Efe Akpınar

28869

Final Report

“Sleep Quality Analysis”

Motivation

The motivation behind this project was to gain a deeper understanding of personal sleep patterns and physical activity metrics over a nearly one-year period. By analyzing the impact of academic schedules, exams, and vacations on sleep quality and physical health, the goal was to identify actionable insights for optimizing sleep and maintaining a healthy lifestyle. This analysis was driven by the desire to improve stress management, enhance academic performance, and ensure overall well-being.

Data Source

Wearable Devices:

Metrics such as Total_Steps, Total_Calories_Burned, and sleep phases (REM, Deep, and Core sleep) were collected using a wearable device.

Manual Logging:

Contextual factors, including School Semester, Vacation, and Finals Week, Sleep_Quality were manually recorded to enrich the dataset.

Data Export:

The data was exported from the Apple Health app (export.xml) and export.xml processed into a structured format for analysis.

“The dataset spans from October 30, 2023, to October 14, 2024.”

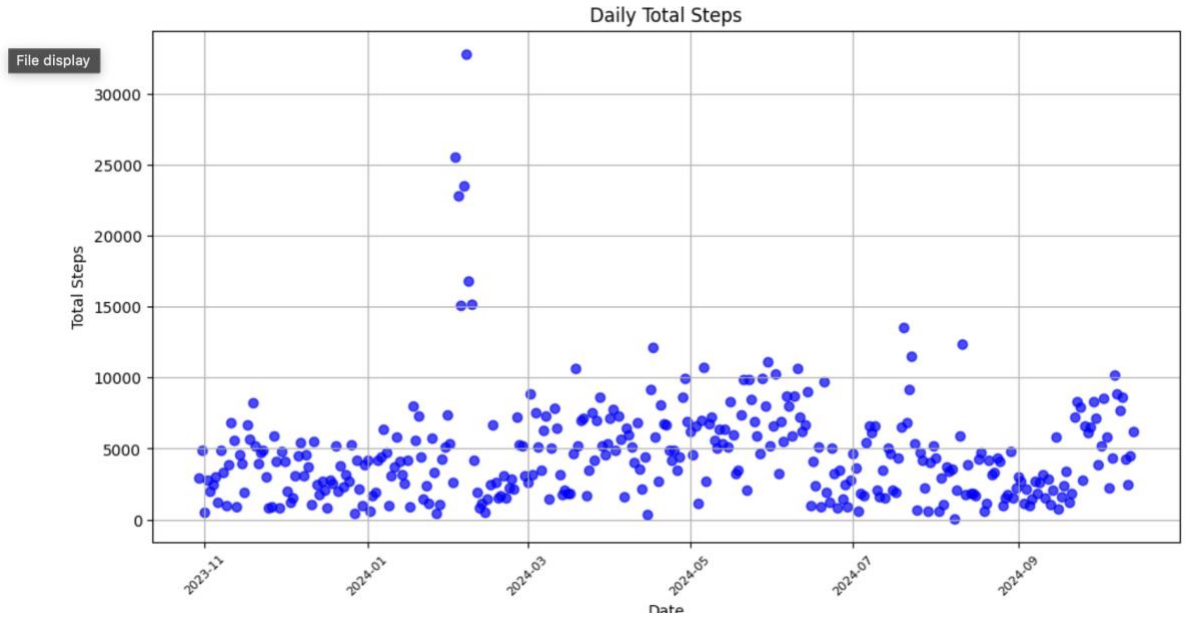
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 351 entries, 0 to 350
Data columns (total 11 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   Unnamed: 0             351 non-null    int64
 1   Date                   351 non-null    datetime64[ns]
 2   Total_Steps            351 non-null    int64
 3   Total_Calories_Burned  351 non-null    float64
 4   Total_Sleep_Hours      351 non-null    float64
 5   Context                351 n           category
 6   REM_Sleep              351 non-null    float64
 7   Deep_Sleep             351 non-null    float64
 8   Core_Sleep             351 non-null    float64
 9   Sleep_Quality          351 non-null    object
 10  Month                  351 non-null    period[M]
dtypes: category(1), datetime64[ns](1), float64(5), int64(2), object(1), period[M](1)
memory usage: 28.0+ KB
```

File display

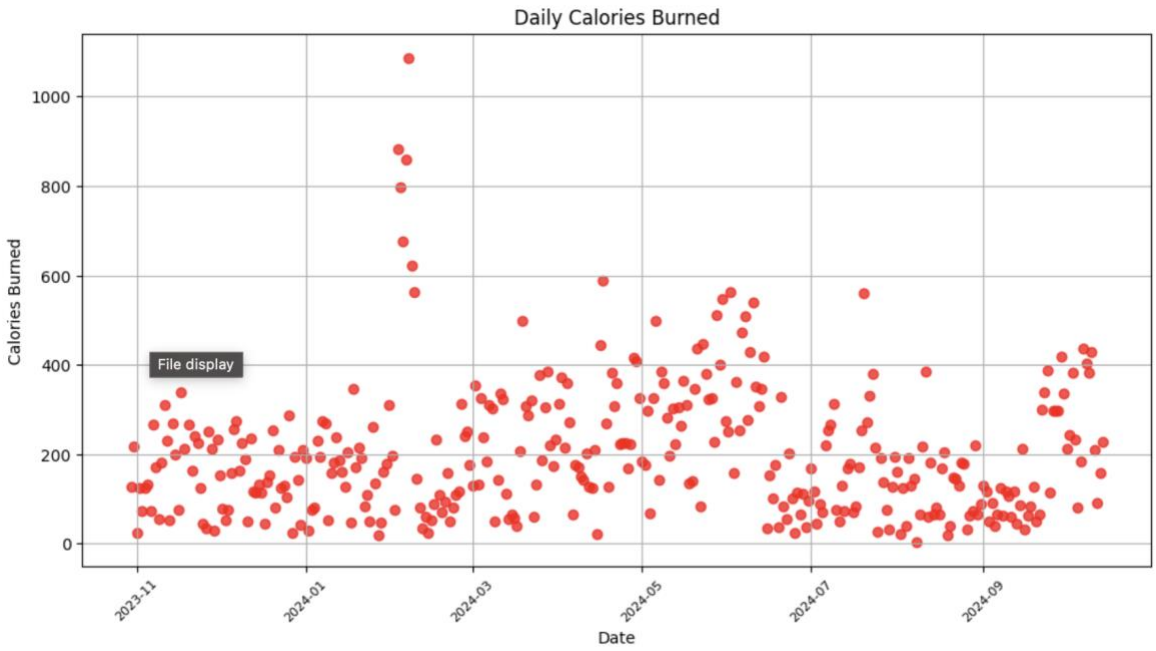
Data Analysis

First EDA conducted on each primary column:

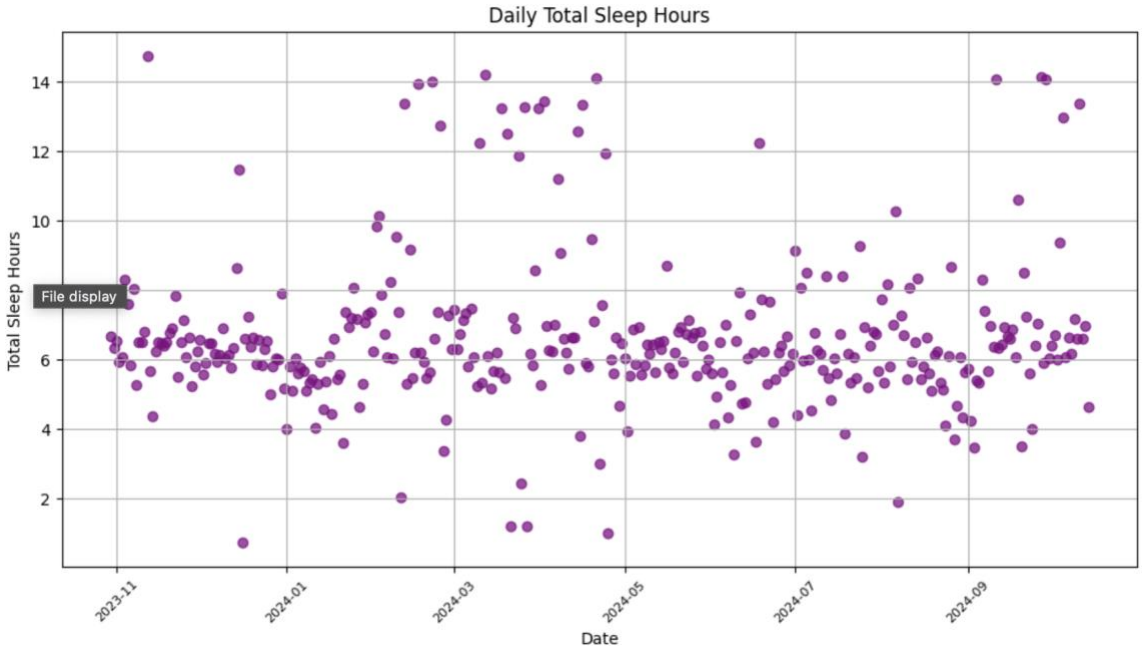
Plot total steps day by day as a scatter plot:



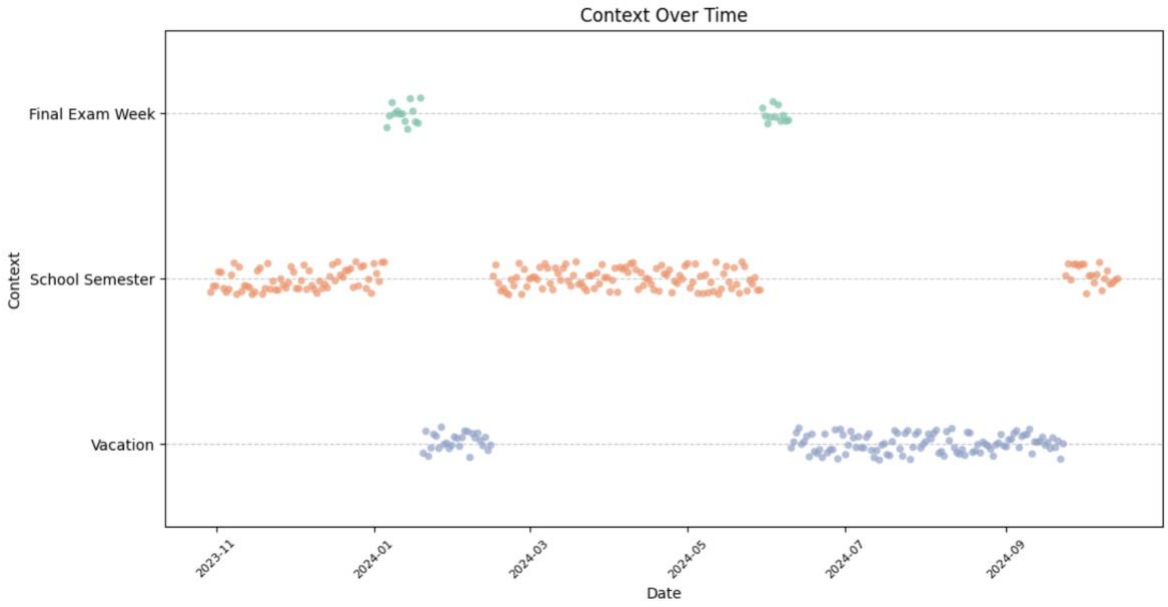
Plot total calories burned day by day as a scatter plot:



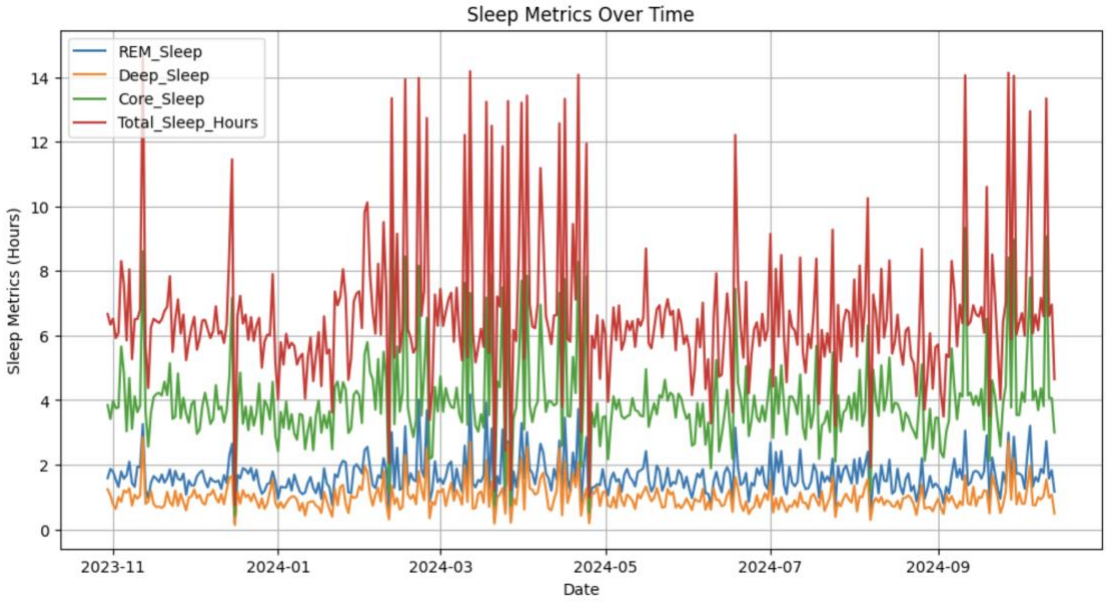
