In [1]:
```
import pandas as pd
import matplotlib.pyplot as plt
```

In [2]:
```
data1=pd.read_csv(r"C:\Users\Vatsal_Fast\Desktop\DataSciencePractice\keithgali
data1
```

Out[2]:

	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
...	...	...	...	...	...	...
25112	319666	Lightning Charging Cable	1	14.95	12/11/19 20:58	14 Madison St, San Francisco, CA 94016
25113	319667	AA Batteries (4-pack)	2	3.84	12/01/19 12:01	549 Willow St, Los Angeles, CA 90001
25114	319668	Vareebadd Phone	1	400	12/09/19 06:43	273 Wilson St, Seattle, WA 98101
25115	319669	Wired Headphones	1	11.99	12/03/19 10:39	778 River St, Dallas, TX 75001
25116	319670	Bose SoundSport Headphones	1	99.99	12/21/19 21:45	747 Chestnut St, Los Angeles, CA 90001

25117 rows × 6 columns

In [3]:
```
import os

#for loop that could create a list by iterating over the directory file list
files=[file for file in os.listdir(r"C:\Users\Vatsal_Fast\Desktop\DataScienceP
files
```

Out[3]:
```
['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 [4]:
```
ces\Data_science\keithgaillypython\Pandas\analysis\data\"+file)
tsal_Fast\\Desktop\\DataSciencePractice\\keithgaliElectronicRetailAnalysis\\da
_data]
data])



a_science\keithgaillypython\Pandas\analysis\data\Sales_2019.csv")
esktop\DataSciencePractice\keithgaliElectronicRetailAnalysis\Agregateddata\Sal
```

```
      Order ID                    Product Quantity Ordered Price Each  \
0       176558        USB-C Charging Cable                2      11.95
1          NaN                         NaN              NaN        NaN
2       176559  Bose SoundSport Headphones                1      99.99
3       176560                Google Phone                1        600
4       176560            Wired Headphones                1      11.99
...        ...                         ...              ...        ...
11681   259353      AAA Batteries (4-pack)                3       2.99
11682   259354                      iPhone                1        700
11683   259355                      iPhone                1        700
11684   259356      34in Ultrawide Monitor                1     379.99
11685   259357        USB-C Charging Cable                1      11.95

          Order Date                      Purchase Address
0      04/19/19 08:46             917 1st St, Dallas, TX 75001
```

In [4]:
```
ces\Data_science\keithgaillypython\Pandas\analysis\data\"+file)
tsal_Fast\\Desktop\\DataSciencePractice\\keithgaliElectronicRetailAnalysis\\da
_data]
data])



a_science\keithgaillypython\Pandas\analysis\data\Sales_2019.csv")
esktop\DataSciencePractice\keithgaliElectronicRetailAnalysis\Agregateddata\Sal
```

```
      Order ID                     Product Quantity Ordered Price Each  \
0       176558        USB-C Charging Cable                2      11.95
1          NaN                         NaN              NaN        NaN
2       176559  Bose SoundSport Headphones                1      99.99
3       176560                Google Phone                1        600
4       176560             Wired Headphones               1      11.99
...        ...                         ...              ...        ...
11681   259353      AAA Batteries (4-pack)                3       2.99
11682   259354                      iPhone                1        700
11683   259355                      iPhone                1        700
11684   259356      34in Ultrawide Monitor                1     379.99
11685   259357        USB-C Charging Cable                1      11.95

           Order Date                      Purchase Address
0      04/19/19 08:46          917 1st St, Dallas, TX 75001
1                 NaN                                   NaN
2      04/07/19 22:30      682 Chestnut St, Boston, MA 02215
3      04/12/19 14:38   669 Spruce St, Los Angeles, CA 90001
4      04/12/19 14:38   669 Spruce St, Los Angeles, CA 90001
...              ...                                   ...
11681  09/17/19 20:56   840 Highland St, Los Angeles, CA 90001
11682  09/01/19 16:00  216 Dogwood St, San Francisco, CA 94016
11683  09/23/19 07:39    220 12th St, San Francisco, CA 94016
11684  09/19/19 17:30   511 Forest St, San Francisco, CA 94016
11685  09/30/19 00:18   250 Meadow St, San Francisco, CA 94016

[186850 rows x 6 columns]
```

In [6]:
```
atsal_Fast\Desktop\DataSciencePractice\keithgaliElectronicRetailAnalysis\Agreg
```

```
  Order ID                     Product Quantity Ordered Price Each  \
0   176558        USB-C Charging Cable                2      11.95
1      NaN                         NaN              NaN        NaN
2   176559  Bose SoundSport Headphones                1      99.99
3   176560                Google Phone                1        600
4   176560             Wired Headphones               1      11.99

       Order Date                      Purchase Address
0  04/19/19 08:46          917 1st St, Dallas, TX 75001
1             NaN                                   NaN
2  04/07/19 22:30      682 Chestnut St, Boston, MA 02215
3  04/12/19 14:38   669 Spruce St, Los Angeles, CA 90001
4  04/12/19 14:38   669 Spruce St, Los Angeles, CA 90001
```

In [9]:
```
# print(data[data['Month']=='NaN'])????????????????????????????
```

Finding the best month for sales

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

In [7]:
```
data['Month']=data['Order Date'].str[0:2]
data.head()
```

Out[9]:
Out[7]:

	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
1	NaN	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	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

In [10]:
```
# data['Month']=data['Month'].astype('int16')
#for the header data cleaning
```

In [11]:
```
print(data[data['Month']=='Or'])
data=data[data['Month']!='Or']
print(data[data['Month']=='Or'])
```

In [9]:
```
# print(data[data['Month']=='NaN'])?????????????????????????
```

Finding the best month for sales

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

In [7]:
```
data['Month']=data['Order Date'].str[0:2]
data.head()
```

Out[9]:

Out[7]:

	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 [10]:
```
# data['Month']=data['Month'].astype('int16')
#further cleaning required
```

In [11]:
```
print(data[data['Month']=='Or'])
data=data[data['Month']!='Or']
print(data[data['Month']=='Or'])
```

```
        Order ID  Product  Quantity Ordered  Price Each  Order Date  \
519     Order ID  Product  Quantity Ordered  Price Each  Order Date
1149    Order ID  Product  Quantity Ordered  Price Each  Order Date
1155    Order ID  Product  Quantity Ordered  Price Each  Order Date
2878    Order ID  Product  Quantity Ordered  Price Each  Order Date
2893    Order ID  Product  Quantity Ordered  Price Each  Order Date
...          ...      ...               ...         ...         ...
185164  Order ID  Product  Quantity Ordered  Price Each  Order Date
185551  Order ID  Product  Quantity Ordered  Price Each  Order Date
186563  Order ID  Product  Quantity Ordered  Price Each  Order Date
186632  Order ID  Product  Quantity Ordered  Price Each  Order Date
186738  Order ID  Product  Quantity Ordered  Price Each  Order Date

        Purchase Address Month
519     Purchase Address    Or
1149    Purchase Address    Or
1155    Purchase Address    Or
2878    Purchase Address    Or
2893    Purchase Address    Or
...                  ...   ...
185164  Purchase Address    Or
185551  Purchase Address    Or
186563  Purchase Address    Or
186632  Purchase Address    Or
186738  Purchase Address    Or

[355 rows x 7 columns]
Empty DataFrame
Columns: [Order ID, Product, Quantity Ordered, Price Each, Order Date, Purch
ase Address, Month]
Index: []
```

In [13]:
```
data['Net_price']=data['Quantity Ordered']*data['Price Each']
```

In [12]:
```
data['Month']=data['Month'].astype('int16')
#further data cleaning required
print(data.head())
#success finally
```

```
  Order ID                    Product Quantity Ordered Price Each  \
0   176558        USB-C Charging Cable                2      11.95
2   176559  Bose SoundSport Headphones                1      99.99
3   176560                Google Phone                1        600
4   176560            Wired Headphones                1      11.99
5   176561            Wired Headphones                1      11.99

       Order Date                      Purchase Address  Month
0  04/19/19 08:46          917 1st St, Dallas, TX 75001      4
2  04/07/19 22:30     682 Chestnut St, Boston, MA 02215      4
3  04/12/19 14:38  669 Spruce St, Los Angeles, CA 90001      4
4  04/12/19 14:38  669 Spruce St, Los Angeles, CA 90001      4
5  04/30/19 09:27     333 8th St, Los Angeles, CA 90001      4
```

In [13]:
```
data['Net_price']=data['Quantity Ordered']*data['Price Each']
data['Month']=data['Month'].astype('int16')
#further data cleaning required
print(data.head())
#success finally
```

```
  Order ID                    Product Quantity Ordered Price Each  \
0   176558        USB-C Charging Cable                2      11.95
2   176559  Bose SoundSport Headphones                1      99.99
3   176560                Google Phone                1        600
4   176560            Wired Headphones                1      11.99
5   176561            Wired Headphones                1      11.99

       Order Date                      Purchase Address  Month
0  04/19/19 08:46          917 1st St, Dallas, TX 75001      4
2  04/07/19 22:30     682 Chestnut St, Boston, MA 02215      4
3  04/12/19 14:38  669 Spruce St, Los Angeles, CA 90001      4
4  04/12/19 14:38  669 Spruce St, Los Angeles, CA 90001      4
5  04/30/19 09:27     333 8th St, Los Angeles, CA 90001      4
```

```
---------------------------------------------------------------------------
TypeError                                 Traceback (most recent call last)
File ~\AppData\Local\Programs\Python\Python311\Lib\site-packages\pandas\core
\ops\array_ops.py:171, in _na_arithmetic_op(left, right, op, is_cmp)
    170 try:
--> 171     result = func(left, right)
    172 except TypeError:

File ~\AppData\Local\Programs\Python\Python311\Lib\site-packages\pandas\core
\computation\expressions.py:239, in evaluate(op, a, b, use_numexpr)
    237     if use_numexpr:
    238         # error: "None" not callable
--> 239         return _evaluate(op, op_str, a, b)  # type: ignore[misc]
    240 return _evaluate_standard(op, op_str, a, b)

File ~\AppData\Local\Programs\Python\Python311\Lib\site-packages\pandas\core
\computation\expressions.py:70, in _evaluate_standard(op, op_str, a, b)
     69     _store_test_result(False)
---> 70 return op(a, b)

TypeError: can't multiply sequence by non-int of type 'str'

During handling of the above exception, another exception occurred:

TypeError                                 Traceback (most recent call last)
Cell In[13], line 1
----> 1 data['Net_price']=data['Quantity Ordered']*data['Price Each']
      2 #further data cleaning required

File ~\AppData\Local\Programs\Python\Python311\Lib\site-packages\pandas\core
```

```
---------------------------------------------------------------------------
TypeError                                 Traceback (most recent call last)
File ~\AppData\Local\Programs\Python\Python311\Lib\site-packages\pandas\core
\ops\array_ops.py:171, in _na_arithmetic_op(left, right, op, is_cmp)
    170 try:
--> 171     result = func(left, right)
    172 except TypeError:

File ~\AppData\Local\Programs\Python\Python311\Lib\site-packages\pandas\core
\computation\expressions.py:239, in evaluate(op, a, b, use_numexpr)
    237     if use_numexpr:
    238         # error: "None" not callable
--> 239         return _evaluate(op, op_str, a, b)  # type: ignore[misc]
    240 return _evaluate_standard(op, op_str, a, b)

File ~\AppData\Local\Programs\Python\Python311\Lib\site-packages\pandas\core
\computation\expressions.py:70, in _evaluate_standard(op, op_str, a, b)
     69     _store_test_result(False)
---> 70 return op(a, b)

TypeError: can't multiply sequence by non-int of type 'str'

During handling of the above exception, another exception occurred:

TypeError                                 Traceback (most recent call last)
Cell In[13], line 1
----> 1 data['Net_price']=data['Quantity Ordered']*data['Price Each']
      2 #further data cleaning required

File ~\AppData\Local\Programs\Python\Python311\Lib\site-packages\pandas\core
\ops\common.py:81, in _unpack_zerodim_and_defer.<locals>.new_method(self, ot
her)
     77             return NotImplemented
     79 other = item_from_zerodim(other)
---> 81 return method(self, other)

File ~\AppData\Local\Programs\Python\Python311\Lib\site-packages\pandas\core
\arraylike.py:202, in OpsMixin.__mul__(self, other)
    200 @unpack_zerodim_and_defer("__mul__")
    201 def __mul__(self, other):
--> 202     return self._arith_method(other, operator.mul)

File ~\AppData\Local\Programs\Python\Python311\Lib\site-packages\pandas\core
\series.py:6112, in Series._arith_method(self, other, op)
   6110 def _arith_method(self, other, op):
   6111     self, other = ops.align_method_SERIES(self, other)
-> 6112     return base.IndexOpsMixin._arith_method(self, other, op)

File ~\AppData\Local\Programs\Python\Python311\Lib\site-packages\pandas\core
\base.py:1348, in IndexOpsMixin._arith_method(self, other, op)
   1345 rvalues = ensure_wrapped_if_datetimelike(rvalues)
   1347 with np.errstate(all="ignore"):
-> 1348     result = ops.arithmetic_op(lvalues, rvalues, op)
   1350 return self._construct_result(result, name=res_name)

File ~\AppData\Local\Programs\Python\Python311\Lib\site-packages\pandas\core
\ops\array_ops.py:232, in arithmetic_op(left, right, op)
    228     _bool_arith_check(op, left, right)
    230     # error: Argument 1 to "_na_arithmetic_op" has incompatible type
    231     # "Union[ExtensionArray, ndarray[Any, Any]]"; expected "ndarray
[Any, Any]"
--> 232     res_values = _na_arithmetic_op(left, right, op)  # type: ignore
    234 return res_values

File ~\AppData\Local\Programs\Python\Python311\Lib\site-packages\pandas\core
\ops\array_ops.py:178, in _na_arithmetic_op(left, right, op, is_cmp)
    174         # For object dtype, fallback to a masked operation (only ope
rating
    175         #  on the non-missing values)
    176         # Don't do this for comparisons, as that will handle complex
numbers
    177         #  incorrectly, see GH#32047
--> 178         result = _masked_arith_op(left, right, op)
    179     else:
    180         raise

File ~\AppData\Local\Programs\Python\Python311\Lib\site-packages\pandas\core
\ops\array_ops.py:116, in _masked_arith_op(x, y, op)
    114     # See GH#5284, GH#5035, GH#19448 for historical reference
    115     if mask.any():
--> 116         result[mask] = op(xrav[mask], yrav[mask])
    118 else:
    119     if not is_scalar(y):

TypeError: can't multiply sequence by non-int of type 'str'
```

In [14]:
```
# data['Quantity Ordered']=data['Quantity Ordered'].astype('int32')
# data['Price Each']=data['Price Each'].astype('float')

#or
data['Quantity Ordered']=pd.to_numeric(data['Quantity Ordered'])
data['Price Each']=pd.to_numeric(data['Price Each'])
```

In [15]:
```
data['Net_price']=data['Quantity Ordered']*data['Price Each']
print(data.head())
print(data.groupby('Month').sum().sort_values(by=['Net_price'],ascending=False
```

```
  Order ID                     Product  Quantity Ordered  Price Each  \
0   176558        USB-C Charging Cable                 2       11.95
2   176559  Bose SoundSport Headphones                 1       99.99
3   176560                Google Phone                 1      600.00
4   176560            Wired Headphones                 1       11.99
5   176561            Wired Headphones                 1       11.99

       Order Date                Purchase Address  Month  Net_price
0  04/19/19 08:46          917 1st St, Dallas, TX 75001      4      23.90
2  04/07/19 22:30      682 Chestnut St, Boston, MA 02215      4      99.99
3  04/12/19 14:38  669 Spruce St, Los Angeles, CA 90001      4     600.00
4  04/12/19 14:38  669 Spruce St, Los Angeles, CA 90001      4      11.99
5  04/30/19 09:27      333 8th St, Los Angeles, CA 90001      4      11.99
                                                  Order ID  \
Month
```

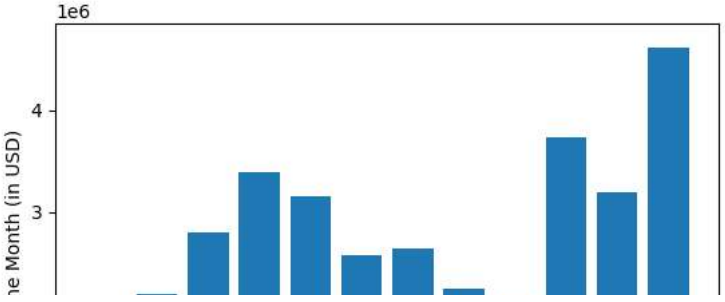
In [14]:
```
#data['Quantity Ordered']=data['Quantity Ordered'].astype('int32')
# data['Price Each']=data['Price Each'].astype('float')

#File ~\AppData\Local\Programs\Python\Python311\Lib\site-packages\pandas\core
\ops\array_ops.py:178, in _na_arithmetic_op(left, right, op, is_cmp)
data['Quantity Ordered']=pd.to_numeric(data['Quantity Ordered'])
data['Price Each']=pd.to_numeric(data['Price Each'])
    174         # For object dtype, fallback to a masked operation (only ope
rating
    175         #  on the non-missing values)
    176         # Don't do this for comparisons, as that will handle complex
numbers
    177         #  incorrectly, see GH#32047
--> 178         result = _masked_arith_op(left, right, op)
    179     else:
    180         raise
File ~\AppData\Local\Programs\Python\Python311\Lib\site-packages\pandas\core
\ops\array_ops.py:116, in _masked_arith_op(x, y, op)
    114     # See GH#5284, GH#5035, GH#19448 for historical reference
    115     if mask.any():
--> 116         result[mask] = op(xrav[mask], yrav[mask])
    118 else:
    119     if not is_scalar(y):
TypeError: can't multiply sequence by non-int of type 'str'
```

In [15]:
```
data['Net_price']=data['Quantity Ordered']*data['Price Each']
print(data.head())
print(data.groupby('Month').sum().sort_values(by=['Net_price'],ascending=False
```

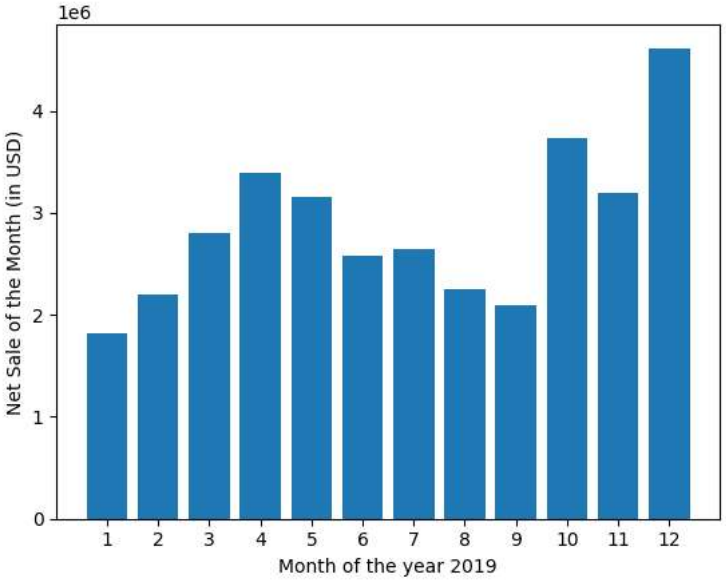
```
  Order ID                     Product  Quantity Ordered  Price Each  \
0   176558        USB-C Charging Cable                 2       11.95
2   176559  Bose SoundSport Headphones                 1       99.99
3   176560                Google Phone                 1      600.00
4   176560            Wired Headphones                 1       11.99
5   176561            Wired Headphones                 1       11.99

