Patterns of Practice: Modeling Well-Being Change and Engagement Over Time

Sadaf Nasir

University of Wisconsin - Madison

Psychology: Data Science in Human Behavior

December 16th, 2025

Introduction

Digital mental health applications have rapidly expanded in scope and adoption, offering scalable tools for stress reduction, mindfulness training, and psychological well-being (Goldberg et al., 2020). Among these tools, mindfulness-based digital interventions have gained particular attention due to their adaptability, low marginal cost, and alignment with evidence-based contemplative practices (Goldberg et al., 2020). At Healthy Minds Innovations (HMI), scientists have honed in on these potential benefits and developed a science-based mental health application that is designed to support well-being through guided digital meditation and mindfulness practices/activities (Goldberg et al., 2020).

HMI is a nonprofit organization associated with the Center for Healthy Minds and University of Wisconsin - Madison, and is driven by the mission to translate science into tools to cultivate well-being (Healthy Minds Innovations, n.d., 2025). The Healthy Minds Program application is a science-based and widely accessible tool geared towards cultivating well-being (Healthy Minds Innovations, n.d., 2025). The guided meditation and mindfulness activities within the Healthy Minds Program app are designed to enhance four pillars of well-being, Awareness, Connection, Insight, & Purpose (Goldeberg et al., 2020). These four pillars make up the Healthy Minds Index, which is a validated, 17-item, self-report measure of well-being, based on a theory that each pillar is trainable and improvable (Kral et al., 2024).

Within the app, there are four separate modules, (“Awareness”, “Connection”, “Insight”, & “Purpose”), making up a recommended path for users to follow in the app. Each module contains activities pertaining to cultivating that respective pillar, as well as a fifth “Explore” module, which allows users to customize their app usage however they wish.

Through exploring and analyzing user data from the Healthy Minds app, HMI can further improve the practices and interventions contained in the app, thus providing a highly effective, accessible tool for the average individual to improve well-being. In accordance with the concepts and skills taught in the Data Science in Human Behavior (DSHB) M.S. program at the University of Wisconsin - Madison, this project aims to uncover valuable insights into app usage patterns and benefits.

Previous analyses conducted by students of the DSHB program focused heavily on translating app engagement data into insights on well-being improvements (Simental, 2024). Little research was conducted translating well-being patterns into engagement trends. Exploring which factors promote or hinder engagement with the Healthy Minds app could help mold changes in the future geared towards increasing user engagement.

Additionally, this project was motivated by explorations in the individual well-being pillars (ACIP). Previous analyses with Healthy Minds data evaluated an exploratory factor analysis with Healthy Minds Index data, uncovering that all items in the Healthy Minds Index load onto four separate factors corresponding to the pillars being measured (Awareness, Connection, Insight, & Purpose).

This led to the guiding research question for this project: Do changes in well-being and stress have measurable effects on engagement? Additionally, the question of individual pillars and their predictive strength in engagement was explored. Two main hypotheses were proposed in this analysis. First it was hypothesized that changes in ACIP pillars will differ in their predictive strength, such that changes in some pillars will be stronger predictors of engagement than others. Second, it was hypothesized that the relationship between changes in well-being and engagement will vary as a function of stress levels, and that this interaction effect between changes in well-being and stress will remain statistically significant with the addition of demographic variables in the statistical model.

Methods

Data

Data collection for this analysis is rather complex. As data collection relies heavily on real-world application usage, the structure of data files is multidimensional. Data are collected longitudinally, across three time points, (T0, T1, & T2). Collection begins as users download the application, at which point they are prompted to complete a demographic survey, a baseline (T0) well-being survey (measured through validated Healthy Minds Index), as well as a baseline perceived stress survey (measured through validated Perceived Stress Scale). All surveys are optional, and users are free to skip straight to the meditation/mindfulness activities if they choose to do so. As a result, this project tackles paramount missing data.

Approximately one month after downloading the Healthy Minds Program app, users are prompted to complete the well-being Healthy Minds Index survey for a second time (T1), as well as another perceived stress survey. Finally, users are prompted to do this for a third (T2) time approximately two months after downloading the Healthy Minds Program app. To reiterate, survey completion is entirely optional.

Between T0 and T2 timepoints, all app usage data is collected and saved. All activities attempted and completed were made available for analysis. All data were stored in six separate data files: demographics, stress, user activity, T0, T1, and T2, which can be seen in Table 1 below.

Table 1: Data Sources Used in the HMI Capstone Project

Dataset	Source	Unit	Content	Role in Analysis
Demographics	SQLite database	User	Age group, gender, race, education, relationship status, prior meditation	Baseline covariates and sample characterization
User Engagement History	CSV application logs	Event (aggregated to user)	Session timestamps, activity type, duration	Derivation of engagement metrics (sessions, active days, duration days)
Well-Being Survey (T0)	SQLite database	User	Awareness, connection, insight, purpose; composite HMI	Baseline well-being assessment
Well-Being Survey (T1)	SQLite database	User	Awareness, connection, insight, purpose; composite HMI	One-month longitudinal follow-up
Well-Being Survey (T2)	SQLite database	User	Awareness, connection, insight, purpose; composite HMI	Two-month longitudinal follow-up
Perceived Stress Survey	CSV survey export	User	Individual stress item scores	Derivation of mean perceived psychological stress

Demographic Data

Demographic data were collected at the baseline (T0) timepoint if users wished to complete the demographic survey. Data collected pertained to age, gender, race, highest education completed, relationship status, reasons for using the Healthy Minds Program app, and meditation habits prior to downloading the app. The demographic dataset contained 174,097 observations, with one row per user.

User Engagement History

The user engagement dataset contains event-level records capturing individual interactions with the well-being application. Each record includes a unique user identifier, activity identifier, and timestamps indicating when an activity was completed, allowing engagement to be temporally ordered at the user level. Additional variables describe task classification fields including task type, application module, and activity series. Content-specific characteristics are also recorded, including activity name, practice posture, practice speaker, and practice duration, providing contextual information about the nature of each engagement. An elapsed time variable reflects the duration of user interaction with each activity, while activity status indicates completion outcomes. For analytic purposes, completion timestamps were converted from epoch time to calendar dates, enabling aggregation of engagement events into user-level metrics such as total sessions, active days, and duration of use. The user engagement history dataset contains 5,273,258 activity records from 217,303 users.

