**Habu Project 1: Optimal marketing mix and budget optimization**

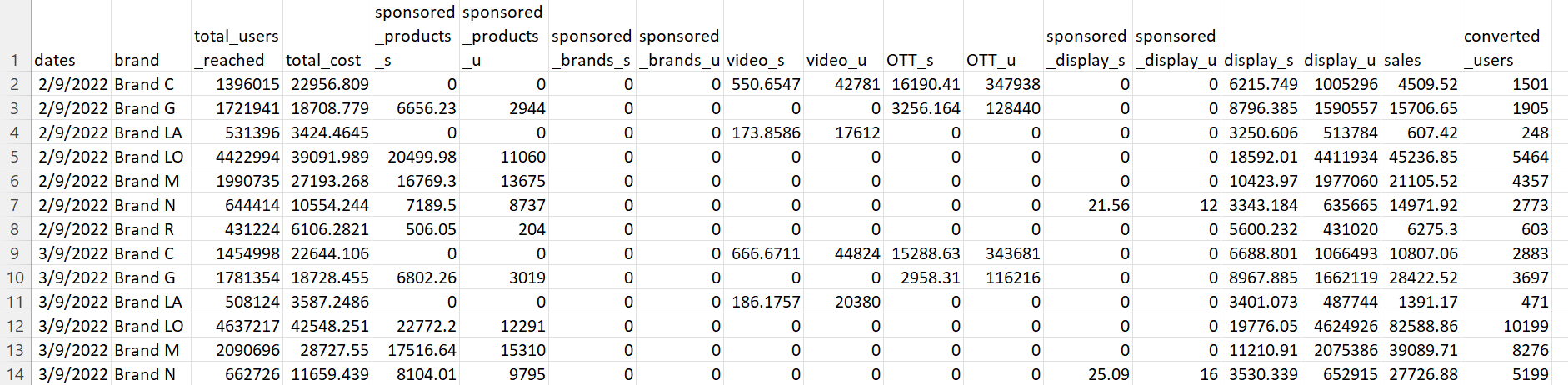
**Project Progress Report**

**Oct 26, 2023**

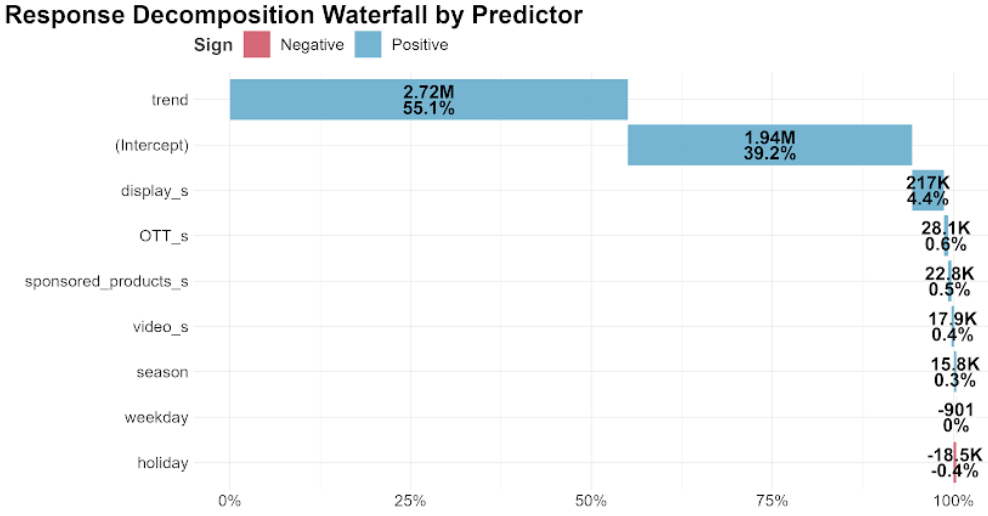
1. **Project Overview**

This project aims to optimize advertising strategies for businesses in order to maximize their impact and return on investment (ROI) in the evolving landscape of digital marketing. Our project involves the analysis of aggregate data sourced from Amazon Marketing Cloud, encompassing daily data from seven brands spanning a duration of 12.5 months (from early September 2022 to mid-September 2023). This includes data on sales and converted users, as well as the spend and user reach of six media channels (namely display, sponsored products, sponsored brands, sponsored display, videos, and over-the-top (OTT)) for each brand. By leveraging this comprehensive dataset, we aim to delve deeper into understanding advertisements' true effectiveness at various granularity levels and develop a budget optimization methodology for the strategic allocation of advertising budgets. Ultimately, the project aims to reveal actionable insights, enhance ROI, and identify optimal channel and tactic combinations for advertising success.

1. **Project Progress**
2. Completed work
   1. Understand the business context and objectives of the project
   2. Draft the Project Proposal with a detailed project plan
   3. Create a data frame using the two datasets provided by Habu (conversion.csv and media.csv), and merge data based on brands and dates.
      1. The merged data frame’s columns include dates, brands, total\_users\_reached, total\_cost, display\_s, sponsored\_products\_s, sponsored\_brands\_s, video\_s, OTT\_s, sponsored\_display\_s, display\_u, sponsored\_products\_u, sponsored\_brands\_u, video\_u, OTT\_u, sponsored\_display\_u (where the “s” and “u” suffix represents the spend and user reach of each media channel respectively), sales, and converted\_users.
      2. The sample data frame is appended below -

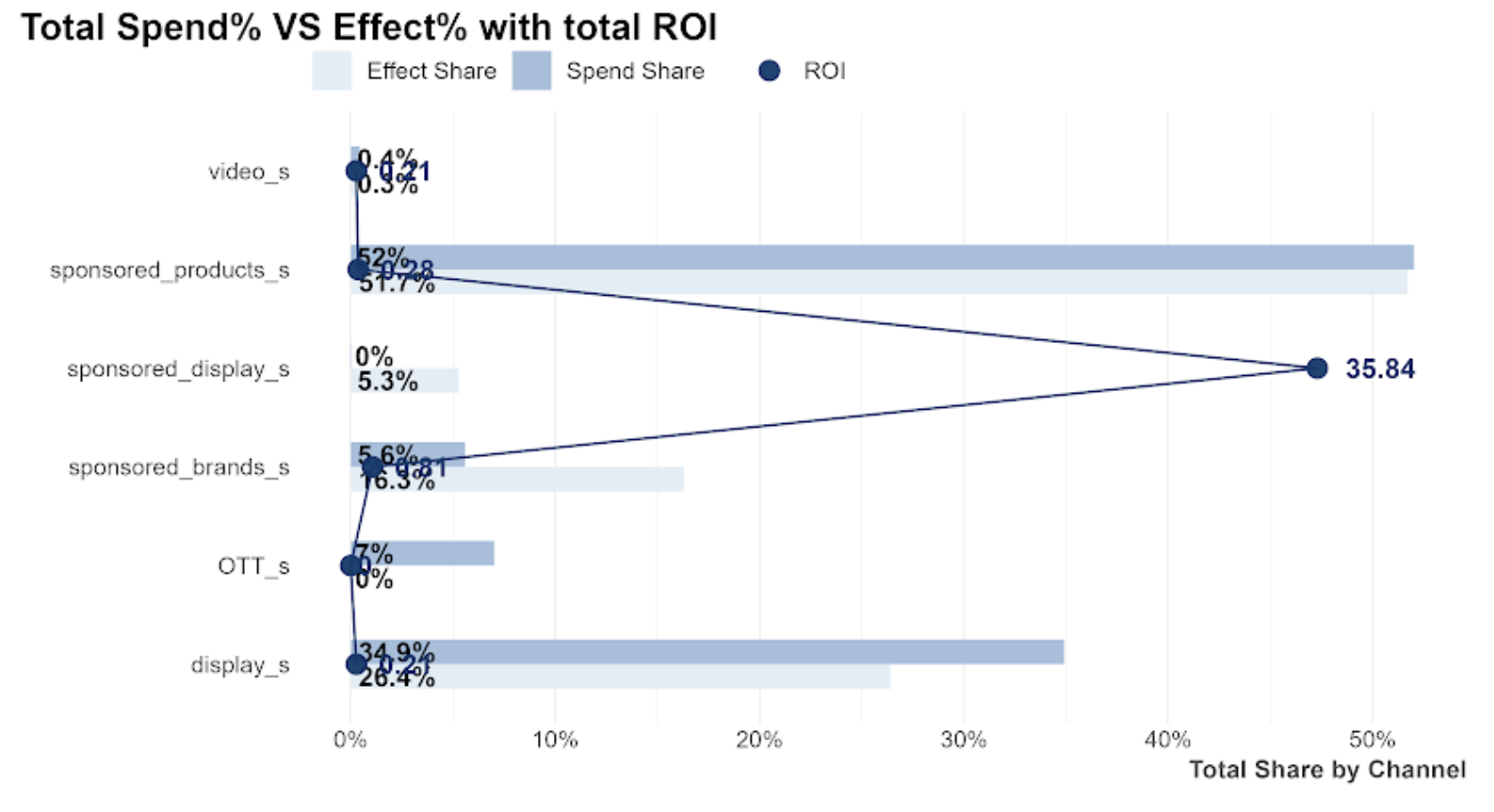


* 1. Build a Marketing Mix Model with Robyn[[1]](#footnote-0) (an experimental, AI/ML-powered and open-sourced MMM package from Meta Marketing Science) in R
     1. About Robyn:
* The function robyn\_inputs() mainly captures all model specifications for the dataset. The model allows inclusion of either revenue or conversion as the dependent variable, paid media variables (i.e. the various media channels), organic variables (i.e. marketing activities without direct spend), contextual variables (e.g. competitor sales or events), variables controlling for trends and seasonality, and hyperparameters for adstock and saturation transformations.
* Robyn carries out a multi-objective hyperparameter optimization with Nevergrad, which mutates the hyperparameter values based on which values have better scores (objective functions)
* Robyn implements three objective functions as the “goals” for hyperparameter optimization:
  + - * Normalized Root Mean Square Error (NRMSE), referred to as the prediction error
      * Decomposition Root Sum of Square Distance (DECOMP.RSSD), referred to as the business error, representing the difference between the share of spend and share of effect for paid media variables
      * Mean Absolute Percentage Error (MAPE.LIFT), referred to as calibration error (not applicable in our model because calibration is not carried out)
* A time series validation is carried out by splitting the data 3-way for training, validation and testing
* After running the model trials, k-means clustering is used to further reduce the model choice. Robyn then automatically exports the model outputs for each pareto-optimal model as a onepager
* Robyn offers a budget allocator for a selected model with two scenarios -
  + - * “Maximum response” calculates the optimum cross-media budget split by maximizing the response, given a total budget and media-level constraints
      * “Target efficiency” calculates the optimum cross-media budget split and the total budget by maximizing the response, given a target return on advertising spend (ROAS) or cost per action (CPA)
  1. Replicate Robyn's demo with their simulated data to understand the model better
  2. Run the model on each of the brands and for all the brands combined
     1. For paid media variables, we used the spending of each media channel
     2. To account for carry-over effects of advertising, we conducted geometric adstock transformation
* media\_adstocked\_ij = media\_raw\_ij + decay\_rate\_j \* media\_raw\_i-1\_j, where i is a given time period and j depicts a media variable. The decay\_rate\_j is a constant value per media variable and equal to the Geometric parameter theta
  + 1. To account for diminishing returns of advertising (the more is spent on a channel, the less marginal return), we conducted a saturation transformation
* media\_saturated\_j = 1 / (1 + (gamma\_j / media\_adstocked\_j) ^ alpha\_j), where j depicts a media variable
  + 1. Meta’s time-series forecast package Prophet is used to account for trend, season, holiday and weekday
    2. Build and select the best performing model
    3. Apply budget allocator to the model to obtain optimized budget allocation for each brand and all the brands combined under three scenarios -
* Maximum response given the original budget of the brand
* Maximum response given an arbitrary budget
* Target ROAS

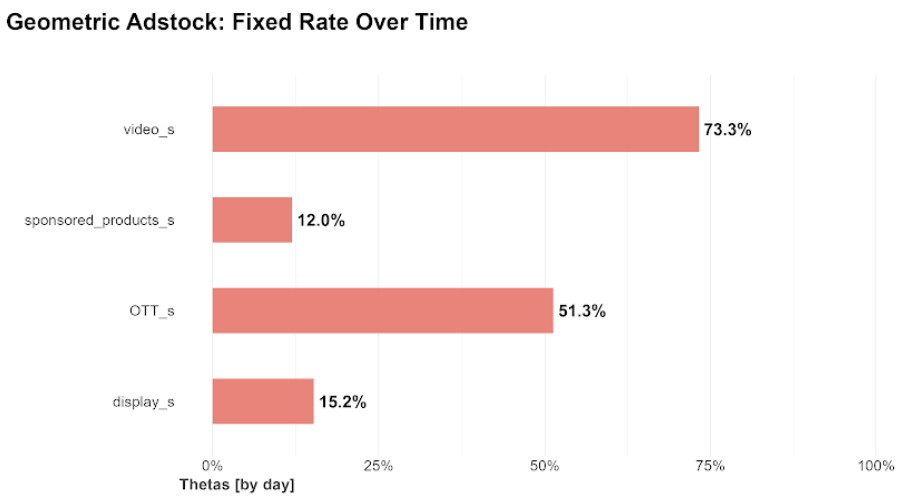
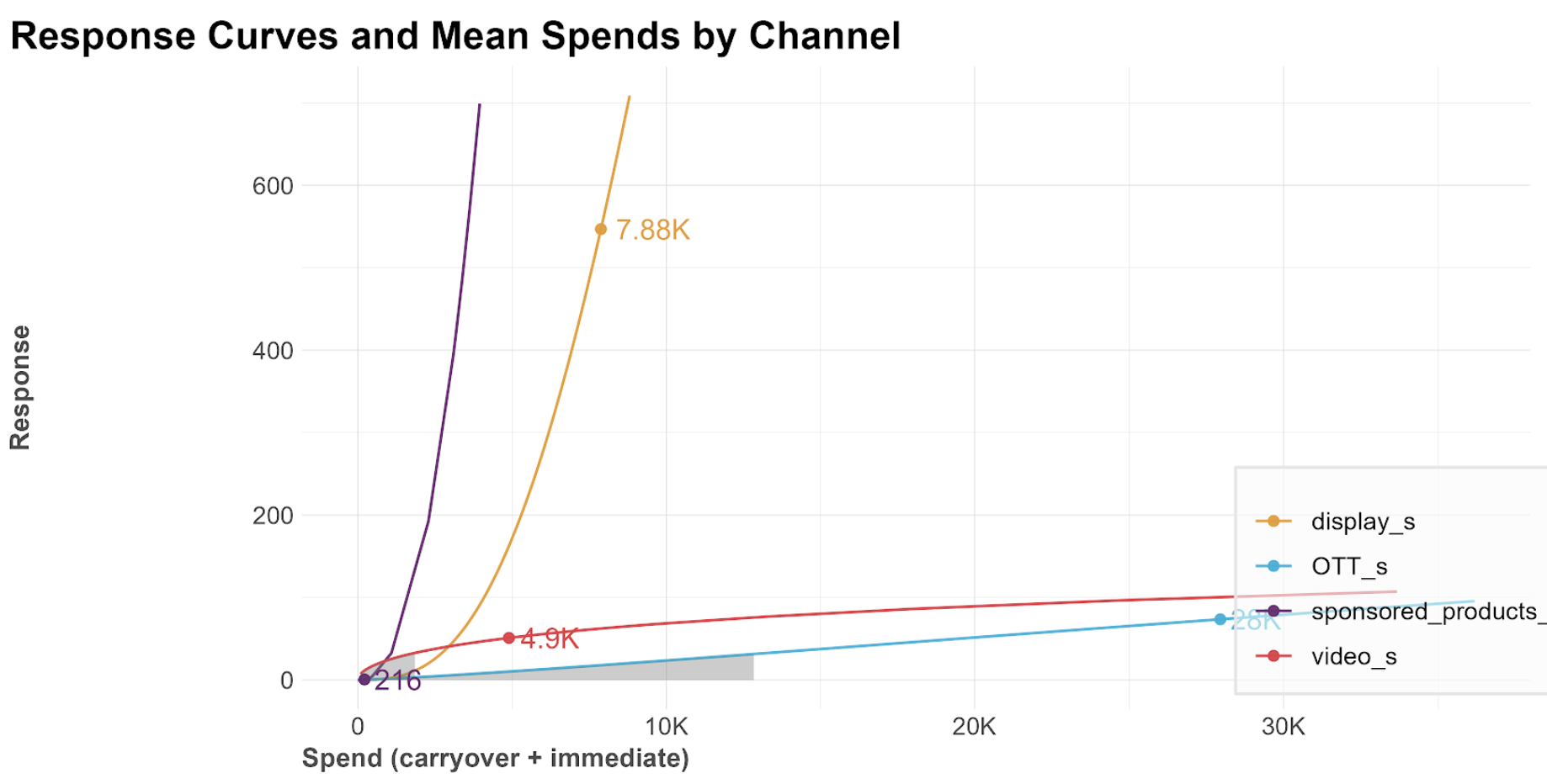
1. Current and potential blockers
   1. NRMSE and DECOMP.RSSD are not converging (as they should) for some of the brands
   2. The adjusted R-squared values for our models do not meet our expectations
   3. Uncertain about the selection of hyperparameters
   4. Budget optimization unable to hit target ROAS for most of the brands (budget allocator 3)
2. **Preliminary Findings for Further Investigation**
   1. **ROI**: ROIs for some channels, such as sponsored products, are generally better than others.
      1. For Brand N, sponsored\_display's ROI reaches 35.84.
      2. Sponsored products generally have a better ROI than others.
   2. **ROAS**:
      1. After budget allocation, ROAS does not meet our expectations since most of the ROAS are less than 1. However, for some brands, such as Brand M and Brand LO, when the budget allocation among channels is adjusted drastically, skewing towards sponsored\_products, the ROASs become better. Display and sponor\_products have 0% spending but an extremely large ROAS.
      2. The channel having the largest proportion of spend and response usually does not have an expected ROAS, and vice versa.
   3. **R-squared**: Some of our models have a train R squared < 0.8, which is not ideal so we need to improve our model. (Brand C, Brand LA, Brand LO, and Brand M have an R squared value of around 0.76; Brand G and N have an R-squared = 0.85)
   4. **Indication of ROI with contribution:** Based on Brand C, sponsored\_products has the highest ROI of 0.28 and a low contribution, which suggests a high potential increase in spend, given that it is delivering good returns and is likely to not be saturated due to the low spends. However, the budget reallocation is 0% for sponsored\_products. Also, low ROI and high spend means underperforming and hence spend should be moved away from it. The potential increase in spend can also be observed from the Total Spend % VS Effect %. 

|  |
| --- |

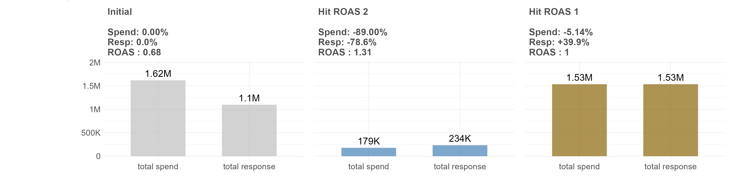
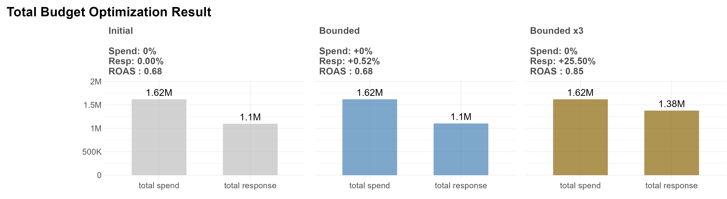
* 1. **Spend share vs. Effect share**: Based onBrand N, sponsored\_display has a significantly greater ROI than other channels, and the reason for this might need more investigation; sponsored\_brands channel has a larger effect share compared to the corresponding spend share while the display channel has a larger spend share compared to corresponding effect share. Based on that, we can consider reallocating the budget spending for sponsored\_brands and display by increasing the allocation for sponsored\_brands and decreasing the allocation for display.



* 1. From the **Response curve** (saturation curve), we can find whether the channel’s spend is at an optimal level or if it is approaching saturation. The faster the curves reach an inflection point and a horizontal/flat slope, the quicker the media channel will saturate with each extra ($) spent. For example, in Brand C, video and OTT almost reach their saturation point, but the model still allocates a large percentage of spend to these two channels causing a lower response rate.



* 1. From the **Geometric Adstock: Fixed Rate Over Time**, we can find the decay rate for each channel. A high decay rate means the longer the effect that a specific media channel has after the initial exposure.
  2. Compared to Oct 2022, for all brands except for Brand LA, there is a slight drop in **actual vs. predicted response** in Oct 2023 due to poor performance.
  3. **Ad spending should be much lower** to make a profit**.** Expected return rarely recovers the initial budget spent. Money invested in advertising on Amazon is usually around $1M per month but the best ROAS that brands can reach is generally attained when the initial budget is smaller, i.e. around $200k. In the picture below, the gold graph shows the best ROAS that can be expected with current spend levels. The blue graph shows the best ROAS achieved by the brand if it were to change its original spend. Two main findings can be inferred:
     1. Brands largely underestimate the effect of saturation on their ad spend and should consequently drastically reduce their ad spend levels.
     2. Our model is not capturing the full relationship between spend and sales and should incorporate additional data to uncover hidden links and predict new budget allocations accurately.



1. **Remaining work to be done**
2. Increase the model’s accuracy
   1. Ask Habu for more data to enhance the prediction of our model
      1. include organic\_vars and context\_vars
   2. Try to run the model on conversion instead of sales
3. Build a clear and simple model allowing replicability and ease of explanation to clients
   1. Take a closer look at Robyn’s source code
   2. Adapting Robyn’ source code with our own findings a simpler model
   3. Summarizing all outputs to create one simple new one
4. Present deliverables in a report

1. https://facebookexperimental.github.io/Robyn/ [↑](#footnote-ref-0)