       Order Date                   Purchase Address  Month  Net_price
0  04/19/19 08:46          917 1st St, Dallas, TX 75001      4      23.90
2  04/07/19 22:30      682 Chestnut St, Boston, MA 02215      4      99.99
3  04/12/19 14:38  669 Spruce St, Los Angeles, CA 90001      4     600.00
4  04/12/19 14:38  669 Spruce St, Los Angeles, CA 90001      4      11.99
5  04/30/19 09:27      333 8th St, Los Angeles, CA 90001      4      11.99
                                                Order ID  \
Month
12     295665295666295667295668295669295670295671295 6...
10     259358259359259360259361259362259363259364259 3...
4      176558176559176560176561176562176563176564176 5...
11     278797278798278799278800278801278802278803278 8...
5      176978177551177778177778179076179076179134180 0...
3      150533150541150812151749152036153284153544153 8...
7      222910222911222912222913222914222915222916222 9...
6      209921209922209923209924209925209926209927209 9...
8      236670236671236672236673236674236675236676236 6...
2      150502150503150504150505150506150507150508150 5...
9      238834239285240636241054242343242865243667244 9...
1      297150297817297817297969298344299049299125300 3...

                                                 Product  Quantity Ordered
\
Month
12     Macbook Pro LaptopLG Washing MachineUSB-C Char...             28114
10     34in Ultrawide Monitor27in 4K Gaming MonitorAA...             22703
4      USB-C Charging CableBose SoundSport Headphones...             20558
11     Wired HeadphonesUSB-C Charging CableApple Airp...             19798
5      Apple Airpods Headphones27in FHD MonitoriPhone...             18667
3      AAA Batteries (4-pack)AAA Batteries (4-pack)Wi...             17005
7      Apple Airpods HeadphonesFlatscreen TVAA Batter...             16072
6      USB-C Charging CableMacbook Pro LaptopThinkPad...             15253
8      Wired HeadphonesBose SoundSport HeadphonesiPho...             13448
2      iPhoneAA Batteries (4-pack)27in 4K Gaming Moni...             13449
9      Apple Airpods Headphones34in Ultrawide Monitor...             13109
1      Lightning Charging CableiPhoneLightning Chargi...             10903

       Price Each                                         Order Date  \
Month
12     4588415.41  12/30/19 00:0112/29/19 07:0312/19 18:2112/2...
10     3715554.83  10/28/19 10:5610/28/19 17:2610/24/19 17:2010/1...
4      3367671.02  04/19/19 08:4604/07/19 22:3004/12/19 14:3804/1...
11     3180600.68  11/21/19 09:5411/17/19 10:0311/19/19 14:5611/2...
5      3135125.13  05/01/19 03:2905/01/19 00:1305/01/19 00:4805/0...
3      2791207.83  03/01/19 03:0603/01/19 01:0303/01/19 02:1803/0...
7      2632539.56  07/26/19 16:5107/05/19 08:5507/29/19 12:4107/2...
6      2562025.61  06/23/19 19:3406/30/19 10:0506/24/19 20:1806/0...
8      2230345.42  08/31/19 22:2108/15/19 15:1108/06/19 14:4008/2...
2      2188885.72  02/18/19 01:3502/13/19 07:2402/18/19 09:4602/0...
10     2084992.09  09/01/19 04:1309/01/19 01:0909/01/19 02:0709/0...
1      1811768.38  01/01/20 00:3801/20 00:2201/20 00:2201/0...

                                   Purchase Address   Net_price
Month
12     136 Church St, New York City, NY 10001562 2nd ...  4613443.34
10     193 Cherry St, Dallas, TX 75001225 5th St, Los...  3736726.88
4      917 1st St, Dallas, TX 75001682 Chestnut St, B...  3390670.24
```

In [16]:
```
monthly_sale=data.groupby('Month').sum()['Net_price']
months=range(1,13)
print(months)
plt.xlabel('Month number')
plt.ylabel('Sales in USD')
plt.xticks(months)
plt.bar(months,monthly_sale)
plt.show()
```

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



In [16]:
```
12      4588415.41  12/30/19 00:0112/29/19 07:0312/12/19 18:2112/2...
10      3715554.83  10/28/19 10:5610/28/19 17:2610/24/19 17:2010/1...
4       3661642.01  04/19/19 08:4604/07/19 22:3004/12/19 14:3804/1...
11      3180600.68  11/21/19 09:5411/17/19 10:0311/19/19 14:5611/2...
5       2916925.64  05/01/19 03:2905/01/19 00:1305/01/19 00:4805/0...
3       2791207.83  03/01/19 03:0603/01/19 01:0303/01/19 02:1803/0...
7       2632539.56  07/26/19 16:5107/05/19 08:5507/29/19 12:4107/2...
6       2562025.61  06/23/19 19:3406/30/19 10:0506/24/19 20:1806/0...
8       2230345.42  08/31/19 22:2108/15/19 15:1108/06/19 14:4008/2...
2       2188884.72  02/18/19 01:3502/13/19 07:2402/18/19 09:4602/0...
9       2084992.09  09/01/19 04:1309/01/19 01:0909/01/19 02:0709/0...
1       1811768.38  01/01/20 00:3801/01/20 00:2201/01/20 00:2201/0...
plt.bar(months,monthly_sale)
                                          Purchase Address   Net_price
plt.show()
12      136 Church St, New York City, NY 10001562 2nd ...  4613443.34
10      609 Cherry St, Dallas, TX 75001225 5th St, Los...  3736726.88
4       917 1st St, Dallas, TX 75001682 Chestnut St, B...  3390670.24
```



Finding the City with Highest Sale

In [17]:
```
data

# data['Purchase Address'].apply(lambda x: x.split(','))[1]
data['Purchase Address'].apply(lambda x: x.split(',')[1])

data['City']=data['Purchase Address'].apply(lambda x: x.split(',')[1])


print(data)
```

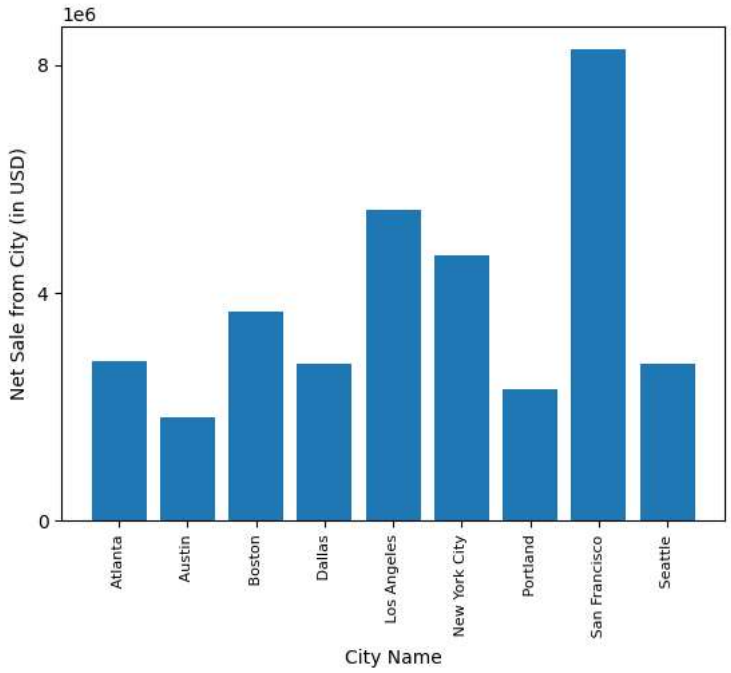
```
       Order ID                     Product  Quantity Ordered  Price Each  \
0        176558        USB-C Charging Cable                 2       11.95
2        176559  Bose SoundSport Headphones                 1       99.99
3        176560                Google Phone                 1      600.00
4        176560            Wired Headphones                 1       11.99
5        176561            Wired Headphones                 1       11.99
...         ...                         ...               ...         ...
186845   259353      AAA Batteries (4-pack)                 3        2.99
186846   259354                      iPhone                 1      700.00
186847   259355                      iPhone                 1      700.00
186848   259356      34in Ultrawide Monitor                 1      379.99
186849   259357        USB-C Charging Cable                 1       11.95

            Order Date                      Purchase Address  Month  \
0       04/19/19 08:46           917 1st St, Dallas, TX 75001      4
2       04/07/19 22:30       682 Chestnut St, Boston, MA 02215      4
3       04/12/19 14:38    669 Spruce St, Los Angeles, CA 90001      4
4       04/12/19 14:38    669 Spruce St, Los Angeles, CA 90001      4
5       04/30/19 09:27       333 8th St, Los Angeles, CA 90001      4
...                ...                                     ...    ...
```

In [17]:
```
data

# data['Purchase Address'].apply(lambda x: x.split(','))[1]
data['Purchase Address'].apply(lambda x: x.split(',')[1])

data['City']=data['Purchase Address'].apply(lambda x: x.split(',')[1])


print(data)
```

```
       Order ID                     Product  Quantity Ordered  Price Each  \
0        176558        USB-C Charging Cable                 2       11.95
2        176559  Bose SoundSport Headphones                 1       99.99
3        176560                Google Phone                 1      600.00
4        176560            Wired Headphones                 1       11.99
5        176561            Wired Headphones                 1       11.99
...         ...                         ...               ...         ...
186845   259353      AAA Batteries (4-pack)                 3        2.99
186846   259354                      iPhone                 1      700.00
186847   259355                      iPhone                 1      700.00
186848   259356      34in Ultrawide Monitor                 1      379.99
186849   259357        USB-C Charging Cable                 1       11.95

            Order Date                      Purchase Address  Month  \
0       04/19/19 08:46             917 1st St, Dallas, TX 75001      4
2       04/07/19 22:30        682 Chestnut St, Boston, MA 02215      4
3       04/12/19 14:38     669 Spruce St, Los Angeles, CA 90001      4
4       04/12/19 14:38     669 Spruce St, Los Angeles, CA 90001      4
5       04/30/19 09:27        333 8th St, Los Angeles, CA 90001      4
...                ...                                     ...    ...
186845  09/17/19 20:56   840 Highland St, Los Angeles, CA 90001      9
186846  09/01/19 16:00  216 Dogwood St, San Francisco, CA 94016      9
186847  09/23/19 07:39     220 12th St, San Francisco, CA 94016      9
186848  09/19/19 17:30   511 Forest St, San Francisco, CA 94016      9
186849  09/30/19 00:18   250 Meadow St, San Francisco, CA 94016      9

        Net_price            City
0           23.90          Dallas
2           99.99          Boston
3          600.00     Los Angeles
4           11.99     Los Angeles
5           11.99     Los Angeles
...           ...             ...
186845       8.97     Los Angeles
186846     700.00   San Francisco
186847     700.00   San Francisco
186848     379.99   San Francisco
186849      11.95   San Francisco

[185950 rows x 9 columns]
```

In [18]:
```
data.groupby('City').sum()

city_sale=data.groupby('City').sum()['Net_price']


# city=pd.distinct[data['City']] #no
# city=data['City'].unique()                                          #with char
#or
# city=[city for city in data.groupby('City')] #no
city=[city for city,data in data.groupby('City')]


plt.xlabel('City Name')
plt.ylabel('Net Sale from City (in USD)')
# plt.xticks(city)?????????????????????????????????????????????????????????????
plt.xticks(rotation='vertical',size=8)

# y=float(range(1,5))?????????????????????????????????????????????????????????????
# yticks=y*0.4

plt.yticks([0,4000000,8000000,12000000,16000000])

plt.bar(city,city_sale)

plt.show()
data.groupby('City').sum()
```

In [18]:
```
data.groupby('City').sum()

city_sale=data.groupby('City').sum()['Net_price']


# city=pd.distinct[data['City']] #no
# city=data['City'].unique()                                    #with char
#or
# city=[city for city in data.groupby('City')] #no
city=[city for city,data in data.groupby('City')]


plt.xlabel('City Name')
plt.ylabel('Net Sale from City (in USD)')
# plt.xticks(city)?????????????????????????????????????????????????????????
plt.xticks(rotation='vertical',size=8)

# y=float(range(1,5))?????????????????????????????????????????????????????????
# yticks=y*0.4

plt.yticks([0,4000000,8000000,12000000,16000000])

plt.bar(city,city_sale)

plt.show()
data.groupby('City').sum()
```



Out[18]:

City	Order ID	Product	Quantity Ordered	I
Atlanta	176564176571176582176589176623176627176634176 6...	USB-C Charging CableLightning Charging CableBo...	16602	2
Austin	176591176609176674176677176686176701176705176 7...	Apple Airpods HeadphonesApple Airpods Headphon...	11153	1
Boston	176559176566176578176581176585176585176600176 6...	Bose SoundSport HeadphonesWired HeadphonesAppl...	22528	3
Dallas	176558176569176570176577176596176602176610176 6...	USB-C Charging Cable27in 4K Gaming MonitorAA B...	16730	2

Out[18]:

City	Order ID	Product	Quantity Ordered	I
Atlanta	176564176571176582176589176623176627176634176 6...	USB-C Charging CableLightning Charging CableBo...	16602	2
Austin	176591176609176674176677176686176701176705176 7...	Apple Airpods HeadphonesApple Airpods Headphon...	11153	1
Boston	176559176566176578176581176585176585176600176 6...	Bose SoundSport HeadphonesWired HeadphonesAppl...	22528	3
Dallas	176558176569176570176577176596176602176610176 6...	USB-C Charging Cable27in 4K Gaming MonitorAA B...	16730	2
Los Angeles	176560176560176561176567176574176574176576176 5...	Google PhoneWired HeadphonesWired HeadphonesGo...	33289	5
New York City	176572176575176579176590176599176613176637176 6...	Apple Airpods HeadphonesAAA Batteries (4-pack)...	27932	4
Portland	176583176593176617176642176646176648176662176 7...	AAA Batteries (4-pack)Lightning Charging Cable...	14053	2
San Francisco	176562176565176573176584176586176586176594176 5...	USB-C Charging CableMacbook Pro LaptopUSB-C Ch...	50239	8
Seattle	176563176568176588176612176624176654176663176 6...	Bose SoundSport HeadphonesLightning Charging C...	16553	2

In [20]:
```
#there could be multiple cities with same name hence state is also important
```

In [19]:
```
data['City'].unique()
```

Out[19]:
```
array([' Dallas', ' Boston', ' Los Angeles', ' San Francisco', ' Seattle',
       ' Atlanta', ' New York City', ' Portland', ' Austin'], dtype=object)
```

```
def getcity(x):
    return x.split(',')[1]

def getstate(x):
    y=x.split(',')[2]
    return y.split(' ')[1]

def getpin(x):
    y=x.split(',')[2]
    return y.split(' ')[2]

# data['CitywithState']=f"{getcity(data['Purchase Address'])} {getstate(data['
data['CitywithState']=data['Purchase Address'].apply(lambda x: f"{getcity(x)}

data['City Pin']=data['Purchase Address'].apply(lambda x: f"{getpin(x)}")

print(data)
data.groupby('CitywithState').sum()
# data.groupby('City').sum()
# data.groupby('City Pin').sum()
```

```
4       04/12/19 14:38      669 Spruce St, Los Angeles, CA 90001      4
5       04/30/19 09:27        333 8th St, Los Angeles, CA 90001       4
...               ...                                         ...    ...
186845  09/17/19 20:56   840 Highland St, Los Angeles, CA 90001       9
186846  09/01/19 16:00  216 Dogwood St, San Francisco, CA 94016       9
186847  09/23/19 07:39     220 12th St, San Francisco, CA 94016       9
186848  09/19/19 17:30   511 Forest St, San Francisco, CA 94016       9
```

In [20]:
```
#there could be multiple cities with same name hence state is also important
```

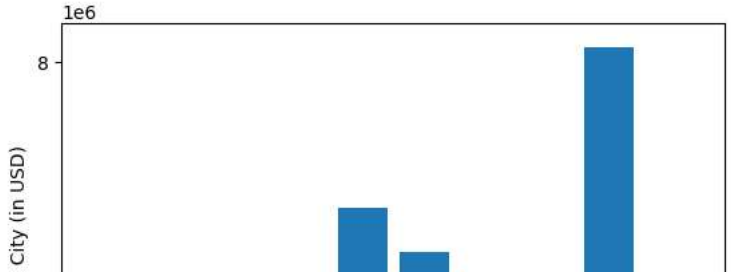
In [19]:
```
data['City'].unique()
```

Out[19]:
```
array([' Dallas', ' Boston', ' Los Angeles', ' San Francisco', ' Seattle',
       ' Atlanta', ' New York City', ' Portland', ' Austin'], dtype=object)
```

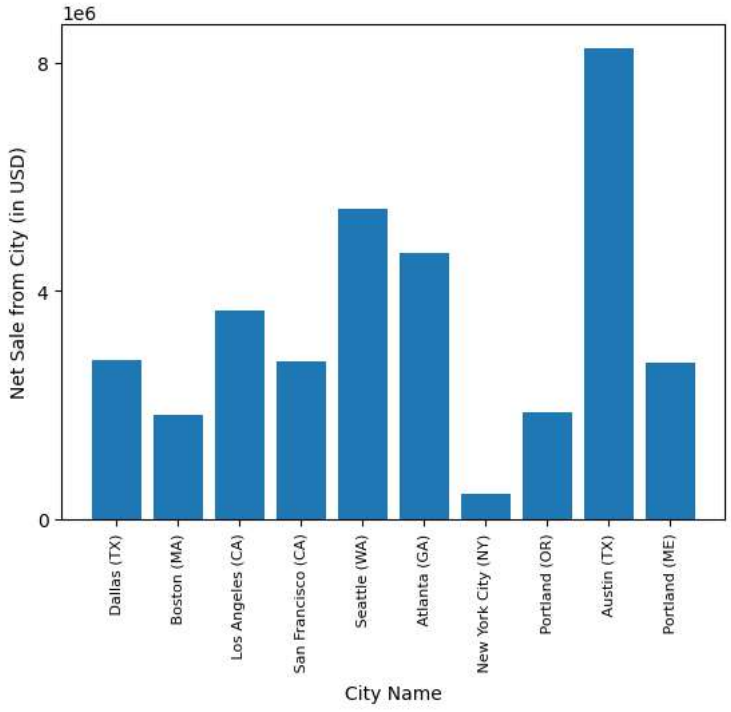
```
def getcity(x):
    return x.split(',')[1]

def getstate(x):
    y=x.split(',')[2]
    return y.split(' ')[1]

def getpin(x):
    y=x.split(',')[2]
    return y.split(' ')[2]

# data['CitywithState']=f"{getcity(data['Purchase Address'])} {getstate(data['
data['CitywithState']=data['Purchase Address'].apply(lambda x: f"{getcity(x)}

data['City Pin']=data['Purchase Address'].apply(lambda x: f"{getpin(x)}")

print(data)
data.groupby('CitywithState').sum()
# data.groupby('City').sum()
# data.groupby('City Pin').sum()
```

```
4       04/12/19 14:38     669 Spruce St, Los Angeles, CA 90001      4
5       04/30/19 09:27       333 8th St, Los Angeles, CA 90001      4
...               ...                                      ...    ...
186845  09/17/19 20:56   840 Highland St, Los Angeles, CA 90001      9
186846  09/01/19 16:00  216 Dogwood St, San Francisco, CA 94016      9
186847  09/23/19 07:39     220 12th St, San Francisco, CA 94016      9
186848  09/19/19 17:30   511 Forest St, San Francisco, CA 94016      9
186849  09/30/19 00:18   250 Meadow St, San Francisco, CA 94016      9

        Net_price            City        CitywithState City Pin
0           23.90          Dallas          Dallas (TX)    75001
2           99.99          Boston          Boston (MA)    02215
3          600.00     Los Angeles     Los Angeles (CA)    90001
4           11.99     Los Angeles     Los Angeles (CA)    90001
5           11.99     Los Angeles     Los Angeles (CA)    90001
...           ...             ...                 ...      ...
186845       8.97     Los Angeles     Los Angeles (CA)    90001
186846     700.00   San Francisco   San Francisco (CA)    94016
186847     700.00   San Francisco   San Francisco (CA)    94016
```

In [21]:
```
city_sale=data.groupby('CitywithState').sum()['Net_price']


city=data['CitywithState'].unique()                                    #wi


plt.xlabel('City Name')
plt.ylabel('Net Sale from City (in USD)')
plt.xticks(rotation='vertical',size=8)

plt.yticks([0,4000000,8000000,12000000,16000000])

plt.bar(city,city_sale)

plt.show()
data.groupby('CitywithState').sum()
```



In [21]:
```
city_sale=data.groupby('CitywithState').sum()['Net_price']


city=data['CitywithState'].unique()                                    #wi


plt.xlabel('City Name')
plt.ylabel('Net Sale from City (in USD)')
plt.xticks(rotation='vertical',size=8)

plt.yticks([0,4000000,8000000,12000000,16000000])

plt.bar(city,city_sale)

plt.show()
data.groupby('CitywithState').sum()
```



Out[21]:

CitywithState	Order ID	Product	Quantity Ordered
Atlanta (GA)	176564176571176582176589176623176627176634176 6...	USB-C Charging CableLightning Charging CableBo...	16602
Austin (TX)	176591176609176674176677176686176701176705176 7...	Apple Airpods HeadphonesApple Airpods Headphon...	11153
Boston (MA)	176559176566176578176581176585176585176600176 6...	Bose SoundSport HeadphonesWired HeadphonesAppl...	22528
Dallas (TX)	176558176569176570176577176596176602176610176 6...	USB-C Charging Cable27in 4K Gaming MonitorAA B...	16730

Out[21]:

CitywithState	Order ID	Product	Quantity Ordered
Atlanta (GA)	176564176571176582176589176623176627176634176 6...	USB-C Charging CableLightning Charging CableBo...	16602
Austin (TX)	176591176609176674176677176686176701176705176 7...	Apple Airpods HeadphonesApple Airpods Headphon...	11153
Boston (MA)	176559176566176578176581176585176585176600176 6...	Bose SoundSport HeadphonesWired HeadphonesAppl...	22528
Dallas (TX)	176558176569176570176577176596176602176610176 6...	USB-C Charging Cable27in 4K Gaming MonitorAA B...	16730
Los Angeles (CA)	176560176560176561176567176574176574176576176 5...	Google PhoneWired HeadphonesWired HeadphonesGo...	33289
New York City (NY)	176572176575176579176590176599176613176637176 6...	Apple Airpods HeadphonesAAA Batteries (4-pack)...	27932
Portland (ME)	176773176879176909177072177230177233177299177 3...	AAA Batteries (4-pack)Flatscreen TVWired Headp...	2750
Portland (OR)	176583176593176617176642176646176648176662176 7...	AAA Batteries (4-pack)Lightning Charging Cable...	11303
San Francisco (CA)	176562176565176573176584176586176586176594176 5...	USB-C Charging CableMacbook Pro LaptopUSB-C Ch...	50239
Seattle (WA)	176563176568176588176612176624176654176663176 6...	Bose SoundSport HeadphonesLightning Charging C...	16553

here is a great error of respectiveness of city name list and group by sales sum list

In [22]:
```
#Seattle (WA)  sorted by status
city_sale=data.groupby('CitywithState').sum().sort_values(by=['CitywithState']
```

```
# city=data['CitywithState'].sort_values(by=['CitywithState'],ascending=False)
city=data['CitywithState'].sort_values(ascending=False).unique()


plt.xlabel('City Name')
plt.ylabel('Net Sale from City (in USD)')
plt.xticks(rotation='vertical',size=8)

plt.yticks([0,4000000,8000000,12000000,16000000])

plt.bar(city,city_sale)

plt.show()
data.groupby('CitywithState').sum()
```

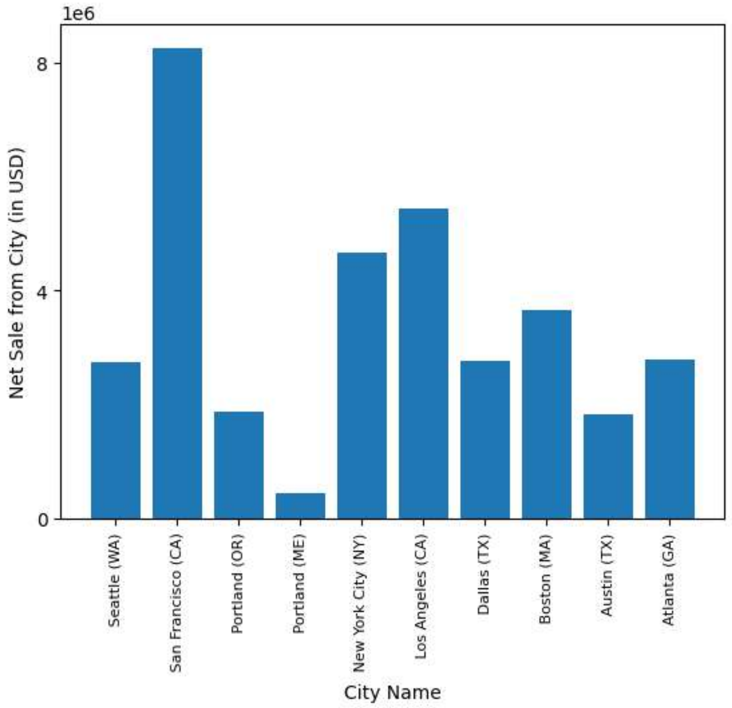
(CA)		LaptopUSB-C Ch...	
Seattle (WA)	1765631765681765881766121766241766541766631766...	Bose SoundSport HeadphonesLightning Charging C...	16553

here is a great error of respectiveness of city name list and group by sales sum list

In [22]:

```
#hence used sort by clause

city_sale=data.groupby('CitywithState').sum().sort_values(by=['CitywithState']
```

