

# **CHALLENGE REPORT FOR DATA SCIENCE POSITION**

2025

Farzam Nazari

# Table of Contents

- 
- How do you model spend carry over?
  - Explain your choice of prior inputs to the model?
  - How are your model results based on prior sampling vs. posterior sampling?

3

- 
- How good is your model performing?  
How you do measure it?
  - What are your main insights in terms of channel performance/ effects?

4

- 
- Can you derive ROI (return on investment) estimates per channel?  
What is the best channel in terms of ROI?
- 

5

- **How do you model spend carry over?**

The spend carry over is modeled using an adstock transformation. This transformation captures the lagged effect of marketing spend over time, represented by an adstock rate that determines how quickly the impact of past spend decays.

- **Explain your choice of prior inputs to the model?**

The prior inputs were chosen based on domain knowledge and empirical evidence. Normal priors were used for the regression coefficients, assuming that the effects of the channels are centered around zero with some variance. This choice helps regularize the model and incorporate prior beliefs about the parameters.

- **How are your model results based on prior sampling vs. posterior sampling?**

Prior sampling involves drawing samples from the prior distributions before observing the data, which reflects initial assumptions. Posterior sampling, however, updates these priors with observed data to generate posterior distributions. The posterior sampling results are more informed as they incorporate the actual data, leading to more accurate parameter estimates.

- **How good is your model performing? How you do measure it?**

The model's performance is evaluated using MAE, MSE, and RMSE. These metrics indicate that while the model does not have a significant bias (mean residual close to zero), it has substantial variability in its predictions (high standard deviation of residuals and RMSE). Although the dataset should ideally be split into training and test sets for proper evaluation, due to time constraints and other issues, the model was evaluated using the entire dataset (base x and y). Comparing these metrics to a baseline model, such as predicting the mean revenue, can help assess if the model provides additional value.

If we split the data with a larger dataset, it can be optimized further by spending more time testing other models. Initially, I used a linear model for cross-validation, but it took a significantly longer time due to the large errors produced by the linear model. Therefore, I switched to the current model to achieve better performance.

- **What are your main insights in terms of channel performance/effects?**

The main insights from the model indicate which marketing channels are more effective in driving revenue. By analyzing the coefficients of the channels, we can determine which channels have a positive impact and which do not. Channels with higher positive coefficients are more effective, while those with negative or smaller coefficients might be less effective or indicate overspend.

- **Can you derive ROI (return on investment) estimates per channel? What is the best channel in terms of ROI?**

Based on the ROI calculations, the channel with the highest ROI is identified as the best channel. From the visualizations, it is observed that Channel 2 not only has the highest ROI but also shows higher median revenue compared to the other channels. This makes Channel 2 the most effective channel in terms of return on investment and revenue generation.