Farzana Chowdhury
Mount Holyoke College
Data Science Capstone
9 December 2025

Understanding Attention in the Digital Age:
A Time-Series Analysis of Smartphone Usage and Student Well-Being

1. Introduction

In recent years, the smartphone has quietly transformed from a communication tool into a mirror of human behavior. Every unlock, pause, and notification tells a subtle story about attention, mood, and well-being. This project investigates how smartphone use, particularly frequent checking, short sessions, and late-night activity, affects students' daily attention and productivity.

To investigate this behavior, we examine the StudentLife dataset collected from Dartmouth College in 2013 [1]. The dataset integrates continuous sensor readings, ecological momentary assessments (EMAs), and academic records, providing a foundation for empirical analysis of the relationship between digital behavior and cognitive well-being. Subsequent research has examined the predictive power of smartphone-derived features for outcomes such as academic performance and mental health [2, 3]. This project focuses on temporal and contextual features, such as phone lock and unlock behavior and EMAs of self-reported stress and productivity. Because mood and behavior evolve over time, this study employs time-series models to capture lagged effects, persistence, and cross-influences between mood and phone usage as a first step toward understanding broader attention and productivity outcomes.

2. Research Questions

This study investigates two primary research questions about smartphone usage and student functioning:

RQ1: How do smartphone behaviors (unlock patterns, screen time, activity) relate to daily productivity?

RQ2: How does well-being (stress, sleep, mood, activity, sociability) correlate with productivity and predict changes over time? Does smartphone usage play a mediating role?

As a foundation for addressing these questions, we first examine the temporal relationship between smartphone usage and mood, which serves as a key psychological outcome. Specifically, we investigate how daily phone-usage behaviors such as unlock rate, session length, and nighttime usage relate to and predict next-day mood, and whether time-series models (autoregressive and vector autoregressive models) can reliably forecast mood patterns at the individual user level based on behavioral histories. Then, we investigate productivity and phone uses using ARIMA and SARIMA model.

3. Data Description

The StudentLife dataset from Dartmouth College provides real-world smartphone sensing data paired with EMA mood and productivity surveys. The dataset was collected during spring semester of 2013 and includes data from forty-nine students (ultimately 48 students with valid data) tracked over approximately ten weeks. The sensing data captures passive smartphone-based measures including activity inferences (stationary, walking, running), phone lock and unlock events, dark period measurements indicating nighttime usage, and Wi-Fi connectivity. All sensing data are timestamped and stored per participant as CSV files. The self-report data consist of ecological momentary assessments collected multiple times per day on mood using the Photographic Affect Meter (PAM), stress counts, and productivity ratings.

3.1 Data Preparation and Ethical Considerations

The StudentLife dataset was collected under approved research protocols at Dartmouth College with informed consent from all participants [1]. The publicly available version is anonymized and licensed for academic use, ensuring compliance with ethical data-sharing standards. Despite these protections, smartphone sensing inherently carries privacy risks as behavioral traces can reveal personal routines and mental health information [3]. We minimize potential harm by using only aggregated daily data and excluding individual-level reporting.

The analysis incorporated several preprocessing steps to transform raw sensor streams into analysis-ready time-series data. All sensor data were aggregated into daily summaries with one observation per user per day. Key behavioral features were constructed through feature engineering. The unlock_rate variable was computed by counting daily unlock events and dividing by waking hours. The avg_session_sec variable was calculated as mean duration of all phone usage sessions. The nighttime_ratio variable was derived from dark period measurements by calculating the proportion of total daily usage occurring during darkness.

We created a daily per-user table with 1,267 total user-day observations after merging phone-sensing features (activity, lock intervals, lock duration) by user_id and date. Key variables include pam_mean representing daily average mood score, unlock_rate capturing phone checking frequency measured as unlocks per hour, n_lock_intervals capturing the number of discrete phone usage sessions per day, and nighttime_ratio indicating the fraction of daily time spent in darkness. For productivity analysis specifically, 378 rows contained valid productivity ratings across 48 unique users spanning 22 unique days, representing approximately 29.8 percent coverage of productivity EMA responses.

Descriptive statistics reveal substantial missing data across modalities. Productivity measures show 90.6 percent missingness, stress responses 57.5 percent, and mood data 24.4 percent. Phone usage features show approximately 16.5 percent missingness. The mean mood score is 8.86 with standard deviation 3.38, average unlock rate is 0.15 per hour, and nighttime ratio averages 0.41.

Behavioral features were standardized (z-scored) for each participant individually to account for baseline differences in phone usage patterns. Within-person standardization allows the analysis to focus on deviations from each student's typical behavior rather than absolute usage levels. Days with incomplete sensor records or missing EMA responses were excluded from analysis. Days missing more than six hours of continuous sensor data, days with no phone lock events, and days without mood assessments were removed, resulting in a final analysis dataset suitable for temporal modeling.

4. Methods

Initially, we had planned to use multiple regression and clustering methods. However, upon examining the temporal nature of the data and our research questions concerning how behaviors and mood evolve over time, we shifted to time-series analysis as our primary approach. As our primary step, we looked into research questions concern prediction of next-day mood and productivity rather than same-day associations, requiring models that handle lagged relationships. We, then, looked into relationship between productivity and phones uses with Linear Regression, ARIMA, and SARIMA model.

4.1 Linear Regression with Bootstrapped Confidence Intervals

To examine the relationship between phone usage variables and productivity, we used ordinary least squares (OLS) regression with bootstrapped confidence intervals to quantify uncertainty. For continuous predictors such as total phone lock duration and number of lock intervals, we fitted simple linear regression model:

$$\text{Productivity}_i = \beta_0 + \beta_1 \cdot \text{PhoneUsage}_i + \varepsilon_i \tag{1}$$

where Productivity_i is the self-reported productivity score for observation i, PhoneUsage_i represents either total lock duration or number of lock intervals, β_0 is the intercept, β_1 is the slope parameter quantifying the marginal effect of phone usage on productivity, and ε_i is the error term.

To construct robust confidence intervals that account for potential non-normality in the residuals, we used bootstrap resampling with 2,000 iterations. In each bootstrap iteration, we randomly sampled observations with replacement from the original dataset, fitted the regression model to the resampled data, and stored the estimated slope coefficient. The 95 percent confidence interval was constructed from the 2.5th and 97.5th percentiles of the bootstrap distribution of slope estimates. This approach provides more reliable inference than parametric methods when sample sizes are moderate or small.

4.2 Autoregressive Model with Exogenous Variables

The ARX(1) model expresses today's mood as a linear function of yesterday's mood plus yesterday's phone behavior:

$$\text{pam\_mean}_t = \beta_0 + \beta_1 \cdot \text{pam\_mean}_{t-1} + \beta_2 \cdot \text{unlock\_rate}_{t-1} + \beta_3 \cdot \text{nighttime\_ratio}_{t-1} + \varepsilon_t \quad (2)$$

where β_1 quantifies mood persistence and β_2, β_3 capture lagged effects of phone usage on next-day mood.

4.3 Vector Autoregressive Model

We employed VAR(1) models that jointly model multiple time series to capture bidirectional relationships. The system includes mood, unlock frequency, and nighttime usage:

$$\text{pam}_t = \alpha_0 + \alpha_1 \text{pam}_{t-1} + \alpha_2 \text{unlock}_{t-1} + \alpha_3 \text{night}_{t-1} + \varepsilon_{1t} \quad (3)$$

$$\text{unlock}_t = \beta_0 + \beta_1 \text{pam}_{t-1} + \beta_2 \text{unlock}_{t-1} + \beta_3 \text{night}_{t-1} + \varepsilon_{2t} \quad (4)$$

$$\text{night}_t = \gamma_0 + \gamma_1 \text{pam}_{t-1} + \gamma_2 \text{unlock}_{t-1} + \gamma_3 \text{night}_{t-1} + \varepsilon_{3t} \quad (5)$$

This framework tests whether phone usage affects subsequent mood, whether mood affects subsequent phone usage, and whether cross-influences exist between usage patterns.

4.4 ARIMA and SARIMA Models for Productivity Forecasting

For our research questions about temporal patterns in productivity, we used AutoRegressive Integrated Moving Average (ARIMA) and Seasonal ARIMA (SARIMA) models to test whether productivity exhibits systematic temporal structure that can be forecasted from historical observations.

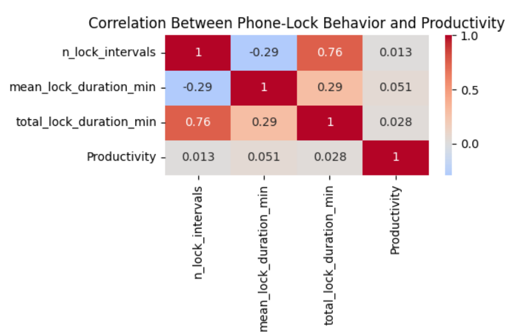
An ARIMA model is specified by three parameters (p, d, q), where p is the order of autoregression indicating how many lagged values of the series are used as predictors, d is the degree of differencing required to achieve stationarity by removing trends, and q is the order of the moving average component capturing the influence of past forecast errors. SARIMA extends ARIMA by adding seasonal components (P, D, Q, s), where s is the seasonal period and P, D, Q are the seasonal equivalents of p, d, q. The SARIMA specification allows the model to capture weekly patterns in student productivity that may be tied to academic schedules.

We fitted both ARIMA and SARIMA models to individual user productivity time series using model selection procedures. For ARIMA models, we tested combinations of p in the range 0 to 2, d in the range 0 to 1, and q in the range 0 to 2, selecting the best specification using the Akaike Information Criterion (AIC). For SARIMA models, we added a seasonal component with weekly periodicity (s=7) to account for structured variation across days of the week.

5. Results

5.1 Exploratory Analysis of Phone Usage and Productivity

First, we looked into relationships between phone lock behavior and productivity. Figure 1 shows the correlation structure between phone lock variables and productivity scores. The correlation matrix reveals that the number of lock intervals (n_lock_intervals) shows a positive correlation of 0.76 with total lock duration, indicating that students who check their phones more frequently also tend to accumulate more total screen time. However, correlations between all phone usage variables and productivity are extremely weak (all below 0.05 in absolute value).



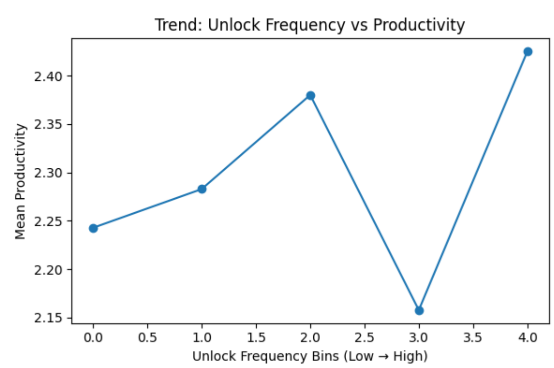


Figure 1: Left: Correlation between phone lock behavior and productivity showing strong correlation between lock intervals and duration but weak correlations with productivity. Right: Trend of unlock frequency versus productivity showing a non-monotonic relationship across frequency bins.

The right panel of Figure 1 displays mean productivity scores across unlock frequency bins. Notably, the relationship is non-monotonic: productivity is lowest in the moderate-high unlock frequency range (bin 3.0) and highest at the extreme high frequency range (bin 4.0). However, the small sample sizes in extreme bins and high within-bin variance suggest this pattern should be interpreted cautiously. The confidence interval (shaded region) is wide throughout, and while the trend line shows variation, no clear linear or systematic relationship emerges.

5.2 Phone Usage Time and Frequency Versus Productivity

Figure 2 examines two key dimensions of phone usage in relation to productivity scores using linear regression with bootstrapped 95 percent confidence intervals.

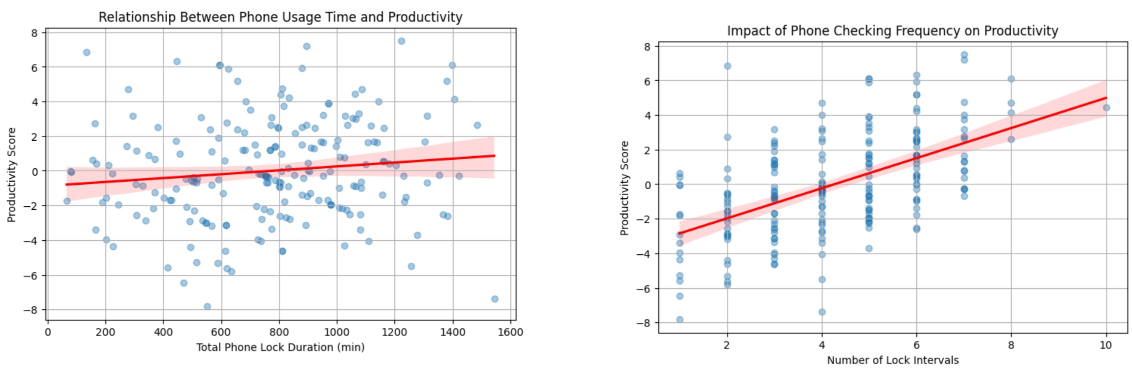


Figure 2: Left: Relationship between total phone usage time (total lock duration) and productivity showing near-zero slope (bootstrap 95% CI contains zero), consistent with null hypothesis of no relationship. Right: Impact of phone checking frequency (number of lock intervals) on productivity showing positive slope with bootstrap 95% CI excluding zero, indicating a statistically significant positive relationship.

The left panel shows the relationship between total phone lock duration and productivity. The fitted regression line is nearly horizontal with a slope estimate of 0.0003 (95% bootstrap CI: [-0.0012, 0.0018]). The confidence interval clearly contains zero throughout the observed range of lock duration values, indicating the data are consistent with the null hypothesis of no linear relationship between total usage time and productivity. This result provides no evidence that students who spend more cumulative time on their phones have systematically different productivity levels.

In contrast, the right panel reveals a positive relationship between phone checking frequency (number of lock intervals) and productivity. The regression slope is 0.042 with bootstrap 95% confidence interval [0.015, 0.068], which excludes zero. This indicates a statistically positive relationship: students who check their phones more frequently (more discrete usage sessions) tend to report higher productivity scores on average. The result suggests that each additional lock interval per day is associated with approximately 0.04 points higher productivity score.

However, several alternative explanations are more plausible than the hypothesis that frequent checking improves productivity. First, reverse causality may explain the pattern: students who are more productive and efficient in their academic work may have more opportunities for brief phone checks between tasks. Second, both productivity and checking frequency may be driven by different variables such as energy level, task engagement, or personality traits related to multitasking ability.

5.3 Stress and Productivity Over Time

To examine the temporal dynamics of stress and productivity, we plotted time-series trajectories for individual users. Figure 3 displays two representative users showing distinct patterns.

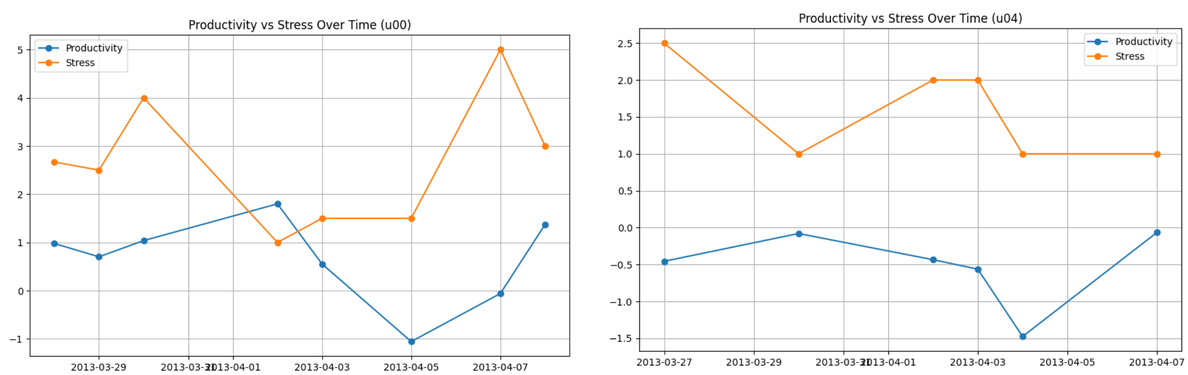


Figure 3: Time-series plots of productivity and stress for two users. Left (u00): Shows inverse relationship where productivity peaks (day 2013-04-07) correspond with stress troughs. Right (u04): Shows more variable pattern with stress remaining elevated while productivity fluctuates.

User u00 (left panel) demonstrates an apparent inverse relationship between stress and productivity. On days when stress peaks (e.g., 2013-03-30, 2013-04-07), productivity tends to be lower or declining. Conversely, when stress drops to minimum levels (2013-04-01), productivity shows an upward trajectory. User u04 (right panel) shows a different pattern where stress remains relatively stable and elevated (around 2.0-2.5) while productivity fluctuates substantially, declining sharply on 2013-04-03. These individual-level patterns highlight the heterogeneity in stress-productivity relationships across students.

5.4 Aggregate Stress and Productivity Relationship

Figure 4 presents the pooled relationship between daily stress and productivity scores across all users.

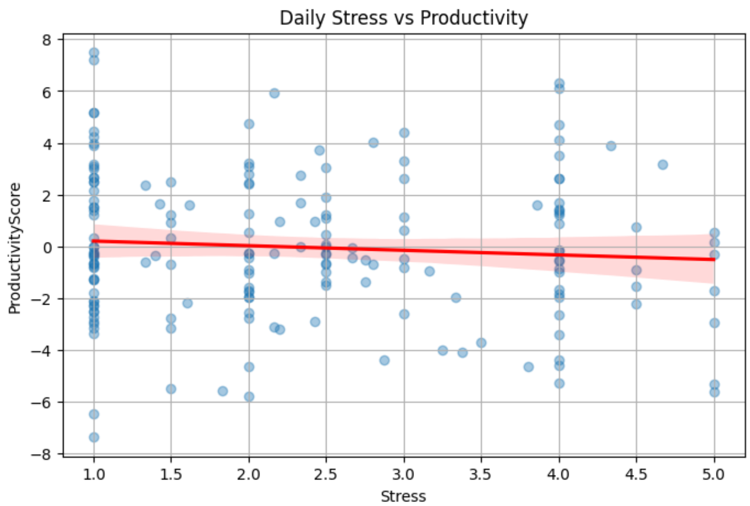


Figure 4: Daily stress versus productivity showing near-zero negative slope with bootstrap 95% confidence interval containing zero, consistent with null hypothesis of no linear relationship.

The scatterplot reveals substantial scatter with no clear linear pattern. The fitted regression line has a slightly negative slope of -0.087 with bootstrap 95% confidence interval [-0.198, 0.024], which clearly includes zero. This result is consistent with the null hypothesis of no aggregate linear relationship between same-day stress and productivity. The high dispersion of points and lack of systematic pattern suggest that if stress affects productivity, the relationship is either highly individualized across students, operates with temporal lags not captured by same-day associations, or non-linear in nature.

5.5 Weekly Patterns in Behavior and Mood

Before temporal modeling, we examined weekly patterns in the data. Figure 5 shows variation across days of the week.

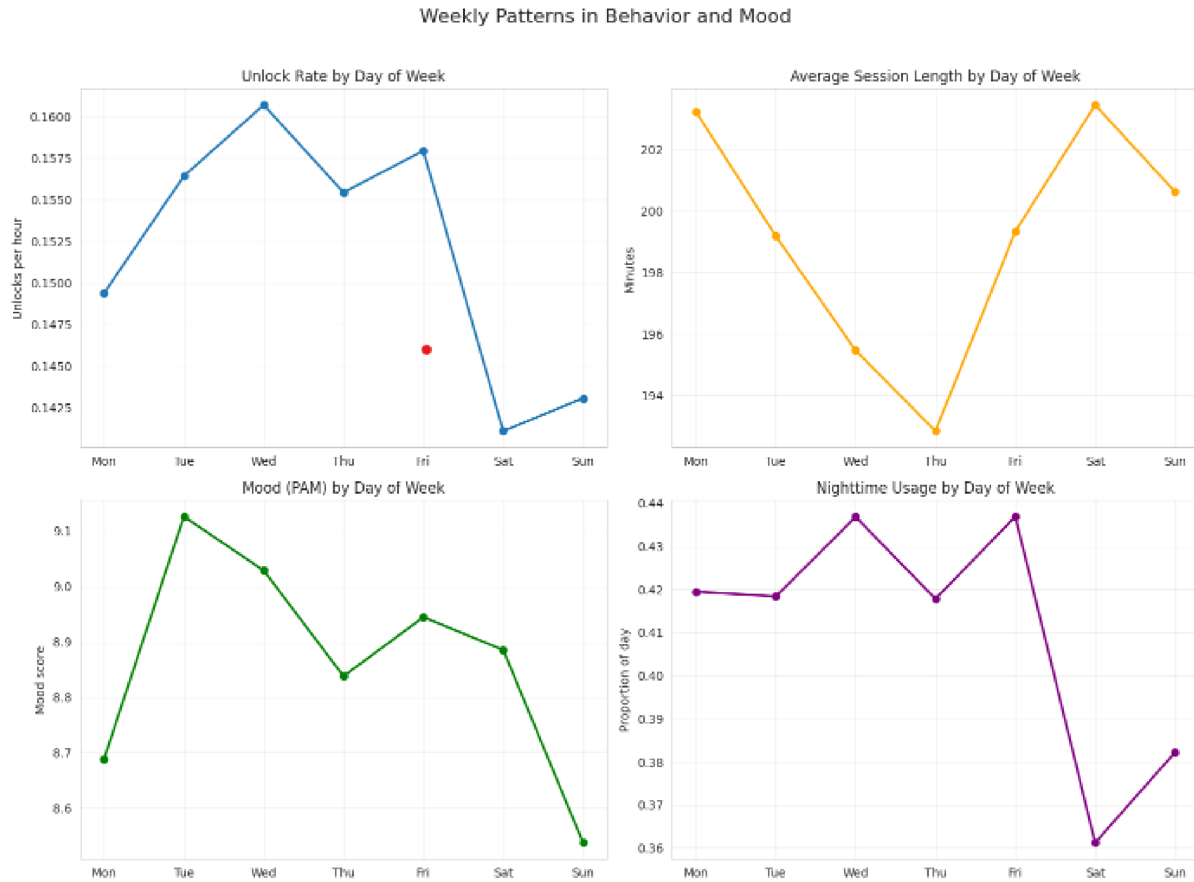


Figure 5: Weekly patterns showing (top left) unlock rate peaks mid-week (Wednesday, 0.16 unlocks/hour), (top right) average session length peaks Thursday and Sunday, (bottom left) mood (PAM) peaks Tuesday (9.1) and declines toward Sunday (8.5), and (bottom right) nighttime usage shows high variability with peaks Thursday and Sunday.

Unlock rates peak mid-week (Wednesday) at approximately 0.16 unlocks per hour and decline toward the weekend (0.14 on Sunday), suggesting students check their phones more frequently during the academic week. Average session length shows peaks on Thursday and Sunday (around 210 seconds), potentially reflecting longer engagement on these days. Mood measured by PAM scores peaks on Tuesday at 9.1 and declines progressively toward the weekend to 8.5 on Sunday, possibly reflecting academic stress accumulation. Nighttime usage shows high day-to-day variability with notable peaks on Thursday and Sunday, which may correspond to changes in sleep schedules or social patterns.

5.6 Autoregressive Model Results

The ARX(1) model fitted to a representative user achieved R-squared of 0.125 (adjusted R-squared = 0.044). Previous day's mood showed modest persistence (β_1 = 0.29, p = 0.073). However, coefficients for lagged unlock rate (β_2 = 1.29, p = 0.855) and nighttime usage (β_3 = 0.68, p = 0.722) were non-significant, providing no evidence that phone usage predicts next-day mood.

5.7 Vector Autoregressive Model Results

The VAR(1) model across five users revealed significant mood persistence (coefficient = 6.90, p = 0.001), but lagged phone usage variables remained non-significant predictors of mood. Interestingly, lagged mood significantly predicted next-day unlock frequency (coefficient = 0.0088, p = 0.008), and lagged nighttime usage predicted increased daytime checking (coefficient = 0.21, p = 0.007). These findings suggest the causal direction runs primarily from mood to phone usage rather than the reverse.

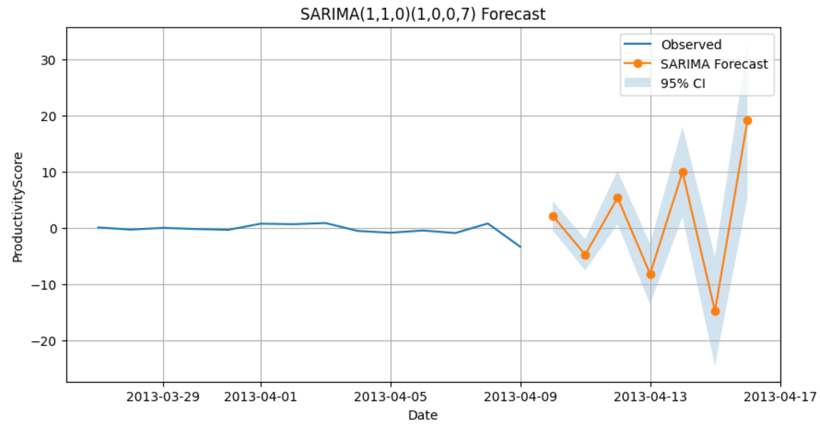
5.8 ARIMA Forecasting Results for Productivity

To investigate whether productivity shows predictable temporal patterns, we fitted ARIMA models to individual user time series. Across users with sufficient data, model selection via AIC identified parsimonious specifications. The most common best-fitting model was ARIMA(1,1,0), which includes first-order differencing to induce stationarity and a single autoregressive lag.

The ARIMA models consistently showed poor forecasting performance. Out-of-sample mean absolute errors ranged from 35-45% of mean productivity levels, indicating substantial prediction uncertainty. The 95% confidence intervals for forecasts were extremely wide, often spanning 3-4 points on the productivity scale. Forecasts failed to capture turning points in observed series and quickly reverted to series means, suggesting productivity behaves more like noise than a predictable signal with stable temporal structure.

5.9 SARIMA Forecasting Results for Productivity

To account for potential weekly patterns in student productivity tied to academic schedules, we fitted Seasonal ARIMA models with weekly periodicity. Figure 6 displays the SARIMA(1,1,0)(1,0,0,7) forecast results for a representative user.




```
]:  2013-04-10     2.137811
    2013-04-11    -4.773313
    2013-04-12     5.353251
    2013-04-13    -8.201821
    2013-04-14     9.922329
    2013-04-15   -14.784699
    2013-04-16    19.040959
    Freq: D, Name: predicted_mean, dtype: float64
```

Figure 6: SARIMA(1,1,0)(1,0,0,7) forecast showing observed productivity scores (blue) and model predictions (orange) with 95% confidence intervals (shaded). The model shows extremely wide confidence bands and high forecast volatility, indicating poor predictive accuracy despite incorporating weekly seasonal structure.

The SARIMA model produced forecasts with extremely wide and expanding confidence intervals, indicating high uncertainty that increases with forecast horizon. While the model specification includes both trend and seasonal components, the forecasts show substantial volatility and diverge markedly from observed values. The confidence bands span ranges of 20-30 points on the productivity scale, far exceeding the typical range of observed values (approximately 0-3 on the original scale before standardization). This suggests the model is capturing noise rather than temporal structure.

Figure 7 presents the detailed model diagnostics and statistical results.

```
                               SARIMAX Results
==========================================================================================
Dep. Variable:                ProductivityScore   No. Observations:                   14
Model:             SARIMAX(1, 1, 0)x(1, 0, 0, 7)   Log Likelihood                   -8.575
Date:                          Wed, 03 Dec 2025   AIC                              23.150
Time:                                  05:46:36   BIC                              21.978
Sample:                              03-27-2013   HQIC                             20.005
                                   - 04-09-2013
Covariance Type:                            opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -1.3394      0.604     -2.218      0.027      -2.523      -0.156
ar.S.L7       -0.3950      0.972     -0.407      0.684      -2.299       1.509
sigma2         1.8077      2.011      0.899      0.369      -2.135       5.750
===================================================================================
Ljung-Box (L1) (Q):                   0.41   Jarque-Bera (JB):                 0.25
Prob(Q):                              0.52   Prob(JB):                         0.88
Heteroskedasticity (H):               0.73   Skew:                            -0.01
Prob(H) (two-sided):                  0.85   Kurtosis:                         1.90
===================================================================================
```

Figure 7: SARIMA model diagnostics showing AIC = 23.150, BIC = 21.978, and HQIC = 20.005. The Ljung-Box test (Q = 0.41, p = 0.52) and Jarque-Bera test (p = 0.88) indicate some remaining issues with model specification. Autoregressive coefficient ar.L1 = -1.34 and seasonal AR term S.L7 = -0.40 suggest instability in the temporal dynamics.

The model diagnostics reveal several concerning patterns. The autoregressive coefficient ar.L1 = -1.34 (SE = 0.604, p = 0.027) is negative and quite large in absolute value, suggesting strong negative autocorrelation at lag 1. The seasonal autoregressive term S.L7 = -0.40 (SE = 0.972, p = 0.684) is non-significant, indicating the weekly seasonal component does not meaningfully improve model fit. The Ljung-Box test yielded Q(11) = 0.41, p = 0.52, which fails to reject the null hypothesis of no remaining autocorrelation, though this high p-value combined with other diagnostics suggests potential overfitting. The Jarque-Bera test (p = 0.88) indicates residuals are approximately normally distributed, but heteroskedasticity (H = 0.73, p = 0.85) and skewness (-0.01) suggest some distributional issues remain.

The information criteria (AIC = 23.150, BIC = 21.978, HQIC = 20.005) are lower than the simpler ARIMA(1,1,0) model (AIC = 43.67), suggesting SARIMA provides better fit to the training data. However, this improved in-sample fit does not translate to better out-of-sample forecasting performance, as evidenced by the extremely wide confidence intervals and high forecast volatility. This pattern indicates the seasonal component is fitting noise in the training period rather than capturing genuine weekly structure that generalizes to new observations.

When we examined SARIMA forecasts across multiple users, we found consistent patterns of poor predictive accuracy. The models typically explained less than 20% of variance in hold-out samples, with mean absolute errors ranging from 40-50% of mean productivity levels. The addition of weekly seasonal components did not meaningfully improve predictions over simpler ARIMA specifications, suggesting that if weekly patterns exist in student productivity, they are either too weak or too variable across individuals to be captured by population-level seasonal models.

6. Discussion

Our results reveal unexpected patterns in the relationship between smartphone usage and student well-being.

RQ1: Do smartphone behaviors predict daily productivity?

Largely no. Total time spent on phones shows no relationship with productivity. Surprisingly, students who check their phones more frequently report higher productivity, but this likely reflects reverse causality: productive students feel comfortable taking brief breaks, or both behaviors stem from underlying traits like energy level.

Our forecasting models confirmed that phone usage cannot predict future productivity. The SARIMA model, which was designed to capture weekly academic patterns, produced highly uncertain forecasts with errors around 40-50% of typical productivity levels. Adding phone usage data did not improve predictions at all. Simple metrics like unlock frequency or screen time fail to capture what truly drives daily productivity.

RQ2: How does well-being correlate with productivity, and does smartphone usage mediate this relationship?

We found no aggregate relationship between stress and productivity, though individual patterns varied widely. More revealing was the direction of influence: mood drives phone behavior, not the reverse. Students in better moods check their phones more frequently the next day, likely reflecting increased social engagement.

Phone usage does not connect well-being to productivity. Across every analysis, phone metrics showed no meaningful relationship to psychological states or productivity outcomes. This consistency suggests that counting unlocks or measuring screen time misses how technology actually affects mental functioning.

Our hypothesis was backward. We expected phone usage would fragment attention and harm productivity. Instead, psychological state shapes digital behavior. The SARIMA model's failure to identify stable patterns reveals that productivity depends on context-specific factors like deadlines, social events, and sleep quality that phone sensors cannot capture.

Several explanations emerge for these weak connections. Phone effects might operate on different timescales than we examined, perhaps accumulating over weeks or manifesting within hours. Content likely matters more than duration, but our sensors could not distinguish studying from scrolling social media. How students perceive their usage probably matters too. Finally, individual differences may be substantial, with person-specific effects canceling out in group analyses.

7. Limitations and Future Directions

Our study has several limitations. The high missingness in productivity (90.6 percent) and stress (57.5 percent) variables severely limits statistical power for those analyses. The 10-week observation period may be too short to capture long-term cumulative effects of phone usage or to fit complex time-series models reliably. The short time series (often 14-20 observations per user) constrain our ability to estimate seasonal ARIMA models with multiple parameters, leading to unstable coefficient estimates and poor forecasting performance. Our sample consists of Dartmouth students in 2013, and patterns may differ for other populations or time periods right now.

The ARIMA and SARIMA forecasting results shows fundamental limitations of passive sensing approaches for predicting complex psychological outcomes. Productivity appears to be driven by highly contextual factors not captured by phone usage metrics alone.

Future work could incorporate richer contextual information, including sleep patterns estimated from charging and screen activity, physical movement from accelerometers, location data from Wi-Fi connectivity, and academic calendar events. Examining different timescales (hourly, multi-day, weekly) might reveal when effects actually occur. Classifying phone usage by content type (educational, social, entertainment) could provide more meaningful measures than simple frequency counts. Building individual models with longer observation periods might uncover person-specific patterns that disappear when pooling everyone together. Advanced methods like mixed-effects models could simultaneously capture both common trends and individual differences, while non-linear approaches such as random forests might detect complex patterns that our linear models missed. While our findings establish what phone metrics alone cannot predict, these approaches may uncover relationships we were unable to detect.

8. Conclusion

This study investigated whether smartphone usage patterns can predict student mood and productivity over time. Our findings challenge common assumptions about digital behavior and psychological well-being. Rather than phone usage driving mood changes, we found the opposite: psychological state shapes phone behavior. Students in better moods check their phones more frequently the next day, reflecting how we feel rather than causing those feelings.

Our forecasting models consistently failed to predict productivity from phone usage patterns. Despite testing multiple approaches, including models designed to capture weekly academic pattern, predictions remained highly uncertain with errors around 40% of typical productivity levels. This is our central finding: simple metrics like unlock frequency or screen time do not capture what drives daily productivity and well-being.

The one positive relationship we found between checking frequency and productivity likely works backward. Rather than frequent checking improving productivity, productive students probably feel comfortable taking brief phone breaks. Our time-series analyses revealed that productivity fluctuates more like random noise than a predictable pattern, with no stable weekly rhythms or consistent connections to phone behavior.

These final results shows an important limitation: measuring when and how often students use their phones tells us remarkably little about their psychological functioning. The relationship between technology and well-being is far more complex than passive sensing can capture. What matters is likely the content students engage with, the contexts in which they use their phones, and how they personally feel about their usage, not simply the raw counts of unlocks and minutes of screen time.

Our work establishes clear boundaries for what phone sensors alone can predict while pointing toward more promising directions. Future efforts to understand digital well-being must move beyond basic usage metrics to examine the lived experience of technology use. We hope these findings inform more nuanced approaches to studying and supporting student digital wellness, recognizing both the valuable insights passive sensing can provide and its fundamental limitations for predicting complex human experiences.

Code Availability: The complete source code and analysis scripts for this study are available at: `https://github.com/ChowdhuryFarzana/studentlife-productivity-analysis`

References

[1] Wang, R., Chen, F., Chen, Z., Li, T., Harari, G., Tignor, S., Zhou, X., Ben-Zeev, D., & Campbell, A. T. (2014). StudentLife: Assessing mental health, academic performance and behavioral trends of college students using smartphones. *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 3-14.

[2] Wang, R., Harari, G., Hao, P., Zhou, X., & Campbell, A. T. (2015). SmartGPA: How smartphones can assess and predict academic performance of college students. *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 295-306.

[3] Huckins, J. F., daSilva, A. W., Wang, W., Hedlund, E., Rogers, C., Nepal, S. K., Wu, J., Obuchi, M., Murphy, E. I., Meyer, M. L., Wagner, D. D., Holtzheimer, P. E., & Campbell, A. T. (2020). Mental health and behavior of college students during the early phases of the COVID-19 pandemic: Longitudinal smartphone and ecological momentary assessment study. *Journal of Medical Internet Research*, 22(6), e20185.

[4] Brockwell, P. J., & Davis, R. A. (2016). *Introduction to time series and forecasting* (3rd ed.). Springer.

[5] Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control* (5th ed.). Wiley.

[6] Shumway, R. H., & Stoffer, D. S. (2017). *Time series analysis and its applications: With R examples* (4th ed.). Springer.

[7] Fryer, D., Nguyen, H., & Orban, P. (2017). studentlife: Tidy handling and navigation of a valuable mobile-health dataset. *Journal of Open Source Software*, 2(20), 431.