Well-Being Surveys (T0, T1, & T2)

Well-being surveys represent three of the six total datasets used in this analysis. These are longitudinal measures of well-being collected at three time points, each one month apart (T0, T1, & T2). Since well-being is measured through the four ACIP pillars, each of these datasets contains 4 columns for well-being, measuring Awareness, Connection, Insight, & Purpose. A date column is also included in each dataset indicating when these surveys were completed. A general trend of reduced observations is noted in the progression from T0 to T2, with 279,625 observations collected at T0, 53,667 observations collected at T1, and 28,430 observations collected at T2.

Perceived Stress Data

The final dataset utilized in this analysis contained data measuring perceived stress in users of the Healthy Minds Program app. The perceived stress scale is a validated self-report stress measure that users were prompted to take upon completion of the well-being surveys at each time point. The survey contains 10 items relating to stressful events and ability to manage stressful situations, as well as a date column, indicating when the survey was completed. This dataset contains 181,356 observations from 176,875 users.

Derivation of Key Features

The data are not inherently ready for analyses, so basic manipulations were performed to have functional features for regression analysis.

Quantifying Engagement

Quantifying engagement from the user history dataset was a monumental task. Previous analyses on this topic involved creating a ratio of the number of active days of using the Healthy Minds app over the total duration of days using the app. This created a very tight distribution of most of the sample hovering very close to 0. To simplify the analysis, engagement was quantified exclusively as the total number of sessions completed per user. However, the number of active days and duration days were still retained. After grouping the user history dataset by user with the number of sessions completed, the dataset was significantly shortened to 151,968 observations, with only one row per user. This inherently switched the dataset from long format into wide format, resulting in all datasets being in wide format.

Stress Measure

The perceived stress data were provided in a .csv file and contained 10 individual stress items as well as a completion date column. These columns were renamed and shortened to make them workable (e.g. "In_the_past_month_how_often_have_you_found_that_you_could_not_cope_with_all_the_things_you_had_to_do" was shortened to "unable_to_cope"). Responses from the stress items were on a 5-point Likert scale, ranging from "never" to "very often". These were converted into numeric responses ranging from 0-4, with reverse-coded items converted accordingly.

The perceived stress dataset came with a few issues. First, there were some exact duplicate rows, where the user_id, completion date, and individual stress responses were identical. The duplicates of these rows were removed. The second problem came with the temporal order of responses. While the well-being surveys were separated out by time point, the stress surveys were not. As a result, many users had completed multiple surveys on different dates, but all survey responses were contained in one dataset. Unfortunately, the stress surveys and the well-being surveys were not completed on the same dates (either due to app design or user choice), and the time points could not be aligned with the stress surveys. To retain as much data as possible, users who completed multiple stress surveys had all surveys averaged together. Ultimately each user's 10 responses to the stress survey were averaged together as well, resulting in only one stress score per user.

Merging of Datasets

To continue deriving features, the datasets had to be merged together. This allowed T0, T1, and T2 data to be called simultaneously. To merge, a full_join was used. While this created an overall dataset much larger than what ended up in the final models, a full-join was still the preferred joining method, as data retention was the goal. The merged dataset was written as a new .csv file and loaded into another notebook for further analysis.

Difference Scores

The aim of this project was to analyze the effect of *change in well-being* on engagement, and as such, difference scores had to be calculated for well-being across different time points. With four measures of well-being at each of the three time points, this resulted in the following 12 calculations for difference scores (4 ACIP difference scores for T1-T0, 4 ACIP difference scores for T2-T1, and 4 ACIP difference scores for T2-T0).

Table 2: Well-Being Pillar Change Scores

Subscale	Baseline to T1 (Δ_{1-0})	T1 to T2 (Δ_{2-1})	Baseline to T2 (Δ_{2-0})
Awareness	$\text{Awareness}_1 - \text{Awareness}_0$	$\text{Awareness}_2 - \text{Awareness}_1$	$\text{Awareness}_2 - \text{Awareness}_0$
Connection	$\text{Connection}_1 - \text{Connection}_0$	$\text{Connection}_2 - \text{Connection}_1$	$\text{Connection}_2 - \text{Connection}_0$
Insight	$\text{Insight}_1 - \text{Insight}_0$	$\text{Insight}_2 - \text{Insight}_1$	$\text{Insight}_2 - \text{Insight}_0$
Purpose	$\text{Purpose}_1 - \text{Purpose}_0$	$\text{Purpose}_2 - \text{Purpose}_1$	$\text{Purpose}_2 - \text{Purpose}_0$

For models containing average measures of well-being, difference scores were calculated for all three model comparisons (T1-T0, T2-T1, & T2-T0). These calculations can be seen below.

Table 3: Average Well-Being (HMI) Change Scores

Change Interval	Definition
Baseline to T1 (Δ_{1-0})	$\text{Avg HMI}_1 - \text{Avg HMI}_0$
T1 to T2 (Δ_{2-1})	$\text{Avg HMI}_2 - \text{Avg HMI}_1$
Baseline to T2 (Δ_{2-0})	$\text{Avg HMI}_2 - \text{Avg HMI}_0$

Preparation for Modeling

Numeric variables were mean-centered to prepare for regression analysis. Categorical variables were converted to factor variables to make them compatible with linear regression in R.

Exploratory Data Analysis

Exploratory data analysis (EDA) was conducted to characterize the structure, completeness, and distributions of the analytic dataset prior to statistical modeling. The final wide-format dataset contained 334,492 unique users and 35 variables spanning demographics, engagement metrics, perceived stress, and longitudinal well-being outcomes. Variables included a mixture of categorical demographic fields, date variables capturing survey and engagement timing, and continuous measures reflecting engagement intensity and well-being scores.

