

EDUARDO OHASHI DE SOUZA

PROPOSTA DE MODELO PREDITIVO DE RECEITA PARA ESTABELECIMENTO DE  
METAS PARA COMÉRCIO ELETRÔNICO

São Paulo

2022



EDUARDO OHASHI DE SOUZA

PROPOSTA DE MODELO PREDITIVO DE RECEITA PARA ESTABELECIMENTO DE  
METAS PARA COMÉRCIO ELETRÔNICO

Trabalho de Formatura apresentado à Escola  
Politécnica da Universidade de São Paulo para  
obtenção do Diploma de Engenheiro de  
Produção.

Orientador: Professor Davi Noboru Nakano

São Paulo

2022

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

### **Catálogo-na-publicação**

Souza, Eduardo

PROPOSTA DE MODELO PREDITIVO DE RECEITA PARA  
ESTABELECIMENTO DE METAS PARA COMÉRCIO ELETRÔNICO / E. Souza  
- São Paulo, 2022.

111 p.

Trabalho de Formatura - Escola Politécnica da Universidade de São  
Paulo. Departamento de Engenharia de Produção.

1.Modelos Preditivos 2.Previsão de Demanda 3.Comércio Eletrônico  
4.Ciência de Dados I.Universidade de São Paulo. Escola Politécnica.  
Departamento de Engenharia de Produção II.t.

À minha mãe e ao meu pai por sempre acreditarem em mim. Minha família por me ensinar tanto e minha sobrinha por motivar a busca por um mundo melhor.



## **AGRADECIMENTOS**

Inicialmente, gostaria de agradecer a oportunidade e privilégio de poder ter estudado na Escola Politécnica da Universidade de São Paulo. Compreendo a importância que o ensino superior público possui para o avanço científico, econômico e democrático no Brasil. Por isso sinto-me honrado de poder ter feito parte dessa escola tão essencial para o País e ao Estado de São Paulo. Além disso, gostaria de enaltecer todos os professores que contribuíram para minha formação como cidadão e engenheiro ao longo dos últimos anos.

Meus profundos agradecimentos aos meus pais (Célia Miyako e Mario Luiz), irmãs (Camila Gabriela, Marina), sobrinhas (Ariane e Tainá) e demais familiares por sempre apoiarem as minhas decisões e torcerem pelo meu sucesso. Foram essenciais em todas as minhas conquistas e essa, definitivamente, não foi diferente.

Na universidade, o Centro Acadêmico da Engenharia de Produção (CAEP) foi essencial para o desenvolvimento de meu senso de responsabilidade e de trabalho em equipe, foi onde conheci pessoas (veteranos e calouros) que me ensinaram muito e permanecerão sempre em minhas memórias mais alegres. Destaque especial para a Cristina e Osni, que representam todo o sentimento de carinho que tenho por essa entidade. Agradecimentos aos meus amigos ingressantes de 2016, que dividiram comigo a experiência de gestão no CAEP com incontáveis aprendizados e memórias inesquecíveis. Daniela S, Gustavo N, Shelly B, Lucas A, Daniela C, Francisco E, Manuel P, Pedro B, Pedro C, Rafael A, Thiago T, Vinicius C e Victor M, muito obrigado pela jornada e que nossa amizade só cresça.

Na Poli pude fazer amizades com alunos de outras faculdades, engenharias e entidades acadêmicas que também pretendo levar daqui para sempre. Adriana W, Bruno C, Daniel H, Diana R, Henry C, Fernando M, Flávia B, Gabriel C, Gabriela H, Gustavo R, Mariana Z, Matheus V, Nicholas B, Olivia C, Paola S, Rafaela G, Valter B e Vitor E, obrigado por me acompanharem nestes anos tão marcantes.

Por fim, agradeço ao meu orientador Profº Davi Noboru Nakano por me acompanhar e aconselhar ao longo da graduação e no presente trabalho.



“We study history not to know the future but to widen our horizons, to understand that our present situation is neither natural nor inevitable, and that we consequently have many more possibilities before us than we imagine.”

Yuval Noah Harari



## RESUMO

O trabalho foi desenvolvido sob o contexto de uma empresa de cosméticos masculinos com operação no comércio eletrônico. O intuito do projeto é elaborar um modelo preditivo capaz de estimar valores de receita futuros a fim de melhorar o processo de estabelecimento de metas pela liderança da empresa. Não existe um método padronizado, baseado em dados, para definição dos valores de vendas previstos. Isso fez com que as metas fossem muito distorcidas em alguns períodos, fator que atrapalha a organização interna da empresa, além de gerar ansiedade e estresse entre os colaboradores. No estudo foram aplicados quatro métodos preditivos à base histórica da empresa e seus resultados foram comparados. Ao fim espera-se que o leitor seja capaz de compreender todas as etapas de desenvolvimento, testagem, validação, escolha e implementação do modelo preditivo final. O projeto foi construído com o intuito de ser incorporado como um projeto da área de Data-Science da organização.

**Palavras Chave:** Modelos Preditivos. Previsão de Demanda. Comércio Eletrônico. Ciência de Dados.



## ABSTRACT

This project was developed under the context of a male cosmetics company operating in e-commerce. The purpose of the analysis is to develop a predictive model capable of estimating future revenue values in order to improve the process of setting goals by the company's leadership. There is no standardized, data-based method for defining forecast sales figures. This meant that goals were very distorted in some periods, a factor that disrupts the company's internal organization, in addition to generating anxiety and stress among employees. In the study, four predictive methods were applied to the company's historical database and their results were compared. At the end, the reader is expected to be able to understand all stages of development, testing, validation, choice method and implementation of the final predictive model. The work was built with the intention of being incorporated as a project of the Data-Science sector of the organization.

**Key Word:** Predictive Models. Demand Forecasting. E-Commerce. Data-Science.



## **LISTA DE FIGURAS**

Figura 1: Exemplo suavização exponencial simples	34
Figura 2: Exemplo suavização exponencial dupla	35
Figura 3: Exemplo suavização exponencial de Holt-Winters	36
Figura 4: Ciclo de rotinas para manutenção do modelo preditivo	47



## LISTA DE TABELAS

Tabela 1: Base de dados simplificada	49
Tabela 2: Base de dados com distinção de origem de receita	49
Tabela 3: Erros entre previsões e receita faturada	53
Tabela 4: Erros para média móvel ( $p=3$ ) aplicado sobre a receita total	60
Tabela 5: Erros para média móvel ( $p=6$ ) aplicado sobre a receita total	60
Tabela 6: Erros para média móvel ( $p=9$ ) aplicado sobre a receita total	61
Tabela 7: Erros para média móvel em receita de canais grátis	63
Tabela 8: Erros para média móvel em receita de canais pagos	64
Tabela 9: Erros do modelo com regressão linear aplicada na previsão da receita total	68
Tabela 10: Erros do modelo com regressão linear aplicada na previsão da receita de origem gratuita	70
Tabela 11: Erros do modelo com regressão linear aplicada na previsão da receita de origem paga	72
Tabela 12: Erros do modelo com suavização exponencial simples aplicada na previsão da receita total	75
Tabela 13: Erros do modelo com suavização exponencial simples aplicada na previsão da receita de origem grátis	77
Tabela 14: Erros do modelo de suavização exponencial simples aplicada na previsão da receita de origem paga	79
Tabela 15: Erros do modelo de suavização exponencial de Holt-Winters aplicada na previsão da receita total	81
Tabela 16: Erros do modelo de suavização exponencial de Holt-Winters aplicada na previsão da receita gratuita	84
Tabela 17: Erros do modelo de suavização exponencial de Holt-Winters aplicada na previsão da receita paga	86
Tabela 18: Comparação de modelos: erros percentuais absolutos máximos	88
Tabela 19: Comparação de modelos: erros percentuais absolutos médios	88
Tabela 20: Comparação de modelos: erros percentuais médios	88
Tabela 21: Comparação de modelos aplicados na receita grátis: erros percentuais absolutos máximos	91

Tabela 22: Comparação de modelos aplicados na receita grátis: erros percentuais absolutos médios	91
Tabela 23: Comparação de modelos aplicados na receita grátis: erros percentuais médios	92
Tabela 24: Comparação de modelos aplicados na receita paga: erros percentuais absolutos máximos	93
Tabela 25: Comparação de modelos aplicados na receita paga: erros percentuais absolutos médios	93
Tabela 26: Comparação de modelos aplicados na receita paga: erros percentuais médios	94
Tabela 27: Comparação de erros: modelo de suavização exponencial simples e modelo modelos de suavização exponencial simples ajustado com regressão linear	100
Tabela 28: Erros obtidos pelo modelo final aplicado nos meses de Setembro e Outubro de 2022	102

## LISTA DE GRÁFICOS

Gráfico 1: Diferença percentual entre previsão e receita obtida	46
Gráfico 2: Receita por tipo de canal e investimento em marketing	51
Gráfico 3: Comparação entre receitas por tipo de canal	51
Gráfico 4: Receita por tipo de canal e investimento em marketing	52
Gráfico 5: Distribuição de receita por categoria de produto	52
Gráfico 6: Previsão da receita total com média móvel de três períodos	58
Gráfico 7: Previsão da receita total com média móvel de seis períodos	58
Gráfico 8: Previsão da receita total com média móvel de nove períodos	59
Gráfico 9: Erros percentuais para diferentes valores de p	59
Gráfico 10: Erros percentuais absolutos para diferentes valores de p	60
Gráfico 11: Erros percentuais para diferentes valores de p (canais grátis)	62
Gráfico 12: Erros percentuais absolutos para diferentes valores de p (canais grátis)	62
Gráfico 13: Erros percentuais para diferentes valores de p (canais pagos)	63
Gráfico 14: Erros percentuais absolutos para diferentes valores de p (canais pagos)	64
Gráfico 15: Regressão linear da receita total e investimento	66
Gráfico 16: Previsão da receita total com regressão linear da receita e investimento	66
Gráfico 17: Erro percentual absoluto no modelo com regressão linear aplicada na previsão da receita total	67
Gráfico 18: Erro percentual absoluto no modelo com regressão linear aplicada na previsão da receita total	67
Gráfico 19: Regressão linear da receita de canais grátis e investimento	68
Gráfico 20: Previsão da receita gratuita com regressão linear da receita de canais grátis e investimento	69
Gráfico 21: Erro percentual absoluto no modelo com regressão linear aplicada na previsão da receita de canais grátis	69
Gráfico 22: Erro percentual absoluto no modelo com regressão linear aplicada na previsão da receita de canais grátis	70
Gráfico 23: Regressão linear da receita de canais pagos e investimento	71

Gráfico 24: Previsão da receita paga com regressão linear da receita de canais pagos e investimento	71
Gráfico 25: Erro percentual absoluto no modelo com regressão linear aplicada na previsão da receita de canais pagos	72
Gráfico 26: Previsões da receita total com suavização exponencial simples	74
Gráfico 27: Erro percentual em valores suavização exponencial simples para receita total	74
Gráfico 28: Erro percentual absoluto em valores da suavização exponencial simples para receita total	75
Gráfico 29: Previsões da receita de origem gratuita com suavização exponencial simples	76
Gráfico 30: Erro percentual nos valores obtidos com a suavização exponencial simples para receita de origem gratuita	77
Gráfico 31: Previsões da receita de origem paga com suavização exponencial simples	78
Gráfico 32: Erro percentual em valores suavização exponencial simples para receita de origem paga	78
Gráfico 33: Previsão da receita total com suavização exponencial (Holt-Winters)	80
Gráfico 34: Erro percentual em valores obtidos para receita total com suavização exponencial de Holt-Winters	80
Gráfico 35: Erro percentual absoluto em valores obtidos para receita total com suavização exponencial de Holt-Winters	81
Gráfico 36: Previsões da receita gratuita com suavização exponencial de Holt-Winters	82
Gráfico 37: Erro percentual em valores obtidos para receita gratuita com suavização exponencial de Holt-Winters	83
Gráfico 38: Erro percentual absoluto em valores obtidos para receita gratuita com suavização exponencial de Holt-Winters	83
Gráfico 39: Previsões da receita paga com suavização exponencial de Holt-Winters	84
Gráfico 40: Erro percentual em valores obtidos para receita paga com suavização exponencial de Holt-Winters	85
Gráfico 41: Erro percentual absoluto em valores obtidos para receita paga com suavização exponencial de Holt-Winters	85

Gráfico 42: Previsões dos diferentes métodos, receita obtida e investimento realizado em marketing	86
Gráfico 43: Erros percentuais obtidos pelos diferentes métodos	87
Gráfico 44: Erros percentuais absolutos obtidos pelos diferentes métodos	87
Gráfico 45: Previsões dos diferentes métodos aplicados sobre a receita de origem gratuita, receita obtida e investimento realizado em marketing	90
Gráfico 46: Erros percentuais obtidos pelos diferentes métodos aplicados sobre a receita de origem gratuita	91
Gráfico 47: Previsões dos diferentes métodos aplicados sobre a receita de origem paga, receita obtida e investimento realizado em marketing	92
Gráfico 48: Erros percentuais obtidos pelos diferentes métodos aplicados sobre a receita de origem paga	93
Gráfico 49: Evolução dos erros percentuais obtidos mensalmente pela suavização exponencial simples	95
Gráfico 50: Evolução da variação de receita mês a mês	96
Gráfico 51: Evolução da variação de investimento em marketing mês a mês	96
Gráfico 52: Variação de receita, investimento e erros percentuais mês a mês	97
Gráfico 53: Variação de receita, previsões do modelo escolhido e previsões do modelo ajustado	99
Gráfico 54: Comparação de erros: modelo escolhido x modelo ajustado	100
Gráfico 55: Modelo final aplicado nos meses de Setembro e Outubro de 2022	101
Gráfico 56: Proporção de itens vendidos por sku desagregada	103



## SUMÁRIO

1. INTRODUÇÃO	27
1.1. A empresa	27
1.2. Descrição do problema	28
1.3. Justificativa e importância do problema	28
2. FUNDAMENTAÇÃO TEÓRICA: MODELOS DE PREVISÃO DE DEMANDA	29
2.1 Importância de Modelos de Previsão	29
2.2 Implementação de Modelos de Previsão de Demanda	30
2.3 Tipos de Modelos de Previsão de Demanda	31
2.4 Métodos de Previsão Quantitativos	31
2.4.1 Média Móvel	32
2.4.2 Regressão Linear	32
2.4.3 Suavização Exponencial Simples	33
2.4.4 Suavização Exponencial Dupla	34
2.4.5 Suavização Exponencial de Holt-Winters	35
2.5 Erro em Modelos Preditivos	37
2.5.1 Erro Médio (ME)	37
2.5.2 Erro Absoluto Médio (MAD)	38
2.5.3 Erro Percentual Médio	38
2.5.4 Erro Percentual Absoluto (MAPE)	39
2.5.5 Erro Quadrático Médio (MSD)	39
3. CONTEXTUALIZAÇÃO DO TRABALHO	41
3.1 Mercado de Cosméticos Brasileiro	41
3.2 Posicionamento da Empresa no mercado	41
3.3 Área de Ciência de Dados para o E-Commerce	42
4. PLANEJAMENTO	45
4.1 Objetivos do Modelo Preditivo	45
4.2 Padronização, Segurança e Tecnologias Necessárias	47

4.2.2 Extração e Coleta dos Dados	48
4.2.3 Ferramenta de Análise Escolhida	49
4.3 Análise Prévia	50
4.4 Conclusões Iniciais	53
4.5 Estrutura de Testes Para Comparação dos Modelos	54
5. APLICAÇÃO E AVALIAÇÃO DAS SOLUÇÕES	57
5.1 Média Móvel	57
5.1.1 Média Móvel Aplicada na Receita Total	57
5.1.2 Média Móvel Aplicada na Receita de Origem Gratuita	61
5.1.3 Média Móvel Aplicada na Receita de Origem Paga	63
5.2 Regressão Linear	65
5.2.1 Regressão Linear Aplicada na Receita Total	65
5.2.2 Regressão Linear Aplicada na Receita de Origem Gratuita	68
5.2.3 Regressão Linear Aplicada na Receita de Origem Paga	71
5.3 Suavização Exponencial Simples	73
5.3.1 Suavização Exponencial Simples Aplicada na Receita Total	73
5.3.2 Suavização Exponencial Simples Aplicada na Receita de Origem Gratuita	76
5.3.3 Suavização Exponencial Simples Aplicada na Receita de Origem Paga	77
5.4 Suavização Exponencial de Holt-Winters	79
5.4.1 Suavização de Holt-Winters Aplicada na Receita Total	79
5.4.2 Suavização de Holt-Winters Aplicada na Receita de Origem Gratuita	82
5.4.3 Suavização de Holt-Winters Aplicada na Receita de Origem Paga	84
5.5 Avaliação dos modelos para Receita Total	86
5.6 Avaliação dos modelos com Distinção por Origem de Receita	90
5.6.1 Avaliação dos modelos para Receita de Origem Gratuita	90
5.6.2 Avaliação dos modelos para Receita de Origem Paga	92
5.6.3 Viabilidade dos Modelos Distintos por Origem de Receita	94

6. ESCOLHA E IMPLEMENTAÇÃO DO MODELO	95
6.1 Modelo Escolhido	95
6.2 Ajuste no modelo	98
6.2.1 Incorporação da Regressão Linear	98
6.3 Implementação do Modelo na Empresa	101
6.4 Previsão de Vendas de Produtos Através da Previsão da Receita	102
7. CONCLUSÕES	105
7.1 Aprendizados e Objetivos Alcançados	105
7.2 Acompanhamentos e Possibilidades de Mudança	106
8. REFERÊNCIAS BIBLIOGRÁFICAS	109



## 1. INTRODUÇÃO

Este trabalho de formatura para a engenharia de produção tem como assunto o planejamento e controle de produção, com um foco na previsão de receita para estabelecimento de metas. O trabalho foi realizado em uma empresa do setor de cosméticos na qual o autor estagiou de Agosto de 2021 até Setembro de 2022. Com relação à estrutura do documento, ele será dividido em sete capítulos principais: introdução, fundamentação teórica, contextualização do trabalho, planejamento, aplicação e avaliação das soluções, escolha e implementação do modelo, conclusões e referências bibliográficas.

### 1.1. A empresa

A Dr Jones é uma empresa de cosméticos e produtos de higiene masculinos, fundada em 2013. Inicialmente o modelo de negócio da empresa era focado em canais de venda tradicionais como o varejo físico e marketplaces de *e-commerce* (amazon, submarino, etc). Em 2019 a empresa passou por uma grande transformação, focando seus esforços em construir um canal de vendas próprio por e-commerce, a fim de obter controle sobre a experiência do cliente com produtos da marca. Após a transição a empresa conseguiu levantar aportes que totalizaram R\$11,5 milhões em 2022, que foram investidos principalmente no desenvolvimento de novos produtos e comunicação para promover a expansão da marca.

A organização conta com 35 colaboradores distribuídos em 8 células organizacionais, sendo elas: Planejamento/Financeiro, People & Culture, Comunicação & Design, Growth & Digital Product, CRM & Performance, Customer Happiness, Logística & Operações e Product & Supply. Cada área possui um coordenador sendo supervisionados por 3 lideranças, os dois sócio-fundadores e o líder de Growth & Digital Product.

## 1.2. Descrição do problema

No início de 2022, a empresa passou por ajustes de metas, passando a priorizar a redução de custos e aumento de rentabilidade de suas operações. No passado a marca utilizava uma parcela considerável de recursos para promover sua expansão, principalmente através de gastos em canais de marketing digital.

Como forma de reduzir despesas e evitar uma perda no volume de vendas e aquisição de novos clientes, a empresa passou a ser mais cautelosa e mais certa nas despesas de marketing, e valorizar a área de CRM, que traz receitas com custos baixos. As áreas responsáveis por estabelecer e executar as estratégias de marketing são a de Performance e CRM, que baseiam suas iniciativas em previsões de vendas. Porém, as estimativas de vendas futuras são, na maior parte das vezes, obtidas sem um método estabelecido, ou através de pequenas análises pontuais. Porém, com a meta de reduzir gastos e realizar ações de marketing mais direcionadas, é preciso que as tomadas de decisão sejam embasadas em dados, a fim de otimizar a aquisição de clientes e o lucro.

## 1.3. Justificativa e importância do problema

O objetivo do trabalho é o de embasar as áreas de “Growth & Digital Product” e “Planejamento & Financeiro” com dados, para facilitar o processo de decisão. Com a implementação de modelos preditivos mais sofisticados, será possível definir metas mais precisas e um planejamento de marketing mais eficiente. Além disso, será necessário definir formas de mensurar e exibir a evolução dos indicadores chaves de desempenho, (*Key Performance Indicators* - KPI's) selecionadas para acompanhamento da empresa e projeto.

Todas essas melhorias são muito importantes para a empresa, especialmente por conta do foco atual de se aumentar o retorno sobre investimento (*Return Over Investment* - ROI) e reduzir os custos desnecessários da área. Com essas melhorias a área de “Produto&Performance” pode entregar mais por menos.

## 2. FUNDAMENTAÇÃO TEÓRICA: MODELOS DE PREVISÃO DE DEMANDA

### 2.1 Importância de Modelos de Previsão

As decisões gerenciais e de planejamento estratégico de uma empresa são, em grande parte, baseadas em expectativas de resultados. Um modelo de previsão bem construído pode auxiliar na tomada de decisões dentro da empresa, bem como nivelar as expectativas de todos os envolvidos de acordo com a realidade do mercado e fatores internos do negócio.

Os modelos preditivos têm um grande potencial para afetar diversas áreas da empresa. Com os departamentos cada vez mais conectados através da troca de informações facilitada, uma previsão de demanda bem ajustada pode influenciar de forma positiva o desempenho de diversas áreas da organização. Existem três principais categorias de atividades afetadas pela previsão de demanda: o planejamento operacional, a aquisição de recursos e a definição da necessidade de recursos futuros (Makridakis, 1998). Essas atividades afetam diretamente o desempenho de curto e longo prazo da empresa, e são portanto, essenciais para o bom funcionamento da mesma. Dessa forma, um modelo preditivo pouco calibrado ou inexistente, pode impactar de forma negativa diversas áreas de uma organização. Prejudicando a formulação de metas, estratégias e planejamentos da mesma.

O desenvolvimento de um modelo preditivo deve ser independente dos planos e metas da organização, que deverão ser alinhados a partir do cenário previsto. Um erro comum em organizações é o de constituir meios preditivos de vendas baseados em esperanças e pontos de vista, um modelo preditivo quantitativo pode ajudar a estabelecer valores imparciais através do comportamento de vendas nos períodos passados.(Armstrong, 2001). Arnold, Chapman e Clive (2011) levantaram quatro princípios que permeiam modelos preditivos de demanda, sendo eles:

- Toda previsão apresentará um erro, idealmente o modelo que apresentar o menor erro ou erro satisfatório deverá ser escolhido.
- Toda previsão deve conter a estimativa do erro calculado, importante para definir a precisão do modelo e verificar possíveis adaptações.

- Previsões de objetos individuais tendem a ser menos precisas, em geral o agrupamento de itens com características similares promove um modelo com menor erro.
- Quanto mais distante o período a ser previsto, maior será o erro esperado.

## 2.2 Implementação de Modelos de Previsão de Demanda

Segundo Pellegrini (2001), a implementação de um modelo de previsão de demanda pode ser dividida em seis etapas. Apesar de tratar especificamente de modelos para previsão de demanda, essas etapas podem ser generalizadas para outros modelos preditivos. Sendo elas:

1. Definição de um problema: Nesta fase, a análise de diversos fatores deverá acontecer, de forma a decidir onde o modelo será aplicado, o nível de precisão necessário para atender as necessidades do projeto e as limitações impostas por cada modelo.
2. Coleta de informações: Um modelo preditivo necessita de no mínimo dois tipos de dados: quantitativos e qualitativos. Os dados quantitativos serão utilizados para a modelagem matemática da previsão, já os dados subjetivos servem como forma de validação das previsões geradas.
3. Seleção do pacote computacional: Determinado o modelo a ser construído, é importante escolher qual ferramenta será utilizada para consolidar os dados e gerar as previsões. Isso deverá ser definido tendo em mente o volume de dados e habilidades necessárias para implementação.
4. Análise Preliminar: Essa etapa consiste na retirada e agrupamento dos dados históricos da empresa para identificar possíveis padrões e tendências presentes na série observada. Isso dará suporte para definição da modelagem matemática a ser realizada.
5. Escolha e validação dos modelos: Com as características da série temporal definidas anteriormente, deverá se escolher um modelo a ser testado. Essa escolha levará em consideração o problema a ser resolvido, o grau de agrupamento dos dados, os padrões identificados na série histórica, o intervalo de previsão e precisão desejada.

6. Verificação: Com o processo de escolha e modelagem inicial finalizados, é necessário acompanhar os resultados para possível manutenção. Além disso, é importante ir incorporando os novos dados na base a fim de atualizá-la e repetir o processo de validação com o passar do tempo.

### 2.3 Tipos de Modelos de Previsão de Demanda

Existem diversos métodos para se construir um modelo de previsão, que podem ser divididos em duas categorias: métodos qualitativos e métodos quantitativos (Armstrong 2001). Os métodos qualitativos são feitos através de pesquisas e entrevistas que levantarão experiências, opiniões e julgamentos de pessoas envolvidas ou de especialistas. Já os métodos quantitativos utilizam ferramentas estatísticas para processar dados numéricos históricos, de forma a projetar previsões com alguma margem de erro.

A quantidade de dados históricos disponíveis, bem como a adequação ao problema sendo trabalhado serão os principais fatores a serem considerados na hora de decidir se deverão ser empregados métodos qualitativos ou quantitativos. Para a escolha dos métodos quantitativos a serem testados, deverá ser considerada a disponibilidade de dados que irão alimentar o modelo.

### 2.4 Métodos de Previsão Quantitativos

A aplicação de um método de previsão quantitativo exige a existência de uma base histórica de dados que serão utilizados como parâmetros dentro do modelo. Uma das formas mais comuns de base histórica é a série temporal, onde os dados são dispostos ao longo de intervalos de tempo iguais. Para aplicação de séries temporais em métodos quantitativos de previsão de demanda, é preciso que ao menos três condições sejam atendidas (Makridakis, 1998):

- Disponibilidade de dados históricos
- Dados históricos dispostos como dados numéricos
- Assumir que comportamentos do passado devem se repetir no futuro

Ainda de acordo com Makridakis (1998), as técnicas de previsão quantitativas variam consideravelmente entre si. Desenvolvidas por disciplinas diversas para propósitos diferentes, cada uma apresenta suas propriedades internas, precisão e custos atrelados que devem ser considerados no processo de escolha.

#### 2.4.1 Média Móvel

Uma das formas mais simples para construir uma previsão. Consiste na retirada de um valor médio de um intervalo de períodos a ser definido, a partir desse valor médio é possível estimar o valor esperado para os próximos períodos.

A definição da quantidade de períodos a serem analisados é decisiva para o modelo. Períodos menores irão priorizar informações recentes e que refletem melhor a situação atual, porém períodos maiores proporcionam menores variações e flutuações pontuais. Ambas podem gerar previsões imprecisas, por isso é importante entender qual caso melhor se aplica ao problema.

Segundo Makridakis (1998), comparando com o uso da média simples, a média móvel apresenta as seguintes características: o método lida apenas com os últimos  $k$  períodos, número fixos de pontos ao longo do tempo, demanda um armazenamento de dados maior (todos os dados das últimas  $k$  observações) e não performa bem com dados que apresentem tendências ou sazonalidades.

#### 2.4.2 Regressão Linear

A regressão linear é um método estatístico para investigar e modelar a relação entre variáveis que pode ser utilizada para prever o valor de uma variável com base no valor de outra (Montgomery, 2021). Essa regressão se ajusta a uma linha reta que minimiza as discrepâncias entre os valores de entrada e saída. É uma ferramenta muito documentada desenvolvida no final do século XIX, além de ser um recurso relativamente simples com uma fórmula matemática de fácil compreensão.

A regressão linear tem o seguinte formato:

$$f = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \dots + \beta_n * X_n$$

Onde  $X_n$  representa a  $n$ ésima variável independente e  $b_n$  representa o coeficiente calculado através do método de mínimos quadrados referente à  $n$ ésima variável independente.

Seguem alguns requisitos para utilização correta da regressão:

- Variáveis com medições a nível contínuo
- Observações independentes umas das outras
- Dados sem valores discrepantes significativos
- Homocedasticidade da amostra
- Erros da linha de regressão de melhor ajuste seguem uma distribuição normal

#### 2.4.3 Suavização Exponencial Simples

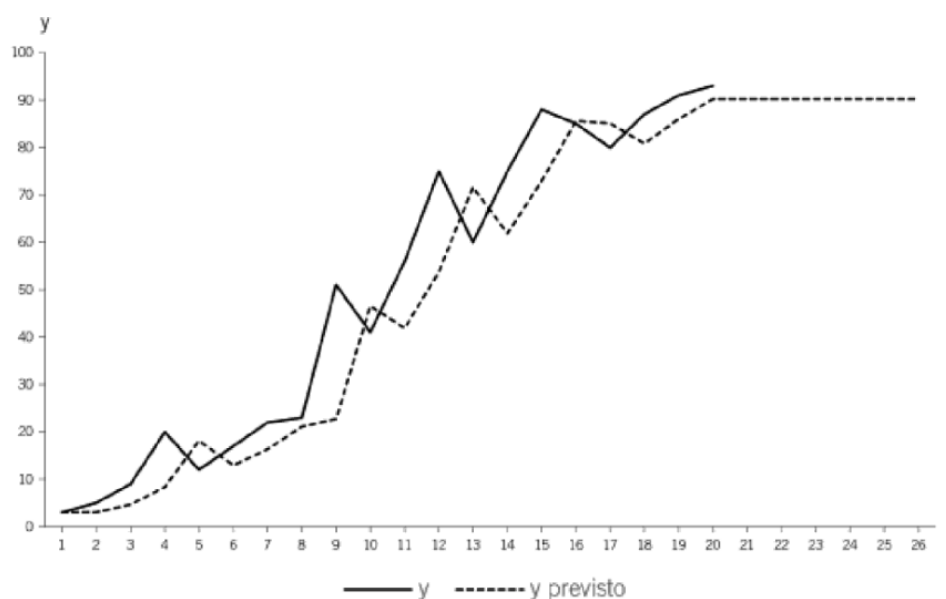
Técnicas de suavização exponencial são as mais utilizadas quando comparadas a outras formas, por sua simplicidade, acurácia e facilidade de ajuste. O uso da suavização exponencial simples é indicado quando os dados não apresentam tendência ou sazonalidade. (Pellegrini, 2001)

Esse método pondera todos os dados históricos com pesos exponencialmente decrescentes em direção ao passado, dessa forma, geralmente os dados mais recentes apresentam maior peso. A fórmula expandida pode ser escrita na forma:

$$P = \alpha Y(t - 1) + \alpha(1 - \alpha)Y(t - 2) + \alpha(1 - \alpha)^2 Y(t - 3) + \dots$$

A constante de suavização  $\alpha$  assume valores de 0 a 1, períodos mais distantes apresentam uma constante menor por possuir menor relevância. Como os valores dessas constantes irão determinar o peso dos períodos sendo analisados, é importante testar algumas variações e ir observando os valores gerados até que se encontre um alfa que minimize o erro final da previsão.

Figura 1: Exemplo suavização exponencial simples



Fonte: ARONNE et al. (2021)

#### 2.4.4 Suavização Exponencial Dupla

A suavização exponencial dupla utiliza uma nova constante  $\gamma$  (gama), que permite a consideração de possíveis tendências sazonais da série, adicionando essas variações ao construir uma previsão, consiste em suavizar a série temporal duas vezes a partir das seguintes construções.

$$P_t = \alpha Y_t + (1 - \alpha)(P_{t-1} + b_{t-1})$$

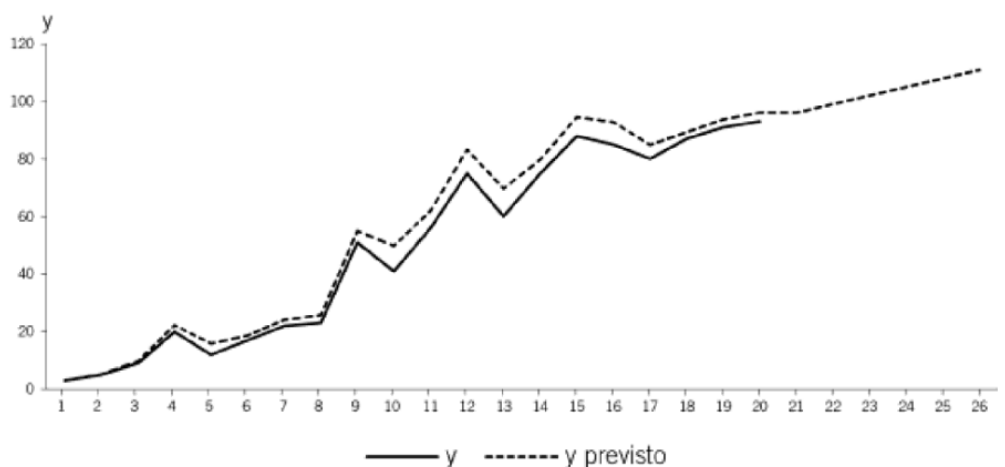
$$b_t = \gamma(P_t - P_{t-1}) + (1 - \gamma)b_{t-1}$$

A série  $P_t$  será suavizada de acordo com as tendências observadas no período anterior,  $b_t$  possui uma estrutura muito similar com a suavização simples, atualizando a tendência com base nos últimos valores obtidos. Para inicialização,  $b_1$  pode ser definido como o valor médio da variação dos períodos anteriores:

$$b_1 = (Y_n - Y_{n-1}) / (n - 1)$$

Técnicas de otimização não lineares podem ser utilizadas para encontrar o valor ótimo para as constantes, minimizando o erro.

Figura 2: Exemplo suavização exponencial dupla



Fonte: ARONNE et al. (2021)

#### 2.4.5 Suavização Exponencial de Holt-Winters

Esse método é uma extensão da suavização exponencial de Holt, apropriado para séries com tendência linear e um componente sazonal. O método produz valores exponencialmente suavizados para o nível da previsão, a tendência e o ajuste sazonal. Apresenta melhor desempenho para dados que apresentem tendências e sazonalidades que não aumentem com o tempo. O método de Holt-Winters exige um mínimo de dados de períodos completos para que seja possível determinar a estimativa inicial dos índices de sazonalidade.

A equação e os coeficientes podem ser determinados pelas seguintes construções:

$$P_t = \left( \frac{Y_t}{I_{t-L}} \right) + (1 - \alpha)(P_{t-1} + b_{t-1}) \quad (A)$$

$$b_t = \gamma(P_t - P_{t-1}) + (1 - \gamma)b_{t-1} \quad (B)$$

$$I_t = \beta \left( \frac{Y_t}{S_t} \right) + (1 - \beta)I_{t-L} \quad (C)$$

$$F_{t+m} = (S_t + mb_t) I_{t-L+m} \quad (D)$$

Onde (A) é a equação de suavização geral, (B) é a equação de suavização de tendência, (C) é a equação de suavização de sazonalidade e (D) é a equação de previsão. As constantes presentes no método devem ser determinadas de forma a reduzir os erros e softwares estatísticos podem ser utilizados para isso.

O fator de tendência pode ser definido através da fórmula:

$$b_1 = 1/L \left( \frac{Y_{L+1} - Y_1}{L} + \frac{Y_{L+2} - Y_2}{L} + \dots + \frac{Y_{L+L} - Y_L}{L} \right)$$

Para iniciar o índice de sazonalidade é necessário calcular o valor médio para cada período e depois dividir cada observação pela média do período em questão.

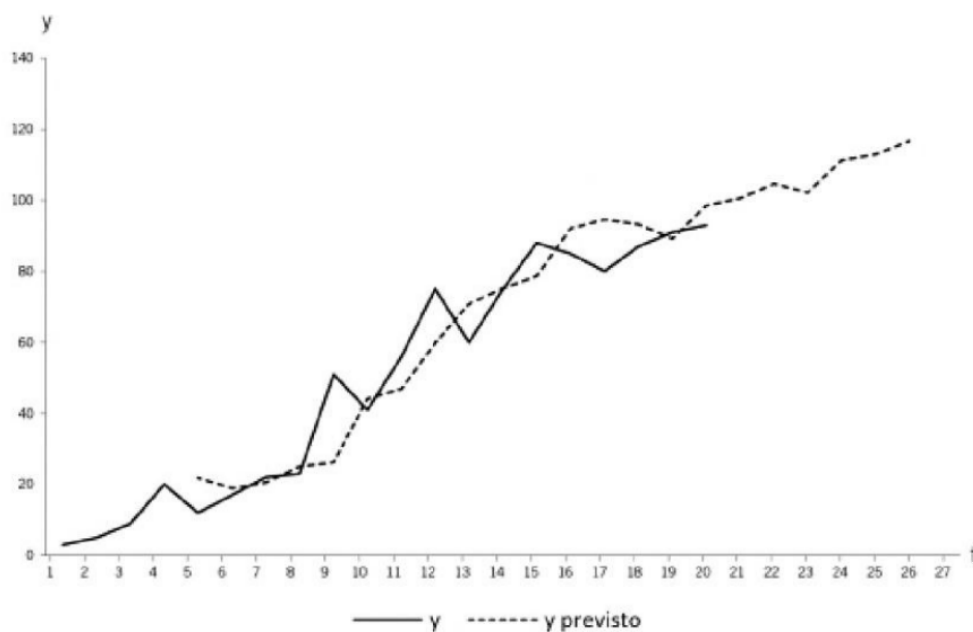
$$Ap = \sum_i^p \frac{Y_i}{p}$$

Ao fim, os índices sazonais serão definidos como a média de cada linha, que representam as mesmas estações presentes no período analisado, na forma:

$$I_1 = \left( \frac{Y_1}{A_1} + \frac{Y_5}{A_2} + \frac{Y_9}{A_3} + \dots \right)$$

$$I_x = \left( \frac{Y_x}{A_1} + \frac{Y_{x+4}}{A_2} + \frac{Y_{x+8}}{A_3} + \dots \right)$$

Figura 3: Exemplo suavização exponencial de Holt-Winters



Fonte: ARONNE et al. (2021)

## 2.5 Erro em Modelos Preditivos

De acordo com Makridakis e Carbone (1985) os métodos preditivos assumem que relações e padrões estabelecidos não irão se alterar após o estabelecimento do modelo, e afirmam como isso não é uma suposição realista para séries econômicas de negócios. Portanto, nem o modelo mais preciso estará imune a erros. Todo modelo apresenta um erro de previsão, por isso é importante encontrar a melhor opção e calibrá-la de forma a reduzir o erro (Arnold, Chapman e Clive, 2011). Por definição, o erro pode ser definido como a demanda real subtraída da demanda prevista. Seguem algumas formas de se medir o desvio de precisão (Daniel Mota, 2021):

### 2.5.1 Erro Médio (ME)

O erro médio cumpre o papel de oferecer uma visão global da precisão do método utilizado e pode ser calculado através da média dos erros observados em uma amostra da previsão. Com a seguinte formulação:

$$ME = \sum_{t=1}^n \frac{(Dt - Ft)}{n}$$

$Dt$  = Demanda no período  $t$

$Ft$  = Previsão para o período  $t$

$n$  = Número de períodos sendo analisado

É importante notar que essa forma considera a diferença de valores negativos e positivos, que podem acabar se anulando. Fator que pode comprometer a avaliação da precisão dos resultados. Por outro lado, um valor médio sinalizado pode indicar uma previsão subestimada ou superestimada, mostrando o viés do modelo.

### 2.5.2 Erro Absoluto Médio (MAD)

O erro absoluto tem uma construção similar ao erro médio, porém elimina a possibilidade de valores se anularem com adição do módulo do erro na somatória, por isso fornece uma melhor visualização da acurácia real do método. Segue a formulação:

$$MAD = \sum_{t=1}^n \frac{|Dt - Ft|}{n}$$

$Dt$  = Demanda no período  $t$

$Ft$  = Previsão para o período  $t$

$n$  = Número de períodos sendo analisado

### 2.5.3 Erro Percentual Médio

O erro percentual médio representa a diferença média dos erros observados em uma determinada amostra com representação percentual, retirando o fator de escala do erro e considerando o sinal dos erros calculados. Esse formato permite uma comparação mais justa entre períodos diferentes, uma vez que valores absolutos podem variar muito com a passagem do tempo. O erro é calculado pela seguinte fórmula:

$$MPE = \sum_{t=1}^n \frac{\left(\frac{Dt - Ft}{Dt}\right)}{n}$$

$Dt$  = Demanda no período  $t$

$Ft$  = Previsão para o período  $t$

$n$  = Número de períodos sendo analisado

#### 2.5.4 Erro Percentual Absoluto (MAPE)

O erro percentual absoluto é uma medida relativa que dimensiona o erro absoluto médio em unidade percentual. Assim como no MAD, esse método impede que valores positivos e negativos sejam cancelados e permite a comparação da precisão da previsão entre métodos de séries de tempo. O erro é calculado pela seguinte fórmula:

$$MAPE = \sum_{t=1}^n \frac{\left| \frac{Dt - Ft}{Dt} \right|}{n}$$

$Dt$  = Demanda no período  $t$

$Ft$  = Previsão para o período  $t$

$n$  = Número de períodos sendo analisado

#### 2.5.5 Erro Quadrático Médio (MSD)

O erro quadrático médio fornece a média de diferença quadrática entre a previsão do modelo e o valor de destino, e pode ser utilizada como medida da qualidade de um estimador. A tendência do erro pode revelar uma necessidade de ajuste ou manutenção do modelo: uma crescente indica que a métrica está se deteriorando, uma tendência decrescente indica que a métrica está melhorando e uma variação irregular pode indicar uma necessidade de aumento da amostra (IBM, 2022). Segue a formulação:

$$MSD = \sum_{t=1}^n \frac{(Dt - Ft)^2}{n}$$

$Dt$  = Demanda no período  $t$

$Ft$  = Previsão para o período  $t$

$n$  = Número de períodos sendo analisado



### 3. CONTEXTUALIZAÇÃO DO TRABALHO

#### 3.1 Mercado de Cosméticos Brasileiro

De acordo com o relatório da consultoria Research & Markets o mercado global de beleza masculina atingiu o valor de \$69 bilhões em 2020 e tem a previsão de atingir \$79 bi em 2024. É um valor pequeno se comparado ao mercado de cosméticos feminino, que gira em torno de \$630 bilhões segundo o mesmo relatório, diferença que pode indicar um potencial de crescimento do mercado.

No Brasil o mercado de cosméticos masculino se mostra promissor, de acordo com a Associação Brasileira da Indústria de Higiene Pessoal, Perfumaria e Cosméticos (ABIHPEC) o mercado nacional cresceu 527% entre 2014 e 2019. Além disso, segundo relatório da Euromonitor, o Brasil representa cerca de 13% do mercado global de beleza masculina.

O crescimento recente do setor foi alavancado por uma expansão de estabelecimentos de beleza destinados ao público masculino, além do surgimento de comércios digitais voltados ao segmento. A expansão pode ser explicada por uma mudança de comportamento do público masculino que, segundo pesquisas realizadas pela ABIHPEC, se mostra cada vez mais preocupado com a aparência e mais disposto a investir tempo e dinheiro no setor de beleza e cosméticos.

Atualmente a maior parte do *Market Share* pertence a fábricas grandes com marcas estabelecidas como a Natura, Unilever, Boticário, Procter & Gamble, etc. Empresas menores e startups apostam em estratégias de marketing inovadoras, atendimento personalizado, marketing provocativo, proximidade com o consumidor e diferenciação de produtos para conquistar espaço no mercado.

#### 3.2 Posicionamento da Empresa no mercado

As novas tendências de mercado dentre o público masculino permitiram a expansão da Dr Jones, que pivotou sua operação para um modelo Direct to Consumer (D2C) em 2019. A empresa iniciou suas atividades no e commerce apostando no marketing provocativo, cutucando concorrentes grandes, e através de

uma atenção especial no atendimento e pós compra dos consumidores. Além de tratar de diversos tópicos de beleza e saúde masculina em seu blog e redes sociais.

A empresa teve seu crescimento impulsionado através de diversos aportes que recebeu de fundos como: Astella, Igah Ventures e Ignite XL. Essas rodadas de investimento bem sucedidas possibilitaram maiores investimentos em expansão de portfólio, estrutura e investimentos em marketing.

Os produtos disponíveis em seu catálogo, em sua grande maioria, são de formulações próprias da empresa. Muitos deles foram desenvolvidos durante o período B2B. A empresa ainda não possui um amplo catálogo de produtos e realiza lançamentos bem espaçados.

O marketing de performance foi essencial para o crescimento da DRJ, que alinhado a uma comunicação despojada e direta, proporcionou uma maior visibilidade aos produtos e serviços de assinatura da empresa. Atualmente a DRJ conta com um público muito fiel e que divulga a marca de forma orgânica, indicador disso é o elevado NPS (Net Promoter Score) da empresa.

A DRJ pretende seguir com o crescimento observado nos anos anteriores, utilizando novos métodos de aquisição de clientes, expansão do catálogo de produtos e parcerias com outras marcas, além de buscar novas fontes para captação de recursos a fim de possibilitar maior agressividade em investimentos.

### 3.3 Área de Ciência de Dados para o E-Commerce

O comércio eletrônico brasileiro cresceu muito em volume na última década, com diversas marcas consolidadas expandindo seus negócios para o ambiente virtual e o surgimento de inúmeras empresas que operam apenas pela internet. A tendência foi impulsionada pela pandemia de Covid-19, com o isolamento social consumidores passaram a comprar mais, além de expandirem as categorias de produtos comprados pela internet. Diversos públicos, com destaque às faixas etárias superiores, foram incentivados a experimentar a compra online pela primeira vez. (Lara et al., 2020).

Uma área de dados estruturada será capaz de construir modelos e análises baseadas em dados, para consolidar um ambiente de inteligência de negócios e ciência de dados é necessário verificar as necessidades a serem atendidas e a

disponibilidade de dados que podem ser manipulados. Alguns benefícios da implementação da área de dados no comércio eletrônico são: previsão de demanda, mensuração de indicadores chave de desempenho (KPI's) do marketing, agrupamento de clientes, sistemas de recomendação e análises de comportamento do consumidor. (Fedirko Et al., 2021)

A diversidade de plataformas digitais e ferramentas integrativas de dados disponíveis no mercado hoje em dia permitem com que empresas consigam coletar grandes quantidades de dados relativos à operação. Apesar disso, muitas organizações possuem inúmeros dados inutilizados. O armazenamento de dados se torna inútil até que eles sejam utilizados para analisar tendências, padrões e relações que forneçam insights para a definição de estratégias da empresa.(Albright Et al, 2011)



## 4. PLANEJAMENTO

### 4.1 Objetivos do Modelo Preditivo

Antes de iniciar a construção e teste dos modelos é primordial a definição do problema para que os objetivos do trabalho sejam elencados (Pellegrini 2001), possibilitando o manejo dos recursos necessários, levantar as especificidades do negócio e determinar as variáveis que devem ser levadas em consideração para melhor adequação ao projeto.

A Dr Jones realiza o planejamento de receita através de metas mensais, que são estabelecidas pelos sócios com a ajuda da área de Growth. Os métodos utilizados para determinação não são padronizados e acabam sendo parcialmente embasados em achismos e expectativas das lideranças da empresa. Além disso, como as metas de meses seguidos são definidas de uma vez, é comum que a distância entre a meta e a receita real vá aumentando no acumulado mensal, gerando frustração dos times com relação à performance esperada e obtida.

O modelo preditivo será aplicado na receita da empresa com o intuito de estimar valores futuros que servirão como base para o estabelecimento de metas para os próximos períodos. Espera-se que o modelo permita uma estimativa precisa das metas, levando em conta possíveis flutuações sazonais e outras variáveis que geram impacto na receita, como o budget disponível para investimento em marketing. O trabalho será bem sucedido se for capaz de gerar valores de metas com erro menor que os observado no modelo atual, para isso será necessário testar valores na base histórica e comparar os resultados com os valores reais observados na empresa.

Por se tratar de uma empresa com um catálogo enxuto, com poucos lançamentos de produtos ao longo do tempo e com estabilidade na distribuição de origem da receita entre os skus, ao se estimar a receita em períodos futuros talvez seja possível extrapolar os valores para gerar uma previsão de demanda dos produtos. Isso auxiliaria o gerenciamento de estoques, possibilitando maior previsibilidade e evitando esgotamentos indesejados. Para implementação, essa hipótese precisará ser testada a fim de verificar a margem de erro da previsão de demanda dos produtos.

Gráfico 1: Diferença percentual entre previsão e receita obtida



Fonte: elaborado pelo autor.

No gráfico 1 é possível visualizar de forma mensal a diferença percentual entre a previsão estabelecida e a receita obtida, nele fica evidenciado que a meta costuma ser superestimada e que houve uma melhora da precisão dos valores a partir do ano de 2022. Isso aconteceu devido a uma estabilização no crescimento mensal da empresa, aumentando a previsibilidade de demanda, e também porque as metas passaram a ser definidas mais vezes ao longo do ano, com maior enfoque no curto prazo. Dessa forma as previsões têm maior êxito ao levar em conta o momento atual da empresa e do mercado, além da possibilidade de se levantar valores mais realistas de investimentos a serem realizados em marketing. Esses fatores favoreceram o melhor desempenho na definição de metas observado em 2022.

Em 2021 o erro percentual médio observado entre meta e receita foi cerca de 22%, em 2022 esse valor caiu para cerca de 5%. Espera-se que o modelo preditivo seja capaz de determinar metas com um erro médio menor do que o observado anteriormente, além disso, é importante que o modelo identifique e corrija os valores que podem sofrer distorção devido à sazonalidades como o Dia dos Pais e Black Friday.

Como o modelo preditivo irá embasar decisões estratégicas da empresa, é primordial que uma rotina padronizada seja estabelecida. A rotina será composta por

quatro etapas e terá o intuito de avaliar os resultados progressivamente e permitir a continuidade do modelo proposto. As fases estão detalhadas na figura abaixo.

Figura 4: Ciclo de rotinas para manutenção do modelo preditivo

#### 1. Previsão do próximo período

Modelo preditivo realiza o cálculo de receita esperada para os períodos futuros. Levando em conta a base histórica, o investimento em marketing esperado e a sazonalidade do período.

#### 2. Aprovação da meta

Após o estabelecimento da meta, ela deverá ser aprovada pelos sócios e time de Growth. Caso necessário, pequenos ajustes manuais podem ser feitos nessa etapa.

#### 3. Avaliação dos resultados

Avaliação dos resultados no fim do mês, a fim de verificar a conformidade do modelo e levantar o erro obtido entre receita esperada e obtida.

#### 4. Atualização da base

Adicionar novos valores na base de cálculo do modelo. No caso, as entradas são: a receita obtida no período e o investimento esperado para o próximo mês.



Fonte: elaborado pelo autor.

## 4.2 Padronização, Segurança e Tecnologias Necessárias

Para que a implementação do modelo possa ser incorporada à empresa é necessário estabelecer uma padronização em todos os acessos e plataformas que serão utilizadas no projeto. Isso é importante pois dados sensíveis irão alimentar a base do modelo, justificando a cautela com relação aos acessos, vulnerabilidades e medidas de segurança. Além disso, é crucial que o projeto não seja corrompido por má gestão dos dados e plataformas envolvidas. Caso o modelo ou a base de dados precise ser verificada, caberá ao time de Data-Science executar a tarefa.

O acesso aos resultados do modelo preditivo será restrito às lideranças da empresa, time de Growth e time de Data-Science. O projeto tem o intuito de auxiliar a definição de metas e estimativa de vendas para controle de estoque, porém a meta a ser compartilhada com os colaboradores da organização será definida pela liderança. Esse ajuste manual é válido uma vez que o modelo não leva em conta

fatores externos às bases de dados, como flutuações de mercado, novas estratégias da empresa, lançamento de produtos, etc.

#### 4.2.1 Disponibilidade dos Dados

Os dados históricos da empresa são armazenados em um repositório central de informações (Data Warehouse) na plataforma BigQuery do Google. Os dados de receita e investimento se encontram em uma base com variáveis já padronizadas e formatadas. O acesso ao Banco de Dados da empresa é limitado à área de Data-Science & BI, isso porque a plataforma contém dados de desempenho de vendas do e-commerce, bem como os dados pessoais dos clientes. O acesso não é liberado dentro da organização por conta da facilidade de extração e manipulação de dados que a plataforma proporciona a seus usuários. A fim de preservar dados sensíveis da empresa, os valores absolutos exibidos em gráficos e tabelas serão alterados através da multiplicação de um fator que não será divulgado no relatório.

#### 4.2.2 Extração e Coleta dos Dados

A plataforma BigQuery permite que os usuários realizem consultas através de instruções de consulta através de uma Linguagem de Consulta Estruturada (SQL). Os dados que irão alimentar o modelo serão extraídos no formato diário. Essa forma permite uma maior maleabilidade para testar métodos preditivos, pois os dados podem ser aglomerados de forma semanal e mensal.

O código SQL será utilizado para criar uma tabela virtual (visualização) dentro do repositório. A visualização desempenha o papel de uma tabela comum, mas pode ser configurada como fonte para alimentar bases de diversas ferramentas de análise, o que permite a automatização na extração e contribui para uma maior integridade do modelo.

Para testar os modelos preditivos, serão construídas duas bases. Uma base simplificada com valores de receita total e investimento em marketing por dia. A outra base trará os valores de receita de acordo com a origem, que pode ser gratuita ou paga, além dos valores de investimento em marketing. A separação de acordo com a origem da receita pode ser interessante pois o investimento em marketing de

forma geral atrai receita de origem paga. O modelo mais adequado pode ser uma mescla de modelos diferentes, que calcule a previsão de receita de origem gratuita e paga de forma independente.

Ao fim da extração teremos duas bases com os seguintes formatos:

Tabela 1: Base de dados simplificada

dia	receita	investimento	semana	mês
2021-01-03	x	a	2021-01-03	2021-01
2021-01-04	y	b	2021-01-03	2021-01
2021-01-05	z	c	2021-01-03	2021-01
...	...	...	...	...

Fonte: elaborado pelo autor.

Tabela 2: Base de dados com distinção de origem de receita

dia	receita gratuita	receita paga	investimento	semana	mês
2021-01-03	x	m	a	2021-01-03	2021-01
2021-01-04	y	n	b	2021-01-03	2021-01
2021-01-05	z	o	c	2021-01-03	2021-01
...	...	...	...	...	...

Fonte: elaborado pelo autor.

#### 4.2.3 Ferramenta de Análise Escolhida

O programa de análise que será utilizado para realizar os testes dos modelos preditivos será a plataforma gratuita Google Sheets. O editor já faz parte do conjunto de programas que a empresa utiliza em suas rotinas e por isso o modelo preditivo final também será construído na plataforma. Todas as previsões, gráficos e tabelas de erros que serão utilizadas para comparação entre modelos serão calculados na ferramenta. Outra conveniência da plataforma é a possibilidade de conexão direta com visualizações do BigQuery, de forma gratuita, permitindo a automatização das entradas de dados na base do modelo.

Com relação às permissões de acesso do arquivo final do projeto, apenas o time de Data-Science terá a permissão de edição em todas as abas. As lideranças e o time de Growth terão apenas a permissão para visualizar a aba de previsão de receita, a fim de evitar confusão e edições não intencionais.

#### 4.3 Análise Prévia

Dando sequência às etapas de implementação de um modelo preditivo proposto por Pellegrini (2001). Após a definição do processo de coleta de dados é necessário que ocorra uma análise preliminar do problema, para que seja possível visualizar a evolução das variáveis que serão utilizadas nos modelos a fim de identificar padrões que podem auxiliar na testagem e construção da proposta final. Os valores absolutos que serão expostos no segmento a seguir foram alterados para preservar dados sensíveis.

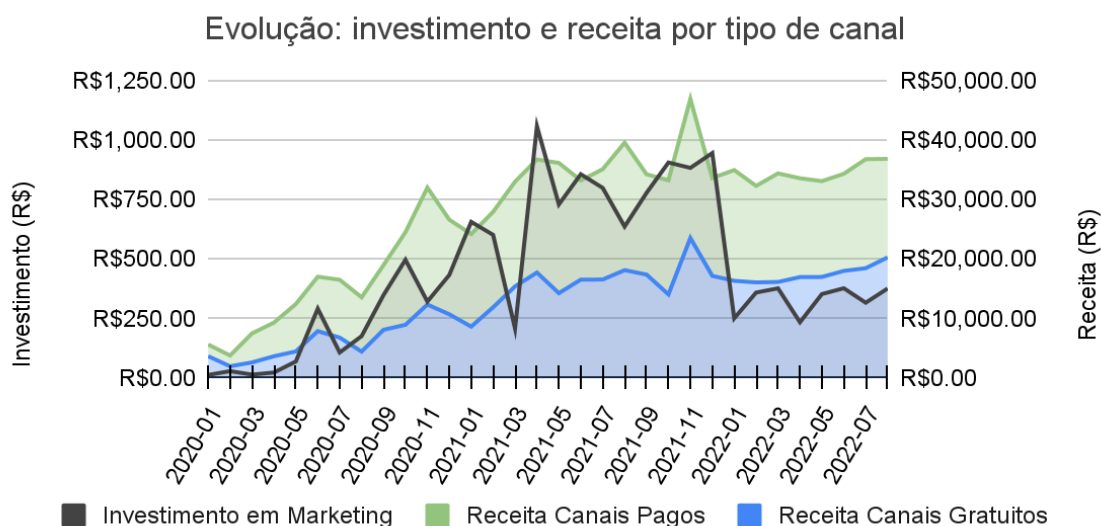
Para essa análise inicial serão considerados dados a partir de 2020 até o mês de Agosto de 2022. No período anterior a receita da empresa era diminuta e não possuía valores de investimento em marketing relevantes. A partir do ano de 2020, viabilizada por rodadas de investimentos bem sucedidos, a empresa passou a investir na expansão da marca de forma agressiva através do marketing de performance, se aproximando do volume de vendas atual.

Ao analisar o histórico de receitas e investimentos em marketing da empresa, verifica-se que o negócio passou por três fases distintas ao longo do tempo. Em 2020 a empresa foi aumentando os investimentos de forma progressiva e observou um crescimento de receita ao longo do período, em 2021 a receita se manteve em curva crescente com investimentos mais agressivos em canais pagos, em 2022 a empresa reduziu drasticamente os investimentos e observou uma redução na velocidade de crescimento. Além disso, percebe-se uma semelhança entre as variações de receita de origem paga e gratuita, com subidas e quedas síncronas entre os canais.

Na organização os canais pagos sempre foram muito importantes para a aquisição de novos clientes. Isso se deve ao foco dado nas plataformas Google Ads e Facebook Business, com campanhas agressivas de aquisição através de promoções para novos clientes, direcionada a novas audiências. A redução nos

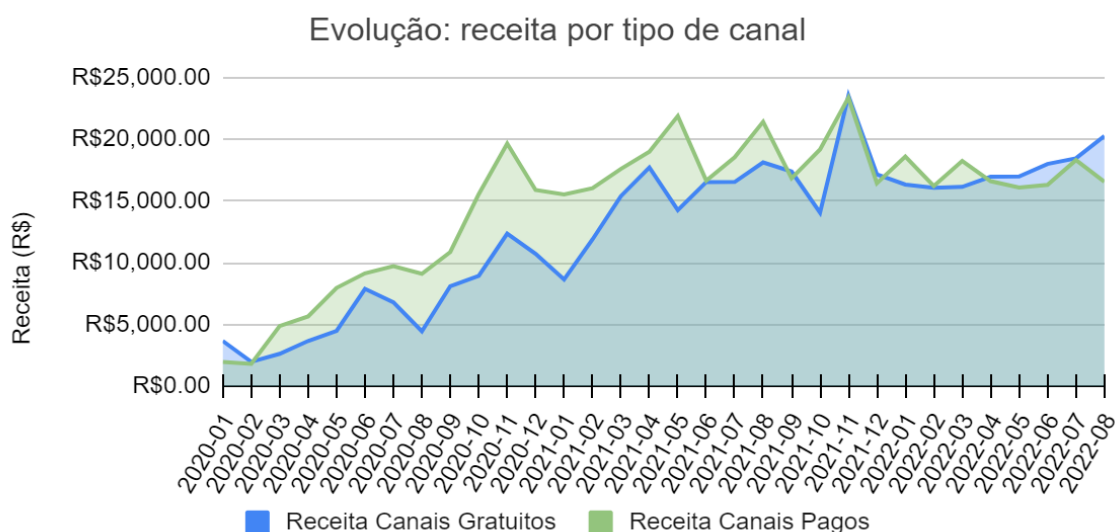
valores de investimento em marketing de performance em 2022, somada a uma base de clientes cada vez maior, fez com que a participação de clientes já adquiridos na receita mensal aumentasse, estabilizando em torno de 60%.

Gráfico 2: Receita por tipo de canal e investimento em marketing



Fonte: elaborado pelo autor.

Gráfico 3: Comparação entre receitas por tipo de canal



Fonte: elaborado pelo autor.

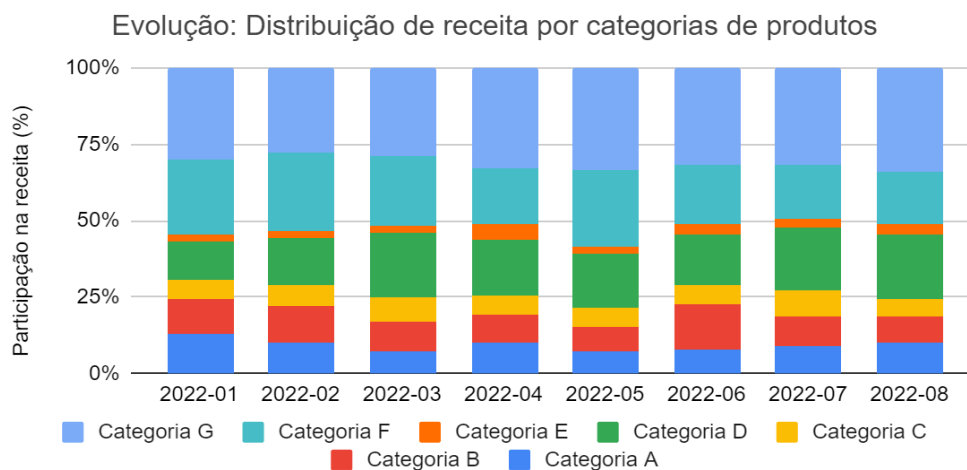
Gráfico 4: Receita por tipo de canal e investimento em marketing



Fonte: elaborado pelo autor.

Na definição dos objetivos do modelo preditivo foi elencada a possibilidade de utilizar as previsões de receita para estimar as vendas dos produtos, fornecendo maior previsibilidade para o time responsável pelo controle de estoques. Ao verificar a distribuição de receita de acordo com as categorias dos produtos nota-se uma estabilidade, especialmente a partir de 2022. A precisão e análise de viabilidade dessa extensão do projeto será testada após a escolha do modelo preditivo, afinal o objetivo primário do modelo é a estimativa dos valores de receita e esse será o critério levado em conta para determinação do melhor método.

Gráfico 5: Distribuição de receita por categoria de produto



Fonte: elaborado pelo autor.

Com relação ao erro entre o método de previsão de demanda atual e a receita, ele é a variável mais importante para determinar a implementação do modelo na organização, uma vez que o modelo preditivo só será incorporado se fornecer valores mais precisos do que os métodos manuais utilizados atualmente. No Gráfico 1 é possível visualizar a evolução do erro ao longo dos meses. O erro absoluto não foi utilizado nesse caso pois as diferenças ocorreram devido a valores superestimados das previsões, com apenas uma exceção de valor desprezível (-0,10%).

Tabela 3: Erros entre previsões e receita faturada

Ano	Erro Percentual Máximo (mensal)	Erro Percentual Médio (mensal)	Erro Percentual Acumulado no Ano
2021	55,9%	22,2%	28,0%
2022	11,9%	4,9%	5,1%

Fonte: elaborado pelo autor.

#### 4.4 Conclusões Iniciais

Com a análise preliminar foi possível constatar o impacto que o investimento em marketing produz sobre o crescimento da empresa e a receita faturada, períodos com menor investimento apresentam maior estabilidade. Essa característica pode impactar a escolha do modelo, pois alguns métodos serão melhores para cálculo de valores em receita crescente e outros serão mais adequados para períodos estáveis. O propósito do projeto é ser capaz de estimar valores futuros de acordo com o desempenho passado e investimento previsto, independente da fase que a empresa se encontra, por isso o modelo que apresentar melhor acurácia em ambas as situações, de receita crescente e estável, será priorizado.

Um comportamento identificado no Gráfico 2 é a semelhança entre as flutuações de receita por tipo canal, que não acompanham a variação de investimentos necessariamente. Com relação a sazonalidade, os dois principais picos de vendas e promoções da empresa são impulsionados pelo dia dos pais e black friday, o primeiro impactando as vendas de Julho e Agosto e o segundo de Novembro. Por conta das sazonalidades, talvez seja necessário realizar ajustes no

modelo a fim de evitar um aumento do erro ao subestimar valores para meses com maior volume de vendas e, também, para que os próximos períodos não sejam superestimados em razão desse aumento sazonal.

A análise dos erros foi importante para determinar valores base de desempenho esperado dos métodos a serem testados. Estabelece-se assim um requisito mínimo para a escolha de um modelo, o de apresentar um erro percentual médio inferior a 4,9% ao mês. Se atentando para que o erro máximo não ultrapasse 11,9% no mês. Dessa forma a empresa conseguirá alinhar suas expectativas com relação às vendas futuras e se planejar com maior embasamento, evitando potenciais frustrações.

#### 4.5 Estrutura de Testes Para Comparação dos Modelos

Com as características observadas na análise preliminar podemos determinar um planejamento para os testes a serem realizados. Optou-se por utilizar métodos quantitativos para realizar a previsão, portanto só serão realizados testes de métodos que consigam estimar valores a partir da base de dados disponível (Armstrong, 2001). A implementação de modelos de inteligência artificial e *machine learning* é inviável para a base atual, que é pequena demais para o propósito.

Não podemos aferir uma relação direta entre a receita e o investimento a partir dos dados disponíveis, mas espera-se de forma intuitiva que um maior investimento em marketing de performance irá trazer mais receita para a empresa, podendo haver deterioração do indicador chave de performance que representa a receita obtida sobre o investimento ou ROI. Considerando a possível relação entre os dados, métodos que consideram a correlação entre variáveis também serão testados.

Para facilitar o controle de sazonalidades da amostra, os valores serão agrupados em uma série temporal com valores mensais. Valores diários serão utilizados para identificar sazonalidades proporcionadas pelo dia da semana e período do mês, que poderão ser usadas para calcular uma diarização da previsão de receita.

Considerando os padrões observados no histórico da empresa, pontos levantando na análise prévia, disponibilidade de dados e a existência de períodos de receita crescente e estável, serão testados os seguintes métodos:

- Média Móvel
- Regressão Linear
- Suavização Exponencial Simples
- Suavização Exponencial de Holt-Winters

Como já mencionado anteriormente, a manipulação dos dados será realizada na plataforma de análise Google Sheets. Alguns modelos podem ser estruturados através da aplicação de fórmulas já existentes na ferramenta, outros serão levantados através da extensão gratuita para o Sheets chamada de “XLMiner Analysys ToolPak”, capaz de calcular a regressão linear múltipla e suavização exponencial simples por exemplo. No próximo capítulo cada aplicação de método será explicado individualmente, destacando suas particularidades.

Para avaliar a precisão de cada teste, serão levantados o Erro Percentual Médio e o Erro Percentual Absoluto mês a mês. A primeira forma pode ser útil para indicar uma tendência dos erros de previsão, a segunda para avaliar a acurácia geral do método. O modelo, ou composição de modelos, que cumprir os requisitos mínimos levantados na Tabela 3 e apresentar os menores valores de erro será escolhido para ser implementado na empresa.



## 5. APLICAÇÃO E AVALIAÇÃO DAS SOLUÇÕES

### 5.1 Média Móvel

A primeira aplicação a ser testada é a da previsão de demanda através da média móvel. É o método com cálculo mais simples dentre os que serão verificados, e considera o valor da receita média dos últimos  $p$  períodos. Através da seguinte formulação:

$$P_R = \frac{D_{t-p} + \dots + D_{t-2} + D_{t-1}}{p}$$

$P_R$  = Previsão de demanda para o próximo período

$D_t$  = Receita faturada no período  $t$

$p$  = Número de períodos sendo analisado

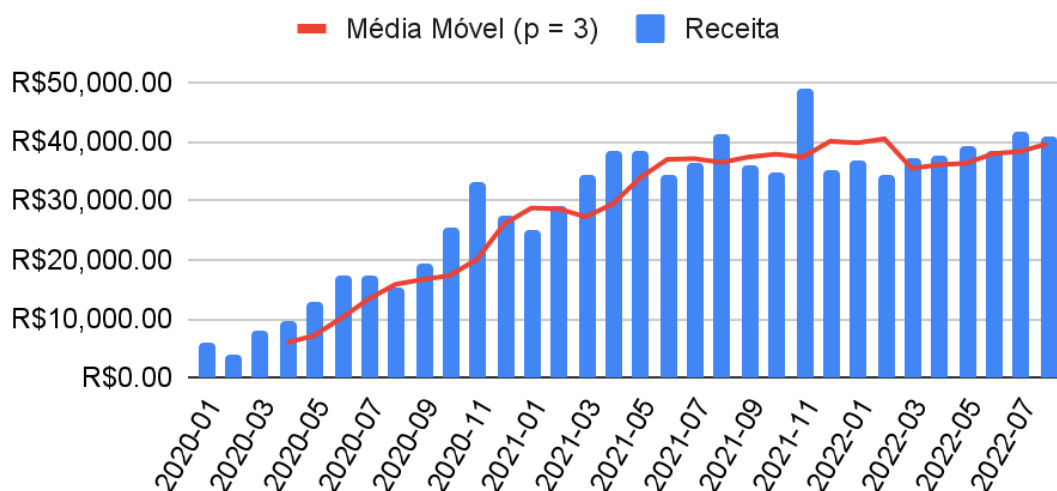
O valor atribuído a  $p$  determinará quantos meses serão considerados no cálculo da média, um número de períodos maior reduz a probabilidade de uma flutuação pontual afetar a previsão mas também reduz o impacto que os meses recentes desempenham na amostra total. Por isso, serão testados 3 valores distintos para  $p$  (3, 6 e 9). Cada teste de aplicação será realizado de duas formas distintas, a primeira utilizando a receita agregada total e outra separando a receita por tipo de canal.

#### 5.1.1 Média Móvel Aplicada na Receita Total

Iniciando o primeiro teste sobre a receita total, será aplicado o método da média móvel, com diferentes valores para  $p$  a fim de se encontrar o que mais se adapta ao histórico da empresa. De antemão é possível afirmar que é esperado um modelo que apresente uma performance melhor em períodos de maior estabilidade na receita mensal. Uma vez que o método considera os últimos valores da amostra para gerar uma previsão a partir da média. Obteve-se os seguintes gráficos:

Gráfico 6: Previsão da receita total com média móvel de três períodos

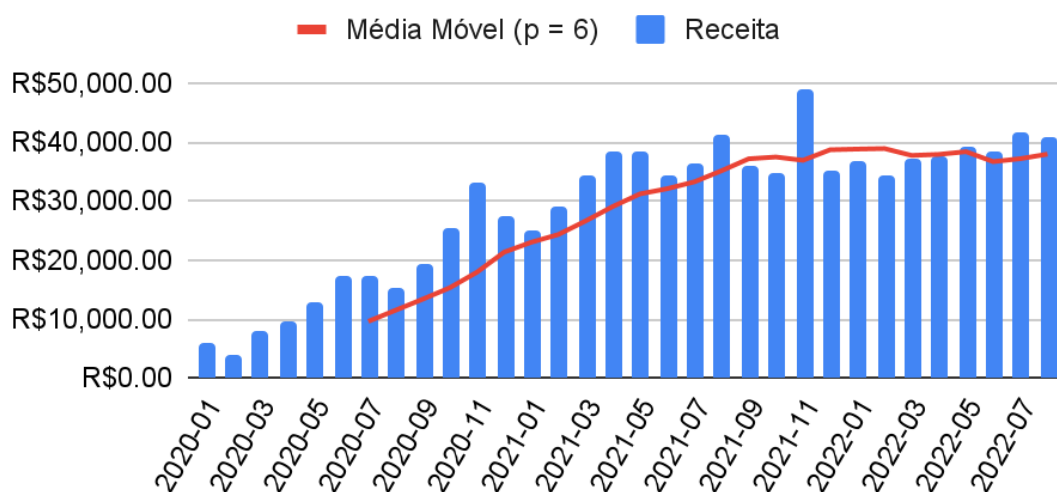
Evolução: Previsão da média móvel e receita faturada (p = 3)



Fonte: elaborado pelo autor.

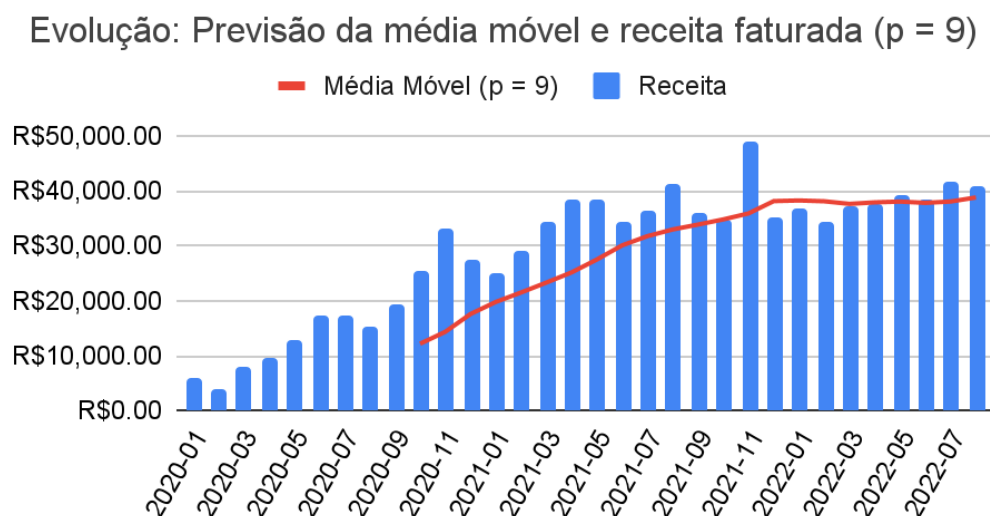
Gráfico 7: Previsão da receita total com média móvel de seis períodos

Evolução: Previsão da média móvel e receita faturada (p = 6)



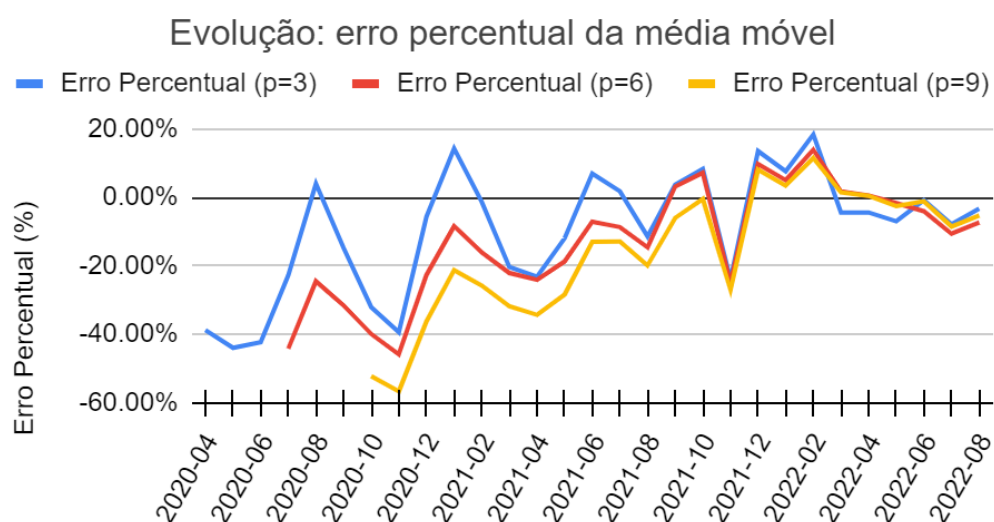
Fonte: elaborado pelo autor.

Gráfico 8: Previsão da receita total com média móvel de nove períodos



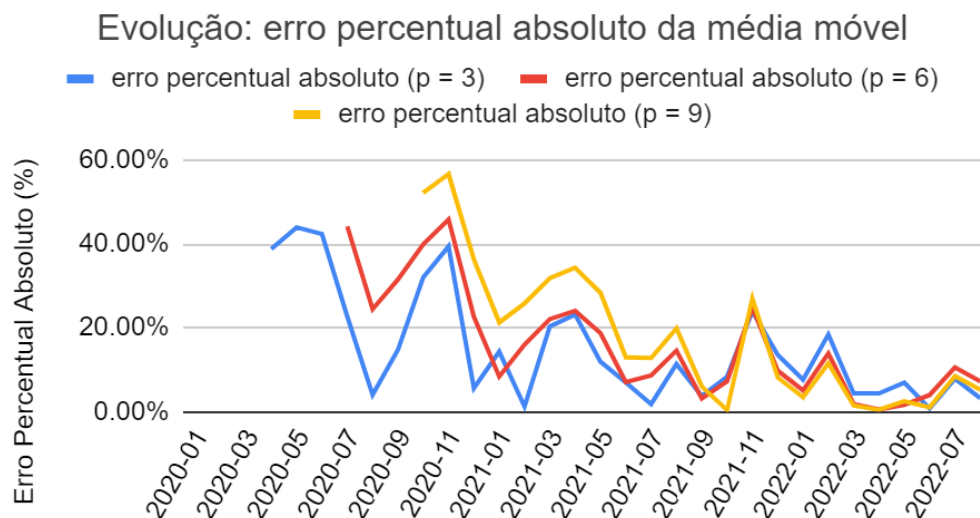
Fonte: elaborado pelo autor.

Com os valores calculados, é possível levantar a evolução dos erros observados entre a previsão estabelecida e a receita obtida. Seguem os gráficos de erro percentual e erro percentual absoluto:

Gráfico 9: Erros percentuais para diferentes valores de  $p$ 

Fonte: elaborado pelo autor.

Gráfico 10: Erros percentuais absolutos para diferentes valores de p



Fonte: elaborado pelo autor.

Assim, o modelo de média móvel aplicado a receita total seguintes tabelas resultado:

Tabela 4: Erros para média móvel (p=3) aplicado sobre a receita total

Ano	Erro Percentual Absoluto Máximo (mensal)	Erro Percentual Absoluto Médio (mensal)	Erro Percentual Médio (mensal)
2020	43,96%	27,07%	-26,17%
2021	23,96%	11,74%	3,60%
2022	18,44%	6,71%	0,18%

Fonte: elaborado pelo autor.

Tabela 5: Erros para média móvel (p=6) aplicado sobre a receita total

Ano	Erro Percentual Absoluto Máximo (mensal)	Erro Percentual Absoluto Médio (mensal)	Erro Percentual Médio (mensal)
2020	45,83%	34,77%	-34,76%
2021	24,88%	13,73%	-10,35%
2022	13,94%	5,61%	0,26%

Fonte: elaborado pelo autor.

Tabela 6: Erros para média móvel (p=9) aplicado sobre a receita total

Ano	Erro Percentual Absoluto Máximo (mensal)	Erro Percentual Absoluto Médio (mensal)	Erro Percentual Médio (mensal)
2020	56,69%	48,38%	-48,38%
2021	34,32%	19,07%	-17,71%
2022	11,55%	4,30%	-0,03%

Fonte: elaborado pelo autor.

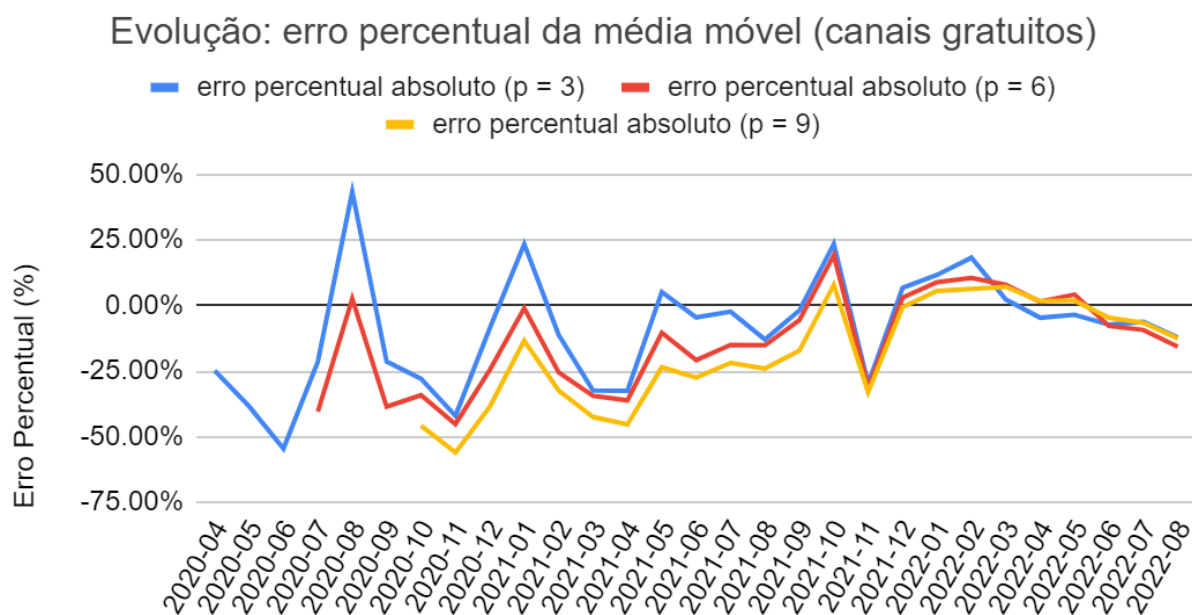
Ao analisar os gráficos e tabelas de erros verifica-se que o método apresenta uma tendência de subestimar a previsão, com uma melhora de precisão notável com o passar do tempo. A razão para isso é a velocidade de crescimento da empresa, que era superior nos períodos iniciais da análise. Portanto, o modelo com período de três meses é o que mais se adequa à base de forma geral.

Um fator que contribuiu para um maior erro percentual do modelo indicado foi o desempenho da Black Friday, que inflou a estimativa de receita dos meses subsequentes, além de ter sido subestimado inicialmente. Caso esse método seja escolhido ao fim dos testes, é necessário realizar um ajuste personalizado, para reduzir os erros gerados pela sazonalidade de vendas. O modelo com  $p = 6$  foi o que apresentou melhor desempenho geral em erros percentuais absolutos, e por isso, será utilizado para a comparação de modelos.

### 5.1.2 Média Móvel Aplicada na Receita de Origem Gratuita

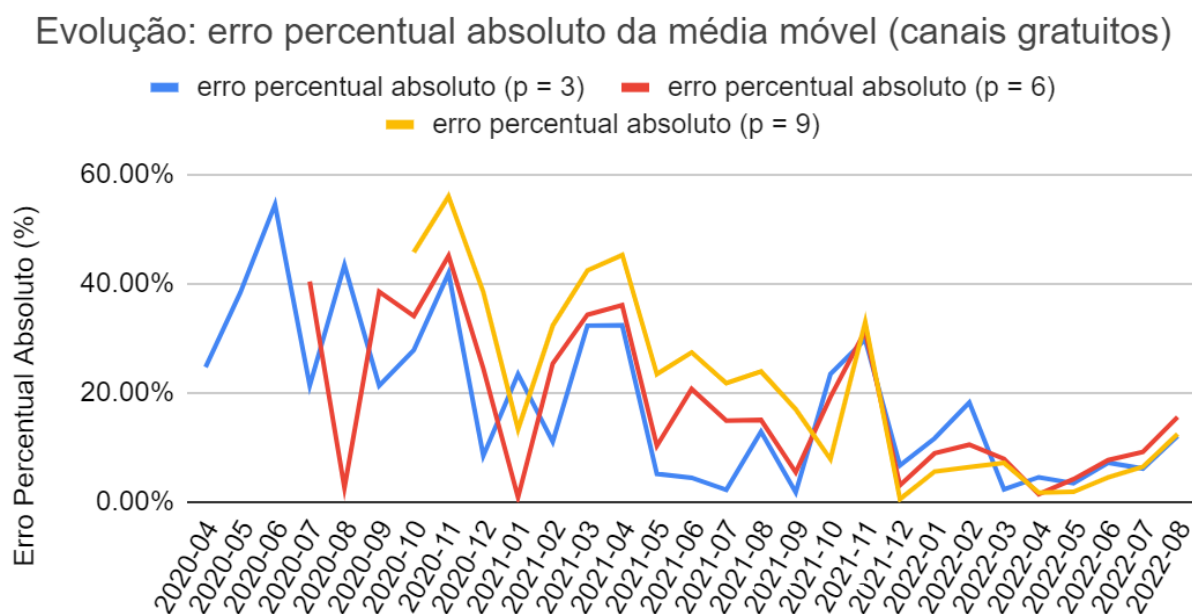
Ao aplicarmos a média móvel na receita de canais gratuitos e pagos de forma independente obtemos a evolução dos erros observados. Esses testes serão úteis para determinar se uma mescla de modelos para cada tipo de receita pode ser benéfica. Segue os resultados para a previsão da receita de canais grátis:

Gráfico 11: Erros percentuais para diferentes valores de p (canais grátis)



Fonte: elaborado pelo autor.

Gráfico 12: Erros percentuais absolutos para diferentes valores de p (canais grátis)



Fonte: elaborado pelo autor.

Tabela 7: Erros para média móvel em receita de canais grátis

p	Ano	Erro Percentual Absoluto Máximo (mensal)	Erro Percentual Absoluto Médio (mensal)	Erro Percentual Médio (mensal)
3	2021	32,42%	15,51%	-5,68%
3	2022	18,30%	8,25%	-0,15%
6	2021	36,10%	18,10%	-14,36%
6	2022	15,63%	8,23%	0,07%
9	2021	45,27%	24,05%	-22,73%
9	2022	12,49%	5,82%	-0,07%

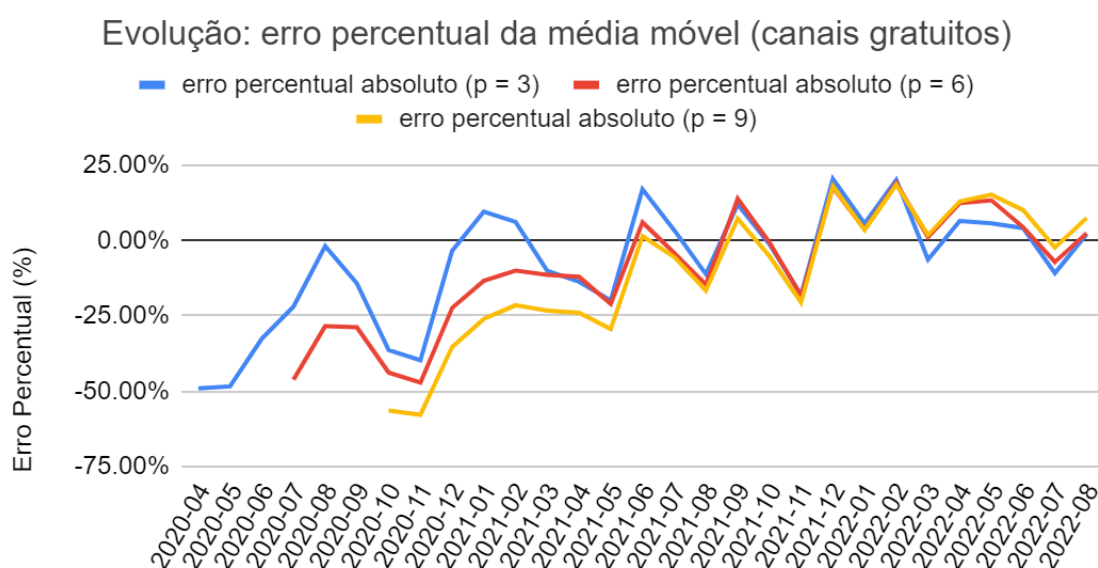
Fonte: elaborado pelo autor.

A previsão por média móvel aplicada apenas na receita de origem gratuita não proporcionou nenhum ganho relevante na precisão do modelo e, portanto, não deve ser usado em uma mescla de modelos.

### 5.1.3 Média Móvel Aplicada na Receita de Origem Paga

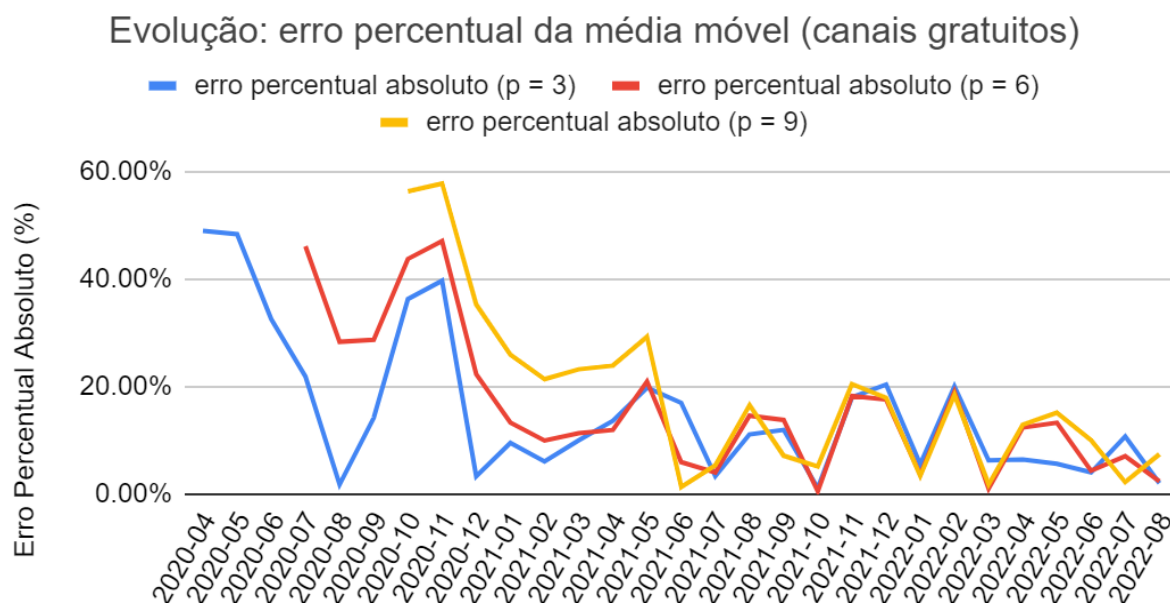
O resultado da aplicação da média móvel na receita de origem paga está colocado a seguir:

Gráfico 13: Erros percentuais para diferentes valores de p (canais pagos)



Fonte: elaborado pelo autor.

Gráfico 14: Erros percentuais absolutos para diferentes valores de p (canais pagos)



Fonte: elaborado pelo autor.

Tabela 8: Erros para média móvel em receita de canais pagos

p	Ano	Erro Percentual Absoluto Máximo (mensal)	Erro Percentual Absoluto Médio (mensal)	Erro Percentual Médio (mensal)
3	2021	20,44%	11,89%	-0,44%
3	2022	20,09%	7,66%	3,37%
6	2021	21,02%	11,91%	-5,65%
6	2022	19,10%	7,95%	6,17%
9	2021	29,31%	16,52%	-12,09%
9	2022	18,52%	8,99%	8,40%

Fonte: elaborado pelo autor.

A previsão por média móvel aplicada apenas na receita de origem paga não demonstrou nenhum ganho relevante na precisão do modelo e, portanto, também não deve ser utilizado em uma possível mescla de modelos.

## 5.2 Regressão Linear

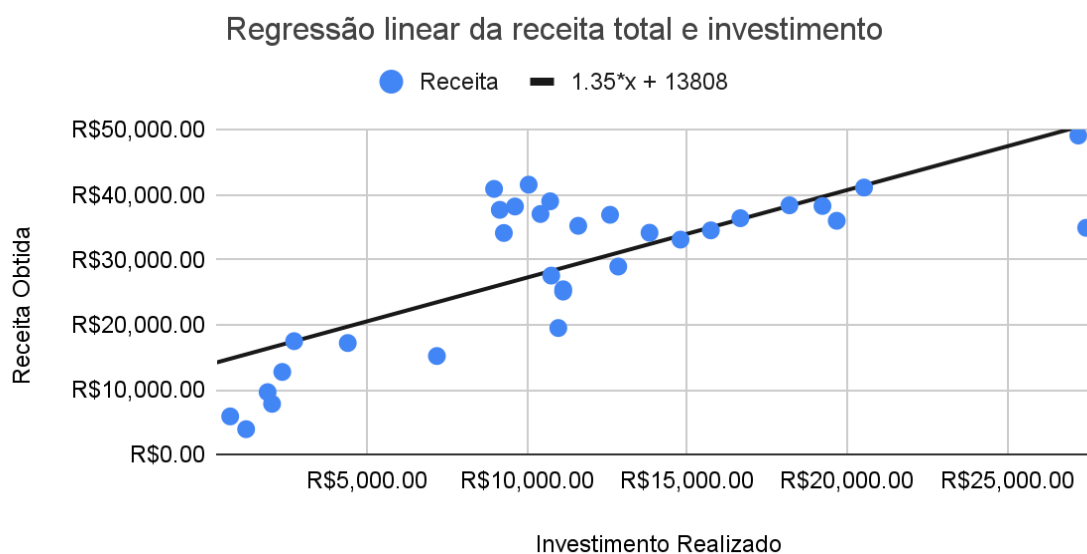
O segundo teste será a aplicação da regressão linear para a previsão de receita a partir dos valores de investimento. A Análise de regressão é uma ferramenta muito útil por ser capaz de gerar uma equação que demonstra a relação entre a variável interesse e um ou mais conjuntos de variáveis preditoras. (Montgomery, 2021). O método tem o potencial de minimizar os erros oriundos dos picos de vendas recorrentes, uma vez que investimentos nesses períodos também costumam ser maiores. É esperado que o modelo apresente erros absolutos menores nos meses de maior receita e investimento (ex. black friday), se contrapondo ao modelo da média móvel.

Para realizar o cálculo, basta plotar os valores mensais em pontos de um gráfico de dispersão com o eixo horizontal representando o investimento realizado e o eixo vertical a receita obtida. Dessa forma será possível estabelecer uma linha de tendência reta, que poderá ser usada para estimar valores de receita a partir do investimento previsto para o mês seguinte.

### 5.2.1 Regressão Linear Aplicada na Receita Total

Considerando os valores de receita total e investimento em marketing na aplicação da regressão, foi modelada uma linha tendência linear representada pela equação  $y = 1,35x + 13808$  com o valor de receita (y) em função do investimento (x). O coeficiente de determinação da linha gerada é de  $R^2 = 0.593$ . A métrica, que varia de 0 a 100%, indica o nível de ajuste entre o modelo e os dados, como o  $R^2$  observado é relativamente baixo, é provável que ocorram picos de erros em alguns meses.

Gráfico 15: Regressão linear da receita total e investimento



Fonte: elaborado pelo autor.

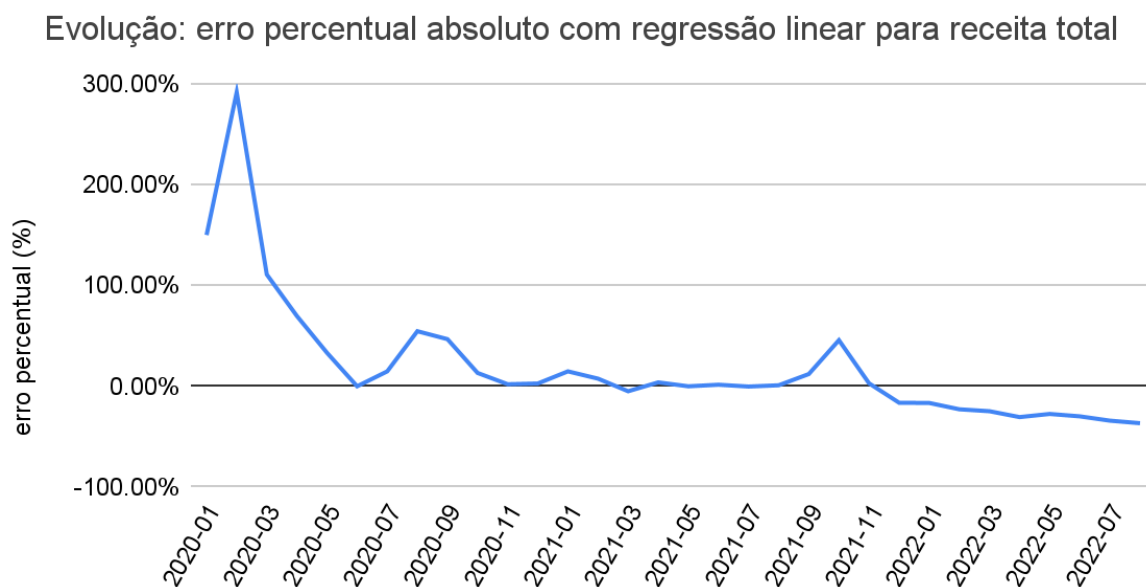
Aplicando a linha de tendência encontrada para a construção de um modelo de previsão de receita a partir dos valores esperados de investimento encontra-se os seguintes resultados:

Gráfico 16: Previsão da receita total com regressão linear da receita e investimento



Fonte: elaborado pelo autor.

Gráfico 17: Erro percentual absoluto no modelo com regressão linear aplicada na previsão da receita total



Fonte: elaborado pelo autor.

Gráfico 18: Erro percentual absoluto no modelo com regressão linear aplicada na previsão da receita total



Fonte: elaborado pelo autor.

Tabela 9: Erros do modelo com regressão linear aplicada na previsão da receita total

Ano	Erro Percentual Absoluto Máximo (mensal)	Erro Percentual Absoluto Médio (mensal)	Erro Percentual Médio (mensal)
2021	45,69%	9,25%	5,63%
2022	36,66%	27,88%	-27,87%

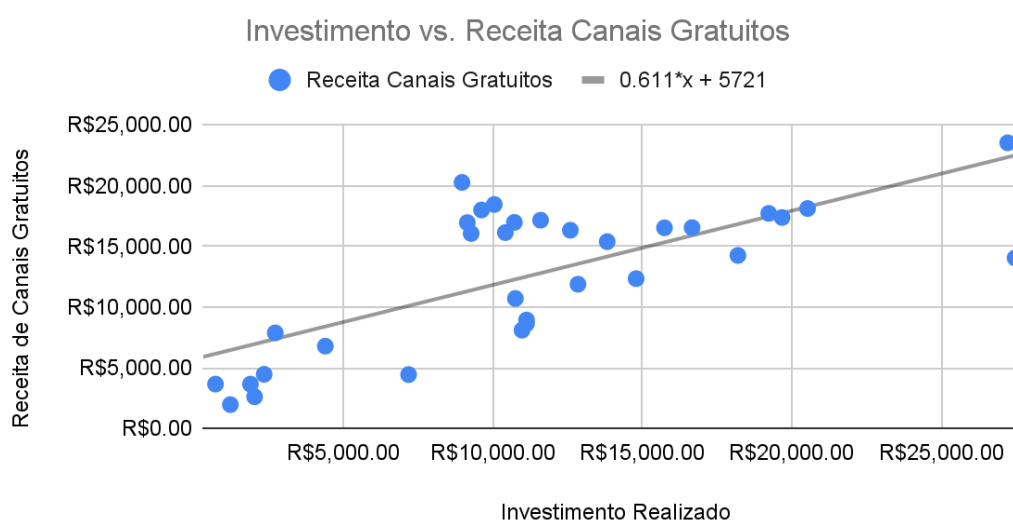
Fonte: elaborado pelo autor.

Com os valores de erros absolutos percentuais apontados no Gráfico 17 e Tabela 9 é possível concluir que o modelo de regressão linear aplicado na receita total não possui uma acurácia boa o suficiente para ser implementado. Entretanto, vale ressaltar que o modelo aparenta performar melhor durante os meses de maior investimento, em especial nos períodos sazonais de expansão de vendas.

### 5.2.2 Regressão Linear Aplicada na Receita de Origem Gratuita

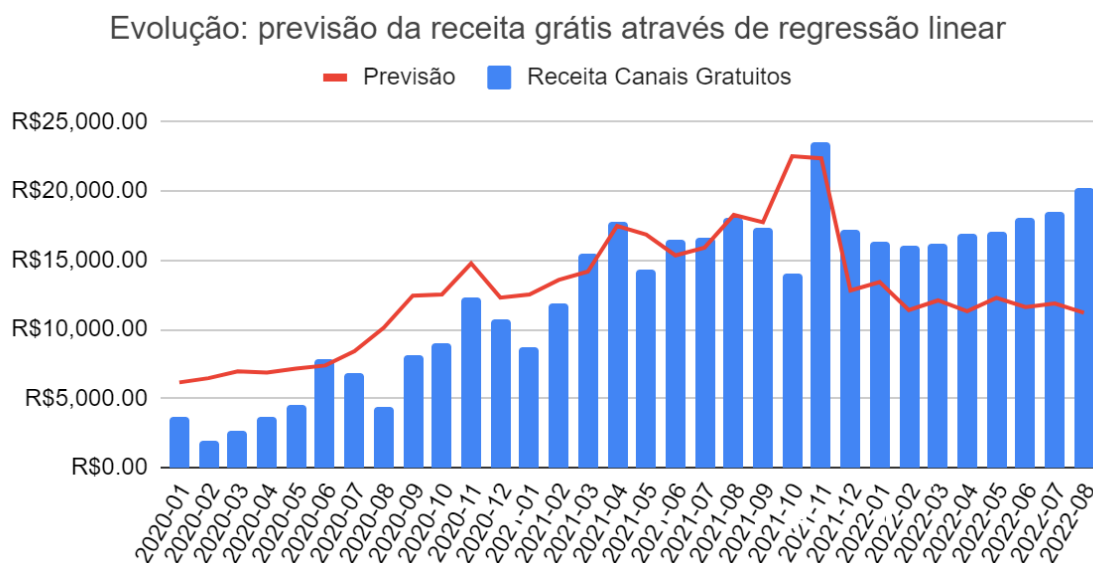
Utiliza-se o processo de regressão linear aplicado apenas para previsão da receita de canais gratuitos. Assim, obtém-se a linha de tendência representada pela equação  $y = 0,611x + 5721$  e com coeficiente de determinação de  $R^2 = 0.512$ . Ao utilizar a equação obtida temos:

Gráfico 19: Regressão linear da receita de canais grátis e investimento



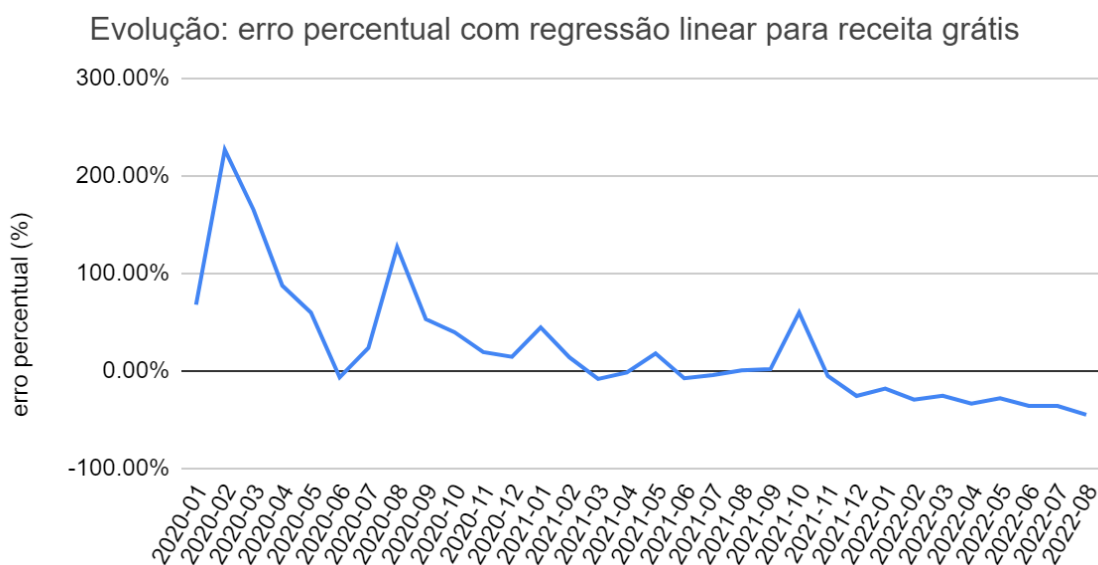
Fonte: elaborado pelo autor.

Gráfico 20: Previsão da receita gratuita com regressão linear da receita de canais grátis e investimento



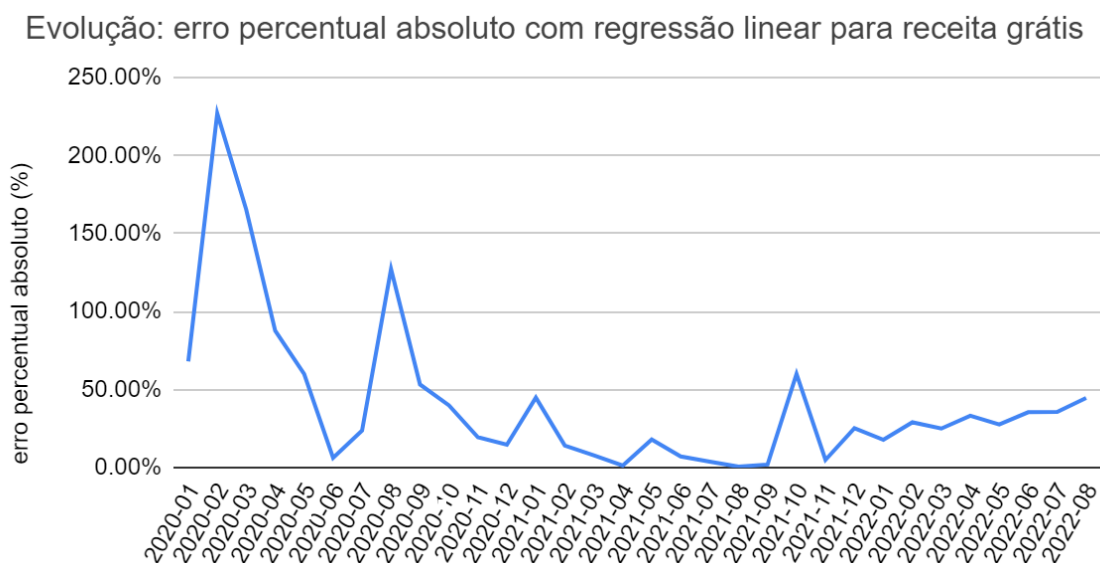
Fonte: elaborado pelo autor.

Gráfico 21: Erro percentual absoluto no modelo com regressão linear aplicada na previsão da receita de canais grátis



Fonte: elaborado pelo autor.

Gráfico 22: Erro percentual absoluto no modelo com regressão linear aplicada na previsão da receita de canais grátis



Fonte: elaborado pelo autor.

Tabela 10: Erros do modelo com regressão linear aplicada na previsão da receita de origem gratuita

Ano	Erro Percentual Absoluto Máximo (mensal)	Erro Percentual Absoluto Médio (mensal)	Erro Percentual Médio (mensal)
2021	60,14%	15,92%	7,43%
2022	44,74%	31,15%	-31,15%

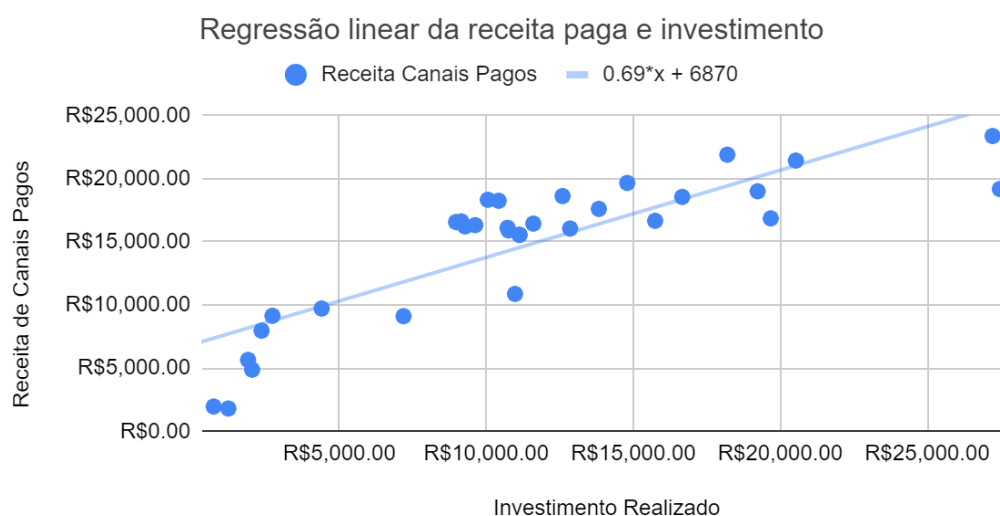
Fonte: elaborado pelo autor.

A previsão realizada através da regressão linear aplicada apenas na receita de origem gratuita apresentou acurácia pior do que quando aplicada na receita total. Além disso, apresentou valores de erros máximos bem elevados. Com isso podemos descartar a possibilidade de uso do modelo em questão.

### 5.2.3 Regressão Linear Aplicada na Receita de Origem Paga

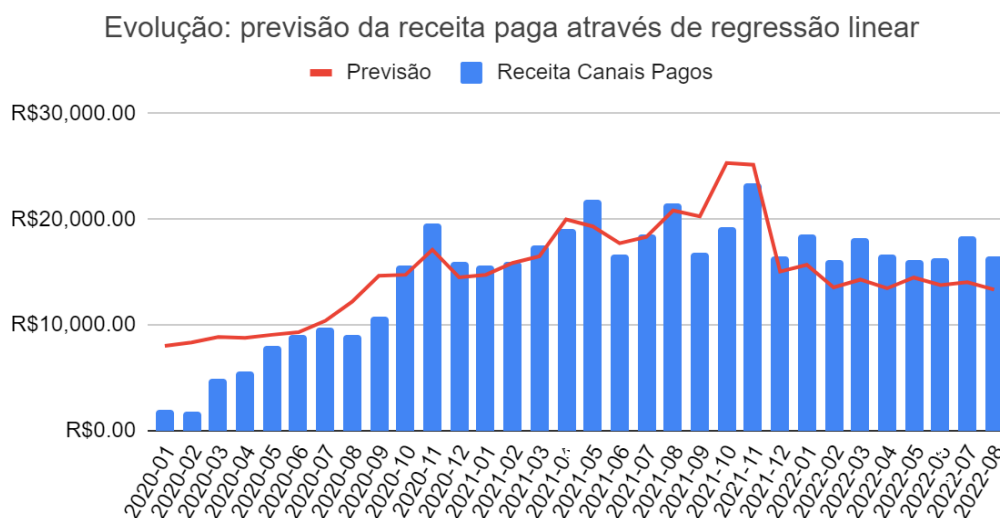
A regressão linear com valores de receita de origem paga a investimentos gera uma linha de tendência com a equação  $y = 0,646x + 7560$  e com coeficiente de determinação de  $R^2 = 0.696$ . Segue os resultados obtidos:

Gráfico 23: Regressão linear da receita de canais pagos e investimento



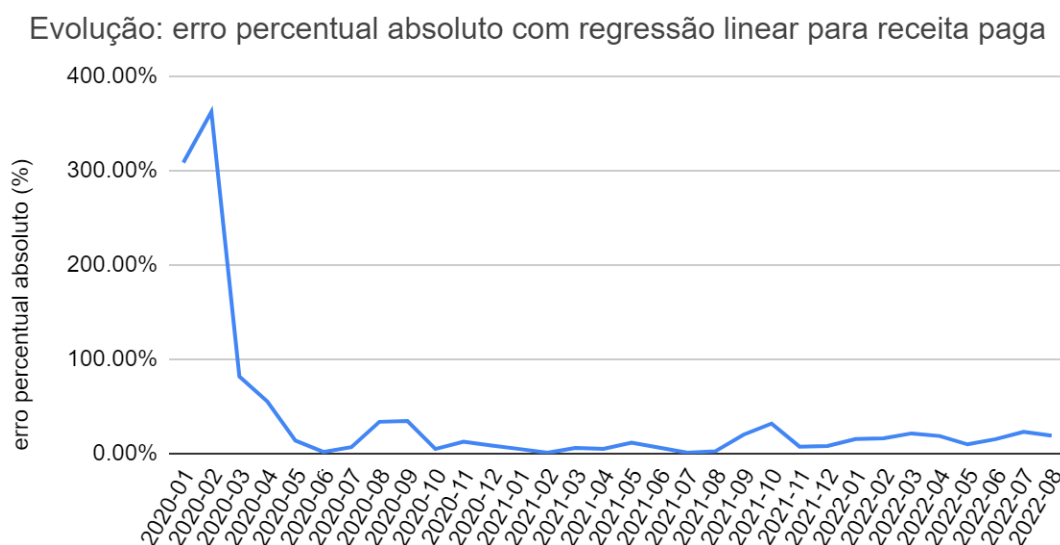
Fonte: elaborado pelo autor.

Gráfico 24: Previsão da receita paga com regressão linear da receita de canais pagos e investimento



Fonte: elaborado pelo autor.

Gráfico 25: Erro percentual absoluto no modelo com regressão linear aplicada na previsão da receita de canais pagos



Fonte: elaborado pelo autor.

Tabela 11: Erros do modelo com regressão linear aplicada na previsão da receita de origem paga

Ano	Erro Percentual Absoluto Máximo (mensal)	Erro Percentual Absoluto Médio (mensal)	Erro Percentual Médio (mensal)
2021	32,00%	9,02%	2,92%
2022	23,28%	17,60%	-17,59%

Fonte: elaborado pelo autor.

O modelo de regressão linear aplicado na receita oriunda de canais pagos foi o que apresentou melhor adequação. Esse resultado era esperado uma vez que a linha de tendência do modelo foi o que apresentou maior coeficiente de determinação ( $R^2$ ). Apesar disso, o método não demonstrou precisão suficiente, com valores de erros maiores que os adquiridos pela aplicação da média móvel.

### 5.3 Suavização Exponencial Simples

O método exponencial é uma das formas mais utilizadas para a suavização de séries temporais discretas a fim de prever o futuro imediato. A ideia do método é suavizar a série original de forma similar a aplicação da média móvel, porém a suavização exponencial determina uma maior influência de valores recentes em detrimento de valores distantes (Ostertagova, 2012). A fórmula a ser utilizada na ferramenta de análise Google Sheets seguirá o modelo:

$$P_t = \alpha \cdot y_{t-1} + (1 - \alpha) \cdot P_{t-1}$$

$P_t$ : *Previsão de demanda para o próximo período*

$P_{t-1}$ : *Previsão de demanda no período anterior*

$y_{t-1}$ : *Receita faturada no período anterior*

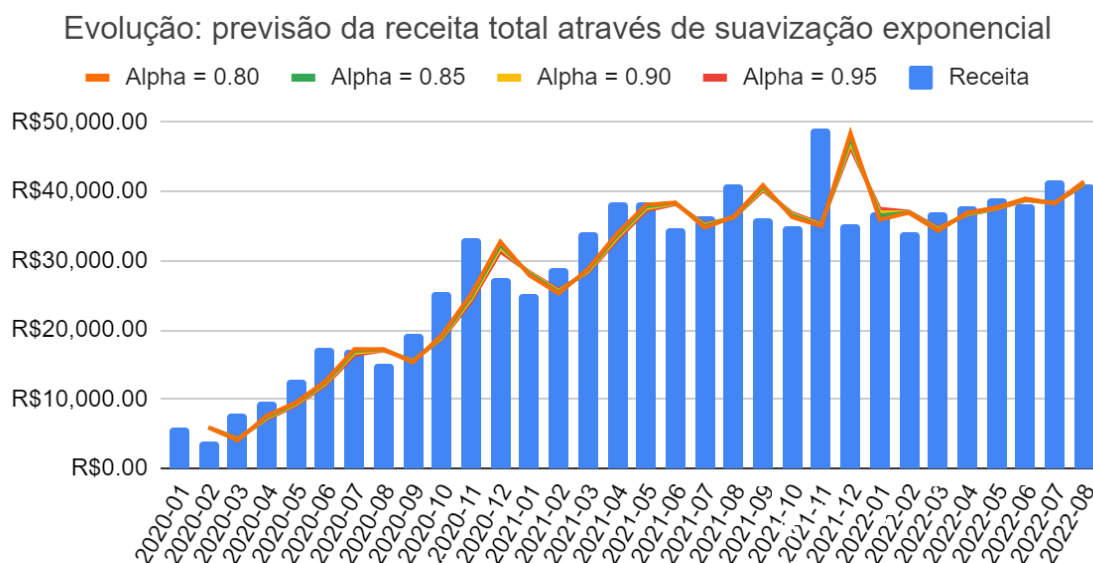
$\alpha$ : *Coefficiente de suavização*

No total, 19 coeficientes de suavização serão testados a fim de encontrar a opção que mais se adequa à base de dados. A lista de coeficientes varia de 5% a 95%, com intervalo de 5% entre cada valor. Os 4 coeficientes que apresentarem os menores valores de erros médios absolutos serão escolhidos para integrar os gráficos e as tabelas comparativas.

#### 5.3.1 Suavização Exponencial Simples Aplicada na Receita Total

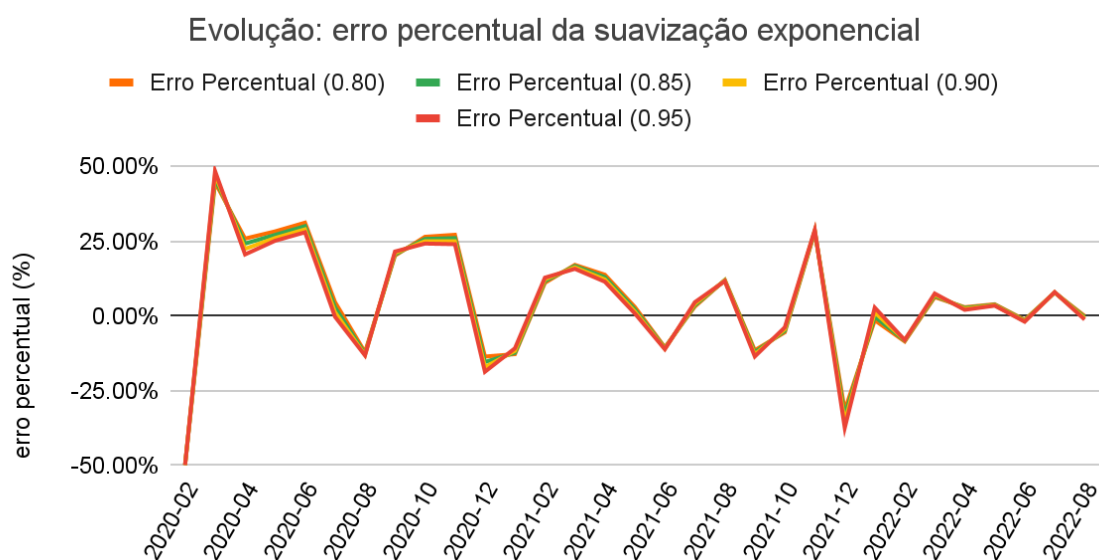
Utilizando o método da suavização exponencial simples na receita total, constata-se que os coeficientes mais adequados foram os 4 maiores: 0,95; 0,9; 0,85 e 0,80. Seguem os resultados:

Gráfico 26: Previsões da receita total com suavização exponencial simples



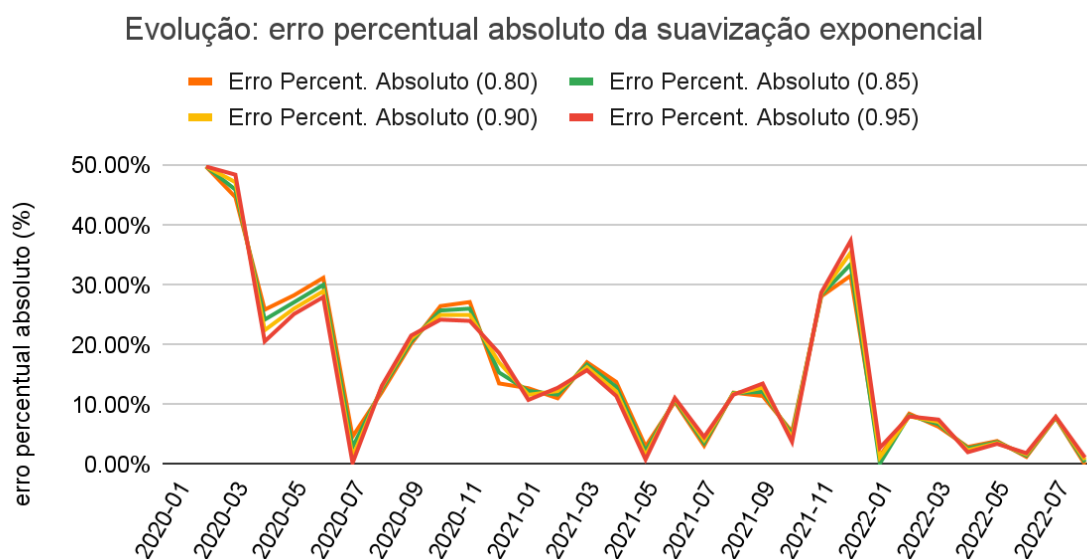
Fonte: elaborado pelo autor.

Gráfico 27: Erro percentual em valores suavização exponencial simples para receita total



Fonte: elaborado pelo autor.

Gráfico 28: Erro percentual absoluto em valores da suavização exponencial simples para receita total



Fonte: elaborado pelo autor.

Tabela 12: Erros do modelo com suavização exponencial simples aplicada na previsão da receita total

$\alpha$	Ano	Erro Percentual Absoluto Máximo (mensal)	Erro Percentual Absoluto Médio (mensal)	Erro Percentual Médio (mensal)
0,80	2020	49,82%	25,83%	12,10%
0,80	2021	31,57%	13,34%	1,38%
0,80	2022	8,50%	4,03%	0,00%
0,85	2020	49,82%	25,51%	11,35%
0,85	2021	33,48%	13,44%	1,15%
0,85	2022	8,23%	3,92%	1,34%
0,90	2020	49,82%	25,18%	10,65%
0,90	2021	35,42%	13,52%	0,95%
0,90	2022	8,06%	4,10%	1,46%
0,95	2020	49,82%	24,90%	10,01%
0,95	2021	37,40%	13,55%	0,77%
0,95	2022	8,04%	4,37%	1,59%

Fonte: elaborado pelo autor.

Os modelos de suavização exponenciais que apresentaram os menores valores de erros para a amostra possuem um formato bem similar entre si. Os erros percentuais desse método flutuam com uma tendência de alternância entre valores positivos e negativos vide o Gráfico 24.

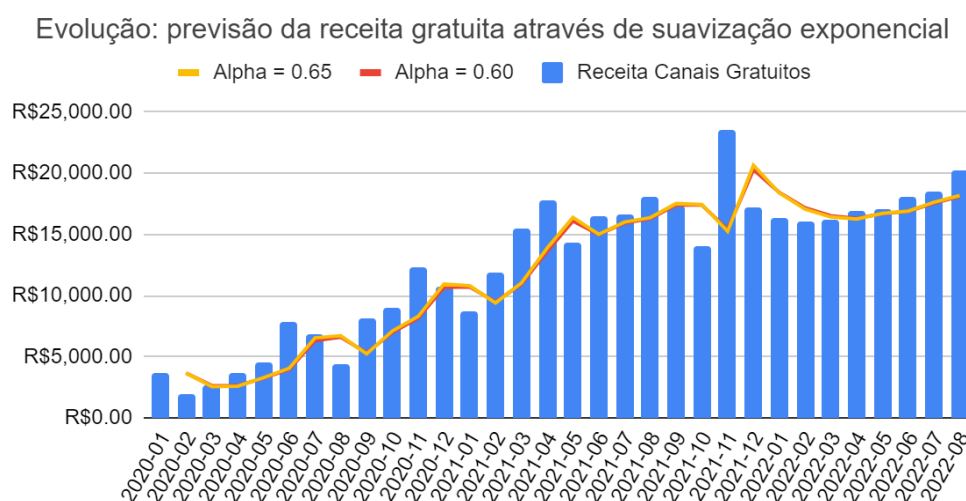
O erro absoluto demonstra uma tendência decrescente ao longo do tempo, por conta da maior estabilidade na receita nos períodos finais. O modelo, assim como na média móvel, também tem uma performance pior nos períodos de pico de vendas, ao subestimar o pico e superestimar os meses seguintes. O modelo com coeficiente  $\alpha = 0,85$  obteve maior adequação e só poderia ser utilizado na implementação final caso ajustes para tratar as sazonalidades sejam realizados.

Os coeficientes diferentes não influenciaram o modelo de forma significativa, por possuírem valores próximos e exercerem influência similar nas previsões. Por isso, serão testados apenas dois valores para os modelos exponenciais simples aplicados nas receitas separadas por origem.

### 5.3.2 Suavização Exponencial Simples Aplicada na Receita de Origem Gratuita

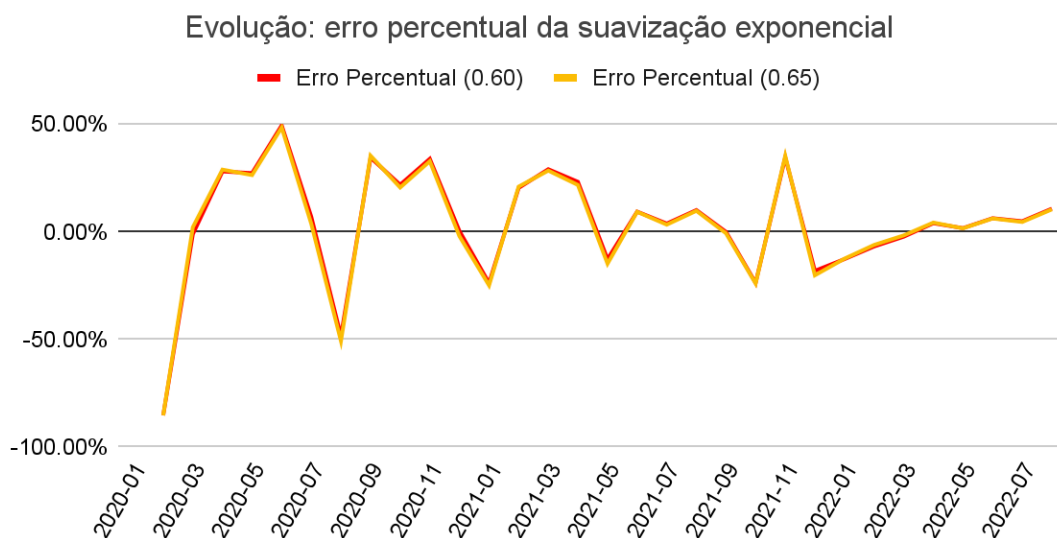
Para a receita de origem gratuita, os coeficientes que apresentaram os melhores resultados foram  $\alpha = 0,60$  e  $\alpha = 0,65$ . Os resultados foram colocados a seguir:

Gráfico 29: Previsões da receita de origem gratuita com suavização exponencial simples



Fonte: elaborado pelo autor.

Gráfico 30: Erro percentual nos valores obtidos com a suavização exponencial simples para receita de origem gratuita



Fonte: elaborado pelo autor.

Tabela 13: Erros do modelo com suavização exponencial simples aplicada na previsão da receita de origem grátis

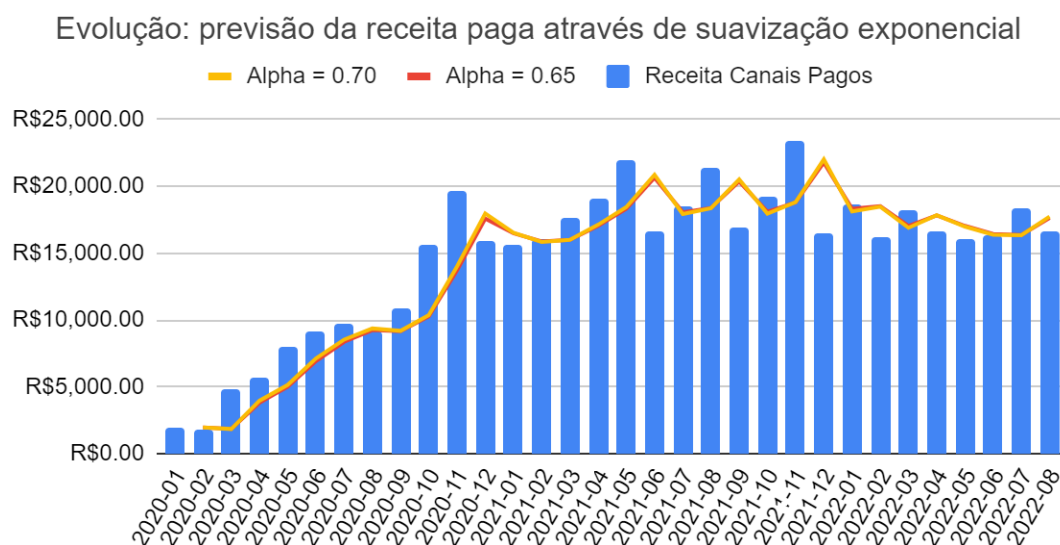
$\alpha$	Ano	Erro Percentual Absoluto Máximo (mensal)	Erro Percentual Absoluto Médio (mensal)	Erro Percentual Médio (mensal)
0,60	2020	85,30%	30,65%	6,13%
0,60	2021	34,62%	17,42%	4,35%
0,60	2022	12,68%	6,18%	0,74%
0,65	2020	85,30%	30,58%	5,48%
0,65	2021	35,27%	17,79%	3,69%
0,65	2022	12,49%	5,92%	0,85%

Fonte: elaborado pelo autor.

### 5.3.3 Suavização Exponencial Simples Aplicada na Receita de Origem Paga

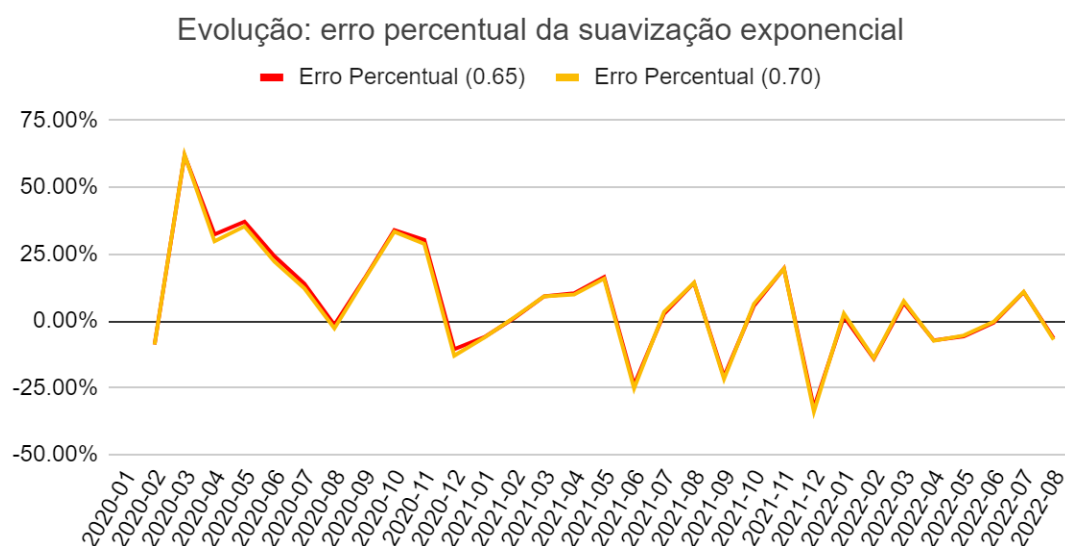
Para a receita de origem gratuita, os coeficientes que apresentaram os melhores resultados foram  $\alpha = 0,60$  e  $\alpha = 0,65$ . Os resultados foram colocados a seguir:

Gráfico 31: Previsões da receita de origem paga com suavização exponencial simples



Fonte: elaborado pelo autor.

Gráfico 32: Erro percentual em valores suavização exponencial simples para receita de origem paga



Fonte: elaborado pelo autor.

Tabela 14: Erros do modelo de suavização exponencial simples aplicada na previsão da receita de origem paga

$\alpha$	Ano	Erro Percentual Absoluto Máximo (mensal)	Erro Percentual Absoluto Médio (mensal)	Erro Percentual Médio (mensal)
0,65	2020	61.85%	24.56%	20,78%
0,65	2021	32.41%	13.52%	-0,33%
0,65	2022	14.12%	6.68%	-1,89%
0,70	2020	62.01%	23.96%	19,50%
0,70	2021	33.83%	13.90%	-0,59%
0,70	2022	13.86%	6.88%	-1,62%

Fonte: elaborado pelo autor.

#### 5.4 Suavização Exponencial de Holt-Winters

A suavização de Holt-Winters é uma das variantes do método exponencial e é indicado para produzir previsões de curto prazo (Chatfield, 1988). Uma das dificuldades impostas pelo método é a determinação de valores para as constantes que irão ajustar a equação de suavização geral ( $\alpha$ ), a equação de suavização da tendência ( $\beta$ ) e equação de suavização de sazonalidade ( $\gamma$ ). O modelo foi construído no Google Sheets, e para estimação dos valores ótimos das constantes foi utilizado a extensão do Solver para Google Drive, de forma a reduzir o erro quadrático médio das previsões do modelo.

##### 5.4.1 Suavização de Holt-Winters Aplicada na Receita Total

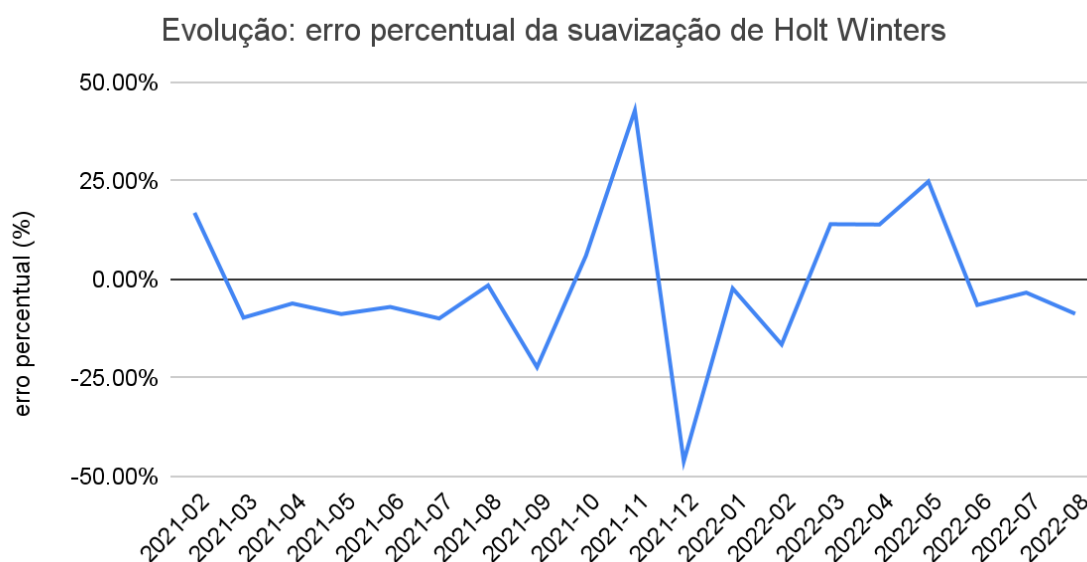
Ao utilizar o Solver para encontrar os valores otimizados para as constantes, determina-se:  $\alpha = 0,99$ ;  $\beta = 0,33$  e  $\gamma = 0,75$ . Com as constantes definidas o modelo obtém os seguintes resultados:

Gráfico 33: Previsões da receita total com suavização exponencial de Holt-Winters



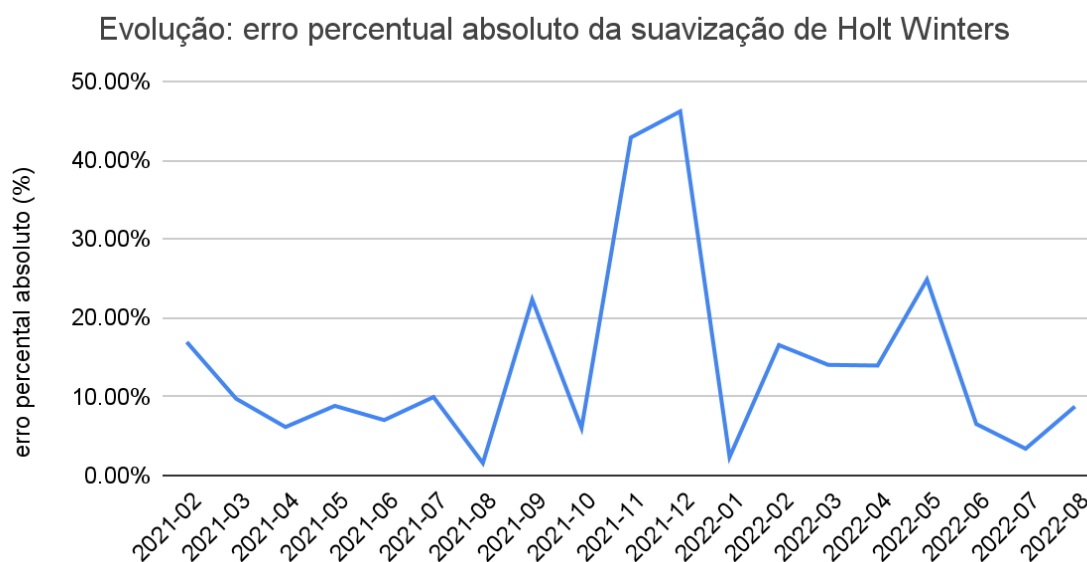
Fonte: elaborado pelo autor.

Gráfico 34: Erro percentual em valores obtidos para receita total com suavização exponencial de Holt-Winters



Fonte: elaborado pelo autor.

Gráfico 35: Erro percentual absoluto em valores obtidos para receita total com suavização exponencial de Holt-Winters



Fonte: elaborado pelo autor.

Tabela 15: Erros do modelo de suavização exponencial de Holt-Winters aplicada na previsão da receita total

Ano	Erro Percentual Absoluto Máximo (mensal)	Erro Percentual Absoluto Médio (mensal)	Erro Percentual Médio (mensal)
2021	46,20%	16,13%	-4,17%
2022	24,85%	11,28%	1,92%

Fonte: elaborado pelo autor.

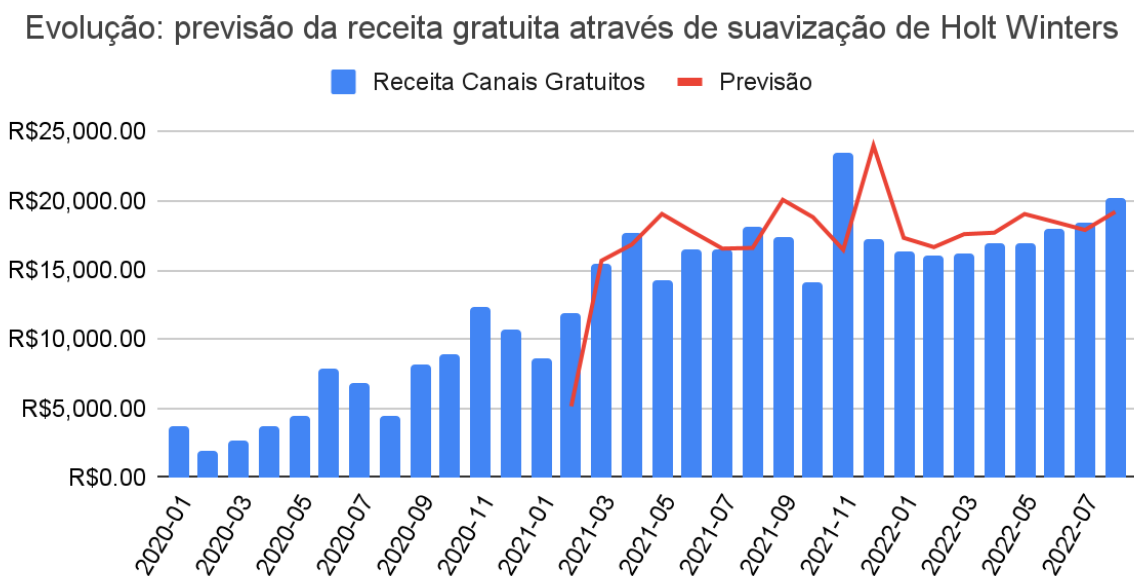
O modelo não demonstrou uma acurácia boa apesar de ser o único, dentre os testados, que considera fatores de sazonalidade e tendência da amostra. Uma causa provável para a má adequação do método é a base de dados limitada, falta de constância nas elevações de receita proporcionadas por picos de vendas sazonais e diferentes velocidades de crescimento ao longo dos anos.

Assim como os outros modelos, com exceção da regressão linear, os valores de investimento não estão sendo considerados. Isso faz com que erros em períodos de picos de vendas sejam acentuados. Com isso, o método de Holt-Winter aplicado à receita total não deverá ser incorporado no modelo final.

#### 5.4.2 Suavização de Holt-Winters Aplicada na Receita de Origem Gratuita

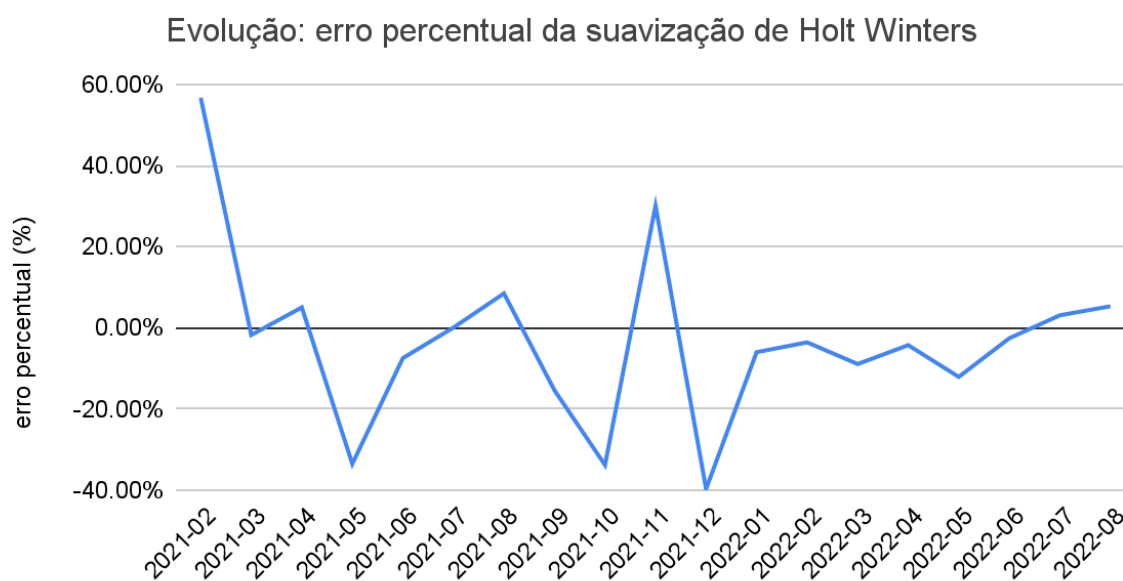
Para a receita de origem gratuita, com a otimização realizada pelo Solver, obteve as seguintes constantes:  $\alpha = 0,916$  ;  $\beta = 0,001$  e  $\gamma = 0,497$ . Os resultados podem ser conferidos a seguir:

Gráfico 36: Previsões da receita gratuita com suavização exponencial de Holt-Winters



Fonte: elaborado pelo autor.

Gráfico 37: Erro percentual em valores obtidos para receita gratuita com suavização exponencial de Holt-Winters



Fonte: elaborado pelo autor.

Gráfico 38: Erro percentual absoluto em valores obtidos para receita gratuita com suavização exponencial de Holt-Winters



Fonte: elaborado pelo autor.

Tabela 16: Erros do modelo de suavização exponencial de Holt-Winters aplicada na previsão da receita gratuita

Ano	Erro Percentual Absoluto Máximo (mensal)	Erro Percentual Absoluto Médio (mensal)	Erro Percentual Médio (mensal)
2021	56,74%	21,11%	-2,87%
2022	12,13%	5,75%	-3,65%

Fonte: elaborado pelo autor.

#### 5.4.3 Suavização de Holt-Winters Aplicada na Receita de Origem Paga

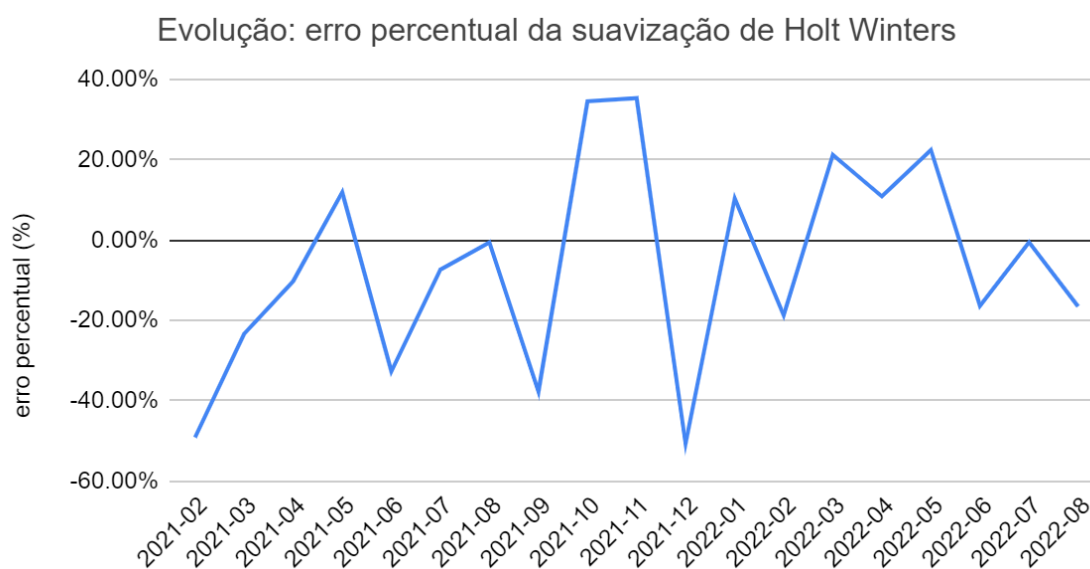
Para a receita de origem paga, o Solver encontrou os seguintes valores:  $\alpha = 0,999$  ;  $\beta = 0,421$  e  $\gamma = 0,999$ . Seguem os resultados do modelo:

Gráfico 39: Previsões da receita paga com suavização exponencial de Holt-Winters



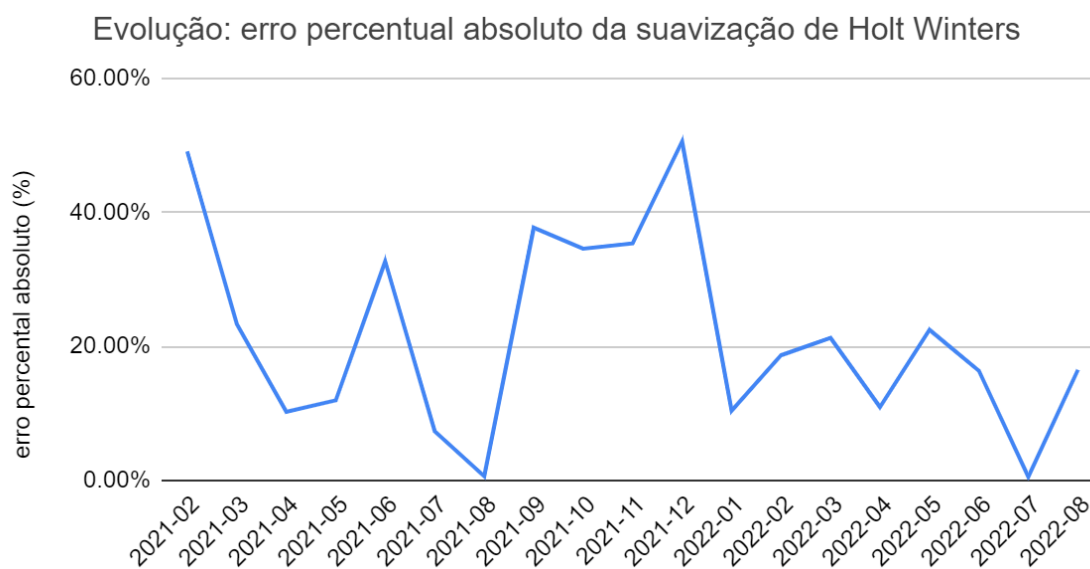
Fonte: elaborado pelo autor.

Gráfico 40: Erro percentual em valores obtidos para receita paga com suavização exponencial de Holt-Winters



Fonte: elaborado pelo autor.

Gráfico 41: Erro percentual absoluto em valores obtidos para receita paga com suavização exponencial de Holt-Winters



Fonte: elaborado pelo autor.

Tabela 17: Erros do modelo de suavização exponencial de Holt-Winters aplicada na previsão da receita paga

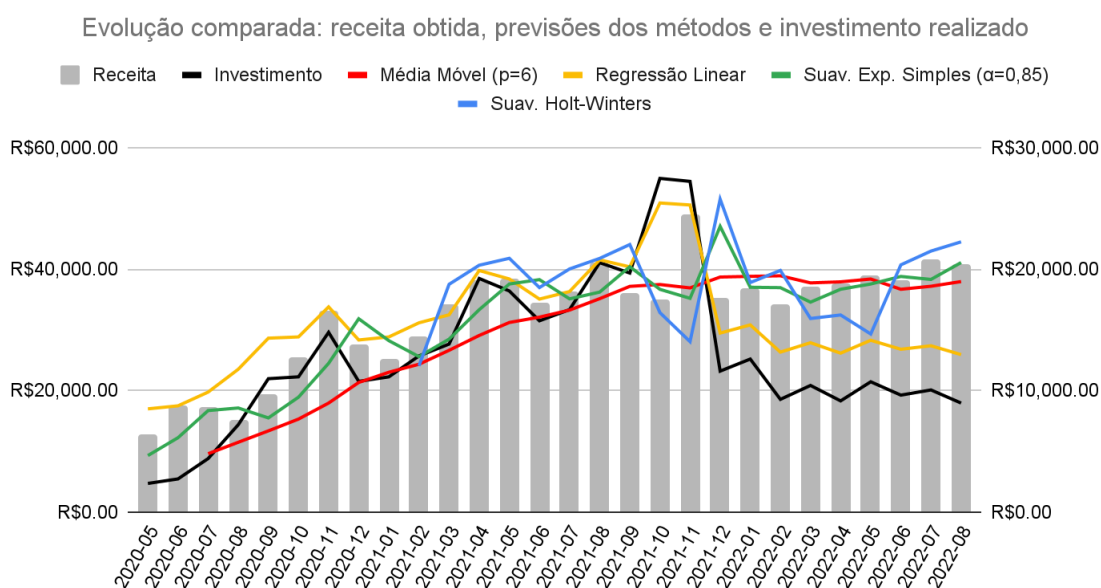
Ano	Erro Percentual Absoluto Máximo (mensal)	Erro Percentual Absoluto Médio (mensal)	Erro Percentual Médio (mensal)
2021	50,59%	26,68%	-11,78%
2022	22,47%	14,62%	1,63%

Fonte: elaborado pelo autor.

### 5.5 Avaliação dos modelos para Receita Total

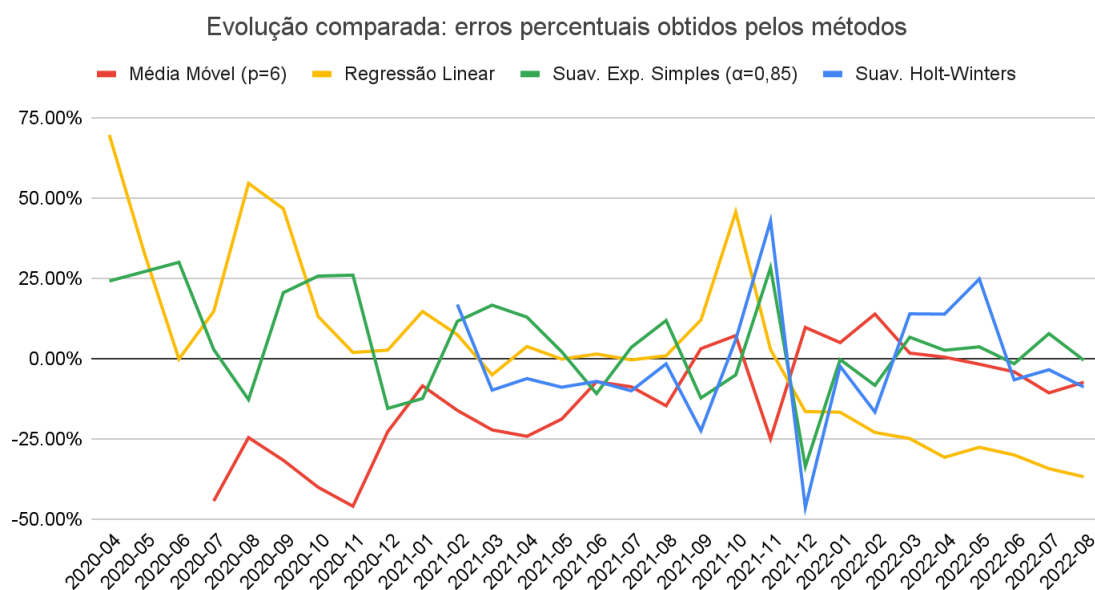
Nessa sessão se buscará comparar os valores obtidos em cada um a fim de facilitar o processo de escolha para implementação do(s) método(s). Nessa etapa serão analisados os valores de previsão sem considerar a distinção por origem da receita.

Gráfico 42: Previsões dos diferentes métodos, receita obtida e investimento realizado em marketing



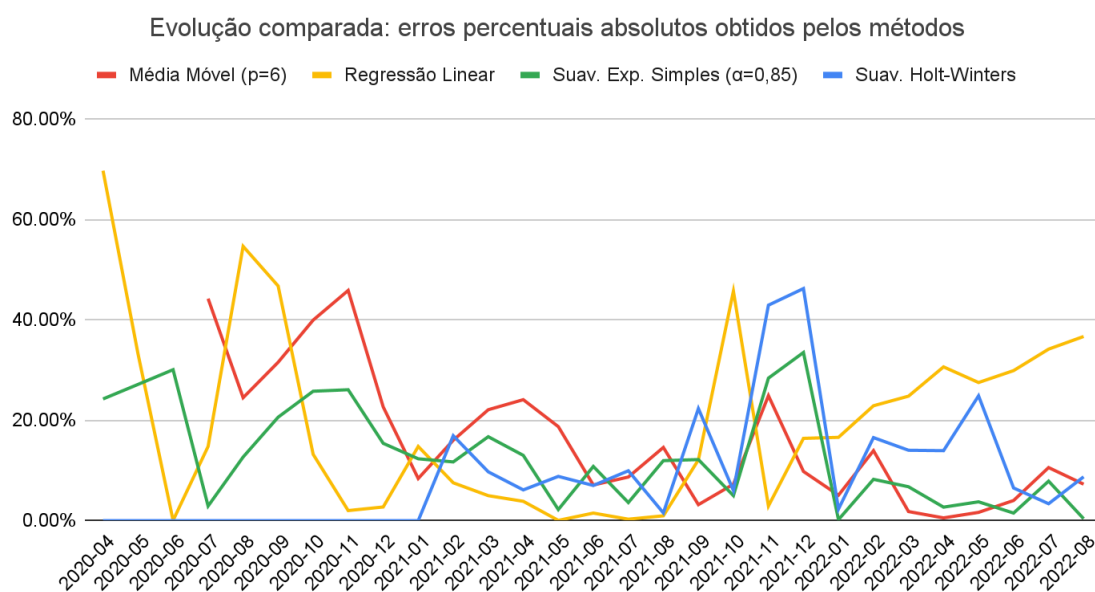
Fonte: elaborado pelo autor.

Gráfico 43: Erros percentuais obtidos pelos diferentes métodos



Fonte: elaborado pelo autor.

Gráfico 44: Erros percentuais absolutos obtidos pelos diferentes métodos



Fonte: elaborado pelo autor.

Tabela 18: Comparação de modelos: erros percentuais absolutos máximos

Ano	Média Móvel (p=6)	Regressão Linear	Suavização Exponencial Simples	Suav. Exponencial Holt-Winters
2021	24,88%	45,69%	33,48%	46,20%
2022	13,94%	36,66%	8,23%	24,85%

Fonte: elaborado pelo autor.

Tabela 19: Comparação de modelos: erros percentuais absolutos médios

Ano	Média Móvel (p=6)	Regressão Linear	Suavização Exponencial Simples	Suav. Exponencial Holt-Winters
2021	13,73%	9,25%	13,44%	16,13%
2022	5,61%	27,88%	3,92%	11,28%

Fonte: elaborado pelo autor.

Tabela 20: Comparação de modelos: erros percentuais médios

Ano	Média Móvel (p=6)	Regressão Linear	Suavização Exponencial Simples	Suav. Exponencial Holt-Winters
2021	-10,35%	5,63%	1,15%	-4,17%
2022	0,26%	-27,88%	1,34%	1,92%

Fonte: elaborado pelo autor.

Ao compilar os resultados é verifica-se que os modelos apresentam características bem distintas entre si, porém é possível elencar as seguintes similaridades: um desempenho geral pior no início das previsões em 2020, um aumento significativo em picos dos erros entre os meses próximos ao dia dos pais e black friday (agosto e novembro) e uma tendência relativamente similar de previsões antes da virada para o ano de 2022. A partir desse ano, a empresa reduziu significativamente o investimento realizado em marketing (Gráfico 40), que colaborou para uma redução na tendência de crescimento da receita mensal da

empresa. Esses fatores colaboraram para uma maior diferenciação entre o comportamento dos métodos.

O método da média móvel com  $p = 6$  foi o que obteve maior melhora da precisão no último ano da base histórica, a razão para isso foi a maior estabilidade de receita a partir de agosto de 2021. Porém no período de crescimento o modelo apresentava uma tendência de subestimar a previsão. Como a empresa pretende voltar a assumir maiores investimentos em marketing no futuro e com isso pode apresentar uma maior velocidade de crescimento, o método poderia apresentar resultados descalibrados.

O modelo construído através da regressão linear obteve maior precisão no ano de 2021, época da empresa com tendência de crescimento e maiores investimentos em marketing. Uma característica positiva do modelo foi a precisão das previsões nos períodos de aumento de venda sazonal, com destaque para os erros percentuais de 2,00% e 2,91% para os meses de novembro (black friday) do ano de 2020 e 2021 respectivamente. Apesar disso, o modelo superestimou as previsões no período de maior crescimento (2020) e subestimou os valores a partir do momento de estabilidade de receita e supressão dos investimentos em marketing (2022).

A suavização exponencial simples apresentou uma constância nos valores de erros percentuais absolutos, com uma tendência de erro decrescente ao longo do tempo observado. Os picos de erro do modelo são ocasionados por conta de períodos com expansão de vendas sazonais. Foi o método que obteve o menor erro percentual absoluto médio e seus erros percentuais alternam entre valores positivos e negativos (Gráficos 43 e 44).

O método de suavização exponencial de Holt-Winters possui um histórico de previsões reduzido, uma vez que o modelo utiliza o primeiro ano para iniciar as variáveis que irão modelar a tendência e sazonalidade. Apesar disso, as previsões não se adequaram às sazonalidades da empresa, a redução da tendência de crescimento a partir do fim do segundo semestre de 2021 também impactou os erros percentuais de forma negativa. O modelo poderia apresentar uma performance melhor se houvesse uma base histórica de dados mais ampla, já que as variáveis de tendência e sazonalidade vão sendo atualizadas conforme os períodos se repetem, permitindo uma maior adequação às tendências da empresa.

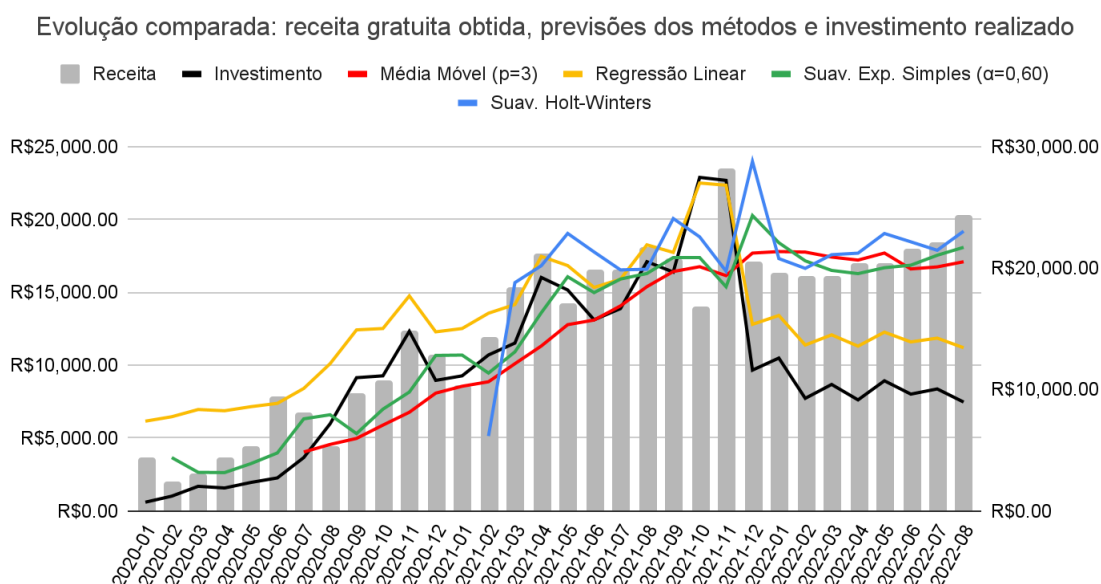
Ao analisar todos os métodos aplicados sobre a receita total, considerou-se o método da suavização exponencial simples o mais adequado para ser implementado em um modelo de previsão de receita total.

## 5.6 Avaliação dos modelos com Distinção por Origem de Receita

Essa seção busca avaliar a possibilidade de realizar a previsão da receita total através de dois métodos aplicados independentemente, um para obter valores de receita de origem de canais gratuitos e outro para a receita de canais pagos. Essa avaliação é relevante pois ao separar as receitas por origem, pequenas diferenças de tendência e sazonalidade das amostras podem influenciar a performance dos métodos de forma geral.

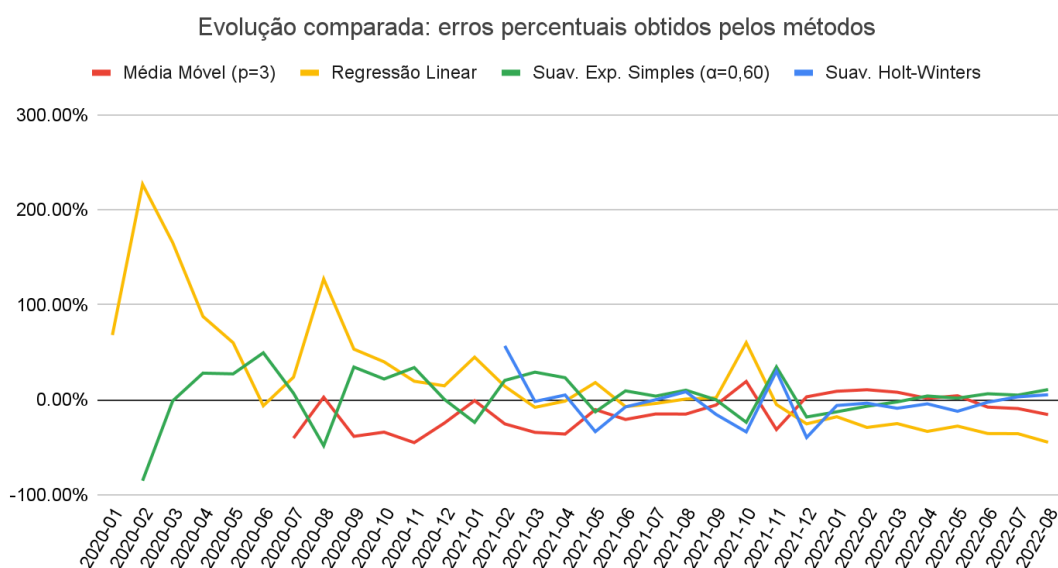
### 5.6.1 Avaliação dos modelos para Receita de Origem Gratuita

Gráfico 45: Previsões dos diferentes métodos aplicados sobre a receita de origem gratuita, receita obtida e investimento realizado em marketing



Fonte: elaborado pelo autor.

Gráfico 46: Erros percentuais obtidos pelos diferentes métodos aplicados sobre a receita de origem gratuita



Fonte: elaborado pelo autor.

Tabela 21: Comparação de modelos aplicados na receita grátis: erros percentuais absolutos máximos

Ano	Média Móvel (p=6)	Regressão Linear	Suavização Exponencial Simples	Suav. Exponencial Holt-Winters
2021	36,10%	60,14%	34,62%	56,74%
2022	15,63%	44,74%	12,68%	12,13%

Fonte: elaborado pelo autor.

Tabela 22: Comparação de modelos aplicados na receita grátis: erros percentuais absolutos médios

Ano	Média Móvel (p=6)	Regressão Linear	Suavização Exponencial Simples	Suav. Exponencial Holt-Winters
2021	18,10%	15,92%	17,42%	21,11%
2022	8,23%	31,15%	6,18%	5,75%

Fonte: elaborado pelo autor.

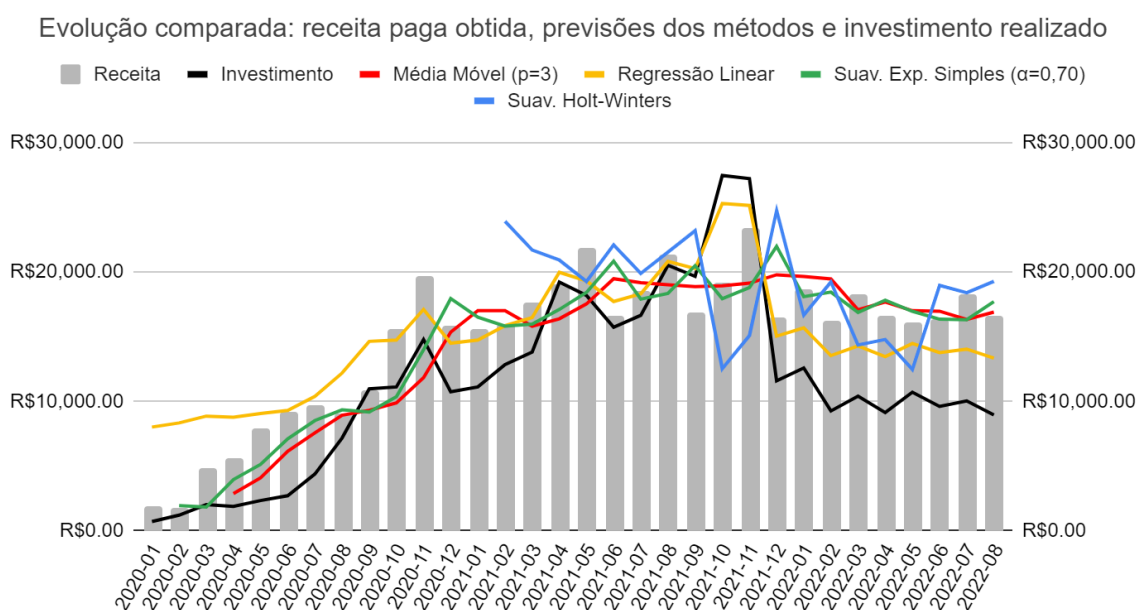
Tabela 23: Comparação de modelos aplicados na receita gratis: erros percentuais médios

Ano	Média Móvel (p=6)	Regressão Linear	Suavização Exponencial Simples	Suav. Exponencial Holt-Winters
2021	-14,37%	7,44%	4,35%	-2,87%
2022	0,07%	-31,15%	0,74%	-3,65%

Fonte: elaborado pelo autor.

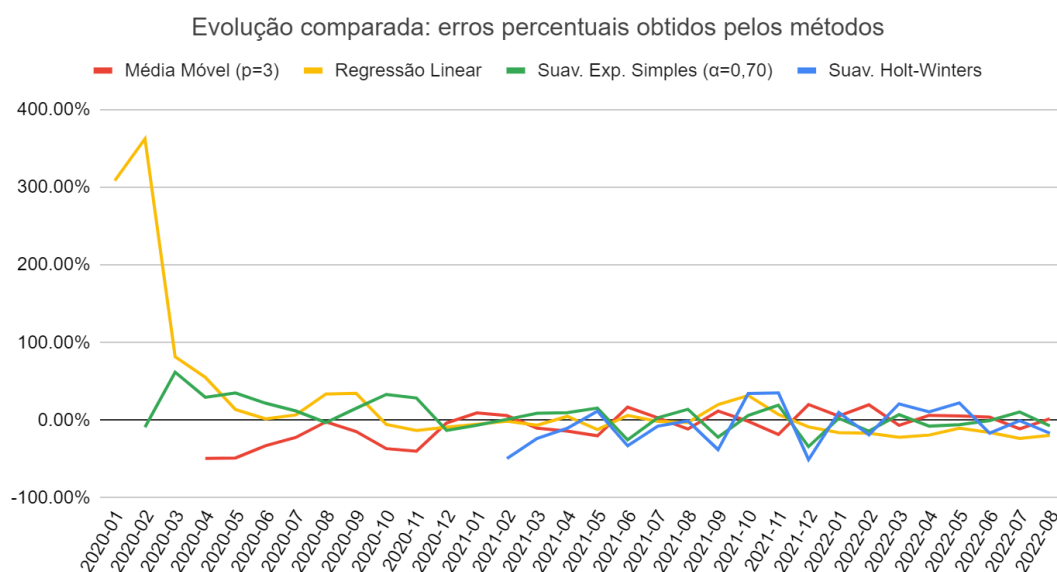
### 5.6.2 Avaliação dos modelos para Receita de Origem Paga

Gráfico 47: Previsões dos diferentes métodos aplicados sobre a receita de origem paga, receita obtida e investimento realizado em marketing



Fonte: elaborado pelo autor.

Gráfico 48: Erros percentuais obtidos pelos diferentes métodos aplicados sobre a receita de origem paga



Fonte: elaborado pelo autor.

Tabela 24: Comparação de modelos aplicados na receita paga: erros percentuais absolutos máximos

Ano	Média Móvel (p=3)	Regressão Linear	Suavização Exponencial Simples	Suav. Exponencial Holt-Winters
2021	20,44%	32,00%	33,83%	50,59%
2022	20,09%	23,28%	13,86%	22,47%

Fonte: elaborado pelo autor.

Tabela 25: Comparação de modelos aplicados na receita paga: erros percentuais absolutos médios

Ano	Média Móvel (p=6)	Regressão Linear	Suavização Exponencial Simples	Suav. Exponencial Holt-Winters
2021	11,89%	9,02%	13,90%	26,68%
2022	7,66%	17,60%	6,88%	14,62%

Fonte: elaborado pelo autor.

Tabela 26: Comparação de modelos aplicados na receita paga: erros percentuais médios

Ano	Média Móvel (p=3)	Regressão Linear	Suavização Exponencial Simples	Suav. Exponencial Holt-Winters
2021	-0,45%	2,92%	-0,59%	-11,78%
2022	3,38%	-17,60%	-1,62%	1,63%

Fonte: elaborado pelo autor.

### 5.6.3 Viabilidade dos Modelos Distintos por Origem de Receita

Ao comparar as tabelas de erros percentuais absolutos médios dos métodos aplicados à receita total, gratuita e paga, verifica-se que de forma geral os métodos aplicados de forma distinta não proporcionaram melhorias na precisão dos resultados. Com exceção à regressão linear aplicada para a previsão dos canais pagos, isso era esperado uma vez que o modelo em questão apresentou o maior coeficiente de determinação ( $R^2 = 0.696$ ) para a regressão, indicando maior proximidade entre a linha de tendência e os pontos observados. Essa maior conformidade do modelo ocorre porque os investimentos em marketing de performance, em sua maioria, são revertidos em receita de origem paga.

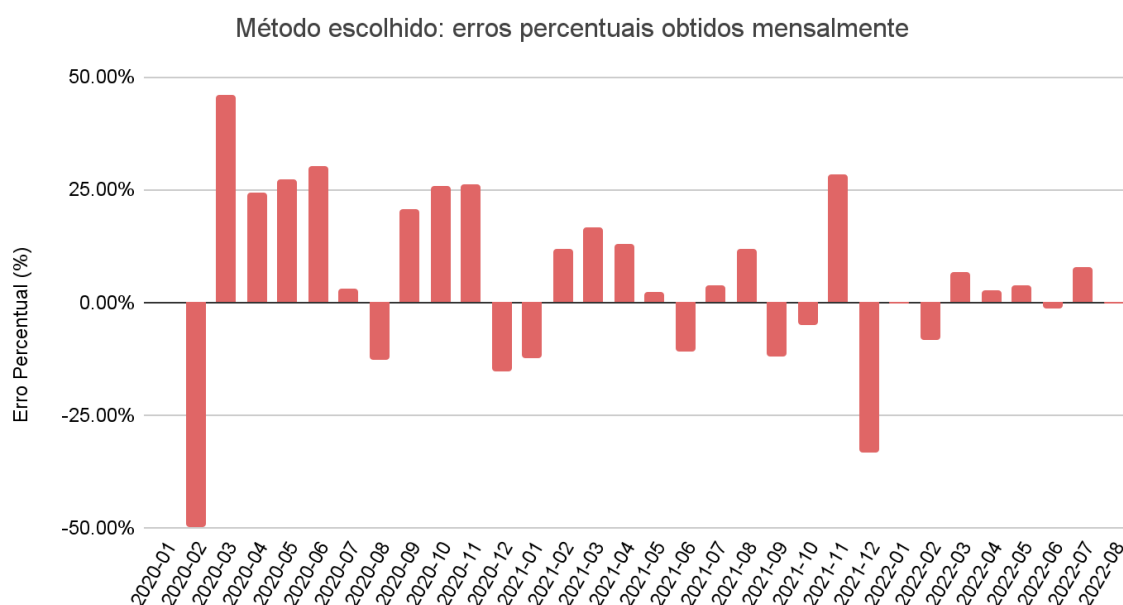
Se algum modelo aplicado aos canais de origem gratuita apresentasse boa performance, um método combinado poderia ser testado para estimar a previsão de receita total. Utilizando a regressão linear para calcular a receita de origem paga. Além disso, como exibido no Gráfico 3, as duas curvas de receita (canais grátis e pagos) apresentam valores com tendências, sazonalidades e variações similares, reduzindo a necessidade de tratar as previsões de forma independente, por isso optou-se por descartar essa possibilidade.

## 6. ESCOLHA E IMPLEMENTAÇÃO DO MODELO

### 6.1 Modelo Escolhido

Dentre os modelos testados, o que apresentou melhor conformação foi o que utiliza o método da Suavização Exponencial Simples com coeficiente de suavização  $\alpha = 0,85$  aplicado na receita total (Capítulo 5.3.1). É necessário aprofundar a análise do método, para que ajustes para melhoria do projeto possam ser levantados, além de elencar fragilidades e possíveis pontos que merecem atenção ao implementar o modelo final na empresa. Para isso, novas visualizações serão criadas, endereçando um foco maior na observação dos erros e relações entre variáveis.

Gráfico 49: Evolução dos erros percentuais obtidos mensalmente pela suavização exponencial simples



Fonte: elaborado pelo autor.

Gráfico 50: Evolução da variação de receita mês a mês



Fonte: elaborado pelo autor.

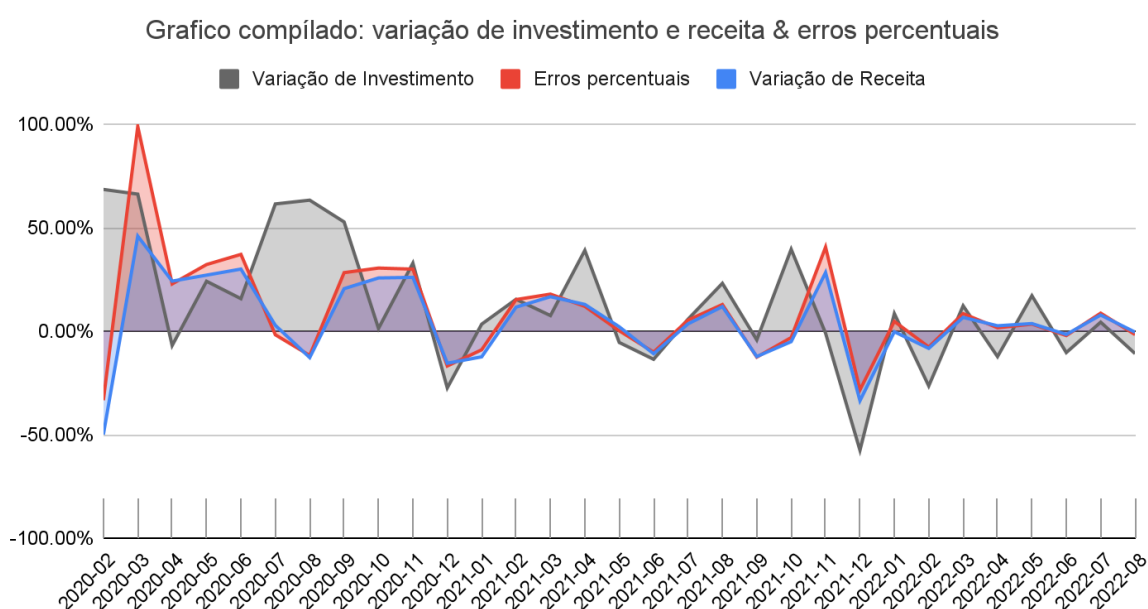
Gráfico 51: Evolução da variação de investimento em marketing mês a mês



Fonte: elaborado pelo autor.

Os maiores erros absolutos observados aconteceram no início do modelo, por conta da velocidade de crescimento de receita mensal e falta de dados prévios para o método poder basear suas estimativas. Além disso, é interessante notar as similaridades entre os 3 gráficos colocados anteriormente, indicando que a variação de receita (Gráfico 50) e a variação de investimento em marketing (Gráfico 51) impactam a precisão da previsão do modelo (Gráfico 49). Segue um compilado a fim de facilitar a comparação:

Gráfico 52: Variação de receita, investimento e erros percentuais mês a mês



Fonte: elaborado pelo autor.

O modelo se mostrou suficientemente apurado para realizar previsões de receita a partir da diminuição do ritmo de crescimento observado a partir de 2021. Apesar disso, será preciso realizar ajustes para diminuir os picos de erros. Caso a manutenção seja bem sucedida, será possível reduzir a magnitude do erro percentual absoluto médio, além de calibrar as expectativas da empresa para períodos de vendas mais importantes para a organização.

## 6.2 Ajuste no modelo

O foco da manutenção a ser realizada é reduzir os picos de erros, para que não sejam feitas muitas alterações no modelo original a atenção será direcionada em corrigir valores para meses anteriores e posteriores à Black Friday, incluindo de Outubro à Dezembro. Isso poderia ser feito através de uma atribuição de constantes ou variáveis nesses meses de forma subjetiva embasada em expectativas com base no momento atual da empresa. Porém o método da regressão linear foi eficaz em estimar meses com expansão de investimento em marketing (Gráfico 42), e portanto, será utilizado numa tentativa de calibrar o modelo construído através da suavização exponencial.

Os ajustes deverão ser mínimos, para que a performance e credibilidade do modelo não sejam impactados negativamente. A ideia é incorporar a regressão de forma sutil para reduzir o erro de forma tênue, e não zerá-lo através de valores convenientemente selecionados.

### 6.2.1 Incorporação da Regressão Linear

Agora que a empresa se encontra em um patamar de maior solidez com uma base mais ampla de clientes fidelizados, se reduziu a necessidade de investir agressivamente em marketing como nos períodos iniciais de expansão da Startup. Além disso, a fim de evitar uma saturação das audiências de plataformas de mídia e da base de clientes, a empresa passou a investir de forma mais intensa apenas em meses promocionais. Por isso, foi definido que os meses a serem corrigidos pela regressão linear serão novembro (Black Friday) e dezembro (Rebote Black Friday).

A manutenção será realizada através de um fator multiplicador de correção (fc) que levará em conta a diferença entre as previsões realizadas pelo método da suavização exponencial simples e da regressão linear para os meses indicados e se houve uma queda ou aumento brusco do valor a ser investido, e assumirá valor de 1 para os outros períodos.

Segue a formulação da adaptação nos meses a serem ajustados:

$$fc = 1 + \frac{(P_{RL} - P_{SES})}{P_{SES}}$$

$$P_2 = fc \cdot P_{SES}$$

$fc$ : Fator de correção

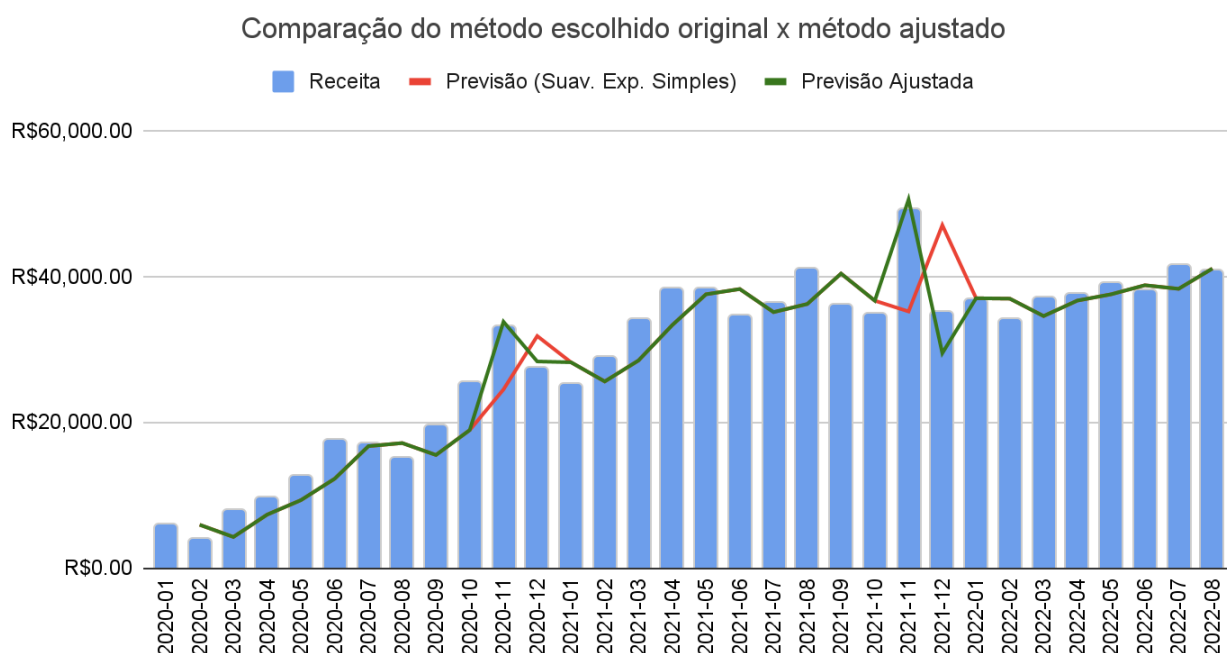
$P_{RL}$ : Previsão da regressão linear

$P_{SES}$ : Previsão inicial (suavização exponencial simples)

$P_2$ : Previsão Ajustada

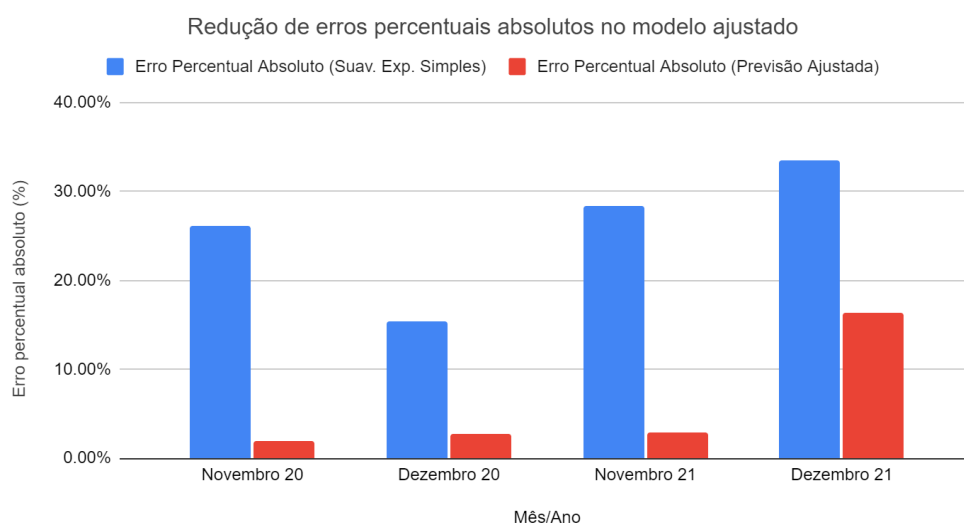
O modelo ajustado obteve os seguintes resultados:

Gráfico 53: Variação de receita, previsões do modelo escolhido e previsões do modelo ajustado



Fonte: elaborado pelo autor.

Gráfico 54: Comparação de erros: modelo escolhido x modelo ajustado



Fonte: elaborado pelo autor.

Tabela 27: Comparação de erros: modelo de suavização exponencial simples e modelo modelos de suavização exponencial simples ajustado com regressão linear

Ano	Erro Percentual Absoluto Médio		Erro Percentual Médio	
	Modelo Suavização Exp. Simples	Modelo Ajustado	Modelo Suavização Exp. Simples	Modelo Ajustado
2020	25,51%	22,17%	11,35%	9,94%
2021	13,44%	9,89%	1,15%	2,69%
2022	3,92%	3,92%	1,34%	1,34%

Fonte: elaborado pelo autor.

O ajuste através da implementação da regressão linear para gerar um fator de correção na previsão do modelo escolhido reduziu de forma significativa o erro observado nos meses de Novembro e Dezembro. Agora o comportamento do modelo está mais próximo dos resultados reais, e apresenta uma boa precisão para o mês da Black Friday. Isso é especialmente positivo pois novembro gera uma certa incerteza da empresa com relação aos resultados, costuma ser o período mais forte do ano e com expectativas altas por parte dos colaboradores.

Vale lembrar que esse novo modelo deverá receber como entrada manual o orçamento para o marketing em performance para o mês seguinte para que os

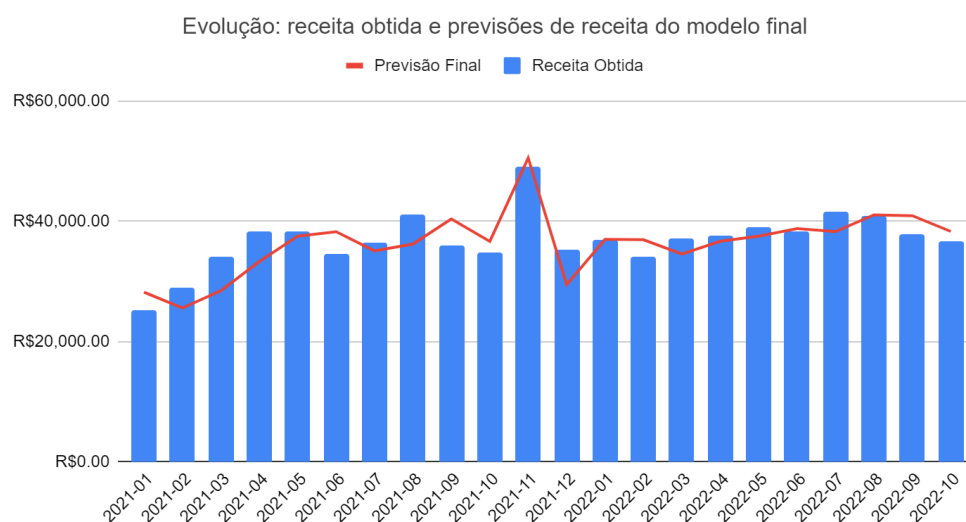
ajustes possam ser executados através da regressão linear entre a receita e investimento. Com isso, o modelo será construído em arquivo à parte para que possa ser incorporado à empresa.

### 6.3 Implementação do Modelo na Empresa

A implementação na empresa se dará a partir do seguinte fluxo: a área de dados deverá fornecer as entradas para o modelo rodar a previsão para o mês seguinte, os resultados serão levados para a equipe de Growth e para as lideranças da empresa, que utilizarão os valores para determinar a meta de vendas do mês seguinte. Espera-se que o modelo seja preciso suficiente a ponto de alinhar as expectativas das lideranças de acordo com o momento atual da empresa, tratando os cenários de forma mais realista e embasada em dados.

A parte técnica necessária para implementação já foi detalhada nos capítulos 4.2.2 e 4.2.3. Como um teste final, iremos simular a sua implementação, utilizando os valores dos meses de Setembro e Outubro do ano de 2022. Isso será feito pois as bases utilizadas na construção dos testes entre modelos só continham dados até o fim de Agosto de 2022.

Gráfico 55: Modelo final aplicado nos meses de Setembro e Outubro de 2022



Fonte: elaborado pelo autor.

Tabela 28: Erros obtidos pelo modelo final aplicado nos meses de Setembro e Outubro de 2022

Mês	Erro percentual	Erro percentual absoluto
Setembro 2022	8,13%	8,13%
Outubro 22	4,56%	4,56%

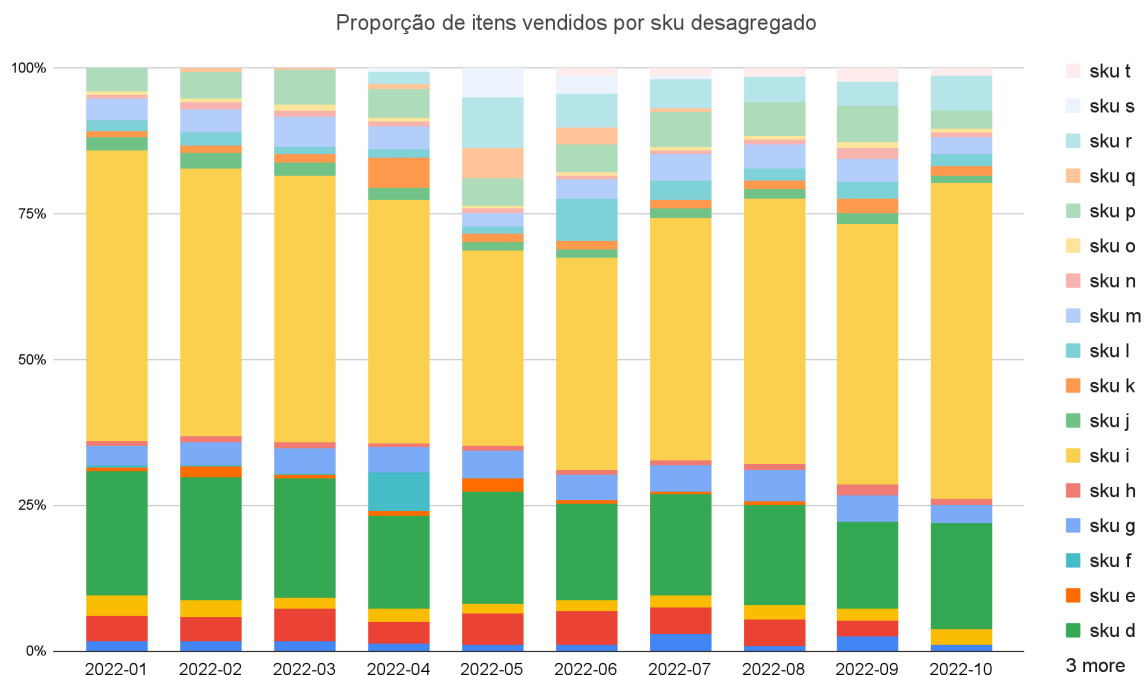
Fonte: elaborado pelo autor.

As previsões para esses dois meses tiveram um bom desempenho, com erros positivos proporcionados pelo pequeno encolhimento do volume de vendas nos períodos. Pode-se esperar que o método apresente uma piora na precisão em meses em que se inverta a tendência dos valores de receita. Com essa validação final, o modelo está pronto para ser utilizado na rotina da organização. Se os erros permanecerem em torno dos valores observados neste relatório, no processo de estabelecimento de metas de vendas pelas equipes responsáveis é importante definir valores próximos às previsões levantadas. Isso para que as diferenças entre meta e receita não se aproximem de valores observados nos anos anteriores (Gráfico 1).

#### 6.4 Previsão de Vendas de Produtos Através da Previsão da Receita

Como constatado no gráfico 5, a distribuição de receita por categoria de produto se mostrou bem estável ao longo do tempo. Para aprofundar a análise, agora será avaliada a distribuição de receita por skus desagregadas. Skus desagregadas podem ser entendidas como os produtos base da empresa, que podem compor outros skus (como kits e combos) ou serem comprados individualmente. Ou seja, são os produtos oferecidos diretamente pelos fornecedores da organização. Segue a distribuição:

Gráfico 56: Proporção de itens vendidos por sku desagregada



Fonte: elaborado pelo autor.

No gráfico acima percebe-se que, apesar da estabilidade por categoria, a proporção dos itens vendidos por sku desagregado mensalmente não apresentam a mesma consistência. Como o modelo preditivo aplicado na receita total já apresenta flutuações de precisão por si só, a aplicação do método para prever a saída de produtos não será aplicada pois apresentaria inconsistências ainda maiores. Como principais causas para essa flutuação podemos citar: o lançamento e descontinuação de kits e combos com composições de produtos diferentes e promoções aplicadas a produtos distintos ao longo do tempo.



## 7. CONCLUSÕES

### 7.1 Aprendizados e Objetivos Alcançados

O trabalho surgiu para auxiliar o estabelecimento de metas de receita da Dr Jones, fornecendo um embasamento através da análise dos dados com a aplicação de um modelo preditivo. Como parte do trabalho consistia em testar métodos diferentes para a obtenção do modelo mais adequado, foi interessante levantar os pontos positivos e negativos de cada modelo. Outro ponto de aprendizado interessante é o fato dos modelos testados terem apresentado maior dificuldade, de forma geral, em prever valores de receita para períodos de grande crescimento de receita mensal, especialmente se a base histórica for limitada.

O modelo final foi construído através da suavização exponencial simples (Capítulo 6.1), com o método de regressão linear sendo utilizado para ajustar valores em períodos de venda expandida e com grande variação de receita (Capítulo 6.2.1). A partir de 2021, o modelo apresentou um erro percentual absoluto médio de 7,40% e erro percentual médio máximo de 16,71%.

Como o projeto foi realizado para ser aplicado dentro da área de data-science, a implementação e manutenção do modelo se torna simples. Através das ferramentas e plataformas apresentadas anteriormente. Esse projeto aumentará o valor agregado pela área de dados à empresa que almeja, cada vez mais, tomar suas decisões baseadas em dados e análises a fim de diminuir as incertezas e reduzir o impacto que opiniões e pontos de vistas incorretos podem implicar sobre as estratégias da empresa.

É importante destacar que o modelo realiza suas previsões utilizando apenas dados históricos de receita total e investimento em marketing previsto para o mês sendo estimado (para meses que passam por ajuste). Existem diversos outros fatores que podem impactar as vendas e o crescimento da empresa, que podem influenciar com maior ou menor intensidade os resultados. Por isso é importante que acompanhamentos sejam estabelecidos para que se possa ir avaliando a performance do método ao longo do tempo e, talvez, encontrar potenciais novos ajustes a serem implementados ou até mesmo a reformulação completa do modelo em busca de um mais adequado para o momento presente da organização.

Por fim, pode-se afirmar que os resultados foram bem sucedidos. O modelo foi construído a partir de métodos relativamente simples, e atingiu o objetivo de fornecer uma ideia do que esperar da receita para o mês seguinte. Para a aplicação do projeto não será necessário implementação de nenhuma ferramenta paga e, portanto, não acarretará em nenhum custo a mais. Com desempenho suficiente em dois momentos distintos da empresa, de leve crescimento e estabilização nos anos de 2021 e 2022 respectivamente (Gráfico 53). Portanto, espera-se que o modelo proporcione uma melhora no estabelecimento das metas, de forma a não gerar uma quebra de expectativa entre os colaboradores ao não atingi-las, além de facilitar a organização interna da Dr Jones.

## 7.2 Acompanhamentos e Possibilidades de Mudança

Nos próximos meses será necessário um olhar vigilante para com o modelo, para que a progressão dos erros percentuais e erros percentuais absolutos sejam acompanhados. Além de verificar se o fator de correção proposto para incorporar a regressão linear no modelo está sendo eficaz em reduzir os erros de Novembro e Dezembro. No futuro será preciso montar uma nova regressão linear entre a receita total obtida e investimento realizado, para atualizar, com as novas entradas, a equação que define a linha de tendência e relação entre as duas variáveis.

Outro desdobramento é a verificação de como o modelo está sendo utilizado pelos gestores para definir as metas, se os dados fornecidos estão realmente sendo considerados. Além de observar se as metas passam a ser mais precisas após a implementação plena do modelo na empresa.

Ao longo da execução do presente estudo o Grupo Boticário adquiriu a Dr Jones, a fim de ampliar o portfólio ao público masculino, além de reforçar sua presença dentro do ambiente de comércio digital. Essa aquisição pode transformar a evolução da Dr Jones de forma ainda desconhecida, potencializando maiores investimentos, uma estrutura logística robusta, produção de novos produtos e através dos diversos canais de venda que o Grupo possui. A estrutura da Dr Jones será preservada, bem como seu modelo de negócios, com a direção sendo mantida nas mãos dos fundadores da empresa.

Dependendo do alavanque proporcionado por essa aquisição, o modelo proposto pode apresentar uma variação de performance. Caso o método escolhido não seja capaz de atender o novo cenário, uma nova rodada de testes deverá ser realizada com o intuito de averiguar a viabilidade de manutenção ou necessidade de troca de método preditivo para se adequar ao novo momento da marca.



## 8. REFERÊNCIAS BIBLIOGRÁFICAS

ALBRIGHT, S., WAYNE, L., Christopher, J., Mark B. Data analysis and decision making. Vol. 577. South-Western/Cengage Learning, 2011.

ARNOLD, J. R. T.; CHAPMAN, S. N.; CLIVE, L. M. Introduction to Materials Management. 6. ed. New Jersey: Pearson Prentice Hall, 2011

ARMSTRONG, J. S. Principles of Forecasting: A Handbook for Researchers and Practitioners. Springer Illustrated, 2001.

ARONNE, A.; BRESSAN, A.; BRASIL, H. Mensuração e gerenciamento de riscos corporativos. Saint Paul Editora, 29 de abril de 2021.

BONFIM, B. Boticário compra Dr. Jones e amplia portfolio para atender o público masculino. Exame, 7 de outubro de 2022. Disponível em: <https://exame.com/negocios/boticario-compra-dr-jones-e-amplia-portfolio-para-atender-o-publico-masculino/>. Acesso em: 06/11/2022.

CHATFIELD, C., YAR, M. (1988), Holt-Winters Forecasting: Some Practical Issues. Journal of the Royal Statistical Society: Series D (The Statistician), 37: 129-140. <https://doi.org/10.2307/2348687>

CARBINATTO, B. Beleza masculina: um mercado em alta. 8 de abril de 2022. Abril Você S/A. Disponível em: <https://vocesa.abril.com.br/empreendedorismo/beleza-masculina-um-mercado-em-alta/>. Acesso em: 11/10/2022.

FEDIRKO, O., ZATONATSKA, T., WOLOWIEC, T., & SKOWRON, S. (2021). Data science and marketing in E-commerce amid COVID-19 pandemic. European Research Studies, 24, 3-16.

LARA, J., AFONSO, T., ELIAS, R. K., AFONSO, B. P. D., AUGUSTO, T. T. (2020). Marketing e logística: em busca da satisfação e lealdade do consumidor de compras on-line. Revista Horizontes Interdisciplinares da Gestão, Setembro 2020.

MAJOR, Leslie. How to perform Holt Winter's Method on Monthly Data (in Excel). Youtube, 9 de maio de 2020. Disponível em: [https://youtu.be/4\\_ciGzvrQl8](https://youtu.be/4_ciGzvrQl8). Acesso em: 16/10/2022.

MAKRIDAKIS, S.; WHEELWRIGHT, S. C.; MCGEE, V. E. Forecasting: Methods and Applications. 3. ed. New York: John Wiley and Sons, 1998.

MAKRIDAKIS, S.; CARBONE, R. Forecasting when pattern changes occur beyond the historical data. INSEAD, Abril 1985.

MONTGOMERY, D., JOHNSON, L. A., GARDINER, J. Forecasting and Time Series Analysis. 2 ed. New York: McGraw-Hill, 1990.

MONTGOMERY, D., PECK, E., VINING, G. Introduction to linear regression analysis. John Wiley & Sons, 2021.

OSTERTAGOVA, E., OSTERTAG, O. (2011). The Simple Exponential Smoothing Model.

PELLEGRINI, F. R.; FOGLIATTO, F. S. Passos para Implantação de Sistemas de Previsão de Demanda - Técnicas e Estudo de Caso. Revista PRODUÇÃO, v. 11, n. 1, 2001.

SANTORO, M. C. Previsão de Demanda. Fevereiro de 2013. Apresentação das aulas 1, 2, 3 e 4 de PRO 3445 (Planejamento e Controle da Produção) da Escola Politécnica da Universidade de São Paulo. Acesso: 08/09/2022.

Sem autor. Como Trabalhar com Planejamento Preditivo no Smart View - Aditivo de Holt-Winters. Oracle Documentation. Sem data de publicação. Pág. 84. Disponível em:

[https://docs.oracle.com/cloud/help/pt\\_BR/pbcs\\_common/CSPPU/holt-winters\\_additive.htm](https://docs.oracle.com/cloud/help/pt_BR/pbcs_common/CSPPU/holt-winters_additive.htm). Acesso em: 12/09/2022.

Sem autor. Como Trabalhar com o Planning - MAPE (Erro Percentual Absoluto Médio). Oracle Docs. Sem data de publicação. Pág. 252. Disponível em: [https://docs.oracle.com/cloud/help/pt\\_BR/pbcs\\_common/PFUSU/insights\\_metrics\\_MAPE.htm#PFUSU-GUID-C33B0F01-83E9-468B-B96C-413A12882334](https://docs.oracle.com/cloud/help/pt_BR/pbcs_common/PFUSU/insights_metrics_MAPE.htm#PFUSU-GUID-C33B0F01-83E9-468B-B96C-413A12882334). Acesso em: 08/09/2022.

Sem autor. Erro de quadrado médio. IBM Documentation. 14 de abril de 2021. Disponível em: [https://docs.oracle.com/cloud/help/pt\\_BR/pbcs\\_common/CSPPU/holt-winters\\_additive.htm](https://docs.oracle.com/cloud/help/pt_BR/pbcs_common/CSPPU/holt-winters_additive.htm). Acesso em: 12/09/2022.

Sem autor. Introdução ao SQL no BigQuery. Google Cloud Documentation. Sem data de publicação. Disponível em: <https://cloud.google.com/bigquery/docs/reference/standard-sql/introduction#:~:text=BigQuery%20supports%20the%20Google%20Standard,supported%20using%20Google%20Standard%20SQL>. Acesso em: 12/09/2022.

SILVA, W. M.; MORAIS, L. A.; FRADE, C. M.; PESSOA, M. F. Digital marketing, E-commerce and pandemia: a bibliographic review on the brazilian panorama. *Research, Society and Development*, [S. l.], v. 10, n. 5, p. e45210515054, 2021.

Disponível em: <https://rsdjournal.org/index.php/rsd/article/view/15054>. Acesso em: 29 sep. 2022.