HNG 13 WORKSPACE DATA ANALYSIS, MACHINE LEARNING PREDICTION & DASHBOARD REPORT

Prepared by Winner Obayomi

INTRODUCTION

The HNG 13 internship program generated an extensive amount of activity across its Slack workspace between October 15, 2025 the official start date and December 6, 2025. With thousands of members, dozens of channels, and millions of messages, the workspace provided a rich dataset capable of revealing interaction patterns, engagement levels, and predictive indicators of future behavior.

The objective of this analysis was threefold:

1. To extract and analyze Slack activity data.

2. To build a predictive machine learning model capable of identifying low-engagement or at-risk members.

3. To develop and publish an interactive Power BI dashboard that communicates insights in a clear and actionable way.

This report provides a detailed explanation of the dataset, the analytical methods used, the engineered features, the modeling approach, and the insights derived from each dashboard.

DATA OVERVIEW

The dataset contains member activity across the entire internship window (Oct 15 – Dec 6, 2025). The main components include:

- Member profile information (name, role, account type)
- Messaging activity across public channels, private channels, and DMs
- Daily active user counts
- Channel-level engagement statistics
- HNG 13 Product Leader-board analysis

The dataset ultimately contained:

- 13,000+ unique members
- 3,000,000+ messages
- 56 channels
- Daily activity logs across 53 days
- 12 products of HGN 13

KEY ENGINEERED FEATURES

To improve the predictive power of the model, several engineered features were created:

- Total_messages: total messages posted by a user
- Days_active: number of days a user posted at least one message
- Avg_messages_per_day: total messages divided by active days
- Dm_ratio, private_ratio, public_ratio: distribution of communication types
- Engagement_probability (prediction output)
- Consistency_score: messages_posted / (days_active + 1)

These features enabled the model to capture both volume and consistency of member activity.

EXPLORATORY DATA ANALYSIS (EDA)

EDA focused on trends in:

- Daily message volumes
- Daily active users
- Channel usage patterns
- Engagement distribution by Member Role, Account Type, and Activity Level
- Identification of the most active members and channels

Immediate insights included:

- A sharp spike in messages in early November, likely tied to key program activities such as several HNG products engagements.
- Engagement was heavily dominated by public channel interactions.
- A large proportion of members were inactive due to deactivation of not proceeding to the next stage.
- A small fraction of "super-engaged" members contributed the majority of total messages.

MACHINE LEARNING MODELING

Two models were trained and evaluate. Logistic Regression and Random Forest because they provide complementary strengths needed for reliable engagement prediction.

- **Logistic Regression** served as a baseline model.
  It is simple, interpretable, and statistically transparent, allowing to see the direction and weight of each behavioural factor. It also helps confirm whether the relationship between features and engagement is largely linear.
- **Random Forest Classifier** was included as a non-linear, high-performance model.
  It captures complex patterns, handles interactions between user behaviours, and is more robust to noise and irregular activity patterns. It also provides feature importance insights that help explain what drives engagement.

Comparing both using AUC ensures the final model choice is evidence-based.
Random Forest outperformed Logistic Regression, demonstrating it captured behavioural nuances better, so it was selected for the final engagement-risk predictions.

The final predictions included:

- A binary class (High/Low Engagement)
- A continuous probability score
- A risk category (High, Medium, Low)

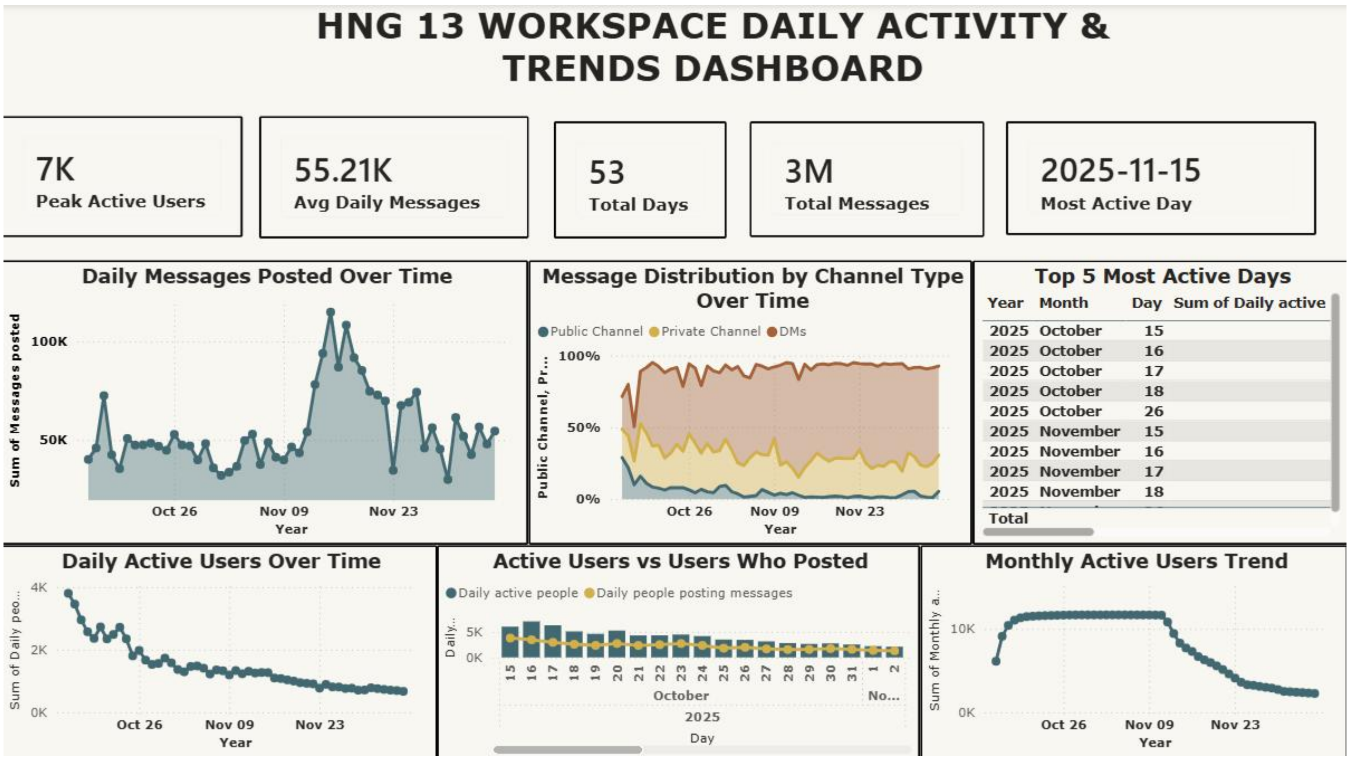
These predictions were exported to Power BI to enable visual exploration.

DASHBOARD INSIGHTS

The project resulted in five dashboard sections, each offering unique insights.

1. DAILY ACTIVITY & TRENDS DASHBOARD

This dashboard examines overall workspace activity over the 53-day period.



Key insights:

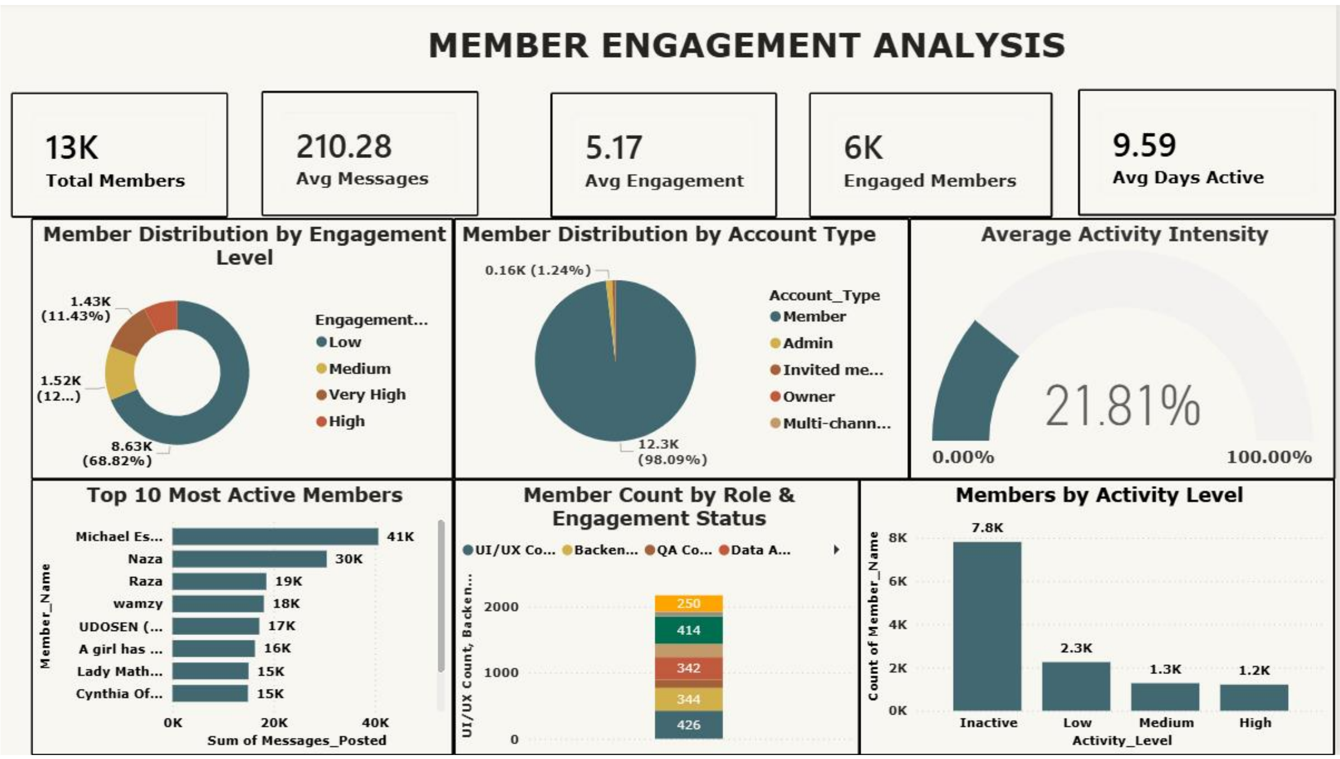
- Daily messages fluctuated between 43K and 115K peaking around November 11–15.
- DMs and private channels contributed more messages than public channels.
- Daily active users steadily declined over the internship, from over 3,000 on opening week to about 700 by December 2025.
- Top active days were clustered in late October and mid-November likely tied to major deadlines and announcements.

- Monthly active trend shows a decline in user trend which is tied to deactivation of users that failed their track task.

This dashboard provides a strong understanding of engagement momentum across the internship timeline.

2. MEMBER ENGAGEMENT ANALYSIS DASHBOARD

This dashboard focuses on member-level behaviors.



Key insights:

- Out of 13,000 members, about 6,000 engaged at least once, implying a near 50% engagement rate.

- Average messages per member were around 210.

- Engagement Levels distribution:

 - Low engagement dominated (68%)

 - Medium (12%)

- Very High (11%)

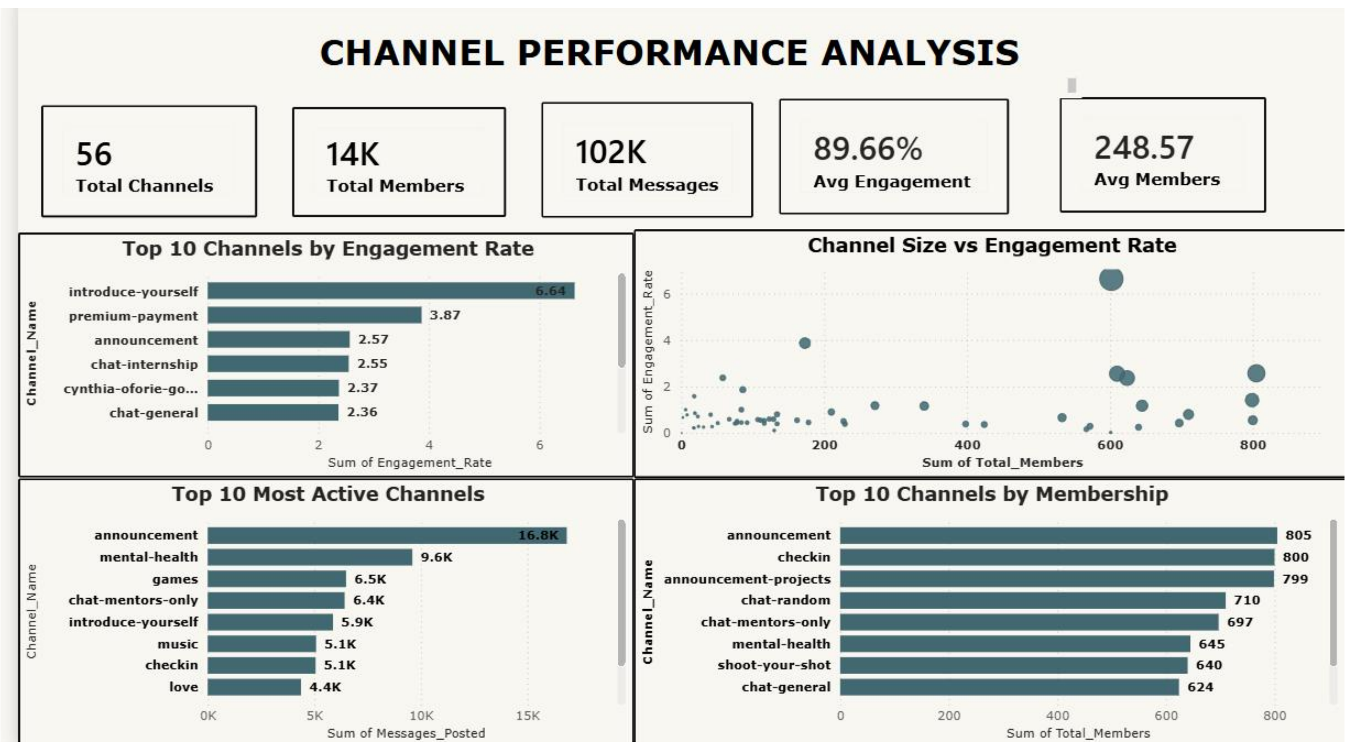
- High (8%)

- Most users used standard Member accounts; Admins and Owners represented an extremely small fraction.
- Activity intensity averaged 21.81%, indicating moderate activity among engaged users.
- Top 10 members posted exceptionally high volumes, one user exceeding 41K messages.

This dashboard highlights the “engagement inequality” common in online communities: a few members drive most of the activity.

3. CHANNEL PERFORMANCE ANALYSIS DASHBOARD

This dashboard evaluates channel-level engagement.



Key insights:

