Sales Analysis

**** analyze and answer business questions about 12 months worth of sales data.**

The data contains hundreds of thousands of electronics store purchases broken down by month, product type, cost, purchase address, etc.

We start by cleaning our data. Tasks during this section include:

- **Drop NaN values from DataFrame**
- **Removing rows based on a condition**
- **Change the type of columns (to_numeric, to_datetime, astype)**

Once we have cleaned up our data a bit, we move the data exploration section. In this section we explore 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? ****

In [ ]:

```
```

Import necessary libraries

In [1]:

```
import os
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

In [2]:

```
os.getcwd()
```

Out[2]:

```
'C:\\Users\\Ad\\WORKING\\EDA\\1. Sales Analysis-Pandas-Data-Science-Tasks\\SalesAnalysis'
```

In [7]:

```
os.listdir("C:\\Users\\Ad\\WORKING\\EDA\\1. Sales Analysis-Pandas-Data-Science-Tasks\\SalesAnalysis")
```

Out[7]:

```
['.ipynb_checkpoints',
 'Output',
 'SalesAnalysis - Copy.ipynb',
 'SalesAnalysis.ipynb',
 'Sales_Data',
 'Untitled.ipynb']
```

In [ ]:

```
```

Merge data from each month into one CSV

In [8]:

```
path="./Sales_Data/"
```

In [9]:

```
os.listdir(path)
```

Out[9]:

```
['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 [10]:

```
os.getcwd()
```

Out[10]:

```
'C:\\Users\\Ad\\WORKING\\EDA\\1. Sales Analysis-Pandas-Data-Science-Tasks\\SalesAnalysis'
```

In [20]:

```
pd.set_option('display.max_columns', 500)
path="./Sales_Data/"
files=[file for file in os.listdir(path) if not file.startswith('.')]# Ignore hidden fil
es

#creating an empty dataframe to concat all files
all_months_data= pd.DataFrame()

for file in files:
    current_data = pd.read_csv(path+"/"+file)
    print("*"*10)
    print(path+"/"+file)
    all_months_data=pd.concat([all_months_data,current_data])
    print("length after each iteration",len(all_months_data))
    print("*"*10)
    print()

all_months_data.to_csv("all_data.csv", index=False)
```

```
**********
./Sales_Data//Sales_April_2019.csv
length after each iteration 18383
**********

**********
./Sales_Data//Sales_August_2019.csv
length after each iteration 30394
**********

**********
./Sales_Data//Sales_December_2019.csv
length after each iteration 55511
**********

**********
./Sales_Data//Sales_February_2019.csv
length after each iteration 67547
**********
```

```
**********
./Sales_Data//Sales_January_2019.csv
length after each iteration 77270
**********

**********
./Sales_Data//Sales_July_2019.csv
length after each iteration 91641
**********

**********
./Sales_Data//Sales_June_2019.csv
length after each iteration 105263
**********

**********
./Sales_Data//Sales_March_2019.csv
length after each iteration 120489
**********

**********
./Sales_Data//Sales_May_2019.csv
length after each iteration 137124
**********

**********
./Sales_Data//Sales_November_2019.csv
length after each iteration 154785
**********

**********
./Sales_Data//Sales_October_2019.csv
length after each iteration 175164
**********

**********
./Sales_Data//Sales_September_2019.csv
length after each iteration 186850
**********
```

In [18]:

```
all_months_data.shape
```

Out[18]:

```
(186850, 6)
```

In [19]:

```
all_months_data.columns.values
```

Out[19]:

```
array(['Order ID', 'Product', 'Quantity Ordered', 'Price Each',
       'Order Date', 'Purchase Address'], dtype=object)
```

Read in updated dataframe

In [28]:

```
all_data = pd.read_csv("all_data.csv")
all_data.head()
```

Out[28]:

Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address

	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

Clean up the data!

The first step in this is figuring out what we need to clean. I have found in practice, that you find things you need to clean as you perform operations and get errors. Based on the error, you decide how you should go about cleaning the data

Drop rows of NAN

In [29]:

```
nan_df_any=all_data[all_data.isnull().any(axis=1)]
nan_df_any.shape
```

Out[29]:

```
(545, 6)
```

In [30]:

```
nan_df_all=all_data[all_data.isnull().all(axis=1)]
nan_df_all.shape
```

Out[30]:

```
(545, 6)
```

In [31]:

```
all_data.shape
```

Out[31]:

```
(186850, 6)
```

In [32]:

```
# Find NAN
#nan_df = all_data[all_data.isna().any(axis=1)]
#display(nan_df.head())

all_data = all_data.dropna(how='all')
all_data.head()
```

Out[32]:

	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

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

In [33]:

```
all_data.shape
```

Out[33]:

```
(186305, 6)
```

In [34]:

```
all_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 186305 entries, 0 to 186849
Data columns (total 6 columns):
Order ID            186305 non-null object
Product             186305 non-null object
Quantity Ordered    186305 non-null object
Price Each          186305 non-null object
Order Date          186305 non-null object
Purchase Address    186305 non-null object
dtypes: object(6)
memory usage: 9.9+ MB
```

All columns are in string format. We need to change the datatype as needed

In [35]:

```
all_data['Quantity Ordered']=pd.to_numeric(all_data['Quantity Ordered'])
```

```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
pandas\_libs\lib.pyx in pandas._libs.lib.maybe_convert_numeric()

ValueError: Unable to parse string "Quantity Ordered"

During handling of the above exception, another exception occurred:

ValueError                                Traceback (most recent call last)
<ipython-input-35-1444eb392605> in <module>()
----> 1 all_data['Quantity Ordered']=pd.to_numeric(all_data['Quantity Ordered'])

c:\users\ad\anaconda3\lib\site-packages\pandas\core\tools\numeric.py in to_numeric(arg, e
rrors, downcast)
    133             coerce_numeric = False if errors in ('ignore', 'raise') else True
    134             values = lib.maybe_convert_numeric(values, set(),
--> 135                                                coerce_numeric=coerce_numeric)
    136
    137     except Exception:

pandas\_libs\lib.pyx in pandas._libs.lib.maybe_convert_numeric()

ValueError: Unable to parse string "Quantity Ordered" at position 517
```

In [36]:

```
all_data['Quantity Ordered']=all_data['Quantity Ordered'].astype('int32')
```

```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
<ipython-input-36-2ee28a1cbe20> in <module>()
----> 1 all_data['Quantity Ordered']=all_data['Quantity Ordered'].astype('int32')

c:\users\ad\anaconda3\lib\site-packages\pandas\core\generic.py in astype(self, dtype, cop
y, errors, **kwargs)
   5689             # else, only a single dtype is given
   5690             new_data = self._data.astype(dtype=dtype, copy=copy, errors=errors,
-> 5691                                          **kwargs)
```

```
   5692             return self._constructor(new_data).__finalize__(self)
   5693
```

```
c:\users\ad\anaconda3\lib\site-packages\pandas\core\internals\managers.py in astype(self, dtype, **kwargs)
    529
    530     def astype(self, dtype, **kwargs):
--> 531         return self.apply('astype', dtype=dtype, **kwargs)
    532
    533     def convert(self, **kwargs):
```

```
c:\users\ad\anaconda3\lib\site-packages\pandas\core\internals\managers.py in apply(self, f, axes, filter, do_integrity_check, consolidate, **kwargs)
    393                                                 copy=align_copy)
    394
--> 395             applied = getattr(b, f)(**kwargs)
    396             result_blocks = _extend_blocks(applied, result_blocks)
    397
```

```
c:\users\ad\anaconda3\lib\site-packages\pandas\core\internals\blocks.py in astype(self, dtype, copy, errors, values, **kwargs)
    532     def astype(self, dtype, copy=False, errors='raise', values=None, **kwargs):
    533         return self._astype(dtype, copy=copy, errors=errors, values=values,
--> 534                             **kwargs)
    535
    536     def _astype(self, dtype, copy=False, errors='raise', values=None,
```

```
c:\users\ad\anaconda3\lib\site-packages\pandas\core\internals\blocks.py in _astype(self, dtype, copy, errors, values, **kwargs)
    631
    632                     # _astype_nansafe works fine with 1-d only
--> 633                     values = astype_nansafe(values.ravel(), dtype, copy=True)
    634
    635                 # TODO(extension)
```

```
c:\users\ad\anaconda3\lib\site-packages\pandas\core\dtypes\cast.py in astype_nansafe(arr, dtype, copy, skipna)
    681         # work around NumPy brokenness, #1987
    682         if np.issubdtype(dtype.type, np.integer):
--> 683             return lib.astype_intsafe(arr.ravel(), dtype).reshape(arr.shape)
    684
    685         # if we have a datetime/timedelta array of objects
```

```
pandas\_libs\lib.pyx in pandas._libs.lib.astype_intsafe()
```

```
ValueError: invalid literal for int() with base 10: 'Quantity Ordered'
```

In [37]:

```
all_data['Quantity Ordered'].unique()
```

Out[37]:

```
array(['2', '1', '3', '5', 'Quantity Ordered', '4', '7', '6', '8', '9'],
      dtype=object)
```

'Quantity Ordered' value is creating problem

In [38]:

```
all_data=all_data[all_data['Quantity Ordered'] !='Quantity Ordered']
all_data.shape
```

Out[38]:

```
(185950, 6)
```

Get rid of text in order date column

In [40]:

```
all_data = all_data[all_data['Order Date'].str[0:2]!='Or']
```

Make columns correct type

In [41]:

```
all_data['Quantity Ordered'] = pd.to_numeric(all_data['Quantity Ordered'])
all_data['Price Each'] = pd.to_numeric(all_data['Price Each'])
```