Plot total sleep hours day by day as a scatter plot:



Plot Context by Date using a strip plot:



Plot REM, DEEP_SLEEP, CORE_SLEEP, TOTAL_SLEEP_HOURS over time:



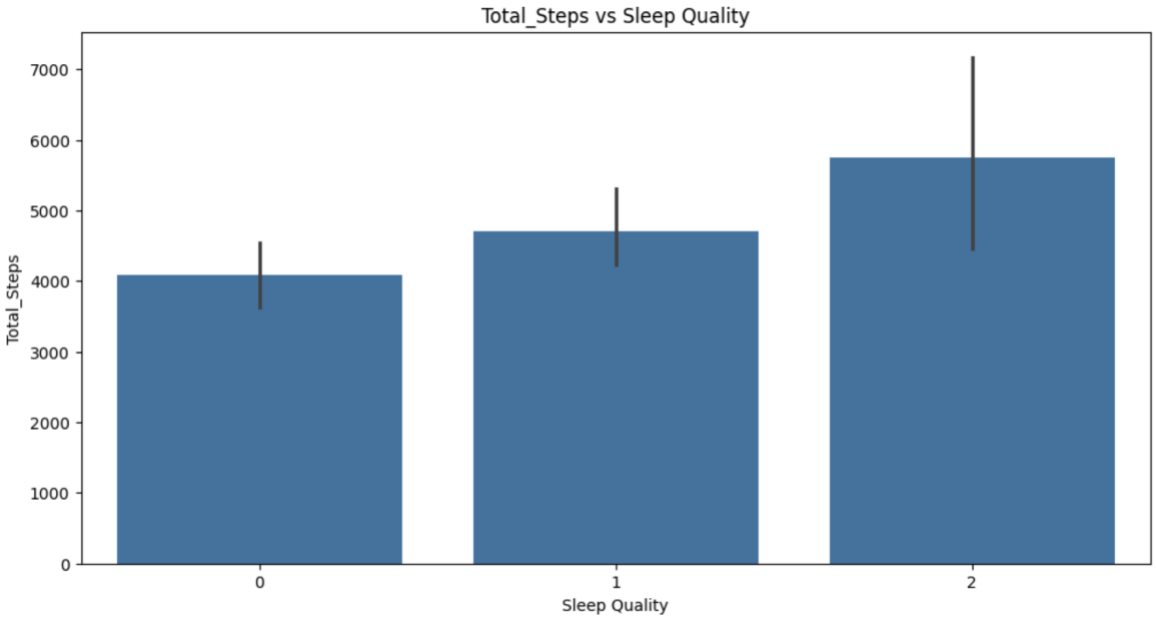
Analysis of Sleep Quality

1. Physical Activity vs Sleep Quality:

Correlation:

A strong correlation was observed between Total_Steps and Total_Calories_Burned, leading to the decision to use Total_Steps as the representative metric for physical activity.

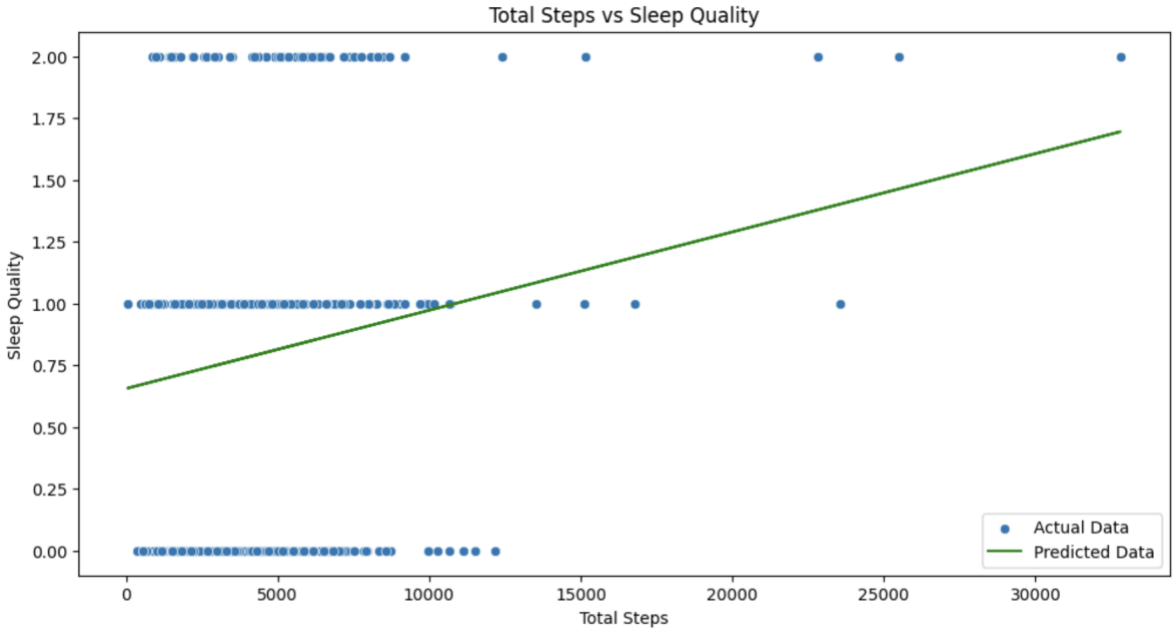
A positive correlation was also found between physical activity and sleep quality.



Regression Analysis:

Physical activity was shown to contribute positively to both sleep duration and sleep quality, although the strength of the relationship varied.

The scatter plot and regression line visually confirmed these relationships.



R^2 Score:

The regression model explained a portion of the variance in sleep quality, highlighting physical activity as a contributing factor.

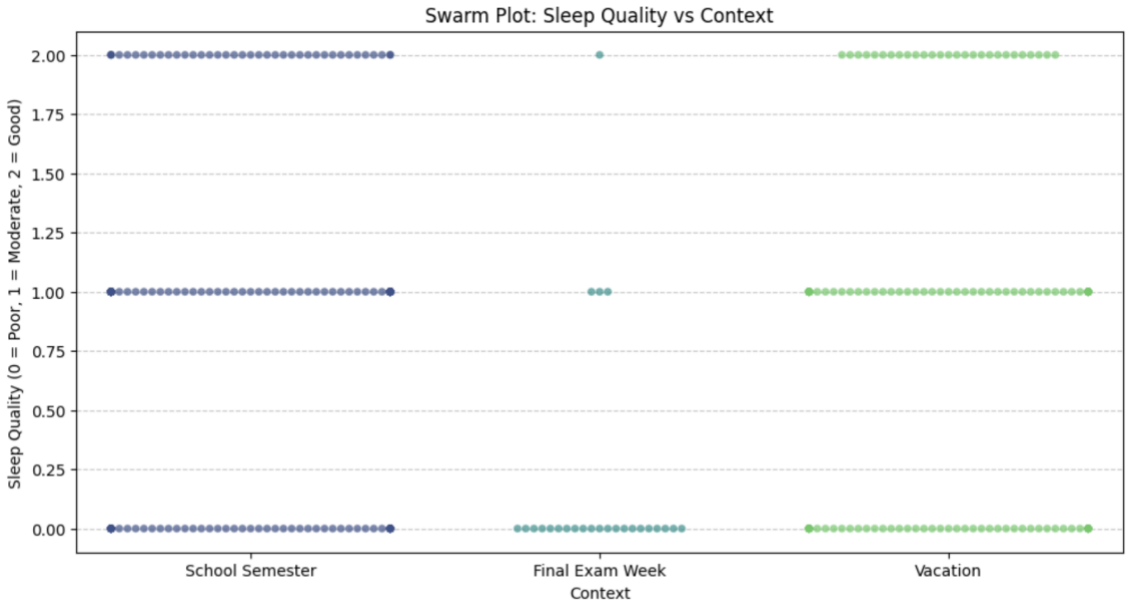
2. Context (School Semester/Vacation/Finals Week) vs Sleep Quality:

Swarm Plot:

Sleep quality was highest during vacations, with most data points clustered around 2 (Good Sleep Quality).

Sleep quality dropped significantly during finals week, with many observations at 0 (Poor Sleep Quality).

During school semesters, sleep quality showed a balanced distribution across 0, 1, and 2.



Insights:

Vacations positively impacted sleep quality due to reduced stress, while finals week had a clear negative impact.

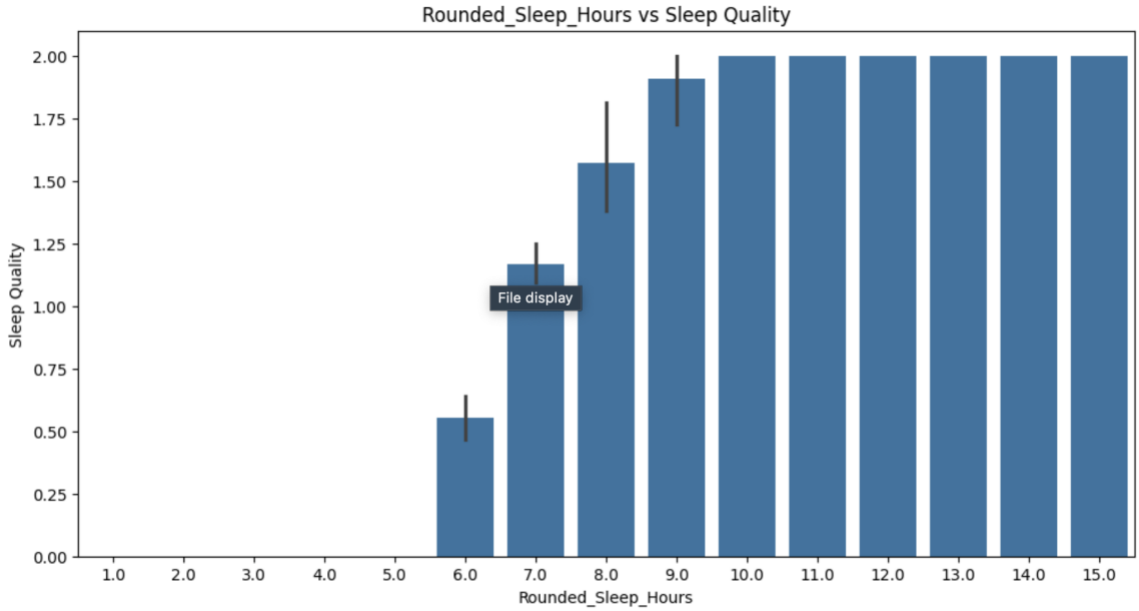
3. Time of Sleep vs Sleep Quality:

Rounded Sleep Hours:

Sleep hours were rounded to the nearest whole number to simplify the analysis.

Bar Plot:

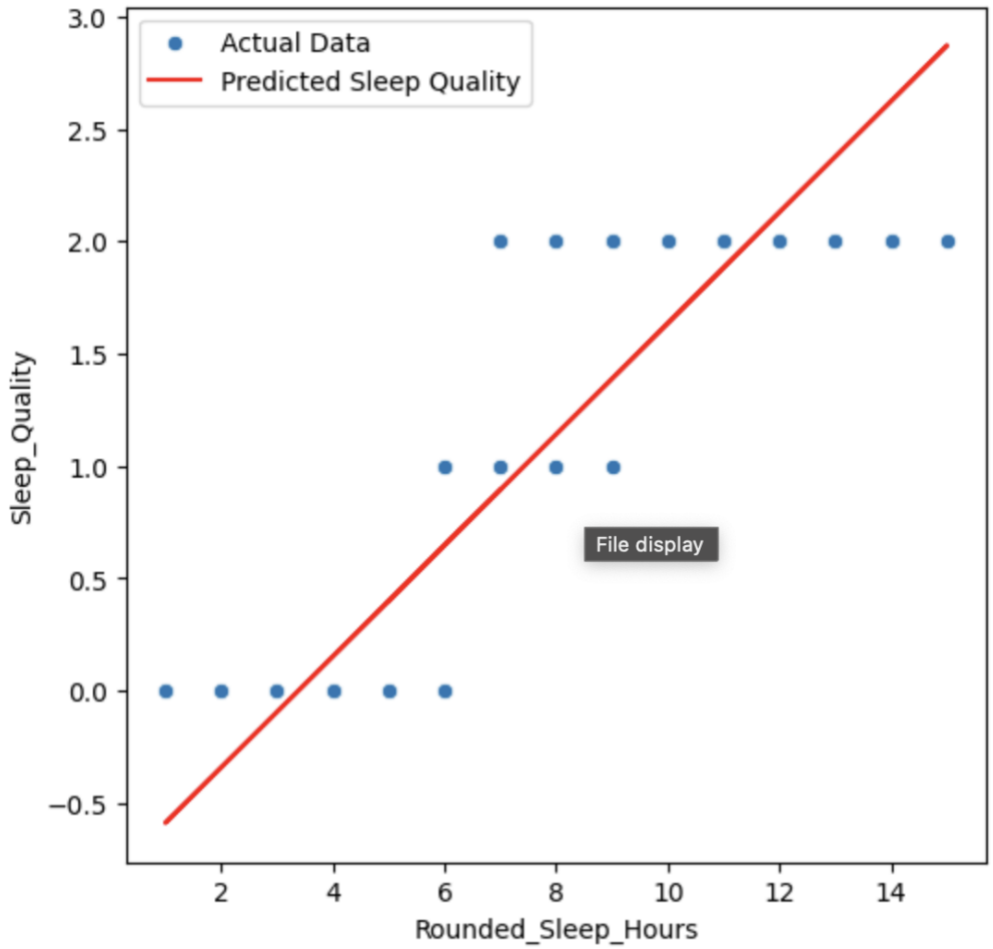
Shorter sleep durations (≤5 hours) were associated with poor sleep quality (0), while longer durations (≥8 hours) were linked to good sleep quality (2).



Linear Regression:

A positive trend was observed between rounded sleep hours and sleep quality.

The regression model showed that longer sleep durations significantly improved sleep quality.



4. Predictive Modeling for Sleep Quality:

Machine Learning Model:

A Random Forest Classifier was trained to predict Sleep_Quality using features such as:

Total_Steps

Total_Sleep_Hours

Contextual variables like School Semester and Vacation.

Results:

```
Classification Report:
              precision    recall  f1-score   support

           0       1.00      0.93      0.96        27
           1       0.81      0.86      0.83        29
           2       0.73      0.73      0.73        15

    accuracy                           0.86        71
   macro avg       0.85      0.84      0.84        71
weighted avg       0.86      0.86      0.86        71

Confusion Matrix:
[[25  2  0]
 [ 0 25  4]
 [ 0  4 11]]
Feature Importances:
                   Feature  Importance
1        Total_Sleep_Hours    0.892920
0              Total_Steps    0.092337
2  Context_School Semester    0.009330
3         Context_Vacation    0.005413
```

Overall Accuracy: 86%

Key Metrics:

Precision: 1.00 for Poor, 0.81 for Moderate, 0.73 for Good.

Recall: 0.93 for Poor, 0.86 for Moderate, 0.73 for Good.

Feature Importance:

Total_Sleep_Hours contributed most significantly to predictions, followed by Total_Steps, while contextual factors had minor impacts.

Insights:

The model demonstrated strong predictive ability, particularly for Poor and Moderate sleep quality.

The confusion matrix highlighted occasional misclassifications, especially between Moderate and Good sleep quality, reflecting overlapping patterns.

Findings

Physical Activity:

Increased physical activity, measured by Total_Steps, positively impacts both sleep duration and quality.

Contextual Impact:

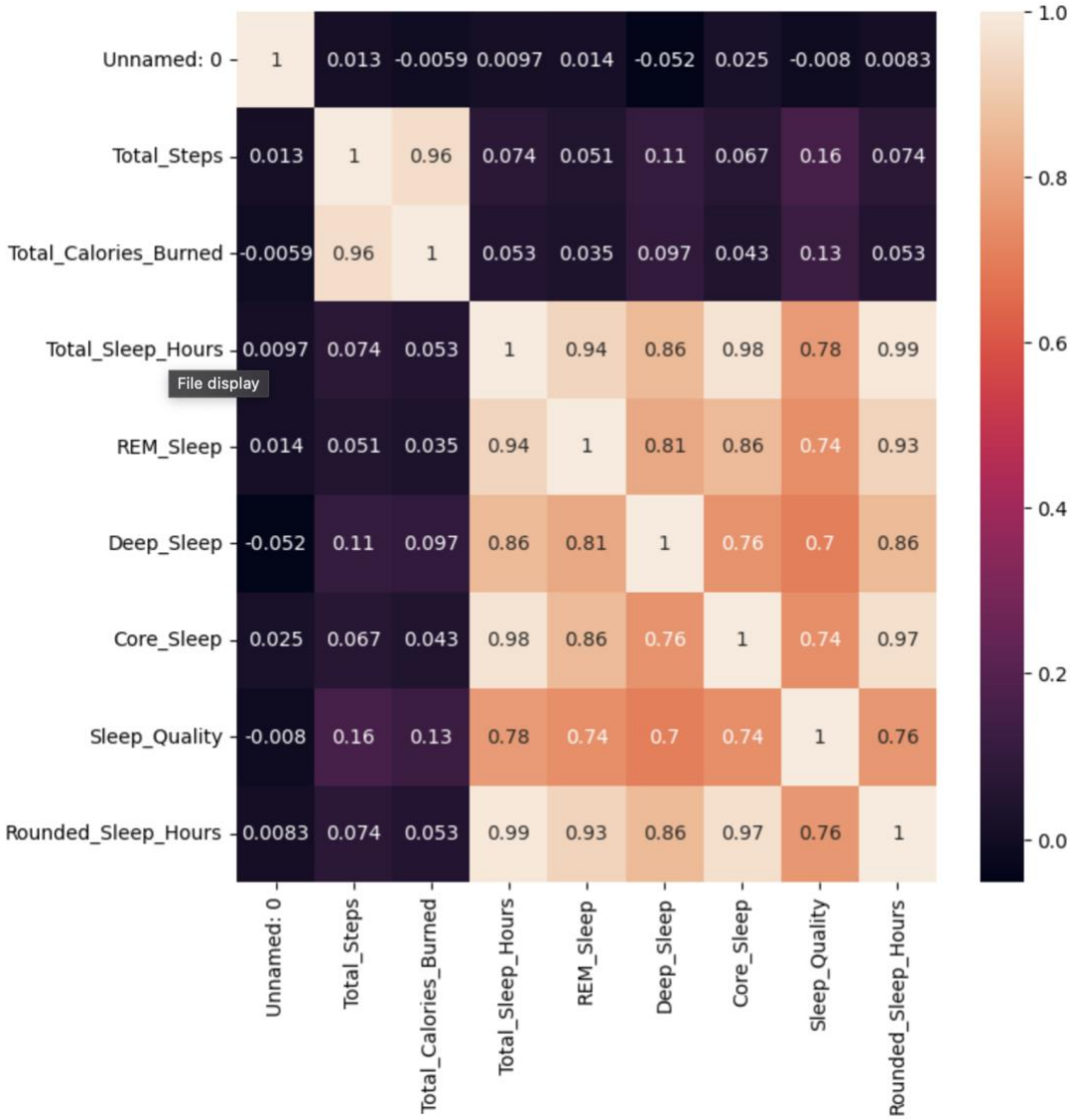
Vacations lead to better sleep quality, while finals week significantly reduces it. School semesters reflect a balanced pattern depending on individual schedules and stress levels.

Sleep Duration:

Sleep duration is the most significant predictor of sleep quality. Consistently longer sleep hours improve sleep quality outcomes.

Machine Learning:

Predictive modeling provided actionable insights, with sleep duration emerging as the most critical factor for determining sleep quality.


Unnamed: 0
Total_Steps
Total_Calories_Burned
Total_Sleep_Hours
REM_Sleep
Deep_Sleep
Core_Sleep
Sleep_Quality
Rounded_Sleep_Hours
1 0.013 -0.0059 0.0097 0.014 -0.052 0.025 -0.008 0.0083
0.013 1 0.96 0.074 0.051 0.11 0.067 0.16 0.074
-0.0059 0.96 1 0.053 0.035 0.097 0.043 0.13 0.053
0.0097 0.074 0.053 1 0.94 0.86 0.98 0.78 0.99
0.014 0.051 0.035 0.94 1 0.81 0.86 0.74 0.93
-0.052 0.11 0.097 0.86 0.81 1 0.76 0.7 0.86
0.025 0.067 0.043 0.98 0.86 0.76 1 0.74 0.97
-0.008 0.16 0.13 0.78 0.74 0.7 0.74 1 0.76
0.0083 0.074 0.053 0.99 0.93 0.86 0.97 0.76 1
1.0
0.8
0.6
0.4
0.2
0.0

Limitations and Future Work

Limitations:

Contextual factors were manually logged, which may introduce errors.

External influences such as diet, screen time, and stress levels were not comprehensively captured.

The analysis was limited to one year, restricting long-term trend evaluation.

Future Work

Expand Data Collection: Incorporate additional factors like stress, diet, and screen time for a holistic analysis.

Advanced Modeling: Employ additional machine learning models (e.g., Gradient Boosting, Neural Networks) for better predictions.

Behavioral Simulations: Simulate the impact of lifestyle changes, such as increasing physical activity or maintaining consistent sleep schedules, to identify actionable recommendations.

Conclusion

This project provided critical insights into the interplay between sleep patterns, physical activity, and contextual factors. By understanding the impact of academic schedules and lifestyle habits, actionable strategies can be developed to enhance sleep quality and overall well-being. Predictive modeling further reinforced the importance of sufficient sleep duration and physical activity in achieving good sleep quality for me. Future iterations of the project will aim to expand data sources and refine modeling techniques for even deeper insights.