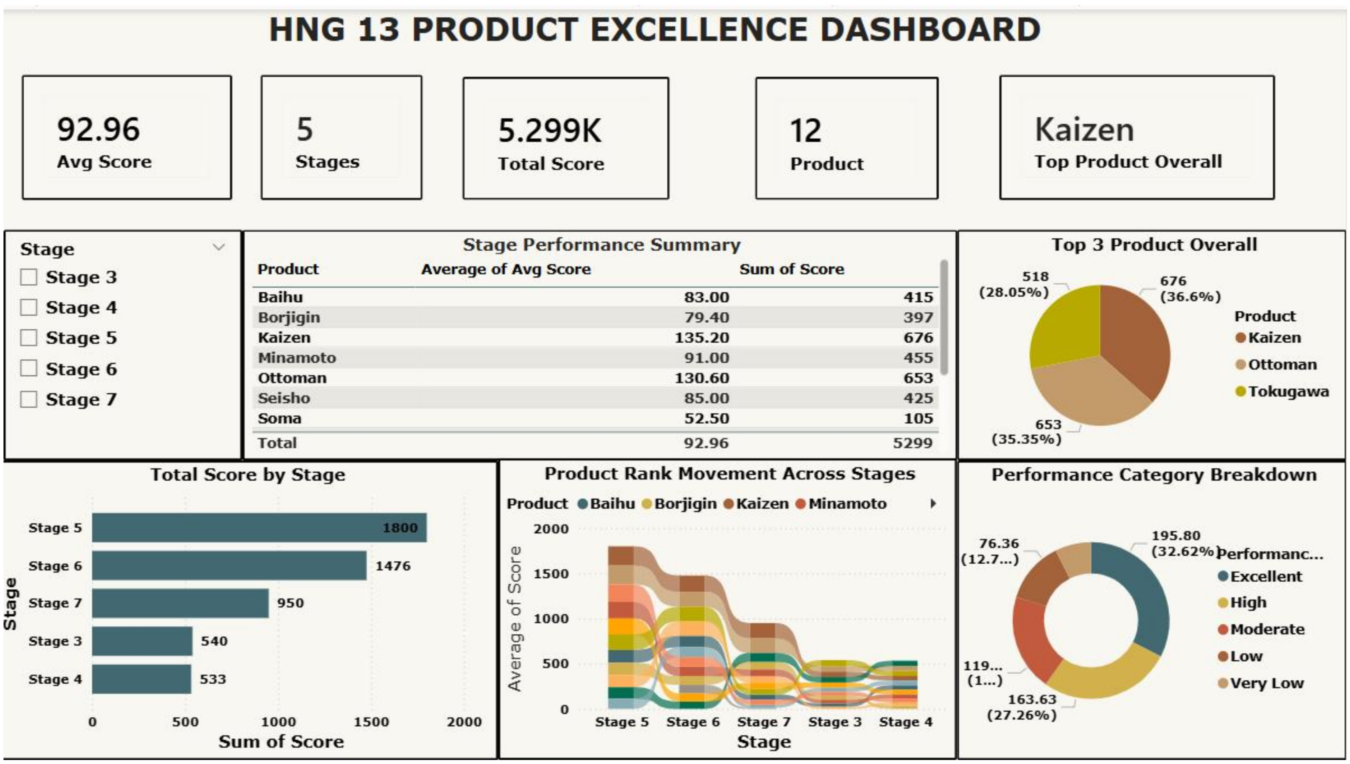
- 56 total channels, but a small subset of them drove most activity.
- **“announcement”** channel was the most active with 16.8K messages.

- Engagement Rate was highest in “introduce-yourself” and “premium-payment” channels.
- Channel Size vs. Engagement Rate scatter plot shows no strong correlation larger channels did not necessarily have higher engagement.
- The most populated channels included **“announcement”, “checkin”,** and **“announcement-projects”.**

This dashboard helps identify which channels were most effective and which required improvement.

4. PRODUCT EXCELLENCE DASHBOARD

This dashboard examines performance across HNG stages and products.



Key insights:

- There were 12 unique products evaluated across stages.
- Total program score was 5,299 with an average score of 92.96.
- Stage 5 recorded the highest total score (1,800), followed by Stage 6 (1,476).

- Product Ranked movement across stages which also compared the scores as the product progress. For instance: Kaizen had -15.36% average score change as it progress to stage 6 which means that Kaizen did better in stage 5 than stage 6 but still rank 1st.
- Top 3 products based on total scores were:

- Kaizen

- Ottoman

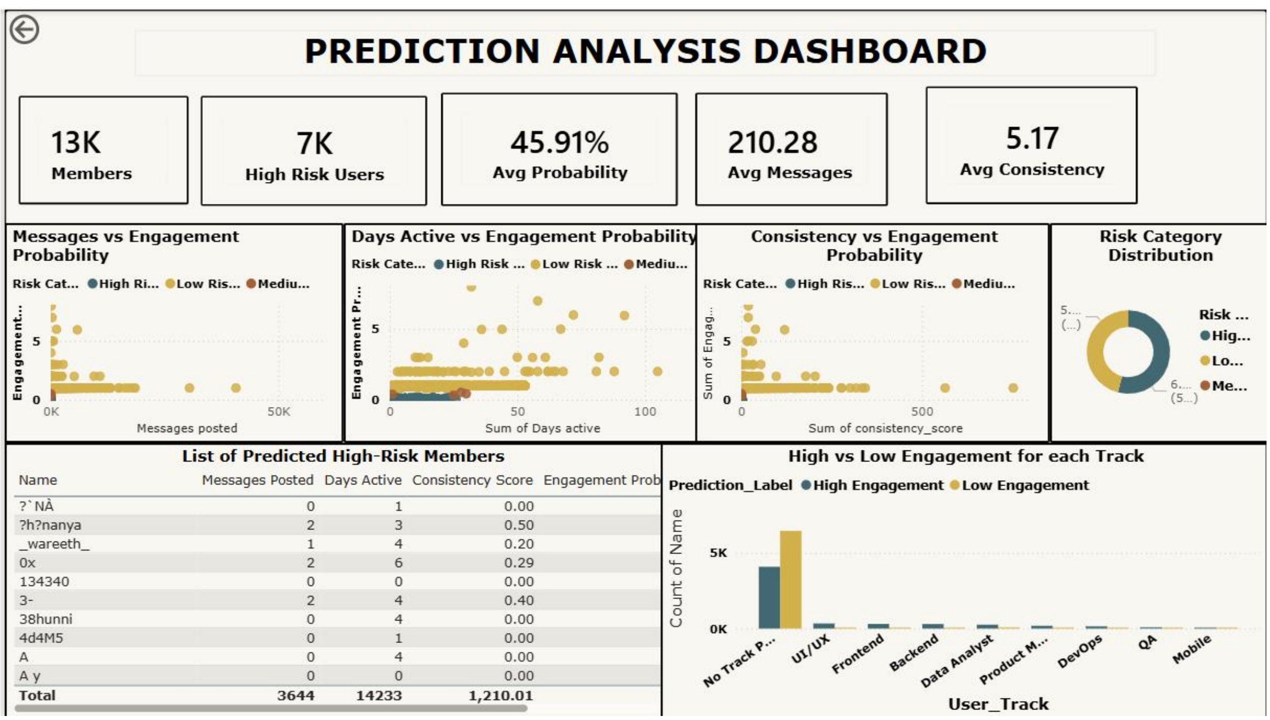
- Tokugawa

- Performance categories were distributed across Excellent, High, Moderate, Low, and Very Low, with "Excellent" being the largest category.

This dashboard illustrates how projects performed across the internship and which teams excelled.

5. PREDICTION ANALYSIS DASHBOARD

This is the core of the machine learning section.



Key insights:

- 7,000+ members were classified as high-risk (low predicted engagement).
- Average predicted engagement probability was 45.91%.
- Scatter plots illustrate relationships:

- Messages Posted vs. Engagement Probability → higher message counts correlating with higher predicted engagement.

- Days Active vs. Engagement Probability → more active days resulted in higher engagement likelihood.

- Consistency Score vs. Engagement Probability → consistent posting behavior improved predicted engagement.

- The Risk Category distribution showed a predominance of High-Risk users, confirming earlier EDA findings that many members were inactive.

- The User Track comparison shows that most tracks have more Low-Engagement users than High-Engagement users, indicating that engagement challenges cut across all roles.
- A large group of members fell under "No Track Provided," and this category had the highest count of Low-Engagement users, suggesting many users might have been deactivated. Also the "No Track Provided," are the users that did not specify their track on their slack profile.

This dashboard enables managers to spot members who may drop off so corrective steps (reminders, mentorship, nudges) can be taken.

CONCLUSION

The analysis of the HNG 13 Slack workspace provides deep insight into user engagement trends, channel performance, and program activity patterns. The machine learning model successfully predicts at-risk members using behavior-based features such as messages posted, days active, and consistency score. The dashboards provide a powerful tool for monitoring engagement, identifying patterns, and supporting data-driven decision-making.

Key conclusions:

- Engagement was strong early on but declined steadily.
- A small proportion of users contributed most activity.

- Channels varied widely in engagement, with some serving as key collaboration hubs.
- Machine learning allowed accurate identification of low-engagement members.

RECOMMENDATIONS

1. Early Engagement Onboarding:

New members should receive reminders, tutorials, and structured introductions to boost early participation.

2. Mid-Program Engagement Boosters:

Since activity drops mid-program, introducing checkpoints, competitions, or prompts could sustain engagement.

3. Use Predictive Analytics Continuously:

The model should be updated periodically to detect disengagement early, allowing intervention.

4. Improve Channel Strategy:

Channels with low engagement should be consolidated or restructured.

5. Reward High-Contributing Members:

Recognizing top contributors could encourage more balanced engagement.

Overall, the combination of analytics, machine learning, and interactive dashboards provides a robust foundation for improving the internship experience and managing large online communities effectively.