Retail & Marketing Analytics

INDIVIDUAL ASSIGNMENT REPORT

Nikita Bhilare | MSc. Business Analytics | Imperial College London

Table of Contents

Introduction

Marketing stands at the heart of a business's interaction with its consumer base, serving as the link between customer needs and the company's offerings. In today's retail landscape, investing in the optimal mix of marketing channels is not just about amplifying user engagement or boosting short-term sales; it's a strategic operation critical for sustainable revenue growth, building brand equity, and ensuring a cohesive brand narrative across all consumer touchpoints. Omnichannel marketing highlights the necessity for an integrated approach that blends online and offline experiences to increase consumer reach. By leveraging detailed analytics, businesses can fine-tune their marketing expenditures, align inventory levels with promotional activities, and set prices that resonate with consumers while driving profitability. This report seeks to explore these dynamics using a dataset that captures various facets of marketing and sales, providing a comprehensive view of the impacts of marketing activities and operational decisions on sales performance.

Omnichannel Marketing

Marketing Opportunity

In today's competitive and rapidly evolving retail sector, companies across industries are continuously seeking ways to optimize their marketing strategies and enhance overall performance. On the other hand, numerous companies with high-quality products struggle to convert their offerings into sales, often hindered by less effective brand awareness and suboptimal pricing strategies. Customer unawareness, unattractive pricing, and inadequate stock management obstructs market share growth, further complicated by volatile consumer preferences.

By leveraging advanced analytics techniques, businesses can gain deep insights into the drivers of sales performance and understand their short-term and long-term effects on overall business outcomes. The identified marketing opportunity lies in harnessing the power of analytics to optimize channel budget allocation, refine pricing and discounting strategies, and enhance inventory management practices. This approach enables marketers to prioritize high-impact marketing channels and fine-tune pricing, discounting, and inventory strategies to maximize revenue and profitability, cultivate strong brand presence, thereby improving customer satisfaction and gaining a competitive edge in the market. Employing Vector Autoregression (VAR) modeling on our dataset will allow us to examine the nuances of how each marketing channel and operational decision affects sales, potentially resulting in increased market share or sales growth.

Understanding the data and the context

This dataset encapsulates a detailed record of sales and marketing metrics for a specific product by an undisclosed brand over an extensive period of 992 weeks. However, for the analysis, we are focusing on the most recent 104 weeks of data, that is from week 889 to week 992. This timeframe, equivalent to two years, is strategically selected to reflect relevant insights of current market trends and the effectiveness of recent marketing strategies.

Each entry in this dataset corresponds to one week of operation for a local e-commerce startup based in Berlin and documents the following attributes:

Variable	Description
Week	Week Number
Sale	Numerical data representing the number of product sales for each observed week
Price	The observed week's base price for the product
Discount	The discount rate applicable for the observed week
StockRate	The stock-out rate, calculated as the number of times the product was out of stock divided by the total number of product visits
InStrSpending	The average expenses associated with promoting the product in stores for the observed week
OnlineAdsSpending	The total amount of spend on online advertising
TvSpending	The average expenditure on television campaigns during the observed week
Radio	The number of radio advertisements or campaigns promoting the product for the observed week

All financial figures are standardized in USD, ensuring uniformity and comparability across the different variables.

The dataset provides a granular view of the interplay between sales volume and the dimensions of marketing—spanning digital, broadcast, and in-store campaigns, along with pricing strategies and stock availability. This information thereby presents a unique opportunity to dissect and understand the impacts of different marketing levers on sales, yielding insights that are actionable for optimizing marketing investment and operational decisions.

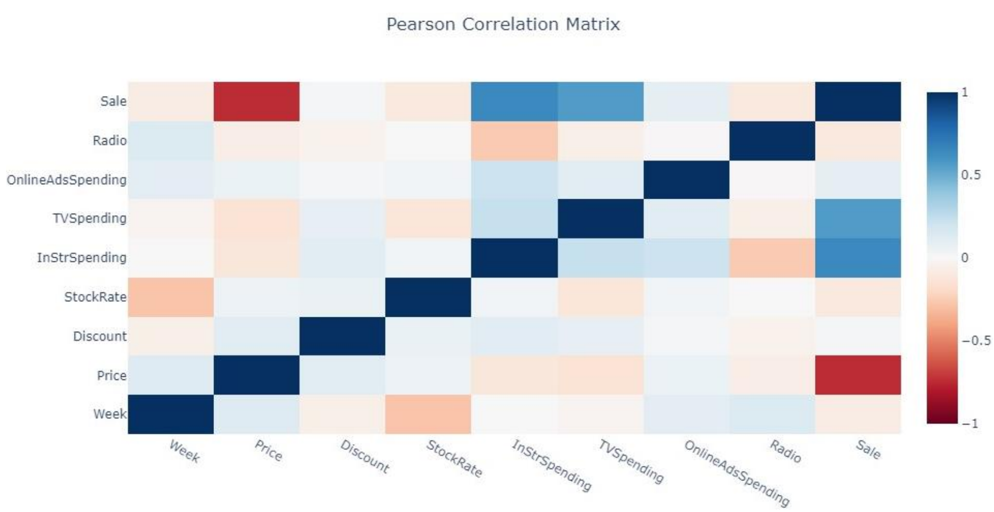
Exploratory Data Analysis

EDA will help us establish a baseline understanding of the data's structure, the relationships within, and the potential marketing opportunities.

Dataset Structure:

	Week	Price	Discount	StockRate	InStrSpending	TVSpending	OnlineAdsSpending	Radio	Sale
0	889	8.36	0.41	0.78	52.04	90.87	3187.92	454	264661
1	890	26.82	0.35	0.46	6.58	85.99	1694.51	1580	23336
2	891	21.37	0.36	0.64	41.62	45.92	786.62	996	123992
3	892	14.54	0.45	0.69	57.46	174.81	1223.56	177	294288
4	893	12.55	0.44	0.47	43.09	13.40	2504.79	1042	170391
...	...	...	...	...	...	...	...	...	...
99	988	26.22	0.18	0.81	19.69	32.60	2238.49	1840	26191
100	989	17.53	0.23	0.21	19.48	116.64	1564.64	2156	132714
101	990	29.23	0.49	0.26	22.79	39.33	1020.22	503	32894
102	991	23.13	0.14	0.33	13.84	37.71	1201.03	2902	39091
103	992	23.40	0.19	0.62	7.81	78.55	653.62	1820	37290

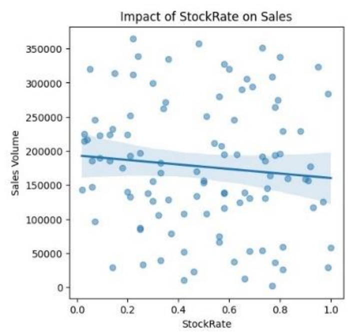
Correlation Matrix:

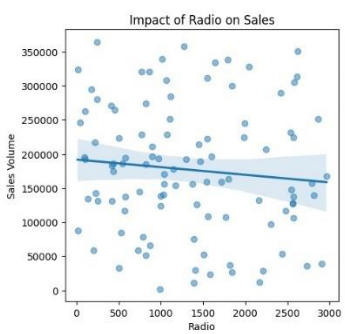


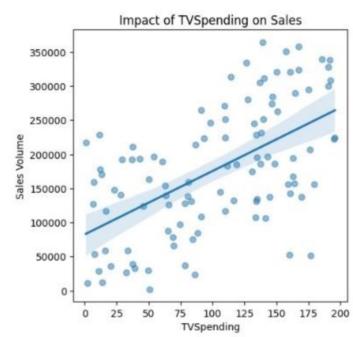
This reveals significant positive correlations between sales and In-Store Spending, as well as TV Spending, suggesting these channels may be potent drivers of sales; conversely, Price exhibits a moderate negative correlation with sales, indicating potential price sensitivity. Online Ads Spending and Radio show weaker correlations with sales, suggesting limited impact from these channels.

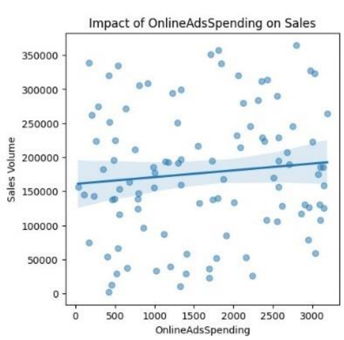
Scatter Plots:

They suggest a negative relationship between Price and sales, indicative of typical demand sensitivity, while positive associations for In-Store and TV Spending with sales imply their potential effectiveness in driving sales volume. The plots for rest show more dispersed points, indicating fewer clear relationships with sales.







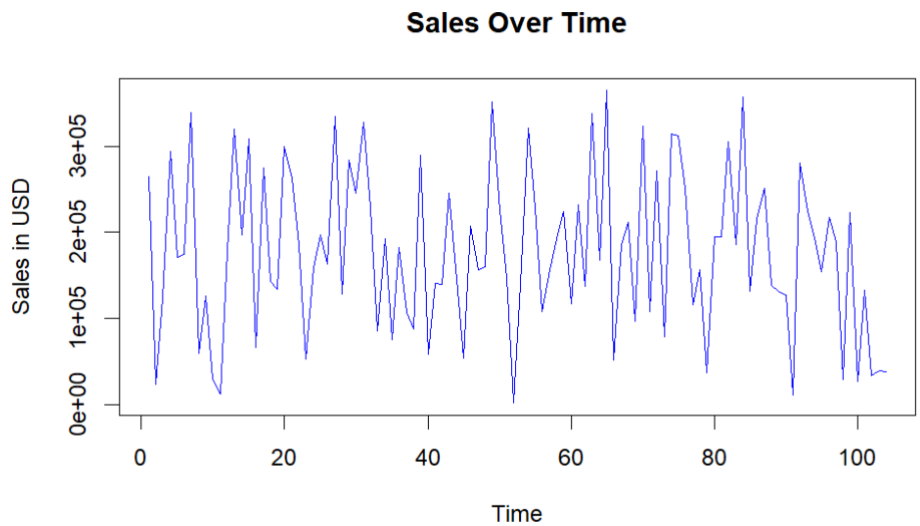


For a comprehensive view of other visualizations and statistical summaries, please refer to the appendices section of this report.

Preliminary Analyses and Stationarity Testing

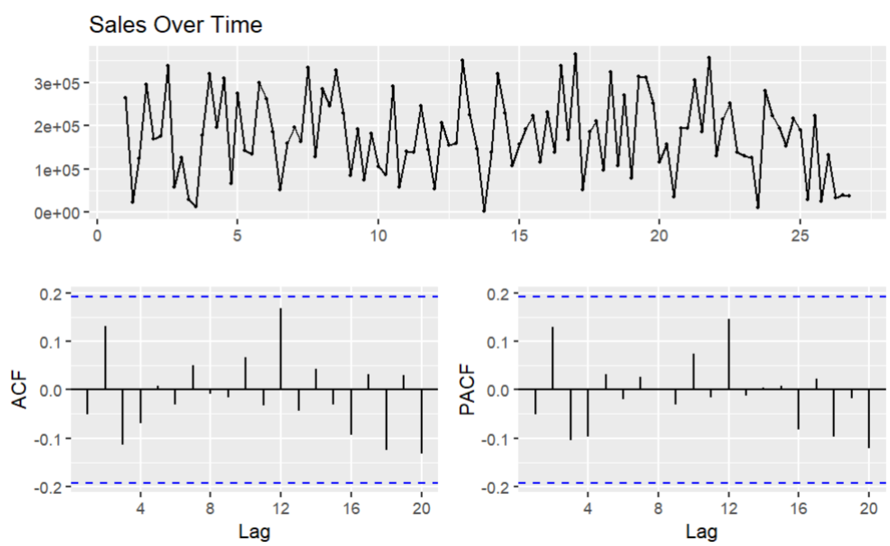
Time Series Plots:

They help us to visualize patterns, unusual observations, changes over time, and relationships between variables. Here, we have plotted the time series of Sales over time and no evident patterns are seen in this case.



ACF & PACF plots:

These plots help in autocorrelation analysis to help detect patterns and check for randomness. The plots for sales do not exhibit significant autocorrelation at most lags, suggesting that sales are largely random over time without strong dependencies on their past values.



These analyses were performed for all variables within the dataset to evaluate their individual time-dependent behaviors and the extent to which they are influenced by their own past values, setting a foundation for the VAR model selection. The graphs are available in the appendices section.

Data Transformation:

Logarithmic transformation was applied to the variables to stabilize variance, reduce skewness, and convert multiplicative relationships into additive ones.

Unit Root Stationarity Testing:

Stationarity implies that the statistical properties of our series, such as the mean and variance, remain constant over time.

Table below presents the results of the stationarity testing:

Variable	ADF (p-value)	PP (p-value)	KPSS (p-value)	Stationarity
Sale	0.01	0.01	0.1	Stationary
InStrSpending	0.01	0.01	0.1	Stationary
OnlineAdsSpending	0.01131	0.01	0.1	Stationary
TVSpending	0.01378	0.01	0.1	Stationary
Radio	0.02534	0.01	0.1	Stationary
Discount	0.01	0.01	0.1	Stationary
StockRate	0.01	0.01	0.1	Stationary
Price	0.01	0.01	0.1	Stationary

In our dataset, all variables are stationary. The stationarity of all variables ensures that the relationships we model are stable over time, which is a prerequisite for drawing meaningful conclusions from VAR analysis.

Vector Autoregressive Model

VAR model can provide valuable insights into how marketing activities across different channels affect sales and each other over time, making it a powerful tool for understanding the dynamics at play in a multivariate time series dataset like ours. It allows us to analyze the impact of various variables on sales.

The Akaike Information Criterion (AIC) helped determine the optimal lag structure, balancing model complexity and fit, which, in this case, was found to be 1 lag.

The output of VAR model is as follows:

	Dependent variable:							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sale.l1	-0.145	-0.168	-0.016	-0.054	-0.060	0.081	0.009	-0.098
	(0.174)	(0.211)	(0.185)	(0.215)	(0.174)	(0.182)	(0.185)	(0.234)
TVSpending.l1	0.093	0.104	-0.062	-0.053	0.001	0.044	0.093	0.002
	(0.095)	(0.115)	(0.101)	(0.117)	(0.095)	(0.099)	(0.100)	(0.127)
OnlineAdsSpending.l1	-0.066	0.032	0.022	0.275**	-0.078	0.052	0.018	-0.047
	(0.096)	(0.117)	(0.103)	(0.119)	(0.096)	(0.101)	(0.102)	(0.130)
Radio.l1	-0.027	-0.058	-0.010	-0.165	0.130	-0.065	-0.055	0.030
	(0.083)	(0.101)	(0.088)	(0.102)	(0.083)	(0.087)	(0.088)	(0.112)
InStrSpending.l1	0.292**	0.284*	0.194	0.043	0.255*	-0.192	-0.228	-0.165
	(0.138)	(0.167)	(0.147)	(0.170)	(0.138)	(0.144)	(0.146)	(0.186)
Discount.l1	-0.028	-0.010	-0.042	-0.119	-0.017	-0.054	0.069	-0.017
	(0.099)	(0.120)	(0.105)	(0.122)	(0.099)	(0.103)	(0.105)	(0.133)
StockRate.l1	0.090	-0.004	0.059	0.117	-0.107	-0.051	-0.032	-0.090
	(0.096)	(0.117)	(0.103)	(0.119)	(0.096)	(0.101)	(0.102)	(0.130)
Price.l1	0.059	-0.025	0.048	-0.070	0.102	0.022	0.008	-0.245*
	(0.097)	(0.118)	(0.103)	(0.120)	(0.097)	(0.102)	(0.103)	(0.131)
const	12.794***	5.155**	6.733***	6.837***	2.381	-2.199	-0.459	4.483
	(2.041)	(2.474)	(2.170)	(2.520)	(2.041)	(2.135)	(2.165)	(2.747)
Observations	103	103	103	103	103	103	103	103
R2	0.096	0.055	0.041	0.104	0.082	0.032	0.048	0.084
Adjusted R2	0.020	-0.025	-0.040	0.028	0.004	-0.050	-0.033	0.006
Residual Std. Error (df = 94)	0.848	1.028	0.901	1.047	0.848	0.887	0.900	1.141
F Statistic (df = 8; 94)	1.254	0.687	0.506	1.368	1.045	0.390	0.592	1.078
Note:					*p<0.1; **p<0.05; ***p<0.01			

Carry over effects:

Positive carryover effects suggest that investments in these areas may have a sustained impact beyond the initial period; for instance, a unit increase in TV spending, online ads spending, in-store promotions, and price leads to increases in their subsequent values by 0.104, 0.022, 0.255, and 0.008 units respectively. Conversely, negative carryover effects imply a decline in the variable's future value following an initial increase, signaling diminishing returns over time.

Cross Effects:

Customers who interact with one channel are more likely to interact with another, according to positive cross-over effects. For instance, if OnlineAdsSpending is increased by one unit, there's a suggestion (not statistically significant) that Radio might increase by 0.275 units in the next period. An increase in InStrSpending seems to potentially raise TVSpending by 0.284 units. This could be interpreted as an effect where investment in in-store promotions might lead to an increased emphasis or budget for television advertising.

