

# Business Problem

Effective customer segmentation is essential for boosting retention, reducing churn, and maximizing customer lifetime value. Organisations often fail to leverage transactional data effectively, leading to generic marketing campaigns that do not engage different customer groups. This results in missed revenue opportunities and higher churn risk.

This analysis aims to:

- Identify high-value customers who contribute most to revenue and profit.
- Detect at-risk customer segments for targeted re-engagement.
- Understand purchasing patterns to predict future behaviour.
- Optimize marketing strategies to deliver personalized offers.
- Improve campaign ROI through data-driven segmentation.

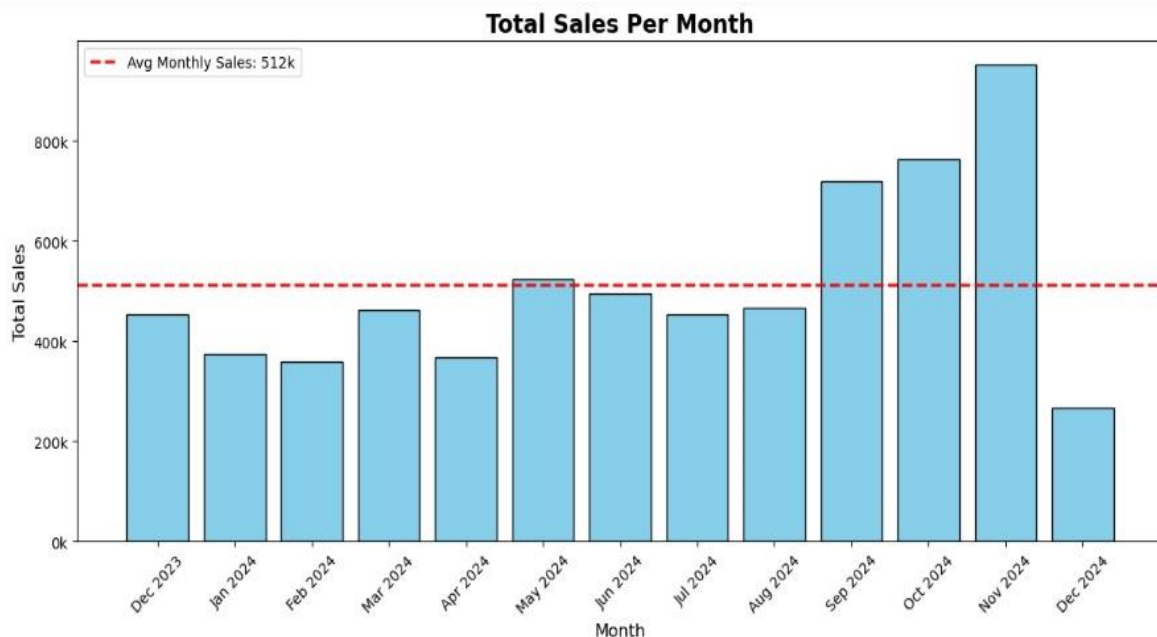
## Data Cleaning & Preparation

To enhance the reliability of the customer segmentation insights, we cleaned the dataset by:

- Removed missing **CustomerID** rows ( $\approx 24.93\%$ ) *(to ensure all transactions are linked to identifiable customers)*.
- Dropped duplicate rows (5,526) *(to avoid double-counting in analysis)*.
- Excluded cancelled orders (InvoiceNo starting with 'C';  $\approx 8,872$  rows) *(to focus only on valid sales)*.
- Filtered non-product **StockCode** entries – 'POST', 'M', 'PADS', and 'DOT' (to remove shipping charges, adjustments, or other non-merchandise items).

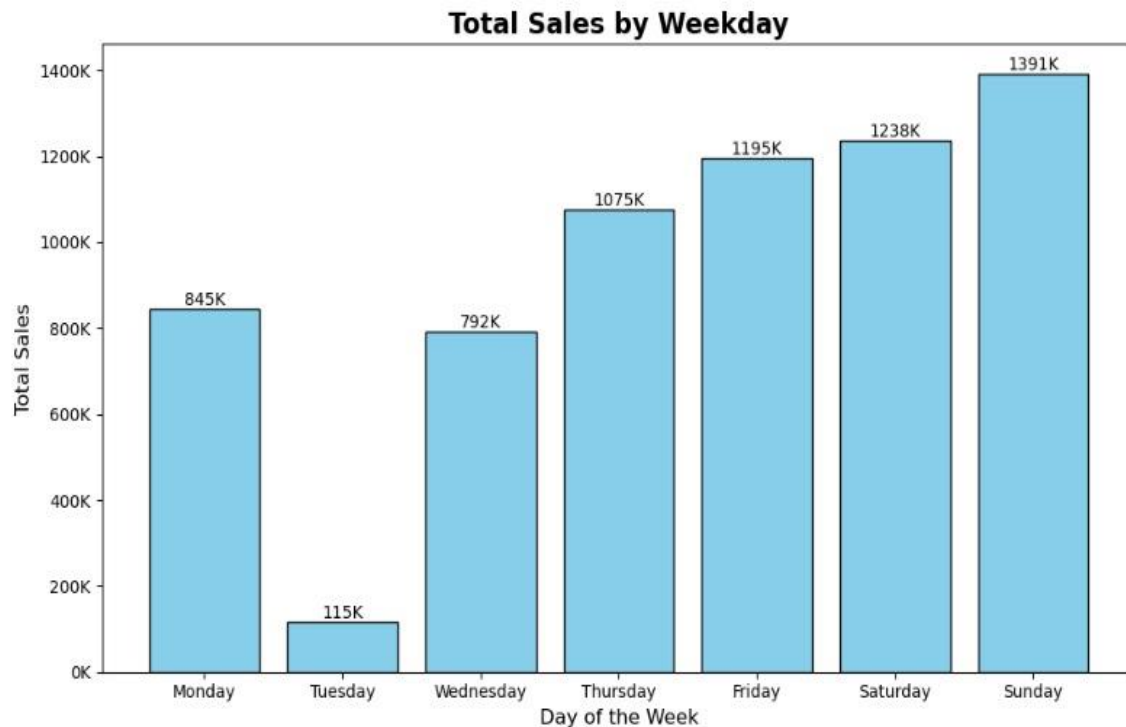
# Exploratory Data Analysis

## Total Sales per Month



- **Average monthly sales** stand at **512K**, serving as a benchmark for performance.
- Sales remained **below average** from Dec 2023 to Aug 2024, with the lowest in **Feb 2024** and **Apr 2024** (~355K–370K).
- A strong **upward trend** began in **Sep 2024**, with consistent growth peaking in **Nov 2024** (~950K), almost double the average.
- **Dec 2024** shows a sharp drop (~260K), possibly due to incomplete data or seasonal slowdown.
- The significant Q4 surge (Sep–Nov) suggests **seasonal demand** or **successful marketing campaigns**, which could be replicated in future periods.

## Total Sales by Weekday



- **Sunday** records the highest sales (~1391K), followed closely by **Saturday** (~1238K) and **Friday** (~1195K), indicating strong weekend demand.
- **Tuesday** has the lowest sales (~115K), significantly below all other days, representing a potential opportunity for targeted mid-week promotions.
- Sales gradually increase from **Wednesday** (~792K) through the weekend, suggesting a buildup in purchasing activity later in the week.
- **Monday** (~845K) and **Thursday** (~1075K) perform moderately, above mid-week levels but below weekend peaks.
- Concentrating marketing and promotions toward **mid-week low performers (Tuesday, Wednesday)** could help balance sales distribution and improve overall weekly revenue.

## Pareto Principle

The Pareto Principle, also known as the 80/20 rule, states that roughly **80% of outcomes are driven by 20% of causes**. In business, this often means a small percentage of customers or products generate the majority of revenue or profit.



In our RFM analysis:

- **31.51% of customers** generate **80% of total sales**.
- **23.75% of products** generate **80% of total sales**.

These insights highlight the need to focus retention, marketing, and inventory strategies on high-value customers and top-performing products to maximise revenue and operational efficiency.

## RFM Analysis & Customer Segmentation

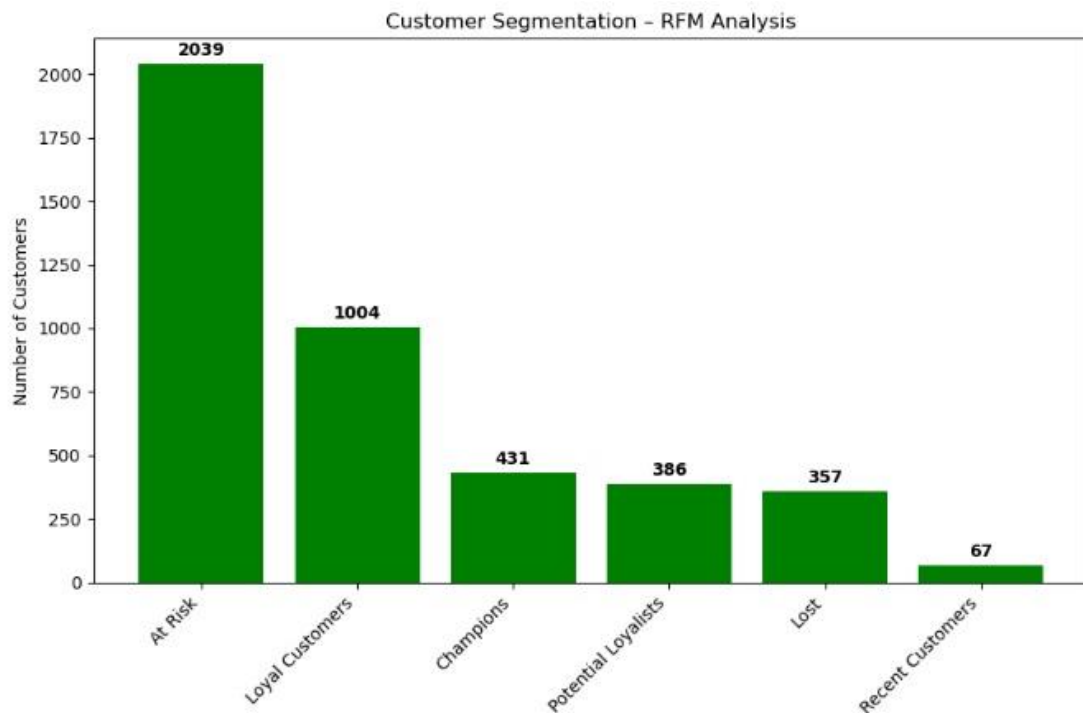
RFM (Recency, Frequency, Monetary) analysis is a data-driven technique used to evaluate customer value based on:

- **Recency (R)** – how recently a customer made a purchase.
- **Frequency (F)** – how often a customer makes purchases.
- **Monetary (M)** – how much a customer spends in total.

Customers are segmented into **five equal buckets** for each RFM metric – **Recency, Frequency, and Monetary**. Each customer is ranked within each metric, assigned a score from **1 (lowest)** to **5 (highest)**. These three scores are then summed to calculate an overall **RFM score**, which is used to classify customers into actionable segments

Based on this RFM Score Analysis, we have six customer segments:

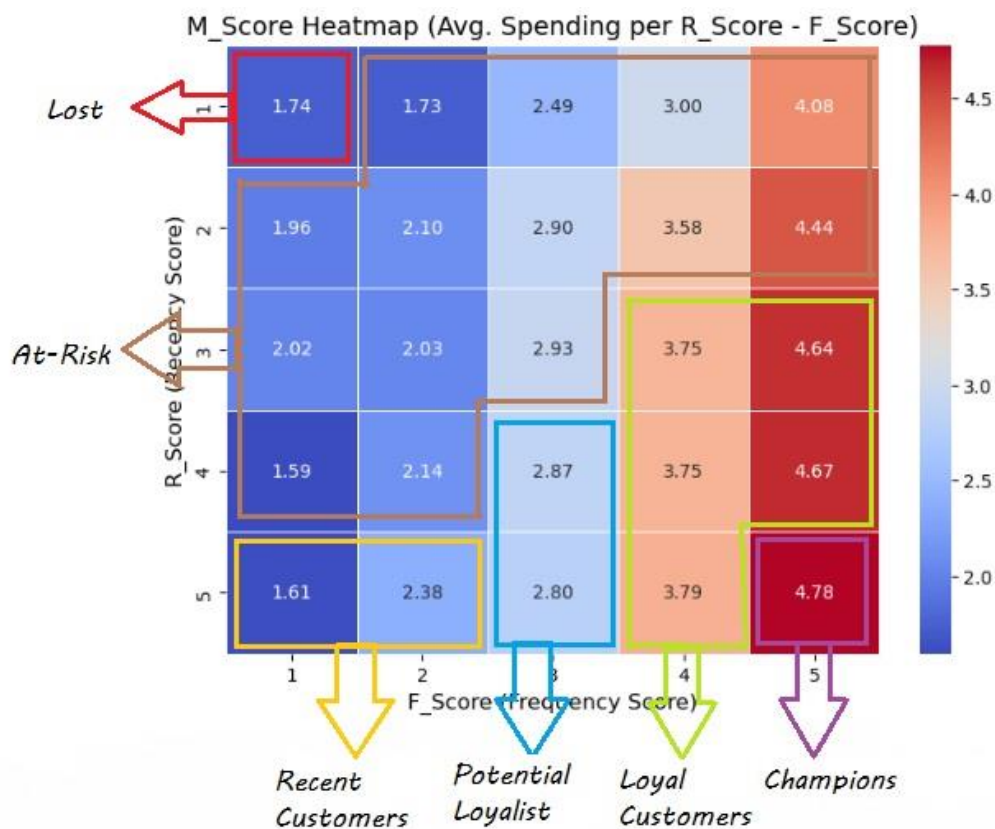
1. Champions
2. Loyal Customers
3. Potential Loyalist
4. Recent Customers
5. At-Risk Customers
6. Lost



- **At-Risk customers** (2,039; **47% of base**) represent the biggest revenue leakage risk. Immediate, targeted win-back efforts could protect a significant portion of recurring sales.
- **Loyal Customers** (1,004) and **Champions** (431) together form the **most profitable 33%** of our customer base. These groups should be prioritised for retention programs, loyalty perks, and upsell opportunities.
- **Potential Loyalists** (386) are in the conversion "sweet spot." Strategic nurturing could graduate them into Loyal Customers or Champions within the next purchase cycle.

- **Lost Customers** (357) are disengaged, but selective reactivation campaigns — especially during seasonal peaks — could recover some value without heavy marketing spend.
- **Recent Customers** (67) are at a critical first-purchase stage. Onboarding success here will determine whether they move into growth segments or churn early.

## Heat Map Based on RFM Analysis



# Recommendations

## 1. Champions

- **Characteristics:** Top-tier customers with exceptional engagement and spending.
- **Behaviour:** recent purchases, very frequent, highest spenders.
- **Strategy:** Major revenue contributors and brand advocates. Retain through VIP programs, early product access, and exclusive offers. Encourage referrals to leverage their advocacy.

## 2. Loyal Customers

- **Characteristics:** Consistent repeat buyers with high brand trust.
- **Behaviour:** Purchase regularly, moderate-to-high spend.
- **Strategy:** Reliable revenue stream with potential for higher AOV. Use loyalty rewards, personalised recommendations, and bundle promotions to deepen engagement.

## 3. Potential Loyalists

- **Characteristics:** Recent or early repeat buyers showing promise.
- **Behaviour:** Recent activity, moderate spend and frequency.
- **Insights & Strategy:** Growth pipeline for future Loyal Customers or Champions. Nurture with onboarding campaigns, targeted discounts, and recommendations to boost repeat purchases.

#### 4. At-Risk Customers

- **Characteristics:** Previously engaged customers now showing declining activity.
- **Behaviour:** Strong historical engagement, overdue for purchase.
- **Strategy:** At high risk of churn but easier to win back than acquire new. Deploy win-back campaigns, highlight new arrivals, and offer limited-time deals.

#### 5. Lost Customers

- **Characteristics:** Long inactive and disengaged customers.
- **Behaviour:** No recent activity, low historical contribution.
- **Strategy:** Low reactivation probability but still worth targeted attempts. Use strong reactivation incentives, seasonal promotions, and alternative product suggestions.

#### 6. Recent Customers

- **Characteristics:** Newly acquired customers in the early stage of their journey.
- **Behaviour:** Very recent purchase, low frequency and spend so far.
- **Strategy:** Onboarding experience is critical for retention. Follow up post-purchase, provide usage tips, and offer upsell opportunities to encourage repeat buying.