

DATASET ANALYSIS REPORT

Dataset: D:/Data/period_03/period_03/2024_fb_ads_president_scored_anon.csv
Generated: 2025-06-30 08:28:08

DATASET OVERVIEW:

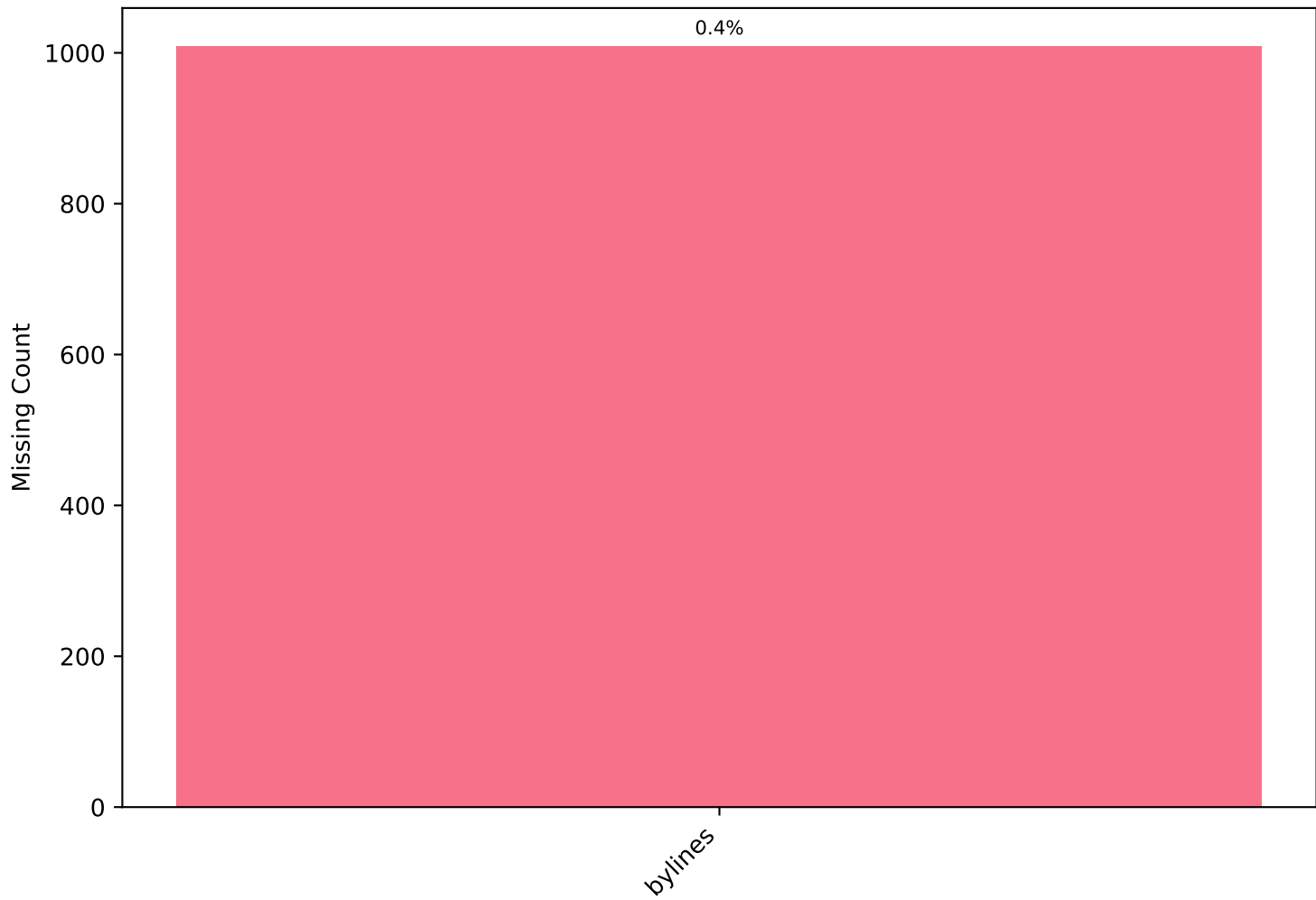
- Total Rows: 246,745
- Total Columns: 41
- Memory Usage: 627.73 MB

COLUMN TYPES:

- int64: 31 columns
- object: 10 columns

Missing Data Analysis

Missing Data by Column



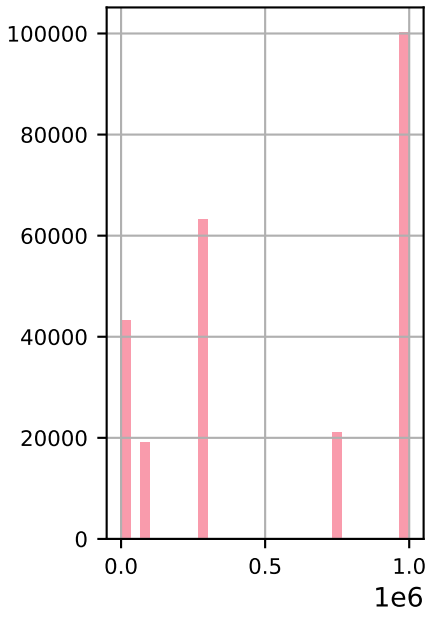
Missing Data Summary

| Column | Missing Count | Percentage |
|---------|---------------|------------|
| bylines | 1,009 | 0.41% |

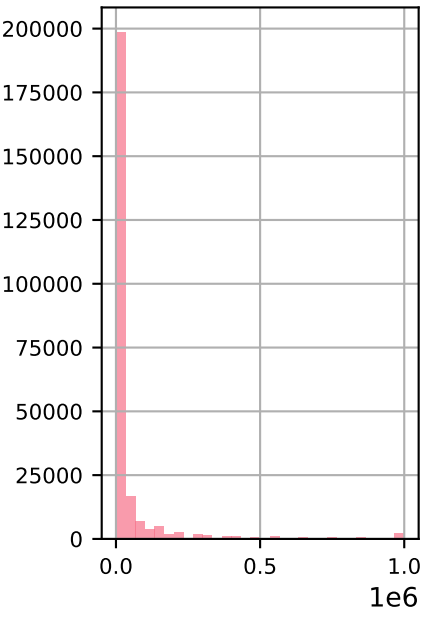
Numeric Columns - Descriptive Statistics

| | estimated_audience_size | estimated_impressions | estimated_spend | scam_illuminating | election_integrity_Truth_illuminating | advocacy_msg_type_illuminating |
|----------|-------------------------|-----------------------|-----------------|-------------------|---------------------------------------|--------------------------------|
| count | 246745 | 246745 | 246745 | 246745 | 246745 | 246745 |
| mean | 5564526 | 101061.2 | 0.072005 | 0.540380 | 0.270120 | 0.220140 |
| std | 4098536799 | 2.50210 | 0.490480 | 0.440410 | 0.490330 | 0.420350 |
| min | 0.0499 | 49.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 25% | 75000 | 49.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 50% | 300000 | 49.0 | 0.0 | 1.0 | 0.0 | 0.0 |
| 75% | 1000000 | 49.0 | 0.0 | 1.0 | 1.0 | 0.0 |
| max | 1000000000 | 499.0 | 1.0 | 1.0 | 1.0 | 1.0 |
| mode | 1000000 | 49.0 | 0.0 | 1.0 | 0.0 | 1.0 |
| skew | -0.05507 | 19.03324 | 1.120190 | 0.481021 | 0.330292 | 0.271292 |
| kurtosis | 1.748376 | 759.0315 | 0.011961 | 1.760940 | 0.221913 | 0.150322 |

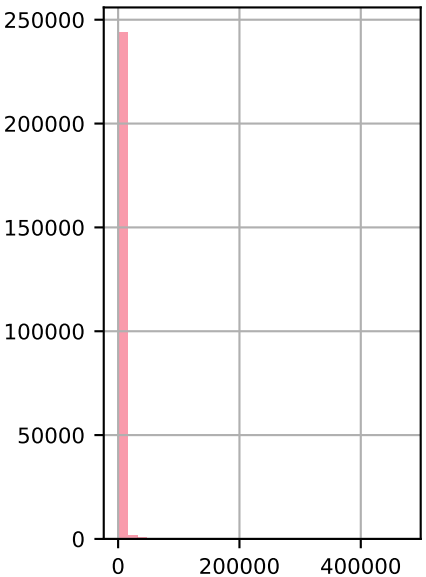
estimated_audience_size
Distribution



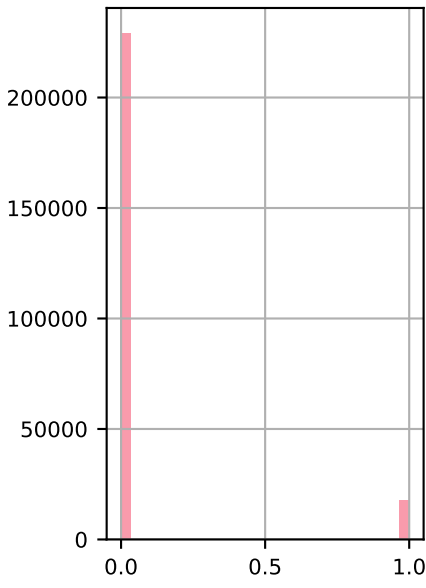
estimated_impressions
Distribution



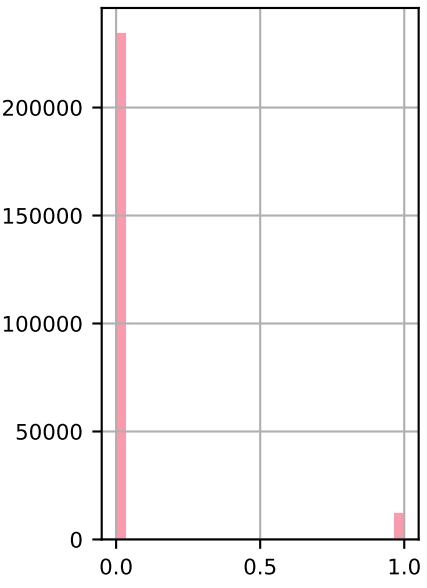
estimated_spend
Distribution



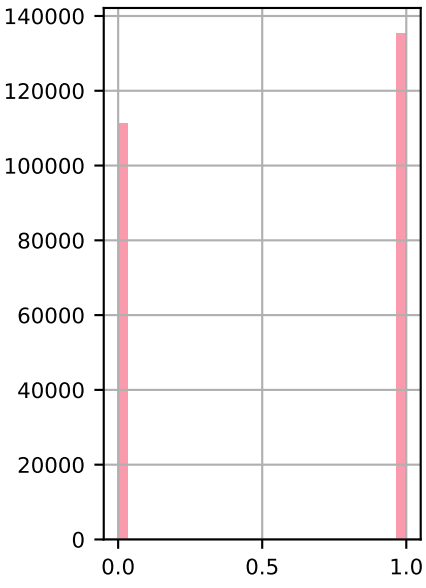
scam_illuminating
Distribution



election_integrity_Truth_illuminating
Distribution



advocacy_msg_type_illuminating
Distribution

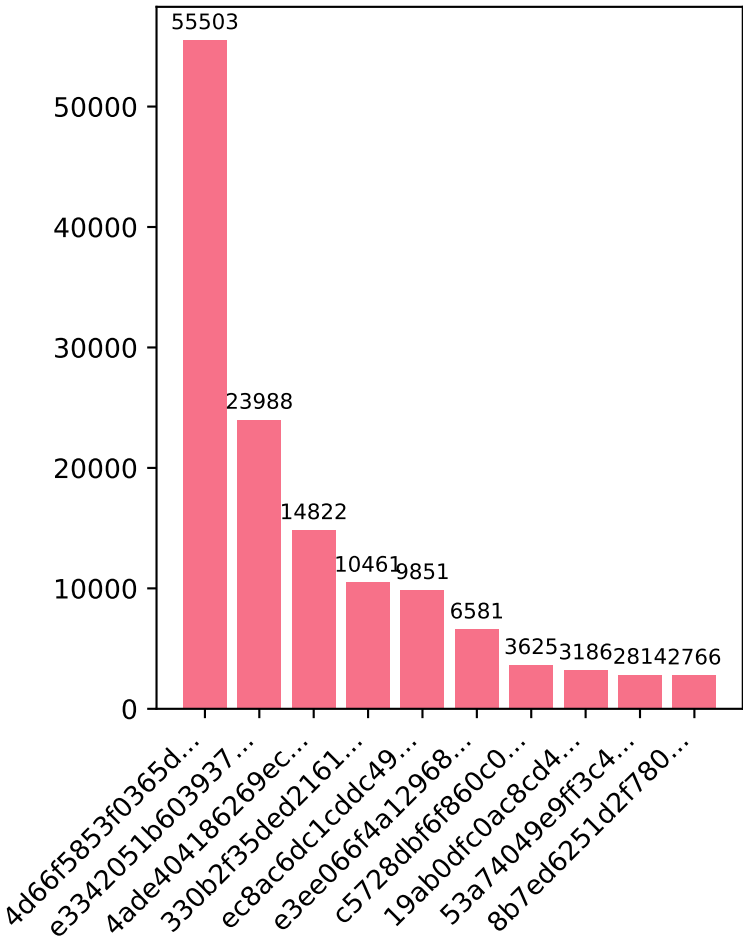


Categorical Columns Analysis

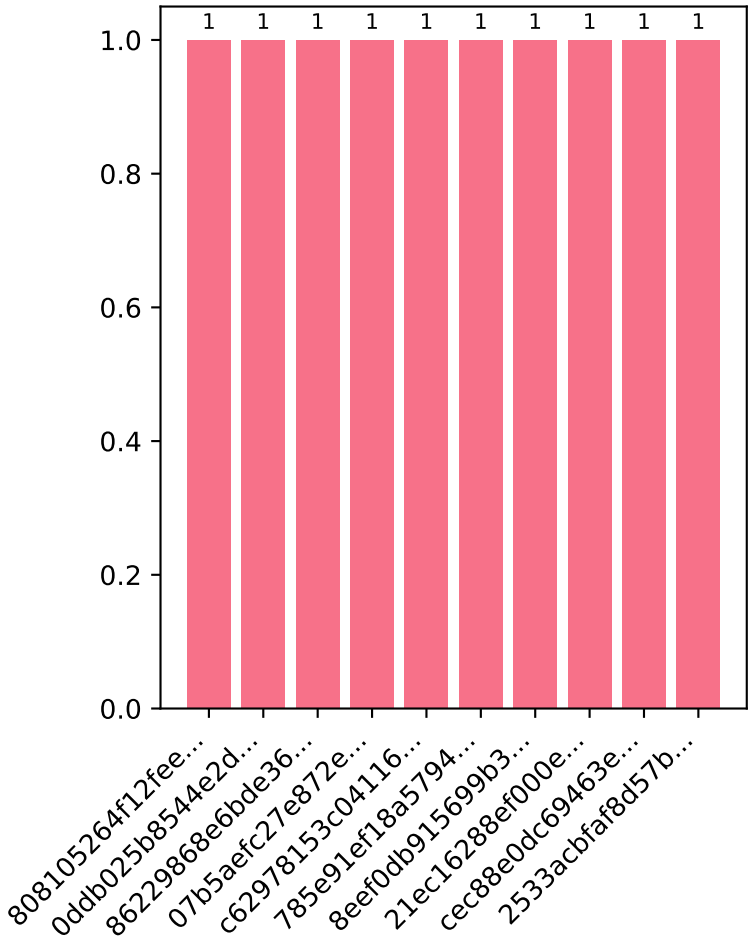
Summary Statistics

| Column | Unique Count | Most Frequent | Count |
|--------------------------|--------------|----------------------|--------|
| page_id | 4475 | 4d66f5853f0365dba0 | 55503 |
| ad_id | 2467458081 | 05264f12fee56 | 1 |
| ad_creation_time | 547 | 2024-10-27 | 8619 |
| bylines | 3790 | HARRIS FOR PRESID | 49788 |
| currency | 18 | USD | 246599 |
| delivery_by_region | 141122 | {} | 30989 |
| demographic_distribution | 215622 | {} | 30989 |
| publisher_platform | 9 | ['facebook', 'instag | 214434 |
| illuminating_score_mean | 26338 | e3b0c44298fc1c149a | 5972 |
| illuminating_mentions | 278 | [] | 73205 |

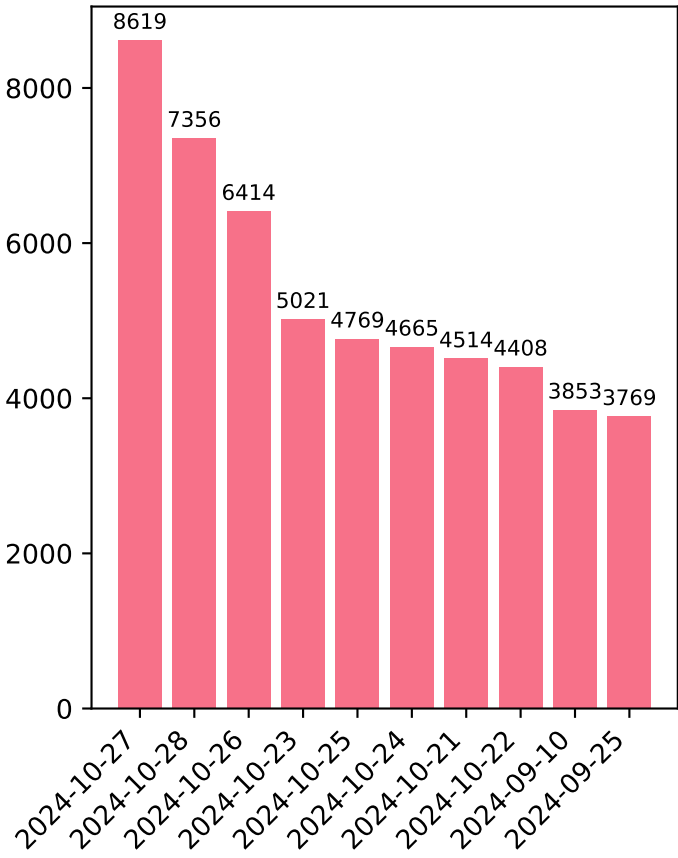
page_id - Top Values



ad_id - Top Values



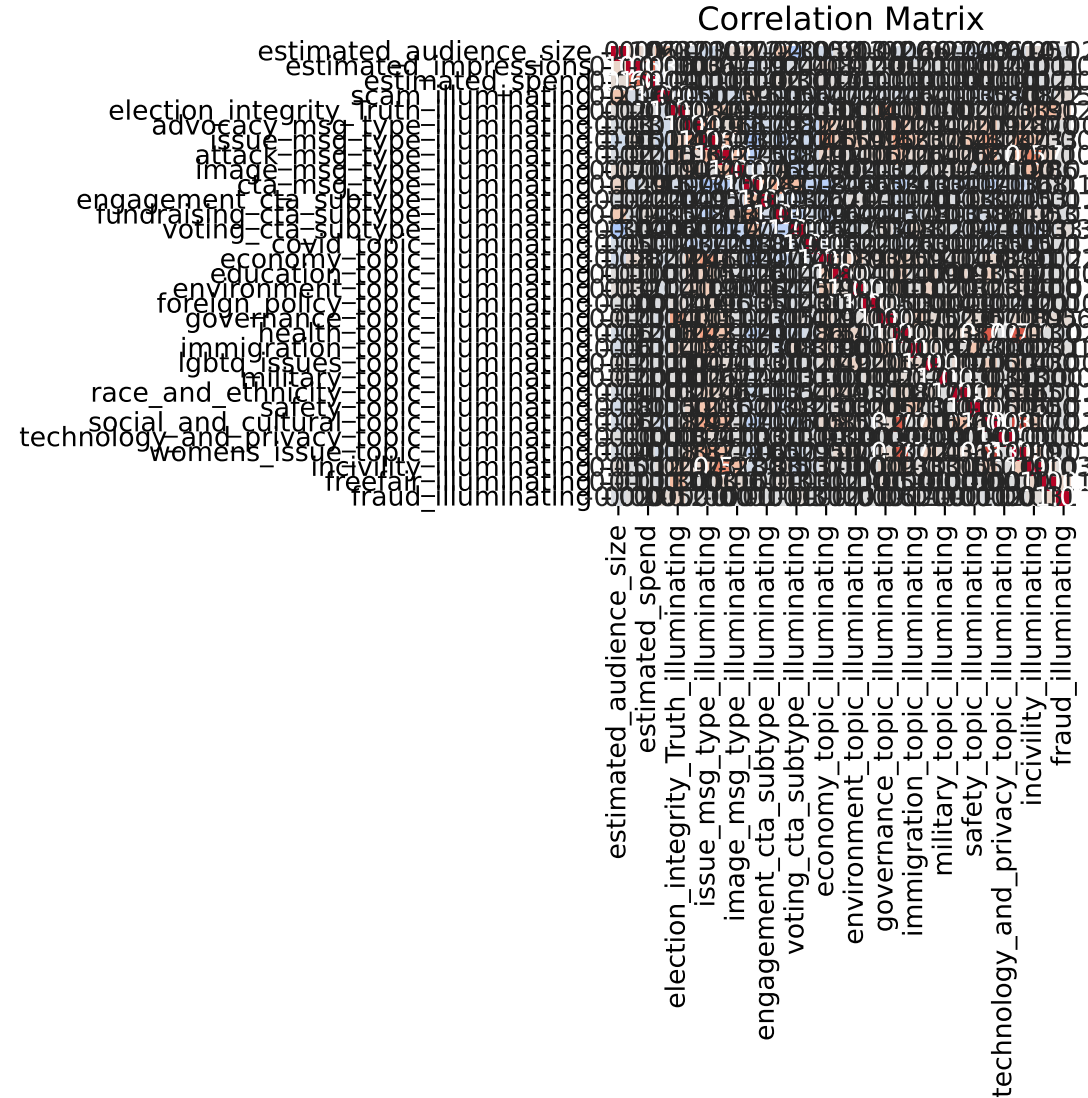
ad_creation_time - Top Values



Correlation Analysis

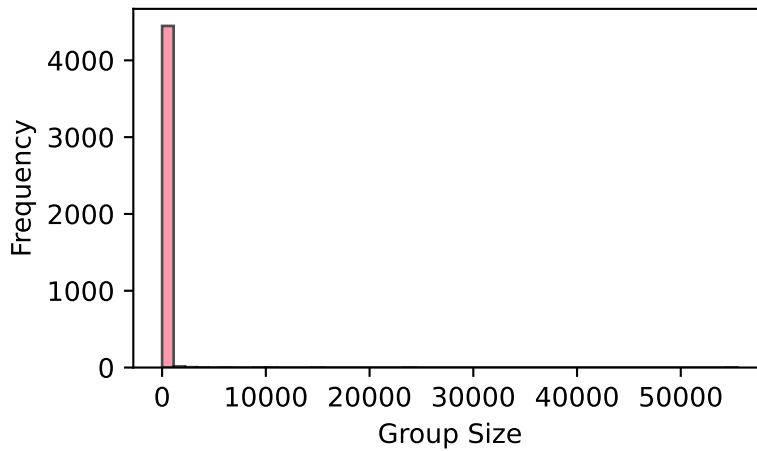
High Correlation Pairs ($|r| > 0.7$)

health_topic_illuminating ↔ social_and_cultural_topic_illuminating: 0.767
health_topic_illuminating ↔ womens_issue_topic_illuminating: 0.765
social_and_cultural_topic_illuminating ↔ womens_issue_topic_illuminating: 0.830



Analysis by Page ID

Group Size Distribution



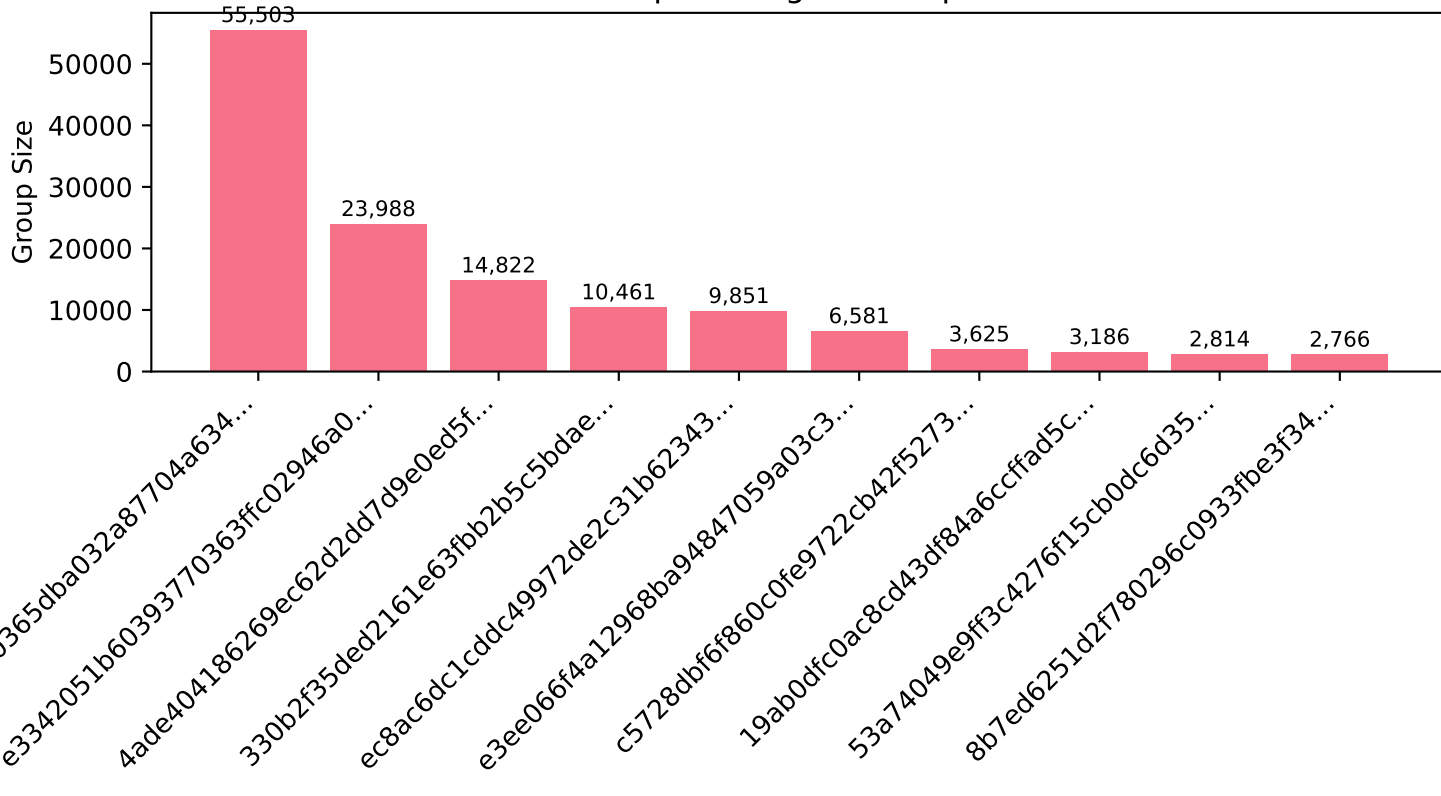
Group Size Statistics

Total Groups: 4,475

Group Size Statistics:

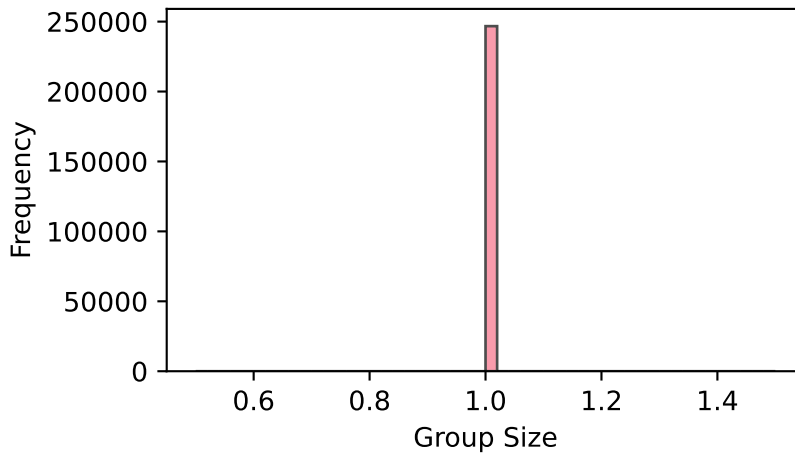
- Mean: 55.1
- Median: 3.0
- Min: 1
- Max: 55,503
- Std Dev: 971.9

Top 10 Largest Groups



Analysis by Page ID + Ad ID

Group Size Distribution



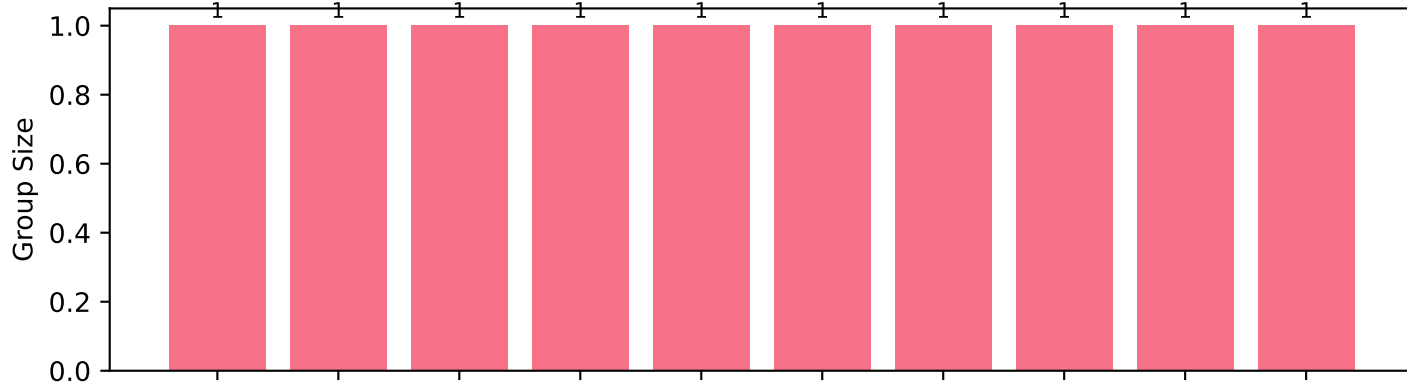
Group Size Statistics

Total Groups: 246,745

Group Size Statistics:

- Mean: 1.0
- Median: 1.0
- Min: 1
- Max: 1
- Std Dev: 0.0

Top 10 Largest Groups



(('0005687f4ac876c135dda2bd5aa6...

(('0005687f4ac876c135dda2bd5aa6...

(('00157b7e28846e0510e1d265c5ea...

(('00157b7e28846e0510e1d265c5ea...

(('00157b7e28846e0510e1d265c5ea...

(('002415d34a27e69e6b9c0a2d1b0b...

(('002a7fde0f1b2e9116e80f42d7df...

(('002a7fde0f1b2e9116e80f42d7df...

(('002a7fde0f1b2e9116e80f42d7df...

(('002a7fde0f1b2e9116e80f42d7df...