In [1]:
```
import pandas as pd
import numpy as np
```

In [7]:
```
df=pd.read_csv('C:/Users/raviy/OneDrive/Desktop/MOCK_DATA.csv')
```

In [8]:
```
df
```

Out[8]:

	id	first_name	last_name	email	gender	Date of birth
0	1	Skyler	Thornhill	sthornhill0@merriam-webster.com	Male	02-06-2025
1	2	Cary	Buston	cbuston1@shutterfly.com	Male	27-08-2025
2	3	Sonnie	NaN	speace2@time.com	Male	11-07-2025
3	4	NaN	Dummett	ldummett3@blogtalkradio.com	Male	16-01-2025
4	5	Juana	Cearley	jcearley4@mail.ru	Female	07-01-2025
5	6	Patrizius	Skates	pskates5@stanford.edu	Male	02-05-2025
6	7	Carmelia	Ferran	cferran6@npr.org	Female	05-05-2025
7	8	Izaak	Meiklejohn	imeiklejohn7@prnewswire.com	Male	04-01-2025
8	9	Randene	Forsythe	rforsythe8@github.com	Female	10-08-2025
9	10	Adela	Dikels	adikels9@businessinsider.com	Polygender	06-07-2025
10	11	Phillie	Locarno	plocarnoa@sitemeter.com	Female	05-01-2025
11	12	Gretchen	Laydon	glaydonb@harvard.edu	Female	NaN

In [9]:
```
df.dropna()
```

34	35	Willy	Cornilleau	wcornilleauy@digg.com	Female	01-06-2025
35	36	Joey	Havard	jhavardz@plala.or.jp	Female	02-08-2025
36	37	Tucker	Heffy	theffy10@eepurl.com	Male	22-09-2025
37	38	Svend	Coaster	scoaster11@about.me	Male	18-02-2025
38	39	Fayina	Baythrop	fbaythrop12@hp.com	Female	21-10-2025
39	40	Alec	Coleson	acoleson13@miibeian.gov.cn	Male	23-06-2025
40	41	Danita	Mitton	dmitton14@example.com	Female	28-11-2024
41	42	Nollie	Baruch	nbaruch15@chronoengine.com	Female	08-03-2025
42	43	Raddy	Braxay	rbraxay16@europa.eu	Male	13-06-2025
43	44	Vita	Rudd	vrudd17@bing.com	Female	29-03-2025
45	46	Marcelle	Littlecote	mlittlecote19@blog.com	Female	01-10-2025
48	49	Lotti	Farlam	lfarlam1c@hc360.com	Female	12-10-2025
49	50	Anne-corinne	Penright	apenright1d@zimbio.com	Female	09-09-2025

In [10]:
```
df.dropna(inplace=True)
```

In [12]:
```
df["gender"].fillna('male')
```

```
23     Genderqueer
24          Female
26            Male
28            Male
32            Male
33            Male
34          Female
35          Female
36            Male
37            Male
38          Female
39            Male
40          Female
41          Female
42            Male
43          Female
45          Female
48          Female
49          Female
Name: gender, dtype: object
```

In [2]:
```
df1=pd.read_csv('C:/Users/raviy/OneDrive/Desktop/sample_data.csv')
```

In [15]:
```
df1
```

Out[15]:

	Product name	sale price	MRP	Ratings	sale date
0	Whisk Set	12.99	4.99	37.0	03-10-2025
1	Maple Syrup	6.99	2.99	22.0	08-01-2025
2	Vegetable Chips	3.99	79.99	46.0	10-09-2025
3	Hummus Variety Pack	5.99	2.49	46.0	12/25/2024
4	Ice Cream Maker	79.99	2.49	20.0	02-07-2025
5	Portable Air Compressor	44.99	1.89	92.0	07-05-2025
6	Silicone Ice Cube Tray	10.99	5.49	20.0	7/30/2025
7	Lightweight Backpacking Tent	129.99	5.49	46.0	4/29/2025
8	Cushion Covers	35.00	4.29	25.0	03-01-2025
9	Sriracha Hot Chili Sauce	2.99	4.99	33.0	11/28/2024
10	Compact Refrigerator	199.99	89.99	49.0	02-07-2025
11	Dark Chocolate Covered Almonds	7.19	3.29	44.0	06-04-2025

In [21]:

```
#replace with mean ,median,mode
x=df1['sale price'].mean()
df1['sale price'].fillna(x)
```

```
31      39.990000
32      15.990000
33      34.990000
34      34.990000
35      49.990000
36       3.990000
37       5.990000
38       2.890000
39      25.502708
40      19.990000
41      24.990000
42       2.490000
43      24.990000
44       5.990000
45       9.990000
46       5.490000
47       5.990000
48      34.990000
49       3.990000
Name: sale price, dtype: float64
```

In [3]:

```
x=df1['sale price'].median()
df1['sale price'].fillna(x)
```

Out[3]:

```
0      12.99
1       6.99
2       3.99
3       5.99
4      79.99
5      44.99
6      10.99
7     129.99
8      35.00
9       2.99
10    199.99
11      7.19
12     10.49
13      2.09
14      3.99
15     39.99
16     34.99
17      3.99
18      2.29
19      2.99
```

In [5]:

```
x=df1['MRP'].mode()[0]
df1['MRP'].fillna(x)
```

```
32     10.99
33      2.79
34      3.99
35     18.99
36      3.49
37     15.99
38     34.99
39      7.99
40     24.99
41     44.99
42      1.99
43    199.99
44      1.99
45      3.29
46      3.19
47     49.99
48      8.99
49     14.99
Name: MRP, dtype: float64
```

In [10]:

```
df1['MRP']
```

Out[10]:

```
0       4.99
1       2.99
2      79.99
3       2.49
4       2.49
5       1.89
6       5.49
7       5.49
8       4.29
9       4.99
10     89.99
11      3.29
12     54.99
13     99.99
14      2.99
15      2.99
16     29.99
17      2.79
18      1.99
```

In [11]:
```
df1.loc[2,'MRP']=3.99 # replacing vals
df1['MRP']
```

Out[11]:
```
0       4.99
1       2.99
2       3.99
3       2.49
4       2.49
5       1.89
6       5.49
7       5.49
8       4.29
9       4.99
10     89.99
11      3.29
12     54.99
13     99.99
14      2.99
15      2.99
16     29.99
17      2.79
18      1.99
```

In [12]:
```
for x in df1.index:
    if df1.loc[x,'MRP']<2.00:
        df1.loc[x,'MRP']=2.00
print(df1['MRP'])
```

```
30     12.99
31      2.00
32     10.99
33      2.79
34      3.99
35     18.99
36      3.49
37     15.99
38     34.99
39      7.99
40     24.99
41     44.99
42      2.00
43    199.99
44       NaN
45      3.29
46      3.19
47     49.99
48      8.99
49     14.99
```

Duplicated-To find duplicate rows

In [14]:
```
print(df1.duplicated())
```

```
0     False
1     False
2     False
3     False
4     False
5     False
6     False
7     False
8     False
9     False
10    False
11    False
12    False
13    False
14    False
15    False
16    False
17    False
18    False
19    False
```

Drop_duplicates()-to remove duplicates

In [16]:
```
df1.drop_duplicates()
```

Out[16]:

	Product name	sale price	MRP	Ratings	sale date
0	Whisk Set	12.99	4.99	37.0	03-10-2025
1	Maple Syrup	6.99	2.99	22.0	08-01-2025
2	Vegetable Chips	3.99	3.99	46.0	10-09-2025
3	Hummus Variety Pack	5.99	2.49	46.0	12/25/2024
4	Ice Cream Maker	79.99	2.49	20.0	02-07-2025
5	Portable Air Compressor	44.99	2.00	92.0	07-05-2025
6	Silicone Ice Cube Tray	10.99	5.49	20.0	7/30/2025
7	Lightweight Backpacking Tent	129.99	5.49	46.0	4/29/2025
8	Cushion Covers	35.00	4.29	25.0	03-01-2025
9	Sriracha Hot Chili Sauce	2.99	4.99	33.0	11/28/2024
10	Compact Refrigerator	199.99	89.99	49.0	02-07-2025
11	Dark Chocolate Covered Almonds	7.19	3.29	44.0	06-04-2025

Data correlations

```
finding relationship between column in dataset

the number varies from -1 to 1
```

```
1 means one to one relationship (perfect correlation)
```

In [17]:
```
df1[['sale price','MRP','Ratings']].corr()
```

Out[17]:

	sale price	MRP	Ratings
sale price	1.000000	0.134812	-0.100232
MRP	0.134812	1.000000	0.055787
Ratings	-0.100232	0.055787	1.000000

Describe-to view basic statistical details

In [18]:
```
df1.describe()
```

Out[18]:

	sale price	MRP	Ratings
count	48.000000	49.000000	49.000000
mean	25.502708	21.878571	49.714286
std	35.405632	36.103907	24.076268
min	1.290000	2.000000	3.000000
25%	3.990000	3.290000	33.000000
50%	10.490000	5.490000	46.000000
75%	34.990000	24.990000	69.000000
max	199.990000	199.990000	96.000000

In [19]:
```
#adding new column
df1['new']=4
df1
```

Out[19]:

	Product name	sale price	MRP	Ratings	sale date	new
0	Whisk Set	12.99	4.99	37.0	03-10-2025	4
1	Maple Syrup	6.99	2.99	22.0	08-01-2025	4
2	Vegetable Chips	3.99	3.99	46.0	10-09-2025	4
3	Hummus Variety Pack	5.99	2.49	46.0	12/25/2024	4
4	Ice Cream Maker	79.99	2.49	20.0	02-07-2025	4
5	Portable Air Compressor	44.99	2.00	92.0	07-05-2025	4
6	Silicone Ice Cube Tray	10.99	5.49	20.0	7/30/2025	4
7	Lightweight Backpacking Tent	129.99	5.49	46.0	4/29/2025	4
8	Cushion Covers	35.00	4.29	25.0	03-01-2025	4
9	Sriracha Hot Chili Sauce	2.99	4.99	33.0	11/28/2024	4
10	Compact Refrigerator	199.99	89.99	49.0	02-07-2025	4
11	Dark Chocolate Covered Almonds	7.19	3.29	44.0	06-04-2025	4

In [27]:

```
#col deletion
df1.drop(['new'],axis=1,inplace=True)
df1
```

Out[27]:

	Product name	sale price	MRP	Ratings	sale date
0	Whisk Set	12.99	4.99	37.0	03-10-2025
1	Maple Syrup	6.99	2.99	22.0	08-01-2025
2	Vegetable Chips	3.99	3.99	46.0	10-09-2025
3	Hummus Variety Pack	5.99	2.49	46.0	12/25/2024
4	Ice Cream Maker	79.99	2.49	20.0	02-07-2025
5	Portable Air Compressor	44.99	2.00	92.0	07-05-2025
6	Silicone Ice Cube Tray	10.99	5.49	20.0	7/30/2025
7	Lightweight Backpacking Tent	129.99	5.49	46.0	4/29/2025
8	Cushion Covers	35.00	4.29	25.0	03-01-2025
9	Sriracha Hot Chili Sauce	2.99	4.99	33.0	11/28/2024
10	Compact Refrigerator	199.99	89.99	49.0	02-07-2025
11	Dark Chocolate Covered Almonds	7.19	3.29	44.0	06-04-2025

In [28]:

```
a=pd.DataFrame({'product name':2,'sale price':3,
                'MRP':5,'Ratings':8,'sale date':7},index=[0])
a
```

Out[28]:

	product name	sale price	MRP	Ratings	sale date
0	2	3	5	8	7

In [29]:

```
df1=pd.concat([a,df1])
df1
```

Out[29]:

	product name	sale price	MRP	Ratings	sale date	Product name
0	2.0	3.00	5.00	8.0	7	NaN
0	NaN	12.99	4.99	37.0	03-10-2025	Whisk Set
1	NaN	6.99	2.99	22.0	08-01-2025	Maple Syrup
2	NaN	3.99	3.99	46.0	10-09-2025	Vegetable Chips
3	NaN	5.99	2.49	46.0	12/25/2024	Hummus Variety Pack
4	NaN	79.99	2.49	20.0	02-07-2025	Ice Cream Maker
5	NaN	44.99	2.00	92.0	07-05-2025	Portable Air Compressor
6	NaN	10.99	5.49	20.0	7/30/2025	Silicone Ice Cube Tray
7	NaN	129.99	5.49	46.0	4/29/2025	Lightweight Backpacking Tent
8	NaN	35.00	4.29	25.0	03-01-2025	Cushion Covers
9	NaN	2.99	4.99	33.0	11/28/2024	Sriracha Hot Chili Sauce

In [30]:

```
#row deletion
df1.drop(0,inplace=True)
df1
```

Out[30]:

	product name	sale price	MRP	Ratings	sale date	Product name
1	NaN	6.99	2.99	22.0	08-01-2025	Maple Syrup
2	NaN	3.99	3.99	46.0	10-09-2025	Vegetable Chips
3	NaN	5.99	2.49	46.0	12/25/2024	Hummus Variety Pack
4	NaN	79.99	2.49	20.0	02-07-2025	Ice Cream Maker
5	NaN	44.99	2.00	92.0	07-05-2025	Portable Air Compressor
6	NaN	10.99	5.49	20.0	7/30/2025	Silicone Ice Cube Tray
7	NaN	129.99	5.49	46.0	4/29/2025	Lightweight Backpacking Tent
8	NaN	35.00	4.29	25.0	03-01-2025	Cushion Covers
9	NaN	2.99	4.99	33.0	11/28/2024	Sriracha Hot Chili Sauce
10	NaN	199.99	89.99	49.0	02-07-2025	Compact Refrigerator
11	NaN	7.19	3.29	44.0	06-04-2025	Dark Chocolate Covered Almonds

In [31]:

```
df1[2:6]
```

Out[31]:

	product name	sale price	MRP	Ratings	sale date	Product name
3	NaN	5.99	2.49	46.0	12/25/2024	Hummus Variety Pack
4	NaN	79.99	2.49	20.0	02-07-2025	Ice Cream Maker
5	NaN	44.99	2.00	92.0	07-05-2025	Portable Air Compressor
6	NaN	10.99	5.49	20.0	7/30/2025	Silicone Ice Cube Tray

In [32]:

```
df1[1:7:2]
```

Out[32]:

	product name	sale price	MRP	Ratings	sale date	Product name
2	NaN	3.99	3.99	46.0	10-09-2025	Vegetable Chips
4	NaN	79.99	2.49	20.0	02-07-2025	Ice Cream Maker
6	NaN	10.99	5.49	20.0	7/30/2025	Silicone Ice Cube Tray

In [33]:

```
df1[['MRP']][4:11]
```

Out[33]:

	MRP
5	2.00
6	5.49
7	5.49
8	4.29
9	4.99
10	89.99
11	3.29

In [ ]: