Pandas Sales Analysis

November 5, 2022

0.1 Pandas Sales Analysis.

0.1.1 Import required libraries.

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

0.1.2 Merge 12 months of data into a single file.

My first task is to merge all 12 files corresponding to a different month into a single CSV for easy cleaning. I used a simple list comprehension here as well as imported the os library.

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

files = [file for file in os.listdir('SalesAnalysis/Sales_Data')]

for file in files:
    df = pd.read_csv('SalesAnalysis/Sales_Data/'+file)
    all_months_data = pd.concat([all_months_data, df])

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

0.1.3 Read in updated dataframe

Here, I read the CSV file and displayed the head so we can find out what kind of structure the table has.

[3]:
```
all_data = pd.read_csv('all_data.csv')
all_data.head()
```

[3]:

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address
0	176558	USB-C Charging Cable	2	11.95	04/19/19 08:46	917 1st St, Dallas, TX 75001
1	NaN	NaN	NaN	NaN		
2	176559	Bose SoundSport Headphones	1	99.99		
3	176560	Google Phone	1	600		
4	176560	Wired Headphones	1	11.99		

```
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
```

There are NANs in our data so we should address them.

0.1.4 Drop NANs

We can see there are multiple NANs so I decided to drop them so that we can begin some analysis. I decided to use all since this would drop the rows where ALL the data is missing.

[4]:
```
nan_df = all_data[all_data.isna().any(axis=1)]
nan_df.head()
```

[4]:
```
     Order ID Product Quantity Ordered Price Each Order Date Purchase Address
1         NaN     NaN               NaN        NaN        NaN              NaN
356       NaN     NaN               NaN        NaN        NaN              NaN
735       NaN     NaN               NaN        NaN        NaN              NaN
1433      NaN     NaN               NaN        NaN        NaN              NaN
1553      NaN     NaN               NaN        NaN        NaN              NaN
```

[5]:
```
all_data = all_data.dropna(how='all')
all_data.head()
```

[5]:
```
  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
0  04/19/19 08:46          917 1st St, Dallas, TX 75001
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
5  04/30/19 09:27     333 8th St, Los Angeles, CA 90001
```

0.1.5 Add month column

There was initially some trouble here since the column titles were being repeated so I deicded to redefine the dataframe as a dataframe that includes all the data where the first two string indexes of Order Data is not equal to 'Or'. This might not seem intuitive here but since the column titles were being duplicated as values (from the initial merge at the start is what I suspect). This would indeed take out all of those data points.

[6]:
```
all_data = all_data[all_data['Order Date'].str[0:2] != 'Or']
all_data.head()
```

[6]:
```
  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
0  04/19/19 08:46          917 1st St, Dallas, TX 75001
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
5  04/30/19 09:27     333 8th St, Los Angeles, CA 90001
```

I wanted to add a month column so that when I do add a graph or visualisation I can do it on a per month basis. Here I filtered the first two string characters of Order Data.

[7]:
```
all_data['Month'] = all_data['Order Date'].str[0:2]
all_data['Month'] = all_data['Month'].astype('int32')
all_data.head()
```

[7]:
```
  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
```

0.1.6 What is the best month for sales?

0.1.7 Convert data type, create column and rearrange columns

Before we can add a price column we need to convert the data types to our desired type.

[8]:
```
all_data['Quantity Ordered'] = all_data['Quantity Ordered'].astype('int32')
all_data['Price Each'] = all_data['Price Each'].astype('float')
```

[9]:
```
all_data.head()
```

[9]:
```
  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
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
```

[10]:
```
all_data['Sales'] = all_data['Quantity Ordered'] * all_data['Price Each']
```

[11]:
```
all_data = all_data[['Order ID', 'Product', 'Quantity Ordered', 'Price Each',␣
↪'Sales', 'Order Date', 'Purchase Address', 'Month']]
all_data.head()
```

[11]:
```
  Order ID                     Product  Quantity Ordered  Price Each   Sales  \
0   176558        USB-C Charging Cable                 2       11.95   23.90
2   176559  Bose SoundSport Headphones                 1       99.99   99.99
3   176560                Google Phone                 1      600.00  600.00
4   176560            Wired Headphones                 1       11.99   11.99
5   176561            Wired Headphones                 1       11.99   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
```

Rearranging the columns was not necessary but I did it anyway.

0.1.8 Question 1: Best month for sales, how much was earned?

[12]:
```
result = all_data.groupby('Month').sum('numeric')
```

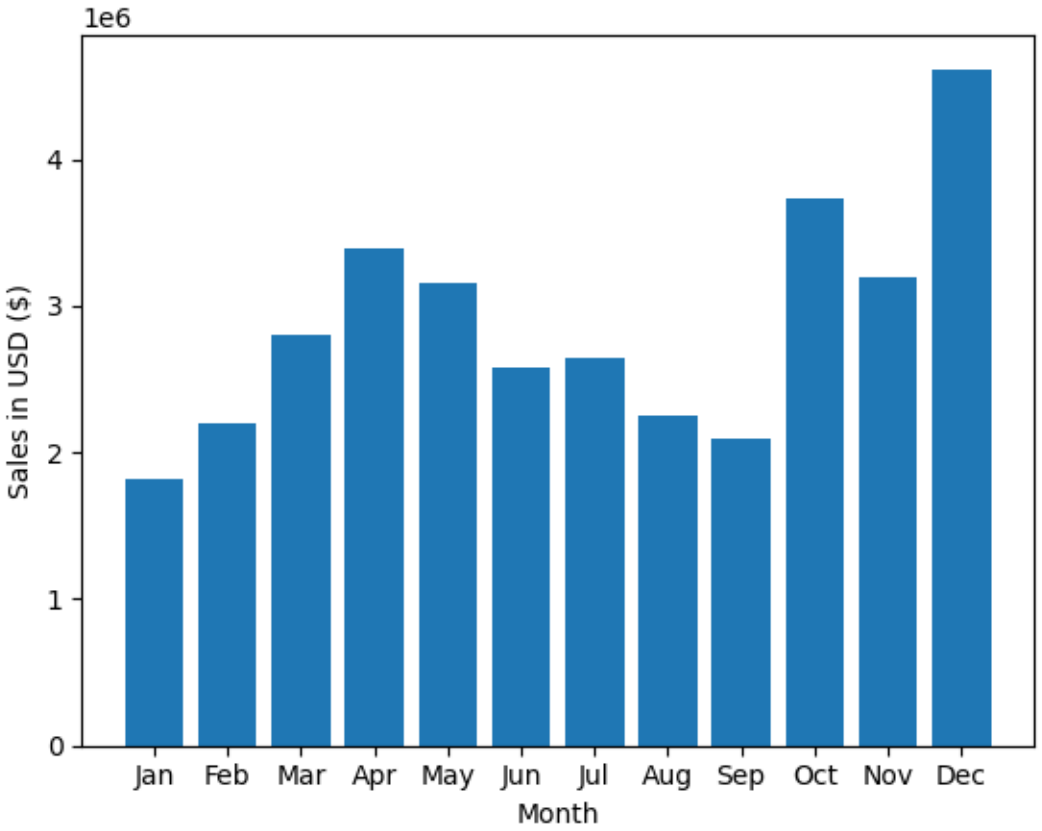
[13]:
```
import matplotlib.pyplot as plt
import datetime
```

This actually took a little bit of fiddling around but I managed to import the datetime function so that I could map numbers to string months.

[14]:
```
month_num = range(1, 13)
months = []

for i in month_num:
    i = str(i)
    datetime_object = datetime.datetime.strptime(i, "%m")
    month_name = datetime_object.strftime('%b')
    months.append(month_name)

plt.bar(months, result['Sales'])
plt.ylabel('Sales in USD ($)')
plt.xlabel('Month')
plt.show()
```



Now we know that December is the month with the most sales, but that is a fairly obvious statement given that December is when some parts of the world celebrate Christmas. I am now going to add a city column so that we can take a look at the data on a city by city basis.

[15]:
```
all_data.head()
```

[15]:
```
  Order ID                    Product  Quantity Ordered  Price Each   Sales  \
0   176558        USB-C Charging Cable                 2       11.95   23.90
2   176559  Bose SoundSport Headphones                 1       99.99   99.99
3   176560                Google Phone                 1      600.00  600.00
4   176560            Wired Headphones                 1       11.99   11.99
5   176561            Wired Headphones                 1       11.99   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
```

[16]:
```
temp_df = all_data['Purchase Address'].str.split(',', n=2, expand=True)
```

[17]:
```
temp_df.head()
```

[17]:
```
                0            1          2
0       917 1st St       Dallas   TX 75001
2  682 Chestnut St       Boston   MA 02215
3    669 Spruce St  Los Angeles   CA 90001
4    669 Spruce St  Los Angeles   CA 90001
5       333 8th St  Los Angeles   CA 90001
```

Creating a new city column and rearranging colulmn order.

[18]:
```
all_data['City'] = temp_df[1]
all_data.head()
```

[18]:
```
  Order ID                    Product  Quantity Ordered  Price Each   Sales  \
0   176558        USB-C Charging Cable                 2       11.95   23.90
2   176559  Bose SoundSport Headphones                 1       99.99   99.99
3   176560                Google Phone                 1      600.00  600.00
4   176560            Wired Headphones                 1       11.99   11.99
5   176561            Wired Headphones                 1       11.99   11.99

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

[19]:
```
all_data = all_data[['Order ID', 'Product', 'Quantity Ordered', 'Price Each',␣
↪'Sales', 'Order Date', 'Month', 'City', 'Purchase Address']]
all_data.head()
```

[19]:
```
  Order ID                    Product  Quantity Ordered  Price Each   Sales  \
0   176558        USB-C Charging Cable                 2       11.95   23.90
2   176559  Bose SoundSport Headphones                 1       99.99   99.99
3   176560                Google Phone                 1      600.00  600.00
4   176560            Wired Headphones                 1       11.99   11.99
5   176561            Wired Headphones                 1       11.99   11.99

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

American city names will not suffice since there are duplicate city names all over the place. Washington and Maine come to mind. I am sure there are others. I will make a city column inclusive of this added complexity.

[20]:
```
def get_state(address):
    return address.split(',')[2].split(' ',)[1]

def get_city(address):
    return address.split(',')[1]

all_data['City'] = all_data['Purchase Address'].apply(lambda x: f'{get_city(x)} ({get_state(x)})')
all_data.head()
```

[20]:
```
  Order ID                    Product  Quantity Ordered  Price Each   Sales  \
0   176558        USB-C Charging Cable                 2       11.95   23.90
2   176559  Bose SoundSport Headphones                 1       99.99   99.99
3   176560                Google Phone                 1      600.00  600.00
4   176560            Wired Headphones                 1       11.99   11.99
5   176561            Wired Headphones                 1       11.99   11.99

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

                      Purchase Address
0          917 1st St, Dallas, TX 75001
2     682 Chestnut St, Boston, MA 02215
3  669 Spruce St, Los Angeles, CA 90001
4  669 Spruce St, Los Angeles, CA 90001
```

```
5     333 8th St, Los Angeles, CA 90001
```

0.1.9 Question 2: Which city had the highest number of sales

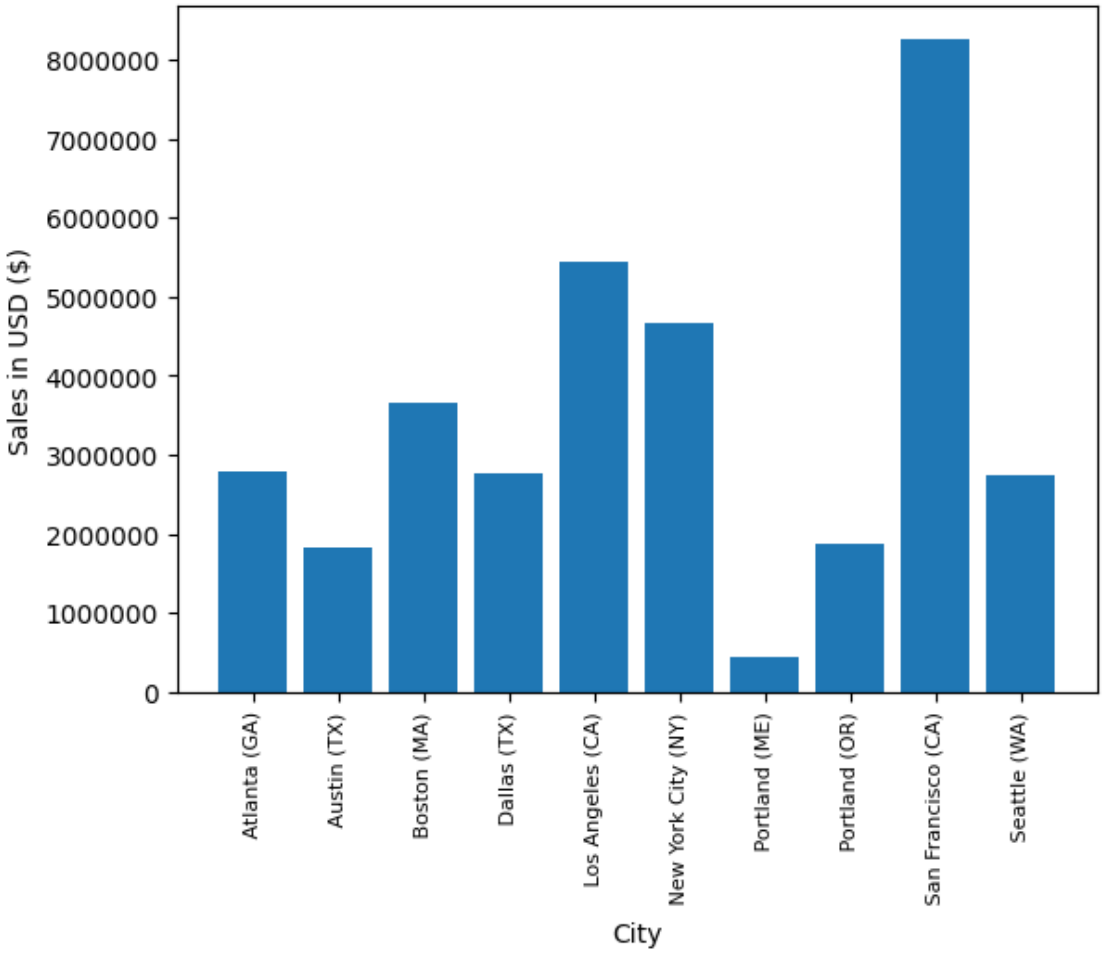
[21]:
```
result = all_data.groupby('City').sum('Numeric')
result.head()
```

[21]:

City	Quantity Ordered	Price Each	Sales	Month
Atlanta (GA)	16602	2779908.20	2795498.58	104794
Austin (TX)	11153	1809873.61	1819581.75	69829
Boston (MA)	22528	3637409.77	3661642.01	141112
Dallas (TX)	16730	2752627.82	2767975.40	104620
Los Angeles (CA)	33289	5421435.23	5452570.80	208325

[22]:
```
cities = [city for city, df in all_data.groupby('City')]

plt.ticklabel_format(style='plain')
plt.bar(cities, result['Sales'])
plt.xticks(cities, rotation='vertical', size=8)
plt.ylabel('Sales in USD ($)')
plt.xlabel('City')
plt.show()
```



0.1.10 Question 3: What time should we display advertisements to maximise purchases?

The calculation actually took about 7 seconds to run too.

[23]:
```
all_data['Order Date'] = pd.to_datetime(all_data['Order Date'])
all_data.head()
```

[23]:

	Order ID	Product	Quantity Ordered	Price Each	Sales	\
0	176558	USB-C Charging Cable	2	11.95	23.90	
2	176559	Bose SoundSport Headphones	1	99.99	99.99	
3	176560	Google Phone	1	600.00	600.00	
4	176560	Wired Headphones	1	11.99	11.99	
5	176561	Wired Headphones	1	11.99	11.99	

	Order Date	Month	City	\
0	2019-04-19 08:46:00	4	Dallas (TX)	

```
2 2019-04-07 22:30:00      4        Boston (MA)
3 2019-04-12 14:38:00      4   Los Angeles (CA)
4 2019-04-12 14:38:00      4   Los Angeles (CA)
5 2019-04-30 09:27:00      4   Los Angeles (CA)

                      Purchase Address
0          917 1st St, Dallas, TX 75001
2     682 Chestnut St, Boston, MA 02215
3  669 Spruce St, Los Angeles, CA 90001
4  669 Spruce St, Los Angeles, CA 90001
5     333 8th St, Los Angeles, CA 90001
```

[24]:
```
all_data['Hour'] = all_data['Order Date'].dt.hour
all_data['Minute'] = all_data['Order Date'].dt.minute
all_data.head()
```

[24]:
```
  Order ID                    Product  Quantity Ordered  Price Each   Sales  \
0   176558        USB-C Charging Cable                 2       11.95   23.90
2   176559  Bose SoundSport Headphones                 1       99.99   99.99
3   176560                Google Phone                 1      600.00  600.00
4   176560            Wired Headphones                 1       11.99   11.99
5   176561            Wired Headphones                 1       11.99   11.99

           Order Date  Month              City  \
0 2019-04-19 08:46:00      4        Dallas (TX)
2 2019-04-07 22:30:00      4        Boston (MA)
3 2019-04-12 14:38:00      4   Los Angeles (CA)
4 2019-04-12 14:38:00      4   Los Angeles (CA)
5 2019-04-30 09:27:00      4   Los Angeles (CA)

                      Purchase Address  Hour  Minute
0          917 1st St, Dallas, TX 75001     8      46
2     682 Chestnut St, Boston, MA 02215    22      30
3  669 Spruce St, Los Angeles, CA 90001    14      38
4  669 Spruce St, Los Angeles, CA 90001    14      38
5     333 8th St, Los Angeles, CA 90001     9      27
```

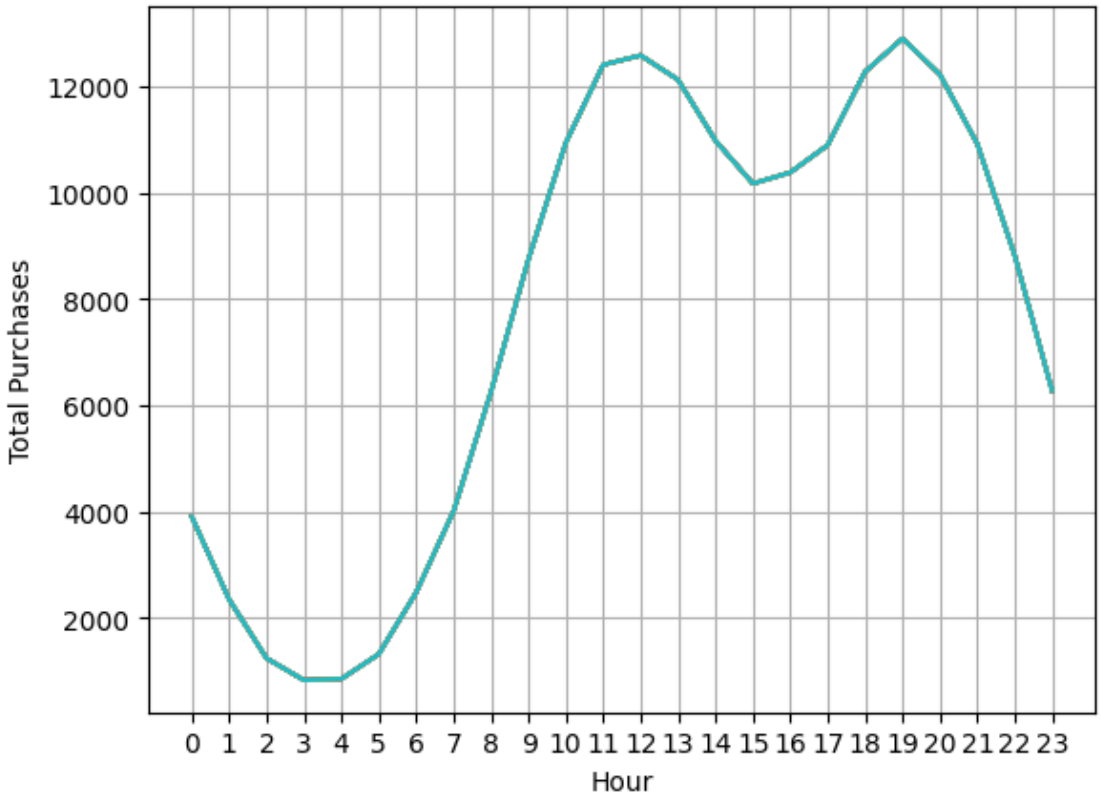
[25]:
```
hours = [hour for hour, df in all_data.groupby('Hour')]
all_data.groupby(['Hour']).count().head(24)
```

[25]:
```
      Order ID  Product  Quantity Ordered  Price Each  Sales  Order Date  \
Hour
0         3910     3910              3910        3910   3910        3910
1         2350     2350              2350        2350   2350        2350
2         1243     1243              1243        1243   1243        1243
3          831      831               831         831    831         831
4          854      854               854         854    854         854
```

5	1321	1321	1321	1321	1321	1321
6	2482	2482	2482	2482	2482	2482
7	4011	4011	4011	4011	4011	4011
8	6256	6256	6256	6256	6256	6256
9	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

Hour	Month	City	Purchase Address	Minute
0	3910	3910	3910	3910
1	2350	2350	2350	2350
2	1243	1243	1243	1243
3	831	831	831	831
4	854	854	854	854
5	1321	1321	1321	1321
6	2482	2482	2482	2482
7	4011	4011	4011	4011
8	6256	6256	6256	6256
9	8748	8748	8748	8748
10	10944	10944	10944	10944
11	12411	12411	12411	12411
12	12587	12587	12587	12587
13	12129	12129	12129	12129
14	10984	10984	10984	10984
15	10175	10175	10175	10175
16	10384	10384	10384	10384
17	10899	10899	10899	10899
18	12280	12280	12280	12280
19	12905	12905	12905	12905
20	12228	12228	12228	12228
21	10921	10921	10921	10921
22	8822	8822	8822	8822
23	6275	6275	6275	6275

[26]:
```
plt.plot(hours, all_data.groupby(['Hour']).count())
plt.ylabel('Total Purchases')
plt.xlabel('Hour')
plt.xticks(hours)
plt.grid()
plt.show()
```



Hours that are most busy are 12PM and 7PM. Displaying ads an hour before this time up until these peak purchase hours might be wise.

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

[27]:
```
pair_purchase = all_data[all_data['Order ID'].duplicated(keep=False)]
pair_purchase['Bundled'] = pair_purchase.groupby('Order ID')['Product'].
  ↪transform(lambda x: ','.join(x))
pair_purchase = pair_purchase[['Order ID', 'Bundled']].drop_duplicates()
pair_purchase.head()
```

```
/tmp/ipykernel_6785/203834445.py:2: 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
  pair_purchase['Bundled'] = pair_purchase.groupby('Order
ID')['Product'].transform(lambda x: ','.join(x))
```

[27]:

	Order ID	Bundled
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

[28]:
```
from itertools import combinations
from collections import Counter
```

[29]:
```
count = Counter()

for row in pair_purchase['Bundled']:
    row_list = row.split(',')
    count.update(Counter(combinations(row_list, 2)))

print(count)
```

```
Counter({('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, ('Lightning Charging Cable', 'Apple Airpods
Headphones'): 81, ('Vareebadd Phone', 'Bose SoundSport Headphones'): 80, ('USB-C
Charging Cable', 'Bose SoundSport Headphones'): 77, ('Apple Airpods Headphones',
'Wired Headphones'): 69, ('Lightning Charging Cable', 'USB-C Charging Cable'):
58, ('Lightning Charging Cable', 'AA Batteries (4-pack)'): 55, ('Lightning
Charging Cable', 'Lightning Charging Cable'): 54, ('Bose SoundSport Headphones',
'Wired Headphones'): 53, ('AA Batteries (4-pack)', 'Lightning Charging Cable'):
51, ('AAA Batteries (4-pack)', 'USB-C Charging Cable'): 50, ('Apple Airpods
Headphones', 'AAA Batteries (4-pack)'): 48, ('AA Batteries (4-pack)', 'AAA
Batteries (4-pack)'): 48, ('USB-C Charging Cable', 'USB-C Charging Cable'): 48,
('AAA Batteries (4-pack)', 'AAA Batteries (4-pack)'): 48, ('USB-C Charging
Cable', 'AAA Batteries (4-pack)'): 45, ('Wired Headphones', 'USB-C Charging
Cable'): 45, ('AA Batteries (4-pack)', 'Wired Headphones'): 44, ('AAA Batteries
(4-pack)', 'Lightning Charging Cable'): 44, ('AAA Batteries (4-pack)', 'Wired
Headphones'): 43, ('Wired Headphones', 'AAA Batteries (4-pack)'): 43, ('USB-C
Charging Cable', 'Lightning Charging Cable'): 42, ('AA Batteries (4-pack)',
'Apple Airpods Headphones'): 41, ('AAA Batteries (4-pack)', 'AA Batteries
(4-pack)'): 39, ('Wired Headphones', 'AA Batteries (4-pack)'): 39, ('Lightning
Charging Cable', 'Bose SoundSport Headphones'): 39, ('USB-C Charging Cable', 'AA
```

Batteries (4-pack)'): 38, ('Bose SoundSport Headphones', 'AAA Batteries (4-pack)'): 37, ('AA Batteries (4-pack)', 'USB-C Charging Cable'): 37, ('Wired Headphones', 'Lightning Charging Cable'): 37, ('Lightning Charging Cable', 'AAA Batteries (4-pack)'): 36, ('Apple Airpods Headphones', 'Lightning Charging Cable'): 35, ('Wired Headphones', 'Wired Headphones'): 35, ('AA Batteries (4-pack)', 'AA Batteries (4-pack)'): 35, ('USB-C Charging Cable', 'Apple Airpods Headphones'): 34, ('Bose SoundSport Headphones', 'Lightning Charging Cable'): 33, ('AAA Batteries (4-pack)', 'Apple Airpods Headphones'): 33, ('Apple Airpods Headphones', 'Bose SoundSport Headphones'): 32, ('Wired Headphones', 'Apple Airpods Headphones'): 31, ('USB-C Charging Cable', '27in FHD Monitor'): 31, ('Apple Airpods Headphones', 'USB-C Charging Cable'): 29, ('Apple Airpods Headphones', 'AA Batteries (4-pack)'): 29, ('AA Batteries (4-pack)', 'Bose SoundSport Headphones'): 28, ('Bose SoundSport Headphones', 'Bose SoundSport Headphones'): 27, ('Bose SoundSport Headphones', 'AA Batteries (4-pack)'): 27, ('Bose SoundSport Headphones', 'USB-C Charging Cable'): 25, ('Apple Airpods Headphones', 'Apple Airpods Headphones'): 24, ('AAA Batteries (4-pack)', '27in FHD Monitor'): 22, ('27in FHD Monitor', 'AAA Batteries (4-pack)'): 21, ('Wired Headphones', 'Bose SoundSport Headphones'): 21, ('AAA Batteries (4-pack)', 'Bose SoundSport Headphones'): 20, ('34in Ultrawide Monitor', 'AA Batteries (4-pack)'): 19, ('Lightning Charging Cable', '27in 4K Gaming Monitor'): 18, ('AA Batteries (4-pack)', 'iPhone'): 18, ('27in FHD Monitor', 'Lightning Charging Cable'): 18, ('Lightning Charging Cable', '27in FHD Monitor'): 18, ('34in Ultrawide Monitor', 'Lightning Charging Cable'): 18, ('Bose SoundSport Headphones', 'Apple Airpods Headphones'): 18, ('Wired Headphones', '27in 4K Gaming Monitor'): 18, ('iPhone', 'AAA Batteries (4-pack)'): 17, ('Wired Headphones', '34in Ultrawide Monitor'): 17, ('ThinkPad Laptop', 'AAA Batteries (4-pack)'): 16, ('Lightning Charging Cable', 'Google Phone'): 16, ('27in 4K Gaming Monitor', 'Lightning Charging Cable'): 16, ('34in Ultrawide Monitor', 'USB-C Charging Cable'): 15, ('27in FHD Monitor', 'AA Batteries (4-pack)'): 15, ('Wired Headphones', 'iPhone'): 15, ('AAA Batteries (4-pack)', '27in 4K Gaming Monitor'): 15, ('iPhone', 'USB-C Charging Cable'): 15, ('20in Monitor', 'USB-C Charging Cable'): 15, ('Lightning Charging Cable', '20in Monitor'): 15, ('27in 4K Gaming Monitor', 'AAA Batteries (4-pack)'): 15, ('Lightning Charging Cable', '34in Ultrawide Monitor'): 15, ('Google Phone', 'AA Batteries (4-pack)'): 14, ('Apple Airpods Headphones', 'Google Phone'): 14, ('USB-C Charging Cable', 'iPhone'): 14, ('Bose SoundSport Headphones', '27in FHD Monitor'): 14, ('AA Batteries (4-pack)', '27in 4K Gaming Monitor'): 14, ('AAA Batteries (4-pack)', 'iPhone'): 14, ('iPhone', 'AA Batteries (4-pack)'): 14, ('AA Batteries (4-pack)', 'Flatscreen TV'): 13, ('AA Batteries (4-pack)', '34in Ultrawide Monitor'): 13, ('AAA Batteries (4-pack)', '34in Ultrawide Monitor'): 13, ('Wired Headphones', 'Macbook Pro Laptop'): 13, ('Apple Airpods Headphones', 'iPhone'): 13, ('Apple Airpods Headphones', '27in 4K Gaming Monitor'): 12, ('Apple Airpods Headphones', '27in FHD Monitor'): 12, ('27in FHD Monitor', 'Bose SoundSport Headphones'): 12, ('27in FHD Monitor', 'USB-C Charging Cable'): 12, ('Google Phone', 'Lightning Charging Cable'): 12, ('Apple Airpods Headphones', 'Macbook Pro Laptop'): 12, ('Wired Headphones', '27in FHD Monitor'): 12, ('USB-C Charging Cable', 'Google Phone'): 12, ('Macbook Pro Laptop', 'USB-C Charging Cable'): 12, ('34in Ultrawide Monitor', 'AAA Batteries (4-pack)'): 12, ('Lightning Charging

Cable', 'Flatscreen TV'): 12, ('27in 4K Gaming Monitor', 'USB-C Charging Cable'): 12, ('27in FHD Monitor', 'Apple Airpods Headphones'): 12, ('20in Monitor', 'Wired Headphones'): 12, ('27in 4K Gaming Monitor', 'AA Batteries (4-pack)'): 12, ('AAA Batteries (4-pack)', 'Google Phone'): 11, ('AAA Batteries (4-pack)', 'Macbook Pro Laptop'): 11, ('USB-C Charging Cable', '27in 4K Gaming Monitor'): 11, ('USB-C Charging Cable', 'ThinkPad Laptop'): 11, ('34in Ultrawide Monitor', 'Wired Headphones'): 11, ('20in Monitor', 'Lightning Charging Cable'): 11, ('AA Batteries (4-pack)', '27in FHD Monitor'): 11, ('Bose SoundSport Headphones', '34in Ultrawide Monitor'): 11, ('ThinkPad Laptop', 'Lightning Charging Cable'): 11, ('Google Phone', 'AAA Batteries (4-pack)'): 11, ('USB-C Charging Cable', '34in Ultrawide Monitor'): 11, ('Flatscreen TV', 'AAA Batteries (4-pack)'): 11, ('Macbook Pro Laptop', 'Bose SoundSport Headphones'): 11, ('Macbook Pro Laptop', 'Lightning Charging Cable'): 11, ('27in 4K Gaming Monitor', 'Wired Headphones'): 11, ('AA Batteries (4-pack)', 'Google Phone'): 11, ('AAA Batteries (4-pack)', 'ThinkPad Laptop'): 11, ('Flatscreen TV', 'Lightning Charging Cable'): 10, ('Wired Headphones', 'ThinkPad Laptop'): 10, ('USB-C Charging Cable', '20in Monitor'): 10, ('27in 4K Gaming Monitor', 'Apple Airpods Headphones'): 10, ('USB-C Charging Cable', 'Flatscreen TV'): 10, ('27in FHD Monitor', 'Wired Headphones'): 10, ('AA Batteries (4-pack)', '20in Monitor'): 10, ('AAA Batteries (4-pack)', 'Flatscreen TV'): 10, ('Lightning Charging Cable', 'iPhone'): 10, ('Bose SoundSport Headphones', 'Flatscreen TV'): 10, ('Bose SoundSport Headphones', '27in 4K Gaming Monitor'): 10, ('Lightning Charging Cable', 'Macbook Pro Laptop'): 10, ('Apple Airpods Headphones', 'ThinkPad Laptop'): 9, ('Wired Headphones', 'Google Phone'): 9, ('27in 4K Gaming Monitor', 'Bose SoundSport Headphones'): 9, ('20in Monitor', 'Bose SoundSport Headphones'): 9, ('Macbook Pro Laptop', 'AA Batteries (4-pack)'): 9, ('ThinkPad Laptop', 'USB-C Charging Cable'): 9, ('ThinkPad Laptop', 'Bose SoundSport Headphones'): 9, ('Vareebadd Phone', 'AA Batteries (4-pack)'): 9, ('USB-C Charging Cable', 'Macbook Pro Laptop'): 9, ('Lightning Charging Cable', 'ThinkPad Laptop'): 9, ('AA Batteries (4-pack)', 'ThinkPad Laptop'): 9, ('27in FHD Monitor', '27in FHD Monitor'): 9, ('AA Batteries (4-pack)', 'Macbook Pro Laptop'): 8, ('Flatscreen TV', 'AA Batteries (4-pack)'): 8, ('Apple Airpods Headphones', 'Flatscreen TV'): 8, ('ThinkPad Laptop', 'AA Batteries (4-pack)'): 8, ('34in Ultrawide Monitor', 'Apple Airpods Headphones'): 8, ('AAA Batteries (4-pack)', '20in Monitor'): 8, ('Bose SoundSport Headphones', 'Google Phone'): 8, ('20in Monitor', 'Apple Airpods Headphones'): 7, ('Macbook Pro Laptop', 'Apple Airpods Headphones'): 7, ('Wired Headphones', 'Flatscreen TV'): 7, ('Wired Headphones', '20in Monitor'): 7, ('Macbook Pro Laptop', 'Wired Headphones'): 7, ('USB-C Charging Cable', 'Vareebadd Phone'): 7, ('Google Phone', '27in FHD Monitor'): 7, ('Macbook Pro Laptop', 'AAA Batteries (4-pack)'): 7, ('34in Ultrawide Monitor', 'iPhone'): 7, ('34in Ultrawide Monitor', '34in Ultrawide Monitor'): 7, ('Bose SoundSport Headphones', 'iPhone'): 7, ('Flatscreen TV', 'USB-C Charging Cable'): 7, ('Google Phone', 'Apple Airpods Headphones'): 7, ('Macbook Pro Laptop', '27in 4K Gaming Monitor'): 7, ('ThinkPad Laptop', 'Apple Airpods Headphones'): 7, ('iPhone', '27in 4K Gaming Monitor'): 6, ('Flatscreen TV', 'Flatscreen TV'): 6, ('Apple Airpods Headphones', '34in Ultrawide Monitor'): 6, ('iPhone', '34in Ultrawide Monitor'): 6, ('Vareebadd Phone', 'Apple Airpods Headphones'): 6, ('27in 4K

Gaming Monitor', '34in Ultrawide Monitor'): 6, ('27in 4K Gaming Monitor', 'Macbook Pro Laptop'): 6, ('Apple Airpods Headphones', 'Vareebadd Phone'): 6, ('20in Monitor', 'AA Batteries (4-pack)'): 6, ('27in FHD Monitor', 'Macbook Pro Laptop'): 6, ('Wired Headphones', 'Vareebadd Phone'): 6, ('Apple Airpods Headphones', '20in Monitor'): 6, ('iPhone', 'Flatscreen TV'): 6, ('34in Ultrawide Monitor', 'Bose SoundSport Headphones'): 6, ('Google Phone', 'iPhone'): 6, ('Bose SoundSport Headphones', '20in Monitor'): 6, ('iPhone', 'Bose SoundSport Headphones'): 5, ('27in 4K Gaming Monitor', 'Google Phone'): 5, ('Flatscreen TV', 'Apple Airpods Headphones'): 5, ('27in 4K Gaming Monitor', '27in 4K Gaming Monitor'): 5, ('Flatscreen TV', '34in Ultrawide Monitor'): 5, ('34in Ultrawide Monitor', '27in FHD Monitor'): 5, ('27in FHD Monitor', '34in Ultrawide Monitor'): 5, ('Macbook Pro Laptop', '34in Ultrawide Monitor'): 4, ('iPhone', 'Vareebadd Phone'): 4, ('Bose SoundSport Headphones', 'ThinkPad Laptop'): 4, ('20in Monitor', 'Macbook Pro Laptop'): 4, ('Vareebadd Phone', '34in Ultrawide Monitor'): 4, ('Flatscreen TV', 'Wired Headphones'): 4, ('Flatscreen TV', '27in FHD Monitor'): 4, ('iPhone', 'ThinkPad Laptop'): 4, ('Flatscreen TV', 'iPhone'): 4, ('Bose SoundSport Headphones', 'Macbook Pro Laptop'): 4, ('Flatscreen TV', 'Macbook Pro Laptop'): 4, ('ThinkPad Laptop', 'Wired Headphones'): 4, ('27in 4K Gaming Monitor', '27in FHD Monitor'): 4, ('27in 4K Gaming Monitor', 'ThinkPad Laptop'): 4, ('27in FHD Monitor', 'ThinkPad Laptop'): 4, ('Macbook Pro Laptop', 'Google Phone'): 4, ('AAA Batteries (4-pack)', 'Vareebadd Phone'): 4, ('ThinkPad Laptop', 'Flatscreen TV'): 4, ('LG Dryer', 'AA Batteries (4-pack)'): 4, ('27in FHD Monitor', '27in 4K Gaming Monitor'): 4, ('Vareebadd Phone', 'Google Phone'): 4, ('LG Washing Machine', 'AAA Batteries (4-pack)'): 4, ('Lightning Charging Cable', 'LG Washing Machine'): 4, ('Macbook Pro Laptop', 'ThinkPad Laptop'): 3, ('ThinkPad Laptop', 'Google Phone'): 3, ('34in Ultrawide Monitor', 'Macbook Pro Laptop'): 3, ('Lightning Charging Cable', 'Vareebadd Phone'): 3, ('Google Phone', 'ThinkPad Laptop'): 3, ('20in Monitor', '20in Monitor'): 3, ('20in Monitor', 'AAA Batteries (4-pack)'): 3, ('AA Batteries (4-pack)', 'Vareebadd Phone'): 3, ('iPhone', 'Google Phone'): 3, ('Apple Airpods Headphones', 'LG Dryer'): 3, ('27in 4K Gaming Monitor', 'iPhone'): 3, ('Macbook Pro Laptop', 'Macbook Pro Laptop'): 3, ('Flatscreen TV', 'Google Phone'): 3, ('Macbook Pro Laptop', 'iPhone'): 3, ('34in Ultrawide Monitor', 'ThinkPad Laptop'): 3, ('iPhone', 'Macbook Pro Laptop'): 3, ('27in FHD Monitor', '20in Monitor'): 3, ('Vareebadd Phone', 'Flatscreen TV'): 3, ('Flatscreen TV', 'Bose SoundSport Headphones'): 3, ('ThinkPad Laptop', 'iPhone'): 3, ('34in Ultrawide Monitor', 'Flatscreen TV'): 3, ('Google Phone', 'Macbook Pro Laptop'): 3, ('Google Phone', 'Google Phone'): 3, ('Macbook Pro Laptop', '27in FHD Monitor'): 3, ('Vareebadd Phone', 'iPhone'): 3, ('Wired Headphones', 'LG Washing Machine'): 3, ('Google Phone', '34in Ultrawide Monitor'): 3, ('27in 4K Gaming Monitor', 'Flatscreen TV'): 3, ('34in Ultrawide Monitor', 'Google Phone'): 2, ('Macbook Pro Laptop', '20in Monitor'): 2, ('Lightning Charging Cable', 'LG Dryer'): 2, ('Flatscreen TV', '27in 4K Gaming Monitor'): 2, ('ThinkPad Laptop', 'Macbook Pro Laptop'): 2, ('ThinkPad Laptop', '27in FHD Monitor'): 2, ('Vareebadd Phone', 'AAA Batteries (4-pack)'): 2, ('27in FHD Monitor', 'LG Dryer'): 2, ('27in FHD Monitor', 'Flatscreen TV'): 2, ('Macbook Pro Laptop', 'LG Washing Machine'): 2, ('Google Phone', '20in Monitor'): 2, ('20in Monitor', 'Google Phone'): 2, ('iPhone', '20in Monitor'):

```
2, ('iPhone', 'iPhone'): 2, ('LG Washing Machine', 'Bose SoundSport
Headphones'): 2, ('AA Batteries (4-pack)', 'LG Dryer'): 2, ('Bose SoundSport
Headphones', 'Vareebadd Phone'): 2, ('Flatscreen TV', 'ThinkPad Laptop'): 2,
('LG Washing Machine', 'Lightning Charging Cable'): 2, ('20in Monitor', '27in
FHD Monitor'): 2, ('ThinkPad Laptop', 'ThinkPad Laptop'): 2, ('Vareebadd Phone',
'ThinkPad Laptop'): 2, ('20in Monitor', 'ThinkPad Laptop'): 2, ('Vareebadd
Phone', '27in 4K Gaming Monitor'): 2, ('27in 4K Gaming Monitor', '20in
Monitor'): 2, ('27in 4K Gaming Monitor', 'Vareebadd Phone'): 1, ('27in FHD
Monitor', 'iPhone'): 1, ('Vareebadd Phone', 'Lightning Charging Cable'): 1,
('Vareebadd Phone', '27in FHD Monitor'): 1, ('ThinkPad Laptop', '27in 4K Gaming
Monitor'): 1, ('20in Monitor', 'Flatscreen TV'): 1, ('USB-C Charging Cable', 'LG
Dryer'): 1, ('LG Washing Machine', '20in Monitor'): 1, ('Flatscreen TV', '20in
Monitor'): 1, ('LG Dryer', 'AAA Batteries (4-pack)'): 1, ('ThinkPad Laptop',
'34in Ultrawide Monitor'): 1, ('iPhone', 'LG Washing Machine'): 1, ('AAA
Batteries (4-pack)', 'LG Dryer'): 1, ('LG Dryer', '27in 4K Gaming Monitor'): 1,
('Google Phone', 'Vareebadd Phone'): 1, ('20in Monitor', 'iPhone'): 1, ('LG
Dryer', 'Vareebadd Phone'): 1, ('34in Ultrawide Monitor', 'LG Washing Machine'):
1, ('Google Phone', '27in 4K Gaming Monitor'): 1, ('LG Washing Machine',
'iPhone'): 1, ('LG Dryer', 'Wired Headphones'): 1, ('LG Dryer', 'Flatscreen
TV'): 1, ('27in FHD Monitor', 'Google Phone'): 1, ('iPhone', '27in FHD
Monitor'): 1, ('27in FHD Monitor', 'LG Washing Machine'): 1, ('LG Dryer', '27in
FHD Monitor'): 1, ('20in Monitor', '34in Ultrawide Monitor'): 1, ('34in
Ultrawide Monitor', '20in Monitor'): 1, ('Macbook Pro Laptop', 'Flatscreen TV'):
1, ('ThinkPad Laptop', 'Vareebadd Phone'): 1, ('Google Phone', 'Flatscreen TV'):
1, ('LG Washing Machine', 'Google Phone'): 1, ('LG Washing Machine', 'Wired
Headphones'): 1, ('27in FHD Monitor', 'Vareebadd Phone'): 1, ('LG Washing
Machine', '27in 4K Gaming Monitor'): 1, ('LG Washing Machine', 'Apple Airpods
Headphones'): 1, ('27in 4K Gaming Monitor', 'LG Dryer'): 1, ('20in Monitor', 'LG
Washing Machine'): 1, ('LG Dryer', 'Google Phone'): 1, ('LG Dryer', 'Lightning
Charging Cable'): 1, ('ThinkPad Laptop', 'LG Dryer'): 1, ('LG Washing Machine',
'AA Batteries (4-pack)'): 1})
```

[30]:
```
for key, value in count.most_common(15):
    print(key, value)
```

```
('iPhone', 'Lightning Charging Cable') 1005
('Google Phone', 'USB-C Charging Cable') 987
('iPhone', 'Wired Headphones') 447
('Google Phone', 'Wired Headphones') 414
('Vareebadd Phone', 'USB-C Charging Cable') 361
('iPhone', 'Apple Airpods Headphones') 360
('Google Phone', 'Bose SoundSport Headphones') 220
('USB-C Charging Cable', 'Wired Headphones') 160
('Vareebadd Phone', 'Wired Headphones') 143
('Lightning Charging Cable', 'Wired Headphones') 92
('Lightning Charging Cable', 'Apple Airpods Headphones') 81
('Vareebadd Phone', 'Bose SoundSport Headphones') 80
('USB-C Charging Cable', 'Bose SoundSport Headphones') 77
```

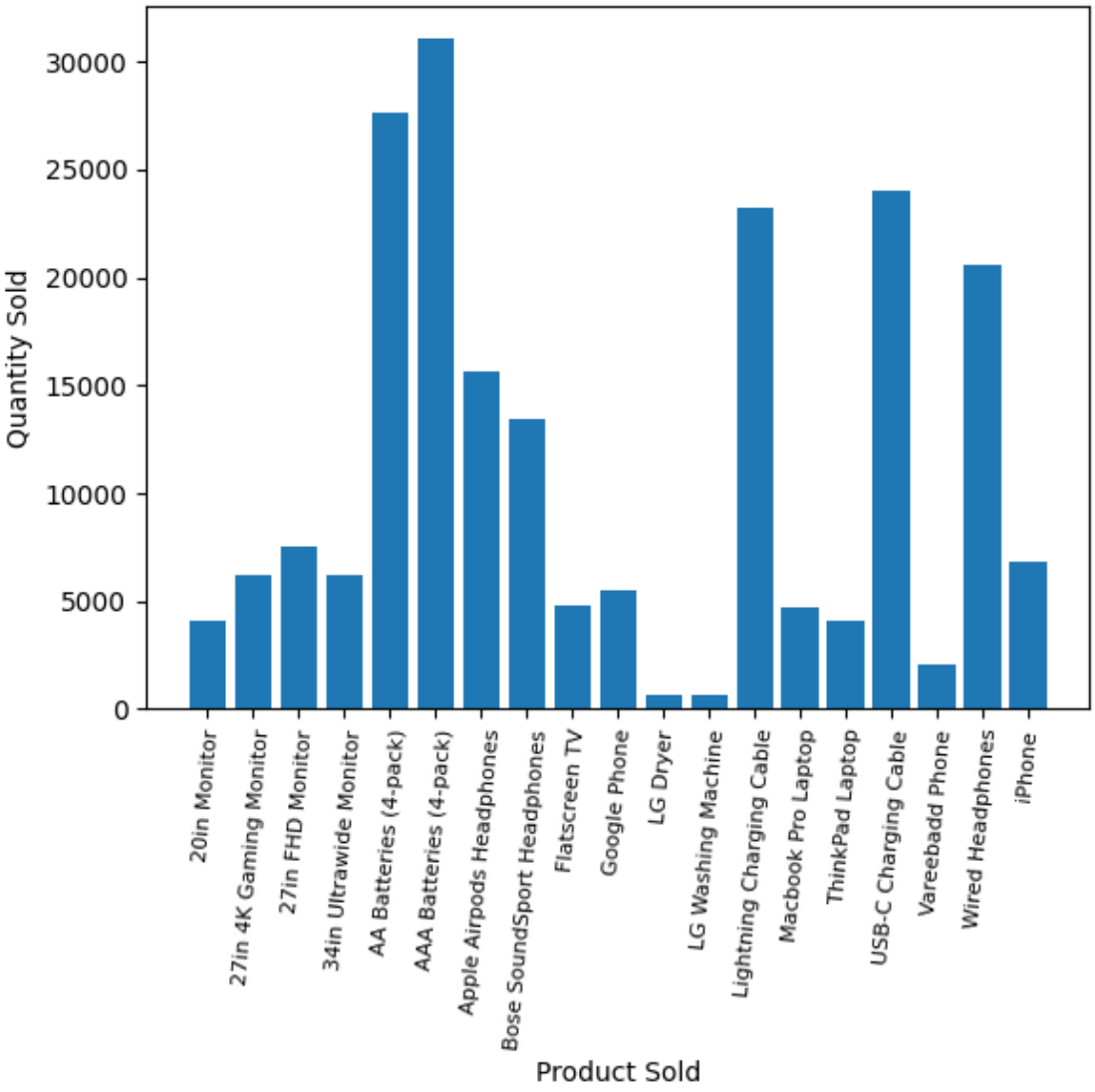
```
('Apple Airpods Headphones', 'Wired Headphones') 69
('Lightning Charging Cable', 'USB-C Charging Cable') 58
```

0.1.12 Question 5: What products sold the most?

[31]:
```
product_group = all_data.groupby('Product')
quantity_ordered = product_group.sum('Numeric')['Quantity Ordered']
```

[32]:
```
products = [product for product, df in product_group]

plt.xticks(rotation=85, size=8)
plt.xlabel('Product Sold')
plt.ylabel('Quantity Sold')
plt.bar(products, quantity_ordered)
plt.show()
```



Overlay with prices on graph

[33]:
```
prices = all_data.groupby('Product').mean('Numeric')['Price Each']
print(prices)
```

```
Product
20in Monitor                  109.99
27in 4K Gaming Monitor        389.99
27in FHD Monitor              149.99
34in Ultrawide Monitor        379.99
AA Batteries (4-pack)           3.84
AAA Batteries (4-pack)          2.99
Apple Airpods Headphones      150.00
Bose SoundSport Headphones     99.99
Flatscreen TV                 300.00
Google Phone                  600.00
LG Dryer                      600.00
LG Washing Machine            600.00
Lightning Charging Cable       14.95
Macbook Pro Laptop           1700.00
ThinkPad Laptop               999.99
USB-C Charging Cable           11.95
Vareebadd Phone               400.00
Wired Headphones               11.99
iPhone                        700.00
Name: Price Each, dtype: float64
```

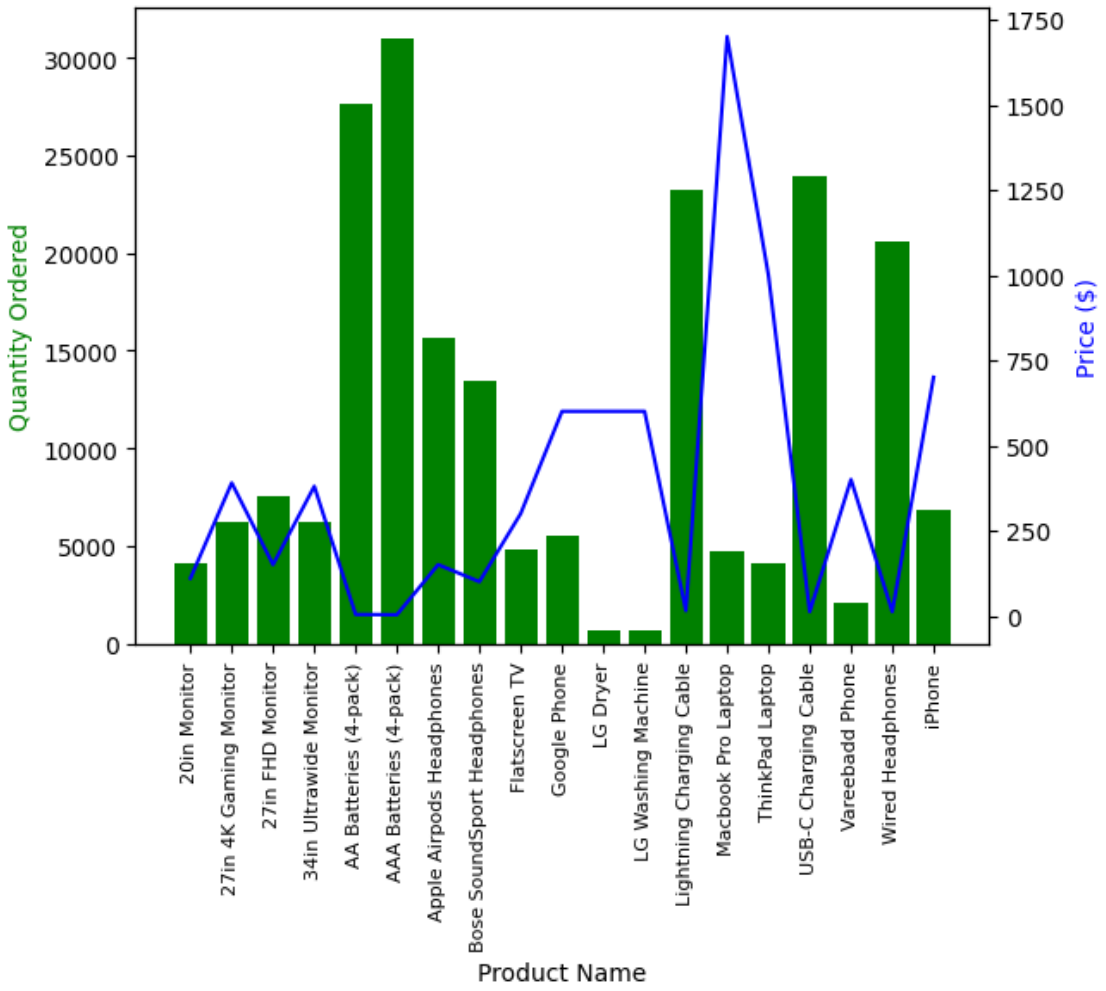
[35]:
```
fig, ax1 = plt.subplots()

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

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


plt.show()
```

```
/tmp/ipykernel_6785/2046822335.py:10: UserWarning: FixedFormatter should only be
used together with FixedLocator
  ax1.set_xticklabels(products, rotation='vertical', size=8)
```


Quantity Ordered
Price ($)
Product Name
0
5000
10000
15000
20000
25000
30000
0
250
500
750
1000
1250
1500
1750
20in Monitor
27in 4K Gaming Monitor
27in FHD Monitor
34in Ultrawide Monitor
AA Batteries (4-pack)
AAA Batteries (4-pack)
Apple Airpods Headphones
Bose SoundSport Headphones
Flatscreen TV
Google Phone
LG Dryer
LG Washing Machine
Lightning Charging Cable
Macbook Pro Laptop
ThinkPad Laptop
USB-C Charging Cable
Vareebadd Phone
Wired Headphones
iPhone