



Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Estatística

Dissertação de Mestrado

Previsão de séries temporais aplicada a dados de venda de uma grande varejista do Brasil

por

Aïtcheou Gauthier Zountchégnon

Brasília, 19 de dezembro de 2023

Previsão de séries temporais aplicada a dados de venda de uma grande varejista do Brasil

por

Aïtcheou Gauthier Zountchégnon

Dissertação apresentada ao Departamento de Estatística da Universidade de Brasília, como requisito parcial para obtenção do título de Mestre em Estatística.

Orientador: Prof. Dr. Jose Augusto Fiorucci

Brasília, 19 de dezembro de 2023

Dissertação submetida ao Programa de Pós-Graduação em Estatística do Departamento de Estatística da Universidade de Brasília como parte dos requisitos para a obtenção do título de Mestre em Estatística.

Texto aprovado por:

Prof. Dr José Augusto Fiorucci
Orientador, EST/UnB

Prof. Dr. Eduardo Nakano
EST/UnB

Prof. Dr. MARINHO G. ANDRADE
USP

Para toda minha família Zountchégnon.

Agradecimentos

À minha família, que é a minha fonte de motivação, que mesmo longe se esforça para me ajudar de todas as formas possíveis.

Ao meu orientador, Prof. Dr. Jose Augusto Fiorucci, que me ensinou bastante coisas durante a dissertação, e teve a paciência e a sabedoria necessária para orientar esse trabalho.

A todos os meus Professores do departamento, pelos ensinamentos, pelas ajudas nas matérias e pelas motivações.

A todas as pessoas que me ajudaram em algum momento durante o mestrado.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Resumo

O comércio varejista desempenha um papel crucial na economia brasileira, sendo o planejamento do volume de vendas e outros fatores associados a esse setor de extrema importância para seu crescimento. Para a previsão e planejamento eficazes das vendas, métodos relacionados a séries temporais surgem como ferramentas fundamentais. Este estudo concentra-se no desenvolvimento e avaliação de modelos preditivos, os quais devem levar em consideração características típicas desses dados, como a estrutura hierárquica, a presença de múltiplas sazonalidades nas séries de níveis mais elevados e o comportamento intermitente nas séries de níveis mais baixos.

Palavras-Chave: Previsão de demanda, séries temporais hierárquicas, séries temporais intermitentes, modelo TBATS, modelo de regressão dinâmica, reconciliação ótima.

Abstract

Retail trade plays a crucial role in the Brazilian economy, and planning for sales volume and other factors related to the retail sector is of great importance for its growth. To effectively forecast and plan sales quantities, methodologies related to time séries can be employed. This study focuses on the development and evaluation of predictive models, which should consider typical characteristics of such data, such as hierarchical structure, the presence of multiple seasonalities in higher-level séries, and intermittent behavior in lower-level séries.

Key Words: Demand forecasting, hierarquical time séries, intermittent time séries, TBATS model, dinamic regression model, optimal reconciliation. .

Title: Time series forecasting applied to data sale of a large retailer in Brazil.

Sumário

1	Introdução	1
2	Revisão de literatura	4
2.1	Séries temporais hierárquicas e agrupadas	4
2.1.1	A abordagem Bottom-up	7
2.1.2	Abordagens Top-down	8
2.1.3	Abordagem de reconciliação ótima	9
2.2	Modelos para previsão de séries temporais	11
2.2.1	Modelo ARIMA	11
2.2.2	Modelo Croston	16
2.2.3	Modelo Regressão Dinâmica e Harmônica	17
2.2.4	Modelo TBATS	19
3	Metodologia	22
3.1	Conjunto de dados	22
3.1.1	Dados de venda	22
3.1.2	Dados de feriados para o modelo Prophet	27
3.1.3	Dados de feriados para o modelo Arima	29
3.2	Algumas técnicas utilizadas nos experimentos	30
3.2.1	A Validação Cruzada	30

3.2.2	Teste de ranking	33
3.3	Experimentos	35
3.3.1	Modelagem no nível <i>subclasses</i> (nível 1)	35
3.3.2	Modelagem no nível <i>item-parent</i> (nível 2)	37
4	Resultados	41
4.1	séries do nível <i>subclasses</i> (nível 1)	41
4.2	séries do nível <i>item-parent</i> (nível 2)	45
5	Conclusões	51
5.1	Considerações Finais	51
	Referências Bibliográficas	53
	Apêndice - Análises dos resíduos modelos do nível 1	55

Lista de Tabelas

1	Teste estatístico	58
---	------------------------------------	----

Lista de Figuras

2.1	Estrutura hierárquica de dois níveis	5
2.2	Estrutura agrupada de dois níveis	6
2.3	Propriedades teóricas da FAC e FACP.	13
3.1	Imagem da base de dados	23
3.2	Hierarquia dos dados	24
3.3	Decomposição da série de total de venda	25
3.4	Decomposição da série de venda de uma subclasse	26
3.5	Decomposição de uma série de venda de um Item-parent	26
3.6	Decomposição de uma série de venda de um sku	27
3.7	Imagem das variáveis exógenas do modelo Prophet	28
3.8	Imagem das variáveis exógenas do modelo ARIMA	29
3.9	Janela deslizando	32
3.10	séries temporais do nível 1	36
4.1	Erros dos modelos na subclasse 6189	42
4.2	Desempenho dos modelos	44
4.3	Imagem dos dados para o ranking	46
4.4	Desempenho dos modelos no nível 2	48
4.5	Gráfico do teste de ranking das 3 datas.	49

1	Análises visuais dos resíduos.	56
---	--	----

Abreviações e Siglas

AIC	Critério de Informação de Akaike (Akaike information criterion).
AR(p)	Autorregressivo de ordem p
ARIMA(p,d,q)	Autorregressivo integrado de médias móveis de ordem p, d, q
ARMA(p,q)	Autorregressivo de médias móveis de ordem p, q
BIC	Critério de Informação Bayesiano
FCA	Função de autocorrelação
FCAP	Função de autocorrelação parcial
H_0	hipótese nula
H_1	hipótese alternativa
KPSS	Kwiatkowski, Phillips, Schmidt e Shin
MA(q)	Média móvel de ordem q
MAE	Média absoluta dos erros de previsão
MSE	<i>mean squared errors</i>
<i>P-value</i>	nível descritivo de um teste de hipóteses

Capítulo 1

Introdução

Uma série temporal pode ser considerada como uma coleção de observações feitas sequencialmente ao longo do tempo, em intervalos iguais. Neste tipo de dados, as informações vizinhas são dependentes e o objetivo é encontrar padrões não aleatórios nos dados, analisar e modelar a dependência das observações para a realização de previsões sobre o futuro, ajudando na tomada de decisões (Ord, Fildes e Kourentzes, 2017).

Modelos de séries temporais são usados em várias áreas, como Engenharia, Ciências Médicas, Indústria, Mercado Financeiro, Economia, Setor Varejista, dentre outros. Este tipo de modelagem geralmente é empregado para realizar planejamento, estimar antecipadamente as demandas, as capacidades e necessidades de recursos, entre outros (Box et al., 2015; Ord, Fildes e Kourentzes, 2017; Hyndman e Athanasopoulos, 2018). Por exemplo, para uma empresa varejista planejar as vendas, ela precisa das previsões da demanda futura. A previsão é uma ferramenta muito importante para um planejamento eficaz e eficiente.

Dependendo da situação, do problema e do tipo de negócio, as séries temporais podem ser agregadas ou desagregadas. Por exemplo, o número total de sapatos produzidos ao longo do tempo pode ser dividido por cores, tamanhos, etc. Cada um desses grupos pode ainda ser desagregado em outros subgrupos, como regiões geográficas, lojas. Esse tipo de dado caracteriza as séries temporais hierárquicas (Hyndman e Athanasopoulos, 2018, cap. 10).

Geralmente, os artigos sobre séries temporais focam em um determinado nível de agregação, mas o desafio surge quando lidamos com séries com níveis hierárquicos diferentes. Nesse caso, por questões de coerência, fazer previsões para as granularidades mais baixas não é igual à previsão agregada. Portanto, é necessário utilizar técnicas que gerem previsões coerentes (ou reconciliem previsões incoerentes), permitindo que as previsões de séries temporais individuais sejam feitas de forma independente, mantendo os relacionamentos dentro da hierarquia (Wickramasuriya, Athanasopoulos e Hyndman, 2015; Wickramasuriya, Athanasopoulos e Hyndman, 2019). Atualmente, as técnicas de previsão de séries temporais hierárquicas são amplamente aplicadas em diversas áreas, principalmente no setor do varejo.

Este estudo foca no desenvolvimento e avaliação de modelos preditivos para dados de vendas de uma grande varejista de modas do Brasil, a qual forneceu o conjunto de dados e o modelo atualmente utilizado, este baseado em *Machine Learning* de nome comercial *Prophet* (Taylor e Letham, 2018). Esses dados são apresentados com frequência semanal de observações em uma estrutura que é naturalmente hierárquica, pois contem o histórico de vendas de diversos produtos, que são vendidos em diferentes setores dentro de diversas lojas. Além de metodologias para lidar com a estrutura hierárquica, por se tratar de dados de venda semanais, é natural que esses apresentem múltipla sazonalidade nos níveis mais alto da hierarquia, o qual demanda de modelagem específica (De Livera, Hyndman e Snyder, 2011).

O estudo desenvolvido consiste primeiramente em identificar as melhores metodologias conforme o nível hierárquico das séries temporais, em seguida é proposta a utilização de uma metodologia de reconciliação para equilibrar as previsões geradas para os diferentes níveis da hierarquia. O desempenho preditivo dos modelos é avaliado por meio de janela deslizante, de modo que algumas formulações propostas apresentam resultados de acurácia melhores que o modelo atualmente utilizado pela varejista.

Os capítulos a seguir estão distribuídas da seguinte forma, no capítulo 2 apresentamos as principais técnicas para séries temporais hierárquicas e agrupadas, as principais metodologias para previsão de séries temporais que são adequadas para determinados níveis da hierarquia

dos dados; no capítulo 3 apresentamos os conjuntos de dados, os experimentos realizados, os resultados dos testes foram apresentados no capítulo 4; por fim, no capítulo 5 apresentamos uma finalização dessa monografia e por ultimo a bibliografia utilizada.

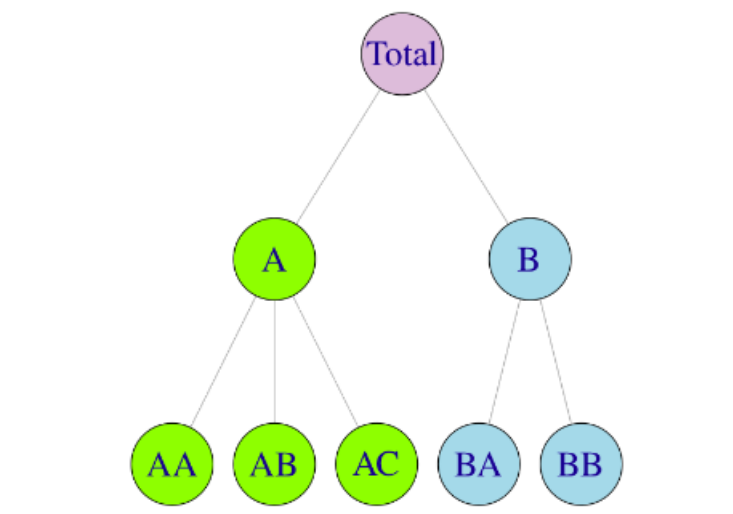
Capítulo 2

Revisão de literatura

2.1 Séries temporais hierárquicas e agrupadas

As séries temporais geralmente podem ser agregadas e desagregadas por diversas propriedades de interesse. Ao comentar de séries temporais agregadas e desagregadas, é possível distinguir dois casos e é simples de explicar usando um exemplo; consideramos uma empresa varejista de E-commerce que comercializa uma diversidade de itens e produtos no Brasil. Quando olhamos para a venda total de todos os produtos, temos o caso de séries temporais agregadas. Essa série pode ser desagregada por tipo de produto, por regiões geográficas, por lojas etc. Essas divisões são ligadas às categorias de grupos superiores e, portanto, a coleção de séries temporais tem uma estrutura de agregação hierárquica.

A Figura 2.1 ilustra uma estrutura hierárquica de dois níveis ($k=2$). No topo da estrutura está o *Total* o nível de agregação mais completa dos dados. O topo da hierarquia *Total* é chamado de nível 0, geralmente denominado como nó raiz e denotado por y_t em que $t = 1, \dots, n$. No nível 1 temos a desagregação de *Total* em dois grupos (A e B), e os nós podem ser considerados como intermediários e dependendo do contexto eles podem ser chamados de nós-pai ou nós-folhas. Por exemplo, o nó A é um nó-folha em relação ao nó *Total* e um nó-pai em relação ao nó AB. No nível 2, as séries A e B são divididas em três e dois grupos respectivamente no nível inferior



Fonte: (Hyndman e Athanasopoulos, 2018)

Figura 2.1: Estrutura hierárquica de dois níveis

da hierarquia. Esse nível corresponde a desagregação máxima dos dados e são considerados como nós-folha.

A partir do nível 1 a notação $y_{j,t}$ será usada para indicar a observação t do nó j . Considerando a Figura 2.1, $y_{A,t}$ corresponde a t -ésima observação do grupo A no nível 1, $y_{BB,t}$ indica a t -ésima observação do nó BB no nível 2, e assim sucessivamente.

Considerando o exemplo da Figura 2.1, temos uma série no nível 0, duas no nível 1 e cinco séries no nível 3 então $m=5$ (m é quantidade total de serie no nível mais baixo), ao total temos $n=8$ séries (n é a quantidade total de série). Em todos os níveis é possível perceber que $n > m$.

Os dados do nível inferior da estrutura serão acrescentados às observações da série do nível superior, para qualquer t . Podemos traduzir essa relação pelas seguintes equações:

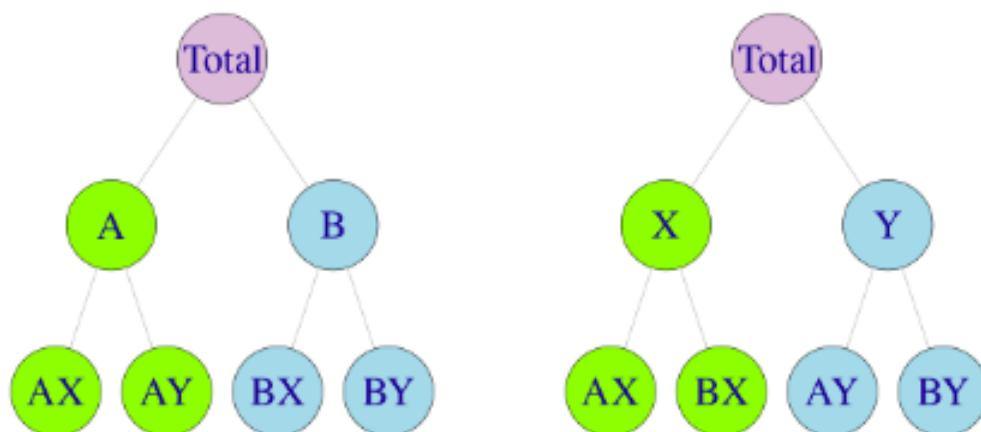
$$y_t = y_{AA,t} + y_{AB,t} + y_{AC,t} + y_{BA,t} + y_{BB,t}$$

$$y_{A,t} = y_{AA,t} + y_{AB,t} + y_{AC,t}$$

$$y_{B,t} = y_{BA,t} + y_{BB,t}$$

Séries temporais agrupadas apresentam formas de agregação mais abrangentes em com-

paração com as séries temporais hierárquicas, e há várias maneiras de desagregar a estrutura hierárquica. Tomando como exemplo o varejista de E-commerce, é possível observar vários níveis de detalhamento, como tipo de produto, fabricante, cores, tamanho, região geográfica, lojas, faixa de preço, entre outros. No caso das séries temporais agrupadas, pode não existir uma única maneira “correta” de agregar os dados, proporcionando flexibilidade na análise e modelagem. Isso ocorre porque diferentes níveis de agregação podem ser mais apropriados para diferentes contextos ou objetivos analíticos.



Fonte: (Hyndman e Athanasopoulos, 2018)

Figura 2.2: Estrutura agrupada de dois níveis

A Figura 2.2 apresenta uma estrutura agrupada de nível 2. O nível 0 representado por *Total* está no topo da estrutura agrupada. O *Total* é o nível mais agregado dos dados e pode ser desagregado em duas séries temporais A e B ou em X e Y. O raciocínio e anotações são semelhantes às séries temporais hierárquicas. A Figura 2.2 mostra também que diferente das séries temporais hierárquicas, neste caso pode existir uma variedade de caminhos para a desagregação dos dados.

Previsão de séries temporais hierárquica/agrupadas é geralmente desafiador porque é preciso gerar previsões coerentes em toda a estrutura de agregação. A previsão é dita coerente quando ela soma de forma consistente com a estrutura de agregação subjacente. Por exemplo, considerando a Figura 1, as previsões de todas as séries do nível 2 devem ser coincidir com às previsões das séries do nível 1, e todas do nível 1 ao nível mais alto.

2.1.1 A abordagem Bottom-up

A abordagem Bottom-up é a mais simples para gerar previsões coerentes, nessa abordagem as previsões são feitas de baixo para cima. Um modelo adequado ao problema é usado para prever cada série do nível mais baixo da hierarquia de forma independente e, em seguida as previsões são somadas levando em consideração a hierarquia para chegar nas previsões dos níveis superiores. Considerando o exemplo da Figura 1, suponha que h previsões foram geradas para as séries do nível mais baixo $\hat{y}_{AA,h}, \hat{y}_{AB,h}, \hat{y}_{CA,h}, \hat{y}_{BA,h}, \hat{y}_{BB,h}$ e depois as previsões foram somadas para chegar nos níveis mais altos da hierarquia segundo as seguintes relações:

$$\begin{aligned}\bar{y}_h &= \hat{y}_{AA,h} + \hat{y}_{AB,h} + \hat{y}_{CA,h} + \hat{y}_{BA,h} + \hat{y}_{BB,h} \\ \bar{y}_{A,h} &= \hat{y}_{AA,h} + \hat{y}_{AB,h} + \hat{y}_{CA,h} \\ \bar{y}_{B,h} &= \hat{y}_{BA,h} + \hat{y}_{BB,h}\end{aligned}$$

Desta forma, não há perda de informação por causa da agregação, porque as previsões são feitas no nível inferior da hierarquia. No entanto, a desvantagens é que algumas relações entre as séries como diferença entre regiões, pais, lojas, etc não são levadas em consideração. Além que dependendo da quantidade de observação nos níveis mais baixos, a abordagem pode ser muito custoso computacionalmente. Por causa dos ruídos nos dados de níveis mais baixos, a modelagem pode ficar complicada. Por exemplo, e comum que as séries temporais desses níveis mais baixos sejam inflacionadas de zeros e não apresentem características como tendência e sazonalidade evidentes.

2.1.2 Abordagens Top-down

A abordagem Top-down as previsões são realizadas de cima para baixo, neste caso em um primeiro momento, é calculada a previsão para o nível mais agregado para então ser feita a divisão para os níveis inferiores de hierarquia. Com essa abordagem precisamos definir como as previsões do topo devem ser divididas para chegar nas previsões da base. Geralmente as proporções históricas são usadas na desagregação. Na literatura, existem alguns métodos para calcular essas proporções de modo a manter as características do período de interesse.

A biblioteca *hts* (Hyndman, Athanasopoulos e Shang, 2015) do *R* disponibiliza três métodos para obter as proporções nessa abordagem, as quais estão listadas abaixo.

- *Proporções históricas das médias*: Suponha $j = 1, \dots, m$. Cada proporção p_j reflete a média das proporções históricas da série de nível inferior $y_{j,t}$ ao longo do período $t = 1, \dots, T$ em relação à série agregada total y_t . A proporção p_j pode ser representada pela relação seguinte :

$$p_j = \frac{1}{T} \sum_{t=1}^T \frac{y_{j,t}}{y_t}.$$

- *Proporções das médias históricas*: Cada proporção p_j captura o valor histórico médio da série do nível baixo $y_{j,t}$ em relação ao valor médio da série agregada total e é dada por

$$p_j = \frac{\sum_{t=1}^T \frac{y_{j,t}}{T}}{\sum_{t=1}^T \frac{y_t}{T}}.$$

- *Previsão de proporções*: Nesta abordagem, são realizadas em primeiro lugar previsões para as séries de nível inferior, depois é calculada a proporção de cada previsão. As proporções encontradas são usadas para desagregar a previsão da série do topo. A formula da proporção pode ser expressa da forma seguinte:

$$p_j = \prod_{l=0}^{K-1} \frac{\hat{y}_j^{(l)},h}{S^{(l+1)}}.$$

em que K indica a quantidade do nível da hierarquia, $\hat{y}_j^{(l)},_h$ é a previsão de h passos a frente da série correspondente ao nó que está l níveis acima de j e $S^{(\hat{l}+1)}$ representa a soma das previsões h passos a frente abaixo do nó que está l níveis acima do nó j .

A vantagem da abordagem *top-down* é a simplicidade, sendo esta possivelmente uma das abordagens mais utilizadas na prática pelos varejistas. No entanto, fica claro que essa abordagem trabalha em um nível bastante resumido dos dados, desprezando assim um grande volume de dados e informação.

2.1.3 Abordagem de reconciliação ótima

A abordagem de reconciliação ótima, conforme proposta por Wickramasuriya, Athanasopoulos e Hyndman, 2019, faz uso de todas as informações disponíveis nas estruturas das séries hierárquicas ou agrupadas. Isso é crucial, uma vez que os usuários podem identificar padrões e características relevantes para modelagem em determinados níveis de agregação e agrupamento. Nessa metodologia, as previsões são inicialmente realizadas de forma independente para todos os níveis, sem considerar o fato de que os resultados podem ser incoerentes. Posteriormente, as previsões independentes são ajustadas pela reconciliação ótima, gerando resultados coesos e consistentes com a estrutura hierárquica.

Na prática, a previsão atualizada de cada nível é calculada pela média ponderada das previsões de todos os níveis. Os pesos utilizados nesse cálculo são determinados por meio de regressão linear, na qual as previsões de todos os nós são regredidas contra variáveis exógenas que indicam a contribuição de cada série do nível inferior em cada nó. Essa abordagem é considerada “ótima” devido ao fato de que utiliza o menor erro quadrático médio na reconciliação, garantindo que os pesos permitam a atualização das previsões de todos os níveis pelo valor mais baixo possível para alcançar coerência nas previsões.

As abordagens de previsão para as estruturas hierárquicas ou agrupadas podem ser repre-

sentadas pela seguinte formula (Hyndman e Athanasopoulos, 2018):

$$\tilde{y}_h = SG\hat{y}_h$$

em que \hat{y}_h denota o vetor de previsão para todas as séries calculada de modo independente, esta é chamada de previsão base, h é o horizonte de previsão, G é uma matriz que mapeia as previsões da bases para o nível mais inferior, e a matriz S agrega os valores usando a estrutura de agregação para gerar um conjunto de previsões coerentes \tilde{y}_h .

O objetivo da abordagem é encontrar a matriz G que consiga minimizar as variações do erro das previsões coerentes. Segundo (Hyndman e Athanasopoulos, 2018), é importante ter previsões não viesadas. Nesse caso, se \hat{y}_h é não viesado então \tilde{y}_h é não viesado sob restrição de que $SGS = S$. Para encontrar os erros das previsões, Wickramasuriya, Athanasopoulos e Hyndman, 2019 mostra que a matriz de variância-covariância dos erros de previsões coerentes de h passos à frente é dada por:

$$V_h = Var[y_{T+h} - \tilde{y}_h] = SGW_hG'S'$$

em que $W_h = Var[y_{T+h} - \hat{y}_h]$ representa a matriz de variância-covariância dos erros de previsões de base.

A diagonal da matriz V_h contém as variações de erro das previsões coerentes, então o traço de V_h seria a soma dos elementos da diagonal. Wickramasuriya, Athanasopoulos e Hyndman, 2019 mostra que a matriz G que minimiza o traço de V_h tal que $SGS = S$ é dada por:

$$G = (S'W_h^{-1}S)^{-1}S'W_h^{-1}$$

e as previsões da reconciliação são dadas pela equação:

$$\tilde{y}_h = S(S'W_h^{-1}S)^{-1}S'W_h^{-1}\hat{y}_h.$$

Desta forma, é necessário estimar primeiro o W_h e a variância dos erros de previsões coerentes de h passos à frente para viabilizar o uso da fórmula. A abordagem mais simples consiste em considerar igual a zero as covariâncias entre as previsões. Mas existem outras abordagens, as quais podem ser consultadas em Wickramasuriya, Athanasopoulos e Hyndman, 2019.

Geralmente, essa abordagem gera resultados mais precisos que a *bottom-up* e a *top-down*, pois trabalha com todos os níveis hierárquicos do conjunto de dados. No entanto esta abordagem pode ser lenta computacionalmente devido ao cálculo dos pesos de reconciliações usados para gerar uma média ponderada das previsões em cada nó da estrutura hierárquica.

2.2 Modelos para previsão de séries temporais

Dada a natureza do problema em estudo, é comum que as séries nos níveis mais baixos da hierarquia exibam comportamento intermitente com excesso de zeros, enquanto as séries nos níveis mais altos podem apresentar mais de um ciclo sazonal.

Neste estudo, serão considerados vários modelos de séries temporais, os quais são detalhados nas subseções a seguir.

2.2.1 Modelo ARIMA

O modelo Auto Regressivo Integrado de Média Móvel (ARIMA) (Box et al., 2015), também chamado de metodologia de Box-Jenkins, destaca-se como um dos modelos de previsão de séries temporais mais simples e amplamente utilizados, sendo de fácil aplicação e conhecido por fornecer resultados confiáveis. No entanto, sua eficácia é mais pronunciada em séries temporais não intermitentes que apresentam apenas um ciclo sazonal.

O modelo ARIMA é o resultado da mistura de três componentes: o componente auto-regressivo (AR), o componente de Integração (I) e o componente de Médias Móveis (MA).

Um processo $\{y_t\}$ é dito ser ARIMA de ordem (p,d,q) se ele é descrito pela seguinte equação

$$\Phi_p(B) (1 - B)^d y_t = \Theta_q(B) \varepsilon_t$$

em que,

- B denota o operador de retardo, definido como $B^k y_t = y_{t-k}$, $k = 1, 2, 3, \dots$;
- $\Phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ é o polinômio autoregressivo de ordem p , sendo ϕ_1, \dots, ϕ_p os parâmetros autoregressivos;
- $\Theta_q(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q$ é o polinômio de média móvel de ordem q , sendo $\theta_1, \dots, \theta_q$ os parâmetros de médias móveis;

O termo $x_t := (1 - B)^d y_t$ é utilizado para anular d raízes unitárias do processo $\{y_t\}$, resultando em um processo $\{x_t\}$ estacionário. Se $d = 0$, então o processo ARIMA se resume a um processo estacionário conhecido como $ARMA(p, q)$.

No caso do processo $\{y_t\}$ apresentar sazonalidade, pode-se utilizar uma formulação semelhante para tratar a sazonalidade, conhecida como ARIMA Sazonal ou simplesmente SARIMA. Seja m o tamanho do ciclo sazonal, por exemplo, se o processo $\{y_t\}$ descreve uma série mensal com ciclo sazonal anual, então teríamos $m = 12$.

O modelo $SARIMA(p, d, q) \times (P, D, Q)_m$ é definido como

$$\Phi_P(B^m) \Phi_p(B) (1 - B^m)^D (1 - B)^d y_t = \Theta_Q(B^m) \Theta_q(B) \varepsilon_t$$

em que, os termos adicionais são:

- $\Phi_P(B^m) = 1 - \varphi_1 B^m - \varphi_2 B^{2m} - \dots - \varphi_P B^{Pm}$ é o polinômio autoregressivo sazonal de ordem P , sendo $\varphi_1, \dots, \varphi_P$ os parâmetros autoregressivos sazonal;
- $\Theta_Q(B^m) = 1 + \vartheta_1 B^m + \vartheta_2 B^{2m} + \dots + \vartheta_Q B^{Qm}$ é o polinômio de média móvel sazonal de ordem Q , sendo $\vartheta_1, \dots, \vartheta_Q$ os parâmetros de médias móveis sazonal;

O termo $x_t := (1 - B^m)^D (1 - B)^d y_t$ é utilizado para anular d raízes unitárias e D raízes unitárias sazonais do processo $\{y_t\}$, resultando em um processo $\{x_t\}$ estacionário.

Por definição o termo $\{\varepsilon_t\}$ é um processo independente e identicamente distribuído com média zero e variância contante (σ^2), no qual geralmente é assumido $\varepsilon_t \sim Normal(0, \sigma^2)$.

Identificação

Para utilização de um modelo ARIMA (ou SARIMA) a primeira etapa é determinar a ordem $(p, d, q) \times (P, D, Q)_m$. Para isso, é preciso verificar em primeiro lugar se a série é estacionaria ou não, no qual geralmente é empregado testes estatísticos para verificar a presença de raiz unitária simples e sazonal. Geralmente são utilizados os testes de conhecidos como Dickey Fuller ou KPSS (Dickey e Fuller, 1981; Kwiatkowski et al., 1992) para verificação de raízes unitárias simples e para verificação raízes unitárias sazonal geralmente são utilizados os testes propostos em Hylleberg et al., 1990 e Wang, Smith e Hyndman, 2006. Desta forma são determinados os termos d e D .

Para os demais termos pode-se tentar interpreta os gráficos FAC e FACP produzidos após a remoção das raízes unitárias da série, conforme apresentado na Figura 2.3. No entanto, muitas vezes essa interpretação não gera resultados claros e a seleção da ordem do modelo é feita testando um conjunto de possibilidades e então escolhe-se com base em critérios de parcimônia ou acurácia.

Processo	FAC	FACP
série aleatória	0	0
AR(1), $\alpha > 0$	decaimento exponencial	0, $k \geq 2$
AR(1), $\alpha < 0$	decaimento oscilatório	idem
AR(p)	decaimento para zero	0, $k > p$
MA(1)	0, $k > 1$	decaimento oscilatório
ARMA(p, q)	decaimento a partir de q	decaimento a partir de p

Fonte: Ehlers (2009, p. 31)

Figura 2.3: Propriedades teóricas da FAC e FACP.

Estimação

Dado uma série temporal observada y_1, \dots, y_n , seja x_1, \dots, x_n a série temporal estacionária, isto é, $x_t = (1 - B^m)^D (1 - B)^d y_t$, desta forma a estimação do vetor de parâmetros $\theta = (\varphi_1, \dots, \varphi_P, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \vartheta_1, \dots, \vartheta_Q)'$ geralmente é feita utilizando a função de verossimilhança condicional, a qual é definida como,

$$L(\theta) = f(x_{k+1}, \dots, x_n | x_1, \dots, x_k) = \prod_{t=k+1}^n f_{t-1}(x_t)$$

em que k denota o tamanho do conjunto de observações iniciais que são necessárias para inicializar o modelo, $f(x_{k+1}, \dots, x_n | x_1, \dots, x_k)$ denota a função densidade de probabilidade conjunta condicionada aos valores de inicialização x_1, \dots, x_k e $f_{t-1}(x_t)$ denota a função densidade de probabilidade de x_t condicionada em x_1, \dots, x_{t-1} .

Sob a hipótese de $\{\varepsilon_t\}$ ser um ruído branco gaussiano, é fácil ver que $f_{t-1}(x_t)$ consiste na função densidade de probabilidade da *Normal*(μ_t, σ^2), em que $\mu_t = E[x_t | x_1, \dots, x_{t-1}]$ e $\sigma^2 = Var[\varepsilon_t]$.

Ainda sob a hipótese de normalidade, pode-se mostrar que o estimador de máxima verossimilhança condicional coincide com o estimador de mínimos quadrados condicionais, isto é,

$$\hat{\theta} = \operatorname{argmin}_{\theta} \sum_{t=1}^n \varepsilon_t^2$$

em que $\varepsilon_1, \dots, \varepsilon_k$ são assumidos como zero.

Adequação

Nesta etapa, a adequação do modelo e a consistência são avaliadas. É importante verificar os itens seguintes:

- **Resíduos estacionários** : O teste de KPSS pode ser realizado para verificar a estacionariedade dos resíduos.

- **Independência dos resíduos:** Os gráficos ACF e PACF podem ser usados para examinar a independência dos resíduos. A independência é confirmada quando ACF e PACF estão dentro das bandas. Geralmente o teste conhecido como Ljung-Box (Ljung e Box, 1978) também é utilizado para testar a hipótese de independência.
- **Normalidade dos resíduos:** O gráfico qqplot é uma das ferramentas usada para verificar a normalidade dos resíduos. O teste de Shapiro-Wilk também pode ser utilizado nesse processo. No entanto, vale destacar que mesmo que a normalidade seja rejeitada, o modelo ainda pode ser aproveitado utilizando a partir da abordagem conhecida como Bootstrap não paramétrico (Hyndman e Athanasopoulos, 2018).
- **Aleatoriedade dos resíduos:** O gráfico dos resíduos pode ser realizado e os pontos têm que ter um comportamento aleatório em torno do zero para confirmar a aleatoriedade dos resíduos.

Critérios de parcimônia

Caso seja identificado mais de um modelo adequado da mesma família, geralmente os critérios de parcimônia são utilizados para escolher um deles.

O critério de Akaike (AIC) pode ser calculado da forma seguinte :

$$AIC = -2 \log(L(\hat{\theta})) + 2K$$

em que K denota o número de parâmetros do modelo.

O critério de Akaike corrigido (AICc) consiste em uma proposta que visa melhorar o desempenho quando o conjunto de dados é pequeno, este é definido como:

$$AIC_c = AIC + \frac{2K(K+1)}{n-K-1}$$

em que $\frac{2K(K+1)}{n-K-1}$ é o viés de correção

2.2.2 Modelo Croston

O modelo Croston (Croston, 1972) é comumente empregado para previsão de demanda intermitente. Mais especificamente, essas são séries temporais de contagem com valores esporádicos, passando por vários períodos com demanda igual a zero. Casos de demandas intermitentes são frequentes em setores como varejo, automotivo, manufatura, entre outros. Séries intermitentes não exibem padrões sazonais e contêm muitos valores zero. Devido à natureza desses dados, modelos clássicos muitas vezes não produzem previsões precisas, e o método de Croston pode ser mais eficaz em tais situações.

Segundo o Capítulo 12.2 de Hyndman e Athanasopoulos, 2018, o método Croston segue o seguinte procedimento:

- Separar a série temporal original em duas novas séries com base nos períodos com os valores nulos e nos intervalos de tempos com valores diferentes de zero.
- Considerar q_i como a i -ésima quantidade diferente de zero e a_i como a diferença de tempo entre q_{i-1} e q_i .
- Modelar as duas novas séries a e q usando suavização exponencial simples. A série q é chamada de demanda e a série a o tempo entre chegadas.
- Se $\hat{q}_{i+1|i}$ e $\hat{a}_{i+1|i}$ são as previsões de uma etapa do $(i + 1)$ -ésima demanda (q) e o tempo entre chegadas (a), respectivamente, com base nos dados até a demanda i , o método Croston é igual a:

$$\hat{q}_{i+1|i} = (1 - \alpha)\hat{q}_{i+1|i} + \alpha q_i$$

$$\hat{a}_{i+1|i} = (1 - \alpha)\hat{a}_{i+1|i} + \alpha a_i$$

Nas duas equações, α é denominado parâmetro de suavização e pertence ao intervalo 0 a 1. Ele é usado para atribuir pesos às observações mais recentes ou históricas.

- A previsão da demanda no tempo $(T + h)$ é dada pela equação:

$$\hat{y}_{T+H|T} = q_{j+1|j}/a_{j+1|j},$$

em que h é o tamanho da janela e j a data em que foi observada a última observação positiva.

Em resumo o modelo croston :

- Estima o nível médio de demanda quando a demanda é diferente de zero.
- Estima o tempo médio entre duas ocorrências de demanda.
- Prever a demanda como o nível de demanda (quando há ocorrência) multiplicado pela probabilidade de ocorrência.

2.2.3 Modelo Regressão Dinâmica e Harmônica

A regressão dinâmica (DR) é considerada como uma regressão clássica com variáveis explicativas mas tendo um processo ARIMA em vez de ruído branco. Ela pode ser representada pela seguinte equação (Hyndman e Athanasopoulos, 2018):

$$y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \dots + \beta_k x_{k,t} + \eta_t$$

$$\eta_t = \phi_1 \eta_{t-1} + \dots + \phi_p \eta_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

em que:

- y_t é a variável dependente.
- $x_t = (x_{1,t}, \dots, x_{k,t})$ é o vetor de variáveis regressoras ou variáveis independentes.
- $\{\varepsilon_t\}$ é um processo independente e identicamente distribuído com média zero e variância constante.

- $(\beta = \beta_0, \beta_1, \dots, \beta_k)' \in \mathbb{R}^{k+1}$ é o vetor de parâmetros das variáveis regressoras.
- $\{\eta_t\}$ é um processo ARMA(p,q).
- (ϕ_1, \dots, ϕ_p) é o vetor dos parâmetros autoregressores.
- $(\theta_1, \dots, \theta_p)$ é o vetor dos parâmetros de médias móveis.

Sob hipótese de normalidade, o estimador de Mínimos Quadrados Condicional e Máxima Verossimilhança condicional são obtidos minimizando $\sum_{t=1}^n \varepsilon_t^2$. Conhecendo os valores futuros das variáveis regressoras a previsão pontual pode ser encontrada pela seguinte equação (Hyndman e Athanasopoulos, 2018):

$$\hat{y}_{n+h|n} = \hat{\beta}_0 + \hat{\beta}_1 x_{1,n+h} + \dots + \hat{\beta}_k x_{k,n+h} + \hat{\eta}_{n+h|n},$$

em que, $\hat{\eta}_{n+h|n}$ representa a previsão obtida pelo modelo ARMA(p,q) da série $\hat{\eta}_1, \dots, \hat{\eta}_n$.

O termo regressivo sendo uma constante somada na previsão de η_{n+h} , a previsão intervalar é encontrada da mesma maneira como nos modelos ARMA/ ARIMA.

Quando estamos interessados em modelar a sazonalidade nas séries temporais adicionamos componentes de sazonalidade ao modelo ARIMA (modelo SARIMA). Mas o modelo SARIMA não consegue lidar com caso em que a série possui múltiplas sazonalidades, ou ciclos sazonais complexos (não inteiros) ou sazonalidade com grandes ciclos. Nesses casos a transformação de Fourier é usada na modelagem. Segundo a série de Fourier, qualquer função periódica pode ser decomposta numa soma de ondas senoidais e cosseno. Matematicamente, essa teoria pode ser escrita da seguinte forma :

$$f(t) = A_0 + \sum_{k=1}^{\infty} \left(A_k \cos \left(\frac{2kt\pi}{m} \right) + B_k \sin \left(\frac{2kt\pi}{m} \right) \right)$$

em que:

- A_0 é o valor médio da função periódica fornecida.

- A_k é o coeficientes dos componentes do cosseno.
- B_k é o coeficiente das componentes senoidais.
- k é a ordem que é a frequência da onda senoidal ou cosseno.
- m é o período da função.

Os termos m e k são conhecidos previamente, mas os coeficientes (A_0, A_k, B_k) precisam ser encontrados e geralmente são calculados computacionalmente.

A série de Fourier permite modelar múltiplos ciclos sazonais, ela precisa de poucos parâmetros mesmo para ciclos sazonais grandes, a sensibilidade da sazonalidade pode ser facilmente ajustada através da ordem e amplitudes dos componentes seno e cosseno.

O modelo conhecido como regressão harmônica consiste basicamente em incluir um conjuntos séries harmônicas para um ciclo sazonal como variáveis regressoras no modelo de regressão dinâmica. Desta forma, a sazonalidade é tratada como processo estático, sendo importante estabilizar a amplitude sazonal antes da utilização desse modelo, o que geralmente é feito utilizando a transformação de Box-Cox.

2.2.4 Modelo TBATS

Podemos interpretar os termos da sigla do modelo TBATS da seguinte forma: T significa termos trigonométricos para sazonalidade, B para transformações Box-Cox para heterogeneidade, A para erros ARMA para dinâmica de curto prazo, T para tendência e por fim S para períodos sazonais.

O modelo TBATS foi proposto por De Livera, Hyndman e Snyder, 2011 e permite modelar séries temporais com diversos períodos sazonais não inteiros, tendências, transformação Box-Cox para melhorar a distribuição dos dados e a correção de erros pelo ARMA. Ele é baseado em uma compilação de termos de Fourier com um modelo de espaço de estados de suavização exponencial e uma transformação Box-Cox.

O modelo TBATS pode ser formulado pelo seguinte conjunto de equações:

- Transformação Box-Cox:

$$y_t^{(w)} = \begin{cases} \frac{y_t^w - 1}{w}, & w \neq 0 \\ \log y_t, & w = 0 \end{cases}$$

- Períodos sazonais:

$$y_t^w = l_{t-1} + \phi b_{t-1} + \sum_{i=1}^M S_{t-m_i}^i + d_t$$

- Tendência global e local:

$$l_t = l_{t-1} + \phi b_{t-1} + \alpha d_t$$

$$b_t = (1 - \phi)b + \phi b_{t-1} + \beta d_t$$

- Erro ARMA(p,q):

$$d_t = \sum_{i=1}^p \phi d_{t-1} + \sum_{j=1}^q + \varepsilon_t$$

- Termos sazonais de Fourier:

$$S_t^i = \sum_{j=i}^{k_i} S_{j,t}^{(i)}$$

$$S_{j,t}^{(i)} = S_{j,t-1}^{(i)} \cos \lambda_j^i + S_{j,t-1}^{*(i)} \sin \lambda_j^i + \gamma_1^i d_t$$

$$S_{j,t}^{*(i)} = -S_{j,t-1}^i \sin \lambda_j^i + S_{j,t-1}^{*(i)} \cos \lambda_j^i + \gamma_2^i d_t$$

em que y_t representa a série temporal, $y_t^{(w)}$ indica a transformação Box-Cox da série y_t , l_t representa o nível local no tempo t , b_t representa a tendência de curto prazo no tempo t , S_t^i é o i^a componente sazonal no tempo t , ϕ é o parâmetro de amortecimento da tendência, m_1, \dots, m_T

representa os períodos sazonais, d_t se refere a um processo ARMA(p,q), ε_t é o ruído branco com distribuição normal de média zero e variância constante σ^2 .

Os parâmetros de suavização são: $\gamma_1^i, \gamma_2^i, \alpha, \beta$ e γ_i , em que $i = 1, \dots, T$. O termo $\lambda_j^i = 2\pi/m_i$ determina o ciclo da série harmônica j da ciclo sazonal m_i . $S_{j,t}^{(i)}$ é o nível estocástico de i^a componente sazonal. A mudança no nível estocástico de i^a componente sazonal no decorrer do tempo é representada por $S_{j,t}^{*(i)}$. O componente k_i é basicamente o número de harmônicos preciso para i^a componente sazonal. Essa equação pode modelar simultaneamente múltiplos ciclos sazonais com periodicidades não inteiras.

O ordem do modelo é identificada na seguinte forma

$$TBATS(w, \phi, p, q, \{m_1, k_1\}, \{m_2, k_2\}, \dots, \{m_T, k_T\}),$$

em as parcelas $\{m_i, k_i\}$ denotam por m_i o tamanho do ciclo sazonal e por k_i número de pares de séries harmônicas utilizados para modelar a i -ésima sazonalidade.

Assim como os modelos anteriores, a estimação do TBATS é feita geralmente por mínimos quadrados, o qual coincide com o estimador de máxima verossimilhança sob a hipótese do termo de erro ter distribuição Normal.

Este modelo não possui formas fechadas para os cálculos do valor esperado condicional e da variância condicional, de modo que previsões geralmente são calculadas utilizando simulações de trajetórias futuras do processo e tomando o estimador de Monte Carlo.

Capítulo 3

Metodologia

Os dados usados neste projeto foram disponibilizados por uma empresa varejista de moda do Brasil. Esses dados foram obtidos após pedido formal feito para a empresa.

Nas seções a seguir apresentamos em detalhes o conjunto de dados, tratamentos e geração das variáveis de calendário, bem como apresentamos os experimentos e as métricas utilizadas nesse estudo.

3.1 Conjunto de dados

3.1.1 Dados de venda

O banco de dados inclui registros de vendas semanais e informações relacionadas a 540 produtos comercializados no Brasil, distribuídos entre 11 lojas físicas e online. Essas lojas estão localizadas em diversos estados do país, e a hierarquia dos produtos é consistente em todas elas. Os dados abrangem o período de janeiro de 2018 a setembro de 2022, fornecendo um histórico de 4 anos para análises e previsões. O arquivo é composto por 9 colunas e 1.018.646 linhas. Algumas variáveis, como *stock*, presentes nesse banco de dados, serão utilizadas como variáveis exógenas durante a modelagem. A Figura 3.1 exibe uma representação visual de algumas linhas do arquivo.

sku	loja	item_parent	subclass	semana	venda	estado	stock	cluster_clima
550636290	478	550636222	6189.0	2021-07-17	0.0	PR	2.0	RIO
539800431	478	539800415	6793.0	2019-05-11	0.0	PR	6.0	RIO
550636290	478	550636222	6189.0	2020-03-14	0.0	PR	1.0	RIO
539800431	478	539800415	6793.0	2020-10-24	1.0	PR	6.0	RIO
550636290	478	550636222	6189.0	2021-12-18	0.0	PR	2.0	RIO
...
541018668	435	541018650	6827.0	2021-03-27	0.0	SP	22.0	SPO
541018668	435	541018650	6827.0	2021-12-04	3.0	SP	46.0	SPO
541018668	435	541018650	6827.0	2022-07-02	0.0	SP	7.0	SPO
541018668	435	541018650	6827.0	2022-03-19	2.0	SP	16.0	SPO
541018668	435	541018650	6827.0	2021-03-13	0.0	SP	18.0	SPO

Fonte: (Autor)

Figura 3.1: Imagem da base de dados

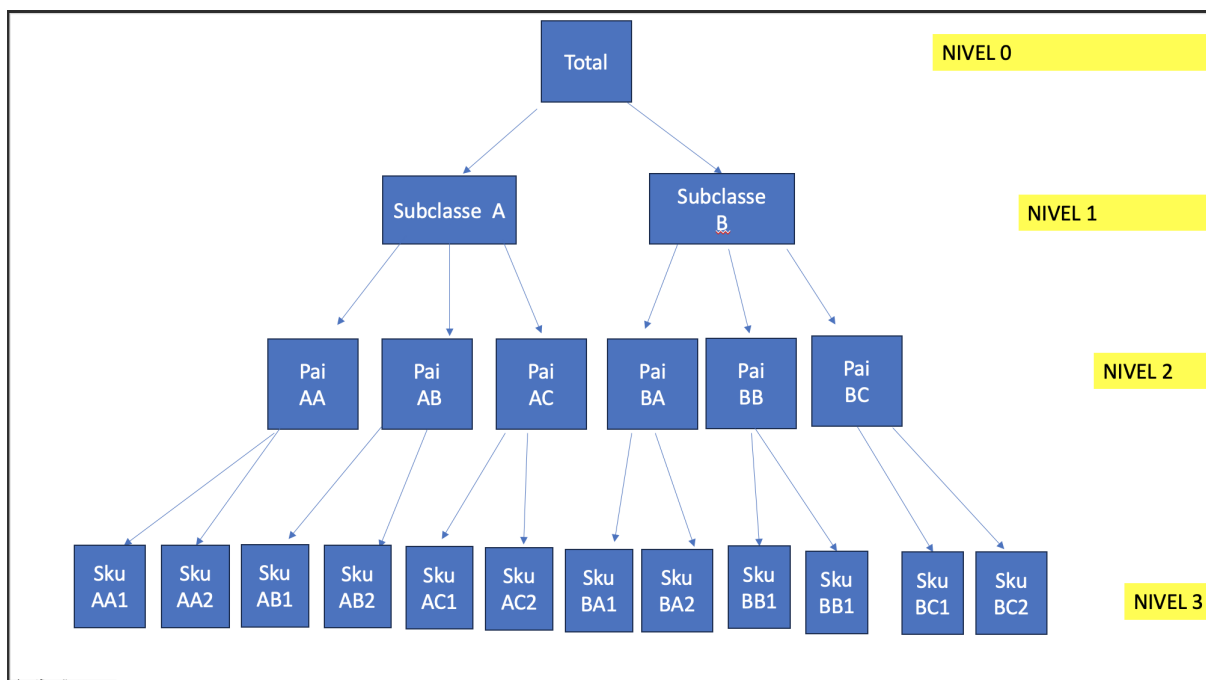
Segue abaixo o significado das colunas mais importantes :

- *sku* : Essa sigla significa *Stock Keeping Unit* a qual pode ser traduzido em português como “unidade de manutenção de estoque”. É o produto do nível mais baixo da hierarquia;
- *loja*: essa variável representa o código da loja;
- *item-parent*: essa coluna contém o código dos itens-pai. O "item-parent" é geralmente um agrupamento de sku;
- *semana*: essa variável representa a semana da venda realizada;
- *venda*: é a quantidade de venda realizada por semana;
- *estado*: é o estado onde a loja é localizada;

- *Stock*: A quantidade de itens disponíveis no estoque da loja;
- *Subclass*: é o código da subclasse. Considerando os dados disponíveis no arquivo, subclasse é o nível mais alto da hierarquia;
- *cluster-clima*: Indica o agrupamento das lojas segundo critérios climáticos.

No topo da hierarquia temos o total de venda de todos os produtos da empresa. O total é dividido em vários grupos chamado *subclasse*, na base de dados temos 5 subclasses. Cada *subclasse* é dividida em subgrupos chamados *item-pai*, que por sua vez podem ser desagregados em *sku*. Os *skus* são encontrados no nível mais baixo da hierarquia. Desta forma, o conjunto de dados possui 74 séries temporais de *item-pai* e 540 séries temporais de *sku*, 55 séries temporais no nível *subclasse/loja*, 814 no nível *item-pai/loja* e 5940 no nível *sku/loja*.

A Figura 3.2 apresenta um esboço da hierarquia dos dados:

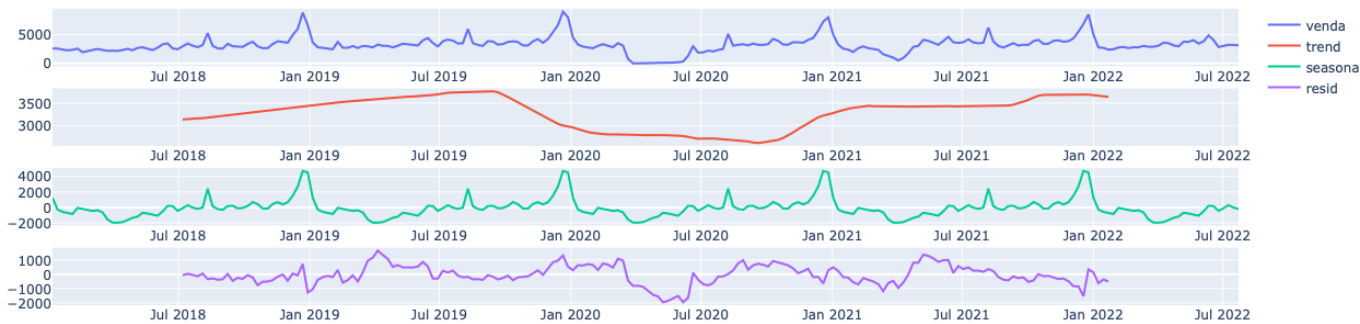


Fonte: (Autor)

Figura 3.2: Hierarquia dos dados

Para entender melhor os dados seria necessário realizar a decomposição das séries em tendência, sazonalidade e resíduos em cada nível da hierárquica. Foram realizados 5 gráficos e cada gráfico apresenta na seguinte ordem a representação da série, da tendência, da sazonalidade e dos resíduos.

A Figura 3.3 abaixo apresenta a decomposição STL da série *Total*, mostrando assim o volume de vendas no período em estudo.



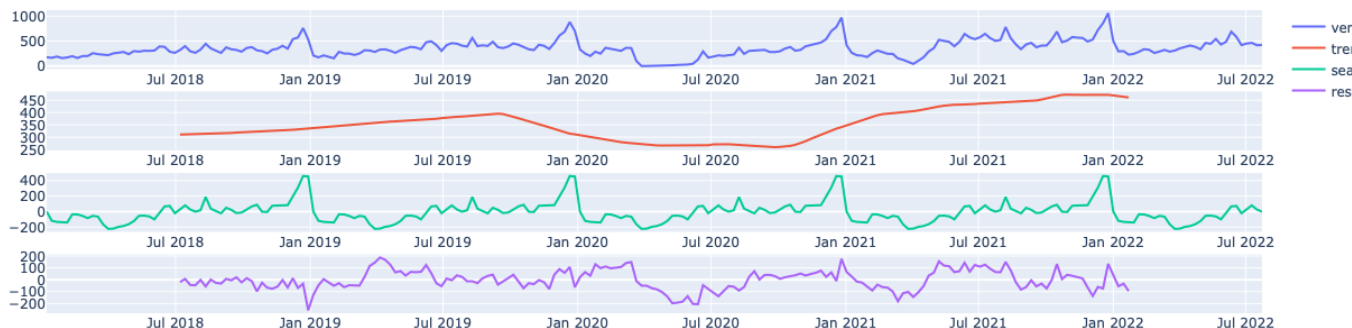
Fonte: (Autor)

Figura 3.3: Decomposição da série de total de venda

Analisando a Figura 3.3, podemos perceber que o gráfico de tendência apresenta um crescimento ao longo dos anos. A queda observada na curva da tendência ao longo de 2020 foi causada pelo fechamento das lojas no período do COVID19. Observando a curva da sazonalidade podemos perceber alguns picos sazonais nos meses de julho e de dezembro/janeiro. Geralmente nesses meses a empresa realiza várias promoções provocando o aumento considerável das vendas.

A Figura 3.4 apresenta a decomposição de uma série de venda de uma *subclasse* em todas as lojas.

O nível subclasse representa o nível 1 da hierarquia dos dados. Podemos perceber que a

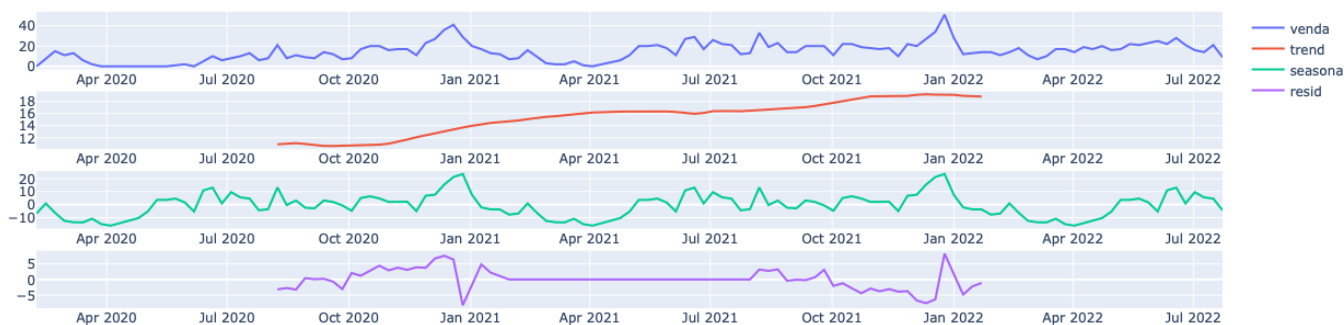


Fonte: (Autor)

Figura 3.4: Decomposição da série de venda de uma subclasse

Figura 3.4 apresenta quase as mesmas características que a Figura 3.3. As vendas no nível subclasse são menores que as vendas no nível 0 da hierarquia. Observando a curva da tendência é possível perceber que os picos consideráveis estão nos meses de dezembro/janeiro.

A Figura 3.5 mostra a decomposição de uma série de venda de um *Item-parent* em todas as lojas. O nível *Item-parent* representa o nível 2 da hierarquia dos dados.

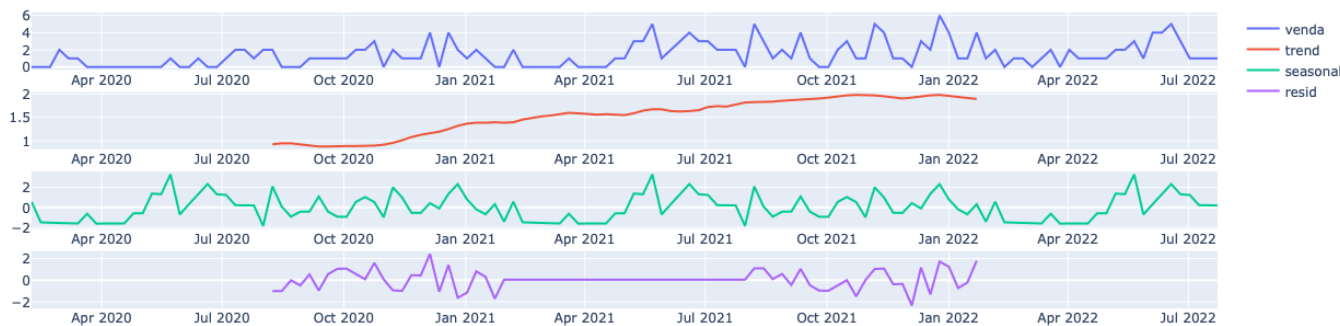


Fonte: (Autor)

Figura 3.5: Decomposição de uma série de venda de um *Item-parent*

Analisando a Figura 3.5, podemos perceber que o gráfico de tendência apresenta um crescimento ao longo dos anos. Observando a curva da sazonalidade podemos perceber alguns picos sazonais nos meses de dezembro/janeiro. As vendas no nível 2 da hierarquia são bem menores que as vendas nos outros níveis da hierarquia. O pico máximo foi de 40 sendo que o pico máximo no nível 0 foi de 5000.

A Figura 3.6 mostra a decomposição de uma série de venda de um *sku* em todas as lojas.



Fonte: (Autor)

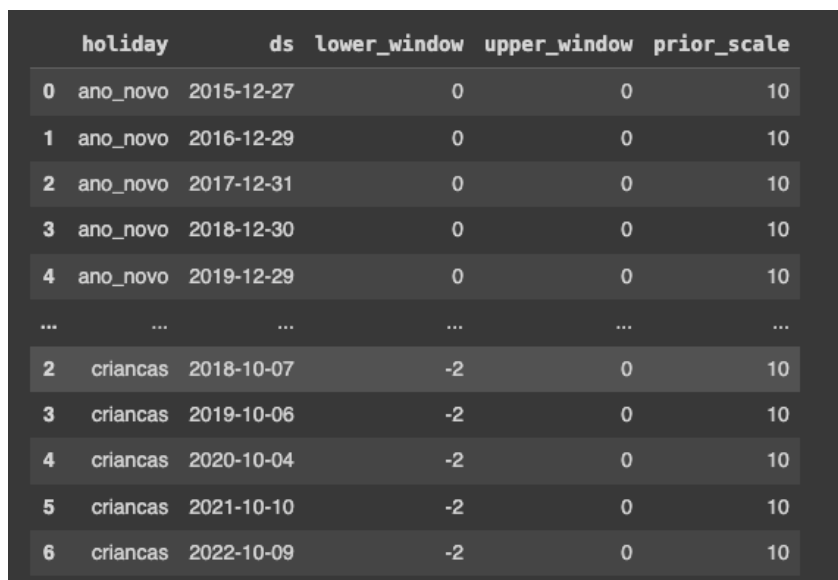
Figura 3.6: Decomposição de uma série de venda de um sku

O nível *sku* representa o nível 3 da hierarquia dos dados. As vendas nesse nível são as mais baixas da hierarquia, a venda máxima foi de 6 peças. As séries apresentam bastante ruídos, sazonalidade menos clara. A modelagem nesse nível é mais difícil de ser realizada por causa das características das séries. Podemos perceber que o gráfico de tendência apresenta um crescimento ao longo dos anos.

3.1.2 Dados de feriados para o modelo Prophet

O modelo Prophet será utilizado neste estudo como referência, sendo notável por sua capacidade de lidar com datas comemorativas do calendário gregoriano. Ele apresenta uma estrutura que facilita a incorporação de variáveis exógenas na modelagem dos dados, visando gerar previsões

mais precisas. O arquivo foi criado considerando o padrão de *dataframe* de feriados compatível com a estrutura do modelo. Possui 71 linhas e 5 colunas, abrangendo o período de dezembro de 2015 a setembro de 2022. A Figura 3.7 ilustra um resumo da base de dados.



	holiday	ds	lower_window	upper_window	prior_scale
0	ano_novo	2015-12-27	0	0	10
1	ano_novo	2016-12-29	0	0	10
2	ano_novo	2017-12-31	0	0	10
3	ano_novo	2018-12-30	0	0	10
4	ano_novo	2019-12-29	0	0	10
...
2	criancas	2018-10-07	-2	0	10
3	criancas	2019-10-06	-2	0	10
4	criancas	2020-10-04	-2	0	10
5	criancas	2021-10-10	-2	0	10
6	criancas	2022-10-09	-2	0	10

Fonte: (Autor)

Figura 3.7: Imagem das variáveis exógenas do modelo Prophet

Segue a interpretação das variáveis do arquivo:

- *holiday*: é o nome do feriado, ou do evento comemorativo;
- *ds*: esta coluna representa a semana do feriado;
- *lower-window*: esta coluna indica o ponto de referência inicial, ela é usada para dar mais flexibilidade ao modelo para ele incluir outras datas na modelagem além da data do feriado. Defini-lo como zero consideraria apenas o data do evento. Para incorporar o dia anterior ao evento, será necessário definir como -1, 2 dias antes do evento o parâmetro será igual a -2 e assim por diante;
- *upper-window*: esta variável se refere ao ponto de referência final, ela é igual 0 se apenas a data do feriado será considerada. Para incluir a data seguinte o parâmetro será igual a

1, 2 se precisar incluir 2 datas seguintes e assim por diante;

- *pior-scale*: é o parâmetro de escala, ele é definido separadamente para feriados individuais, ele regula a força e o tamanho do efeito do regressor. Por padrão ele é definido como 10.

3.1.3 Dados de feriados para o modelo Arima

Um conjunto de variáveis exógenas foi elaborado, levando em consideração as datas comemorativas mais relevantes para o setor varejista. Cinco eventos ao longo do ano foram contemplados: Carnaval, Páscoa, Dia dos Pais, Dia das Mães e Black Friday, além de uma variável indicando períodos de *lockdown* por conta da pandemia do COVID-19. Os dados cobrem o período de janeiro de 2018 a setembro de 2022, totalizando 238 entradas e 11 colunas. Na Figura 3.8, é apresentada uma representação visual desse arquivo.

	ds	carnaval	pascoa	pais	maes	black	carnaval_ves	pascoa_ves	maes_ves	pais_ves	covid
0	2018-01-07	0	0	0	0	0	0	0	0	0	0
1	2018-01-14	0	0	0	0	0	0	0	0	0	0
2	2018-01-21	0	0	0	0	0	0	0	0	0	0
3	2018-01-28	0	0	0	0	0	0	0	0	0	0
4	2018-02-04	0	0	0	0	0	1	0	0	0	0
...
233	2022-06-26	0	0	0	0	0	0	0	0	0	0
234	2022-07-03	0	0	0	0	0	0	0	0	0	0
235	2022-07-10	0	0	0	0	0	0	0	0	0	0
236	2022-07-17	0	0	0	0	0	0	0	0	0	0
237	2022-07-24	0	0	0	0	0	0	0	0	0	0

Fonte: (Autor)

Figura 3.8: Imagem das variáveis exógenas do modelo ARIMA

Abaixo listamos a descrição das variáveis que aparecem no arquivo:

- *ds*: esta coluna representa a semana comercial, geralmente é o domingo da semana;

- *carnaval*: esta variável representa a semana do Carnaval, ela assume o valor 1 quando estamos na semana do evento e 0 quando estamos fora;
- *Pascoa*: esta coluna indica a semana de Pascoa, ela é igual a 1 quando estamos na semana de pascoa e 0 quando estamos fora;
- *maes*: esta variável representa a semana de comemoração do dia das mães, ela assume 1 na semana do evento, e 0 fora;
- *pais*: É a coluna indicadora do dia dos Paes, essa variável é 1 quando está na semana do dia dos Paes, e 0 quando está fora;
- *black*: esta variável representa o Black Friday. A semana do Black Friday costuma registrar um aumento de vendas. A variável é igual a 1 quando está na semana do Black Friday, e 0 caso contrário;
- *Covid*: Essa coluna indica se naquela semana as lojas estavam fechadas por causa do covid; Ela assume o valor 1 se as lojas estavam fechadas, e 0 caso contrário.

As colunas terminando por “ves” são variáveis indicadoras da semana véspera da semana do evento comemorativo. Ela assume o valor 1 quando está na véspera, e 0 quando está fora. Essa variável é importante para o modelo pelo fato que o aumento das vendas ocorre geralmente uma semana antes da semana do evento.

3.2 Algumas técnicas utilizadas nos experimentos

3.2.1 A Validação Cruzada

A validação cruzada é um dos métodos de reamostragem de dados mais amplamente utilizados para avaliar, comparar e ajustar parâmetros de modelos de aprendizado de máquina, além de prevenir overfitting (Hyndman e Athanasopoulos, 2018). Na construção do modelo, a validação

cruzada é empregada para obter uma estimativa do desempenho do modelo final em dados futuros.

Dentre as técnicas de reamostragem, a mais simples é a abordagem de validação única, que divide aleatoriamente os dados em conjuntos de treinamento e validação. Geralmente, o conjunto de treinamento usado para ajustar o modelo representa 80% dos dados, enquanto 20% são reservados para avaliar o desempenho do modelo. Os resultados podem variar significativamente dependendo da escolha do conjunto de teste, e a conclusão pode ser afetada se uma janela de validação inadequada for selecionada. No caso de séries temporais, a escolha de uma data específica ou de um ponto fixo no tempo pode ser feita para realizar a divisão entre treino e teste.

Outra técnica de validação cruzada amplamente utilizada é a validação por janela deslizante. Neste método, várias janelas de treinamento e teste são previamente determinadas. O modelo é ajustado de forma iterativa nos diferentes conjuntos de treino e, em seguida, avaliado, de modo que em cada iteração, um conjunto é utilizado para validar o modelo. Dessa forma, cada horizonte de previsão é avaliado várias vezes. Esse procedimento de avaliação do modelo usando várias janelas fornece estimativas mais robustas do desempenho do modelo, agregando métricas sobre vários horizontes e compensando as flutuações naturais das séries, auxiliando na escolha cuidadosa do modelo mais adequado.

A Figura 4.9 ilustra o processo de validação por janela deslizante, mostrando a divisão do conjunto de dados em cada iteração.

	1	2	3	...												n
Passo 1:								1	2	3	...	h				
Passo 2:									1	2	3	...	h			
Passo 3:										1	2	3	...	h		
...											1	2	3	...	h	
...												1	2	3	...	
...													1	2	3	
...														1	2	
...															1	



Fonte: (Slides do Prof. Dr. José Augusto Fiorucci)

Figura 3.9: Janela deslizante

Na figura, cada coluna representa um índice da série temporal, e cada linha representa uma iteração do algoritmo. Em cada linha, o modelo é ajustado na parte destacada em azul, e suas previsões são avaliadas na parte destacada em amarelo. Ao final da execução, as células numeradas correspondem ao número de passos à frente daquela previsão. Por exemplo, todas as células com valor igual a 1 representam previsões de um passo à frente, enquanto todas as células com valor igual a 2 correspondem a previsões de dois passos à frente, e assim por diante. Desta forma, um métrica pode ser empregada para resumir a magnitude dos erros conforme o número de passos da previsão. Nesse estudo, adotaremos a métrica conhecida como Erro Absoluto Médio (MAE), a qual pode ser definida como

$$MAE(i) = \frac{1}{n_i} \sum_{j=1}^{n_i} |y_{ij} - \hat{y}_{ij}|,$$

em que:

- y_{ij} é o valor real para o j-ésimo valor do i-ésimo horizonte de previsão;

- \hat{y}_{ij} é o valor da previsão do j -ésimo valor do i -ésimo horizonte de previsão;
- n_i é o número de previsões realizadas para o i -ésimo horizonte de previsão, onde $i = 1, \dots, h$.

A métrica MAE (Erro Absoluto Médio) é frequentemente utilizada devido à sua vantagem na interpretação intuitiva em termos de erros relativos, sendo muito útil na comparação de resultados em diferentes escalas.

3.2.2 Teste de ranking

Para evitar conclusões equivocadas, é crucial empregar testes estatísticos rigorosos para auxiliar na escolha do melhor método de previsão de séries temporais entre uma variedade de modelos.

A metodologia do teste de ranking foi apresentada por Koning et al., 2005. Esse teste foi desenvolvido para determinar se existe uma diferença estatisticamente significativa na precisão de vários modelos de previsão de séries temporais. Originalmente concebido para avaliar o desempenho de modelos propostos em uma competição de previsão chamada M3, o teste baseia-se nas classificações dos métodos de previsão e nas diferenças significativas entre eles.

Os rankings e as comparações múltiplas são conduzidos por meio de metodologias estatísticas não paramétricas, conhecidas por serem fáceis de compreender, aplicar e interpretar. Não há necessidade de pressupostos sobre a distribuição de probabilidade, e a identificação do melhor e pior modelo torna-se uma tarefa simples.

Considerando que existem K modelos de previsão ($k = 1, 2, \dots, K$) e N séries temporais ($n = 1, 2, \dots, N$), os K métodos são utilizados para modelar N séries em uma janela de previsão H ($h = 1, 2, \dots, H$). Para cada k e h , o erro de previsão é calculado e utilizado para criar os rankings.

A metodologia empregada para a comparação estatística dos diferentes métodos de previsão de séries temporais é o teste MCB (comparações múltiplas com o melhor teste). Em testes de comparação múltipla, quando a hipótese H_0 não é rejeitada, geralmente todas as hipóteses im-

plícitas em H_0 são aceitas. Caso contrário, quando a hipótese é rejeitada, o objetivo é determinar quais componentes também podem ser rejeitados.

O teste estatístico pode ser descrito da forma seguinte (Koning et al., 2005):

$$H_0 : \tau_{k_1} = \tau_{k_2};$$

em que, $k_1 = 1, 2, \dots, k_2 - 1$, $k_2 = 1, 2, \dots, K$, e τ_k é o efeito do k -ésimo método. Os componentes da hipótese nula são rejeitados somente se:

$$|\bar{R}_{k_1} - \bar{R}_{k_2}| \geq r_{\alpha, K, N};$$

em que \bar{R}_k é a posição média do modelo k no ranking, $r_{\alpha, K, N} \approx q_{\alpha, k} \sqrt{\frac{K(K+1)}{12N}}$ é o valor crítico escolhido, e $q_{\alpha, k}$ é o ponto do percentil superior a α no intervalo de K variáveis normais padrão independentes.

Um gráfico pode ser usado para visualizar o processo de comparação múltipla de todos os modelos de previsões de séries temporais. O gráfico pode ser gerado seguindo as etapas seguintes:

- Construir um intervalo de tamanho $r_{\alpha, K, N}$ e centralizado em R_K para cada modelo usado (k).
- A hipótese nula H_0 é rejeitada em favor da hipótese alternativa se os intervalos de classificações para os modelos k_1 e k_2 não se sobrepõem.
- Desenhar uma reta como linha de referência no nível do limite superior do intervalo do melhor modelo, que conseqüentemente tem o menor limite superior. A linha de referência delimita a região dos melhores modelos. .
- Desenhar uma reta como linha de referência no nível do limite superior do intervalo do melhor modelo, que conseqüentemente tem o menor limite superior. A linha de referen-

cia delimita a fronteira entre os melhores modelos os os piores. Então os modelos com desempenho significativamente pior do que o melhor método, têm intervalos de confiança acima da linha de referência, e não devem ser usados.

3.3 Experimentos

Todas as análises, manipulações de dados e experimentos foram feitos no *software* estatístico *Rstudio*. Antes de iniciar a modelagem, foram feitas uma limpeza e transformações nos 3 conjuntos de dados. Dados faltantes foram preenchidos, por exemplo no conjunto de dados de vendas foi observada a ausência de algumas datas. Isto é devido ao fato que o sistema não registre nenhuma informação na base de dados nas datas que não há venda. Por esse fato, foi necessário preencher o arquivo com as datas faltantes e as informações de vendas (venda = 0). Algumas variáveis sofreram algumas transformações, por exemplo a coluna “venda” foi transformada no formato *integer*, e a coluna “semana” em formato *date*. É importante realizar esse procedimento antes da modelagem para que os dados sejam facies de interpretação para os modelos e para evitar a modelagem de dados com ruídos.

3.3.1 Modelagem no nível *subclasses* (nível 1)

Antes de iniciar a modelagem para as séries do nível 1, foi necessário gerar um *dataframe* contendo as informações das 5 subclasses do nível 1. Os dados foram agrupados por semana, subclasse e as vendas foram somadas, assim cada linha possui as informações da subclasse (numero da subclasse), da semana de venda e da quantidade de venda realizada naquela semana. Os dados de subclasses e de feriados foram concatenados na intenção de obter todas as informações em uma base única. A Figura 3.10 apresenta a estrutura dos dados usada para modelagem das 5 séries no nível 1.

ds	carnaval	pascoa	pais	maes	black	carnaval_ves	pascoa_ves	maes_ves	pais_ves	covid	subclass	sum_venda
2018-01-07	0	0	0	0	0	0	0	0	0	0	6189	176
2018-01-14	0	0	0	0	0	0	0	0	0	0	6189	164
2018-01-21	0	0	0	0	0	0	0	0	0	0	6189	189
2018-01-28	0	0	0	0	0	0	0	0	0	0	6189	158
2018-02-04	0	0	0	0	0	1	0	0	0	0	6189	171
2018-02-11	1	0	0	0	0	0	0	0	0	0	6189	196
2018-02-18	0	0	0	0	0	0	0	0	0	0	6189	161
2018-02-25	0	0	0	0	0	0	0	0	0	0	6189	199
2018-03-04	0	0	0	0	0	0	0	0	0	0	6189	200
2018-03-11	0	0	0	0	0	0	0	0	0	0	6189	257
2018-03-18	0	0	0	0	0	0	0	0	0	0	6189	237
2018-03-25	0	0	0	0	0	0	1	0	0	0	6189	230

Fonte: (Autor)

Figura 3.10: séries temporais do nível 1

O arquivo possui todas as informações necessárias para a modelagem das séries do nível 1. O objetivo desse experimento é testar vários modelos estatísticos nos dados e escolher o modelo com o melhor desempenho preditivo.

Foram testados ao todo 4 modelos, a saber: TBATS, ARIMA, ARIMA com variáveis exógenas e Regressão Dinâmica Harmônica. Vale ressaltar que as séries das subclasses podem apresentar múltiplas sazonalidades, e os modelos TBATS e Regressão Dinâmica Harmônica possuem capacidade e estrutura adequadas para lidar com esse tipo de dados

Para cada subclasse foram aplicados os 4 modelos da seguinte forma:

- **TBATS/Regressão Dinâmica Harmônica:**

Os dois modelos são adequados para modelar séries temporais com múltiplas sazonalidades. Os dados de subclasses podem apresentar sazonalidades mensal e anual, então o parâmetro de períodos sazonais do comando *msts* foi definido como *seasonal.periods=c(4.35, 52.18)*. Para cada subclasse foram criadas as séries de Fourier para modelar os múltiplos ciclos sazonais. As séries de Fourier entraram no modelo de Regressão Dinâmica como variáveis exógenas juntamente com as variáveis de calendário.

O modelo TBATS não recebe dados exógenas, então apenas os dados de vendas das subclasses entraram no modelo.

- **ARIMA/ARIMA com exogenas :**

Os modelos ARIMA e ARIMAX conseguem modelar apenas uma sazonalidade, então o parâmetro de períodos sazonais foi definido considerando somente uma sazonalidade ($seasonal.periods = c(52)$). O modelo ARIMAX foram incluídas as variáveis de calendário.

Afim de comparar esses modelos foi utilizado um estudo de desempenho preditivo por janela deslizante, no qual foi utilizada a função $tsCV()$ da biblioteca *forecast* (Hyndman et al., 2023). O estudo foi inicializado a partir da semana 156 e foi avaliado um horizonte de previsão de 12 semanas. Os resultado serão apresentados no capítulo 4.

3.3.2 Modelagem no nível *item-parent* (nível 2)

Inicialmente, foi necessário criar um *dataframe* com as informações das séries do nível 2. O arquivo foi criado agrupando os dados pelas colunas “semana” e “item-parent” e por fim somando as quantidades de vendas. O *dataframe* final contém os dados de item pai (numero de item-parent), a semana de vendas e a quantidade de item pai vendida naquela semana. O nível 2 tem 74 *item-parent* então 74 séries a serem modeladas, e algumas séries apresentam características de demanda intermitente.

Os experimentos no nível 2 envolveram os modelos seguintes: Prophet, Croston e suas variações, Regressão Dinâmica Harmônica com as abordagens *Top-down* e *reconciliação ótima*. O Prophet foi escolhido para fazer parte desse grupo por ser um modelo de *machine learning* e foi usado para comparar com os modelos estatísticos tradicionais. O modelo Croston é conhecido por lidar com séries intermitentes, e a Regressão Dinâmica Harmônica foi utilizada na intenção de testar as abordagens top-down e optimal reconciliation.

No objetivo de ter estimativas robustas, foi decidido realizar as previsões considerando 3 datas (2021-11-21, 2022-02-13, 2022-05-08), e uma janela de previsão de 12 semanas. Para cada série do *item-parent* foi aplicado cada modelo considerando as 3 datas de previsões, e a

métrica MAE foi usada para avaliar o desempenho dos modelos. Para cada previsão o MAE é calculado, armazenado em um arquivo e por fim foi usado o teste de ranking (subseção 3.2.2) para ranquear os modelos segundo as performances afim de escolher o melhor modelo.

As avaliações foram realizados da seguinte forma:

- **Regressão Dinâmica Harmônica:**

Antes de começar a modelar os dados, foi preciso criar as séries de Fourier, transformar as séries do nível 2 em matriz e depois em formato *hts*. O comando *hts* do *Rstudio* realiza essa tarefa e para isso é necessário informar no parâmetro *nodes* a lista contendo a estrutura hierárquica dos dados. As séries de Fourier e os dados de feriados são considerados no modelo como variáveis exógenas. As abordagens top-down e optimal reconciliation são indicadas ao modelo, respectivamente, pelos parâmetros *method = "tdfp"* e *method = "comb"* da função *forecast*.

- **Croston e suas variações:** A modelagem das séries dos itens pai com os modelos Croston foi realizada usando as funções *crost.ma* e *crost*. *crost.ma* é uma variação do método Croston, ela é baseada na média móvel com decomposição do método de Croston para séries de demanda intermitente com parâmetros fixos ou otimizados.

- **Prophet:** O modelo Prophet foi desenvolvido em 2018 pelo Facebook, no objetivo de prever dados diários com sazonalidade semanal e anual, além de efeitos de feriados. O modelo foi criado para ser rápido, preciso e fácil de usar. Ele é reconhecido por ter um desempenho impressionante com séries temporais com múltiplas sazonalidade e é um dos modelos de séries temporais mais usados para previsão de venda no varejo. O FBProphet basicamente atua modelando os dados de séries temporais como uma combinação de tendências, sazonalidade e componentes de ruído. Segundo (Hyndman e Athanasopoulos, 2018), o algoritmo Prophet pode ser considerado como uma regressão não linear descrita pela equação seguinte:

$$y_t = g(t) + s(t) + h(t) + \varepsilon_t.$$

em que,

- $\mathbf{g}(\mathbf{t})$, descreve uma tendência linear por partes. O componente de tendência captura os pontos de mudanças da série temporal. Os pontos são selecionados automaticamente se não foram especificados e um modelo de regressão linear por partes pode ser usado para modelar a tendência.
- $\text{textbfs}(\mathbf{t})$ captura os vários padrões sazonais. O componente sazonal é modelado usando a série de Fourier, facilitando assim a modelagem dos diferentes padrões sazonais. Por padrão, o parâmetro $\text{fourier.order} = 10$ é usado para sazonalidade anual e $\text{fourier.order} = 3$ é usado para sazonalidade semanal.
- $\mathbf{h}(\mathbf{t})$ descreve os efeitos do feriado. Os efeitos são acrescentados no modelo como variáveis dummy simples.
- ε_t representa o erro do ruído branco. Esse componente captura todas as variações aleatórias presentes nos dados e que não podem ser descritas pelos componentes de tendência ou de sazonalidade.

A estimação do modelo FBProphet é feita usando a abordagem bayesiana para facilitar a seleção automática dos pontos de mudança e outras características do modelo.

Antes de usar o modelo Prophet é importante ajustar alguns parâmetros, em especial:

- $n.\text{changepoints}$: geralmente, as séries temporais apresentam mudanças inesperadas nos seus históricos, por meio desse parâmetro o modelo é capaz de detectar os pontos de mudança e permitirá que a tendência se adapte adequadamente;
- changepoint.range : representa a proporção do histórico dos dados em que os pontos de mudança de tendência serão estimados;
- $\text{changepoint.prior.scale}$: por meio desse parâmetro é possível regular a sensibilidade do modelo em relação á tendência. Se $\text{changepoint.prior.scale}$ for um valor

muito alto, o modelo será muito flexível na modelagem da tendência, e o histórico será bem ajustado. Isso pode impactar negativamente nas previsões. O ideal é escolher um valor adequado, nem muito alto e nem muito baixo;

- *weekly.seasonality* : indica a sazonalidade semanal;
- *yearly.seasonality* : indica a sazonalidade anual;
- *holidays* : recebe os dados de feriados.

A variável do *lockdown* relacionada ao COVID19 foi introduzida no modelo como variável regressora. A função *fit.prophet* foi usada para ajustar o modelo, e as previsões foram feitas pela função *predict*.

Capítulo 4

Resultados

Neste capítulo, os resultados dos experimentos serão apresentados e discutidos.

4.1 séries do nível *subclasses* (nível 1)

Conforme descrito na subseção 3.3.1, apresentamos aqui os resultados do estudo de desempenho preditivo por janela deslizante para as séries temporais do nível de *subclasses*.

