DATASET ANALYSIS REPORT

Dataset: 2024_fb_ads_president_scored_anon.csv
Generated: 2025-06-29 20:13:57

DATASET OVERVIEW:
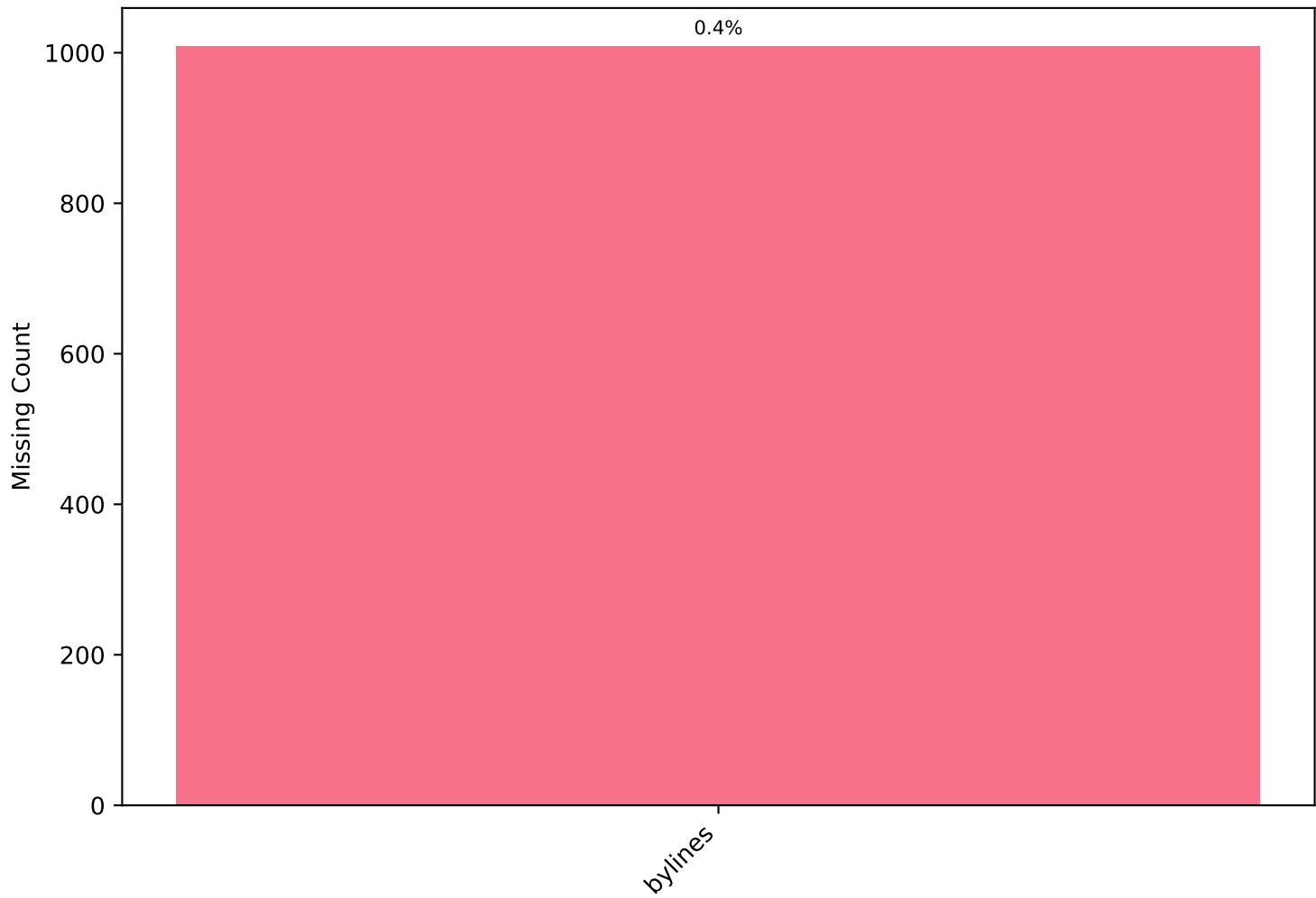• Total Rows: 246,745
• Total Columns: 41
• Memory Usage: 627.73 MB

COLUMN TYPES:
• int64: 31 columns
• object: 10 columns

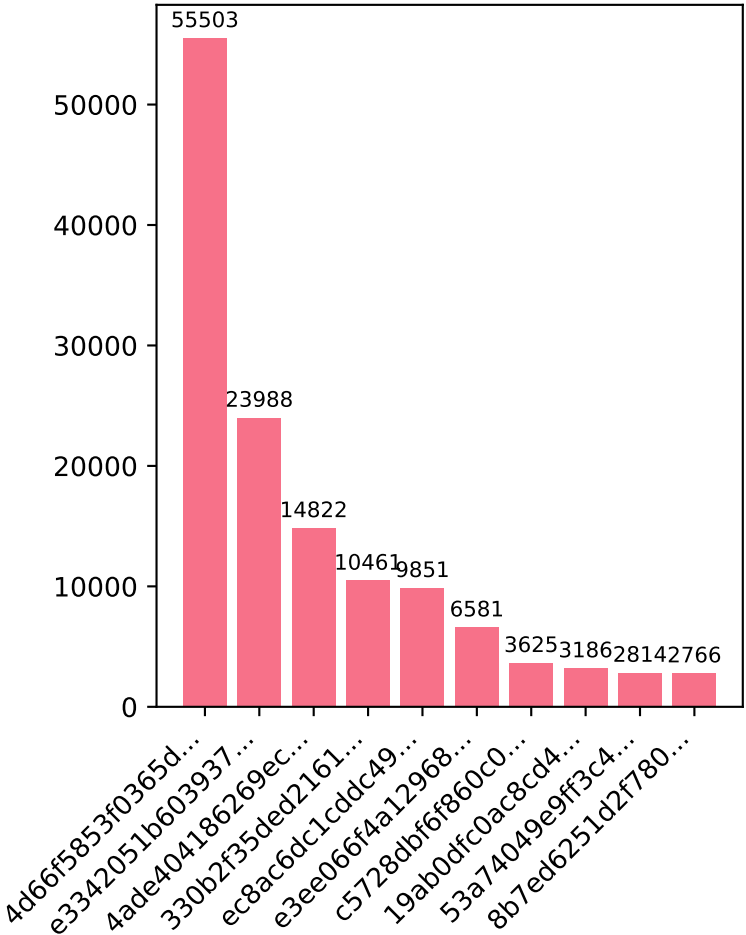# Missing Data Analysis

Missing Data by Column

0.4%

Missing Count

bylines

Missing Data Summary

Column	Missing Count	Percentage
bylines	1,009	0.41%

Numeric Columns - Descriptive Statistics

	estimated_audience_size	estimated_impressions	estimated_spend	...	illumin...
count	246724...				246745.0
mean	5564...	261061.2	0.07	0.05 0.54 0.38 0.27 0.22 0.57 0.12 0.22 0.14 0.02 0.12 0.01 0.02 0.00 0.02 0.10 0.03 0.00 0.00 0.01 0.03 0.10 0.00 0.08 0.18 0.00	0.003
std	40986...	367992.5	0.25 0.21 0.49 0.48 0.44 0.41 0.49 0.33 0.42 0.35 0.15 0.32 0.11 0.14 0.07 0.15 0.31 0.18 0.05 0.04 0.11 0.18 0.30 0.03 0.27 0.39 0.08	0.051
min	0.0	499.0 49.0	0.0 (all 0.0)		0.0
25%	75000	499.0 49.0	0.0 (all 0.0)		0.0
50%	300003	499.0 49.0	0.0 0.0 1.0 0.0 0.0 0.0 1.0 0.0 ...		0.0
75%	1000002	2499.0 49.0	0.0 0.0 1.0 1.0 1.0 0.0 1.0 0.0 ...		0.0
max	1000100...	499999	1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0	1.0
mode	1000000	499.0 49.0	0.0 0.0 1.0 0.0 0.0 0.0 1.0 0.0 ...		0.0
skew	-0.05	5.07 19.03 3.32 4.12 0.19 0.48 1.02 0.33 -0.29 2.27 1.29 2.03 6.10 2.30 8.17 6.64 3.67 6.00 2.50 6.17 7.52 1.36 8.85 1.62 2.56 9.37 3.07 4.60 2.36	9.392
kurtosis	-1.74	28.37 767.59 9.03 15.01 1.96 1.76 0.94 0.22 1.91 3.15 -0.32 2.12 35.22 3.32 64.84 42.08 34.93 34.02 4.28 34.83 54.68 4.47 75.42 24.68 4.58 60.77 9.44 50.86 4.03	4.035

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	246745808	105264f12fee56b	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_scored_me	26338	e3b0c44298fc1c149a	5972
illuminating_mention	278	[]	73205

page_id - Top Values

Bar values (left to right):
- 4d66f5853f0365d...: 55503
- e3342051b603937...: 23988
- 4ade404186269ec...: 14822
- 330b2f35ded2161...: 10461
- ec8ac6dc1cddc49...: 9851
- e3ee066f4a12968...: 6581
- c5728dbf6f860c0...: 3625
- 19ab0dfc0ac8cd4...: 3186
- 53a74049e9ff3c4...: 2814
- 8b7ed6251d2f780...: 2766

ad_id - Top Values

Bar values (left to right), all equal to 1:
- 808105264f12fee...: 1
- 0ddb025b8544e2d...: 1
- 86229868e6bde36...: 1
- 07b5aefc27e872e...: 1
- c62978153c04116...: 1
- 785e91ef18a5794...: 1
- 8eef0db915699b3...: 1
- 21ec16288ef000e...: 1
- cec88e0dc69463e...: 1
- 2533acbfaf8d57b...: 1

ad_creation_time - Top Values

Bar values (left to right):
- 2024-10-27: 8619
- 2024-10-28: 7356
- 2024-10-26: 6414
- 2024-10-23: 5021
- 2024-10-25: 4769
- 2024-10-24: 4665
- 2024-10-21: 4514
- 2024-10-22: 4408
- 2024-09-10: 3853
- 2024-09-25: 3769
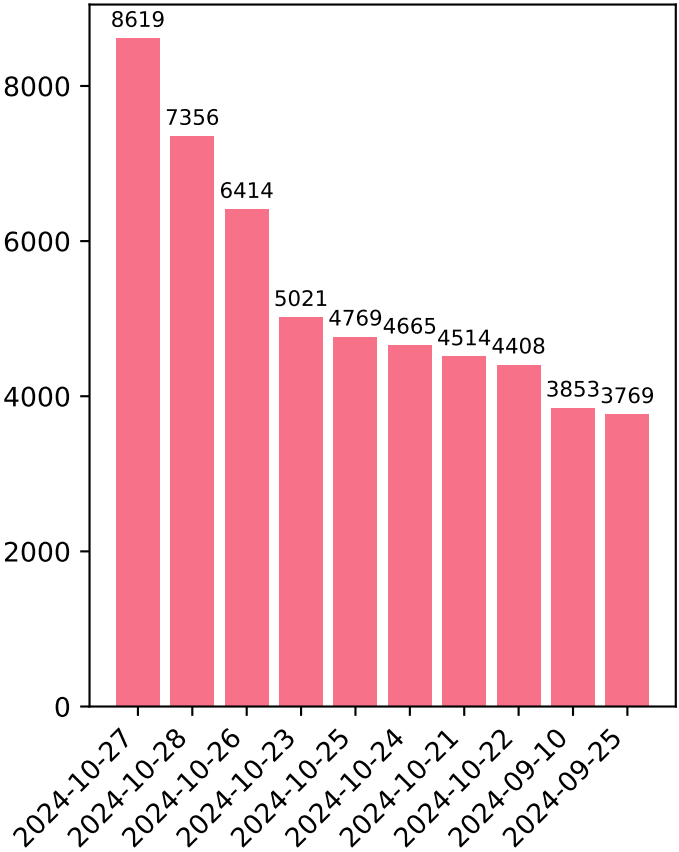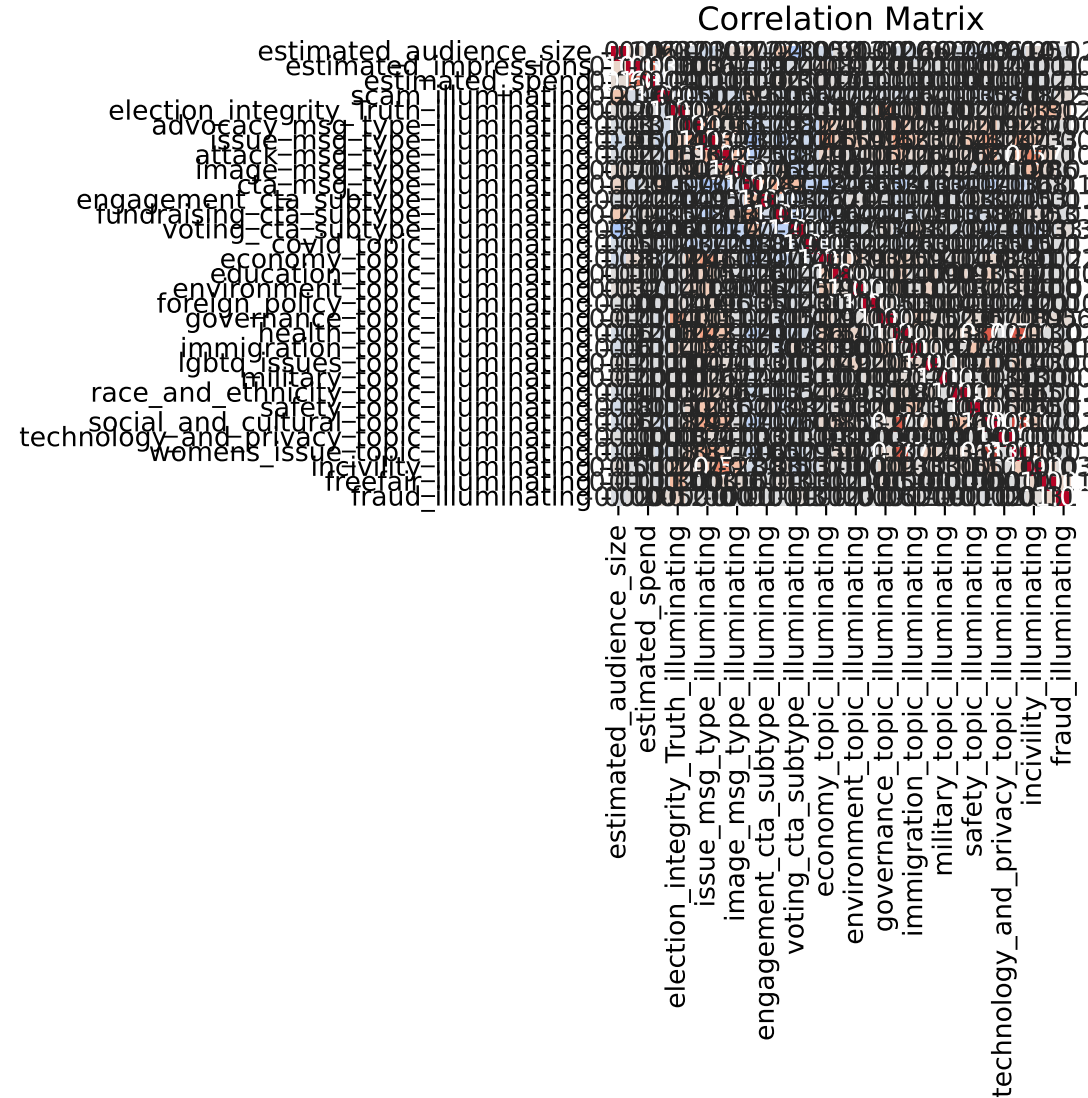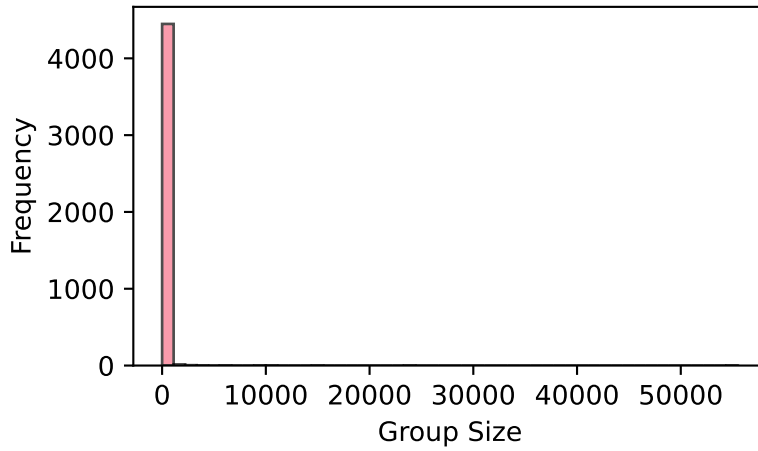
Correlation Analysis

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

Correlation Matrix

Analysis by Page ID

Group Size Distribution

Group Size Statistics
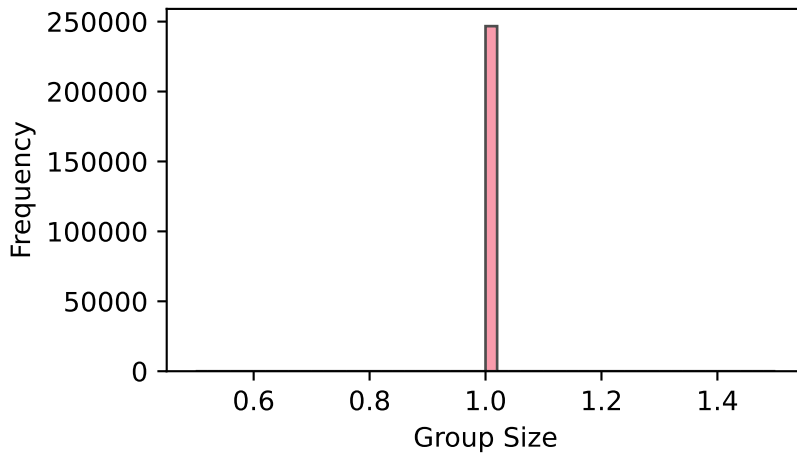
Total Groups: 4,475

Group Size Statistics:
• Mean: 55.1
• Median: 3.0
• Min: 1
• Max: 55,503
• Std Dev: 971.9

Top 10 Largest Groups
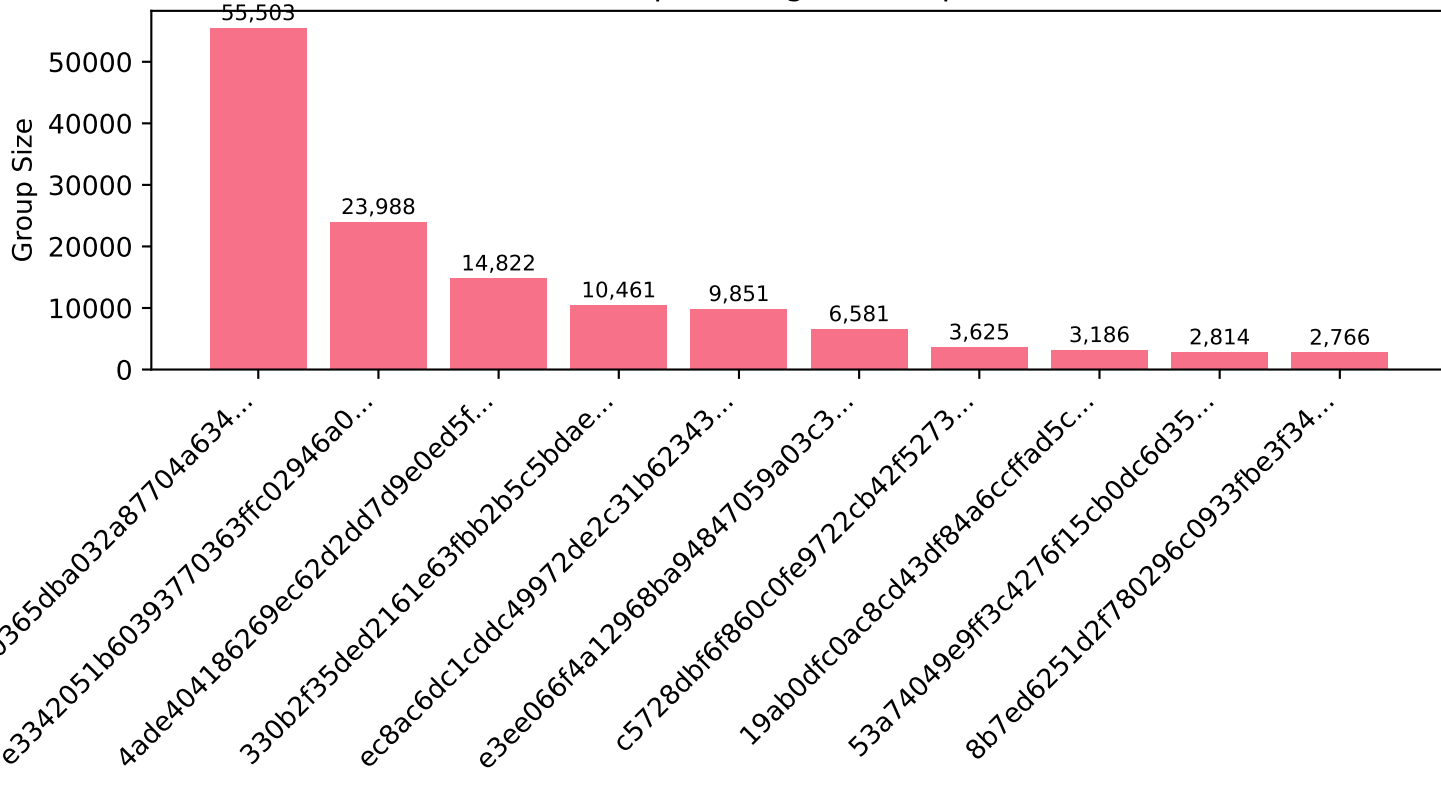
Analysis by Page ID + Ad ID

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