APPLICATIONS OF CUSTOMER SEGMENTATION AND BAYESIAN MODELLING IN MARKET MIX MODELS

Submitted in partial fulfilment of the requirements for the degree of

Bachelor of Technology
in
CSE with Business Systems

by

Aditya Malpani	Rachita Singh	Akshita Jain
20BBS0062	20BBS0115	20BBS0127

Under the guidance of

Prof. Anny Leema A

School of Computer Science and Engineering,

VIT, Vellore.

May, 2024

DECLARATION

We hereby declare that the thesis entitled "Applications of Customer Segmentation and Bayesian Modelling in Market Mix Models" submitted by us, for the award of the degree of *Bachelor of Technology in CSE with Business Systems* to VIT is a record of bonafide work carried out by us under the supervision of Prof. Anny Leema A.

We further declare that the work reported in this thesis has not been submitted and will not be submitted, either in part or in full, for the award of any other degree ordiploma in this institute or any other institute or university.

Place : Vellore

Date :

Signature of the Candidate(s)

Aditya Malpani (20BBS0062)

Rachita Singh (20BBS0115)

Akshita Jain (20BBS0127)

CERTIFICATE

This is to certify that the thesis entitled "Applications of Customer Segmentation and Bayesian Modelling in Market Mix Models" submitted by **Aditya Malpani (20BBS0062), Rachita Singh (20BBS0115) & Akshita Jain (20BBS0127)**, **School of Computer Science and Engineering**, VIT, for the award of the degree of *Bachelor of Technology in CSE with Business Systems*, is a record of bonafide work carried out by him / her under my supervision during the period, 01. 12. 2023 to 30.04.2024, as per the VIT code of academic and research ethics.

The contents of this report have not been submitted and will not be submitted either in part or in full, for the award of any other degree or diploma in this institute or any other institute or university. The thesis fulfils the requirements and regulations of the University and in my opinion meets the necessary standards for submission.

Place : Vellore

Date :

Signature of the Guide

Internal Examiner

External Examiner

Head of the Department

Programme

ACKNOWLEDGEMENTS

No project is completed by just the members. There always are numerous guides who enlighten their path. Ours is no different. The success and the final outcome of this project required a lot of guidance and assistance, and we are extremely privileged to have got this all along with the completion of this project. All that we have accomplished is only due to excellent supervision and assistance, and we would not forget to thank them.

First and foremost, we would like to express our gratitude to our academic institution, Vellore Institute of Technology, where we were constantly encouraged to think creatively and leverage the power of technology in various domains. Their commitment to advancing knowledge provided the foundation upon which this research was built. We thank them for providing us the opportunity to build this project and improve our understanding of real-world problems.

We are deeply thankful to our project supervisor, Prof. Anny Leema A, for her expert guidance, insightful feedback, and unwavering encouragement at every stage of the project. It was her suggestions during the early stages of out project that lead us to this topic. Her mentorship and expertise in machine learning played a crucial role in shaping the direction and methodology of our research.

We are also grateful to our esteemed panel, who guided us during the both the reviews and highlighted important aspects. Their thought-provoking questions led us to improvisions in our project, which ultimately improved our outcomes.

This research would not have been possible without the collective efforts and support of all those mentioned above. We are truly grateful for their contributions and commitment to the success of this endeavour.

Aditya Malpani

Rachita Singh

Akshita Jain

EXECUTIVE SUMMARY

Market mix modelling (MMM) is a statistical analysis technique used by businesses to understand how different marketing variables affect sales and other key performance indicators. It typically involves the use of frequentist statistical techniques, such as regression analysis, to quantify the impact of each of these elements on sales or other performance metrics. Bayesian models are also becoming increasingly popular as they are useful in incorporating prior knowledge and handling uncertainty in a more nuanced way compared to traditional frequentist statistical methods. A Bayesian framework approach helps incorporate prior knowledge within the model description, as opposed to the traditional Frequentist media mix modelling approach that fails to account for uncertainties that might occur within multiple advertisement channels. Ad stock and Saturation effect are two main advertising principles that constitute as the driving force for such models. Understanding each media channel's contribution to ROI and optimising budget at channel level helps businesses understand the impact of their campaigns better and drives strategic and optimal allocation of marketing finances. Customer Segmentation is yet another technique that helps marketers identify specific customer groups and understand their requirements. This helps businesses to run targeted campaigns that can cater to maximum segments and increase conversions for their products/services. Understanding how different media channels impact different customer segments can improve ROI and enhance customer experience. We propose a system that groups customers into different groups based on their annual income and spending score, analyse media data and the impact of each media channel on those segments, and suggest an optimal channel-wise budget allocation via a gradient based optimisation technique to maximise sales.

Keywords: Market mix modelling, frequentist statistical techniques, Bayesian models, Ad stock, Saturation effect, ROI, optimal budget, customer segmentation, gradient based optimisation

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS AND ABBREVIATIONS

MMM	**M**arket **M**ix **M**odelling
WCSS	**W**ithin **C**luster **S**um of **S**quares
MCMC	**M**arkov **C**hain **M**onte **C**arlo
ROI	**R**eturn **O**n **I**nvestment
SLSQP	**S**equential **L**east **S**quares **Q**uadratic **P**rogramming

1. INTRODUCTION

1.1 Objectives

Our project aims to understand market mix modelling and consumer psychology, study existing methods, and implement a significant model that enhances insights that help marketers and analysts make strategic decisions regarding advertising campaigns and products. Our end goal caters to budget optimization for different marketing channels to maximize sales, get a better reach, and improve customer acquisition. This can be achieved by fulfilling the following objectives:

- Identify the effect of different media channels on different customer segments.
- Generate media insights, and study media channels' Return on Investment (ROI) contribution to sales via Bayesian MMM.
- Perform budget optimization and allocation for greater ROI.
- Perform a comparative analysis of the Bayesian model with a frequentist method (linear regression).

1.2 Motivation

Businesses typically allocate a significant portion of their resources towards managing customer relations and advertising. Within a company, there are various channels of advertising, all aimed at maximizing customer reach or encouraging their engagement. It is essential for companies to promote their products or services to establish their identity and enhance brand recognition, leading them to allocate a specific budget for marketing endeavours. Without a clear understanding of the correlation between marketing efforts and sales, there's a risk of significant revenue losses. Determining how to distribute funds among marketing channels like social media, TV, direct mail, or radio can be daunting. To prevent guesswork and capital drain in budget allocation, optimization is crucial. This Involves estimating the effectiveness of each marketing channel to ensure the marketing budget is utilized optimally.

Market Mix Modelling is a solution that allows marketers and analysts to assess the influence of marketing elements on product or service performance. We propose a model based on Bayesian inference, and determine the individual media channel contributions to sales and their respective ROI using Bayesian Regression. A gradient based optimisation algorithm (SLSQP) is used to determine the optimal budget allocation across media channels. Additionally, we address a crucial gap by analysing the impact of different media channels on distinct customer segments and conducting a comparative analysis with frequentist models like OLS regression during implementation.

1.3 Background

In the digital age, where data drives decisions, the sanctity and accuracy of data have become paramount. However, with the advent of stringent data privacy regulations and the increasing use of ad blockers by consumers, the gap between captured data and actual user behaviour has widened, posing a significant challenge for marketers. This discrepancy, which can be as high as 30% in conversion attribution, undermines the effectiveness of tracking tools and leaves marketers navigating in the dark, unable to gauge the real impact of their campaigns. This misalignment not only obscures the visibility of campaign performance but also leads to inefficient budget allocation across various online mediums, culminating in substantial financial waste and missed opportunities. The marketing budget of a company is a precious asset, but with numerous advertisement streams and marketing channels, assessing their impact becomes challenging. It's essential to comprehend how marketing efforts influence sales, as this understanding can guide the development of market strategies that yield optimal results, i.e. maximizing sales with the most accurate use of the budget in hand.

Market Mix Modeling emerges as a robust alternative in this scenario, offering a way to eliminate the limitations posed by data privacy issues and the inherent shortcomings of attribution models. MMM employs statistical techniques to analyse historical data and quantify the impact of various marketing elements, both online and offline, on sales or other key performance indicators (KPIs). By aggregating data over time, MMM provides insights into how different factors, including marketing channels, economic conditions, seasonality, and competitive actions, contribute to business outcomes.

	Purpose	Example of business question
1	To measure efficiency	How efficient was our spending on TV last year?
2	To simulate	How would our sales change if we spent more/less on TV next year?
3	To optimize media budget	How should my media budget be allocated to maximize sales?

Fig 1.1 Purpose of MMMs

Unlike multi touch attribution, which relies on user-level data, MMM operates at a higher, more aggregated level, making it less susceptible to data privacy constraints. This characteristic allows MMM to incorporate offline marketing efforts into its analysis, offering a comprehensive view of the marketing mix's effectiveness. MMM can guide strategic decisions on budget allocation, channel optimization, and marketing mix strategy by identifying the incremental impact of each marketing input on sales.

2. PROJECT DESCRIPTION AND GOALS

2.1 Survey of the existing system

S.no	Name	Year	Journal	Authors	Technique	Limitations
1.	Bayesian Methods for Media Mix Modelling with shape and funnel effects	2023	Arxiv	Javier Marín (NextBrain.ai)	The authors use Bayesian methods to model the relationships between different advertising channels and consumer behaviour. The authors also use the Maxwell-Boltzmann equation and the Michaelis-Menten model to model the flow of consumers between different advertising channels.	The paper is based on a simulation study, so it is not clear how well the approach would work in real-world data. The paper only considers a limited number of advertising channels and consumer segments. It is possible that the approach would not work as well with more complex data.
2.	Hierarchical marketing mix models with sign constraints	2021	Journal of Applied Statistics	Hao Chen, Minguang Zhang, Lanshan Han & Alvin Lim	Hierarchical marketing mix models allow for the estimation of the effects of marketing activities at different levels, such as the product level, the	It is difficult to capture the carryover effect of marketing activities beyond a certain period. It is difficult to model the saturation effect of

					brand level, and the aggregate level, incorporating carryover and shape effects as well. Sign restrictions are used to ensure that the estimated parameters have the expected signs. Hamiltonian Monte Carlo is a MCMC algorithm that is used to sample from complex probability distributions.	advertisements, where additional advertising has no impact on sales. The scaling factor needs to be taken into account when maximizing the likelihood function, but this can be computationally expensive. The carryover effect is assumed to be constant for each advertisement, but it could vary across regions.

3.	Marketing Mix Modeling (MMM) – Concepts and Model Interpretation	2021	International Journal of Engineering Research and Technology	Sandeep Pandey, Snigdha Gupta, Shubham Chhajed	The paper adopts an econometric approach to Marketing Mix Modeling using statistical models to assess how various marketing tactics affect sales. This method aims to refine marketing budget allocations to enhance investment returns by deeply understanding strategic interactions within the marketing mix.	Econometric models presume stable variable relationships, which can miss the volatile nature of market dynamics influenced by unforeseen external factors. Moreover, their complexity may result in overlooking some market nuances, potentially skewing the strategic insights derived from the analysis. External factors.

4.	Media Mix Modelling – A Monte Carlo simulation study	2014	Journal of Marketing Analytics	Yong Liu, Jorge Laguna, Matt Wright, Hua He	The paper introduces a novel algorithm for media mix modelling using Monte Carlo simulation, focusing on simultaneous time and revenue response extraction from media spending data. This approach is	Key limitations include the reliance on simulation data for validation and the potential complexity in applying the algorithm to real-world data with inherent uncertainties. The methodology's success hinges on accurate parameter reconstruction, a

					generalizable across various budget allocation optimizations, showing potential for a significant increase in revenue through channel spend optimization.	challenging task given the model's sensitivity to data quality and the high dimensionality of the optimization problem.
5.	Dynamic budget allocation for social media advertising campaigns: optimization and learning	2021	European Journal of Operations Research	Yossi Luzon , Rotem Pinchover , Eugene Khmelnitsky	The paper introduces a novel methodology for optimizing social media advertising budgets, focusing on maximizing audience exposure through a dynamic optimization problem and an effectiveness function to assess bid impacts. It leverages the Maximum Principle for optimality conditions and proposes an algorithm for real-time strategy adjustments, emphasizing the integration of learning and optimization to	Limitations include the model's reliance on assumptions that may not apply universally, and the complexity of real-time implementation. The effectiveness varies across platforms due to different user behaviors and platform algorithms, with outcomes highly sensitive to the precision of parameter estimations within the effectiveness function, potentially leading to suboptimal strategies.

					adapt to market changes.	

6.	Allocation of advertising budget between multiple channels to support sales in multiple markets	2016	Journal of the Operational Research Society	Vahideh Sadat Abedi	The methodology centers on formulating the challenge as a nonlinear and non-separable knapsack problem. The authors employ a 'branch and cut' solution method augmented by heuristic strategies to manage the dynamic allocation of marketing budgets across channels with varying marketability and cost-effectiveness. This approach incorporates a threshold effect, requiring a minimum investment to generate noticeable sales impacts, which adds complexity to the modelling.	The study recognizes several limitations, including the computational intensity due to the problem's complex nature, which may restrict its use to scenarios with substantial computational resources. While designed to adapt to different market conditions, the reliance on heuristic adjustments might not fully capture real-world market dynamics, possibly resulting in less optimal marketing strategies.

7.	Bayesian Methods for Media Mix Modelling with Carryover and Shape effects	2017	Google Inc.	Yuxue Jin, Yueqing Wang, Yunting Sun, David Chan, Jim Koehler	The paper proposes a sophisticated Bayesian approach for media mix modelling that specifically addresses carryover and shape effects of advertising, which are typically challenging to model with traditional linear regression. The methodology employs (MCMC) algorithms for estimating the model parameters, leveraging a Bayesian framework to incorporate prior knowledge from previous models.	The model's performance is significantly influenced by the choice of priors, which can introduce bias, especially in cases of small data samples. The computational complexity due to the use of MCMC algorithms may also limit the applicability of the model in environments where computational resources are constrained.

8.	A Hierarchical Bayesian Approach to Improve Media Mix Models using Category Data	2017	Google Inc.	Yuxue Jin, Yueqing Wang, Yunting Sun, David Chan, Jim Koehler	The paper introduces a hierarchical Bayesian approach to enhance MMMs by pooling data across various brands within the same product category. This method utilizes category-level data to generate informative priors, which are then applied to brand-specific models. The approach aims to leverage the increased data variability from pooling to improve the robustness and accuracy of parameter estimates, enabling more effective predictions of media impacts on sales performance.	Limitations include a reliance on the availability and consistency of category-level data, which can vary significantly. The assumption that media effects are uniform across brands may not always be valid, potentially leading to inaccurate predictions.

2.2 Research Gaps

The integration of digital and traditional media within contemporary marketing strategies often lacks a cohesive and unified approach, presenting several challenges. One prominent issue lies in the oversight of synergistic effects between various marketing channels, leading to missed opportunities for amplifying campaign impact. Moreover, the granularity and quality of data present ongoing obstacles, impeding the generation of actionable insights crucial for informed decision-making. Another significant challenge emerges from the struggle of existing models to effectively incorporate the dynamic nature of consumer behaviour and focus user- centric media campaigns to maximise key marketing objectives.

2.3 Problem Statement

Businesses are undergoing a digital transformation. It is through this digital trans- formation that companies have developed digital methods to reach customers through advertising. Therefore advertisers seek to properly measure the effectiveness of the different marketing/media channels. Since the adoption of additional advertisement streams, it has become necessary to be able to measure the effectiveness of online and offline media and the contributions they have towards achieving business goals.

Marketing Mix Models (MMM) is a solution implemented to achieve the goal of measuring market effectiveness. There are challenges and limitations involved with regards to the development of a model that can be used. Essentially a well-defined MMM is determined by how well it can contribute towards the decision making process of a business. MMMs can guide strategic decisions on budget allocation, channel optimization, and marketing mix strategy by identifying the incremental impact of each marketing input on sales. An implementation of Bayesian Models along with understanding customer segment level impact of media channels can help generate more granular insights and drive business decisions effectively.

3. TECHNICAL SPECIFICATION

3.1 Requirements

3.1.1 Functional

Functional requirements describe what the system should do, encompassing all the features and functions the system must perform.

- Data Integration: Ability to integrate and ingest data from multiple sources, including sales figures, marketing spend, customer demographics, and macroeconomic indicators.
- Clustering: Perform K-means clustering to generate required customer segments that will be beneficial for targeted marketing.
- Bayesian Analysis: Perform Bayesian statistical analysis to estimate the effect of different marketing channels on sales or other KPIs, taking into account prior knowledge and evidence.
- Budget Optimization: Implement optimization algorithms that utilize the Bayesian analysis outputs to recommend how to allocate marketing budget across different channels for maximum efficiency and effectiveness.
- User Interface (UI): Provide an interactive UI that allows users to specify parameters for the Bayesian model and budget optimization, run analyses, and view recommendations.
- Scenario Planning: Enable users to create and evaluate different budgeting scenarios based on various assumptions and constraints.
- Insights Reporting: Generate dashboards that present the results of the Bayesian MMM and budget optimization, including ROI, channel effectiveness, and suggested budget allocations.

3.1.2 Non-Functional

Non-functional requirements detail how the system performs certain operations and under what constraints, ensuring the system's reliability, efficiency, and overall quality.

- Performance: Ensure the system can process large datasets and complex calculations swiftly to produce timely results.
- Scalability: Design the system to handle an increasing number of marketing channels, larger datasets, and more complex models as the organization grows.
- Reliability: Guarantee a high level of accuracy in the model's predictive capabilities and the optimization recommendations.
- Usability: Ensure that the system is intuitive and easy to use, with minimal training required for end users.
- Security: Protect sensitive marketing data and proprietary models with robust security measures, including data encryption and secure authentication protocols.
- Maintainability: Facilitate easy updates and maintenance of the system with clear documentation and a modular design.
- Integration: The system should seamlessly integrate with other business intelligence tools and data warehouses used by the organization.
- Compliance: Adhere to legal and regulatory standards regarding data privacy, such as GDPR or CCPA, in handling and storing user and marketing data.

3.2 Feasibility Study

3.2.1 Technical Feasibility

- Data Availability and Quality
  Feasible: If high-quality historical data on marketing spends, sales, and other relevant variables are available.

  Challenge: Insufficient or poor-quality data could lead to inaccurate models and recommendations.

- Computational Resources
  Feasible: Access to adequate computational power to perform complex Bayesian calculations and optimization routines.
  Challenge: Bayesian computation can be resource-intensive, requiring advanced IT infrastructure.
- Expertise
  Feasible: Availability of skilled personnel in Bayesian statistics, data science, and software development to build and maintain the system.
  Challenge: Scarcity of experts could lead to implementation delays or increased training costs.
- Software and Tools
  Feasible: Existing tools and libraries (e.g., LightweightMMM, Numpyro, PyMC3) can facilitate the development of Bayesian models.
  Challenge: Integration of these tools into a seamless system may require significant customization.

3.2.2 Economic Feasibility

- Cost-Benefit Analysis
  Feasible: The expected increase in ROI from optimized marketing spend could outweigh the costs of developing and running the system.
  Challenge: High initial development costs or lower-than-expected ROI improvements could make the project less economically viable.
- Development Costs
  Feasible: If the budget for software development, data infrastructure, and personnel is within the company's financial capability.
  Challenge: Cost overruns or extended development timelines could impact economic viability.
- Market Position Improvement
  Feasible: Enhanced marketing efficiency could lead to a better competitive position and increased market share.
  Challenge: If the market response is slower or less receptive than anticipated, projected gains may not materialize.

3.2.3 Social Feasibility

- User Adoption
  Feasible: If the system is user-friendly and integrates well into the existing workflow, leading to high adoption rates among marketers and analysts.
  Challenge: Resistance to change or lack of trust in automated recommendations could hinder adoption.
- Client Satisfaction
  Feasible: Clients could see improved performance from more effective marketing campaigns, leading to increased satisfaction.
  Challenge: Over-reliance on the system without human insight could lead to generic campaigns and reduced customer engagement.

3.3 System Specification

3.3.1 Hardware Specifications

- Laptop/Desktop: The system should be equipped with a laptop or desktop computer.
- 8GB RAM: The system must have a minimum of 8 gigabytes (GB) of Random Access Memory (RAM) to ensure smooth performance.
- Minimum Storage of 256 GB or More: A storage capacity of at least 256 gigabytes (GB) or higher is required for storing project files, datasets, and software installations.
- Processor: The system should have a processor meeting the following specifications:
  - For Intel-based systems: Intel Pentium or above.
  - For Apple based systems: M1 chip or above.

3.3.2 Software Specifications

- Compatible with an OS (Windows, Linux, Mac): The software must be compatible with major operating systems such as Windows, Linux, and Mac.
- Python Version 3.7 or Above: The system should have Python programming language version 3.7 or a higher version installed.

- Anaconda Navigator: Anaconda Navigator, a graphical user interface (GUI) included with the Anaconda distribution, is required for managing packages, environments, and channels.
- Jupyter Notebook: Jupyter Notebook, an open-source web application, is necessary for creating and sharing documents containing live code, equations, visualizations, and narrative text.

- Web Browser (Chrome, Firefox, Safari, Microsoft Edge): A modern web browser such as Chrome, Firefox, Safari, or Microsoft Edge is required for accessing online resources, documentation, and web-based tools.

3.3.3 Standards and Policies

- Clean Code: All code must adhere to clean coding practices, following industry-standard conventions and guidelines.
- Proper Documentation of All Modules Throughout the Project: Each module and component of the project should be thoroughly documented, including explanations of functionality, inputs, outputs, and usage instructions.
- Efficient Data Management: Effective data management practices must be implemented to ensure data integrity, security, and efficiency throughout the project lifecycle.

- Making Note of Any Project Specific Additions, Exceptions, or Enhancements as Necessary: Any project-specific additions, exceptions, or enhancements should be documented and communicated as necessary to ensure clarity and alignment among team members and stakeholders.

4. DESIGN APPROACH AND DETAILS

4.1 System Architecture

Our project is essentially divided into two phases – in phase 1, we gather customer data to perform customer segmentation using K-means clustering to get five different customer segments as outputs, based on the annual income and spending scores of the customers. The clusters are then used in the MMM dataset as dummy variables to run phase 2 of the model, which is essentially performing market mix modelling.

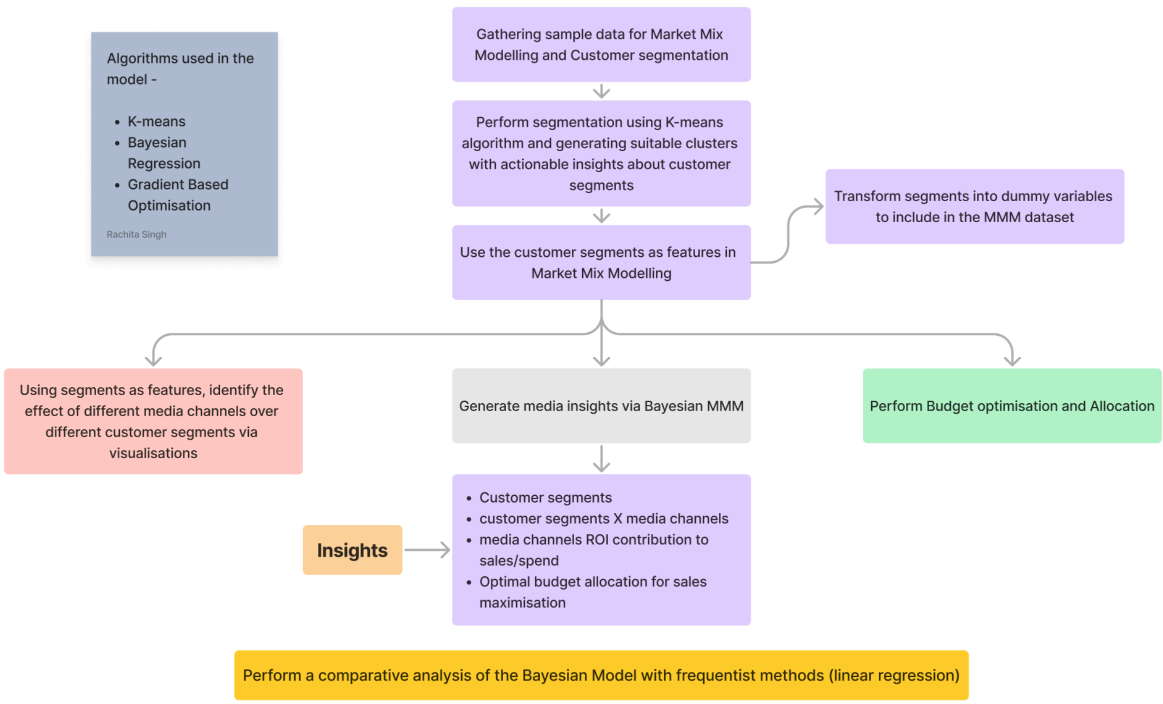


Fig 4.1 Project workflow

In phase 2, the dataset is a time series data with four media channels with their spends, sales as the target variable along with holiday, seasonality and clusters as dummy variables. We focus on 3 processes – understand the impact of different media channels on different customer segments via regression, estimation of media channel contribution and ROI of each media channel with respect to sales generated, and optimal allocation of budget across media channels to maximise sales. The project output will be implemented via a web application and a comparative analysis of the Bayesian model with frequentist models such as linear regression is also proposed.

A detailed system architecture of the Bayesian model is discussed below, with detailed explanation about each component.

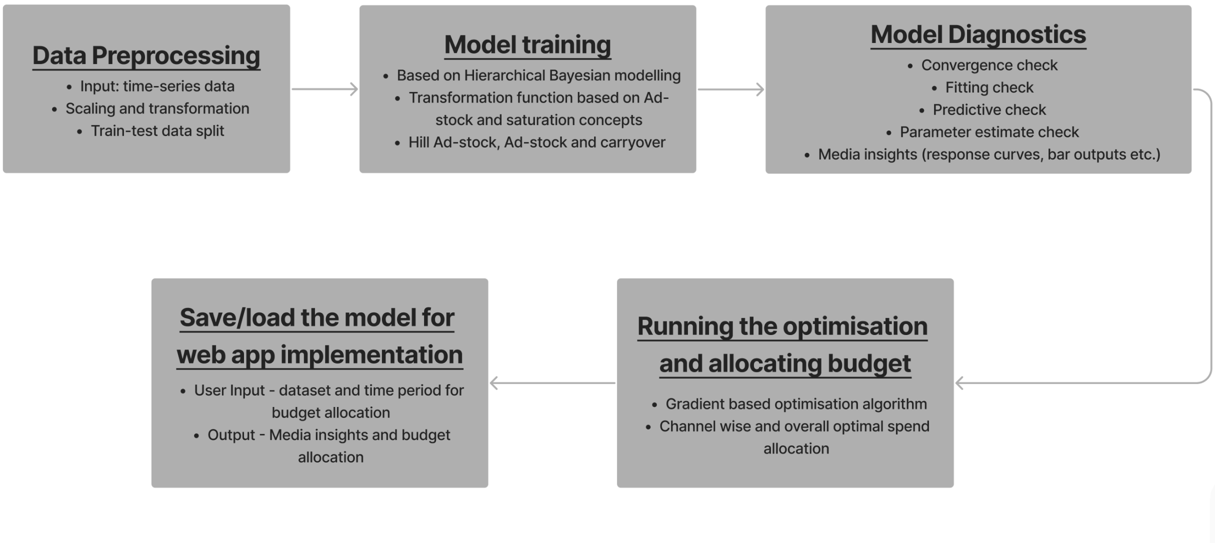


Fig 4.2 System Architecture

1. *Data Preprocessing*

The input data used is time-series data, which does not include any privacy related data. In terms of time, MMMs often require two to three years of weekly-level data. However, if that data is not available, daily data is also acceptable, but in that case, reviewing the outliers becomes crucial step. Talking about granularity when looking at media spending data, a common granularity is the media channel level, such as TV, Print, OOH, and digital.

2. *Model Training*

Hierarchical Bayesian modelling in the training phase allows the model to learn not only at the individual media channel level but also to understand and leverage patterns at higher levels of aggregation. Ad-stock transforms raw advertising spend into a metric that represents the 'memory' of past advertising, addressing the idea that the effects of advertising spend accumulate and decay over time. Saturation functions model the diminishing returns of advertising; as spend increases, it may become less effective at driving additional sales.

3. *Model Diagnostics*

Model diagnostics involve a series of checks and analyses to ensure the robustness and accuracy of the statistical model. These diagnostics assess whether the model has correctly converged, fits the historical data well, makes accurate predictions, and provides reliable parameter estimates. Essentially, they are a quality control measure, confirming that the model is valid and can be trusted for making informed decisions about marketing strategy and budget allocation.

4. *Running the optimisation and allocating budget*

After constructing a Bayesian model that offers a nuanced understanding of each media channel's contribution to ROI, the next critical step involves determining the optimal distribution of marketing spend to enhance overall campaign effectiveness. Sequential Least Squares Quadratic Programming, a gradient based optimisation algorithm plays a pivotal role in this process by helping marketers understand not just how much to spend in total, but precisely how to distribute that spend across various channels—such as newspaper, radio, TV, and social media—to ensure the highest possible return.

5. *Save/load the model for web app implementation*

For easy accessibility, we plan to implement a web application that allows users to input the dataset and specify a forecast time period. Based on these inputs, the model first displays the media impacts on customer segments and media insights and then an optimal spend allocation that aims to maximise sales is presented across each media channel. This is essentially based on the predictions that result after running the Bayesian Regression model, followed by gradient based optimisation.

4.2 Design

System design involves outlining the architecture, components, interfaces, and data for a system to meet specified requirements, focusing on how these elements work together to achieve functional goals of the project efficiently and reliably. The relevant diagrams are discussed further.

4.2.1 Data Flow Diagram

A Data Flow Diagram (DFD) is a graphical representation of the flow of data through an information system, modelling its process aspects. It illustrates how data enters the system, is processed by the system, and is outputted. DFDs are composed of four main components: data sources and destinations (external entities), data flows, data processes, and data stores.

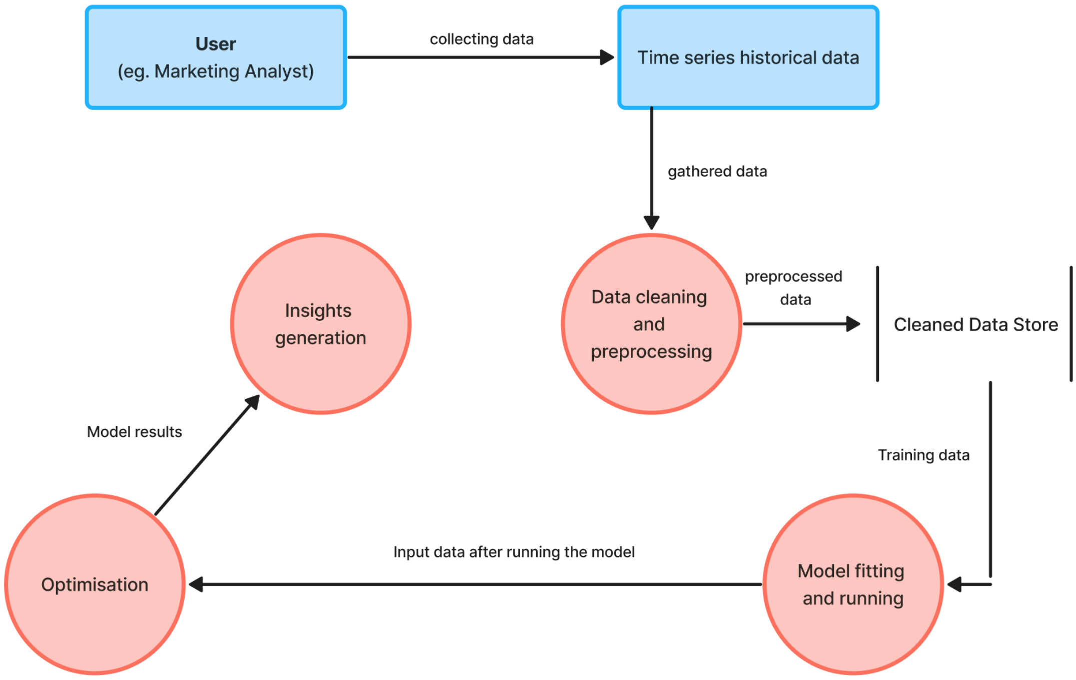


Fig 4.3 Data Flow Diagram

A DFD aids in understanding, communicating, and refining the MMM process by providing a clear and concise visualization of how data moves through the system. This visualization supports better planning, execution, and optimization of the marketing mix to achieve business objectives.

4.2.2 Use Case Diagram

A Use Case Diagram is a type of behavioural diagram defined by and created from a Use-Case analysis. Its main purpose is to show the relationships among actors (users or any other system) and use cases (system functionalities or processes) within a system. A use case diagram includes use cases, actors, and the interactions between these actors and the use cases, illustrating how different users interact with the system to achieve a goal.

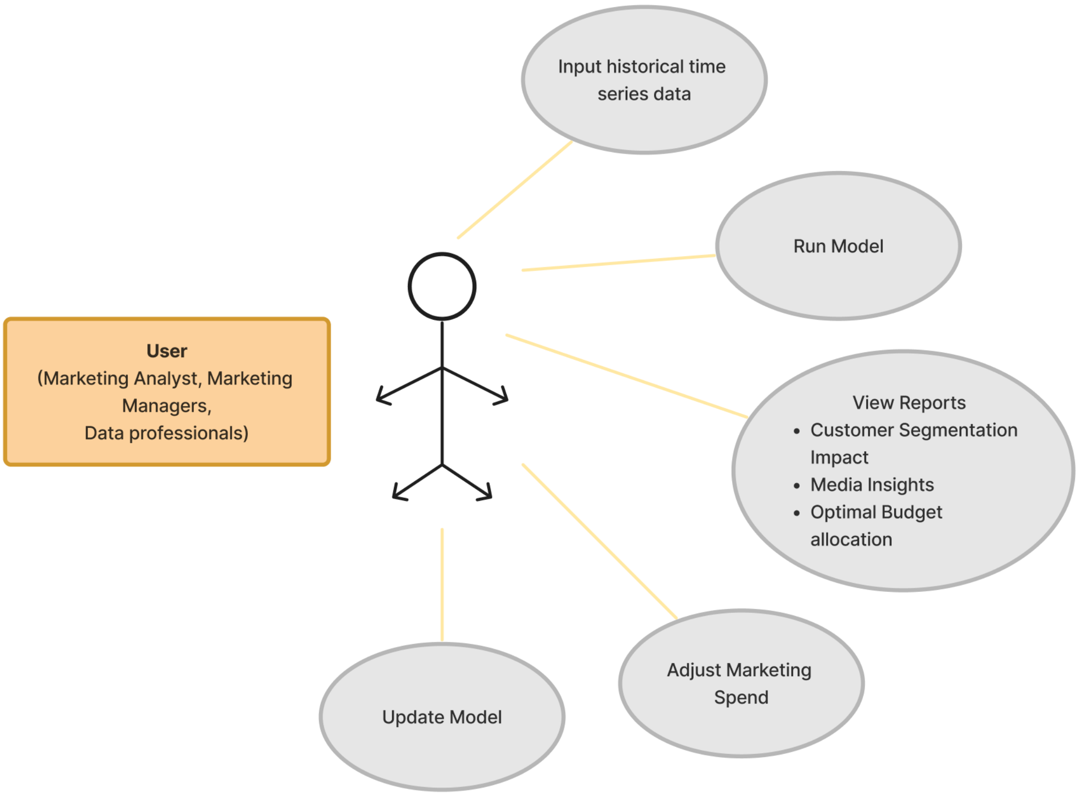


Fig 4.4 Use Case Diagram

Use case diagrams are instrumental in planning and designing market mix models, ensuring that the system is user-centric and meets the diverse needs of all stakeholders involved in marketing analysis and decision-making processes.

4.2.3 Class Diagram

A class diagram is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among the classes. It's a fundamental component of the Unified Modeling Language (UML), used in object-oriented software engineering.

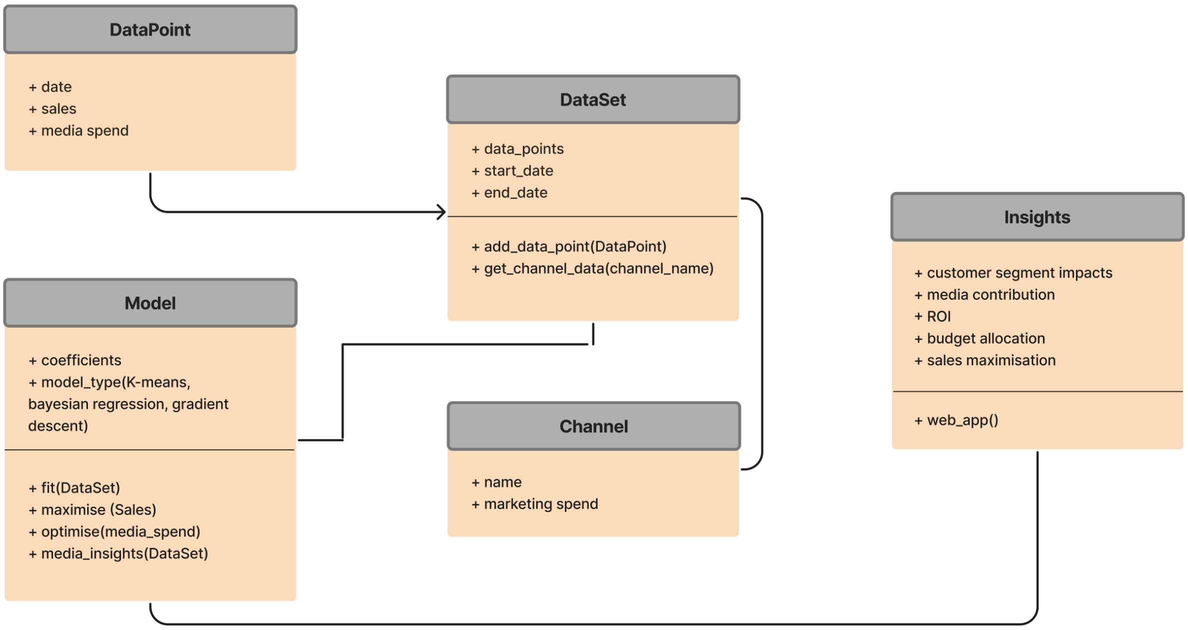


Fig 4.5 Class Diagram

Class diagrams are invaluable in the planning, design, and implementation phases of market mix models. They offer a clear structure for the system, promote better understanding and communication among team members, and ensure that the final system aligns with the intended goals and requirements of marketing analysis.

4.2.4 Sequence Diagram

A sequence diagram is a type of interaction diagram that shows how objects operate with one another and in what order. It is a construct of a Message Sequence Chart that illustrates the objects participating in a specific interaction and the sequence of messages exchanged between them to carry out the functionality of the scenario. Sequence diagrams are used in software development and systems engineering to detail the interactions between classes, systems, components, or actors in a process.

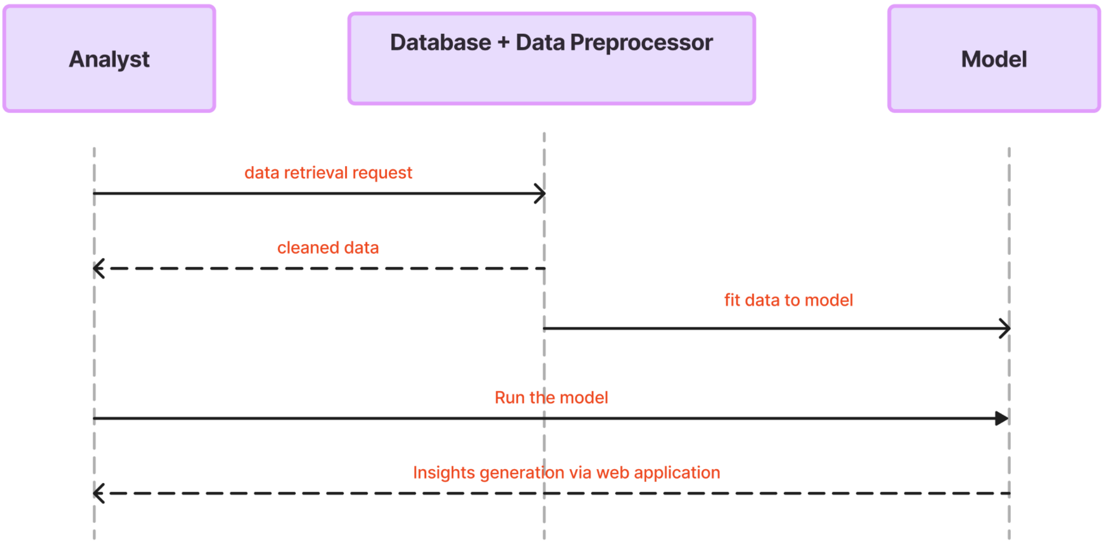


Fig 4.6 Sequence Diagram

Sequence diagrams are invaluable in designing, implementing, and optimizing market mix models. They provide a clear and detailed view of the dynamic behaviour of the system, facilitating better planning, communication, and understanding among all parties involved in the development and use of Marketing Mix Models.

4.3 Constraints, Alternatives and Trade-Offs

Knowing the restrictions of a project is equally important as knowing it's objectives and functionalities, especially one as complex as market mix modelling. And understanding the constraints, alternatives and trade-offs help us do exactly that.

One of the major constraints is data availability. Limited availability or quality of data can constrain the effectiveness of the analysis. Ensuring data collection methods are robust and comprehensive is a good way to handle it. We also deal with time and budget constraints. Time constraints may limit the depth of analysis or the complexity of the models implemented. Efficient project management is crucial to meet deadlines. The project budget, when implemented in the real world, may limit the resources available for data collection, software/tools acquisition, or hiring specialized expertise. Availability of skilled personnel in Bayesian statistics, machine learning, and marketing analytics could be a constraint for organisations from non tech background. They might have to train existing team members or outsource specific tasks.

Keeping our primary objective unchanged, the project offers a few alternatives as well. There are alternatives in selection of models, methodology and data sources.

Alternative modelling techniques apart from Bayesian regression, can be considered such as machine learning algorithms like random forests or gradient boosting, which may offer different insights. Instead of Bayesian methods, alternative statistical approaches like time series analysis or econometric modelling could also be the potential methodologies for market mix modelling. Organisations could explore alternative data sources beyond internal sales and marketing data, such as third-party data or industry reports, to enrich analysis and insights.

Alternative optimization algorithms like genetic algorithms or simulated annealing could also be used for budget allocation.

No matter how useful a project might be, it won't offer all the insights organisations might want. In improving the primary motives, there might be trade-offs with other chain of ideas. Some typical trade-offs of our project include:

- Model Complexity vs Interpretability: More complex models may provide better accuracy but could be harder to interpret. It can also be looked at as a trade-off between model complexity and ease of understanding for stakeholders.
- Accuracy vs Resource Intensity: Achieving higher accuracy may require more sophisticated models or extensive data processing, which could consume more resources.
- Short-term vs Long-term Focus: Optimizing for short-term sales may neglect long-term brand building. Balancing immediate sales impact with long-term brand equity need to be considered.
- Model Flexibility vs Rigidity: Flexible models may adapt better to changing market dynamics but could be more prone to overfitting, whereas rigid models may generalize better but struggle to capture nuanced relationships.

- Resource Allocation vs Risk Mitigation: Allocate resources between maximizing returns (e.g., investing heavily in high-ROI channels) and mitigating risks (e.g., diversifying investments across multiple channels to hedge against volatility).

5. SCHEDULE, TASKS AND MILESTONES

5.1 Gantt Chart

In this section, we present the project timeline in the form of a Gantt chart, outlining the sequential phases and tasks involved in our research project. The Gantt chart serves as a visual representation of the project's scope, timeline, and key milestones, providing a roadmap for the execution of tasks and deliverables. The Gantt chart serves as a vital tool for project management, enabling stakeholders to gain insights into the project's progression, identify potential bottlenecks or delays, and make informed decisions to mitigate risks and optimize resource utilization. Creating a Gantt chart for the project involves outlining the various tasks and their dependencies over time. The Gantt chart for our project is given below:

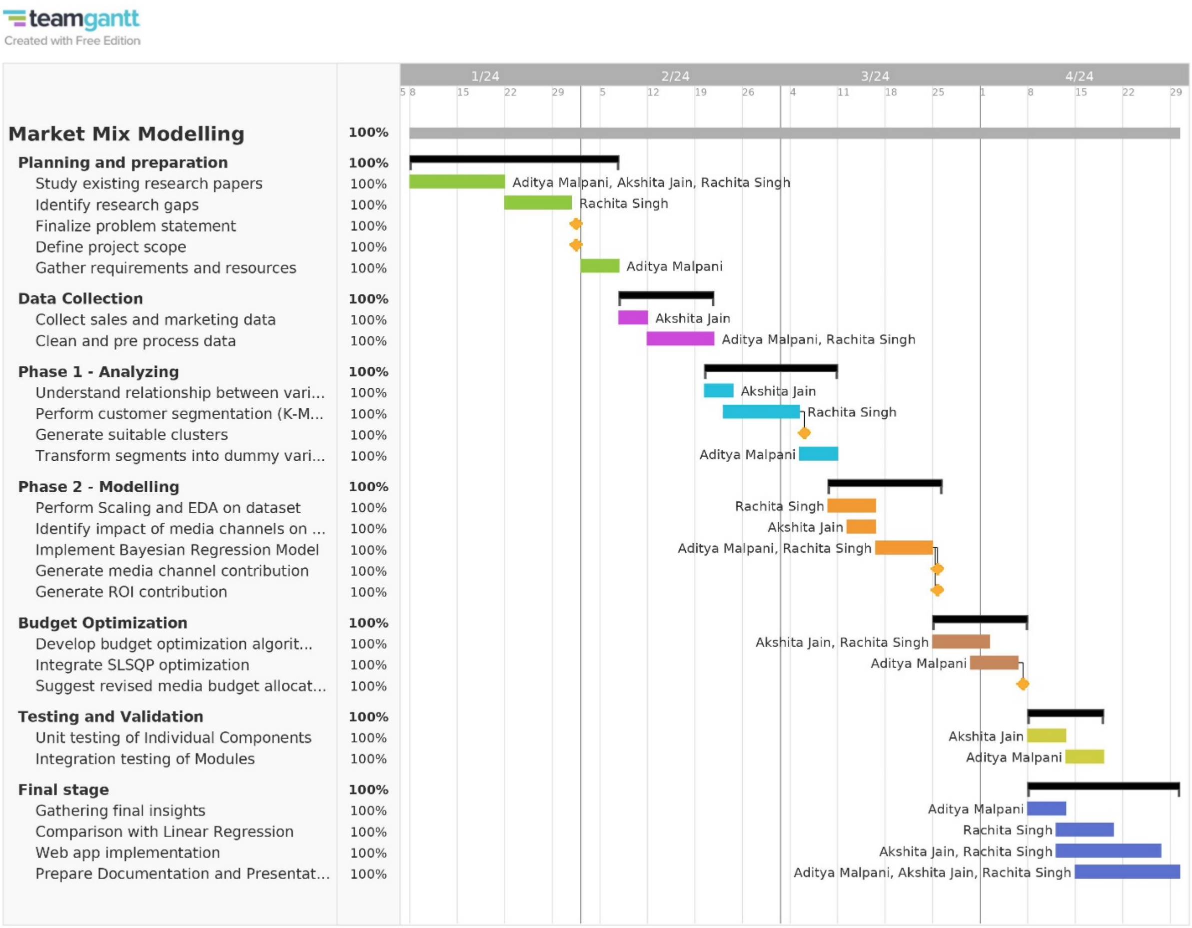


Fig 5.1 Gantt Chart

5.2 Modules

The project essentially comprises of 4 modules, with one module under phase 1 of the project and the remaining 3 under phase 2. Results of each module are integrated in the next module, and the functioning of each module is crucial for the overall implementation of the project.

5.2.1 Module 1: K-means Clustering

To perform customer segmentation, K-means clustering algorithm is used. A sample of the customer dataset shown in Fig 4.7 shows the following features – Customer ID, Customer Age, Customer Gender, Annual Income of the customer (in thousand dollars) and Spending score of the customer (based on customer behaviour and spending nature). The dataset contains around 200 datapoints.

Table 5.1 Customer Dataset

```
In [2]: df= pd.read_csv('/Users/rachitasingh/Desktop/Capstone/Customer_Data.csv', index_col = 0)
        df.head()
Out[2]:
```

CustomerID	Gender	Age	Annual Income (k$)	Spending Score (1-100)
1	Male	19	15	39
2	Male	21	15	81
3	Female	20	16	6
4	Female	23	16	77
5	Female	31	17	40

Data pre-processing and exploratory data analysis is performed on the dataset to identify and understand the relationships between different variables. Annual Income and Spending Score are the two features that form the basis for clustering. Feature normalisation helps to adjust all the data elements to a common scale in order to improve the performance of the clustering algorithm. The MinMaxScaler normalization technique was used to normalize the features before running the k-Means algorithm on the dataset.

To determine the number of clusters, WCSS method is used to generate an elbow graph. The WCSS is the sum of the variance between the observations in each cluster. It measures the distance between each observation and the centroid and calculates the squared difference between the two.

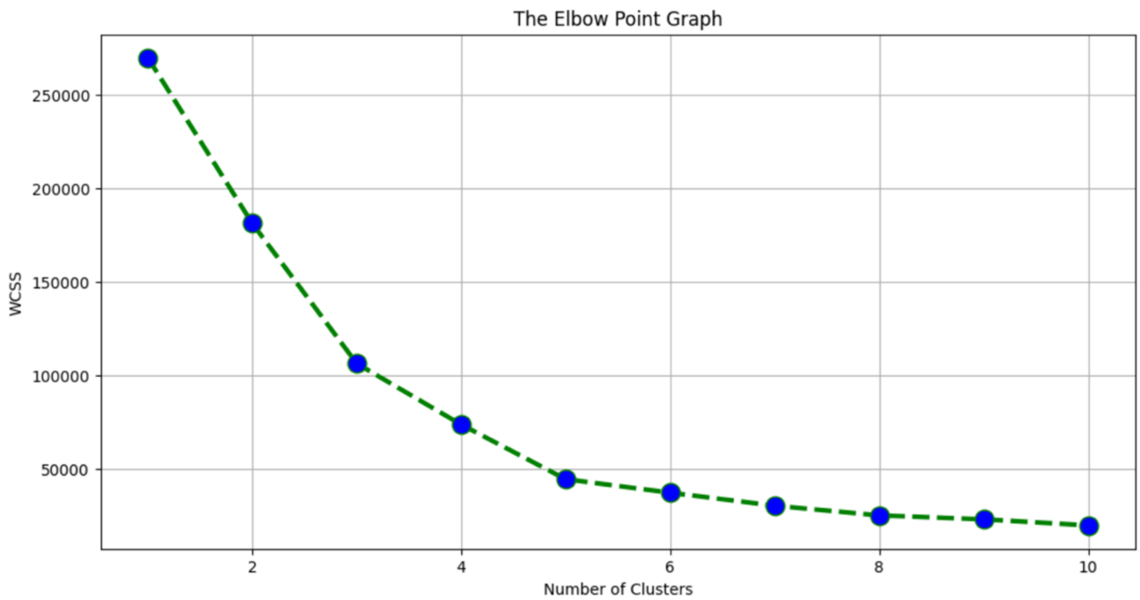


Fig 5.2 K-means Elbow graph

From the above graph we can observe that between number of cluster = 4 to number of cluster = 6 there has been substantial decrease(an elbow) hence, we chose the K value for our dataset as 5. After determining the optimal number of clusters, K Means algorithms is run to obtain the five unique clusters or customer segments.

5.2.2 Module 2: Impact of media channels on customer segments

The identified customer segments are incorporated in the MMM dataset as dummy variables. The MMM dataset consists of the following features – Media channels (Newspaper, TV, Radio, Social Media), Sales, Holidays, Seasonality and Clusters. A sample of the MMM dataset containing around 200 datapoints is shown in Fig 4.9.

Table 5.2 MMM dataset

```
In [2]: df_main = pd.read_csv("/Users/rachitasingh/Desktop/Capstone/MMM_Data.csv")
        df_main.head()
```

Out[2]:

	Week Start Date	TV	Radio	Newspaper	Social Media	Sales	hldy_Black Friday	hldy_Christmas Day	hldy_Christmas Eve	hldy_Columbus Day	...	seas_week_44	seas_week_45	seas_v
0	03/08/14	230.1	37.8	69.2	129.76	22100.0	0.0	0.0	0.0	0.0	...	0.0	0.0	
1	10/08/14	44.5	39.3	45.1	70.36	10400.0	0.0	0.0	0.0	0.0	...	0.0	0.0	
2	17/08/14	17.2	45.9	69.3	107.72	9300.0	0.0	0.0	0.0	0.0	...	0.0	0.0	
3	24/08/14	151.5	41.3	58.5	141.71	18500.0	0.0	0.0	0.0	0.0	...	0.0	0.0	
4	31/08/14	180.8	10.8	58.4	160.69	12900.0	0.0	0.0	0.0	0.0	...	0.0	0.0	

5 rows × 52 columns

Table 5.3 Clusters in MMM dataset

Cluster_0.0	Cluster_1.0	Cluster_2.0	Cluster_3.0	Cluster_4.0
0	0	0	0	1
1	0	0	0	0
0	0	0	0	1
1	0	0	0	0
0	0	0	0	1

To understand the impact of different media channels on different customer segments, we ran a regression model by initialising interaction terms (media x cluster) and assessed the impact on the basis of coefficient values.

