



Encontrando Perfis de Consumidores a partir da Redução de Dimensionalidade por Componentes Principais

Gilsiley H. Darú¹
Selma R. A. Cantelli²
Felipe D. Motta³
Roseli Romero⁴
Afonso Paiva⁵
ICMC-USP

1 Introdução

Este presente trabalho é sobre o comportamento atual do consumidor no setor varejista brasileiro. No ápice da pandemia, as expectativas de crescimento para o segmento atacadista eram de -5,89% abaixo do consumo anual do ano anterior. Segundo a análise da Associação Brasileira de Atacadistas e Distribuidores (ABAD) [1], no entanto, até outubro de 2020 o setor acumulou uma alta de 4,43%.

Para explicar este crescimento faz-se necessário entender o comportamento do consumidor. A partir desta necessidade, [2], é possível observar o crescimento do e-commerce, consolidando-se como um canal importante do setor. Ainda de [2] pode-se perceber que, mesmo em meio a instabilidades econômicas e medidas de saúde pública, o segmento atacadista distribuidor se consolidou garantindo a distribuição e o abastecimento em todos os Estados e regiões do Brasil. Analisamos a partir da redução de dimensionalidade, [3] os itens que mais são comprados em conjunto, os quais explicam a variabilidade dos dados. A partir destes padrões de consumo inferimos tipos de perfis de clientes.

A análise foi feita com as seguintes classificações [4]:

- Categoria de produto
- Canal
- Itens mais comprados por brasileiros

¹ghdaru@gmail.com

²selma_andriatto@hotmail.com

³daltrozo@gmail.com

⁴rafrance@icmc.usp.br

⁵apneto@icmc.usp.br

Para exemplificar tipos de perfil de consumidores podemos citar o “consumidor responsável”, o qual tende a mostrar forte interesse em sustentabilidade, delineando compromissos ambientais, sociais e de governança (ESG), com métricas de consumo inclinadas a produtos ligados a questões de sustentabilidade. Um dos resultados esperados deste estudo é a contribuição ao setor varejista com relação a suas ofertas e itens promocionais, uma vez que identificados os conjuntos de produtos que são comprados de acordo com cada grupo de consumidor, estes tendem a ter frequência cíclicas como semanais, quinzenais e mensais.

Escolhemos utilizar os dados de um grande hipermercado de Santa Catarina pela importância nacional que o estado tem nesse setor, pois Santa Catarina possui 8 dentre os 100 maiores varejistas do Brasil [5], segundo estudo divulgado em 10 de maio de 2022, realizado pela Associação Brasileira de Atacadistas e Distribuidores (Abad) [2] em parceria com a consultoria NielsenIQ.

2 Trabalhos Relacionados

No trabalho [6], utilizando-se de análise fatorial confirmatória, foi feito um estudo referente aos fatores que afetam a atividade de compra dos clientes em lojas de varejo.

Aplicando análise fatorial nos dados coletados o trabalho de [7] buscou descobrir quais os fatores mais importantes que afetam as vendas em uma região.

O estudo de [8] utiliza técnicas de PCA e constrói um modelo para marketing de precisão baseado em rede neural.

O artigo [6] contribuiu com um estudo do impacto e de fatores que influenciam no poder de compra dos consumidores por meio de análise fatorial.

O trabalho [9] traz um estudo sobre fatores que influenciam no poder de compra de consumidores em lojas de varejo.

3 Metodologia

A Análise de Componentes Principais ou PCA (Principal Component Analysis) é uma técnica de análise multivariada [10], que foi utilizada nesse trabalho com objetivo de encontrar padrões de grupos de produtos vendidos, assim como identificar quais são os itens mais vendidos no período analisado. Estas informações são valiosas na manutenção do fluxo de estoque e na colocação em destaque dos produtos preferidos pelo consumidor nas melhores prateleiras. Analisamos durante um período de 30 dias as compras realizadas em um grande Hipermercado de Santa Catarina, para extraír padrões de consumo desse cenário de dados. O varejo selecionado localiza-se em uma região de classe média alta na cidade mais populosa de Santa Catarina. Os dados (Figura 1) extraídos e analisados são do ano de 2021, e contêm os produtos vendidos nesta loja dentre 7.968 diferentes itens. Foi possível identificar que nove (9) fatores (novas variáveis encontradas no modelo e que resumem os dados) representam 70% dos dados. Destes foi possível associar um perfil de consumidor para cinco fatores.

```

[ ] import pandas as pd

[ ] from google.colab import drive
drive.mount('/content/drive')

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).

[ ] ###Carga dos dados
data = pd.read_excel("/content/drive/MyDrive/SCC5948/Trabalho/vendasloja.xlsx")
dim_item = pd.read_excel("/content/drive/MyDrive/SCC5948/Trabalho/basecompleta.xlsx")

[ ] ###Preparando o arquivo de informações dos itens
dim_item.groupby(['id_item','nm_item','cd_ean']).agg({'vl_quantity_sold':[sum,'count',lambda x: sum((x>0))]}))

[ ] 
          vl_quantity_sold
                    sum  count <lambda_0>
id_item nm_item      cd_ean
5676488  CARGA GILLETTE MACH3 C/2 47400179240 400.0 222 222
5676495  PO GELATINA ROYAL CEREJA ZERO PCT 12G 7622300172879 387.0 141 141
5676503  SHAMPOO ELSEVE RT5 200ML 7899026459960 84.0 70 70
5676508  GOMA MASCAR TRIDENT INTENSE PCT C/4 7895800412794 206.0 145 145
5676512  ESMALTE COLORAMA CREM INCOLOR 7899026400863 36.0 34 34
...
33096351  CERVEJA 1906 RESERVA ESPECIAL LN 330ML 8412598034213 35.0 1 1
33096352  CERVEJA CUSQUENA TRIGO LN 330ML 7753749003025 17.0 9 9
33096353  CERVEJA ESTRELLA GALICIA LT 350ML 7898953990065 28.0 5 5
33239217  ALGODAO BELLACOTTON QUADRADO C/250 7898959957864 5.0 5 5
33250963  NECTAR DEL VALLE UVA TP 200ML 7894900660432 9.0 3 3
7968 rows × 3 columns

