

```
In [1]: import pandas as pd
```

```
import matplotlib.pyplot as plt
```

```
In [2]: data1=pd.read_csv(r"C:\Users\Vatsal_Fast\Desktop\DataSciencePractice\keithgaliE1.csv")
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\DataSciencePractice")]
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]: year_data=pd.DataFrame()

for file in files:
#    month_data=pd.read_csv(r"H:\IT Courses\Data_science\keithgaillypython\Pana
    month_data=pd.read_csv("C:\\Users\\Vatsal_Fast\\Desktop\\DataSciencePractice\\
#    year_data=pd.concat[year_data,month_data]
    year_data=pd.concat([year_data,month_data])

# net records should be around 186850
print(year_data)
# year_data=pd.to_csv(r"H:\IT Courses\Data_science\keithgaillypython\Pandas\anal
year_data.to_csv(r"C:\\Users\\Vatsal_Fast\\Desktop\\DataSciencePractice\\keithgaliEle
```

	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]: data=pd.read_csv(r"C:\\Users\\Vatsal_Fast\\Desktop\\DataSciencePractice\\keithgaliEle
print(data.head())
```

	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	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			

Finding the best month for sales

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

	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
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 [8]: # data['Month']=data['Month'].astype('int16')
# error hence data cleaning
```

```
In [9]: # print(data[data['Month']=='NaN'])?????????????????????????????
data=data.dropna(how='all')
data.head()
```

Out[9]:

	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, Purchase Address, Month]

Index: []

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

	Order ID	Product	Quantity Ordered	Price Each	Month
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 [42]: #data['Net_price']=data['Quantity Ordered']*data['Price Each']
#further data cleaning required
```

```
In [43]: # 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	

Month

12	2956652956662956672956682956692956702956712956...
10	2593582593592593602593612593622593632593642593...
4	1765581765591765601765601765611765621765631765...
11	2787972787982787992788002788012788022788032788...
5	1769781775511777781777781790761790761791341800...
3	1505331505411508121517491520361532841535441538...
7	222910222911222912229132229142229152229162229...
6	2099212099222099232099242099252099262099272099...
8	2366702366712366722366732366742366752366762366...
2	1505021505031505041505051505061505071505081505...
9	2388342392852406362410542423432428652436672449...
1	2971502978172978172979692983442990492991253003...

	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 07:0312/12/19 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	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...	

	Purchase Address	Net_price	\
Month			

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
11	46 Park St, New York City, NY 10001962 Hickory...	3199603.20
5	589 Lake St, Portland, OR 97035615 Lincoln St,...	3152606.75
3	270 Dogwood St, San Francisco, CA 94016683 Ada...	2807100.38
7	389 South St, Atlanta, GA 30301590 4th St, Sea...	2647775.76
6	950 Walnut St, Portland, ME 0410180 4th St, Sa...	2577802.26
8	359 Spruce St, Seattle, WA 98101492 Ridge St, ...	2244467.88
2	866 Spruce St, Portland, ME 0410118 13th St, S...	2202022.42
9	761 Forest St, San Francisco, CA 94016373 1st ...	2097560.13
1	427 Wilson St, Dallas, TX 75001519 13th St, Ne...	1822256.73

```
In [16]: monthly_sale=data.groupby('Month').sum()['Net_price']
```

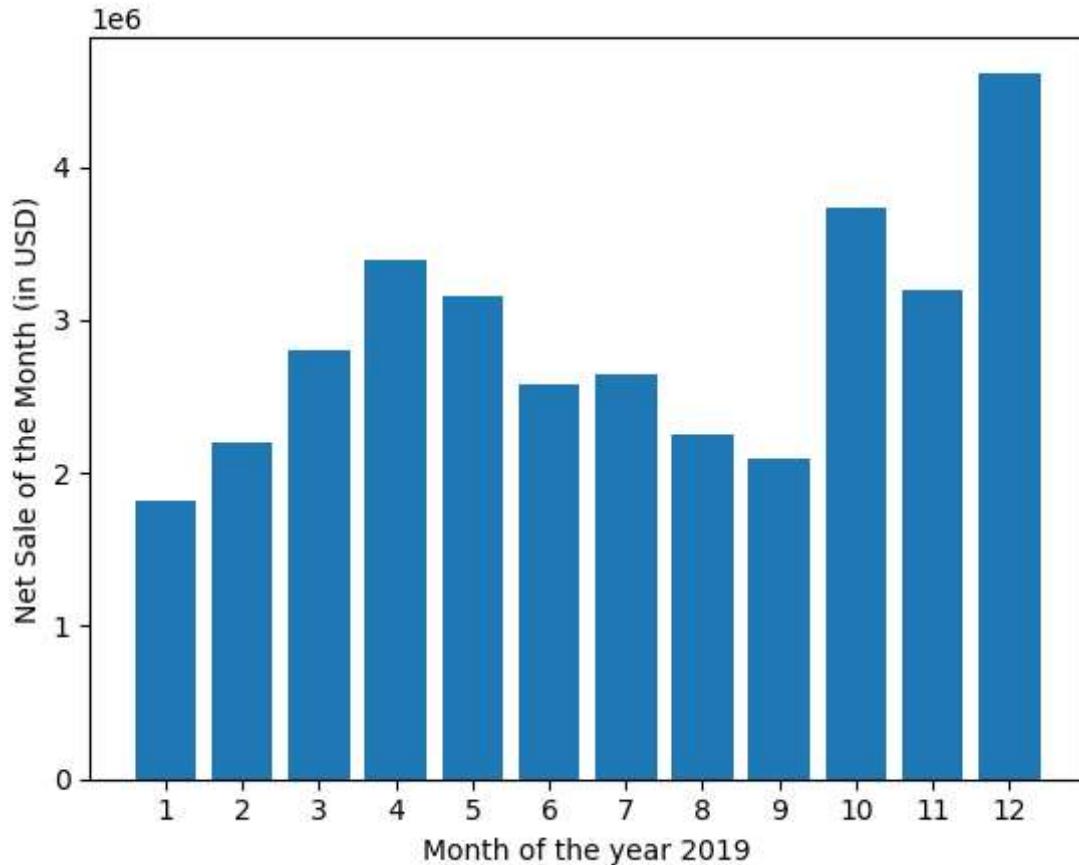
```
# months=arange(1,13)
months=range(1,13)
print(months)

plt.xlabel('Month of the year 2019')
plt.ylabel('Net Sale of the Month (in USD)')
plt.xticks(months)

plt.bar(months,monthly_sale)

plt.show()
```

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



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	
...
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	

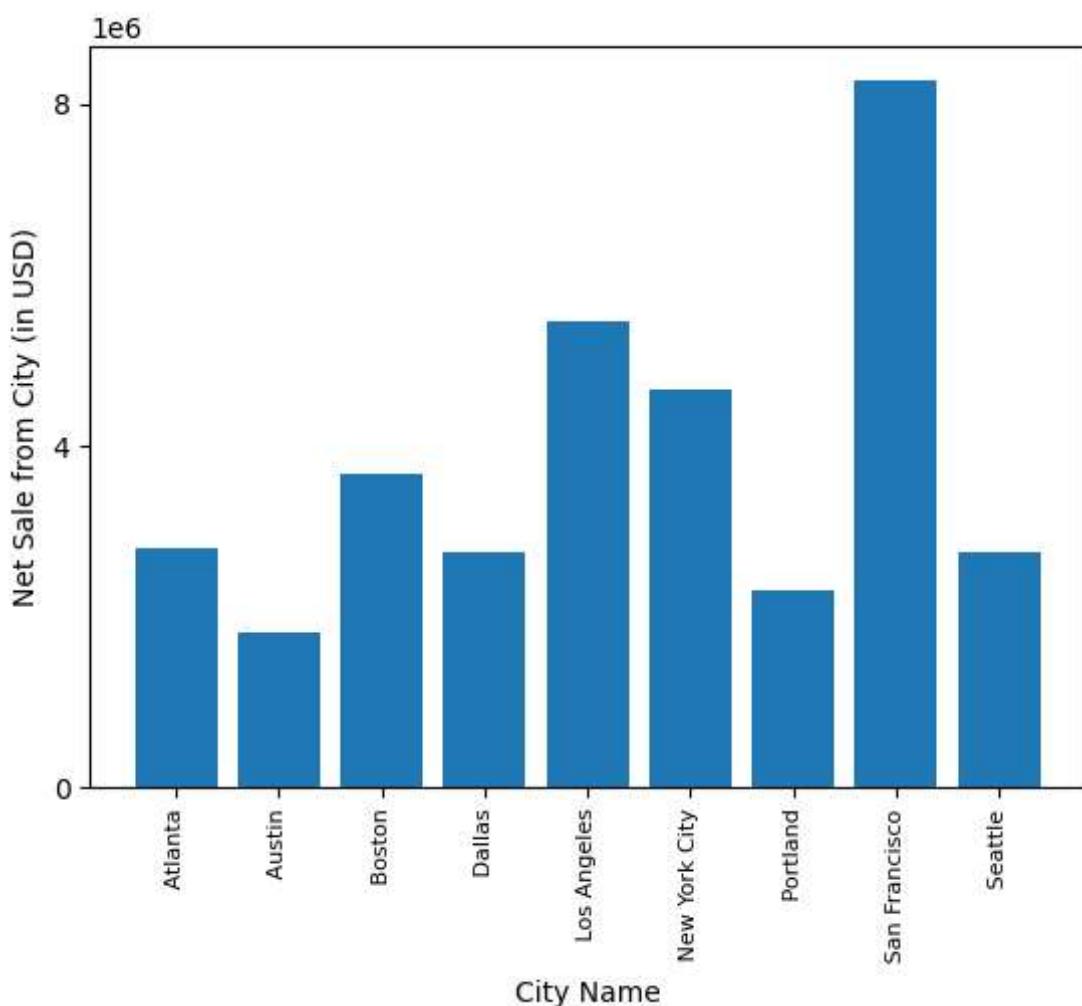
	Net_price	City
0	23.90	Dallas
2	99.99	Boston
3	600.00	Los Angeles
4	11.99	Los Angeles
5	11.99	Los Angeles
...
186845	8.97	Los Angeles
186846	700.00	San Francisco
186847	700.00	San Francisco
186848	379.99	San Francisco
186849	11.95	San Francisco

[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 chance
#or
# city=[city for city in data.groupby('City')] #no
```



Out[18]:

City	Order ID	Product	Quantity
Atlanta	1765641765711765821765891766231766271766341766...	USB-C Charging CableLightning Charging CableBo...	1
Austin	1765911766091766741766771766861767011767051767...	Apple Airpods HeadphonesApple Airpods Headphon...	1
Boston	1765591765661765781765811765851765851766001766...	Bose SoundSport HeadphonesWired HeadphonesAppl...	1
Dallas	1765581765691765701765771765961766021766101766...	USB-C Charging Cable27in 4K Gaming MonitorAA B...	1
Los Angeles	1765601765601765611765671765741765741765761765...	Google PhoneWired HeadphonesWired HeadphonesGo...	1
New York City	1765721765751765791765901765991766131766371766...	Apple Airpods HeadphonesAAA Batteries (4-pack)...	1
Portland	1765831765931766171766421766461766481766621767...	AAA Batteries (4-pack)Lightning Charging Cable...	1

Order ID	Product	Quantity	Order Date
City			
San Francisco	1765621765651765731765841765861765861765941765...	USB-C Charging Cable	Macbook Pro Laptop
Seattle	1765631765681765881766121766241766541766631766...	Bose SoundSport Headphones	Lightning Charging C...

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

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

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

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

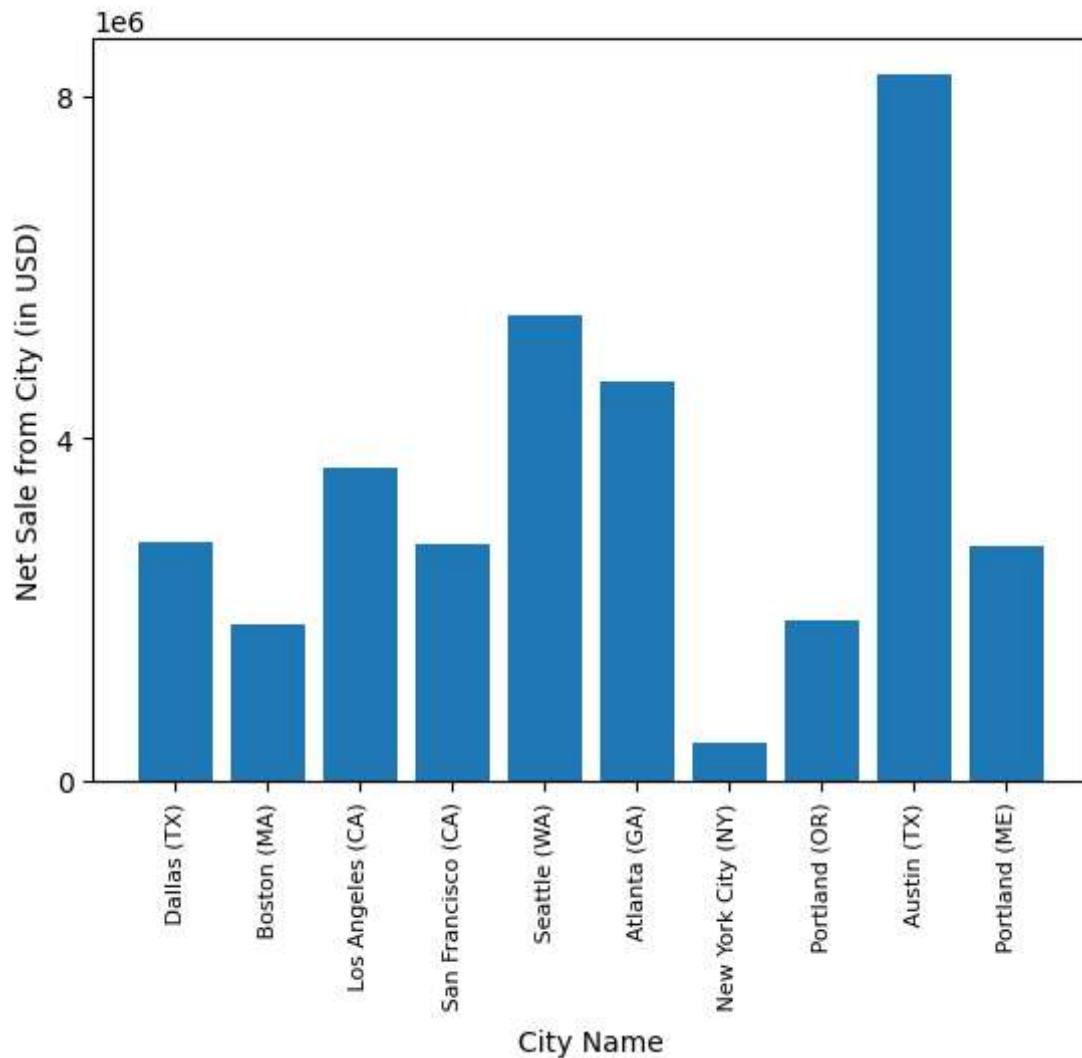
	Order ID	Product	Quantity Ordered	Price Each	\
0	176558	USB-C Charging Cable	2	11.95	
2	176559	Bose SoundSport Headphones	1	99.99	
3	176560	Google Phone	1	600.00	
4	176560	Wired Headphones	1	11.99	
5	176561	Wired Headphones	1	11.99	
...
186845	259353	AAA Batteries (4-pack)	3	2.99	
186846	259354	iPhone	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	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	
186848	379.99	San Francisco	San Francisco (CA)	94016	
186849	11.95	San Francisco	San Francisco (CA)	94016	

[185950 rows x 11 columns]

Out[20]:

CitywithState	Order ID	Product
Atlanta (GA)	1765641765711765821765891766231766271766341766...	USB-C Charging CableLightning Charging CableBo...
Austin (TX)	1765911766091766741766771766861767011767051767...	Apple Airpods HeadphonesApple Airpods Headphon...
Boston (MA)	1765591765661765781765811765851765851766001766...	Bose SoundSport HeadphonesWired HeadphonesAppl...
Dallas (TX)	1765581765691765701765771765961766021766101766...	USB-C Charging Cable27in 4K Gaming MonitorAA B...
Los Angeles (CA)	1765601765601765611765671765741765741765761765...	Google PhoneWired HeadphonesWired HeadphonesGo...
New York City (NY)	1765721765751765791765901765991766131766371766...	Apple Airpods HeadphonesAAA Batteries (4-pack)...
Portland (ME)	1767731768791769091770721772301772331772991773...	AAA Batteries (4-pack)Flatscreen TVWired Headp...

		Order ID	Product
CitywithState			
Portland (OR)	1765831765931766171766421766461766481766621767...		AAA Batteries (4-pack) Lightning Charging Cable...
San Francisco (CA)	1765621765651765731765841765861765861765941765...		USB-C Charging Cable Macbook Pro Laptop USB-C Ch...
Seattle (WA)	1765631765681765881766121766241766541766631766...	Bose SoundSport Headphones Lightning Charging C...	
In [21]:			
city_sale = data.groupby('CitywithState').sum()['Net_price']			
city = data['CitywithState'].unique() #with			
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
Atlanta (GA)	1765641765711765821765891766231766271766341766...	USB-C Charging CableLightning Charging CableBo...
Austin (TX)	1765911766091766741766771766861767011767051767...	Apple Airpods HeadphonesApple Airpods Headphon...
Boston (MA)	1765591765661765781765811765851765851766001766...	Bose SoundSport HeadphonesWired HeadphonesAppl...
Dallas (TX)	1765581765691765701765771765961766021766101766...	USB-C Charging Cable27in 4K Gaming MonitorAA B...
Los Angeles (CA)	1765601765601765611765671765741765741765761765...	Google PhoneWired HeadphonesWired HeadphonesGo...
New York City (NY)	1765721765751765791765901765991766131766371766...	Apple Airpods HeadphonesAAA Batteries (4-pack)...
Portland (ME)	1767731768791769091770721772301772331772991773...	AAA Batteries (4-pack)Flatscreen TVWired Headp...

		Order ID	Product
CitywithState			
Portland (OR)	1765831765931766171766421766461766481766621767...		AAA Batteries (4-pack) Lightning Charging Cable...
San Francisco (CA)	1765621765651765731765841765861765861765941765...		USB-C Charging Cable Macbook Pro Laptop USB-C Ch...
Seattle (WA)	1765631765681765881766121766241766541766631766...	Headphones	Bose SoundSport Lightning Charging C...

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'], ascending=False)

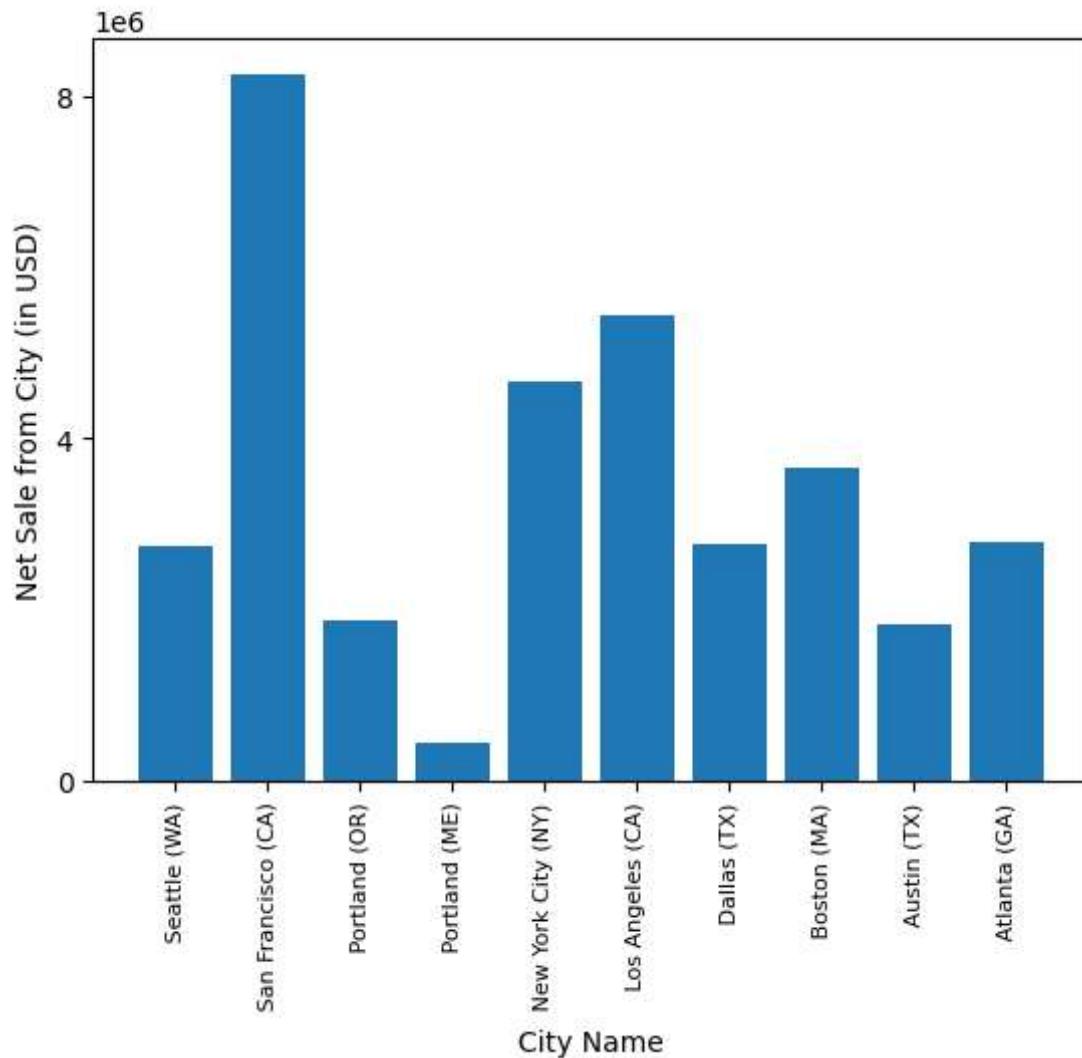
# city=data['CitywithState'].sort_values(by=['CitywithState'], ascending=False).unique()
city=data['CitywithState'].sort_values(ascending=False).unique()

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]:

CitywithState	Order ID	Product
Atlanta (GA)	1765641765711765821765891766231766271766341766...	USB-C Charging CableLightning Charging CableBo...
Austin (TX)	1765911766091766741766771766861767011767051767...	Apple Airpods HeadphonesApple Airpods Headphon...
Boston (MA)	1765591765661765781765811765851765851766001766...	Bose SoundSport HeadphonesWired HeadphonesAppl...
Dallas (TX)	1765581765691765701765771765961766021766101766...	USB-C Charging Cable27in 4K Gaming MonitorAA B...
Los Angeles (CA)	1765601765601765611765671765741765741765761765...	Google PhoneWired HeadphonesWired HeadphonesGo...
New York City (NY)	1765721765751765791765901765991766131766371766...	Apple Airpods HeadphonesAAA Batteries (4-pack)...
Portland (ME)	1767731768791769091770721772301772331772991773...	AAA Batteries (4-pack)Flatscreen TVWired Headp...

		Order ID	Product
CitywithState			
Portland (OR)	1765831765931766171766421766461766481766621767...		AAA Batteries (4-pack) Lightning Charging Cable...
San Francisco (CA)	1765621765651765731765841765861765861765941765...		USB-C Charging Cable Macbook Pro Laptop USB-C Ch...
Seattle (WA)	1765631765681765881766121766241766541766631766...	Bose SoundSport Headphones	Lightning Charging C...

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

185950 rows × 11 columns

```
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()
```

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

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

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

Order Time	Order ID \
19	1765691765741765741765771765871766121766251766...
12	1765681765821765831765891766071766081766621766...
11	1765761765851765851765901766291766631766721766...
20	1765721765841765961765981766161766221766361766...
18	1765731766131766191766311766421766451766641766...
13	1765621765921765931765971766111766511766831767...
17	1765671765861765861766601766891766891767391767...
14	1765601765601765661765711766061766271766321766...
21	1765811766141766181766261766411766431766701767...
10	1765641765651765791766041766151766201766241766...
15	1765701765991766001766861767401767471767481767...
16	1766051766091766211766391766501766561766661766...
09	1765611765951766021766351766751766931767051767...
22	1765591766011766401766781766851766971767341768...
08	1765581766171766341766381766491766581766731767...
23	1765781765941766231767001767081767141767351767...
07	1765631765911766031766461767281767491767811767...
00	1765751765801766101766671766761766841767681768...
01	1766481767371768091769071769911770371771711771...
06	1766821769001769431769441770421770491770681770...
02	1766441768181769451770821772141772541772541774...
05	1766301767121768191768731769761770751771021771...
04	1765881766281767881776871779061780941783771786...
03	1769781775011778311780771782881787171794551794...

Order Time	Product \
19	27in 4K Gaming MonitorGoogle PhoneUSB-C Chargi...
12	Lightning Charging CableBose SoundSport Headph...
11	Apple Airpods HeadphonesBose SoundSport Headph...
20	Apple Airpods HeadphonesFlatscreen TVApple Air...
18	USB-C Charging CableUSB-C Charging CableFlatsc...
13	USB-C Charging CableUSB-C Charging CableLightn...
17	Google PhoneAAA Batteries (4-pack)Google Phone...
14	Google PhoneWired HeadphonesWired HeadphonesLi...
21	iPhoneAA Batteries (4-pack)Lightning Charging ...
10	USB-C Charging CableMacbook Pro LaptopAA Batte...
15	AA Batteries (4-pack)Lightning Charging Cable2...
16	USB-C Charging CableApple Airpods HeadphonesLi...
09	Wired HeadphonesWired HeadphonesLightning Char...
22	Bose SoundSport HeadphonesUSB-C Charging Cable...
08	USB-C Charging CableApple Airpods HeadphonesLi...
23	Apple Airpods HeadphonesWired Headphones27in F...
07	Bose SoundSport HeadphonesApple Airpods Headph...
00	AAA Batteries (4-pack)USB-C Charging CableAAA ...
01	27in FHD MonitorApple Airpods HeadphonesBose S...
06	Lightning Charging CableBose SoundSport Headph...
02	USB-C Charging CableWired HeadphonesUSB-C Char...
05	Google PhoneGoogle PhoneAA Batteries (4-pack)A...
04	20in MonitorLightning Charging CableBose Sound...
03	Apple Airpods HeadphonesWired HeadphonesLightn...

Order Time	Quantity Ordered	Price Each \
19	14470	2398588.31
12	14202	2299876.68
11	14005	2288855.18
20	13768	2268185.16

18	13802	2207696.93
13	13685	2139743.86
17	12229	2116777.02
14	12362	2072194.77
21	12244	2030763.83
10	12308	1932665.62
15	11391	1931174.99
16	11662	1892454.54
09	9816	1628498.49
22	9899	1599464.44
08	7002	1185970.62
23	7065	1172625.87
07	4556	740568.11
00	4428	709296.70
01	2619	458490.00
06	2810	445000.11
02	1398	233833.64
05	1493	229621.21
04	937	162058.18
03	928	144726.42

Order Time	Order Date \
19	04/16/19 19:23 04/03/19 19:42 04/03/19 19:42 04/0...
12	04/15/19 12:18 04/27/19 12:20 04/20/19 12:00 04/0...
11	04/28/19 11:42 04/07/19 11:31 04/07/19 11:31 04/1...
20	04/04/19 20:30 04/24/19 20:39 04/27/19 20:11 04/2...
18	04/27/19 18:41 04/26/19 18:58 04/16/19 18:37 04/1...
13	04/29/19 13:03 04/27/19 13:04 04/15/19 13:45 04/1...
17	04/18/19 17:18 04/10/19 17:00 04/10/19 17:00 04/1...
14	04/12/19 14:38 04/12/19 14:38 04/08/19 14:05 04/1...
21	04/09/19 21:38 04/07/19 21:01 04/18/19 21:44 04/0...
10	04/12/19 10:58 04/24/19 10:38 04/11/19 10:23 04/0...
15	04/22/19 15:09 04/11/19 15:25 04/30/19 15:54 04/1...
16	04/15/19 16:29 04/11/19 16:59 04/11/19 16:09 04/2...
09	04/30/19 09:27 04/02/19 09:11 04/30/19 09:39 04/2...
22	04/07/19 22:30 04/02/19 22:32 04/27/19 22:54 04/1...
08	04/19/19 08:46 04/25/19 08:03 04/10/19 08:00 04/2...
23	04/09/19 23:35 04/17/19 23:04 04/20/19 23:51 04/0...
07	04/02/19 07:46 04/21/19 07:21 04/01/19 07:46 04/2...
00	04/27/19 00:30 04/05/19 00:35 04/26/19 00:40 04/1...
01	04/24/19 01:17 04/07/19 01:28 04/19/19 01:14 04/1...
06	04/27/19 06:19 04/21/19 06:17 04/09/19 06:50 04/2...
02	04/23/19 02:09 04/09/19 02:19 04/24/19 02:17 04/1...
05	04/10/19 05:01 04/20/19 05:21 04/15/19 05:44 04/0...
04	04/02/19 04:00 04/15/19 04:03 04/15/19 04:39 04/0...
03	05/01/19 03:29 04/10/19 03:43 04/01/19 03:09 04/0...

Order Time	Purchase Address \ Month
19	657 Hill St, Dallas, TX 75001 20 Hill St, Los A... 91389
12	438 Elm St, Seattle, WA 98101 178 Lincoln St, A... 89161
11	771 Ridge St, Los Angeles, CA 90001 823 Highland... 87654
20	149 Dogwood St, New York City, NY 10001 936 Church... 86375
18	214 Chestnut St, San Francisco, CA 94016 405 4th ... 86421
13	381 Wilson St, San Francisco, CA 94016 352 4th ... 85808
17	444 7th St, Los Angeles, CA 90001 365 Center St... 77454
14	669 Spruce St, Los Angeles, CA 90001 669 Spruce... 77836
21	84 Jackson St, Boston, MA 02215 647 Jackson St,... 77103
10	790 Ridge St, Atlanta, GA 30301 915 Willow St, ... 76928

15	186 12th St, Dallas, TX 75001279	Maple St, New... 72060
16	717 Pine St, San Francisco, CA 94016267	11th S... 72939
09	333 8th St, Los Angeles, CA 90001383	6th St, L... 60981
22	682 Chestnut St, Boston, MA 02215754	10th St, ... 62088
08	917 1st St, Dallas, TX 75001319	8th St, Portla... 43626
23	513 Church St, Boston, MA 0221563	Maple St, Sa... 44364
07	668 Center St, Seattle, WA 98101600	Maple St, ... 28850
00	433 Hill St, New York City, NY 10001886	Willow... 27554
01	732 2nd St, Portland, OR 97035109	Center St, L... 16657
06	593 Adams St, Boston, MA 02215356	North St, Bo... 17539
02	962 7th St, Los Angeles, CA 90001642	13th St, ... 8507
05	770 Maple St, San Francisco, CA 94016910	Washi... 9301
04	765 Cherry St, Seattle, WA 98101628	Wilson St,... 6148
03	589 Lake St, Portland, OR 97035888	14th St, Se... 5904

Order Time	Net_price	City \
19	2412938.54	Dallas Los Angeles Los Angeles Dallas Los Ang...
12	2316821.34	Seattle Atlanta Portland Atlanta Boston San F...
11	2300610.24	Los Angeles Boston Boston New York City San F...
20	2281716.24	New York City San Francisco Dallas San Franci...
18	2219348.30	San Francisco New York City Los Angeles San F...
13	2155389.80	San Francisco Los Angeles Portland San Franci...
17	2129361.61	Los Angeles San Francisco San Francisco Bosto...
14	2083672.73	Los Angeles Los Angeles Boston Atlanta San Fr...
21	2042000.86	Boston San Francisco San Francisco Boston Bos...
10	1944286.77	Atlanta San Francisco New York City San Franc...
15	1941549.60	Dallas New York City Boston Austin Dallas San...
16	1904601.31	San Francisco Austin Boston San Francisco Bos...
09	1639030.58	Los Angeles Los Angeles Dallas San Francisco ...
22	1607549.21	Boston Los Angeles San Francisco San Francisc...
08	1192348.97	Dallas Portland Atlanta Boston San Francisco ...
23	1179304.44	Boston San Francisco Atlanta Dallas New York ...
07	744854.12	Seattle Austin San Francisco Portland Atlanta...
00	713721.27	New York City Los Angeles Dallas Los Angeles ...
01	460866.88	Portland Los Angeles San Francisco Los Angele...
06	448113.00	Boston Boston San Francisco Austin San Franci...
02	234851.44	Los Angeles San Francisco Austin Seattle Seat...
05	230679.82	San Francisco San Francisco Portland Los Ange...
04	162661.01	Seattle Dallas New York City San Francisco Ne...
03	145757.89	Portland Seattle San Francisco San Francisco ...

Order Time	CitywithState \
19	Dallas (TX) Los Angeles (CA) Los Angeles (CA)...
12	Seattle (WA) Atlanta (GA) Portland (OR) Atlan...
11	Los Angeles (CA) Boston (MA) Boston (MA) New ...
20	New York City (NY) San Francisco (CA) Dallas ...
18	San Francisco (CA) New York City (NY) Los Ang...
13	San Francisco (CA) Los Angeles (CA) Portland ...
17	Los Angeles (CA) San Francisco (CA) San Franc...
14	Los Angeles (CA) Los Angeles (CA) Boston (MA)...
21	Boston (MA) San Francisco (CA) San Francisco ...
10	Atlanta (GA) San Francisco (CA) New York City...
15	Dallas (TX) New York City (NY) Boston (MA) Au...
16	San Francisco (CA) Austin (TX) Boston (MA) Sa...
09	Los Angeles (CA) Los Angeles (CA) Dallas (TX)...
22	Boston (MA) Los Angeles (CA) San Francisco (C...
08	Dallas (TX) Portland (OR) Atlanta (GA) Boston...
23	Boston (MA) San Francisco (CA) Atlanta (GA) D...

07	Seattle (WA)	Austin (TX)	San Francisco (CA)	P...
00	New York City (NY)	Los Angeles (CA)	Dallas (T...	
01	Portland (OR)	Los Angeles (CA)	San Francisco ...	
06	Boston (MA)	Boston (MA)	San Francisco (CA) Au...	
02	Los Angeles (CA)	San Francisco (CA)	Austin (T...	
05	San Francisco (CA)	San Francisco (CA)	Portlan...	
04	Seattle (WA)	Dallas (TX)	New York City (NY) S...	
03	Portland (OR)	Seattle (WA)	San Francisco (CA)...	

City Pin					
Order Time	Order ID	Product	Quantity Ordered	Price Each	Order Date \
19	7500190001900017500190001981019401690001940169...				
12	9810130301970353030102215940169703575001100019...				
11	9000102215022151000194016981011000110001733010...				
20	1000194016750019401694016940160221510001303017...				
18	9401610001900019401697035750019000194016940169...				
13	9401690001970359401694016022159000173301100010...				
17	9000194016940160221510001100017330173301900011...				
14	9000190001022153030194016303019401690001750017...				
21	02215940169401602215022159401694016970359...				
10	3030194016100019401675001900019810110001900019...				
15	7500110001022157330175001940160221590001900019...				
16	9401673301022159401602215100019810110001940169...				
09	9000190001750019401698101940167330198101750019...				
22	0221590001940169401610001100011000194016981019...				
08	7500197035303010221594016940160221510001900019...				
23	0221594016303017500110001900017330173301940163...				
07	9810173301940169703530301981017500175001940160...				
00	1000190001750019000190001900019000194016940161...				
01	9703590001940169000194016940169401673301022159...				
06	0221502215940167330194016940169000102215940169...				
02	9000194016733019810198101022150221573301970359...				
05	9401694016970359000190001940161000110001940169...				
04	9810175001100019401610001750019401694016750017...				
03	9703598101940169401675001940167500130301750019...				
Order Time	Order ID	Product	Quantity Ordered	Price Each	Order Date \
00	3910	3910	3910	3910	3910
01	2350	2350	2350	2350	2350
02	1243	1243	1243	1243	1243
03	831	831	831	831	831
04	854	854	854	854	854
05	1321	1321	1321	1321	1321
06	2482	2482	2482	2482	2482
07	4011	4011	4011	4011	4011
08	6256	6256	6256	6256	6256
09	8748	8748	8748	8748	8748
10	10944	10944	10944	10944	10944
11	12411	12411	12411	12411	12411
12	12587	12587	12587	12587	12587
13	12129	12129	12129	12129	12129
14	10984	10984	10984	10984	10984
15	10175	10175	10175	10175	10175
16	10384	10384	10384	10384	10384
17	10899	10899	10899	10899	10899
18	12280	12280	12280	12280	12280
19	12905	12905	12905	12905	12905
20	12228	12228	12228	12228	12228
21	10921	10921	10921	10921	10921
22	8822	8822	8822	8822	8822

	23	6275	6275	6275	6275	6275	6275	
Order	Time	Purchase	Address	Month	Net_price	City	CitywithState	City Pin
00			3910	3910	3910	3910	3910	3910
01			2350	2350	2350	2350	2350	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	10944
11			12411	12411	12411	12411	12411	12411
12			12587	12587	12587	12587	12587	12587
13			12129	12129	12129	12129	12129	12129
14			10984	10984	10984	10984	10984	10984
15			10175	10175	10175	10175	10175	10175
16			10384	10384	10384	10384	10384	10384
17			10899	10899	10899	10899	10899	10899
18			12280	12280	12280	12280	12280	12280
19			12905	12905	12905	12905	12905	12905
20			12228	12228	12228	12228	12228	12228
21			10921	10921	10921	10921	10921	10921
22			8822	8822	8822	8822	8822	8822
23			6275	6275	6275	6275	6275	6275

```
In [26]: hours=[hour for hour,df in data.groupby('Order Time')]

order_count_hour=data.groupby('Order Time').count()

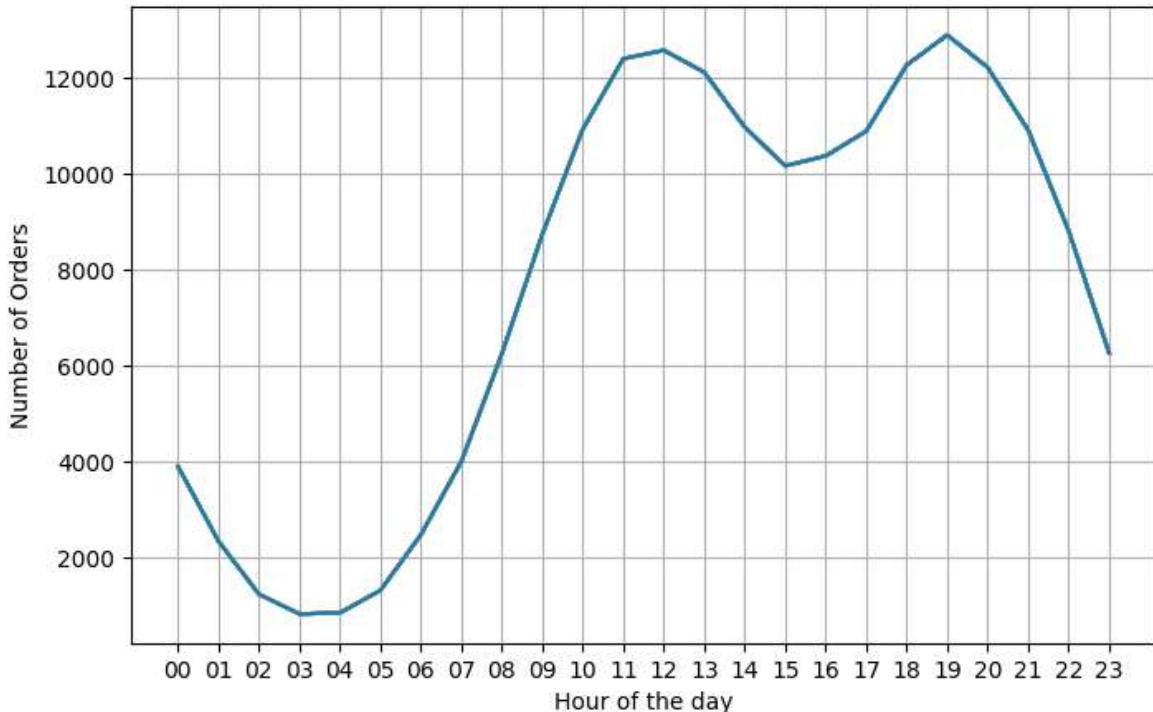
plt.figure(figsize=(8,5),dpi=100)

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 Date',
       'Net price']]
data2

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

Grouped
0 USB-C Charging Cable
2 Bose SoundSport Headphones
3 Google Phone,Wired Headphones
4 Google Phone,Wired Headphones
5 Wired Headphones

```
C:\Users\Vatsal_Fast\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
    data2['Grouped']=data2.groupby('Order ID')['Product'].transform(lambda x: ','.join(x))
```

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

Out[28]:

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

385 rows × 8 columns

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 Date', '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 have
#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
3	176560	Google Phone	1	600.00	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001	4	600.00 A
4	176560	Wired Headphones	1	11.99	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001	4	11.99 A
18	176574	Google Phone	1	600.00	04/03/19 19:42	20 Hill St, Los Angeles, CA 90001	4	600.00 A
19	176574	USB-C Charging Cable	1	11.95	04/03/19 19:42	20 Hill St, Los Angeles, CA 90001	4	11.95 A
30	176585	Bose SoundSport Headphones	1	99.99	04/07/19 11:31	823 Highland St, Boston, MA 02215	4	99.99 I
31	176585	Bose SoundSport Headphones	1	99.99	04/07/19 11:31	823 Highland St, Boston, MA 02215	4	99.99 I
32	176586	AAA Batteries (4-pack)	2	2.99	04/10/19 17:00	365 Center St, San Francisco, CA 94016	4	5.98 Fra
33	176586	Google Phone	1	600.00	04/10/19 17:00	365 Center St, San Francisco, CA 94016	4	600.00 Fra
119	176672	Lightning Charging Cable	1	14.95	04/12/19 11:07	778 Maple St, New York City, NY 10001	4	14.95 Yo
120	176672	USB-C Charging Cable	1	11.95	04/12/19 11:07	778 Maple St, New York	4	11.95 Yo

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	Month	Net_price
						City, NY 10001		
129	176681	Apple Airpods Headphones	1	150.00	04/20/19 10:39	331 Cherry St, Seattle, WA 98101	4	150.00
130	176681	ThinkPad Laptop	1	999.99	04/20/19 10:39	331 Cherry St, Seattle, WA 98101	4	999.99
138	176689	Bose SoundSport Headphones	1	99.99	04/24/19 17:15	659 Lincoln St, New York City, NY 10001	4	99.99
139	176689	AAA Batteries (4-pack)	2	2.99	04/24/19 17:15	659 Lincoln St, New York City, NY 10001	4	5.98
189	176739	34in Ultrawide Monitor	1	379.99	04/05/19 17:38	730 6th St, Austin, TX 73301	4	379.99
190	176739	Google Phone	1	600.00	04/05/19 17:38	730 6th St, Austin, TX 73301	4	600.00
225	176774	Lightning Charging Cable	1	14.95	04/25/19 15:06	372 Church St, Los Angeles, CA 90001	4	14.95
226	176774	USB-C Charging Cable	1	11.95	04/25/19 15:06	372 Church St, Los Angeles, CA 90001	4	11.95
233	176781	iPhone	1	700.00	04/03/19 07:37	976 Hickory St, Dallas, TX 75001	4	700.00
234	176781	Lightning Charging Cable	1	14.95	04/03/19 07:37	976 Hickory St, Dallas, TX 75001	4	14.95

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	Month	Net_price	
250	176797	Google Phone	1	600.00	04/21/19 08:54	923 Elm St, Los Angeles, CA 90001	4	600.00	A
251	176797	Bose SoundSport Headphones	1	99.99	04/21/19 08:54	923 Elm St, Los Angeles, CA 90001	4	99.99	A
252	176797	Wired Headphones	1	11.99	04/21/19 08:54	923 Elm St, Los Angeles, CA 90001	4	11.99	A
260	176805	Google Phone	1	600.00	04/01/19 15:50	91 Lincoln St, Portland, OR 97035	4	600.00	Pc
261	176805	USB-C Charging Cable	1	11.95	04/01/19 15:50	91 Lincoln St, Portland, OR 97035	4	11.95	Pc
264	176808	Google Phone	1	600.00	04/28/19 18:03	933 Meadow St, San Francisco, CA 94016	4	600.00	Fra
265	176808	Wired Headphones	1	11.99	04/28/19 18:03	933 Meadow St, San Francisco, CA 94016	4	11.99	Fra
270	176813	Google Phone	1	600.00	04/28/19 18:01	269 Hill St, Atlanta, GA 30301	4	600.00	A
271	176813	Wired Headphones	1	11.99	04/28/19 18:01	269 Hill St, Atlanta, GA 30301	4	11.99	A
394	176935	AAA Batteries (4-pack)	1	2.99	04/03/19 21:31	315 1st St, Dallas, TX 75001	4	2.99	

```
In [31]: data3['Grouped']=data3.groupby('Order ID')[['Product']].transform(lambda x: ', '.join(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
19	Google Phone,USB-C Charging Cable
30	Bose SoundSport Headphones,Bose SoundSport He...

C:\Users\Vatsal_Fast\AppData\Local\Temp\ipykernel_21308\2648751239.py:1: 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

```
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]:

	Order ID	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

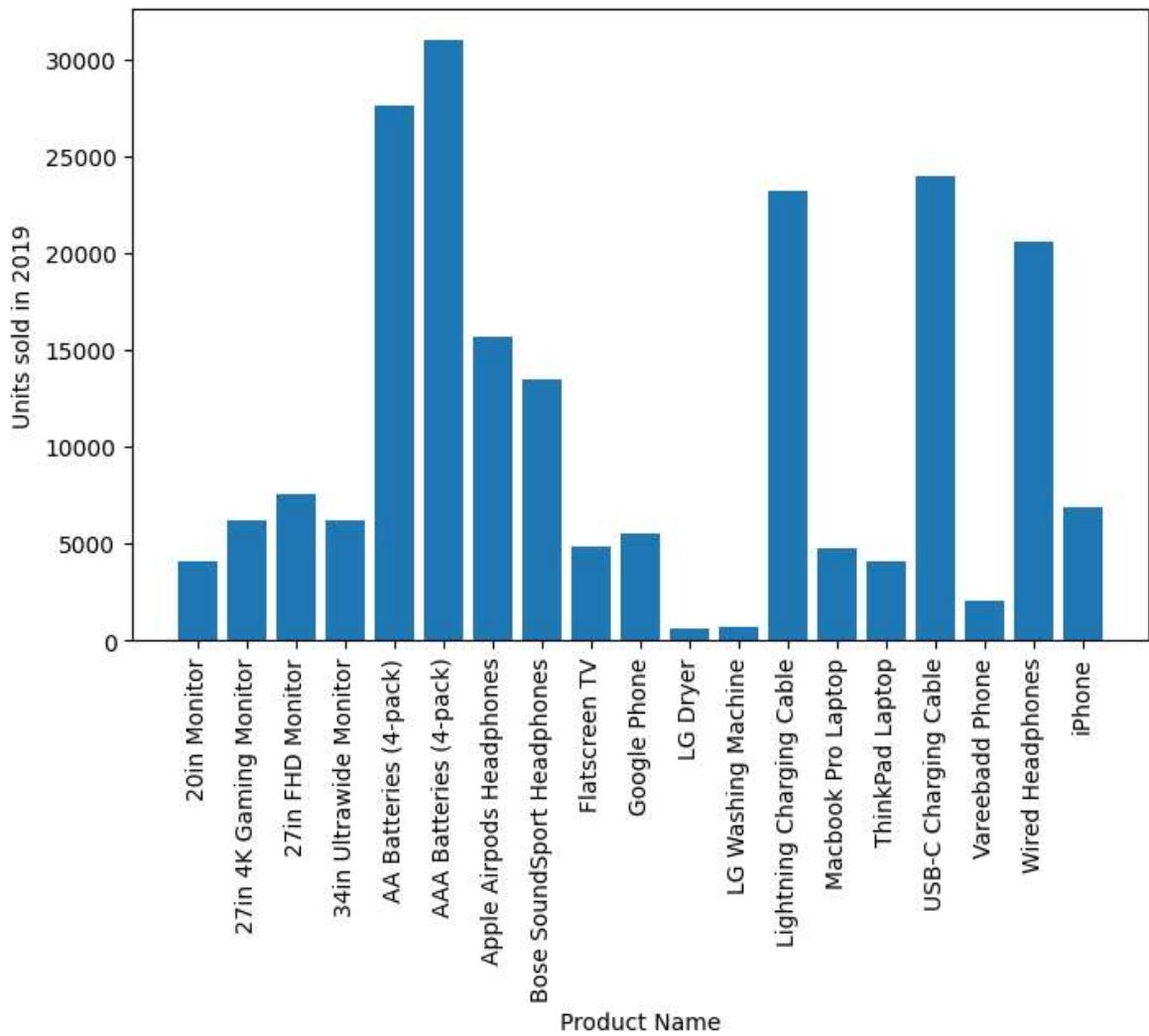
Finding the product Sold the most and why

```
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()
```



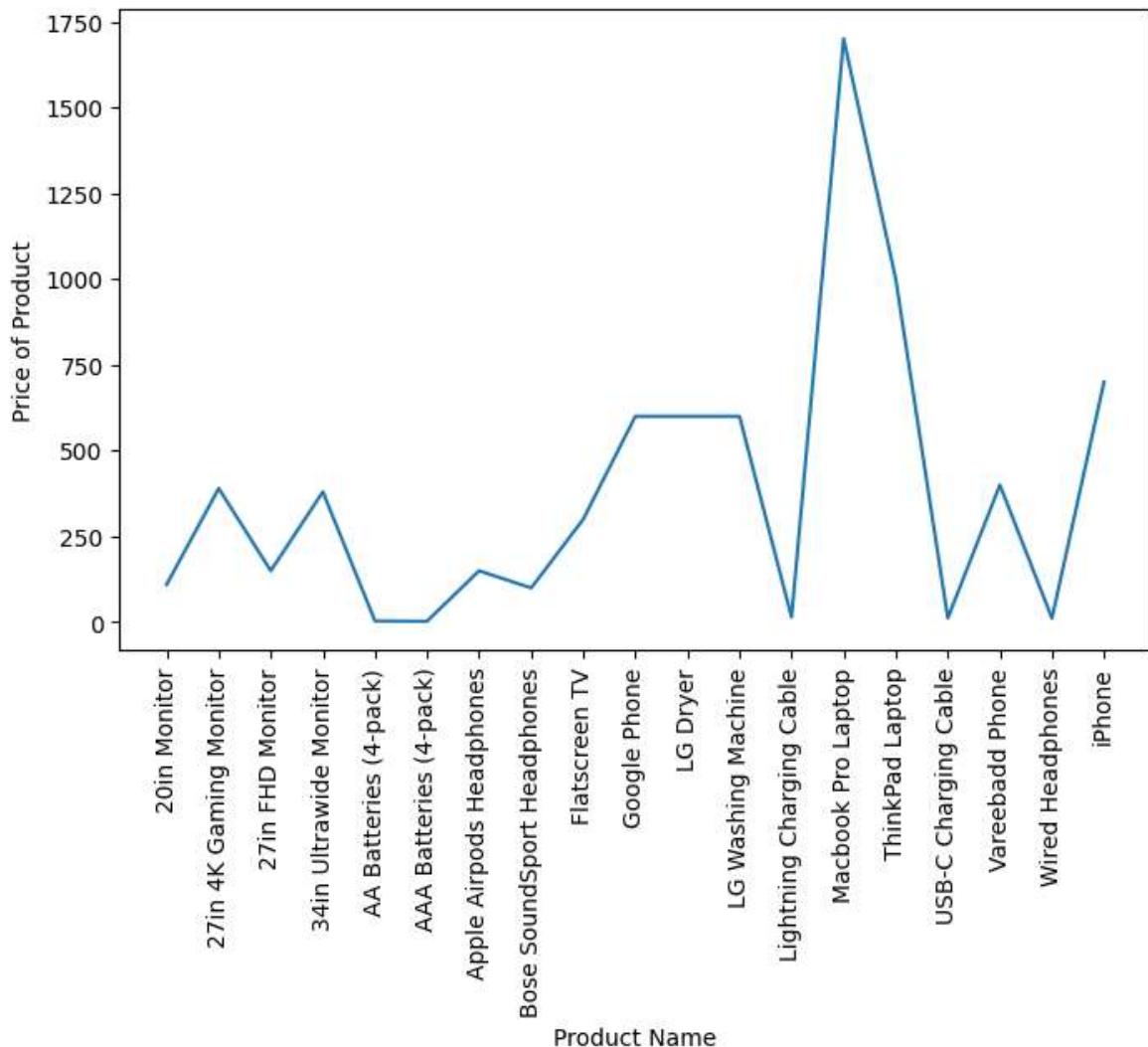
```
In [39]: # price_each=[price for price, df in data.groupby('Product').mean()]
# price_each=[price for price, df in data.groupby('Price Each').mean()]
#price_each=[price for price in data.groupby('Product').mean()['Price Each']]
price_each = [df['Price Each'].mean() for _, df in data.groupby('Product')]

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 in products]

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

fig,ax1=plt.subplots()

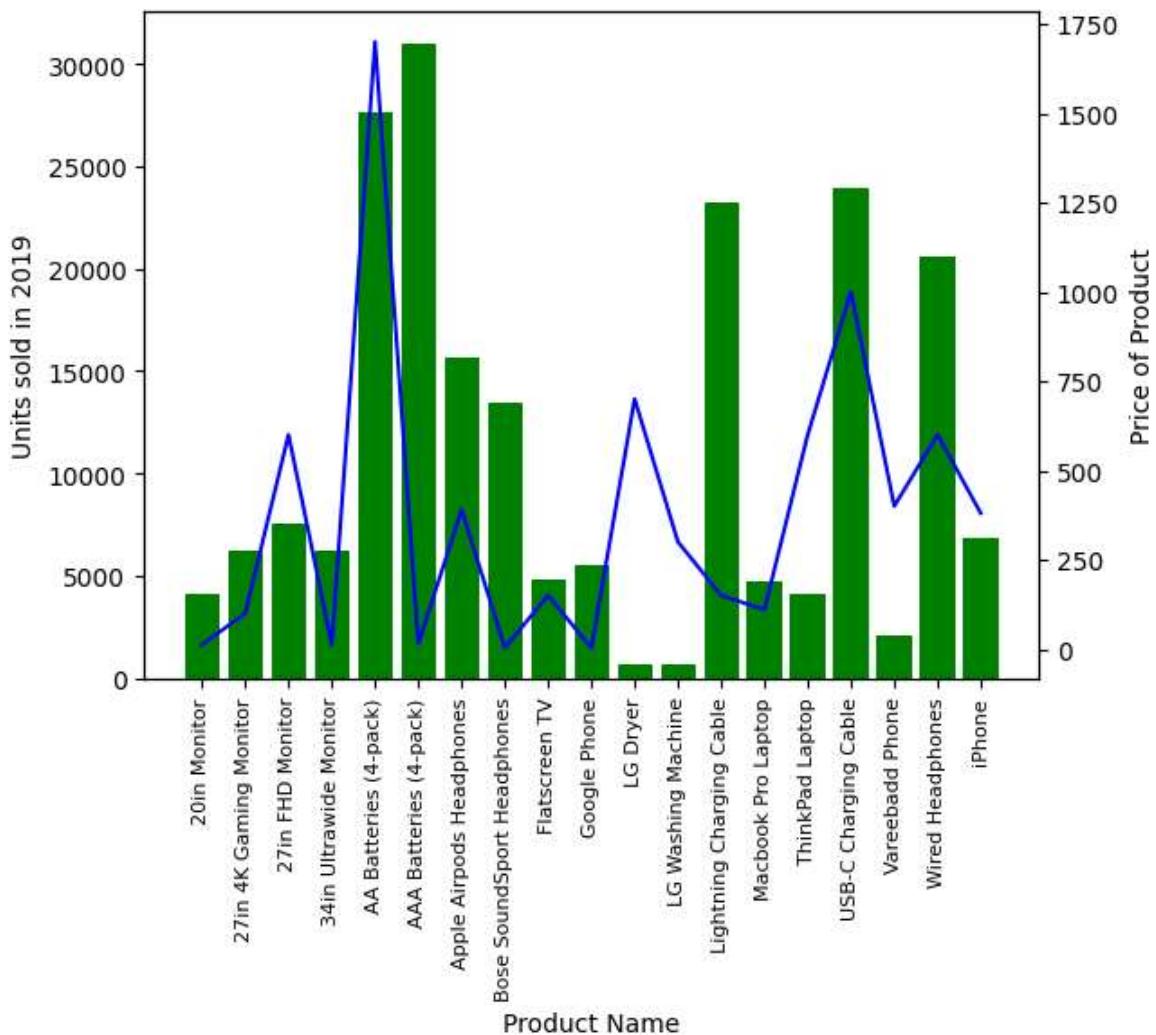
ax2=ax1.twinx()

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: UserWarning: FixedFormatter should only be used together with FixedLocator
    ax1.set_xticklabels(products, rotation='vertical', size=8)
```



In []:

In []: