Data Scientist MMM challenge results

Joao Domingues Semeano

12 April 2023

This report presents the answers to the questions made in the Data Scientist MMM challenge. For further comments on the model development and findings, please refer to the accompanying Jupyter notebook.

1 - Carryover modelling:

Carryover, or adstock, is modelled with the geometric function proposed in the paper by Jin and Wang, et. al:

$$adstock(x_{t,m}; \alpha_m^l, L) = \frac{\sum_{l=0}^{L-1} \alpha_m^l x_{t-l,m}}{\sum_{l=0}^{L-1} \alpha_m^l}$$

where $x_{t,m}$ is the channel spend at time t for channel m and α is the geometric decay parameter. At each point, the adstock is the average sum of the geometrically decaying channel spending. This assumes that the maximum effect of any channel is contemporaneous with the spending. The effects of different α and L are visible below:

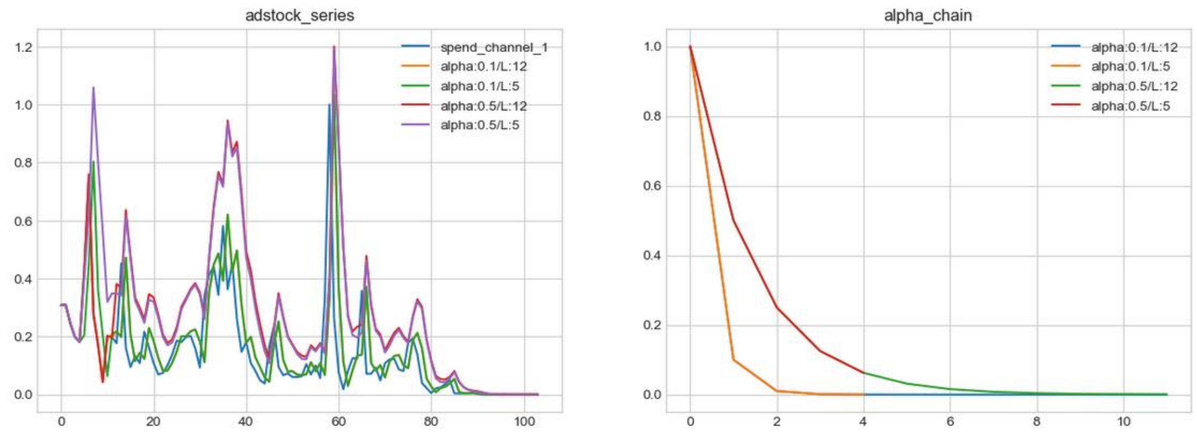


Fig. 1 – adstock effects

2 – Prior choices:

Choosing the priors for a model relates to the prior information about the parameters underlying the process generating the real data. The table below shows the prior choices for each parameter:

Parameter	Prior	Comment
base (value at time 0)	Exponential(λ = 0.0001)	We can see from the charts that the value at 0 is positive
b_trend (time trend)	Normal(μ = 0, σ = 2)	The trend is most likely negative, but with a normal distribution it can take on any value. σ is left large to allow for different values
b_fourier (seasonality)	Normal(μ = 0, σ = 2)	The seasonality terms may be positive or negative. σ is left large to allow for different values

Parameter	Prior	Comment
coef_ (channel 1 to 7)	Exponential(λ = 0.001)	Assuming that each channel affects the revenue positively, an exponential prior was chosen. λ is left larger than in the base to allow for more variation
alpha_ (channel 1 to 7)	Beta(σ = 2, β = 2)	It's important that the alpha of the geometric decay in adstock is strictly between 0 and 1. So a Beta distribution was chosen. The initial parameters place the peak of the distribution at 0.5 which is an agnostic approach
Noise	Exponential(λ = 1)	The noise around the revenue must be positive. Since the magnitude is unknown, λ is set to 1

Tab. 1 – prior choices and reasoning

Several other priors were tried, for example allowing the channel coefficients to be negative (normal distribution), regularizing the seasonality contributions with Laplace distributions and trying different initial parameters. No material difference was found in model performance.

3 – Prior vs Posterior sampling:

The sampling method used is NUTS (the default in PyMC). This is a MCMC algorithm and works by drawing samples from the posterior distribution given some data, iteratively updating the prior. Therefore, we are mainly using posterior sampling.

4 – Model performance:

There are several ways to check model performance. In the current case, a check of whether the data falls within 2 standard deviations of the path obtained through the NUTS sampling and a comparison between the implied posterior distribution and the data posterior distribution are performed:

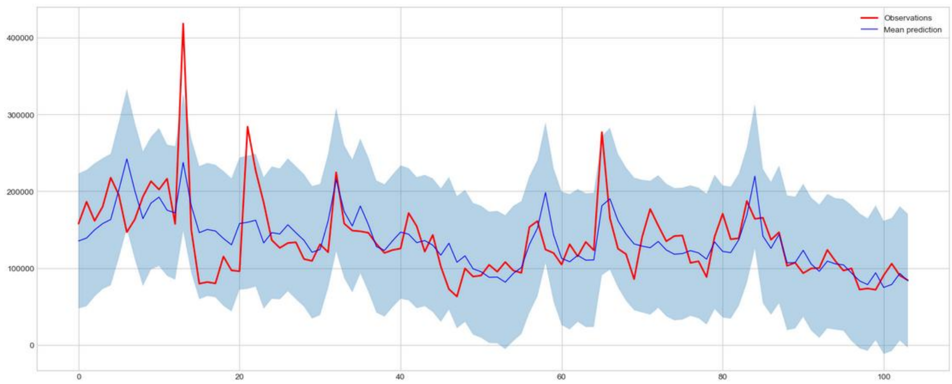


Fig. 2 – path implied by posterior distribution and real data path

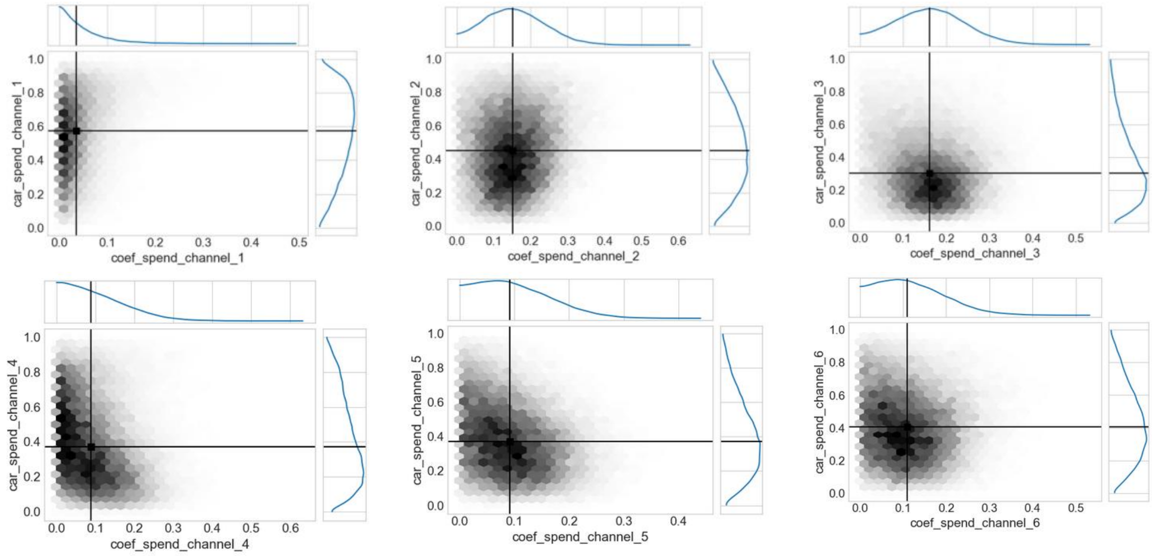
Fig. 3 – posterior distribution implied by the model and implied by the real data

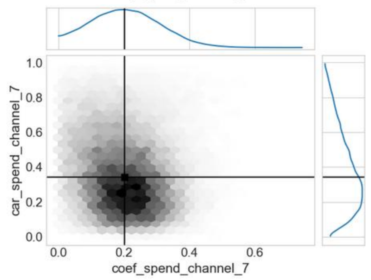
On the one hand, most of the real path does fall within 2 standard deviations of the path implied by the posterior distribution, with only extreme spikes laying outside. The comparison of the posterior distributions does show that the model overestimates revenues and their dispersion.

Another method, especially useful if the model is to be used for forecasting, is to divide the data into training and test data. Then we can fit the model in the training data and apply it in the testing set. We can then check for overfitting.

5 – Channel performance:

Channel performance depends on the coefficient of the channel (how much it affects the revenue) and the adstock alpha parameter (how quickly the effect tapers off). A good channel would then have both a large effect and a large adstock alpha. The charts below plot the posteriors for both coefficients and alphas for each of the 7 channels:





Comments on the channels:
Channel 1: largest permanence through adstock, lowest coefficient
Channel 2: very high permanence, second highest coefficient
Channel 3: lowest permanence, third highest coefficient
Channel 4: low permanence, second lowed coefficient
Channel 5: similar to channel 4, but with a better coefficient distribution
Channel 6: good permanence, but relatively small coefficient
Channel 7: highest coefficient, but rather low permanence

Fig. 4 – posterior distribution for the channel coefficients and the adstock alpha

Referring to Fig.4, we can point out that channels 4, 5 and 6 have similar performances, with channel 6 being the best of these 3, due to an alpha distribution tilted to higher values. Channels 3 and 7 have the best coefficients but low permanence through adstock. We could consider increasing the investment here. Channel 2 seems to work well due to a large coefficient but already has the largest permanence, perhaps there's no need to invest more. Finally, channel 1 seems to have very low impact and a very large permanence, we should either consider it underperforming or invest in it decisively but sparsely.

6 – Return on Investment per channel:

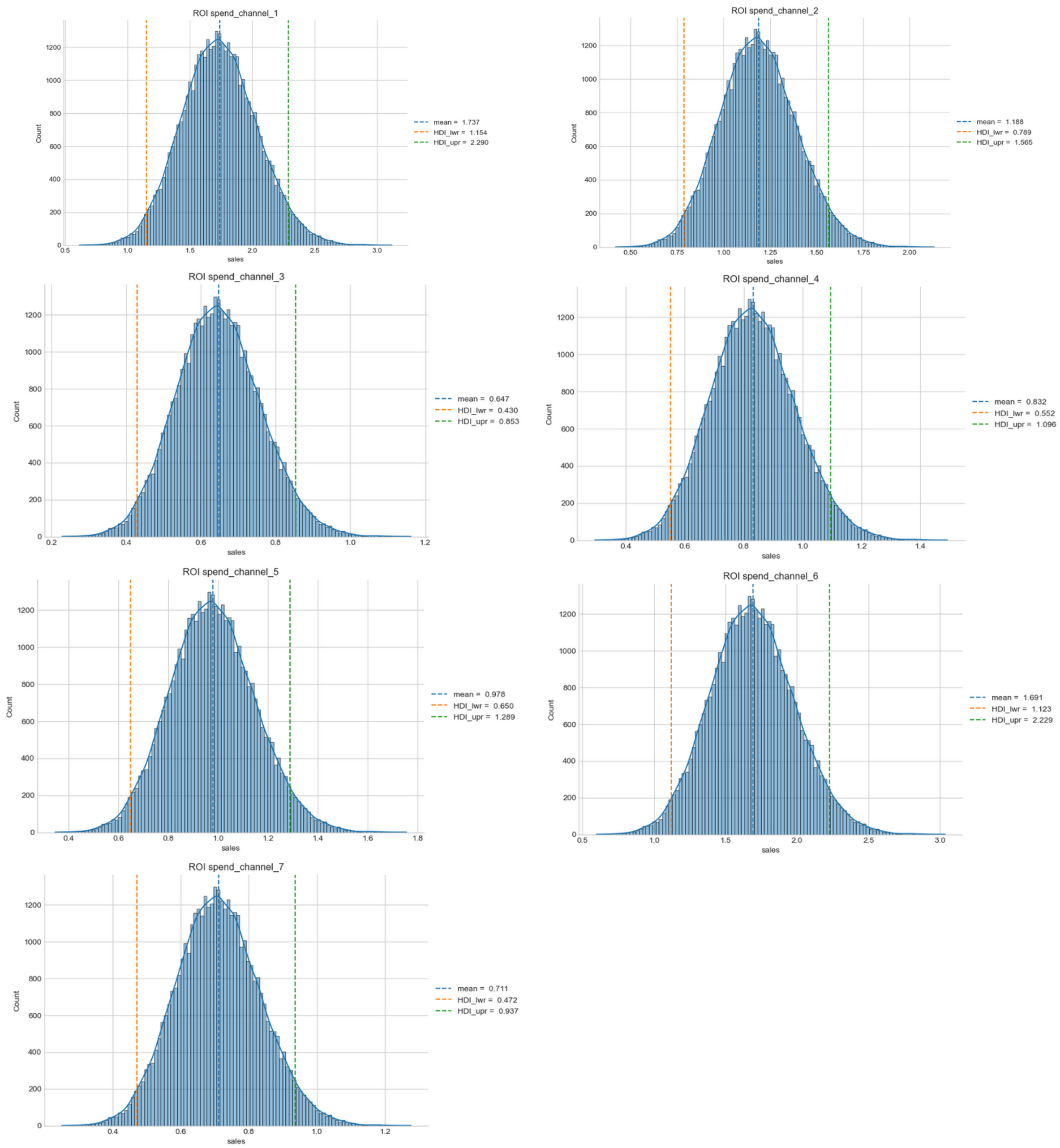


Fig. 5 – Return On Investment for the 7 channels

The Return On Investment estimates are all above zero, confirming the assumption that all publicity is good publicity.

Interestingly enough, some of the estimates contradict some of the assertions made before on this very section.

Ordered by mean ROI, the channels from bottom to top performer are: 3, 7, 4, 5, 2, 6, 1!

Apparently, the return on investment for channel 1 is the best (more than twice the worst channel). The adstock permanence seems to be more relevant than the coefficient itself!