In [1]:

```
import pandas as pd
import os
```

## Task 1: Merging 12 months of sales data into a single file

In [2]:

```
df = pd.read_csv("./Documents/SalesAnalysis/Sales_Data/Sales_April_2019.csv")

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

for file in files:
    print(file)
```

```
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 [3]:

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

all_months_data = pd.DataFrame()

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

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

## Read in updated dataframe

In [4]:

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

Out[4]:

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

## Clean the data

In [5]:

```
all_data.head(100)
```

Out[5]:

| | Order ID | Product | Quantity Ordered | Price Each | Order Date | Purchase Address |
|---|---|---|---|---|---|---|
| **0** | 176558 | USB-C Charging Cable | 2 | 11.95 | 04/19/19 08:46 | 917 1st St, Dallas, TX 75001 |
| **1** | NaN | NaN | NaN | NaN | NaN | NaN |
| **2** | 176559 | Bose SoundSport Headphones | 1 | 99.99 | 04/07/19 22:30 | 682 Chestnut St, Boston, MA 02215 |
| **3** | 176560 | Google Phone | 1 | 600 | 04/12/19 14:38 | 669 Spruce St, Los Angeles, CA 90001 |
| **4** | 176560 | Wired Headphones | 1 | 11.99 | 04/12/19 14:38 | 669 Spruce St, Los Angeles, CA 90001 |
| **...** | ... | ... | ... | ... | ... | ... |
| **95** | 176648 | 27in FHD Monitor | 1 | 149.99 | 04/24/19 01:17 | 732 2nd St, Portland, OR 97035 |
| **96** | 176649 | USB-C Charging Cable | 1 | 11.95 | 04/09/19 08:49 | 702 11th St, San Francisco, CA 94016 |
| **97** | 176650 | Lightning Charging Cable | 1 | 14.95 | 04/12/19 16:47 | 153 River St, Boston, MA 02215 |
| **98** | 176651 | iPhone | 1 | 700 | 04/07/19 13:14 | 997 South St, Boston, MA 02215 |
| **99** | 176652 | LG Washing Machine | 1 | 600.0 | 04/09/19 20:04 | 502 14th St, New York City, NY 10001 |

100 rows × 6 columns

## Drop rows of NaN

In [6]:
```
nan_df = all_data[all_data.isna().any(axis=1)]
nan_df.head()
```

Out[6]:

| | 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 |

In [7]:
```
all_data = all_data.dropna(how='all')
all_data.head()
```

Out[7]:

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

## Task 2: Add month column

In [8]:
```
all_data = all_data[all_data['Order Date'].str[0:2] != 'Or']    # error caused due to invalid string.
```

In [9]:

```
all_data['Month'] = all_data['Order Date'].str[0:2]
all_data['Month'] = all_data['Month'].astype('int32')
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 | 4 |
| **2** | 176559 | Bose SoundSport Headphones | 1 | 99.99 | 04/07/19 22:30 | 682 Chestnut St, Boston, MA 02215 | 4 |
| **3** | 176560 | Google Phone | 1 | 600 | 04/12/19 14:38 | 669 Spruce St, Los Angeles, CA 90001 | 4 |
| **4** | 176560 | Wired Headphones | 1 | 11.99 | 04/12/19 14:38 | 669 Spruce St, Los Angeles, CA 90001 | 4 |
| **5** | 176561 | Wired Headphones | 1 | 11.99 | 04/30/19 09:27 | 333 8th St, Los Angeles, CA 90001 | 4 |

## Task 3: Add Sales column

In [10]:

```
all_data['Quantity Ordered'] = pd.to_numeric(all_data['Quantity Ordered'])       # making integer
all_data['Price Each'] = pd.to_numeric(all_data['Price Each'])             # making float
all_data.head()


all_data['Sales'] = all_data['Quantity Ordered'] * all_data['Price Each']
all_data.head()
```

Out[10]:

| | Order ID | Product | Quantity Ordered | Price Each | Order Date | Purchase Address | Month | Sales |
|---|---|---|---|---|---|---|---|---|
| **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 |

## 1: What was the best month for sale? How much was earned that month?

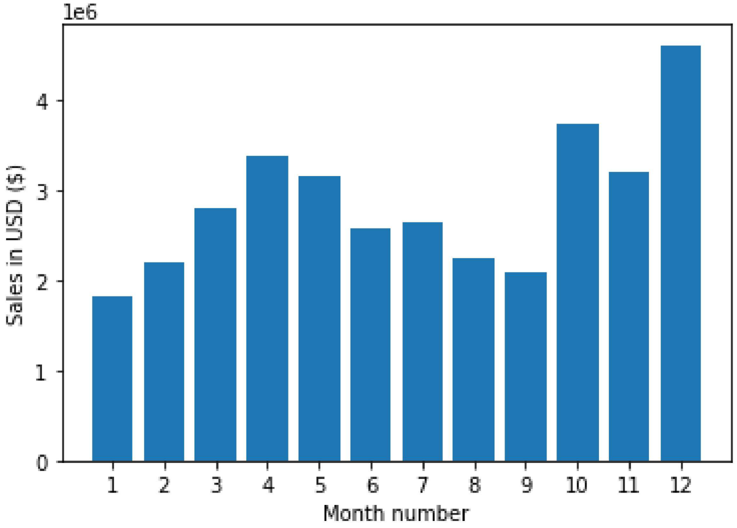
In [11]:

```
result = all_data.groupby('Month').sum()
result
```

Out[11]:

| | Quantity Ordered | Price Each | Sales |
|---|---|---|---|
| **Month** | | | |
| **1** | 10903 | 1811768.38 | 1822256.73 |
| **2** | 13449 | 2188884.72 | 2202022.42 |
| **3** | 17005 | 2791207.83 | 2807100.38 |
| **4** | 20558 | 3367671.02 | 3390670.24 |
| **5** | 18667 | 3135125.13 | 3152606.75 |
| **6** | 15253 | 2562025.61 | 2577802.26 |
| **7** | 16072 | 2632539.56 | 2647775.76 |
| **8** | 13448 | 2230345.42 | 2244467.88 |
| **9** | 13109 | 2084992.09 | 2097560.13 |
| **10** | 22703 | 3715554.83 | 3736726.88 |
| **11** | 19798 | 3180600.68 | 3199603.20 |
| **12** | 28114 | 4588415.41 | 4613443.34 |

In [12]:
```
import matplotlib.pyplot as plt

months = range(1, 13)
plt.bar(months, result['Sales'])
plt.xticks(months)
plt.ylabel('Sales in USD ($)')
plt.xlabel('Month number')
plt.show()
```



## 2: Which city had the highest no. of sales?

## Task 4: Add city column

In [13]:

```
def get_city(address):
    return address.split(',')[1]                     # function to get only city name

def get_state(address):
    return address.split(',')[2].split(' ')[1]      # function to get city name along with country code.

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

Out[13]:

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

In [14]:

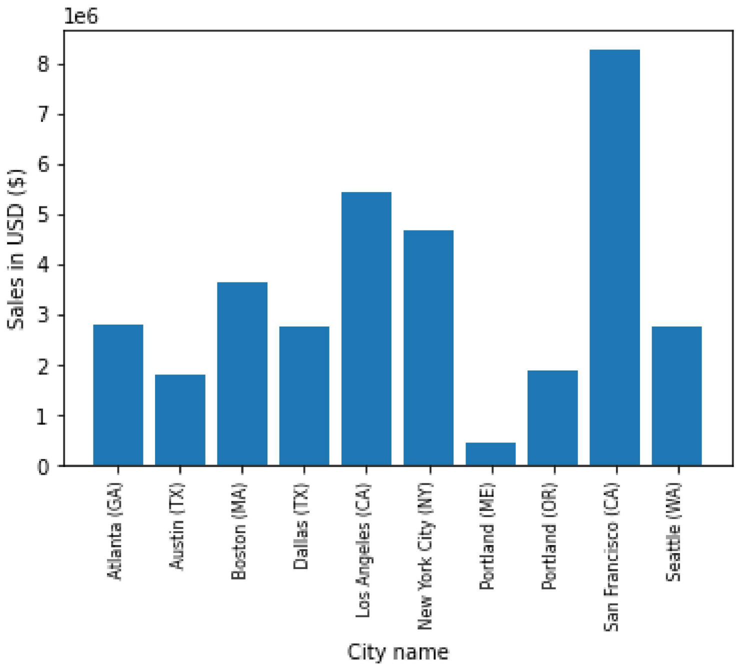
```
results = all_data.groupby('City').sum()
results
```

Out[14]:

| | Quantity Ordered | Price Each | Month | Sales |
|---|---|---|---|---|
| **City** | | | | |
| **Atlanta (GA)** | 16602 | 2779908.20 | 104794 | 2795498.58 |
| **Austin (TX)** | 11153 | 1809873.61 | 69829 | 1819581.75 |
| **Boston (MA)** | 22528 | 3637409.77 | 141112 | 3661642.01 |
| **Dallas (TX)** | 16730 | 2752627.82 | 104620 | 2767975.40 |
| **Los Angeles (CA)** | 33289 | 5421435.23 | 208325 | 5452570.80 |
| **New York City (NY)** | 27932 | 4635370.83 | 175741 | 4664317.43 |
| **Portland (ME)** | 2750 | 447189.25 | 17144 | 449758.27 |
| **Portland (OR)** | 11303 | 1860558.22 | 70621 | 1870732.34 |
| **San Francisco (CA)** | 50239 | 8211461.74 | 315520 | 8262203.91 |
| **Seattle (WA)** | 16553 | 2733296.01 | 104941 | 2747755.48 |

In [15]:

```
import matplotlib.pyplot as plt

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

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



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

```
In [16]: all_data.head()         # what data looks like.
```

Out[16]:

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

In [17]:

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

Out[17]:

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

In [18]:
```
all_data[all_data["Hour"] == all_data["Hour"].max()]
```

Out[18]:

| | Order ID | Product | Quantity Ordered | Price Each | Order Date | Purchase Address | Month | Sales | City | Hour | Minute |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **23** | 176578 | Apple Airpods Headphones | 1 | 150.00 | 2019-04-09 23:35:00 | 513 Church St, Boston, MA 02215 | 4 | 150.00 | Boston (MA) | 23 | 35 |
| **41** | 176594 | Wired Headphones | 1 | 11.99 | 2019-04-17 23:04:00 | 63 Maple St, San Francisco, CA 94016 | 4 | 11.99 | San Francisco (CA) | 23 | 4 |
| **70** | 176623 | 27in FHD Monitor | 1 | 149.99 | 2019-04-20 23:51:00 | 807 12th St, Atlanta, GA 30301 | 4 | 149.99 | Atlanta (GA) | 23 | 51 |
| **150** | 176700 | 34in Ultrawide Monitor | 1 | 379.99 | 2019-04-07 23:10:00 | 967 Walnut St, Dallas, TX 75001 | 4 | 379.99 | Dallas (TX) | 23 | 10 |
| **158** | 176708 | Wired Headphones | 1 | 11.99 | 2019-04-28 23:03:00 | 542 Ridge St, New York City, NY 10001 | 4 | 11.99 | New York City (NY) | 23 | 3 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **186787** | 259299 | Macbook Pro Laptop | 1 | 1700.00 | 2019-09-30 23:59:00 | 240 Chestnut St, Los Angeles, CA 90001 | 9 | 1700.00 | Los Angeles (CA) | 23 | 59 |
| **186816** | 259326 | AAA Batteries (4-pack) | 3 | 2.99 | 2019-09-15 23:01:00 | 163 Church St, New York City, NY 10001 | 9 | 8.97 | New York City (NY) | 23 | 1 |
| **186830** | 259339 | USB-C Charging Cable | 2 | 11.95 | 2019-09-12 23:43:00 | 509 Park St, Austin, TX 73301 | 9 | 23.90 | Austin (TX) | 23 | 43 |
| **186835** | 259344 | AAA Batteries (4-pack) | 2 | 2.99 | 2019-09-10 23:33:00 | 721 Madison St, San Francisco, CA 94016 | 9 | 5.98 | San Francisco (CA) | 23 | 33 |
| **186836** | 259345 | ThinkPad Laptop | 1 | 999.99 | 2019-09-21 23:12:00 | 406 Dogwood St, San Francisco, CA 94016 | 9 | 999.99 | San Francisco (CA) | 23 | 12 |

6275 rows × 11 columns

In [19]:

```
all_data["Hour"].value_counts()
```

Out[19]:

```
19    12905
12    12587
11    12411
18    12280
20    12228
13    12129
14    10984
10    10944
21    10921
17    10899
16    10384
15    10175
22     8822
9      8748
23     6275
8      6256
7      4011
0      3910
6      2482
1      2350
5      1321
2      1243
4       854
3       831
Name: Hour, dtype: int64
```

## 4: what product are most often sold together?

In [20]:
```
all_data.head()
```

Out[20]:

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

In [21]:

```
df = all_data[all_data['Order ID'].duplicated(keep=False)]     # to find the products which have same order ID.
df.head(10)
```

Out[21]:

|  | Order ID | Product | Quantity Ordered | Price Each | Order Date | Purchase Address | Month | Sales | City | Hour | Minute |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **3** | 176560 | Google Phone | 1 | 600.00 | 2019-04-12 14:38:00 | 669 Spruce St, Los Angeles, CA 90001 | 4 | 600.00 | Los Angeles (CA) | 14 | 38 |
| **4** | 176560 | Wired Headphones | 1 | 11.99 | 2019-04-12 14:38:00 | 669 Spruce St, Los Angeles, CA 90001 | 4 | 11.99 | Los Angeles (CA) | 14 | 38 |
| **18** | 176574 | Google Phone | 1 | 600.00 | 2019-04-03 19:42:00 | 20 Hill St, Los Angeles, CA 90001 | 4 | 600.00 | Los Angeles (CA) | 19 | 42 |
| **19** | 176574 | USB-C Charging Cable | 1 | 11.95 | 2019-04-03 19:42:00 | 20 Hill St, Los Angeles, CA 90001 | 4 | 11.95 | Los Angeles (CA) | 19 | 42 |
| **30** | 176585 | Bose SoundSport Headphones | 1 | 99.99 | 2019-04-07 11:31:00 | 823 Highland St, Boston, MA 02215 | 4 | 99.99 | Boston (MA) | 11 | 31 |
| **31** | 176585 | Bose SoundSport Headphones | 1 | 99.99 | 2019-04-07 11:31:00 | 823 Highland St, Boston, MA 02215 | 4 | 99.99 | Boston (MA) | 11 | 31 |
| **32** | 176586 | AAA Batteries (4-pack) | 2 | 2.99 | 2019-04-10 17:00:00 | 365 Center St, San Francisco, CA 94016 | 4 | 5.98 | San Francisco (CA) | 17 | 0 |
| **33** | 176586 | Google Phone | 1 | 600.00 | 2019-04-10 17:00:00 | 365 Center St, San Francisco, CA 94016 | 4 | 600.00 | San Francisco (CA) | 17 | 0 |
| **119** | 176672 | Lightning Charging Cable | 1 | 14.95 | 2019-04-12 11:07:00 | 778 Maple St, New York City, NY 10001 | 4 | 14.95 | New York City (NY) | 11 | 7 |
| **120** | 176672 | USB-C Charging Cable | 1 | 11.95 | 2019-04-12 11:07:00 | 778 Maple St, New York City, NY 10001 | 4 | 11.95 | New York City (NY) | 11 | 7 |

In [22]:

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

df['Grouped'] = df.groupby('Order ID')['Product'].transform(lambda x: ','.join(x))    # Creeating a new column t
df = df[['Order ID', 'Grouped']].drop_duplicates()                                    # ID products in one colum
df.head(10)
```

```
C:\Users\Shashank\AppData\Local\Temp\ipykernel_8764\727416753.py:3: 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#re
turning-a-view-versus-a-copy (https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-
a-view-versus-a-copy)
  df['Grouped'] = df.groupby('Order ID')['Product'].transform(lambda x: ','.join(x))    # Creeating a new colu
mn to get same order
```

Out[22]:

| | 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 |
| **129** | 176681 | Apple Airpods Headphones,ThinkPad Laptop |
| **138** | 176689 | Bose SoundSport Headphones,AAA Batteries (4-pack) |
| **189** | 176739 | 34in Ultrawide Monitor,Google Phone |
| **225** | 176774 | Lightning Charging Cable,USB-C Charging Cable |
| **233** | 176781 | iPhone,Lightning Charging Cable |

In [23]:

```
from itertools import combinations
from collections import Counter
                                       # snippet to get the products which are ordered together. (2 products at
count = Counter()

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

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

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

In [24]:

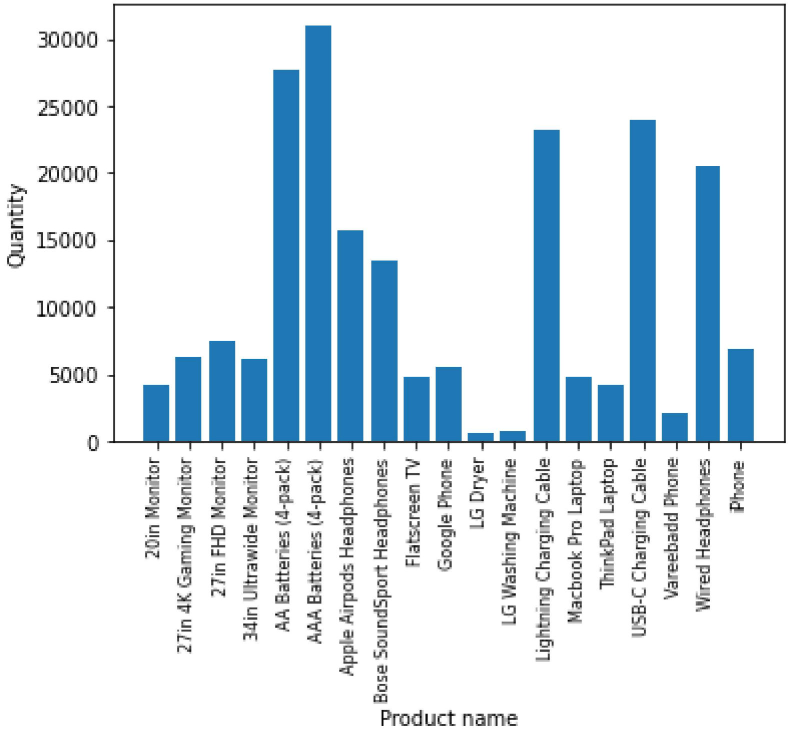
```
from itertools import combinations
from collections import Counter
                                       # snippet to get the products which are ordered together. (3 products at
count = Counter()

for row in df['Grouped']:
    row_list = row.split(',')
    count.update(Counter(combinations(row_list, 3)))

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

```
('Google Phone', 'USB-C Charging Cable', 'Wired Headphones') 87
('iPhone', 'Lightning Charging Cable', 'Wired Headphones') 62
('iPhone', 'Lightning Charging Cable', 'Apple Airpods Headphones') 47
('Google Phone', 'USB-C Charging Cable', 'Bose SoundSport Headphones') 35
('Vareebadd Phone', 'USB-C Charging Cable', 'Wired Headphones') 33
('iPhone', 'Apple Airpods Headphones', 'Wired Headphones') 27
('Google Phone', 'Bose SoundSport Headphones', 'Wired Headphones') 24
('Vareebadd Phone', 'USB-C Charging Cable', 'Bose SoundSport Headphones') 16
('USB-C Charging Cable', 'Bose SoundSport Headphones', 'Wired Headphones') 5
('Vareebadd Phone', 'Bose SoundSport Headphones', 'Wired Headphones') 5
```

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

In [27]:

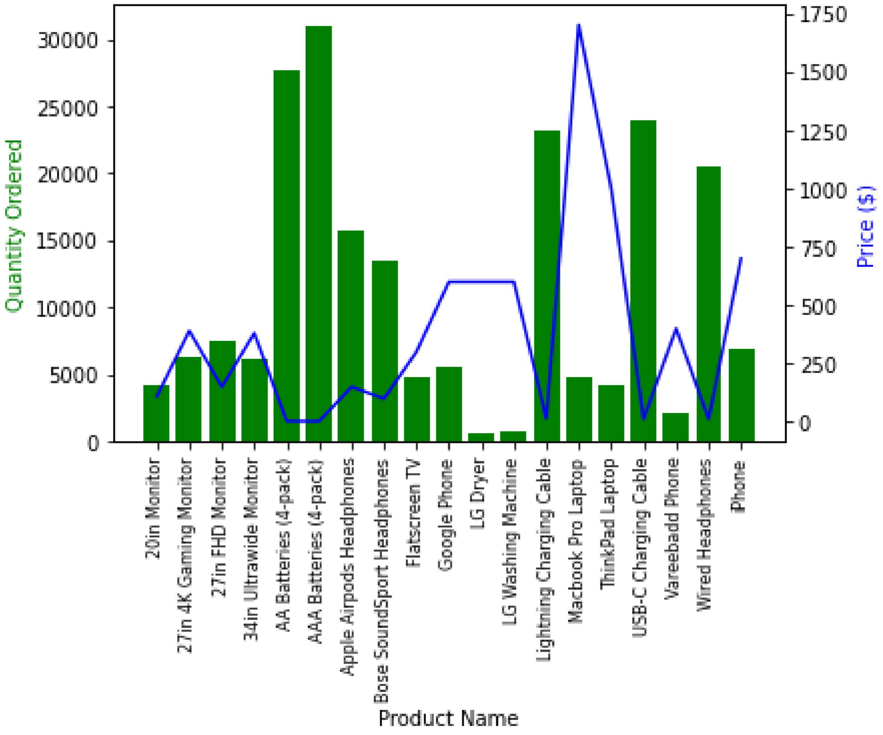
```
product_group = all_data.groupby('Product')
quantity_ordered = product_group.sum()['Quantity Ordered']

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



In [28]:

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

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

fig, ax1 = plt.subplots()

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

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

fig.show()
```

```
C:\Users\Shashank\AppData\Local\Temp\ipykernel_8764\136096346.py:14: UserWarning: FixedFormatter should only b
e used together with FixedLocator
  ax1.set_xticklabels(keys, rotation='vertical', size=8)
C:\Users\Shashank\AppData\Local\Temp\ipykernel_8764\136096346.py:16: UserWarning: Matplotlib is currently usin
g module://matplotlib_inline.backend_inline, which is a non-GUI backend, so cannot show the figure.
  fig.show()
```



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

Out[30]:

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

```
In [41]: all_data['Product'].value_counts()
```

```
Out[41]: USB-C Charging Cable          21903
         Lightning Charging Cable      21658
         AAA Batteries (4-pack)        20641
         AA Batteries (4-pack)         20577
         Wired Headphones              18882
         Apple Airpods Headphones      15549
         Bose SoundSport Headphones    13325
         27in FHD Monitor               7507
         iPhone                         6842
         27in 4K Gaming Monitor         6230
         34in Ultrawide Monitor         6181
         Google Phone                   5525
         Flatscreen TV                  4800
         Macbook Pro Laptop             4724
         ThinkPad Laptop                4128
         20in Monitor                   4101
         Vareebadd Phone                2065
         LG Washing Machine              666
         LG Dryer                        646
         Name: Product, dtype: int64
```

```
In [ ]:
```