

```
import pandas as pd
import numpy as np

# In[23]:

my_data=pd.read_csv('diabetes.csv')

# In[24]:

my_data

# In[25]:

my_data.shape

# In[26]:

types=my_data.dtypes
types

# In[27]:

my_data.columns

# In[28]:

my_data.head()

# In[29]:

my_data.isnull().sum()

# In[30]:

my_data.info()

# In[31]:

my_data.describe()

# In[32]:

my_data.groupby('Outcome').size()

# In[33]:

my_data.groupby('Outcome').mean()

# In[34]:

my_data.groupby('Outcome').median()

# In[35]:

my_data.groupby('Outcome').std()

# In[36]:

my_data.groupby('Outcome').skew()
```

```

# In[37]:

import warnings
warnings.filterwarnings('ignore')

# In[38]:

import seaborn as sns
sns.countplot(my_data['Outcome'], label="count")

# In[39]:

corr = my_data.corr()
corr

# In[40]:

sns.heatmap(corr, annot=True)

# In[41]:

print("Total: ", my_data[my_data.BloodPressure == 0].shape[0])
print(my_data[my_data.BloodPressure == 0].groupby('Outcome')['Age'].count())

# In[42]:

print("Total: ", my_data[my_data.Insulin == 0].shape[0])
print(my_data[my_data.Insulin == 0].groupby('Outcome')['Age'].count())

# In[43]:

print("Total: ", my_data[my_data.SkinThickness == 0].shape[0])
print(my_data[my_data.SkinThickness == 0].groupby('Outcome')['Age'].count())

# In[44]:

print("Total: ", my_data[my_data.BMI == 0].shape[0])
print(my_data[my_data.BMI == 0].groupby('Outcome')['Age'].count())

# In[45]:

print("Total: ", my_data[my_data.Glucose == 0].shape[0])
print(my_data[my_data.Glucose == 0].groupby('Outcome')['Age'].count())

# In[46]:

my_data=my_data[(my_data.BloodPressure !=0) & (my_data.BMI !=0) & (my_data.Glucose !=0)]
print(my_data.shape)

# In[52]:

from matplotlib import pyplot
import matplotlib.pyplot as plt

# In[60]:

fig, axes=plt.subplots(4,2,figsize=(8,8))

sns.histplot(data=my_data["Pregnancies"], kde=True, ax=axes[0,0], color='violet').set(title='Pregnancies Histogram')
sns.histplot(data=my_data["Glucose"], kde=True, ax=axes[0,1], color='indigo').set(title='Glucose Histogram')
sns.histplot(data=my_data["BloodPressure"], kde=True, ax=axes[1,0], color='blue').set(title='BloodPressure Histogram')
sns.histplot(data=my_data["SkinThickness"], kde=True, ax=axes[1,1], color='green').set(title='SkinThickness Histogram')
sns.histplot(data=my_data["Insulin"], kde=True, ax=axes[2,0], color='yellow').set(title='Insulin Histogram')
sns.histplot(data=my_data["BMI"], kde=True, ax=axes[2,1], color='orange').set(title='BMI Histogram')
sns.histplot(data=my_data["DiabetesPedigreeFunction"], kde=True, ax=axes[3,0], color='red').set(title='DiabetesPedigreeFunction Histogram')
sns.histplot(data=my_data["Age"], kde=True, ax=axes[3,1], color='black').set(title='Age Histogram')

```

```
plt.tight_layout()
plt.show()
```

```
# In[66]:
```

```
fig, axes=plt.subplots(4,2,figsize=(10,10))

sns.boxplot(x=my_data['Pregnancies'], ax=axes[0,0]).set(title='Boxplot for Pregnancies')
sns.boxplot(x=my_data['Glucose'],ax=axes[0,1]).set(title='Boxplot For Glucose')
sns.boxplot(x=my_data['BloodPressure'],ax=axes[1,0]).set(title='Boxplot For Bloodpressure')
sns.boxplot(x=my_data['SkinThickness'],ax=axes[1,1]).set(title='Boxplot of SkinThickness')
sns.boxplot(x=my_data['Insulin'],ax=axes[2,0]).set(title='Boxplot for Insulin')
sns.boxplot(x=my_data['BMI'],ax=axes[2,1]).set(title='Boxplot for BMI')
sns.boxplot(x=my_data['DiabetesPedigreeFunction'],ax=axes[3,0]).set(title='Boxplot for DiabetesPedigreeFunction')
sns.boxplot(x=my_data['Age'],ax=axes[3,1]).set(title='Boxplot for Age')
plt.tight_layout()
plt.show()
```

```
# In[83]:
```

```
my_data.plot(kind='density',subplots=True,layout=(3,3),sharex=False,figsize=(15,15))
pyplot.show()
```

```
# In[90]:
```

```
import pandas
from pandas.plotting import scatter_matrix

dataCorr = my_data.corr()
pandas.plotting.scatter_matrix(dataCorr,figsize=(15,15))
pyplot.show()
```

```
# In[ ]:
```