

PROBLEMS

- How was the sales trend over the months?
- What are the most frequently purchased products?
- How many products does the customer purchase in each transaction?
- What are the most profitable segment customers?
- Based on your findings, what strategy could you recommend to the business to gain more profit?

Importing libraries

```
In [1]: import pandas as pd
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings('ignore')

import cufflinks as cf
from cufflinks.offline import download_plotlyjs, init_notebook_mode, plot, iplot

init_notebook_mode(connected=True)
cf.offline()
cf.set_config_title(theme='space', sharings='public', offline=True)
```

Data Loading

```
In [2]: datapd.read_csv("Sales Data.csv")

In [3]: data.head()
```

	TransactionNo	Date	ProductNo	ProductName	Price	Quantity	CustomerNo	Country
Out[3]:	0	581482	12/9/2019	22485	Set Of 2 Wooden Market Crates	21.47	12	17490.0 United Kingdom
	1	581475	12/9/2019	22596	Christmas Star Wish List Chalkboard	18.86	36	13069.0 United Kingdom
	2	581475	12/9/2019	23235	Storage Tin Vintage Lead	11.53	12	13069.0 United Kingdom
	3	581475	12/9/2019	23272	Tree T-Light Holder White Winkle	10.65	12	13069.0 United Kingdom
	4	581475	12/9/2019	23239	Set Of 4 Knock Knock Tin Poppies	11.94	6	13069.0 United Kingdom

```
In [4]: data.tail()
```

	TransactionNo	Date	ProductNo	ProductName	Price	Quantity	CustomerNo	Country
Out[4]:	536345	C536548	12/12/2018	22168	Organiser Wood Antique White	18.96	-2	12472.0 Germany
	536346	C536548	12/12/2018	23218	Red Spotty Blotcut Tin	14.09	-3	12472.0 Germany
	536347	C536548	12/12/2018	20967	Porcelain Hanging Bell Small	11.74	-1	12472.0 Germany
	536348	C536548	12/12/2018	22560	Advent Calendar Gingham Sack	16.35	-4	12472.0 Germany
	536349	C536548	12/12/2018	22767	True Photo Frame Cornice	20.45	-2	12472.0 Germany

```
In [5]: data.shape

Out[5]: (536358, 8)
```

```
In [6]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 536358 entries, 0 to 536349
Data columns (total 8 columns):
#   Column              Non-Null Count  Dtype
--  --
0   TransactionNo        536358 non-null  object
1   Date                 536358 non-null  object
2   ProductNo            536358 non-null  object
3   ProductName          536358 non-null  object
4   Price                536358 non-null  float64
5   Quantity             536358 non-null  int64
6   CustomerNo           536295 non-null  float64
7   Country              536358 non-null  object
dtypes: float64(2), int64(1), object(5)
memory usage: 32.7+ MB
```

Additional info

- TransactionNo (categorical): A six-digit unique number that defines each transaction. The letter "C" in the code indicates a cancellation.
- Date(numeric): The date when each transaction was generated.
- ProductNo (categorical): A five or six-digit unique character used to identify a specific product.
- Product (categorical): Productitem name.
- Price (numeric): The price of each product per unit in pound sterling.
- Quantity (numeric): The quantity of each product per transaction. Negative values related to cancelled transactions.
- CustomerNo (categorical): A five-digit unique number that defines each customer.
- Country (categorical): Name of the country where the customer resides.

Turn the Date column to DateTime type

```
In [7]: data['Date']=pd.to_datetime(data['Date'])

In [8]: data.head()
```

	TransactionNo	Date	ProductNo	ProductName	Price	Quantity	CustomerNo	Country
Out[8]:	0	581482	2019-12-09	22485	Set Of 2 Wooden Market Crates	21.47	12	17490.0 United Kingdom
	1	581475	2019-12-09	22596	Christmas Star Wish List Chalkboard	18.86	36	13069.0 United Kingdom
	2	581475	2019-12-09	23235	Storage Tin Vintage Lead	11.53	12	13069.0 United Kingdom
	3	581475	2019-12-09	23272	Tree T-Light Holder White Winkle	10.65	12	13069.0 United Kingdom
	4	581475	2019-12-09	23239	Set Of 4 Knock Knock Tin Poppies	11.94	6	13069.0 United Kingdom

```
In [9]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 536358 entries, 0 to 536349
Data columns (total 8 columns):
#   Column              Non-Null Count  Dtype
--  --
0   TransactionNo        536358 non-null  object
1   Date                 536358 non-null  datetime64[ns]
2   ProductNo            536358 non-null  object
3   ProductName          536358 non-null  object
4   Price                536358 non-null  float64
5   Quantity             536358 non-null  int64
6   CustomerNo           536295 non-null  float64
7   Country              536358 non-null  object
dtypes: datetime64[ns](1), float64(2), int64(1), object(4)
memory usage: 32.7+ MB
```

```
In [10]: data.describe()
```

	Price	Quantity	CustomerNo
count	536358.000000	536358.000000	536295.000000
mean	12.862382	9.809487	3822.989178
std	8.490480	216.662300	1716.952932
min	5.130000	-80995.000000	12004.000000
25%	10.900000	1.000000	1307.000000
50%	11.940000	3.000000	15152.000000
75%	14.090000	10.000000	16729.000000
max	660.620000	80995.000000	18287.000000

```
In [11]: data.isnull().sum()
```

```
Out[11]: TransactionNo      0
Date                  0
ProductNo             0
ProductName            0
Price                 0
Quantity              0
CustomerNo            85
Country               0
dtype: int64
```

```
In [12]: data[data['CustomerNo'].isnull()]

Out[12]:
```

	TransactionNo	Date	ProductNo	ProductName	Price	Quantity	CustomerNo	Country
Out[12]:	6531	C581406	2019-12-06	46000M	Polyester Filler Pad 45x45cm	6.19	-340	NaN United Kingdom
	6532	C581406	2019-12-06	46000S	Polyester Filler Pad 45x45cm	6.19	-300	NaN United Kingdom
	90098	C575153	2019-11-08	22947	Wooden Advent Calendar Red	44.25	-1	NaN United Kingdom
	102871	C574288	2019-11-03	22178	Victorian Glass Hanging T-Light	25.37	-1	NaN United Kingdom
	117263	C573180	2019-10-28	23048	Set Of 10 Lanterns Fairy Light Star	14.50	-1	NaN United Kingdom
	163160	C569415	2019-10-04	21843	Red Retrospot Cake Stand	21.47	-1	NaN United Kingdom
	190988	C567588	2019-09-20	22846	Bread Bin Orner Style Red	27.62	-1	NaN United Kingdom
	192294	C567518	2019-09-20	21871	Save The Planet Mug	11.94	-12	NaN United Kingdom
	242405	C563015	2019-08-11	46000M	Polyester Filler Pad 45x45cm	10.25	-160	NaN United Kingdom
	242406	C563015	2019-08-11	46000S	Polyester Filler Pad 45x45cm	10.25	-220	NaN United Kingdom
	243574	C562617	2019-08-08	23243	Set Of Tea Coffee Sugar Tins Pottery	15.32	-2	NaN United Kingdom
	254250	C561966	2019-08-01	22371	Airline Bag Vintage Tokyo TN	14.48	-1	NaN United Kingdom
	254254	C562030	2019-08-01	84595E	Large Tortilla Design Red Bowl	13.62	-1	NaN United Kingdom
	260633	C561239	2019-07-26	22427	Enamel Flower Jug Cream	16.35	-1	NaN United Kingdom
	260634	C561239	2019-07-26	23426	Enamel Wash Bowl Cream	14.09	-1	NaN United Kingdom
	260635	C561239	2019-07-26	84755	Colour Glass T-Light Holder Hanging	10.92	-4	NaN United Kingdom
	270366	C559949	2019-07-14	22099	Roses Regency Teacup And Saucer	13.71	-2	NaN United Kingdom
	285494	C559296	2019-07-07	23001	Travel Card Wallet Dotzomzhangtop	10.25	-200	NaN United Kingdom
	285495	C559296	2019-07-07	22618	Pack Of 12 London Tissues	10.25	-204	NaN United Kingdom
	285496	C559296	2019-07-07	22286	Spotty Starling	10.25	-170	NaN United Kingdom
	297959	C558584	2019-06-28	22667	Jam Making Set With Jar	18.75	-1	NaN United Kingdom
	298352	558245	2019-06-27	22734	Set Of 6 Ribbons Vintage Christmas	10.25	8	NaN United Kingdom
	301655	C567694	2019-06-21	22847	Bread Bin Orner Style Ivory	25.57	-1	NaN United Kingdom
	304635	C557664	2019-06-21	23183	Mother's Kitchen Spoon Rest	16.09	-1	NaN United Kingdom
	311943	C556939	2019-06-15	46000M	Polyester Filler Pad 45x45cm	10.25	-80	NaN United Kingdom
	311844	C556939	2019-06-15	46000S	Polyester Filler Pad 45x45cm	10.25	-160	NaN United Kingdom
	333735	C554716	2019-05-26	22662	Lunch Bag Dolly Girl Design	5.97	-1	NaN United Kingdom
	333736	C554716	2019-05-26	23206	Lunch Bag Apple Design	5.97	-1	NaN United Kingdom
	342808	C538362	2019-05-19	84945	Multi Colour Silver T-Light Holder	21.38	-1	NaN United Kingdom
	342953	C538364	2019-05-19	22623	Engines Design Rosette	11.53	-1	NaN United Kingdom
	342954	C538364	2019-05-19	22667	Bag 12g Saily Marden	10.86	-1	NaN United Kingdom
	342955	C538364	2019-05-19	23891	Yan For One Pinkstar	14.61	-1	NaN United Kingdom
	361482	C552230	2019-05-06	20765	Black Min Tape Measure	10.25	-1	NaN United Kingdom
	361483	C552230	2019-05-06	22617	Baking Set Spacecity Design	10.25	-1	NaN United Kingdom
	361484	C552230	2019-05-06	37327	Assol Metallic Colour Mug	10.25	-1	NaN United Kingdom
	361485	C552230	2019-05-06	84375	Set Of 20 Kids Cookie Cutters	10.25	-1	NaN United Kingdom
	361486	C552230	2019-05-06	84969	Box Of 6 Assorted Cook Teapoons	10.25	-1	NaN United Kingdom
	395771	C548805	2019-04-01	22453	Measuring Tape Babushka Blue	13.27	-1	NaN United Kingdom
	412419	C547091	2019-03-20	22766	Photo Frame Cornice	13.27	-12	NaN United Kingdom
	412420	C547091	2019-03-20	21316	Small Churny Glass Roman Bowl	13.27	-1	NaN United Kingdom
	412421	C547091	2019-03-20	22900	Antique Tail Swirlgess Trestle Pot	14.09	-4	NaN United Kingdom
	427674	C547008	2019-03-18	22267	20 Dolly Pops Retrospat	11.53	-1	NaN United Kingdom
	427675	C547059	2019-03-16	22212	Four Hook White Lovetools	12.40	-1	NaN United Kingdom
	429733	C545506	2019-03-03	22907	Pack Of 20 Napkins Party Design	11.12	-12	NaN United Kingdom
	454936	C543384	2019-02-07	48173C	Dormat Black Floor	18.40	-1	NaN United Kingdom
	455511	C543185	2019-02-04	22233	Retrospot Party Bag + Sticker Set	11.94	-14	NaN United Kingdom
	455512	C543185	2019-02-04	22332	Skulls Party Bag + Sticker Set	11.94	-14	NaN United Kingdom
	459432	C542860	2019-02-01	22580	Advent Calendar Gingham Sack	16.35	-1	NaN United Kingdom
	462407	C542540	2019-01-28	21058	Glass Beurre Dish	18.75	-1	NaN United Kingdom
	483337	C540854	2019-01-12	22461	Savvy Art Deco Clock	23.32	-1	NaN United Kingdom
	483650	C540599	2019-01-10	21888	Bingo Set	14.09	-4	NaN United Kingdom
	483651	C540185	2019-01-05	72828	Ocean Scott Candle In Jewellery Box	14.16	-54	NaN United Kingdom
	483701	C539948	2018-12-23	21868	Bingo Set	14.09	-4	NaN United Kingdom
	487950	C538756	2018-12-21	22720	Set Of 3 Cakes Tins Party Design	15.32	-1	NaN United Kingdom
	507547	C536986	2018-12-14	22467	Gumball Coat Rack	12.86	-1	NaN United Kingdom

All the null values have the same country and negative Quantity (except one, whose TransactionNo does not start with the letter 'C' and does not have a negative Quantity).

Let's drop all null and duplicated values

```
In [13]: data.dropna(inplace=True)
data.isnull().sum()
```

```
Out[13]: TransactionNo      0
Date                  0
ProductNo             0
ProductName            0
Price                 0
Quantity              0
CustomerNo            0
Country               0
dtype: int64
```

```
In [14]: data.duplicated().sum()
```

```
Out[14]: 5288
```

```
In [15]: data.drop_duplicates(inplace=True)
data.duplicated().sum()
```

```
Out[15]: 0
```

Answering the Questions

1. How was the sales trend over the months?

```
In [16]: data['Total']=data['Quantity']*data['Price']
total_Month_columns=[data['Date'],total]
total_Month_columns=[data['Date'],total]
data.head()
```

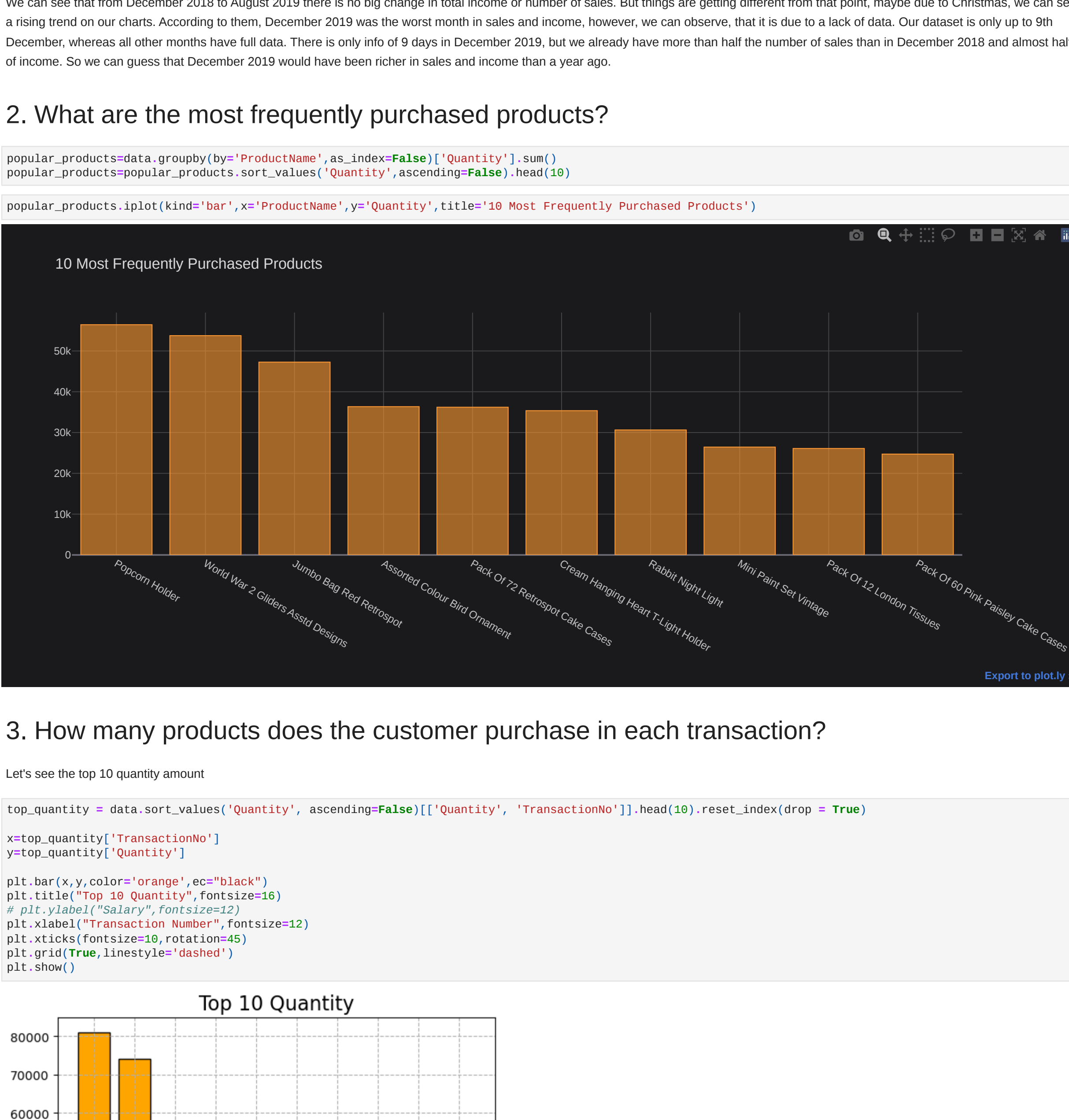
	TransactionNo	Date	ProductNo	ProductName	Price	Quantity	CustomerNo	Country	Total	Month_Year
Out[16]:	0	581482	2019-12-09	22485	Set Of 2 Wooden Market Crates	21.47	12	17490.0 United Kingdom	257.64	2019-12
	1	581475	2019-12-09	22596	Christmas Star Wish List Chalkboard	18.86	36	13069.0 United Kingdom	383.40	2019-12
	2	581475	2019-12-09	23235	Storage Tin Vintage Lead	11.53	12	13069.0 United Kingdom	138.36	2019-12
	3	581475	2019-12-09	23272	Tree T-Light Holder White Winkle	10.65	12	13069.0 United Kingdom	127.80	2019-12
	4	581475	2019-12-09	23239	Set Of 4 Knock Knock Tin Poppies	11.94	6	13069.0 United Kingdom	71.64	2019-12

```
In [17]: total_month=data.groupby(['Month_Year'],as_index=False)['Total'].sum()
total_month.head()
```

```
Out[17]:
```

	Date	Total
0	2019-12	4217381.53
1	2019-01	3638948.12
2	2019-02	3292324.69
3	2019-03	4345005.60
4	2019-04	3406002.17

```
In [18]: total_month.plot(kind='line',x='Date',y='Total',title='Sales Trend Over The Months',
ytitle='Total Money',xtitle='Month',mode='lines+markers',size=6)
```

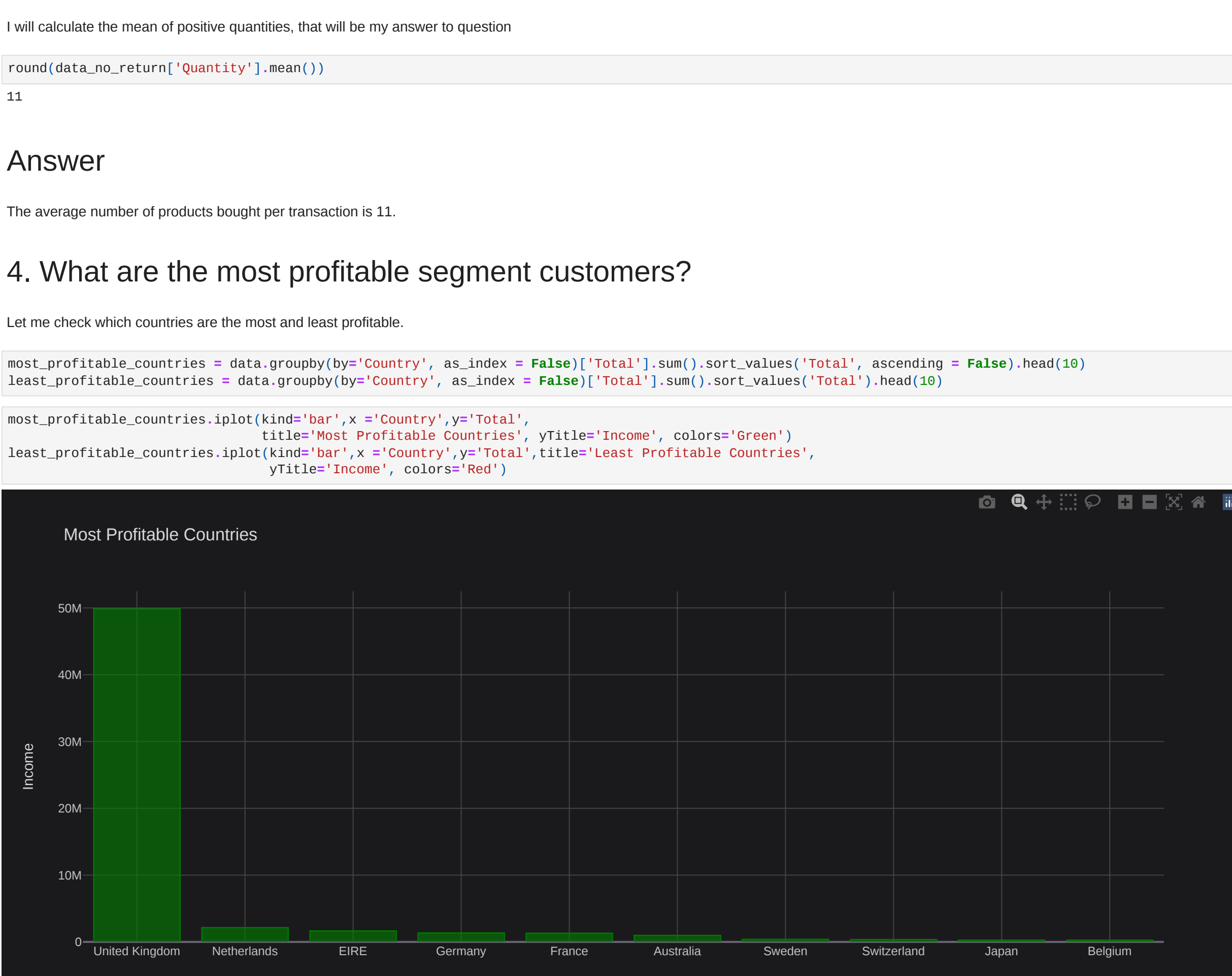


```
In [19]: data_no_return=data[data['Quantity']>0]
number_of_sales=data_no_return.groupby(by=['Month_Year'],as_index=False)['Total'].count()
number_of_sales
```

```
Out[19]:
```

	Month_Year	Total
0	2019-12	4217381.53
1	2019-01	3638948.12
2	2019-02	3292324.69
3	2019-03	3638948.12
4	2019-04	28707.00
5	2019-05	35756.00
6	2019-06	35633.00
7	2019-07	38138.00
8	2019-08	34092.00
9	2019-09	48835.00
10	2019-10	50008.00
11	2019-11	91810.00
12	2019-12	24616.00

```
In [20]: number_of_sales.columns=['date','Sum of Sales']
number_of_sales.plot(kind='line',x='date',y='Sum of Sales',title='Number of Sales Each Month',
ytitle='Number of Sales',xtitle='Month',mode='lines+markers',
size=6)
```



```
In [21]: data.sort_values('Date',ascending=False).head(5)
```

	TransactionNo	Date	ProductNo	ProductName	Price	Quantity	CustomerNo	Country	Total	Month_Year
Out[21]:	0	581482	2019-12-09	22485	Set Of 2 Wooden Market Crates	21.47	12	17490.0 United Kingdom	257.64	2019-12
	1076	581498	2019-12-09	21920	Jumbo Storage Bag Skulls	6.19	2	14480.0 United Kingdom	12.38	2019-12
	1086	581498	2019-12-09	22041	Record Frame 7" Single Size	6.19	2	14480.0 United Kingdom	12.38	2019-12
	1085	581498	2019-12-09	21895	Pack Of 20 Skull Paper Napkins	6.19	1	14480.0 United Kingdom	6.19	2019-12
	1084	581498	2019-12-09	21887	Pack Of 3 Cakes Tins Party Design	6.19	2	14480.0 United Kingdom	12.38	2019-12

Answer

We can see that from December 2019