DATA VISUALIZATION LAB

(KG21CD608)

LABORATORY MANUAL

B.TECH CSE - DS

(III YEAR–II SEM)(2023-2024)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING – DATA SCIENCE

KG REDDY COLLEGE OF ENGINEERING AND TECHNOLOGY

Affiliated to JNTUH, Chilkur,(V), Moinabad(M) R. R Dist, TS-501504

DEPARTMENT OF COMPUTER SCIENCE & INFORMATION TECHNOLOGY

Vision & Mission

Vision

* To achieve high quality in technical education that provides the skills and attitude to adapt to the global needs of the Information Technology sector, through academic and research excellence.

Mission

* To equip the students with the cognizance for problem solving and to improve the teaching learning pedagogy by using innovative techniques.

* To strengthen the knowledge base of the faculty and students with motivation towards possession of effective academic skills and relevant research experience.

* To promote the necessary moral and ethical values among the engineers, for the betterment of the society.

Quality Policy

* Strives to inculcate the students with the world class Technical Knowledge, Entrepreneurial Competence and Social Ethics by providing continual improvement and innovation in the curriculum; based upon well-defined measurements and best practices.

* Develop faculty competencies, creativity, empowerment and accountability through faculty development programs and show strong management involvement and commitment.

PROGRAMME EDUCATIONAL OBJECTIVES (PEOs)

PEO1 – ANALYTICAL SKILLS:

- To facilitate the graduates with the ability to visualize, gather information, articulate, analyze, solve complex problems, and make decisions. These are essential to address the challenges of complex and computation intensive problems increasing their productivity.

PEO2 – TECHNICAL SKILLS:

- To facilitate the graduates with the technical skills that prepare them for immediate employment and pursue certification providing a deeper understanding of the technology in advanced areas of computer science and related fields, thus encouraging to pursue higher education and research based on their interest.

PEO3 – SOFT SKILLS:

- To facilitate the graduates with the soft skills that include fulfilling the mission, setting goals, showing self-confidence by communicating effectively, having a positive attitude, get involved in team-work, being a leader, managing their career and their life.

PEO4 – PROFESSIONAL ETHICS:

- To facilitate the graduates with the knowledge of professional and ethical responsibilities by paying attention to grooming, being conservative with style, following dress codes, safety codes, and adapting to technological advancements.

PROGRAM SPECIFIC OUTCOMES (PSOs)

After the completion of the course, B. Tech Information Technology, the graduates will have the following Program Specific Outcomes:

1. **Fundamentals and critical knowledge of the Computer System:-** Able to Understand the working principles of the computer System and its components , Apply the knowledge to build, asses, and analyze the software and hardware aspects of it .

2. **The comprehensive and Applicative knowledge of Software Development:** Comprehensive skills of Programming Languages, Software process models, methodologies, and able to plan, develop, test, analyze, and manage the software and hardware intensive systems in heterogeneous platforms individually or working in teams.

3. **Applications of Computing Domain & Research:** Able to use the professional, managerial, interdisciplinary skill set, and domain specific tools in development processes, identify the research gaps, and provide innovative solutions to them.

PROGRAM OUTCOMES (POs)

Engineering Graduates will be able to:

1. Engineering knowledge: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.

2. Problem analysis: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

3. Design / development of solutions: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.

4. Conduct investigations of complex problems: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

5. Modern tool usage: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.

6. The engineer and society: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.

7. Environment and sustainability: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

8. Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

9. Individual and team work: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.

10. Communication: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.

11. Project management and finance : Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multi disciplinary environments.

12. Life-long learning: Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change

K G REDDY COLLEGE OF ENGINEERING AND TECHNOLOGY

III Year B. TECH- CSE-DS – II - SEM **L/T/P/C**

-/-/2/1

(KG21CD608) DATA VISUALIZATION LAB

COURSE OBJECTIVES:

Students will be able to

1. Understand the importance of data visualization for business intelligence and decision making.
2. Know approaches to understand visual perception
3. Learn about categories of visualization and application areas
4. Familiarize with the data visualization tools
5. Gain knowledge of effective data visuals to solve workplace problems

1. Introduction to various Data Visualization tools
2. Basic Visualization in Python
3. Basic Visualization in R
4. Introduction to Tableau and Installation
5. Connecting to Data and preparing data for visualization in Tableau
6. Data Aggregation and Statistical functions in Tableau
7. Data Visualizations in Tableau
8. Basic Dashboards in Tableau

COURSE OUTCOMES:

At the end of the course, Students will be able to:

1. Use Python, R and Tableau for data visualization
2. Apply data visuals to convey trends in data over time using tableau
3. Construct effective data visuals to solve workplace problems
4. Explore and work with different plotting libraries
5. Learn and create effective visualizations

Reference Books:

1. 1. Data visualization with python: create an impact with meaningful data insights using interactive and engaging visuals, Mario Dobler, Tim Grobmann, Packt Publications, 2019
2. Practical Tableau: 100 Tips, Tutorials, and Strategies from a Tableau Zen Master, Ryan Sleeper, Oreilly Publications, 2018
3. Data Visualization with R: 111 Examples by Thomas Rahlf, Springer, 2020

TABLE OF CONTENTS

Week 1: Introduction to various Data Visualization tools

Data visualization is the practice of translating information into a visual context, such as a map or graph, to make data easier for the human brain to understand and pull insights from. It is the representation of information and data through use of common graphics, such as charts, plots, infographics, and animations. Data visualization is a powerful way for people, especially data professionals, to display data so that it can be interpreted easily.

Data Visualization enables decision-makers of any enterprise or industry to look into analytical reports and understand concepts that might otherwise be difficult to grasp.

Benefits of Data Visualization:

1. It is easy to understand the information with graphics
2. It made data to be represented in attractive way
3. Shows complex relationships
4. Helps to process large datasets
5. Useful for identifying trends
6. Minimizes ambiguity

Data visualization tools provide the ability to see and understand data trends, outliers, and patterns in an easy, intuitive way. There are various data visualization tools available. One must choose the tool based on various factors such as its ease of use, types of graphical representations the tool can produce, size of the dataset the tool can handle etc. some of Data Visualization tools are Tableau, Power BI, Google Charts, JupyteR, Grafana etc.

The following are some common types of data visualizations:

- **Table:** A table is data displayed in rows and columns, which can be easily created in a Word document or Excel spreadsheet.
- **Chart or graph:** Information is presented in tabular form with data displayed along an x and y axis, usually with bars, points, or lines, to represent data in comparison.
- **Geospatial visualization:** Data is depicted in map form with shapes and colours that illustrate the relationship between specific locations, such as a choropleth or heat map.
- **Dashboard:** Data and visualizations are displayed, usually for business purposes, to help analysts understand and present data.

Week 2: Basic Visualization in Python

Python has different modules for visualizing data such as matplotlib, seaborn. Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python. It presents data in 2D graphics. Seaborn is a visualization library that is built on top of Matplotlib. It provides data visualizations that are typically more aesthetic and statistically sophisticated. Matplotlib can be installed using the following command:

pip install matplotlib

Once the module installed, it must be imported into the program using the following command

import matplotlib as mpl, where mpl is the alias name given to matplotlib library.

matplotlib.pyplot is a state-based interface to matplotlib. **matplotlib.pyplot** is a collection of functions that make matplotlib work like MATLAB. Each pyplot function makes some change to a figure: e.g., creates a figure, creates a plotting area in a figure, plots some lines in a plotting area, decorates the plot with labels etc. pyplot can be imported into the program using following command

import matplotlib.pyplot as plt

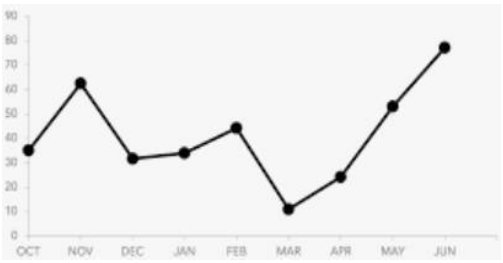
Following are some of the basic data visualization plots

1. Line plots
2. Area plots
3. Histograms
4. Bar charts
5. Pie charts
6. Box plots
7. Scatter plots

Line Plots:

A line plot is used to represent quantitative values over a continuous interval or time period. It is generally used to depict trends on how the data has changed over time.

Example:



Program:

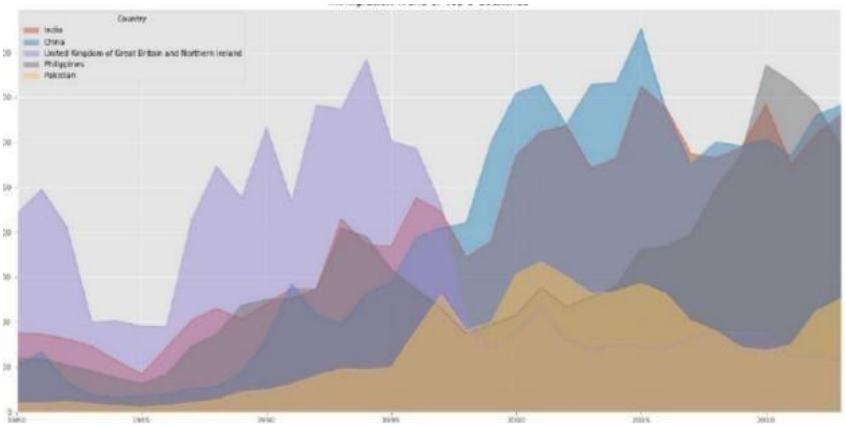
```
import matplotlib.pyplot as plt
x = [1, 2, 3, 4, 5, 6]
y = [1, 5, 3, 5, 7, 8]
plt.plot(x, y)
plt.show()
```

Output:

Area Plots:

An Area Plot is also called as Area Chart which is used to display magnitude and proportion of multiple variables.

Example:



Program:

```
import matplotlib.pyplot as plt
days = [1,2,3,4,5]
sleeping =[7,8,6,11,7]
eating = [2,3,4,3,2]
working =[7,8,7,2,2]
playing = [8,5,7,8,13]
plt.plot([],[],color='m', label='Sleeping', linewidth=5)
plt.plot([],[],color='c', label='Eating', linewidth=5)
```

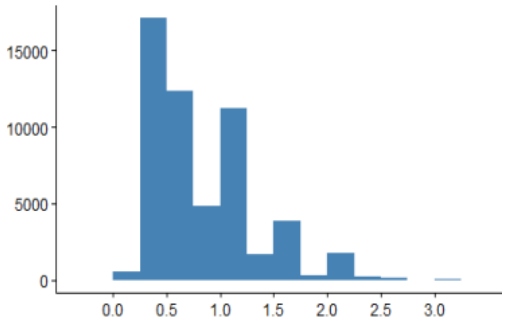
```
plt.plot([],[],color='r', label='Working', linewidth=5)
plt.plot([],[],color='k', label='Playing', linewidth=5)
plt.stackplot(days, sleeping,eating,working,playing, colors=['m','c','r','k'])
plt.xlabel('x')
plt.ylabel('y')
plt.title('Stack Plot')
plt.legend()
plt.show()
```

Output:

Histograms:

Histograms represents the frequency distribution of a dataset. It is a graph showing the number of observations within each given interval.

Example:



Program:

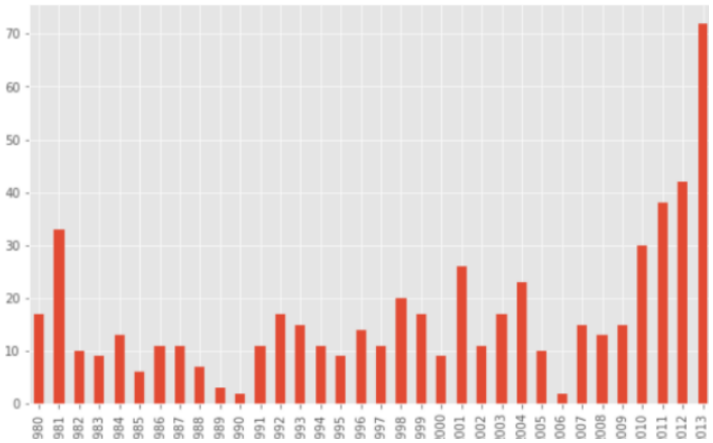
```
import matplotlib.pyplot as plt
population_age=[22,55,62,45,21,22,34,42,42,4,2,102,95,85,55,110,120,70,65,55,111,115,80]
bins = [0,10,20,30,40,50,60,70,80,90,100]
plt.hist(population_age, bins, histtype='bar', rwidth=0.8)
plt.xlabel('age groups')
plt.ylabel('Number of people')
plt.title('Histogram')
plt.show()
```

output:

Bar Charts:

A Bar chart or bar graph is a chart or graph that presents categorical data with rectangular bars with heights or lengths proportional to the values that they represent. A bar plot is a way of representing data where the length of the bars represents the magnitude/size of the feature/variable.

Example:



Program:

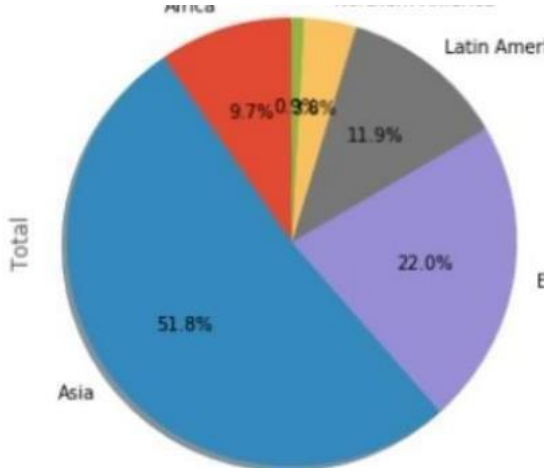
```
from matplotlib import pyplot as plt
plt.bar([0.25,1.25,2.25,3.25,4.25],[50,40,70,80,20],label="BMW",width=.5)
plt.bar([.75,1.75,2.75,3.75,4.75],[80,20,20,50,60],label="Audi", color='r',width=.5)
plt.legend()
plt.xlabel('Days')
plt.ylabel('Distance (kms)')
plt.title('Information')
plt.show()
```

Output:

Pie Charts:

A Pie chart is a circular statistical chart, which is divided into sectors to illustrate numerical proportion.

Example:



Program:

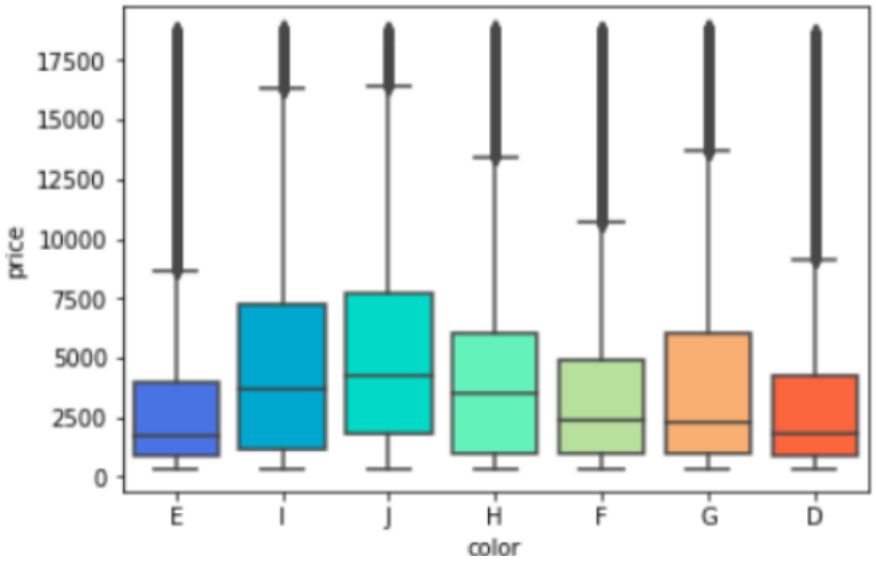
```
import matplotlib.pyplot as plt
days = [1,2,3,4,5]
sleeping =[7,8,6,11,7]
eating = [2,3,4,3,2]
working =[7,8,7,2,2]
playing = [8,5,7,8,13]
slices = [7,2,2,13]
activities = ['sleeping','eating','working','playing']
cols = ['c','m','r','b']
plt.pie(slices, labels=activities, colors=cols, startangle=90, shadow= True, explode=(0,0.1,0,0), autopct='%1.1f%%')
plt.title('Pie Plot')
plt.show()
```

Output:

Box Plots:

A Box plot (or box-and-whisker plot) shows the distribution of quantitative data in a way that facilitates comparisons between variables or across levels of a categorical variable. Box plot shows the quartiles of the dataset while the whiskers extend encompass the rest of the distribution but leave out the points that are the outliers.

Example:



Program:

```
import matplotlib.pyplot as plt
x=[1,2,3,4,5,6,7]
y=[1,2,4,5,3,6,9]
z=[x,y]
plt.boxplot(z,labels=[“A”,”B”],showmeans=True)
plt.show()
```

Output:

Scatter Plots:

A Scatter chart, also called a scatter plot, is a chart that shows the relationship between two variables.

Program:

```
import matplotlib.pyplot as plt
x=[1,1.5,2,2.5,3,3.5,3.6]
y=[7.5,8,8.5,9,9.5,10,10.5]
x1=[8,8.5,9,9.5,10,10.5,11]
y1=[3,3.5,3.7,4,4.5,5,5.2]
plt.scatter(x,y, label='high income low saving',color='r')
plt.scatter(x1,y1,label='low income high savings',color='b') plt.xlabel('saving*100')
plt.ylabel('income*1000')
plt.title('Scatter Plot')
plt.legend()
plt.show()
```

Output:

Week 3: Basic Visualization in R

1. Scatter plots
2. Line plots
3. Box plots
4. Histograms
5. Bar charts

ggplot2 is an open-source data visualization package for the statistical programming language R. ggplot is enriched with customized features to make visualization better. ggplot2 is a system for declaratively creating graphics, based on The Grammar of Graphics. ggplot2 can greatly improve the quality and aesthetics of graphics.

The ggplot2 package can be easily installed using the following R function:

install. packages(ggplot2)

then the following command must be used in program to use ggplot package:

library(ggplot2)

Consider the following dataset named surveys. All the visualizations mentioned above are applied on this dataset.

```
Surveys<-data.frame(record_id=c(1,2,3,4,5),
month=c(7,7,7,7,7),day=c(16,16,16,17,17),year=c(1977,1977,1977,1977,1977),plot_id=c(2,3
,2,7,3),species_id=c(NL,NL,DM,DM,DM),sex=c(M,M,F,M,M),hindfoot_length=c(32,33,37,
36,35))
```

Scatter plot:

```
ggplot(data = surveys, mapping = aes(x = weight, y = hindfoot_length)) + geom_point(alpha
= 0.1, color = "blue“)
```

Output:

Histogram:

```
ggplot(surveys, aes(species) + geom_histogram(binwidth = 2)+ labs(title = "Histogram")
```

Output:

bar chart:

ggplot(surveys, aes(species.id)) + geom_bar(fill = "red")+ labs(title = "Bar Chart")

Output:

Box plot:

ggplot(data = surveys, mapping = aes(x = species_id, y = weight)) + geom_boxplot()

Output:

Line plot:

ggplot(data = yearly_counts, aes(x = year, y = n, group = species_id, colour = species_id)) + geom_line()

Output:

Week 4: Introduction to Tableau and Installation

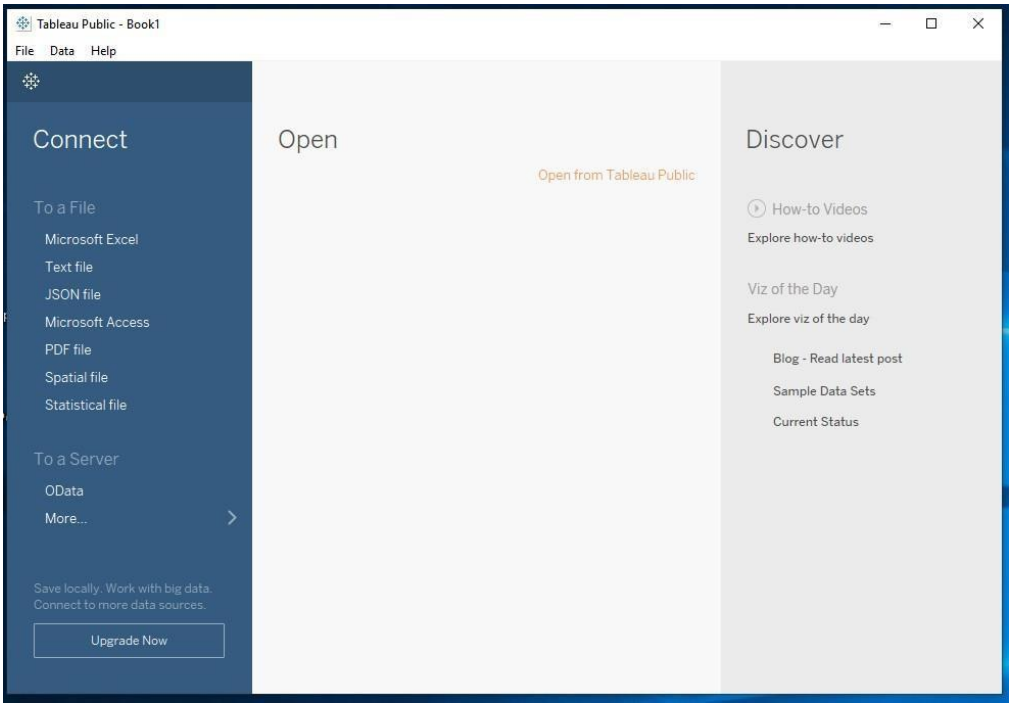
Tableau is a data visualization tool that provides pictorial and graphical representations of data. It is used for data analytics and business intelligence. Tableau provides limitless data exploration without interrupting flow of analysis. With an intuitive drag and drop interface, user can uncover hidden insights in data and make smarter decisions faster.

