the best chart that performs well in this scenario is **Box Chart.** Each sub point in the column box represents the year i.e., 2014, 2015, 2016, 2017. As we hover over the data points then we get into more detailing. Each data point defines the sales made by individual customers in each year.

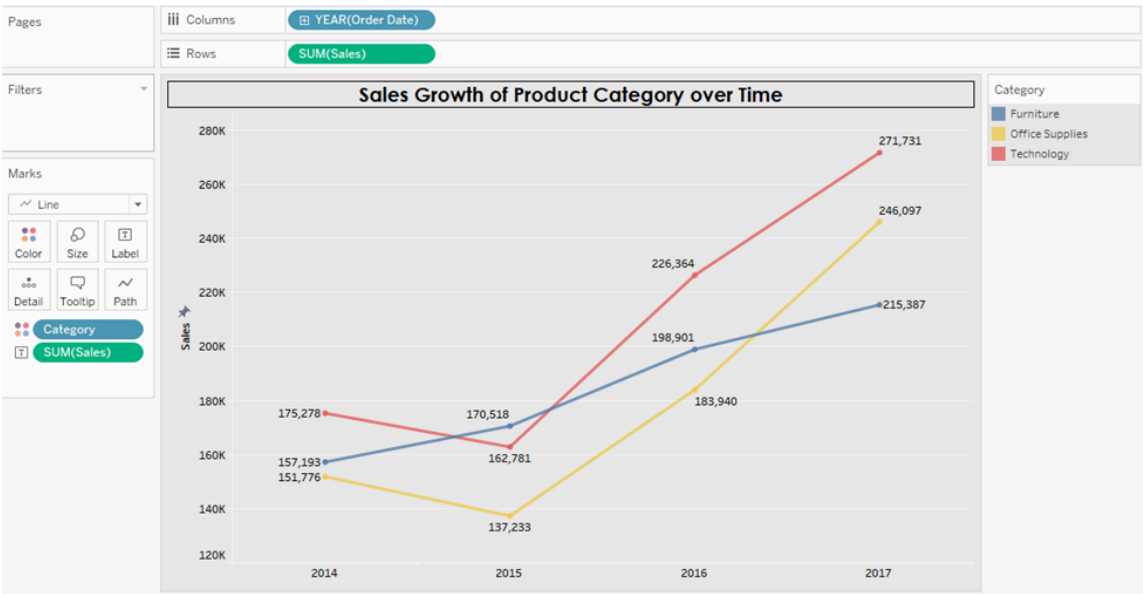
5. How do sales vary based on different days of the week and product categories?

The best chart to understand time analysis is **Line Chart**. Here also we have tried to represent sales variation for each individual product category over time, more specifically based on days. We can clearly analyze from the below graph that sales for each category has seen fluctuation if counted on a day basis.



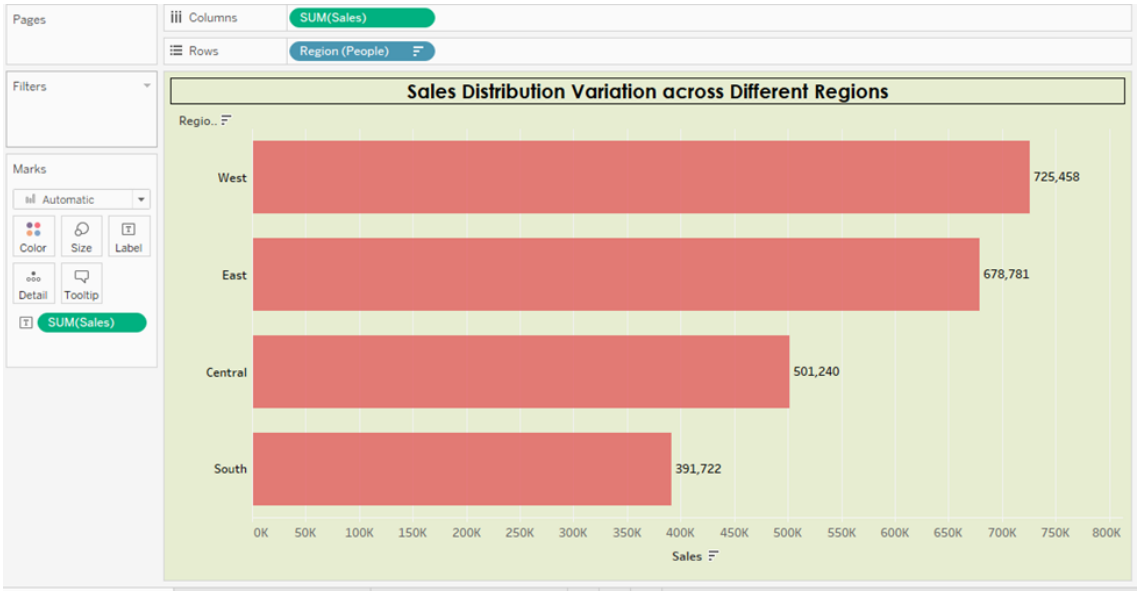
6. Can we visualize the sales growth of different product categories over time?

Here we have taken **Line Chart** into consideration as it gives better understanding of time variation analysis of how sales varied for different product categories over time. We have applied filters and reduced sales range to scale the graph more precisely. Furniture has shown an incremental sales growth over time, whereas both technology and office supplies have seen a sudden downfall in 2015.



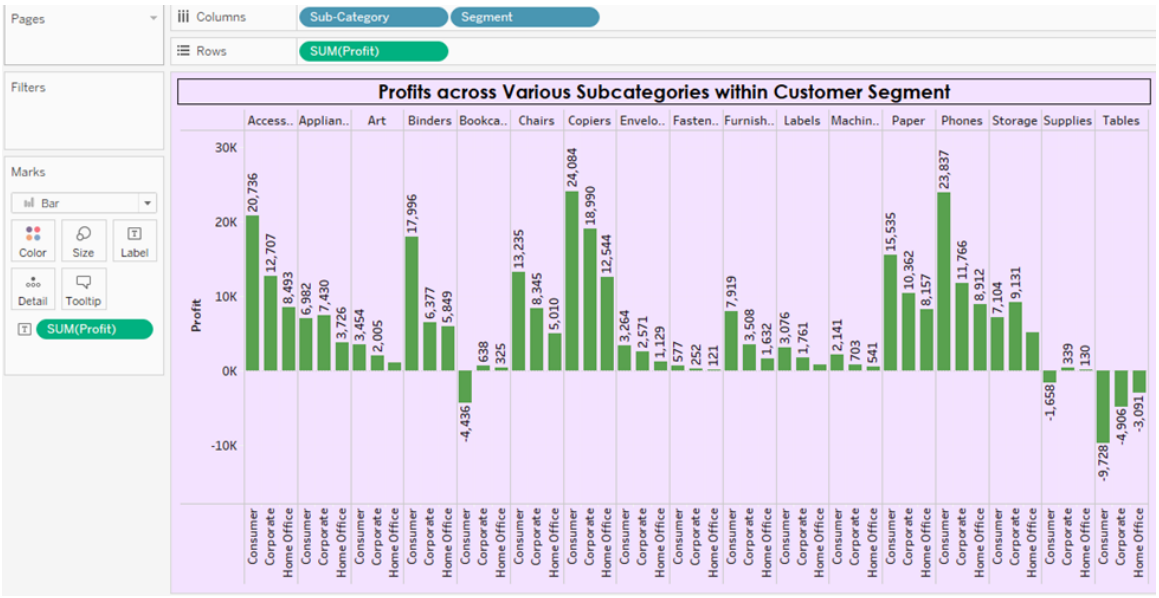
7. How does the sales distribution vary across different regions in the "Superstore" dataset?

