Unveiling Marketing Campaign Insights and Visualization

Sneha Naidu

Abstract—Put here a brief summary of your work: analysis task, data, approach, main findings. Length: up to 200 words.

1 Problem Statement

The analysis aims to investigate the performance of marketing campaigns with a focus on identifying key factors influencing profitability, return on investment (ROI), and customer acquisition. We'll explore how different campaign factors—like budget allocation, subscription tiers, keywords, and customer satisfaction—affect overall success. By addressing key questions, the goal is to uncover insights that can help improve future marketing strategies.

The main focus here is the performance of marketing campaigns, which is essential for businesses aiming to boost ROI while growing their customer base sustainably. Some key questions include:

1. How does the budget influence the revenue generated across different subscription tiers?
2. What factors contribute to higher ROI and customer satisfaction?
3. Which subscription tiers and campaign keywords drive the most sales?
4. How do different campaigns perform in terms of overall metrics?

The dataset used in this analysis includes key variables like campaign budgets, subscription tiers, ROI, revenue, keywords, and customer satisfaction. These variables are perfect for answering our research questions since they cover the most important aspects of marketing performance. The structure of the dataset allows for a thorough exploration of relationships and trends, with visual analytics helping to spot hidden patterns and insights.

Ultimately, this analysis will give a clearer picture of what drives marketing success, helping decision-makers fine-tune their strategies for better profitability and customer engagement.

2 State of the Art

The foundation of my project was built upon insights gained from an article *"Visualization Approaches Integrating Real-Time Market Data"* by Mark J. Laufenberg [1], who analyzed real-time market data in power systems, focusing on visualization techniques to aid decision-making. This study explored the usability of visual tools such as pie charts, contour maps, and dynamic visualizations, aiming to enhance understanding among market participants. Inspired by their emphasis on user-friendly visualizations, I adopted similar principles in my Marketing Campaign Analysis project. By creating scatter plots to examine budget versus revenue and bar plots to compare units sold across subscription tiers, we aimed to ensure our visualizations were both intuitive and impactful. The article's exploration of data integration and usability profoundly shaped my approach to presenting campaign performance insights.

Building on this foundation, the paper *"Visualizing Results of Promoting Campaigns"* [2]focused on sales representatives' performance data, with an emphasis on exploratory data analysis (EDA) and the detection of anomalies such as outliers and fraudulent reporting. This study highlighted the importance of effective visualization in communicating findings clearly. Taking inspiration from their work, we applied similar EDA and data cleaning techniques to my analysis. Our visualizations—ranging from scatter plots and bar charts to heatmaps and pie charts—were designed to uncover insights into campaign performance, such as budget efficiency, revenue trends, and customer satisfaction by subscription tier. Their methodological approach also encouraged me to adopt advanced visualizations, such as Sankey diagrams, to track the flow of budget allocation and its impact on ROI.

Another study, "*Improving Data Analysis and Visualization in Market Research with Tableau"* by Jatin Thakur and Prasenjit Das [3], focused on consumer behavior and market trends using Tableau, employing interactive visualizations and descriptive statistics to identify patterns in sales data and customer preferences. The study demonstrated the importance of interactive dashboards in fostering real-time collaboration and improving decision-making processes. Inspired by this approach, Python libraries such as Plotly and Bokeh were utilized to create dynamic visualizations for the Marketing Campaign Analysis project. These tools enabled stakeholders to explore data interactively, enhancing the clarity and engagement of the insights. The emphasis on interactivity highlighted the value of making data accessible and actionable for diverse audiences.

Finally, the study by Alaa Abu-Srhan and Sanaa Al Zghoul, *"Visualization and Analysis in Bank Direct Marketing Prediction"* [4], analyzed data from a Portuguese banking institution's direct marketing campaigns, consisting of 4,521 instances and 17 features. The research addressed challenges posed by imbalanced datasets and explored various visualization techniques and oversampling methods. The Naive Bayes classifier was identified as the most effective predictive model. Drawing from these insights, a multiclass classification approach using the Random Forest classifier was employed to analyze customer behavior in marketing campaigns. The study's focus on robust preprocessing and visualization techniques informed the methodology, ensuring the accurate handling of data complexities and effective prediction of customer engagement.

Together, these studies provided a comprehensive foundation for my project, influencing not only my technical approach but also my commitment to clear, insightful, and actionable visualizations. Each paper offered unique perspectives that collectively enriched my analysis, enabling me to bridge the gap between raw data and meaningful marketing insights.