Tableau can be downloaded from the following website:

https://www.tableau.com/products/public/download

after downloading, the following is the screen appears.

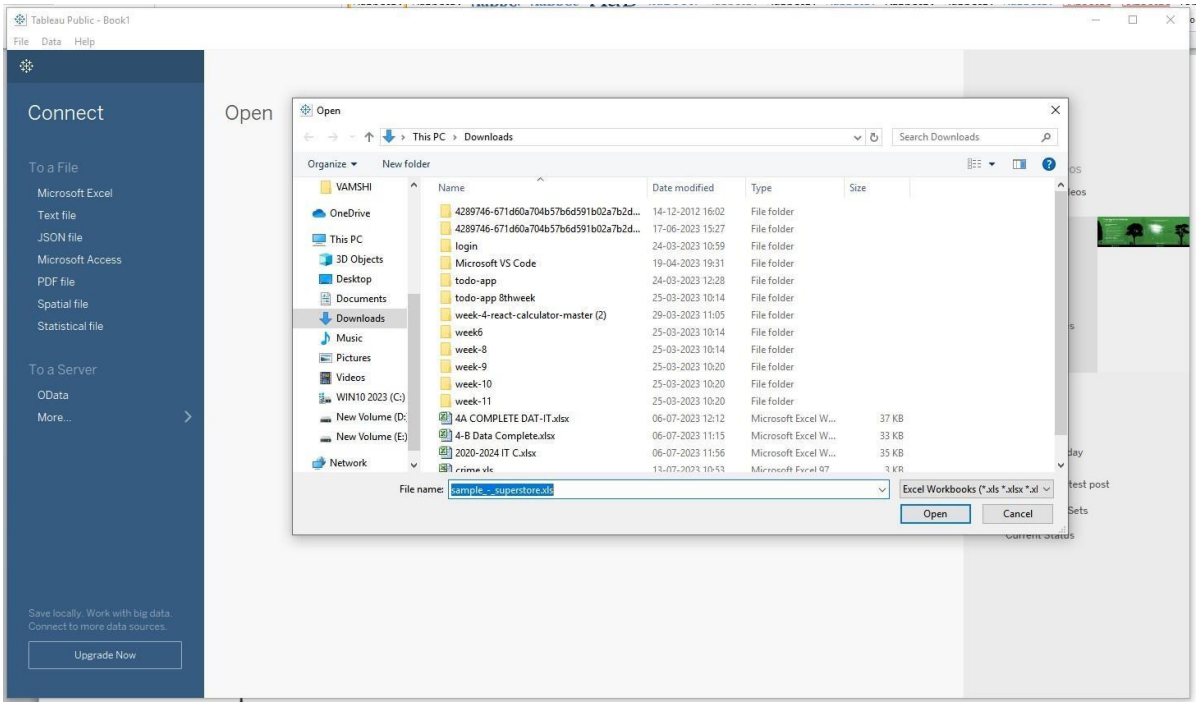
Click the licence agreement checkbox and then click on install button. After installation, click on Tableau Public icon to run Tableau. Following is the Tableau Public home screen.



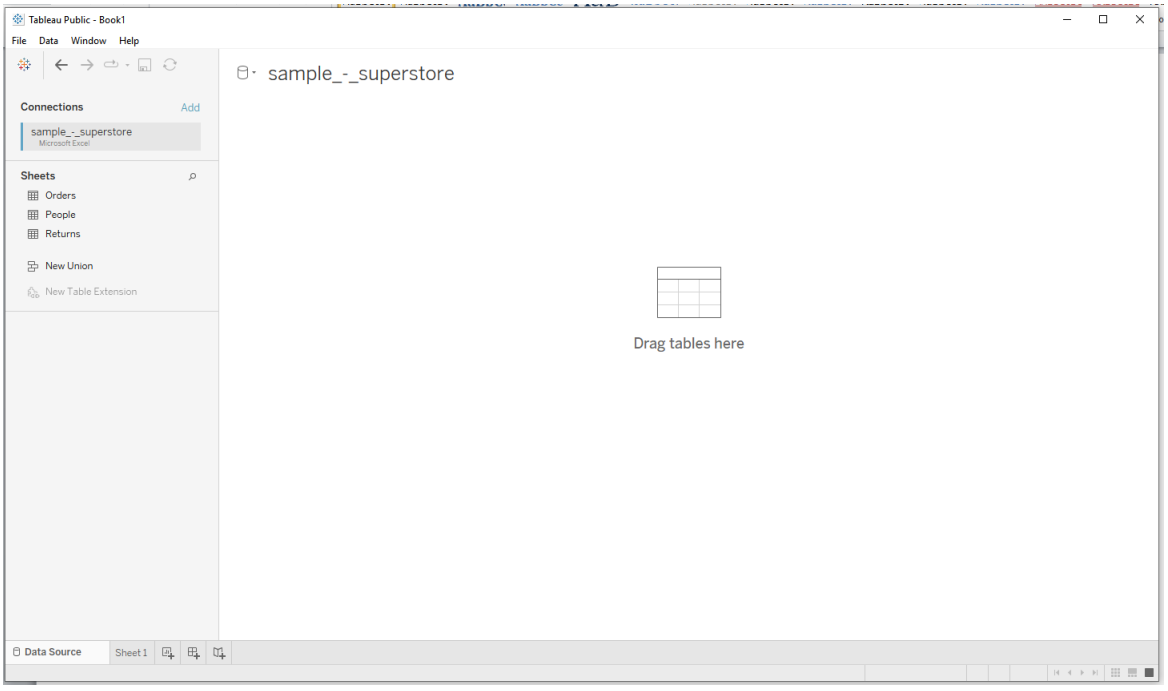
Week 5: Connecting to Data and preparing data for visualization in Tableau

Tableau supports connecting to a wide variety of data, stored in a variety of places. For example, data might be stored on computer in a spread sheet or a text file, or in a big data, relational, or cube (multidimensional) database on a server in enterprise or the data can be from a public domain available on the web.

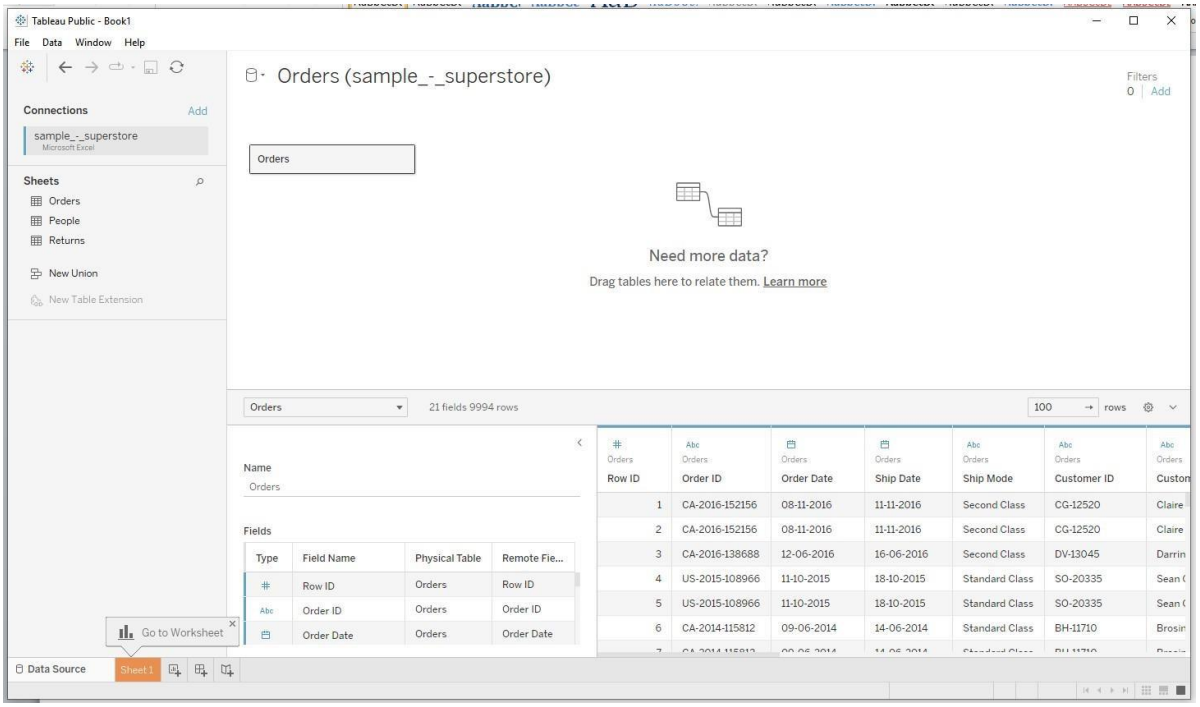
Data can be imported in Tableau Public from Connect panel on left side. For example, an Excel sample data set was loaded into Tableau as follows:



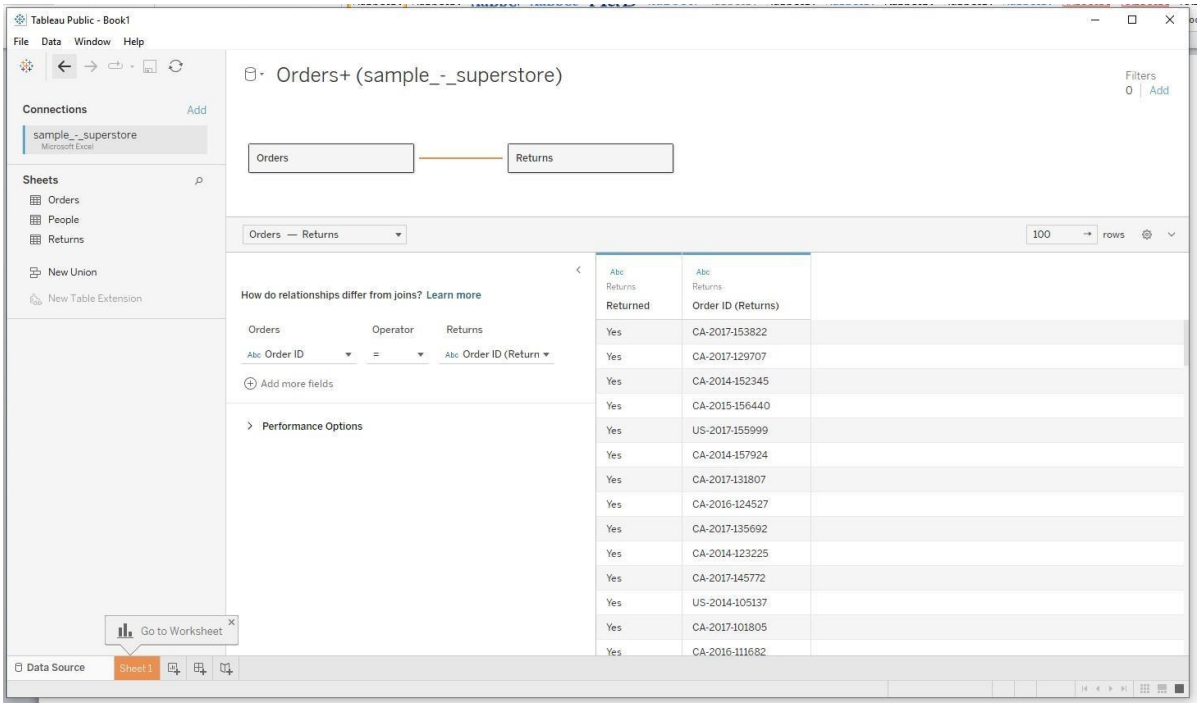
After clicking on open, screen is as follows:



The data store page appears as above. The left pan shows that above dataset consists of 3 worksheets. If we drag orders table, screen appears as follows: Tableau automatically identifies the data type of each column.

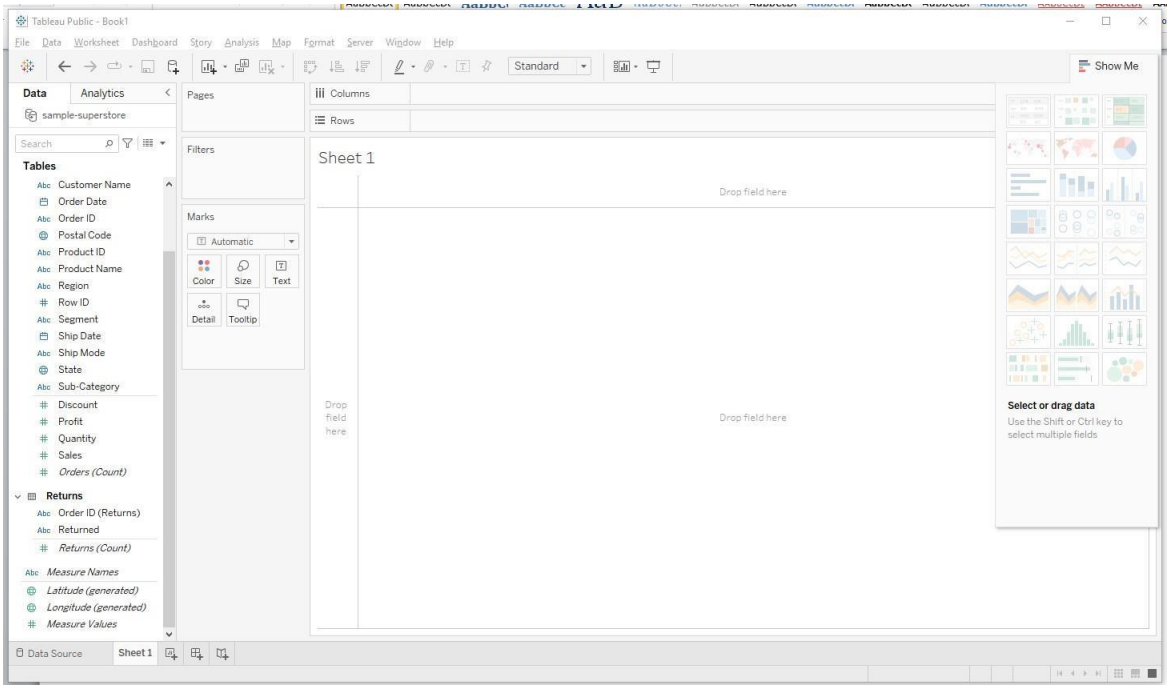


Now drag Returns table onto the Canvas to the right of Orders table. This shows the relation between the two tables Orders and Returns.



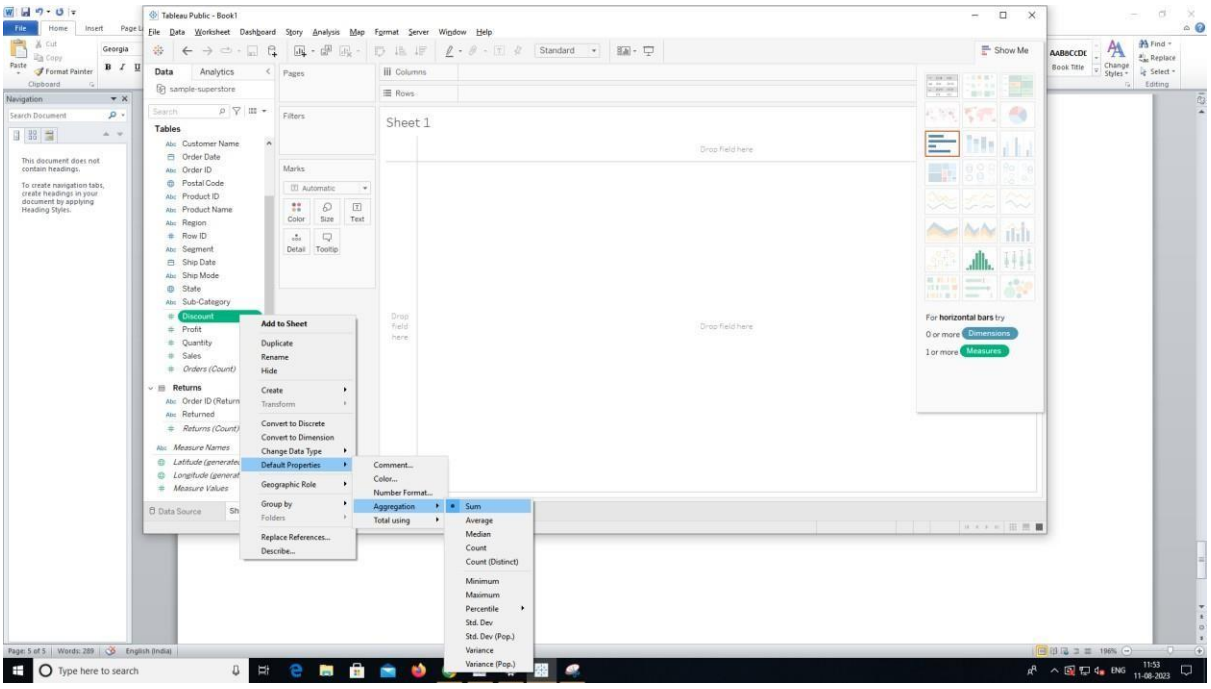
If we click on the link between Orders and Returns table names at the top gives the summary of the relationship between the tables. Now rename the data store and click on Sheet1 at the

bottom left to proceed. This step creates a data extract which improves query performance.