Bar chart shows efficiency when analyzing the difference in sales based on regions. Here we can see that the West region has made the largest sales of 725,458 followed by east with sales of 678,781.



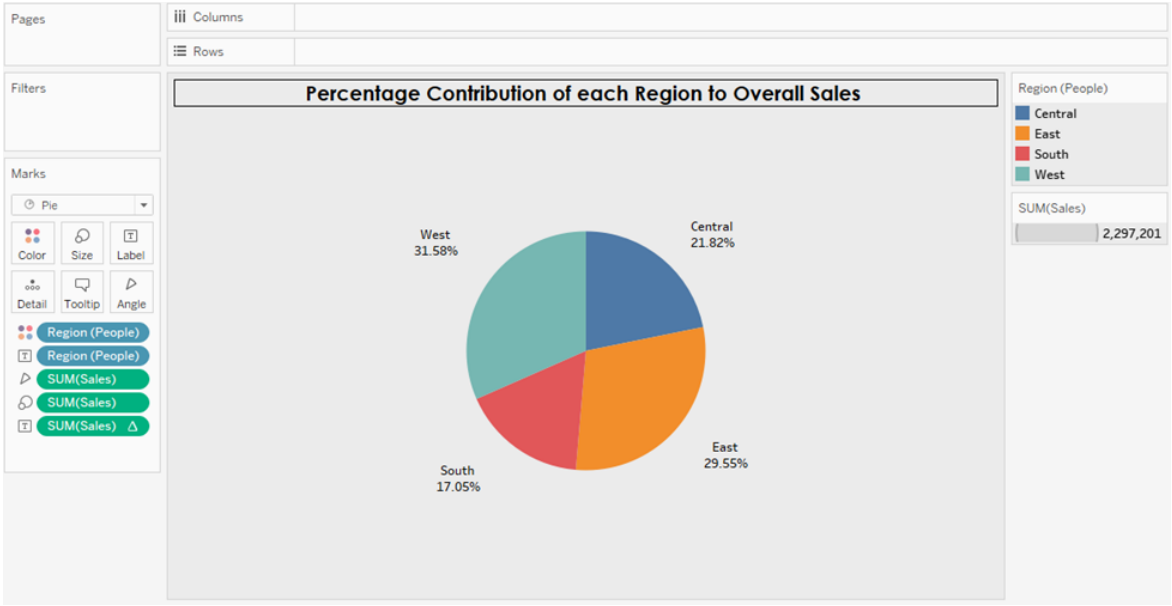
8. Can we visualize the composition of profits across various subcategories within different customer segments ?

We can visualize the composition of profits across various subcategories within different customer segments via **Label Chart** as it represents complex data in a very simple and visually understandable way. Segmenting the Subcategories and then calculating profit for respective fields.



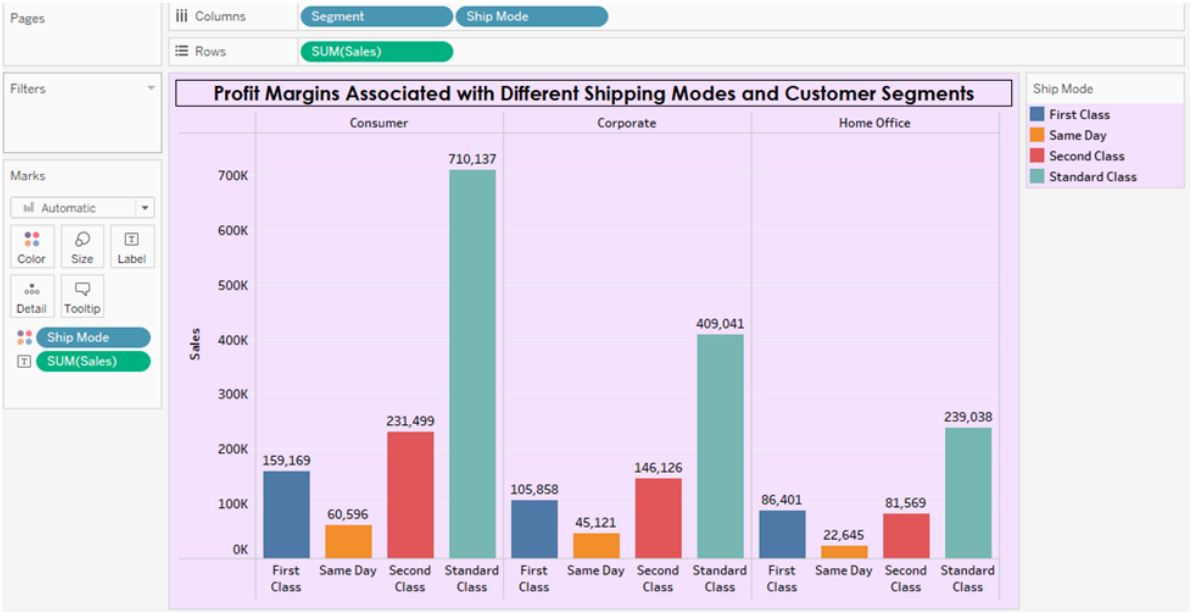
9. What is the percentage contribution of each region to the overall sales?

Bar Chart displays the distribution percentage of each region with respect to the sales contribution made. The West region contributed the highest percentage of 31.58% followed by East's 29.55%, Central's 21.82%, South's 17.05% .



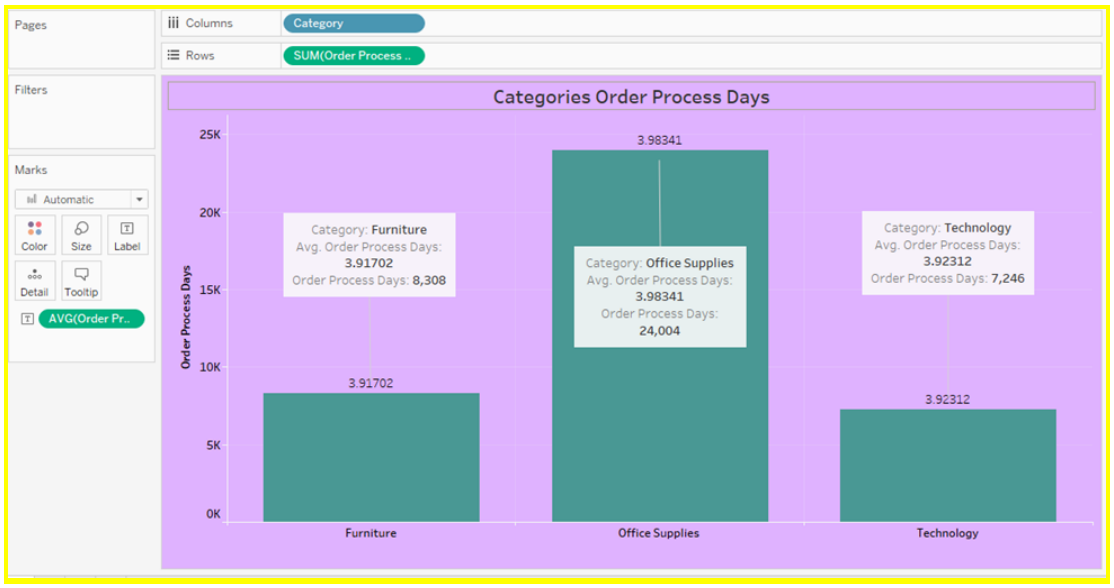
10. Can we visualize the profit margins associated with different shipping modes and customer segments?

To highlight sales drawn by different shipping modes based on customer segment, **Label Chart** fits into the criteria the most. Standard class contributes to the maximum sales in every customer segment. While same day shipping mode shows the least sales contribution in every segment.



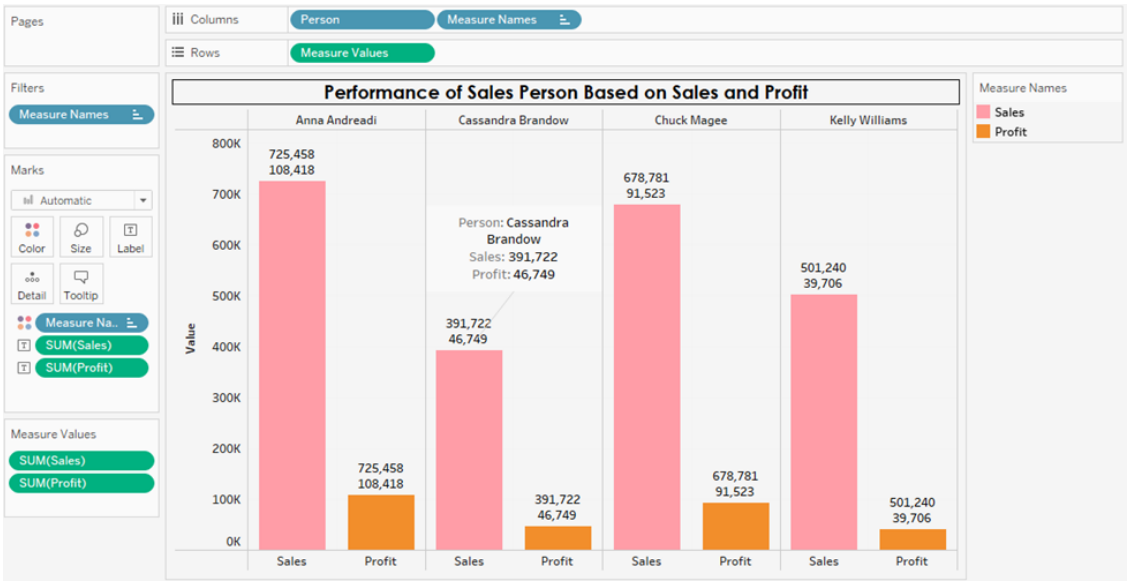
11. How long does it take to process orders for different product categories?

Here we have utilized **Label Chart** to compare the average orders day and total no.of orders days for different product categories. As we can visualize that the order process days is maximum for office supplies i.e., 24,004 but minimum for technology i.e., 7,246 with average being almost similar for all i.e., 3.



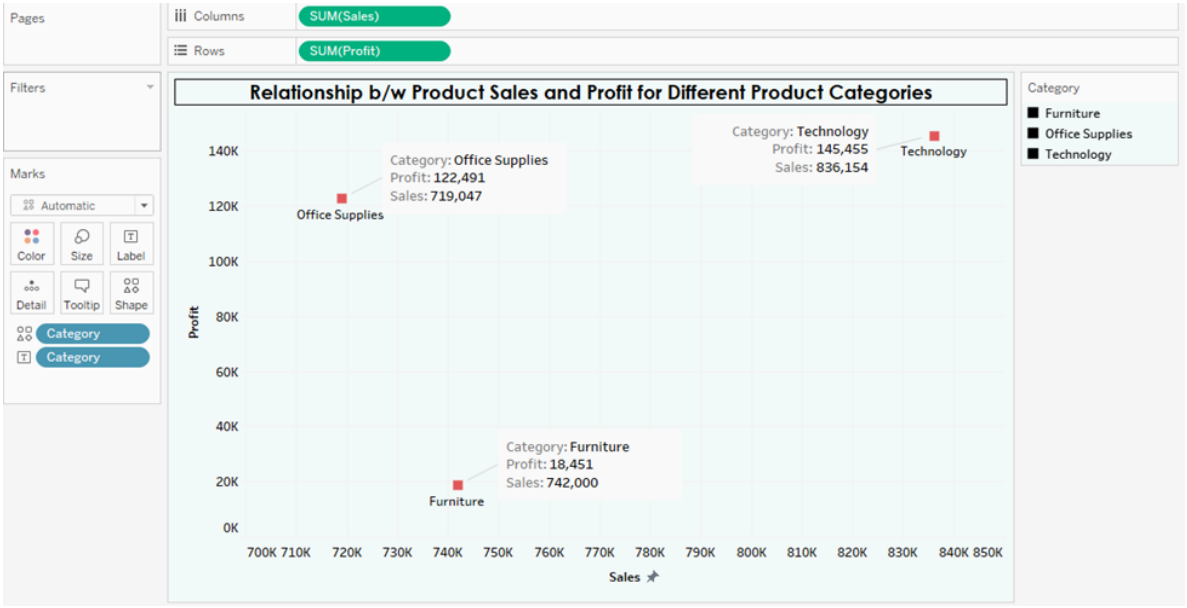
12. How does the performance of different salespeople compare in terms of sales targets, actual sales, and profitability?

Label Chart depicts the performance of a sales person based on sales and profit. The key finding shows that profit made by every sales person is quite low as compared to sales. Kelly with 39,706 contributes least sales and Anna with 725,458 make highest.



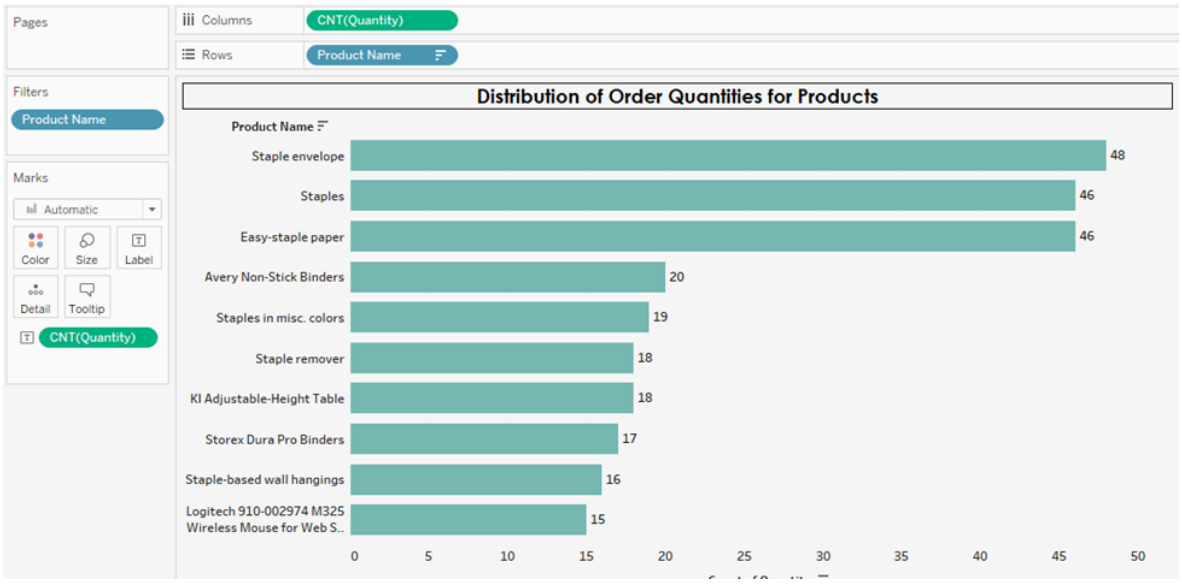
13. Can we visualize the relationship between product sales and profitability for different product categories?

To understand the relationship b/w sales and profit made by different product categories we have chosen **Scatter Plot** as it justifies the need. The pointers have been annotated to give more detailed information of the findings. Technology makes the most profit as well as sales.



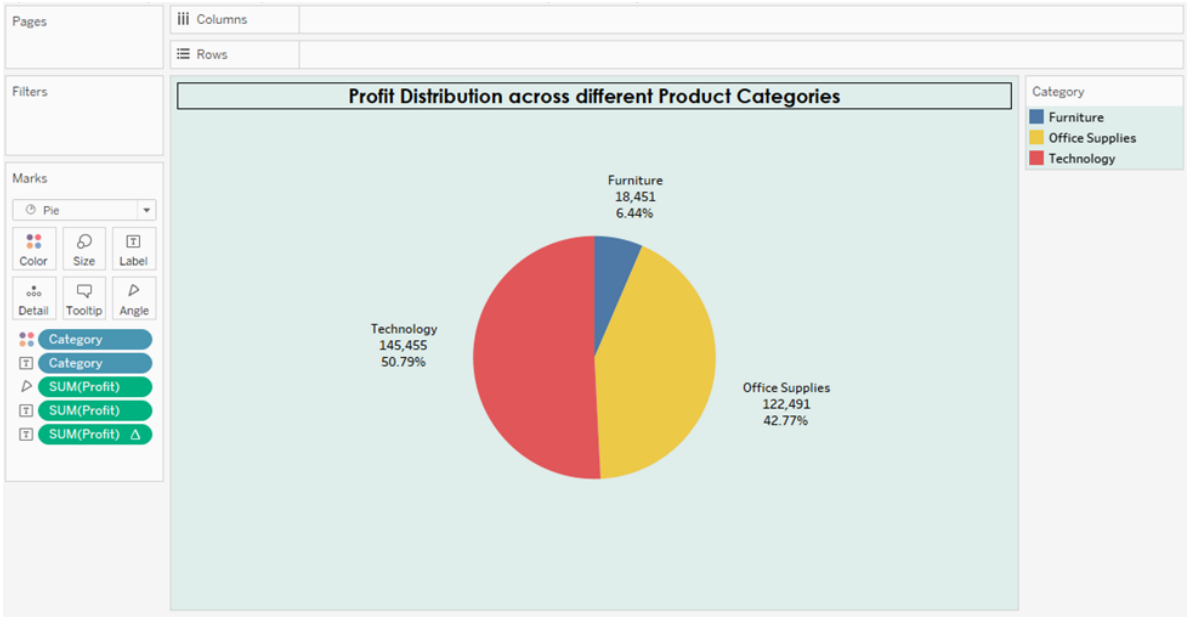
14. What is the distribution of order quantities for products in the dataset?

We have chosen a **Bar Chart** here to define the key difference in the distribution of order quantities for products. Since Products are pretty huge in numbers we have taken Top 10 amongst them. Where the staple envelope marks the highest distribution of 48.



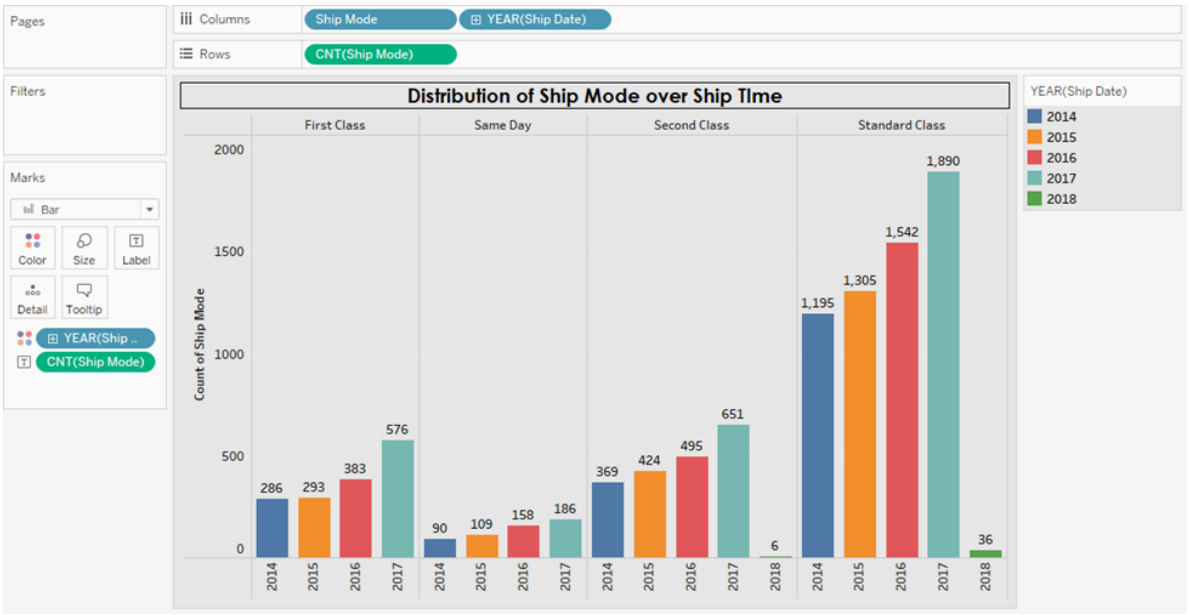
15. How do the profit distributions vary across different product categories?

The profit distribution across different product categories has been showcased via **Pie Chart**. We can pretty clearly say that the profit share of technology is the most whereas furniture is the least. With profit of 145,455 technology contributes for 50.79%.



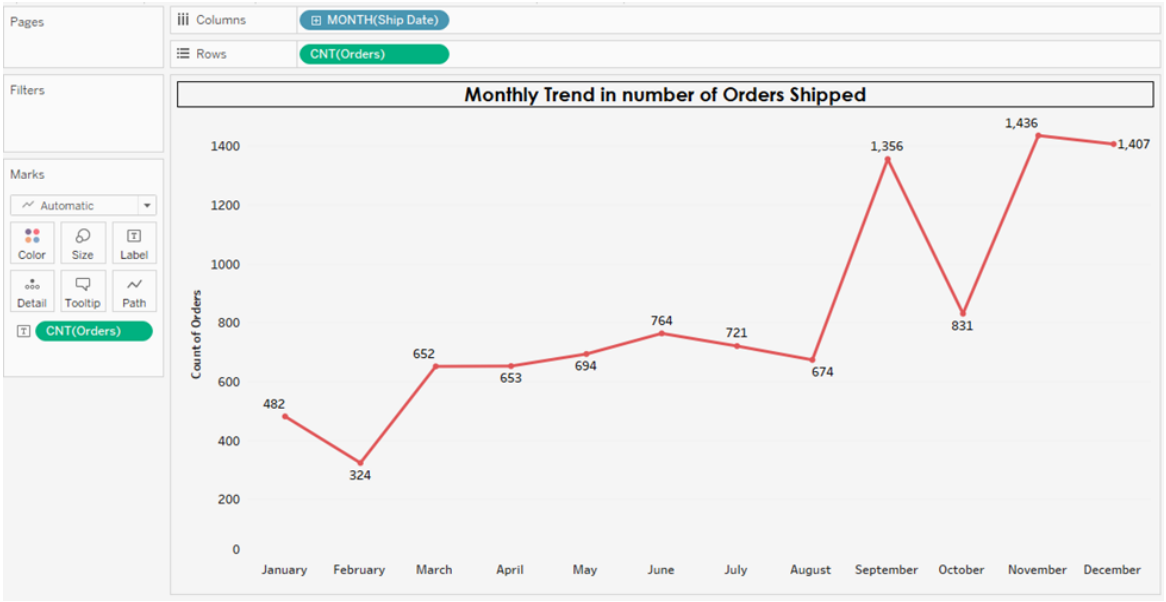
16. Can we compare the shipping time distributions for different shipping modes?

Yes we can compare the shipping time distributions for different shipping modes via **Label Chart**. We can investigate from the below chart that standard class has made the highest shipping of 1890 that to in 2017 and second class the least of 6 in 2018.



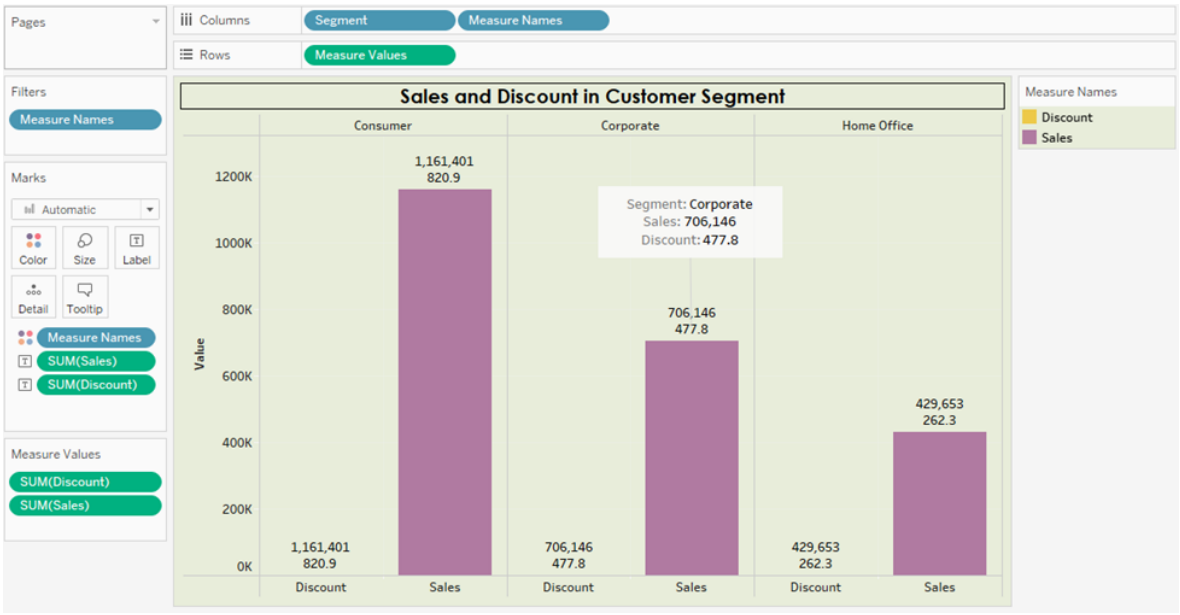
17. What is the monthly trend in the number of orders shipped?

We can analyze the time trend of the count of orders done per month with the help of **Line Chart.** It can be observed that November has seen the highest count of orders of 1436 and February the least of 324.



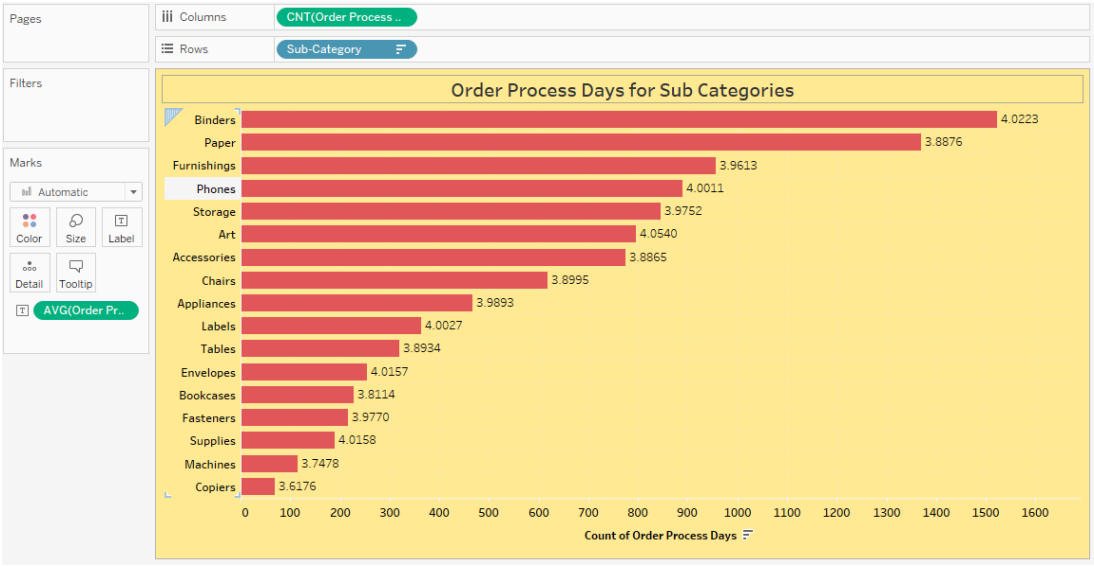
18. How do different customer segments perform in terms of sales and discount rates?

Label Chart helps to understand how discount and sales have seen growth in different customer segments. Consumer segment has made the highest sales in conjunction with the highest discount. Hence, we can conclude from the below chart that higher the discount the higher is the sales.



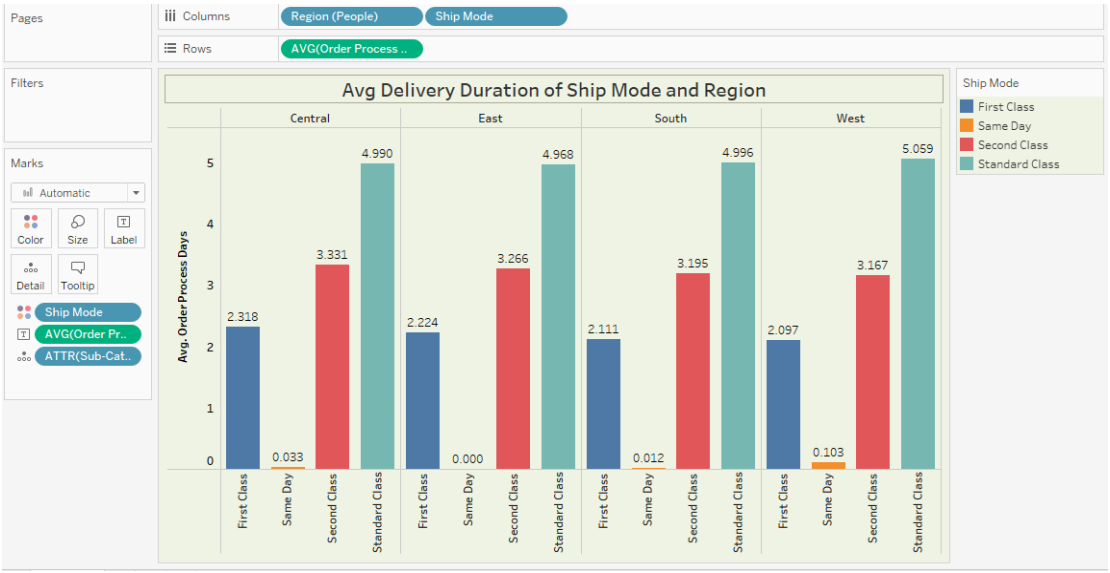
19. How efficiently are different product subcategories being fulfilled in terms of order processing time and on-time delivery?

To understand the case thoroughly we have chosen a **Bar Chart**. We can clearly see from the graph that the average order process days for different sub categories fall somewhere b/w 3-4 with Blinders making the highest of 4 and Copiers the least of 3.6.



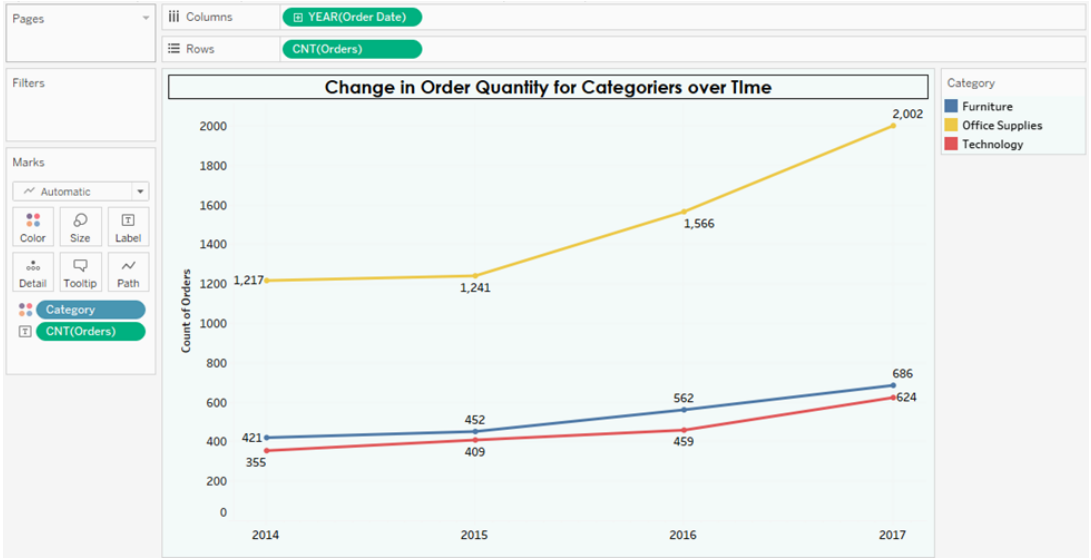
20. What is the average delivery duration for different regions and ship modes?

To understand the distribution of average delivery duration for different regions and ship modes here have selected Side by **Side by Side Chart** as it portrays the information quite clearly. From the graph we can see that average delivery duration varies from 2.3-5.0. Under the East same day serves the least average delivery duration whereas standard class in West serves the most.



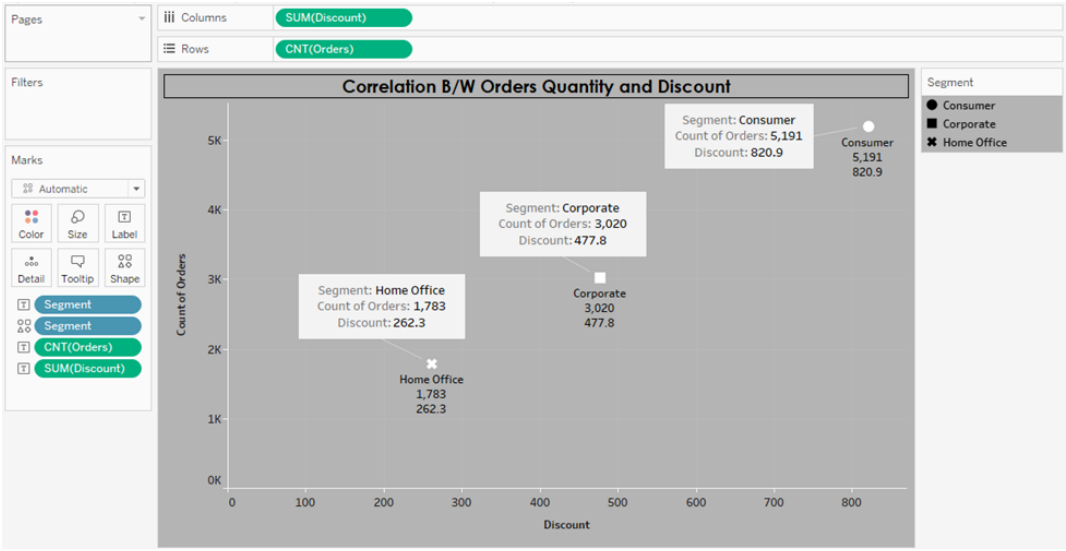
21. How has the average order quantity changed over the years for various product categories?

Average order quantity change over years for various product categories could be understood via **Line Chart**. Office supplies have drawn maximum orders of 2002 in 2017. Overall, all the product categories have shown a significant uplift in orders count every passing year.



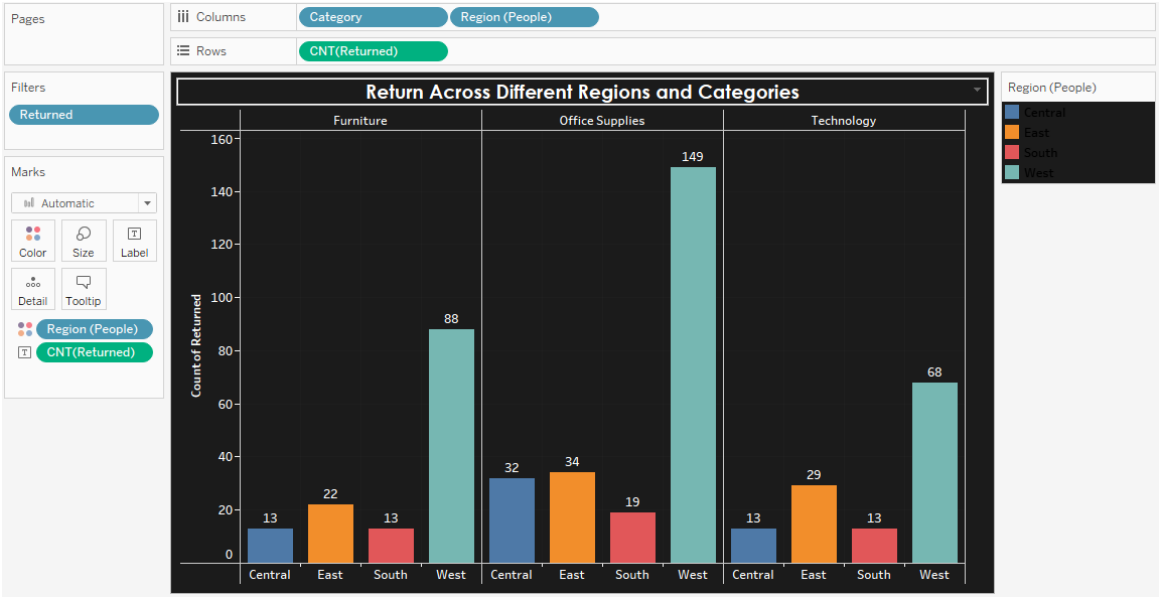
22. Can we visualize the correlation between discount rates and order quantities for different customer segments?

To deduce correlation b/w orders quantity and discount we have chosen **Scatter Plot.** From the below chart we can study that the consumer segment has made 5191 highest no.of orders at an average discount of 820. While the home office had made the least.



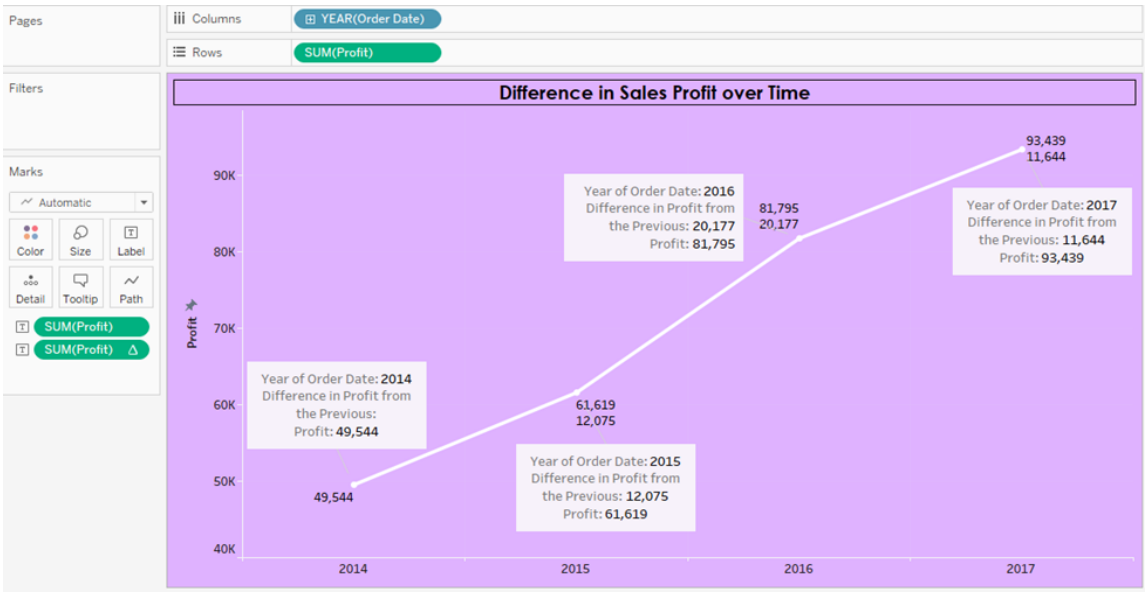
23. What is the trend of returns and refunds across different regions and product Categories?

To analyze returns across different regions and product categories we have chosen a **Column Chart**. We can clearly figure out from the chart that West in office supplies have shown the highest no.of returns, whereas South has shown the least returns.



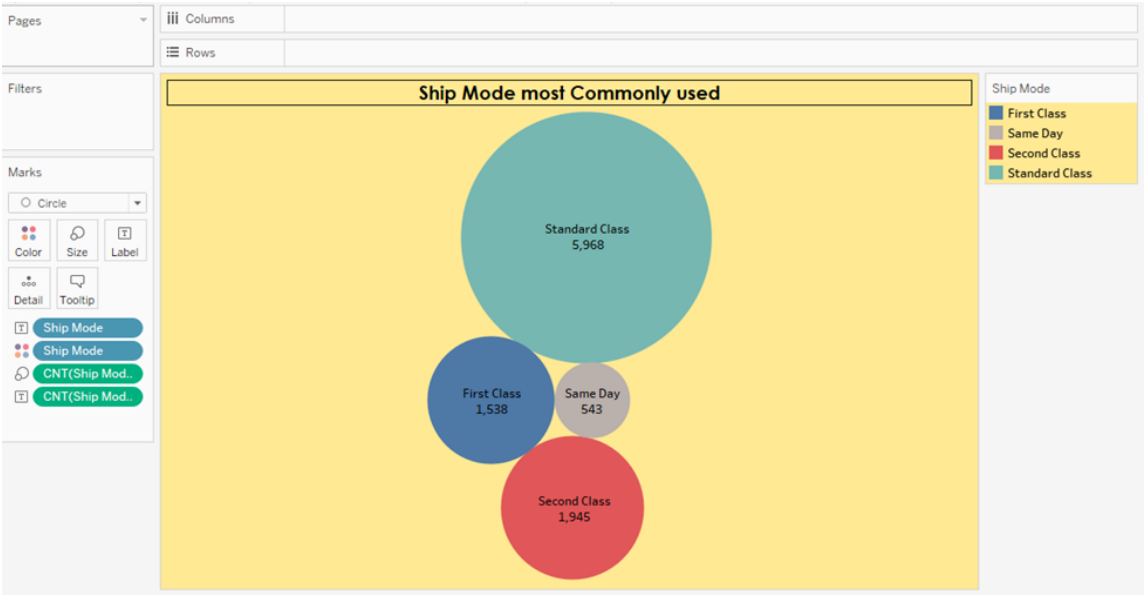
24. How do the sales of high-profit products compare with low-profit products over time?

The graph that best shows the sales profit and difference from the previous one over time is a **Line Chart.** Here we can analyze that every passing year has seen incremental growth from the previous one, but the difference in sales profit is dynamic throughout.



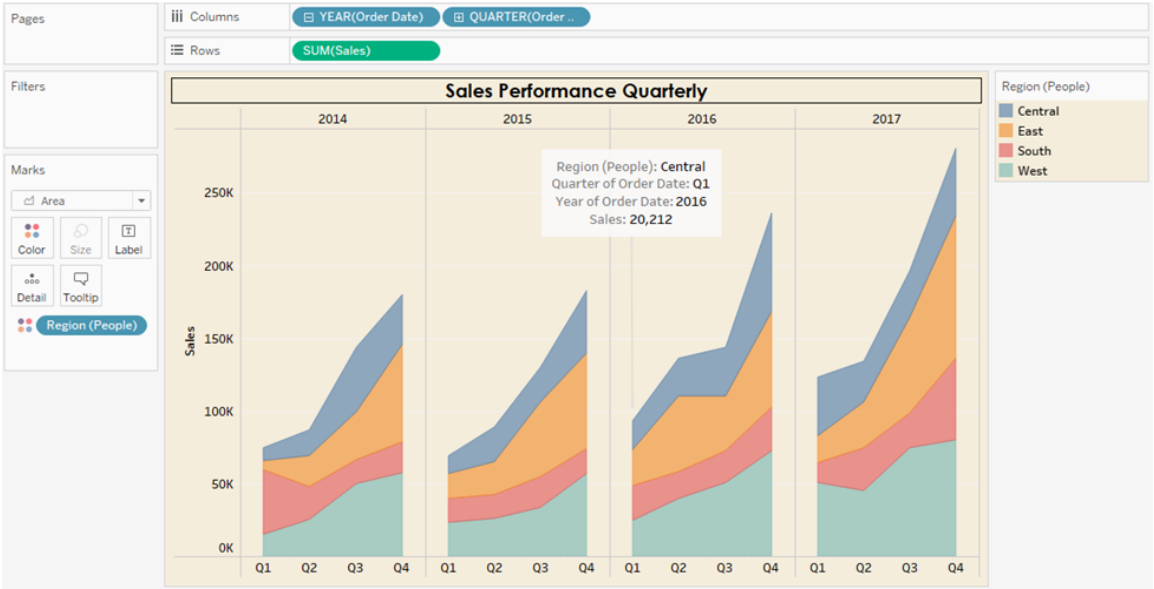
25. Which shipping mode is the most commonly used in the Sample Superstore dataset?

Based on the ask of the question the best chart that matches the criteria is **Packed Bubble Chart** as it gives a clear description of shipping modes accessed most often. Amongst all shipping modes, Second class (5968) is used the most, whereas same day (543) the least.



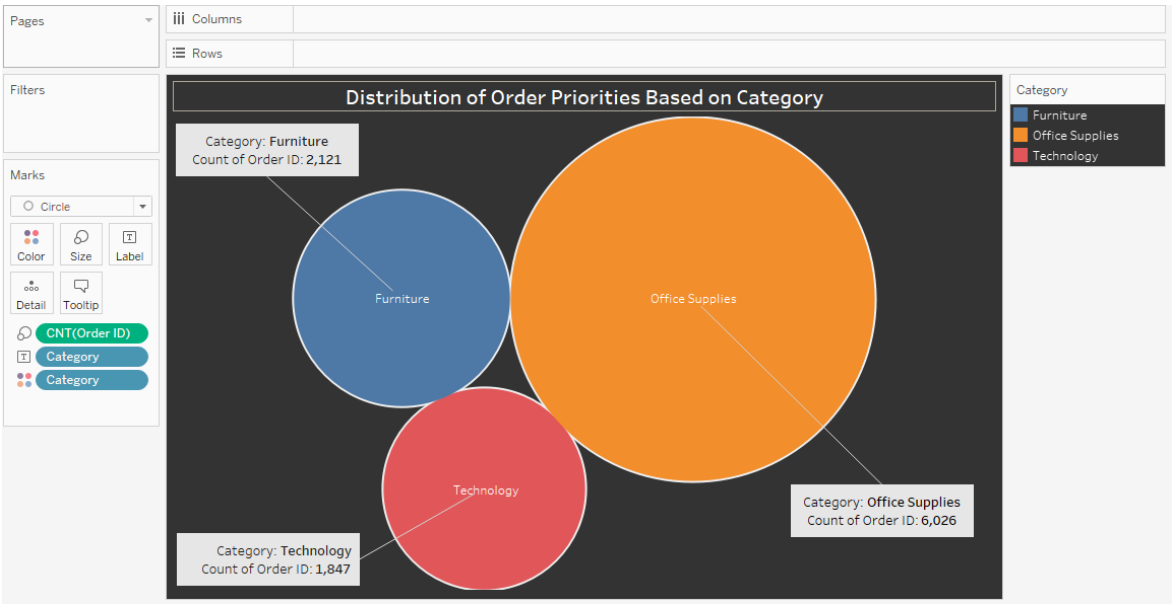
26. How does the sales performance of different regions evolve throughout the quarters of a year?

To understand sales performance of each region throughout the quarters of the year we have chosen **Area Chart.** Adding annotation gives a better and detailed description of the sales done. As we hover on the data points we get to see more detailed information.



27. What is the distribution of order priorities across different product categories?

To show the distribution of order priorities across different product categories we have chosen **Packed Bubble Chart**. As we can see the office supplies hold the highest priority whereas technology the least.

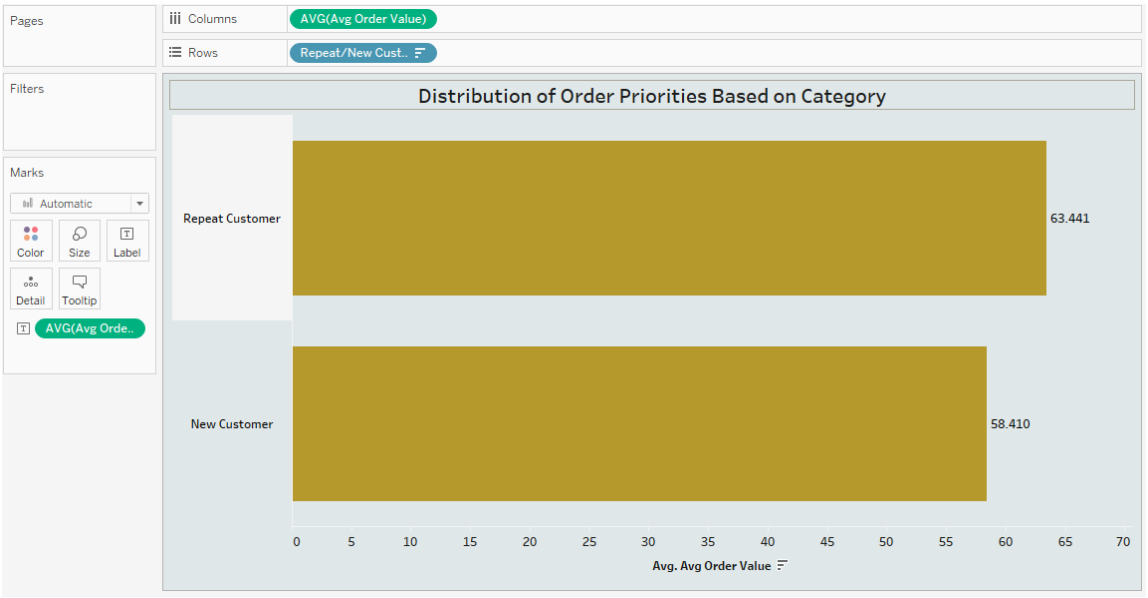


28. What is the relationship between discounts and sales?

To interpret relationship b/w discounts and sales **Scatter Plot** justifies the case most. As we can see from the chart that the avg sales done is 229.9 at a discount of 0.1562.

29. How does the average order value differ between repeat customers and new Customers?

To show the difference in average order value b/w repeated customers and new customers we have chosen **Bar Chart**. It is clearly visible from the chart that the average order value for new customers is 58.4 whereas for new customers it's 63.4.



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

The below **Map chart** shows the distribution of returns and its overall impact on profitability from the Country, the United States of America and its corresponding states. The profit and returns are in quite distributed form.