Social Media Advertisement Performance

Machine Learning Project

Ron Levi

Table of Contents

This project analyzes and predicts user interactions using the Social Media Advertisement Performance dataset[1], a synthetic yet realistic dataset simulating engagement with advertising campaigns on platforms such as Facebook and Instagram.
All data were generated with Python libraries such as Faker and NumPy, ensuring complete privacy with no personally identifiable information (PII).
The dataset follows a relational structure composed of four main tables: Users, Campaigns, Ads, and Ad Events.

Users: demographic and interest-based profiles (age, gender, country, interests).

Campaigns: strategic campaign details such as budget, duration, and start/end dates.

Ads: creative assets linked to campaigns with targeting parameters (age, gender, interests).

Ad Events: the central log of user interactions from Impression to Purchase.

The main predictive goal is to forecast the type of user event (event_type) based on demographic, campaign, and ad attributes.

What Are We Trying to Find Out?

The project aims to identify the behavioral and contextual factors influencing user engagement and to build a predictive model that classifies the type of event a user performs when exposed to an ad.
The target variable event_type represents the user's interaction type, which may include: Impression, Click, Like, Share, Comment, or Purchase.
By predicting this variable, we aim to understand what drives users from simple exposure to deeper engagement or conversion.
The model is designed to help advertisers optimize campaign strategy, deciding which campaign type to launch, which audience to target, what format to use (story, post, etc.), and when to publish (seasonally or by time of day).
Such insights can improve efficiency, reduce marketing costs, and maximize the impact and ROI of advertising campaigns.

[1] The data used in this project was obtained from the Social Media Advertisement Performance dataset available on Kaggle: https://www.kaggle.com/datasets/alperenmyung/social-media-advertisement-performance

What Are We Aiming to Achieve?

The goal is to develop a supervised machine learning model capable of accurately predicting the event_type variable.
Success is measured through two main criteria:

Model Performance: evaluated with metrics such as Accuracy, Precision, Recall, and F1-score.
Analytical Insight: identifying the most influential user, campaign, and ad-level factors affecting engagement.

Additionally, the project seeks to define the key success criteria for advertising campaigns and to identify the types of users most likely to engage, whether through clicks, shares, likes, comments, or purchases.
Ultimately, the project aims to create a practical and interpretable framework for predicting digital ad outcomes and guiding data-driven marketing strategy.

What Do We Already Know?
Users have demographic and interest-based profiles, including age, gender, country, and personal interests.
Campaigns contain strategic information such as total budget, duration, start and end dates, and targeting strategy.
Ads include creative and placement details like format, platform, and audience targeting parameters.
Ad events record real user interactions ranging from Impressions to Purchases, reflecting different levels of engagement.

Prior research and marketing analytics indicate that user engagement is influenced by several recurring factors: demographic relevance, timing, creative quality, frequency of exposure, and content personalization.
These known relationships serve as the foundation for this project, guiding the construction of features and the interpretation of the predictive model's results.

What Factors Affect Our Results?

Several factors directly influence the model's predictive performance and interpretation:

- **Data Imbalance:** The dataset is highly skewed toward Impression events (≈85%), causing the model to prioritize majority-class predictions and lowering precision for minority events.

- **User Behavior Patterns:** Engagement varies primarily by age and gender: younger users and female audiences demonstrate higher interaction rates.
- **Temporal Factors:** Purchases and active engagements tend to occur in the early morning and late evening, suggesting time-of-day influences on conversion behavior.
- **Interest Alignment:** Campaigns that align closely with user interests (e.g., Health, Finance, Gaming) produce stronger engagement, confirming that targeting precision drives better results.

- **Feature Importance:** Factors such as campaign duration, daily budget, and platform type were consistently among the most influential predictors.

These factors combined explain why recall remained high while precision was limited, emphasizing the trade-off between identifying most positive events and avoiding false positives.

Is There Something New We Can Use?

This project revealed new insights that can enhance future digital marketing strategies:

- **Engagement Timing:** The discovery of clear temporal patterns, with peak purchase activity before and after typical work hours - can guide scheduling for future campaigns.

- **Targeting Strategy:** The user- campaign interest overlap analysis highlighted which content categories are most and least effective, offering actionable direction for campaign optimization.

- **Model Optimization Insights:** The threshold calibration process demonstrated how small probability adjustments can dramatically shift recall and precision, providing a practical lesson for real-world model deployment.

- **Data-Driven Marketing:** The combination of behavioral, demographic, and temporal data created a predictive framework that can be reused to improve audience segmentation and ad delivery efficiency.

Together, these findings go beyond confirming prior knowledge - they introduce measurable, actionable patterns that connect ad design, user behavior, and engagement outcomes in a meaningful, data-driven way.

Part 1 - Data Preparation

In this stage, the main objective was to combine all data sources into a unified, high-quality dataset while ensuring accuracy, consistency, and readiness for analysis and modeling.

The process began with importing the required Python libraries and loading the four dataset files from the *Social Media Advertisement Performance* collection. These included ad_events, ads, campaigns, and users. Together, they represent the complete structure of digital advertising activity, from user demographics to campaign and event-level details.

A major focus of this phase was cleaning and validating the users table, which contained duplicate and inconsistent entries. Several targeted corrections were applied. For users with multiple entries and one record labeled “Other” in the gender column, that single row was removed to keep valid gender information.

In cases where both Male and Female entries existed for the same user, the Female record was deleted to preserve one consistent gender per user. This adjustment was made because the dataset contained a noticeably higher proportion of female users overall, and removing the female duplicates helped achieve a more balanced gender distribution. Similarly, rows containing a single inconsistent “United States” record were removed, since the majority of users originated from the United States and duplicate entries from this country were overrepresented in the data. A few individual records were manually dropped after verification. Following these cleaning steps, a validation check confirmed that each user_id appeared only once, ensuring data reliability.

Next, the datasets were merged into a single comprehensive table. The ad_events and ads tables were merged using an outer join on ad_id to include all available records. This was followed by an inner join with the campaigns table on campaign_id to retain only matching campaign data. Finally, the merged dataset was joined with the cleaned users table using an inner join on user_id, creating a complete and consistent dataset that connects user, ad, and campaign information. The unified file was saved as merged_df1_2_3_4.csv.

To support temporal analysis, all date columns (start_date, end_date, and timestamp) were converted to datetime format. From the timestamp column, new features were extracted, including year, month, month name, day, hour, minute, and second. This allowed deeper exploration of seasonal and behavioral trends over time. Additionally, a new feature named is_weekend was created, identifying whether each event occurred between Saturday and Monday. This variable is expected to help analyze user engagement patterns by day of the week.

The final prepared dataset, named df_after_prep.csv, was saved both locally and to Google Drive, ensuring accessibility for the next phases of analysis.

Overall, this phases uccessfully established a reliable and well-structured foundation for the project. All datasets were accurately merged, inconsistencies were resolved, and valuable new features were created. The resulting data is now clean, unified, and ready for Exploratory Data Analysis and model development.

Part 2 - Exploratory Data Analysis (EDA)

The purpose of this phase was to conduct a comprehensive exploration of the dataset in order to understand its structure, quality, distribution patterns, relationships between features, and potential factors influencing user engagement events.

The analysis began by loading the prepared dataset *df_after_prep.csv* from Google Drive and generating an initial metadata summary that included data types, missing values, unique value counts, and numerical ranges. This ensured a clear overview of the dataset's structure and confirmed that it was suitable for deeper analysis. The results showed that the dataset was clean, well-formatted, and contained no significant missing data.

Automated exploration using the **AutoViz** library provided an overall understanding of variable distributions, correlations, and relationships between features and the target variable event_type. Following that, a detailed manual EDA was performed using Seaborn and Matplotlib to extract key insights from both categorical and numerical variables.

A first observation concerned the **distribution of event types**, visualized through a bar chart and a pie chart. The results revealed a strong imbalance: approximately 85% of the events were *Impressions*, while all other types (Clicks, Likes, Shares, Comments, Purchases) together made up only about 15%. This imbalance highlighted the need for data balancing techniques in later modeling stages.

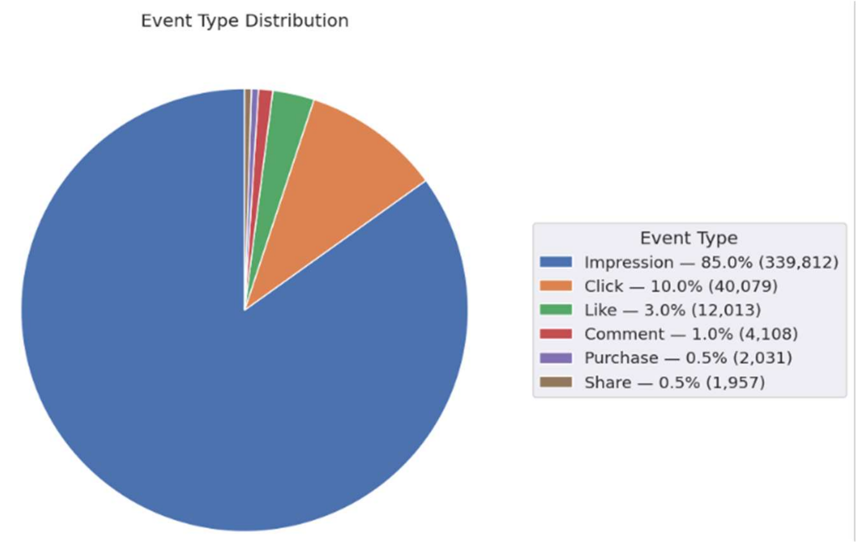


Figure 1: Distribution of Event Types - clear dominance of "Impression" events (≈85%) compared to other engagement types.

To simplify analysis, the event_type variable was encoded into two groups: 0 for Impressions and 1 for all other events.

Additional insights were uncovered from the exploratory analysis, highlighting patterns related to user activity over time and demographic characteristics.

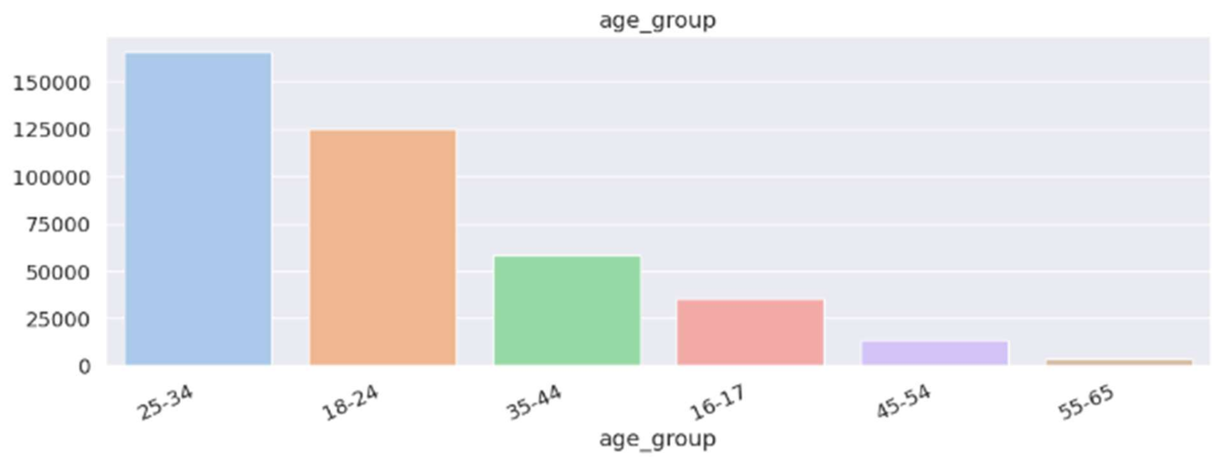


Figure 2 : User Age Distribution - right-skewed pattern showing most users are aged 18- 34

Age distribution analysis revealed that the majority of users belong to younger age groups, primarily between 18 and 34 years old. This demographic segment represents the most responsive audience to online advertising and is typically associated with higher engagement rates across social media platforms. Younger users are also more likely to interact actively with visual and interactive content such as stories, short videos, and targeted ad campaigns.

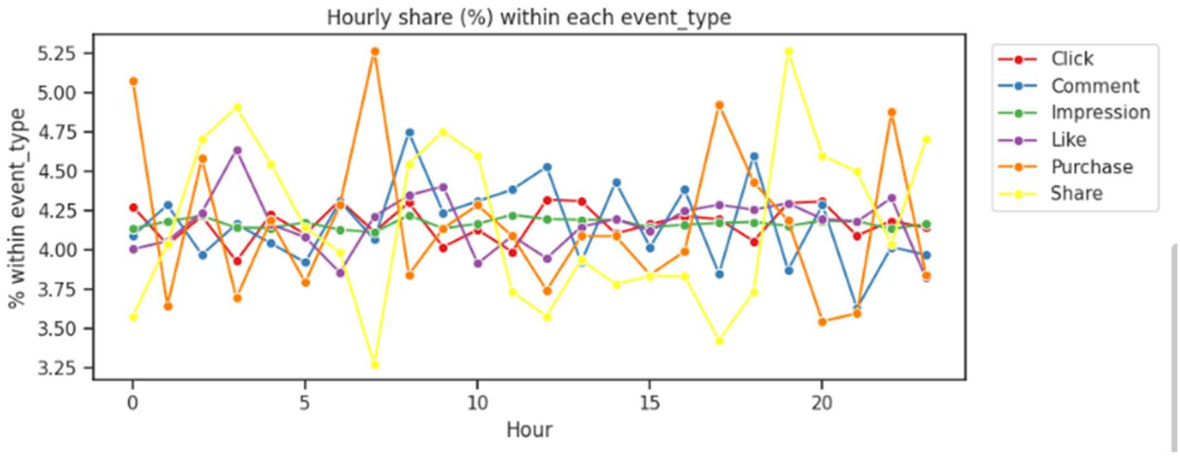
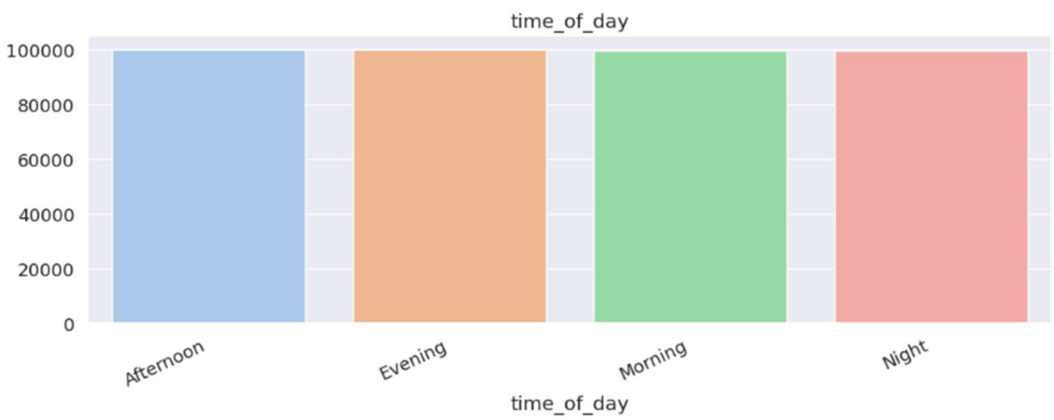


Figure 3: Hourly Engagement Pattern - higher activity observed during early mornings and late evenings

The hourly engagement chart (Hourly Share % within each event_type) revealed a clear temporal pattern in purchasing behavior. Purchases tended to occur primarily during early morning hours (around 6- 7 AM) and late in the evening (from approximately 9 PM until midnight). This suggests that users are more likely to make purchase-related interactions during periods associated with the beginning or end of their day times when people are typically more relaxed and exposed to social media before or after work.



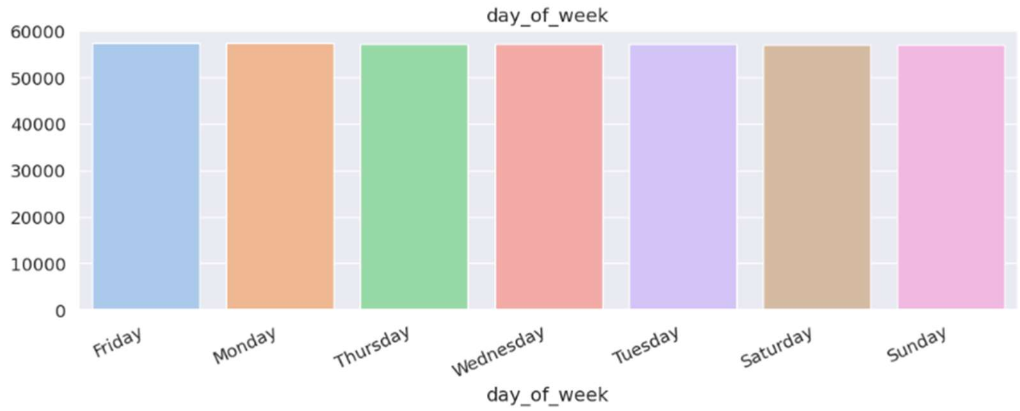


Figure 4: Distribution of User Activity by Time of Day and Day of Week

When analyzing engagement by both time of day and day of the week, no substantial variations were observed across the different temporal categories. Engagement levels remained relatively consistent throughout the week, with similar patterns across mornings, afternoons, evenings, and nights. This uniform distribution indicates that user activity and responsiveness are not strongly dependent on the specific day or time, suggesting that exposure to digital ads maintains steady effectiveness across the entire week.

