Applicant: Emre Gürsoy

Dear Dr. Alwin Haensel,

Thank you for the interesting data science challenge, I truly enjoyed experimenting with PyMC. Please note that my results represent a first-order approximation. The current model is overfitting and would require further refinement for production use. That said, I believe the results demonstrate a solid understanding of the core concepts. I hope the results are sufficient to move on to the next stage of the process. My salary expectation is 65,000 € gross annually, and I am available to start on May 1st, 2025.

Question 1:

"How do you model spend carryover?"

Response: Implemented a geometric adstock with normalized retention weights.

$$\text{adstock}_t = \sum_{l=0}^{L-1} w_l \cdot \text{spend}_{t-l}, \quad w_l = \frac{\alpha^l}{\sum_{k=0}^{L-1} \alpha^k}$$

Here, $\alpha \sim \text{Beta}(2, 2)$ controls the retention rate. I chose this Beta prior as it is centered at 0.5 and avoids extreme cases such as 0 or 1. I selected $L = 8$, assuming that carryover effects become negligible after approximately two months.

Question 2:

"Explain your choice of prior inputs to the model."

Response: Priors were selected to balance model flexibility with business constraints.

Base revenue: Normal distribution centered around historical revenue with some flexibility.

Channel effectiveness: Half-normal distribution to ensure media impacts remain non-negative.

Adstock retention: Beta(2, 2), centered at 0.5 and away from extremes (0 or 1).

Trend slope: Normal distribution allowing both growth and decline.

Dual seasonality: I included both annual and semi-annual components to capture patterns at multiple timescales.

Question 3:

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

Response: The model parameters showed clear convergence after fitting.

Initially, prior samples for channel effectiveness were wide and uncertain. After running the model with data, the posterior distributions narrowed significantly, reflecting updated and more confident estimates.

Question 4:

"How good is your model performing? How do you measure it?"

Response: The model works reasonably well, though it's not perfect.

To evaluate performance, I used MAPE (mean absolute percentage error), which was around 20%, which is an acceptable considering the model being a first-order approximation. The R^2 value was 0.43, meaning the model explains less than half of the variance in revenue, suggesting that while some key drivers are captured, others are missing.

Question 5:

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

Response: The channel effects show large variation.

The results reveal one superstar (channel 2 at 32.46), two dead channels (1 and 5 near zero), and four moderate performers. Channel 2's 300x boost from its prior seems unrealistic, either it's truly revolutionary, which I highly doubt, or the model is overfitting its small spend. The near-zero channels (channels 1 and 5) should likely be cut, while the others (channels 3, 4, 6, 7) deliver stable but modest returns.

Question 6 (Bonus):

"Can you derive ROI estimates per channel?"

Response: The ROI estimates also reflect large variation.

The ROI analysis shows extreme differences in performance. Channel 2 delivers a return of 9.45%, orders of magnitude higher than the others, surely overfitting. The remaining channels show negligible ROI (0.002-0.014%), with channels 1 and 5 effectively dead (0% return).