



José Pedro Alexandre Canoso

**Impacto da COVID-19 no Retalho dos
Estados Unidos da América: Comparação por
Análise de *Clusters* para Séries Temporais**

Coimbra, outubro de 2025



José Pedro Alexandre Canoso

**Impacto da COVID-19 no Retalho dos
Estados Unidos da América: Comparação por
Análise de Clusters para Séries Temporais**

Dissertação submetida(o) ao Instituto Superior de Contabilidade e Administração de Coimbra para cumprimento dos requisitos necessários à obtenção do grau de **Mestre em Análise de Dados e Sistemas de Apoio à Decisão**, realizada sob a orientação da Professora Doutora Joana Jorge de Queiroz Leite.

Coimbra, outubro de 2025

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

TERMO DE RESPONSABILIDADE

Declaro ser o autor desta dissertação, que constitui um trabalho original e inédito, que nunca foi submetido a outra Instituição de ensino superior para obtenção de um grau académico ou outra habilitação. Atesto ainda que todas as citações estão devidamente identificadas e que tenho consciência de que o plágio constitui uma grave falta de ética, que poderá resultar na anulação da presente dissertação.

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

PENSAMENTO

“Hard work beats talent when talent doesn't work hard”

Tim Notke

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

DEDICATÓRIA

À minha mãe

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

AGRADECIMENTOS

Gostaria de agradecer a todas as pessoas que me apoiaram e contribuíram para a conclusão de mais uma etapa académica.

Em primeiro lugar, o meu sincero agradecimento à minha orientadora, a Professora Doutora Joana Jorge de Queiroz Leite, por todo o acompanhamento semanal, dedicação e total disponibilidade para me ajudar na conclusão deste projeto. Sem o seu apoio e orientação esta conclusão não teria sido possível.

Agradecer a todos os meus colegas, amigos e amigas que me acompanharam ao longo deste percurso.

Por último, um especial obrigado à minha família, em particular à minha mãe, ao meu irmão e à minha namorada, que sempre estiveram presentes, nos bons e nos maus momentos.

Muito obrigado a todos!

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

RESUMO

O Setor do Retalho nos Estados Unidos da América constitui um pilar económico fundamental, que, apesar da sua robustez, se revelou vulnerável aos impactos da COVID-19. A sua elevada dimensão e diversidade tornam-no um caso de estudo ideal e representativo para a deteção de novas tendências e vulnerabilidades globais. Neste enquadramento, o presente trabalho visa compreender a influência da COVID-19 na estrutura do retalho tendo por base as séries temporais de vendas dos seus setores e subsectores, obtidas do *Federal Reserve Economic Data*. Foi realizada uma análise de *clusters*, considerando o *clustering* hierárquico baseado em quatro características extraídas destas séries, concretamente, força da tendência, força da sazonalidade, entropia e autocorrelação de primeira ordem. Esta abordagem permitiu uma avaliação comparativa dos padrões de consumo no pré e no pós-COVID. Concluiu-se que a grande maioria das séries transitou para *clusters* diferentes dos inicialmente formados no período pré-COVID, o que reflete uma alteração nas preferências e nos padrões de consumo dos consumidores. Esta reorganização revela o profundo impacto que a COVID-19 teve no mercado de retalho dos EUA e a forma como estas séries reagiram ao período de recessão.

Palavras-chave: COVID-19, Retalho, Séries Temporais, Análise de Clusters, EUA

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

ABSTRACT

The Retail Sector in the United States of America constitutes a fundamental economic pillar which, despite its robustness, proved vulnerable to the impacts of COVID-19. Its large scale and diversity make it an ideal and representative case study for detecting new global trends and vulnerabilities. In this context, the present work aims to understand the influence of COVID-19 on the retail structure based on the time series of sales from its sectors and sub-sectors, obtained from the Federal Reserve Economic Data. A cluster analysis was performed, considering hierarchical clustering based on four characteristics extracted from these series, namely: trend strength, seasonality strength, entropy, and first-order autocorrelation. This approach allowed for a comparative assessment of consumption patterns in the pre- and post-COVID periods. It was concluded that most of the series transitioned to clusters different from those initially formed in the pre-COVID period, which reflects a change in consumer preferences and consumption patterns. This reorganization reveals the profound impact that COVID-19 had on the US retail market and how these series reacted to the period of recession.

Keywords: COVID-19, Retail, Time Series, Cluster Analysis, USA.

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

ÍNDICE GERAL

INTRODUÇÃO.....	1
1 Setor de Retalho dos EUA.....	3
1.1 O Setor de Retalho dos EUA antes da COVID-19.....	3
1.2 Impacto da COVID-19 no Retalho.....	5
1.2.1 Alterações no Comportamento do Consumidor: <i>E-commerce</i>	7
1.2.2 Respostas e Estratégias dos Retalhistas Durante a Pandemia.....	9
1.2.3 Retalho nos EUA Atualmente.....	10
2 Séries Temporais e Análise de <i>Clusters</i>	15
2.1 Séries Temporais.....	15
2.1.1 Alguns Conceitos Base.....	15
2.1.2 <i>Features</i> das Séries Temporais.....	16
2.2 Análise de <i>Clusters</i>	19
2.2.1 Processo de <i>Clustering</i>	19
2.2.2 Seleção da Métrica de Distância.....	21
2.2.3 Algoritmos Hierárquicos.....	22
2.2.4 Análise do Dendrograma.....	23
2.3 Processo Geral das <i>Features</i> para <i>Clustering</i> de Séries Temporais.....	24
3 Entendimento e Preparação dos Dados.....	27
3.1 Entendimento dos Dados.....	27
3.2 Normalização e Tratamento dos Dados.....	31
4 Modelação e Avaliação.....	38

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

4.1	Análise Pré COVID	38
4.2	Análise Pós COVID	46
4.3	Análise comparativa: Pré-covid e Pós-covid.....	53
CONCLUSÃO.....		58
REFERÊNCIAS BIBLIOGRÁFICAS		60
APÊNDICES		65
APÊNDICE 1. <i>Features</i> disponíveis na biblioteca <i>tsfeatures</i> do Python.....		66
APÊNDICE 2. Hierarquia das categorizações das vendas mensais no retalho e restauração nos EUA		70
APÊNDICE 3. Lista com todas as séries por <code>unique_id</code> e nome.....		71

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

ÍNDICE DE FIGURAS

Figura 1.1. Vendas do retalho em milhões de dólares.....	4
Figura 1.2. Evolução da variação percentual anual PIB nos EUA	6
Figura 1.3. Vendas de <i>social commerce</i> nos EUA	12
Figura 2.1. Processo de clustering	20
Figura 2.2. Feature Engineering	25
Figura 3.1. Estatísticas descritivas das séries temporais.	30
Figura 3.2 Série original, tendência, sazonalidade e resíduos	33
Figura 3.3 Distribuição e <i>Outliers</i>	33
Figura 3.4 Normalização pré COVID.....	35
Figura 3.5 Normalização pós COVID	36
Figura 4.1 Dataframe com setor_principal_id.....	38
Figura 4.2 Dendograma clustering hierarquico pré COVID	39
Figura 4.3 PCA pré COVID	45
Figura 4.4 Dendograma clustering pós COVID	46
Figura 4.5 PCA pós-COVID	53

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

ÍNDICE DE TABELAS

Tabela 4.1 Séries Cluster 1 pré-COVID	40
Tabela 4.2 Estatísticas descritivas cluster 1 pré-COVID	40
Tabela 4.3 Séries cluster 2 pré-COVID.....	41
Tabela 4.4 Estatísticas descritivas cluster 2 pré-COVID	41
Tabela 4.5 Séries cluster 3 pré-COVID.....	42
Tabela 4.6 Estatísticas descritivas cluster 3 pré-COVID	42
Tabela 4.7 Séries cluster 4 pré-COVID.....	43
Tabela 4.8 Estatísticas descritivas cluster 4 pré-COVID	43
Tabela 4.9 Séries cluster 5 pré-COVID.....	44
Tabela 4.10 Estatísticas descritivas cluster 5 pré-COVID	44
Tabela 4.11 Setores cluster 1 pós-COVID	47
Tabela 4.12 Estatísticas descritivas cluster 1 pós-COVID	48
Tabela 4.13 Setores cluster 2 pós-COVID	49
Tabela 4.14 Estatísticas descritivas cluster 2 pós- COVID.....	49
Tabela 4.15 Setores cluster 3 pós-COVID	50
Tabela 4.16 Estatísticas descritivas cluster 3 pós-COVID.....	50
Tabela 4.17 Setores cluster 4 pós-COVID	51
Tabela 4.18 Estatísticas descritivas cluster 4 pós-COVID.....	51
Tabela 4.19 Setores cluster 5 pós COVID.....	52
Tabela 4.20 Estatísticas descritivas cluster 5 pós-COVID.....	52
Tabela 4.21 Cluster pré COVID 1 e o pós-COVID.....	54
Tabela 4.22 Cluster pré COVID 2 e o pós-COVID.....	54

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

Tabela 4.23 Cluster pré-COVID 3 e o pós COVID.....	55
Tabela 4.24 Cluster pré-COVID 4 e o pós-COVID	56
Tabela 4.25 Cluster pré COVID 5 e o pós COVID.....	57

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

Lista de abreviaturas, acrónimos e siglas

AIC-Akaike Information Criterion

CRISP-DM – Cross Industry Standard Process for Data Mining

EUA – Estados Unidos da América

IA-Inteligência Artificial

PCA- Principal Component Analysis

PIB – Produto Interno Bruto

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

INTRODUÇÃO

O setor do retalho nos Estados Unidos da América (EUA), constitui um pilar económico fundamental, refletindo a evolução dos padrões de consumo e as tendências globais. No primeiro trimestre de 2020, este setor foi afetado pela COVID-19, um acontecimento que desencadeou mudanças profundas no setor do retalho afetando os níveis de vendas, alterando os padrões de consumo. O impacto da COVID-19 foi mais notório nos Estados Unidos da América, cuja dimensão económica e diversidade de segmentos tornam o país um caso de estudo relevante e representativo.

A principal motivação deste estudo é a necessidade de compreender de que forma a COVID-19 influenciou a estrutura do setor do retalho norte-americano, através da análise comparativa entre o período pré e pós COVID-19. Ao agrupar séries temporais com comportamentos semelhantes antes e depois do COVID-19, procura-se detetar alterações nos padrões de consumo e avaliar a resiliência ou vulnerabilidade de determinados subsectores.

Neste contexto, o objetivo deste trabalho reside na aplicação de uma técnica de agrupamento hierárquico às *features* de cada período, pré e pós COVID, de modo a ser possível analisar os *clusters* formados com as séries do retalho nos EUA.

De modo que todo o trabalho possua uma abordagem robusta e sistemática, a metodologia CRISP-DM (*Cross Industry Standard Process for Data Mining*) será adotada como a estrutura orientadora do projeto. Esta metodologia é caracterizada por ser uma metodologia flexível e amplamente reconhecida para a realização de projetos de análise de dados. “Em termos de um modelo de processo hierárquico, é constituída por conjuntos de tarefas descritas em quatro níveis de abstração (do geral ao específico): fase, tarefa genérica, tarefa especializada e instância de processo” (Chapman et al., 2000).

Os dados e *notebooks* usados para o desenvolvimento do trabalho estão disponíveis a partir do link:

https://drive.google.com/drive/folders/1MSsQdlfJ31oPtE_Vr3DsmHfSsl_xgTJE.

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

Em relação à estrutura, no que se segue o trabalho está dividido em quatro capítulos.

No primeiro capítulo, é feito um enquadramento teórico do setor do retalho dos EUA subdividido em duas secções. A primeira ilustra o setor do retalho dos EUA antes da COVID-19. A segunda secção que aborda o impacto da COVID-19 no retalho. No segundo capítulo são analisadas as séries temporais e a análise de *clusters*. O terceiro capítulo, será abordado o entendimento e preparação dos dados. Por último, no quarto capítulo, temos a modelação e avaliação dos *clusters* formados para o período pré e pós COVID.

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

1 Setor de Retalho dos EUA

O mercado do retalho consiste, essencialmente, nas vendas de bens acabados diretamente aos consumidores e nos serviços necessários para apoiar essas vendas (University of Texas, 2017). Este mercado inclui uma grande diversidade de operadores, desde comerciantes diretos, proprietários de unidades de pequena e média dimensão, a operadores de grandes superfícies. De acordo com a *Retail Trade Industry (2025)*, podemos caracterizar esta indústria por três grandes setores. Primeiro, o setor do retalho com loja física, que abrange concessionários de veículos motorizados e peças, lojas de vestuário e acessórios de roupa, lojas de alimentação e bebidas, e postos de combustível. Segundo, o setor do retalho sem loja física, que engloba as lojas de *e-commerce*, os operadores de máquinas de venda automática e os estabelecimentos de venda direta. Por último, temos o sistema de *franchising*, que é um modelo muito popular nos EUA, pois proporciona um plano estabelecido para a expansão do negócio e um meio de distribuição do risco.

1.1 O Setor de Retalho dos EUA antes da COVID-19

O retalho nos EUA é um dos maiores e mais dinâmicos do mundo. Salvo o ano de 2008 para 2009, é um mercado que tem vindo a aumentar constantemente o seu valor de vendas de ano para ano.

A Figura 1.1 mostra que o mercado de vendas a retalho nos EUA tem tido um crescimento notável desde 1992. Houve, no entanto, uma exceção com uma diminuição de 8,28% de 2008 para 2009, que pode ser atribuída aos efeitos da crise das hipotecas *Subprime* de 2008. Este evento financeiro, ao provocar um aumento do desemprego, diminuiu drasticamente o poder de compra dos americanos. Apesar desse revés, o setor recuperou e continuou a sua trajetória ascendente, de forma consistente, com o volume de vendas a atingir valores acima dos 5 biliões de dólares em 2018 e 2019.

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

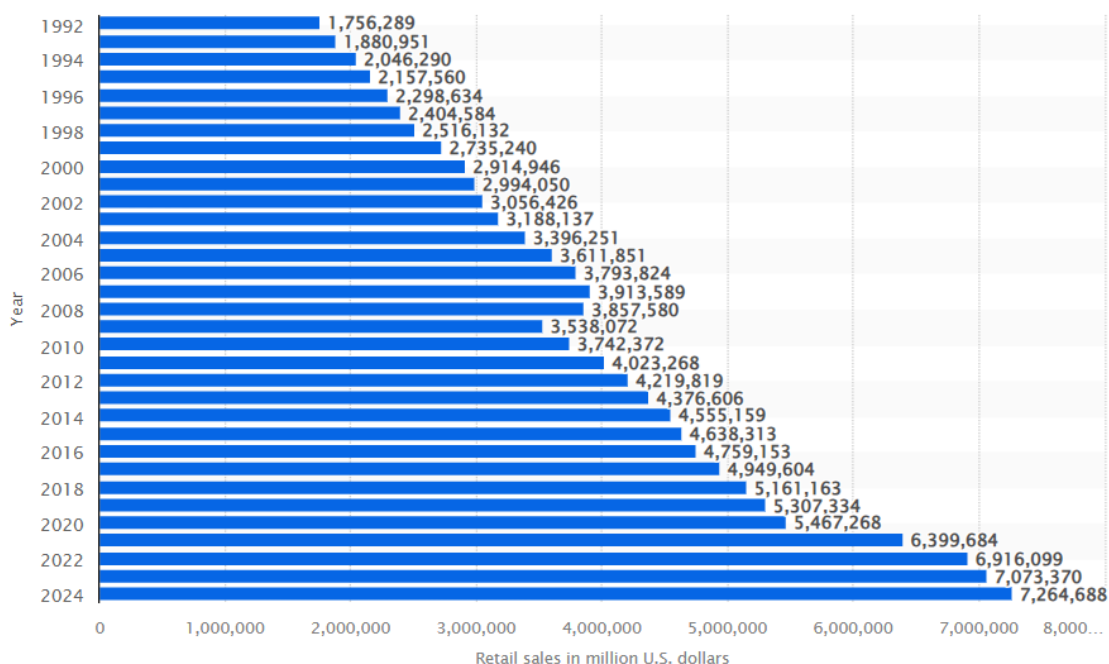


Figura 1.1. Vendas do retalho em milhões de dólares

Fonte: *United States Census Bureau* a partir da plataforma Statista ¹

No período anterior à pandemia de COVID-19, o mercado de retalho nos EUA foi moldado por forças destrutivas internas. Uma delas foi a incapacidade de retalhistas tradicionais, como a Sears, a J.C. Penney e a Toys"R"Us, de se adaptarem, o que levou a um declínio que o especialista em retalho Mark Cohen descreveu como "um desastre anunciado" (Cohen, 2018).

O declínio destes modelos de negócio tradicionais também se deveu a um excesso de espaço físico de retalho nos EUA, um problema estrutural que tornou o mercado insustentável. Por exemplo, só em 2017, o encerramento de lojas mais do que triplicou em comparação com o ano anterior, com aproximadamente 7000 lojas a fecharem nos EUA. Esta tendência foi mais visível no mercado norte-americano, ainda que não fosse exclusiva deste país (Helm et al., 2020).

¹ <https://www.statista.com/statistics/197576/annual-retail-sales-in-the-us-since-1992/>.

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

Os fatores responsáveis por estas mudanças incluíram um insustentável e descontrolado crescimento do espaço de retalho. O excesso de endividamento, a mudança no poder para os consumidores, que agora conseguiam comparar facilmente produtos e preços *online* foi um fator crucial. Com os consumidores a atuarem como os agentes de mudança na indústria.

Os consumidores começaram a assumir um papel cada vez mais importante, com mais informação e uma voz mais ativa. Queixavam-se de problemas como a falta de competência dos funcionários, espaços sobrelotados, longas filas de espera para pagar e a má qualidade dos produtos. Consequentemente, muitos viam com bons olhos um futuro com muito poucas lojas físicas, encarando a mudança como uma transformação inevitável da sociedade (Helm et al., 2020).

1.2 Impacto da COVID-19 no Retalho

Em dezembro de 2019, na cidade de Wuhan, na província de Hubei, China, surgiram os primeiros casos da COVID-19, uma doença causada pelo novo coronavírus SARS-CoV-2. Os primeiros casos foram associados ao mercado de frutos do mar de Huanan e, apesar do encerramento do mercado em janeiro de 2020, o vírus espalhou-se rapidamente.

Em 30 de janeiro de 2020, a organização mundial da saúde declarou o surto como uma emergência de saúde pública de âmbito internacional e, a 11 de março do mesmo ano, classificou-o oficialmente como uma pandemia global.

As medidas de confinamento, de acordo com (Houvèssou et al., 2021), nos Estados Unidos, foram implementadas de forma gradual e descentralizada. Inicialmente, o governo emitiu recomendações de restrição de movimento e distanciamento social nas regiões mais afetadas. Após esta fase, foi emitida uma recomendação de auto quarentena, em que a política se focou na manutenção da atividade económica.

Progressivamente, estas medidas foram suspensas, foi novamente permitido o acesso a eventos e encontros pessoais, e houve suspensão das políticas de isolamento domiciliar. Em geral, o país adotou o fecho de fronteiras, restrição de viagens e uso de máscaras, mas, ao contrário de outros países, não implementou um *lockdown* total. Devido às

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

medidas de confinamento, as pessoas passaram mais tempo em casa, o que levou a um menor consumo de bens. Isto resultou que o desemprego nos EUA foi de 14,7% em abril de 2020, que foi a taxa mais alta desde a grande depressão (Rahman et al., 2020).

A pandemia afetou o mundo inteiro, com reflexos na economia. Apesar do esforço que todos os governos fizeram em termos de política monetária, o produto interno bruto (PIB) – indicador macroeconómico que quantifica o valor monetário total dos bens e serviços finais produzidos dentro das fronteiras de um país, geralmente ao longo de um ano civil – teve uma queda abruta. As políticas fiscais, o orçamento de estado e os gastos públicos são os principais determinantes para um crescimento do PIB. Apesar disso, uma boa parte também está relacionada com o consumo privado, os gastos das famílias em bens e serviços e o investimento das empresas em novos equipamentos, edifícios e tecnologias.

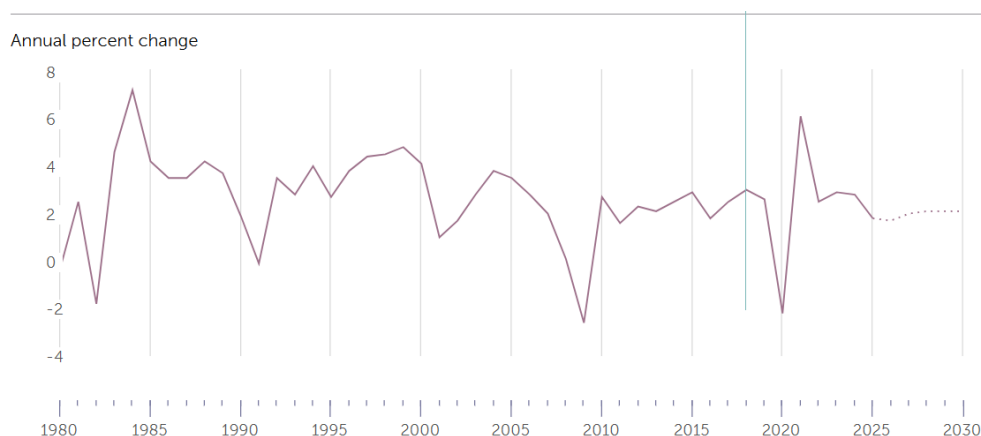


Figura 1.2. Evolução da variação percentual anual PIB nos EUA

Fonte: Fundo Monetário Internacional²A

Na figura anteriormente representada, Figura 1.2, podemos observar que os EUA, em 2019, tinham uma variação percentual do PIB de 2,6 pontos percentuais positivos e que em 2020 caiu para 2,2 pontos percentuais negativos.

Em março de 2020, foram confirmados 90870 casos de COVID-19, o que levou muitos governos a decretar o confinamento geral. O objetivo era travar as taxas de infeção

² https://www.imf.org/external/datamapper/NGDP_RPCH@WEO/USA.

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

exponenciais e evitar a sobrelotação dos hospitais, uma vez que, à época, não existia nenhum tratamento antiviral eficaz nem vacina disponível (Ferraz, 2022). O impacto da COVID-19, acelerou tendências e introduziu disrupções sem precedentes no setor do retalho. Embora os efeitos a longo prazo da COVID-19 ainda estejam por ser determinados, o seu impacto imediato no retalho foi significativo (Redda, 2021). A interrupção global das cadeias de abastecimento foi um dos maiores desafios no retalho. O fecho de fábricas na China e na Índia causou atrasos na entrega de produtos. Isto levou a um congestionamento logístico, com atraso nos produtos o aumento dos custos de transporte e uma escassez de produtos no mercado.

De acordo com Biggs et al. (2020), em novembro de 2020, a projeção era de que recuperação demoraria entre 15 a 20 meses, dada a necessidade de conter o vírus, havendo uma forte probabilidade de recessão. No entanto, ao olharmos para os números das vendas a retalho nos EUA, não se observa uma recessão, como aconteceu de 2008 para 2009. Segundo a informação publicada pelo Statista Research Department, 2023, foi possível observar um aumento do consumo, com uma grande parte dos consumidores a armazenar mais alimentos do que o normal nos EUA. As grandes superfícies generalistas, como a Walmart e a Target, foram as que mais lucraram, com as vendas a retalho de bens de grande consumo a aumentarem 10% face ao ano anterior.

A transição para o comércio eletrónico foi massiva a um alto nível, levando muitas empresas a entrar neste meio, por uma questão de sobrevivência, ou, no caso das que já estavam, a verem-se obrigadas a melhorar o seu serviço ao cliente. Estas ações alteraram radicalmente a forma como a sociedade funciona, mesmo em países e zonas onde a pandemia de COVID-19 ainda não era grave (Laato et al., 2020).

1.2.1 Alterações no Comportamento do Consumidor: *E-commerce*

O confinamento e o receio do futuro levaram a uma rápida e massiva mudança do comportamento do consumidor. Uma nova tendência que teve um grande impulso com a pandemia foi o mercado de *e-commerce*, que se refere à compra e venda de bens e serviços através da *internet* (Arlandis, n.d.). Esta nova forma começou a ser vista como uma opção

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

mais segura, rápida e cómoda para os consumidores. Se observarmos os dados, a tendência do *e-commerce* começou a ganhar mais força antes da COVID-19.

Em 2018, os EUA reportaram que 9,9% das vendas a retalho tinham sido originadas *online* (Liu & Lipsman, 2020).

Mesmo com o aparecimento desta nova tendência de mercado do *e-commerce*, e segundo dados de 2019 de Sayyida et al. (2021), quase 80% dos americanos que compravam *online* não abandonaram a compra em lojas físicas. De acordo com estes autores, emergiram dois perfis de consumidores: *showroomers* e *webroomers*. Um *showroomer* é um consumidor que procura a informação sobre um produto numa loja física e faz a compra final *online*. Um *webroomer* é um consumidor que usa os meios digitais para procurar um produto e no fim faz a sua compra numa loja física. Ambos os consumidores que se enquadram nestes perfis têm a necessidade de ver, tocar e sentir um produto, algo que só pode ser satisfeito em lojas físicas.

Embora já existissem consumidores com estes perfis antes da COVID-19, a pandemia impulsionou-os significativamente, com as pessoas a usarem a *internet* de forma mais intensa para pesquisar e comparar produtos. A percentagem de vendas *online* no retalho em 2020 foi de 14,5%, que foi um resultado superior em 3,5% a 2019 (Sayyida et al., 2021). Uma das justificações é o fácil acesso à informação *online* e o facto de os consumidores procurarem cada vez mais a opinião de outros compradores *online* (Dincer & Dincer, 2023). Outro motivo para o aumento exponencial nas compras *online* foram os constrangimentos que se faziam sentir no mercado de retalho. Adicionalmente, muitos destes consumidores alteraram o seu comportamento para evitarem o contacto pessoal. Algumas das implicações mais visíveis da COVID-19 no setor do retalho alimentar incluíram as compras por pânico, as mudanças nos padrões de compra de alimentos, as entregas de alimentos e os serviços digitais (Burgos & Ivanov, 2021).

O consumo *online* é mais visível nos *Millennials* do que em outra geração (Greg Petro, 2020). Geração nascida entre o início dos anos 80 e meados dos anos 90, foi onde se observou uma maior alteração de comportamento, em parte devido à sua familiaridade

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

com a tecnologia. Sendo nativos digitais, já habituados a comprar *online*, tornou a adoção deste comportamento mais fácil.

Outro fator que influencia a mudança de comportamento foi a população ir de áreas altamente urbanizadas para menos urbanizadas, devido ao elevado risco de contaminação. Com menos ofertas físicas, comerciais e com um aumento do tempo em casa, verificou-se uma diminuição do volume de negócios para restaurantes e ginásios. Em oposição, houve um aumento do *e-commerce* para a alimentação em casa (Biggs et al., 2020).

1.2.2 Respostas e Estratégias dos Retalhistas Durante a Pandemia

A pandemia de COVID-19 afetou os setores de retalho de forma diferente. Por um lado, segundo Steven Begley et al. (2020), a procura por bens e serviços em áreas de luxo, como roupas e sapatos de luxo, viagens aéreas, hotelaria, aeroespacial comercial, petróleo e gás, teve uma queda entre 20% a 45%. Por outro lado, devido ao seu papel fundamental os retalhistas alimentares tiveram de se adaptar rapidamente (Burgos & Ivanov, 2021). A alteração no comportamento dos consumidores, que passaram a cozinhar mais em casa, resultou um aumento significativo da faturação neste setor. Apesar desta exceção, a grande maioria das empresas necessitou de ajudas do estado. Segundo dados de (Roman et al. (2022), foi solicitada assistência financeira do *Paycheck Protection Program* por 61,7% das empresas, do programa *Economic Injury Disaster Loan* por 21,6% das empresas e, por último, do programa *Small Business Administration Forgiveness Loan* por 21,0% das empresas.

Os retalhistas tiveram de responder a perguntas como: que tipo de bens e serviços os consumidores querem e precisam neste ambiente alterado? Que fatores macroeconómicos irão influenciar as suas decisões de compra? Que tipo de novos hábitos de consumo irão permanecer, e por quanto tempo?

Estes precisaram de se reinventar, revendo os seus modelos de operação e encontrando novas formas de otimizar a sua estrutura, seja através de cortes de pessoal ou de ajustes na cadeia de abastecimento, para agilizar todo o processo. Aliado a isto, uma forte capacidade digital tornou-se crucial. Devido ao facto de os consumidores passarem mais

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

tempo em casa, foi necessário criar opções de recolha em loja e de entrega ao domicílio, bem como desenvolver novas aplicações digitais.

As despesas por parte dos retalhistas aumentaram, entre elas, a necessidade de maiores investimentos em intervenções para reforçar ou manter as cadeias de abastecimento, salvaguardar a saúde de funcionários e clientes, e manter os níveis de pessoal (Steven Begley et al., 2020). A logística de última milha foi um problema que estas empresas tiveram de enfrentar. Esta é a ponte final entre o retalhista e o consumidor, fundamental para a perceção de qualidade por parte do cliente final. Para o combater, tiveram de otimizar rotas, aumentar o *stock* em armazém e reforçar o pessoal para este tipo de tarefas. Aliado a isto, tiveram de desenvolver novos planos de *marketing* para contrariar as interrupções na lealdade à marca e melhorar a configuração das lojas físicas, bem como as experiências de compra. Esta mudança, em superfícies comerciais como supermercados, foi sustentada por um aumento de consumidores, uma vez que as pessoas em confinamento compravam alimentos e outros bens de primeira necessidade, muitas vezes fazendo *stock* para longos períodos de isolamento (Burgos & Ivanov, 2021).

1.2.3 Retalho nos EUA Atualmente

Presentemente, o mercado de retalho dos EUA é uma indústria dinâmica e de grande escala, com um valor estimado de vendas de aproximadamente 7,073 biliões de dólares em 2023. As projeções indicam que este valor crescerá para cerca de 8,38 biliões de dólares em 2028, com uma taxa de crescimento anual composta de 3,46% (Choudhury, 2025). Estes dados são importantes, pois revelam que as pessoas ou estão a comprar mais bens (aumentou o volume) ou bens mais caros devido ao aumento da inflação.

O retalho é um dos mercados mais importantes e relevantes dos EUA. As mais recentes estimativas de emprego indicam que 15,6 milhões de trabalhadores estavam diretamente empregados na indústria do retalho em meados de 2024. De acordo com a *National Retail Federation* ³, estima-se que o setor do retalho apoie, direta e indiretamente, 52 milhões

³ <https://www.trade.gov/selectusa-retail-trade-industry>.

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

de empregos a tempo inteiro e a tempo parcial, fazendo desta indústria o maior empregador do setor privado nos EUA. É um mercado bastante importante para os EUA, pois permite a entrada de jovens inexperientes no mercado de trabalho, assim como também proporciona oportunidades para posições de pessoas com um elevado grau académico.

Especificamente, para as vendas *online*, a Deloitte para 2024 previu um aumento de 7,0% a 9,0% em termos homólogos, e o crescimento observado situou-se em linha, entre 6,7% e 8,7% (Stephens et al., 2025). Este resultado significa que a mudança estrutural para o digital não foi temporária, mas sim algo que se está a tornar cada vez mais robusto. Os consumidores, neste momento, já não procuram apenas o preço mais baixo. Só esta característica não é suficiente; é necessário também que o produto tenha qualidade, mesmo que o seu preço seja superior ao do artigo mais barato. A liderança de preço, por si só, já não garante o consumo e a fidelidade do cliente.

Os retalhistas devem agora investir em gamas de produtos de valor acrescentado, oferecendo variedade e opções *premium* para criar diferenciação no mercado (Nokes & Henshaw, 2025). O mercado tem de estar preparado para os novos consumidores, que apresentam uma maneira diferente de consumir e uma perceção distinta do retalho. Apesar de a maioria dos consumidores (mais de 50%) ter idades acima de 45 anos, uma nova fatia (mais de 25%) é composta pela Geração Z (nascidos entre 1997 e 2012, atualmente com 13 a 28 anos) e pela Geração Alpha (nascidos entre 2013 e 2025, atualmente com 0 a 12 anos) (Nokes & Henshaw, 2025). Neste contexto, o *social commerce* tornou-se o próximo grande fator de disrupção no setor do retalho. Através de plataformas como o Instagram e o TikTok, os jovens fazem cada vez mais as suas compras. Devido ao poderoso algoritmo, cada *feed* é altamente personalizado e mostra os produtos de acordo com o gosto de cada pessoa. Geralmente, os utilizadores não estão a navegar com uma intenção específica de comprar; essa decisão é despertada pelo conteúdo que veem. Este é um mercado em forte crescimento, pois 52% dos consumidores americanos já fizeram, pelo menos uma vez, uma compra numa plataforma de *social media* (Martin, 2025).

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

Conforme podemos observar na Figura 1.3, o mercado de *social commerce* tem aumentado ao longo dos anos. Em 2027, prevê-se que atinja um valor de 1223,70 dólares por consumidor. Embora não se observe um crescimento acentuado no número de consumidores, é notório que cada um irá gastar mais em cada compra.

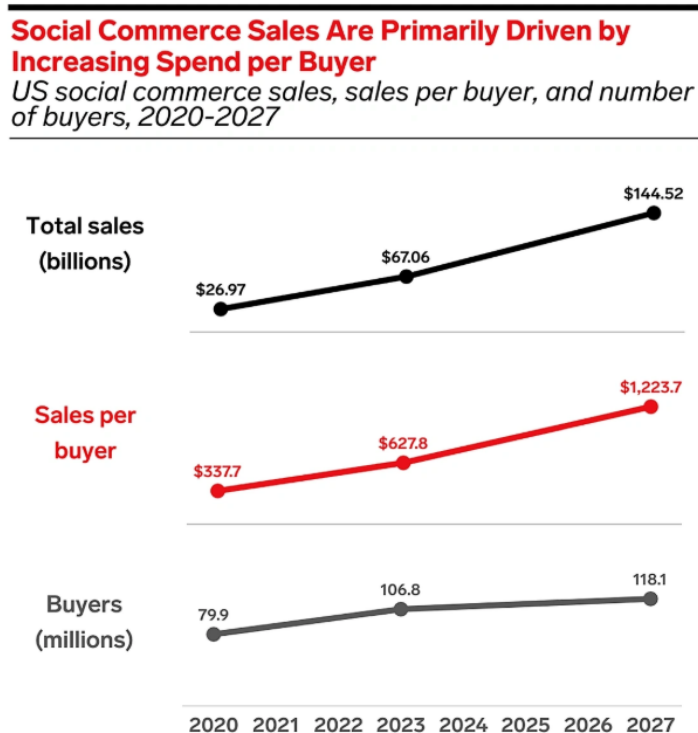


Figura 1.3. Vendas de *social commerce* nos EUA

Fonte: Insider Intelligence | eMarketer Forecast, Oct 2023 ⁴

Muitas empresas estão a alterar a sua estratégia de vendas e *marketing* para uma experiência omnicanal. O seu principal objetivo é proporcionar ao cliente uma jornada de compra fluida e consistente, independentemente do canal que ele utilize. A experiência omnicanal significa que todos os pontos de contacto – lojas físicas, *e-commerce*, *apps*, redes sociais e *contact centres* – estão totalmente integrados e comunicam entre si.

⁴ <https://www.emarketer.com/content/social-commerce-sales-rising-faster-than-number-of-social-buyers>.

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

Com mais de 80% da população dos EUA a realizar compras online, é de notar que, dessas compras feitas na internet, apenas 25% são realizadas com recurso a um computador, ao passo que 73% são feitas através de um *smartphone* (Ismail et al., 2025). É através do *smartphone* que os consumidores tendem a informar-se antes de tomar uma decisão. Os consumidores procuram cada vez mais informação antes de realizar uma compra, sendo que essa informação já é, muitas vezes, transmitida por Inteligência Artificial (IA).

Segundo o estudo realizado por Wright et al. (2025), metade dos consumidores inquiridos tomou uma decisão de compra com o apoio de IA generativa. Atualmente, 72% dos consumidores interagem com IA, e, destes, 30% confiam nas suas sugestões. Esta tendência transformou a IA na principal fonte de aconselhamento de compra que mais cresceu no último ano. Além disso, para os utilizadores ativos, é agora a segunda principal fonte de recomendações de produtos, superada apenas pelas lojas físicas.

Além disso, os sistemas de *self-service* e de pagamento autónomo tornaram-se a norma nas lojas de retalho, transformando radicalmente as condições de compra tanto para consumidores como para retalhistas (Choudhury, 2025). Esta automatização permite que os retalhistas otimizem os seus processos, reduzindo custos operacionais e possibilitando que os funcionários se concentrem em tarefas de maior valor. Para o consumidor, esta tecnologia oferece maior conveniência, rapidez e controlo sobre a sua experiência de compra.

Atualmente, as pessoas estão muito mais preocupadas com o ambiente. O facto de se saber cada vez mais sobre o impacto ambiental dos hábitos de consumo está a fazer crescer a procura por produtos amigos do ambiente e por formas de fazer negócios mais sustentáveis. Esta mudança para um consumidor mais exigente e informado está a forçar uma alteração em todo o mercado. Os produtos que seguem esta tendência são conhecidos como produtos ecológicos, produtos sustentáveis ou produtos verdes. Isso está a levar a inovações em áreas como produção, transporte e embalagem, bem como em novos produtos e serviços que atendem a essas necessidades. Contudo, apesar do bom desempenho, os resultados não foram os esperados, pois este é um setor que enfrenta

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

desafios, incluindo a inflação, a volatilidade do poder de compra do consumidor e outras necessidades que competem pela despesa das famílias (Sousa, 2023).

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

2 Séries Temporais e Análise de Clusters

Nesta secção, será feita primeiro a introdução aos conceitos fundamentais da análise de séries temporais. Em seguida, será dada uma breve explicação das características das séries temporais, quais são e qual a sua importância. Por último, aborda-se o procedimento *pipeline* de *clustering* de séries temporais que irá ser útil para a aplicação no capítulo seguinte de análise e agrupamento de séries (*clustering*).

2.1 Séries Temporais

2.1.1 Alguns Conceitos Base

Uma série temporal representa um conjunto de observações de dados registados em intervalos de tempo regulares ou irregulares, oferecendo um registo cronológico de fenómenos observados, como sinais vitais, vendas, preços no mercado de ações, ou mudanças climáticas (Kong et al., 2025).

Um dos objetivos da análise de séries temporais é explorar as suas componentes: tendência, sazonalidade, componente cíclica e componente irregular (ou residual). A tendência representa as mudanças de longo prazo nos dados, refletindo o crescimento ou o declínio ao longo de um período prolongado. A sazonalidade representa as variações periódicas que ocorrem em intervalos regulares e previsíveis, podendo ser semanais, mensais ou anuais. Os ciclos são caracterizados por flutuações que ocorrem em períodos irregulares e geralmente mais longos do que a sazonalidade, estando frequentemente ligados a fenómenos económicos, como os ciclos de negócio. Por fim, a componente irregular corresponde à parte dos dados que resta e que não pode ser explicada pelas outras componentes, representando as flutuações aleatórias e não estruturadas da série.

Neste sentido, a análise de uma série temporal envolve, frequentemente, a sua caracterização através da extração de *features*. Uma seleção de características elegante pode diminuir significativamente a carga de trabalho e simplificar o processo de design subsequente (Xu & Wunsch, 2005).

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

As propriedades e os padrões intrínsecos encontrados nos dados históricos podem fornecer informações valiosas para descrever ocorrências futuras (Kong et al., 2025). No caso do retalho, a análise de dados de vendas históricas pode revelar padrões sazonais de crescimento ou declínio, permitindo antecipar a procura de produtos. Precisamente, a quantificação destas e de outras características inerentes à série que irá fornecer informação relevante e estruturada aos modelos, para a descrever e prever de forma eficaz.

2.1.2 *Features* das Séries Temporais

Com base no artigo (Kang et al., 2017), e com o recurso à biblioteca *tsfeatures* no apêndice 1, foram selecionadas *features* que forneceram resultados interessantes em outros estudos já realizados. As *features* são as seguintes: força da tendência, hurst, força sazonal, autocorrelação, entropia e estabilidade. De seguida, vamos aprofundar cada uma delas.

A força da tendência, que permite entender a evolução subjacente de um fenómeno ao longo de um prazo de uma série temporal. Esta pode ser ascendente em que os valores aumentam de forma consistente ao longo do tempo, descendente com os valores a diminuírem, ou constante com os valores a permanecerem inalterados ao longo do tempo.

Existe uma tendência quando há uma mudança de longo prazo no nível médio de uma série temporal. Medimos a força da tendência através da decomposição inicial de uma série temporal x_t nas componentes de tendência, força sazonalidade e residual, utilizando uma decomposição STL (*Seasonal and Trend decomposition using Loess*): $x_t = S_t + T_t + R_t$. A força da tendência, representada na seguinte fórmula, da série temporal é então medida comparando as variâncias da série sem tendência R_t e da série sem sazonalidade $x_t - S_t$ (Kang et al., 2017).

$$\text{Força da tendência} = 1 - \frac{\text{var}(R_t)}{\text{var}(x_t - S_t)}.$$

O expoente de Hurst apresenta valores entre zero e um, e revela a persistência dos valores da série temporal. Este parâmetro indica se é provável que a série continue a subir,

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

mantendo uma tendência persistente, ou se, pelo contrário, tende a regressar ao seu valor médio. A interpretação dos valores é a seguinte: um valor igual a 0,5 indica um movimento aleatório, o que significa que os valores futuros não são previsíveis através dos valores do passado por outro lado, valores superiores a 0,5 indicam persistência na tendência da serie temporal. Esta é uma ferramenta importante, sendo um indicador da eficiência do mercado e amplamente considerada para testar o impacto de eventos históricos (por exemplo, crises financeiras) e a liberalização do mercado (Vogl, 2023).

A força sazonal mede a importância da componente sazonal em comparação com toda a série. É um valor que contribui para a variação total dos dados. Uma elevada força de sazonalidade indica que a série é dominada por ciclos de repetição, enquanto uma baixa força de sazonalidade indica que a série é mais afetada por tendências ou eventos aleatórios.

Identificar o período sazonal é essencial para entender a repetição de padrões em um conjunto de dados temporais. Sem essa informação, pode-se usar a função *findfrequency()* do pacote *forecast* em R. Essa função estima a frequência ao detetar o pico da densidade espectral em um modelo autorregressivo otimizado pelo critério AIC (*Akaike Information Criterion*), depois de remover tendências. O resultado final é ajustado para o número inteiro mais próximo. Embora essa técnica seja útil, Kang et al., (2017) destacam que, na prática, o analista geralmente já conhece o período, tornando a estimação automática desnecessária em muitos casos.