```
# city=data['CitywithState'].sort_values(by=['CitywithState'],ascending=False)
city=data['CitywithState'].sort_values(ascending=False).unique()


plt.xlabel('City Name')
plt.ylabel('Net Sale from City (in USD)')
plt.xticks(rotation='vertical',size=8)

plt.yticks([0,4000000,8000000,12000000,16000000])

plt.bar(city,city_sale)

plt.show()
data.groupby('CitywithState').sum()
```



Out[22]:

	Order ID	Product	Quantity Ordered
CitywithState			
Atlanta (GA)	1765641765711765821765891766231766271766341766...	USB-C Charging CableLightning Charging CableBo...	16602
Austin (TX)	1765911766091766741766771766861767011767051767...	Apple Airpods HeadphonesApple Airpods Headphon...	11153
Boston (MA)	1765591765661765781765811765851765851766001766...	Bose SoundSport HeadphonesWired HeadphonesAppl...	22528
Dallas (TX)	1765581765691765701765771765961766021766101766...	USB-C Charging Cable27in 4K Gaming MonitorAA B...	16730

Out[22]:

CitywithState	Order ID	Product	Quantity Ordered
Atlanta (GA)	176564176571176582176589176623176627176634176 6...	USB-C Charging CableLightning Charging CableBo...	16602
Austin (TX)	176591176609176674176677176686176701176705176 7...	Apple Airpods HeadphonesApple Airpods Headphon...	11153
Boston (MA)	176559176566176578176581176585176585176600176 6...	Bose SoundSport HeadphonesWired HeadphonesAppl...	22528
Dallas (TX)	176558176569176570176577176596176602176610176 6...	USB-C Charging Cable27in 4K Gaming MonitorAA B...	16730
Los Angeles (CA)	176560176560176561176567176574176574176576176 5...	Google PhoneWired HeadphonesWired HeadphonesGo...	33289
New York City (NY)	176572176575176579176590176599176613176637176 6...	Apple Airpods HeadphonesAAA Batteries (4-pack)...	27932
Portland (ME)	176773176879176909177072177230177233177299177 3...	AAA Batteries (4-pack)Flatscreen TVWired Headp...	2750
Portland (OR)	176583176593176617176642176646176648176662176 7...	AAA Batteries (4-pack)Lightning Charging Cable...	11303
San Francisco (CA)	176562176565176573176584176586176586176594176 5...	USB-C Charging CableMacbook Pro LaptopUSB-C Ch...	50239

In [23]:
```
data
```

Out[23]:

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	Month	Net_price	City
~~Seattle (WA)~~	~~176563176568176589176612176624176654176663176 6...~~	~~Bose SoundSport HeadphonesLightning Charging C...~~							~~16553~~
0	176558	USB-C Charging Cable	2	11.95	04/19/19 08:46	917 1st St, Dallas, TX 75001	4	23.90	Dallas
2	176559	Bose SoundSport Headphones	1	99.99	04/07/19 22:30	682 Chestnut St, Boston, MA 02215	4	99.99	Boston
3	176560	Google Phone	1	600.00	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001	4	600.00	Los Angeles
4	176560	Wired Headphones	1	11.99	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001	4	11.99	Los Angeles
5	176561	Wired Headphones	1	11.99	04/30/19 09:27	333 8th St, Los Angeles, CA 90001	4	11.99	Los Angeles
...	...	...	...	...	...	...	...	...	...
		AAA				840 Highland			

Finding the Best time for advertisement

In [23]:
```
data
```

Out[23]:

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	Month	Net_price	City
0	176558	USB-C Charging Cable	2	11.95	04/19/19 08:46	917 1st St, Dallas, TX 75001	4	23.90	Dallas
2	176559	Bose SoundSport Headphones	1	99.99	04/07/19 22:30	682 Chestnut St, Boston, MA 02215	4	99.99	Boston
3	176560	Google Phone	1	600.00	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001	4	600.00	Los Angeles
4	176560	Wired Headphones	1	11.99	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001	4	11.99	Los Angeles
5	176561	Wired Headphones	1	11.99	04/30/19 09:27	333 8th St, Los Angeles, CA 90001	4	11.99	Los Angeles
...	...	...	...	...	...	...	...	...	...
186845	259353	AAA Batteries (4-pack)	3	2.99	09/17/19 20:56	840 Highland St, Los Angeles, CA 90001	9	8.97	Los Angeles
186846	259354	iPhone	1	700.00	09/01/19 16:00	216 Dogwood St, San Francisco, CA 94016	9	700.00	San Francisco
186847	259355	iPhone	1	700.00	09/23/19 07:39	220 12th St, San Francisco, CA 94016	9	700.00	San Francisco
186848	259356	34in Ultrawide Monitor	1	379.99	09/19/19 17:30	511 Forest St, San Francisco, CA 94016	9	379.99	San Francisco
186849	259357	USB-C Charging Cable	1	11.95	09/30/19 00:18	250 Meadow St, San Francisco, CA 94016	9	11.95	San Francisco

185950 rows × 11 columns

Finding the Best time for advertisement

In [24]:
```
# data['Order Date'].split(' ').split(':')[0]
data['Order Time']=data['Order Date'].apply(lambda x: (x.split(' ')[1]).split(

#OR
#data['Order Date']=pd.to_datetime(data['Order Date'])

# data['Hour']=data['Order Date'].dt.hour
# data['Minutes']=data['Order Date'].dt.minute


data.head()
```

Out[24]:

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

In [24]:
```
# data['Order Date'].split(' ').split(':')[0]
data['Order Time']=data['Order Date'].apply(lambda x: (x.split(' ')[1]).split(

#OR
#data['Order Date']=pd.to_datetime(data['Order Date'])

# data['Hour']=data['Order Date'].dt.hour
# data['Minutes']=data['Order Date'].dt.minute


data.head()
```

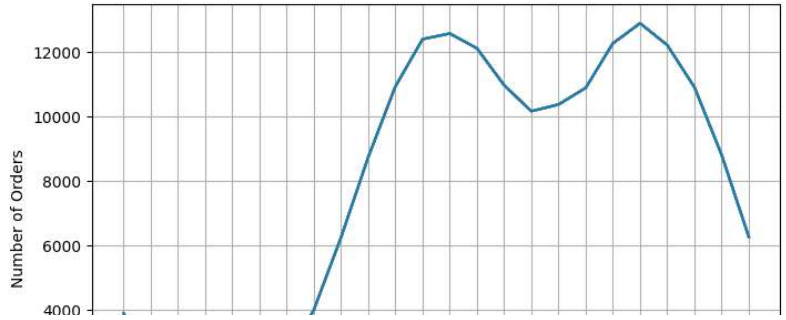
Out[24]:

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

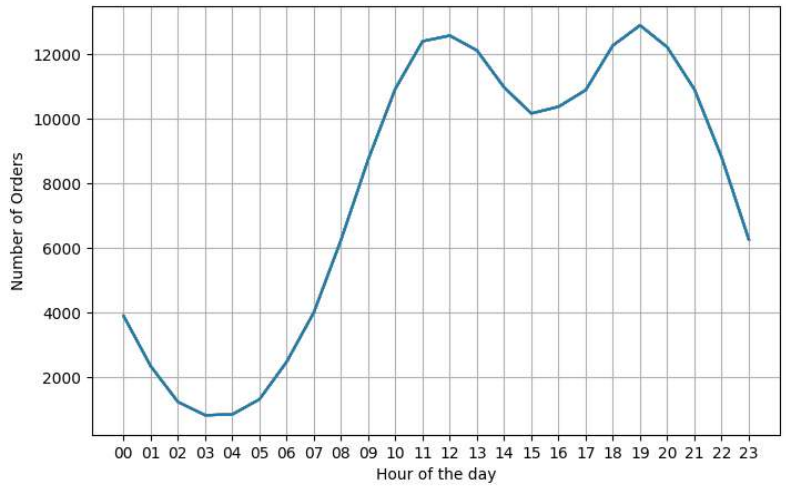
In [25]:
```
print(data.groupby('Order Time').sum().sort_values(by='Net_price',ascending=Fa
# print(data.groupby('Order Time').sum().sort_values(by='Quantity Ordered',asc


print(data.groupby('Order Time').count())
```

```
2350
02                    1243   1243        1243   1243            1243
1243
03                     831    831         831    831             831
831
04                     854    854         854    854             854
854
05                    1321   1321        1321   1321            1321
1321
06                    2482   2482        2482   2482            2482
2482
07                    4011   4011        4011   4011            4011
4011
08                    6256   6256        6256   6256            6256
6256
09                    8748   8748        8748   8748            8748
8748
10                   10944  10944       10944  10944           10944      1
10944
11                   13411  13411       13411  13411           13411      1
```

In [26]:
```
hours=[hour for hour,df in data.groupby('Order Time')]
order_count_hour=data.groupby('Order Time').count()
plt.figure(figsize=(8,5),dpi=100)

plt.ylabel("Number of Orders")
plt.xlabel("Hour of the day")
# plt.xticks(hours)

plt.grid()

plt.plot(hours,order_count_hour)

plt.show()
```



In [26]:
```
hours=[hour for hour,df in data.groupby('Order Time')]
order_count_hour=data.groupby('Order Time').count()
plt.figure(figsize=(8,5),dpi=100)

plt.ylabel("Number of Orders")
plt.xlabel("Hour of the day")
# plt.xticks(hours)

plt.grid()

plt.plot(hours,order_count_hour)

plt.show()
```



Finding the products sold together

In [27]:
```
data.columns
data2=data[['Order ID', 'Product', 'Quantity Ordered', 'Price Each', 'Order Da
       'Net_price']]
data2

data2['Grouped']=data2.groupby('Order ID')['Product'].transform(lambda x: ','.
print(data2.head())
```

```
  Order ID                    Product  Quantity Ordered  Price Each  \
0   176558        USB-C Charging Cable                 2       11.95
2   176559  Bose SoundSport Headphones                 1       99.99
3   176560                Google Phone                 1      600.00
4   176560            Wired Headphones                 1       11.99
5   176561            Wired Headphones                 1       11.99

       Order Date  Month      CitywithState  Net_price  \
0  04/19/19 08:46      4         Dallas (TX)      23.90
2  04/07/19 22:30      4         Boston (MA)      99.99
3  04/12/19 14:38      4    Los Angeles (CA)     600.00
4  04/12/19 14:38      4    Los Angeles (CA)      11.99
5  04/30/19 09:27      4    Los Angeles (CA)      11.99
```

In [28]:
```
data2['Grouped'].count()
```

Out[28]: 5

