DATA ANALYSIS

**TOPIC:**

***Student Sleeping Patterns***

**SECTION:**

**BS INFOTECH – 3B**

**MEMBERS:**

Alora, Mia

Roncales, Iris S.

Calimotan, jazel C.

Lacadue, Jhonmark

Palomas, Kent Karl

Maribojoc, Marry Jane

Nacional, Melinda

Dumalag, Royette

**SUBMITTED TO:**

**Ms. Myka Pilpa**

**Instructor**

INTRODUCTION 1

DATA EXPLORATION 2

METHODS 3

MODELLING AND RESULTS 4

CONCLUSION 5

REFERENCES 6

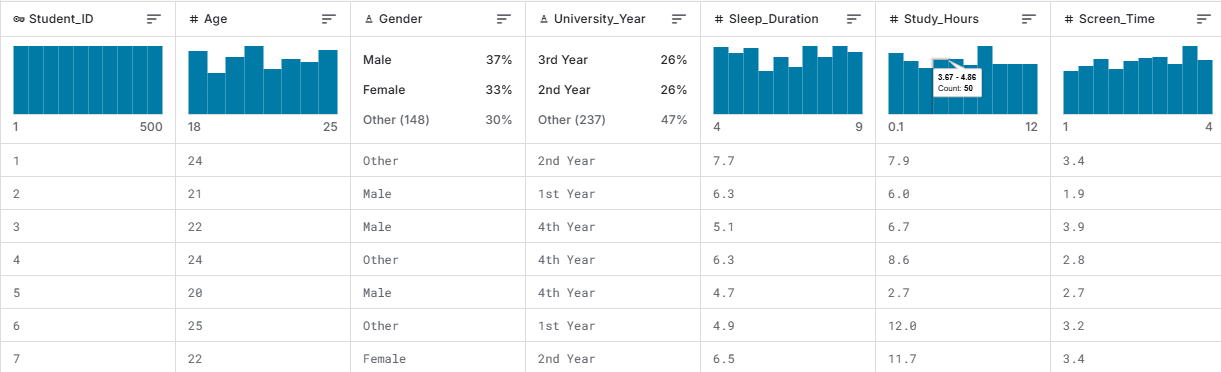
**INTRODUCTION**

This report investigates student sleep patterns and their potential impact on sleep quality. The dataset contains information on factors such as age, gender, university year, screen time, and caffeine intake. The primary goal of this analysis is to explore the relationship between these factors and to segment students into distinct clusters based on shared characteristics.

**DATA EXPLORATION**

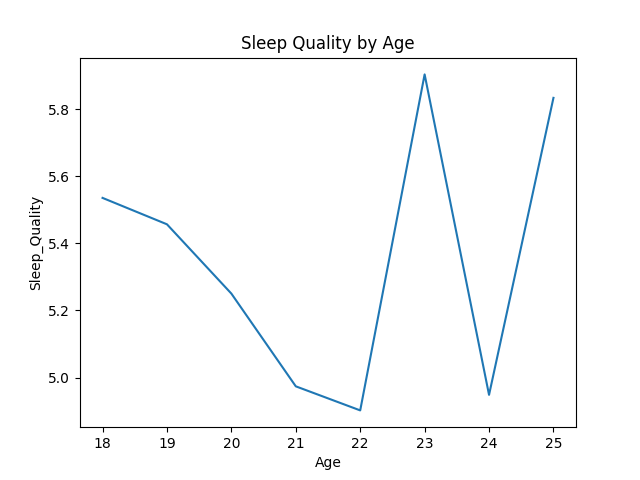
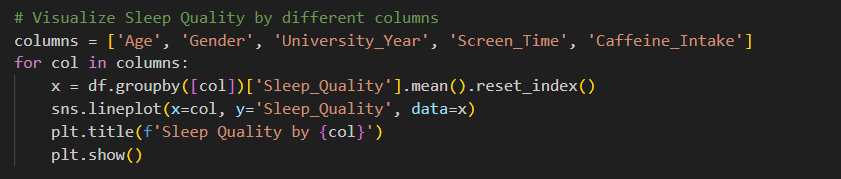
**Dataset Overview**

The dataset was provided in CSV format and contains anonymized records of students. Key features include:

* **Student\_ID**
* **Age**
* **Gender**
* **University\_Year**
* **Sleep\_Duration**
* **Sudy\_hours**
* **Screen\_Time**

**How the Data Was Collected**

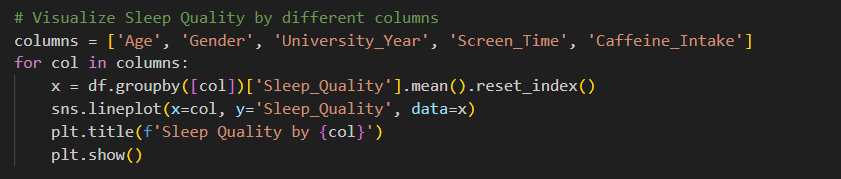
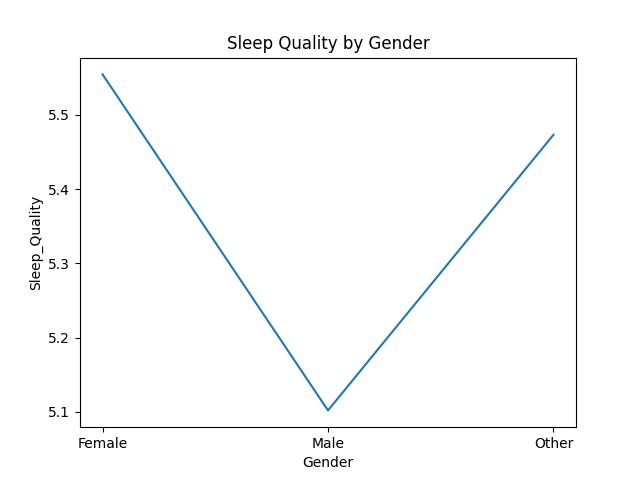
The dataset was sourced from a Kaggle repository titled "Student Sleep Patterns." It was downloaded as a CSV file and loaded into a Pandas DataFrame for analysis. The features were chosen to understand how lifestyle factors influence sleep quality.

**Features Identified for Analysis**

**Fig 1: Sleep Quality by Age**

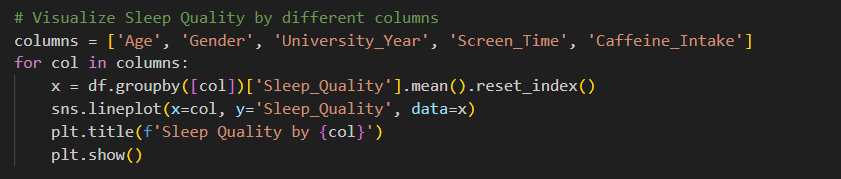
**Insight:** Sleep quality decreases from age 18 to 22, shows a sharp spike at 23, drops at 24, and then rebounds at 25. This suggests inconsistent sleep quality with notable improvements at ages 23 and 25.

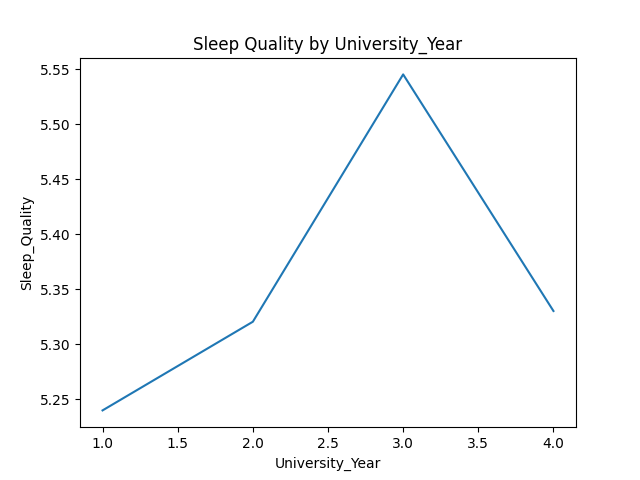
**Visualization:** Use markers on data points and a smoother trend line to emphasize fluctuations and critical changes in sleep quality.

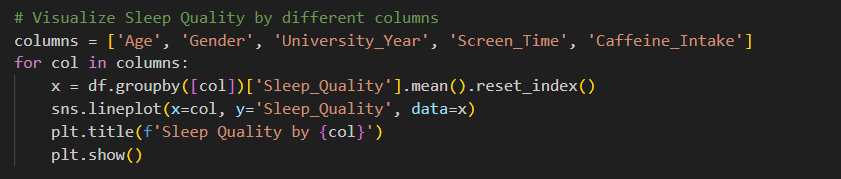
**Fig 2: Sleep Quality by Gender**  
 **Insight:** Sleep quality is highest for females, lowest for males, and slightly improves for the "Other" category. This suggests a gender-based disparity in sleep quality.

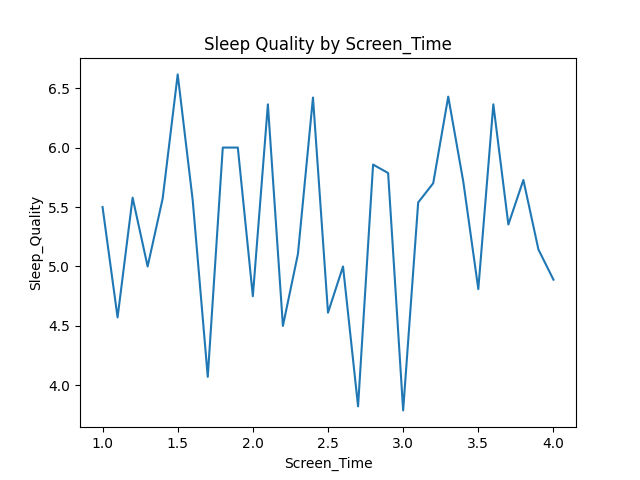
**Visualization:** Use markers at each gender point and add labels to clarify sleep quality values across categories.



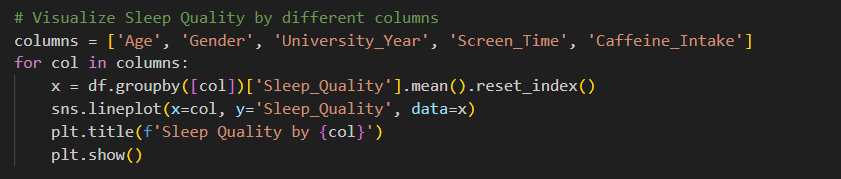


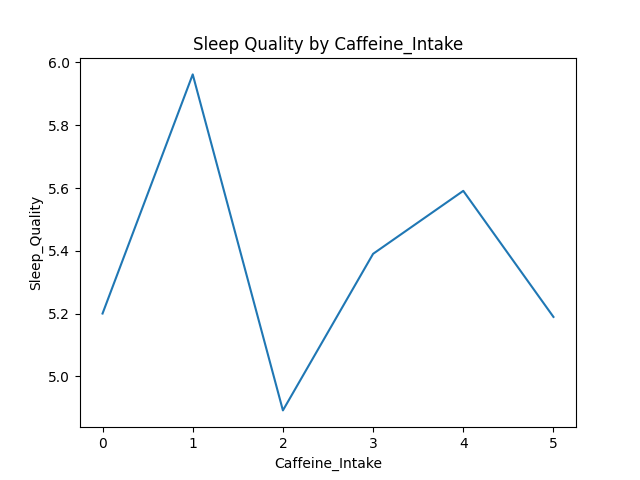
**Fig 3: Sleep Quality by University\_Year**  
  
**Insights:** Sleep quality may improve initially as students adjust to university life but declines later, possibly due to increased workload or stress in senior years.  
**Visualization:** Sleep quality gradually improves until Year 3, where it peaks, but then declines slightly in Year 4.





**Fig 4: Sleep Quality by Screen\_Time**  
  
**Insights:** Although no clear pattern emerges, prolonged screen time may introduce variability in sleep quality, potentially disrupting consistent rest.  
**Visualization:** Sleep quality fluctuates greatly with screen time, showing no clear trend and considerable inconsistency.





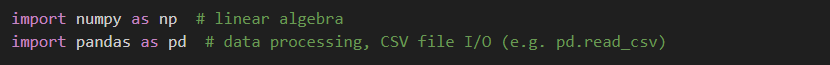
**Fig 5: Sleep Quality by Caffeine\_Intake**  
  
**Insights:** A peak in sleep quality occurs at low caffeine intake, while higher intake appears to disrupt sleep quality, suggesting a negative impact from excessive caffeine.  
**Visualization:** Sleep quality decreases as caffeine intake rises beyond 1 unit, with significant variability between levels.

**METHODS**

**Pre-Processing Techniques Used**

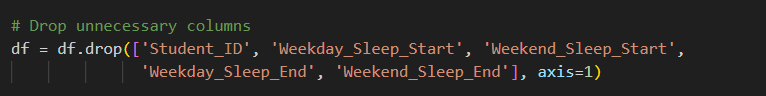
The following pre-processing techniques were implemented to prepare the dataset for analysis:

**Data Importation**: The dataset was imported using the **pandas** library, ensuring that all necessary libraries (Pandas, NumPy) were installed for data analysis and visualization. The dataset was loaded using the **read\_csv** function from Pandas, which is essential for handling CSV files.



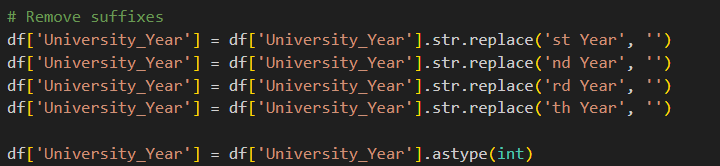
**Data Cleaning:**

1. **Column Dropping:** Unnecessary columns such as Student\_ID, Weekday\_Sleep\_Start, and Weekend\_Sleep\_End were removed.



Data cleaning processes, such as column dropping and data transformation, were applied to optimize the dataset for analysis. Unnecessary columns like **Student\_ID**, **Weekday\_Sleep\_Start**, and **Weekend\_Sleep\_End** were removed because they did not contribute directly to analyzing sleep quality, and their presence could add noise and complexity.

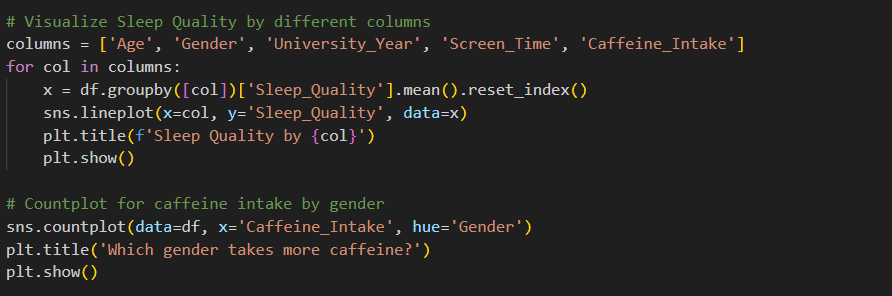
1. **Data Transformation:** University year suffixes (e.g., "st Year") were removed and converted to integers.



University year suffixes (e.g., "1st Year" or "2nd Year") were removed and converted to integers to standardize the data format, making it easier to perform numerical comparisons, calculations, and visualizations. These steps ensure the dataset is simplified, consistent, and focused on meaningful features, improving efficiency and accuracy for further analysis.

**Data Visualization**

1. **Sleep Quality Analysis:** Sleep quality was analyzed by grouping data by age, gender, university year, screen time, and caffeine intake.
2. **Countplots and Pie Charts:** Visualizations included:
   * Gender distribution of caffeine intake.
   * University year and gender distributions.



Data visualization is applied for gaining insights into the dataset and understanding complex patterns, trends, and relationships. It highlights trends, such as changes in sleep quality across age or university years, and explores interactions between variables, like caffeine intake and gender differences. Visual tools, including line plots, countplots, pie charts, and scatter plots, make it easier to identify data distributions, recognize patterns, and communicate findings effectively. Additionally, visualizations support decision-making by guiding further analysis steps, such as feature engineering or model selection, making raw data more accessible and actionable.

**CONCLUSION**

In conclusion, this analysis underscores the complex interplay between lifestyle factors and sleep quality among students. Key findings reveal that high screen time and excessive caffeine intake are associated with lower sleep quality, while age, gender, and university year also show varying degrees of influence. The insights obtained from this study emphasize the need for targeted interventions to promote healthier sleep habits, particularly in addressing screen time and caffeine consumption.

**REFERENCES**

Arsalan J, 2024, Student Sleep Patterns <https://www.kaggle.com/datasets/arsalanjamal002/student-sleep-patterns/data>