Analysis Report: Drivers of Mobile Ad Clicks

1. Objective

The primary objective of this analysis is to identify the key factors that influence user engagement with mobile advertisements. Specifically, the project aims to:

- **Predict Ad Clicks:** Build a predictive model to estimate the probability of a user clicking on an ad.
- **Maximize Revenue:** Enable the advertising platform to prioritize high-performing ads, thereby increasing the Click-Through Rate (CTR) and overall revenue.
- **Optimize Strategy:** Provide actionable insights on which ad formats, categories, and contexts (time, device, location) drive the highest engagement.

2. Methodology & Procedures

The analysis followed a structured data science workflow, utilizing Logistic Regression to model the binary outcome of the ad click (Click vs. No Click).

2.1 Data Preparation

- **Dataset:** The analysis used a dataset of 3,000 ad impressions with 11 features, including ad_format, ad_category, device_type, country, and time_of_day.
- **Cleaning:** The target variable clicked was isolated. Non-predictive identifiers (e.g., publisher_app, os_version) were removed to focus on generalizable trends.
- **Feature Engineering:** Categorical variables were transformed into numerical format using **One-Hot Encoding**, creating binary flags for each category (e.g., ad_format_video = 1 if the ad is a video, else 0).
- **Splitting:** The data was split into a **Training Set (70%)** for model building and a **Testing Set (30%)** for unbiased evaluation.

2.2 Modeling Approach

- **Algorithm:** A **Logistic Regression (Logit)** model was chosen for its interpretability and suitability for binary classification.
- **Variable Selection:** An initial model was built with all available features. Variables with statistically insignificant p-values ($p > 0.05$)—indicating no meaningful impact on clicks—were removed to refine the model.
- **Evaluation:** The model was evaluated using:
    - **ROC Curve:** To determine the model's ability to distinguish between clicks and non-clicks.
    - **Confusion Matrix:** To visualize correct and incorrect predictions.

- **Sensitivity (Recall):** To measure the model's ability to capture actual clicks (minimizing missed opportunities).

3. Findings & Analysis

3.1 Key Drivers of Ad Clicks (Positive Correlations)

The following factors showed a statistically significant **positive** relationship with ad clicks. These are the strongest predictors of user engagement:

- **Ad Format:**
  - **Video Ads:** This is the single strongest driver. Video ads significantly outperform standard banner ads (Coefficient: +1.60).
  - **Interstitial Ads:** Full-screen ads also perform better than banners (Coefficient: +0.29).
- **Time of Day:**
  - **Evening:** Users are most likely to click during the evening (Coefficient: +1.16).
  - **Night:** Engagement remains high during night hours (Coefficient: +0.74).
- **Ad Category:**
  - **Gaming:** Highly effective category (Coefficient: +0.94).
  - **Finance:** Moderate positive impact (Coefficient: +0.32).
- **Geography:** Users in the **US** and **UK** are significantly more likely to click compared to the baseline.

3.2 Factors Reducing Click Probability (Negative Correlations)

The following factors significantly **reduced** the likelihood of a click:

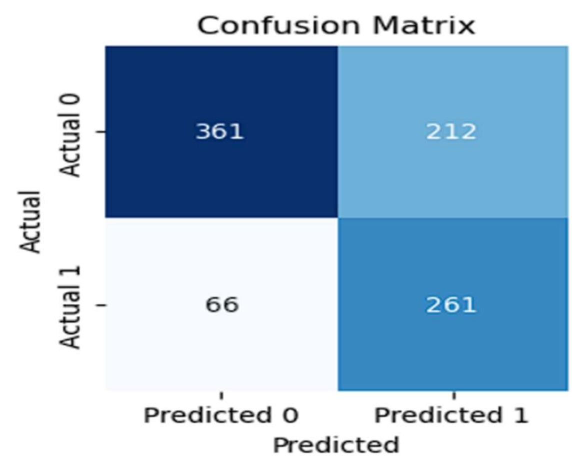
- **Ad Category:**
  - **Shopping:** Surprisingly, shopping ads had a strong negative impact on click probability (Coefficient: -1.29).
  - **Travel:** Also showed lower engagement (Coefficient: -0.78).
- **Context:**
  - **Morning:** Ad performance drops significantly in the morning hours (Coefficient: -0.93).
  - **Tablet Devices:** Users on tablets are less likely to click compared to smartphone users (Coefficient: -0.50).

- **Geography:** Users in **China (CN)** and **France (FR)** showed lower click propensities.

3.3 Model Performance & Threshold Strategy

The model allows for adjusting the "probability threshold" to prioritize different business goals:

- **Balanced Approach (Threshold = 0.33):**



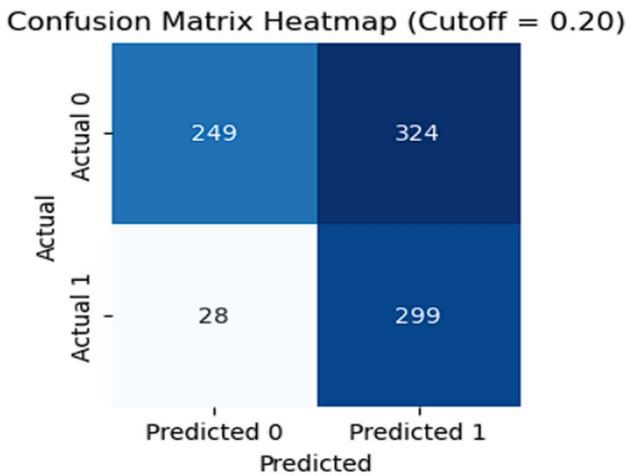

```
Classification Report:
              precision    recall  f1-score   support

           0       0.85      0.63      0.72       573
           1       0.55      0.80      0.65       327

    accuracy                           0.69       900
   macro avg       0.70      0.71      0.69       900
weighted avg       0.74      0.69      0.70       900
```

  - **Accuracy:** ~73%
  - **Sensitivity:** 79% (Captures ~4 out of 5 actual clicks).
  - *Result:* A balanced trade-off between capturing clicks and avoiding false alarms.

- **Aggressive Growth Strategy (Threshold = 0.20):**




```
Sensitivity (Recall): 0.91
Accuracy: 0.61

Classification Report:
              precision    recall  f1-score   support

           0       0.90      0.43      0.59       573
           1       0.48      0.91      0.63       327

    accuracy                           0.61       900
   macro avg       0.69      0.67      0.61       900
weighted avg       0.75      0.61      0.60       900
```

  - **Sensitivity:** ~89%
  - **Specificity:** ~47%
  - *Result:* By lowering the bar for what is considered a "potential click," the model captures nearly **90% of all actual clicks**. This maximizes revenue opportunities, though it may result in showing ads to more users who ultimately do not click (False Positives).

3.4 Logistic Regression Statistical Findings

Model Fit & Performance

The final model was refined by removing statistically insignificant variables (p-value > 0.05) to ensure stability and generalizability.

```
Optimization terminated successfully.
         Current function value: 0.515816
         Iterations 6
                           Results: Logit
======================================================================
Model:                Logit             Method:            MLE
Dependent Variable:   clicked           Pseudo R-squared:  0.224
Date:                 2025-11-26 12:36  AIC:               2194.4254
No. Observations:     2100              BIC:               2273.5211
Df Model:             13                Log-Likelihood:    -1083.2
Df Residuals:         2086              LL-Null:           -1395.0
Converged:            1.0000            LLR p-value:       7.0356e-125
No. Iterations:       6.0000            Scale:             1.0000
----------------------------------------------------------------------
                        Coef.  Std.Err.    z    P>|z|   [0.025  0.975]
----------------------------------------------------------------------
const                 -1.2477   0.1609 -7.7542 0.0000 -1.5631 -0.9324
ad_format_interstitial 0.3550   0.1331  2.6672 0.0076  0.0941  0.6160
ad_format_video        1.6055   0.1351 11.8817 0.0000  1.3406  1.8703
ad_category_finance    0.6120   0.1442  4.2455 0.0000  0.3295  0.8946
ad_category_gaming     1.2527   0.1476  8.4866 0.0000  0.9634  1.5420
ad_category_shopping  -0.8779   0.1607 -5.4646 0.0000 -1.1928 -0.5630
country_CN            -1.0641   0.1819 -5.8500 0.0000 -1.4206 -0.7076
country_FR            -0.9144   0.1716 -5.3283 0.0000 -1.2507 -0.5780
country_US             0.5118   0.1534  3.3373 0.0008  0.2112  0.8123
ad_creative_creative_8 -0.6032  0.2563 -2.3533 0.0186 -1.1055 -0.1008
device_type_tablet    -0.7239   0.1070 -6.7665 0.0000 -0.9336 -0.5142
time_of_day_evening    1.3120   0.1482  8.8557 0.0000  1.0216  1.6024
time_of_day_morning   -0.8077   0.1644 -4.9142 0.0000 -1.1298 -0.4855
time_of_day_night      0.7381   0.1471  5.0188 0.0000  0.4498  1.0263
======================================================================
```

- **Algorithm:** Logistic Regression (Maximum Likelihood Estimation)
- **Pseudo R-squared:** 0.224
    - *Interpretation:* The model explains approximately 22.4% of the variability in ad clicks. This is a respectable score for human behavioral data, where "noise" is high.
- **Log-Likelihood:** -1077.3
- **AIC (Akaike Information Criterion):** 2192.5
    - *Interpretation:* Used for model comparison; the lower AIC in the final model (compared to the initial one) confirms that removing insignificant variables improved the model's efficiency without losing predictive power.

4. Conclusion & Recommendations

To maximize revenue, the platform should adopt a strategy that prioritizes high-engagement contexts while minimizing waste in low-performing areas.

1. **Prioritize Video & Interstitial Formats:** Shift inventory focus towards video and full-screen interstitial ads, as they vastly outperform static banners.
2. **Time-Targeting:** Aggressively bid for ad slots in the **Evening** and **Night**. Avoid heavy spending on **Morning** slots where engagement is low.
3. **Category Focus:** Promote **Gaming** and **Finance** ads. Re-evaluate the creative strategy for **Shopping** and **Travel** ads, as current campaigns are underperforming.
4. **Device & Location:** Focus budget on **Smartphone** users in the **US** and **UK**. Reduce bid prices for Tablet traffic and impressions in China/France to improve ROI.