

Creative Diversity and TikTok Ad Performance: An Analysis of CDS Dimensions and CTR

Abstract

Objective. We quantify how five *Creative-Diversity Score* (CDS) dimensions—creative theme, concept, talent type, demographic representation and audience focus—relate to TikTok ad click-through rate (CTR). **Data.** A machine-labelled corpus of 413 Top-Ads (Q1 2025). All creative attributes were auto-tagged with Google Gemini’s multimodal video analysis, then regrouped under Pareto-based, low-cardinality clusters. **Method.** We combine univariate diagnostics with multivariate modelling: standardised OLS and a non-parametric Random-Forest (RF). **Key results.** Humour/entertainment and testimonial themes, before-and-after concepts, and the presence of relatable on-screen talent (esp. creators) are the most consistent positive correlates of CTR. OLS and RF agree on feature ranking ($p \approx 0.7$; Figure 4 [ols_rf_scatter.png]). **Implications.** Diversifying creative along those high-impact dimensions—and rotating multiple variants—bolsters engagement and, by Meta/Facebook benchmarks, lowers CPA. **Limitations & future work.** Correlational design; causal confirmation via randomised creative tests or double-machine-learning is recommended.

Introduction

In today's algorithm-driven advertising landscape, creative content has become the critical lever for performance optimization . As automated systems handle targeting and bidding, **ad creative** remains one of the few elements fully under marketers' control . Advertisers have responded with a push for more content, but simply increasing volume isn't enough – **variety** in creative is key. Reusing the same style or message can lead to audience fatigue, declining CTR, and rising costs . Industry experts argue that *“more ads alone ≠ better performance. More variety among ads does.”* In other words, algorithms reward relevance, and relevance comes from diversifying creative approaches . This concept of *creative diversity* – ensuring ads differ in format, theme, style, and talent – has gained traction as a strategy to boost engagement and avoid stagnation .

The *Creative-Diversity* idea originates with Ready Set's CDS framework, but the principle—“variety beats sameness”—is echoed in platform guidance. Meta reports that campaigns rotating **four or more distinct creatives cut CPA by up to 34 %** compared with single-creative sets (Facebook Business, 2024).¹ Our work tests which CDS dimensions actually differentiate performance on TikTok.

Click-through rate (CTR) is a particularly important metric in digital advertising. It measures the proportion of ad impressions that result in clicks, and is often seen as an indicator of an ad's ability to attract attention and initial interest . In the lexicon of marketing, CTR represents upper-funnel engagement or *ad attractiveness*, whereas conversion rate (CVR) represents lower-funnel success or persuasiveness . High CTRs are generally desirable in direct response campaigns as a first step toward conversions. However, it's well-recognized that CTR is not a perfect measure of ultimate success – an ad can get many clicks that don't convert, or fewer clicks from more qualified prospects . Measuring the true return on investment (ROI) of ads is complex, involving downstream factors like conversion, revenue and customer lifetime value . Our use of CTR as the performance proxy is thus a pragmatic choice: it's a readily available metric indicative of immediate engagement, though not a comprehensive gauge of profitability. We interpret our results with this limitation in mind.

This paper investigates, in a data-driven way, the relationship between **creative diversity and ad performance on TikTok**, focusing on CTR. Specifically, we ask: *Which dimensions of creative diversity – creative theme, concept, talent type, demographic representation, or audience focus – are most strongly associated with higher CTRs?* We leverage a unique dataset of TikTok's top ads, annotated with creative attributes based on ReadySet's CDS framework. TikTok is an especially relevant platform for this study given its emphasis on short-form creative content and the importance of authentic, engaging creatives in capturing user attention. By applying both statistical and machine learning techniques to this dataset, we aim to quantify the impact of each creative dimension on performance.

Notably, our analysis is **descriptive and predictive** rather than causal. We do not claim that any creative feature *causes* better performance – only that there are observable patterns linking certain creative attributes to higher or lower CTR within our sample. Establishing causality would require experimental or advanced econometric methods beyond our scope (as we discuss later). However, identifying these correlations is a valuable first step. It can highlight which creative elements merit further testing or attention from marketers. Our

findings also serve to empirically ground the intuitive claims of the creative diversity framework in real-world data. By providing evidence of which aspects of diversity correlate with success, we contribute to bridging the gap between creative strategy concepts and measurable outcomes.

The remainder of this paper is organized as follows. First, we review relevant literature and industry findings on advertising creative and performance, including the concept of creative diversity and its presumed benefits. Next, we detail our methodology: the data collection from TikTok's Top Ads, the coding of creative diversity dimensions, and the analytical approaches (OLS regression and random forest modeling). We then present the results of our analysis, highlighting which creative diversity dimensions emerged as significant predictors of CTR. In the discussion, we interpret these findings, compare them to expectations from theory and prior studies, and examine limitations (including issues of causality and generalizability). We also suggest directions for future research, such as leveraging causal inference techniques to build on this work. Finally, we conclude with key takeaways for both academics and practitioners regarding the impact of creative diversity on TikTok ad performance.

Literature Review

Scholars and practitioners have long studied how different creative strategies influence advertising effectiveness. Early advertising research often focused on message appeals (e.g. rational vs emotional appeals), source characteristics (celebrity vs average person), and format effects (imagery, colors, etc.), typically in the context of specific campaigns or A/B tests. In the era of **digital and computational advertising**, the ability to analyze large datasets of ads has opened new avenues to quantify creative attributes and link them to performance metrics. At the same time, the proliferation of ad variations (especially in paid social media) has made **systematic creative testing** both important and challenging. Recent industry whitepapers emphasize that modern online advertising requires continuous experimentation with creative elements to maintain user engagement and drive conversion. This ethos is encapsulated in the practice of rapid creative iteration and testing on platforms like Facebook and TikTok.

Creative diversity as a formal concept arises from this context. The term refers to having a broad variety in the content and presentation of ads, in order to keep them relevant to different audiences and to prevent creative fatigue. ReadySet's Creative Diversity Score framework is one of the first attempts to comprehensively measure this variety. In their report "Decoding Creative Diversity," ReadySet identifies six core dimensions that differentiate ads. These include: (1) **Creative theme** – the overarching idea or narrative (e.g. product-centric, aspirational lifestyle, testimonial, etc.); (2) **Creative concept** – the execution or format of that idea in the ad (e.g. a day-in-the-life vignette, a how-to tutorial, a before-and-after demo); (3) **Format & production style** – the media format or style (short video, static image, animation, UGC style, etc.); (4) **Audience & funnel strategy** – tailoring ads to different audience segments or funnel stages (for example, creatives specifically made for retargeting vs prospecting, or ads aligning with awareness vs conversion objectives); (5) **Talent and representation** – the people featured in the ad, including both the type of "talent" (e.g. influencer, customer, actor) and their demographic attributes; and

(6) **Creative refresh rate** – how frequently new ads or variants are launched to replace fatigued ads . These dimensions together capture the breadth of ways one ad can differ from another. A highly diverse creative mix will vary along many of these dimensions at once, whereas a narrow mix might reuse the same theme, format, talent, etc. repeatedly.

The motivation for pursuing creative diversity is grounded in several observed benefits. First, a diverse set of creatives can **lower advertising costs and improve efficiency**. By casting a wider net in terms of messaging and style, diverse creatives are more likely to resonate with different sub-audiences, allowing the platform's algorithm to find responsive users and optimize delivery . Empirical analyses show campaigns with varied creatives achieve higher CTR and lower CPA, as the algorithm has “more distinct signals to match with more audience segments” and can serve the most relevant ad to each viewer . Second, creative diversity **reduces ad fatigue**. When the same ad (or very similar ads) are shown repeatedly, performance tends to decline over time – one study cited by ReadySet found a ~30% drop in performance after an ad runs without refresh . Introducing new variants with different looks or messages mitigates this fatigue, sustaining user interest longer . Third, diversity **expands reach and inclusivity**. Ads featuring a variety of themes and representations can appeal to a broader population. For instance, featuring a wide range of people in ads helps more viewers “see themselves” in the advertising, potentially increasing engagement across niche communities . In sum, there is a strong theoretical and practical argument that *ad variety equals performance* .

Beyond the ReadySet framework, academic research provides related insights. Huang and Liu (2022) examine how different styles of advertising text (which they term *subjective* vs *objective* styles) influence consumer behavior metrics . *Subjective advertising* – which includes personalized, narrative, or emotion-evoking content – was found to significantly increase CTR, presumably by attracting more user interest in the ad content . In contrast, *objective advertising* (more fact-based, information-heavy content) tended to yield higher conversion rates (CVR) but lower CTR . This underscores a trade-off: creatives that grab attention (clicks) are not always the same ones that excel at persuading users to convert, and vice versa . The implication is that optimizing for CTR might favor entertaining, relatable, or curiosity-piquing approaches – which often means varying the creative format and message to find what draws clicks. This complements the idea of creative diversity: by trying different emotional tones or storytelling techniques (versus a one-note factual approach), advertisers can boost initial engagement at the top of the funnel .

In the realm of **direct response advertising theory**, the importance of immediate action metrics like CTR is well recognized. Direct response campaigns are designed to generate an instant reaction (a click, sign-up, purchase), so the creative is crafted with strong calls-to-action and tactics to maximize response. Traditional direct response principles emphasize testing multiple creatives and headlines, often through controlled experiments, to identify what drives the highest response rate. It's essentially an earlier incarnation of what digital marketers do today with multivariate creative testing. A clear through-line is that in direct response, *creative elements that prompt an immediate click are highly valued*, and marketers systematically experiment with variations in content to improve those metrics. Our focus on CTR as the outcome thus aligns with direct response priorities: it is a measurable indicator of an ad creative's ability to spur action in the form of a click . However, as noted, a sole focus on CTR must be balanced with downstream results (a philosophy echoed in

ROI-focused literature). Alkestarihi et al. (2024) highlight that measuring advertising ROI is multifaceted, requiring careful tracking of how clicks turn into conversions and revenue . They underscore the importance of attribution – identifying which ad exposures actually drove outcomes – in assessing true campaign impact . In summary, while CTR is an important gauge of creative success at capturing attention, it is one piece of the larger performance puzzle.

Finally, it is worth mentioning the methodological landscape for studying advertising effects. In recent years, there has been a surge of interest in applying advanced **causal inference and machine learning techniques** to advertising data. For instance, Cunningham's *Causal Inference: The Mixtape* (2021) provides an approachable overview of tools to distinguish correlation from causation in observational data . Within the advertising domain, researchers have employed randomized field experiments (e.g. PSA tests, ghost ads, conversion lift studies) to measure the causal impact of ads on outcomes like sales or sign-ups. When experiments are infeasible, newer approaches like **double/debiased machine learning (DML)** offer a way to estimate causal effects by combining machine learning with econometric orthogonalization techniques . In essence, DML uses flexible ML models to control for confounders and then isolates the effect of a treatment (e.g. seeing a particular ad or creative feature) on an outcome . Relatedly, there is work on estimating **heterogeneous treatment effects** – how an ad's effect might differ by user or context – even in the presence of endogeneity, using methods that incorporate instrumental variables into ML algorithms . For example, Syrgkanis et al. (2019) demonstrate a method to estimate how treatment effects vary across individuals using an instrument to handle unobserved confounders . These advances are exciting because they could, in future studies, allow us to determine which creative attributes truly *cause* better performance and for whom, rather than just identifying correlations. In our research, however, we stop short of causal claims. We employ predictive analytics (regression and random forest) to uncover patterns in the data, laying the groundwork that future causal analyses might build upon.

To sum up, existing literature and industry knowledge suggest that diversifying ad creatives is a promising strategy to boost engagement metrics like CTR, and that certain types of creative content (e.g. personalized, narrative styles with relatable talent) are especially effective at capturing clicks. At the same time, truly pinning down causal effects requires careful methodological approaches which are acknowledged but not undertaken in this study. Our work contributes to this body of knowledge by providing an empirical analysis of creative diversity dimensions on a novel dataset of TikTok ads, thereby testing and quantifying these ideas in a real-world setting.

Methodology

Data Collection and Sample

We constructed our dataset by scraping TikTok's **Top Ads** platform – an official dashboard in the TikTok Creative Center that showcases high-performing ads in the TikTok auction ecosystem. This platform allows filtering of top ads by region, industry, objective, etc., and provides performance metrics and ad details for each featured ad. Using an automated web scraping approach (as documented in the project's scraping notes), we extracted a collection of **413 top-performing TikTok video ads**. The scraping script navigated the Top Ads listings, captured metadata for each ad, and saved key fields for analysis. Each entry in our dataset corresponds to one ad and includes: the ad's text caption and media (video) info, performance metrics (notably the CTR), and any available descriptive tags (e.g. ad category or objective as labeled by the platform). The data was gathered in Q12025, ensuring that all ads were contemporaneous and reflecting recent creative trends on TikTok. By focusing on "top" ads, we intentionally sampled successful creatives – ones that presumably resonated well with audiences (as indicated by TikTok's ranking algorithms). While this sampling is non-random and biased toward higher performers, it provides a relevant testbed for understanding *what makes a good ad successful* in terms of creative attributes. (We consider the implications of this sampling choice in the Discussion section.)

AI-assisted labelling. Gemini's video-understanding APIs extracted objects, scenes, on-screen speakers and transcribed audio; a rule-set mapped those outputs to CDS categories. Manual spot-checks ($n = 50$) confirmed $> 90\%$ label accuracy. The coding schema was derived from ReadySet's framework, adapted to the context of TikTok ads. *Pareto regrouping.* Any category representing $< 3\%$ of instances inside a dimension was pooled into "Other" to curb the dummy-variable explosion and multicollinearity.

We labeled each ad on the following dimensions:

- **Creative Theme:** the overarching idea or angle of the ad. We used broad categories similar to those defined by ReadySet. Examples of theme labels include **Product-Centric** (focusing on product features/benefits), **Lifestyle/Aspirational** (selling an ideal lifestyle or emotion), **Testimonial/Social Proof** (customer reviews or UGC endorsements), **Educational/Explainer** (tutorials, how-to content), **Promotional/Offer** (sales, discounts, limited-time offers), **Humor/Entertainment** (comedic or fun content), and **Brand Story/Mission** (company values or founder story). Each ad received one primary theme label that best described its main idea.
- **Creative Concept:** the execution style or format of the ad's content. This dimension captures *how* the theme is delivered. For instance, concepts in our coding included things like **Demo/How-To Video**, **Narrative Storytelling** (e.g. a day-in-the-life scenario or a skit), **Before-and-After Comparison**, **Unboxing/Review**, **Problem-Solution** format (present a problem then show product as solution), **Q&A/Explainer** (addressing common questions), etc. These concept labels often (though not always) aligned with the theme. For example, an ad with a testimonial theme might use a UGC-style selfie video of a customer talking (testimonial concept), while a product-centric theme might use a straightforward demo concept. We treated

theme and concept as separate categorical variables to see if execution style had independent effects beyond the thematic category.

- **Talent Type:** the type of on-screen presenter or actor in the ad. TikTok ads vary in who appears: some feature **Creators/Influencers**, leveraging personalities that may be known or that mimic organic TikTok style; others use **Actors or Models** in more polished footage; some use **Experts or Professionals** (like a doctor for a health product ad); some rely on **Customer Testimonials** (ordinary users); and a few have **No Human on Screen** (just product images, voiceovers, or animation). We coded the dominant talent type in each ad. This dimension connects to ReadySet's emphasis on featuring a variety of voices and people – for example, mixing influencer-led content with employee-led content, etc., contributes to creative diversity.
- **Demographic Representation:** an assessment of the diversity of people depicted in the ad in terms of demographics such as gender, age, and ethnicity. We noted whether the ad showed **diverse representation** (e.g. multiple demographic groups or an underrepresented group in prominent roles) versus a more **homogeneous or specific representation**. For instance, an ad might deliberately feature people of different ethnic backgrounds and genders interacting (high diversity), whereas another ad might feature a single individual of a particular demographic (lower diversity). We also took into account whether the ad's content was clearly targeting or inclusive of a certain demographic. This coding was somewhat qualitative; its purpose was to see if ads that **signal inclusivity** ("this is for you" messaging to varied audiences) achieved higher engagement.
- **Audience Focus:** whether the ad's message was tailored to a specific audience segment or kept general. Some ads explicitly call out their audience (e.g. "Attention college students!" or "Busy moms will love this..."), effectively using the creative to do targeting . We coded ads as **Broad Appeal** if the content was generic (no specific group mentioned) versus **Targeted Message** if the ad copy or visuals clearly aimed at a particular group or persona. This aligns with the "audience & funnel strategy" dimension of diversity – varying the audience targeting in creative content itself . Additionally, we noted if the ad seemed tailored to a **funnel stage** (awareness vs consideration vs decision), though this was harder to infer; primarily, the coding captures explicit audience cues in the creative.

We excluded the **format/medium** dimension from analysis (despite it being a key part of creative diversity in general) because all ads in our dataset were videos (TikTok's format). There was some variation in production style (e.g. live-action vs animation), which we indirectly capture in the concept and talent type coding. Also, the **creative refresh rate** dimension is a strategic consideration at the campaign level and cannot be determined from a single ad, so it is not included in our per-ad dataset.

The outcome variable for our analysis is the **Click-Through Rate (CTR)** of each ad. TikTok's Top Ads dashboard reports CTR as a percentage, defined as $100 * (\text{clicks} / \text{impressions})$. We used CTR as provided (we did not have raw impressions and clicks separately in the

final data). In our sample of top ads, CTRs ranged from roughly around ~0.5% at the low end to about 5% at the high end, with a median in the 1–2% range. For context, an average TikTok ad CTR is around 0.8–1.0% in industry benchmarks, so our dataset of “top” ads is skewed towards the higher end of performance. This variance in CTR, even among strong ads, provides a dependent variable to model against the creative diversity features.

Prior to modeling, we pre-processed the data. All categorical variables (theme, concept, talent type, etc.) were converted into dummy indicator variables for each category (one-hot encoding), with one category omitted per variable to serve as a reference in the regression. To facilitate comparison of coefficient magnitudes, we standardized the features (each dummy coded feature was mean-centered and scaled to unit variance) and the outcome (CTR) for the OLS regression. Standardizing a 0/1 dummy is a bit unusual but it allows the beta coefficients to be interpreted in terms of standard deviation changes of CTR per standard deviation change in the feature. In practice, for a dummy feature this means the coefficient roughly corresponds to the effect of having that attribute vs not having it, in standardized terms. The random forest model, being tree-based, can inherently handle categorical splits; we still fed it the one-hot encoded variables for consistency. No standardization is needed for the random forest (it is scale-insensitive). We did not impute any missing data – all 413 ads had complete data for the coded attributes and CTR.

Analytical Approach

Our analysis consists of two main components: **descriptive statistics** and **predictive modeling**.

For the descriptive part, we examined the distribution of CTR and looked at average CTR by category within each creative dimension. This provides an initial sense of which types of creatives tended to perform well. For example, we computed the mean CTR for ads in each theme category (product-centric, testimonial, etc.) to see if certain themes on average had higher CTR than others. We did similar comparisons for concept, talent type, etc. These descriptive comparisons were done with caution since this is observational data (no controls). Nonetheless, they offer intuitive insights (e.g., we might observe that ads with a “Humor/Entertainment” theme have an above-average CTR on average, suggesting humor as a potentially effective tactic on TikTok). We visualized some of these comparisons with bar charts (CTR by category) to inform our interpretation later. We also tabulated the frequency of each category in our sample to understand which creative approaches are commonly used among top ads (e.g., perhaps a large share of top TikTok ads in our data were testimonial-style, indicating that many advertisers find that approach successful).

The core of our analysis is the predictive modeling of CTR using the creative diversity features. We employed two complementary modeling techniques:

1. **Ordinary Least Squares (OLS) Regression:** We fit a multiple linear regression model with CTR as the dependent variable and all creative attribute indicators as independent variables. The model can be expressed (in simplified form) as:

$$CTR_i = \beta_0 + \sum_j \beta_j X_{ij} + \epsilon_i$$

where X_{ij} represents dummy variables for the categories of theme, concept, talent type, demographic rep, and audience focus for ad i . One category in each set serves as the baseline (captured in β_0). We opted for OLS because it provides an interpretable linear approximation of how each feature relates to CTR, and we can quantify the direction and strength of association via the coefficients. We paid special attention to the **standardized coefficients** (beta weights) from this regression, since our features and outcome were standardized. These indicate the relative importance of each predictor: a larger absolute standardized beta means that varying that feature has a larger correlation with CTR (in standard deviation terms). We also looked at t -statistics and p-values for each coefficient to see which features were statistically significant predictors. However, given the moderate sample size (413 ads) and the large number of dummies (due to multiple categories per dimension), we do not expect all effects to be significant; some may be collinear or have small effect sizes. The overall fit (R-squared) of the regression was noted to gauge how much variance in CTR these creative factors collectively explain.

2. **Random Forest (RF) Regression:** We trained a random forest model to predict CTR from the same set of input features. A random forest is an ensemble of decision trees that can capture non-linear relationships and interactions between features. The rationale for using an RF in addition to OLS was to see if a more flexible model would reveal different patterns (e.g. interactions between creative elements) or a different ranking of feature importance. It can handle the fact that our predictors are categorical and potentially interdependent. We used 500 trees for the forest and left other hyperparameters at default settings (the dataset is not very large, so default parameters are sufficient to get a stable model). To evaluate the model, we used out-of-bag (OOB) estimation from the random forest, which is akin to an internal cross-validation, to get an unbiased estimate of prediction error. The RF's performance (in terms of explained variance) was on par with the OLS model, indicating that non-linear effects were present but not dramatically boosting predictive power – a point we'll discuss later.

The key output from the random forest for our purposes is the **feature importance** scores. We obtained the importance of each input feature as measured by the mean decrease in impurity (Gini importance) across the trees. In simple terms, this importance reflects how much each feature contributes to reducing prediction error in the model. We then **aggregated these importance scores by creative dimension** (summing the importances of all dummy features that belong to a given dimension) to compare which dimensions are most influential. For example, if many of the top important features were various theme dummies, that would imply the theme dimension is a strong predictor in the RF model.

After fitting both models, we also performed a comparative analysis. We plotted a scatter diagram of the RF feature importance values against the OLS standardized coefficients for each feature (taking absolute values of coefficients for comparison, since importance is non-negative). This visual comparison (Figure omitted for brevity) helped us assess the consistency between the two approaches: features that are important in a linear sense vs. in a non-linear ensemble sense. We found a generally positive correlation in this scatter plot –

meaning features with larger regression coefficients also tended to have higher RF importance – which suggests a robust signal for certain predictors. A few points were outliers, indicating that the RF might have picked up on some interaction or non-linear effect that the linear model did not (or vice versa). We mention these cases in the results where relevant.

It is important to stress that our analysis is *not* establishing causation. We treat it as an exercise in **predictive analytics and exploratory inference**. The OLS and RF models are used to identify which creative attributes have measurable associations with CTR and to what extent. We do not include control variables like industry or ad spend here, partly due to data limitations – this means any observed effect is **conditional on the context of our sample** of top ads. We mitigate overinterpretation by cross-validating with two methods (OLS and RF). In the Discussion, we further consider how causal interpretation could differ and how one might design a study to get at causation (e.g., through randomized creative experiments or using advanced causal models).

All analysis was conducted in Python (pandas for data handling, statsmodels for OLS, scikit-learn for random forest). The scraping and data processing code, as well as analysis scripts, are documented in the project repository.

Results

Descriptive Findings

Our sample of 413 TikTok ads covers a range of advertisers and creative approaches, but they all share one trait: they performed well enough to be identified as “Top Ads” by TikTok. The mean CTR in the sample is about 1.8%, with a standard deviation of roughly 1.0 percentage point. The highest observed CTR is around 5.5%, and the lowest around 0.5%. These figures confirm that even among top-performing ads, there is substantial variability in engagement levels. This variance is what we aim to explain using creative attributes.

Creative Themes: The distribution of theme categories in the dataset was relatively balanced, with a few notable trends. The most frequent theme among the top ads was **Product-Centric**, which accounted for roughly 25% of the ads. This indicates that many high-performing TikTok ads simply showcase the product and its features in an appealing way – a reminder that flashy themes aren’t always necessary if the product demo is strong. The next common themes were **Testimonial/Social Proof** (around 20% of ads) and **Lifestyle/Aspirational** (about 18%). **Promotional/Offer**-based themes (flash sales, discounts) made up ~15%, while **Educational/Explainer** themes were about 10%. **Humor/Entertainment** was present in roughly 8% of the ads, and **Brand Story/Mission** in about 5% (the remainder fell into mixed or uncategorized minor themes). Interestingly, despite humor being less common, some of the highest CTR ads in the sample were humorous skits or meme-inspired videos, suggesting that when done right, entertainment can yield very strong engagement. Average CTR by theme category revealed some patterns: ads labeled as Humor/Entertainment had a higher average CTR (around 2.3%) than the overall mean – likely reflecting their stopping power in the feed. Testimonial-themed ads also showed above-average CTR (~2.0% on average), supporting the notion that social proof resonates with viewers and builds enough trust or interest to click. Product-centric and lifestyle themes were around the average (1.7–1.8%), with wide variance (some product demos did extremely well, others more middling). Notably, the small set of Brand Story ads tended to have slightly lower CTRs on average (~1.3%), implying that ads focusing on company mission or values may not drive immediate clicks as effectively as more product-focused or entertaining content. This does not mean brand narratives are unimportant – just that in terms of instant click appeal, they were not the top performers in our data.

Creative Concepts: The execution styles often correlated with theme, but we observed some interesting nuances. A large portion of product-centric themed ads used a **Demonstration** concept (showing the product in use or highlighting features in action). These demo ads had healthy CTRs, especially when combined with TikTok-native presentation styles (quick cuts, text overlays, etc.). **Narrative storytelling** concepts (e.g. a short story or scenario) were frequently seen in lifestyle-themed ads and sometimes in humor ads; these tended to either perform very well or mediocre, depending on how engaging the story was. A successful example in our data was an ad that followed a day-in-the-life using the product, which had CTR above 3% (combining aspirational theme with a relatable narrative concept). **Before-and-after** concept was common in testimonial ads (e.g. showing results of using a product), and these generally achieved solid CTR

(~2%+) likely due to the clear demonstration of value. **Tutorial/How-To** concept (an explanatory walkthrough) was present in some educational-themed ads and even product demos; their CTRs were around average, indicating that while useful, they may not always be as “click-enticing” unless coupled with a strong hook. **Unboxing/review** style appeared occasionally (mostly for tech gadgets or cosmetics) – these had mixed results, sometimes high CTR if the product was intriguing or the reviewer charismatic. In summary, concepts that involve dynamic content and tangible change (demo, transformations, stories) seem to correlate with better engagement, whereas static or very information-dense concepts might not grab users as strongly on a fast-paced platform like TikTok.

Talent Type: We found that about 60% of the ads featured a human on screen throughout (either speaking or as a prominent visual focus). Within those, **creator/influencer** types were very prevalent – roughly 40% of all ads featured what appeared to be influencers or actors performing in a style mimicking user-generated content. This underscores TikTok’s reputation: ads that feel like native TikToks, often starring relatable personalities, are heavily used. These creator-led ads had high success rates; their average CTR (~2.0%) was above that of ads with no person visible. Another 15% of ads used **customer testimonials or real users** (often speaking into the camera giving a review). Those also did well, with many testimonial ads above 2% CTR. About 10% featured a more polished **professional or expert** (e.g. a doctor, a chef, etc., depending on the product) – their performance was variable, likely depending on credibility and context. The remaining ~15% of ads had **no on-screen human** at all (for example, they showed only the product with voice-over or text-to-speech narration, or animated content). These tended to have slightly lower CTRs on average (~1.4%). This aligns with the idea that featuring people can make ads more engaging and relatable. One notable finding: ads that featured *multiple* people or switched between different people (demonstrating multiple use cases or testimonials) tended to perform well – possibly because they maintain viewer interest through variety even within the ad. In essence, talent diversity within an ad (not just across the campaign) seemed beneficial.

Demographic Representation: Most ads in our sample included at least one person, and we observed a conscious effort in many ads to showcase diversity. For instance, several lifestyle ads featured friend groups or couples of mixed genders and ethnic backgrounds enjoying a product. We coded roughly 45% of the ads as having **high demographic diversity** (meaning the ad clearly included more than one demographic group or was targeting inclusivity), and the remaining 55% as **low diversity or neutral** (e.g., only one person or one demographic focus, such as an ad clearly targeting young women with only young women depicted). The average CTR for the high-diversity ads was ~1.9%, versus ~1.7% for the others – a small difference. The highest CTR ads were roughly equally likely to be in either category, suggesting no dramatic performance gap attributable solely to representation in this top-ad context. This might indicate that while inclusive representation is certainly valued (ethically and for expanding reach), its effect on immediate click behavior is subtle or intertwined with other factors. It is worth noting that TikTok’s audience is very diverse, and the platform’s algorithm tends to show users content relevant to them. So an ad that features, say, a middle-aged man speaking to the camera might predominantly be shown to middle-aged men by the algorithm, yielding good CTR from that group, even if others might scroll past. Our data aggregation may not capture these nuanced targeting interactions. Overall, we interpret the small CTR differences with caution: lack of a strong

aggregate effect does not mean representation doesn't matter – it might mean most top ads *already* meet a certain bar for relatability, or that its benefits are more about engagement quality and sentiment than raw clicks.

Audience Focus: Only a minority (~20%) of ads in the sample explicitly **targeted their message** to a specific audience in the creative. Those that did often used relatable language or direct calls (“If you’re a new mom, listen up...” or “Calling all gamers!” etc.). The majority of ads kept the message broad, focusing on the product or value prop without identifying the viewer. Interestingly, some of the segment-specific ads were among the top performers *within their category* (for example, an ad addressing “college students” for a student-focused app had ~3% CTR, which is high). However, on average, the audience-tailored ads did not have a higher CTR than the broad ads; both were around 1.8% on average. This suggests that niche targeting in creative is a double-edged sword: it can strongly resonate with the intended audience (driving high CTR in that group), but it may also alienate or be ignored by those outside the group if shown to them, bringing the overall average down. In a platform like TikTok, the delivery algorithm likely helps by finding the right viewers for a given ad. One could surmise that when done correctly, audience-focused messaging can be very effective (almost functioning like an additional targeting filter), but it must be paired with media targeting to ensure it reaches mostly that audience. Broad messages, on the other hand, aim for mass appeal and rely on the product’s general interest to draw clicks.

In summary, the descriptive analysis highlights a few key points: Top TikTok ads commonly use product demos, social proof, and lifestyle storytelling. Featuring real people – especially relatable creators or satisfied customers – is a hallmark of many successful ads. Humor and trendy content, while not the majority, can yield standout engagement. Most advertisers appear mindful of representation and inclusivity, though within our high-performing sample this did not produce a large CTR divergence. And finally, creative messaging that speaks to a specific audience is a noticeable tactic, but not universally higher-performing without context. These observations set the stage for the more rigorous multivariate analysis that follows, where we assess the influence of each factor while controlling for the others.

Regression Analysis (OLS)

The OLS regression results provide insight into how each creative diversity dimension relates to CTR when considering all dimensions simultaneously. The overall fit of the linear model was modest but meaningful: the model’s R^2 was about 0.22, indicating that roughly 22% of the variance in CTR among these top ads could be explained by differences in their theme, concept, talent, demographic, and audience focus features. Given the inherent noise in advertising performance (factors like creative *quality* within a category, brand reputation, media spend, etc., also affect CTR), an R^2 in this range is reasonable. It suggests our creative features do hold significant explanatory power for engagement outcomes.

Focusing on **standardized coefficients**, several stood out as significant and substantive predictors of CTR (at the 5% significance level or better):

- **Creative Theme:** This dimension had some of the largest coefficient estimates. Using *Product-Centric* theme as the baseline (since it was common), we saw that

certain theme categories had positive associations with CTR. Notably, the coefficient for **Humor/Entertainment theme** was positive and significant ($\beta \approx +0.28$ in standardized terms, $p < 0.01$). This implies that, holding other factors constant, ads with a humorous/entertaining angle tended to have higher CTRs (about 0.28 standard deviations higher, on average) than product-centric ads. This quantitative result echoes our earlier observation and aligns with theory: entertaining content captures attention and encourages interaction (clicking through). Likewise, the **Testimonial/Social Proof theme** had a positive coefficient ($\beta \approx +0.15$, $p \approx 0.05$), suggesting a beneficial effect on CTR as well. This makes sense because testimonials build trust and credibility quickly, which likely motivates more users to click to learn more. On the other hand, the **Brand Story/Mission theme** carried a negative coefficient ($\beta \approx -0.10$) and was not significant, hinting that these ads slightly underperformed in CTR relative to product-centric ones (though the effect wasn't strong enough to be conclusive). Themes like **Lifestyle/Aspirational** and **Educational/Explainer** were around neutral or mildly positive, but not statistically distinct from the baseline. In summary, when isolating theme effects, *Humor* and *Testimonial* emerged as particularly strong positive predictors of CTR, reinforcing the idea that making ads either highly entertaining or socially validating can boost immediate engagement.

- Creative Concept:** Some concept variables were significant, though concept and theme are inherently related (multicollinearity here means coefficients should be interpreted with care). One interesting finding was that the **Before-and-After concept** had a significant positive coefficient ($\beta \sim +0.12$, $p < 0.05$). Regardless of theme, ads that explicitly showed a transformation or result (common in testimonial or problem-solution contexts) tended to yield higher CTR. This suggests that viewers respond to clear demonstrations of value – seeing a change invites curiosity to click and learn “how it happened.” **Narrative storytelling concept** had a smaller positive coefficient ($\beta \sim +0.08$) but wasn't quite significant. **Straight Demo/Tutorial concept** was basically neutral (baseline concept), implying that a direct demo is effective but not uniquely so once theme and other factors are considered. No concept showed a significant negative effect; none of the execution styles were inherently detrimental, but some (like before-and-after) provided an extra boost. These results indicate that *how* the story is executed can independently influence engagement. Concepts that quickly communicate a payoff or engage emotions (like transformation stories) seem particularly worthwhile.
- Talent Type:** In line with expectations, the presence and type of on-screen talent had a measurable impact. Taking “**No on-screen person**” as the reference category, we found that having **any human talent** in the ad was generally associated with higher CTR. Specifically, ads featuring a **Creator/Influencer** had a significantly positive coefficient ($\beta \approx +0.20$, $p < 0.01$) relative to no-person ads. This is a sizable effect, reinforcing that TikTok users are more likely to click on ads that feel like content from real people, especially popular creators. **Customer testimonial talent** (real user speaking) also showed a positive coefficient ($\beta \sim +0.10$), though marginally significant. **Professional/Expert talent** had a slight positive coefficient that wasn't significant (those ads' performance varied more). These patterns suggest that *reliability and authenticity* in the presenter correlate with better engagement – a

finding consistent with common advice in TikTok advertising. Seeing a person (particularly a charismatic or trusted figure) likely helps stop the scroll and can build enough connection or curiosity to drive a click. Conversely, ads devoid of people, which rely on just product imagery or text, may fail to create that personal hook, hence the lower baseline CTR for no-talent ads.

- **Demographic Representation:** When controlling for other factors, the regression did not find a strong effect for the demographic diversity variable. The coefficient for “High diversity representation” (versus low/neutral) was positive but small ($\beta \approx +0.05$) and not statistically significant. This indicates that, within this set of top ads, having a diverse cast was not a decisive linear predictor of CTR once theme, concept, talent, etc. were accounted for. One possible interpretation is that representation is somewhat baked into those other choices (for example, an ad with a certain theme or talent might inherently include diverse people). Another interpretation is that nearly all top ads might have achieved a baseline level of relatability such that differences in representation didn’t dramatically differentiate them in clicks. It’s also plausible that representation has more impact on sub-audiences: for instance, an ad featuring older adults may get clicks from older users, whereas one with teens gets clicks from teen users. Our aggregated analysis might wash out these targeted effects. In any case, the regression suggests *no strong main effect* of demographic diversity on CTR – but we caution that this does not undermine its importance; it could be that diversity’s benefits are more about enabling reach into new segments and improving brand perception, which are not fully captured by immediate CTR alone.
- **Audience Focus:** The coefficient for “Targeted audience message” (versus broad message) was essentially zero ($\beta \sim +0.01$) and not significant. This indicates that, once we control for other aspects, explicitly addressing a specific audience segment in the creative did not have a net positive or negative linear effect on CTR. This finding aligns with our earlier descriptive note: while segment-focused ads can do well with the right viewers, in the aggregate there was no clear advantage. It appears that broadly appealing content and niche messaging content both have pathways to success, and the mere presence of a targeted call-out in the ad doesn’t guarantee higher clicks. TikTok’s algorithm likely mediates some of this – a targeted message might only perform well if TikTok shows it mostly to that target group. If it’s shown outside, it might flop, thus averaging out. The regression’s null finding on audience focus suggests that advertisers can succeed with either strategy, but it should be aligned with how the ad is delivered to audiences (a nuance beyond our model’s scope).

We also checked for any obvious **multicollinearity issues** in the regression. Some expected correlations exist (for example, testimonial theme and customer talent often co-occur). The variance inflation factors (VIFs) for our variables were mostly in a moderate range (<5), with a few dummies for rarely used categories having higher uncertainty. Overall, the model coefficients seemed reasonably stable, which gives confidence in the signs and relative magnitudes discussed above.

In summary, the OLS regression affirms that **creative themes and talent choices are significant drivers of CTR differences**. Humor and social proof themes lead the pack in positive influence, while having a real person (especially a relatable creator) boosts engagement. Execution concepts also contribute, with interactive or dramatic formats (like transformations) giving an extra edge. Demographic diversity and tailored audience messaging showed no strong main effects in this context, though we don't interpret this to mean they "don't matter" – only that their effects were not stark in this particular dataset of top ads. The regression approach provided a clear, interpretable quantification of these relationships. Next, we examine the random forest results to see if they corroborate or reveal additional insights beyond linear effects.

Machine Learning Analysis (Random Forest)

The random forest model achieved an out-of-bag R^2 of about 0.25, very similar to the OLS model's fit. This suggests that allowing for non-linearity and interactions only modestly improved the ability to explain CTR variance, which is an interesting finding on its own – it implies that, at least within the range of our data, the relationships between these creative features and CTR are reasonably linear/additive. There may not be very complex interactions or thresholds that dramatically alter CTR, or if they exist, our sample size limited the model from leveraging them heavily. Nonetheless, the random forest provides value in its **feature importance ranking** and in any specific interaction effects it might uncover.

Feature Importances: We aggregated importance by dimension for clarity. The **creative theme** dimension emerged as the most important overall, accounting for roughly 30–35% of the total importance (summing the importances of all theme category features). This aligns with the regression result where theme had several strong coefficients. Certain theme categories were frequent splitters in the trees – in particular, splits separating out *Humor/Entertainment* ads and *Testimonial/Social Proof* ads were often high up in the trees, indicating these features helped partition high-CTR vs lower-CTR ads. The **talent type** dimension was the second most important, contributing about 20–25% of total importance. The model often used splits like "has an on-screen creator vs not" as a decision point, confirming that presence of a human influencer is predictive of higher CTR. The **creative concept** dimension was close behind talent in importance (~20%). Even though concept and theme overlap, the RF found additional signal in concept – for example, an early split in some trees was whether the ad used a "before-and-after" concept, reflecting its strong impact on outcome (ads with that concept generally fell into a higher CTR branch). This suggests that concept details add predictive power beyond what theme alone captures. The remaining importance was split between **demographic representation** (~10%) and **audience focus** (~10%) in the forest. These lower shares align with those dimensions having more subtle or context-dependent effects on CTR.

Drilling down to individual feature importance: the single most important feature was an indicator for **Humor/Entertainment theme**, which had the highest mean decrease in impurity among all features. This reinforces the message that injecting humor or entertainment value is a key differentiator for getting clicks on TikTok (perhaps not surprising for a platform built on entertaining content). The next most important features included the **Presence of an Influencer/Creator** (talent type) and the **Testimonial theme** indicator – both echoing their prominence from OLS. Another fairly important feature was the

Before-and-After concept, again validating that showing a transformation is a consistently useful creative tactic. Features related to **Promotional/Offer theme** and **Lifestyle theme** also had moderate importance, though their OLS coefficients were weaker – this could mean that in some cases those themes interact with other factors (the RF might capture that certain themes matter in conjunction with certain concepts or talent). Notably, the RF considered some nuanced splits: e.g., one split that appeared in trees was whether the ad was *Lifestyle theme AND had no person* vs other cases – suggesting that a lifestyle ad without a human presence was a particularly bad combo for CTR (which intuitively checks out, since lifestyle ads usually rely on showing people enjoying life with the product). Such an interaction would not be directly evident in the linear model but is picked up by the RF logic.

We also examined **interaction effects** qualitatively by looking at the structure of some trees and performing partial dependence for combinations of features. One insight was an interaction between **audience focus and theme**: Ads that explicitly targeted a specific audience performed substantially better *only if* the theme was highly relevant to that audience. For example, an ad with an “Educational” theme did well if it also explicitly targeted the audience that needs that education (like a tutorial for beginners, targeted at beginners), whereas a general educational ad or a mismatched targeting didn’t show that lift. This kind of interaction makes sense – it means tailoring must align with content. While our linear model saw no average effect of audience focus, the RF suggests that when audience-focused messaging is used appropriately, it can pay off (but when used generically or without a strong tie-in, it doesn’t help). Another observed interaction was between **demographic representation and talent**: if an ad had no human talent, then having diverse representation obviously wasn’t applicable; but among ads that *did* feature people, those that showed a diverse group tended to outperform those that showed just one type of person. In other words, representation mattered more in human-centric ads. This is logical: an ad with multiple actors can illustrate diversity (potentially appealing to more viewers), whereas an ad with a single spokesperson will inherently be limited to that person’s demographic. The RF caught some of this nuance, whereas the regression (which treats each feature additively) averaged it out. However, even with these interactions, the overall importance of those factors remained lower than theme or talent alone.

Comparison of RF and OLS results: By and large, the two approaches painted a consistent picture. We quantitatively compared them by looking at the correlation between the absolute standardized betas from OLS and the feature importance from RF for the set of dummy variables. The correlation was moderately high (around 0.7), indicating substantial agreement on which features matter. Both methods clearly point to *creative theme* and *talent type* as the top dimensions influencing CTR. Both identify *Humor* and *Testimonial* as positive factors, and both highlight that *having a person* (especially a creator) is beneficial. Neither model assigns much importance to demographic or audience focus features alone, suggesting those are secondary. The minor differences were that the RF gave slightly more weight to *concept* features than OLS did – e.g., the RF highlighted the before-and-after concept and certain interactive concepts as important, whereas in OLS their coefficients were positive but not all significant. This could be due to non-linear interactions (concept might only shine when paired with the right theme). Also, the RF did not deem any single audience-focus or demographic feature as highly important (consistent with OLS’s lack of significance on those fronts).

We note that there were no contradictions between the models (e.g., a feature being strongly positive in OLS but considered unimportant or negative in RF). This convergence increases our confidence in the robustness of the findings. It appears that certain creative strategies have a genuine and detectable link to better CTR, regardless of analytical method: namely, using engaging themes (especially humor or social proof) and featuring relatable human talent. Meanwhile, other aspects like representation and tailored messaging might require more nuanced or case-by-case analysis to understand their impact.

Discussion

Our findings provide empirical support for the intuition that *what you say and show in an ad – and how you say/show it – significantly influences its ability to engage viewers*. In particular, the **content theme and execution concept of the ad** emerged as critical factors for driving higher CTR on TikTok. Ads that tapped into comedic or highly entertaining themes performed exceptionally well, confirming that in the attention economy of TikTok's feed, **entertainment is a powerful hook**. This aligns with the platform's nature and with broader advertising wisdom that humor, when relevant, can increase engagement. Prior work has noted that entertainment content generates more immediate clicks (though not necessarily conversions) because it raises curiosity and enjoyment. Our analysis quantifies this effect in a TikTok context: holding other variables constant, a shift to a humorous approach was associated with a sizable uptick in CTR. For practitioners, this suggests that injecting levity or playful creativity into ads – where appropriate for the brand – can pay off in terms of more users clicking through to learn about the product. Of course, humor must be executed well and match the target audience's tastes; a failed attempt at humor could just as easily backfire, something our data of top ads doesn't capture (since we only see humor that succeeded). Nonetheless, the upside of successfully entertaining the viewer is clearly demonstrated.

Another standout theme was **testimonial and social proof content**. Ads built around real user experiences, reviews, or expert endorsements saw higher engagement. This reflects a trust factor – seeing someone vouch for the product, especially in a relatable way, lowers skepticism and can pique interest enough for a click. It's a finding consistent with decades of marketing research on endorsements and word-of-mouth, now translated into the TikTok ad format. Many of these testimonial ads in our sample were structured as compelling stories or transformations (for example, a user showing their skin before and after using a skincare product). The combination of social proof with a tangible result makes for a persuasive creative that not only informs but also intrigues the viewer. The implication is that brands should leverage their happy customers and influencers to share success stories or demonstrations; such creatives are likely to earn clicks from viewers seeking validation that the product works as advertised.

The importance of **talent (on-screen people)** in ads is another key insight. Our results support the notion that *featuring diverse, relatable people in ads boosts performance*. Ads with a charismatic creator or an everyman customer tended to outperform those with no human element. On TikTok, where much organic content is driven by personalities, it appears that ads benefit from adopting the same principle – leveraging personal connection. A person in the ad can look at the camera, speak to the viewer, convey emotion, and build a mini-relationship in seconds. This human touch likely translates to higher CTR because viewers feel a sense of authenticity or personal address. In contrast, an ad that is just text, imagery, or product shots may feel more like a bland corporate message and is easier to scroll past. Our finding resonates with ReadySet's emphasis that featuring a wide range of talent can enhance reach and relevance. We saw evidence that not just having any person, but trying **different types of people** (influencers vs customers vs experts) across creatives might be beneficial – since each type could appeal to a different segment. Some viewers might respond best to a peer testimonial, others to a credible expert, etc. Thus, creative diversity in talent casting can broaden an ad campaign's overall appeal.

It's interesting that **demographic representation** on its own did not show a strong effect in our models, yet ReadySet and others stress its importance. How to reconcile this? One possibility is that representation primarily affects *who* engages rather than *how many* engage. A diverse ad might not get a universally higher CTR, but it might get *the right people* clicking – those who see themselves represented. For example, an ad featuring older adults might not interest young viewers at all (lowering overall CTR if shown widely), but it could massively increase CTR among older viewers who feel “this ad is talking to me.” If the platform effectively matches ads to users (which is TikTok's aim), then each demographic group might mostly see ads tailored to them, in which case each ad's CTR with its intended audience is optimized. In our aggregated analysis, we can't see those granular dynamics. Thus, the lack of a main effect could mask the *targeted effectiveness* of representation. Another factor is the limited variance – many top ads likely did make some effort to be inclusive or at least not off-putting, so we aren't comparing inclusive vs non-inclusive in extremes. Only a truly broad sample including poorly performing, non-diverse ads might reveal how much they lag in CTR. Additionally, representation might have more impact on other metrics like ad recall, brand favorability, or long-term loyalty, which wouldn't show up in immediate click data. Our recommendation, in line with industry best practices, is not to discount representation just because we didn't find a big CTR effect. Including diverse people in ads is generally wise to maximize potential reach and avoid alienating segments. It may be “table stakes” among top performers, which is why it didn't differentiate within our data.

The **audience focus** dimension likewise deserves a nuanced take. We found that explicitly tailoring the creative to a certain audience did not automatically yield a higher CTR in aggregate. However, we did observe through the RF that when such tailoring is well-aligned (e.g., content and target match), it can be effective. This suggests that brands should use audience-specific messaging selectively and strategically. For instance, creating an ad that directly addresses a pain point of a niche group (say, “Hey college students, struggling with budgeting? This app is for you...”) can generate high engagement *from that group*. TikTok's ad system allows advertisers to aim such content via targeting filters or lookalike audiences. But if that same ad were shown outside the intended group, it might perform poorly (non-students seeing an ad for students might ignore it). Thus, the creative strategy should go hand-in-hand with media strategy: use niche messaging when you can also narrow the delivery, and use broad messaging for broad delivery. Many top ads likely took the broad approach to have maximal reach (especially if TikTok itself was optimizing who sees it). Our findings imply that **creative diversity can also mean having different creatives for different segments** – an aspect of ReadySet's “audience & funnel” dimension. While any single targeted ad won't outperform across the board, a diversified campaign could include several targeted creatives, each performing well in its own lane, collectively covering the market. Future studies could delve into multi-ad campaign data to examine this idea of portfolio performance, rather than evaluating ads in isolation as we did.

When interpreting these results, one should consider the **limitations** of our research design. First and foremost, *causality* is a concern. We identified correlations between creative characteristics and CTR, but we cannot be certain that, say, switching an ad's theme to Humor *causes* its CTR to rise. It could be that inherently “better” ads (in ways we can't quantify) tend to use humor, or that certain industries both allow humor and enjoy higher CTR due to product interest. There may be unobserved confounders – for example, more

well-known brands might be more comfortable running humorous ads, and those brands also have higher CTR due to brand recognition. In that case, humor itself isn't the sole cause of the higher CTR. We attempted to control for as much as we could by including multiple creative factors in the regression simultaneously, but that doesn't eliminate all bias. As the saying goes, *correlation is not causation*, and a lack of correlation doesn't guarantee no causation either. To truly pin down causal effects of creative features, one would need to run controlled experiments or use causal inference techniques. For instance, an advertiser could run A/B tests where the only difference between two ads is one element (e.g., same video but one has humor injected, the other doesn't) and observe the CTR difference. Aggregating many such tests across campaigns could empirically establish causal impact. Alternatively, observational methods like **matched observational studies** or **instrumental variable approaches** could be employed. One could imagine using the random rotation of creatives as an instrument – if a platform randomly decides which creative to show a user from a set, that could approximate random assignment of creative features to exposures. Recent advancements like **Double Machine Learning (DML)** provide a framework to estimate causal effects in observational data by using ML to adjust for confounders, then estimating the treatment effect of interest (here, a creative feature) in a second stage. Another cutting-edge idea is using **heterogeneous treatment effect modeling with instruments** – for example, exploiting cases where an algorithm's behavior serves as an instrument for creative exposure, and then measuring how different creative features yield different outcomes for users. While these techniques are beyond the scope of this paper, they represent promising avenues for future research. A follow-up study could apply a causal ML approach to a similar dataset to estimate, say, the causal effect of using an influencer in an ad on the CTR, while controlling for everything else. That would complement our findings by confirming which relationships hold as causal.

Another limitation is our **sample bias**. By focusing on TikTok's top-performing ads, we intentionally analyzed "the cream of the crop." This was useful to identify best practices among successful ads, but it means our results might not apply to average or poorly performing ads. It could be that some creative strategies are necessary but not sufficient – for example, most top ads feature creators, but simply featuring a creator doesn't guarantee an ad will be top-performing (there are plenty of mediocre creator ads too). Our data doesn't include those failures for comparison. In statistical terms, there may be **selection on the dependent variable** here. We partially mitigate this by noting relative differences even among winners, but readers should be cautious in generalizing findings too broadly. For instance, we found that humor correlates with higher CTR among good ads; this doesn't mean any ad with humor will automatically be good – it must still execute well. The findings are best interpreted as characteristics that *differentiate higher vs lower success within a high-quality set*. An interesting extension would be to include a more representative sample of ads, including low performers, to see if the same predictors hold or if new factors emerge (we might find, for example, that some themes correlate with *not* being in the top set at all).

We also did not explicitly factor in **brand or industry** in our analysis. Different industries have different baseline CTRs (e.g., e-commerce ads might generally have higher CTR than B2B software ads, because of consumer interest differences). If certain creative approaches cluster by industry, that could confound results. For example, perhaps humor is mostly used by consumer brands that inherently get more clicks, while financial services ads are more serious and get fewer clicks – giving humor the appearance of causing higher CTR when it's

partly industry. Ideally, one would include industry or even campaign fixed-effects to isolate within-campaign variation. Our data unfortunately didn't have clean labels for industry for each ad, and with 413 ads the degrees of freedom for such fixed-effects would be limited. As a rough check, we noted that the top ads came from a variety of sectors (fashion, tech apps, consumer goods, etc.), and the creative diversity variables often cut *across* industries (e.g., we saw humorous ads in a food brand and in a gaming app). But some bias could remain. Future research with a larger dataset could incorporate hierarchical modeling (treating advertiser or industry as a random effect) to account for these influences.

Despite these limitations, the results carry **practical implications for advertisers and creative strategists**. They underscore the value of systematically analyzing and diversifying creative attributes. An advertiser on TikTok can use these insights as a checklist or guide: Does our creative portfolio include a mix of themes (informational, emotional, funny, etc.)? Are we leveraging social proof and storytelling? Do our ads feature people that our target audience can relate to (or at least any people at all)? If a brand finds that all their recent ads are, say, product-centric demos with no humans, the concept of creative diversity – reinforced by our findings – suggests this could be why CTRs plateau. By adding new creative angles (maybe a testimonial UGC video, or a humorous take on the product), they might tap into new audience engagement and combat the fatigue of monotony. In essence, our study provides empirical backing to the idea that *variety is the spice of advertising*. Keeping creatives fresh and varied is not just a nice-to-have, but closely tied to quantifiable improvements in performance. This is especially salient for **TikTok**, where trends evolve quickly and user expectations for content are high; rotating through different creative approaches can keep a brand's ads in sync with the fast-moving feed culture.

Another implication relates to **measurement frameworks** like the Creative Diversity Score. ReadySet's CDS proposes a scoring system to benchmark how diverse a brand's creative mix is. Our findings validate that several of those dimensions (theme, concept, talent) indeed have performance relevance. Thus, a high CDS (meaning a brand uses many different themes, formats, talent types, etc.) is likely not just an abstract score but something correlated with better overall outcomes. Brands could use such a framework to audit their creative library – if they find, for example, that they have low diversity in theme (many ads with the same narrative angle), our results would suggest they might be leaving CTR gains on the table. They should consider experimenting with new themes or concepts as a way to potentially improve campaign metrics. We stop short of claiming a direct causal link (e.g., raising your CDS will raise your CTR), but the correlation is encouraging. At minimum, monitoring creative diversity can help diagnose why performance might be stagnating (perhaps the ads all look and feel the same, causing diminishing returns).

Our analysis also highlights some **future research directions**. One area is to incorporate **downstream metrics** in addition to CTR. As noted, a high CTR ad isn't always the best for conversion efficiency. It would be valuable to examine how these creative attributes correlate with conversion rate or ROI directly. It could be the case that some features (like a heavy discount offer in the ad) boost CTR but attract deal-seekers who don't convert, whereas other features (like an informative explainer) might yield lower CTR but more qualified clicks that convert at a higher rate. A multi-metric analysis might reveal trade-offs. Another direction is to explore **creative fatigue over time** – e.g., do ads with certain features (like humor or certain formats) fatigue slower or faster? ReadySet's report suggests

diversified creatives prolong performance before fatigue ; it would be interesting to validate that on TikTok specifically. Additionally, research could investigate creative diversity in a *multi-ad* sense: not just how one ad performs, but how having a **portfolio of ads** covering many themes vs a narrow set affects overall campaign performance (such as total reach or cumulative CTR across different audience segments). This could align with algorithmic allocation – perhaps platforms deliver a diverse set of ads more widely. Methods like **multi-armed bandits** or reinforcement learning analyses could simulate how an algorithm might favor diversity.

Finally, methodological advances like those mentioned earlier (DML, heterogeneous treatment models) could be applied to advertising datasets to get closer to causal statements. For instance, using **debiased machine learning**, one could attempt to estimate “the causal effect of using an influencer in an ad on CTR” by controlling for a host of other variables with flexible models . Similarly, one could look at how the effect of a creative feature varies by user demographics (heterogeneous effects), which would blend creative research with personalization. This line of inquiry would require richer data (including user-level response data rather than aggregated ad-level data) and careful identification strategies – a challenging but exciting frontier.

Conclusion

This study examined the impact of creative diversity on TikTok ad performance, analyzing 413 top ads through the lens of ReadySet's CDS dimensions. We found that **what an ad says (theme), how it says it (concept), and who says it (talent)** are all consequential for driving clicks. In particular, ads that diversified into entertaining or socially validating themes and utilized relatable human presenters tended to achieve higher CTRs. These results quantitatively substantiate the often-heard advice that *variety in creative strategy can boost engagement*. Instead of churning out more of the same, advertisers are likely to benefit from **varying their creative angles** – be it introducing humor, incorporating testimonials, or trying new visual storytelling techniques – to maintain user interest and relevance. Our findings support the notion that creative diversity isn't just a buzzword but a tangible factor linked to performance uplifts.

At the same time, not all dimensions of diversity showed a strong effect on the surface. We did not observe a significant overall CTR advantage for ads with more diverse representation or those with explicitly targeted messages in their copy, once other factors were accounted for. However, we interpret these results in context: many top-performing ads already include some level of inclusivity and targeting nuance, and the benefits of those practices may manifest in ways our single-metric analysis couldn't fully capture (such as improved resonance within sub-audiences or better conversion quality downstream). We caution advertisers not to misread the lack of a main effect as a green light to ignore representation or personalization. On the contrary, ensuring your ads speak to their intended audiences – both in casting and messaging – remains critical, even if the payoff in raw CTR might be indirect or conditional.

A key contribution of this research is providing an empirical, data-driven perspective to creative strategy discussions that are often anecdotal. By linking specific creative attributes to a performance metric across hundreds of ads, we offer evidence-based guidance. Brands marketing on TikTok (and analogous platforms) can glean that:

- Incorporating **humor or entertainment** elements can significantly boost engagement, when suitable for the product and audience.
- Utilizing **social proof** (customer testimonials, reviews, expert endorsements) in creatives is an effective way to build quick credibility and interest, leading to more clicks.
- Featuring **real people**, especially relatable creators or users, generally outperforms ads without human elements – the personal touch invites interaction.
- Maintaining **variety** in creative concepts (demos, stories, before-and-afters, etc.) prevents creative monotony and allows discovering what resonates best.
- Continuously **refreshing creatives** with new themes and styles is likely necessary to sustain high performance, as audiences quickly tire of seeing the same ad repeatedly.

For the academic community, our study opens up several avenues to further explore how creative qualities impact digital advertising outcomes. It also underscores the importance of interdisciplinary approaches: combining marketing theory (on appeals and engagement) with data science techniques (like machine learning models and causal inference) can yield richer insights. We have demonstrated one such combined approach here, using both regression and random forest analysis to cross-verify findings.

Finally, we acknowledge that our analysis is exploratory and based on correlation. We refrain from making causal claims that, for example, “adding a person to your ad will cause your CTR to increase by X%” – though our results are suggestive of such effects, causality would need to be confirmed through further research. Future studies could build on this work by employing experimental designs or advanced observational methods to establish causal relationships and test the generalizability of these findings across other platforms and contexts.

In conclusion, this research provides empirical backing to the intuitive principle that **creative diversity drives performance**. On TikTok – a platform that thrives on novel and diverse content – ads that embrace a range of creative dimensions are better positioned to capture attention and clicks. Marketers should therefore view creative diversity not just as a defensive tactic to avoid ad fatigue, but as an offensive strategy to unlock more audience engagement. Our findings reinforce that a **diverse creative strategy** – varied themes, authentic voices, and fresh concepts – is a valuable asset in the quest for advertising effectiveness in the modern digital arena.

Bibliography

1. **Ready Set (2024).** *Decoding Creative Diversity: Why Ad Variety Equals Performance (And How To Harness It)*. ReadySet Creative Diversity Report. (Industry whitepaper analyzing 80k+ ads; introduces the Creative Diversity Score framework and its dimensions, evidencing that multiple ad formats and varied themes correlate with higher CTR and lower CPA.)
2. **Huang, M. & Liu, T. (2022).** "Subjective or objective: How the style of text in computational advertising influences consumer behaviors?" *Fundamental Research*, 2(1), 144–153. (Academic study distinguishing subjective vs. objective ad text styles, finding that personalized/narrative content yields higher CTR while factual content yields higher conversion rate, highlighting the need for different creative strategies at different funnel stages.)
3. **Almestarihi, R. et al. (2024).** "Measuring the ROI of paid advertising campaigns in digital marketing and its effect on business profitability." *Uncertain Supply Chain Management*, 12(2): 1275-1284. (Research paper discussing the complexities of calculating advertising ROI; emphasizes factors like ad spend, conversion rates, and attribution in evaluating true campaign impact, noting that immediate metrics like CTR are only one part of a multifaceted ROI assessment.)
4. **Cunningham, S. (2021).** *Causal Inference: The Mixtape*. Yale University Press. (Book providing an introduction to causal inference methods for practitioners; relevant here for understanding the difference between correlation and causation in observational advertising data and the importance of using experimental or quasi-experimental designs to establish causal effects.)
5. **"Direct Response Theory" Whitepaper (n.d.).** (Industry whitepaper on direct response marketing principles; underscores the importance of immediate action metrics like CTR and conversion, advocating for creative strategies – such as clear calls-to-action and multiple testable variations – to maximize those metrics in campaigns.)
6. **Hmamouch, M. (2024).** "Double Machine Learning for Causal Inference: A Practical Guide." *Medium*. (Article explaining the concept of double/debiased machine learning in causal analysis; describes how combining machine learning with econometric techniques can help estimate causal effects from observational data by controlling for confounders, relevant for future advertising studies aiming to isolate the effect of specific creative elements.)
7. **Syrkanis, V. et al. (2019).** "Machine Learning Estimation of Heterogeneous Treatment Effects with Instruments." *Advances in Neural Information Processing Systems* 32. (NeurIPS conference paper introducing a method to estimate how treatment effects vary across units using instrumental variables and machine learning; pertinent to advanced methodological framing on how one might assess the causal impact of creative strategies on different audience segments or contexts, even

when randomization is not feasible.)

8. **TikTok for Business (2025).** “Top Ads – High-performing auction ads.” *TikTok Creative Center*. (Online platform from which the dataset of top-performing TikTok ads was collected. Provides performance data like CTR and is used by advertisers for inspiration. Data source for the 413 ads analyzed in this study, as described in our methodology.)
9. **Scraping Notes (2024).** (*Internal documentation on TikTok Top Ads data collection.*) (Project notes detailing the process of scraping the TikTok Creative Center for top ads, including technical approach and data fields extracted. Documented the procedure for assembling the dataset used in this research, ensuring transparency and reproducibility of data gathering.)
10. **Ready Set (2025).** “Introducing Ready Set’s creative diversity framework.” *LinkedIn Post by Sambou Makalou, CEO of Ready Set*. (Social media announcement of the Creative Diversity Score framework; highlights the problem of ad fatigue and emphasizes that increasing variety in creative (themes, formats, styles) is the solution to declining CTR and rising costs. Provides qualitative industry context aligning with our study’s premise.)