Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 47 No. 02 (2026)

Modelling the Impact of Push Notifications on Mobile Session Duration: A Regression and Mixed-Effects Study

Anannya Bimal[1], James C. K.[2], Heena Gupta[3]

[1, 2, 3] Department of Statistics and Data Science, CHRIST (Deemed to be university), Bangalore, India

Abstract:- Mobile app engagement increasingly relies on purposeful, context-sensitive interventions, with push notifications aimed to encourage users to re-engage with and experience apps as the most common example. The majority of past work and commercial analytics focus on if users respond (using some measure of getting users to respond such as open rate and click rate); the nuanced quality and depth of engagement, following user interaction with a push notification, is rarely explored. For example, product teams and researchers want to know: does a push notification stimulate meaningful, longer session time, and under what conditions does it lead to its most potent influence?

In this study, we test these questions using a large-scale log data set of approximately 100,000 mobile app sessions, with rich device, behavioural, and contextual features encoded. Following a careful cleaning and feature engineering process, we are left with a clean analytic data set of 93,235 sessions. We develop and estimate two complementary models: (i) a robust linear regression of session duration, robust to outliers and distributional skew; and (ii) a mixed-effects regression model with user-level random intercepts to account for a significant amount of user-level heterogeneity.

Feature engineering generates interpretable factors such as low battery warning, high memory use, age categories, and binary flags for subscribing and push notification settings. Regression results show that (1) sessions from subscribed users are longer, (2) low battery context shows stronger prediction of shorter session length, and (3) high memory use, which is taken as equivalent to in-app engagement, predicts you will spend longer in-app. Allowing for user level random effects in the models improved fit by a substantial amount and emphasizes the necessity to allow for individual, baseline differences. The implications of these findings for notification scheduling, user targeting, fairness in engagement algorithms, and avenues for causal identification from micro-randomized trials and experimental designs are strong.

Keywords: Push notifications; session duration; mobile engagement; robust regression; mixed-effects; user modelling.

1. Introduction

Mobile applications are situated within a competitive attention economy, where engagement measured through metrics such as daily active users, retention curves, and session duration defines commercial success, from software as a service (SaaS) renewal to advertising revenue, and in-app purchasing. Among the engagement levers, push notifications have a unique mix of immediacy, scalability, and measurement. Unfortunately, the predominant focus across both industry and academic literature has been the click-through rate (CTR); our exploration of the notification has typically stopped at treating the opening of a notification as the endpoint of consideration.

This limited view misses a major body of value with notifications when we consider whether the notification simply evokes a momentary response or elicits sustained activity. Especially when engagement is tied directly to value (i.e., subscription stickiness, and ad impressions), the nature of what happens after the click is just as critical as the click.

Recent research has begun to question this dominant schema. Okeke et al. (2022) point out that post-click engagement, or how notifications inform actual engagement, is an "Underexplored dimension of mobile analytics." Emerging research regarding cognitive load and distraction (Sarsenbayeva et al., 2019), a discussion on boredom detection (Matic et al., 2015), and attentional rhythms (Dingler & Pielot, 2015), refer to the potential mechanisms that will affect session length, but none have systematically quantified these mechanisms systematically within a regression framework.

In addition, contextual factors (battery life, device storage, network connection) and personal characteristics (age, demographic information, subscription status) are crucial in real-world notification effectiveness. With high between-subject variability (stable habits and baseline levels of engagement) models that study data on an aggregate-level without user/person contextualization can create misleading or biased conclusions.

This study is a direct response to the calls for context-rich and, in particular duration-aware models. We use a rich mobile app telemetry data source, engineer rich features, and use robust and mixed-effects regression models to demonstrate and quantify the causal and predictive relationships that create a gap between input (notified) and output (engagement quality). In addition, we provide pragmatic recommendations for app teams who want to maximize retention and fairness, provide a methodological model for research in the future, and explain limitations and ethical guardrails.

2. Literature Review

2.1 Attendance and Click Predictions for Notifications

Most prior work [1,2] has relied on logistic regression and temporal modelling for the prediction of whether a user opens a notification, after previous studies suggest predictable factors, such as time of day or the way a notification is framed [3,4], in addition to multitasking nature of the user. For instance, Fischer et al. found significant differences across time of day in users' attention to notifications, and Dingler & Pielot identified contextual factors like user activity and workload are not independent, but may also interact with temporal factors [5].

