Weather and Productivity Analysis

# Project Overview

This project explores the impact of daily weather conditions on productivity, defined by time spent on work-related applications (e.g., IDEs, Google Docs, email). By combining personal screen time, sleep data, and weather statistics, the study reveals how environmental factors—alongside academic events like exams—affect behavior and focus levels.

# Datasets

1. Weather Data: Includes temperature, humidity, precipitation, wind speed, cloud cover, and weather type.  
2. Screen Time & Productivity Data: Tracks time spent on work apps, social media, entertainment, and includes sleep duration and exam periods.

# Research Questions

- Does weather affect time spent on work-related applications?  
- Is social media usage correlated with temperature?  
- Does cloudy weather increase gaming time?  
- Does more sunlight increase productivity?  
- Do exam periods influence behavior?

# Hypotheses

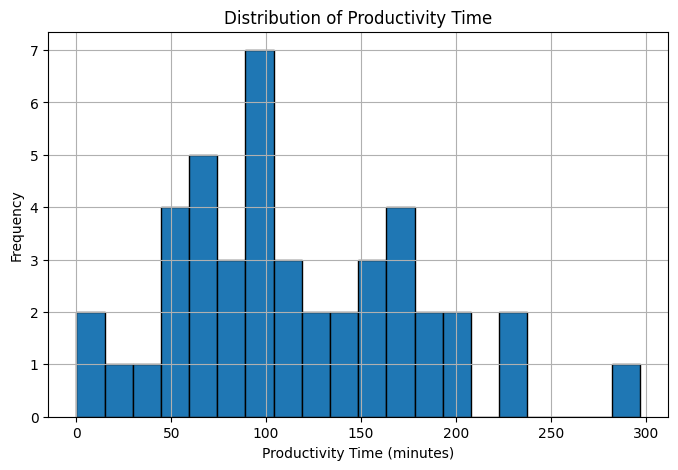
Null Hypothesis (H₀): Weather conditions and exam periods have no significant impact on productivity.  
Alternative Hypothesis (H₁): Weather conditions and exam periods significantly affect productivity.

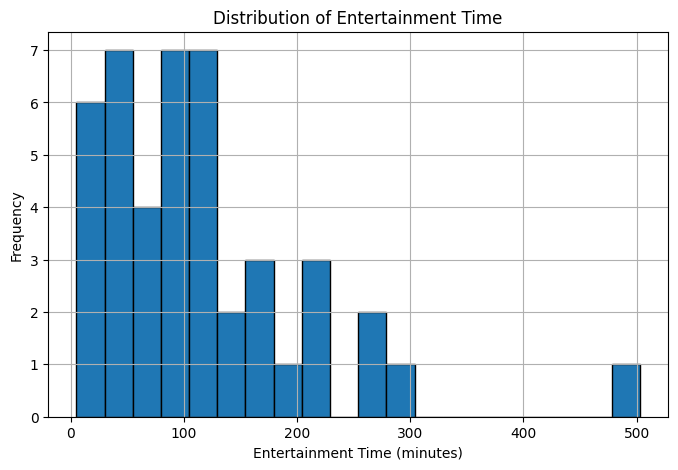
# Methods

Daily data was collected over a 14-week academic term. After cleaning and merging, exploratory data analysis, statistical testing, and machine learning models were applied.

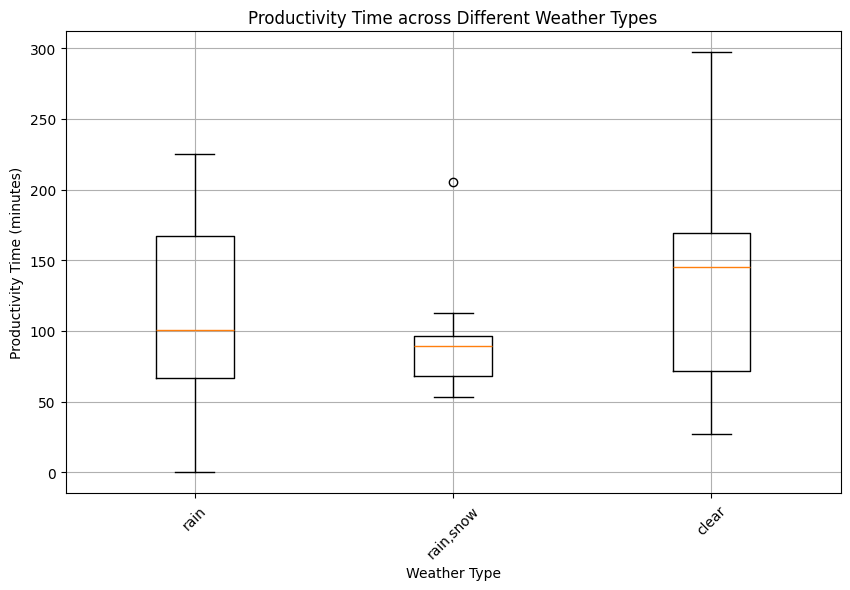
# Exploratory Data Analysis (EDA)

Histograms of productivity and entertainment time show user engagement distributions.

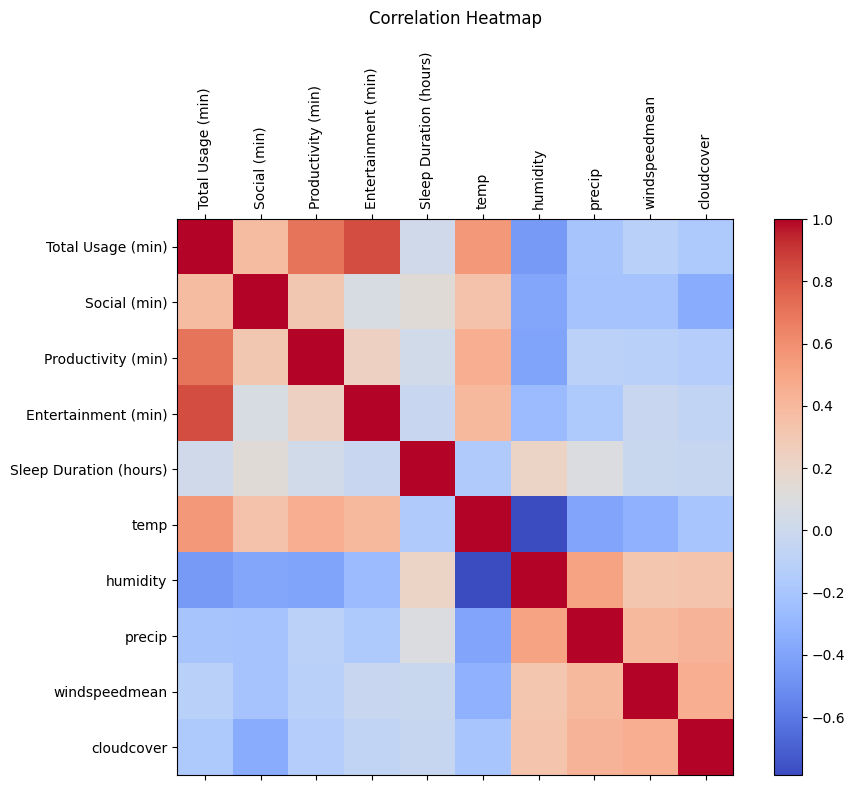




Productivity grouped by weather condition type shows subtle variations.



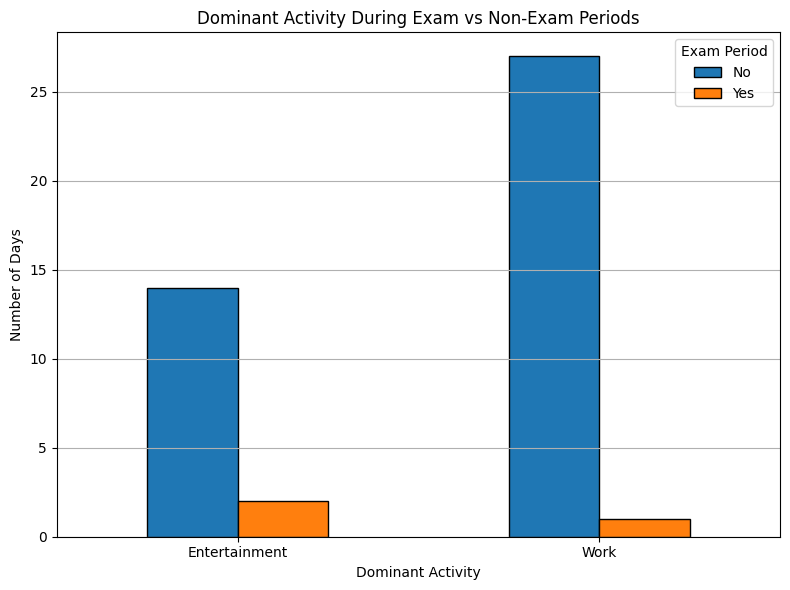
Correlation heatmap reveals interdependencies among features.



# Statistical Analysis

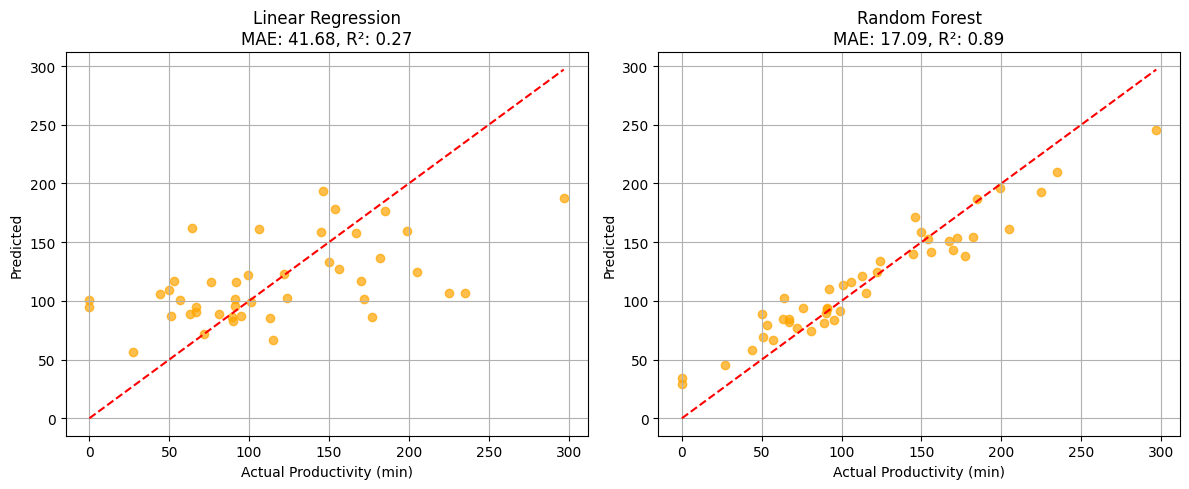
ANOVA: Weather vs Productivity  
F-statistic: 1.131, p-value: 0.3325  
Interpretation: No significant difference. H₀ not rejected.

Chi-Square Test: Exam Period vs Dominant Activity  
Chi-square statistic: 0.259, p-value: 0.6110  
Interpretation: No significant relationship. H₀ not rejected.



# Machine Learning Results

Objective: Predict daily productivity using weather, sleep, and exam indicators.  
  
Models: Linear Regression and Random Forest Regressor  
  
Results:  
Linear Regression – MAE: 41.7, R²: 0.27  
Random Forest – MAE: 17.1, R²: 0.89  
Interpretation: Random Forest outperformed Linear Regression, showing strong predictive ability.



# Limitations and Future Work

Limitations:  
- Small sample size (44 days)  
- Productivity based on screen usage only  
- Possible unmeasured variables  
  
Future Work:  
- Collect a larger dataset  
- Include academic load, mood tracking  
- Apply time-series modeling or deep learning

# Analysis & Visual Insights

The visualizations offer rich insights into user behavior patterns:

- The distribution plots indicate high variability in both productivity and entertainment time, with several days showing intensive screen usage and others nearly idle.  
- The boxplot of productivity across weather types suggests that although averages differ slightly, the spread and overlap are large, explaining the non-significant ANOVA result.  
- The correlation heatmap highlights strong relationships between some digital activity types, but weaker ones between weather metrics and productivity—pointing to more complex, possibly non-linear interactions.  
- The exam period bar chart shows that user focus does not systematically shift toward work during exams, challenging assumptions about academic pressure's effect on behavior.  
- Machine learning visual comparisons of predicted vs actual productivity reveal that the Random Forest model closely approximates actual behavior with far less error than the linear model.

# Conclusion

This study concludes that although weather conditions and exam periods do not have a statistically significant effect on productivity when analyzed independently, they hold predictive value when combined with sleep and contextual features. While simple statistical methods fell short, machine learning models—especially Random Forest—were effective at modeling productivity patterns.  
  
These findings reinforce the value of integrating behavioral, environmental, and academic data to understand and predict personal productivity. Future studies with larger and more diverse datasets may uncover additional hidden patterns and improve model generalization.