

Bayesian MMM Project Documentation

Project Summary

A Bayesian Marketing Mix Modeling (MMM) pipeline was implemented using simulated daily media and revenue data. The model was designed to identify optimal marketing spend levels across four paid channels — Google Ads, Facebook Ads, TV Ads, and Influencer Marketing — by evaluating each channel's contribution to revenue while accounting for seasonal effects, competitor activity, and organic growth trends.

Tools & Libraries

- Python
 - NumPy / pandas
 - CmdStanPy (Stan for Bayesian regression)
 - Google BigQuery (data warehouse)
 - Google Looker (reporting layer)
-

Step-by-Step Pipeline

1. Data Generation

- Synthetic daily data was generated from 2022–2025.
- Each channel's spend followed a distinct trend or distribution to mimic real-world marketing behavior.
- Revenue was calculated as a function of each channel's spend, seasonal multipliers, competitor activity, organic growth, and injected noise.

2. Feature Engineering

- Constructed a daily dataset including features such as:
 - Google, Facebook, TV, and Influencer spend
 - Competitor activity (categorical)
 - Monthly seasonality index
 - Organic trend (upward over time)
 - Black Friday boost
- Target variable: total revenue

3. Bayesian Regression Modeling

- A simple Bayesian linear regression model was defined using Stan:
 - Prior: Normal(0, 1) for each coefficient
 - Output: posterior distributions for each feature's coefficient (β)
- Model compiled and sampled using CmdStanPy with 4 chains, 1000 samples each

4. Posterior Inference

- Extracted the posterior samples and mapped beta values to corresponding features
- Summarized posterior distributions using 95% confidence intervals
- Calculated channel-level coefficient means and standard deviations

5. Optimal Spend Recommendation

- A grid search was performed over all four paid channels to determine the best weekly spend allocation for April 2025:
 - Allowed spend range: 50–150% of current weekly average
 - Total April spend capped at 20% above current total

- Added noise to coefficients to avoid flat or identical recommendations
- Best plan selected based on expected revenue calculated using mean β values

6. Executive-Level Summary Table

- Summarized:
 - Channel
 - Current vs. Suggested Spend
 - % Change in Spend
 - Expected Revenue Impact
 - Confidence, Math Logic, and Human-readable Explanation
- Logic was based on the magnitude and sign of β and associated uncertainty

Example Recommendations:

- “The model sees room to grow here — with 77% confidence, it recommends investing \$2,470 more, expecting to generate \$1,625 in additional net revenue.”
- “This spend is returning only \$0.64 per \$1 — with 82% confidence, the model recommends reallocating \$1,950 to better-performing channels.”

7. Data Warehouse Integration

- Final outputs were pushed to Google BigQuery using the `pandas_gbq` package
 - Table 1: Daily generated dataset (`bayesianMMM_daily_data`)
 - Table 2: Final summary (`bayesianMMM_april_summary`)

8. Visualization in Looker

- The BigQuery tables were connected to Looker

- Charts were created to:
 - Compare current vs. suggested spend
 - Show estimated revenue lift
 - Display confidence per channel
 - Surface executive-readable recommendations
-