3 PROPERTIES OF THE DATA

3.1 DATA COLLECTION:

The dataset for this project was sourced from Kaggle and is composed of synthetic data related to marketing campaigns. It includes 17 fields and a total of 10,000 records, offering insights into various campaign aspects like budgets, clicks, conversions, revenues, and customer satisfaction.

3.2 DATA QUALITY AND PREPROCESSING:

1. Missing Value

A few missing values were identified in columns like Budget (6 values), Subscription Tier (2 values), and Common Keywords (6 values). Numerical gaps, such as those in the Budget column, were filled using the median or mean, while categorical gaps, like in Subscription Tier, were imputed with the mode.

2. Outlier Detection:

The dataset did not exhibit major outliers due to its synthetic nature. To gain better insight into the data spread, boxplots with overlaid red dots representing individual data points were created. These swarmplots provided a detailed view of the data distribution and confirmed the absence of extreme values.

3.3 DATA EXPLORATION AND INSIGHTS:

1. Correlation Analysis: Fig(1)

We created a correlation matrix to find relationships between different features. Notably, a strong positive correlation exists between Budget and Profitability Level, indicating that higher budgets are generally associated with increased profitability. Additionally, we observe a moderate positive correlation between Clicks and ROI, suggesting that campaigns with higher click-through rates tend to have better returns on investment. However, a weak negative correlation exists between Revenue Generated and Common Keywords, suggesting that increasing the number of common keywords may have a slightly negative impact on revenue generation.

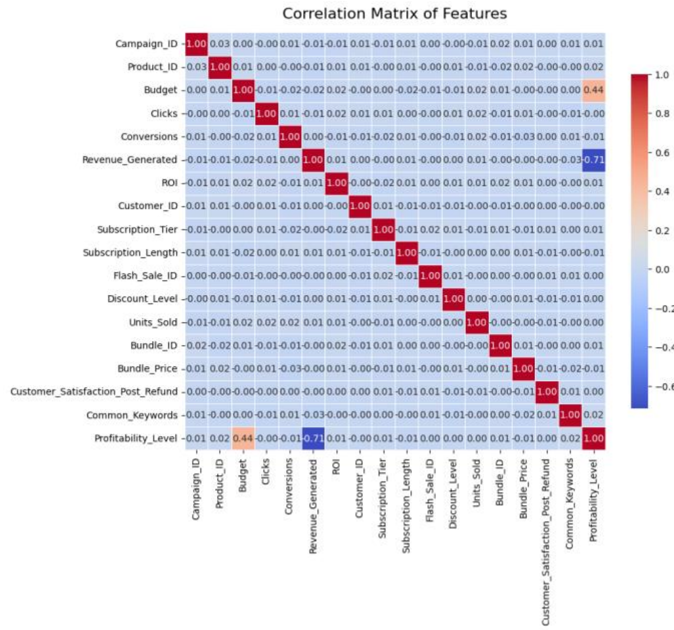

Fig. 1 Correlation Matrix of Features

2. Principal Component Analysis (PCA):

To understand the data, we applied Principal Component Analysis (PCA) to reduce dimensionality and visualize it in two dimensions. This revealed a clear clustering structure, with three distinct groups separated primarily along the first principal component. The PCA projection simplified the data, making it easier to interpret and identify potential segments, while highlighting important features aligned with the first principal component.

Next, I used KMeans clustering on the PCA-reduced data to group the campaigns into three clusters based on profitability levels, enabling more effective analysis of feature relationships.

Finally, I created a heatmap to visualize feature importance within each cluster (Fig 2). The heatmap revealed:

- Cluster 0: Negative importance for profitability and budget, suggesting these factors are negatively correlated with performance.
- Cluster 1: Positive importance for revenue-related features and negative importance for common keywords, indicating that excessive keywords may be harmful.
- Cluster 2: Strong positive importance for profitability but negative importance for revenue, suggesting that high revenue doesn't always lead to high profitability.

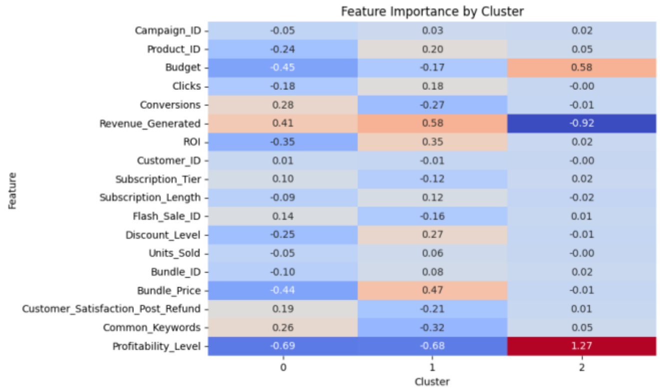

Fig. 2 Heatmap: Feature Importance by Cluster

In summary, PCA helped reduce the data to a manageable form, KMeans clustering segmented the data, and the heatmap provided insight into the relationships between features within each cluster.

3.4 Feature Engineering :

Additionally, a new feature, "Profitability Levels," was created by calculating the Revenue-to-Budget ratio. Campaigns were categorized as Not Profitable, Moderately Profitable, or Highly Profitable. This feature offers a clear perspective on the financial success of campaigns and is highly relevant for deeper analysis and modelling.

3.5 Feature Transformation:

- Categorical columns were converted to numerical values using Label Encoding.
- All features were normalized using MinMaxScaler to ensure uniform scaling.

4 Analysis

4.1 Analysis Approach

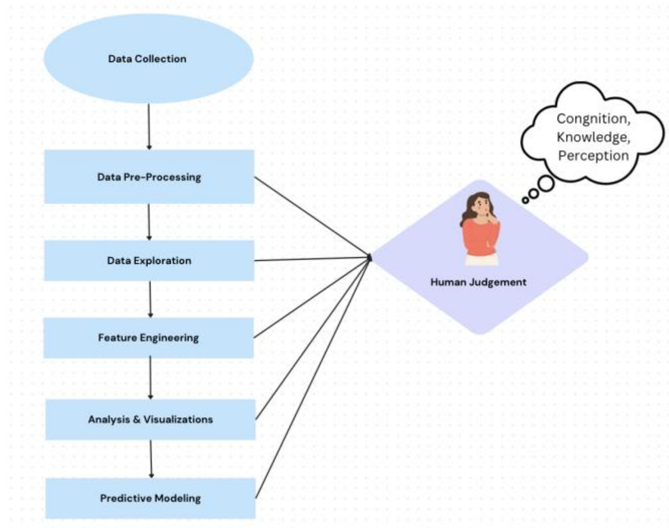

Fig. 3 Workflow Diagram of Analysis Approach

The analytical workflow for this project was designed to integrate human reasoning and computational methods, ensuring that each step was purposeful, data-driven, and involved informed decision-making. The complexity of the problem required iterative exploration, where visualizations and computational methods provided insights, but critical decisions and interpretations were driven by human judgment. Below is the approach, emphasizing the flow of reasoning and how each decision informed subsequent steps.

1. Research Questions Framing

Human Reasoning:
The analysis began by defining key research questions based on the business problem. For example:

- How does budget allocation impact revenue across subscription tiers?
- What are the most effective campaign keywords driving revenue?

This stage required understanding the business context, stakeholder priorities, and the data available. Human reasoning was central to identifying which questions were relevant and actionable, ensuring the analysis would yield meaningful insights.

2. Visualization-Driven Exploration

Human Reasoning and Decision-Making:
To answer these questions, appropriate visualizations were chosen iteratively. For instance:

- **Why Scatter Plots First?**
  Scatter plots were selected to examine relationships between numerical variables such as budget, revenue, and ROI. The decision to use scatter plots was driven by the need to visually assess patterns or correlations that could guide further analysis.

- **What Informed the Next Step?**
  Upon observing a strong relationship between budget and revenue, bar plots were created to further break down revenue by subscription tier and campaign keywords. This choice was informed by the human analyst’s interpretation that categorical variables might provide additional insights into specific tiers or campaigns.

- **Why Use Sankey Diagrams?**
  The need to understand flows between subscription tiers and campaign keywords led to the selection of a Sankey diagram. This visualization was specifically chosen to represent the interconnectedness and relative impact of keywords on tiers, which could not be conveyed effectively through other visualizations.

Each visualization was chosen not merely for its computational feasibility but based on human judgment about what patterns or trends might answer the research questions or suggest new ones.

3. Predictive Modeling

Why Multiclass Classification?
The final stage involved developing a multiclass classification model using Random Forest to predict campaign profitability. The model’s purpose was to provide a scalable and systematic way to classify campaigns into profitability tiers, using features such as budget, subscription tier, and keywords.

At each stage, human judgment was required to evaluate whether the patterns seen in the visualizations warranted further exploration or were sufficient for decision-making.

This workflow deliberately separated tasks suited to computational methods from those requiring human reasoning:

- **Computers excelled** at processing large datasets, generating visualizations, and building predictive models.

- **Humans excelled** at defining the problem, selecting appropriate methods, interpreting complex patterns, and making strategic decisions.

4.2 PROCESS

In this section, we detail the step-by-step approach used in our analysis, combining visual and computational techniques with human judgment to answer our research questions. Each step was carefully chosen for its relevance to the objectives, ensuring a logical flow informed by data insights. Visualizations played a key role in refining our understanding, while human reasoning guided interpretations and actionable conclusions.

4.2.1 RESEARCH QUESTION 1: HOW DOES THE BUDGET INFLUENCE THE REVENUE GENERATED ACROSS DIFFERENT SUBSCRIPTION TIERS?

Scatter Plot of Budget vs. Revenue by Subscription Tier

To explore the relationship between budget allocation and revenue generation, we began with a scatter plot segmented by subscription tiers (Basic, Standard, and Premium). This approach helped identify trends or anomalies in how budget size affects revenue across these tiers.
The plot visualizes the relationship between "Budget" and "Revenue Generated" for three different subscription tiers: Standard, Basic, and Premium.

Observations: Fig(4)

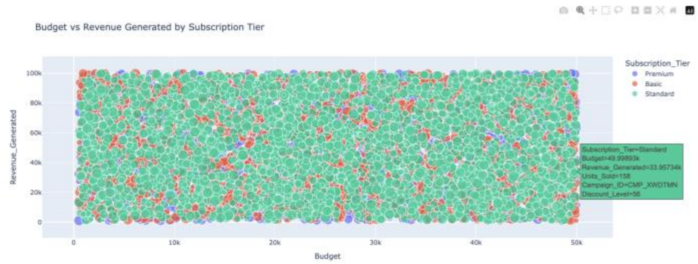

Fig. 4 Scatter Plot: Budget vs Revenue Generated by Subscription Tier

Tier-wise Distribution:

- **Standard:** The majority of data points appear to belong to the Standard tier. These points are spread across the plot, suggesting a wide range of revenue generation for different budget levels.
- **Basic:** The Basic tier has a moderate number of data points, also spread across the plot with a similar pattern as Standard.
- **Premium:** The Premium tier has the fewest data points, suggesting that it might be the least subscribed or generate less revenue compared to Standard and Basic.

The scatter plot revealed that revenue generation isn’t a simple linear function of budget increases—it’s likely influenced by other factors. Based on this, we explored keyword contributions to revenue through treemaps.

Treemap of Revenue by Subscription Tier and Keywords

In this step, treemaps were used to uncover which keywords drove the most revenue within each subscription tier (Basic, Standard, and Premium). This visual representation allowed us to identify patterns and trends in keyword performance, providing insights into consumer preferences across different tiers.

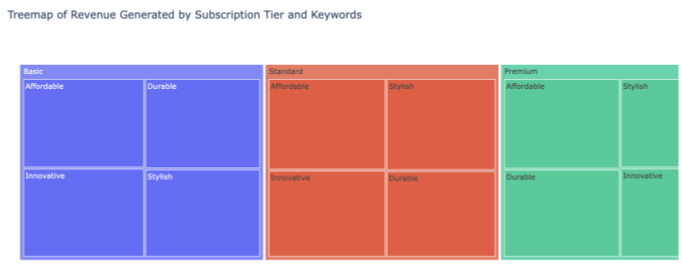

Fig. 5 Treemap of Revenue Generated by Subscription Tier and Keywords

Key Insights from the Visualization: Fig(5)

- **Basic Tier:**
  "Affordable" emerged as the top-performing keyword, generating $44.1 million in revenue, indicating its strong appeal to budget-conscious customers. "Durable" followed closely at $42 million, suggesting that customers in this tier value longevity and quality in products. Meanwhile, "Innovative" and "Stylish" each generated $41 million, showing comparable interest in modern features and design.
- **Standard Tier:**
  In this tier, "Affordable" once again took the lead, generating $43 million in revenue. This suggests that affordability continues to be a critical factor, even for mid-tier subscribers. "Durable" also performed well, contributing $42 million, while "Innovative" and "Stylish" brought in $41 million and $40 million, respectively. The close revenue figures reflect a balanced interest in affordability, quality, and design for Standard subscribers.
- **Premium Tier:**
  Among Premium subscribers, "Affordable" and "Durable" both led the way, each generating $42 million in revenue. This finding indicates that even high-tier subscribers prioritize value and durability in their purchasing decisions. "Innovative" and "Stylish" followed closely, each contributing $40 million, highlighting that Premium customers also appreciate cutting-edge features and aesthetics.

While the revenue generated by keywords was relatively consistent across the tiers, slight differences in emphasis were observed. For example, "Affordable" consistently performed well across all tiers, but it had the strongest impact in the Basic Tier, likely due to cost-conscious subscribers. Similarly, "Durable" maintained strong performance across all tiers, reflecting a universal preference for products with long-lasting value.

These insights suggest opportunities for targeted strategies. For instance, focusing on high-performing keywords like "Affordable" and "Durable" could help enhance revenue in underperforming tiers. Additionally, the consistent performance of "Innovative" and "Stylish" highlights the importance of maintaining these qualities across campaigns. Building on these findings, we moved on to analyze ROI and customer satisfaction metrics to refine our overall strategy further.

4.2.2 Research Question 2: What factors contribute to higher ROI and customer satisf action?

Scatter Plot of Revenue, ROI, and Customer Satisfaction
To examine the relationships between ROI, customer satisfaction, and revenue generation, we created a scatter plot with a color gradient representing satisfaction levels. This helped us visualize how financial and customer-related outcomes were interconnected.

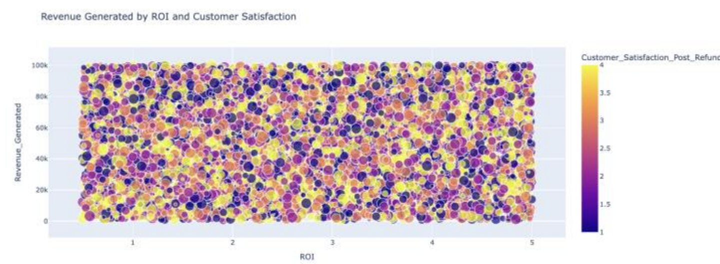

Fig. 6 Scatter Plot: Revenue Generated by ROI and Customer Satisfaction

Key Insights from Visualization: Fig (6)

- **Data Distribution:** The data points are densely clustered, showing a wide range of ROI values.
- **Revenue Variation:** Revenue Generated varies significantly, ranging from near $0 to close to $100k.
- **ROI and Revenue Relationship:** There's a general trend where higher ROI values often align with higher revenue. However, this relationship isn't consistent—some high ROI campaigns have low revenue, and vice versa.
- **Customer Satisfaction:** The color-coding for customer satisfaction post-refund shows no clear pattern or correlation with ROI or revenue, suggesting satisfaction may not directly drive revenue in this context.

The lack of a strong pattern implies customer satisfaction may not be the primary driver of revenue, though further analysis is needed to confirm this.
The dense clustering of points makes it difficult to analyze individual data characteristics, indicating the need for refined visualization or aggregation techniques.

Bar Plot: Total Revenue Generated by Subscription Tier and Campaign Keywords

This analysis visualizes the total revenue generated by each subscription tier and the performance of key campaign keywords.

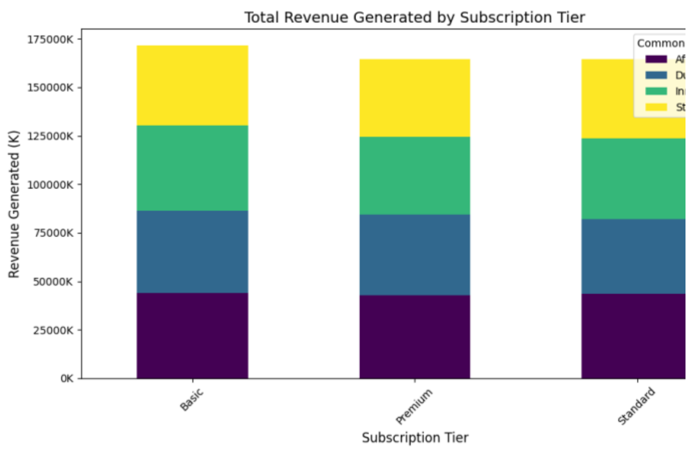

Fig. 7 Stacked Bar Chart: Total Revenue Generated by Subscription Tier

Observations: Fig(7)