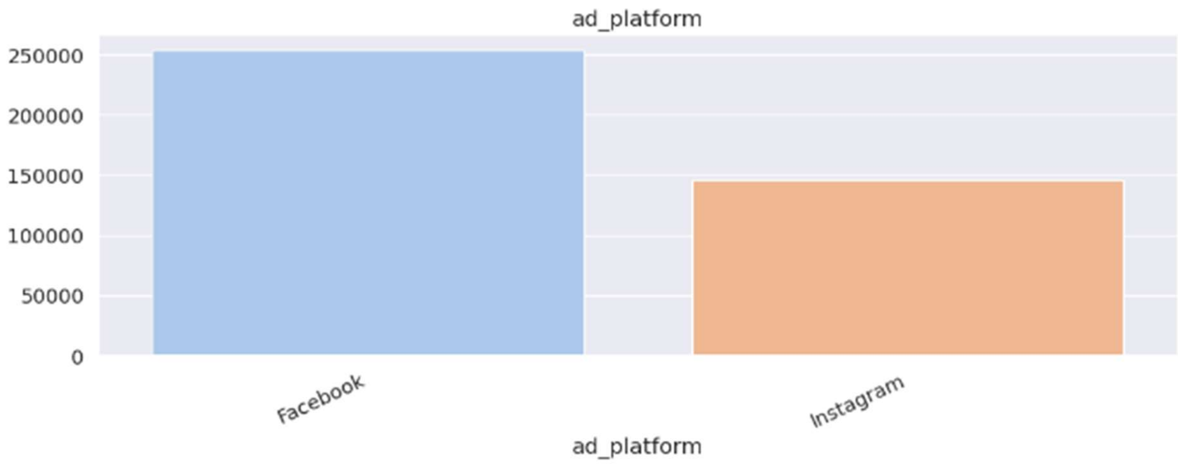
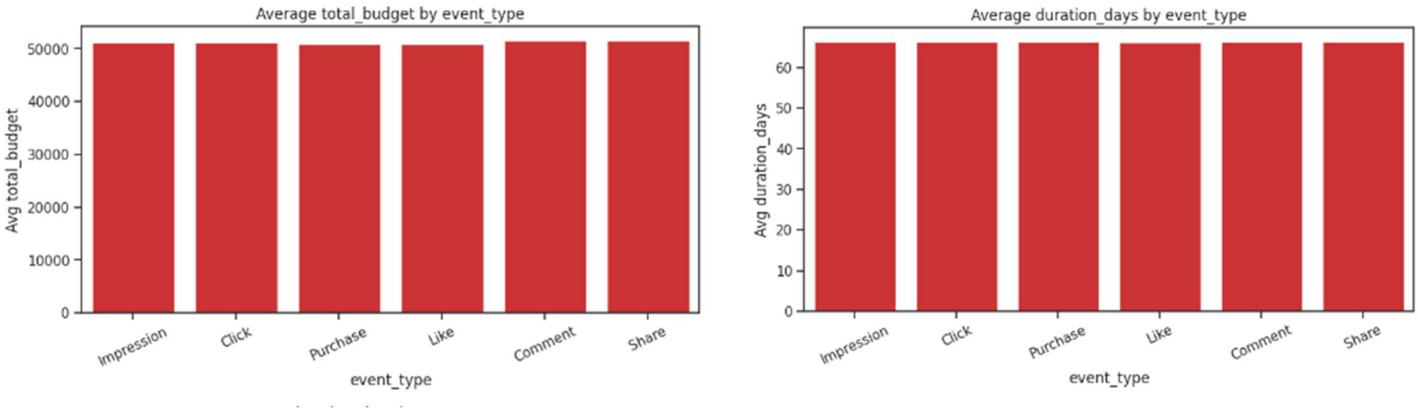


Figure 5: Distribution of Advertisements by Platform

From a demographic perspective, there was a noticeably higher proportion of female users participating in ad-related events compared to male users. This imbalance is consistent with broader social media engagement trends, where female audiences tend to demonstrate higher interaction levels with digital advertising content.

These findings emphasize the combined influence of temporal and demographic factors on user engagement behavior and will serve as valuable context for developing targeted and data-driven features in the next modeling phase .



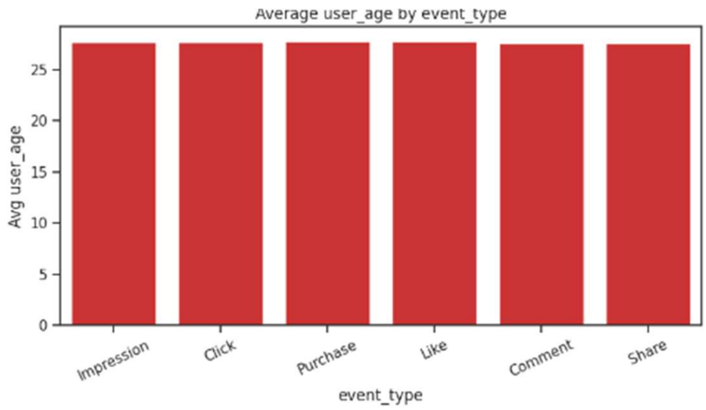


Figure 6: Average Campaign Attributes by Event Type

Next, visualizations were used to explore the relationships between **numerical variables** (user_age, total_budget, and duration_days) and the target variable. Bar charts and boxplots showed no distinct pattern in how these numerical attributes differed across event types, suggesting that engagement behavior is not primarily driven by these continuous variables. The **user age distribution** was found to be right-skewed, with most users being relatively young and fewer older participants.

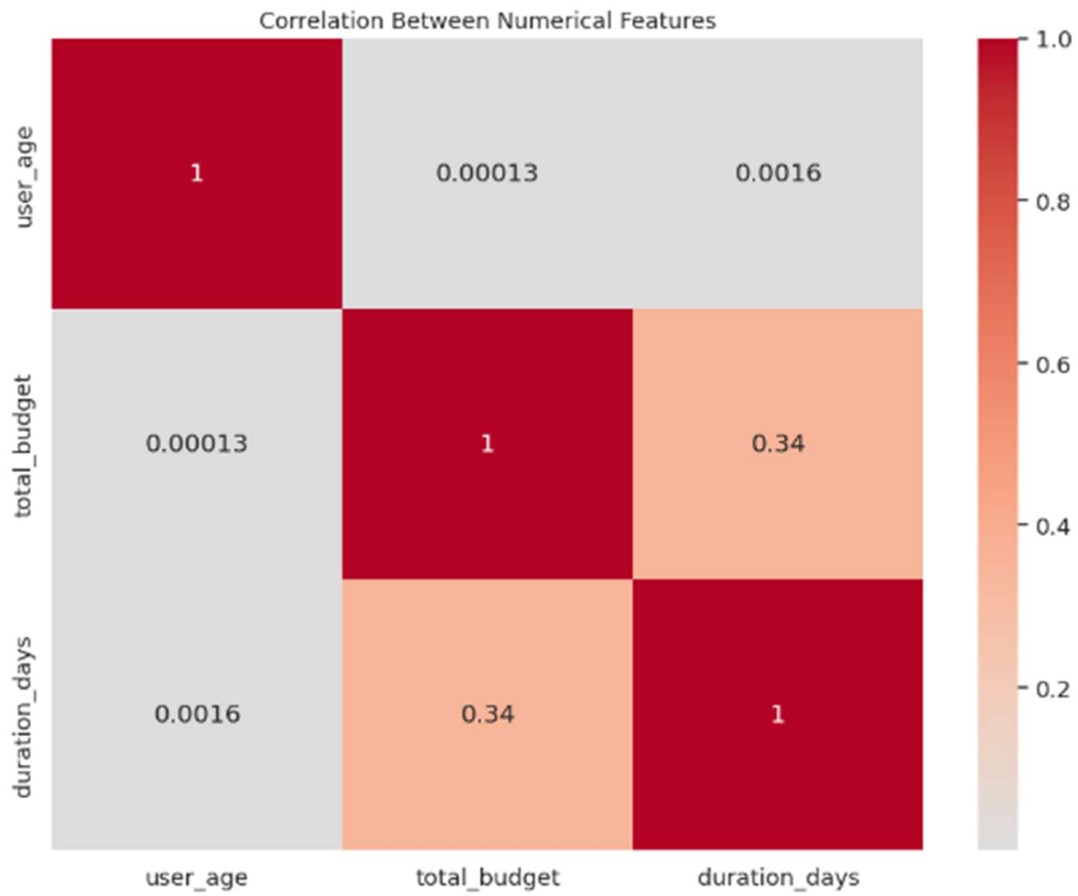


Figure 7: Correlation Heatmap - showing near-zero correlations between numerical features and event type.

Correlation analysis confirmed that all numerical features had **very weak linear relationships** with event_type. The correlation heatmap indicated near-zero values, meaning that the event outcome is likely influenced by categorical or contextual variables such as ad characteristics, platform, or timing, rather than numeric measures like age or budget.

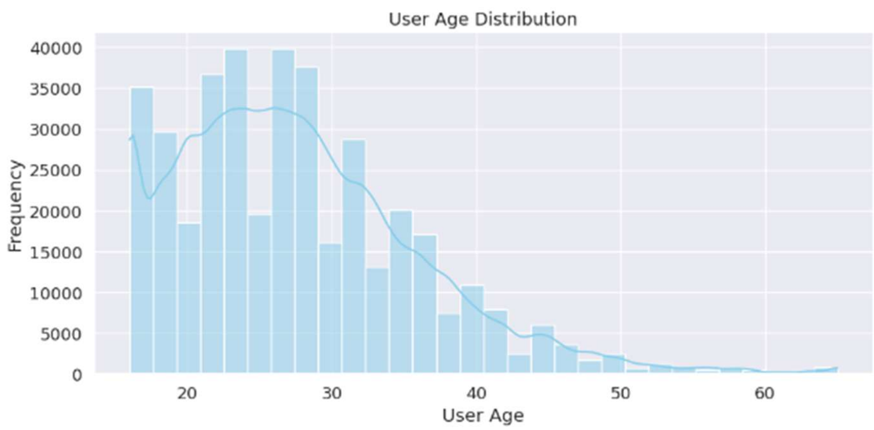
Further statistical tests were performed to validate these observations. Independent samples **T-tests** were conducted for each numeric variable to compare mean values between the two event groups. None of the tests showed statistically significant differences, reinforcing the earlier finding that numeric variables do not meaningfully separate event types.

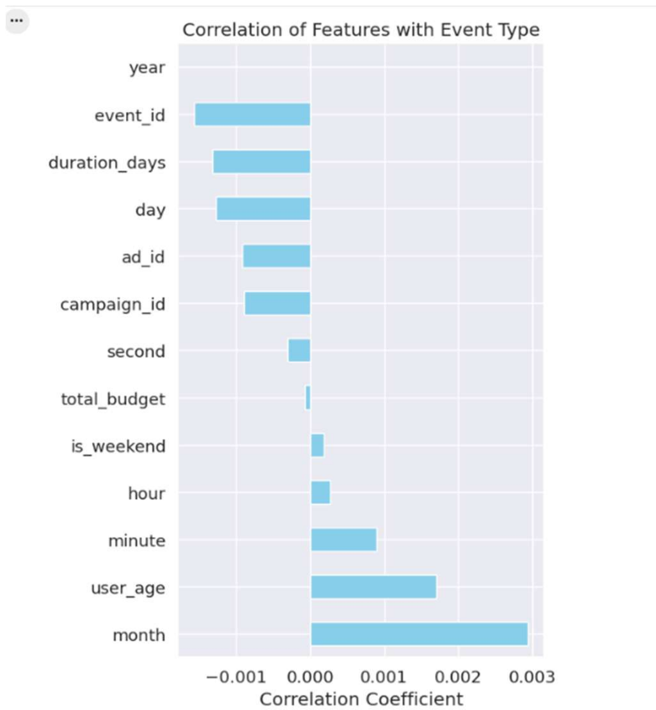
Similarly, **Chi-Square tests of independence** were applied to key categorical variables including age_group, ad_type, target_gender, day_of_week, and time_of_day. The results revealed no statistically significant associations between these features and event_type, suggesting that engagement behavior in this dataset is relatively uniform across demographic and temporal categories.

Additional time-based visualizations provided insight into hourly engagement trends. The **Hourly Activity Line Chart** showed variation in user activity throughout the day, with noticeable peaks in the morning and evening hours depending on the event type. Moreover, **interest-based heatmaps** illustrated average engagement levels across both user interests (such as fashion, gaming, health) and targeted interests defined by campaigns. These visualizations highlighted differences in how certain topics resonated more with active engagement events (Likes, Shares, Purchases) compared to passive ones (Impressions).

Finally, skewness was analyzed for all numeric features. Most variables displayed mild to moderate skewness, but not to an extreme that would require immediate transformation. This analysis confirmed that the data distribution was acceptable for subsequent modeling, while indicating that user_age had a slightly right-skewed distribution, consistent with the younger audience observed earlier.

The histogram below illustrates the distribution of user ages across the dataset. The curve shows a clear right-skewed pattern, with the highest concentration of users between the ages of 18 and 34.
This visualization confirms that the audience is predominantly young, consistent with the earlier observation that younger users are the most active and responsive





group within social media advertising campaigns.

Figure 8: Distribution and Correlation Analysis of Numerical Features

The bar chart shows the correlation coefficients between all features and the target variable event_type.
All correlation values are very close to zero, indicating that no single numeric feature has a strong linear relationship with event type.
This reinforces the conclusion that engagement behavior is likely influenced by categorical or contextual variables rather than by numerical ones.

Part 3 - Data Cleansing

At this stage, a comprehensive data validation process was conducted to ensure that the dataset was complete, accurate, and ready for further modeling. Since the dataset had already undergone multiple preparation and merging steps in earlier stages, the data was found to be generally clean, consistent, and free from major quality issues. Therefore, only validation and verification procedures were required rather than extensive cleaning.

The process began by checking for missing values across all columns. The results showed **zero missing entries**, confirming that all features were fully populated and no imputation or record removal was necessary. This indicated that the dataset was complete and structurally sound.

Next, a duplicate check was performed to identify any repeated rows that could distort model training. The analysis revealed **no duplicate records**, confirming that each observation was unique and correctly represented.

Statistical validation of potential outliers was then performed using the **Z-score method** on numerical features (user_age, total_budget, and duration_days). While a few observations were flagged as outliers (notably in user_age), a closer inspection revealed that these were legitimate data points reflecting natural variation within the population rather than true anomalies. Consequently, **no records were removed or altered**, and the dataset was deemed valid for modeling.

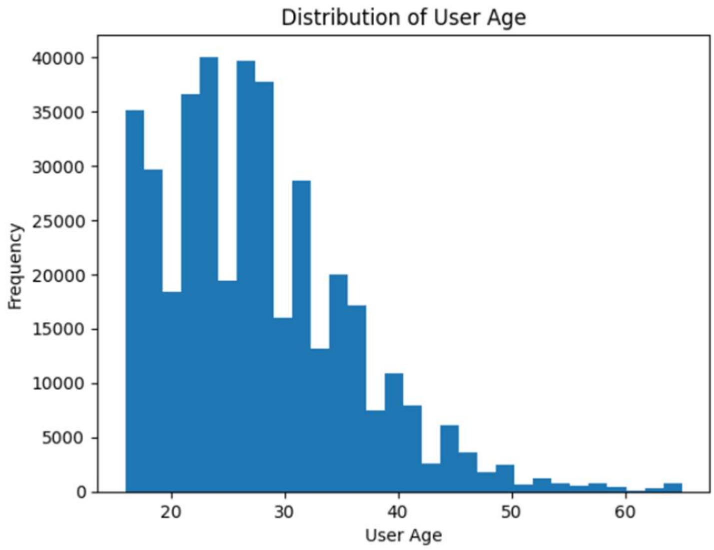


Figure 9:Distribution of User Age

Overall, this phase confirmed that the dataset was already well-structured, balanced, and reliable. The minimal presence of irregularities reflects the robustness of the data preparation process, allowing the focus to shift from cleaning to analytical modeling in the next stage.

Part 4 - One-Hot Encoding and Feature Transformation

This phase focused on encoding categorical variables into a machine-readable numerical format suitable for modeling. Since most machine learning algorithms require numerical input, all categorical features were systematically converted into encoded values, ensuring that no information was lost while maintaining interpretability.

The process began with loading the prepared dataset (df_after_eventtype_encoding.csv) from Google Drive. Using **Label Encoding**, categorical columns such as day_of_week, time_of_day, ad_platform, ad_type, target_gender, location, target_age_group, age_group, user_gender, and month_name were transformed into numeric representations. Each unique category within these variables was assigned an integer label, producing corresponding encoded columns (e.g., ad_type_encoded, user_gender_encoded, etc.). After encoding, the original categorical columns were dropped to prevent redundancy.

Following this, additional feature transformations were performed on the *interest-related* variables to enhance their analytical and predictive usefulness. The interests and target_interests columns originally contained comma-separated lists of multiple topics. To make this information usable for machine learning, each individual interest (e.g., *fashion*, *fitness*, *travel*, *technology*) was transformed into a separate binary column. Each new column takes the value 1 if the user or campaign includes that interest, and 0 otherwise.
For example, if a user's interest list contains "fashion, art, travel," the columns user_interest_fashion, user_interest_art, and user_interest_travel receive the value 1, while all others are set to 0. The same logic was applied to the campaign's target_interests, generating variables such as target_interests_fashion, target_interests_health, and target_interests_food.

This transformation significantly increased the feature space and allowed the model to analyze relationships between specific interests and engagement types. It also made it possible to measure alignment between user interests and campaign targeting an important factor in understanding engagement behavior.

Finally, the fully encoded dataset was saved as **df_after_encoding.csv**, ensuring a clean, numeric, and model-ready dataset for subsequent training and evaluation phases.

Overall, this encoding phase ensured that all categorical and multi-label fields were converted into a structured numerical format without information loss, providing the foundation for robust and interpretable modeling.

Part 5 - Feature Engineering & Feature Selection

This phase focused on enhancing the dataset through targeted **feature engineering** and **selecting** the most informative predictors to improve model performance and interpretability.

Feature Engineering

A new feature named **daily_budget** was created to represent the average daily investment per campaign.

This variable was derived by dividing each campaign's total budget by its duration in days, providing a more interpretable measure of campaign intensity and scale.

Infinite or missing values resulting from zero-duration campaigns were replaced with zeros to maintain numerical consistency.

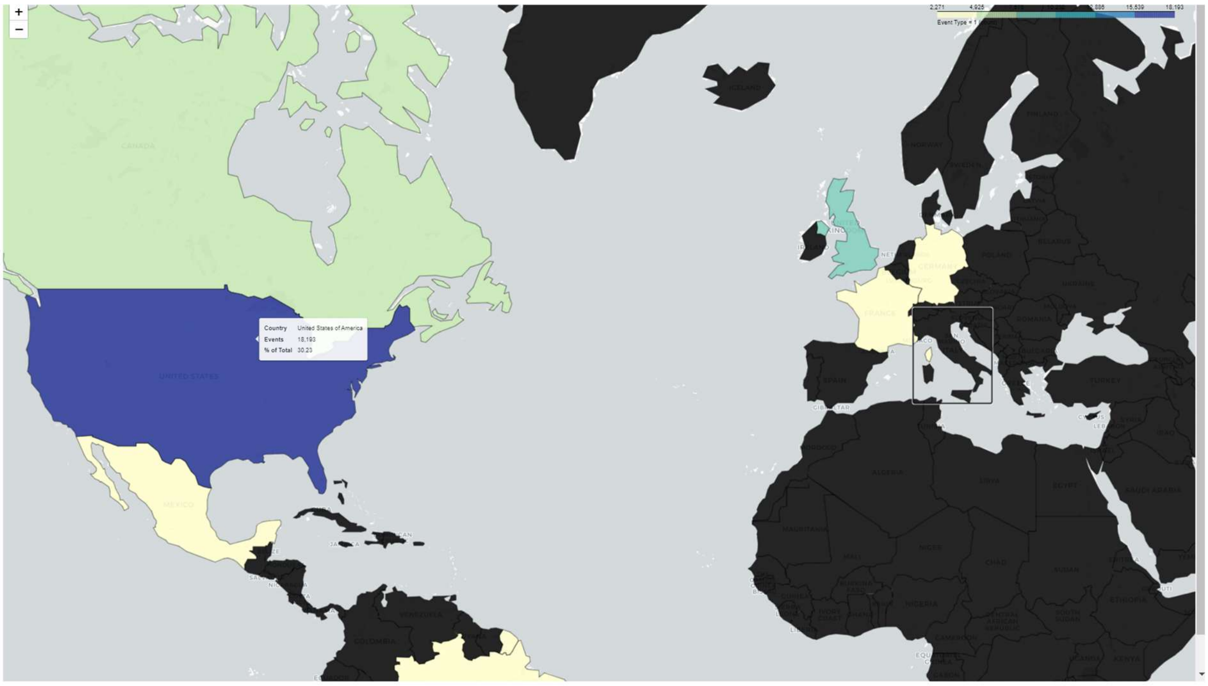


Figure 10: Global Distribution of Active Engagement Events (Event Type = 1)

To visually explore the global distribution of user engagement, a **choropleth world map** was generated, illustrating where *Event Type = 1* (active engagement events such as clicks, likes, shares, and purchases) occurred most frequently.
The visualization is **interactive**, allowing the user to hover over each country to view the exact number of active events recorded in that location.
The map revealed notable concentrations of engagement in countries such as the **United States**, **Canada**, and the **United Kingdom**, suggesting that these markets are more responsive to digital advertising campaigns.
This insight can inform future segmentation strategies and geographical targeting.

Feature Selection Process

To identify the most relevant predictors of user engagement, a comprehensive **multi-model feature selection** approach was applied.

Four models were used to evaluate feature importance: **Lasso**, **Ridge**, **Gradient Boosting**, and **Random Forest**.

Each model applied a different criterion for measuring feature contribution, from penalized regression coefficients to tree-based importance scores, allowing for a robust comparison.

The feature importance results were compiled into a unified summary table, where each feature received a score based on how many models selected it.
Features with a total score of **3 or higher** were retained as the most consistently important predictors.
After this filtering, 43 features remained.

To further improve model efficiency, non-informative attributes such as unique identifiers (IDs) and timestamp components (year, month, day, hour, etc.) were removed.

Next, an in-depth analysis of **user vs. campaign interests** was performed to assess audience alignment.

The matching analysis showed that categories such as **Health**, **Finance**, and **Gaming** had the highest overlap between user and campaign interests, indicating well-targeted audiences.
Conversely, **Food**, **Lifestyle**, and **Fitness** displayed the weakest alignment, suggesting opportunities to refine targeting in those areas.
Consequently, the low-overlap interests were excluded from the final feature set.

Finally, redundant and minimally informative variables, such as is_weekend, month_name_encoded, and overlapping demographic encodings (target_gender_encoded, target_age_group_encoded), were removed.
This manual refinement ensured that the final dataset contained approximately **23 well-defined, non-redundant features** representing the most meaningful predictors for modeling user engagement.

This phase resulted in a refined and well-structured dataset optimized for model development.
Through thoughtful feature engineering, the data gained a new explanatory variable (*daily_budget*), while the multi-model feature selection ensured that only the most consistent and meaningful predictors were retained.
The combination of statistical rigor and domain-driven filtering produced a compact yet informative dataset of 23 features: balanced, interpretable, and ready for the subsequent modeling phase .

Part 6 - Model Selection and FineTuning

This phase focused on building, evaluating, and optimizing machine learning models to predict the type of user engagement (event_type) based on the selected features from the previous stage.
Given the significant class imbalance in the target variable, where "Impression" events dominate the dataset, a careful strategy was applied to handle this imbalance and ensure fair model evaluation.

Data Splitting and Resampling

The dataset was divided into training (70%), development (15%), and test (15%) sets using stratified sampling to preserve the ratio of both classes across all splits.
To address class imbalance, SMOTE (Synthetic Minority Oversampling Technique) was applied only to the training set.
This ensured balanced learning while keeping the validation and test sets in their original unbalanced form, maintaining realistic evaluation conditions.

To further address the class imbalance problem, the **Synthetic Minority Oversampling Technique (SMOTE)** was applied to the training data.
The dataset exhibited a significant imbalance between the majority class (Impressions) and the minority class (Clicks, Likes, Shares, Purchases, and Comments).
To prevent the model from being biased toward the dominant class, SMOTE generated synthetic samples of the minority class by interpolating between existing observations, effectively increasing its representation without duplication.
This approach maintained feature diversity and allowed the classifier to better learn minority-class patterns.

The balanced training set after applying SMOTE significantly improved the model's ability to recognize positive engagement events and reduced bias toward the majority class.

Figure 11: Visualization of oversampling with SMOTE, balancing the majority and minority classes in the training data.
https://pub.towardsai.net/handling-imbalanced-datasets-in-machine-learning-smote-oversampling-undersampling-explained-6a86116c1f41

Model Benchmarking

Seven baseline models were trained and compared on the balanced training data:

- Decision Tree Classifier
- Random Forest Classifier
- XGBoost (XGBClassifier)
- Gradient Boosting Classifier (GBM)
- AdaBoost Classifier
- Logistic Regression
- Support Vector Machine (SVM)

Each model was evaluated using a consistent set of metrics: Accuracy, Precision, Recall, F1-score, Log-loss, and AUC (Area Under the ROC Curve).

The tree-based models (Decision Tree and Random Forest) achieved almost perfect performance on the training data, clearly indicating overfitting.
In contrast, XGBoost and Gradient Boosting achieved strong but realistic results with high F1 and AUC scores, while Logistic Regression and SVM performed relatively poorly.

These results identified XGBoost (primary) and Gradient Boosting (backup) as the most promising candidates for the next stage.

	Model	Accuracy	Precision	Recall	f1-score	Log-loss	AUC
0	Decision Tree	0.999416	0.999937	0.998894	0.999415	0.021062	0.999416
1	RandomForest	0.999397	0.999588	0.999205	0.999397	0.021744	0.999397
2	XGB	0.834049	0.878748	0.775039	0.823642	5.981491	0.834049
3	GBM	0.806666	0.815846	0.792133	0.803815	6.968468	0.806666
4	ADAboost	0.763457	0.736298	0.820926	0.776312	8.525872	0.763457
5	Logistic Regression	0.550421	0.548916	0.565801	0.557231	16.204476	0.550421
6	SVM	0.499626	0.499744	0.731120	0.593685	18.035313	0.499626

Figure 12: Model Benchmarking Results on Balanced Training Data

Cross-Validation for Model Robustness

A 5-Fold Stratified Cross-Validation was conducted for both XGB and GBM to measure stability and generalization.
The evaluation used four main metrics: F1, AUC, Precision, and Recall across all folds.

Results showed that XGBoost consistently outperformed Gradient Boosting, achieving:
Mean F1 ≈ 0.819
Mean AUC ≈ 0.885
Low standard deviations across folds, indicating high stability and reliability.

Therefore, XGBoost was selected as the final candidate for hyperparameter tuning.

	Model	F1_mean	F1_std	AUC_mean	AUC_std	Prec_mean	Prec_std	Rec_mean	Rec_std
0	XGB	0.819409	0.001580	0.885493	0.001012	0.874031	0.001576	0.771218	0.002616
1	GBM	0.803746	0.001556	0.869758	0.001567	0.815915	0.002358	0.791939	0.001937

Figure 13: Cross-Validation Performance Comparison - XGBoost vs. GBM

Fine-Tuning with Random and Grid Search

To further enhance performance, a **Randomized Search (RandomizedSearchCV)** was applied to optimize XGBoost hyperparameters using 5-fold cross-validation and F1-score as the primary metric.
The search explored key hyperparameters:
n_estimators (100- 400)
max_depth (3- 10)
learning_rate (0.01- 0.2)
subsample (0.7- 1.0)
colsample_bytree (0.7- 1.0)
min_child_weight (1- 6)

The best configuration achieved **F1 ≈ 0.846** in cross-validation, reflecting a noticeable improvement over the baseline model.
Evaluation on the unbalanced development set confirmed that while the model maintained strong predictive power, the recall remained low, a typical behavior for imbalanced datasets using a fixed threshold of 0.5.

To validate and refine these results, a **Grid Search** was subsequently performed on a narrower parameter range around the optimal configuration.
This final fine-tuning step confirmed the model's stability and achieved a slightly higher cross-validation F1 score, ensuring the XGBoost model was fully optimized before proceeding to threshold calibration and final evaluation.

Part 7 - Model Evaluation and Threshold Optimization

This phase focused on evaluating the final tuned XGBoost model using the development (DEV) and test (TEST) datasets.

The main goal was to identify an optimal decision threshold that balances precision and recall, especially considering the strong class imbalance in the data where the majority of events are "Impressions."

Threshold Search on Development Set

Instead of relying on the default 0.5 probability threshold, multiple cutoff values between **0.05 and 0.95** were systematically tested on the DEV set.
For each threshold, the F1-score was computed, and the threshold that maximized F1 was selected as the optimal one.

The optimal threshold was found to be **0.05**, achieving **F1 = 0.252** on the development set.
This very low threshold indicates that the model assigns relatively small predicted probabilities to the positive class and that a higher sensitivity (recall) can only be achieved by classifying many low-probability cases as positives.

Final Evaluation on Test Set

The final evaluation was performed on the unseen TEST data using the selected threshold (0.05).

The following metrics summarize the model's performance:

```
=== TEST Metrics (thresholded) ===
Accuracy:  0.2977
Precision: 0.1506
Recall:    0.7901
F1-score:  0.2529
ROC-AUC:   0.5020
Log-loss:  0.5416

Confusion Matrix (TEST):
[[10728 40244]
 [ 1895  7133]]

Classification Report (TEST):
              precision    recall  f1-score   support

           0       0.85      0.21      0.34     50972
           1       0.15      0.79      0.25      9028

    accuracy                           0.30     60000
   macro avg       0.50      0.50      0.30     60000
weighted avg       0.74      0.30      0.32     60000
```

Figure 14: Final Model Evaluation on Test Set (After Threshold Optimization)

These results indicate that while the model is effective in capturing positive engagement events (high recall of 0.79), its **precision is low (0.15)**, meaning that many of the predicted positives were incorrect.
The overall **F1-score (0.25)** and **ROC-AUC (0.50)** suggest limited discriminative ability and generalization beyond the oversampled training distribution.

The **low optimal threshold (0.05)** confirms that the model compensates for class imbalance by lowering its decision boundary to detect more positives.
This behavior is common when models are trained on oversampled data, as they

tend to overpredict the minority class when evaluated on the original unbalanced distribution.

This project successfully demonstrated how data-driven analysis and machine learning can uncover meaningful behavioral patterns in social media advertising. By combining demographic, temporal, and interest-based features, the model was able to predict user engagement levels and highlight the key factors that drive interaction.
Although precision remained limited due to class imbalance, the high recall and interpretability of results provided valuable insights into audience behavior and campaign optimization.
The analytical process established a scalable framework that can be applied to future datasets, offering a foundation for continuous improvement and smarter marketing decisions.

Deployment and Beneficiaries of Machine Learning

How will we deploy the Machine Learning system?

The trained and fine-tuned XGBoost model can be integrated as part of a **digital marketing analytics platform** or a **campaign management dashboard**.
It will operate as a predictive engine that evaluates new ad events in real time and assigns engagement likelihood scores.
These predictions can help marketing teams optimize **ad placement, timing, and audience targeting** dynamically during campaign execution.
The system can be deployed via a REST API or integrated into existing **advertisement performance monitoring systems**, allowing automated insights to be generated directly from live campaign data.

Additionally, the threshold calibration and explainability components can be included in the deployment pipeline to allow continuous model monitoring and adjustment based on live feedback.
Future iterations could also integrate **automated retraining** using updated event data, ensuring that the model adapts to changes in user behavior and platform algorithms.

Who will use and benefit from the Machine Learning

The primary users and beneficiaries of this system include:

- **Digital Marketing Analysts** - will use the model's engagement predictions to understand which audiences and time windows produce the highest response rates.
- **Campaign Managers** - can leverage the system to allocate budgets more efficiently and design better-targeted, data-driven campaigns.

- **Creative and Content Teams** - can adjust messaging and ad formats based on predicted user engagement levels.
- **Business Decision Makers** - gain access to interpretable engagement forecasts, enabling more accurate strategic planning and investment decisions.

By translating historical engagement data into predictive insights, this machine learning solution empowers teams to move from reactive campaign analysis to **proactive optimization**, leading to improved conversion rates, smarter ad spending, and higher overall marketing ROI.