Assignment -5 -(Data science and Analysis)

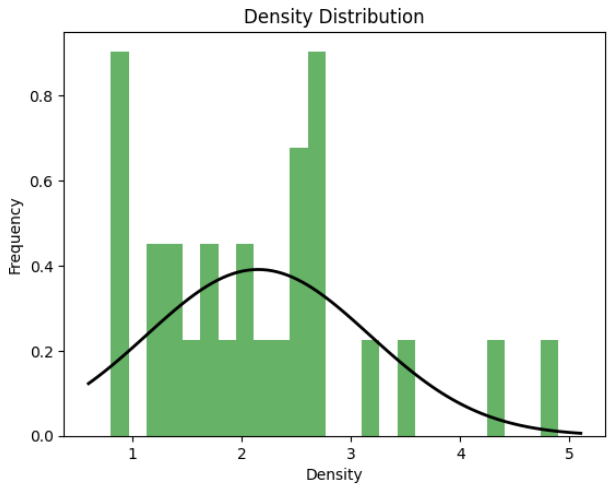
Name:- Ashwin Kumar

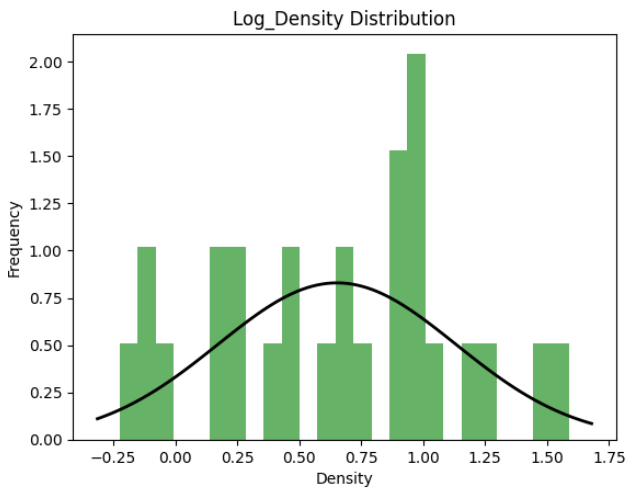
Roll:- CE21BTECH11008

Q1.

```
import numpy as np
import pandas as pd
import scipy.stats as st
import matplotlib.pyplot as plt

data = pd.read_csv('/content/Asteroid _ Dens +-.txt', sep=' ',
skipinitialspace=True, error_bad_lines=False)
print(data)
density= data['Dens']
log_density=np.log(density)
value=data['value']
print(value)
density
params= st.shapiro(density)
print(f"P vaue is {params.pvalue}  ")
print(params.statistic)


mu, std = st.norm.fit(density)

# Plot the histogram of the data
plt.hist(density, bins=25, density=True, alpha=0.6, color='g')

# Plot the PDF of the fitted normal distribution
xmin, xmax = plt.xlim()
x = np.linspace(xmin, xmax, 100)
p = st.norm.pdf(x, mu, std)
plt.plot(x, p, 'k', linewidth=2)

# Add labels and title to the plot
plt.xlabel('Density')
plt.ylabel('Frequency')
plt.title('Density Distribution')
plt.show()

mu1, std1 = st.norm.fit(log_density)

# Plot the histogram of the data
plt.hist(log_density, bins=25, density=True, alpha=0.6, color='g')
```

```
# Plot the PDF of the fitted normal distribution
xmin, xmax = plt.xlim()
x1 = np.linspace(xmin, xmax, 100)
p1 = st.norm.pdf(x1, mu1, std1)
plt.plot(x1, p1, 'k', linewidth=2)

# Add labels and title to the plot
plt.xlabel('Density')
plt.ylabel('Frequency')
plt.title('Log_Density Distribution')
plt.show()
```





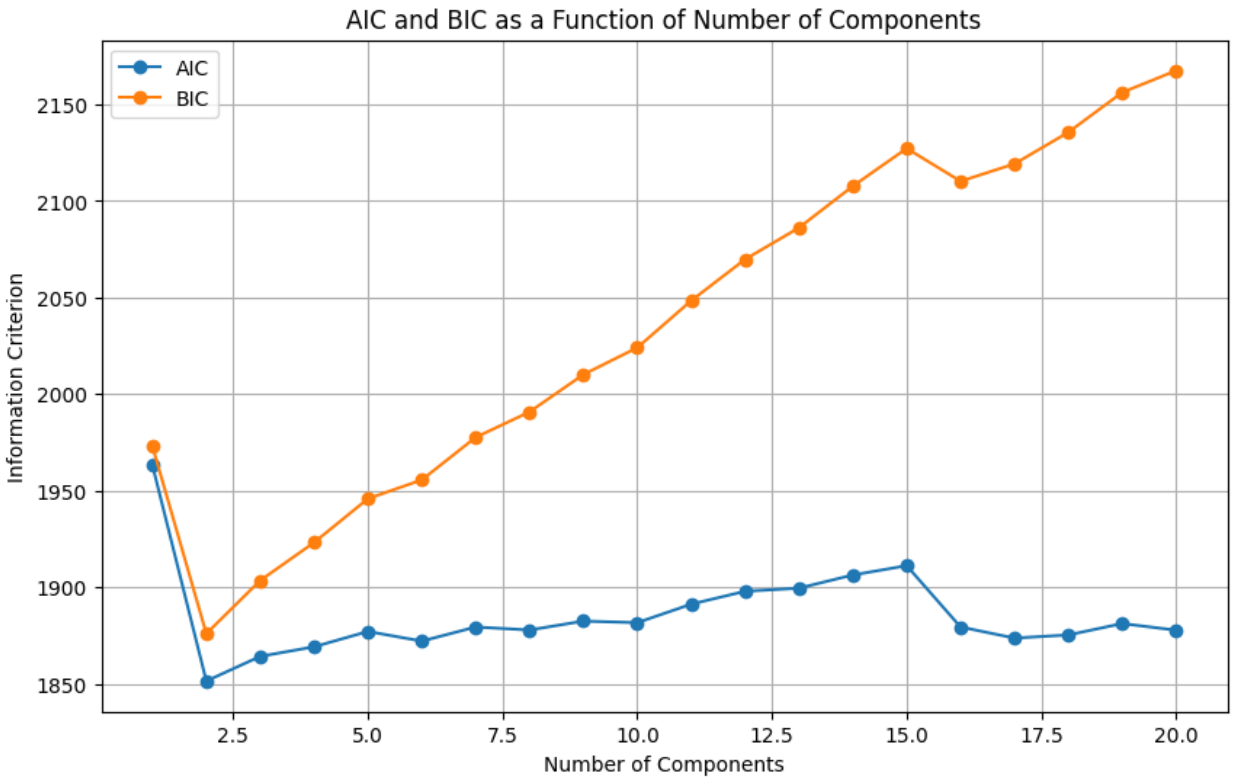
Q2.

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.mixture import GaussianMixture

# Load the T90 data
data = np.loadtxt('/content/dsa_data_assign6-2.txt')

# Take the log10 of the data
log10_data = np.log10(data)

print(log10_data)
```

```
# Reshape the data to a 2D array with a single feature
log10_data_reshaped = log10_data.reshape(-1, 1)

# Initialize lists to store AIC and BIC values
AIC_values = []
BIC_values = []

# Fit GMM models with different number of components
for n_components in range(1, 21):
    gmm = GaussianMixture(n_components=n_components, random_state=0)

    #applying log to the data
    gmm.fit(log10_data_reshaped)


    AIC_values.append(gmm.aic(log10_data_reshaped))
    BIC_values.append(gmm.bic(log10_data_reshaped))

# Plot AIC and BIC as a function of number of components
plt.figure(figsize=(10, 6))
plt.plot(range(1, 21), AIC_values, marker='o', label='AIC')
plt.plot(range(1, 21), BIC_values, marker='o', label='BIC')
plt.xlabel('Number of Components')
plt.ylabel('Information Criterion')
plt.title('AIC and BIC as a Function of Number of Components')
plt.legend()
plt.grid(True)
plt.show()

# Find the optimum number of components using AIC and BIC
optimal_n_components_AIC = np.argmin(AIC_values) + 1
optimal_n_components_BIC = np.argmin(BIC_values) + 1
print(f'Optimum components according to AIC: {optimal_n_components_AIC}')
print(f'Optimum components according to BIC: {optimal_n_components_BIC}')
```

Optimum number of components according to AIC: 2

Optimum number of components according to BIC: 2



Q3.

```
import pandas as pd
from scipy.stats import ttest_ind

# Read data from the text file into a DataFrame
df = pd.read_csv('/content/HIP_star.txt', delim_whitespace=True)

# Filter data for Hyades and non-Hyades stars
hyades_stars = df[(df['RA'] >= 50) & (df['RA'] <= 100) & (df['DE'] >= 0) &
(df['DE'] <= 25) & (df['pmRA'] >= 90) & (df['pmRA'] <= 130) & (df['pmDE']
>= -60) & (df['pmDE'] <= -10)]['B-V']
non_hyades_stars = df[~((df['RA'] >= 50) & (df['RA'] <= 100) & (df['DE']
>= 0) & (df['DE'] <= 25) & (df['pmRA'] >= 90) & (df['pmRA'] <= 130) &
(df['pmDE'] >= -60) & (df['pmDE'] <= -10))]['B-V']

# Perform two-sample t-test
t_statistic, p_value = ttest_ind(hyades_stars, non_hyades_stars)

# general significane level
alpha = 0.05
```

```
# Interpret results
if p_value < alpha:
    print("Reject the null hypothesis: The color (B-V) of the Hyades stars
differs from the non-Hyades ones.")
else:
    print("Accepting the null hypothesis: The color (B-V) of the Hyades
stars does not differ from the non-Hyades ones.")
```

Reject the null hypothesis: The color (B-V) of the Hyades stars differs from the non-Hyades ones.