

Ex. No:	6
Date:	31/7/24

Data aggregation and statistical functions

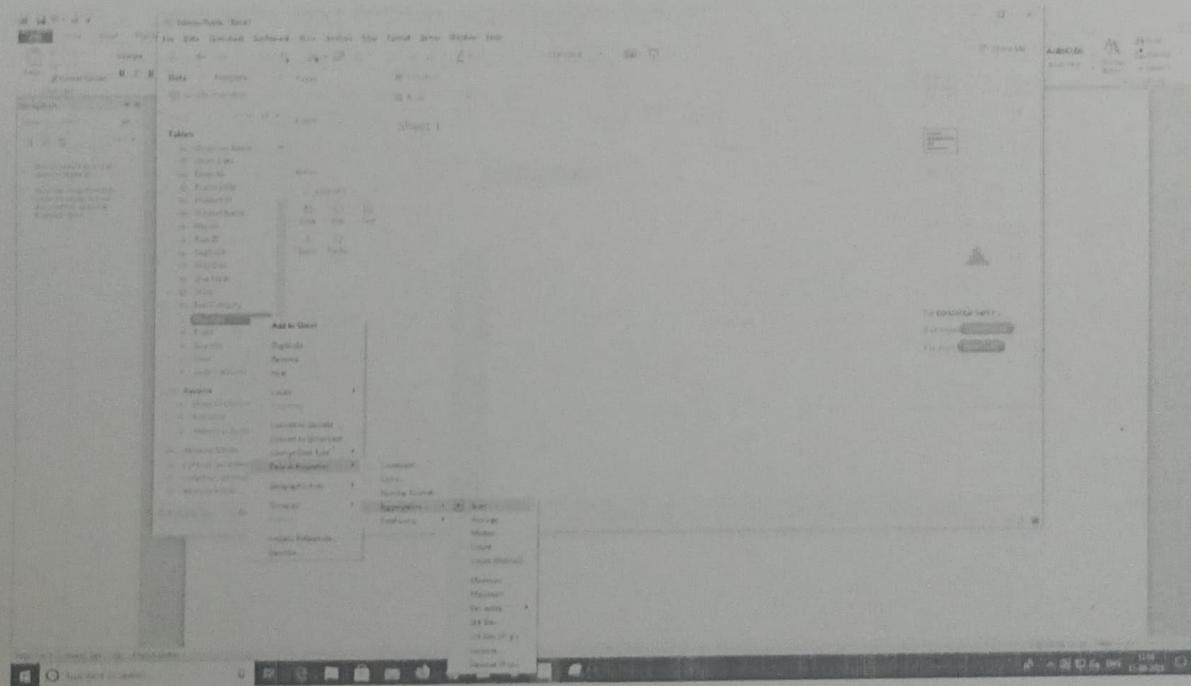
AIM

To implement the data aggregation and statistical functions

DESCRIPTION:

We can apply various aggregation and statistical functions on data such as count, minimum, maximum, standard deviation, variance etc. This is shown below.

This can be done by right clicking on the required field of dataset, click on Default properties and click on aggregation.



Or the above operation can be done by creating a calculated field as shown below. To create a calculated field, click on the down arrow button beside search tab above Tables panel, drag a field to that calculated field window.

The screenshot shows the Tableau Public interface with the 'Customer' data source loaded. A calculated field named 'Total Sales' has been created and is being edited. The formula is set to `Sum([Sales])`. The description is "Return the sum of all the values in the expression, and the result will always be limited to one decimal place". The 'Data or flag data' checkbox is checked.

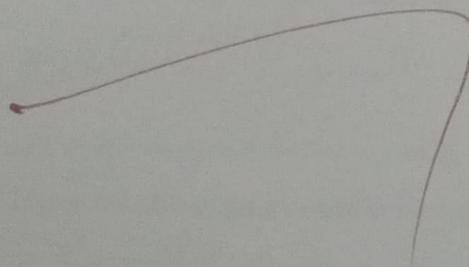
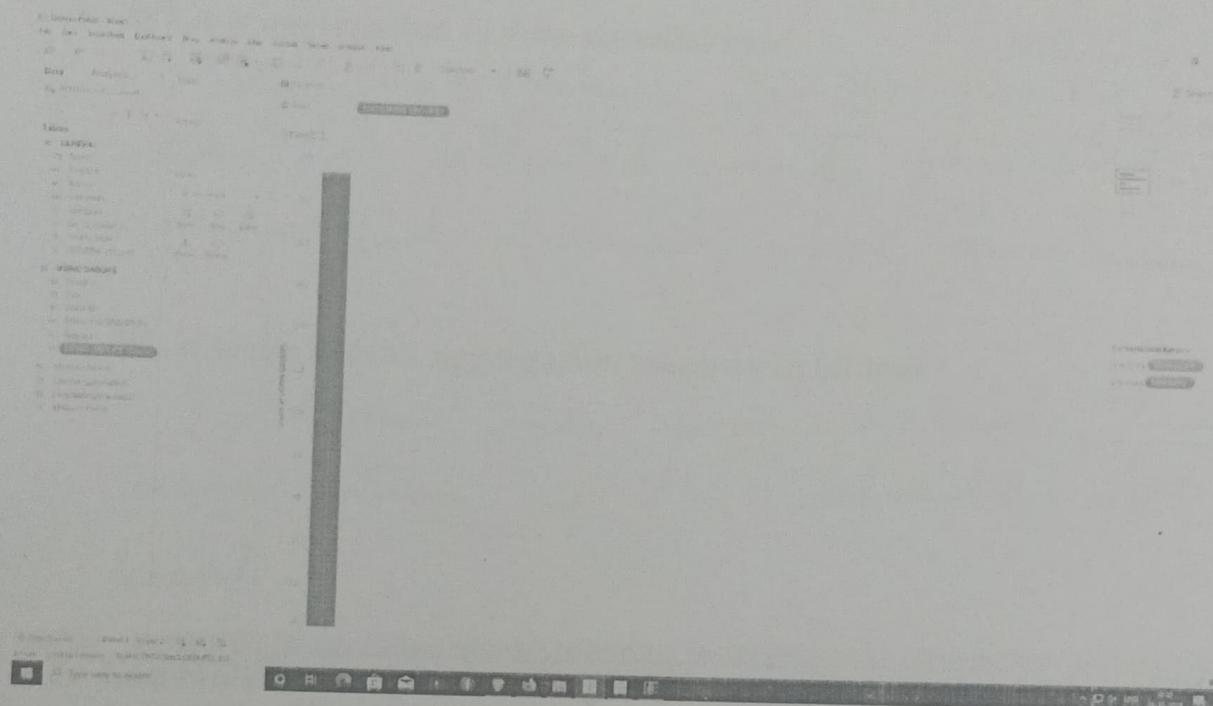
Then click on apply and results are shown below:

The screenshot shows the results of the calculated field 'Total Sales'. The data table now includes a new column 'Total Sales' which contains the sum of the 'Sales' column for each row. The total value is displayed as 2200.

Customer Name	Product ID	Order ID	Postal Code	Product ID	Product Name	Region	Row ID	Region	Sales	Total Sales
Customer 1	Prod 1	Ord 1	10001	Prod 1	Prod 1	Region 1	Row 1	Region 1	100	100
Customer 2	Prod 2	Ord 2	10002	Prod 2	Prod 2	Region 2	Row 2	Region 2	200	200
Customer 3	Prod 3	Ord 3	10003	Prod 3	Prod 3	Region 3	Row 3	Region 3	300	300
Customer 4	Prod 4	Ord 4	10004	Prod 4	Prod 4	Region 4	Row 4	Region 4	400	400
Customer 5	Prod 5	Ord 5	10005	Prod 5	Prod 5	Region 5	Row 5	Region 5	500	500
Customer 6	Prod 6	Ord 6	10006	Prod 6	Prod 6	Region 6	Row 6	Region 6	600	600
Customer 7	Prod 7	Ord 7	10007	Prod 7	Prod 7	Region 7	Row 7	Region 7	700	700
Customer 8	Prod 8	Ord 8	10008	Prod 8	Prod 8	Region 8	Row 8	Region 8	800	800
Customer 9	Prod 9	Ord 9	10009	Prod 9	Prod 9	Region 9	Row 9	Region 9	900	900
Customer 10	Prod 10	Ord 10	10010	Prod 10	Prod 10	Region 10	Row 10	Region 10	1000	1000
Customer 11	Prod 11	Ord 11	10011	Prod 11	Prod 11	Region 11	Row 11	Region 11	1100	1100
Customer 12	Prod 12	Ord 12	10012	Prod 12	Prod 12	Region 12	Row 12	Region 12	1200	1200
Customer 13	Prod 13	Ord 13	10013	Prod 13	Prod 13	Region 13	Row 13	Region 13	1300	1300
Customer 14	Prod 14	Ord 14	10014	Prod 14	Prod 14	Region 14	Row 14	Region 14	1400	1400
Customer 15	Prod 15	Ord 15	10015	Prod 15	Prod 15	Region 15	Row 15	Region 15	1500	1500
Customer 16	Prod 16	Ord 16	10016	Prod 16	Prod 16	Region 16	Row 16	Region 16	1600	1600
Customer 17	Prod 17	Ord 17	10017	Prod 17	Prod 17	Region 17	Row 17	Region 17	1700	1700
Customer 18	Prod 18	Ord 18	10018	Prod 18	Prod 18	Region 18	Row 18	Region 18	1800	1800
Customer 19	Prod 19	Ord 19	10019	Prod 19	Prod 19	Region 19	Row 19	Region 19	1900	1900
Customer 20	Prod 20	Ord 20	10020	Prod 20	Prod 20	Region 20	Row 20	Region 20	2000	2000

In the same way we can apply any aggregate or statistical function on data with the help of calculated fields.

Output:



VIVA VOICE

1. What is aggregation in data visualization?

Data aggregation is the process where data is collected and presented in a summarized format for analysis.

2. What is the default aggregation measure in tableau?

Tableau uses sum because that aggregation is the default for the field.

3. Can tableau do standard deviation?

Finding the standard deviation in tableau just involves changing the aggregation of a measure.

4. How to show percentile in tableau?

Create a new worksheet.

Drag and drop fields.

Add a measure to the view.

Configure percentile calculation.

5. How to calculate formula in tableau?

Open your tableau workbook.

Create a calculated field

Define the calculation.

Use the calculated field.

Ex. No:	7	Data Visualization
Date:	7/2/24	

AIM

To implement various types of Visualization using tableau

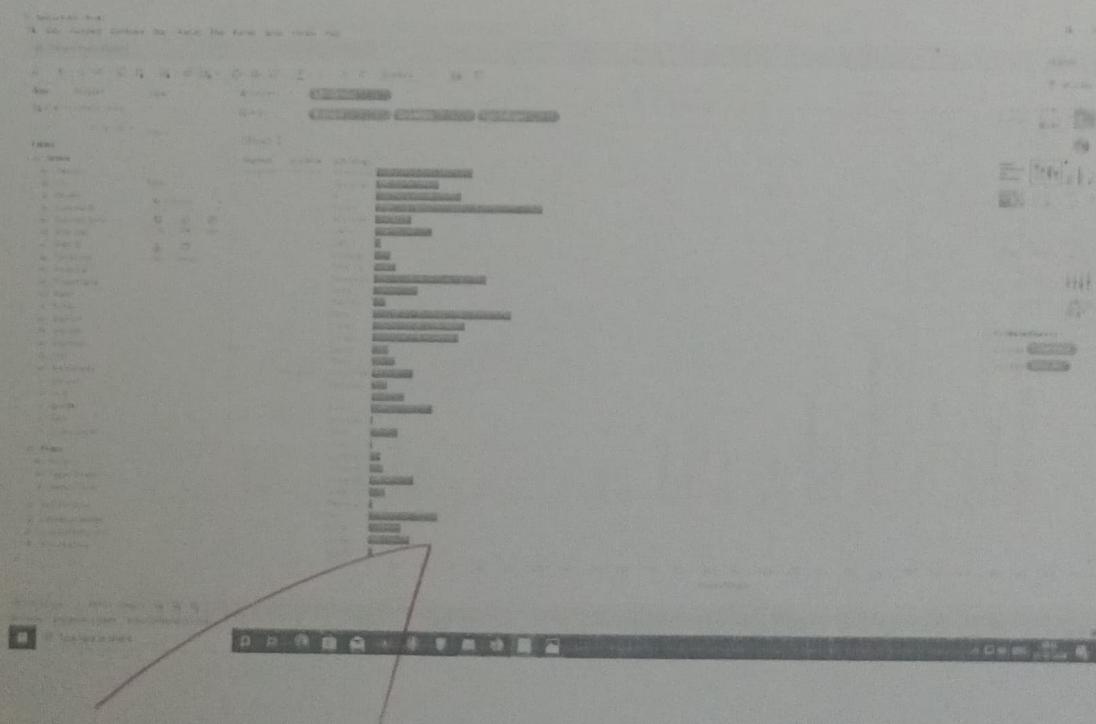
Description:

We can perform various visualization operations on data in Tableau. Some of them are bar chart, histogram, bubble chart, gantt chart, scatter plot, heat map etc.

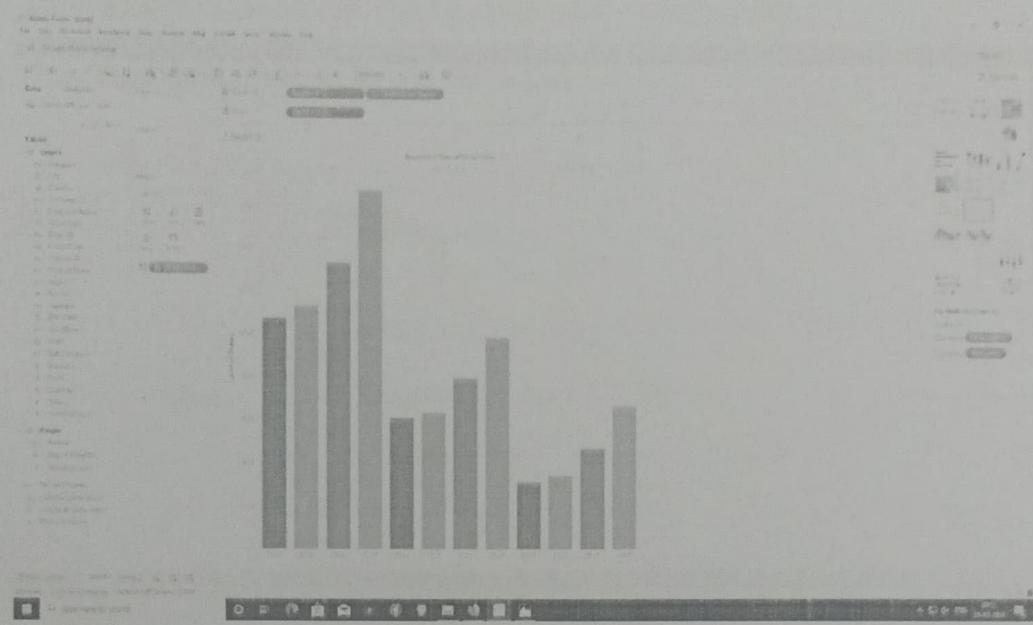
Bar chart:

Bar charts can be created in 3 variations in Tableau: Horizontal bars, stacked bars, side-by-side bars.

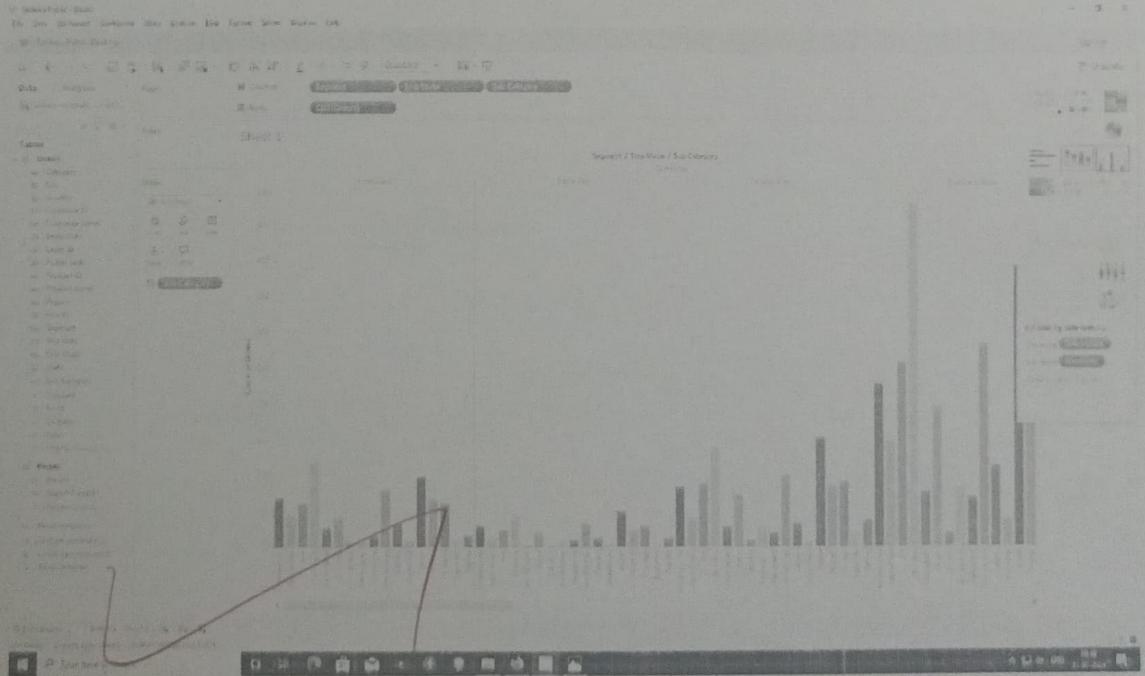
Horizontal bars can be created by selecting that type of chart from Show Me menu on right hand side of Canvas. The type of chart in box on right hand side represents horizontal bar graph.



In similar to above, stacked bar graph can be created and the result is shown below

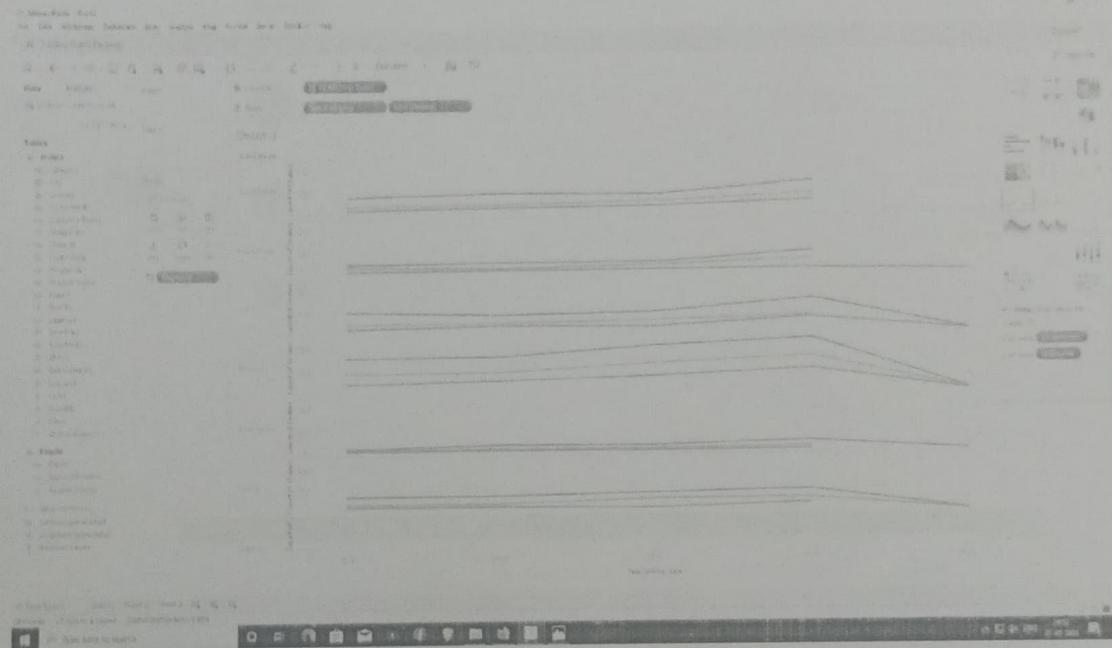


Side-by-side bar chart can be created in following way.

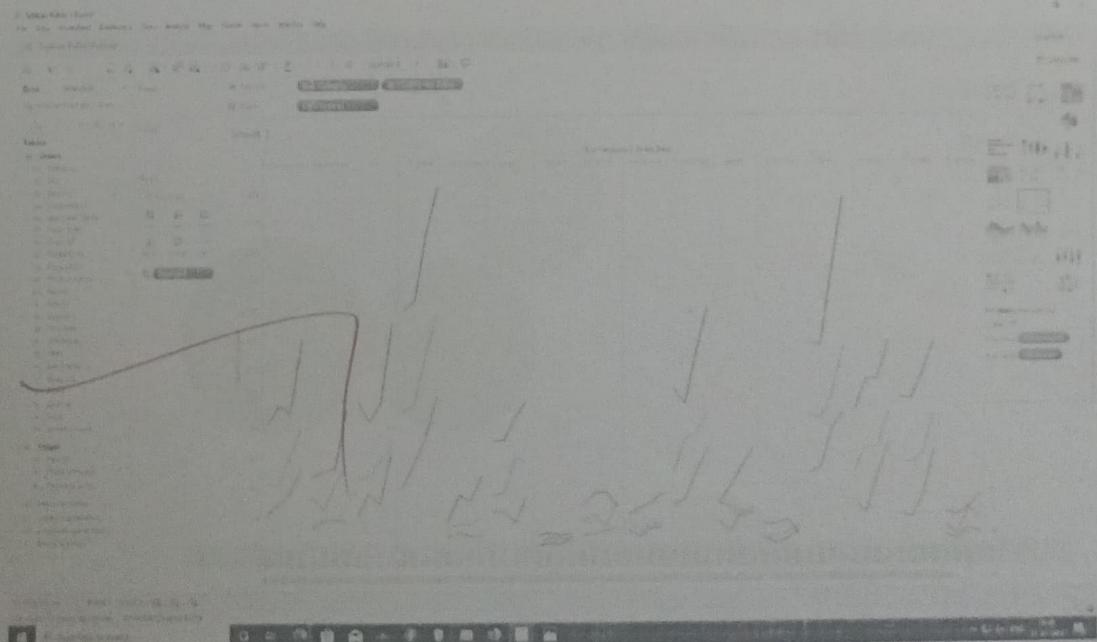


Line graph: Line graph can be continuous or discrete.

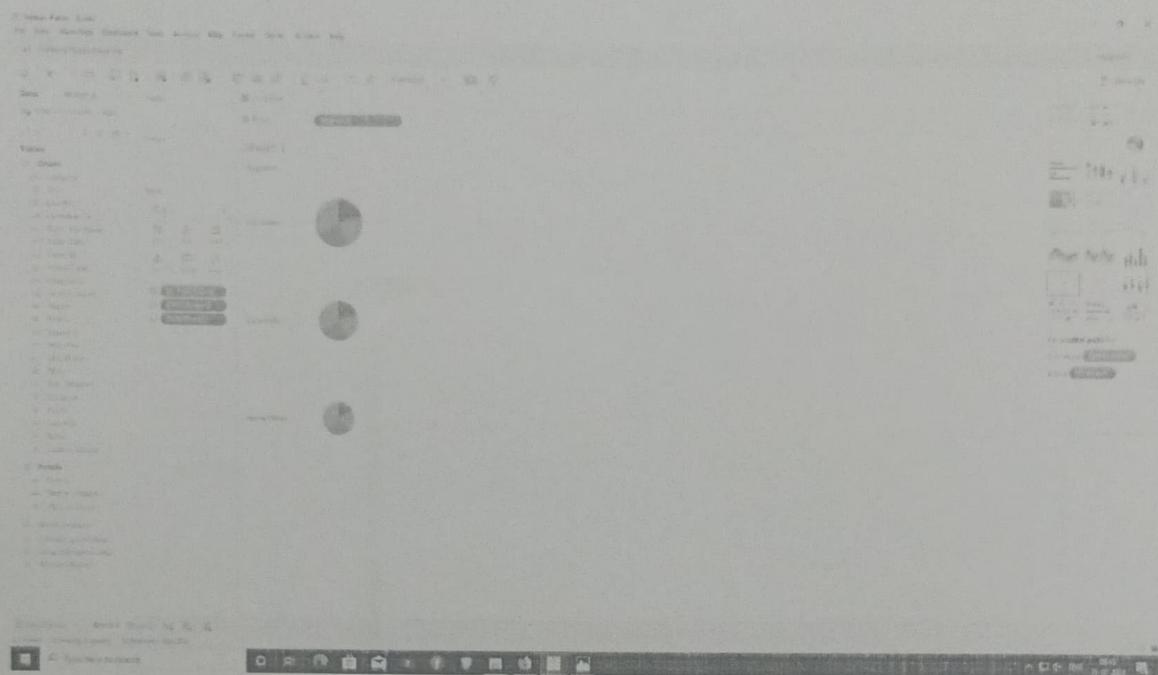
Continuous line graph is shown below:



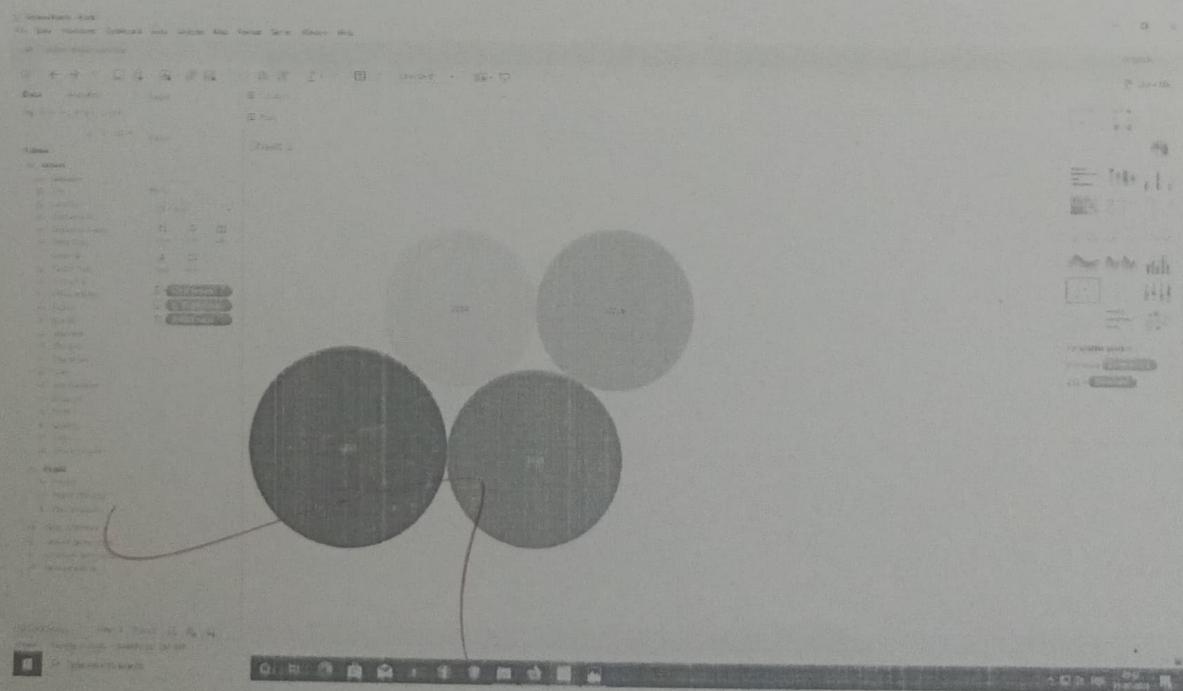
Discrete line graph is shown below:



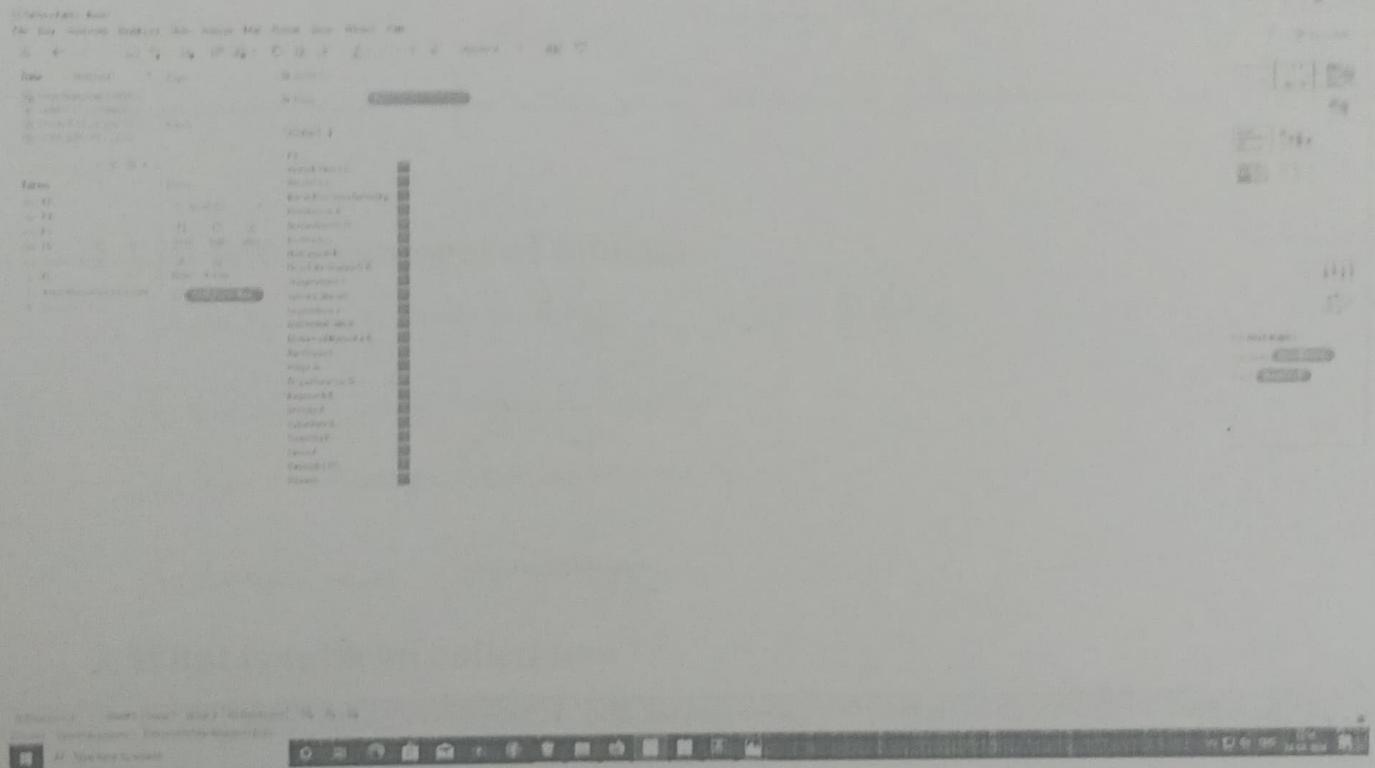
Pie chart:



Bubble chart:



Heat map:



7

1. Is tableau a tool or a software?

Tableau software is an american interactive data visualization software company formed on business intelligence.

2. List few advantages of tableau.

User friendly interface

Data integration

Data exploration

Advance analysis

3. What is tableau called now?

Tableau is known as CRM analytics.

4. Give some tableau dashboard examples

Google analytics is the most well known example of an analytics dashboard.

5. how many visualizations are there in tableau

The tableau chart catalog list 100 chart types with links to actual visualization created by 74 unique authors.

Ex. No:	8	Dashboards
Date:	14/8/24	

AIM

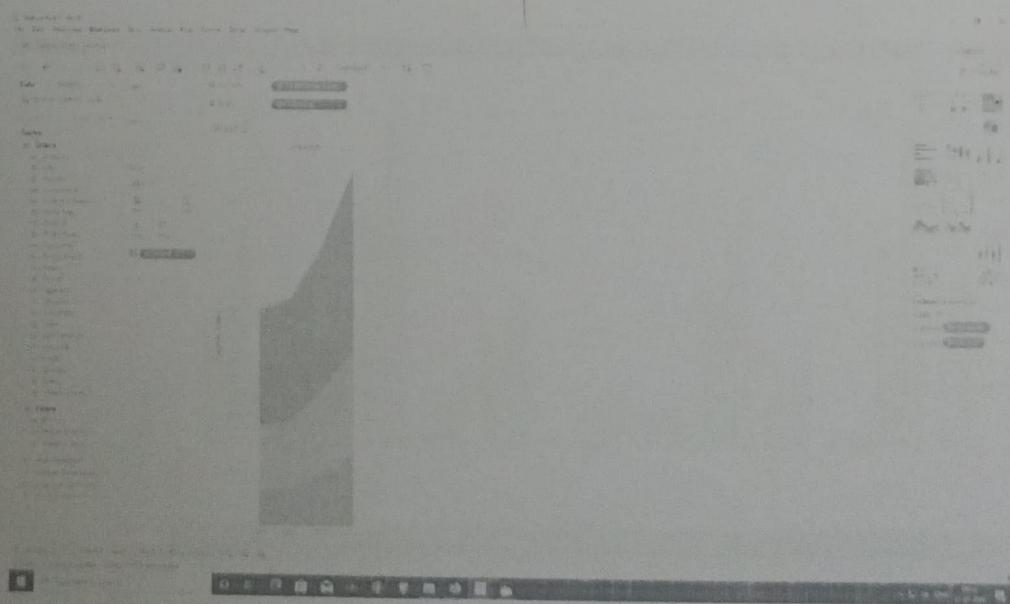
To implements Dashboards using tableau

Description:

A dashboard is a way of displaying various types of visual data in one place. Usually, a dashboard is intended to convey different, but related information in an easy-to-digest form. And oftentimes, this includes things like key performance indicators (KPI)s or other important business metrics that stakeholders need to see and understand at a glance.

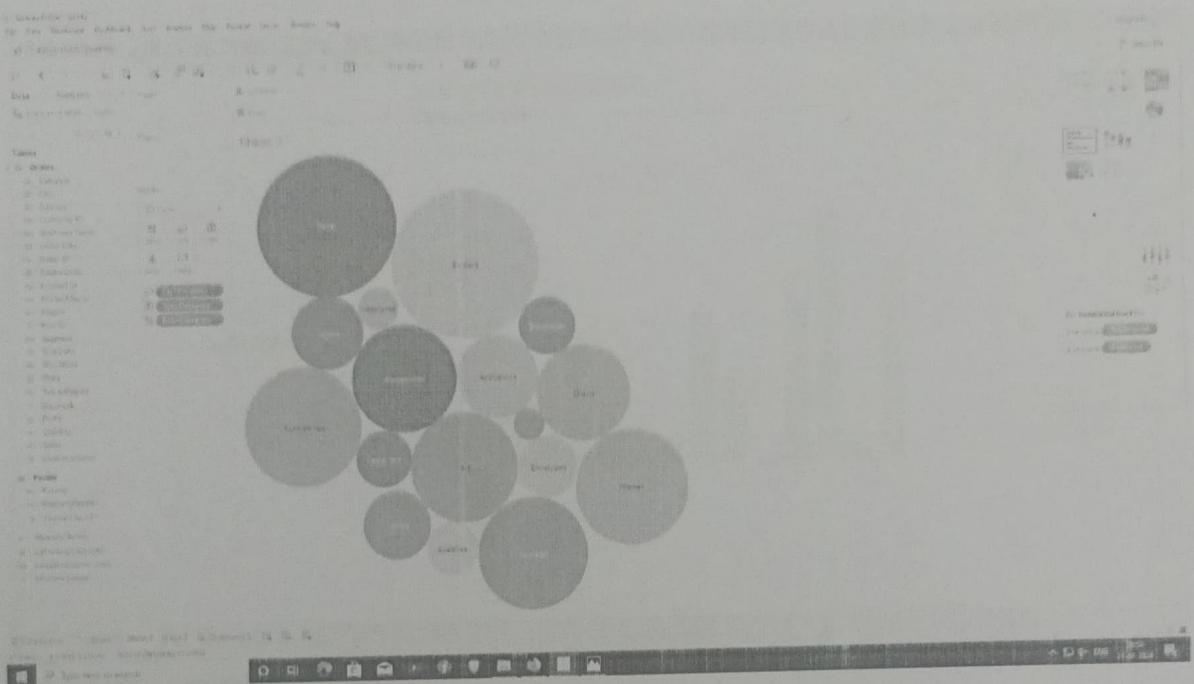
Dashboards are useful across different industries and verticals because they're highly customizable. They can include data of all sorts with varying date ranges to help you understand: what happened, why it happened, what may happen, and what action should be taken.

For example, category of sales across months in a year, region is the field added. The first view is shown below. This can be renamed at the bottom of the screen.

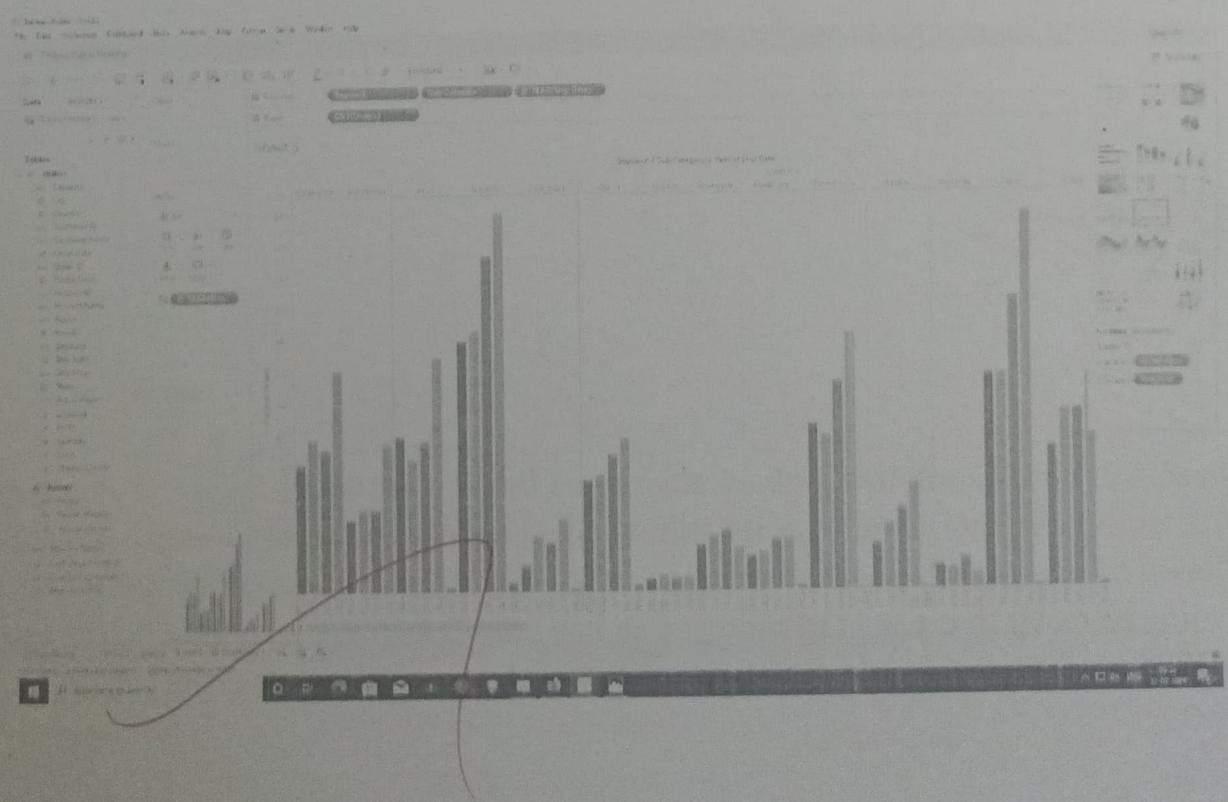


Now go to 2nd sheet for creating the 2nd view. The second view is shown below. A bubble chart was drawn between profit and subcategory. Then

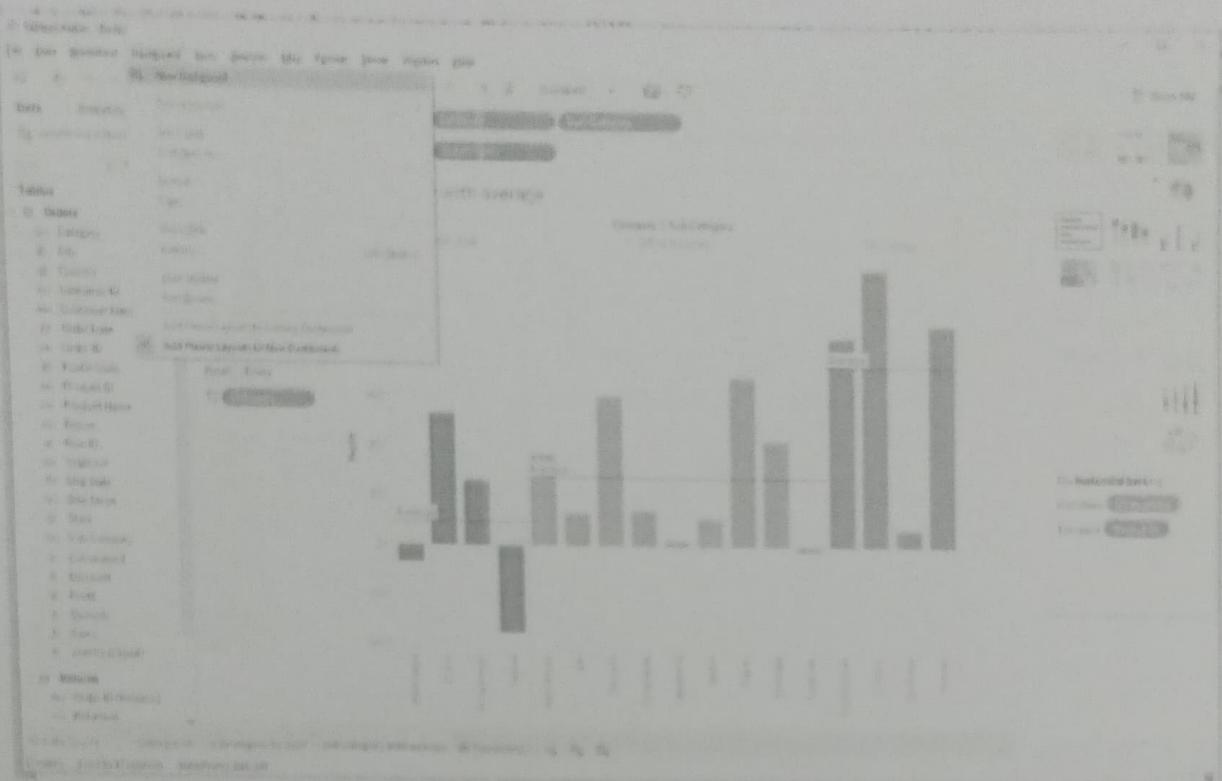
rename the sheet.



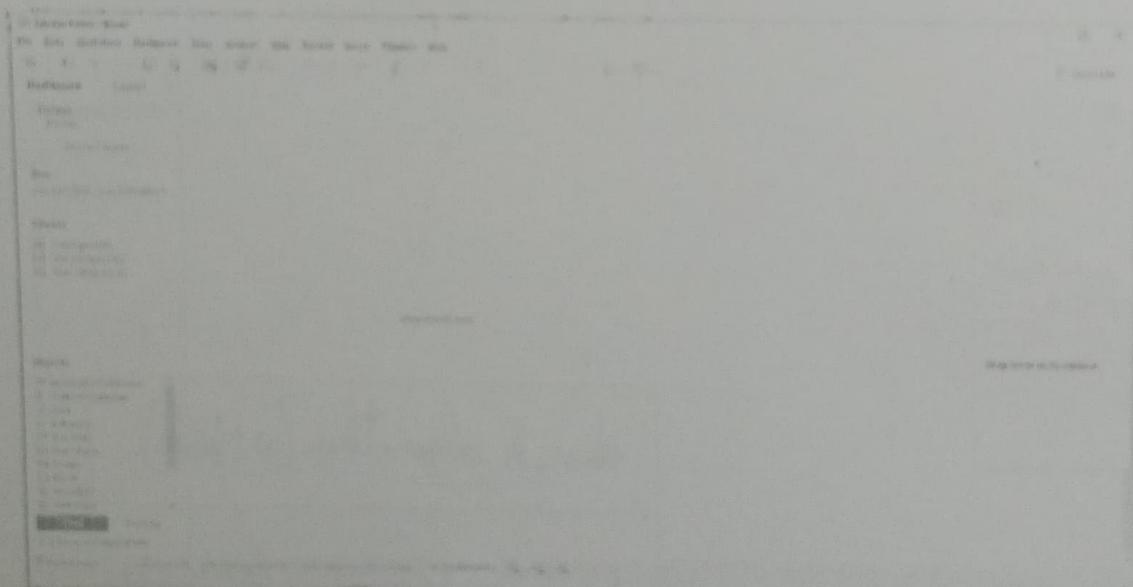
Next 3rd view is created as follows for profit for each subcategory in the category with averages.



After creating individual views, now a65Dashboard can be created by clicking on create dashboard at the toolbar.

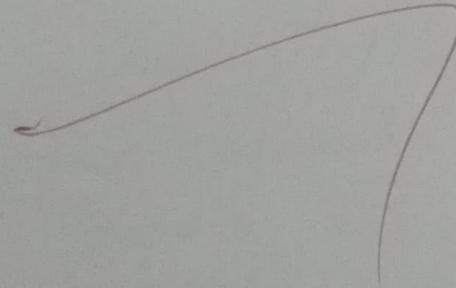
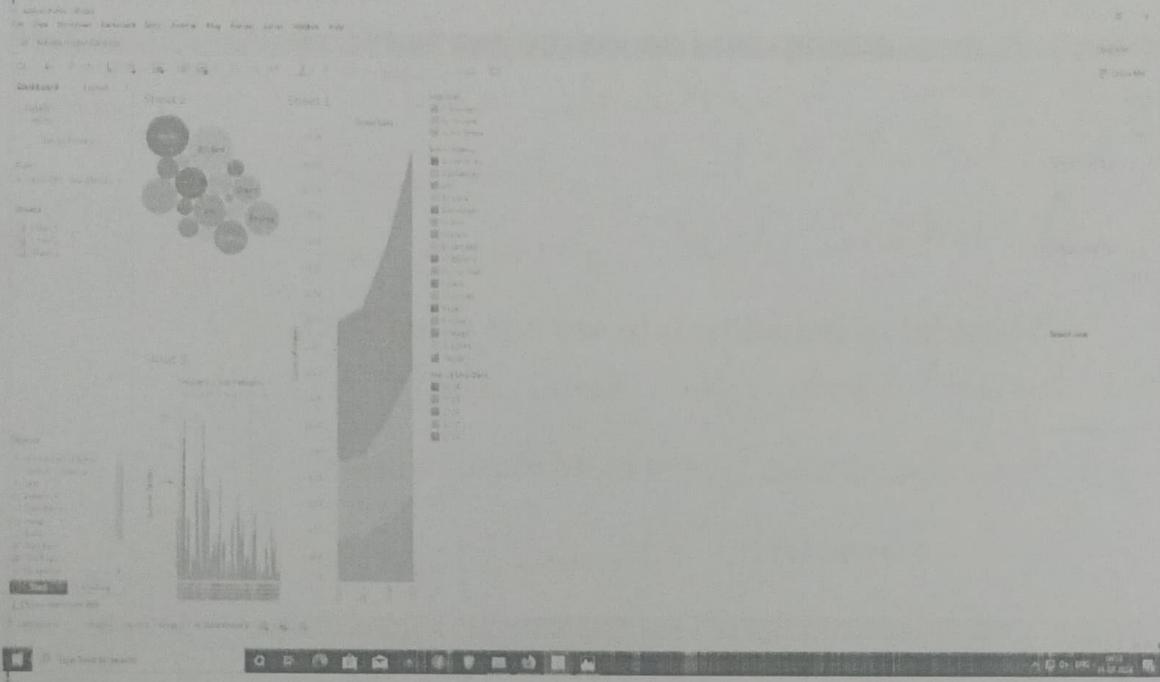


after clicking on new dashboard option, the screen is shown below.



now the sheets or views which are created earlier can be drag and dropped on this dashboard. The above three created views are placed in the dashboard as follows. One can follow their own way of importing sheets on the dashboard. After creating dashboard, title can be given to the dashboard from Dashboard tab. Dahsboard can be customized in terms of its appearance by the user if required. Dashboard once created can be saved on users system and can be retrieved whenever required.

OUTPUT



VIVA VOCE:

1. What are dashboards in tableau?

Data dashboards are a summary of different, but related datasets, presented in a way that makes the related information.

2. What is the use of dashboard in tableau?

The use of dashboard is to show a comprehensive overview of data from different source.

3. List the difference between dashboard and report in tableau?

Dashboard is to visual and interactive allowing the user to engage with the information and create own analysis.

Reports are static they other details lost the end users.

4. What are the components of tableau?

- * Application server
- * Visual server
- * Data server
- * Backgrounder

5. List any 5 benefits of dashboard?

- * Real time monitoring
- * Better decision making
- * Increased probability
- * Data visualization
- * Efficiency . 68

Ex. No:	9
Date:	21/8/24

TIME SERIES ANALYSIS

AIM

To Analysis Time series visualization

What Is a Time Series?

A time series is a set of data points that are collected over a period of time, usually at regular intervals. The most common type of time series data is financial data, such as stock prices or exchange rates. However, time series can also be used to track other types of information, such as meteorological data or sales figures.

Time series can be either univariate or multivariate.

- Univariate Time Series Data: Based on one variable, such as stock prices or the number of cases of a disease
- Multivariate Time Series Data: Based on multiple variables, such as weather data (which could include variables such as temperature, humidity, and rainfall)

Time series are often graphed to visualize the data, and they can be analyzed using statistical methods. Time series analysis can be used for forecasting future values, and it is a powerful tool for understanding complex data.

Types of Time Series Data

There are two main types of time series data:

Continuous data: This type of data is collected at regular intervals and can be represented by a line on a graph. For example, data from a thermometer would be considered continuous data.

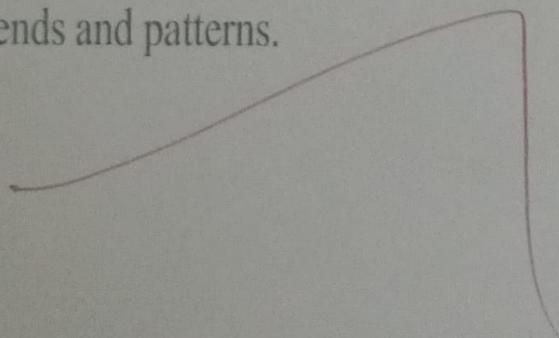
Discrete data: This type of data is collected at specific points in time and can be represented by a dot on a graph. For example, data from a survey would be considered discrete data.

Continuous data is more common than discrete data since most real-world phenomena are continuous.

For example, the population of a country is a continuous variable that changes over time—it doesn't jump from one value to another. In contrast, the results of an election are discrete since there are only a finite number of outcomes (e.g., candidate A wins, candidate B wins, etc.).

The Best Way to Visualize Time Series Data

Time series line graphs are the best way to visualize data that changes over time. This is because line graphs show how a variable changes from one point in time to another, making it easy to see trends and patterns.



Tesla Inc

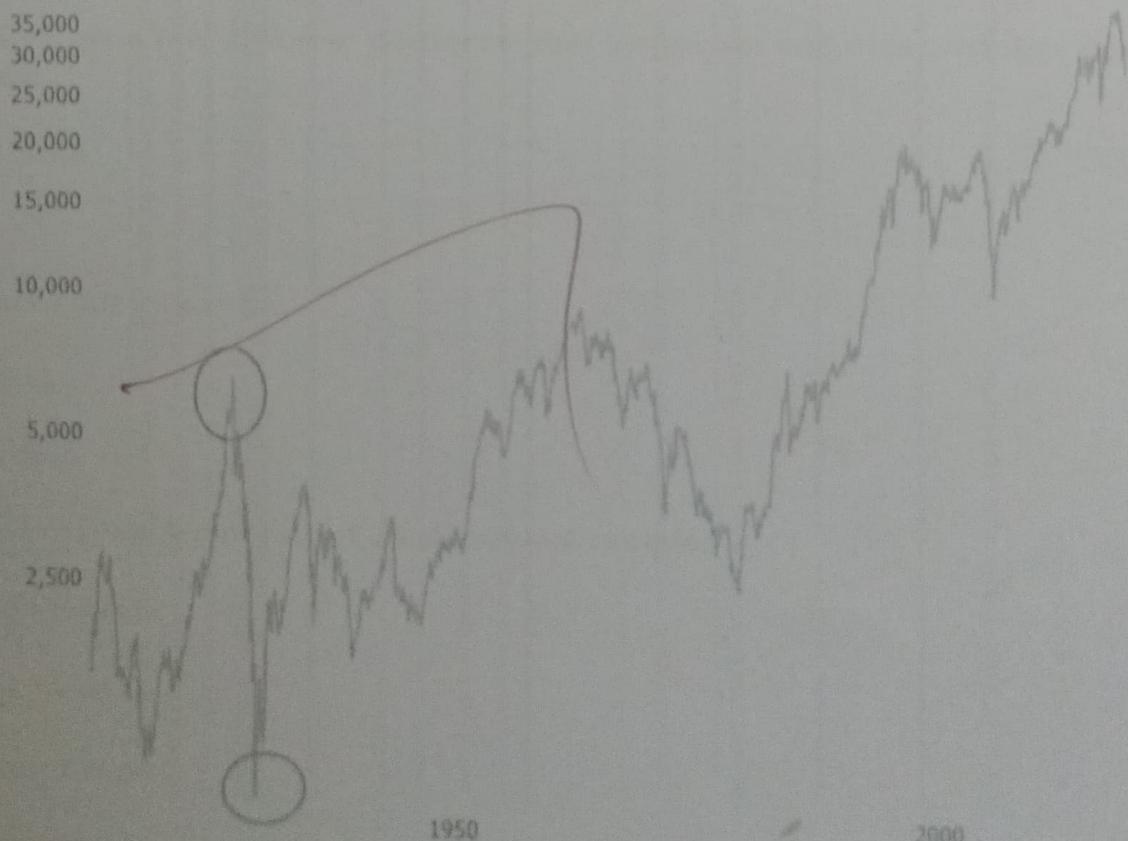
\$228.52 +38.46% 142.81 1Y

After Hours: \$228.35 (+0.074%) -0.17

Closed: Oct 28, 7:59:45 PM UTC 4· USD · NASDAQ · Disclaimer

1D 5D 1M 6M YTD 1Y 5Y MAX

Key events



Line graphs are also useful for identifying specific points in time when there was a sudden change in the data (known as an anomaly).

When to Use Other Temporal Visualizations

Other types of graphs can be used to visualize time series data, but they are less common.

For example, you could use a scatter plot to visualize how two variables are related. This would happen in cases where you have multivariate time series data.

For example, consider the following scatter plot of stock prices and interest rates:

The scatter plot shows that there is a positive relationship between stock prices and interest rates—as one variable increases, the other variable also tends to increase.

This relationship would be much harder to see if the data were presented in a line graph.

Another example would be a bar graph, which could be used to compare different time periods. For example, you could use a bar graph to compare the stock prices of two different companies:

The bar graph makes it easy to see that Company A's stock price is higher than Company B's stock price.

However, the bar graph is not as useful for seeing trends over time since it doesn't show how the data changes from one point in time to another. Best Platforms to Visualize Data

There are a few different platforms that businesses and data scientists use to visualize data:

Tableau

R

Microsoft Power BI

Excel

Python

Seven Time Series Data Visualization Examples

Line Graphs

Bar Graphs

Gantt Charts

Heat Maps

Time Series Data Visualization using Python

We will use Python libraries for visualizing the data.

Importing the Libraries

We will import all the libraries that we will be using throughout this article in one place so that do not have to import every time we use it this will save both our time and effort.

- **Numpy** – A Python library that is used for numerical mathematical computation and handling multidimensional ndarray, it also has a very large collection of mathematical functions to operate on this array.
- **Pandas** – A Python library built on top of NumPy for effective matrix multiplication and dataframe manipulation, it is also used for data cleaning, data merging, data reshaping, and data aggregation.
- **Matplotlib** – It is used for plotting 2D and 3D visualization plots, it also supports a variety of output formats including graphs for data.

```
import pandas as pd  
import numpy as np  
import seaborn as sns  
import matplotlib.pyplot as plt  
from statsmodels.graphics.tsaplots import plot_acf  
from statsmodels.tsa.stattools import adfuller
```

Loading The Dataset

To load the dataset into a dataframe we will use the pandas `read_csv()` function. We will use `head()` function to print the first five rows of the dataset. Here we will use the ‘`parse_dates`’ parameter in the `read_csv` function to convert the ‘Date’ column to the DatetimeIndex format. By default, Dates are stored in string format which is not the right format for time series data analysis.

```
# reading the dataset using read_csv  
df = pd.read_csv("stock_data.csv",  
                  parse_dates=True,  
                  index_col="Date")  
  
# displaying the first five rows of dataset  
df.head()
```

OutPut:



Result :

Thus the Time series analysis for visualization using python has been executed successfully.

21/8/24