Exploratory Data Analysis Report

1. Data Inventory and Quality Assessment

Dataset Overview

The dataset provides a comprehensive view of marketing performance and external drivers across 40 unique geographic regions. It contains 6,240 total observations covering 156 weeks from January 25, 2021, to January 15, 2024.

Completeness and Integrity

The dataset exhibits exceptional data hygiene, with a 0.0% missing data rate across all 18 variables. This provides a robust foundation for econometric modeling.

Outlier Analysis

IQR-based outlier detection identified specific channels with high variance that may require normalization or holiday-effect adjustments during the modeling phase:

- **Channel 2:** Highly volatile, with 15.1% of observations flagged as outliers for both spend and impressions.
- **Organic Channel 0:** Exhibited 4.9% outliers in impressions. Notably, this channel carries no spend, serving as a pure organic baseline.
- **Channel 1:** Showed 3.4% outliers in both spend and impressions.

Business Logic Validation

A rigorous validation check was performed to cross-reference media spend against impressions. The analysis confirms 100% logic alignment: there are zero records where spend was recorded without corresponding impressions, ensuring the integrity of the media cost data.

2. Descriptive Statistics and Revenue Performance

Aggregate Revenue Metrics

The high-level performance metrics for the audited period are as follows:

- **Total Revenue: $1,318,699,437.74**
- **Average Weekly Revenue: $211,330.04**
- **Total Conversions: 65,930,714**

Revenue Efficiency

The Average Revenue per Conversion is $0.02. It is important to note that this metric is nearly constant across the dataset (standard deviation of 0.000099), suggesting a fixed revenue-to-conversion relationship. Weekly performance volatility is represented by a revenue standard deviation of $121,605.91, indicating significant fluctuations that the MMM must account for through seasonality and media drivers.

3. Media Spend Distribution and Channel Analysis

Total Investment Breakdown

Media investment is heavily skewed toward a single channel, creating a potential concentration risk.

Channel	Total Spend	Percent of Total
Channel 3	$87,860,360	51.15%
Channel 0	$40,501,710	23.58%
Channel 1	$31,320,300	18.23%
Channel 2	$12,079,350	7.03%

Dominant Channels

Channel 3 is the primary investment vehicle, commanding over half of the total budget ($87.86M). **Channel 0** represents the second-largest investment at 23.58%.

Spending Characteristics

Distribution analysis reveals that **Channel 0** and **Channel 1** utilize a more stable spending cadence. Conversely, **Channel 3** and **Channel 2** exhibit high variance and frequent outliers, suggesting these channels are likely used for tactical "burst" campaigns or highly seasonal scaling.

4. Marketing Efficiency Metrics (Naive ROI and CPM)

Note: Naive ROI reflects revenue generated in the same week as the spend and does not account for time-lagged or carryover effects.

Cost Efficiency (CPM)

Cost Per Mille (CPM) analysis identifies **Channel 0** as the most cost-effective placement for reach.

Channel	CPM
Channel 1	$9.64
Channel 3	$7.79
Channel 2	$7.43
Channel 0	$7.33

Return on Investment (Naive ROI)

Efficiency metrics show a dramatic disparity between paid channels.

Channel	Naive ROI
Channel 2	38.08

Channel 1	30.25
Channel 0	28.73
Channel 3	14.55

Efficiency Gap

A critical **misallocation of resources** is evident: Channel 3 receives 51.15% of the total budget despite yielding the lowest Naive ROI (14.55). Channel 3's efficiency is less than half that of Channel 2 (38.08). This represents a major optimization opportunity to shift capital from low-performing reach to high-performing conversion drivers.

5. Time-Series and Correlation Analysis

Temporal Trends

Revenue and media spend exhibit synchronized fluctuations over the three years. Clear peaks and troughs suggest that while baseline revenue exists, aggressive spending in high-variance channels like Channel 3 significantly influences weekly performance spikes.

Revenue Drivers

Correlation analysis identifies the following variables with the strongest relationship to `total_revenue`:

1. **Conversions (0.99):** Note that this represents a mathematical identity (Revenue = Conversions × Rate) rather than an independent driver.
2. **Channel 3 Spend (0.47):** The strongest paid media driver.
3. **Channel 0 Spend (0.41):** The second-strongest media driver.

Multicollinearity Check

Variance Inflation Factor (VIF) scores were calculated for all spend variables to ensure model stability. All scores remain below the 5.0 threshold, indicating that multicollinearity will not compromise the coefficient reliability in the final MMM.

Variable	VIF
Channel 3 Spend	4.10
Channel 0 Spend	2.94
Channel 1 Spend	2.16
Channel 2 Spend	1.64

6. Geographic and External Factor Impact

Geographic Performance

Revenue performance is concentrated in key markets. The following table highlights the Top 10 regions by volume.

Rank	Geo	Total Revenue	Revenue per Capita
1	Geo36	$65,816,480	67.86
2	Geo39	$64,482,450	80.71
3	Geo10	$61,910,790	62.28
4	Geo23	$56,731,840	57.80

5	Geo31	$53,677,540	61.46
6	Geo7	$51,532,210	70.21
7	Geo2	$49,370,240	55.32
8	Geo25	$48,727,410	62.54
9	Geo17	$44,085,240	56.70
10	Geo15	$43,162,740	63.45

While Geo36 leads in total volume, **Geo39** is the most efficient market with a **Revenue per Capita of 80.71**.

Promotional Impact

The analysis of promotion periods reveals a 6.8% lift in average revenue ($217,832 vs $203,946.90). However, the **T-test P-value is 1.0000**, which is statistically anomalous. This indicates that the current promotion strategy, as executed, provides no verifiable incremental value, and the groups are statistically indistinguishable.

External Controls

- **Sentiment Score:** Shows a negligible correlation (-0.009) with revenue, suggesting brand sentiment is not a primary driver of short-term sales.
- **Competitor Sales:** Exhibits a negative correlation (-0.117), confirming that aggressive competitor activity actively erodes our market share.

7. Strategic Marketing Recommendations

Budget Reallocation

The current budget distribution is fundamentally inefficient. We are over-investing in Channel 3 (51% of spend) while it delivers the lowest ROI.

Recommendation: Aggressively reallocate a portion of the Channel 3 budget to **Channel 2**. Despite receiving only 7% of the budget, Channel 2 is our most efficient vehicle. Shifting funds here will capitalize on superior ROI and likely drive higher aggregate revenue.

Geographic Targeting

Volume should not be the only metric for geographic investment.

Recommendation: While maintaining Geo36 for scale, pivot incremental investment toward **Geo39** and **Geo16** (80.71 and 80.34 per capita, respectively). These regions demonstrate a higher propensity to spend and offer better returns on localized marketing efforts.

Promotion Strategy

The current promotional framework is failing to drive statistical significance.

Recommendation: Halt current spending on statistically insignificant promotions. The P-value of 1.0000 suggests the current promo mechanics are effectively "noise." We recommend a complete overhaul of promotional offers and a move toward A/B testing different mechanics (e.g., discount vs. bundle) to identify a strategy that actually moves the needle.