OIT367: Course Project Final Report

Leveling up Learning - What Drives Engagement on an EdTech Professional Development Platform?

By: Eric Bradford, Nikhil Gupta, and Saksham Gupta

12 March 2025

Executive Summary

The platform faces a significant engagement challenge in driving sustained use of its professional development content. While the current 70% completion rate for learning challenges is promising, understanding what drives sustained engagement is critical for optimizing the platform's educational impact. Our analysis of ~10,000 user records from June 2022 reveals distinct user segments and key engagement drivers that can inform targeted interventions.

We identified three distinct user segments based on engagement patterns (four if we include inactive users): Occasional Users (low activity despite varied time on platform), High-Intensity Users (high activity over short durations), and Long-Term Engaged Users (moderate-to-high activity over extended periods). Through multiple analytical approaches - including clustering, decision trees, logistic regression, LASSO regression, and neural network - we consistently found that skill completion is the strongest predictor of sustained engagement, followed by a few other features such as notification preferences and reputation scores.

Our analysis reveals that teachers who complete more than 8.5 skills are significantly more likely to become Long-Term Engaged Users, especially when they maintain higher reputation scores (>4.05). Weekly notification preferences for new messages and evidence feedback correlate strongly with sustained engagement, suggesting users who customize their platform communications are more invested in their learning journey. Additionally, we found notable differences in engagement patterns across teaching levels and countries, as shown in the Appendix.

We recommend the platform implement targeted interventions focused on guiding users to complete their first skills as quickly as possible, enhance notification systems to encourage sustained interaction, and highlight growing reputation scores to reinforce engagement. The decision tree and neural network models developed can be implemented on the platform's current dataset to refine intervention strategies for each user segment, ultimately supporting the platform's mission of improving teaching practice globally.

I. Introduction

The platform addresses the global shortage of high-quality teachers through a blend of technology and pedagogical innovation. The platform enables teachers to complete learning challenges and submit evidence from their teaching practice, which then receives feedback from peers, school leaders, and AI models.

This study analyzes user behavior patterns on the platform to identify key engagement drivers and predict which users are likely to maintain consistent platform usage. By examining interaction data from ~10,000 users, we sought to answer the key question: **Which user behaviors and metrics on the platform are indicative of the "tipping point" of sustained engagement?** The "tipping point" is the critical moment when a product crosses a threshold, leading to a sudden and significant shift in usage.

These insights are particularly valuable for product managers and customer success teams who can leverage them to refine notification strategies, enhance user experience elements, and develop targeted interventions for users who might otherwise disengage. Improving engagement not only benefits the platform's metrics but ultimately supports the core mission of enhancing teaching quality and addressing educational challenges globally.

II. Data

Our key data source is user engagement data that we received directly from the platform's team. The dataset contained 12,365 rows, each corresponding to a platform user as of June 9, 2022. During our data preparation phase, we identified and removed 1,493 duplicate user IDs. The client reported data migration issues that prevented newer data from being extracted from the past 3 years. The dataset provides a variety of information including teaching levels, country, activity scores, evidence interactions, chat history, skill completion, reputation scores, and account details such as registration date and last access date. To ensure meaningful analysis, the data required additional cleaning to filter out users who registered after April 9, 2022 due to insufficient time on the platform to establish usage patterns, or never accessed the platform.

III. Methods

A) Segmentation and Statistical Correlation Analysis

We applied K-means clustering to segment platform users based on engagement duration (difference between registration and last accessed date) and activity score. The elbow method identified three optimal clusters.

We also performed statistical correlation analysis to identify relationships between metrics, conducting chi-square tests to evaluate categorical associations, and creating visualizations including distribution charts, box plots, etc. We also analyzed country-specific behavior patterns and notification preferences to provide additional context to our user segments.

B) Decision Tree

To predict user segments, we implemented a multi-class decision tree classifier using engagement metrics (reputation, completed skills, chats, evidence interactions) and categorical notification preferences. Decision trees provide clear, interpretable rules which can help predict which segment a relatively new user might fall into based on their initial behaviors and settings, enabling proactive engagement strategies to drive sustained usage.

A preprocessing pipeline handled mixed data types, using passthrough for numerical features and one-hot encoding for categorical variables. We limited tree depth to three levels to maintain interpretability. The model was trained on 70% of the dataset and validated on 30%, with performance assessed through precision, recall, F1-score, and a confusion matrix.

After developing a decision tree for all users, we replicated the same methodology for just the active users (i.e., after filtering out the inactive users) to focus on understanding the distinguishing characteristics among active user segments. Given the greater interpretability of those results, we applied an approach of excluding users with activity scores of 5 (no activity) for subsequent analyses.

C) Logistic Regression

We employed logistic regression to quantify the impact of specific user features on high engagement. We created a binary outcome variable called "high_engagement" by applying specific thresholds to our continuous engagement metrics. Users were classified as highly engaged (1) if they had been active on the platform for more than 365 days and maintained an activity score above 50, and as not highly engaged (0) otherwise. We included the same feature set as our decision tree model, encompassing both numerical engagement metrics and categorical notification preferences. Model performance was evaluated using an ROC curve and AUC score.

D) LASSO Regression

To comprehensively analyze user engagement beyond binary classifications, we implemented a LASSO regression model. LASSO regression predicted a continuous engagement score by normalizing and weighting engagement duration and activity level on a 0-100 scale.

After applying cross-validation to identify the optimal alpha (regularization parameter), the model systematically eliminated less significant features by driving their coefficients to zero, providing a streamlined view of the truly influential engagement drivers. This complemented our decision tree and logistic regression analyses by offering a third analytical lens: while decision trees identified the hierarchical importance of engagement factors and logistic regression quantified their impact on high engagement probability, LASSO helped isolate the strongest predictors of engagement intensity while controlling more for overfitting.

E) Neural Network

This study employs a deep learning approach to model user engagement on the platform. We implemented a three-layer feedforward neural network using PyTorch. This feeds into the first hidden layer consisting of 64 neurons with ReLU activation and 20% dropout. The second hidden layer reduces to 32 neurons, also with ReLU activation and 20% dropout, before connecting to a single-neuron output layer for regression prediction. The architecture incorporates dropout regularization to prevent overfitting, particularly important given the limited dataset size.

We partitioned the dataset as follows: 64% for training, 16% for validation, and 20% for testing. The model was trained using Mean Squared Error as the loss function and the Adam optimizer with a learning rate of 0.001. We implemented training with a batch size of 32 and incorporated early stopping based on validation loss performance, with a maximum cap of 100 epochs. We use a similar evaluation criteria as the lasso regression model.

IV. Results

User Clustering and Correlation Analysis

Our cluster analysis identified three distinct user segments (in addition to the Inactive Users) based on engagement duration and activity scores (Figure 1). The "High-Intensity Users" demonstrate high activity levels (150-370) with relatively short engagement durations (~< 250 days), suggesting intense but potentially shorter/newer platform usage. "Long-Term Engaged Users" maintain moderate to high activity levels (50-270) over extended periods (500-2000 days), indicating sustained engagement. "Occasional Users" show the lowest activity levels (~< 80) despite varying time on the platform, suggesting minimal interaction regardless of account age.

The correlation matrix (Figure 2) revealed moderate associations between completed skills and other engagement metrics, with the strongest correlations observed between completed skills and evidence consulted (0.67) and between evidence consulted and evidence evaluated (0.93).

Further analyses with rich insights on gaps in sustained usage are in the Appendix.

Fig 1: 3 active user segments identified through k-means clustering

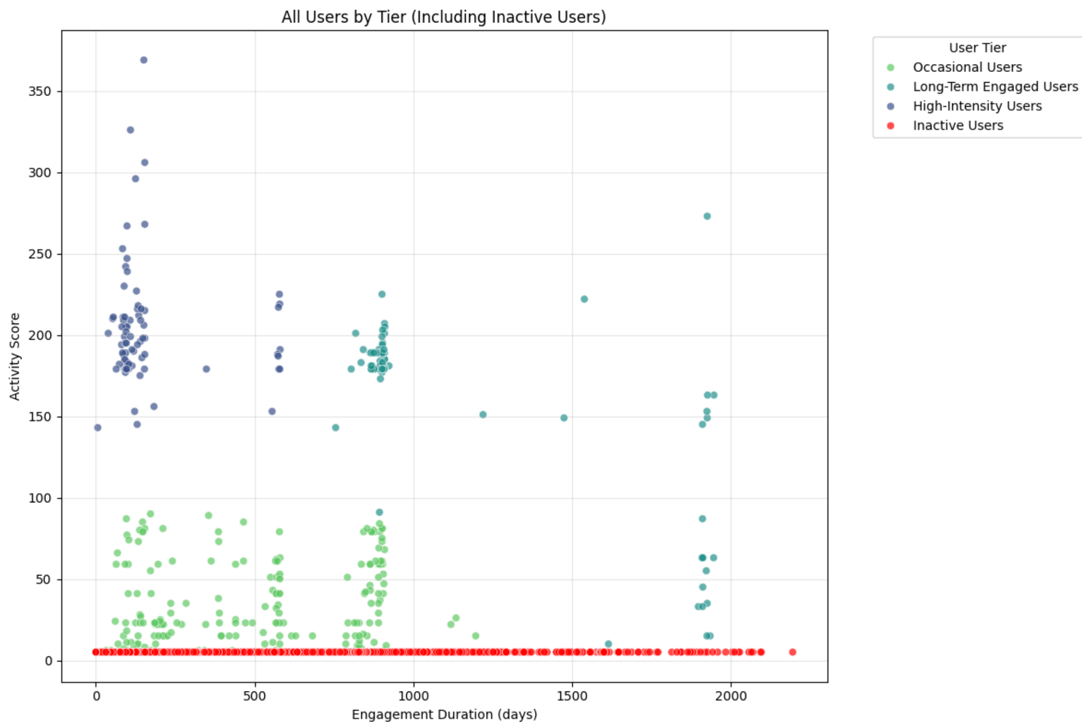
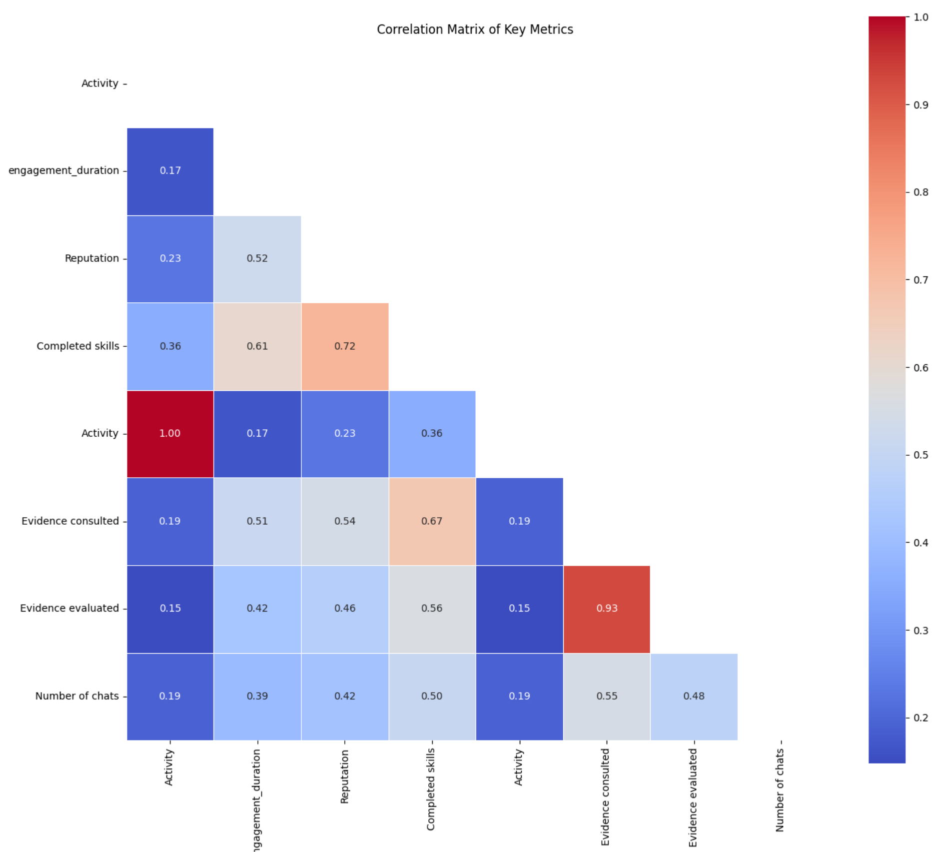


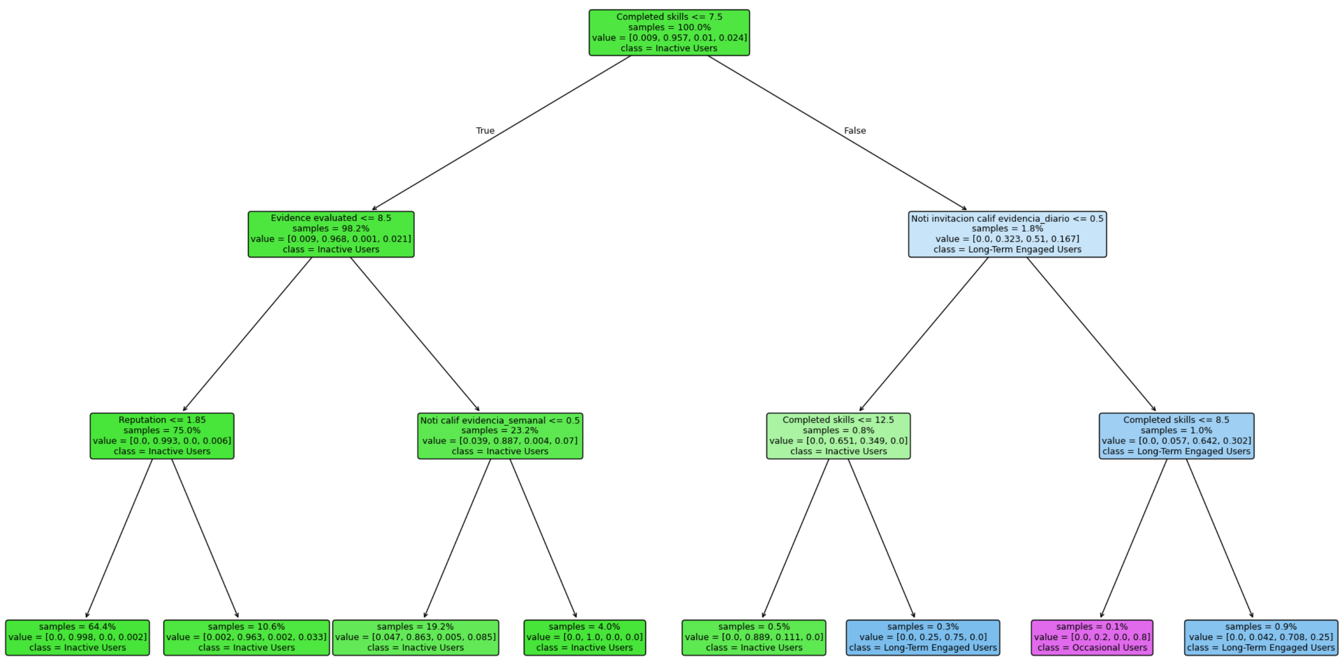
Fig 2: Correlation matrix between all the numerical features corresponding to an user



Decision Tree Analysis

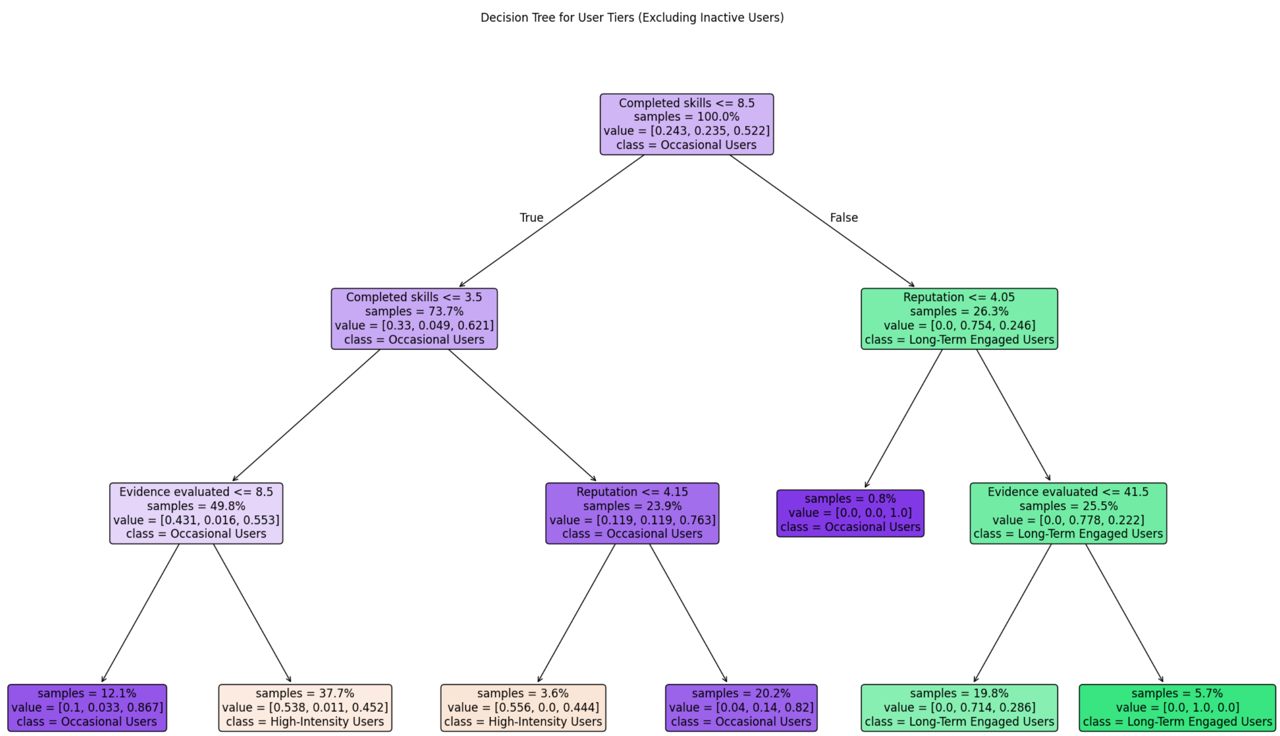
Our initial decision tree analysis included all users, but the high proportion of inactive users (activity score = 5) heavily skewed the results, with most decision paths leading to the inactive classification (Figure 3). This limited the model's ability to distinguish meaningful patterns among active users. We subsequently filtered out inactive users, resulting in a more insightful tree (Figure 4).

Fig 3: Decision tree for user tiers (full set with inactive users)



This refined decision tree revealed that completed skills is the primary determinant of user tier classification. Users who completed more than 8.5 skills were further segmented based on reputation scores, with higher reputation strongly associated with Long-Term Engaged Users. For users with fewer completed skills, evidence evaluation activity and reputation scores served as secondary distinguishing factors, with higher values in either category increasing the likelihood of classification as High-Intensity Users. This hierarchical structure confirms that active skill completion is fundamental to sustained platform engagement, with reputation and evidence evaluation serving as important supporting factors.

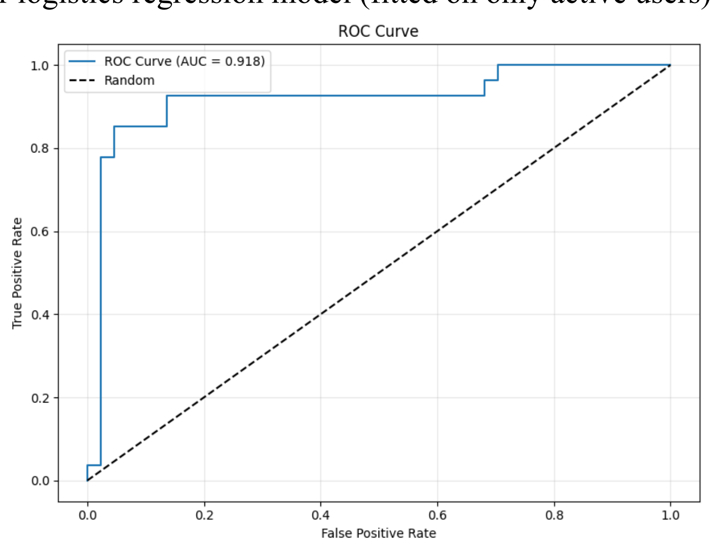
Fig 4: Decision tree for user tiers (excluding inactive users)



Logistic Regression

The model achieved a high AUC score of 0.918 (Figure 5), suggesting strong predictive power. However, this performance metric should be interpreted cautiously. The step-like pattern of the ROC curve showed the model is making predictions for a very small number of samples. The confusion matrix revealed that while the model accurately classified most low-engagement users, there is class imbalance inflating the AUC.

Fig 5: ROC curve for logistics regression model (fitted on only active users)



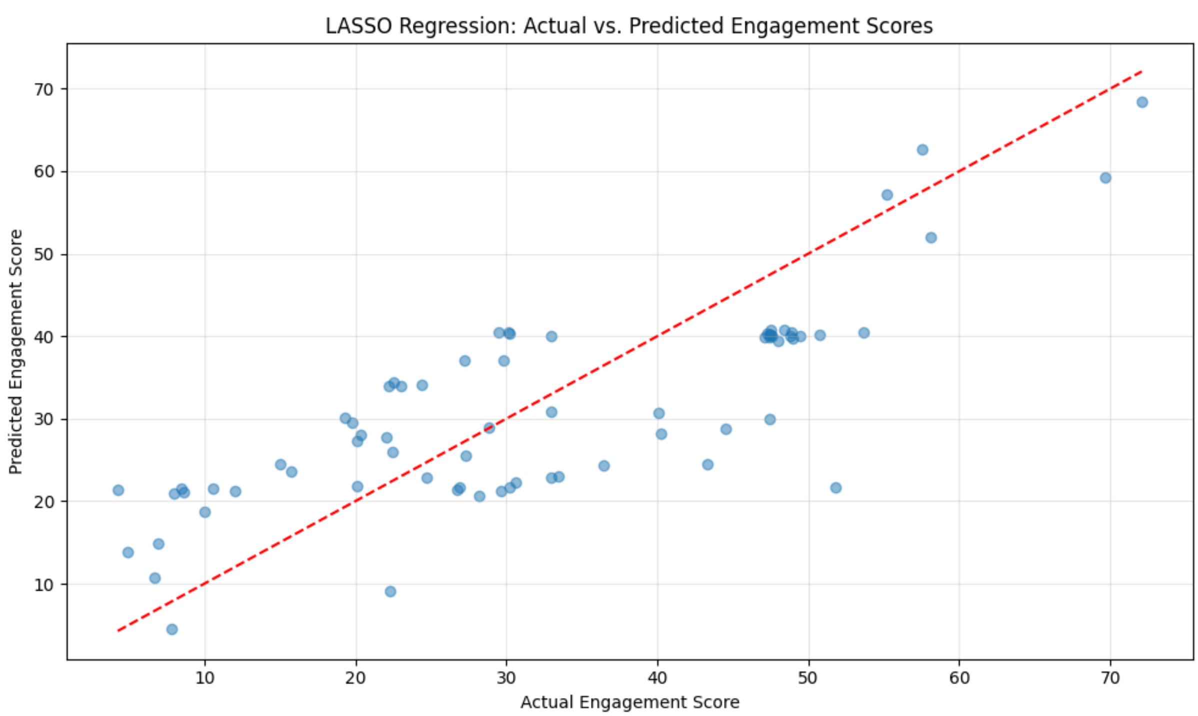
Despite this limitation, the model's feature coefficients indicated that completed skills, weekly new message notifications, and reputation scores were the strongest positive predictors of high engagement. The significant impact of notification settings suggests that users who customize their platform communications are more likely to maintain engagement, potentially indicating a relationship between platform investment and sustained usage.

LASSO Regression

Our LASSO regression model predicted a continuous engagement score combining normalized engagement duration and activity metrics. The actual versus predicted plot (Figure 6) shows a moderate correlation between predicted and actual values, with better prediction accuracy at the extremes of the engagement spectrum. The R^2 value of approximately 0.63 indicates the model explains a moderate portion of engagement score variance.

The primary strength of the LASSO analysis was its automatic feature selection capabilities. The non-zero coefficients highlight completed skills as the dominant predictor of engagement, with a coefficient nearly twice as large as the next most important factor (weekly new message notifications). The model also retained the number of chats, reputation score, instantaneous evidence, feedback notifications, and evidence interaction metrics as meaningful predictors, although with substantially smaller coefficients.

Fig 6: Regression curve for actual vs. predicted engagement scores (R^2 = 0.6259)

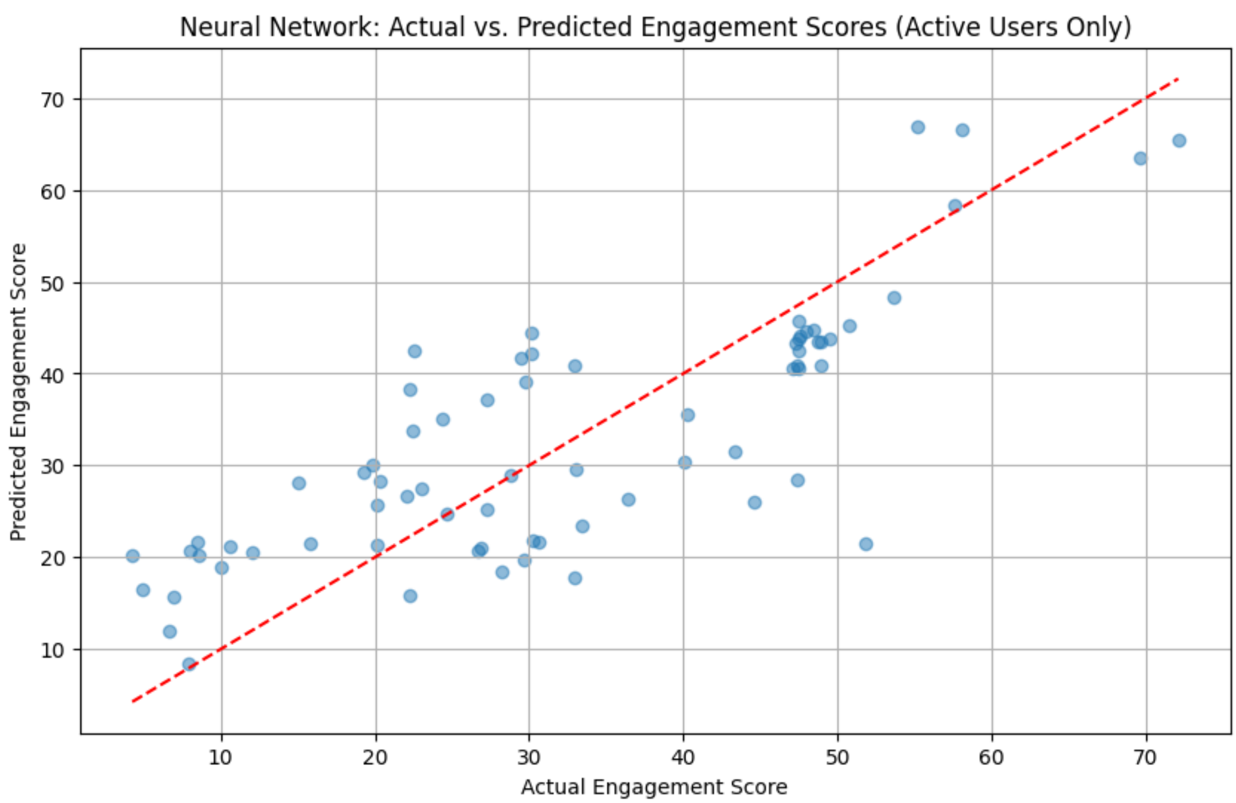


Neural Network

Our neural network analysis built upon previous modeling approaches by focusing exclusively on active users to avoid skewing effects. The model demonstrated effective convergence during training, with appropriate generalization despite our limited dataset size. The model's predictions show a strong linear relationship with actual engagement scores (Figure 7), achieving an R^2 value of approximately 0.63 – similar to the LASSO regression.

The neural network captured nuanced relationships than the decision tree’s rigid boundaries and provided more granular engagement predictions than logistic regression’s binary classification. Despite these methodological advantages, the consistent importance of completed skills across all modeling approaches reinforces this metric as the most reliable predictor of sustained platform engagement.

Fig 7: Regression curve for actual vs. predicted engagement scores ($R^2 = 0.6259$)



Generalizability and Limitations

While our models provide valuable insights, several factors limit their generalizability. First, the dataset consists of only 10,871 unique users as of June 2022, of which most are inactive, which may not represent current platform usage patterns. Second, the strong reliance on completed skills as a predictor may be redundant, as skill completion inherently would contribute to the activity score used in our dependent variables. Third, the significance of notification settings might indicate selection bias, where engaged users are more likely to customize notifications rather than notifications driving engagement.

Applying these models to the platform's expanded user base post-data migration could enhance robustness, revealing deeper engagement drivers. Future models would also benefit from temporal data (which was not available in the current dataset) to track engagement trends over time.

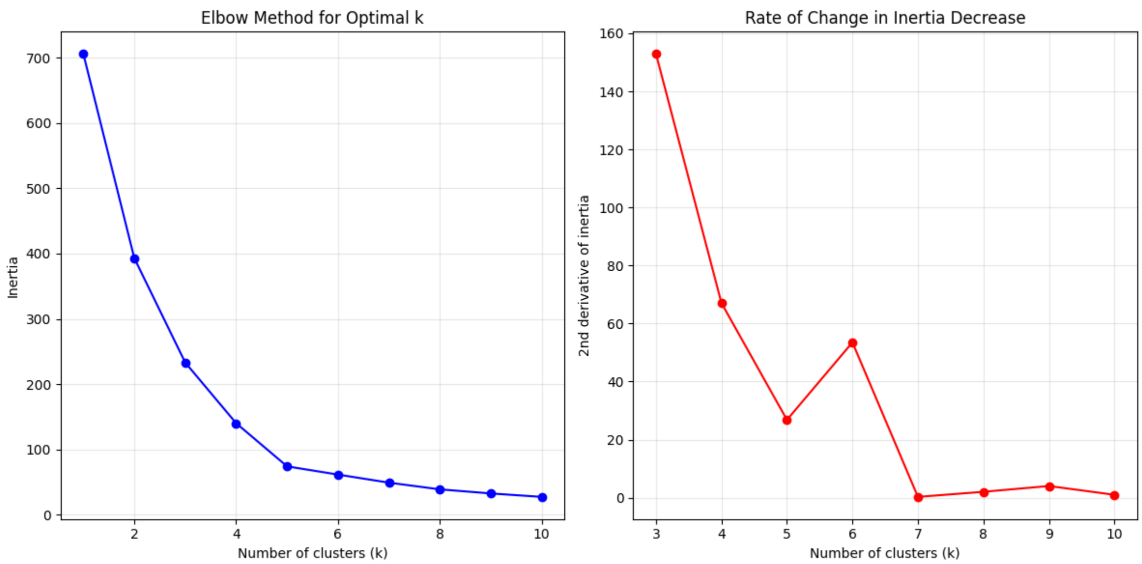
V. Conclusion

Our analysis revealed that while a significant portion of the platform's user base remains inactive, focusing on the subset of users who demonstrate engagement provides valuable insights into behavior patterns that drive sustained platform usage. Across multiple modeling approaches, skill completion emerged as the tipping point of engagement, followed by notification preferences—particularly for new messages—and reputation scores. The correlation between evidence interactions (both consulting and evaluating) further suggests that active participation in the platform's collaborative feedback mechanisms is integral to user retention. These findings highlight the importance of not just initial onboarding but facilitating early successes through skill completion and meaningful peer interactions.

Based on these insights, we recommend the platform implement targeted interventions to guide users toward completing their first skills, which appears to be a critical activation threshold. Specifically, enhancing notification systems—especially for new messages and evidence feedback—could significantly boost engagement among moderately active users. Additionally, highlighting users' growing reputation scores may reinforce engagement, leveraging the positive correlation between reputation and platform commitment. The platform could implement our decision tree and neural network models on their current, larger dataset to develop more precise intervention strategies tailored to each user segment's specific engagement patterns and needs. While correlation does not equate to causation, the aforementioned indicators represent a strong, positive signal and a good starting point for experimentation.

Appendix

Exhibit 1: Performing elbow method to find optimal number of clusters




```
Using 3 clusters for user segmentation...

Cluster Statistics:
   engagement_duration    Activity  Count  Percentage  avg_rank
0           508.000000   37.375000  184.0   52.124646       1.5
1           162.271605  202.765432   81.0   22.946176       2.0
2          1126.363636  161.897727   88.0   24.929178       2.5
```

Exhibit 2: Long-term engaged users have a slightly higher mean reputation score than other active user categories

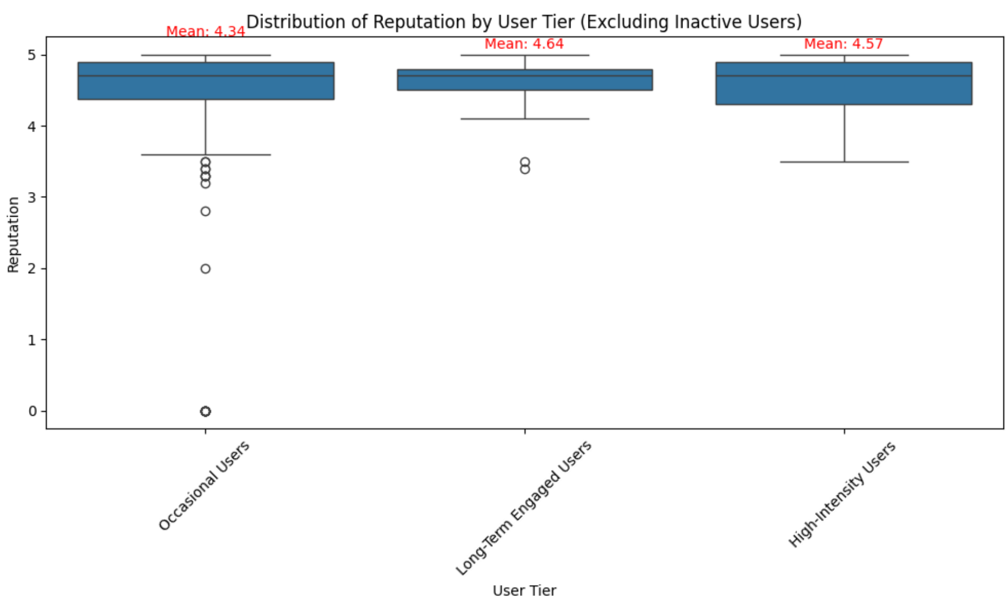


Exhibit 3: Nursery and primary school teachers tend to have more sustained usage patterns versus university teachers, who have high activity levels over a shorter duration of time

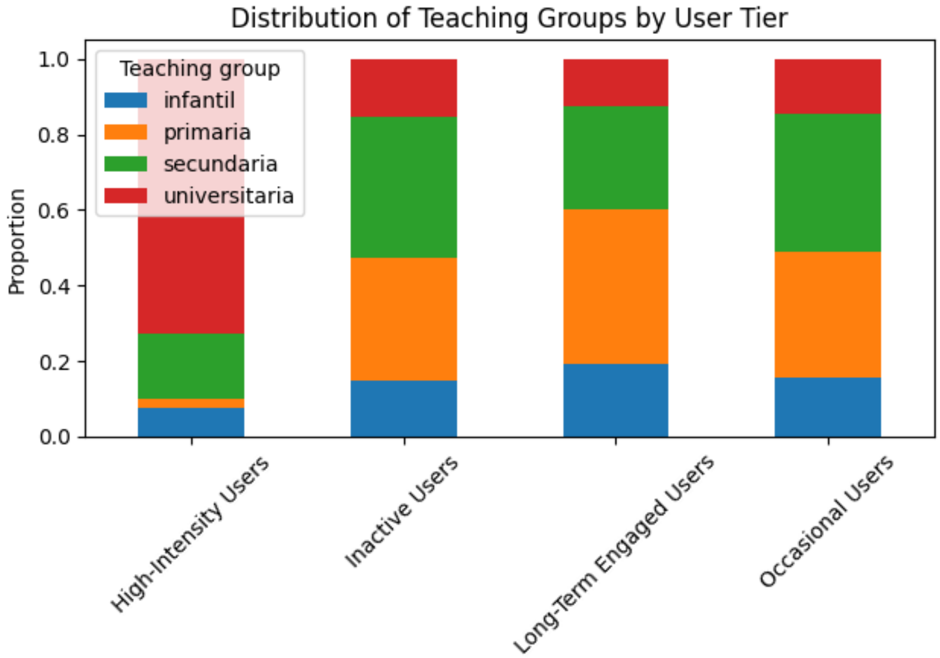
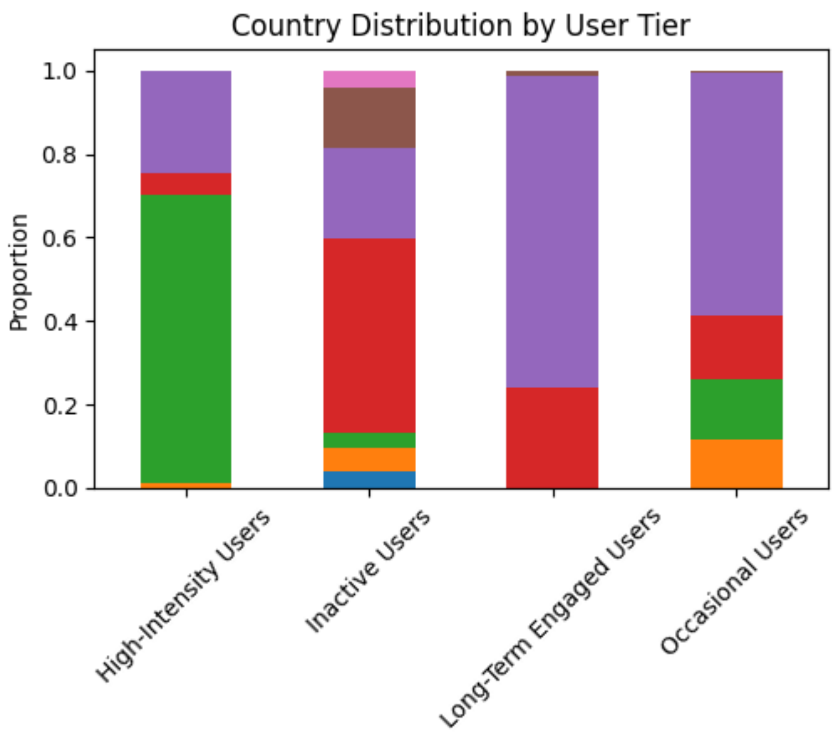
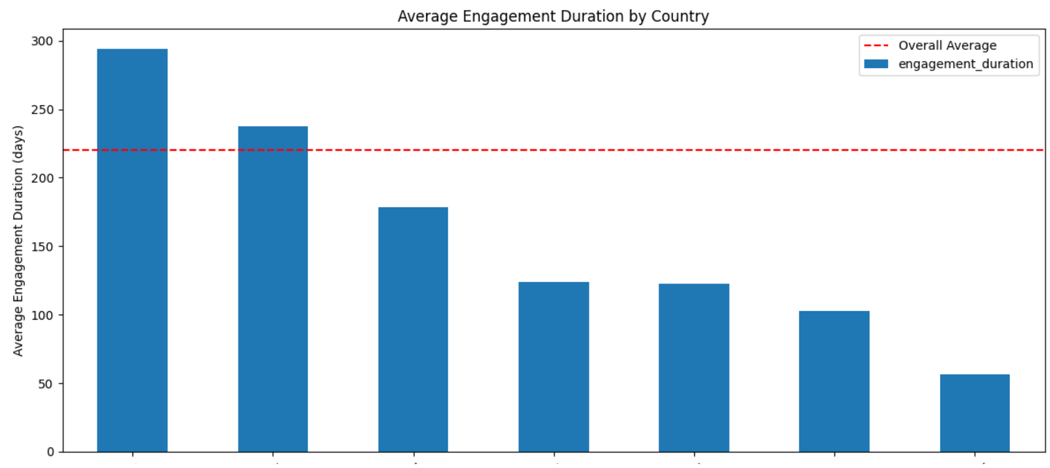


Exhibit 4: Teachers in [REDACTED] have higher usage levels [SOME GRAPH AXES SANITIZED]





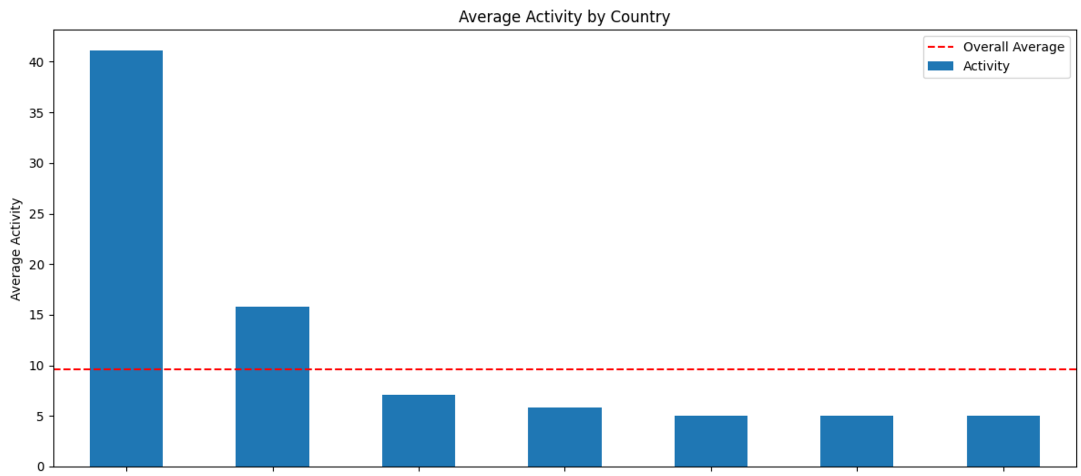
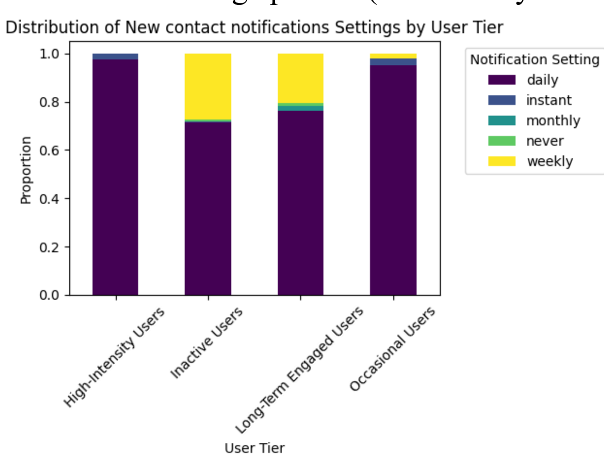
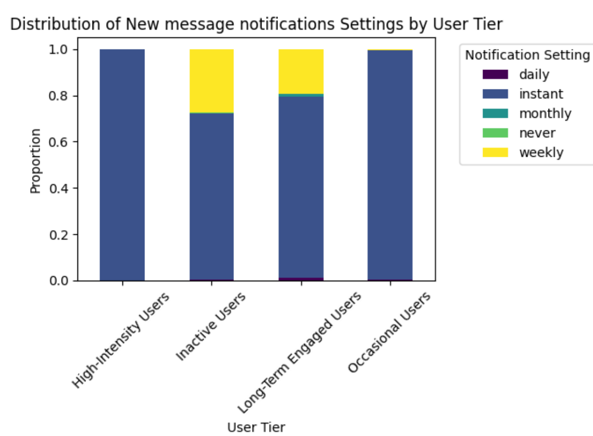
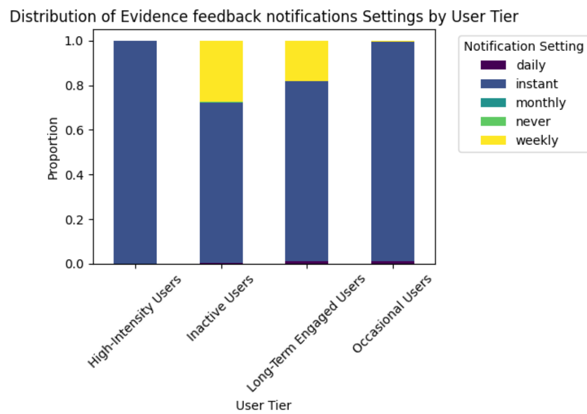


Exhibit 5: Long-term engaged users tend to have a higher weekly notification preference given their sustained usage pattern (versus daily for others who may be active only for a short period)