- **Revenue by Tier:**
  The **Basic** tier generates the highest revenue, with a total close to **$100M**, followed by the **Standard** tier at approximately **$40M**, and the **Premium** tier at **$20M**.
- **Keyword Performance:** Stylish: The keyword "Stylish" consistently contributes to a significant portion of revenue across all tiers. This suggests that style is a key factor influencing subscription choices. Durable: The keyword "Durable" also performs well, indicating a preference for products that are built to last. Innovative: The keyword "Innovative" shows strong performance in the Premium and Standard tiers, suggesting that subscribers in these tiers are more willing to pay for cutting-edge features. Affordable: The keyword "Affordable" contributes the least to revenue across all tiers. This could be due to the perception of higher-priced subscriptions being associated with better quality or features.

These observations set the stage for understanding how subscription tiers and keywords relate to higher sales.

4.2.3 Research Question 3: Which Subscription Tiers and Campaign Keywords Drive the Most Sales?

Sankey Diagram: Subscription Tiers and Campaign Keywords

The Sankey diagram illustrates the relationship between subscription tiers (Basic, Standard, and Premium) and campaign keywords (Affordable, Innovative, and Durable). The thickness of the connecting lines represents the relative volume of traffic or conversions associated with each keyword and subscription tier.

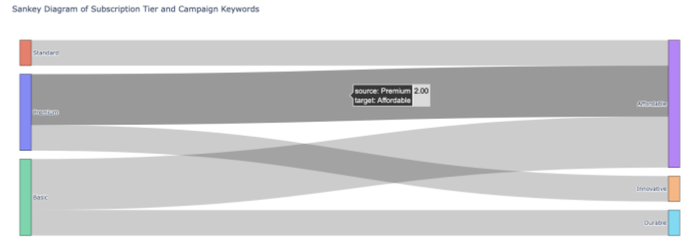

Fig. 8 San Sankey Diagram: Subscription Tier and Campaign Keywords

Observations: Fig(8)

- The **Premium Tier** has the strongest affinity with the **Affordable** keyword, suggesting that customers in this tier prioritize affordability even when subscribing to higher-priced plans.
- The **Basic Tier** is closely associated with the **Innovative** keyword, indicating that customers in this group are drawn to cutting-edge or unique products.
- The **Standard Tier** has a balanced distribution across all three keywords, suggesting that this group values a mix of affordability, innovation, and durability.

A significant overlap exists between **Affordable** and **Durable** keywords, especially within the **Premium Tier**. This highlights that customers in this group are seeking products that offer both value for money and long-lasting quality. There's also some overlap between **Innovative** and **Durable** keywords, predominantly within the **Basic Tier**, suggesting that these customers want products that are not only innovative but also reliable and built to last.

These insights illustrated the complex interactions between keywords and subscription tiers, aiding in understanding how to target campaigns effectively.

4.2.4 Research Question 4: How do different campaigns perform in terms of overall metrics?

Parallel Coordinates of Campaign Performance

We visualized campaign performance metrics such as budget, clicks, conversions, revenue, and ROI using a parallel coordinates plot. This allowed us to assess how different campaigns compared across various metrics.

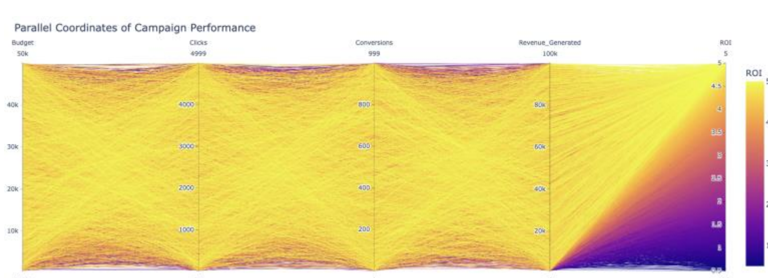

Fig. 9 Parallel Coordinates of Campaign Performance

Observations: Fig(9)

- ***Budget vs. ROI:*** Higher budgets don't always lead to better ROI. Some high-budget campaigns exhibit lower ROI (**purple region**). A few campaigns with moderate budgets achieve high ROI (**yellow lines**).
- *Clicks and Conversions:* Campaigns with higher clicks generally correlate with higher conversions, though this doesn't always lead to high ROI. Some campaigns achieve high conversions at relatively lower click levels.
- *Revenue Generated:* Campaigns generating higher revenue tend to have higher ROI (**bright yellow region** on the far right). However, a portion of low-revenue campaigns is associated with lower ROI (**purple region**), indicating inefficiencies.

