

RFM-Based Churn Risk Analysis & Intervention Prioritization

INTRODUCTION

Project Background

1. **Business Problem:** We faced a challenge in identifying **B2B Retailers at risk of churn** since explicit "churn data" (like account closures) was unavailable. Our sales data was recent, showing most retailers were still active (Recency<30 days). This made traditional churn analysis ineffective.
2. **Analytical Solution:** This project was initiated to build an "**Intervention Priority Index**" using a modified RFM model. This model aims to measure **prospective risk** based on behavioural scores and provide actionable insights.
3. **Specific Project Goals:**
 - **Business Goal:** To deliver a clear **priority action list (Intervention Priority Index)** to Sales and Account Managers, ensuring resources are focused on High-Risk Retailers (e.g., Amazon, Kohl's, Walmart).
 - **Analytical Goal:** To demonstrate proficiency in: 1) Building a **Custom Risk Model** that combines business logic (Recency days) and statistical thresholds (Quartiles for F & M). 2) Applying **Weighting and Root Cause Analysis** for action-driven insights.

METHODOLOGY

We used a modified RFM model to transform transaction data into standardized, weighted risk scores.

1. **RFM (Recency, Frequency, Monetary):** Used as a proxy for unavailable explicit churn data. RFM identifies changes in retailer behaviour based on when they last purchased (Recency), how often they purchase (Frequency), and how much they spend (Monetary).
2. **Quartiles (Q2 & Q3):** Used as statistical thresholds to calculate F-Score and M-Score. This ensures performance is benchmarked relative to the retailer's own customer base (e.g., retailers above Quartile 3 receive the best performance score of 3).
3. **Weighted Scoring (2:1:1):** Applied to calculate the **Total Risk Score**. **Recency is weighted 2×** because it is the most representative and time-sensitive metric for short-term churn risk.
4. **Root Cause Analysis (Risk Composition):** This measures the percentage contribution of each metric (R, F, M) to the **Total Risk Score**. This is the "**Why**" behind the risk percentage (e.g., is the risk driven by low Frequency or low Monetary value?).

ANALYSIS PROCESS

1. & 2. Data Preparation

- Datasets were sourced and cleaned. ([source](#))
- Raw data was copied to a new worksheet for data integrity.

3. Calculating Recency (R-Score)

| 1. Last Purchase (Recency) | | | |
|----------------------------|---------------------|---------|-------|
| Row Labels | Max of Invoice Date | Recency | Score |
| Amazon Foot Locker | 31/12/2021 | 0 | 0 |
| Kohl's Sports Direct | 25/12/2021 | 6 | 0 |
| Walmart | 16/12/2021 | 15 | 0 |
| West Gear | 23/12/2021 | 8 | 0 |

- **Latest Purchase:** A Pivot Table was created to find the MAX Invoice Date per retailer.
- **Recency Days:** Calculated the number of days between the retailer's latest purchase and the latest date in the overall dataset.
- **Scoring (0-3):** An R-Score was assigned based on risk logic. Scores closer to 0 are best (Ascending Risk).

Logic: If Recency>90 → 3; If Recency>60 → 2; If Recency>30 → 1; Otherwise → 0

4. Calculating Frequency (F-Score)

| 2. Total Purchase (Frequency) | | | | | |
|-------------------------------|-----------------------|------------|--------|-------|--|
| Row Labels | Count of Invoice Date | Quartile 2 | Values | Score | |
| Amazon | 949 | 1531 | 949 | 1 | |
| Foot Locker | 2637 | Quartile 3 | 2637 | 3 | |
| Kohl's Sports | 1030 | | 1030 | 1 | |
| Direct | 2032 | | 2032 | 2 | |
| Walmart | 626 | | 626 | 1 | |
| West Gear | 2374 | | 2374 | 3 | |

- **Purchase Count:** Used the `COUNT` of `Invoice Date` as the frequency metric.
- **Scoring (1-3):** An F-Score was assigned based on Quartiles (Descending Risk: 3=Best Performance).

Logic:

If $\text{Freq.} \geq Q3 \rightarrow 3$;

If $Q2 \leq \text{Freq.} < Q3 \rightarrow 2$;

Otherwise $\rightarrow 1$

5. Calculating Monetary (M-Score)

| 3. Total Spending (Monetary) | | | | | |
|------------------------------|--------------------|------------|--------|-------|--|
| Row Labels | Sum of Total Sales | Quartile 2 | Values | Score | |
| Amazon | 78 M | 142292875 | 78 M | 1 | |
| Foot Locker | 220 M | Quartile 3 | 220 M | 3 | |
| Kohl's Sports | 102 M | | 102 M | 1 | |
| Direct | 182 M | | 182 M | 2 | |
| Walmart | 75 M | | 75 M | 1 | |
| West Gear | 243 M | | 243 M | 3 | |

- **Total Spending:** Used the `SUM` of `Total Sales` as the monetary value.
- **Scoring (1-3):** The Quartile calculation and scoring logic are **identical to the Frequency process**.

6. Calculating Relative Risk Score

| Retailer | R-Score | F-Score | M-Score | Total Score | Relative Risk Score (0-100%) | Risk Category |
|----------------------|---------|---------|---------|-------------|------------------------------|---------------|
| <i>Amazon</i> | 0 | 3 | 3 | 6 | 50% | Medium |
| <i>Foot Locker</i> | 0 | 1 | 1 | 2 | 17% | Low |
| <i>Kohl's</i> | 0 | 3 | 3 | 6 | 50% | Medium |
| <i>Sports Direct</i> | 0 | 2 | 2 | 4 | 33% | Low |
| <i>Walmart</i> | 0 | 3 | 3 | 6 | 50% | Medium |
| <i>West Gear</i> | 0 | 1 | 1 | 2 | 17% | Low |

1. **Score Inversion:** F and M Scores were inverted to align with the R-Score (3 = Bad).

Formula: F/M Risk Score = 4 – F/M Score

2. **Total Risk Score:**

Total Risk Score = (R-Score×2) + F-Risk + M-Risk

3. **Relative Risk Score (Prioritization):** Total Risk is normalized against the Max Risk score (12).

Relative Risk Score = (Total Risk Score/12) × 100

7. Calculating Risk Composition

| 6. Which Affects Each Risk Percentage | | | | |
|---------------------------------------|--------|--------|--------|-------------------|
| Retailer | R-Risk | F-Risk | M-Risk | Sum of Percentage |
| <i>Amazon</i> | 0% | 50% | 50% | 100% |
| <i>Foot Locker</i> | 0% | 50% | 50% | 100% |
| <i>Kohl's</i> | 0% | 50% | 50% | 100% |
| <i>Sports Direct</i> | 0% | 50% | 50% | 100% |
| <i>Walmart</i> | 0% | 50% | 50% | 100% |
| <i>West Gear</i> | 0% | 50% | 50% | 100% |

- **Purpose:** To determine the percentage contribution of R, F, or M to the retailer's Total Risk Score.

Component % Formula = Component Score (R x 2, F, or M) / Total Risk Score

- **Visualization:** This result is shown in the **100% Stacked Column Chart**.

INSIGHTS AND RECOMMENDATIONS

Key Findings

1. **Intervention Priority:** Based on the **Intervention Priority Index**, the three highest-risk retailers (Amazon, Kohl's, Walmart) fall into the High-Risk category with a 50% Relative Risk Score.
2. **Root Cause:** The **Composition of Each Risk Score** chart reveals that 100% of this 50% risk is **driven by F-Risk and M-Risk** (50% F-Risk, 50% M-Risk).
3. **Recency Status:** The R-Risk factor is 0%. This indicates that while these retailers are technically still active (Recency<15 days), their high intrinsic risk (low F and M values) suggests they are highly likely to churn once they pass the 30-day Recency threshold.

Actionable Recommendations

- **Sales & Account Management Action:** Teams must **shift their focus away from Recency to increasing Frequency and Monetary value**.
- **Program Focus:** Implement targeted incentives (e.g., volume discounts or new product offerings) specifically for High-Risk Retailers to boost their **purchase frequency** and/or **average transaction value**, thereby increasing their F-Score and M-Score.