DS210 Final Project: Analyzing Student Sleep Patterns

Project Overview

This project aims to analyze a dataset of student sleep patterns to explore the impact of academic and lifestyle factors on sleep duration, quality, and habits. Using Rust, the project demonstrates data analysis techniques, including descriptive statistics, graph-based modeling, and clustering, to uncover meaningful patterns. The findings provide insights into how students manage their sleep during academic pressures and different lifestyle choices.

Research Questions

How do academic and lifestyle factors influence students' sleep patterns?

Sub-questions:

- Is there a relationship between study hours and sleep duration?
- How do screen time and other lifestyle factors correlate with bedtime and wake-up times?
- Are there specific "clusters" of students with similar sleep patterns?

To address these questions, the project integrates data science techniques:

- Descriptive analysis to summarize patterns and variability.
- Graph modeling to represent relationships between sleep patterns.
- Clustering to group students with similar habits.

Dataset

The dataset has 500 rows, representing an individual student's sleep and lifestyle metrics.

Key attributes :

- Student_ID: A unique identifier for each student
- Demographics: Age, gender, and university year
- Sleep and Lifestyle Metrics: Sleep duration, study hours, screen time, caffeine intake, physical activity, and sleep quality

Analysis and Results

Descriptive Statistics:

The mean sleep duration was 6.47 hours, and the median was 6.50 hours. The average sleep duration is below the recommended 7–9 hours, suggesting widespread sleep deprivation among students. A narrow gap between mean and median indicates a relatively symmetric distribution

of sleep duration. The data shows that academic pressures and lifestyle habits like screen time are likely contributing to reduced sleep duration.

Graph Representation:

- Number of Nodes: 4 (university years)
- Number of Edges: 12 (connections based on similar sleep patterns)

The graph shows high interconnectivity between university years, highlighting shared sleep habits across grades. Graph analysis reveals how sleep habits evolve across academic stages and identifies commonalities in student behaviors.

Clustering Analysis:

- Cluster 1: 155 students, likely balanced habits (moderate sleep, study hours, screen time)
- Cluster 2: 173 students, likely long study hours with reduced sleep duration
- Cluster 3: 172 students, likely high screen time and irregular sleep schedules

The clusters clearly segments students' habits, showing some patterns and behaviors. The clustering demonstrates how academic and lifestyle factors create distinct sleep pattern groups, addressing the diversity in student experiences. Students in Cluster 2 demonstrate a likely inverse relationship between study hours and sleep duration, aligning with the hypothesis that longer study hours reduce sleep. Students in Cluster 3 exhibit high screen time, correlating with irregular sleep schedules and potentially poorer sleep quality.

How to Run the Project

Have rust installed and place the dataset in the directory. Then you need to clone lone the repository with git clone (repository link). Then use cd in terminal to switch to the directory of the project. Then use cargo run to run the code and can do cargo test to test it as well.

Looking at the Code

The project exceeds 150 lines of code and has modules:

- data.rs: Data parsing and struct definition.
- analysis.rs: Statistical calculations.
- graph.rs: Graph construction and analysis.
- clustering.rs: Clustering implementation.

I also made tests for the data parsing part, statistical calculations, graph constructions, and the clustering module. We also have multiple git commits reflecting my incremental progress.

Future Areas for Improvement

I could do some correlation analysis by computing the relationships between study hours, screen time, and sleep duration using correlation coefficients. Using centrality measures like closeness could identify the most influential sleep patters. Data visualization with graphs and clusters can enhance how I interpret the data.

Conclusion

ovuncovering patterns and relationships in student sleep data. The findings highlight the interplay between academic pressures, lifestyle habits, and sleep quality. Rust's performance and safety make it an excellent choice for handling such data-intensive tasks, demonstrating its capabilities in real-world data science applications.