High ROI campaigns (yellow lines) dominate the data, reflecting efficient utilization of budgets, clicks, and conversions for most campaigns. **Low ROI campaigns (purple lines)** are less frequent, primarily observed in cases of low revenue or inefficiencies in cost management.
The analysis process concluded when consistent patterns emerged across multiple visualizations, aligning with the research questions. Each step refined the findings, and the insights from one visualization informed the next, ensuring a comprehensive understanding of the data.

After conducting various visualizations and uncovering key insights, we decided to proceed with modeling to predict profitability levels more effectively. The purpose of modeling is to create a predictive model that can generalize the patterns we observed, ultimately providing actionable insights and decision-making tools for future campaigns.

4.2.5 MODELING:

For our modeling, we used the **Profitability Level** feature, which was created during feature engineering, as the target variable. Since the feature has three distinct values, it naturally became a multiclass classification problem. To evaluate its predictive performance, we applied a **Random Forest Classifier**, a robust and widely used machine learning model known for its ability to handle complex datasets and provide valuable insights into feature importance.

Model Evaluation: Random Forest Classifier
Classification Report:

	Precision	Recall	F1-Score	Support
Class 0	0.98	1.00	0.99	1246
Class 1	0.95	0.83	0.88	253
Class 2	0.97	0.98	0.97	501
Accuracy			0.97	2000
Macro Avg	0.96	0.94	0.95	2000
Weighted Avg	0.97	0.97	0.97	2000

- **Recall Score (weighted)**: 0.97
- **F1-Score (weighted)**: 0.97
- **AUC Score (weighted, OVR)**: 1.00

Observations from the Matrix:

- **Class 0**: This class is predicted with high accuracy, showing very few misclassifications.
- **Class 1**: This class has a higher rate of misclassifications, particularly into Class 0 and Class 2.
- **Class 2**: Shows strong performance with a high number of correct predictions, similar to Class 0.

These results indicate that the Random Forest model performs well overall, with some challenges in predicting Class 1

4.3 Results

Our analysis revealed key insights into how budget, subscription tiers, and keywords impact revenue generation, ROI, and customer satisfaction. Through visualizations, we observed that while higher budgets generally align with higher revenue, the relationship isn’t always straightforward, as some low-budget campaigns achieved high ROI. The scatter plot of budget vs. revenue by subscription tier revealed that the Standard tier had the broadest distribution of data points, indicating a more balanced revenue generation across budget levels. Keywords such as "Affordable" and "Durable" performed strongly across all tiers, with subtle variations in their impact on different customer groups.

The Sankey diagram illustrated the affinity between subscription tiers and keywords, suggesting that the Premium tier values affordability, while the Basic tier prioritizes innovation. Our Random Forest classification model showed strong performance in predicting profitability levels, with high accuracy for Class 0 and Class 2, although Class 1 exhibited more misclassifications.
These findings suggest targeted strategies could focus on optimizing keywords and budgets for different subscription tiers to maximize revenue.

Additionally, we have a **Sunburst Chart of Units Sold by Subscription Tier and Keywords** Fig(10), which provides a detailed view of how keyword performance correlates with units sold across different subscription tiers. This allows us to further refine our understanding of how consumer preferences influence sales within each tier.

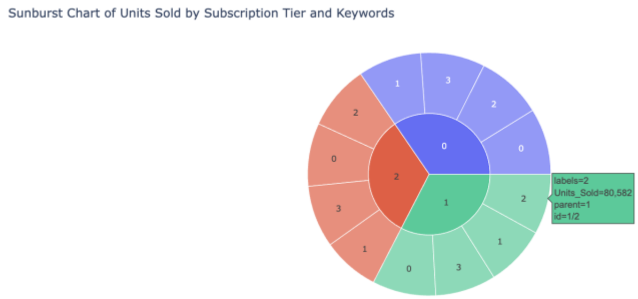

Fig. 10 Sunburst Chart of Units Sold by Subscription Tier and Keywords

In the Data Set Understanding and Preprocessing stage, we created a boxplot to identify and handle outliers. The normal boxplots showed that the scaled data lacked major variance or extreme outliers, which is expected due to normalization. To gain a more granular view, we created a swarmplot in combination with the boxplot, plotting individual data points on it. This detailed visualization Fig(11), helped us identify any potential outliers that might have been overlooked and provided a more comprehensive view of the data distribution, supporting more accurate decision-making for future strategies.

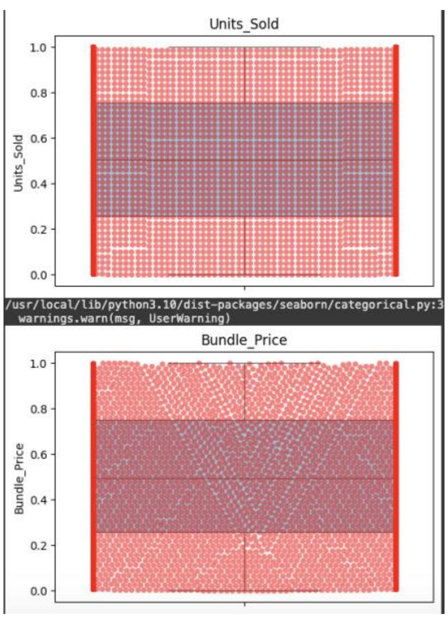

Fig. 11 Boxplot and Swarmplot for Outlier Detection

5 Critical reflection

The analysis of the marketing campaign dataset, while insightful, had certain limitations. The synthetic nature of the data limited the depth of insights, as it lacked the variability seen in real-world data. This impacted the accuracy of patterns identified, especially the relationship between budget and profitability. Real-world data would have introduced more complexity, leading to more robust conclusions.

A key challenge was the uniform distribution of features, which didn't fully capture the natural fluctuations found in real-world marketing campaigns. For instance, "Affordable" emerged as a significant factor across all subscription tiers, but in actual campaigns, seasonality and market trends would likely add more layers of complexity. Real-time data would have accounted for these variations, making the findings more practical and actionable.

Time-Series Analysis:
Marketing campaigns are influenced by temporal factors, which could have been better captured through time-series analysis. This would have highlighted trends and seasonality, providing a clearer understanding of how campaigns evolve over time.

Advanced Techniques:
While PCA and KMeans clustering were useful for dimensionality reduction, employing more advanced methods like hierarchical clustering or deep learning could have offered deeper insights. These techniques would enable a more granular analysis of campaign performance, providing a richer perspective.

Visualizations:
Python proved to be an excellent tool for creating visualizations, with libraries like Matplotlib, Seaborn, and Plotly offering straightforward ways to explore the data. These visualizations, such as scatter plots, treemaps, and heatmaps, were effective in identifying key correlations between features. However, exploring additional visualization tools like Tableau or Power BI could have enhanced the analysis further. These platforms offer interactive dashboards, allowing users to explore data dynamically and offering a more engaging and insightful experience.

In conclusion, the analysis effectively identified broad patterns but was limited by the synthetic nature of the data and the simplicity of the methods used. Future research should focus on leveraging real-world data, advanced techniques, and interactive visualizations to uncover deeper insights and improve prediction accuracy.

Table of word counts

Problem statement	225
State of the art	515
Properties of the data	500
Analysis: Approach	439
Analysis: Process	1580
Analysis: Results	220
Critical reflection	350

References

Bibliography

[1] M. J. Laufenberg, "Visualization Approaches Integrating Real-Time Market Data".

[2] A. Vladova, "Research Gate," 2021. [Online]. Available: https://www.researchgate.net/publication/356449519_Visualizing_Results_of_Promoting_Campaigns.

[3] P. D. Jatin Thakur1, "Improving Data Analysis and Visualization in Market Research with Tableau," IEEE .

[4] S. A. z. B. A. Alaa Abu-Srhan, "Visualization and Analysis in Bank Direct Marketing Prediction," (IJACSA) International Journal of Advanced Computer Science and Applications.

https://ieeexplore.ieee.org/abstract/document/1509166

https://link.springer.com/chapter/10.1007/978-981-15-2282-6_5