Mayur Sejpal (C2071044)

University of Law

Data Visualisation

Prof. Varun Gupta

May 18th, 2023

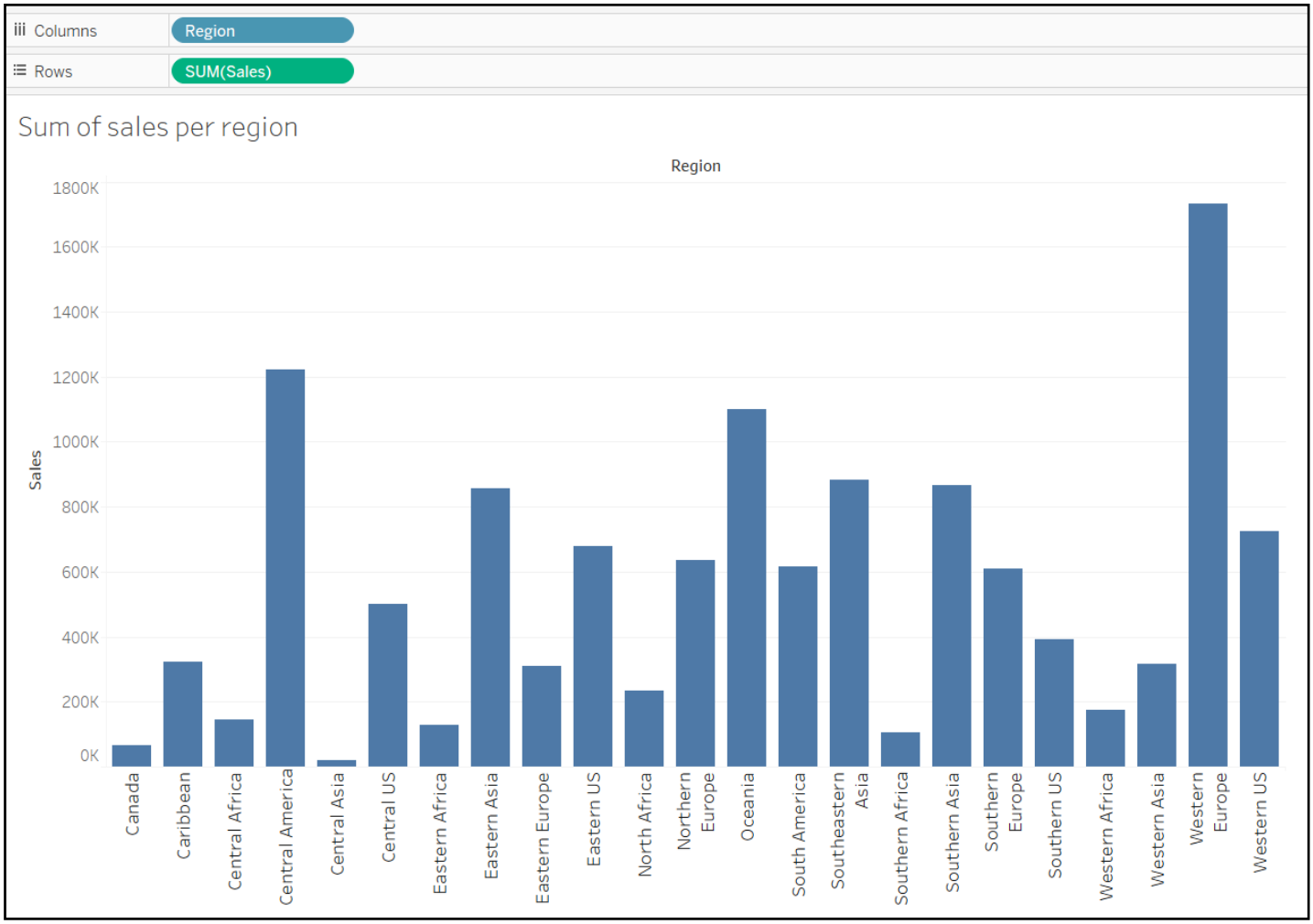
Contents

Portfolio activity 1

Your task is to produce the following diagrams in Tableau. For this activity, you will need to use the Global Superstore 2018.

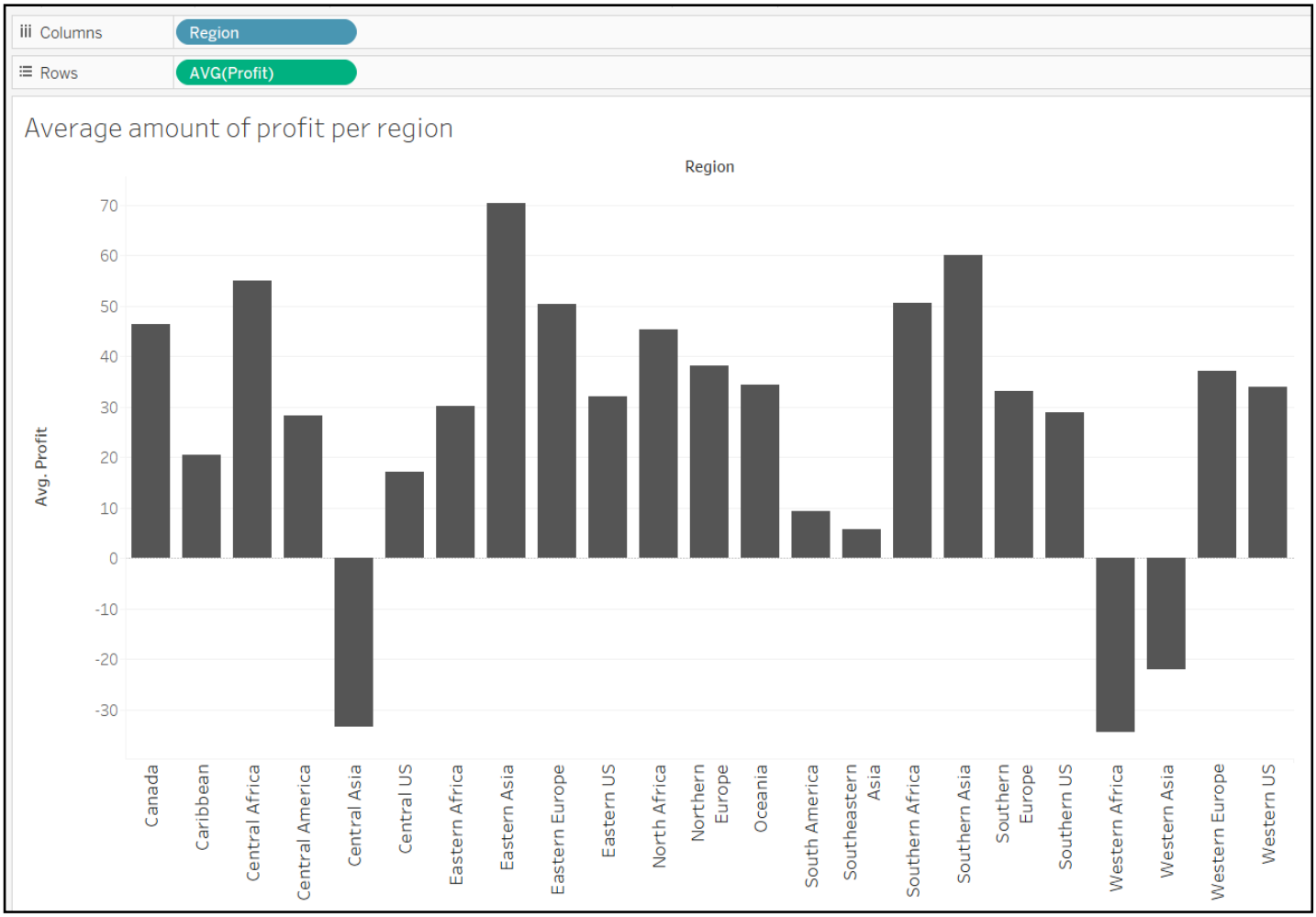
- Plot the regions and the sum of sales per region in a bar chart as shown below.

Bar Chart



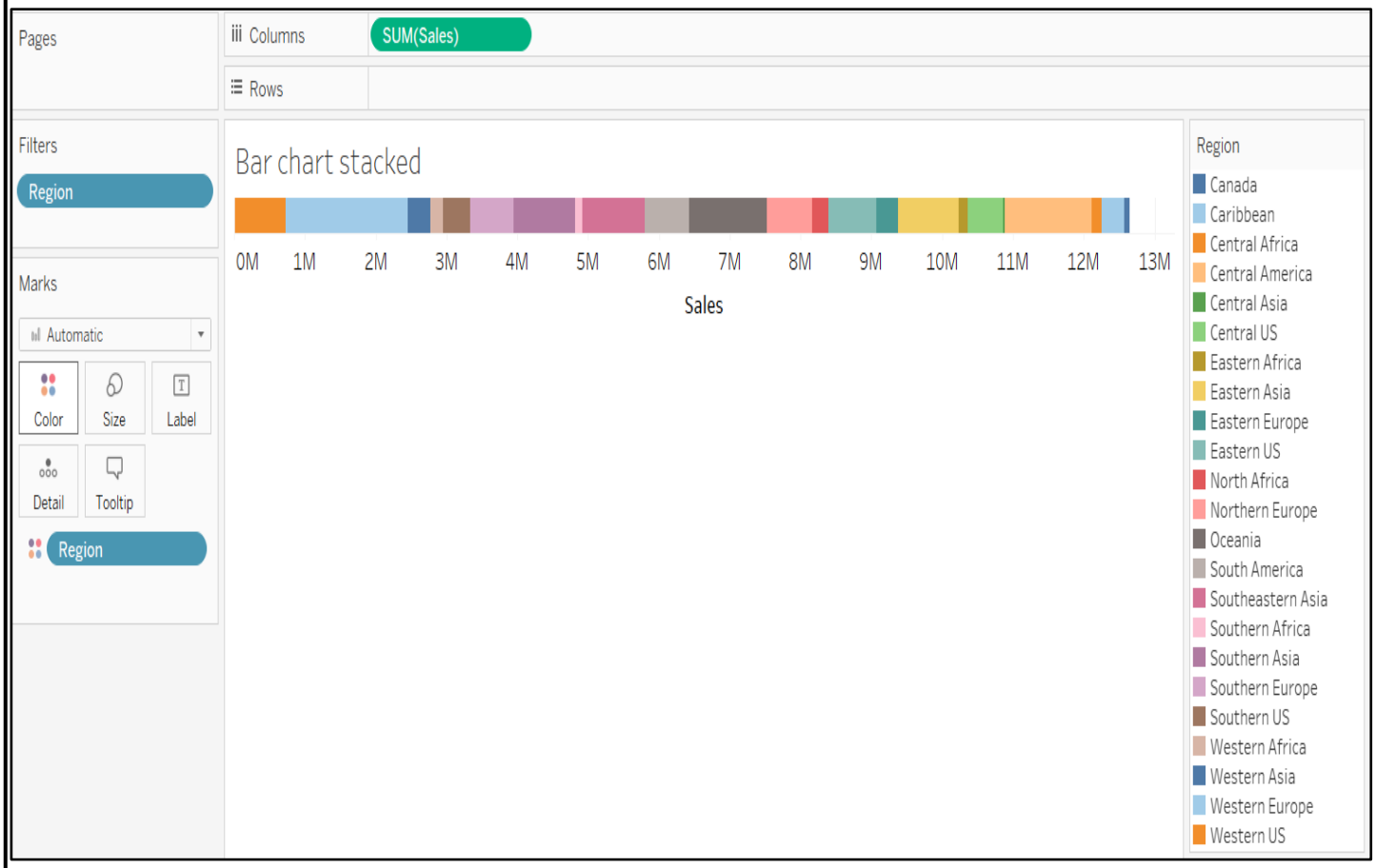
❖ Plot a diagram to demonstrate the correlation between the average amount of profit per region.

Bar Chart (Average basis)



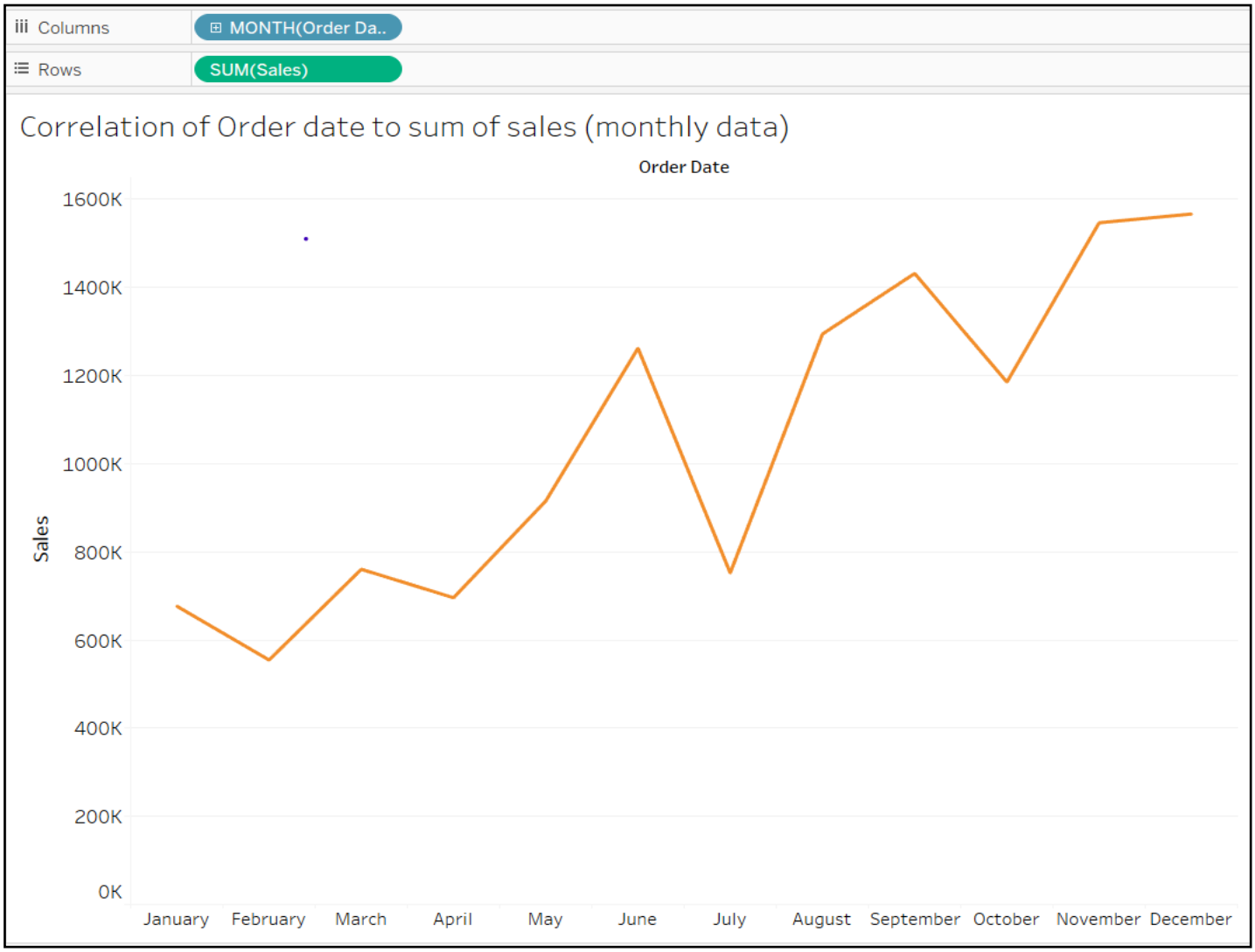
❖ Create a bar chart stacked using the amount of sales per region as shown below.

Stacked Bar Chart



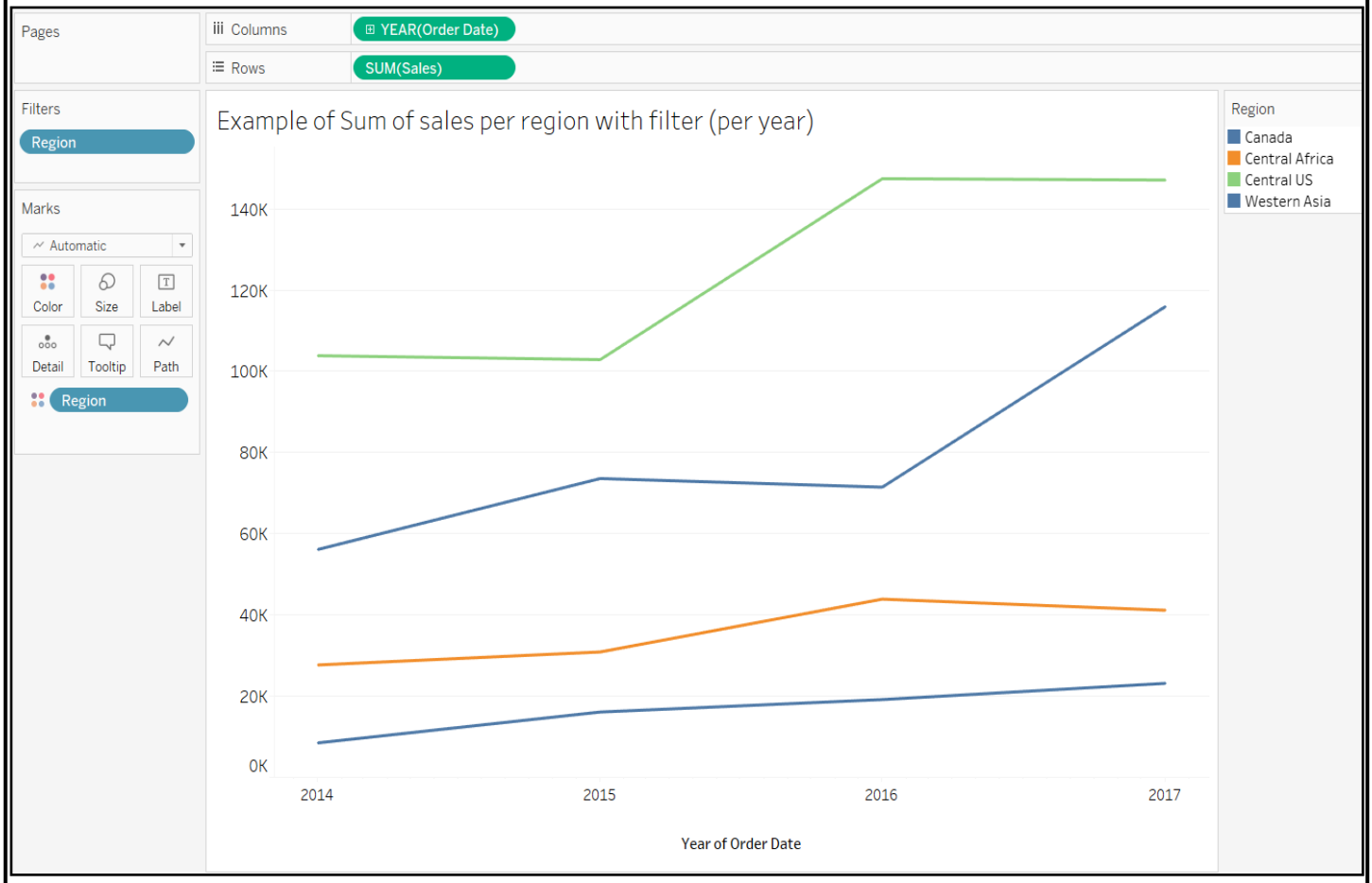
❖ Create a line graph to plot of Correlation of the order of date to the sum of sales (monthly data), as shown below.

Single Line Graph



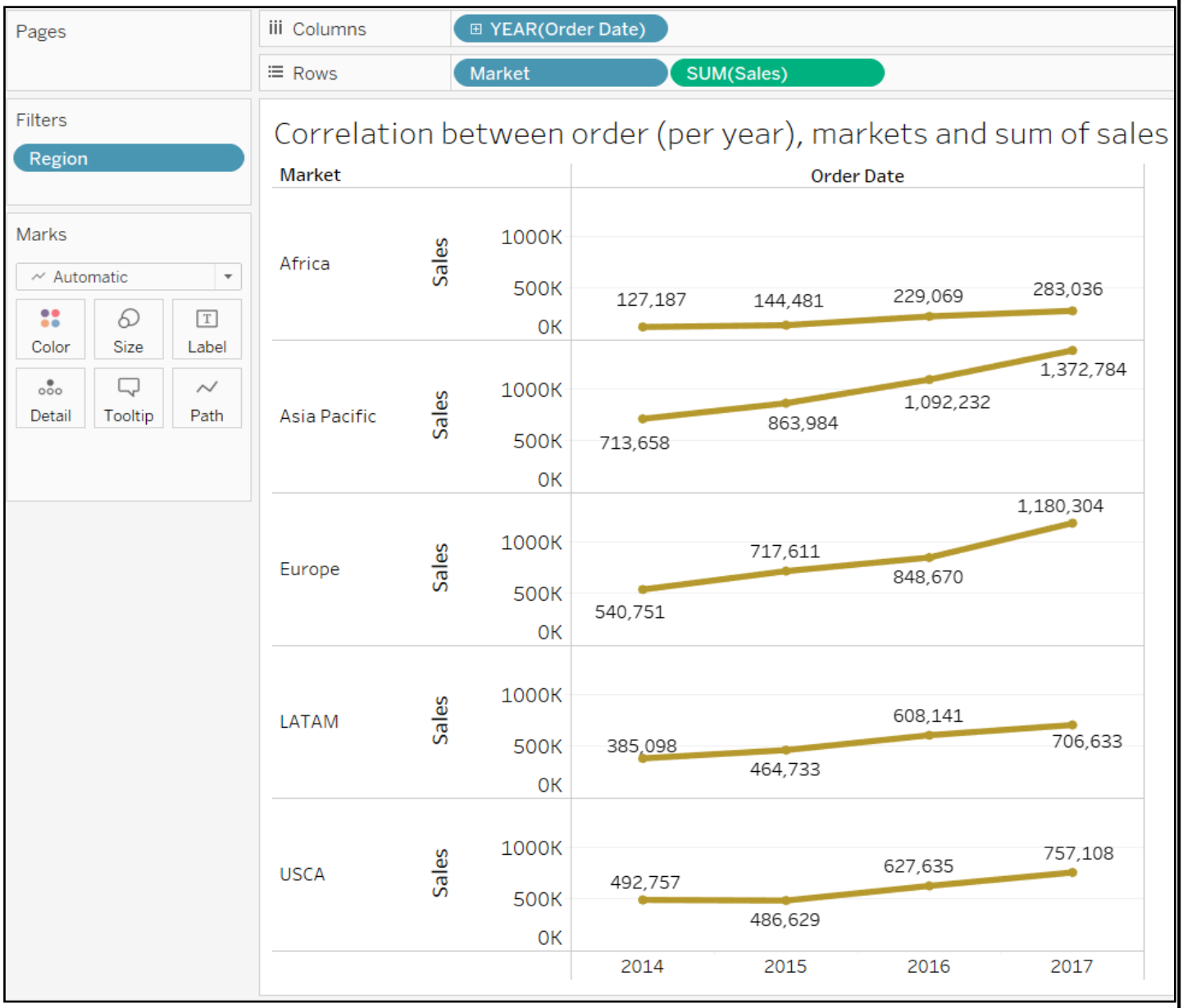
❖ Plot a diagram to demonstrate the correlation of sales per order date (year) for different markets, as shown below.

Continues Line Graph



❖ Plot a diagram to demonstrate the correlation of sales per order date (year) for different markets, as shown below.

Multiple Continues Line Graph



Portfolio activity 2

Your task is to produce the relational model using the Portfolio Dataset file. Generate a relational model in the Data Source of Tableau.

- You will need to utilise the following Excel sheets (tables):
  1) Admins
  2) Classrooms
  3) Enrolments
  4) Modules
  5) Students
  6) Teaching assistants
  7) Teaching staff modules
  8) Tutors

- Provide a screenshot of your data model in Tableau and discuss each relationship between the associated sheets. Include screenshots to demonstrate the specific associations for the fields used in the relationships.

- Provide an analysis of your rationality in developing the data model. Since there might be various ways to relate datasets, analyse your decision to use the selected fields.

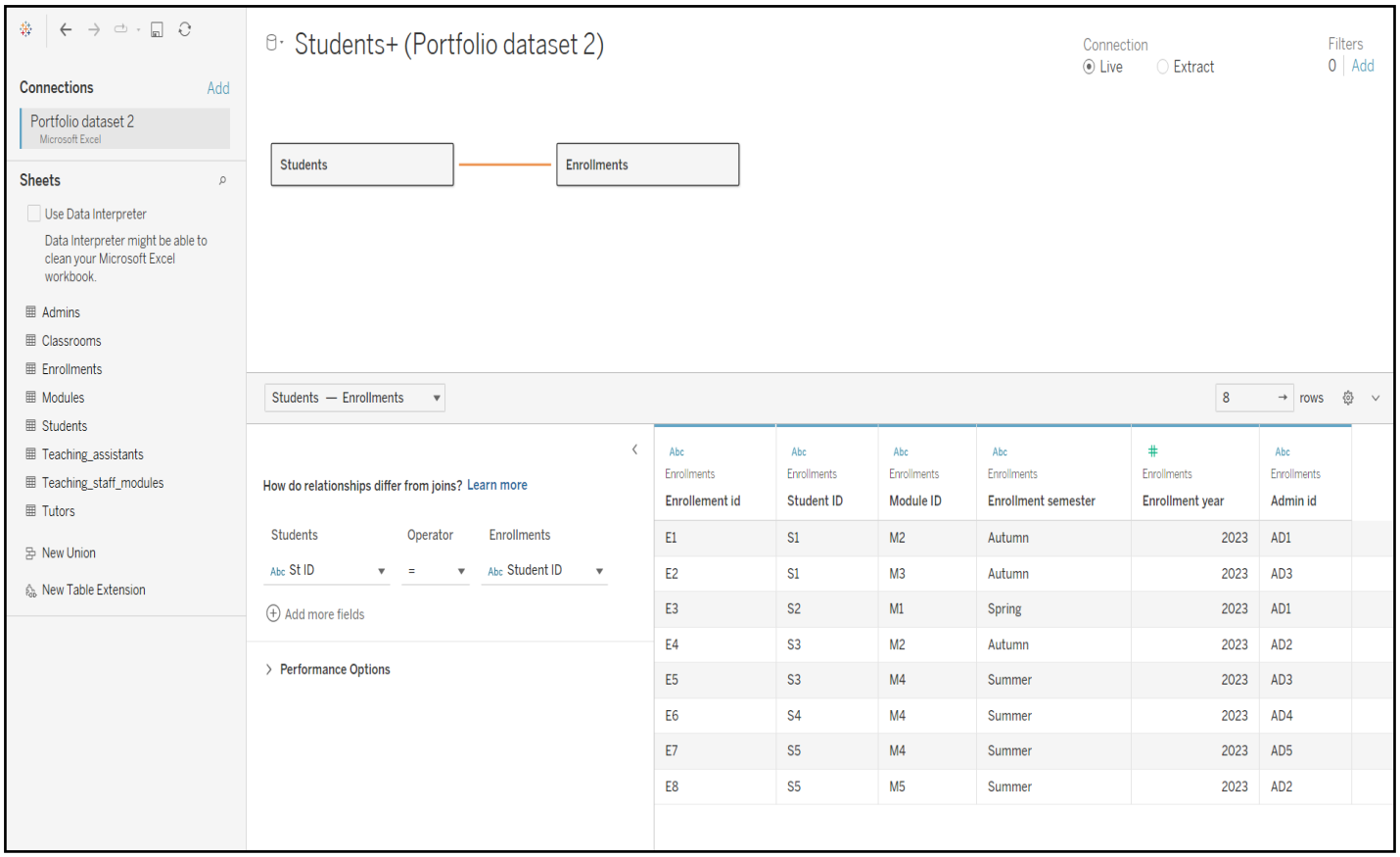
1st Relational model

Relationship between ***Students*** and ***enrolments*** is defined using "Student_Id"

The student table specify the detail about the student (i.e Student Name, address, city, postal code, Student id as a general details). Where as in Enrollment table we can find all other detail for which student can enrol (i.e enrolment year, enrolment semester and enrolment id along with student id).

We can see that student id is common in the both the table which tries to link the information between two separate table using primary key called" Student_id" thus, I will link these two tables which would be synchronised a common data used for Visualization.

After linking both the tables, it would get synchronized, and thus one can see the complete data about the student: Enrollment id, module id, Enrollment year, Enrollment semester, along with the individual details of the student like name, address, city, and postal code. This would allow one to see and analyse data under one parametric component in a simple and easier way under one database used for visualization instead of preparing two separate visualization graphs and attempting to link them to interpret the final outcome.

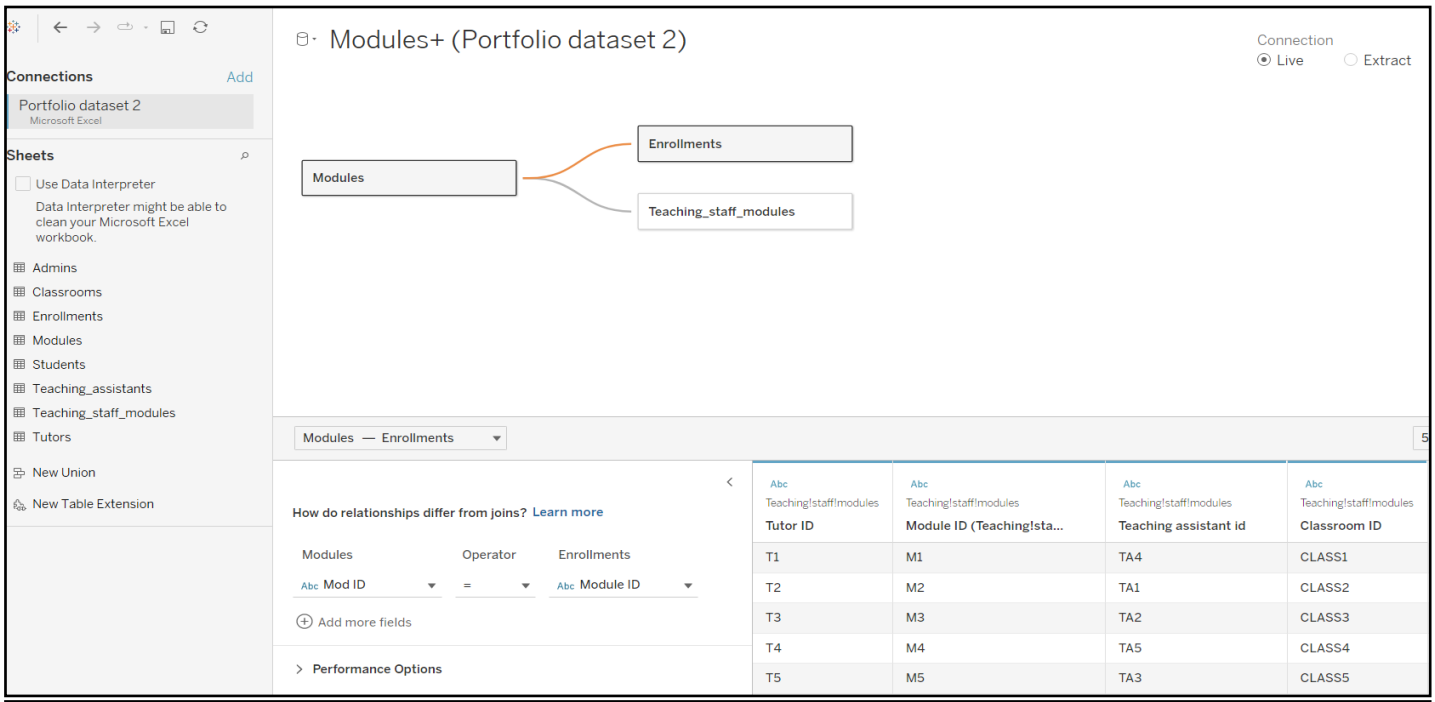


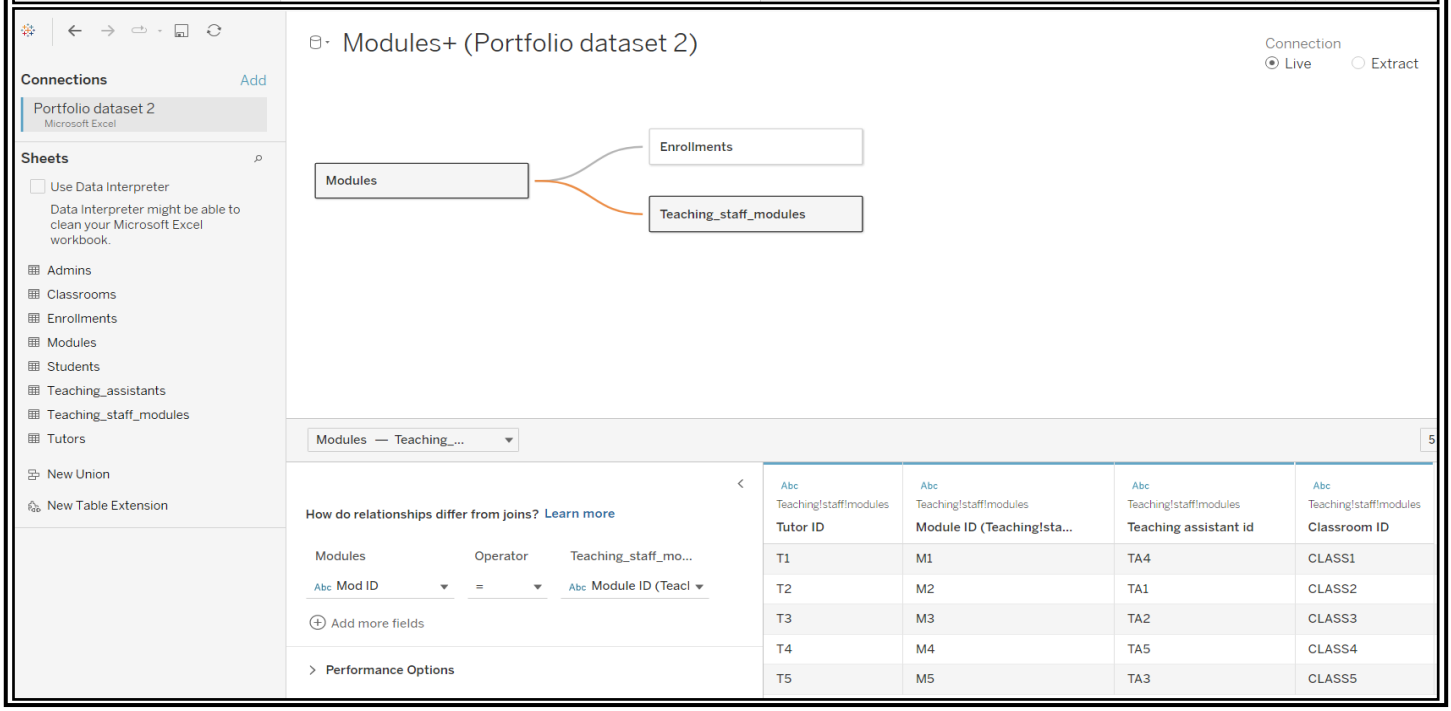
2nd Relational model

Relationship between (*Modules* & *Enrollment)* + (*Modules* & *Teaching staff modules)*

The module table specifies details about the module (i.e., module title, module id, module description). Whereas, in the Enrollment table, we can find all other details for which students can enrol (i.e., enrolment year and semester, enrolment ID, student ID, and module ID). Thus, we can link both tables using a common attribute called "**Module_id**" to create a single common database table for visualization. Now we can implement a relationship between two tables in a relational database using the primary key as the common attribute specified above as "**Module_id**."

The module table specifies the details about the module (i.e., module title, module id, module description) and the teaching staff module table (i.e., tutor id, module id, teaching assistance id, classroom id). By linking both tables using a common parameter, one can see all the details of the module and teaching assistance. By attributing them together into one common parameter called "**Module_id**" one can use both tables together in visualization.



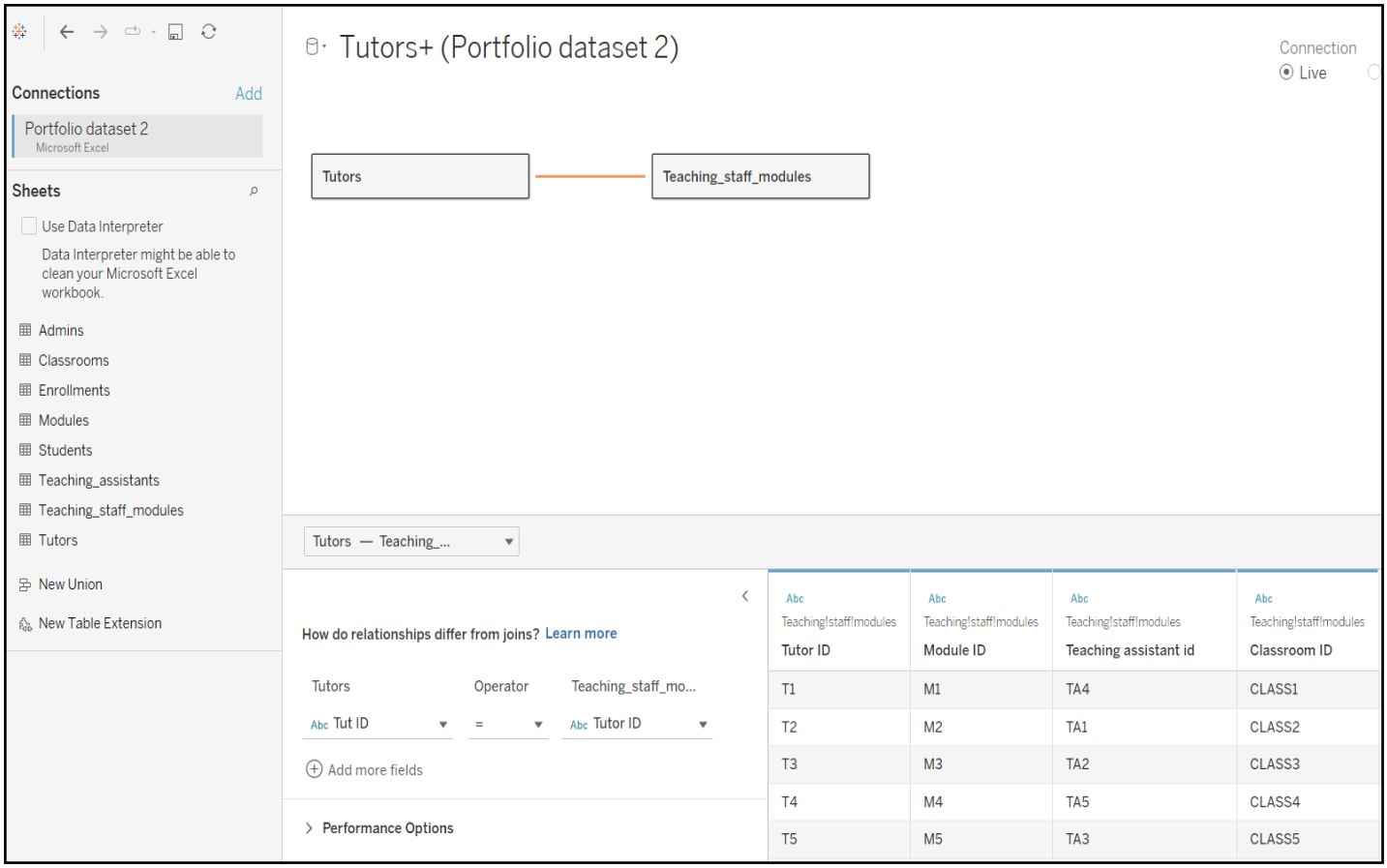


3rd Relational model

Relationship between ***Tutors*** and ***Teaching staff modules*** is defined using "**Tutor_Id**"

Tutor contains all the details about the tutor (i.e., tutor id, tutor name, tutor last name, tutor address, tutor post code), whereas teaching staff modules table have details of (i.e., tutor id, module id, teaching assistant id, classroom id). Both tables are unique, specifying details about the tutor and teaching staff modules taken by the tutor. But both tables act independently; therefore, in order to do visualization, one can use one common database or table by combining two tables using common attributes using primary key call "**Tutor_id**".

Now one common database exists where one sees all the combined information about tutor and teaching staff modules under one parametric condition used in visualization.

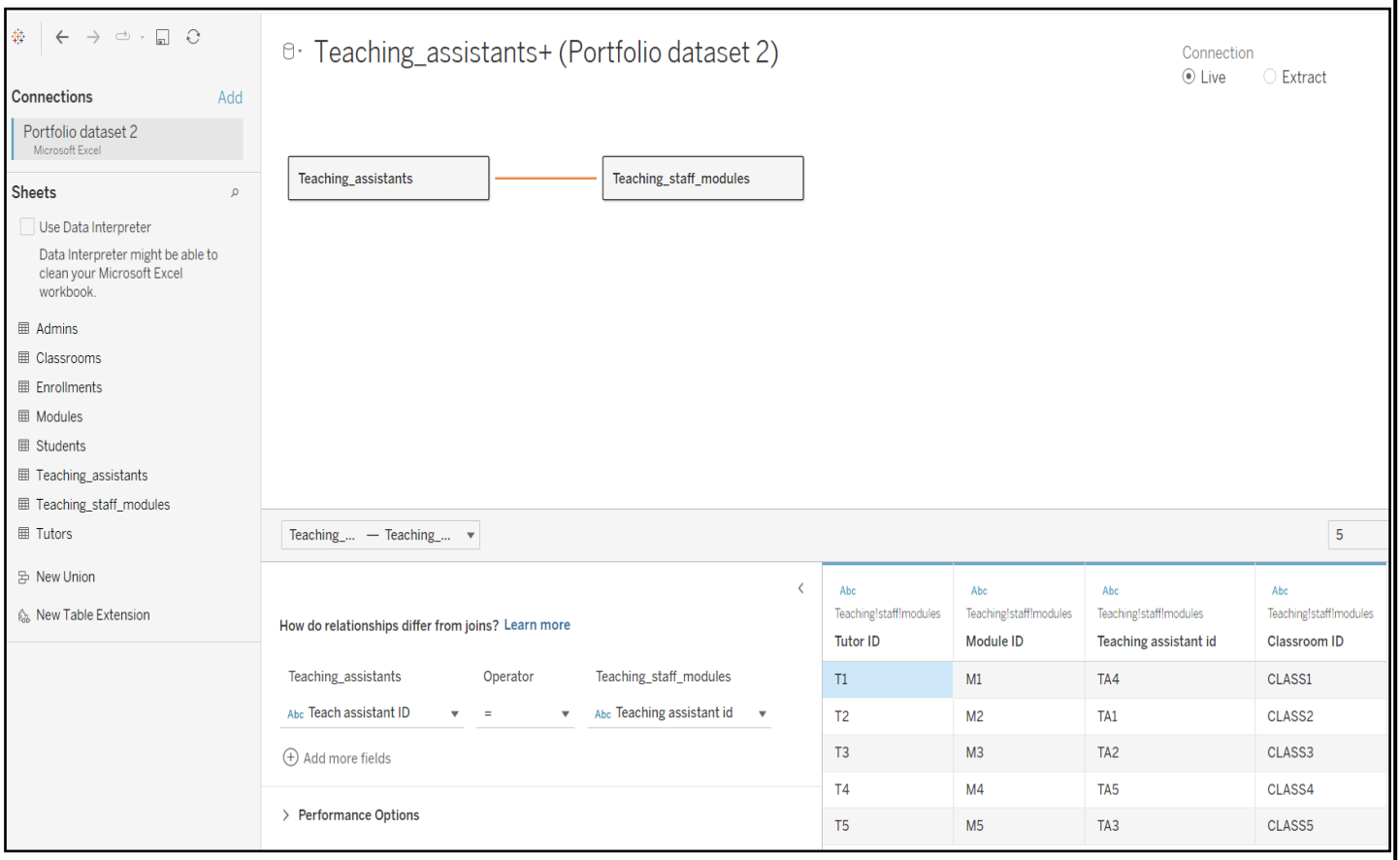


4th Relational model

Relationship between ***Teaching assistants*** and ***Teaching staff modules*** is defined using "**Teach_assistant_id**"

The teaching assistant table specifies detail about the teaching assistant (i.e., teaching assistant name, address, postal code, city, and teaching assistant ID). But teaching staff module tables specify details about (i.e., tutor id, module id, teaching assistant id, classroom id).

Both the tables are unique and have a similar resemblance to staff information, but both tables can't be used in visualization together until a synchronized database is prepared to combine two separate tables using one common parameter called the "Teaching Assistant ID." Now, using the primary key, we can establish a common synchronized database using one single table that contains the combined information of both teaching assistants and staff, which can be used for visualization.

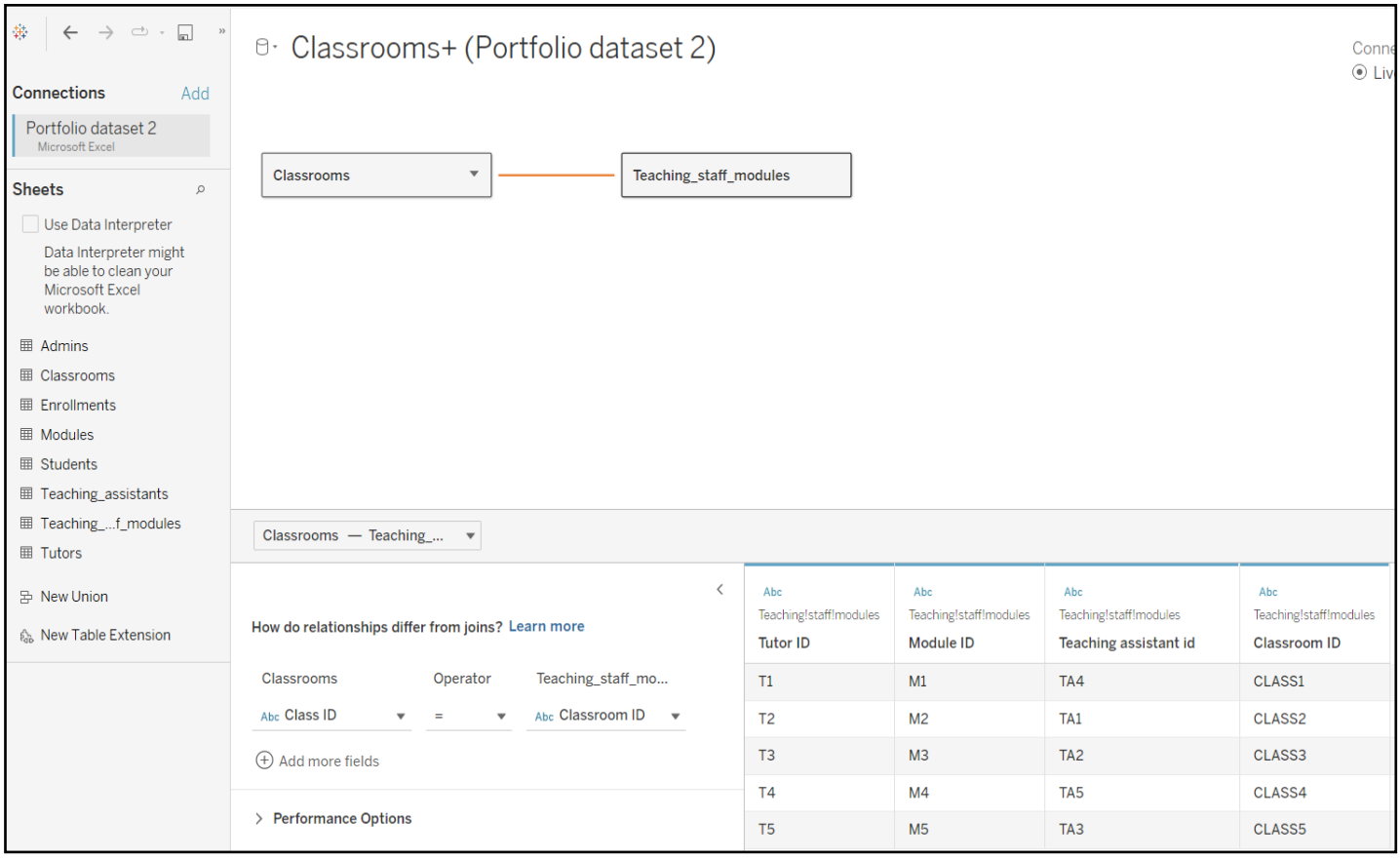


5th Relational model

Relationship between ***Classroom*** and ***Teaching staff modules*** is defined using "**Class_id**" or "**Classroom_id"**

Classrooms specifies detail about (class id, Class capacity, class floor, Lab room) and teaching staff module tables specify details about (i.e., tutor id, module id, teaching assistant id, classroom id). In the both the table we can see that it is connect in terms of subject in class. But both the table has to be used separately for visualization until unless a common parameter is established to create synchronised database.

We can use common parameter from both the table using primary key instead of using two separate table and thus we can create one common table by linking using "**Classroom_id**" or "**Class_id**". Thus, now common database is established to use one common table for visualization

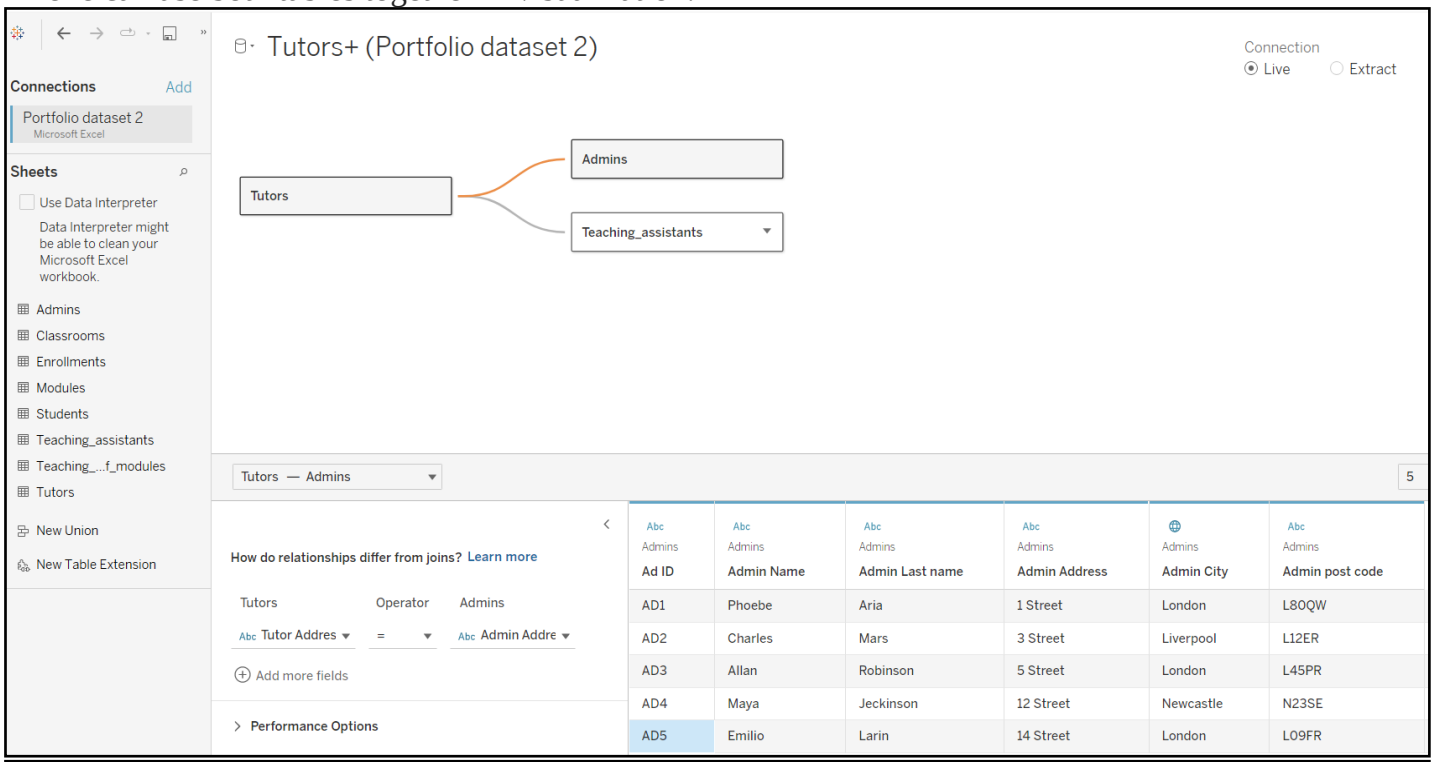


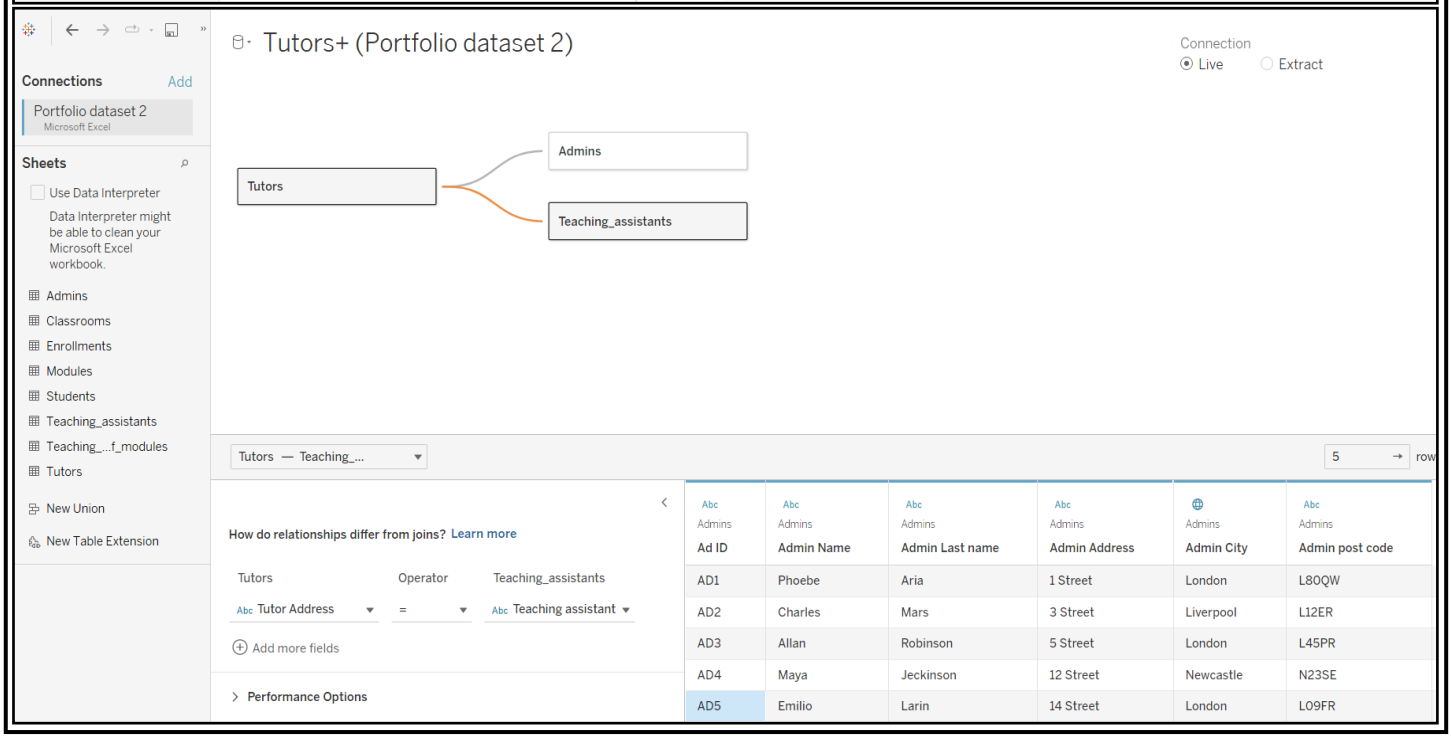
6th Relational model

Relationship between (*Tutors* & *Admins)* + (*Tutors* & *Teaching assistants)*

Tutor contains all the details about the tutor (i.e., tutor id, tutor name, last name, tutor address and post code), whereas admin table have details of (i.e., Admin id, Admin first name; last name; Admin address and Admin city and postal code). Thus, we can link both tables using a common attribute called "**Address**" to create a single common database table for visualization. Now we can implement a relationship between two tables in a relational database using the primary key as the common attribute as "**admin address**" & "**tutor address**"

Tutor contains all the details about the tutor (i.e., tutor id, tutor name, last name, **Tutor address** and post code), whereas the teaching assistant table specifies detail about the teaching assistant (i.e., teaching assistant name, **Teaching assistant address**, postal code, city, and teaching assistant id). By linking both tables using a common parameter, called "**Address**" one can see all the details of the both the table into one synchronised table. By attributing them together into one common parameter one can use both tables together in visualization.





Portfolio activity 3

Your task is to produce the following list of diagrams using the following two datasets:

- Sample superstore or superstore dataset
- Hospital Visits dataset

Part 1: For the following tasks you will need to use the Superstore data

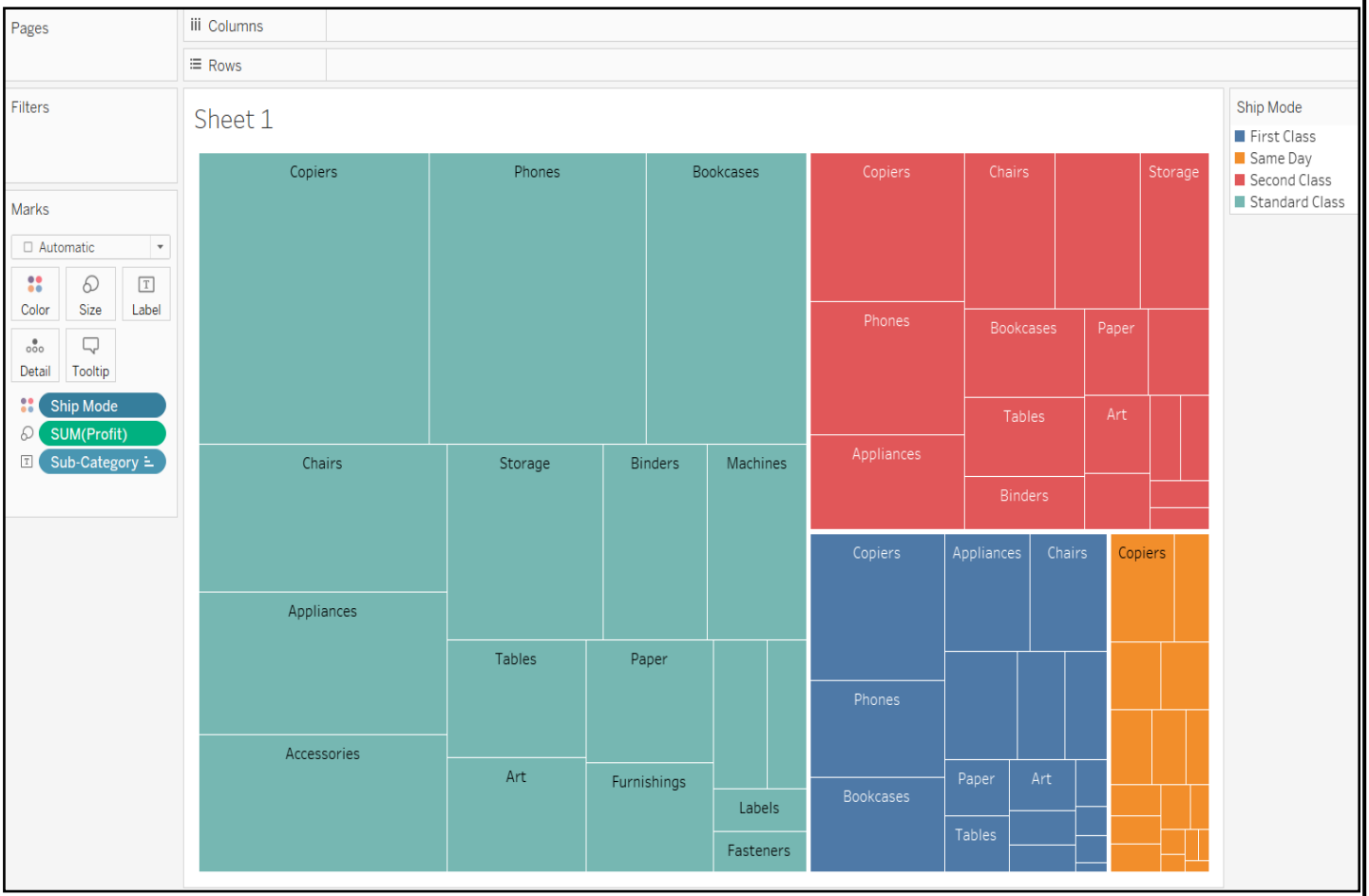
❖ Create a tree map diagram using the Sum of profit, the sub-category of products and the shipping mode. Discuss shortly your understanding on the diagram.

Explanation: A tree map is a sort of visualization of data used by Tableau that shows structured data as a collection of nested rectangles. Each rectangle's size and colour correspond to various data values or categories. The placement of the rectangles reveals the connections between them as well as the general organization of the data.

In the below diagram, we can see that each colour is classified for each class or category from the ship mode classification on the right side of the picture. The standard class data has a larger box size, which represents a higher amount or percentage of shipping done. For example: "Copiers" the copiers are shipped at a higher percentage as compared to chairs, bookcases, and other items in the standard class. This same rule applies to the other three classes as well.

This also means that there is a higher amount of shipping done for the various items in the standard class as compared to second class, first class, and same day shipping because the overall rectangle is larger in the standard class as compared to other rectangles in other classes.

Tree Map

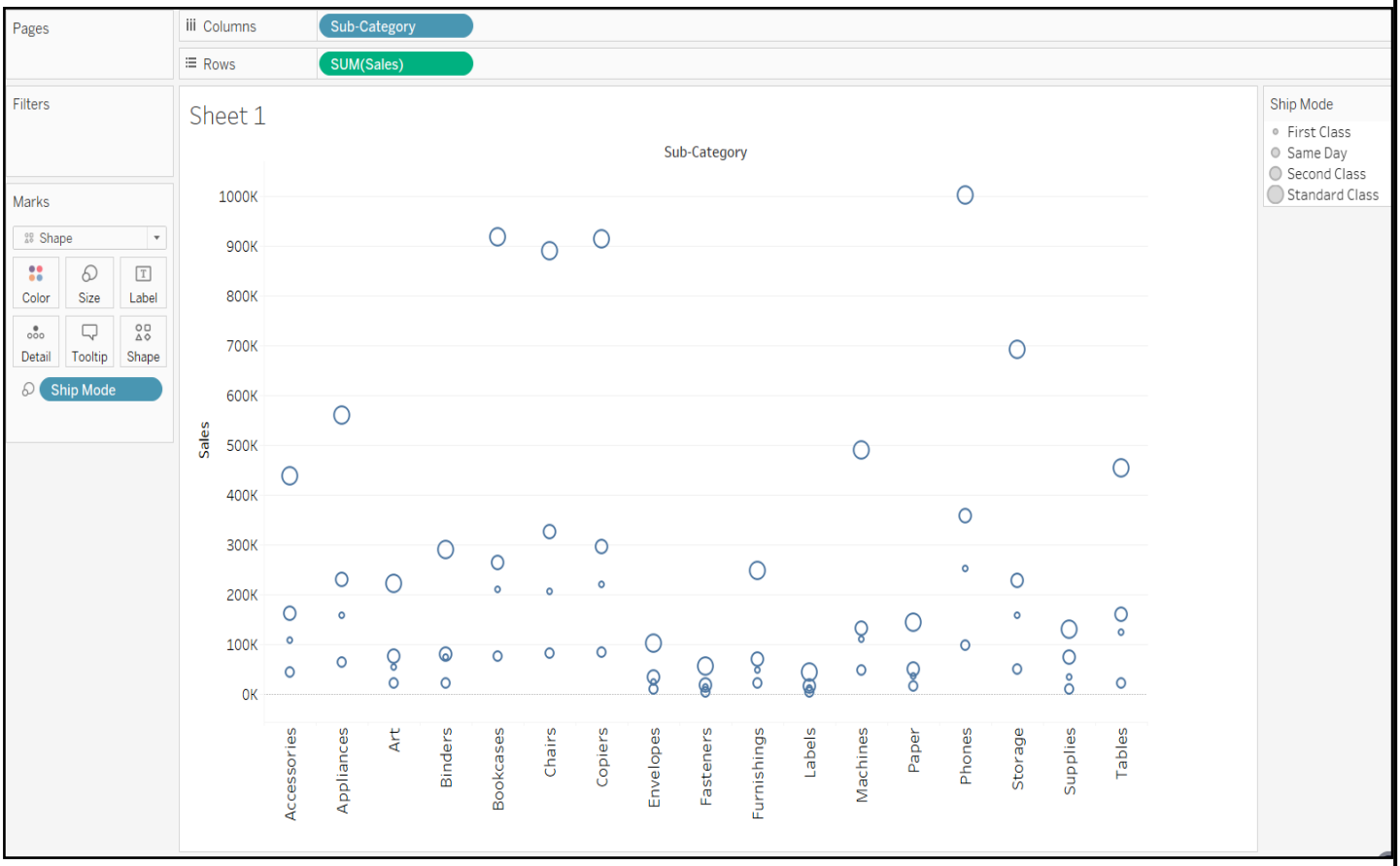


❖ Create a circle chart to demonstrate the shipping mode in terms of sum of the sales and sub-categories. Discuss shortly your understanding on the diagram.

Explanation: The "Circle Chart" allows users to create interactive dashboards and perform visualizations in which viewers can select a data point or category and specify the related information. Specifically, the Circle Chart feature enables users to select a mark on a visualization and help them to specify the specific information on each circle. The selected mark is highlighted marked with certain information.

In the below case, we can see Standard Class is making more sales as compared to Same Day and First Class (example: bookcases, chairs, copiers, etc.). Just after standard class, we can see that second class with a smaller circle size is making less sales after standard class but more sales as compared to first class. Whereas first class is making the least amount of sales compared to all other classes, it is marked at the bottom of the x-axis indicating smaller circle size as compared to all other.

Circle Chart



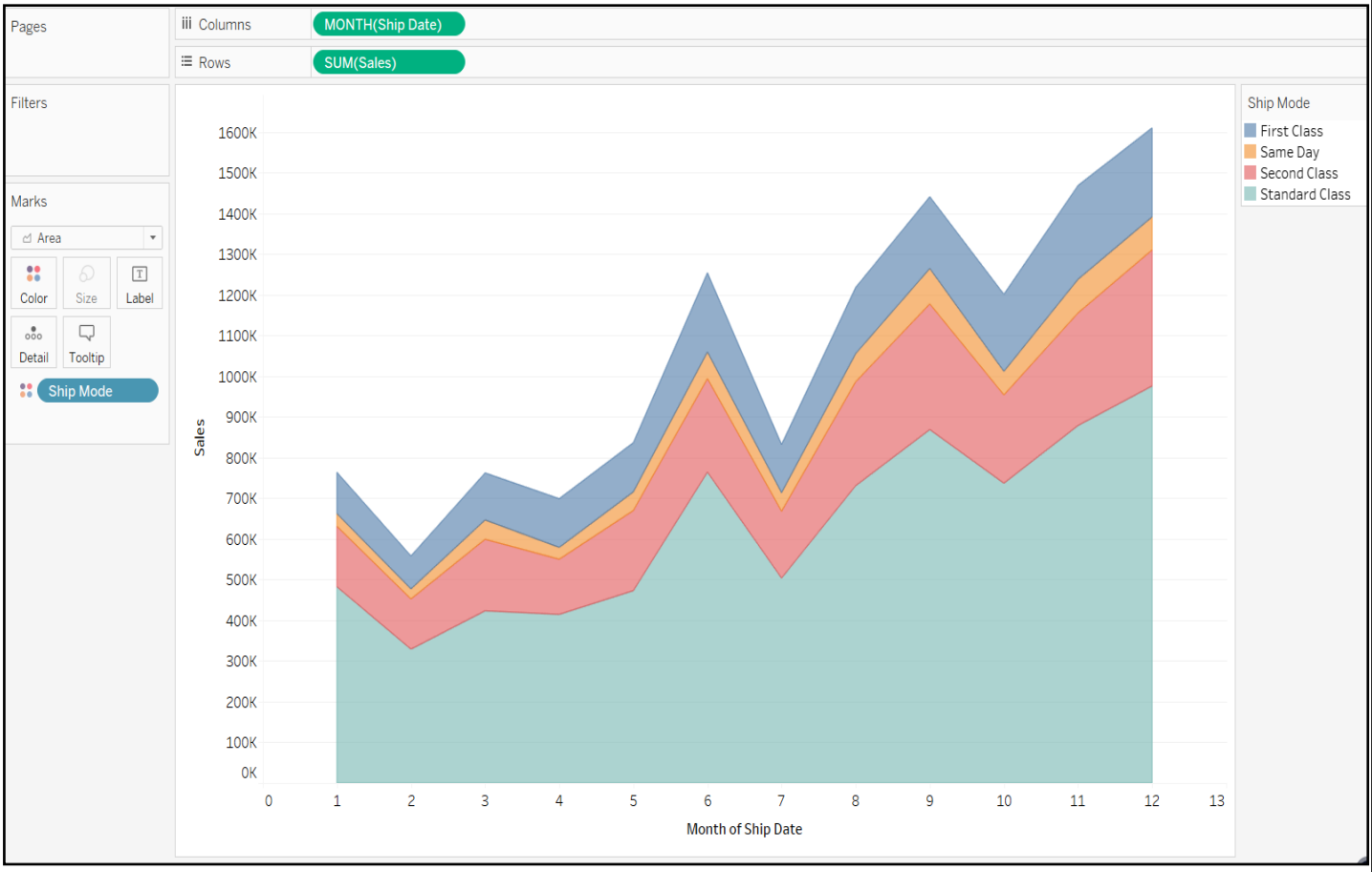
❖ Create an area chart using the monthly shipping date and the sum of sales. Demonstrate the data based on the shipping mode. Discuss shortly your understanding on the diagram.

Explanation: The "Area Chart" is a style of data visualization in which quantitative data is displayed as a series of data points linked by line segments, with the area below the line coloured in with different colours or shading. The x-axis of an area chart typically indicates time or a category variable, whereas the y-axis reflects a quantitative variable. Area charts are excellent for analysing the relative sizes of different categories and for displaying how data changes over time or between categories.

In the below Area chat, we can see that there is an uneven trend formed via the category of ship mode on various ship dates. In the month of February, the lowest quantity of sales was done for same-day shipping mode as compared to other categories of shipping mode.

The larger size area represents a higher amount of sales that we can see in the standard class of shipping mode in the month of December as compared to other shipping modes. However, in the month of December, there were higher sales in all categories as compared to other months. The sale figure in various shipping modes was never increasing or decreasing. It was quite uneven from month to month. Also, we can notice that in June and September, a quite good amount of sales happened via all four categories of ship modes as compared to other months exceptionally strong performance in the last month of the year.

Area Chart

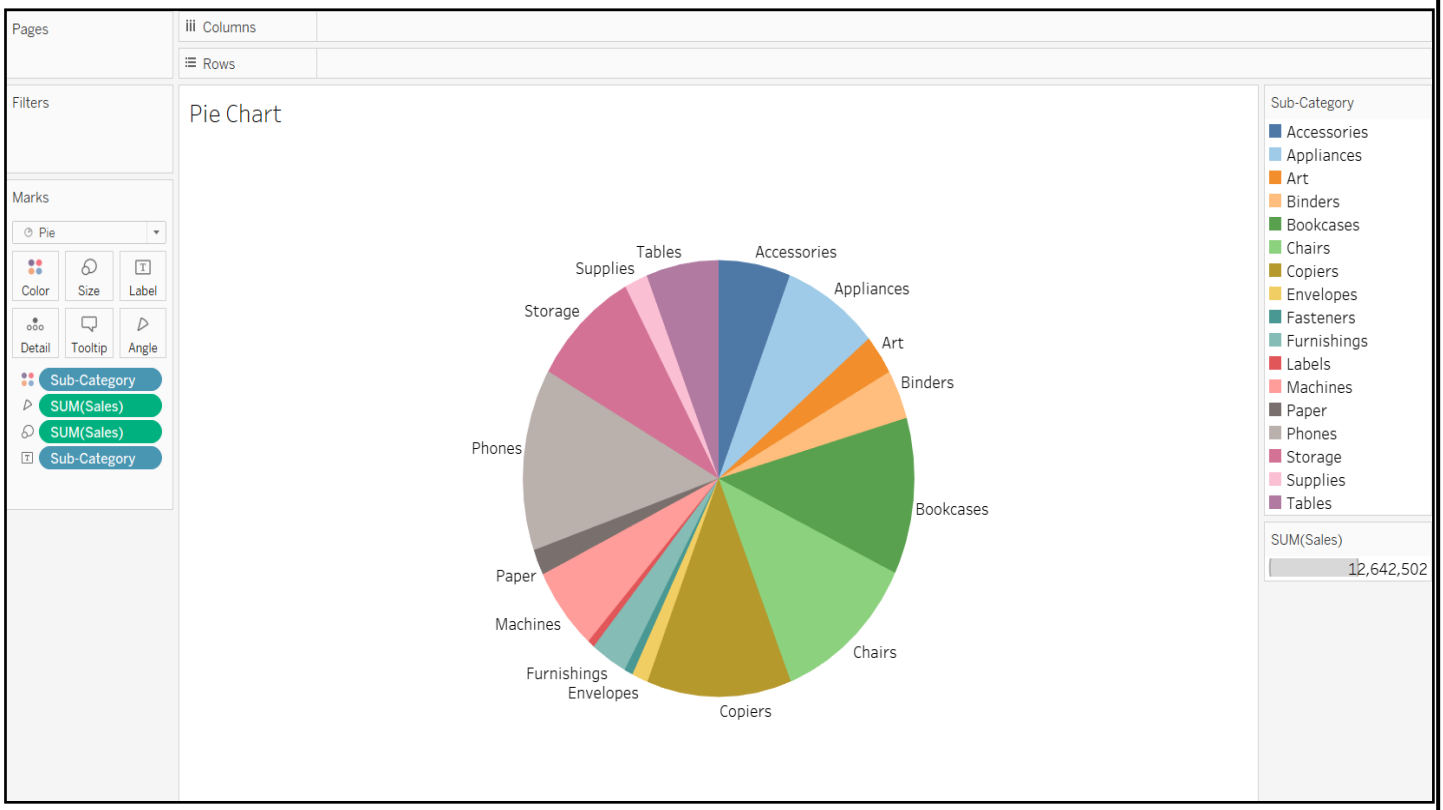


❖ Create a pie to demonstrate the sum of the sales per sub-category. Discuss shortly your understanding on the diagram.

Explanation: A Pie Chart is a circular chart divided into slices to illustrate data proportions. A pie chart's slices reflect a category or fraction of the data being evaluated. Each slice's size is defined by the value of the data it represents, with bigger slices representing a greater proportion of the data. Tableau frequently employs pie charts to depict the distribution of data that is categorical. They are especially effective for displaying data with a limited number of categories or for emphasizing one or a few categories that comprise the majority of the data.

In the below case, we can see that phones represent a higher percentage or amount of sales, followed by bookcases, chairs, copiers, storage, appliances, and machines in decreasing order. Whereas labels and fasteners have the lowest amount of sales in the complete subcategory, which is even below 10%. Tables, accessories, binders, and furnishings have a mid-level of sales between 20% and 40%. By the size of the slice in the pie chart for various categories, we can analyse how much amount or percentage of sales have been generated in various sub-categories in the table on the right side of the chart.

Pie Chart



❖ Create a box-and-whisker diagram to demonstrate the shipping mode, profit per sub-category. Discuss shortly your understanding on the diagram.

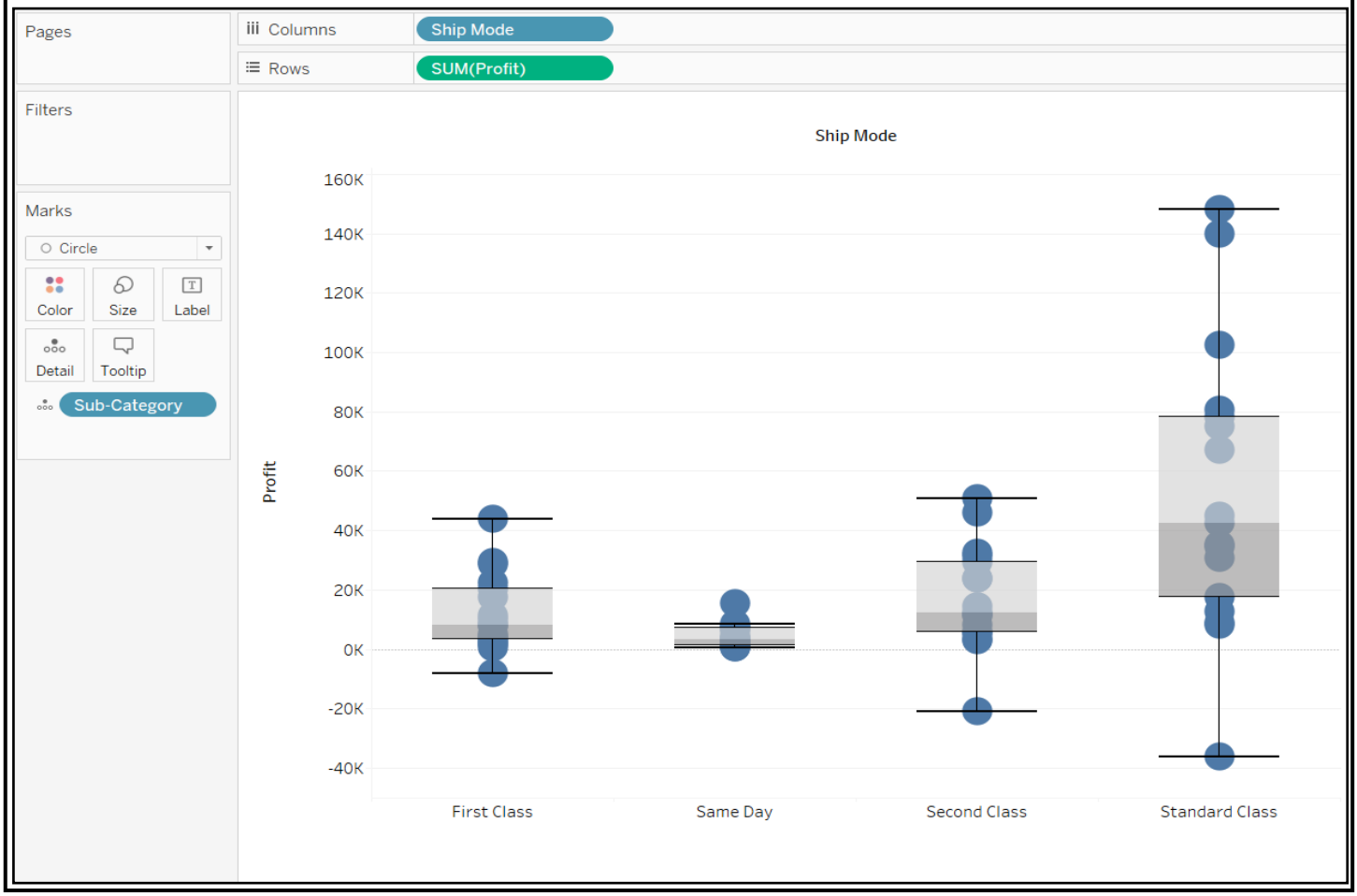
Explanation: Box-and-whisker diagram shows the distribution of a continuous variable across its quartiles, commonly referred to as a box plot. The plot's box displays the data's interquartile range (IQR), which is the range among the first and third quartiles. The line that runs vertically inside the box represents the median. The whiskers extend from the box's edges to represent the data range, excluding any outliers. Individual points beyond of the whiskers are used to illustrate outliers.

In the below case we can see that….

- ✓ ***First class:*** The copier made the highest profit, and tables made a loss. whereas accessories, binders, and arts made an average level of profit at mid-level in the 2nd quartile. The phone and appliances are above the 3rd quartile, and supplies, tables, & machines are all in the 1st quartile.
- ✓ ***Same day:*** we can see that copiers had a higher profit and envelopes made the least profit. Whereas most of the data lies above 50% in the box at the 2nd quartile.
- ✓ ***Second class***: The copier made a higher profit, and tables made a loss under this category. 50% of the data lies in the 2nd quartile within the box; those include envelops, art, binders, bookcases, and accessories.
- ✓ ***Standard class*:** In this class, we can see that a higher amount of profit has been generated as compared to all other classes, but not more than 40% of the data lies in the box at the 2nd quartile, and more than 50% of the data lies outside the box (in the 1st quartile and below the 3rd quartile). Like usual, the copier made a higher profit above the 1st quartile, and tables made a loss below the 3rd quartile.

Note: *Here, data means various items shipped under a sub-category via ship mode*.

Box-and-whisker diagram



Part 2: *For the following task you will need to use the Hospital Visits dataset*

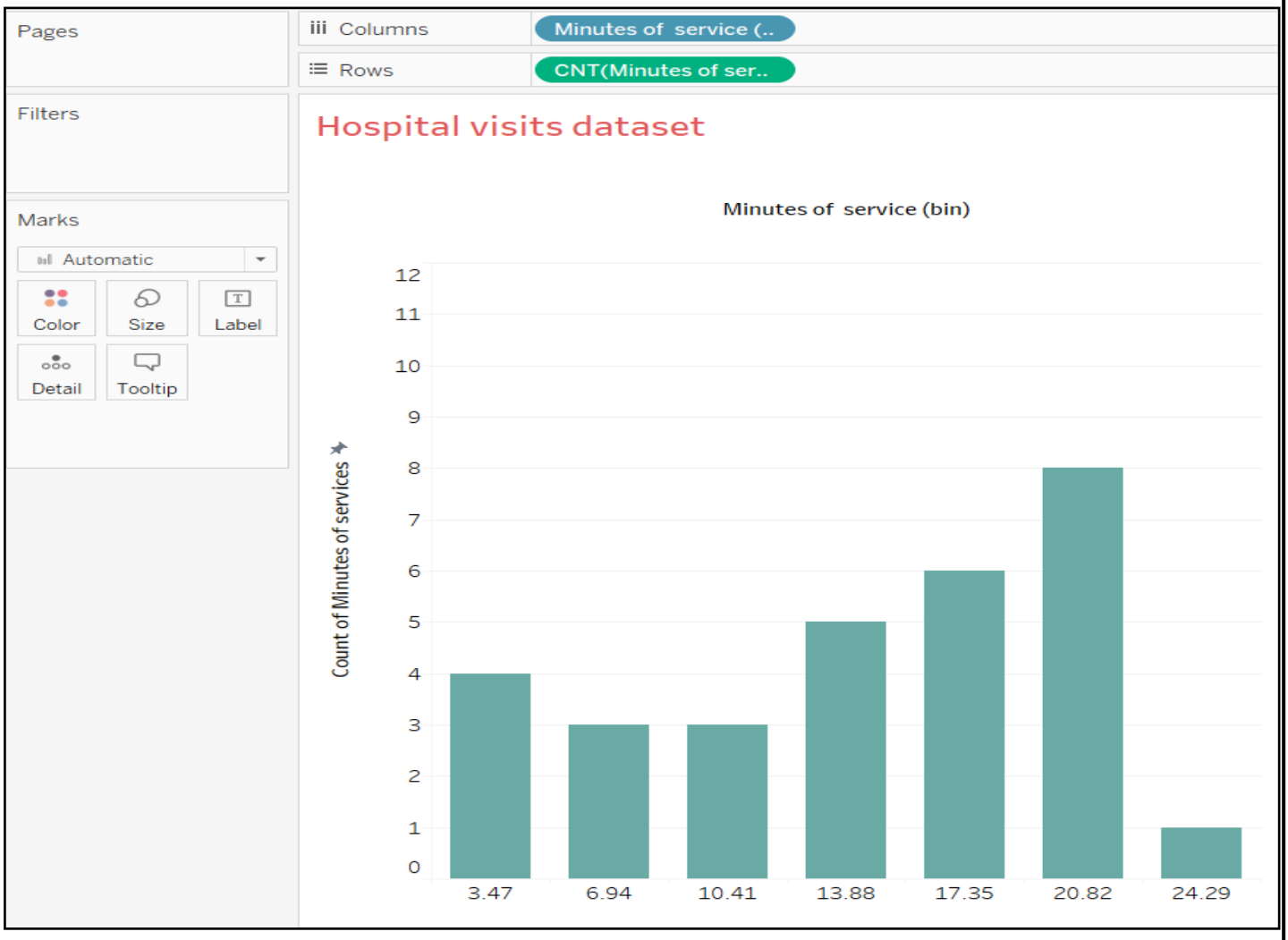
Create a histogram to demonstrate the minutes of service (bins of 3.47) in relation to the count of the minutes of service. Discuss shortly your understanding on the diagram.

Explanation: In the below Histogram graph we can see that the hospital visit data is not normally distributed, it is uneven dataset having positive skewness having one outlier at right side.

Each bin size is been distanced at 3.47 and height of the bar represent the number of counts. In the below case we can see that there is higher minute of service (8) has been provided between (17.35 – 20.82) and lowest minute of service (1) has been provided between (20.82 – 24.29). However, there is stable services offered from (6.94 to 10.41). The minutes of service has been increasing level from (13.88 to 20.82).

This resembles that there are better minutes of service provided from (3.47 to 20.82) exceptionally worst service offered at (20.82 – 24.29) for all the patients who visited hospital.

Histogram Diagram



Portfolio activity 4

- What is digital transformation and how can digital transformation support an IT company to face modern challenges?

The process of leveraging digital technology to radically alter how businesses function, provide value to consumers, and engage with stakeholders is known as "digital transformation." Utilizing technologies like cloud computing, data analytics, AI, and the Internet of Things (also known as the IoT) to boost operational effectiveness, improve the client experience, and spur innovation

The implementation of digital transformation can have a substantial positive impact on an IT company's productivity, agility, and resilience in the face of contemporary difficulties. An IT company can benefit from digital transformation in a number of ways, such as:

- ✓ *Increased agility and flexibility*: Thanks to digital transformation, IT organizations are now better equipped to adapt swiftly to shifting market conditions, consumer needs, and newly developed technologies. IT organizations can create and roll out novel goods and services more quickly, respond to changing circumstances in the market, and scale activities more effectively by utilizing technological advances.
- ✓ *Enhanced operating effectiveness*: Digital change can help IT organizations simplify operations, automate manual procedures, and save money. IT firms can streamline their processes, make the most of their resources, and increase their general effectiveness by utilizing cloud computing, information analytics, and other technological advances.
- ✓ *Improved client experience*: digitization can assist IT organizations in providing their clients with more individualized, captivating, and frictionless experiences. IT organizations may learn about the behaviour, tastes, and wants of their customers and create solutions that are specifically suited to their needs by utilizing information about their clients, data analysis, and AI technology.
- ✓ *Innovation growth*: Digital evolution can assist IT organizations in driving innovation and investigating new business structures, goods, and services. IT firms can find new opportunities, try out new concepts, and create fresh sources of income by utilizing evolving technology and data-driven knowledge.

In conclusion, IT organizations may gain a lot from digital transformation, allowing them to increase productivity, improve consumer service, and spur innovation. IT organizations may thrive in today's quickly changing business environment by integrating digital technology and utilizing information-driven conclusions.

- Discuss the advantages of visualisations and the use of visualisation software to the digital transformation era.

In the age of digitization, visualization and visualization applications are essential because they offer a number of benefits that help businesses use information and technological resources to stimulate innovation, boost productivity, and improve the client experience. In the age of digital change, some of the main benefits of visualization and tools for visualization are:

- ✓ *Making better choices*: People who make decisions may rapidly and easily gain understanding from complicated sets of information with the aid of visualizations and visualization software, enabling them to make well-informed decisions in the present. Decision-makers can examine data from many viewpoints and points of view using interactive visualizations, picking up patterns, developments, and anomalies that might otherwise go unreported.
- ✓ *Greater interaction and cooperation*: By offering a single platform for data exchange and analysis, visualizations and visualization tools can assist teams in collaborating more successfully. Members of the team may collaborate to analyse data, exchange insights, and create ideas using dynamic visuals, which improves productivity and yields better results.
- ✓ *Effectiveness and efficiency are acquired*: By automating repetitive processes like data cleansing and presentation, visualization programs can allow stakeholders like data scientists to devote themselves to more worthwhile endeavours. Reduced time and effort are needed to evaluate data and decide since users can simply and rapidly obtain the information they seek via dynamic representations.
- ✓ *Better consumer experience*: Companies can better understand their customers' needs by using visualization technology. Organizations may determine the wants and preferences of their customers and provide services that are tailored to their individual needs by evaluating their information and generating dynamic representations.
- ✓ *Improved innovation*: Businesses can explore fresh concepts and prospects as well as create cutting-edge goods and services with the aid of visualization tools. Corporations may see developing trends, forecast demand in the future, and create novel company structures that promote growth and a competitive edge by utilizing visuals that are interactive and using sophisticated analytics.

In conclusion, representations and visualization applications benefit organizations in the digital transformation era in a number of ways, allowing companies to make better choices, cooperate more successfully, boost efficiency and effectiveness, enhance the user experience, and spur innovation. Companies may survive and grow in today's rapidly changing marketplace by utilizing these advances in technology.

References

- Tableau: - https://sso.online.tableau.com/public/idp/SSO
- Hospital visit data: - https://github.com/topics/hospital-data
- Marc Benioff, Chairman and CO-CEO, Salesforce (Year 28-04-2.23) (url: https://www.salesforce.com/eu/products/platform/what-is-digital-transformation/)
- Jay Ferro, chief information and technology officer of Clario (2016) (url: https://enterprisersproject.com/what-is-digital-transformation)