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

In [3]:
```
all_data=pd.read_csv("all_data.csv")
```

In [4]:
```
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 8:46	917 1st St, Dallas, TX 75001
1	NaN	NaN	NaN	NaN	NaN	NaN
2	176559	Bose SoundSport Headphones	1	99.99	04/07/19 22:30	682 Chestnut St, Boston, MA 02215
3	176560	Google Phone	1	600	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001
4	176560	Wired Headphones	1	11.99	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001

Clean up the data

Drop rows of NAN

In [5]:
```
# Find NAN
nan_df = all_data[all_data.isna().any(axis=1)]
display(nan_df.head())

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

	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

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 8:46	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 9:27	333 8th St, Los Angeles, CA 90001

Get rid of text in order date column¶

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

Make columns correct type

In [7]:
```
all_data['Quantity Ordered'] = pd.to_numeric(all_data['Quantity Ordered'])
all_data['Price Each'] = pd.to_numeric(all_data['Price Each'])
```

Augment data with additional columns

In [8]:
```
all_data['Month'] = all_data['Order Date'].str[0:2]
all_data['Month'] = all_data['Month'].astype('int32')
all_data.head()
```

Out[8]:

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

Add month column (alternative method)

In [9]:
```
all_data['Month 2'] = pd.to_datetime(all_data['Order Date']).dt.month
all_data.head()
```

Out[9]:

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

Add city column

In [10]:
```
def get_city(address):
    return address.split(",")[1].strip(" ")

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

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

Out[10]:

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

Data Exploration!

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

In [11]:
```
all_data['Sales'] = all_data['Quantity Ordered'].astype('int') * all_data['Price Each'].astype('float')
```

In [12]:
```
all_data.groupby(['Month']).sum()
```

Out[12]:

Month	Quantity Ordered	Price Each	Month 2	Sales
4	17739	2899439.68	63088	2918954.40
5	26	8851.62	125	8855.46

Question 2: What city sold the most product?

In [13]:
```
city_max=all_data.groupby(['City']).sum()
print(max(city_max))
```

```
Sales
```

Question 4: What products are most often sold together?

In [14]:
```
df = all_data[all_data['Order ID'].duplicated(keep=False)]

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

```
3                        Google Phone,Wired Headphones
4                        Google Phone,Wired Headphones
18                    Google Phone,USB-C Charging Cable
19                    Google Phone,USB-C Charging Cable
30       Bose SoundSport Headphones,Bose SoundSport Hea...
31       Bose SoundSport Headphones,Bose SoundSport Hea...
32                     AAA Batteries (4-pack),Google Phone
33                     AAA Batteries (4-pack),Google Phone
119         Lightning Charging Cable,USB-C Charging Cable
120         Lightning Charging Cable,USB-C Charging Cable
129            Apple Airpods Headphones,ThinkPad Laptop
130            Apple Airpods Headphones,ThinkPad Laptop
138      Bose SoundSport Headphones,AAA Batteries (4-pack)
139      Bose SoundSport Headphones,AAA Batteries (4-pack)
189                 34in Ultrawide Monitor,Google Phone
190                 34in Ultrawide Monitor,Google Phone
225         Lightning Charging Cable,USB-C Charging Cable
226         Lightning Charging Cable,USB-C Charging Cable
233                        iPhone,Lightning Charging Cable
234                        iPhone,Lightning Charging Cable
250      Google Phone,Bose SoundSport Headphones,Wired ...
251      Google Phone,Bose SoundSport Headphones,Wired ...
252      Google Phone,Bose SoundSport Headphones,Wired ...
260                       Google Phone,USB-C Charging Cable
261                       Google Phone,USB-C Charging Cable
264                          Google Phone,Wired Headphones
265                          Google Phone,Wired Headphones
270                          Google Phone,Wired Headphones
271                          Google Phone,Wired Headphones
394              AAA Batteries (4-pack),27in FHD Monitor
                                ...
15525       AA Batteries (4-pack),Lightning Charging Cable
15577                     Google Phone,USB-C Charging Cable
15578                     Google Phone,USB-C Charging Cable
15591    iPhone,Lightning Charging Cable,Apple Airpods ...
15592    iPhone,Lightning Charging Cable,Apple Airpods ...
15593    iPhone,Lightning Charging Cable,Apple Airpods ...
15609          AA Batteries (4-pack),AA Batteries (4-pack)
15610          AA Batteries (4-pack),AA Batteries (4-pack)
15614                        iPhone,Wired Headphones
15615                        iPhone,Wired Headphones
15659                     Google Phone,USB-C Charging Cable
15660                     Google Phone,USB-C Charging Cable
15675          USB-C Charging Cable,Apple Airpods Headphones
15676          USB-C Charging Cable,Apple Airpods Headphones
15702                     Google Phone,USB-C Charging Cable
15703                     Google Phone,USB-C Charging Cable
15712                        34in Ultrawide Monitor,iPhone
15713                        34in Ultrawide Monitor,iPhone
15727    Bose SoundSport Headphones,AAA Batteries (4-pack)
15728    Bose SoundSport Headphones,AAA Batteries (4-pack)
15775                     Google Phone,USB-C Charging Cable
15776                     Google Phone,USB-C Charging Cable
15778          AAA Batteries (4-pack),AA Batteries (4-pack)
15779          AAA Batteries (4-pack),AA Batteries (4-pack)
15786                  USB-C Charging Cable,Wired Headphones
15787                  USB-C Charging Cable,Wired Headphones
15818             Vareebadd Phone,Lightning Charging Cable
15819             Vareebadd Phone,Lightning Charging Cable
15874              Google Phone,Bose SoundSport Headphones
15875              Google Phone,Bose SoundSport Headphones
Name: Grouped, Length: 1269, dtype: object
```

```
C:\Users\student\Anaconda3\lib\site-packages\ipykernel_launcher.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy
  after removing the cwd from sys.path.
```

In [15]:
```
from itertools import combinations
from collections import Counter

count = Counter()

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

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

```
('iPhone', 'Lightning Charging Cable') 94
('Google Phone', 'USB-C Charging Cable') 92
('Google Phone', 'Wired Headphones') 34
('iPhone', 'Wired Headphones') 33
('Vareebadd Phone', 'USB-C Charging Cable') 32
('iPhone', 'Apple Airpods Headphones') 29
('Google Phone', 'Bose SoundSport Headphones') 20
('Vareebadd Phone', 'Wired Headphones') 15
('USB-C Charging Cable', 'Wired Headphones') 11
('AA Batteries (4-pack)', 'Apple Airpods Headphones') 7
```

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

In [16]:
```
product_group = all_data.groupby('Product')
quantity_ordered = product_group.sum()['Quantity Ordered']
```

In [17]:
```
print(quantity_ordered)
```

```
Product
20in Monitor                 345
27in 4K Gaming Monitor       491
27in FHD Monitor             633
34in Ultrawide Monitor       563
AA Batteries (4-pack)       2446
AAA Batteries (4-pack)      2559
Apple Airpods Headphones    1303
Bose SoundSport Headphones  1110
Flatscreen TV                398
Google Phone                 497
LG Dryer                      69
LG Washing Machine            56
Lightning Charging Cable    2027
Macbook Pro Laptop           400
ThinkPad Laptop              329
USB-C Charging Cable        1938
Vareebadd Phone              185
Wired Headphones            1823
iPhone                       593
Name: Quantity Ordered, dtype: int64
```

In [18]:
```
prices = all_data.groupby('Product').mean()['Price Each']
```

In [19]:
```
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
```

In [ ]:
```
```