Data Analysis (Sales.csv)

In [137…

In [138…

In [139…

In [140…

In [ ]:

In [3]:
```
import pandas as pd
from pandas import Series
from pandas import DataFrame
```

In [143…
```
pd.read_csv('sales.csv')
```

Out[143]:

	Invoice ID	Date	Time	Gender	Location	City	Member	Category	Price	Quant
0	460489604	1/25/2018	16:46	Male	Brookfield	NewYork	Yes	Groceries	30	
1	471006167	3/19/2018	16:48	Female	Water tower	Chicago	Yes	Fashion	35	
2	411909258	2/25/2018	13:33	Male	Water tower	Chicago	No	Clothing	57	
3	487313402	1/22/2018	13:38	Female	Park lane	Dallas	Yes	Sporting	89	
4	197763430	2/18/2018	15:31	Female	Park lane	Dallas	No	Books	82	
...	...	...	...	...	...	...	...	...	...	
995	818829599	3/26/2018	11:19	Male	Park lane	Dallas	No	Groceries	31	
996	556589713	2/20/2018	17:17	Male	Water tower	Chicago	Yes	Groceries	3	
997	82324424	2/6/2018	11:44	Male	Brookfield	NewYork	No	Clothing	71	
998	783661702	1/29/2018	15:44	Female	Brookfield	NewYork	No	Clothing	89	
999	759171975	1/31/2018	10:13	Female	Water tower	Chicago	Yes	Clothing	31	

1000 rows × 13 columns

In [144…
```
pd.read_csv('sales.csv')
```

Out[144]:

	Invoice ID	Date	Time	Gender	Location	City	Member	Category	Price	Quant
0	460489604	1/25/2018	16:46	Male	Brookfield	NewYork	Yes	Groceries	30	
1	471006167	3/19/2018	16:48	Female	Water tower	Chicago	Yes	Fashion	35	
2	411909258	2/25/2018	13:33	Male	Water tower	Chicago	No	Clothing	57	
3	487313402	1/22/2018	13:38	Female	Park lane	Dallas	Yes	Sporting	89	
4	197763430	2/18/2018	15:31	Female	Park lane	Dallas	No	Books	82	
...	...	...	...	...	...	...	...	...	...	
995	818829599	3/26/2018	11:19	Male	Park lane	Dallas	No	Groceries	31	
996	556589713	2/20/2018	17:17	Male	Water tower	Chicago	Yes	Groceries	3	
997	82324424	2/6/2018	11:44	Male	Brookfield	NewYork	No	Clothing	71	
998	783661702	1/29/2018	15:44	Female	Brookfield	NewYork	No	Clothing	89	
999	759171975	1/31/2018	10:13	Female	Water tower	Chicago	Yes	Clothing	31	

1000 rows × 13 columns

In [145…

```
pd.read_csv('sales (3).csv')
```

Out[145]:

	Invoice ID	Date	Time	Gender	Location	City	Member	Category	Price	Quant
0	460489604	1/25/2018	16:46	Male	Brookfield	NewYork	Yes	Groceries	30	
1	471006167	3/19/2018	16:48	Female	Water tower	Chicago	Yes	Fashion	35	
2	411909258	2/25/2018	13:33	Male	Water tower	Chicago	No	Clothing	57	
3	487313402	1/22/2018	13:38	Female	Park lane	Dallas	Yes	Sporting	89	
4	197763430	2/18/2018	15:31	Female	Park lane	Dallas	No	Books	82	
...	...	...	...	...	...	...	...	...	...	
995	818829599	3/26/2018	11:19	Male	Park lane	Dallas	No	Groceries	31	
996	556589713	2/20/2018	17:17	Male	Water tower	Chicago	Yes	Groceries	3	
997	82324424	2/6/2018	11:44	Male	Brookfield	NewYork	No	Clothing	71	
998	783661702	1/29/2018	15:44	Female	Brookfield	NewYork	No	Clothing	89	
999	759171975	1/31/2018	10:13	Female	Water tower	Chicago	Yes	Clothing	31	

1000 rows × 13 columns

In [ ]:

In [ ]:
```
sales=pd.read_csv('sales.csv')
```

In [147…
```
sales.head(10)
```

Out[147]:

	Invoice ID	Date	Time	Gender	Location	City	Member	Category	Price	Quantity
0	460489604	1/25/2018	16:46	Male	Brookfield	NewYork	Yes	Groceries	30	1
1	471006167	3/19/2018	16:48	Female	Water tower	Chicago	Yes	Fashion	35	5
2	411909258	2/25/2018	13:33	Male	Water tower	Chicago	No	Clothing	57	2
3	487313402	1/22/2018	13:38	Female	Park lane	Dallas	Yes	Sporting	89	4
4	197763430	2/18/2018	15:31	Female	Park lane	Dallas	No	Books	82	5
5	263634050	3/9/2018	17:55	Male	Brookfield	NewYork	Yes	Clothing	52	4
6	99646662	3/4/2018	13:21	Male	Water tower	Chicago	No	Clothing	81	3
7	188869875	3/8/2018	13:24	Male	Park lane	Dallas	No	Books	80	2
8	325637547	1/18/2018	15:33	Male	Park lane	Dallas	Yes	Clothing	91	3
9	562942936	2/24/2018	16:05	Female	Park lane	Dallas	No	Books	80	2

In [148…
```
sales['Invoice ID']
```

Out[148]:
```
0      460489604
1      471006167
2      411909258
3      487313402
4      197763430
         ...
995    818829599
996    556589713
997     82324424
998    783661702
999    759171975
Name: Invoice ID, Length: 1000, dtype: int64
```

In [ ]:
```
sales['Category']
```

In [149…
```
sales['Category'].unique()
```

Out[149]:
```
array(['Groceries', 'Fashion', 'Clothing', 'Sporting', 'Books',
       'Furniture'], dtype=object)
```

In [150…
```
sales.iloc[201]
```

Out[150]:
```
Invoice ID      414119922
Date             1/30/2018
Time                 11:07
Gender                Male
Location       Water tower
City               Chicago
Member                  No
Category          Sporting
Price                   44
Quantity                 7
Total                  308
Payment               Card
Rating                   3
Day                     30
Month                    1
Name: 201, dtype: object
```

In [151…
```
sales[200:211]
```

Out[151]:

	Invoice ID	Date	Time	Gender	Location	City	Member	Category	Price	Quant
200	576523018	2/15/2018	12:35	Male	Water tower	Chicago	No	Groceries	82	
201	414119922	1/30/2018	11:07	Male	Water tower	Chicago	No	Sporting	44	
202	123247697	2/8/2018	20:03	Male	Brookfield	NewYork	No	Groceries	22	
203	40419058	1/26/2018	14:53	Male	Water tower	Chicago	Yes	Books	95	
204	890497298	2/25/2018	18:06	Female	Water tower	Chicago	No	Clothing	69	
205	885702268	1/28/2018	12:45	Male	Brookfield	NewYork	No	Groceries	89	
206	172252194	3/22/2018	20:10	Female	Park lane	Dallas	No	Books	41	
207	866140211	2/25/2018	13:34	Female	Park lane	Dallas	No	Clothing	54	
208	597678548	3/15/2018	17:35	Male	Water tower	Chicago	Yes	Sporting	34	
209	428818456	2/25/2018	18:50	Male	Park lane	Dallas	Yes	Books	94	
210	236950025	3/9/2018	12:40	Female	Brookfield	NewYork	Yes	Furniture	76	

In [152…
```
sales.tail(9)
```

Out[152]:

	Invoice ID	Date	Time	Gender	Location	City	Member	Category	Price	Quant
991	699906043	2/15/2018	12:58	Male	Brookfield	NewYork	Yes	Furniture	96	
992	927909627	3/2/2018	13:35	Male	Brookfield	NewYork	No	Books	54	
993	455112063	3/19/2018	17:30	Female	Brookfield	NewYork	Yes	Clothing	3	
994	446582560	1/31/2018	19:00	Male	Water tower	Chicago	No	Fashion	29	
995	818829599	3/26/2018	11:19	Male	Park lane	Dallas	No	Groceries	31	
996	556589713	2/20/2018	17:17	Male	Water tower	Chicago	Yes	Groceries	3	
997	82324424	2/6/2018	11:44	Male	Brookfield	NewYork	No	Clothing	71	
998	783661702	1/29/2018	15:44	Female	Brookfield	NewYork	No	Clothing	89	
999	759171975	1/31/2018	10:13	Female	Water tower	Chicago	Yes	Clothing	31	

In [153…

```
sales[sales["Gender"]=='Male']
```

Out[153]:

	Invoice ID	Date	Time	Gender	Location	City	Member	Category	Price	Quant
0	460489604	1/25/2018	16:46	Male	Brookfield	NewYork	Yes	Groceries	30	
2	411909258	2/25/2018	13:33	Male	Water tower	Chicago	No	Clothing	57	
5	263634050	3/9/2018	17:55	Male	Brookfield	NewYork	Yes	Clothing	52	
6	99646662	3/4/2018	13:21	Male	Water tower	Chicago	No	Clothing	81	
7	188869875	3/8/2018	13:24	Male	Park lane	Dallas	No	Books	80	
...	...	...	...	...	...	...	...	...	...	...
992	927909627	3/2/2018	13:35	Male	Brookfield	NewYork	No	Books	54	
994	446582560	1/31/2018	19:00	Male	Water tower	Chicago	No	Fashion	29	
995	818829599	3/26/2018	11:19	Male	Park lane	Dallas	No	Groceries	31	
996	556589713	2/20/2018	17:17	Male	Water tower	Chicago	Yes	Groceries	3	
997	82324424	2/6/2018	11:44	Male	Brookfield	NewYork	No	Clothing	71	

499 rows × 15 columns

In [154…

```
sales[sales['Total']>100]
```

Out[154]:

	Invoice ID	Date	Time	Gender	Location	City	Member	Category	Price	Quant
1	471006167	3/19/2018	16:48	Female	Water tower	Chicago	Yes	Fashion	35	
2	411909258	2/25/2018	13:33	Male	Water tower	Chicago	No	Clothing	57	
3	487313402	1/22/2018	13:38	Female	Park lane	Dallas	Yes	Sporting	89	
4	197763430	2/18/2018	15:31	Female	Park lane	Dallas	No	Books	82	
5	263634050	3/9/2018	17:55	Male	Brookfield	NewYork	Yes	Clothing	52	
...	...	...	...	...	...	...	...	...	...	
994	446582560	1/31/2018	19:00	Male	Water tower	Chicago	No	Fashion	29	
995	818829599	3/26/2018	11:19	Male	Park lane	Dallas	No	Groceries	31	
997	82324424	2/6/2018	11:44	Male	Brookfield	NewYork	No	Clothing	71	
998	783661702	1/29/2018	15:44	Female	Brookfield	NewYork	No	Clothing	89	
999	759171975	1/31/2018	10:13	Female	Water tower	Chicago	Yes	Clothing	31	

658 rows × 15 columns

In [155…
```
sales[sales['Payment']== 'Card']
```

Out[155]:

	Invoice ID	Date	Time	Gender	Location	City	Member	Category	Price	Quant
1	471006167	3/19/2018	16:48	Female	Water tower	Chicago	Yes	Fashion	35	
8	325637547	1/18/2018	15:33	Male	Park lane	Dallas	Yes	Clothing	91	
11	846531724	3/15/2018	12:12	Male	Park lane	Dallas	No	Sporting	22	
15	355981315	2/15/2018	15:47	Female	Brookfield	NewYork	Yes	Clothing	78	
21	443198801	2/24/2018	19:31	Female	Brookfield	NewYork	No	Clothing	48	
...	...	...	...	...	...	...	...	...	...	
987	245049016	1/20/2018	10:03	Female	Park lane	Dallas	No	Furniture	74	
991	699906043	2/15/2018	12:58	Male	Brookfield	NewYork	Yes	Furniture	96	
992	927909627	3/2/2018	13:35	Male	Brookfield	NewYork	No	Books	54	
995	818829599	3/26/2018	11:19	Male	Park lane	Dallas	No	Groceries	31	
997	82324424	2/6/2018	11:44	Male	Brookfield	NewYork	No	Clothing	71	

311 rows × 15 columns

In [156…
```
sales['Payment'].unique()
```

Out[156]:
```
array(['Cash', 'Card', 'Gpay'], dtype=object)
```

In [157…
```
sales.sum()['Total']
```

Out[157]:
```
207788
```

In [158…
```
sales.sum()['Quantity']
```

Out[158]:
```
4059
```

In [159…
```
sales.max()
```

Out[159]:
```
Invoice ID      999665753
Date              3/9/2018
Time                 20:59
Gender                Male
Location      Water tower
City              NewYork
Member                Yes
Category          Sporting
Price                  100
Quantity                 7
Total                  693
Payment              Gpay
Rating                   5
Day                     31
Month                    3
dtype: object
```

In [160…
```
sales.max()["Total"]
```

Out[160]:
```
693
```

In [161…
```
sales[sales["Total"]== sales.max()["Total"]]
```

Out[161]:

	Invoice ID	Date	Time	Gender	Location	City	Member	Category	Price	Quantit
47	315624039	3/30/2018	19:45	Female	Water tower	Chicago	Yes	Groceries	99	
679	270918350	3/27/2018	13:21	Female	Park lane	Dallas	Yes	Groceries	99	

In [162…
```
sales[sales["Total"]== sales.min()["Total"]]
```

Out[162]:

	Invoice ID	Date	Time	Gender	Location	City	Member	Category	Price	Quantity
184	801334050	1/21/2018	20:18	Female	Park lane	Dallas	No	Furniture	1	1
587	646176653	3/10/2018	19:39	Male	Park lane	Dallas	Yes	Sporting	1	1

In [ ]:
```
```

In [163…
```
sales.mean()["Total"]
```

```
C:\Users\janna\AppData\Local\Temp\ipykernel_7348\2507315148.py:1: FutureWarning: D
ropping of nuisance columns in DataFrame reductions (with 'numeric_only=None') is
deprecated; in a future version this will raise TypeError.  Select only valid colu
mns before calling the reduction.
  sales.mean()["Total"]
```

Out[163]:
```
207.788
```

In [164…
```
sales[sales["Total"]== sales.min()["Total"]]
```

Out[164]:

	Invoice ID	Date	Time	Gender	Location	City	Member	Category	Price	Quantity
184	801334050	1/21/2018	20:18	Female	Park lane	Dallas	No	Furniture	1	1
587	646176653	3/10/2018	19:39	Male	Park lane	Dallas	Yes	Sporting	1	1

In [165…
```
sales.groupby('Gender').sum()
```

Out[165]:

Gender	Invoice ID	Price	Quantity	Total	Rating	Day	Month
Female	253138028847	25474	2069	106452	1517	7700	1012
Male	241355024494	25638	1990	101336	1455	7556	981

In [166…
```
sales.groupby('Gender').sum()['Total']
```

Out[166]:
```
Gender
Female    106452
Male      101336
Name: Total, dtype: int64
```

In [167…
```
sales.groupby('Payment').sum()['Total']
```

Out[167]:
```
Payment
Card    64264
Cash    72643
Gpay    70881
Name: Total, dtype: int64
```

In [168…
```
sales.groupby('City').sum()['Total']
```

Out[168]:
```
City
Chicago    65215
Dallas     70432
NewYork    72141
Name: Total, dtype: int64
```

In [169…
```
sales.max()['City']
```

Out[169]:
```
'NewYork'
```

In [171…
```
li=sales.groupby('Location')
```

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

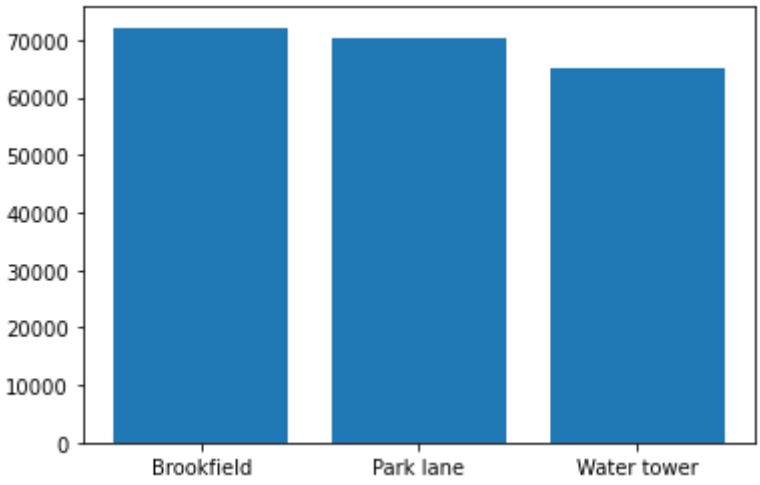
In [172…
```
lo=[x for x,y in li]
```

In [173…
```
lo
```

Out[173]:
```
['Brookfield', 'Park lane', 'Water tower']
```

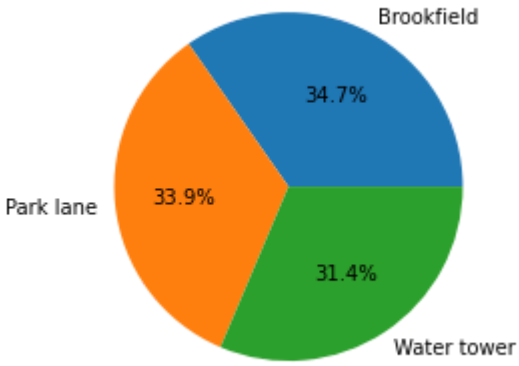
In [174…
```
plt.bar(lo,sales.groupby('Location').sum()['Total'])
```

Out[174]:
```
<BarContainer object of 3 artists>
```



In [175…
```
plt.pie(sales.groupby('Location').sum()['Total'],labels=lo,autopct='%1.1f%%')
```

Out[175]:
```
([<matplotlib.patches.Wedge at 0x26de0bd3160>,
  <matplotlib.patches.Wedge at 0x26de0bd3880>,
  <matplotlib.patches.Wedge at 0x26de0bd3fa0>],
 [Text(0.5080357843151303, 0.9756534435214743, 'Brookfield'),
  Text(-1.0939745592878725, -0.11497679605427102, 'Park lane'),
  Text(0.6072323961649048, -0.917207074246502, 'Water tower')],
 [Text(0.27711042780825285, 0.5321746055571678, '34.7%'),
  Text(-0.596713395975203, -0.06271461602960236, '33.9%'),
  Text(0.33121767063540253, -0.5002947677708193, '31.4%')])
```



In [ ]:
```
#which location have more male & female
```

In [176…
```
loca=sales.groupby(['Gender','Location']).count()['Invoice ID']
```

In [ ]:
```

```

In [177…
```
loca
```

Out[177]:
```
Gender  Location
Female  Brookfield     179
        Park lane      179
        Water tower    143
Male    Brookfield     161
        Park lane      153
        Water tower    185
Name: Invoice ID, dtype: int64
```

In [ ]:
```
unsta=loca.unstack(level=0)
```

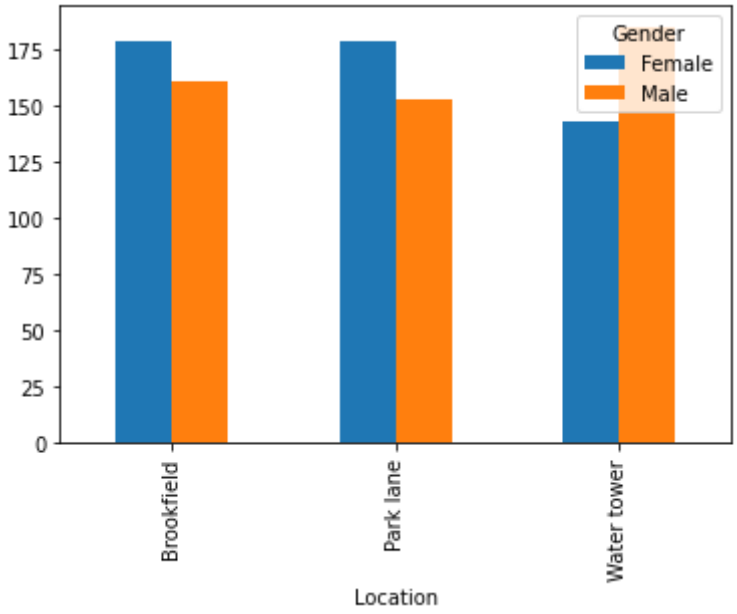
In [178…
```
unsta
```

Out[178]:

Gender	Female	Male
Location		
Brookfield	179	161
Park lane	179	153
Water tower	143	185

In [179…
```
unsta.plot(kind='bar')
```

Out[179]:
```
<AxesSubplot:xlabel='Location'>
```



In [ ]:
```
#what days make most sales?
```

In [180…
```
maxsale=sales.groupby(['Date']).max()['Total']
```

In [ ]:
```
maxsale
```

In [ ]:
```

```

In [ ]:
```
pd.to_datetime(sales['Date']).dt.day
```

In [ ]:
```
sales['Day']=pd.to_datetime(sales['Date']).dt.day
```

In [182…
```
sales.head(10)
```

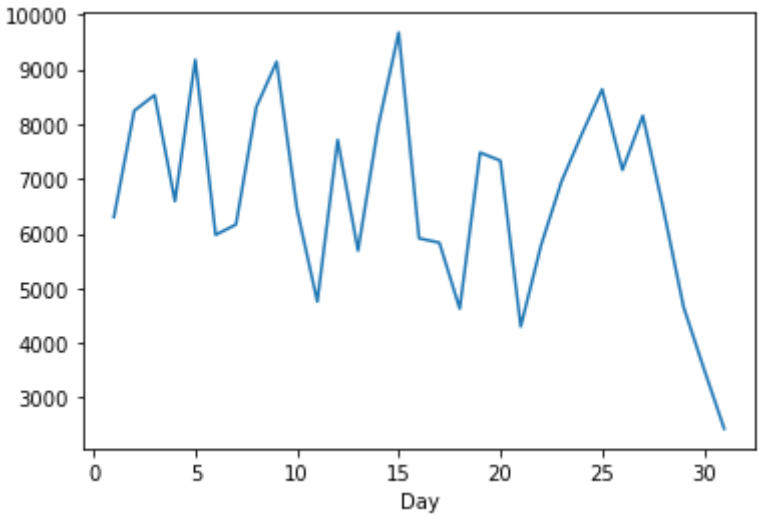
Out[182]:

	Invoice ID	Date	Time	Gender	Location	City	Member	Category	Price	Quantity
0	460489604	1/25/2018	16:46	Male	Brookfield	NewYork	Yes	Groceries	30	1
1	471006167	3/19/2018	16:48	Female	Water tower	Chicago	Yes	Fashion	35	5
2	411909258	2/25/2018	13:33	Male	Water tower	Chicago	No	Clothing	57	2
3	487313402	1/22/2018	13:38	Female	Park lane	Dallas	Yes	Sporting	89	4
4	197763430	2/18/2018	15:31	Female	Park lane	Dallas	No	Books	82	5
5	263634050	3/9/2018	17:55	Male	Brookfield	NewYork	Yes	Clothing	52	4
6	99646662	3/4/2018	13:21	Male	Water tower	Chicago	No	Clothing	81	3
7	188869875	3/8/2018	13:24	Male	Park lane	Dallas	No	Books	80	2
8	325637547	1/18/2018	15:33	Male	Park lane	Dallas	Yes	Clothing	91	3
9	562942936	2/24/2018	16:05	Female	Park lane	Dallas	No	Books	80	2

In [ ]:
```
saletot=sales.groupby(['Day']).sum()['Total']
```

In [183…
```
saletot.plot()
```

Out[183]:
```
<AxesSubplot:xlabel='Day'>
```



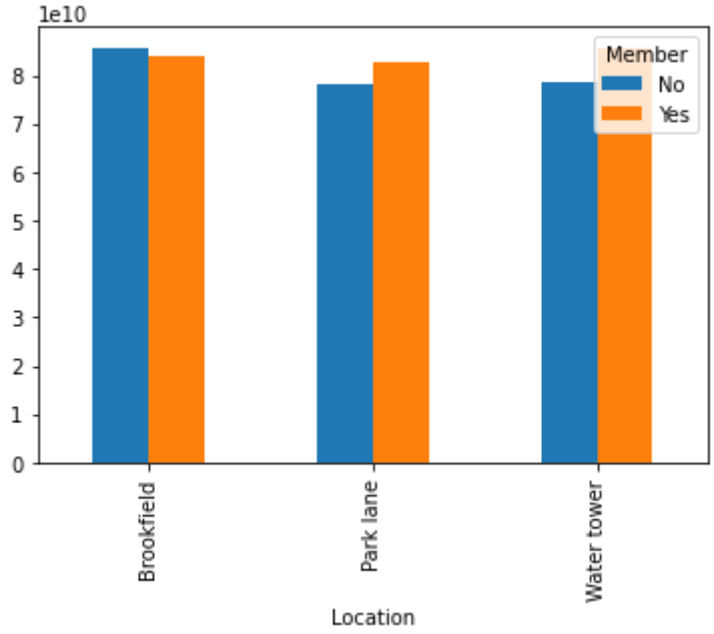
In [ ]:
```
#highest member location vs low member lpcation?
```

In [ ]:
```
memb=sales.groupby(['Member','Location']).sum()['Invoice ID']
```

In [ ]:
```
finalout=memb.unstack(level=0)
```

In [184…
```
finalout.plot(kind='bar')
```

Out[184]:
```
<AxesSubplot:xlabel='Location'>
```

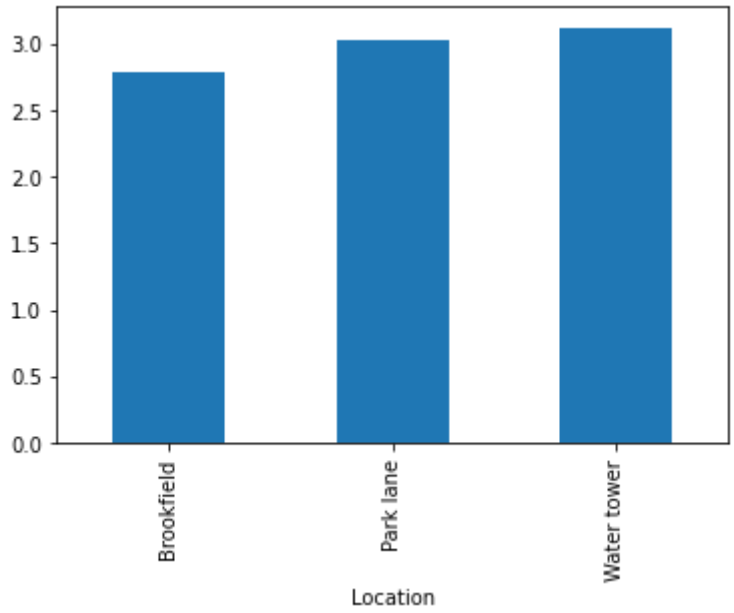


In [ ]:
```
#which have highest rating ?
```

In [ ]:
```
rate=sales.groupby(['Location']).mean()['Rating']
```

In [185…
```
rate.plot(kind="bar")
```

Out[185]:
```
<AxesSubplot:xlabel='Location'>
```

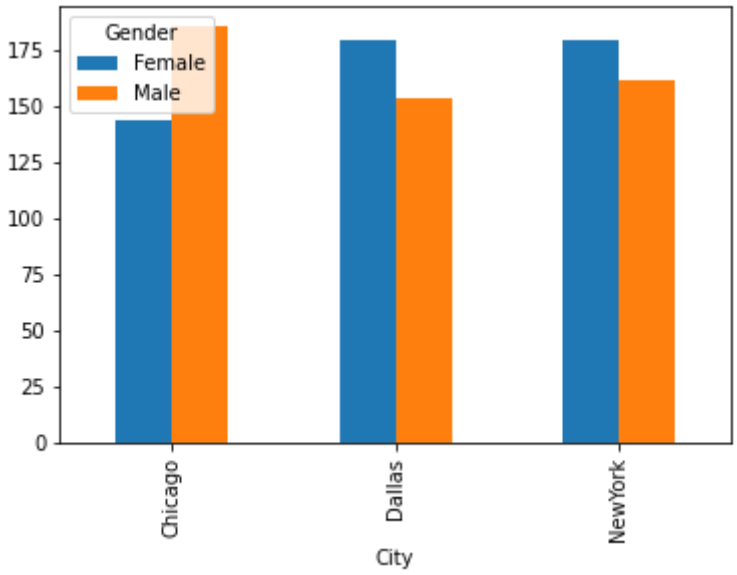


which city have more female shoping ?

In [ ]:
```
shop=sales.groupby(['Gender','City']).count()['Invoice ID']
```

In [186…
```
shop.unstack(level=0).plot(kind='bar')
```

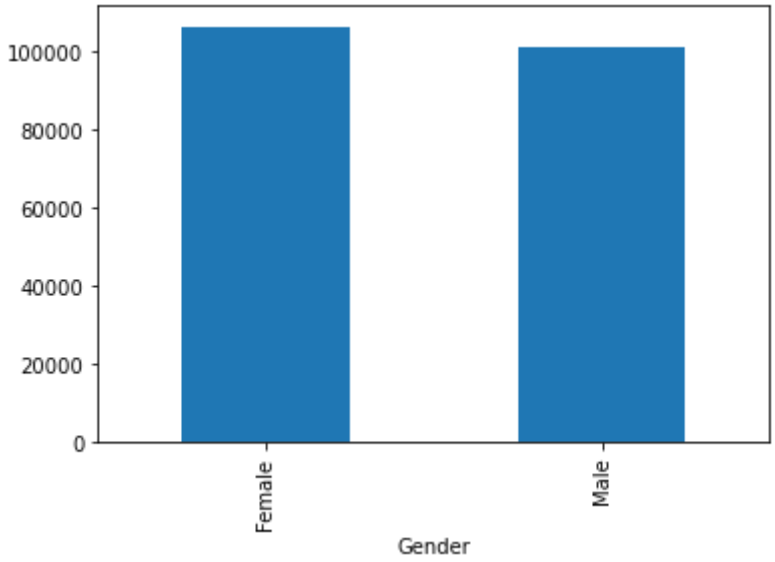
Out[186]:
```
<AxesSubplot:xlabel='City'>
```



In [ ]:
```
#who spens more?
```

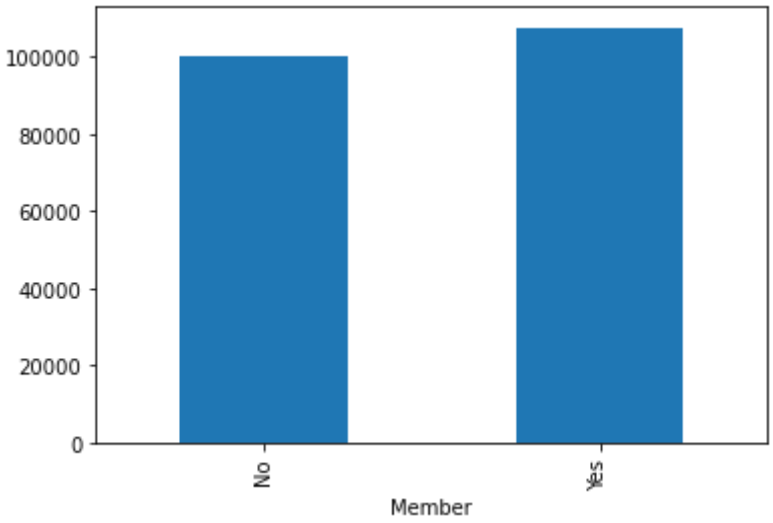
In [187…
```
sales.groupby(['Gender']).sum()['Total'].plot(kind='bar')
```

Out[187]:
```
<AxesSubplot:xlabel='Gender'>
```



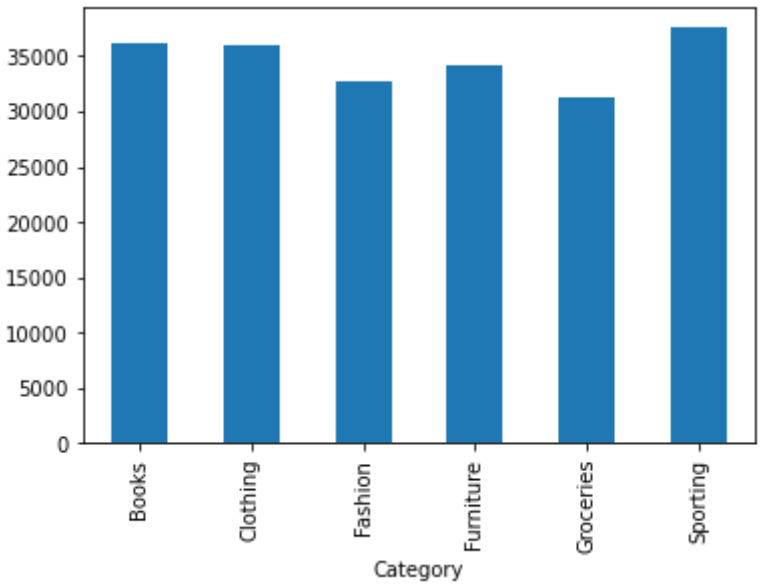
In [188…
```
sales.groupby(['Member']).sum()['Total'].plot(kind='bar')
```

Out[188]:
```
<AxesSubplot:xlabel='Member'>
```



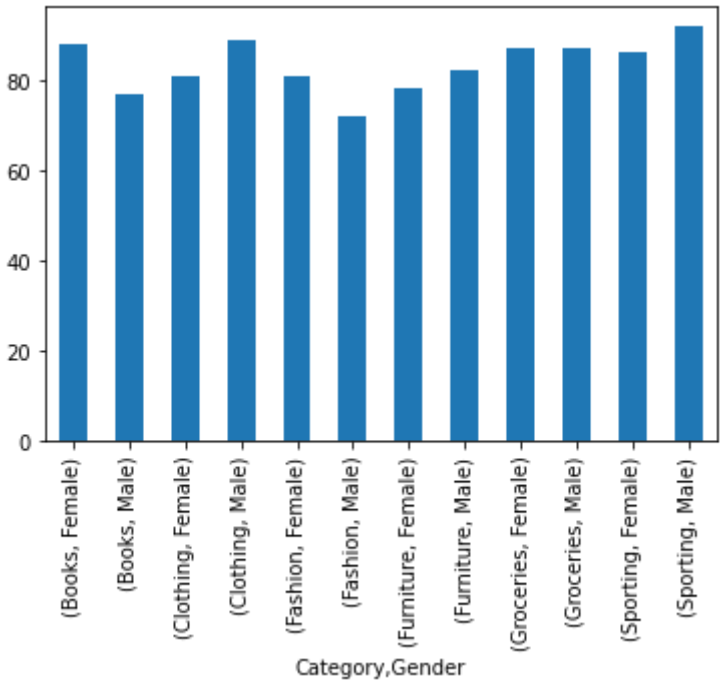
In [189…
```
sales.groupby(['Category']).sum()['Total'].plot(kind='bar')
```

Out[189]:
```
<AxesSubplot:xlabel='Category'>
```



In [190…
```
sales.groupby(['Category','Gender']).count()['Invoice ID'].plot(kind='bar')
```

Out[190]:
```
<AxesSubplot:xlabel='Category,Gender'>
```



In [191…
```
sales['Month']=pd.to_datetime(sales['Date']).dt.month
```

In [192…
```
sales.head()
```

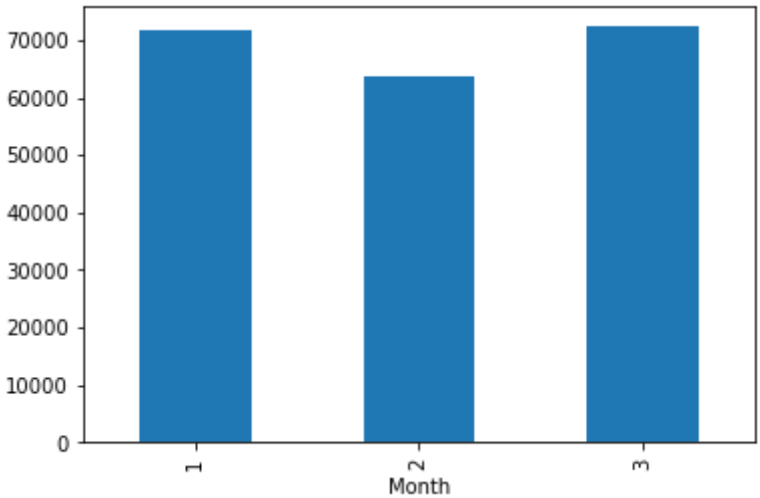
Out[192]:

	Invoice ID	Date	Time	Gender	Location	City	Member	Category	Price	Quantity
0	460489604	1/25/2018	16:46	Male	Brookfield	NewYork	Yes	Groceries	30	1
1	471006167	3/19/2018	16:48	Female	Water tower	Chicago	Yes	Fashion	35	5
2	411909258	2/25/2018	13:33	Male	Water tower	Chicago	No	Clothing	57	2
3	487313402	1/22/2018	13:38	Female	Park lane	Dallas	Yes	Sporting	89	4
4	197763430	2/18/2018	15:31	Female	Park lane	Dallas	No	Books	82	5

In [193…
```
salofm=sales.groupby(['Month']).sum()['Total']
```

In [194…
```
salofm.plot(kind='bar')
```

Out[194]:
```
<AxesSubplot:xlabel='Month'>
```



In [ ]: