

# SUPERSTORE ANALYSIS



PRESENTED BY-  
**KRISH GARG (24042)**

# CONTENT

*01*

ABOUT

*02*

APPLICATION OF R

*03*

RAW DATA

*04*

KEY INSIGHTS AND  
RECOMMENDATIONS

# ABOUT

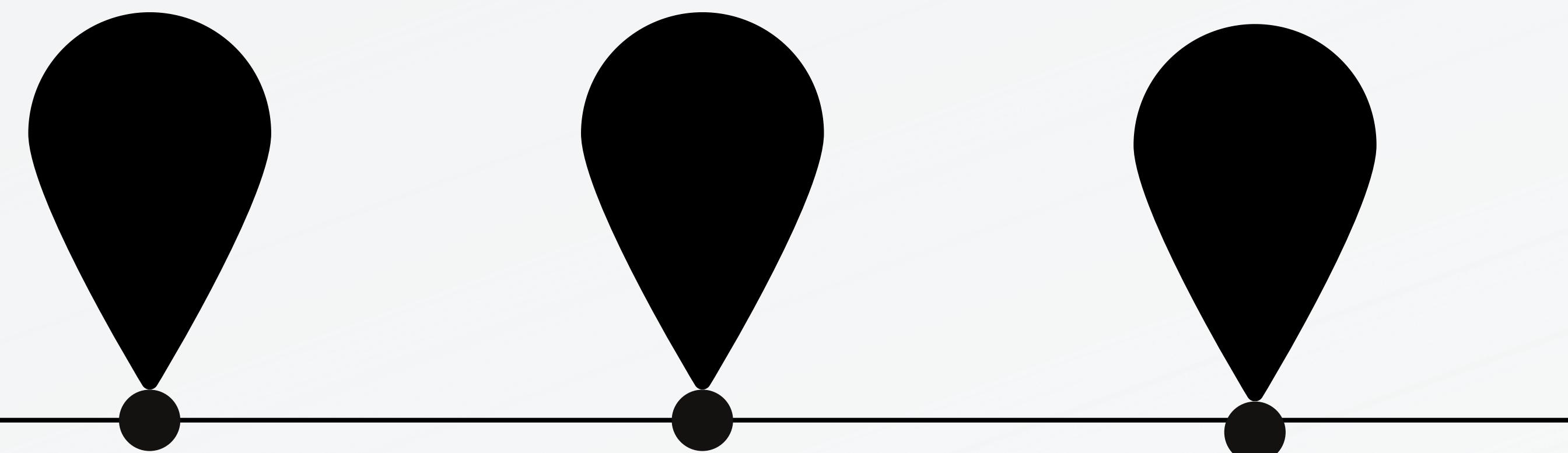


R is a programming language and free software environment used primarily for **statistical computing, data analysis, and graphical representation**



R is an **interpreted programming language** (a scripting language), meaning your code does not have to be compiled before running it

# APPLICATION OF R

- 
- **Data science and analytics:** R is used to analyze and visualize data, and is a key tool for data scientists.
  - **Academic research:** Researchers use R to perform statistical analysis and create reproducible research.
  - **Finance:** Financial analysts use R to build models and analyze financial data. For example, banks use R to create credit risk models, conduct risk analysis, and detect fraud.