Augment data with additional columns

Add month column

In [42]:

```
all_data.head()
```

Out[42]:

	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	600.00	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 [43]:

```
all_data['Month'] = all_data['Order Date'].str[0:2]
all_data['Month'] = all_data['Month'].astype('int32')
all_data.head()
```

Out[43]:

	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	600.00	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001	4
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

Add month column (alternative method)

In [44]:

```
all_data['Month 2'] = pd.to_datetime(all_data['Order Date']).dt.month
```

```
all_data.head()
```

Out[44]:

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

Add city column

In [60]:

```
all_data['City_1']=all_data['Purchase Address'].apply(lambda x:x.split(',')[1])
all_data['City_2']=all_data['Purchase Address'].apply(lambda x:x.split(',')[2].split()[0
])
all_data['City']=all_data['City_1']+"("+all_data['City_2']+")"
all_data.drop(['City_1','City_2'],axis=1,inplace = True)
all_data.head()
```

Out[60]:

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

Add city column (alternative method)

In [61]:

```
def get_city(address):
    return address.split(",")[1].strip(" ")

def get_state(address):
    return address.split(",")[2].split(" ")[1]

all_data['City_2'] = all_data['Purchase Address'].apply(lambda x: f"{get_city(x)}  ({get_
state(x)})")
all_data.head()
```

Out[61]:

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	Month	Month 2	City	City_2
0	176558	USB-C Charging	2	11.95	04/19/19	917 1st St, Dallas, TX	4	4	Dallas(TX)	Dallas

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	Month	Month 2	City	City_2
0	176558	Cable	2	11.95	08:46	75001	4	4	Dallas(TX)	(TX)
2	176559	Bose SoundSport Headphones	1	99.99	04/07/19 22:30	682 Chestnut St, Boston, MA 02215	4	4	Boston(MA)	Boston (MA)
3	176560	Google Phone	1	600.00	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001	4	4	Los Angeles(CA)	Los Angeles (CA)
4	176560	Wired Headphones	1	11.99	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001	4	4	Los Angeles(CA)	Los Angeles (CA)
5	176561	Wired Headphones	1	11.99	04/30/19 09:27	333 8th St, Los Angeles, CA 90001	4	4	Los Angeles(CA)	Los Angeles (CA)

In [ ]:

```
```

Data Exploration!

Question 1: What was the best month for sales? How much was earned that month?

In [62]:

```
all_data['Sales'] = all_data['Quantity Ordered'].astype('int') * all_data['Price Each'].
astype('float')
```

In [63]:

```
all_data.groupby(['Month']).sum()
```

Out[63]:

Month	Quantity Ordered	Price Each	Month 2	Sales
1	10903	1.811768e+06	9709	1.822257e+06
2	13449	2.188885e+06	23950	2.202022e+06
3	17005	2.791208e+06	45459	2.807100e+06
4	20558	3.367671e+06	73116	3.390670e+06
5	18667	3.135125e+06	82830	3.152607e+06
6	15253	2.562026e+06	81324	2.577802e+06
7	16072	2.632540e+06	100051	2.647776e+06
8	13448	2.230345e+06	95688	2.244468e+06
9	13109	2.084992e+06	104589	2.097560e+06
10	22703	3.715555e+06	202820	3.736727e+06
11	19798	3.180601e+06	193303	3.199603e+06
12	28114	4.588415e+06	299808	4.613443e+06

In [66]:

```
all_data.groupby(['Month']).sum().sort_values(by='Sales',ascending=False)
```

Out[66]:

Month	Quantity Ordered	Price Each	Month 2	Sales
12	28114	4.588415e+06	299808	4.613443e+06

Month	Quantity Ordered	Price Each	Month 2	Sales
10	22703	3.715554e+06	202820	3.736727e+06
4	20558	3.367671e+06	73116	3.390670e+06
11	19798	3.180601e+06	193303	3.199603e+06
5	18667	3.135125e+06	82830	3.152607e+06
3	17005	2.791208e+06	45459	2.807100e+06
7	16072	2.632540e+06	100051	2.647776e+06
6	15253	2.562026e+06	81324	2.577802e+06
8	13448	2.230345e+06	95688	2.244468e+06
2	13449	2.188885e+06	23950	2.202022e+06
9	13109	2.084992e+06	104589	2.097560e+06
1	10903	1.811768e+06	9709	1.822257e+06

To overcome the scientific notationissue for float values

In [70]:

```
#https://re-thought.com/how-to-suppress-scientific-notation-in-pandas/
pd.set_option('display.float_format', lambda x: '%.2f' % x)
all_data.groupby(['Month']).sum().sort_values(by='Sales',ascending=False)
```

Out[70]:

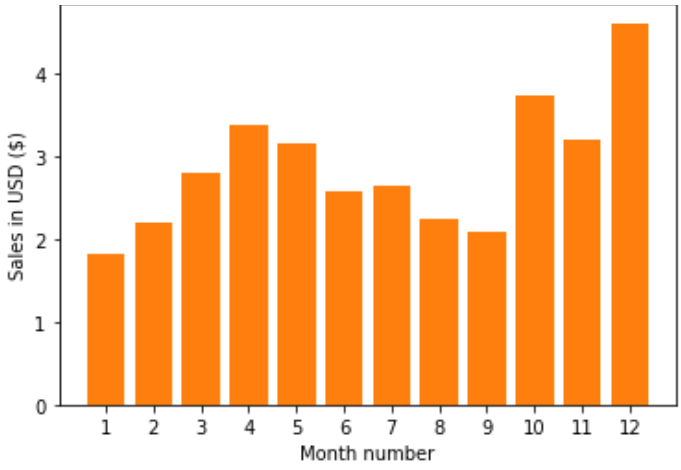
Month	Quantity Ordered	Price Each	Month 2	Sales
12	28114	4588415.41	299808	4613443.34
10	22703	3715554.83	202820	3736726.88
4	20558	3367671.02	73116	3390670.24
11	19798	3180600.68	193303	3199603.20
5	18667	3135125.13	82830	3152606.75
3	17005	2791207.83	45459	2807100.38
7	16072	2632539.56	100051	2647775.76
6	15253	2562025.61	81324	2577802.26
8	13448	2230345.42	95688	2244467.88
2	13449	2188884.72	23950	2202022.42
9	13109	2084992.09	104589	2097560.13
1	10903	1811768.38	9709	1822256.73

Maximum sales happenedin Dec and sale amount is 4613443.34

In [72]:

```
import matplotlib.pyplot as plt

months = range(1,13)
print(months)

plt.bar(months,all_data.groupby(['Month']).sum()['Sales'])
plt.xticks(months)
plt.ylabel('Sales in USD ($)')
plt.xlabel('Month number')
plt.show()
```

```
range(1, 13)
```



Question 2: What city sold the most product?

In [75]:

```
pd.set_option('display.float_format', lambda x: '%.2f' % x)
all_data.groupby(['City']).sum()['Sales']
```

Out[75]:

```
City
 Atlanta(GA)          2795498.58
 Austin(TX)           1819581.75
 Boston(MA)           3661642.01
 Dallas(TX)           2767975.40
 Los Angeles(CA)      5452570.80
 New York City(NY)    4664317.43
 Portland(ME)          449758.27
 Portland(OR)         1870732.34
 San Francisco(CA)    8262203.91
 Seattle(WA)          2747755.48
Name: Sales, dtype: float64
```

In [82]:

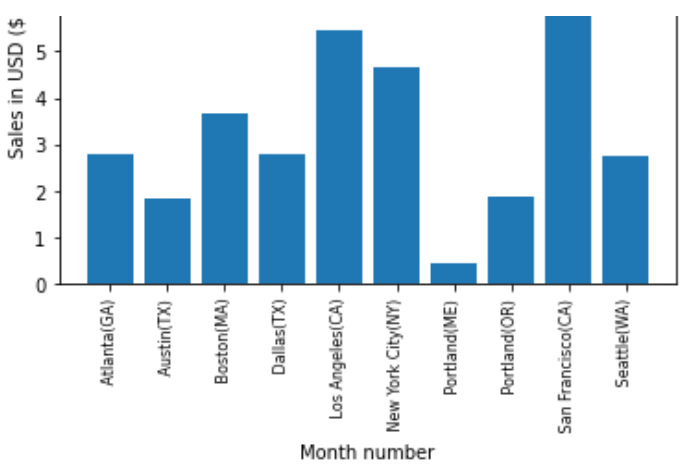
```
#we see above that Citynames are arranged inascending order of name
City=sorted(all_data['City'].unique())
City
```

Out[82]:

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

In [83]:

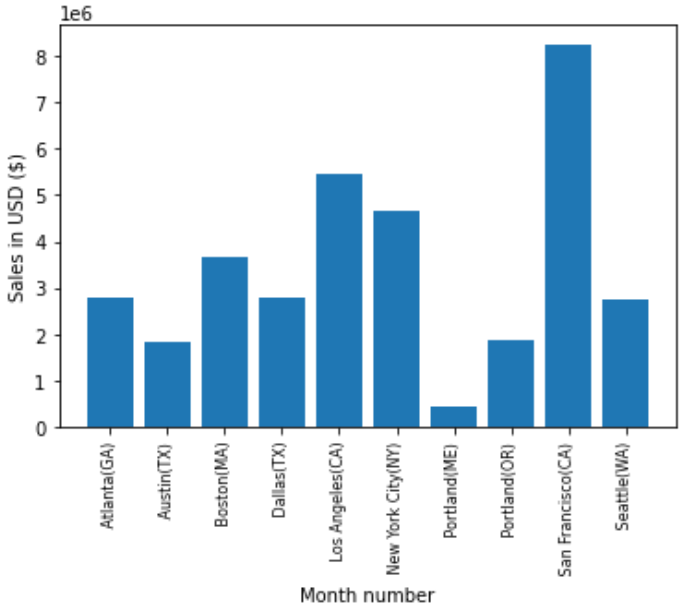
```
keys=City
plt.bar(keys,all_data.groupby(['City']).sum()['Sales'])
plt.xticks(keys, rotation='vertical', size=8)
plt.ylabel('Sales in USD ($)')
plt.xlabel('Month number')
plt.show()
```



altrnate way

In [84]:

```
import matplotlib.pyplot as plt

keys = [city for city, df in all_data.groupby(['City'])]

plt.bar(keys,all_data.groupby(['City']).sum()['Sales'])
plt.ylabel('Sales in USD ($)')
plt.xlabel('Month number')
plt.xticks(keys, rotation='vertical', size=8)
plt.show()
```



Question 3: What time should we display advertisements to maximize likelihood of customer's buying product?