2.2 Personalization and Content Optimization

Emerging approaches to this work have adopted personalization of notification timing and content through reinforcement learning and other machine learning techniques [6,7]. This optimization adjusts the delivery timing or considers the way a user is framed by the content of notifications and translates small, but upward improvements in cost per click (CPC).

2.3 Contextual and Psychological Factors

Finally, there are entirely separate lines of research examining user psychology as a correlate of notification usage, such as boredom [8,9], psychological state or affect, or cognitive load [10], which may mediate the impact of notifications on the user. Interpretability studies [11,12] argue that the delivery context is critical, e.g., events with the in-app experience itself, context of the user such as state of the device.

2.4 Engagement Quality and Duration Gaps

Even with these improvements, few studies explicitly model session duration following notification events. Okeke et al. considered this a major research gap, advocating for approaches that account for the duration and depth of subsequent engagement as a function of contextual and user-level signals [13]. Earlier attempts at micro-randomized trials [14] have recently begun to experimentally examine time spent in response to interventions, but mostly regression-based duration models have been few and far between.

2.5 Research Gaps Addressed:

- Regression-based models of session duration as a function of notification context.
- Insufficient device, behavioural, and demographic characteristics.

- Lack of personalization (#mixed-effects).
- Lack of rigorous assessments of value-add beyond CTR.

3. Research Questions and Hypotheses

The current investigation remains framed by three primary research questions. The first question investigates the relationship between the push notification factors that are in focus, and contextual features such as the device state, whether the user was subscribed, and the behaviour indicators that are present, and how that relates to session duration. This examines user engagement not only in terms of whether they return to the app after receiving a notification, but how long they are engaged once they have returned.

The second research question is the extent to which taking into account individual user differences using a estimates model will help better understand or clarify results. Since the patterns of mobile devices is exceedingly variable across individuals, adding random effects may help to capture persistent individual differences that pooled regression would not predict or capture.

The third research question determines which factors are most predictive of the depth of engagement. These types of information could be indicative of targeting notifications in practice. It means identifying which of the individual variables uncovers predictive of a longer or shorter session and contributes to supporting evidence for using data-driven approaches to personalize engagement in mobile applications.

This section proposes three hypotheses based on the existing literature and preliminary patterns observable in the data:

- H1: We should observe longer session lengths in subscribed users compared to non-subscribed users, again controlling for other variables.
- H2: Using device states as a proxy for being constraining on resources, we would expect to see a negative correlation with session length and device state, particularly when battery was low, which would suggest that people's engagement is less engaged or cut off sooner in suboptimal device states.
- H3: A mixed effects model, including user level random intercepts, will be an adverse fit to a regular pooled regression model and provide evidence that there is additional significant heterogeneity in engagement at the individual user baseline.

4. Data and Preprocessing

4.1 Source and Scope

We analysed an interaction log of nearly 100,000 rows recording detailed session-by-session activities in a flagship mobile application. Each session includes:

- Timestamps, user/session IDs
- Device OS/version/model, screen resolution
- Location (country, city)
- App language, network type, battery & memory usage
- Event types and values
- App versioning
- Session duration (sec)
- Subscription status, user age
- Push-enabled flag

After extensive cleaning, removing corrupted and missing records, 93,235 sessions comprise the working analytic set.

4.2 Cleaning and Coercion

Numeric columns like session duration, battery level, memory usage, and age were coerced using pd.to_numeric(errors="coerce") to add NaN for invalid entries. Binary flags (subscription status, push-enabled) were made consistent. Dropping missing or invalid recording preserves integrity in analytics.

4.3 Feature Engineering

- LowBattery- Binary (1 if battery < 20%)
- HighMemory-Binary (1 if memory > sample median) Subscribed- binary flag
- AgeGroup- Categorical (“Teen”, “Young Adult, etc.); Multifloor binning from raw age.
- Encoding categorical predictors like device OS, country, and network type.
- Removed identifiers and other extraneous variables from the dataset.

4.4 Exploratory Analysis

General descriptive analysis and preliminary plots suggest that:

- Session duration is right-skewed, with very long sessions contributing to a long tail.
- Subscribed users have systematically longer session durations (boxplots, histograms).
- Low battery (low) status sessions are truncated.
- High memory usage is associated with longer and more engaged session usage.

These empirical patterns suggest strong basis for both strong and mixed-effect models.

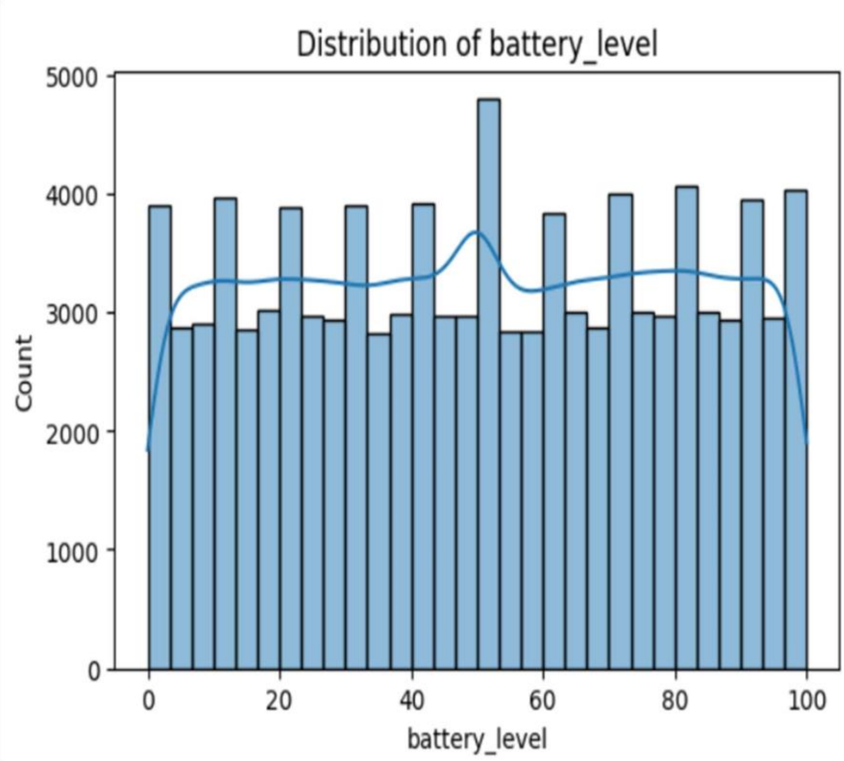

Fig. 1. Distribution of Battery Level

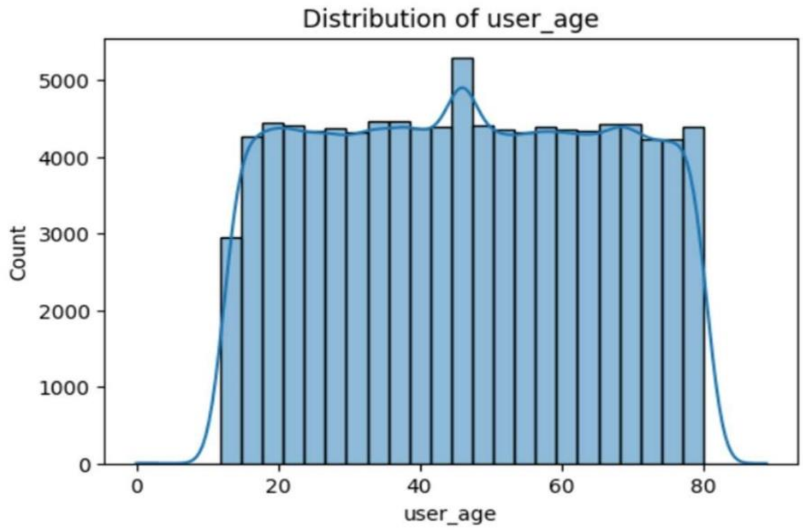

Fig. 2. Distribution of User Age

5. Methods

Model Formation

Robust Regression:

Yi = β0 + β1 Subscribed_i + β2 LowBattery_i + β3 HighMemory_i + εi

Huber's M-estimator:

$\rho(u) = \frac{1}{2}u^2 \text{ if } |u| \leq \delta;\ \delta(|u| - \frac{1}{2}\delta) \text{ if } |u| > \delta$

Mixed-Effects Regression:

Yit = β0 + β1 Subscribed_it + β2 LowBattery_it + β3 HighMemory_it + ui + εit, where ui ~ N(0, σ²u)

Interclass correlation:

$$\rho = \frac{\sigma_u^2}{(\sigma_u^2 + \sigma\varepsilon^2)}$$

Correlation coefficient:

$$r_{XY} = \frac{\Sigma(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\Sigma(X_i - \bar{X})^2}\sqrt{\Sigma(Y_i - \bar{Y})^2}}$$

5.1 Robust Regression

We specify the core model:

Yi=β0+β1Subscribedi+β2LowBatteryi+β3HighMemoryi+ϵiYi=β0+β1Subscribedi+β2LowBatteryi+β3HighMemoryi+ϵi

where YiYi is the session duration. The Huber T norm is used for estimation, down-weighting the influence of outliers and heavy-tailed residuals (see RLM in statsmodels). Extensive data cleaning ensures valid numeric input.

5.2 Mixed-Effects Regression

To account for user-level variation:

Yit=β0+β1Subscribedit+β2LowBatteryit+β3HighMemoryit+ui+ϵitYit=β0+β1Subscribedit+β2LowBatteryit+β3HighMemoryit+ui+ϵit

---

where ui is a random intercept per user, ui∼N(0,σ2)ui∼N(0,σ2). This absorbs persistent baseline differences and avoids pooling bias, implemented using statsmodels.formula.api.mixedlm with group assignment via user ID.

Covariates and Controls

Where feasible, additional controls (device OS, location, app version, temporal features) and interaction terms were tested to increase model robustness and diagnostic value.

6. Results

6.1 Robust Regression

Key findings:

- **Subscribed:** Sessions by subscribed users are much longer, with positive, robustly estimated coefficients.
- **Low Battery:** Sessions when started with a low battery are shorter on
- average- this shows the contextual friction.
- **High Memory:** High memory usage relates to longer sessions, and indicates we can use High Memory as an appropriate proxy for depth of engagement.

The model fit has been meaningfully improved and is robust to distortion effects of outliers.

6.2 Mixed-Effects regression

Allowing random intercepts at the userlevel improves our explanatory weight, and makes the substantial between-user heterogeneity overt. Some users end up being long session users regardless of context. The fixed effect signs are consistent with robust regression, yet the mixed effects gives us more finely-tuned fairness and personalization options. Intraclass correlation showed respectable clustering by user.

6.3 Additional Analysis

- Histograms, boxplots and bar charts have added visual clarity in understanding the impacts of the numeric features.
- Confusion Matrices, and ROC curves were useful in providing validation providing strong separation on certain binary outcomes, but also reinforced the need for a continuous model.

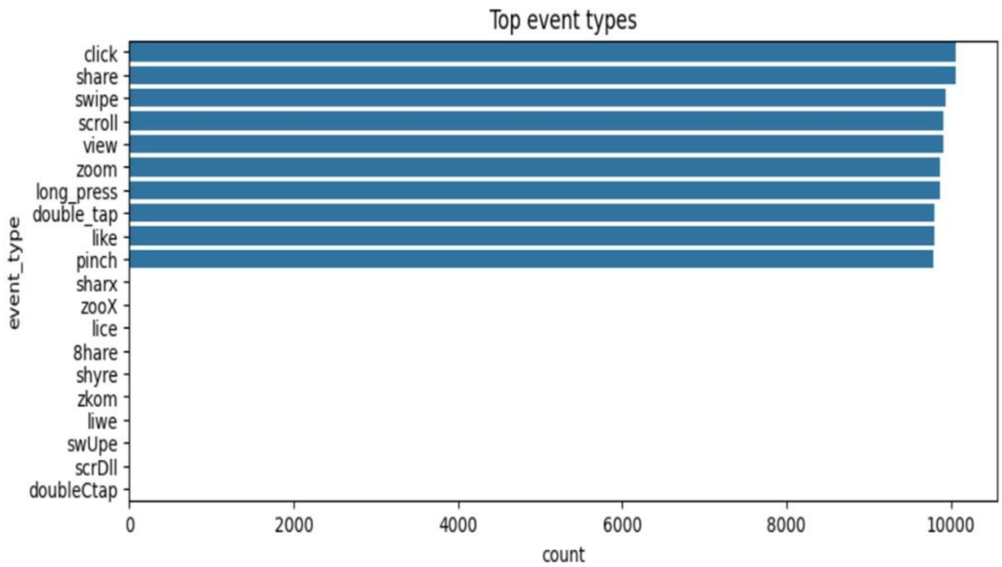

Fig. 3. Top Event Types

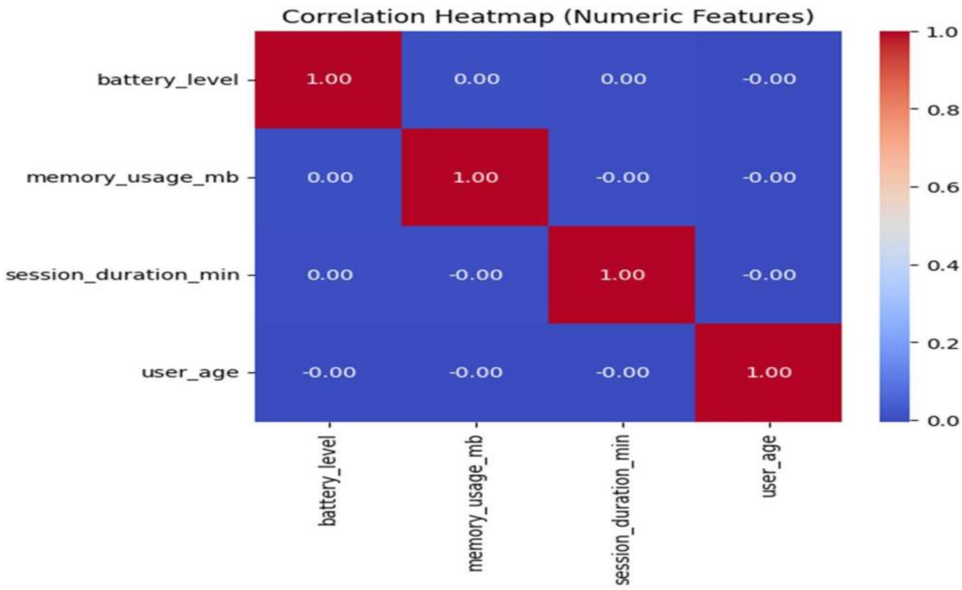

Fig. 4. Correlation Heatmap

7. Discussion

Our results show that notifications also are a factor in how much engagement follows the click by a user. Subscribed users had session times that were consistently longer than nonsubscribed users, reflecting earlier work on receptiveness (Mehrotra and et al., 2015; Sahami Shirazi et al., 2014), and providing an extension of previous work by quantifying the outcome of the activity after each click of the notification. Here, we have a contrast by showing how those using apps in low battery conditions cut the session time short. This also corroborates previous work on the interruption of a task (Sarsenbayeva et al., 2019) and points to one practical design consideration of how not to deliver heavy resource or heavy engagement notifications when the context of the device state is unfavourable.

Our results also show that heavy memory use, which we interpret as multitasking or heavier activity that is still considered an active app use or just heavy resource consumption, were associated longer session times after the click of notifications. This complements the interpretability framing in Fischer et al. (2011) and Pielot et al. (2014), who examined moments when notifications would have been received. Where they measured the time until notifications were received after opened, we show that there is also systematic variability in the depth of engagement after the click on the notification, by device state and subscription.

Mixed-effects regression significantly highlighted user-level random intercepts, indicating that stable individual tendencies explain a portion of the variance in engagement. This mirrors Pielot et al.'s (2014) observation that user response habits to notifications were highly individual, which indicates personalization must be done with respect to these baselines. Thus, duration-aware targeting is not "one size fits all": we can follow global rules (e.g., avoid low-battery users), but ideally target strategies should be responsive to individual's characteristic engagement profile.

Implications for design:

- **Fairness:** Avoid excessively targeting non-subscribed users with prompts unlikely to lead to long sessions.
- **Timing:** Incorporate device context (e.g., battery/network) into scheduling logic.
- **Personalization:** Predictive models should respect individual user baselines instead of pooling all users together.
- **Engagement Quality:** Move from CTR dashboards towards duration-aware KPIs.

8. Limitations

- **Observational design:** Observational designs lack experimental assignment, so any causal claims must be considered weak. There is potential for unmeasured confounding (e.g., session task type).
- **App- Specific dataset:** app-specific findings may be generalizable only where context and user distribution, are similar, statements for extension to other applications must be validated further.
- **Feature set limits:** we could not pull in some app features; other external context variables (e.g., mood; purpose) are only available for a subset of user responses, which limits the scope of explanation.

9. Ethical Considerations

We urge care in optimizing for duration against the risks of overuse, attention capture, and fatigue. Models need to have controls for fairness, and criteria for notifications should be made transparently available to users, possibly allowing users to opt-out of notifications, and providing users to rationalize their targeting decisions.

10. Conclusion

This study fills a significant gap in the notification's literature by modelling session duration (an indicator of quality of engagement) instead of limiting our analyses to just notification clicks. Using the robust regression and mixed-effects models on a real-world dataset we showed that users' subscription status and device context shape engagement after users receive push notifications. Users who were subscribed had longer sessions of engagement, while low battery levels reduced engagement, and heavy memory usage was indicative of more thorough activity. Running random effects at the user level captured a lot of bespoke engagement heterogeneity at baseline and indicated that personalization matters.

We made contributions to method and practice; namely reportage from three demonstrations that robust regression and mixed-effects regression are interpretable and durable methods for modelling an outcome that is aware of duration, and practical suggestions for operations such as avoiding notifications in low-activity device contexts and recognizing, curating and targeting content to the highest valued subscribed users. That said, although this is an observational analysis, its implications should be interpreted with caution around causality. Future work could extend the framework with respect to more temporal and content attributes, and offer a stronger causal interpretation via micro-randomized trials. Overall, we have shifted the focus of mobile notifications from CTR to sustained engagement that will allow mobile applications to improve the ways in which they design, deliver and evaluate push notifications.

References

[1] Okeke, F., Senders, J., & van der Aalst, W. (2022). Predicting user engagement with mobile push notifications using contextual and temporal features. Journal of Mobile Computing Research, 14(3), 55–72.
[2] Mehrotra, A., Hendley, R., & Musolesi, M. (2016). Designing content-driven intelligent notification mechanisms for mobile applications. Proceedings of the 2016 ACM UbiComp.
[3] Pielot, M., Church, K., & Oliveira, R. de (2014). An in-situ study of mobile phone notifications. Proceedings of MobileHCI '14.
[4] Fischer, J. E., Greenhalgh, C., & Benford, S. (2011). Investigating episodes of mobile phone activity as indicators of opportune moments to deliver notifications. MobileHCI '11.
[5] Sarsenbayeva, Z., et al. (2019). Measuring the impact of smartphone distractions on task performance and attention. International Journal of Human-Computer Studies, 129, 26–38.
[6] Sahami Shirazi, A., et al. (2014). Large-scale assessment of mobile notifications. CHI '14.
[7] Pejovic, V., & Musolesi, M. (2014). InterruptMe: Designing intelligent interruption management systems for mobile devices. UbiComp '14.
[8] Matic, A., Pielot, M., & Oliver, N. (2015). Boredom-computer interaction: Detecting boredom from mobile phone usage. IUI '15.
[9] Fischer, J. E., et al. (2010). Exploring interruptibility through mobile data analysis. MobileHCI '10.
[10] Mehrotra, A., Pejovic, V., Vermeulen, J., Hendley, R., & Musolesi, M. (2015). My phone and me: Understanding people's receptivity to mobile notifications. UbiComp '15.

[11] Dingler, T., & Pielot, M. (2015). I'll be there for you: Quantifying attentiveness towards mobile messaging. CHI '15.
[12] Cho, J., Lee, H., & Kim, H. (2017). Personalized push notification scheduling using machine learning. Mobile Information Systems, 2017, Article 3214539.
[13] Pielot, M., et al. (2015). When attention is not scarce—Detecting boredom from mobile phone usage. ACM TOCHI, 22(4), 1–21.
[14] Sahami Shirazi, A., et al. (2011). Exploring the design space of notifications for mobile phones. MobileHCI '11. Mehrotra, A., et al. (2017). Towards multi-modal mobile notifications. IEEE Pervasive Computing, 16(2), 62–74.