In [1]:
```
import pandas as pd
url="https://raw.githubusercontent.com/svkarthik86/Assignment/main/Sales_Data/Sales_January_2019.csv"
sales_data=pd.read_csv(url)
sales_data.head(4)
```

Out[1]:

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address
0	141234	iPhone	1	700	01/22/19 21:25	944 Walnut St, Boston, MA 02215
1	141235	Lightning Charging Cable	1	14.95	01/28/19 14:15	185 Maple St, Portland, OR 97035
2	141236	Wired Headphones	2	11.99	01/17/19 13:33	538 Adams St, San Francisco, CA 94016
3	141237	27in FHD Monitor	1	149.99	01/05/19 20:33	738 10th St, Los Angeles, CA 90001

In [2]:
```
sales_data.columns=sales_data.columns.str.replace(" ","_")
sales_data.dropna(inplace=True)
sales_data
```

Out[2]:

	Order_ID	Product	Quantity_Ordered	Price_Each	Order_Date	Purchase_Address
0	141234	iPhone	1	700	01/22/19 21:25	944 Walnut St, Boston, MA 02215
1	141235	Lightning Charging Cable	1	14.95	01/28/19 14:15	185 Maple St, Portland, OR 97035
2	141236	Wired Headphones	2	11.99	01/17/19 13:33	538 Adams St, San Francisco, CA 94016
3	141237	27in FHD Monitor	1	149.99	01/05/19 20:33	738 10th St, Los Angeles, CA 90001
4	141238	Wired Headphones	1	11.99	01/25/19 11:59	387 10th St, Austin, TX 73301
...	...	...	...	...	...	...
9718	150497	20in Monitor	1	109.99	01/26/19 19:09	95 8th St, Dallas, TX 75001
9719	150498	27in FHD Monitor	1	149.99	01/10/19 22:58	403 7th St, San Francisco, CA 94016
9720	150499	ThinkPad Laptop	1	999.99	01/21/19 14:31	214 Main St, Portland, OR 97035
9721	150500	AAA Batteries (4-pack)	2	2.99	01/15/19 14:21	810 2nd St, Los Angeles, CA 90001
9722	150501	Google Phone	1	600	01/13/19 16:43	428 Cedar St, Boston, MA 02215

9697 rows × 6 columns

In [3]:
```
sales_data.drop_duplicates(inplace=True)
sales_data[sales_data.duplicated()]
sales_data=sales_data[~(sales_data.Price_Each=="Price Each")]
sales_data
```

Out[3]:

	Order_ID	Product	Quantity_Ordered	Price_Each	Order_Date	Purchase_Address
0	141234	iPhone	1	700	01/22/19 21:25	944 Walnut St, Boston, MA 02215
1	141235	Lightning Charging Cable	1	14.95	01/28/19 14:15	185 Maple St, Portland, OR 97035
2	141236	Wired Headphones	2	11.99	01/17/19 13:33	538 Adams St, San Francisco, CA 94016
3	141237	27in FHD Monitor	1	149.99	01/05/19 20:33	738 10th St, Los Angeles, CA 90001
4	141238	Wired Headphones	1	11.99	01/25/19 11:59	387 10th St, Austin, TX 73301
...	...	...	...	...	...	...
9718	150497	20in Monitor	1	109.99	01/26/19 19:09	95 8th St, Dallas, TX 75001
9719	150498	27in FHD Monitor	1	149.99	01/10/19 22:58	403 7th St, San Francisco, CA 94016
9720	150499	ThinkPad Laptop	1	999.99	01/21/19 14:31	214 Main St, Portland, OR 97035
9721	150500	AAA Batteries (4-pack)	2	2.99	01/15/19 14:21	810 2nd St, Los Angeles, CA 90001
9722	150501	Google Phone	1	600	01/13/19 16:43	428 Cedar St, Boston, MA 02215

9671 rows × 6 columns

In [4]:
```
sales_data["order_price"]=sales_data.Price_Each.astype(float)*sales_data.Quantity_Ordered.astype(float)
sales_data
```

```
C:\Users\HAPPYHOME\AppData\Local\Temp\ipykernel_8020\2453370470.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  sales_data["order_price"]=sales_data.Price_Each.astype(float)*sales_data.Quantity_Ordered.astype(float)
```

Out[4]:

	Order_ID	Product	Quantity_Ordered	Price_Each	Order_Date	Purchase_Address	order_price
0	141234	iPhone	1	700	01/22/19 21:25	944 Walnut St, Boston, MA 02215	700.00
1	141235	Lightning Charging Cable	1	14.95	01/28/19 14:15	185 Maple St, Portland, OR 97035	14.95
2	141236	Wired Headphones	2	11.99	01/17/19 13:33	538 Adams St, San Francisco, CA 94016	23.98
3	141237	27in FHD Monitor	1	149.99	01/05/19 20:33	738 10th St, Los Angeles, CA 90001	149.99
4	141238	Wired Headphones	1	11.99	01/25/19 11:59	387 10th St, Austin, TX 73301	11.99
...	...	...	...	...	...	...	...
9718	150497	20in Monitor	1	109.99	01/26/19 19:09	95 8th St, Dallas, TX 75001	109.99
9719	150498	27in FHD Monitor	1	149.99	01/10/19 22:58	403 7th St, San Francisco, CA 94016	149.99
9720	150499	ThinkPad Laptop	1	999.99	01/21/19 14:31	214 Main St, Portland, OR 97035	999.99
9721	150500	AAA Batteries (4-pack)	2	2.99	01/15/19 14:21	810 2nd St, Los Angeles, CA 90001	5.98
9722	150501	Google Phone	1	600	01/13/19 16:43	428 Cedar St, Boston, MA 02215	600.00

9671 rows × 7 columns

In [5]:
```
c=sales_data.drop('Purchase_Address',1)
c
```

```
C:\Users\HAPPYHOME\AppData\Local\Temp\ipykernel_8020\623450919.py:1: FutureWarning: In a future version of pandas all arguments of DataFrame.drop except for t
he argument 'labels' will be keyword-only.
  c=sales_data.drop('Purchase_Address',1)
```

Out[5]:

	Order_ID	Product	Quantity_Ordered	Price_Each	Order_Date	order_price
0	141234	iPhone	1	700	01/22/19 21:25	700.00
1	141235	Lightning Charging Cable	1	14.95	01/28/19 14:15	14.95
2	141236	Wired Headphones	2	11.99	01/17/19 13:33	23.98
3	141237	27in FHD Monitor	1	149.99	01/05/19 20:33	149.99
4	141238	Wired Headphones	1	11.99	01/25/19 11:59	11.99
...	...	...	...	...	...	...
9718	150497	20in Monitor	1	109.99	01/26/19 19:09	109.99
9719	150498	27in FHD Monitor	1	149.99	01/10/19 22:58	149.99
9720	150499	ThinkPad Laptop	1	999.99	01/21/19 14:31	999.99
9721	150500	AAA Batteries (4-pack)	2	2.99	01/15/19 14:21	5.98
9722	150501	Google Phone	1	600	01/13/19 16:43	600.00

9671 rows × 6 columns

In [6]:
```
sales_data=pd.concat((sales_data,pd.DataFrame([[100001,"OnePhone",1,600,"01/22/19 21:25",600]],columns=['Order_ID', 'Product', 'Quantity_Ordered', 'Price_Each',
       'order_price'])),ignore_index=True)
sales_data
```

Out[6]:

	Order_ID	Product	Quantity_Ordered	Price_Each	Order_Date	Purchase_Address	order_price
0	141234	iPhone	1	700	01/22/19 21:25	944 Walnut St, Boston, MA 02215	700.00
1	141235	Lightning Charging Cable	1	14.95	01/28/19 14:15	185 Maple St, Portland, OR 97035	14.95
2	141236	Wired Headphones	2	11.99	01/17/19 13:33	538 Adams St, San Francisco, CA 94016	23.98
3	141237	27in FHD Monitor	1	149.99	01/05/19 20:33	738 10th St, Los Angeles, CA 90001	149.99
4	141238	Wired Headphones	1	11.99	01/25/19 11:59	387 10th St, Austin, TX 73301	11.99
...	...	...	...	...	...	...	...
9667	150498	27in FHD Monitor	1	149.99	01/10/19 22:58	403 7th St, San Francisco, CA 94016	149.99
9668	150499	ThinkPad Laptop	1	999.99	01/21/19 14:31	214 Main St, Portland, OR 97035	999.99
9669	150500	AAA Batteries (4-pack)	2	2.99	01/15/19 14:21	810 2nd St, Los Angeles, CA 90001	5.98
9670	150501	Google Phone	1	600	01/13/19 16:43	428 Cedar St, Boston, MA 02215	600.00
9671	100001	OnePhone	1	600	01/22/19 21:25	NaN	600.00

9672 rows × 7 columns

In [7]:
```
import numpy as np
sales_data.groupby("Product")["order_price"].agg([min,max,np.mean])
```

Out[7]:

Product	min	max	mean
20in Monitor	109.99	219.98	111.022770
27in 4K Gaming Monitor	389.99	389.99	389.990000
27in FHD Monitor	149.99	299.98	150.348828
34in Ultrawide Monitor	379.99	759.98	381.204026
AA Batteries (4-pack)	3.84	19.20	5.267977
AAA Batteries (4-pack)	2.99	20.93	4.410388
Apple Airpods Headphones	150.00	300.00	150.928218
Bose SoundSport Headphones	99.99	199.98	100.447271
Flatscreen TV	300.00	300.00	300.000000
Google Phone	600.00	1200.00	601.898734
LG Dryer	600.00	600.00	600.000000
LG Washing Machine	600.00	600.00	600.000000
Lightning Charging Cable	14.95	44.85	16.114024
Macbook Pro Laptop	1700.00	1700.00	1700.000000
OnePhone	600.00	600.00	600.000000
ThinkPad Laptop	999.99	1999.98	1004.619583
USB-C Charging Cable	11.95	47.80	13.106122
Vareebadd Phone	400.00	800.00	403.225806
Wired Headphones	11.99	35.97	12.898514
iPhone	700.00	700.00	700.000000

In [ ]: