Smt. Indira Gandhi College of Engineering
Computer Engineering Department
Ghansoli – Navi Mumbai
Academic Year 2023-24 (Even Sem)

Student Name: Khyati Garude **Roll No.:** 13 **Class:** BE **Sem:** VIII

Course Name: Applied Data Science Lab

Course Code: CSL8023

Experiment No. 05

Experiment Title: Implementing Data Visualisation On The Selected Dataset

Date of Performance	Date of Submission	Marks (10)					Sign / Remark
		A	B	C	D	E	
		2	3	2	2	1	
1/2/24	8/2/24	2	2	2	2	1	20/3/24.
		Total Marks					
		09					

A: Prerequisite Knowledge
B: Implementation
C: Oral
D: Content
E: Punctuality & Discipline

Khyati Garude
BE comps
ADS

Date:

DATE	EXPERIMENT-5	SIGN
8/2/24	Implementing Data Visualization on the selected dataset.	20/3/24.

AIM: Implementing data visualization on the selected dataset.

THEORY:

Data Visualization is the process of representing data in a graphical or visual format to facilitate understanding, analysis, and interpretation.

The primary purpose of data visualization is to communicate information effectively. By presenting data visually, complex patterns, trends and relationships can be easily identified and understood by users.

Types of Data visualization:

(i) Line Plot:

Line Plots are used to visualize the trend or relationship between two continuous variables over time or another continuous dimension.

Date:_____

(ii) Bar plot:

Bar plots are used to compare categorical data or to display the distribution of a categorical variable.

(iii) Histogram:

Histograms are used to visualize the distribution of a continuous variable by dividing the data into bins and displaying the frequency of observations in each bin.

(iv) Scatter plot:

Scatter plots are used to visualize the relationship between two continuous variables to identify patterns or correlations.

(v) Box plot:

Box plots are used to visualize the distribution of a continuous variable and identify outliers and variability within the data.

(vi) Heatmap:

Heatmaps are used to visualize the magnitude of a variable across two dimensions typically as a grid of coloured squares.

(vii) Pie chart:

Pie charts are used to visualize the proportion of different categories in a dataset.

Date:_____

(viii) Area plot:
Area plots are similar to line plots but filled with colour between the lines, making them suitable for visualizing cumulative values over time.

(ix) Violin plot:
Violin plots are used to visualize the distribution of continuous variable, providing insights into both the distribution's shape and its central tendency.

(x) Bubble chart:
Bubble charts are similar to scatter plots but with an additional dimension represented by the size of points (bubbles).

Principles of Data Visualization:

(i) Simplify: simplify the visual elements to focus on the most important aspects of the data.
(ii) Clarity: Ensure clarity in the presentation of information by using clear labels, titles and legends.
(iii) Accuracy: Ensure that the visualization accurately represents the underlying data without distortion or misinterpretation.
(iv) Relevance: Present only relevant information that supports the intended message or analysis.
(v) Engagement: Design visually appealing and engaging visualizations to capture the audience's attention and maintain interest.

Date:_____

Data visualization is used across various domains and industries, including business intelligence, finance, healthcare, marketing, social sciences, and more. It helps analysts, researchers, and decision-makers gain insights, identify trends, make data-driven decisions, and communicate findings effectively.

ALGORITHM:

(i) Import the necessary libraries: 'pandas', 'matplotlib.pyplot', and 'seaborn'.
(ii) Load the dataset using 'pd.read_csv()'.
(iii) Plot a line plot using 'sns.lineplot()' for sleep duration Vs sleep efficiency.
(iv) Plot a scatter plot using 'sns.scatterplot()' for sleep duration Vs Sleep efficiency.
(v) Plot a Box plot using 'sns.boxplot()' for sleep duration Vs sleep efficiency.
(vi) Plot a histogram using 'plt.hist()' for the distribution of sleep duration.
(vii) Plot a bar plot using 'sns.countplot()' for Gender distribution.
(viii) Plot a heatmap using 'sns.heatmap()' to visualize the correlation between attributes.
(ix) Plot a pie chart using 'plt.pie()' for the distribution of sleep duration.
(x) Plot an area plot using 'sns.lineplot()' and 'plt.fill_between()' for Age Vs Sleep duration.

Date:_____

(xi) Plot a violin plot using 'sns.violinplt()' for sleep efficiency by Gender.

(xii) Plot a bubble chart using 'sns.scatterplot()' for sleep duration vs sleep efficiency with bubble size representing Age.

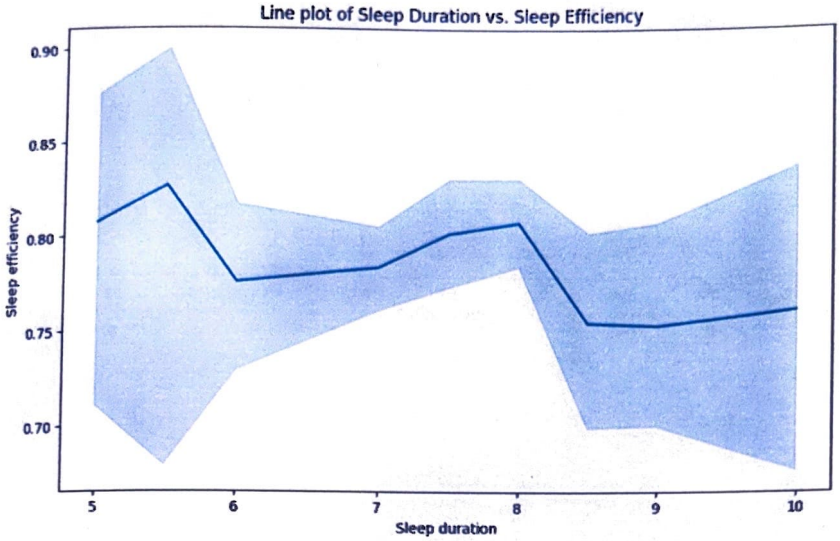
(xiii) Show the visualizations using 'plt.show()'.

CONCLUSION:

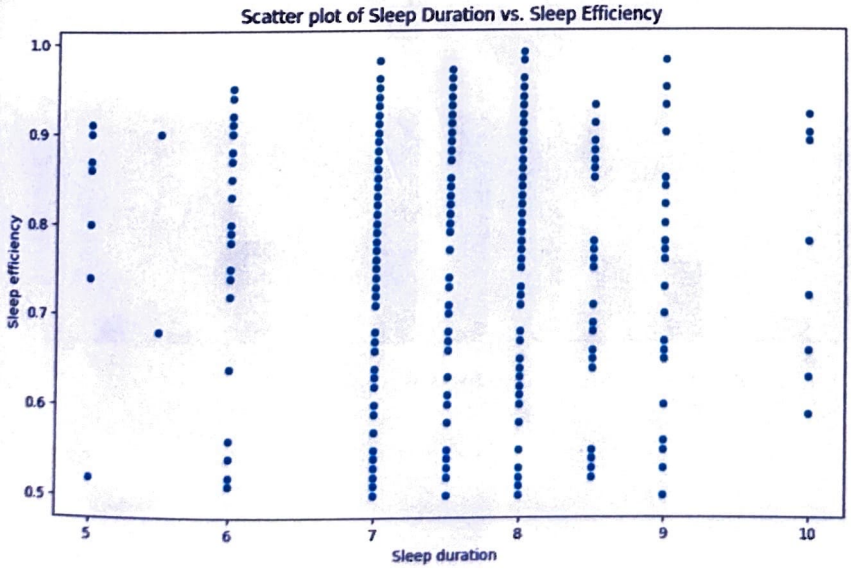
Data visualization is a powerful tool for exploring and communicating insights from data effectively. It offers a wide range of visualization types and customization options, empowering users to create informative visual representations of complex data relationships and patterns. With its flexibility, ease of use, and extensive libraries, data visualization facilitates better understanding and decision-making across various domains and industries.

20/3/24

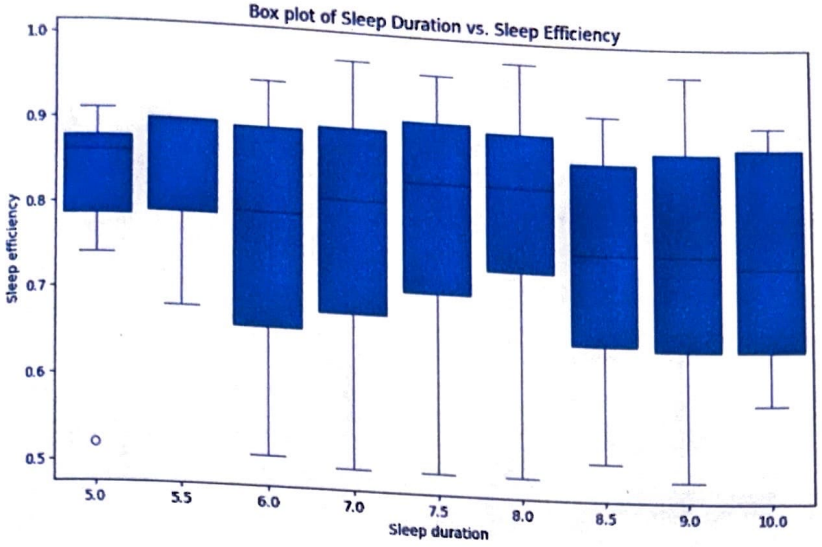
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Read the data
data = pd.read_csv("/content/Sleep_Efficiency.csv")

# Line plot
plt.figure(figsize=(10, 6))
sns.lineplot(x='Sleep duration', y='Sleep efficiency', data=data)
plt.title('Line plot of Sleep Duration vs. Sleep Efficiency')
plt.show()
```



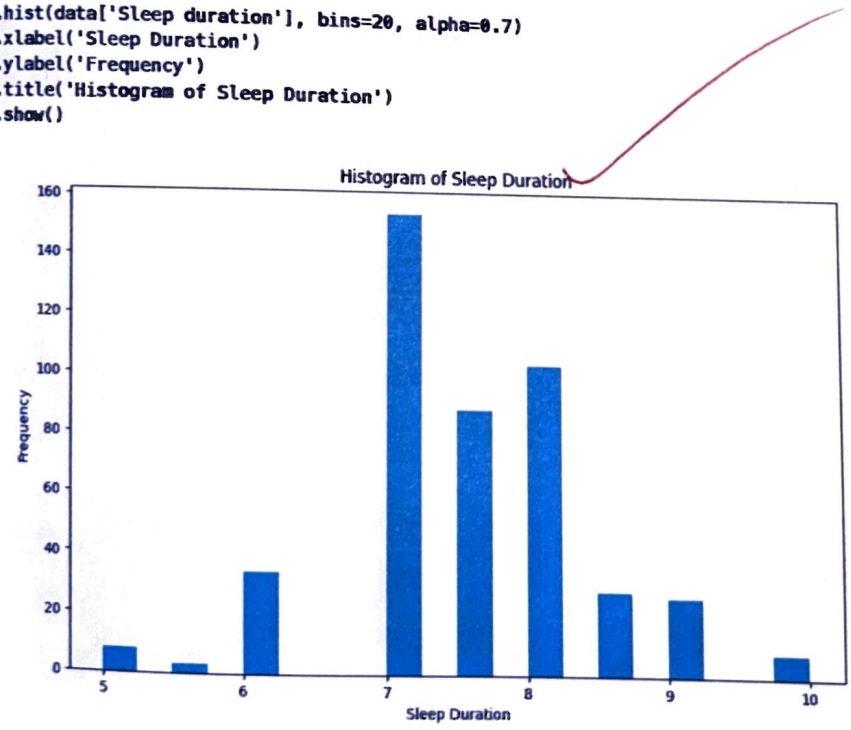

```
Scatter plot
lt.figure(figsize=(10, 6))
ns.scatterplot(x='Sleep duration', y='Sleep efficiency', data=data)
lt.title('Scatter plot of Sleep Duration vs. Sleep Efficiency')
lt.show()
```



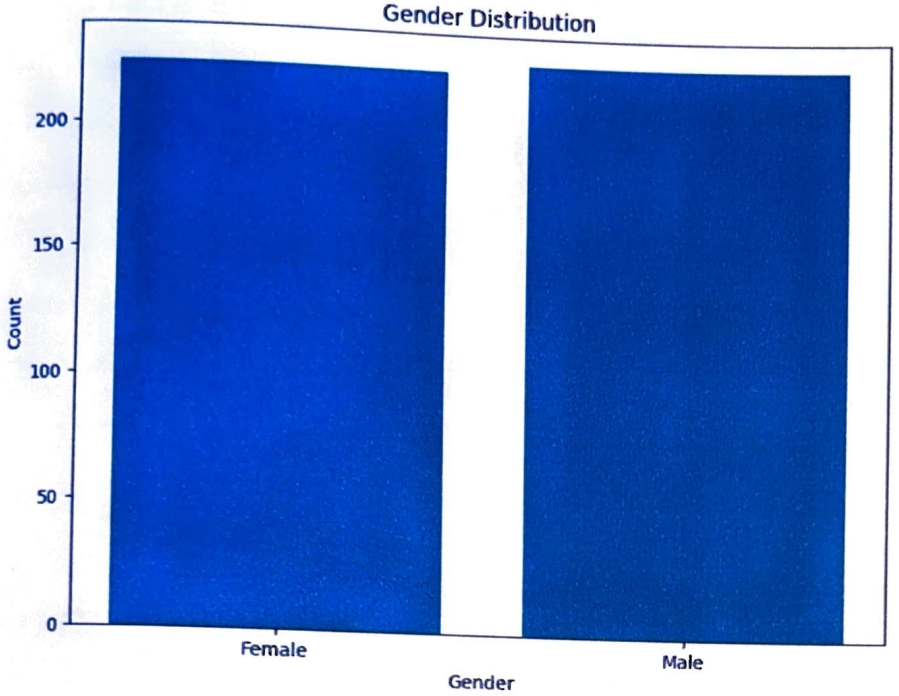
```
#Box plot
plt.figure(figsize=(10, 6))
sns.boxplot(x='Sleep duration', y='Sleep efficiency', data=data)
plt.title('Box plot of Sleep Duration vs. Sleep Efficiency')
plt.show()
```



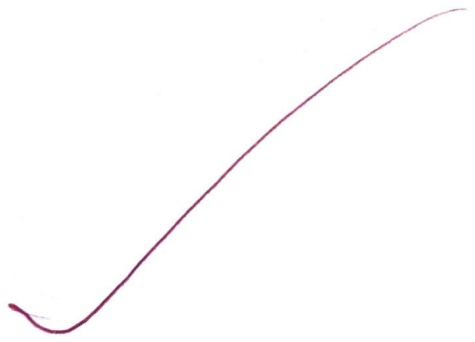

```
# Histogram
plt.figure(figsize=(10, 6))
plt.hist(data['Sleep duration'], bins=20, alpha=0.7)
plt.xlabel('Sleep Duration')
plt.ylabel('Frequency')
plt.title('Histogram of Sleep Duration')
plt.show()
```



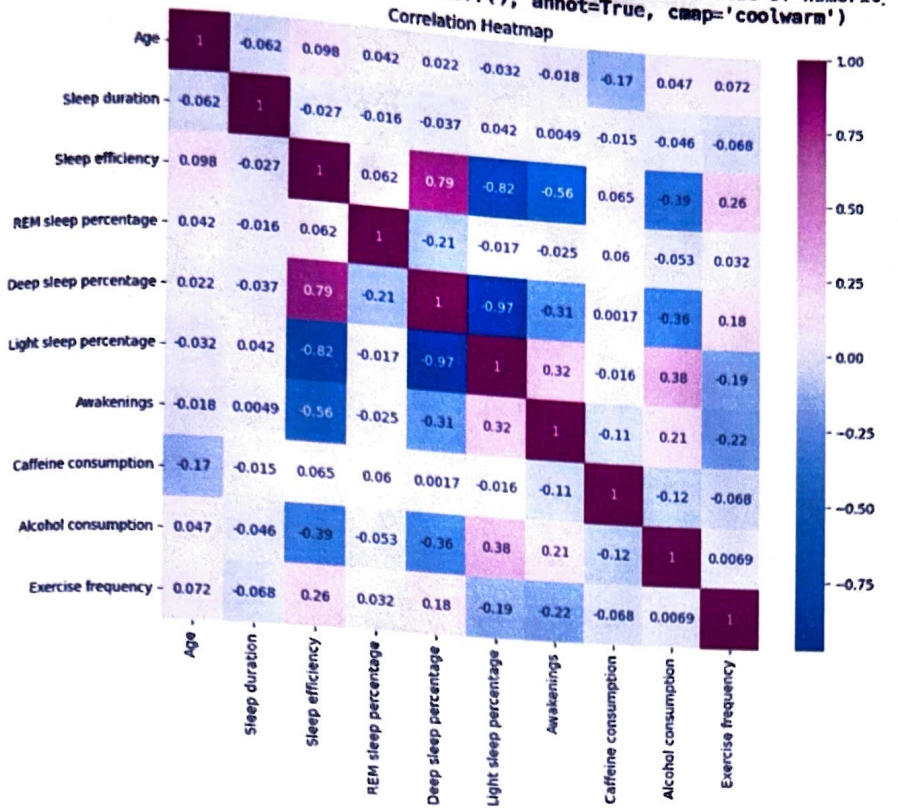
```
# Bar plot
plt.figure(figsize=(8, 6))
sns.countplot(x='Gender', data=data)
plt.title('Gender Distribution')
plt.xlabel('Gender')
plt.ylabel('Count')
plt.show()
```



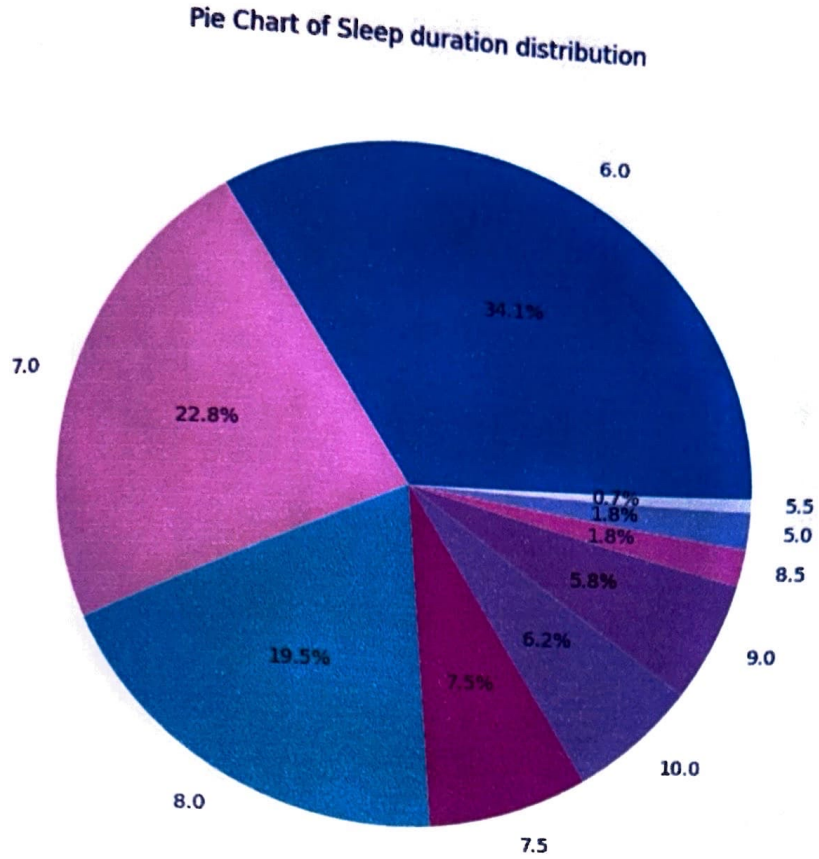

```
# Heatmap for all attributes except ID
plt.figure(figsize=(10, 8))
sns.heatmap(data.drop('ID', axis=1).corr(), annot=True, cmap='coolwarm')
plt.title('Correlation Heatmap')
plt.show()
```



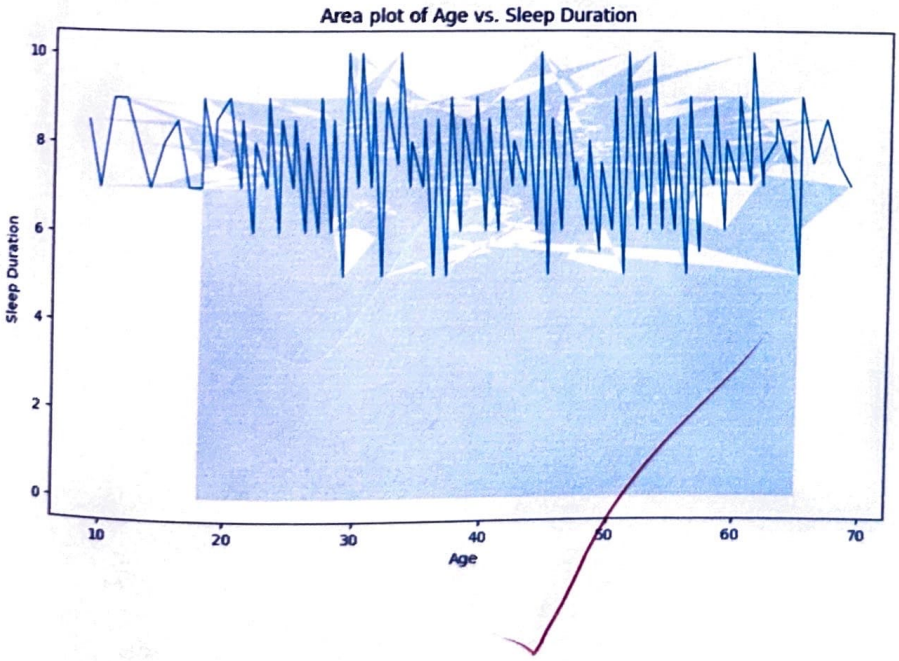
```
<ipython-input-6-ee32f16a59bf>:3: FutureWarning: The default value of numeric
  sns.heatmap(data.drop('ID', axis=1).corr(), annot=True, cmap='coolwarm')
```



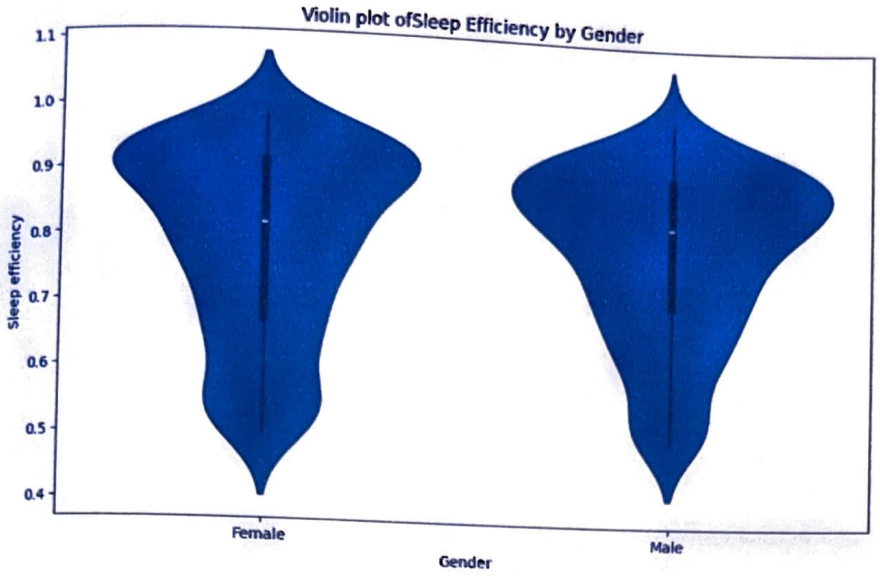

```
# Pie chart for Sleep duration
plt.figure(figsize=(8, 8))
plt.pie(data['Sleep duration'].value_counts(), labels=data['Sleep duration'].unique(), autopct='%1.1f%%')
plt.title('Pie Chart of Sleep duration distribution')
plt.show()
```




```
# Area plot for Age and Sleep duration
plt.figure(figsize=(10, 6))
sns.lineplot(x='Age', y='Sleep duration', data=data, estimator=None, lw=1)
plt.fill_between(data['Age'], data['Sleep duration'], alpha=0.2)
plt.title('Area plot of Age vs. Sleep Duration')
plt.xlabel('Age')
plt.ylabel('Sleep Duration')
plt.show()
```



```
# Violin plot for Sleep efficiency by Gender
plt.figure(figsize=(10, 6))
sns.violinplot(x='Gender', y='Sleep efficiency', data=data)
plt.title('Violin plot ofSleep Efficiency by Gender')
plt.show()
```



20/3/24.