Homework-10

**Out Date:** 11/19/2019 (Tuesday)

**Due Date:** 12/01/2019 (Sunday) 11:59PM

Team#: \_\_\_

Team Member-1:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Member’s Contribution (in %) \_\_

Team Member-2:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Member’s Contribution (in %) \_\_

**Submission**

1. Work on the homework requirements.
2. Prepare your Python file (e.g., HW10\_P1\_Team#.py).
3. Upload the file to blackboard.

**Problem Statement:** A dataset for a superstore (Superstore.xls) is available with this document. The dataset is organized in 23 attributes and 8399 records, containing information about a superstore’s transactions. The following table briefly describes the dataset’s attributes.

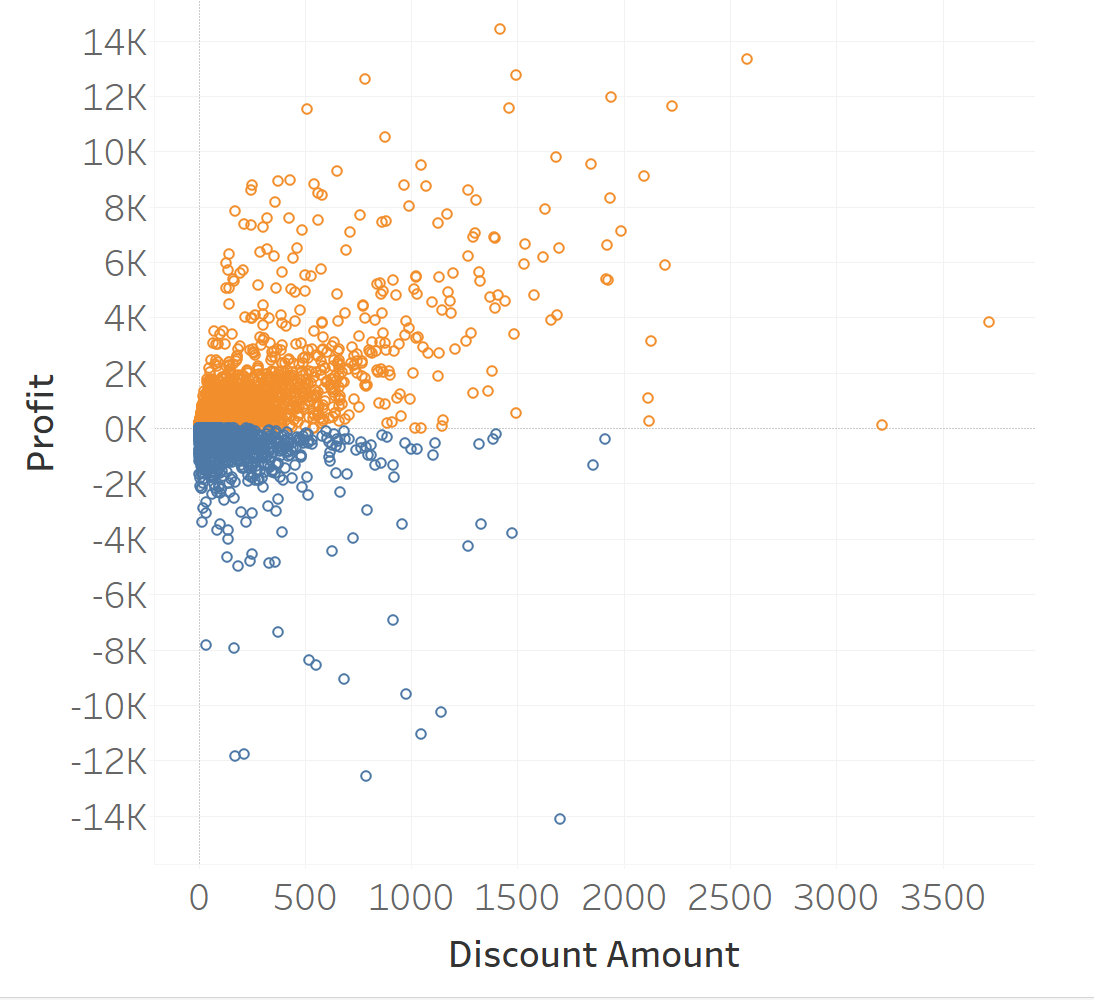
|  |  |
| --- | --- |
| **Attribute** | **Description** |
| Row ID | Transaction ID (Unique numbers) |
| Order ID | Order ID |
| Order Date | Date when the order was received |
| Order Priority | Attribute values: Critical, High, Low, Medium, Not Specified |
| Order Quantity | Quantity of the order |
| Sales | Sales amount |
| Discount | Discount percentage |
| Ship Mode | Attribute values: Delivery Truck, Express Air, Regular Air |
| Profit | Profit amount |
| Unit Price | Cost per item |
| Shipping Cost | Shipping cost per order |
| Customer Name |  |
| City |  |
| Zip Code |  |
| State |  |
| Region |  |
| Customer Segment | Attribute values: Consumer, Corporate, Home Office, Small Business |
| Product Category | Attribute values: Furniture, Office Supplies, Technology |
| Product Sub-Category | Attribute values: 17 values (Appliances, Bookcases, etc.) |
| Product Name |  |
| Product Container | Attribute values: 7 values (Jumbo Box, Jumbo Drum, etc.) |
| Product Base Margin |  |
| Ship Date | Date when the order was shipped |

Write a Python script to visualize the following the information:

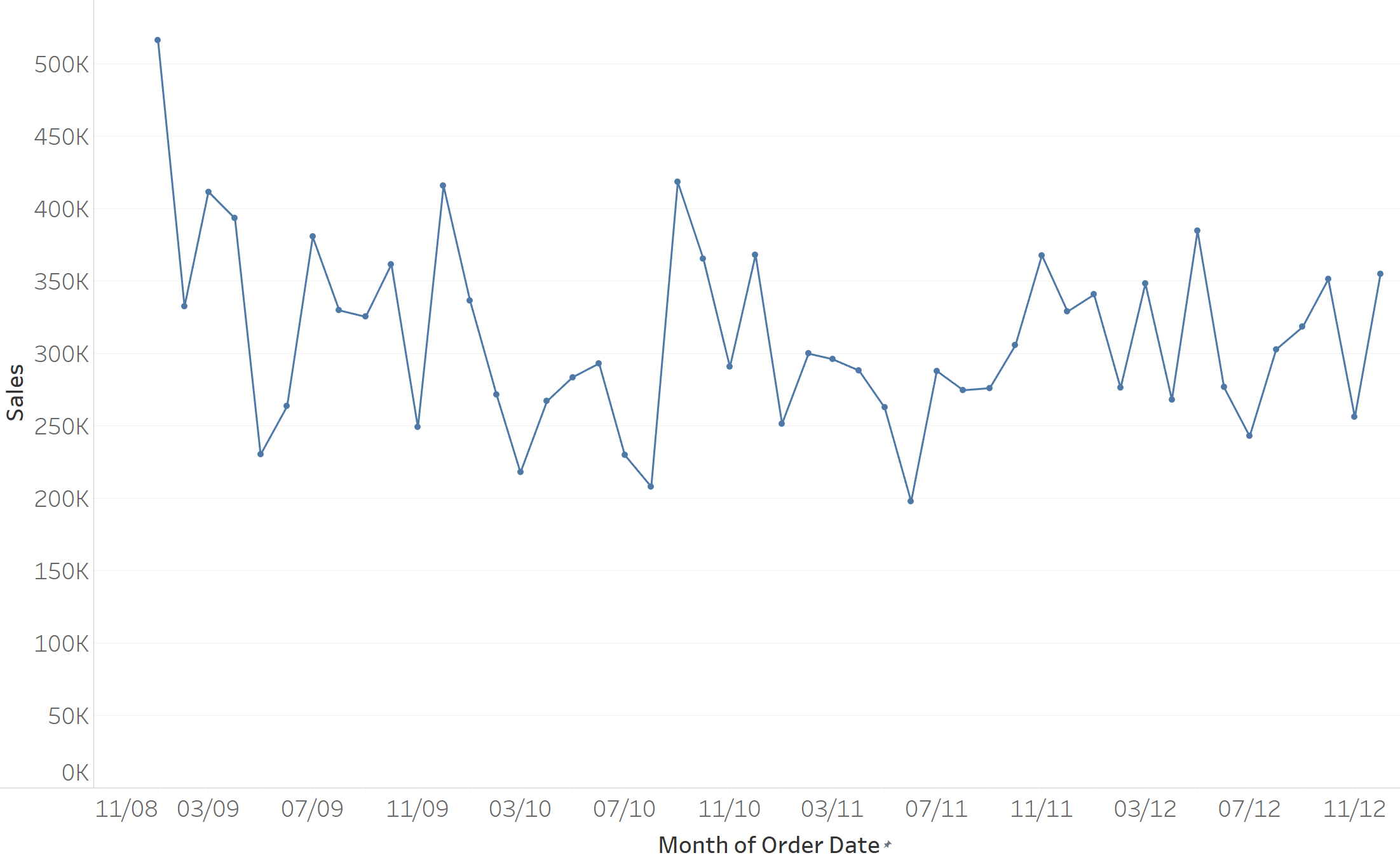
1. The discount is given in percentage of the sales amount. Compute the discount in dollar amount for each transaction, and create a scatter plot of profit and discount amount. Each dot represents one transaction (one row). **[10 points]**



