Temporal Patterns in User Conversions: Investigating the Impact of Ad Scheduling in Digital Marketing

Satrya Fajri Pratama[1,*], Dwi Sugianto[2]

[1]Department of Computer Science, School of Physics, Engineering and Computer Science, University of Hertfordshire, Hatfield AL10 9AB, United Kingdom

[1,2]Magister of Computer Science, Computer Science Faculty, Universitas Amikom Purwokerto, Indonesia

ABSTRACT

This study explores the impact of ad scheduling on user conversions by analyzing temporal patterns in user behavior. In the increasingly competitive landscape of digital marketing, optimizing the timing of ad placements is critical for maximizing user engagement and conversion rates. Utilizing a comprehensive dataset from Kaggle, which includes variables such as user ID, ad exposure details, and conversion outcomes, we employed both time series analysis and survival analysis to uncover insights into how different ad scheduling strategies affect conversion rates. The ARIMA model, used for time series analysis, provided reasonable predictive accuracy with a Mean Absolute Error (MAE) of 389.92, Root Mean Squared Error (RMSE) of 463.97, and Mean Absolute Percentage Error (MAPE) of 2.26%. This model effectively identified specific hours and days with higher likelihoods of conversion, particularly during evenings and weekends. On the other hand, the Cox Proportional Hazards model, used for survival analysis, demonstrated superior performance with a concordance index of 0.97, indicating its exceptional ability to predict the timing of user conversions based on various covariates such as the number of ads seen and the specific hours of exposure. The findings suggest that strategic ad scheduling, tailored to align with user temporal behavior, can significantly enhance marketing effectiveness by targeting users during peak conversion periods. These insights offer practical implications for digital marketers aiming to refine their ad delivery strategies to achieve higher conversion rates and improve return on investment.

Keywords Ad Scheduling, User Conversions, Temporal Patterns, ARIMA, Cox Proportional Hazards, Digital Marketing, Time Series Analysis, Survival Analysis, Conversion Rates, Digital Advertising, Predictive Modeling

INTRODUCTION

Digital advertising has significantly transformed how businesses engage with their target audiences. Through digital marketing strategies, companies can create personalized messages tailored to the specific needs and preferences of their target demographics [1]. This personalized approach is facilitated by the connectivity and interactivity inherent in digital marketing, enabling businesses to engage in many-to-many communications that are timely, relevant, and cost-effective [2].

Digital advertising encompasses various strategies and platforms, from social media ads to search engine marketing, providing unparalleled opportunities for precise targeting and real-time interaction. To enhance ad delivery in the competitive digital advertising landscape, businesses should implement

Submitted 22 June 2024
Accepted 20 July 2024
Published 1 September 2024

Corresponding author
Satrya Fajri Pratama,
s.f.pratama@herts.ac.uk

Additional Information and Declarations can be found on page 180

DOI: 10.47738/jdmdc.v1i2.10

How to cite this article: S. F. Pratama and D. Sugianto, “Temporal Patterns in User Conversions: Investigating the Impact of Ad Scheduling in Digital Marketing,” *J. Digit. Mark. Digit. Curr.*, vol. 1, no. 2, pp. 165-182, 2024.

strategies that effectively capture and retain consumer attention. The transition from offline to online consumer behavior has resulted in a proliferation of digital advertisements competing for consumer attention [3]. This optimization includes determining the best times to display ads to maximize user engagement and conversions [4].

Ad scheduling, or the strategic timing of ad placements, plays a vital role in digital marketing effectiveness [5]. Properly scheduled ads can significantly increase the chances of reaching users at the moments they are most likely to engage with the content and convert into customers. The timing of ad delivery can influence a user's likelihood to interact with an ad and make a purchase, highlighting the necessity of understanding and leveraging temporal patterns in user behavior.

The primary objective of this study is to investigate the impact of ad scheduling on user conversions by analyzing temporal patterns. By employing both time series analysis and survival analysis, this research aims to uncover insights into how different ad scheduling strategies affect conversion rates. The findings of this study will provide valuable guidance for digital marketers seeking to enhance their ad scheduling tactics to improve user conversions and overall campaign performance.

This study seeks to answer several critical questions regarding the impact of ad scheduling on user conversions. Firstly, it aims to understand how different ad scheduling strategies affect user conversion rates. By examining various scheduling approaches, the study will identify which strategies are most effective in driving conversions. Secondly, the research aims to uncover the temporal patterns in user conversions. Analyzing these patterns will provide insights into the specific times and conditions under which users are most likely to convert. Understanding these patterns is essential for optimizing ad delivery and enhancing marketing effectiveness. Finally, the study will evaluate the effectiveness of different modeling approaches—specifically, time series analysis and survival analysis—in predicting user conversions. By comparing these methodologies, the research will determine which approach provides more accurate and actionable insights for optimizing ad scheduling strategies.

The significance of this study lies in its practical implications for digital marketers and advertisers, as well as its contribution to the existing literature on ad scheduling and user behavior. For practitioners in the field, the findings offer actionable insights into how ad scheduling can be optimized to maximize user conversions. By identifying the most effective times and conditions for ad delivery, marketers can enhance the efficiency and effectiveness of their campaigns, ultimately driving higher conversion rates and improving return on investment.

Additionally, this study contributes to the academic discourse by providing empirical evidence on the impact of ad scheduling on user conversions. It extends the current understanding of temporal patterns in user behavior and the effectiveness of different modeling approaches, such as time series analysis and survival analysis, in predicting conversions. This research fills a gap in the literature by offering a comprehensive analysis of how scheduling strategies influence user engagement and conversion outcomes, thereby informing future studies and advancing knowledge in the field of digital marketing.

Literature Review

Digital Advertising and User Behavior

Digital advertising has become a cornerstone of modern marketing strategies, leveraging the internet's vast reach to connect with potential customers in ways that were previously unimaginable. The landscape of digital advertising encompasses a wide array of strategies, including search engine marketing (SEM), social media advertising, display advertising, email marketing, and influencer partnerships, among others. Each of these strategies offers unique benefits and challenges, allowing businesses to tailor their marketing efforts to specific audiences and goals.

Search engine marketing, for instance, involves placing ads on search engine results pages (SERPs) to capture user interest at the moment of search intent. Social media advertising utilizes platforms like Facebook, Instagram, Twitter, and LinkedIn to engage users through targeted ads that appear in their social feeds. Display advertising involves placing banner ads on websites, leveraging visuals to attract attention and drive traffic. Email marketing, a more direct approach, involves sending promotional messages directly to users' inboxes, fostering a personalized connection. Influencer partnerships leverage the credibility and reach of social media influencers to promote products authentically to their followers.

Several factors influence user behavior and conversions in digital advertising. First and foremost is the relevance of the ad content to the user [6]. Highly targeted ads that align with user interests and needs are more likely to capture attention and drive engagement. The quality and design of the ad itself also play crucial roles; visually appealing ads with clear calls to action can significantly enhance user response rates. Timing and frequency of ad exposure are critical factors as well [7]. Repeated exposure to an ad can build familiarity and trust, increasing the likelihood of conversion, but too much repetition can lead to ad fatigue and diminished returns. The timing of ad delivery—whether the user is exposed to the ad during peak engagement times or in moments of decision-making—can also significantly impact effectiveness.

Additionally, the platform on which the ad is displayed can influence user behavior. Users may respond differently to ads on social media compared to search engines or email, depending on the context in which they encounter the ad. The device used to view the ad (e.g., mobile, desktop, tablet) can also affect user engagement and conversion rates, as mobile users may exhibit different browsing and purchasing behaviors compared to desktop users.

User behavior is also influenced by external factors such as seasonality, economic conditions, and cultural trends. Seasonal campaigns that align with holidays or special events can see higher engagement due to the timely relevance of the ads. Economic conditions, such as consumer confidence and disposable income levels, can affect purchasing behavior and response to promotional offers. Cultural trends and social issues can also shape user attitudes and responses to advertising. Understanding these factors is essential for designing effective digital advertising campaigns. By leveraging insights into user behavior and preferences, marketers can create targeted, engaging ads that resonate with their audience and drive higher conversion rates. This understanding also forms the basis for optimizing ad scheduling strategies, ensuring that ads are delivered at the most opportune times to maximize impact and efficiency.

Ad Scheduling

Ad scheduling, also known as dayparting, refers to the strategic timing of ad placements to optimize their effectiveness. The importance of ad scheduling in digital marketing cannot be overstated, as it directly influences the likelihood of user engagement and conversions. By carefully planning when ads are displayed, marketers can target audiences at the times they are most receptive, thereby maximizing the return on investment (ROI) for their advertising campaigns. The primary goal of ad scheduling is to ensure that advertisements are delivered to the right audience at the right time. This involves analyzing user behavior patterns to identify peak engagement periods [8]. For example, e-commerce platforms may find that users are more likely to browse and make purchases during evenings and weekends, while B2B services might see higher engagement during weekday working hours. By scheduling ads to coincide with these periods, marketers can increase the visibility and impact of their campaigns.

Several strategies are employed in ad scheduling to enhance ad performance. One common approach is to use historical data and analytics to determine the optimal times for ad delivery. Marketers can analyze metrics such as click-through rates (CTR), conversion rates, and user activity patterns to identify the times when users are most likely to interact with ads. This data-driven approach ensures that ads are shown when they are most likely to achieve the desired outcomes.

Another strategy is to use real-time bidding (RTB) platforms that allow advertisers to bid for ad placements in real-time. RTB platforms enable advertisers to dynamically adjust their bids based on the time of day, ensuring that ads are delivered during high-value periods. This approach not only optimizes ad spending but also enhances the relevance and effectiveness of the ads. Automated ad scheduling tools, provided by platforms like Google Ads and Facebook Ads Manager, offer sophisticated options for dayparting. These tools allow advertisers to set specific times and days for ad delivery, automate bid adjustments, and track performance metrics. By leveraging these tools, marketers can implement precise ad scheduling strategies without manual intervention, ensuring consistent and optimized ad performance.

Temporal Patterns in User Behavior

Temporal patterns in user behavior refer to the trends and cycles that characterize how users interact with online content over time. Understanding these patterns is crucial for optimizing digital marketing strategies, particularly ad scheduling. By analyzing when users are most active and receptive, marketers can tailor their campaigns to coincide with peak engagement periods, thereby maximizing the effectiveness of their advertising efforts. Temporal patterns in user behavior are crucial for recommendation systems as they capture evolving preferences and behaviors [9]. These patterns encompass user preference drifts, seasonal effects, and fluctuations in user and item activity rates [10].

Users' online activities often follow predictable temporal patterns influenced by various factors such as daily routines, work schedules, and seasonal trends. For instance, internet usage typically peaks during certain times of the day, such as morning hours when people check their emails and social media, lunchtime breaks, and evenings when they relax and browse the internet. Weekdays and weekends also exhibit distinct usage patterns, with weekends often showing increased leisure-related browsing and weekdays reflecting more work-related searches.

Advanced analytics and big data technologies enable marketers to dissect these patterns at a granular level. Tools like Google Analytics, Facebook Insights, and custom data analytics solutions can track user interactions across different platforms and timeframes. By aggregating and analyzing this data, marketers can identify the specific times when their target audience is most active and engaged. For example, an e-commerce website might find that traffic spikes in the evening hours, suggesting that ads scheduled during this period could achieve higher engagement and conversion rates.

Numerous studies have explored the impact of ad timing on user engagement and conversion rates, revealing significant insights into how temporal factors influence advertising effectiveness. One key finding is that ad delivery timing can dramatically affect click-through rates (CTR) and conversion rates. Ads delivered during high-traffic periods are more likely to be seen and interacted with, resulting in better performance metrics.

The timing of advertisements is crucial in determining user engagement with ad content. Research has shown that excessive ad clutter can lead to a reduction in user attention and engagement [11]. Advertisers should consider factors such as ad exposure time, user characteristics, and contextual information to optimize engagement rates [12], [13]. Additionally, the quality and relevance of ads are essential for maximizing long-term user engagement [14], [15]. Studies have indicated that user engagement not only increases the effectiveness of ads but also positively impacts sales and productivity [16].

Modeling Approaches

In the field of digital marketing, understanding and predicting user behavior are essential for optimizing ad campaigns and improving conversion rates. To achieve this, various modeling approaches can be employed, each offering unique insights and advantages. This section provides an overview of two prominent modeling techniques: time series analysis with ARIMA models and survival analysis with Cox Proportional Hazards models. It also compares these approaches in the context of predicting user behavior.

Time series analysis involves the examination of data points collected or recorded at specific time intervals. This type of analysis is particularly useful for identifying patterns, trends, and seasonal effects in temporal data. In digital marketing, time series analysis can help understand how user behavior changes over time and how these changes correlate with ad scheduling and other marketing activities.

One of the most widely used models in time series analysis is the Auto-Regressive Integrated Moving Average (ARIMA) model. The ARIMA model is designed to capture the underlying patterns in time series data by combining three components: Auto-Regressive (AR), Integrated (I), and Moving Average

(MA). The Auto-Regressive component uses the dependency between an observation and a number of lagged observations, essentially leveraging past values to predict future ones. The Integrated component involves differencing the observations to make the time series stationary, ensuring that its statistical properties, like mean and variance, remain constant over time. Finally, the Moving Average component utilizes the dependency between an observation and a residual error from a moving average model applied to lagged observations. Together, these components allow the ARIMA model to effectively analyze and forecast time series data. The ARIMA model is flexible and can be adjusted to suit various types of time series data, making it a powerful tool for forecasting future user behavior based on historical trends. For example, it can predict the number of user conversions in the coming days based on past conversion patterns, helping marketers to optimize their ad scheduling strategies accordingly.

Survival analysis, originally developed for medical research, is used to analyze the time until the occurrence of an event of interest, such as user conversion in the context of digital marketing. This approach is particularly useful when dealing with time-to-event data and can provide insights into the factors that influence the timing of user conversions. The Cox Proportional Hazards model is a popular method in survival analysis. It assesses the effect of several variables on the hazard or risk of the event occurring. The model assumes that the hazard ratio between two individuals is constant over time, making it possible to evaluate the impact of different covariates (e.g., ad exposure, user demographics) on the likelihood of conversion. The Cox model provides a semi-parametric approach, meaning it does not assume a specific baseline hazard function, making it flexible for various types of data. This model can help marketers understand not only whether users convert but also when they are likely to convert, based on different influencing factors. For instance, it can reveal how the timing and frequency of ad exposures impact the probability of conversion over time.

Both ARIMA and Cox Proportional Hazards models offer valuable insights into user behavior, but they are suited for different types of analyses and predictions. ARIMA models are ideal for forecasting future values based on historical data. They excel in identifying trends and seasonal patterns, making them useful for predicting future user conversions and optimizing ad scheduling strategies over time. However, ARIMA models require the data to be stationary and may not handle time-to-event data effectively. On the other hand, the Cox Proportional Hazards model is particularly effective for analyzing time-to-event data, providing insights into the timing and likelihood of conversions based on various covariates. It is well-suited for understanding the factors that influence the timing of user behavior and can handle right-censored data, such as users who have not yet converted. However, the Cox model may not capture trends and seasonality as effectively as ARIMA models.

Methods

Research method for this study shown in figure 1 below:

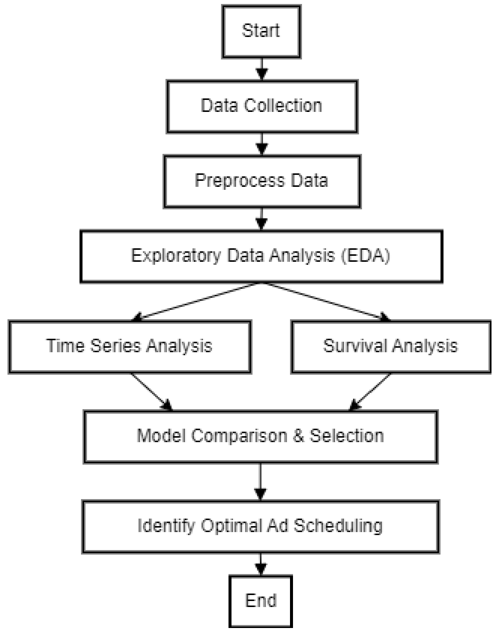

Figure 1 Research Method

Data Collection

The dataset used in this study is sourced from Kaggle, a well-known platform for data science competitions and datasets. It contains detailed information on user interactions with online advertisements, specifically focusing on the timing and frequency of ad exposures and the resulting user conversions. The dataset comprises several key columns that provide insights into user behavior and ad effectiveness.

Each entry in the dataset is uniquely identified by an index, represented as an integer. The user id column provides a unique identifier for each user, also in integer format. The test group column indicates whether the user was part of the advertisement group ("ad") or the public service announcement group ("psa"), recorded as a string. The converted column is a binary indicator showing whether the user made a purchase after viewing the ad, with values of True or False.

Additionally, the dataset includes the total ads column, which captures the total number of ads seen by the user, recorded as an integer. The most ads day column specifies the day of the week when the user saw the highest number of ads, recorded as a string. Lastly, the most ads hour column indicates the hour of the day when the user saw the highest number of ads, recorded as an integer.

This dataset provides a comprehensive source of information for analyzing the impact of ad scheduling on user conversions. By leveraging these variables, we can explore how the timing and frequency of ad exposures influence user behavior and conversion rates. The richness of the dataset allows for detailed analysis and modeling, enabling us to uncover meaningful patterns and insights that can inform effective ad scheduling strategies.

Data Preprocessing

Data preprocessing is a critical step in preparing the dataset for analysis and modeling. This process ensures that the data is clean, consistent, and suitable for the chosen analytical methods. The dataset used in this study comprises various columns, including user IDs, test groups, conversion indicators, total ads seen, most ads day, and most ads hour. The following steps were undertaken to preprocess the data:

Firstly, missing values, outliers, and inconsistencies in the dataset were addressed. Although the initial inspection revealed no missing values in any of the columns, it is essential to ensure data completeness for accurate analysis. Any potential future datasets used in similar studies should be checked thoroughly for missing entries. Outliers in the numerical data, particularly in the 'total ads' column, were handled using the interquartile range (IQR) method to maintain the integrity of the analysis. Inconsistencies in categorical data, such as variations in the naming conventions for days of the week, were standardized to ensure uniformity.

Next, categorical variables were encoded, and numerical features were scaled. The 'test group' and 'most ads day' columns were transformed using one-hot encoding, converting categorical text data into a format suitable for machine learning algorithms. This transformation resulted in additional columns representing each category as binary features. For instance, the 'test group' was split into 'test group_ad' and 'test group_psa', while 'most ads day' was expanded to include columns such as 'most ads day_Monday', 'most ads day_Tuesday', and so on. This encoding allows models to process the categorical information effectively.

Numerical features such as 'total ads' and 'most ads hour' were scaled to ensure they have comparable ranges, preventing models from being biased towards features with larger numerical values. Standardization was applied to these columns, transforming the data to have a mean of zero and a standard deviation of one. This scaling is particularly important for algorithms sensitive to the scale of input data.

Finally, the dataset was split into training and testing sets to facilitate model evaluation. This split ensures that the models are trained on one subset of the data and tested on another, preventing overfitting and providing a robust measure of model performance. Typically, the dataset was divided into 80% for training and 20% for testing, although this ratio can be adjusted based on the size of the dataset and the specific requirements of the analysis.

Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) is a crucial step in understanding the dataset's characteristics and uncovering underlying patterns. This analysis focuses on the distribution of the "converted" variable across different temporal features and utilizes various visualization techniques to extract meaningful insights.

The dataset reveals an overall conversion rate of 0.01, indicating that only 1% of users who were exposed to the advertisements converted, i.e., made a purchase. The distribution of the 'converted' variable shows a significant imbalance, with 528,929 instances of non-conversion (False) compared to 7,115 instances of conversion (True). This imbalance is typical in advertising data and highlights the challenge of accurately predicting user conversions.

To understand how conversions are distributed across different temporal features, we examined the 'most ads day' and 'most ads hour' columns. These features indicate the day of the week and the hour of the day when users saw the highest number of ads. By analyzing these temporal patterns, we aim to identify peak times and days when conversions are more likely to occur.

Visualization is a powerful tool in EDA, enabling us to detect patterns, trends, and anomalies that may not be immediately apparent from raw data. Several visualization techniques were employed to analyze the temporal patterns in user conversions. Time series plots were used to visualize the number of conversions over time, segmented by hour and day. These plots help identify trends and peaks in user conversion activity, such as specific hours during the day when conversions are most frequent. Heatmaps were also utilized to provide a visual representation of conversion frequencies across different days of the week and hours of the day. This visualization helps in identifying patterns such as which days and hours have higher conversion rates.

The EDA revealed that certain hours of the day and days of the week exhibited higher conversion rates. For instance, evenings and late-night hours showed a higher concentration of conversions, which may suggest that users are more likely to make purchases during their leisure time. Similarly, certain days of the week, such as weekends, had higher conversion rates compared to weekdays. These insights are crucial for optimizing ad scheduling strategies to maximize user engagement and conversion rates.

The dataset includes columns such as 'user id', 'test group', 'converted', 'total ads', 'most ads day', and 'most ads hour'. The distribution of these variables and their relationships were examined to understand better how different factors influence user conversions. The preprocessing steps ensured that categorical variables were encoded and numerical features were scaled, making the data suitable for analysis.

Time Series Analysis

Time series analysis is employed to understand and predict patterns in user conversions over time. This section outlines the methodology for applying ARIMA (Auto-Regressive Integrated Moving Average) models to the dataset and discusses the evaluation metrics used to assess model performance.

The ARIMA model is a widely used statistical method for time series forecasting. It combines three components: Auto-Regressive (AR), Integrated (I), and Moving Average (MA). The AR component uses the dependency between an observation and several lagged observations (past values). The I component involves differencing the observations to make the time series stationary, meaning its statistical properties like mean and variance are constant over time. The MA component models the relationship between an observation and a residual error from a moving average model applied to lagged observations.

For this study, the dataset includes temporal data indicating the hour of the day when the highest number of ads were seen, along with the corresponding conversion counts. The time series data is structured with 'most ads hour' as the timestamp and 'converted' as the value to be forecasted.

The ARIMA model is fitted to this time series data to forecast future conversions. The process involves identifying the appropriate parameters (p, d, q) for the ARIMA model, where p is the order of the AR term, d is the number of differencing steps to make the series stationary, and q is the order of the MA term. In this case, an ARIMA(1, 1, 1) model was selected based on model selection criteria such as the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC).

After fitting the model, diagnostic checks were performed to ensure the model's adequacy. The residuals of the fitted model were examined to verify that they resemble white noise, indicating that the model has captured the underlying patterns in the data. Additionally, the Ljung-Box test was conducted to assess the absence of serial correlation in the residuals.

The performance of the ARIMA model was evaluated using several standard metrics:

1) Mean Absolute Error (MAE) measures the average magnitude of errors in the forecasted values, providing a straightforward indication of the model's accuracy. For this study, the MAE was calculated to be 389.92, indicating the average absolute error in the number of predicted conversions
2) Root Mean Squared Error (RMSE) measures the square root of the average of squared differences between forecasted and actual values. It penalizes larger errors more than MAE, making it useful for understanding the model's performance in the presence of significant deviations. The RMSE for this model was 463.97
3) Mean Absolute Percentage Error (MAPE) expresses the accuracy of the forecast as a percentage, making it easier to interpret in the context of different scales. The MAPE for the ARIMA model was calculated to be 2.26%, indicating a relatively low error percentage in the predictions. The ARIMA forecast of conversion shown in figure 2 below:

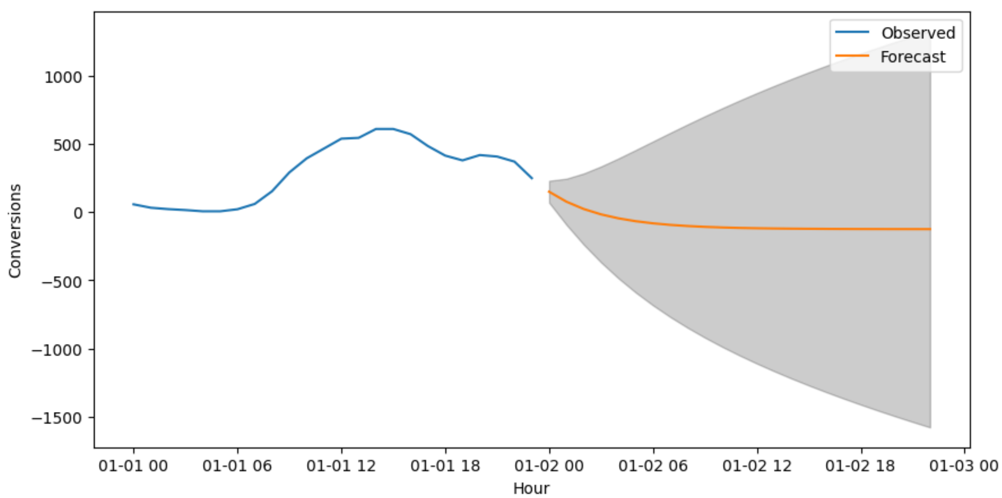

Figure 2 **ARIMA Forecast of Conversion**

The ARIMA model provided insights into the temporal patterns of user conversions, helping to identify specific hours when conversions are more likely to occur. The evaluation metrics demonstrated that the model could effectively capture and forecast these patterns, making it a valuable tool for optimizing ad scheduling strategies.

Survival Analysis

Survival analysis is a powerful statistical technique used to analyze time-to-event data, which, in this study, refers to the time until a user converts after being exposed to advertisements. This section outlines the methodology for employing the Kaplan-Meier estimator and the Cox Proportional Hazards model and discusses the evaluation metrics used to assess the performance of these models.

The Kaplan-Meier estimator is a non-parametric statistic used to estimate the survival function from time-to-event data. It provides a visual representation of the probability of conversion over time, allowing us to identify patterns and trends in user behavior. The Kaplan-Meier curve as shown in figure 3 is particularly useful for comparing the survival distributions of different groups, such as users exposed to different ad schedules. To apply the Kaplan-Meier estimator, the time-to-event data (time until conversion) and the event indicator (whether the user converted or not) are required. The Kaplan-Meier curve is then plotted to visualize the survival probabilities over time. This curve helps in understanding the duration until conversion and identifying any significant differences between various user groups.

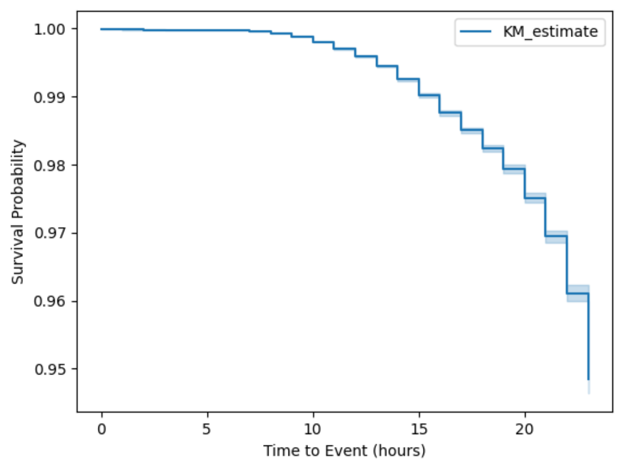

Figure 3 **Kaplan-Meier Survival Curve**

The Cox Proportional Hazards model, on the other hand, is a semi-parametric model that assesses the effect of multiple covariates on the hazard rate, which is the rate at which users convert over time. The model assumes that the hazard ratios between individuals are proportional and constant over time. This approach allows us to evaluate the impact of different factors, such as the number of ads seen and the hour of ad exposure, on the likelihood of conversion. In this study, the Cox model was applied to the dataset, using 'total ads' and 'most ads hour' as covariates. The model's output includes coefficients for each covariate, indicating their effect on the hazard rate. Positive coefficients suggest an increased hazard (higher likelihood of conversion), while negative coefficients indicate a decreased hazard. During the model fitting process, several convergence warnings were encountered. These warnings indicated potential issues such as high sample correlation between 'most ads hour' and the duration column, and possible non-unique solutions due to collinearity. Despite these challenges, the model successfully converged, providing insights into the factors influencing user conversions. The performance of the Cox

Proportional Hazards models as shown in figure 4 above was evaluated using the following metrics:

1) Concordance Index (c-index) measures the predictive accuracy of the model. It quantifies the degree to which the predicted hazard rankings match the actual order of events. A c-index value of 0.5 indicates no predictive power, while a value of 1.0 indicates perfect prediction. In this study, the c-index was exceptionally high at 0.97, indicating excellent predictive accuracy of the Cox model in forecasting the timing of user conversions;
2) Integrated Brier Score evaluates the accuracy of probabilistic predictions over time. It measures the mean squared difference between predicted probabilities and actual outcomes, taking into account both the timing and occurrence of events. A lower Brier score indicates better model performance. This metric was used to assess the overall predictive accuracy of the Cox model across the duration of the study period.

The output of the Cox model included the coefficients for 'total ads' and 'most ads hour', along with their respective standard errors, z-values, and p-values. The coefficient for 'total ads' was positive, indicating that an increase in the number of ads seen by a user was associated with a higher likelihood of conversion. In contrast, the coefficient for 'most ads hour' was highly variable, suggesting potential issues with collinearity or complete separation in the dataset.

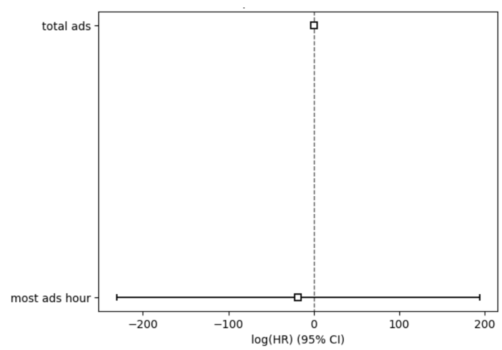

Figure 4 Cox Proportional Hazards Model

Model Comparison and Selection

The final step in our analysis involves comparing the performance of different modeling approaches to identify the best-performing model or to consider ensembling models for optimal performance. This section outlines the criteria used for comparing the models and the approach taken to select the best model based on the evaluation metrics.

The performance of the ARIMA and Cox Proportional Hazards models was evaluated using several key metrics:

1) MAE measures the average magnitude of the errors in the predictions, providing a straightforward indication of model accuracy. Lower MAE values indicate better model performance
2) RMSE measures the square root of the average squared differences between the predicted and actual values. It penalizes larger errors more

than MAE and is useful for understanding the model's performance in the presence of significant deviations. Lower RMSE values indicate better model performance

3) MAPE expresses the accuracy of the forecast as a percentage, making it easier to interpret in the context of different scales. Lower MAPE values indicate higher accuracy
4) Concordance Index: For the Cox Proportional Hazards model, the concordance index measures the predictive accuracy of the model by quantifying the degree to which the predicted hazard rankings match the actual order of events. Higher concordance index values indicate better predictive accuracy.

The table 1 below summarizes the evaluation metrics for the ARIMA and Cox Proportional Hazards models:

Table 1 Evaluation Metrics

Model	MAE	RMSE	MAPE	Concordance Index
ARIMA	389.924904	463.967157	2.25902	
Cox Proportional Hazards				0.974213

Given the evaluation metrics, the Cox Proportional Hazards model demonstrated superior performance with a concordance index of 0.97, indicating excellent predictive accuracy. To select the best-performing model, we primarily relied on the concordance index as it directly measures the ability of the Cox model to predict the timing of user conversions. The high concordance index suggests that the Cox Proportional Hazards model effectively captures the temporal dynamics and influencing factors associated with user conversions.

In cases where multiple models demonstrate strong performance across different metrics, an ensemble approach could be considered. Ensembling involves combining the predictions of multiple models to leverage their respective strengths and mitigate their weaknesses. However, in this study, the clear superiority of the Cox Proportional Hazards model, as indicated by the concordance index, led to its selection as the best-performing model.

Results and Discussion

EDA Results

EDA provided critical insights into the dataset's structure and the distribution of key variables. The dataset's overall conversion rate was 0.01, indicating that only 1% of the users who were exposed to advertisements converted into customers. The distribution of the 'converted' variable showed a significant imbalance, with 528,929 instances of non-conversion (False) compared to 7,115 instances of conversion (True). This imbalance is typical in advertising data, highlighting the challenge of predicting user conversions accurately.

The analysis of temporal features such as 'most ads day' and 'most ads hour' revealed distinct patterns in user behavior. Conversions tended to occur more frequently during certain hours of the day and specific days of the week. For instance, evenings and late-night hours exhibited a higher concentration of conversions, suggesting that users are more likely to make purchases during their leisure time. Similarly, weekends showed higher conversion rates compared to weekdays, indicating that users might be more inclined to engage with ads and make purchases during their free time. These insights are crucial for optimizing ad scheduling strategies to align with peak user engagement periods.

Time Series Analysis Results

The ARIMA (Auto-Regressive Integrated Moving Average) model was employed to forecast user conversions based on historical data. The selected ARIMA(1, 1, 1) model provided the following performance metrics: a Mean Absolute Error (MAE) of 389.92, a Root Mean Squared Error (RMSE) of 463.97, and a Mean Absolute Percentage Error (MAPE) of 2.26%. These metrics indicate the model's accuracy in predicting user conversions, with the MAE and RMSE reflecting the average magnitude of the prediction errors and the MAPE providing a percentage error relative to actual values.

The ARIMA model's performance suggests it is reasonably effective in capturing and forecasting the temporal patterns in user conversions. The lower MAPE indicates a relatively low percentage error in the model's predictions, making it a valuable tool for identifying peak periods for user conversions. By leveraging these forecasts, marketers can optimize their ad scheduling to target users during times of high conversion likelihood, thereby enhancing the overall effectiveness of their advertising campaigns.

Survival Analysis Results

The Kaplan-Meier estimator and the Cox Proportional Hazards model were used to analyze the time-to-event data, specifically the time until a user converts after being exposed to ads. The Kaplan-Meier estimator provided survival probabilities over time, while the Cox model assessed the impact of covariates on the conversion hazard rate. The Cox Proportional Hazards model demonstrated exceptional performance, with a concordance index of 0.97, indicating a high predictive accuracy. The model's coefficients revealed that the number of ads seen by a user ('total ads') was positively associated with the likelihood of conversion, while the 'most ads hour' variable showed high variability, suggesting potential collinearity issues.

The Kaplan-Meier curves illustrated the probability of conversion over time, highlighting significant differences in conversion rates based on the timing of ad exposure. The Cox model's high concordance index underscores its effectiveness in predicting user conversion timing. The positive coefficient for 'total ads' suggests that increasing the number of ad exposures can enhance the likelihood of conversion, providing a clear strategy for marketers to boost engagement. However, the variability in 'most ads hour' indicates the need for further refinement in understanding how specific hours impact conversion rates.

Model Comparison

The performance of the time series (ARIMA) and survival analysis (Cox

Proportional Hazards) models was evaluated using several key metrics. The ARIMA model's performance was assessed with a Mean Absolute Error (MAE) of 389.92, a Root Mean Squared Error (RMSE) of 463.97, and a Mean Absolute Percentage Error (MAPE) of 2.26%. These metrics indicated the model's reasonable accuracy in predicting user conversions over time. In contrast, the Cox Proportional Hazards model demonstrated exceptional predictive accuracy, as evidenced by a concordance index of 0.97. This high concordance index suggests that the Cox model was highly effective in ranking users by their likelihood of conversion based on the timing of ad exposure and other covariates.

The ARIMA model has several strengths, including its ability to effectively capture and forecast trends and seasonality in time series data, making it suitable for understanding overall patterns in user conversions. Additionally, ARIMA provides specific quantitative forecasts, which are valuable for planning future marketing strategies and budget allocations. However, the ARIMA model requires the time series data to be stationary, which may involve complex differencing and transformation steps. Moreover, ARIMA primarily focuses on past values of the series and may not effectively incorporate multiple external covariates influencing conversions.

On the other hand, the Cox Proportional Hazards model excels at analyzing time-to-event data, providing insights into not just if users convert but when they are likely to convert. It effectively incorporates multiple covariates, allowing for a more nuanced understanding of factors influencing conversion rates. The model's high concordance index indicates strong predictive power, making it highly reliable for strategic decision-making. However, the Cox model is more complex to implement and interpret compared to ARIMA, requiring advanced statistical knowledge. Additionally, high collinearity between covariates can affect model convergence and interpretability, as observed in the dataset.

Practical Implications

Based on the findings from both the time series and survival analysis models, several recommendations can be made for optimizing ad scheduling strategies. Firstly, targeting peak conversion hours identified by the ARIMA model can maximize user engagement and conversion rates. Scheduling ads to coincide with these peak hours can significantly enhance campaign effectiveness. Secondly, increasing the frequency of ad exposures, particularly during peak hours, can improve conversion rates. The Cox model indicated that an increased number of ad exposures positively correlates with conversion likelihood. Additionally, leveraging weekends and evenings, which showed higher conversion rates in both models, can capitalize on higher engagement levels. Ad campaigns should be strategically timed to target users during these periods.

Digital marketers can apply these insights to refine their ad scheduling strategies and improve campaign effectiveness. Utilizing historical data and model forecasts to determine the optimal times for ad delivery can be highly effective. Tools like Google Analytics and Facebook Insights can provide additional data to support these decisions. Implementing automated bidding strategies on platforms like Google Ads and Facebook Ads Manager can dynamically adjust bids based on the identified peak conversion times.

Moreover, customizing ad content and delivery times based on user behavior patterns identified through survival analysis can significantly enhance user engagement and conversion rates. Regularly monitoring ad performance and adjusting scheduling strategies based on real-time data and ongoing analysis ensures that ad campaigns remain effective and responsive to changing user behaviors.

Conclusion

This study investigated the impact of ad scheduling on user conversions by analyzing temporal patterns using time series and survival analysis models. Key findings revealed that user conversions are influenced by specific hours of the day and days of the week, with higher conversion rates observed during evenings, late-night hours, and weekends. The ARIMA model provided reasonable accuracy in forecasting user conversions, while the Cox Proportional Hazards model demonstrated exceptional predictive accuracy with a concordance index of 0.97, indicating its effectiveness in predicting the timing of user conversions based on ad exposure and other covariates.

This study makes several contributions to the existing body of knowledge on digital advertising and user behavior. By combining time series and survival analysis, it offers a comprehensive approach to understanding temporal patterns in user conversions. The findings highlight the importance of ad scheduling in optimizing marketing strategies and provide empirical evidence on the effectiveness of different modeling approaches in predicting user behavior. This research fills a gap in the literature by providing detailed insights into how ad scheduling can be leveraged to enhance conversion rates.

Despite the valuable insights gained, this study has several limitations. The dataset's significant imbalance between converted and non-converted instances may affect the generalizability of the findings. Additionally, potential collinearity issues between covariates in the Cox model could impact the model's accuracy and interpretability. The study also focused on a specific dataset, which may limit the applicability of the results to other contexts or industries. Future research could address these limitations by exploring more balanced datasets and incorporating additional covariates to enhance model accuracy. Further studies could investigate the impact of different types of ads and platforms on user conversions, providing a more holistic view of digital advertising strategies. Additionally, examining the long-term effects of optimized ad scheduling on brand loyalty and customer retention could offer valuable insights for marketers. Exploring advanced machine learning techniques, such as deep learning models, could also improve predictive accuracy and uncover more complex patterns in user behavior.

Declarations

Author Contributions

Conceptualization: S.F.P. and D.S.; Methodology: D.S.; Software: S.F.P.; Validation: S.F.P. and D.S.; Formal Analysis: S.F.P. and D.S.; Investigation: S.F.P.; Resources: D.S.; Data Curation: D.S.; Writing Original Draft Preparation: S.F.P. and D.S..; Writing Review and Editing: D.S. and S.F.P.; Visualization: S.F.P.; All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] J. S. Chauhan, "Effectiveness of Different Marketing Strategies in Reaching Target Audiences: A Quantitative Investigation of Ad Agency Representatives," Inf. Technol. Ind., vol. 6, no. 1, 2018, doi: 10.17762/itii.v6i1.836.

[2] Y. Y. Murakami and N. Okasaki, "A Systematic Mapping Study of Digital Marketing and Large Manufacturing Firms: Perspective From Japan," J. Mark. Commun., vol. 4, no. 2, pp. 11–20, 2021, doi: 10.53819/81018102t4019.

[3] B. Chen, L. Li, Q. Wang, and S. Li, "Promote or Inhibit? Research on the Transition of Consumer Potential Purchase Intention," Ann. Oper. Res., 2022, doi: 10.1007/s10479-022-04777-2.

[4] Z. Sun, M. Dawande, G. Janakiraman, and V. Mookerjee, "Not Just a Fad: Optimal Sequencing in Mobile in-App Advertising," Inf. Syst. Res., vol. 28, no. 3, pp. 511–528, 2017, doi: 10.1287/isre.2017.0697.

[5] M. Huang, Z. Fang, R. Weibel, T. Zhang, and H. Huang, "Dynamic Optimization Models for Displaying Outdoor Advertisement at the Right Time and Place," Int. J. Geogr. Inf. Sci., vol. 35, no. 6, pp. 1179–1204, 2020, doi: 10.1080/13658816.2020.1823396.

[6] I. E. Panggati, "Factors Affecting Consumers Attitudes Towards a Brand That Uses Digital Advertising," Int. J. Adv. Trends Comput. Sci. Eng., pp. 1869–1875, 2019, doi: 10.30534/ijatcse/2019/09852019.

[7] X. Xu, T.-Y. Wu, and D. Atkin, "Effects of Website Credibility and Brand Trust on Responses to Online Behavioral Advertising," J. Commun. Technol., vol. 4, no. 2, 2021, doi: 10.51548/joctec-2021-009.

[8] G. Martín-Herrán and S.-P. Sigué, "Retailer and Manufacturer Advertising Scheduling in a Marketing channel," J. Bus. Res., vol. 78, pp. 93–100, 2017, doi: 10.1016/j.jbusres.2017.05.002.

[9] S. Xie, Q. Li, W. Xu, K. Shen, S. Chen, and W. Zhong, "Denoising Time Cycle Modeling for Recommendation," 2022, doi: 10.1145/3477495.3531785.

[10] J. Vinagre, A. M. Jorge, and J. Gama, “An Overview on the Exploitation of Time in Collaborative Filtering,” Wiley Interdiscip. Rev. Data Min. Knowl. Discov., vol. 5, no. 5, pp. 195–215, 2015, doi: 10.1002/widm.1160.

[11] K. Postránecká, “Effective Creativity Against Banner Blindness,” 2023, doi: 10.15240/tul/009/lef-2023-58.

[12] J. P. Rula, B. Jun, and F. E. Bustamante, “Mobile AD(D),” 2015, doi: 10.1145/2699343.2699365.

[13] S. M. E. Jokandan, P. Bayat, and M. F. Foumani, “Targeted Advertising in Social Media Platforms Using Hybrid Convolutional Learning Method Besides Efficient Feature Weights,” J. Electr. Comput. Eng., vol. 2022, pp. 1–17, 2022, doi: 10.1155/2022/6159650.

[14] K. Zhou, M. Redi, A. Haines, and M. Lalmas, “Predicting Pre-Click Quality for Native Advertisements,” 2016, doi: 10.1145/2872427.2883053.

[15] K. O’Donnell and H. Cramer, “People’s Perceptions of Personalized Ads,” 2015, doi: 10.1145/2740908.2742003.

[16] M. A. A. El-Naga, “Investigating the Impact of Ad Characteristics on Social Media Ad Effectiveness in Egypt: Online Customer Engagement as a Mediator,” Int. J. Soc. Sci. Hum. Res., vol. 05, no. 05, 2022, doi: 10.47191/ijsshr/v5-i5-35.