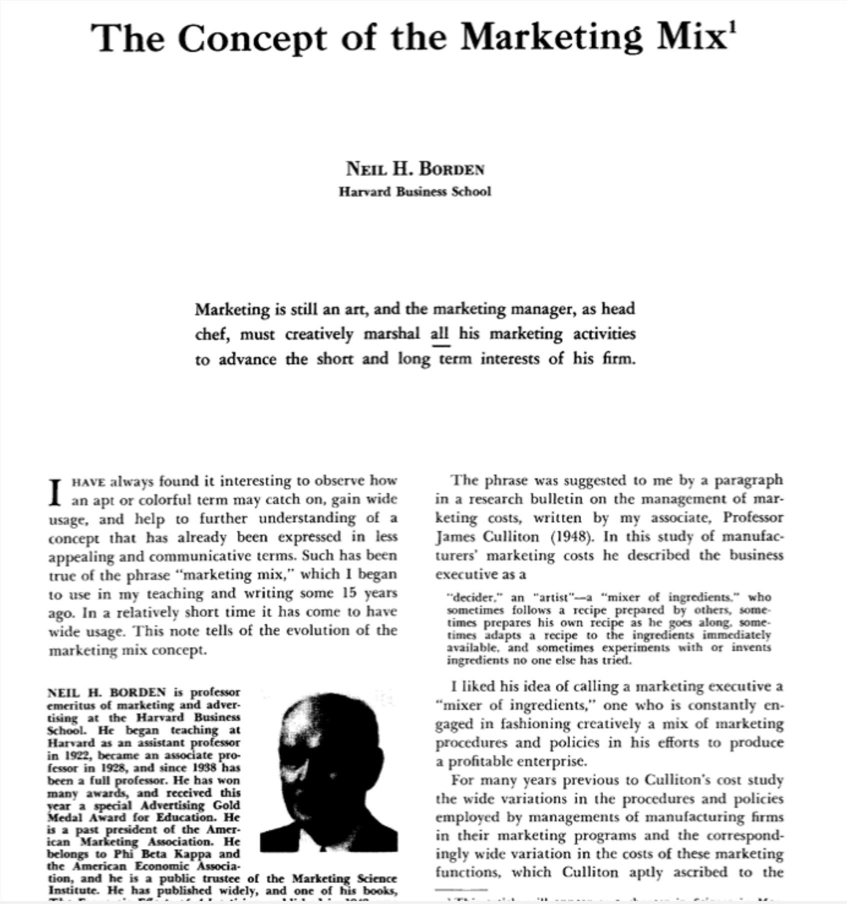
ArŶmα Labs

Decoding the past, Encoding the future

A – Z of MMM Workshop

Introduction to Marketing Mix Modeling (MMM)



The Concept of the Marketing Mix[1]

NEIL H. BORDEN
Harvard Business School

Marketing is still an art, and the marketing manager, as head chef, must creatively marshal all his marketing activities to advance the short and long term interests of his firm.

I HAVE always found it interesting to observe how an apt or colorful term may catch on, gain wide usage, and help to further understanding of a concept that has already been expressed in less appealing and communicative terms. Such has been true of the phrase "marketing mix," which I began to use in my teaching and writing some 15 years ago. In a relatively short time it has come to have wide usage. This note tells of the evolution of the marketing mix concept.

NEIL H. BORDEN is professor emeritus of marketing and advertising at the Harvard Business School. He began teaching at Harvard as an assistant professor in 1922, became an associate professor in 1928, and since 1938 has been a full professor. He has won many awards, and received this year a special Advertising Gold Medal Award for Education. He is a past president of the American Marketing Association. He belongs to Phi Beta Kappa and the American Economic Association, and he is a public trustee of the Marketing Science Institute. He has published widely, and one of his books,

The phrase was suggested to me by a paragraph in a research bulletin on the management of marketing costs, written by my associate, Professor James Culliton (1948). In this study of manufacturers' marketing costs he described the business executive as a

> "decider," an "artist"—a "mixer of ingredients," who sometimes follows a recipe prepared by others, sometimes prepares his own recipe as he goes along, sometimes adapts a recipe to the ingredients immediately available, and sometimes experiments with or invents ingredients no one else has tried.

I liked his idea of calling a marketing executive a "mixer of ingredients," one who is constantly engaged in fashioning creatively a mix of marketing procedures and policies in his efforts to produce a profitable enterprise.

For many years previous to Culliton's cost study the wide variations in the procedures and policies employed by managements of manufacturing firms in their marketing programs and the correspondingly wide variation in the costs of these marketing functions, which Culliton aptly ascribed to the

"In all the above illustrative situations it should be recognized that advertising is not an operating method to be considered as something apart, as something whose profit value is to be judged alone.

An able management does not ask, "Shall we use or not use advertising, without consideration of the product and of other management procedures to be employed.

Rather the question is always one of finding a management formula giving advertising its due place in the combination of manufacturing methods, product form, pricing, promotion and selling methods, and distribution methods.

As previously pointed out different formulae, i.e., different combinations of methods, may be profitably employed by competing manufacturers." - **Neil H Borden**

What is MMM? Why is it such a powerful tool?

ArŶma Labs
Decoding the past, Encoding the future

Business Metric*

- Brand Equity
- Sales Volume
- Sales Value
- Market Share
- Brand Penetration
- Orders/Quotes
- TOMA
- Spont. Awareness

*could be any one of these metrics

= f {

Marketing Inputs

Distribution + Price + Promotions (ATL, BTL) + TV Spends/GRP

Halo TV Spends/GRP + Radio/Podcast Spends + Outdoor Spends + Print Spends

YouTube Spends + Facebook Spends + Twitter Spends + Instagram Spends

Competitor Price + Competitor Media Spends + Impressions, clicks + Macro-economic factors

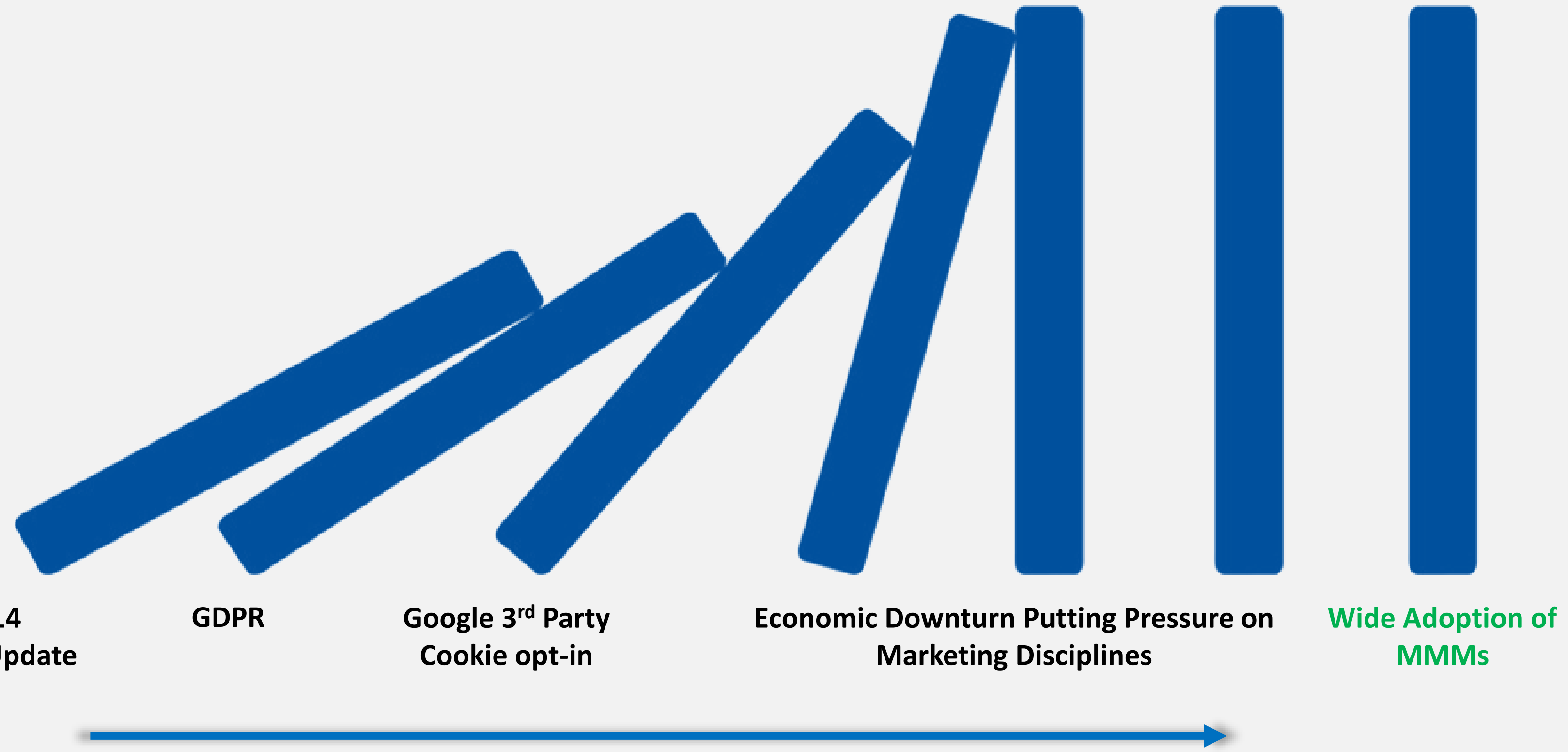
}

“Market Mix Modeling (MMM) is a technique which helps in quantifying the impact of several marketing inputs on sales or Market Share. The purpose of using MMM is to understand how much each marketing input contributes to sales, and how much to spend on each marketing input.” – Aryma Labs

Why MMM is gaining prominence (again)

ArYma Labs
Decoding the past, Encoding the future

<<Index



ArŶmα Labs

Decoding the past, Encoding the future

Statistics needed for MMM

What is correlation?

ArŶmα Labs
Decoding the past, Encoding the future

Image by Hansueli Krapf, licensed under [CC BY-SA 3.0]

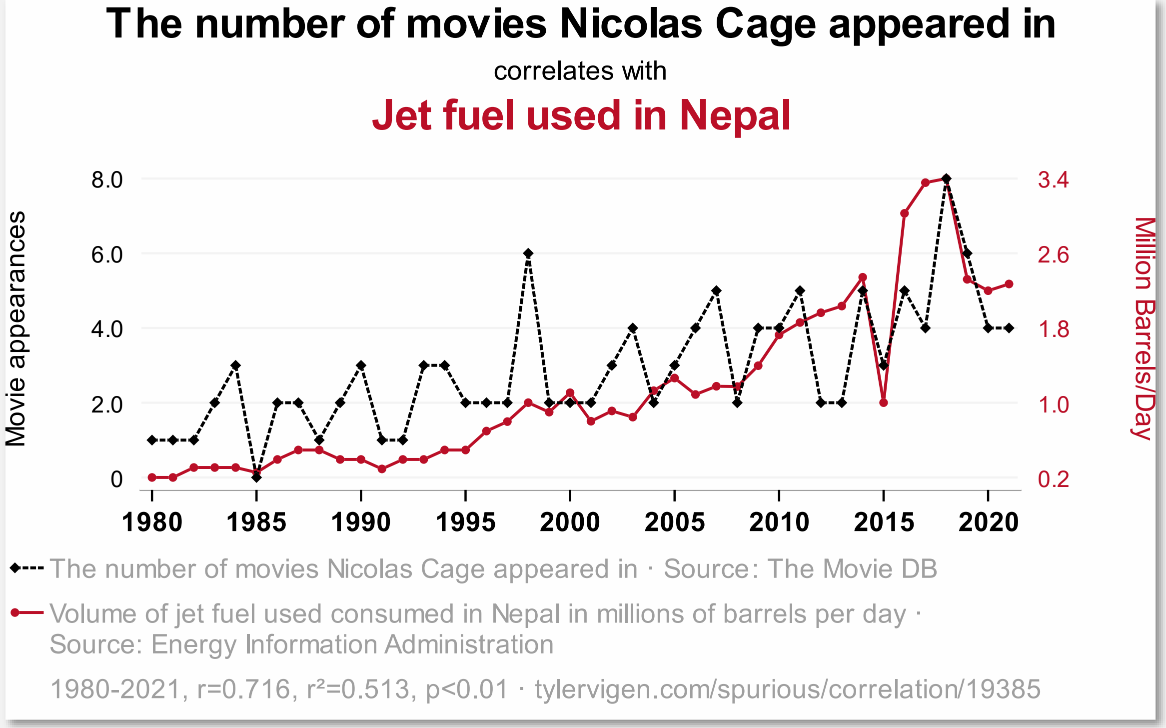


Image by Tyler Vigen

What is correlation?

- A statistical measure that expresses the extent to which two variables are linearly related.

- **It is derived from Covariance .**

$$\text{Cov(X,Y)}=\frac{1}{n}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y}).$$

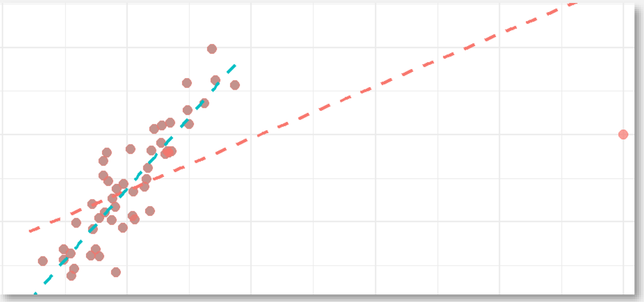
- The formula for the Pearson Product–Moment Correlation Coefficient is

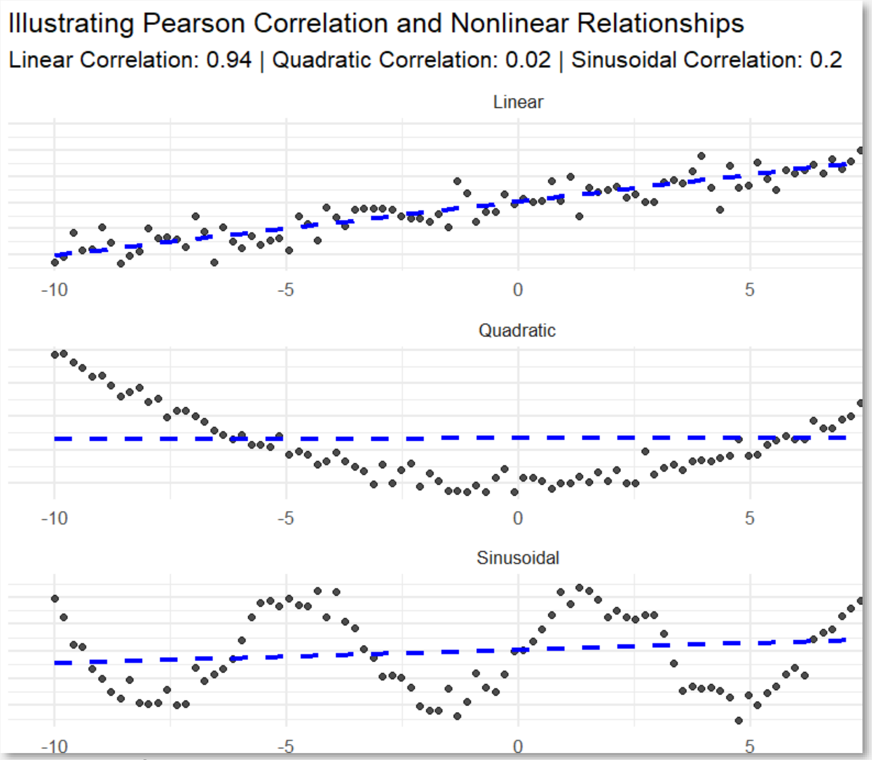
$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Where x_i and y_i are data points, $\bar{x}$ and $\bar{y}$ are mean of the variables and n is the number of data points.

Why Correlation is Not a Good Metric Alone

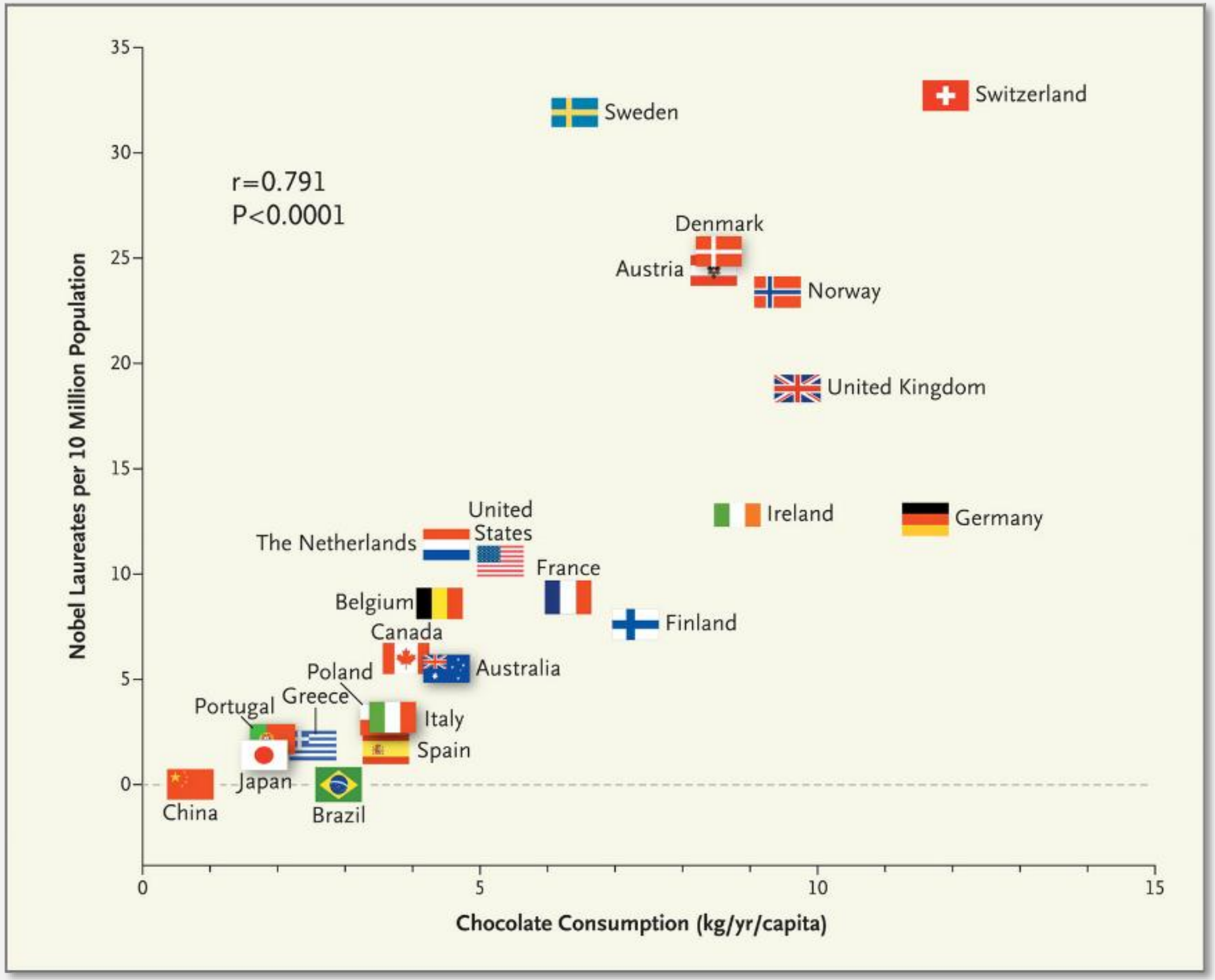
- **Does not imply causation!**
- **Sensitive to outliers (Pearson).**
- **Fails with non-linear relationships (Pearson).**



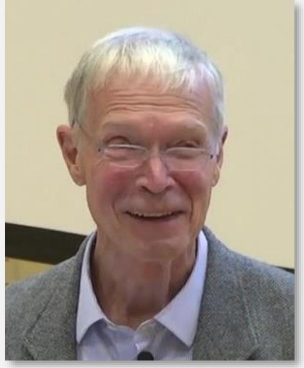


Correlation does not imply causation

ArŶmα Labs
Decoding the past, Encoding the future



- Correlation measures the relationship between two variables, but it doesn't tell us if one variable causes the other to change.



+

=

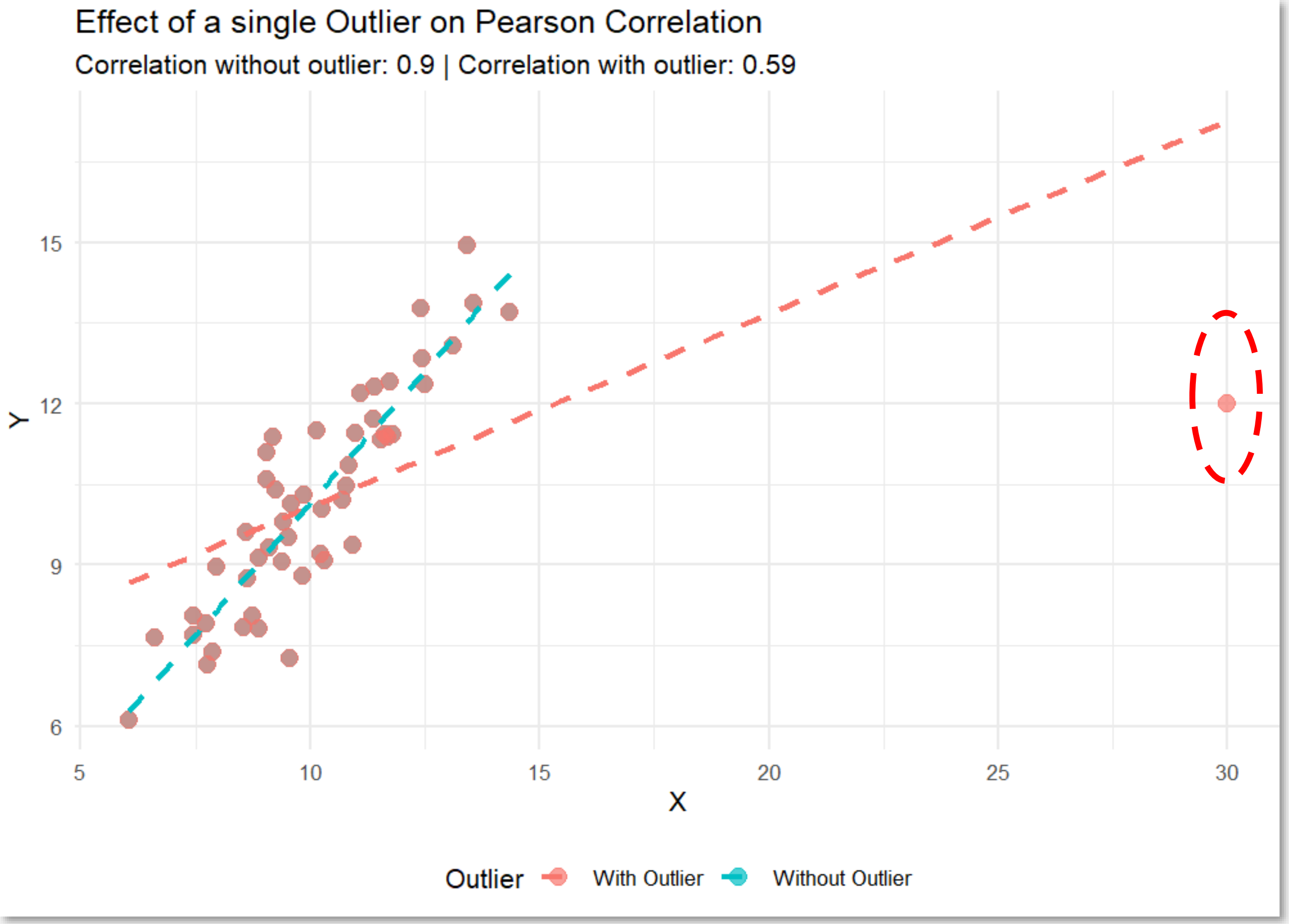
?

Correlation does not imply causation

ArŶmα Labs
Decoding the past, Encoding the future

Sensitive to outliers (Pearson)

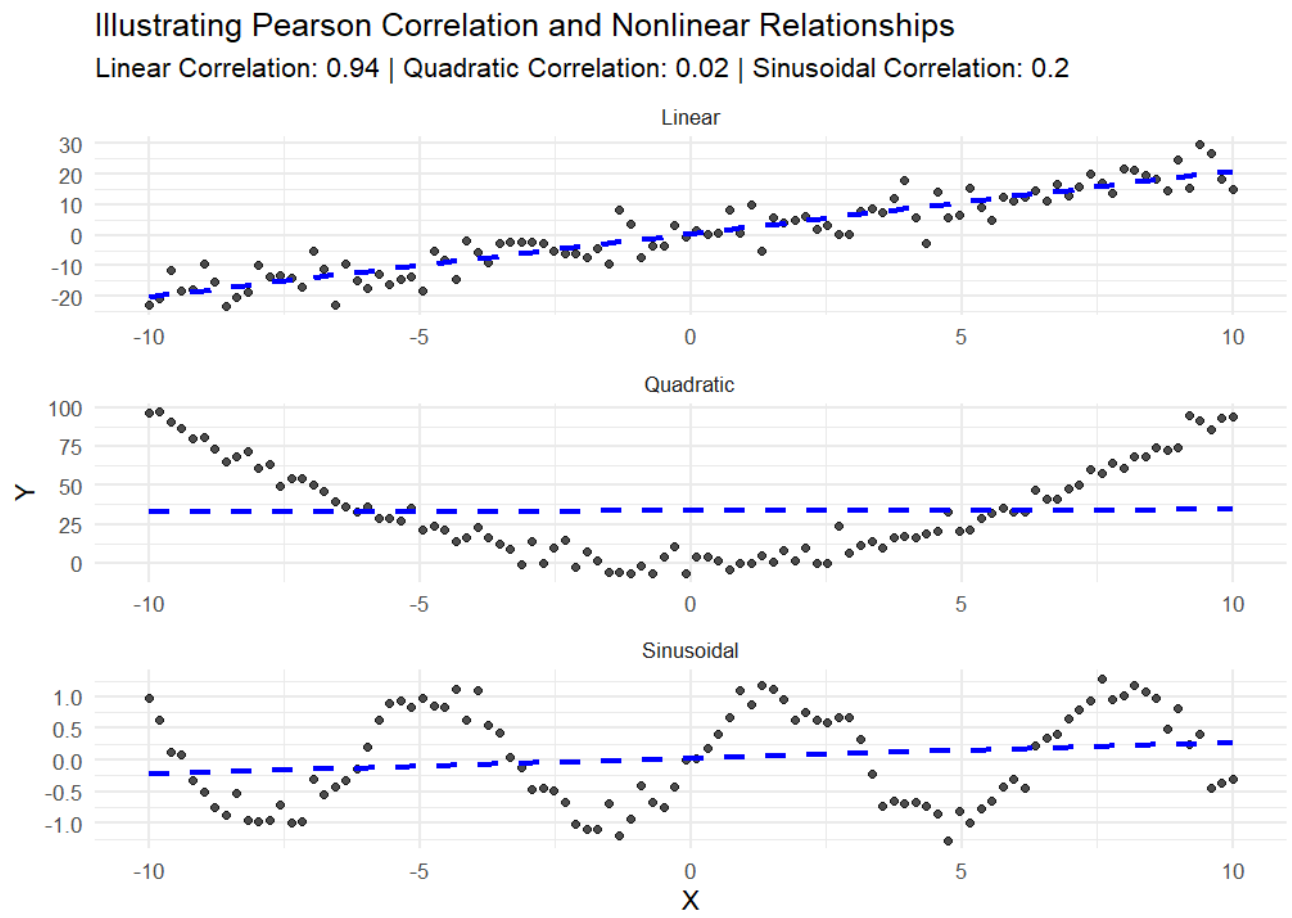
ArŶmα Labs
Decoding the past, Encoding the future



- Correlation is very sensitive to outliers in the data.
- Even a single point can have a large impact on the Pearson correlation coefficient.

Failure with non-linear relationships

ArŶmα Labs
Decoding the past, Encoding the future



- Correlation is only sensitive to monotonic or linear relations in the data.

Types of Correlation

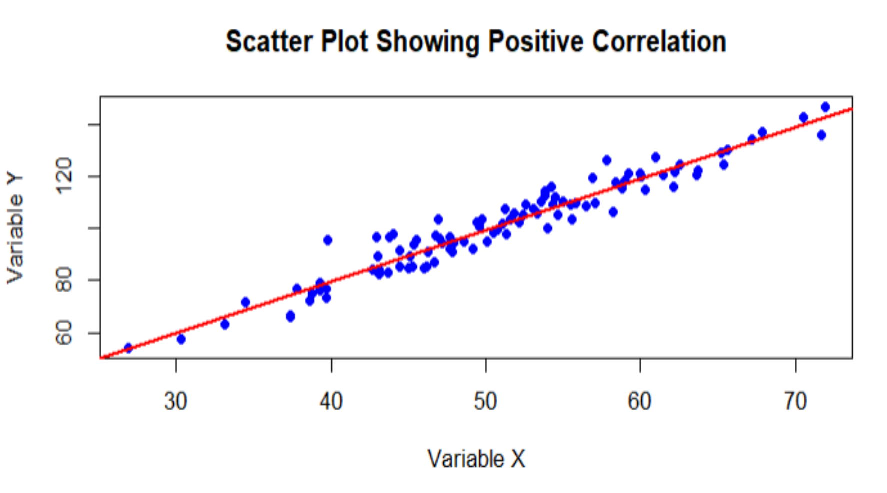
ArŶmα Labs
Decoding the past, Encoding the future

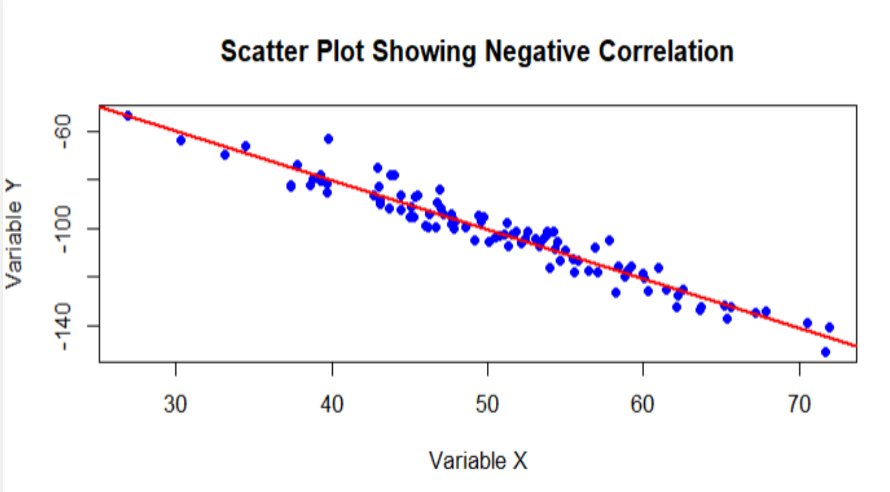
- Pearson Product-moment Correlation (r)
- Spearman Rank Correlation (ρ)
- Kendall Tau Correlation (τ)

Correlation or Pearson Product-moment Correlation (r)

ArŶmα Labs
Decoding the past, Encoding the future

- Pearson correlation measures the strength and direction of only the linear relationship between two continuous variables.
- It is bounded in the closed interval from -1 to 1.
- Positive correlation has r > 0.
- Negative correlation has r < 0.
- Zero for independent variables.





Pearson Product-moment Correlation (r)

ArŶmα Labs
Decoding the past, Encoding the future

- The formula for the Pearson Product−Moment Correlation Coefficient is

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Where x_i and y_i are data points, $\bar{x}$ and $\bar{y}$ are mean of the variables and n is the number of data points.

Spearman Rank Correlation (ρ)

ArŶmα Labs
Decoding the past, Encoding the future

- Spearman's rank correlation measures the strength and direction of the monotonic relationship between two variables.
- This method ranks the data and then calculates the correlation between those ranks.
- It is bounded in the closed interval from -1 to 1.
- Zero for independent variables.
- For the example, -0.1 is the Spearman rank correlation.

X	Y
15	13
47	562
78	2
45	78
96	52

Spearman Rank Correlation (ρ)

ArŶmα Labs
Decoding the past, Encoding the future

- Spearman's rank correlation measures the strength and direction of the monotonic relationship between two variables.
- This method ranks the data and then calculates the correlation between those ranks.
- It is bounded in the closed interval from -1 to 1.
- Zero for independent variables.
- For the example, -0.1 is the Spearman rank correlation.

X	Y	Rank of observation in X	Rank of observation in Y
15	13	5	4
47	562	3	1
78	2	2	5
45	78	4	2
96	52	1	3

Spearman Rank Correlation (ρ)

ArŶmα Labs
Decoding the past, Encoding the future

- Spearman's rank correlation measures the strength and direction of the monotonic relationship between two variables.
- This method ranks the data and then calculates the correlation between those ranks.
- It is bounded in the closed interval from -1 to 1.
- Zero for independent variables.
- For the example, -0.1 is the Spearman rank correlation.

X	Y	Rank of observation in X	Rank of observation in Y	Difference
15	13	5	4	1
47	562	3	1	2
78	2	2	5	3
45	78	4	2	2
96	52	1	3	2

Spearman Rank Correlation (ρ)

ArŶmα Labs
Decoding the past, Encoding the future

- The formula for Spearman Rank Correlation is:

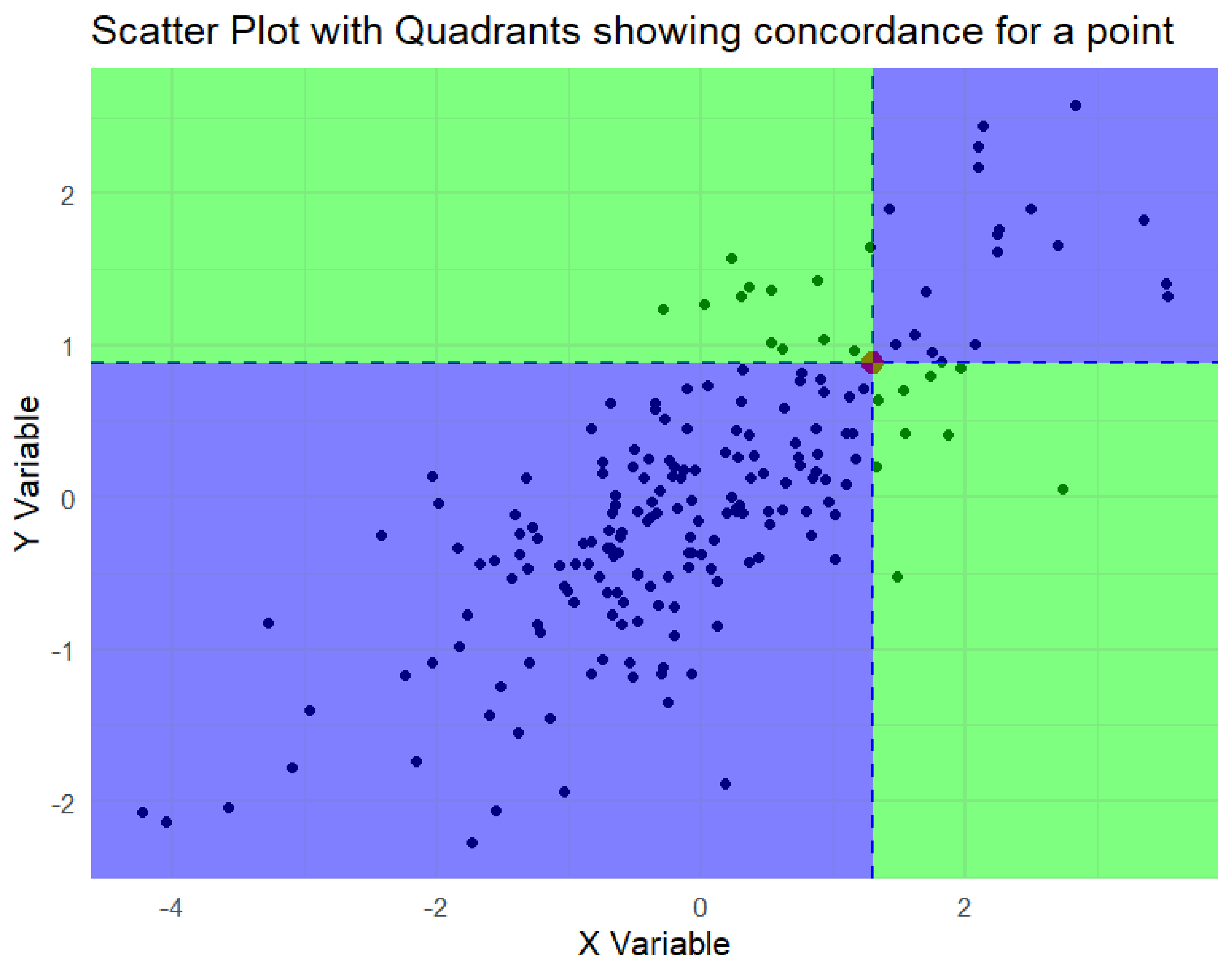
$$\rho = 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{n(n^2 - 1)}$$

- Where, $d_i = R(x_i) - R(y_i)$ is the difference in the ranks of corresponding values of x and y,
- n is the number of data points,
- $R(x_i)$ and $R(y_i)$ represent the corresponding ranks of the x and y variables.

Kendall Tau Correlation (τ)

ArŶmα Labs
Decoding the past, Encoding the future

- **Kendall Tau correlation measures the strength of agreement between two ranked variables by comparing the number of concordant and discordant pairs in the data.**
- **It is bounded in the closed interval from -1 to 1.**
- **It is zero for independent random variables.**
- **Kendall Tau compares the concordance (both rankings agree) and discordance (rankings disagree) between the two event rankings.**



Kendall Tau Correlation (τ)

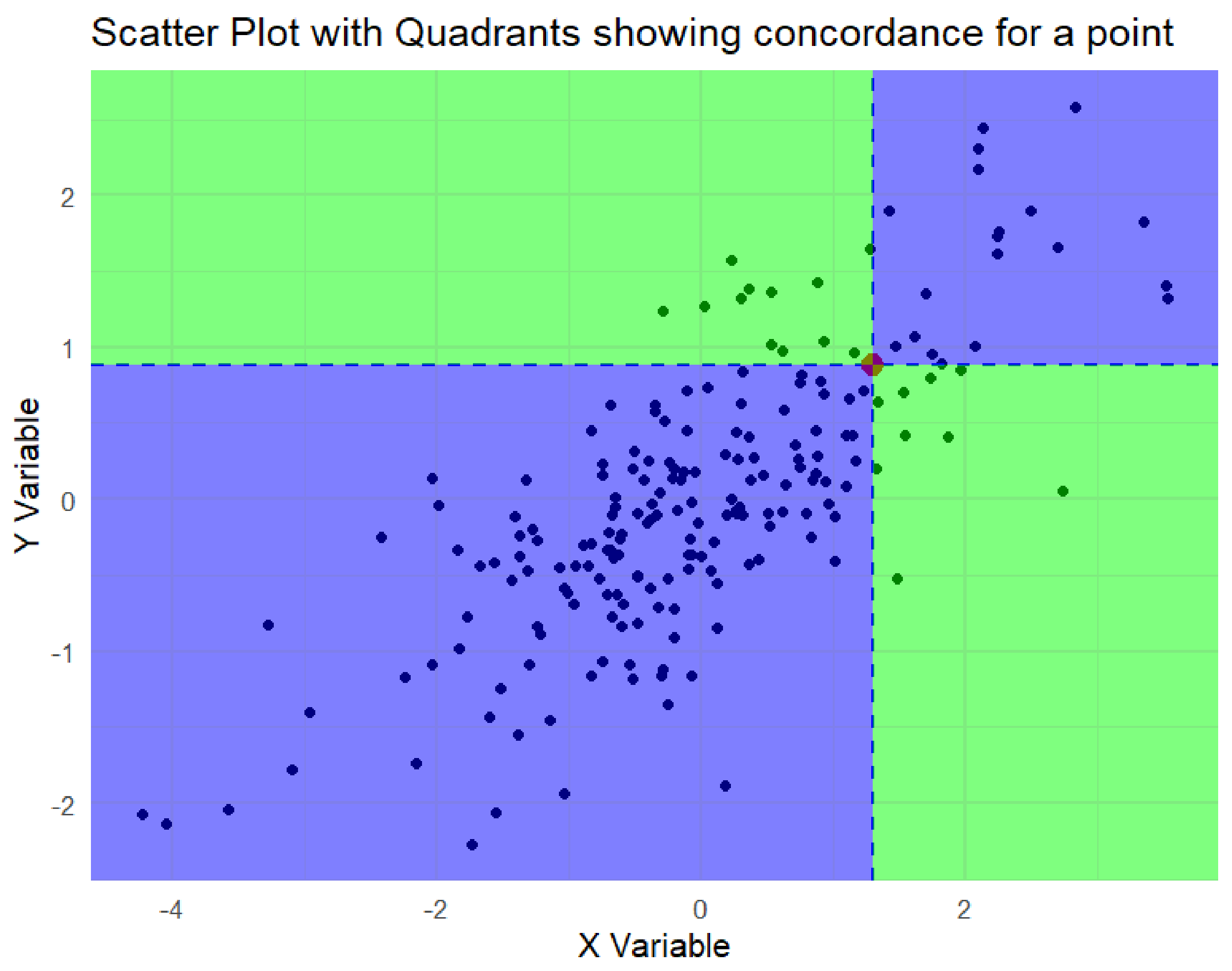
ArŶmα Labs
Decoding the past, Encoding the future

- If we have two observations (x_i, y_i) and (x_j, y_j), such that $(x_i - x_j)(y_i - y_j)$ is positive, such a pair is said to be concordant.

$$(x_i - x_j)(y_i - y_j) > 0$$

- If we have two observations (x_i, y_i) and (x_j, y_j), such that $(x_i - x_j)(y_i - y_j)$ is negative, such a pair is said to be discordant.

$$(x_i - x_j)(y_i - y_j) < 0$$



Kendall’s Tau Correlation (τ)

ArŶmα Labs
Decoding the past, Encoding the future

- The formula for Kendall's Tau Correlation Coefficient is

$$\tau = \frac{C - D}{\binom{n}{2}}$$

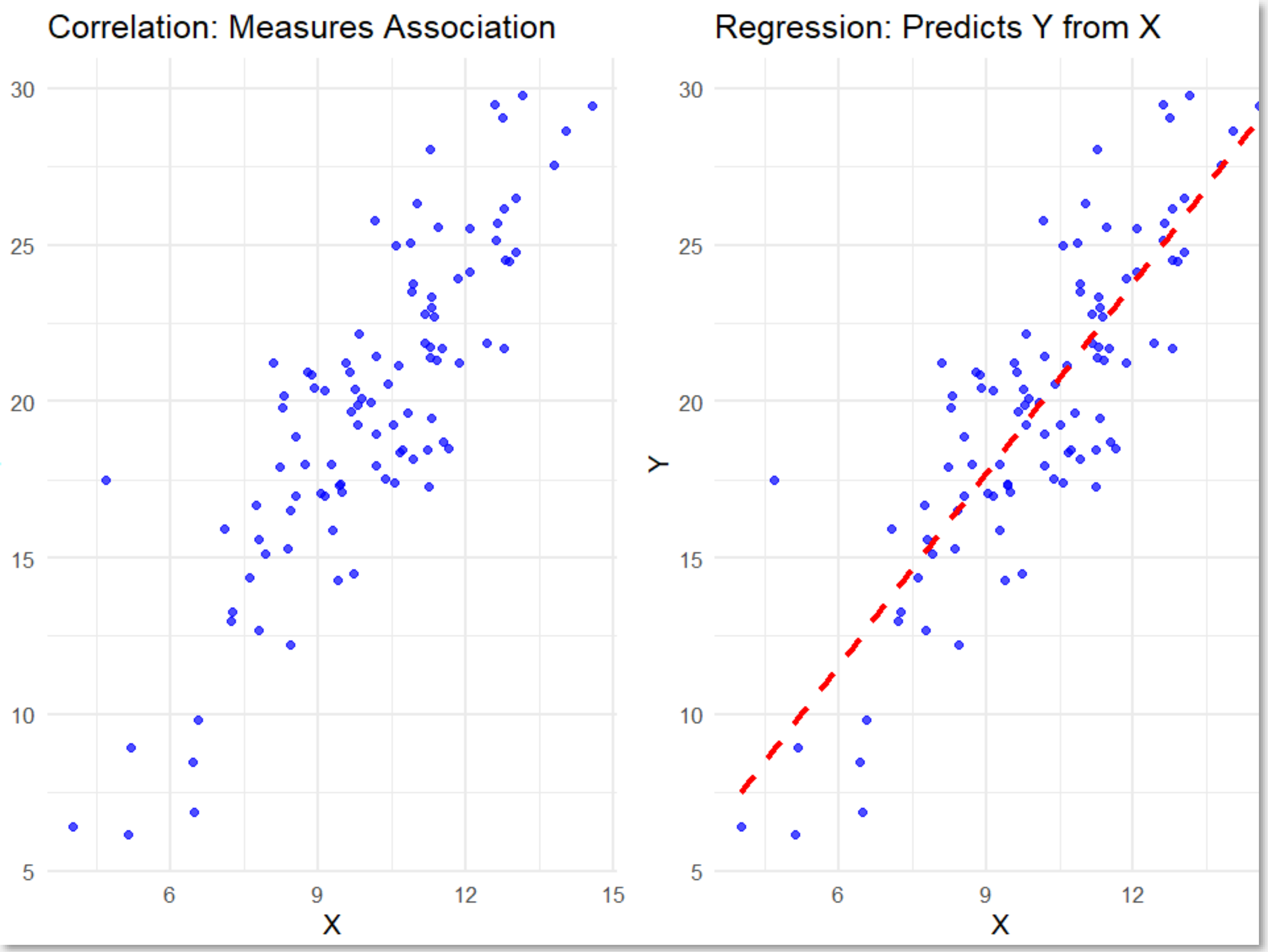
- Where C is the number of Concordant pairs
- D is the number of discordant pairs
- n is the number of data points

Is Correlation = Regression?

ArŶmα Labs
Decoding the past, Encoding the future

Correlation and Regression: Connections and Similarities

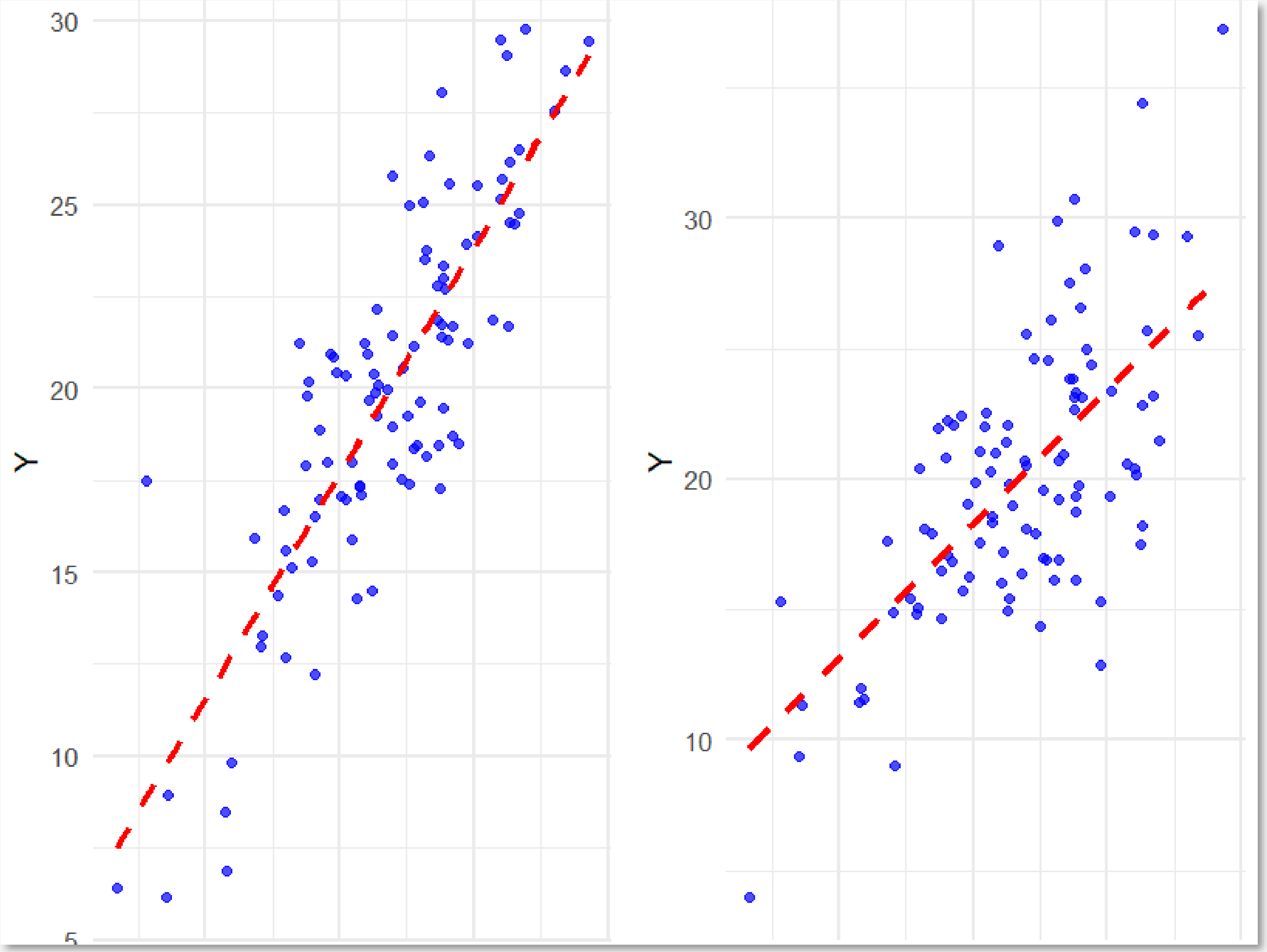
ArŶmα Labs
Decoding the past, Encoding the future



- For two variables with the same variance, both the correlation coefficient(Pearson) and slope in linear regression are the same.
- In case of two variables, both asses the linear relationship between them.
- The sign of the correlation coefficient and slope in linear regression is the same.
- Both assume constant variance for the variables.

Correlation and Regression: Key Differences

ArŶmα Labs
Decoding the past, Encoding the future



- Correlation is symmetric in the variables, but regression is not.
- Regression describes/predicts the relationship in addition to measuring the strength and direction of the relationship.
- **Correlation coefficient is unitless, while the regression slope has units.**
- **Regression can be used to define the causal effect of one variable on another, while correlation cannot.**

Difference between Multivariate and Multi-variable Regression (MVR)

ArŶmα Labs
Decoding the past, Encoding the future

Multivariate Regression has a multivariate response.

$$\begin{pmatrix} Sales \\ \begin{pmatrix} Brand \\ Awareness \end{pmatrix} \end{pmatrix} = \begin{pmatrix} \beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n \\ \alpha_0 + \alpha_1 x_1 + \ldots + \alpha_n x_n \end{pmatrix} + \vec{\epsilon}$$

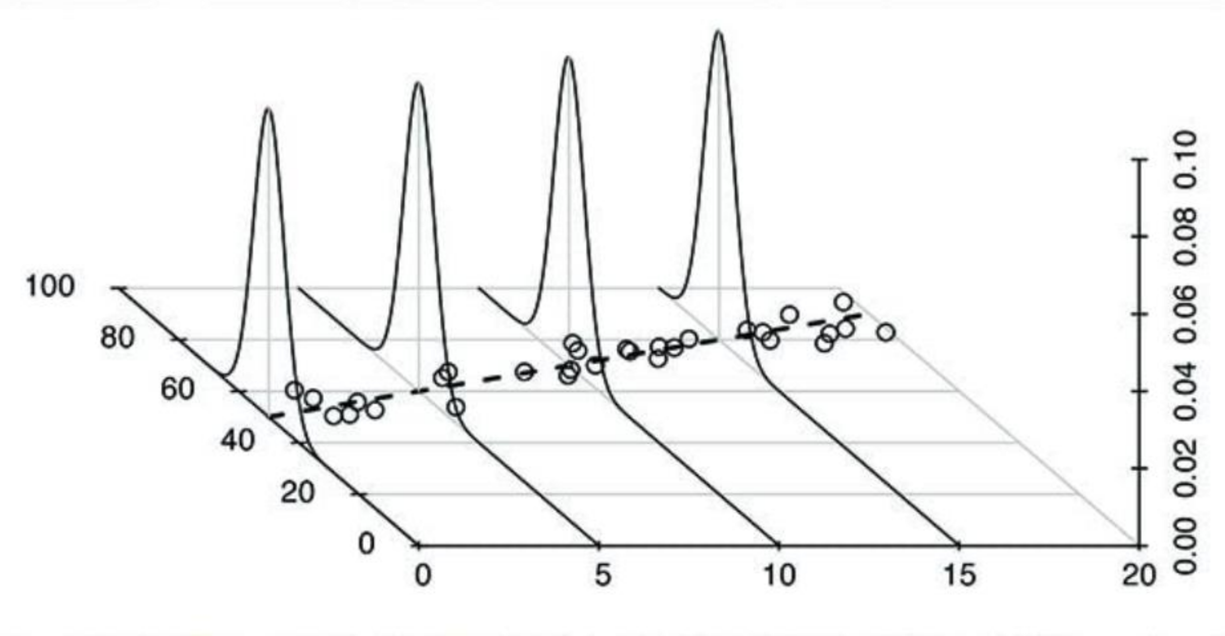
Multivariable regression has a univariate response.

$$Sales = \beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n + \epsilon$$

Multi Linear Regression is all about conditional expectation

ArŶma Labs
Decoding the past, Encoding the future

- Fitted values $\hat{y} = X\hat{\beta}$ are explicitly calculated based on the observed covariates X.
- The estimated coefficients $\hat{\beta}$ are derived by optimizing the model fit, making them dependent on the structure and values of X.



Book cover of '**Understanding Regression Analysis**' in the **Paperback Edition** by Andrea Arias and Peter Westfall

Assumptions in Statistics – Why are they required?

When it comes to statistical tests or ML algorithms, we make many assumptions.

For e.g. in Linear regression, we make assumptions like:

"Errors need to be normally distributed"
"Independence of errors"
"Linearity"
"Homoscedasticity"

Why do we make these assumptions ? What purpose do they serve ?

One might be right in thinking that it is for mathematical / statistical convenience.

But the deeper answer is that:

We make assumptions because in a way it means that we have less parameters to estimate.

The more assumptions we make, the lesser parameters needs to be estimated.

Assumptions in Statistics – Some Caveats

ArŶmα Labs
Decoding the past, Encoding the future

- Breaking assumptions is common.
- We can test the degree to which assumptions are violated.
- Violation of assumptions to some degree can be managed, while extreme cases can require newer methods like robust regression.

ArŶmα Labs

Decoding the past, Encoding the future

Assumptions of Multi Variable Regression

Assumptions of Multi-variable regression

ArŶmα Labs

Decoding the past, Encoding the future

Linearity

(Linear In terms of coefficients)

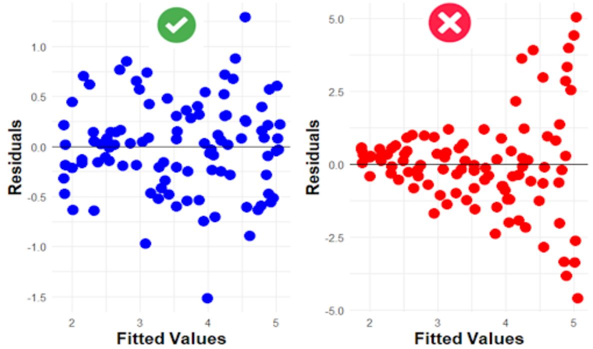
$y = \beta_0 + \beta_1 x + \text{error}$ ✔

$y = \beta_0 + \beta_1 \sqrt{x} + \text{error}$ ✔

$y = \beta_0 x^{\beta_1} + \text{error}$ ✖

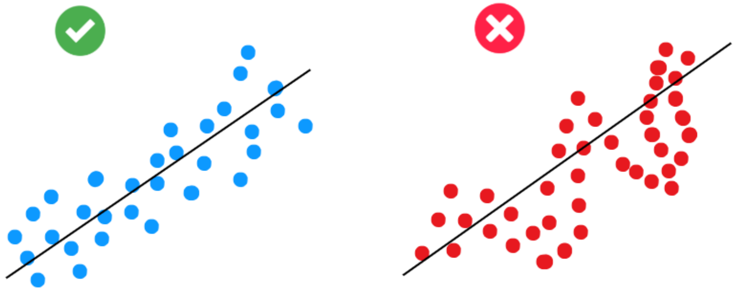
Homoscedasticity

(Equal variance of error terms)



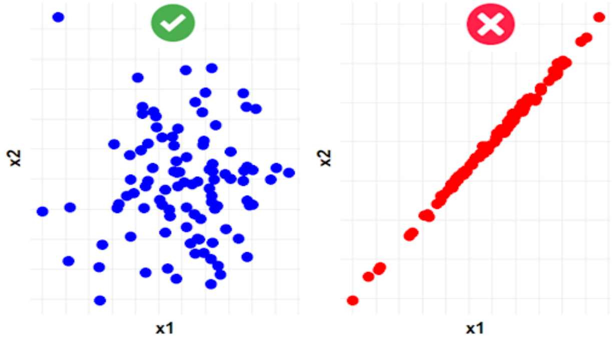
Independence

(Error terms are Independent)



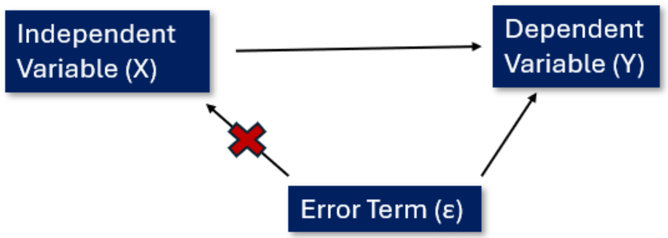
Lack of Multicollinearity

(Predictors are uncorrelated with each other)



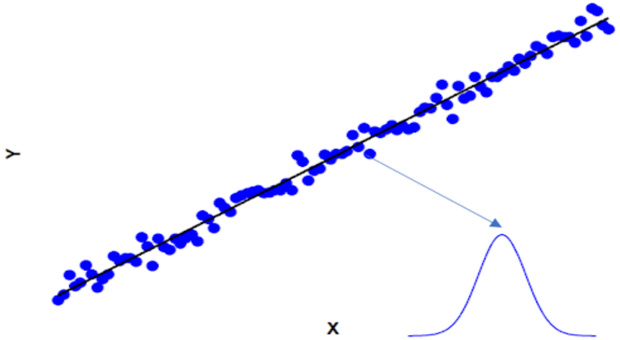
Absence of Endogeneity

(No correlation with Predictors and error)



Normality

(Errors are normal)



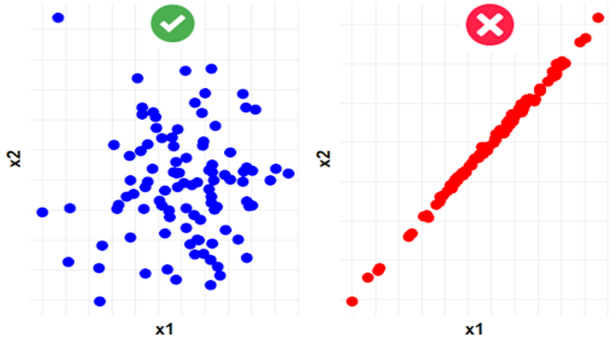
Important assumptions for MMM

ArYma Labs
Decoding the past, Encoding the future

- Most important assumptions for MMM are
  - Lack of Multicollinearity,
  - Absence of Endogeneity
  - Homoscedasticity

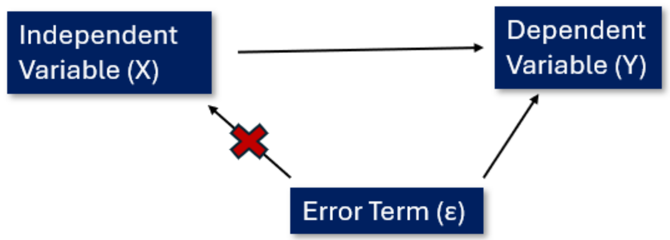
Lack of Multicollinearity

(Predictors are uncorrelated with each other)



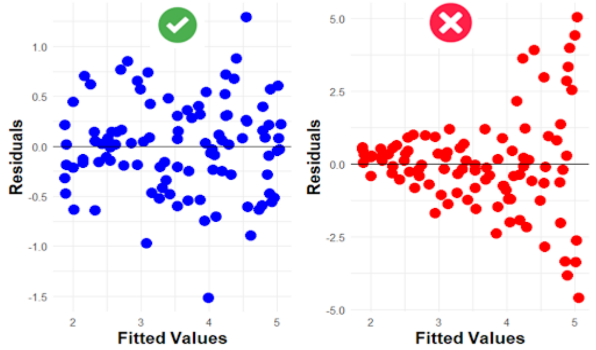
Absence of Endogeneity

(No correlation with Predictors and error)



Homoscedasticity

(Equal variance of error terms)



Important assumptions for MMM – Multicollinearity

What problems it can cause

ArŶmα Labs
Decoding the past, Encoding the future

In the marketing mix modeling space, attribution is everything. Failure to do so is a huge downer.

stat_daddy • 4y ago • Edited 4y ago •
Statistician

For the purposes of statistical tests that are designed to measure differences between quantities, I like to think of "power" as being analogous to the *magnifying power* of a magnifying glass.

Suppose you have two objects positioned so close to another that you can't tell by the naked eye whether they are physically connected or not. You know that the objects are either connected or not, but you need a magnifying glass in order to visualize the space between them (if it exists).

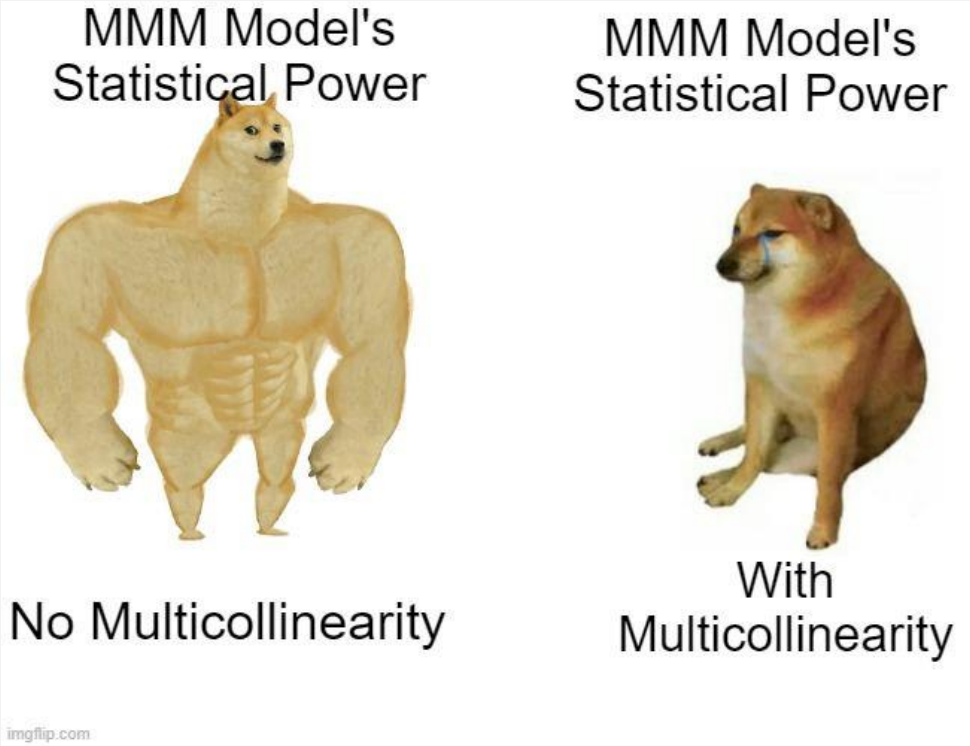
(Suppose we're talking VERY tiny distances here)

In this analogy, the distance between the objects is analogous to an effect size, and you are attempting to show that the distance is >0.

If the distance is relatively large, then a weak magnifying glass would be sufficient to show that there is a difference between the two objects. However, if the distance is very small, you might need a very powerful magnifying glass before you could see any difference. Similarly the more "powerful" a statistical test is, the smaller the difference between two quantities it can resolve (for some allowable degree of uncertainty).

However, This can sometimes backfire because a small difference might be meaningful in a statistical sense but not a practical one!, As another example, suppose I have an identical twin. He and I are the same height for all practical purposes, but if you used a powerful enough magnifying glass I'm sure you would find that our heights differ by some small amount. (Part of the issue here is that our hypothesis doesn't have any understanding of what a "practical" difference is or even what "practical" means - the difference in our heights is greater than zero, after all)

In the same way, KS tests are prone to concluding that samples differ from normality even when those differences may be of no practical concern. Your advisor is cautioning you that even though your KS tests may indicate a difference from "normal", it doesn't address the question you really need the answer to, which is "is my sample TOO non-normal for my analytical strategy?"



Multicollinearity is a signal redundancy problem rather than a signal deficit problem.

Important assumptions for MMM – Homoscedasticity

What problems it can cause

ArŶmα Labs
Decoding the past, Encoding the future

Residuals offer telltale signs of how consistent your model is. Ideally you would want some kind of stability in your model.

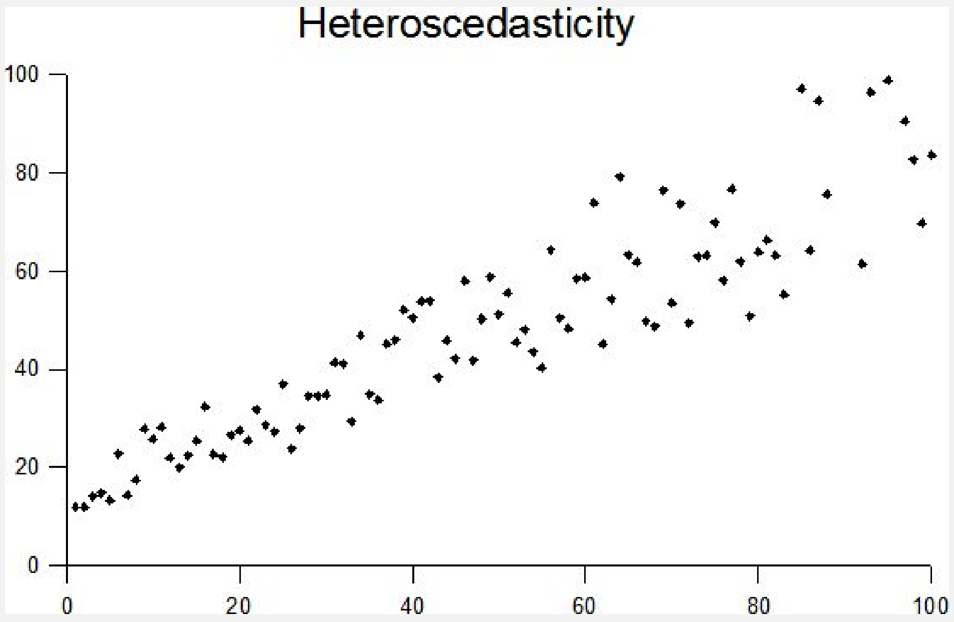
In case of Heteroscedasticity, the residuals by and large have some pattern. Let's take an example of one of the popular pattern.

When you have fan/funnel shape of the residuals, it means the model is getting worse over time since it indicates inflation of error.

How it affects MMM?

The name of the game in MMM is inference. You want your estimates to be precise and unbiased. However, Heteroscedasticity makes your estimates less precise even though it may not bias them.

Heteroscedasticity reduces the trust factor in your MMM. Given that companies make million-dollar decisions to spend on certain marketing variables, it becomes imperative that the MMM model you build is trust worthy and accurate.



Why Heteroscedasticity happens?

Heteroscedasticity often happens because of outliers or huge disparity in the range of your independent variables. For e.g. a company spends in the range of 10-15k USD every month on YouTube ads. But in few instances, say during BFCM the company decided to really ramp up their spends. Let's say this in the range of 85k-100k. Data like these would lead Heteroscedasticity in the model.

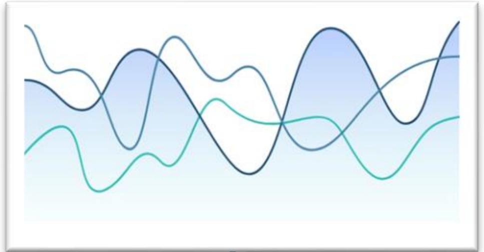
ArŶmα Labs

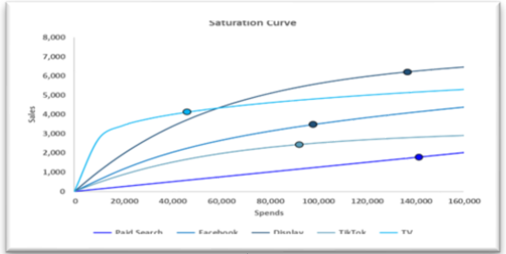
Decoding the past, Encoding the future

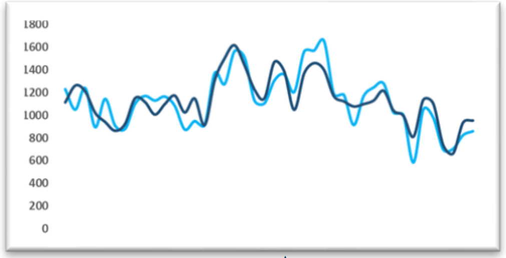
EDA

MMM Process

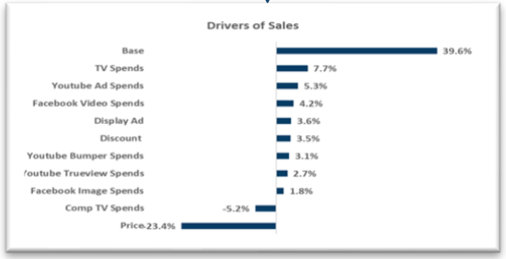
ArŶma Labs
Decoding the past, Encoding the future

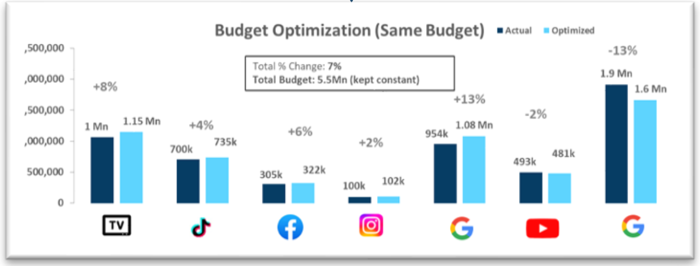






EDA → Model Building → Model Evaluation → Budget Allocator → Implementing Models on the ground → Causal Lift Validation → Frequent Model Updates

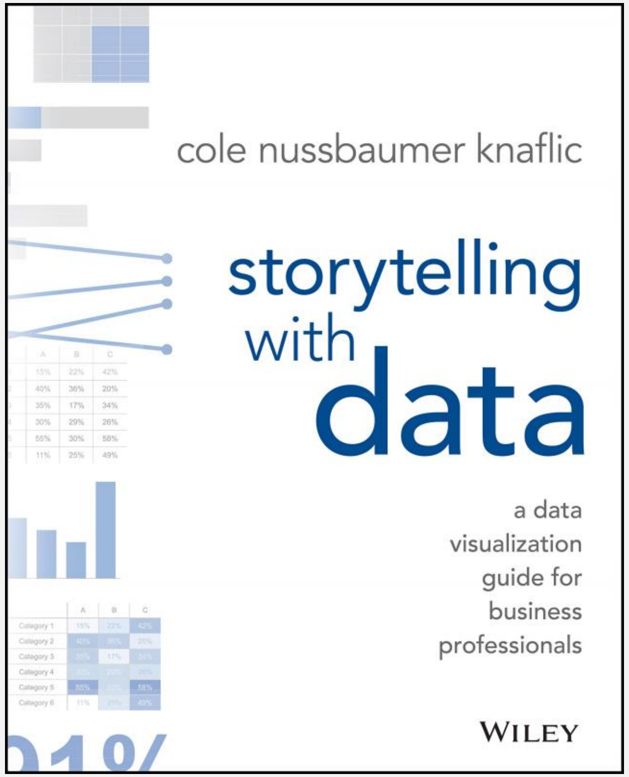


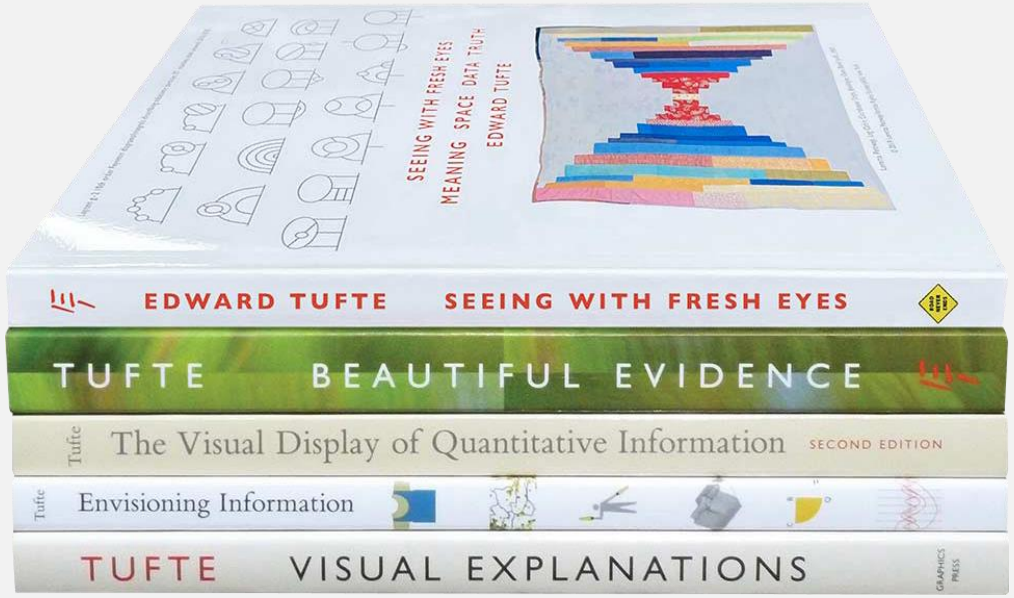


Data Visualization

ArŶmα Labs
Decoding the past, Encoding the future

Clear and effective data visualization relies on simplicity and minimizing cognitive effort.





Exploratory Data Analysis

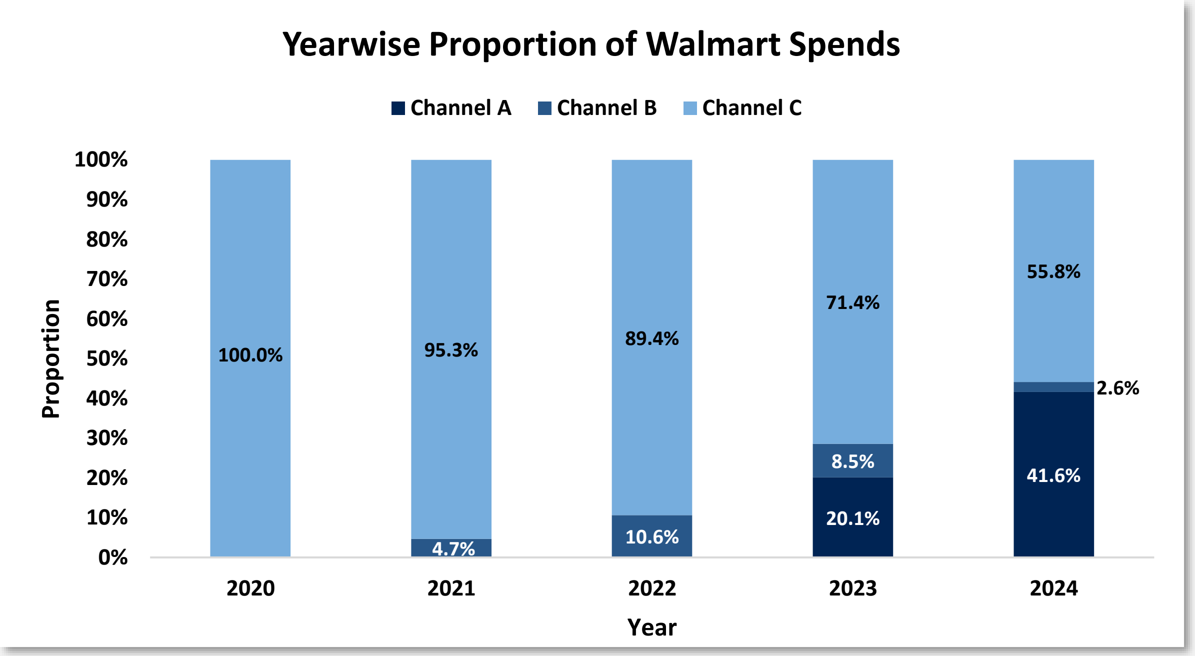
ArŶma Labs
Decoding the past, Encoding the future

What is EDA?

Exploratory Data Analysis *involves* ***summarizing main characteristics, and patterns*** *in data using statistical and visual methods.*

Why do we conduct EDA?

- It helps identify trends, outliers, and missing values.
- Provides a foundation for further analysis or modeling.

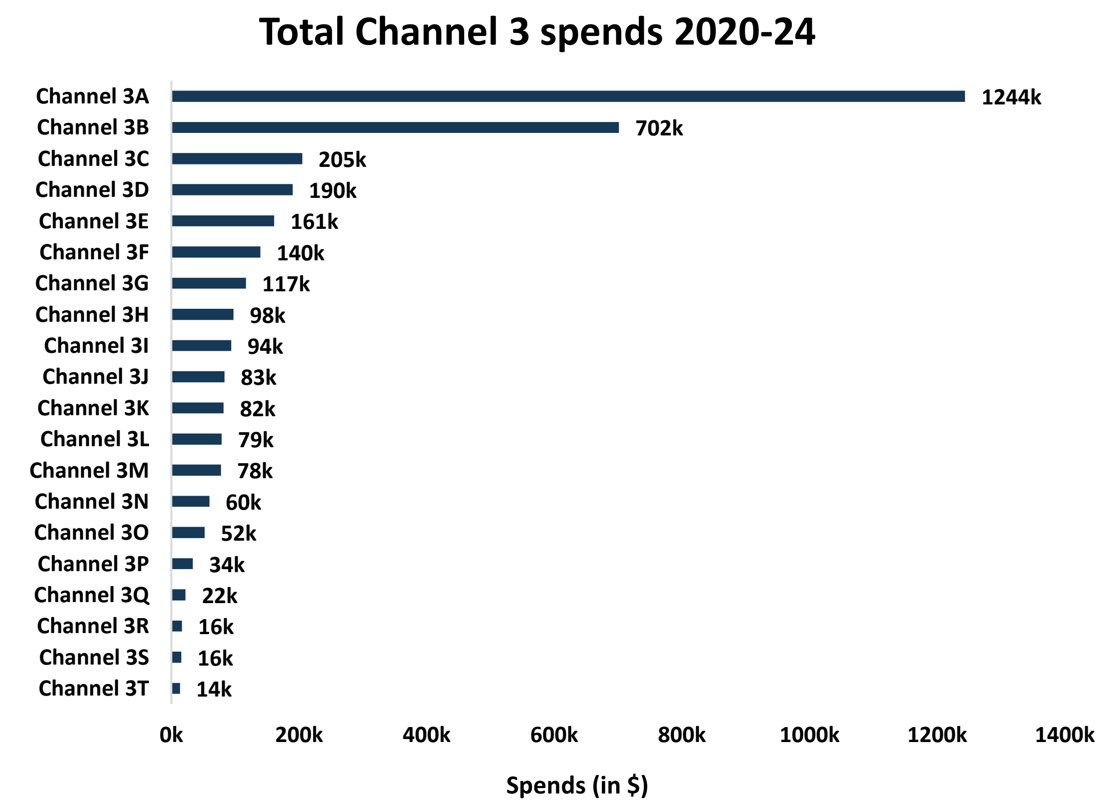


EDA – Common practices

ArŶmα Labs
Decoding the past, Encoding the future

Common EDA practices that should be followed:

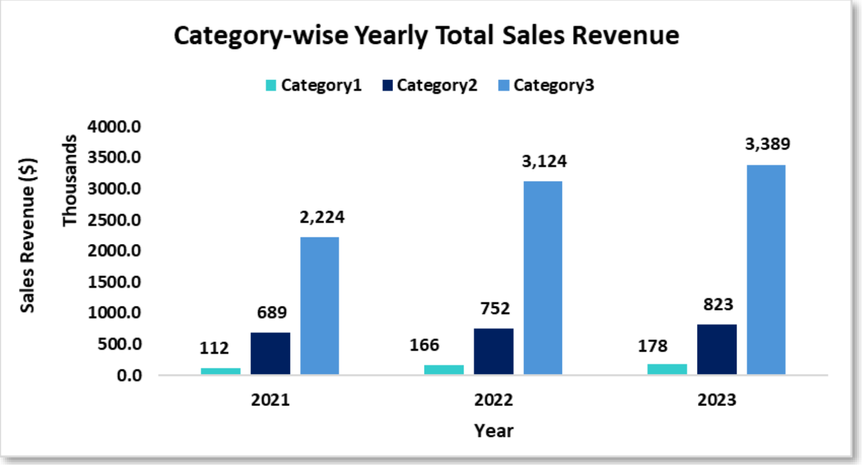
- ✓ Check for duplicates, Handle missing values,
- ✓ Analyze the distribution of individual variables,
- ✓ Identify anomalies that could skew results,
- ✓ Decompose time series into trend, seasonality, and residual components and analyze patterns over time,
- ✓ Use techniques like standardization and normalization to scale and transform the data,
- ✓ Make a summary of the documented findings and record insights.



List of EDA Tasks (1/2)

ArYma Labs
Decoding the past, Encoding the future

- Dual axis line charts
- CCF Plots
- Correlation Summary: Overall vs year-wise
- Trend charts for Sales Revenue, Volume and Price
- Category-wise Yearly Comparison charts for Rev, Vol and Price : Clustered column charts
- Channel-wise Yearly Comparison charts for Rev, Vol and Price : Clustered column charts
- Market share comparison chart : Pie Chart
- Trend chart of Inflation rate with sales revenue and volume



- Yearly comparison of total media spends : Stacked column charts
- Yearly comparison of proportion of media spends : Stacked column charts
- Yearly comparison of total and proportion of impressions, views, clicks : Clustered Columns charts
- Comparison of ATL Spends : Clustered column charts

List of EDA Tasks (2/2)

ArŶmα Labs
Decoding the past, Encoding the future

CAGR computation :

$$CAGR = \left(\frac{Avg.of\ last\ year}{Avg.of\ first\ year}\right)^{\frac{1}{n-1}} - 1$$,where n is the number of years available in the data

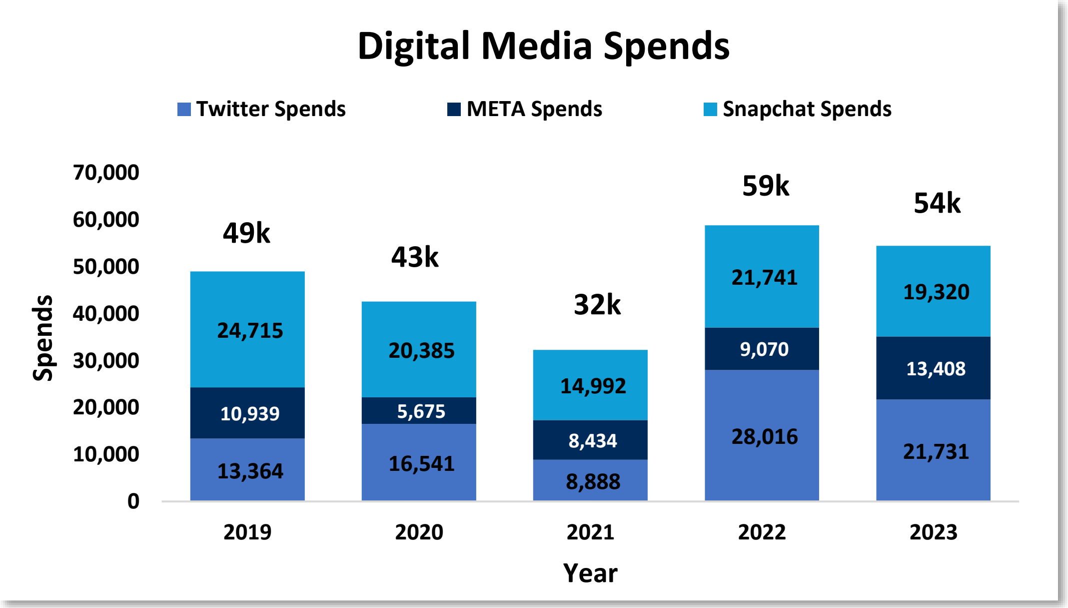
Computed Annual Growth Rate (CAGR) measures the annual growth rate of a brand. CAGR is computed for variables like Sales revenue, sales volume, price etc.

Metrics Traffic Drivers - Category1 - Year wise				
Metrics	Jan'21-Dec'21	Jan'22-Dec'22	Jan'23-Dec'23	CAGR
Sales Revenue	$ 112,309	$ 166,326	$ 178,305	26%
Sales Volume	38,444	54,279	51,215	15%
Average Price	$3.0	$3.1	$3.5	9%

General Visualization Practices

Ar**Ŷ**ma Labs
Decoding the past, Encoding the future

- Use a white background for all charts. Remove gridlines for clarity.
- Include a descriptive chart title and axis titles. Add a legend and data labels where relevant.
- Ensure charts are correctly linked to the underlying data.
- Remove underscores or technical naming conventions (e.g., "variable_names").

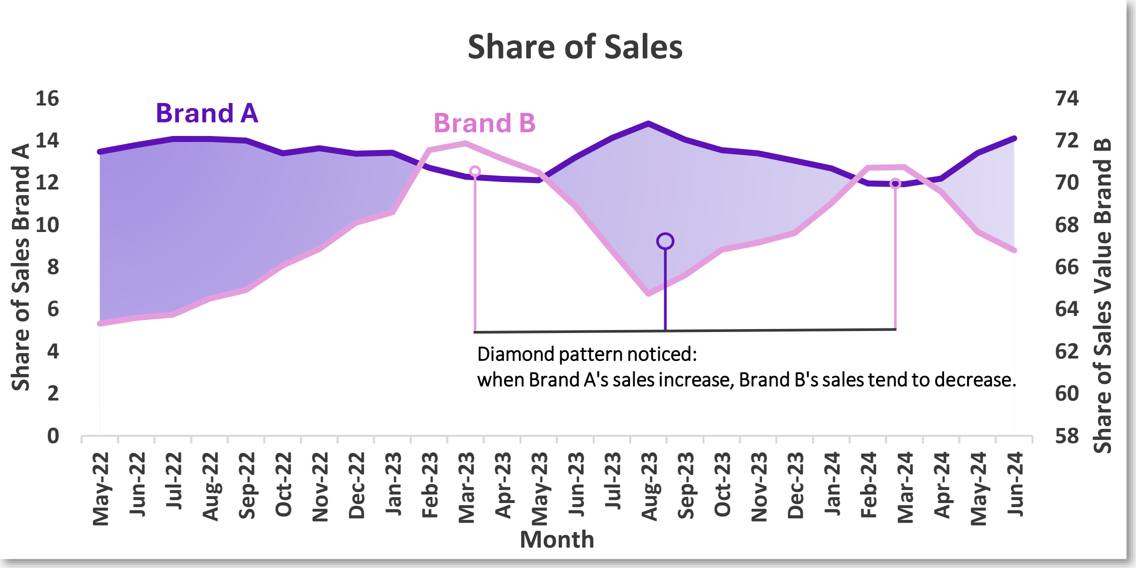


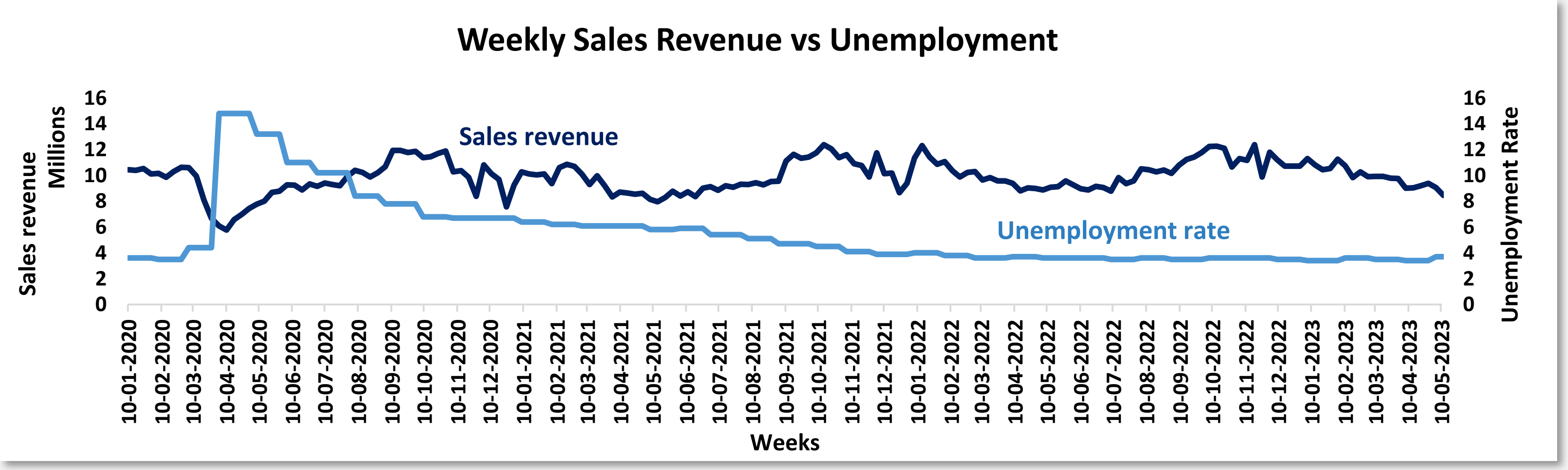
- Use the appropriate number formats. Round off values to two decimal places for consistency and readability.
- Maintain consistent fonts and colors for the same variables across all charts. Assign one color per variable and use it uniformly in all visualizations.
- Ensure all charts are labeled clearly with meaningful legends and titles.

EDA Charts (1/4)

ArŶmα Labs
Decoding the past, Encoding the future

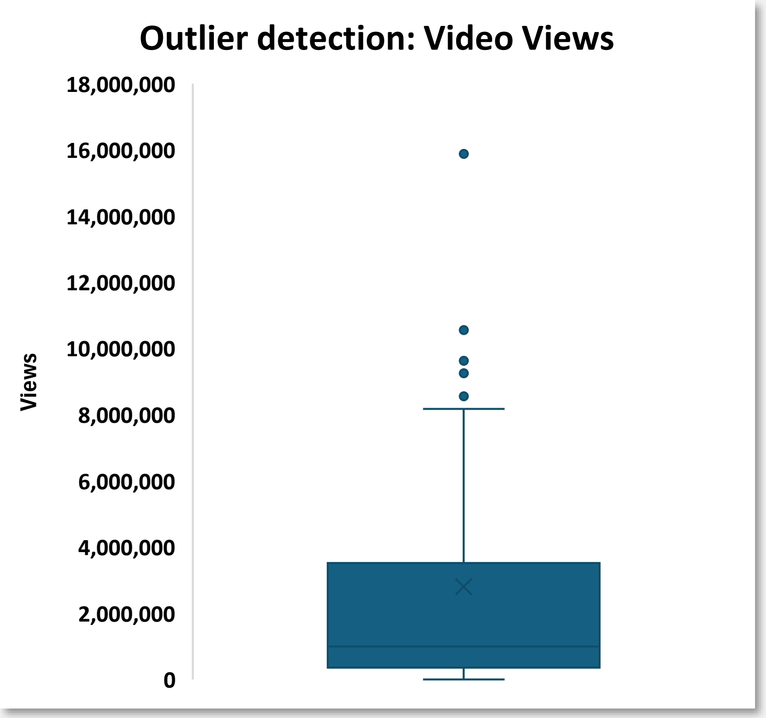
Trend Line Charts: To depict changes over time, for analyzing trends in the KPI (e.g., Revenue/Conversion, price, market share etc.)



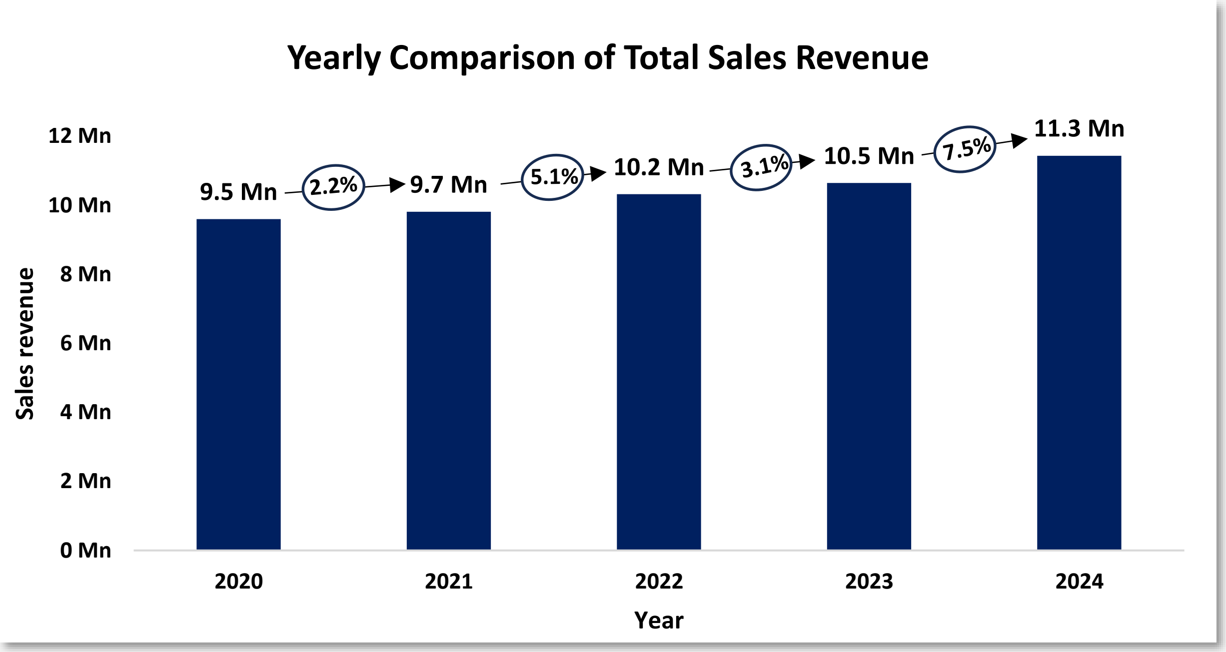


EDA Charts (2/4)

ArŶmα Labs
Decoding the past, Encoding the future



Boxplots: To identify outliers in data columns.

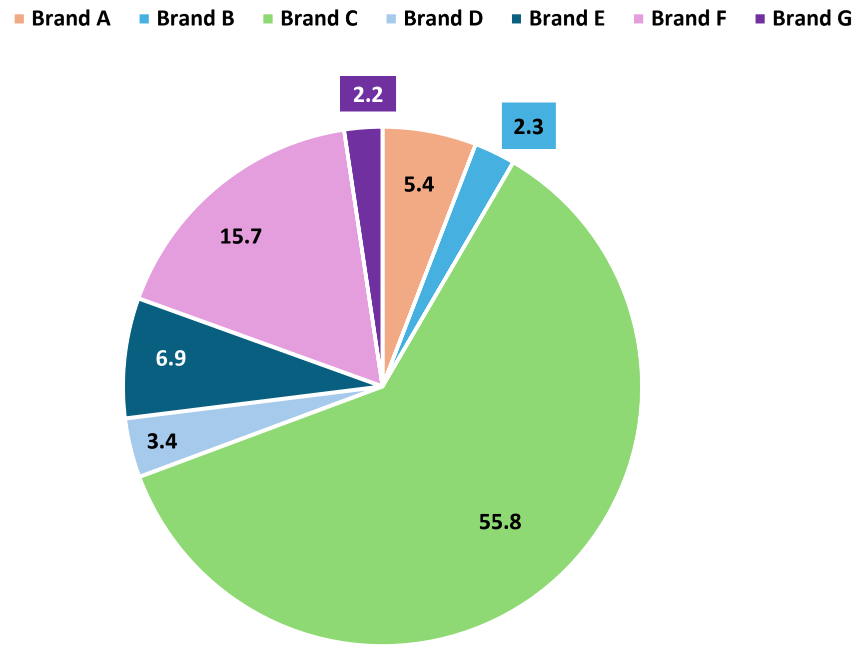


Column Charts: To analyze and compare sales revenue across different years.

EDA Charts (3/4)

ArŶmα Labs
Decoding the past, Encoding the future

Brand Market Share with Competitors



Correlation Summary

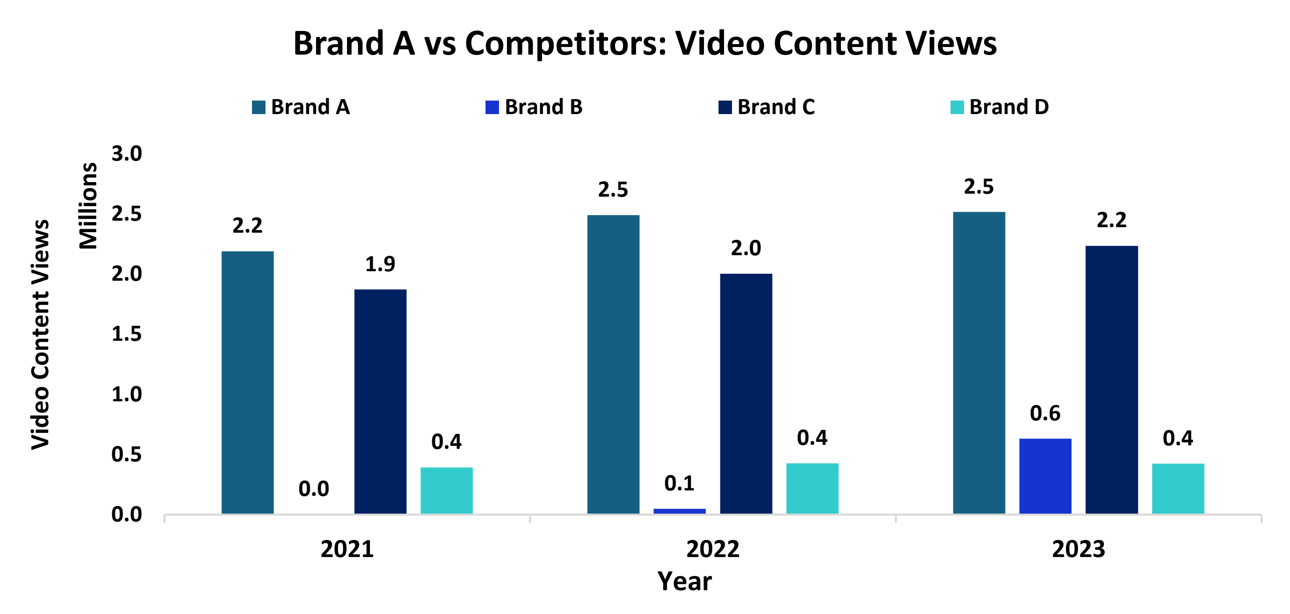
	Sales	Marketing Spend	Online Ads	TV Ads	Radio Ads	Email Campaigns	Social Media Engagement	Website Visits	Customer Retention Rate	Discounts
Sales	1									
Marketing Spend	0.094	1								
Online Ads	-0.027	0.004	1							
TV Ads	0.004	0.018	0.018	1						
Radio Ads	-0.001	-0.004	-0.017	-0.017	1					
Email Campaigns	-0.009	0.019	0.025	-0.001	0.001	1				
Social Media Engagement	0.011	0.016	0.029	-0.012	-0.005	-0.027	1			
Website Visits	-0.003	0.026	0.004	0.008	-0.002	0.007	0.002	1		
Customer Retention Rate	0.008	0.030	-0.006	-0.010	0.006	-0.001	0.000	0.013	1	
Discounts	-0.024	-0.020	0.005	0.024	0.017	-0.001	-0.003	0.000	-0.013	1

Correlation heatmaps: A color-coded matrix where each cell shows the correlation coefficient, helping identify strong positive, negative, or neutral relationships between variables.

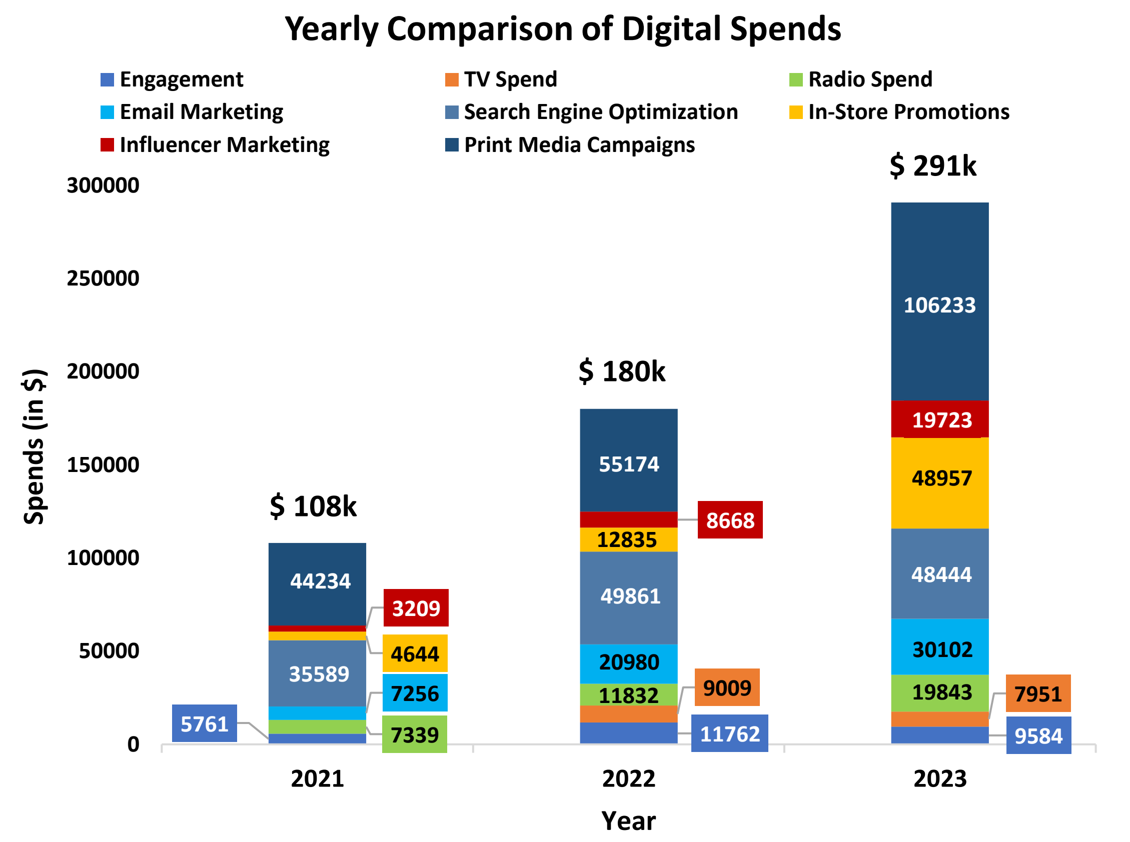
Pie Charts: Use to compare brand market shares against competitors, these charts ensure all categories sum to 100%.

EDA Charts (4/4)

ArYma Labs
Decoding the past, Encoding the future

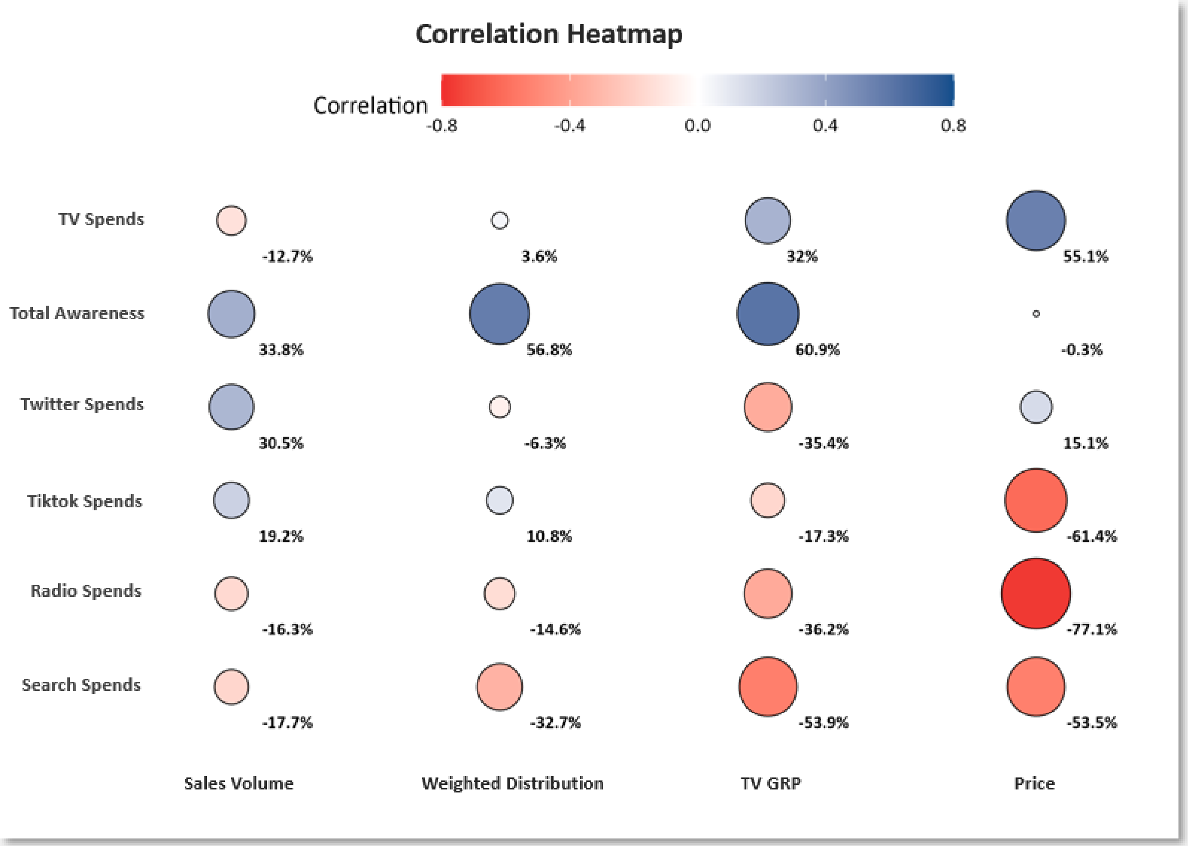


Clustered Column Charts: To compare own brand content views vs competitor video content views.



Stacked Column Charts: For yearly sum and proportion of media spend variables. (digital and traditional media variables, comparison of impressions)

Correlation Heatmap

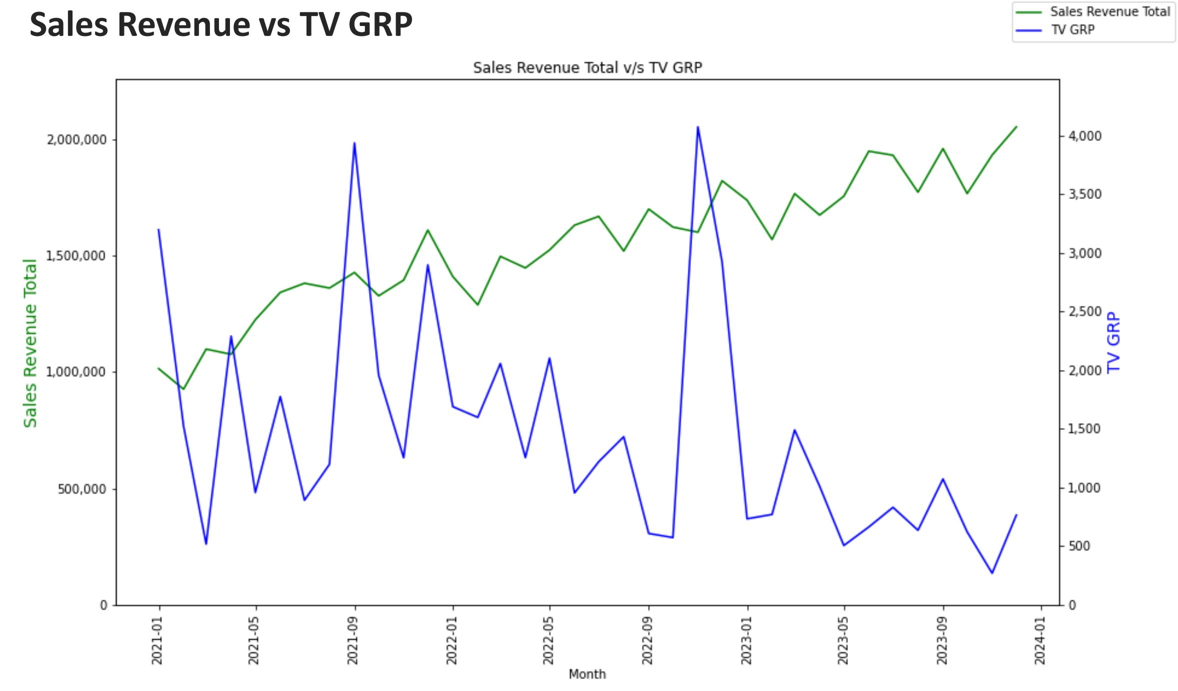


Correlation heatmaps are visual tools used to display the relationships between numerical variables.

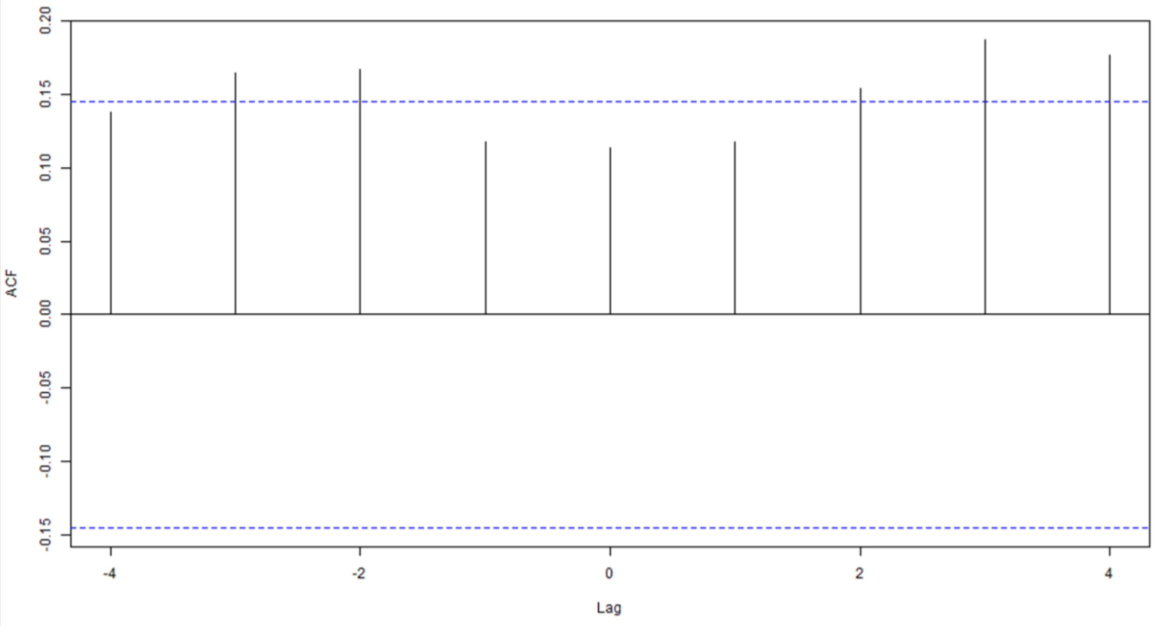
Important charts for Feature Analysis and Model Building

ArYma Labs
Decoding the past, Encoding the future

Sales Revenue vs TV GRP



Sales Revenue vs Impressions



Dual axis line charts: To show the dual-axis comparison between KPI (green line) and independent variable (blue line) over time.

CCF Plots: To identify significant lags between KPI and independent variables.

ArŶmα Labs

Decoding the past, Encoding the future

Understanding the Concept of Adstock

What is Adstock?

How many of you remember an ad that you saw say 5yrs or 10yrs ago?

Do you recall the specifics of the ad?

How many of you remember an ad that you saw couple of months ago?

Did you make a purchase decision in both the above cases?

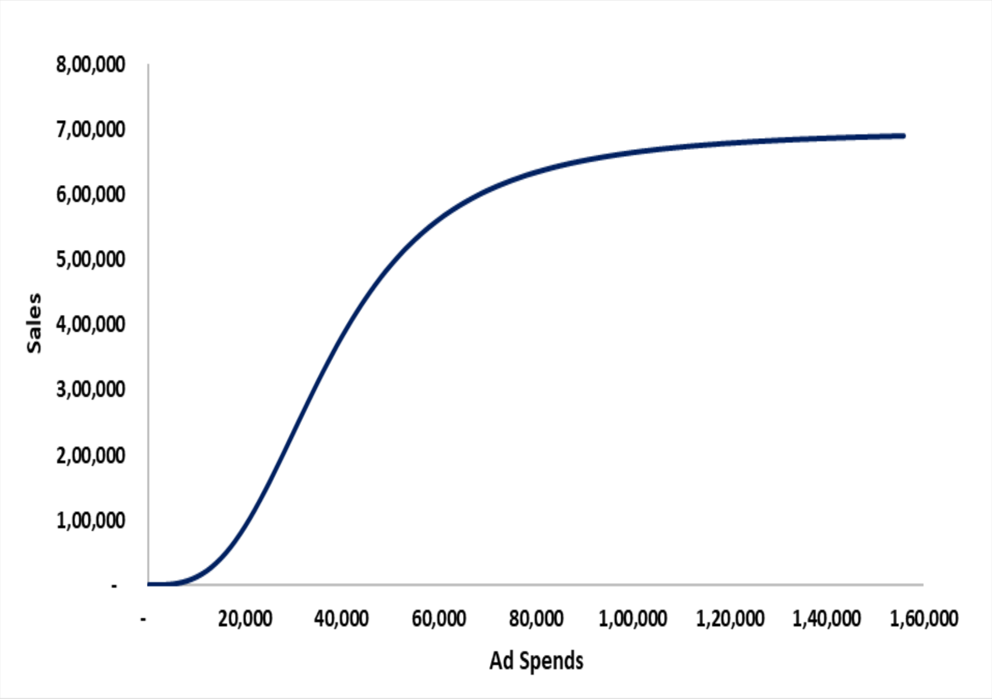
Would you develop a negative feeling towards a brand if they bombarded you with their ads?

got milk?®

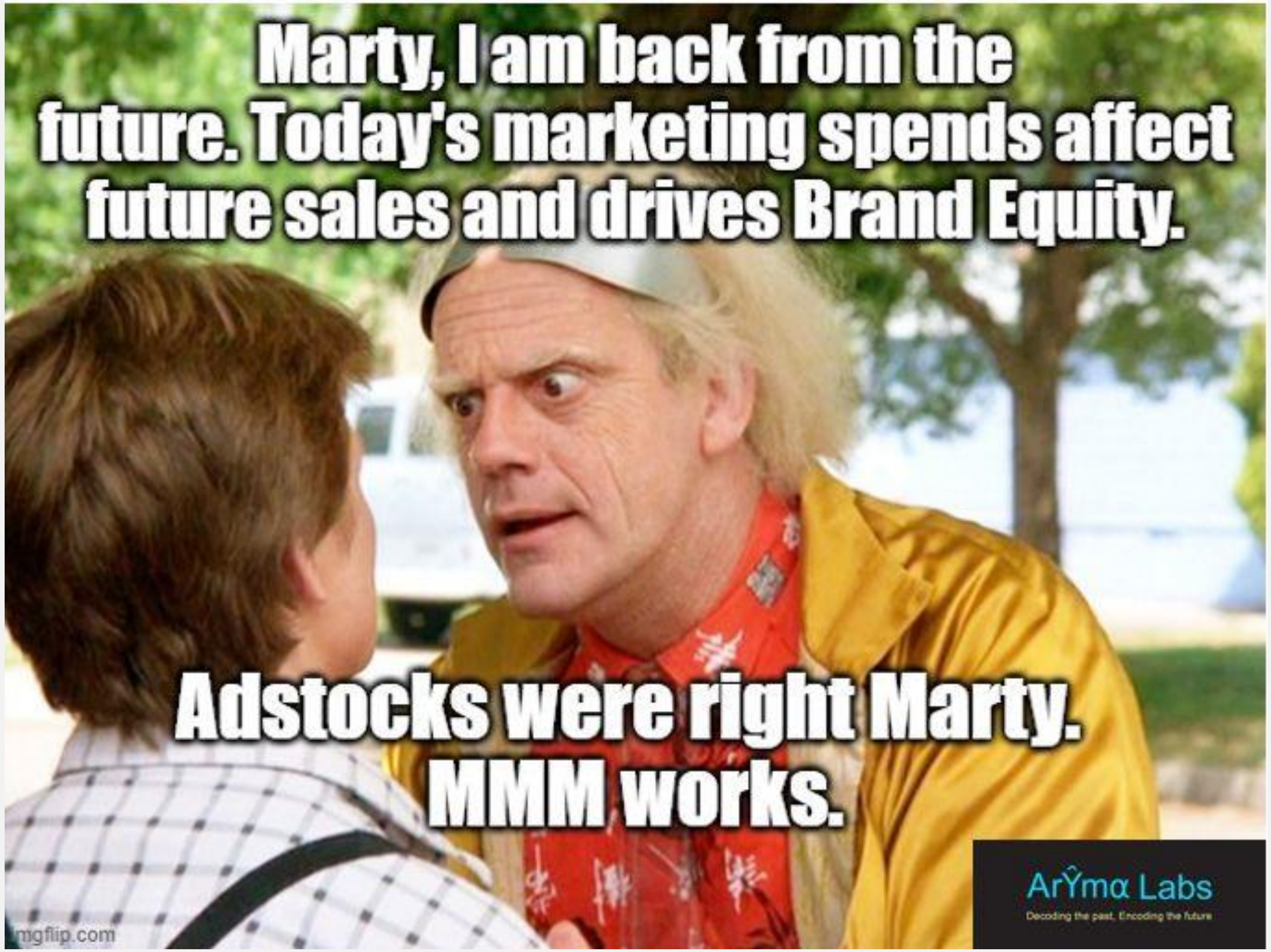
What is Adstock?

ArŶmα Labs
Decoding the past, Encoding the future

- Adstock refers to the cumulative impact of past advertising on current sales.
- Adstock has two components: **Carryover effect** and **Diminishing Returns.**
- Carryover Effect: The impact of past advertising on the present sales. Also called the decay effect as the impact of the past advertisement decays over time.
- Diminishing Returns: the impact of advertisement starts diminishing over time after a certain extent.
- Adstock transformation helps to capture the non-linear relationship of ads with the sales.



What is Adstock?



What is Adstock? - Seminal Work

ArŶmα Labs
Decoding the past, Encoding the future

One way TV advertisements work

Simon Broadbent
Leo Burnett Ltd

Abstract

TV advertising affects consumers' awareness that a brand has been advertised. A model is proposed for the way these two measures are related. It includes both response and decay or forgetting. It separates advertising spend from advertisement content.

The model has been used on a number of brands in different markets and gives reasonable fits. It illuminates their strategies and efficiencies, the results of changing campaigns and advertisements, the effect of other individual brands and of competition generally. It also gives insights into the general effect of advertising over time.

The analyses reported here investigate how certain survey measures of consumers' awareness are related to TV advertising.

These measures come from regular tracking studies or from a panel; they include awareness that the brand has been advertised recently, either spontaneous or prompted, and may require proof of recall. For convenience we call all such measures simply 'awareness'.

The main objectives of such analyses are:
(1) to separate the effect of advertising *spend* from that of the advertisements' *content*, and so to evaluate the creative work,
(2) to provide information on the response and decay of advertising's effects, as measured by instrusiveness and memorability, so helping in scheduling decisions.

There is also the objective of learning about 'how advertisements work'. That subject is really concerned with sales effectiveness, but we then come up against the well-known difficulty that advertising is usually a relatively small factor in sales results. It turns out that the mechanisms applicable in this simpler case suggest what may happen in the more complex one. We therefore get insights from this work into more important issues.

It is assumed from now on that advertising's objective is to keep our brand's advertising awareness high, compared with other products in the same group. Advertising has other objectives and may have sales efficiency without creating advertising awareness but this point is not pursued here.

Download This Paper | Open PDF in Browser | ☆ Add Paper to My Library

Understanding Advertising Adstock Transformations

7 Pages • Posted: 16 Aug 2006

Joy V. Joseph
Syneractiv

Date Written: May 15, 2006

Abstract

Advertising effectiveness and Return on Investment (ROI) are typically measured through econometric models that measure the impact of varying levels of advertising Gross Ratings Points (GRPs) on sales or on purchase decision and choice. TV advertising has both dynamic and diminishing returns effects on sales, different models capture these dynamic and nonlinear effects differently. This paper focuses on reviewing the econometric rationale behind the popularized Adstock transformation model that allows the inclusion of lagged and non-linear effects in linear models based on aggregate data.

Keywords: Advertising, Adstock Model, Non-linear transformation, Marketing-Mix

JEL Classification: M37

Suggested Citation:

Joseph, Joy V., Understanding Advertising Adstock Transformations (May 15, 2006). Available at SSRN: https://ssrn.com/abstract=924128 or http://dx.doi.org/10.2139/ssrn.924128

Carryover or Decay Effect

ArŶmα Labs
Decoding the past, Encoding the future

- The general formula for Adstock transformation to capture the carryover or decay effect is:

$$Adstock_t = Raw\_spend_t + decay\_rate_t * Adstock_{t-1}$$

- Mainly there are three types of Adstocks:
  - **Geometric**: Decay rate (theta) is fixed. Theta = 0.5 means 50% of the ads in the previous period is carried over to present period.
  - **Weibull CDF**: The decay rate at time t is determined by the Cumulative Distribution Function of Weibull Distribution with given shape and scale parameters.
  - **Weibull PDF**: The decay rate at time t is determined by the Density Function of Weibull Distribution with given shape and scale parameters. This method also incorporates lagged effect.

Visualizing the Carryover Effect

ArŶmα Labs
Decoding the past, Encoding the future

https://www.linkedin.com/posts/ridhima-kumar7_visualizing-the-media-carryover-effect-i-activity-7213797490209218560-GEI5?utm_source=share&utm_medium=member_desktop

Geometric Adstock

ArŶmα Labs
Decoding the past, Encoding the future

```
adstock_geometric <- function(x, theta) {
  stopifnot(length(theta) == 1)
  if (length(x) > 1) {
    x_decayed <- c(x[1], rep(0, length(x) - 1))
    for (xi in 2:length(x_decayed)) {
      x_decayed[xi] <- x[xi] + theta * x_decayed[xi - 1]
    }
    thetaVecCum <- theta
    for (t in 2:length(x)) {
      thetaVecCum[t] <- thetaVecCum[t - 1] * theta
    } # plot(thetaVecCum)
  } else {
    x_decayed <- x
    thetaVecCum <- theta
  }
  inflation_total <- sum(x_decayed) / sum(x)
  return(list(x = x, x_decayed = x_decayed, thetaVecCum = thetaVecCum, inflation_total = inflation_total))
}
```

- Here, x is the raw spends and theta is the decay parameter, the decayed series is calculated by:

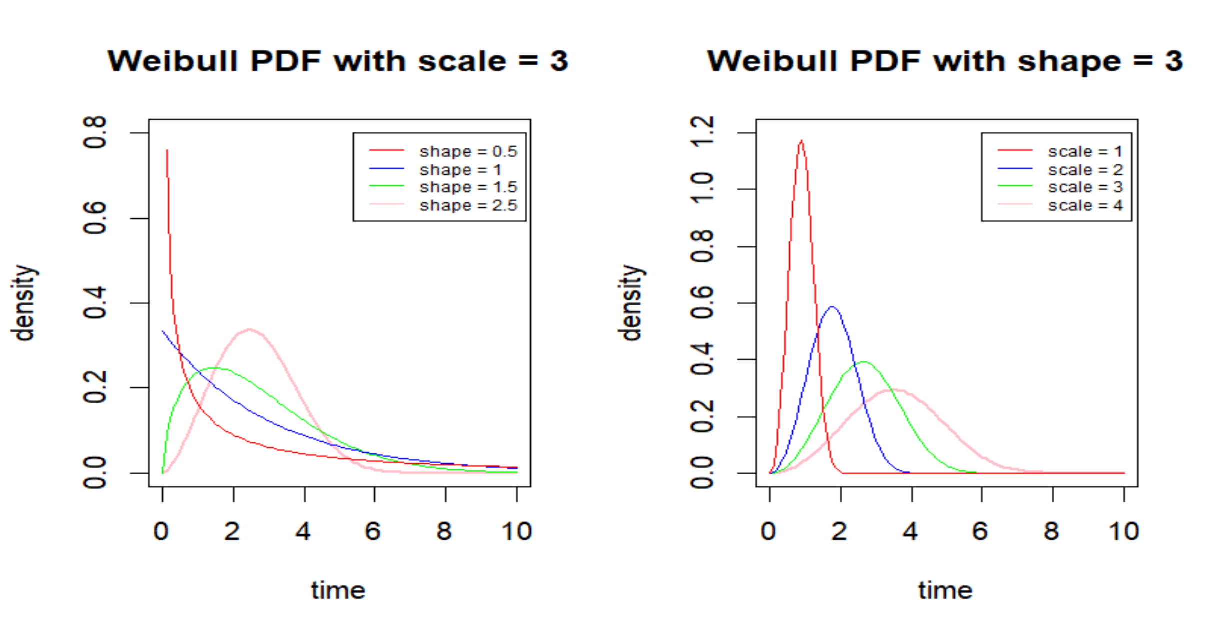
$$x\_decayed_t = x_t + \theta * x\_decayed_{t-1} = \sum_{i=0}^{t} \theta^i \, x_{t-i}$$

- The vector 'thetaVecCum' captures the decaying impact of the ad over time. Starting at time 0, the impact is highest with a value of 1, then decreases progressively as time advances, taking values of θ at time 1, θ^2 at time 2, and so on, reflecting the decaying effect over time.

- The recommended ranges for theta are:
  - TV: 0.3 – 0.8
  - OOH/Print/Radio: 0.1 – 0.4
  - Digital: 0 – 0.3

Weibull Distribution

ArŶmα Labs
Decoding the past, Encoding the future



- Two parameter Weibull distribution has the density function:

$$f(t) = \left(\frac{\beta}{\alpha}\right)\left(\frac{t}{\alpha}\right)^{\beta-1} \exp\left\{-\left(\frac{t}{\alpha}\right)^{\beta}\right\}; t \geq 0, \alpha, \beta > 0$$

- The cumulative distribution function of Weibull distribution is:

$$F(t) = 1 - \exp\left\{-\left(\frac{t}{\alpha}\right)^{\beta}\right\}$$

- Where, α is the scale parameter and β is the shape parameter.

Weibull CDF Adstock

ArŶmα Labs
Decoding the past, Encoding the future

```r
adstock_weibull <- function(x, shape, scale, windlen = length(x), type = "cdf") {
  stopifnot(length(shape) == 1)
  stopifnot(length(scale) == 1)
  if (length(x) > 1) {
    check_opts(tolower(type), c("cdf", "pdf"))
    x_bin <- 1:windlen
    scaleTrans <- round(quantile(1:windlen, scale), 0)
    if (shape == 0 | scale == 0) {
      x_decayed <- x
      thetaVecCum <- thetaVec <- rep(0, windlen)
      x_imme <- NULL
    } else {
      if ("cdf" %in% tolower(type)) {
        thetaVec <- c(1, 1 - pweibull(head(x_bin, -1), shape = shape, scale = scaleTrans)) # plot(thetaVec)
        thetaVecCum <- cumprod(thetaVec) # plot(thetaVecCum)
      } else if ("pdf" %in% tolower(type)) {
        thetaVecCum <- .normalize(dweibull(x_bin, shape = shape, scale = scaleTrans)) # plot(thetaVecCum)
      }
      x_decayed <- mapply(function(x_val, x_pos) {
        x.vec <- c(rep(0, x_pos - 1), rep(x_val, windlen - x_pos + 1))
        thetaVecCumLag <- lag(thetaVecCum, x_pos - 1, default = 0)
        x.prod <- x.vec * thetaVecCumLag
        return(x.prod)
      }, x_val = x, x_pos = seq_along(x))
      x_imme <- diag(x_decayed)
      x_decayed <- rowSums(x_decayed)[seq_along(x)]
    }
  } else {
    x_decayed <- x_imme <- x
    thetaVecCum <- 1
  }
```

Weibull CDF Adstock

$$Adstock_t = \sum_{j=0}^{t} \tilde{S}(j) \times Raw\_spend_{t-j} \ ; \ \tilde{S}(j) = \prod_{i=0}^{j} S(i)$$

- $S(i) = 1 - F(i)$ and $F(i)$ values are generated by the "pweibull" function. $S(i)$ is called the survival function.
- The recommended shape and scale parameter ranges are:
  - Shape: 0 – 2
  - Scale: 0 – 0.1
- Scale parameter α is transformed to ScaleTrans, which is the α^{th} quantile of time indices, rounded off to nearest integer.
- "thetaVecCum" stores the cumulative product of the survival probabilities ($\tilde{S}(j)$ in the formula) which are our time varying decay rates to compute the adstock series.

Weibull PDF Adstock

ArŶmα Labs
Decoding the past, Encoding the future

```
adstock_weibull <- function(x, shape, scale, windlen = length(x), type = "cdf") {
  stopifnot(length(shape) == 1)
  stopifnot(length(scale) == 1)
  if (length(x) > 1) {
    check_opts(tolower(type), c("cdf", "pdf"))
    x_bin <- 1:windlen
    scaleTrans <- round(quantile(1:windlen, scale), 0)
    if (shape == 0 | scale == 0) {
      x_decayed <- x
      thetaVecCum <- thetaVec <- rep(0, windlen)
      x_imme <- NULL
    } else {
      if ("cdf" %in% tolower(type)) {
        thetaVec <- c(1, 1 - pweibull(head(x_bin, -1), shape = shape, scale = scaleTrans)) # plot(thetaVec)
        thetaVecCum <- cumprod(thetaVec) # plot(thetaVecCum)
      } else if ("pdf" %in% tolower(type)) {
        thetaVecCum <- .normalize(dweibull(x_bin, shape = shape, scale = scaleTrans)) # plot(thetaVecCum)
      }
      x_decayed <- mapply(function(x_val, x_pos) {
        x.vec <- c(rep(0, x_pos - 1), rep(x_val, windlen - x_pos + 1))
        thetaVecCumLag <- lag(thetaVecCum, x_pos - 1, default = 0)
        x.prod <- x.vec * thetaVecCumLag
        return(x.prod)
      }, x_val = x, x_pos = seq_along(x))
      x_imme <- diag(x_decayed)
      x_decayed <- rowSums(x_decayed)[seq_along(x)]
    }
  } else {
    x_decayed <- x_imme <- x
    thetaVecCum <- 1
  }
```

Weibull PDF Adstock

$$Adstock_t = \sum_{j=0}^{t} \tilde{f}(j+1) \times Raw\_spend_{t-j-1}$$

- $\tilde{f}(j)$ is the density function generated using "dweibull" function and normalized as below:

$$\tilde{f}(t) = \frac{f(t) - \min\{f(t)\}}{\max\{f(t)\} - \min\{f(t)\}}$$

- "thetaVecCum" stores these normalized density values which are our time varying decay rates.
- The recommended shape parameter ranges to capture various cases are:
  - 0-10 in general
  - >2 to capture lagged effects
  - 0-1 for no lagged effect
  - 1 for exponential decay
- Scale parameter ranges from 0 to 0.1 and transformed into ScaleTrans as in Weibull CDF adstock.

Summary of Adstock Transformations

ArŶmα Labs
Decoding the past, Encoding the future

Geometric	Weibull CDF	Weibull PDF
One hyperparameter	Two hyperparameters	Two hyperparameters
Fixed decay rate	Flexible decay rate	Flexible decay rate
Does not capture lagging effect	Does not capture lagging effect	Captures lagging effect
Computationally easy	Computationally difficult	Computationally difficult

Diminishing Returns using Power Transformation

Power Transformation

$$saturated_t = (raw\_media_t)^n + \theta * saturated_{t-1}$$

- Where, n is the diminishing parameter and θ is the decay parameter.
- The ranges of both n and θ are 0.1 – 0.9.
- Used together with the Geometric adstock.

Diminishing Returns using Hill Transformation

ArŶmα Labs
Decoding the past, Encoding the future

```r
saturation_hill <- function(x, alpha, gamma, x_marginal = NULL) {
  stopifnot(length(alpha) == 1)
  stopifnot(length(gamma) == 1)
  inflexion <- c(range(x) %*% c(1 - gamma, gamma)) # linear interpolation by dot product
  if (is.null(x_marginal)) {
    x_scurve <- x**alpha / (x**alpha + inflexion**alpha) # plot(x_scurve) summary(x_scurve)
  } else {
    x_scurve <- x_marginal**alpha / (x_marginal**alpha + inflexion**alpha)
  }
  return(x_scurve)
}
```

Hill Transformation

$$saturated_t = \frac{(adstocked_t)^{\alpha}}{(adstocked_t)^{\alpha} + (inflexion)^{\alpha}}$$

- α controls the shape of the curve. $\alpha < 1$ gives a C-shaped curve, whereas $\alpha > 1$ gives an S-shaped curve.
- γ controls the inflexion point and the relationship between them is given by:

$$inflexion = (1 - \gamma) * \min(x\_decayed) + \gamma * \max(x\_decayed)$$

- The recommended ranges for α and γ are:
  - alpha: 0.5 – 3
  - gamma: 0.3 - 1

Hands-on Demo using R Shiny App

ArYma Labs
Decoding the past, Encoding the future

ArŶmα Labs

Decoding the past, Encoding the future

Feature Selection in MMM

Feature Selection in MMM

- In MMM, identifying and including the features that best predict or explain the dependent variable is key to building an effective model.

- Traditional Feature Selection Approach relies on correlation analysis and domain expertise. Variables are first shortlisted based on their correlation with the KPI, followed by domain experts refining the selection using a correlation threshold and domain knowledge.

- Correlation measures only linear relationships with the KPI, which can lead to the exclusion of important variables that have significant non-linear impacts and contribute valuable information.

Information Theoretic Approaches

ArŶmα Labs
Decoding the past, Encoding the future

Transfer Entropy (TE):

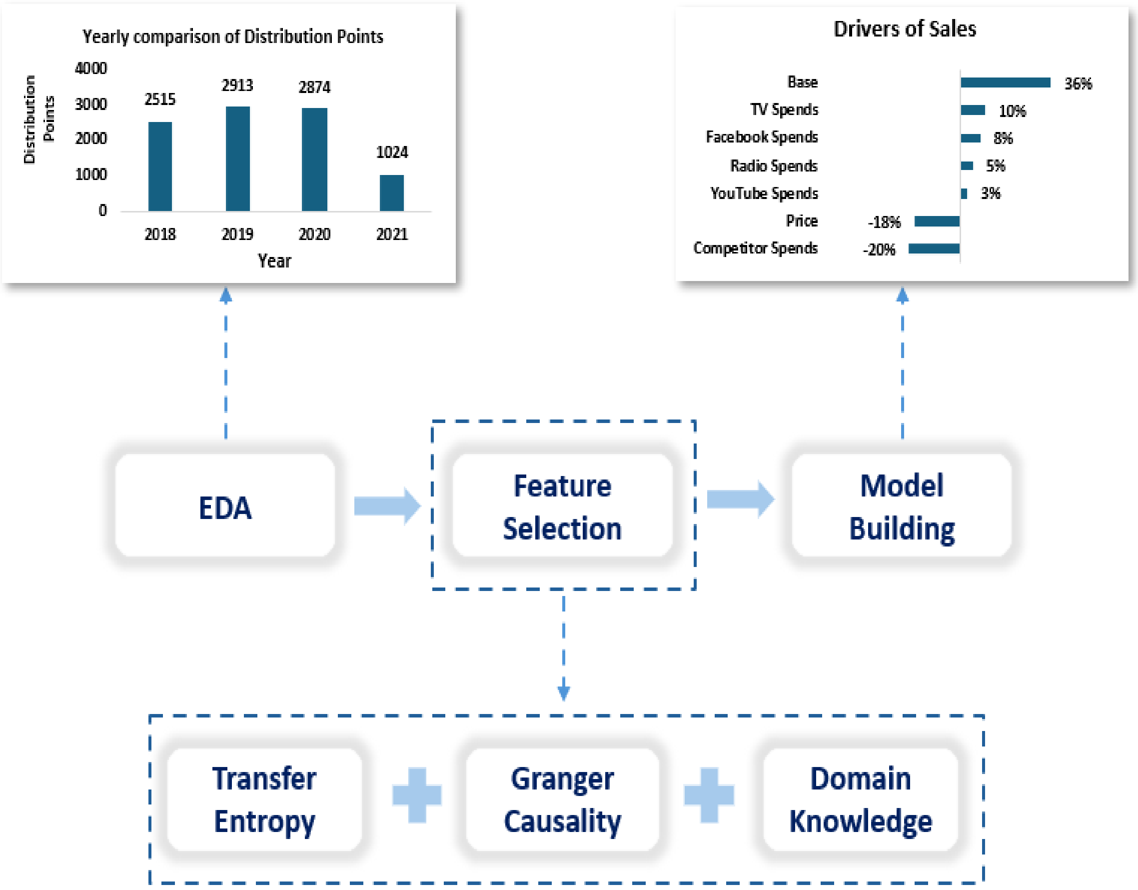
- Transfer entropy is a non-parametric method to measure the amount of directed (time-asymmetric) transfer of information between two random processes.
- Transfer entropy from a process X to another process Y is the amount of uncertainty reduced in future values of Y by knowing the past values of X given past values of Y.
- Transfer Entropy captures linear as well as non-linear relationships.

Granger Causality (GC):

- Granger causality test is a statistical hypothesis test for determining that whether one time series is useful in forecasting another time series.
- If we have two time series variables X and Y, then X is said to “Granger cause” Y, if predictions of Y based on the past values of X and past values of Y are better than the predictions of Y based only on the past values of Y.
- Therefore, we see if X contains some useful information for the prediction of Y which is not present in the past values of Y.

How Aryma Labs Leverages the Trifecta Approach?

ArŶmα Labs
Decoding the past, Encoding the future



The Trifecta Approach is done using the following steps:

- The variables are divided into media and non-media categories. Media variables undergo adstock transformation using our proprietary techniques, while non-media variables remain untransformed.
- TE and GC values are computed for both transformed and untransformed variables with respect to the KPI.
- Based on the calculated TE and GC results, two separate lists of candidate variables are created. A third list is simultaneously prepared by domain experts based on their insights and knowledge.
- The three lists of variables are put into the trifecta algorithm that we have devised. This algorithm is based on iterative weighting of variables and the weights will help to choose the variables which carries maximum information about the KPI.

ArŶmα Labs

Decoding the past, Encoding the future

Building MMM Models

What are we modelling ?

Is there one true MMM model out there?

ArŶmα Labs
Decoding the past, Encoding the future

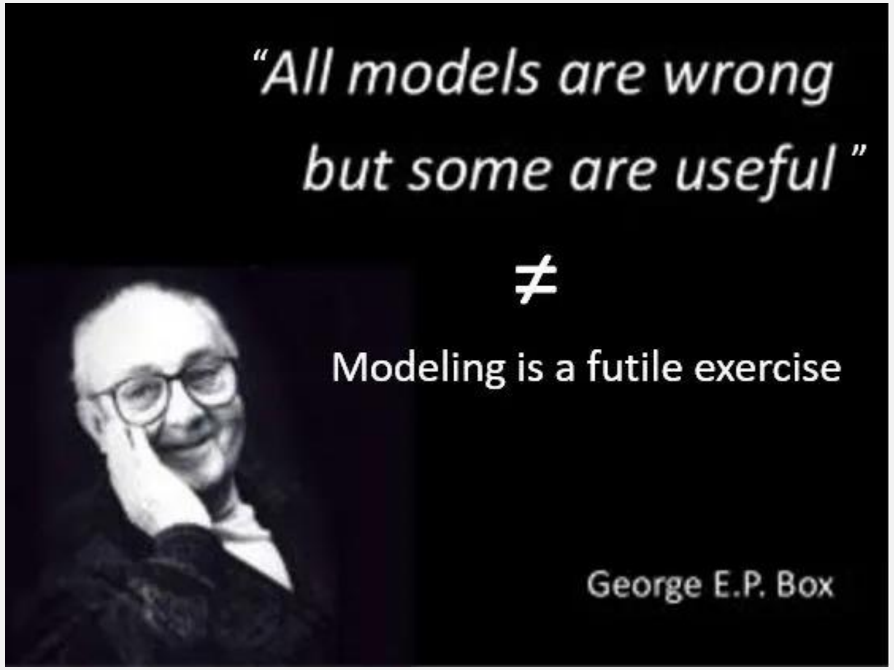
"All models are wrong, some are useful" is an aphorism (meaning it is a concise expression of general truth). But the aphorism in this case leads to misinterpretation.

Firstly, it is important to understand what modeling is.
The purpose of modeling is to provide an abstraction of real process. Basically, a good approximation of reality.

Anybody who mistakes the abstraction for the real, commits the Fallacy of Reification (yup, one more fallacy to add to the list of all fallacies which we data scientists/statisticians commit).

In an exact sense, a map is also wrong because it does not provide 1:1 mapping of the real world.

So, George Box's phrase should be construed the same way as a map is considered wrong because it does not represent the real world.



What are we modelling ?
Is there one true MMM model out there?

ArŶmα Labs
Decoding the past, Encoding the future

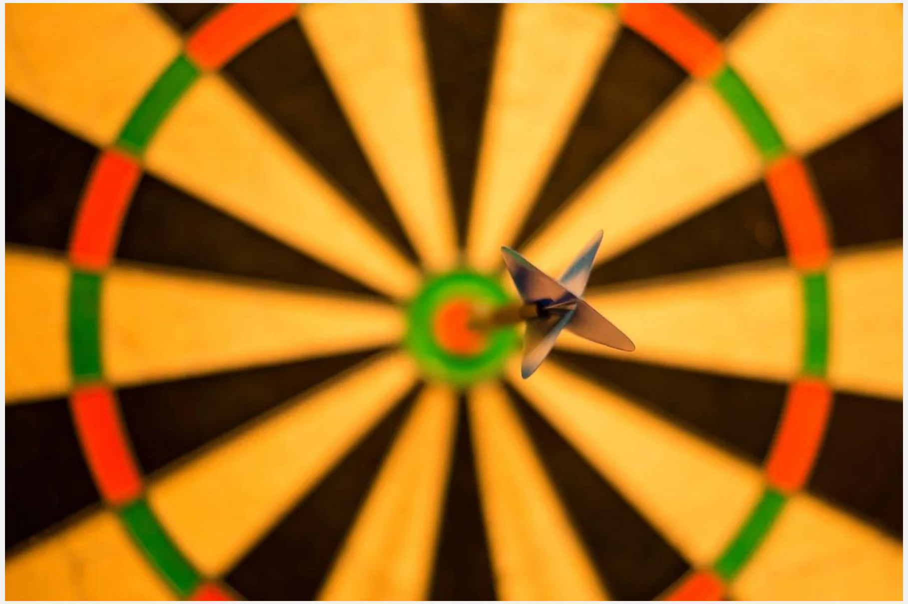
In MMM we deal with historical data, meaning the sales (or any KPI) has already been realized. Now the million-dollar (literally) question is what lead to this sales? What combination of factors led to the sales.

Now because this is all in the past, there is only one set of combination which could have resulted in the sales. The job of a MMM vendor is to find out this combination of factors that led to the observed sales. So technically it is honing on the truth.

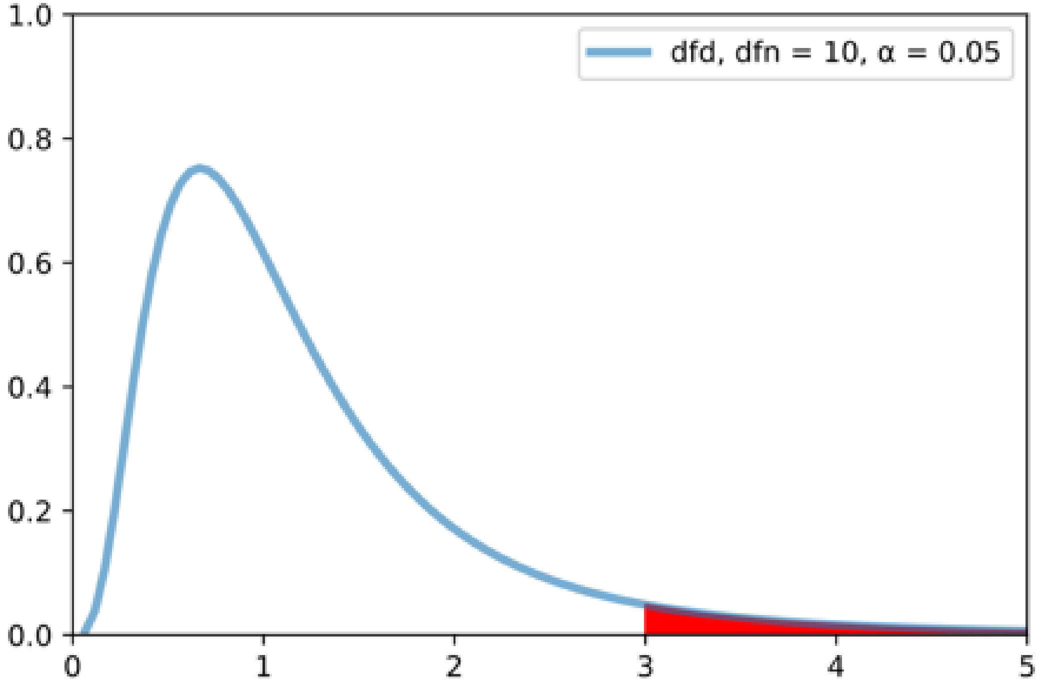
There are no multiple ways through which sales could have been gotten. This part is the scenario planning and budget planning exercise where once we have a ground truth understanding of what moved the needle of sales, we then try to see how we can tweak the combinations of marketing inputs to yield a better sales number.

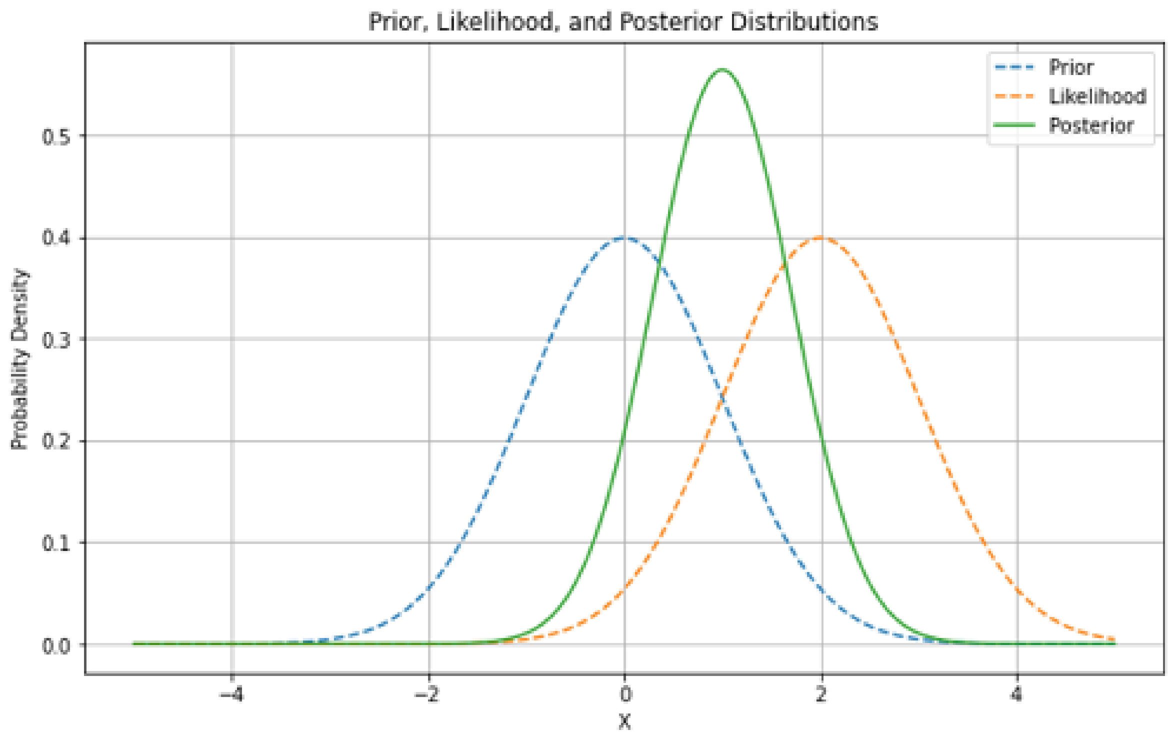
In the latter, many different models are possible.

To summarize:
•There is one true MMM model. Vendors should help clients find that model. We at Aryma Labs always do.
•In scenario planning / Budget optimization, there could be multiple models.



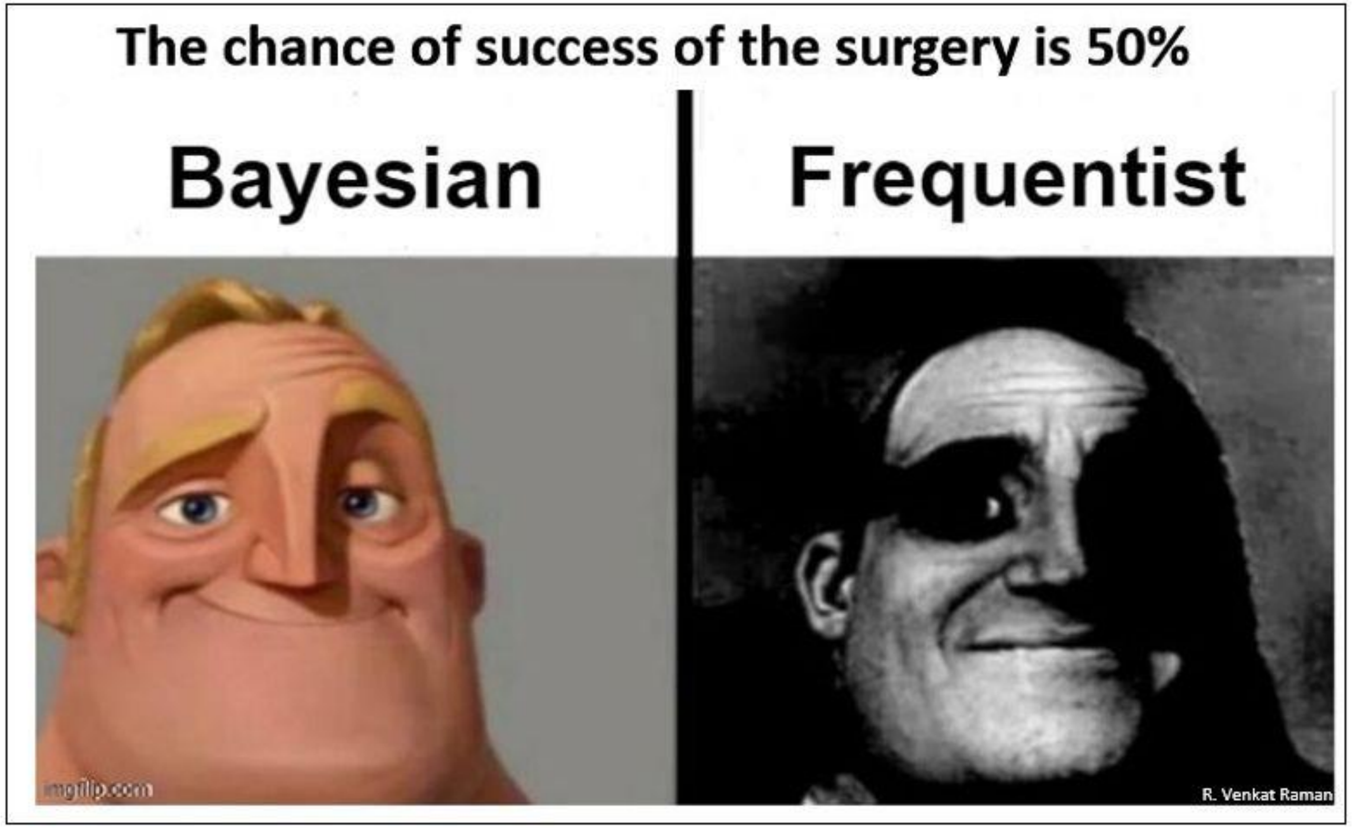
Types of MMM models – Frequentist vs Bayesian





Types of MMM models – Frequentist vs Bayesian

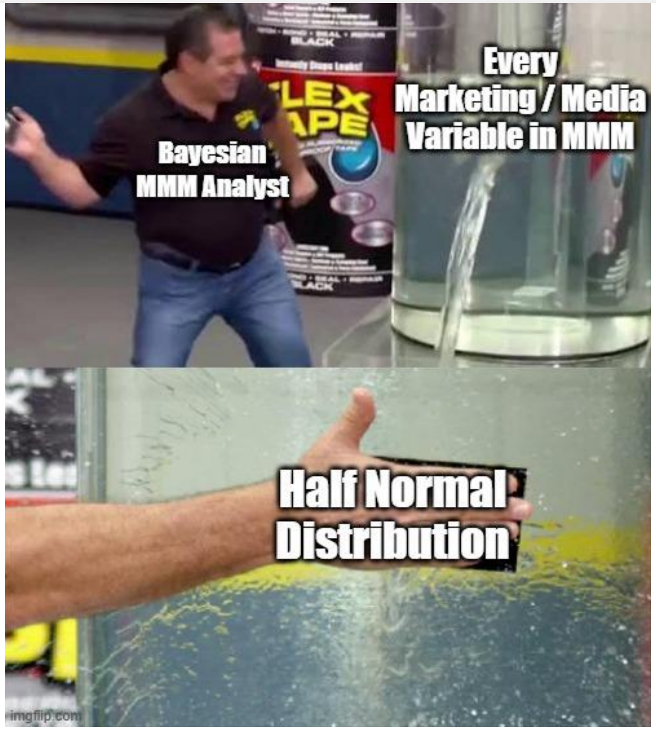
ArŶmα Labs
Decoding the past, Encoding the future

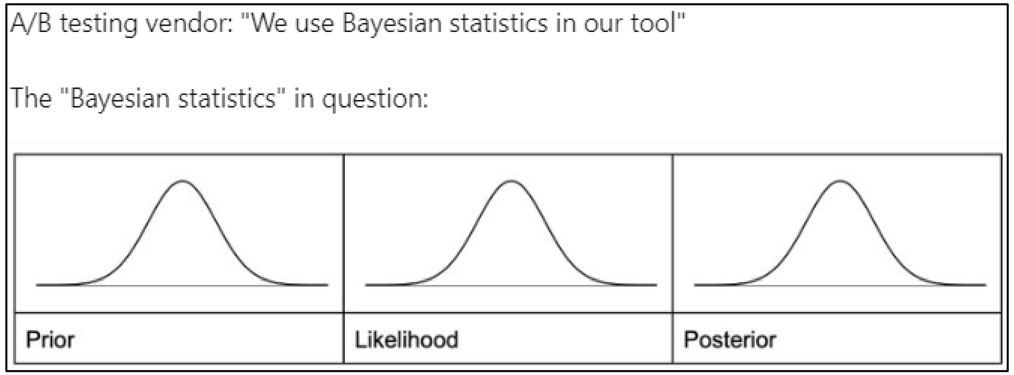


The Bayesian World - Everything is a probability distribution

But we humans are not good at encoding information through probability distributions

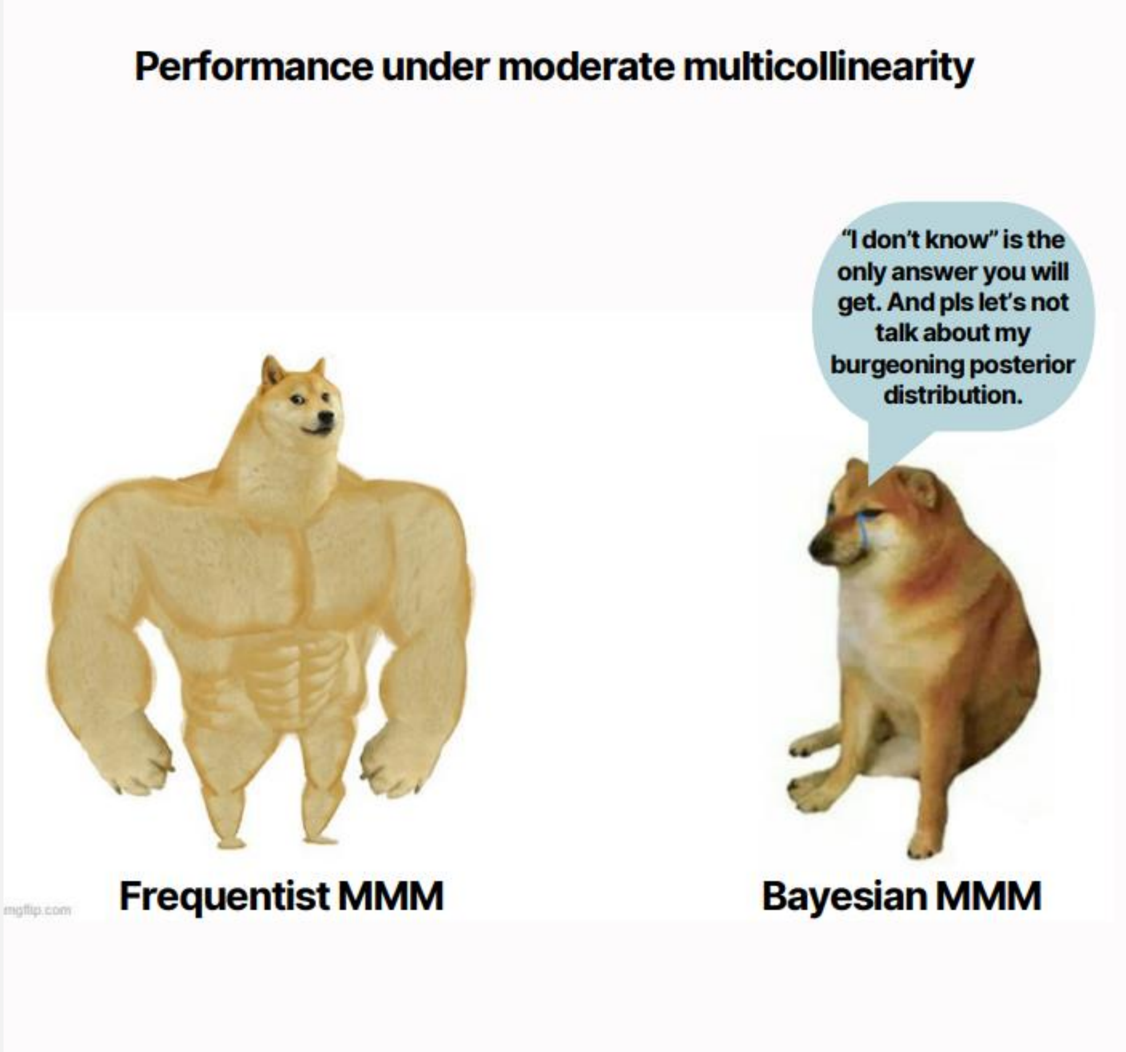
ArŶmα Labs
Decoding the past, Encoding the future

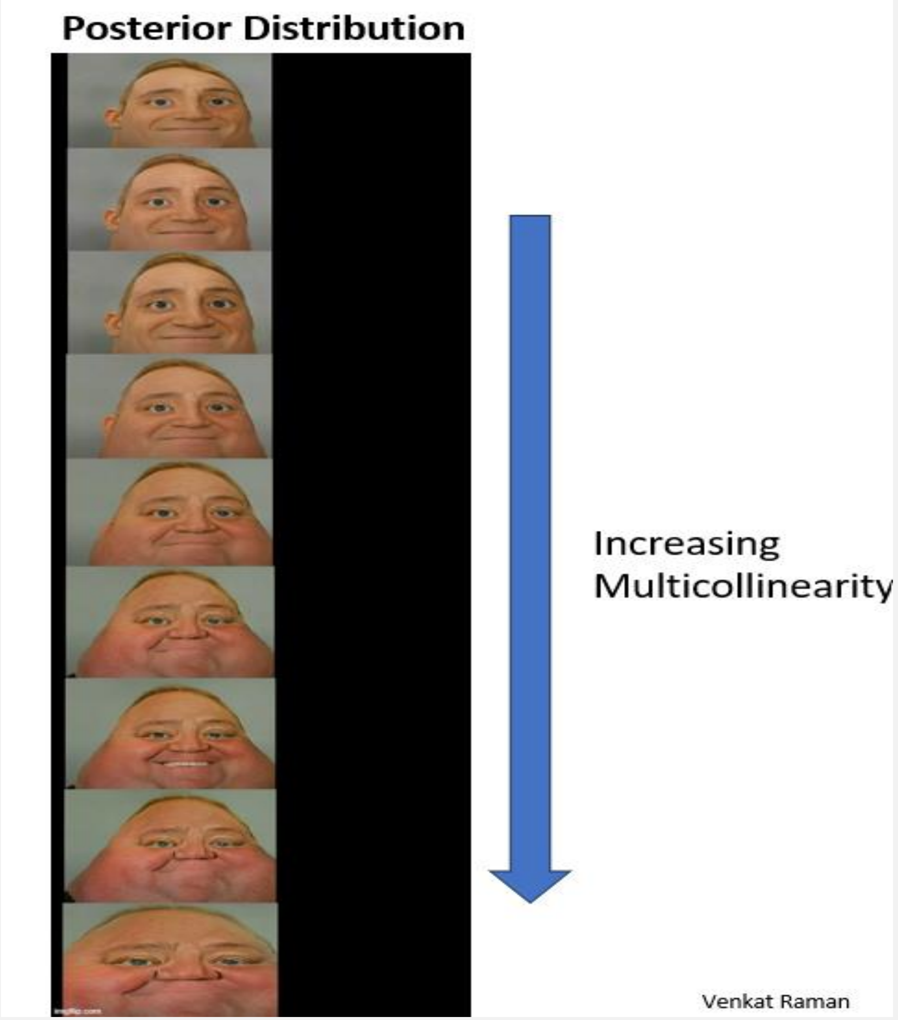




Why Frequentist MMM triumphs over Bayesian MMM?

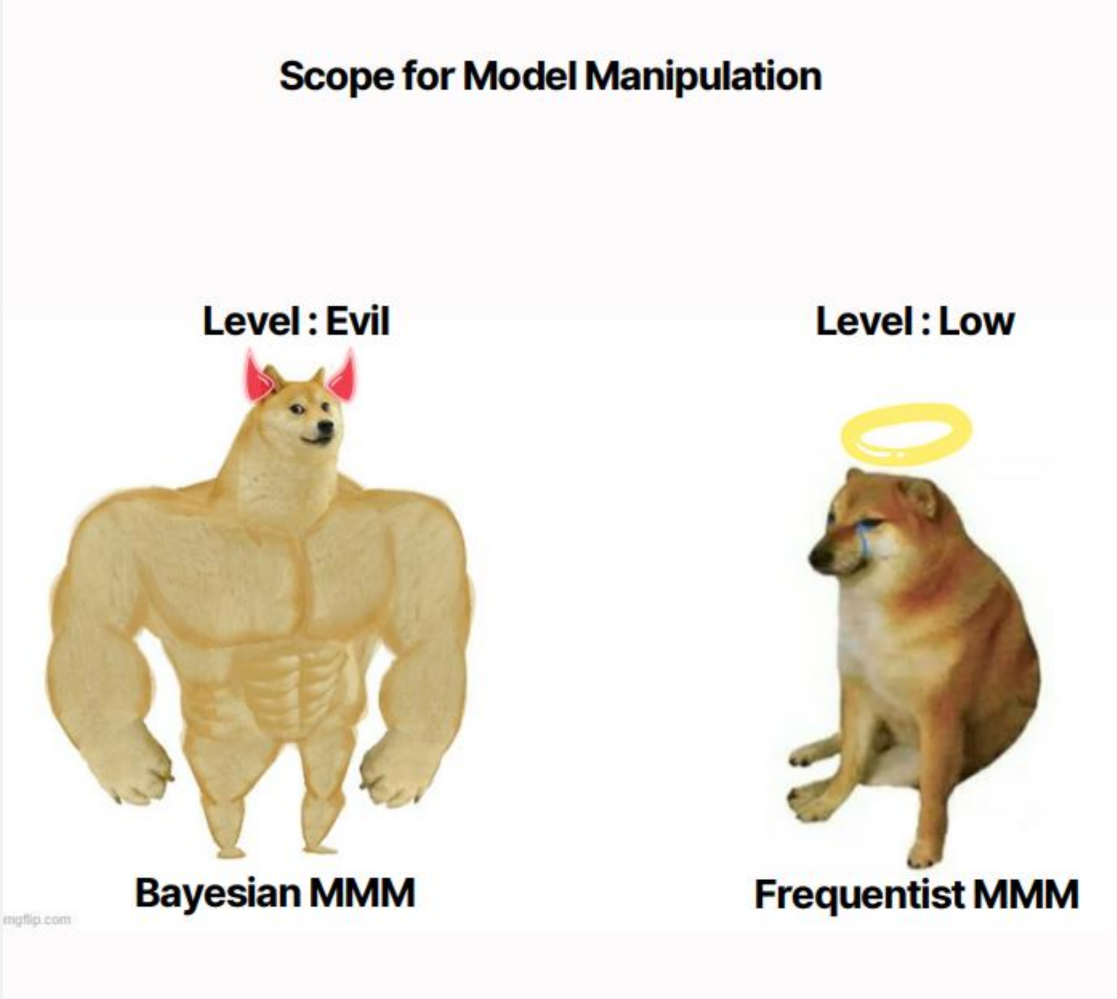
ArŶmα Labs
Decoding the past, Encoding the future

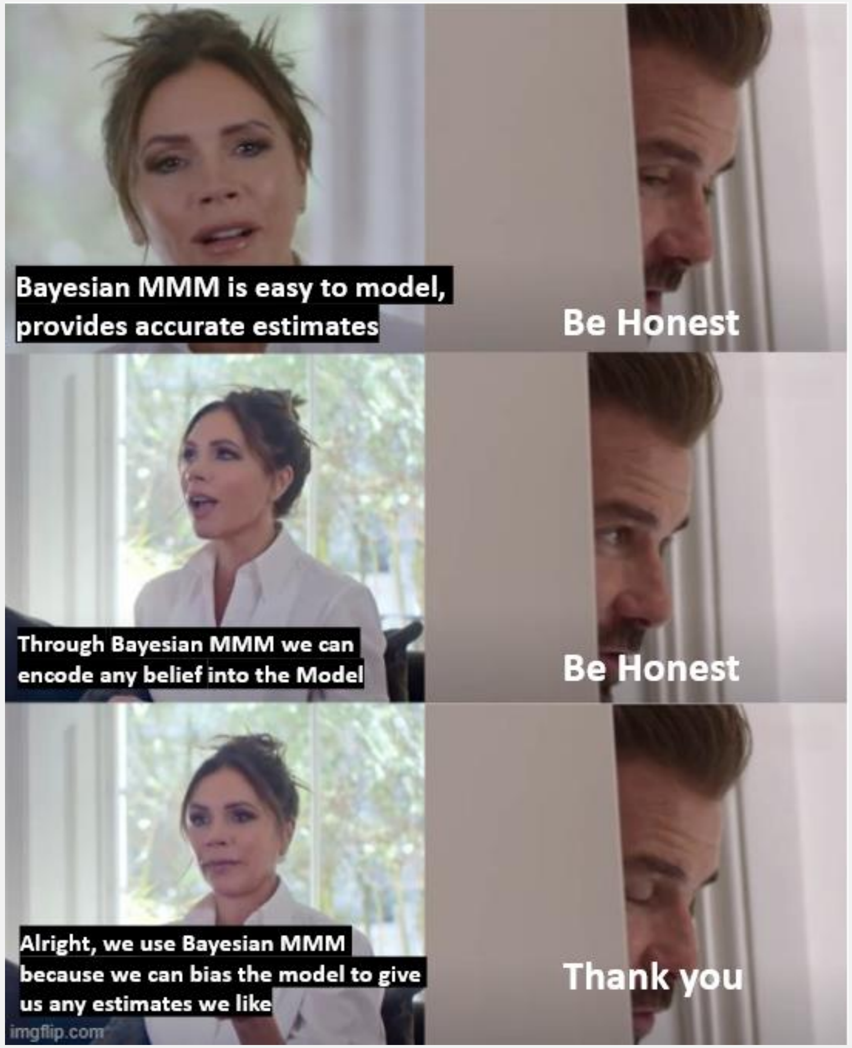




Why Frequentist MMM triumphs over Bayesian MMM?

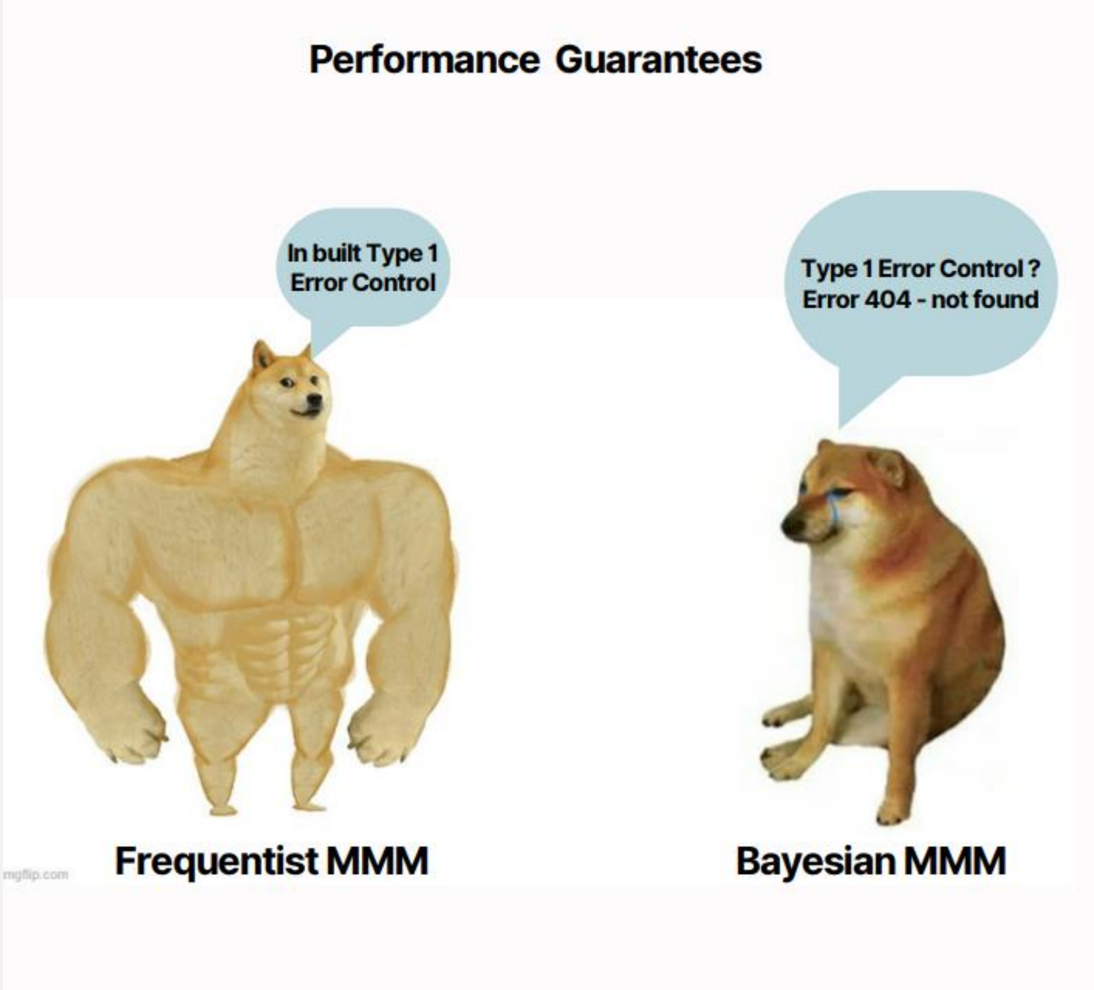
ArŶmα Labs
Decoding the past, Encoding the future



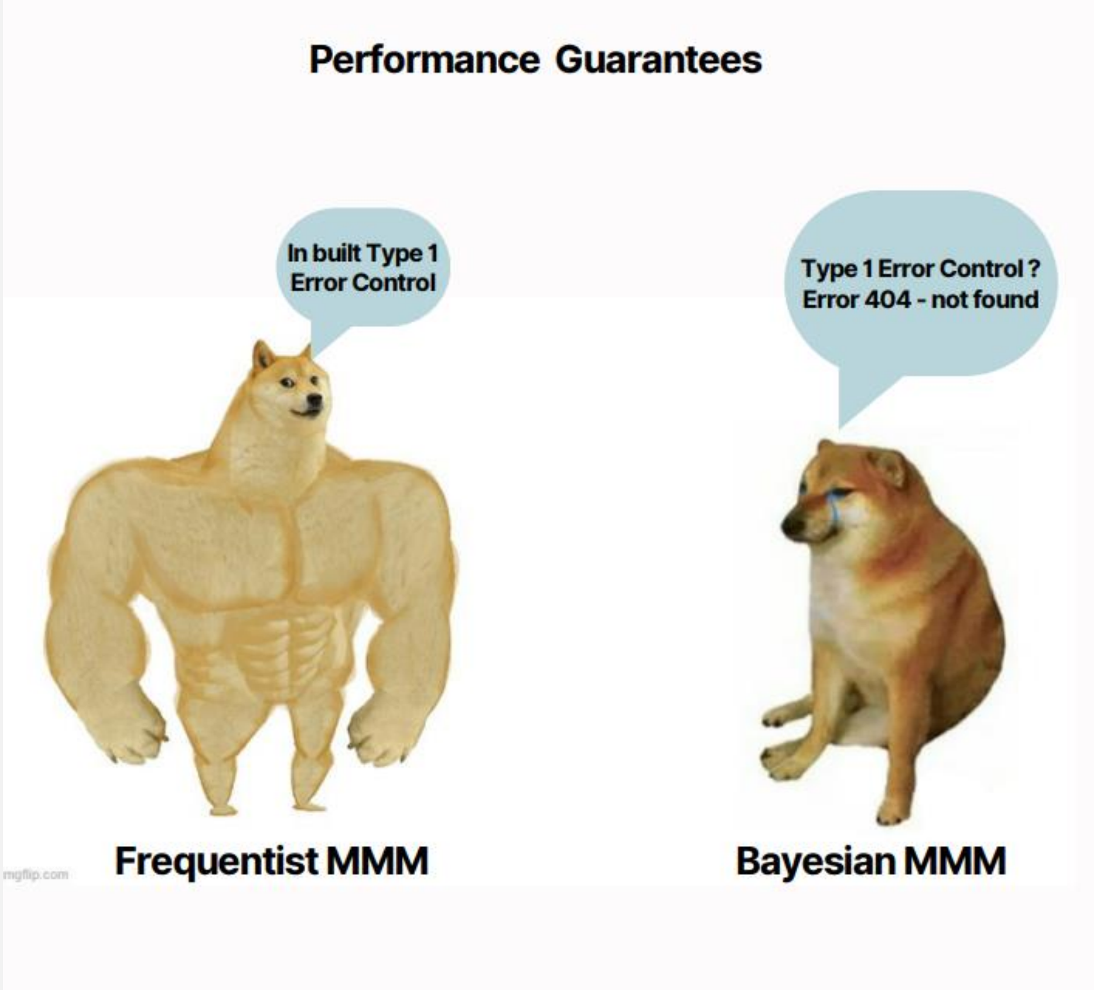


Why Frequentist MMM triumphs over Bayesian MMM?

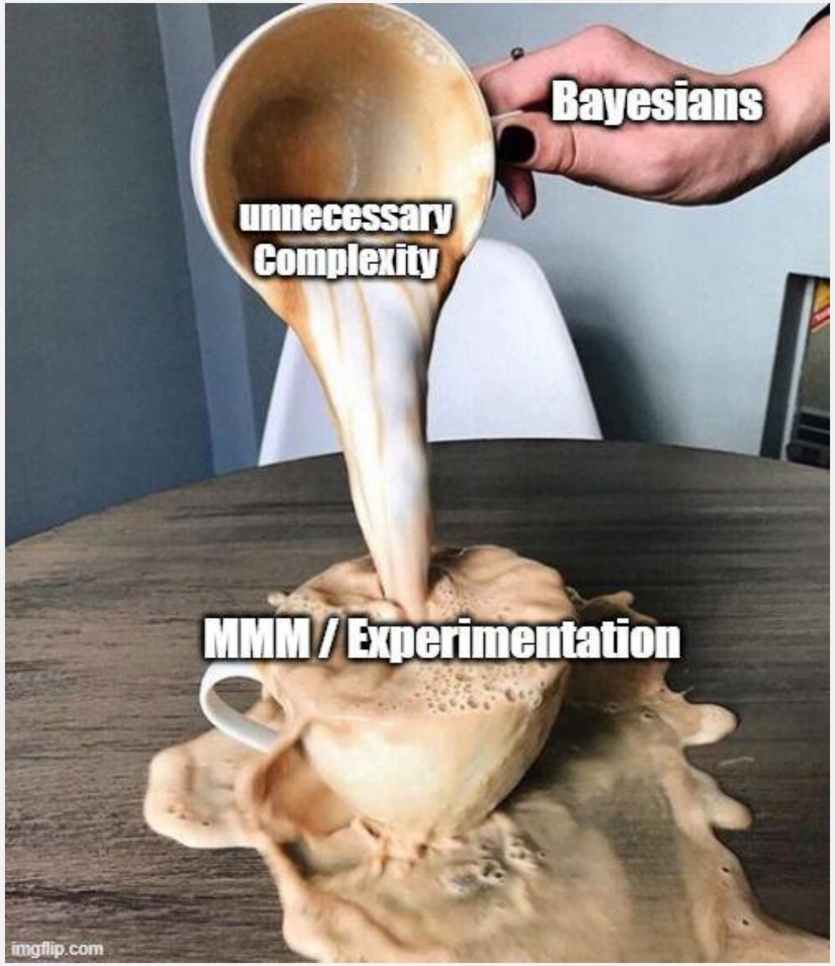
ArŶmα Labs
Decoding the past, Encoding the future



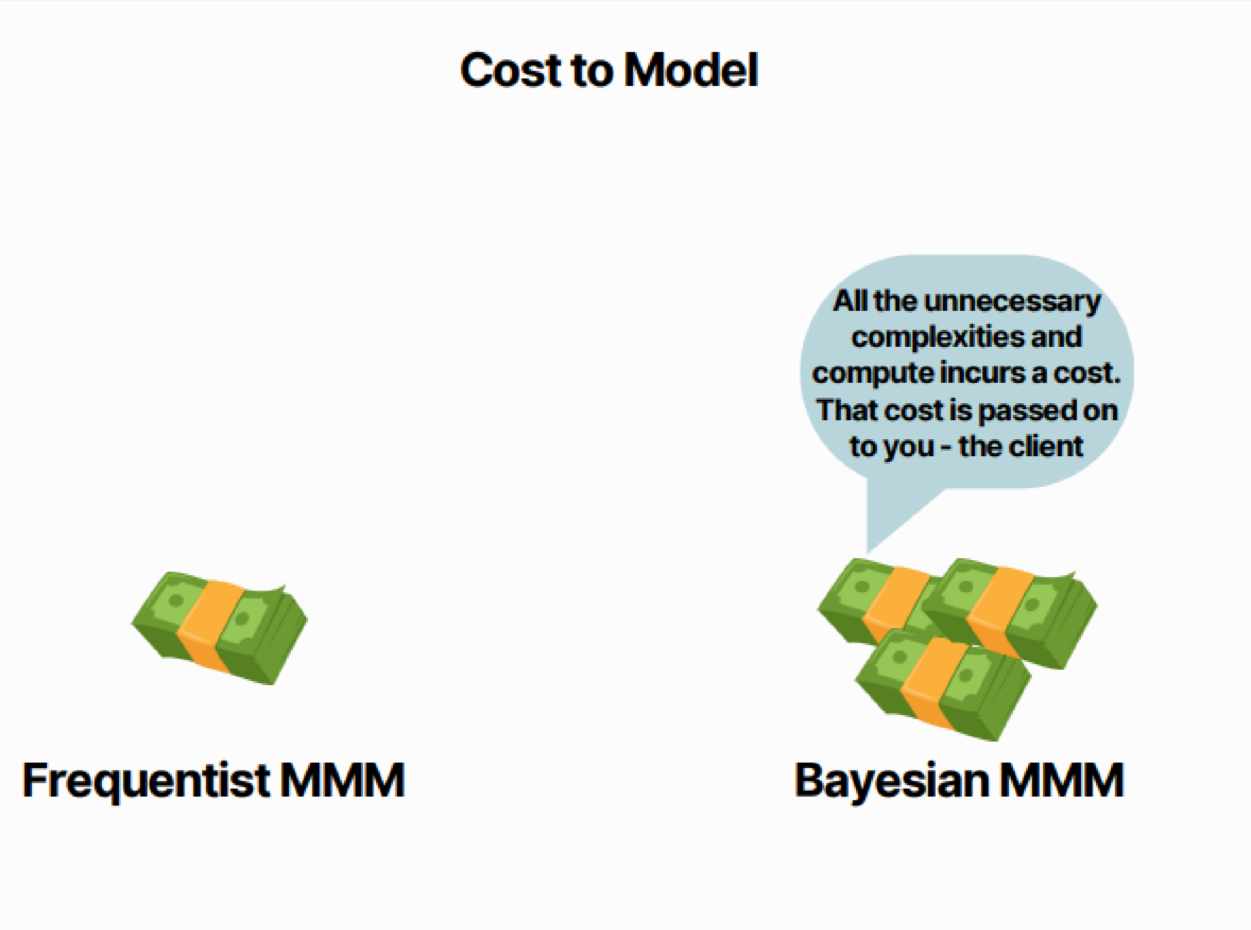
Why Frequentist MMM triumphs over Bayesian MMM?



Why Frequentist MMM triumphs over Bayesian MMM?



Why Frequentist MMM triumphs over Bayesian MMM?



Why Step-wise Regression is bad?

ArŶmα Labs
Decoding the past, Encoding the future

MMM is all about attribution.

1. It yields R-squared values that are badly biased to be high.

2. The F and chi-squared tests quoted next to each variable on the printout do not have the claimed distribution.

3. The method yields confidence intervals for effects and predicted values that are falsely narrow; see Altman and Andersen (1989).

4. It yields p-values that do not have the proper meaning, and the proper correction for them is a difficult problem.

5. It gives biased regression coefficients that need shrinkage (the coefficients for remaining variables are too large; see Tibshirani [1996]).

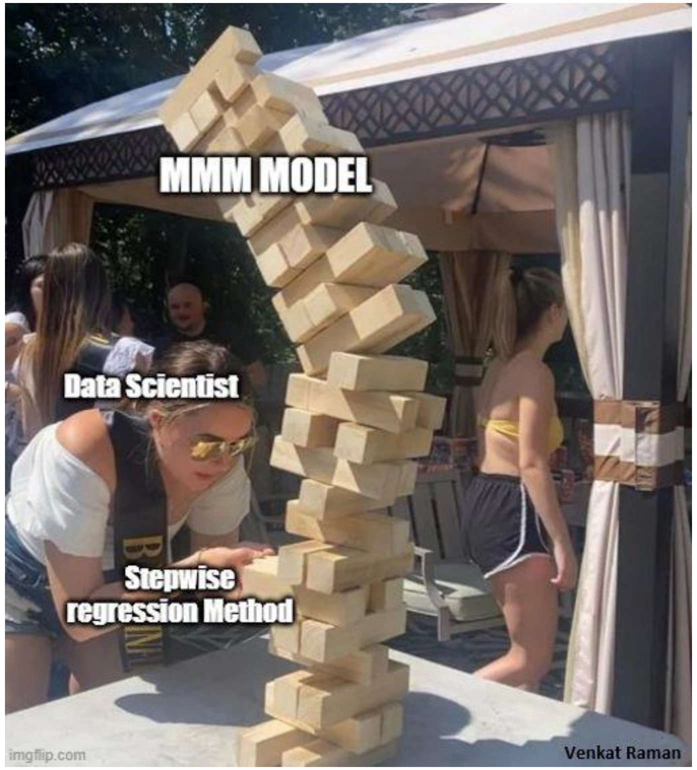
6. It has severe problems in the presence of collinearity.

7. It is based on methods (e.g., F tests for nested models) that were intended to be used to test prespecified hypotheses.

8. Increasing the sample size does not help very much; see Derksen and Keselman (1992).

Finally this one is my favourite

9. It allows us to not think about the problem.



Unpacking a Linear Regression output table

```
Call:
lm(formula = mpg ~ cyl + disp + am, data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-5.0863 -1.7831 -0.4842  1.5987  6.6358

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 32.91686    2.77914  11.844 2.03e-12 ***
cyl         -1.61822    0.69937  -2.314   0.0282 *
disp        -0.01559    0.01065  -1.463   0.1545
am           1.92873    1.33973   1.440   0.1611
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3 on 28 degrees of freedom
Multiple R-squared:  0.7761,	Adjusted R-squared:  0.7522
F-statistic: 32.36 on 3 and 28 DF,  p-value: 3.06e-09
```

Unpacking a Linear Regression output table – F test

```
Call:
lm(formula = mpg ~ cyl + disp + am, data = data)

Residuals:
    Min      1Q  Median      3Q     Max 
-5.0863 -1.7831 -0.4842  1.5987  6.6358 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 32.91686    2.77914  11.844 2.03e-12 ***
cyl         -1.61822    0.69937  -2.314   0.0282 *  
disp        -0.01559    0.01065  -1.463   0.1545    
am           1.92873    1.33973   1.440   0.1611    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3 on 28 degrees of freedom
Multiple R-squared:  0.7761,	Adjusted R-squared:  0.7522 
F-statistic: 32.36 on 3 and 28 DF,  p-value: 3.06e-09
```

ArŶmα Labs
Decoding the past, Encoding the future

Hands-on Session