Como uma ilustração, a Figura 4.1 apresenta os resultados dos 4 modelos para uma das 5 séries do nível de *subclasses* (subclasse 6189). Os valores do MAE de cada modelo para cada um dos 12 horizontes de previsões estão nas colunas do arquivo. Os erros do modelo TBATS estão na primeira coluna, os erros do modelo ARIMA estão na segunda coluna, os da Regressão Harmônica Dinâmica estão na terceira coluna e, por fim, os MAEs do modelo ARIMA com variáveis exógenas estão na última coluna.

	tbats	arima	reg	exo
1	138.8568	117.4256	114.1406	122.1705
2	174.1597	163.8853	150.2451	183.8378
3	190.2086	198.5918	158.4893	204.9689
4	194.3716	224.4760	166.3646	214.5525
5	221.6298	247.4546	174.6508	220.4999
6	228.2043	279.7302	174.7769	223.1888
7	235.1579	290.1559	180.8054	225.5521
8	247.4679	307.3765	184.3555	225.8759
9	235.0635	327.1739	184.2289	224.7076
10	255.9480	347.9059	184.5200	222.9249
11	252.1117	359.1518	181.2551	218.5613
12	242.8812	378.7024	176.9929	212.1722

Fonte: (Autor)

Figura 4.1: Erros dos modelos na subclasse 6189

Para simplificar a visualização dos resultados, a Figura 4.2 apresenta graficamente o desempenho dos modelos nas *subclasses* 6732, 6189, 6583, 6793 e 6827. Utilizamos quatro modelos distintos: ARIMA (em vermelho), TBATS (em preto), Regressão Dinâmica Harmônica (em verde) e ARIMAX (em azul).

Na Figura 4.2a, observamos que na *subclasse* 6732, o ARIMA teve o desempenho menos favorável, enquanto o ARIMAX se destacou como o mais eficaz, seguido pela Regressão Dinâmica Harmônica.

Os resultados para a *subclasse* 6189, conforme ilustrado na Figura 4.2b, indicam um desempenho menos favorável do ARIMA, com TBATS e Regressão Harmônica Dinâmica apresentando os melhores resultados.

A Figura 4.2c reflete a avaliação na *subclasse* 6583, destacando que TBATS e Regressão Dinâmica Harmônica obtiveram os menores erros, enquanto que o ARIMA teve o pior desem-

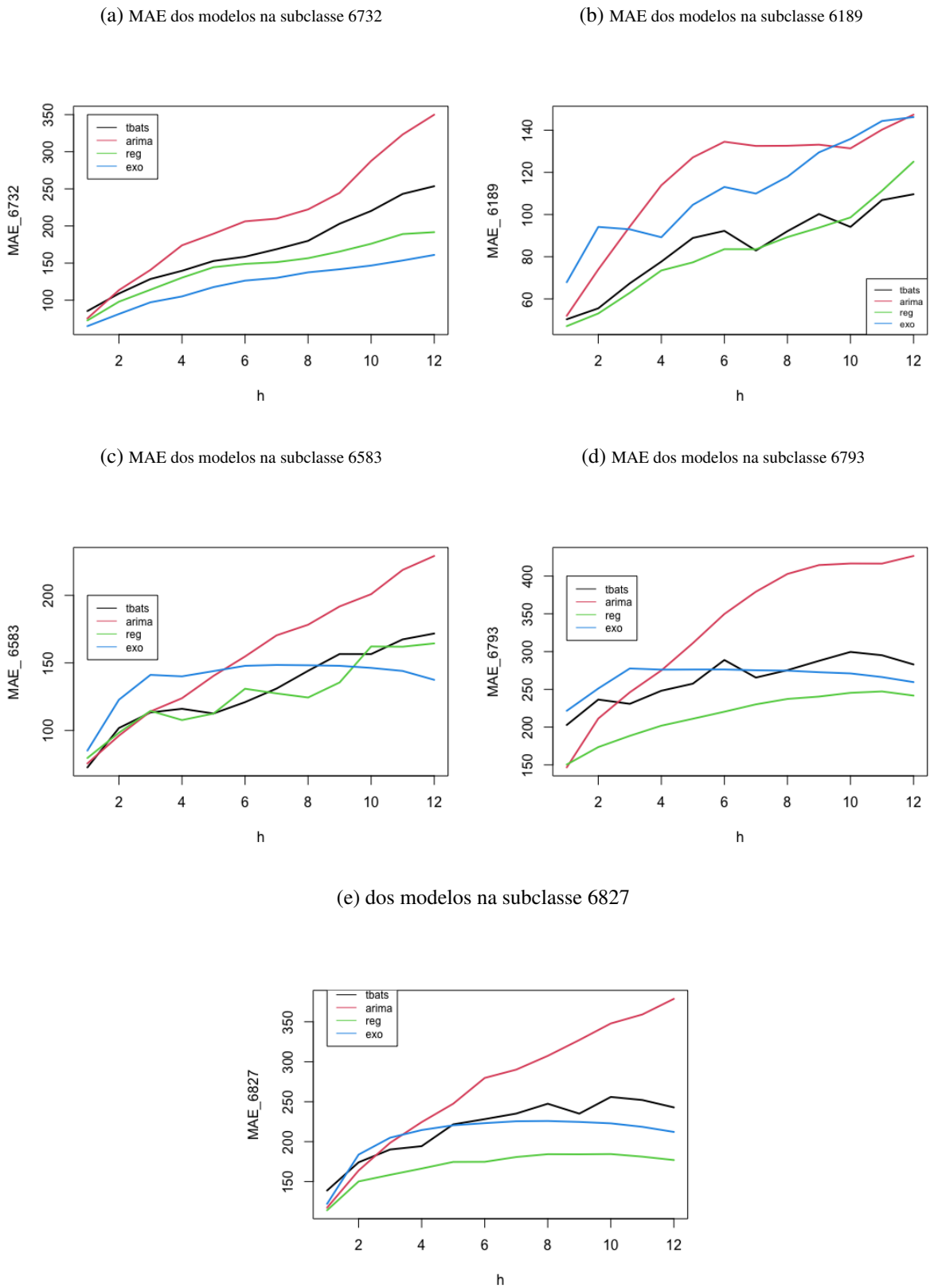
penho.

No contexto da *subclasse* 6793 (Figura 4.2d), a Regressão Harmônica Dinâmica gerou as melhores previsões, enquanto o ARIMA teve o MAE mais alto.

Finalmente, na *subclasse* 6827 (último gráfico), notamos que a Regressão Harmônica Dinâmica ajustou-se de maneira mais eficaz aos dados.

Esses resultados proporcionam uma avaliação abrangente do desempenho dos modelos em diferentes *subclasses*, facilitando a escolha do modelo mais adequado para cada caso.

Figura 4.2: Desempenho dos modelos



Em geral, o ARIMAX e a Regressão Dinâmica Harmônica mostraram-se consistentemente mais acurados que os demais modelos, destacando-se como opções preferenciais para previsões precisas. No entanto, para a *subclasse* de código 6189, o ARIMAX teve resultado fraco, ao longo que a Regressão Dinâmica Harmônica mostrou-se consistente para todas as séries temporais. Por conta disso, no segue, adotaremos a Regressão Dinâmica Harmônica como o modelo oficial desse nível hierárquico.

4.2 séries do nível *item-parent* (nível 2)

Conforme descrito na subseção 3.3.2, apresentamos aqui os resultados do estudo de desempenho preditivo para as 74 séries temporais do nível de *item-parent*. Este estudo contempla os modelos Prophet, Croston, Regressão Dinâmica Harmônica com as abordagens *Top-down* e reconciliação ótima.

A Figura 4.3 apresenta uma amostra dos dados usados para gerar o gráfico de teste de ranking para a primeira janela de previsão (21/11/2021):

- mae-pro : é referente ao MAE do modelo Prophet;
- item-pai: representa o numero da série temporal(o numero do produto item pai);
- ds : é a data de corte para janela de previsão;
- mae-cro-ma : representa o MAE da variação do modelo Croston;
- mae-cro: refere-se ao MAE da do modelo Croston;
- mae-cmb : trata-se do MAE da abordagem *optimal reconciliation*;
- mae-top : refere-se ao MAE da abordagem top-down.

O gráfico do teste de ranking foi feito para cada data, e outro gráfico foi gerado para todas as 3 datas somando assim 4 gráficos. Os gráficos apresentam os rankings dos modelos com base

↕	mae_pro ↕	item_pai ↕	ds ↕	mae_cro_ma ↕	mae_cro ↕	mae_cmb ↕	mae_top ↕
1	23.7145	502030066	2021-11-21	28.333333	28.495097	21.9646633	22.0046078
2	9.5712	502969719	2021-11-21	15.416667	15.764121	10.1607439	10.3831106
3	22.1584	507921511	2021-11-21	25.657631	25.537121	17.9216766	17.4285209
4	16.1541	509553530	2021-11-21	26.675414	24.764880	10.6984490	10.5328941
5	26.0969	516276761	2021-11-21	30.601852	30.357377	20.4884406	20.2827267
6	81.2633	519879980	2021-11-21	112.377451	111.804863	61.8881668	59.6908140
7	38.0970	519940293	2021-11-21	51.370550	53.591445	59.1383355	57.7587658
8	38.0773	522688800	2021-11-21	31.802857	31.734079	26.9037172	29.7249037
9	12.6989	528808809	2021-11-21	16.938713	16.709133	9.0823134	8.9581183
10	31.6956	529347493	2021-11-21	21.916667	22.019881	41.1925843	40.0446164
11	38.6811	530136474	2021-11-21	42.083333	42.083333	28.6641121	28.6937164
12	14.4989	530367053	2021-11-21	10.083333	10.835952	12.6176607	12.5518646
13	121.4799	530466318	2021-11-21	63.500000	63.500000	86.9514574	84.2608437
14	19.5221	531875682	2021-11-21	19.078231	18.307663	22.2808451	22.0604415
15	25.8724	533175567	2021-11-21	30.587674	30.474526	22.7283883	22.9494255
16	12.0787	533406679	2021-11-21	12.583333	12.583333	11.3624214	11.7565542
17	43.4955	533750200	2021-11-21	8.916667	13.388281	8.5932269	8.4890514

Fonte: (Autor)

Figura 4.3: Imagem dos dados para o ranking

no Erro Absoluto Médio (MAE). A media dos ranking é apresentada no eixo Y e os nomes dos modelos estão no eixo da seguinte maneira:

- mae-top: refere-se à abordagem top-down;
- mae-cro: refere-se ao modelo Croston;
- mae-cmb: trata-se da abordagem reconciliação ótima;
- mae-cro-ma: representa a variação do modelo Croston;
- mae-pro: É referente ao modelo Prophet.

Os gráficos do teste de ranking para cada data considerada estão apresentados na Figura 4.4. Na Figura 4.4a, a abordagem reconciliação ótima teve o ranking médio, mas não difere

estatisticamente da abordagem *top-down*, no entanto foi estatisticamente melhor que os outros 3 outros modelos.

Para a segunda data (figura b), o modelo Prophet obteve o melhor desempenho, mas checou a é estatisticamente melhor que a reconciliação ótima. O *top-down*, mas foi estatisticamente melhor apenas do que o método de Croston.

Na ultima (figura c), as abordagens *Top-down*, reconciliação ótima e o modelo Prophet tiveram ranking médio muito próximos, com uma ligeira vantagem para o Croston, estes foram então estatisticamente equivalentes.

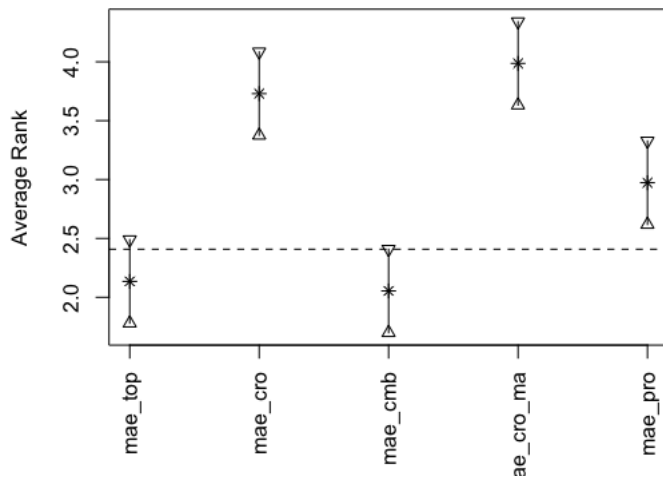
Observando os 3 gráficos, podemos destacar a abordagem *Top-down* e a reconciliação ótima como as mais robustas para esses cenários estudados, com ligeira vantagem para a reconciliação ótima, por ter sido a única que ficou entre as 3 melhores em todas as datas.

Figura 4.4: Desempenho dos modelos no nível 2

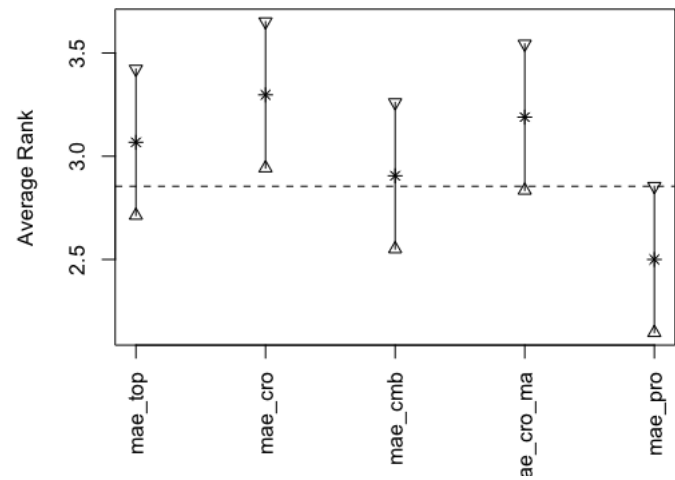
(a) ranking dos modelos na primeira data

(b) ranking dos modelos na segunda data

Rank of MAE results for first week time series

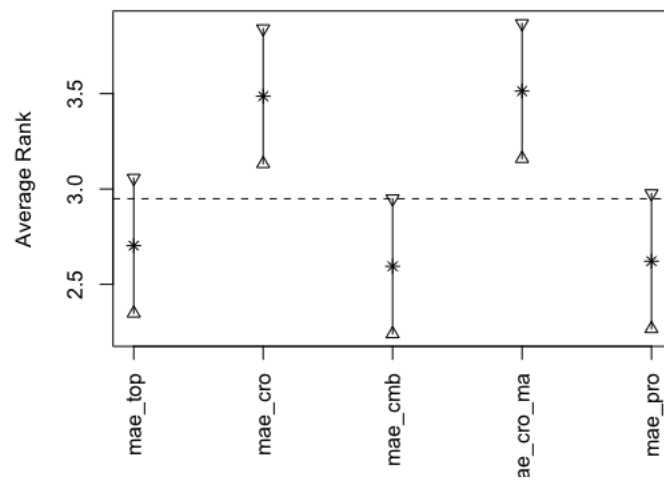


Rank of MAE results for second week time series

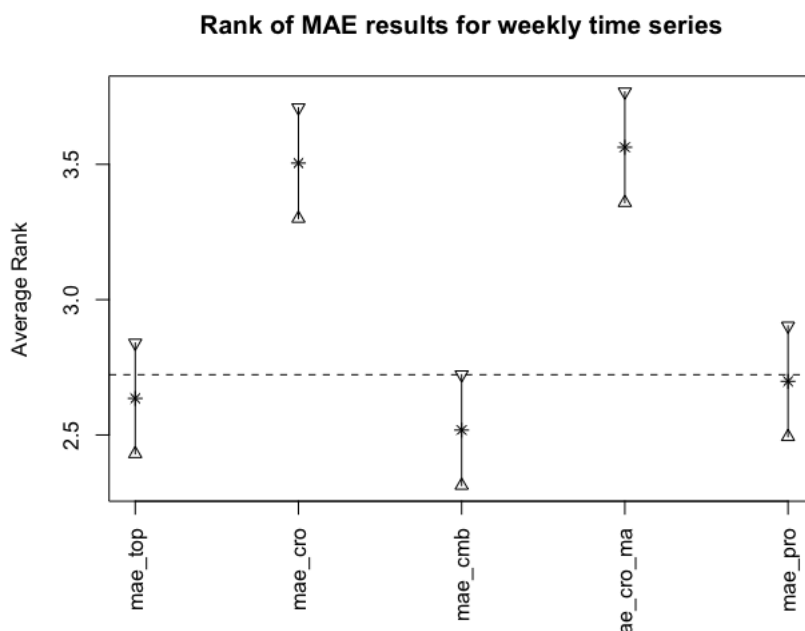


(c) ranking dos modelos na terceira data

Rank of MAE results for the third week time series



Para ter uma conclusão mais sólida foi necessário avaliarmos o gráfico ranking para as 3 datas em conjunto, o qual é apresentado na Figura 4.5.



Fonte: (Autor)

Figura 4.5: Gráfico do teste de ranking das 3 datas.

Examinando a Figura 4.5, a reconciliação ótima de fato apresenta o melhor ranking médio entre os modelos estudados, mas não chega a diferir estatisticamente das abordagens *Top-down* e *Prophet*. Os modelos Croston apresentaram os piores resultados, sendo estatisticamente inferior a reconciliação ótima.

Em resumo, os modelos Croston não entregaram resultados satisfatórios, tiveram os erros mais altos em todos os casos. O modelo Prophet e a abordagem *top-down* apresentaram um desempenho intermediário, em alguns casos se ajustaram bem nos dados e em outros casos não. A *reconciliação ótima* foi a abordagem que gerou as previsões com os menores erros, ela foi uma das mais bem ranqueadas em todos os experimentos com os dados do nível 2.

Capítulo 5

Conclusões

5.1 Considerações Finais

Na primeira parte deste estudo, apresentou-se a definição e o conceito de séries temporais hierárquicas e agrupadas, discutindo os diferentes métodos de resolução e como aplicar esses métodos em dados reais. Além disso, foram destacados os pontos positivos e negativos de cada abordagem, bem como as situações em que cada uma delas é mais adequada.

Na segunda parte, abordou-se modelos de previsões de séries temporais, detalhando as características e ajustes de vários modelos, como ARIMA, TBATS, ARIMA com variáveis exógenas, Regressão Harmônica Dinâmica e método Croston. Também discutiu-se em qual nível hierárquico cada modelo deve ser aplicado para obter as previsões mais precisas.

Após essa etapa, foram conduzidos experimentos de modelagem com os dados. Os modelos ARIMA, TBATS, ARIMA com variáveis exógenas e Regressão Harmônica Dinâmica foram ajustados nos dados do nível 1, e a validação cruzada e o MAE foram utilizados para avaliar o desempenho dos modelos, visando a escolha do método mais eficaz. Nas séries temporais do nível 2, os métodos Croston, o modelo Prophet e a Regressão Harmônica Dinâmica, com abordagem *Top-down* e abordagem reconciliação ótima, foram aplicados para identificar o modelo mais adequado. Para essa tarefa, foi necessário utilizar o teste estatístico de ranking para

classificar os modelos de acordo com seu desempenho durante o experimento.

Após os experimentos, o modelo de Regressão Harmônica Dinâmica obteve o melhor desempenho para as séries temporais do nível 1. Os dados desse nível podem apresentar vários padrões sazonais, e a inclusão da série de Fourier junto com os dados de variáveis exógenas provavelmente auxiliou o modelo a fornecer previsões mais precisas. No nível 2 da hierarquia, a abordagem de reconciliação ótima se destacou pelo seu desempenho, apresentando uma performance superior em quase todos os experimentos desse nível. Essa abordagem costuma fornecer previsões mais assertivas do que outras técnicas de resolução de séries temporais hierárquicas e agrupadas.

Como maior contribuição, pode-se destacar que o modelo de Regressão Harmônica Dinâmica com abordagem de reconciliação ótima formulado nesse estudo superou o modelo de *Machine Learning Prophet*, o qual atualmente é utilizado pela empresa que forneceu os dados para o planejamento de demanda.

Para estudos futuros, alguns pontos podem ser considerados:

- Testar o modelo Prophet nas séries do nível 1.
- Experimentar os métodos Croston nos dados do nível mais baixo.
- Explorar outros modelos de previsão de séries intermitentes para o nível mais baixo da hierarquia.
- Utilizar a abordagem reconciliação ótima para as séries do nível 3.
- Testar outras variáveis exógenas, como Estado e quantidade de estoque do item disponível na loja.
- Experimentar outros modelos de *Machine Learning*.

Referências Bibliográficas

- Box, George EP et al. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.
- Croston, John D (1972). “Forecasting and stock control for intermittent demands”. *Journal of the Operational Research Society* 23.3, pp. 289–303.
- De Livera, Alysha M, Hyndman, Rob J e Snyder, Ralph D (2011). “Forecasting time series with complex seasonal patterns using exponential smoothing”. *Journal of the American statistical association* 106.496, pp. 1513–1527.
- Dickey, David A e Fuller, Wayne A (1981). “Likelihood ratio statistics for autoregressive time series with a unit root”. *Econometrica: journal of the Econometric Society*, pp. 1057–1072.
- Hylleberg, Svend et al. (1990). “Seasonal integration and cointegration”. *Journal of econometrics* 44.1-2, pp. 215–238.
- Hyndman, Rob et al. (2023). *forecast: Forecasting functions for time series and linear models*. R package version 8.21.1. URL: <https://pkg.robjhyndman.com/forecast/>.
- Hyndman, Rob J e Athanasopoulos, George (2018). *Forecasting: principles and practice*. OTexts.
- Hyndman, Rob J, Athanasopoulos, George e Shang, Han Lin (2015). “hts: An R Package for Forecasting Hierarchical or Grouped Time Series”. *R Package (Available at <http://cran.unej.ac.id/web/packages/hts/vignettes/hts.pdf>)*.
- Koning, Alex J et al. (2005). “The {M3} competition: Statistical tests of the results”. *International Journal of Forecasting* 21 (3), pp. 397–409.
- Kwiatkowski, Denis et al. (1992). “Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?” *Journal of econometrics* 54.1-3, pp. 159–178.
- Ljung, Greta M e Box, George EP (1978). “On a measure of lack of fit in time series models”. *Biometrika* 65.2, pp. 297–303.
- Ord, Keith, Fildes, Robert Alan e Kourentzes, Nikolaos (2017). “Principles of business forecasting”.
- Taylor, Sean J e Letham, Benjamin (2018). “Forecasting at scale”. *The American Statistician* 72.1, pp. 37–45.
- Wang, Xiaozhe, Smith, Kate e Hyndman, Rob (2006). “Characteristic-based clustering for time series data”. *Data mining and knowledge Discovery* 13, pp. 335–364.
- Wickramasuriya, Shanika L, Athanasopoulos, George e Hyndman, Rob J (2015). “Forecasting hierarchical and grouped time series through trace minimization”. *Department of Econometrics and Business Statistics, Monash University* 105.

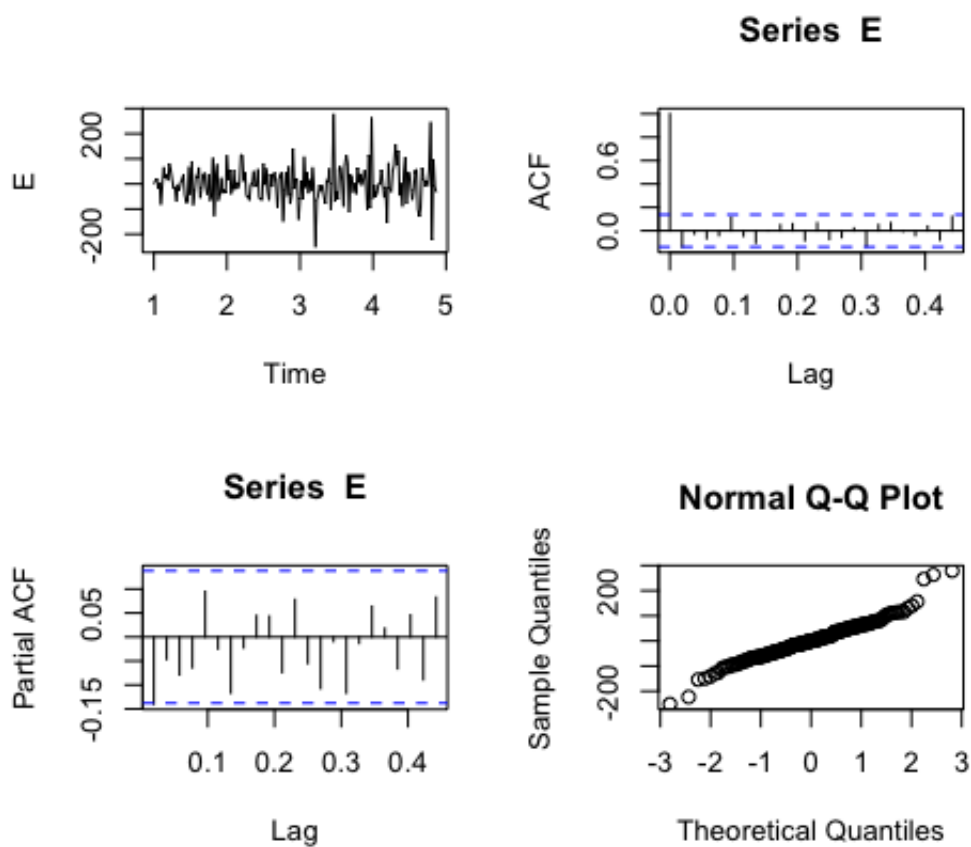
Wickramasuriya, Shanika L, Athanasopoulos, George e Hyndman, Rob J (2019). “Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization”. *Journal of the American Statistical Association* 114.526, pp. 804–819.

Apêndice - Análises dos resíduos modelos do nível 1

Neste apêndice , serão apresentados os resultados da análises dos resíduos. O objetivo dessa análise é verificar se o melhor modelo no nível 1 da hierarquia atende todos os critérios de adequação dos resíduos. Esse estudo foi realizado considerando "2021-11-21" como a data de corte dos dados em treino e teste. Com base no modelo ajustado nos dados de treino da série temporal da subclasse 6189, os resíduos do modelo foram extraídos para avaliar se eles atendem os critérios seguintes :

- Resíduos estacionários.
- Independência dos resíduos.
- Normalidade dos resíduos.
- Aleatoriedade dos resíduos.

A Figura 1 a seguir apresenta umas Análises visuais dos resíduos:



Fonte: (Autor)

Figura 1: Análises visuais dos resíduos.

Analisando a Figura 1, é possível perceber que:

- os resíduos têm um comportamento aleatório em torno de zero;
- ACF e PACF ficaram dentro da banda azul confirmando a hipótese de independência dos resíduos;
- no gráfico do qqplot, todos os pontos parecem seguir a linha oblíqua. É possível concluir que os resíduos apresentam um comportamento normal.

Com base nas análises visuais, é razoável afirmar que os resíduos atendem todos os critérios de adequação. Para confirmar e completar essas análises visuais, foi preciso realizar alguns testes hipóteses.

O **teste KPSS** foi usado para determinar se os resíduos são estacionários. O P-valor é igual a 0,1. Como este valor não é inferior a 0,05, então a hipótese nula do teste KPSS não foi rejeitada. Isso significa que é possível assumir que os resíduos são estacionários.

O **Teste de Box-Ljung** verifica se existe autocorrelação em uma série temporal. A estatística do teste é $Q = 14,509$ e o p-valor do teste é 0.1053, que é muito maior que 0,05. Assim, a hipótese nula do teste não foi rejeitada e concluiu-se que os resíduos são independentes. A tabela 1 mostra o resumo dos resultados dos testes estatísticos

Tabela 1: **Teste estatístico**

Teste	P-valor
KPSS Test	0.1
Box-Ljung test	0.1053