Missingness and Data Completeness

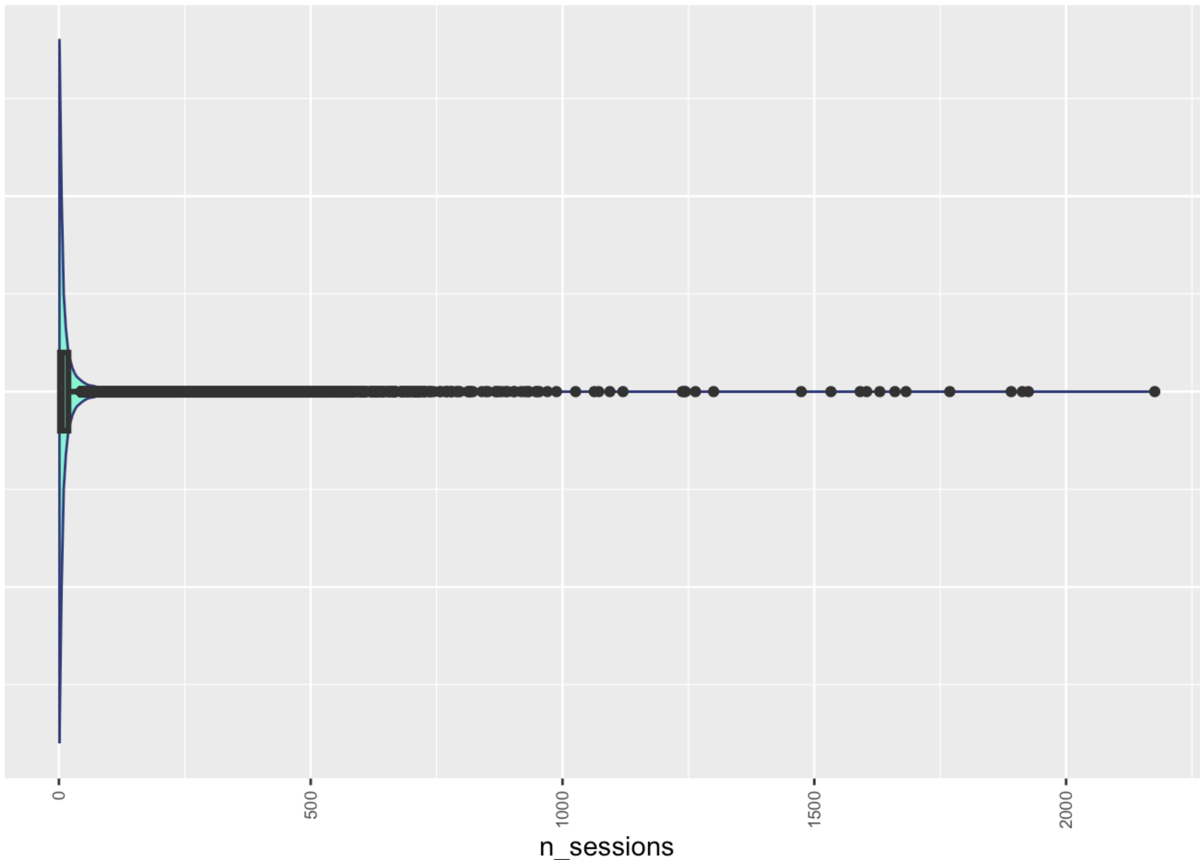
EDA revealed substantial and systematic missingness across multiple domains. Demographic variables exhibited moderate to high levels of missing data, with approximately half of users missing information on age, gender, race, education, and relationship status. This pattern is consistent with optional surveys common in consumer-facing digital health applications. In contrast, baseline well-being scores (T0) demonstrated relatively high completeness, with approximately 83–84% of users providing valid responses across the composite well-being/HMI score and its four subscales. Follow-up well-being assessments showed markedly lower completion rates, with only 16% of users completing T1 surveys and approximately 8% completing T2 surveys, resulting in an unbalanced longitudinal structure.

Engagement metrics also exhibited missingness, as these variables were only defined for users who engaged with the application at least once. Approximately 45% of users had valid session-level engagement summaries, reflecting a large proportion of users who did not proceed to active use. Perceived stress data were available for approximately 53% of users, typically based on a single completed stress survey.

Distribution of Engagement Metrics

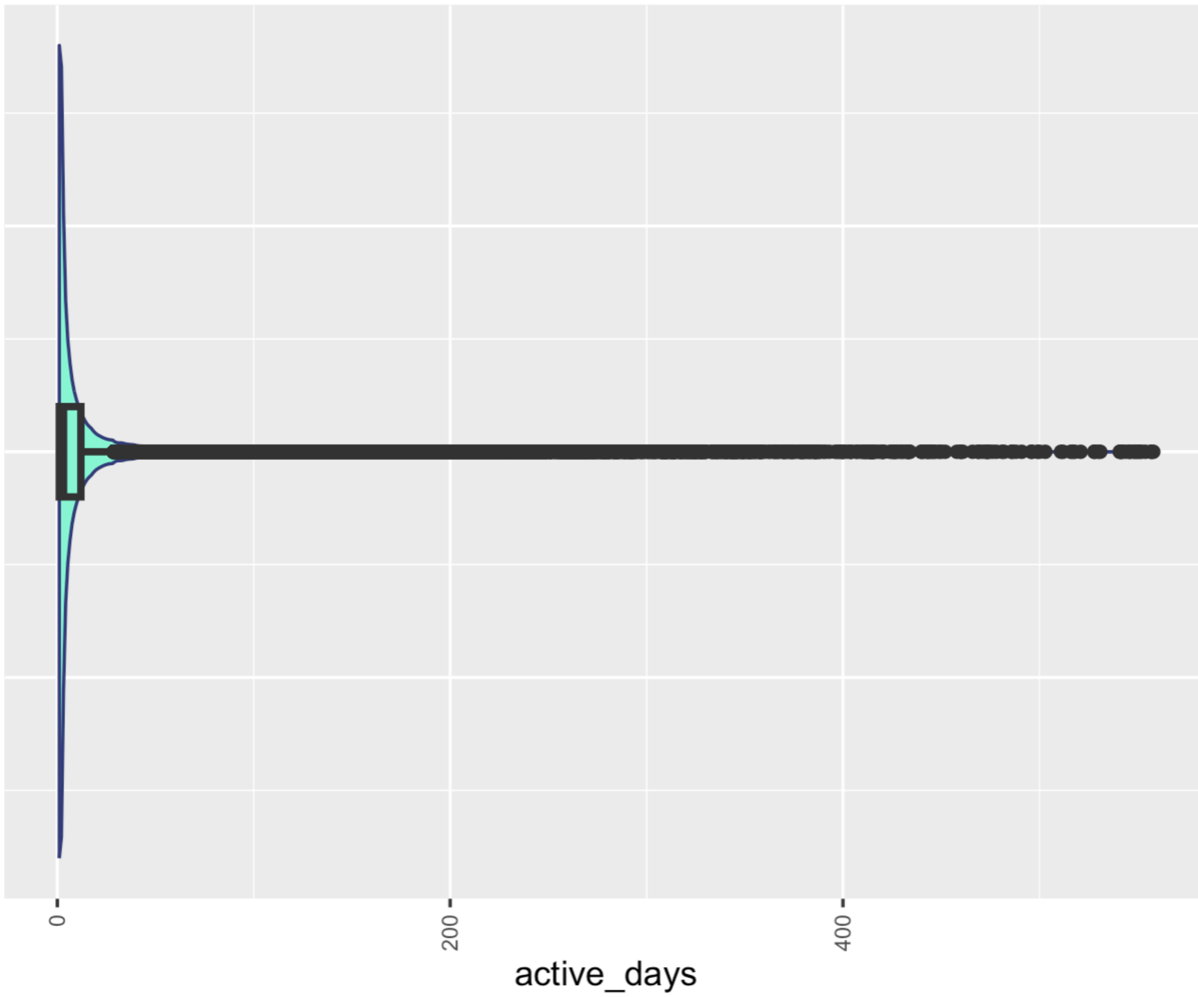
User engagement variables displayed pronounced right-skewed distributions. The number of completed sessions ranged from 1 to over 2,000, with a median of 6 sessions and an interquartile range of 2 to 19 sessions. A box-violin plot showing the univariate distribution of n_sessions can be seen below in Figure 1.

Figure 1: Univariate Distribution of Total Number of Sessions Completed



Similarly, the number of active days showed a median of 3 days, while the maximum exceeded 500 days, indicating a small subset of highly persistent users. A distribution of active_days can be seen below in Figure 2. Duration of use, defined as the number of days between first and last recorded sessions, exhibited extreme skewness, with a median of 15 days and a maximum exceeding 7,500 days. These distributions suggest substantial heterogeneity in engagement behavior, with most users engaging briefly and a minority demonstrating sustained long-term use.

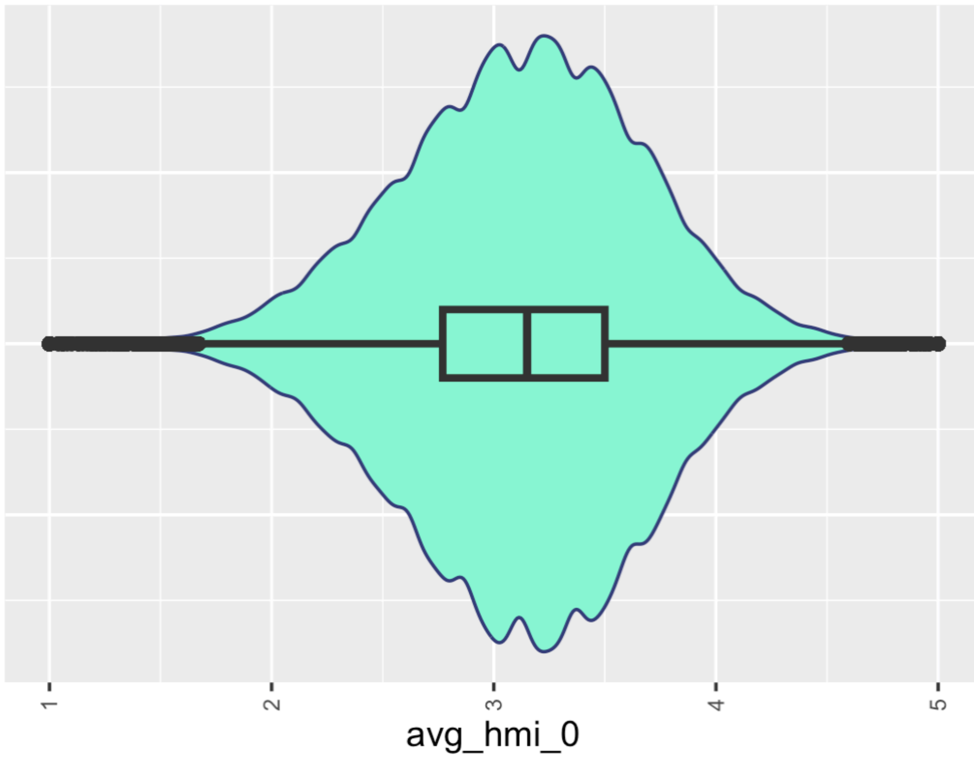
Figure 2: Univariate Distribution of Active Days of Use



Well-Being Scores at Baseline and Follow-Up

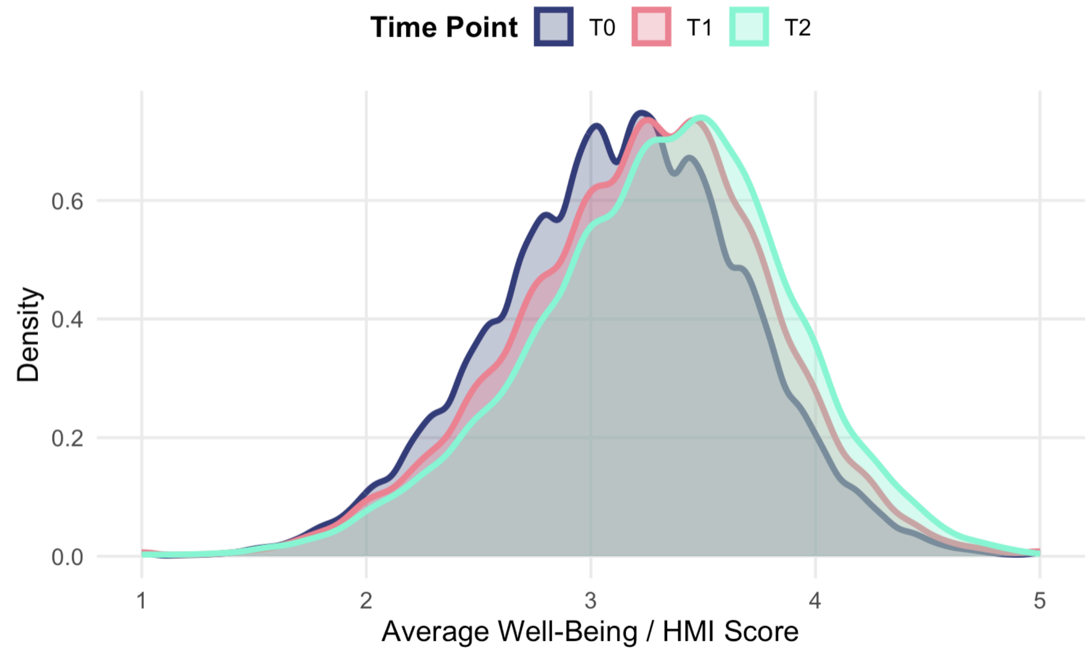
Baseline composite well-being scores (Avg HMI_0) were approximately normally distributed with slight positive skew, centered around a mean of 3.13 (SD = 0.55) on a 1–5 scale (shown in Figure 3). Subscale distributions showed similar patterns, with connection scores tending to be higher on average than insight scores, which exhibited greater dispersion. Purpose scores demonstrated the widest variability among subscales at baseline.

Figure 3: Univariate Distribution of Avg HMI_0



At follow-up time points, mean composite well-being scores increased modestly, with avg_hmi_1 averaging 3.22 and avg_hmi_2 averaging 3.31 among users who completed these assessments (seen below in Figure 4). Subscale distributions at T1 and T2 retained similar shapes to baseline, suggesting stability in scale properties across time. However, due to selective attrition, these summaries likely reflect a subset of users with higher engagement and continued participation.

Figure 4: Distribution of Average Well-Being/HMI Scores Across T0, T1, &T2



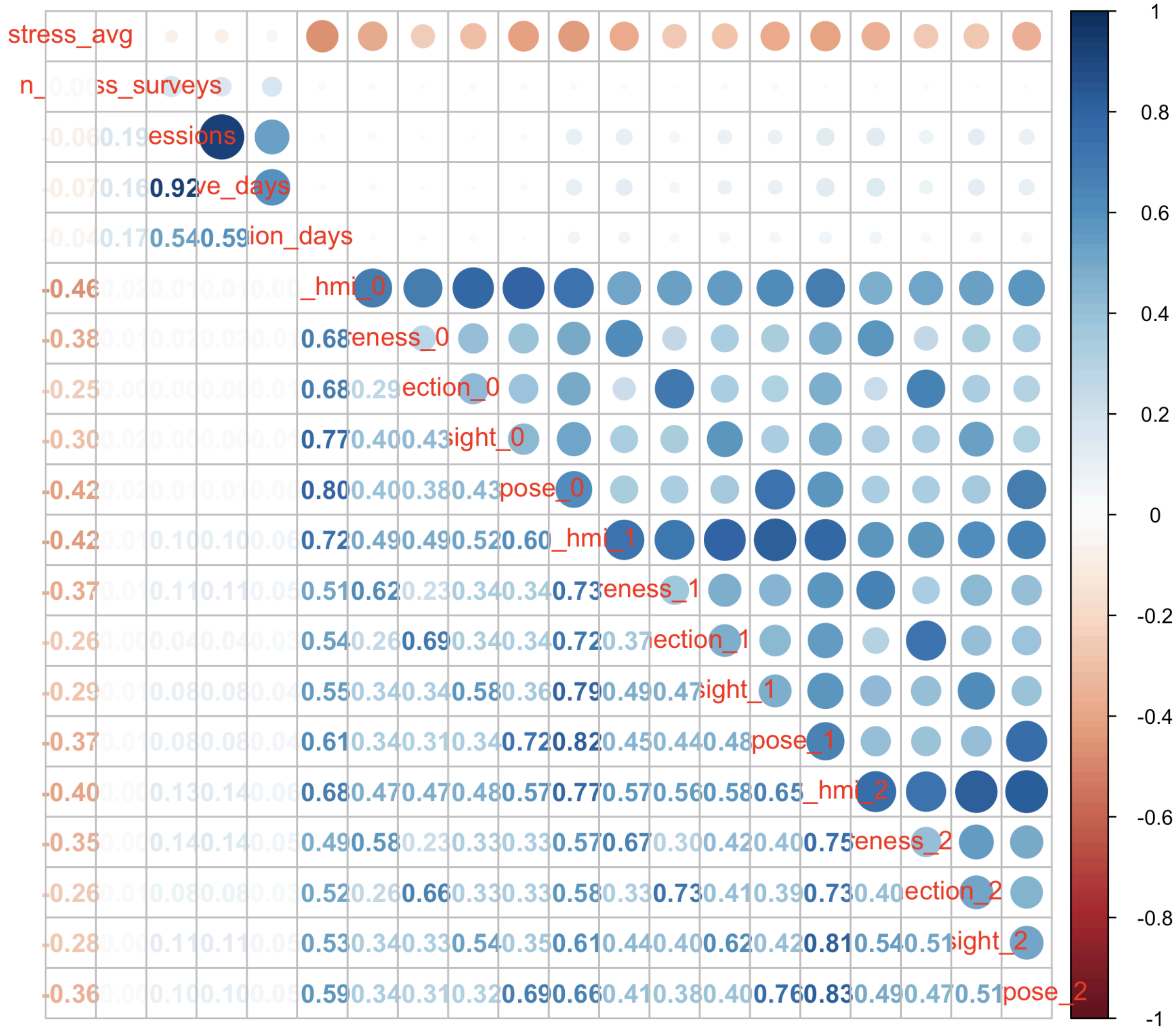
Perceived Stress Measures

Perceived stress scores were moderately normally distributed, with a mean of 2.19 (SD = 0.69) on a bounded scale ranging from 0 to 4. Most users completed only a single stress survey, as reflected by a median and mode of one completed survey. The limited repeated measurement of stress constrained the use of longitudinal stress modeling but allowed for its inclusion as a cross-sectional covariate in regression analyses.

Correlation Structure and Variable Relationships

Correlation analyses conducted during EDA indicated moderate positive correlations among well-being subscales at each time point, supporting the use of a composite HMI score. Engagement metrics were weakly to moderately correlated with one another, suggesting that session count, active days, and duration capture related but nonredundant dimensions of use. No evidence of severe multicollinearity was observed among candidate predictors, supporting their joint inclusion in multivariable models. The correlation matrix can be seen below in Figure 5.

Figure 5: Correlation Matrix of all Numeric Variables



Implications for Modeling

Findings from EDA informed several analytic decisions. The heavy right skew of engagement variables motivated cautious interpretation of linear model assumptions. The unbalanced longitudinal structure and substantial missingness at follow-up supported the use of complete-case analyses rather than imputation. Overall, EDA confirmed that while the dataset is large and information-rich, its structure reflects real-world usage patterns that necessitate conservative and transparent analytic approaches.

Results

Descriptive Statistics

The analytic dataset contained 334,492 unique users and 35 variables, including demographics (e.g., education, age category, gender, race), engagement metrics derived from app use (e.g., number of sessions, active days, duration in days), repeated timepoint well-being measures (T0, T1, T2), and a perceived stress summary score (stress_avg).

Missingness and completeness

Completeness varied substantially across variable families. Demographic fields were frequently missing (e.g., education complete rate = 0.48; age = 0.52; gender = 0.52; race = 0.48), while the baseline Well-Being/HMI composite (avg_hmi_0) was comparatively more complete (complete rate = 0.83).

Engagement metrics were also incomplete (complete rate = 0.45 for n_sessions, active_days, and duration_days), consistent with the fact that some users completed survey components without recorded usage in the engagement window (or vice versa).

The perceived stress average (stress_avg) had a complete rate of 0.53. A dataset overview and missingness summary can be seen below in Table 4.

Table 4: Analytic dataset overview and missingness summary (N = 334,492 users)

Variable group / example fields	Example variables (as named in dataset)	Missing n	Complete rate
Demographics	education; age; gender; race	education: 174,251; age: 160,440; gender: 161,052; race: 172,270	education: 0.48; age: 0.52; gender: 0.52; race: 0.48

Perceived stress	stress_avg	157,617	0.53
Engagement	n_sessions; active_days; duration_days	182,524 (each)	0.45 (each)
Well-Being/HMI (baseline)	avg_hmi_0	58,303	0.83

Note. “Missing n” and “complete rate” are reported directly from the project’s skim summaries of the analytic dataset.

Distributional features of key continuous measures

Among users with observed values, perceived stress scores were centered near the midrange (M = 2.19, SD = 0.69; median = 2.20).

Engagement variables were strongly right-skewed, with a median of 6 sessions but a maximum of 2,176 sessions, indicating a small subset of very high-use users (M = 23.81, SD = 55.69).

Similarly, active_days had a median of 3 days with a maximum of 558 days (M = 15.16, SD = 34.11), again consistent with a long right tail.

Baseline well-being scores (avg_hmi_0) were concentrated in the upper midrange (M = 3.13, SD = 0.55; median = 3.15). Descriptive Statistics can be seen below in Table 5.

Table 5: Descriptive statistics for key continuous variables

Variable	Mean	SD	Min	25th	Median	75th	Max
stress_avg	2.19	0.69	0	1.70	2.20	2.70	4

n_sessions	23.81	55.69	1	2.00	6.00	19.00	2176
active_days	15.16	34.11	1	1.00	3.00	12.00	558
avg_hmi_0	3.13	0.55	1	2.77	3.15	3.50	5

Note. Percentiles (25th/median/75th) correspond to p25/p50/p75 from the skim summaries.

Overview of Model Sets

Nine linear regression models were fit with number of sessions completed (n_sessions) as the outcome. Models 1–3 were additive linear models evaluating whether domain-level well-being change scores (Awareness, Connection, Insight, & Purpose) over each interval predicted engagement, adjusting for average perceived stress (stress_avg_c). Models 4–6 evaluated whether overall average well-being change (avg_hmi_*) predicted engagement and whether this association differed by stress via an interaction term. Models 7–9 repeated Models 4–6 while additionally adjusting for age, gender, and race.

Across models, stress was consistently associated with fewer sessions (negative coefficients), while improvement in well-being change scores tended to be associated with more sessions (positive coefficients). Model fit was modest overall (R^2 ranging from ~.013 to ~.044).

Models 1–3: Domain-level Well-Being change predicting engagement (plus stress)

Model 1 (T1–T0 domain changes + stress). The baseline domain-change model explained 2.24% of variance in sessions ($R^2 = .022$). All parameter estimates were statistically significant at $p < .001$ for this model, with adjusted $R^2 = .022$.

Model 2 (T2–T1 domain changes + stress). This model explained 1.36% of variance ($R^2 = .0136$). Awareness, Connection, and Insight change scores were statistically significant positive predictors, while Purpose change was not significant. Stress was a strong negative predictor ($b = -14.14$, $p < .001$).

Model 3 (T2–T0 domain changes + stress). This model showed the strongest fit among the domain-change set ($R^2 \approx .0302$), suggesting that longer-interval change scores captured more variance in engagement than adjacent-interval changes.

Interpretively, across the domain-level models, stress was consistently associated with lower engagement, and positive well-being changes, especially Connection-related change, tended to align with higher engagement, with the clearest pattern in the T2–T0 window.

Models 4–6: Average Well-Being change × stress interactions predicting engagement

These models tested whether the relationship between average well-being change and sessions differed by stress level.

Model 4 (avg_hmi_1_0 × stress). Average well-being change from T1–T0 was positively associated with sessions ($b = 25.19$, $p < .001$), stress was negatively associated ($b = -10.45$, $p < .001$), and the interaction was statistically significant and negative ($b = -6.33$, $p < .001$).

Model 5 (avg_hmi_2_1 × stress). Average well-being change from T2–T1 was positively associated with sessions ($b = 18.09$, $p < .001$) and stress was negative ($b = -14.17$, $p < .001$. The interaction term was not significant ($p = .162$).

Model 6 (avg_hmi_2_0 × stress). Average well-being change from T2–T0 showed the strongest association with sessions ($b = 35.32$, $p < .001$), stress remained negative ($b = -14.36$, $p < .001$), and the interaction was statistically significant and negative ($b = -9.53$, $p < .001$).

Across the interaction models, the negative interaction (when significant) indicates that as stress increases, the positive association between well-being improvement and engagement becomes smaller (i.e., stress dampens the well-being→engagement relationship).

Models 7–9: Average Well-Being change × stress, adjusted for demographics

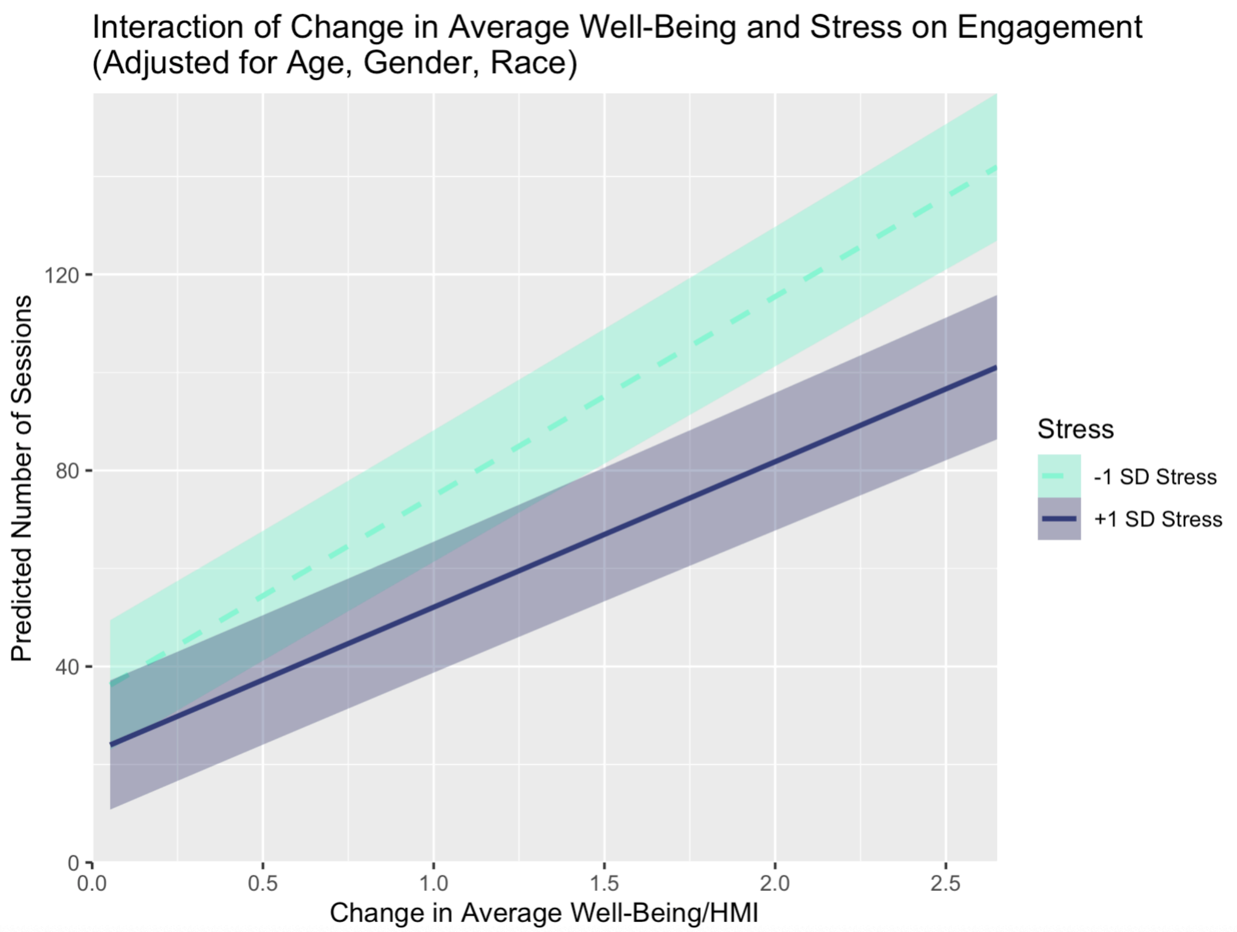
These models added age, gender, and race covariates.

Model 7 (avg_hmi_1_0 × stress + age + gender + race). Adding demographics increased explained variance modestly versus Model 4 ($R^2 \approx .0275$). The interaction term remained statistically significant ($p < .05$), indicating the stress-moderation effect persisted after adjustment.

Model 8 (avg_hmi_2_1 × stress + age + gender + race). Model 8 (the demographics-adjusted T2–T1 model) had $R^2 = .025$ and the interaction remained non-significant compared to Model 5 (avg_hmi_2_1_c:stress_avg_c $p = .159440$).

Model 9 (avg_hmi_2_0 × stress + age + gender + race). This was the best-fitting model overall ($R^2 = .044$; adjusted $R^2 = .043$). Average well-being change (T2–T0) was strongly positively associated with sessions ($b = 35.20$, $p < .001$), stress was negatively associated ($b = -10.03$, $p < .001$), and the interaction was statistically significant ($b = -8.04$, $p < .01$). Several age categories and male gender were also significant positive predictors in this adjusted model. The interaction effect in this model can be seen below in Figure 6.

Figure 6: Interaction between change in well-being & stress from T0 to T2



Notably, in Model 9, race indicators were not statistically significant in the displayed output, while age and gender showed clearer associations (e.g., several age categories positive vs the reference group, and gendermale positive).

A summary of all models is shown below in Table 6.

Table 6: Summary of the nine engagement models (n_sessions as outcome)

Model	Predictor set	Time window	Interaction included	Covariates	R^2 (Adj. R^2)
1	Awareness/Connection/Insight/ Purpose change + stress	T1–T0	No	None	.0224 (.0222)
2	Awareness/Connection/Insight/ Purpose change + stress	T2–T1	No	None	.0136 (.0131)
3	Awareness/Connection/Insight/ Purpose change + stress	T2–T0	No	None	.0302 (.0297)
4	Avg HMI change + stress	T1–T0	Yes	None	.0224 (.0223)
5	Avg HMI change + stress	T2–T1	Yes	None	.0129 (.0126)
6	Avg HMI change + stress	T2–T0	Yes	None	.0334 (.0332)
7	Avg HMI change + stress	T1–T0	Yes	Age + Gender + Race	.0275 (.0264)
8	Avg HMI change + stress	T2–T1	Yes	Age + Gender + Race	.0248 (.0221)

9	Avg HMI change + stress	T2–T0	Yes	Age + Gender + Race	.0444 (.0426)

Key comparison takeaway (across all nine)

The T2–T0 (longer-interval) average well-being change model with demographics (Model 9) showed the strongest overall fit and the most consistent pattern: higher well-being improvement predicted more sessions, higher stress predicted fewer sessions, and stress significantly moderated (reduced) the strength of the well-being–engagement relationship.

Discussion

The primary aim of this capstone project was to evaluate methodological strategies for modeling the relationship between well-being-related change, perceived stress, and user engagement within a large-scale digital well-being platform. Using a sequence of nine regression models, the project examined how different operationalizations of well-being change, time windows, and covariate structures influenced observed associations with engagement behavior, measured as the number of completed sessions. Rather than prioritizing causal claims, the project focused on comparing model specifications and assessing the robustness and interpretability of results across analytic choices.

Across all model families, several consistent patterns emerged. First, improvements in composite well-being scores were positively associated with engagement, particularly when change was defined over longer time intervals. Second, perceived stress was consistently associated with lower engagement, and in models including interaction terms, stress moderated the relationship between well-being change and engagement. Third, models incorporating longer cumulative change windows (T2–T0) explained more variance than those limited to adjacent assessment intervals (T1–T0 or T2–T1). Together, these findings highlight the importance of temporal framing, stress context, and model specification when analyzing real-world digital mental health data.

Interpretation of Engagement as the Outcome

A defining feature of this project is the decision to model engagement as the outcome rather than as a predictor of well-being change. This choice reflects both practical and conceptual considerations. From a practical standpoint, engagement metrics such as session count are measured with high granularity and temporal precision, whereas well-being assessments are sparse and irregular. Conceptually, engagement in digital well-being platforms can be understood as an adaptive behavioral response that may be influenced by perceived benefit, psychological readiness, and contextual stressors.

The results suggest that users who demonstrate greater improvements in well-being, particularly over extended periods, also tend to engage more frequently with the platform. This association may reflect a feedback process in which perceived benefit reinforces continued use. Importantly, the project does not assert that improvements in well-being cause increased engagement, or vice versa. Instead, the observed associations are interpreted as mutually reinforcing patterns within a self-selected user population.

