TASK 1

Create a bar chart or histogram to visualize the distribution ofa categorical or continuous variable, such as the distribution ofages or genders in a population.

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

df = pd.read_csv("C:\\Users\\hp\\Internship\\healthcare_dataset.csv")

df.head(5)
```

```
             Name  Age  Gender Blood Type Medical Condition Date of
Admission  \
0  Bobby JacksOn   30    Male         B-            Cancer
2024-01-31
1   LesLie TErRy   62    Male         A+           Obesity
2019-08-20
2    DaNnY sMitH   76  Female         A-           Obesity
2022-09-22
3   andrEw waTtS   28  Female         O+          Diabetes
2020-11-18
4  adrIENNE bEll   43  Female        AB+            Cancer
2022-09-19

              Doctor                     Hospital Insurance Provider  \
0     Matthew Smith             Sons and Miller         Blue Cross
1   Samantha Davies                      Kim Inc           Medicare
2  Tiffany Mitchell                    Cook PLC              Aetna
3       Kevin Wells  Hernandez Rogers and Vang,           Medicare
4    Kathleen Hanna                 White-White              Aetna

   Billing Amount  Room Number Admission Type Discharge Date
Medication  \
0    18856.281306          328         Urgent     2024-02-02
Paracetamol
1    33643.327287          265      Emergency     2019-08-26
Ibuprofen
2    27955.096079          205      Emergency     2022-10-07
Aspirin
3    37909.782410          450       Elective     2020-12-18
Ibuprofen
4    14238.317814          458         Urgent     2022-10-09
Penicillin

   Test Results
0        Normal
1  Inconclusive
```

```
2        Normal
3      Abnormal
4      Abnormal
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 55500 entries, 0 to 55499
Data columns (total 15 columns):
 #   Column              Non-Null Count  Dtype  
---  ------              --------------  -----  
 0   Name                55500 non-null  object 
 1   Age                 55500 non-null  int64  
 2   Gender              55500 non-null  object 
 3   Blood Type          55500 non-null  object 
 4   Medical Condition   55500 non-null  object 
 5   Date of Admission   55500 non-null  object 
 6   Doctor              55500 non-null  object 
 7   Hospital            55500 non-null  object 
 8   Insurance Provider  55500 non-null  object 
 9   Billing Amount      55500 non-null  float64
 10  Room Number         55500 non-null  int64  
 11  Admission Type      55500 non-null  object 
 12  Discharge Date      55500 non-null  object 
 13  Medication          55500 non-null  object 
 14  Test Results        55500 non-null  object 
dtypes: float64(1), int64(2), object(12)
memory usage: 6.4+ MB
```

```
df.describe()
```

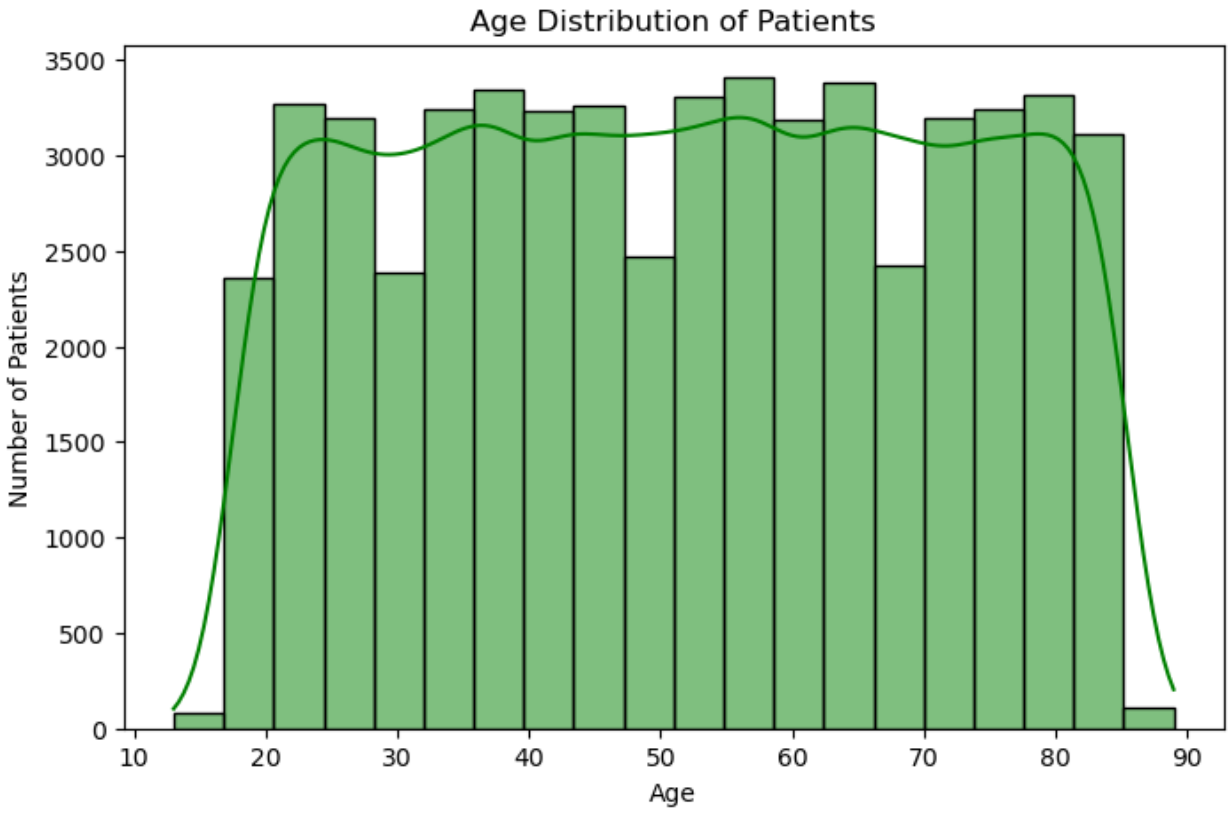
```
                Age  Billing Amount   Room Number
count  55500.000000    55500.000000  55500.000000
mean      51.539459    25539.316097    301.134829
std       19.602454    14211.454431    115.243069
min       13.000000    -2008.492140    101.000000
25%       35.000000    13241.224652    202.000000
50%       52.000000    25538.069376    302.000000
75%       68.000000    37820.508436    401.000000
max       89.000000    52764.276736    500.000000
```

```
df.isna().sum()
```

```
Name                  0
Age                   0
Gender                0
Blood Type            0
Medical Condition     0
Date of Admission     0
Doctor                0
Hospital              0
```

```
Insurance Provider    0
Billing Amount        0
Room Number           0
Admission Type        0
Discharge Date        0
Medication            0
Test Results          0
dtype: int64
```

```
# Histogram for Age
plt.figure(figsize=(8,5))
sns.histplot(df['Age'], bins=20, kde=True, color='green')  #parameter
in histogram decides how finely you divide the data range
plt.title('Age Distribution of Patients')                  #adds a
smooth density curve.
plt.xlabel('Age')
plt.ylabel('Number of Patients')
plt.show()
```




```
# Bar chart for Blood Type
plt.figure(figsize=(6,4))
sns.countplot(x='Blood Type', data=df)
plt.title('Blood Type Distribution')
plt.xlabel('Blood Type')
```

```
plt.ylabel('Count')
plt.show()
```