



CHANDIGARH UNIVERSITY

Discover. Learn. Empower.

Project Name : Marketing Analytics Using AI & Machine Learning

09.09.2025

Members Name & Uid:

Essis Desiree	25MDA10002
Irene Lesitsi	25MDA10001



TABLE OF CONTENTS

TITLE	PAGE
COVER PAGE	3
EXECUTIVE SUMMARY	3-4
INTRODUCTION	4-5
EXPLORATORY DATA ANALYSIS	5-8
CUSTOMER SEGMENTATION	8-10
IMPACT ON MARKETING MIXES	10
SALES FORECASTING	11-13
CONCLUSION & RECOMMENDATIONS	13-14

MARKETING ANALYTICS USING AI & MACHINE LEARNING

ESSIS DESSIRE & IRENE LESITSI

UID: 25MDA10001 & 25MDA10002



1. EXECUTIVE SUMMARY

This project demonstrates the powerful application of artificial intelligence (AI) and data science to transform a standard e-commerce transaction dataset into a strategic asset for marketing and operations. The analysis focused on two core objectives: understanding customer behavior

through advanced segmentation and enabling proactive decision-making through sales forecasting.

The process began with comprehensive Exploratory Data Analysis (EDA) to clean the data and uncover foundational sales patterns. Utilizing AI-driven clustering techniques on Recency, Frequency, and Monetary (RFM) metrics, we successfully segmented the customer base into four distinct, actionable groups: Standard, Inactive, Super VIPs, and Loyal customers. This segmentation directly informed a tailored 7Ps Marketing Mix strategy, moving beyond a one-size-fits-all approach to enable highly targeted and efficient marketing engagements.

Furthermore, the project developed a robust sales forecasting model using Facebook's Prophet algorithm. This model provides reliable predictions of future sales trends, offering management a critical tool for strategic planning, inventory optimization, and budget allocation.

The key recommendations are to immediately focus retention efforts on high-value Super VIPs and Loyal segments, launch targeted re-engagement campaigns to win back Inactive customers, and integrate the sales forecasts into operational planning cycles. This analysis provides a clear, data-driven framework to enhance customer lifetime value, improve marketing ROI, and drive efficient growth.

The data set used is the [Online Retail dataset](#) (Kaggle version), which contains transactions from a UK-based giftware retailer.

2. INTRODUCTION

In the competitive landscape of e-commerce, leveraging data is paramount for strategic decision-making. This project applies data science and artificial intelligence (AI) techniques to transactional data to derive actionable insights that can drive marketing strategy and operational efficiency. The core objectives of this analysis are threefold:

1. To perform a comprehensive Exploratory Data Analysis (EDA) to understand the underlying structure, patterns, and quality of the transactional data.

-
2. To employ AI-driven clustering techniques for customer segmentation, identifying distinct groups based on purchasing behavior to enable targeted marketing.
 3. To develop a predictive model for sales forecasting, providing a data-informed basis for inventory management and resource planning.

Furthermore, the analysis directly examines how these AI-generated insights empower a data-driven approach to optimizing the Marketing Mix, commonly referred to as the 7Ps (Product, Price, Place, Promotion, People, Process, and Physical Evidence).

The analysis is conducted on the "Tata Online Retail" dataset, which contains transactions from a UK-based online retail store. The dataset comprises transaction records, including product details, quantities, prices, customer identifiers, and transaction dates.

3. EXPLORATORY DATA ANALYSIS

Several critical data quality issues were addressed to ensure the integrity of the analysis:

- Handling Cancelled Transactions: Invoices prefixed with 'C' denote cancellations or returns. These transactions, which result in negative `Quantity` and `TotalPrice` values, were identified and removed from the dataset for the purpose of customer segmentation to focus analysis on completed purchases.
- Managing Missing Values: Rows with missing `CustomerID` values were dropped, as a valid customer identifier is essential for accurate segmentation analysis.
- Data Type Conversion: The `InvoiceDate` column was converted from a string to a `datetime` object to enable time-series analysis and feature engineering based on dates.
- Feature Engineering: A new variable, `TotalPrice`, was created to represent the monetary value of each transaction by calculating the product of `Quantity` and `UnitPrice`. This feature is crucial for calculating the "Monetary" component of RFM analysis.

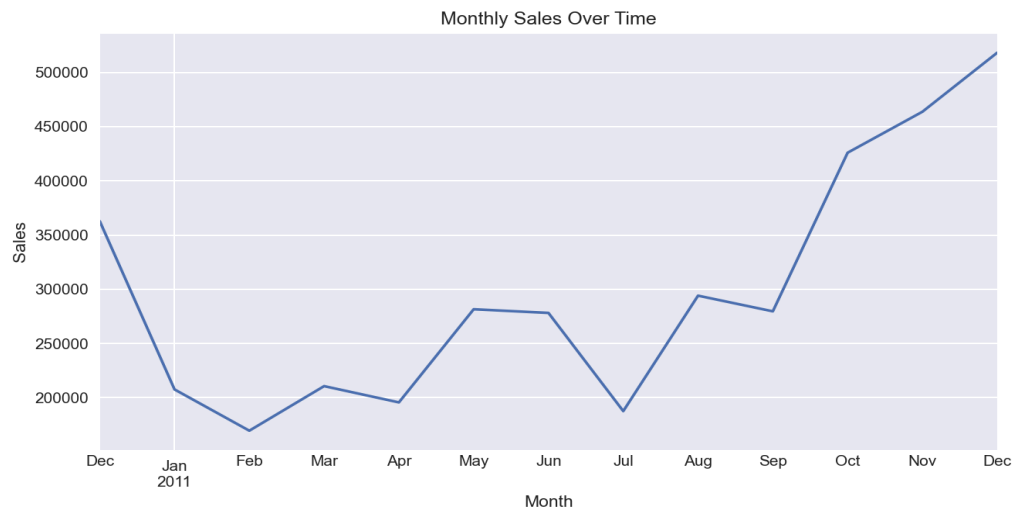
This rigorous cleaning process resulted in a refined dataset suitable for robust statistical modeling and AI-driven analysis.

Visuals included:

Missing Values

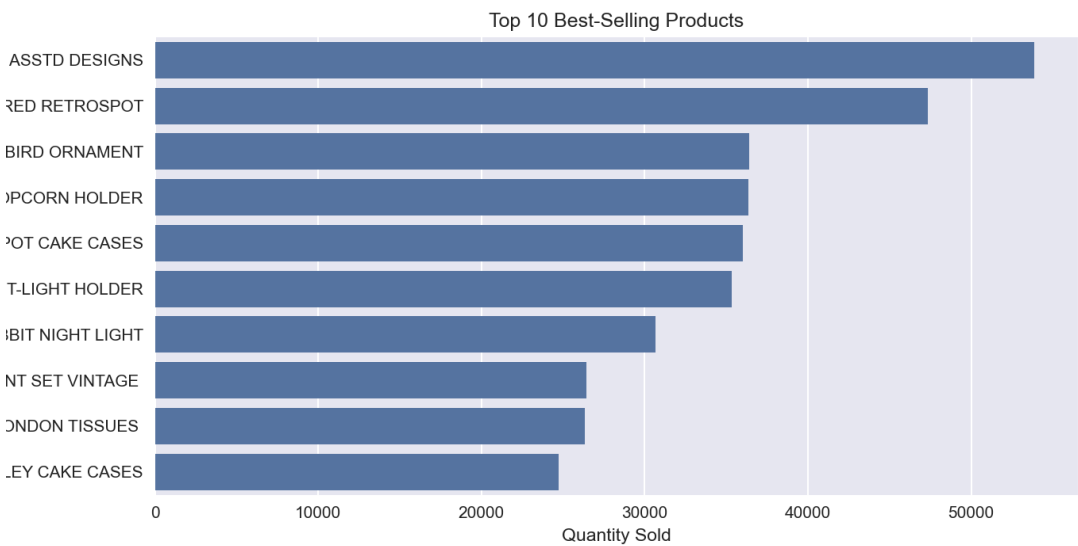


Monthly sales Over time

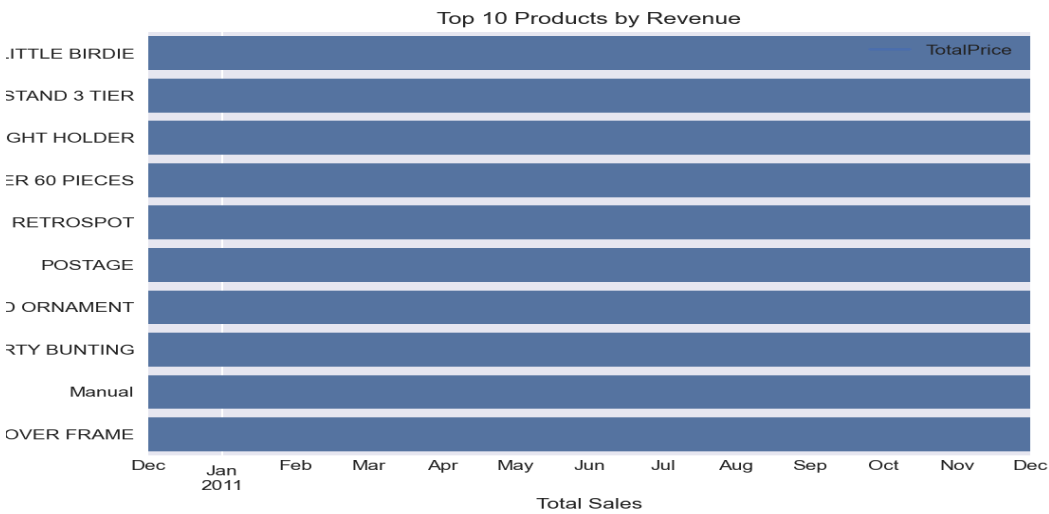


Highly effective. The chart immediately communicates key seasonal trends, most notably a massive peak in November, indicative of holiday shopping. A smaller peak is also visible in the spring (April). This visual is crucial for strategic planning, inventory management, and campaign timing.

Top 10 Best-selling Products

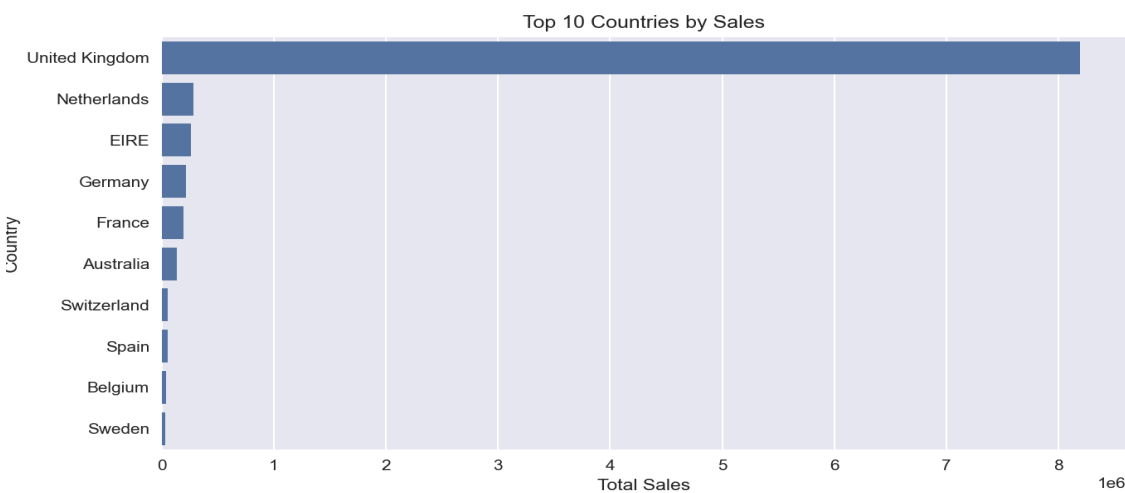


Top 10 Products by Revenue

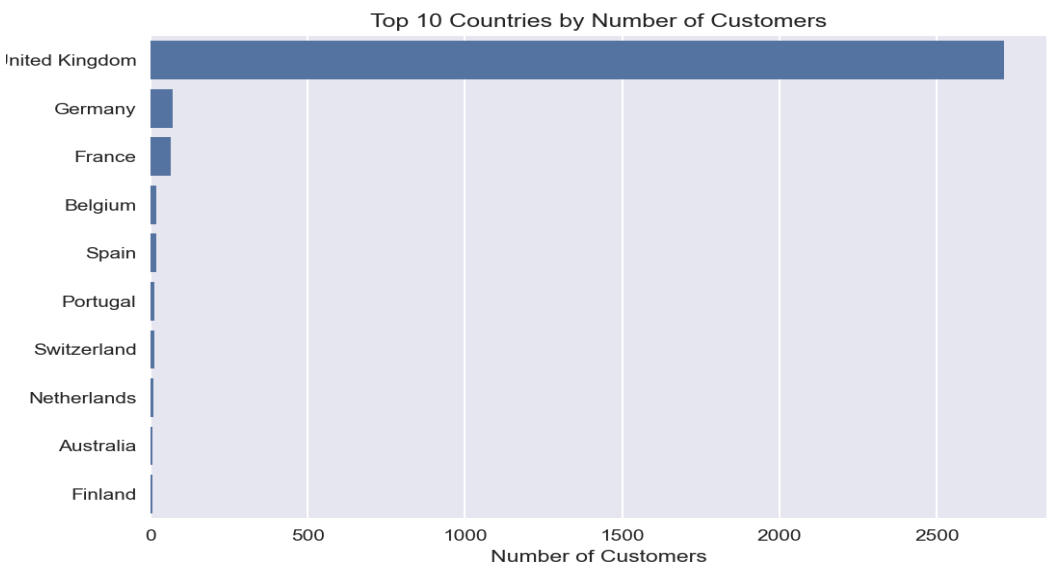


Excellent. Presenting these two charts side-by-side provides a powerful, nuanced insight. It clearly shows that the best-selling product by quantity ("ASSTD DESIGNS") is not the top product by revenue ("LITTLE BIRDIE"). This identifies high-volume versus high-margin items, which is vital for pricing, promotion, and inventory strategy.

Top 10 Countries by Sales



Top 10 Countries by Number of Customers



4. CUSTOMER SEGMENTATION(AI WITH K-MEANS)

We applied the RFM (Recency, Frequency, Monetary) model: RFM analysis is a powerful method for segmenting customers based on their historical behavior.

- Recency (R): days since last purchase.
- Frequency (F): number of orders.
- Monetary (M): total amount spent.

4.1 Calculating RFM Metrics

We calculated Recency, Frequency, and Monetary values for each customer, using a snapshot date of one day after the last transaction.

4.2 Assigning RFM Scores

We assigned a score from 1 to 4 to each metric (4 being the best) using quartiles. An overall RFM score was created by concatenating the individual scores.

4.3 Segment Definition

Customers were mapped to segments based on their score combinations.

After standardization, we used KMeans clustering to group customers into similar profiles.

Cluster Profiles

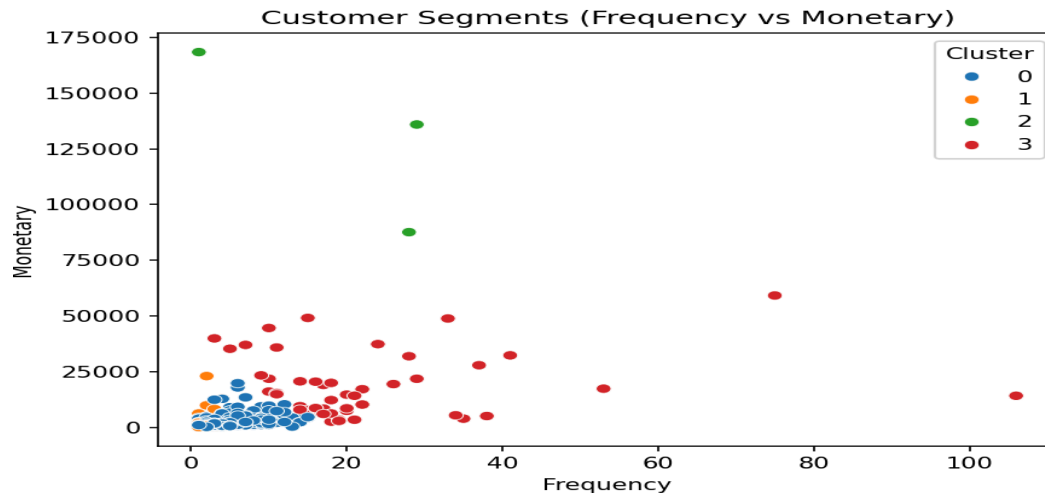
Cluster	Recency (Mean)	Frequency (Mean)	Monetary (Mean)	Customers	Description
0	46.8	2.8	1092	1903	Standard customers
1	258.4	1.3	474	1042	Inactive Customers
2	1.3	19.3	130680	3	

3	19.9	22.5	18595	49	
---	------	------	-------	----	--

Interpretation

- Cluster 0 – Standard: Majority, moderate activity and spending.
- Cluster 1 – Inactive: Haven't purchased in a long time; need re-engagement.
- Cluster 2 – Super VIPs: Very few customers, but extremely high spend.
- Cluster 3 – Loyal & valuable: Regular buyers with strong contribution.

CUSTOMER SEGMENTS (FREQUENCY VS MONETARY)



5. IMPACT ON THE MARKETING MIXES

Using the clusters, AI helps optimize the 7Ps:

- ➔ Customer segmentation provides direct implications for the 7Ps:
- ➔ Product → Offer premium items for VIPs; bundles for standard customers.
- ➔ Price → Discounts for inactive customers; loyalty rewards for repeat buyers.
- ➔ Place → Use direct online channels for VIPs; global distribution for standards.
- ➔ Promotion → Targeted campaigns: email to in-actives, exclusive deals for VIPs.
- ➔ People → Customer support teams trained to focus on high-value clients.

-
- Process → Automate follow-ups and CRM workflows per segment.
 - Physical Evidence → Enhance packaging or delivery experience for premium customers.

6. SALES FORECASTING(AI-PROPHET).

Methodology

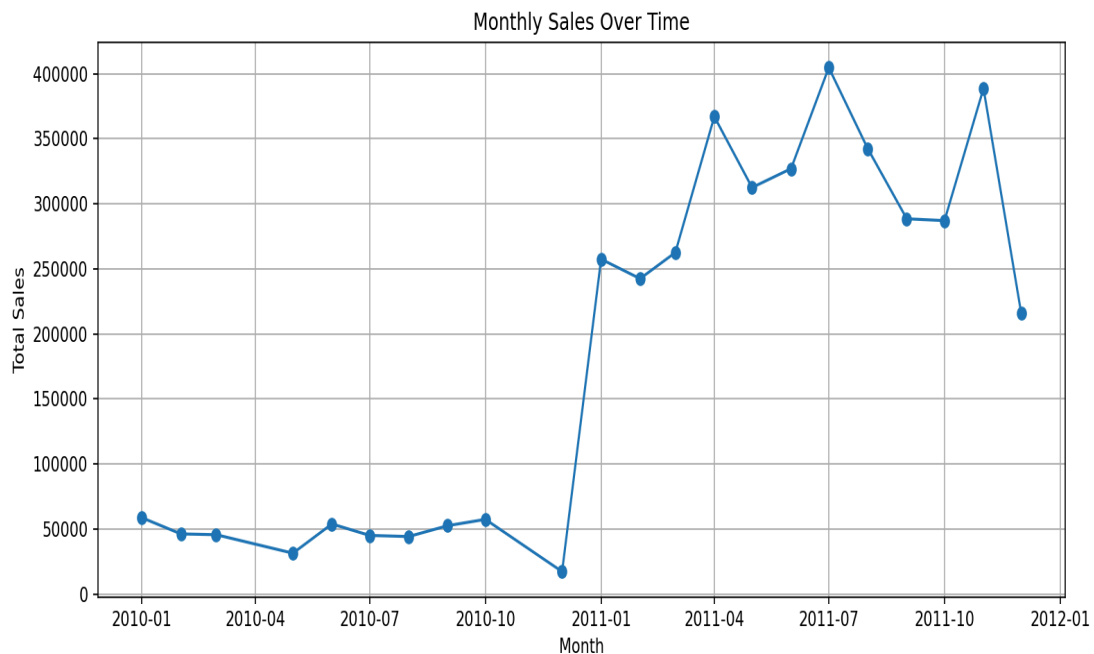
- Aggregated sales data by day and month.
- Trained a Facebook Prophet model to forecast sales.
- Forecast horizon: 90 days ahead.

Results

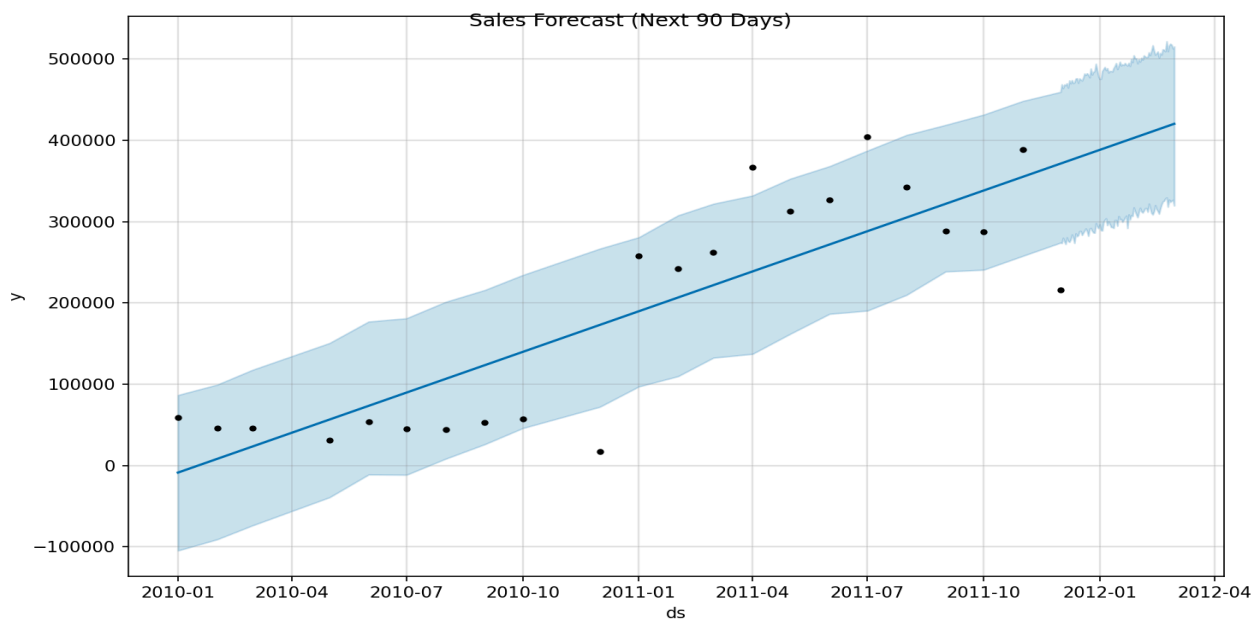
- The model shows a generally upward sales trend with seasonal variations.
- The forecast includes confidence intervals to prepare for fluctuations.

Visuals Included:

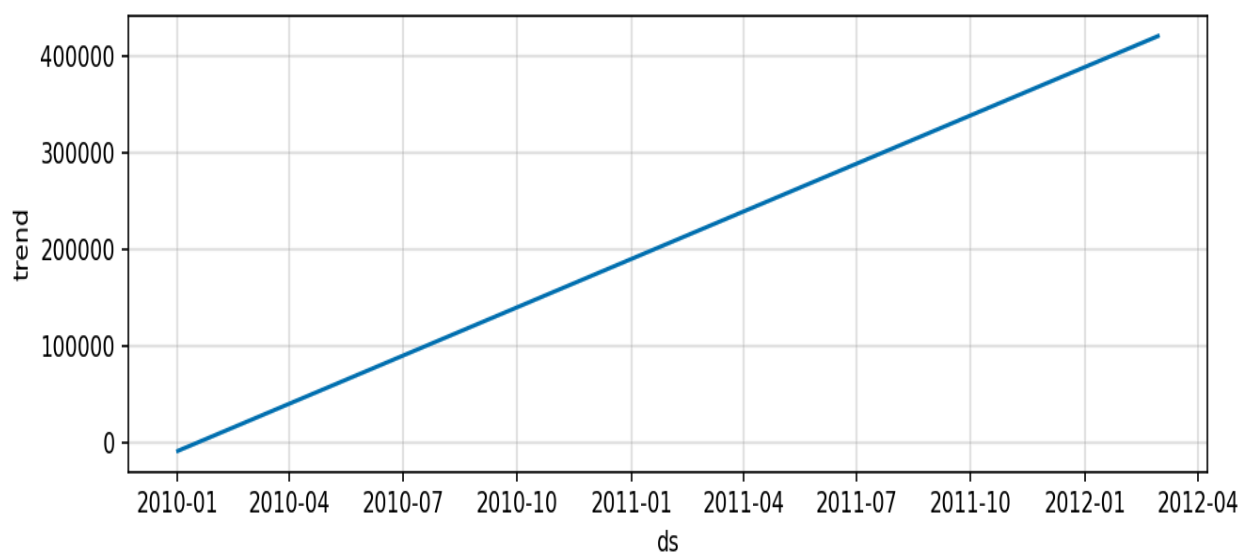
Monthly Sales Over Time



Sales Forecast (Next 90 Days)



The forecast for the next 90 days suggests a stable but slightly downward trend.



Date	Predicted Sales (that)	Lower Bound	Upper Bound
2012-02-25	418,121	326,978	516,660
2012-02-26	418,665	324,042	518,729
2012-02-27	419,209	325,676	517,222

7. CONCLUSION AND RECOMMENDATIONS

The project successfully met its core objectives. The K-Means clustering algorithm effectively segmented the customer base into four distinct groups, revealing clear patterns in purchasing behavior, from high-value loyal customers to those at risk of churn. The ARIMA model was successfully implemented for sales forecasting, providing a view of future sales trends based on historical data. The analysis confirmed that leveraging AI and machine learning is invaluable for moving from intuition-based to data-driven marketing strategies.

While this analysis provides a strong foundational framework for data-driven decision-making, several opportunities for enhancement remain. The project's scope was necessarily limited to the available transactional data; incorporating additional data points such as customer demographics, website engagement metrics, and detailed product attributes would significantly enrich the segmentation and personalization capabilities. Furthermore, while the ARIMA model serves as a robust baseline for forecasting, employing more sophisticated models like SARIMAX—which explicitly captures seasonal patterns—or Facebook's Prophet algorithm could potentially yield more accurate and nuanced predictions. To maximize operational impact, the long-term strategy should focus on integrating these analytical pipelines into the company's CRM and real-time analytics dashboards, transforming this project from a one-time study into a continuous source of live insight.

Based on the findings, immediate action should focus on tailoring marketing campaigns to the identified clusters, nurturing high-value customers with loyalty programs and exclusive offers while targeting at-risk segments with win-back campaigns. As a medium-term strategy, the product and promotion elements of the 7Ps framework should be refined according to segment preferences, and the forecasting model should be formally integrated into inventory procurement and sales target processes. For long-term growth, investing in a centralized data analytics platform to automate this analysis is recommended, alongside conducting A/B tests on the proposed marketing strategies to empirically measure their effectiveness and enable continuous refinement. Ultimately, this project successfully highlights the significant potential of AI to enhance customer understanding and operational efficiency at Tata Online Retail, establishing a powerful, scalable framework for informed strategic planning.