Automated Ingestion & Reporting Pipeline

Ingest • Transform • Auto-generate Executive PDFs

Generated on: 03 December 2025

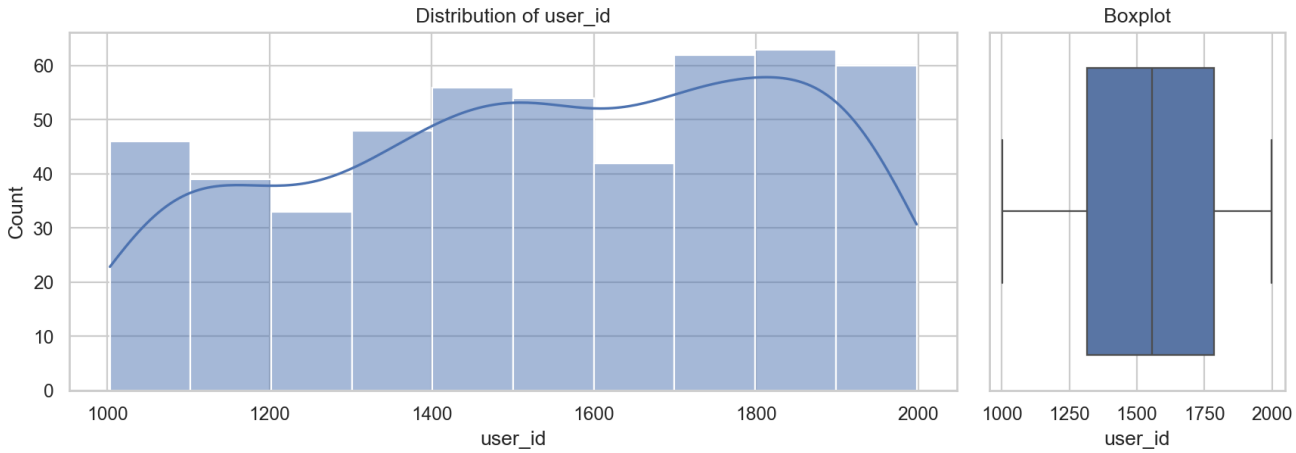
normalized 0c41e5badfd0d4b2f7753149dc878f28 ff0bfe6d

Summary

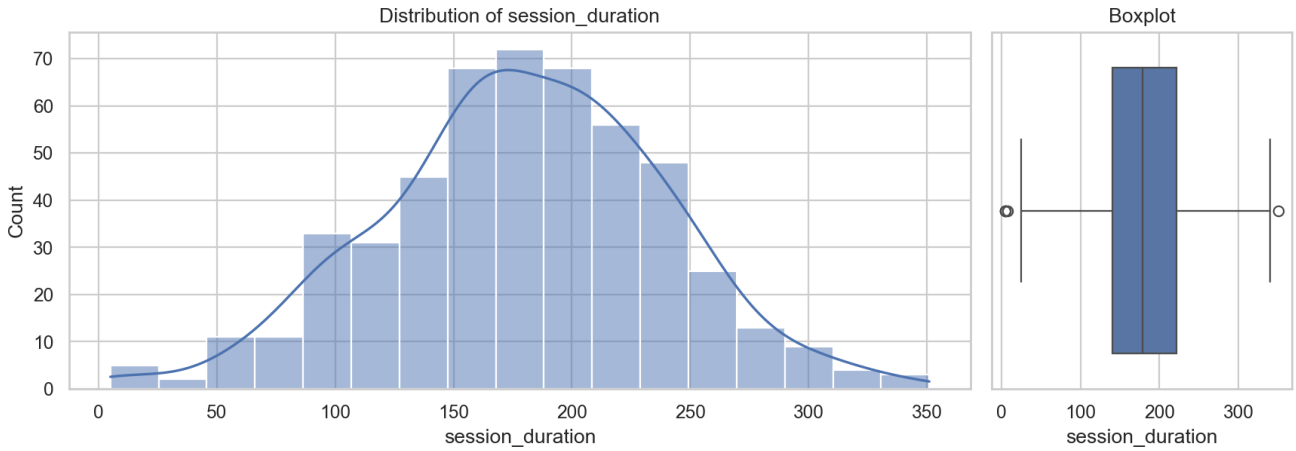
Overall Description:
The dataset captures 505 web□session records with timestamps, user identifiers, device types, page□view counts, session durations, and optional campaign tags. The visualizations collectively illustrate the data’s size, the distributions of key numeric fields, weak inter□variable relationships, temporal stability of user identifiers, and the composition of device usage, while highlighting the high missingness of the campaign identifier.

Rows	Cols	Numeric	Categorical
505	**8**	**3**	**5**

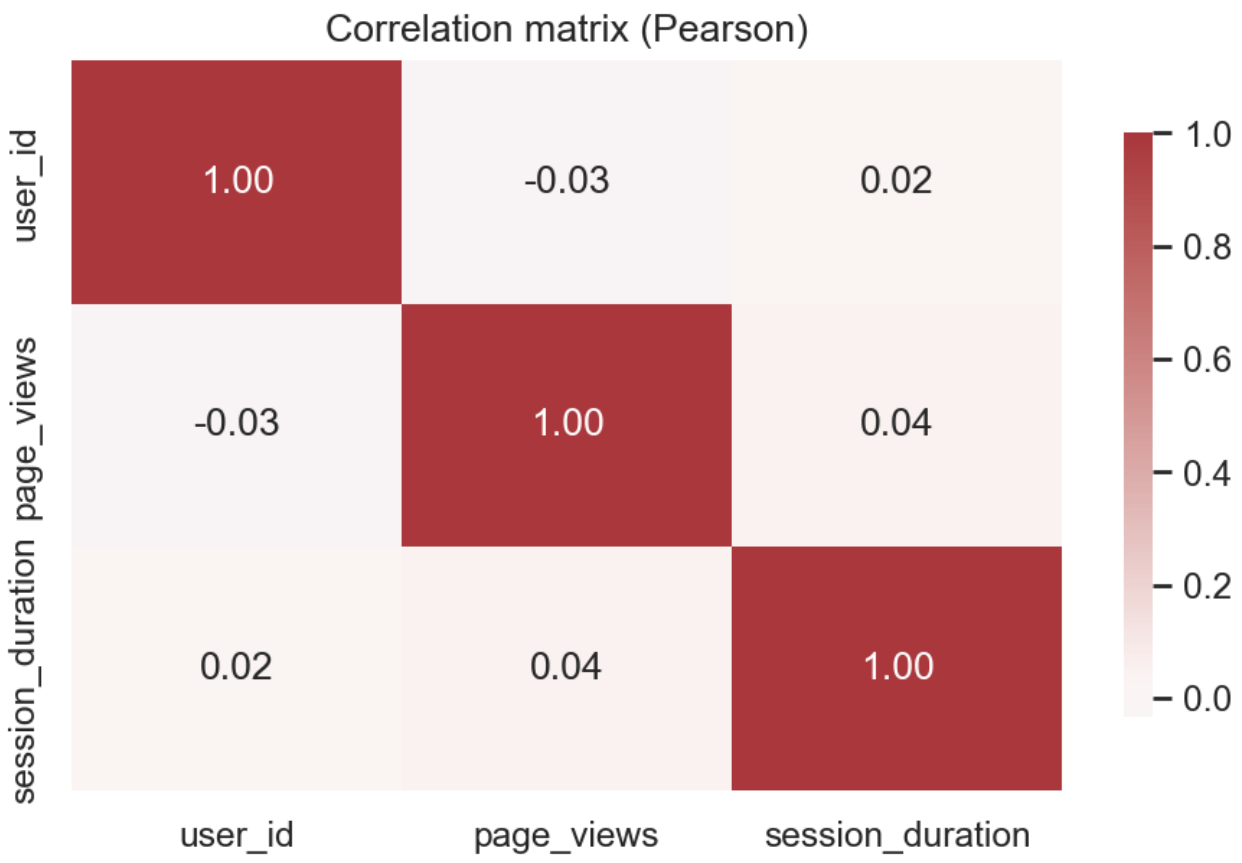
Shows the dataset’s basic metrics: 505 rows, 8 columns, and the count of numeric variables, confirming a modest□sized, primarily quantitative table.



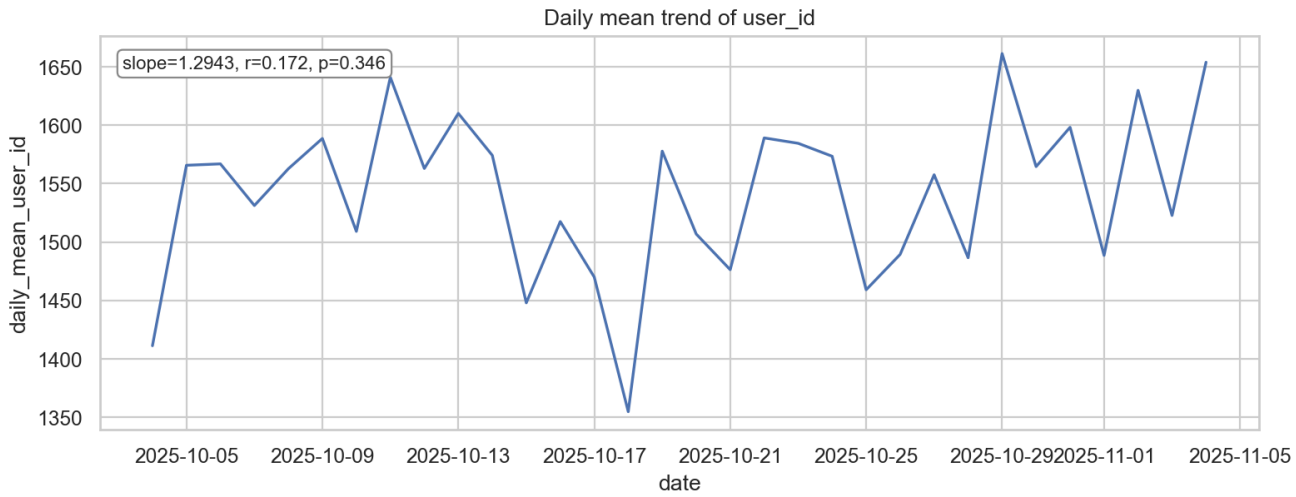
The `user_id` values (n = 503) are roughly symmetric (skew = □0.21) with mean ≈ 1541 and standard deviation ≈ 288, indicating a moderately dispersed identifier set.



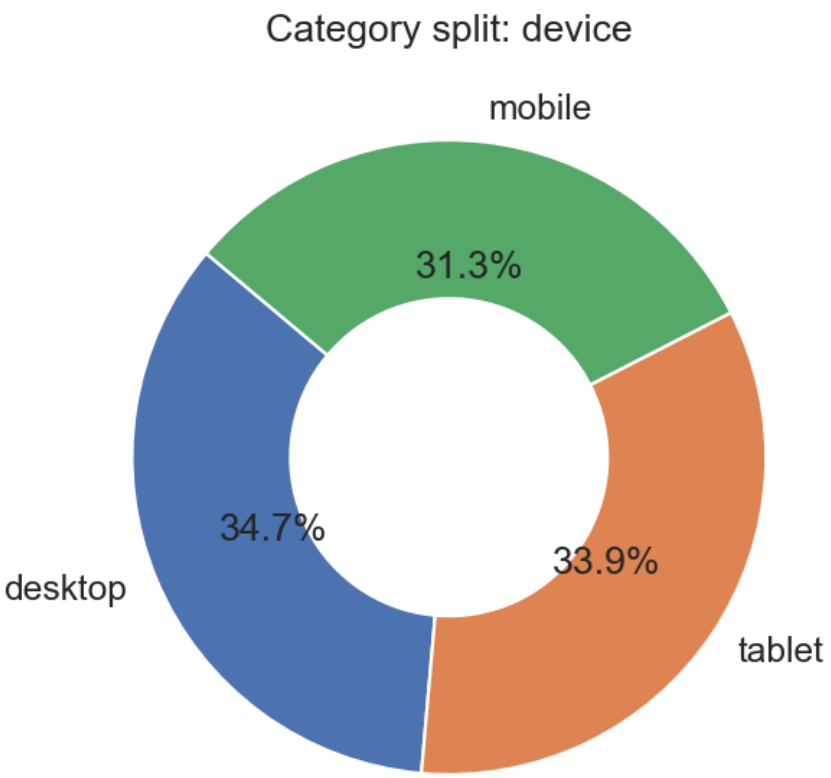
`session_duration` (n = 504) exhibits a near□normal shape (skew = □0.13) centered at a mean of ≈ 179 seconds with a standard deviation of ≈ 60 seconds.



Numeric correlations are negligible; the strongest pair is `page_views` vs. `session_duration` ($r \approx 0.044$, $p = 0.323$), suggesting no meaningful linear association among the measured metrics.



The daily mean of `user_id` shows a shallow positive trend (slope ≈ 1.29) with low explanatory power (r = 0.172, p = 0.346), indicating no significant temporal drift in user identifiers.



Depicts the distribution of device categories, revealing the dominant platforms used during sessions; exact proportions are not specified but the chart emphasizes the relative popularity of each device type.

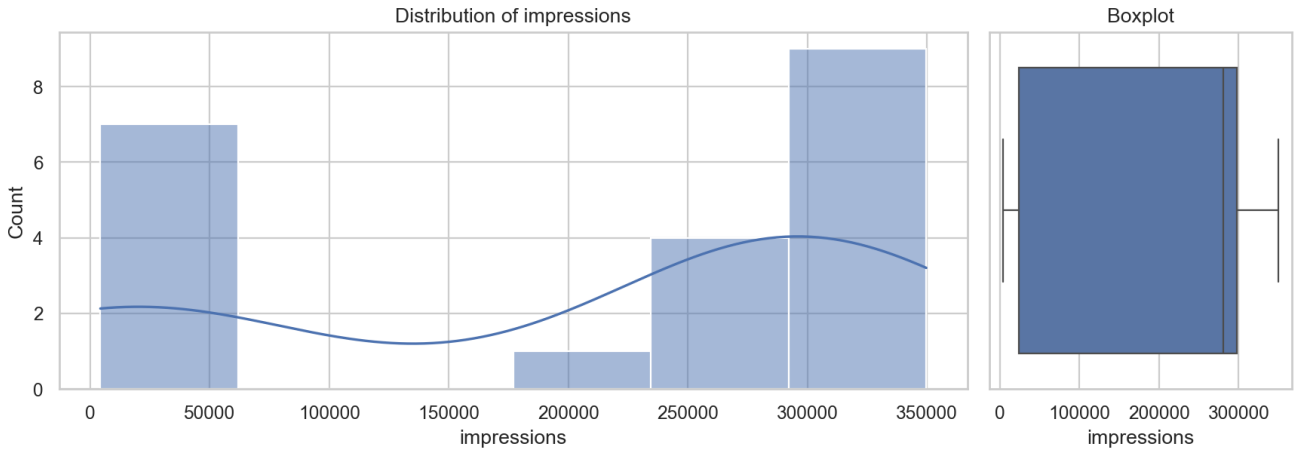
normalized 6337d65c03840b4f4a7f1152428ed033 d3c50389

Summary

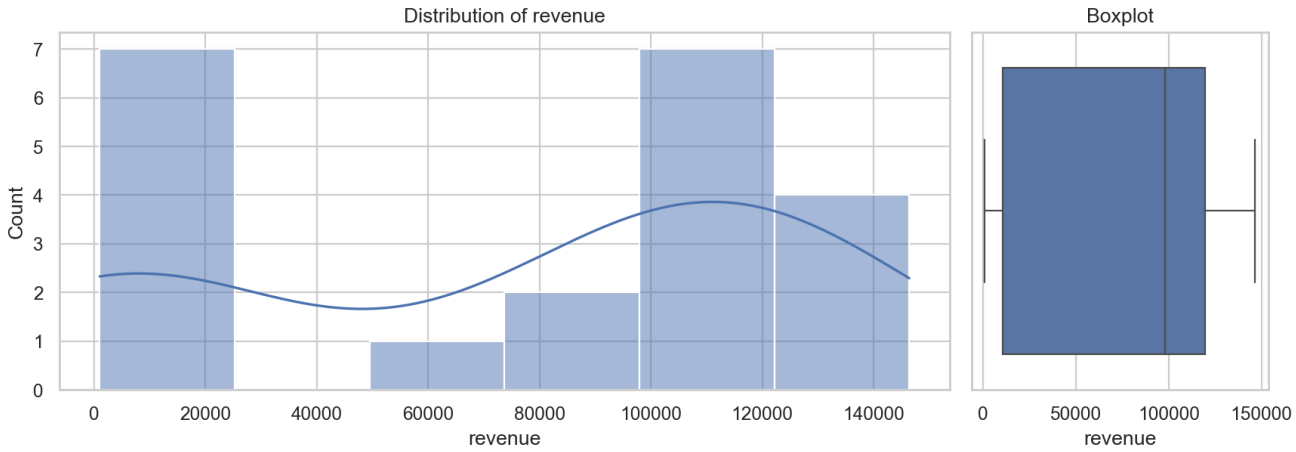
Overall Description:
The dataset captures monthly advertising performance across 21 campaigns, detailing impressions, clicks, spend, and revenue. Visualizations collectively illustrate the distribution of key metrics, the temporal composition of campaigns, and exceptionally strong positive relationships—particularly between spend and revenue—indicating consistent return on advertising investment.

Rows	Cols	Numeric	Categorical
21	**6**	**4**	**2**

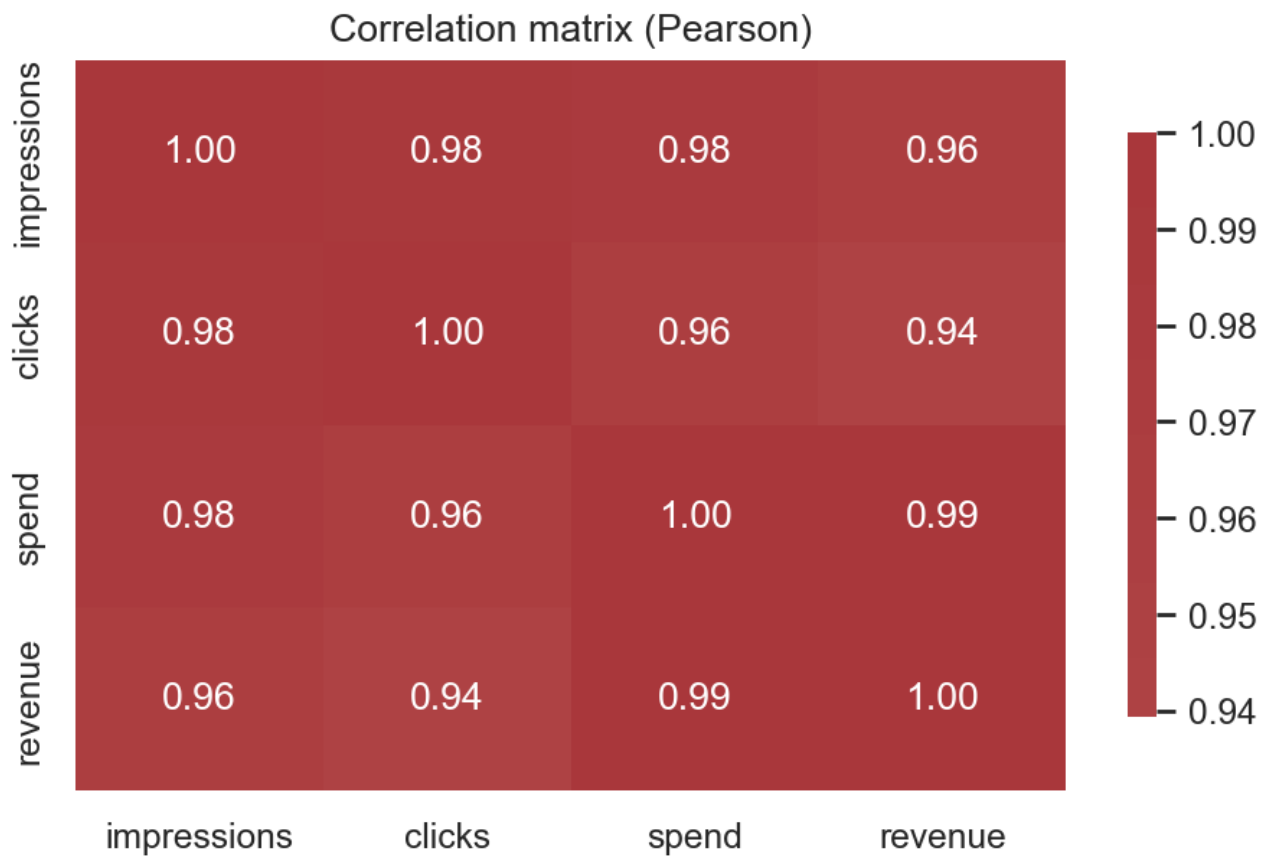
The KPI card confirms a compact dataset with 21 rows and 6 fully populated columns, all numeric fields free of missing values.



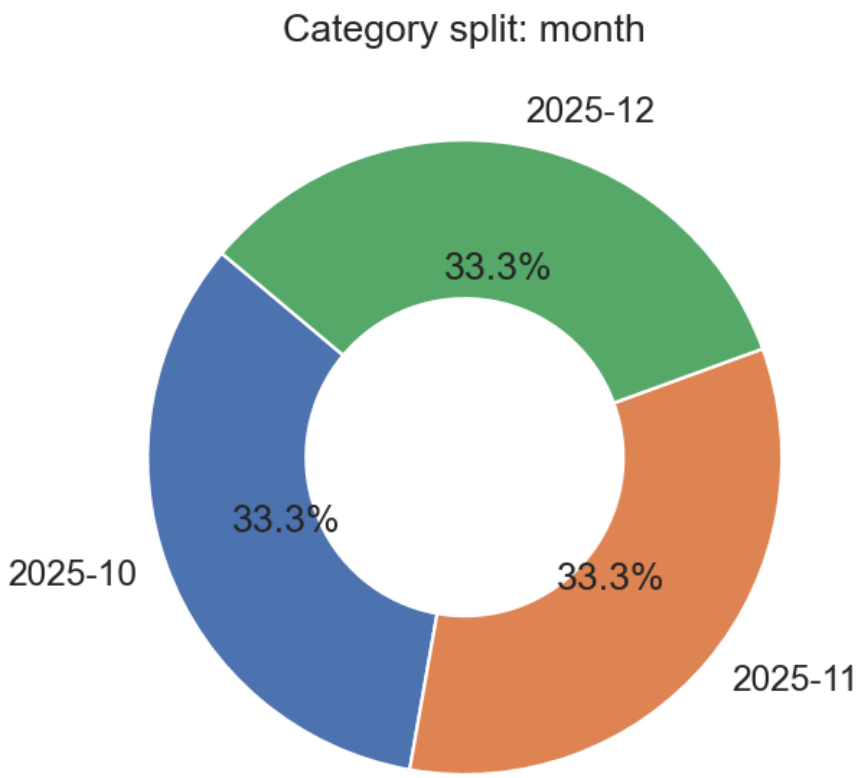
Impressions exhibit a mean of 202,409 with moderate dispersion (σ ≈ 135,758) and a slight left□skew (□0.65), suggesting most campaigns generate high exposure with a few lower□impression outliers.



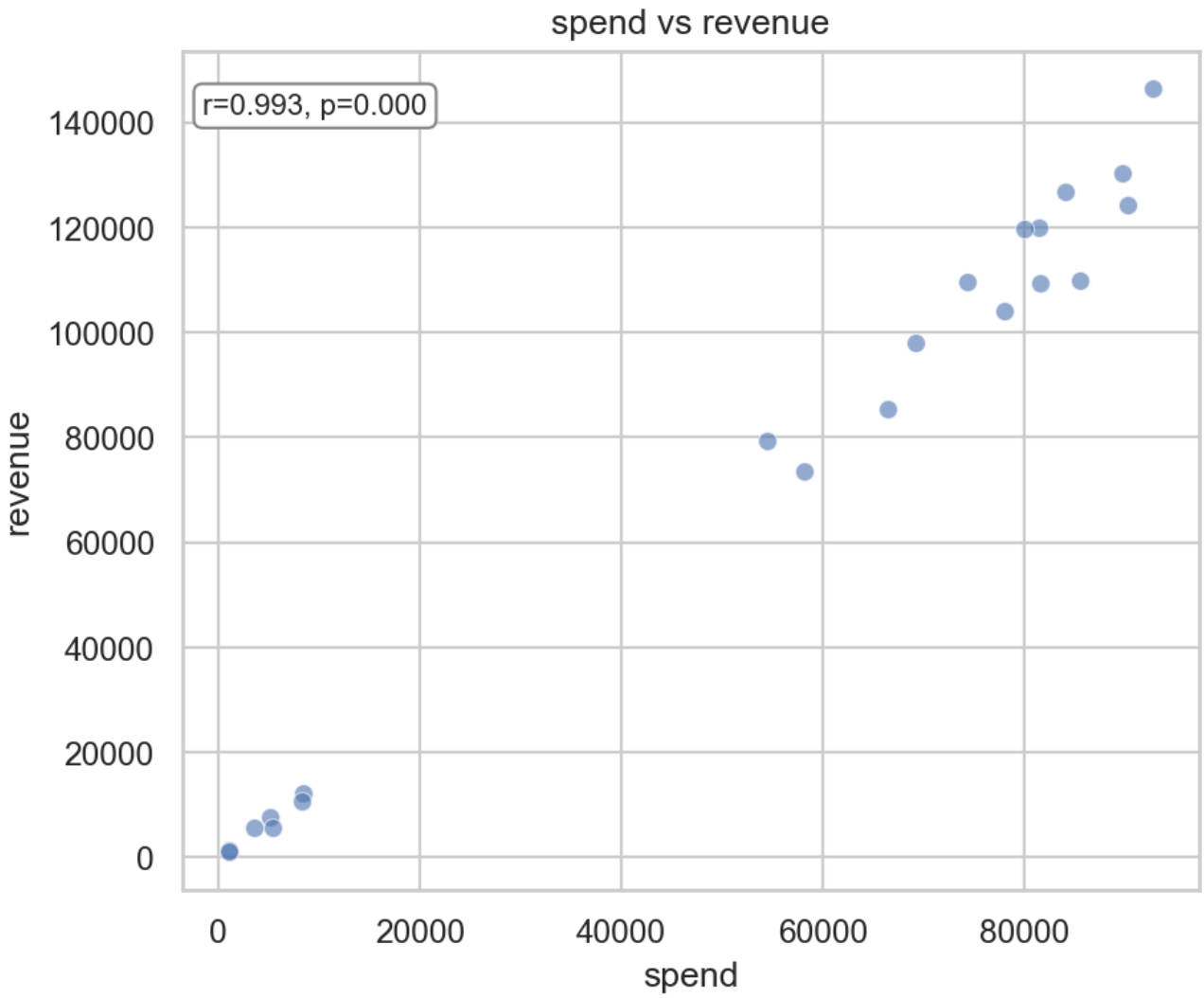
Revenue shows a mean of 75,267 and a standard deviation of 52,706, also left□skewed (□0.45), indicating that while average earnings are substantial, a minority of campaigns contribute disproportionately lower revenue.



The correlation heatmap highlights an almost perfect positive link between spend and revenue ($r = 0.993$, $p < 0.001$) and strong associations among impressions, clicks, spend, and revenue ($r \geq 0.96$, $p < 0.001$), confirming that higher exposure and investment translate directly into higher earnings.



The month pie chart reveals the distribution of campaigns across months, emphasizing which periods dominate the dataset’s activity.



The scatter plot of spend versus revenue displays a tight linear trend (Pearson r = 0.993, p < 0.001), reinforcing the near□one□to□one proportionality between advertising expenditure and generated revenue.

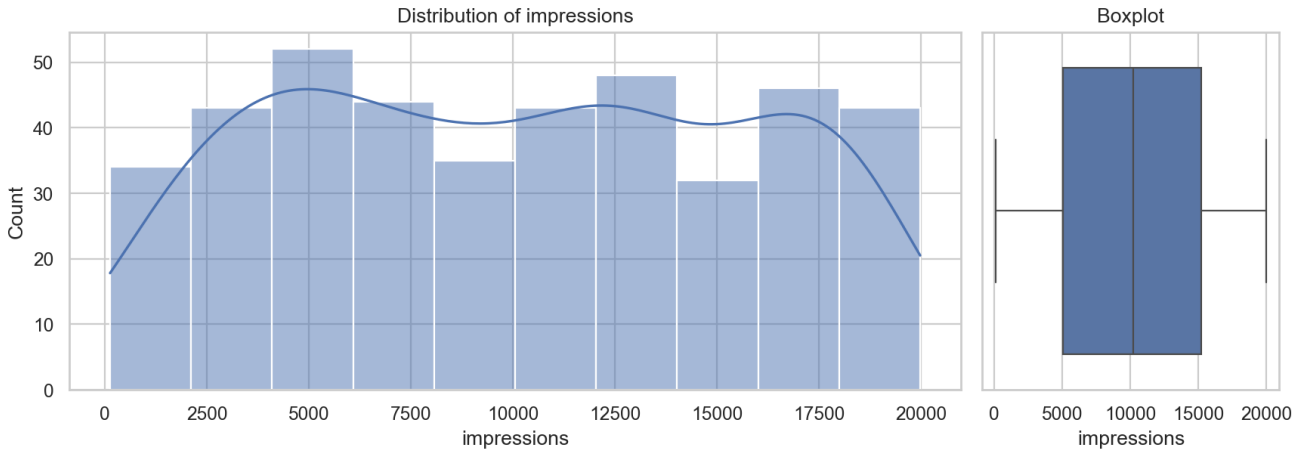
normalized 99681b1f583da3ffeb0f6c5cf73f8ead f785e8ad

Summary

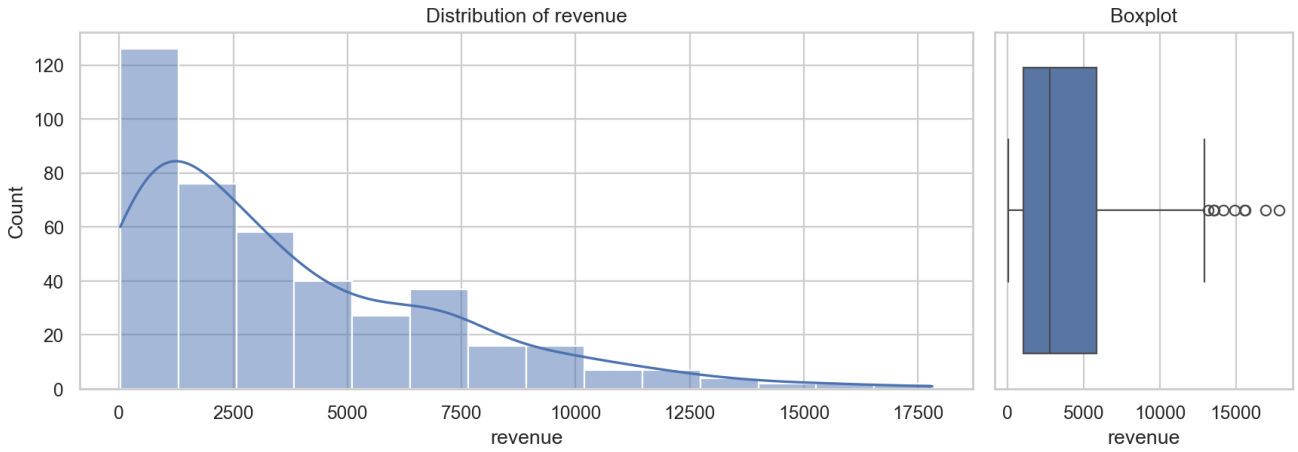
Overall Description:
The dataset captures daily performance metrics for 420 advertising records across six fields, including spend, revenue, impressions, clicks, and campaign identifiers. Visualizations reveal generally normal impression volumes, a right□skewed revenue distribution, a very strong positive relationship between spend and revenue, minimal temporal trend in impressions, and a balanced representation of the top five campaigns.

Rows	Cols	Numeric	Categorical
420	**6**	**4**	**2**

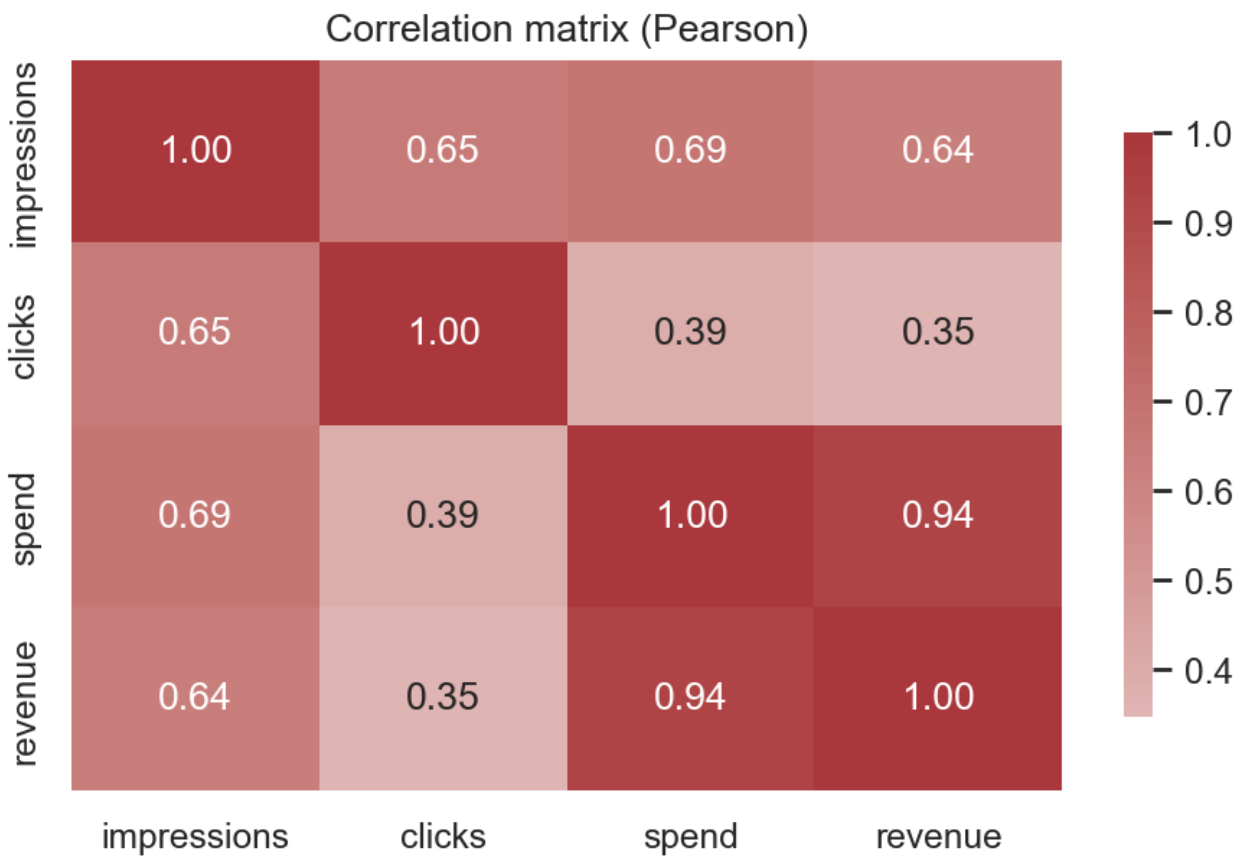
Shows the dataset dimensions (420 rows × 6 columns) and confirms that all columns are numeric or datetime with zero missing values.



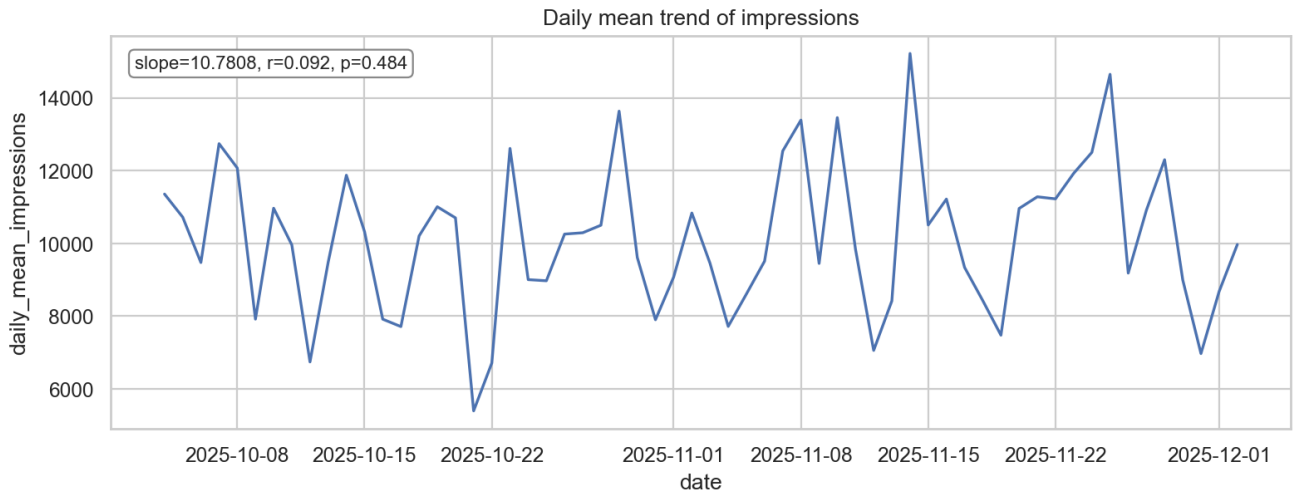
Impressions are approximately normally distributed (mean ≈ 10,120; σ ≈ 5,647) with negligible skew (0.02), indicating consistent exposure levels across records.



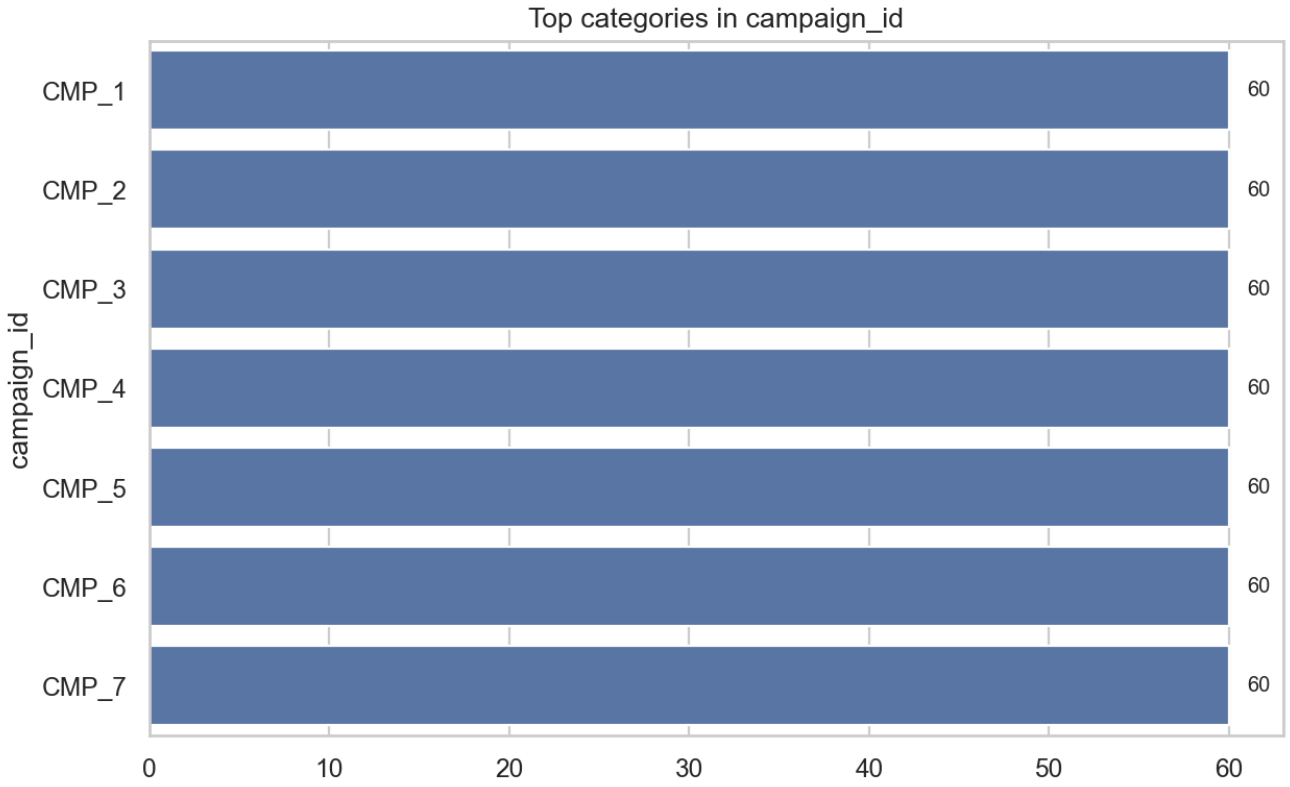
Revenue exhibits a right□skewed pattern (mean ≈ 3,763; σ ≈ 3,476; skew = 1.26), suggesting a minority of records generate disproportionately high earnings.



The strongest correlation is between spend and revenue (r = 0.936, p < 0.001), confirming that higher advertising spend translates into higher revenue. Moderate positive links also exist: impressions□spend (r = 0.685), impressions□clicks (r = 0.653), and impressions□revenue (r = 0.637).



Daily mean impressions show no significant upward or downward trend (slope ≈ 10.78, r = 0.092, p = 0.484), indicating stable exposure over time.



The top five campaign IDs (CMP_1 to CMP_5) each appear 60 times, reflecting an evenly distributed allocation of records among the leading campaigns.

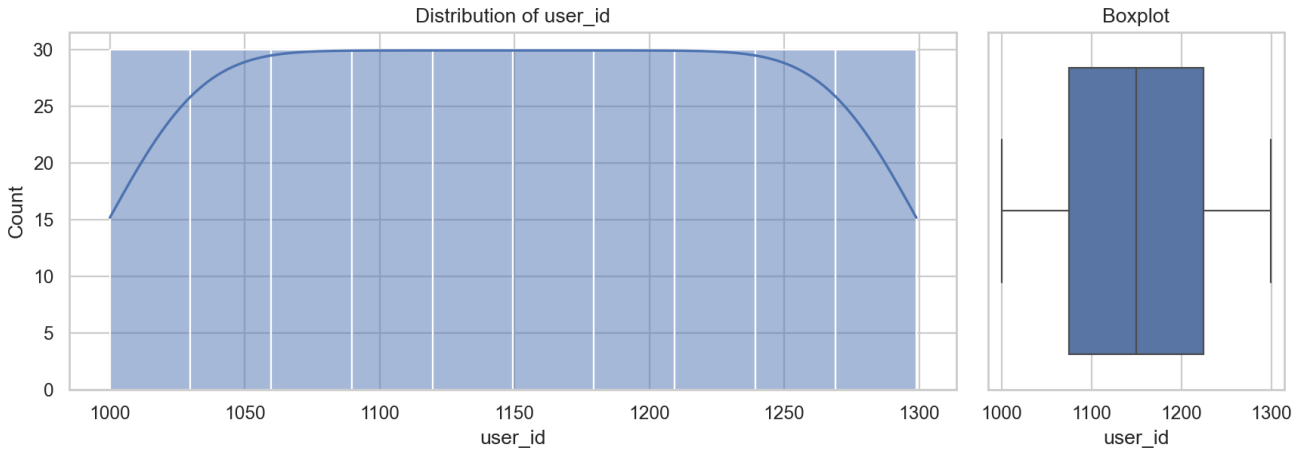
normalized b979cd0f7abb29b77fcbe0607e52a2b0 b8617133

Summary

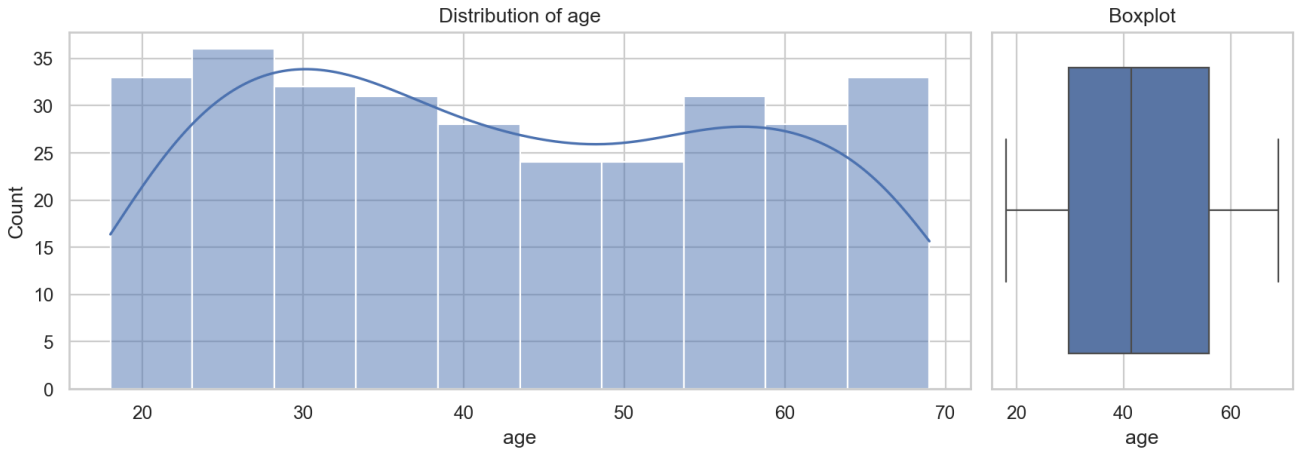
Overall Description:
The dataset captures 300 user records with identifiers, signup timestamps, demographic attributes (country, age, gender), and exhibits minimal missingness except for gender (33%). Visualizations reveal near□normal distributions for numeric fields, negligible linear relationships, and a gender composition dominated by a few categories, providing a concise baseline for further demographic or temporal analyses.

Rows	Cols	Numeric	Categorical
300	**5**	**2**	**3**

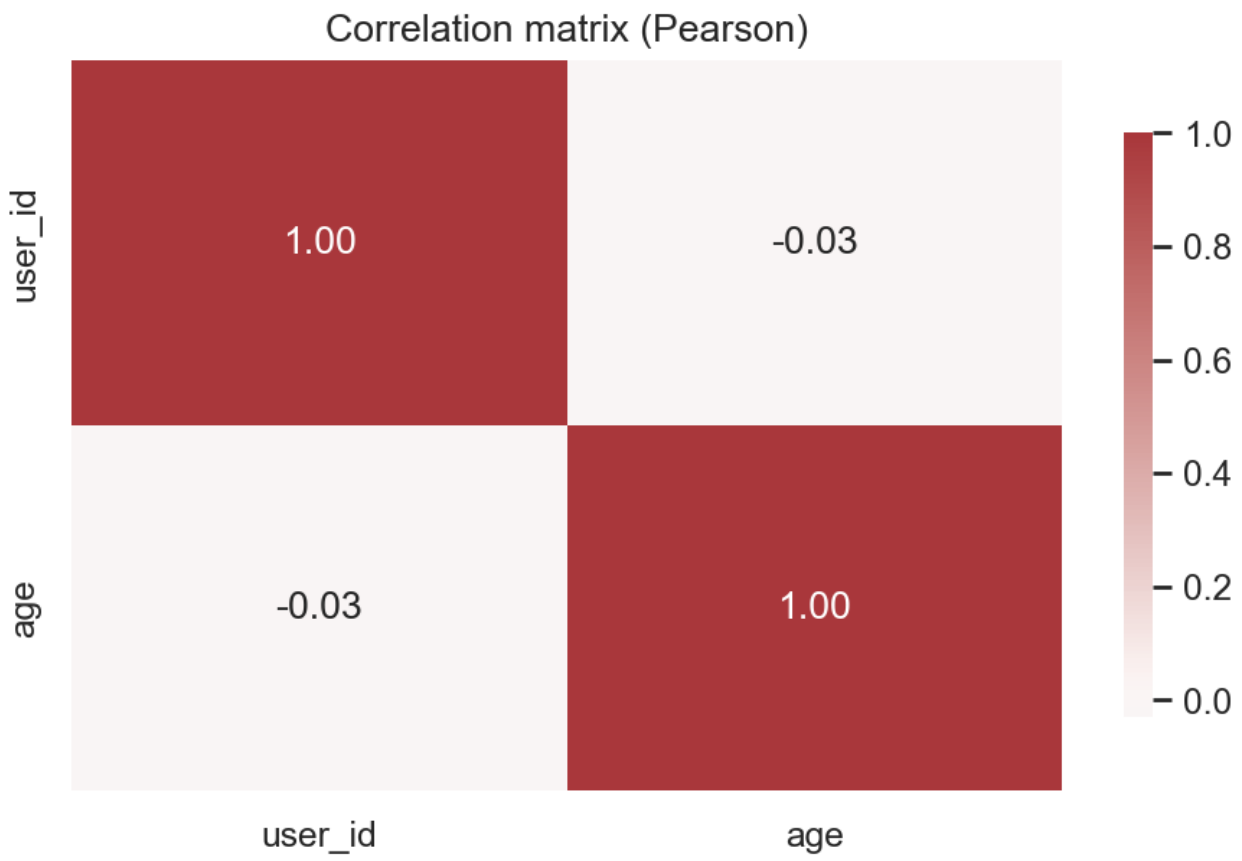
Summarizes the dataset size (300 rows, 5 columns) and highlights that only one column (gender) exceeds the 20 % missing□value threshold.



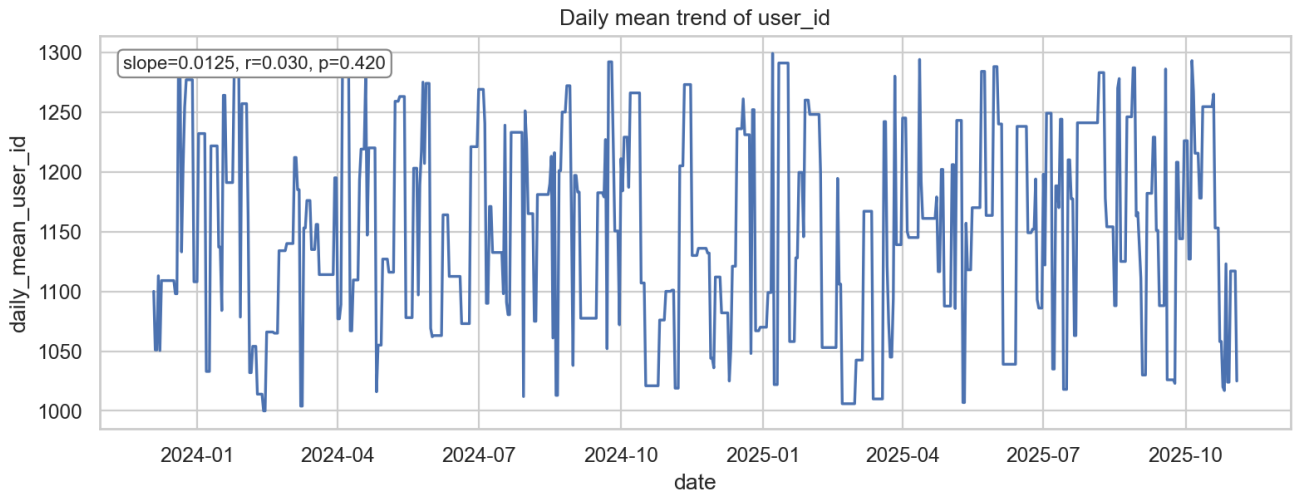
Shows a tightly clustered user_id distribution (mean ≈ 1149.5, σ ≈ 86.7) with zero skew, indicating a roughly symmetric spread around the central value.



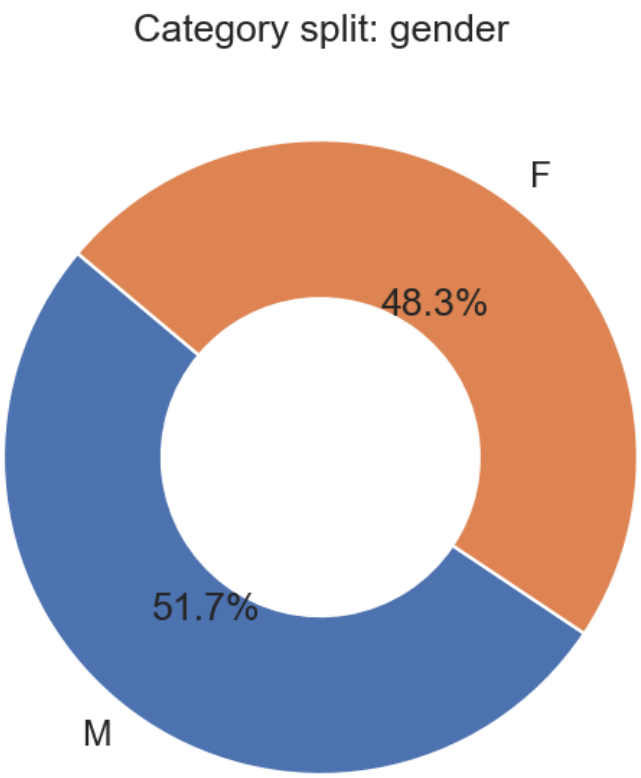
Depicts the age profile (mean ≈ 42.7 years, σ ≈ 15.2) with slight positive skew (0.12), suggesting a modest tail toward older ages.



Highlights the strongest numeric association: user_id vs. age (r = □0.031, p = 0.592), a negligible and non□significant correlation.



Plots daily mean user_id over signup_date, yielding a flat trend (slope = 0.0125, r = 0.030, p = 0.420), confirming no temporal drift in identifier values.



Illustrates the gender composition, with a few categories comprising the majority of the observed (67 %) non□missing entries.

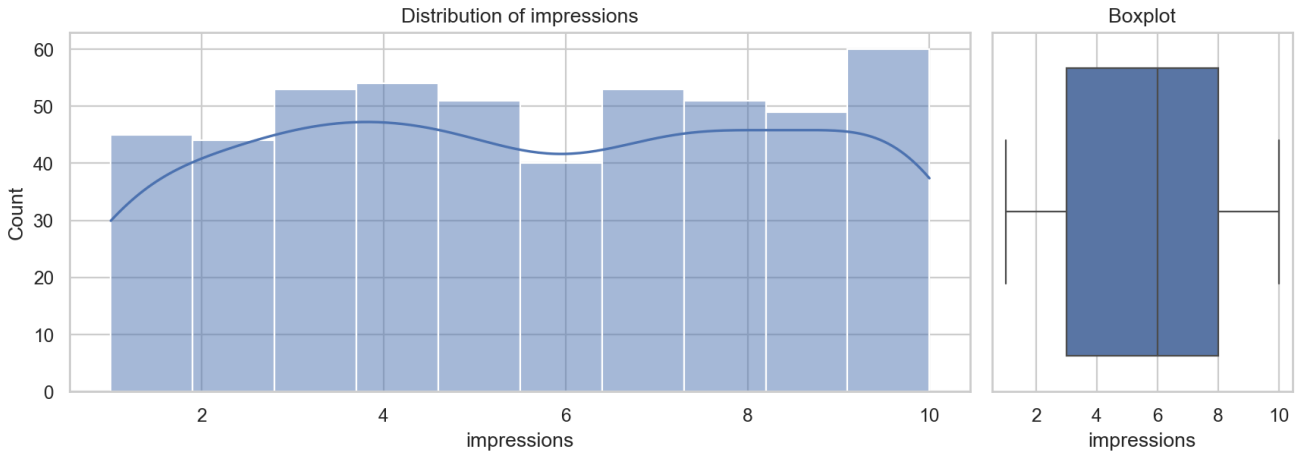
normalized bba12fa2661e2b358250b095d4c2031a 80eb786a

Summary

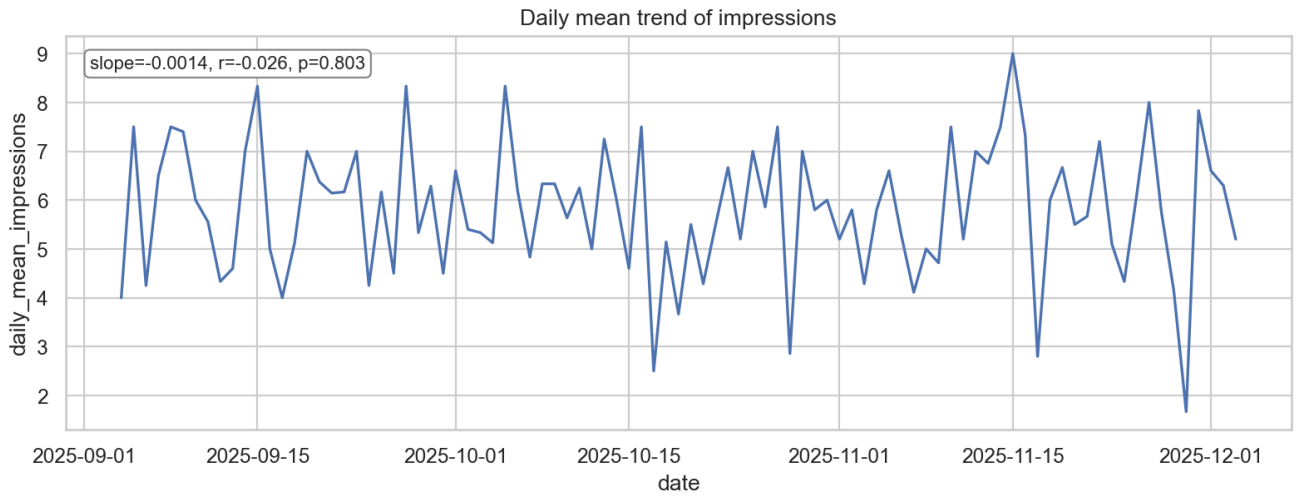
Overall Description:
The dataset captures 500 ad impression records across five fields, including timestamps, campaign and ad identifiers, and the count of impressions. Visualizations reveal a modest, roughly symmetric distribution of impression counts, negligible temporal trend, and a concentration of activity among a few dominant campaigns, while the KPI summary confirms a complete, clean dataset with no missing values.

Rows	Cols	Numeric	Categorical
500	**5**	**1**	**4**

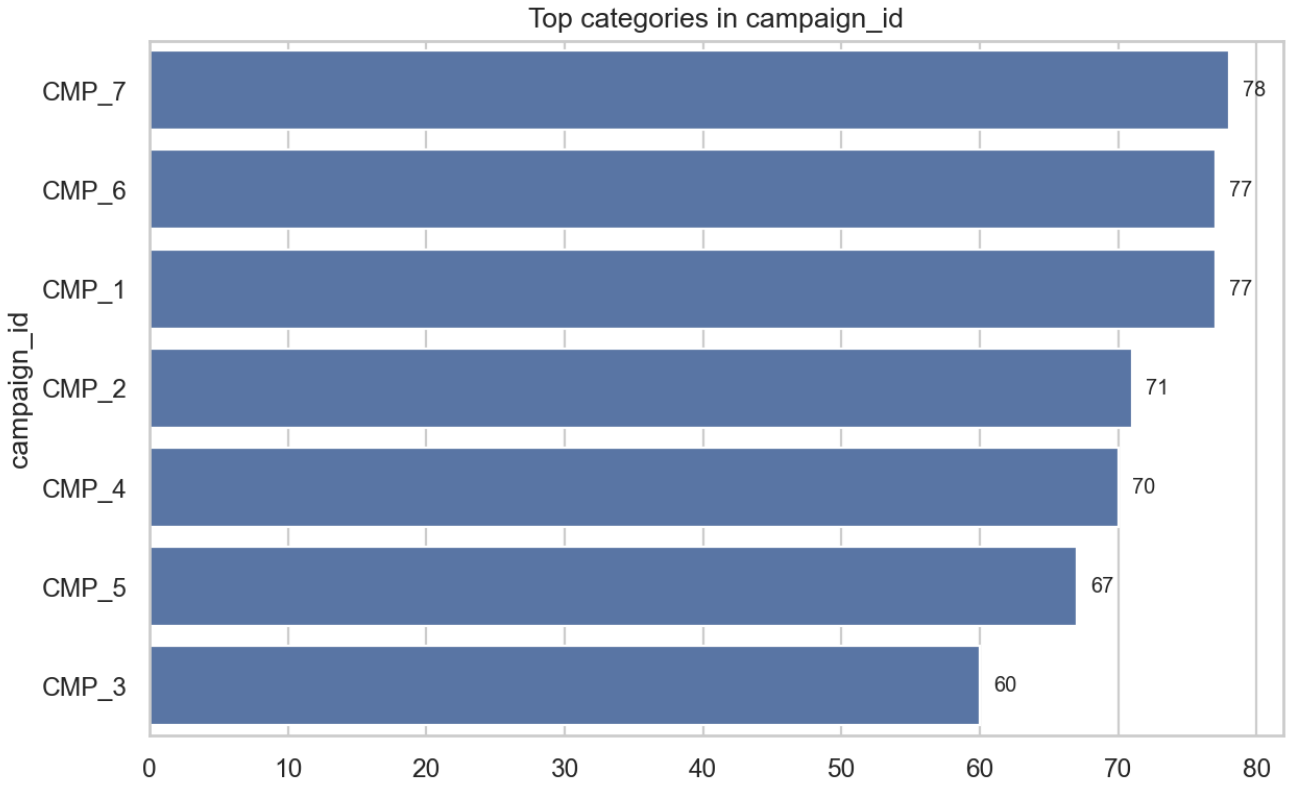
The KPI card confirms the dataset’s integrity: 500 rows, 5 columns, and zero missing values across all fields.



Impression counts have a mean of 5.65 with a standard deviation of 2.89; the distribution is nearly symmetric (skew ≈ □0.02) and shows no extreme outliers.



The daily mean of impressions exhibits a flat trend (slope = □0.0014, r = □0.026, p = 0.803), indicating no statistically significant change over time.



Campaign activity is dominated by a few IDs: CMP_7 (78 records), CMP_6 (77), CMP_1 (77), CMP_2 (71), and CMP_4 (70), highlighting a skewed distribution of impressions across campaigns.

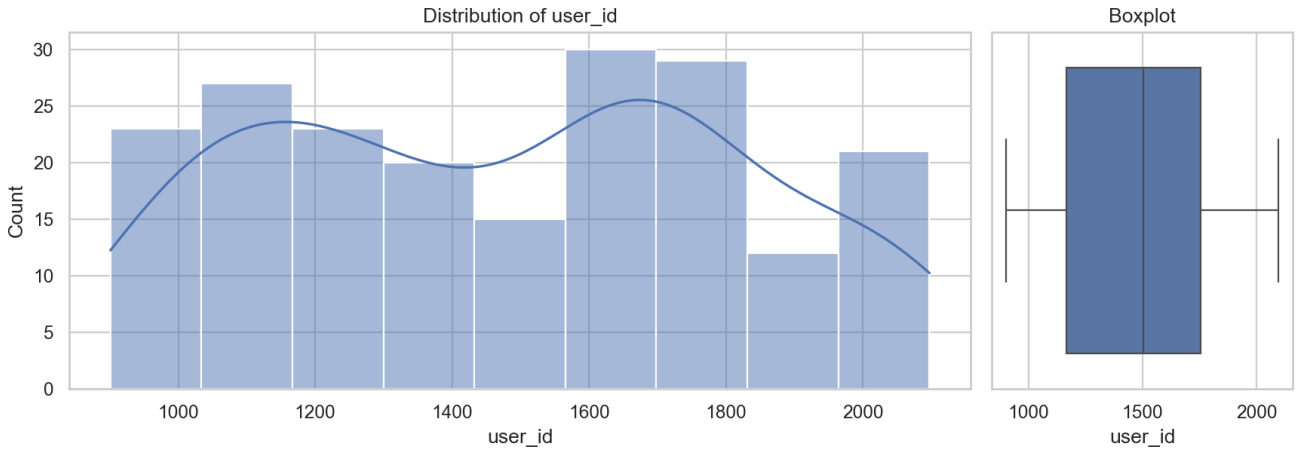
normalized e29a88cdc6ab94e02a88f2ba16ccc750 d86e9d81

Summary

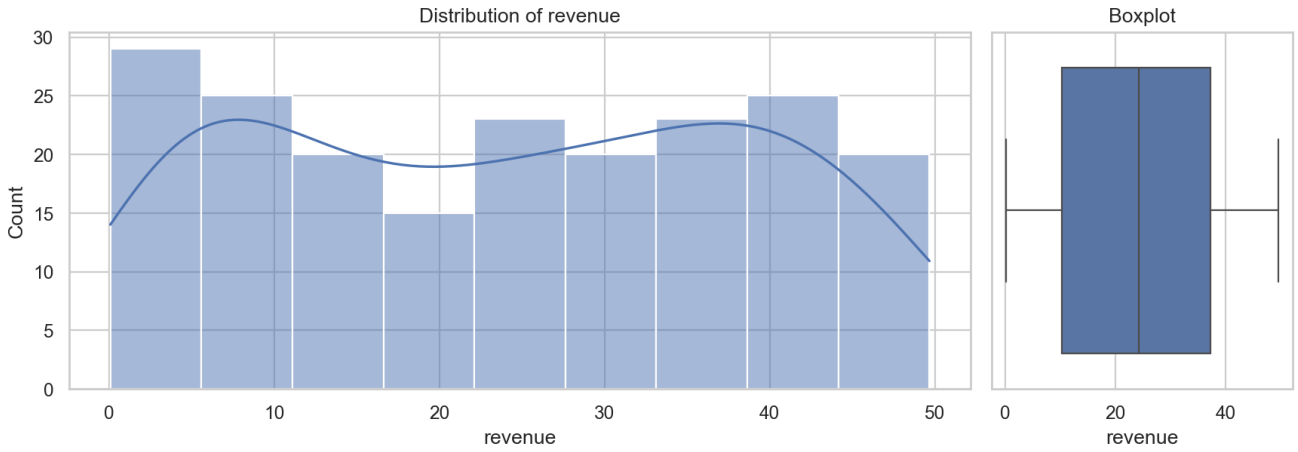
Overall Description:
The dataset captures 200 conversion events, each identified by a unique `conversion_id` and timestamped, with associated user, campaign, and revenue information. Visualizations reveal the distributions of `user_id` and `revenue`, a negligible linear relationship between them, stable daily user□id averages over time, and the predominance of a few campaign categories. No substantial missing data or strong numeric correlations are present, indicating a clean but modestly informative snapshot of conversion activity.

Rows	Cols	Numeric	Categorical
200	5	2	3

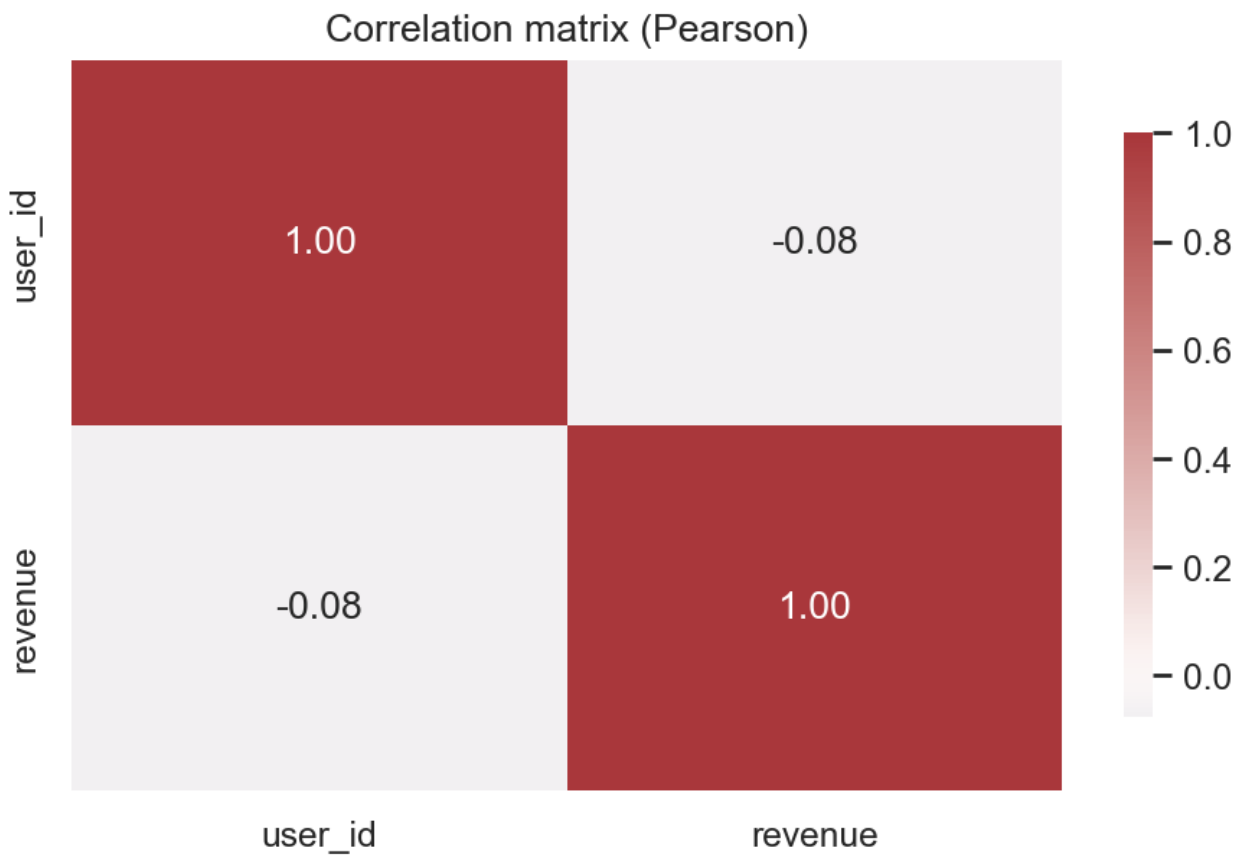
Shows the basic dimensions of the data (200 rows, 5 columns) and confirms that all columns are fully populated with numeric summaries available.



