This Project aims to analyze and answer business questions about 12 months of E-commerce sales data. The data contains hundreds of thousands of electronics store purchases broken down by month, product type, cost, purchase address, etc.

Data Exploration Analysis is executed to answer 5 high level business questions related to our data:

- What was the best month for sales? How much was earned that month?
- What city sold the most product?
- What time should we display advertisemens to maximize the likelihood of customer's buying product?
- What products are most often sold together?
- What product sold the most? Why do you think it sold the most?

First: Data Cleaning methods are applied to prepare my data

In [4]:
```
import pandas as pd
```

In [5]:
```
april = pd.read_csv('Sales_April_2019.csv')
```

In [7]:
```
april.head()
```

Out[7]:

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address
0	176558	USB-C Charging Cable	2	11.95	04/19/19 08:46	917 1st St, Dallas, TX 75001
1	NaN	NaN	NaN	NaN	NaN	NaN
2	176559	Bose SoundSport Headphones	1	99.99	04/07/19 22:30	682 Chestnut St, Boston, MA 02215
3	176560	Google Phone	1	600	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001
4	176560	Wired Headphones	1	11.99	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001

In [8]:
```
august = pd.read_csv('Sales_August_2019.csv')
```

In [9]:
```
august.head()
```

Out[9]:

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address
0	236670	Wired Headphones	2	11.99	08/31/19 22:21	359 Spruce St, Seattle, WA 98101
1	236671	Bose SoundSport Headphones	1	99.99	08/15/19 15:11	492 Ridge St, Dallas, TX 75001
2	236672	iPhone	1	700.0	08/06/19 14:40	149 7th St, Portland, OR 97035
3	236673	AA Batteries (4-pack)	2	3.84	08/29/19 20:59	631 2nd St, Los Angeles, CA 90001
4	236674	AA Batteries (4-pack)	2	3.84	08/15/19 19:53	736 14th St, New York City, NY 10001

In [10]:
```
dec = pd.read_csv('Sales_December_2019.csv')
```

In [11]:
```
dec.head()
```

Out[11]:

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address
0	295665	Macbook Pro Laptop	1	1700	12/30/19 00:01	136 Church St, New York City, NY 10001
1	295666	LG Washing Machine	1	600.0	12/29/19 07:03	562 2nd St, New York City, NY 10001
2	295667	USB-C Charging Cable	1	11.95	12/12/19 18:21	277 Main St, New York City, NY 10001
3	295668	27in FHD Monitor	1	149.99	12/22/19 15:13	410 6th St, San Francisco, CA 94016
4	295669	USB-C Charging Cable	1	11.95	12/18/19 12:38	43 Hill St, Atlanta, GA 30301

In [12]:
```
feb = pd.read_csv('Sales_February_2019.csv')
```

In [13]:
```
feb.head()
```

Out[13]:

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address
0	150502	iPhone	1	700	02/18/19 01:35	866 Spruce St, Portland, ME 04101
1	150503	AA Batteries (4-pack)	1	3.84	02/13/19 07:24	18 13th St, San Francisco, CA 94016
2	150504	27in 4K Gaming Monitor	1	389.99	02/18/19 09:46	52 6th St, New York City, NY 10001

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address
3	150505	Lightning Charging Cable	1	14.95	02/02/19 16:47	129 Cherry St, Atlanta, GA 30301
4	150506	AA Batteries (4-pack)	2	3.84	02/28/19 20:32	548 Lincoln St, Seattle, WA 98101

In [14]:
```
jan = pd.read_csv('Sales_January_2019.csv')
```

In [15]:
```
jan.head()
```

Out[15]:

	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

In [16]:
```
july= pd.read_csv('Sales_July_2019.csv')
```

In [17]:
```
july.head()
```

Out[17]:

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address
0	222910	Apple Airpods Headphones	1	150	07/26/19 16:51	389 South St, Atlanta, GA 30301
1	222911	Flatscreen TV	1	300	07/05/19 08:55	590 4th St, Seattle, WA 98101
2	222912	AA Batteries (4-pack)	1	3.84	07/29/19 12:41	861 Hill St, Atlanta, GA 30301
3	222913	AA Batteries (4-pack)	1	3.84	07/28/19 10:15	190 Ridge St, Atlanta, GA 30301
4	222914	AAA Batteries (4-pack)	5	2.99	07/31/19 02:13	824 Forest St, Seattle, WA 98101

In [18]:
```
june = pd.read_csv('Sales_June_2019.csv')
```

In [19]:
```
june.head()
```

Out[19]:

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address
0	209921	USB-C Charging Cable	1	11.95	06/23/19 19:34	950 Walnut St, Portland, ME 04101
1	209922	Macbook Pro Laptop	1	1700.0	06/30/19 10:05	80 4th St, San Francisco, CA 94016
2	209923	ThinkPad Laptop	1	999.99	06/24/19 20:18	402 Jackson St, Los Angeles, CA 90001
3	209924	27in FHD Monitor	1	149.99	06/05/19 10:21	560 10th St, Seattle, WA 98101
4	209925	Bose SoundSport Headphones	1	99.99	06/25/19 18:58	545 2nd St, San Francisco, CA 94016

In [20]:
```
march = pd.read_csv('Sales_March_2019.csv')
```

In [21]:
```
march.head()
```

Out[21]:

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address
0	162009	iPhone	1	700	03/28/19 20:59	942 Church St, Austin, TX 73301
1	162009	Lightning Charging Cable	1	14.95	03/28/19 20:59	942 Church St, Austin, TX 73301
2	162009	Wired Headphones	2	11.99	03/28/19 20:59	942 Church St, Austin, TX 73301
3	162010	Bose SoundSport Headphones	1	99.99	03/17/19 05:39	261 10th St, San Francisco, CA 94016
4	162011	34in Ultrawide Monitor	1	379.99	03/10/19 00:01	764 13th St, San Francisco, CA 94016

In [22]:
```
may = pd.read_csv('Sales_May_2019.csv')
```

In [23]:
```
may.head()
```

Out[23]:

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address
0	194095	Wired Headphones	1	11.99	05/16/19 17:14	669 2nd St, New York City, NY 10001
1	194096	AA Batteries (4-pack)	1	3.84	05/19/19 14:43	844 Walnut St, Dallas, TX 75001

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address
2	194097	27in FHD Monitor	1	149.99	05/24/19 11:36	164 Madison St, New York City, NY 10001
3	194098	Wired Headphones	1	11.99	05/02/19 20:40	622 Meadow St, Dallas, TX 75001
4	194099	AAA Batteries (4-pack)	2	2.99	05/11/19 22:55	17 Church St, Seattle, WA 98101

In [25]:
```
nov = pd.read_csv('Sales_November_2019.csv')
```

In [26]:
```
nov.head()
```

Out[26]:

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address
0	278797	Wired Headphones	1	11.99	11/21/19 09:54	46 Park St, New York City, NY 10001
1	278798	USB-C Charging Cable	2	11.95	11/17/19 10:03	962 Hickory St, Austin, TX 73301
2	278799	Apple Airpods Headphones	1	150.0	11/19/19 14:56	464 Cherry St, Los Angeles, CA 90001
3	278800	27in FHD Monitor	1	149.99	11/25/19 22:24	649 10th St, Seattle, WA 98101
4	278801	Bose SoundSport Headphones	1	99.99	11/09/19 13:56	522 Hill St, Boston, MA 02215

In [27]:
```
octob = pd.read_csv('Sales_October_2019.csv')
```

In [28]:
```
octob.head()
```

Out[28]:

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address
0	259358	34in Ultrawide Monitor	1	379.99	10/28/19 10:56	609 Cherry St, Dallas, TX 75001
1	259359	27in 4K Gaming Monitor	1	389.99	10/28/19 17:26	225 5th St, Los Angeles, CA 90001
2	259360	AAA Batteries (4-pack)	2	2.99	10/24/19 17:20	967 12th St, New York City, NY 10001
3	259361	27in FHD Monitor	1	149.99	10/14/19 22:26	628 Jefferson St, New York City, NY 10001
4	259362	Wired Headphones	1	11.99	10/07/19 16:10	534 14th St, Los Angeles, CA 90001

In [29]:
```
sep = pd.read_csv('Sales_September_2019.csv')
```

In [30]:
```
sep.head()
```

Out[30]:

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address
0	248151	AA Batteries (4-pack)	4	3.84	09/17/19 14:44	380 North St, Los Angeles, CA 90001
1	248152	USB-C Charging Cable	2	11.95	09/29/19 10:19	511 8th St, Austin, TX 73301
2	248153	USB-C Charging Cable	1	11.95	09/16/19 17:48	151 Johnson St, Los Angeles, CA 90001
3	248154	27in FHD Monitor	1	149.99	09/27/19 07:52	355 Hickory St, Seattle, WA 98101
4	248155	USB-C Charging Cable	1	11.95	09/01/19 19:03	125 5th St, Atlanta, GA 30301

Merging Sales Monthly data into one DataFrame

In [34]:
```
all_months_data = pd.DataFrame()
```

In [35]:
```
import os
```

In [37]:
```
files = [file for file in os.listdir('./Sales Data 2019')]
```

In [38]:
```
files
```

Out[38]:
```
['Sales_April_2019.csv',
 'Sales_August_2019.csv',
 'Sales_December_2019.csv',
 'Sales_February_2019.csv',
 'Sales_January_2019.csv',
 'Sales_July_2019.csv',
 'Sales_June_2019.csv',
 'Sales_March_2019.csv',
 'Sales_May_2019.csv',
 'Sales_November_2019.csv',
 'Sales_October_2019.csv',
 'Sales_September_2019.csv']
```

In [40]:
```
for file in files:
    df = pd.read_csv('./Sales Data 2019/' + file)
    all_months_data = pd.concat([all_months_data, df])
```

In [42]:
```
all_months_data.to_csv('concat_data.csv', index=False)
```

In [78]:
```
all_data2 = pd.read_csv('concat_data.csv')

all_data2.head()
```

Out[78]:

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address
0	176558	USB-C Charging Cable	2	11.95	04/19/19 08:46	917 1st St, Dallas, TX 75001
1	NaN	NaN	NaN	NaN	NaN	NaN
2	176559	Bose SoundSport Headphones	1	99.99	04/07/19 22:30	682 Chestnut St, Boston, MA 02215
3	176560	Google Phone	1	600	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001
4	176560	Wired Headphones	1	11.99	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001

In [81]:
```
all_data2=all_data2.dropna(how='all')
all_data2.head()
```

Out[81]:

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address
0	176558	USB-C Charging Cable	2	11.95	04/19/19 08:46	917 1st St, Dallas, TX 75001
2	176559	Bose SoundSport Headphones	1	99.99	04/07/19 22:30	682 Chestnut St, Boston, MA 02215
3	176560	Google Phone	1	600	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001
4	176560	Wired Headphones	1	11.99	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001
5	176561	Wired Headphones	1	11.99	04/30/19 09:27	333 8th St, Los Angeles, CA 90001

In [86]:
```
all_data2 = all_data2.drop_duplicates()
```

In [87]:
```
all_data2.dtypes
```

Out[87]:
```
Order ID            object
Product             object
Quantity Ordered    object
Price Each          object
Order Date          object
Purchase Address    object
```

```
Month             object
dtype: object
```

In [ ]:
```
all_data2
```

In [88]:
```
all_data2['Month'] = all_data2['Order Date'].str[0:2]
all_data2.head()
```

Out[88]:

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	Month
0	176558	USB-C Charging Cable	2	11.95	04/19/19 08:46	917 1st St, Dallas, TX 75001	04
2	176559	Bose SoundSport Headphones	1	99.99	04/07/19 22:30	682 Chestnut St, Boston, MA 02215	04
3	176560	Google Phone	1	600	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001	04
4	176560	Wired Headphones	1	11.99	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001	04
5	176561	Wired Headphones	1	11.99	04/30/19 09:27	333 8th St, Los Angeles, CA 90001	04

In [91]:
```
new_data = all_data2[all_data2['Order Date'].str[0:2] !='Or']
```

In [92]:
```
new_data['Month'] = new_data['Month'].astype('int32')
```

```
C:\Users\MOHMAE~1\AppData\Local\Temp/ipykernel_6472/3350714747.py:1: SettingWithCopyWarn
ing:
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
  new_data['Month'] = new_data['Month'].astype('int32')
```

In [93]:
```
new_data.head()
```

Out[93]:

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	Month
0	176558	USB-C Charging Cable	2	11.95	04/19/19 08:46	917 1st St, Dallas, TX 75001	4
2	176559	Bose SoundSport Headphones	1	99.99	04/07/19 22:30	682 Chestnut St, Boston, MA 02215	4
3	176560	Google Phone	1	600	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001	4

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	Month
4	176560	Wired Headphones	1	11.99	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001	4
5	176561	Wired Headphones	1	11.99	04/30/19 09:27	333 8th St, Los Angeles, CA 90001	4

Second: Exploratory Data Analysis

Question #1: What was the best month for sales?

In [96]:
```
new_data['Quantity Ordered'] = new_data['Quantity Ordered'].astype('int32')
new_data['Price Each'] = new_data['Price Each'].astype('float')
new_data['Revenue'] = new_data['Quantity Ordered'] * new_data['Price Each']
```

```
C:\Users\MOHMAE~1\AppData\Local\Temp/ipykernel_6472/119196687.py:1: SettingWithCopyWarni
ng:
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
  new_data['Quantity Ordered'] = new_data['Quantity Ordered'].astype('int32')
C:\Users\MOHMAE~1\AppData\Local\Temp/ipykernel_6472/119196687.py:2: SettingWithCopyWarni
ng:
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
  new_data['Price Each'] = new_data['Price Each'].astype('float')
C:\Users\MOHMAE~1\AppData\Local\Temp/ipykernel_6472/119196687.py:3: SettingWithCopyWarni
ng:
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
  new_data['Revenue'] = new_data['Quantity Ordered'] * new_data['Price Each']
```

In [97]:
```
new_data.head()
```

Out[97]:

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	Month	Revenue
0	176558	USB-C Charging Cable	2	11.95	04/19/19 08:46	917 1st St, Dallas, TX 75001	4	23.90
2	176559	Bose SoundSport Headphones	1	99.99	04/07/19 22:30	682 Chestnut St, Boston, MA 02215	4	99.99
3	176560	Google Phone	1	600.00	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001	4	600.00

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	Month	Revenue
4	176560	Wired Headphones	1	11.99	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001	4	11.99
5	176561	Wired Headphones	1	11.99	04/30/19 09:27	333 8th St, Los Angeles, CA 90001	4	11.99

In [107…
```
X = new_data.groupby('Month').sum()
```

In [108…
```
X['Revenue'].max()
```

Out[108…
```
4608295.7
```

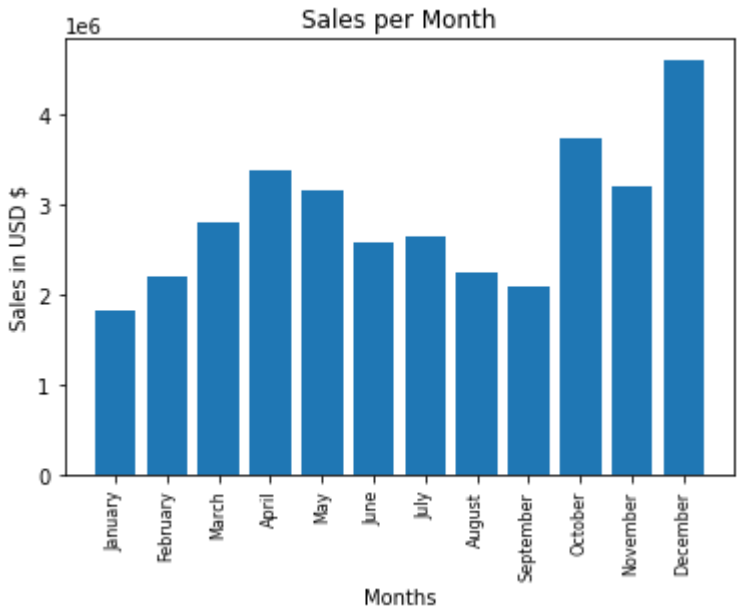
In [110…
```
X
```

Out[110…

Month	Quantity Ordered	Price Each	Revenue
1	10893	1810924.81	1821413.16
2	13431	2186940.38	2200078.08
3	16979	2789084.64	2804973.35
4	20536	3366218.76	3389217.98
5	18653	3133134.61	3150616.23
6	15234	2560503.50	2576280.15
7	16054	2631225.12	2646461.32
8	13429	2226963.90	2241083.37
9	13091	2081897.65	2094465.69
10	22669	3713608.80	3734777.86
11	19769	3178872.53	3197875.05
12	28074	4583267.77	4608295.70

In [192…
```
import matplotlib.pyplot as plt
import calendar
a = new_data['Month'].unique()
a.sort()
aa=[]
for i in a:
      aa.append(calendar.month_name[i])

plt.bar(aa, X['Revenue'])
plt.xticks(aa, rotation = 90, size=8)
plt.xlabel('Months')
plt.ylabel('Sales in USD $')
```

```
plt.title('Sales per Month')
plt.show()
```



Question #2: What city sold the most product?

In [174…

```
new_data['city name'] = new_data['Purchase Address'].apply(lambda x: x.split(',')[1] +
new_data.head()
```

```
C:\Users\MOHMAE~1\AppData\Local\Temp/ipykernel_6472/263891634.py:1: SettingWithCopyWarni
ng:
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
  new_data['city name'] = new_data['Purchase Address'].apply(lambda x: x.split(',')[1] +
' ' + '(' + x.split(',')[2].split(' ')[1] + ')')
```

Out[174…

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	Month	Revenue	City	city name
0	176558	USB-C Charging Cable	2	11.95	04/19/19 08:46	917 1st St, Dallas, TX 75001	4	23.90	TX	Dallas (TX)
2	176559	Bose SoundSport Headphones	1	99.99	04/07/19 22:30	682 Chestnut St, Boston, MA 02215	4	99.99	MA	Boston (MA)
3	176560	Google Phone	1	600.00	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001	4	600.00	CA	Los Angeles (CA)
4	176560	Wired Headphones	1	11.99	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001	4	11.99	CA	Los Angeles (CA)

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	Month	Revenue	City	city name
5	176561	Wired Headphones	1	11.99	04/30/19 09:27	333 8th St, Los Angeles, CA 90001	4	11.99	CA	Los Angeles (CA)

In [176…
```
b = new_data.groupby('city name').sum()
b
```

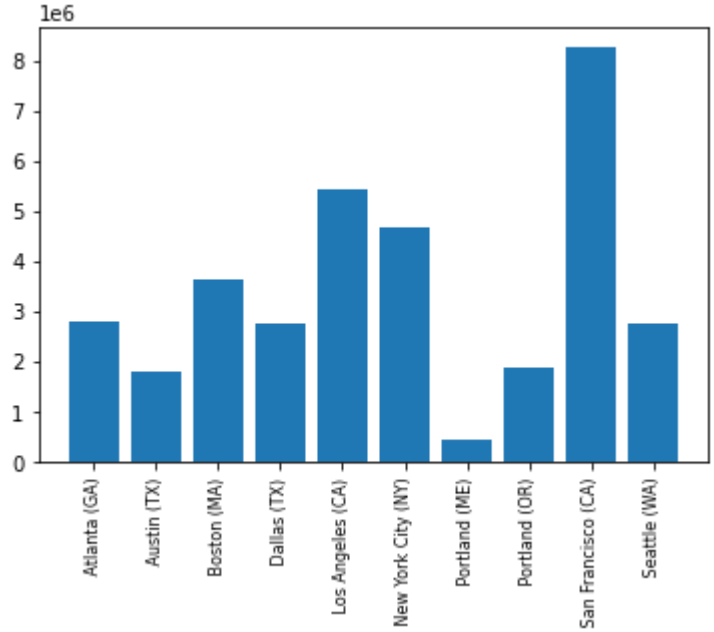
Out[176…

city name	Quantity Ordered	Price Each	Month	Revenue
Atlanta (GA)	16584	2778608.69	104649	2794199.07
Austin (TX)	11137	1808340.03	69720	1818044.33
Boston (MA)	22494	3634398.40	140898	3658627.65
Dallas (TX)	16707	2750026.38	104447	2765373.96
Los Angeles (CA)	33247	5417171.70	208020	5448304.28
New York City (NY)	27903	4632920.54	175557	4661867.14
Portland (ME)	2746	446752.36	17119	449321.38
Portland (OR)	11291	1859836.44	70517	1870010.56
San Francisco (CA)	50169	8204001.38	314949	8254743.55
Seattle (WA)	16534	2730586.55	104817	2745046.02

In [177…
```
c=new_data['city name'].unique()
c.sort()
c
```

Out[177…
```
array([' Atlanta (GA)', ' Austin (TX)', ' Boston (MA)', ' Dallas (TX)',
       ' Los Angeles (CA)', ' New York City (NY)', ' Portland (ME)',
       ' Portland (OR)', ' San Francisco (CA)', ' Seattle (WA)'],
      dtype=object)
```

In [179…
```
plt.bar(c, b['Revenue'])
plt.xticks(c, rotation = 'vertical', size =8)
plt.show()
```



Question #3: What time should we display advertisements to maximize the likelihood of purchases?

In [193…
```
new_data['Order Date'] = pd.to_datetime(new_data['Order Date'])
```

```
C:\Users\MOHMAE~1\AppData\Local\Temp/ipykernel_6472/473725514.py:1: SettingWithCopyWarni
ng:
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
  new_data['Order Date'] = pd.to_datetime(new_data['Order Date'])
```

In [194…
```
new_data.head()
```

Out[194…

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	Month	Revenue	City	city name
0	176558	USB-C Charging Cable	2	11.95	2019-04-19 08:46:00	917 1st St, Dallas, TX 75001	4	23.90	TX	Dallas (TX)
2	176559	Bose SoundSport Headphones	1	99.99	2019-04-07 22:30:00	682 Chestnut St, Boston, MA 02215	4	99.99	MA	Boston (MA)
3	176560	Google Phone	1	600.00	2019-04-12 14:38:00	669 Spruce St, Los Angeles, CA 90001	4	600.00	CA	Los Angeles (CA)
4	176560	Wired Headphones	1	11.99	2019-04-12 14:38:00	669 Spruce St, Los Angeles, CA 90001	4	11.99	CA	Los Angeles (CA)

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	Month	Revenue	City	city name
5	176561	Wired Headphones	1	11.99	2019-04-30 09:27:00	333 8th St, Los Angeles, CA 90001	4	11.99	CA	Los Angeles (CA)

In [195…
```
new_data['Hour'] = new_data['Order Date'].dt.hour
```

```
C:\Users\MOHMAE~1\AppData\Local\Temp/ipykernel_6472/3524801654.py:1: SettingWithCopyWarn
ing:
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
  new_data['Hour'] = new_data['Order Date'].dt.hour
```

In [196…
```
new_data.head()
```

Out[196…

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	Month	Revenue	City	city name	Hour
0	176558	USB-C Charging Cable	2	11.95	2019-04-19 08:46:00	917 1st St, Dallas, TX 75001	4	23.90	TX	Dallas (TX)	8
2	176559	Bose SoundSport Headphones	1	99.99	2019-04-07 22:30:00	682 Chestnut St, Boston, MA 02215	4	99.99	MA	Boston (MA)	22
3	176560	Google Phone	1	600.00	2019-04-12 14:38:00	669 Spruce St, Los Angeles, CA 90001	4	600.00	CA	Los Angeles (CA)	14
4	176560	Wired Headphones	1	11.99	2019-04-12 14:38:00	669 Spruce St, Los Angeles, CA 90001	4	11.99	CA	Los Angeles (CA)	14
5	176561	Wired Headphones	1	11.99	2019-04-30 09:27:00	333 8th St, Los Angeles, CA 90001	4	11.99	CA	Los Angeles (CA)	9

In [199…
```
new_data['Minutes'] = new_data['Order Date'].dt.minute
```

```
C:\Users\MOHMAE~1\AppData\Local\Temp/ipykernel_6472/4268269669.py:1: SettingWithCopyWarn
ing:
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
  new_data['Minutes'] = new_data['Order Date'].dt.minute
```

In [200…

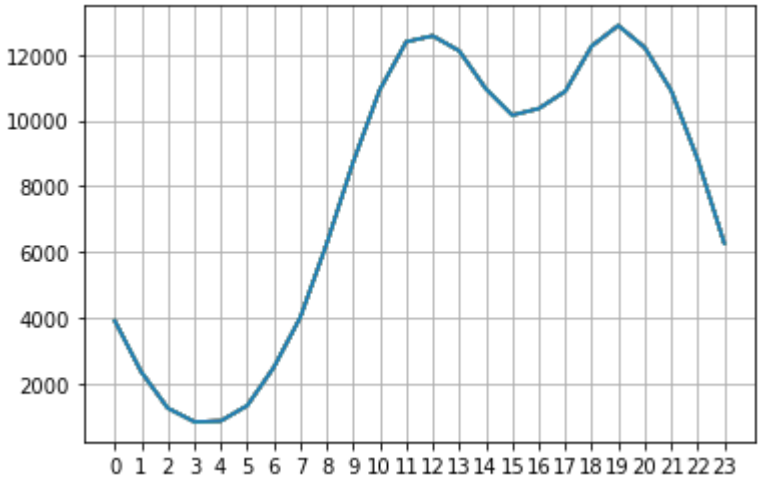
```
new_data.head()
```

Out[200…

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	Month	Revenue	City	city name	Hour	M
0	176558	USB-C Charging Cable	2	11.95	2019-04-19 08:46:00	917 1st St, Dallas, TX 75001	4	23.90	TX	Dallas (TX)	8	
2	176559	Bose SoundSport Headphones	1	99.99	2019-04-07 22:30:00	682 Chestnut St, Boston, MA 02215	4	99.99	MA	Boston (MA)	22	
3	176560	Google Phone	1	600.00	2019-04-12 14:38:00	669 Spruce St, Los Angeles, CA 90001	4	600.00	CA	Los Angeles (CA)	14	
4	176560	Wired Headphones	1	11.99	2019-04-12 14:38:00	669 Spruce St, Los Angeles, CA 90001	4	11.99	CA	Los Angeles (CA)	14	
5	176561	Wired Headphones	1	11.99	2019-04-30 09:27:00	333 8th St, Los Angeles, CA 90001	4	11.99	CA	Los Angeles (CA)	9	

In [211…

```
hour = range(0,24)
plt.plot(hour, new_data.groupby('Hour').count())
plt.xticks(hour)
plt.grid()
plt.show()
```



Question #4: What products are most often sold together?

In [212…

```
df = new_data[new_data['Order ID'].duplicated(keep=False)]
df['Bought together'] = df.groupby('Order ID')['Product'].transform(lambda x: ','.join(
```

```
C:\Users\MOHMAE~1\AppData\Local\Temp/ipykernel_6472/3688563883.py:2: SettingWithCopyWarn
ing:
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
  df['Bought together'] = df.groupby('Order ID')['Product'].transform(lambda x: ','.join
(x))
```

In [213…

```
df.head()
```

Out[213…

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	Month	Revenue	City	city name	Hour
3	176560	Google Phone	1	600.00	2019-04-12 14:38:00	669 Spruce St, Los Angeles, CA 90001	4	600.00	CA	Los Angeles (CA)	14
4	176560	Wired Headphones	1	11.99	2019-04-12 14:38:00	669 Spruce St, Los Angeles, CA 90001	4	11.99	CA	Los Angeles (CA)	14
18	176574	Google Phone	1	600.00	2019-04-03 19:42:00	20 Hill St, Los Angeles, CA 90001	4	600.00	CA	Los Angeles (CA)	19
19	176574	USB-C Charging Cable	1	11.95	2019-04-03 19:42:00	20 Hill St, Los Angeles, CA 90001	4	11.95	CA	Los Angeles (CA)	19

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	Month	Revenue	City	city name	Hour
32	176586	AAA Batteries (4-pack)	2	2.99	2019-04-10 17:00:00	365 Center St, San Francisco, CA 94016	4	5.98	CA	San Francisco (CA)	17

In [215…
```
df = df[['Order ID', 'Bought together']].drop_duplicates()
```

In [216…
```
df.head()
```

Out[216…

	Order ID	Bought together
3	176560	Google Phone,Wired Headphones
18	176574	Google Phone,USB-C Charging Cable
32	176586	AAA Batteries (4-pack),Google Phone
119	176672	Lightning Charging Cable,USB-C Charging Cable
129	176681	Apple Airpods Headphones,ThinkPad Laptop

In [219…
```
k=df.groupby('Bought together').count()
```

In [224…
```
j=k.sort_values('Order ID', ascending=False)
```

In [225…
```
j.head(10)
```

Out[225…

Bought together	Order ID
iPhone,Lightning Charging Cable	886
Google Phone,USB-C Charging Cable	857
iPhone,Wired Headphones	361
Vareebadd Phone,USB-C Charging Cable	312
Google Phone,Wired Headphones	303
iPhone,Apple Airpods Headphones	286
Google Phone,Bose SoundSport Headphones	161
Vareebadd Phone,Wired Headphones	104
Google Phone,USB-C Charging Cable,Wired Headphones	79
Vareebadd Phone,Bose SoundSport Headphones	60

Question #5: What product sold the most? Why do you think it did?

In [226…
```
new_data.head()
```

Out[226…

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	Month	Revenue	City	city name	Hour	M
0	176558	USB-C Charging Cable	2	11.95	2019-04-19 08:46:00	917 1st St, Dallas, TX 75001	4	23.90	TX	Dallas (TX)	8	
2	176559	Bose SoundSport Headphones	1	99.99	2019-04-07 22:30:00	682 Chestnut St, Boston, MA 02215	4	99.99	MA	Boston (MA)	22	
3	176560	Google Phone	1	600.00	2019-04-12 14:38:00	669 Spruce St, Los Angeles, CA 90001	4	600.00	CA	Los Angeles (CA)	14	
4	176560	Wired Headphones	1	11.99	2019-04-12 14:38:00	669 Spruce St, Los Angeles, CA 90001	4	11.99	CA	Los Angeles (CA)	14	
5	176561	Wired Headphones	1	11.99	2019-04-30 09:27:00	333 8th St, Los Angeles, CA 90001	4	11.99	CA	Los Angeles (CA)	9	

In [258…
```
v=new_data.groupby('Product').sum()
v
```

Out[258…

Product	Quantity Ordered	Price Each	Month	Revenue	Hour	Minutes
20in Monitor	4126	450739.02	29319	453818.74	58729	122096
27in 4K Gaming Monitor	6239	2427687.75	44396	2433147.61	90841	184129
27in FHD Monitor	7541	1124625.02	52484	1131074.59	107422	219659
34in Ultrawide Monitor	6192	2346058.26	43258	2352898.08	88957	183318
AA Batteries (4-pack)	27615	78942.72	145423	106041.60	298077	608560
AAA Batteries (4-pack)	30986	61629.88	146149	92648.14	296904	611350
Apple Airpods Headphones	15637	2328750.00	109304	2345550.00	222938	454935
Bose SoundSport Headphones	13430	1329667.02	93918	1342865.70	192077	391667

Product	Quantity Ordered	Price Each	Month	Revenue	Hour	Minutes
Flatscreen TV	4813	1438200.00	34172	1443900.00	68719	142629
Google Phone	5529	3313200.00	38286	3317400.00	79438	162678
LG Dryer	646	387600.00	4383	387600.00	9326	19043
LG Washing Machine	666	399600.00	4523	399600.00	9785	19462
Lightning Charging Cable	23169	323069.50	152741	346376.55	311815	632915
Macbook Pro Laptop	4725	8025700.00	33526	8032500.00	68225	137440
ThinkPad Laptop	4128	4125958.74	28926	4127958.72	59713	121447
USB-C Charging Cable	23931	261215.05	154520	285975.45	313980	646428
Vareebadd Phone	2068	826000.00	14309	827200.00	29472	61835
Wired Headphones	20524	225999.51	133123	246082.76	271245	552928
iPhone	6847	4788000.00	47933	4792900.00	98617	201604

In [230…

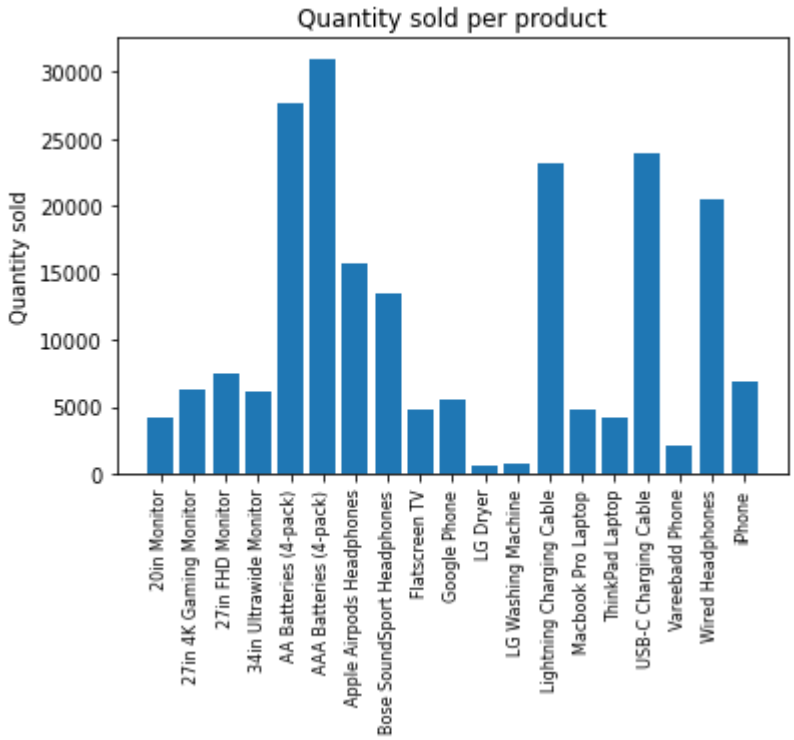
```
s= new_data['Product'].unique()
s.sort()
s
```

Out[230…

```
array(['20in Monitor', '27in 4K Gaming Monitor', '27in FHD Monitor',
       '34in Ultrawide Monitor', 'AA Batteries (4-pack)',
       'AAA Batteries (4-pack)', 'Apple Airpods Headphones',
       'Bose SoundSport Headphones', 'Flatscreen TV', 'Google Phone',
       'LG Dryer', 'LG Washing Machine', 'Lightning Charging Cable',
       'Macbook Pro Laptop', 'ThinkPad Laptop', 'USB-C Charging Cable',
       'Vareebadd Phone', 'Wired Headphones', 'iPhone'], dtype=object)
```

In [259…

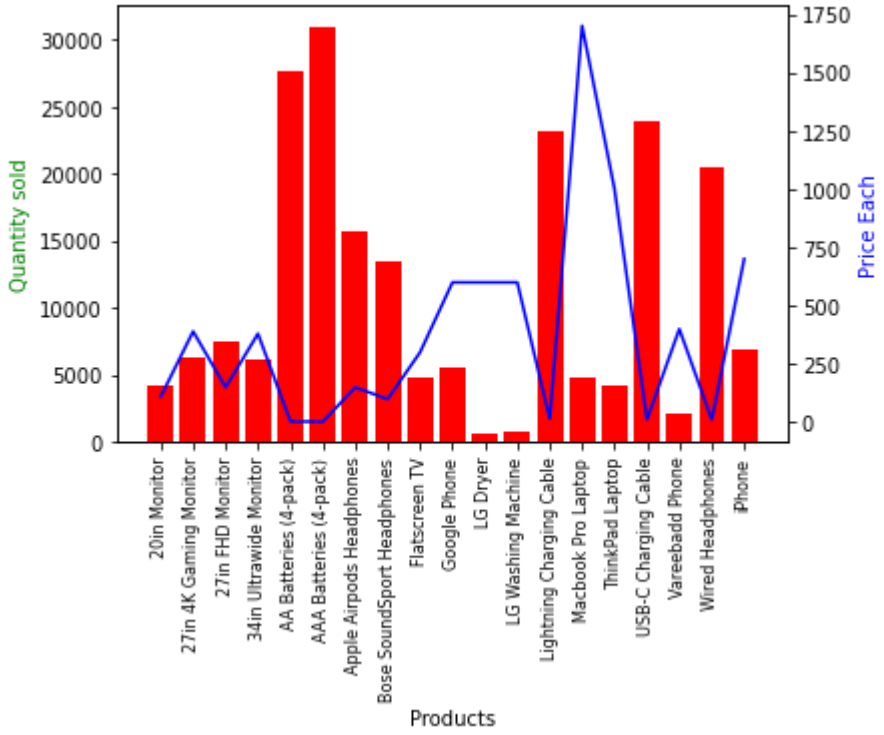
```
plt.bar(s, v['Quantity Ordered'])
plt.xticks(s, rotation=90, size=8)
plt.ylabel('Quantity sold')
plt.title('Quantity sold per product')
plt.show()
```



In [261…

```
p1 = new_data.groupby('Product').mean()

fig, ax1 = plt.subplots()
ax2 = ax1.twinx()
ax1.bar(s, v['Quantity Ordered'], color='r')
ax2.plot(s, p1['Price Each'],'b-')

ax1.set_xticklabels(s, rotation ='vertical', size=8)


ax1.set_xlabel('Products')
ax1.set_ylabel('Quantity sold', color = 'g')
ax2.set_ylabel('Price Each', color ='b')

plt.show()
```

```
C:\Users\MOHMAE~1\AppData\Local\Temp/ipykernel_6472/2137544820.py:8: UserWarning: FixedF
ormatter should only be used together with FixedLocator
  ax1.set_xticklabels(s, rotation ='vertical', size=8)
```



Results Section:

- What was the best month for sales? How much was earned that month?

  `December with sales = 4608295.7`
- What city sold the most product?

  `San Francisco (CA)`
- What time should we display advertisemens to maximize the likelihood of customer's buying product?

  `19 (7:00 pm)`
- What products are most often sold together?

  ```
  iPhone,Lightning Charging Cable
  Google Phone,USB-C Charging Cable
  iPhone,Wired Headphones
  Vareebadd Phone,USB-C Charging Cable
  Google Phone,Wired Headphones
  iPhone,Apple Airpods Headphones
  Google Phone,Bose SoundSport Headphones
  Vareebadd Phone,Wired Headphones
  Google Phone,USB-C Charging Cable,Wired Headphones
  Vareebadd Phone,Bose SoundSport Headphones
  ```
- What product sold the most? Why do you think it sold the most?

  `AAA Batteries (4-pack), Quantity sold increased for cheaper prices`

In [ ]: