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
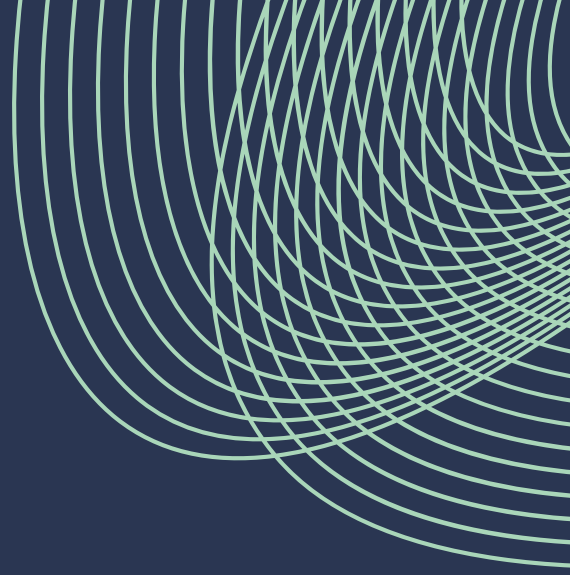
04 Exploratory Analysis

- Univariate Analysis: e.g., Ratings, Installs, Prices
- Bivariate Analysis: Rating vs Reviews
- Multivariate Analysis: Reviews vs Rating vs Installs

05 Insights & Reporting

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

KEY QUESTIONS EXPLORED

- Which categories dominate the Play Store, and where is competition highest?

- What is the distribution of Free vs Paid apps across categories and ratings?

- Who are the target audiences of apps (content rating distribution)?

- How do app sizes vary, and what is the optimal size range?

- How do app ratings behave across categories, review volumes, and time?

- What monetization models work best by category and audience?

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

9660

Records

13

Features

App Diversity

8196

Apps covered

33

App Categories

115

App Genres

DATA OVERVIEW

Below is a detailed description of the feature set:

Dataset Features	Type	Feature Description
App	String	Name of the application as listed on the Google Play Store
Category	Categorical	Primary category under which the app is classified
Rating	Numerical (Continuous)	Average user rating of the app on a scale of 1 to 5
Reviews	Numerical (Discrete)	Total number of user reviews submitted for the app
Size	Numerical (Continuous)	Storage size of the app as reported on the Play Store
Installs	Numerical (Discrete)	Approximate number of times the app has been installed by users
Type	Categorical	Indicates whether the app is Free or Paid
Price	Numerical (Continuous)	Price of the app (0 indicates a free app)
Content Rating	Categorical	Target age group suitability (e.g., Everyone, Teen, Mature)
Genres	Categorical	Specific genre or sub-genres associated with the app
Last Updated	Date/time	Date when the app was most recently updated
Current Ver	Categorical	Current version of the application
Android Ver	Categorical	Minimum Android OS version required to run the app

DATA OVERVIEW

Below is a detailed description of the feature set:

Dataset Features	Type	Feature Description
App	String	Name of the application as listed on the Google Play Store
Category	Categorical	Primary category under which the app is classified
Rating	Numerical (Continuous)	Average user rating of the app on a scale of 1 to 5
Reviews	Numerical (Discrete)	Total number of user reviews submitted for the app
Size	Numerical (Continuous)	Storage size of the app as reported on the Play Store
Installs	Numerical (Discrete)	Approximate number of times the app has been installed by users
Type	Categorical	Indicates whether the app is Free or Paid
Price	Numerical (Continuous)	Price of the app (0 indicates a free app)
Content Rating	Categorical	Target age group suitability (e.g., Everyone, Teen, Mature)
Genres	Categorical	Specific genre or sub-genres associated with the app
Last Updated	Date/time	Date when the app was most recently updated
Current Ver	Categorical	Current version of the application
Android Ver	Categorical	Minimum Android OS version required to run the app

DATA OVERVIEW

Below is a detailed description of the feature set:

Dataset Features	Type	Feature Description
App	String	Name of the application as listed on the Google Play Store
Category	Categorical	Primary category under which the app is classified
Rating	Numerical (Continuous)	Average user rating of the app on a scale of 1 to 5
Reviews	Numerical (Discrete)	Total number of user reviews submitted for the app
Size	Numerical (Continuous)	Storage size of the app as reported on the Play Store
Installs	Numerical (Discrete)	Approximate number of times the app has been installed by users
Type	Categorical	Indicates whether the app is Free or Paid
Price	Numerical (Continuous)	Price of the app (0 indicates a free app)
Content Rating	Categorical	Target age group suitability (e.g., Everyone, Teen, Mature)
Genres	Categorical	Specific genre or sub-genres associated with the app
Last Updated	Date/time	Date when the app was most recently updated
Current Ver	Categorical	Current version of the application
Android Ver	Categorical	Minimum Android OS version required to run the app

DATA QUALITY CHALLENGES & ANOMALIES

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

DATA ANOMALIES

- **Missing values** are mainly present in the Rating column, with very few missing entries in Current Ver, Android Ver, Content Rating, and Type.
- **Columns requiring cleaning or type conversion**: Reviews, Size, Installs, Type, Price, Last Updated, Android Ver
- **Columns with notable anomalies or invalid entries**:
    - Category (e.g., 1.9)
    - Rating (e.g., 19.)
    - Type (e.g., 0)
    - Installs (e.g., Free)
    - Price (e.g., Everyone)
    - Size and Android Ver (e.g., Varies with device)

DATA CLEANING & TREATMENT

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

DATA CLEANING SUMMARY

- Missing values, mainly in **Rating** and a few version-related fields, were handled using median imputation.
- Features such as **Reviews, Size, Installs, Price, Type, Last Updated, and Android Ver** were cleaned and standardized.
- Inconsistent and invalid entries (e.g., incorrect categories, unrealistic ratings, mixed text values) were identified and corrected.
- A new feature, **Minimum Android Version Required**, was created from Android version data.
- **Current Ver** was removed as it is a technical detail with limited business relevance.
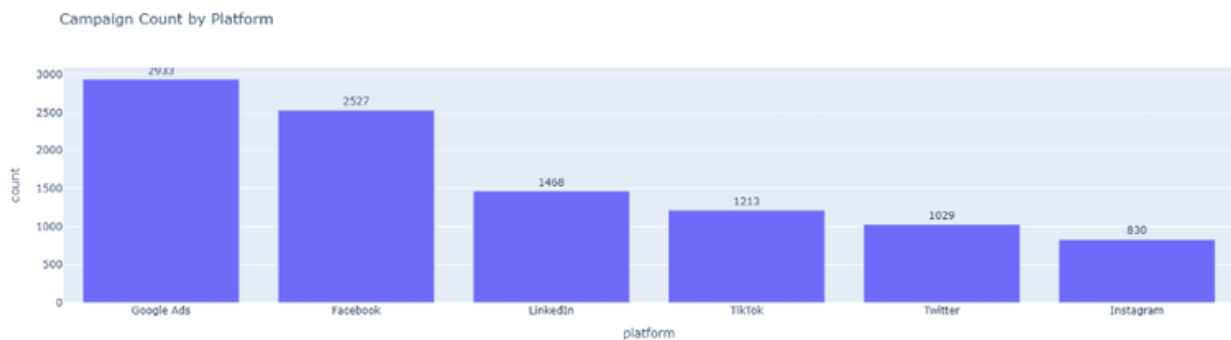
INSIGHTS

Platform
&
Market Landscape

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.
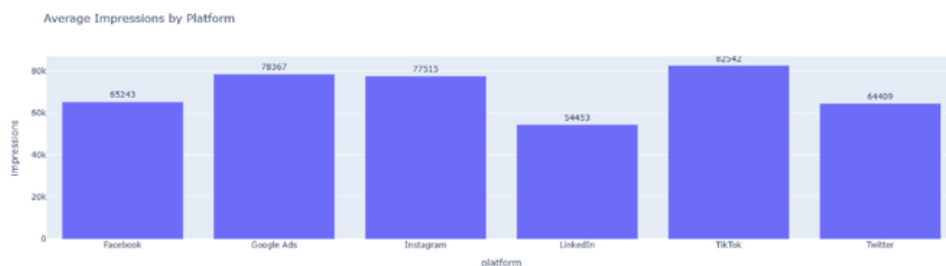- Instagram has the lowest campaign count (~830), while LinkedIn, TikTok, and Twitter fall in a mid-to-lower tier range.

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.
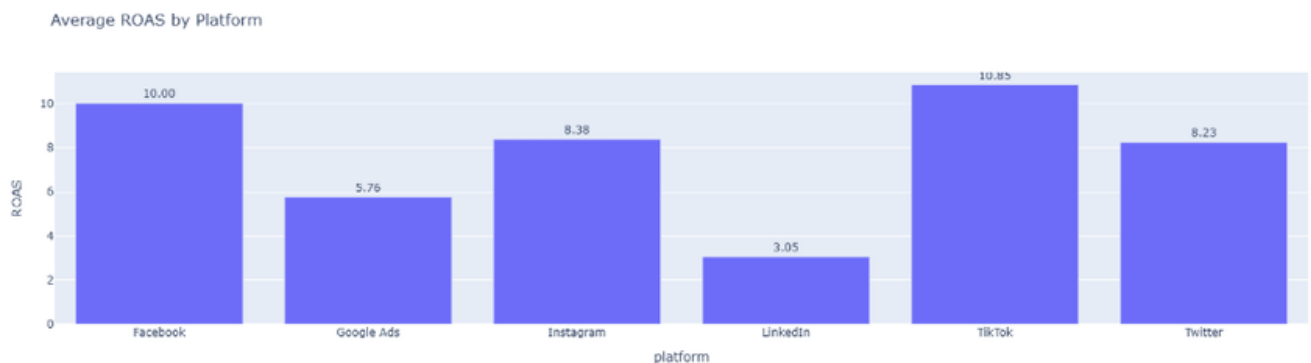
Business Insights

- 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).
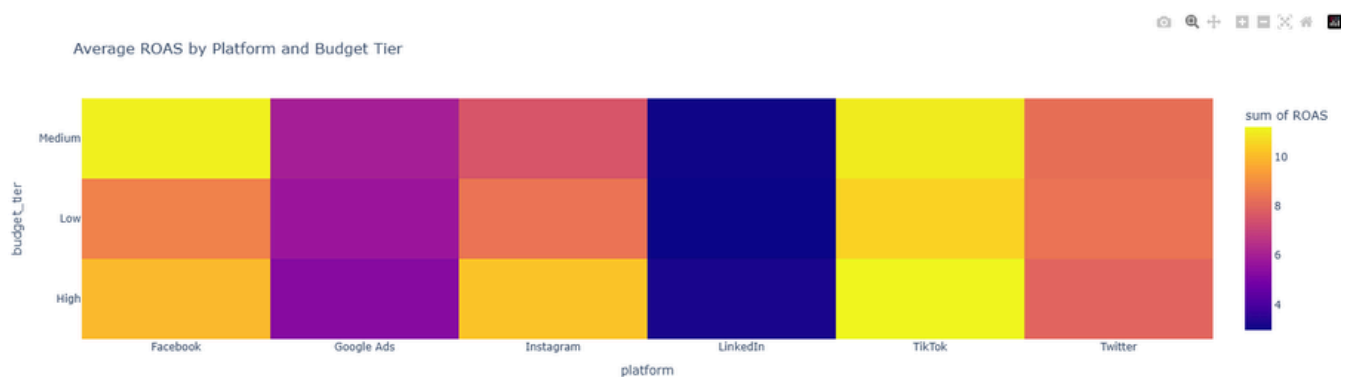- LinkedIn records the lowest average ROAS (~3.05), substantially below all other platforms.

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.
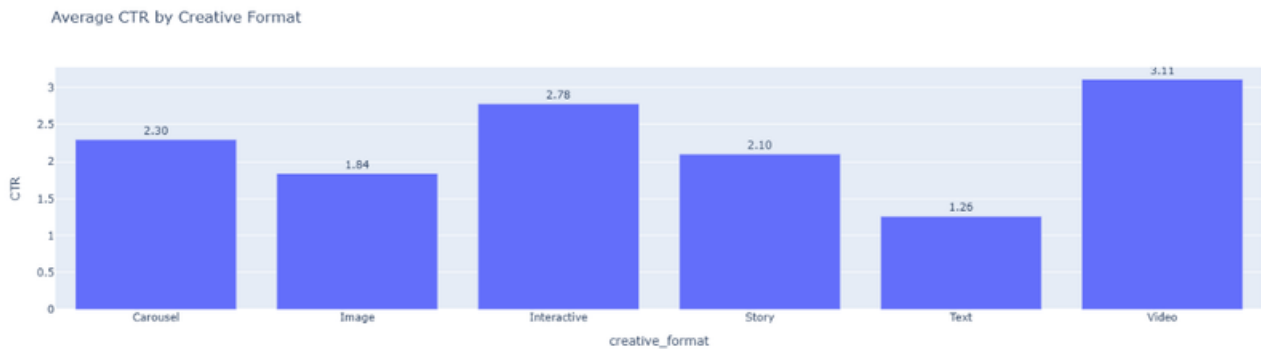
Creative Strategy
&
Engagement

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.
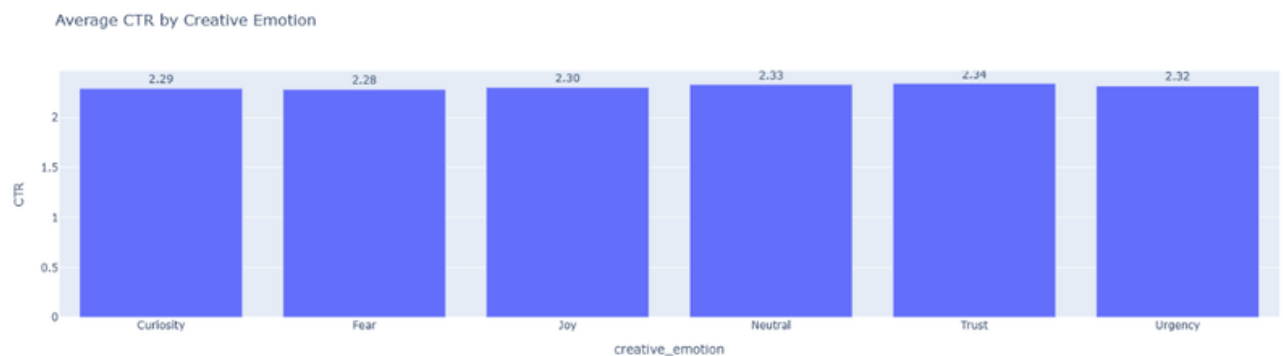
Business Insights

- 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.
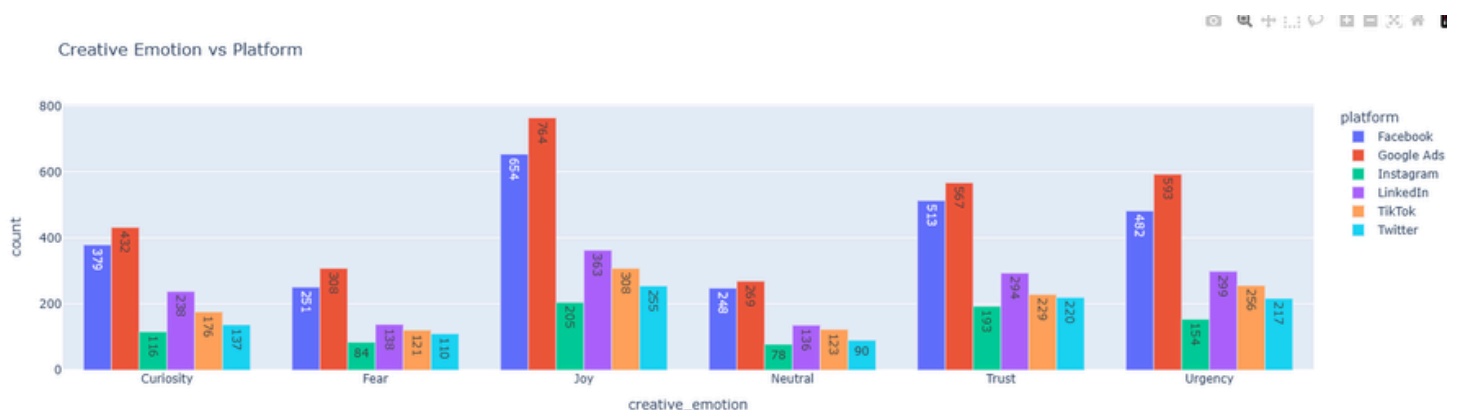
Business Insights

- 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)

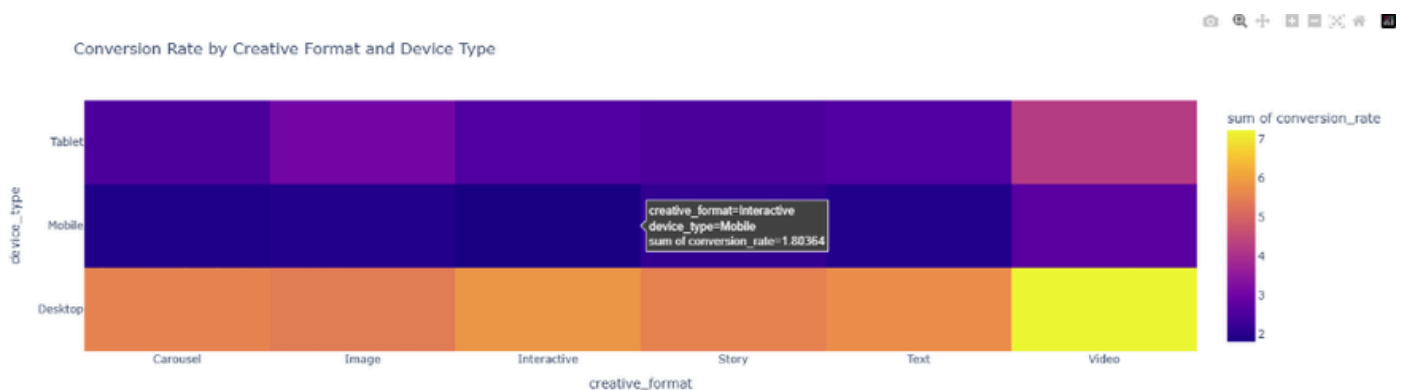Creative Emotion vs Platform

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.
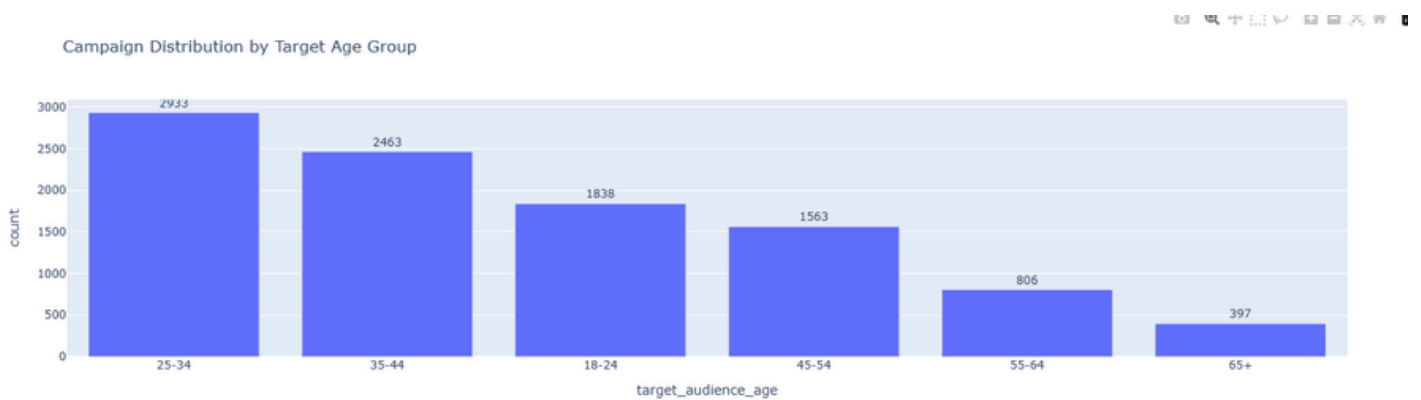
Audience Targeting
&
Intent

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.
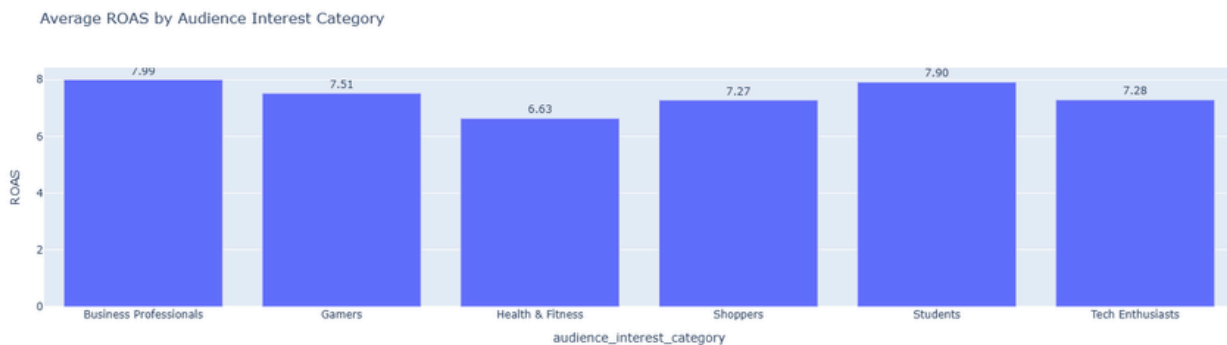
Business Insights

- 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.
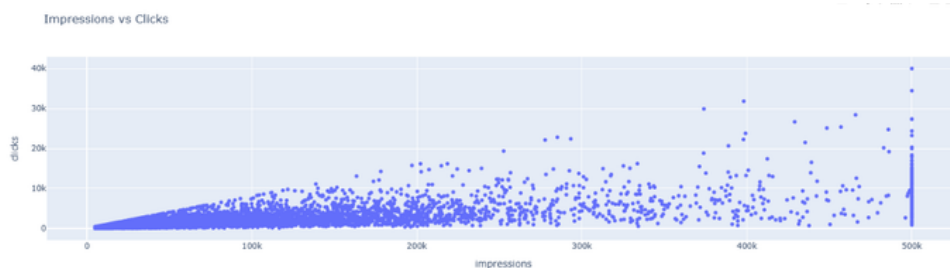
Funnel
Performance
&
Optimization

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.
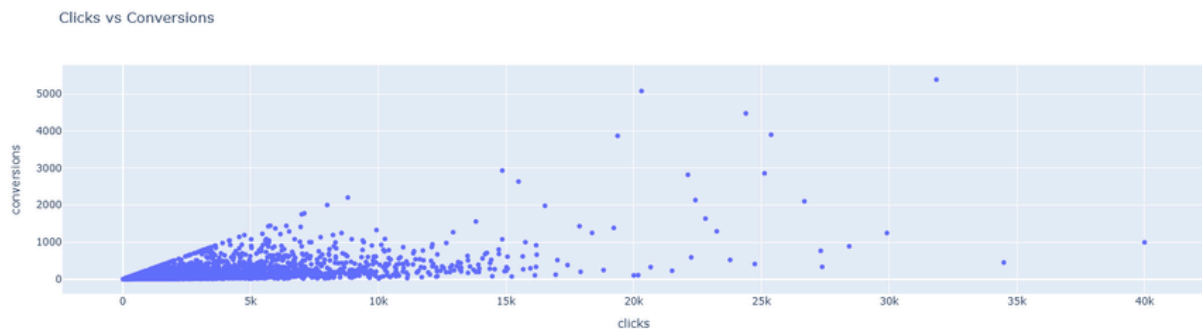
Business Insights

- 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.
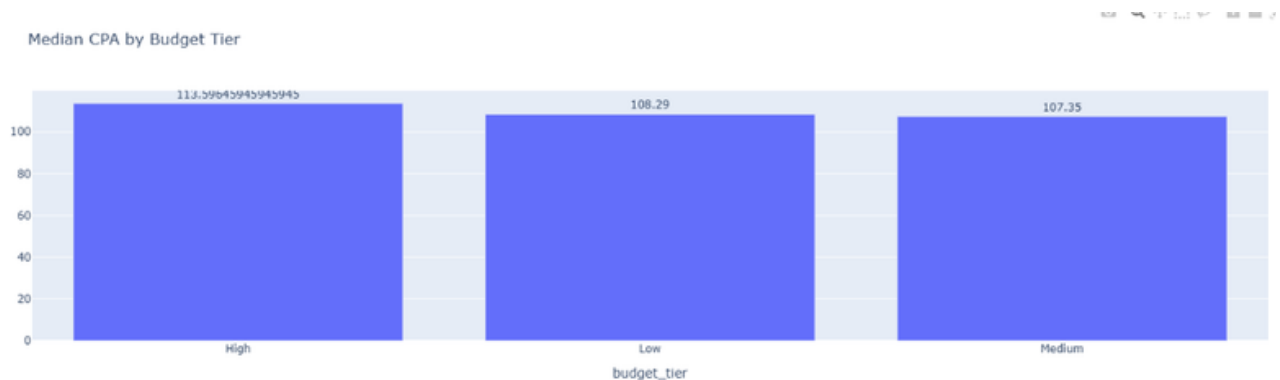
Business Insights

- 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.
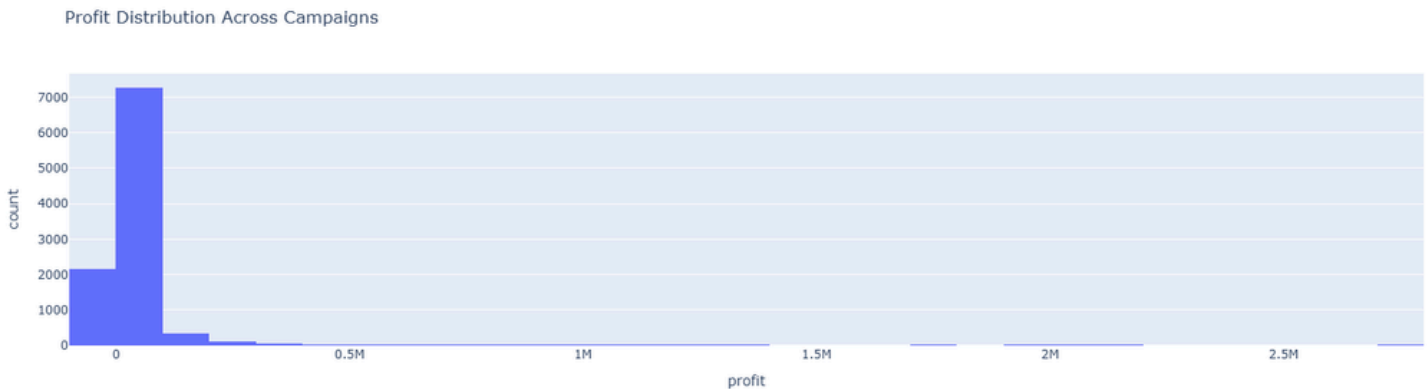
Budgeting, Profitability & Retargeting

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.
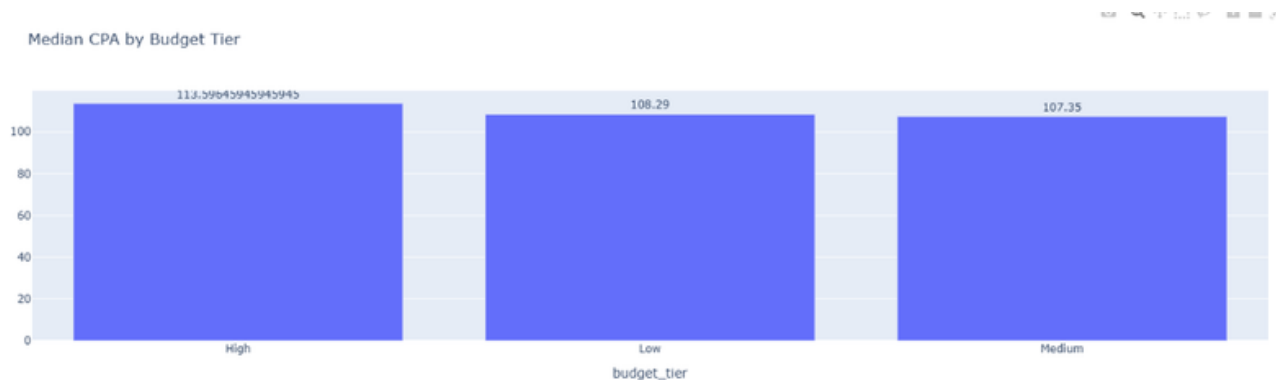
Time & Performance Stability

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

5,930

Typical Number of Reviews (Median)
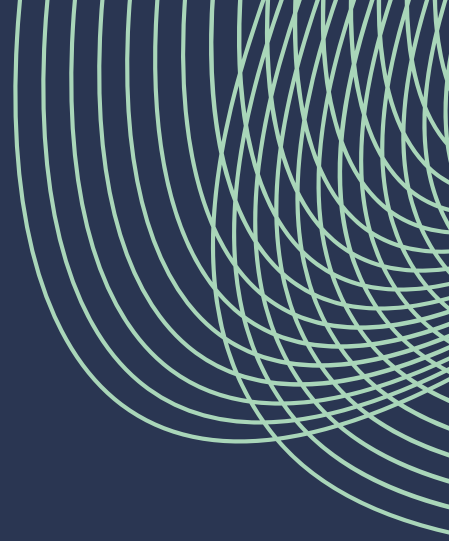

Median CPA by Budget Tier

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

- **High-Competition Categories**: Family, Games, and Entertainment dominate. Consider exploring less crowded categories like Auto & Vehicles, Events, and Beauty to stand out.

- **Free Apps Lead the Market**: Most apps are free. Freemium or ad-based models help reduce risk while maximizing reach.

- **Free vs. Paid Strategy by Category:**
  - **Free apps dominate most categories**. A free-first or freemium approach works best.
  - **Paid apps thrive in high-trust, high-value segments**. A paid- or premium-first strategy builds credibility and attracts willing buyers.

- **Target Broad Audiences:** "Everyone"-rated apps dominate. Focusing on general users helps boost downloads and ad revenue.

BUSINESS / DEVELOPER TAKEAWAYS

- **Optimize App Size & UX:** With a median size of ~14 MB and a mean of ~23 MB, balancing rich features with storage efficiency is key to avoiding download friction.

- **Ratings Require Volume & Quality:** Apps with few reviews show unreliable ratings. Early feedback helps stabilize perceived quality, while a consistent UX builds user trust.

- **Platform Maturity & Opportunity:** Ratings have stabilized over time. Leverage platform improvements to maintain quality, while exploring less crowded categories to increase visibility.

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