# Assessment 2: MMM Modeling with Mediation Assumption

## Context

You'll receive a **2-year weekly dataset** (above) containing paid media metrics, direct response levers (email/SMS), price, followers, promotions, and **revenue**.

## Task

Using this dataset, **build and document a machine learning model** that explains **Revenue** as a function of the input variables. You are free to choose the modeling approach (e.g., regularized regression, tree-based models, Bayesian regression, etc.), feature engineering, and validation strategy.
Keep a **causal perspective** in mind: assume **Google spend is a mediator** between **Facebook/TikTok/Snapchat** and **Revenue** (i.e., social/display can stimulate **search intent**, which influences Google spend, which then affects revenue).

You should treat this as a **product/measurement problem** rather than a purely academic fit exercise.

## What to deliver

- A **reproducible notebook or repo** (code + README) and a **short write-up** explaining:
    1. **Data preparation**: handling weekly seasonality, trend, zero-spend periods, and feature scaling/transformations.
    2. **Modeling approach**: the model you chose and why; hyperparameter choices; regularization/feature selection; validation plan.
    3. **Causal framing**: explicit treatment of the **mediator** assumption (e.g., structure your features, or use a two-stage approach; discuss back-door paths and leakage).
    4. **Diagnostics**: out-of-sample performance, stability checks (e.g., rolling/blocked CV to respect time), residual analysis, and sensitivity to Average Price and Promotions.
    5. **Insights & recommendations**: defend your interpretation of drivers of revenue; identify risks (e.g., collinearity and mediated effects).

## Evaluation criteria

1. **Technical Rigor**

    - Correct time-series CV (no look-ahead), robust preprocessing, handling of zeros and sparsity, well-reasoned hyperparameters.
2. **Causal Awareness**
    - Thoughtful handling of the **Google-as-mediator** assumption (e.g., staged models, mediation analysis proxies, or DAG-consistent feature design).
3. **Interpretability & Communication**
    - Clarity in how variables influence revenue (price elasticity, promo lift, email/SMS effects, diminishing returns if modeled).
    - Clear plots/tables; avoid vanity metrics.
4. **Product Thinking**
    - Are recommendations practical and defensible for a growth/marketing team?
    - Do you identify decision boundaries and trade-offs (e.g., price vs. demand, search vs. social)?
5. **Reproducibility & Craft**
    - Clean repo/notebook, environment instructions, deterministic results, and professional documentation.