In [85]:

```
all_data.head()
```

Out[85]:

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	Month	Month 2	City	City_2	Sales
0	176558	USB-C Charging Cable	2	11.95	04/19/19 08:46	917 1st St, Dallas, TX 75001	4	4	Dallas(TX)	Dallas (TX)	23.90
2	176559	Bose SoundSport Headphones	1	99.99	04/07/19 22:30	682 Chestnut St, Boston, MA 02215	4	4	Boston(MA)	Boston (MA)	99.99
										Los	

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	Month	Month 2	City	City_2	Sales
3	176560	Google Phone	1	600.00	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001	4	4	Los Angeles(CA)	Los Angeles (CA)	600.00
4	176560	Wired Headphones	1	11.99	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001	4	4	Los Angeles(CA)	Los Angeles (CA)	11.99
5	176561	Wired Headphones	1	11.99	04/30/19 09:27	333 8th St, Los Angeles, CA 90001	4	4	Los Angeles(CA)	Los Angeles (CA)	11.99

In [86]:

```
# Add hour column
all_data['Hour'] = pd.to_datetime(all_data['Order Date']).dt.hour
all_data['Minute'] = pd.to_datetime(all_data['Order Date']).dt.minute
all_data['Count'] = 1
all_data.head()
```

Out[86]:

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	Month	Month 2	City	City_2	Sales	Hour	Minute	C
0	176558	USB-C Charging Cable	2	11.95	04/19/19 08:46	917 1st St, Dallas, TX 75001	4	4	Dallas(TX)	Dallas (TX)	23.90	8	46	
2	176559	Bose SoundSport Headphones	1	99.99	04/07/19 22:30	682 Chestnut St, Boston, MA 02215	4	4	Boston(MA)	Boston (MA)	99.99	22	30	
3	176560	Google Phone	1	600.00	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001	4	4	Los Angeles(CA)	Los Angeles (CA)	600.00	14	38	
4	176560	Wired Headphones	1	11.99	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001	4	4	Los Angeles(CA)	Los Angeles (CA)	11.99	14	38	
5	176561	Wired Headphones	1	11.99	04/30/19 09:27	333 8th St, Los Angeles, CA 90001	4	4	Los Angeles(CA)	Los Angeles (CA)	11.99	9	27	

In [87]:

```
pd.set_option('display.float_format', lambda x: '%.2f' % x)
all_data.groupby(['Hour']).sum()['Count']
```

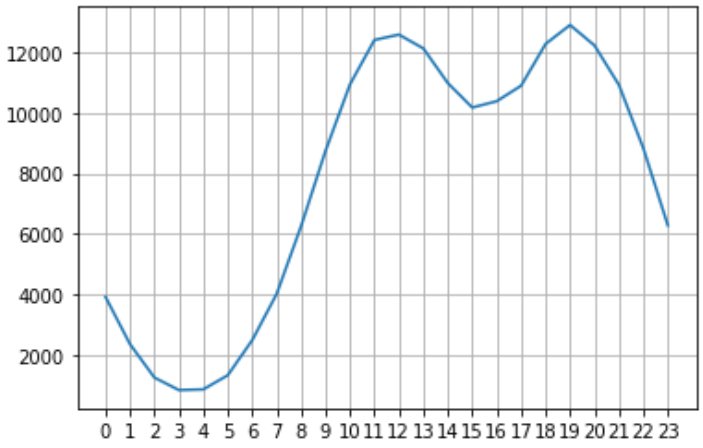
Out[87]:

```
Hour
0       3910
1       2350
2       1243
3        831
4        854
5       1321
6       2482
7       4011
8       6256
9       8748
10     10944
```

```
11    12411
12    12587
13    12129
14    10984
15    10175
16    10384
17    10899
18    12280
19    12905
20    12228
21    10921
22     8822
23     6275
Name: Count, dtype: int64
```

In [89]:

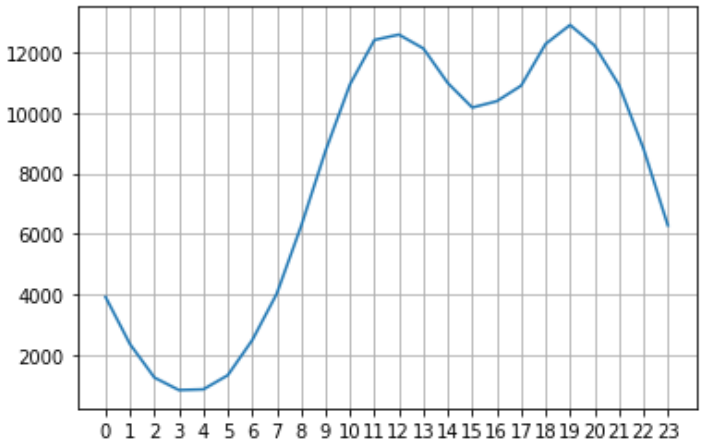
```
keys=sorted(all_data['Hour'].unique())
plt.plot(keys, all_data.groupby(['Hour']).count()['Count'])
plt.xticks(keys)
plt.grid()
plt.show()
```



alternate way

In [90]:

```
keys = [pair for pair, df in all_data.groupby(['Hour'])]

plt.plot(keys, all_data.groupby(['Hour']).count()['Count'])
plt.xticks(keys)
plt.grid()
plt.show()

# My recommendation is slightly before 11am or 7pm
```



Question 4: What products are most often sold together?

In [95]:

```
#https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.duplicated.html
all_data[all_data['Order ID'].duplicated(keep='first')]
```

Out[95]:

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	Month	Month 2	City	City_2	Sales	Hour
4	176560	Wired Headphones	1	11.99	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001	4	4	Los Angeles(CA)	Los Angeles (CA)	11.99	14
19	176574	USB-C Charging Cable	1	11.95	04/03/19 19:42	20 Hill St, Los Angeles, CA 90001	4	4	Los Angeles(CA)	Los Angeles (CA)	11.95	19
31	176585	Bose SoundSport Headphones	1	99.99	04/07/19 11:31	823 Highland St, Boston, MA 02215	4	4	Boston(MA)	Boston (MA)	99.99	11
33	176586	Google Phone	1	600.00	04/10/19 17:00	365 Center St, San Francisco, CA 94016	4	4	San Francisco(CA)	San Francisco (CA)	600.00	17
120	176672	USB-C Charging Cable	1	11.95	04/12/19 11:07	778 Maple St, New York City, NY 10001	4	4	New York City(NY)	New York City (NY)	11.95	11
130	176681	ThinkPad Laptop	1	999.99	04/20/19 10:39	331 Cherry St, Seattle, WA 98101	4	4	Seattle(WA)	Seattle (WA)	999.99	10
139	176689	AAA Batteries (4-pack)	2	2.99	04/24/19 17:15	659 Lincoln St, New York City, NY 10001	4	4	New York City(NY)	New York City (NY)	5.98	17
190	176739	Google Phone	1	600.00	04/05/19 17:38	730 6th St, Austin, TX 73301	4	4	Austin(TX)	Austin (TX)	600.00	17
226	176774	USB-C Charging Cable	1	11.95	04/25/19 15:06	372 Church St, Los Angeles, CA 90001	4	4	Los Angeles(CA)	Los Angeles (CA)	11.95	15
234	176781	Lightning Charging Cable	1	14.95	04/03/19 07:37	976 Hickory St, Dallas, TX 75001	4	4	Dallas(TX)	Dallas (TX)	14.95	7
251	176797	Bose SoundSport Headphones	1	99.99	04/21/19 08:54	923 Elm St, Los Angeles, CA 90001	4	4	Los Angeles(CA)	Los Angeles (CA)	99.99	8
252	176797	Wired Headphones	1	11.99	04/21/19 08:54	923 Elm St, Los Angeles, CA 90001	4	4	Los Angeles(CA)	Los Angeles (CA)	11.99	8
261	176805	USB-C Charging Cable	1	11.95	04/01/19 15:50	91 Lincoln St, Portland, OR 97035	4	4	Portland(OR)	Portland (OR)	11.95	15
265	176808	Wired Headphones	1	11.99	04/28/19 18:03	933 Meadow St, San Francisco, CA 94016	4	4	San Francisco(CA)	San Francisco (CA)	11.99	18
271	176813	Wired Headphones	1	11.99	04/28/19 18:01	269 Hill St, Atlanta, GA	4	4	Atlanta(GA)	Atlanta (GA)	11.99	18

