

Assignment-3

Adithya Vardhan.S(21BAI1535)

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

```
df=pd.read_csv('/penguins_size.csv')
df
```

	species	island	culmen_length_mm	culmen_depth_mm	flipper_length_mm	body_mass_g
0	Adelie	Torgersen	39.1	18.7	181.0	3750
1	Adelie	Torgersen	39.5	17.4	186.0	3800
2	Adelie	Torgersen	40.3	18.0	195.0	3250
3	Adelie	Torgersen	NaN	NaN	NaN	NaN
4	Adelie	Torgersen	36.7	19.3	193.0	3450
...
339	Gentoo	Biscoe	NaN	NaN	NaN	NaN
340	Gentoo	Biscoe	46.8	14.3	215.0	4850
341	Gentoo	Biscoe	50.4	15.7	222.0	5750
342	Gentoo	Biscoe	45.2	14.8	212.0	5200
343	Gentoo	Biscoe	49.9	16.1	213.0	5400

244 rows x 7 columns

```
from matplotlib import rcParams
rcParams['figure.figsize']=8,8
fig,axes=plt.subplots(2,2)
sns.histplot(data=df['body_mass_g'],ax=axes[0,0])
sns.distplot(df['culmen_depth_mm'],ax=axes[1,1])
sns.barplot(x=df['culmen_length_mm'],y=df['culmen_length_mm'],ax=axes[0,1])
sns.boxplot(data=df['flipper_length_mm'],ax=axes[1,0])
```

```
<ipython-input-3-5906f08a3de5>:5: UserWarning:
```

```
`distplot` is a deprecated function and will be removed in seaborn v0.14.0.
```

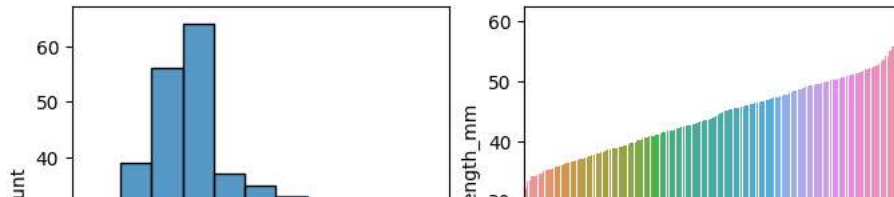
Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see

<https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751>

```
sns.distplot(df['culmen_depth_mm'],ax=axes[1,1])
```

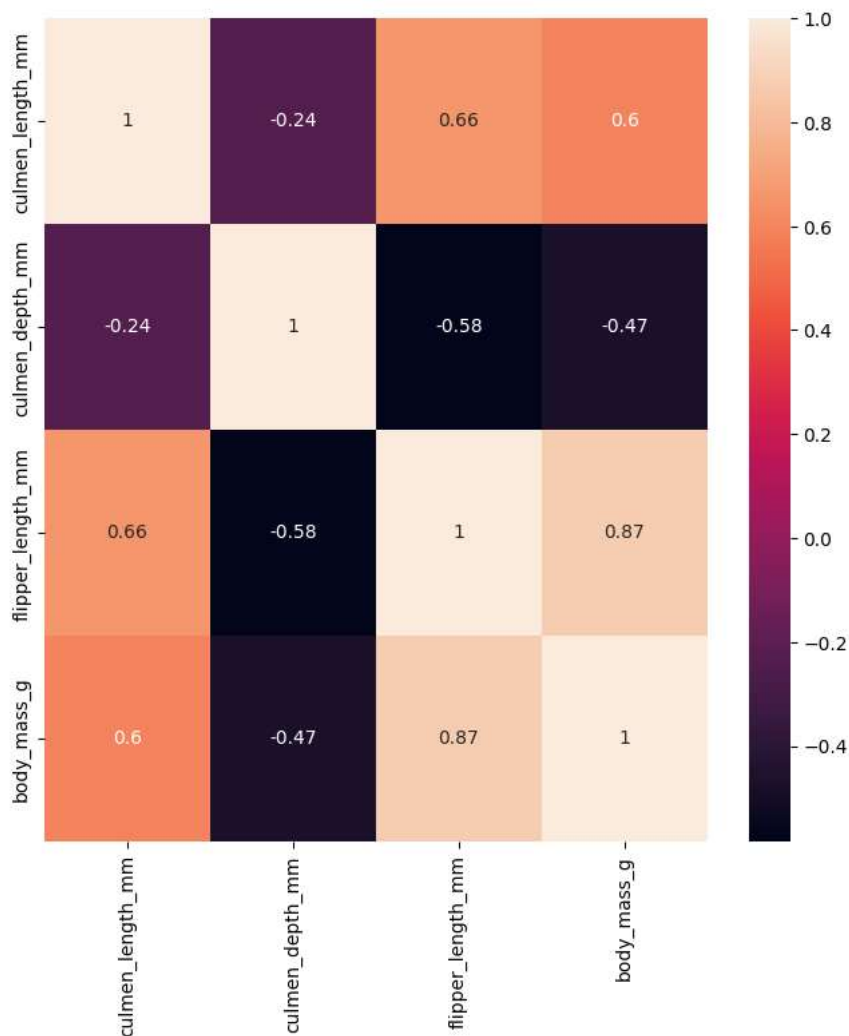
```
<Axes: >
```





```
sns.heatmap(df.corr(),annot=True)
```

```
<ipython-input-4-8df7bcac526d>:1: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a fu
sns.heatmap(df.corr(),annot=True)
```



```
<Axes: >
```



```
df.describe()
```

	culmen_length_mm	culmen_depth_mm	flipper_length_mm	body_mass_g	
count	342.000000	342.000000	342.000000	342.000000	
mean	43.921930	17.151170	200.915205	4201.754386	
std	5.459584	1.974793	14.061714	801.954536	
min	32.100000	13.100000	172.000000	2700.000000	
max	58.900000	21.900000	231.000000	8550.000000	

```
df.isnull().any()
df['culmen_length_mm'].fillna(df['culmen_length_mm'].median(),inplace=True)
df['flipper_length_mm'].fillna(df['flipper_length_mm'].median(),inplace=True)
df['culmen_depth_mm'].fillna(df['culmen_depth_mm'].median(),inplace=True)
df['body_mass_g'].fillna(df['body_mass_g'].median(),inplace=True)
most_frequent_category = df['sex'].mode()[0]
df['sex'].fillna(most_frequent_category, inplace=True)
df
```

	species	island	culmen_length_mm	culmen_depth_mm	flipper_length_mm	body_mass_g	sex	
0	Adelie	Torgersen	39.10	18.7	181.0	3750.0	MALE	
1	Adelie	Torgersen	39.50	17.4	186.0	3800.0	FEMALE	
2	Adelie	Torgersen	40.30	18.0	195.0	3250.0	FEMALE	
3	Adelie	Torgersen	44.45	17.3	197.0	4050.0	MALE	
4	Adelie	Torgersen	36.70	19.3	193.0	3450.0	FEMALE	
...	
339	Gentoo	Biscoe	44.45	17.3	197.0	4050.0	MALE	
340	Gentoo	Biscoe	46.80	14.3	215.0	4850.0	FEMALE	
341	Gentoo	Biscoe	50.40	15.7	222.0	5750.0	MALE	
342	Gentoo	Biscoe	45.20	14.8	212.0	5200.0	FEMALE	
343	Gentoo	Biscoe	49.90	16.1	213.0	5400.0	MALE	

344 rows × 7 columns

```
sns.boxplot(df.body_mass_g)
```

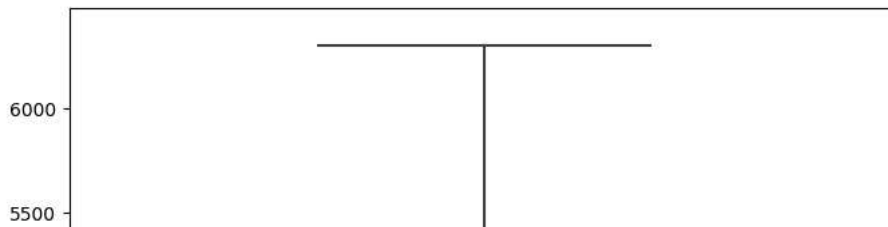
--	--

```
<ipython-input-8-ef88e168ffba>:6: FutureWarning: The default value of numeric_only in DataFrame.median is deprecated. In a future versio
df.median()
culmen_length_mm      44.45
culmen_depth_mm       17.30
flipper_length_mm     197.00
body_mass_g          4050.00
dtype: float64
```

Government	Percentage
Current government	75%
Previous government	25%


```
sns.boxplot(df.body_mass_g)
```

<Axes: >



```

from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
df.sex=le.fit_transform(df.sex)
df.species=le.fit_transform(df.species)
df.island=le.fit_transform(df.island)
correlation_matrix = df.corr()
correlation_with_target = correlation_matrix['species']
print(correlation_with_target)

```

```

species          1.000000
island          -0.635659
culmen_length_mm  0.728706
culmen_depth_mm  -0.741282
flipper_length_mm 0.850819
body_mass_g      0.747547
sex             -0.003823
Name: species, dtype: float64

```

```

X=df.drop('body_mass_g',axis=1)
y=df['body_mass_g']
from sklearn.preprocessing import StandardScaler
sc=StandardScaler()
X_scaled= pd.DataFrame(sc.fit_transform(X),columns =X.columns)
X_scaled.head()
from sklearn.model_selection import train_test_split
X_train,X_test,y_train,y_test=train_test_split(X_scaled,y,test_size=0.2,random_state=42)
X_train.shape

```

```
(275, 6)
```

```
X_test.shape
```

```
(69, 6)
```

[Colab paid products](#) - [Cancel contracts here](#)

✓ 0s completed at 11:27 PM

