

Automated Data Analysis Report

Data Summary

dataset_overview:

total_rows: 1000
total_columns: 18
memory_usage_mb: 0.6

column_information:

user_id:

data_type: int64
non_null_count: 1000
null_count: 0
null_percentage: 0.0

statistics:

mean: 555.96
median: 568.0
std_dev: 259.61
min: 100
max: 999
q1: 327.0
q3: 774.75
unique_values: 610

timestamp:

data_type: object
non_null_count: 1000
null_count: 0
null_percentage: 0.0
unique_values: 1000

top_5_values:

2025-11-18 22:20:05: 1
2025-05-09 15:31:25: 1
2025-06-08 13:08:27: 1
2025-01-23 18:52:03: 1
2025-10-31 22:55:38: 1

device_type:

data_type: object
non_null_count: 1000
null_count: 0
null_percentage: 0.0
unique_values: 3

top_5_values:

Desktop: 355
Mobile: 338

Tablet: 307

location:

data_type: object
non_null_count: 1000
null_count: 0
null_percentage: 0.0
unique_values: 5

top_5_values:

India: 213
Germany: 210
USA: 201
UK: 192
Canada: 184

age_group:

data_type: object
non_null_count: 1000
null_count: 0
null_percentage: 0.0
unique_values: 5

top_5_values:

25-34: 217
45-54: 214
35-44: 196
18-24: 192
55+: 181

gender:

data_type: object
non_null_count: 1000
null_count: 0
null_percentage: 0.0
unique_values: 2

top_5_values:

Female: 509
Male: 491

ad_id:

data_type: object
non_null_count: 1000
null_count: 0
null_percentage: 0.0
unique_values: 948

top_5_values:

A5205: 3
A5630: 2
A1728: 2
A2631: 2

A8094: 2

content_type:

data_type: object
non_null_count: 1000
null_count: 0
null_percentage: 0.0
unique_values: 3

top_5_values:

Video: 337
Image: 335
Text: 328

ad_topic:

data_type: object
non_null_count: 1000
null_count: 0
null_percentage: 0.0
unique_values: 5

top_5_values:

Travel: 220
Fashion: 214
Electronics: 195
Health: 191
Automotive: 180

ad_target_audience:

data_type: object
non_null_count: 1000
null_count: 0
null_percentage: 0.0
unique_values: 5

top_5_values:

Family Oriented: 206
Travel Lovers: 203
Fitness Lovers: 201
Tech Enthusiasts: 197
Young Adults: 193

click_through_rate:

data_type: float64
non_null_count: 1000
null_count: 0
null_percentage: 0.0

statistics:

mean: 0.08
median: 0.08
std_dev: 0.04
min: 0.01

max: 0.15

q1: 0.04

q3: 0.12

unique_values: 1000

conversion_rate:

data_type: float64

non_null_count: 1000

null_count: 0

null_percentage: 0.0

statistics:

mean: 0.08

median: 0.08

std_dev: 0.04

min: 0.01

max: 0.15

q1: 0.04

q3: 0.11

unique_values: 717

engagement_level:

data_type: object

non_null_count: 1000

null_count: 0

null_percentage: 0.0

unique_values: 4

top_5_values:

Ignored: 272

Liked: 257

Shared: 240

Commented: 231

view_time:

data_type: int64

non_null_count: 1000

null_count: 0

null_percentage: 0.0

statistics:

mean: 36.16

median: 37.0

std_dev: 14.8

min: 10

max: 60

q1: 23.0

q3: 49.0

unique_values: 51

cost_per_click:

data_type: float64

non_null_count: 1000

null_count: 0

null_percentage: 0.0

statistics:

mean: 0.81

median: 0.84

std_dev: 0.4

min: 0.1

max: 1.5

q1: 0.47

q3: 1.15

unique_values: 141

click_through_rate.1:

data_type: float64

non_null_count: 1000

null_count: 0

null_percentage: 0.0

statistics:

mean: 0.08

median: 0.08

std_dev: 0.04

min: 0.01

max: 0.15

q1: 0.04

q3: 0.12

unique_values: 1000

conversion_rate.1:

data_type: float64

non_null_count: 1000

null_count: 0

null_percentage: 0.0

statistics:

mean: 0.08

median: 0.08

std_dev: 0.04

min: 0.01

max: 0.15

q1: 0.04

q3: 0.11

unique_values: 717

ROI:

data_type: float64

non_null_count: 1000

null_count: 0

null_percentage: 0.0

statistics:

mean: 1.19
median: 0.67
std_dev: 1.54
min: 0.02
max: 11.57
q1: 0.3
q3: 1.42

unique_values: 316

numeric_summary:

total_numeric_columns: 8
correlation_insights: High correlation detected

categorical_summary:

total_categorical_columns: 10
high_cardinality_columns: ['timestamp', 'ad_id']

AI Insights

Executive Summary: Analysis of Digital Ad Performance Dataset

This analysis provides a comprehensive overview of digital advertising campaign performance, drawing insights from a dataset of 1000 ad interactions. The data reveals a robust conversion funnel post-click but a highly skewed Return on Investment (ROI), indicating significant opportunities for optimization by replicating successful strategies and addressing underperforming assets. User engagement is diverse, yet a substantial portion of ads are 'Ignored,' highlighting a need for improved initial capture. The current broad targeting across demographics, devices, and content types, while inclusive, may be suboptimal without segment-specific optimization.

1. Key Statistical Findings

- * **Dataset Completeness and Quality:** The dataset is remarkably complete with zero null values across all 18 columns, ensuring high data reliability for analysis. However, a significant data quality issue is present: `click_through_rate` and `click_through_rate.1`, along with `conversion_rate` and `conversion_rate.1`, are exact duplicates. This redundancy needs to be addressed.
- * **Ad Performance Metrics Consistency:**
 - * **Click-Through Rate (CTR):** Mean CTR is 0.08 (8%), with a median of 0.08, indicating a relatively consistent performance across campaigns. The standard deviation of 0.04 suggests moderate variability.
 - * **Conversion Rate (CR):** Mean CR is 0.08 (8%), with a median of 0.08, mirroring the CTR statistics very closely. This strong alignment between CTR and CR implies that once an ad is clicked, the subsequent conversion process is generally effective and consistent.
 - * **Cost Per Click (CPC):** Mean CPC is \$0.81, with a median of \$0.84, showing reasonable consistency.
- * **Return on Investment (ROI) Skewness:** The ROI distribution is highly skewed and characterized by

significant variability. The mean ROI is 1.19, but the median is notably lower at 0.67. This large difference, coupled with a high standard deviation (1.54) and a maximum value of 11.57, strongly suggests that a few high-performing campaigns are significantly inflating the average, while the majority perform closer to or below the median.

- * **User Engagement Diversity:** Engagement levels are distributed across "Ignored" (27.2%), "Liked" (25.7%), "Shared" (24.0%), and "Commented" (23.1%). While "Ignored" is the most frequent outcome, the substantial positive engagement (Liked, Shared, Commented) signifies a receptive audience for many ads.

- * **User Interaction Frequency:** With 610 unique `user_id`s out of 1000 total rows, approximately 39% of interactions come from repeat users (on average, ~1.6 interactions per user). This indicates a segment of users engaging with multiple ad events.

- * **Even Distribution of Key Categorical Variables:** `device_type` (Desktop, Mobile, Tablet), `location` (5 countries), `age_group` (5 segments), `gender`, `content_type` (Video, Image, Text), `ad_topic`, and `ad_target_audience` all show relatively even distributions. This suggests a broad, diversified advertising strategy across multiple dimensions.

- * **View Time:** The average `view_time` is 36.16 seconds, with a median of 37 seconds, indicating typical user attention spans for the content served. The range from 10 to 60 seconds provides room for analysis of engagement duration.

2. Notable Patterns & Trends

- * **Strong Click-to-Conversion Funnel:** The near-identical statistical profiles for CTR and Conversion Rate indicate that ads are effective at converting users *after* they click. This suggests that the landing pages, offers, or user experience post-click are largely consistent and well-optimized. The primary bottleneck might therefore lie in generating the initial click.

- * **High-Performing Ad Outliers Driving Average ROI:** The significant disparity between mean and median ROI, coupled with the exceptionally high maximum ROI, points to a clear Pareto principle. A small subset of ad campaigns are achieving disproportionately high returns. Identifying the characteristics of these campaigns is paramount.

- * **Cross-Platform and Demographic Reach:** The even distribution across device types, locations, age groups, and genders demonstrates a successful broad reach strategy. This implies that the advertising efforts are not overly concentrated on a single segment, providing a wide net for potential customers.

- * **Balanced Content and Topic Strategy:** The equal prevalence of `Video`, `Image`, and `Text` content types, along with diverse `ad_topic` and `ad_target_audience` distributions, indicates a varied content and targeting approach. This broadness could be an asset for general brand awareness but may lack the specificity for hyper-optimized performance within niche segments.

- * **Opportunity in Repeat User Engagement:** The presence of users interacting multiple times with ads (610 unique users in 1000 interactions) highlights an opportunity to study repeat engagement behavior and potentially develop retargeting or loyalty programs.

3. Anomalies & Outliers

- * Significant ROI Outliers: The most prominent anomaly is the maximum ROI of 11.57, which is an extreme outlier compared to the Q3 value of 1.42 and a median of 0.67. These exceptionally high-performing instances represent significant positive outliers that warrant immediate investigation. Conversely, the minimum ROI of 0.02 suggests some campaigns are performing extremely poorly.

- * Redundant Columns: The presence of `click_through_rate.1` and `conversion_rate.1` as exact duplicates of `click_through_rate` and `conversion_rate` is a data quality anomaly. While not affecting the statistical values, it represents redundant information and potential for confusion or inefficient data processing.

4. Business Implications

- * Untapped ROI Potential: The highly skewed ROI distribution implies that overall campaign profitability could be significantly boosted by systematically identifying and scaling the characteristics of the top-performing ads. Conversely, underperforming ads are likely consuming budget with minimal returns, dragging down the aggregate ROI.

- * Focus Shift: From Conversion Optimization to Click Generation: Given the strong alignment between CTR and Conversion Rate, the primary lever for increasing conversions is likely to be improving the initial ad click-through. This suggests that ad creative, targeting precision, and placement strategies are critical areas for intervention.

- * Strategic Segmentation Imperative: While broad targeting ensures wide reach, the even distribution across various demographic and platform segments suggests a "one-size-fits-all" approach may be in play. This could mean missed opportunities for hyper-segmentation where specific combinations of demographics, devices, content types, or ad topics could yield significantly higher ROI.

- * Data Integrity and Efficiency: The duplicate columns (.1 suffix) indicate a potential data pipeline or recording issue. If not addressed, this could lead to inefficient data management, larger data storage, and potential analytical errors or redundant efforts.

- * Customer Lifetime Value (CLV) Opportunity: The existence of repeat user interactions provides a foundation for understanding customer journeys, predicting future engagement, and developing strategies to increase CLV through personalized content and retargeting for repeat users.

- * Engagement Strategy Refinement: The ~27% 'Ignored' rate signals that a quarter of ad impressions are not capturing user attention. Improving this initial engagement could unlock significant gains in CTR and subsequent conversions.

5. Recommended Actions

1. Conduct a Deep Dive into High-ROI Performers:

- * Action: Immediately isolate and analyze all ad interactions with an ROI significantly above the third quartile (e.g., > 1.42), especially those approaching or exceeding the maximum of 11.57. Correlate these high performers with `ad_id`, `ad_topic`, `ad_target_audience`, `content_type`, `device_type`, `location`, `age_group`, `gender`, and `view_time`.

- * Goal: Identify the specific attributes, combinations, and contexts that consistently drive exceptional ROI.

- * Output: A "Success Playbook" outlining best practices for ad creative, targeting, platform placement,

and content strategy to scale high-performing campaigns.

2. Optimize Initial Ad Engagement to Boost CTR:

- * Action: Perform a detailed analysis of ads categorized as `engagement_level: "Ignored"` (~27%). Investigate correlations between 'Ignored' status and `ad_topic`, `content_type`, `ad_target_audience`, `device_type`, `cost_per_click`, and `view_time`.

- * Goal: Develop actionable insights to reduce the "Ignored" rate and significantly improve `click_through_rate` across campaigns.

- * Output: Recommendations for A/B testing different ad creatives, headlines, call-to-actions, and initial targeting parameters to better capture user attention and drive clicks.

3. Implement Granular Segment-Specific Optimization:

- * Action: Move beyond broad targeting by analyzing `click_through_rate`, `conversion_rate`, and `ROI` across specific intersections of `device_type`, `location`, `age_group`, `gender`, `content_type`, `ad_topic`, and `ad_target_audience`. For instance, analyze "Video content for Young Adults on Mobile in India" versus "Image content for Fitness Lovers on Desktop in USA."

- * Goal: Identify the most profitable and responsive segments for each ad characteristic to enable precise budget allocation and customized campaign development.

- * Output: A strategic segmentation matrix with performance benchmarks and tailored recommendations for content, targeting, and budget distribution for each identified high-potential segment.

4. Resolve Data Quality Issues:

- * Action: Engage with data engineering or analytics teams to immediately remove the redundant `click_through_rate.1` and `conversion_rate.1` columns from the dataset. Ensure future data ingestion processes prevent such duplications.

- * Goal: Maintain data integrity, reduce storage overhead, streamline analytical processes, and prevent potential reporting errors.

- * Output: A clean, de-duplicated dataset for ongoing analysis.

5. Develop a Repeat User Strategy:

- * Action: Segment users into "first-time" and "repeat" categories based on `user_id` uniqueness. Analyze the differential ad engagement patterns, view times, and conversion behaviors of repeat users compared to new users.

- * Goal: Understand the characteristics and journey of repeat engagers to foster loyalty and maximize their lifetime value through targeted re-engagement strategies.

- * Output: Recommendations for personalized ad sequencing, loyalty programs, or exclusive content tailored for recurring users.