fejsuqj84

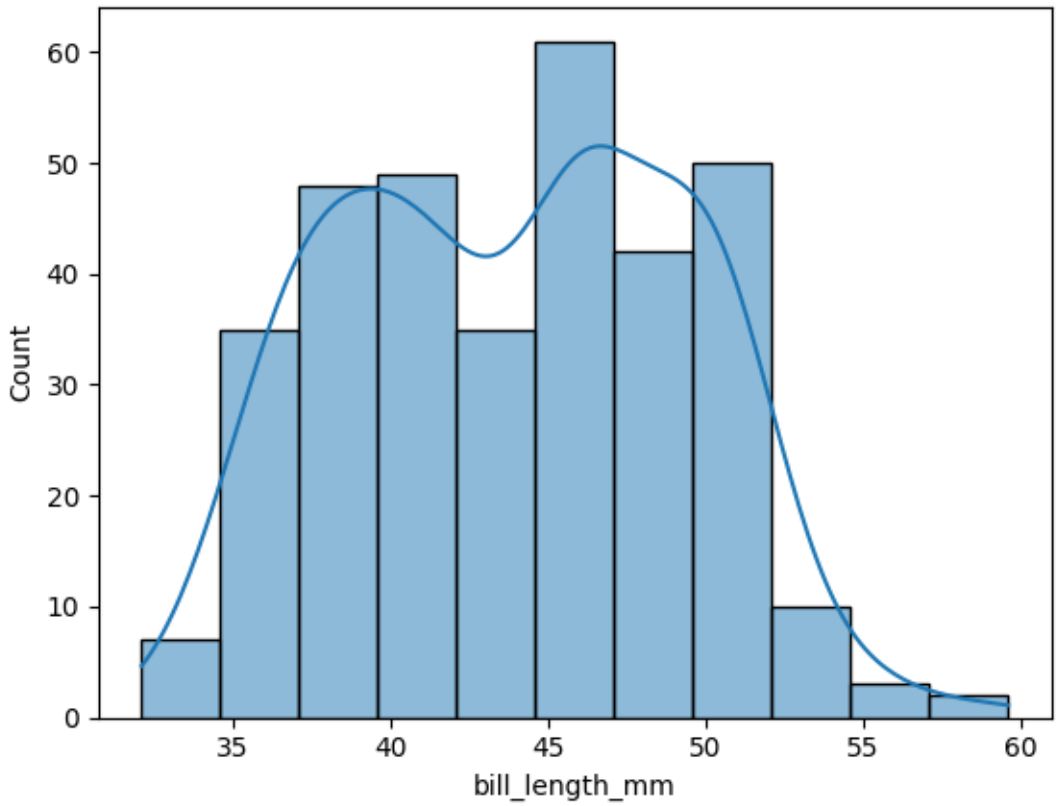
October 5, 2023

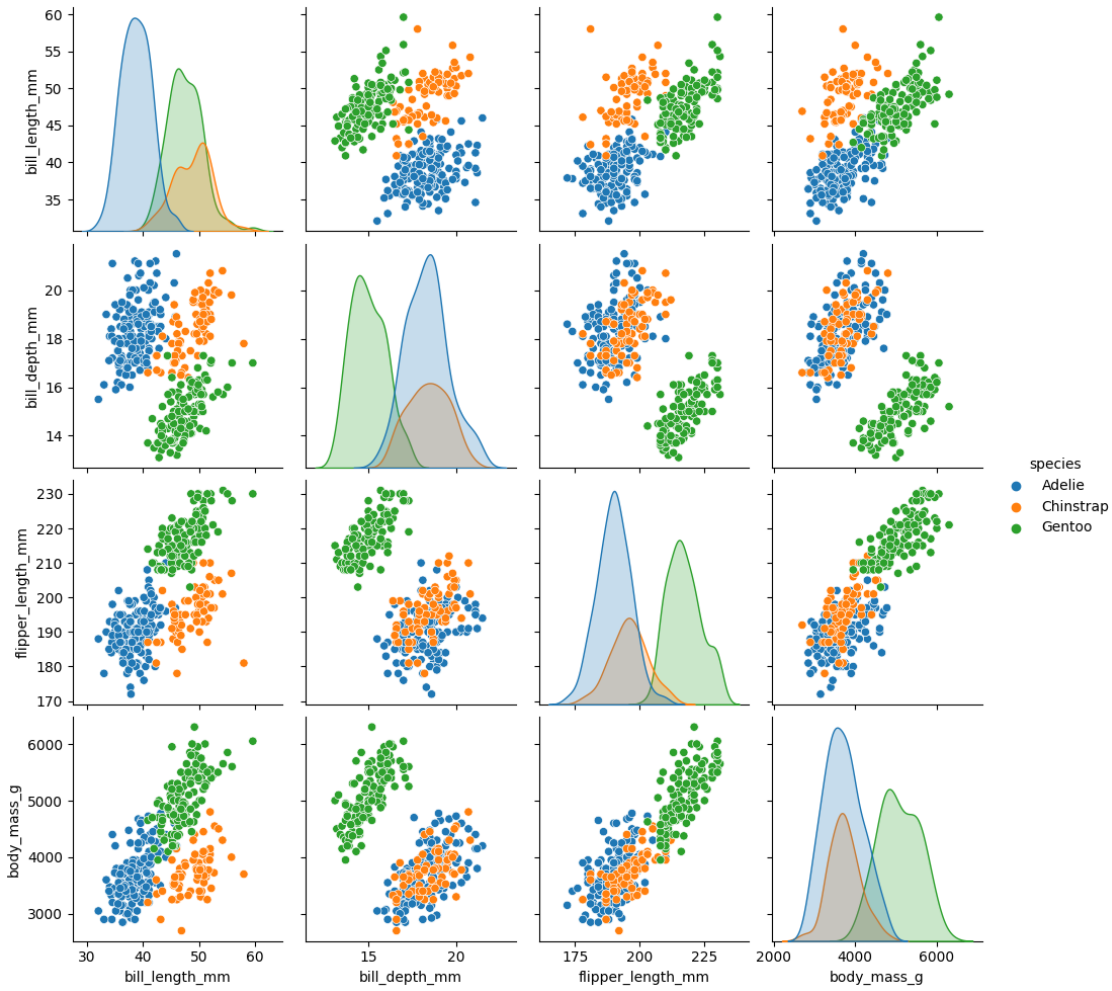
AIML DIGITAL ASSIGNMENT - 3

NAME - AKARSHIT MISRA REG.NO - 21BAI1597

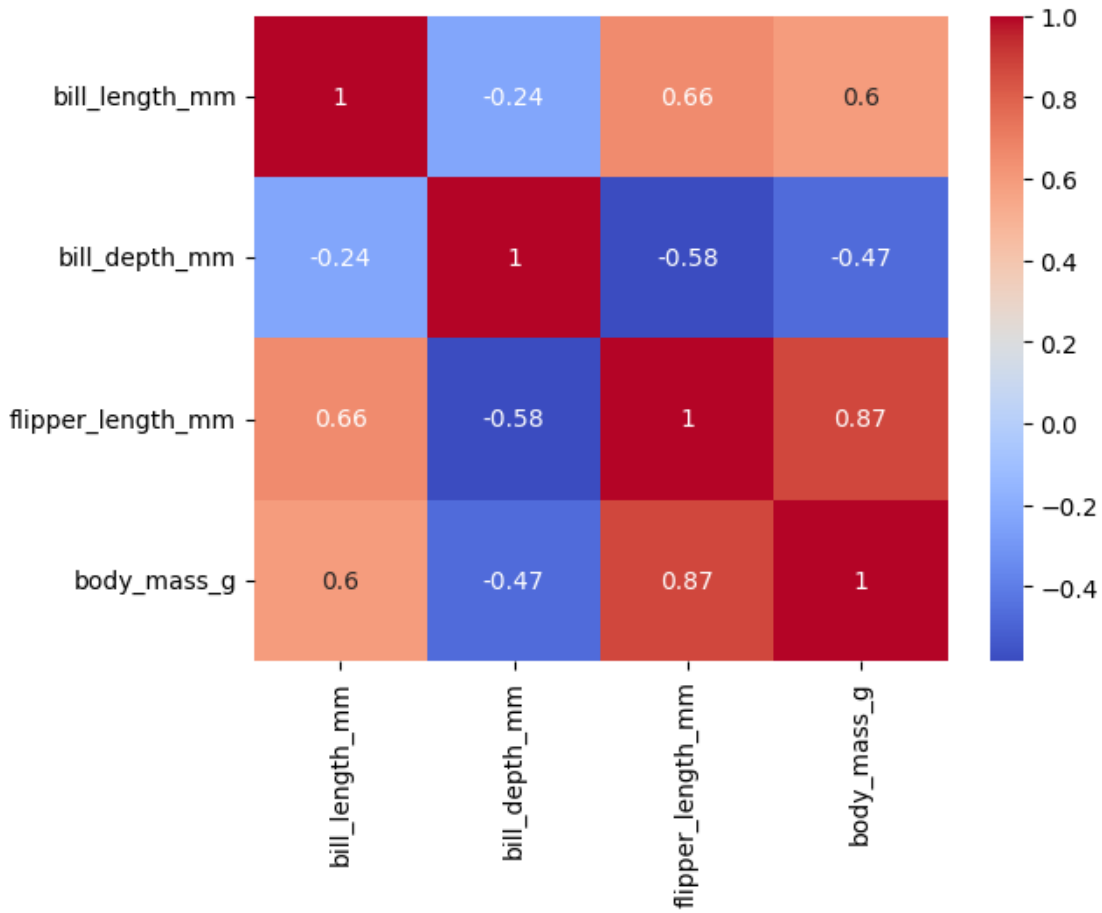
[13]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler, LabelEncoder
from sklearn.cluster import KMeans
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score, classification_report

df = pd.read_csv('/content/penguins.csv')
df.head()

# Perform Univariate Analysis
sns.histplot(data=df, x="bill_length_mm", kde=True)
plt.show()

# Perform Bi-Variate Analysis
sns.pairplot(df, hue="species")
plt.show()

# Perform Multi-Variate Analysis
correlation_matrix = df.corr()
sns.heatmap(correlation_matrix, annot=True, cmap="coolwarm")
plt.show()

# Perform descriptive statistics
descriptive_stats = df.describe()

df = df.dropna()

correlation_with_target = df.corr()['body_mass_g'].sort_values()

le = LabelEncoder()
df['sex'] = le.fit_transform(df['sex'])
```

```
df = pd.get_dummies(df, columns=['island'])

X = df.drop('species', axis=1)
y = df['species']

scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

X_train, X_test, y_train, y_test = train_test_split(X_scaled, y, test_size=0.2,␣
↪random_state=42)

print("Training data shape:", X_train.shape, y_train.shape)
print("Testing data shape:", X_test.shape, y_test.shape)
```






```
<ipython-input-13-4edb49f12a71>:23: FutureWarning: The default value of
numeric_only in DataFrame.corr is deprecated. In a future version, it will
default to False. Select only valid columns or specify the value of numeric_only
to silence this warning.
  correlation_matrix = df.corr()
```




```
Training data shape: (266, 8) (266,)
Testing data shape: (67, 8) (67,)

<ipython-input-13-4edb49f12a71>:32: FutureWarning: The default value of
numeric_only in DataFrame.corr is deprecated. In a future version, it will
default to False. Select only valid columns or specify the value of numeric_only
to silence this warning.
  correlation_with_target = df.corr()['body_mass_g'].sort_values()
<ipython-input-13-4edb49f12a71>:35: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df['sex'] = le.fit_transform(df['sex'])
```