

```
import numpy as np
import pandas as pd
url="https://raw.githubusercontent.com/AmenaNajeeb/Data/master/bigmart_data.csv"
df=pd.read_csv(url)
df.head(10)
df.isnull().sum()
```

```
Item_Identifier      0
Item_Weight          1463
Item_Fat_Content      0
Item_Visibility      0
Item_Type            0
Item_MRP             0
Outlet_Identifier     0
Outlet_Establishment_Year  0
Outlet_Size          2410
Outlet_Location_Type  0
Outlet_Type           0
Item_Outlet_Sales     0
dtype: int64
```

```
weight_null_random_list = np.random.randint(0,20.000000)
df['Item_Weight'][np.isnan(df['Item_Weight'])] = weight_null_random_list
df['Item_Weight'] = df['Item_Weight'].astype(float)

df['Item_Weight']
```

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame
```

See the caveats in the documentation: <https://pandas.pydata.org/pandas-docs/stable/user>

```
0      9.300
1      5.920
2     17.500
3     19.200
4      8.930
...
8518    6.865
8519    8.380
8520   10.600
8521    7.210
8522   14.800
Name: Item_Weight, Length: 8523, dtype: float64
```

```
df.isnull().sum()
df.head(20)
```

```
df['Outlet_Size']=df['Outlet_Size'].fillna('Medium')
df.isnull().sum()
```

```

Item_Identifier      0
Item_Weight          0
Item_Fat_Content     0
Item_Visibility      0
Item_Type            0
Item_MRP             0
Outlet_Identifier    0
Outlet_Establishment_Year  0
Outlet_Size          0
Outlet_Location_Type 0
Outlet_Type          0
Item_Outlet_Sales    0
dtype: int64

```

```
del df['Item_Identifier']
```

```
del df['Outlet_Identifier']
```

```
df=df.dropna("Outlet_Size"== "NaN")
df ["Outlet_Size"]
```

```

0      Medium
1      Medium
2      Medium
4      High
5      Medium
...
8517   Small
8518   High
8520   Small
8521   Medium
8522   Small
Name: Outlet_Size, Length: 4650, dtype: object

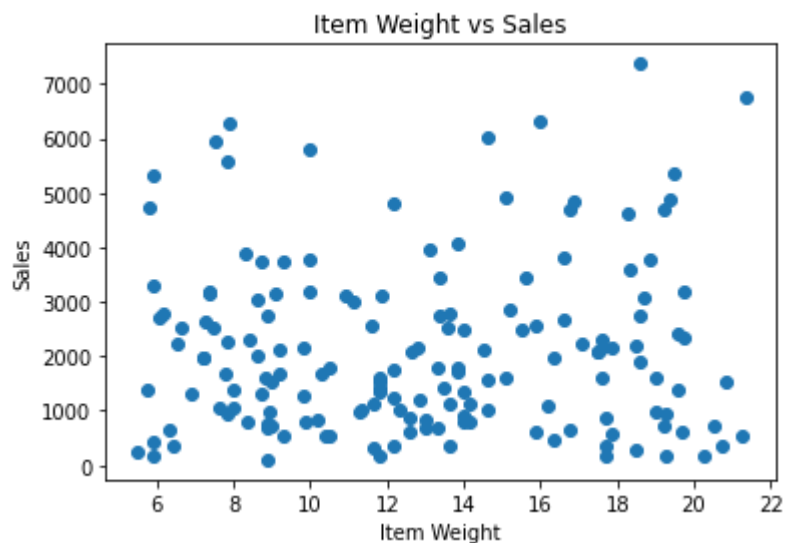
```

```

import matplotlib.pyplot as plt
plt.scatter(df["Item_Weight"][:200],df["Item_Outlet_Sales"][:200])
plt.xlabel('Item Weight')
plt.ylabel('Sales')
plt.title('Item Weight vs Sales')

