Mixed marketing modelling using Robyn

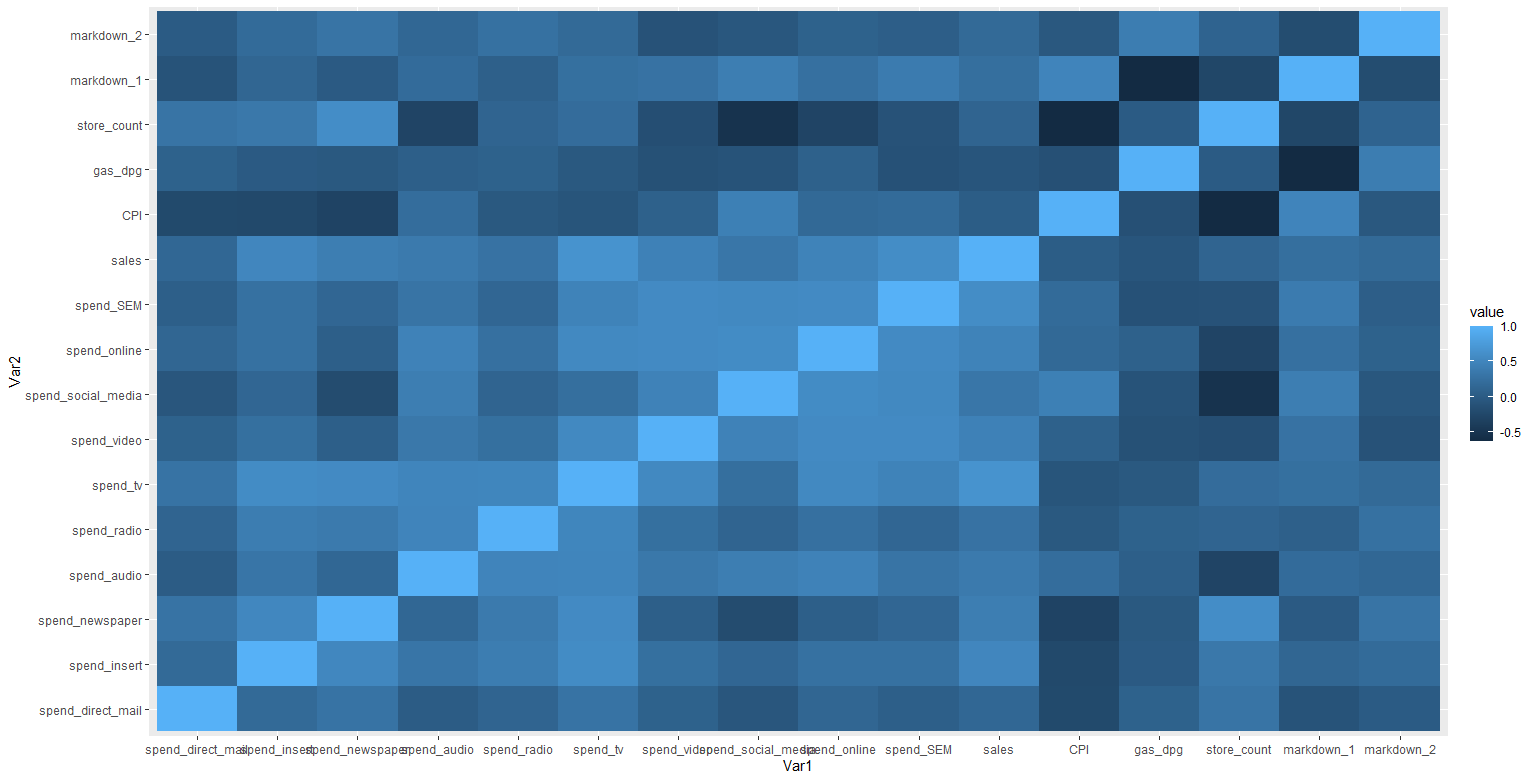
*About the dataset*

Anonymised dataset originally provided by University with sales data, a number of macro variables and 10 different media spend channels. For mixed marketing modelling, we split the dataset as follows:

* Dependent variable – sales
* Date variable – wk\_strt\_dt (week start date)
* Core variables – CPI, gas\_dpg, store\_count, markdown\_1 and markdown\_2
* Media spend variables (10 different channels) – newspaper, audio, social media etc

*Data pre-processing*

Key issues with mixed marketing modelling tend to be issues with the dataset (primarily around missing data) and multicollinearity. Both these in the code have been investigated with no missing values and ggplot showing limited evidence of multicollinearity.



Multicollinearity is somewhat neutralised by the fact that Robyn uses ridge regression.

*Initiate robyn model*

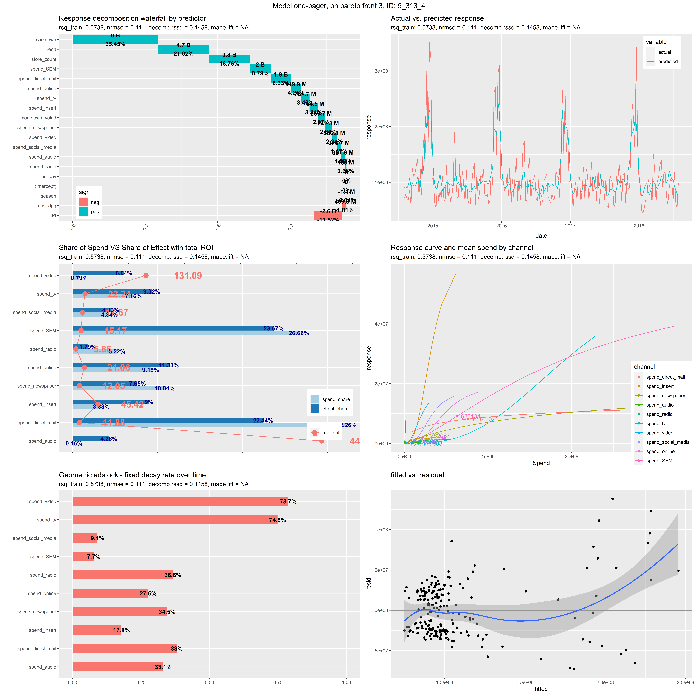
Initial inputs are collected. We also initialise categorical variables related to seasonality, public holidays and trend term using prophet\_vars.

The adstock model used in this example is geometric, with decay controlled by theta. This is initialised with the rule of thumb: TV c(0.3, 0.8), OOH/Print/Radio c(0.1, 0.4), digital c(0, 0.3). These ranges provided by the user are subsequently optimised when the model is actually initialised.

*Selecting appropriate model*

Robyn provides a range of solutions with the user selecting the best model. This is best by analysing the fitted vs residual chart (bottom left). This is saved.

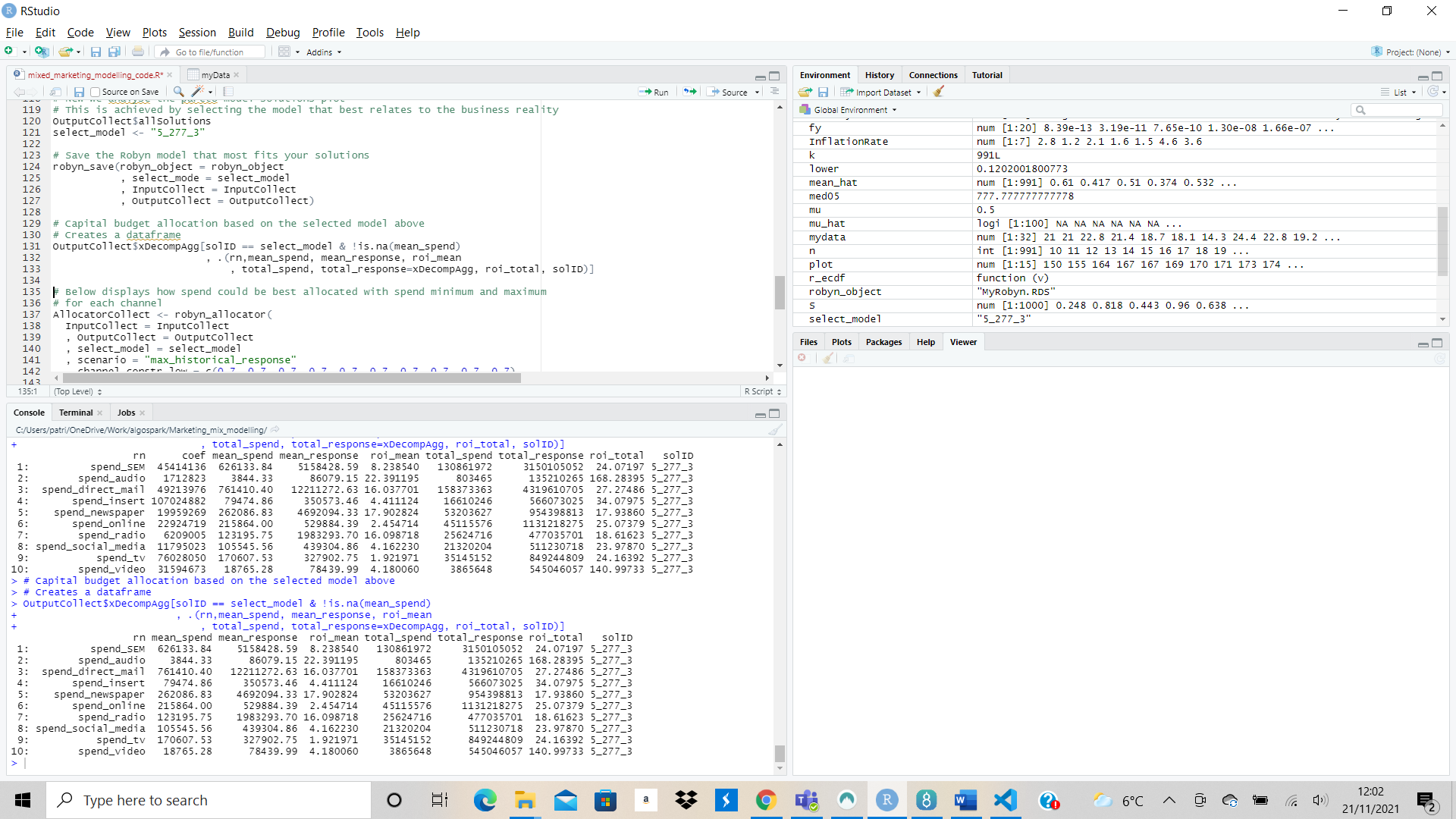
*N.B. A more official solution would need more refined analysis of decay rates for the adstock model with some of the ranges for some variables likely too large.*



*Example output for one graph*

The output also produces 4 CSV files (pareto\_hyperparameters.csv, pareto\_aggregated.csv, pareto\_media\_transform\_matrix.csv, pareto\_alldecomp\_matrix.csv), which contains all the results. These can be read in separately and analysed if appropriate.

However, for the purposes of this analysis we will just focus on the ROI from different channels.



*Future areas of development*

While this represents a suitable starting point, some of the channels that received low spend see a particularly high roi (e.g. spend\_audio). This is likely a reflection of diminishing returns which holds that each additional unit of advertising increases the response, but at a declining rate. This will be mitigated by a better understanding of the saturation curve by better tuning the hyperparameters.