```
C:\Users\Vedasri Last\AppData\Local\Temp\ipykernel_21308\3172522887.py:6: Set
tingWithCopyWarning:
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 (https://pand
as.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-v
ersus-a-copy)
  data2['Grouped']=data2.groupby('Order ID')['Product'].transform(lambda x:
','.join(x))
```

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Month	CitywithState	Net_price
Grouped								
20in Monitor	3942	3942	3942	3942	3942	3942	3942	3942
20in Monitor,20in Monitor	6	6	6	6	6	6	6	6
20in Monitor,27in FHD Monitor	4	4	4	4	4	4	4	4
20in Monitor,34in Ultrawide Monitor	2	2	2	2	2	2	2	2
20in Monitor,AA Batteries (4-pack)	12	12	12	12	12	12	12	12
20in Monitor,USB-C Charging Cable	24	24	24	24	24	24	24	24
iPhone,Vareebadd Phone	6	6	6	6	6	6	6	6
iPhone,Wired Headphones	722	722	722	722	722	722	722	722
iPhone,Wired Headphones,Lightning Charging Cable	3	3	3	3	3	3	3	3
iPhone,iPhone	2	2	2	2	2	2	2	2

In [28]:
```
data2['Grouped'].count()
```

Out[28]:

Product	Grouped
20in Monitor	3942
20in Monitor,20in Monitor	6
20in Monitor,27in FHD Monitor	4
20in Monitor,34in Ultrawide Monitor	2
20in Monitor,AA Batteries (4-pack)	12
iPhone,USB-C Charging Cable	24
iPhone,Vareebadd Phone	6
iPhone,Wired Headphones	722
iPhone,Wired Headphones,Lightning Charging Cable	3
iPhone,iPhone	2

385 rows × 8 columns

```
2  04/07/19 22:30       4       Boston (MA)      99.99
3  04/12/19 14:38       4  Los Angeles (CA)     600.00
4  04/12/19 14:38       4  Los Angeles (CA)      11.99
5  04/30/19 09:27       4  Los Angeles (CA)      11.99

C:\Users\Vreddast\AppData\Local\Temp\ipykernel_21308\3172522887.py:6: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)
  data2['Grouped']=data2.groupby('Order ID')['Product'].transform(lambda x: ','.join(x))
```

In [29]:
```
#these have some duplication errors so restart with elimination of duplicates
```

In [30]:
```
# data3=data[['Order ID', 'Product', 'Quantity Ordered', 'Price Each', 'Order
#         'Net_price']]
# data3.head()

data3=data[data['Order ID'].duplicated(keep=False)]
#did it to only consider the record which are having double Order id
#that the products which are going to be sold/ordered together are going to ha
#hence only those records will be necessary for analysis of this question

data3.head(30)
```

Out[30]:

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	Month	Net_price	City
3	176560	Google Phone	1	600.00	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001	4	600.00	Los Angeles
4	176560	Wired Headphones	1	11.99	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001	4	11.99	Los Angeles
18	176574	Google Phone	1	600.00	04/03/19 19:42	20 Hill St, Los Angeles, CA 90001	4	600.00	Los Angeles
19	176574	USB-C Charging Cable	1	11.95	04/03/19 19:42	20 Hill St, Los Angeles, CA 90001	4	11.95	Los Angeles
30	176585	Bose SoundSport Headphones	1	99.99	04/07/19 11:31	823 Highland St, Boston, MA 02215	4	99.99	Boston

In [31]:
```
data3['Grouped']=data3.groupby('Order ID')['Product'].transform(lambda x: ','.
print(data3.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  Net_price
\
3   04/12/19 14:38  669 Spruce St, Los Angeles, CA 90001      4     600.00
4   04/12/19 14:38  669 Spruce St, Los Angeles, CA 90001      4      11.99
18  04/03/19 19:42     20 Hill St, Los Angeles, CA 90001      4     600.00
19  04/03/19 19:42     20 Hill St, Los Angeles, CA 90001      4      11.95
30  04/07/19 11:31     823 Highland St, Boston, MA 02215      4      99.99

            City      CitywithState City Pin Order Time  \
3    Los Angeles   Los Angeles (CA)    90001          14
4    Los Angeles   Los Angeles (CA)    90001          14
18   Los Angeles   Los Angeles (CA)    90001          19
19   Los Angeles   Los Angeles (CA)    90001          19
30        Boston        Boston (MA)    02215          11

                                    Grouped
3                  Google Phone,Wired Headphones
4                  Google Phone,Wired Headphones
18               Google Phone,USB-C Charging Cable
```

In [31]:
```
data3['Grouped']=data3.groupby('Order ID')['Product'].transform(lambda x: ','.
print(data3.head(7))
```

```
   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  Net_price
\
3   04/12/19 14:38  669 Spruce St, Los Angeles, CA 90001      4     600.00
4   04/12/19 14:38  669 Spruce St, Los Angeles, CA 90001      4      11.99
18  04/03/19 19:42     20 Hill St, Los Angeles, CA 90001      4     600.00
19  04/03/19 19:42     20 Hill St, Los Angeles, CA 90001      4      11.95
30  04/07/19 11:31     823 Highland St, Boston, MA 02215      4      99.99

            City      CitywithState City Pin Order Time  \
3    Los Angeles   Los Angeles (CA)    90001          14
4    Los Angeles   Los Angeles (CA)    90001          14
18   Los Angeles   Los Angeles (CA)    90001          19
19   Los Angeles   Los Angeles (CA)    90001          19
30        Boston        Boston (MA)    02215          11

                                        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...

C:\Users\Vatsal_Fast\AppData\Local\Temp\ipykernel_21308\2648751239.py:1: Set
tingWithCopyWarning:
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 (https://pand
as.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-v
ersus-a-copy)
  data3['Grouped']=data3.groupby('Order ID')['Product'].transform(lambda x:
','.join(x))
```

In [32]:
```
data4=data3[['Order ID','Grouped']].drop_duplicates()
data4
```

Out[32]:

	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
...	...	...
186781	259296	Apple Airpods Headphones,Apple Airpods Headphones
186783	259297	iPhone,Lightning Charging Cable,Lightning Char...
186791	259303	34in Ultrawide Monitor,AA Batteries (4-pack)
186803	259314	Wired Headphones,AAA Batteries (4-pack)
186841	259350	Google Phone,USB-C Charging Cable

7136 rows × 2 columns

In [33]:
```
data4.groupby('Grouped').count()
```

Out[33]:

Grouped	Order ID
20in Monitor,20in Monitor	3
20in Monitor,27in FHD Monitor	2
20in Monitor,34in Ultrawide Monitor	1
20in Monitor,AA Batteries (4-pack)	6
20in Monitor,AAA Batteries (4-pack)	3
...	...
iPhone,USB-C Charging Cable	12
iPhone,Vareebadd Phone	3
iPhone,Wired Headphones	361
iPhone,Wired Headphones,Lightning Charging Cable	1
iPhone,iPhone	1

366 rows × 1 columns

imp to do library

In [33]:
```
data4.groupby('Grouped').count()
```

Out[33]:

Order ID	Grouped
186783	259297 iPhone,Lightning Charging Cable,Lightning Char...
186791	259303 34in Ultrawide Monitor,AA Batteries (4-pack)
186803	259314 Wired Headphones,AAA Batteries (4-pack)
186841	259350 Google Phone,USB-C Charging Cable

7136 rows × 2 columns

Grouped	
20in Monitor,20in Monitor	3
20in Monitor,27in FHD Monitor	2
20in Monitor,34in Ultrawide Monitor	1
20in Monitor,AA Batteries (4-pack)	6
20in Monitor,AAA Batteries (4-pack)	3
...	...
iPhone,USB-C Charging Cable	12
iPhone,Vareebadd Phone	3
iPhone,Wired Headphones	361
iPhone,Wired Headphones,Lightning Charging Cable	1
iPhone,iPhone	1

366 rows × 1 columns

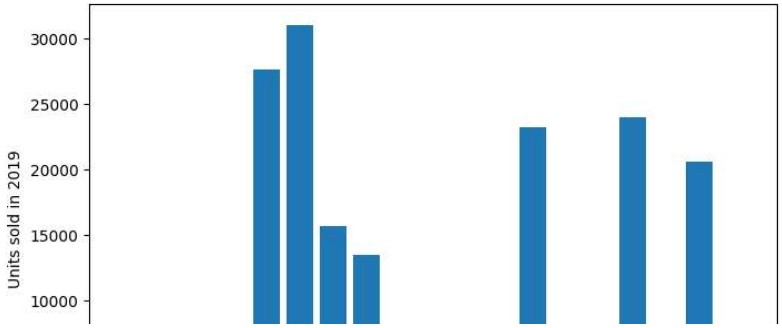
imp to do library

In [34]:
```
from collections import Counter
from itertools import combinations

count=Counter()

for r in data4['Grouped']:
    row_list=r.split(',')
    count.update(Counter(combinations(row_list,2)))


count

count.most_common(10)

for k,v in count.most_common(10):
    print(k," :    ",v)
```

```
('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
```

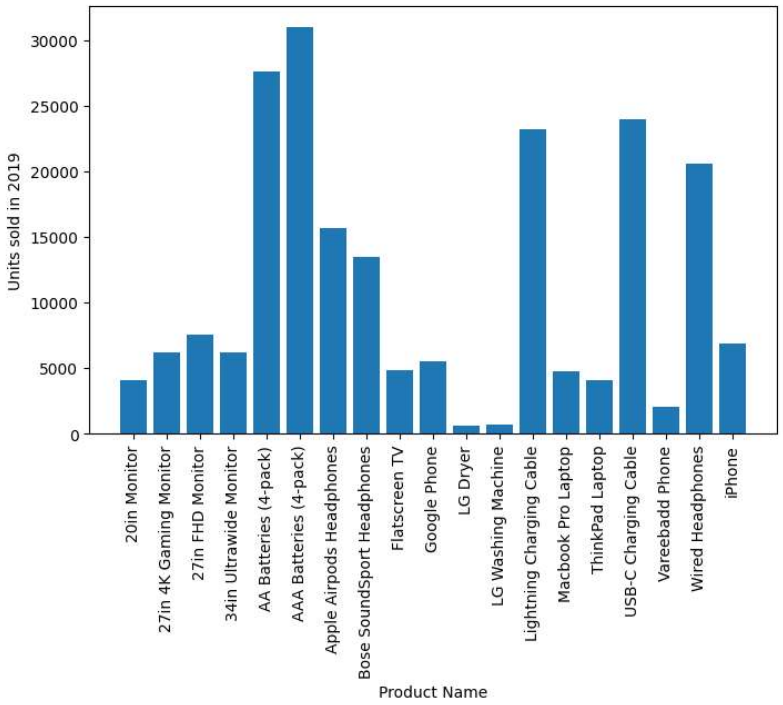
Finding the product Sold the most and why

In [35]:
```
product_count=data.groupby('Product').sum()['Quantity Ordered']
products=[prod for prod,df in data.groupby('Product')]

plt.figure(figsize=(8,5),dpi=100)

plt.xlabel('Product Name')
plt.ylabel('Units sold in 2019')
plt.xticks(rotation='vertical')

plt.bar(products,product_count)
plt.show()
```

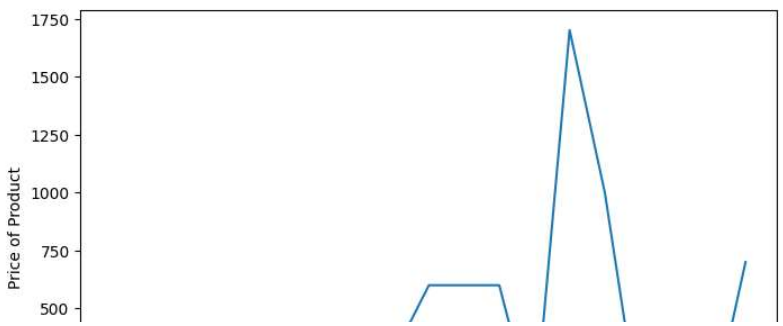


Finding the product Sold the most and why

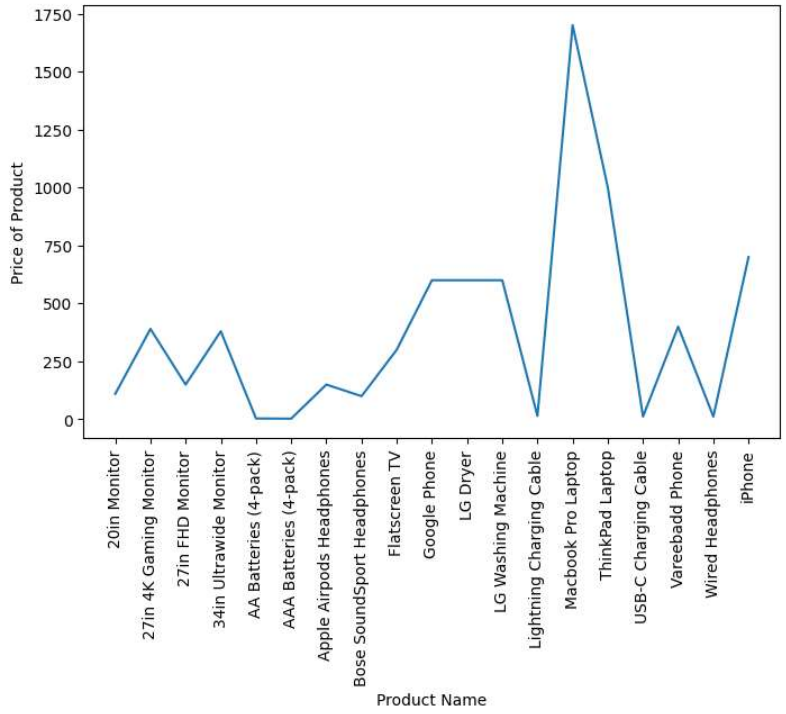
In [35]:
```
product_count=data.groupby('Product').sum()['Quantity Ordered']
products=[prod for prod,df in data.groupby('Product')]

plt.figure(figsize=(8,5),dpi=100)

plt.xlabel('Product Name')
plt.ylabel('Units sold in 2019')
plt.xticks(rotation='vertical')

plt.bar(products,product_count)
plt.show()
```



In [39]:
```
# price_each=[price for price, df in data.groupby('Product').mean()]
# price_each=[price for price, df in data.groupby('Price Each').mean()]
#price_each=[price for price in data.groupby('Product').mean()['Price Each']]
price_each = [df['Price Each'].mean() for _, df in data.groupby('Product')]

products=[prod for prod,df in data.groupby('Product')]

plt.figure(figsize=(8,5),dpi=100)

plt.xlabel('Product Name')
plt.ylabel('Price of Product')
plt.xticks(rotation='vertical')

plt.plot(products,price_each)
plt.show()
```



In [39]:
```
# price_each=[price for price, df in data.groupby('Product').mean()]
# price_each=[price for price, df in data.groupby('Price Each').mean()]
#price_each=[price for price in data.groupby('Product').mean()['Price Each']]
price_each = [df['Price Each'].mean() for _, df in data.groupby('Product')]

products=[prod for prod,df in data.groupby('Product')]

plt.figure(figsize=(8,5),dpi=100)

plt.xlabel('Product Name')
plt.ylabel('Price of Product')
plt.xticks(rotation='vertical')

plt.plot(products,price_each)
plt.show()
```

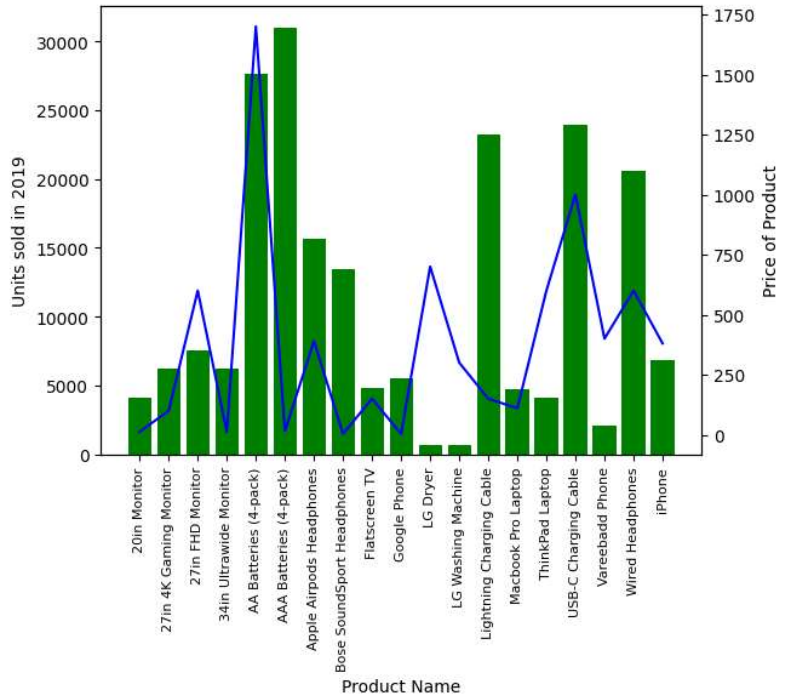


In [41]:
```
products=[prod for prod,df in data.groupby('Product')]

product_count=data.groupby('Product').sum()['Quantity Ordered']

#price_each=[price for price in data.groupby('Product').mean()['Price Each']]
price_each = [data.groupby('Product')['Price Each'].mean().loc[prod] for prod


#plt.figure(figsize=(8,5),dpi=100)??????????????????????????????????????????


fig,ax1=plt.subplots()


ax2=ax1.twinx()

ax1.bar(products,product_count,color='g')
ax2.plot(products,price_each,'b-')

ax1.set_xlabel('Product Name')
ax1.set_ylabel('Units sold in 2019')
ax2.set_ylabel('Price of Product')
# ax1.set_xticks(rotation='vertical',size=8)
ax1.set_xticklabels(products,rotation='vertical',size=8)


plt.show()
```

```
C:\Users\Vatsal_Fast\AppData\Local\Temp\ipykernel_21308\1978691006.py:24: Us
```

In [41]:
```
products=[prod for prod,df in data.groupby('Product')]

product_count=data.groupby('Product').sum()['Quantity Ordered']

#price_each=[price for price in data.groupby('Product').mean()['Price Each']]
price_each = [data.groupby('Product')['Price Each'].mean().loc[prod] for prod


#plt.figure(figsize=(8,5),dpi=100)??????????????????????????????????????????


fig,ax1=plt.subplots()


ax2=ax1.twinx()

ax1.bar(products,product_count,color='g')
ax2.plot(products,price_each,'b-')

ax1.set_xlabel('Product Name')
ax1.set_ylabel('Units sold in 2019')
ax2.set_ylabel('Price of Product')
# ax1.set_xticks(rotation='vertical',size=8)
ax1.set_xticklabels(products,rotation='vertical',size=8)


plt.show()
```

```
C:\Users\Vatsal_Fast\AppData\Local\Temp\ipykernel_21308\1978691006.py:24: Us
erWarning: FixedFormatter should only be used together with FixedLocator
  ax1.set_xticklabels(products,rotation='vertical',size=8)
```



In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]: