SWE 485 Project

Student Sleep Pattern Analysis and Recommendation System

Group #2 | Section: 56751
Prepared By:

Reem Alomair	444204383
Arwa Altheeb	443200566
Sultanah Almutairi	443200444
Ruba Alrabiah	443200453
Lama Albogami	443200175

Outline

- ✓ **Introduction**
- ✓ **Methodology & Data**
- ✓ **Findings & Analysis**
- ✓ **Conclusion & Recommendations**
- ✓ **Lessons Learned**
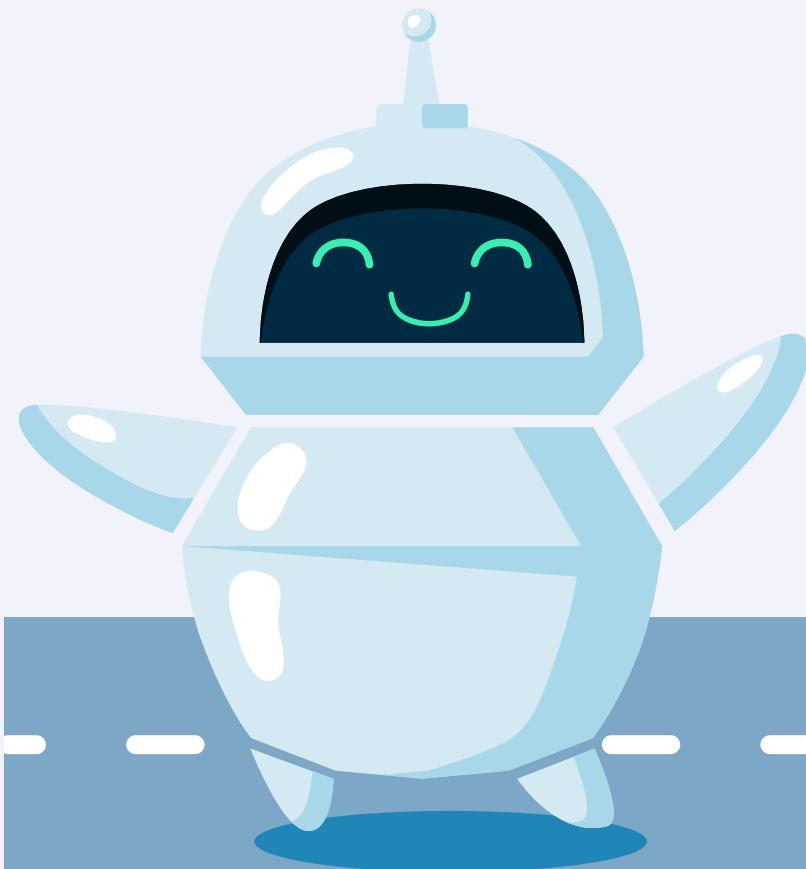
1

Introduction

Project Overview

Problem Statement

University students often struggle to maintain a healthy sleep routine due to academic workload, stress, screen exposure, and inconsistent daily habits. These factors contribute to poor sleep quality, reduced concentration, and lower academic performance.

Objective
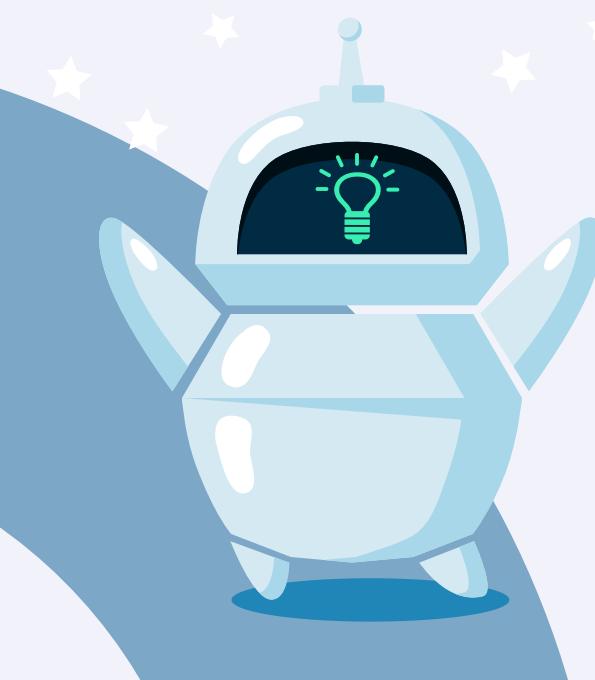
This project aims to analyze the key lifestyle factors that influence students' sleep quality, build predictive models using supervised and unsupervised learning, and generate personalized sleep recommendations using Generative AI.

Business / Social Value
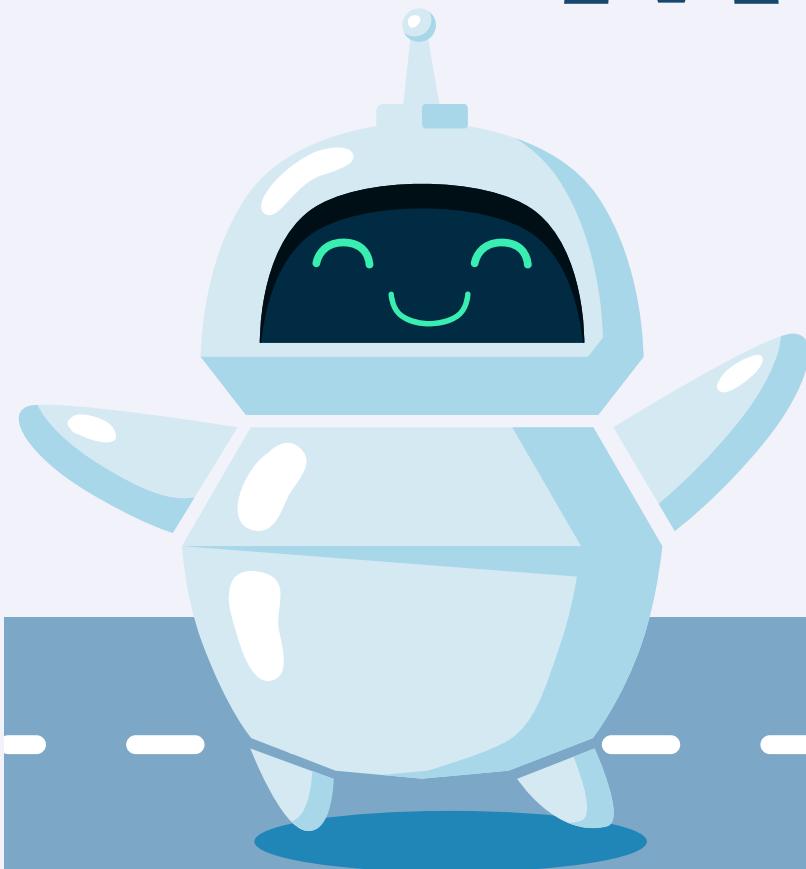
- Boosts student performance.
- Enables data-driven recommendations that promote long-term well being.
- Supports university wellness decisions.
- Scalable for broader health application.

2

Methodology & Data

Dataset Overview

Dataset Selection
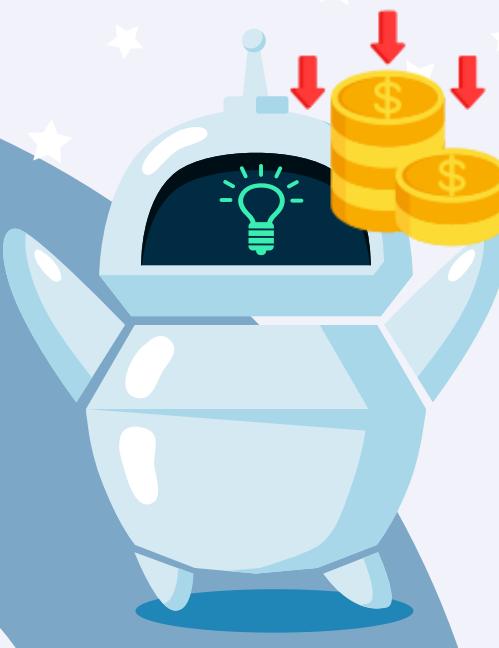
The **Student Sleep Patterns dataset** was chosen because it provides structured, real-world information about university students. It contains no missing values or duplicate records, making it a clean and reliable source for analyzing factors that influence sleep quality.

of Observations and Features

The dataset contains **500** observations and **14** features, including Age, Gender, University Year, Sleep Duration, Study Hours, Screen Time, Caffeine Intake, Physical Activity, Sleep Quality, Weekday Sleep Start, Weekday Sleep End, Weekend Sleep Start, Weekend Sleep End, and Student ID.

Target Variable Description

The target variable in this dataset is **Sleep Quality.** This variable reflects how well the student is sleeping and serves as the main outcome the project aims to analyze and predict. Its distribution is balanced across the dataset, making it suitable for modeling

Data Processing Overview

Generative AI

- Brief Recommendation Template
- Detailed Recommendation Template

Data Cleaning

• Ensured the dataset had no missing or duplicated records.
• Verified data consistency and corrected formatting issues.

Supervised Learning

- Random Forest
- Support Vector Machine – SVM

Exploratory Data Analysis (EDA)

• Analyzed distributions of key variables (sleep, study hours, screen time).
• Identified patterns and early indicators of sleep behavior.

Unsupervised Learning

- K-Means Clustering
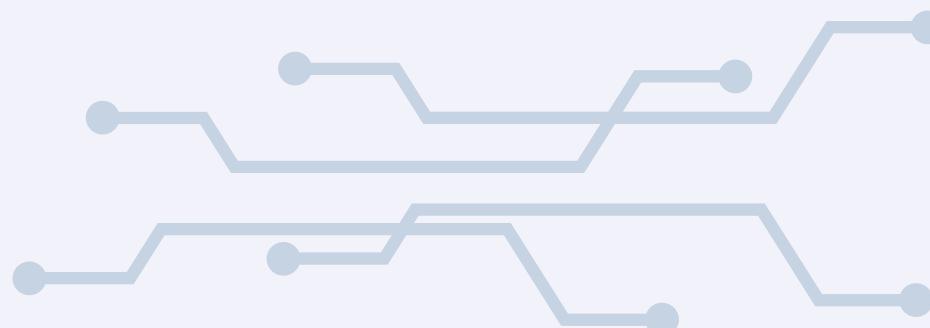
Preparing Data for ML

• Encoded categorical variables for model compatibility.
• Scaled & transformed numeric variables when needed.
• Prepared separate datasets for supervised and unsupervised tasks.

Preprocessing Techniques

Variable Transformation

Categorical fields like Gender and University Year were converted into numeric values for proper model interpretation.

Discretization

Numeric values were grouped into labeled categories (e.g., sleep quality and times) to simplify pattern interpretation.

Variable Removal

Student ID was removed since it adds no predictive value.

Tools Used

> **Python**

> **Pandas**

> **NumPy**

> **Matplotlib / Seaborn**
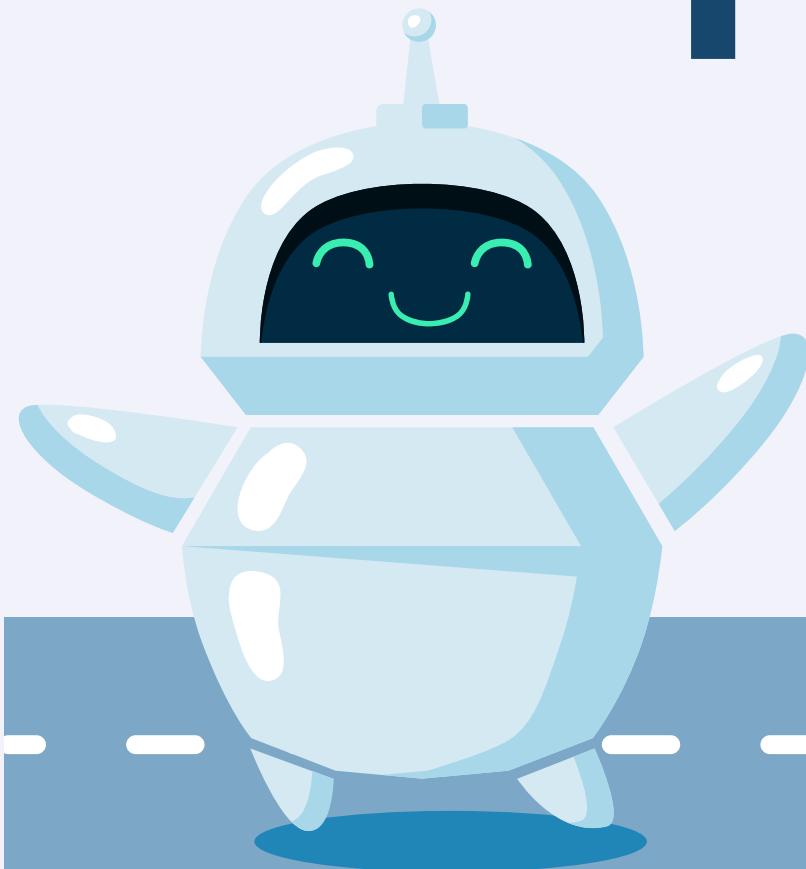
> **Scikit-learn**

> **Groq API + LLaMA Model**

> **Jupyter Notebook + VS Code**

> **GitHub**

3

Findings & Analysis

# Supervised Learning Results: Random Forest vs. SVM
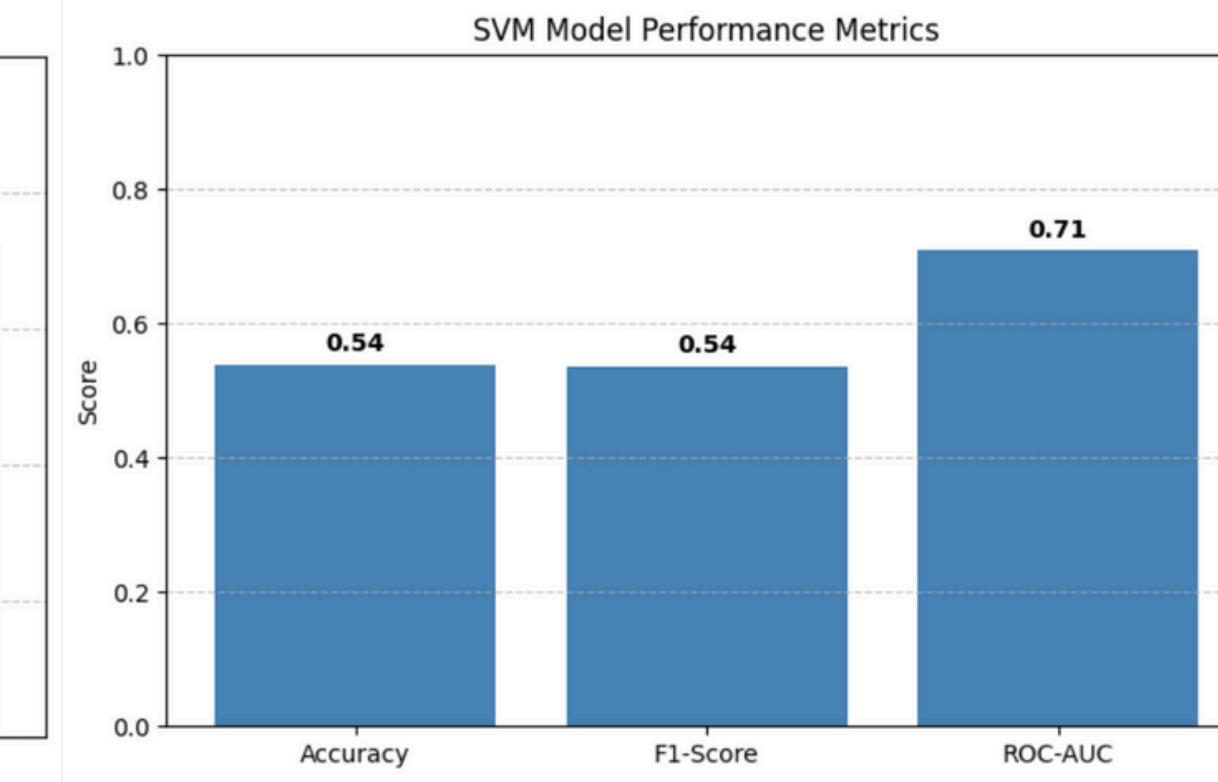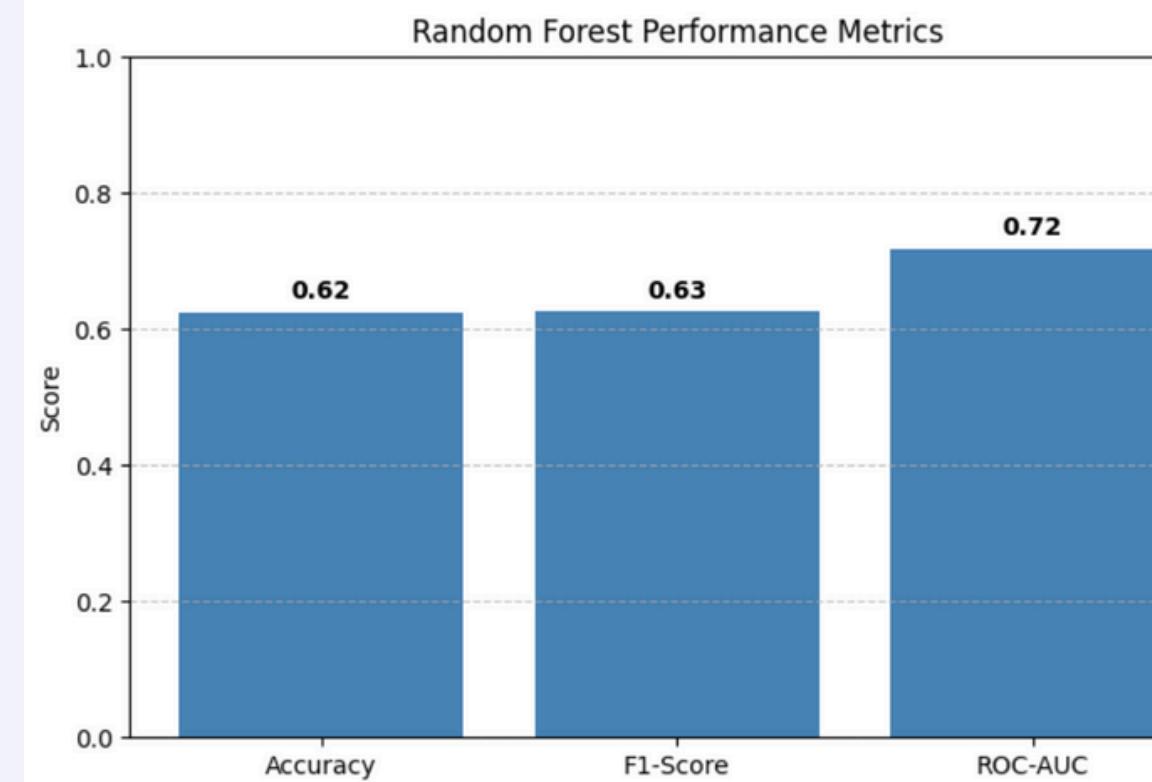
Model Performance Comparison

Random Forest outperformed SVM across all metrics, achieving higher accuracy, F1-score, and ROC-AUC.

Final Model Selection

As a result, **Random Forest** was selected as the **final model** for predicting students' sleep quality because it produced more stable and balanced predictions.

Random Forest Performance Metrics
- Accuracy: 0.62
- F1-Score: 0.63
- ROC-AUC: 0.72

SVM Model Performance Metrics
- Accuracy: 0.54
- F1-Score: 0.54
- ROC-AUC: 0.71
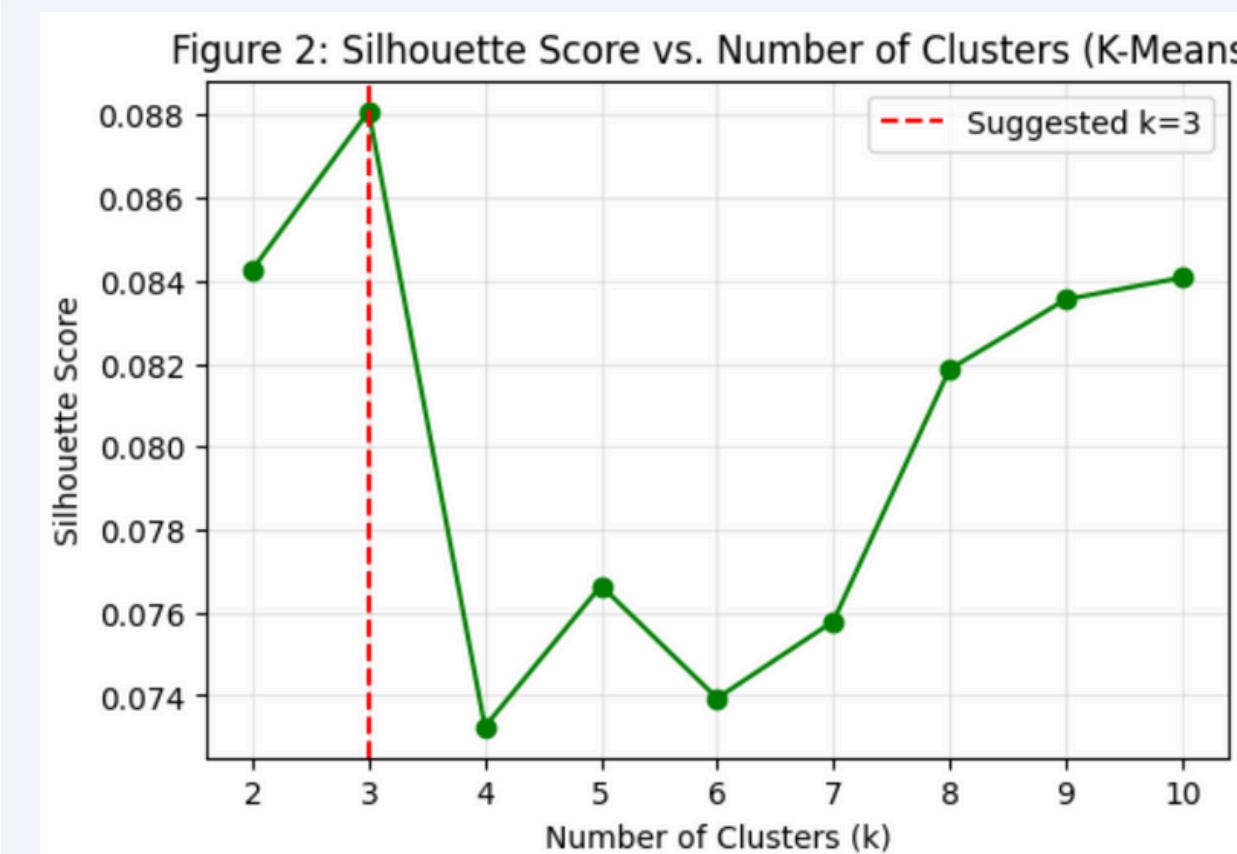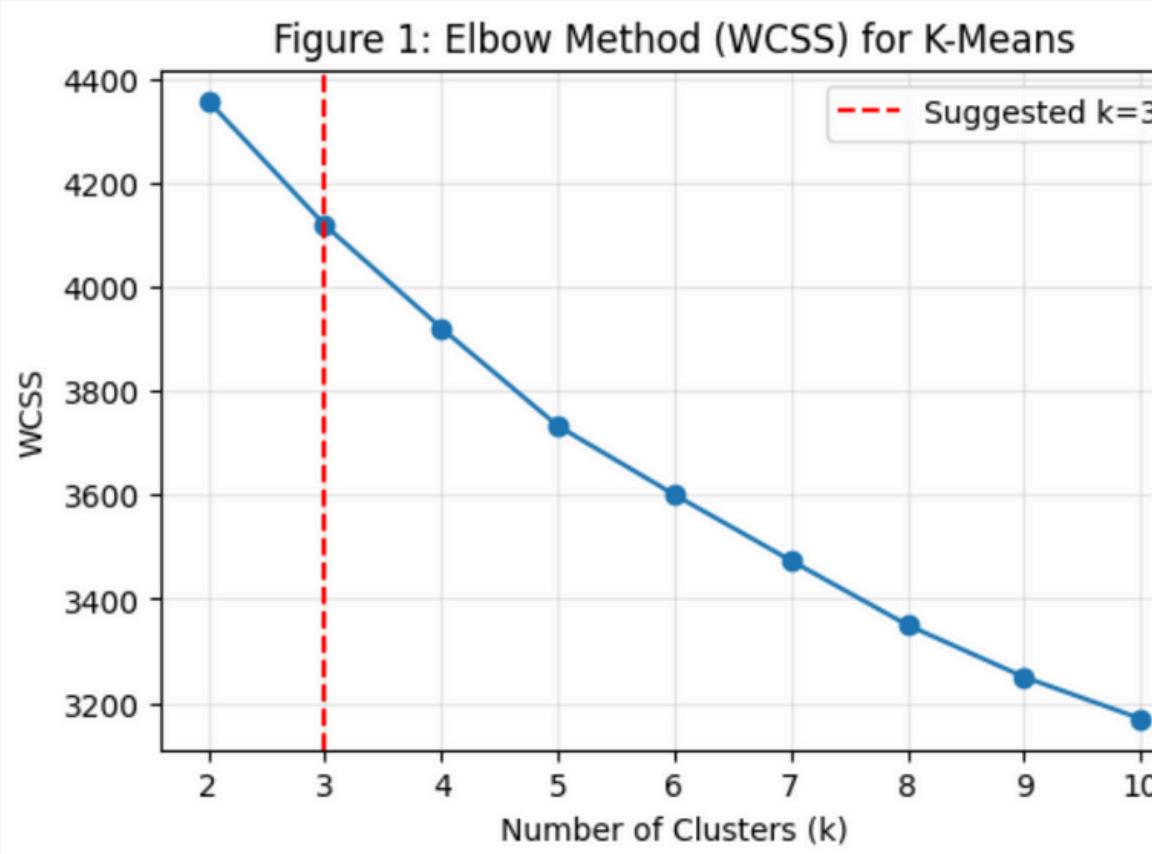
Unsupervised Learning: Clustering Insights (K-Means)

Optimal Number of Clusters
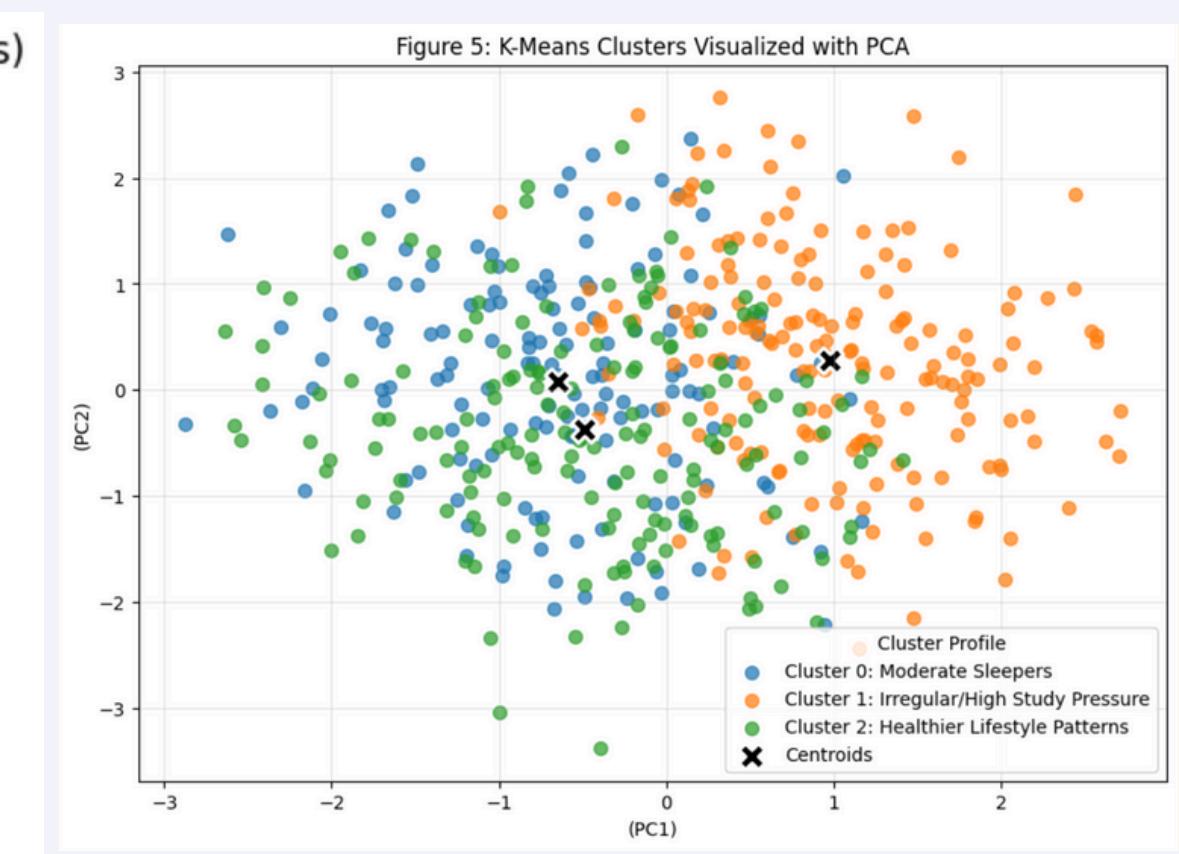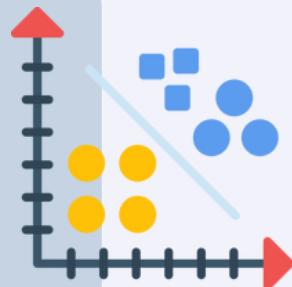
The Elbow Method and Silhouette Score both indicated an inflection point at k = 3, showing that three clusters offer the best balance between compactness and meaningful separation.

Cluster Visualization

The clustering results revealed three student behavior groups. One group showed clear irregular habits under higher study pressure, while the other two groups overlapped, reflecting similar lifestyle patterns.


Figure 1: Elbow Method (WCSS) for K-Means

--- Suggested k=3


Figure 2: Silhouette Score vs. Number of Clusters (K-Means)

--- Suggested k=3

Figure 5: K-Means Clusters Visualized with PCA

Cluster Profile
● Cluster 0: Moderate Sleepers
● Cluster 1: Irregular/High Study Pressure
● Cluster 2: Healthier Lifestyle Patterns
✕ Centroids

Generative AI: Comparing Template Outputs

This example shows the outputs generated by both templates after the user entered their sleep profile. The **brief template** provides a concise recommendation with 1–2 suggestions, while the **detailed template** offers a full analysis with habit evaluation and a multi-step action plan.

```
=========================================================
          TEMPLATE 1: BRIEF RECOMMENDATION
=========================================================
Based on your sleep profile, it seems like you're struggling with not getting enough sleep, especially on weekdays. This can affect your concentration and performance in university.

To improve your sleep, I recommend:

- Try to establish an earlier bedtime routine on weekdays, aiming for 10-11 PM to get at least 7 hours of sleep.
- Limit your screen time before bed by stopping it at least 30 minutes before sleep, and opt for a relaxing activity instead, such as reading or meditation.
```

```
=========================================================
          TEMPLATE 2: DETAILED RECOMMENDATION
=========================================================
**Sleep Recommendation for University Student**

**Short Overview:**
This 24-year-old female university student is facing sleep challenges. Her average sleep duration of 5 hours per night is significantly lower than the dataset average of 6.5 hours. Th

**Analysis:**

- **Sleep Duration**: The student's sleep duration of 5 hours/night is 1.5 hours shorter than the dataset average. This chronic sleep deprivation can lead to impaired cognitive funct:
- **Study Hours and Screen Time**: With study hours of 4 hours/day and screen time of 7 hours/day, the student is likely experiencing a stimulatory environment that interferes with h
- **Caffeine Intake**: Consuming 3 cups of caffeine per day is above the dataset average of 2.5 cups. Caffeine can interfere with sleep quality, especially when consumed in large amo
- **Physical Activity**: While the student engages in 140 minutes of physical activity per week, which is above the dataset average, her overall physical activity level may still be
- **Weekday vs Weekend Schedule**: The student's inconsistent sleep schedule, switching from late sleep to early wake on weekdays and late sleep to late wake on weekends, can disrupt

**Action Plan:**

1. **Gradually increase sleep duration**: Aim for 7-8 hours of sleep each night to align with the dataset average. Establish a consistent sleep schedule by setting a regular bedtime
2. **Set screen-free hours before bed**: Limit screen time to 2 hours before bedtime, and replace screens with relaxing activities like reading or meditation to promote melatonin pro
3. **Reduce caffeine intake**: Limit caffeine consumption to 1-2 cups per day, avoiding it altogether 4-6 hours before bedtime.
4. **Increase physical activity**: Aim for at least 150 minutes of moderate-intensity exercise or 75 minutes of vigorous-intensity exercise per week, and incorporate more physical ac
5. **Create a bedtime routine**: Establish a calming pre-sleep routine, including activities like journaling, deep breathing, or a warm bath, to signal to your body that it's time to

**Final Encouragement:**
By implementing these changes, you'll be well on your way to improving your sleep quality, increasing your energy levels, and enhancing your overall academic performance.
```
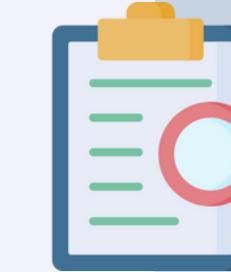
Template Comparison & Final Choice

Quality

- **Template 1 (Brief):** Clear, simple, and suitable for quick guidance.
- **Template 2 (Detailed):** Well-structured and provides deeper, more meaningful insights.

Level of Detail

- **Template 1 (Brief):** Limited detail; focuses on one main issue with 1–2 suggestions.
- **Template 2 (Detailed):** High detail; uses dataset averages and provides a multi-step action plan.

Relevance

- **Template 1 (Brief):** Highly relevant but limited to addressing only the most critical inputs.
- **Template 2 (Detailed):** Highly relevant and contextual.
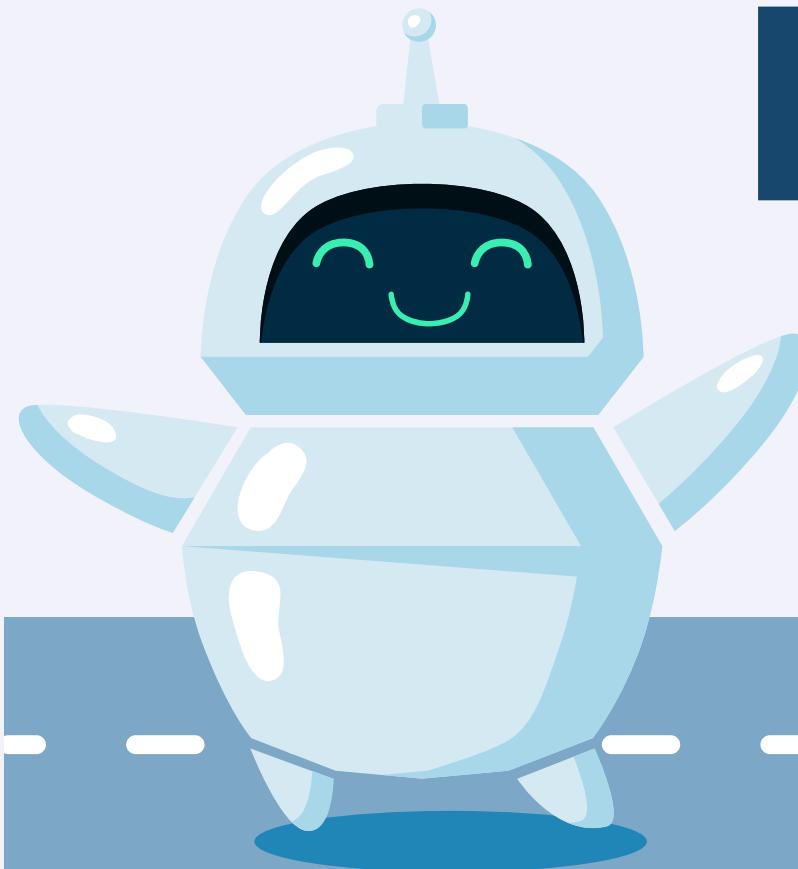
Final Choice
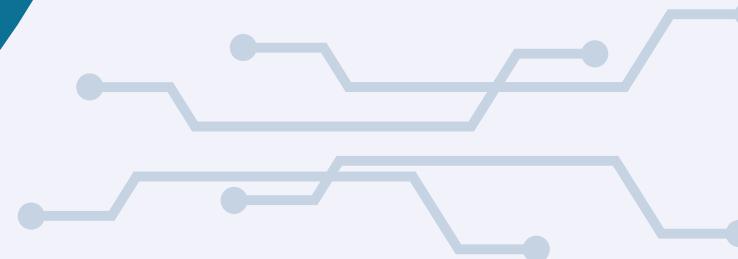
Template 2 was selected because it provides more comprehensive, data-driven, and personalized guidance. Its structured analysis and clear action plan make it more effective than Template 1 for meaningful sleep recommendations.

4

# Conclusion & Recommendations
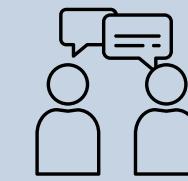
Conclusion
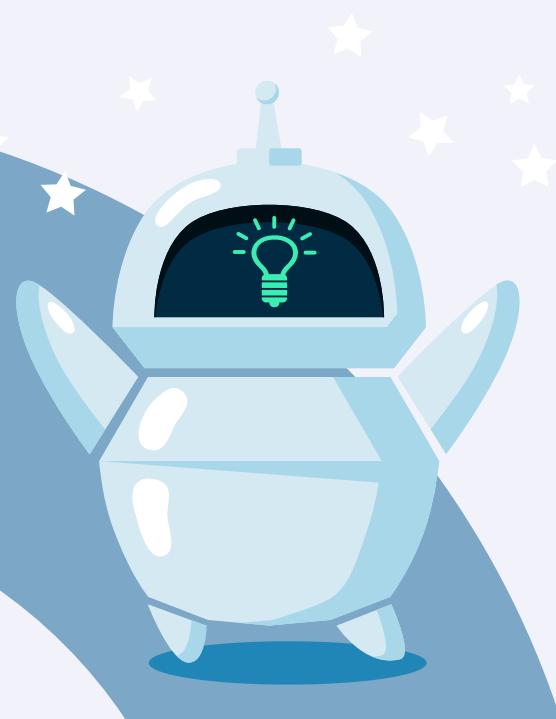
- The analysis shows that students' sleep quality is shaped by their habits especially screen time, study workload, caffeine intake, and consistency in sleep schedules.

- The Random Forest model provided the strongest predictions, while clustering revealed three distinct student behavior patterns.

- These results show that personalized help through AI is more effective than giving everyone the same advice.

Recommendations

Students should keep a stable sleep routine, reduce night screen use, and avoid caffeine
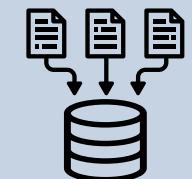
Universities can support healthier habits through quick awareness sessions.

Students can use the AI insights to track and improve their sleep.

Future Work

Expand the dataset with additional factors like stress levels for deeper insights.
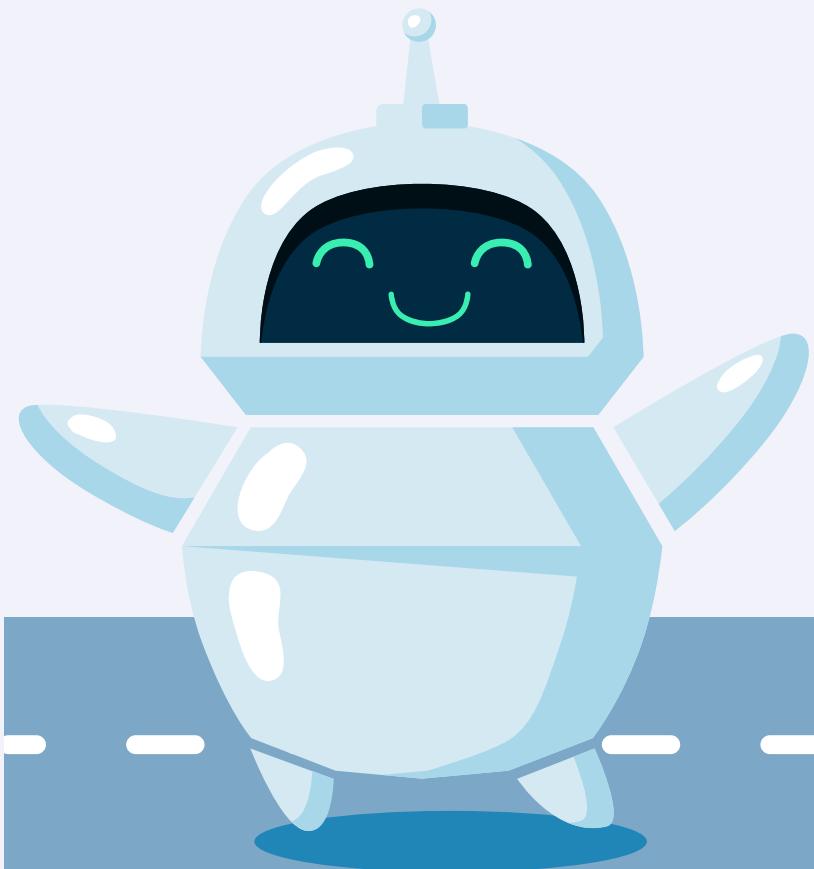
Improve the AI system to deliver more adaptive and interactive guidance.

Develop a dashboard that lets students track their sleep habits and progress over time.

5

Lessons Learned

Project Insights

Exploratory Data Analysis (EDA)
Visualizing data early uncovers trends and guides modeling decisions

Model Selection & Evaluation
Comparing algorithms (Random Forest vs. SVM) shows which model works best for the data
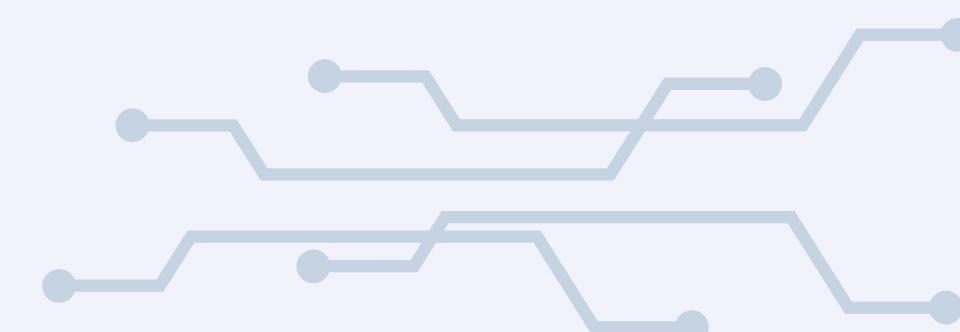
Generative AI Integration (Groq + LLaMA)
Combining AI with data analysis generates practical, actionable recommendations
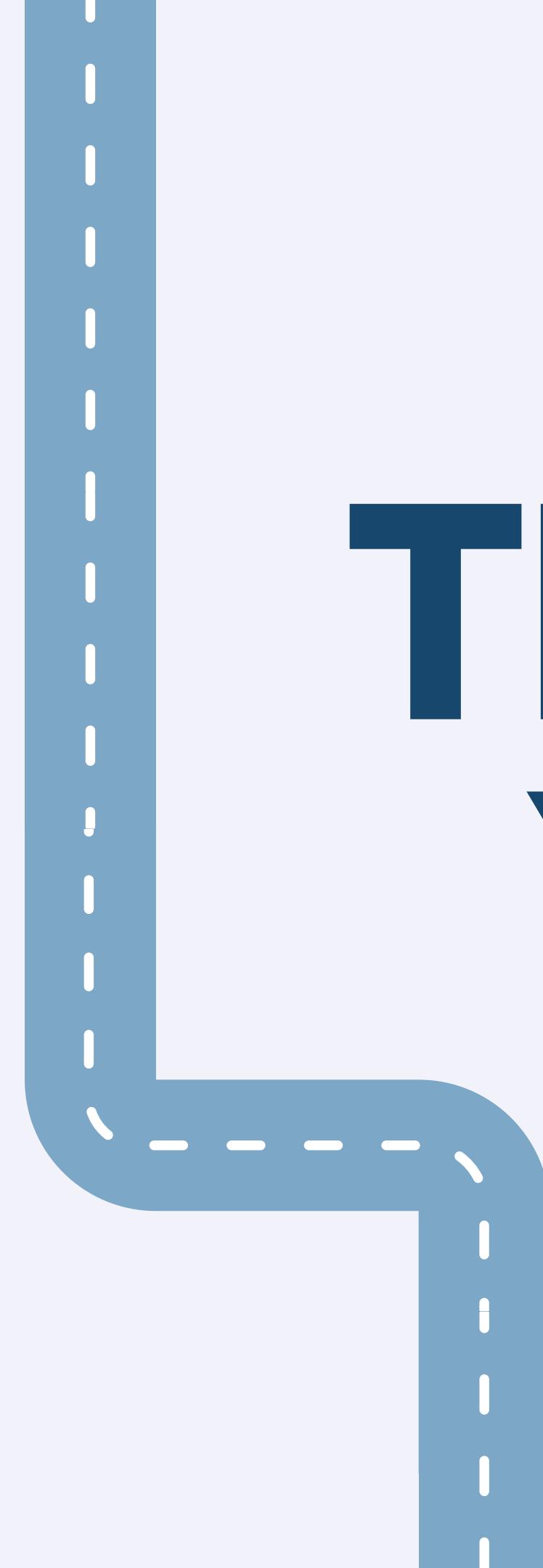
Data Quality & Preprocessing
Clean, well-prepared data significantly improves model accuracy and efficiency

Clustering Insights
Clustering Insights
K-Means reveals hidden student behavior patterns affecting sleep quality

THANK YOU