

E-commerce Customer Segmentation & Sales Analytics Report

End-to-End Data Analytics using Python, RFM Analysis & Power BI

Kirthan Kumar

Tools used

Python | Pandas | Power BI | DAX | RFM Analysis

Purpose

Portfolio & Learning Project

EXECUTIVE SUMMARY

Business Objective:

The purpose of this project is to understand how an e-commerce business is performing by closely analysing its transaction-level data.

Instead of looking only at overall revenue, the focus is on how customers behave, what they buy, and when they buy, so that meaningful business decisions can be made.

Through this analysis, the project aims to identify high-value customers, highlight customers at risk of churn, uncover product and time-based sales patterns, and provide clear, data-driven recommendations that can help improve revenue growth and customer retention.

The approach and analysis closely reflect how business intelligence and analytics teams in consulting and corporate environments work with real business data.

Key Business Questions Addressed:

- How is the business performing overall, and is revenue growing over time?
- Which customers contribute the most to total revenue?
- Are there customer groups that show signs of reduced engagement or churn risk?
- Which products and product categories generate the highest sales?
- At what times (days and hours) do customers tend to purchase the most?
- Based on historical trends, what does near-term revenue growth look like?

Data Summary:

The analysis is based on cleaned and validated transactional data with the following scope:

- Total Transactions (after cleaning): 354,321
- Total Unique Customers: 3,920
- Time Period Covered: December 2010 to December 2011
- Geographic Focus: United Kingdom
- Total Revenue Analyzed: Approximately \$7.3 million

Only high-quality, valid transactions were used to ensure accurate insights and reliable business conclusions.

Key Insights & Findings:

Several important patterns emerged from the analysis:

- Revenue is not evenly distributed across customers; a relatively small group of customer's accounts for a large share of total sales.
- The Champions segment (1,025 customers) represents the most valuable group, contributing around 65% of overall revenue.
- A noticeable portion of customers fall into At Risk and Hibernating segments, suggesting clear opportunities for targeted retention efforts.
- Sales activity tends to peak mid-week, especially on Thursdays, and during late morning to early afternoon hours, which are ideal windows for promotions and campaigns.
- Certain product categories consistently outperform others and frequently appear in multi-item orders, indicating strong cross-sell and bundling potential.
- Revenue shows a steady upward trend over time, with recent months reflecting positive month-over-month growth.

Forecast & Growth Outlook:

Using historical sales trends:

- Latest Month Revenue: ~\$980.65K
- Average Month-over-Month Growth: ~4%
- Forecasted Revenue for the Next Month: ~\$1M

If current customer behavior and demand patterns continue, the business is expected to maintain a positive short-term growth trajectory.

Business Impact & Value:

The insights generated from this analysis can help business stakeholders to:

- Focus retention and loyalty efforts on high-value customers
- Proactively re-engage customers showing early signs of churn
- Make informed decisions around product assortment and inventory planning
- Identify the best timing for promotions and campaigns
- Support strategic planning through revenue forecasting and trend analysis

Summary Statement

Overall, this project demonstrates how combining customer segmentation, sales performance analysis, and forecasting can translate raw transaction data into actionable insights that support smarter business decisions and sustainable growth.

DATA OVERVIEW & PREPARATION

Dataset Overview:

This analysis is based on a publicly available e-commerce transactional dataset that captures customer purchase behavior over a one-year period. After performing data quality checks and cleaning, only valid and meaningful transactions were retained for analysis.

The final dataset used for analysis includes:

- **Total Records (after cleaning):** 354,321 transactions
- **Time Period Covered:** December 2010 to December 2011
- **Geographic Scope:** United Kingdom
- **Level of Detail:** Individual transaction and customer-level data

Focusing on a single geography helped ensure consistency in customer behavior patterns and avoided noise caused by regional differences.

Data Cleaning & Preparation Approach:

Before performing any analysis, the dataset was carefully cleaned to ensure that the insights generated were reliable and business-ready. Each cleaning step was driven by a specific analytical need rather than technical convenience.

Key preparation steps included:

- **Removing transactions with missing Customer IDs**
These records could not be linked to individual customers and therefore were not suitable for customer-level analysis or segmentation.
- **Excluding negative quantities and cancelled orders**
Such transactions typically represent returns or order cancellations and can distort revenue calculations if included.
- **Creating derived and analytical fields** to support deeper insights, including:
 - **TotalPrice** ($\text{Quantity} \times \text{UnitPrice}$) for accurate revenue calculation
 - **Day of Week** and **Hour of Purchase** to analyze time-based buying behavior
 - **Month** and **Year-Month** to support trend analysis and forecasting

These steps ensured that the dataset was both **clean and analytically enriched**, forming a strong foundation for segmentation, visualization, and forecasting.

CUSTOMER SEGMENTATION (RFM ANALYSIS)

Why RFM Analysis?

Not all customers contribute equally to a business. Some purchase frequently and spend more, while others may have purchased only once or stopped engaging altogether.

To understand these differences, **RFM (Recency, Frequency, Monetary) analysis** was used. This approach is widely adopted in retail and consulting environments because it provides a simple yet powerful way to measure customer value based on actual purchasing behavior.

RFM helps answer questions such as:

- Who are our most valuable customers?
- Which customers are becoming inactive?
- Where should retention efforts be focused?

RFM Metrics Explained:

Each customer was evaluated using three key dimensions:

- **Recency** – How recently the customer made a purchase
- **Frequency** – How often the customer placed orders
- **Monetary** – How much the customer spent in total

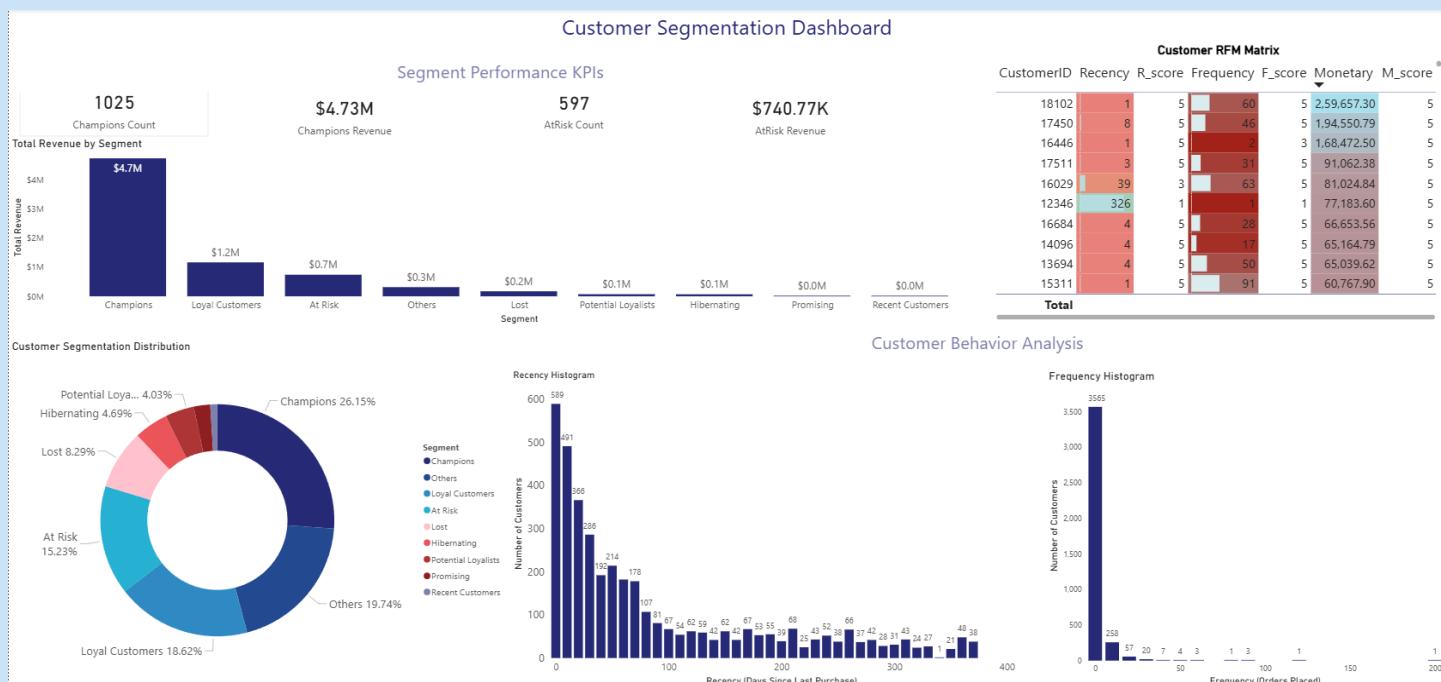
Customers were scored on a **1 to 5 scale** for each metric, where higher scores indicate stronger engagement and higher value. These scores were then combined to create meaningful customer segments.

Customer Segments Created:

Based on RFM scores, customers were grouped into intuitive and business-friendly segments, including:

- **Champions** – Highly recent, frequent, and high-spending customers
- **Loyal Customers** – Regular buyers with consistent spending patterns
- **Potential Loyalists** – Customers showing early signs of loyalty
- **At Risk** – Previously valuable customers with declining recent activity
- **Lost Customers** – Customers with long inactivity and low engagement
- **Others** – Remaining customers with mixed or low engagement

This segmentation enables the business to move beyond generic customer views and adopt **targeted, data-driven strategies**.



SEGMENT INSIGHTS & BUSINESS IMPACT

Segment Performance Insights:

Analyzing customer segments revealed clear differences in how customers contribute to revenue:

- **Champions**, although representing only about **26% of customers**, contribute **approximately 65% of total revenue**, making them the most valuable group.
- **At Risk customers** show strong historical value but reduced recent activity, indicating a high potential for churn if not addressed.
- **Lost customers** have minimal recent engagement and low purchase frequency, contributing very little to current revenue.

This highlights that **revenue is heavily concentrated within a small portion of the customer base**, making targeted retention strategies critical.

Business Recommendations:

Based on these insights, the following actions are recommended:

- **Protect and retain Champions** through loyalty programs, exclusive offers, or early access to promotions.
- **Re-engage At Risk customers** with personalized campaigns, reminders, or limited-time incentives.
- **Nurture Potential Loyalists** by encouraging repeat purchases through discounts or bundled offers.
- **Reduce marketing spend on Lost customers**, or target them selectively with low-cost reactivation campaigns.

These actions allow the business to **allocate marketing and retention budgets more efficiently**, focusing on customers who have the highest impact on revenue.

FORECASTING & TRENDS

Revenue Trend Analysis

An analysis of revenue over time shows a **clear and consistent upward trend** across the year. While short-term fluctuations are present as expected in transactional data the overall direction indicates steady business growth. In the later months of the year, noticeable **seasonal spikes** are observed. These peaks align with typical retail behavior, where customer spending increases due to festive periods and higher promotional activity. This seasonality highlights the importance of proactive planning around inventory, marketing, and operations during high-demand periods. Overall, the trend analysis suggests that the business is not only growing, but doing so in a **predictable and measurable way**, which is essential for strategic planning.

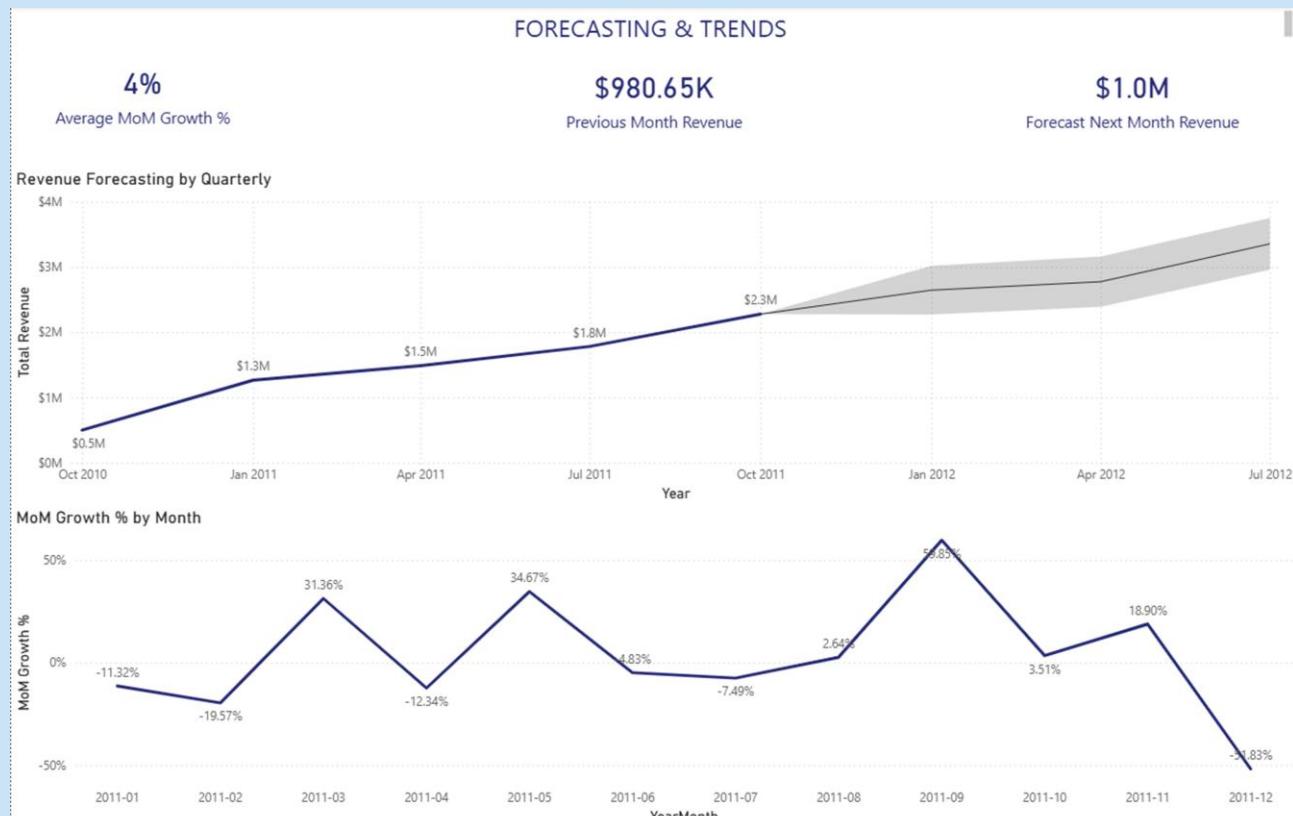
Forecasting Overview

To understand what the near future might look like, a **time-series forecasting approach** was applied using historical revenue data. The goal of this forecast is not to predict exact numbers, but to provide a **data-backed estimate of future performance** based on past trends.

The forecast is supported by the following key indicators:

- **Latest Month Revenue:** Approximately **\$980.65K**, reflecting strong recent performance
- **Average Month-over-Month Growth:** Around **4%**, indicating consistent revenue expansion
- **Forecasted Next Month Revenue:** Approximately **\$1.0M**, assuming similar demand and customer behavior

The forecasting results suggest that if current trends continue, the business is likely to maintain its **positive growth momentum** in the short term.



CONCLUSION & NEXT STEPS

Key Takeaways

This project demonstrates how transaction-level data can be transformed into meaningful business insights when analyzed thoughtfully. The key takeaways from the analysis include:

- Customer segmentation reveals that **revenue is highly concentrated**, with a relatively small group of customers driving a majority of sales.
- Using data-driven targeting strategies can significantly improve **customer retention and lifetime value**.
- Product and time-based sales analysis provide valuable guidance for **inventory planning, promotions, and operational decision-making**.
- Trend analysis and forecasting offer a **forward-looking perspective**, supporting better short-term planning and performance monitoring.

Together, these insights highlight the importance of combining **descriptive, diagnostic, and predictive analytics** to support well-rounded business decisions.

Future Enhancements

While this analysis provides strong foundational insights, several enhancements could further increase its business value:

- **Customer Lifetime Value (CLV) modelling** to better understand long-term customer profitability
- **Marketing campaign effectiveness analysis** to measure ROI and optimize promotional strategies
- **Region-wise or market-level expansion analysis** to support geographic growth decisions
- **Advanced forecasting techniques**, such as ARIMA or machine learning-based models, for longer-term predictions

These enhancements would allow the business to move from **descriptive analytics toward more advanced predictive and prescriptive analytics**.

Final Note

This project showcases an end-to-end analytics workflow, from data preparation and segmentation to visualization and forecasting, reflecting real-world analytics practices used in consulting and business intelligence teams.