

```

In [13]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

file_path = r'C:\Users\was11\Music\anaconda\healthcare-dataset-stroke-data.csv'
df = pd.read_csv(file_path)

print(df.head())

print("Missing values in each column:")
print(df.isnull().sum())

for column in df.select_dtypes(include=['float64', 'int64']).columns:
    mean_value = df[column].mean()
    df[column] = df[column].fillna(mean_value)

for column in df.select_dtypes(include=['object']).columns:
    mode_value = df[column].mode()[0]
    df[column] = df[column].fillna(mode_value)

print("After handling missing values:")
print(df.isnull().sum())

def detect_outliers_iqr(column):
    Q1 = column.quantile(0.25)
    Q3 = column.quantile(0.75)
    IQR = Q3 - Q1
    lower_bound = Q1 - 1.5 * IQR
    upper_bound = Q3 + 1.5 * IQR
    return (column < lower_bound) | (column > upper_bound)

for column in df.select_dtypes(include=['float64', 'int64']).columns:
    df[column] = df[column].where(~detect_outliers_iqr(df[column]), np.nan)

for column in df.select_dtypes(include=['float64', 'int64']).columns:
    mean_value = df[column].mean()
    df[column] = df[column].fillna(mean_value)

print("After handling outliers:")
print(df.isnull().sum())

plt.figure(figsize=(10, 8))
numeric_df = df.select_dtypes(include=['float64', 'int64'])
sns.heatmap(numeric_df.corr(), annot=True, cmap='viridis', annot_kws={"size": 12, "weight": "bold"})
plt.title('CORRELATION HEATMAP', fontsize=16, weight='bold')
plt.xticks(rotation=45, ha='right', fontsize=12, weight='bold')
plt.yticks(fontsize=12, weight='bold')
plt.show()

plt.figure(figsize=(8, 5))
sns.boxplot(x=df['age'])
plt.title('Boxplot for Age', fontsize=16, weight='bold')
plt.show()

plt.figure(figsize=(8, 5))
sns.countplot(x='gender', data=df)
plt.title('Gender Distribution', fontsize=16, weight='bold')
plt.show()

output_file_path = r'C:\Users\was11\Music\anaconda\modified_healthcare_data.csv'
df.to_csv(output_file_path, index=False)

print("Modified data saved to CSV file.")

```

	id	gender	age	hypertension	heart_disease	ever_married	\
0	9046	Male	67.0	0	1	Yes	
1	51676	Female	61.0	0	0	Yes	
2	31112	Male	80.0	0	1	Yes	
3	60182	Female	49.0	0	0	Yes	
4	1665	Female	79.0	1	0	Yes	

	work_type	Residence_type	avg_glucose_level	bmi	smoking_status	\
0	Private	Urban	228.69	36.6	formerly smoked	
1	Self-employed	Rural	202.21	NaN	never smoked	
2	Private	Rural	105.92	32.5	never smoked	
3	Private	Urban	171.23	34.4	smokes	
4	Self-employed	Rural	174.12	24.0	never smoked	

	stroke
0	1
1	1
2	1
3	1
4	1

Missing values in each column:

id	0
gender	0
age	0
hypertension	0
heart_disease	0
ever_married	0
work_type	0
Residence_type	0
avg_glucose_level	0
bmi	201
smoking_status	0
stroke	0

dtype: int64

After handling missing values:

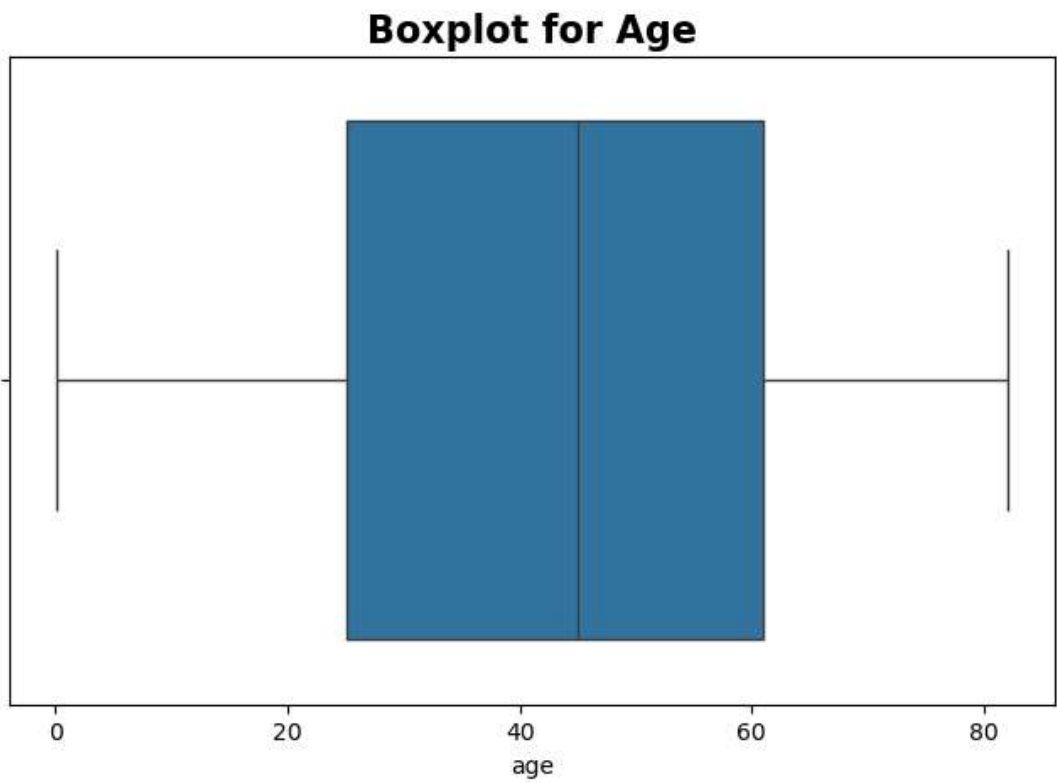
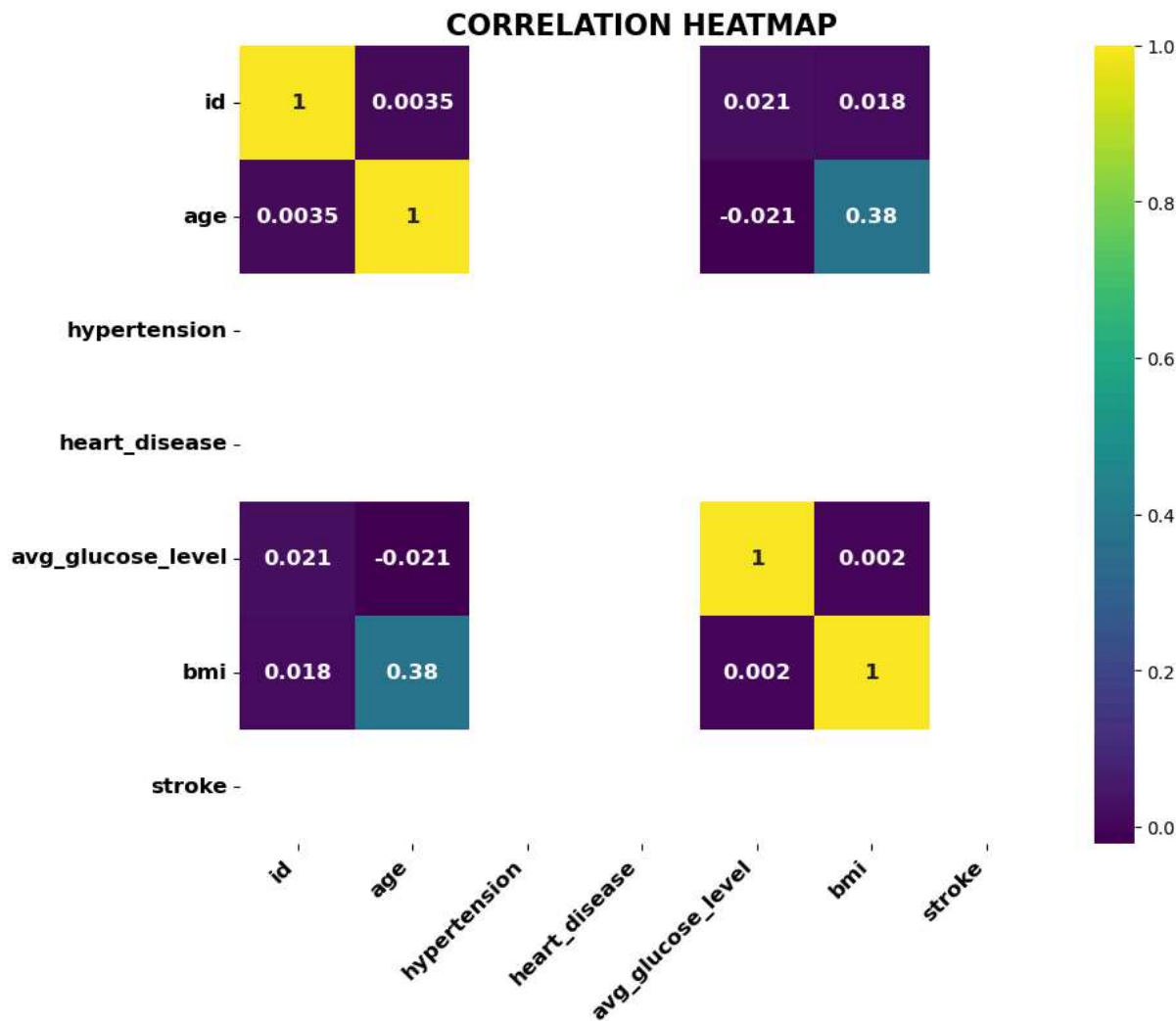
id	0
gender	0
age	0
hypertension	0
heart_disease	0
ever_married	0
work_type	0
Residence_type	0
avg_glucose_level	0
bmi	0
smoking_status	0
stroke	0

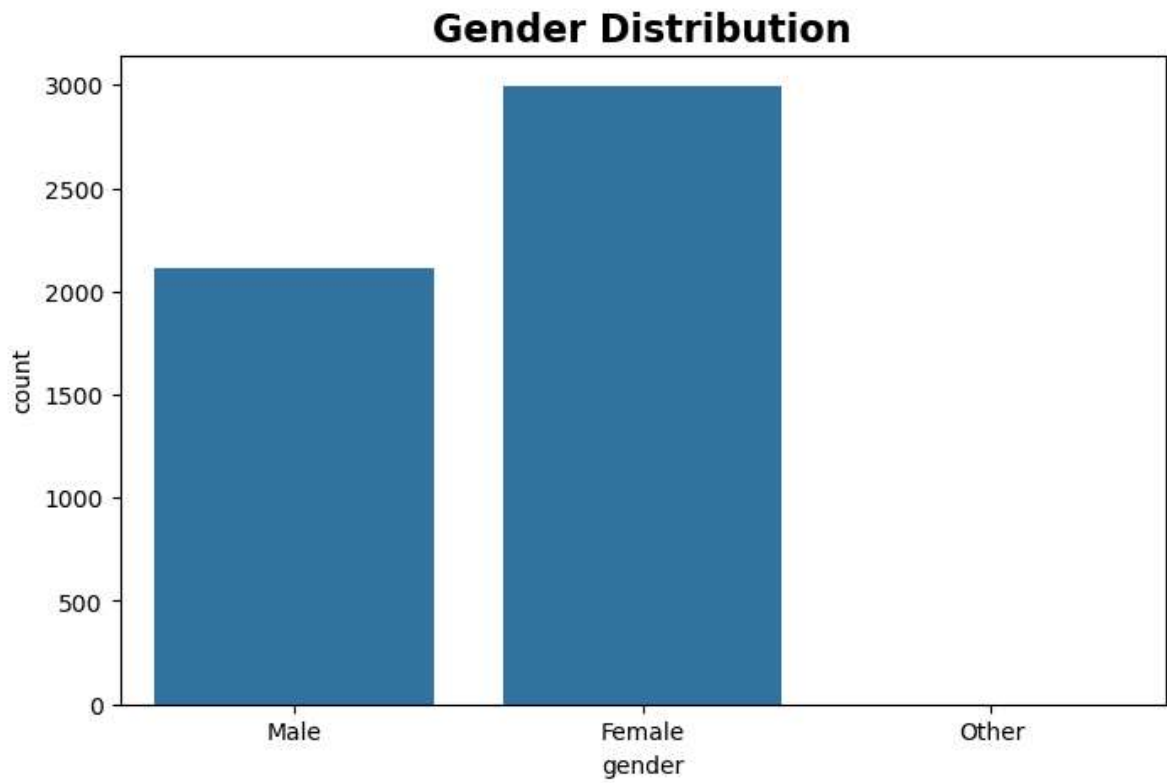
dtype: int64

After handling outliers:

id	0
gender	0
age	0
hypertension	0
heart_disease	0
ever_married	0
work_type	0
Residence_type	0
avg_glucose_level	0
bmi	0
smoking_status	0
stroke	0

dtype: int64





Modified data saved to CSV file.

In []: