

Experiment-1

Date: _____

AIM

Understanding Data, what is data, where to find data, data wrangling, data clean up basics - formatting, outliers, duplicates, normalizing and standardizing data.

PROCEDURE

Step-1: Study about data and its importance

Step-2: Understanding about data wrangling procedure

Step-3: Understanding the importance of normalizing and standardizing the data

SOURCE CODE

~~Data:- Data is a collection of facts and figures. Data contains text, numbers, images, graphs, audio, video, special characters, etc.~~

~~Data wrangling:- Data Wrangling is the process of gathering, accessing, cleaning and transforming and organizing the raw data into a easily understandable format and analyzable format.~~

Data wrangling is also known as Data Managing

Data wrangling is a preprocessing technique

Data wrangling is used in data warehouse and data mining, data analytics, data visualization and data science applications.

Data wrangling includes

→ Identifying gaps in data and either filling the data or deleting them.

→ Deleting the unwanted data

6 Steps of Warangling Data :-

- 1) Discovering
- 2) Structuring
- 3) Cleaning
- 4) Enriching
- 5) Validating
- 6) publishing

Q) import pandas as pd

import seaborn as sns

from sklearn.preprocessing import MinMaxScaler, StandardScaler

data = {

'Name': ['Mahi', 'Dhoni', 'King', 'Chiru', 'Kohli'],

'Age': [15, 20, 25, 30, 35],

'Income': [200, 300, 400, 600, 1000],

'height': [165, 170, 180, 160, 175]

}
data

O/P:- { 'Name': ['Mahi', 'Dhoni', 'King', 'Chiru', 'Kohli'],

'Age': [15, 20, 25, 30, 35],

'Income': [200, 300, 400, 600, 1000],

'height': [165, 170, 180, 160, 175]}

2) print(data)

create a data frame

df = pd.DataFrame(data)

df
o/p:- { 'Name': ['Mahi', 'Dhoni', 'King', 'Chiru', 'Kohli'], 'Age': [15, 20, 25, 30, 35], 'Income': [200, 300, 400, 600, 1000], 'height': [165, 170, 180, 160, 175] }

	Name	Age	Income	height
0	Mahi	15	200	165
1	Dhoni	20	300	170
2	King	25	400	180
3	Chiru	30	600	160
4	Kohli	35	1000	175

3) #duplicates

df.duplicated()

O/P: 0 False

1 False

2 False

3 False

4 False

dtype: bool

4) df[df.duplicated()]

O/P Name Age Income height

5) df.duplicated().sum()

O/P 0

6) df.drop_duplicates()

df.drop

O/P:-

	Name	Age	Income	height
0	Mahi	15	200	165
1	Dhoni	20	300	170
2	King	25	400	180
3	Chiru	30	600	160
4	Kohli	35	1000	175

OUTPUT

7) # Normalization

min-max-scaler = MinMaxScaler()

df['Normalized-income'] = min_max_Scaler.fit_transform(df[['income']])

df['Normalized-income']

df

	Name	Age	Income	height	Normalized-income
0	Mahi	15	200	165	0.000
1	Dhoni	20	300	170	0.125
2	King	25	400	180	0.250
3	Chiru	30	600	160	0.500
4	Kohli	35	1000	175	1.000

8) # Standardization

StandardScaler = StandardScaler()

df['Standard-income'] = StandardScaler.fit_transform(df[['income']])

df

	Name	Age	Income	height	Normalized-income	Standard-income
0	Mahi	15	200	165	0.000	-1.060660
1	Dhoni	20	300	170	0.125	-0.707107
2	King	25	400	180	0.250	-0.353535
3	Chiru	30	600	160	0.500	-0.353535
4	Kohli	35	1000	175	1.000	-1.767767

9) # Formatting

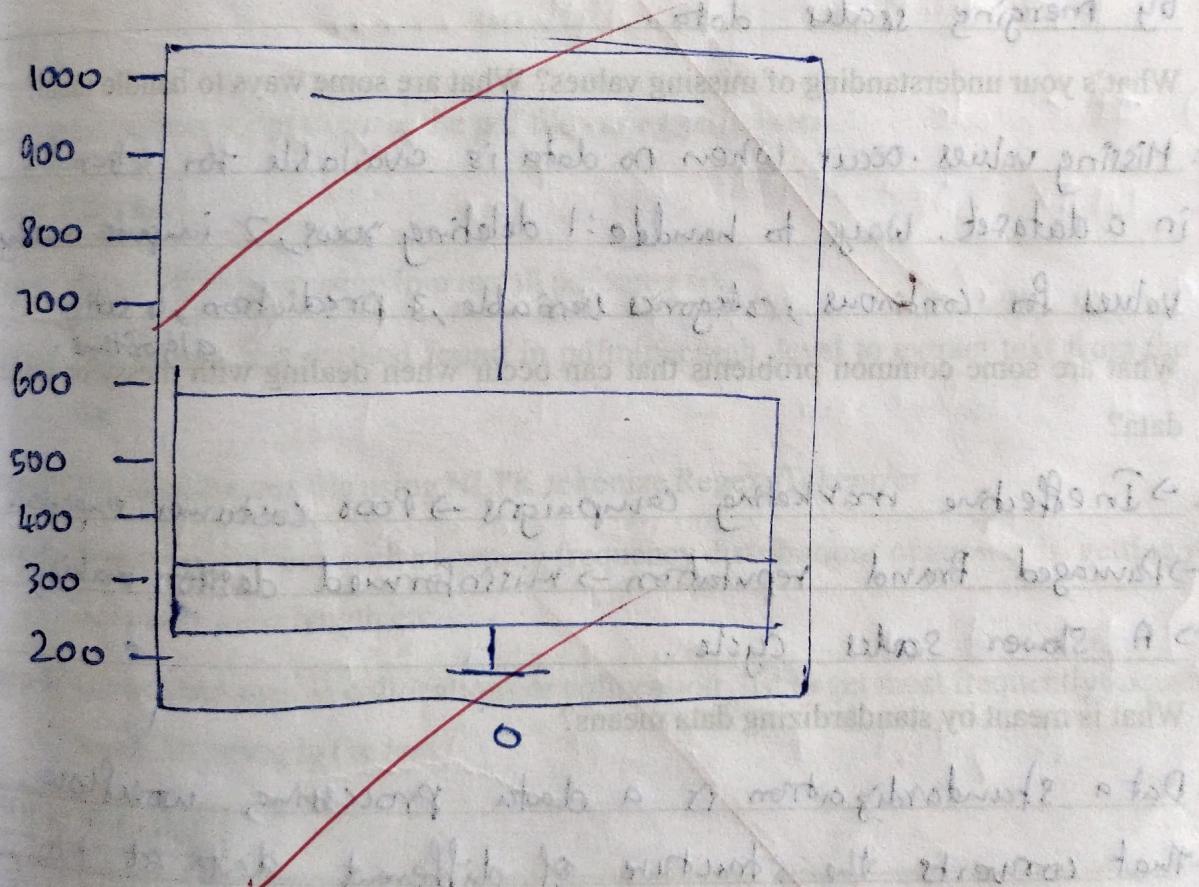
df['format-income'] = df['income'].apply(lambda x: "\${:,.2f}".format(x))

df['format-income']

df.

O/P:-	Name	Age	Income	height	Normalized-income	Standard-income	Format-income
0	Mahi	15	200	165	0.000	-1.60660	\$ 200.00
1	Dhoni	20	300	170	0.125	-0.707107	\$ 300.00
2	King	25	400	180	0.250	-0.353553	\$ 400.00
3	Chim	30	600	160	0.500	0.353553	\$ 600.00
4	Kohli	35	1000	175	1.00	-0.167767	\$ 1000.00

(2) #outline
Sbn. boxplot (df[, 'Income'])



Experiment-2

Date: _____

AIM

Develop the python script to parse the pdf files using pdfminer.

PROCEDURE

Step-1: Set up PDFMiner using !pip install pdfminer.six.

Step-2: Use extract_text method found in pdfminer.high_level to extract text from the PDF file

Step-3: Tokenize the text file using NLTK.tokenize RegexpTokenizer

Step-4: Perform operations such as getting frequency distributions of the words, getting words more than some length etc.

Step-5: Use method such as collocations or collocation_list to get most frequently sequence of words occurring in the text

SOURCE CODE

1) Pip install pdfminer.six
already satisfied

2) Pip install nltk

already satisfied

- 3) from pdfminer.high_level import extract_text
- 4) from nltk.tokenize import RegexpTokenizer
- 5) from nltk.probability import FreqDist
- 6) # Extract the text from PDF file
text = extract_text('python.pdf')
- 7) print(text)

Introduction to Python ...

- 8) # Create an instance of tokenizer using NLTK RegexpTokenizer
tokenizer = RegexpTokenizer(r'\w+')
- 9) # Tokenize the text read from pdf
tokens = tokenizer.tokenize(text)
- 10) # Find frequency distribution
freqdist = FreqDist(tokens)

11) #Find words whose length is greater than 5
and frequency greater than 20 long-frequent-words
= [words for words in tokens if
if len(words) > 6 and freq dist[words] > 5]

12) long-frequent-words

O/P:- ['Introduction',
'Programming',
'Modules',
'Exception',
'Constants',
'Strings',
'Indexing',
'Functions',
'Numeric',
'Dictionaries',
-----]

OUTPUT

13) FreqDist (long - frequent - words).plot()

Experiment-3

Date: _____

AIM

Develop the python Shell Script to do the basic data cleanup on child labour and child marriage data.xlsx
a) check duplicates and missing data b) eliminate mismatches c) cleans line breaks, spaces, and special characters.

PROCEDURE

Step-1: Import required libraries like pandas and numpy

Step-2: Read data using read.csv method in pandas

Step-3: Find basic info about data

Step-4: check duplicates and missing data.

Step-5: Eliminate mismatches

Step-6: Clean line breaks, spaces and special characters

Step-7: Detect outliers

Step-8: Normalize Casing

SOURCE CODE

C.S.		State	Agrir	Manuf	Const	C.SPS	Other	Lab Record Total
0	Non-Special category	A.P	69.0	9.7	3.2	7.7	2.0	100
1	Non-Special category Assam States	Assam	69.3	8.4	1.8	7.8	11.0	100
2	Non-Special category States	Bihar	71.8	11.8	0.0	15.5	1.1	100
3	Non-Special category States	Chhattisgarh	87.9	2.4	0.9	7.2	NAN	100
							Total	25.0
							0.0	0.0
							0.0	0.0
							0.0	0.0
2)	df.shape	(25, 9)						
3)	df.columns							
		Index(['Category of states', 'States', 'Agriculture', 'Manufacturing', 'Construction', 'Trade Hotels & Restaurants', 'Community, Social and personal Services', 'Others', 'Total'], dtype='object')						
4)	df.dtypes							
	Category of states				object			
	States				object			
	Agriculture				float 64			
5)	df.info()							
		<class 'pandas.core.frame.DataFrame'>						
		Range Index : 25 entries, 0 to 24						
		Data columns (total 9 columns):						
	# column		Non-null count		Dtype			
0	Category	25	Non-null		object			
1	States	25	non-null		object			
2	Agriculture	24	non-null		float 64			
	dtypes:	float 64 (5), int 64 (1), object (3)						
	memory usage:	1.9 + KB						
6)	df.describe()							
	Agriculture	Count	24.000000	Construction	Traders, Hotels Restaurants	Community Social & Personnel Service	Others	Total
	mean	69.000000	22.000000	2.542409	25.000000	23.000000	19.000000	25.0
	std	16.613536	2.046753	11.720000	5.795852	7.119658	1.436842	100.0
	min	19.200000	0.000000	11.189748	0.300000	0.916643	0.0	100.0
				3.300000		0.000000		

7. df.head()

	Category of States	States	Agriculture	Manufacturing	Trade, H&R Restaurants	Construction	Community Social & Personal Service
0	Non-Special Category States	AP	98.7	9.7	3.2	9.2	7.1
1	Non-Special	Assam	69.3	8.4	18	7.8	
2	Non-Special	Bihar	71.8	11.2	0.0	15.5	11.0
	Others	Total					61
		2.0	100				
		1.8	100				
		0.5	100				

8) df.tail()

	C.O.S	S	A	M	C	T, H&R	C,S,P,S	Others	Total
20	Special	Uttara khand	80.7	4.7	5.2	9.3	NaN	NaN	100
21	Special	Uttarakhand	80.7	4.7	5.2	9.3	NaN	NaN	100
22	All India	AP	80.7	16.6	2.0	8.5	3.4	1.5	100

9) df.head(1)

	C.O.S	S	A	M	C	T,H&R	C,S,P,S	Others	Total
0	Non-Special	AP	69.0	9.7	3.2	9.0	7.1	2.0	100

10) df.tail(1)

	C.O.S	S	A	M	C	T,HRR	C,S,P,S	Others	Total
24	Non-Special	Bihar	71.8	11.2	0.0	15.5	1.1	0.5	100

11. df.duplicated()

- 0 False
- 1 False
- 2 False

d type : bool

12. df.duplicated.sum()

2

13 duplicated = df['df duplicated']
duplicated

	C.O.S	States	A	M	C	T, H&R	C, S, O, & P.S	Others	Total
17	Non-Special	U.P.	61.02	25.3	0.4	9.7	2.1	1.2	100
21	Special	Uttarakhand	80.7	4.7	5.2	9.3	NaN	NaN	100

14. df['states'].duplicated()

0 False

1 False

2 False

Name : states, dtype = bool

15. df['states'].duplicated().sum()

16. ~~duplicated - States = df[df['states'].duplicated()]~~
~~duplicated - States~~

	C.O.S	States	A	M	C	T, H&R	C, S, O, & P.S	Others	Total
17	Non-Special	U.P.	61.2	25.3	0.4	9.7	2.1	2.7	100
19	Special	Uttarakhand	80.7	4.7	5.2	9.3	NaN	NaN	100
23	Non-Special	A.P.	69.0	9.7	4.0	8.0	7.1	2.0	100

17 df.drop.duplicates()

	C.O.S	States	A	M	C	T, H&R	C, S, O, & P.S	Others	Total
0	Non-Special	A.P.	69.0	9.7	3.2	9.0	7.1	2.0	100
1	Non-Special	Assam	69.3	8.4	1.8	7.8	11.0	1.8	100
2	Non-Special	Bihar	71.8	11.2	0.0	15.3	1.1	0.5	100

18. df['states'].drop.duplicates()

0 Andra Pradesh

1 Assam

2 Bihar

Name: states, dtype: object

OUTPUT

19. `a = df['states'].drop_duplicates()`

```
a
0 Andhra pradesh
1 Assam
2 Bihar
```

Name: States, dtype: object

20. ~~`df['states'].drop_duplicates()`~~

```
0 Andhra pradesh
1 Assam
2 Bihar
```

Name: States, dtype: object

21. ~~`states_match = df[df['states'] == 'Andhra Pradesh']`~~

States_match

C.O.S	States	A	M	C	T, H & R	C, S & P.S	Others	Total
0 Non-Special	A.P	69.0	9.7	3.2	9.0	7.1	2.0	100
2 Non-Special	A.P	69.0	9.7	4.0	8.0	7.1	2.0	100

22. ~~`states_not_match = df[df['states'] != 'Andhra Pradesh']`~~

States_not_match

C.O.S	S	A	M	C	T, H & R	C, S & P.S	Others	Total
1 Non-Special	Assam	89.3	8.4	1.8	7.8	11.0	1.0	100
2 Non-Special	Bihar	71.8	11.2	0.0	15.5	1.1	0.5	100

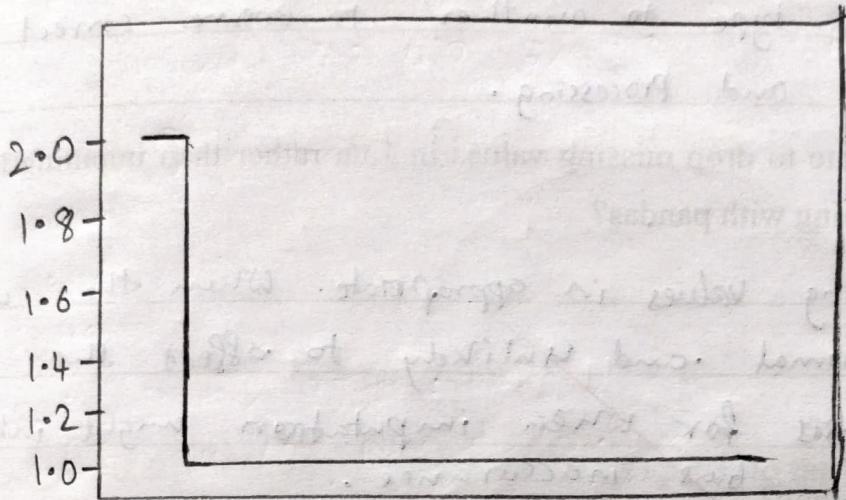
23. ~~`df['states'].value_counts()`~~

Andhra Pradesh 2

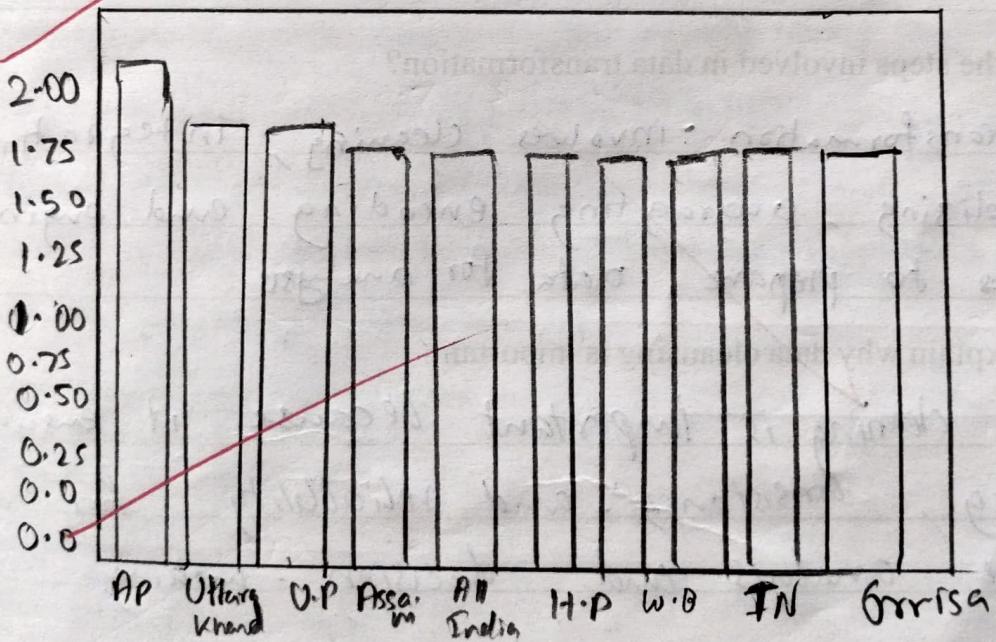
Uttarakhand 2

Name States; dtype : int 64

24 df ['States'].value_counts().plot()
(Axes: >



25. df ['States'].value_counts().plot(kind = 'bar')
(Axes: >



Experiment-4

Date: _____

AIM

Draw the chart between perceived corruption scores compared to the child labour percentages using matplotlib.

PROCEDURE

Step-1: Install the Matplotlib package

Step-2: Read the required data using read.csv method of pandas library

Step-3: Define the x-axis and corresponding y-axis values as lists.

Step-4: Plot them on canvas using .plot() function.

Step-5: Give a name to x-axis and y-axis using .xlabel() and .ylabel() functions.

Step-6: Give a title to your plot using .title() function.

Step-7: To view your plot, use .show() function.

SOURCE CODE

```
1) import pandas as pd  
import numpy as np  
import matplotlib.pyplot as plt  
2) df = pd.read_csv('Child Labour in India.csv')  
3) df
```

OP:	Category of States	Agriculture	Manufacturing	Construction	Trade Hotel Restaurant	Community social personal services	Others	Total
19	Non-Special Category States	Andhra Pradesh	69.0	9.7	3.2	9.0	2.1	110.1
20	Non-Special Category States	Assam	69.3	8.4	1.8	7.8	1.0	110.1

4) df['category'] = df[['category' of states']].replace(['non special category states', 'Special category states'], ['Non special', 'Special'])

5) df['category'], dtype=object

OP: 0

1a

22

23

24

Non special

special

All India

Non special

Non-special category states *

Name: Category, dtype: object

6. df['category'] == 'Special'

OP: 0

False

19 True

22 False

Name: Category, dtype: bool

7) df[df['category'], dtype=bool] == 'Special']

OP:

Category of States	States	Agriculture	Manufacturing	Construction	Trade Hotel Restaurant	Community social personal services	Others	Total
Special Category States	Himachal	87.4	Nan	Nan	6.7	4.7	7.2	100
Special Category States	Uttarakhand	80.7	4.7	5.2	9.3	Nan	Nan	100
Special Category States	Uttarakhand	80.7	4.7	5.2	9.3	Nan	Nan	100

20

8) len (df[df['Category'] == 'Special'])

OP: 3

9) len (df[df['Category'] == 'Non Special'])

OP: 20

10) # Percentage of special and non special and All India
Category of states

Special = len (df[df['Category'] == 'Special'])

non-Special = len (df[df['Category'] == 'Non Special'])

All_India = len (df[df['Category'] == 'All India'])

Sizes = [Special, Non-Special, All_India]

Labels = 'Special', 'Non special', 'All_India'

Colors = ['skyblue', 'yellowgreen', 'red']

Explode = [0.2, 0, 0.3]

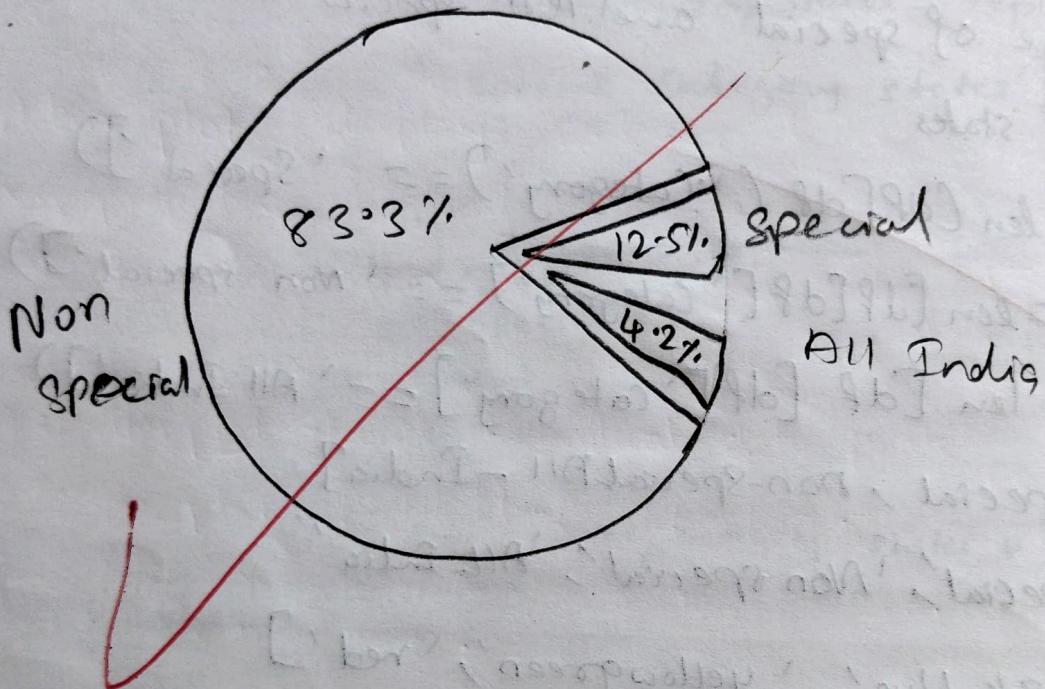
plt.pie (Sizes, labels=Labels, colors=Colors, explode=

Explode, autopct='%.1f%%', shadow=True)

States and All India in the data set')

Text (0.5, 1.0, 'Percentage of special and Non special
and all India in the dataset')

Percentage of special and non-special states and all India in the dataset.



Experiment-5

Date: 10/9/24

AIM

Write a python program to download & display content of robot.txt for en.wikipedia.org.

PROCEDURE

Step-1: Use Requests (HTTP for Humans) Library for Web Scraping

Step-2: Scrape the robots.txt file

Step-3: print the data

SOURCE CODE

```
import requests  
response = requests.get("https://en.wikipedia.org/robots.txt")  
test = response.text  
print(test)
```

OUTPUT

```
# robots.txt for http://www.wikipedia.org/ and friends  
#  
# please note: There are a lot of pages on this site, and  
# some misbehaved spiders out there that go-way-too-fast. If you're  
# irresponsible, your access to the site may be blocked  
#
```

Experiment-6

Date: _____

AIM

Foundations for building data visualizations, Creating first visualization.

PROCEDURE

Step-1: Study about basics of data visualization

Step-2: Study about prerequisites of data visualization

Step-3: Study about different visualization charts for data visualization

Step-4: read the required data into tableau using source connection in Tableau

Step-5: Study different visualization charts available in Tableau on the data

SOURCE CODE

Foundation for building data visualization creating an effective visualization requires a strong foundation in several areas. They are

- 1.) Data Analysis
- 2.) Statistical knowledge
- 3.) Domain knowledge
- 4.) Visualization Tools
- 5.) Design principles
- 6.) Interactivity

Data analysis: Before creating visualizations, you should thoroughly analyze your data and understand its structure, relationships, etc.

Statistical knowledge: - Understanding basic statistical essentials for making meaningful interpretations of data. Concepts like mean, median, standard deviation are commonly used in data visualization.

Domain knowledge: Having knowledge of specific domain (or) subject matter related to your data is crucial for creating visualization. It helps you to ask right questions and provide valuable insights.

Visualize tools: - Familiarize yourself with data visualization tools and libraries such as matplotlib, which can be used for different types of visualization.

Design principles: - Study design principle to create effective visualization. Avoid common pitfalls like misleading visualizations.

Interacting: Learn how to add interactive elements to your visualization to engage users and allow them to explore the data.

Creating your first visualization:-

Select your data: Choose a data set that aligns with your goals and interests. Ensure that data is clean and structural.

Define your objective:- (Clearly define what you want to communicate or explore with your visualization are you looking to show trends, comparison or distributions.)

Choose the right visualization type:- Select a visualization type that suits your data and objectives like bar chart, pie charts, plots, histograms

Create the visualization:- Use a suitable tool or library to create your visualization. Customize it with labels, colours etc.

Interactivity:- Add interactive feature to your visualization to allow users to interact with the data.

Test and Interact:- Preview your visualization for accuracy and clarity. Get feedback from others and make improvements as necessary.

Publish or share:- Once you're satisfied with your visualization, publish it or share it with audience.

Document and explain:- Provide context and explanation for your visualizations.

OUTPUT

Maintain and update :- If the data changes can
new data added update your visualization
accordingly.

Experiment-7

Date: _____

AIM

Getting started with tableau software using data file formats, connecting data to tableau, creating basic charts (line, bar charts, tree maps) using the show me panel.

PROCEDURE

Step-1: Study about overview of Tableau

Step-2: To open the application, click the Tableau icon on your desktop (or in your Start menu)

Step-3: In the Connect panel at the left side of the Start page, click the Excel link under the "File" heading to the open file selection option.

Step-4: Using the file selection box, select the Excel worksheet that you want to open, and click the Open button to continue

Step-5: Drag the dimension and measure in row and column input field and it will automatically suggest a graph best fitted on data.

Step-6: You can change the graph by clicking on the show me button and select whichever graph you want.

SOURCE CODE

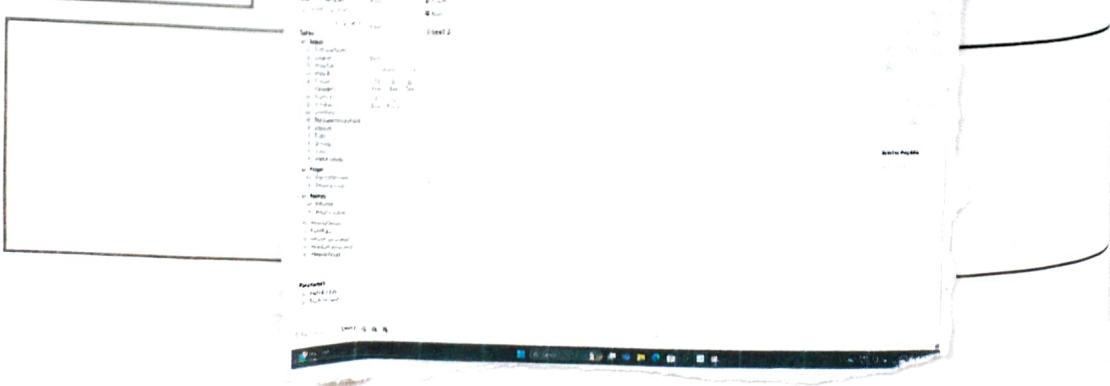


Tableau: Tableau is a software it is used to create a data visualization we have five types of tableau

- 1) Desktop
- 2) server
- 3) reader
- 4) online
- 5) public

i) Tableau desktop: Tableau desktop is used to create and analyse the data. We can create worksheet dashboards, stories by using tableau desktop

~~Title bar:- This title bar indicate the name of application along with file name.~~

~~menu bar:- In menu bar has a file, data, worksheet, etc.~~

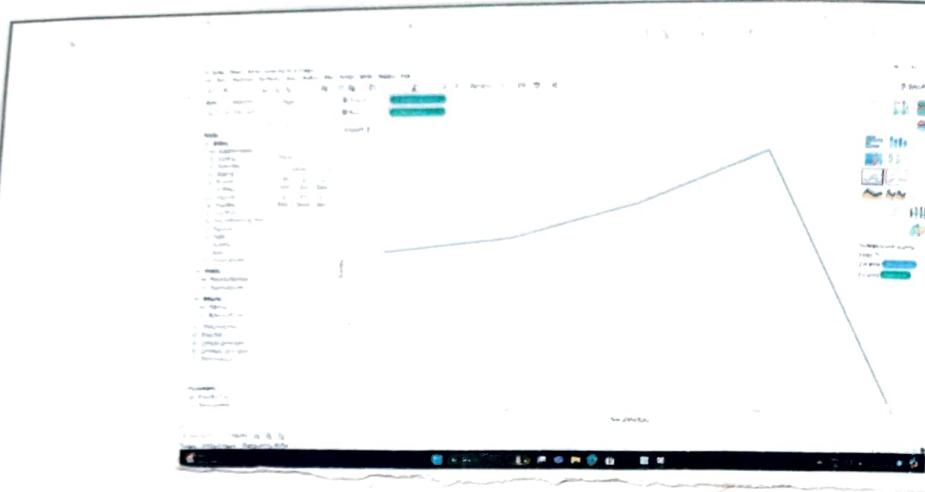
~~tool bar: It has a short cut tools.~~

~~data window: It has a dimensions and measures~~

~~→ discrete data is called dimensions (in the form of bullet)~~

~~→ continuous data is called as measures (in the form of green colour)~~

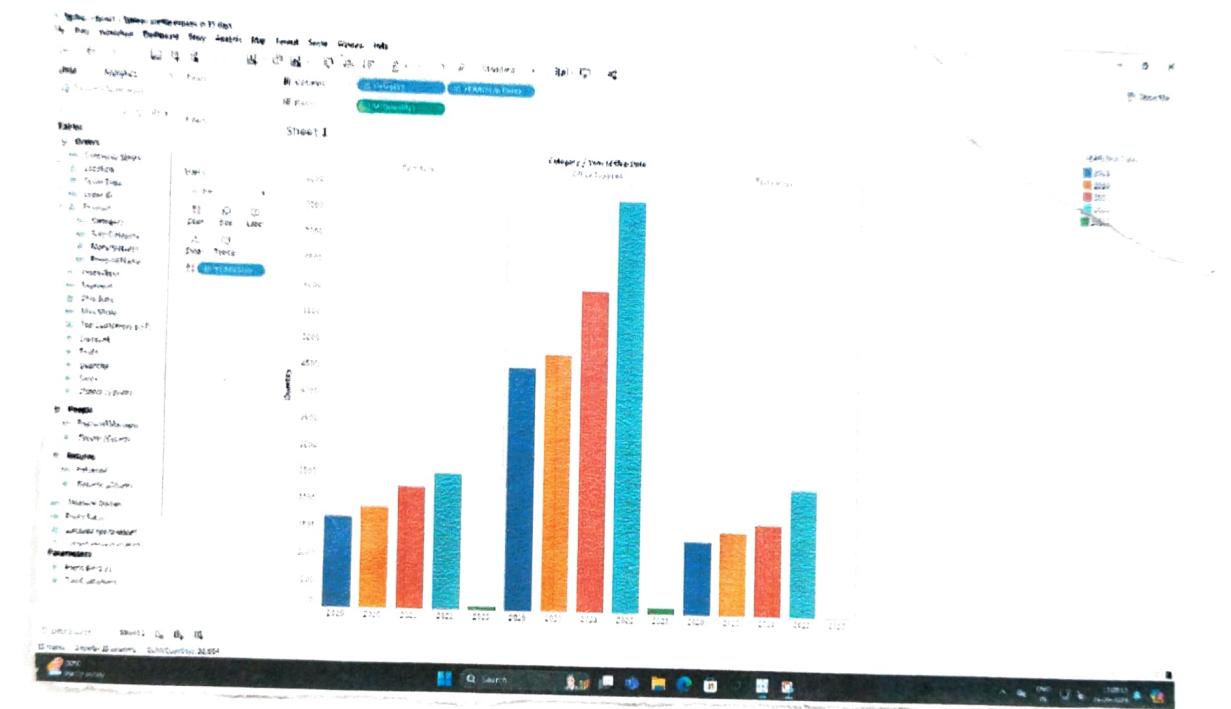
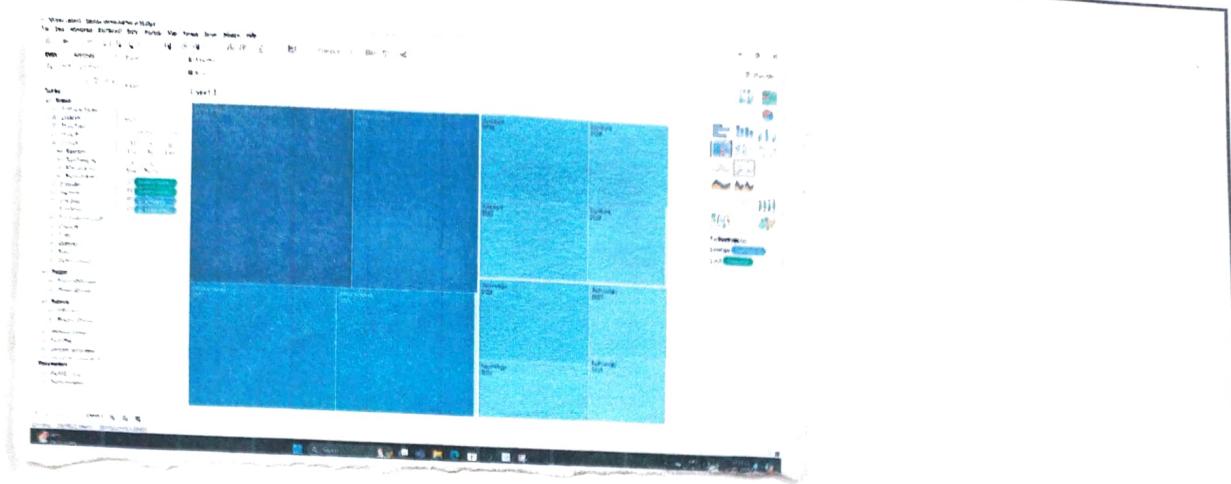
Line chart:- This line chart is used for create a line (1 date, 0 or more dimensions (or) more measures are required for line chart.



~~Bar chart: It is represented in rectangular shape. In this bar chart we can measure the values based on height. Bar chart has horizontal and vertical bar charts. → This bar chart is used for comparisons.~~

~~Tree map: It is used for create a tree for more dimensions, 1 or more measures are required for tree map.~~

OUTPUT



U

Experiment-8

Date: _____

AIM

Tableau calculations, overview of SUM, AVG and aggregate features, creating custom calculation and fields.

PROCEDURE

Step-1: create the calculated field

- In a worksheet in Tableau, select Analysis > Create Calculated Field.
- In the Calculation Editor that opens, give the calculated field a name

Step-2: Enter a formula

- In the Calculation Editor, enter a formula.

Example for how to enter formula:

$\text{SUM}([\text{Profit}])/\text{SUM}([\text{Sales}])$

- When finished, click OK.

The new calculated field is added to the Data pane

Step-3: study the overview of SUM, AVG and aggregate functions

SOURCE CODE

Types of calculations :-

You create calculated fields using calculations. There are three main types of calculations you can use to create calculated fields in Tableau.

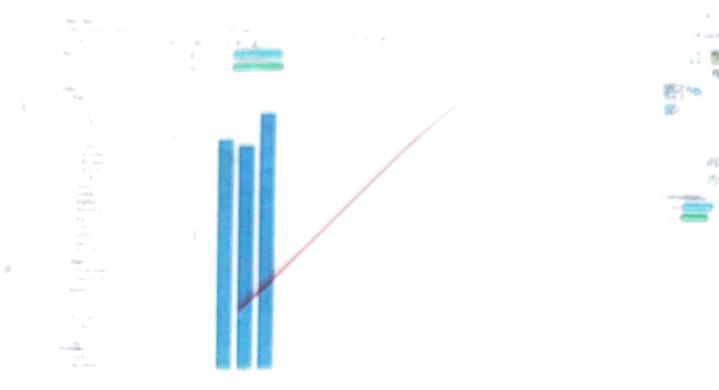
→ **Basic calculations**: Basic calculations allow you to transform values or members at data source level of detail (a row-level calculation) or at the visualization level of detail (an aggregate calculation).

→ **Level of detail (LOD) Expression**: Just like calculation LOD calculations allow you to compute values at the data source level, and the visualization level. However, LOD calculations give you even more control over the granularity you want to compute. They can be performed at a more granular level (INCLUDE), a less granular level (EXCLUDE) or an entity independent level (FIXED) with respect to the granularity of the visualization.

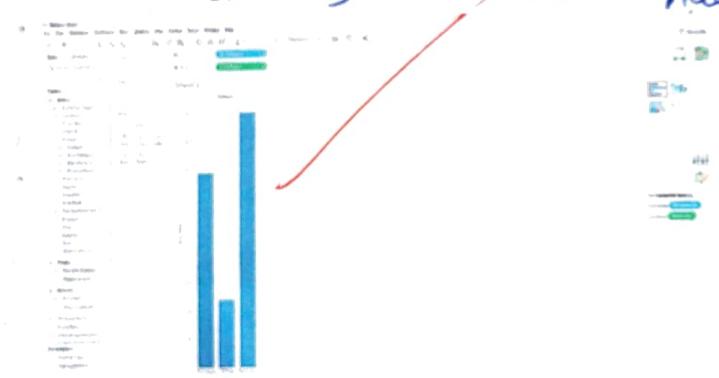
Table calculations:- Table calculations allow you to transform values at the level of details of the visualization only.

→ **Sum and Avg ~~etc~~ FUNCTIONS**:-

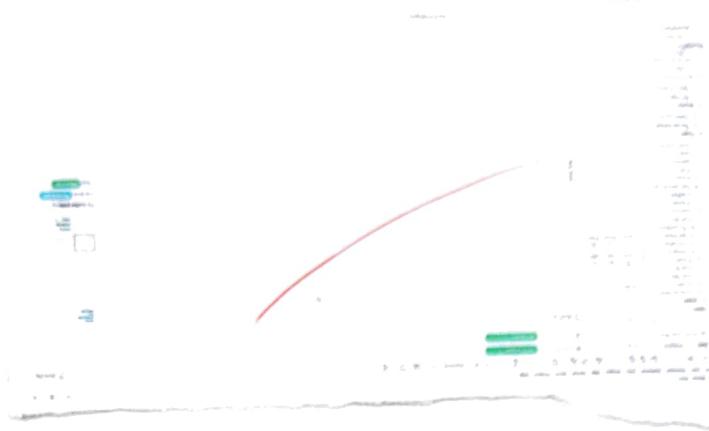
→ **Sum Function**: The sum function in Tableau calculates the total sum of numeric field. You can use it to find the sum of values in a column or as part of a more complex calculations. To use sum simply drag and drop a numeric field into the 'Sum' , or you can create a calculated field using the sum function.



Avg (average) function: - The AVG function calculates the avg (mean) value of a numeric field, like sum. You can use it by dragging numeric field into the "AVG" shelf or creating a calculated field with this function.



Aggregate Function: Tableau provides a range of aggregate functions that allows you to perform calculations on group of data. Common aggregate function include sum, avg, count, min (minimum value) and max. These functions are useful when you want to analyse data at different levels of granularity.

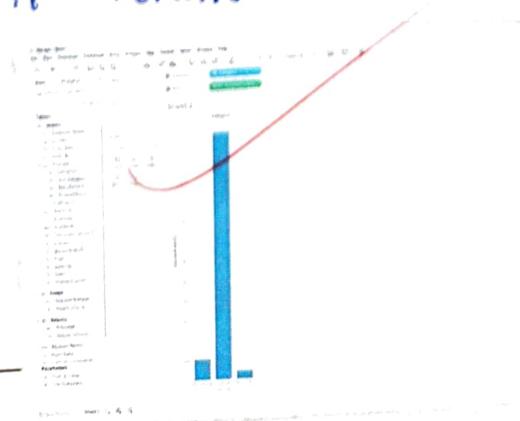


Create a calculated field:

- In tableau , select analysis > 'Create calculated field'.
- In the calculation Editor that opens , do the following -
- Enter a name for the calculated field . In this example , the field is called , discount ratio.
- Enter a formula : This example was the following formula :

$=IF([Sales] >= 0, [Discount]/[Sales]), 0)$

This formula checks if sales is not equal to zero. If true , it returns the discount ratio (Discount / Sales) . If false , it returns zero



Experiment-9

Date: _____

AIM

Applying new data calculations to visualizations, formatting visualizations, formatting tools and menus, formatting specific parts of the view.

PROCEDURE

I. Applying new data calculations to visualizations

Step-1: Build the view

- a. From Dimensions, drag Required field to the Columns shelf.
- b. From Dimensions, drag Required field to the Rows shelf.
- c. On the Rows shelf, click the plus icon (+) on the Category field to drill-down Subcategory.

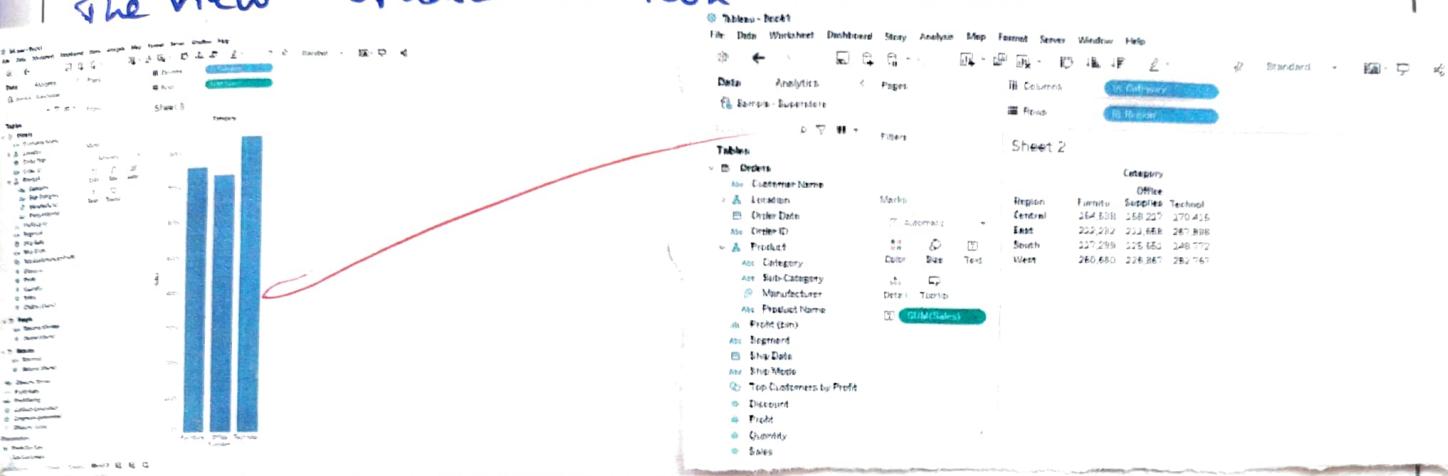
Step-2: Add the calculated field to the view

From Measures, drag Required Field to Color on the Marks card.

On the Rows shelf, right-click select required field and
select Measure (Sum) > Average.

SOURCE CODE

- From Dimensions, drag Category to the rows shelf.
- On the rows shelf, click plus icon (+) on category field to drill-down to subcategory. The view updates to look like this:



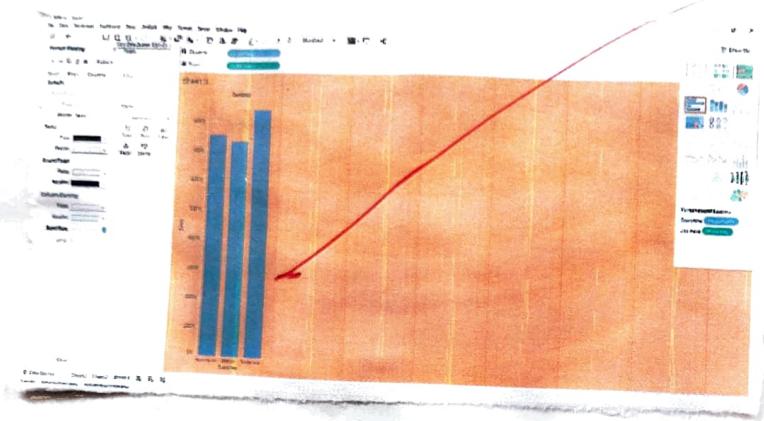
Step 2: Add the calculated field to the view. From measures drag discount ratio to colors on the marks card. The view update to highlight table.



Filter with calculated fields:

Create filters using calculated fields to control which data points are displayed in your visualization.

You can use calculated fields to filter by specific criteria, such as a calculated range or a custom ranking.



II. Formatting Visualizations, Formatting Tools and Menus

Step-1: Display a worksheet or dashboard.

Step-2: From the Format menu, choose the part of the view that you want to format, such as Font, Borders, or Filters.

Step-3: To format interactive controls go to **Format > Worksheet > Interactive Controls**.

Step-4: To control the background color of the worksheet, pane, and headers.

to go **Format > Worksheet > Shading**.

Step-5: To format Filters and Sets, formatting by either going into **Format > Filters and Sets**

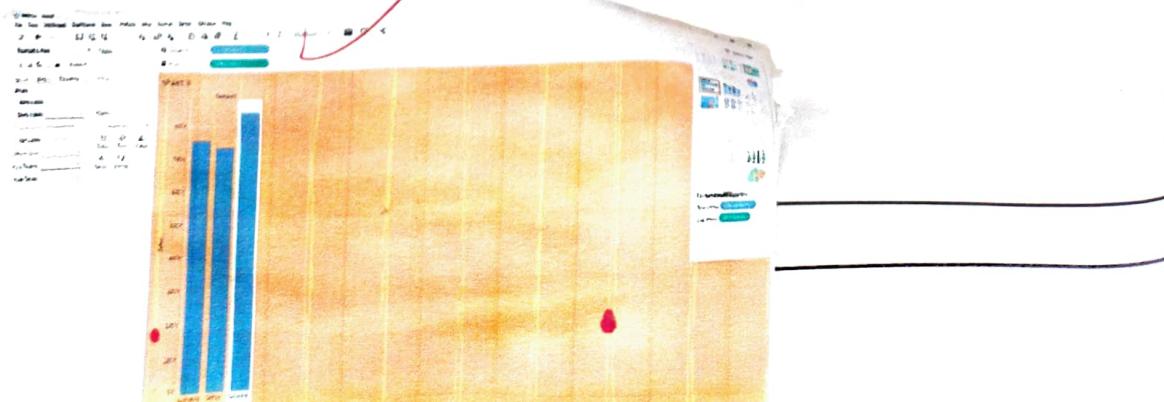
Step-6: To access highlighter formatting by either going into **Format > Highlighters**

SOURCE CODE

Formatting Tools and Menus

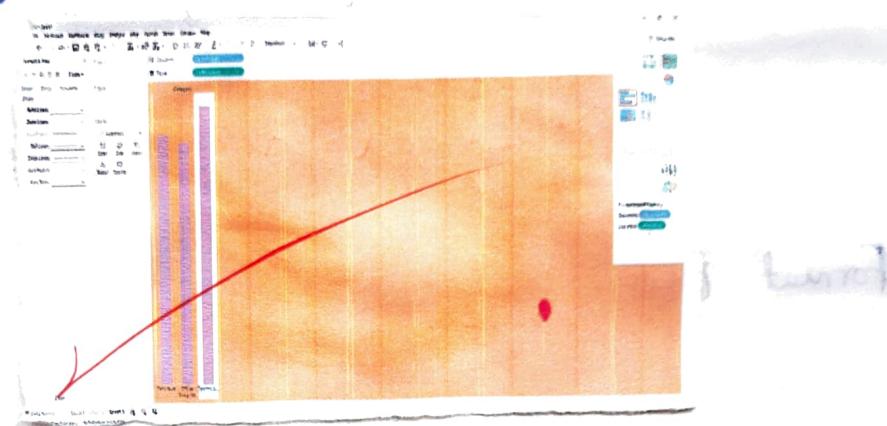
Tableau provides several formatting tools and means to help you refine the appearance of your visualizations.

- 1) Format menu:- The format menu at the top of the tableau interface provides access to various formatting options . including fonts, styles, shading, borders, more.
- 2) Worksheet menu:- In this menu , you'll find option to format the entire worksheet, including back-ground , border & worksheet.
- 3) Dashboard menu: If you're working with dashboard , this menu allows you.

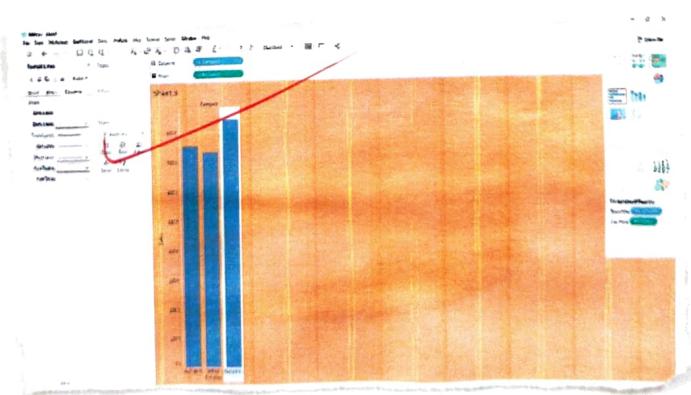


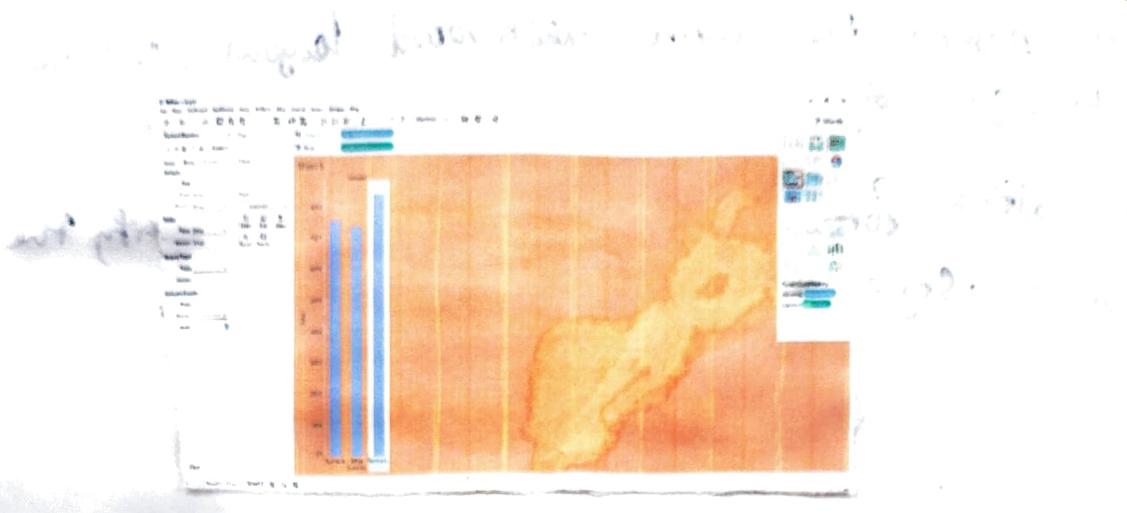
To format the entire dashboard layout, including background, size and title.

Formatting fonts: For a view, you can specify the font, style, size and color to either:



Format borders: Borders are the lines that surround and header in a view you can specify the border style width and colour for the cell and header areas.





Format Lines :-

You can control the appearance of the lines that one parts of the view such as grid and zero lines as well as lines that helps that you inspect data, as trendlines, reference lines and drop lines.

Eg:- One can set trend lines and use a red color and an increased thickness.

III. Formatting Specific Parts of The View:

DESCRIPTION

Every view has a table in some form, which may include rows, columns, headers, axes, panes, cells, and marks. Views can optionally include tooltips, titles, captions, field labels, and legends.

PROCEDURE

Step-1: To show and unshow headers:

- a. Right-click (control-click on Mac) the headers in the view and select Show Header.
- b. Select the field in the view whose headers you want to show and select Show Header on the field menu.

Step-2: To show and unshow axis:

- a. Right-click (control-click on MOUSE) the axis in the view and select Show Header to clear the check mark next to this option.
- b. Right-click (control-click on MOUSE) the measure in the view whose axis you want to show and select Show Header on the field menu.

Step-3: To Disable tooltip commands:

- a. If you don't want users to be able to access tooltip commands, you can disable them. Click Tooltip on the Marks card or select **Worksheet > Tooltip**.
- b. In the Edit Tooltip dialog box, clear the Include command buttons check box

Step-4: To show or hide titles in a worksheet

- a. From the toolbar menu, click **Worksheet > Show Title**.
- b. On the toolbar, click the drop-down arrow on the Show/Hide Cards button and select **Title from the context menu**.

Toggle the check mark on or off to show or hide the title.

SOURCE CODE

Headers and titles :-

Format headers, titles and subtitles for clarity and consistency. Use the format pane or the format menu to adjust text formatting, alignment and spacing.

Experiment-10

Date: _____

AIM

Editing and formatting axes, manipulating data in tableau data, pivoting tableau data.

PROCEDURE

I. Editing and Formatting Axes

Step-1: Double-click an axis to open the Edit Axis dialog box and change the axis configuration and formatting.

Step-2: To select the marks associated with the axis, right-click the axis and select Marks.

Step-3: To hide an axis

Right-click (control-click on windows) the axis in the view, and then clear the check mark next to the Show Header option.

Step-4: Try remaining options of axis

SOURCE CODE

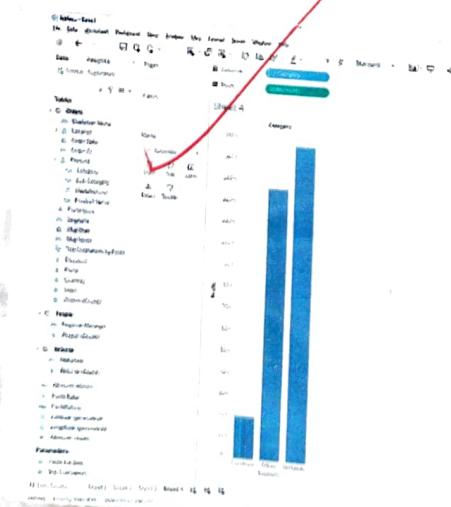
Change on axis range:

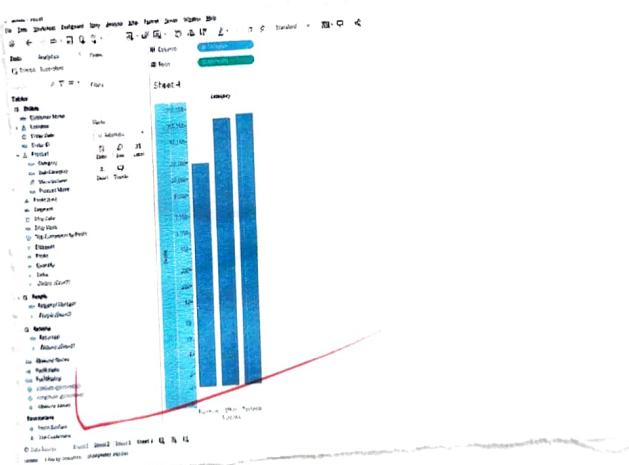
You can limit the axis range to focus the view on certain data points.

To change an axis range:

- 1) Double click the axis that you want to edit. You can also right click the axis, and then select "Edit axis".
- 2) In the Edit Axis dialog box, select one of the following options:
 - Automatic : Automatically bases the axis range on the data used in the view.
 - Uniform axis range for all rows or columns : set the axis range uniformly to the maximum data range for all the panels in the view.
 - Independent axis ranges for each row or column : make the axis range independent for each pane in the view.
 - Fixed : specifies to start and end the axis at the specific value.
- 3) You can also specify where to include zero. When you clear the include zero check box, the axis range adjust to show the range of values.

Let's check it.





II. Manipulating Data in Tableau Data, Pivoting Tableau Data

DESCRIPTION

Multiple data manipulation operations can be carried out to make data fit for visualization in the data preview pane or the data pane in the worksheet view. The data manipulation options vary by the data type of the columns. For example, the split option only comes up for string datatype columns, the create bins option only comes up for numeric datatype columns, etc.

All data manipulation is done at the metadata level, which is limited to Tableau only.

The changes are not reflected in the original data file or data source.

PROCEDURE

Step-1: The data type can be changed by clicking on the data type symbol in the column header

Step-2: Right-click on the column header and select hide to hide an unwanted column from the data view

Step-3: To create a group for a column, right click on its header and choose create group.

Next, a create group window pops up in which you can select multiple values.

Then click on Group

Step-4: Columns consisting of string values can be easily split into multiple columns using the split or custom split options

Step-5: Pivoting data with Tableau Desktop

- a. Once you have connected your data, on the data source page.
- b. Select the data fields you would like to pivot.
- c. On the drop-down menu of one of the fields select pivot.
- d. Executing this, we've a view with products packed on one column and values on another column

SOURCE CODE

Description:-

multiple data manipulation operations can be carried out to make data fit for visualization in the data preview pane on the data pane in the worksheet view. The data manipulation options vary by the datatype of the columns. For example, the split option only comes up for the string datatype columns. The create bins options only come up for the numeric datatype columns etc.

All the data manipulations is done ~~at~~ at the metadata level which is limited to table only. The changes are not reflected in the original data file or data source.

Procedure :-

Step 1: The datatype can be changed by clicking on datatype symbol in the column header.

Step 2: Right click on the column header and select hide to hide an unwanted column from the data view.

Step 3: To create a group for a column, right click on it's header and choose create group.

Step 4: A column consisting of strings of values can be easily split into multiple columns using the split or custom split options.

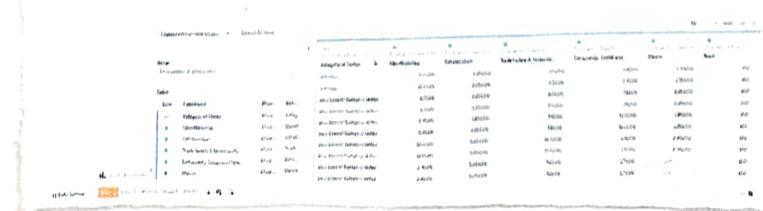
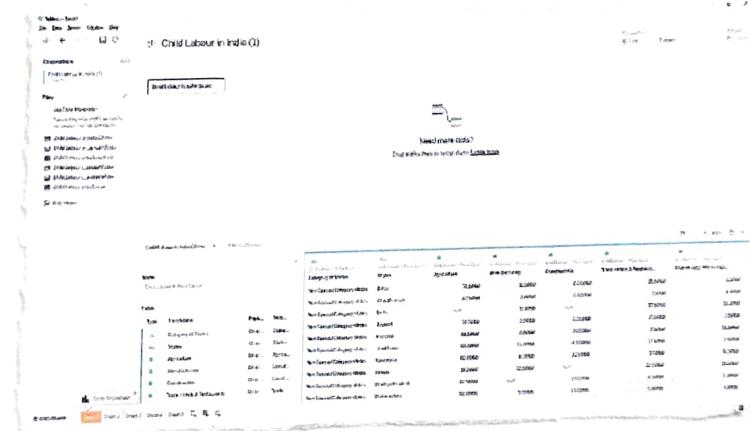
Steps:- Pivoting data with tableau desktop

a) Once you have connected your data, on the last a source page.

② Select the data fields you would like to pivot

b) On the drop down menu of one of the fields
prof.

d) Executing this, we've a views with products
in one column and values on another
column.



Experiment-11

Date: 5/11/24

AIM

Structuring the data, sorting and filtering tableau data, pivoting tableau data.

PROCEDURE

Step-1: To create a filter, drag a field directly from the Data pane to the Filters shelf.

Step-2: To filter entire rows or columns of data from your view, select the header in the view on the tooltip that appears, select to Exclude or Keep Only the selected data.

Step-3: Table calculation filter to apply to the totals, you can select Apply to totals in the drop-down menu for that filter (on the Filters shelf). This option lets you decide when a table calculation filter should be applied to totals.

Step-4: To select a filter card mode, in the view, click the drop-down menu on the filter card and then select a mode from the list.

Step-5: To sort specific field in viz

Right-click (Windows) or control-click (Mac) the field you want to sort, and select Sort. Select a Sort By option, and configure its behavior

Step-6: To remove the sort on a specific field, right click to open the menu and select Clear Sort.

SOURCE CODE

Structuring the data:

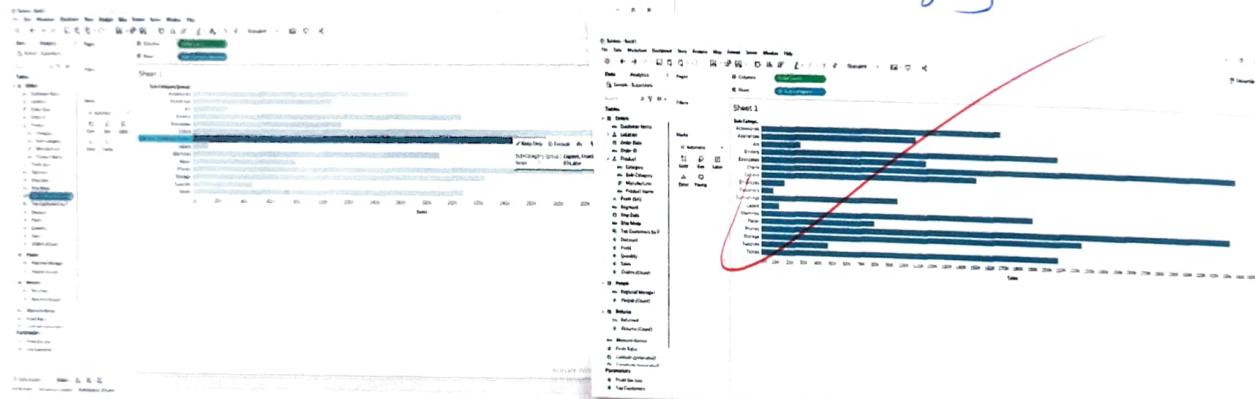
- 1) Groups
- 2) sets
- 3) Hierarchy
- 4) Bins
- 5) Sort

Grouping :- We can perform grouping in many ways

like :-

- Group by header
- Group by data window
- Group by visualization

Group by header's col : sum(sales)
 rows : subcategory

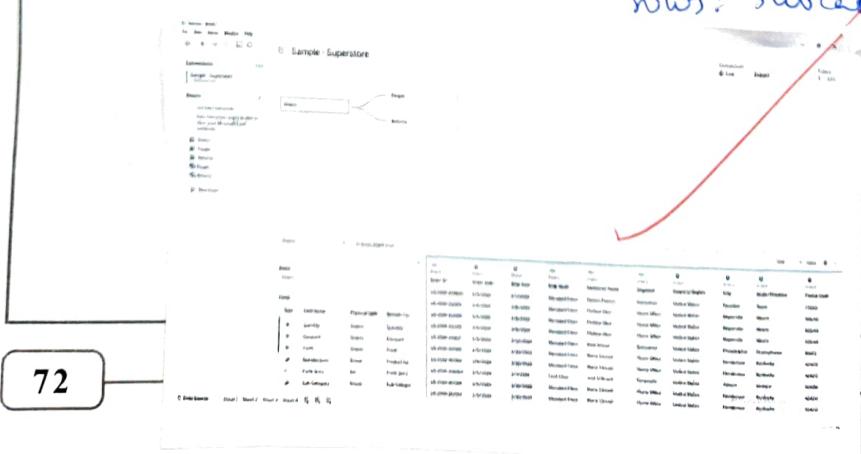


Select the required columns for grouping and

Click on the grouping symbol.

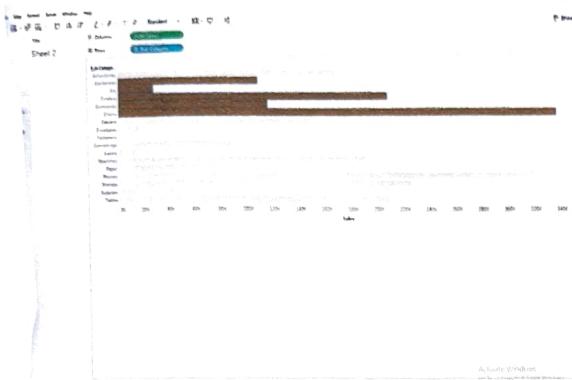
Group by data window:-

col : sum(sales)
 rows : subcategory



Under the data window select any dimensions field and click on create then select group.

Group by visualization: col: sales
or subcategory

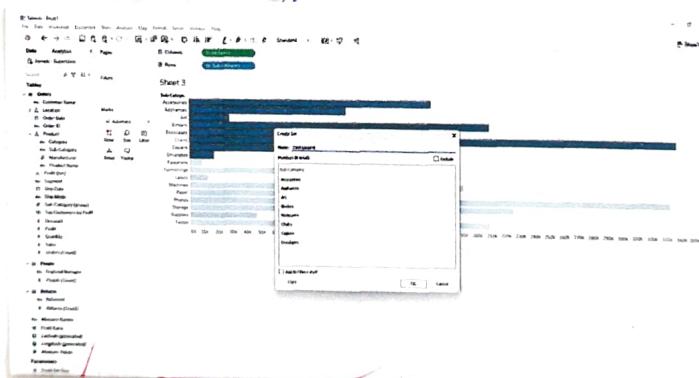


groups and click on group icon

Select required columns for group and click on group

- Steps:- Set with header
→ Set with dataview
→ Set with visualization

Same procedure like grouping



Hierarchy: We can create hierarchy - select one field from data window drag and drop field on another field.



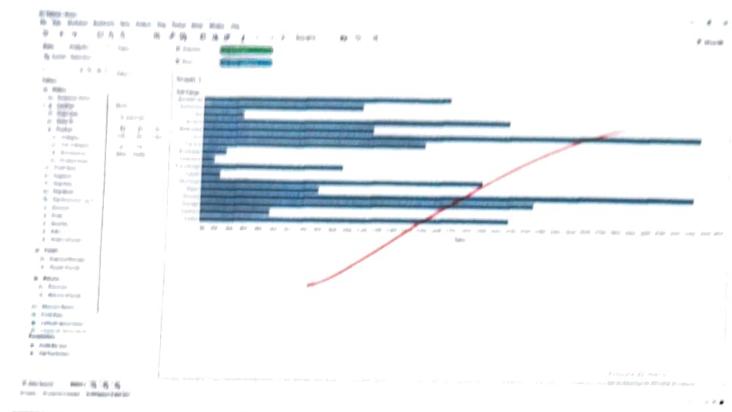
Bins:- We can create bins on measures. We can't create on dimensions.

→ Bin field will be used in histogram charts.

→ Bins are useful to create Range of data.

Sort:- We can perform sorting in many ways.

- 1) Quick sort
- 2) Sort from tool bar
- 3) Sort from pills



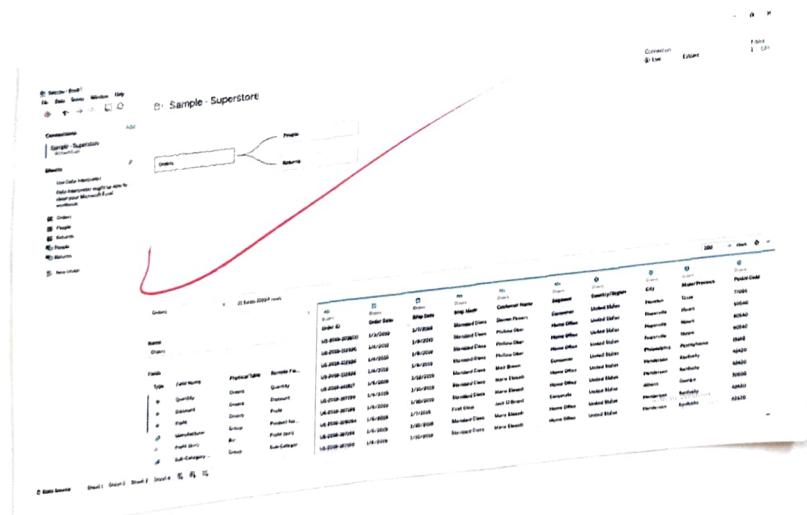
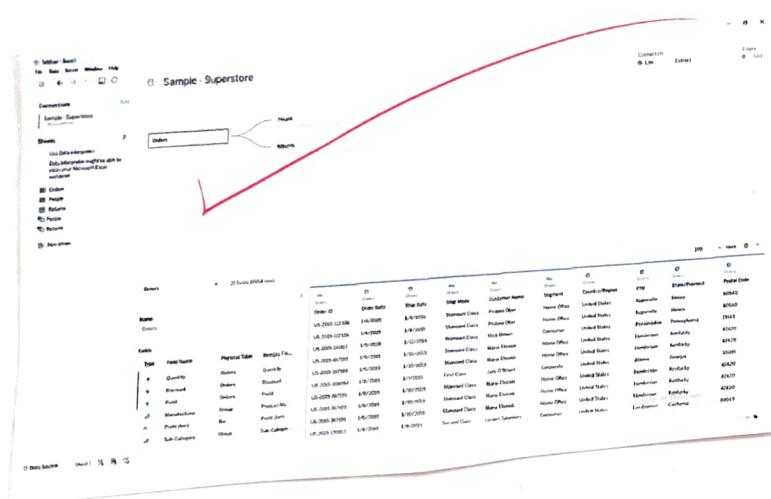
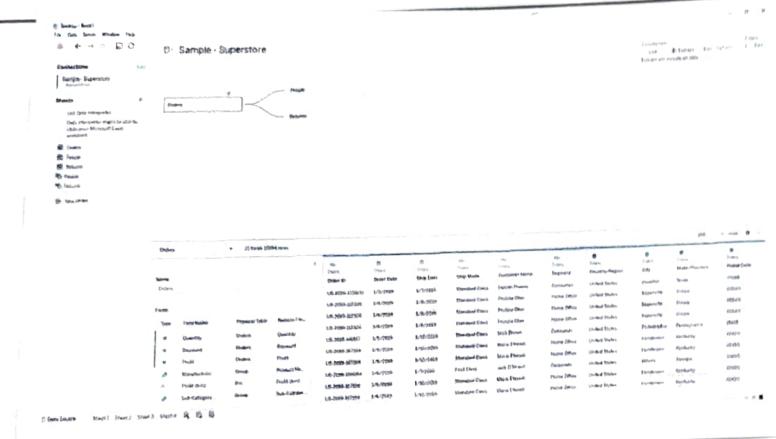
Filtering:- We can perform filtering in ~~many~~^{many} ways through dimensions measures data sources, etc.

You can place subcategory

Types of filters:-

- 1) Data source filters
- 2) Extract filters
- 3) Dimensions filters
- 4) Measure filters
- 5) Context + filters
- 6) Table calculations filter*

OUTPUT



Experiment-12

Date: 22/11/24

AIM

Advanced visualization tools: using filters, using the detail panel, using the size panels, customizing filters, using and customizing tooltips, formatting data with colors.

PROCEDURE

Step-1: To assign a color to marks in the view, do one of the following:

- a. On the Marks card, click Color, and then select a color from the menu.
- b. From the Data pane, drag a field to Color on the Marks card.

Step-2: To change the color palette or customize how color is applied to your marks:

- a. On the Marks card, click **Color > Edit Colors**.

Step-3: To change the size of marks in the view, do one of the following:

- a. On the Marks card, click Size, and then move the slider to the left or right.
- b. From the **Data** pane, drag a field to **Size** on the Marks card.

Step-4: To edit the size of marks, or change how size is being applied to marks in the view:

- a. On the Size legend card (which appears when you add a field to Size on the Marks card), click the drop-down arrow in the right-hand corner and select **Edit Sizes**.
- b. In the **Edit Sizes** dialog box that appears, make your changes and then click **OK**.

Step-5: To add a tooltip

- Drag a field to Tooltip on the Marks card.
- Click Tooltip on the Marks card to open the Edit Tooltip dialog box, where you can add text, rearrange the tooltip contents, and insert more fields.

SOURCE CODE

Advanced Visualization Tools:-

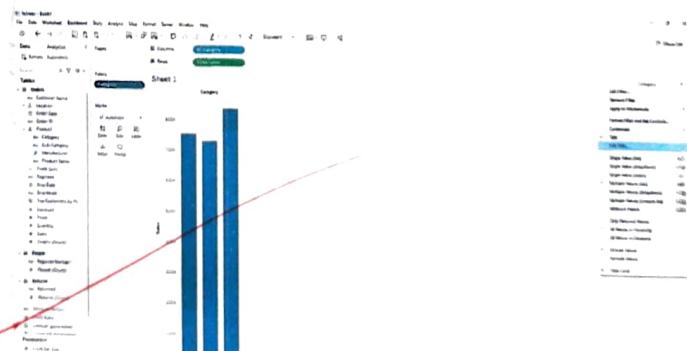
- Interactive Filter
- Advanced interactive filter
- Action filters
- customizing filters
- Colours
- Size
- Labels
- details
- Tooltips
- size

Interactive filters :-

Col: Sum (Sales)

Rows: Category , Subcategory

Apply filters on category & click on show filter

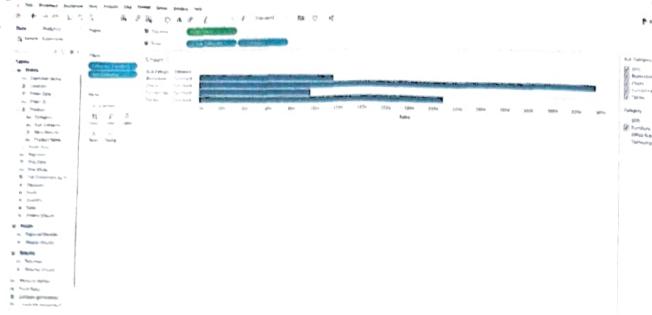


Advanced interactive filter :-

Col: Sum (Sales)

Rows: Category , subcategory

Apply filters on category & subcategory columns



So suppose if you select furniture in the category, that furniture related sub category will be displayed under the subcategory filter not after subcategory.

Action filters:-

We will take 2 worksheets

→ First worksheet in category wise Sales

→ Second worksheet is category and subcategory wise.

worksheet 1

cols: Category

rows: sum (Sales)

Here, source is worksheet 1

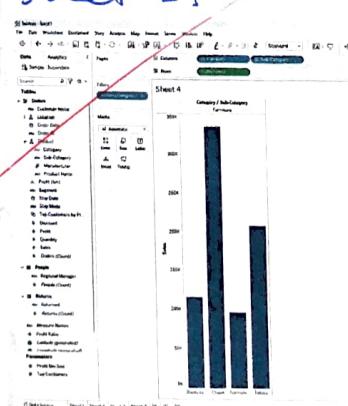
target is worksheet 2

worksheet 2

cols: category, sub category

rows: sum (Sales)

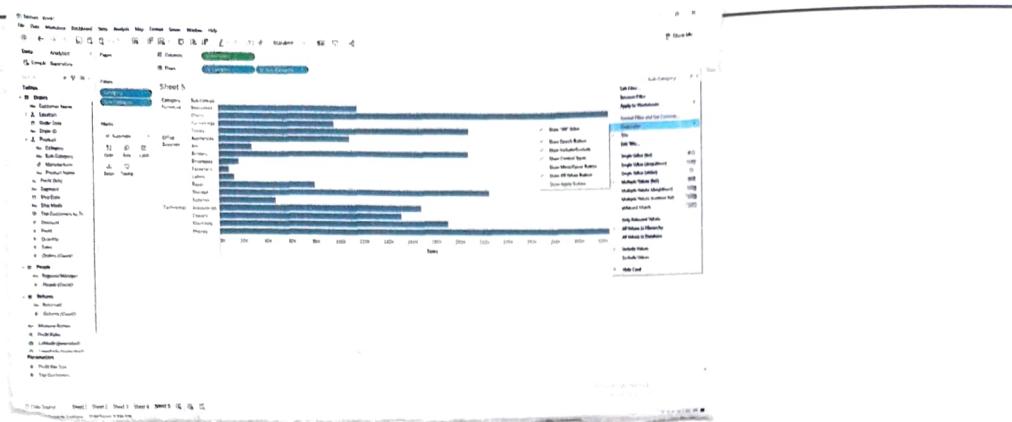
If you Select anyone category on sheet1, the related
sheet 2.



cols: sum (Sales)

rows: Category → sub category

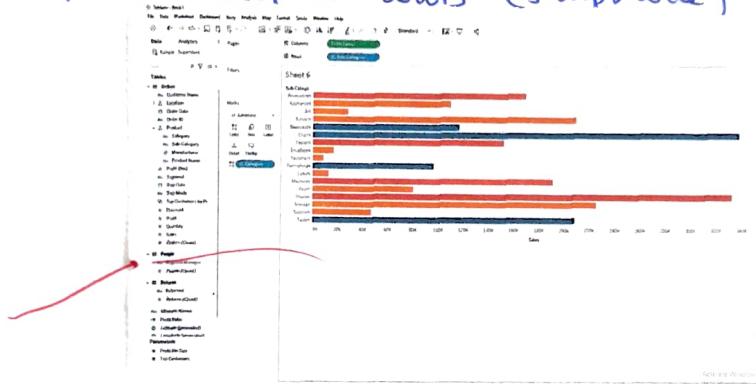
Apply filters on category column and click
on Show filters.



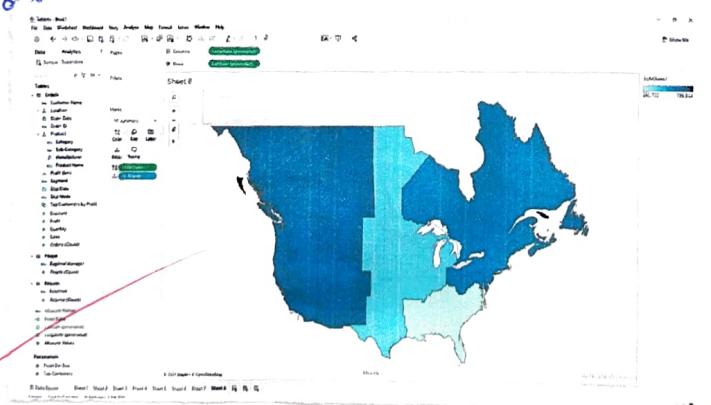
Colors

Colors with bar :- Col: sum (Sales)
row : sub-category

Drag and drop category in colors (8 levels mode)

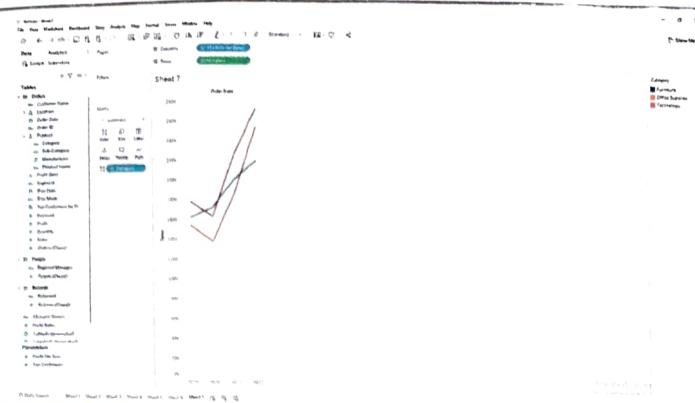


Color with map; Double click on region
drag and drop sales in colors



~~Color~~ ~~With~~ line: Col: order (date)
rows: sum(sales)

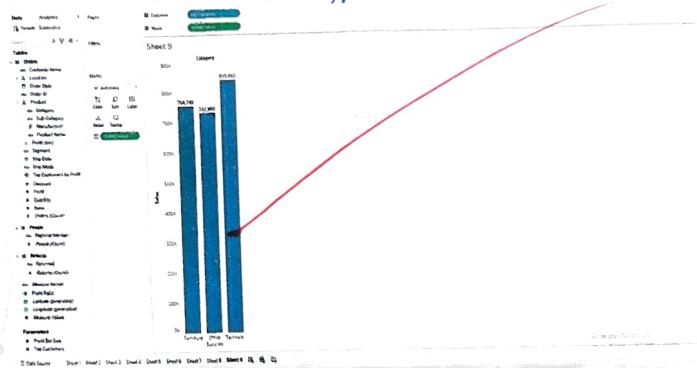
OUTPUT



Size:

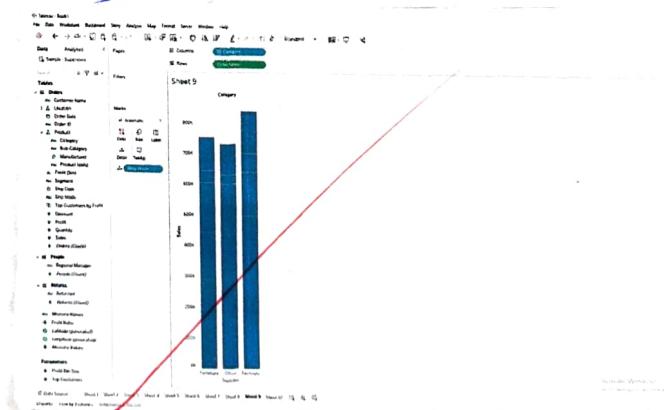
size with bar : col: Category
rows: sum(sales)

Click on size under marks card you can play.



size with line : col: & year (order date)
rows: sum (sales)

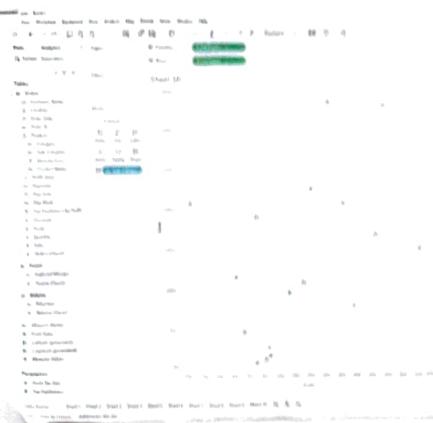
Click on size and play



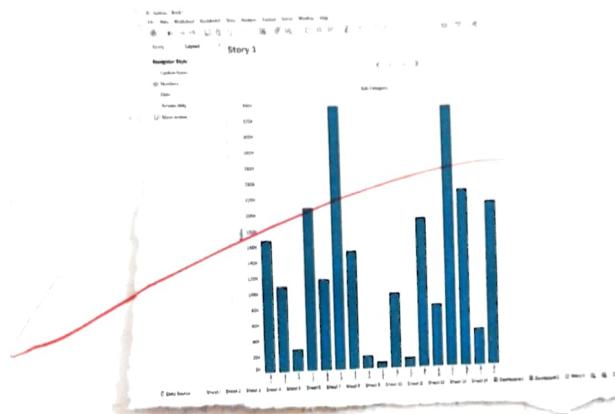
Label :

label with bar: col: Category
rows: sum (sales)

drag and drop Sales in sales > click on labels show mark



details and color both are similar, but only the difference is color showing difference in colors whereas details showing the difference in different sections. Drag and drop ship mode in details.



Tool Tip :- In visualization if you place cursor on any place you will see the tool tip we can edit tool tip. If you drag and drop column in the tooltip we can see those values under tooltip in visualization.

Shapes = - If you take two measures then only you can see the shapes under marks card
col : sum (profit); rows : sum (sales)
~~drag and drop subcategory on shapes.~~



Step-3: To publish to Tableau Server or Tableau Cloud, your server or site administrator must grant you the following capabilities:

- A site role of Creator (formerly Publisher) on the site you're publishing to.
- View and Save capabilities set to Allowed on the project into which you publish.

Step-4: In Tableau Desktop, open the workbook you want to publish.

- Select Server > Publish Workbook.
- If the Publish Workbook option does not appear on the Server menu, make sure a worksheet or dashboard tab is active (not the Data Source tab).
- If necessary, sign in to a server. For Tableau Cloud, enter <https://online.tableau.com>.
- In the Publish Workbook dialog box, select the project, enter a name for the workbook, and add search tags.

SOURCE CODE

Dashboard : Collections of multiple worksheets (charts, graphs) and other displays arranged together on a single screen.
→ Dashboard are used to analyse data efficiently
⇒ Sales of states → Sales by month → Sales by category



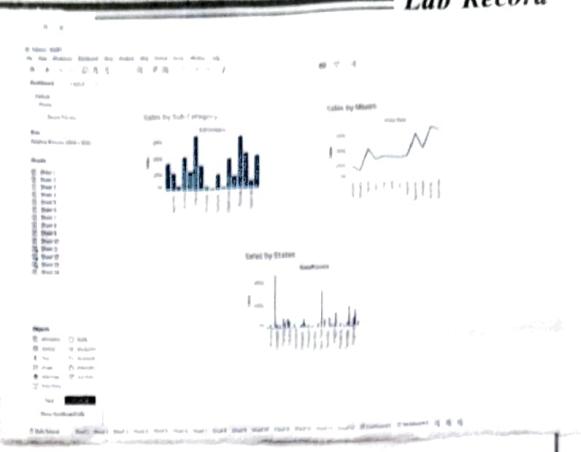
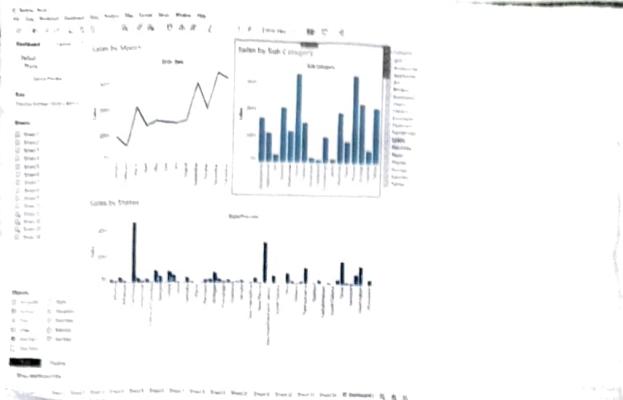
Create dashboard with above worksheets

drag & drop those sheets on dashboard

Dashboard Format :

If you select fitted then drag and drop worksheet on dashboard we are getting big images

If you select floating we will get small images, we will give proper alignment with the help of floating.

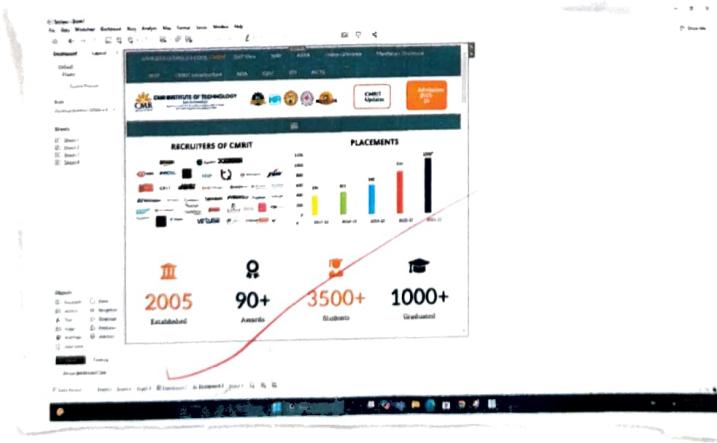


Dashboard Auto Adder:

- Select vertical container
- Double click on sales by state
- Double click on sales by sub category



Dashboard title change by dynamically



Dashboard
homepage

Drag and drop webpage on dashboard
→ Enter the URL

Go to Dashboard → actions → addaction → click on URL

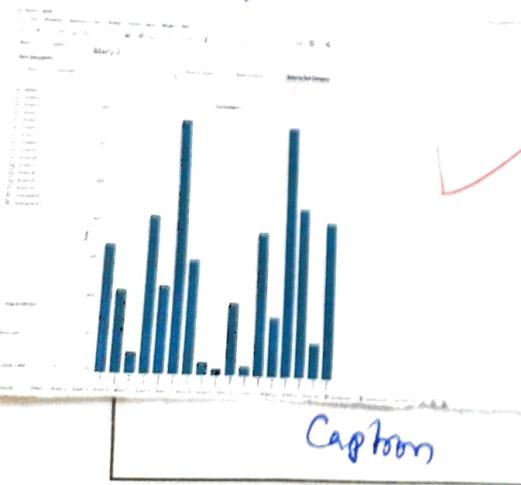
Tableau Device Manager : We will be able to design dashboard for different layouts → desktop → tablet → phone
 → Drag and drop worksheet on dashboard
 → Size change automatically
 → Click on device preview
 Now you can add it and customize different device types



Phone

Story point : - Sequential way to visualization analyzed data.

- Select Story for the Workbook
- Drag and drop each sheet on the caption.
- You can edit caption
- You can add the sheets whenever you want beginning, middle or ending.
- You can give title to the story.
- You can change navigation style.
- Caption boxes → Numbers → date → arrows



Captions

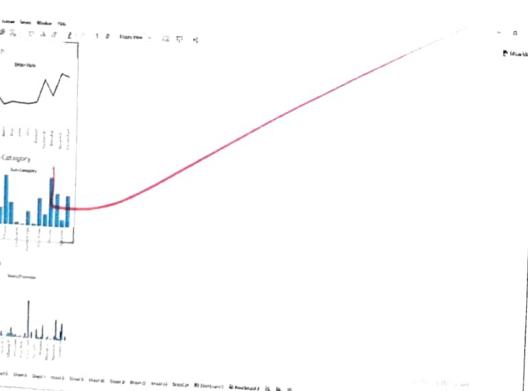
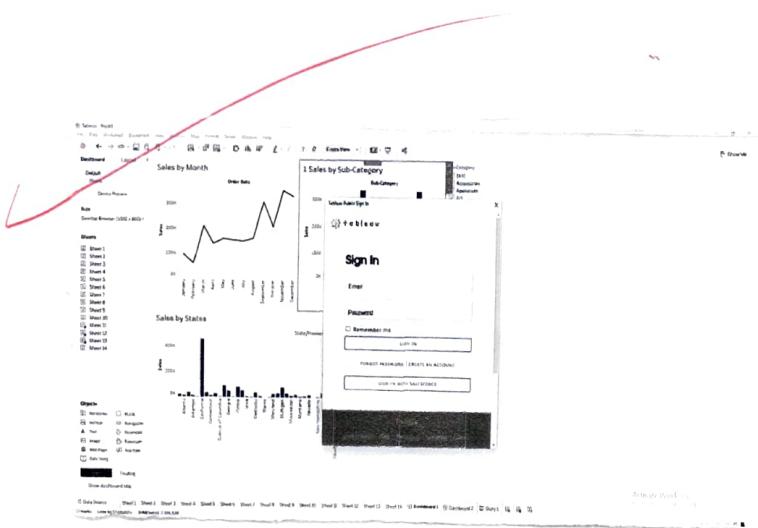


Numbers

OUTPUT

Publishing data:

- Proto server → Click on tableau-public → save to tableau public
- keep data source · connections extract.
- you need tableau public credentials then only you can publish.
- once publish the data · we can it share · it · and download it.
- you can see the backend code.



Experiment-14

Date: 22/11/24

AIM

Creating custom charts, cyclical data and circular area charts, dual axis charts.

PROCEDURE

Step-1: To create dual axis charts use the following steps:

- Make a graph for one of the measures
- Drag the second measure onto the opposite axis
- Create a dual-axis combination chart by changing one of the mark types

Step-2: To create the Circular area chart:

- First drag and drop required fields from data pane to both rows and columns
- On the Marks card, click the Mark Type drop-down and select **circular area chart**.

Step-3: To create custom funnel chart

- Add Measures to the Rows Section
- Select Dimensions in the Marks Section
- Convert Standard View to Entire View
- Add Labels in the Marks Section

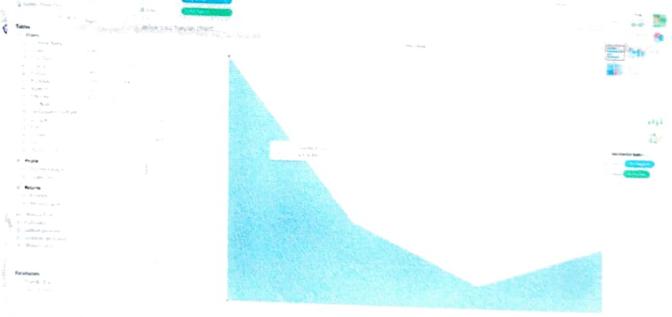
SOURCE CODE

Funnel chart: double click on sales. Drag and drop segment on colors. Drag and drop sales on size. View changed from Standard to entire view. Drag and drop segment and sales on labels. Apply table calculation of sales (% of total).



Advanced funnel chart: Double click on sales. Double click on shipmode apply sweeping sort descending order col: sum (sales) rows: Shipmode

Under marks menu change to area chart
Change view from size



Lollipop chart: double click on shipmode. Double click on sales take one more sales. We will get multiple axis.

Convert multiple axis to dual axis: Sales → click on dual axis.

Now the entire view shown as circle. Under marks card you can see

- All
- Sum(Sales) → Sum(Sales): Decrease the size
- Sum(Sales)(2) → Sum(Sales)(2): Increase

→ Select Sum(Sales) change to bar (marks card)

→ Select Sum(Sales)(2) change & to circle.

Drag and drop profit in Sum(Sales)(2) & colors.

Donut chart

Donut chart :- Under marks card change Automatic to pie now we can drag and drop Sales on angle / category on colors. Now change view from standard to entire view.

- AVU → select AVH(0)(2), remove all options. Drag and drop
- AVH(0) drag Sales on labels (AVH(0)(2)) Change color to
- & AVG(0)2 White Increase the size and number size.
- Convert multiple axis to dual axis.



Butterfly chart :- Drag - drop Sales on column shelf, again drag & drop Sales. Drag & Drop sub categories to row shelf.

Create zero axis AVH(0) one column shelf.

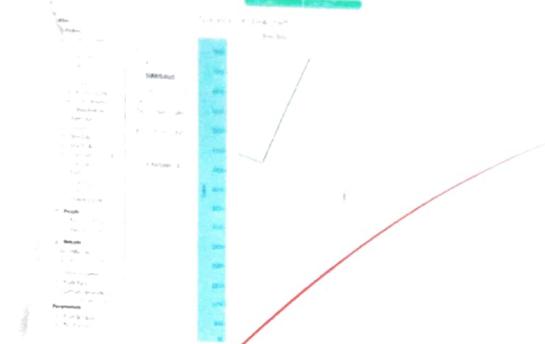
Sum(Sales) → Right Click → edit axis → Reserve

- Sum(Sales) → Drag and drop sub categories on ~~color~~ and sales on label in both sum(Sales)
- AVH(0)
- Sum(Sales) sum(Sales)(2) AVH(0) changed from automatic to next.



Dual Axis

With line chart : col: Year(orderdate), rows: sum(sales) sum(profit). Drag and drop Sales on row shelf. Drag and drop profit on rows and years on columns. Convert multiple axes to dual axis. Synchronize the axis Right click on profit axis → Select Synchronize axis.



With Bar chart :- Drag and drop cols: Subcategory rows: Sales & Subcategory

Convert multiple axes to dual axis.

Right click on profit axis → Select Synchronize Axis change Automatic to bar in mark card menu.



~~Cyclical data~~ Cyclical data: Cyclical data means waveform repeats intervals Ex: Quarter, month, week, day

Cyclical data with Linechart: If you want to see trends and compares different cycles use linechart

cols: Day(order_date)

rows: Sum(profit)

OUTPUT