# RAW SALES DATA OF SUPERSTORE

Row	Order ID	Order Date	Ship Date	Ship Mode	Customer	Customer Name	Segment	Country	City	State	Postal Code	Region	Product ID	Category	Sub-Category
1	CA-2016-152156	08-11-2016	11-11-2016	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420	South	FUR-BO-10001798	Furniture	Bookcases
2	CA-2016-152156	08-11-2016	11-11-2016	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420	South	FUR-CH-10000454	Furniture	Chairs
3	CA-2016-138688	12-06-2016	16-06-2016	Second Class	DV-13045	Darrin Van Huff	Corporate	United States	Los Angeles	California	90036	West	OFF-LA-10000240	Office Supplies	Labels
4	US-2015-108966	11-10-2015	18-10-2015	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	Florida	33311	South	FUR-TA-10000577	Furniture	Tables
5	US-2015-108966	11-10-2015	18-10-2015	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	Florida	33311	South	OFF-ST-10000760	Office Supplies	Storage
6	CA-2014-115812	09-06-2014	14-06-2014	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	FUR-FU-10001487	Furniture	Furnishings
7	CA-2014-115812	09-06-2014	14-06-2014	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	OFF-AR-10002833	Office Supplies	Art
8	CA-2014-115812	09-06-2014	14-06-2014	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	TEC-PH-10002275	Technology	Phones
9	CA-2014-115812	09-06-2014	14-06-2014	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	OFF-BI-10003910	Office Supplies	Binders
10	CA-2014-115812	09-06-2014	14-06-2014	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	OFF-AP-10002892	Office Supplies	Appliances
11	CA-2014-115812	09-06-2014	14-06-2014	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	FUR-TA-10001539	Furniture	Tables
12	CA-2014-115812	09-06-2014	14-06-2014	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	TEC-PH-10002033	Technology	Phones
13	CA-2017-114412	15-04-2017	20-04-2017	Standard Class	AA-10480	Andrew Allen	Consumer	United States	Concord	North Carolina	28027	South	OFF-PA-10002365	Office Supplies	Paper
14	CA-2016-161389	05-12-2016	10-12-2016	Standard Class	IM-15070	Irene Maddox	Consumer	United States	Seattle	Washington	98103	West	OFF-BI-10003656	Office Supplies	Binders
15	US-2015-118983	22-11-2015	26-11-2015	Standard Class	HP-14815	Harold Pawlan	Home Office	United States	Fort Worth	Texas	76106	Central	OFF-AP-10002311	Office Supplies	Appliances
16	US-2015-118983	22-11-2015	26-11-2015	Standard Class	HP-14815	Harold Pawlan	Home Office	United States	Fort Worth	Texas	76106	Central	OFF-BI-10000756	Office Supplies	Binders
17	CA-2014-105893	11-11-2014	18-11-2014	Standard Class	PK-19075	Pete Kriz	Consumer	United States	Madison	Wisconsin	53711	Central	OFF-ST-10004186	Office Supplies	Storage
18	CA-2014-167164	13-05-2014	15-05-2014	Second Class	AG-10270	Alejandro Grove	Consumer	United States	West Jordan	Utah	84084	West	OFF-ST-10000107	Office Supplies	Storage
19	CA-2014-143336	27-08-2014	01-09-2014	Second Class	ZD-21925	Zuschuss Donatelli	Consumer	United States	San Francisco	California	94109	West	OFF-AR-10003056	Office Supplies	Art
20	CA-2014-143336	27-08-2014	01-09-2014	Second Class	ZD-21925	Zuschuss Donatelli	Consumer	United States	San Francisco	California	94109	West	TEC-PH-10001949	Technology	Phones
21	CA-2014-143336	27-08-2014	01-09-2014	Second Class	ZD-21925	Zuschuss Donatelli	Consumer	United States	San Francisco	California	94109	West	OFF-BI-10002215	Office Supplies	Binders
22	CA-2016-137330	09-12-2016	13-12-2016	Standard Class	KB-16585	Ken Black	Corporate	United States	Fremont	Nebraska	68025	Central	OFF-AR-10000246	Office Supplies	Art
23	CA-2016-137330	09-12-2016	13-12-2016	Standard Class	KB-16585	Ken Black	Corporate	United States	Fremont	Nebraska	68025	Central	OFF-AP-10001492	Office Supplies	Appliances
24	US-2017-156909	16-07-2017	18-07-2017	Second Class	SF-20065	Sandra Flanagan	Consumer	United States	Philadelphia	Pennsylvania	19140	East	FUR-CH-10002774	Furniture	Chairs
25	CA-2015-106320	25-09-2015	30-09-2015	Standard Class	EB-13870	Emily Burns	Consumer	United States	Orem	Utah	84057	West	FUR-TA-10000577	Furniture	Tables
26	CA-2016-121755	16-01-2016	20-01-2016	Second Class	EH-13945	Eric Hoffmann	Consumer	United States	Los Angeles	California	90049	West	OFF-BI-10001634	Office Supplies	Binders
27	CA-2016-121755	16-01-2016	20-01-2016	Second Class	EH-13945	Eric Hoffmann	Consumer	United States	Los Angeles	California	90049	West	TEC-AC-10003027	Technology	Accessories
28	US-2015-150630	17-09-2015	21-09-2015	Standard Class	TB-21520	Tracy Blumstein	Consumer	United States	Philadelphia	Pennsylvania	19140	East	FUR-BO-10004834	Furniture	Bookcases
29	US-2015-150630	17-09-2015	21-09-2015	Standard Class	TB-21520	Tracy Blumstein	Consumer	United States	Philadelphia	Pennsylvania	19140	East	OFF-BI-10000474	Office Supplies	Binders
30	US-2015-150630	17-09-2015	21-09-2015	Standard Class	TB-21520	Tracy Blumstein	Consumer	United States	Philadelphia	Pennsylvania	19140	East	FUR-FU-10004848	Furniture	Furnishings
31	US-2015-150630	17-09-2015	21-09-2015	Standard Class	TB-21520	Tracy Blumstein	Consumer	United States	Philadelphia	Pennsylvania	19140	East	OFF-EN-10001509	Office Supplies	Envelopes
32	US-2015-150630	17-09-2015	21-09-2015	Standard Class	TB-21520	Tracy Blumstein	Consumer	United States	Philadelphia	Pennsylvania	19140	East	OFF-AR-10004042	Office Supplies	Art
33	US-2015-150630	17-09-2015	21-09-2015	Standard Class	TB-21520	Tracy Blumstein	Consumer	United States	Philadelphia	Pennsylvania	19140	East	OFF-BI-10001525	Office Supplies	Binders
34	US-2015-150630	17-09-2015	21-09-2015	Standard Class	TB-21520	Tracy Blumstein	Consumer	United States	Philadelphia	Pennsylvania	19140	East	OFF-AR-10001683	Office Supplies	Art
35	US-2015-150630	17-09-2015	21-09-2015	Standard Class	TB-21520	Tracy Blumstein	Consumer	United States	Philadelphia	Pennsylvania	19140	East	OFF-AR-1000249	Office Supplies	Paper
36	CA-2017-107727	19-10-2017	23-10-2017	Second Class	MA-17560	Matt Ahelman	Home Office	United States	Houston	Texas	77095	Central	OFF-PA-10000249	Office Supplies	Paper

R Superstore.R x

Source on Save | Run | Source

```
1 library("readxl")          # To Read the Sample Superstore Excel Workbook
2 Dataset=read_excel("C:/Users/HP/OneDrive/Desktop/MABS/Trimester-2/Business Analytics-2/Assignment and Project/Assignment-4/Superstore.xlsx")
3 library(dplyr)              # Using dplyr for Data Manipulation and Transformation
4
5 #To view the Dataset
6 View(Dataset)               # To view the complete Dataset
7
8 #To show the structure of Dataset
9 str(Dataset)                # To view the Structure of Dataset in Console Area
10
11 #To view the Top 20 rows of the Dataset
12 head(Dataset,20)            # To view top 20 rows of the Dataset in Console Area
13 View(head(Dataset,20))       # To view top 20 rows of the Dataset
14
15 #To view the bottom 6 rows of the Dataset
16 tail(Dataset)               # To view Bottom 6 rows of the Dataset in Console Area
17 View(tail(Dataset))         # To view Bottom 6 rows of the Dataset
18
19 #To show the statistical summary of the Dataset
20 summary(Dataset)            # To view Summary of the Dataset in Console Area
21 View(summary(Dataset))      # To view Summary of the Dataset
22
23 #To select specific columns
24 selected_data=Dataset%>%
25   select(`Ship Mode`,`Customer Name`)
26 View(selected_data)           # To View the Ship Mode and Customer Name in Dataset
27
28 #How to Rename the Ship Mode and Postal Code Column?
29 renamed_data=Dataset%>%
30   rename(
31     Mode='Ship Mode',
32     Pin_Code='Postal Code'
33   )
34 View(renamed_data)           # To View the Dataset after renaming the Ship Mode and Postal Code Columns
35
36 #Q.1) How can we view the top 6 rows of the renamed dataset?
37 View(head(renamed_data))     # To View the Top 6 rows of the Renamed Dataset
38
```

# HOW CAN WE VIEW THE TOP 6 ROWS OF THE RENAMED DATASET?

```
#How to Rename the Ship Mode and Postal Code Column?  
renamed_data=Dataset%>%  
  rename(  
    Mode='Ship Mode',  
    Pin_Code='Postal Code'  
)  
View(renamed_data)      # To View the Dataset after renaming the Ship Mode and Postal Code Columns
```

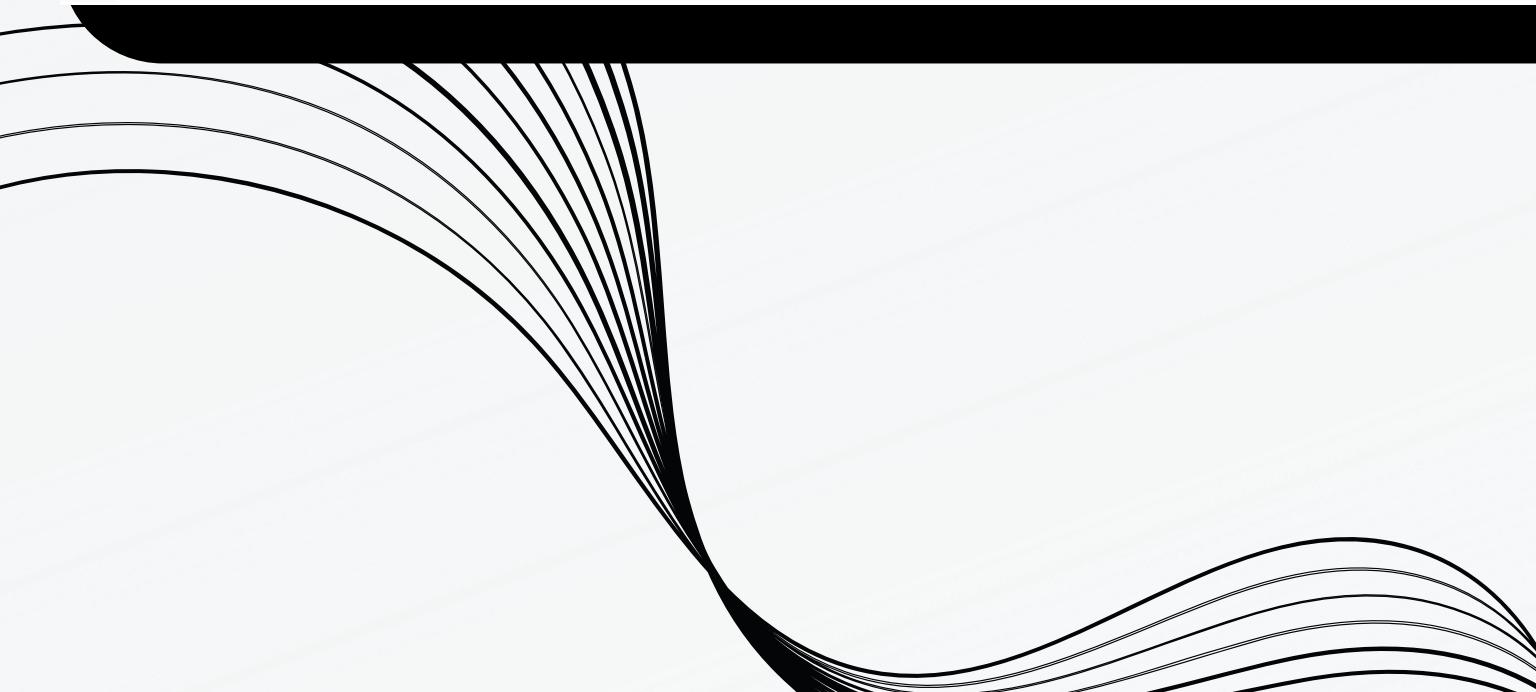
	Order ID	Order Date	Ship Date	Mode	Customer ID	Customer Name	Segment	Country	City	State	Pin_Code	Region	P
1	CA-2016-152156	2016-11-08	2016-11-11	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420	South	F
2	CA-2016-152156	2016-11-08	2016-11-11	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420	South	F
3	CA-2016-138688	2016-06-12	2016-06-16	Second Class	DV-13045	Darrin Van Huff	Corporate	United States	Los Angeles	California	90036	West	C
4	US-2015-108966	2015-10-11	2015-10-18	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	Florida	33311	South	F
5	US-2015-108966	2015-10-11	2015-10-18	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	Florida	33311	South	C
6	CA-2014-115812	2014-06-09	2014-06-14	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	F
7	CA-2014-115812	2014-06-09	2014-06-14	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	C
8	CA-2014-115812	2014-06-09	2014-06-14	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	T
9	CA-2014-115812	2014-06-09	2014-06-14	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	C
10	CA-2014-115812	2014-06-09	2014-06-14	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	C
11	CA-2014-115812	2014-06-09	2014-06-14	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	F
12	CA-2014-115812	2014-06-09	2014-06-14	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	T
13	CA-2017-114412	2017-04-15	2017-04-20	Standard Class	AA-10480	Andrew Allen	Consumer	United States	Concord	North Carolina	28027	South	C
14	CA-2016-161389	2016-12-05	2016-12-10	Standard Class	IM-15070	Irene Maddox	Consumer	United States	Seattle	Washington	98103	West	C
15	US-2015-118983	2015-11-22	2015-11-26	Standard Class	HP-14815	Harold Pawlan	Home Office	United States	Fort Worth	Texas	76106	Central	C
16	US-2015-118983	2015-11-22	2015-11-26	Standard Class	HP-14815	Harold Pawlan	Home Office	United States	Fort Worth	Texas	76106	Central	C
17	CA-2014-105893	2014-11-11	2014-11-18	Standard Class	PK-19075	Pete Kriz	Consumer	United States	Madison	Wisconsin	53711	Central	C
18	CA-2014-167164	2014-05-13	2014-05-15	Second Class	AG-10270	Alejandro Grove	Consumer	United States	West Jordan	Utah	84084	West	C
19	CA-2014-143336	2014-08-27	2014-09-01	Second Class	ZD-21925	Zuschuss Donatelli	Consumer	United States	San Francisco	California	94109	West	C

- Here the name of columns **Ship Mode** and **Postal Code** changes to **Mode** and **Pin Code**



# Q.1) HOW CAN WE VIEW THE TOP 6 ROWS OF THE RENAMED DATASET?

```
#Q.1) How can we view the top 6 rows of the renamed dataset?  
View(head(renamed_data)) # To View the Top 6 rows of the Renamed Dataset
```



Row ID	Order ID	Order Date	Ship Date	Mode	Customer ID	Customer Name	Segment	Country	City	State	
1	1	CA-2016-152156	2016-11-08	2016-11-11	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky
2	2	CA-2016-152156	2016-11-08	2016-11-11	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky
3	3	CA-2016-138688	2016-06-12	2016-06-16	Second Class	DV-13045	Darrin Van Huff	Corporate	United States	Los Angeles	California
4	4	US-2015-108966	2015-10-11	2015-10-18	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	Florida
5	5	US-2015-108966	2015-10-11	2015-10-18	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	Florida
6	6	CA-2014-115812	2014-06-09	2014-06-14	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California

- Here are the top 6 rows of the **renamed columns**.

## Q.2) HOW CAN WE CHECK FOR MISSING VALUES?

```
#Q.2) How can we Check for missing values?
```

```
colSums(is.na(Dataset)) # To Check the Missing value in the Dataset  
View(colSums(is.na(Dataset))) # To View the Missing value in the Dataset
```

Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment
0	0	0	0	0	0	0	0
Country	City	State	Postal Code	Region	Product ID	Category	Sub-Category
0	0	0	0	0	0	0	0
Product Name	Sales	Quantity	Discount	Profit			
0	0	0	0	0			

Name	Type	Value
colSums(is.na(Dataset))	double [21]	0 0 0 0 0 0 ...
Row ID	double [1]	0
Order ID	double [1]	0
Order Date	double [1]	0
Ship Date	double [1]	0
Ship Mode	double [1]	0
Customer ID	double [1]	0
Customer Name	double [1]	0
Segment	double [1]	0
Country	double [1]	0
City	double [1]	0
State	double [1]	0
Postal Code	double [1]	0
Region	double [1]	0
Product ID	double [1]	0
Category	double [1]	0
Sub-Category	double [1]	0
Product Name	double [1]	0

- Here are the **missing values** of the dataset

### Q.3) HOW CAN WE FILTERED THE ROWS BASED AS PER OUR REQUIREMENTS?

```
#Q.3) How can we Filtered the Rows based as per our requirements?  
filtered_data=Dataset%>%  
  filter(`Ship Mode` == "First Class", Segment=="Corporate", State=="Texas")  
View(filtered_data) # To View the Filter data that contains value having Ship Mode= First Class, Segment=Corporate, State= Texas
```

The screenshot shows an RStudio interface with a code editor window containing the provided R code. Below it is a data grid titled 'filtered\_data' showing a subset of the Superstore dataset. The columns include Row ID, Order ID, Order Date, Ship Date, Ship Mode, Customer ID, Customer Name, Segment, Country, City, State, Postal Code, Region, Product ID, and Category. The data is filtered to show rows where Ship Mode is 'First Class', Segment is 'Corporate', and State is 'Texas'. The first few rows of the filtered data are:

Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	State	Postal Code	Region	Product ID	Category	
1	36	CA-2016-117590	2016-12-08	2016-12-10	First Class	GH-14485	Gene Hale	Corporate	United States	Richardson	Texas	75080	Central	TEC-PH-10004977	Technology
2	37	CA-2016-117590	2016-12-08	2016-12-10	First Class	GH-14485	Gene Hale	Corporate	United States	Richardson	Texas	75080	Central	FUR-FU-10003664	Furniture
3	76	US-2017-118038	2017-12-09	2017-12-11	First Class	KB-16600	Ken Brennan	Corporate	United States	Houston	Texas	77041	Central	OFF-BI-10004182	Office
4	77	US-2017-118038	2017-12-09	2017-12-11	First Class	KB-16600	Ken Brennan	Corporate	United States	Houston	Texas	77041	Central	FUR-FU-10000260	Furniture
5	78	US-2017-118038	2017-12-09	2017-12-11	First Class	KB-16600	Ken Brennan	Corporate	United States	Houston	Texas	77041	Central	OFF-ST-10000615	Office
6	611	CA-2016-161816	2016-04-28	2016-05-01	First Class	NB-18655	Nona Balk	Corporate	United States	Dallas	Texas	75217	Central	TEC-PH-10003012	Technology
7	612	CA-2016-161816	2016-04-28	2016-05-01	First Class	NB-18655	Nona Balk	Corporate	United States	Dallas	Texas	75217	Central	OFF-LA-10004345	Office
8	1417	CA-2015-126697	2015-09-21	2015-09-24	First Class	SV-20815	Stuart Van	Corporate	United States	Houston	Texas	77041	Central	TEC-PH-10002922	Technology
9	1418	CA-2015-126697	2015-09-21	2015-09-24	First Class	SV-20815	Stuart Van	Corporate	United States	Houston	Texas	77041	Central	TEC-AC-10004353	Technology
10	1419	CA-2015-126697	2015-09-21	2015-09-24	First Class	SV-20815	Stuart Van	Corporate	United States	Houston	Texas	77041	Central	FUR-FU-10001706	Furniture
11	1463	CA-2016-152289	2016-08-26	2016-08-28	First Class	LC-16930	Linda Cazamias	Corporate	United States	Pasadena	Texas	77506	Central	TEC-AC-10004571	Technology
12	1464	CA-2016-152289	2016-08-26	2016-08-28	First Class	LC-16930	Linda Cazamias	Corporate	United States	Pasadena	Texas	77506	Central	FUR-CH-10002126	Furniture
13	2307	CA-2015-160794	2015-08-06	2015-08-08	First Class	MS-17980	Michael Stewart	Corporate	United States	Houston	Texas	77041	Central	OFF-PA-10004156	Office
14	3186	CA-2014-123498	2014-11-07	2014-11-09	First Class	TC-20980	Tamara Chand	Corporate	United States	Houston	Texas	77041	Central	OFF-BI-10000632	Office

- Here are the rows that are filtered based on **Ship Mode=First Class, Segment=Corporate, State=Texas**

#### Q.4) HOW CAN WE VIEW THE TOP 6 ROWS OF THE FILTERED DATASET?

```
#Q.4) How can we view thw top 6 rows of the Filtered dataset?  
view(head(filtered_data)) # To View the Top 6 rows of the Filtered Dataset
```

Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	State	Postal Code	Region	Product ID	Category
36	CA-2016-117590	2016-12-08	2016-12-10	First Class	GH-14485	Gene Hale	Corporate	United States	Richardson	Texas	75080	Central	TEC-PH-10004977	Technology
37	CA-2016-117590	2016-12-08	2016-12-10	First Class	GH-14485	Gene Hale	Corporate	United States	Richardson	Texas	75080	Central	FUR-FU-10003664	Furniture
76	US-2017-118038	2017-12-09	2017-12-11	First Class	KB-16600	Ken Brennan	Corporate	United States	Houston	Texas	77041	Central	OFF-BI-10004182	Office Supp
77	US-2017-118038	2017-12-09	2017-12-11	First Class	KB-16600	Ken Brennan	Corporate	United States	Houston	Texas	77041	Central	FUR-FU-10000260	Furniture
78	US-2017-118038	2017-12-09	2017-12-11	First Class	KB-16600	Ken Brennan	Corporate	United States	Houston	Texas	77041	Central	OFF-ST-10000615	Office Supp
611	CA-2016-161816	2016-04-28	2016-05-01	First Class	NB-18655	Nonna Balk	Corporate	United States	Dallas	Texas	75217	Central	TEC-PH-10003012	Technology

- Here are the top 6 rows of the **Filtered Dataset**.

## Q.5) HOW CAN WE VIEW THE SUMMARY OF CATEGORY, SUB-CATEGORY,SEGMENT,TOTAL SALES AND TOTAL PROFIT?

```
#Q.5) How can we view the Summary of Category, Sub-Category,Segment,Total Sales and Total Profit?  
Short_Summary=Dataset%>%  
group_by(Category, `Sub-Category`,Segment) %>%  
summarise(Total_Sales=sum(Sales, na.rm = TRUE),  
          Total_Quantity=sum(Quantity, na.rm=TRUE),  
          Total_Profit=sum(Profit, na.rm = TRUE))  
View(Short_Summary)           # To View the Short Summary of the Dataset
```

The screenshot shows the RStudio interface with the code for generating a summary dataset. Below the code, the 'Short\_Summary' dataset is displayed as a table. The table has columns: Category, Sub-Category, Segment, Total\_Sales, Total\_Quantity, and Total\_Profit. The data consists of 18 rows, each numbered from 1 to 18, representing different products or categories. The 'Category' column includes Furniture, Office Supplies, and Art. The 'Sub-Category' column includes Bookcases, Chairs, Furnishings, Tables, Appliances, and Art. The 'Segment' column includes Consumer, Corporate, and Home Office. The 'Total\_Sales' column shows values like 68632.729, 34005.924, etc. The 'Total\_Quantity' column shows values like 496, 271, etc. The 'Total\_Profit' column shows values like -4435.6382, 638.4502, etc.

	Category	Sub-Category	Segment	Total_Sales	Total_Quantity	Total_Profit
1	Furniture	Bookcases	Consumer	68632.729	496	-4435.6382
2	Furniture	Bookcases	Corporate	34005.924	271	638.4502
3	Furniture	Bookcases	Home Office	12241.343	101	324.6320
4	Furniture	Chairs	Consumer	172862.742	1234	13235.3319
5	Furniture	Chairs	Corporate	99140.878	719	8344.6565
6	Furniture	Chairs	Home Office	56445.483	403	5010.1779
7	Furniture	Furnishings	Consumer	49620.046	1834	7919.4227
8	Furniture	Furnishings	Corporate	25001.266	1086	3508.2077
9	Furniture	Furnishings	Home Office	17083.852	643	1631.5132
10	Furniture	Tables	Consumer	99933.795	602	-9728.0378
11	Furniture	Tables	Corporate	70871.717	419	-4906.4986
12	Furniture	Tables	Home Office	36160.020	220	-3090.9447
13	Office Supplies	Appliances	Consumer	52819.581	908	6981.9282
14	Office Supplies	Appliances	Corporate	36588.683	569	7429.8952
15	Office Supplies	Appliances	Home Office	18123.897	252	3726.1820
16	Office Supplies	Art	Consumer	14251.930	1625	3454.3011
17	Office Supplies	Art	Corporate	8590.448	850	2004.6477
18	Office Supplies	Art	Home Office	4276.414	525	1068.8382

- Here is the summary of **Category, Sub-Category and Segment wise Total Sales and Total Profit.**



## Q.6) HOW CAN WE FIND THE MOST PROFITABLE CATEGORY?

```
#Q.6) How can we find the most profitable category?  
most_profitable_category=Dataset%>%  
  group_by(Category)%>%  
  summarise(Total_Profit = sum(Profit, na.rm = TRUE))%>%  
  arrange(desc(Total_Profit)) %>%  
  head(1)  
View(most_profitable_category)      # To View the Most Profitable Category
```

	Category	Total_Profit
1	Technology	145454.9

- So, **Technology** is the Most Profitable Category with Total Profit **145454.9**



## Q.7) HOW CAN WE FIND THE LEAST PROFITABLE CATEGORY?

```
#Q.7) How can we find the least profitable category?  
least_profitable_category=Dataset%>%  
  group_by(Category)%>%  
  summarise(Total_Profit = sum(Profit, na.rm = TRUE))%>%  
  arrange(Total_Profit) %>%  
  head(1)  
view(least_profitable_category) # To View the Least Profitable Category
```

	Category	Total_Profit
1	Furniture	18451.27

- So, **Furniture** is the **Least Profitable Category** with Total Profit **18451.27**

## Q.8) HOW TO CREATE A NEW COLUMN 'UNIT PRICE OF A PRODUCT' BY USING MUTATING AND TRANSFORMING DATA?

```
#Q.8) How to create a new Column 'Unit Price of a Product' by using Mutating and Transforming Data?  
mutated_data=Short_Summary%>%  
  mutate(  
    Unit_Price=Total_Sales/Total_Quantity  
  )  
View(mutated_data)          # To View the new column of the Dataset  
head(mutated_data)  
View(head(mutated_data))    # To View the top 6 rows of the Dataset
```

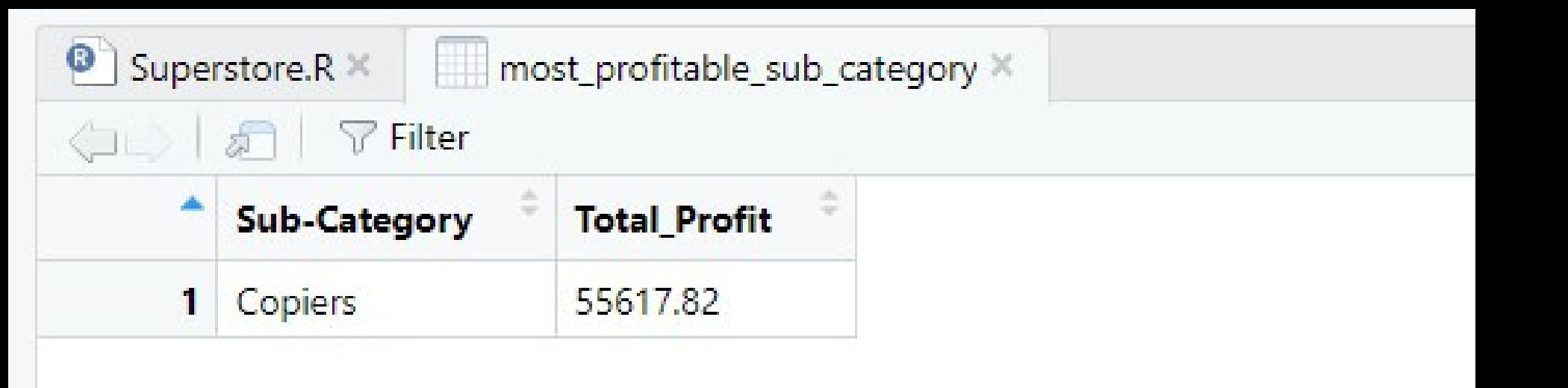
The screenshot shows an RStudio interface with a code editor and a data viewer. The code editor contains the R code provided above. The data viewer shows a table titled 'mutated\_data' with 18 rows of data. The columns are: Category, Sub-Category, Segment, Total\_Sales, Total\_Quantity, Total\_Profit, and Unit\_Price. The Unit\_Price column is the result of the mutation, calculated as Total\_Sales divided by Total\_Quantity.

	Category	Sub-Category	Segment	Total_Sales	Total_Quantity	Total_Profit	Unit_Price
1	Furniture	Bookcases	Consumer	68632.729	496	-4435.6382	138.372438
2	Furniture	Bookcases	Corporate	34005.924	271	638.4502	125.483115
3	Furniture	Bookcases	Home Office	12241.343	101	324.6320	121.201416
4	Furniture	Chairs	Consumer	172862.742	1234	13235.3319	140.083259
5	Furniture	Chairs	Corporate	99140.878	719	8344.6565	137.887174
6	Furniture	Chairs	Home Office	56445.483	403	5010.1779	140.063233
7	Furniture	Furnishings	Consumer	49620.046	1834	7919.4227	27.055641
8	Furniture	Furnishings	Corporate	25001.266	1086	3508.2077	23.021424
9	Furniture	Furnishings	Home Office	17083.852	643	1631.5132	26.568977
10	Furniture	Tables	Consumer	99933.795	602	-9728.0378	166.002982
11	Furniture	Tables	Corporate	70871.717	419	-4906.4986	169.144911
12	Furniture	Tables	Home Office	36160.020	220	-3090.9447	164.363725
13	Office Supplies	Appliances	Consumer	52819.581	908	6981.9282	58.171345
14	Office Supplies	Appliances	Corporate	36588.683	569	7429.8952	64.303485
15	Office Supplies	Appliances	Home Office	18123.897	252	3726.1820	71.920226
16	Office Supplies	Art	Consumer	14251.930	1625	3454.3011	8.770418
17	Office Supplies	Art	Corporate	8590.448	850	2004.6477	10.106409
18	Office Supplies	Art	Home Office	4276.414	525	1068.8382	8.145550

- Here is the creation of new column by using **Mutating** and **Transforming Data**.

## Q.9) HOW CAN WE FIND THE MOST PROFITABLE SUB-CATEGORY?

```
#Q.9) How can we find the most profitable Sub-Category?  
most_profitable_sub_category=Dataset%>%  
  group_by(`Sub-Category`)%>%  
  summarise(Total_Profit = sum(Profit, na.rm = TRUE))%>%  
  arrange(desc(Total_Profit)) %>%  
  head(1)  
View(most_profitable_sub_category)      # To View the Most Profitable Sub-Category
```



	Sub-Category	Total_Profit
1	Copiers	55617.82

- So, **Copier** is the **Most Profitable Sub-Category** with **Total Profit 55617.82**

## Q.10) HOW CAN WE FIND THE LEAST PROFITABLE SUB-CATEGORY?

```
#Q.10) How can we find the least profitable Sub-Category?  
least_profitable_sub_category=Dataset%>%  
  group_by(`Sub-Category`)%>%  
  summarise(Total_Profit = sum(Profit, na.rm = TRUE))%>%  
  arrange(Total_Profit) %>%  
  head(1)  
View(least_profitable_sub_category) # To View the Least Profitable Sub-Category
```

	Sub-Category	Total_Profit
1	Tables	-17725.48

- So, **Tables** is the **Least Profitable Sub-Category** with **Total Profit -17725.48** i.e Loss of 17725.48

# THANK YOU