Role of Perceived Stress as a Moderator

One of the most informative findings of the project concerns the role of perceived stress. Across all relevant models, higher average stress was associated with fewer completed sessions. Moreover, in models using cumulative average well-being change (T2–T0), the interaction between stress and well-being change was statistically significant and negative. This indicates that the positive association between well-being improvement and engagement was weaker among users reporting higher stress.

This moderation effect has important interpretive implications. Rather than functioning solely as a confounder, perceived stress appears to shape how well-being-related changes translate into engagement behavior. For users experiencing higher stress, improvements in well-being may be less likely to result in increased engagement, possibly due to competing demands, cognitive load, or emotional exhaustion. Conversely, users with lower stress may be better positioned to convert well-being improvements into sustained platform use.

From a methodological perspective, these results underscore the importance of explicitly testing interaction effects rather than assuming uniform associations across subgroups. Models that omitted stress interactions would have overstated the strength and generalizability of the well-being–engagement relationship.

Importance of Time Window Selection

A key comparative insight from the nine-model framework is the influence of time window selection on model performance and interpretability. Models using cumulative change from baseline to the second follow-up (T2–T0) consistently demonstrated stronger associations and higher explanatory power than models restricted to adjacent intervals. This pattern suggests that well-being-related change may accumulate gradually and that short-term change scores may be too noisy or unstable to capture meaningful trajectories in real-world settings.

This finding is particularly relevant for digital mental health research, where follow-up assessments are often sparse and completion is voluntary. Short-term change measures may be disproportionately influenced by measurement error, transient mood states, or selective attrition. By contrast, cumulative change measures integrate information across longer periods, potentially yielding more stable indicators of sustained change.

Domain-Level Versus Composite Measures

The project also compared domain-level well-being change models with composite average well-being change models. While domain-level models demonstrated statistically significant associations in some cases, their explanatory power was generally lower and less consistent than that of composite models. This suggests that aggregating across domains may better capture the multidimensional nature of well-being.

That said, domain-level analyses remain valuable for hypothesis generation and content-specific evaluation. The stronger and more stable performance of composite measures in this project likely reflects both statistical advantages (reduced measurement error) and conceptual alignment with holistic well-being outcomes emphasized by the platform.

Contribution to Methodological Practice

The primary contribution of this project lies in its demonstration of a transparent, stepwise modeling strategy for large, observational digital health datasets. By systematically varying predictors, time windows, and covariate structures, the project illustrates how analytic conclusions can shift depending on modeling decisions. This approach helps guard against overinterpretation of single-model results and encourages more nuanced, comparative reasoning.

In addition, the project highlights the importance of aligning analytic choices with data-generating processes. The unbalanced longitudinal structure, substantial missingness, and right-skewed engagement distributions observed in the dataset are not methodological flaws but intrinsic features of real-world digital platforms. Addressing these features through careful model specification, rather than attempting to force idealized assumptions, strengthens the credibility of findings.

Limitations

Several limitations must be acknowledged when interpreting the results of this project. Most fundamentally, the observational nature of the data precludes causal inference. Associations between well-being change, stress, and engagement may reflect self-selection, unmeasured confounding, or bidirectional processes. Users who are more motivated, less stressed, or more psychologically resourced may both improve more and engage more, independent of any platform effect.

Missing data represent a second major limitation. Follow-up well-being assessments were completed by a relatively small subset of users, resulting in an unbalanced longitudinal design. While complete-case analysis was appropriate given the project's methodological aims, it may introduce bias if missingness is systematically related to unobserved variables. Similarly, engagement metrics were unavailable for users who registered but did not actively use the platform, limiting generalizability to engaged users.

Measurement limitations also warrant consideration. Perceived stress was operationalized as an average score across completed surveys, often based on a single observation. This limits the ability to model stress dynamics over time or to distinguish between chronic and acute stress. Engagement was measured primarily through session count, which captures frequency but not necessarily depth, quality, or contextual relevance of use.

Finally, model fit across all specifications was modest, with R^2 values indicating that a large proportion of variance in engagement remains unexplained. This is expected in behavioral data drawn from heterogeneous, real-world populations but underscores the need for cautious interpretation and the inclusion of additional predictors in future work.

Future Directions

Future extensions of this project could proceed along several complementary directions. First, richer longitudinal modeling approaches, such as mixed-effects or latent growth models, could be employed to explicitly model within-user change over time. These approaches would allow for more nuanced characterization of individual trajectories while accommodating irregular measurement intervals.

Second, engagement could be operationalized using alternative or multidimensional metrics, such as time spent, content type diversity, or adherence patterns. Examining whether different forms of engagement relate differently to well-being outcomes would deepen understanding of how users interact with digital well-being tools.

Third, stress could be modeled dynamically rather than as a static average. Incorporating time-varying stress measures or ecological momentary assessments would allow for more precise examination of how fluctuations in stress interact with engagement and well-being change. This approach would require a change in data collection or an improved process of cross-referencing dates.

Fourth, future work could explore stratified or subgroup analyses to assess whether associations differ by baseline well-being, demographic characteristics, or prior experience with contemplative practices. Such analyses would support more personalized interpretations and potentially inform adaptive intervention design.

Finally, integrating qualitative or mixed-methods approaches could provide valuable context for interpreting quantitative patterns. User narratives, feedback, or in-app reflections may help explain why some users translate well-being improvements into sustained engagement while others do not.

Conclusion

In summary, this capstone project demonstrates the value of comparative, methodologically transparent modeling approaches for analyzing complex digital mental health data. By examining multiple specifications across nine models, the project highlights how engagement, well-being change, and stress interact over time in a real-world setting. While findings are necessarily associative and context-dependent, they provide a defensible foundation for future analytic refinement and platform evaluation. More broadly, the project illustrates how careful methodological choices can enhance interpretability and rigor in the study of large-scale digital well-being interventions.

References

Goldberg SB, Imhoff-Smith T, Bolt DM, Wilson-Mendenhall CD, Dahl CJ, Davidson RJ, Rosenkranz MA. Testing the Efficacy of a Multicomponent, Self-Guided, Smartphone-Based Meditation App: Three-Armed Randomized Controlled Trial. JMIR Ment Health. 2020 Nov 27;7(11):e23825. doi: 10.2196/23825. PMID: 33245288; PMCID: PMC7732708.

Kral, T. R. A., Kesebir, P., Redford, L., Dahl, C. J., Wilson-Mendenhall, C. D., Hirshberg, M. J., Davidson, R. J., & Tatar, R. (2024). Healthy Minds Index: A brief measure of the core dimensions of well-being. *PloS one*, *19*(5), e0299352. https://doi.org/10.1371/journal.pone.0299352

Wellbeing is Our Mission. About | Healthy Minds Innovations. (n.d.). https://hminnovations.org/about/