Analyse des ventes de gadgets

December 4, 2021

1 Découverte des données

[62]:
```
import os #Manipuler les fichiers
import pandas as pd #Traiter les données
import numpy as np
import matplotlib.pyplot as plt #Visualisation des données
import seaborn as sns
```

[81]:
```
files =  [file for file in os.listdir(r'C:/Users/boudi/Documents/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
```

[82]:
```
path =r'C:/Users/boudi/Documents/Sales_Data'
print(path)
#Création de la base de données vide
all_data = pd.DataFrame()
print(all_data)
```

```
C:/Users/boudi/Documents/Sales_Data
Empty DataFrame
Columns: []
Index: []
```

[83]:
```
#Alimenter la bdd
for file in files :
    current_file = pd.read_csv(path+'/'+file)
```

```
#     print(current_file)
    all_data = pd.concat([all_data,current_file])
print(all_data)
```

```
      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]
```

[84]:
```
all_data.to_csv(path+'/all_data.csv',index=False)
```

[85]:
```
all_data.dtypes
```

[85]:
```
Order ID            object
Product             object
Quantity Ordered    object
Price Each          object
Order Date          object
Purchase Address    object
dtype: object
```

[86]:
```
all_data.head()
```

[86]:
```
  Order ID                    Product Quantity Ordered Price Each  \
0   176558        USB-C Charging Cable                2      11.95
```

```
1      NaN                            NaN              NaN        NaN
2   176559  Bose SoundSport Headphones                    1      99.99
3   176560                Google Phone                    1        600
4   176560             Wired Headphones                   1      11.99

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

[87]:
```
all_data.isnull().sum()
```

[87]:
```
Order ID            545
Product             545
Quantity Ordered    545
Price Each          545
Order Date          545
Purchase Address    545
dtype: int64
```

[88]:
```
all_data = all_data.dropna(how='all')
all_data.shape
```

[88]:
```
(186305, 6)
```

[89]:
```
all_data.head()
```

[89]:
```
  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
```

2 Quel est le mois ou on a réalisée le meilleur chiffre d'affaire ?

[91]:
```
def mouth(x):
    return x.split('/')[0]
mouth('04/19/19 ')
```

[91]:
```
'04'
```

[92]:
```
all_data['mounth'] = all_data['Order Date'].apply(mouth)
all_data
```

[92]:

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	mounth
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
…	…	…	…	…	…	…	…
11681	259353	AAA Batteries (4-pack)	3	2.99	09/17/19 20:56	840 Highland St, Los Angeles, CA 90001	09
11682	259354	iPhone	1	700	09/01/19 16:00	216 Dogwood St, San Francisco, CA 94016	09
11683	259355	iPhone	1	700	09/23/19 07:39	220 12th St, San Francisco, CA 94016	09
11684	259356	34in Ultrawide Monitor	1	379.99	09/19/19 17:30	511 Forest St, San Francisco, CA 94016	09
11685	259357	USB-C Charging Cable	1	11.95	09/30/19 00:18	250 Meadow St, San Francisco, CA 94016	09

[186305 rows x 7 columns]

[105]:
```
all_data['Price Each'] = all_data['Price Each'].astype(float)
all_data['Quantity Ordered'] = all_data['Quantity Ordered'].astype(int)
```

```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
<ipython-input-105-c93377c265cd> in <module>
----> 1 all_data['Price Each'] = all_data['Price Each'].astype(float)
      2 all_data['Quantity Ordered'] = all_data['Quantity Ordered'].astype(int)
```

```
~\anaconda3\lib\site-packages\pandas\core\generic.py in astype(self, dtype,␣
↪copy, errors)
   5875         else:
   5876             # else, only a single dtype is given
-> 5877             new_data = self._mgr.astype(dtype=dtype, copy=copy,␣
↪errors=errors)
   5878             return self._constructor(new_data).__finalize__(self,␣
↪method="astype")
   5879

~\anaconda3\lib\site-packages\pandas\core\internals\managers.py in astype(self,␣
↪dtype, copy, errors)
    629         self, dtype, copy: bool = False, errors: str = "raise"
    630     ) -> "BlockManager":
--> 631         return self.apply("astype", dtype=dtype, copy=copy,␣
↪errors=errors)
    632
    633     def convert(

~\anaconda3\lib\site-packages\pandas\core\internals\managers.py in apply(self,␣
↪f, align_keys, ignore_failures, **kwargs)
    425                     applied = b.apply(f, **kwargs)
    426                 else:
--> 427                     applied = getattr(b, f)(**kwargs)
    428             except (TypeError, NotImplementedError):
    429                 if not ignore_failures:

~\anaconda3\lib\site-packages\pandas\core\internals\blocks.py in astype(self,␣
↪dtype, copy, errors)
    671             vals1d = values.ravel()
    672             try:
--> 673                 values = astype_nansafe(vals1d, dtype, copy=True)
    674             except (ValueError, TypeError):
    675                 # e.g. astype_nansafe can fail on object-dtype of strin s

~\anaconda3\lib\site-packages\pandas\core\dtypes\cast.py in astype_nansafe(arr,
↪dtype, copy, skipna)
   1095     if copy or is_object_dtype(arr) or is_object_dtype(dtype):
   1096         # Explicit copy, or required since NumPy can't view from / to␣
↪object.
-> 1097         return arr.astype(dtype, copy=True)
   1098
   1099     return arr.view(dtype)

ValueError: could not convert string to float: 'Price Each'
```

[104]:
```
all_data.dtypes
```

[104]:
```
Order ID            object
Product             object
Quantity Ordered    object
Price Each          object
Order Date          object
Purchase Address    object
mounth              object
dtype: object
```

[107]:
```
all_data['Quantity Ordered'].unique()
```

[107]:
```
array(['2', '1', '3', '5', 'Quantity Ordered', '4', '7', '6', '8', '9'],
      dtype=object)
```

[110]:
```
all_data = all_data[all_data['Quantity Ordered']!=  'Quantity Ordered']
all_data['Quantity Ordered'].unique()
```

[110]:
```
array(['2', '1', '3', '5', '4', '7', '6', '8', '9'], dtype=object)
```

[111]:
```
all_data['Quantity Ordered'] = all_data['Quantity Ordered'].astype(int)
all_data.dtypes
```

```
<ipython-input-111-6b0cd9762d1c>: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
  all_data['Quantity Ordered'] = all_data['Quantity Ordered'].astype(int)
```

[111]:
```
Order ID            object
Product             object
Quantity Ordered     int32
Price Each          object
Order Date          object
Purchase Address    object
mounth              object
dtype: object
```

[113]:
```
all_data['mounth'].unique()
```

[113]:
```
array(['04', '05', '08', '09', '12', '01', '02', '03', '07', '06', '11',
       '10'], dtype=object)
```

[114]:
```
all_data['mounth'] = all_data['mounth'].astype(int)
```

```
<ipython-input-114-5e202ed0772c>: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
  all_data['mounth'] = all_data['mounth'].astype(int)
```

[115]:
```
all_data['Price Each'].unique()
```

[115]:
```
array(['11.95', '99.99', '600', '11.99', '1700', '14.95', '389.99',
       '3.84', '150', '2.99', '700', '300', '149.99', '109.99', '600.0',
       '999.99', '400', '379.99', '700.0', '1700.0', '150.0', '300.0',
       '400.0'], dtype=object)
```

[116]:
```
all_data['Price Each'] = all_data['Price Each'].astype(float)
```

```
<ipython-input-116-f28fec3bc38b>: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
  all_data['Price Each'] = all_data['Price Each'].astype(float)
```

[117]:
```
all_data.dtypes
```

[117]:
```
Order ID             object
Product              object
Quantity Ordered      int32
Price Each          float64
Order Date           object
Purchase Address     object
mounth                int32
dtype: object
```

[118]:
```
all_data['Sales'] = all_data['Quantity Ordered']*all_data['Price Each']
```

```
<ipython-input-118-d28a9cd72cf6>: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
  all_data['Sales'] = all_data['Quantity Ordered']*all_data['Price Each']
```

[119]:
```
all_data
```

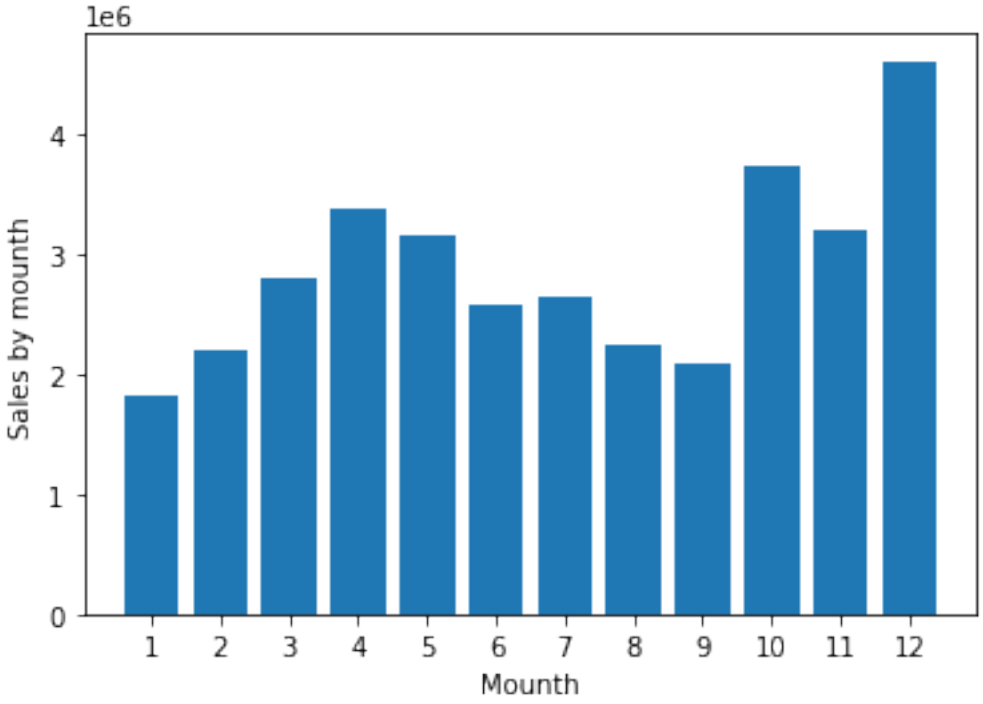
[119]:
```
      Order ID                    Product  Quantity Ordered  Price Each  \
0       176558        USB-C Charging Cable                 2       11.95
2       176559  Bose SoundSport Headphones                 1       99.99
3       176560                Google Phone                 1      600.00
4       176560            Wired Headphones                 1       11.99
5       176561            Wired Headphones                 1       11.99
…          …                           …                 …         …
11681   259353      AAA Batteries (4-pack)                 3        2.99
11682   259354                      iPhone                 1      700.00
11683   259355                      iPhone                 1      700.00
11684   259356      34in Ultrawide Monitor                 1      379.99
11685   259357        USB-C Charging Cable                 1       11.95

           Order Date                      Purchase Address  mounth   Sales
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
…                 …                                       …      …      …
11681  09/17/19 20:56   840 Highland St, Los Angeles, CA 90001       9    8.97
11682  09/01/19 16:00  216 Dogwood St, San Francisco, CA 94016       9  700.00
11683  09/23/19 07:39     220 12th St, San Francisco, CA 94016       9  700.00
11684  09/19/19 17:30   511 Forest St, San Francisco, CA 94016       9  379.99
11685  09/30/19 00:18   250 Meadow St, San Francisco, CA 94016       9   11.95

[185950 rows x 8 columns]
```

[120]:
```
all_data.groupby('mounth')['Sales'].sum()
```

[120]:
```
mounth
1     1.822257e+06
2     2.202022e+06
3     2.807100e+06
4     3.390670e+06
5     3.152607e+06
6     2.577802e+06
7     2.647776e+06
8     2.244468e+06
9     2.097560e+06
10    3.736727e+06
11    3.199603e+06
12    4.613443e+06
Name: Sales, dtype: float64
```

[121]:
```
mounth = range(1,13)
```

[122]:
```
plt.bar(mounth,all_data.groupby('mounth')['Sales'].sum())
plt.xticks(mounth)
plt.xlabel('Mounth')
plt.ylabel('Sales by mounth')
plt.show()
```



3 Quelle est la ville qui réalise le plus de vente ?

[123]:
```
def city(x):
    return x.split(',')[1]
all_data['City'] = all_data['Purchase Address'].apply(city)
all_data
```

```
<ipython-input-123-6822252b3cc8>: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#returning-a-view-versus-a-copy
  all_data['City'] = all_data['Purchase Address'].apply(city)
```

[123]:

	Order ID	Product	Quantity Ordered	Price Each	\
0	176558	USB-C Charging Cable	2	11.95	

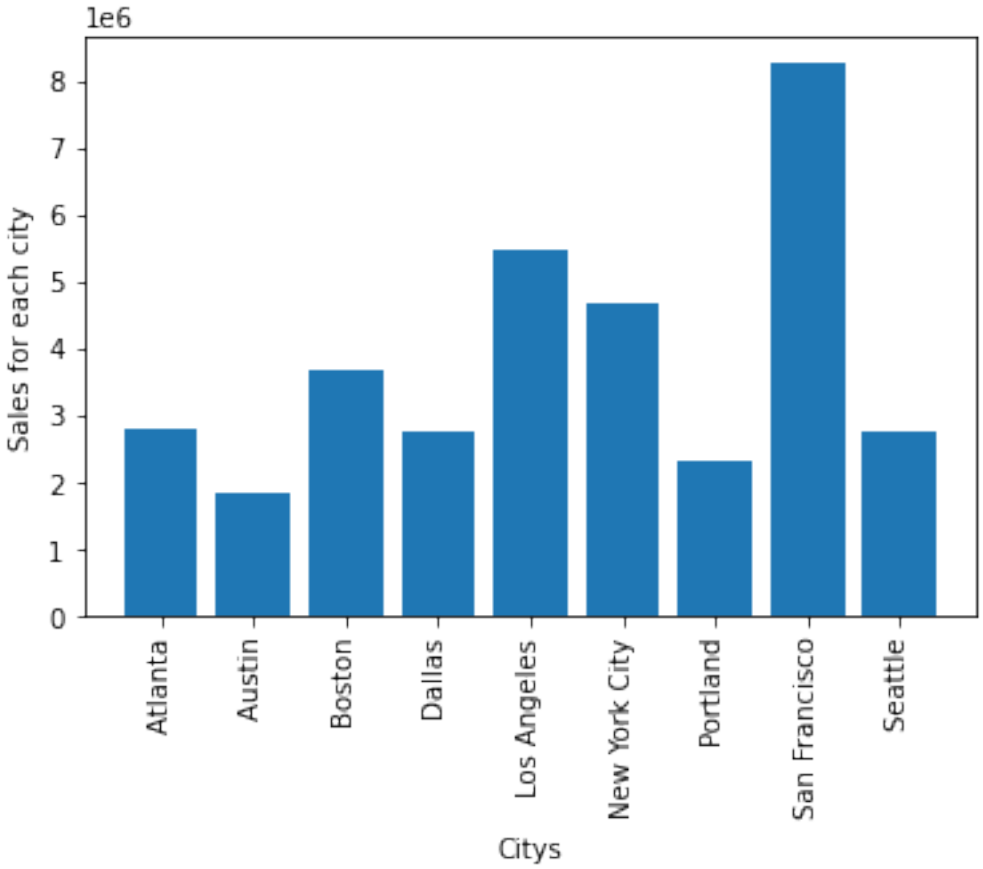
```
2       176559  Bose SoundSport Headphones                1       99.99
3       176560                Google Phone                1      600.00
4       176560            Wired Headphones                1       11.99
5       176561            Wired Headphones                1       11.99
…          …                           …                …          …
11681   259353      AAA Batteries (4-pack)                3        2.99
11682   259354                      iPhone                1      700.00
11683   259355                      iPhone                1      700.00
11684   259356      34in Ultrawide Monitor                1      379.99
11685   259357        USB-C Charging Cable                1       11.95

           Order Date                        Purchase Address  mounth  \
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
…                 …                                       …     …
11681  09/17/19 20:56   840 Highland St, Los Angeles, CA 90001       9
11682  09/01/19 16:00  216 Dogwood St, San Francisco, CA 94016       9
11683  09/23/19 07:39     220 12th St, San Francisco, CA 94016       9
11684  09/19/19 17:30   511 Forest St, San Francisco, CA 94016       9
11685  09/30/19 00:18   250 Meadow St, San Francisco, CA 94016       9

        Sales            City
0       23.90          Dallas
2       99.99          Boston
3      600.00     Los Angeles
4       11.99     Los Angeles
5       11.99     Los Angeles
…         …               …
11681    8.97     Los Angeles
11682  700.00   San Francisco
11683  700.00   San Francisco
11684  379.99   San Francisco
11685   11.95   San Francisco

[185950 rows x 9 columns]
```

[127]:
```
all_data.groupby('City')['Sales'].sum()
```

[127]:
```
City
 Atlanta          2.795499e+06
 Austin           1.819582e+06
 Boston           3.661642e+06
 Dallas           2.767975e+06
 Los Angeles      5.452571e+06
```

```
 New York City    4.664317e+06
 Portland         2.320491e+06
 San Francisco    8.262204e+06
 Seattle          2.747755e+06
Name: Sales, dtype: float64
```

[134]:
```
plt.bar(all_data.groupby('City')['Sales'].sum().index,all_data.
 ↪groupby('City')['Sales'].sum().values)
plt.xticks(rotation='vertical')
plt.xlabel('Citys')
plt.ylabel('Sales for each city')
plt.show()
```



[135]:
```
all_data['Order Date']
```

[135]:
```
0        04/19/19 08:46
2        04/07/19 22:30
3        04/12/19 14:38
4        04/12/19 14:38
```

```
5        04/30/19 09:27
              …
11681    09/17/19 20:56
11682    09/01/19 16:00
11683    09/23/19 07:39
11684    09/19/19 17:30
11685    09/30/19 00:18
Name: Order Date, Length: 185950, dtype: object
```

4 En quel moment peut faire une compagne publicitaire pour avoir plus de commandes ?

[137]:
```
all_data['hour'] = pd.to_datetime(all_data['Order Date']).dt.hour
```

```
<ipython-input-137-41150794c191>: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
  all_data['hour'] = pd.to_datetime(all_data['Order Date']).dt.hour
```

[138]:
```
all_data
```

[138]:

	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	
…	…	…	…	…	
11681	259353	AAA Batteries (4-pack)	3	2.99	
11682	259354	iPhone	1	700.00	
11683	259355	iPhone	1	700.00	
11684	259356	34in Ultrawide Monitor	1	379.99	
11685	259357	USB-C Charging Cable	1	11.95	

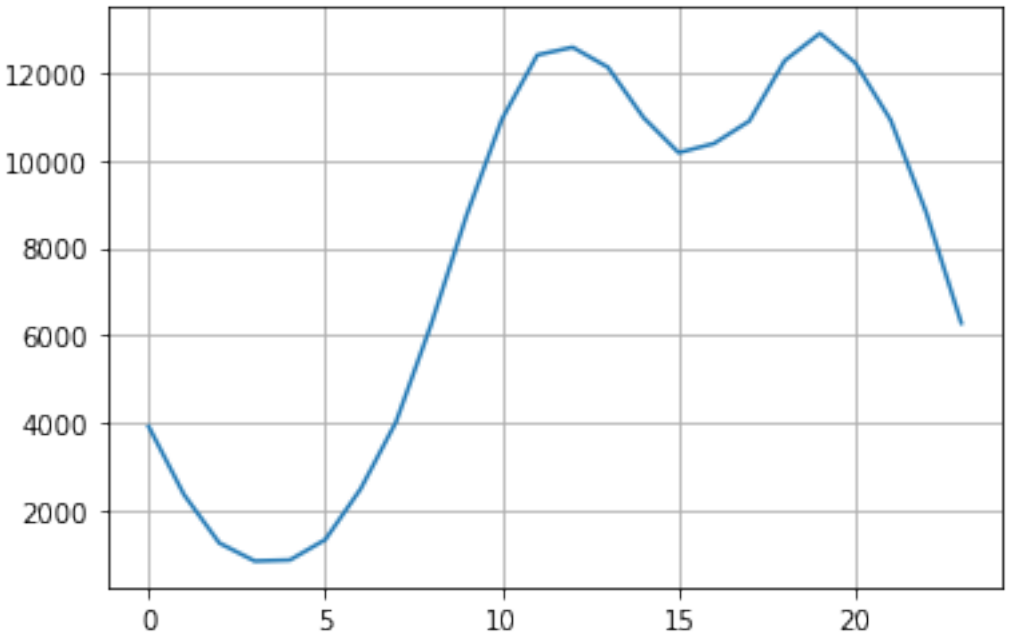
	Order Date	Purchase Address	mounth	\
0	04/19/19 08:46	917 1st St, Dallas, TX 75001	4	
2	04/07/19 22:30	682 Chestnut St, Boston, MA 02215	4	
3	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001	4	
4	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001	4	
5	04/30/19 09:27	333 8th St, Los Angeles, CA 90001	4	
…	…	…	…	
11681	09/17/19 20:56	840 Highland St, Los Angeles, CA 90001	9	
11682	09/01/19 16:00	216 Dogwood St, San Francisco, CA 94016	9	

```
11683  09/23/19 07:39     220 12th St, San Francisco, CA 94016       9
11684  09/19/19 17:30   511 Forest St, San Francisco, CA 94016       9
11685  09/30/19 00:18   250 Meadow St, San Francisco, CA 94016       9

        Sales            City  hour
0       23.90          Dallas     8
2       99.99          Boston    22
3      600.00     Los Angeles    14
4       11.99     Los Angeles    14
5       11.99     Los Angeles     9
…         …              …    …
11681    8.97     Los Angeles    20
11682  700.00   San Francisco    16
11683  700.00   San Francisco     7
11684  379.99   San Francisco    17
11685   11.95   San Francisco     0

[185950 rows x 10 columns]
```

[139]:
```
all_data.groupby('hour')['Quantity Ordered'].count()
```

[139]:
```
hour
0      3910
1      2350
2      1243
3       831
4       854
5      1321
6      2482
7      4011
8      6256
9      8748
10    10944
11    12411
12    12587
13    12129
14    10984
15    10175
16    10384
17    10899
18    12280
19    12905
20    12228
21    10921
22     8822
23     6275
Name: Quantity Ordered, dtype: int64
```

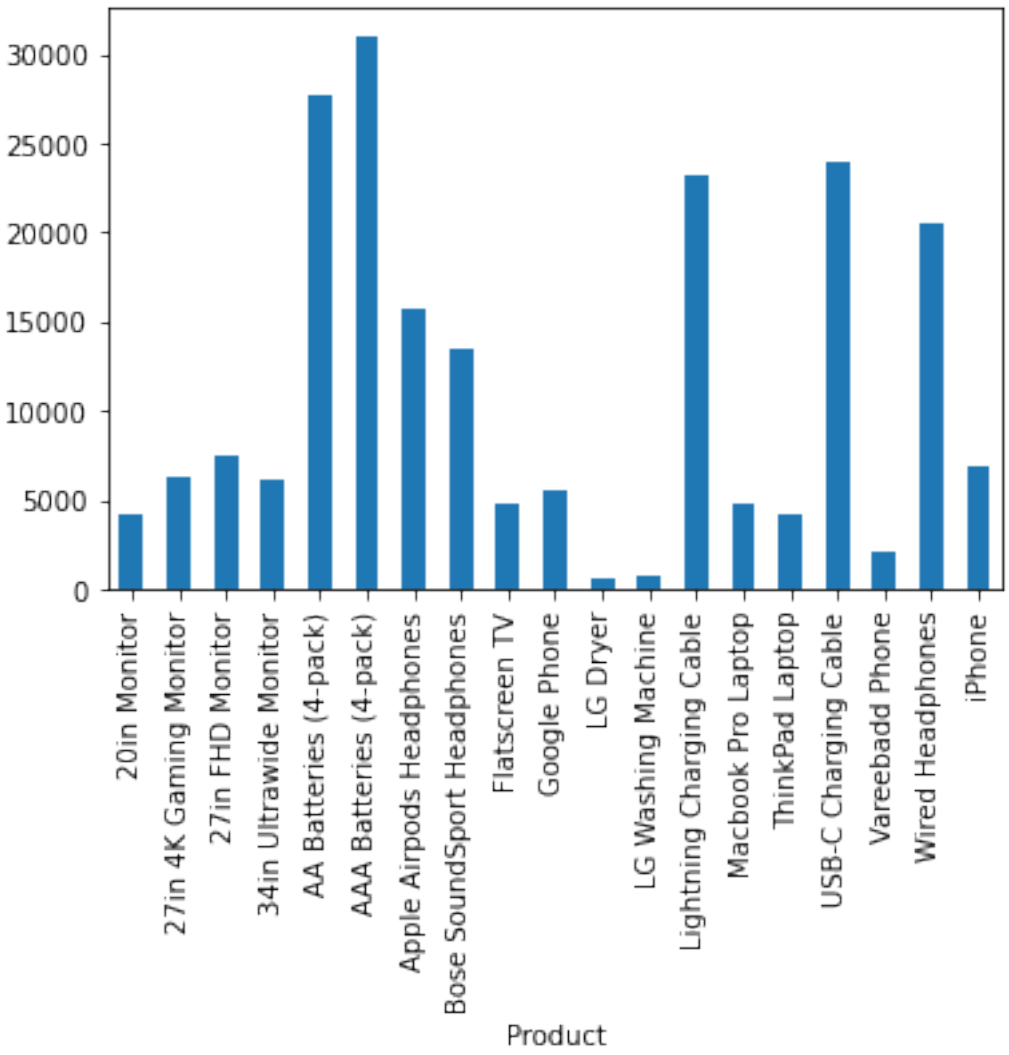
[141]:
```
plt.grid()
plt.plot(all_data.groupby('hour')['Quantity Ordered'].count().index,all_data.
 ↪groupby('hour')['Quantity Ordered'].count().values)
plt.show()
```



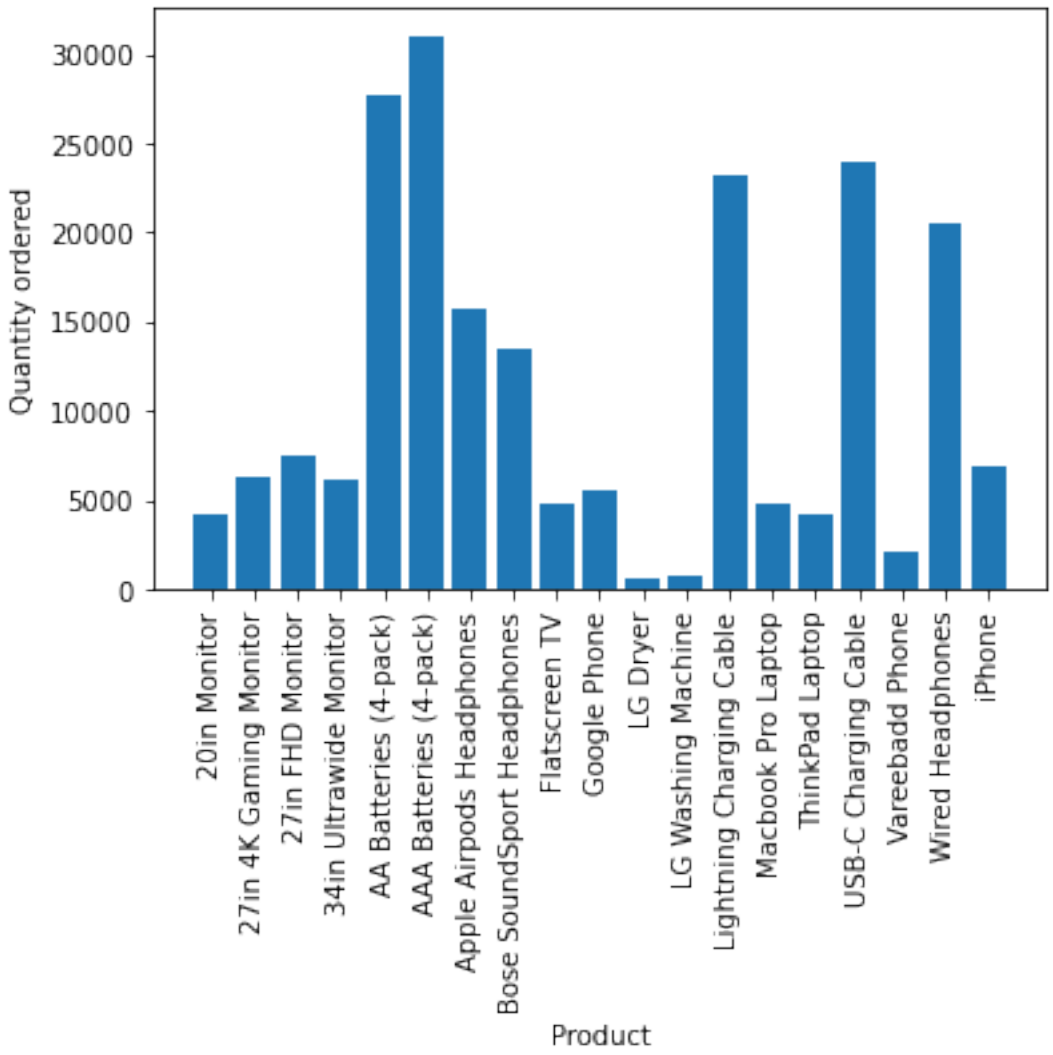
5 Quel est le produit qui réalise le plus de vente ?

[149]:
```
all_data.groupby('Product')['Quantity Ordered'].sum().plot(kind='bar')
```

[149]:
```
<AxesSubplot:xlabel='Product'>
```



[150]:
```
plt.bar(all_data.groupby('Product')['Quantity Ordered'].sum().index,all_data.
 ↪groupby('Product')['Quantity Ordered'].sum().values)
plt.xlabel('Product')
plt.xticks(rotation='vertical')
plt.ylabel('Quantity ordered')
plt.show()
```



[151]:
```
all_data.groupby('Product')['Price Each'].mean()
```

[151]:
```
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
```

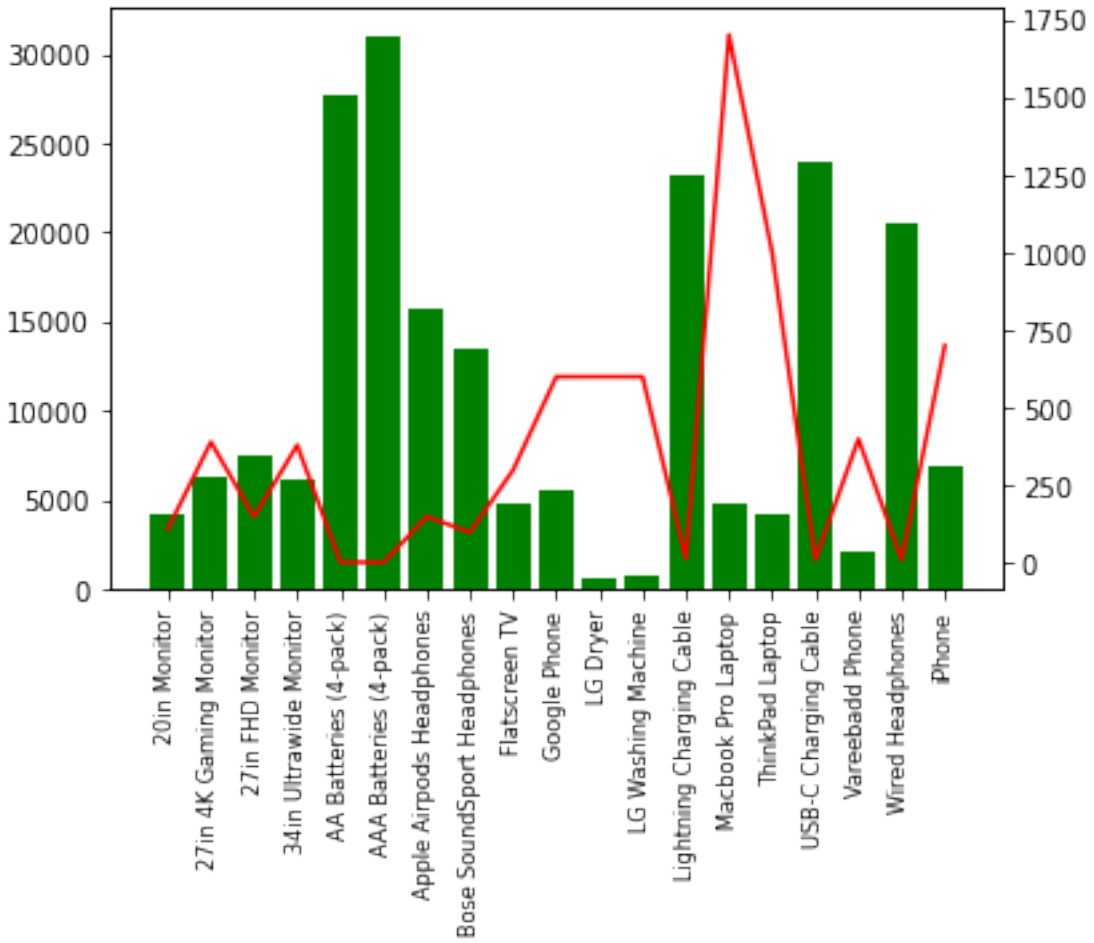
[152]:
```
Products = all_data.groupby('Product')['Quantity Ordered'].sum().index
quandtity = all_data.groupby('Product')['Quantity Ordered'].sum().values
Mean_Price = all_data.groupby('Product')['Price Each'].mean()
```

[159]:
```
plt.figure(figsize=(40,24))
fig,ax1 = plt.subplots()
ax2 = ax1.twinx()
ax1.bar(Products,quandtity,color='g')
ax2.plot(Products,Mean_Price,color='r')
ax1.set_xticklabels(Products,rotation='vertical',size=8)
```

```
<ipython-input-159-0c07eb64a477>:6: UserWarning: FixedFormatter should only be
used together with FixedLocator
  ax1.set_xticklabels(Products,rotation='vertical',size=8)
```

[159]:
```
[Text(0, 0, '20in Monitor'),
 Text(1, 0, '27in 4K Gaming Monitor'),
 Text(2, 0, '27in FHD Monitor'),
 Text(3, 0, '34in Ultrawide Monitor'),
 Text(4, 0, 'AA Batteries (4-pack)'),
 Text(5, 0, 'AAA Batteries (4-pack)'),
 Text(6, 0, 'Apple Airpods Headphones'),
 Text(7, 0, 'Bose SoundSport Headphones'),
 Text(8, 0, 'Flatscreen TV'),
 Text(9, 0, 'Google Phone'),
 Text(10, 0, 'LG Dryer'),
 Text(11, 0, 'LG Washing Machine'),
 Text(12, 0, 'Lightning Charging Cable'),
 Text(13, 0, 'Macbook Pro Laptop'),
 Text(14, 0, 'ThinkPad Laptop'),
 Text(15, 0, 'USB-C Charging Cable'),
 Text(16, 0, 'Vareebadd Phone'),
 Text(17, 0, 'Wired Headphones'),
 Text(18, 0, 'iPhone')]
```

```
<Figure size 2880x1728 with 0 Axes>
```



6 Quelle est la combinaison de produits la plus vendu ?

[161]:
```
df = all_data[all_data['Order ID'].duplicated(keep=False)]
```

[162]:
```
fd
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-162-c1b1d8137e2e> in <module>
----> 1 fd

NameError: name 'fd' is not defined
```

[163]:
```
df
```

[163]:
```
      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
…          …                          …                 …          …
11628   259303       AA Batteries (4-pack)                  1        3.84
11639   259314             Wired Headphones                 1       11.99
11640   259314      AAA Batteries (4-pack)                  2        2.99
11677   259350                 Google Phone                 1      600.00
11678   259350         USB-C Charging Cable                 1       11.95

           Order Date                      Purchase Address  mounth   Sales  \
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
…                 …                                     …      …      …
11628  09/20/19 20:18          106 7th St, Atlanta, GA 30301       9    3.84
11639  09/16/19 00:25     241 Highland St, Atlanta, GA 30301       9   11.99
11640  09/16/19 00:25     241 Highland St, Atlanta, GA 30301       9    5.98
11677  09/30/19 13:49  519 Maple St, San Francisco, CA 94016       9  600.00
11678  09/30/19 13:49  519 Maple St, San Francisco, CA 94016       9   11.95

                 City  hour
3         Los Angeles    14
4         Los Angeles    14
18        Los Angeles    19
19        Los Angeles    19
30             Boston    11
…                 …   …
11628         Atlanta    20
11639         Atlanta     0
11640         Atlanta     0
11677   San Francisco    13
11678   San Francisco    13

[14649 rows x 10 columns]
```

[166]:
```
df['Grouped'] = df.groupby('Order ID')['Product'].transform(lambda x:','.
↪join(x))
```

```
<ipython-input-166-0b35994e80a8>: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
  df['Grouped'] = df.groupby('Order ID')['Product'].transform(lambda
x:','.join(x))
```

[167]:
```
df['Grouped']
```

[167]:
```
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…
                                  …
11628         34in Ultrawide Monitor,AA Batteries (4-pack)
11639              Wired Headphones,AAA Batteries (4-pack)
11640              Wired Headphones,AAA Batteries (4-pack)
11677                    Google Phone,USB-C Charging Cable
11678                    Google Phone,USB-C Charging Cable
Name: Grouped, Length: 14649, dtype: object
```

[168]:
```
df
```

[168]:

	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	
…	…	…	…	…	
11628	259303	AA Batteries (4-pack)	1	3.84	
11639	259314	Wired Headphones	1	11.99	
11640	259314	AAA Batteries (4-pack)	2	2.99	
11677	259350	Google Phone	1	600.00	
11678	259350	USB-C Charging Cable	1	11.95	

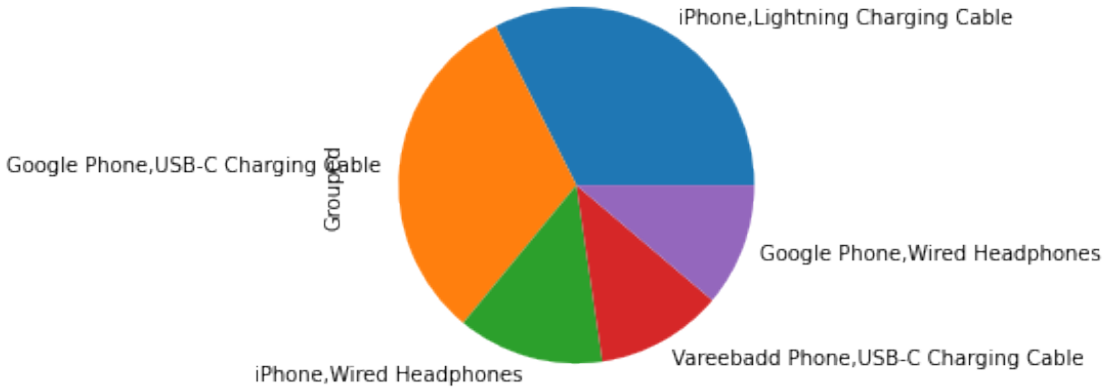
	Order Date	Purchase Address	mounth	Sales	\
3	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001	4	600.00	
4	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001	4	11.99	
18	04/03/19 19:42	20 Hill St, Los Angeles, CA 90001	4	600.00	
19	04/03/19 19:42	20 Hill St, Los Angeles, CA 90001	4	11.95	
30	04/07/19 11:31	823 Highland St, Boston, MA 02215	4	99.99	
…	…	…	…	…	
11628	09/20/19 20:18	106 7th St, Atlanta, GA 30301	9	3.84	
11639	09/16/19 00:25	241 Highland St, Atlanta, GA 30301	9	11.99	
11640	09/16/19 00:25	241 Highland St, Atlanta, GA 30301	9	5.98	
11677	09/30/19 13:49	519 Maple St, San Francisco, CA 94016	9	600.00	
11678	09/30/19 13:49	519 Maple St, San Francisco, CA 94016	9	11.95	

```
                  City  hour                                                Grouped
3          Los Angeles    14                    Google Phone,Wired Headphones
4          Los Angeles    14                    Google Phone,Wired Headphones
18         Los Angeles    19                 Google Phone,USB-C Charging Cable
19         Los Angeles    19                 Google Phone,USB-C Charging Cable
30              Boston    11  Bose SoundSport Headphones,Bose SoundSport Hea…
…                  …   …                                                   …
11628          Atlanta    20       34in Ultrawide Monitor,AA Batteries (4-pack)
11639          Atlanta     0            Wired Headphones,AAA Batteries (4-pack)
11640          Atlanta     0            Wired Headphones,AAA Batteries (4-pack)
11677    San Francisco    13                 Google Phone,USB-C Charging Cable
11678    San Francisco    13                 Google Phone,USB-C Charging Cable

[14649 rows x 11 columns]
```

[171]:
```
df1 = df.drop_duplicates(subset='Order ID')
```

[174]:
```
df1['Grouped'].value_counts()[0:5].plot.pie()
```

[174]:
```
<AxesSubplot:ylabel='Grouped'>
```



[175]:
```
df1['Grouped'].value_counts()[0:5]
```

[175]:
```
iPhone,Lightning Charging Cable         882
Google Phone,USB-C Charging Cable       856
iPhone,Wired Headphones                 361
Vareebadd Phone,USB-C Charging Cable    312
Google Phone,Wired Headphones           303
Name: Grouped, dtype: int64
```

[176]:
```
df1['Grouped']
```

[176]:
```
3                              Google Phone,Wired Headphones
18                          Google Phone,USB-C Charging Cable
30         Bose SoundSport Headphones,Bose SoundSport Hea…
32                       AAA Batteries (4-pack),Google Phone
119            Lightning Charging Cable,USB-C Charging Cable
                                    …
11617      Apple Airpods Headphones,Apple Airpods Headphones
11619      iPhone,Lightning Charging Cable,Lightning Char…
11627           34in Ultrawide Monitor,AA Batteries (4-pack)
11639                Wired Headphones,AAA Batteries (4-pack)
11677                       Google Phone,USB-C Charging Cable
Name: Grouped, Length: 7136, dtype: object
```

[177]:
```
!export PATH=/Library/TeX/texbin:$PATH
```

```
'export' n'est pas reconnu en tant que commande interne
ou externe, un programme exécutable ou un fichier de commandes.
```

[179]:
```
jupyter nbconvert your_notebook.ipynb --to pdf
```

```
  File "<ipython-input-179-34652316e6e6>", line 1
    jupyter nbconvert your_notebook.ipynb --to pdf
            ^
SyntaxError: invalid syntax
```

[180]:
```
import os
print(os.environ['PATH'])
```

```
C:\Users\boudi\anaconda3;C:\Users\boudi\anaconda3\Library\mingw-w64\bin;C:\Users
\boudi\anaconda3\Library\usr\bin;C:\Users\boudi\anaconda3\Library\bin;C:\Users\b
oudi\anaconda3\Scripts;C:\Users\boudi\anaconda3\bin;C:\Users\boudi\anaconda3\con
dabin;C:\Users\boudi\anaconda3;C:\Users\boudi\anaconda3\Library\mingw-w64\bin;C:
\Users\boudi\anaconda3\Library\usr\bin;C:\Users\boudi\anaconda3\Library\bin;C:\U
sers\boudi\anaconda3\Scripts;C:\Program Files\Common
Files\Oracle\Java\javapath;C:\Program Files (x86)\Common Files\Oracle\Java\javap
ath;C:\WINDOWS\system32;C:\WINDOWS;C:\WINDOWS\System32\Wbem;C:\WINDOWS\System32\
WindowsPowerShell\v1.0;C:\WINDOWS\System32\OpenSSH;C:\Program
Files\Java\jdk1.8.0_241\bin;C:\Program Files\PuTTY;C:\Program
Files\nodejs;C:\Program Files\Java\jdk-15.0.1\bin;C:\Program
Files\Git\cmd;C:\sqlite3;C:\Program Files\MySQL\MySQL Shell
8.0\bin;C:\Windows\System32\chcp.com;C:\Program Files\MySQL\MySQL Server
8.0\bin;C:\Windows\System32;C:\Program Files\MySQL\MySQL Server
8.0\bin;C:\Program Files\MySQL\MySQL Shell 8.0\bin;C:\Program Files\MySQL\MySQL
Router 8.0\bin;C:\Windows\System32;C:\ProgramData\Microsoft\Windows\Start
Menu\Programs;C:\Users\boudi\AppData\Roaming\Microsoft\Windows\Start
Menu\Programs\pip;C:\Program Files\MySQL\MySQL Workbench 8.0 CE;C:\Program
```

```
Files\MySQL\MySQL Shell 8.0\bin;C:\Program Files\nodejs;C:\Program
Files\MongoDB\Server\5.0\bin;C:\Windows\System32;C:\Program Files\MySQL\MySQL
Server 8.0\bin;C:\WINDOWS\system32;C:\WINDOWS;C:\Users\boudi\AppData\Local\Progr
ams\Microsoft VS Code\bin;C:\Users\boudi\AppData\Roaming\Microsoft\Windows\Start
Menu\Programs\pip;C:\Users\boudi\AppData\Local\Programs\Python\Python37\Lib\site
-packages\pip;C:\Users\boudi\AppData\Local\Programs\Python\Python37\Scripts;C:\P
rogram Files\MySQL\MySQL Workbench 8.0 CE;.
```

[181]:
```
pip install nbconvert
```

```
Requirement already satisfied: nbconvert in c:\users\boudi\anaconda3\lib\site-
packages (6.0.7)
Requirement already satisfied: traitlets>=4.2 in
c:\users\boudi\anaconda3\lib\site-packages (from nbconvert) (5.0.5)
Requirement already satisfied: mistune<2,>=0.8.1 in
c:\users\boudi\anaconda3\lib\site-packages (from nbconvert) (0.8.4)
Requirement already satisfied: jupyterlab-pygments in
c:\users\boudi\anaconda3\lib\site-packages (from nbconvert) (0.1.2)
Requirement already satisfied: pygments>=2.4.1 in
c:\users\boudi\anaconda3\lib\site-packages (from nbconvert) (2.8.1)
Requirement already satisfied: pandocfilters>=1.4.1 in
c:\users\boudi\anaconda3\lib\site-packages (from nbconvert) (1.4.3)
Requirement already satisfied: jinja2>=2.4 in c:\users\boudi\anaconda3\lib\site-
packages (from nbconvert) (2.11.3)
Requirement already satisfied: defusedxml in c:\users\boudi\anaconda3\lib\site-
packages (from nbconvert) (0.7.1)
Requirement already satisfied: entrypoints>=0.2.2 in
c:\users\boudi\anaconda3\lib\site-packages (from nbconvert) (0.3)
Requirement already satisfied: jupyter-core in
c:\users\boudi\anaconda3\lib\site-packages (from nbconvert) (4.7.1)
Requirement already satisfied: testpath in c:\users\boudi\anaconda3\lib\site-
packages (from nbconvert) (0.4.4)
Requirement already satisfied: nbformat>=4.4 in
c:\users\boudi\anaconda3\lib\site-packages (from nbconvert) (5.1.3)
Requirement already satisfied: bleach in c:\users\boudi\anaconda3\lib\site-
packages (from nbconvert) (3.3.0)
Requirement already satisfied: nbclient<0.6.0,>=0.5.0 in
c:\users\boudi\anaconda3\lib\site-packages (from nbconvert) (0.5.3)
Requirement already satisfied: MarkupSafe>=0.23 in
c:\users\boudi\anaconda3\lib\site-packages (from jinja2>=2.4->nbconvert) (1.1.1)
Requirement already satisfied: jupyter-client>=6.1.5 in
c:\users\boudi\anaconda3\lib\site-packages (from
nbclient<0.6.0,>=0.5.0->nbconvert) (6.1.12)
Requirement already satisfied: async-generator in
c:\users\boudi\anaconda3\lib\site-packages (from
nbclient<0.6.0,>=0.5.0->nbconvert) (1.10)
Requirement already satisfied: nest-asyncio in
c:\users\boudi\anaconda3\lib\site-packages (from
```

```
nbclient<0.6.0,>=0.5.0->nbconvert) (1.5.1)
Requirement already satisfied: pyzmq>=13 in c:\users\boudi\anaconda3\lib\site-
packages (from jupyter-client>=6.1.5->nbclient<0.6.0,>=0.5.0->nbconvert)
(20.0.0)
Requirement already satisfied: python-dateutil>=2.1 in
c:\users\boudi\anaconda3\lib\site-packages (from jupyter-
client>=6.1.5->nbclient<0.6.0,>=0.5.0->nbconvert) (2.8.1)
Requirement already satisfied: tornado>=4.1 in
c:\users\boudi\anaconda3\lib\site-packages (from jupyter-
client>=6.1.5->nbclient<0.6.0,>=0.5.0->nbconvert) (6.1)
Requirement already satisfied: pywin32>=1.0 in
c:\users\boudi\anaconda3\lib\site-packages (from jupyter-core->nbconvert) (227)
Requirement already satisfied: jsonschema!=2.5.0,>=2.4 in
c:\users\boudi\anaconda3\lib\site-packages (from nbformat>=4.4->nbconvert)
(3.2.0)
Requirement already satisfied: ipython-genutils in
c:\users\boudi\anaconda3\lib\site-packages (from nbformat>=4.4->nbconvert)
(0.2.0)
Requirement already satisfied: six>=1.11.0 in c:\users\boudi\anaconda3\lib\site-
packages (from jsonschema!=2.5.0,>=2.4->nbformat>=4.4->nbconvert) (1.15.0)
Requirement already satisfied: attrs>=17.4.0 in
c:\users\boudi\anaconda3\lib\site-packages (from
jsonschema!=2.5.0,>=2.4->nbformat>=4.4->nbconvert) (20.3.0)
Requirement already satisfied: pyrsistent>=0.14.0 in
c:\users\boudi\anaconda3\lib\site-packages (from
jsonschema!=2.5.0,>=2.4->nbformat>=4.4->nbconvert) (0.17.3)
Requirement already satisfied: setuptools in c:\users\boudi\anaconda3\lib\site-
packages (from jsonschema!=2.5.0,>=2.4->nbformat>=4.4->nbconvert)
(52.0.0.post20210125)
Requirement already satisfied: webencodings in
c:\users\boudi\anaconda3\lib\site-packages (from bleach->nbconvert) (0.5.1)
Requirement already satisfied: packaging in c:\users\boudi\anaconda3\lib\site-
packages (from bleach->nbconvert) (20.9)
Requirement already satisfied: pyparsing>=2.0.2 in
c:\users\boudi\anaconda3\lib\site-packages (from packaging->bleach->nbconvert)
(2.4.7)
Note: you may need to restart the kernel to use updated packages.
```

[182]:
```
pip install --upgrade --user nbconvert
```

```
Requirement already satisfied: nbconvert in c:\users\boudi\anaconda3\lib\site-
packages (6.0.7)
Collecting nbconvert
  Downloading nbconvert-6.3.0-py3-none-any.whl (556 kB)
Requirement already satisfied: nbformat>=4.4 in
c:\users\boudi\anaconda3\lib\site-packages (from nbconvert) (5.1.3)
Requirement already satisfied: jinja2>=2.4 in c:\users\boudi\anaconda3\lib\site-
packages (from nbconvert) (2.11.3)
```

```
Requirement already satisfied: bleach in c:\users\boudi\anaconda3\lib\site-
packages (from nbconvert) (3.3.0)
Requirement already satisfied: pandocfilters>=1.4.1 in
c:\users\boudi\anaconda3\lib\site-packages (from nbconvert) (1.4.3)
Requirement already satisfied: traitlets>=5.0 in
c:\users\boudi\anaconda3\lib\site-packages (from nbconvert) (5.0.5)
Requirement already satisfied: jupyterlab-pygments in
c:\users\boudi\anaconda3\lib\site-packages (from nbconvert) (0.1.2)
Requirement already satisfied: jupyter-core in
c:\users\boudi\anaconda3\lib\site-packages (from nbconvert) (4.7.1)
Requirement already satisfied: testpath in c:\users\boudi\anaconda3\lib\site-
packages (from nbconvert) (0.4.4)
Requirement already satisfied: defusedxml in c:\users\boudi\anaconda3\lib\site-
packages (from nbconvert) (0.7.1)
Requirement already satisfied: nbclient<0.6.0,>=0.5.0 in
c:\users\boudi\anaconda3\lib\site-packages (from nbconvert) (0.5.3)
Requirement already satisfied: mistune<2,>=0.8.1 in
c:\users\boudi\anaconda3\lib\site-packages (from nbconvert) (0.8.4)
Requirement already satisfied: pygments>=2.4.1 in
c:\users\boudi\anaconda3\lib\site-packages (from nbconvert) (2.8.1)
Requirement already satisfied: entrypoints>=0.2.2 in
c:\users\boudi\anaconda3\lib\site-packages (from nbconvert) (0.3)
Requirement already satisfied: MarkupSafe>=0.23 in
c:\users\boudi\anaconda3\lib\site-packages (from jinja2>=2.4->nbconvert) (1.1.1)
Requirement already satisfied: nest-asyncio in
c:\users\boudi\anaconda3\lib\site-packages (from
nbclient<0.6.0,>=0.5.0->nbconvert) (1.5.1)
Requirement already satisfied: async-generator in
c:\users\boudi\anaconda3\lib\site-packages (from
nbclient<0.6.0,>=0.5.0->nbconvert) (1.10)
Requirement already satisfied: jupyter-client>=6.1.5 in
c:\users\boudi\anaconda3\lib\site-packages (from
nbclient<0.6.0,>=0.5.0->nbconvert) (6.1.12)
Requirement already satisfied: tornado>=4.1 in
c:\users\boudi\anaconda3\lib\site-packages (from jupyter-
client>=6.1.5->nbclient<0.6.0,>=0.5.0->nbconvert) (6.1)
Requirement already satisfied: pyzmq>=13 in c:\users\boudi\anaconda3\lib\site-
packages (from jupyter-client>=6.1.5->nbclient<0.6.0,>=0.5.0->nbconvert)
(20.0.0)
Requirement already satisfied: python-dateutil>=2.1 in
c:\users\boudi\anaconda3\lib\site-packages (from jupyter-
client>=6.1.5->nbclient<0.6.0,>=0.5.0->nbconvert) (2.8.1)
Requirement already satisfied: pywin32>=1.0 in
c:\users\boudi\anaconda3\lib\site-packages (from jupyter-core->nbconvert) (227)
Requirement already satisfied: jsonschema!=2.5.0,>=2.4 in
c:\users\boudi\anaconda3\lib\site-packages (from nbformat>=4.4->nbconvert)
(3.2.0)
Requirement already satisfied: ipython-genutils in
```

```
c:\users\boudi\anaconda3\lib\site-packages (from nbformat>=4.4->nbconvert)
(0.2.0)
Requirement already satisfied: setuptools in c:\users\boudi\anaconda3\lib\site-
packages (from jsonschema!=2.5.0,>=2.4->nbformat>=4.4->nbconvert)
(52.0.0.post20210125)
Requirement already satisfied: attrs>=17.4.0 in
c:\users\boudi\anaconda3\lib\site-packages (from
jsonschema!=2.5.0,>=2.4->nbformat>=4.4->nbconvert) (20.3.0)
Requirement already satisfied: six>=1.11.0 in c:\users\boudi\anaconda3\lib\site-
packages (from jsonschema!=2.5.0,>=2.4->nbformat>=4.4->nbconvert) (1.15.0)
Requirement already satisfied: pyrsistent>=0.14.0 in
c:\users\boudi\anaconda3\lib\site-packages (from
jsonschema!=2.5.0,>=2.4->nbformat>=4.4->nbconvert) (0.17.3)
Requirement already satisfied: packaging in c:\users\boudi\anaconda3\lib\site-
packages (from bleach->nbconvert) (20.9)
Requirement already satisfied: webencodings in
c:\users\boudi\anaconda3\lib\site-packages (from bleach->nbconvert) (0.5.1)
Requirement already satisfied: pyparsing>=2.0.2 in
c:\users\boudi\anaconda3\lib\site-packages (from packaging->bleach->nbconvert)
(2.4.7)
Installing collected packages: nbconvert
Successfully installed nbconvert-6.3.0
Note: you may need to restart the kernel to use updated packages.

  WARNING: The scripts jupyter-dejavu.exe and jupyter-nbconvert.exe are
installed in 'C:\Users\boudi\AppData\Roaming\Python\Python38\Scripts' which is
not on PATH.
  Consider adding this directory to PATH or, if you prefer to suppress this
warning, use --no-warn-script-location.
ERROR: pip's dependency resolver does not currently take into account all the
packages that are installed. This behaviour is the source of the following
dependency conflicts.
spyder 4.2.5 requires pyqt5<5.13, which is not installed.
spyder 4.2.5 requires pyqtwebengine<5.13, which is not installed.
```

[183]:
```
import sys
if "\your\path\to\xelatex" not in sys.path:
    print('adding path') # I just add this to know if the path was present or␣
 ↪not.
    sys.path.append("\your\path\to\xelatex")
```

```
  File "<ipython-input-183-3711ee5406d8>", line 2
    if "\your\path\to\xelatex" not in sys.path:
       ^

SyntaxError: (unicode error) 'unicodeescape' codec can't decode bytes in␣
 ↪position 13-15: truncated \xXX escape
```

[ ]: