

15º Congresso de Inovação, Ciência e Tecnologia do IFSP - 2024

RFV NA PRÁTICA: ESTRATÉGIAS DE SEGMENTAÇÃO E PREVISÃO DE COMPRA COM MINERAÇÃO DE DADOS

Rafael Andrade¹, Sergio Shimura²

¹ Pós-Graduando em Indústria 4.0, IFSP, Campus Sorocaba, rafael.andrade1@aluno.ifsp.edu.br

² Orientador em Indústria 4.0, IFSP, Campus Sorocaba, Doutor em Engenharia Elétrica. Universidade de São Paulo
sergio.shimura@ifsp.edu.br

Área de conhecimento (Tabela CNPq): 1.02.02.08-0 Análise de Dados

RESUMO:

Este trabalho aborda a análise do comportamento dos clientes utilizando o modelo RFV (Recência, Frequência e Valor), uma metodologia amplamente aplicada para segmentação de clientes com base em dados transacionais. A análise RFV permite entender melhor o engajamento dos clientes ao medir a última compra (recência), a frequência de compras em um período específico e o valor gasto nas transações. Além disso, foram calculadas medidas descritivas, como a média e o desvio padrão dos intervalos de compras, o que aprimora a precisão das análises. Com o advento do Big Data, a capacidade das empresas de coletar grandes volumes de dados transacionais aumentou significativamente. Isso torna a extração de informações relevantes e a identificação de padrões ocultos essenciais para estratégias de marketing e retenção de clientes. A mineração de dados desempenha um papel fundamental na análise de grandes volumes de dados, auxiliando na segmentação de clientes e na previsão de comportamentos futuros. Este trabalho demonstra como a análise RFV, combinada à mineração de dados, pode prever hábitos de consumo e segmentar clientes. A análise utiliza a média dos intervalos de compra e o desvio padrão para identificar padrões. Clientes com menor desvio padrão mostram maior consistência, permitindo estratégias de retenção eficazes.

PALAVRAS-CHAVE: Mineração de Dados; RFV; Análise de Dados; Segmentação.

Segmentation Strategies and Purchase Prediction Using Data Mining

ABSTRACT:

This work addresses customer behavior analysis using the RFM (Recency, Frequency, and Monetary) model, a widely applied methodology for customer segmentation based on transactional data. The RFM analysis provides a better understanding of customer engagement by measuring the last purchase (recency), the frequency of purchases within a specific period, and the amount spent in transactions. Additionally, descriptive measures, such as the average and standard deviation of purchase intervals, were calculated to enhance the accuracy of the analysis. With the advent of Big Data, companies' ability to collect large volumes of transactional data has significantly increased. This makes extracting relevant information and identifying hidden patterns essential for marketing strategies and customer retention. Data mining plays a fundamental role in analyzing large data sets, helping with customer segmentation and forecasting future behaviors. This study demonstrates how RFM analysis, combined with data mining, can predict consumption habits and segment customers. The analysis uses the average of purchase intervals and the standard deviation to identify patterns. Customers with lower standard deviation show more consistent behavior, enabling effective retention strategies.

KEYWORDS: Data Mining; RFM; Data Analysis; Segmentation

INTRODUÇÃO

Cada vez mais o mundo está orientado a dados, tanto na área fabril quanto administrativa, A partir de grandes volumes de dados, as empresas buscam extrair informações valiosas tornando-se uma vantagem competitiva essencial para os diversos setores. Uma das técnicas utilizadas para extrair as informações é a mineração de dados. Segundo Silva (2017) mineração de dados é definida em termos de esforços para descoberta de padrões em bases de dados, descobrindo relações ocultas, padrões e gerando regras para prever e correlacionar dados, que podem ajudar as instituições nas tomadas de decisões mais rápidas ou, até mesmo, a atingir maior confiança.

Dentro desse contexto, a análise RFV (Recência, Frequência e Valor) destaca-se como uma abordagem poderosa para entender o comportamento dos clientes. É uma metodologia amplamente utilizada para a análise do comportamento dos clientes (Christy et al, 2018). Originalmente popularizada no marketing e na gestão de relacionamento com o cliente (CRM), a análise RFV é utilizada para segmentar clientes com base em três dimensões: a recência da última interação, a frequência de interações ou compras, e o valor monetário gerado por essas transações. Ao categorizar os clientes com base nesses fatores, as empresas conseguem direcionar campanhas de marketing, personalizar ofertas e otimizar seus esforços para melhorar o engajamento e aumentar a lealdade.

O trabalho proposto é aplicado a dados reais de uma companhia do Reino Unido. Objetivo é demonstrar como análise RFV combinada com mineração de dados podem melhorar a segmentação de clientes, demonstrando o hábito de compras e identificando com mais precisão aqueles com maior potencial de novas compras. Essa abordagem visa demonstrar como é possível otimizar estratégias de marketing e vendas, resultando em campanhas mais eficazes e maior engajamento e retenção de clientes.

MATERIAL E MÉTODOS

Para alcançar os resultados necessários para a análise e discussão, foram aplicadas técnicas de mineração de dados no conjunto RFV, utilizando as ferramentas Microsoft Excel, Google Colab e a linguagem de programação Python.

Dados

Para realização do estudo foi extraído a base de dados Online Retail no Kaggle, plataforma online e comunidade de data scientists e profissionais de machine learning que oferece competições de data science, no formato CSV. O arquivo contém transações realizadas entre 1º de dezembro de 2010 e 9 de dezembro de 2011 por uma empresa de varejo online registrada no Reino Unido. A tabela possui 8 categorias: InvoiceNo, StockCode, Description, Quantity, InvoiceDate, UnitPrice, CustomerID e Country. Para o estudo foram considerados as seguintes variáveis:

- InvoiceNo: representa o número da fatura em um conjunto de dados de transações.
- Quantity: Quantidade de itens comprados.
- CustomerID: código de identificação do cliente.
- UnitPrice: Preço unitário do item.

Na Figura 1 abaixo é um exemplo simplificado das dez primeiras linhas filtradas no “dataframe” [quadro de dados].

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
93	536378	20725	LUNCH BAG RED RETROSPOT	10	01-12-2010 09:37	1.65	14688.0	United Kingdom
95	536378	22352	LUNCH BOX WITH CUTLERY RETROSPOT	6	01-12-2010 09:37	2.55	14688.0	United Kingdom
31	536370	10002	INFLATABLE POLITICAL GLOBE	48	01-12-2010 08:45	0.85	12583.0	France
66	536375	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	01-12-2010 09:32	2.55	17850.0	United Kingdom
47	536372	22632	HAND WARMER RED POLKA DOT	6	01-12-2010 09:01	1.85	17850.0	United Kingdom

FIGURA 1. Visualização do dataframe resumido.

Tratamento dos Dados

Após a extração dos dados, utilizou-se o Google Colab, uma plataforma gratuita do Google Research que fornece notebooks interativos baseados em Jupyter, permitindo a execução de código Python sem necessidade de configuração (Google Colaboratory, 2024). Os dados foram carregados e

manipulados com o suporte da biblioteca Pandas, uma ferramenta eficiente, flexível e de fácil uso, amplamente utilizada para análise e manipulação de dados em Python (Pandas, 2024).

Algumas transformações da estrutura dos dados iniciais foram necessárias para construção do modelo RFV. Na análise da variável 'Quantity' revelou a presença de valores negativos, e foi identificado que a variável 'UnitPrice' contém valores iguais a zero. Como ambas devem representar valores inteiros positivos diferentes de zero, esses registros foram excluídos utilizando a função "query", sem que haja prejuízo no desenvolvimento do estudo. Também foi calculado o valor total de cada linha, criando uma categoria chamada 'InvoiceRevenue', resultante da multiplicação das colunas 'Quantity' e 'UnitPrice', a qual foi utilizada posteriormente no cálculo do campo 'Valor'.

Cálculo do RFV, média dos intervalos de compra e desvio-padrão

O modelo RFV mede quando as pessoas compram, com qual frequência compram e quanto compram, amplamente aplicado no marketing e na gestão de vendas e clientes (Wei; Lin; Wu, 2010). Segundo Kandeil (2014) e Ernawati (2021), esse modelo continua sendo amplamente utilizado para compreender e analisar o comportamento dos clientes. Na sequência realizou os cálculos da recência, frequência e valor no Google Colab utilizando a biblioteca Pandas.

Recência: número de dias desde a última compra registrada no conjunto de dados. Na equação 1 é apresentado o cálculo da

$$R = DA - DUC \quad (1)$$

em que,

R – Recência;

DA – Data da análise;

DUC – Data da Última Fatura do Cliente (InvoiceNo).

Frequência: quantidade de dias únicos em que o cliente realizou uma compra.

$$F = \sum_{i=1}^n 1 \quad (2)$$

em que,

n - número de faturas (InvoiceNo) emitidas para o cliente.

Valor Monetário: gasto médio por data de compra.

$$V = \sum_{i=1}^n \frac{\text{InvoiceRevenue}_i}{n} \quad (3)$$

em que,

n - número de faturas (InvoiceNo) emitidas para o cliente;

InvoiceRevenue - valor total cobrado em cada fatura.

Neste estudo, além de calcular a média do intervalo entre as compras realizadas pelos clientes, também foi calculado o desvio padrão desses intervalos, permitindo uma análise mais detalhada da variação no comportamento de compra.

$$\text{Média de Intervalos de Compras} = \sum_{i=1}^{n-1} \frac{(\text{Data da Fatura}_{i+1} - \text{Data da Fatura}_i)}{n - 1} \quad (4)$$

em que,

n - número de faturas (InvoiceNo) emitidas para o cliente;

$$DP = \sqrt{\frac{1}{n-2} \sum_{i=1}^{n-1} (\text{Data da Fatura}_{i+1} - \text{Data da Fatura}_i) - \text{MIC}} \quad (5)$$

em que,

n - número de faturas (InvoiceNo) emitidas para o cliente;

MIC – Média de Intervalos de Compras.

Após a realização do processamento e a mineração dos dados, gerou-se a tabela de saída com o 'CustomerID, Recency, Frequency, Monetary, AveragePurchaseInterval e PurchaseIntervalStdDev, respectivamente Identificação do cliente, recência, frequência, valor, média do intervalo de compra e desvio padrão do intervalo de compra. Nos cálculos não considerou a pontuação do cliente, o foco é analisar padrões de comportamento de compra com base em variáveis mais diretamente relacionadas às transações. O *score* (pontuação), embora útil em algumas análises, pode introduzir complexidades adicionais que não são necessárias para os objetivos deste trabalho, como segmentação e identificação de hábitos de consumo.

RESULTADOS E DISCUSSÃO

Na Figura 2 são apresentadas as informações de saída do cálculo do RFV, a média dos intervalos de compras descrito como AveragePurchaseInterval e o desvio padrão como PurchaseIntervalStdDev geradas através dos dados processados.

CustomerID	Recency	Frequency	Monetary	AveragePurchaseInterval	PurchaseIntervalStdDev
0	12364.0	0	4 1313.10	34.333333	2.516611
1	12488.0	1	3 1298.66	32.500000	4.949747
2	12628.0	0	3 786.46	29.500000	2.121320
3	12748.0	0	209 33719.73	1.425837	3.129497
4	12856.0	0	6 2179.93	10.800000	3.701351
5	12916.0	5	4 3006.15	60.000000	4.358899
6	12971.0	0	86 11189.91	4.060976	3.862921
7	13038.0	3	3 908.46	104.500000	2.121320
8	13089.0	0	97 58825.83	3.619565	4.359762
9	13334.0	3	3 3536.82	111.000000	4.242641

FIGURA 2. Resultado do processamento de dados saída do cálculo do RFV, a média dos intervalos de compras e o desvio padrão.

Observa-se que as medidas descritivas, a média dos intervalos de compras e o desvio padrão, fornecem uma visão clara sobre o comportamento de compra dos clientes somando com os dados processados do RFV. A média dos intervalos de compras, associada à recência, reflete o tempo médio entre as compras, indicando o grau de atividade dos clientes. Um intervalo menor sugere uma frequência de compras mais alta, enquanto intervalos maiores podem indicar clientes menos ativos ou que correm o risco de se tornarem inativos. O desvio padrão complementa essa análise, revelando a variação no comportamento dos clientes. Um desvio padrão elevado indica uma grande dispersão nos intervalos de compra, sugerindo que alguns clientes compram frequentemente enquanto outros têm períodos mais longos de inatividade. Já um desvio padrão menor sugere que a maioria dos clientes segue um padrão de compra mais consistente.

Com os dados descritivos processados, nota-se na Figura 3 que foram geradas 4.338 informações RFV de clientes únicos. Desses, 2.832 tiveram a média calculada e 1.999 tiveram o desvio padrão processado. Observa-se que o menor desvio padrão foi de 0,707, o que indica que há clientes que realizam compras de maneira bastante regular, com pouca variação nos intervalos de tempo entre uma compra e outra. Esse comportamento sugere um padrão de consumo estável, o que pode ser utilizado para direcionar estratégias mais personalizadas de retenção e fidelização.

	CustomerID	Recency	Frequency	Monetary	AveragePurchaseInterval	PurchaseIntervalStdDev
count	4338.000000	4338.000000	4338.000000	4338.000000	2832.000000	1999.000000
mean	15300.408022	3.222453	4.272015	2054.266460	72.510606	41.781571
std	1721.808492	3.284536	7.697998	8989.230441	65.353258	35.759220
min	12346.000000	0.000000	1.000000	3.750000	0.000000	0.707107
25%	13813.250000	1.000000	1.000000	307.415000	29.500000	17.625880
50%	15299.500000	2.000000	2.000000	674.485000	53.225000	31.595358
75%	16778.750000	5.000000	5.000000	1661.740000	92.000000	54.466129
max	18287.000000	12.000000	209.000000	280206.020000	365.000000	242.537626

FIGURA 3. Dados descritivos dos dados processados.

Verifica-se que clientes que não apresentam um desvio padrão significativo, ou seja, aqueles com apenas duas compras ou menos, possuem um comportamento de compra menos previsível. Esses clientes podem ser mais desafiadores para análises futuras, pois a falta de dados históricos consistentes limita a identificação de padrões de compra. Nesse caso, outras técnicas de aprendizado de máquina podem ser consideradas para uma eventual inferência para prever a potencial chance de recompra pelo cliente.

Outra análise gerada através dos processamentos de dados, deu-se através da análise de correlação entre as variáveis RFV e a desvio padrão. Na figura 3 Mapa de calor da correlação RFV, a média dos intervalos de compras e o desvio padrão onde +1 ou -1 indica que as variáveis possuem forte correlação entre elas, respectivamente de diretamente proporcional e indiretamente proporcional. Entende-se que a média e o desvio são duas novas variáveis que foram mineradas e que agrega valor na análise e não possui uma relação direta ou indireta com outras variáveis do RFV.

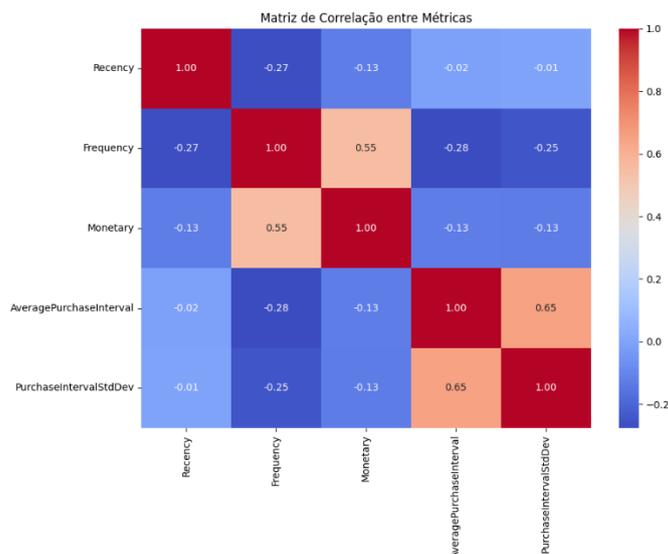


FIGURA 4. Mapa de calor da correlação RFV, a média dos intervalos de compras e o desvio padrão.

CONCLUSÃO

A análise RFV, quando combinada com técnicas de mineração de dados, revelou-se eficaz para compreender e segmentar o comportamento dos clientes de maneira mais precisa. A observação das medidas descritivas, como a média dos intervalos de compra e o desvio padrão, permitiu identificar padrões de compra, evidenciando clientes com comportamentos estáveis e aqueles com maior variabilidade. Essas informações permitem as áreas de marketing e gestão de vendas desenvolver estratégias de retenção e reengajamento mais direcionadas.

A inclusão dessas variáveis na análise proporciona uma visão mais abrangente, permitindo identificar oportunidades de fidelização e entender melhor as flutuações no comportamento de compra. Para clientes com poucas transações, que não apresentam um desvio padrão mensurável, surge um

desafio adicional. No entanto, em trabalhos futuros, técnicas de inferência em aprendizado de máquina poderão ser aplicadas para prever o comportamento de compra desses clientes com base em semelhanças com outras variáveis. Isso ampliará o potencial da análise e permitirá uma abordagem ainda mais estratégica para a gestão do relacionamento com o cliente.

CONTRIBUIÇÕES DOS AUTORES

Rafael Andrade desenvolveu o trabalho, sobre orientação do Sergio Shimura, o desenvolvimento e a revisão foi dado desde curadoria, pré-processamentos, análise de dados. A supervisão, revisão e aprovação da versão submetida deu-se pelo Sergio Shimura.

AGRADECIMENTOS

A todo corpo docente do Instituto Federal de São Paulo campos Sorocaba pela contribuição no desenvolvimento de temas inovadores dentro do curso Industria 4.0.

REFERÊNCIAS

CHRISTY, A. J.; UMAMAKESWARI, A.; PRIYATHARSINI, L.; NEYAA, A. RFM ranking – An effective approach to customer segmentation. *Journal of King Saud University – Computer and Information Sciences*, v. 33, p. 1251-1257, 2021. Disponível em: <https://doi.org/10.1016/j.jksuci.2018.09.004>. Acesso em: 21 out. 2024

ERNAWATI, E.; BAHARIN, S. S. K.; KASMIN, F. A review of data mining methods in RFM-based customer segmentation. *Journal of Physics: Conference Series*, v. 1869, 2nd Annual Conference of Science and Technology (ANCOSET 2020), 28 nov. 2020, Malang, Indonésia. IOP Publishing Ltd, 2021. p. 1-9. Disponível em: <https://iopscience.iop.org/article/10.1088/1742-6596/1869/1/012085>. Acesso em: 21 jul. 2024.

GOLDSCHMIDT, R.; PASSOS, E. *Data mining: um guia prático*. 4ª reimpressão. Rio de Janeiro: Editora Elsevier, 2005.

GOOGLE COLABORATORY. Perguntas frequentes. Disponível em: <https://research.google.com/colaboratory/intl/pt-BR/faq.html>. Acesso em: 25 jan. 2023.

KANDEIL, D. A.; SAAD, A. A.; YOUSSEF, S. M. A two-phase clustering analysis for B2B customer segmentation. In: 2014 International Conference on Intelligent Networking and Collaborative Systems (INCoS), 2014, Salerno. Anais... New York: IEEE, 2014. p. 1-8. Disponível em: <https://ieeexplore.ieee.org/xpl/conhome/7056704/proceeding>. Acesso em: 22 jul. 2024.

MOGHADDAM, Saman Qadaki; ABDOLVAND, Neda; HARANDI, Saeedeh Rajae. A RFMV Model and Customer Segmentation Based on Variety of Products. *Journal of Information Systems and Telecommunication*, v. 5, n. 3, p. 131-155, jul./set. 2017. Disponível em: <http://www.jst.ir/Article/Download/14983> Acesso em: 21 jul. 2024.

PANDAS. Disponível em: <https://pandas.pydata.org/>. Acesso em: 25 jul. 2024.

SILVA, L. A.; PERES, S. M.; BOSCARIOLI, C. *Introdução à Mineração de Dados com R*. 1. ed. Rio de Janeiro: Editora Elsevier, 2017.

WEI, Jo-Ting; LIN, Shih-Yen; WU, Hsin-Hung. A review of the application of RFM model. *African Journal of Business Management*, v. 4, n. 19, p. 4199-4206, dez. 2010. Disponível em: <http://www.academicjournals.org/AJBM>. Acesso em: 13 out. 2024.