

```

import numpy as np
import pandas as pd
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score

df=pd.read_csv("/content/drive/MyDrive/Shadi Al Minawi/Project 1/healthcare-dataset-stroke-
data.csv")

display(df.head())
display(df.info())
display(df.describe(include='all'))
df['bmi'] = df['bmi'].fillna(df['bmi'].median())
display(df.info())

def remove_outliers(df, column):
    Q1 = df[column].quantile(0.25)
    Q3 = df[column].quantile(0.75)
    IQR = Q3 - Q1
    lower_bound = Q1 - 1.5 * IQR
    upper_bound = Q3 + 1.5 * IQR
    return df[(df[column] >= lower_bound) & (df[column] <= upper_bound)]

df = remove_outliers(df, 'age')
df = remove_outliers(df, 'bmi')
df = remove_outliers(df, 'avg_glucose_level')


import matplotlib.pyplot as plt
import seaborn as sns

plt.figure(figsize=(16, 4))

# Age Distribution
plt.subplot(1, 3, 1)
sns.histplot(df['age'], bins=30, kde=True, color='skyblue')
plt.title("Age Distribution")

# BMI Boxplot
plt.subplot(1, 3, 2)
sns.boxplot(y=df['bmi'], color='orange')
plt.title("BMI Boxplot")

# Glucose Level by Stroke
plt.subplot(1, 3, 3)
sns.boxplot(x='stroke', y='avg_glucose_level', data=df)

```

```
plt.title("Glucose Level by Stroke")
```

```
plt.tight_layout()
```

```
plt.savefig("stroke_data_visualizations.png")
```

```
plt.show()
```