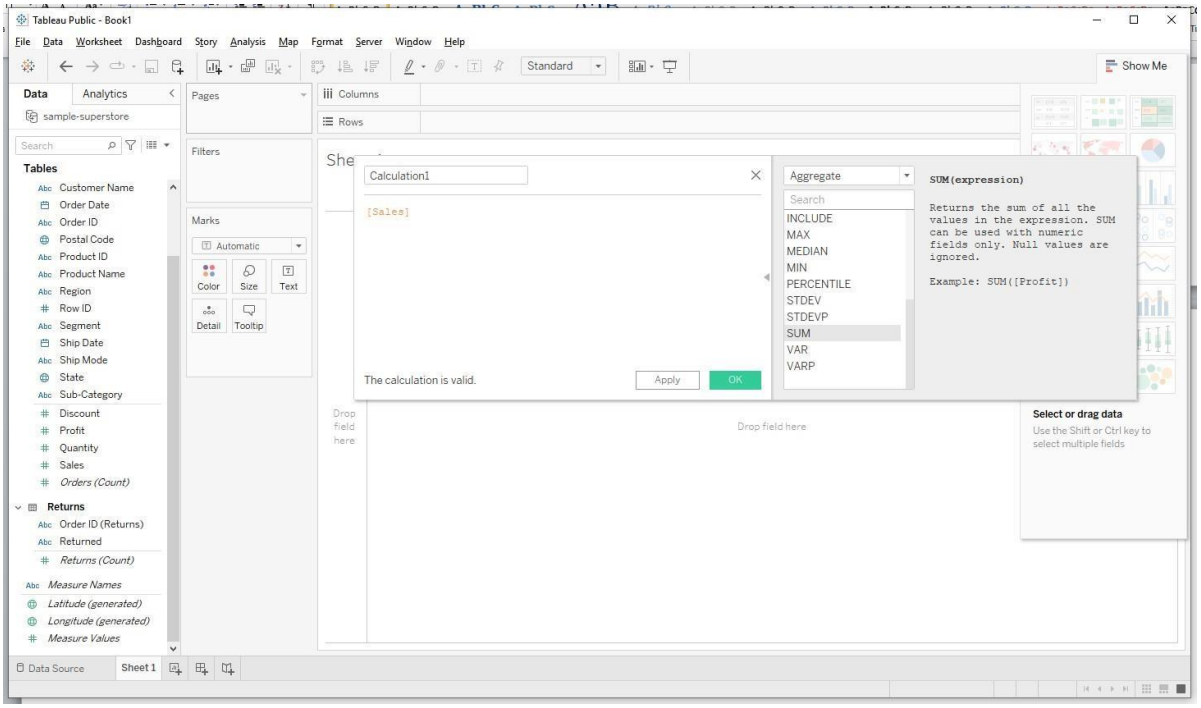


Week 6: Data aggregation and statistical functions

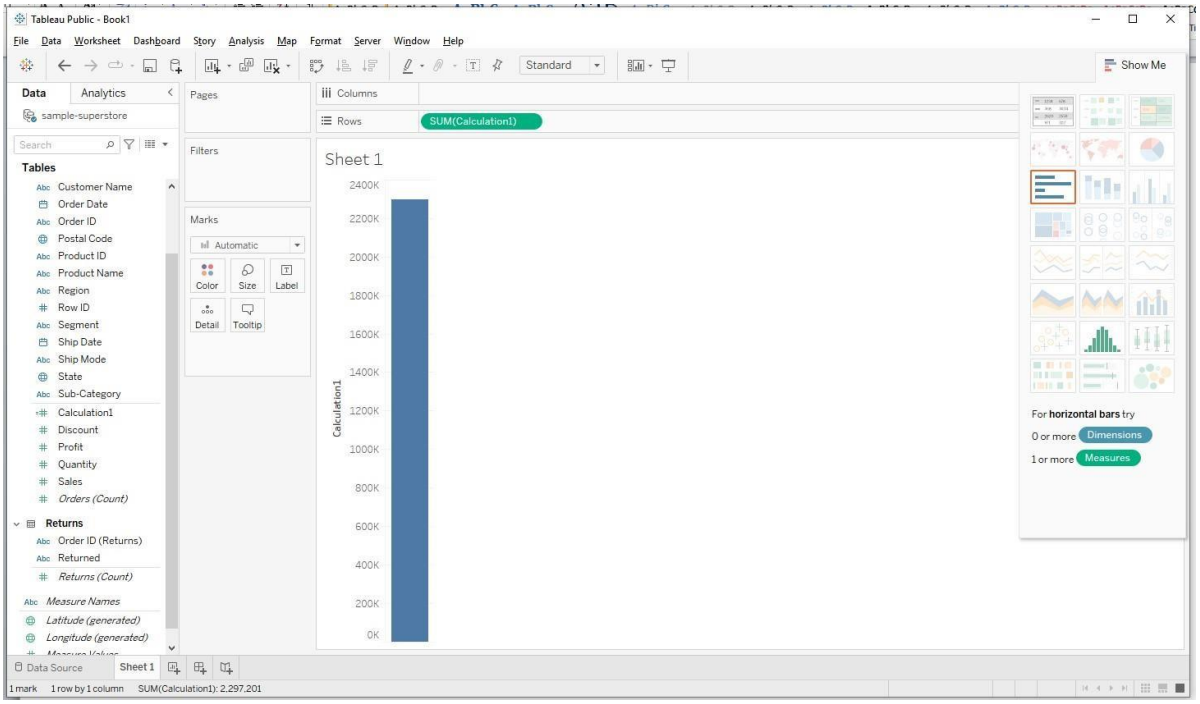
We can apply various aggregation and statistical functions on data such as count, minimum, maximum, standard deviation, variance etc. This is shown below. This can be done by right clicking on the required field of dataset, click on Default properties and click on aggregation.



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.



Then click on apply and results are shown below:



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

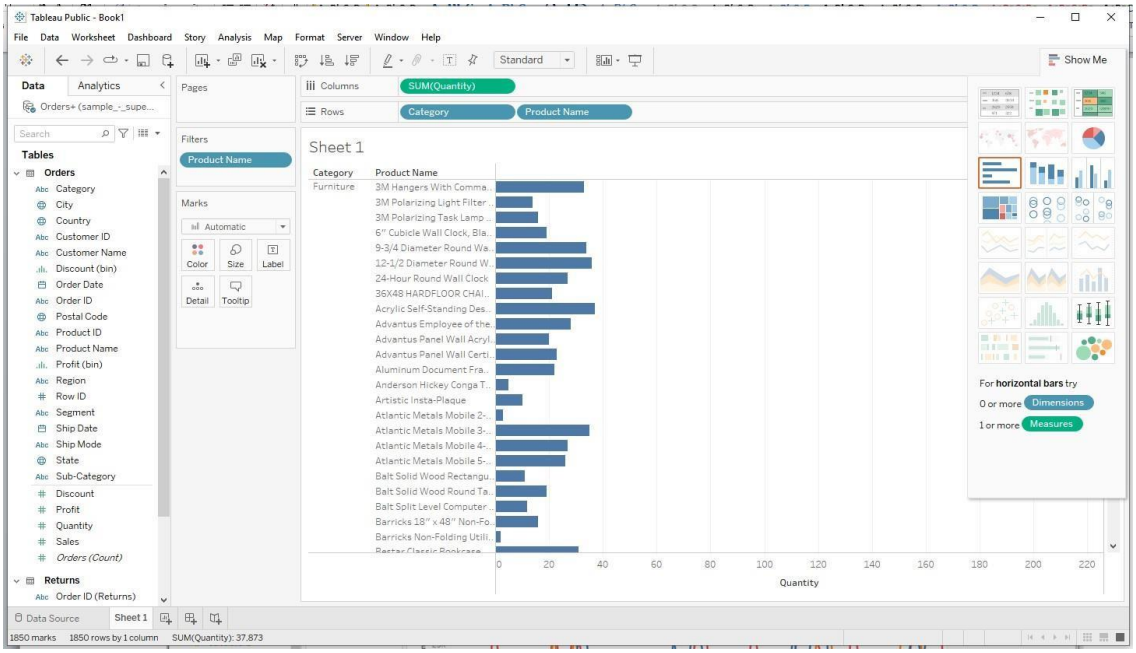
Week 7: Data Visualization

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

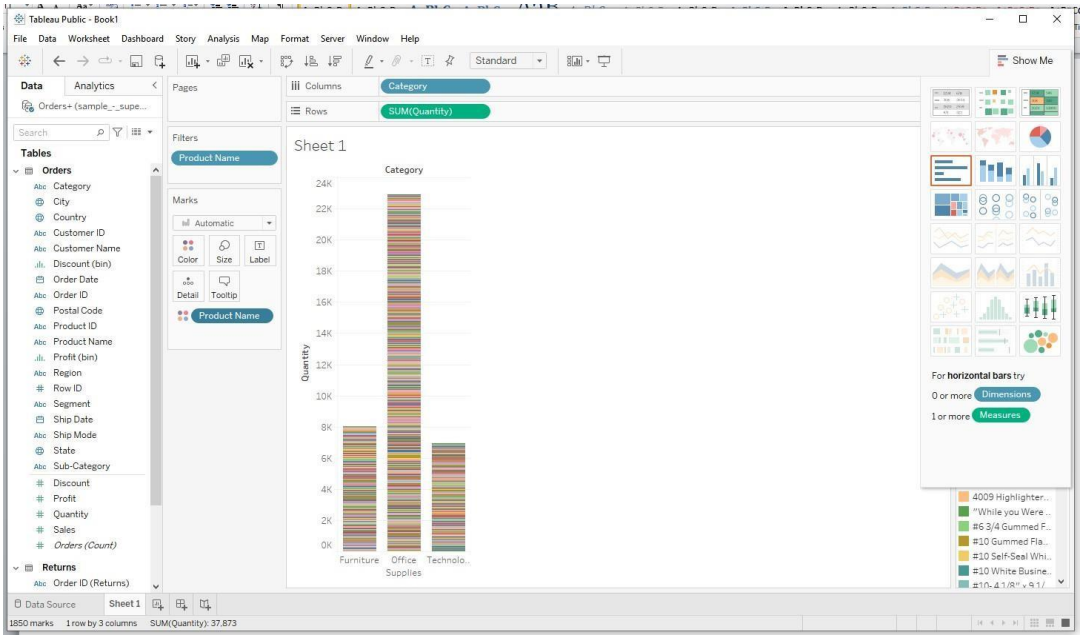
Bar chart:

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

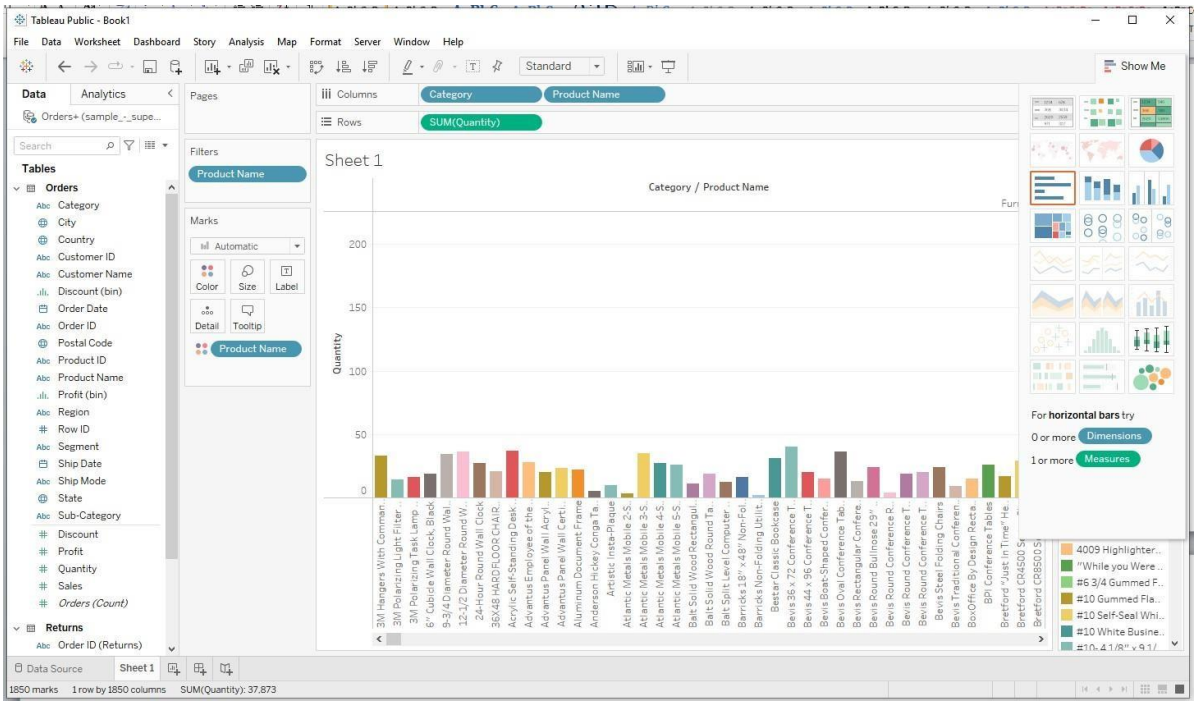
Horizontal bars can be created by selecting that type of chart from Show Me menu on right hand side of Canvas. The type of chart in box on right hand side represents horizontal bar graph.



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

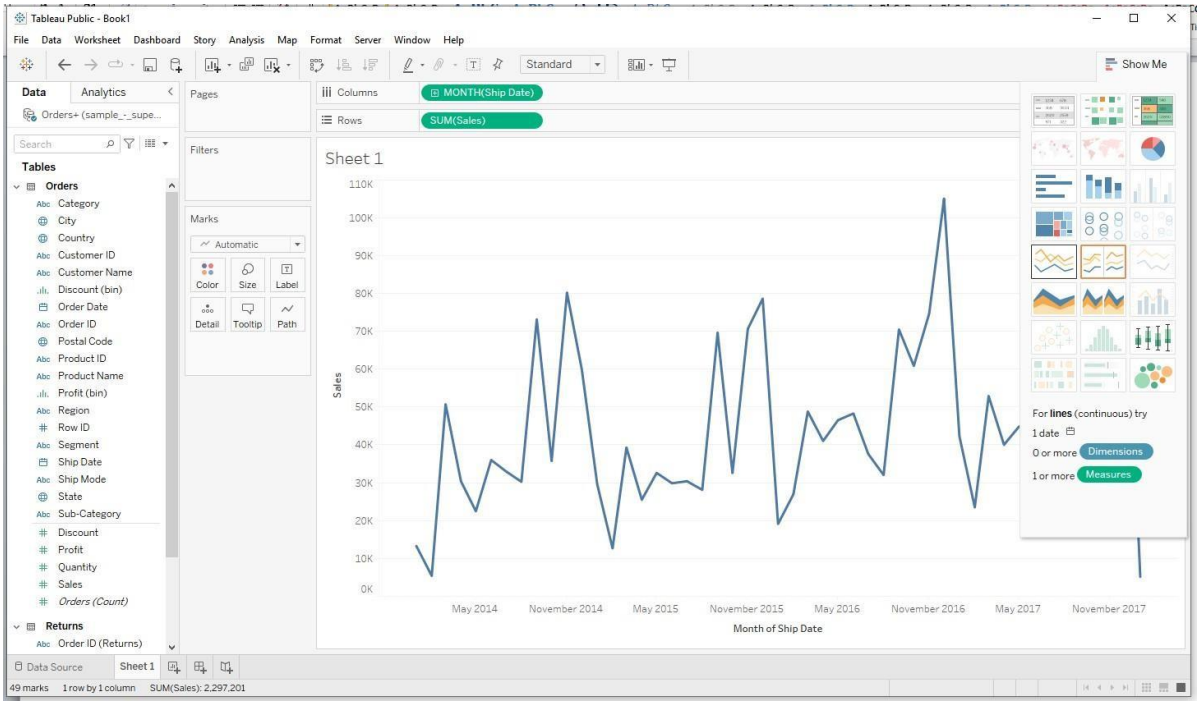


Side-by-side bar chart can be created in following way.



Line graph: Line graph can be continuous or discrete.

Continuous line graph is shown below:



Discrete line graph is shown below:

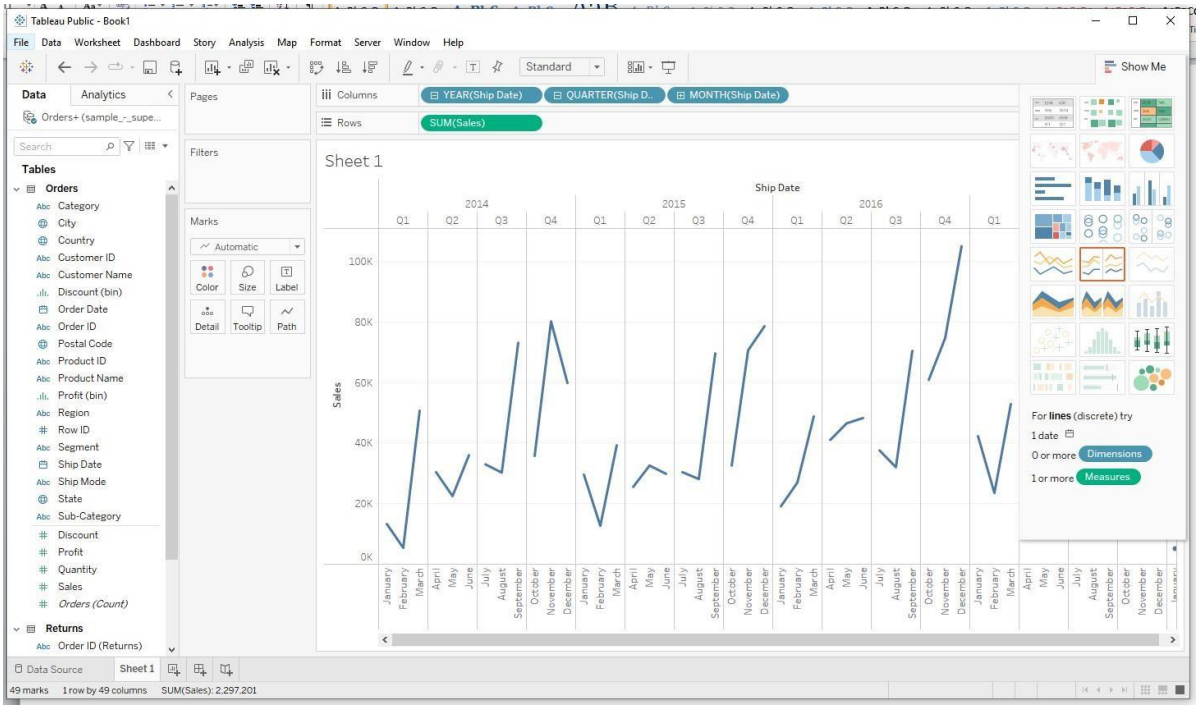


Tableau Public - Book1
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Standard
Show Me
Data Analytics
Orders+ (sample_-_supe...
Search
Tables
Orders
Category
City
Country
Customer ID
Customer Name
Discount (bin)
Order Date
Order ID
Postal Code
Product ID
Product Name
Profit (bin)
Region
Row ID
Segment
Ship Date
Ship Mode
State
Sub-Category
Discount
Profit
Quantity
Sales
Orders (Count)
Returns
Order ID (Returns)
Pages
Filters
Marks
Automatic
Color
Size
Label
Detail
Tooltip
Path
Columns
YEAR(Ship Date)
QUARTER(Ship D..
MONTH(Ship Date)
Rows
SUM(Sales)
Sheet 1
Ship Date
2014
2015
2016
Q1
Q2
Q3
Q4
Sales
100K
80K
60K
40K
20K
0K
January
February
March
April
May
June
July
August
September
October
November
December
For lines (discrete) try
1 date
0 or more
Dimensions
1 or more
Measures
Data Source
Sheet 1
49 marks
1 row by 49 columns
SUM(Sales): 2,297,201

Pie chart:

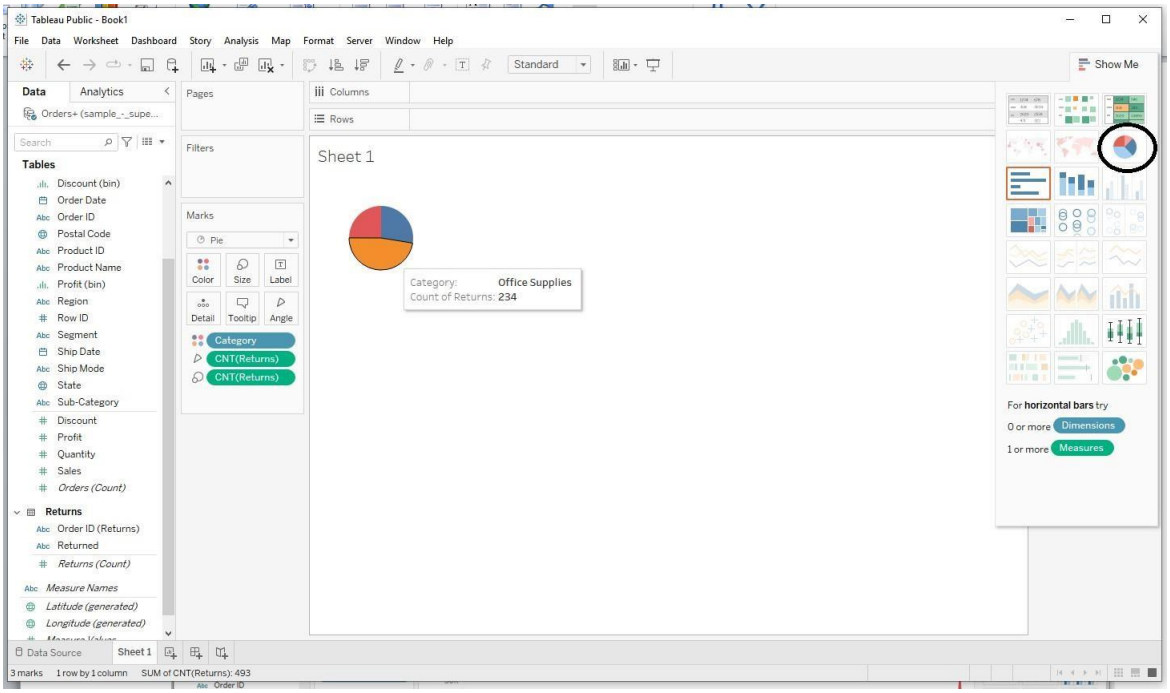
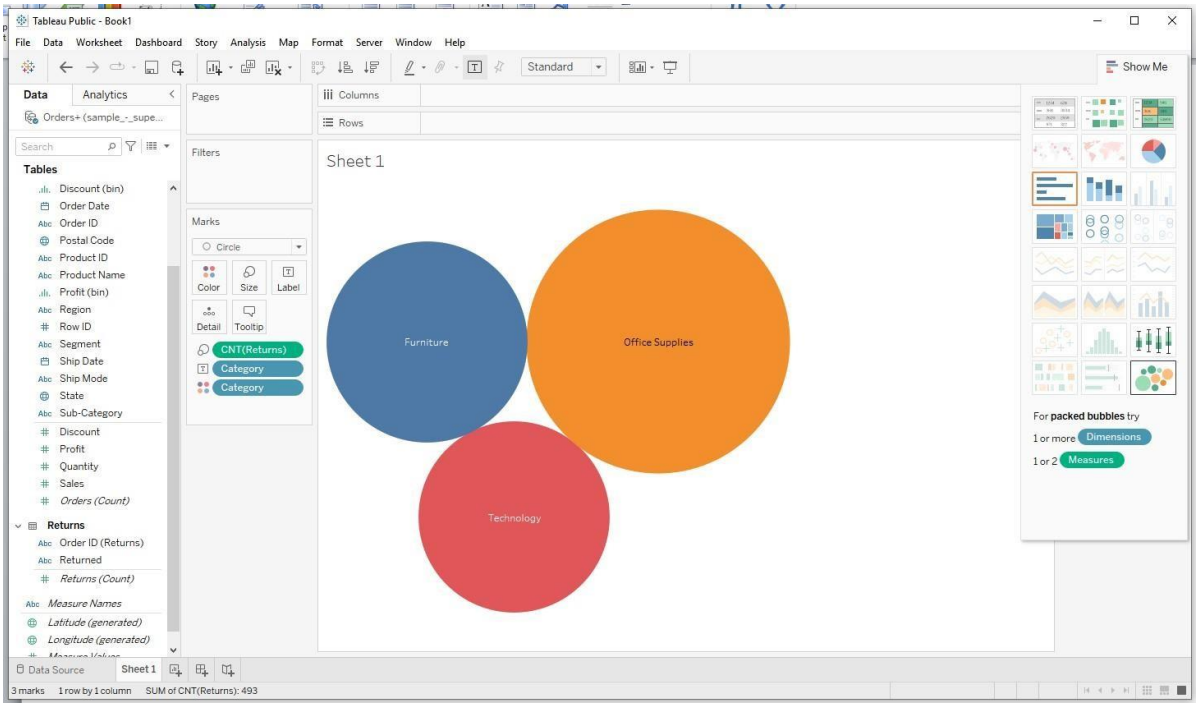
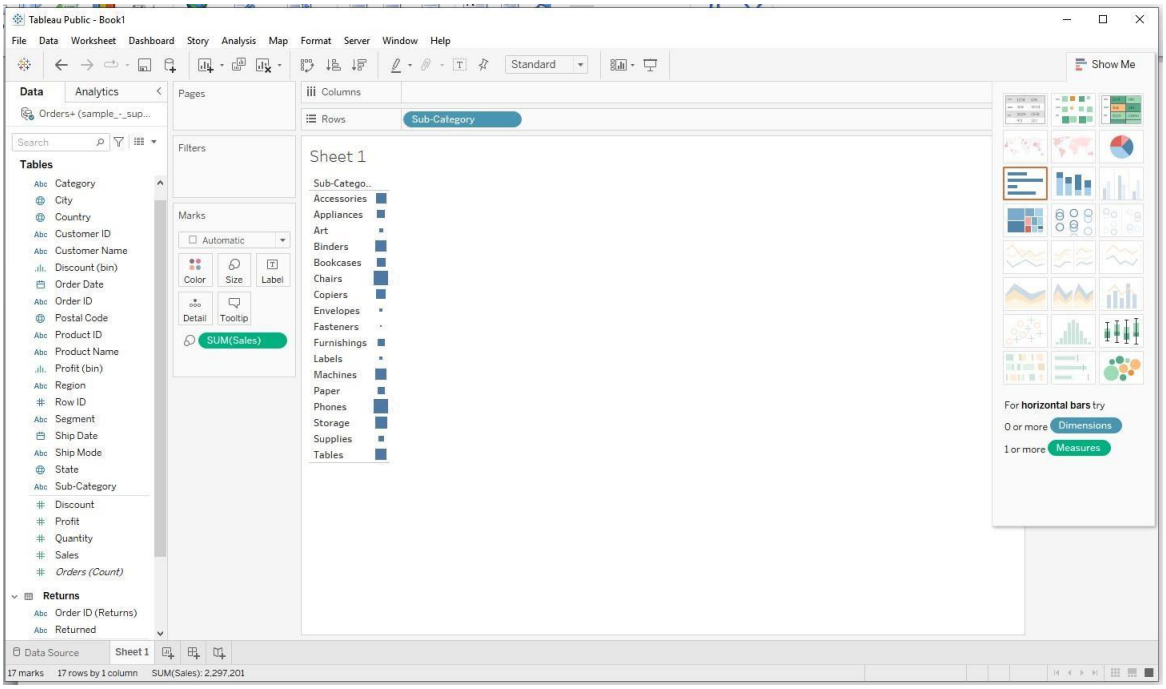


Tableau Public - Book1
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Standard
Show Me
Data Analytics
Orders+ (sample_-_supe...
Search
Tables
Discount (bin)
Order Date
Order ID
Postal Code
Product ID
Product Name
Profit (bin)
Region
Row ID
Segment
Ship Date
Ship Mode
State
Sub-Category
Discount
Profit
Quantity
Sales
Orders (Count)
Returns
Order ID (Returns)
Returned
Returns (Count)
Measure Names
Latitude (generated)
Longitude (generated)
Pages
Filters
Marks
Pie
Color
Size
Label
Detail
Tooltip
Angle
Category
CNT(Returns)
CNT(Returns)
Columns
Rows
Sheet 1
Category: Office Supplies
Count of Returns: 234
For horizontal bars try
0 or more
Dimensions
1 or more
Measures
Data Source
Sheet 1
3 marks
1 row by 1 column
SUM of CNT(Returns): 493

Bubble chart:



Heat map:

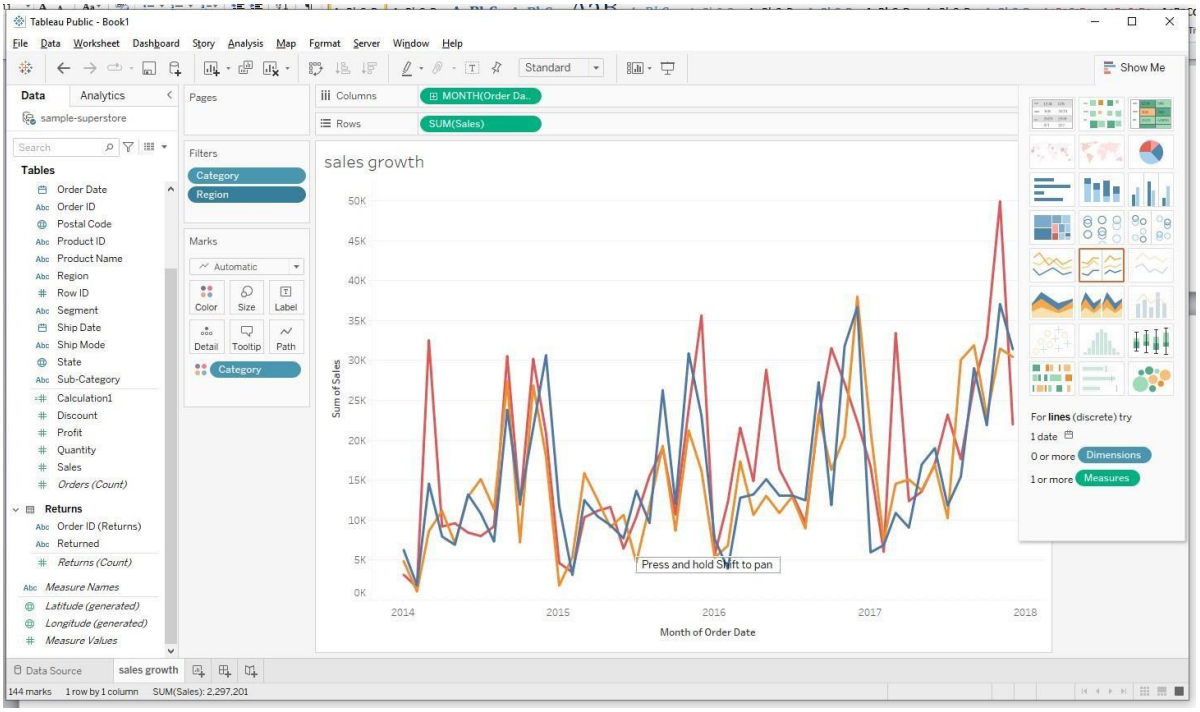


Week 8: Dashboards

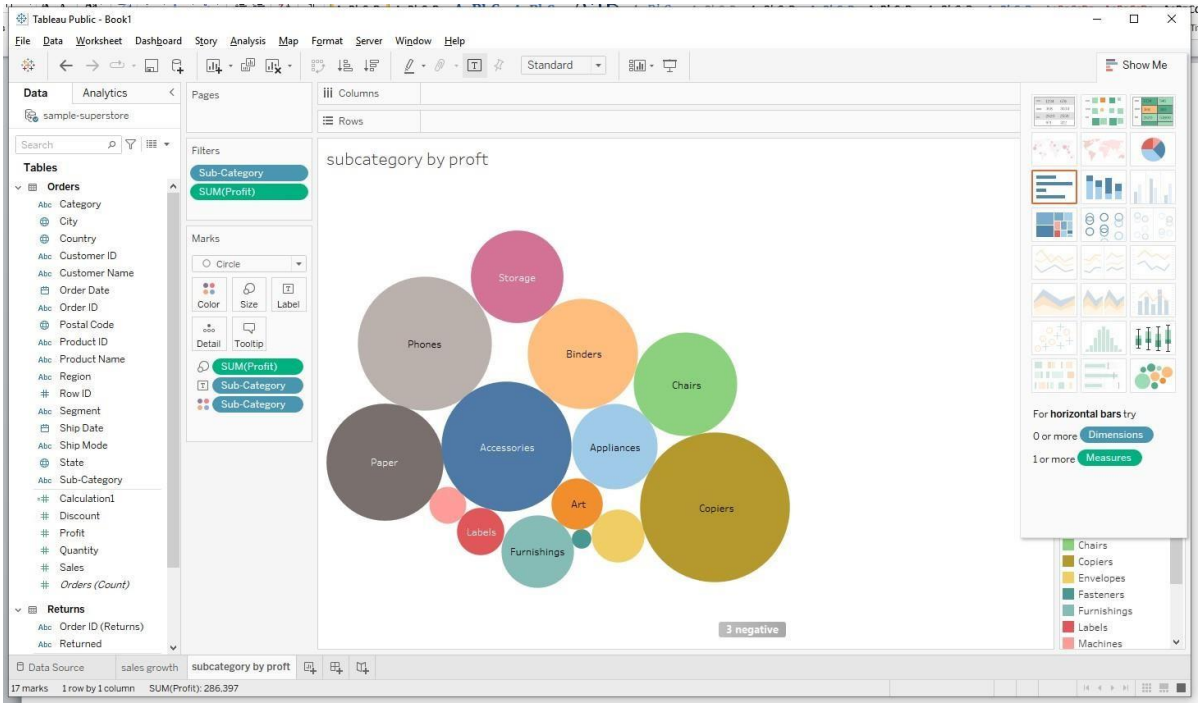
A dashboard is a way of displaying various types of visual data in one place. Usually, a dashboard is intended to convey different, but related information in an easy-to-digest form. And oftentimes, this includes things like key performance indicators (KPI)s or other important business metrics that stakeholders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

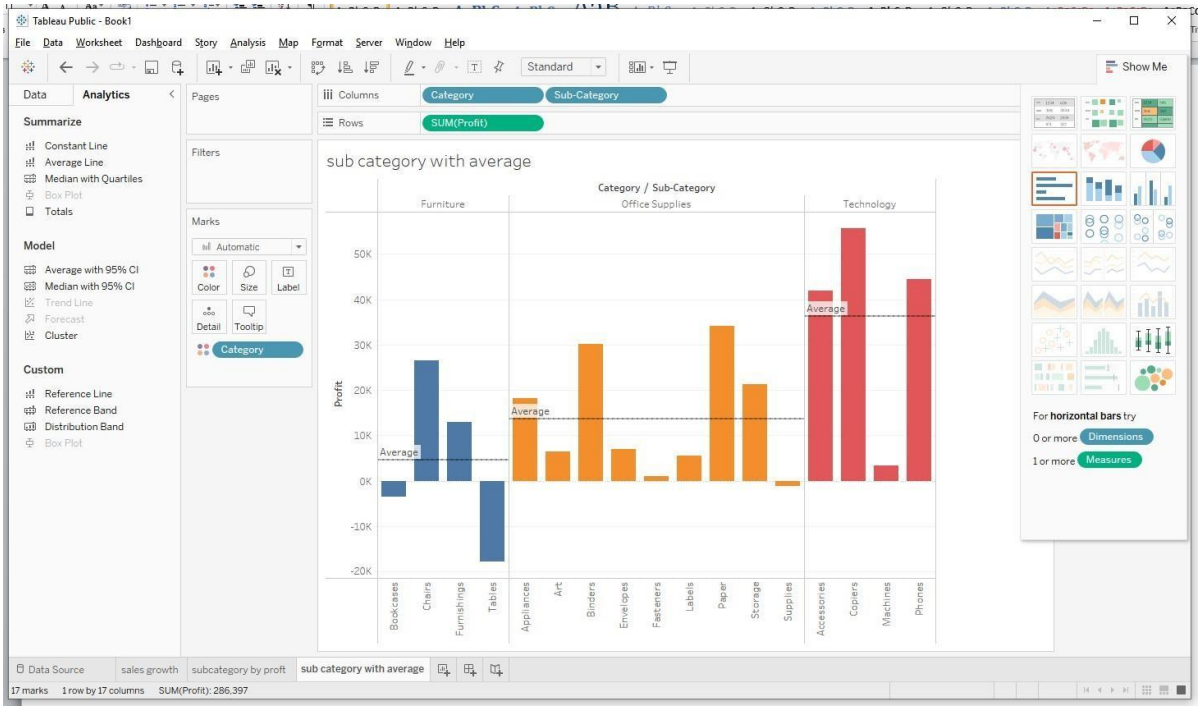
For example, category of sales across months in a year, region is the field added. The first view is shown below. This can be renamed at the bottom of the screen.



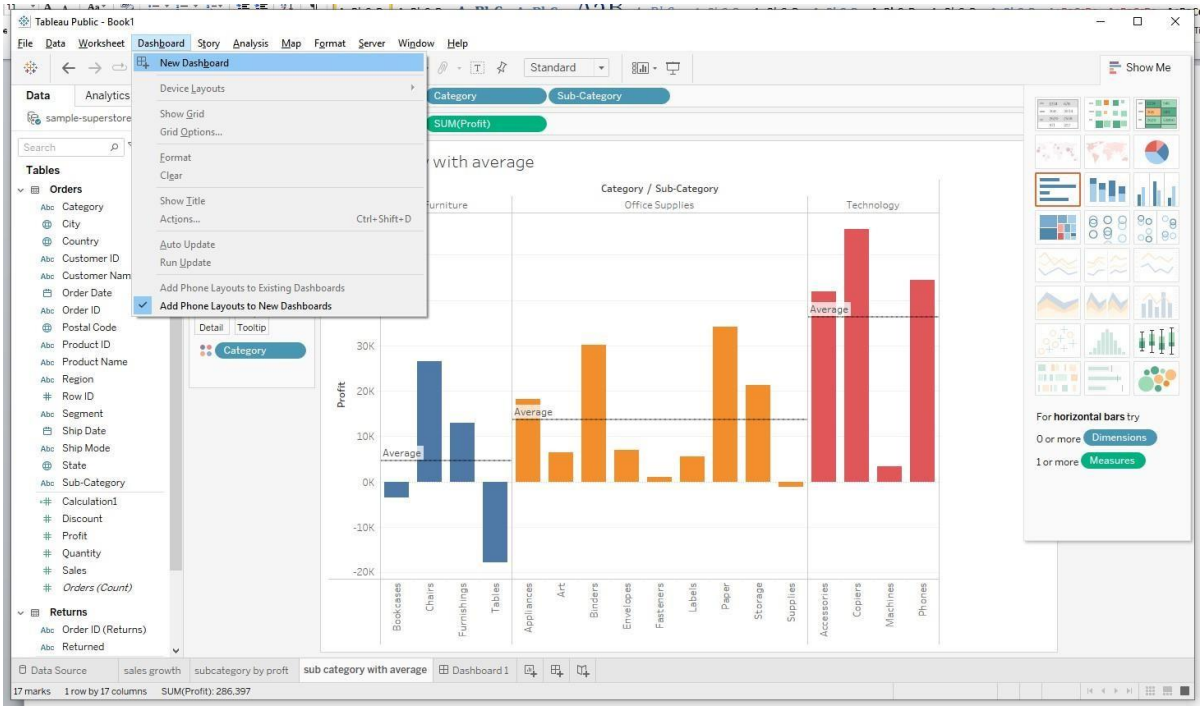
Now go to 2nd sheet for creating the 2nd view. The second view is shown below. A bubble chart was drawn between profit and subcategory. Then rename the sheet.



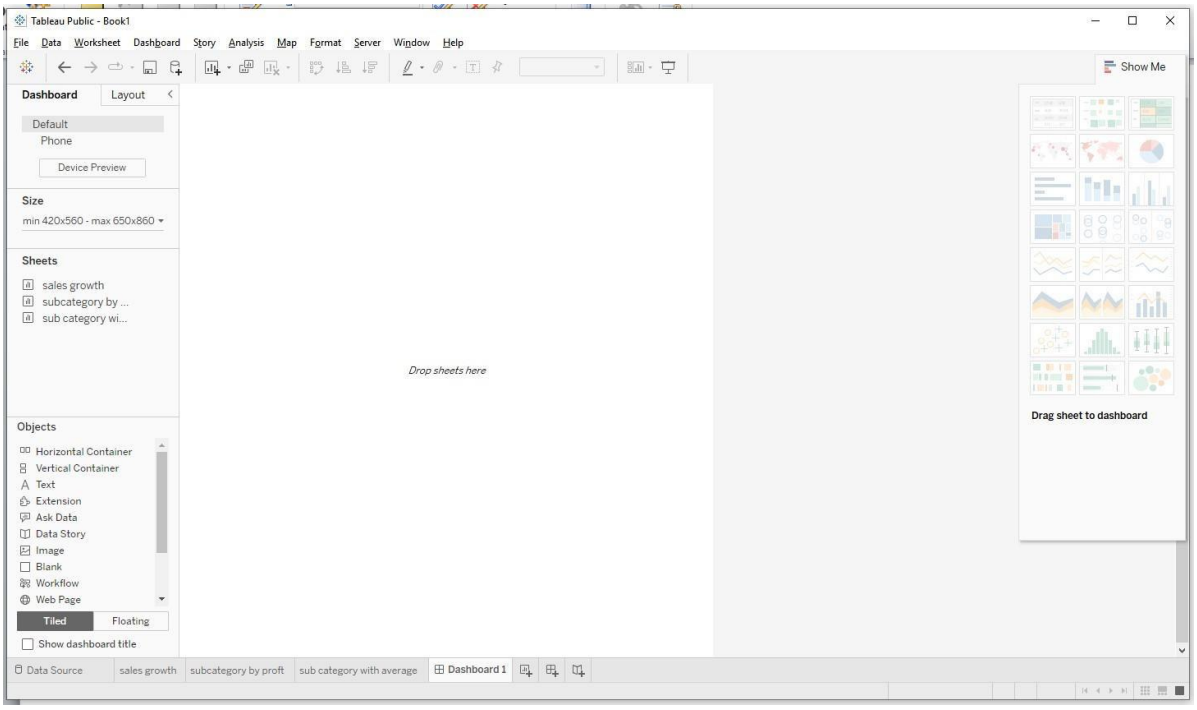
Next 3rd view is created as follows for profit for each subcategory in the category with averages.



After creating individual views, now a Dashboard 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 dahsboard, title can be given to the dashboard from Dashboard tab. Dahsboard can be customized in terms of its appearance by the user if requied. Dashboard once created can be saved on users system and can be retrieved whenever required.

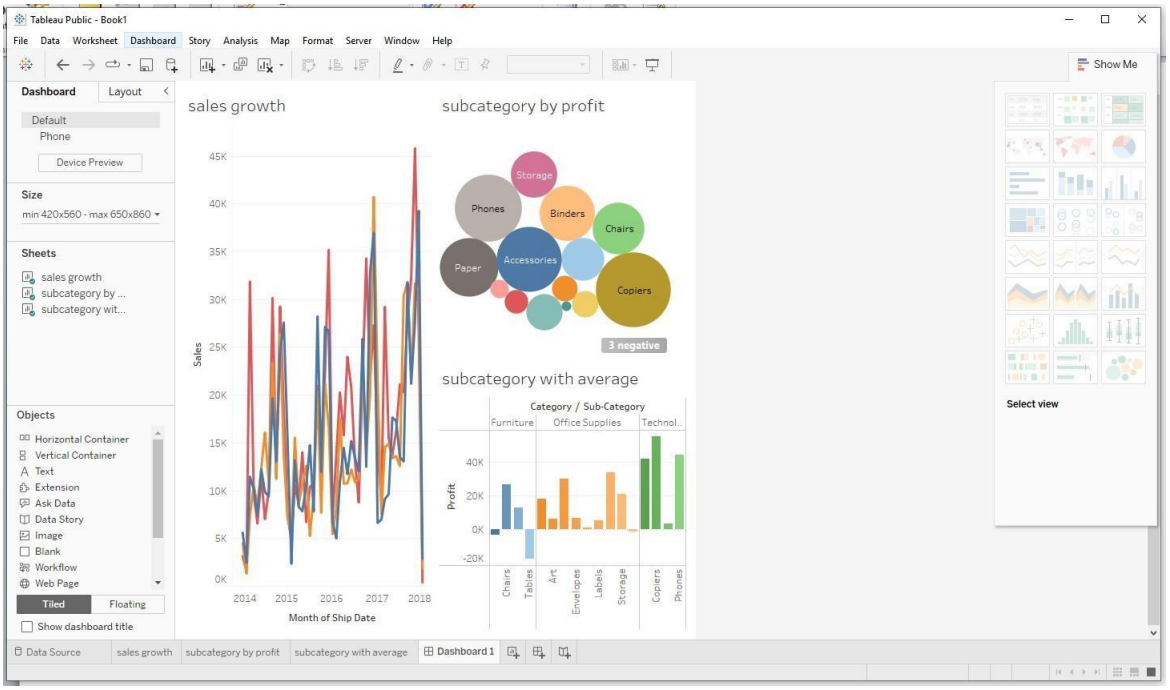


Tableau Public - Book1
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Dashboard Layout
Default
Phone
Device Preview
Size
min 420x560 - max 650x860
Sheets
sales growth
subcategory by ...
subcategory wit...
Objects
Horizontal Container
Vertical Container
Text
Extension
Ask Data
Data Story
Image
Blank
Workflow
Web Page
Tiled Floating
Show dashboard title
sales growth
Sales
Month of Ship Date
subcategory by profit
Storage
Phones
Binders
Chairs
Paper
Accessories
Copiers
3 negative
subcategory with average
Category / Sub-Category
Furniture
Office Supplies
Profit
Chairs
Tables
Art
Envelopes
Labels
Storage
Copiers
Phones
Show Me
Select view
Data Source
sales growth
subcategory by profit
subcategory with average
Dashboard 1