Direct Effects:

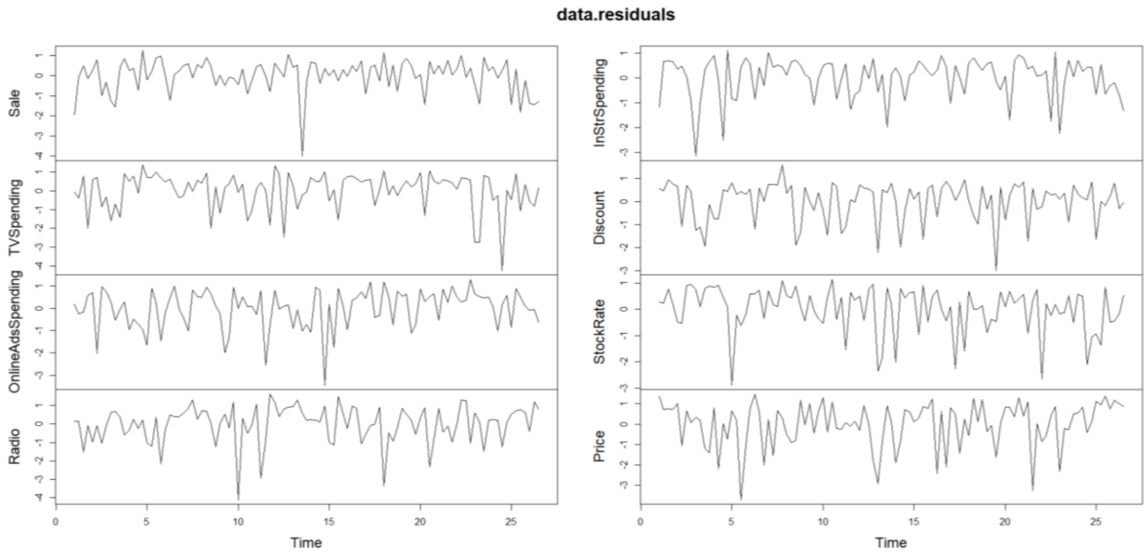
Even though many of the coefficients are not statistically significant at conventional levels, we can still observe the magnitude of direct effects, which could offer some marketing insights.

1. The positive impact on sales indicates TV ads are a powerful tool for driving sales and should be considered as a part of the marketing strategy.
2. In-store spending's substantial positive effect on sales underscores the importance of investing in the physical retail experience to boost purchase rates.
3. The non-significant, negative impact of online ad spend calls for a thorough review of digital marketing tactics to enhance their contribution to sales.
4. Discounts do not significantly affect sales, suggesting the need for a more targeted discount strategy, possibly focusing on customer segmentation and purchasing behaviors.
5. The price's non-significant positive relationship with sales offers the potential for strategic pricing without major sales volume loss, hinting at a less price-sensitive market or a strong brand perception.

The lack of statistical significance calls for further investigation, possibly through additional data. Hence, these findings should be taken as preliminary conclusions.

Residual Analysis:

This analysis is crucial as it helps ensure that the model is performing well and that the forecasts or insights it generates are reliable.



From these plots, we can make these observations:

1. The residuals do not display any obvious patterns, which indicates that the model has captured the underlying processes in the data reasonably well.
2. There seem to be some sharp spikes, which could be outliers or unexpected shocks to the system that the model hasn't accounted for.

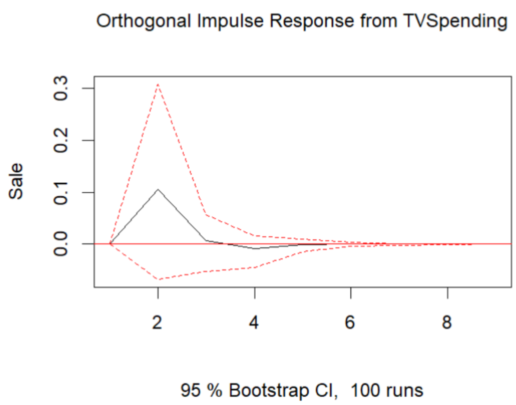
The outliers in residuals imply that marketing mix model can be improved by additional variables.

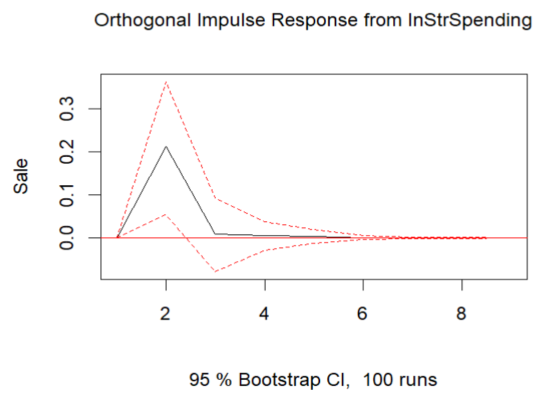
Granger Causality:

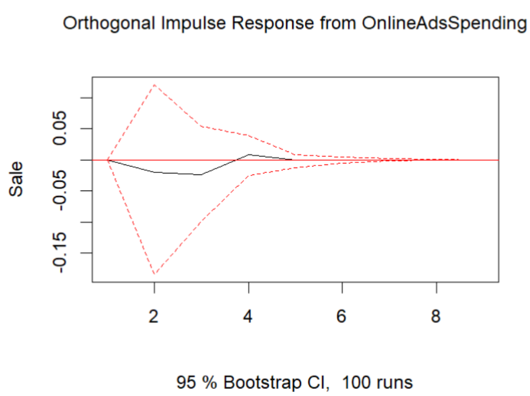
They are particularly important in VAR models to understand if past marketing activities and retail dynamics hold predictive power for future sales, beyond what is already explained by sales' own historical values. While there was no significant evidence of Granger causality between other variables with sales, the tests revealed that in-store spending does indeed Granger-cause sales and has instantaneous causality. This implies that changes in in-store advertising spending may precede and influence changes in sales.

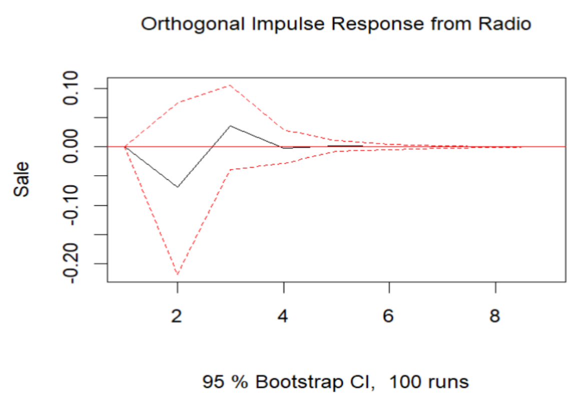
Impulse Response Function Tests:

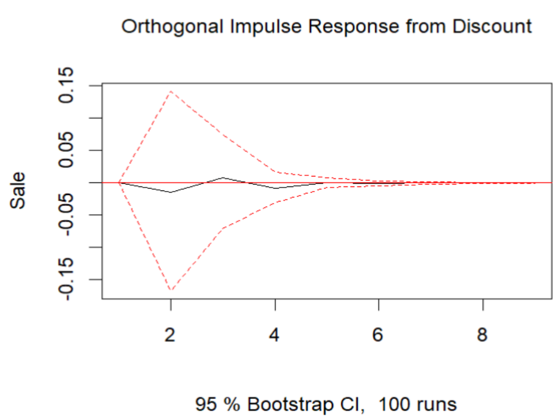
We can analyze the impact of the impulse series on the response series and how it progresses over time, here 8 periods ahead, through IRF plots below:

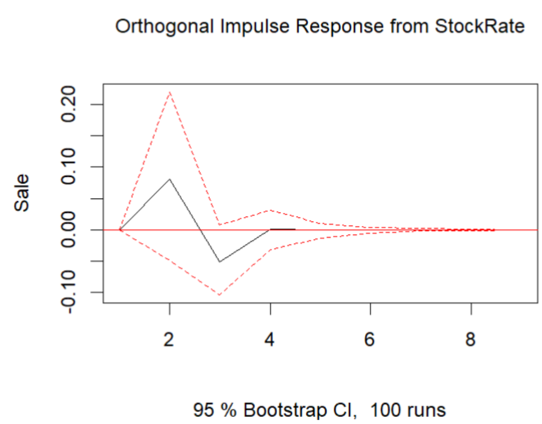


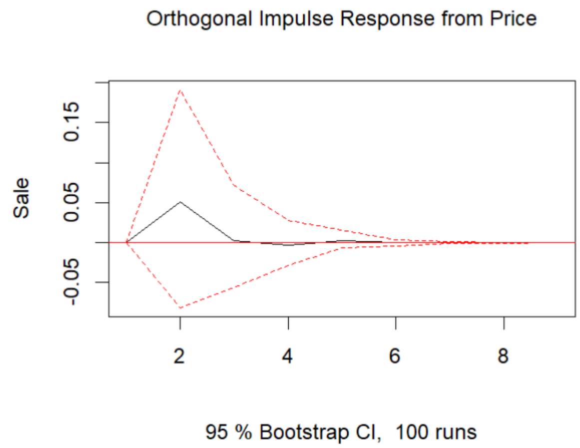












TV and In-Store Spending exhibit prompt positive impacts on sales, indicating their effectiveness as immediate sales drivers. In contrast, Online Ads, Radio shows an initial dip in sales, suggesting delayed customer response. However, the influence of all marketing activities, tends to neutralize within eight weeks, suggesting that while some strategies offer a quick boost, their long-term effectiveness on sales is limited.

Immediate & Long-term Effects

This analysis begins by examining the immediate impact of marketing and retail efforts on sales and tracks these effects over time. I used the t>1 criteria to determine coefficient significance and calculate long-term elasticities of variables. This suggests that t-statistic for a coefficient must be greater than 1 for it to be considered statistically significant.

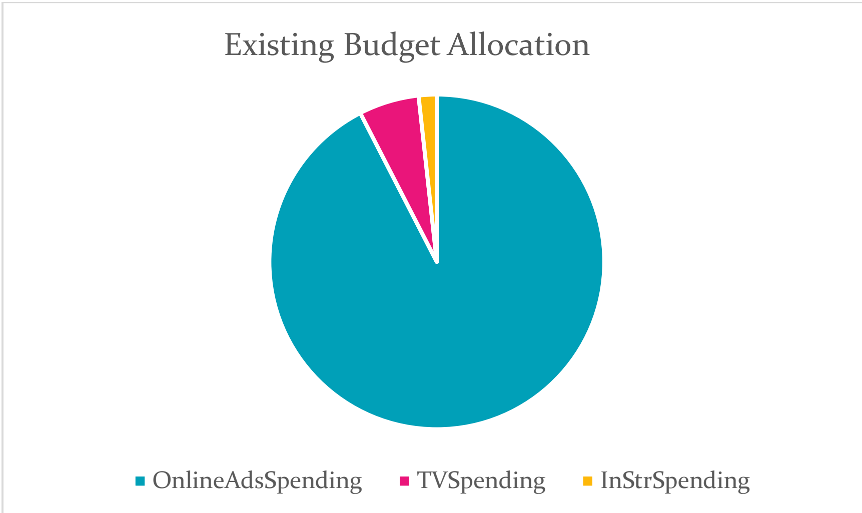
The values can be seen below:

Period	TV Spending	InStr Spending	OnlineAds Spending	Radio	Discount	StockRate	Price
1	0.1061484	0.2123423	0	0	0	0.0805438	0
2	0	0	0	0	0	-0.0514252	0
3	0	0	0	0	0		0
4	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0

In the first period, a shock in TV Spending and In-Store Spending yields immediate increases in sales by approximately 0.106 and 0.213 units, respectively. Stock Rate indicates an initial positive impact on sales which then turns negative in the second period, suggesting that stock availability might initially satisfy demand but persistent stockouts could damage sales. The zeroes in subsequent periods for all variables suggest that effects of shocks are not sustained over the long-term, emphasizing need for varied marketing efforts to maintain sales momentum.

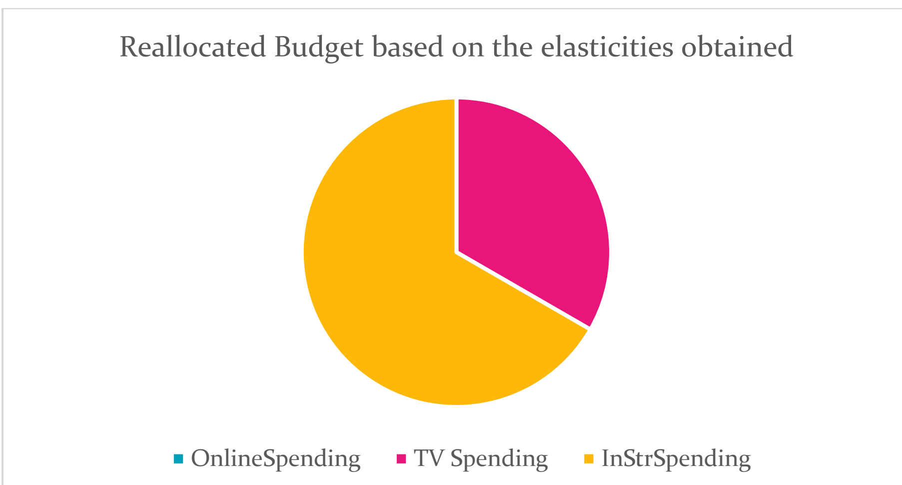
Budget Allocations

Now that our time series analysis is complete, it is necessary to analyse the existing budget prior to putting forward any marketing recommendations.



The current budget allocation is predominantly directed towards online marketing, constituting 92% of total spend. However, analysis through VAR and subsequent IRF tests suggest that company may not be harnessing budget to its fullest potential. The data also indicates that TV and In-Store spending is 6% and 2% of the budget respectively. However, they demonstrated a significant and immediate positive impact on sales. Given these insights, it is necessary to reassess the budget distribution.

The reallocated budget from the elasticities obtained through our analysis is as follows:

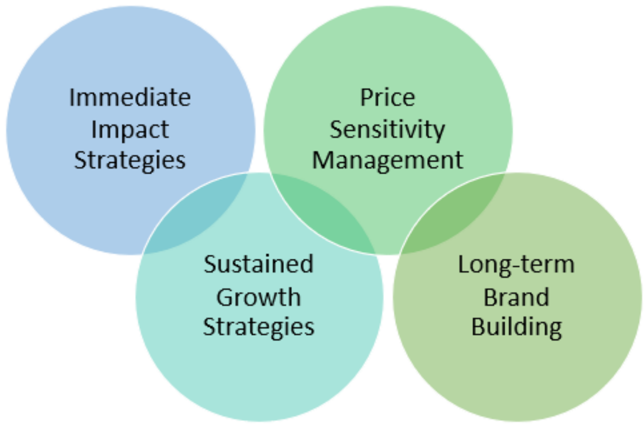


The reallocated budget suggests to have 66.67% for In-store Spending, about 33% for Tv Spending, and no budget for Online Ads. Since, the effects are seen only for period 1 and not for long-term, it won't be feasible to diminish the role of online marketing yet.

Hence, rather than making drastic cuts to online advertising, company shall maximize immediate returns from TV and In-Store while preserving the long-term engagement and brand-building potential of online marketing efforts. It's a strategic decision that balances the need for immediate ROI with the importance of long-term planning and customer engagement.

Strategic Recommendations

The aim here is to help the company enhance immediate and long-term sales of their product while fostering a resilient brand presence and cultivating a robust customer base, using the results from our analysis.



Immediate impact Strategies

TV Advertising: To optimize the immediate sales impact, allocate a larger budget for crafting captivating TV commercials that focus on hours with substantial audience.

Advancement of the In-Store Experience: To bring about rapid sales growth, invest in developing an immersive in-store environment while making benefit of physical retail's intuitive advantage.

Inventory Management: The immediate effect of Stock Rate is notable. Invest in predictive analytics for more intelligent inventory planning, making sure that stock rate optimization is in line with customer demand trends, to avoid the detrimental effects of stockouts. `

Sustainable Growth Strategies

Diversification of Digital and Broadcast: Although channels such as Radio and Online Ads do not exhibit rapid elasticity, they shouldn't be overlooked. Sustain allocating a significant amount for online advertisements and maintain radio advertising to promote the brand.

Over time, these platforms may become increasingly important as they boost consumer engagement and brand exposure. Long-term planning may include SEO, content marketing, and other techniques that build brand equity and sales over time.

Price Sensitivity Management

Dynamic Pricing Implementation: With the understanding that sales are not significantly hindered by price changes, there's an opportunity to optimize pricing to improve profit margins or competitive positioning. adjust prices in real-time in response to various external market conditions, such as changes in consumer demand, competitor pricing, or inventory levels.

Selective Discounting: To maximize sales volume and profitability, implement discount tactics that are specific to the consumer segments that show the greatest adaptability to price adjustments. Offer discounts specifically to these high-elasticity segments to maximize the total sales volume and profitability. Because these consumers are more price-sensitive, targeted discounts can effectively stimulate additional purchases that might not have occurred at the original price.

Long-term Brand Building

Community Investment and Content: Allocate resources to community engagement and content marketing in order to foster organic growth and strengthen brand equity.

Loyalty Programs: Launch loyalty programs with the goal of raising customer retention rates and lifetime value.

Adaptive Budget Reallocation: Agile budget management allows for flexible budget modifications to increase returns by introducing a methodical framework for continuous marketing performance assessment across all channels. Set aside a budget reserve to experiment with new marketing channels and approaches.

Adopting this dual strategy of short-term action and long-term vision, the company will be in line with our suggested marketing initiatives. They can build a loyal client base and a lasting brand identity in addition to sales growth. The firm may maintain the agility and efficiency of its marketing operations by keeping a flexible budget that can be dynamically altered based on market response. By doing this, they can ensure that they are constantly at the forefront of marketing innovation and prepared to take advantage of the shifting customer landscape.

References

1. Understanding VAR models: A comprehensive guide. Medium. Available at: https://medium.com/@data-overload/understanding-var-models-a-comprehensive-guide-782ded47dcf4

2. Chapter 16.1 - Vector Autoregressions (n.d.) Econometrics with R [Online]. Available at: https://www.econometrics-with-r.org/16.1-vector-autoregressions.html

3. Vector Autoregressive models VAR(p) models.
   Available at: https://online.stat.psu.edu/stat510/lesson/11/11.2

4. Vector Autoregressions. Available at:
   https://www.princeton.edu/~mwatson/papers/Stock_Watson_JEP_2001.pdf

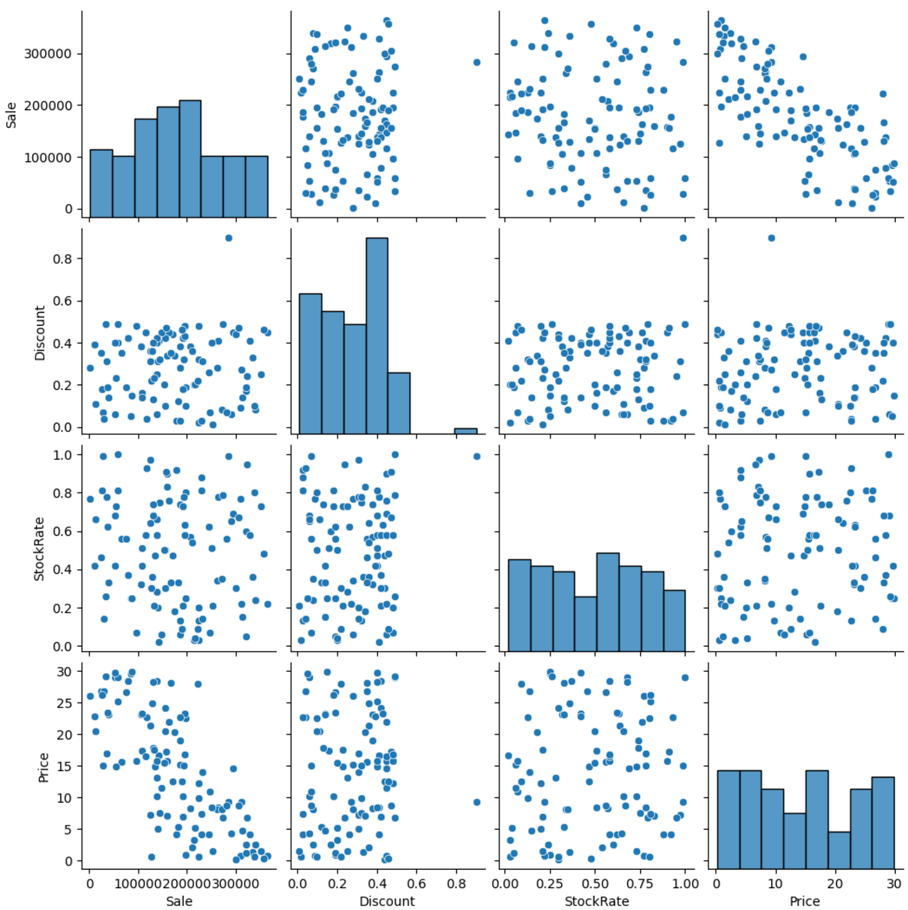
Appendix

1) Summary statistics

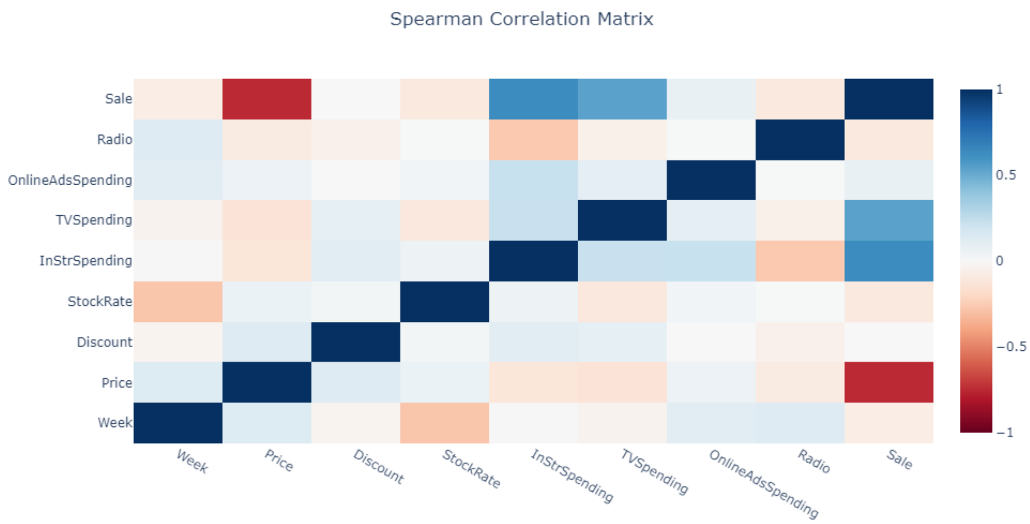
```
summary_statistics
```

	Week	Price	Discount	StockRate	InStrSpending	TVSpending	OnlineAdsSpending	Radio	Sale
count	104.000000	104.000000	104.000000	104.000000	104.000000	104.000000	104.000000	104.000000	104.000000
mean	940.500000	14.095769	0.279519	0.492981	30.499231	101.796538	1618.552788	1324.557692	177013.153846
std	30.166206	9.208078	0.160299	0.278881	16.070590	57.398769	967.064827	844.873436	94049.811187
min	889.000000	0.210000	0.010000	0.020000	0.660000	1.060000	32.150000	13.000000	1992.000000
25%	914.750000	6.807500	0.140000	0.250000	17.320000	50.450000	735.007500	578.500000	116720.750000
50%	940.500000	14.675000	0.310000	0.510000	33.755000	109.725000	1627.405000	1156.000000	172585.000000
75%	966.250000	22.672500	0.410000	0.732500	43.000000	147.355000	2517.437500	1985.500000	245591.750000
max	992.000000	29.900000	0.900000	1.000000	57.460000	195.670000	3187.920000	2962.000000	364483.000000

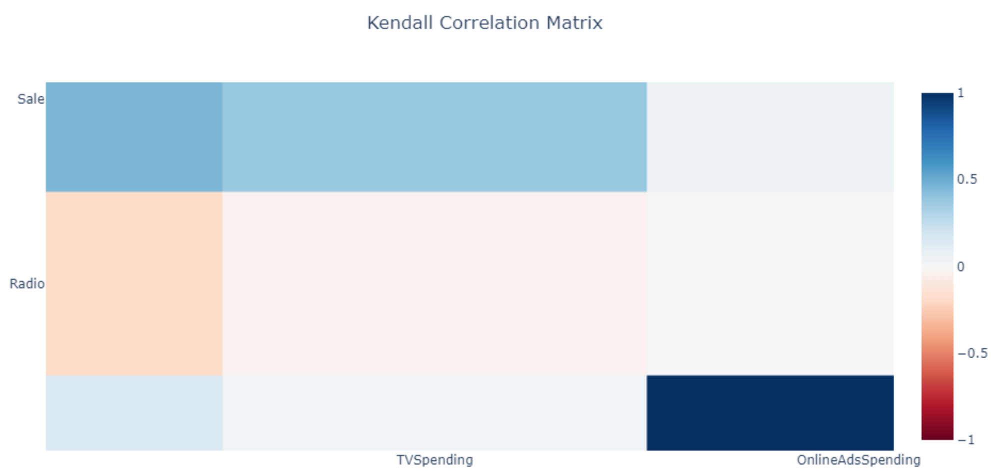
2) PairPlots



3) Spearman Correlation Matrix

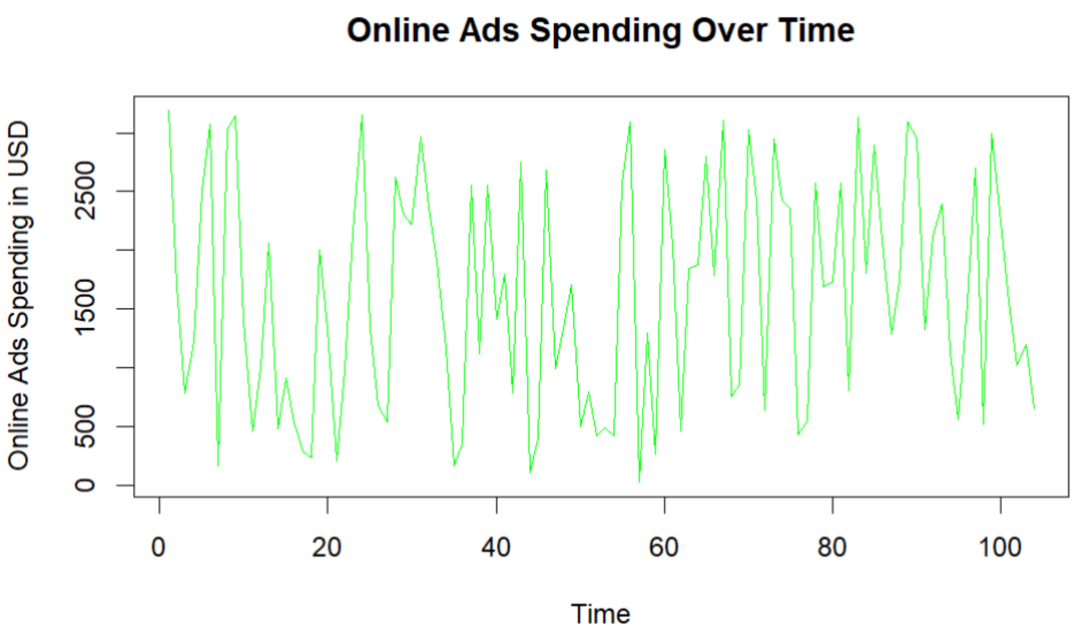


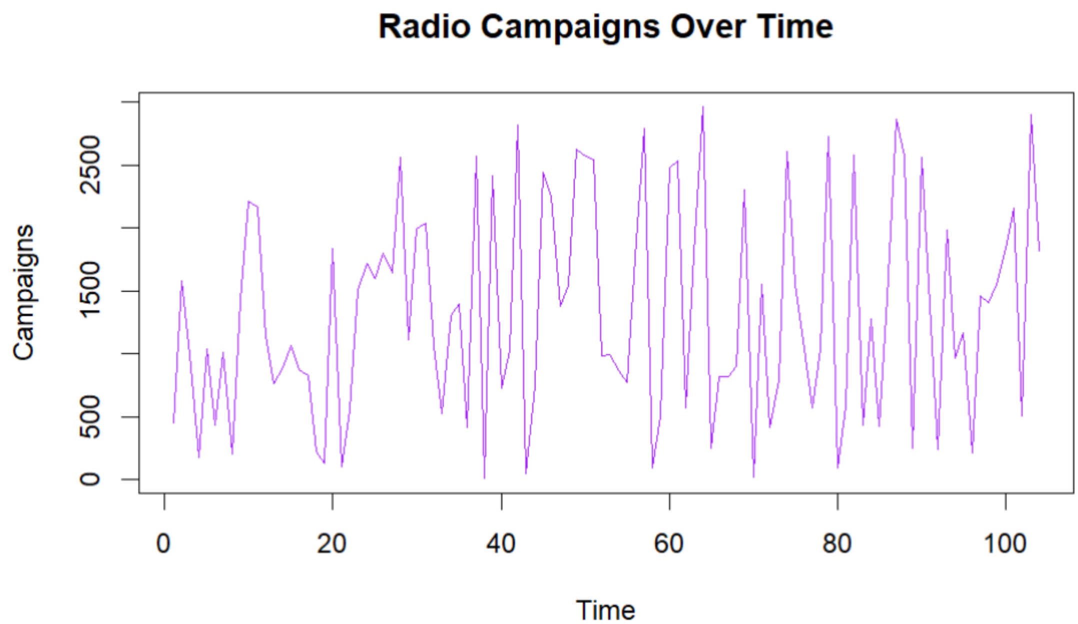
4) Kendall Correlation Matrix

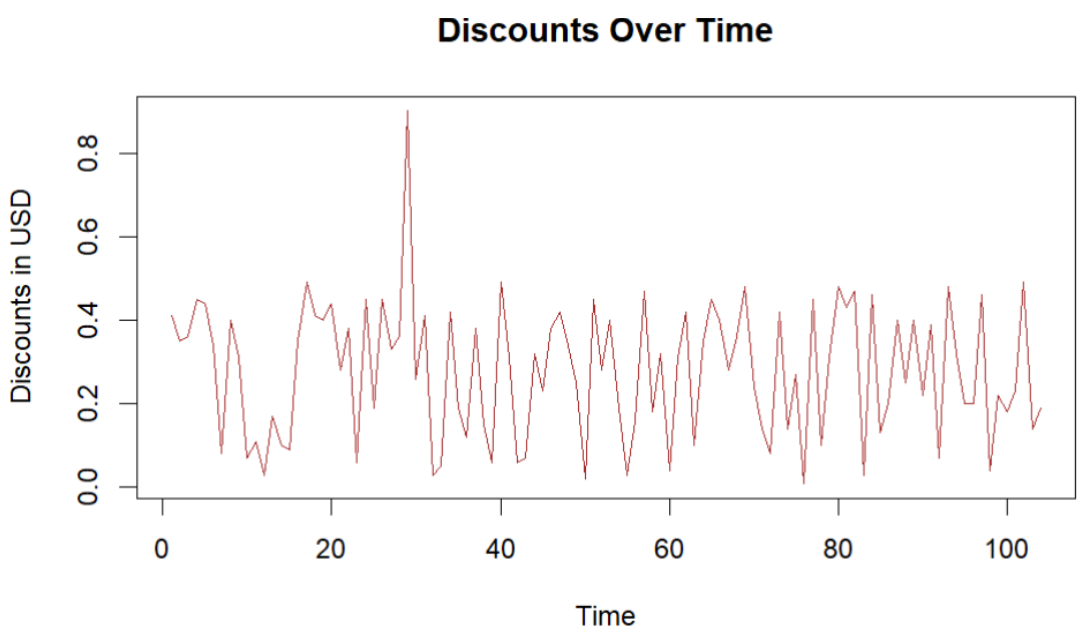


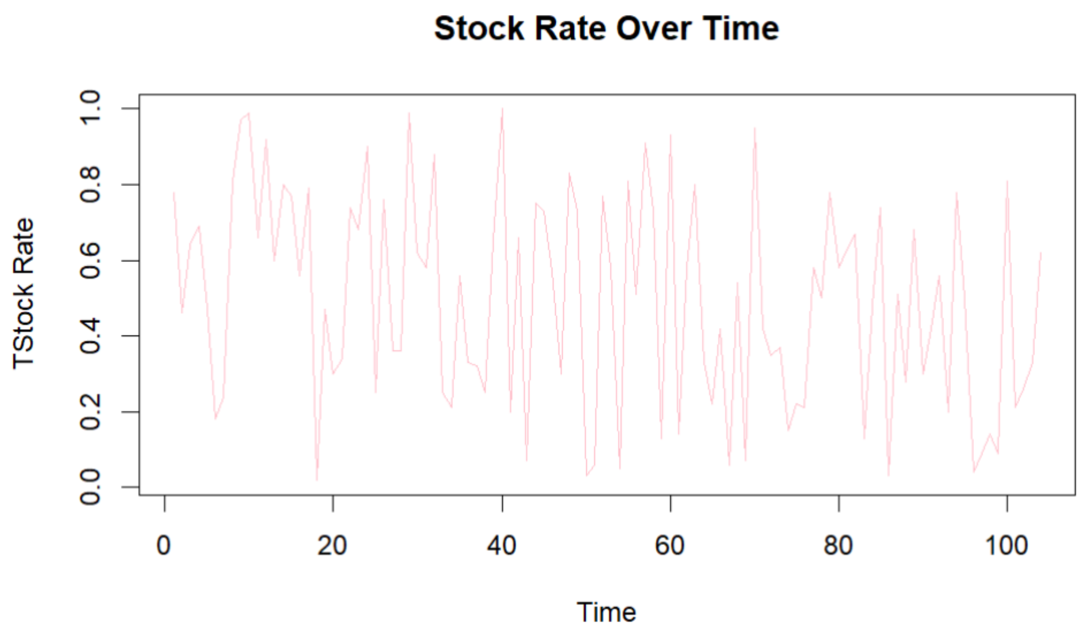
5) Time series plots

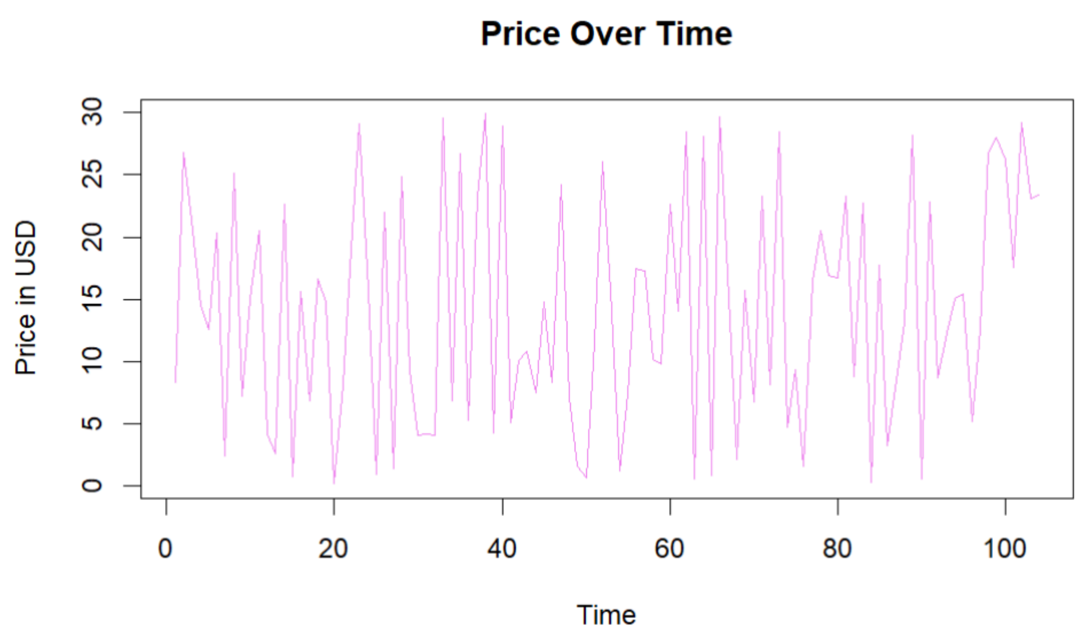

In-store Spending Over Time
In-Store Spending in USD
0
10
20
30
40
50
0
20
40
60
80
100
Time


Online Ads Spending Over Time
Online Ads Spending in USD
0
500
1500
2500
0
20
40
60
80
100
Time

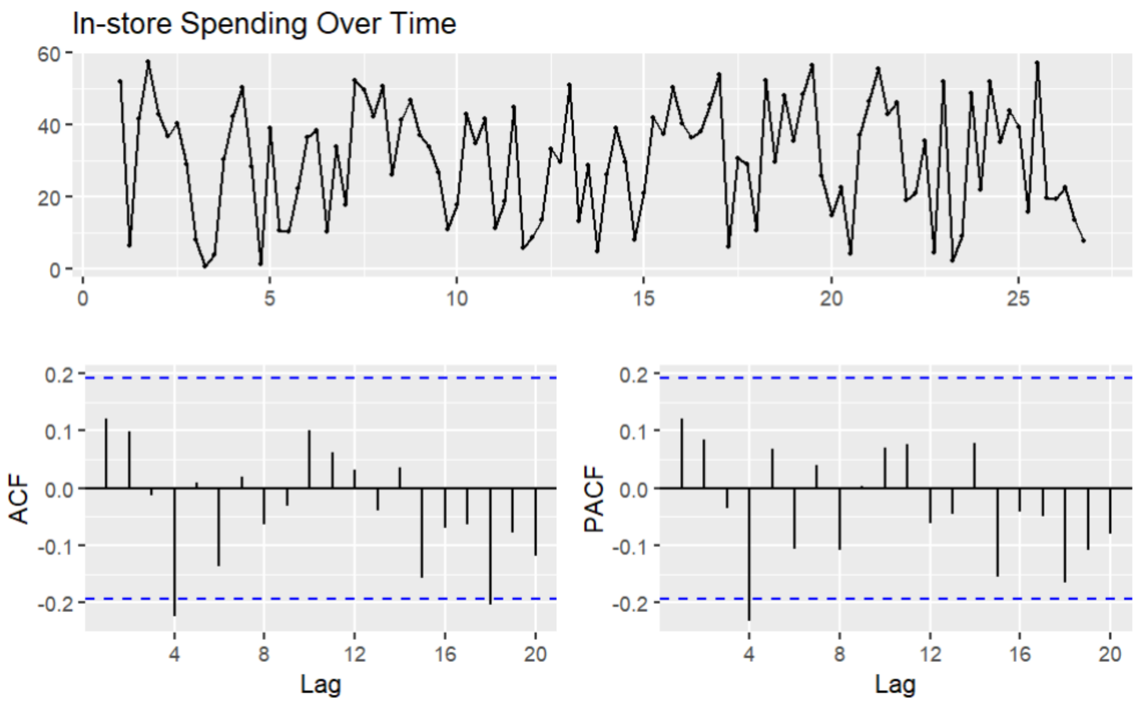

Radio Campaigns Over Time
Campaigns
0
500
1500
2500
0
20
40
60
80
100
Time

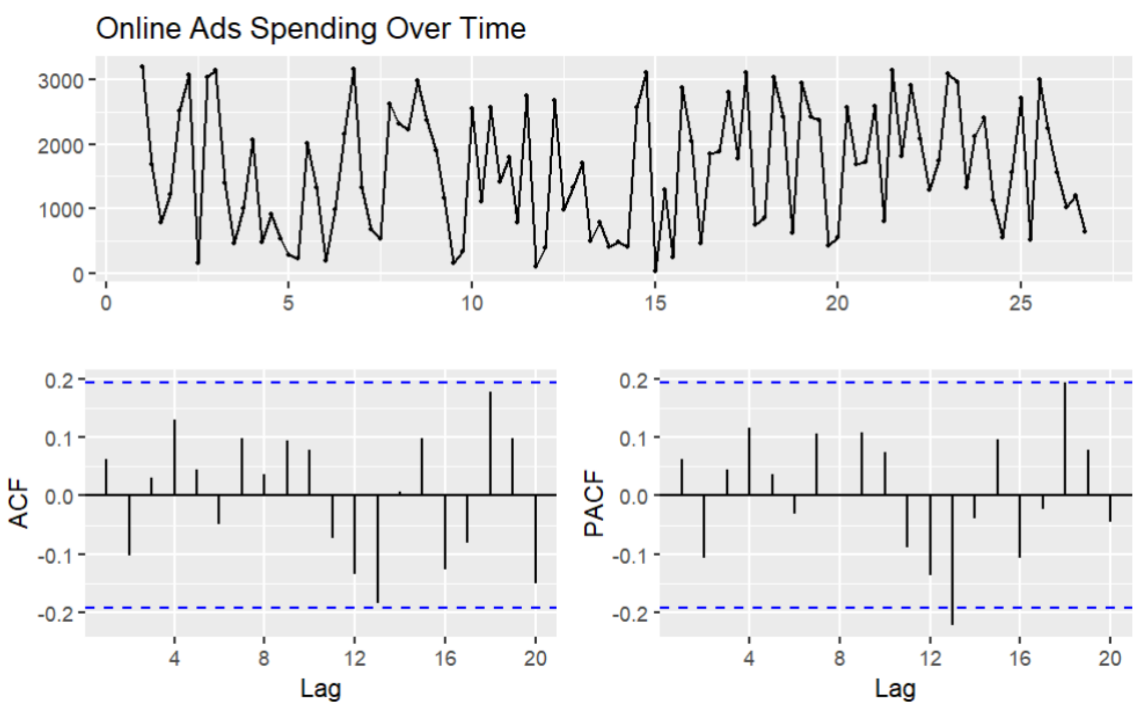

Discounts Over Time
Discounts in USD
0.0
0.2
0.4
0.6
0.8
0
20
40
60
80
100
Time

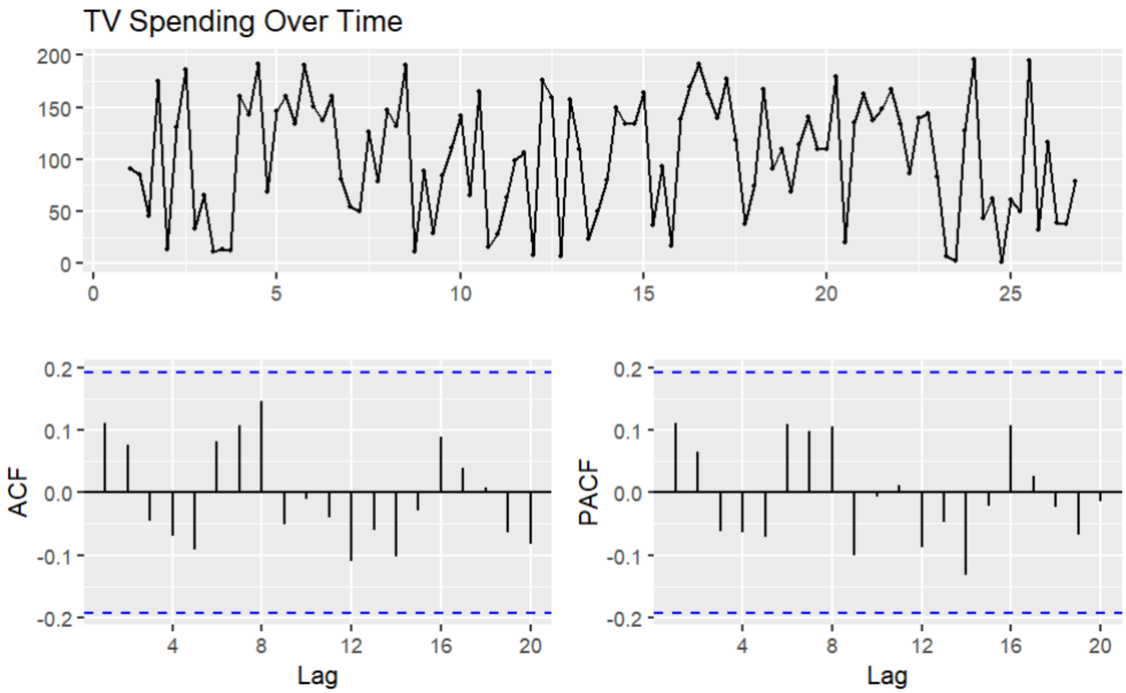


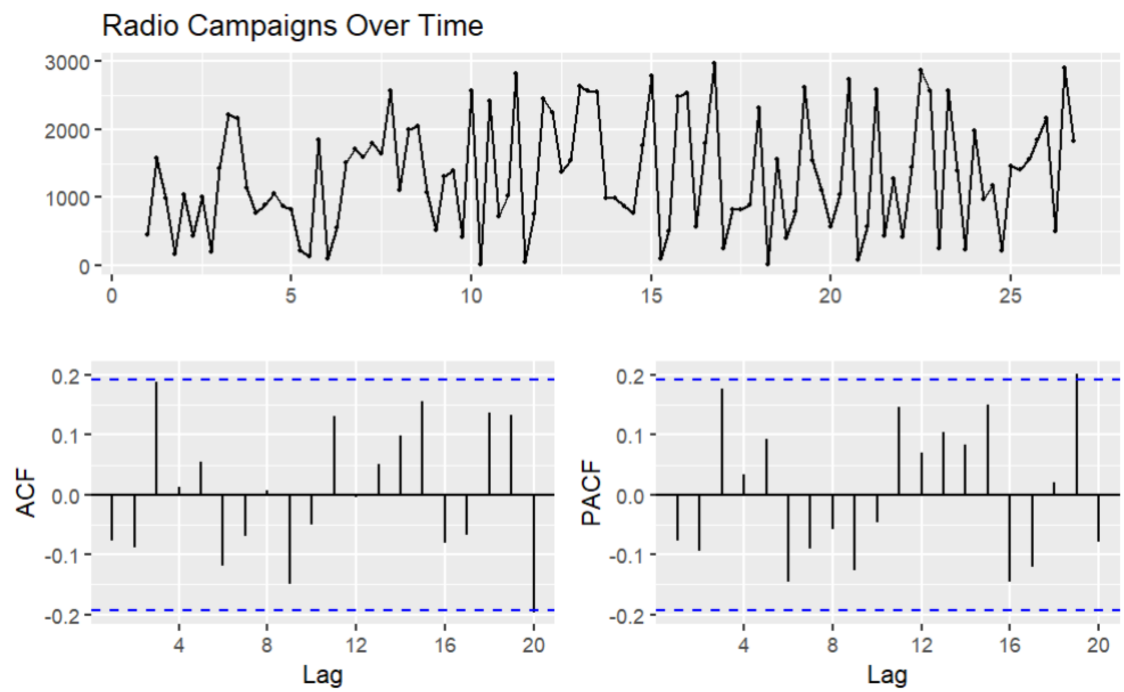


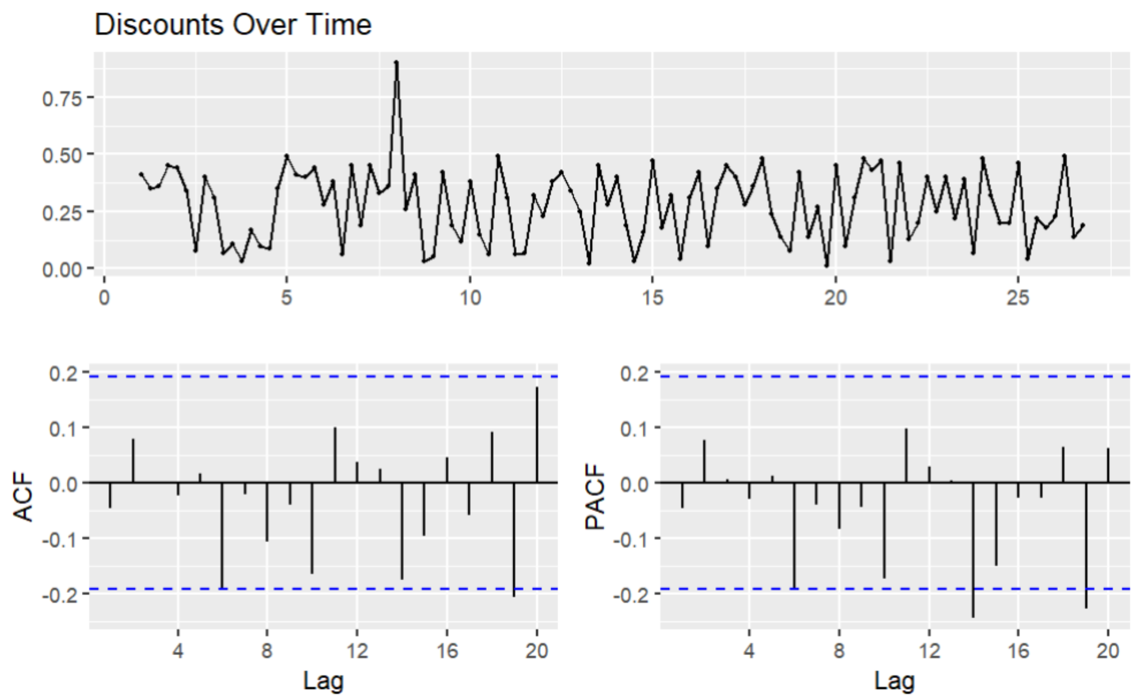
6) ACF and PACF Plots

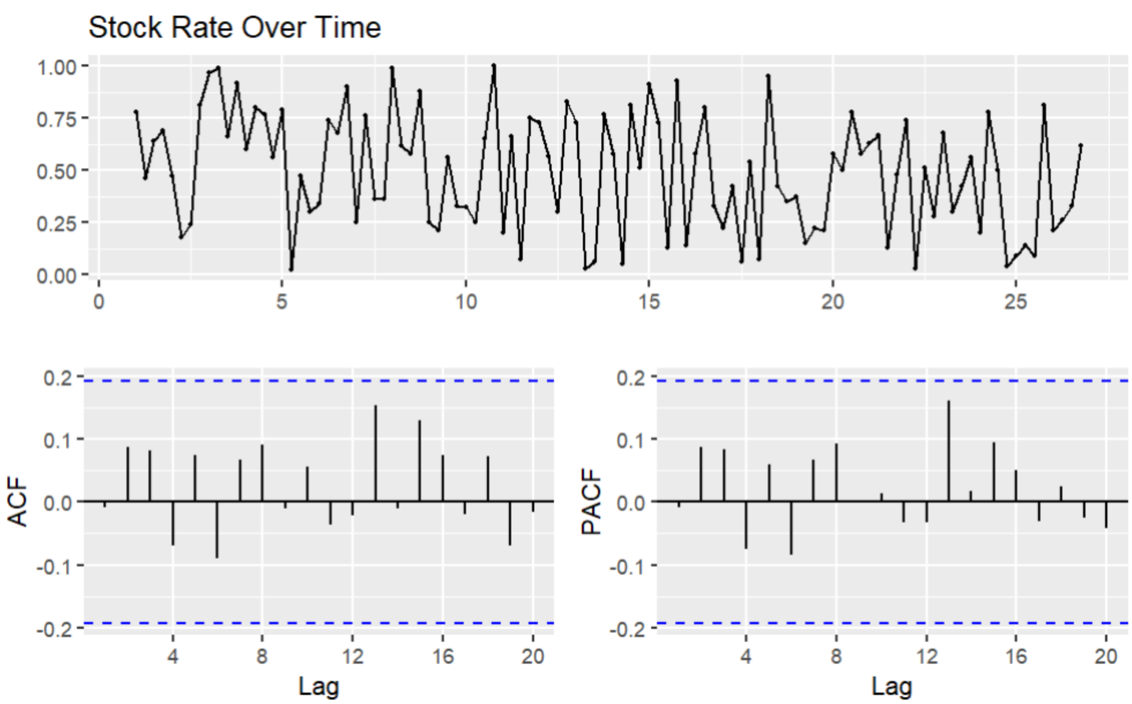

In-store Spending Over Time
ACF
PACF
Lag
Lag

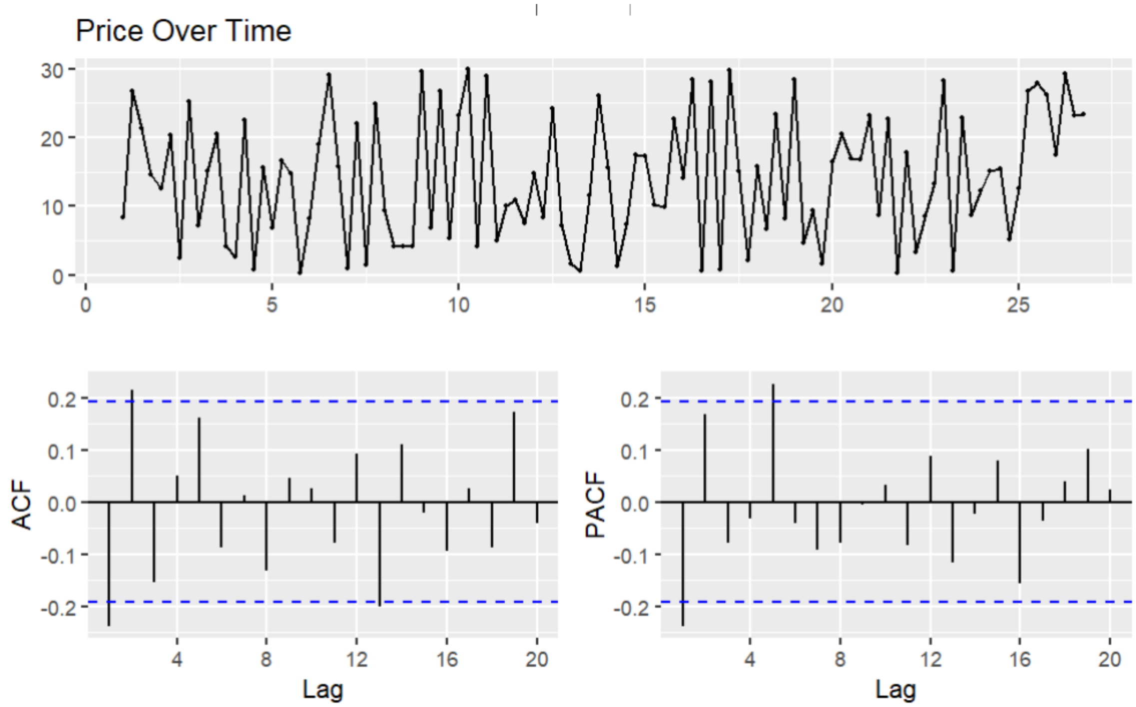

Online Ads Spending Over Time
ACF
PACF
Lag
Lag


TV Spending Over Time
ACF
PACF
Lag


Radio Campaigns Over Time
ACF
PACF
Lag


Discounts Over Time
ACF
Lag
PACF
Lag


Stock Rate Over Time
ACF
Lag
PACF
Lag


Price Over Time
30
20
10
0
0
5
10
15
20
25
0.2
0.1
0.0
-0.1
-0.2
ACF
4
8
12
16
20
Lag
PACF
Lag