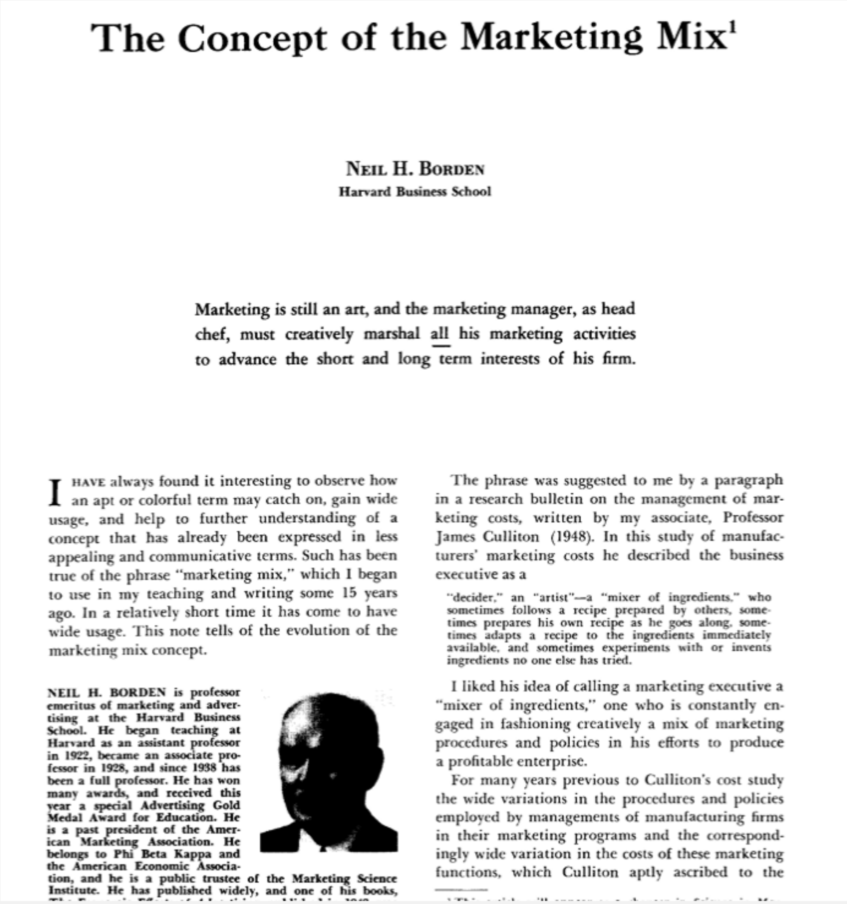
ArŶmα Labs

Decoding the past, Encoding the future

A – Z of MMM Workshop

Introduction to Marketing Mix Modeling (MMM)

ArŶmα Labs
Decoding the past, Encoding the future



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?

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

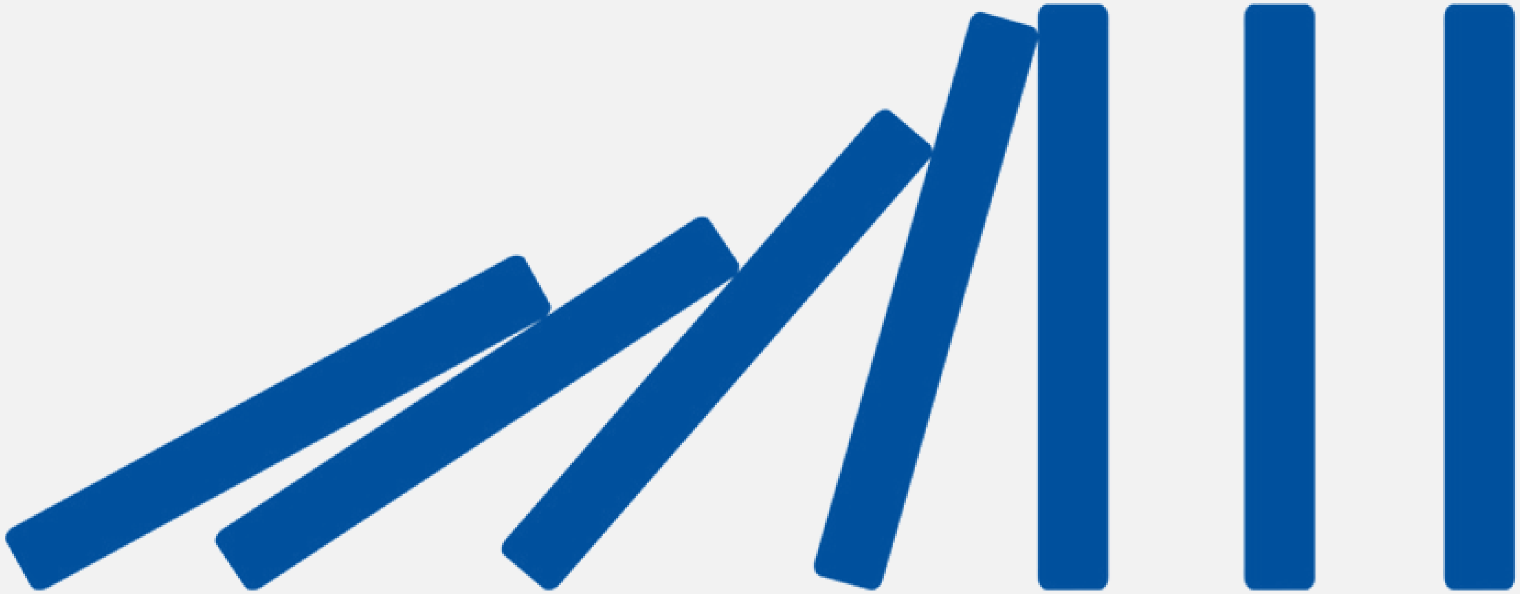
Competitor Price + Competitor Media Spends + Impressions, clicks + Macro-economic factors }

"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)

ArŶmα Labs
Decoding the past, Encoding the future

<<Index



iOS 14 Privacy Update

GDPR

Google 3rd Party Cookie opt-in

Economic Downturn Putting Pressure on Marketing Disciplines

Wide Adoption of MMMs

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]

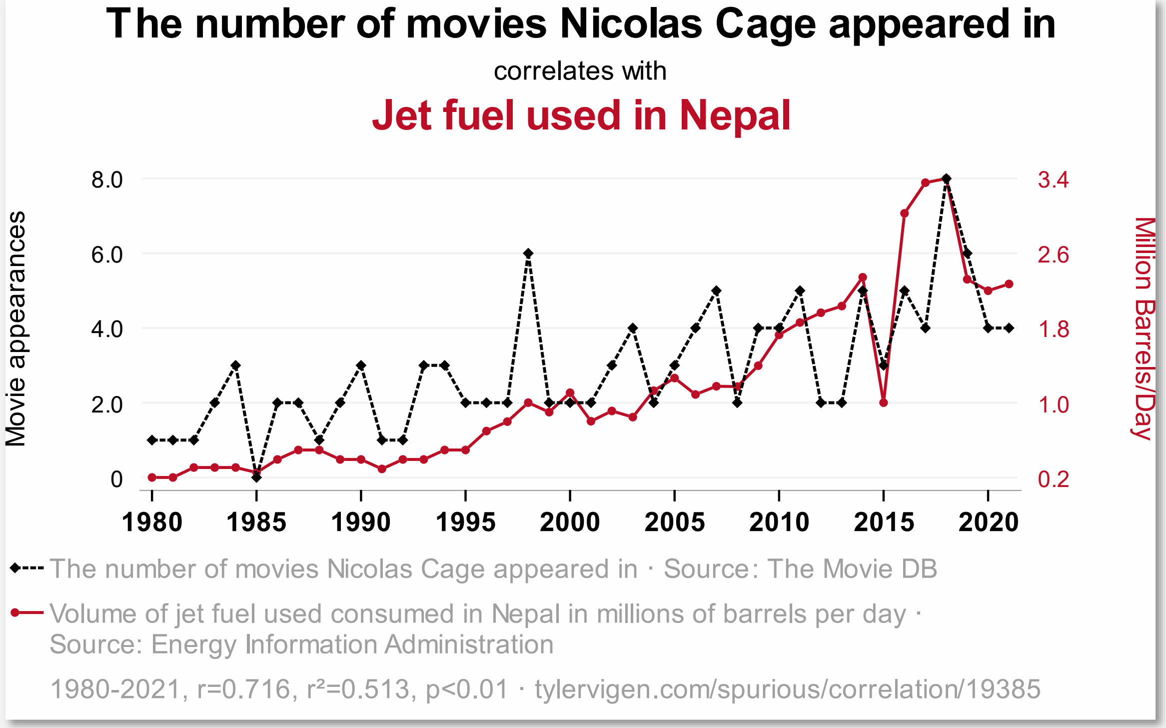


Image by Tyler Vigen

What is correlation?

ArŶmα Labs
Decoding the past, Encoding the future

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

- **It is derived from Covariance .**

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

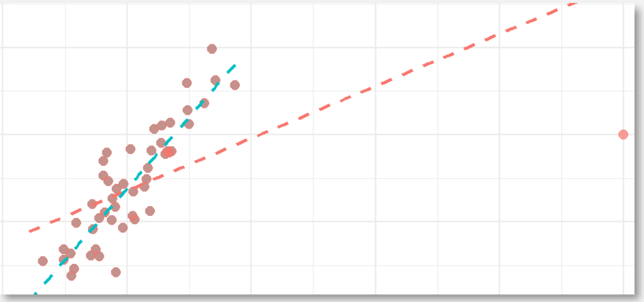
- The formula for the Pearson Product–Moment Correlation Coefficient is

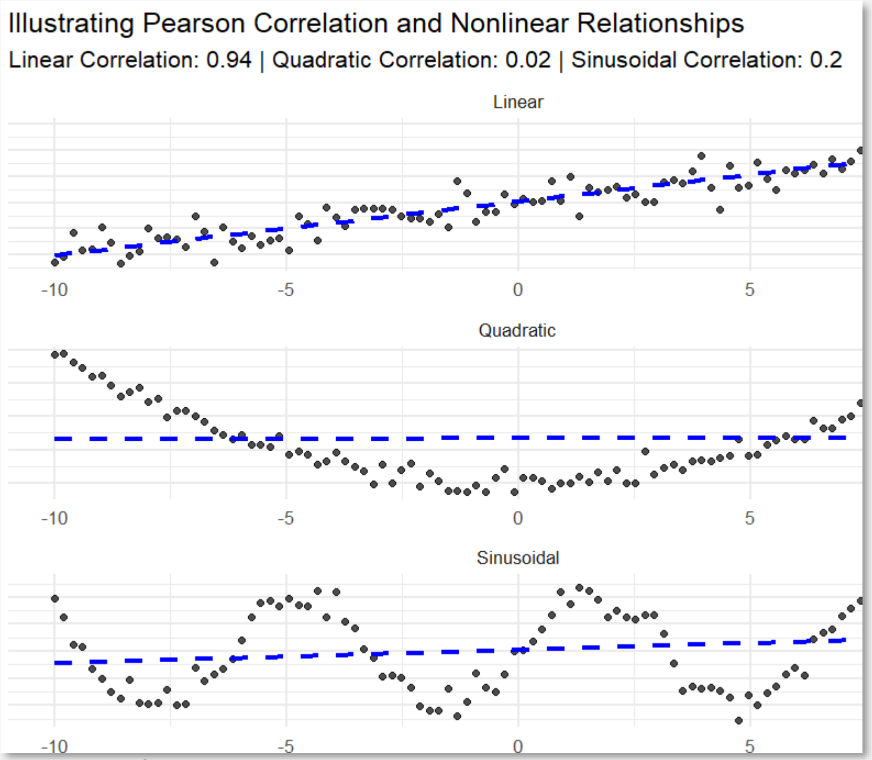
$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

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

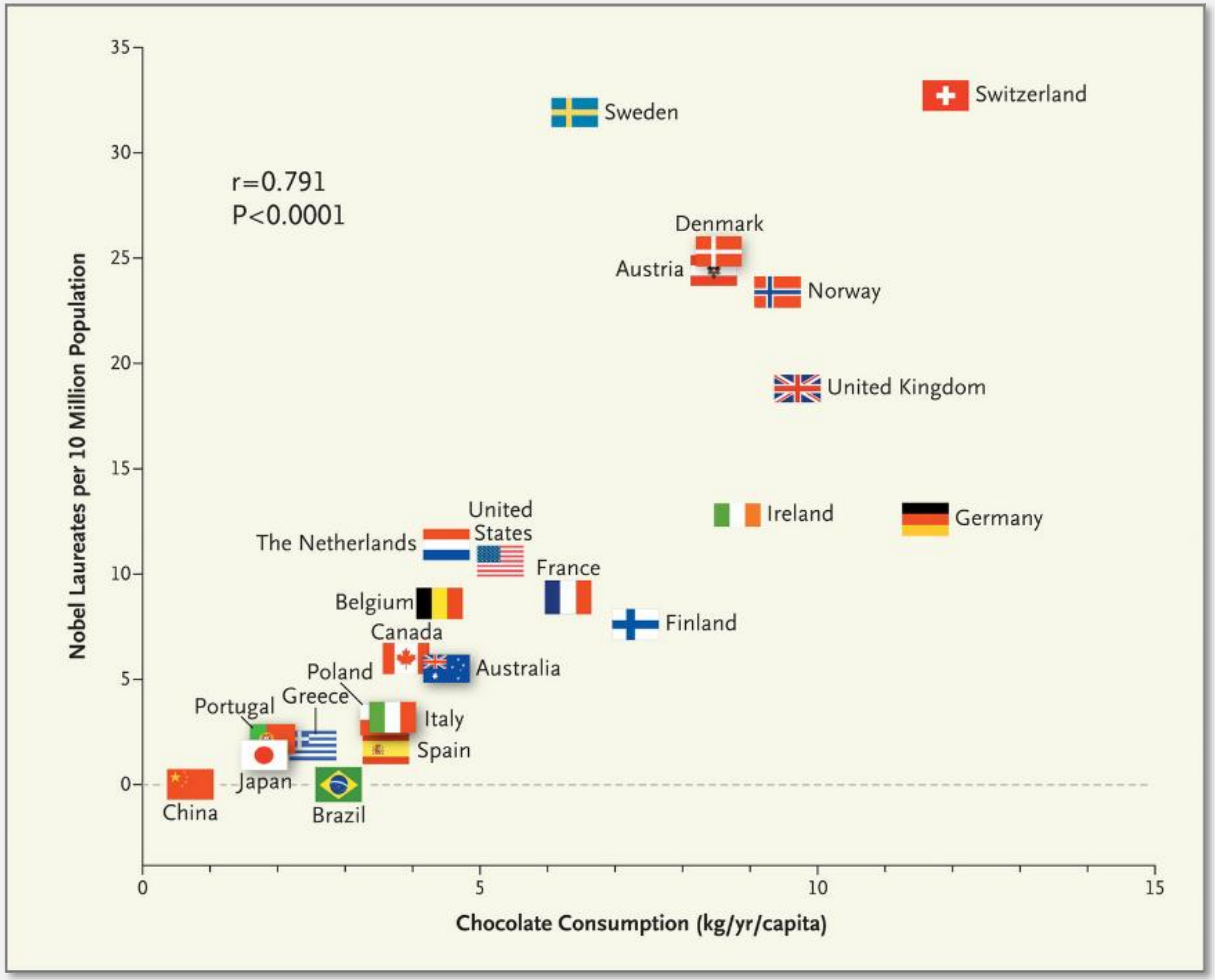
- **Does not imply causation!**
- **Sensitive to outliers (Pearson).**
- **Fails with non-linear relationships (Pearson).**



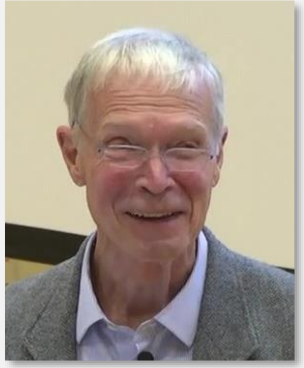


Correlation does not imply causation

ArYma 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.



+

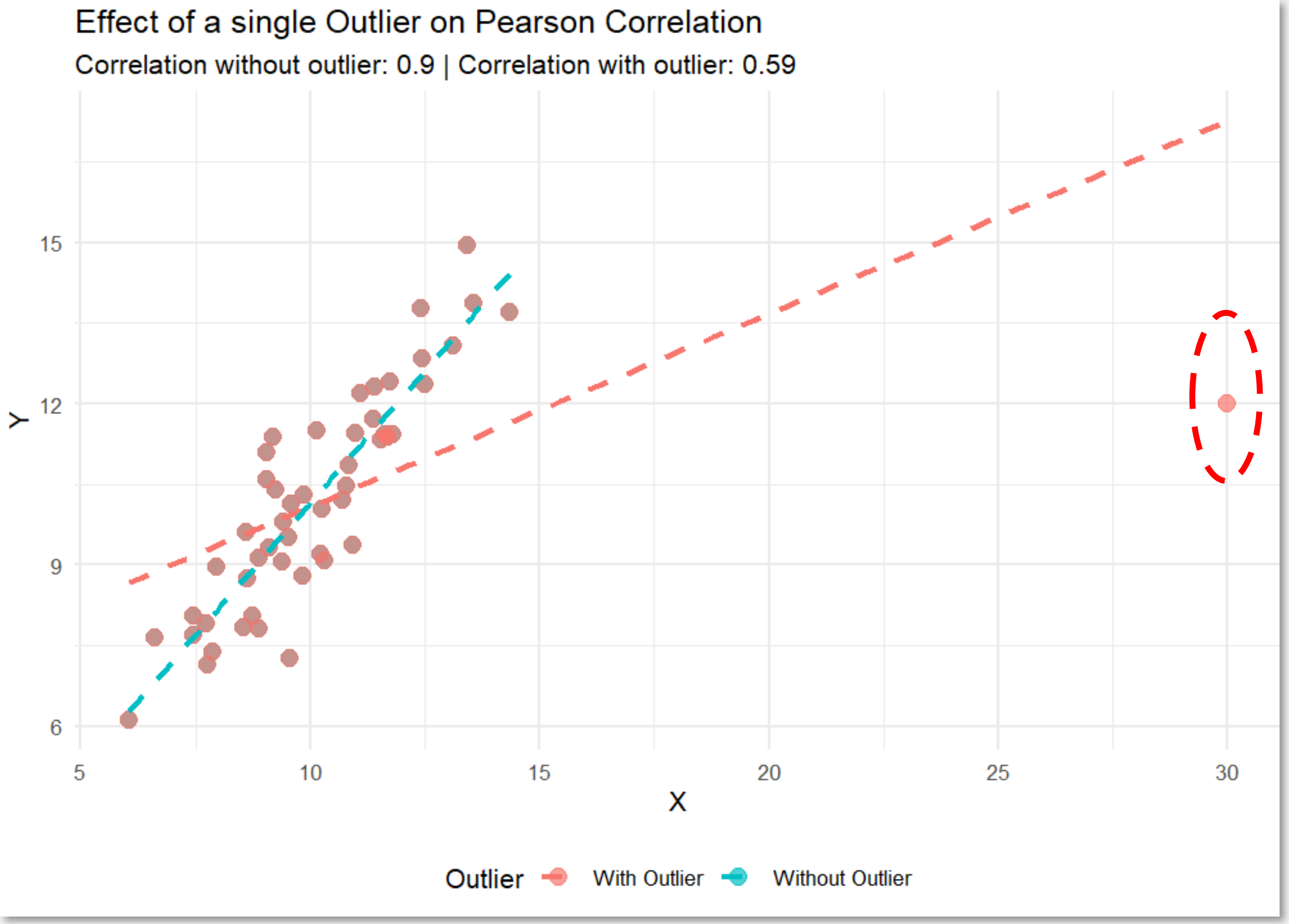
=

?

Correlation does not imply causation

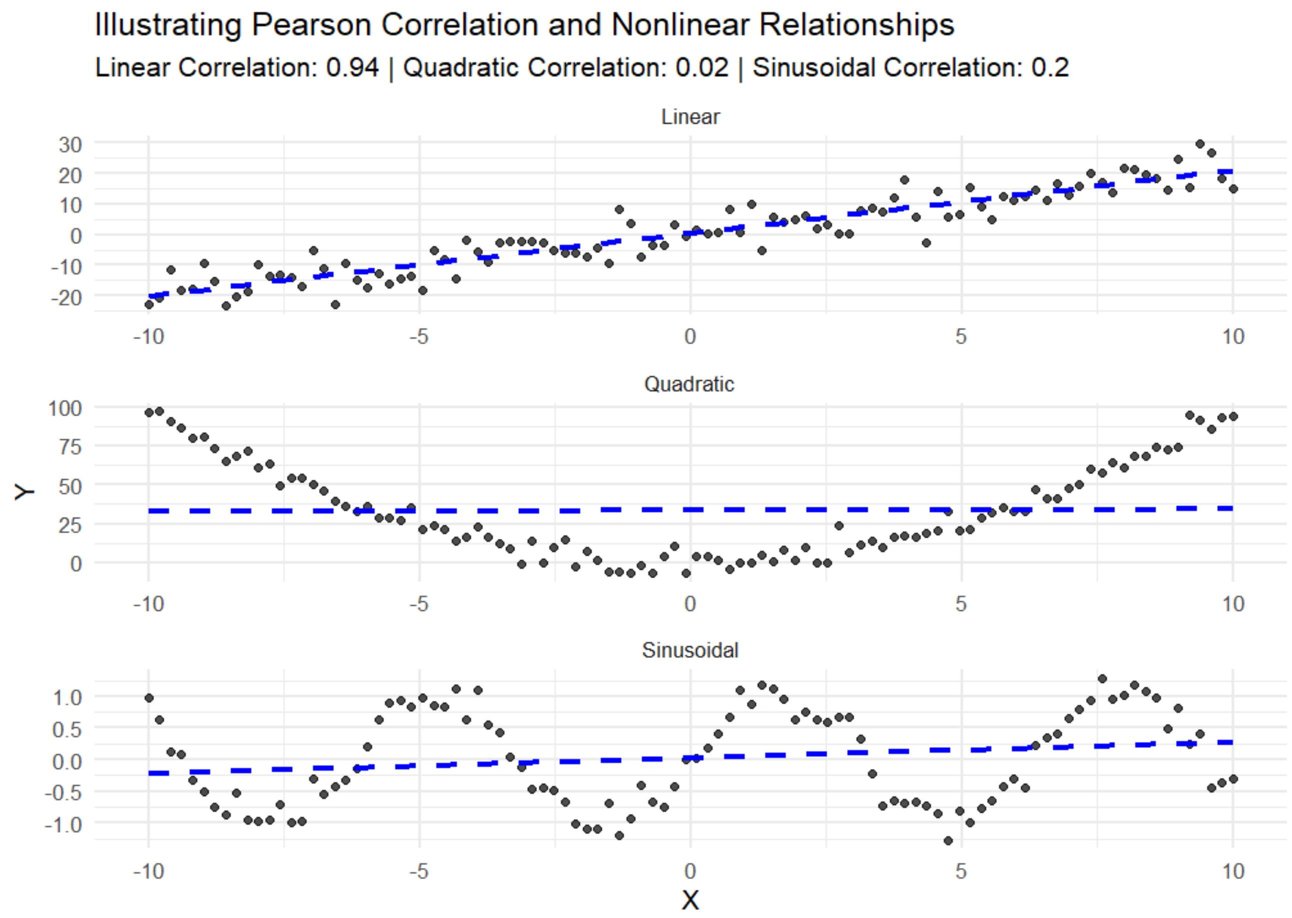
ArŶmα Labs
Decoding the past, Encoding the future

Sensitive to outliers (Pearson)



- 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



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

Types of Correlation

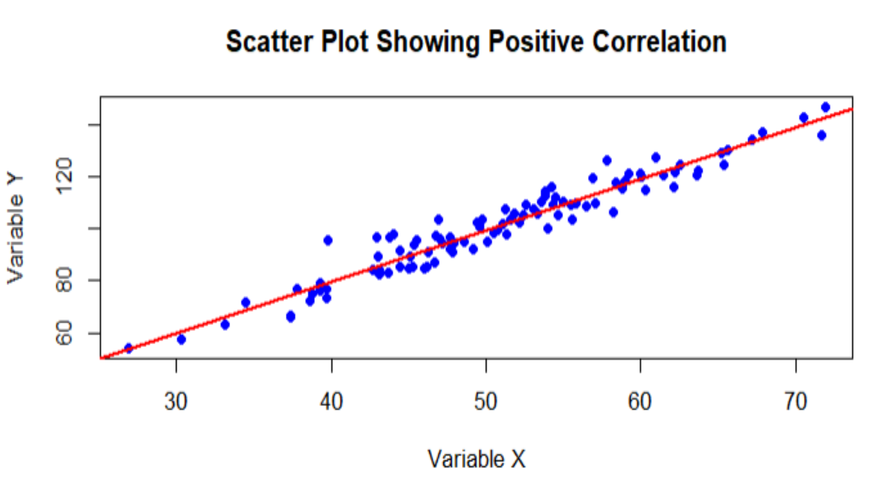
ArŶmα Labs
Decoding the past, Encoding the future

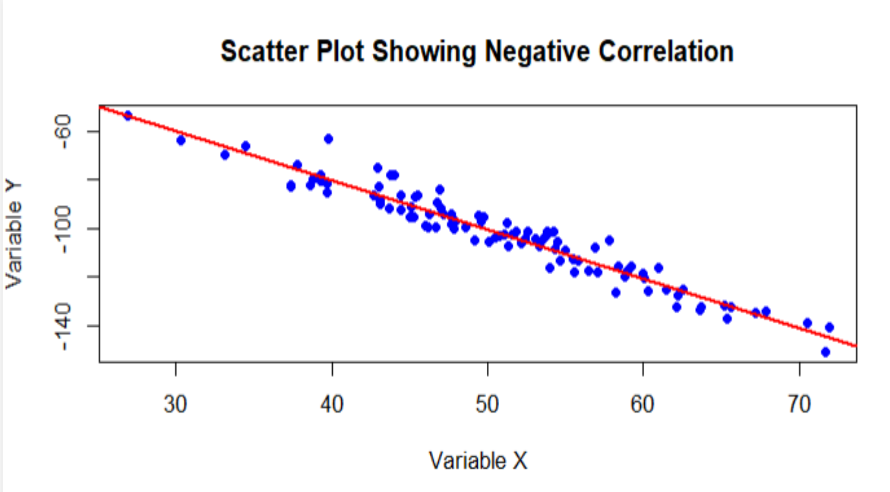
- Pearson Product-moment Correlation (r)
- Spearman Rank Correlation (ρ)
- Kendall Tau Correlation (τ)

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:

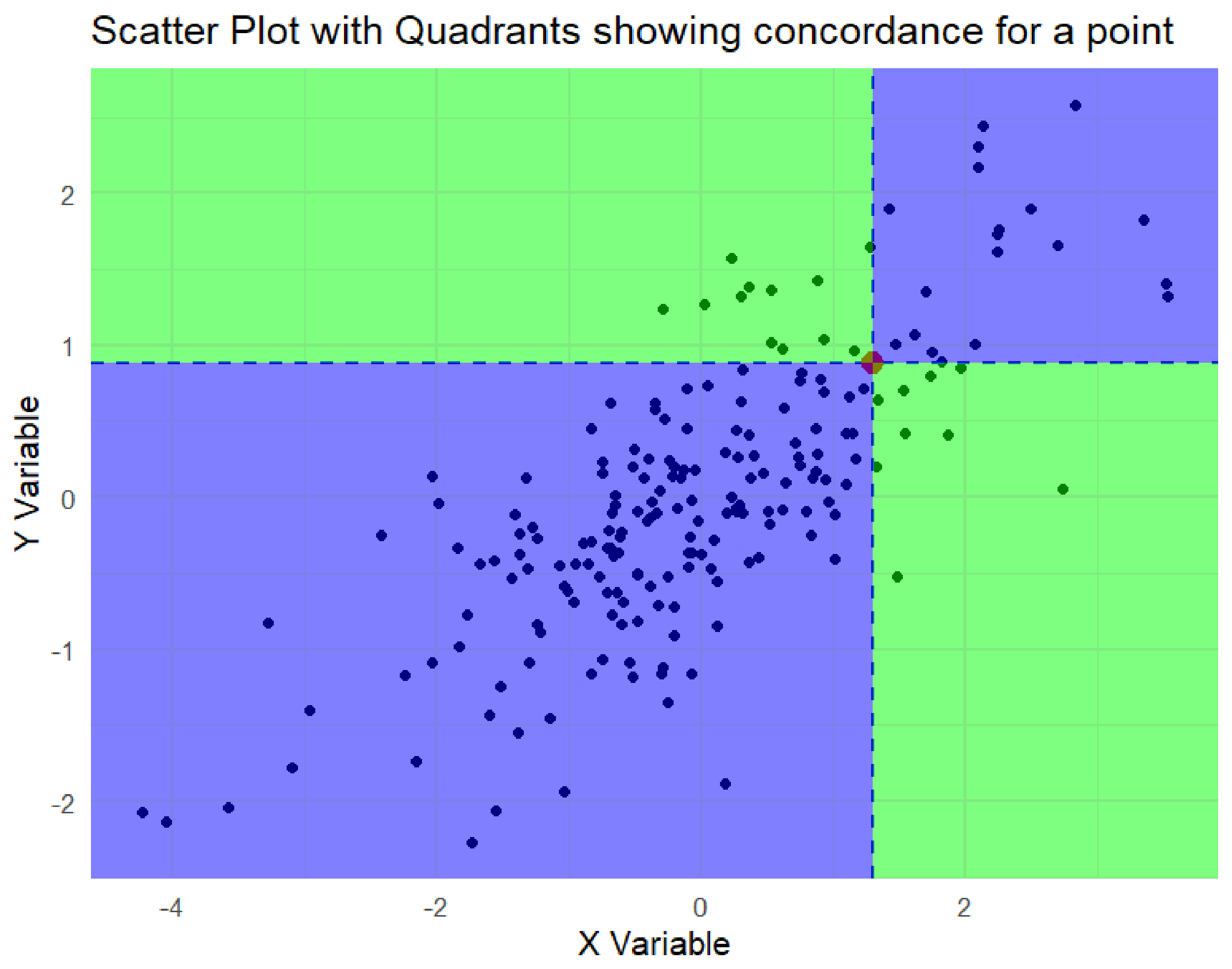
$$\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2 - 1)}$$

- 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)

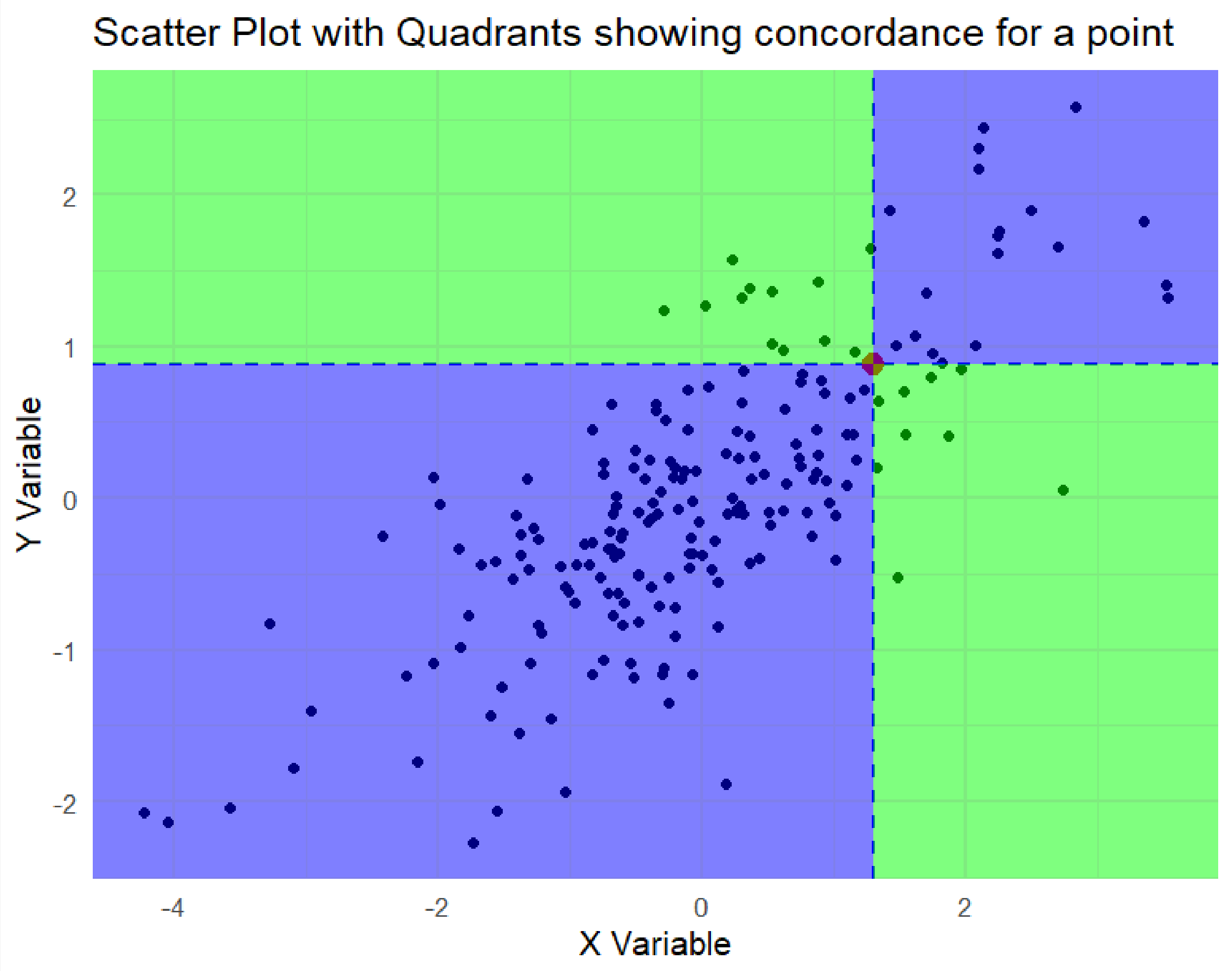
ArŶmα Labs
Decoding the past, Encoding the future

- 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)

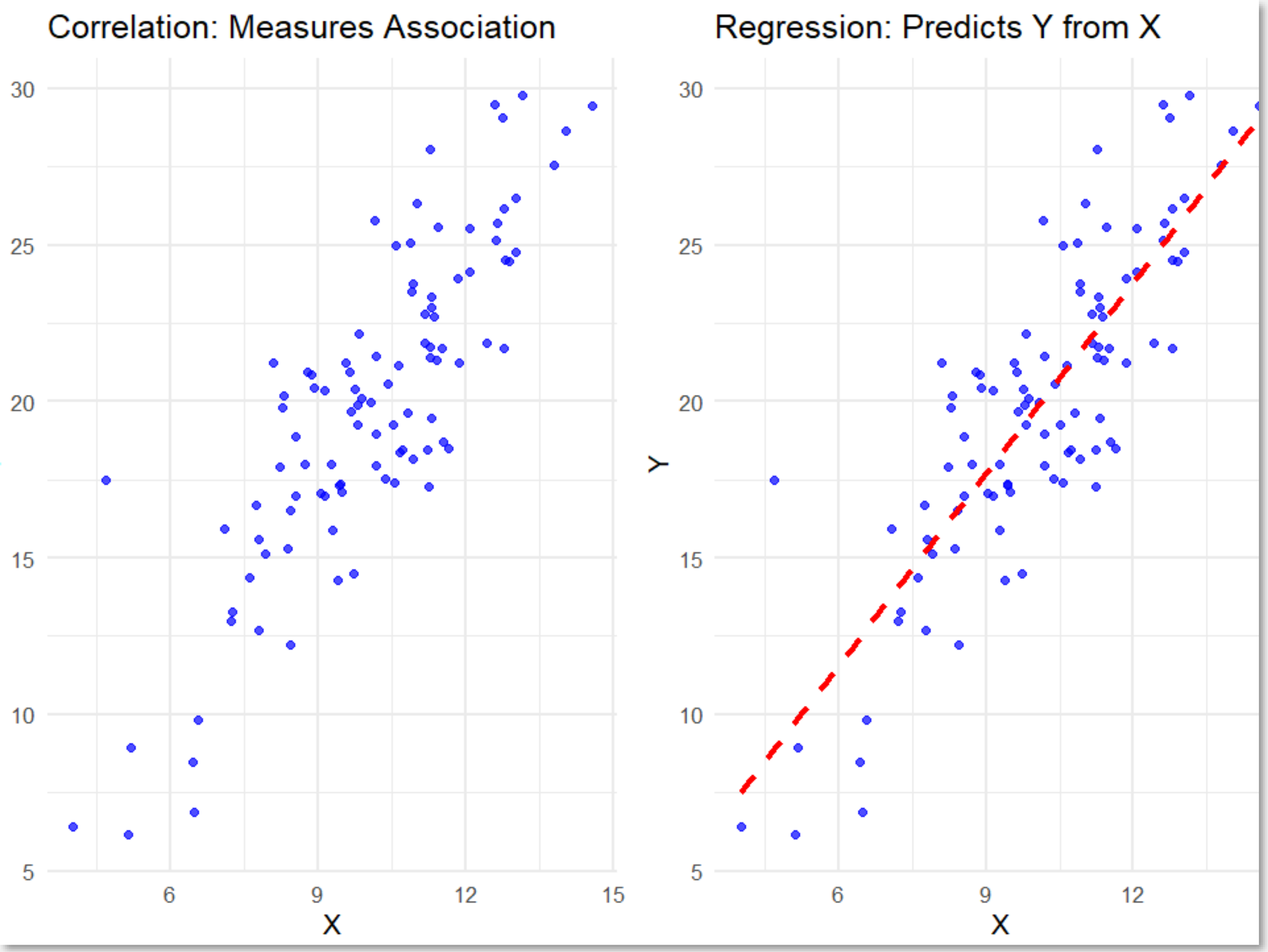
- The formula for Kendall's Tau Correlation Coefficient is

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

- 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?

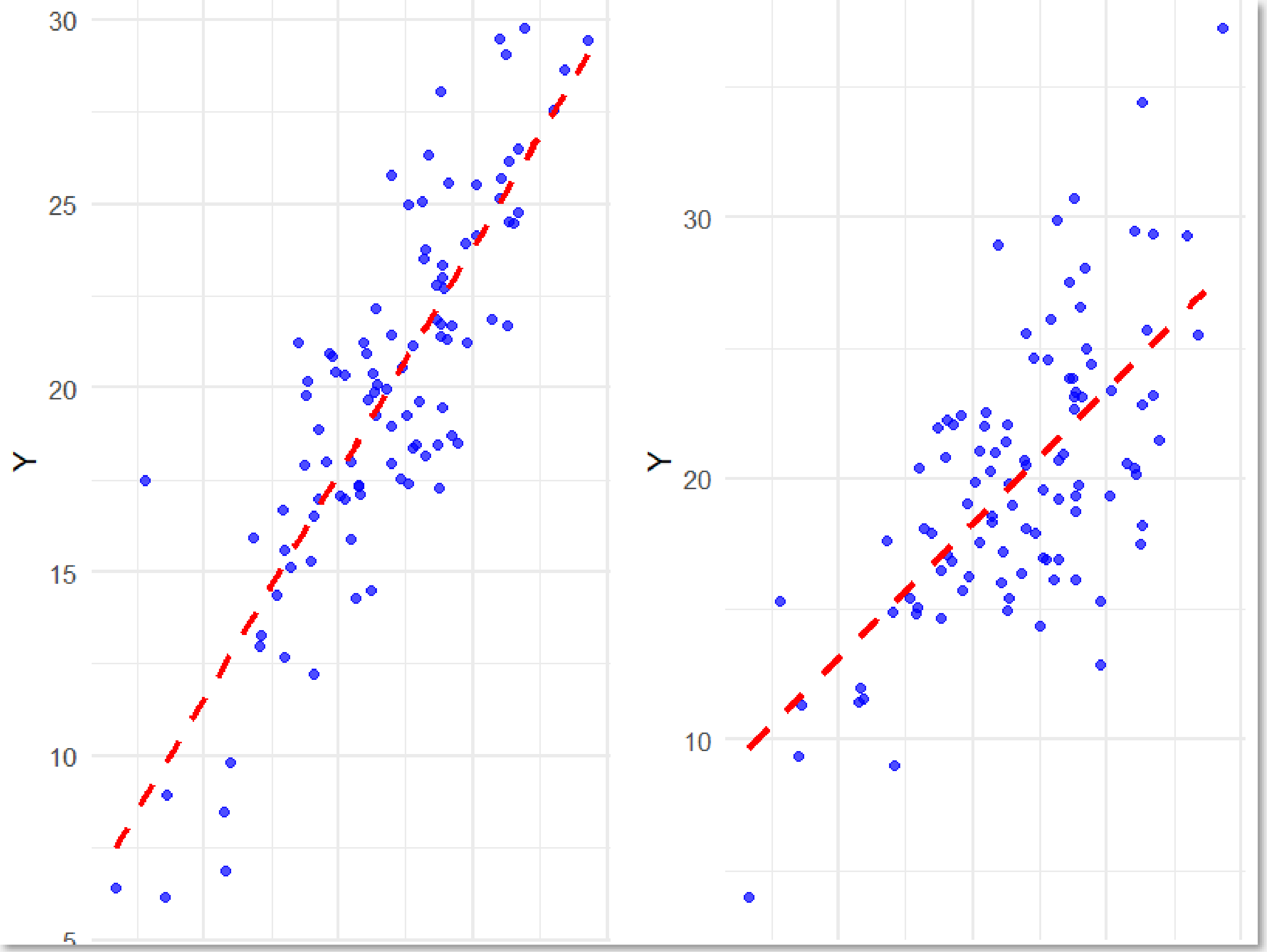
Correlation and Regression: Connections and Similarities



- 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.

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

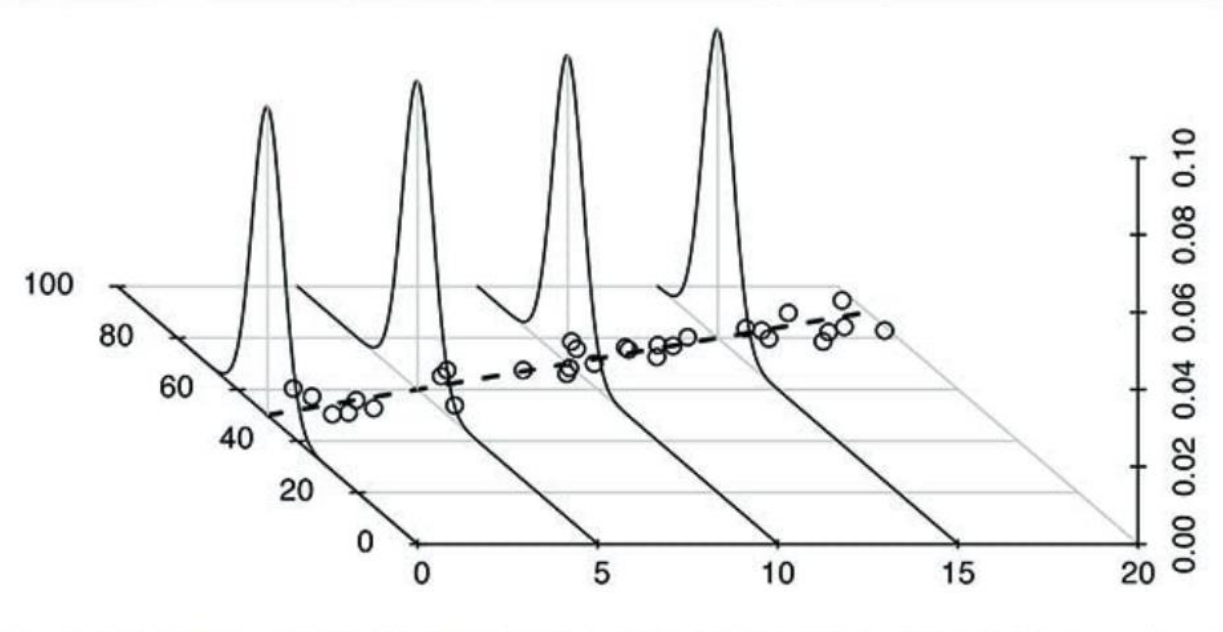
Multivariable regression has a univariate response.

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

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?

ArŶmα Labs
Decoding the past, Encoding the future

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

ArYma Labs
Decoding the past, Encoding the future

Linearity

(Linear In terms of coefficients)

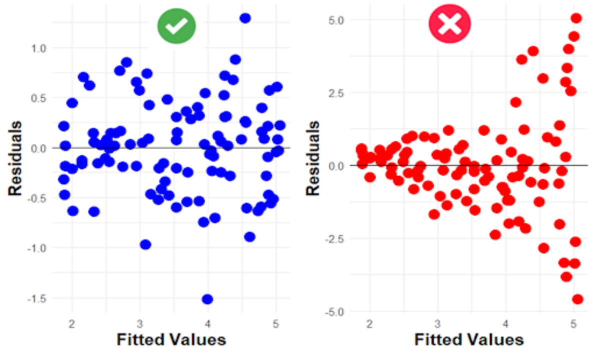
$y = \beta_0 + \beta_1 x + \text{error}$

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

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

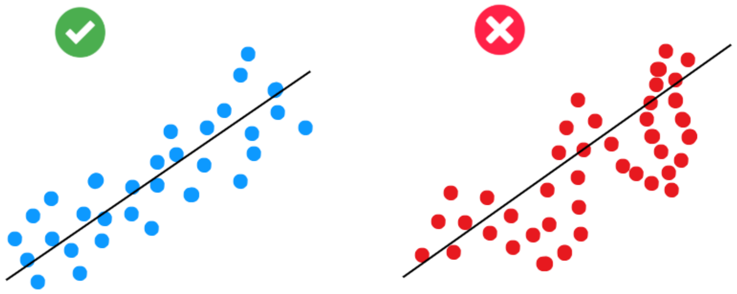
Homoscedasticity

(Equal variance of error terms)



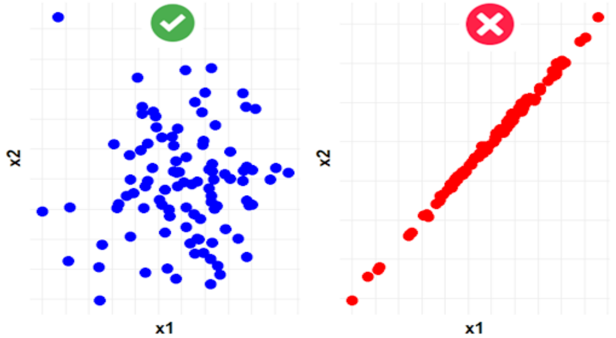
Independence

(Error terms are Independent)



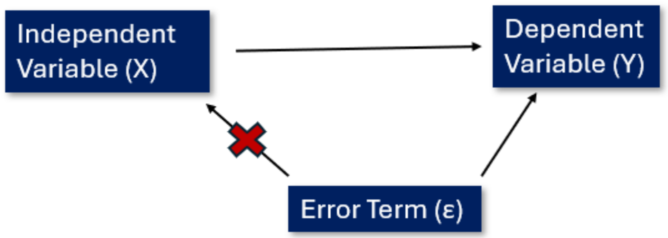
Lack of Multicollinearity

(Predictors are uncorrelated with each other)



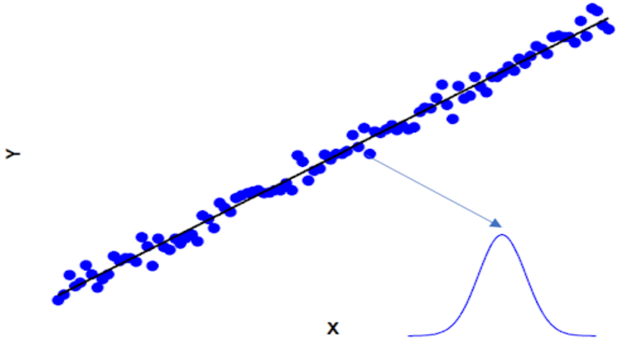
Absence of Endogeneity

(No correlation with Predictors and error)



Normality

(Errors are normal)

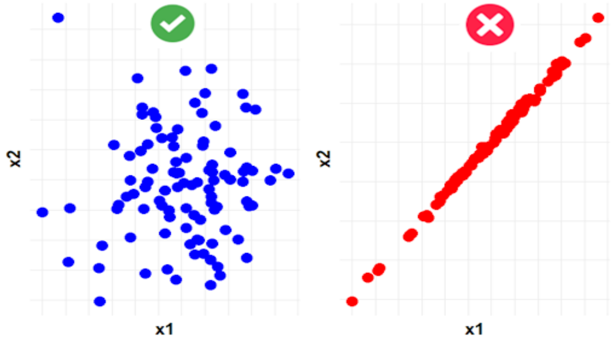


Important assumptions for MMM

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

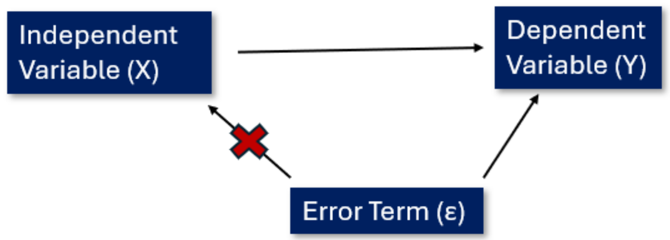
Lack of Multicollinearity

(Predictors are uncorrelated with each other)



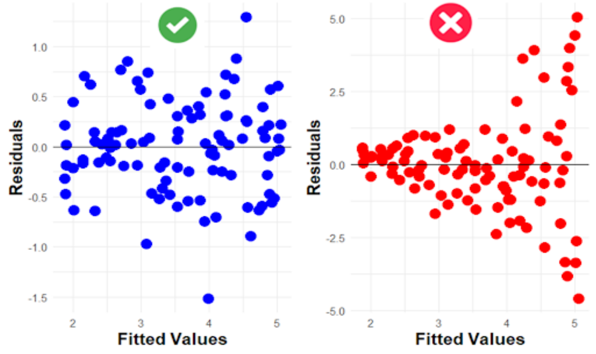
Absence of Endogeneity

(No correlation with Predictors and error)



Homoscedasticity

(Equal variance of error terms)



Important assumptions for MMM – Multicollinearity

What problems it can cause

Ar𝛾ma 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).

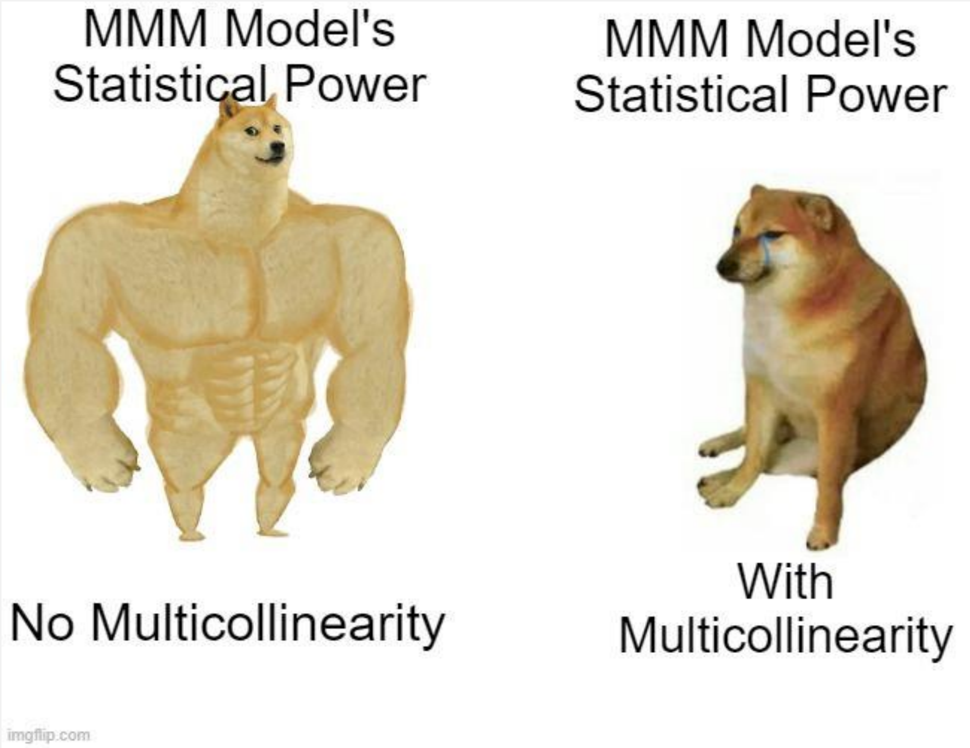
(Suppose we're talking VERY tiny distances here)

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

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

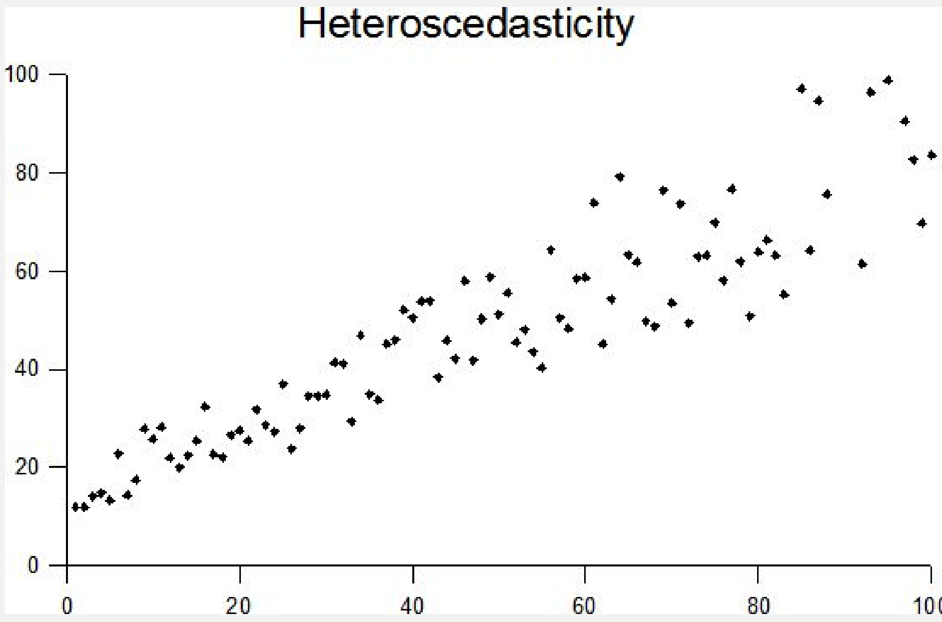
In case of Heteroscedasticity, the residuals by and large have some pattern. Let's take an example of one of the popular pattern.

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.

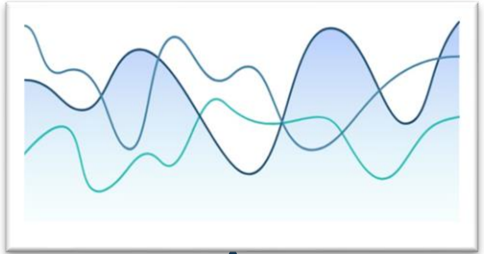
ArŶmα Labs

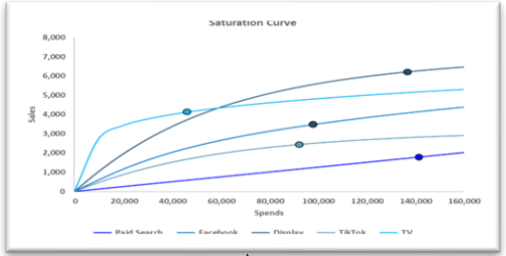
Decoding the past, Encoding the future

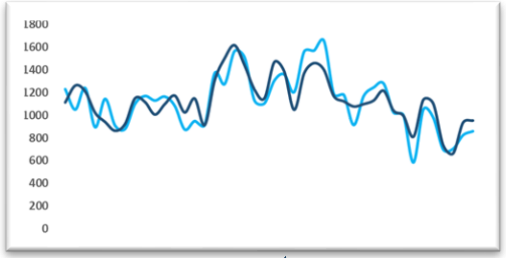
EDA

MMM Process

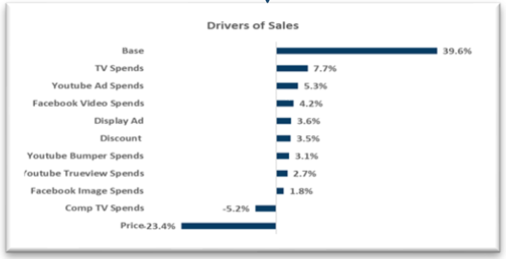
ArŶmα Labs
Decoding the past, Encoding the future

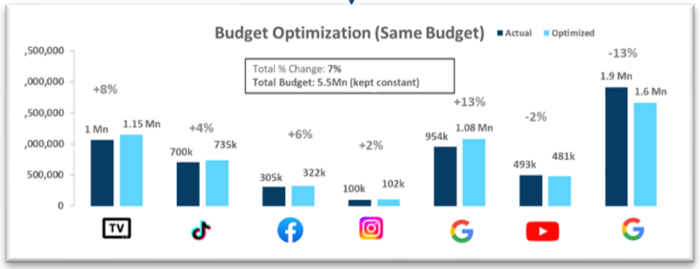






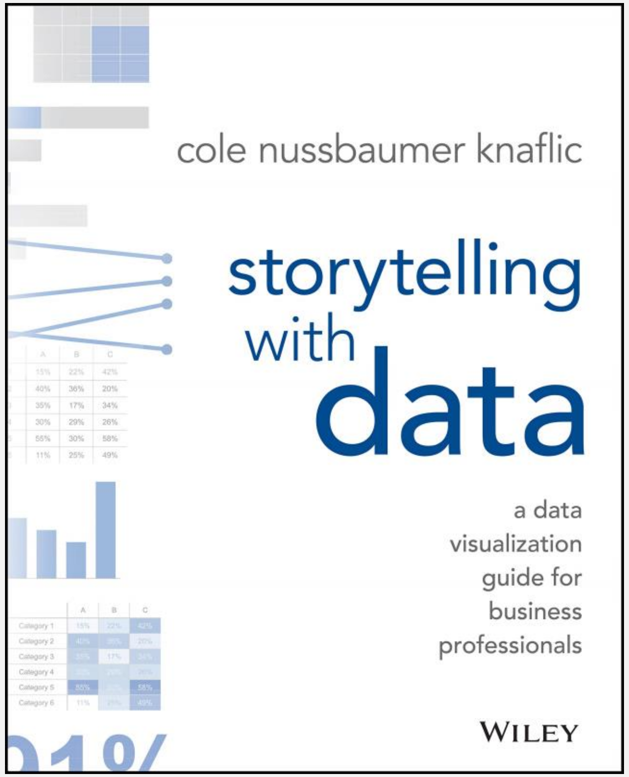
EDA → Model Building → Model Evaluation → Budget Allocator → Implementing Models on the ground → Causal Lift Validation → Frequent Model Updates

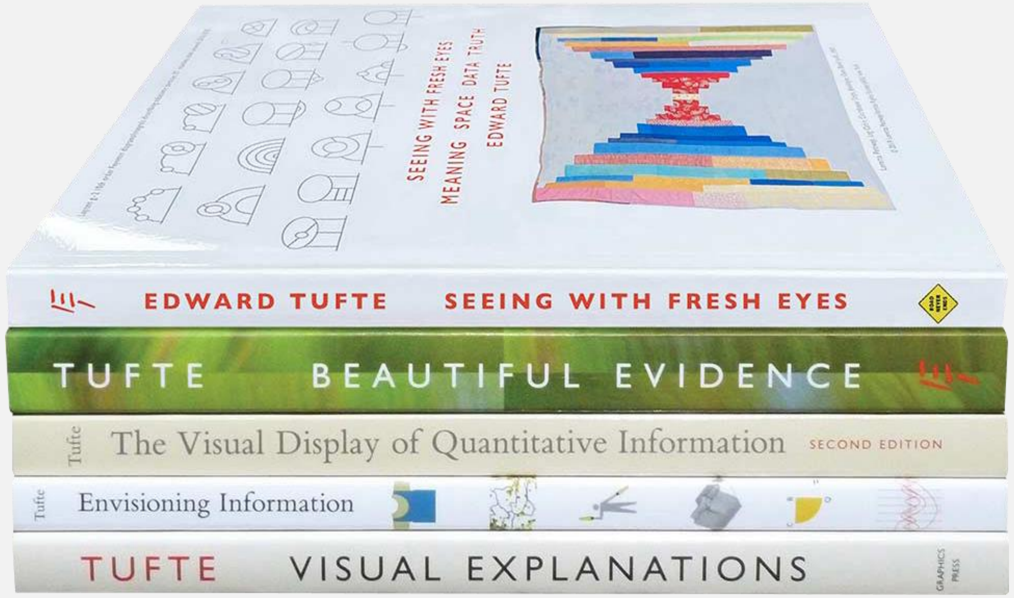




Data Visualization

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





Exploratory Data Analysis

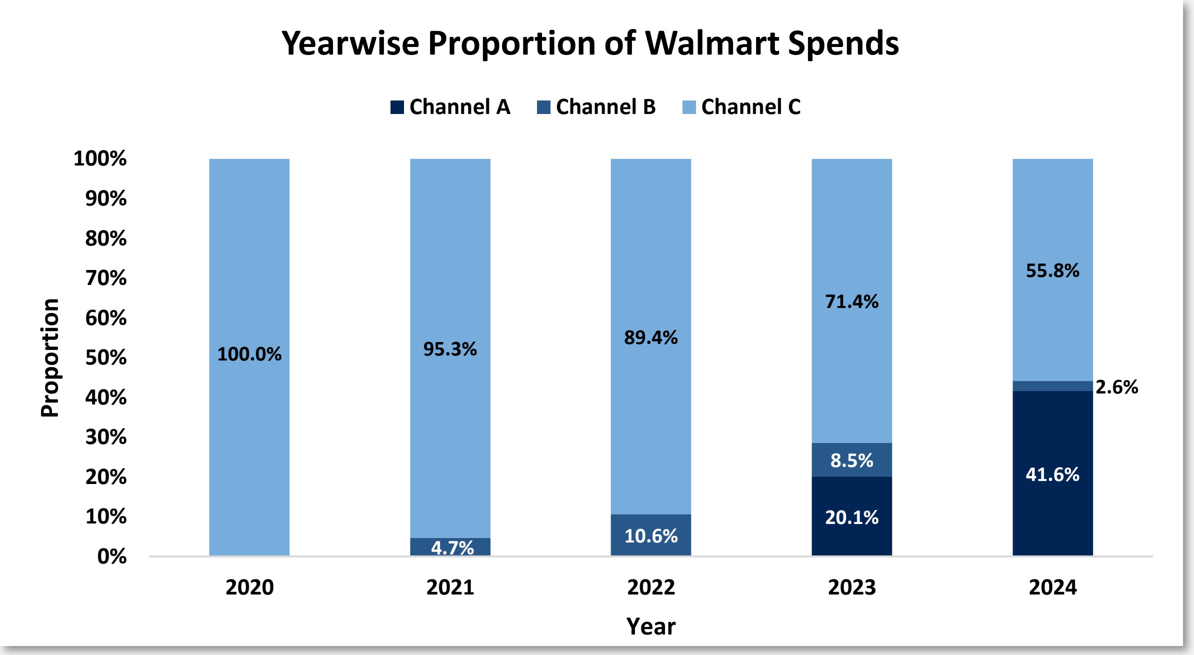
ArYma Labs
Decoding the past, Encoding the future

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.

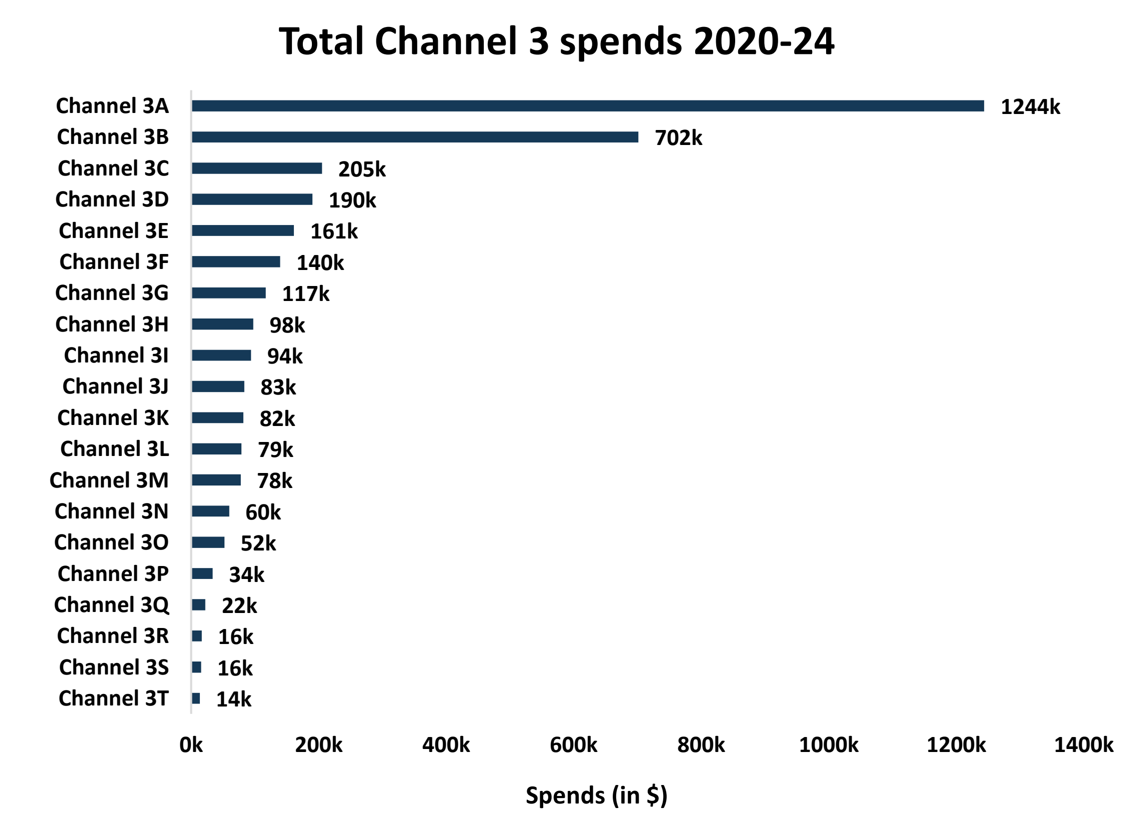


EDA – Common practices

ArŶmα Labs
Decoding the past, Encoding the future

Common EDA practices that should be followed:

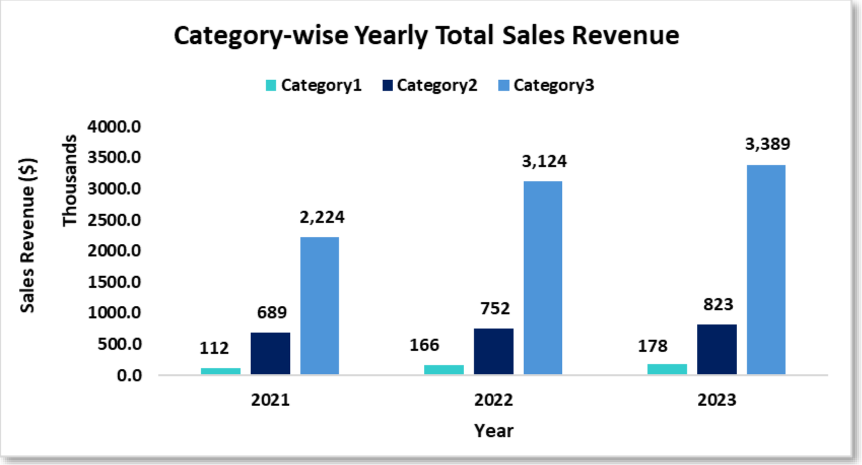
- ✓ Check for duplicates, Handle missing values,
- ✓ Analyze the distribution of individual variables,
- ✓ Identify anomalies that could skew results,
- ✓ Decompose time series into trend, seasonality, and residual components and analyze patterns over time,
- ✓ Use techniques like standardization and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



List of EDA Tasks (1/2)

ArŶmα 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

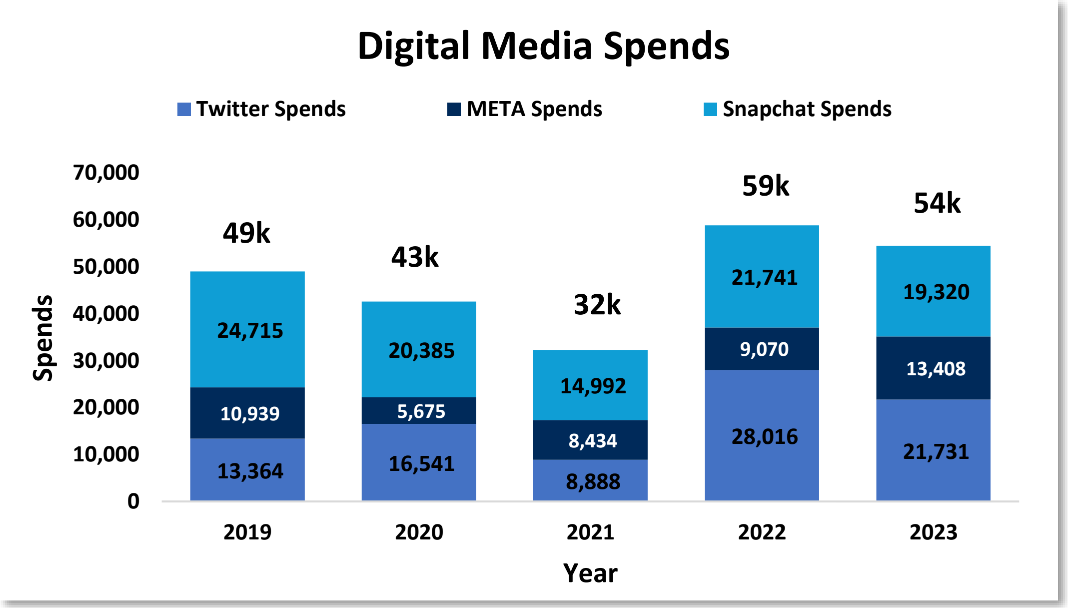
ArŶmα Labs
Decoding the past, Encoding the future

- 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.

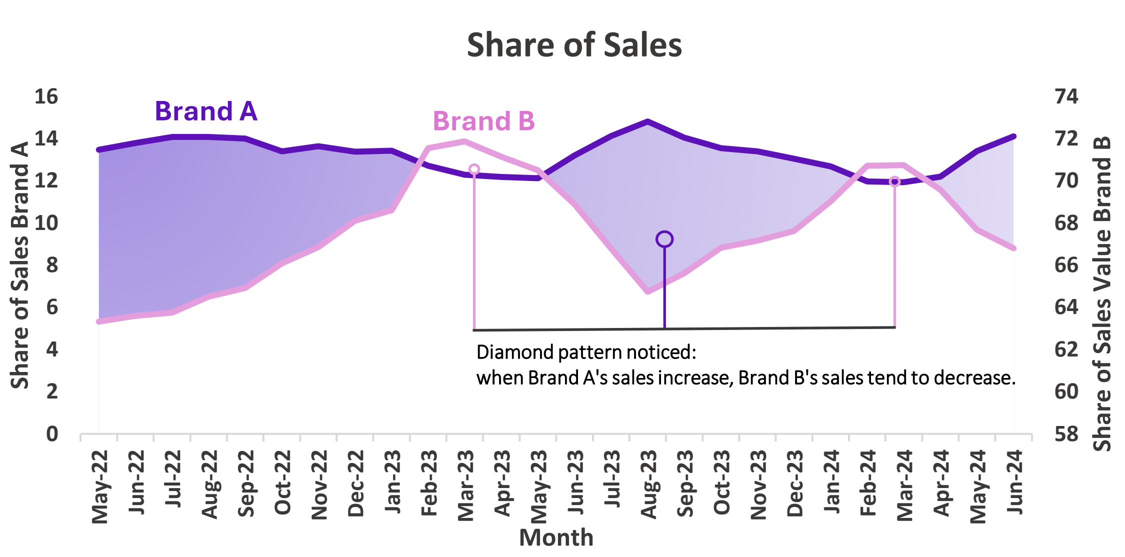
- Maintain consistent fonts and colors for the same variables across all charts. Assign one color per variable and use it uniformly in all visualizations.

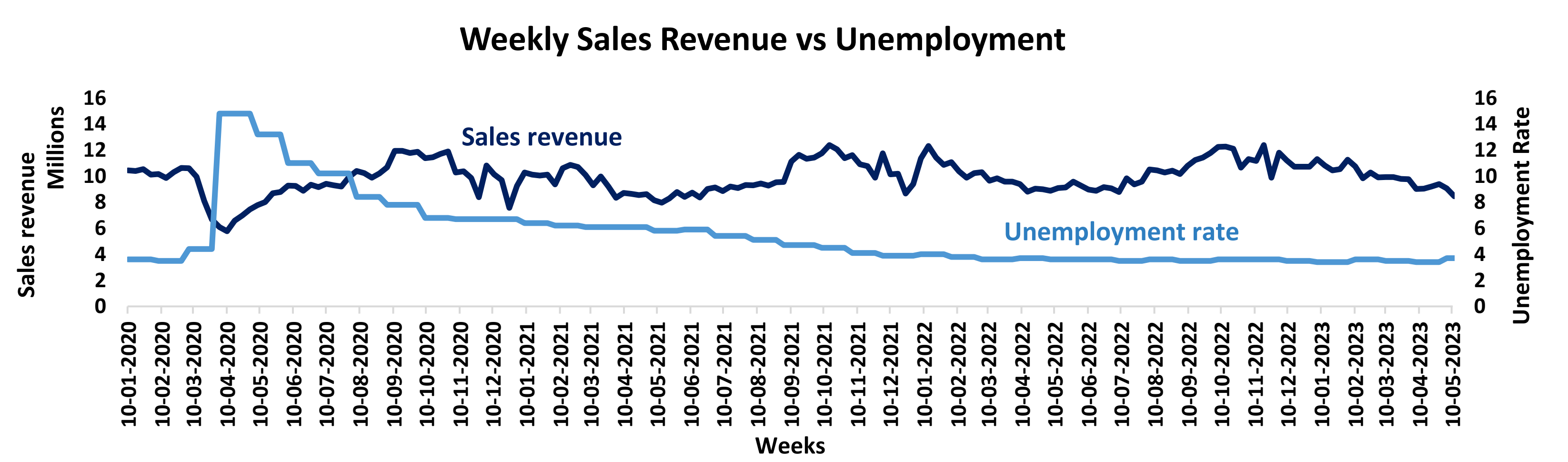
- Ensure all charts are labeled clearly with meaningful legends and titles.

EDA Charts (1/4)

ArŶmα Labs
Decoding the past, Encoding the future

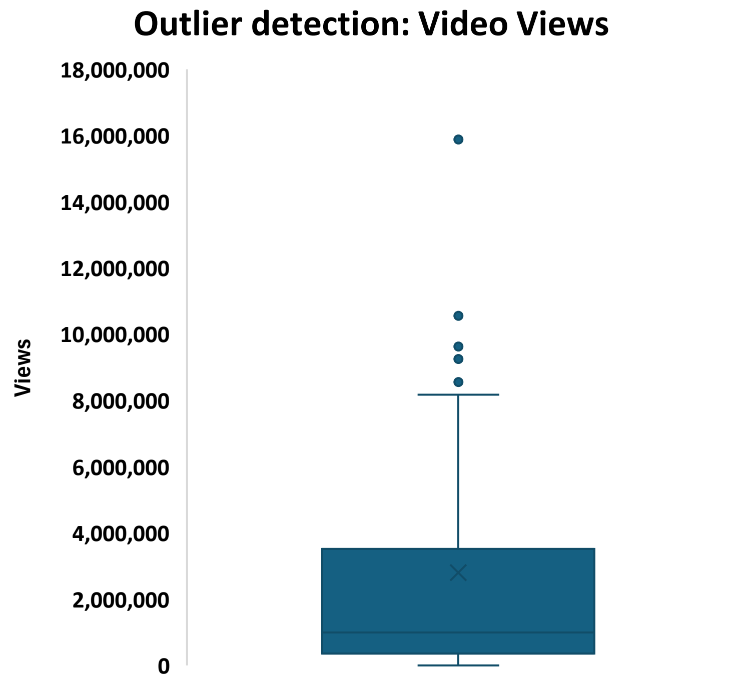
Trend Line Charts: To depict changes over time, for analyzing trends in the KPI (e.g., Revenue/Conversion, price, market share etc.)



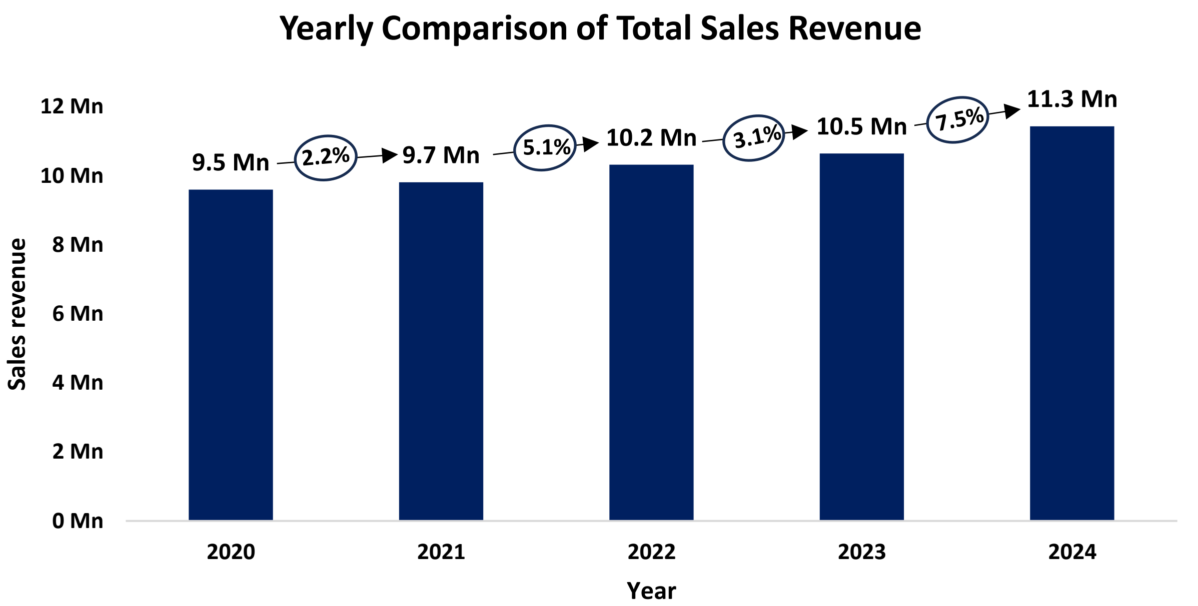


EDA Charts (2/4)

ArŶmα Labs
Decoding the past, Encoding the future



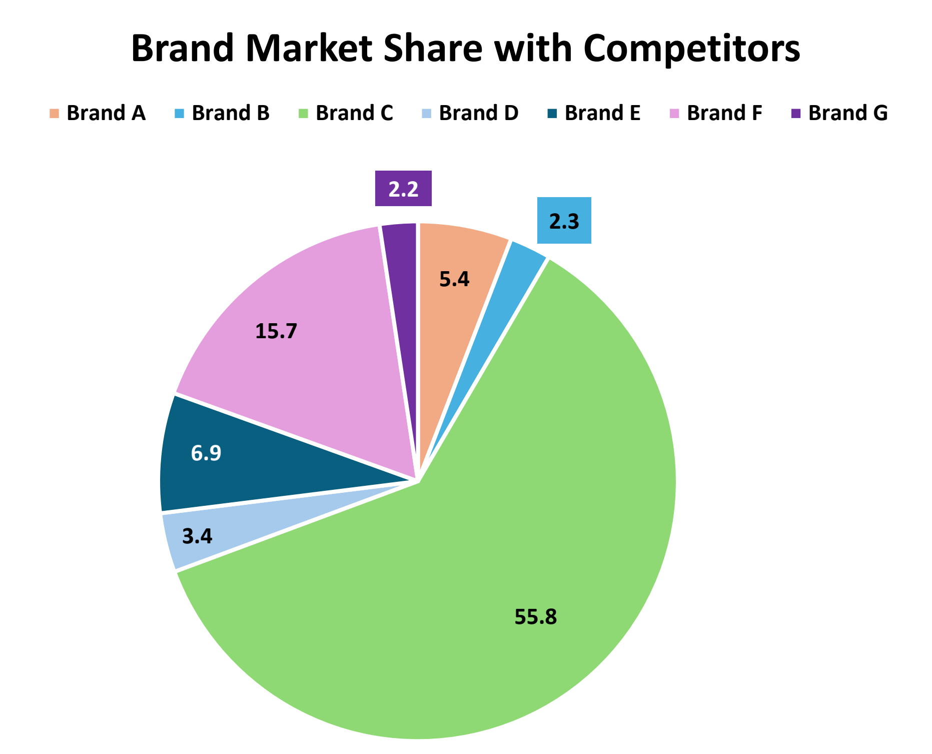
Boxplots: To identify outliers in data columns.



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

EDA Charts (3/4)

ArŶmα Labs
Decoding the past, Encoding the future



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

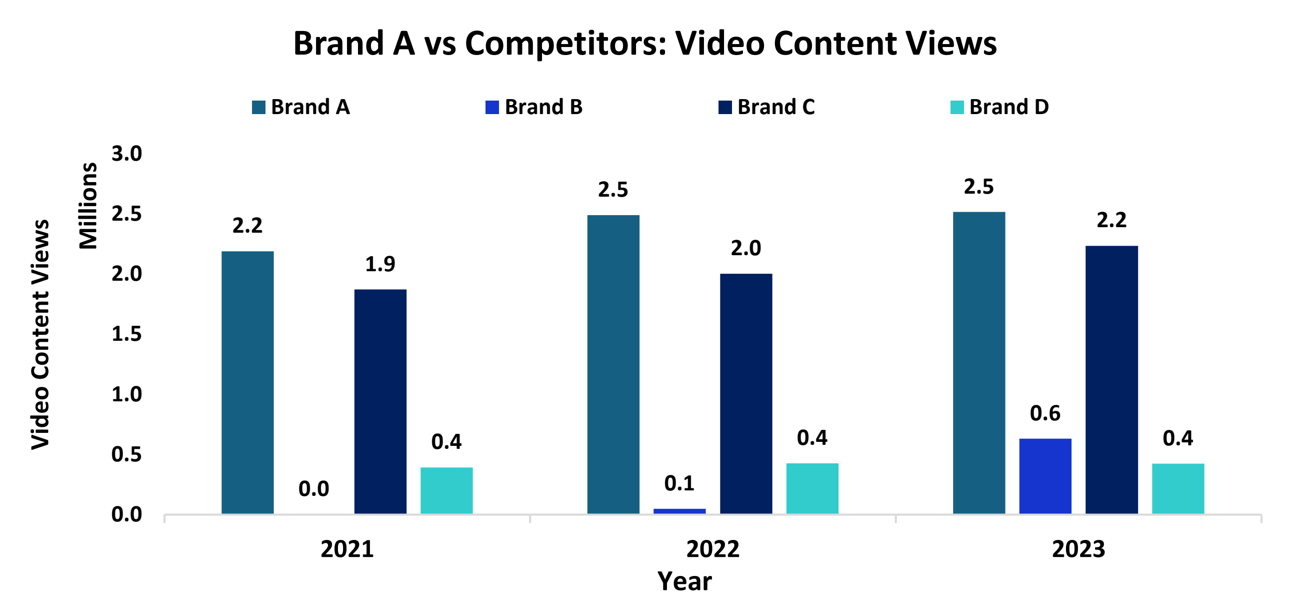
Correlation Summary

	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

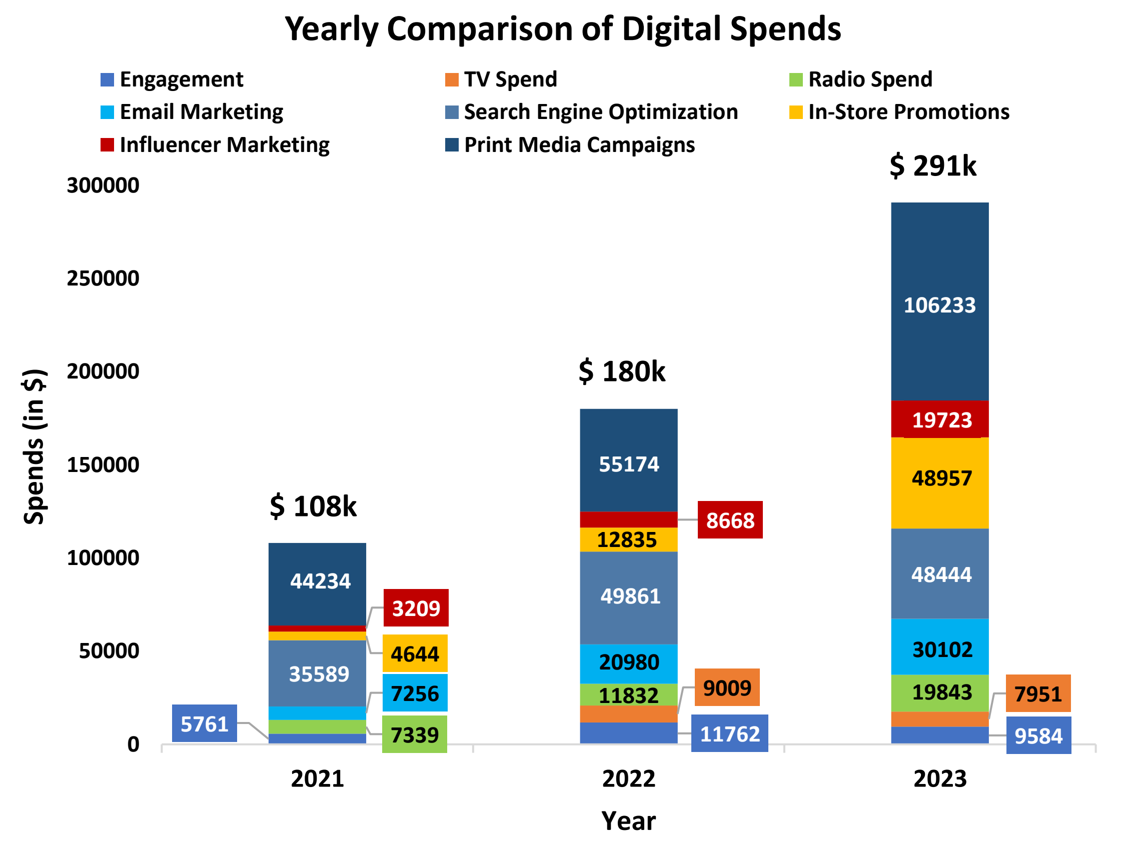
Correlation heatmaps: A color-coded matrix where each cell shows the correlation coefficient, helping identify strong positive, negative, or neutral relationships between variables.

EDA Charts (4/4)

ArŶmα Labs
Decoding the past, Encoding the future

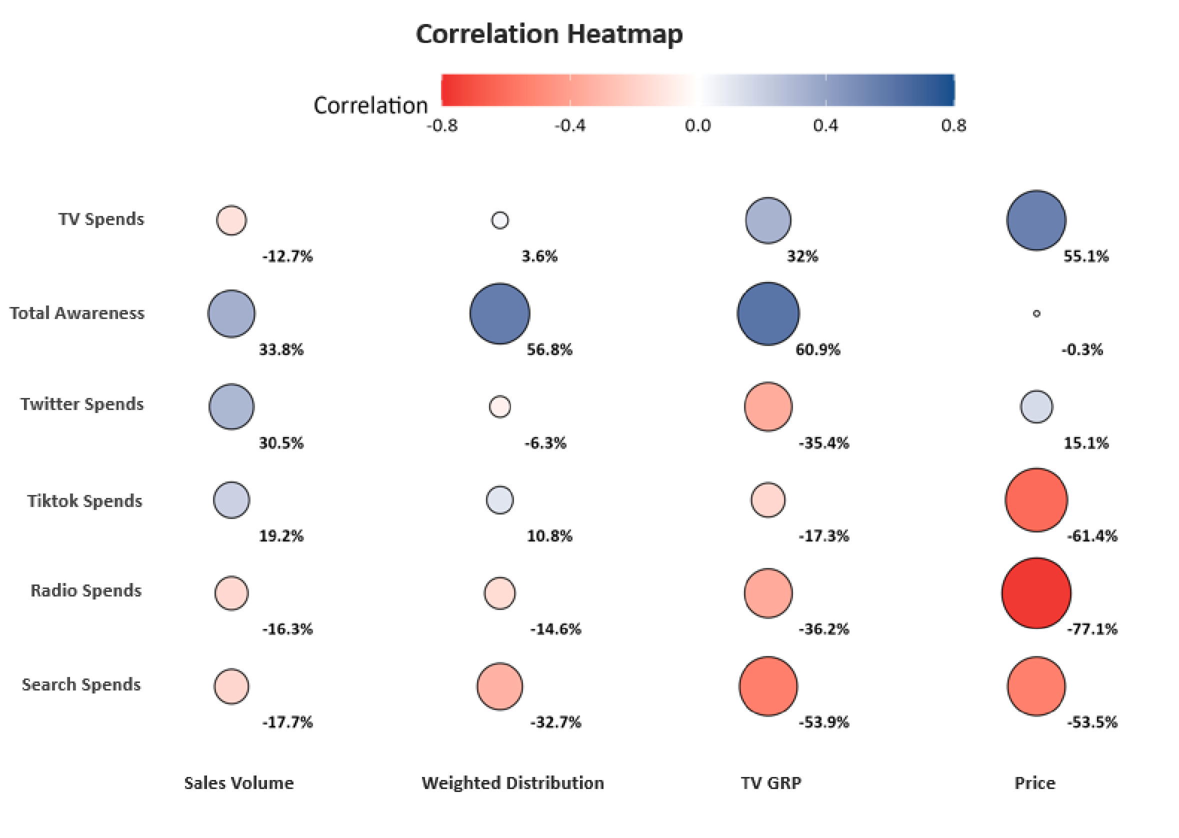


Clustered Column Charts: To compare own brand content views vs competitor video content views.



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

Correlation Heatmap

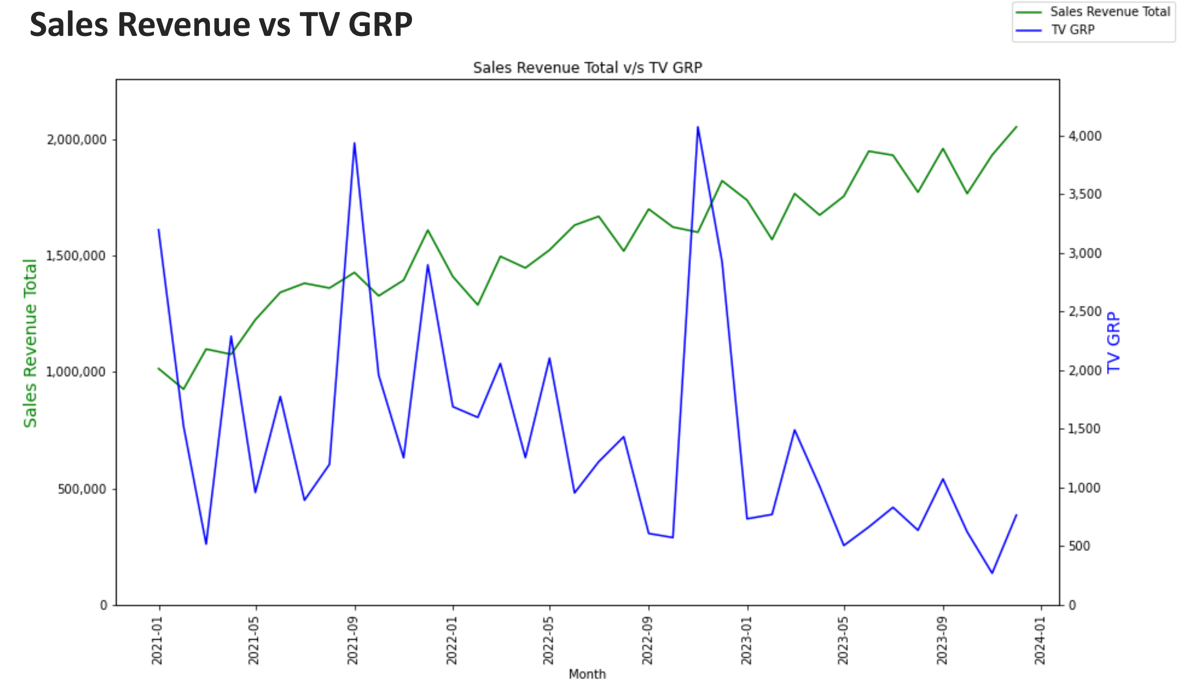


Correlation heatmaps are visual tools used to display the relationships between numerical variables.

Important charts for Feature Analysis and Model Building

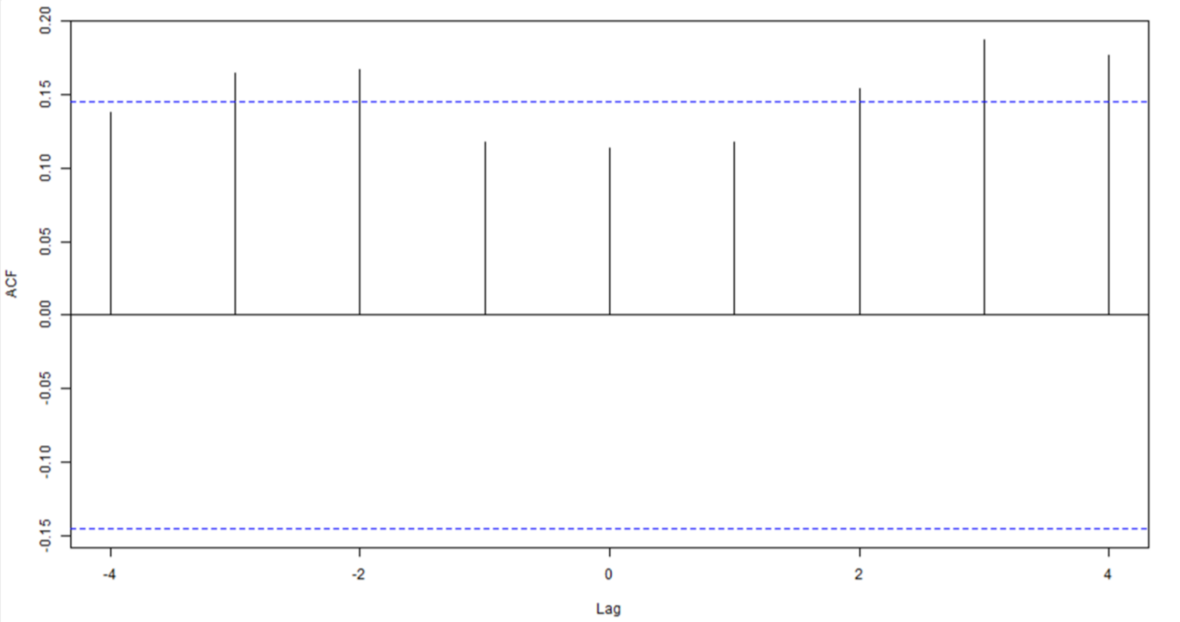
ArŶmα Labs
Decoding the past, Encoding the future

Sales Revenue vs TV GRP



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

Sales Revenue vs Impressions



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