EXPLORATORY DATA ANALYSIS OF ADVERTISING CAMPAIGNS ON TECH PLATFORMS

Campaign Trends, Performance & Audience Behavior

TOOLS USED

Python, Pandas, Matplotlib, Seaborn & Plotly

PRESENTED BY

Tejas Jadhav

TABLE OF CONTENTS

PROBLEM STATEMENT & OBJECTIVE

Problem Statement:

The Digital Advertising landscape features thousands of campaigns across diverse platforms and formats, making it challenging to identify trends, understand audience engagement and determine which campaigns achieve the greatest impact.

Objective:

To explore the digital advertising campaign performance dataset and uncover patterns, insights, and trends that can help marketers and businesses optimize their campaigns and improve ROI.

EDA WORKFLOW

For this analysis, a structured workflow was followed, involving data collection, understanding, cleaning, exploration, and summarization of insights to allow a clear understanding of the dataset and its trends.

01 Data Collection

- Gathered the Google Play Store dataset from Kaggle

02 Data Understanding & Anomaly Detection

- Looked at data distributions
- Found missing values, outliers, and unusual patterns

03 Data Cleaning & Treatment

- Fixed missing or incorrect values
- Standardized formats for consistency

04 Exploratory Analysis

- Univariate Analysis: eg. Platform, Device_Type, Audience Category etc.
- Bivariate Analysis: Impressions vs Clicks
- Multivariate Analysis: Conversion Rate vs Creative Format vs Device

05 Insights & Reporting

- Summarized patterns and trends
- Highlighted key findings for developers and businesses

KEY QUESTIONS EXPLORED
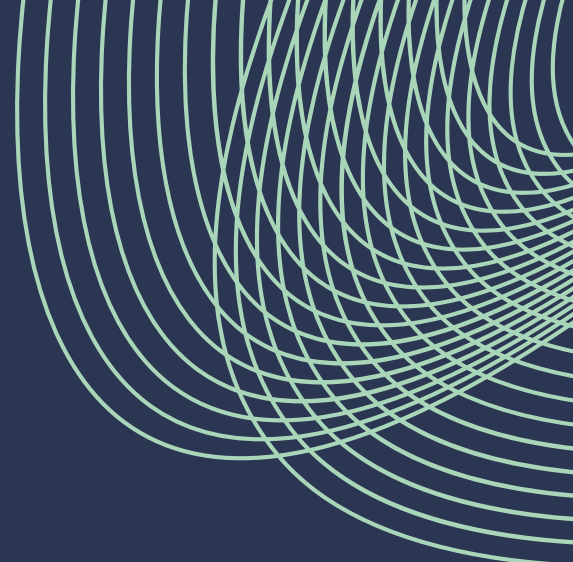
- Which platforms drive the highest revenue efficiency and ROAS and where should budget be reallocated?

- How does TikTok compare to Google and Facebook in terms of reach, revenue generation & scalability potential?

- Which creative formats (Video, Interactive, Text) generate the strongest engagement, conversions & cost efficiency?

- Which audience segments deliver the highest return on investment?

- How does retargeting impact profitability compared to non-retargeted campaigns?

- Where do the biggest funnel drop-offs occur and how can conversion efficiency be improved?

- Are higher budgets actually driving better performance, or is profitability concentrated among a few high-performing campaigns?

- How do campaign objectives (Conversion, Lead Gen, Brand Awareness) differ in terms of profitability & return efficiency?

DATA OVERVIEW

The dataset provides key information about Google Play Store apps, including ratings, reviews, installs, price, category, and type. Additional details like content rating, genres, app size, last update, and Android version capture app characteristics and trends.

Data Source: Kaggle

Dataset Size

10000

Records

41

Features

Campaign Diversity

70,53,92,230

Impressions

6

Platforms

6

industry Verticals

DATA OVERVIEW

Below is a detailed description of the feature set:

Dataset Features	Type	Feature Description
campaign_id	String	Unique identifier for each advertising campaign
campaign_objective	Categorical	Primary goal of the campaign
platform	Categorical	Advertising platform where the campaign ran (Google Ads, Facebook, LinkedIn, TikTok, Twitter, Instagram)
ad_placement	Categorical	Specific location of the ad display (Feed, Stories, Search, Display Network, In-Stream Video, Sidebar)
device_type	Categorical	User device used to view the ad (Desktop, Mobile, Tablet)
operating_system	Categorical	Device operating system (iOS, Android, Windows, macOS, Other)
creative_format	Categorical	Format of the ad creative (Video, Image, Carousel, Text, Interactive, Story)
creative_size	Categorical	Dimensions of the ad in pixels (e.g., 1080x1080, 1920x1080, 300x250)
ad_copy_length	Categorical	Length category of the ad text (Short, Medium, Long)
has_call_to_action	Boolean	Indicates whether the ad contains a CTA button (True/False)
creative_emotion	Categorical	Emotional tone conveyed by the ad (Fear, Joy, Urgency, Trust, Curiosity, Neutral)
creative_age_days	Numerical (Discrete)	Number of days since the creative was launched (1-90)
target_audience_age	Categorical	Age bracket of the targeted audience (18-24, 25-34, etc.)

DATA OVERVIEW

Continued…

Dataset Features	Type	Feature Description
target_audience_gender	Categorical	Gender of the targeted audience (Male, Female, All)
audience_interest_category	Categorical	Interest segment of the targeted audience
income_bracket	Categorical	Household income level of the audience (<$50K, $50K-$100K, etc.)
purchase_intent_score	Categorical	Behavioral purchase intent level (Low, Medium, High)
retargeting_flag	Boolean	Indicates whether the campaign is a retargeting effort (True/False)
quarter	Numerical (Discrete)	Calendar quarter of the campaign (1–4)
day_of_week	Categorical	Day of the week the ad was shown (Monday–Sunday)
hour_of_day	Numerical (Continuous)	Hour of the day the ad was displayed (0–23)
campaign_day	Numerical (Discrete)	Day number within the campaign lifecycle (1–90)
quality_score	Numerical (Discrete)	Platform-assigned quality rating of the ad (1–10)
actual_cpc	Numerical (Continuous)	Actual cost per click paid for the ad ($0.25–$17.00)
impressions	Numerical (Discrete)	Number of times the ad was displayed (5,000–500,000)
clicks	Numerical (Discrete)	Number of ad clicks (≥10 and ≤ impressions)

DATA OVERVIEW

Continued…

Dataset Features	Type	Feature Description
conversions	Numerical (Discrete)	Number of completed actions from the ad (0+ and ≤ clicks)
ad_spend	Numerical (Continuous)	Total campaign spend calculated as clicks × actual_cpc
revenue	Numerical (Continuous)	Revenue generated from the campaign ($)
bounce_rate	Numerical (Discrete)	Percentage of immediate exits (10-90%)
avg_session_duration_seconds	Numerical (Continuous)	Average time users spend on the site after clicking (10-600 seconds)
pages_per_session	Numerical (Discrete)	Average number of pages viewed per session (1.0-10.0)
industry_vertical	Categorical	Business sector of the advertiser
budget_tier	Categorical	Campaign budget classification (Low, Medium, High)
CTR	Numerical (Continuous)	Click-through rate (%) = (clicks / impressions) × 100
CPC	Numerical (Continuous)	Cost per click ($) = ad_spend / clicks
conversion_rate	Numerical (Continuous)	Conversion rate (%) = (conversions / clicks) × 100
CPA	Numerical (Continuous)	Cost per acquisition ($) = ad_spend / conversions
ROAS	Numerical (Continuous)	Return on ad spend = revenue / ad_spend
profit	Numerical (Continuous)	Profit ($) = revenue − ad_spend

DATA QUALITY CHALLENGES & ANOMALIES

Few inconsistencies were found in the dataset, which could have affected the analysis if left unaddressed.

DATA ANOMALIES

- Columns requiring cleaning or type conversion: Start_Date.
- Derived metrics (CTR, conversion rate, CPA, ROAS) may not perfectly match base metrics due to pre-computation or rounding, so recalculating them from raw data is recommended for accurate analysis.

DATA CLEANING & TREATMENT

Inconsistent and missing values were addressed, and key features were cleaned and standardized for analysis.

DATA CLEANING SUMMARY

- Start_Date was cleaned and converted to proper datetime format to ensure accurate temporal analysis.
- Derived performance metrics (CTR, conversion rate, CPA, ROAS) were recalculated from base fields to resolve rounding or pre-computation inconsistencies and ensure analytical accuracy.
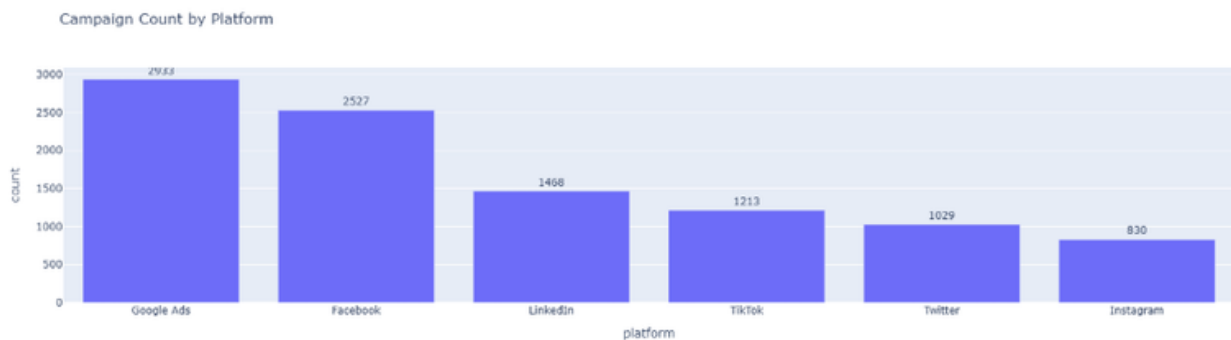
INSIGHTS

Platform & Market Landscape Overview

Google Ads & Facebook dominate campaign allocation, capturing the majority of advertising activity

3,578

Total Apps in Top 3 Categories

Campaign Count by Platform

Key observations

- Google Ads has the highest campaign count (~2,933), followed by Facebook (~2,527), significantly ahead of other platforms.
- Instagram has the lowest campaign count (~830), while LinkedIn, TikTok, and Twitter fall in a mid-to-lower tier range.

Business Insights

- Budget concentration appears heavily skewed toward Google Ads and Facebook, indicating these platforms are primary acquisition channels.
- Performance optimization efforts should prioritize Google and Facebook first, while smaller platforms may require selective or experimental investment rather than equal budget distribution.

TikTok delivers the highest average reach despite not leading in campaign volume

8,719

Apps are free to download

Average Impressions by Platform
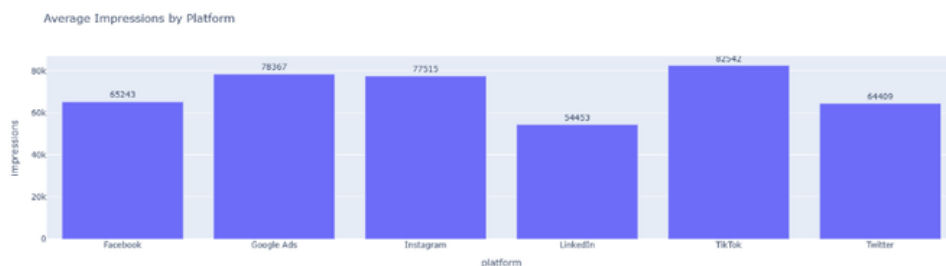
Key observations

- TikTok has the highest average impressions (~82,542), outperforming all other platforms.

- LinkedIn records the lowest average impressions (~54,453), while Google Ads and Instagram are close behind TikTok but still lower.
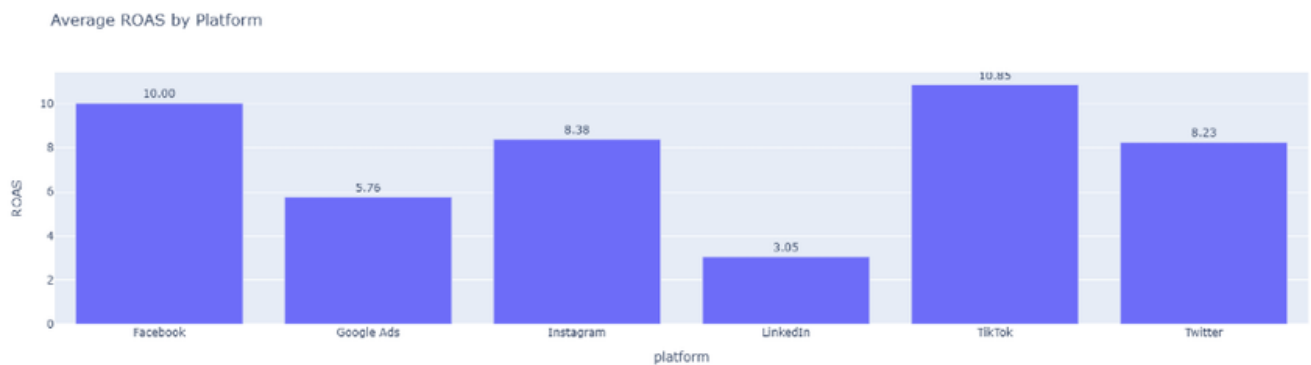
Business Insights

- TikTok appears to generate stronger reach per campaign, indicating higher exposure efficiency at the campaign level.

- If awareness and reach are primary objectives, reallocating incremental budget toward TikTok could increase total visibility without proportionally increasing campaign count.

TikTok and Facebook generate the strongest return efficiency, while LinkedIn significantly underperforms

79.2%

Apps targeted general audience

Average ROAS by Platform

Key observations

- TikTok has the highest average ROAS (~10.85), followed closely by Facebook (~10.00).
- LinkedIn records the lowest average ROAS (~3.05), substantially below all other platforms.

Business Insights

- Revenue efficiency is concentrated on TikTok and Facebook, indicating these platforms convert spend into revenue more effectively.
- Budget reallocation from lower-ROAS platforms (particularly LinkedIn) toward higher-performing ones could improve overall marketing profitability and blended ROAS.

Higher budgets amplify ROAS on TikTok & Instagram, while LinkedIn consistently underperforms regardless of spend

4.3

Typical App Rating (Median)

Average ROAS by Platform and Budget Tier
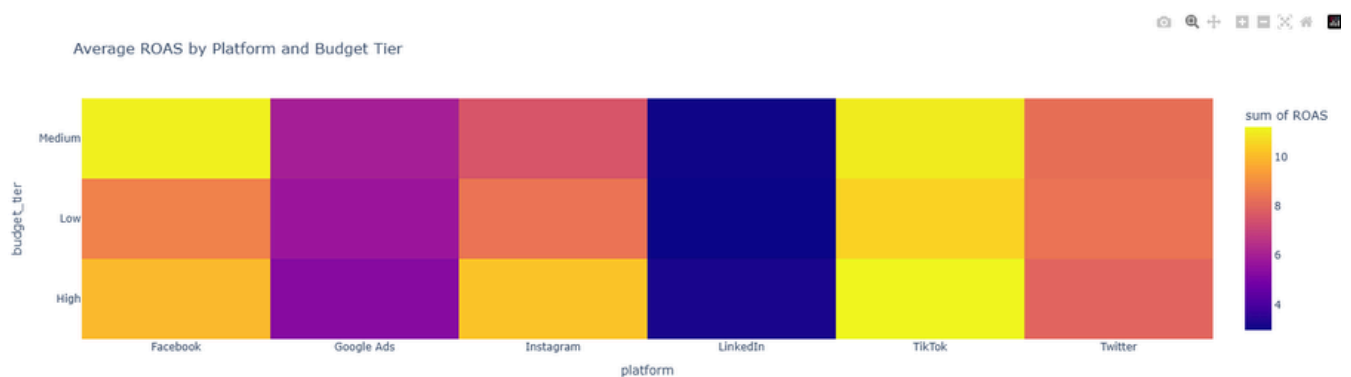
Key observations

- TikTok and Instagram show the highest ROAS at the High budget tier, outperforming other platforms at the same spend level.

- LinkedIn records the lowest ROAS across all budget tiers, with minimal improvement even as budget increases.

Business Insights

- Scaling budgets on TikTok and Instagram is likely to generate stronger revenue efficiency compared to other platforms.

- Budget allocation toward LinkedIn should be reassessed or optimized strategically, as increased spend does not translate into proportional ROAS gains.
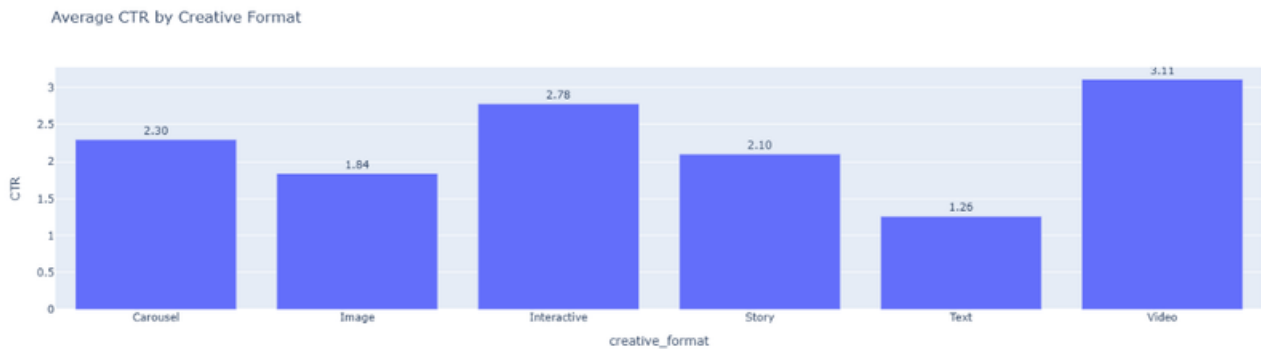
Creative Strategy & Engagement Analysis

Video & interactive creatives drive the strongest engagement, while text-based formats significantly underperform

4.3

Typical App Rating (Median)

Average CTR by Creative Format

Key observations

- Video has the highest average CTR (~3.11), followed by Interactive (~2.78), outperforming all other creative formats.
- Text ads record the lowest CTR (~1.26), substantially below visual and dynamic formats like Carousel and Story.
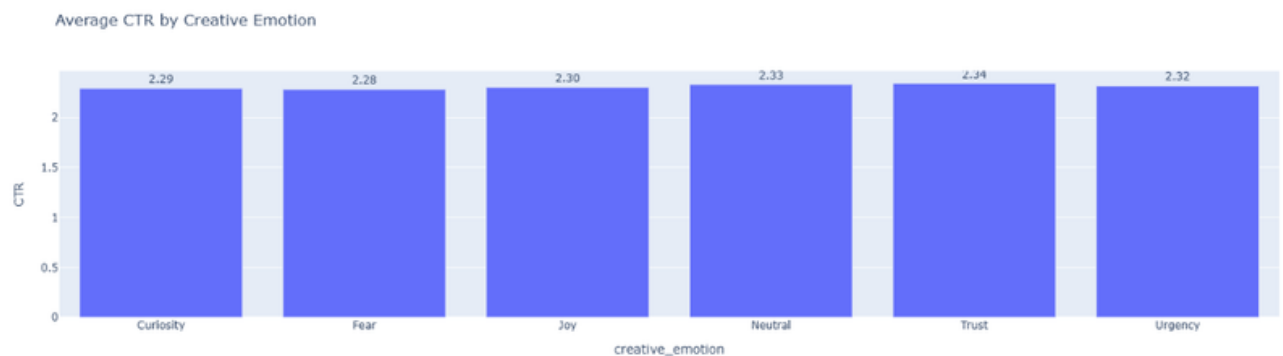
Business Insights

- Shifting creative investment toward Video and Interactive formats can meaningfully improve engagement and traffic efficiency.
- Text-heavy formats should be deprioritized or redesigned with stronger visual elements to prevent wasted impressions and low engagement rates.

Creative emotion has minimal impact on CTR, with performance remaining nearly uniform across tones

4.3

Typical App Rating (Median)

Average CTR by Creative Emotion

Key observations

- All creative emotions cluster tightly between ~2.28 and ~2.34 CTR, indicating very small performance variation.
- Trust (~2.34) and Neutral (~2.33) are marginally higher, but the gap versus the lowest (Fear ~2.28) is negligible.

Business Insights

- Emotional tone alone is unlikely to materially improve engagement; optimization efforts should prioritize format, placement, or targeting instead.
- Creative testing budgets can be consolidated rather than fragmented across emotional variants, as returns appear relatively consistent across tones.

Google Ads and Facebook campaigns leverage joyful and urgent emotions most effectively.

5,930

Typical Number of Reviews (Median)

Creative Emotion vs Platform
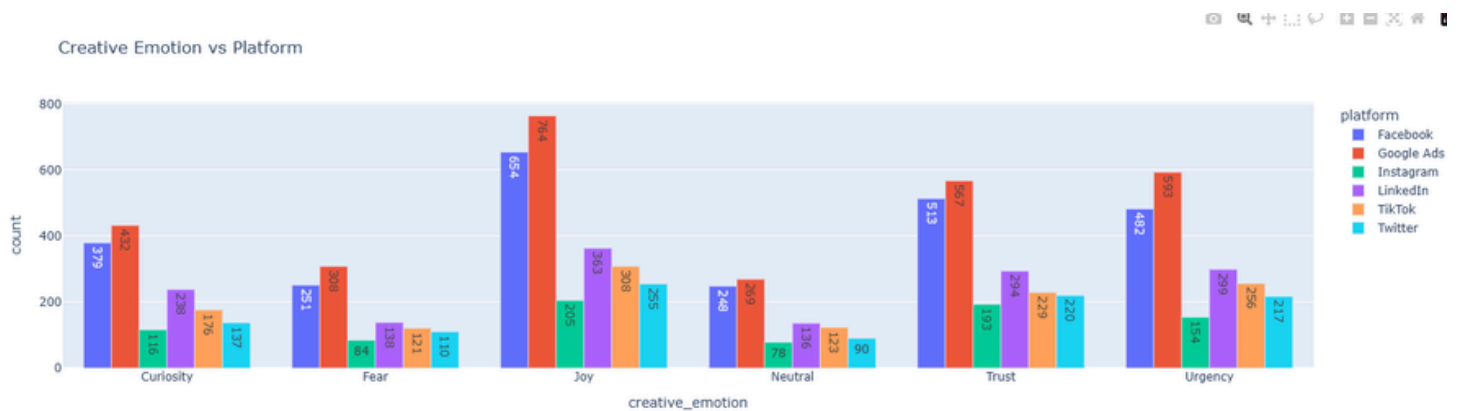
Key observations

- "Joy" and "Urgency" are the most frequently used emotions on Google Ads and Facebook, with counts of 764 and 593 respectively for Google Ads.
- Platforms like Instagram, TikTok, and Twitter show lower overall usage across all emotions, with comparatively higher counts for "Joy" and "Trust" than "Fear" or "Neutral."

Business Insights

- Prioritize "Joy" and "Urgency"-driven creatives on Google Ads and Facebook to maximize engagement, given their strong adoption and apparent platform alignment.
- For platforms with lower volume (Instagram, TikTok, Twitter), focus on emotionally resonant content like "Trust" and "Joy" to stand out, rather than spreading effort across less effective emotions.

Desktop video creatives drive the highest conversion rates across devices.

Conversion Rate by Creative Format and Device Type
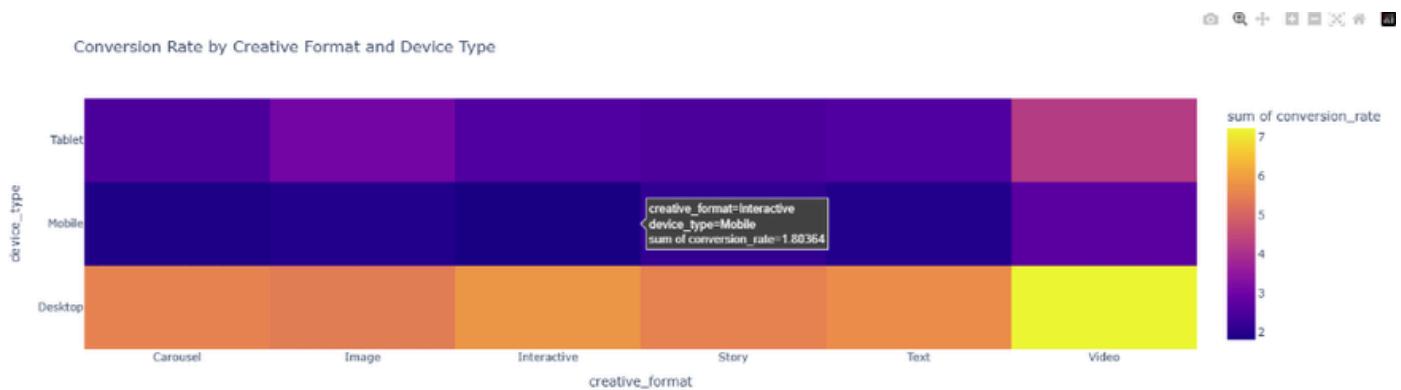
Key observations

- Desktop devices consistently outperform Mobile and Tablet in conversion rate for all creative formats, with Video creatives reaching the peak.

- Mobile and Tablet show much lower conversion rates, with minor variation across creative formats.

Business Insights

- Prioritize video creatives for Desktop campaigns to maximize conversions and ROAS.

- Consider device-specific optimization: focus higher budgets on Desktop while tailoring creative formats for Mobile and Tablet to improve their performance.
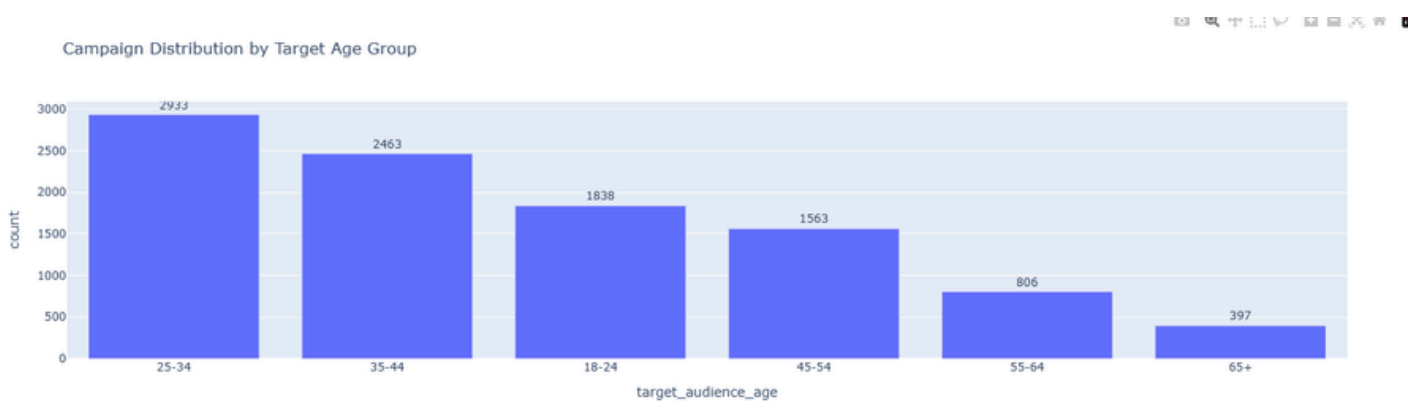
Audience Targeting
& Intent Evaluation

Young adults aged 25-34 are the primary focus of advertising campaigns.

5,930

Typical Number of Reviews (Median)

Campaign Distribution by Target Age Group

Key observations

- The 25-34 age group accounts for the highest number of campaigns (2,933), followed by 35-44 (2,463).
- Older age groups, particularly 55-64 (806) and 65+ (397), receive significantly fewer campaigns.

Business Insights

- Marketing efforts and budgets should prioritize the 25-34 and 35-44 segments for maximum reach and engagement.
- Tailored messaging or alternative channels may be required to effectively target older audiences who are currently underrepresented.

Higher purchase intent audiences convert at more than double the rate of low-intent segments

5,930

Typical Number of Reviews (Median)

Average Conversion Rate by Purchase Intent Score

Key observations

- High-intent audiences deliver the highest average conversion rate (~6.87), more than 2x Medium (~3.08) and nearly 4x Low (~1.83).
- There is a clear step-change improvement in conversion rate as intent level increases from Low → Medium → High.
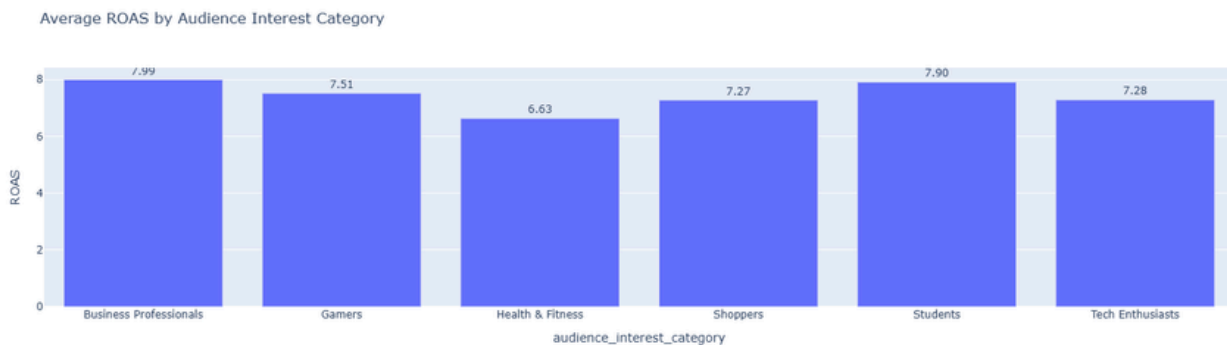
Business Insights

- Purchase intent scoring should be a primary driver of budget allocation, with aggressive investment in High-intent segments to maximize efficiency.
- Low-intent audiences should be repositioned into upper-funnel or retargeting pipelines rather than treated as direct conversion targets.

Students and Business Professionals generate the strongest return, while Health & Fitness underperforms in ROAS

5,930

Typical Number of Reviews (Median)

Average ROAS by Audience Interest Category

Key observations

- Students (~7.90) and Business Professionals (~7.99) show the highest average ROAS among all interest categories.
- Health & Fitness records the lowest ROAS (~6.63), notably below the top-performing segments.

Business Insights

- Scaling spend toward Students and Business Professionals could improve overall portfolio-level ROAS.
- Campaigns targeting Health & Fitness may require creative, offer, or targeting optimization before additional budget allocation.
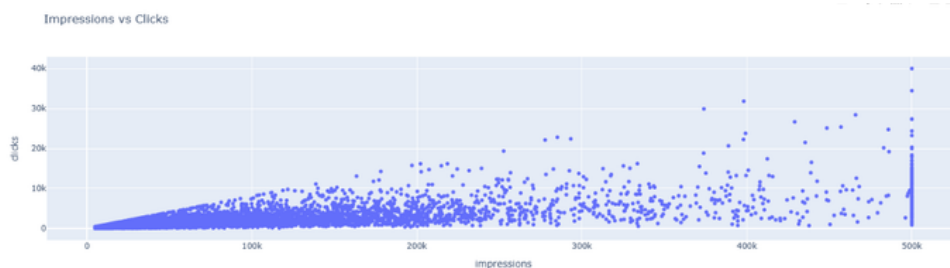
Funnel Performance & Optimization Insights

Higher impressions generally drive more clicks, but click growth becomes highly inconsistent at scale

5,930

Typical Number of Reviews (Median)

Impressions vs Clicks

Key observations

- There is a clear positive relationship between impressions and clicks — as impressions increase, clicks also increase.
- At very high impression levels (especially near 500K), click volumes vary widely, indicating large performance dispersion across campaigns.

Business Insights

- Scaling impressions alone does not guarantee proportional click growth; performance monitoring must accompany budget scaling.
- High-impression campaigns should be optimized for creative, targeting, and placement efficiency to reduce variability and stabilize CTR performance.

Conversion growth increases with clicks, but efficiency varies significantly across campaigns

5,930

Typical Number of Reviews (Median)

Clicks vs Conversions
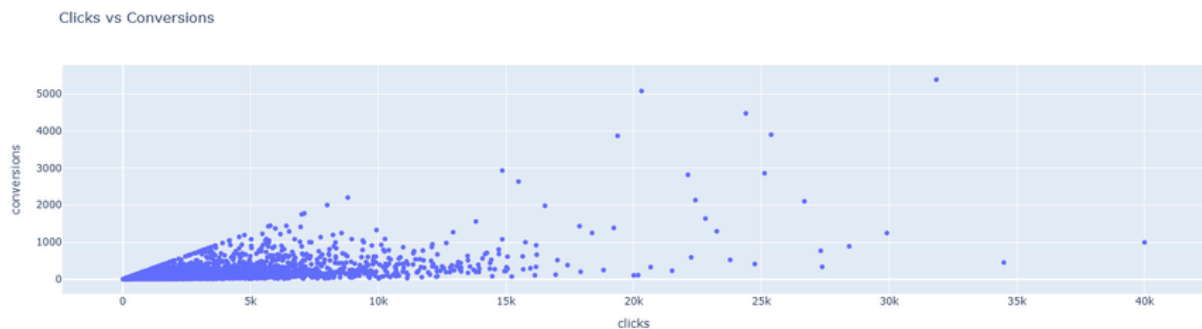
Key observations

- There is a positive relationship between clicks and conversions — campaigns with more clicks generally generate more conversions.
- At higher click volumes, conversion outcomes vary widely, with some campaigns achieving strong conversion counts while others remain relatively low.

Business Insights

- Driving traffic alone is insufficient; conversion optimization (landing pages, audience alignment, offer strength) must be prioritized to improve efficiency.
- High-click, low-conversion campaigns should be audited to reduce wasted spend and improve overall CPA performance.

Higher budgets do not translate into lower acquisition costs

5,930

Typical Number of Reviews (Median)

Median CPA by Budget Tier
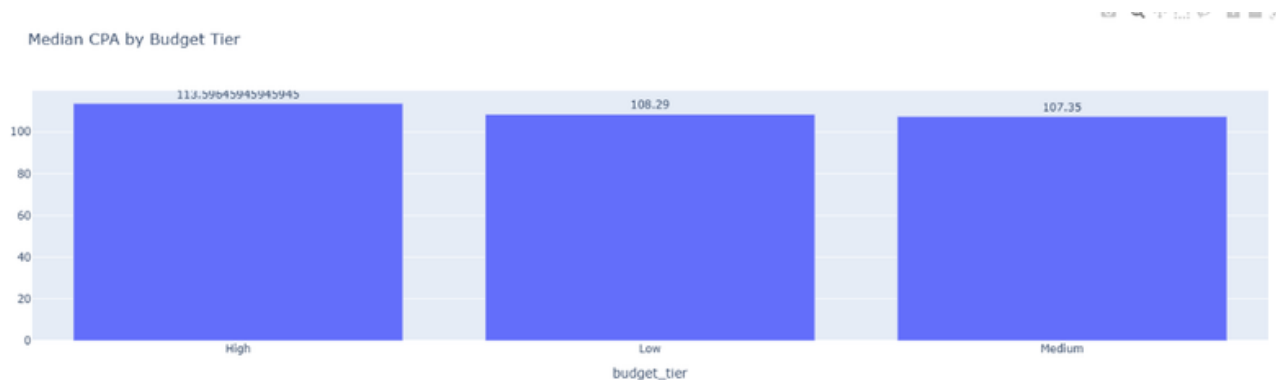
Key observations

- Median CPA is slightly highest for High budget campaigns (~113.6) compared to Low (~108.3) and Medium (~107.4).
- The difference in CPA across budget tiers is relatively small, indicating no strong cost advantage at higher spend levels.

Business Insights

- Increasing budget alone does not improve cost efficiency; optimization quality matters more than spend size.
- Budget allocation decisions should prioritize campaign performance and targeting effectiveness rather than assuming economies of scale at higher tiers.
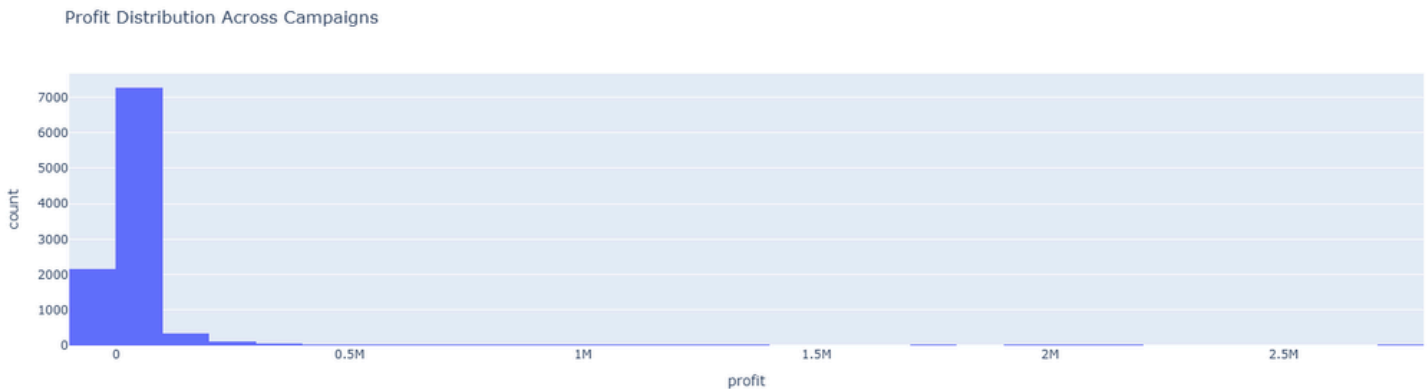
Budgeting, Profitability & Retargeting Strategy

Most campaigns generate modest profits, while a small number of outliers drive extreme gains

5,930

Typical Number of Reviews (Median)

Profit Distribution Across Campaigns

Key observations

- The profit distribution is heavily right-skewed, with the majority of campaigns clustered near low profit values.
- A small fraction of campaigns generate exceptionally high profits, creating a long tail of extreme outliers.

Business Insights

- Overall profitability is likely driven by a limited set of high-performing campaigns rather than consistent gains across all campaigns.
- Budget allocation decisions should prioritize campaign performance and targeting effectiveness rather than assuming economies of scale at higher tiers.

Conversion-led campaigns outperform all other objectives in median profitability

5,930

Typical Number of Reviews (Median)

Key observations

- Campaigns optimized for Conversions show the highest median profit, followed by Lead Generation, indicating stronger revenue outcomes.
- Brand Awareness campaigns have the lowest median profit, substantially below performance-driven objectives.

Business Insights

- Profit-driven strategies should prioritize Conversion and Lead Generation objectives, as they consistently yield higher median returns.
- Brand Awareness campaigns should be evaluated separately on upper-funnel metrics, as their direct profitability appears weaker compared to performance objectives.

Retargeting drives higher profitability regardless of campaign objective

5,930

Typical Number of Reviews (Median)

Key observations

- For every campaign objective, retargeted campaigns show substantially higher median profit than non-retargeted ones.
- The profit uplift from retargeting is most pronounced for Conversions and Lead Generation objectives.

Business Insights

- Retargeting should be treated as a core profitability lever rather than a tactical add-on, especially for performance-driven campaigns.
- Budget allocation for conversion and lead-gen campaigns can be optimized by systematically pairing them with retargeting strategies.
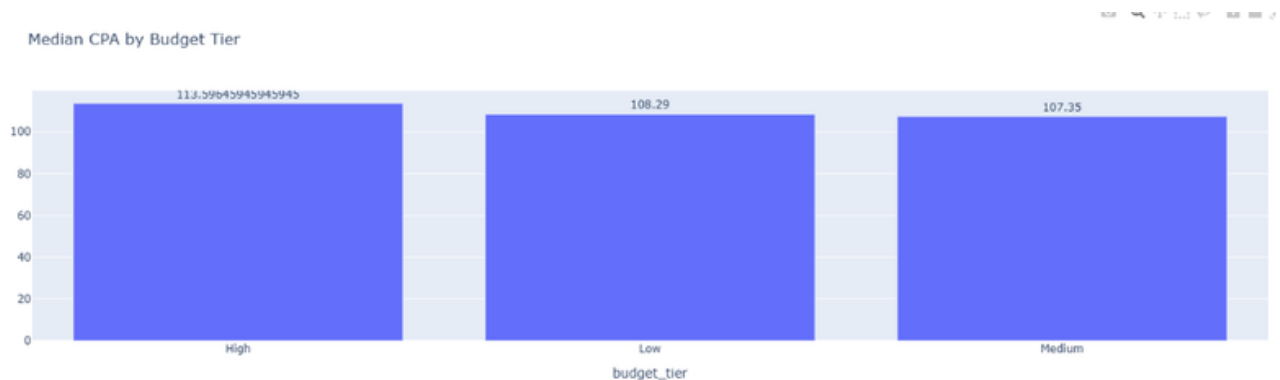
Temporal Performance & Stability Trends

TikTok and Instagram Drive Peak ROAS, While LinkedIn Consistently Underperforms Across Months

5,930

Typical Number of Reviews (Median)

Median CPA by Budget Tier

Key observations

- TikTok records the highest ROAS spikes across the timeline (peaking above 20), followed by strong but volatile performance from Instagram and Twitter.
- LinkedIn maintains the lowest ROAS throughout the period, rarely exceeding mid-range values compared to other platforms.

Business Insights

- Budget allocation should prioritize TikTok and Instagram during high-performing months, leveraging their strong upside potential for revenue acceleration.
- LinkedIn campaigns may require strategic repositioning (objective, creative, or targeting refinement) or tighter budget control due to consistently lower return efficiency.

BUSINESS / DEVELOPER TAKEAWAYS
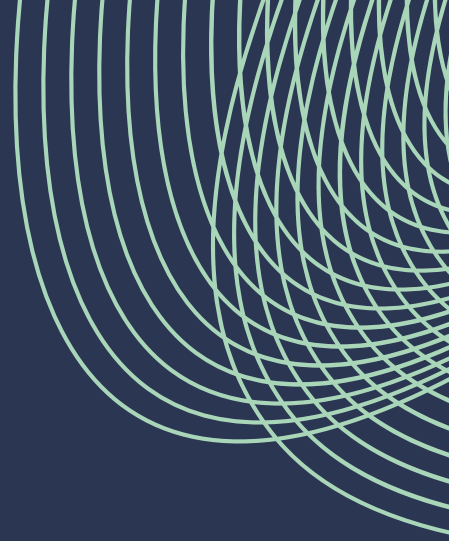
- **Double Down on Core & High-Return Platforms**: Google Ads and Facebook are primary acquisition engines, while TikTok and Instagram show strong revenue & reach efficiency. Reallocate budget from underperforming platforms like LinkedIn toward higher ROAS channels.

- **Use TikTok Strategically for Scalable Reach**: TikTok offers strong exposure and revenue efficiency. Incremental investment can boost visibility and blended ROAS without increasing campaign volume proportionally.

- **Prioritize Creative Format Over Emotional Variants:** Video and Interactive creatives outperform text-heavy formats. Emotional tone alone has limited impact. The focus should on optimizing format, placement, and audience alignment.

- **Focus on High-Intent & Core Age Segments (25-44):** High purchase-intent audiences and the 25–44 age group drive the best engagement and returns. Low-intent segments should be nurtured through upper-funnel or retargeting strategies.

BUSINESS / DEVELOPER TAKEAWAYS
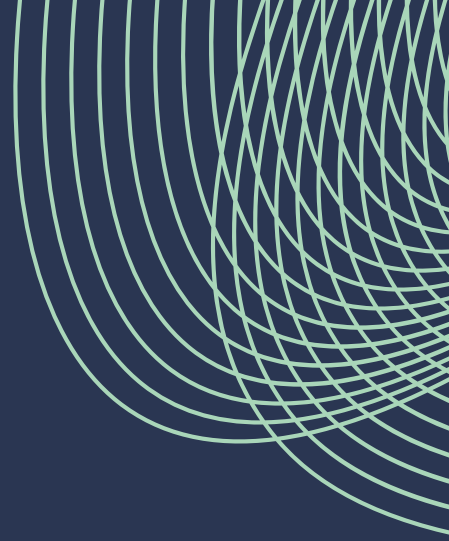
- **Treat Retargeting as a Core Profit Lever**: Retargeting improves profitability and should be systematically paired with Conversion and Lead Generation campaigns rather than used tactically.

- **Optimize the Funnel, Not Just Spend**: Scaling impressions or clicks alone doesn't guarantee results; conversion optimization through landing pages, targeting, and offer strength is key to improving CPA and stabilizing ROAS.

- **Profitability Is Concentrated, Scale What Works:** A limited set of high-performing campaigns drive the majority of profits. Budget allocation should prioritize proven performers rather than assuming efficiency improves with scale.

- **Align Budget with Objectives:** Conversion and Lead Generation campaigns generate stronger direct returns, while Brand Awareness should be evaluated on upper-funnel metrics, not short-term profitability.

CHALLENGES & OPPORTUNITIES

Limitations

- The dataset only includes campaigns from January 2024 to January 2026, so longer-term trends are not captured.
- No geographic information is included, so regional performance insights cannot be derived.
- Campaigns are limited to certain advertisers & industries, which could bias insights toward specific sectors.

Future Work / Opportunities:

- Develop models to predict which campaigns are likely to succeed and optimize strategies for maximum ROI automatically.
- Detect irregularities in engagement and conversions, flagging potential fraudulent activity or ineffective campaign tactics.
- Explore which creative approaches, messaging, or formats resonate most with audiences, informing future content strategies.
- Create simulations to predict outcomes under different campaign strategies, budgets, or market conditions, helping in decision-making.

CONCLUSION

This analysis of digital advertising campaigns provided a clear view of performance patterns, audience engagement, and ROI trends. Key insights were uncovered, showing how data-driven approaches can inform strategic decisions. Future work could focus on building predictive models for campaign success, detecting anomalies in engagement, analyzing effective creative strategies, and simulating outcomes under different campaign scenarios.

This analysis was conducted using Python, Pandas, Matplotlib, Seaborn and Plotly.

Author:

Tejas Jadhav

GitHub: @tejas-jadhav
LinkedIn: Tejas Jadhav