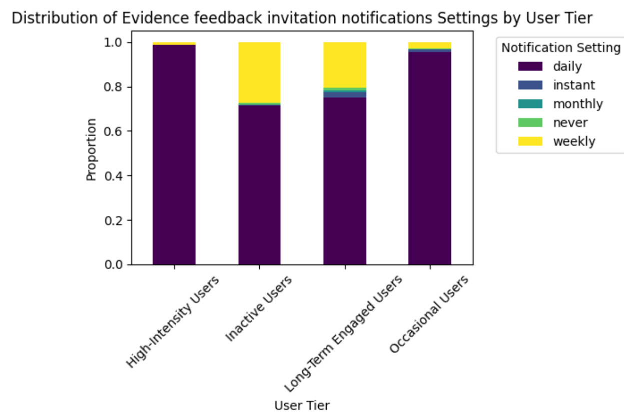
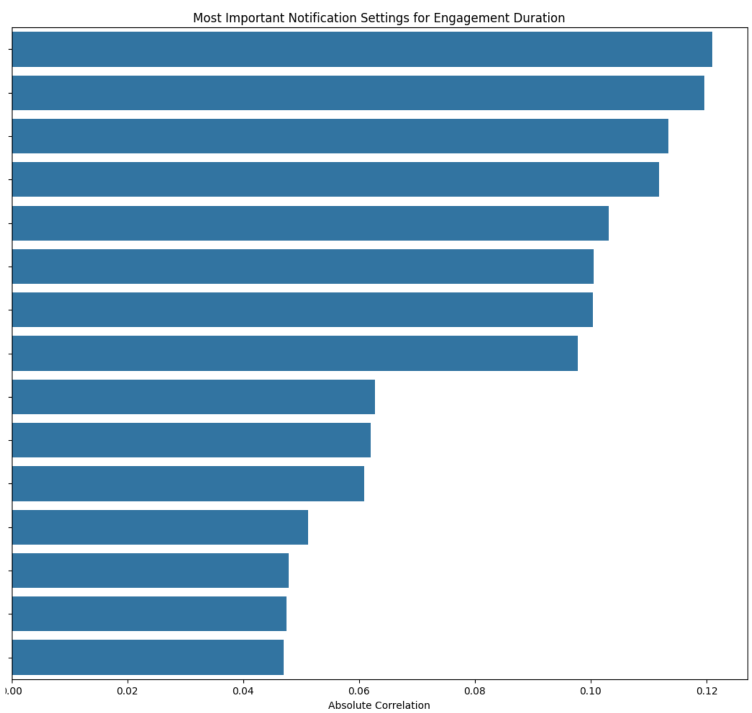


Exhibit 6: Correlation of notification preference features with our key usage metrics: activity score and engagement duration [SOME GRAPH AXES SANITIZED]



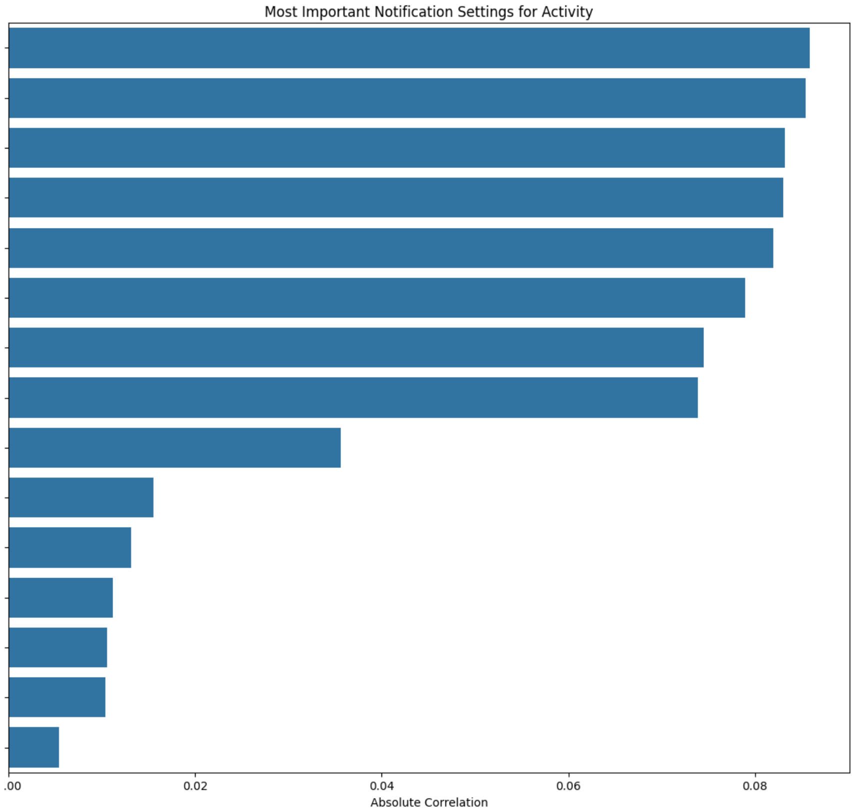
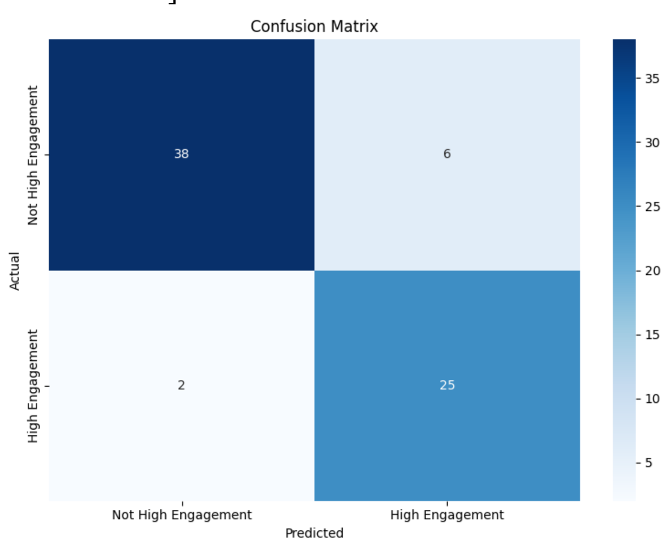
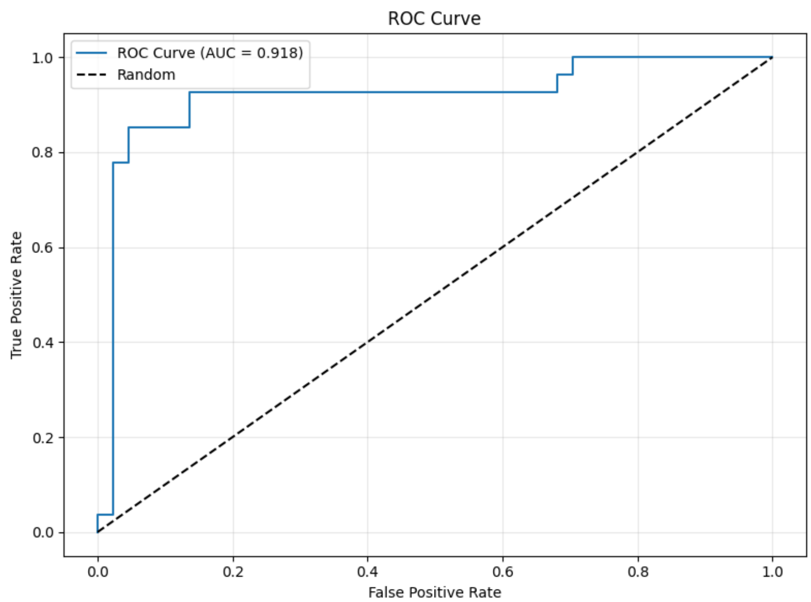


Exhibit 8: Logistic regression results without inactive users [SOME GRAPH AXES SANITIZED]





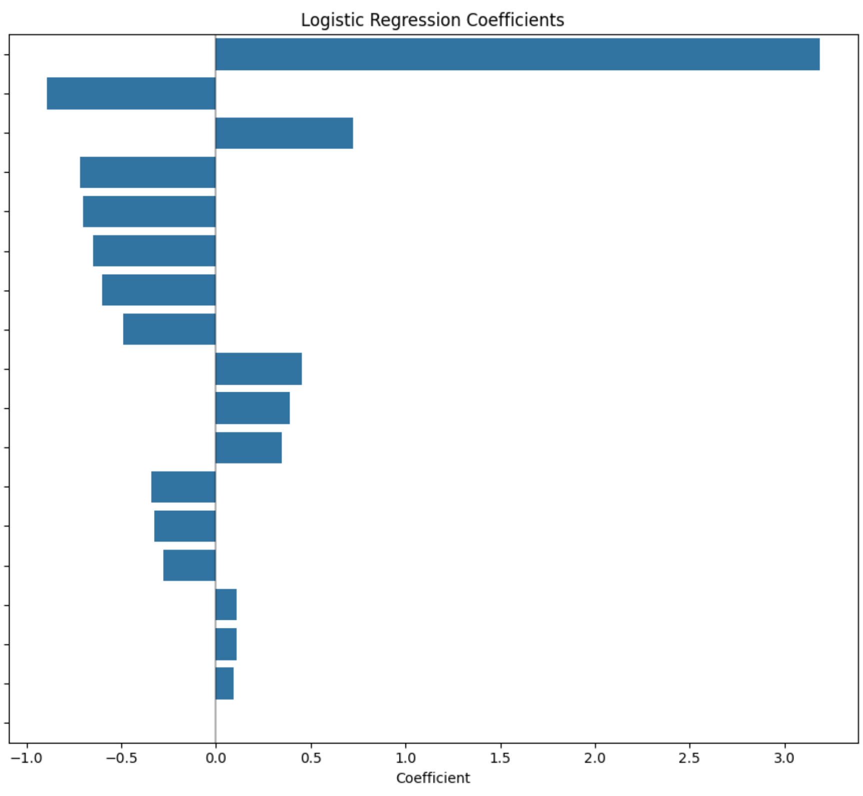
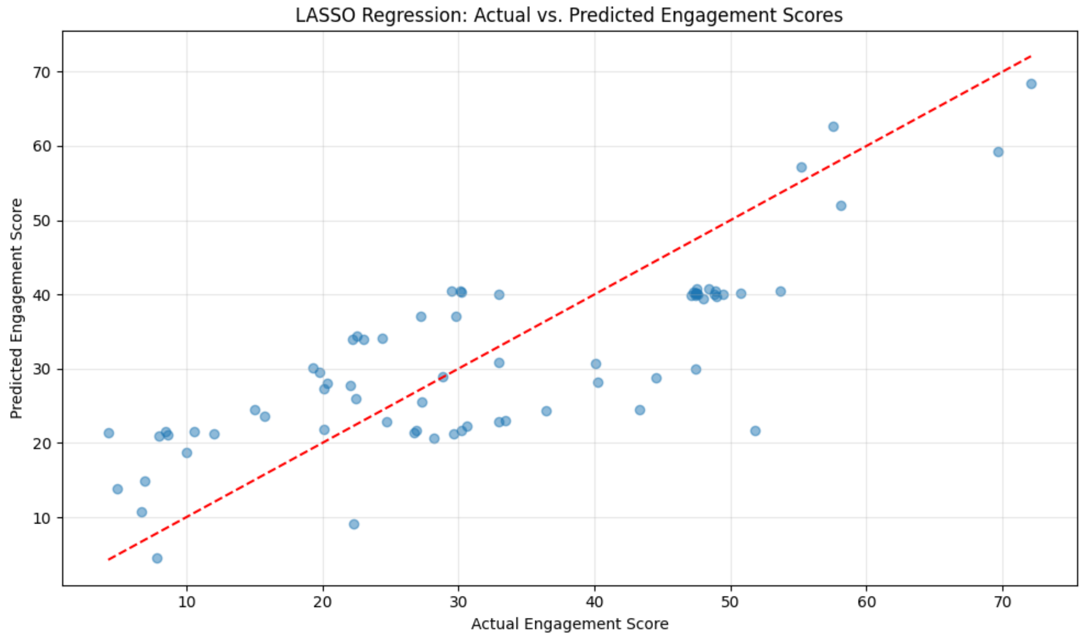
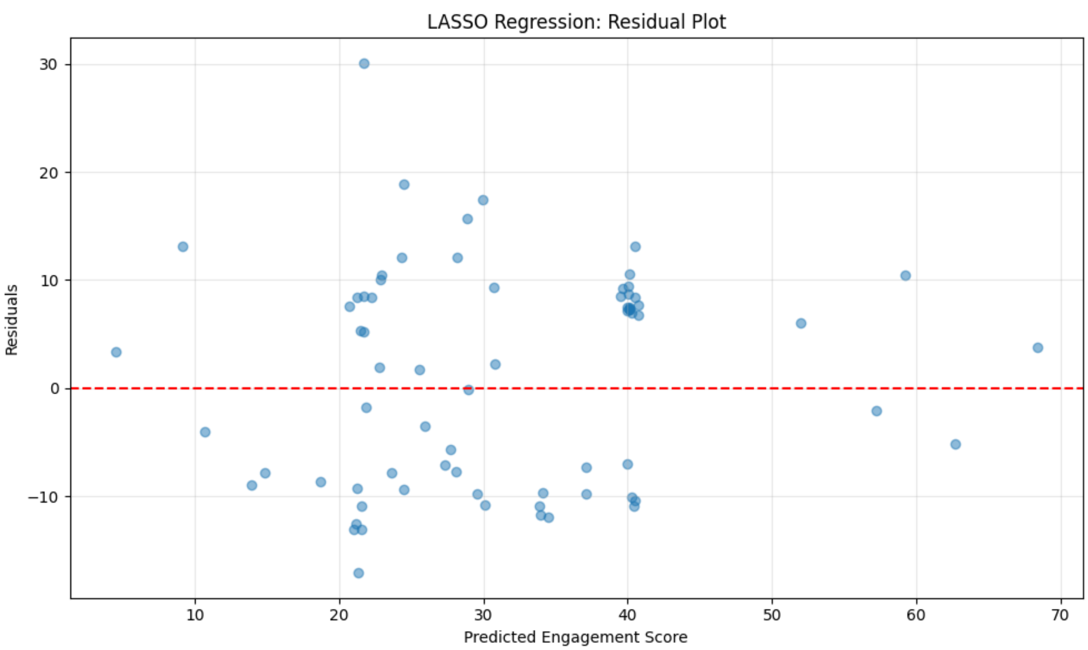
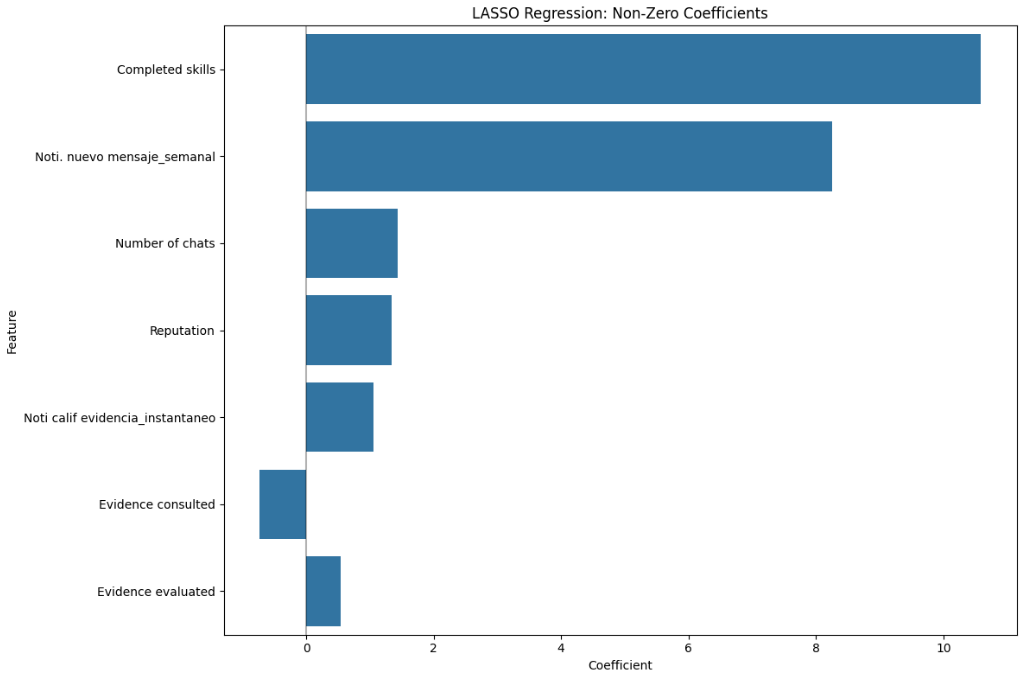
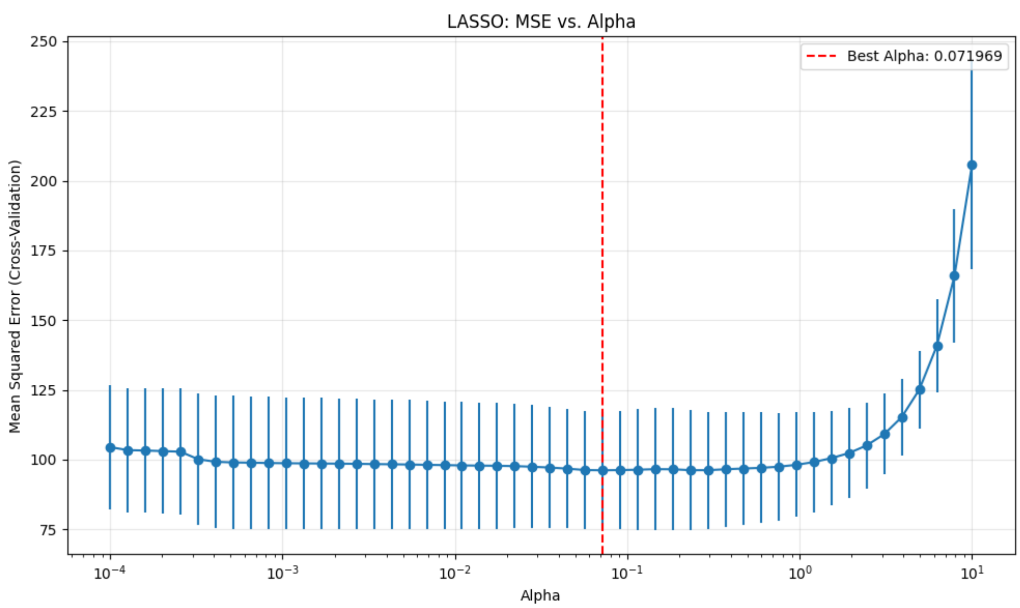


Exhibit 9: LASSO regression results without inactive users


LASSO Regression: Actual vs. Predicted Engagement Scores
Predicted Engagement Score
Actual Engagement Score
10
20
30
40
50
60
70


LASSO Regression: Residual Plot
Residuals
Predicted Engagement Score
−10
0
10
20
30
40
50
60
70


LASSO Regression: Non-Zero Coefficients
Completed skills
Noti. nuevo mensaje_semanal
Number of chats
Reputation
Noti calif evidencia_instantaneo
Evidence consulted
Evidence evaluated
Feature
0
2
4
6
8
10
Coefficient


LASSO: MSE vs. Alpha
Best Alpha: 0.071969
Mean Squared Error (Cross-Validation)
250
225
200
175
150
125
100
75
10^{-4}
10^{-3}
10^{-2}
10^{-1}
10^{0}
10^{1}
Alpha

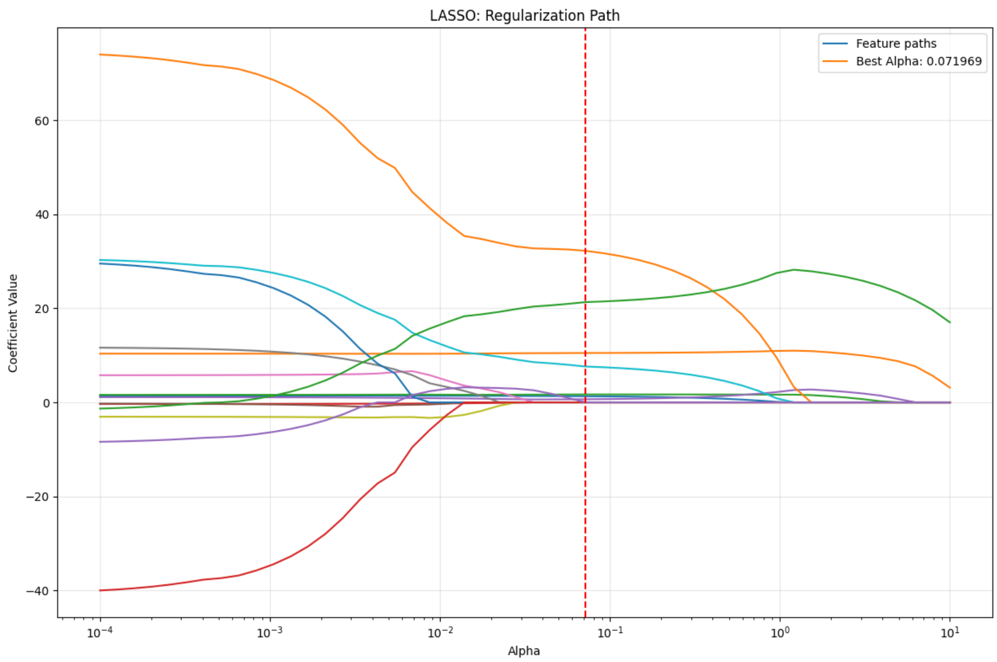
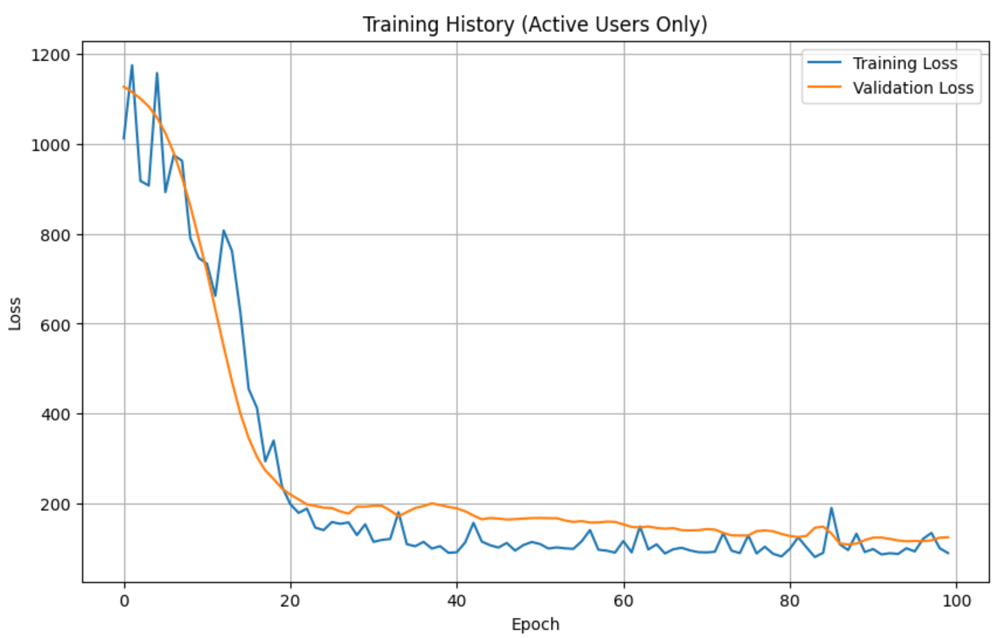
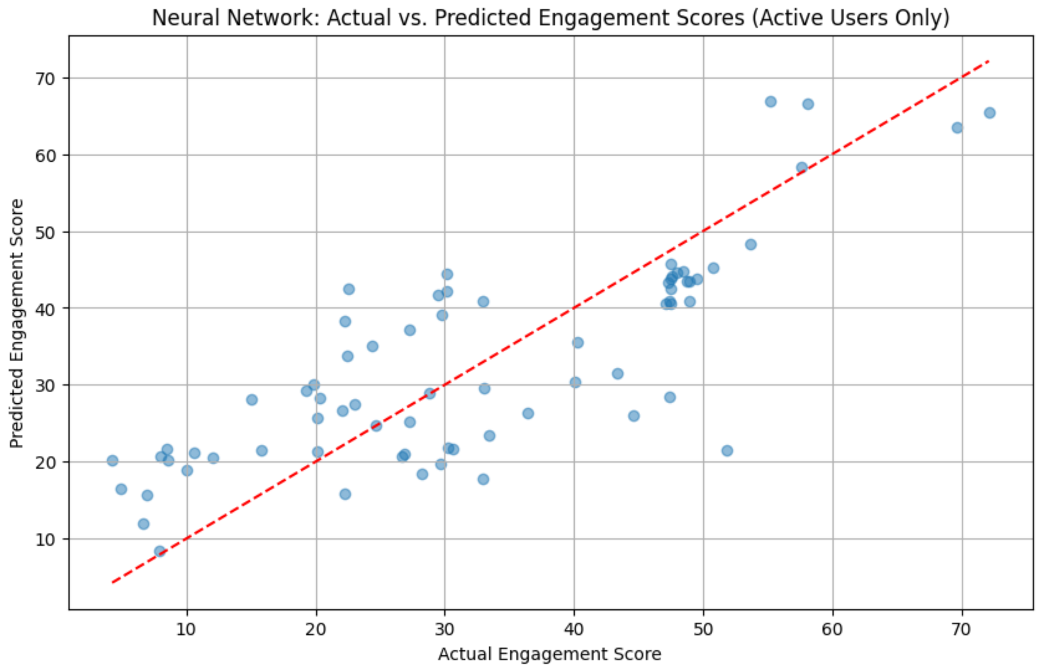
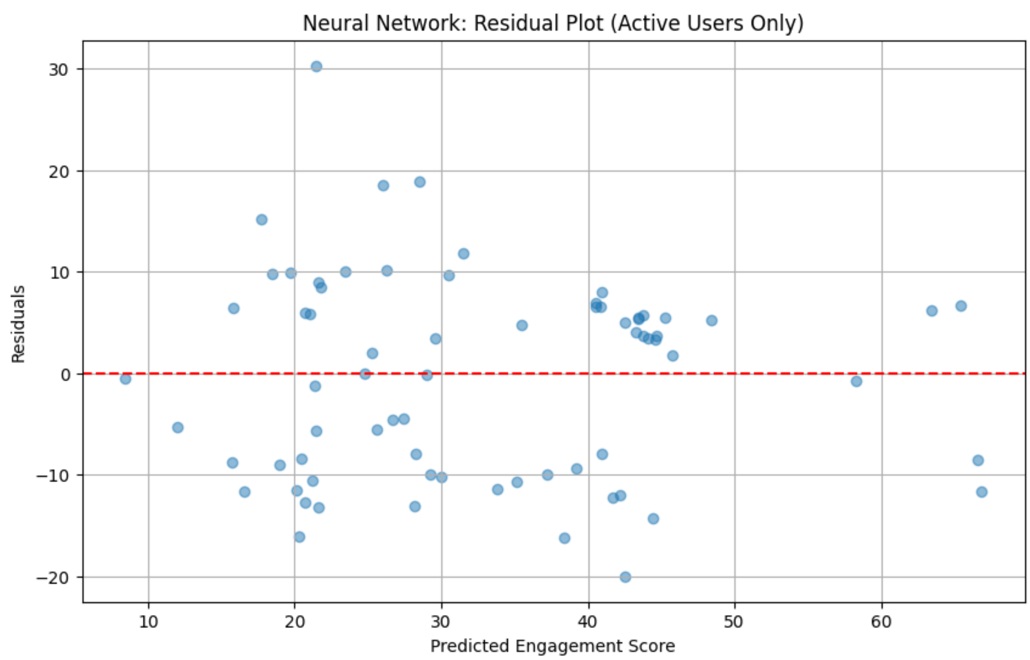


Exhibit 10: Neural Network model results without inactive users




Neural Network: Actual vs. Predicted Engagement Scores (Active Users Only)
Predicted Engagement Score
Actual Engagement Score
10
20
30
40
50
60
70


Neural Network: Residual Plot (Active Users Only)
Residuals
Predicted Engagement Score
30
20
10
0
−10
−20
10
20
30
40
50
60