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	Month	Month 2	City	City_2	Sales	Hour
		Headphones				30301				(GA)		
395	176935	27in FHD Monitor	1	149.99	04/03/19 21:31	315 1st St, Dallas, TX 75001	4	4	Dallas(TX)	Dallas (TX)	149.99	21
436	176975	AAA Batteries (4-pack)	1	2.99	04/23/19 15:46	28 13th St, San Francisco, CA 94016	4	4	San Francisco(CA)	San Francisco (CA)	2.99	15
451	176989	USB-C Charging Cable	1	11.95	04/09/19 13:43	346 9th St, Los Angeles, CA 90001	4	4	Los Angeles(CA)	Los Angeles (CA)	11.95	13
456	176993	Wired Headphones	1	11.99	04/07/19 09:43	28 South St, Los Angeles, CA 90001	4	4	Los Angeles(CA)	Los Angeles (CA)	11.99	9
486	177022	Wired Headphones	1	11.99	04/17/19 18:14	216 Chestnut St, Los Angeles, CA 90001	4	4	Los Angeles(CA)	Los Angeles (CA)	11.99	18
568	177102	27in 4K Gaming Monitor	1	389.99	04/24/19 05:19	542 Center St, New York City, NY 10001	4	4	New York City(NY)	New York City (NY)	389.99	5
582	177115	Lightning Charging Cable	1	14.95	04/19/19 19:10	71 6th St, San Francisco, CA 94016	4	4	San Francisco(CA)	San Francisco (CA)	14.95	19
585	177117	AAA Batteries (4-pack)	1	2.99	04/16/19 19:34	564 Hill St, San Francisco, CA 94016	4	4	San Francisco(CA)	San Francisco (CA)	2.99	19
636	177167	Apple Airpods Headphones	1	150.00	04/28/19 16:40	39 Lakeview St, Boston, MA 02215	4	4	Boston(MA)	Boston (MA)	150.00	16
637	177167	AAA Batteries (4-pack)	3	2.99	04/28/19 16:40	39 Lakeview St, Boston, MA 02215	4	4	Boston(MA)	Boston (MA)	8.97	16
649	177178	Lightning Charging Cable	1	14.95	04/07/19 00:02	342 Dogwood St, Seattle, WA 98101	4	4	Seattle(WA)	Seattle (WA)	14.95	0
653	177181	Apple Airpods Headphones	1	150.00	04/09/19 00:49	309 8th St, Los Angeles, CA 90001	4	4	Los Angeles(CA)	Los Angeles (CA)	150.00	0
655	177182	ThinkPad Laptop	1	999.99	04/10/19 22:00	332 10th St, Los Angeles, CA 90001	4	4	Los Angeles(CA)	Los Angeles (CA)	999.99	22
658	177184	Flatscreen TV	1	300.00	04/01/19 16:49	289 Hill St, Atlanta, GA 30301	4	4	Atlanta(GA)	Atlanta (GA)	300.00	16
673	177198	USB-C Charging Cable	2	11.95	04/06/19 16:07	154 6th St, New York City, NY 10001	4	4	New York City(NY)	New York City (NY)	23.90	16
...	...	...	...	...	...	...	...	...	...	...	...	...
186173	258715	Lightning Charging Cable	1	14.95	09/15/19 16:50	550 10th St, Portland, OR 97035	9	9	Portland(OR)	Portland (OR)	14.95	16
						146						

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	Month	Month 2	City	City_2	Sales	Hour
186190	258791	Lightning Charging Cable	1	14.95	09/20/19 13:00	146 Jackson St, Los Angeles, CA 90001	9	9	Los Angeles(CA)	Los Angeles (CA)	14.95	13
186215	258755	Lightning Charging Cable	1	14.95	09/03/19 21:41	17 Highland St, New York City, NY 10001	9	9	New York City(NY)	New York City (NY)	14.95	21
186220	258759	USB-C Charging Cable	1	11.95	09/01/19 14:00	344 4th St, Boston, MA 02215	9	9	Boston(MA)	Boston (MA)	11.95	14
186226	258764	Apple Airpods Headphones	1	150.00	09/04/19 15:18	691 Hill St, New York City, NY 10001	9	9	New York City(NY)	New York City (NY)	150.00	15
186270	258807	Lightning Charging Cable	1	14.95	09/07/19 18:12	809 Willow St, Seattle, WA 98101	9	9	Seattle(WA)	Seattle (WA)	14.95	18
186312	258848	USB-C Charging Cable	1	11.95	09/13/19 18:59	340 South St, Portland, OR 97035	9	9	Portland(OR)	Portland (OR)	11.95	18
186343	258878	27in 4K Gaming Monitor	1	389.99	09/17/19 15:49	812 Ridge St, Seattle, WA 98101	9	9	Seattle(WA)	Seattle (WA)	389.99	15
186366	258900	Bose SoundSport Headphones	1	99.99	09/11/19 12:24	746 Jefferson St, New York City, NY 10001	9	9	New York City(NY)	New York City (NY)	99.99	12
186370	258903	AAA Batteries (4-pack)	2	2.99	09/12/19 10:05	781 Park St, Los Angeles, CA 90001	9	9	Los Angeles(CA)	Los Angeles (CA)	5.98	10
186398	258930	Apple Airpods Headphones	1	150.00	09/13/19 14:31	494 6th St, Portland, OR 97035	9	9	Portland(OR)	Portland (OR)	150.00	14
186432	258963	Lightning Charging Cable	1	14.95	09/07/19 14:37	120 Madison St, Los Angeles, CA 90001	9	9	Los Angeles(CA)	Los Angeles (CA)	14.95	14
186448	258978	USB-C Charging Cable	1	11.95	09/18/19 01:00	663 Dogwood St, Portland, OR 97035	9	9	Portland(OR)	Portland (OR)	11.95	1
186477	259006	USB-C Charging Cable	1	11.95	09/19/19 10:30	494 11th St, New York City, NY 10001	9	9	New York City(NY)	New York City (NY)	11.95	10
186480	259008	Apple Airpods Headphones	1	150.00	09/07/19 13:31	441 Meadow St, Atlanta, GA 30301	9	9	Atlanta(GA)	Atlanta (GA)	150.00	13
186508	259035	27in FHD Monitor	1	149.99	09/29/19 13:52	327 Lake St, San Francisco, CA 94016	9	9	San Francisco(CA)	San Francisco (CA)	149.99	13
186542	259068	USB-C Charging Cable	1	11.95	09/10/19 18:39	175 Washington St, Boston, MA 02215	9	9	Boston(MA)	Boston (MA)	11.95	18
186598	259121	Wired	1	11.99	09/08/19	190 Meadow St,	9	9	Austin(TX)	Austin	11.99	22

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	Month	Month 2	City	City_2	Sales	Hour
		Headphones	1		22:48	Austin, TX 73301			Austin(TX)	(TX)		22
186600	259122	AA Batteries (4-pack)	1	3.84	09/08/19 21:15	764 Hickory St, Boston, MA 02215	9	9	Boston(MA)	Boston (MA)	3.84	21
186612	259133	Wired Headphones	1	11.99	09/01/19 12:54	507 Church St, Boston, MA 02215	9	9	Boston(MA)	Boston (MA)	11.99	12
186685	259204	USB-C Charging Cable	1	11.95	09/17/19 18:06	850 Meadow St, Boston, MA 02215	9	9	Boston(MA)	Boston (MA)	11.95	18
186690	259208	ThinkPad Laptop	1	999.99	09/20/19 14:31	375 Pine St, New York City, NY 10001	9	9	New York City(NY)	New York City (NY)	999.99	14
186754	259270	USB-C Charging Cable	1	11.95	09/06/19 15:27	940 10th St, San Francisco, CA 94016	9	9	San Francisco(CA)	San Francisco (CA)	11.95	15
186762	259277	Wired Headphones	2	11.99	09/28/19 13:07	795 Willow St, New York City, NY 10001	9	9	New York City(NY)	New York City (NY)	23.98	13
186782	259296	Apple Airpods Headphones	1	150.00	09/28/19 16:48	894 6th St, Dallas, TX 75001	9	9	Dallas(TX)	Dallas (TX)	150.00	16
186784	259297	Lightning Charging Cable	1	14.95	09/15/19 18:54	138 Main St, Boston, MA 02215	9	9	Boston(MA)	Boston (MA)	14.95	18
186785	259297	Lightning Charging Cable	1	14.95	09/15/19 18:54	138 Main St, Boston, MA 02215	9	9	Boston(MA)	Boston (MA)	14.95	18
186792	259303	AA Batteries (4-pack)	1	3.84	09/20/19 20:18	106 7th St, Atlanta, GA 30301	9	9	Atlanta(GA)	Atlanta (GA)	3.84	20
186804	259314	AAA Batteries (4-pack)	2	2.99	09/16/19 00:25	241 Highland St, Atlanta, GA 30301	9	9	Atlanta(GA)	Atlanta (GA)	5.98	0
186842	259350	USB-C Charging Cable	1	11.95	09/30/19 13:49	519 Maple St, San Francisco, CA 94016	9	9	San Francisco(CA)	San Francisco (CA)	11.95	13

7513 rows × 14 columns

In [96]:

```
all_data[all_data['Order ID'].duplicated(keep='last')]
```

Out[96]:

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	Month	Month 2	City	City_2	Sales	Ho
3	176560	Google Phone	1	600.00	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001	4	4	Los Angeles(CA)	Los Angeles (CA)	600.00	
18	176574	Google Phone	1	600.00	04/03/19 19:42	20 Hill St, Los Angeles, CA 90001	4	4	Los Angeles(CA)	Los Angeles (CA)	600.00	
						823						

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	Month	Month 2	City	City 2	Sales	Ho
30	176585	Bose SoundSport Headphones	1	99.99	04/07/19 11:31	Highland St, Boston, MA 02215	4	4	Boston(MA)	Boston (MA)	99.99	
32	176586	AAA Batteries (4-pack)	2	2.99	04/10/19 17:00	365 Center St, San Francisco, CA 94016	4	4	San Francisco(CA)	San Francisco (CA)	5.98	
119	176672	Lightning Charging Cable	1	14.95	04/12/19 11:07	778 Maple St, New York City, NY 10001	4	4	New York City(NY)	New York City (NY)	14.95	
129	176681	Apple Airpods Headphones	1	150.00	04/20/19 10:39	331 Cherry St, Seattle, WA 98101	4	4	Seattle(WA)	Seattle (WA)	150.00	
138	176689	Bose SoundSport Headphones	1	99.99	04/24/19 17:15	659 Lincoln St, New York City, NY 10001	4	4	New York City(NY)	New York City (NY)	99.99	
189	176739	34in Ultrawide Monitor	1	379.99	04/05/19 17:38	730 6th St, Austin, TX 73301	4	4	Austin(TX)	Austin (TX)	379.99	
225	176774	Lightning Charging Cable	1	14.95	04/25/19 15:06	372 Church St, Los Angeles, CA 90001	4	4	Los Angeles(CA)	Los Angeles (CA)	14.95	
233	176781	iPhone	1	700.00	04/03/19 07:37	976 Hickory St, Dallas, TX 75001	4	4	Dallas(TX)	Dallas (TX)	700.00	
250	176797	Google Phone	1	600.00	04/21/19 08:54	923 Elm St, Los Angeles, CA 90001	4	4	Los Angeles(CA)	Los Angeles (CA)	600.00	
251	176797	Bose SoundSport Headphones	1	99.99	04/21/19 08:54	923 Elm St, Los Angeles, CA 90001	4	4	Los Angeles(CA)	Los Angeles (CA)	99.99	
260	176805	Google Phone	1	600.00	04/01/19 15:50	91 Lincoln St, Portland, OR 97035	4	4	Portland(OR)	Portland (OR)	600.00	
264	176808	Google Phone	1	600.00	04/28/19 18:03	933 Meadow St, San Francisco, CA 94016	4	4	San Francisco(CA)	San Francisco (CA)	600.00	
270	176813	Google Phone	1	600.00	04/28/19 18:01	269 Hill St, Atlanta, GA 30301	4	4	Atlanta(GA)	Atlanta (GA)	600.00	
394	176935	AAA Batteries (4-pack)	1	2.99	04/03/19 21:31	315 1st St, Dallas, TX 75001	4	4	Dallas(TX)	Dallas (TX)	2.99	
435	176975	USB-C Charging Cable	1	11.95	04/23/19 15:46	28 13th St, San Francisco, CA 94016	4	4	San Francisco(CA)	San Francisco (CA)	11.95	
450	176989	Google Phone	1	600.00	04/09/19 13:43	346 9th St, Los Angeles, CA 90001	4	4	Los Angeles(CA)	Los Angeles (CA)	600.00	
455	176993	iPhone	1	700.00	04/07/19 09:43	28 South St, Los Angeles, CA 90001	4	4	Los Angeles(CA)	Los Angeles (CA)	700.00	
						216 Chestnut				Los		

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	Month	Month 2	City	City 2	Sales	Ho
485	177022	iPhone	1	700.00	04/17/19 18:14	St, Los Angeles, CA 90001	4	4	Los Angeles(CA)	Angeles (CA)	700.00	
567	177102	iPhone	1	700.00	04/24/19 05:19	542 Center St, New York City, NY 10001	4	4	New York City(NY)	New York City (NY)	700.00	
581	177115	iPhone	1	700.00	04/19/19 19:10	71 6th St, San Francisco, CA 94016	4	4	San Francisco(CA)	San Francisco (CA)	700.00	
584	177117	ThinkPad Laptop	1	999.99	04/16/19 19:34	564 Hill St, San Francisco, CA 94016	4	4	San Francisco(CA)	San Francisco (CA)	999.99	
635	177167	iPhone	1	700.00	04/28/19 16:40	39 Lakeview St, Boston, MA 02215	4	4	Boston(MA)	Boston (MA)	700.00	
636	177167	Apple Airpods Headphones	1	150.00	04/28/19 16:40	39 Lakeview St, Boston, MA 02215	4	4	Boston(MA)	Boston (MA)	150.00	
648	177178	iPhone	1	700.00	04/07/19 00:02	342 Dogwood St, Seattle, WA 98101	4	4	Seattle(WA)	Seattle (WA)	700.00	
652	177181	Wired Headphones	1	11.99	04/09/19 00:49	309 8th St, Los Angeles, CA 90001	4	4	Los Angeles(CA)	Los Angeles (CA)	11.99	
654	177182	Macbook Pro Laptop	1	1700.00	04/10/19 22:00	332 10th St, Los Angeles, CA 90001	4	4	Los Angeles(CA)	Los Angeles (CA)	1700.00	
657	177184	AA Batteries (4-pack)	1	3.84	04/01/19 16:49	289 Hill St, Atlanta, GA 30301	4	4	Atlanta(GA)	Atlanta (GA)	3.84	
672	177198	Vareebadd Phone	1	400.00	04/06/19 16:07	154 6th St, New York City, NY 10001	4	4	New York City(NY)	New York City (NY)	400.00	
...	...	...	...	...	...	...	...	...	...	...	...	
186172	258715	Lightning Charging Cable	1	14.95	09/15/19 16:50	550 10th St, Portland, OR 97035	9	9	Portland(OR)	Portland (OR)	14.95	
186189	258731	iPhone	1	700.00	09/20/19 13:00	146 Jackson St, Los Angeles, CA 90001	9	9	Los Angeles(CA)	Los Angeles (CA)	700.00	
186214	258755	iPhone	1	700.00	09/03/19 21:41	17 Highland St, New York City, NY 10001	9	9	New York City(NY)	New York City (NY)	700.00	
186219	258759	Vareebadd Phone	1	400.00	09/01/19 14:00	344 4th St, Boston, MA 02215	9	9	Boston(MA)	Boston (MA)	400.00	
186225	258764	AA Batteries (4-pack)	1	3.84	09/04/19 15:18	691 Hill St, New York City, NY 10001	9	9	New York City(NY)	New York City (NY)	3.84	
186269	258807	iPhone	1	700.00	09/07/19 18:12	809 Willow St, Seattle, WA 98101	9	9	Seattle(WA)	Seattle (WA)	700.00	

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	Month	Month 2	City	City_2	Sales	Ho
186311	258848	Vareebadd Phone	1	400.00	09/13/19 18:59	340 South St, Portland, OR 97035	9	9	Portland(OR)	Portland (OR)	400.00	
186342	258878	Lightning Charging Cable	1	14.95	09/17/19 15:49	812 Ridge St, Seattle, WA 98101	9	9	Seattle(WA)	Seattle (WA)	14.95	
186365	258900	Google Phone	1	600.00	09/11/19 12:24	746 Jefferson St, New York City, NY 10001	9	9	New York City(NY)	New York City (NY)	600.00	
186369	258903	Apple Airpods Headphones	1	150.00	09/12/19 10:05	781 Park St, Los Angeles, CA 90001	9	9	Los Angeles(CA)	Los Angeles (CA)	150.00	
186397	258930	iPhone	1	700.00	09/13/19 14:31	494 6th St, Portland, OR 97035	9	9	Portland(OR)	Portland (OR)	700.00	
186431	258963	iPhone	1	700.00	09/07/19 14:37	120 Madison St, Los Angeles, CA 90001	9	9	Los Angeles(CA)	Los Angeles (CA)	700.00	
186447	258978	AAA Batteries (4-pack)	1	2.99	09/18/19 01:00	663 Dogwood St, Portland, OR 97035	9	9	Portland(OR)	Portland (OR)	2.99	
186476	259006	Google Phone	1	600.00	09/19/19 10:30	494 11th St, New York City, NY 10001	9	9	New York City(NY)	New York City (NY)	600.00	
186479	259008	Wired Headphones	1	11.99	09/07/19 13:31	441 Meadow St, Atlanta, GA 30301	9	9	Atlanta(GA)	Atlanta (GA)	11.99	
186507	259035	27in FHD Monitor	1	149.99	09/29/19 13:52	327 Lake St, San Francisco, CA 94016	9	9	San Francisco(CA)	San Francisco (CA)	149.99	
186541	259068	Google Phone	1	600.00	09/10/19 18:39	175 Washington St, Boston, MA 02215	9	9	Boston(MA)	Boston (MA)	600.00	
186597	259121	iPhone	1	700.00	09/08/19 22:48	190 Meadow St, Austin, TX 73301	9	9	Austin(TX)	Austin (TX)	700.00	
186599	259122	AAA Batteries (4-pack)	2	2.99	09/08/19 21:15	764 Hickory St, Boston, MA 02215	9	9	Boston(MA)	Boston (MA)	5.98	
186611	259133	iPhone	1	700.00	09/01/19 12:54	507 Church St, Boston, MA 02215	9	9	Boston(MA)	Boston (MA)	700.00	
186684	259204	Vareebadd Phone	1	400.00	09/17/19 18:06	850 Meadow St, Boston, MA 02215	9	9	Boston(MA)	Boston (MA)	400.00	
186689	259208	27in FHD Monitor	1	149.99	09/20/19 14:31	375 Pine St, New York City, NY 10001	9	9	New York City(NY)	New York City (NY)	149.99	
		Google			09/06/19	940 10th St, San			San	San		

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	Month	Month 2	City	City_2	Sales	Ho
186753	259270	Phone	1	600.00	15/27	Francisco, CA 94016	9	9	Francisco(CA)	Francisco (CA)	600.00	
186761	259277	iPhone	1	700.00	09/28/19 13:07	795 Willow St, New York City, NY 10001	9	9	New York City(NY)	New York City (NY)	700.00	
186781	259296	Apple Airpods Headphones	1	150.00	09/28/19 16:48	894 6th St, Dallas, TX 75001	9	9	Dallas(TX)	Dallas (TX)	150.00	
186783	259297	iPhone	1	700.00	09/15/19 18:54	138 Main St, Boston, MA 02215	9	9	Boston(MA)	Boston (MA)	700.00	
186784	259297	Lightning Charging Cable	1	14.95	09/15/19 18:54	138 Main St, Boston, MA 02215	9	9	Boston(MA)	Boston (MA)	14.95	
186791	259303	34in Ultrawide Monitor	1	379.99	09/20/19 20:18	106 7th St, Atlanta, GA 30301	9	9	Atlanta(GA)	Atlanta (GA)	379.99	
186803	259314	Wired Headphones	1	11.99	09/16/19 00:25	241 Highland St, Atlanta, GA 30301	9	9	Atlanta(GA)	Atlanta (GA)	11.99	
186841	259350	Google Phone	1	600.00	09/30/19 13:49	519 Maple St, San Francisco, CA 94016	9	9	San Francisco(CA)	San Francisco (CA)	600.00	

7513 rows × 14 columns

In [97]:

```
all_data[all_data['Order ID'].duplicated(keep=False)]
```

Out[97]:

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	Month	Month 2	City	City_2	Sales	Hour
3	176560	Google Phone	1	600.00	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001	4	4	Los Angeles(CA)	Los Angeles (CA)	600.00	14
4	176560	Wired Headphones	1	11.99	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001	4	4	Los Angeles(CA)	Los Angeles (CA)	11.99	14
18	176574	Google Phone	1	600.00	04/03/19 19:42	20 Hill St, Los Angeles, CA 90001	4	4	Los Angeles(CA)	Los Angeles (CA)	600.00	19
19	176574	USB-C Charging Cable	1	11.95	04/03/19 19:42	20 Hill St, Los Angeles, CA 90001	4	4	Los Angeles(CA)	Los Angeles (CA)	11.95	19
30	176585	Bose SoundSport Headphones	1	99.99	04/07/19 11:31	823 Highland St, Boston, MA 02215	4	4	Boston(MA)	Boston (MA)	99.99	11
31	176585	Bose SoundSport Headphones	1	99.99	04/07/19 11:31	823 Highland St, Boston, MA 02215	4	4	Boston(MA)	Boston (MA)	99.99	11
32	176586	AAA Batteries (4-pack)	2	2.99	04/10/19 17:00	365 Center St, San Francisco,	4	4	San Francisco(CA)	San Francisco (CA)	5.98	17

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	Month	Month 2	City	City_2	Sales	Hour
		pack)				CA 94016				(CA)		
33	176586	Google Phone	1	600.00	04/10/19 17:00	365 Center St, San Francisco, CA 94016	4	4	San Francisco(CA)	San Francisco (CA)	600.00	17
119	176672	Lightning Charging Cable	1	14.95	04/12/19 11:07	778 Maple St, New York City, NY 10001	4	4	New York City(NY)	New York City (NY)	14.95	11
120	176672	USB-C Charging Cable	1	11.95	04/12/19 11:07	778 Maple St, New York City, NY 10001	4	4	New York City(NY)	New York City (NY)	11.95	11
129	176681	Apple Airpods Headphones	1	150.00	04/20/19 10:39	331 Cherry St, Seattle, WA 98101	4	4	Seattle(WA)	Seattle (WA)	150.00	10
130	176681	ThinkPad Laptop	1	999.99	04/20/19 10:39	331 Cherry St, Seattle, WA 98101	4	4	Seattle(WA)	Seattle (WA)	999.99	10
138	176689	Bose SoundSport Headphones	1	99.99	04/24/19 17:15	659 Lincoln St, New York City, NY 10001	4	4	New York City(NY)	New York City (NY)	99.99	17
139	176689	AAA Batteries (4-pack)	2	2.99	04/24/19 17:15	659 Lincoln St, New York City, NY 10001	4	4	New York City(NY)	New York City (NY)	5.98	17
189	176739	34in Ultrawide Monitor	1	379.99	04/05/19 17:38	730 6th St, Austin, TX 73301	4	4	Austin(TX)	Austin (TX)	379.99	17
190	176739	Google Phone	1	600.00	04/05/19 17:38	730 6th St, Austin, TX 73301	4	4	Austin(TX)	Austin (TX)	600.00	17
225	176774	Lightning Charging Cable	1	14.95	04/25/19 15:06	372 Church St, Los Angeles, CA 90001	4	4	Los Angeles(CA)	Los Angeles (CA)	14.95	15
226	176774	USB-C Charging Cable	1	11.95	04/25/19 15:06	372 Church St, Los Angeles, CA 90001	4	4	Los Angeles(CA)	Los Angeles (CA)	11.95	15
233	176781	iPhone	1	700.00	04/03/19 07:37	976 Hickory St, Dallas, TX 75001	4	4	Dallas(TX)	Dallas (TX)	700.00	7
234	176781	Lightning Charging Cable	1	14.95	04/03/19 07:37	976 Hickory St, Dallas, TX 75001	4	4	Dallas(TX)	Dallas (TX)	14.95	7
250	176797	Google Phone	1	600.00	04/21/19 08:54	923 Elm St, Los Angeles, CA 90001	4	4	Los Angeles(CA)	Los Angeles (CA)	600.00	8
251	176797	Bose SoundSport Headphones	1	99.99	04/21/19 08:54	923 Elm St, Los Angeles, CA 90001	4	4	Los Angeles(CA)	Los Angeles (CA)	99.99	8
252	176797	Wired Headphones	1	11.99	04/21/19 08:54	923 Elm St, Los Angeles, CA 90001	4	4	Los Angeles(CA)	Los Angeles (CA)	11.99	8
260	176805	Google Phone	1	600.00	04/01/19 15:50	91 Lincoln St, Portland, OR 97035	4	4	Portland(OR)	Portland (OR)	600.00	15
		USB-C				91 Lincoln				Portland		

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	Month	Month 2	City	City_2	Sales	Hour
261	176805	Charging Cable	1	11.95	04/01/19 15:50	St, Portland, OR 97035	4	4	Portland(OR)	Portland (OR)	11.95	15
264	176808	Google Phone	1	600.00	04/28/19 18:03	933 Meadow St, San Francisco, CA 94016	4	4	San Francisco(CA)	San Francisco (CA)	600.00	18
265	176808	Wired Headphones	1	11.99	04/28/19 18:03	933 Meadow St, San Francisco, CA 94016	4	4	San Francisco(CA)	San Francisco (CA)	11.99	18
270	176813	Google Phone	1	600.00	04/28/19 18:01	269 Hill St, Atlanta, GA 30301	4	4	Atlanta(GA)	Atlanta (GA)	600.00	18
271	176813	Wired Headphones	1	11.99	04/28/19 18:01	269 Hill St, Atlanta, GA 30301	4	4	Atlanta(GA)	Atlanta (GA)	11.99	18
394	176935	AAA Batteries (4-pack)	1	2.99	04/03/19 21:31	315 1st St, Dallas, TX 75001	4	4	Dallas(TX)	Dallas (TX)	2.99	21
...	...	...	...	...	...	...	...	...	...	...	...	...
186480	259008	Apple Airpods Headphones	1	150.00	09/07/19 13:31	441 Meadow St, Atlanta, GA 30301	9	9	Atlanta(GA)	Atlanta (GA)	150.00	13
186507	259035	27in FHD Monitor	1	149.99	09/29/19 13:52	327 Lake St, San Francisco, CA 94016	9	9	San Francisco(CA)	San Francisco (CA)	149.99	13
186508	259035	27in FHD Monitor	1	149.99	09/29/19 13:52	327 Lake St, San Francisco, CA 94016	9	9	San Francisco(CA)	San Francisco (CA)	149.99	13
186541	259068	Google Phone	1	600.00	09/10/19 18:39	175 Washington St, Boston, MA 02215	9	9	Boston(MA)	Boston (MA)	600.00	18
186542	259068	USB-C Charging Cable	1	11.95	09/10/19 18:39	175 Washington St, Boston, MA 02215	9	9	Boston(MA)	Boston (MA)	11.95	18
186597	259121	iPhone	1	700.00	09/08/19 22:48	190 Meadow St, Austin, TX 73301	9	9	Austin(TX)	Austin (TX)	700.00	22
186598	259121	Wired Headphones	1	11.99	09/08/19 22:48	190 Meadow St, Austin, TX 73301	9	9	Austin(TX)	Austin (TX)	11.99	22
186599	259122	AAA Batteries (4-pack)	2	2.99	09/08/19 21:15	764 Hickory St, Boston, MA 02215	9	9	Boston(MA)	Boston (MA)	5.98	21
186600	259122	AA Batteries (4-pack)	1	3.84	09/08/19 21:15	764 Hickory St, Boston, MA 02215	9	9	Boston(MA)	Boston (MA)	3.84	21
186611	259133	iPhone	1	700.00	09/01/19 12:54	507 Church St, Boston, MA 02215	9	9	Boston(MA)	Boston (MA)	700.00	12
186612	259133	Wired Headphones	1	11.99	09/01/19 12:54	507 Church St, Boston, MA 02215	9	9	Boston(MA)	Boston (MA)	11.99	12
		Vareebadd			09/17/19	850 Meadow St,				Boston		

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	Month	Month 2	City	City_2	Sales	Hour
186684	259204	[illegible] Phone	1	400.00	[illegible] 18:06	[illegible] Boston, MA 02215	9	9	Boston(MA)	[illegible] (MA)	400.00	18
186685	259204	USB-C Charging Cable	1	11.95	09/17/19 18:06	850 Meadow St, Boston, MA 02215	9	9	Boston(MA)	Boston (MA)	11.95	18
186689	259208	27in FHD Monitor	1	149.99	09/20/19 14:31	375 Pine St, New York City, NY 10001	9	9	New York City(NY)	New York City (NY)	149.99	14
186690	259208	ThinkPad Laptop	1	999.99	09/20/19 14:31	375 Pine St, New York City, NY 10001	9	9	New York City(NY)	New York City (NY)	999.99	14
186753	259270	Google Phone	1	600.00	09/06/19 15:27	940 10th St, San Francisco, CA 94016	9	9	San Francisco(CA)	San Francisco (CA)	600.00	15
186754	259270	USB-C Charging Cable	1	11.95	09/06/19 15:27	940 10th St, San Francisco, CA 94016	9	9	San Francisco(CA)	San Francisco (CA)	11.95	15
186761	259277	iPhone	1	700.00	09/28/19 13:07	795 Willow St, New York City, NY 10001	9	9	New York City(NY)	New York City (NY)	700.00	13
186762	259277	Wired Headphones	2	11.99	09/28/19 13:07	795 Willow St, New York City, NY 10001	9	9	New York City(NY)	New York City (NY)	23.98	13
186781	259296	Apple Airpods Headphones	1	150.00	09/28/19 16:48	894 6th St, Dallas, TX 75001	9	9	Dallas(TX)	Dallas (TX)	150.00	16
186782	259296	Apple Airpods Headphones	1	150.00	09/28/19 16:48	894 6th St, Dallas, TX 75001	9	9	Dallas(TX)	Dallas (TX)	150.00	16
186783	259297	iPhone	1	700.00	09/15/19 18:54	138 Main St, Boston, MA 02215	9	9	Boston(MA)	Boston (MA)	700.00	18
186784	259297	Lightning Charging Cable	1	14.95	09/15/19 18:54	138 Main St, Boston, MA 02215	9	9	Boston(MA)	Boston (MA)	14.95	18
186785	259297	Lightning Charging Cable	1	14.95	09/15/19 18:54	138 Main St, Boston, MA 02215	9	9	Boston(MA)	Boston (MA)	14.95	18
186791	259303	34in Ultrawide Monitor	1	379.99	09/20/19 20:18	106 7th St, Atlanta, GA 30301	9	9	Atlanta(GA)	Atlanta (GA)	379.99	20
186792	259303	AA Batteries (4-pack)	1	3.84	09/20/19 20:18	106 7th St, Atlanta, GA 30301	9	9	Atlanta(GA)	Atlanta (GA)	3.84	20
186803	259314	Wired Headphones	1	11.99	09/16/19 00:25	241 Highland St, Atlanta, GA 30301	9	9	Atlanta(GA)	Atlanta (GA)	11.99	0
186804	259314	AAA Batteries (4-pack)	2	2.99	09/16/19 00:25	241 Highland St, Atlanta, GA 30301	9	9	Atlanta(GA)	Atlanta (GA)	5.98	0
186841	259350	Google Phone	1	600.00	09/30/19 13:49	519 Maple St, San Francisco, CA 94016	9	9	San Francisco(CA)	San Francisco (CA)	600.00	13
		USB-C				519 Maple				San		

186842	259360	USB-C Charging Cable	1	11.95	09/30/19 13:49	St. San Francisco, CA 94016	9	30	San Francisco (CA)	San Francisco (CA)	11.95	13

14649 rows × 14 columns

In [99]:

```
df = all_data[all_data['Order ID'].duplicated(keep=False)]
df.groupby('Order ID')['Product']
```

Out[99]:

```
<pandas.core.groupby.generic.SeriesGroupBy object at 0x000000A9B6C22390>
```

In [100]:

```
df.groupby('Order ID')['Product'].transform(lambda x: ','.join(x))
```

Out[100]:

```
3                              Google Phone,Wired Headphones
4                              Google Phone,Wired Headphones
18                        Google Phone,USB-C Charging Cable
19                        Google Phone,USB-C Charging Cable
30          Bose SoundSport Headphones,Bose SoundSport Hea...
31          Bose SoundSport Headphones,Bose SoundSport Hea...
32                    AAA Batteries (4-pack),Google Phone
33                    AAA Batteries (4-pack),Google Phone
119           Lightning Charging Cable,USB-C Charging Cable
120           Lightning Charging Cable,USB-C Charging Cable
129                Apple Airpods Headphones,ThinkPad Laptop
130                Apple Airpods Headphones,ThinkPad Laptop
138         Bose SoundSport Headphones,AAA Batteries (4-pack)
139         Bose SoundSport Headphones,AAA Batteries (4-pack)
189                   34in Ultrawide Monitor,Google Phone
190                   34in Ultrawide Monitor,Google Phone
225           Lightning Charging Cable,USB-C Charging Cable
226           Lightning Charging Cable,USB-C Charging Cable
233                       iPhone,Lightning Charging Cable
234                       iPhone,Lightning Charging Cable
250         Google Phone,Bose SoundSport Headphones,Wired ...
251         Google Phone,Bose SoundSport Headphones,Wired ...
252         Google Phone,Bose SoundSport Headphones,Wired ...
260                     Google Phone,USB-C Charging Cable
261                     Google Phone,USB-C Charging Cable
264                          Google Phone,Wired Headphones
265                          Google Phone,Wired Headphones
270                          Google Phone,Wired Headphones
271                          Google Phone,Wired Headphones
394                  AAA Batteries (4-pack),27in FHD Monitor
                                ...
186480             Wired Headphones,Apple Airpods Headphones
186507                    27in FHD Monitor,27in FHD Monitor
186508                    27in FHD Monitor,27in FHD Monitor
186541                    Google Phone,USB-C Charging Cable
186542                    Google Phone,USB-C Charging Cable
186597                               iPhone,Wired Headphones
186598                               iPhone,Wired Headphones
186599          AAA Batteries (4-pack),AA Batteries (4-pack)
186600          AAA Batteries (4-pack),AA Batteries (4-pack)
186611                               iPhone,Wired Headphones
186612                               iPhone,Wired Headphones
186684                  Vareebadd Phone,USB-C Charging Cable
186685                  Vareebadd Phone,USB-C Charging Cable
186689                      27in FHD Monitor,ThinkPad Laptop
186690                      27in FHD Monitor,ThinkPad Laptop
186753                    Google Phone,USB-C Charging Cable
186754                    Google Phone,USB-C Charging Cable
186761                               iPhone,Wired Headphones
186762                               iPhone,Wired Headphones
186781    Apple Airpods Headphones,Apple Airpods Headphones
```

```
186782    Apple Airpods Headphones,Apple Airpods Headphones
186783    iPhone,Lightning Charging Cable,Lightning Char...
186784    iPhone,Lightning Charging Cable,Lightning Char...
186785    iPhone,Lightning Charging Cable,Lightning Char...
186791         34in Ultrawide Monitor,AA Batteries (4-pack)
186792         34in Ultrawide Monitor,AA Batteries (4-pack)
186803              Wired Headphones,AAA Batteries (4-pack)
186804              Wired Headphones,AAA Batteries (4-pack)
186841                    Google Phone,USB-C Charging Cable
186842                    Google Phone,USB-C Charging Cable
Name: Product, Length: 14649, dtype: object
```

In [102]:

```
# https://stackoverflow.com/questions/43348194/pandas-select-rows-if-id-appear-several-time
df = all_data[all_data['Order ID'].duplicated(keep=False)]

# Referenced: https://stackoverflow.com/questions/27298178/concatenate-strings-from-several-rows-using-pandas-groupby
df['Grouped'] = df.groupby('Order ID')['Product'].transform(lambda x: ','.join(x))
print(df.head())
df2 = df[['Order ID', 'Grouped']].drop_duplicates()
print(df2.head())
```

```
   Order ID                     Product  Quantity Ordered  Price Each  \
3    176560                Google Phone                 1      600.00
4    176560            Wired Headphones                 1       11.99
18   176574                Google Phone                 1      600.00
19   176574        USB-C Charging Cable                 1       11.95
30   176585  Bose SoundSport Headphones                 1       99.99

        Order Date                      Purchase Address  Month  Month 2  \
3   04/12/19 14:38  669 Spruce St, Los Angeles, CA 90001      4        4
4   04/12/19 14:38  669 Spruce St, Los Angeles, CA 90001      4        4
18  04/03/19 19:42     20 Hill St, Los Angeles, CA 90001      4        4
19  04/03/19 19:42     20 Hill St, Los Angeles, CA 90001      4        4
30  04/07/19 11:31     823 Highland St, Boston, MA 02215      4        4

                City             City_2  Sales  Hour  Minute  Count  \
3    Los Angeles(CA)  Los Angeles  (CA) 600.00    14      38      1
4    Los Angeles(CA)  Los Angeles  (CA)  11.99    14      38      1
18   Los Angeles(CA)  Los Angeles  (CA) 600.00    19      42      1
19   Los Angeles(CA)  Los Angeles  (CA)  11.95    19      42      1
30        Boston(MA)       Boston  (MA)  99.99    11      31      1

                                       Grouped
3                    Google Phone,Wired Headphones
4                    Google Phone,Wired Headphones
18                Google Phone,USB-C Charging Cable
19                Google Phone,USB-C Charging Cable
30  Bose SoundSport Headphones,Bose SoundSport Hea...
    Order ID                                            Grouped
3     176560                     Google Phone,Wired Headphones
18    176574                  Google Phone,USB-C Charging Cable
30    176585  Bose SoundSport Headphones,Bose SoundSport Hea...
32    176586                AAA Batteries (4-pack),Google Phone
119   176672      Lightning Charging Cable,USB-C Charging Cable
```

```
c:\users\ad\anaconda3\lib\site-packages\ipykernel_launcher.py:5: 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: http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy
  """
```

In [103]:

```
# Referenced: https://stackoverflow.com/questions/52195887/counting-unique-pairs-of-numbers-into-a-python-dictionary
```

```
from itertools import combinations
from collections import Counter

count = Counter()

for row in df2['Grouped']:
    row_list = row.split(',')
    count.update(Counter(combinations(row_list, 2))) # for list of 2 products

for key,value in count.most_common(10):
    print(key, value)
```

```
('iPhone', 'Lightning Charging Cable') 1005
('Google Phone', 'USB-C Charging Cable') 987
('iPhone', 'Wired Headphones') 447
('Google Phone', 'Wired Headphones') 414
('Vareebadd Phone', 'USB-C Charging Cable') 361
('iPhone', 'Apple Airpods Headphones') 360
('Google Phone', 'Bose SoundSport Headphones') 220
('USB-C Charging Cable', 'Wired Headphones') 160
('Vareebadd Phone', 'Wired Headphones') 143
('Lightning Charging Cable', 'Wired Headphones') 92
```

What product sold the most? Why do you think it sold the most?

In [104]:

```
product_group = all_data.groupby('Product')
quantity_ordered = product_group.sum()['Quantity Ordered']
quantity_ordered
```

Out[104]:

```
Product
20in Monitor                   4129
27in 4K Gaming Monitor         6244
27in FHD Monitor               7550
34in Ultrawide Monitor         6199
AA Batteries (4-pack)         27635
AAA Batteries (4-pack)        31017
Apple Airpods Headphones      15661
Bose SoundSport Headphones    13457
Flatscreen TV                  4819
Google Phone                   5532
LG Dryer                        646
LG Washing Machine              666
Lightning Charging Cable      23217
Macbook Pro Laptop             4728
ThinkPad Laptop                4130
USB-C Charging Cable          23975
Vareebadd Phone                2068
Wired Headphones              20557
iPhone                         6849
Name: Quantity Ordered, dtype: int64
```

In [110]:

```
keys=sorted(df['Product'].unique())
keys
```

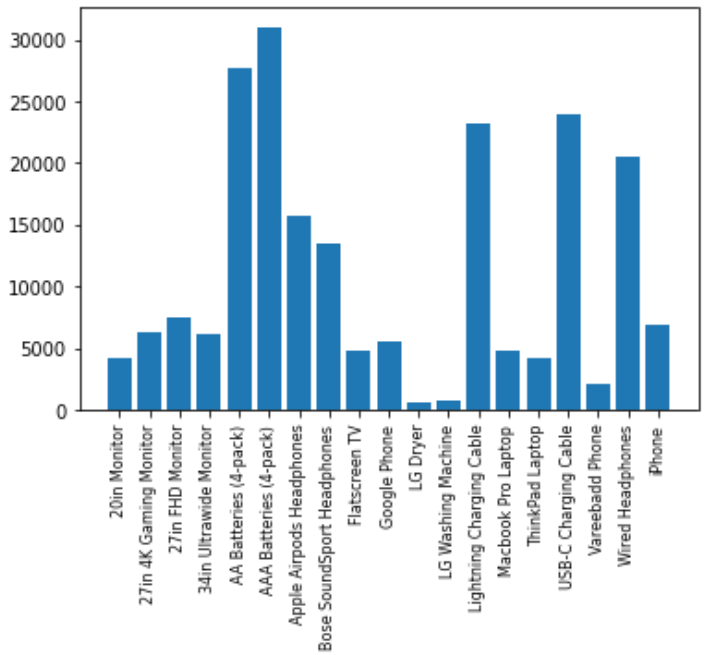
Out[110]:

```
['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']
```

In [111]:

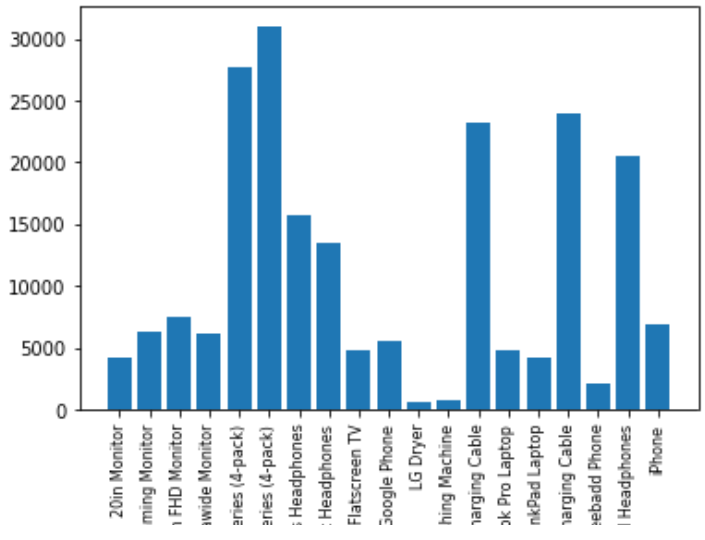
```
product_group = all_data.groupby('Product')
quantity_ordered = product_group.sum()['Quantity Ordered']

plt.bar(keys, quantity_ordered)
plt.xticks(keys, rotation='vertical', size=8)
plt.show()
```



alternate way

In [112]:

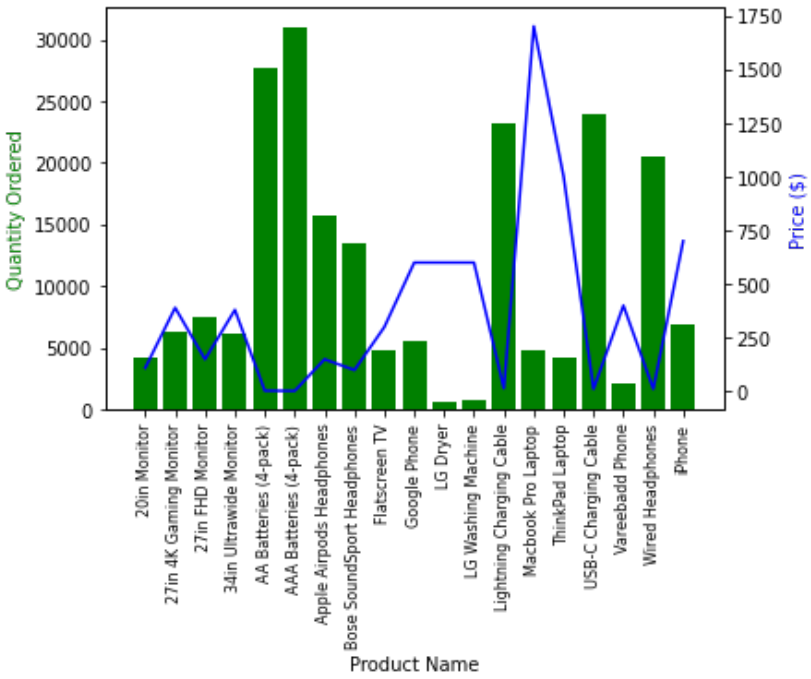
```
product_group = all_data.groupby('Product')
quantity_ordered = product_group.sum()['Quantity Ordered']

keys = [pair for pair, df in product_group]
plt.bar(keys, quantity_ordered)
plt.xticks(keys, rotation='vertical', size=8)
plt.show()
```



In [114]:

```
# Referenced: https://stackoverflow.com/questions/14762181/adding-a-y-axis-label-to-secondary-y-axis-in-matplotlib

%matplotlib inline
prices = all_data.groupby('Product').mean()['Price Each']

fig, ax1 = plt.subplots()

ax2 = ax1.twinx()
ax1.bar(keys, quantity_ordered, color='g')
ax2.plot(keys, prices, color='b')

ax1.set_xlabel('Product Name')
ax1.set_ylabel('Quantity Ordered', color='g')
ax2.set_ylabel('Price ($)', color='b')
ax1.set_xticklabels(keys, rotation='vertical', size=8)

fig.show()
```

```
c:\users\ad\anaconda3\lib\site-packages\ipykernel_launcher.py:15: UserWarning: FixedFormatter should only be used together with FixedLocator
  from ipykernel import kernelapp as app
c:\users\ad\anaconda3\lib\site-packages\ipykernel_launcher.py:17: UserWarning: Matplotlib is currently using module://ipykernel.pylab.backend_inline, which is a non-GUI backend, so cannot show the figure.
```



We can see that ,where price is less ,order is more

In [ ]: