

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.

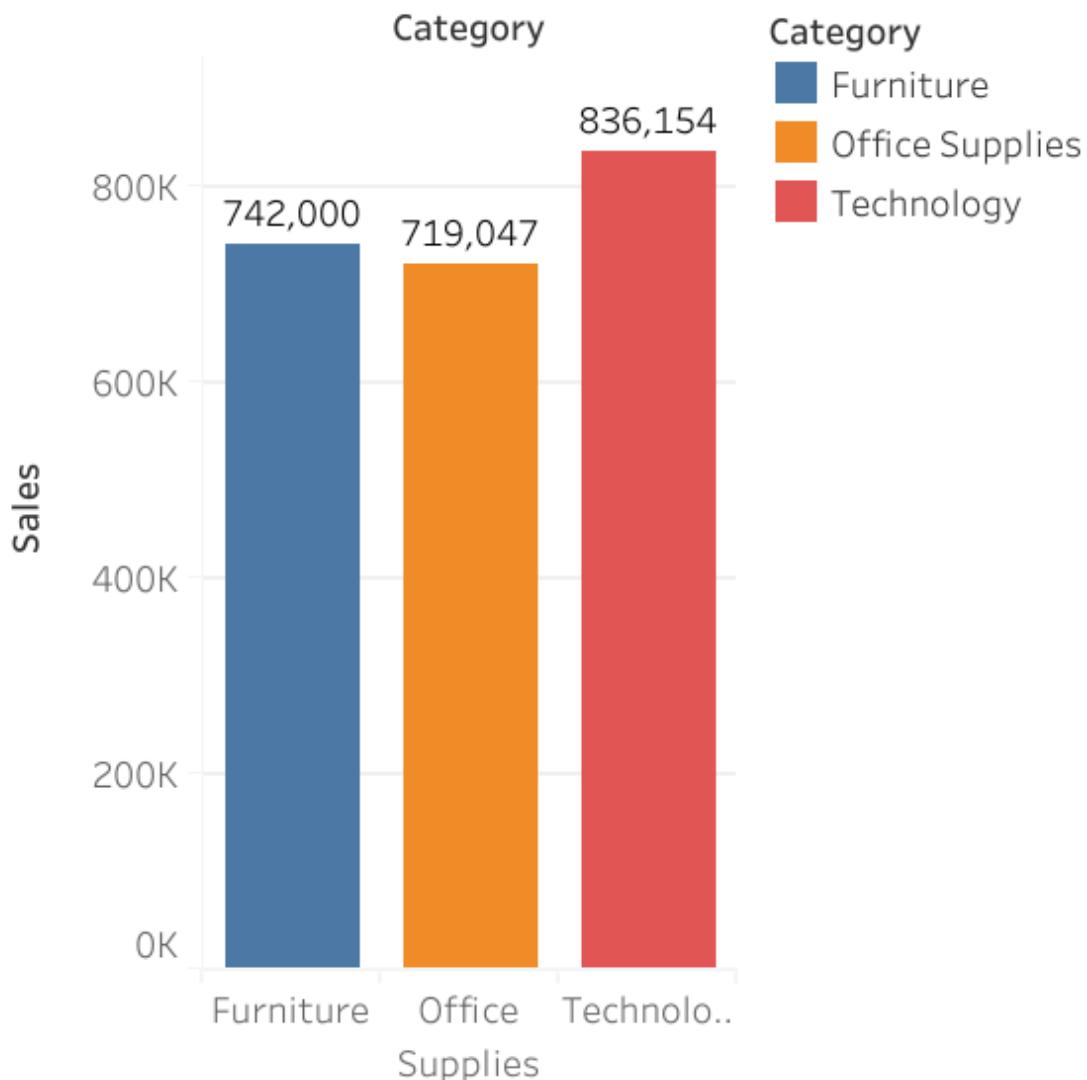
By completing this task effectively, you'll not only demonstrate your proficiency in data visualisation and analysis but also showcase your ability to effectively communicate complex concepts through both text and charts.

Good luck!

Questions:

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

<Total sales by various product categories>



As, total sales across different categories of products is to be compared, A bar graph is preferred for the purpose.

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

Monthly sales amount show a general increasing trend from Jan to Dec. Maximum monthly sales happen in November.

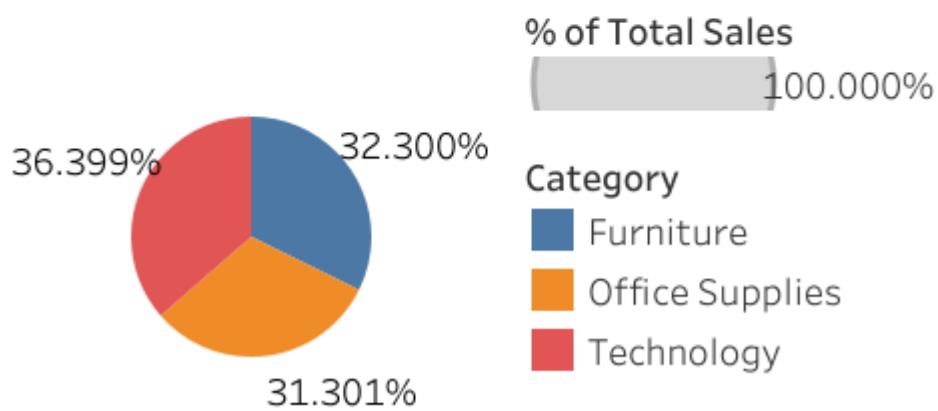
<Monthly sales amount trends over the year>



Trends in Monthly sales amount over the year is shown using a line chart. Line chart is used as it makes the desired information easy to understand.

3. How is the total sales amount distributed among different product categories?
Office supply(31.3%), Furniture(32.3%), Technology(36.4%)

<Distribution of total sales amount among different product categories>



A Pie chart is used to show how the total sales is distributed among various product categories. Pie chart shows individual share as percentage, and is easily understandable, hence is preferred for the purpose.

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

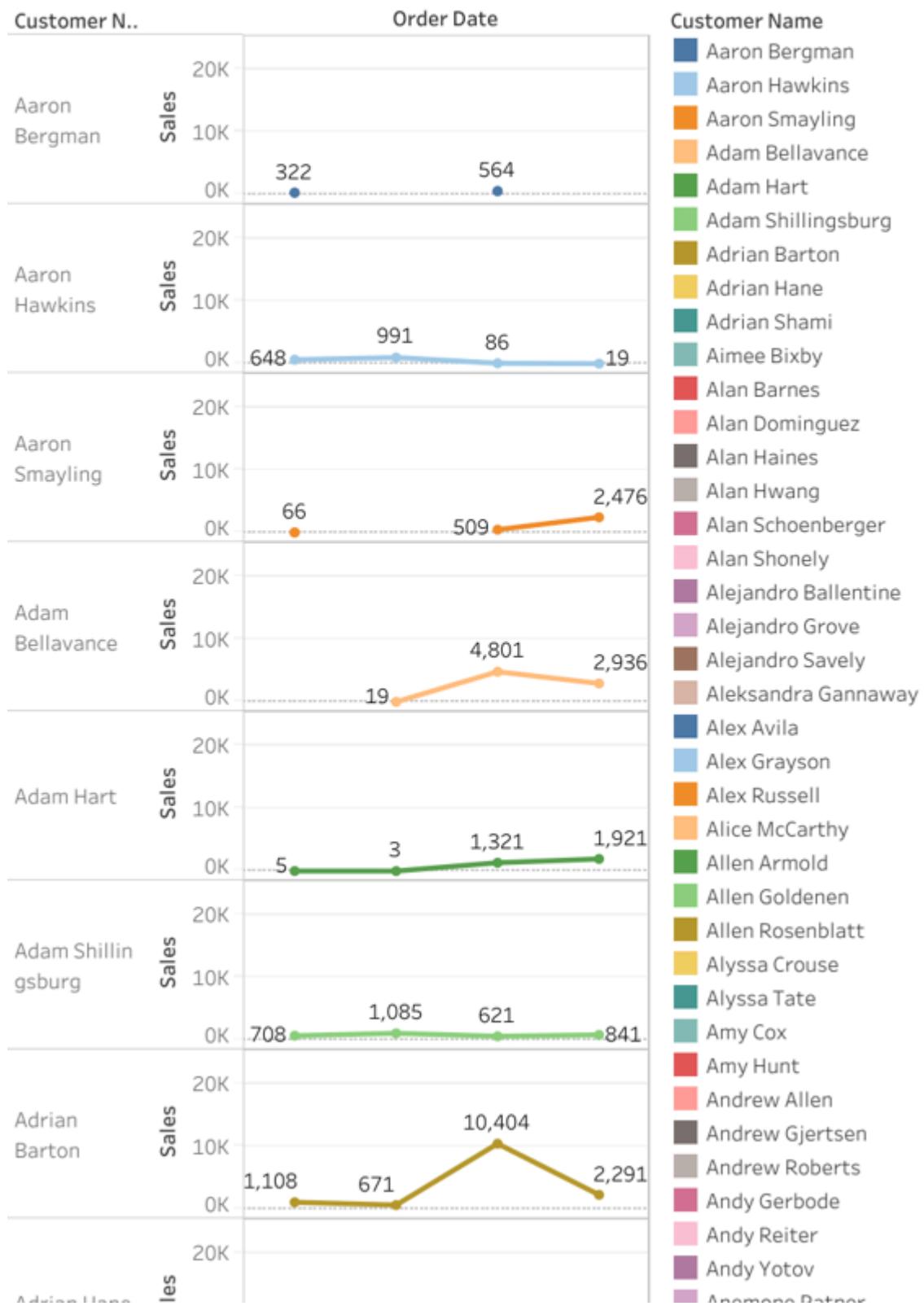
We can analyze the sales performance of individual customers over time (here, done yearly) using highlight tables and line charts. Highlight tables clearly show sales data for every customer yearly. A line chart plots that data and sales trends can be quickly interpreted from them.

Here, is a snapshot:-

<Sales performance of individual customers over time>

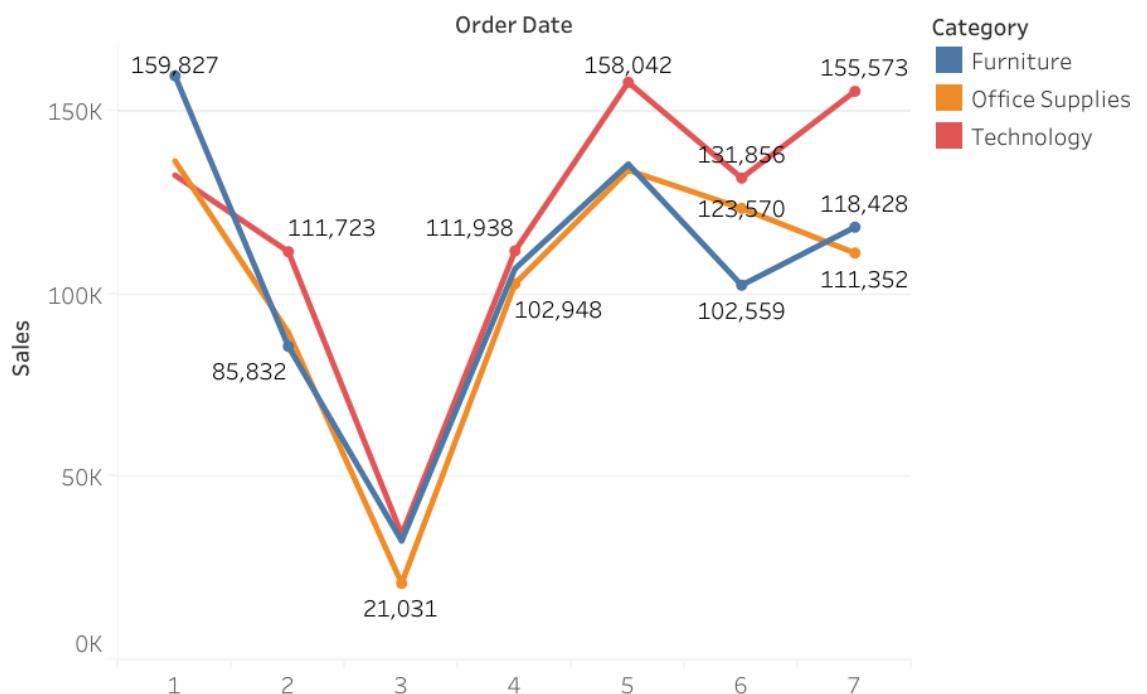
Customer Name	Year of Order Date				Sales
	2014	2015	2016	2017	
Aaron Bergman	322		564		1
Aaron Hawkins	648	991	86	19	23,661
Aaron Smayling	66		509	2,476	
Adam Bellavance		19	4,801	2,936	
Adam Hart	5	3	1,321	1,921	
Adam Shillingsburg	708	1,085	621	841	
Adrian Barton	1,108	671	10,404	2,291	
Adrian Hane	259		661	816	
Adrian Shami			4	55	
Aimee Bixby	706	15	16	231	
Alan Barnes	37	660		416	
Alan Dominguez	606	67		5,434	
Alan Haines		133	1,410	44	
Alan Hwang	606	882	16	3,301	
Alan Schoenberger	2,217	27	166	1,852	
Alan Shonely	378	16	131	59	
Alejandro Ballentine	20	91	65	739	
Alejandro Grove	155	38	2,375	15	
Alejandro Savyly			3,098	116	
Aleksandra Gannaway	249	31	82	6	
Alex Avila	756	27	4,406	374	
Alex Grayson			249	412	
Alex Russell	776	180		100	
Alice McCarthy	15	433	57	309	
Allen Armold	51	269	530	207	
Allen Goldenen	144		37	20	
Allen Rosenblatt	266	240		1,731	
Alyssa Crouse	706		58	162	
Alyssa Tate	17	3	43	1,108	
Amy Cox	742	82	1,244	3,460	
Amy Hunt		1,847	71	577	
Andrew Allen	27		1,748	16	

<Sales performance of individual customers over time>



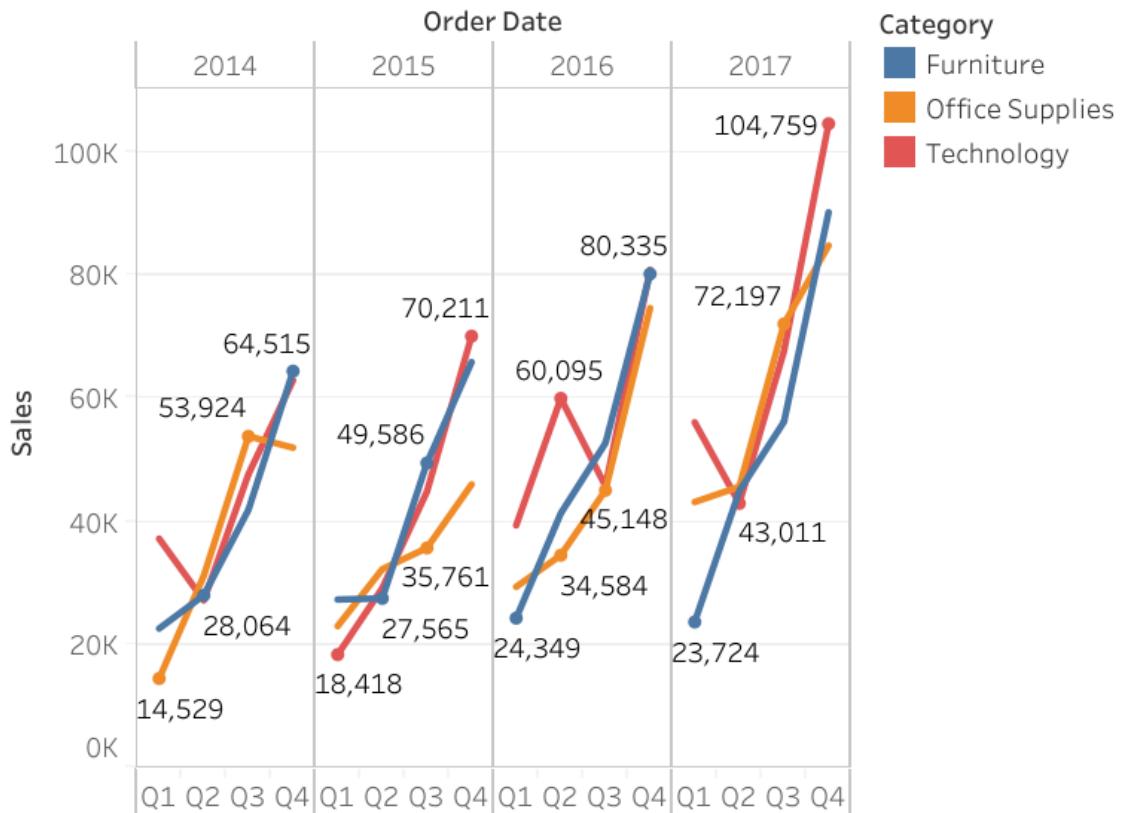
5. How do sales vary based on different days of the week and product categories?
A multi line chart is used to portray weekly sales variation for different product categories. It is easy to understand and compare. From Monday to Wednesday, there is a dip in sales, then the sales rise on Thursday and further increase on Friday to near peak value for all the categories. There is a light blip in sales on Saturday and it rises on Sunday for Technology and Furniture category, however for Office Supplies, the sales continue to decrease on Sunday.

<Sales trend for different categories of products based on different days of the week>



6. Can we visualise the sales growth of different product categories over time?
Yes, we can use multi line chart, where each line chart pertains to one product category, to visualise their sales growth over time. It is easy to understand and compare.

<Sales growth of different product categories over time>

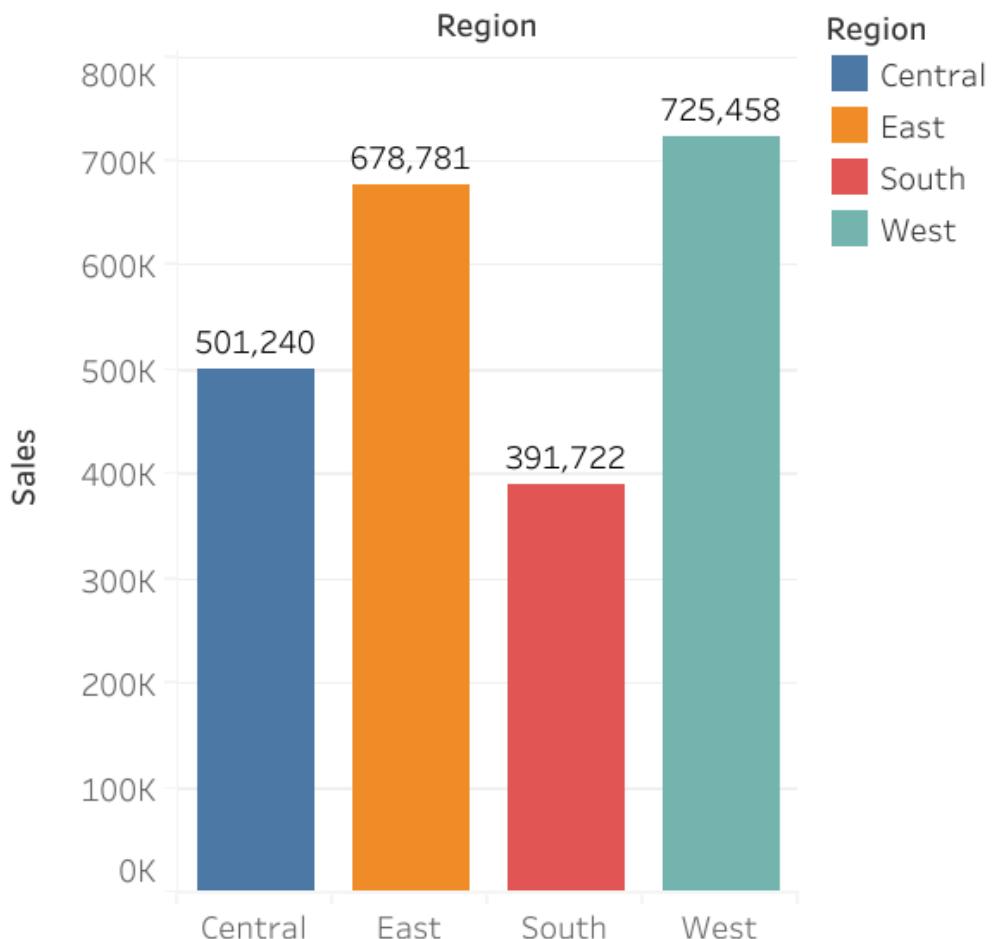


Here, yearly and quarterly sales growth is visualised for each product category.

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

Variation in sales distribution across different regions is best portrayed by bar chart, comparing sales across four regions.

<Sales distribution across different regions>

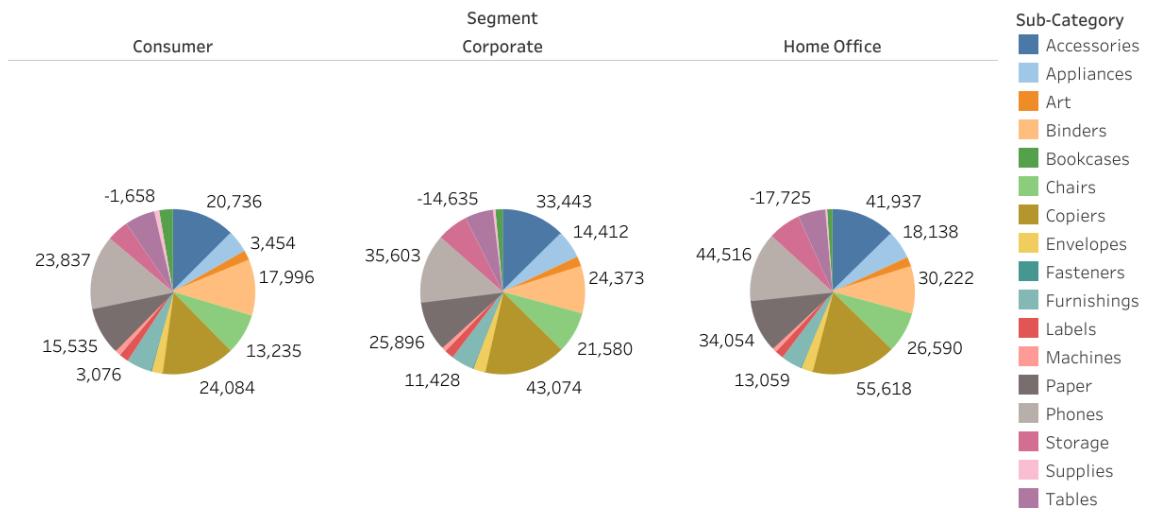


From graph, it can be seen that maximum sales have come from West, followed by the East, Central and South region respectively.

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

Yes, we can use a pie chart to compare composition of profits across various subcategories within different customer segments. It is easy to visualise and interpret.

<Composition of profits across various subcategories among different customer segments>



Maximum profit for Consumer segment is obtained from Copiers subcategory.

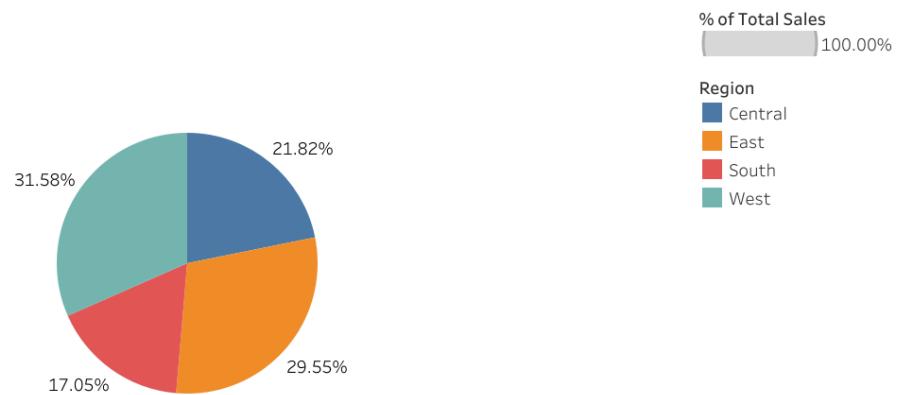
Maximum profit for Corporate segment is obtained from Copiers subcategory.

Maximum profit for Home Office segment is obtained from Copiers subcategory.

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

Pie chart accurately show the percentage contribution of each region to the overall sales. The % share of each region is as follows:-Central-21.82%, East-29.55%, South-17.05%, West-31.58%

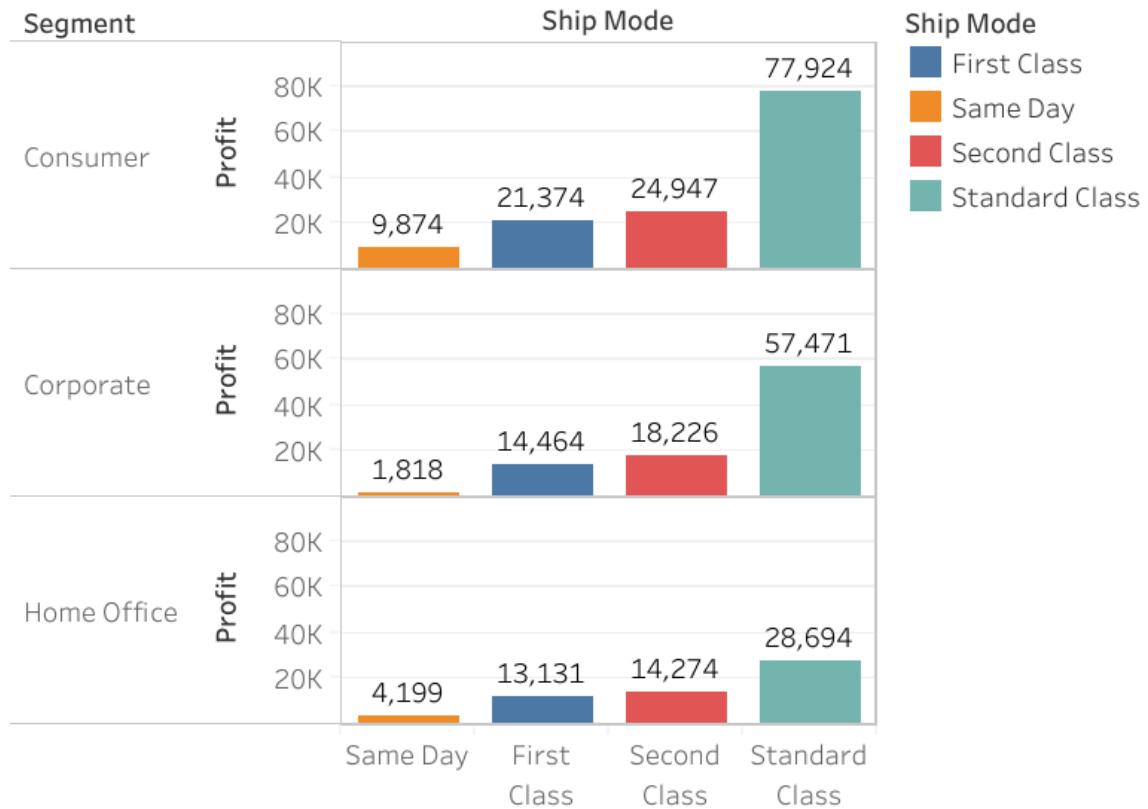
<Percentage contribution of each region to overall sales>



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

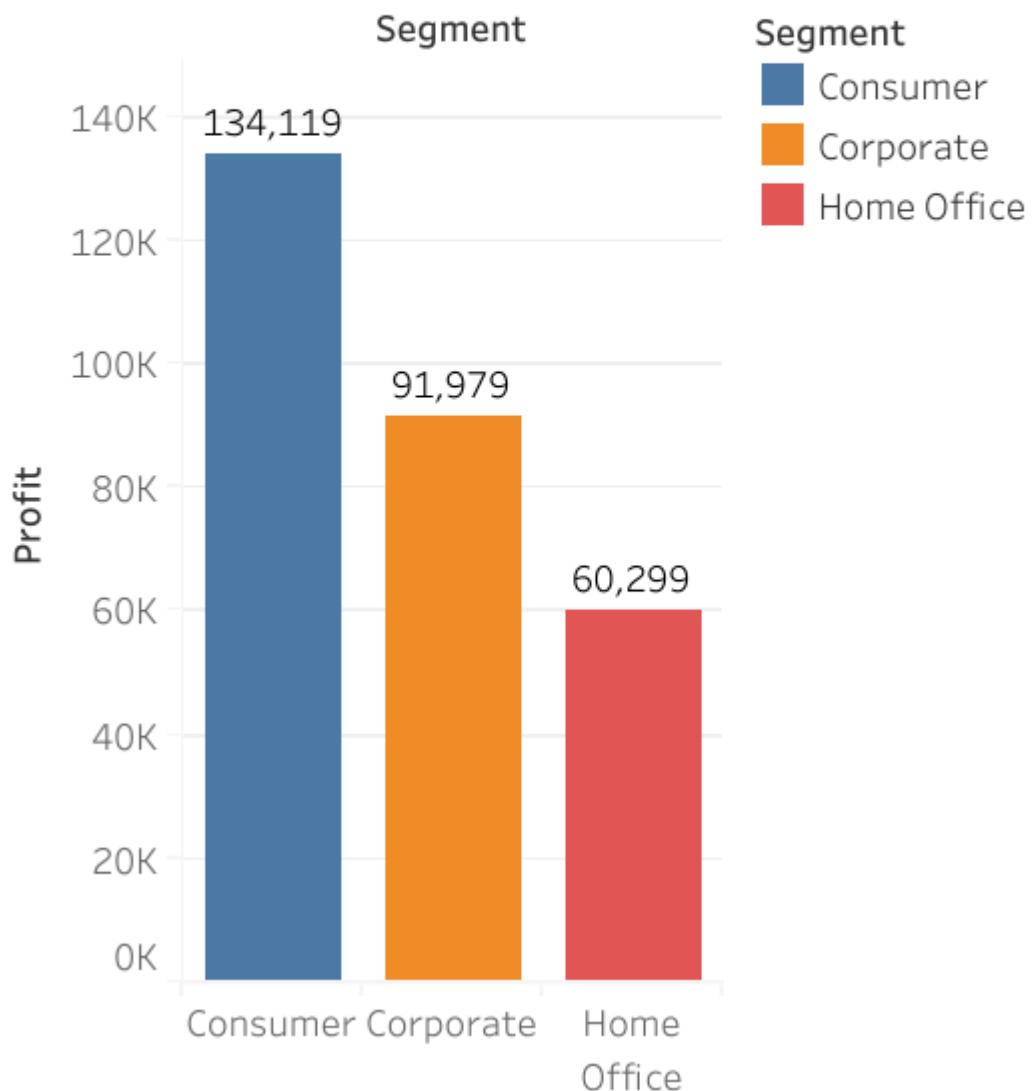
Yes, we can use bar charts to compare profit margins for different shipping modes and customer segments. The data can be clearly seen and compared.

<Profit margins for various shipping modes and customer segments>



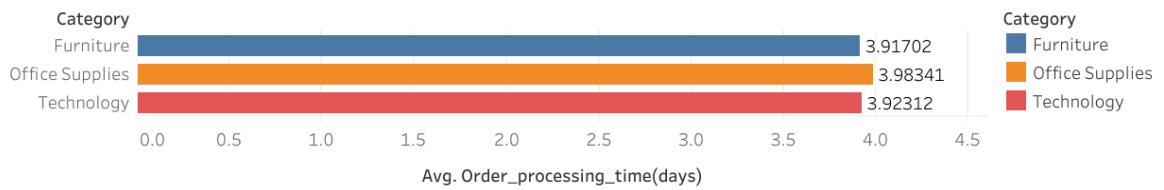
For various customer segments overall, the bar charts give accurate representation of profit margins.

<Profit margins for various customer segments>



11. How long does it take to process orders for different product categories?
Avg orders processing times:-
Furniture: 3.91702 days
Office Supplies: 3.98341 days
Technology: 3.92312 days

<Average order processing time for various categories>

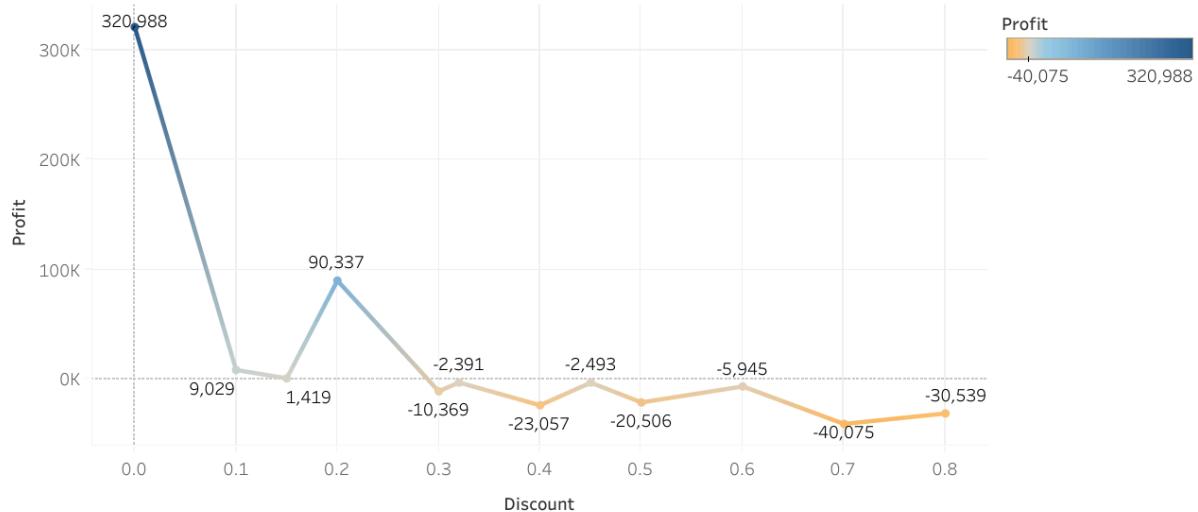


A horizontal bar chart is used to represent Average order processing times for all categories of products. It is easy to interpret and visualise.

12. How do discounts affect overall profit?

Overall profit decreases with increase in discounts. The trends can be seen on a line graph clearly and with ease.

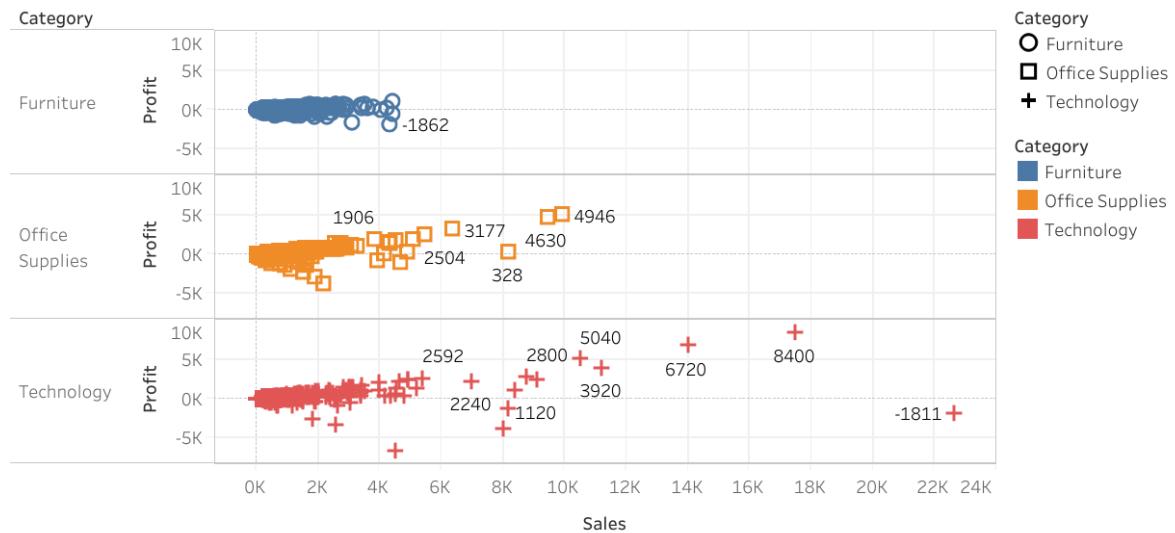
<Effect of discounts on overall profit>



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

Scatter plot is used to visualise the relationship between sales and profitability for different product categories. It gives the relationship clearly between two continuous variables.

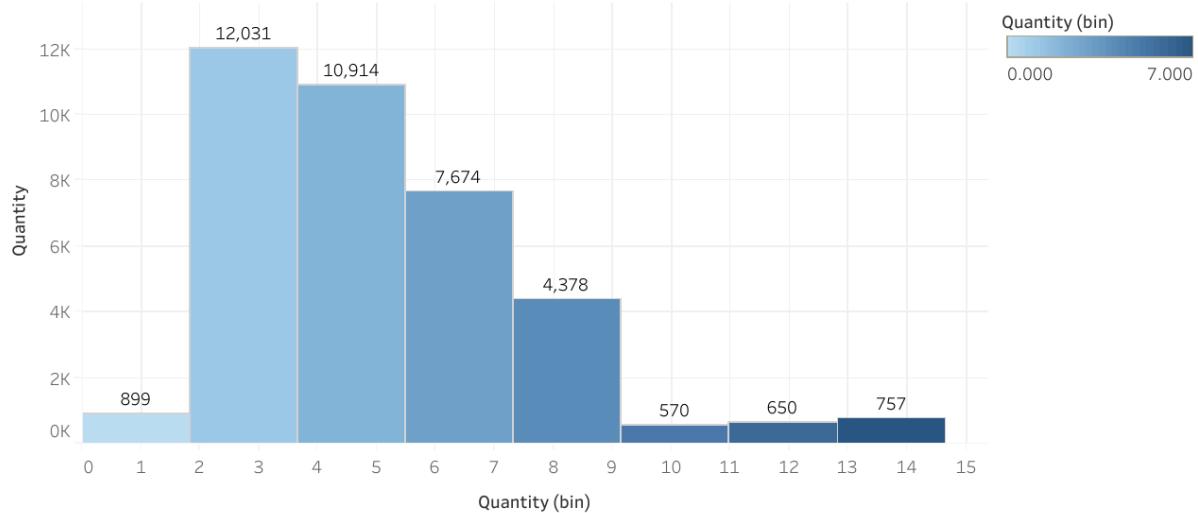
<Relationship between product sales and profitability for various product categories>



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

Histogram is used to plot the distribution of order quantities for products. The frequency of order quantities distributed equally among bins is displayed and visualised with ease. The distribution seems to be slightly right skewed.

<Distribution of order quantities for products in the dataset>

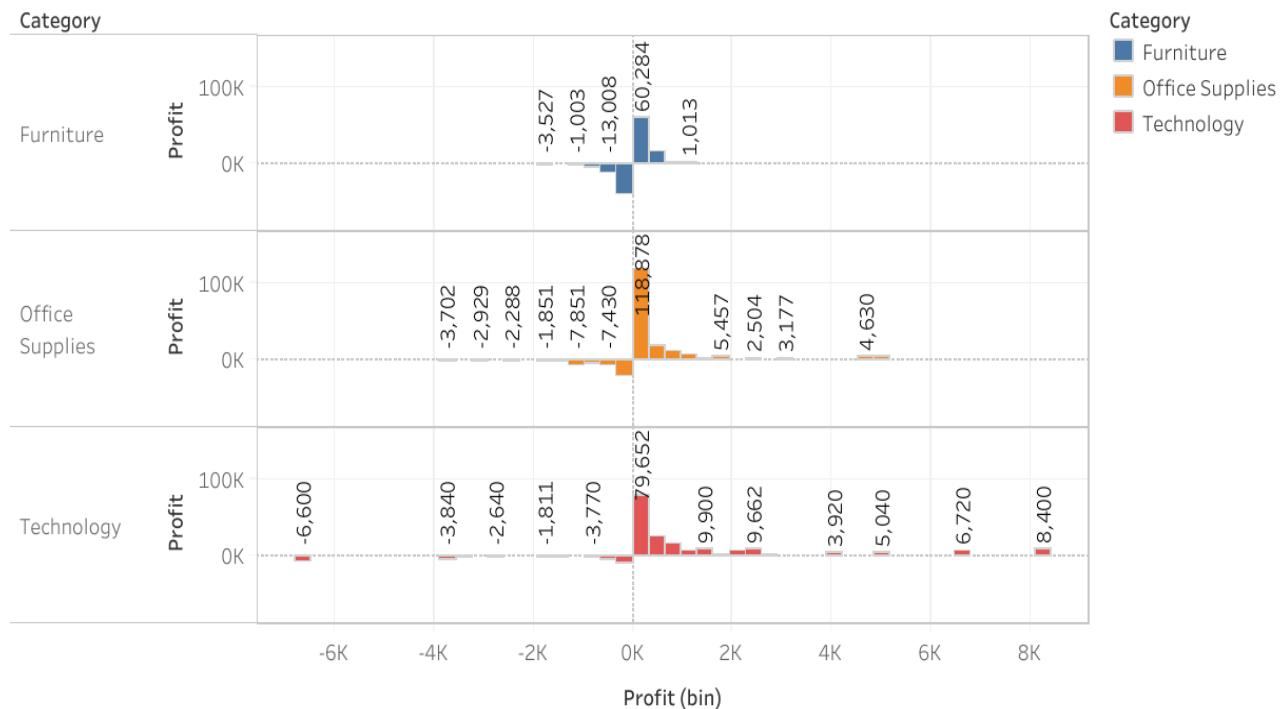


15. How do the profit distributions vary across different product categories?

Histograms for each product category clearly show the variations of profit across different product categories. It is thus used for the visualization.

For all product category, profit in most orders is in range of 0 to 324.

<Profit distributions for various product categories>

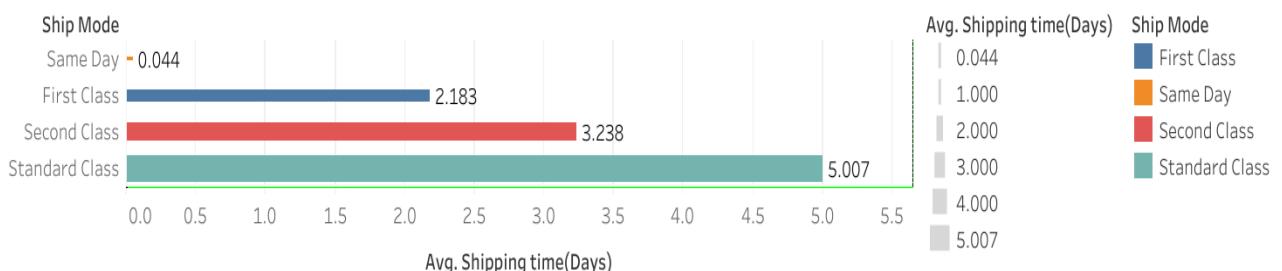


The trend of sum of Profit for Profit (bin) broken down by Category. Color shows details about Category.

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

Yes, we can compare shipping times across different shipping modes, we can use bar graph for it, as it is simple and easy to understand and compare.

<Shipping time distribution for different shipping modes>



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

Line chart is used to plot monthly trend in no. of orders shipped for its ease and clarity in displaying the visual. Also, it is self explanatory. No. of orders shipped decreases from Jan to Feb. From Feb till June, it consistently increases, thereafter consistent decrease from June to

August. A sharp rise is seen in Sep, followed by a precipitous fall in Oct and another rise to highest peak in Nov. Shipments decrease a bit in December.

<Monthly trend in no. of orders shipped>



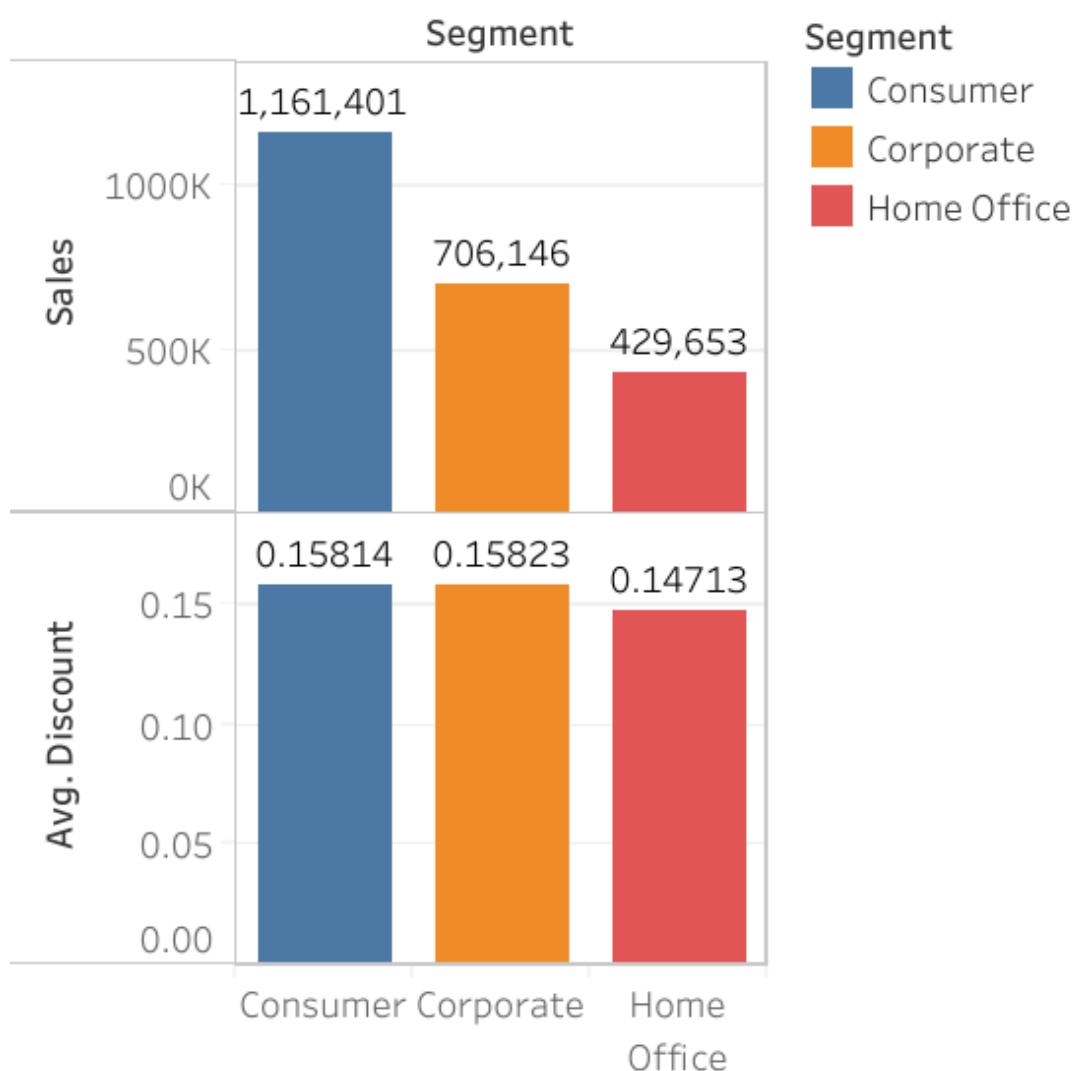
18. How do different customer segments perform in terms of sales and discount rates?

Bar graph is used to compare sales and discount rates for different customers segments. It is easy to visualise and compare.

Maximum sales is obtained from the Consumer segment.

Maximum avg discount is offered to the Corporate segment.

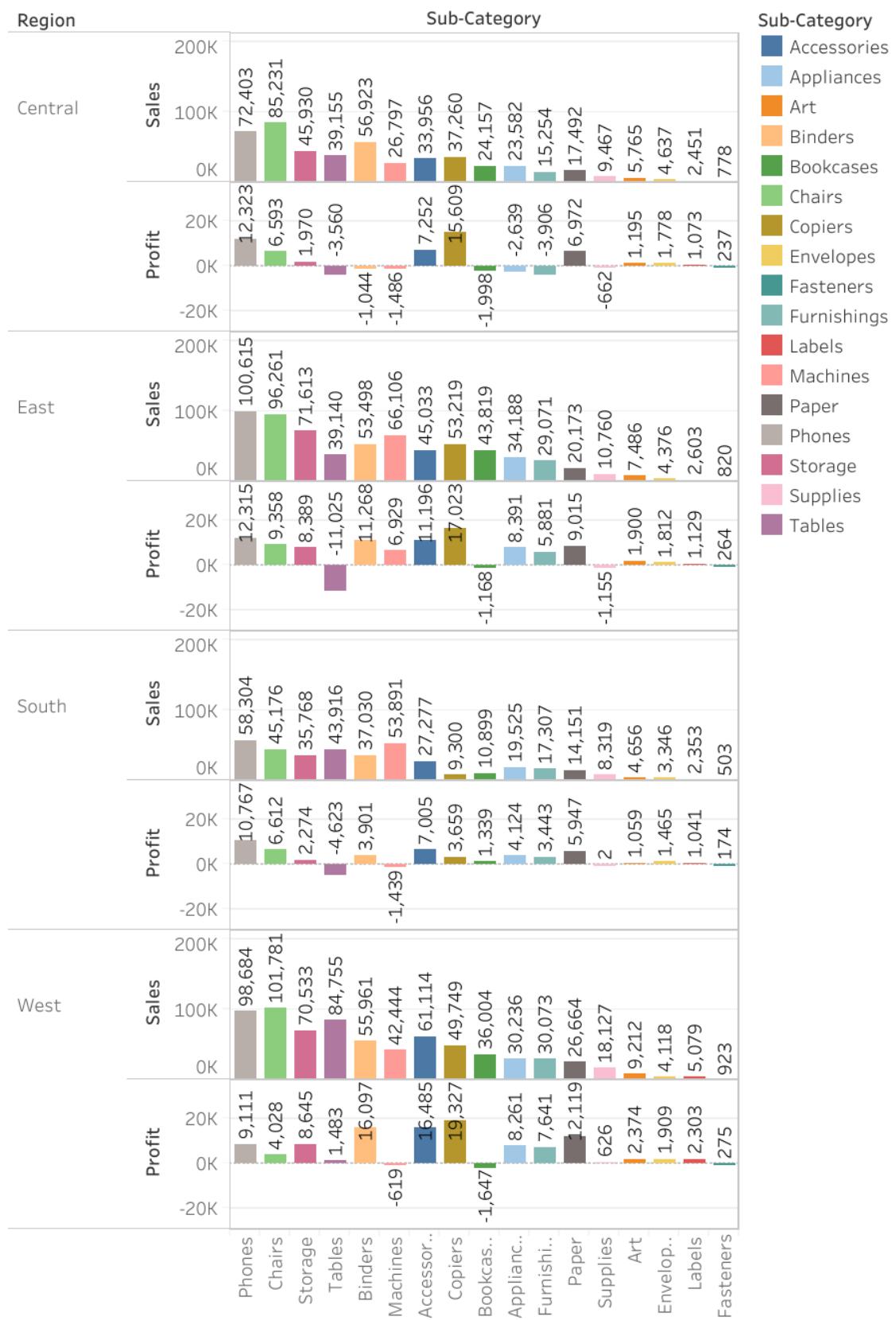
<Sales and discount rates for different customer segments>



19. What are the sales and profit trends across different product subcategories and regions in the Superstore dataset?

Bar graphs clearly show trends in sales and profit for various product subcategories by region.

<Sales and profit trend across different product subcategories and regions>



Central region:-Max sales-Chairs,Max profit-Copiers

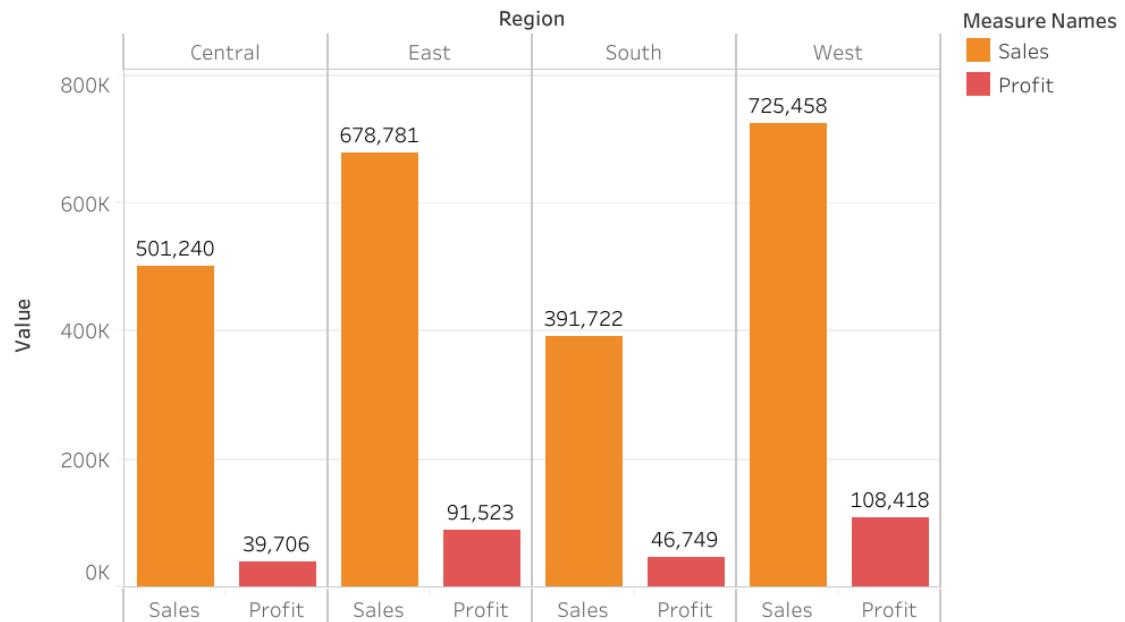
East region:-Max sales-Phones,Max profit-Copiers

South region:-Max sales-Phones,Max profit-Phones

West region:-Max sales-Chairs,Max profit-Copiers

Sales and profit across regions is shown by bar graphs

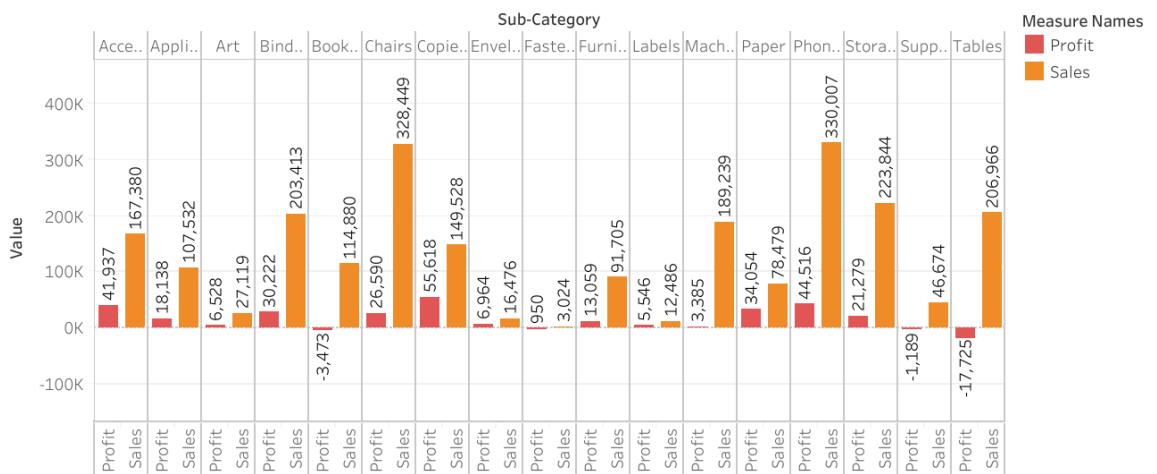
<Sales and profit trends across regions>



Max sales and profit obtained from West region.

Sales and profit subcategories is also shown using bar graphs for ease of comparison

<Sales and profit trends across subcategories>



Max sales from Phones and Chairs.Max profit from Copiers.

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

Bar charts is used for comparisons. Average delivery durations(in days) are as follows.

Central-4.06

West-3.93

East-3.91

South-3.96

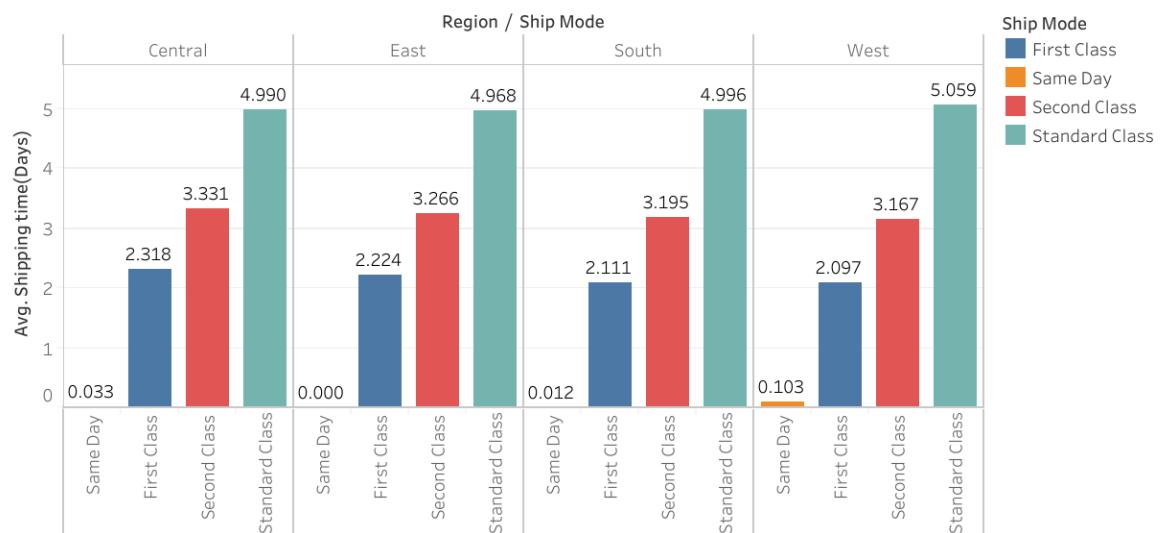
First class-2.183

Second class-3.238

Standard class-5.007

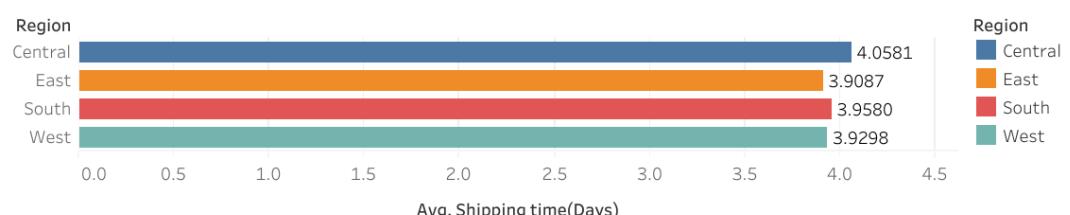
Same Day-0.044

<Average delivery duration for various regions and ship modes>



Taking regions separately,

<Average delivery duration for various regions>



Taking shipping modes separately,

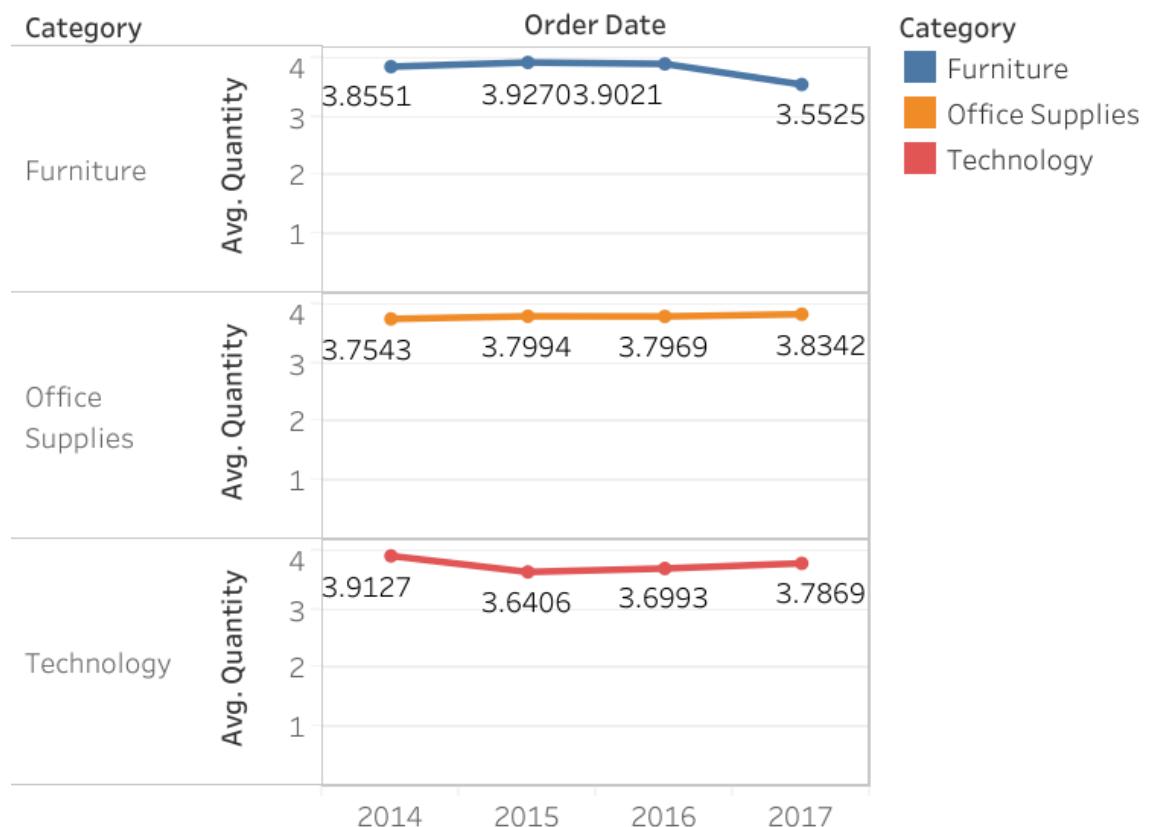
<Average delivery duration for various ship modes>



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

Multi line chart is used to track change in order quantities for each product category.

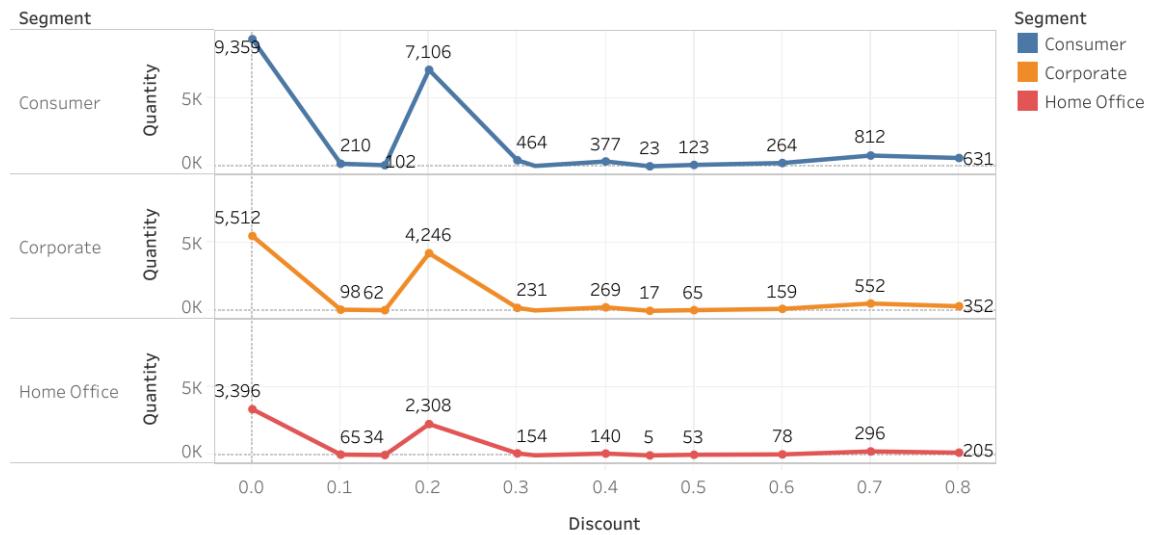
<Avg order quantities over the years for different product categories>



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

Multi line chart is used to correlate discount rates and order quantities for various customer segments. The obtained visualisations are clear and easy to understand and compare.

<Correlation between discount rates and order quantities for different customer segments>



23. What is the proportion of orders returned in each region within the Superstore dataset?

Bar charts are used to compare the proportion of orders returned in each region.

They are as follows:-

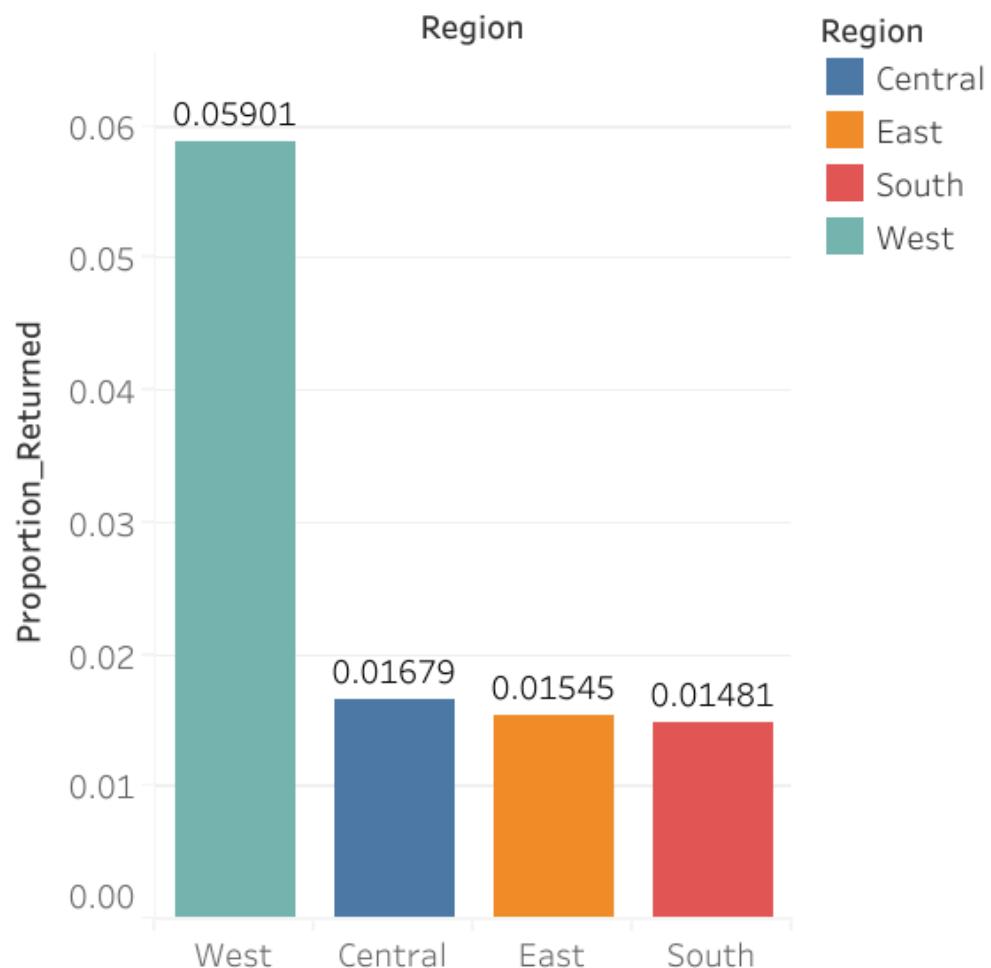
West-0.05901

Central-0.01679

East-0.01545

South-0.01481

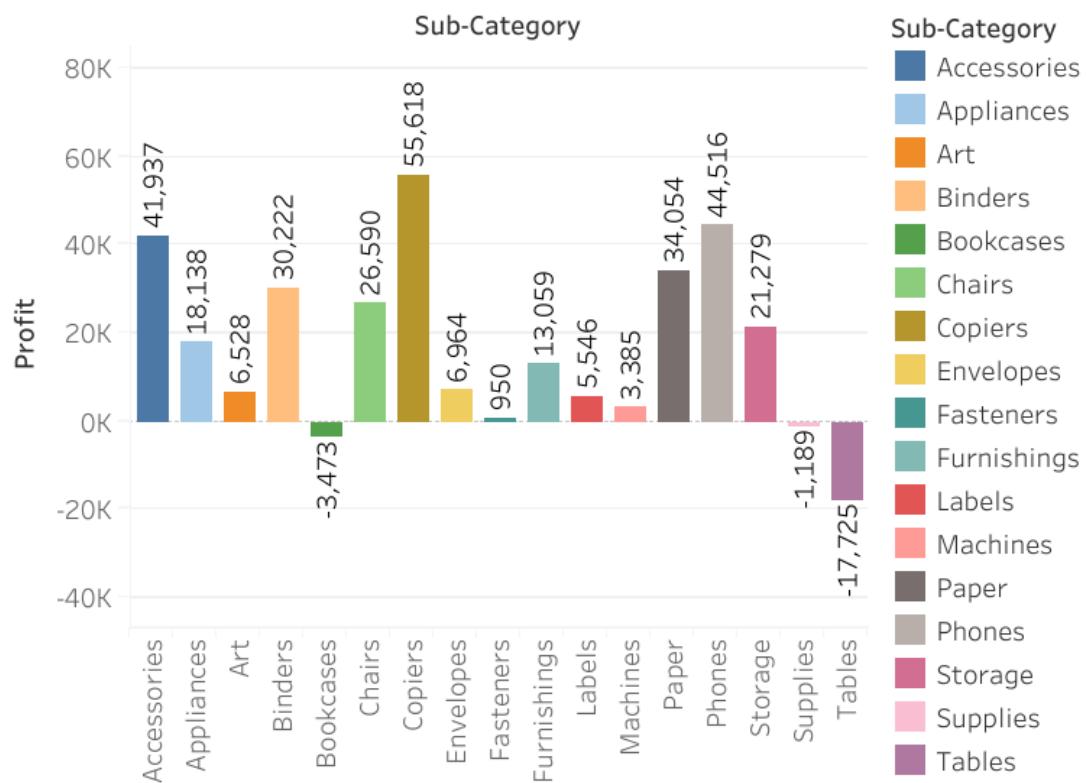
<Proportion of orders returned
in each region>



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

Bar graphs are used to compare profits of different products for different subcategories, it's simple to generate ,visualise and compare.

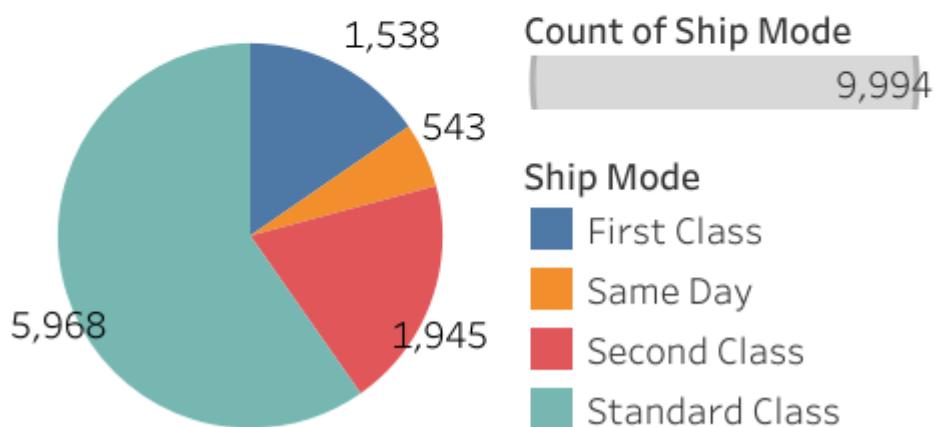
<Profits of different products for different subcategories>



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

Pie chart is used to compare frequency of various shipping modes used. Standard class is the most commonly used shipping mode.

<Frequency of shipping modes>



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

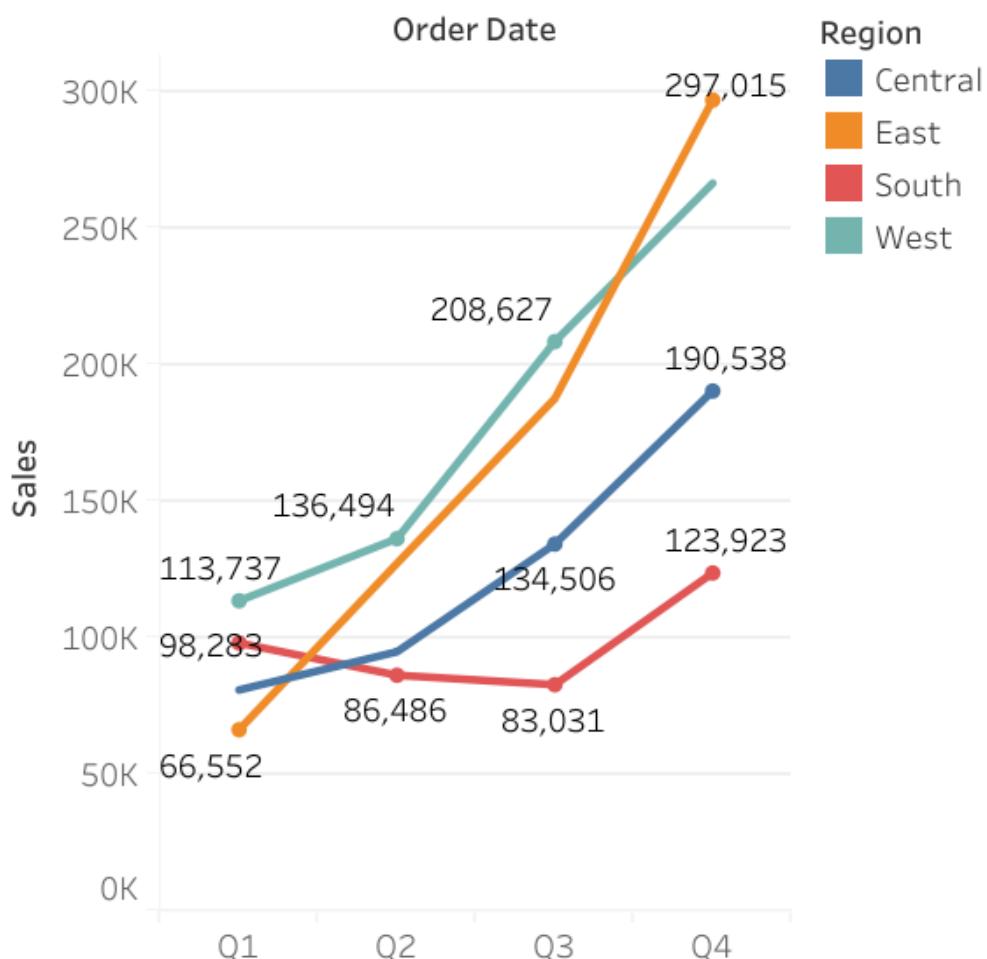
Line charts are used to compare quarterly sales performance of different regions.

In Q1,Q2,Q3 most sales is obtained in west region,in Q4 East gives maximum sales.

West,East and Central regions witness consistent rise in sales from Q1 to Q4.

South experiences consistent fall from Q1 to Q3, but a sharp rise in Q4.

<Quarterly trends in Sales performance of different regions>

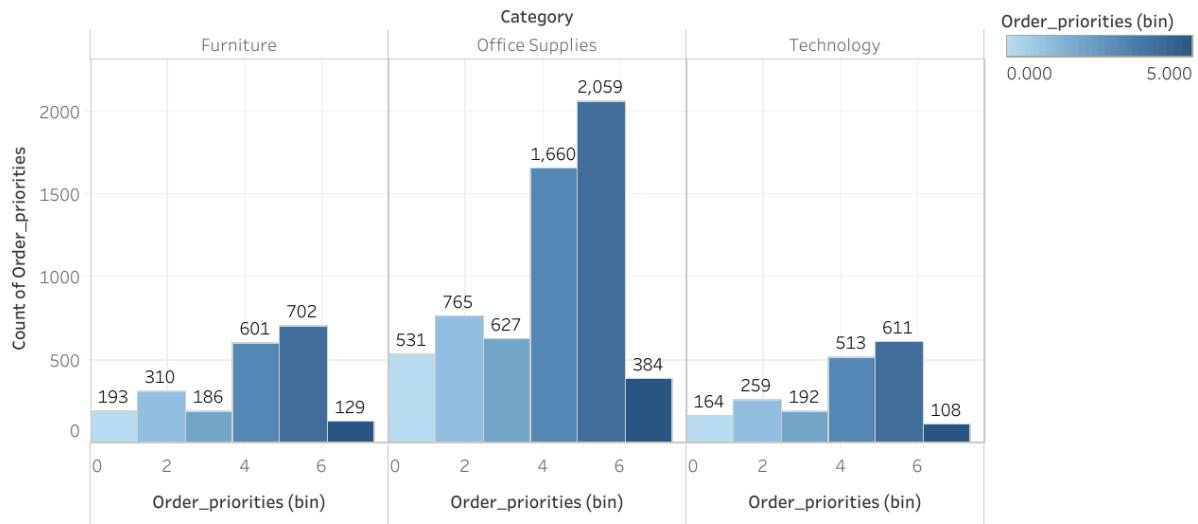


27. What is the distribution of order priorities across different product categories?

Histograms are used to display distribution of order priorities across different categories. It clearly shows the frequency of orders categorised by bins. Most orders are

delivered in between 4.9 to 6.1 days across all categories.

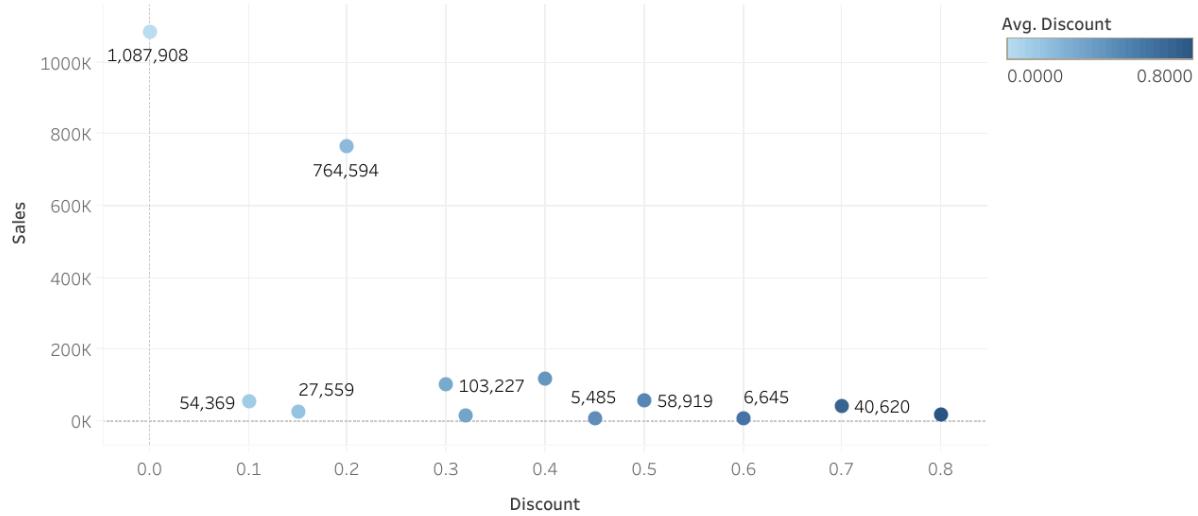
<Order priorities across different product categories>



28. What is the relationship between discounts and sales?

Scatter plot shows the relationship between discounts and sales. From the visual, it's clear that most sales have come when discounts are less.

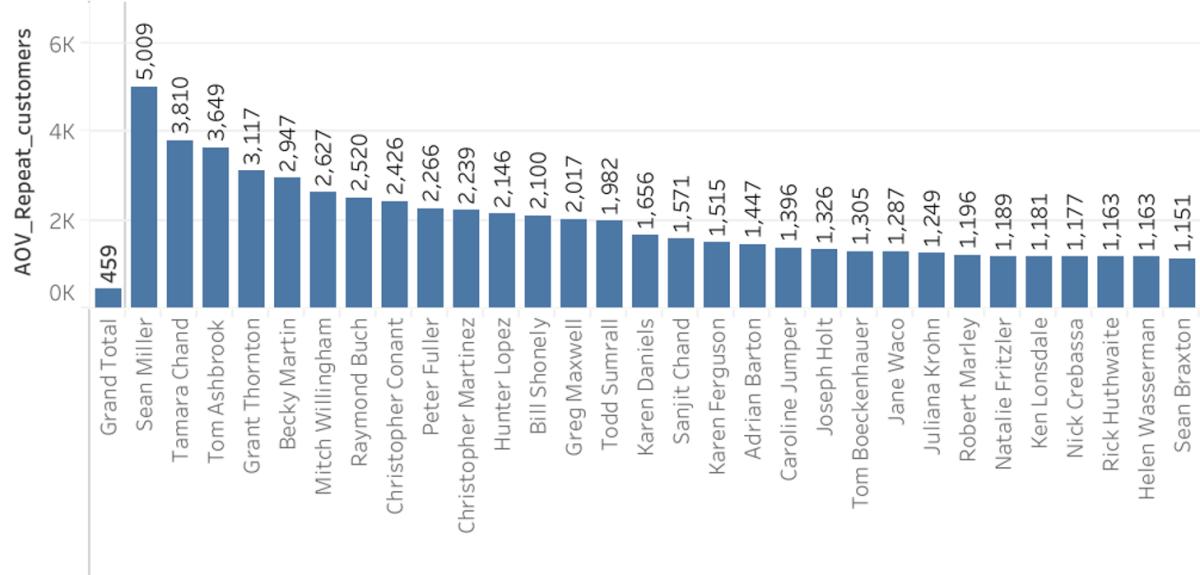
<Relationship between discount and sales>



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

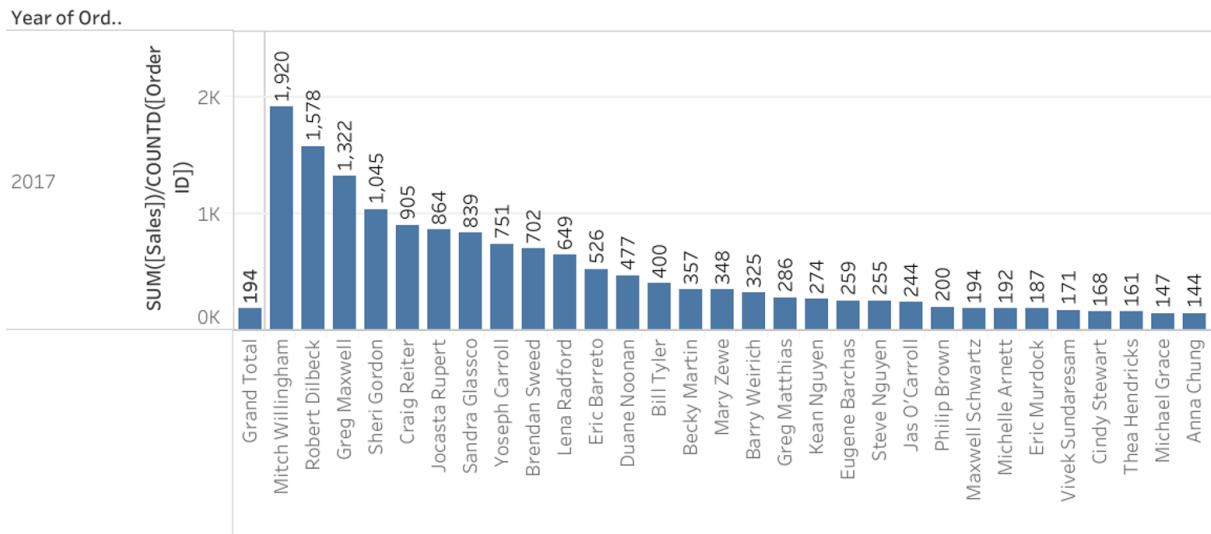
Average order value for repeat customers is displayed using bar graphs for easy comparison. Here is the snap.

<Average order value of repeat customers>



Considering new customers as the ones who ordered their first order in last one year.
We plot average order value of new customers. Here is a snap.

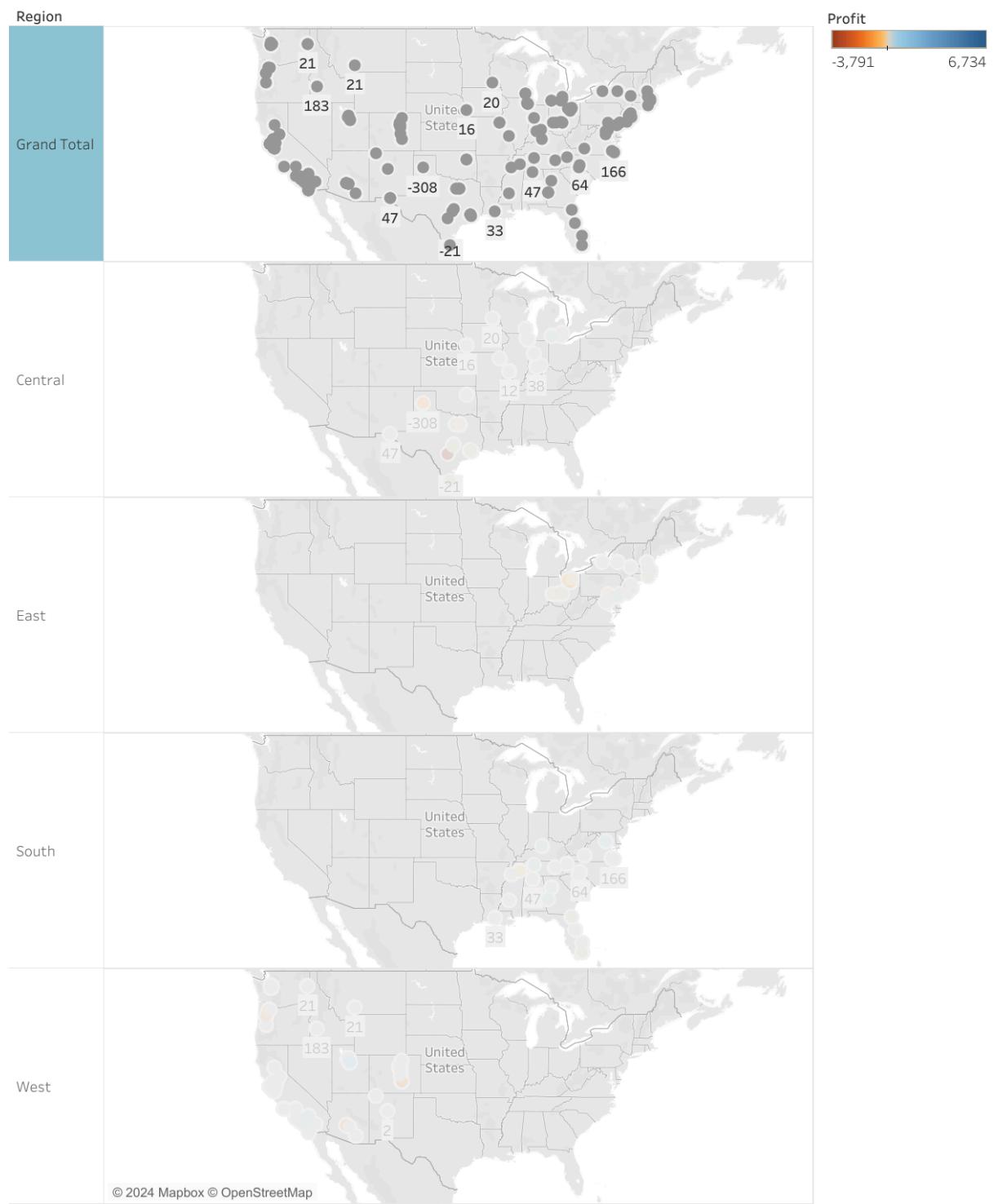
<Avg order value of new customers>



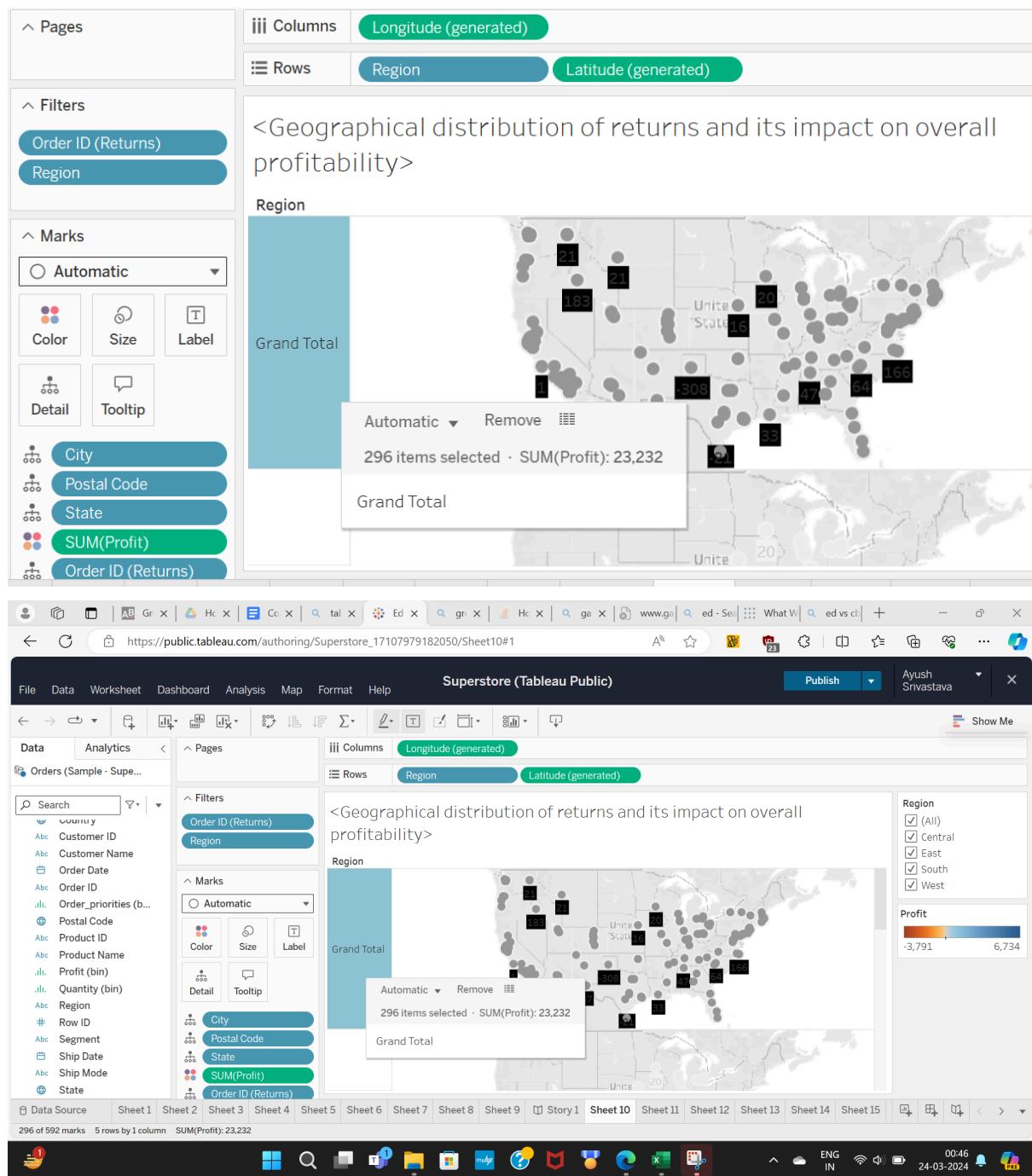
From these two, we can calculate the difference as $459 - 194 = 265$, which is the difference in average order values for repeat customers and new customers.

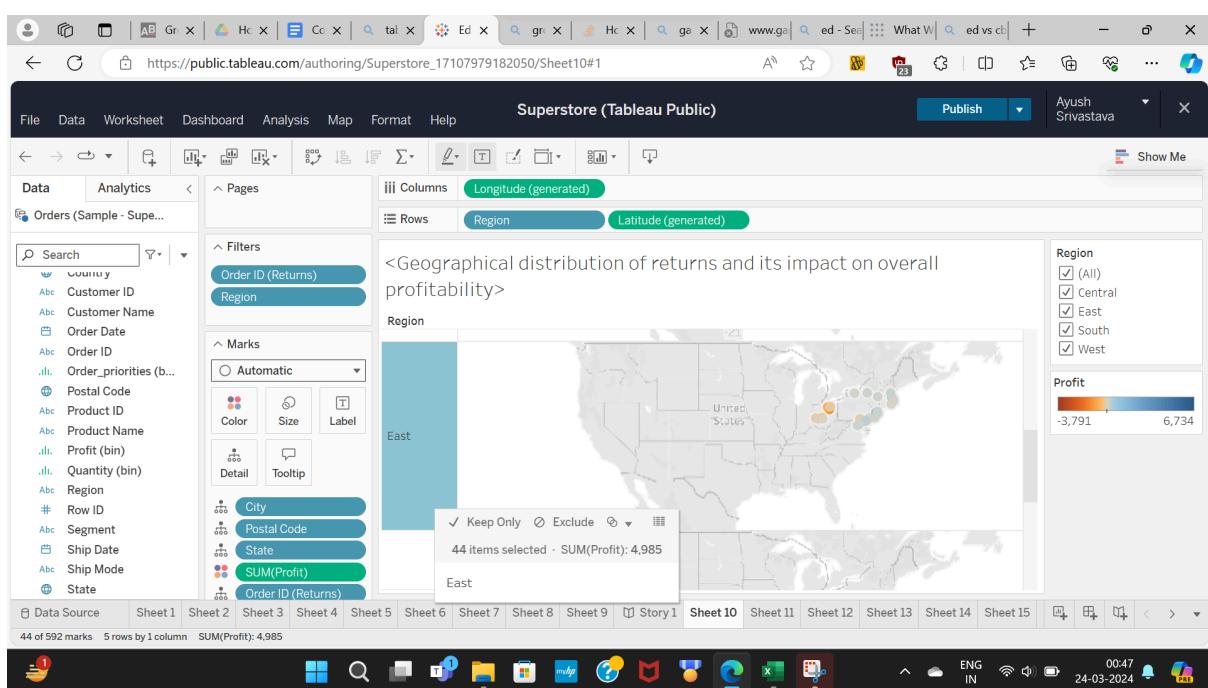
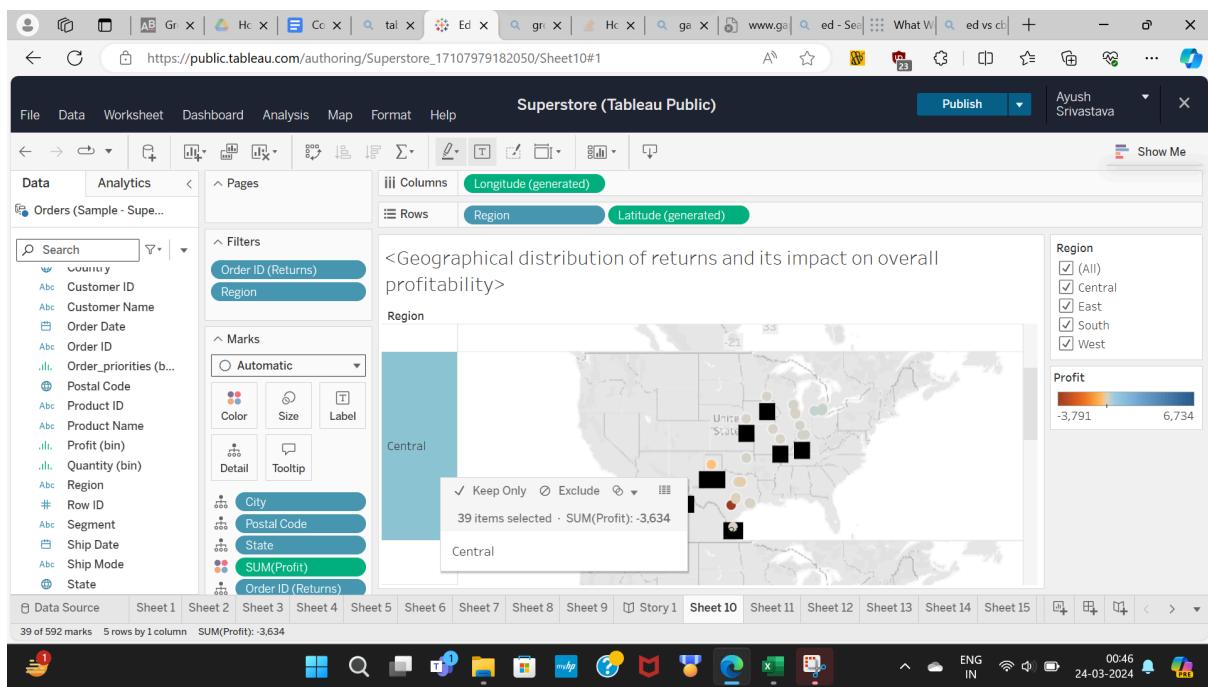
30. What is the geographical distribution of returns and its impact on overall profitability?
A symbol map is used to display geographical distribution of returns and its impact on overall profitability.

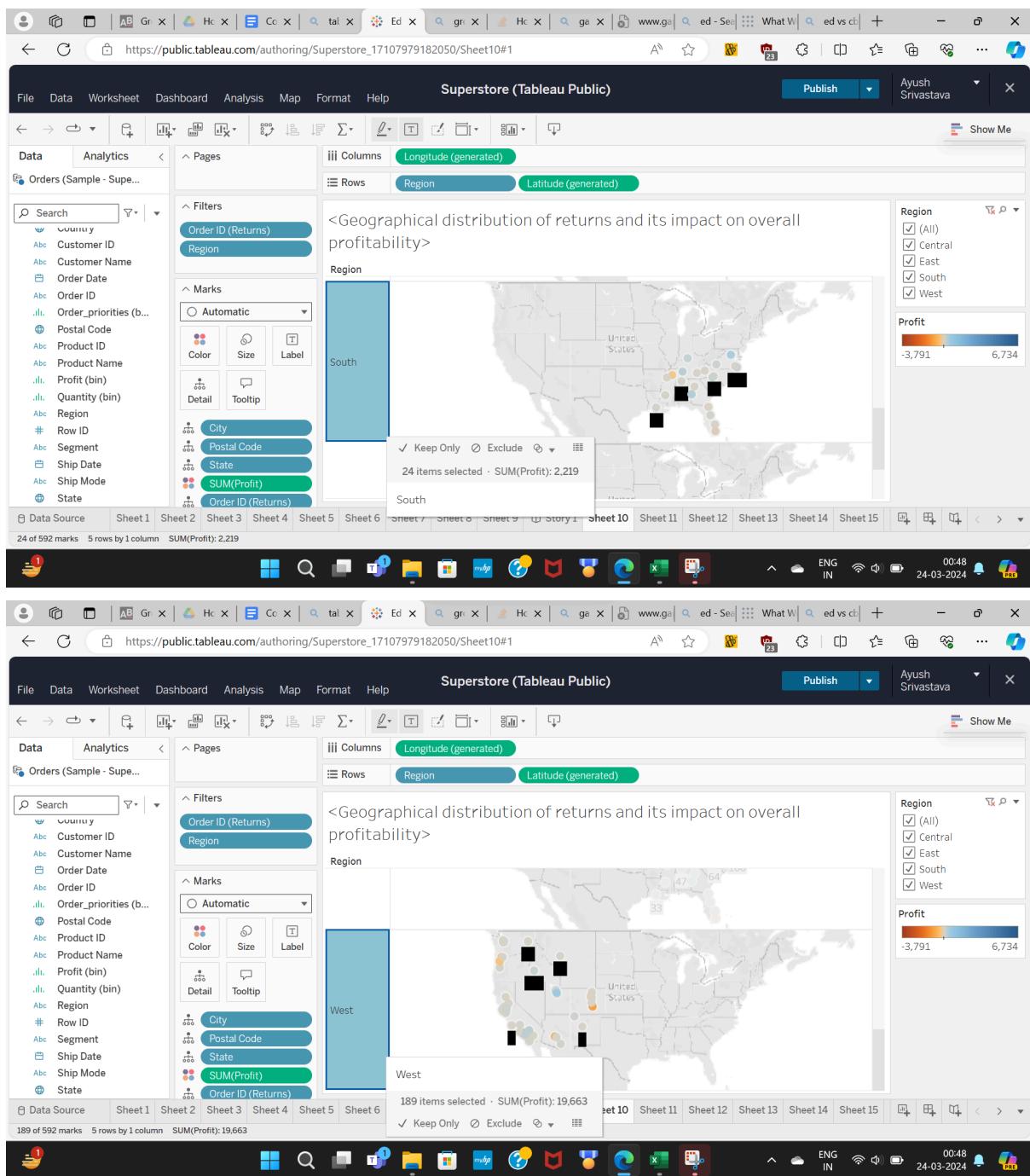
<Geographical distribution of returns and its impact on overall profitability>



Overall effect on profitability is 23,232, which would have added to profits, if orders were not returned.







Regionwise effect on profitability and geographic distribution is shown above.

The End.

Thank You
 Ayush Kr. Srivastava
 Cohort-Chandrayaan