PRACTICAL REPORT ON
PPSCSMAJE305 : DATA VISUALIZATION

SUBMITTED BY

SIDDHESH SANTOSH TELI
ROLL NO: 18

SUBMITTED TO
Mrs. KARISHMA JAIN

MSc. (COMPUTER SCIENCE) SEM - III
2025 – 2026

CONDUCTED AT

CHIKITSAK SAMUHA'S

S. S. & L.S. PATKAR COLLEGE OF ARTS & SCIENCE AND

V. P. VARDE COLLEGE OF COMMERCE & ECONOMICS
An Autonomous college,
Affiliated to University of Mumbai
GOREGAON (W). MUMBAI - 400062

CHIKITSAK SAMUHA'S

SIR SITARAM & LADY SHANTABAI PATKAR COLLEGE OF ARTS & SCIENCE
&
V.P. VARDE COLLEGE OF COMMERCE & ECONOMICS

GOREGAON (WEST), MUMBAI - 400 104.

An Autonomous College, University of Mumbai

CERTIFICATE

Certified that such of the experiments as have been duly signed

were performed by Mr./Miss ______________________________

Roll No. ______________ *of* ______________ *class* __________

Division ______________ *in the* ______________ *Laboratory*

of this college during the year ____________________

Professor-in-Charge　　　　Examiner　　　　Co-ordinator

Date: ______________　　　　______________ *Department*

Expt No.	Date	Experiment	Signature
1	8/07/25	Create one-dimensional data using series and perform various operations on it.	
2	15/07/25	Create Two-dimensional data with the help of data frames and perform different operations on it.	
3	22/07/25	Write a code to read data from the different file formats like JSON, HTML, XML, and CSV files and check for missing data and outlier values and handle them. Note – for reading xml we need xml library (pip install lxml)	
4	02/08/25	Perform Reshaping of the hierarchical data and pivoting data frame data.	
5	05/08/25	Connecting and extracting with various data resources in tableau.	
6	23/08/25	Performing calculations and creating parameters in Tableau.	
7	26/08/25	Create Dashboard and Storytelling using tableau.	
8	02/09/25	Data Visualization Power BI—bar charts, line charts, pie charts, tables, matrices, simple cards	

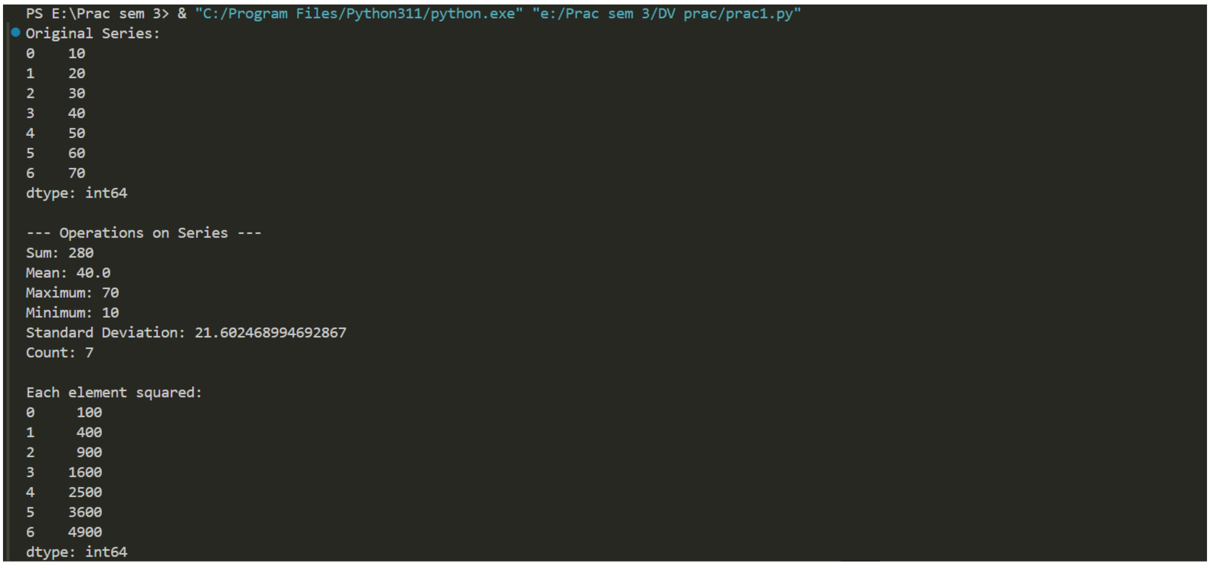
Practical No. 1

Aim: Create one-dimensional data using series and perform various operations on it.

Code:

```
import pandas as pd

import numpy as np


# Create a one-dimensional Series

data = pd.Series([10, 20, 30, 40, 50, 60, 70])

print("Original Series:")

print(data)

print("\n--- Operations on Series ---")

print("Sum:", data.sum())

print("Mean:", data.mean())

print("Maximum:", data.max())

print("Minimum:", data.min())

print("Standard Deviation:", data.std())

print("Count:", data.count())


# Element-wise operations

print("\nEach element squared:")

print(data ** 2)
```

Output:



```
PS E:\Prac sem 3> & "C:/Program Files/Python311/python.exe" "e:/Prac sem 3/DV prac/prac1.py"
Original Series:
0    10
1    20
2    30
3    40
4    50
5    60
6    70
dtype: int64

--- Operations on Series ---
Sum: 280
Mean: 40.0
Maximum: 70
Minimum: 10
Standard Deviation: 21.602468994692867
Count: 7

Each element squared:
0     100
1     400
2     900
3    1600
4    2500
5    3600
6    4900
dtype: int64
```

Practical No. 2

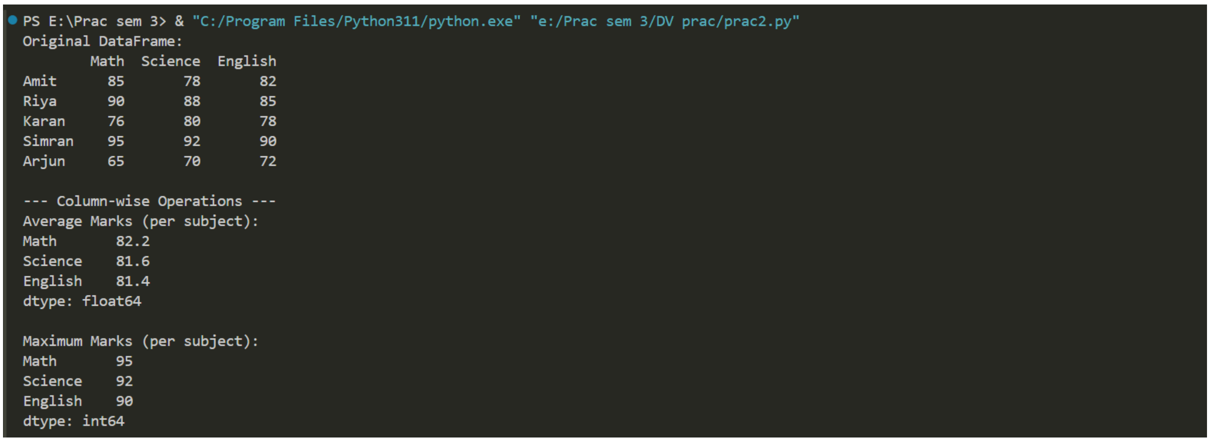
Aim: Create Two-dimensional data with the help of data frames and perform different operations on it.

Code:

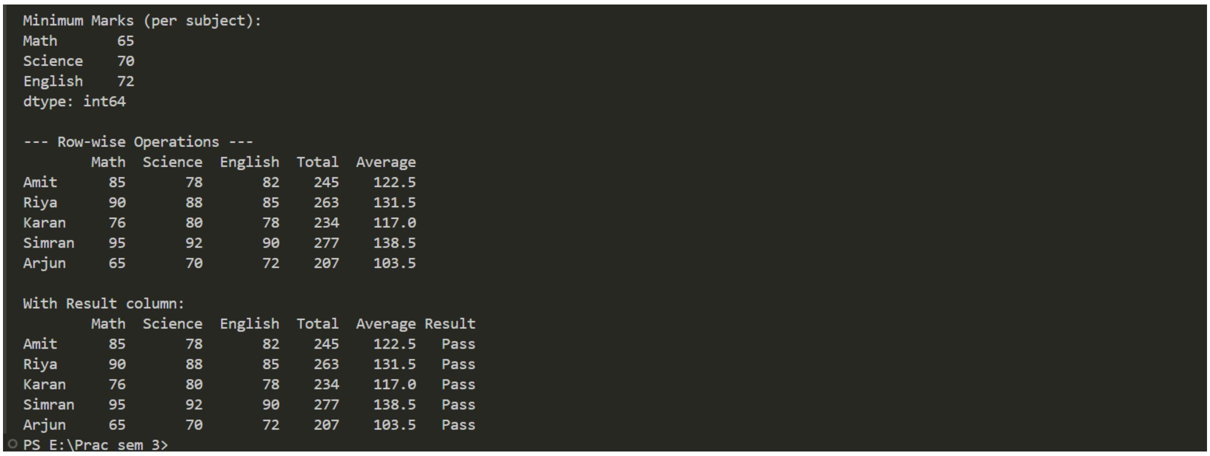
```
import pandas as pd


# Create a two-dimensional DataFrame
data = {
   'Math': [85, 90, 76, 95, 65],
   'Science': [78, 88, 80, 92, 70],
   'English': [82, 85, 78, 90, 72]
}
students = ['Amit', 'Riya', 'Karan', 'Simran', 'Arjun']


df = pd.DataFrame(data, index=students)
print("Original DataFrame:")
print(df)


# Column-wise operations
print("\n--- Column-wise Operations ---")
print("Average Marks (per subject):")
print(df.mean())


print("\nMaximum Marks (per subject):")
print(df.max())


print("\nMinimum Marks (per subject):")
print(df.min())
```

```
# Row-wise operations
print("\n--- Row-wise Operations ---")
df['Total'] = df.sum(axis=1)
df['Average'] = df.mean(axis=1)
print(df)

# Conditional operation
df['Result'] = df['Average'].apply(lambda x: "Pass" if x >= 50 else "Fail")
print("\nWith Result column:")
print(df)
```

Output:



```
PS E:\Prac sem 3> & "C:/Program Files/Python311/python.exe" "e:/Prac sem 3/DV prac/prac2.py"
Original DataFrame:
        Math  Science  English
Amit      85       78       82
Riya      90       88       85
Karan     76       80       78
Simran    95       92       90
Arjun     65       70       72

--- Column-wise Operations ---
Average Marks (per subject):
Math       82.2
Science    81.6
English    81.4
dtype: float64

Maximum Marks (per subject):
Math       95
Science    92
English    90
dtype: int64
```



```
Minimum Marks (per subject):
Math       65
Science    70
English    72
dtype: int64

--- Row-wise Operations ---
        Math  Science  English  Total  Average
Amit      85       78       82    245    122.5
Riya      90       88       85    263    131.5
Karan     76       80       78    234    117.0
Simran    95       92       90    277    138.5
Arjun     65       70       72    207    103.5

With Result column:
        Math  Science  English  Total  Average Result
Amit      85       78       82    245    122.5   Pass
Riya      90       88       85    263    131.5   Pass
Karan     76       80       78    234    117.0   Pass
Simran    95       92       90    277    138.5   Pass
Arjun     65       70       72    207    103.5   Pass
PS E:\Prac sem 3>
```

Practical No. 3

Aim: Write a code to read data from the different file formats like JSON, HTML, XML, and CSV files and check for missing data and outlier values and handle them.

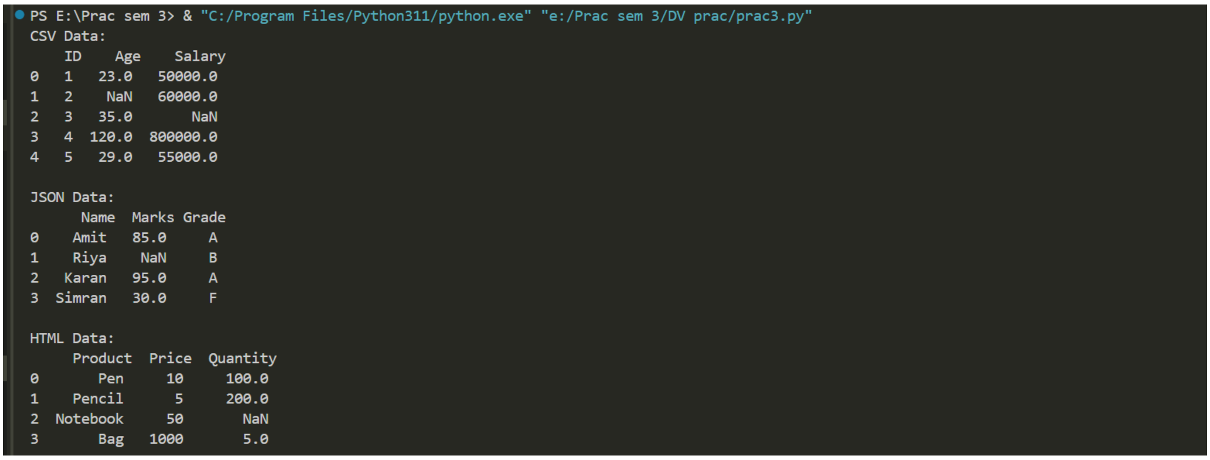
Note – for reading xml we need xml library (pip install lxml)

Code:

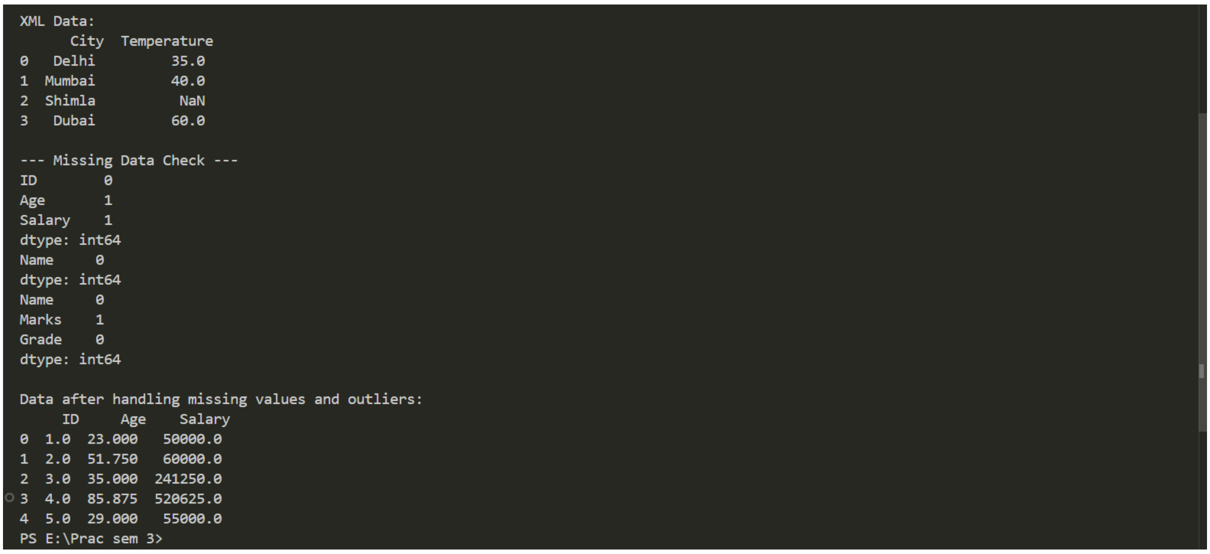
```
import pandas as pd

import numpy as np


# ------------------------

# 1. Reading Different File Formats

# ------------------------


# CSV file

csv_df = pd.read_csv("DV prac\data.csv")

print("CSV Data:\n", csv_df.head())


# JSON file

json_df = pd.read_json("DV prac\data.json")

print("\nJSON Data:\n", json_df.head())


# HTML file (reads tables from HTML page)

html_dfs = pd.read_html("DV prac\data.html")   # returns list of DataFrames

html_df = html_dfs[0]                  # take first table

print("\nHTML Data:\n", html_df.head())


# XML file

xml_df = pd.read_xml("DV prac\data.xml")
```

```
print("\nXML Data:\n", xml_df.head())

# -----------------------
# 2. Check for Missing Data
# -----------------------
print("\n--- Missing Data Check ---")
print(csv_df.isnull().sum())   # number of missing values in CSV
print(json_df.isnull().sum())  # number of missing values in JSON

# Handle missing values
csv_df.fillna(csv_df.mean(numeric_only=True), inplace=True)  # replace with mean
json_df.dropna(inplace=True)  # drop rows with missing values

# -----------------------
# 3. Detect Outliers
# -----------------------
def detect_outliers(df, col):
    """Detect outliers using IQR method"""
    Q1 = df[col].quantile(0.25)
    Q3 = df[col].quantile(0.75)
    IQR = Q3 - Q1
    lower_bound = Q1 - 1.5 * IQR
    upper_bound = Q3 + 1.5 * IQR
    outliers = df[(df[col] < lower_bound) | (df[col] > upper_bound)]
    return outliers

# -----------------------
# 4. Handle Outliers
# ------------------------
# Example: Cap outliers within bounds
for col in csv_df.select_dtypes(include=np.number).columns:
    Q1 = csv_df[col].quantile(0.25)
    Q3 = csv_df[col].quantile(0.75)
    IQR = Q3 - Q1
    lower_bound = Q1 - 1.5 * IQR
    upper_bound = Q3 + 1.5 * IQR
    csv_df[col] = np.where(csv_df[col] < lower_bound, lower_bound, csv_df[col])
    csv_df[col] = np.where(csv_df[col] > upper_bound, upper_bound, csv_df[col])


print("\nData after handling missing values and outliers:\n", csv_df.head())
```

Output:



```
PS E:\Prac sem 3> & "C:/Program Files/Python311/python.exe" "e:/Prac sem 3/DV prac/prac3.py"
CSV Data:
    ID    Age    Salary
0   1   23.0   50000.0
1   2    NaN   60000.0
2   3   35.0       NaN
3   4  120.0  800000.0
4   5   29.0   55000.0

JSON Data:
      Name  Marks Grade
0    Amit   85.0     A
1    Riya    NaN     B
2   Karan   95.0     A
3  Simran   30.0     F

HTML Data:
     Product  Price  Quantity
0       Pen     10     100.0
1    Pencil      5     200.0
2  Notebook     50       NaN
3       Bag   1000       5.0
```



```
XML Data:
      City  Temperature
0   Delhi         35.0
1  Mumbai         40.0
2  Shimla          NaN
3   Dubai         60.0

--- Missing Data Check ---
ID        0
Age       1
Salary    1
dtype: int64
Name     0
dtype: int64
Name     0
Marks    1
Grade    0
dtype: int64

Data after handling missing values and outliers:
     ID     Age    Salary
0  1.0  23.000   50000.0
1  2.0  51.750   60000.0
2  3.0  35.000  241250.0
3  4.0  85.875  520625.0
4  5.0  29.000   55000.0
PS E:\Prac sem 3>
```

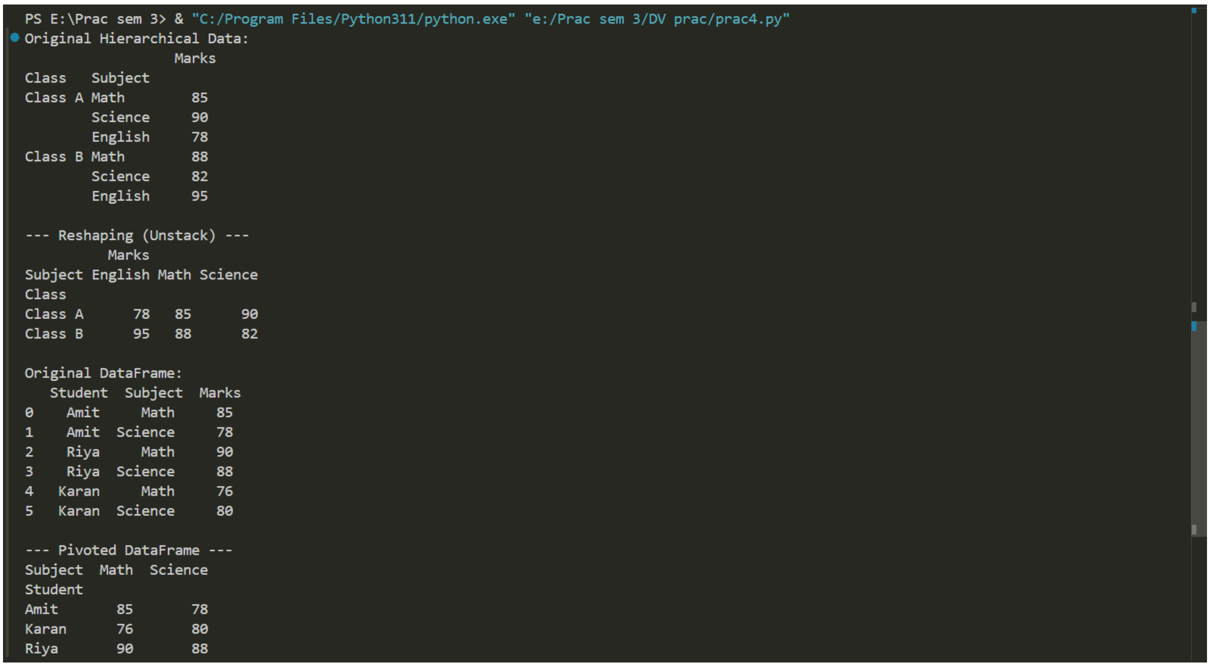
Practical No. 4

Aim: Perform Reshaping of the hierarchical data and pivoting data frame data.

Code:

```
import pandas as pd

# ----------------------------
# 1. Hierarchical Data (MultiIndex)
# ----------------------------

# Create sample hierarchical data
arrays = [
    ["Class A", "Class A", "Class A", "Class B", "Class B", "Class B"],
    ["Math", "Science", "English", "Math", "Science", "English"]
]
index = pd.MultiIndex.from_arrays(arrays, names=("Class", "Subject"))
data = [85, 90, 78, 88, 82, 95]

df_hier = pd.DataFrame({"Marks": data}, index=index)
print("Original Hierarchical Data:\n", df_hier)

# Reshape using unstack
print("\n--- Reshaping (Unstack) ---")
reshaped = df_hier.unstack(level="Subject")
print(reshaped)

# ----------------------------
# 2. Pivoting DataFrame
# ----------------------------


# Create a sample DataFrame
data = {
    "Student": ["Amit", "Amit", "Riya", "Riya", "Karan", "Karan"],
    "Subject": ["Math", "Science", "Math", "Science", "Math", "Science"],
    "Marks": [85, 78, 90, 88, 76, 80]
}
df = pd.DataFrame(data)
print("\nOriginal DataFrame:\n", df)

# Pivot the data: Subjects as columns
pivoted = df.pivot(index="Student", columns="Subject", values="Marks")
print("\n--- Pivoted DataFrame ---")
print(pivoted)
```

Output:



```
PS E:\Prac sem 3> & "C:/Program Files/Python311/python.exe" "e:/Prac sem 3/DV prac/prac4.py"
Original Hierarchical Data:
                  Marks
Class   Subject
Class A Math        85
        Science     90
        English     78
Class B Math        88
        Science     82
        English     95

--- Reshaping (Unstack) ---
          Marks
Subject English Math Science
Class
Class A      78   85      90
Class B      95   88      82

Original DataFrame:
   Student  Subject  Marks
0    Amit     Math     85
1    Amit  Science     78
2    Riya     Math     90
3    Riya  Science     88
4   Karan     Math     76
5   Karan  Science     80

--- Pivoted DataFrame ---
Subject  Math  Science
Student
Amit       85       78
Karan      76       80
Riya       90       88
```

Name – Siddhesh Santosh Teli Roll No. 18
MSc CS

Practical No.5

Aim: Connecting and extracting with various data resources in tableau.

Background Information:

Data Sources Tableau can extract data from all the popular data sources.

These include:

1) File System The simplest data source you can use with Tableau is a file. These could be files like an Excel spreadsheet, a CSV file or a text file.

2) Cloud System You can also source data from popular cloud sources. Some of the options are:

- Google Analytics

- Google BigQuery

- Windows Azure

- Amazon Redshift

3) Relational systems You can connect to many types of relational databases such as SQL Server, Oracle, and DB2.

4) Live Data Sources Connect live is a feature of Tableau that allows you to connect real-time data. Tableau does this by constantly reading the data, so your visualizations are constantly up to date.

Data Extraction Techniques

- Once you've decided on your data sources, the next step is to extract the data you need from those sources.

- Whether you are connecting to a live database or storing your data in memory, you may well want to cut it down to only what you need for your application. This will mean you'll have less data to extract from a live source or a smaller amount of data to store in memory.

