Conestoga College

School of Applied Computer Science & Information Technology

SENG8081 - Case Studies

How student sleep and lifestyle patterns affect academic performance.

Andrews Owusu

Jai Vadhani

Shivani Sharma

Kulwinder Singh

June 26, 2025

**Abstract**

Sleep is a fundamental component of student well-being and academic success, yet it is often overlooked in higher education environments. This project investigates how sleep patterns, along with lifestyle habits such as screen time, physical activity, study hours, and caffeine consumption affect the academic performance of university students. Using three publicly available datasets (from Kaggle, Mendeley, and RPubs), the analysis aims to uncover meaningful relationships between sleep quality and some lifestyle factors; and how these variables correlate with performance metrics like Grade Point Average (GPA).

The data for the project is organized and managed in a SQL Server relational database to support structured querying, integration, and efficient filtering. Exploratory data analysis, statistical methods, and data visualization tools like Python, Tableau, and R are applied to identify trends and correlations. Key research questions include whether higher screen time contributes to poor sleep, how physical activity influences sleep length, and whether students with low sleep quality perform worse academically. The findings aim to support students, educators, and wellness professionals in recognizing the importance of healthy sleep as a driver of academic and personal success.

**Table of contents**

Contents

[Introduction 4](#_Toc201904354)

[Data Research and Data Collection 5](#_Toc201904355)

[Data Storage, Integration and Maintenance 6](#_Toc201904356)

[References 7](#_Toc201904357)

# Introduction

Sleep is a critical yet often overlooked factor in the academic journey of university students. Amid busy class schedules, late-night studying, and high levels of digital engagement, many students experience irregular or insufficient sleep. While it is widely accepted that proper sleep contributes to physical and mental well-being, its role in academic performance deserves closer examination, particularly in the context of student life where stress, lifestyle habits, and environmental pressures constantly interact.

This project explores how sleep patterns relate to various lifestyle choices, including screen time, study duration, physical activity, and caffeine intake. Our goal is to analyze whether these behaviors have a measurable impact on both sleep quality and academic indicators such as grade point average (GPA), stress levels, and overall student performance. By using real-world data, and also synthetic data generated based on real-world scenarios, we aim to uncover patterns and correlations that may explain how quality sleep may contribute to academic excellence.

To achieve this, we have gathered and analyzed data from three publicly available sources: Kaggle’s Student Sleep Patterns dataset, Mendeley’s Insomnia and College Performance dataset, and a structured dataset from RPubs. These datasets have been integrated into a SQL Server relational database for efficient organization and querying. Through statistical analysis and data visualization, we aim to draw insights that not only enhance our understanding of student sleep behavior but also inform future strategies for promoting academic success through improved wellness practices.

# Data Research and Data Collection

In order to determine appropriate datasets, we used search terms such as "college sleep behavior", "student sleeping patterns", and "sleep and academic performance" on Google Dataset Search, Kaggle, GitHub, and RPubs. This helped us to locate datasets that aligned with our project expectations. Having evaluated a few sources, we chose three sets of data that are close to our study objectives:

* Student Sleep Patterns (Kaggle, by Arsalan Jamal):
* Insomnia of students and college performance (Mendeley):
* RPubs Sleep Study Dataset:

The dataset contains information on students’ sleep duration, quality of sleep, screen time, physical activity, and overall health, among other factors. Each row represents a student, with columns which include their age, gender, university year, sleep duration and timing (weekday and weekend start and end times), study hours, screen time, caffeine intake, physical activity levels, and self-reported sleep quality. It includes a manageable number of records and is available in a clean CSV format. It has consistent data types, making it easy to load and process using Python libraries like pandas or visualization tools such as Tableau or Power BI.

* **Student Sleep Patterns Dataset (Kaggle, by Arsalan Jamal):**

The data has a good structure and the variables present in the data are the hours of sleep, quality of sleep, the time spent using screens, stress, study time, and physical activity. It is best suited to conduct statistical and visual analysis to determine some patterns of student sleeping behavior.

* **Insomnia of students and college performance Dataset (Mendeley):**

This is a collection of the survey data of 985 students that covers information regarding the frequency of insomnia, their GPA, stress, and characteristics of their lifestyle. It offers great insights concerning the effect of sleep disorders on academic performance.

* **RPubs Sleep Study Dataset:**

This dataset contains structured variables of sleep in a clean form, which is appropriate in regression, correlation analyses, and visualizations.

# Data Storage, Integration and Maintenance

In this project, the databases will be in a SQL server relational database to make it easy to organize, integrate, and analyze. Each dataset will be stored in separate tables, and relationship between tables will be established by common fields.

**Keeping information in SQL server will have the following benefits:**

* Relational structure: Allows SQL joins to be able to join data on different tables to get more insight.
* Central accessibility: The data source is handled using a single control, without complicating update and queries.
* Data cleaning and filtering: It is possible to clean and filter the data with SQL queries prior to analysis.
* Export flexibility: It is quick and simple to export cleaned and combined datasets to work with Python, R, Power BI, or excel.
* Security and control: SQL Server has features of built-in access control, and backup, versioning.

## 

# References