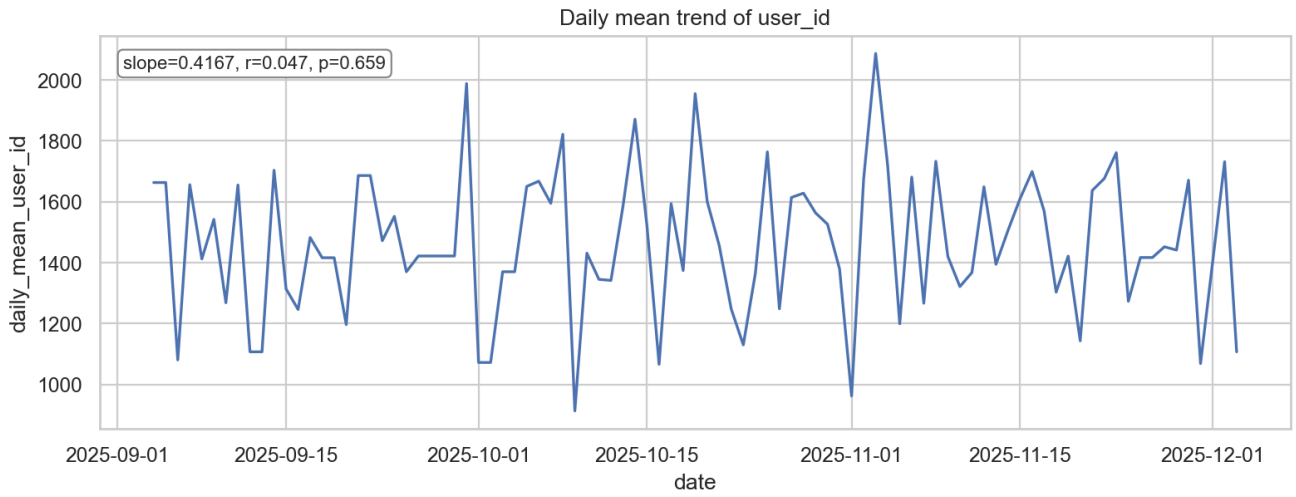
`user_id` values are tightly clustered around a mean of 1,479 (SD = 349) with near□zero skew (0.05), indicating a roughly symmetric distribution without extreme outliers.



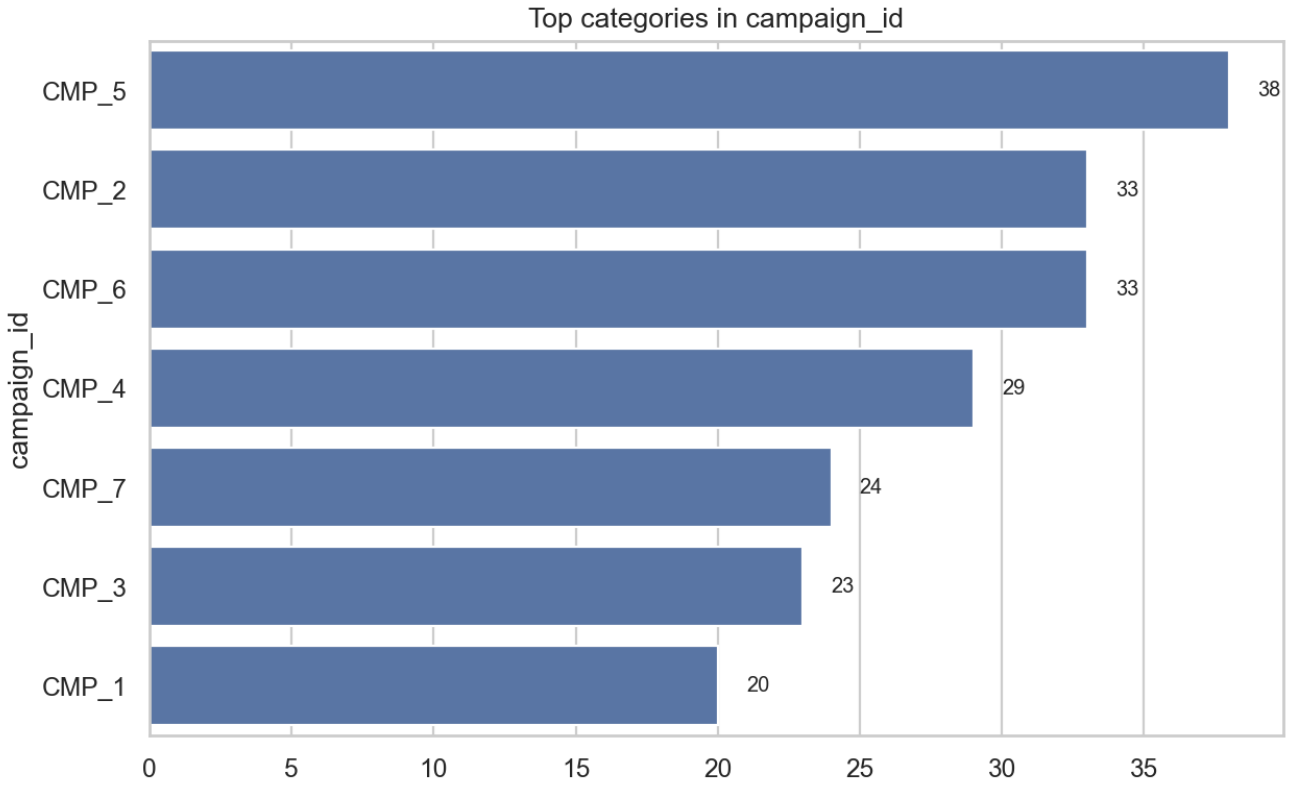
`revenue` exhibits a mean of 24.19 (SD = 15.03) and minimal skew (0.02), suggesting a symmetric spread of monetary values across conversions.



The strongest numeric association is between `user_id` and `revenue` (Pearson r = □0.077, p = 0.280), a weak and statistically non□significant correlation.



The daily mean of `user_id` shows a flat trend (slope ≈ 0.42, r = 0.047, p = 0.659), indicating no meaningful temporal drift in user identifiers.



Campaign participation is dominated by a few categories, with `CMP_5` (38 occurrences) and `CMP_2`/`CMP_6` (33 each) comprising the majority of the 200 records.

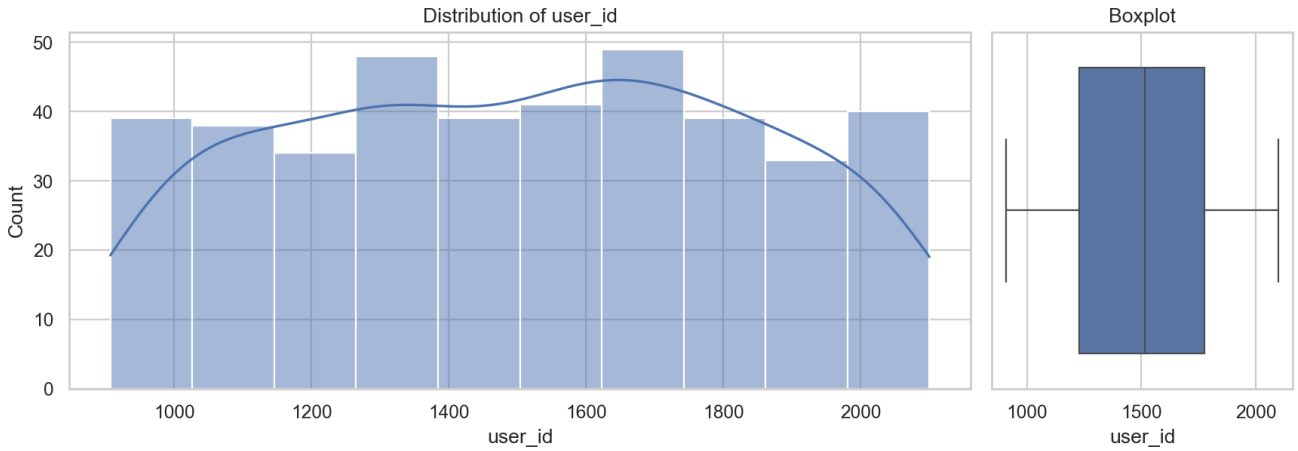
normalized e2b68f980141aba7398d71ae25b17067 4566759e

Summary

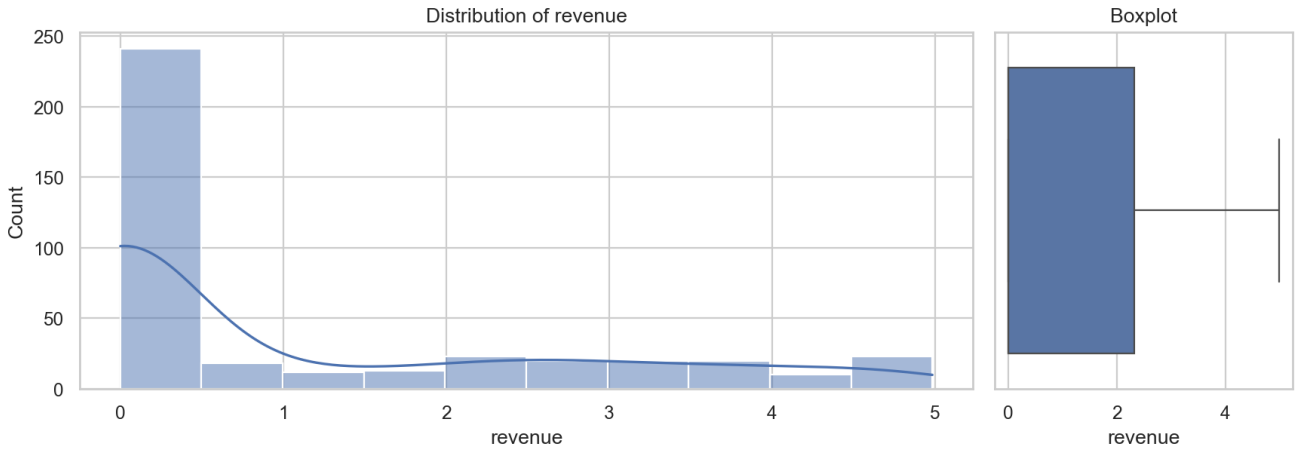
Overall Description:
The dataset captures 400 ad□interaction events, each identified by a unique click and linked to a user, timestamp, ad, campaign, click outcome, and generated revenue. Visualizations reveal the distribution of user identifiers and revenue, a strong positive relationship between clicks and revenue, minimal temporal trends in user activity, and the dominance of a few campaign categories. Together, these plots provide a concise statistical portrait of the ad performance landscape.

Rows	Cols	Numeric	Categorical
400	**7**	**3**	**4**

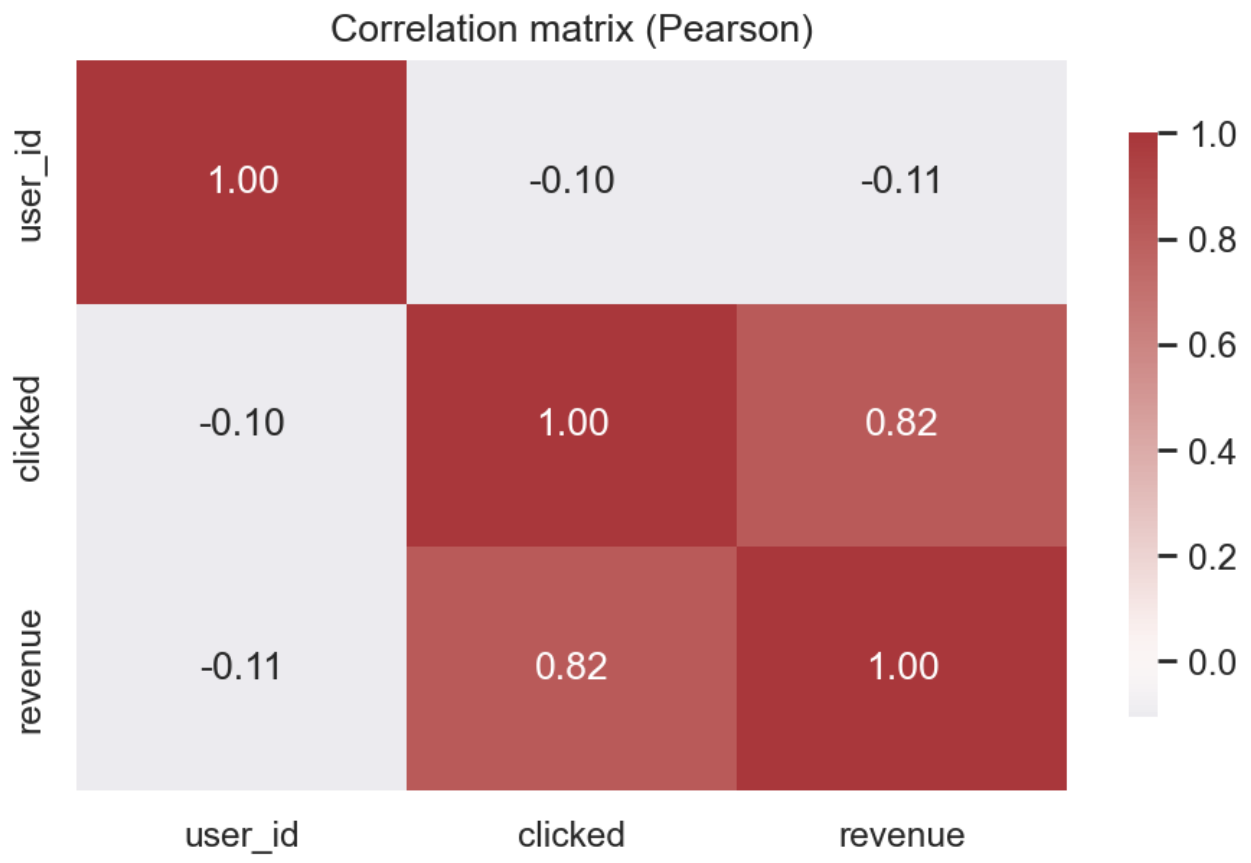
A KPI card confirms the dataset’s modest size (400 rows, 7 columns) and that all variables are fully populated, with numeric fields ready for quantitative analysis.



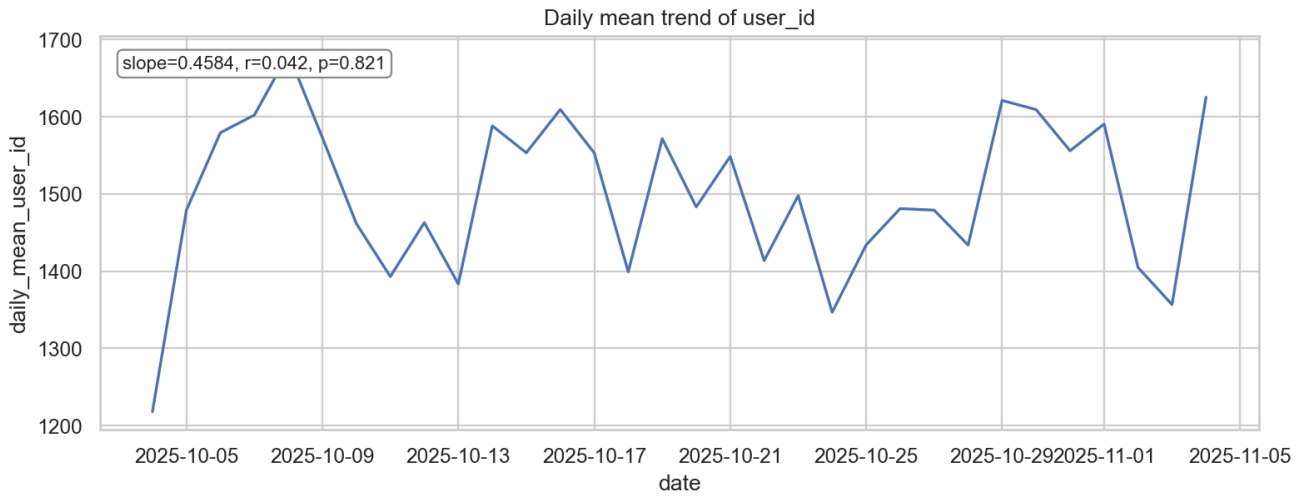
The `user_id` values are approximately normally distributed (mean ≈ 1505, σ ≈ 338) with negligible skew (□0.02), indicating a balanced representation of users across the sample.



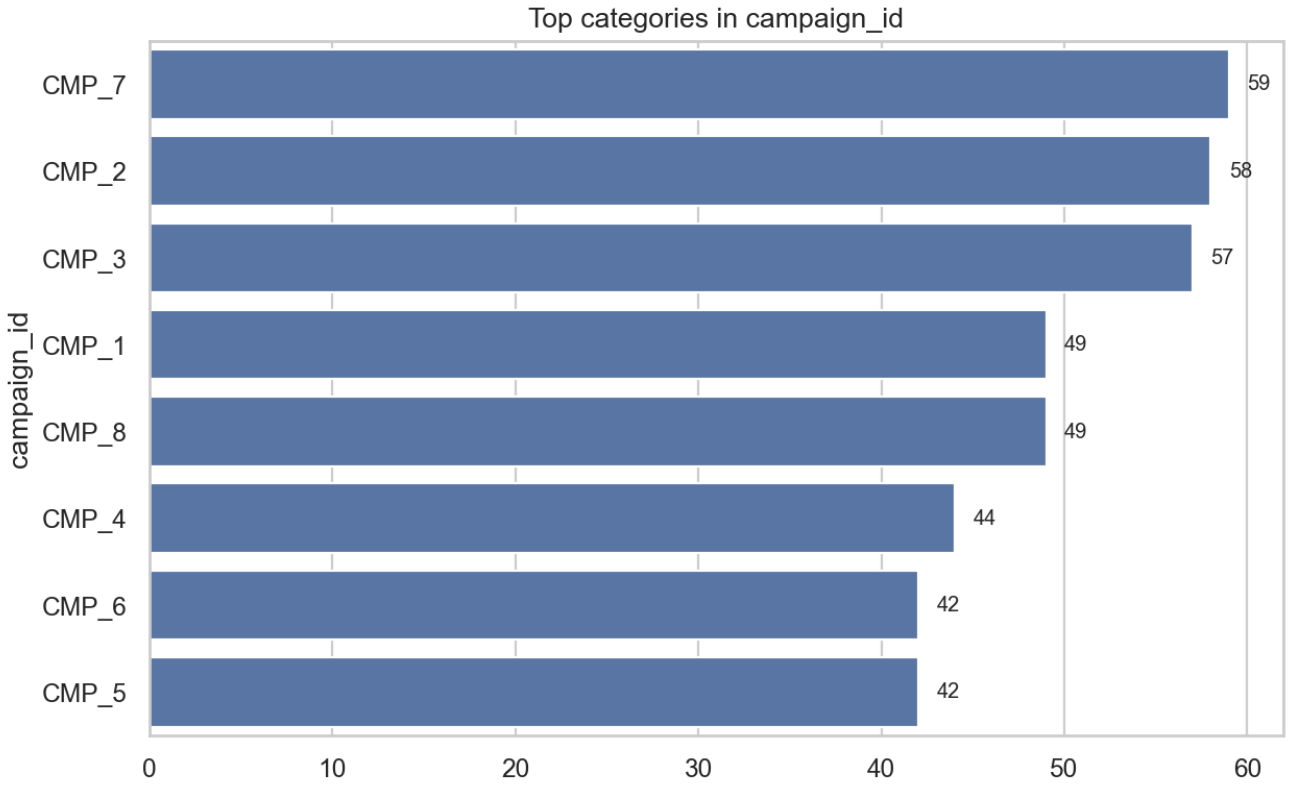
Revenue exhibits a right□skewed distribution (mean ≈ 1.13, σ ≈ 1.60, skew ≈ 1.07), suggesting most clicks generate low revenue while a minority produce higher values.



The strongest numeric association is between `clicked` and `revenue` (r = 0.823, p < 0.001), confirming that clicks are tightly linked to revenue generation. Smaller, statistically significant negative correlations appear for `user_id` with both `revenue` (r = □0.106, p = 0.034) and `clicked` (r = □0.101, p = 0.043).



Daily mean `user_id` shows an almost flat trend (slope = 0.4584, r = 0.042, p = 0.821), indicating no meaningful temporal drift in the user population over the observed period.



Campaign activity is concentrated in a few categories: `CMP_7` (59 clicks) and `CMP_2` (58 clicks) lead, followed closely by `CMP_3`, `CMP_1`, and `CMP_8`, highlighting a skewed distribution of campaign exposure.

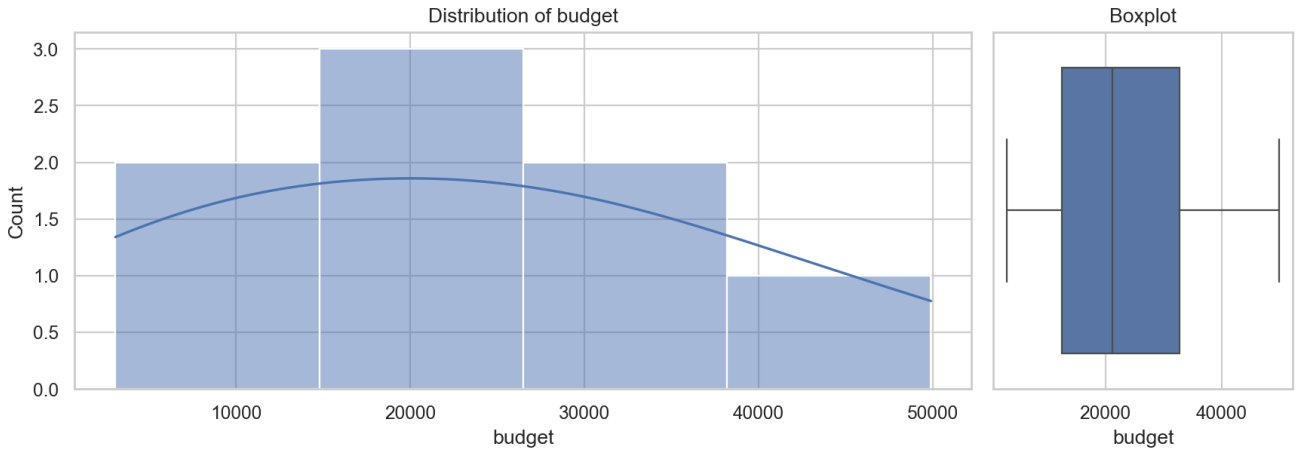
normalized ec699aa6196a357938b1fbb8a7230aae bfa48161

Summary

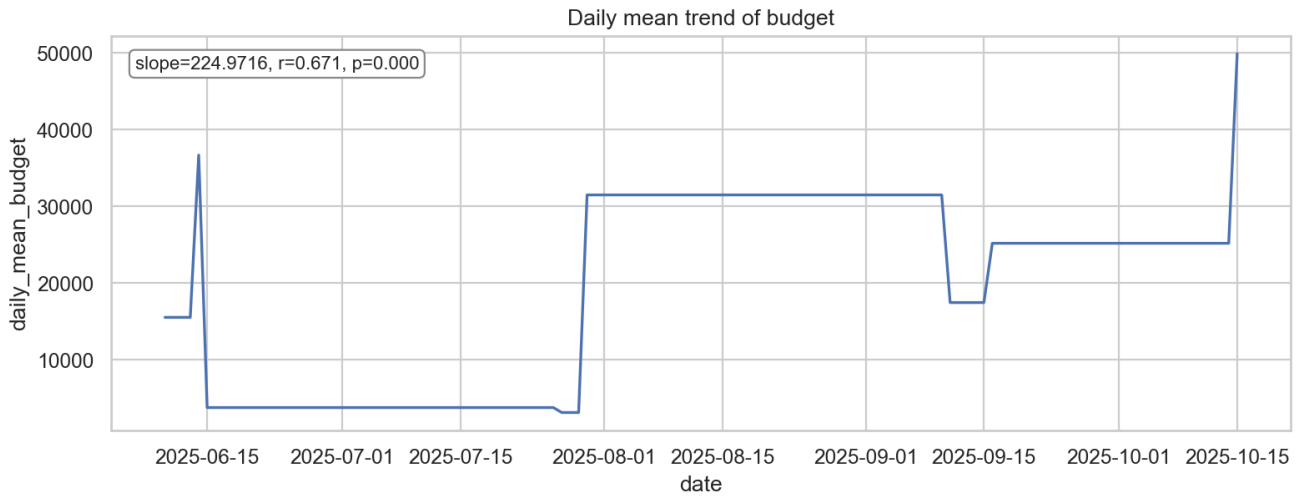
Overall Description:
The dataset captures eight marketing campaigns, each identified by a unique `campaign_id` and `campaign_name`, with associated start dates, end dates, and allocated budgets. Visualizations reveal a modest sample size, a roughly symmetric budget distribution centered around ≈ $22.9 k, a statistically significant upward trend in daily average budget over time, and an even spread of campaign identifiers across the limited records.

Rows	Cols	Numeric	Categorical
8	**5**	**1**	**4**

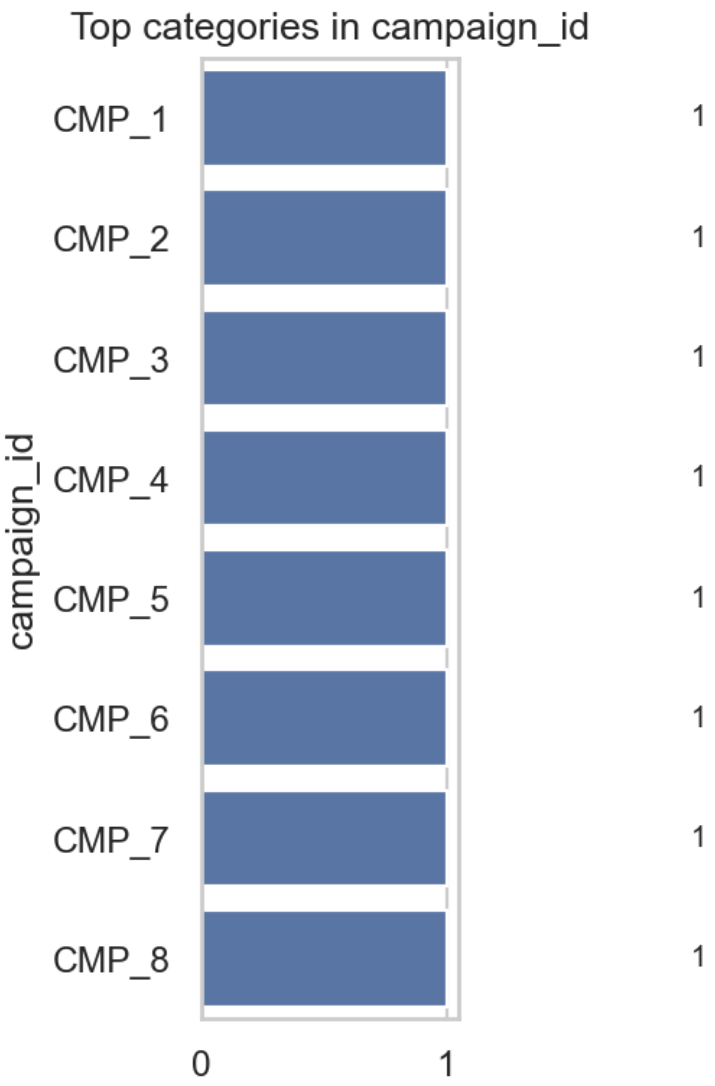
The KPI card confirms the dataset’s compact size (8 rows, 5 columns) and indicates that all columns are numeric or categorical without missing values.



The budget distribution shows a mean of $22,862 with a standard deviation of $16,213 and a slight positive skew (0.34), suggesting most campaigns cluster near the mean with a few higher budget outliers.



The time□series of daily mean budgets exhibits a positive slope of ≈ 225 per day, a correlation of 0.671, and a highly significant p□value ($p < 0.001$), indicating a robust upward budgeting trend across campaign start dates.



The bar chart of `campaign_id` reveals that each of the top five identifiers (CMP_1–CMP_5) appears exactly once, reflecting the dataset’s one□to□one mapping of campaigns to records.

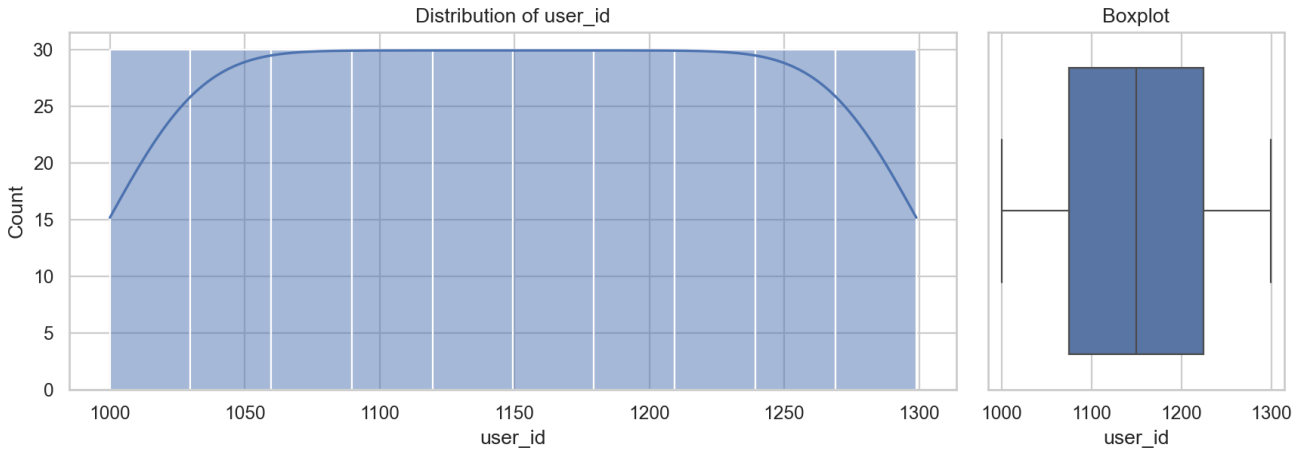
users ea443762

Summary

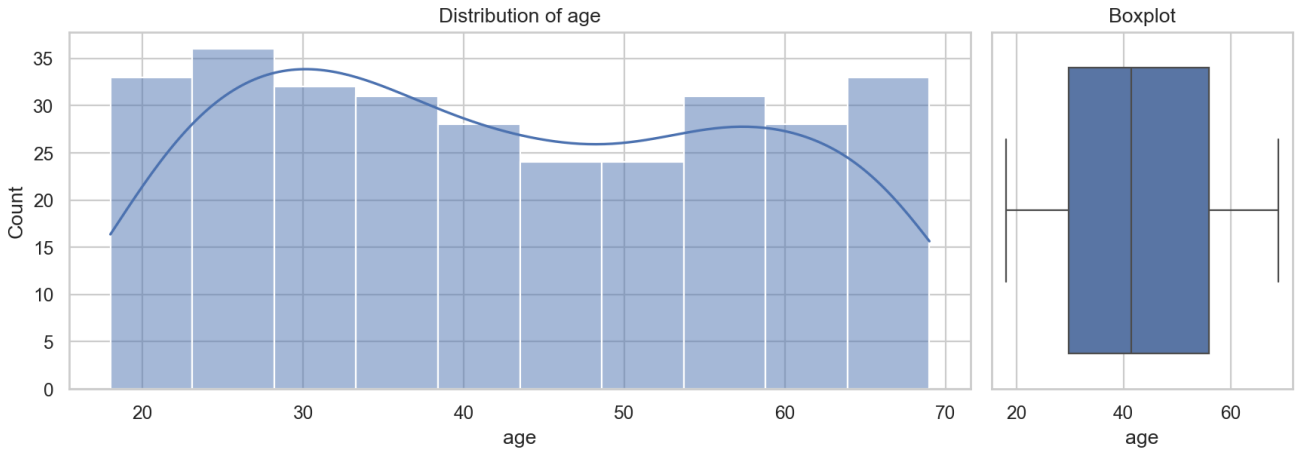
Overall Description:
The dataset captures 300 user records with demographic attributes (age, gender, country) and a signup timestamp. Visualizations reveal the basic distribution of numeric fields, a negligible correlation between user identifiers and age, a flat temporal trend in user IDs, and a gender composition heavily impacted by missing data.

Rows	Cols	Numeric	Categorical
300	**5**	**2**	**3**

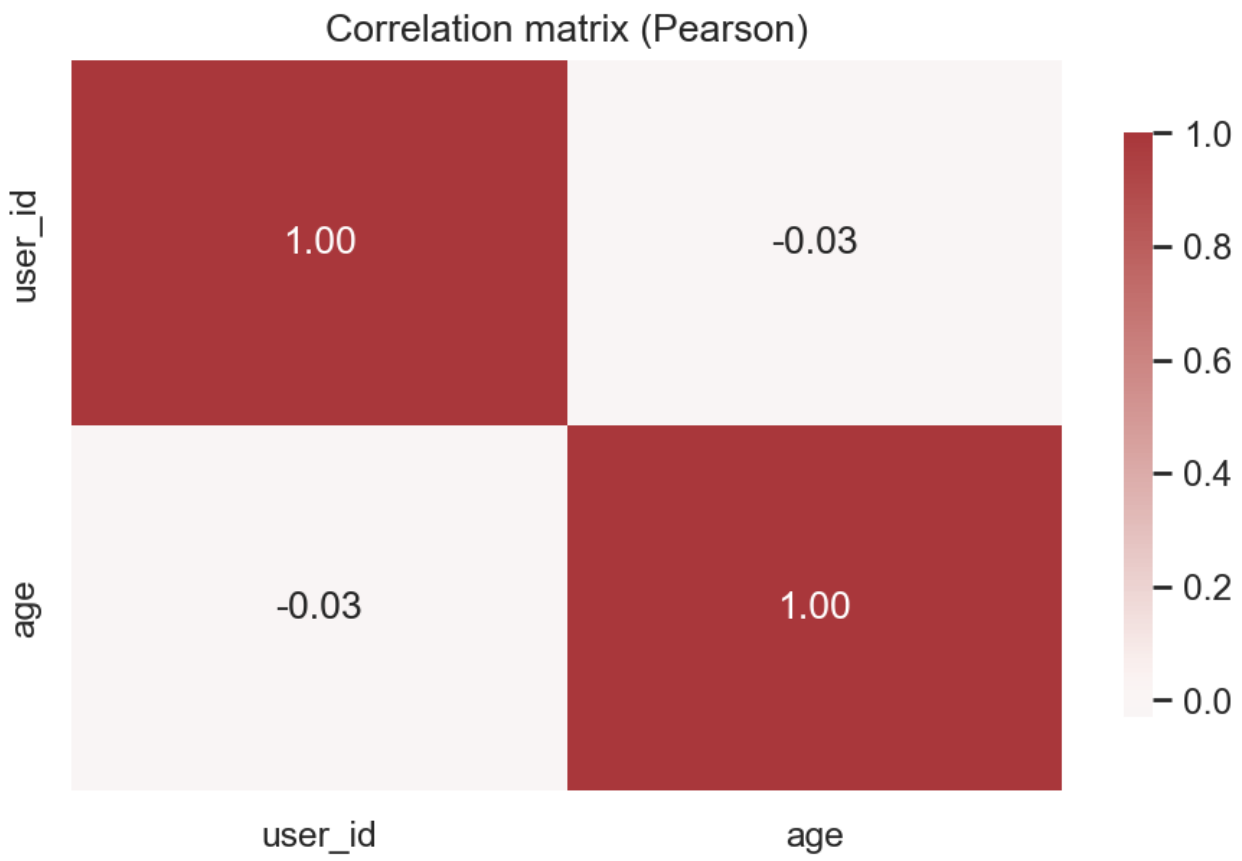
Summarizes dataset dimensions (300 rows, 5 columns) and indicates that only one column (gender) has a notable missing□value proportion (33 %).



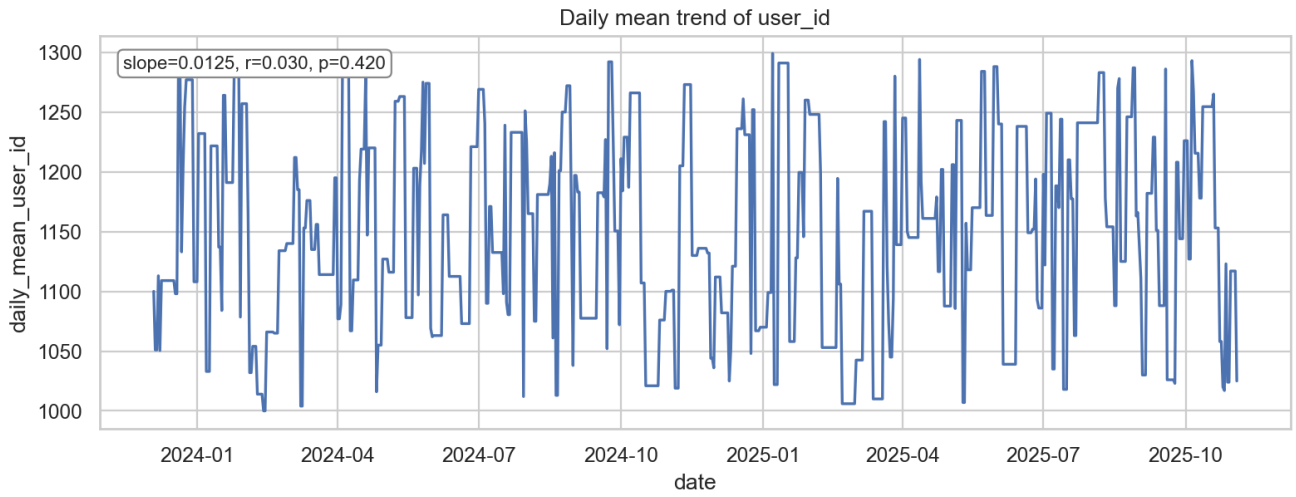
Shows a tightly clustered user_id distribution (mean ≈ 1149.5, σ ≈ 86.7) with zero skew, confirming a uniform identifier range across the sample.



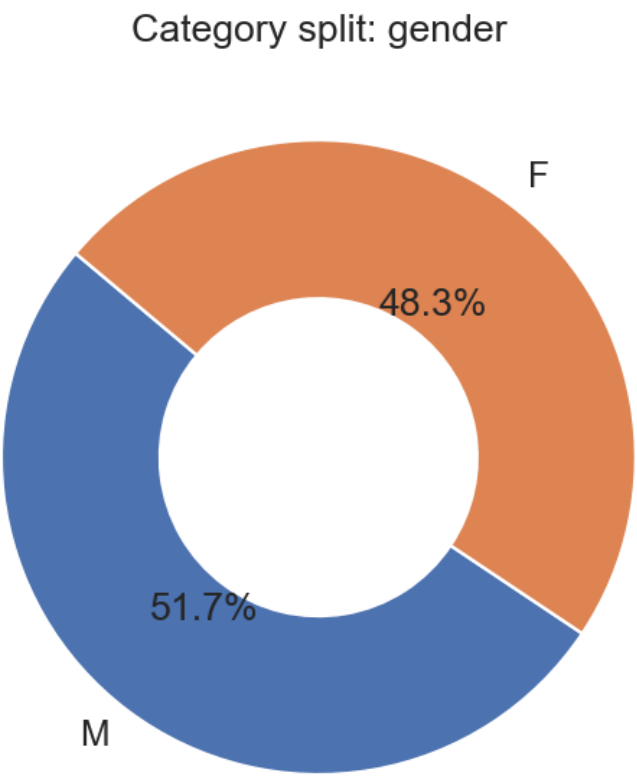
Displays the age profile (mean ≈ 42.7 years, σ ≈ 15.2) with slight positive skew (0.12), suggesting a modest tail toward older users.



Highlights the strongest numeric relationship: user_id vs. age (r = □0.031, p = 0.592), indicating no statistically meaningful linear association.



Plots daily mean user_id over signup_date, yielding a near□zero slope (0.0125) and low correlation (r = 0.030, p = 0.420), implying no temporal drift in identifier values.



Illustrates the gender composition, but with one□third of entries missing, limiting reliable interpretation of the categorical distribution.

Thank you