[10]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
df = pd.read_csv('healthcare-dataset-stroke_ML.csv')
```

[11]:
```
df.head()
```

[11]:

	id	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	avg_glucose_level	bmi	smoking_status	stroke
0	9046	Male	67.0	0	1	Yes	Private	Urban	228.69	36.6	formerly smoked	1
1	51676	Female	61.0	0	0	Yes	Self-employed	Rural	202.21	NaN	never smoked	1
2	31112	Male	80.0	0	1	Yes	Private	Rural	105.92	32.5	never smoked	1
3	60182	Female	49.0	0	0	Yes	Private	Urban	171.23	34.4	smokes	1
4	1665	Female	79.0	1	0	Yes	Self-employed	Rural	174.12	24.0	never smoked	1

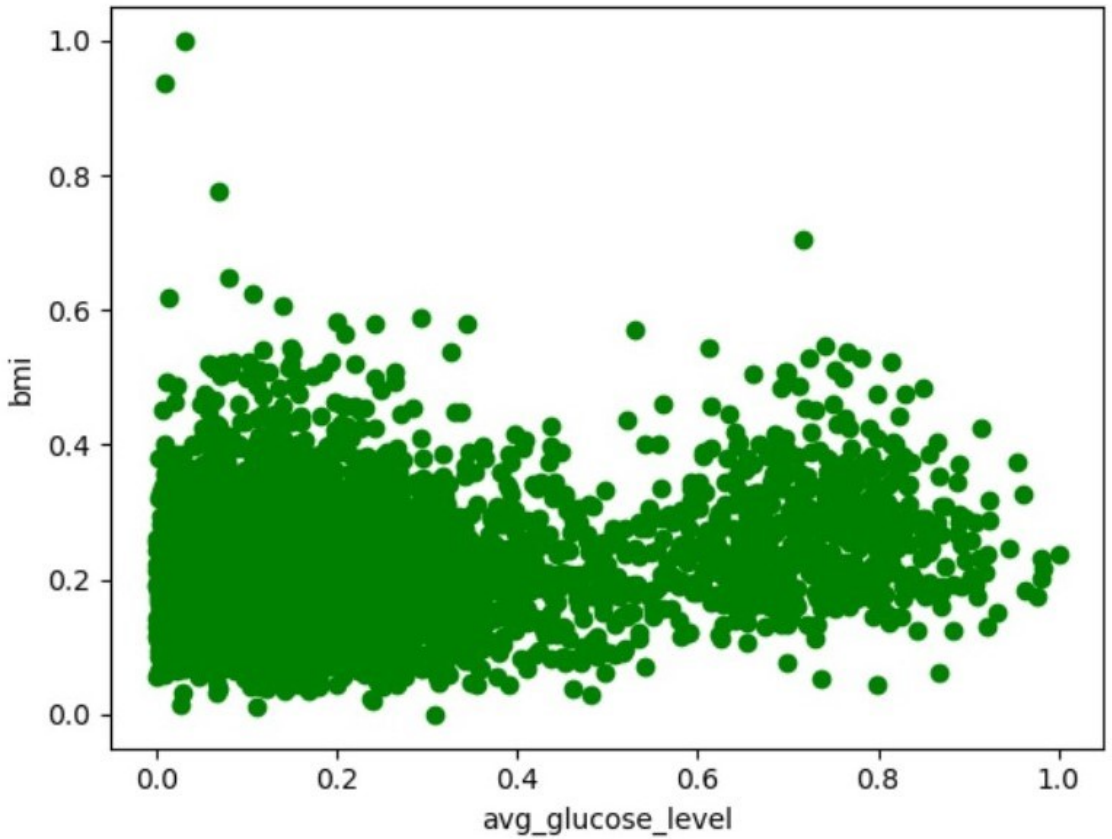
[12]:
```
df= df[df['bmi'].isnull()== False]
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 4909 entries, 0 to 5109
Data columns (total 12 columns):
 #   Column             Non-Null Count  Dtype  
---  ------             --------------  -----  
 0   id                 4909 non-null   int64  
 1   gender             4909 non-null   object 
 2   age                4909 non-null   float64
 3   hypertension       4909 non-null   int64  
 4   heart_disease      4909 non-null   int64  
 5   ever_married       4909 non-null   object 
 6   work_type          4909 non-null   object 
 7   Residence_type     4909 non-null   object 
 8   avg_glucose_level  4909 non-null   float64
 9   bmi                4909 non-null   float64
 10  smoking_status     4909 non-null   object 
 11  stroke             4909 non-null   int64  
dtypes: float64(3), int64(4), object(5)
memory usage: 498.6+ KB
```

```
names = ['age', 'hypertension','heart_disease', 'avg_glucose_level', 'bmi','stroke']
for i in names:
    df[i] = (df[i] - min(df[i])) / (max(df[i]) - min(df[i]))
df.head()
```

	id	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	avg_glucose_level	bmi	smoking_status	stroke
0	9046	Male	0.816895	0.0	1.0	Yes	Private	Urban	0.801265	0.301260	formerly smoked	1.0
2	31112	Male	0.975586	0.0	1.0	Yes	Private	Rural	0.234512	0.254296	never smoked	1.0
3	60182	Female	0.597168	0.0	0.0	Yes	Private	Urban	0.536008	0.276060	smokes	1.0
4	1665	Female	0.963379	1.0	0.0	Yes	Self-employed	Rural	0.549349	0.156930	never smoked	1.0
5	56669	Male	0.987793	0.0	0.0	Yes	Private	Urban	0.605161	0.214204	formerly smoked	1.0

```
plt.plot(df['avg_glucose_level'], df['bmi'], 'go')
plt.xlabel('avg_glucose_level')
plt.ylabel('bmi')
plt.show()
```



```
X = df[['avg_glucose_level']]
y = df[['bmi']]
```

```
from sklearn.model_selection import train_test_split
from sklearn import datasets, linear_model
from sklearn import metrics
from sklearn.metrics import accuracy_score
np.random.seed(7)
X_train, X_test, y_train , y_test = train_test_split(X, y, test_size=.30)
```

```
X_train.shape
```

```
(3436, 1)
```

```
from sklearn import neighbors
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
model = neighbors.KNeighborsClassifier(n_neighbors = 7, p = 2)
model.fit(X_train, y_train)
y_pred = model.predict(X_test)
```

```
C:\Users\ADMIN\AppData\Local\Temp\ipykernel_17204\660069349.py:5: DataConversionWarning: A column-vector y was passed when a 1d array was expecte
d. Please change the shape of y to (n_samples, ), for example using ravel().
  model.fit(X_train, y_train)
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
Cell In [20], line 5
```