# Altman Vilandrie & Co. data challenge

Tools used:

* R-studio (Version 0.98.1091)
* Tableau (Version 9.0)

Methods:

The following steps were taken to complete the challenge:

* Data was imported into R-Studio from the csv files provided
* Sales data was converted to a time series (frequency 52; indicating weekly data) and then was decomposed into a trend, a seasonal and an irregular component.
* Sales data was then plotted without the seasonal component
* The data was normalized for ad-stock effect commonly seen in marketing related metrics (effect of a marketing strategy decreases overtime but still contributes to sales). A ad-stock coefficient of 0.15 (15%) was chosen.
* A multiple regression model was fitted on the data
* The model was checked for multicollinearity using variance inflation factors (VIF) analysis. Low VIF values (<5) were observed.
* The relative importance of the different independent variables was calculated using the package (“relaimpo”) in R. The basis for this ranking is based on the amount of contribution that the variable provides to the decomposed R2 of the model. I used the less computationally intensive metric “car” (the R2 decomposition according to Zuber and Strimmer 2010).
* The relative contribution of each variable was calculated. This was multiplied by the total sales (as the change in sales, before and after the application of marketing strategy is not known!), to get an estimate of the variables contribution to sales.
* The cost/attribute sale was defined as:

***Cost per attributed sale=(relative contribution of each media strategy/spending on the media strategy)***

* Finally, the Return on Investment (ROI) ratio was calculated by:

***Return on investment =((relative contribution of each media strategy - spending on the media strategy) / spending on the media strategy)***

* “LinPro” package in R was used to solve a linear equation to get optimization results (however, the lack of specific constraints resulted in the program suggesting all the money be put in Media 9 strategy, the strategy with the highest ROI).
* The results were plotted onto a tableau dashboard and published on a webpage.

Inferences:

* A strong YEARLY seasonal trend was observed in the data. The trend was replicated in both the years. The data was however not normalized to seasonal trends as the media outlets might be serving season specific advertisements.
* Adstock normalization was performed to account for the memory effect that of advertisements.
* The linear model created did not show multicollinearity and had a high R2 rate.
* Forward variable selection was also considered to isolate important predictors of sales, however, finally the relative importance package (based on decomposing R2) was preferred.
* The highest sales were caused by media strategies: Media 9, Media 14, Media 4
* The highest Return on Investment however was: Media 9, Media 11 and Media 14
* The structural variables with highest impact on sales were: Structural variables 1, 5 and 2
* The optimal investment strategy (given no specific constraints like the minimum investment required in all media strategies would be invest everything in Media 9, as it is the most profitable ROI)