Name: Atharva Suryakant Mali | Class: TY BCA | Roll No.: 04 | DS LAB

Experiment 2

Find the data distributions using box plot and scatter plot and also find the outliers using plot

a) Show the data distribution of total sales against each city in sales dataset using box plot. Also find out the outliers. ¶

In [8]:

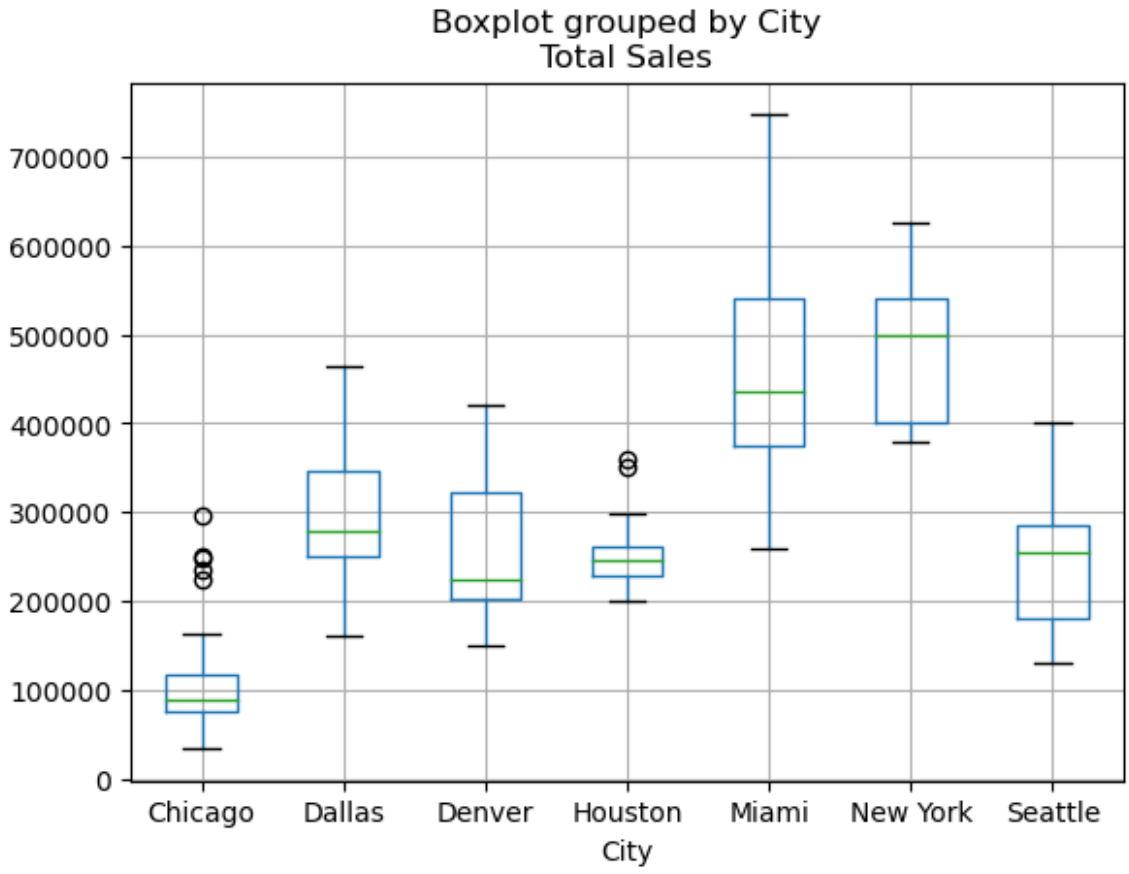
```
import matplotlib.pyplot as pt
import pandas as pd
ds=pd.read_excel('sales.xlsx')
ds.head()
```

Out[8]:

	Retailer	Retailer ID	Invoice Date	Region	State	City	Product	Price per Unit	UnitsSold	Total Sales	Op
0	Foot Locker	1185732	2020-01-01	Northeast	New York	New York	Men's Street Footwear	50.0	1200	600000.0	3
1	Foot Locker	1185732	2020-01-02	Northeast	New York	New York	Men's Athletic Footwear	50.0	1000	500000.0	1
2	Foot Locker	1185732	2020-01-03	Northeast	New York	New York	Women's Street Footwear	40.0	1000	400000.0	1
3	Foot Locker	1185732	2020-01-04	Northeast	New York	New York	Women's Athletic Footwear	45.0	850	382500.0	1
4	Foot Locker	1185732	2020-01-05	Northeast	New York	New York	Men's Apparel	60.0	900	540000.0	1

In [9]:

```
ds.boxplot(by='City',column=['Total Sales']);
```



b) Draw the boxplots for day wise total bill in tips dataset and also find out the outliers.

In [2]:

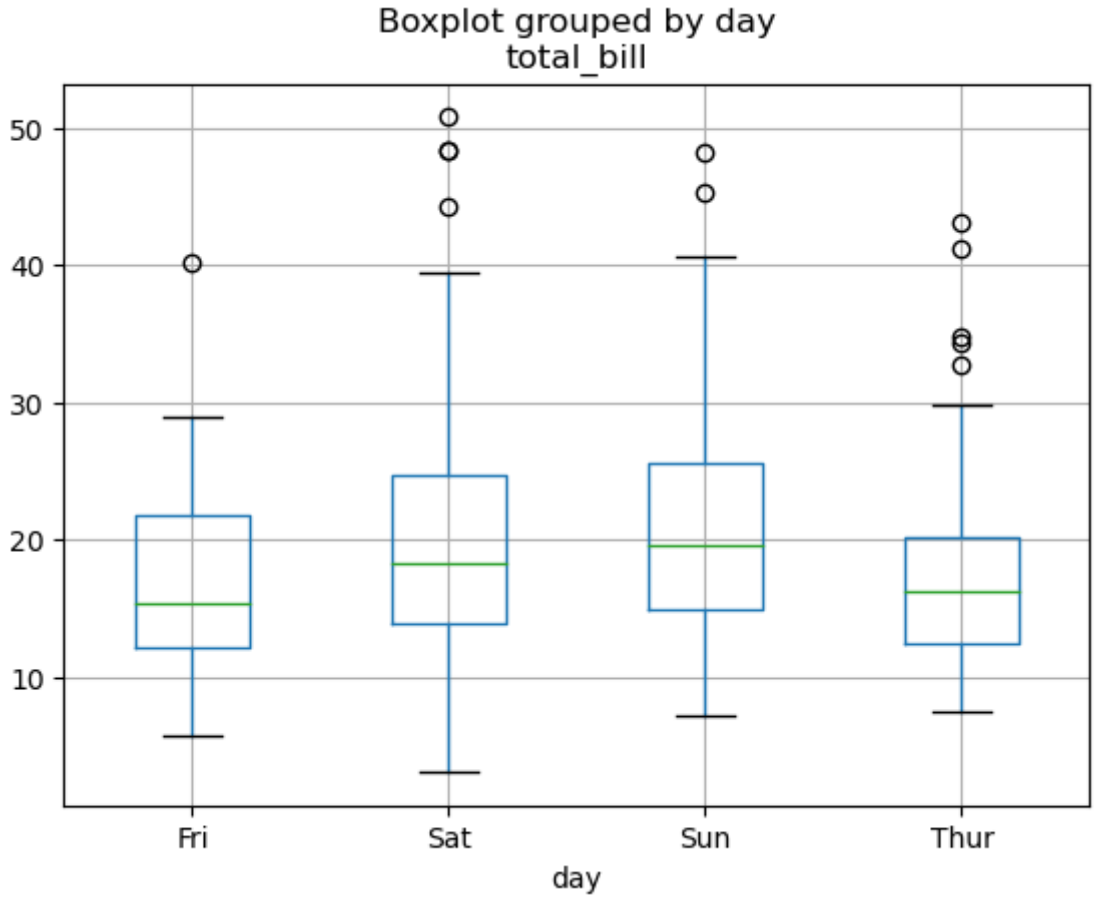
```
import matplotlib.pyplot as pt
import pandas as pd
df=pd.read_csv('tips.csv')
df.head()
```

Out[2]:

	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4

In [4]:

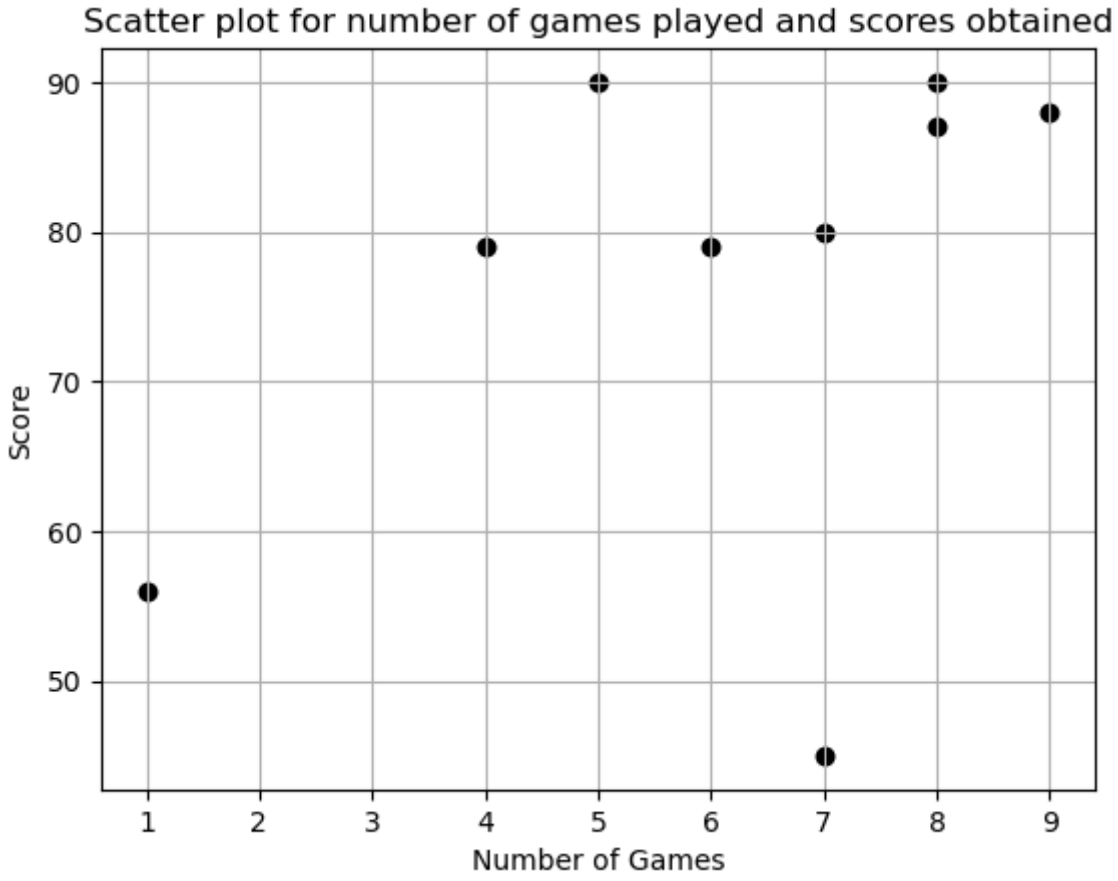
```
df.boxplot(by='day',column=['total_bill']);
```



c) Draw a scatter plot for given data that shows the number of games played and scores obtained. Write outlier data.

In [12]:

```
from matplotlib import pyplot as pt
x=[5,8,1,9,7,4,8,6,7]
y=[90,87,56,88,45,79,90,79,80]
pt.scatter(x,y,marker=".",s=150,c="k")
pt.title("Scatter plot for number of games played and scores obtained")
pt.xlabel("Number of Games")
pt.ylabel("Score")
pt.grid()
pt.show()
```



Outlier data in above scatter plot is

Points (1,56) and (7,45)

d) Draw a scatter plot for total bills against tips. Use tips dataset.

In [2]:

```
import pandas as pd
df=pd.read_csv('tips.csv')
df.head()
```

Out[2]:

	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4

In [14]:

```
from matplotlib import pyplot as pt
x=df['total_bill']
y=df['tip']
pt.scatter(x,y,marker=".",s=100,c="k")
pt.title("Scatter Plot for total_bill against tip")
pt.xlabel("Total Bill")
pt.ylabel("Tips")
pt.grid()
pt.show()
```