5.2.3 Module 3: Media channel contribution and ROI

The MMM model is based on Bayesian Regression, which incorporates prior knowledge or beliefs into the modelling process, allowing for more robust and flexible estimation of marketing channel effects. It adapts to data uncertainty and complexity by updating beliefs with incoming data, offering probabilistic insights into the effectiveness of marketing spends across different channels. This approach provides a deeper understanding of marketing dynamics and decision-making under uncertainty, enhancing strategic planning and optimization in MMM.

A python framework called LightweightMMM has been used to perform Bayesian Modelling, and allows users to choose from 3 different approaches to demonstrate a lagged effect of media channels on sales – Ad stock, Carryover and Hill Ad stock.

Ad stock in marketing describes the phenomenon where advertising continues to influence consumer behaviour long after a campaign has concluded, extending beyond immediate or short-term effects. Ad stock is a component of the broader carryover effect, which refers to the delay between consumers seeing an ad and responding to it. A gradual decrease in the ad's impact represents one form of this effect. A delayed response is another form.

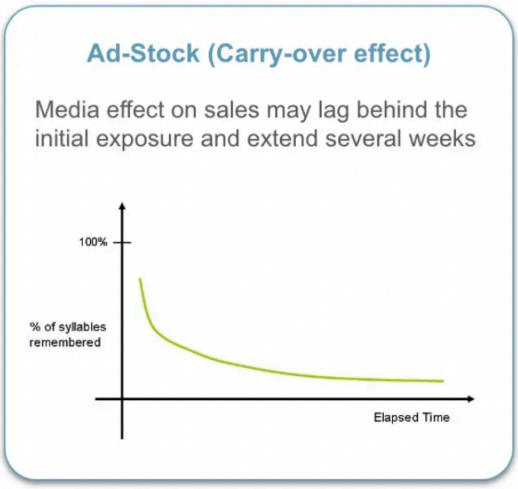


Fig 5.3 Ad stock and Carryover

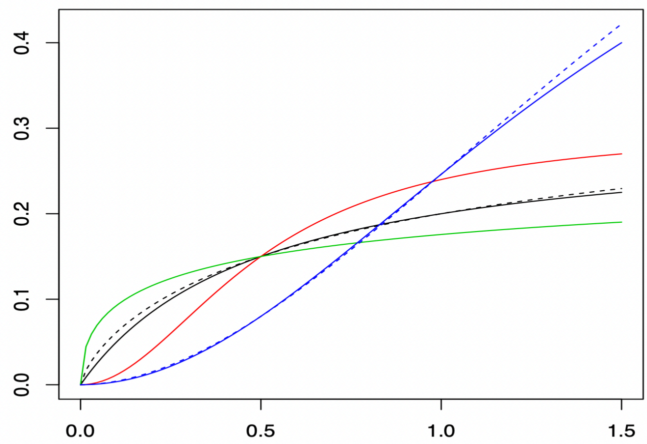


Fig 5.4 Hill Function

Hill Ad stock is a combination of Ad stock and Saturation effect (Each additional investment in advertising increases the response, but at a declining rate. In other words, the more money one spends on one media channel advertisement, the less effective it is). In the model, the Ad stock approach applies an infinite lag that decreases its weight as time passes, the Carryover approach applies a casual convolution giving more weight to the near values than the distant ones, and Hill Ad stock applies a sigmoid like function for diminishing returns to the output of the ad stock function. The choice of approach depends on the specific business use case and requirement. For our model, we have applied the 'Hill Ad stock' approach.

The model fitting is done via Markov Chain Monte Carlo simulation in Bayesian regression.

Markov Chain Monte Carlo (MCMC) is a class of algorithms which is used for sampling from probability distributions, the more steps taken equal more samples from the probability distribution, which results in a higher accuracy. The Markov chain of samples' third equilibrium should be equal to the target distribution that is required. In other words, MCMC tries to construct a set of samples that can search over the range of possible outcomes. The density of the distribution should be proportional to the total amount of time spent in each interval.

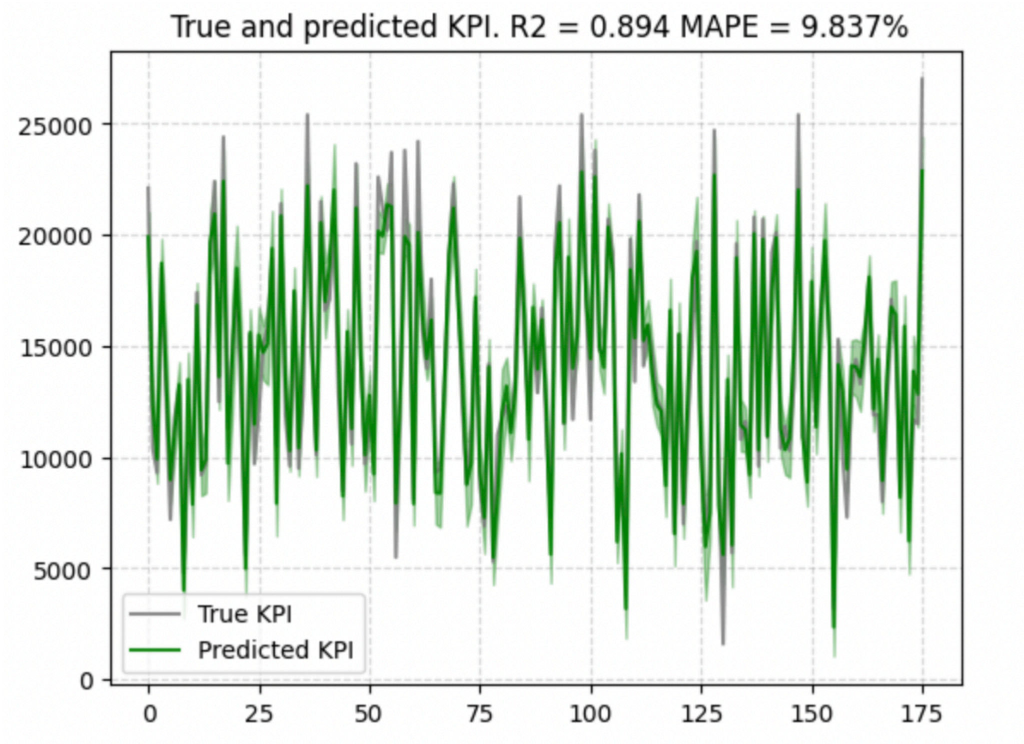


Fig 5.5 Bayesian Model fit results

5.2.4 Module 4: Budget optimisation and Allocation

The next step after understanding the current trends and data through media contribution estimates, ROI estimates and response curves is to perform optimisation. Optimisation is done on the basis of the response curves that are generated for each channel. The response curves are based on the saturation effect in advertising – each additional investment in advertising increases the response, but at a declining rate. In other words, the more money you spend on one media channel advertisement, the less effective it is. This phenomenon is based on the economic principle of diminishing marginal returns, which suggests that as investment in an area continues to rise, the rate of profit from that investment, after a certain point (called the saturation point), starts to decrease.

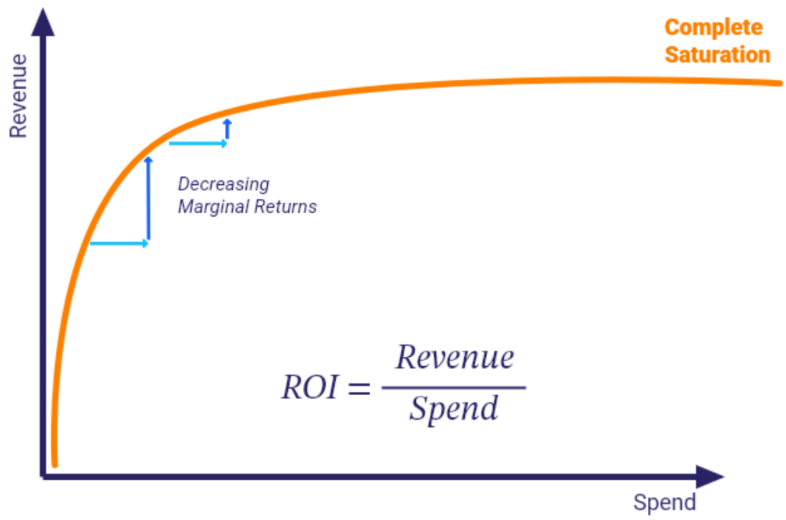


Fig 5.6 Saturation Effect

Optimizing budget using saturation curves involves identifying the optimal spend for each media channel that will result in the highest overall response while keeping the total budget fixed for a selected time period. To initiate optimization, the average spend for a specific time period is generally used as a baseline. The optimizer then uses the budget per channel, which can fluctuate within predetermined minimum and maximum limits (boundaries), for constrained optimization.

Optimisation is about transferring the budget to the media that generates highest incremental/marginal revenue. To perform optimisation in MMM, we have used the Sequential Least Squares Quadratic Programming (SLSQP) algorithm. It is an iterative method for constrained non-linear optimisation problems and is based on gradients. Key features of SLSQP include:

- Sequential – The algorithm solves a sequence of optimisation problems, each of which approximates the full non-linear problem but is easier to solve.
- Least Squares – SLSQP solves least squares problems to fit models to data or minimise the sum of squares of non-linear functions.

- Quadratic Programming – It uses a quadratic approximation of the objective function and a linear approximation of the constants. The QP is then solved at each iteration.
- Handling constraints – SLSQP is well suited for problems with both equality and inequality constraints.

In the context of MMM, the SLSQP algorithm is applied to optimise the allocation of a marketing budget across different channels to maximise the total sales. The objective function here is the sales equation that related the spend across various media channels to sales. Constraints are the budget constraints, spend constraints (min-max spend per channel) and the number of time periods to run the optimisation for. The decision variables in this case are the amounts to be allocated to each marketing channel.

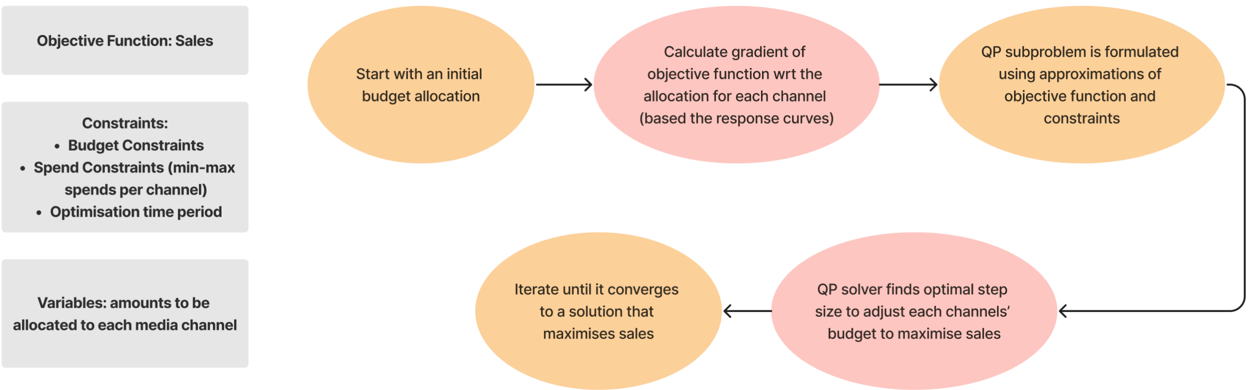


Fig 5.7 Steps for optimisation

The steps are outlined as follows:

- The algorithm starts with an initial budget allocation.
- The gradient of the objective function with respect to the allocation for each channel is calculated. The gradient helps point towards the direction where the objective function can be improved.
- A QP subproblem is formulated using the approximations of the objective function and constraints.
- The QP solver finds the optimal step size to adjust each channels' budget to increase sales. A series of decisions are made to adjust the allocations iteratively based on the step size. After each adjustment, we see if marketing is more profitable and then decide the next move.

- The algorithm iteratively adjusts the budget allocation, respecting the constraints, until it converges to a solution that maximises the objective function.

```
Optimization terminated successfully    (Exit mode 0)
            Current function value: -334577.88393752556
            Iterations: 6
            Function evaluations: 54
            Gradient evaluations: 6
```

Fig 5.8 Gradient Optimisation Results

The objective function iterates over all media channels and calculates the total response based on the sum of individual response levels per media channel. To maximize the response in the optimization function, we need to convert it into a minimization problem. Therefore, we obtain the negative value of the total response, which we then use as the objective for the optimization function. In simple terms, applying this algorithm to MMM is allocating the budget in the best possible way to maximise sales, considering all the budget constraints, with the help of a gradient and quadratic programming.

5.3 Testing

Performing testing for a project like this involves verifying the functionality and integration of individual components and modules, ensuring they work correctly both independently and when combined. Here's how organizations can approach unit testing and integration testing for different aspects of the project:

5.3.1 Unit Testing

Data Collection and Preparation:

- Test Data Retrieval: Verify that data retrieval mechanisms from various sources (e.g., databases, APIs) are functioning correctly and return the expected data.
- Data Cleaning: Test data cleaning functions to ensure missing values are handled appropriately, outliers are identified and addressed, and data is formatted correctly.

Segmentation Analysis:

- Cluster Analysis: Test clustering algorithms to ensure they correctly group customers based on similarities in behaviour or preferences.
- Segmentation Validation: Verify that segments identified align with business intuition and exhibit distinct characteristics.

Bayesian Regression Analysis:

- Model Construction: Test the construction of Bayesian regression models to ensure they are set up correctly with appropriate priors, likelihood functions, and inference methods.
- Parameter Estimation: Validate parameter estimation procedures to ensure they accurately estimate model parameters from data.

Budget Optimization:

- SLSQP Optimization Algorithm: Test the implementation of the SLSQP optimization algorithm to verify its effectiveness in allocating the marketing budget across various channels while maximizing ROI. Verify that the optimization process considers the estimated ROI of each channel and adjusts budget allocations accordingly to achieve the desired sales outcomes.
- Budget Allocation: Validate budget allocation strategies to ensure they distribute funds across channels effectively while maximizing ROI.

Comparative Analysis:

- Model Comparison: Test the implementation of frequentist models (e.g., Linear regression) to ensure they produce results consistent with expectations and are comparable to Bayesian models.
- Statistical Significance: Validate statistical tests used for model comparison to ensure they provide reliable assessments of differences between models.

5.3.2 Integration Testing

End-to-End Workflow:

- Data Flow: Test the flow of data through the entire workflow, from data collection to budget optimization, ensuring that data transformations and analyses are carried out correctly at each step.
- Module Integration: Verify that individual modules integrate seamlessly with each other and exchange data appropriately.

Scenario Testing:

- Use Case Scenarios: Test the system using realistic scenarios representative of typical business use cases, ensuring that the end-to-end workflow functions as expected in practical situations.
- Edge Cases: Test the system with edge cases, such as extreme values or unexpected data formats, to verify robustness and error handling.

Input Validation:

- Data Validation: Validate that the system properly handles invalid or unexpected input data, providing appropriate feedback or error messages.
- Parameter Validation: Ensure that parameters passed between modules are validated to prevent errors due to incorrect inputs.

Performance Testing:

- Scalability: Test the system's performance with varying data volumes to ensure scalability, identifying potential bottlenecks or performance issues.
- Response Time: Measure the response time of critical operations to ensure they meet performance requirements.

Integration with External Systems:

- External Data Sources: Test integration with external data sources (e.g., APIs, third-party databases) to ensure seamless data exchange and compatibility.
- Reporting and Visualization Tools: Verify integration with reporting and visualization tools used for presenting results to stakeholders, ensuring data is accurately represented.

6. PROJECT DEMONSTRATION

We have implemented a web application, that allows users to navigate through tabs/pages, with each tab highlighting a working module of the project. There are six pages in total, with the first page as an "**Introduction**" to the project.

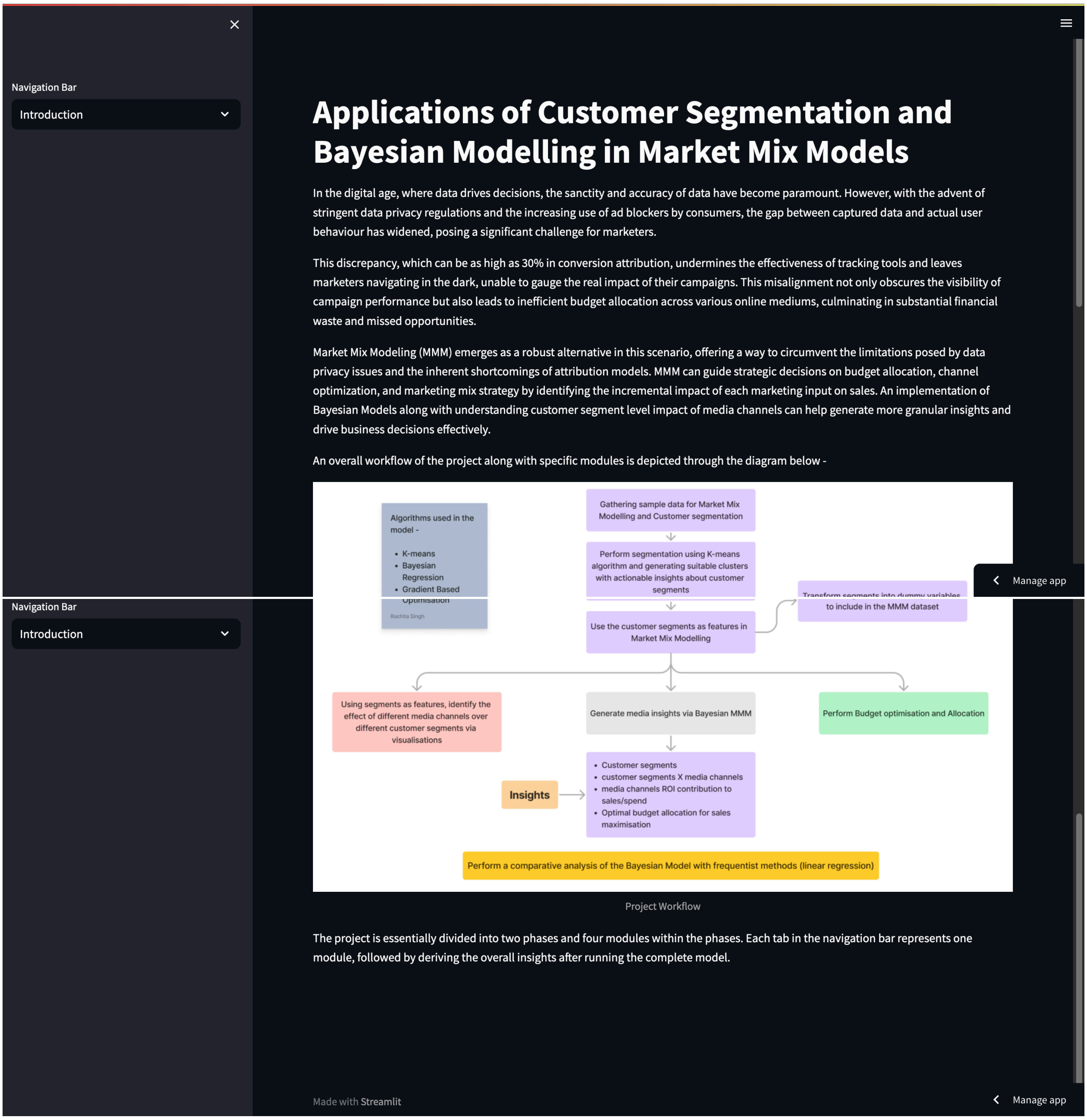
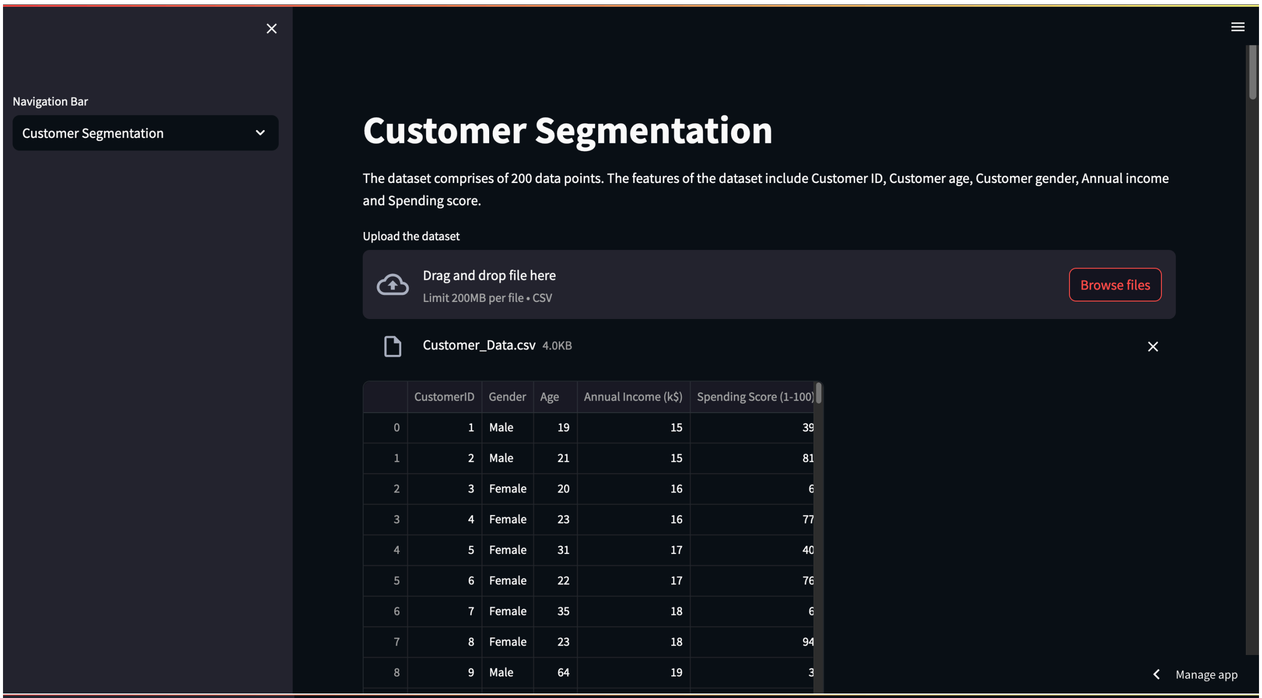
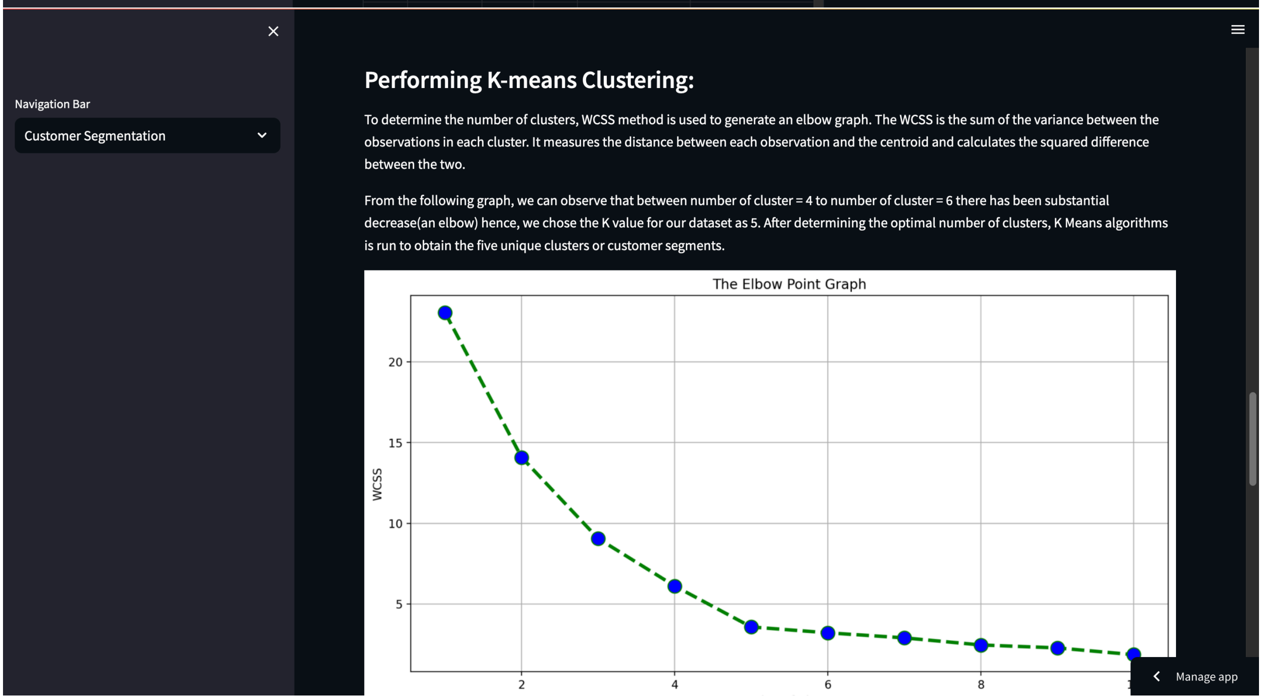


Fig 6.1 Introduction page

The next page shows “**Customer Segmentation**”, where we have uploaded and analysed the customer dataset and performed K-means clustering to obtain five customer segments based on the Annual Income and Spending Score of the customers.



	CustomerID	Gender	Age	Annual Income (k$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40
5	6	Female	22	17	76
6	7	Female	35	18	6
7	8	Female	23	18	94
8	9	Male	64	19	3



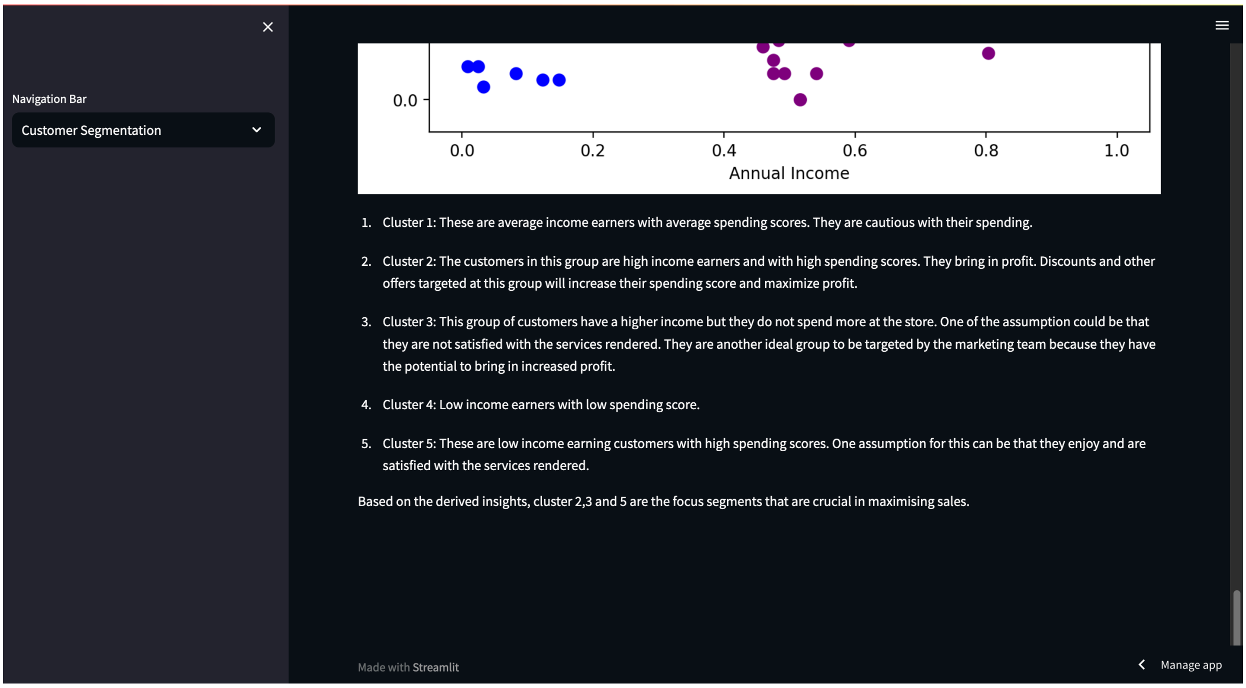
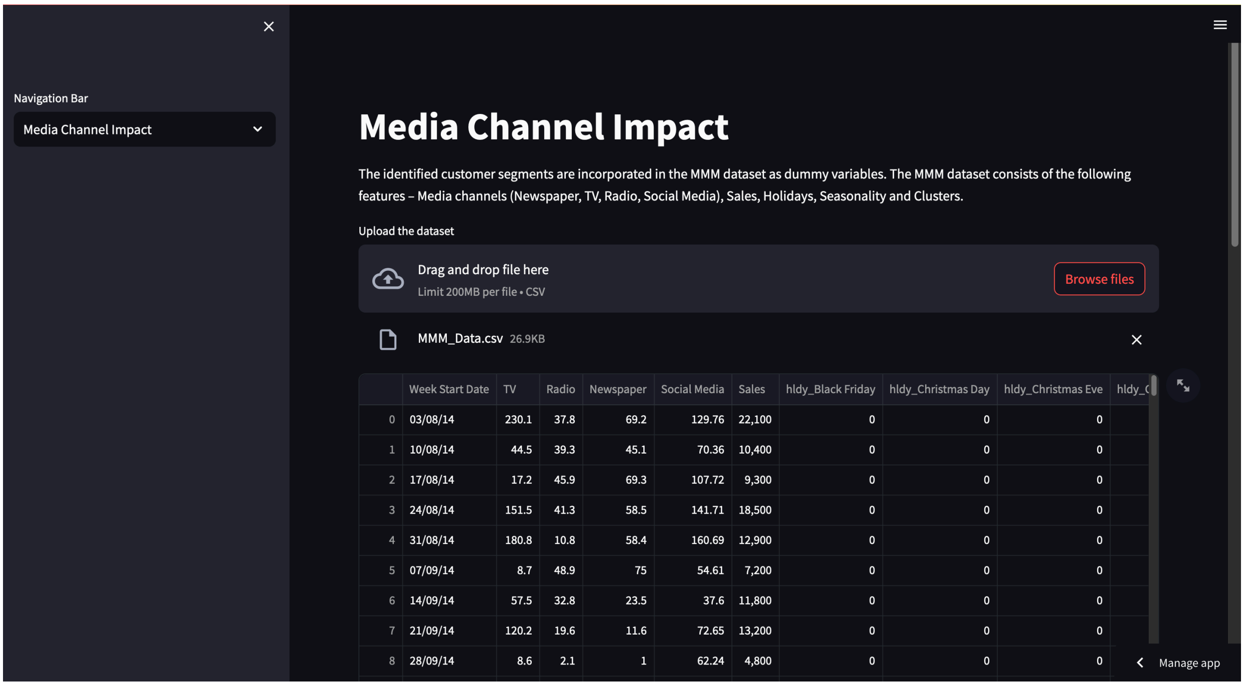


Fig 6.2 Customer Segmentation Page

After identifying the target customer segments, we include them in the MMM dataset in the form of dummy variables and use it to analyse the impact of different media channels (Newspaper, TV, Radio and Social Media) on different customer segments on the "**Media Channel Impact**" page.



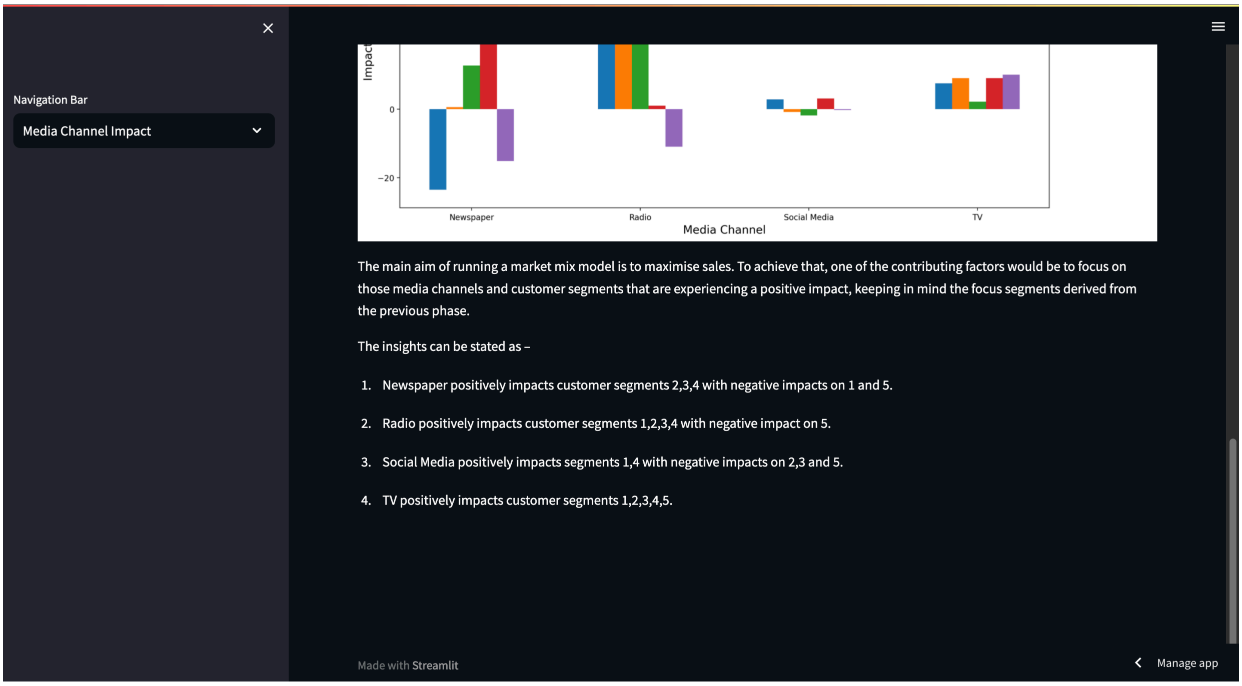
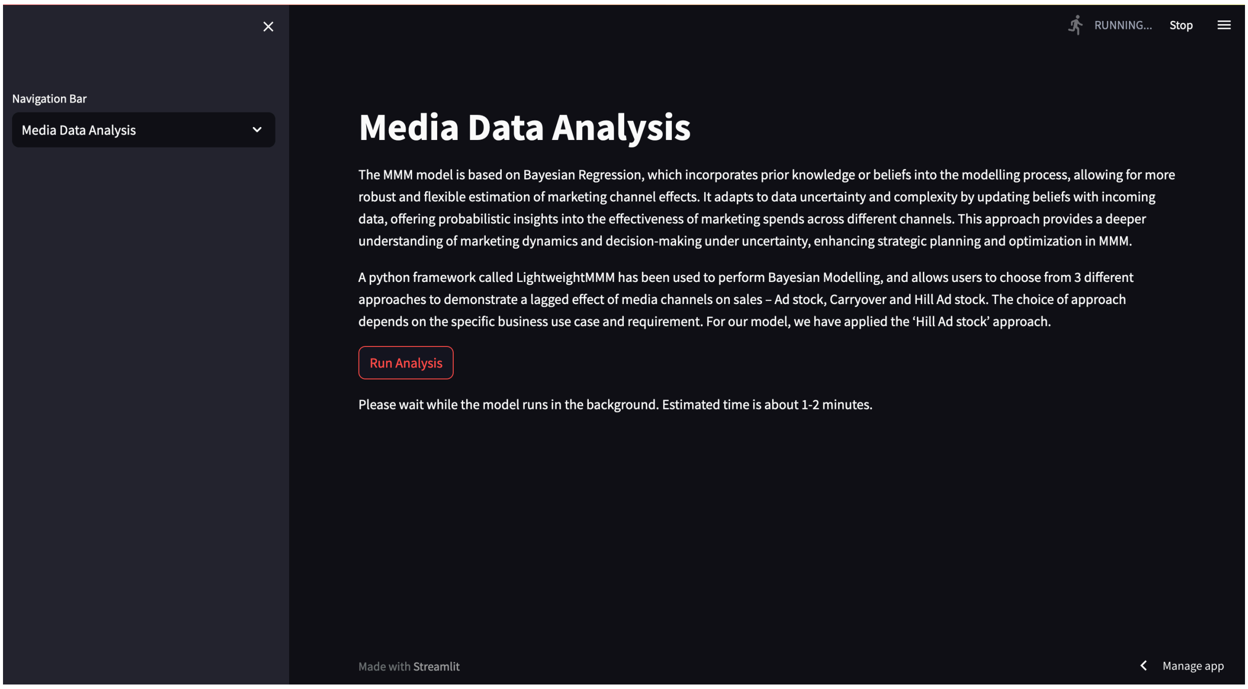


Fig 6.3 Media Channel Impact page

After assessing impact, we now need to understand the prior data and perform an optimal budget allocation based on that. This analysis can be run on the "**Media Data Analysis**" page.



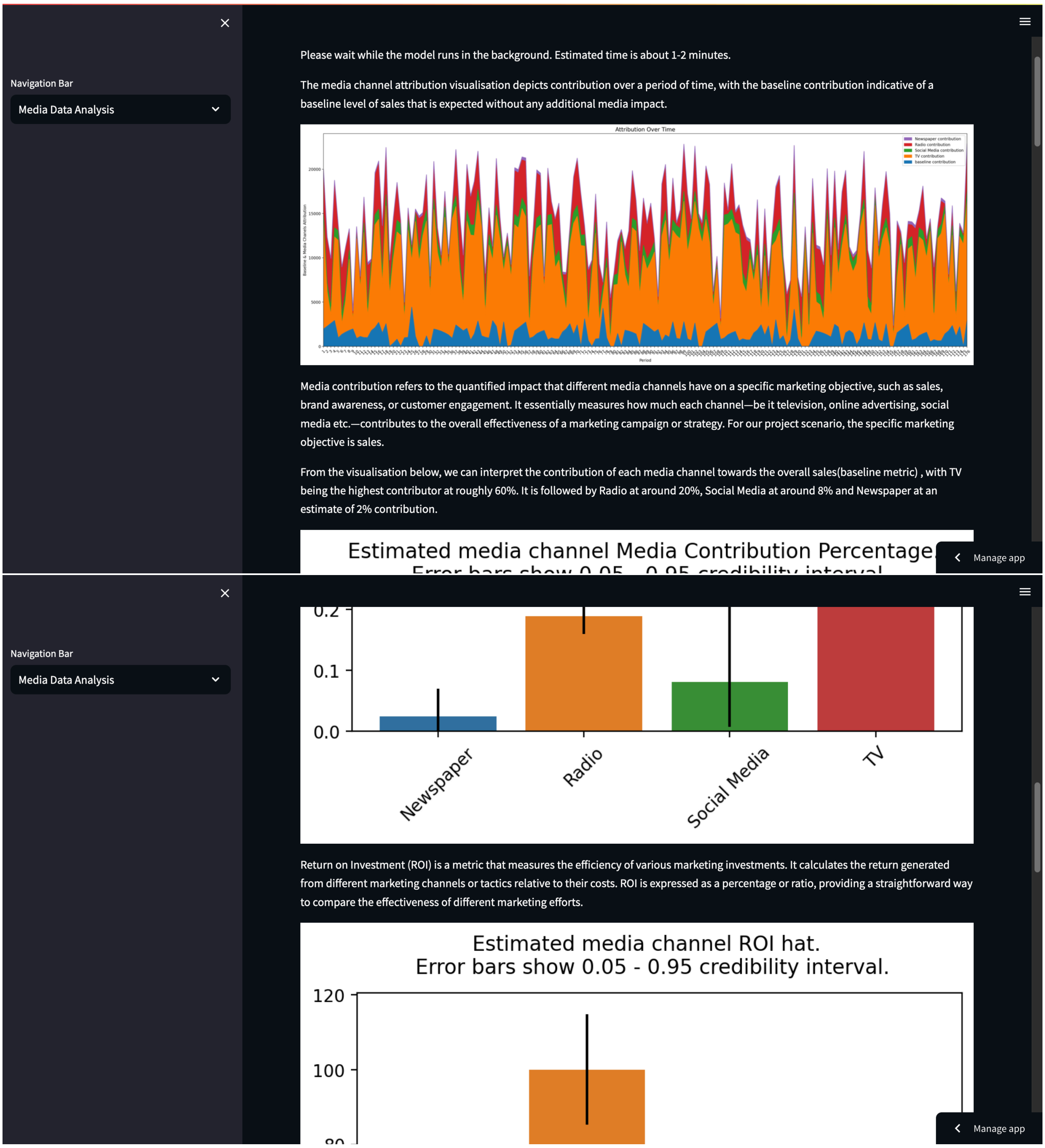


Fig 6.4 Media Data Analysis page

The last module of the project is to perform the optimisation and budget allocation according to a specified number of time periods. The time period is taken as input from the user (20 weeks in this case) on the "**Budget Optimisation and Allocation**" page.

Fig 6.5 Budget Optimisation and Allocation page

Finally, we derive overall insights from the marketing mix model, that provides a suggestive allocation of budget for each of the media channels with a focus on targeted marketing towards specific customer segments on the "**MMM Insights**" page.

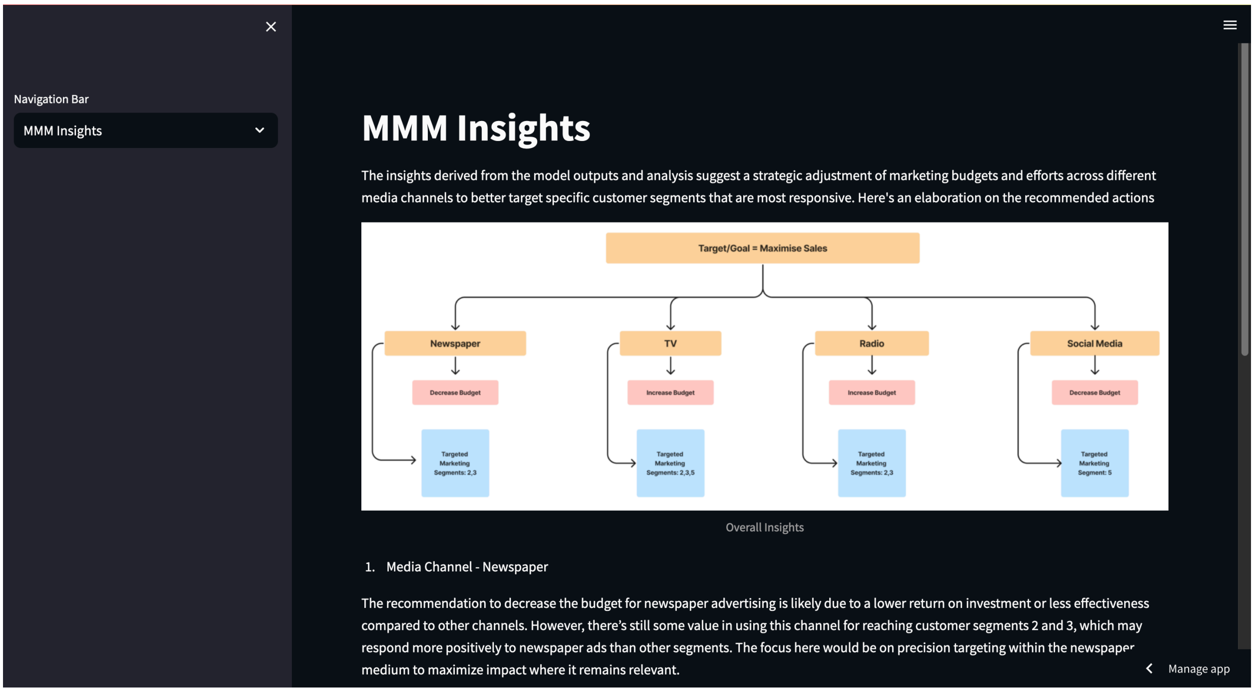


Fig 6.6 MMM Insights page

7. RESULTS AND DISCUSSION

After performing clustering in phase 1, we arrived at the following results -

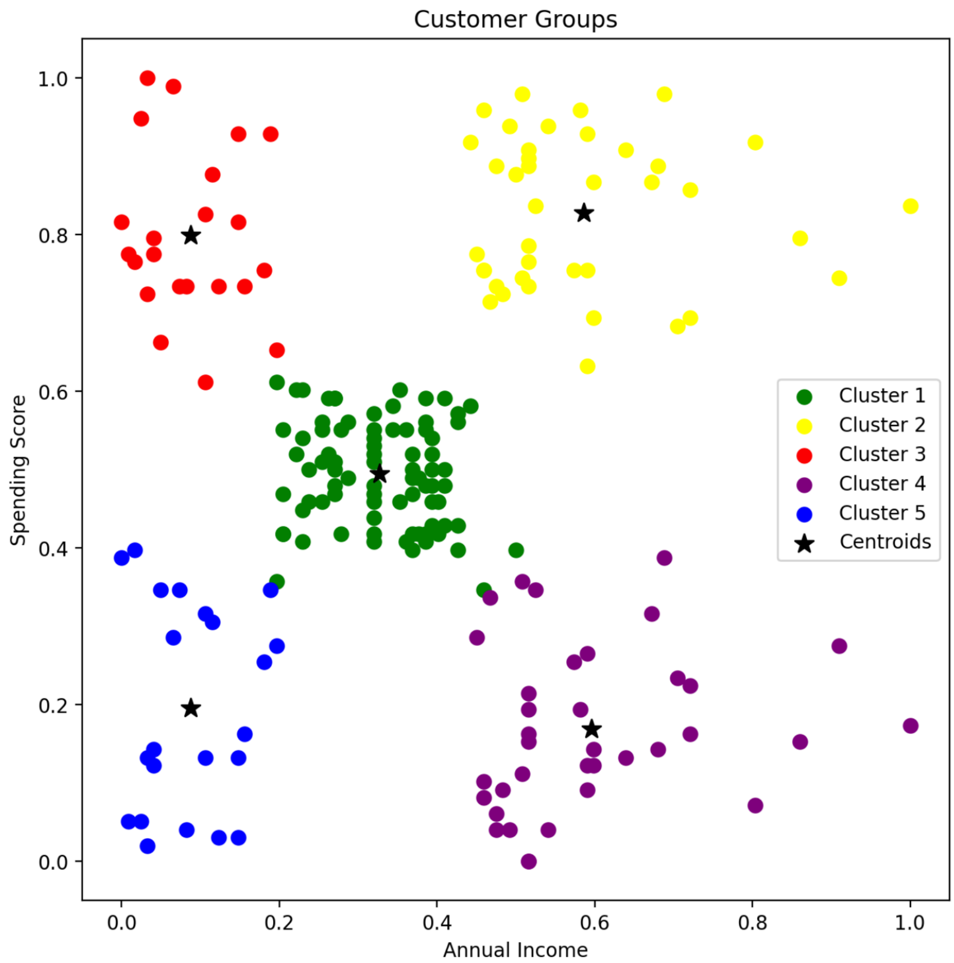


Fig 7.1 Customer Segmentation

- Cluster 1 (green): These are average income earners with average spending scores. They are cautious with their spending.
- Cluster 2 (yellow): The customers in this group are high income earners and with high spending scores. They bring in profit. Discounts and other offers targeted at this group will increase their spending score and maximize profit.
- Cluster 3 (red): This group of customers have a higher income but they do not spend more at the store. One of the assumption could be that they are not satisfied with the

services rendered. They are another ideal group to be targeted by the marketing team because they have the potential to bring in increased profit.

- Cluster 4 (purple): Low income earners with low spending score.
- Cluster 5 (blue): These are low income earning customers with high spending scores. One assumption for this can be that they enjoy and are satisfied with the services rendered.

Based on the derived insights, cluster 2,3 and 5 are the focus segments that are crucial in maximising sales.

In phase 2, our first goal is the understand the impact of different media channels on different customer segments. There are four media channels, namely – newspaper, TV, radio and social media, along with five distinct customer segments. The visualisation depicts either a positive or negative impact of a media channel on different customer segments, and the impacts are based on the coefficient values of the interaction terms (media x cluster). A positive impact estimates that running ads through that media channel could convert that specific group of customers into users of their product/service. A negative impact estimates the opposite.

The main aim of running a market mix model is to maximise sales. To achieve that, one of the contributing factors would be to focus on those media channels and customer segments that are experiencing a positive impact, keeping in mind the focus segments derived from the previous phase.

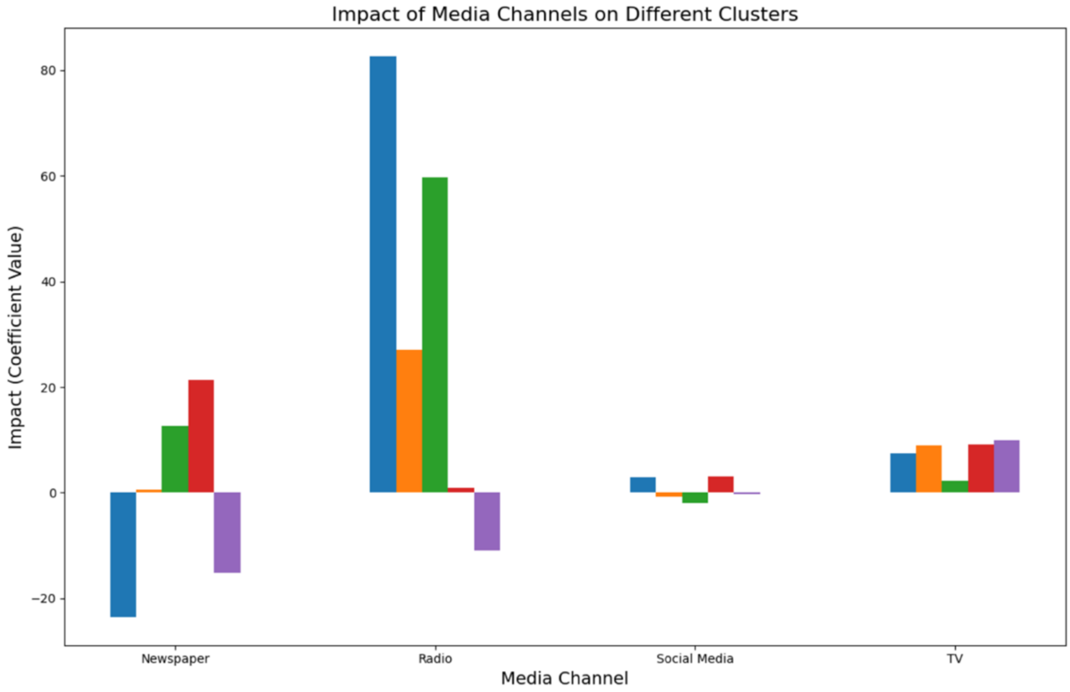


Fig 7.2 Impact of different media on channels on different customer segments

The insights can be stated as –

1. Newspaper positively impacts customer segments 2,3,4 with negative impacts on 1 and 5.

2. Radio positively impacts customer segments 1,2,3,4 with negative impact on 5.

3. Social Media positively impacts segments 1,4 with negative impacts on 2,3 and 5.

4. TV positively impacts customer segments 1,2,3,4,5.

After running the MMM model, we are presented with three key results to analyse and understand – media contribution, ROI and response curves.
Media contribution refers to the quantified impact that different media channels have on a specific marketing objective, such as sales, brand awareness, or customer engagement. It essentially measures how much each channel—be it television, online advertising, social media, etc.—contributes to the overall effectiveness of a marketing campaign or strategy. This is determined by analysing historical data on marketing spends and corresponding outcomes, allowing businesses to understand the return on investment (ROI) for each media channel. For

our project scenario, the specific marketing objective is sales.

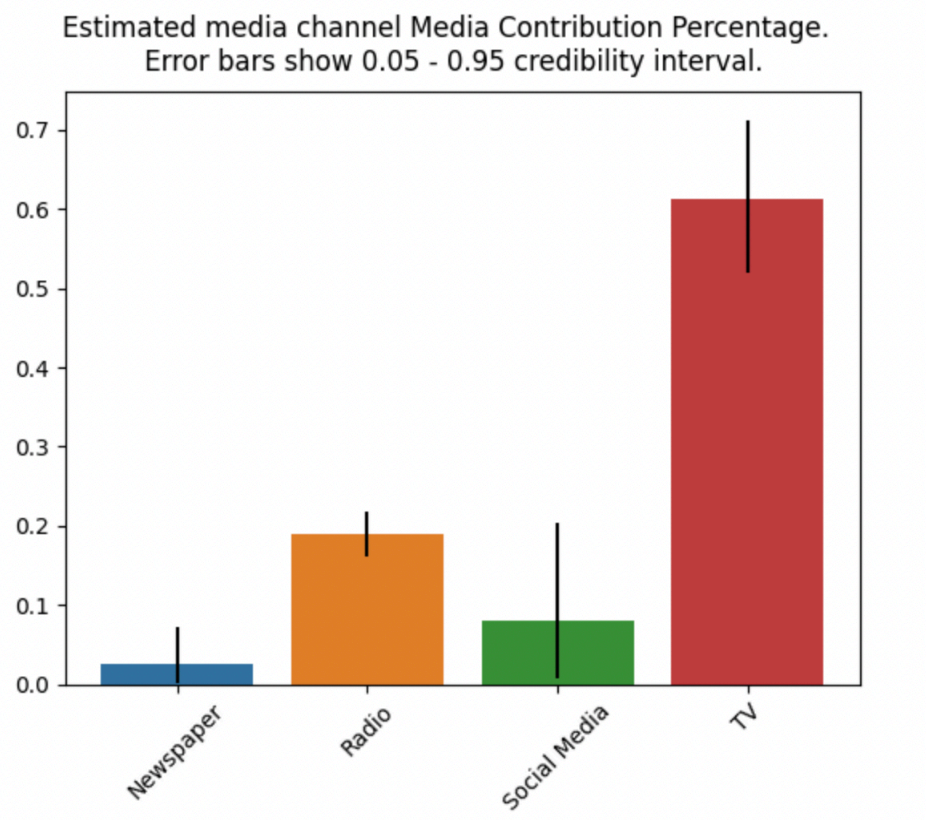


Fig 7.3 Media Contribution

From the above visualisation, we can interpret the contribution of each media channel towards the overall sales(baseline metric) , with TV being the highest contributor at roughly 60%. It is followed by Radio at around 20%, Social Media at around 8% and Newspaper at an estimate of 2% contribution.

The media channel attribution visualisation depicts contribution over a period of time, with the baseline contribution indicative of a baseline level of sales that is expected without any additional media impact.

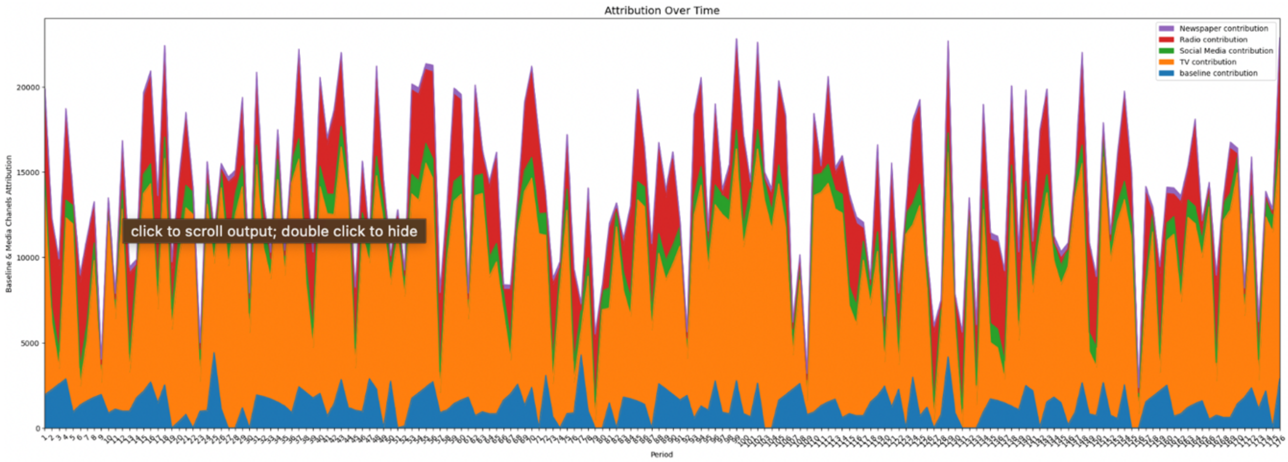


Fig 7.4 Baseline and Media Channel Attribution

Return on Investment (ROI) is a metric that measures the efficiency of various marketing investments. It calculates the return generated from different marketing channels or tactics relative to their costs. ROI is expressed as a percentage or ratio, providing a straightforward way to compare the effectiveness of different marketing efforts.

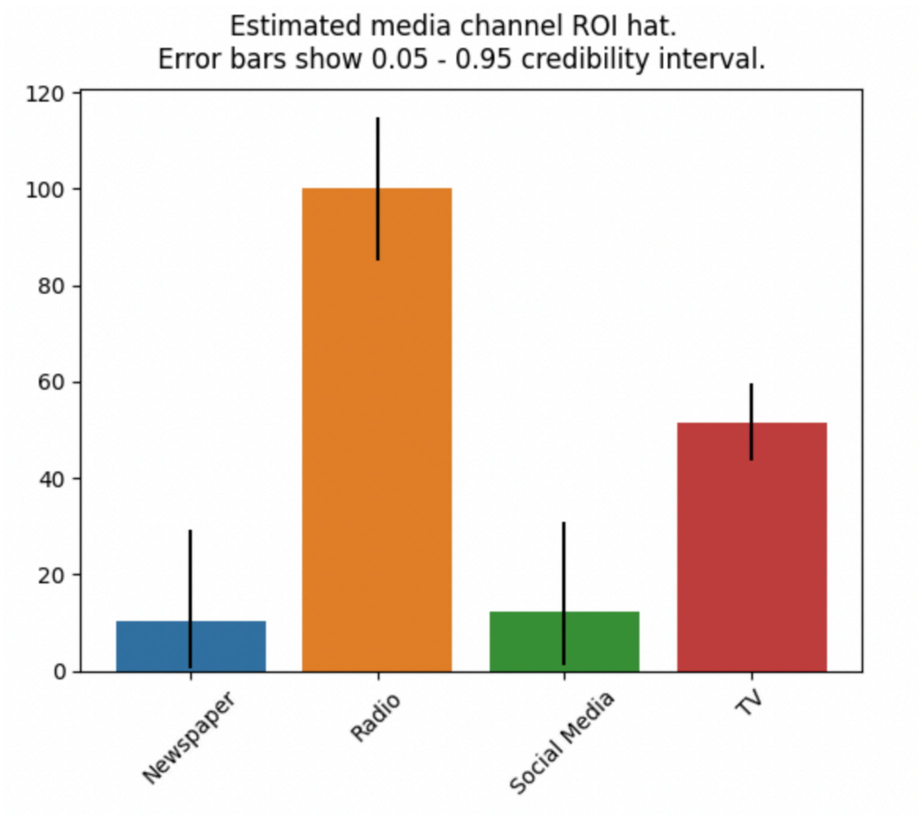


Fig 7.5 ROI of media channels

TV commands the largest share of sales impact, indicating that it is potentially the most effective channel for driving sales volume. This implies that the audience reach and engagement through TV are significant enough to strongly influence the sales figures. On the other hand, Radio, while perhaps not contributing as much to sales volume directly as TV, offers the best ROI. This means that for each dollar spent on Radio advertising, the return in sales is higher compared to TV and other channels. The high ROI on Radio could be due to lower costs associated with this medium or a highly engaged audience that responds well to radio ads, translating to more efficient conversions.

While TV might be essential for visibility and volume, the higher ROI of Radio suggests that increasing the budget allocation there could yield proportionally greater returns.
It is followed by TV, Social Media and Newspaper. This understanding is particularly useful in the budget optimisation process.

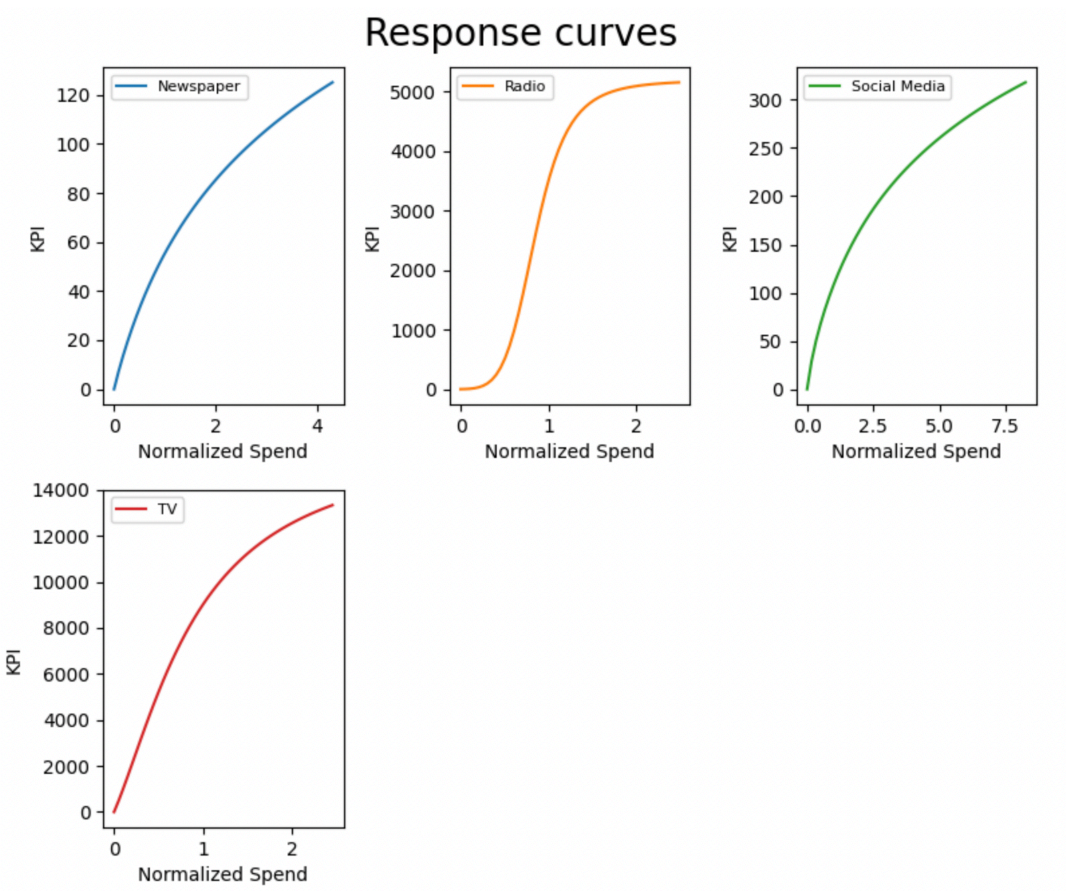


Fig 7.6 Individual response curves of media channels

Marketers make use of marketing response curves and marginal revenue curves that shows the maximum amount required to allocate to a marketing channel before revenue diminishes. The information the response curves provide are translated into budget allocations for each of the marketing channels. This ensures that the maximum ROI is achieved.

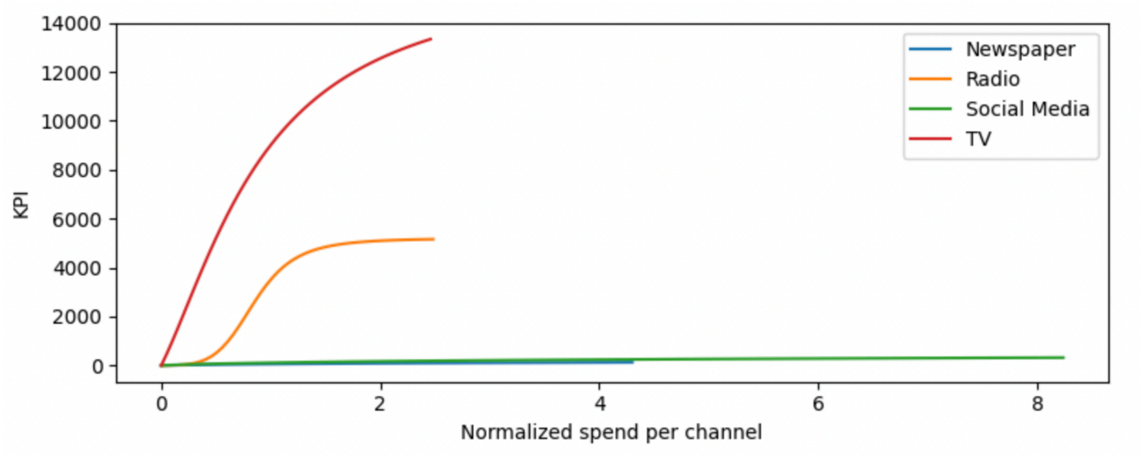


Fig 7.7 Overall response curve

After understanding the input data, the final module is to run the optimisation for sales maximisation. Since the input time series data has weekly data points, we have specified the time period for optimisation as 20 weeks. The optimisation renders a suggestive channel wise budget allocation that would contribute to a significant increase in sales as output.

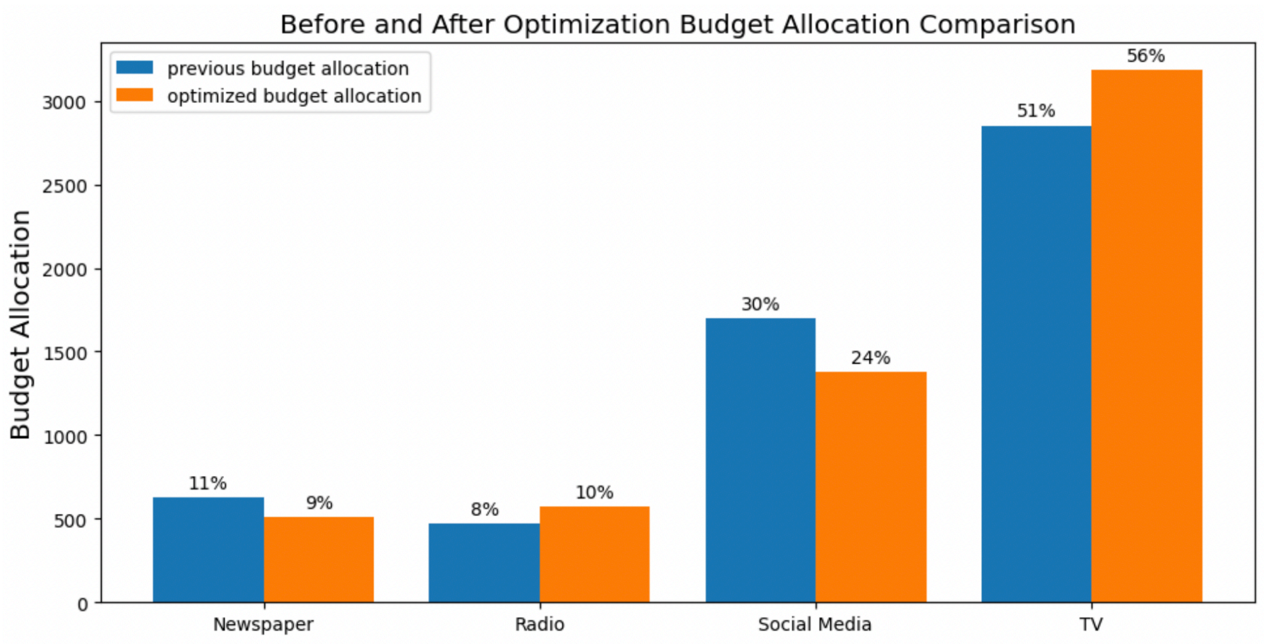


Fig 7.8 Optimal Budget Allocation

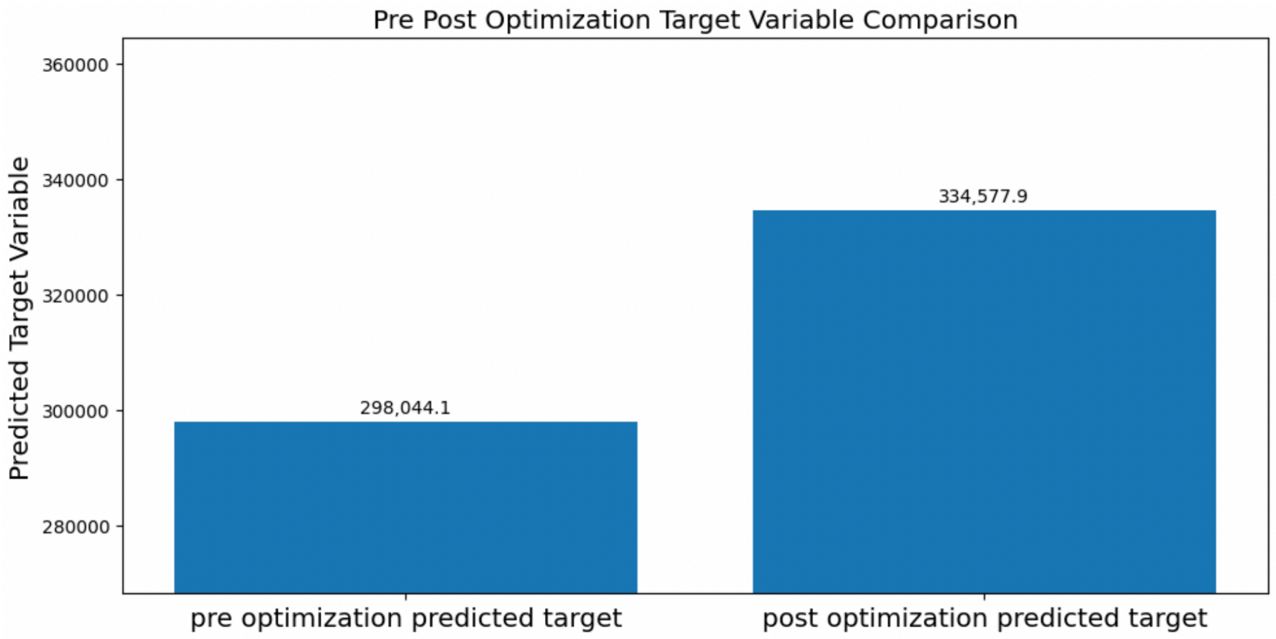


Fig 7.9 Post Optimisation target variable comparison

The insights derived from the model outputs and analysis suggest a strategic adjustment of marketing budgets and efforts across different media channels to better target specific customer segments that are most responsive. Here's an elaboration on the recommended actions:

1. *Media Channel - Newspaper*

The recommendation to decrease the budget for newspaper advertising is likely due to a lower return on investment or less effectiveness compared to other channels. However, there’s still some value in using this channel for reaching customer segments 2 and 3, which may respond more positively to newspaper ads than other segments. The focus here would be on precision targeting within the newspaper medium to maximize impact where it remains relevant.

2. *Media Channel - TV*

TV advertising is shown to be effective and thus warrants an increased budget. This channel's broad reach and influence are particularly significant for customer segments 2, 3, and 5. These segments may have demonstrated a strong engagement with TV ads or a higher likelihood of conversion following exposure to such ads. By boosting investment in TV, the brand aims to capitalize on this channel's strengths to reach and influence these key segments more profoundly.

3. *Media Channel - Radio*

Similar to TV, there is a suggestion to increase the budget for radio advertising. This medium is effective for targeted marketing towards customer segments 2 and 3, which implies these segments may consume radio content frequently or respond well to radio campaigns. Increasing radio spend should focus on times or programs that these segments are most engaged with.

4. *Media Channel - Social Media*

The recommendation to decrease the social media budget indicates that this channel is not performing well in terms of impacting the key customer segments 2, 3, and 5. It's possible these segments are either not as active or receptive on social media, or the social media campaigns are not resonating with them. Given that all focus segments are experiencing a negative impact, efforts here should be minimal and highly strategic. Focusing on segment 5, which is least negatively impacted, suggests trying to refine the approach on social media to this group, possibly by understanding and leveraging the specific aspects of social media that do engage them.

A comparison between our Bayesian model and linear regression (frequentist approach):

We fit a linear regression model to the data and found R^2 to be 0.825 and MAPE to be 0.2716%, as compared to our model with a R^2 value of 0.894 and MAPE = 9.837%. The R^2 value indicates that the Bayesian model was better fit to the data, however the MAPE value is better for linear regression. This indicates that the Bayesian model has a higher percentage of residuals. Bu these metrics are simply not enough to decide which model is better. The Bayesian model incorporates ad stock and saturation advertising principles through transformation functions, along with control variables such as seasonality and holidays. The model is based on Bayesian statistics, which relies on incorporating prior knowledge to generate posterior results. Linear Regression, on the other hand does not incorporate any of these and fails to account for uncertainties within different media channels.
The advantage of Bayesian regression is that, if we use sensible priors, we can still get a decent estimate with a few samples, and the final weights are not a single number, but a distribution composed of every sample drawn during the sampling run.

8. SUMMARY

In this thesis, we looked at using Bayesian Modelling to predict sales based on four advertisement media channel factors (Newspaper, TV, Radio, Social Media), along with understanding consumer psychology and incorporating impact on five identified customer segments based on annual income and spending score. The problem was phased as a machine learning task, with use of algorithms such as K-means, Bayesian Regression and SLSQP. LightweightMMM helped in terms of incorporating our belief, specifying priors and likelihoods.

The ability to predict overall sales can help determine the best amount to allocate to each advertisement channel in order to generate the most interactions or sales. In addition to the standard machine learning models that learn from observations, we successfully applied Bayesian modelling and regression to create a model mapping the features (funds allocated to the marketing channels) to the target (sales). Overall, we were able to derive insights that suggested reallocation of budget across the four media channels to maximise sales, along with a targeted marketing approach that focused on specific customer segments for different media channels.

In conclusion, we were able to successfully implement our proposed model and generate actionable suggestive insights that would prove useful to relevant business stakeholder based on their specific use case. However, we were unable to incorporate control variables such as seasonality and holidays that influence consumer behaviour significantly during customer segmentation due to unavailability of open-source granular datasets. We have assumed that seasonal variation is constant and have thus worked on an additive model, but future work could include working on a multiplicative model which is useful when seasonal variation increases over time. Additionally, there are times when a media channel performs well in one region and does poorly in some other. To address this, implementing a geo-level MMM model is a possibility that can generate more granular insights based on data from specific geographic regions. Dealing with selection bias is crucial for MMM models to improve reliability on results. Selection bias occurs when an input media variable is correlated with an unobserved demand variable (e.g. seasonality, ad targeting), which in turn drives sales. These limitations would require a considerable amount of work, and can be considered for future work and model development.

9. REFERENCES

[1] Yuxue Jin et al. “Bayesian methods for media mix modeling with carryover and shape effects”. In: *Google Research* (2017)

[2] Sandeep Pandey, Snigdha Gupta, Shubham Chhajed, ‘Market Mix Modeling-Concepts and Model Interpretation’, International Journal of Engineering Research and Technology(IJERT), Vol 10 Issue 06 (2021)

[3] Yong Liu, Jorge Laguna, Matt Wright, Hua He, ‘Media Mix Modelling-A Monte Carlo simulation study’, Journal of Marketing Analytics, Vol 2,3 ,176-183

[4] Hao Chen, Minguang Zhang, Lanshan Han, Alvin Ham, ‘Hierarchical marketing Mix Models with Sign constraints’, Journal of Applied Statistics

[5] Sascha Sturze, Markus Hoyer, Claudio Rigetti, Matthias Rasztar, “Multi-touch Attribution and Unified Measurement”, Agile Marketing Performance Management, Springer (2022)

[6] David Chan and Mike Perry, “Challenges and opportunities in media mix modeling”. In: *Google Research* (2017).

[7] Edwin Ng, Zhishi Wang, Athena Dai, “Bayesian Time Varying Coefficient Model with applications to Market Mix Modelling”, In: arXiv:2106.03322v3 (2021)

[8] Kui Zhao, Junhao Hua, Ling Yan, Qi Zhang, Huan Xu, Cheng Yang, “A Unified Framework for Marketing Budget Allocation”, In: KDD’19 (2019)

[9] Yossi Luzon, Rotem Pinchover, Eugene Khmelitsky, “Dynamic budget allocation for social media advertising campaigns: optimisaion and learning”, European Journal of Operations Research, Elsevier (2021)

[10] Vahideh Sadat Abedi, “Allocation of advertising budget between multiple channels to support sales in multiple markets”, Journal of the Operations Research Society (2016)

[11] Tushar Kansal, Suraj Bahuguna, Vishal Singh, Tanupriya Choudhury, “Customer Segmentation using K-Means Clustering”, In: International Conference on Computational Techniques, Electronics and Mechanical Systems (2018)

[12] Musthofa Galih Pradana, Hoang Thi Ha, "Maximising Strategy Improvement in Mall Customer Segmentation Using K-Means clustering", Journal of Applied Data Sciences, Vol 2,1, 19-25 (2021)

[13] Yueqing Wang et al. "A Hierarchical Bayesian Approach to Improve Media Mix Models Using Category Data." In: Google Inc, (2017)

[14] Deloitte, "The future is modeled: A how-to-guide for Advanced Marketing Mix Models" In: Deloitte (2021)

[15] Marketing Evolution. https://www.marketingevolution.com/marketing-essentials/media-mix-modeling

[16] Bayesian Statistics. https://www.analyticsvidhya.com/blog/2016/06/bayesian-statistics-beginners-simple-english/

[17] Mass Analytics. https://mass-analytics.com/marketing-mix-modeling-course/the-marketing-mix-modeling-workflow

[18] Market Mix Modeling. https://towardsdatascience.com/market-mix-modeling-mmm-101-3d094df976f9

[19] Customer Segmentation Using K-Means Clustering. https://sampsonipiankama.medium.com/customer-segmentation-using-k-means-clustering-ae73e3d82934

[20] Media Mix Modelling. https://towardsdatascience.com/media-mix-modeling-how-to-measure-the-effectiveness-of-advertising-with-python-lightweightmmm-b6d7de110ae6

[21] LightweightMMM. https://lightweight-mmm.readthedocs.io/en/latest/api.html#optimize-media

[22] How marketing intelligence is enhancing customer experience. https://www.marketingevolution.com/knowledge-center/how-marketing-intelligence-is-enhancing-customer-experience

[23] Challenges in Marketing Mix Modelling. https://www.lifesight.io/blog/data-challenges-in-marketing-mix-modeling

[24] Optimisation for Marketing Mix Modeling. https://mass-analytics.com/marketing-mix-modeling-blogs/optimization-for-marketing-mix-modeling-what-you-need-to-know/

[26] Understanding Bayesian Marketing Mix Modeling. https://towardsdatascience.com/understanding-bayesian-marketing-mix-modeling-a-deep-dive-into-prior-specifications-af400adb836e

[27] Practical Approaches to Optimising Budget in Marketing Mix Modeling. https://towardsdatascience.com/practical-approaches-to-optimizng-budget-in-marketing-mix-modeling-7816a27f2f71

APPENDIX A- SAMPLE CODE

Phase 1:

```
import numpy as np
import pandas as pd
from matplotlib import pyplot as plt
import seaborn as sns
from sklearn.cluster import KMeans
from sklearn.preprocessing import MinMaxScaler

df= pd.read_csv('/Users/rachitasingh/Desktop/Capstone/Customer_Data.csv', index_col = 0)
df.head()

plt.figure(1 , figsize = (15 , 6))
n = 0
for x in ['Age' , 'Annual Income (k$)' , 'Spending Score (1-100)']:
    n += 1
    plt.subplot(1 , 3 , n)
    plt.subplots_adjust(hspace =0.5 , wspace = 0.5)
    sns.distplot(df[x] , bins = 20)
    plt.title('Distplot of {}'.format(x))
plt.show()
plt.figure(1 , figsize = (15 , 5))
sns.countplot(y = 'Gender' , data = df)
plt.show()
plt.figure(1 , figsize = (15 , 6))
for gender in ['Male' , 'Female']:
    plt.scatter(x = 'Age' , y = 'Annual Income (k$)' , data = df[df['Gender'] == gender] ,
                s = 200 , alpha = 0.5 , label = gender)
plt.xlabel('Age'), plt.ylabel('Annual Income (k$)')
plt.title('Age vs Annual Income w.r.t Gender')
plt.legend()
plt.show()
plt.figure(1 , figsize = (15 , 6))
for gender in ['Male' , 'Female']:
    plt.scatter(x = 'Annual Income (k$)',y = 'Spending Score (1-100)' ,
                data = df[df['Gender'] == gender] ,s = 200 , alpha = 0.5 , label = gender)
plt.xlabel('Annual Income (k$)'), plt.ylabel('Spending Score (1-100)')
plt.title('Annual Income vs Spending Score w.r.t Gender')
plt.legend()
plt.show()
X = df.loc[:,['Annual Income (k$)','Spending Score (1-100)']].values
scaler = MinMaxScaler().fit(X) #It makes an object of the MinMaxScaler and then we fit it on our variable X.
print(scaler)
scaler.transform(X)
plt.figure(figsize = (12,6))
```

```
plt.grid()
plt.plot(range(1,11),wcss, color='green', linestyle='dashed', linewidth = 3,
         marker='o', markerfacecolor='blue', markersize=12)
plt.title('The Elbow Point Graph')
plt.xlabel('Number of Clusters')
plt.ylabel('WCSS')
plt.show
kmeans= KMeans(n_clusters = 5, init = 'k-means++')
label= kmeans.fit_predict(X) #returns a cluster number for each of the data points
plt.figure(figsize=(8,8))
plt.scatter(X[label == 0,0], X[label== 0,1], s=50, c='green', label='Cluster 1')
plt.scatter(X[label == 1,0], X[label== 1,1], s=50, c='yellow', label='Cluster 2')
plt.scatter(X[label == 2,0], X[label== 2,1], s=50, c='red', label='Cluster 3')
plt.scatter(X[label == 3,0], X[label== 3,1], s=50, c='purple', label='Cluster 4')
plt.scatter(X[label == 4,0], X[label== 4,1], s=50, c='blue', label='Cluster 5')
plt.scatter(kmeans.cluster_centers_ [:,0], kmeans.cluster_centers_ [:,1], s= 100, c='black', marker= '*', label='Cen
plt.title('Customer groups')
plt.xlabel('Annual Income')
plt.ylabel('Spending Score (1-100)')
plt.legend()
plt.show()
```

Phase 2:

```
import jax.numpy as jnp
import numpyro
# Import the relevant modules of the library
from lightweight_mmm import lightweight_mmm
from lightweight_mmm import optimize_media
from lightweight_mmm import plot
from lightweight_mmm import preprocessing
from lightweight_mmm import utils
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
import plotly.graph_objects as go
import statsmodels.api as sm

df_main = pd.read_csv("/Users/rachitasingh/Desktop/Capstone/MMM_Data.csv")
df_main.head()
# 1. media variables
mdsp_cols = ["Newspaper", "Radio", "Social Media", "TV"]
# 2. control variables
hldy_cols = [col for col in df_main.columns if 'hldy_' in col]
seas_cols = [col for col in df_main.columns if 'seas_' in col]
control_vars =  hldy_cols + seas_cols
# 3. sales variables
sales_cols =['Sales']
clusters = ['Cluster_0.0', 'Cluster_1.0', 'Cluster_2.0', 'Cluster_3.0', 'Cluster_4.0']
SEED = 105
data_size = len(df_main)
n_media_channels = len(mdsp_cols)
n_extra_features = len(control_vars)
media_data = df_main[mdsp_cols].to_numpy()
extra_features = df_main[control_vars].to_numpy()
target = df_main['Sales'].to_numpy()
costs = df_main[mdsp_cols].sum().to_numpy()
test_data_period_size = 24
split_point = data_size - test_data_period_size
# Media data
media_data_train = media_data[:split_point, ...]
media_data_test = media_data[split_point:, ...]
# Extra features
extra_features_train = extra_features[:split_point, ...]
```

```
extra_features_test = extra_features[split_point:, ...]
# Target
target_train = target[:split_point]
media_scaler = preprocessing.CustomScaler(divide_operation=jnp.mean)
extra_features_scaler = preprocessing.CustomScaler(divide_operation=jnp.mean)
target_scaler = preprocessing.CustomScaler(divide_operation=jnp.mean)
cost_scaler = preprocessing.CustomScaler(divide_operation=jnp.mean, multiply_by=0.15)
media_data_train = media_scaler.fit_transform(media_data_train)
extra_features_train = extra_features_scaler.fit_transform(extra_features_train)
target_train = target_scaler.fit_transform(target_train)
costs = cost_scaler.fit_transform(costs)
# Extracting coefficients for interaction terms from the model
interaction_coeffs = {term: coef for term, coef in model.params.items() if '_X_' in term}
media_channels = ["Newspaper", "Radio", "Social Media", "TV"]
clusters = ['Cluster_1', 'Cluster_2', 'Cluster_3', 'Cluster_4', 'Cluster_5']
n_clusters = len(clusters)
cluster_positions = np.arange(len(media_channels))
bar_width = 0.1
plt.figure(figsize=(14, 8))
for i, cluster in enumerate(clusters):
    impacts = [interaction_coeffs[f'{media}_X_{cluster}'] for media in media_channels]
    plt.bar(cluster_positions + i * bar_width, impacts, width=bar_width, label=cluster)
plt.xlabel('Media Channel', fontsize=14)
plt.ylabel('Impact (Coefficient Value)', fontsize=14)
plt.title('Impact of Media Channels on Different Clusters', fontsize=16)
plt.xticks(cluster_positions + bar_width * (n_clusters - 1) / 2, media_channels)
plt.legend(title='Cluster', bbox_to_anchor=(1.05, 1), loc='upper left')
plt.tight_layout()
plt.show()
mmm = lightweight_mmm.LightweightMMM(model_name="hill_adstock")
number_warmup=1000
number_samples=1000
mmm.fit( media=media_data_train, media_prior=costs, target=target_train, extra_features=extra_features_train, number
plot.plot_model_fit(mmm, target_scaler=target_scaler)
media_contribution, roi_hat = mmm.get_posterior_metrics(target_scaler=target_scaler, cost_scaler=cost_scaler)
plot.plot_media_baseline_contribution_area_plot(media_mix_model=mmm,
                                                target_scaler=target_scaler,
                                                fig_size=(30,10),
                                                channel_names = mdsp_cols
                                                )
plot.plot_bars_media_metrics(metric=media_contribution, metric_name="Media Contribution Percentage", channel_names=m
plot.plot_bars_media_metrics(metric=roi_hat, metric_name="ROI hat", channel_names=mdsp_cols)
```

```
plot.plot_response_curves(media_mix_model=mmm, target_scaler=target_scaler, seed=SEED)
prices = jnp.ones(mmm.n_media_channels)
n_time_periods = 20
budget = jnp.sum(jnp.dot(prices, media_data.mean(axis=0)))* n_time_periods
solution, kpi_without_optim, previous_media_allocation = optimize_media.find_optimal_budgets(
    n_time_periods=n_time_periods,
    media_mix_model=mmm,
    extra_features=extra_features_scaler.transform(extra_features_test)[:n_time_periods],
    budget=budget,
    prices=prices,
    media_scaler=media_scaler,
    target_scaler=target_scaler,
    seed=SEED)
optimal_buget_allocation = prices * solution.x
optimal_buget_allocation
# Plot out pre post optimization budget allocation and predicted target variable comparison.
plot.plot_pre_post_budget_allocation_comparison(media_mix_model=mmm,
                                                kpi_with_optim=solution['fun'],
                                                kpi_without_optim=kpi_without_optim,
                                                optimal_buget_allocation=optimal_buget_allocation,
                                                previous_budget_allocation=previous_budget_allocation,
                                                figure_size=(10,10),
                                                channel_names = mdsp_cols)
```