```

Figura 1: Visualização do dataset em Python.

4 Conclusões

Concluímos que encontramos cinco tipos de perfis de consumidores, incluindo tipos por sazonalidade de compras:

Consumidor 1 - Venda sazonal, consumidor compra em períodos específicos como páscoa, dia das mães, dias dos pais, aniversários, natal, festas de final de ano. Período de compra: trimestral (em média).

Consumidor 2 - Venda ao consumidor que mora sozinho e pertence à geração Z (nascidos de 1997 a 2010). Período de compra: semanal.

Consumidor 3 - Venda ao consumidor com até um filho, pertencente à geração Y ou Millennials (nascidos de 1981 a 1996). Período de compra: mensal.

Consumidor 4 - Venda ao consumidor com grande família, ou mais de um filho e pertencentes à geração X (nascidos de 1965 a 1980) ou geração Y (nascidos de 1981 a 1996). Período de compra: mensal.

Consumidor 5 - Venda aos novos casais ou amigos, geração alfa (nascidos a partir de 2010). Período de compra: esporádico, pois procuram ofertas.

Agradecimentos

Agradecemos aos dados recebidos para o presente estudo ser possível.

Referências

- [1] de Souza, Claudemir José and de Proen, Odair and Soares, Thiago. *Desempenho organizacional e ferramentas de gestão: Uma análise das ações públicas e agenda de pesquisa, 6ª edição, volume 1*. Revista Estudos e Pesquisas em Administração, São Paulo, 2022).
- [2] Souza, Daniel and Rosa, Jonathas. Ranking abad nielseniq (2021), May. <https://abad.com.br/wp-content/uploads/2021/05/Apresentacao-evento-Ranking-Abad-NielsenIQ-2021-V14.pdf>.
- [3] Cantner, Graf, and Rothgang, M., Deloitte - Os poderosos do varejo global (2022), May. <https://www2.deloitte.com/br/pt/pages/consumer-business/articles/poderosos-do-varejo-global.html>.
- [4] Lemos, Artur Bittencourt. Processamento de imagens e machine learning aplicados ao varejo de moda para identificação de produtos vendedores, (2019), *Universidade Federal do Rio de Janeiro*.
- [5] Machado, Pedro. Santa Catarina tem oito empresas entre os 100 maiores atacadistas e distribuidores do país, may, (2002) <https://www.nsctotal.com.br/colunistas/pedro-machado/sc-tem-oito-empresas-entre-os-100-maiores-atacadistas-e-distribuidores-do>.
- [6] Mondal, Subhra and DAS, Subhankar and Musunuru, Kamakshaiah and Dash, Madhusmita, Study on the factors affecting customer purchase activity in retail stores by confirmatory factor analysis *Revista Espacios*, volume 38, number 61, (2017).
- [7] Ananthakumar, Usha and Mridha, Ankita. Application of factor analysis in analyzing sales of the retail stores, *2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, pages: 1-4, (2016).
- [8] Liu, Hongping. Big data precision marketing and consumer behavior analysis based on fuzzy clustering and pca model, *Journal of Intelligent & Fuzzy Systems*, volume 40, pages: 6529–6539 (2021).
- [9] Mondal, Subhra and Mall, Manmohan and Mishra Uma S and Sahoo, Kalyan. Investigating the factors affecting customer purchase activity in retail stores, *Revista Espacios*, volume 38, number 57, (2017).
- [10] Oliveira, Bruno. Análise de componentes principais ou pca, <https://operdata.com.br/blog/analise-de-componentes-principais>, (2019).