AI-Driven Targeted Advertising: Optimizing Campaign Strategy for Marburg Vehicles and Parts

Vipul Singh **Mithun Raj Sekar** **Praneetha Bhupathiraju** **Pranav Chinchore** **Omkar Wadekar**
Philipps-Universität Marburg, Germany
Department of Mathematics and Computer Science
Applications of Artificial Intelligence in Business

1 Abstract

This project utilizes AI-powered analytics to improve Marburg Vehicles and Parts (MVP) advertising by analyzing MVP's sales dataset. Leveraging the Team Data Science Process (TDSP) framework, and analyzing 2,823 records from 2003 to 2005 with Gaussian Mixture Model (GMM) clustering, we aim to uncover meaningful insights. The analysis exposed seasonally recurring sales surges, the most profitable product lines (Classic and Vintage Cars), and the most lucrative markets (USA, Spain, and France). We suggest region-specific messaging tailored to preference and peak season targeted advertising, as well as converting small to medium deal closure strategies. This clearly illustrates the extent to which structured methodologies powered by AI can optimize marketing strategies and improve ROI.

2 Introduction

Marburg Vehicles and Parts (MVP) aims to optimize advertising through data-driven targeting based on 2003-2005 sales records. Following the Team Data Science Process (TDSP) methodology, we analyze product performance, customer behavior, and geographic distribution.

Our approach employs Gaussian Mixture Model (GMM) clustering to identify customer segments despite challenges including missing geographic data (52.7% values) and outdated temporal information. Key findings include:

- **Product Performance:** Classic Cars and Vintage Cars lead sales; Trains underperform
- **Geographic Distribution:** USA, Spain, and France generate highest revenue, with EMEA as top territory
- **Seasonality:** November sales peak indicates optimal promotion timing

This report details how GMM analysis and segmentation techniques can enhance MVP's advertising strategy, aligning with their goals of precision marketing and operational efficiency.

3 Dataset

The MVP sales dataset (Kyanyoga, 2023) comprises 2,823 global sales records spanning January 2003 to May 2005. Each entry captures order-level transactions with key attributes including Order Date, Country, Territory (North America, EMEA, APAC), Product Line, and Sales Amount in USD. The relational structure combines categorical variables with quantitative measures, enabling both time-series and cross-sectional analyses of MVP's international sales performance.

3.1 Geographic Sales Breakdown

Summarizing sales per country shows that the USA and Spain are the biggest markets, followed by France and other prominent European nations. The USA tops the list with sales reaching almost €3.6 million. Spain follows with €1.21 million and France with €1.11 million. With the exception of South Korea, these top markets include all three geo-territories (NA: USA; EMEA: Spain, France, Germany, UK; APAC: Australia). This breakdown shows that North America and Europe are the main spenders.

3.2 Sales Distribution by Territory

EMEA dominates the territorial sales distribution at 49.6% of total sales, followed by NA (38.4%) and APAC (7.4%). EMEA's strong performance stems from high demand in European countries including Germany, UK, and France, while NA benefits from the substantial US market. This distribution identifies EMEA as the region with the highest growth potential, though all territories remain strategically important.

3.3 Product Line Performance

The product portfolio analysis reveals "Classic Cars" as the dominant revenue generator with €3.9 million in sales and the highest average order value at €4.05K. "Vintage Cars" ranks second with €1.9 million in revenue and €3.13K average order value. Lower-performing categories like "Trains" show both smaller revenue totals (€226K) and lower average order values (€2.9K), indicating significant performance disparities across the product range.

3.4 Distribution Trends

Figure 1 reveals distinct patterns in key metrics. Quantity Ordered shows a balanced distribution centered around 35 units, indicating most customers purchase medium-sized orders. Price Each (mean €83.66, median €95.70) and Sales (mean €3,553.89, median €3,184.80) both display right skewness, suggesting frequent lower-priced transactions with occasional high-value outliers. MSRP maintains consistency across products (average €100.72). These patterns indicate MVP's revenue comes primarily from medium and large transactions, with potential opportunity to convert smaller sales to larger ones.

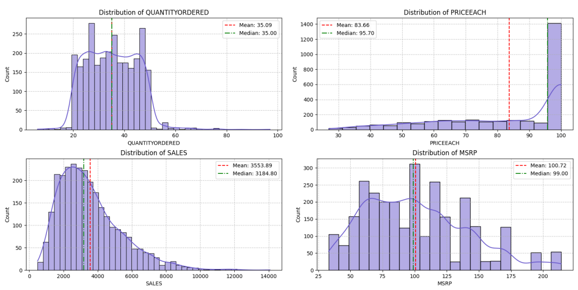

Figure 1: Distribution Trends In Dataset

As shown by these patterns, MVP's sales come mainly from a mix of medium-sized and big transactions. There may be opportunities for increasing revenue by converting smaller sales into bigger ones.

3.5 Total Selling Products by Country

Countries around the world show a persistent demand for Classic Cars, which stand as the most popular product line. France and Spain are quite remarkable due to their significant sales of Vintage Cars, which indicates a latent market opportunity in these countries. The USA continues to dominate in total sales and especially in Classic Cars, reinforcing its role as the major market for MVP's products. Other product lines, such as Motorcycles, Planes, Ships, Trains, Trucks and Buses, and Accessories, also have some contribution to sales in different countries, but not as much as Classic and Vintage Cars.

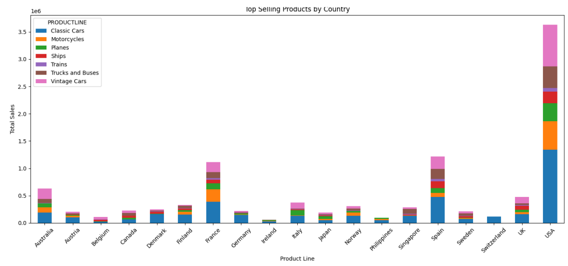

Figure 2: Top Selling Products by Countries

3.6 Fields Missingness

Of the 2,823 sales records, 1,486 (52.6%) lack a `State` value, reflecting gaps in regional data for certain orders. The `AddressLine2` and `PostalCode` fields also exhibit frequent nulls, but since they are not critical to our core analyses, we exclude them from downstream processing.

3.7 Peak Season

Analysis of monthly sales data in Figure 3 uncovers a notable increase towards the end of the year as both November and December stand out as the revenue generating months from 2003 to 2005 along with accounting for almost a third of the annual turnover. There are also secondary peaks in April and July, while January and February are relatively inactive. This seasonality mirrors holiday shopping habits as well as corporate buying at the fiscal year-end. If marketing and inventory plans are synced with these repetitive patterns, ROI can be maximized during peak times and performance during the off-peak months can be raised through targeted incentives.

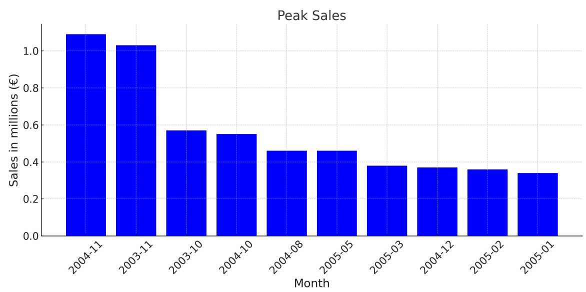

Figure 3: Peak Sales

3.8 Top Customers and RFM Overview

We analyzed customer value using Recency, Frequency, and Monetary metrics for the ten highest-spending accounts as of June 1, 2005, assigning five-point scale scores across these dimensions. Four customers (Euro Shopping Channel, Mini Gifts Distributors Ltd., La Rochelle Gifts, Dragon Souveniers, Ltd.) achieved RFM scores of 14-15, placing them in the "High Value" segment that represents over 65% of the cohort's total spend. Five customers displayed moderate loyalty and spending, categorized as "Mid Value" with significant growth potential through volume incentives or loyalty programs. Only AV Stores, Co. fell into the "Low Value" segment, requiring targeted re-engagement strategies. This segmentation enables MVP to prioritize marketing efforts and optimize resource allocation for maximum customer lifetime value.

4 Business Insights

To help Marburg Vehicles and Parts (MVP) design a smarter, more targeted advertising strategy, we plan to use Gaussian Mixture Models (GMM) to uncover distinct customer segments hidden in the sales data. GMM is particularly useful because it does not just assign customers to one fixed group. Instead, it gives us a probability-based view, allowing a more realistic and flexible understanding of how different customer types behave, where they come from, and what they prefer to buy.

Even before applying GMM, our early exploration of the dataset revealed some clear trends. One of the most striking patterns is the seasonal nature of MVP's sales. Orders consistently spike in November, suggesting that customers are more active towards the end of the year. This is likely driven by the holiday season. This insight points to an important takeaway: MVP's advertising campaigns will likely perform best if they are concentrated in Q4, especially around November, to tap into that natural increase in demand.

When looking at products, Classic Cars and Vintage Cars clearly dominate in both sales volume and revenue. These are not just popular. They represent MVP's premium offerings and are likely to appear prominently in high-value customer clusters. These product lines should be at the forefront of our marketing content. Their collectible and nostalgic appeal aligns well with emotionally driven seasonal advertising.

Geographically, the data tells a consistent story. The United States, Spain, and France are MVP's strongest markets. We expect GMM to highlight distinct clusters around these countries, which means our advertising should be tailored regionally in terms of both timing and content. For example, what appeals to French customers may differ from what works in the United States, and our strategy should reflect that nuance.

Another key insight relates to deal sizes. While most transactions fall into the *Small* category, it is the *Medium*-sized deals that generate the bulk of the revenue. This opens a major opportunity. By using GMM to identify small-deal customers who behave like medium-deal ones, we can target them with special offers or upsell incentives. This could help nudge them into more profitable segments without having to acquire new customers.

We also plan to integrate Recency, Frequency, and Monetary (RFM) scores into our analysis. These scores will help us find MVP's most loyal and high-value customers, such as *Euro Shopping Channel* and *Mini Gifts Distributors Ltd.*, who consistently make valuable purchases. These are the customers worth rewarding with loyalty programs, early product releases, or exclusive promotions to strengthen their connection with the brand.

In short, by applying GMM, we aim to bring more precision and personalization to MVP's advertising strategy. With insights grounded in customer behavior, timing, geography, and product affinity, the campaign can move beyond generic messaging and instead speak directly to the people most likely to respond. This approach will lead to better engagement, stronger loyalty, and a higher return on marketing investment.

5 Methodology: Team Data Science Process (TDSP)

To guide the development of our project in a structured and collaborative way, we decided to follow the Team Data Science Process (TDSP). TDSP, as discussed in our course, is a practical framework designed specifically for projects that rely heavily on working with data. It helped us break the project down into manageable stages, ensuring that we always stayed aligned with our business goals while also keeping our technical work organized.

We began with the **Business Understanding** phase, where we took time as a team to fully understand what MVP (Marburg Vehicles and Parts) needed. The goal was clear: to support their upcoming advertising campaign by helping them understand their sales data better. This involved identifying who their key customers are, which products perform best, and when and where their sales tend to peak. This initial step helped us set clear objectives and ensured that every team member had the same vision.

In the **Data Acquisition and Understanding** phase, we explored the dataset provided to us, which contained over 2,800 sales records from the years 2003 to 2005. We reviewed the structure of the data, noted missing values, and discussed as a team whether the dataset could answer the business questions we had outlined. This phase helped us get comfortable with the data and begin forming ideas about where to focus our analysis.

Next, we moved into **Data Preparation**, where we cleaned and transformed the data to make it suitable for analysis. This included handling missing entries, standardizing formats, and creating additional features like seasonal indicators and customer value metrics. At this point, the dataset was ready for deeper investigation, and we could begin identifying meaningful patterns.

Later in the project, we turned to the final stages of TDSP: **Deployment and Customer Acceptance**. Here, we focused on turning our findings into actionable recommendations for MVP. Rather than stopping at analysis, we developed clear strategies that MVP could use. For example, we identified how to time their ads based on seasonal trends and which customer groups to focus on. Our goal was to deliver insights that would be easy to understand, practical to implement, and valuable from a business perspective.

Overall, TDSP gave us a clear path to follow, helped us work together more effectively, and ensured that the technical work we did stayed grounded in real business needs. It encouraged us to keep checking in as a team, reflect on our progress, and adapt when necessary. All of these elements contributed to a smoother and more focused project journey.

6 AI Technology

Since our data lacks predefined customer categories, we are using clustering to uncover natural behavior patterns, helping MVP target ads effectively. Two of the clustering methods which we can use are K-Means clustering and Gaussian Mixture Model (GMM). K-Means is fast and simple but assumes uniform cluster size and shape, making it less ideal for MVP's diverse customer base. It also struggles with overlapping and noisy data. Hence we choose GMM which handles overlap more effectively.

Gaussian Mixture Models (GMM)

Gaussian Mixture Models (GMM) offer a flexible, probabilistic approach to clustering by modeling data as a combination of Gaussian distributions. Unlike K-Means, GMM does not assume clusters are uniform in size or shape and allows soft assignments, where each point can belong to multiple clusters with varying probabilities. This makes GMM well-suited for capturing overlapping, irregular customer groups and better reflects the complexity of real-world behavior.

Application Plan

We'll use Gaussian Mixture Models (GMM) on the same cleaned dataset and compare the results using log-likelihood and the Bayesian Information Criterion (BIC) to see how well it fits. GMM is especially useful because it helps us understand customers who don't clearly belong to just one group, giving us a way to target ads more intelligently based on the likelihood they belong to different clusters.

Tools/ Libraries

Our implementation leverages **scikit-learn** for GMM clustering with probabilistic assignments, **pandas** and **numpy** for data preparation and manipulation, **seaborn** for cluster distribution visualization, **scipy** for statistical validation metrics, and **plotly** for creating interactive visual explorations of the clustering results.

Use Case

A customer who occasionally buys both car accessories and commercial vehicle parts could be assigned a 70% probability to a "consumer segment" and 30% to a "commercial buyer" segment, this helps MVP tailor dynamic messaging.

References

Kyanyoga. 2023. Sample sales data. Kaggle. Available at https://www.kaggle.com/datasets/kyanyoga/sample-sales-data/.