WEEK 1 TASKS:

NAME: Javeria Iqbal

1.SOURCE CODE:

```
age=25
print("age",age)
float=19.99
print("float")
name="javeria"
print("name", name)
student= True
print("Student",student)
fruits=["apple","banana","mango"]
print("Fruits",fruits)
student={"name":"ali","age":20,"grade":"A"}
print("student",student)
```

OUTPUT:

```
ng/loopw1.py
NUMBER 0
NUMBER 2
NUMBER 3
NUMBER 4
count is 0
positive number
count is 1
positive number
count is 2
positive number
```

2. SOURCE CODE AND OUTPUT(JUPYTER NOTEBOOK):

2]:
```
import pandas as pd
```

5]:
```
df=pd.read_csv("patient_data.csv")
```

6]:
```
df
```

6]:

	PatientID	Gender	Age	CholesterolLevel	BloodPressure	Region	RiskCategory
0	1	Male	37	193.4	129.4	East	High
1	2	Female	45	210.7	107.1	West	Low
2	3	Male	63	244.3	103.9	West	Low
3	4	Male	53	184.5	127.2	West	High
4	5	Male	29	175.7	116.6	South	High
...	...	...	...	...	...	...	...
95	96	Female	58	224.8	126.9	East	High
96	97	Female	68	200.4	123.0	North	Medium
97	98	Female	71	243.6	111.0	West	Low
98	99	Female	51	192.1	121.0	West	High
99	100	Male	23	281.6	114.2	North	High

100 rows × 7 columns

7]:
```
print("data info")
print(df.info())
```

```
data info
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100 entries, 0 to 99
Data columns (total 7 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
```

```
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   PatientID         100 non-null    int64
 1   Gender            100 non-null    object
 2   Age               100 non-null    int64
 3   CholesterolLevel  100 non-null    float64
 4   BloodPressure     100 non-null    float64
 5   Region            100 non-null    object
 6   RiskCategory      100 non-null    object
dtypes: float64(2), int64(2), object(3)
memory usage: 5.6+ KB
None
```

```
print("missing values")
print(df.isnull().sum())
```

```
missing values
PatientID           0
Gender              0
Age                 0
CholesterolLevel    0
BloodPressure       0
Region              0
RiskCategory        0
dtype: int64
```

```
df=df.drop_duplicates()
print("\nAFter REmoving dulicates",df)
```

```
AFter REmoving dulicates     PatientID  Gender  Age  CholesterolLevel  BloodPressure Region  \
0            1    Male   37             193.4          129.4   East
1            2  Female   45             210.7          107.1   West
2            3    Male   63             244.3          103.9   West
3            4    Male   53             184.5          127.2   West
4            5    Male   29             175.7          116.6  South
..         ...     ...  ...               ...            ...    ...
95          96  Female   58             224.8          126.9   East
96          97  Female   68             200.4          123.0  North
97          98  Female   71             243.6          111.0   West
98          99  Female   51             192.1          121.0   West
99         100    Male   23             281.6          114.2  North
```

```
99         100    Male   23             281.6          114.2  North

   RiskCategory
0          High
1           Low
2           Low
3          High
4          High
..          ...
95         High
96       Medium
97          Low
98         High
99         High

[100 rows x 7 columns]
```

```
df=df.dropna(subset=['PatientID','Age'])
```

```
df
```

	PatientID	Gender	Age	CholesterolLevel	BloodPressure	Region	RiskCategory
0	1	Male	37	193.4	129.4	East	High
1	2	Female	45	210.7	107.1	West	Low
2	3	Male	63	244.3	103.9	West	Low
3	4	Male	53	184.5	127.2	West	High
4	5	Male	29	175.7	116.6	South	High
...	...	...	...	...	...	...	...
95	96	Female	58	224.8	126.9	East	High
96	97	Female	68	200.4	123.0	North	Medium
97	98	Female	71	243.6	111.0	West	Low
98	99	Female	51	192.1	121.0	West	High

98	99	Female	51	192.1	121.0	West	High
99	100	Male	23	281.6	114.2	North	High

100 rows × 7 columns

```
numeric_cols=['Age','CholesterolLevel','BloodPressure']
for col in numeric_cols:
    df[col]=df[col].fillna(df[col].mean())
```

```
numeric_cols
```

```
['Age', 'CholesterolLevel', 'BloodPressure']
```

```
catogorial_cols=['Gender','Region','RiskCategory']
for col in catogorial_cols:
    df[col]=df[col].fillna(df[col].mode()[0])
```

```
catogorial_cols
```

```
['Gender', 'Region', 'RiskCategory']
```

```
df['Age']=df['Age'].astype(int)
```

```
df
```

	PatientID	Gender	Age	CholesterolLevel	BloodPressure	Region	RiskCategory
0	1	Male	37	193.4	129.4	East	High
1	2	Female	45	210.7	107.1	West	Low
2	3	Male	63	244.3	103.9	West	Low
3	4	Male	53	184.5	127.2	West	High
4	5	Male	29	175.7	116.6	South	High

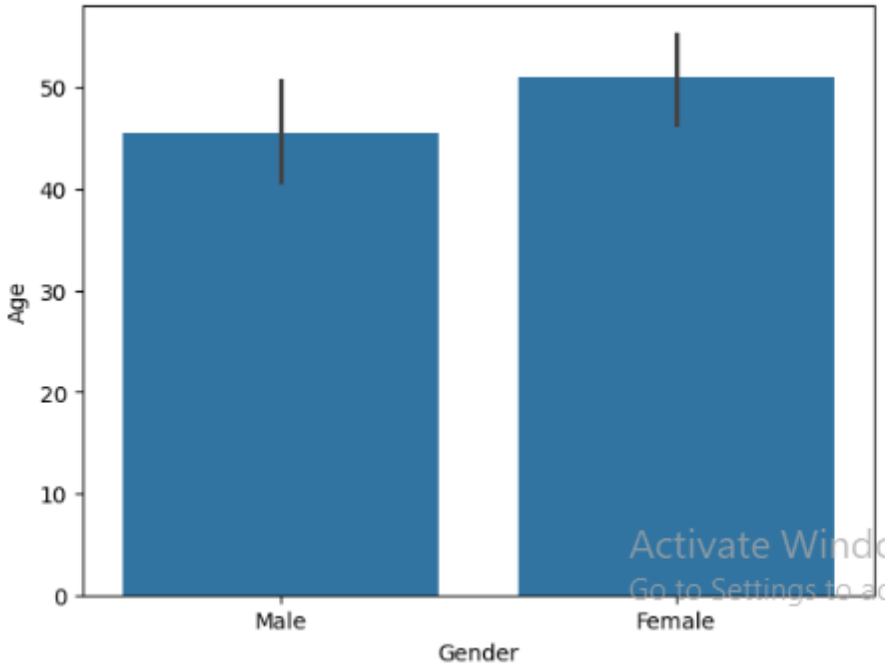
3.

dtype: int64

[1]:
```
import seaborn as sns
```

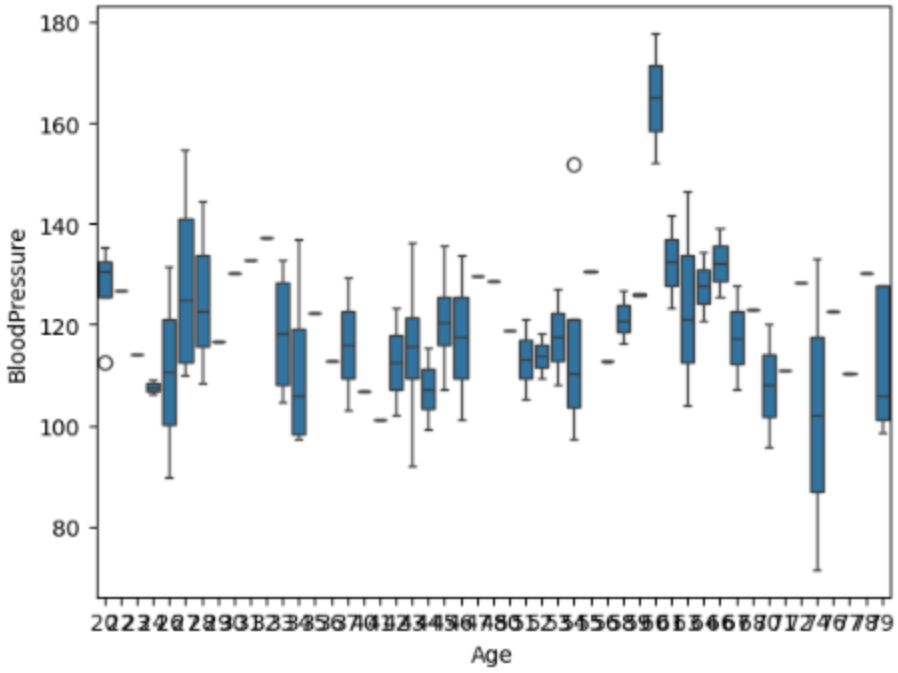
[9]:
```
sns.barplot(x="Gender",y="Age",data=df)
plt.title("Gender by Age")
plt.show()
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
Cell In[9], line 2
      1 sns.barplot(x="Gender",y="Age",data=df)
----> 2 plt.title("Gender by Age")
      3 plt.show()

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



```
sns.boxplot(x="Age", y="BloodPressure", data=df)
plt.title("Total Bill Distribution by Day")
plt.show()
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
Cell In[14], line 2
      1 sns.boxplot(x="Age", y="BloodPressure", data=df)
----> 2 plt.title("Total Bill Distribution by Day")
      3 plt.show()

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



98	99	Female	51	192.1	121.0	West	High
99	100	Male	23	281.6	114.2	North	High

100 rows × 7 columns

]:
```
numeric_cols=['Age','CholesterolLevel','BloodPressure']
for col in numeric_cols:
    df[col]=df[col].fillna(df[col].mean())
```

]:
```
numeric_cols
```

]:
```
['Age', 'CholesterolLevel', 'BloodPressure']
```

]:
```
catogorial_cols=['Gender','Region','RiskCategory']
for col in catogorial_cols:
    df[col]=df[col].fillna(df[col].mode()[0])
```

]:
```
catogorial_cols
```

]:
```
['Gender', 'Region', 'RiskCategory']
```

]:
```
df['Age']=df['Age'].astype(int)
```

]:
```
df
```

]:

	PatientID	Gender	Age	CholesterolLevel	BloodPressure	Region	RiskCategory
0	1	Male	37	193.4	129.4	East	High
1	2	Female	45	210.7	107.1	West	Low
2	3	Male	63	244.3	103.9	West	Low
3	4	Male	53	184.5	127.2	West	High
4	5	Male	29	175.7	116.6	South	High

2	3	Male	63	244.3	103.9	West	Low
3	4	Male	53	184.5	127.2	West	High
4	5	Male	29	175.7	116.6	South	High
...	...	...	...	...	...	...	...
95	96	Female	58	224.8	126.9	East	High
96	97	Female	68	200.4	123.0	North	Medium
97	98	Female	71	243.6	111.0	West	Low
98	99	Female	51	192.1	121.0	West	High
99	100	Male	23	281.6	114.2	North	High

100 rows × 7 columns

25]:
```
print(df.isnull().sum())
```

```
PatientID          0
Gender             0
Age                0
CholesterolLevel   0
BloodPressure      0
Region             0
RiskCategory       0
dtype: int64
```

[ ]: