Import all required all libraries:

In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import os
import warnings
```

In [2]:
```
files=os.listdir()
files
```

Out[2]:
```
['.idea',
 '.ipynb_checkpoints',
 'main.py',
 'sales analysis.ipynb',
 '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']
```

removing files that are not required

In [3]:
```
list=[]
for files in files:
    if(files.endswith('csv')):
        list.append(files)
list
```

Out[3]:
```
['Sales_April_2019.csv',
 'Sales_August_2019.csv',
 'Sales_December_2019.csv',
 'Sales_February_2019.csv',
 'Sales_January_2019.csv',
 'Sales_July_2019.csv',
 'Sales_June_2019.csv',
 'Sales_March_2019.csv',
 'Sales_May_2019.csv',
 'Sales_November_2019.csv',
 'Sales_October_2019.csv',
 'Sales_September_2019.csv']
```

In [4]:
```
edf=pd.DataFrame()
for file in list:
    df=pd.read_csv(file)
    edf=pd.concat([edf,df])
edf
```

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

186850 rows × 6 columns

all file concatenate in one file

In [5]:
```
edf.to_csv('sales.csv',index= False)
```

In [6]:
```
edf
```

Out[6]:

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

186850 rows × 6 columns

removing NaN values in whole row

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

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

186305 rows × 6 columns

checking null values,datatype in dataframe

In [8]:
```
edf.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 186305 entries, 0 to 11685
Data columns (total 6 columns):
 #   Column            Non-Null Count   Dtype 
---  ------            --------------   ----- 
 0   Order ID          186305 non-null  object
 1   Product           186305 non-null  object
 2   Quantity Ordered  186305 non-null  object
 3   Price Each        186305 non-null  object
 4   Order Date        186305 non-null  object
 5   Purchase Address  186305 non-null  object
dtypes: object(6)
memory usage: 9.9+ MB
```

checking which type of string or present in int or float datatype

In [10]:
```
edf['unk']=edf['Quantity Ordered'].apply(str).str.isnumeric()
edf=edf.loc[edf['unk']==True]
edf['Quantity Ordered']=edf['Quantity Ordered'].astype(np.int64)
edf['Order ID']=edf['Order ID'].astype(np.int64)
edf['Price Each']=edf['Price Each'].astype(np.float64)
edf['Order Date']=edf['Order Date'].apply(str).str.replace('/','-')
edf['Order Date']=pd.to_datetime(edf['Order Date'])
edf=edf.drop(columns='unk')
edf.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 185950 entries, 0 to 11685
Data columns (total 6 columns):
 #   Column            Non-Null Count   Dtype         
---  ------            --------------   -----         
 0   Order ID          185950 non-null  int64         
 1   Product           185950 non-null  object        
 2   Quantity Ordered  185950 non-null  int64         
 3   Price Each        185950 non-null  float64       
 4   Order Date        185950 non-null  datetime64[ns]
 5   Purchase Address  185950 non-null  object        
dtypes: datetime64[ns](1), float64(1), int64(2), object(2)
memory usage: 9.9+ MB
```

checking top 5 values and bottom 5 values

In [11]:
```
edf.head()
```

Out[11]:

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

In [12]:
```
edf.tail()
```

Out[12]:

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address
11681	259353	AAA Batteries (4-pack)	3	2.99	2019-09-17 20:56:00	840 Highland St, Los Angeles, CA 90001
11682	259354	iPhone	1	700.00	2019-09-01 16:00:00	216 Dogwood St, San Francisco, CA 94016
11683	259355	iPhone	1	700.00	2019-09-23 07:39:00	220 12th St, San Francisco, CA 94016
11684	259356	34in Ultrawide Monitor	1	379.99	2019-09-19 17:30:00	511 Forest St, San Francisco, CA 94016
11685	259357	USB-C Charging Cable	1	11.95	2019-09-30 00:18:00	250 Meadow St, San Francisco, CA 94016

checking rows and columns

In [13]:
```
edf.shape
```

Out[13]:
```
(185950, 6)
```

checking columns name:

In [14]:
```
edf.columns
```

Out[14]:
```
Index(['Order ID', 'Product', 'Quantity Ordered', 'Price Each', 'Order Date',
       'Purchase Address'],
      dtype='object')
```

data cleaning has been done now analysis is take place

1.what was the total sale by month

In [15]:
```
edf['tsv']=edf['Quantity Ordered']*edf['Price Each']
edf
```

Out[15]:

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	tsv
0	176558	USB-C Charging Cable	2	11.95	2019-04-19 08:46:00	917 1st St, Dallas, TX 75001	23.90
2	176559	Bose SoundSport Headphones	1	99.99	2019-04-07 22:30:00	682 Chestnut St, Boston, MA 02215	99.99
3	176560	Google Phone	1	600.00	2019-04-12 14:38:00	669 Spruce St, Los Angeles, CA 90001	600.00
4	176560	Wired Headphones	1	11.99	2019-04-12 14:38:00	669 Spruce St, Los Angeles, CA 90001	11.99
5	176561	Wired Headphones	1	11.99	2019-04-30 09:27:00	333 8th St, Los Angeles, CA 90001	11.99
...	...	...	...	...	...	...	...
11681	259353	AAA Batteries (4-pack)	3	2.99	2019-09-17 20:56:00	840 Highland St, Los Angeles, CA 90001	8.97
11682	259354	iPhone	1	700.00	2019-09-01 16:00:00	216 Dogwood St, San Francisco, CA 94016	700.00
11683	259355	iPhone	1	700.00	2019-09-23 07:39:00	220 12th St, San Francisco, CA 94016	700.00
11684	259356	34in Ultrawide Monitor	1	379.99	2019-09-19 17:30:00	511 Forest St, San Francisco, CA 94016	379.99
11685	259357	USB-C Charging Cable	1	11.95	2019-09-30 00:18:00	250 Meadow St, San Francisco, CA 94016	11.95

185950 rows × 7 columns

extracting month name from date

In [16]:
```
edf['month']=edf['Order Date'].dt.month_name()
edf
```

Out[16]:

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	tsv	month
0	176558	USB-C Charging Cable	2	11.95	2019-04-19 08:46:00	917 1st St, Dallas, TX 75001	23.90	April
2	176559	Bose SoundSport Headphones	1	99.99	2019-04-07 22:30:00	682 Chestnut St, Boston, MA 02215	99.99	April
3	176560	Google Phone	1	600.00	2019-04-12 14:38:00	669 Spruce St, Los Angeles, CA 90001	600.00	April
4	176560	Wired Headphones	1	11.99	2019-04-12 14:38:00	669 Spruce St, Los Angeles, CA 90001	11.99	April
5	176561	Wired Headphones	1	11.99	2019-04-30 09:27:00	333 8th St, Los Angeles, CA 90001	11.99	April
...	...	...	...	...	...	...	...	...
11681	259353	AAA Batteries (4-pack)	3	2.99	2019-09-17 20:56:00	840 Highland St, Los Angeles, CA 90001	8.97	September
11682	259354	iPhone	1	700.00	2019-09-01 16:00:00	216 Dogwood St, San Francisco, CA 94016	700.00	September
11683	259355	iPhone	1	700.00	2019-09-23 07:39:00	220 12th St, San Francisco, CA 94016	700.00	September
11684	259356	34in Ultrawide Monitor	1	379.99	2019-09-19 17:30:00	511 Forest St, San Francisco, CA 94016	379.99	September

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	tsv	month
11685	259357	USB-C Charging Cable	1	11.95	2019-09-30 00:18:00	250 Meadow St, San Francisco, CA 94016	11.95	September

185950 rows × 8 columns

In [17]:
```
tdf=edf.groupby('month').agg(
     total=('tsv','sum'))
tdf
```

Out[17]:

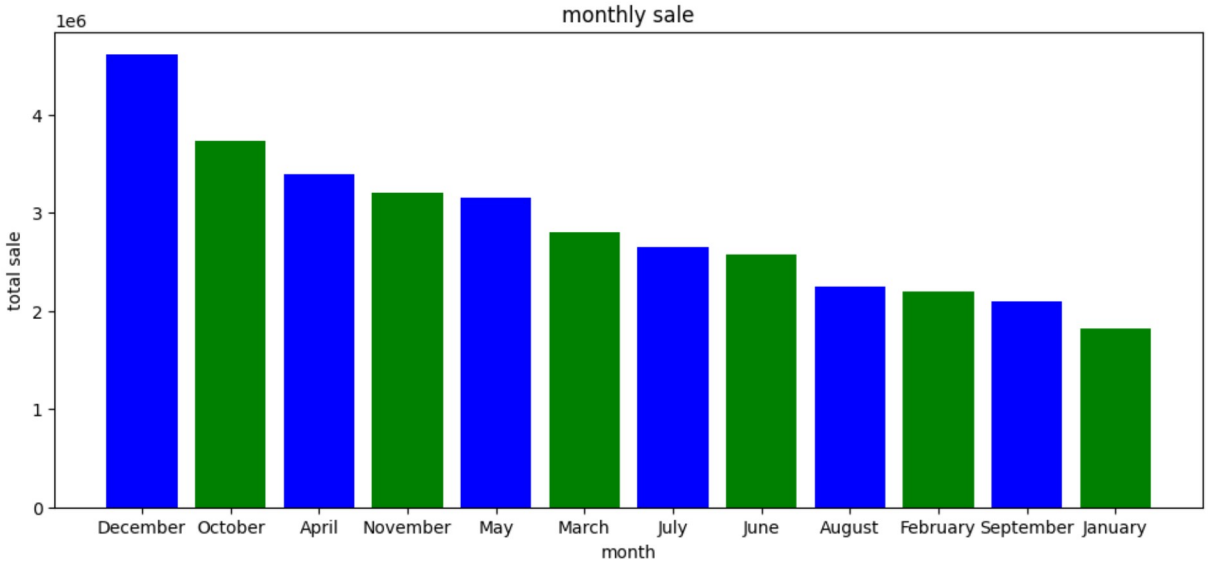
month	total
April	3390670.24
August	2244467.88
December	4613443.34
February	2202022.42
January	1822256.73
July	2647775.76
June	2577802.26
March	2807100.38
May	3152606.75
November	3199603.20
October	3736726.88
September	2097560.13

In [18]:
```
tdf=tdf.sort_values('total',ascending=False).reset_index()
tdf
```

Out[18]:

	month	total
0	December	4613443.34
1	October	3736726.88
2	April	3390670.24
3	November	3199603.20
4	May	3152606.75
5	March	2807100.38
6	July	2647775.76
7	June	2577802.26
8	August	2244467.88
9	February	2202022.42
10	September	2097560.13
11	January	1822256.73

In [19]:
```
plt.figure(figsize=(12,5))
plt.bar(tdf['month'],tdf['total'],color=['blue','green'])
plt.title('monthly sale')
plt.xlabel('month')
plt.ylabel('total sale')
plt.show()
```



2. what city sold the most

In [20]:
```
def fun(string):
    string=string.split(',')
    return string[1]

edf['city']=edf['Purchase Address'].apply(fun)
```

In [21]:
```
edf
```

Out[21]:

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	tsv	month	
0	176558	USB-C Charging Cable	2	11.95	2019-04-19 08:46:00	917 1st St, Dallas, TX 75001	23.90	April	
2	176559	Bose SoundSport Headphones	1	99.99	2019-04-07 22:30:00	682 Chestnut St, Boston, MA 02215	99.99	April	
3	176560	Google Phone	1	600.00	2019-04-12 14:38:00	669 Spruce St, Los Angeles, CA 90001	600.00	April	
4	176560	Wired Headphones	1	11.99	2019-04-12 14:38:00	669 Spruce St, Los Angeles, CA 90001	11.99	April	
5	176561	Wired Headphones	1	11.99	2019-04-30 09:27:00	333 8th St, Los Angeles, CA 90001	11.99	April	
...	...	...	...	...	...	...	...	...	
11681	259353	AAA Batteries (4-pack)	3	2.99	2019-09-17 20:56:00	840 Highland St, Los Angeles, CA 90001	8.97	September	
11682	259354	iPhone	1	700.00	2019-09-01 16:00:00	216 Dogwood St, San Francisco, CA 94016	700.00	September	Fr
11683	259355	iPhone	1	700.00	2019-09-23 07:39:00	220 12th St, San Francisco, CA 94016	700.00	September	Fr
11684	259356	34in Ultrawide Monitor	1	379.99	2019-09-19 17:30:00	511 Forest St, San Francisco, CA 94016	379.99	September	Fr

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	tsv	month	
11685	259357	USB-C Charging Cable	1	11.95	2019-09-30 00:18:00	250 Meadow St, San Francisco, CA 94016	11.95	September	Fr

185950 rows × 9 columns

In [22]:
```
cdf=edf.groupby('city').agg(
    tsc=('tsv','sum')
)
cdf=cdf.reset_index()
cdf
```

Out[22]:

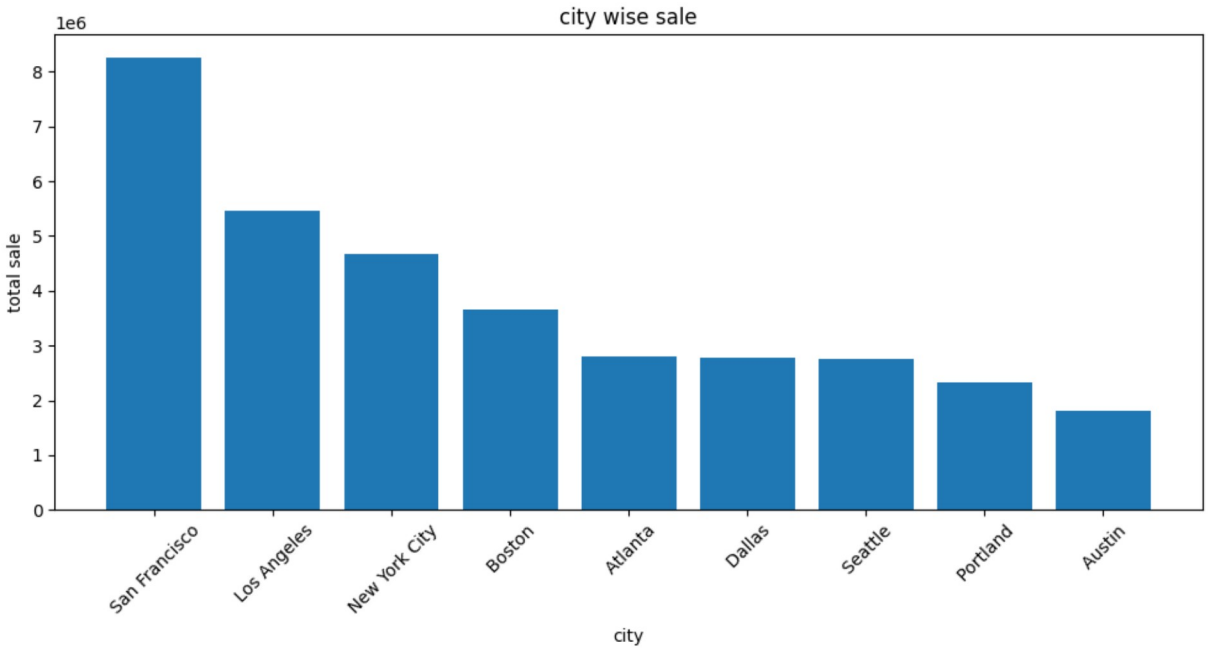
	city	tsc
0	Atlanta	2795498.58
1	Austin	1819581.75
2	Boston	3661642.01
3	Dallas	2767975.40
4	Los Angeles	5452570.80
5	New York City	4664317.43
6	Portland	2320490.61
7	San Francisco	8262203.91
8	Seattle	2747755.48

In [23]:
```
cdf=cdf.sort_values('tsc',ascending=False)
cdf.head()
```

Out[23]:

	city	tsc
7	San Francisco	8262203.91
4	Los Angeles	5452570.80
5	New York City	4664317.43
2	Boston	3661642.01
0	Atlanta	2795498.58

In [24]:
```
plt.figure(figsize=(12,5))
plt.xticks(rotation=45)
plt.bar(cdf['city'],cdf['tsc'])
plt.title('city wise sale')
plt.xlabel('city')
plt.ylabel('total sale')
plt.show()
```



which time to most sold

In [25]:
```
edf['time']=edf['Order Date'].dt.hour
edf
```

Out[25]:

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	tsv	month	
0	176558	USB-C Charging Cable	2	11.95	2019-04-19 08:46:00	917 1st St, Dallas, TX 75001	23.90	April	
2	176559	Bose SoundSport Headphones	1	99.99	2019-04-07 22:30:00	682 Chestnut St, Boston, MA 02215	99.99	April	
3	176560	Google Phone	1	600.00	2019-04-12 14:38:00	669 Spruce St, Los Angeles, CA 90001	600.00	April	
4	176560	Wired Headphones	1	11.99	2019-04-12 14:38:00	669 Spruce St, Los Angeles, CA 90001	11.99	April	
5	176561	Wired Headphones	1	11.99	2019-04-30 09:27:00	333 8th St, Los Angeles, CA 90001	11.99	April	
...	...	...	...	...	...	...	...	...	
11681	259353	AAA Batteries (4-pack)	3	2.99	2019-09-17 20:56:00	840 Highland St, Los Angeles, CA 90001	8.97	September	
11682	259354	iPhone	1	700.00	2019-09-01 16:00:00	216 Dogwood St, San Francisco, CA 94016	700.00	September	Fr
11683	259355	iPhone	1	700.00	2019-09-23 07:39:00	220 12th St, San Francisco, CA 94016	700.00	September	Fr
11684	259356	34in Ultrawide Monitor	1	379.99	2019-09-19 17:30:00	511 Forest St, San Francisco, CA 94016	379.99	September	Fr

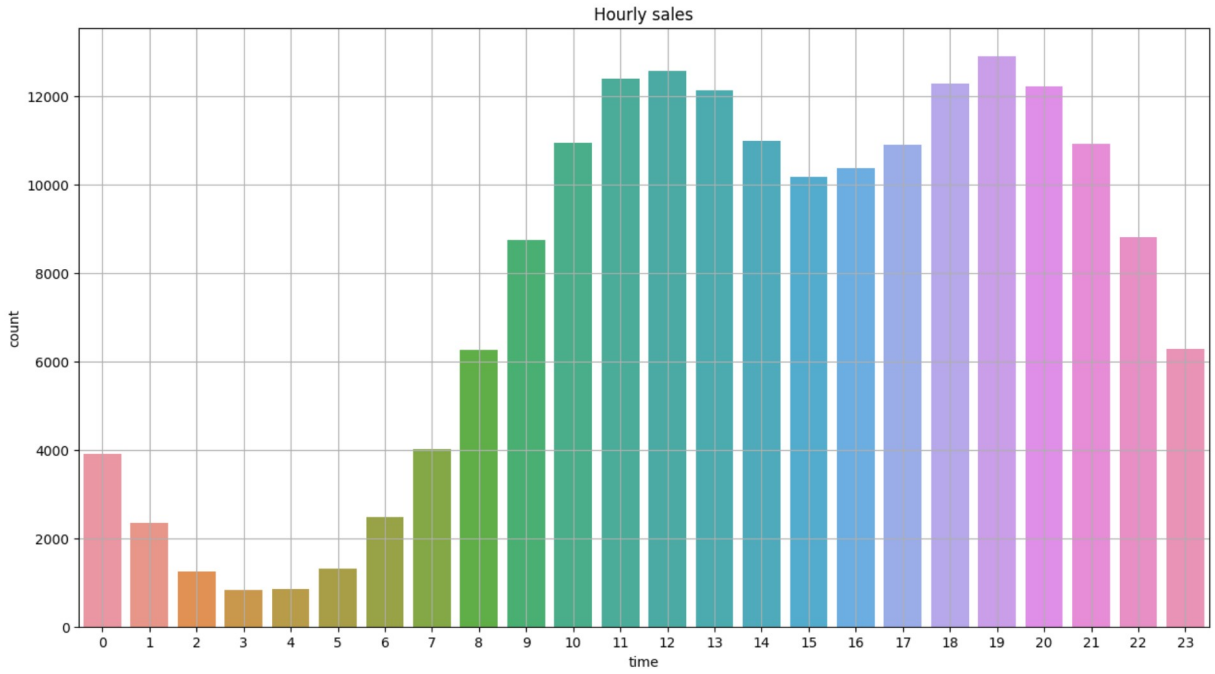
	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	tsv	month	
11685	259357	USB-C Charging Cable	1	11.95	2019-09-30 00:18:00	250 Meadow St, San Francisco, CA 94016	11.95	September	Fı

185950 rows × 10 columns

In [26]:
```
adf=edf['time'].value_counts()
adf
```

Out[26]:
```
time
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: count, dtype: int64
```

In [27]:
```
plt.figure(figsize=(15,8))
plt.title('Hourly sales')
sns.countplot(x='time',data=edf)
plt.grid()
plt.show()
```



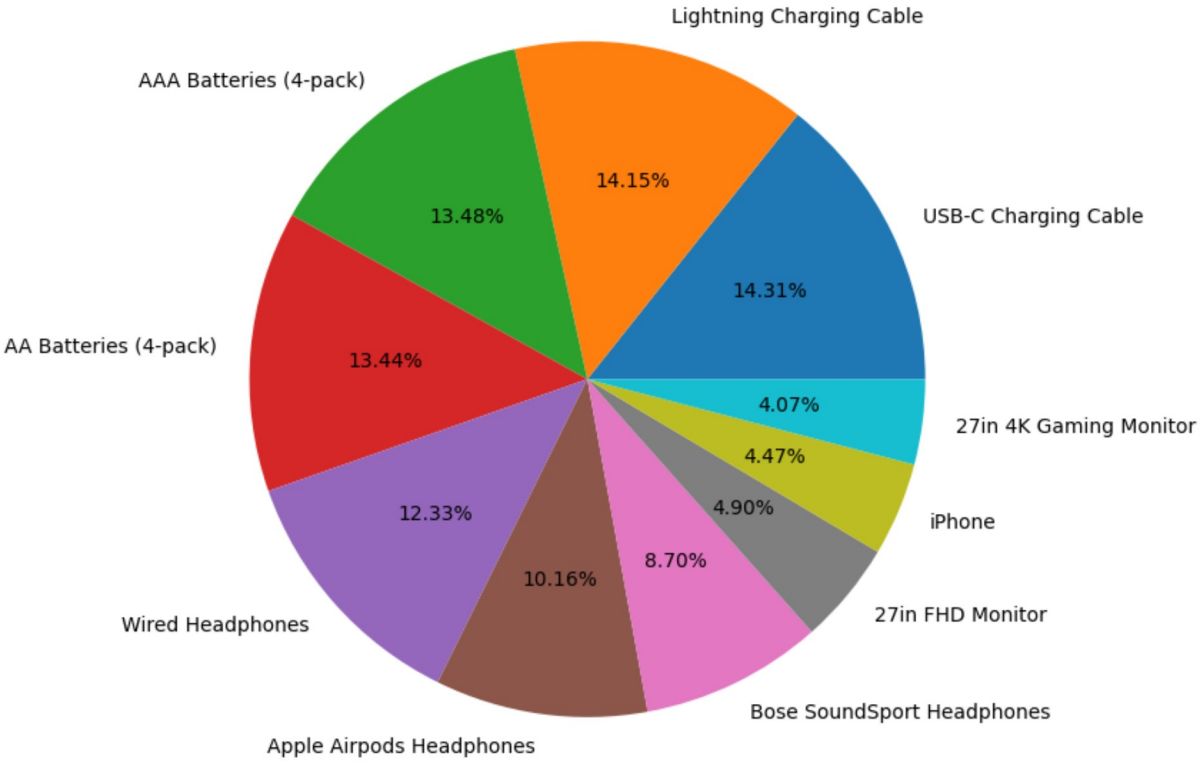
which product sells the most (top 10)

In [28]:
```
vdf=edf['Product'].value_counts().reset_index().sort_values('count',ascending=False
vdf.head(10)
```

Out[28]:

	Product	count
0	USB-C Charging Cable	21903
1	Lightning Charging Cable	21658
2	AAA Batteries (4-pack)	20641
3	AA Batteries (4-pack)	20577
4	Wired Headphones	18882
5	Apple Airpods Headphones	15549
6	Bose SoundSport Headphones	13325
7	27in FHD Monitor	7507
8	iPhone	6842
9	27in 4K Gaming Monitor	6230

In [29]:
```
plt.pie(vdf.head(10)['count'],labels=vdf.head(10)['Product'],autopct='%.2f%%',radiu
plt.show()
```



In [53]:
```
ndf=edf.groupby(['Product']).agg(
            total= ('tsv','sum')
)
ndf=ndf.reset_index().sort_values('total',ascending=False)
ndf.head(10)
```

Out[53]:

	Product	total
13	Macbook Pro Laptop	8037600.00
18	iPhone	4794300.00
14	ThinkPad Laptop	4129958.70
9	Google Phone	3319200.00
1	27in 4K Gaming Monitor	2435097.56
3	34in Ultrawide Monitor	2355558.01
6	Apple Airpods Headphones	2349150.00
8	Flatscreen TV	1445700.00
7	Bose SoundSport Headphones	1345565.43
2	27in FHD Monitor	1132424.50

In [60]:
```
plt.figure(figsize=(15,8))
plt.title('  ')
plt.xticks(rotation=45)
plt.bar(ndf.head(10)['Product'],ndf.head(10)['total'],color='magenta')
plt.xlabel('product')
plt.ylabel('Total Sale')
plt.show()
```