



Internship Report in Business Analytics

as a partial fulfillment of the degree of
Bachelor in Business Administration

Implementing a Marketing Mix Modeling Analysis And Budget Optimization Using CRISP-DM Methodology

By:

Mohamed Majd Ayadi

Academic Supervisor:

Dr. Lassaad Moubarki

Professional Supervisors:

Mr. Fares Ben Mahmoud

Mr. Mens Ayed



Tunis, 2023 - 2024

Abstract

This work was aimed to provide MASS Analytics, a marketing mix modeling company, with insights about the drivers of website traffic for an IT solution provider company called company X. MASS Analytics is a software company that provides MMM solutions and consultancy services to its clients. The project was intended to determine the effect of company X marketing mix on the number of visitors to their website using three years worth of data provided at the national level. In this project, we used two models to run a marketing mix modeling(MMM) project: the first was a log linear model using the company's software, MassTer, and the second was a multiple linear regression model using python. Additionally, we were able to measure the percentage contribution of various factors including media channels, events and others to determine the ones that significantly drive growth for website visitors. After comparing the outputs of each model and selecting the one with a better fit for our project, we run optimization and prediction for year 2024. We created two optimal scenarios aiming to improve website traffic and media budget allocation after receiving the media budget plan for the current year.

Results from this MMM project proved that the log linear model is more suited for this analysis than the multiple linear regression model. Moreover, marketing variables such as : Informative and Educational Books Submissions, Website publications, Newsletter and Youtube outperform other channels in driving website visitors. For the optimal scenarios 1 & 2 built within MassTer with different budget constraints settings, they yield respectively 63.96% and 15.13% increase in traffic growth with the same budget planned for 2024.

Keywords

Marketing Mix Modeling, Website Visitors, KPI, Log Linear Model, Multiple Linear Regression

Contents

Approval	i
Declaration	ii
Acknowledgements	iii
Abstract	iv
List of Abbreviations	viii
List of Tables	ix
List of Figures	x
Introduction	1
1 Internship Overview	2
1.1 Company Presentation	2
1.1.1 General Presentation	2
1.1.2 Mission, Vision and Values	3
1.1.3 Industry and Market Structure	4

1.1.4	Products and Services	7
1.2	Internship Description	9
1.2.1	Job Description	9
1.2.2	Problem Statement	10
1.2.3	Motivation	10
1.3	Conclusion	11
2	Literature Review and Methodology	12
2.1	Literature Review	12
2.1.1	Marketing Mix Summary	12
2.1.2	Marketing Mix Modeling	13
2.1.3	Marketing Mix Modeling Workflow	14
2.1.4	Multiple Linear Regression	14
2.1.5	Log Linear Modeling	15
2.1.6	Conclusion	16
2.2	Methodology	17
3	Implementation and Results	19
3.1	Business Understanding	19
3.2	Data Understanding	20
3.2.1	Data Collection	20
3.2.2	Data Exploration	21
3.3	Data Preparation	28

3.3.1	Data Selection	29
3.3.2	Data Cleaning	29
3.3.3	Data Integration	29
3.3.4	Data Transformation	30
3.4	Model Building	34
3.4.1	Modeling Technique	35
3.4.2	Model Building	35
3.4.3	Model Assessment	36
3.5	Model Evaluation	37
3.6	Results and Findings	37
3.6.1	Log Linear Model (using MassTer)	37
3.6.2	Multiple Linear Regression	46
3.6.3	Recommendations	50
3.7	Log Linear Model Optimization and Prediction	53
3.7.1	Prediction	53
3.7.2	Optimization	53
3.7.3	Prediction and Optimization Recommendations	57
	General Conclusion	58
	References	59
	Appendices	60

List of Abbreviations

- TBS: Tunis Business School
- MMM: Marketing Mix Modeling
- CRM: Customer Relationship Management
- ROI: Return on Investment
- CPC: Cost per click
- CPM: Cost per Thousand Impressions
- CPA: Cost per Acquisition
- DR: Diminishing Returns
- KPI: Key Performance Indicator
- FB: Facebook
- CTR: Click Through Rate

List of Tables

2.1 Elasticity vs Marginal Effects	16
3.1 Website Visitors Year over Year percentage change	22
3.2 Approved Format by MassTer	30
3.3 Outputs of Log Linear Model(using MassTer)	40
3.4 Log Linear Model's Statistical Metrics	41
3.5 Website visitors Decomposition 2021-2023 using Log Linear Model	43
3.6 Group percentage variation and contribution to KPI (2022-2023)	45
3.7 Media Channel Cost per acquisition 2021-2023	46
3.8 Multiple Linear Regression's Statistical Metrics	48
3.9 Website visitors Decomposition 2021-2023 using Multiple Linear Regression	50
3.10 Optimization Scenario 1	56
3.11 Optimization Scenario 2	57
3.12 Metric Explanations	60
3.13 Data collected	61

List of Figures

1.1	Market Analytics Market Size, By Deployment Model, 2021-2032, (USD Billion)	4
1.2	Advertising Media Budget	5
1.3	MassTer Software Logo	8
1.4	MassTer Insight Logo	8
2.1	MMM workflow	14
2.2	Crisp-DM Methodology	17
3.1	Monthly Website Visitors	21
3.2	Monthly Facebook Followers	23
3.3	Monthly Instagram Reach	24
3.4	Yearly Media Budget Percentage Allocations	24
3.5	Monthly Google Paid Search clicks	25
3.6	Monthly Google Display Impressions	25
3.7	Monthly Youtube Clicks	26
3.8	Monthly Linkedin Impressions	26
3.9	Monthly Newsletter Clicks	27

3.10 Monthly Website Publications	27
3.11 Events and Webinars	28
3.12 Monthly Informative and Educational Book Submissions	28
3.13 Youtube Clicks Decay Transformation	31
3.14 Google Display Impressions Diminishing Returns Transformation	32
3.15 Linkedin Lag Transformation	32
3.16 Christmas Dummy Variable	33
3.17 Informative and Educational Book Submissions flat data Window Variable	33
3.18 May22_Website_drop Dummy Variables	34
3.19 Instagram Reach Splitter Transformation	34
3.20 Actual vs Model chart for Log Linear Model	38
3.21 Website Visitors Waterfall Chart 2022-2023	44
3.22 Actual vs Model for Multiple Linear Regression	47
3.23 An example of a diminishing returns curve	54
3.24 Optimization based on diminishing return curves	55

Introduction

This study, framed within the CRISP-DM methodology, aims to evaluate and compare two different regression techniques for a marketing mix modeling project: a log linear model developed using the company's software and a multiple linear regression model built in python. We will evaluate both models from statistical and business perspectives to determine which technique provides the most accurate and robust results. Upon identifying the superior model, we will perform optimization and prediction analysis for year 2024 with two different scenarios(constrained and non-constrained scenario). The results showed that the log linear model is the most appropriate model in our situation. Moreover, outputs from the two optimal scenarios revealed that the number of visitors will rise by 63.96% in the non limited scenario and by 15.13% in the constrained one with the same budget spend planned for 2024.

Our analysis also explores the impact of various internal and external factors including paid and organic online media, events, newsletter, website publications and other variables on the number of website visitors. By using a three years of data at the national level, we identified the most effective marketing channels. The results showed that the contribution of Informative and Educational Books, Website publications, Newsletter and Youtube outperform other channels in driving website visitors. Through assessing the impact of these channels via marketing mix modeling, we aim to significantly improve the company X marketing effectiveness in order to boost their website traffic.

General Conclusion

During this MMM project, we compared the performance of a multiple linear regression model and a log linear model at explaining the variations in website traffic. Both models are statistically robust. However, from a business standpoint, the log linear model proved to be superior due to its ability to account for complex interactions between variables and provide more accurate insights into the factors driving website visits.

To accomplish this, we followed a CRISP-DM methodology that allowed us to systematically analyze the impact of various internal and external factors on website traffic. By applying the two models, we gained a comprehensive understanding of the contribution of various variables to our KPI, enabling us to identify the ones with the greatest influence on website visitors variation.

The insights and recommendations provided will help in making better data driven decisions to enhance overall website performance, ensuring that marketing efforts are aligned with strategic goals. By taking advantage of the superior capabilities of the log linear model, we are well prepared to execute prediction and optimization for year 2024, thus significantly enhancing the long-term growth of the company's website visitors. Offering strategic advice on the best way to allocate marketing budget across various media channels can effectively increase website visits without raising the media budget spend.

While the selection of the log linear model is needed to achieve results that are both statistically and commercially significant, it's important to note that no model is free of limitations. Some issues faced when using this model were related to reporting the individual contribution of each channel. The multiplicative form of the log linear model presents some challenges in separating contributions. This is when approximation methods are used to convert the multiplicative model into an additive one allowing for a clear distinction between the individual contributions of each variable. Obviously, this approximation will yield a decomposition error that will be added to the standard OLS error. Our role as analysts is to make sure that the decomposition error is as small as possible.