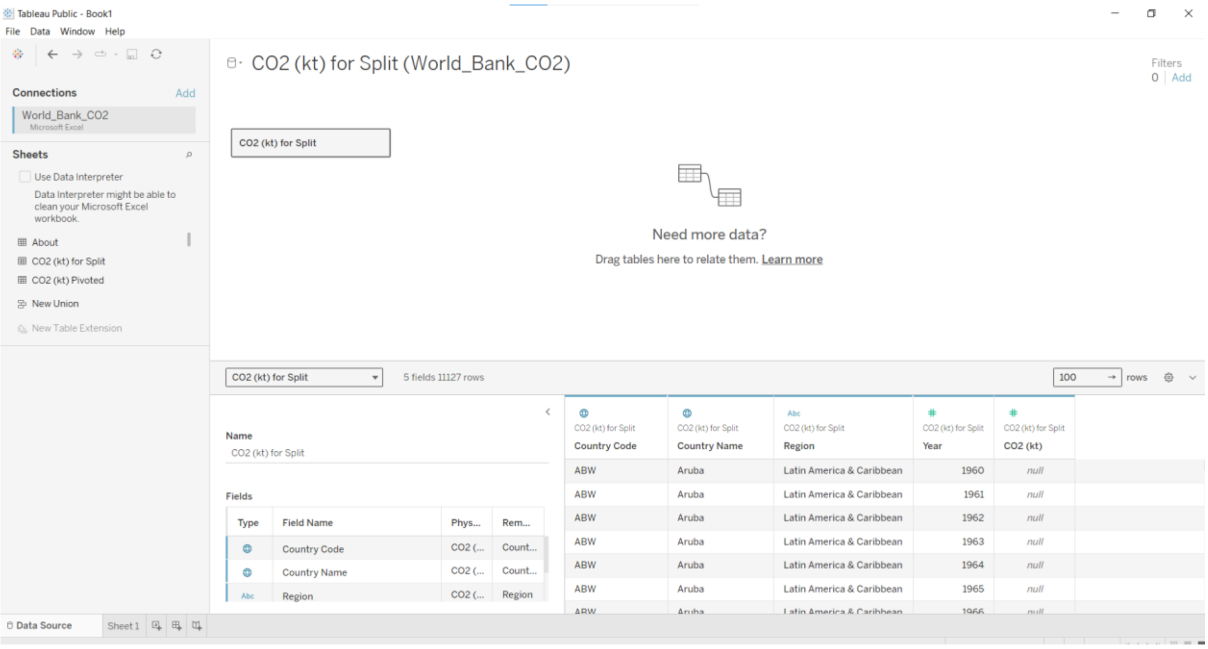
- It also converts the data to a form that works well with the Tableau engine, meaning things will speed up even more.

- With Tableau, this is done with data extracts.

- A data extract is simply a subset of a total data source. When extracting data, you can choose exactly what you want and how much of underlying data to extract using extract data dialog box.

- To create a new Tableau data extract, go to Data -> Extract Data. You'll be presented with many options to limit the number of rows and aggregate for dimensions. Here is where you can use filters to cut down your data to just the things you need.

Excel File Data Source



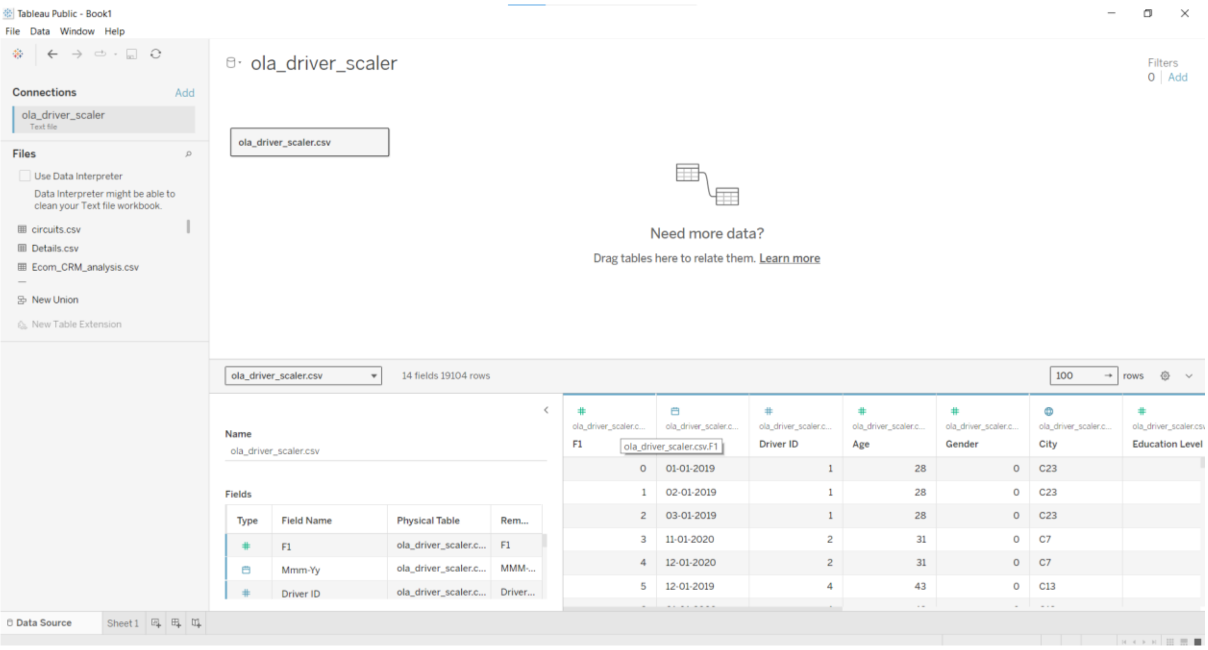
Name
CO2 (kt) for Split

Fields

Type	Field Name	Phys...	Rem...
🌐	Country Code	CO2 (...	Count...
🌐	Country Name	CO2 (...	Count...
Abc	Region	CO2 (...	Region

CO2 (kt) for Split Country Code	CO2 (kt) for Split Country Name	CO2 (kt) for Split Region	CO2 (kt) for Split Year	CO2 (kt) for Split CO2 (kt)
ABW	Aruba	Latin America & Caribbean	1960	null
ABW	Aruba	Latin America & Caribbean	1961	null
ABW	Aruba	Latin America & Caribbean	1962	null
ABW	Aruba	Latin America & Caribbean	1963	null
ABW	Aruba	Latin America & Caribbean	1964	null
ABW	Aruba	Latin America & Caribbean	1965	null
ABW	Aruba	Latin America & Caribbean	1966	null

Text File Data Source



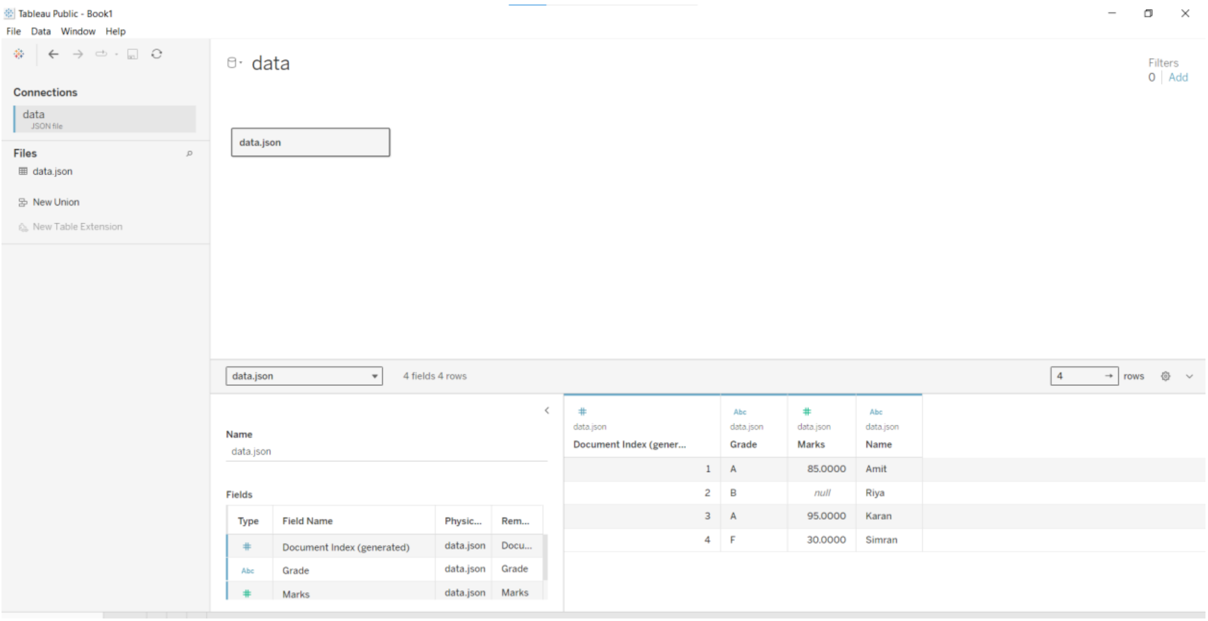
Name
ola_driver_scaler.csv

Fields

Type	Field Name	Physical Table	Rem...
#	F1	ola_driver_scaler.c...	F1
📅	Mmm-Yy	ola_driver_scaler.c...	MMM-...
#	Driver ID	ola_driver_scaler.c...	Driver...

ola_driver_scaler.c... F1	ola_driver_scaler.c... ola_driver_scaler.csv.F1	ola_driver_scaler.c... Driver ID	ola_driver_scaler.c... Age	ola_driver_scaler.c... Gender	ola_driver_scaler.c... City	ola_driver_scaler.csv Education Level
0	01-01-2019	1	28	0	C23	
1	02-01-2019	1	28	0	C23	
2	03-01-2019	1	28	0	C23	
3	11-01-2020	2	31	0	C7	
4	12-01-2020	2	31	0	C7	
5	12-01-2019	4	43	0	C13	

JSON File Data Source



Name
data.json

Fields

Type	Field Name	Physic...	Rem...
#	Document Index (generated)	data.json	Docu...
Abc	Grade	data.json	Grade
#	Marks	data.json	Marks

# data.json Document Index (gener...	Abc data.json Grade	# data.json Marks	Abc data.json Name
1	A	85.0000	Amit
2	B	*null*	Riya
3	A	95.0000	Karan
4	F	30.0000	Simran

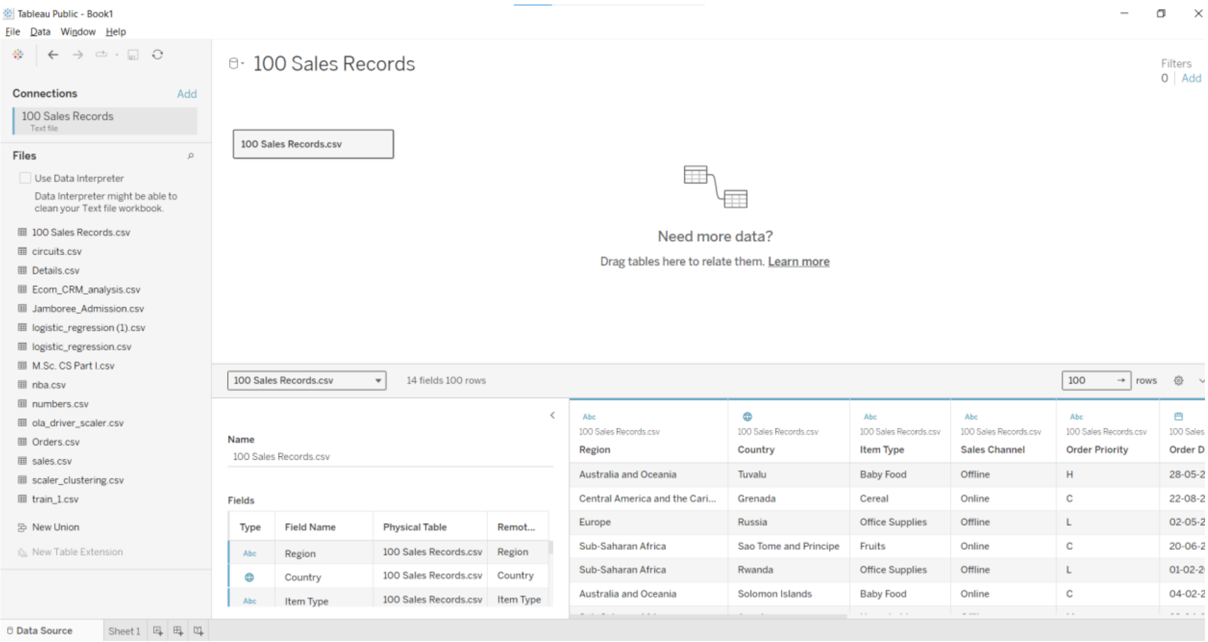
Name – Siddhesh Santosh Teli
MSc CS

Roll No. 18

Practical No. 6

Aim: Performing calculations and creating parameters in Tableau.

Step 1: Load the csv file in Tableau



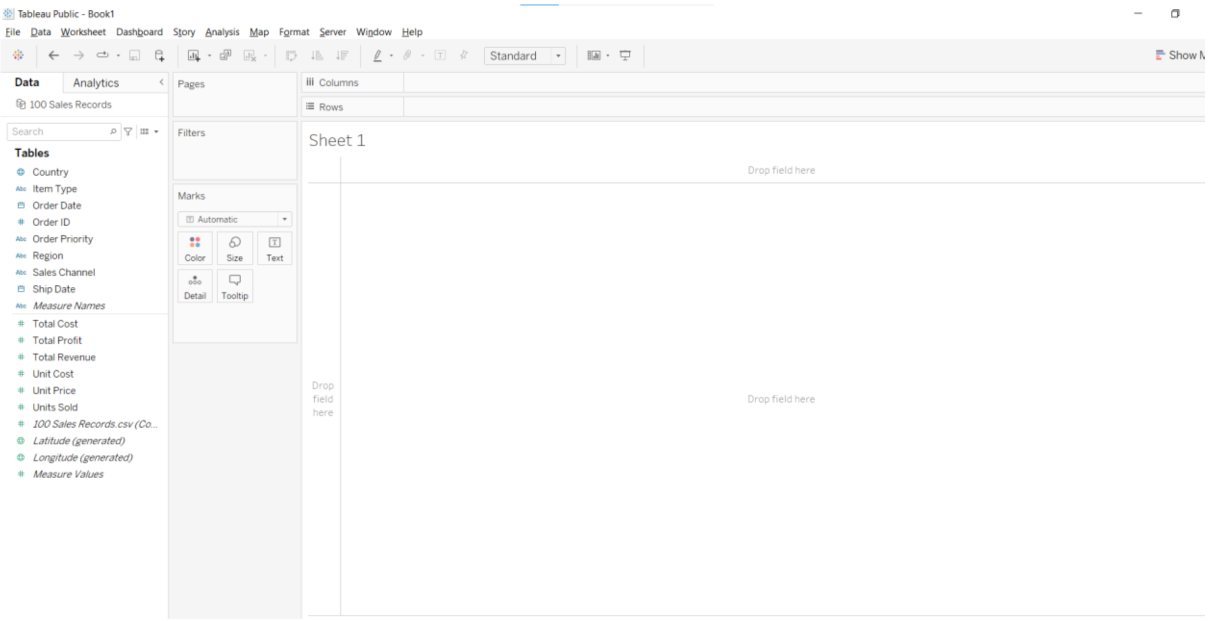
Name
100 Sales Records.csv

Fields

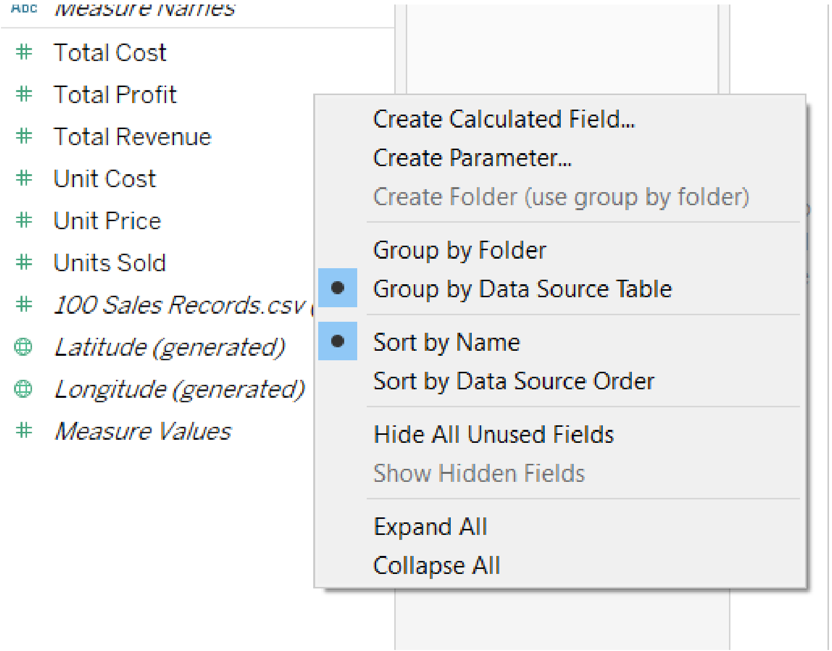
Type	Field Name	Physical Table	Remot...
Abc	Region	100 Sales Records.csv	Region
🌐	Country	100 Sales Records.csv	Country
Abc	Item Type	100 Sales Records.csv	Item Type

Abc 100 Sales Records.csv Region	🌐 100 Sales Records.csv Country	Abc 100 Sales Records.csv Item Type	Abc 100 Sales Records.csv Sales Channel	Abc 100 Sales Records.csv Order Priority	100 Sales Order D
Australia and Oceania	Tuvalu	Baby Food	Offline	H	28-05-2
Central America and the Cari...	Grenada	Cereal	Online	C	22-08-2
Europe	Russia	Office Supplies	Offline	L	02-05-2
Sub-Saharan Africa	Sao Tome and Principe	Fruits	Online	C	20-06-2
Sub-Saharan Africa	Rwanda	Office Supplies	Offline	L	01-02-2
Australia and Oceania	Solomon Islands	Baby Food	Online	C	04-02-2

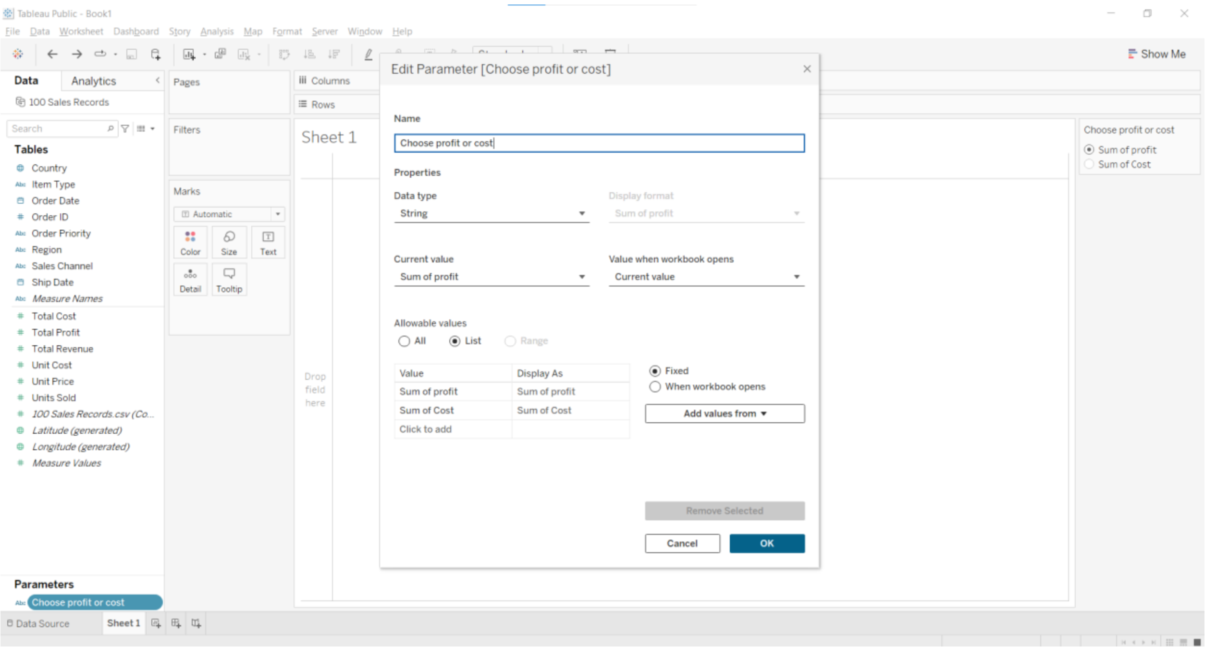
Step 2: Click on sheet1



Step 3: Right Click on left panel and select option create parameter



Step 4: Create a parameter with 2 values as option one as profit and other as cost



Edit Parameter [Choose profit or cost]

Name: Choose profit or cost

Properties

Data type: String

Display format: Sum of profit

Current value: Sum of profit

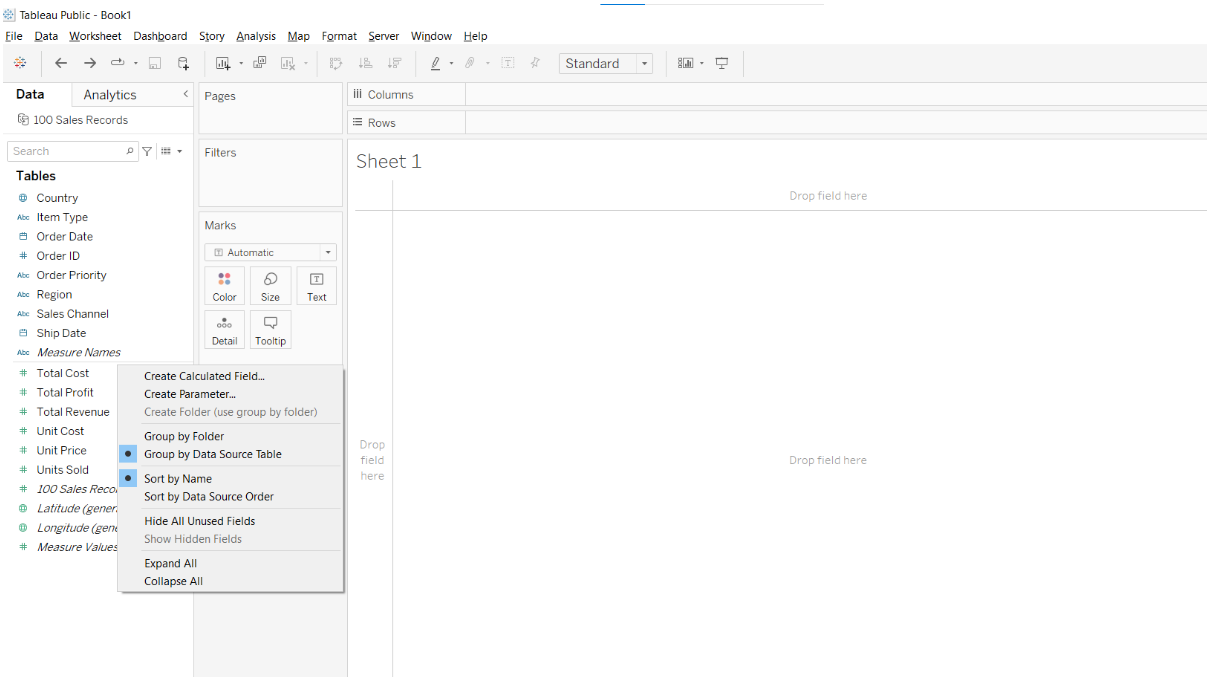
Value when workbook opens: Current value

Allowable values: ○ All ● List ○ Range

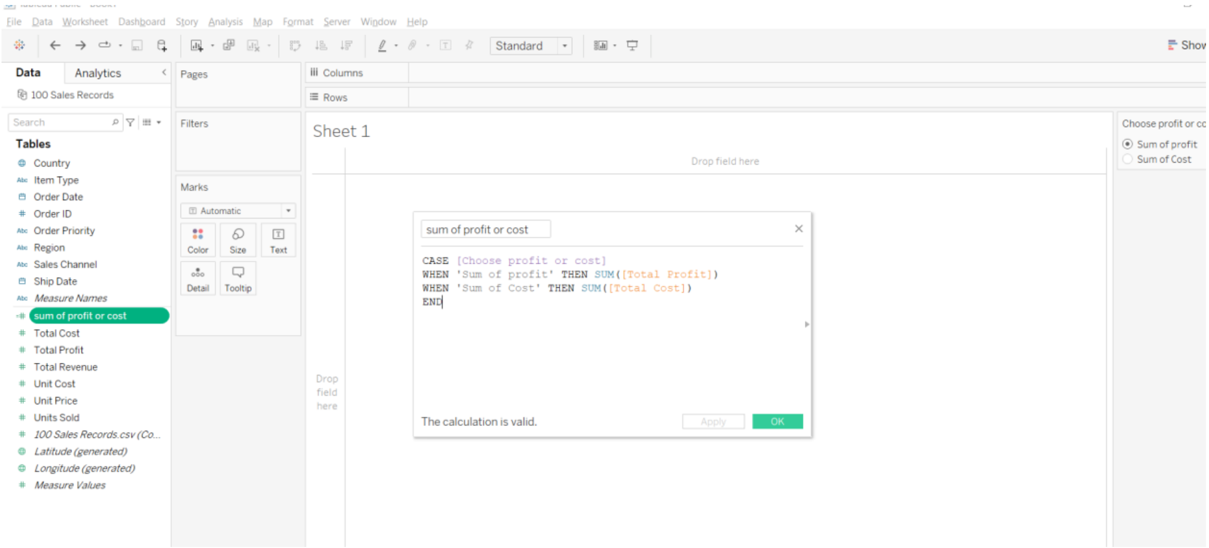
Value	Display As
Sum of profit	Sum of profit
Sum of Cost	Sum of Cost
Click to add	

● Fixed
○ When workbook opens

Step 5: Right click on the left pane and select option create calculation field



Step 6: Create case when statement and add parameter in calculation field



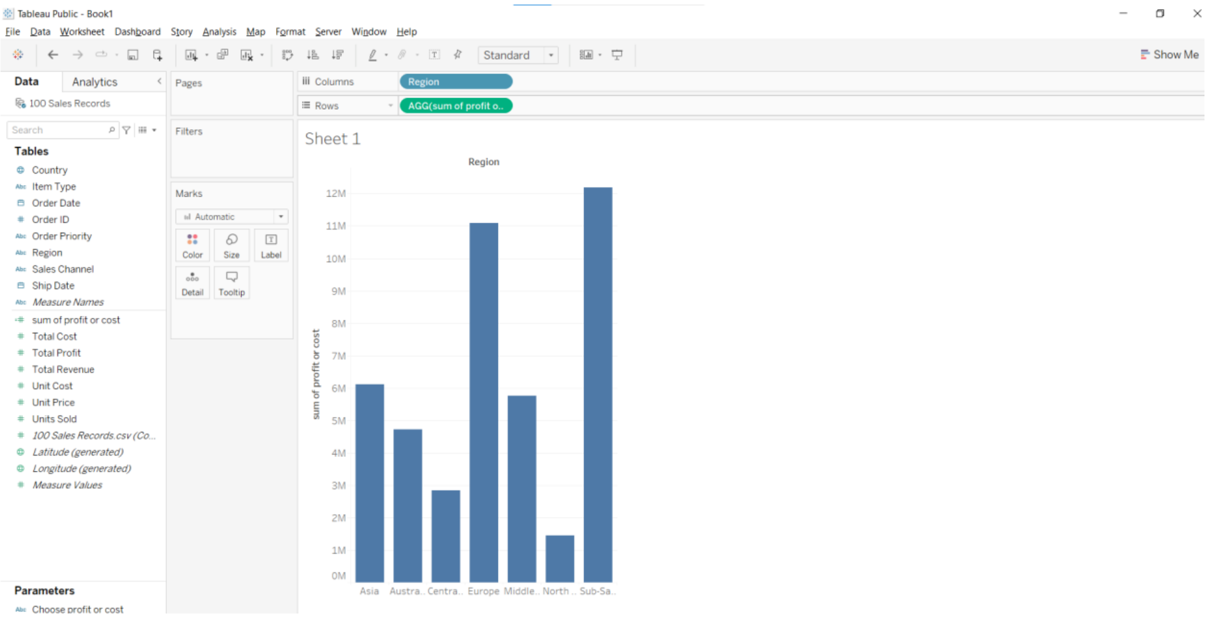
sum of profit or cost

```
CASE [Choose profit or cost]
WHEN 'Sum of profit' THEN SUM([Total Profit])
WHEN 'Sum of Cost' THEN SUM([Total Cost])
END
```

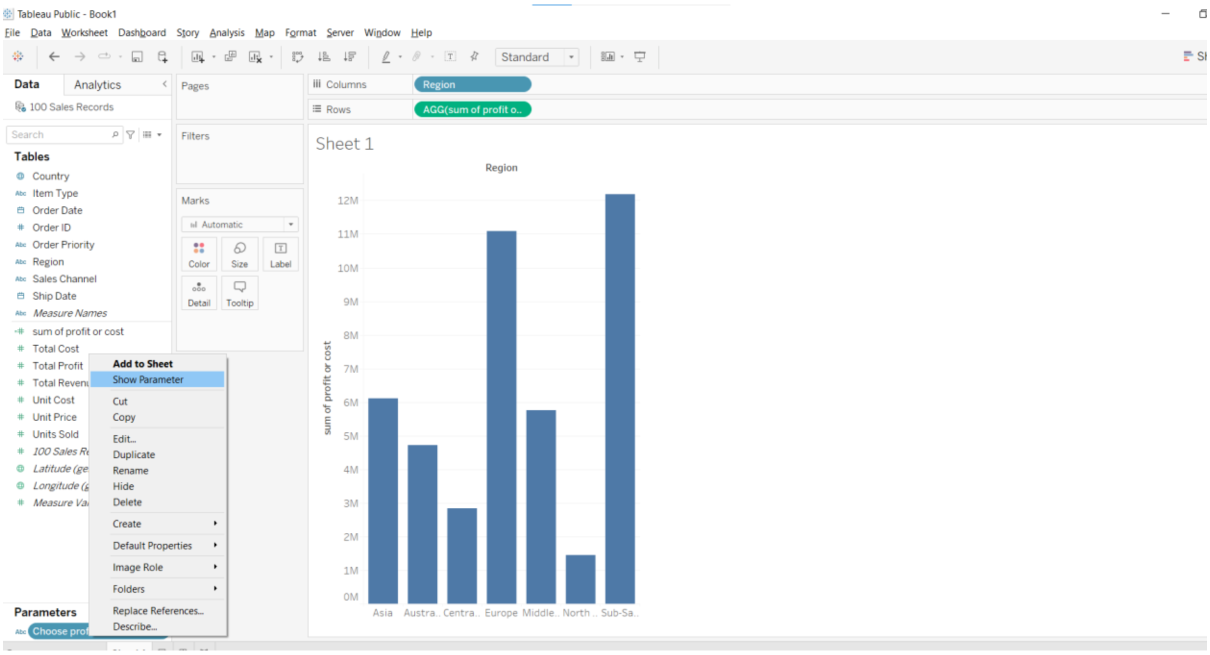
The calculation is valid.

◉ Sum of profit
○ Sum of Cost

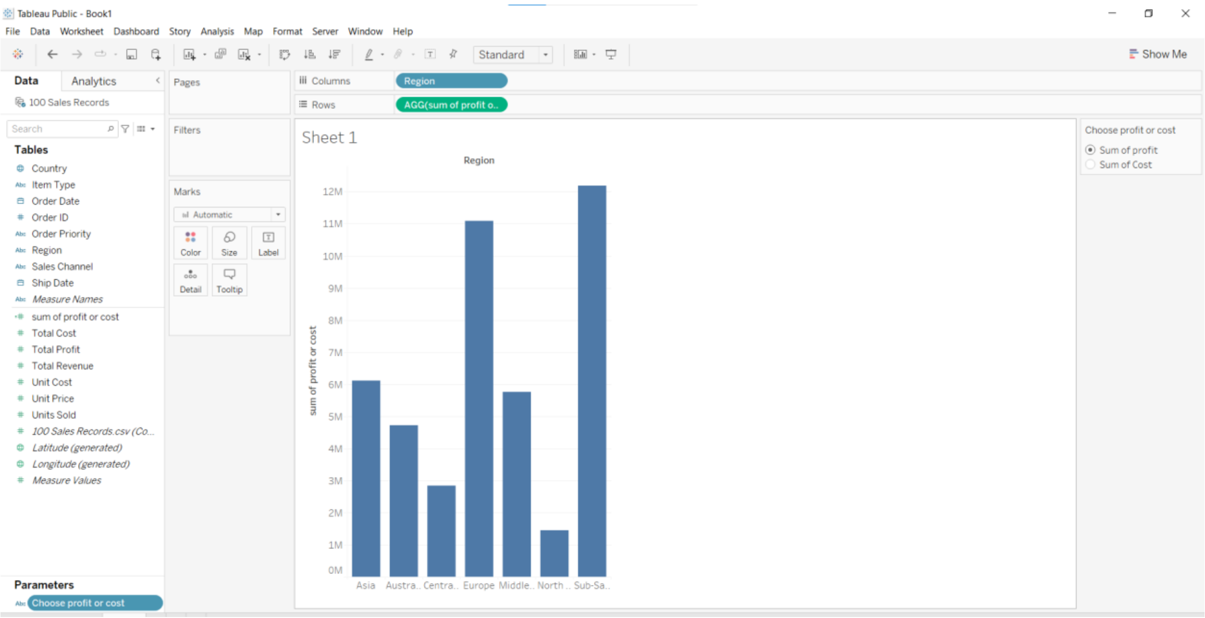
Step 7: Drag Region column to y axis and sum of profit or cost column to x axis



Step 8: Right click on parameter and select option show parameter



Step 9: Final Output



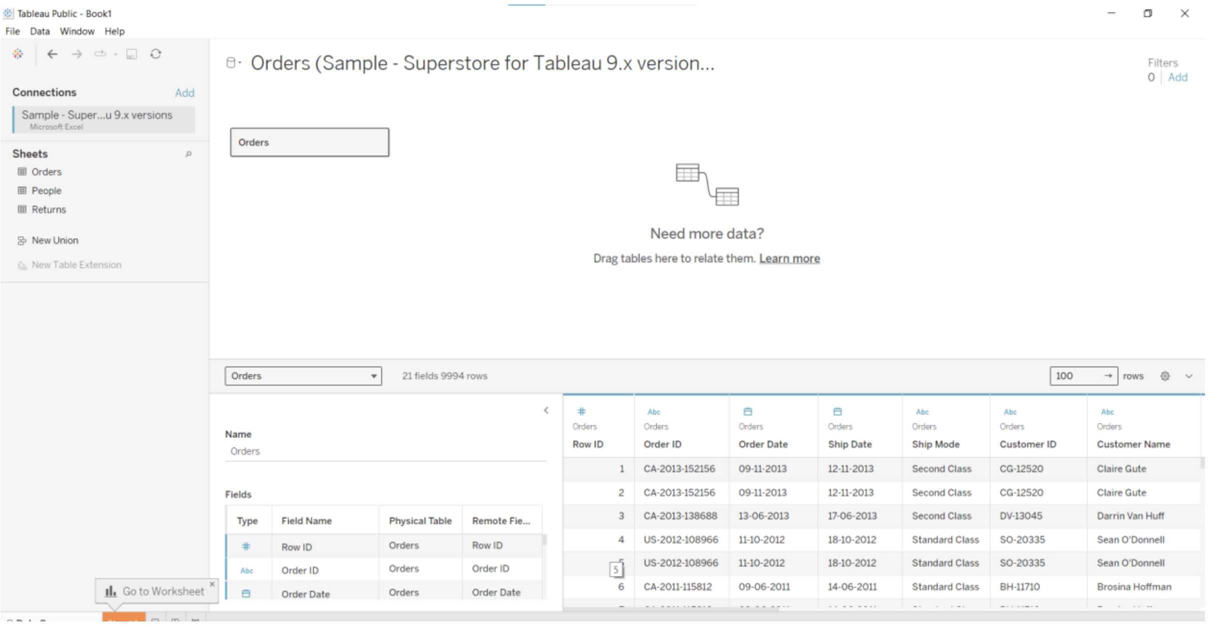
Name – Siddhesh Santosh Teli
MSc CS

Roll No. 18

Practical No. 7

Aim: Create Dashboard and Storytelling using tableau.

Step 1: Open Tableau Desktop and load the sample csv file



Orders ▾ 21 fields 9994 rows — 100 rows

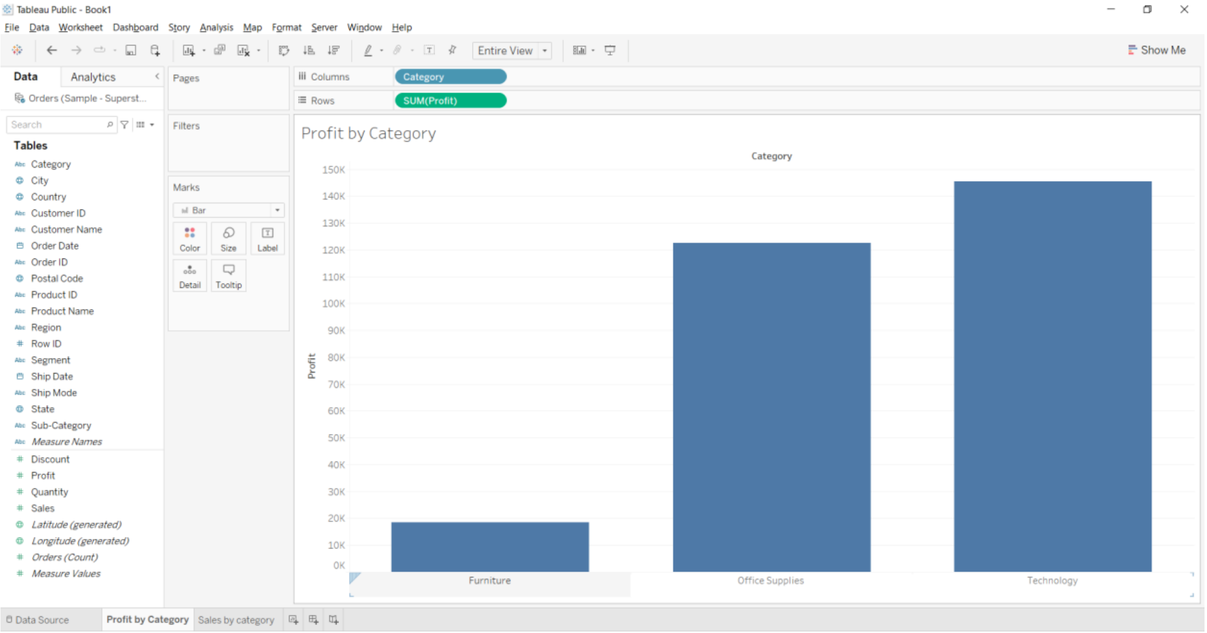
Name: Orders

Fields

Type	Field Name	Physical Table	Remote Fie...
#	Row ID	Orders	Row ID
Abc	Order ID	Orders	Order ID
📅	Order Date	Orders	Order Date

Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name
1	CA-2013-152156	09-11-2013	12-11-2013	Second Class	CG-12520	Claire Gute
2	CA-2013-152156	09-11-2013	12-11-2013	Second Class	CG-12520	Claire Gute
3	CA-2013-138688	13-06-2013	17-06-2013	Second Class	DV-13045	Darrin Van Huff
4	US-2012-108966	11-10-2012	18-10-2012	Standard Class	SO-20335	Sean O'Donnell
5	US-2012-108966	11-10-2012	18-10-2012	Standard Class	SO-20335	Sean O'Donnell
6	CA-2011-115812	09-06-2011	14-06-2011	Standard Class	BH-11710	Brosina Hoffman

Step 2: Create new sheet and name it as profit by category and drag category to column and profit to rows

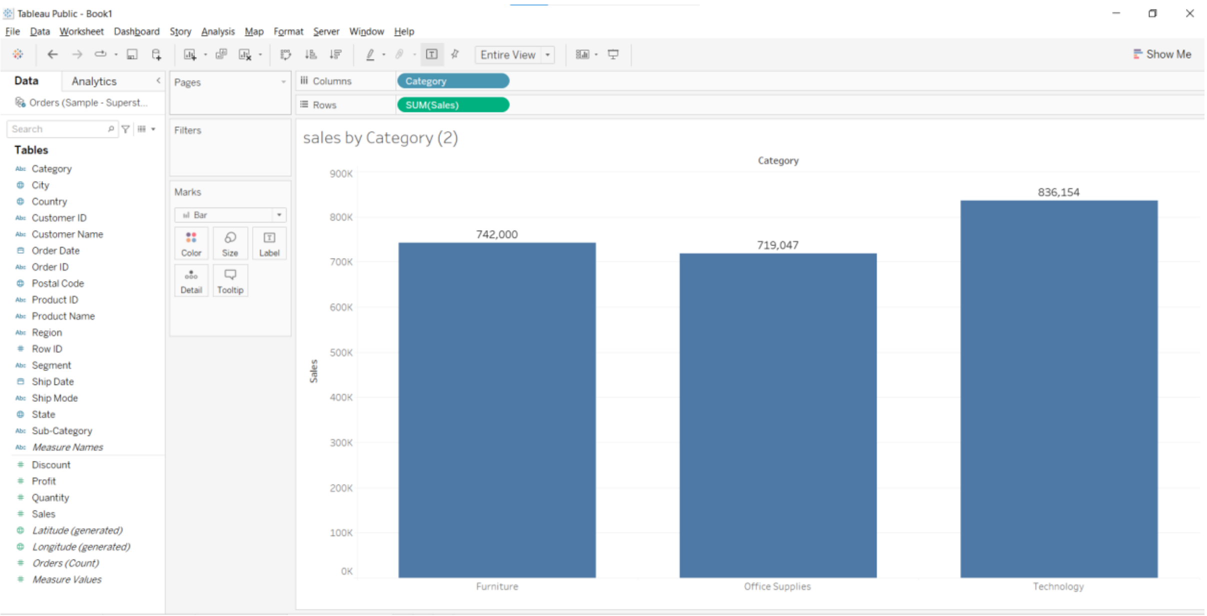


Columns: Category

Rows: SUM(Profit)

Profit by Category

Step 3: Create new sheet and name it as sales by category and drag category to column and sales to rows

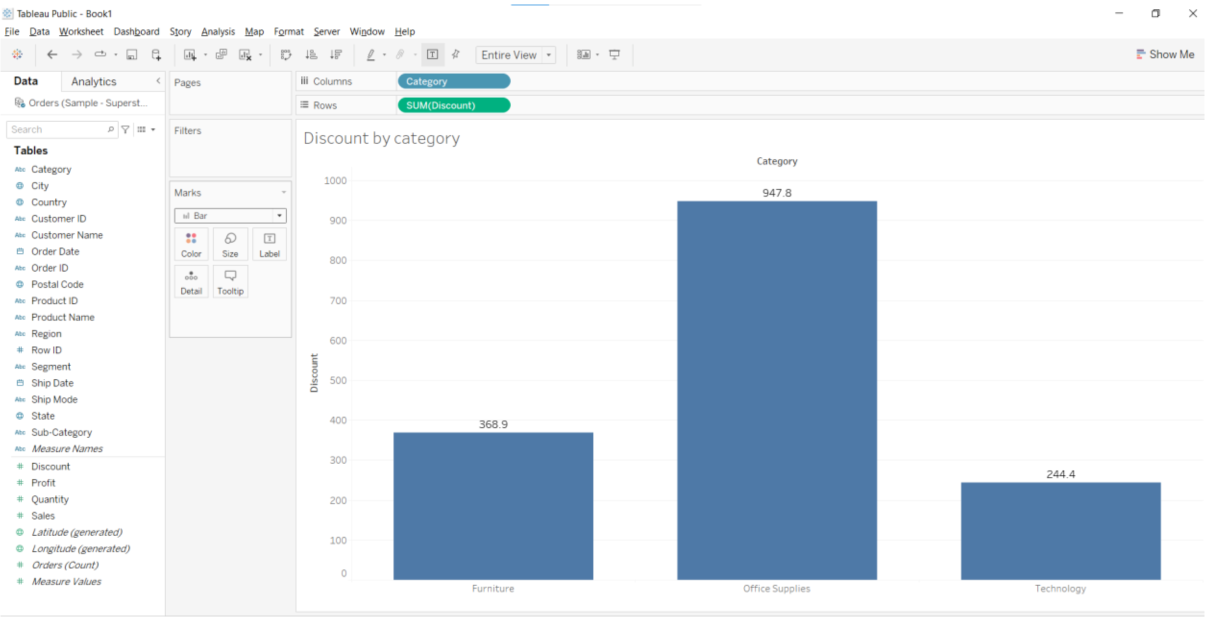


Columns: Category

Rows: SUM(Sales)

sales by Category (2)

Step 4: Create new sheet and name it as discount by category and drag category to column and discount to rows

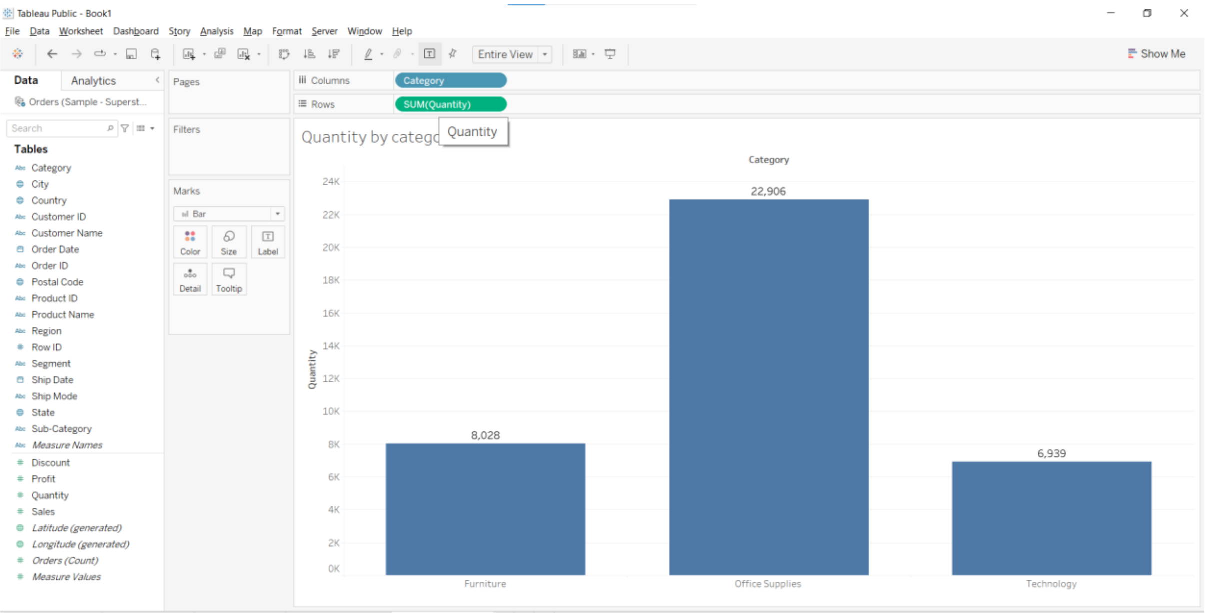


Columns: Category

Rows: SUM(Discount)

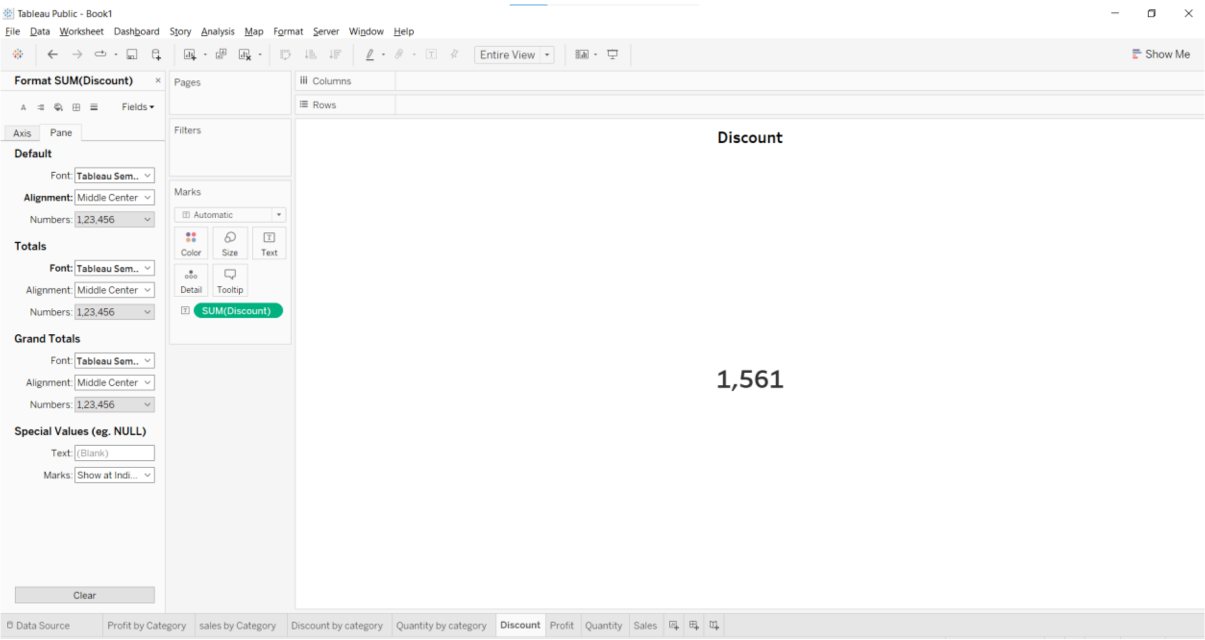
Discount by category

Step 5: Create new sheet and name it as quantity by category and drag category to column and quantity to rows



Step 6: Create KPI to show on dashboard for Sales, Quantity, Profit, Discount

Drag the Discount column to the center of sheet and right click select format increase the text size and align it to center and change the view to Entire View



Format SUM(Discount)

Fields

Axis Pane

Default

Font: Tableau Sem...

Alignment: Middle Center

Numbers: 1,23,456

Totals

Font: Tableau Sem...

Alignment: Middle Center

Numbers: 1,23,456

Grand Totals

Font: Tableau Sem...

Alignment: Middle Center

Numbers: 1,23,456

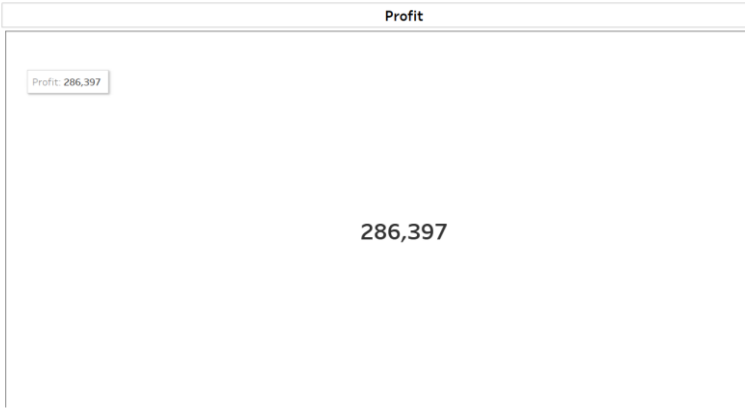
Special Values (eg. NULL)

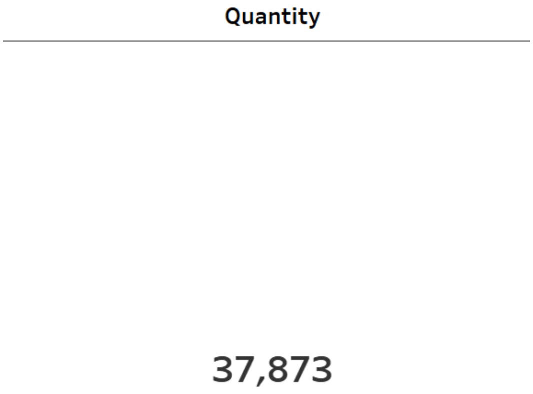
Text: (Blank)

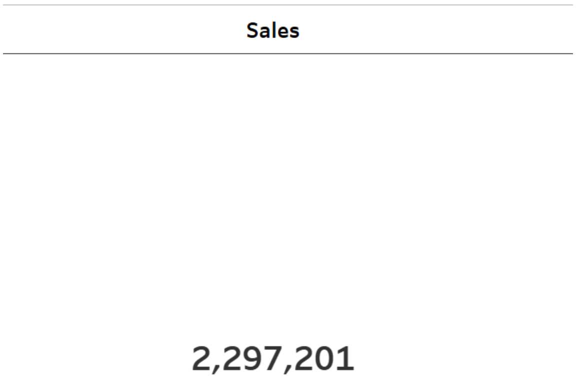
Marks: Show at Indi...

Clear

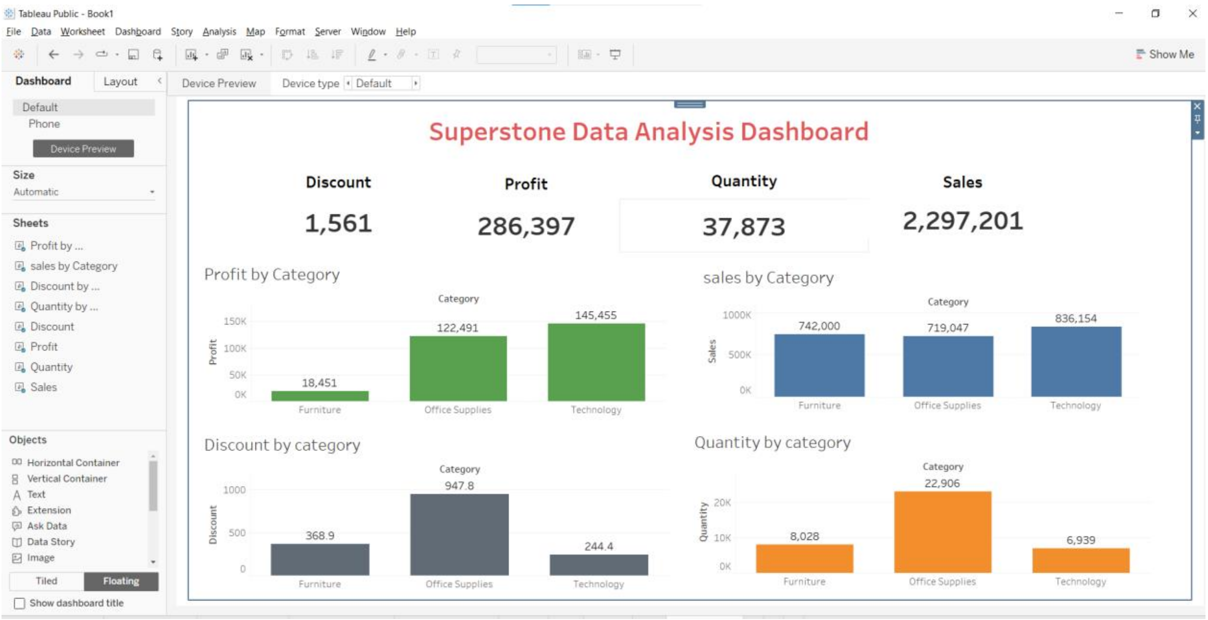
Step 7: Follow same steps for Quantity, Profit and Sales



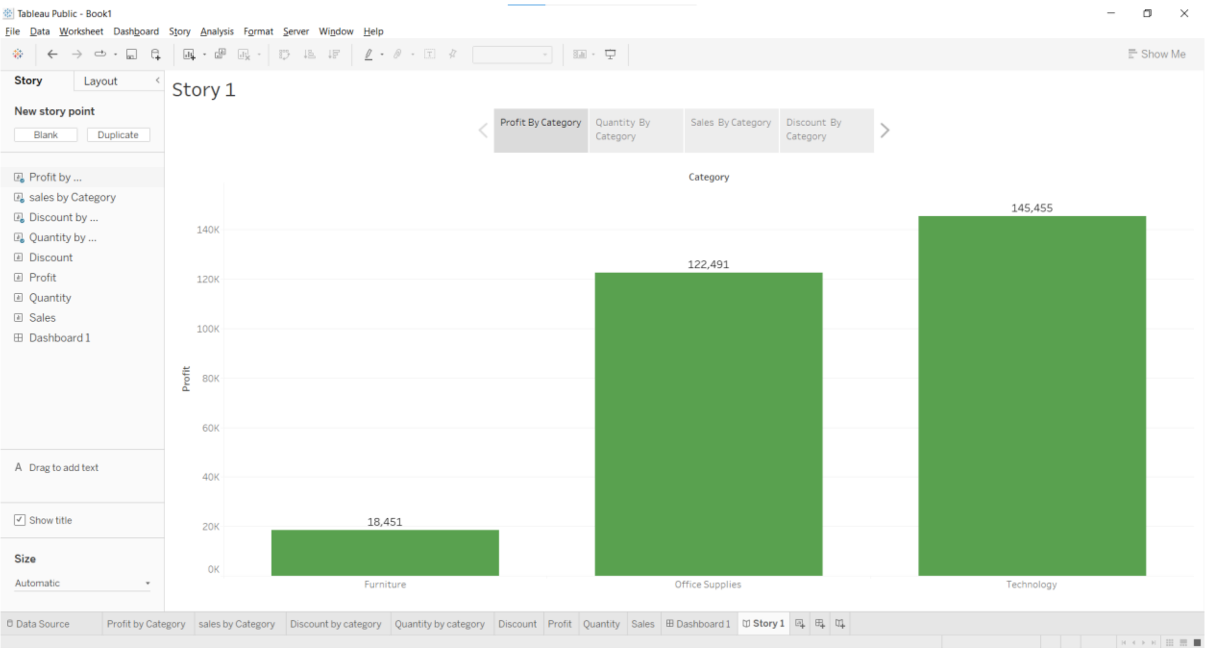


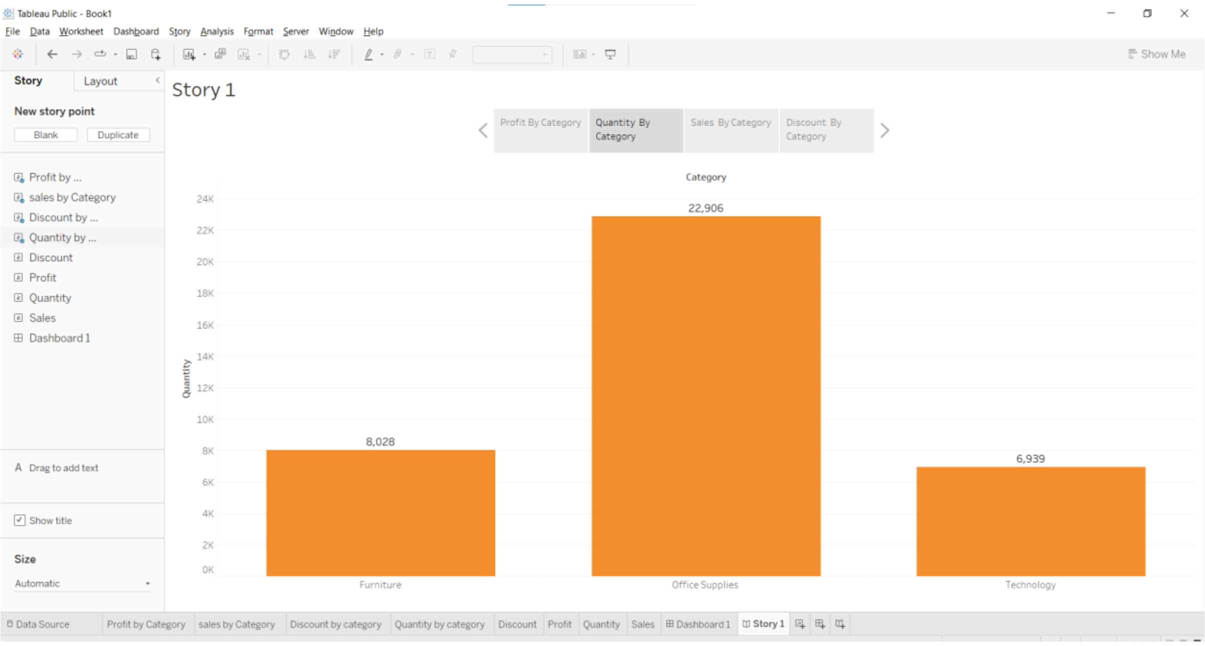


Step 8: Create new Dashboard and Change the type from Tiled to Floating and arrange the sheets in dashboard



Step 9: Create new story and add the story points one by one and also add the sheets



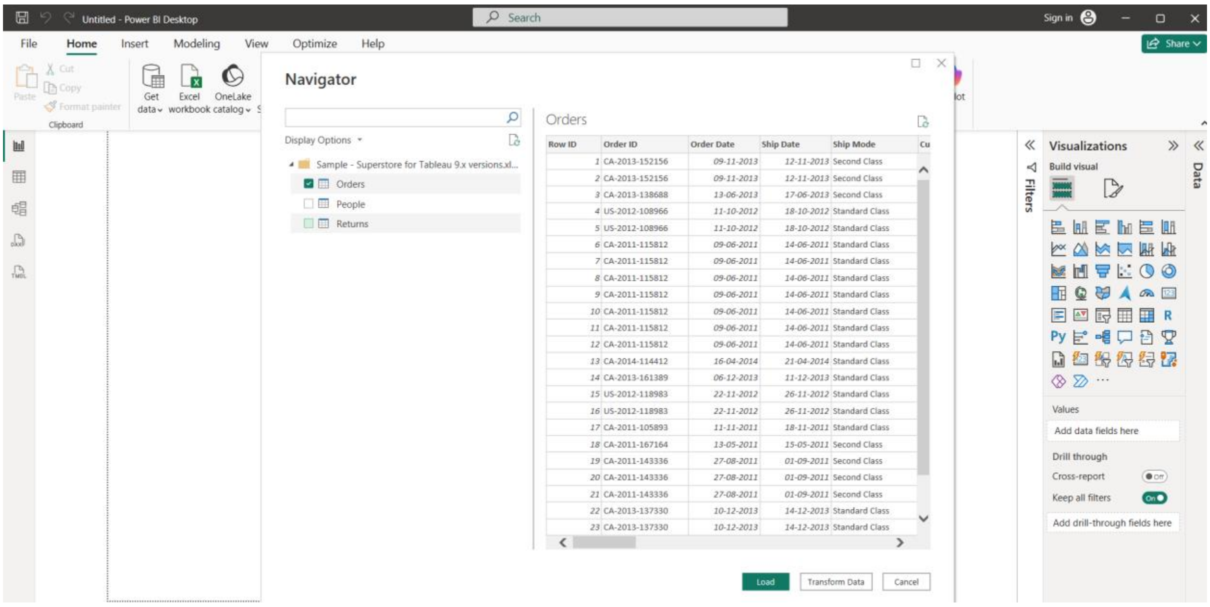


Name – Siddhesh Santosh Teli Roll No. 18
MSc CS

Practical No. 8

Aim: Data Visualization Power BI—bar charts, line charts, pie charts, tables, matrices, simple cards

Step 1: Open Power BI Desktop app and load the data

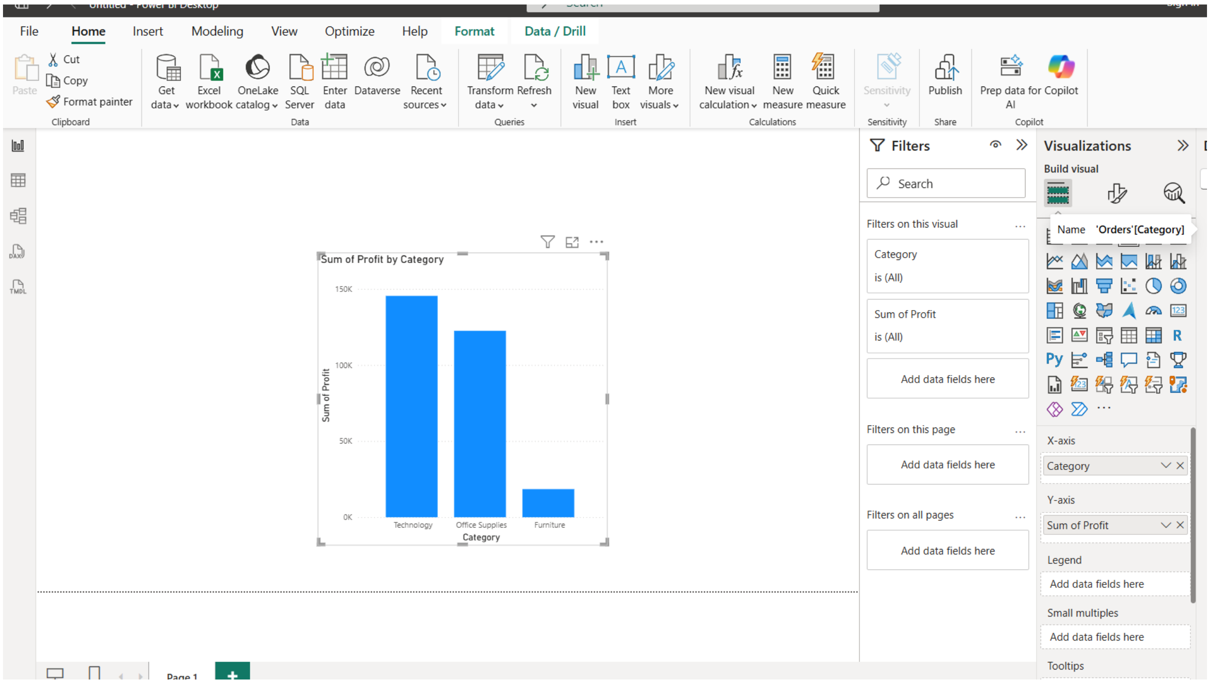


Sample - Superstore for Tableau 9.x versions.xl...
- ☑ Orders
- □ People
- □ Returns

Orders

Row ID	Order ID	Order Date	Ship Date	Ship Mode
1	CA-2013-152156	09-11-2013	12-11-2013	Second Class
2	CA-2013-152156	09-11-2013	12-11-2013	Second Class
3	CA-2013-138688	13-06-2013	17-06-2013	Second Class
4	US-2012-108966	11-10-2012	18-10-2012	Standard Class
5	US-2012-108966	11-10-2012	18-10-2012	Standard Class
6	CA-2011-115812	09-06-2011	14-06-2011	Standard Class
7	CA-2011-115812	09-06-2011	14-06-2011	Standard Class
8	CA-2011-115812	09-06-2011	14-06-2011	Standard Class
9	CA-2011-115812	09-06-2011	14-06-2011	Standard Class
10	CA-2011-115812	09-06-2011	14-06-2011	Standard Class
11	CA-2011-115812	09-06-2011	14-06-2011	Standard Class
12	CA-2011-115812	09-06-2011	14-06-2011	Standard Class
13	CA-2014-114412	16-04-2014	21-04-2014	Standard Class
14	CA-2013-161389	06-12-2013	11-12-2013	Standard Class
15	US-2012-118983	22-11-2012	26-11-2012	Standard Class
16	US-2012-118983	22-11-2012	26-11-2012	Standard Class
17	CA-2011-105893	11-11-2011	18-11-2011	Standard Class
18	CA-2011-167164	13-05-2011	15-05-2011	Second Class
19	CA-2011-143336	27-08-2011	01-09-2011	Second Class
20	CA-2011-143336	27-08-2011	01-09-2011	Second Class
21	CA-2011-143336	27-08-2011	01-09-2011	Second Class
22	CA-2013-137330	10-12-2013	14-12-2013	Standard Class
23	CA-2013-137330	10-12-2013	14-12-2013	Standard Class

Step 2: Step Drag profit into the Page and drag that Category into Profit container and click on bar chart



Filters on this visual

Category
is (All)

Sum of Profit
is (All)

Add data fields here

Filters on this page

Add data fields here

Filters on all pages

Add data fields here

X-axis
Category

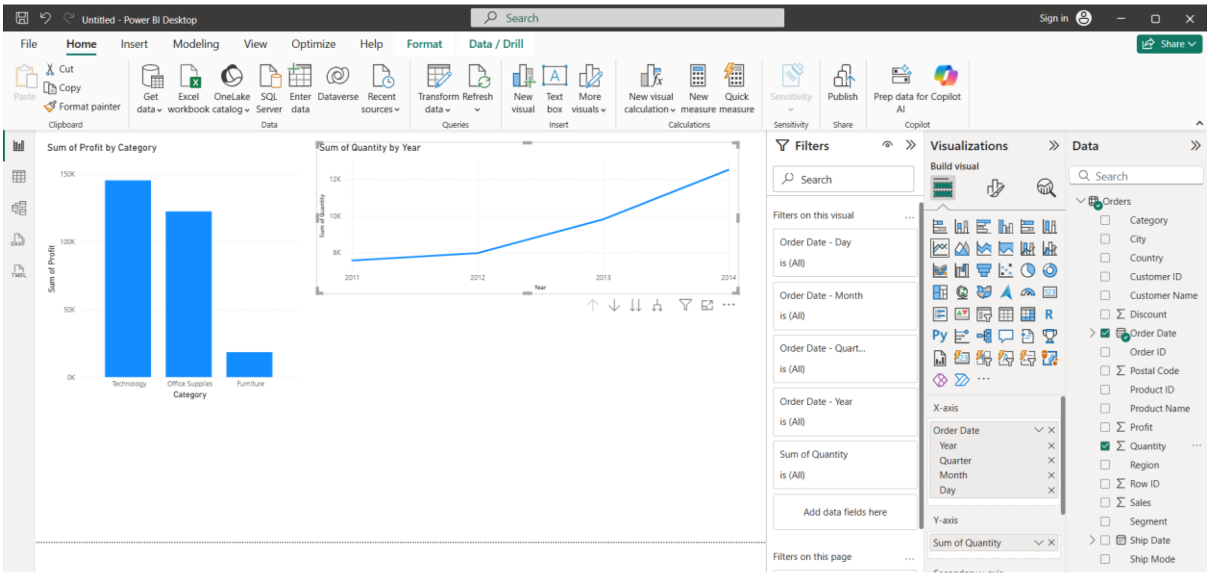
Y-axis
Sum of Profit

Legend
Add data fields here

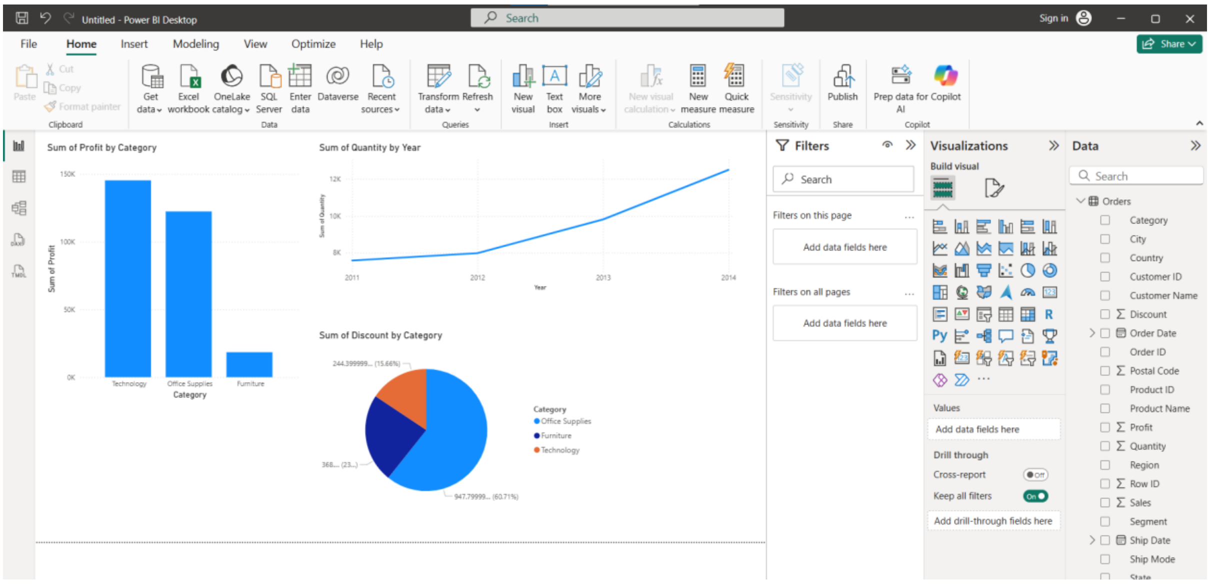
Small multiples
Add data fields here

Tooltips

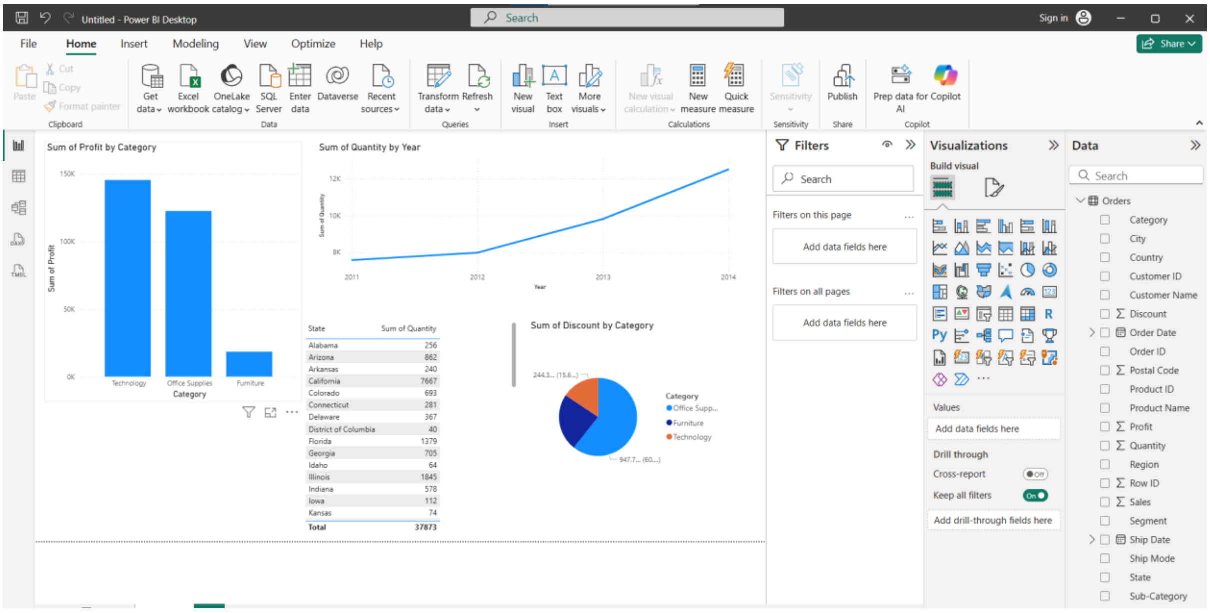
Step 3: Step Drag Quantity into the Page and drag that Order Date into Quantity container and click on line chart



Step 4: Step Drag Discount into the Page and drag that Category into Discount container and click on Pie chart

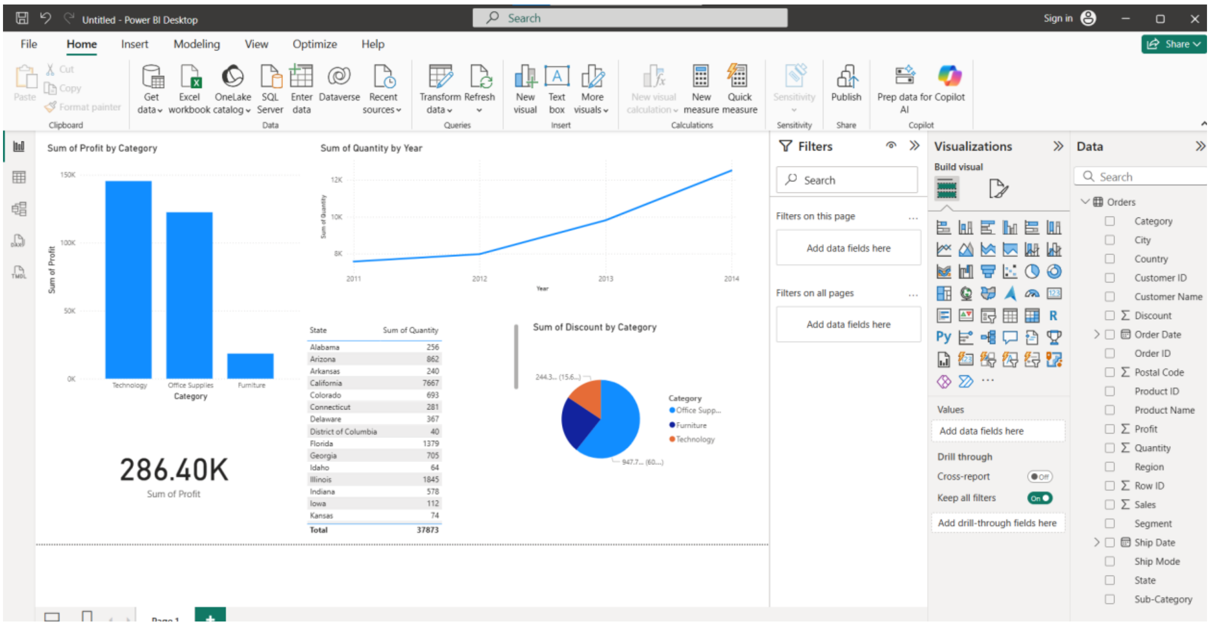


Step 5: Step Drag State into the Page and drag that Quantity into State container and click on Table



State	Sum of Quantity
Alabama	256
Arizona	862
Arkansas	240
California	7667
Colorado	693
Connecticut	281
Delaware	367
District of Columbia	40
Florida	1379
Georgia	705
Idaho	64
Illinois	1845
Indiana	578
Iowa	112
Kansas	74
Total	**37873**

Step 6: Drag card Visualization in page and then drag profit into that card



286.40K
Sum of Profit

State	Sum of Quantity
Alabama	256
Arizona	862
Arkansas	240
California	7667
Colorado	693
Connecticut	281
Delaware	367
District of Columbia	40
Florida	1379
Georgia	705
Idaho	64
Illinois	1845
Indiana	578
Iowa	112
Kansas	74
Total	**37873**