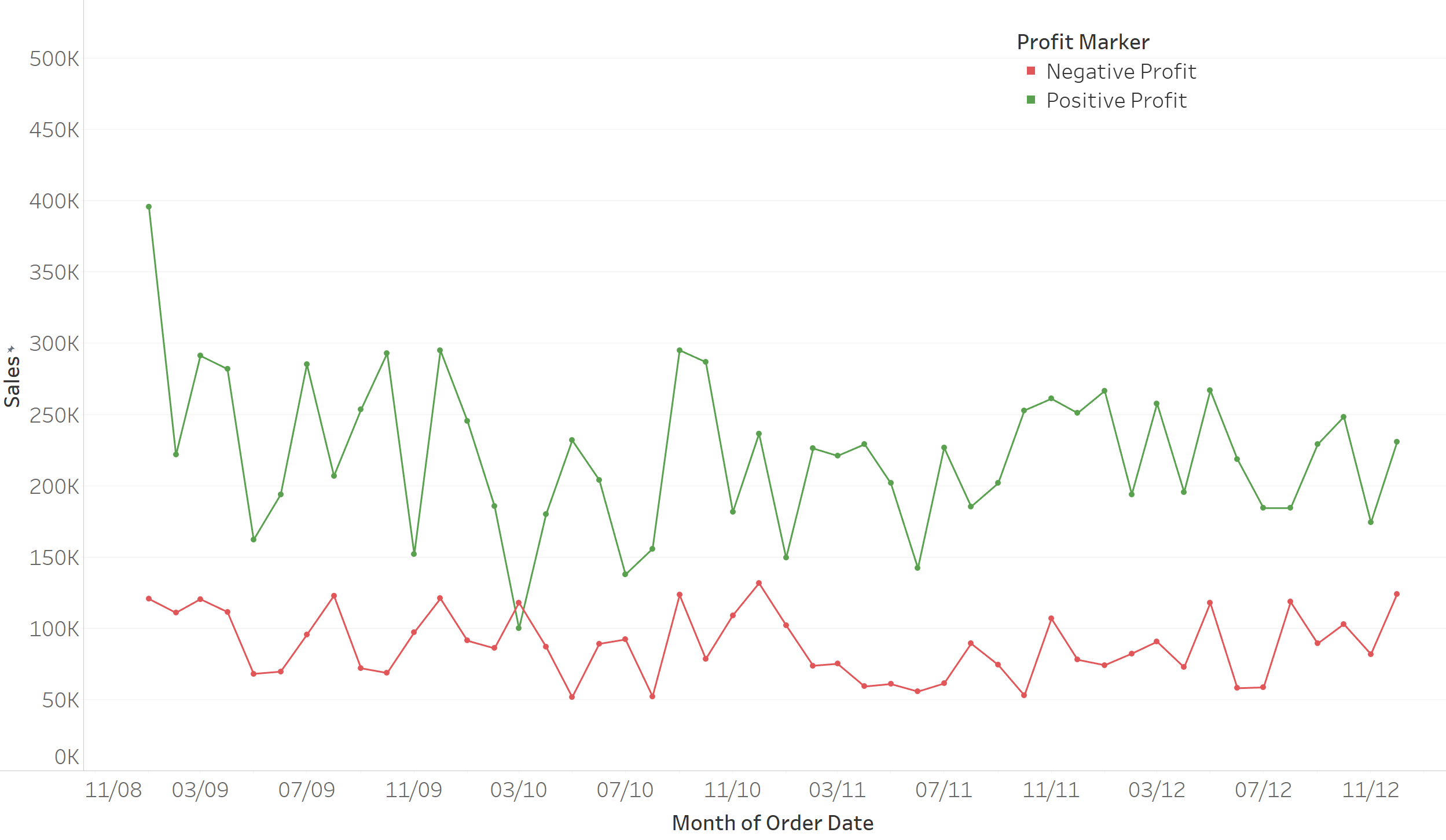
1. Modify the visualization in item-2 to group the transactions according to the profitability. In particular, group all the profitable transactions in an orange color and all unprofitable transactions in a blue color as shown in this figure. **[15 points]**



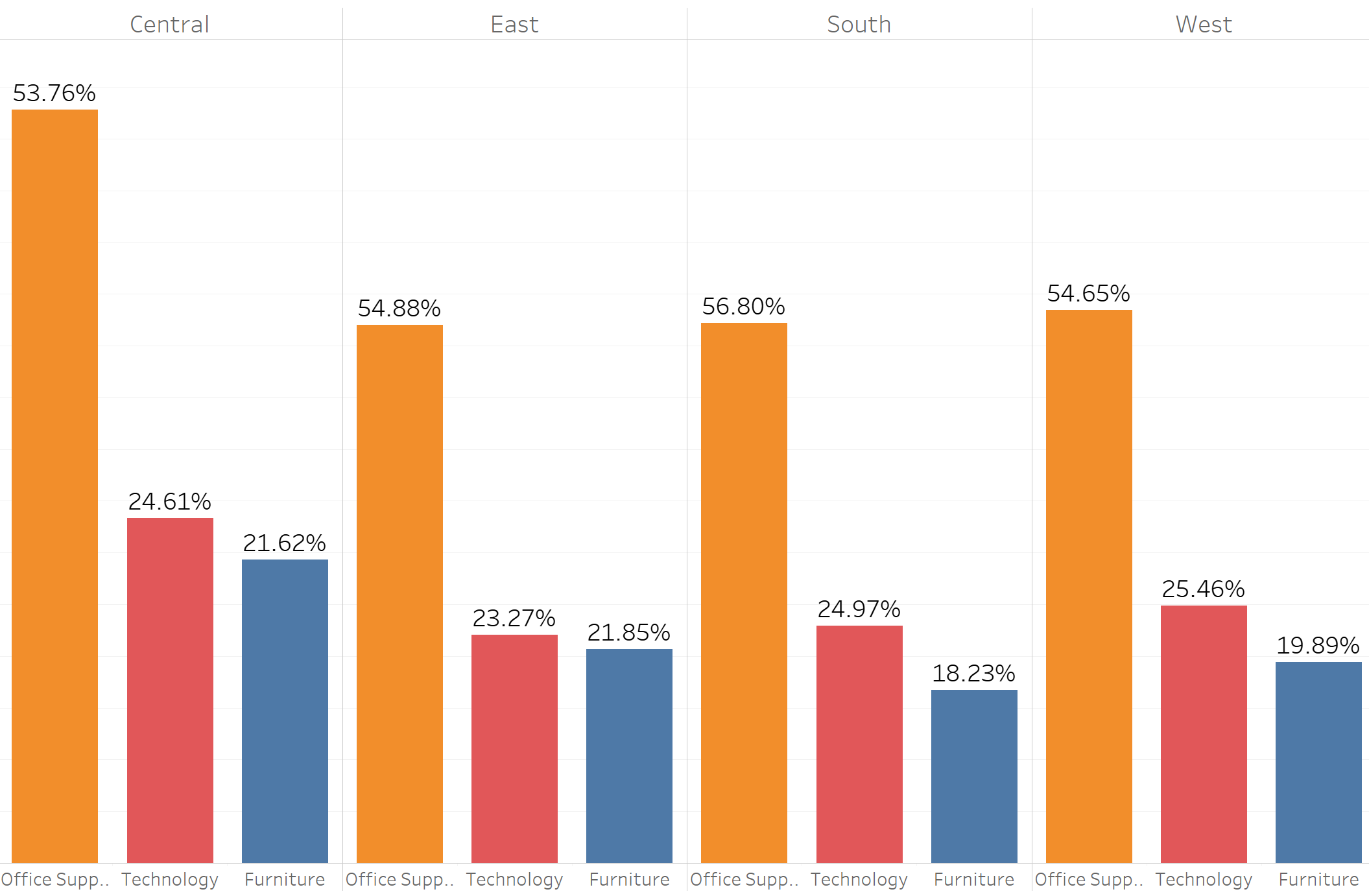
1. Display Sales by the Order Date (Month/Year). **[10 points]**



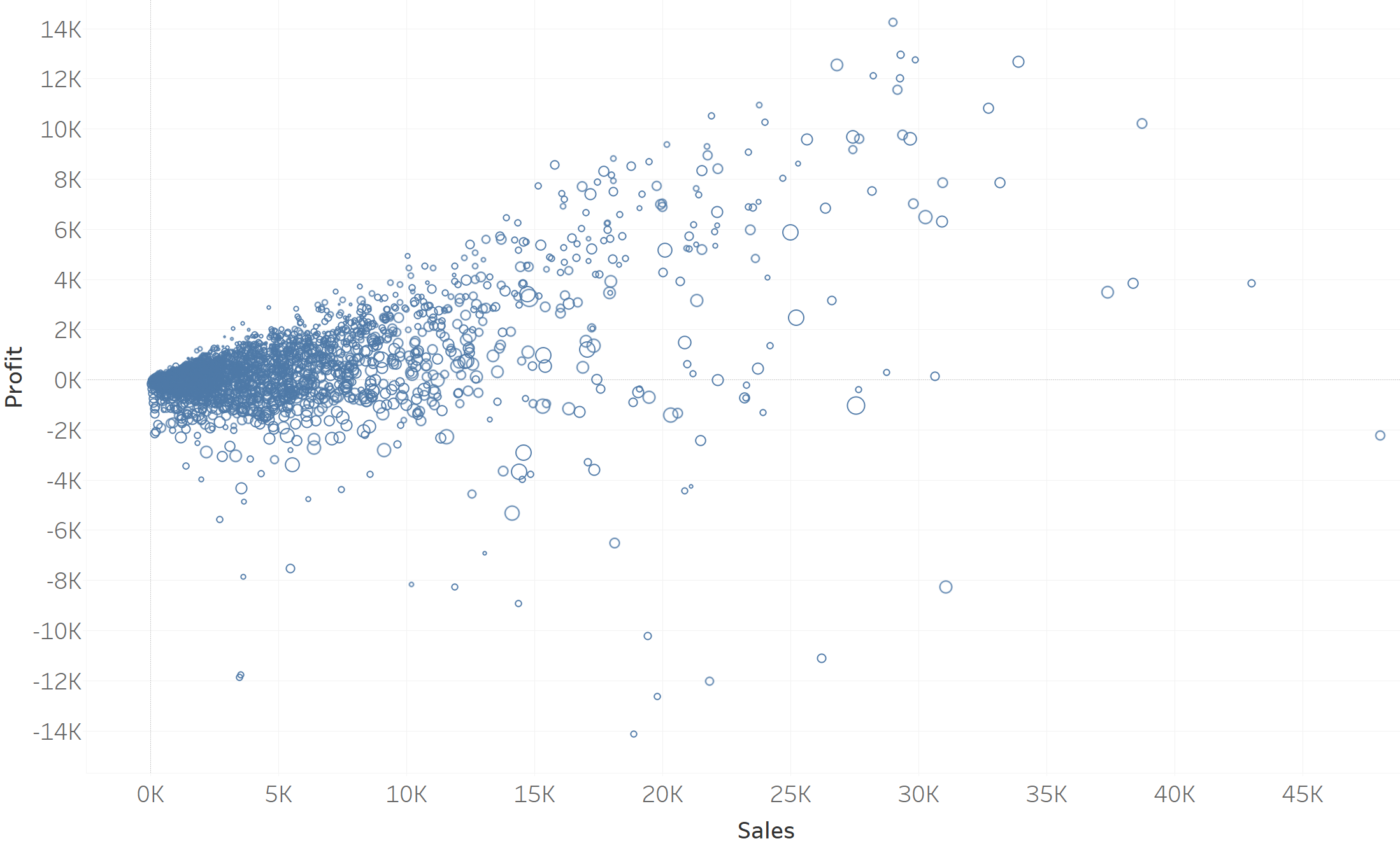
1. Modify the visualization in item-3 to group Sales according to the profitability. In particular, group all the profitable Sales (positive profit) per month and color them in the green color. Group all the unprofitable Sales (negative profit) per month and color them in the red color. **[15 points]**



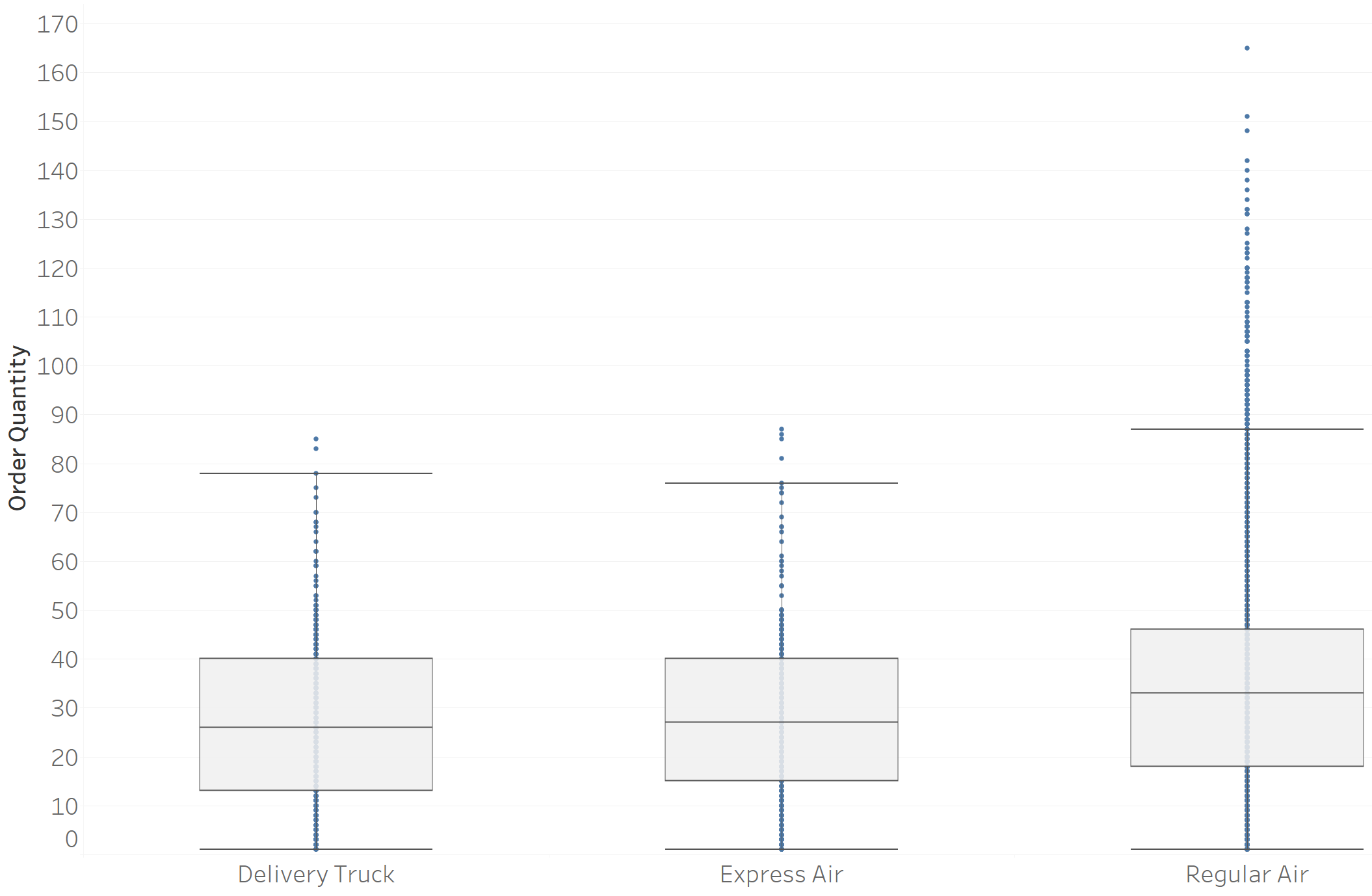
1. Plot a bar graph to show what percentage of the products are Technology, what percentage of the products are Office Supplies and what percentage of the products are Furniture per region in the dataset. **[10 points]**



1. Visualize the relationship between these three variables: profit, sales, and shipping cost. Create a scatter plot as shown below in which the size of the dot represents shipping cost. Each dot represents one transaction (one row). **[10 points]**



1. Display order quantity by shipping mode via boxplots as shown below. **[10 points]**



**Note:** The plots are generated in Tableau. Therefore, use them as reference only. Your plot may look a bit different from these plots.

Use matplotlib, pandas, seaborn, and numpy for data visualization. Submit your Python script on blackboard. The script should be self-contained, meaning it should read the data file (Superstore.xls) from the folder where the script resides and generates plots. The script should save each plot in the vector-graphics format (pdf file). **[10 points]**