**Professional Portfolio: Capstone Project**

A cityscape with text overlay

Description automatically generated

Maximizing Marketing Impact: Returns on investment optimization through Marketing Mix Modelling

**Team members:**

Abhishek Rajendra Mahale

Sukhdeep Singh

Sonjeet Kaur

Nehal Vadoliya

# **1.0 Introduction**

In the words of marketing pioneer John Wanamaker, "Half the money I spend on advertising is wasted. The trouble is I don't know which half." This sentiment reflects a longstanding challenge faced by businesses worldwide: the inability to accurately measure the effectiveness of their marketing efforts. In today's fiercely competitive landscape, where maximizing returns on investment (ROI) is paramount, this lack of insight presents a significant obstacle to success.

To address this challenge, our project focuses on leveraging Marketing Mix Modelling (MMM), a powerful analytical framework that allows businesses to dissect and optimize their marketing strategies. By delving into the intricate interactions between various marketing elements—such as channels, pricing strategies, and promotions—MMM provides invaluable insights into their individual contributions to sales performance. Armed with this knowledge, businesses can make data-driven decisions to allocate resources more effectively and maximize their ROI.

Through a combination of advanced statistical analysis and machine learning techniques, our project aims to develop a MMM framework tailored to the specific needs of our organization. By harnessing the power of MMM, we seek to empower our organization with the tools and insights needed to navigate the complexities of modern marketing effectively. Ultimately, our goal is to help businesses overcome the age-old dilemma highlighted by Wanamaker and unlock new opportunities for growth and success in today's dynamic marketplace.

# **2.0 Project Description**

# 2.1 Framework:

As taught in our professional portfolio course, most of the analytical projects can be solved using the CRISP-DM framework (Cross-Industry Standard Process for Data Mining). We have followed the same for our project as well. It consists of six phases:

1. Business Understanding
2. Data Understanding
3. Data Preparation
4. Modeling
5. Evaluation
6. Deployment

# 2.2 Hypothetical Problem Statement/ Business Understanding:

Skyline Technologies is an e-commerce company in Toronto-Canada selling electronic gadgets. They put a lot of money into marketing, sometimes running big sales events like "Tech Day Extravaganza." Now, they're planning their marketing budget for the next year, including spending on ads, online campaigns, and deals. But looking back over the last year, they see that their marketing money didn't really boost sales like they hoped. They're considering cutting the budget or spending it smarter to make more money next year.

As part of the marketing analytics team, our job is to figure out what went wrong and how to fix it. We're going to create a plan using a marketing model to see what worked and what didn't in the past year. Then, we'll use that info to suggest the best ways to spend the marketing budget next year, aiming for more sales and happier customers.

# 2.3 Business Data Understanding:

We had a limitation to procure publicly available data for our project, which has been stated in the proposal document. Hence, we found a medium.com website from where we could get simulated data, link for it has been provided in the appendices.

As per the 7Ps of Marketing we needed data from the below sources:

* Sales Transactions Data: Detailed information on product sales, including quantities sold, revenue generated, and transactional details.
* Marketing Expenses: Data on marketing expenditures across various channels and activities, such as advertising, promotions, sponsorships, and digital campaigns.
* Product Data: Attributes of the products being sold, such as SKU (Stock Keeping Unit) details, product categories, variations, and pricing tiers.
* Price Data: Historical pricing information for products, including regular prices, promotional prices, discounts, and pricing changes over time.
* Distribution Data: Data on the distribution channels used to sell products, including information on retailers, wholesalers, online platforms, and geographic coverage.
* Promotional Data: Details of promotional activities undertaken, such as discounts, coupons, special offers, and seasonal promotions.
* Competitor Data: Information on competitor activities, pricing strategies, promotional campaigns, and market share dynamics.
* Market Data: External market factors that may influence sales performance, such as economic indicators, consumer trends, seasonality, and demographic data.
* Customer Data: Customer demographics, preferences, purchasing behavior, loyalty program participation, and customer lifetime value.
* External Factors: Additional external factors such as weather patterns, holidays, events, and regulatory changes that may impact consumer behavior and sales performance.

We could get data for the following(These are not full data columns, only relevant have been listed down):

1. Product:

•# of units sold •Delivery days and SLAs •Categories/sub categories & Transactions

1. Price:

•Gmv •Product mrp

1. Place:

•Pin code •Order Payment Type Place (& Time) •Weekofthe year– seasonality •Holiday/Events

1. Promotion:

•Marketing Channel Investments •Customer sentiment (NPS) •Discounts •Adstock

# 2.4 Data Cleaning & Transformation:

**Treating Duplicates and Ensuring Data Consistency**: Duplicate values were meticulously eliminated to maintain data integrity, ensuring each entry remained unique. Columns underwent rigorous verification to uphold uniqueness, minimizing the risk of redundancy and inconsistencies. Non-numeric data within key columns like "GMV" and "Pincode" was removed to enhance data clarity and accuracy.

**Validating Data Scope and Business Logic**: Occurrences of '\N' and data falling outside the specified timeframe were filtered out, focusing solely on relevant data for analysis. Furthermore, columns with a high proportion of null values were eliminated, streamlining the dataset and aligning with business requirements.

**Data Standardization and Integrity Maintenance**: The format of essential columns such as "order\_date" and "GMV" was standardized for uniformity, facilitating seamless analysis. Rows containing negative values or inconsistencies in product MRP, GMV, and units were systematically removed to uphold data integrity.

**Categorization and Segmentation for Targeted Analysis**: Discount percentages were computed, and items were categorized as Luxury or Mass Market based on pricing percentiles, enabling effective segmentation. Irrelevant columns were pruned to focus solely on key variables, enhancing the dataset's relevance and usability. Note here that we only have based the entire analysis on just one product category i.e. Camera Accessory just to limit our scope for targeted analysis.

**Outlier Detection and Removal:** Outliers falling below Q1-1.5IQR or above Q3+1.5IQR were identified and eliminated to ensure statistical robustness, enhancing the reliability of subsequent analyses.

**Data Aggregation and Transformation for Analysis**: Week identifiers were generated to facilitate time-based analysis, and daily Order Data was aggregated to a weekly level for a higher-level view of trends. Master data frames were scaled and partitioned into separate train and test datasets for comprehensive model training and evaluation processes.

By meticulously adhering to these steps, the dataset was cleansed, standardized, and prepared for subsequent analysis, ensuring robust insights and informed decision-making.

# 2.5 Feature Engineering:

In order to facilitate comprehensive analysis, several key transformations and calculations were applied to the dataset. Firstly, a "Week#" column was generated from the order date, allowing for the segmentation of data into weekly intervals, thereby aiding in temporal analysis. Additionally, products were categorized as either Luxury or Mass-market based on whether their GMV value surpassed the 80th percentile, facilitating segmentation for more targeted analysis.

Furthermore, various metrics were computed to provide deeper insights into sales dynamics. The List Price for each transaction was derived by multiplying the GMV by the Units, offering a comprehensive view of the monetary value of products sold. Additionally, the Discount% was calculated for each transaction, allowing for the quantification of applied discounts and their impact on sales performance.

Moreover, temporal patterns in sales behavior were identified by flagging Payday Weeks and Holiday Weeks. Assigning a value of 1 to weeks containing Payday or holidays, and 0 otherwise, enabled the analysis of sales trends around these significant temporal markers. Furthermore, the effectiveness of pricing strategies was assessed through the calculation of SMA# (Sales-Marketing Alignment), comparing List Price to product MRP and expressing the difference as a percentage.

To capture temporal dependencies in the data, Lag Variables were incorporated, including lagged values for key performance indicators (KPIs) by 1, 2, and 3 days. Additionally, 3-week and 5-week Simple Moving Averages (SMAs) were calculated for advertising media channels, Net Promoter Score (NPS), and Stock Index, aiding in the identification of trends and smoothing out fluctuations. Lastly, Adstock Values for all advertising media channels were determined by applying an ad stock rate of 60%, capturing the cumulative effect of past advertising efforts on current sales dynamics. These transformations and calculations collectively lay the groundwork for more robust and insightful analysis of sales performance and marketing effectiveness.

# 2.6 EDA & Visualization:

We have used Tableau desktop to create some EDA visuals for ease of access at later point in time to present the data well.

1. Weekly Net promoter Score(NPS) vs Discount % A graph with blue lines

   Description automatically generated

Figure 1

Nothing significant to note here while discounts can impact customer behaviour and potentially influence Net promoter Score(NPS). Not in our case.

1. Weekly investment for Different Advertisement Channels

A graph with colorful lines

Description automatically generated

Figure 2

We can clearly see that the company spends a lot more on Sponsorship media for Advertising

Considering it might raise the revenue.

1. Product Types vs Units-Sold

A close-up of a graph

Description automatically generated

Figure 3

We can observe that there are two major product types i.e. Luxury and Mass-market which we created as an additional KPI earlier. For Mass Market we can see camera accessory sells the most and similarly Home Audio sells the most for Luxury brands.

1. Weekly Sales vs Monthly Sales

A graph with a line graph

Description automatically generated

Figure 4

The week when the sales were the highest was week 42 and it was a thanks giving week. We can observe the monthly trend as well.

1. Product Category granularity with respect to Gross Merchandise Value(GMV)

A graph with different colored squares

Description automatically generated with medium confidence

Figure 5

Lens , Gamepads and Homeaudio Speakers have the highest GMV amongst all products.

1. GMV vs Discount%

A graph with a line

Description automatically generated

Figure 6

Higher the aggregated discount higher is the GMV as we can observe which holds true.

1. Discount bins vs GMV

A graph with multiple colored bars

Description automatically generated with medium confidence

When the average discount percentage falls between 10% and 20%, the GMV is highest.

8.

A green pie chart with percentages

Description automatically generated

# EDA & Data Visualization:

# **11.0 References**

* Gupta, Sunil. (2019). "Marketing Analytics: Data-Driven Techniques with Microsoft Excel." John Wiley & Sons.
* Hanssens, Dominique M., et al. (2020). "Market Response Models: Econometric and Machine Learning Insights." Springer.
* Neslin, Scott A., et al. (2019). "Sales Promotion and Retailing: Allowance for Marketing Actions." Marketing Science.
* Wang, Wes. (2018). "Marketing Analytics: A Practical Guide to Real Marketing Science." Kogan Page.
* <https://youtu.be/sYn1wuO9BDM>
* [Marketing Mix Modeling White Papers - MASS Analytics (mass-analytics.com)](https://mass-analytics.com/marketing-mix-modeling-white-papers/)