De seguida, temos a autocorrelação, que nos é apresentada pela característica *x_acf1*. Esta é uma correlação entre uma observação e a observação anterior. A autocorrelação varia entre valores -1 e +1, se o valor for próximo de 1, a correlação é positiva, ou seja, indica que se o valor da série temporal aumenta o próximo valor também terá tendência a aumentar. No caso de se observar um valor próximo de -1, a correlação é negativa, indica que se o valor da série temporal aumenta, o próximo tenderá a diminuir. Por último, no caso de ser um valor próximo de 0, isto indica que não existe correlação e que os valores passados não estão linearmente relacionados com os valores futuros.

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

A autocorrelação, representada na seguinte fórmula, mede a relação linear entre uma série temporal e a série defasada em um passo. Esta é fortemente afetada pela tendência e pela sazonalidade, pelo que calculamos as autocorrelações na série sem tendência e sem sazonalidade (Kang et al., 2017).

$$\text{Autocorrelação} = \text{Corr}(R_t, R_t - 1).$$

A magnitude da autocorrelação representa a intensidade da relação linear entre observações passadas e futuras. Portanto, uma autocorrelação elevada sugere maior potencial de previsibilidade na série, desde que componentes determinísticos, como tendência e sazonalidade, sejam removidos (Kang et al., 2017).

A entropia, permite quantificar a complexidade e o grau de organização dos dados. Sendo uma medida de aleatoriedade e imprevisibilidade, uma série com alta entropia é altamente imprevisível com pouco ou nenhum padrão visível. Enquanto uma com baixa entropia é mais previsível e contém padrões repetidos que podem ser facilmente identificados. Com isto, podemos entender o quão desafiador será prever o comportamento futuro de uma série. Existem diversas *features* que podem ser aplicadas na análise de séries temporais.

As medidas baseadas na entropia têm sido amplamente utilizadas em análises não lineares para avaliar a complexidade dos sinais e para medir a “previsibilidade” de uma série temporal. A medida de entropia espectral, representada na fórmula seguinte, *ForeCA* que está incluída no pacote R, a qual é uma estimativa da entropia de *Shannon* da densidade espectral $f_x(\lambda)$ de um processo estacionário x_t . Onde $f_x(\lambda)$ é uma estimativa do espectro da série temporal. Visto que $f_x(\lambda)$ descreve a importância da frequência λ dentro do domínio do período de x_t , a entropia de *Shannon* representa as contribuições relativas de diferentes frequências (Kang et al., 2017).

$$\text{Entropia} = - \int_{-\pi}^{\pi} f_x(\lambda) \log f_x(\lambda) d\lambda,$$

A estabilidade numa série temporal é definida como a consistência da sua variância e estrutura ao longo do tempo. Quando uma série é marcada pela instabilidade, é possível

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

observar que a sua dinâmica foi alterada. Uma alteração na política económica ou a introdução de uma nova tecnologia, por exemplo, pode causar uma quebra na dinâmica, levando a um aumento ou a uma diminuição da volatilidade. Por outro lado, numa série estável, a variância não apresenta variações de forma significativa.

2.2 Análise de *Clusters*

O *Clustering* é uma técnica de *data mining* onde dados semelhantes são colocados em grupos relacionados ou homogéneos, sem um conhecimento prévio das definições desses grupos (Rai et al., 2010). O agrupamento é feito por similaridade, sem um conhecimento prévio das definições dos grupos, ou seja, não é necessário definir à partida quantos *clusters* se pretendem formar. Alguns aspetos podem prejudicar o seu normal funcionamento, como o volume muito grande de dados das séries temporais, que torna complexo o cálculo da sua medida de similaridade, pois os dados contêm *outliers* que podem afetar o comportamento das séries.

Este método é de grande utilidade quando usado para uma análise exploratória, pois ajuda a identificar estruturas e a organizar objetivamente dados não rotulados em grupos semelhantes. Nesta secção, vamos abordar sobre o processo de *clustering*, a seleção da métrica da distância, o algoritmo hierárquico e a análise do dendograma.

2.2.1 Processo de *Clustering*

O *clustering* tem início com um conjunto de dados que não está organizado, que servirá de base para todos os resultados que se pretendem obter. Esta é uma abordagem que irá ajudar a descobrir padrões, tendências e informações úteis em grandes volumes de dados. Quanto mais parecidos forem os dados uns com os outros, maior é a probabilidade para a criação de um *cluster*. Sendo que por definição *clustering* é uma forma de aprendizagem não supervisionada, o que significa que não é necessário providenciar exemplos prévios de como os dados devem ser agrupados.

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

O processo de *clustering*, a preparação dos dados e a respetiva interpretação dos *clusters* obtidos encontra-se representado, conforme a Figura 2.1.

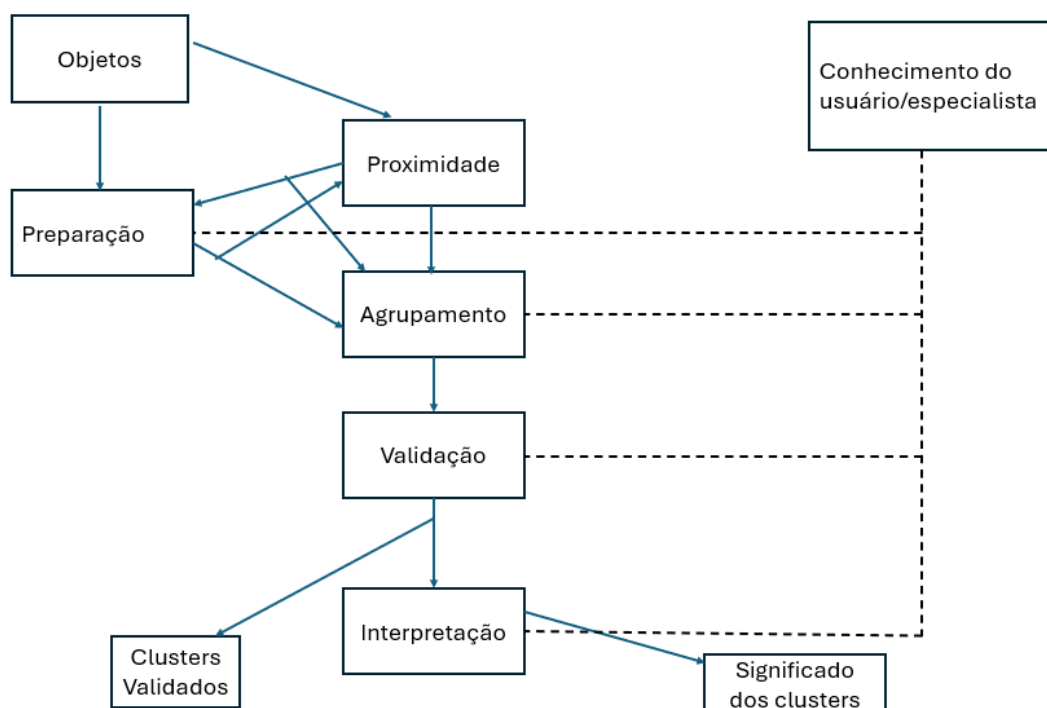


Figura 2.1. Processo de clustering

Fonte: Adaptado de Gama et al., 2017, pág.235

Este processo inclui várias etapas essenciais: a preparação dos dados, o *clustering*, a validação e a interpretação dos resultados, culminando na validação dos *clusters* e na atribuição do seu significado. A primeira fase do processamento interno é a preparação. É aqui que os objetos são tratados e transformados para uma representação adequada. Na fase de proximidade, é determinada a medida de similaridade entre eles. Esta métrica é essencial para que o algoritmo consiga distinguir o quão "próximos" ou "diferentes" são os objetos. De seguida, os *clusters* são formados através do agrupamento, que utiliza a medida de similaridade. Estes *clusters* são também validados, sendo que um *cluster* pode ser visto como uma coleção de objetos próximos ou que satisfazem uma relação espacial (Gama et al., 2017). O objetivo final de todo o processo é a interpretação e validação dos *clusters*, de acordo com o tema dos dados em estudo.

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

2.2.2 Seleção da Métrica de Distância

Para a seleção da métrica de distância, é necessário ter em consideração o conceito de proximidade e a forma como esta deve ser determinada. No contexto do cálculo de distâncias em agrupamentos, podem ser identificados três níveis de proximidade: a relação entre observações individuais, a afinidade entre um ponto e um conjunto de dados, e a semelhança entre distintos grupos de objetos. A distância entre *clusters* é definida precisamente pela proximidade, sendo que *clusters* mais próximos resultarão em distâncias mínimas mais curtas entre eles. O cálculo da distância entre *clusters* pode ser definido por uma ligação mínima (*single linkage*), ou por uma ligação média (*average linkage*) ou por uma ligação máxima (*complete linkage*) (Gama et al., 2017).

A definição da medida de proximidade e do critério de *clustering* utilizados pelos algoritmos geralmente depende implicitamente da imposição de certas hipóteses a respeito da forma dos *clusters* ou da configuração de múltiplos *clusters* (Gama et al., 2017). A proximidade é um conceito fundamental na análise de agrupamentos que se manifesta através de medidas de similaridade ou dissimilaridade (distância). A métrica de distância é a ferramenta que permite quantificar o quão semelhantes ou diferentes são os padrões de comportamento das séries temporais ao longo do tempo. O principal objetivo é agrupar as séries que exibem a maior proximidade.

Pode-se observar a influência da proximidade na forma como esta afeta a formação dos *clusters*. De acordo com Xu & Wunsch, (2005), quase todos os algoritmos de *clustering* estão, explícita ou implicitamente, ligados a alguma definição de medida de proximidade. A escolha do método de cálculo da distância é um passo importante para garantir uma melhor apresentação dos resultados. Existem diversos métodos que podem ser usados para o cálculo das distâncias entre *clusters*, sendo estes os mais comuns:

- *Ward* que minimiza a variância dos *clusters* serem combinados.
- O método *single* que é conhecido como o do vizinho mais próximo, em que a distância entre dois *clusters* é a menor entre quaisquer dois pontos nesses *clusters*.

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

- Outro método é o *complete* que é o oposto do *single* em que a ligação é feita com o vizinho mais distante.
- *Average* é um método em que é usada a média de todas as distâncias entre dois *clusters*.
- *Centroid* que utiliza a distância entre os dois centroides, ou seja, os vetores de média dos dois *clusters*.

2.2.3 Algoritmos Hierárquicos

O agrupamento hierárquico pode ser dividido em duas abordagens: a aglomerativa, que começa com *clusters* e os une sucessivamente; e a divisiva, que começa com um único *cluster* contendo todos os objetos e forma a sequência dividindo os *clusters* sucessivamente (Gama et al., 2017). À semelhança de outros algoritmos, este apresenta aspetos positivos e negativos. Um aspeto positivo deste algoritmo é a sua flexibilidade, que permite ajustar a sua granularidade. Além disso, destaca-se pela sua capacidade de trabalhar com diversos tipos de medidas de similaridade ou distância. Negativamente, o algoritmo tem um critério de determinação vago e, após o *cluster* estar criado, este permanece inalterável até ao final do processo.

Em que ponto importante é que este tipo de algoritmo hierárquico não lida bem com ruídos e *outliers* (Gama et al., 2017). Pontos aleatórios que não pertencem a nenhum *cluster*, podem ser agrupados em *clusters* que não estão muito relacionados, forçando-os a formar uma estrutura imprecisa dos dados. O mesmo acontece com os *outliers*, que pode até formar um *cluster* por si só.

A fase de validação dos *clusters* é uma fase em que verificamos quais os *clusters* mais significativos e que iremos estudar. Apesar de o mesmo algoritmo ser usado em diferentes estudos, a identificação dos parâmetros ou a ordem de apresentação dos padrões de entrada podem afetar os resultados (Xu & Wunsch, 2005).

Iremos recorrer à formação de *clusters* pelo algoritmo de agrupamento hierárquico. Para isto é importante que as similaridades entre objetos e grupos e entre dois grupos façam

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

parte da caracterização de cada algoritmo específico, sendo usadas, geralmente para unir/dividir *clusters*, e dependem do tipo de *cluster* que o algoritmo visa identificar (Gama et al., 2017). Uma característica importante, inerente à análise de agrupamento, e que torna difíceis a análise do desempenho e a comparação dos algoritmos, é a ausência de uma estrutura ideal que seja a resposta esperada para o agrupamento.

O agrupamento é fundamental, pois tem muita influência sobre a estrutura dos *clusters*: qual vai ser a sua forma, tamanho e o número de *clusters* que serão organizados. Os algoritmos de agrupamento são categorizados mais de acordo com os modelos subjacentes do que com os critérios de agrupamento. Assim, é preciso ter atenção redobrada na escolha dos algoritmos e no momento da sua comparação (Gama et al., 2017).

Estes algoritmos são diferentes entre si. Primeiro existem os algoritmos baseados em *grid* dividem o espaço de dados numa estrutura de células ou caixas. Em seguida, contam a densidade de pontos de dados em cada célula. As células com alta densidade são consideradas parte de um *cluster*, enquanto as células com baixa densidade são tratadas como ruído. Estes algoritmos baseados em densidade procuram por áreas de alta densidade de pontos de dados, separadas por áreas de baixa densidade. Estes definem um *cluster* como um conjunto de pontos bem definido e interligado.

Nos algoritmos hierárquicos, as soluções são tipicamente representadas por um dendrograma. Essencialmente, um dendrograma é formado por camadas de nós, cada uma representando um *cluster* (Gama et al., 2017). No caso de métodos aglomeráveis, o dendrograma é construído de baixo para cima, onde cada objeto se vai conectar com o grupo que lhe é mais próximo, vai unindo-se verticalmente ao longo dos ramos verticais. O processo continua unindo grupos de distância maiores até que todos os objetos estejam conectados em um único ponto.

2.2.4 Análise do Dendrograma

Um dendrograma é um diagrama em forma de árvore que ilustra as relações hierárquicas entre os *clusters*. Ele exhibe a ordem em que os *clusters* são unidos e as distâncias entre

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

eles (Boyko & Tkachyk, 2023). Esta é uma ferramenta muito visual e intuitiva, fundamental porque ilustra as relações hierárquicas e as distâncias.

Uma das principais vantagens é que o dendrograma, não exige que o utilizador especifique o número de *clusters* antecipadamente. Em vez disso, permite determinar o número mais apropriado de *clusters* a usar. Após esta análise, podemos interpretar os *clusters* e retirar conclusões. O dendrograma pode ser analisado da seguinte forma, começando pela estrutura de agrupamento, o eixo vertical representa a distância de dissimilaridade entre os *clusters*, e a altura na qual os ramos se unem indica o quão diferentes são os grupos que estão a ser combinados. Por consequência, quanto maior for essa altura da união, maior é a diferença entre os elementos agrupados. No eixo horizontal, podemos encontrar a identificação de cada série. Em oposição, uniões que ocorrem a uma altura inferior significam que os *clusters* são muito semelhantes entre si.

2.3 Processo Geral das Features para *Clustering* de Séries Temporais

As características ou *features* de um modelo são os elementos fundamentais que permitem transformar uma base de dados em informação suscetível de ser analisada. São um aspeto fundamental para que um modelo demonstre um desempenho ótimo. O mesmo algoritmo, quando aplicado a diferentes conjuntos de dados, não é garantia de que apresentará o mesmo resultado. Essa disparidade pode dever-se, frequentemente, ao facto de terem sido selecionadas *features* distintas. Um dos aspetos mais importantes neste processo é a Engenharia de *Features* (*Feature Engineering*), cuja aplicação no processo pode ser observada no diagrama da Figura 2.2.

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

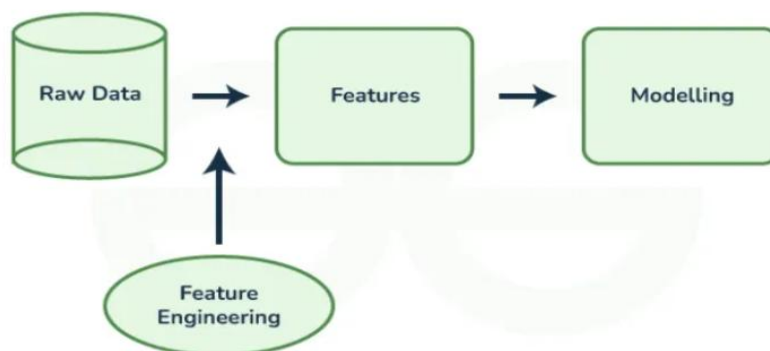


Figura 2.2. Feature Engineering

Fonte: Feature Engineering the Secret Sauce Behind Great Machine Learning Models⁵

A representação refere-se a um dos primeiros passos de transformação do formato dos dados (*raw data*) das séries temporais. Nesta fase onde irá ser feita a limpeza e organização dos dados, em que se pode optar por eliminar alguns dados, que não são relevantes. Após essa fase, existe a aplicação da *features engineering*, onde se pode criar atributos que melhorem a performance ou precisão do modelo. O objetivo é preparar os dados, tornando-os mais adequados ao modelo que será escolhido.

Aplicando as *features* aos dados entramos na última fase de *modeling*, que corresponde no âmbito deste trabalho à consideração das *features* nas séries que irá agrupar. As *features* nas series irão ser agrupadas de acordo com as suas similaridades. A medida de dissimilaridade, ou distância, que é um processo de partição dos dados de modo a que as séries temporais que exibam o mesmo padrão sejam agrupadas (Paparrizos et al., 2024). O objetivo desta medida é atribuir um valor numérico que represente a proximidade ou separação entre duas séries. A dissimilaridade zero significa que as séries são idênticas e

⁵ <https://medium.com/@polighosh7/feature-engineering-the-secret-sauce-behind-great-machine-learning-models-64da676d9b8f>

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

devem estar no mesmo *cluster*. Contrariamente, se a dissimilaridade for alta, significa que as duas séries são muito diferentes e devem pertencer a *clusters* distintos.

Todo este processo, garante que a formação de *clusters* ocorreu de forma coerente e que estes partilham dos mesmos padrões com um elevado grau de similaridade.

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

3 Entendimento e Preparação dos Dados

Na era da informação, tem-se tornado cada vez mais comum o aumento do volume de dados. Isto introduziu uma nova camada de complexidade, exigindo soluções adaptativas eficientes (Paparrizos et al., 2024). Uma destas soluções é o *clustering*, mas para uma boa representação os dados passam por uma análise mais cuidada de modo a percebê-los bem, para que, no final, possam ser devidamente representados para a sua análise e conclusões finais.

Esta fase de entendimento e preparação dos dados é caracterizada pela recolha e entendimento dos dados, a normalização e tratamento dos dados e pela análise exploratória destes.

3.1 Entendimento dos Dados

Os dados em estudo foram seleccionados no repositório do *Federal Reserve Economic Data* (FRED) ⁶. Este é um repositório centralizado de dados económicos, financeiros e demográficos sendo da autoria do Banco da Reserva Federal de St. Louis nos EUA.

A base de dados relativa às ao setor do retalho é alimentada através de um inquérito mensal realizado ao comércio de retalho e serviços de alimentação. Isto permite obter-se uma estimativa preliminar das vendas mensais, por atividade. Os questionários são enviados a uma amostra probabilística de 4700 empresas. Analisando estes resultados, o cálculo é efetuado usando um rácio entre as vendas ponderadas do mês atual e as do mês anterior. A estimativa de vendas totais antecipadas para o mês atual, em cada setor de atividade detalhado, é calculada multiplicando este rácio pela estimativa de vendas preliminar do mês anterior (derivada do mais abrangente *Monthly Retail Trade Survey* MRTS) ao nível do setor de atividade apropriado. Esta também tem em conta a variação que existem em momento sazonais ou feriados.

⁶ [USA, Retail Trade, NSA, Nation, Monthly, Census - Economic Data Series | FRED | St. Louis Fed.](#)

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

As séries são codificadas por um número NAICS (*North American Industry Classification System*), que é um número *standard* usado por esta agência federal norte americana com o fim estatístico de recolher, analisar e publicar informação relacionada com a economia comercial dos EUA (U.S. Census Bureau, 2024). A sua estrutura lógica é a seguinte:

- Os dois primeiros dígitos identificam o setor económico;
- O terceiro dígito indica o subsetor;
- O quarto o grupo de indústria a que pertence esse setor;
- O quinto define a indústria específica;
- O sexto dígito, fornece uma distinção das indústrias dentro do contexto nacional.

Para este estudo, vamos utilizar as séries *Monthly Retail and Food Service Sales by Kind of Business:2024*, que estão associadas ao código NAICS 44X72. A partir deste código, a análise inicia-se com o 4400, que engloba a informação relacionada com o retalho em geral. O código 4400 irá, por sua vez, subdividir-se nos códigos 441, 442 até ao 454, que correspondem às subcategorias do setor de retalho, as quais contêm outras subcategorias mais específicas no seu interior. Relativamente aos serviços de alimentação e locais de bebida, será usada a série 722. A figura, no Apêndice 2, esquematiza de forma detalhada todas as séries e as suas subdivisões, sendo que as assinaladas a vermelho correspondem àquelas que se encontram descontinuadas e que não foram utilizadas neste estudo.

Assim, para esta exploração inicial, começou-se por extrair uma lista com 232 séries utilizando os seguintes filtros:

- - *Concepts: "Retail trade"*
- - *Geography Types: "Nation"*
- - *Geographies: "United States of America"*
- - *Frequencies: "Monthly"*
- - *Sources: "Census"*
- - *Seasonal Adjustments: "Not Seasonally Adjusted"*

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

A primeira filtragem consistiu em ordenar os dados e seleccionar apenas as séries que continham informação sobre “*Retail sales*”, removendo as séries que se encontravam descontinuadas. Ficando uma lista de 44 séries temporais, representadas no apêndice 3, que compõem o conjunto de dados em estudo entre 1 de janeiro de 2015 a 01 de abril de 2025. A inclusão deste período de estudo é essencial para estabelecer um período de estudo robusto.

Relativamente à análise exploratória de dados, foram também analisadas as seguintes medidas que permitem um resumo estatístico dos dados: A Figura 3.1, podemos observar as estatísticas descritivas, de todas as séries assim como os gráficos a ilustrar os diversos conjuntos de dados.

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

	unique_id	nome	count	mean	std	min	25%	50%	75%	max
0	MRTSSM447USN	Gasoline Stations	124.0	45243.4	9214.9	26743.0	38193.5	44082.0	51076.2	73072.0
1	MRTSSM445USN	Food and Beverage Stores	124.0	69084.3	10150.3	51560.0	59893.8	67543.5	79037.8	89736.0
2	MRTSSM4451USN	Grocery Stores	124.0	62151.7	9009.7	46894.0	54071.0	60035.0	71505.8	78601.0
3	MRTSSM4453USN	Beer, Wine, and Liquor Stores	124.0	5033.3	988.2	3273.0	4220.2	4946.5	5738.8	7863.0
4	MRTSSM443USN	Electronics and Appliance Stores	124.0	7504.6	1256.2	2895.0	6982.0	7270.0	7737.8	12203.0
5	MRTSSM44511USN	Supermarkets and Other Grocery (Except Convenience) Stores	124.0	59492.0	8538.9	45222.0	51890.8	57410.5	68049.5	75640.0
6	MRTSSM45111USN	Sporting Goods Stores	124.0	4304.0	1007.5	2454.0	3607.2	4027.5	5048.8	7224.0
7	MRTSSM451211USN	Book Stores	124.0	713.9	247.3	164.0	566.5	633.5	804.2	1485.0
8	MRTSSM45112USN	Hobby, Toy, and Game Stores	124.0	1526.4	515.7	339.0	1189.0	1402.5	1616.2	3438.0
9	MRTSSM443142USN	Electronics Stores	124.0	5938.5	1229.1	1783.0	5411.8	5679.0	6051.5	10624.0
10	MRTSSM452USN	General Merchandise Stores	124.0	63907.4	9587.9	49837.0	56484.2	61396.0	71856.2	92765.0
11	MRTSSM442USN	Furniture and Home Furnishings Stores	124.0	10195.1	1398.7	3780.0	9218.2	10193.5	11250.2	12455.0
12	SM452319USN	All Other General Merchandise Stores	124.0	7185.4	1390.2	4697.0	5999.2	7243.0	8264.5	11071.0
13	SM45221USN	Department Stores	124.0	4710.9	2220.4	205.0	3109.0	4195.0	6072.5	14015.0
14	SM452311USN	Warehouse Clubs and Supercenters	124.0	52011.1	8662.3	37181.0	44642.5	51018.0	59435.5	75748.0
15	SM4523USN	General Merchandise Stores, Including Warehouse Clubs & Supercenters	124.0	59196.5	10021.9	42096.0	50734.8	58549.0	67496.2	86819.0
16	MRTSSM446USN	Health and Personal Care Stores	124.0	30341.9	4422.0	23176.0	26689.8	29191.0	33689.5	42257.0
17	MRTSSM44611USN	Pharmacies and Drug Stores	124.0	25697.8	3798.7	19520.0	22582.0	24732.5	28239.5	34925.0
18	MRTSSM453USN	Miscellaneous Store Retailers	124.0	10834.0	2055.9	6279.0	9188.8	10364.0	12669.5	14910.0
19	MRTSSM45321USN	Office Supplies and Stationery Stores	124.0	885.9	194.8	560.0	754.5	854.0	993.5	1527.0
20	MRTSSM45322USN	Gift, Novelty, and Souvenir Stores	124.0	1418.5	376.7	202.0	1183.8	1357.5	1648.8	2383.0
21	MRTSSM45330USN	Used Merchandise Stores	124.0	1471.8	304.0	386.0	1243.2	1396.0	1744.2	2124.0
22	MRTSSM444USN	Building Materials and Garden Equipment and Supplies Dealers	124.0	34463.4	7179.0	19341.0	29147.5	33443.5	39900.8	50120.0
23	MRTSSM44412USN	Paint and Wallpaper Stores	124.0	1144.0	255.9	663.0	953.8	1124.0	1330.2	1666.0
24	MRTSSM44413USN	Hardware Stores	124.0	2678.4	633.6	1642.0	2133.0	2500.5	3271.5	3997.0
25	MRTSSM454USN	Nonstore Retailers	124.0	75434.7	29871.5	35674.0	47163.8	72266.0	101374.8	154667.0
26	MRTSSM45431USN	Fuel Dealers	124.0	2962.8	1122.3	1353.0	2012.8	2795.0	3750.2	5933.0
27	MRTSSM4421USN	Furniture Stores	124.0	5572.3	789.4	1931.0	5019.2	5507.0	6116.0	7127.0
28	MRTSSM4541USN	Electronic Shopping and Mail-Order Houses	124.0	69225.0	28957.1	28513.0	42363.2	66297.0	93055.8	146567.0
29	MRTSSM448USN	Clothing and Clothing Accessory Stores	124.0	22103.1	5389.0	2631.0	19388.5	21524.0	24153.0	39876.0
30	MRTSSM4482USN	Shoe Stores	124.0	3049.9	641.8	469.0	2732.5	2984.5	3342.5	4679.0
31	MRTSSM44812USN	Women's Clothing Stores	124.0	3019.7	606.5	488.0	2700.2	3061.0	3351.5	4708.0
32	MRTSSM44814USN	Family Clothing Stores	124.0	9176.8	2365.9	645.0	8012.5	9001.5	10084.8	16880.0
33	MRTSSM44819USN	Other Clothing Stores	124.0	1359.3	258.2	227.0	1230.2	1357.0	1478.0	2107.0
34	MRTSSM441USN	Motor Vehicle and Parts Dealers	124.0	109907.5	18244.6	66554.0	97023.2	106965.0	124915.5	152434.0
35	MRTSSM441XUSN	Auto and Other Motor Vehicle Dealers	124.0	101122.0	16777.1	59814.0	89186.5	98698.5	114645.5	140668.0
36	MRTSSM4413USN	Automotive Parts, Accessory, and Tire Stores	124.0	8785.4	1559.9	6464.0	7536.0	8195.0	10075.5	12016.0
37	MRTSSM44111USN	New Car Dealers	124.0	84353.5	13404.6	48730.0	74872.5	81826.5	95588.2	118056.0
38	MRTSSM44112USN	Used Car Dealers	124.0	9953.3	2241.7	5497.0	8136.8	9239.0	11933.5	15309.0
39	MRTSSM722USN	Food Services and Drinking Places	124.0	67689.1	16064.0	30916.0	54603.2	62446.5	82339.2	101934.0
40	SM72251XUSN	Limited Service Eating Places	124.0	30772.4	7315.1	19190.0	24734.0	28481.5	37399.5	45170.0
41	MRTSSM7225USN	Restaurants and Other Eating Places	124.0	60458.8	14734.4	28025.0	48526.5	55088.5	74460.0	92223.0
42	MRTSSM4400CUSN	GAFO (General Merchandise Normally Sold in Department Stores)	124.0	113251.7	17813.8	67664.0	101308.8	109359.0	125162.5	169748.0
43	MRTSSM44X72USN	Retail Trade and Food Services	124.0	553945.6	104128.4	381513.0	466216.8	522961.0	662330.8	786673.0
44	MRTSSM4423XUSN	Furniture, Home Furnishings, Electronics, and Appliance Stores	124.0	17699.4	2281.2	6675.0	16425.2	17225.5	19054.8	23013.0

Figura 3.1. Estatísticas descritivas das séries temporais.

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

A coluna média representa o valor médio das vendas mensais ao longo do período de 1 de janeiro de 2015 a 1 de abril de 2025. Os setores mais dominantes são *retail trade and food services* e *Motor vehicle and part dealers*, que respetivamente apresentam os valores de 553945,6 milhões de dólares e de 109907,5 milhões de dólares. Estes volumes de venda fazem que estas sejam os motores da economia no retalho. Em contraste, encontramos as *book stores, office supplies and stationery stores*, com um volume de vendas respetivamente de 713,9 milhões de dólares e 885,9 milhões de dólares volume esse inferior aos anteriores.

Estas últimas séries (*book stores, office supplies and stationery stores*), são as que têm um desvio padrão menor, de 247,3 milhões de dólares e de 194,8 milhões de dólares respetivamente, como tal menos volatilidade. Em oposição, as séries *nonstore retailers, eletronic shopping and mail order houses, retail trade and food services* são as que apresentam um maior desvio padrão. Estes valores são relativamente esperados, devido a serem séries que continuaram ou aumentaram as suas vendas durante o período de COVID-19.

Relativamente ao valor mínimo, os valores mais baixos são registados pelas séries de *book stores e gift, novelty, and souvenir stores*. O que, nestes casos muito se deve a uma queda acentuada na fase inicial da pandemia ou no período de recessão.

Realizando uma análise interquartil (Q3-Q1), observamos que os valores que apresentam uma maior variação do volume de vendas são *Retail Trade and Food Services* e *Nonstore Retailers*. Enquanto para baixa dispersão temos as séries de *Book Stores* e *Office Supplies and Stationery Stores*.

3.2 Normalização e Tratamento dos Dados

O tratamento dos dados deve ser visto como a etapa inicial e crucial de um projeto de análise, onde os dados brutos são transformados num formato mais estruturado e mais limpo. Este irá permitir que exista uma integração de dados de múltiplas fontes, com o objetivo de que exista uma visão unificada e a redução da quantidade de dados que não

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

são relevantes, isto tudo sem perder a essência do conjunto. Neste caso como existiam vários ficheiros diferentes, realizou-se uma normalização dos dados.

De acordo com o artigo de (Kang et al., 2017), e com base na sua escolha de *features* para a análise das séries temporais, foram consideradas quatro *features* para a análise. Acredita-se que estas possam providenciar a melhor informação útil sobre a análise, sendo elas: a autocorrelação, a tendência e a força sazonal. Estas *features* foram aplicadas tanto para os dados de pré-COVID como para os dados pós- COVID.

A normalização é uma técnica formal que estrutura e organiza os dados e os agrupa na melhor forma que facilite futuras modificações, minimizando o impacto dessas mudanças no funcionamento do sistema.

Esta fase permite garantir a consistência do conjunto de dados antes da sua análise. O seu objetivo é estabelecer relações entre tabelas de acordo com regras concebidas para proteger os dados e tornar a base de dados mais flexível, eliminando redundância e dependência inconsistente⁷.

Antes de começar o processo de normalização, importa fazer uma análise da tendência, sazonalidade, distribuição e *outliers* para uma série. A tendência como já foi referido anteriormente representa o movimento de longo prazo de uma série temporal. A sazonalidade é caracterizada pelos padrões de variação periódica. A distribuição refere-se à maneira como os dados se espalham ou se agrupam.

Por último, um *outlier* é um ponto de dados que difere significativamente das outras observações de um conjunto de dados.

Analisando a série MRTSSM44X72USN, que corresponde aos dados de *Retail Trade and Food Services*, foi realizada uma análise da sua tendência, sazonalidade, distribuição e *outliers*. Esta análise pormenorizada permitiu seleccionar o melhor *scaler* (transformador) para a normalização dos dados. A decisão baseia-se no facto de a série temporal

⁷ <https://learn.microsoft.com/pt-pt/office/troubleshoot/access/database-normalization-description>

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

demonstrar uma clara tendência crescente ao longo do tempo, uma forte sazonalidade e uma distribuição praticamente simétrica, com o valor de assimetria próximo de zero não apresentando nenhum valor de *outliers*.

Conforme podemos observar nas Figura 3.2 e Figura 3.3 a média e o desvio-padrão são estatísticas robustas o suficiente para centrar os dados de forma eficaz. Estes fatores tornam o *StandardScaler* a melhor escolha.

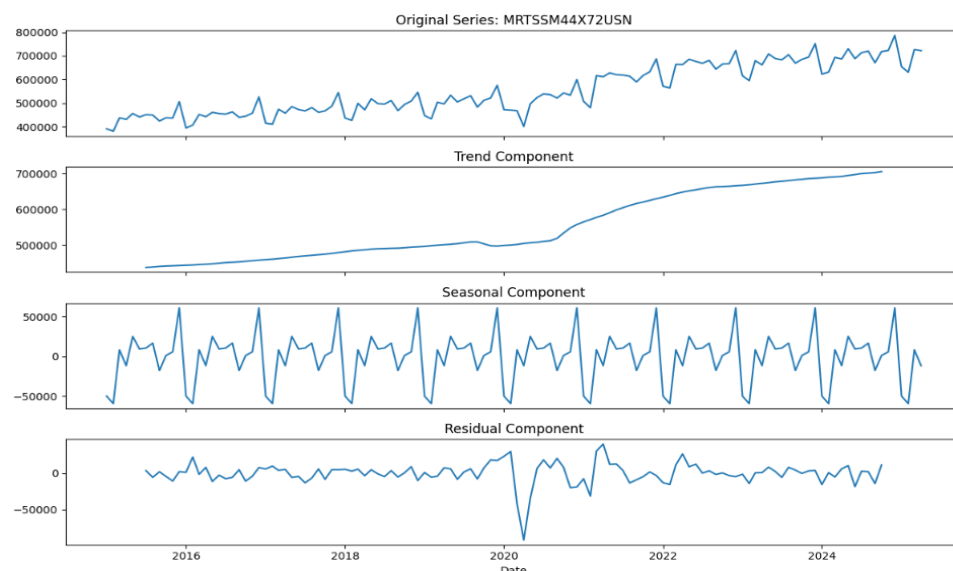


Figura 3.2 Série original, tendência, sazonalidade e resíduos

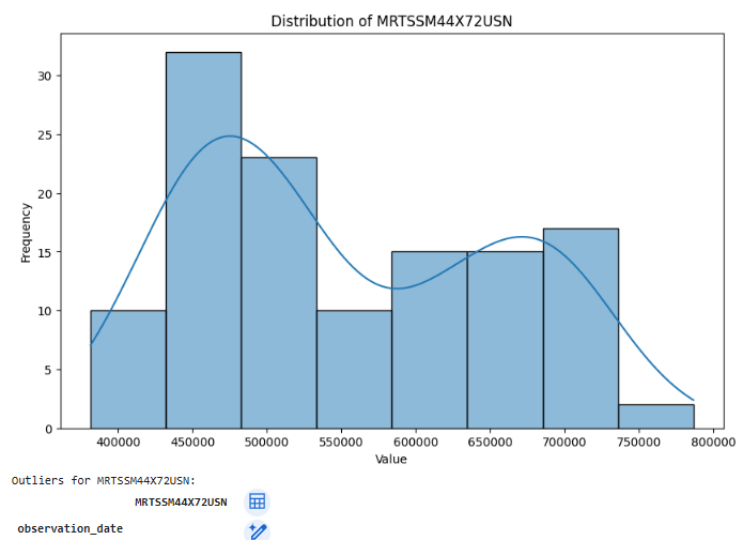


Figura 3.3 Distribuição e *Outliers*

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

Este *scaler* utiliza a média e o desvio-padrão para padronizar os dados. Este é usado quando a distribuição dos dados é aproximadamente normal. Transforma os dados para terem média $\mu=0$ e desvio-padrão $\sigma=1$. A fórmula que é a seguinte:

$$Z = \frac{x - \mu}{\sigma}$$

Em que Z vai ser o valor da observação depois de transformada, x é o valor original que será transformado, μ é a média aritmética e σ sendo o desvio padrão do conjunto de dados relativo à variável a padronizar.

Na Figura 3.4 podemos observar a tabela de normalização dos dados pré- COVID. Não contando para esta análise a série GAFO (*General Merchandise Normally Sold in Department Stores*), sendo ela a autocorrelação da série *new car dealers* tem uma forte correlação negativa, o que se traduz numa volatilidade elevada. O mesmo não acontece com a série *used car dealers*, em que a série, por apresentar um valor positivo de autocorrelação, terá uma volatilidade baixa.

No geral, os valores da entropia são baixos ou próximos de zero, o que nos indica que as séries, em geral, são simples e previsíveis. No caso da tendência, esta apresenta-se como sendo próxima de zero, ou seja, estável a longo prazo, com os valores a oscilarem em torno de uma média fixa.

Por último, a força sazonal, em geral, apresenta valores próximos de zero, o que faz com que as séries não sejam previsíveis nesse aspeto. Já relativamente a séries como a *electronics store* e a *furniture stores*, estas apresentam valores perto de 1 e -1, o que faz com que a sazonalidade seja muito forte e a força sazonal seja previsível.

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

	unique_id	nome	x_acf1	entropy	trend	seasonal_strength
0	MRTSSM4400CUSN	GAFO (General Merchandise Normally Sold in Department Stores)	-1.30	0.65	-0.97	1.02
1	MRTSSM44111USN	New Car Dealers	-0.58	0.70	0.12	-0.18
2	MRTSSM44112USN	Used Car Dealers	0.84	0.13	0.41	-0.82
3	MRTSSM4413USN	Automotive Parts, Accessory, and Tire Stores	-0.04	-0.33	-1.24	-0.85
4	MRTSSM4413USN	Motor Vehicle and Parts Dealers	-0.03	0.29	0.33	0.05
5	MRTSSM441XUSN	Auto and Other Motor Vehicle Dealers	-0.06	0.36	0.35	0.05
6	MRTSSM4421USN	Furniture Stores	-0.13	0.73	0.22	-1.38
8	MRTSSM442USN	Furniture and Home Furnishings Stores	-0.08	0.48	0.44	0.19
9	MRTSSM443142USN	Electronics Stores	-0.30	0.34	0.23	1.02
10	MRTSSM443USN	Electronics and Appliance Stores	-0.40	0.47	0.13	1.00
11	MRTSSM44412USN	Paint and Wallpaper Stores	1.30	-2.69	0.39	0.44
12	MRTSSM44413USN	Hardware Stores	1.07	-1.01	0.73	0.42
13	MRTSSM444USN	Building Materials and Garden Equipment and Supplies Dealers	1.12	-1.76	0.17	0.43
14	MRTSSM44511USN	Supermarkets and Other Grocery (Except Convenience) Stores	0.72	-0.57	0.71	-0.98
15	MRTSSM4451USN	Grocery Stores	0.83	-0.71	0.75	-0.92
16	MRTSSM4453USN	Beer, Wine, and Liquor Stores	-0.66	0.70	0.71	0.91
17	MRTSSM445USN	Food and Beverage Stores	0.65	-0.31	0.74	-0.34
18	MRTSSM44611USN	Pharmacies and Drug Stores	0.29	0.04	0.40	-3.14
19	MRTSSM446USN	Health and Personal Care Stores	0.22	0.24	0.50	-1.85
20	MRTSSM447USN	Gasoline Stations	1.56	-2.68	0.06	-2.14
21	MRTSSM44812USN	Women's Clothing Stores	-1.44	0.18	-2.26	0.72
22	MRTSSM44814USN	Family Clothing Stores	-1.14	0.64	0.03	0.99
23	MRTSSM44819USN	Other Clothing Stores	-1.33	1.44	-3.13	-1.07
24	MRTSSM4482USN	Shoe Stores	-2.37	0.11	-2.31	0.83
25	MRTSSM448USN	Clothing and Clothing Accessory Stores	-1.76	0.60	-1.69	1.04
26	MRTSSM44X72USN	Retail Trade and Food Services	0.11	0.41	0.80	0.52
27	MRTSSM45111USN	Sporting Goods Stores	-1.60	1.04	-0.55	0.84
28	MRTSSM45112USN	Hobby, Toy, and Game Stores	-0.35	0.58	-1.72	0.56
29	MRTSSM451211USN	Book Stores	-0.93	0.13	0.48	1.06
30	MRTSSM452USN	General Merchandise Stores	-1.09	0.72	-0.54	0.80
31	MRTSSM45321USN	Office Supplies and Stationery Stores	0.18	0.56	0.78	0.52
32	MRTSSM45322USN	Gift, Novelty, and Souvenir Stores	-1.36	-0.02	-1.31	0.76
33	MRTSSM45330USN	Used Merchandise Stores	0.73	1.23	0.23	-2.45
34	MRTSSM453USN	Miscellaneous Store Retailers	-0.01	0.25	0.40	0.22
35	MRTSSM4541USN	Electronic Shopping and Mail-Order Houses	1.11	-0.14	0.82	0.09
36	MRTSSM45431USN	Fuel Dealers	1.46	-3.15	-1.40	-0.46
37	MRTSSM454USN	Nonstore Retailers	1.19	-0.33	0.81	0.03
38	MRTSSM7225USN	Restaurants and Other Eating Places	1.30	-0.72	0.88	-0.16
39	MRTSSM722USN	Food Services and Drinking Places	1.35	-0.59	0.87	-0.53
40	SM45221USN	Department Stores	0.42	0.70	0.85	0.55
41	SM452311USN	Warehouse Clubs and Supercenters	-0.32	0.86	0.60	0.54
42	SM452319USN	All Other General Merchandise Stores	-0.39	0.74	0.62	0.57
43	SM4523USN	General Merchandise Stores, Including Warehouse Clubs & Supercenters	-0.32	0.84	0.63	0.59
44	SM72251XUSN	Limited Service Eating Places	1.55	-1.15	0.92	0.49

Figura 3.4 Normalização pré COVID

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

Relativamente à normalização Pós COVID, observamos na Figura 3.5 que existiram séries que viram as suas características alterarem-se significativamente.

	unique_id	nome	x_acf1	entropy	trend	seasonal_strength
0	MRTSSM4400CUSN	GAFO (General Merchandise Normally Sold in Department Stores)	-1.02	0.69	0.64	0.84
1	MRTSSM44111USN	New Car Dealers	0.25	1.48	-0.28	-1.70
2	MRTSSM44112USN	Used Car Dealers	0.70	0.21	-0.73	-0.66
3	MRTSSM4413USN	Automotive Parts, Accessory, and Tire Stores	1.07	-0.63	0.79	-0.04
4	MRTSSM441USN	Motor Vehicle and Parts Dealers	0.34	1.26	-0.18	-0.67
5	MRTSSM441XUSN	Auto and Other Motor Vehicle Dealers	0.26	1.34	-0.40	-0.79
6	MRTSSM4421USN	Furniture Stores	-0.39	1.31	-1.64	-3.08
8	MRTSSM442USN	Furniture and Home Furnishings Stores	-0.63	1.65	-1.16	-1.53
9	MRTSSM443142USN	Electronics Stores	-0.52	0.52	-0.63	0.38
10	MRTSSM443USN	Electronics and Appliance Stores	-0.62	0.78	-0.65	0.24
11	MRTSSM44412USN	Paint and Wallpaper Stores	0.77	-0.99	-0.11	0.48
12	MRTSSM44413USN	Hardware Stores	0.61	-1.09	0.58	0.79
13	MRTSSM444USN	Building Materials and Garden Equipment and Supplies Dealers	0.57	-1.52	-1.20	0.65
14	MRTSSM44511USN	Supermarkets and Other Grocery (Except Convenience) Stores	0.91	-0.88	0.87	0.04
15	MRTSSM4451USN	Grocery Stores	0.94	-1.01	0.89	0.07
16	MRTSSM4453USN	Beer, Wine, and Liquor Stores	-2.06	-0.64	-1.18	0.96
17	MRTSSM445USN	Food and Beverage Stores	0.68	-0.66	0.89	0.41
18	MRTSSM44611USN	Pharmacies and Drug Stores	0.86	-0.25	0.70	-1.43
19	MRTSSM446USN	Health and Personal Care Stores	0.62	0.17	0.75	-0.59
20	MRTSSM447USN	Gasoline Stations	1.43	-1.25	0.41	-1.71
21	MRTSSM44812USN	Women's Clothing Stores	-1.26	1.12	-0.77	-0.12
22	MRTSSM44814USN	Family Clothing Stores	-0.89	1.16	0.35	0.59
23	MRTSSM44819USN	Other Clothing Stores	-0.03	1.10	0.04	-1.05
24	MRTSSM4482USN	Shoe Stores	-2.25	0.23	-1.42	0.41
25	MRTSSM448USN	Clothing and Clothing Accessory Stores	-1.46	1.03	0.19	0.65
26	MRTSSM44X72USN	Retail Trade and Food Services	0.63	0.09	0.97	0.57
27	MRTSSM45111USN	Sporting Goods Stores	-2.26	-0.38	-3.74	0.70
28	MRTSSM45112USN	Hobby, Toy, and Game Stores	-0.32	-1.43	-1.93	0.93
29	MRTSSM451211USN	Book Stores	-0.94	-0.20	0.01	0.76
30	MRTSSM452USN	General Merchandise Stores	-0.43	0.19	0.77	0.79
31	MRTSSM45321USN	Office Supplies and Stationery Stores	-0.03	-0.08	0.42	0.41
32	MRTSSM45322USN	Gift, Novelty, and Souvenir Stores	-1.11	0.69	0.19	0.70
33	MRTSSM45330USN	Used Merchandise Stores	0.32	1.38	-0.17	-3.08
34	MRTSSM453USN	Miscellaneous Store Retailers	0.48	1.25	0.80	0.25
35	MRTSSM4541USN	Electronic Shopping and Mail-Order Houses	0.85	-0.36	0.97	0.74
36	MRTSSM45431USN	Fuel Dealers	1.21	-3.09	-1.02	0.50
37	MRTSSM454USN	Nonstore Retailers	0.90	-0.33	0.97	0.72
38	MRTSSM7225USN	Restaurants and Other Eating Places	1.37	-0.65	0.99	-0.35
39	MRTSSM722USN	Food Services and Drinking Places	1.41	-0.79	1.01	-0.31
40	SM45221USN	Department Stores	-1.36	0.05	-0.21	0.82
41	SM452311USN	Warehouse Clubs and Supercenters	-0.13	0.11	0.78	0.66
42	SM452319USN	All Other General Merchandise Stores	-0.67	-0.54	0.61	0.83
43	SM4523USN	General Merchandise Stores, Including Warehouse Clubs & Supercenters	-0.21	0.05	0.79	0.72
44	SM72251XUSN	Limited Service Eating Places	1.40	-1.09	1.05	0.48

Figura 3.5 Normalização pós COVID

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

Na série *Automotive parts, accessory, and tire stores*, podemos ver que esta agora tem um valor próximo de +1. Ou seja, após a COVID-19, apresentou um comportamento com padrões sazonais mais regulares.

No caso da entropia, as séries *motor vehicle and parts dealers*, *auto and other motor vehicle dealers*, *furniture stores* e *furniture and home furnishings stores* apresentaram valores mais próximos de 1. Isto torna a série mais complexa e imprevisível, algo que não acontecia antes da COVID-19.

A maior mudança na força da tendência foi observada nas séries *furniture stores* e *furniture and home furnishings stores*, em que os valores se foram afastando e tornaram-se negativos. Isto significa que estas séries apresentam um decréscimo forte e consistente ao longo do tempo.

Quanto à força sazonal, os valores para as séries *furniture stores*, *furniture and home furnishings stores* e *electronics stores* indicam que a sazonalidade ficou ainda mais fraca e, como tal, estas séries são muito suscetíveis a eventos imprevisíveis.

Terminada a Normalização dos Dados, onde as *features* foram padronizadas para garantir a igualdade de peso e evitar a distorção do agrupamento, o processo segue para a modelação e avaliação. A normalização dos dados tem um efeito de equalização na forma do espaço de atributos de um conjunto de dados, o que é crucial, pois as *features* agora prontas servirão para a criação do algoritmo de agrupamento hierárquico. O agrupamento em *clustering* é afetado positivamente pela normalização dos dados, resultando em *clusters* mais robustos e baseados em características comportamentais, e não na magnitude dos valores brutos.

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

4 Modelação e Avaliação

A modelação focar-se-á na organização das séries temporais em *clusters* com base na sua similaridade de comportamento. Posteriormente, a avaliação quantificará a qualidade e a robustez desses *clusters*. Nesta secção, irá ser realizada uma análise pré COVID e pós COVID, bem como uma comparação entre as duas.

4.1 Análise Pré COVID

Começou-se por proceder à identificação de que se a série era um subsetor ou um setor principal. Efetuou-se uma indicação de qual série pertencia cada subsetor para melhor organização numerou-se os setores entre os valores 0 e 12. Em que, para tal, foi criada uma coluna com o nome *setor_principal_id* como é possível observar na Figura 4.1.

'DataFrame with new 'setor_principal_id' column:'

	cod	nome	setor principal	subsector (se for subsector sim)	setor_principal_id
0	SM452319USN	All Other General Merchandise Stores	General Merchandise Stores	s	0
1	MRTSSM441XUSN	Auto and Other Motor Vehicle Dealers	Motor Vehicle and Parts Dealers	s	1
2	MRTSSM4413USN	Automotive Parts, Accessory, and Tire Stores	Motor Vehicle and Parts Dealers	s	1
3	MRTSSM4453USN	Beer, Wine, and Liquor Stores	Food and Beverage Stores	s	2
4	MRTSSM451211USN	Book Stores	Sporting Goods, Hobby, Musical Instrument, an...	s	3
5	MRTSSM444USN	Building Materials and Garden Equipment and S...	Building Materials and Garden Equipment and S...	n	4
6	MRTSSM4441USN	Building Materials and Supplies Dealers	Building Materials and Garden Equipment and S...	s	4
7	MRTSSM448USN	Clothing and Clothing Accessory Stores	Clothing and Clothing Accessory Stores	n	5
8	MRTSSM4481USN	Clothing Stores	Clothing and Clothing Accessory Stores	s	5
9	SM45221USN	Department Stores	General Merchandise Stores	s	0

Figura 4.1 Dataframe com *setor_principal_id*

O passo seguinte consistiu na aplicação de um algoritmo hierárquico, conforme ilustrado na Figura 4.2. O objetivo deste procedimento foi criar *clusters* utilizando diversas *features* de séries temporais. Para tal, foram selecionadas a autocorrelação, a força da tendência, a entropia e a força sazonal. É importante notar que todas as séries da seguinte análise incluem apenas dados anteriores a 1 de fevereiro de 2020, o período considerado pré COVID.

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

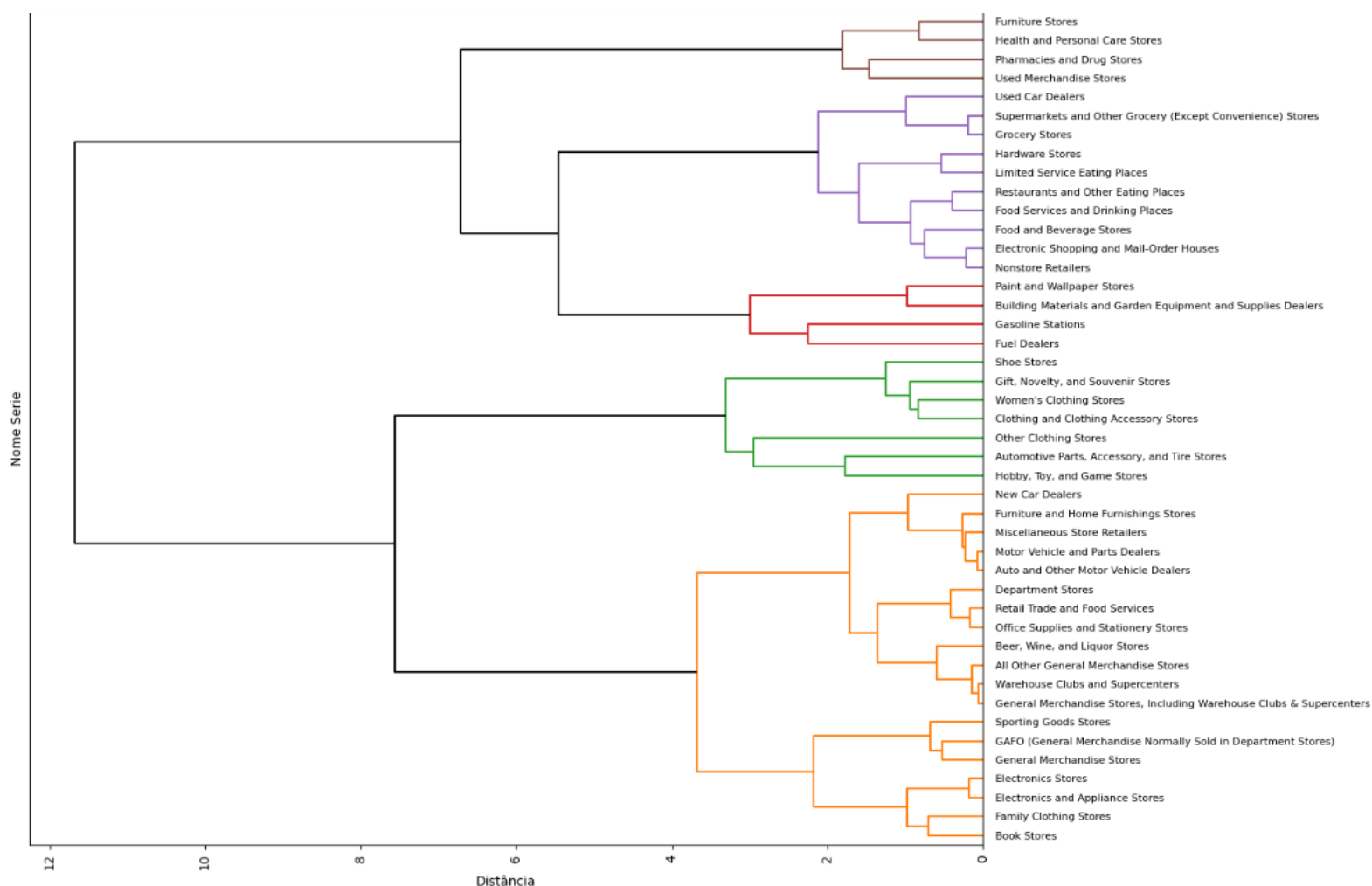


Figura 4.2 Dendrograma clustering hierárquico pré COVID

Efetuada um corte numa distância de quatro vemos que são formados 5 *clusters*. De seguida iremos passar à apresentação e análise de cada *cluster*.

- **Análise Cluster 1-Castanho**

A Tabela 4.1, ilustra a combinação de duas séries de bens essenciais (*Pharmacies and Drug Stores* e *Health and Personal Care Stores*), e bens de consumo duráveis ou de um nicho específico (*Furniture Stores*, *Used Merchandise Stores*). Os setores anteriormente mencionados que, à partida, parecem distintos, mas que partilhavam características estatísticas semelhantes no período pré COVID.

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

unique_id	setor_principal_id	nome
MRTSSM4421USN	7	Furniture Stores
MRTSSM45330USN	11	Used Merchandise Stores
MRTSSM446USN	12	Health and Personal Care Stores
MRTSSM44611USN	12	Pharmacies and Drug Stores

Tabela 4.1 Séries Cluster 1 pré-COVID

Este *cluster* é caracterizado por um comportamento de vendas que combina a estabilidade dos bens essenciais, (*Pharmacies and Drug Stores* e *Health and Personal Care Stores* e duráveis (*Furniture Stores*), com uma forte sazonalidade e uma forte força da tendência. Analisando, a Tabela 4.2, a autocorrelação, o valor é moderado, o que indica que as flutuações a curto prazo ocorriam, mas não eram em elevado número. O valor de entropia é um valor intermédio, o que sugere que as séries não são completamente aleatórias. A força da tendência apresenta, por sua vez, um valor de 0,916, o que indica que as vendas nestes setores tinham um crescimento mensal consistente e previsível. A força da sazonalidade é também elevada, demonstrando que as vendas eram fortemente influenciadas por ciclos anuais ou mensais regulares, como épocas festivas. Setores como o mobiliário e as farmácias têm picos de procura bem definidos.

Estatística	x_acfl	entropy	trend	seasonal_strength
count	4.00	4.00	4.00	4.00
mean	0.46	0.73	0.92	0.90
std	0.09	0.05	0.02	0.02
min	0.36	0.68	0.90	0.87
25%	0.43	0.70	0.90	0.89
50%	0.46	0.72	0.91	0.90
75%	0.50	0.76	0.93	0.91
max	0.58	0.79	0.94	0.92

Tabela 4.2 Estatísticas descritivas cluster 1 pré-COVID

- Análise Cluster 2 -Lilás**

O *cluster 2*, representado na Tabela 4.3 é um grupo em que se caracteriza pela previsibilidade, consistência e séries menos caóticas. É um *cluster* que combina o canal de distribuição mais disruptivo, *e-commerce* (*Electronic Shopping and Mail-Order*

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

Houses), com bens e serviços essenciais na alimentação e restauração (constituído por todas as outras séries).

unique id	setor_principal id	nome
MRTSSM44112USN	1	Used Car Dealers
MRTSSM44511USN	2	Supermarkets and Other Grocery
MRTSSM4451USN	2	Grocery Stores
MRTSSM445USN	2	Food and Beverage Stores
MRTSSM44413USN	4	Hardware Stores
MRTSSM4541USN	6	Electronic Shopping and Mail-Order Houses
MRTSSM454USN	6	Nonstore Retailers
MRTSSM7225USN	8	Restaurants and Other Eating Places
MRTSSM722USN	8	Food Services and Drinking Places
SM72251XUSN	8	Limited Service Eating Places

Tabela 4.3 Séries cluster 2 pré-COVID

Observando a Tabela 4.4, a autocorrelação é forte, indicando que as vendas dependem do valor do mês anterior, o que resulta em flutuações a curto prazo raras. O valor de entropia é baixo, o que torna o padrão de vendas previsível. A força da tendência e a força da sazonalidade são fortes, sugerindo que o comportamento deste *cluster* é regido por padrões anuais muito estáveis.

Estatística	x_acf1	entropy	trend	seasonal_strength
count	10	10	10	10
mean	0.67	0.63	0.97	0.96
std	0.08	0.03	0.02	0.02
min	0.56	0.58	0.92	0.94
25%	0.61	0.62	0.97	0.94
50%	0.68	0.63	0.97	0.96
75%	0.73	0.65	0.98	0.97
max	0.80	0.69	0.99	0.98

Tabela 4.4 Estatísticas descritivas cluster 2 pré-COVID

- Cluster 3- Vermelho**

A definição deste *cluster* é de um grupo de máxima previsibilidade no retalho pré-COVID, constituído pelas séries presentes na Tabela 4.5. É composto por quatro subsetores distintos, dois deles relacionados à habitação (*Paint and Wallpaper Stores*, *Building Materials and Garden Equipment and Supplies Dealers*) e os outros dois com os transportes (*Gasoline Stations*, *Fuel Dealers*).

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

unique_id	setor_principal_id	nome
MRTSSM44412USN	4	Paint and Wallpaper Stores
MRTSSM4444USN	4	Building Materials and Garden Equipment and Supplies Dealers
MRTSSM45431USN	6	Fuel Dealers
MRTSSM447USN	10	Gasoline Stations

Tabela 4.5 Séries cluster 3 pré-COVID

Observando a Tabela 4.6, este *cluster* é caracterizado por uma autocorrelação muito forte e uma baixa entropia, a mais baixa de todos os *clusters*, o que torna o padrão de vendas previsível e o menos aleatório de todos os *clusters*. A sua força sazonal é forte, uma vez que as séries têm períodos sazonais bem definidos, o que faz sentido em materiais de construção e consumo de combustíveis. A força da tendência por sua vez, é alta, que nos indica que o crescimento mensal das vendas para estas séries era rápido ou consistente no período pré-COVID.

Estatística	x_acf1	entropy	trend	seasonal_strength
count	4	4	4	4
mean	0.75	0.46	0.85	0.95
std	0.05	0.05	0.11	0.04
min	0.69	0.41	0.69	0.90
25%	0.72	0.44	0.83	0.94
50%	0.76	0.45	0.89	0.96
75%	0.79	0.47	0.90	0.98
max	0.80	0.53	0.92	0.98

Tabela 4.6 Estatísticas descritivas cluster 3 pré-COVID

- **Cluster 4-Verde**

O cluster 4 é que as séries de vendas, presentes na Tabela 4.7, estão mais estagnadas em termos de crescimento e mais voláteis no curto prazo. Este *cluster* é dominado pelo setor do vestuário (*Clothing and Clothing Accessory Stores*), com quatro séries e inclui os setores de lazer (*Hobby, Toy, and Game Stores e Gift, Novelty, and Souvenir Stores*) e de automóveis (*Automotive Parts, Accessory, and Tire Stores*).

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

unique_id	setor_principal_id	nome
MRTSSM4413USN	1	Automotive Parts, Accessory, and Tire Stores
MRTSSM45112USN	3	Hobby, Toy, and Game Stores
MRTSSM44812USN	5	Women's Clothing Stores
MRTSSM44819USN	5	Other Clothing Stores
MRTSSM4482USN	5	Shoe Stores
MRTSSM448USN	5	Clothing and Clothing Accessory Stores
MRTSSM45322USN	11	Gift, Novelty, and Souvenir Stores

Tabela 4.7 Séries cluster 4 pré-COVID

Na Tabela 4.8 observa-se que a autocorrelação é baixa, sendo o valor mais próximo de zero de todos os *clusters*, indica que as vendas num mês são independentes das vendas do mês anterior, bastante imprevisíveis no curto prazo. A entropia apresenta um valor intermédio a alto, que sugere que as séries eram menos previsíveis na sua estrutura. Por último a força sazonal é elevada, o que sugere que as vendas anuais eram dominadas por ciclos anuais previsíveis (como o natal e as férias). Podemos fazer um balanço geral de que as vendas eram sazonais, com um crescimento moderado.

Estatística	x_acfl	entropy	trend	seasonal_strength
count	7	7	7	7
mean	0.06	0.71	0.62	0.97
std	0.21	0.05	0.09	0.03
min	-0.24	0.65	0.46	0.93
25%	-0.04	0.69	0.57	0.96
50%	0.03	0.70	0.65	0.99
75%	0.17	0.73	0.67	0.99
max	0.38	0.81	0.71	1.00

Tabela 4.8 Estatísticas descritivas cluster 4 pré-COVID

- Análise cluster 5- Laranja**

O *cluster 5* é o que apresenta o maior número de séries no período pré-COVID, representa o núcleo do retalho tradicional (ex: *general merchandise stores*) e de bens com longa duração (ex: *Motor vehicle and parts dealers, furniture and home furnishings stores*). Na Tabela 4.9 podemos observar todas as séries deste cluster.

unique_id	setor_principal_id	nome
MRTSSM452USN	0	General Merchandise Stores
SM45221USN	0	Department Stores
SM452311USN	0	Warehouse Clubs and Supercenters
SM452319USN	0	All Other General Merchandise Stores

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

SM4523USN	0	General Merchandise Stores, Including Warehou...
MRTSSM44111USN	1	New Car Dealers
MRTSSM441USN	1	Motor Vehicle and Parts Dealers
MRTSSM441XUSN	1	Auto and Other Motor Vehicle Dealers
MRTSSM4453USN	2	Beer, Wine, and Liquor Stores
MRTSSM4511USN	3	Sporting Goods Stores
MRTSSM451211USN	3	Book Stores
MRTSSM44814USN	5	Family Clothing Stores
MRTSSM442USN	7	Furniture and Home Furnishings Stores
MRTSSM443142USN	7	Electronics Stores
MRTSSM443USN	7	Electronics and Appliance Stores
MRTSSM44X72USN	9	Retail Trade and Food Services
MRTSSM4400USN	9	GAFO (General Merchandise Normally Sold in Department Stores)
MRTSSM45321USN	11	Office Supplies and Stationery Stores
MRTSSM453USN	11	Miscellaneous Store Retailers

Tabela 4.9 Séries cluster 5 pré-COVID

Na Tabela 4.10, a autocorrelação é baixa, o que implica que as vendas são voláteis e com mudanças rápidas além de não estarem ligadas ao mês anterior. A entropia é alta, que vem reforçar ainda mais que as características deste *cluster* são aleatórias e que não existe um padrão de vendas no curto prazo. A força da tendência de crescimento é forte, impulsionada por um volume de retalho robusto e estável. A força sazonal é também muito alta com as vendas a serem realizadas por ciclos anuais, o que é algo típico de grandes retalhistas.

Estatística	x_acf1	entropy	trend	seasonal_strength
count	19	19	19	19
mean	0.27	0.73	0.91	0.98
std	0.15	0.02	0.07	0.01
min	-0.04	0.69	0.74	0.96
25%	0.18	0.72	0.89	0.98
50%	0.30	0.74	0.92	0.98
75%	0.38	0.74	0.95	0.99
max	0.50	0.77	0.98	1.00

Tabela 4.10 Estatísticas descritivas cluster 5 pré-COVID

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

Devido ao facto de termos quatro *features* e de ser complicado representá-las todas em simultâneo no gráfico optou-se por fazer uma redução de dimensionalidade usando a técnica de PCA (Análise de componentes principais). De acordo com Paparrizos et al., 2024, o PCA é uma medida clássica e bem conhecida para a redução da dimensionalidade. O principal objetivo do PCA é de codificar um conjunto de dados de alta dimensão numa representação de dimensão inferior, mantendo o máximo de informação original possível. Na Figura 4.3 PCA pré COVID podemos ver o PCA para as séries pré-COVID.

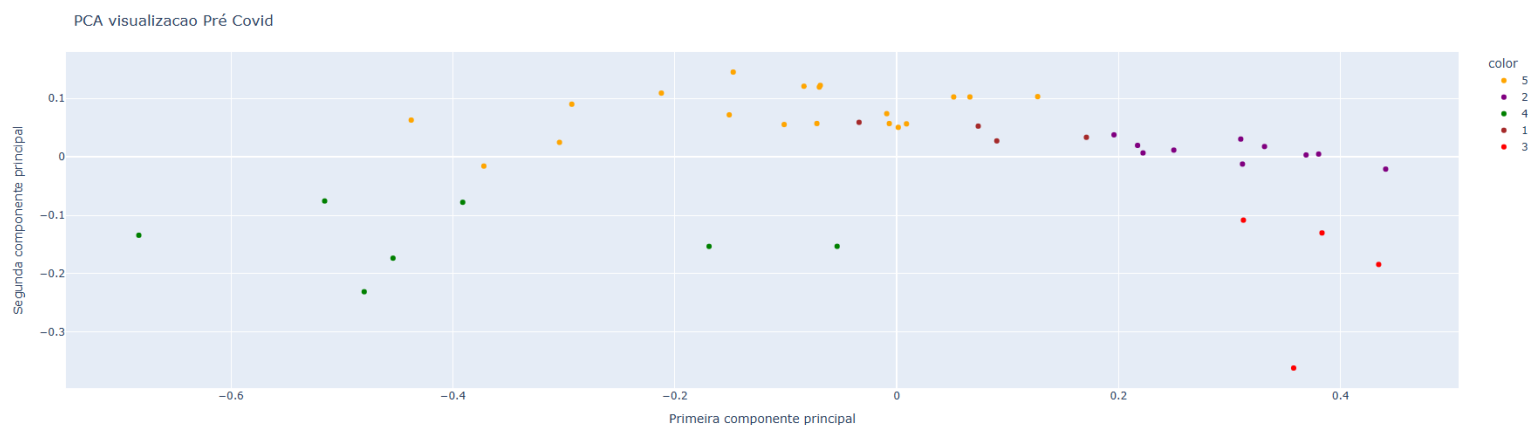


Figura 4.3 PCA pré COVID

Aqui, é possível verificar como as séries se encontram dispersas no espaço e como os *clusters* são formados. A análise PCA ao conjunto de dados do período pré-pandemia revelou que as duas primeiras componentes principais explicam 93,06% da variabilidade dos dados. Esse elevado percentual de variância capturada demonstra que a maior parte das informações presentes nas métricas estatísticas pode ser sintetizada em apenas duas dimensões, sem perda significativa de resolução analítica.

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

4.2 Análise Pós COVID

O dendrograma do *clustering* hierárquico pós-COVID, foi aplicado tendo em conta os valores de todas as séries temporais com data posterior ao dia 1 de maio de 2020. Através do método de ligação *ward*, o objetivo foi agrupar todas as séries com base nas características que foram utilizadas para a análise pré-COVID. Na Figura 4.4 podemos observar o dendrograma para as séries pós COVID.

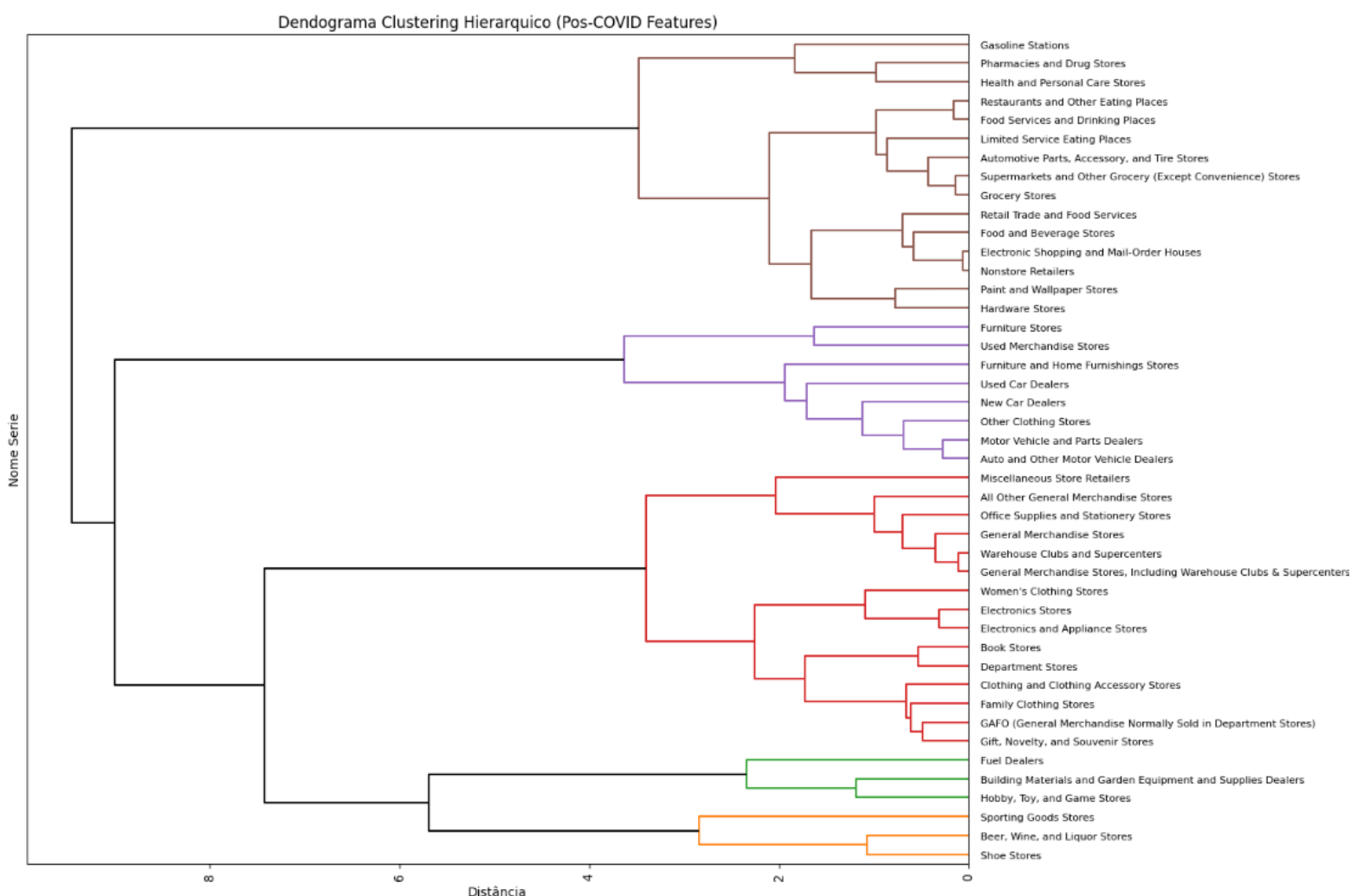


Figura 4.4 Dendrograma clustering pós COVID

A análise resultou em 5 *clusters*, em que foi efetuado um corte pelo valor de distância 4. Este procedimento separou as séries em grupos com características comportamentais distintas, refletindo uma nova dinâmica do mercado do retalho. Nas tabelas seguintes,

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

observamos a nova dinâmica, de *clusters* formados, tendo em conta o comportamento das séries.

- **Análise cluster 1-Castanho**

O primeiro cluster pós-COVID é composto por um enorme número de diferentes setores. Na Tabela 4.11, podemos observar as series saúde/bens essenciais que incluem *Pharmacies and Drug Stores, Health and Personal Care Stores, Supermarkets and Other Grocery, Food and Beverage Stores, Grocery Stores, Gasoline Stations, Retail Trade and Food Services*. Integra também bens duráveis/mobiliário, nomeadamente *Automotive Parts, Accessory, and Tire Stores, Hardware Stores, Paint and Wallpaper Stores*. Também tem presente o novo mercado de *e-commerce* com *Electronic Shopping and Mail-Order Houses, Nonstore Retailers*). Por último o setor da restauração que tem presente *Restaurants and Other Eating Places, Food Services and Drinking Places, Limited Service Eating Places*. Estes são, setores foram massivamente utilizados no período pós pandemia e que tiveram a sua importância na economia refletem a complexa adaptação do consumo após o choque inicial e uma reorganização das prioridades do consumidor.

unique_id	setor_principal_id	nome
MRTSSM4413USN	1	Automotive Parts, Accessory, and Tire Stores
MRTSSM44412USN	4	Paint and Wallpaper Stores
MRTSSM44413USN	4	Hardware Stores
MRTSSM44511USN	2	Supermarkets and Other Grocery
MRTSSM4451USN	2	Grocery Stores
MRTSSM445USN	2	Food and Beverage Stores
MRTSSM4541USN	6	Electronic Shopping and Mail-Order Houses
MRTSSM454USN	6	Nonstore Retailers
MRTSSM7225USN	8	Restaurants and Other Eating Places
MRTSSM722USN	8	Food Services and Drinking Places
SM72251XUSN	8	Limited Service Eating Places
MRTSSM44X72USN	9	Retail Trade and Food Services
MRTSSM447USN	10	Gasoline Stations
MRTSSM44611USN	12	Pharmacies and Drug Stores
MRTSSM446USN	12	Health and Personal Care Stores

Tabela 4.11 Setores cluster 1 pós-COVID

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

As estatísticas descritivas que se observam na Tabela 4.12 permitem caracterizar este *cluster* com uma força da tendência alta, que indica que as séries apresentam valores de venda robustos e lineares. A autocorrelação também forte, indica esta persistência no consumo, assim como uma baixa entropia, que demonstra que o padrão de venda das séries era previsível. A força sazonal também é elevada, mas sem atingir valores demasiado extremos, que reflete a estabilidade dos bens essenciais. Pode-se, portanto, afirmar que este grupo foi o dos mais previsíveis e estáveis após a pandemia.

Este comportamento muito se deve à agregação de series essenciais como (mercearias, farmácias), que são fundamentais. Assim como o crescimento de novos canais, como é o caso do *e-commerce* para a restauração.

Estatística	x_acf1	entropy	trend	seasonal_strength
count	15	15	15	15
mean	0.76	0.55	0.97	0.93
std	0.08	0.04	0.02	0.05
min	0.67	0.49	0.90	0.83
25%	0.70	0.52	0.97	0.91
50%	0.74	0.55	0.98	0.94
75%	0.82	0.58	0.99	0.96
max	0.87	0.64	0.99	0.98

Tabela 4.12 Estatísticas descritivas cluster 1 pós-COVID

- Análise cluster 2 -Lilás**

No *cluster 2*, vemos uma preponderância de bens do setor automóvel (*New Car Dealers, vehicle and parts dealers, Used Car Dealers, Auto and Other Motor Vehicle Dealers, Motor Vehicle and Parts Dealers*) setor do mobiliário (*Furniture Stores e Furniture and Home Furnishings Stores*), por último dois setores de nicho (*Other Clothing Stores, Used Merchandise Stores*). Estes são grupos muito sazonais e sensíveis às flutuações na economia. Encontram-se representados na Tabela 4.13.

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

unique_id	setor_principal_id	nome
MRTSSM44111USN	1	New Car Dealers
MRTSSM44112USN	1	Used Car Dealers
MRTSSM441USN	1	Motor Vehicle and Parts Dealers
MRTSSM441XUSN	1	Auto and Other Motor Vehicle Dealers
MRTSSM44819USN	5	Other Clothing Stores
MRTSSM4421USN	7	Furniture Stores
MRTSSM442USN	7	Furniture and Home Furnishings Stores
MRTSSM45330USN	11	Used Merchandise Stores

Tabela 4.13 Setores cluster 2 pós-COVID

Este grupo, que tem as suas estatísticas descritivas representadas na Tabela 4.14, é caracterizado por uma baixa força de tendência para os *clusters* pós-COVID, assim como uma entropia alta. Isto reflete a dificuldade em manter um padrão de vendas claro e forte, devido a fatores como a escassez que afetou toda a cadeia de abastecimento e a inflação que provocou uma volatilidade nos preços. Sendo a entropia elevada e a autocorrelação moderada, pode-se concluir que, mesmo com uma tendência elevada, é complicado encontrar previsão para futuras vendas das séries.

Estatística	x_acf1	entropy	trend	seasonal_strength
count	8	8	8	8
mean	0.54	0.75	0.87	0.84
std	0.11	0.04	0.04	0.06
min	0.36	0.64	0.79	0.75
25%	0.49	0.75	0.85	0.81
50%	0.58	0.76	0.89	0.86
75%	0.60	0.77	0.90	0.89
max	0.69	0.79	0.92	0.89

Tabela 4.14 Estatísticas descritivas cluster 2 pós- COVID

- Análise cluster 3- Vermelho**

Analisando o *cluster* 3 podemos ver que ele depende maioritariamente de séries não essenciais. Na Tabela 4.15, observamos series relacionadas com o vestuário (*Clothing and Clothing Accessory Stores, Women's Clothing Stores, Family Clothing Stores*), presentes e lazer (*Gift, Novelty, and Souvenir Stores, Book Stores*) e outros nichos como (*Office Supplies and Stationery Stores e Miscellaneous Store Retailers*). Muito destas vendas são relacionadas com forte marketing e uma também elevada sazonalidade (como as lojas de vestuário, livrarias e lojas de presentes).

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

unique_id	setor_principal_id	nome
SM452319USN	0	All Other General Merchandise Stores
SM45221USN	0	Department Stores
MRTSSM452USN	0	General Merchandise Stores
SM4523USN	0	General Merchandise Stores, Including Warehouses
SM452311USN	0	Warehouse Clubs and Supercenters
MRTSSM451211USN	3	Book Stores
MRTSSM448USN	5	Clothing and Clothing Accessory Stores
MRTSSM44814USN	5	Family Clothing Stores
MRTSSM44812USN	5	Women's Clothing Stores
MRTSSM443142USN	7	Electronics Stores
MRTSSM443USN	7	Electronics and Appliance Stores
MRTSSM4400USN	9	GAFO (General Merchandise Normally Sold in Department Stores)
MRTSSM453USN	11	Miscellaneous Store Retailers
MRTSSM45322USN	11	Gift, Novelty, and Souvenir Stores
MRTSSM45321USN	11	Office Supplies and Stationery Stores

Tabela 4.15 Setores cluster 3 pós-COVID

Como referido anteriormente neste *cluster* observa-se uma força sazonal muito elevada, a qual demonstra que as vendas são dominadas por momentos sazonais, por exemplo épocas festivas. Assim como a sua força da tendência é robusta, indicando que o volume de vendas recuperou após o choque do COVID-19. No entanto a sua autocorrelação baixa e entropia moderada indicam que as vendas de um mês para o outro não estão relacionadas e que existem volatilidades e flutuações rápidas. Consequentemente tornam difícil de prever as vendas no curto prazo. Na Tabela 4.16 podemos observar as suas estatísticas descritivas.

Estatística	x_acf1	entropy	trend	seasonal strength
count	15	15	15	15
mean	0.35	0.67	0.93	0.97
std	0.14	0.06	0.04	0.02
min	0.15	0.57	0.85	0.93
25%	0.25	0.63	0.91	0.96
50%	0.35	0.67	0.94	0.97
75%	0.44	0.71	0.97	0.98
max	0.64	0.75	0.97	0.98

Tabela 4.16 Estatísticas descritivas cluster 3 pós-COVID

- **Análise cluster 4-Verde**

O *cluster* 4, tem um alto contraste composto por bens de necessidade (*Fuel Dealers*) e os bens de consumo doméstico (*Building Materials and Garden Equipment and Supplies*

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

Dealers) e bens de lazer, que não são essenciais e de momentos sazonais (*Hobby, Toy, and Game Stores*). Na Tabela 4.17 estão representadas as suas séries.

unique id	setor_principal id	nome
MRTSSM45112USN	3.0	Hobby, Toy, and Game Stores
MRTSSM444USN	4.0	Building Materials and Garden Equipment and Supplies Dealers
MRTSSM45431USN	6.0	Fuel Dealers

Tabela 4.17 Setores cluster 4 pós-COVID

Este *cluster*, conforme pode ser observado na tabela 4.18 está focado em áreas de consumo sazonal, que faz com que o seu valor da força sazonal seja muito elevado, o que é natural devido ao facto de estar ligado a séries como *fuel dealers*, e à volatilidade deste mercado, e de materiais de construção. O crescimento pode ser verificado pelo valor da força da tendência, que é um valor muito alto, indicando o baixo crescimento das vendas após maio de 2020. Relativamente à autocorrelação, esta apresenta um valor forte, que indica que as vendas num mês, estão fortemente ligadas às do mês anterior, realçando a maior inércia ou falta de consumo. Por último, a entropia deste cluster é baixa, o que nos indica que a série é muito previsível.

Estatística	x_acf1	entropy	trend	seasonal_strength
count	3	3	3	3
mean	0.64	0.41	0.81	0.97
std	0.19	0.10	0.04	0.01
min	0.44	0.30	0.76	0.96
25%	0.55	0.38	0.79	0.97
50%	0.66	0.46	0.82	0.97
75%	0.74	0.47	0.83	0.98
max	0.82	0.47	0.83	0.99

Tabela 4.18 Estatísticas descritivas cluster 4 pós-COVID

- Análise cluster 5-Laranja**

O *cluster 5*, que tem as suas séries representadas na Tabela 4.19, é composto pelas séries: bebidas alcoólicas (*Beer, Wine, and Liquor Stores*), calçado (*Shoe Stores*) e artigos desportivos (*Sporting Goods Stores*). Tal como no *cluster* anterior este também é formado por séries bastante diferentes entre si, e muito dependentes de eventos sazonais.

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

unique_id	setor_principal_id	nome
MRTSSM4453USN	2.0	Beer, Wine, and Liquor Stores
MRTSSM4511USN	3.0	Sporting Goods Stores
MRTSSM4482USN	5.0	Shoe Stores

Tabela 4.19 Setores cluster 5 pós COVID

Analisando a Tabela 4.20, observamos um baixo valor da força da tendência. A força da sazonalidade é alta, muito devido às vendas de calçado desportivo e bebidas que são fortemente influenciadas por mudanças sazonais. Este *cluster* também apresenta uma alta volatilidade, que se reflete no valor da autocorrelação ser praticamente nulo/ negativo, o que sugere que as vendas são bastante imprevisíveis no curto prazo. A entropia, por sua vez, apresenta um valor elevado. No entanto, após a remoção da forte sazonalidade, a série torna-se previsível devido à sua rigidez em certas épocas do ano apesar da elevada volatilidade que apresenta de mês para mês.

Estatística	x_acf1	entropy	trend	seasonal_strength
count	3	3	3	3
mean	-0.03	0.59	0.75	0.97
std	0.03	0.05	0.11	0.02
min	-0.05	0.55	0.62	0.96
25%	-0.04	0.57	0.71	0.97
50%	-0.04	0.58	0.80	0.97
75%	-0.02	0.61	0.81	0.98
max	0.00	0.64	0.82	0.99

Tabela 4.20 Estatísticas descritivas cluster 5 pós-COVID

À semelhança dos dados pré-COVID, também para as séries pós- COVID procedemos a uma redução de dimensão por PCA, conforme se encontra na Figura 4.5. Aqui observamos os 5 *clusters* formados no espaço, em que, tal como no agrupamento anterior, apesar de algumas séries se encontrarem distantes em termos de representação espacial, reuniam características estatísticas semelhantes, o que permitiu que formassem um *cluster* coeso com séries dispersas no espaço. Neste caso também se pode observar uma variabilidade de 93,06% das duas componentes principais.

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

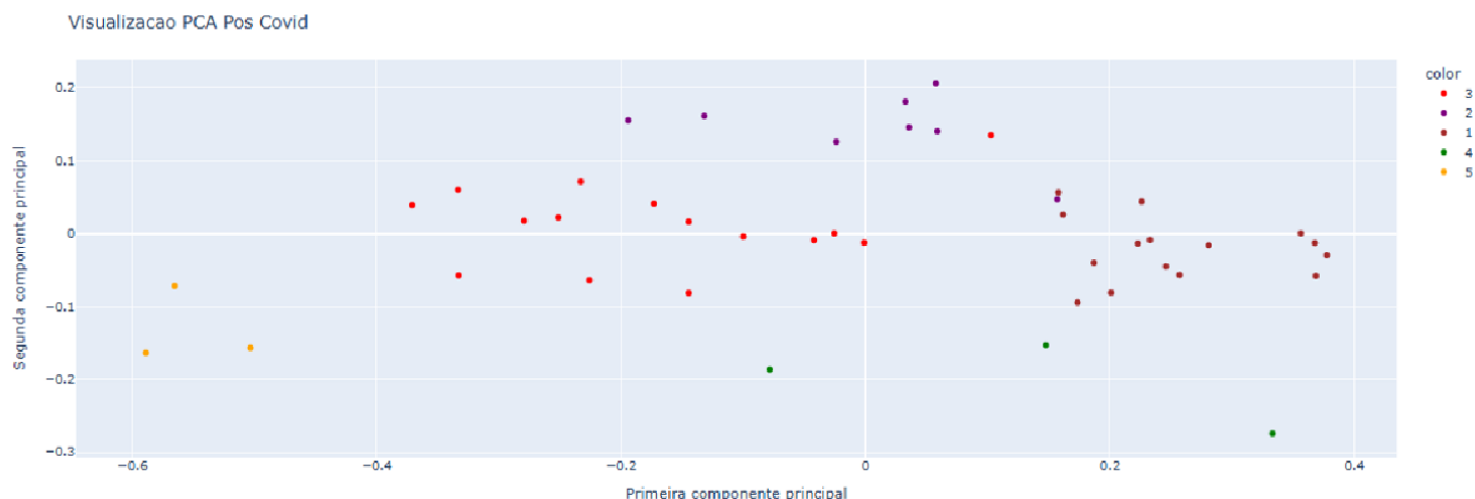


Figura 4.5 PCA pós-COVID

4.3 Análise comparativa: Pré-covid e Pós-covid

No período pré-COVID, os *clusters* eram definidos por padrões de consumo bem estabelecidos, muitas vezes manifestando uma lógica pela categoria da série. Começando pelo *cluster 1*, este apresentava séries, muito estáveis e de consumo moderado com uma procura segura (como os setores da saúde). Após a COVID-19, observou-se que as séries dentro deste grupo reagiram de forma oposta. As séries de saúde *Pharmacies and Drug Stores* e *Health and Personal Care Store* mantiveram-se no mesmo *cluster*. Já as séries de *furniture stores* e de *used merchandise stores* migraram do *cluster 1* para o *cluster 2*. Esta migração, visível na Tabela 4.21, é uma evidencia da reorganização das prioridades de consumo.

Esta reorganização, em conjunto com a análise da entropia no pós-COVID, para as qual os valores do *cluster 1* e *2* pós-COVID, apresentam valores de entropia respetivamente de 0,55 e de 0,74. O valor da entropia é bastante superior no *cluster 2*, esta disparidade faz com que se possa afirmar que as séries *furniture stores* e *used merchandise stores* aumentaram o seu nível de imprevisibilidade, distanciando-se das outras duas séries da área da saúde.

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

nome	setor_principal_id	cluster_pre_COVID	cluster_pos_COVID
Furniture Stores	7	1	2
Used Merchandise Stores	11	1	2
Pharmacies and Drug Stores	12	1	1
Health and Personal Care Stores	12	1	1

Tabela 4.21 Cluster pré COVID 1 e o pós-COVID

O *cluster 2*, no período pré-COVID, agrupava setores de consumo diário estável como mercearias e serviços de restauração, lojas de bens de manutenção e *e-commerce*. Após a COVID-19, viu grande parte das suas séries transitar para o *cluster 1*, um *cluster* de bens essenciais que viu o seu aglomerado de séries reforçado. Podemos afirmar que, para o consumidor, estes setores de atividade ganharam uma maior relevância depois do choque da COVID-19. A única série, como se pode observa na Tabela 4.22, que não sofreu alteração foi a série de *used car dealers* que se manteve no *cluster 2*.

nome	setor_principal_id	cluster_pre_covid	cluster_pos_covid
Used Car Dealers	1	2	2
Food and Beverage Stores	2	2	1
Grocery Stores	2	2	1
Supermarkets and Other Grocery	2	2	1
Hardware Stores	4	2	1
Electronic Shopping and Mail-Order Houses	6	2	1
Nonstore Retailers	6	2	1
Food Services and Drinking Places	8	2	1
Limited Service Eating Places	8	2	1
Restaurants and Other Eating Places	8	2	1

Tabela 4.22 Cluster pré COVID 2 e o pós-COVID

Analisando o *cluster 3* pré-COVID, este é caracterizado por setores ligados aos combustíveis (*Gasoline Stations* e *Fuel Dealers*) e setores ligados ao imobiliário (*Paint and Wallpaper Stores*, *Building Materials* and *Garden Equipment Stores*).

As séries presentes na Tabela 4.23 *Paint and Wallpaper Stores* e *Gasoline Stations* migraram para o *cluster 1*. A entropia para estes setores aumentou, o setor de *Paint and*

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

Wallpaper Stores passou de uma entropia de 0,45 para 0,51 e a série *gasoline stations* de 0,45 para 0,49. Apesar disto no panorama pós-COVID, eles têm mais semelhanças com a entropia do *cluster 1* do que do *cluster 3* (respetivamente 0,55 e 0,67) como tal, foram para um *cluster* mais estável.

Observando as outras duas séries, são compostas por valores de força de sazonalidade muito elevados, analisando os dados pós-COVID continuam com valores de 0,97 para *Building Materials and Garden Equipment Stores* e de 0,96 para *Fuel Dealers*. No cenário pós-COVID, o *cluster 3* apresenta uma força de sazonalidade inferior à do *cluster 4*, portanto as séries agruparam-se ao *cluster 4*.

nome	setor_principal_id	cluster_pre_COVID	cluster_pos_COVID
Paint and Wallpaper Stores	4	3	1
Building Materials and Garden Equipment and S...	4	3	4
Fuel Dealers	6	3	4
Gasoline Stations	10	3	1

Tabela 4.23 Cluster pré-COVID 3 e o pós COVID

No *cluster 4*, representado na Tabela 4.23 verifica-se que quase todas as séries mudaram de *cluster*, com a exceção da série *Hobby, Toy, and Game Stores* que se manteve presente.

Verificou-se, através do valor elevado da força sazonal, que as séries presentes no *cluster 4* são muito dependentes de picos de vendas. A alta força da sazonalidade estatística torna o setor vulnerável porque qualquer quebra durante um pico de vendas (como aconteceu no Natal de 2020) causa uma perda de receita que não pode ser recuperada facilmente.

Analisando o setor do vestuário (setor principal 5) presente no *cluster 4*, todos os setores mudaram de *cluster*. Devido à alteração de comportamentos dos consumidores no pós-COVID estas séries verificaram uma diminuição de valores da força sazonal e entropia e um aumento da força da tendência.

Outra série que ficou mais robusta foi a *Automotive Parts, Accessory, and Tire Stores* que mudou para um *cluster* com menor força sazonal e entropia e uma força da tendência de valor mais elevado.

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

Por último a série *Gift, Novelty, and Souvenir Stores* mudou também para o *cluster* 3. O que mais salta à vista é o aumento da força da tendência que passou de 0,70 para 0,92. Observando os seus valores de vendas, vemos um aumento depois dos meses de recessão, e muito deste crescimento pode estar associado a um aumento do turismo a nível mundial.

nome	setor principal id	cluster pre COVID	cluster pos COVID
Automotive Parts, Accessory, and Tire Stores	1	4	1
Hobby, Toy, and Game Stores	3	4	4
Clothing and Clothing Accessory Stores	5	4	3
Other Clothing Stores	5	4	2
Shoe Stores	5	4	5
Women's Clothing Stores	5	4	3
Clothing and Clothing Accessory Stores	5	4	3
Gift, Novelty, and Souvenir Stores	11	4	3

Tabela 4.24 Cluster pré-COVID 4 e o pós-COVID

O *cluster* 5 pré-COVID, cujas séries se encontram discriminadas na Tabela 4.25, distingue-se por ser o *cluster* mais heterogéneo e com um maior número de setores. Engloba setores do retalho (*Retail Trade and Food Services*), do setor automóvel (*Motor Vehicle and Parts Dealers*), do lazer (*Book stores*) e das bebidas (*Beer, Wine, and Liquor Stores*). As suas estatísticas apresentam uma força sazonal muito elevada, uma força da tendência acentuada, uma baixa autocorrelação e uma entropia moderada.

A observação do período pós-COVID revela uma desagregação deste agrupamento em quatro *clusters* diferentes (1,2,3 e 5).

A série de *Retail Trade and Food Services*, transitou para o *cluster* 1, onde sofreu um aumento na autocorrelação, de 0,42 para 0,67. Este incremento conferiu uma maior estabilidade nos valores de vendas, evidenciando a importância do setor da alimentação na economia. As séries de automóvel (*New Car Dealers, Motor Vehicle and Parts Dealers*) e mobiliário (*Furniture and Home Furnishings Stores*) migraram para o *cluster* 2. Para as séries de automóveis, a análise para o período pós-COVID observa uma diminuição da força sazonal e um aumento da autocorrelação. Isto sugere que estas séries passaram a ser mais previsíveis e menos dependentes de fatores sazonais. Os setores que

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

alteraram para o *cluster 3* são na sua maioria setores tradicionais de retalho, como *Book Stores* e *Electronics Stores*. A grande alteração de valores na entropia fez com que estes setores se aproximassem do *cluster 3*, indicando que este grupo de retalhistas ficou mais vulnerável em termos de vendas devido à ausência de tráfego nas lojas físicas. Por último, as séries de *Beer, Wine, and Liquor Stores* e *Sporting Goods Stores* mantiveram-se no *cluster 5*. As suas *features* não sofreram perturbações suficientes para justificar uma mudança de *cluster*, mantendo-se a força sazonal elevada. Em oposição, os valores entropia e de autocorrelação eram muito baixos. A *feature* de autocorrelação apresenta valores negativos na série *sporting goods stores*. Esta *feature* apresenta valores muito baixos quando comparada com as outras séries.

nome	setor_principal_id	cluster_pre_COVID	cluster_pos_COVID
All Other General Merchandise Stores	0	5	3
Department Stores	0	5	3
General Merchandise Stores	0	5	3
General Merchandise Stores, Including Warehouse clubs & supercenters	0	5	3
Warehouse Clubs and Supercenters	0	5	3
Auto and Other Motor Vehicle Dealers	1	5	2
Motor Vehicle and Parts Dealers	1	5	2
New Car Dealers	1	5	2
Beer, Wine, and Liquor Stores	2	5	5
Book Stores	3	5	3
Sporting Goods Stores	3	5	5
Family Clothing Stores	5	5	3
Electronics Stores	7	5	3
Electronics and Appliance Stores	7	5	3
Furniture and Home Furnishings Stores	7	5	2
GAFO (General Merchandise Normally Sold in Department stores)	9	5	3
Retail Trade and Food Services	9	5	1
Miscellaneous Store Retailers	11	5	3
Office Supplies and Stationery Stores	11	5	3
Miscellaneous Store Retailers	11	5	3

Tabela 4.25 Cluster pré COVID 5 e o pós COVID

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

CONCLUSÃO

O objetivo deste trabalho consistiu em analisar o impacto da COVID-19 no setor do retalho dos Estados Unidos, através da aplicação da metodologia de *clustering* de séries temporais comparando o agrupamento estatístico antes de depois do COVID-19. Esta análise consistiu na aplicação das *features* de força da tendência, força da sazonalidade, entropia e autocorrelação às séries de vendas a retalho. Posteriormente, foi aplicado um algoritmo de *clustering* hierárquico para agrupar as séries com estatísticas semelhantes em *clusters* para os períodos pré e pós-COVID.

O estudo demonstrou que a COVID-19 não apenas afetou os níveis de vendas, mas também reorganizou as relações estatísticas entre os subsectores do retalho. Esta reorganização observada no período pós-COVID, revelou as mudanças de comportamento dos consumidores nos mais variados setores. Uma das alterações mais significativas no pós-COVID é a diminuição da força sazonal em média ao compararmos os valores pré-COVIDe pós-COVID. Esta atenuação permite afirmar que o ruído e a entropia ganharam proeminência no pós-COVID. Consequentemente, a variação das vendas passou a ser impulsionada por menos previsibilidade e maior aleatoriedade inerentes ao impacto da COVID-19. A grande maioria das séries transitou para um *cluster* diferente no pós-COVID, evidenciando setores que anteriormente não eram essenciais, que estavam em *clusters* com maior entropia.

As alterações das séries de *cluster* foram predominantemente causadas por uma força da tendência elevada indicando o crescimento do volume de vendas em valor. A baixa entropia e a força sazonal forte refletiu as séries que eram de maior necessidade para o consumidor. Essas séries transitaram para o *cluster* 1, que estatisticamente era o mais estável e continha as séries da saúde. Exemplos da migração incluem a série *Electronic Shopping and Mail-Order Houses* com o aumento das entregas de produtos em casa, e a série *Gasoline Stations* com a elevada subida dos preços do combustível, que fez aumentar as suas vendas em volume. Estas mudanças fundamentam o objetivo deste estudo e sublinham a importância que a COVID-19 teve na economia, evidenciando na necessidade de um estudo bem aprofundado para esta nova dinâmica.

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

O presente trabalho oferece uma nova perspetiva sobre o impacto da COVID-19, mostrando que o retalho se segmentou em blocos de risco e estabilidade que transcendem a classificação NAICS. Esta nova métrica de análise económica permite identificar setores mais estáveis e menos estáveis em futuras quebras económicas e oportunidade para se prevenirem.

Este trabalho não está isento de limitações. Na escolha do algoritmo, outro poderia ter sido escolhido, como o *k-means* que oferece uma possibilidade de uma maior manipulação de dados. Além disso, as *features* que foram utilizadas para a realização deste estudo, embora adequadas, podem ser complementadas por outras disponíveis na biblioteca *tsfeatures*.

Para trabalhos futuros, sugere-se a utilização de diferentes *features*, que permitam outro tipo de análises. Adicionalmente, dado o carácter contínuo e evolutivo da COVID-19, adicionar novos dados posteriores a abril de 2025. Por fim, o desenvolvimento de modelos de previsão pode ser uma oportunidade para testar a previsibilidade das séries e dos seus *clusters*.

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

REFERÊNCIAS BIBLIOGRÁFICAS

- Arlandis, V. (n.d.). *E-commerce: a revolução do comércio no mundo digital*. Lisbon Digital School. <https://lisbondigitalschool.com/blog/e-commerce-a-revolucao-do-comercio-no-mundo-digital/>
- Biggs, C., Shavdia, D., Kelley, A., Mendu, V., & Avasare, A. (2020). Retail in the Eye of the Storm. In *Boston Consulting Group*. <https://www.bcg.com/publications/2020/long-term-solutions-for-retailers-through-pandemic>
- Boyko, N. I., & Tkachyk, O. A. (2023). Hierarchical clustering algorithm for dendrogram construction and cluster counting. *Informatics and Mathematical Methods in Simulation*, 13(1–2), 5–15. <https://doi.org/10.15276/imms.v13.no1-2.5>
- Burgos, D., & Ivanov, D. (2021). Food retail supply chain resilience and the COVID-19 pandemic: A digital twin-based impact analysis and improvement directions. *Transportation Research Part E: Logistics and Transportation Review*, 152. <https://doi.org/10.1016/j.tre.2021.102412>
- Chapman, P., Clinton, J., & Kerber, R. (2000). *CRISP-DM 1.0*. DaimlerChrysler.
- Choudhury, N. R. (2025). *Food and Grocery Retail Market Analysis - Size, Share, and Forecast Outlook 2025 to 2035*. <https://www.futuremarketinsights.com/reports/food-and-grocery-retail-market>
- Cohen, Mark. (2018). Sears: A Case Study in Business Failure. *Chazen Global Insights-Columbia Business School*. <https://business.columbia.edu/insights/chazen-global-insights/sears-case-study-business-failure>
- Dincer, C., & Dincer, B. (2023). Social Commerce and Purchase Intention: A Brief Look at the Last Decade by Bibliometrics. *Sustainability (Switzerland)*, 15(1). <https://doi.org/10.3390/su15010846>
- Dix, P. (2025). *Time Series Data Analysis Definitions & Best Techniques for 2025*. <https://www.influxdata.com/what-is-time-series-data/#faq>

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

- Ferraz, M. (2022). *The Impact of Covid-19 and Accelerated Growth of e-commerce: The Insignia Case*.
- Gama, J., Carvalho, A. P. de L., Faceli, K., Lorena, A. C., & Oliveira, M. (2017). *Extração de Conhecimento de Dados: Data Mining (3.a)*. Edições Sílabo.
- Greg Petro. (2020). Coronavirus And Shopping Behavior: Men And Women React Differently. *Forbes*.
<https://www.forbes.com/sites/gregpetro/2020/03/13/coronavirus-and-shopping-behavior-men-and-women-react-differently/>
- Helm, S., Kim, S. H., & Van Riper, S. (2020). Navigating the ‘retail apocalypse’: A framework of consumer evaluations of the new retail landscape. *Journal of Retailing and Consumer Services*, 54, 101683.
<https://doi.org/10.1016/j.jretconser.2018.09.015>
- Houvèssou, G. M., Souza, T. P. de, & Silveira, M. F. da. (2021). Medidas de contenção de tipo lockdown para prevenção e controle da COVID-19: estudo ecológico descritivo, com dados da África do Sul, Alemanha, Brasil, Espanha, Estados Unidos, Itália e Nova Zelândia, fevereiro a agosto de 2020. *Epidemiologia e Serviços de Saúde*, 30(1). <https://doi.org/10.1590/s1679-49742021000100025>
- International Trade Administration U.S. Department of Commerce. (2025). *Retail Trade Industry*. <https://www.trade.gov/selectusa-retail-trade-industry>
- Ismail, A., Bora, S., & Uslu, A. (2025). *US Ecommerce Statistics*. <https://analyzify.com/statsup/us-ecommerce#stat-desktop-pcs-used-in-online-shopping-in-the-us>
- Kang, Y., Hyndman, R. J., & Smith-Miles, K. (2017). Visualising forecasting algorithm performance using time series instance spaces. *International Journal of Forecasting*, 33(2), 345–358. <https://doi.org/10.1016/j.ijforecast.2016.09.004>
- Kong, X., Chen, Z., Liu, W., Ning, K., Zhang, L., Muhammad Marier, S., Liu, Y., Chen, Y., & Xia, F. (2025). Deep learning for time series forecasting: a survey.

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

- International Journal of Machine Learning and Cybernetics*, 16(7–8), 5079–5112.
<https://doi.org/10.1007/s13042-025-02560-w>
- Laato, S., Islam, A. K. M. N., Farooq, A., & Dhir, A. (2020). Unusual purchasing behavior during the early stages of the COVID-19 pandemic: The stimulus-organism-response approach. *Journal of Retailing and Consumer Services*, 57. <https://doi.org/10.1016/j.jretconser.2020.102224>
- Liu, C., & Lipsman, A. (2020). US Ecommerce 2020 Coronavirus Boosts Ecommerce Forecast and Will Accelerate Channel-Shift. *Emarketer*. <https://www.emarketer.com/content/us-ecommerce-2020>
- Martin, J. (2025). *Social commerce surges as 1 in 3 young people now shop on social media every week.* Exploding Topics. <https://www.emarketer.com/learningcenter/guides/social-commerce-brand-trends-marketing-strategies/>
- Montero-Manso, P. (2018). *Methodology: Combination of Forecast Methods by Feature-based Learning*. https://htmlpreview.github.io/?https://github.com/robjhyndman/M4metalearning/blob/master/docs/M4_methodology.html#features
- Nokes, C., & Henshaw, R. (2025). *U.S. Retail Insights in 2025*. <https://www.occstrategy.com/usa/article/u-s-retail-insights-in-2025/>
- Paparrizos, J., Yang, F., & Li, H. (2024). *Bridging the Gap: A Decade Review of Time-Series Clustering Methods*. <http://arxiv.org/abs/2412.20582>
- Rahman, M. M., Thill, J. C., & Paul, K. C. (2020). COVID-19 pandemic severity, lockdown regimes, and people's mobility: Early evidence from 88 countries. *Sustainability (Switzerland)*, 12(21), 1–17. <https://doi.org/10.3390/su12219101>
- Rai, P., Singh, S., & Prof, A. (2010). A Survey of Clustering Techniques. In *International Journal of Computer Applications* (Vol. 7, Issue 12).

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

- Redda, E. H. (2021). Initial impact assessment of Covid-19 on retailing: Changing consumer behaviour and retail trade sales. *Journal of Contemporary Management*, 18(2), 22–41. <https://doi.org/10.35683/jcm21033.117>
- Roman, S., Cooke-Hull, S., Dunfee, M., Flaherty, M., Haskell, J., Holland, V., Jackson, C., & Shevlin, C. (2022). *The Coronavirus Pandemic's Economic Impact*. www.census.gov/retail/how_surveys_are_collected.html
- Sayyida, S., Hartini, S., Gunawan, S., & Husin, S. N. (2021). The Impact of the Covid-19 Pandemic on Retail Consumer Behavior. *Aptisi Transactions on Management (ATM)*, 5(1), 79–88. <https://doi.org/10.33050/atm.v5i1.1497>
- Sousa, A. (2023). *Impacto da sustentabilidade no comportamento de compra do consumidor*.
- Statista Research Department. (2023). *Coronavirus: impact on the retail industry worldwide - Statistics & Facts*. . <https://www.statista.com/topics/6239/coronavirus-impact-on-the-retail-industry-worldwide/>
- Stephens, B., Chafetz, S., & Cole, M. (2025). *2024 Holiday Retail Recap: Unwrapping the Results of the Season*.
- Steven Begley, Becca Coggins, Matthew Maloney, & Steve Noble. (2020). The next normal in retail: Charting a path forward. *McKinsey & Company*. <https://www.mckinsey.com/industries/retail/our-insights/the-next-normal-in-retail-charting-a-path-forward>
- University of Texas. (2017). *What Is Retail Marketing?* <https://online.uttyler.edu/degrees/business/mba/marketing/what-is-retail-marketing/>
- U.S. Census Bureau. (2024). *United States Sensos Bureau*.
- Vogl, M. (2023). Hurst exponent dynamics of S&P 500 returns: Implications for market efficiency, long memory, multifractality and financial crises predictability by application of a nonlinear dynamics analysis framework. *Chaos, Solitons & Fractals*, 166, 112884. <https://doi.org/10.1016/j.chaos.2022.112884>

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

Wright, O., Weiss, E., Leary, B., Blackburn, E., & Ducarre, J. (n.d.). *Me, my brand and AI: The new world of consumer engagement Building resilient relationships between consumers, brands and AI in times of uncertainty.*

Xu, R., & Wunsch, D. (2005). Survey of clustering algorithms. In *IEEE Transactions on Neural Networks* (Vol. 16, Issue 3, pp. 645–678).
<https://doi.org/10.1109/TNN.2005.845141>

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

APÊNDICES

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

APÊNDICE 1. *Features* disponíveis na biblioteca *tsfeatures* do Python

A lista de *features* que seguidamente se apresenta foi adaptada de (Montero-Manso, 2018).

<i>Feature</i>	Definição
<i>unique_id</i>	Identificador único para cada série temporal.
<i>hurst</i>	O expoente de Hurst, que mede a dependência de longo prazo da série temporal. Um valor entre 0.5 e 1.0 indica dependência de longo prazo.
<i>series_length</i>	O número de observações na série temporal.
<i>unitroot_pp</i>	Valor observado da estatística teste do teste Phillips-Perron, cuja hipótese nula é que a série temporal é não estacionária. Com raiz unitária, é não estacionária, as suas propriedades como média e variância, não são constantes ao longo do tempo. A estatística pode ser definida como sendo para a versão "Z-alpha" do teste de raiz unitária de Phillips & Perron, com constante, tendência e defasagem 1. A hipótese nula do teste é que a série tem uma raiz unitária (é não estacionária). Rejeição da hipótese nula acontece se a estatística "Z-alpha" for um número negativo grande, existe uma forte evidência para rejeitar a hipótese nula.
<i>unitroot_kpss</i>	Estatística do teste para testar a estacionariedade, onde a hipótese nula é que a série é estacionária. Utilizando o teste de <i>Kwiatkowski-Phillips-Schmidt-Shin</i> (KPSS). A hipótese nula indica que a série temporal é estacionária. Hipótese alternativa indica que a série tem uma raiz unitária e é não estacionária.
<i>hw_alpha</i>	Parâmetro de suavização para o valor médio exponencial de Holts-Winters. Este parâmetro varia entre 0 e 1. Em que próximo de 1 é um modelo altamente reativo a mudanças no nível da série. Atribui um peso muito elevado às observações recentes e baixo a observações passadas. Se for próximo de 0, dá pouca relevância às observações mais recentes torna o modelo estável e resistente a mudanças. Este tem ainda um modelo ETS(A,A,A) com erro aditivo, tendência aditiva e sazonalidade aditiva.
<i>hw_beta</i>	Parâmetro de suavização para a tendência na suavização exponencial de Holts-Winters.
<i>hw_gamma</i>	Parâmetro de suavização para a sazonalidade na suavização exponencial de Holts-Winters.
<i>stability</i>	Estabilidade do desvio padrão da série temporal. A variância das médias com base na divisão da série em porções não sobrepostas. O tamanho das porções é a frequência da série, ou 10 se a série tiver frequência 1. Para efetuar o cálculo faz-se uma divisão das séries com tamanhos iguais. A média é calculada para cada uma dessas porções. Por fim, efetua-se o cálculo da variância das médias. Valores baixos indicam que as médias são muito parecidas entre si, isto sugere que a série temporal é estável. Valores altos indicam que as médias variam, e que a série temporal é considerada instável.
<i>nperiods</i>	O número de períodos sazonais completos na série.
<i>seasonal_period</i>	A duração do período sazonal. É o número de dados que uma série temporal precisa para completar um ciclo sazonal completo e voltar a repetir-se.
<i>trend</i>	A força da tendência linear na série. Refere-se a uma direção de longo prazo de uma série de dados.

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

<i>spike</i>	Presença de valores atípicos (picos e vales) na série temporal. Possível calcular através do intervalo interquartil, vai ser definido um limite inferior e um limite superior. Os dados que estejam abaixo do limite ou acima do limite são respetivamente um <i>outlier</i> ou um pico.
<i>linearity</i>	Em uma decomposição de séries sendo T_t uma componente de tendência linear e β_1 a taxa de linearidade. O parâmetro β_1 quantifica o declive do movimento linear subjacente, enquanto β_2 ajusta o declive ao longo do tempo. O modelo quadrático pode ser dado pelo modelo $T_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \epsilon_t$
<i>curvature</i>	Através do modelo quadrático $T_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \epsilon_t$ é dado por β_2 descreve a mudança desse movimento em que revela se a tendência está a acelerar ou a abrandar. $\beta_2 > 0$ a curvatura é positiva e a taxa de crescimento da tendência está a aumentar. $\beta_2 < 0$ a curvatura é negativa e a taxa de crescimento da tendência está a diminuir. Curvatura $\beta_2 = 0$ a tendência é linear e a taxa de crescimento é constante, sem aceleração ou desaceleração.
<i>e_acf1</i>	Variável que representa a autocorrelação de primeira ordem da uma série temporal. Em que e é a componente resíduo ou erro. Característica fundamental que nos diz quão bem um modelo de previsão se encaixa nos dados. Sendo uma autocorrelação que atua como um diagnóstico de qualidade do modelo, objetivo é isolar a tendência e a sazonalidade, deixando apenas um “ruído branco”. Perto de zero significa que os erros do modelo são aleatórios e não estão relacionados uns com os outros, ou seja, não tem nenhum padrão. Diferente de zero significa que os erros não são aleatórios, que existe um padrão que não foi capturado pelo modelo de decomposição.
<i>e_acf10</i>	Soma dos quadrados das 10 primeiras autocorrelações dos resíduos.
<i>seasonal_strength</i>	Força da sazonalidade, ou seja, refere-se à força do padrão sazonal na série temporal.
<i>peak</i>	A localização do pico (valor máximo) na componente sazonal.
<i>trough</i>	A localização do vale (valor mínimo) na componente sazonal.
<i>x_pacf5</i>	Soma dos quadrados das 5 primeiras autocorrelações parciais da série. O valor obtido for longe de zero, sugere que este é importante para a previsão de um modelo autorregressivo. Pelo contrário, se for perto de zero, indica que não é importante para o cálculo de um modelo de autorregressão.
<i>diff1x_pacf5</i>	A soma dos quadrados dos primeiros 5 coeficientes de autocorrelação parcial da série com primeira diferença. Um valor alto, longe de zero indica que é importante para a previsão de uma série já estacionária.
<i>diff2x_pacf5</i>	Soma dos quadrados das 5 primeiras autocorrelações parciais da série diferenciada duas vezes diferença. Tal como na anterior um valor alto indica que esta é importante para a previsão de uma série estacionária.
<i>seas_pacf</i>	Autocorrelação parcial é utilizada para medir a correlação entre uma série temporal e os seus valores em intervalos sazonais. Um valor significativo indica uma forte correlação do mesmo período no ciclo sazonal anterior.
<i>nonlinearity</i>	Mede o grau de não linearidade de uma série. Significa que a taxa de variação da tendência não é constante pode acelerar ou desacelerar (Dix, 2025). O termo de não linearidade é captada por $\beta_2 t^2$ na fórmula $T_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \epsilon_t$. Este pode ser representado por qualquer valor, positivo, negativo ou zero. Se $\beta_2 > 0$ a tendência está a acelerar se $\beta_2 < 0$ a tendência está a desacelerar.

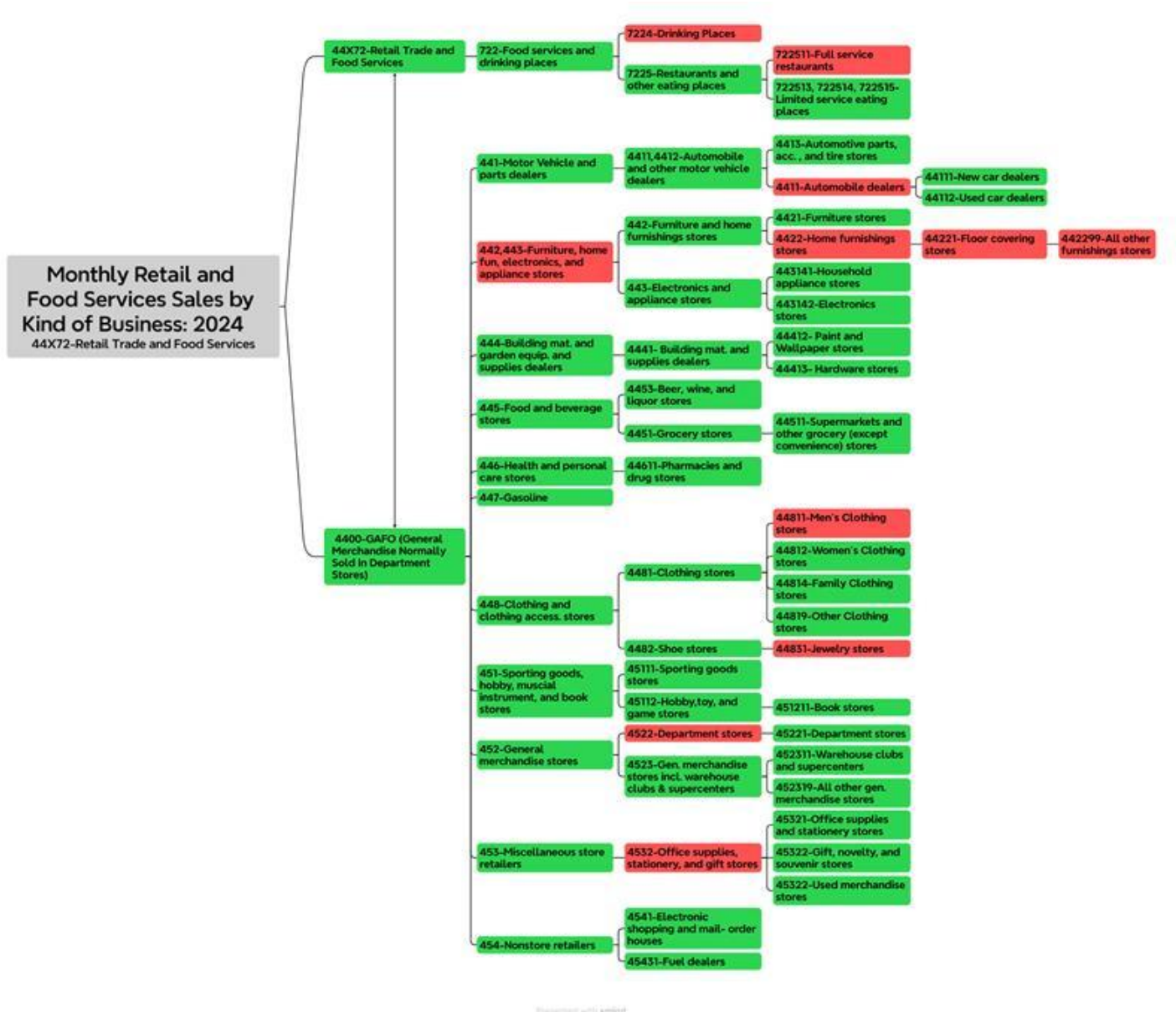
*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

<i>lumpiness</i>	A variância das variâncias baseada numa divisão da série em porções não sobrepostas. Variação do desvio padrão em diferentes janelas de tempo. Um valor <i>lumpiness</i> alto significa uma maior volatilidade na série, ou seja <i>clusters</i> de valores altos ou baixos. Um <i>lumpiness</i> baixo, significa que a série tem uma volatilidade mais consistente e previsível ao longo do tempo.
<i>alpha</i>	Parâmetro de suavização de nível. O seu valor, que está sempre entre 0 e 1, determina o peso que a observação mais recente tem na estimativa do nível da série. Um valor alto indica o quanto volátil e imprevisível é a série. Um valor baixo indica o quão estável e previsível é a série.
<i>beta</i>	Parâmetro de suavização de tendência do modelo de suavização exponencial de Holt. Se o valor de beta é próximo de 1 então a série tem uma tendência volátil, com crescimento ou decréscimo rápido. Em oposição se a série indica valores próximos de 0 tem uma tendência suave.
<i>arch_acf</i>	Autocorrelação dos quadrados dos resíduos do modelo ARCH. Métrica utilizada para medir a volatilidade de uma série temporal. Um valor elevado indica que a sua volatilidade não é constante, isto pode significar períodos de alta volatilidade ocorrem com frequência. Enquanto que um valor próximo de zero indica que não há volatilidade e a série é constante.
<i>garch_acf</i>	Autocorrelação dos quadrados dos resíduos do modelo GARCH. Um valor próximo de zero indica que o modelo conseguiu capturar com sucesso a variabilidade e volatilidade da série temporal. Um valor longe de zero indica que o modelo não conseguiu capturar a volatilidade.
<i>arch_r2</i>	R-quadrado de um modelo ARCH. Indica a percentagem da variância da volatilidade de uma série temporal. Valores próximos de 1 indicam que o valor passado da volatilidade tem um forte poder de previsão sobre a volatilidade atual. Um valor próximo de 0 sugere que os valores passados da série não são bons preditores da volatilidade atual.
<i>garch_r2</i>	R-quadrado, dos resíduos, de um modelo GARCH. Um valor próximo de 0 indica que o modelo foi capaz de explicar a variabilidade na volatilidade da série. Os resíduos que sobraram são aleatórios e sem correlação. Longe de 0 significa que o modelo não foi suficiente para capturar todos os padrões de volatilidade.
<i>flat_spots</i>	O número de pontos estáveis na série, calculado discriminando a série em 10 intervalos de igual tamanho e contando o maior comprimento de sequência ininterrupta dentro de qualquer intervalo. Um valor alto deste indicador resulta num longo período de estabilidade na série temporal. Um baixo valor indica que a série está sempre em constante mudança.
<i>entropy</i>	Mede o grau de imprevisibilidade ou complexidade de uma série. Um valor alto de entropia indica que a série é muito complexa e aleatória. A série tem pouca estrutura e é difícil prever o seu próximo valor. Um valor de entropia baixo, a série tem um padrão definido e é mais previsível. A maioria dos seus valores concentra-se num pequeno número de intervalos. Esta apresenta uma estrutura clara, com tendência e sazonalidade.
<i>crossing_points</i>	O número de vezes que a série cruza a média. Valor alto indica que a série é muito volátil e oscilante, enquanto um valor baixo indica que a série é suave e estável.
<i>arch_lm</i>	Estatística baseada no teste do Multiplicador de Lagrange de Engle (1982) para a heterocedasticidade condicional autorregressiva. Tem a estrutura de uma hipótese nula em que a série não tem efeitos ARCH. Se o p-valor for inferior a 0,05 o resultado é

*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

	significativo e rejeita-se a hipótese nula. Se o p-valor for superior a 0,05 o resultado não é significativo e não se rejeita a hipótese nula.
x_acf1	Autocorrelação de lag 1 da série original.
x_acf10	Soma dos quadrados das 10 primeiras autocorrelações da série original.
$diff1_acf1$	Autocorrelação de lag 1 da série diferenciada uma vez.
$diff2_acf1$	Autocorrelação de lag 1 da série diferenciada duas vezes.
$diff2_acf10$	Soma dos quadrados das 10 primeiras autocorrelações da série diferenciada duas vezes.
$seas_acf1$	Autocorrelação sazonal parcial de lag 1 dos dados sazonais.

APÊNDICE 2. Hierarquia das categorizações das vendas mensais no retalho e restauração nos EUA



*Impacto da COVID-19 no Retalho dos Estados Unidos da América:
Comparação por Análise de Clusters para Séries Temporais*

APÊNDICE 3. Lista com todas as séries por unique_id e nome

	unique_id	nome
0	MRTSSM447USN	Gasoline Stations
1	MRTSSM445USN	Food and Beverage Stores
2	MRTSSM4451USN	Grocery Stores
3	MRTSSM4453USN	Beer, Wine, and Liquor Stores
4	MRTSSM443USN	Electronics and Appliance Stores
5	MRTSSM44511USN	Supermarkets and Other Grocery (Except Convenience) Stores
6	MRTSSM45111USN	Sporting Goods Stores
7	MRTSSM451211USN	Book Stores
8	MRTSSM45112USN	Hobby, Toy, and Game Stores
9	MRTSSM443142USN	Electronics Stores
10	MRTSSM452USN	General Merchandise Stores
11	MRTSSM442USN	Furniture and Home Furnishings Stores
12	SM452319USN	All Other General Merchandise Stores
13	SM45221USN	Department Stores
14	SM452311USN	Warehouse Clubs and Supercenters
15	SM4523USN	General Merchandise Stores, Including Warehouse Clubs & Supercenters
16	MRTSSM446USN	Health and Personal Care Stores
17	MRTSSM44611USN	Pharmacies and Drug Stores
18	MRTSSM453USN	Miscellaneous Store Retailers
19	MRTSSM45321USN	Office Supplies and Stationery Stores
20	MRTSSM45322USN	Gift, Novelty, and Souvenir Stores
21	MRTSSM45330USN	Used Merchandise Stores
22	MRTSSM444USN	Building Materials and Garden Equipment and Supplies Dealers
23	MRTSSM44412USN	Paint and Wallpaper Stores
24	MRTSSM44413USN	Hardware Stores
25	MRTSSM454USN	Nonstore Retailers
26	MRTSSM45431USN	Fuel Dealers
27	MRTSSM4421USN	Furniture Stores
28	MRTSSM4541USN	Electronic Shopping and Mail-Order Houses
29	MRTSSM448USN	Clothing and Clothing Accessory Stores
30	MRTSSM4482USN	Shoe Stores
31	MRTSSM44812USN	Women's Clothing Stores
32	MRTSSM44814USN	Family Clothing Stores
33	MRTSSM44819USN	Other Clothing Stores
34	MRTSSM441USN	Motor Vehicle and Parts Dealers
35	MRTSSM441XUSN	Auto and Other Motor Vehicle Dealers
36	MRTSSM4413USN	Automotive Parts, Accessory, and Tire Stores
37	MRTSSM44111USN	New Car Dealers
38	MRTSSM44112USN	Used Car Dealers
39	MRTSSM722USN	Food Services and Drinking Places
40	SM72251XUSN	Limited Service Eating Places
41	MRTSSM7225USN	Restaurants and Other Eating Places
42	MRTSSM4400CUSN	GAFO (General Merchandise Normally Sold in Department Stores)
43	MRTSSM44X72USN	Retail Trade and Food Services
44	MRTSSM4423XUSN	Furniture, Home Furnishings, Electronics, and Appliance Stores