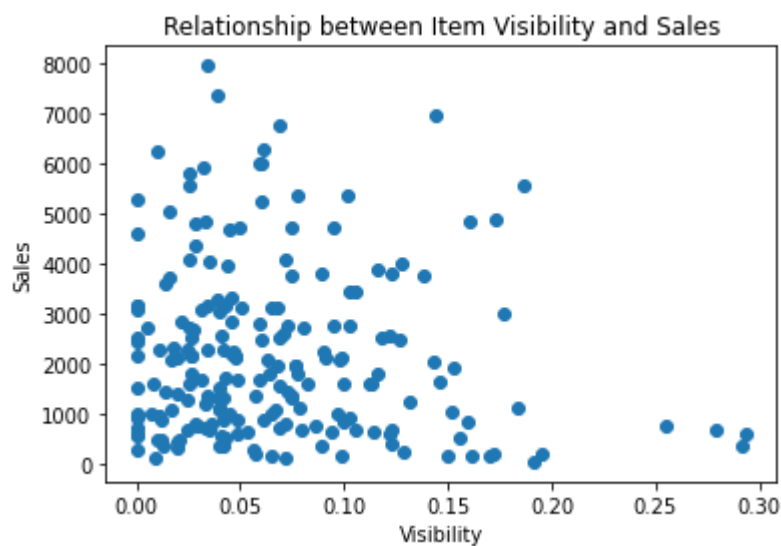
```

```
Text(0.5, 1.0, 'Item Weight vs Sales')
```



```
plt.scatter(df['Item_Visibility'][:200],df['Item_Outlet_Sales'][:200])
plt.title('Relationship between Item Visibility and Sales')
plt.xlabel('Visibility')
plt.ylabel('Sales')
```

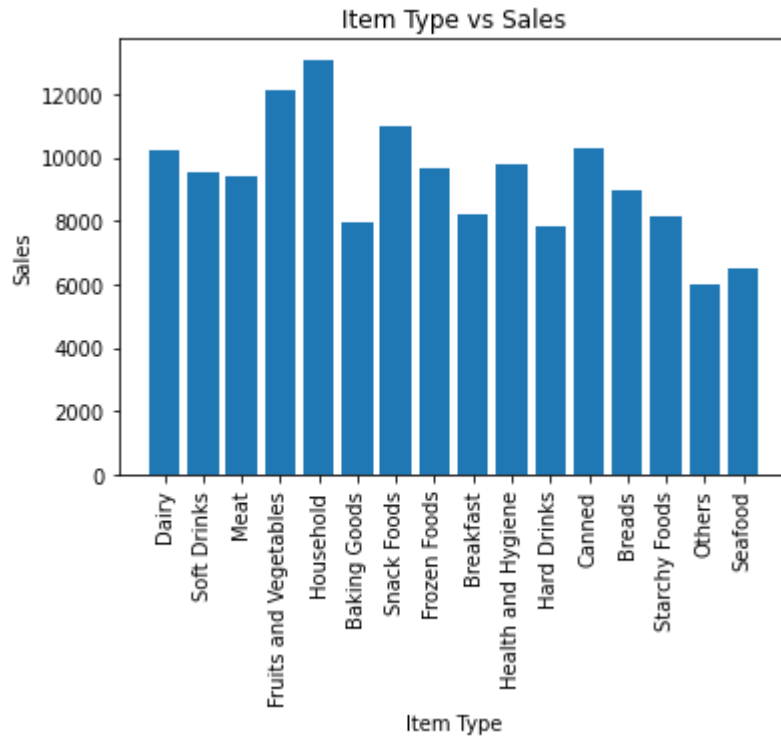
```
Text(0, 0.5, 'Sales')
```



```
plt.bar(df['Item_Type'],df['Item_Outlet_Sales'])
plt.title('Item Type vs Sales')
```

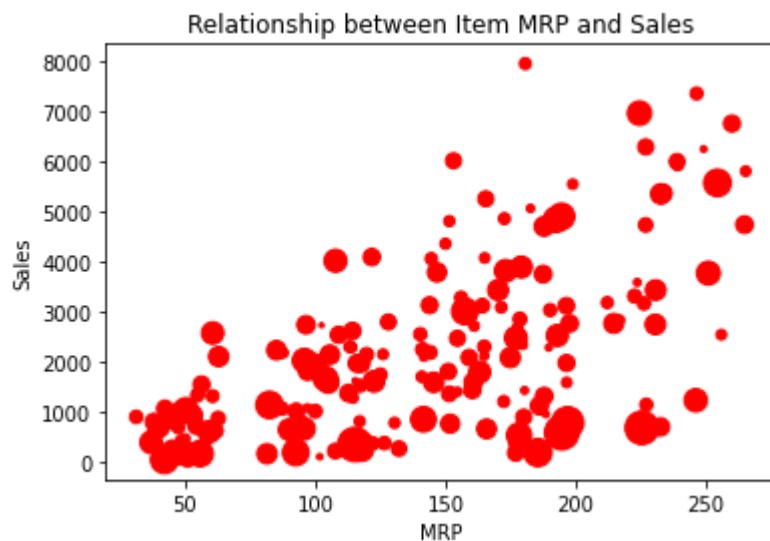
```
plt.xlabel('Item Type')
plt.ylabel('Sales')
plt.xticks(rotation="vertical")
```

```
([0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15],
 <a list of 16 Text major ticklabel objects>)
```



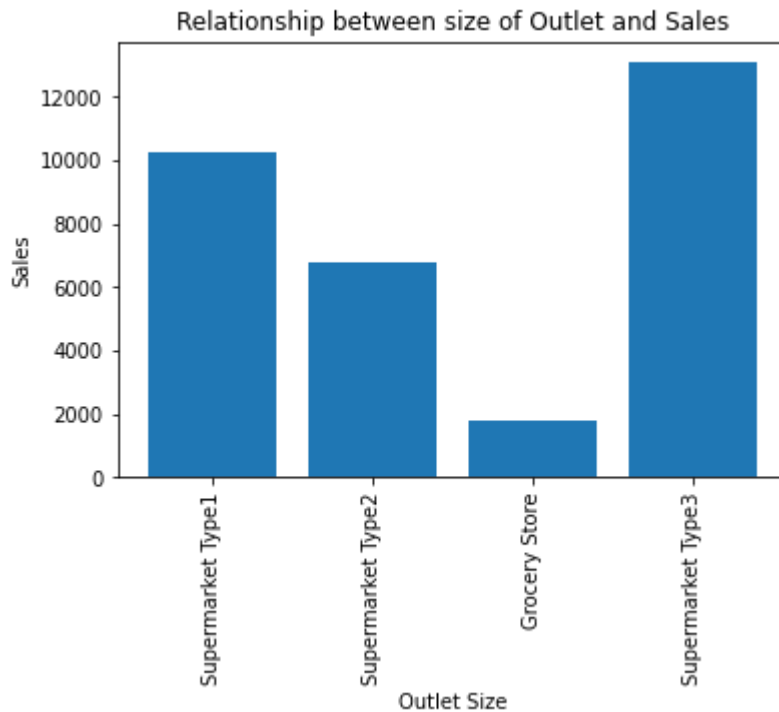
```
plt.scatter(df["Item_MRP"][:200],df["Item_Outlet_Sales"][:200],s=df["Item_Visibility"][:200]*
plt.title('Relationship between Item MRP and Sales')
plt.xlabel('MRP')
plt.ylabel('Sales')
```

```
Text(0, 0.5, 'Sales')
```



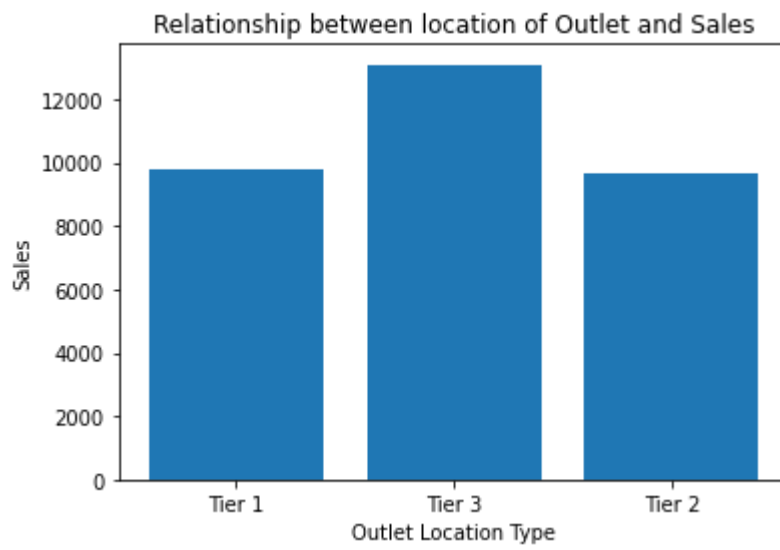
```
plt.bar(df['Outlet_Type'],df['Item_Outlet_Sales'])  
plt.title('Relationship between size of Outlet and Sales')  
plt.xlabel('Outlet Size')  
plt.ylabel('Sales')  
plt.xticks(rotation="vertical")
```

➞ ([0, 1, 2, 3], <a list of 4 Text major ticklabel objects>)



```
plt.bar(df['Outlet_Location_Type'],df['Item_Outlet_Sales'])  
plt.title('Relationship between location of Outlet and Sales')  
plt.xlabel('Outlet Location Type')  
plt.ylabel('Sales')
```

```
Text(0, 0.5, 'Sales')
```



✓ 0s completed at 7:35 PM

