```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt


df = pd.read_csv('shopping.csv')
df
```

	Date	Product	Quantity	Price	Total
0	01/01/2023	Apples	10	0.5	NaN
1	02/01/2023	Bananas	5	0.2	NaN
2	03/01/2023	Oranges	8	0.6	NaN
3	01/02/2023	Apples	12	0.5	NaN
4	05/02/2023	Bananas	7	0.2	NaN
5	10/03/2023	Apples	6	0.5	NaN
6	15/03/2023	Bananas	9	0.2	NaN

```
df.isna().sum()
```

	0
Date	0
Product	0
Quantity	0
Price	0
Total	7

dtype: int64

```
df = pd.read_csv('shopping.csv', parse_dates = ['Date'])
df
```

	Date	Product	Quantity	Price	Total
0	01/01/2023	Apples	10	0.5	NaN
1	02/01/2023	Bananas	5	0.2	NaN
2	03/01/2023	Oranges	8	0.6	NaN
3	01/02/2023	Apples	12	0.5	NaN
4	05/02/2023	Bananas	7	0.2	NaN
5	10/03/2023	Apples	6	0.5	NaN
6	15/03/2023	Bananas	9	0.2	NaN

```
df = pd.read_csv('shopping.csv', parse_dates = ['Date'], dayfirst= True)
df
```

	Date	Product	Quantity	Price	Total
0	2023-01-01	Apples	10	0.5	NaN
1	2023-01-02	Bananas	5	0.2	NaN
2	2023-01-03	Oranges	8	0.6	NaN
3	2023-02-01	Apples	12	0.5	NaN
4	2023-02-05	Bananas	7	0.2	NaN
5	2023-03-10	Apples	6	0.5	NaN
6	2023-03-15	Bananas	9	0.2	NaN

```
df['Date_original'] = df["Date"]
df
```

	Date	Product	Quantity	Price	Total	Date_original
0	2023-01-01	Apples	10	0.5	NaN	2023-01-01
1	2023-01-02	Bananas	5	0.2	NaN	2023-01-02
2	2023-01-03	Oranges	8	0.6	NaN	2023-01-03
3	2023-02-01	Apples	12	0.5	NaN	2023-02-01
4	2023-02-05	Bananas	7	0.2	NaN	2023-02-05
5	2023-03-10	Apples	6	0.5	NaN	2023-03-10
6	2023-03-15	Bananas	9	0.2	NaN	2023-03-15

```
df
```

	Date	Product	Quantity	Price	Total	Date_original
0	2023-01-01	Apples	10	0.5	NaN	2023-01-01
1	2023-01-02	Bananas	5	0.2	NaN	2023-01-02
2	2023-01-03	Oranges	8	0.6	NaN	2023-01-03
3	2023-02-01	Apples	12	0.5	NaN	2023-02-01
4	2023-02-05	Bananas	7	0.2	NaN	2023-02-05
5	2023-03-10	Apples	6	0.5	NaN	2023-03-10
6	2023-03-15	Bananas	9	0.2	NaN	2023-03-15

```
df['Month'] = df['Date_original'].dt.to_period('M')
```

```
df.isnull().sum()
```

	0
Date	0
Product	0
Quantity	0
Price	0
Total	7
Date_original	0
Month	0

dtype: int64

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7 entries, 0 to 6
Data columns (total 7 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   Date           7 non-null      datetime64[ns]
 1   Product        7 non-null      object
 2   Quantity       7 non-null      int64
 3   Price          7 non-null      float64
 4   Total          0 non-null      float64
 5   Date_original  7 non-null      datetime64[ns]
 6   Month          7 non-null      period[M]
dtypes: datetime64[ns](2), float64(2), int64(1), object(1), period[M](1)
memory usage: 520.0+ bytes
```

```
df.isnull().sum()
```

	0
Date	0
Product	0
Quantity	0
Price	0
Total	7
Date_original	0
Month	0

dtype: int64

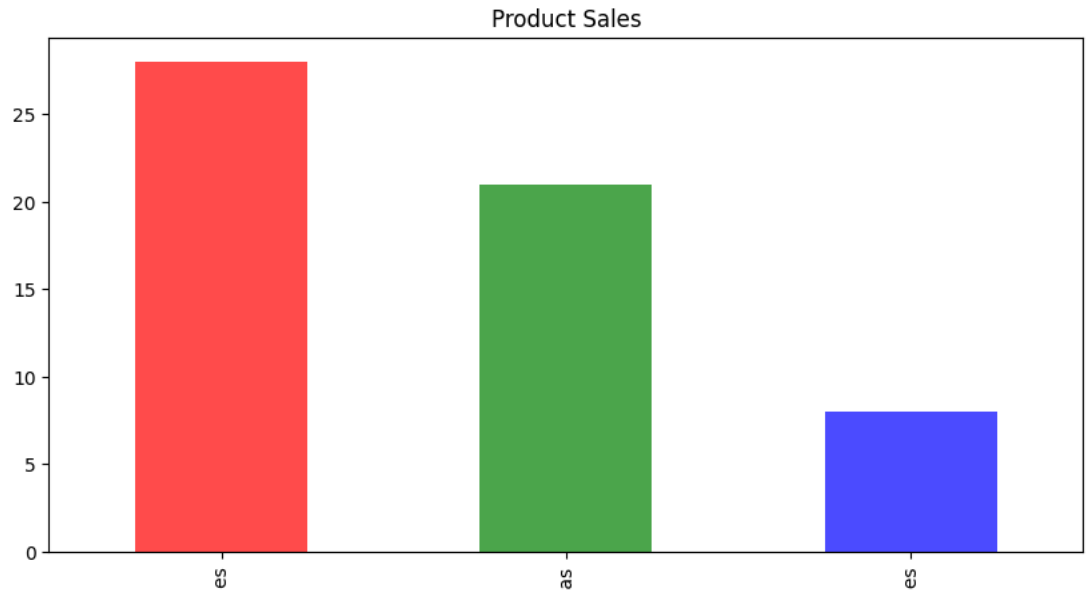
```
df['Total'] = df['Quantity'] * df['Price']
df
```

	Date	Product	Quantity	Price	Total	Date_original	Month
0	2023-01-01	Apples	10	0.5	5.0	2023-01-01	2023-01
1	2023-01-02	Bananas	5	0.2	1.0	2023-01-02	2023-01
2	2023-01-03	Oranges	8	0.6	4.8	2023-01-03	2023-01
3	2023-02-01	Apples	12	0.5	6.0	2023-02-01	2023-02
4	2023-02-05	Bananas	7	0.2	1.4	2023-02-05	2023-02
5	2023-03-10	Apples	6	0.5	3.0	2023-03-10	2023-03
6	2023-03-15	Bananas	9	0.2	1.8	2023-03-15	2023-03

```
df['Quantity'].fillna(df['Quantity'].mean())
df
```

	Date	Product	Quantity	Price	Total	Date_original	Month
0	2023-01-01	Apples	10	0.5	5.0	2023-01-01	2023-01
1	2023-01-02	Bananas	5	0.2	1.0	2023-01-02	2023-01
2	2023-01-03	Oranges	8	0.6	4.8	2023-01-03	2023-01
3	2023-02-01	Apples	12	0.5	6.0	2023-02-01	2023-02
4	2023-02-05	Bananas	7	0.2	1.4	2023-02-05	2023-02
5	2023-03-10	Apples	6	0.5	3.0	2023-03-10	2023-03
6	2023-03-15	Bananas	9	0.2	1.8	2023-03-15	2023-03

```
product_sales = df.groupby('Product')['Quantity'].sum()
color = ['red', 'green', 'blue']
product_sales.plot(kind = "bar", title = "Product Sales", color = color, figsize = (10,5), alpha = 0.7)


plt.show()
```



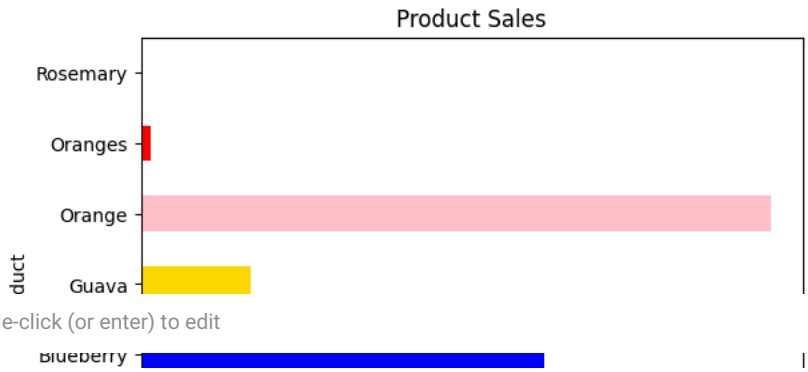

```
#ADD NEW PRODUCTS
df.loc[7,'Product'] = 'Orange'
df.loc[8,'Product'] = 'Orange'
df.loc[9,'Product'] = 'Orange'
df.loc[10,'Product'] = 'Orange'
df.loc[11,'Product'] = 'Orange'
df.loc[12,'Product'] = 'Blueberry'
df.loc[13,'Product'] = 'Rosemary'
df.loc[14,'Product'] = 'Guava'


#ADD NEW PRICES
df.loc[7,'Price'] = 2
df.loc[8,'Price'] =1
df.loc[9,'Price'] = 3
df.loc[10,'Price'] = 4
df.loc[11,'Price'] = 12
df.loc[12,'Price'] = 4.4
df.loc[13,'Price'] = 1
df.loc[14,'Price'] = 2.3

#ADD NEW QUANTITY
df.loc[7,'Quantity'] = 20
df.loc[8,'Quantity'] =12
df.loc[9,'Quantity'] = 30
df.loc[10,'Quantity'] = 4
df.loc[11,'Quantity'] = 12
df.loc[12,'Quantity'] = 44
df.loc[13,'Quantity'] = 1
df.loc[14,'Quantity'] = 23


df
```

	Date	Product	Quantity	Price	Total	Date_original	Month
0	2023-01-01	Apples	10.0	0.5	5.0	2023-01-01	2023-01
1	2023-01-02	Bananas	5.0	0.2	1.0	2023-01-02	2023-01
2	2023-01-03	Oranges	8.0	0.6	4.8	2023-01-03	2023-01
3	2023-02-01	Apples	12.0	0.5	6.0	2023-02-01	2023-02
4	2023-02-05	Bananas	7.0	0.2	1.4	2023-02-05	2023-02
5	2023-03-10	Apples	6.0	0.5	3.0	2023-03-10	2023-03
6	2023-03-15	Bananas	9.0	0.2	1.8	2023-03-15	2023-03
7	NaT	Orange	20.0	2.0	NaN	NaT	NaT
8	NaT	Orange	12.0	1.0	NaN	NaT	NaT
9	NaT	Orange	30.0	3.0	NaN	NaT	NaT
10	NaT	Orange	4.0	4.0	NaN	NaT	NaT
11	NaT	Orange	12.0	12.0	NaN	NaT	NaT
12	NaT	Blueberry	44.0	4.4	NaN	NaT	NaT
13	NaT	Rosemary	1.0	1.0	NaN	NaT	NaT
14	NaT	Guava	23.0	2.3	NaN	NaT	NaT

```
df['Total'] = df['Quantity'] * df['Price']
df
```

	Date	Product	Quantity	Price	Total	Date_original	Month
0	2023-01-01	Apples	10.0	0.5	5.0	2023-01-01	2023-01
1	2023-01-02	Bananas	5.0	0.2	1.0	2023-01-02	2023-01
2	2023-01-03	Oranges	8.0	0.6	4.8	2023-01-03	2023-01
3	2023-02-01	Apples	12.0	0.5	6.0	2023-02-01	2023-02
4	2023-02-05	Bananas	7.0	0.2	1.4	2023-02-05	2023-02
5	2023-03-10	Apples	6.0	0.5	3.0	2023-03-10	2023-03
6	2023-03-15	Bananas	9.0	0.2	1.8	2023-03-15	2023-03
7	NaT	Orange	20.0	2.0	40.0	NaT	NaT
8	NaT	Orange	12.0	1.0	12.0	NaT	NaT
9	NaT	Orange	30.0	3.0	90.0	NaT	NaT
10	NaT	Orange	4.0	4.0	16.0	NaT	NaT
11	NaT	Orange	12.0	12.0	144.0	NaT	NaT
12	NaT	Blueberry	44.0	4.4	193.6	NaT	NaT
13	NaT	Rosemary	1.0	1.0	1.0	NaT	NaT
14	NaT	Guava	23.0	2.3	52.9	NaT	NaT

```
product_sales = df.groupby("Product")['Total'].sum()
color = ['red', 'green', 'blue', 'gold', 'pink']
product_sales.plot(kind = "barh", title = "Product Sales", color = color  )
plt.show()
```



Double-click (or enter) to edit


```
#Plot a. total sales over time using a line chart, grouping by the month

df['Month'] = df['Date'].dt.to_period('M')
monthly_sales = df.groupby('Month')['Total'].sum()
monthly_sales.plot(kind = 'line')
plt.show()
```