

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO

**Análise de impacto do uso de campanhas de
marketing em televisão nas vendas de
supermercado**

Daniel Cunha Rios

PROJETO FINAL DE GRADUAÇÃO

CENTRO TÉCNICO CIENTÍFICO - CTC

DEPARTAMENTO DE INFORMÁTICA

Curso de Graduação em Ciência da Computação

Rio de Janeiro, novembro de 2021



Daniel Cunha Rios

**Análise de impacto do uso de campanhas de marketing em
televisão nas vendas de supermercados**

Relatório de Projeto Final, apresentado ao programa **Ciência da Computação** da PUC-Rio como requisito parcial para a obtenção do título de Bacharel em Ciência da Computação.

Orientador: Hélio Côrtes Vieira Lopes

Rio de Janeiro
novembro de 2021.

Agradecimentos

À minha família por todo o apoio nos anos de graduação.

À minha mãe pelo apoio nos momentos mais complicados na graduação.

A Sergio Antunes pelo apoio no entendimento do negócio para a realização deste projeto.

Resumo

Cunha Rios, Daniel. Lopes, Hélio. Análise de impacto do uso de campanhas de marketing em televisão nas vendas de supermercado. Rio de Janeiro, 2021. Número de páginas 50p. Relatório de Projeto Final – Departamento de Informática. Pontifícia Universidade Católica do Rio de Janeiro.

Este trabalho tem por objetivo a elaboração de uma ferramenta para o estudo do impacto nas vendas de produtos expostos em mídia de televisão aberta em uma rede de supermercados. O estudo dimensiona as variações de volumes conforme as variações de preços aplicados. Os resultados desse estudo permitem avaliar se a escolha do produto é apropriada para alcançar os objetivos, bem como auxilia na previsão volumétrica para aquisição dos produtos junto a fornecedores e distribuição dos mesmos para suas unidades de venda.

Palavras-chave

Comportamento. Demanda. Histórico. Supermercado. Regressão. Dados.

Abstract

Cunha Rios, Daniel. Lopes, Hélio. Analysis of the impact of the use of television marketing campaigns on supermarket sales. Rio de Janeiro, 2021. Número de páginas 50p. Relatório de Projeto Final – Departamento de Informática. Pontifícia Universidade Católica do Rio de Janeiro.

This work aims to develop a tool to study the impact on sales of products displayed on open television media in a supermarket chain. The study scales the volume variations according to the applied price variations. The results of this study make it possible to assess whether the choice of product is appropriate to achieve the objectives, as well as assisting in the volumetric forecast for the acquisition of products from suppliers and their distribution to their sales units.

Keywords

Behavior. Demand. Historical. Supermarket. Regression. Data.

Sumário

1	INTRODUÇÃO	6
2	SITUAÇÃO ATUAL	7
3	OBJETIVOS.....	12
4	ATIVIDADES REALIZADAS.....	14
4.1	ESTUDOS SOBRE O CONTEXTO DO NEGÓCIO.....	14
4.2	ESTUDO DAS TECNOLOGIAS UTILIZADAS.....	15
4.3	ESTUDO DA BASE DE DADOS.....	17
4.4	ESTUDO DA CORRELAÇÃO ENTRE A DEMANDA E O PREÇO	18
4.5	ESTUDO DOS DADOS UTILIZADOS NO MODELO	19
5	PROJETO E ESPECIFICAÇÃO	21
5.1	DESAFIOS A ENFRENTAR	21
5.2	ARQUITETURA.....	22
5.3	SOLUÇÃO	27
5.4	PROPÓSITOS: NOVAS VISÕES.....	29
5.5	USO DA FERRAMENTA.....	30
6	IMPLEMENTAÇÃO E AVALIAÇÃO.....	31
6.1	DESAFIOS ENFRENTADOS.....	31
6.2	AVALIAÇÃO DO MODELO	33
6.3	AVALIAÇÃO PARA O NEGÓCIO	35
7	CONSIDERAÇÕES FINAIS	38
8	REFERÊNCIAS	40
	APÊNDICE A – DEMONSTRAÇÕES VISUAIS DE GRÁFICOS GERADOS.....	41
	APÊNDICE B – PROCESSO DE PREPARAÇÃO DOS DADOS	49

1 Introdução

A motivação do projeto veio devido à importância que mídia da televisão aberta possui para o varejo alimentar hoje. Por ainda ser um veículo de mídia de massa de maior importância, é um dos pilares estratégicos para atrair clientes para o negócio. O objetivo é entender qual o movimento na demanda por um produto que sua oferta na televisão pode proporcionar através de data science [1].

Trata-se de um assunto muito importante e que demanda grande esforço e investimento por parte do negócio. O varejo alimentar é um setor de alta competitividade que baseia sua principal estratégia de atratividade de público no preço, o que torna um setor com baixa lucratividade, chegando em média a 2,4% na sua média mais alta [2].

É de extrema importância ter uma boa estratégia para escolher um produto que seja atraente para ser mais acertado e eficiente na hora de anunciar na televisão. É fundamental também estimar a redução de preço que atraia um grande volume de clientes e ao mesmo tempo seja uma redução viável financeiramente para a empresa e que não comprometa a sua saúde financeira.

Em conjunto a esses objetivos, é importante também identificar a redução de preço que atinge o crescimento de vendas desejado e prever também o crescimento volumétrico de vendas do produto para planejar as atividades de compras e distribuição do produto.

Mesmo sendo um grande setor na sociedade, representando 7,5% do PIB brasileiro em 2020 [2], ainda é um setor onde o uso massivo de dados para geração de inteligência de negócio não seja amplamente difundido pelas redes, e conseqüentemente muitos dados gerados são sequer tratados. Alguns grandes grupos possuem um alto investimento em tecnologia e inteligência de dados, porém não reflete a realidade geral das empresas que estão espalhadas pelo Brasil.

Visando possibilitar escolhas mais precisas de produtos que possam cumprir melhor o objetivo de ser promovido na televisão, e assim podendo também evitar desperdício de investimento na divulgação de produtos que não são interessantes para o tipo de mídia específico, foi construído um ferramental para dar suporte a tais decisões. A ferramenta objetiva que as decisões possam ser tomadas com uma visão mais ampla do cenário e com um suporte estatístico das informações lá presentes. Tendo sido perseguida a busca de trazer sempre entendimento e elementos do negócio para agregar à ferramenta.

2 Situação atual

O setor varejista alimentar, de maneira geral no Brasil, não realiza investimentos significativos em tecnologia de dados, conseqüentemente não tem uma grande maturidade com dados, o que pode acabar diferenciando a operação da rede que levar isso a seu favor. Esse projeto nasceu dentro do estágio em uma empresa que hoje está entre as 50 maiores redes do ranking nacional de supermercados [2], com 22 lojas e 1 centro de distribuição no Rio de Janeiro.

O projeto que nasce nesse contexto visa entender melhor o comportamento e o desempenho de produtos na televisão, pois estes, na média, representam de 10 a 15% do faturamento [3], tornando assim parcela de produtos relevantes que devem ser estudados com atenção para que tenham bom desempenho. Apesar de não só, o peso no faturamento que possui um anúncio de televisão hoje, bem como sua importância estratégica de gerar fluxo de clientes, as escolhas e estudos ao redor do produto ainda são muito manuais. Apesar de já existirem dados estruturados e dados históricos de vendas e de momentos de promoção na televisão, não existe uma métrica ou uma forma de avaliar historicamente o desempenho daquele produto na televisão. A não ser que os dados sejam anotados manualmente, e estudados com cautela, buscando criar relações, o que se torna inviável quando mais de 30 produtos são avaliados de forma conjunta para decidir os produtos para a televisão de uma semana [4].

Hoje as análises são feitas baseadas nos dados que são disponibilizados em uma ferramenta de inteligência de negócios, porém as informações pertencem a diferentes relatórios, o que acaba dificultando o estudo. É possível analisar o preço e a quantidade vendida em uma data, porém a data e a mídia em que foram ofertados está presente em outro relatório. Sendo assim, fica complexo comparar cenários e possibilidades de variação de preços. Então, na maioria dos casos, essas avaliações são comparativas com um histórico mais recente do produto, reduzindo substancialmente a base de dados usada para a tomada de decisão.

São escolhidos os produtos que são considerados mais atrativos para o público geral e os que são melhores vendidos, e com base nisso, caso tenha um histórico recente de anúncio desses produtos na televisão, a venda realizada será usada como base para estimar o volume a ser vendido. Essas análises são eficientes até certo ponto, porém com a inviabilidade de avaliar tamanho volume de dados que foi sugerido acima como ideal e de forma simultânea, as análises ficam mais restritas somente a algumas realidades mais recentes.

Para o entendimento total do projeto, se faz necessário o entendimento de algumas funções operacionais do negócio. Sendo uma delas o papel do comprador, que é o profissional que negocia com o fornecedor quais produtos que serão comprados, suas respectivas quantidades, o preço, o momento da entrega e as condições comerciais de venda, e avalia a competitividade do produto em relação a concorrência. Uma vez que está por dentro das condições comerciais, também figura papel importante na sugestão dos produtos a serem promovidos na televisão. É importante diferenciá-lo do cliente, que é a pessoa que vai a loja fazer suas compras, o consumidor final. No texto será tratado sempre como cliente.

Para que o comprador possa realizar uma boa negociação, é necessário estimar a demanda por determinado produto para que não ultrapasse a capacidade de venda e fique com o estoque muito cheio e nem falte produto para ser vendido.

É importante também explicar alguns termos como a demanda, que é a representação de quantas unidades, ou volume de um determinado produto que será vendido. Para estimá-la então, faremos uso da saída média, que é uma métrica que informa a quantidade média vendida por dia de um determinado produto em um período de 10 semanas, sem considerar as vendas do dia que o produto foi anunciado na televisão, e sem considerar também os dias que eventualmente o produto ficou sem estoque.

Em conjunto com a saída média, é utilizado o conceito de elasticidade que representa o fator multiplicativo pelo qual a saída média é multiplicada em dia que determinado produto foi anunciado na televisão. Um produto que tem saída média de 10 unidades, no dia que foi anunciado na televisão, foram vendidas 50 unidades, naquele dia ele registrou uma elasticidade de 5.

Outro fator que anda em conjunto com a demanda é o preço, e para entender melhor essa variável, será utilizado o conceito de preço médio, que representa a média do preço de venda de um determinado produto nos últimos 30 dias sem considerar o preço de quando ele foi anunciado na televisão. Esse preço médio será utilizado para calcular a diferença percentual do preço médio, que é a diferença entre o preço no dia em que foi anunciado na televisão e seu preço médio. Sendo essa diferença retratada de forma percentual, para evitar falsas impressões que serão explicadas mais à frente.

Além desses conceitos, é importante ressaltar também que existe uma política de estoque, onde todos os produtos devem seguir o padrão de estocagem

de em média no máximo 15 dias no centro de distribuição mais 15 dias nas lojas, dependendo, claro, da perecibilidade do produto.

O varejo alimentar possui uma dinâmica bastante peculiar no seu funcionamento, onde as vendas são fortemente influenciadas por uma série de fatores bem definidos e outros fatores mais difíceis de serem mensurados ou calculados. Uma delas são as datas, onde as vendas do início do mês são bastante diferentes das vendas do final do mês. Dinâmica essa que funciona até mesmo na semana, o início de semana, na segunda, terça e quarta-feira é consideravelmente mais fraco quando comparado ao de quinta, sexta-feira e sábado. Além disso, que são fatores corriqueiros no dia a dia, surge também a influência de datas como feriados e o dia que cai o feriado e a influência de momentos do ano como natal, dia das mães etc. Esses momentos são conhecidos como sazonalidades que acontecem durante o ano e não necessariamente são as mesmas de um ano para o outro, como o carnaval ou páscoa, que são datas móveis.

Além das sazonalidades que são fatores que influenciam na demanda, mas são mais previsíveis e entendíveis, existem outros que não possíveis de medir ou antecipar ou que não estão sobre controle da empresa e podem afetar consideravelmente o negócio. Nessa lista temos como mais expressivo a concorrência, que pode representar a imagem do concorrente geral no mercado, um competidor claro e direto, assim como pode ser um concorrente local, que de loja para loja pode ser muito diferente. Seja um comercio informal próximo a uma loja, uma loja de bairro que compete somente naquela loja, um novo concorrente inaugurando uma loja, enfim, existem muitas situações que podem influenciar nessa direção.

Ainda na concorrência, o concorrente mais geral, que disputa clientes de maneira geral pode adotar medidas que venha também a impactar fortemente na demanda por produtos, independente do preço. Pode ser que o concorrente tenha uma estratégia agressiva nos preços um determinado conjunto de produtos que carregue grande parte dos clientes. Ou que os anúncios na mídia em geral, incluindo da televisão sejam mais agressivos, estando mais presente mais vezes para o consumidor, fazendo com que o consumidor opte pelo concorrente e não pela empresa.

O cenário econômico também é outro grande fator de grande influência no cenário, em casos de inflação por exemplo, independente da porcentagem de diminuição do preço para ser ofertado, pode ser que a elasticidade não aumente tanto, uma vez que o consumo se adapta a nova realidade de inflação, optando

assim por outros produtos, os chamados produtos substitutos. Com o aumento do preço da carne, parte do consumo da carne migra para o frango, caso o frango também aumente o preço, a tendência é a migração para linguiças e outros cortes, podendo chegar até no ovo. O mesmo que ocorre na deflação, quando os hábitos de consumo gerais passam a se adaptar de acordo com uma nova realidade.

Ainda nos fatores externos, temos as condições climáticas, que podem afetar o deslocamento dos consumidores para os supermercados, diminuindo o fluxo em dias de chuva e em finais de semana de muito sol, onde preferem ir para a praia. Além do impacto direto nas vendas, as condições climáticas podem impactar ainda de forma indireta, uma vez que parte dos produtos produzidos dependem das condições climáticas para que a produção esperada seja plenamente atendida. Podendo assim, em casos de desregulação climática gerar aumentos consideráveis no preço desses produtos pela falta deles e afetando diretamente a demanda por eles. Além disso, as estações do ano, em função da temperatura, também influenciam o volume de vendas de alguns produtos como sorvetes, sucos, cervejas, refrigerantes e água durante o verão e vinhos, salgados e chocolates durante o inverno.

Visto todas essas variáveis que não são as únicas, temos ainda também a força da mídia de televisão que foi comprada, uma vez que não é um pacote único, é importante ressaltar que cada anúncio passado pode ter tido um impacto bastante distinto de outro, devido a força do anúncio ou da emissora anunciada, seja por estar presente em programas com mais audiência, ou por mais vezes, ou conjugado em horários diferentes [5]. Todos esses fatores podem influenciar e mudar de forma notável a demanda por tal produto. Como não foi possível levar em consideração essa informação, acaba somando-se as outras variáveis citadas acima, tornando o problema mais complexo.

Somadas as variáveis externas, ainda temos uma importante variável interna que é difícil de mensurar que a execução da promoção em loja. Apesar do produto ter sido anunciado na televisão, é importante que a loja deixe visualmente expostos os produtos com cartazes no interior e no exterior da loja, com anúncios periódicos do locutor, disponibilidade do produto e uma vasta exposição do produto na área de vendas.

Com tantas variáveis possíveis e cenários distintos, impossíveis de prever, é de suma importância a experiência, conhecimento e sentimento de mercado do comprador, para que de alguma forma ele consiga equacionar essas variáveis e fazer uma oferta relevante e impactante. Sendo assim, o intuito do projeto é construir uma ferramenta que possa também auxiliar este profissional nesse

dimensionamento e que a escolha possa ser mais acertada. Com tantas variáveis, somente uma curva que projetasse a demanda de determinado produto poderia ficar muito limitado, com erros muito grandes, desestimulando seu uso. A escolha então foi pela demarcação de uma área de possível demanda com base na curva que será explicada mais à frente no projeto.

3 Objetivos

O objetivo fundamental do trabalho é desenvolver uma ferramenta que auxilie no entendimento da demanda por tipos específicos de produtos quando são anunciados em uma mídia de massa. Nesse caso em específico, na televisão aberta, tornando essas análises de fácil entendimento para o usuário e auxiliando-o no momento de decisão. Para isso foi desenvolvido uma visão clara e mais objetiva possível do comportamento da demanda do produto, para que se torne uma ferramenta de uso diário dos compradores durante seus planejamentos.

Uma vez que a crença é de que o comportamento geral da demanda pelos produtos de supermercado cresce exponencialmente em relação a diminuição do seu preço [6], será possível confirmar com a ferramenta se esse comportamento de fato se reproduz para todos os produtos. O objetivo é demonstrar através de um gráfico o comportamento da demanda de acordo com a redução de preço, e entender se de fato cresce exponencialmente, ou se a partir de determinada redução a demanda tende a estabilizar, ou se possui um outro tipo de comportamento inesperado. Além disso, será possível também estimar o volume a ser vendido com um determinado preço, através de um intervalo mais preciso gerado pela estimativa da demanda.

Ao entender o volume de demanda de determinado produto para a empresa, é possível avaliar se é um produto interessante para ser anunciado na televisão, se tende a atender os objetivos. Um dos pontos avaliados nesse aspecto é: com uma redução de preço em n%, comparando com o potencial de demanda esperado, seria um esforço justificável para empresa? Até que ponto de demanda esperada é justificável o sacrifício de margem daquele produto? Qual será o preço ideal do produto?

Ao estimar o volume, será possível também auxiliar não só na compra do produto, mas também na sua distribuição para as unidades de venda. Ações que ocorrem em momentos diferentes e podendo ser realizadas em diferentes níveis da estimativa.

É importante alcançar a facilidade de entendimento do resultado, a confiança e a aplicabilidade na rotina. O projeto visa trazer uma ferramenta adaptada aos processos da empresa, evitando ao máximo se tornar uma ferramenta engessada, com uma grande curva de aprendizado para que não seja uma grande barreira.

Tratando-se de uma ferramenta de auxílio de tomada de decisão, o nível de acerto é de extrema importância para o sucesso do trabalho. Uma vez que o tema

possui um vasto contexto de variáveis que podem influenciar consideravelmente a demanda que está sendo estudada, será considerada uma faixa de segurança para classificar o acerto da ferramenta. Com a faixa será possível ter uma dimensão de quais níveis de demanda podemos ter de acordo com um dado cenário.

4 Atividades realizadas

Ao início do projeto, diversas frentes que impactaram no mesmo foram estudadas, desde conteúdos inerentes ao negócio, visando a construção de uma base sólida para que pudesse entender melhor a operação e o que seria de fato relevante, quanto a conteúdos técnicos, para operacionalizar a ferramenta e contribuir com os usuários. Grande parte desses estudos iniciaram durante a primeira parte da disciplina e se estenderam até a finalização do desenvolvimento, sempre aperfeiçoando e trazendo novas informações.

4.1 Estudos sobre o contexto do negócio

Para iniciar o projeto, seriam interessantes conhecimentos inerentes do negócio que eu ainda não possuía. Apesar de saber o tema que gostaria de estudar, ainda não conhecia o contexto do tema. Foi então necessária uma visão do tema sobre a perspectiva do negócio.

Com esses estudos consegui entender premissas que seriam fundamentais para o desenvolvimento do projeto e resultados relevantes. A partir deles consegui fazer um levantamento das premissas do negócio que são consideradas relevantes para o tema de demanda por um produto. A exemplo, não somente o preço é fator decisivo para a demanda, existem diversos outros fatores que influenciam na mesma. Sendo alguns mensuráveis e previstos como o dia da semana, dia do mês, força da mídia comprada. E outros que a princípio não são facilmente mensuráveis, como a força da mídia do concorrente, o impacto da mesma sobre as lojas da empresa e outros ainda que não são previsíveis como, qual produto será anunciado pelo concorrente? A que preço?

Levantadas essas particularidades, tive um domínio muito maior do que poderia ser utilizado e do que poderia ser relevante que pudesse impactar positiva ou negativamente no estudo que me propunha a fazer.

Além dos fatores que impactam a demanda, foi necessário estudar o negócio em si, entender o processo de forma geral para que pudesse ter uma clareza maior da importância do estudo para a empresa e no fluxo de trabalho tradicional. Me aprofundei um pouco nos entendimentos de marketing, para entender quais os objetivos das mídias de massa, mais especificamente da televisão, quais são os objetivos a serem cumpridos a partir dela e qual sua relevância na perspectiva de investimento para a empresa. Por ser um investimento de mídia de alto valor, é ideal que os produtos escolhidos sejam

extremamente precisos nos seus propósitos. No caso de mídia de massa como a televisão para o varejo alimentar, o principal objetivo é trazer fluxo de clientes para a loja, sendo necessários assim produtos atrativos para levar o cliente a loja e com preços muito interessantes para o cliente [4], uma vez que este é um dos principais diferenciais no momento de escolha do cliente.

Para que esse alto investimento traga o efeito esperado, é necessária muita atenção e fundamentos no momento de decisão do produto e do respectivo preço. Estudei então um pouco do fluxo da operação que leva o produto desde seu produtor até o cliente final. Entendendo melhor o processo de compra pelos compradores da empresa, quais informações são relevantes para a negociação e compra, em dados que eles se baseiam quando vislumbram o produto para televisão e como é o processo de entrega do produto, seja na loja ou no centro de distribuição, sendo que os produtos de televisão geralmente são entregues no centro de distribuição da empresa em função do grande volume.

Uma vez sabido o volume de compra e fechada a negociação é importante saber como vai ser a distribuição para as lojas, qual o volume que cada uma vai receber daquele produto. Em paralelo, outras informações são relevantes para a perfeita operação, como quanto tempo leva a entrega do produto pelo fornecedor, quanto tempo será necessário para distribuir para as lojas, e a data que será veiculado o anúncio e quanto tempo ficará vigente. Todas essas informações são relevantes para que a operação possa fluir da melhor maneira possível e não limite a venda por falta de produto ou não exposição em loja.

Esse estudo das variáveis e do funcionamento do negócio foi de fundamental importância para que o estudo da demanda pudesse trazer benefícios e ser incorporado ao processo, evitando erros por falta de entendimento que pudessem destorcer os resultados e assim deixar de ser útil.

4.2 Estudo das tecnologias utilizadas

Os dados que foram usados já estavam armazenados como parte da operação do supermercado, então minha função seria consumi-los limpar de acordo com as necessidades e realizar novos cruzamento para chegar nas conclusões que gostaria.

Uma vez que precisava obter os dados de outros bancos através de conexões, precisaria limpar e estruturar para realização do modelo, optei pela escolha do Python por sua tradicional aplicação e suporte a ciência de dados [7].

Em conjunto, optei pelo Pandas [8], biblioteca para facilitar a manipulação e visualização dos dados durante o desenvolvimento. Escolhida também por conta da facilidade de integração com banco de dados, seja para consumo ou para inserção de dados.

Já para a visualização de resultados e tendência fez-se uso Matplotlib [9], biblioteca para a criação de gráficos e representações dos dados que estavam sendo processado. Muito relevante no processo de descoberta durante a busca por correlação entre os dados quanto para a visualização de resultados durante o uso da ferramenta.

Durante estudos preliminares, com um conjunto de dados mais reduzido comecei a ter algumas dificuldades durante a manipulação e limpeza dos dados, apesar de ser um conjunto reduzido, ainda era grande o suficiente para afetar a performance da forma como eu estava fazendo. Comecei a coletar todos os dados brutos a respeito do produto que estava sendo estudado diretamente do banco de dados e armazenando em um grande DataFrame (estrutura da biblioteca Pandas) [10] do produto e fazendo limpeza e ajustes a partir dele. Porém essa estrutura é toda gerada e manipulada em tempo de execução na memória, o que estava me trazendo problemas com o passar do tempo com o aumento da complexidade de operações e busca de novos dados.

Cheguei a cogitar se seria uma ineficiência da biblioteca Pandas e pesquisei por outra que pudesse substituir, onde encontrei o Dask [11], que é construída a partir da estrutura básica do Pandas, o DataFrame [10] e compartilha mesmo uso de funções básicas, porém lida melhor com gerenciamento de memória e permite processamento multi-threads. Essa nova realidade parecia mais adequada ao problema que estava tentando resolver e atendeu de forma considerável o ambiente controlado de testes.

Porém, quando começaram os testes com o volume real de dados, de vários produtos distintos, a performance já não estava interessante novamente, o processamento era demorado e esse tempo de resposta não seria suficiente para atender a demanda de uso real da ferramenta. Comecei então a rever a forma como havia estruturado o consumo e processamento e decidi modificar.

Revisitando a forma como tinha decidido atuar anteriormente, puxando todos os dados possíveis do produto e tratando diretamente no DataFrame [10], percebi que essa não era a melhor abordagem para o problema em específico. Optei então, restritamente pelo uso dos dados que seriam utilizados no momento em específico, delegando assim a performance de busca, separação e contas básicas para o SGBD, uma vez que serão consideravelmente mais eficientes e

preparados para essa função. Com isso trazia para o DataFrame [10] somente os dados que seriam trabalhados no momento. Consegui assim aliviar muito a carga de trabalho anterior que se mostrou desnecessária.

Porém nessa nova estrutura, o uso do banco de dados seria o de produção, o que com o uso indevido, ao até mesmo um erro durante os testes poderia causar um grande problema no desempenho da função para a qual ele foi desenhado para desempenhar, que seria o registro de informações e transações em tempo real. Para contornar a situação, criei um banco de dados paralelo que o único objetivo é alimentar as informações para o estudo proposto. Dessa forma, foi possível planejar um momento mais apropriado para captura dos dados do banco de produção de forma que impactasse o mínimo possível no uso do mesmo.

O novo banco é alimentado e consumido exclusivamente por funções contidas na ferramenta gerada por esse projeto. Dessa forma foi possível garantir que não ocorra duplicidade de informações, nem mesmo inconsistências, uma vez que ele está sendo consumido somente pela ferramenta e é atualizado de forma padronizada com base em uma fonte única de verdade.

Neste novo modelo, o desempenho teve uma melhora significativa. Todas as transformações mais simples e limpezas foram de responsabilidade do SGBD do banco de produção realizar, já trazendo os dados mais preparados para o uso final. Os resultados dessas buscas ficaram originalmente guardados em uma tabela própria, para que, caso seja necessário reprocessar os dados tratados para o modelo, não seja necessário ficar acessando o banco de produção. A partir dessa base então, foram aplicados processamentos para os dados que alimentam meu modelo já estivessem prontos na forma na qual serão usados.

4.3 Estudo da base de dados

O estudo com dados históricos só foi possível por conta da vasta base de dados que já existia e era alimentada, e em cima de informações extraídas desta, consegui executar o modelo.

Porém, é uma vasta base, com as mais diversas informações, de venda, de compra, transferência de produtos, estoques, receita, obrigações legais, contábeis, definição, histórico de preço e diversas outras informações que alimentam o ERP que populam todo o banco e que também ainda não havia sido utilizada com o intuito de extrair informações do gênero que almejava. Logo, não havia nenhuma estrutura com essa finalidade.

Sendo assim me aprofundei inicialmente nos dados de venda somente como data, preço e volume de venda, e com eles iniciei meus estudos para entender quais outros dados poderiam ser interessantes e se eu teria esses dados armazenados historicamente. Em paralelo executei uma busca exploratória pelas tabelas que já estavam sendo usadas e já conhecia. Consegui adicionar informações que pareciam ser relevantes como receita total com o produto em determinada data e o volume total vendido.

Em paralelo com essas buscas realizei algumas entrevistas também com pessoas pela empresa que tem o hábito de consumir mais dados, para que através da experiência deles eu pudesse encontrar outros dados que, à primeira vista, poderiam não me parecer tão óbvios, mas fariam sentido.

Com esse processo então, montei a base final que foi utilizada para o projeto com informações como dia do mês, dia da semana, consolidado de produtos onde o anúncio na televisão correspondem a mais de um código de produto diferentes, volume de saída, receita gerada, estoque do produto e momento que foi anunciado na televisão.

Todas essas informações vieram das mais diferentes fontes que compõem a base de dados total da empresa. Seja do banco que registra as vendas no ponto de venda, seja o estoque que é gerenciado pelo módulo de logística do ERP, pelo anúncio de oferta de produto, o tipo e o código de produto vinculado, que é gerenciado pelo módulo de marketing do ERP e por informações de experiências das vendas e dos anúncios das pessoas que entrevistei.

A composição destas fontes formou uma nova base apartada que é utilizada exclusivamente pelo modelo. A nova base suporta não somente as informações extraídas do banco de produção como também a composição e cálculo de novos índices que são importantes para o funcionamento do modelo.

4.4 Estudo da correlação entre a demanda e o preço

Partindo do princípio que é sabido que a demanda por um produto e seu respectivo preço tem notória correlação, comecei a me aprofundar nos fatores que poderiam influenciar na demanda, em conjunto com o preço em casos de anúncio da oferta na televisão e que tem alta relevância para o setor varejista de alimentos.

Para isso, em entrevistas com compradores e pessoas experientes do mercado levantei um conjunto de hipóteses que nortearam minha pesquisa com os dados. Alguns pontos importantes fortemente ressaltados foram a data, onde mais especificamente é relevante saber se é início de mês ou se já passou da

metade do mês, em conjunto com o dia da semana.

Por conta da grande maioria das pessoas receberem o salário no início do mês, o impacto na demanda por um produto com preço atraente pode ser maior do que o padrão. Uma vez que a pessoa estaria com mais recursos para possivelmente estocar, se for um produto que não seja muito perecível, um determinado produto sendo atraído pelo senso de oportunidade. Em conjunto com esse cenário, soma-se o dia da semana, onde tradicionalmente, início de semana, de segunda a quarta-feira tem movimento mais fraco para o varejo em contraponto à parte final da semana, de quinta-feira a sábado, onde o movimento é mais forte e as pessoas tendem a se preparar para o final de semana.

Além das datas cotidianas, somam-se as sazonalidades, que são épocas ou datas bem definidas, onde o padrão de comportamento e busca por determinados produtos se alteram consideravelmente saindo de um possível padrão. Esses momentos podem ser observados em datas como Natal, Dia das Mães, Páscoa, que para a praça onde a empresa atua são algumas das datas mais importantes do ano em venda. Além das citadas, que são períodos mais curtos, existe também a sazonalidade de carnaval, verão e outras que são mais extensas e ainda assim são bastante impactantes.

A esse cenário são somados outros fatores como o impacto da concorrência, com sua presença física, agressividade na mídia e prática de preços, as condições climáticas da região, o cenário econômico macro que atinge diretamente nossos clientes, a safra no caso de alguns produtos em específico e a execução em loja, disponibilidade, cartazamento, locução e outros. Fatores esses que já foram explicados anteriormente.

A composição de todos esses fatores e micro fatores influenciam diretamente no cliente e no seu consumo, adicionando assim complexidade extra, desafio que o modelo se propõe a estudar.

Todos esses levantamentos foram de extrema importância para que no final o modelo tivesse um resultado aceitável dentro da realidade na qual opera o supermercado hoje. Foi possível estimar se o modelo é capaz de superar esses fatores que saem de um padrão esperado e afetam diretamente a demanda pelo produto.

4.5 Estudo dos dados utilizados no modelo

Com todos os fatores levantados acima, gerou-se um mapeamento dos dados que seriam importantes para a construção do modelo. Já era sabido

também da disponibilidade e do acesso a eles. Cruzando as informações dos levantamentos que fiz dos dados disponíveis com as informações relevantes para o negócio, comecei a consolidar os dados que seriam úteis de todos os produtos que levantei. A escolha dos produtos foi baseada, dentre os produtos que mais foram anunciados na televisão no ano de 2021 e tivessem todos os dados históricos desde o início de 2020 salvos na base, para que a análise histórica pudesse ser mais acertada.

Já com todos os dados salvos, durante o processo de entendimento de qual seria o melhor modelo a ser construído, alguns dados que não estavam salvos começaram a aparecer como relevantes para constituir novas informações para a análise e passaram a fazer parte do banco de dados, assim como outros dados se mostraram não muito efetivos diretamente para a análise como estava sendo construída.

Algumas dessas informações quando também adicionadas diretamente ao modelo acabavam trazendo resultados interessantes em algum produto em específico, porém distorcendo e dificultando na análise dos outros produtos. Por conta disso, algumas informações continuaram armazenadas no banco de dados que suporta o modelo, porém acabaram não sendo utilizadas diretamente no modelo, o que não impede de serem utilizadas em melhorias futuras, uma vez, que do ponto de vista do negócio são relevantes.

Um dado em específico, que foi o registro de preço da concorrência não pode ser levado em consideração por conta da falta de registros históricos. O modelo utiliza dados de todos os produtos desde 01 de janeiro de 2020 e os dados de concorrência estruturados por produto passaram a ser registrados somente em meados de 2021, então por esse motivo ficaram apenas registrados no banco de dados do modelo, porém não puderam ser utilizados em momento nenhum.

Todos esses mapeamentos e estudos puderam nortear o caminho a ser seguido para o desenvolvimento do modelo, se mostrando passo fundamental para completo entendimento do problema, dos dados e da possível forma de solução. Foram indispensáveis para o sucesso do desenvolvimento do projeto.

5 Projeto e especificação

Devido à grande complexidade que envolve a atividade em si, a quantidade de variáveis que influenciam no processo e a dificuldade de antecipação de algumas informações que podem ser cruciais para uma tomada de decisão correta, se mostra um desafio bastante complexo de ser resolvido do alto grau de precisão. Então seria interessante que o projeto encontrasse uma forma, mesmo que mais genérica de auxiliar o processo de tomada de decisão com forte embasamento para os pontos possíveis.

5.1 Desafios a enfrentar

O varejo, principalmente o alimentar tem uma peculiaridade bastante interessante, apesar da alta recorrência de compra, que em tese deveria acarretar uma maior facilidade de previsão de consumo, na prática, essa relação direta não acontece na intensidade que se é esperada. Em estudos iniciais analisando o comportamento de um produto específico, foi possível notar a relevância de quando era anunciado na televisão, onde a demanda pelo produto saltava de forma relativamente padronizada.

Entretanto, como podemos observar no gráfico abaixo, foram registrados casos também onde a demanda superou muitos das expectativas que vinham sendo observadas anteriormente. E casos em que se registrou vendas abaixo do padrão esperado.

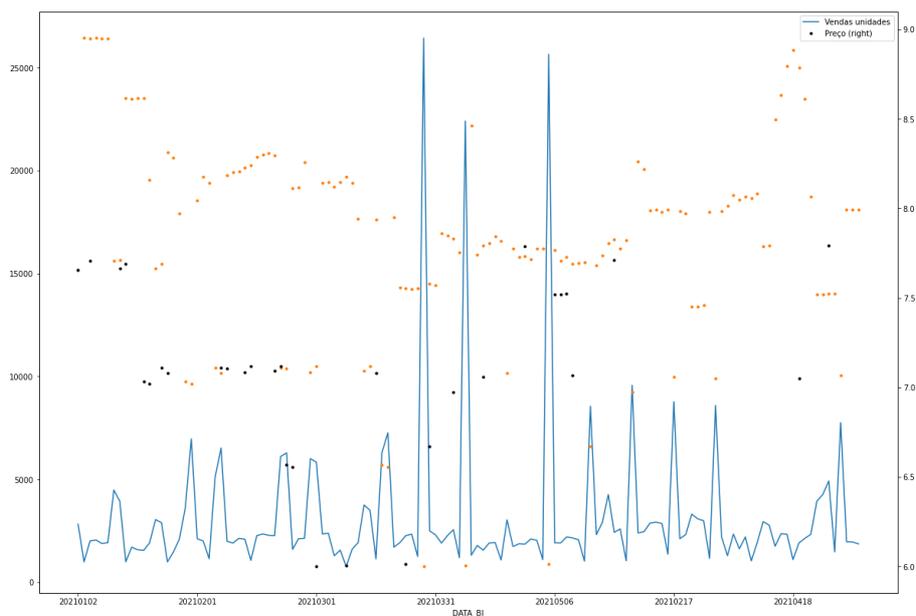


Gráfico 1: Vendas do óleo de soja

No gráfico, nas linhas em azul, temos o volume vendido em unidades, e nos

pontos seus respectivos preços de venda, sendo que os pontos pretos representam que estavam sendo anunciados na televisão e os pontos laranja não estavam. O gráfico traz informações do período de 01 de janeiro de 2021 até 30 de abril de 2021.

Essa diferença retrata bem a complexidade que existe em encontrar os fatores que são decisivos na influência da demanda. Como já discutido anteriormente, esses valores podem ser mensuráveis ou não e podem ser analisados antes ou não.

A finalidade então é encontrar fatores que sejam influências importantes para a demanda, a fim de trazer previsibilidade para a ação, visando todos os mais variados produtos anunciados. Com isso, a abordagem dos valores e dos parâmetros passa a ser um pouco diferente, a análise para a ser mais comparativa com médias anteriores com a finalidade de tentar traçar um padrão de comportamento.

As métricas a serem consideradas passam a ser a elasticidade, que é o comparativo entre o foi vendido em média em um dia, sem anúncio na televisão nas últimas 10 semanas contra a venda do dia analisado, e a diferença percentual em relação a seu preço médio, que é calculado baseado nos últimos 30 dias de venda sem considerar momentos que tenha sido anunciado na televisão.

A partir daí, analisando os produtos de forma separada, foi possível perceber uma tendência na elasticidade das vendas com relação a diferença percentual em relação a seu preço médio, tendência essa que contém alguns pontos relevantemente distantes da tendência original em números absolutos, porém respeitando a curva.

5.2 Arquitetura

Para chegar a tal análise, foi necessário realizar uma engenharia de dados para que o modelo atingisse resultados satisfatórios. Os dados são extraídos do banco de produção da empresa, que tem funções muito bem determinadas. Então esses dados não ficam todos juntos em uma única fonte, uma vez que são informações que são importantes para diferentes áreas da empresa.

Sendo assim, os dados são capturados de suas fontes originais, sempre o mais atualizado possível, coletando informações de venda, marketing, logística e outras fontes. A partir daí, é realizada uma limpeza nesses dados para que sejam

armazenados em um novo banco, onde serão de uso exclusivo para o modelo.

O padrão definido para essa captura foi que seriam gravados dados históricos somente de produtos que tem considerável recorrência na televisão. Objetivando que a análise possa ser embasada em um maior volume de informações e por um período de 2 anos, uma vez que as típicas sazonalidades podem ser bem capturadas nesse período e evitando um excesso de informações. Para que realidades e momentos econômicos distintos não influenciem no modelo, para que não ocorra distorções.

Apesar do padrão definido de dois anos para trás, nessa análise foi realizado um corte a partir de 01 de janeiro de 2020, com o intuito de evitar que o momento anterior a pandemia do Covid-19 influenciasse muito no modelo, podendo distorcê-lo consideravelmente. Uma vez que a forma de consumo e as vendas se modificaram consideravelmente nesse período.

Para o novo banco, foram criadas 4 tabelas distintas conforme o diagrama abaixo, conc, consVendas, historicoProdutos e dadosModelo.

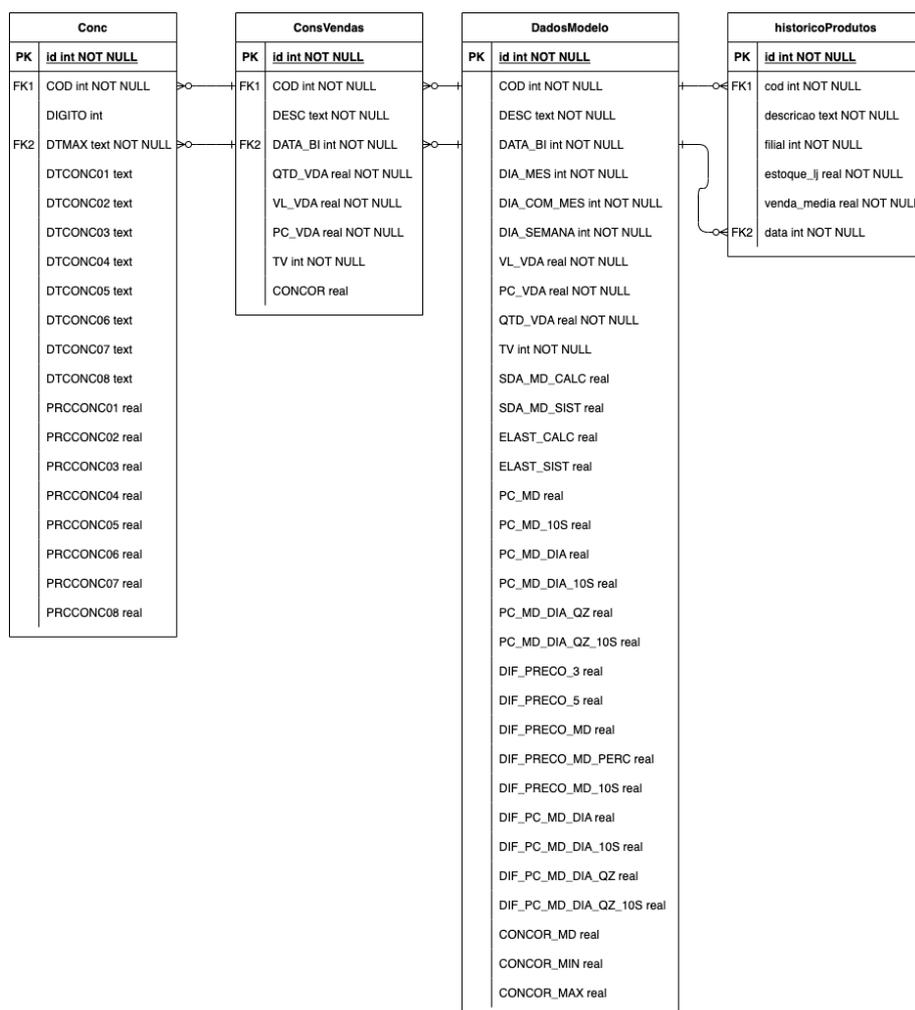


Figura 1: Representação do banco de dados

Dentre as tabelas, a `conc`, `consVendas` e `historicoProdutos` são representações das tabelas armazenadas pelo ERP no banco de produção, utilizando somente os dados pertinentes para a execução do modelo. Já a tabela `dadosModelo` são agrupamentos desses dados acrescidos de processamentos que agregam mais informação para o modelo.

Foi optado por esse formato, para que o banco de dados de produção não ficasse sobrecarregado, uma vez que é possível recriar, modificar e adicionar informações ao `dadosModelo` sem necessariamente ter que fazer novamente uma consulta ao banco de produção, visando interferir o mínimo possível na operação.

Agora detalhando um pouco melhor sobre que tipo de informação que popula cada uma, na tabela `conc` ficam armazenados dados de concorrência de todos os produtos que estão sendo analisados. A data da concorrência, o concorrente e o preço do produto naquele concorrente e a `datamax` representando a data mais recente é que se tem uma concorrência do produto. Para armazenar a concorrência no `dadosModelo`, em ocasiões que se tenha o dado do mesmo produto, na mesma data, opta-se por armazenar o valor mais baixo. Como já explicado anteriormente, esse dado é armazenado no `dadosModelo`, porém não é utilizado pelo modelo por conta da falta de dados históricos do período completo.

Já a `consVendas` armazena dados de venda do produto e de mídia. Nela temos acesso a código do produto, descrição do produto, data, quantidade total vendida do produto, receita total do produto no dia e indicativo se o produto estava na televisão ou não. Tendo as informações de quantidade e receita total do produto naquele dia, é calculado e armazenado também na mesma tabela o preço de venda médio. É importante que seja calculado dessa forma para que todos os produtos sejam englobados com igualdade. Uma vez que existem produtos que podem ter mais de um preço distinto praticado no mesmo dia, devido alterações de preço durante o dia que são exceções ou uma falha operacional ou sistêmica durante a carga de preço no ERP para as lojas.

Em `historicoProdutos` são armazenados dados como código e descrição do produto com data, seu respectivo estoque e um cálculo da saída média realizado pelo próprio ERP para o determinado dia. Essas informações são importantes para posteriormente calcular a elasticidade do produto com base na saída média fornecida. Além disso, fica salvo também a informação de estoque, apesar de não ser usada diretamente no modelo, em situações futuras podem ser interessantes para analisar um ponto que esteja muito abaixo da elasticidade esperada, pois caso a logística não funcione perfeitamente e atrase a entrega do produto e a loja não tenha estoque para vender, a elasticidade pode cair drasticamente, afetando

a análise geral. Sendo que nos casos estudados e apresentados, durante o período escolhido, não tiveram casos em que o estoque fosse mais baixo do que a saída média, ou em que todo o estoque foi vendido naquele dia, trazendo assim indícios de que a elasticidade não foi afetada.

O dadosModelo por sua vez é a tabela onde ficam armazenados todos os dados que são usados pelo modelo ou poderão ser usados, organizados de forma estruturada. Para identificar o produto, podem ser utilizados os campos de código ou descrição em uma determinada data. Assim como os citados, as informações de quantidade e receita total, preço médio de venda, indicação se estava na TV ou não, e a saída média do sistema, são alimentados pelas tabelas conc, consVendas ou historicoProdutos.

Para complementar as informações, temos o campo dia do mês que se mostra relevante para indicar se é primeira ou segunda quinzena do mês, uma vez que esses dois momentos são bem distintos para o varejo de alimentos. O campo dia da semana, que indica qual dia da semana representa aquela data, importante para avaliar se já está na parte final da semana, onde as vendas tendem a crescer, ou não. E o dia com o mês, que pode ser usado para identificar possíveis sazonalidades que se repetem durante os anos, como os feriados com datas fixas.

Visando agora as vendas, foi adicionada uma coluna de saída média calculada, uma vez que, em alguns registros, a saída média do sistema não possui dados, então a calculada é utilizada para substituir o dado faltante. Sendo calculada pelo mesmo padrão, que é a média de venda de um dia considerando o período histórico de 10 semanas sem considerar a venda dos dias em que o produto foi ofertado na televisão. No caso da saída média foi escolhido utilizar o período de 10 semanas pois é o mesmo cálculo que o ERP utiliza para calcular e que é usado para calcular o abastecimento automático de estoque que tem hoje na empresa. Logo, para manter a mesma unidade de raciocínio e visando suavizar possíveis sazonalidades, optou-se por manter o cálculo com as 10 semanas.

Com base nessa informação de saída média, é calculado o índice de elasticidade do produto, que será calculado somente quando o produto for anunciado na televisão. Sendo assim, o fator é calculado pela divisão da venda do dia em questão por sua respectiva saída média naquele mesmo dia, resultando assim um fator multiplicativo das vendas aquele dia.

Para entender a sensibilidade ao preço foram criadas várias métricas que são calculadas de formas parecidas, porém somente é alterado o foco onde são calculadas, com o objetivo de tentar entender se para determinados momentos

uma faz mais sentido uma do que outra.

Começando pelo preço médio, inicialmente foi calculado utilizando o mesmo critério que a saída média, calculava-se o preço médio praticado nas últimas 10 semanas desconsiderando preços de momentos que o produto estava sendo anunciado na televisão. Porém, a média levando em consideração as 10 semanas ficou muito longa para o preço, onde foi perdida a sensibilidade para alterações que são naturais do varejo, como sazonalidades, safra, situação da produção local entre outras.

Por conta disso, essa média passou a ser calculada com base em 30 dias. O preço médio é calculado como a média, de preço praticado dentro de um período de 30 dias, imediatamente anterior a data do momento calculado e desconsiderando os dias em que o produto esteja sendo ofertado na televisão. Mesmo assim, para possíveis efeitos de comparação futuros, o preço médio calculado com base em 10 semanas continuou sendo persistido em uma coluna separada.

Para tentar entender melhor possíveis peculiaridades, o mesmo cálculo de preço médio foi realizado restringindo o conjunto de dados a situações específicas: mesmo dia da semana e ao mesmo dia da semana na mesma quinzena, também somente como registro complementar para estudos futuros. Nesses casos, o conjunto de dados dos quais se calcula a média passa a se restringir a somente os mesmos dias dos quais estão sendo calculados. Por exemplo, caso o cálculo se refira a uma segunda-feira, o conjunto de dados históricos no qual será calculado o preço médio, será composto somente por preços referentes a datas que representam uma segunda-feira dentro do corte de tempo acordado. 30 dias ou 10 semanas. Valendo-se da mesma lógica para cálculos referentes a mesma quinzena, o preço médio é calculado da mesma forma, porém válidos somente dados que representam a mesma quinzena do momento que está sendo calculado. Ambos respeitando a regra de não estar sendo anunciado na televisão.

Após estes processamentos, para complementar, é calculada também a diferença de preço com relação a 3 e 5 dias. Essas diferenças são calculadas da seguinte forma, preço médio de 3 e 5 dias antes da data calculada respectivamente, menos o preço do dia em questão.

Além do recorte mais momentâneo como descrito acima, foi calculada também uma visão mais ampla, onde todos os preços médios calculados anteriormente são utilizados para calcular a diferença contra o preço da data calculada. Ou seja, resultando na diferença de preço médio, diferença de preço

médio dos mesmos dias e diferença de preço médio dos mesmos dias na mesma quinzena. Todos calculados com as bases dos 30 dias e 10 semanas também.

Porém uma coluna em especial sai desse padrão, que é a diferença de preço médio percentual. Ao invés de calcular o valor absoluto que resulta da diferença do preço médio para o preço da data calculada, é calculado o valor percentual. Pois esse formato se mostrou mais interessante para suportar momento de alta de preços expressivas como foi a do momento em que foi realizado o estudo.

5.3 Solução

Com todos os dados então estruturados, a solução começa a ser desenhada. Depois de inúmeras tentativas de abordagens diferentes, consegui identificar uma possível relação da diferença de preço médio percentual com a elasticidade do produto, comportamento esse que se manteve para todos os produtos testados.

Ao plotar os pontos que representam o volume de vendas com anúncio na televisão, a partir da perspectiva da diferença percentual entre preço médio definido (de 30 dias) e o preço praticado na televisão expresso no eixo x e a elasticidade representada no eixo y foi viável observar fatos relevantes. Apesar de não ser um grande volume de dados, foi possível visualizar uma curva que descrevia o comportamento geral. Então optei por realizar uma regressão linear com polinômio de grau 2 [12], onde o dado sabido encontra-se no eixo x e a informação que se pretende estimar está representada no eixo y.

Antes da realização da regressão, foi necessário retirar pontos que estavam muito fora da curva média, provavelmente casos bastante peculiares e específicos. Para isso então, foi aplicado um tratamento de outliers com base em quartis que será aprofundado na seção 6.1.

Apesar de todos os testes e ajustes realizados, ainda assim é um problema que sofre interferências de muitas variáveis, podendo ser imprevisíveis e até mesmo imensuráveis em alguns casos. Sendo assim, construir um modelo com alta precisão e altíssimo nível de confiabilidade é bastante difícil. Resolvi então trazer a visão do negócio do qual estava tentando resolver o problema e entender, mediante as limitações, e a forma como é operado hoje, o que seria bom e aceitável para se tornar de fato uma ferramenta de auxílio de tomada de decisão.

Com essa perspectiva foi criada uma faixa de segurança, onde, com base na variação máxima possível da elasticidade daquele produto, fica definido que

20% desse valor é deslocado para cima e para baixo da curva, formando assim uma região onde a demanda por um determinado produto poderá transitar de acordo com uma diferença de preço médio percentual e trazendo três possíveis cenários, o otimista, o médio e o pessimista. Dentro dessa faixa, foi relatado que a precisão é muito maior do que se é possível estimar hoje, com um nível de acerto calculável com base nos testes realizados.

Nos diversos testes, a regressão utilizando os parâmetros propostos parece se adaptar de forma bastante positiva ao comportamento do produto. Quando um produto é anunciado na televisão com um preço mais chamativo, é natural acreditar que a tendência é ter uma elasticidade de venda exponencialmente crescente de acordo com o aumento da diferença percentual de preço como pode ser observado no gráfico do óleo de soja abaixo.

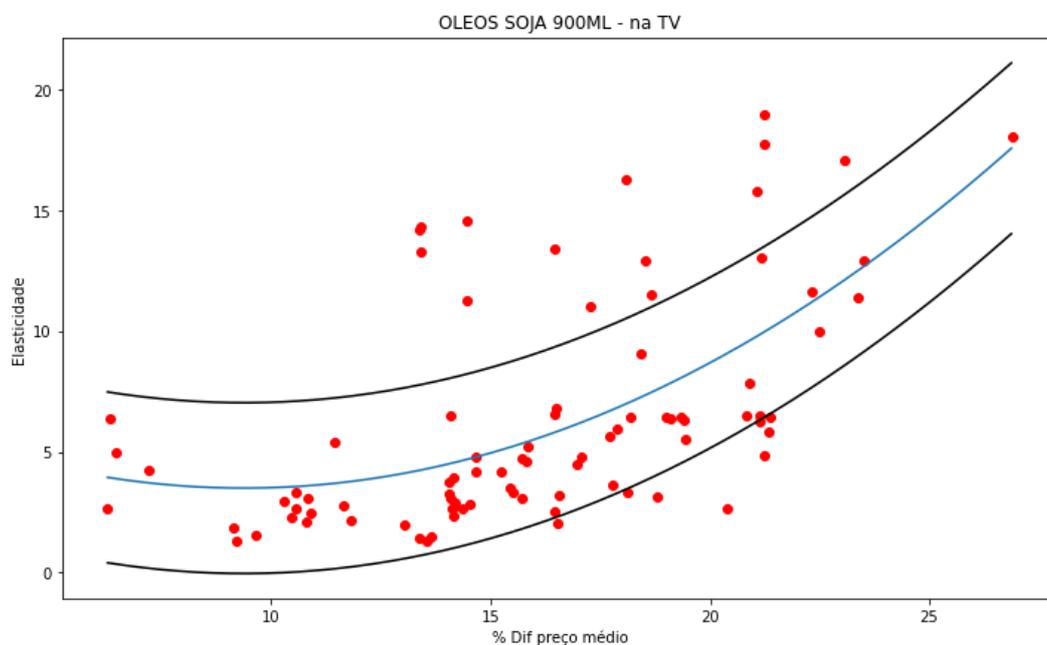


Gráfico 2: Regressão do óleo de soja

Mas apesar de parecer o natural, o modelo também foi capaz de identificar comportamentos que não são exatamente os esperados, porém são o que de fato acontecem, e sem o auxílio de uma ferramenta visual que trabalhe os dados dessa forma, visualizar esses comportamentos somente com os dados de venda, comparando com vezes anteriores fica humanamente muito difícil de perceber essas tendências. Para ilustrar com um exemplo, o gráfico abaixo demonstra a curva do leite UHT 1L, onde apesar da diminuição do preço, a demanda passa a

se estabilizar em determinado nível, conforme demonstra a inclinação da curva.

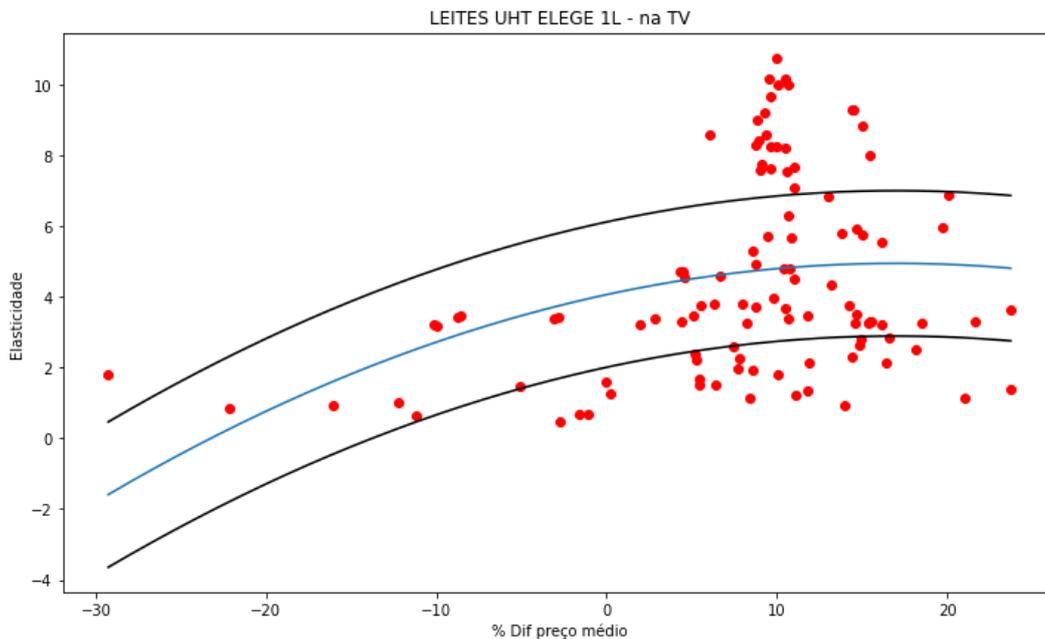


Gráfico 3: Regressão do leite UHT 1L

5.4 Propósitos: Novas visões

A ferramenta de visualização, busca cumprir seu papel de elevar o padrão da tomada de decisão nos quesitos abaixo descritos. Sendo o primeiro deles a escolha do produto, que através da avaliação de desempenho quando promovido na televisão, avalia-se se foi cumprido seu papel ao ser anunciado e se tem uma boa tendência para continuar sendo anunciado. Possibilita avaliar se esse nível de elasticidade que está sendo sugerido pelo modelo é interessante a ponto consiga justificar o significativo investimento que é realizado para promovê-lo na televisão. Com esse tipo de análise, a precisão desses produtos de impacto para os clientes tende a aumentar consequentemente.

Tendo visualmente informações como elasticidade prevista e porcentagem de diferença de preço, é possível também calcular a precificação mais interessante para o determinado produto. Visando o bom fluxo de clientes em loja como o esperado, e equilibrando o nível de sacrifício que a empresa está proposta a fazer no preço. Trazendo assim mais uma possibilidade interessante de otimização possibilitada pela ferramenta.

Como é sabida a saída média do produto no dado momento, passa a ser possível também auxiliar o momento de compra do comprador, uma vez que está sendo comprado com o foco de ação de televisão e se tem um preço estimado para anunciar, se torna possível estimar com um nível maior de confiabilidade a

quantidade que será necessária ser comprada para a determinada ação. E podendo contribuir também na fase de distribuição dos produtos do centro de distribuição para as unidades de venda.

5.5 Uso da ferramenta

No momento de escrita desse texto, a ferramenta ainda é concebida no formato de estudo, onde, para utilização são necessários conhecimentos básicos de programação para que seja rodado o modelo para o produto deseja. Ao rodar a ferramenta, o usuário define a visão e as informações desejadas que serão executadas a partir de uma ferramenta que permita a execução e visualização, como por exemplo o Jupyter Notebook, com todo o ferramental de integração, preparação dos dados, limpeza e montagem já prontos para uso.

Uma vez que está tudo pronto e utilizável para qualquer produto que tem histórico de ser anunciado na televisão, a base e captura desses dados também já está definida. Bastando encontrar o melhor momento de execução no banco de produção e os produtos que serão buscados, se serão os mesmos ou serão adicionados novos ou removidos, e executar a função para que os dados atualizados sejam salvos na base local e a partir daí a base de dados que alimenta o modelo possa se atualizar também consequentemente.

O ideal é que a ferramenta trabalhe sempre com os dados mais atualizados possíveis, para possa ter sempre o histórico mais recente de comportamento do produto, aumentando assim sua tendência ao acerto, e para que tenha o maior número de dados possíveis acerca do produto estudado.

6 Implementação e avaliação

O processo de implementação em si foi uma construção que foi se aperfeiçoando durante todo o projeto, seja na forma de captura dos dados, de tratamento, e até mesmo no modelo, entendendo os melhores parâmetros a serem usados e o tempo que seria levado como histórico.

6.1 Desafios enfrentados

Durante a implementação diversos problemas surgiram, necessitando de soluções durante o processo. Um deles que me chamou a atenção foi decorrente do uso do modelo na prática, no dia a dia da empresa. Por se tratar de dados históricos que já eram armazenados pela empresa, eles precisariam ser consumidos de alguma forma. Porém os dados estavam estruturados no banco de produção, onde o foco e finalidade nunca foi o consumo da forma que estava sendo proposto e sim a inserção de dados para possíveis análises futuras mais estáticas. Onde geralmente as informações são consolidadas via ETL ou até mesmo txt para serem consumidas e interpretadas por outras ferramentas.

Porém esse formato de operar não atenderia ao projeto, uma vez que podem ser grandes volumes de informações para serem consultados durante a rotina do dia a dia, e não havia a possibilidade de criar uma tabela com dados consolidados no banco de produção, pois por mais que o acesso pudesse ser mais rápido, dependendo do volume de dados, o processamento poderia afetar as atividades principais do banco de dados, que no momento se encontrava sem muita ociosidade de processamento.

Já a utilização via txt também não seria viável pela quantidade de informação que eventualmente poderia atingir, tornando esse processamento muito mais demorado do que o necessário, tanto para exportar quanto para importar. A ideia então foi realizar a integração para que pudesse rodar durante a madrugada, onde rodam também outros processos, mas teve possibilidade de adicionar mais esse sem maiores problemas.

Uma vez que essa integração era realizada de forma mais esporádica, e dependeria dela para realizar ajustes, e adicionar possíveis novas informações, optei por criar um banco de dados para uso exclusivo do modelo como foi descrito na seção 5.2 Arquitetura com mais detalhes. Dessa forma foi possível ter mais liberdade para trabalhar com os dados vindo da base de produção e de forma

mais isolada, e conseqüentemente mais segura.

Para a resolução do modelo, optei pelo Pandas [8], somente para auxiliar na organização e visualização dos dados vindos do banco durante o desenvolvimento e, também pela facilidade de integração com o banco de dados. Além disso, foi escolhido o Scikit learn [13], uma vez que o objetivo seria encontrar uma possível relação, optei pelo uso de uma biblioteca já exaustivamente testada e aprimorada para encontrar a relação entre as informações escolhidas.

Depois de muitos testes com diferentes parâmetros para tentar entender o impacto na demanda dos produtos anunciados na televisão, a relação entre a elasticidade e a diferença percentual do preço médio pareceu ser a mais interessantes visando o conjunto de produtos. Gráficamente, ao plotar os pontos, pareciam ter indicativos interessantes para realização de uma regressão, então uma vez que já haviam sido testados outros parâmetros e outros formatos e não tinha obtido resultados interessantes, resolvi optar pela regressão linear com polinômio de grau 2 com os parâmetros citados, sendo cada produto com a sua regressão, uma vez que dada um terá um comportamento específico.

Uma questão que precisou ser tratada durante a implementação foi a remoção de pontos que destoavam muito da amostra e acabavam distorcendo a regressão. Para tratar esse comportamento indesejado foi necessário descartar tais pontos antes de realizar a regressão, para isso utilizei a amplitude interquartil, retirando dados que estivessem consideravelmente abaixo do primeiro quartil ou também consideravelmente acima do terceiro quartil [14]. Sendo este um tratamento padrão aplicado a todos os produtos.

Me deparei também com a quantidade de dados, que talvez não fosse grande o suficiente para regressões mais apuradas caso fosse utilizado somente períodos históricos de somente um ano por exemplo. Para isso decidi trabalhar com histórico de até dois anos, onde é possível observar sazonalidades com mais clareza e os dados se mostram mais volumoso, pois aumentam as chances de determinado produto ter sido anunciado mais vezes na televisão, trazendo assim mais informação para o modelo. No modelo usado no projeto, os dados foram desde o início do ano de 2020, para que o ano ficasse completo de forma equivalente e para evitar que a rápida mudança de consumo e de preço de produtos proporcionada pela pandemia do Covid-19 interferisse nos modelos gerados.

Com dados prontos e estruturados e modelo definido, era necessário testar e entender se de fato era um bom modelo, então a escolha foi separar o conjunto de dados em dados de treino e outro conjunto como dados de teste, sendo essa

separação realizada na proporção de 80% para treino e 20% para teste, divididos sempre de forma aleatória, para que pudesse garantir uma maior confiança e não fosse afetado por nenhum tipo de viés.

6.2 Avaliação do modelo

Os primeiros testes, durante o desenvolvimento do modelo foram em sua grande maioria visuais, para entender a disposição dos pontos e da curva em relação aos pontos. A partir do momento que o modelo pareceu estar mais coerente, resolvi estruturar testes mais específicos. Como se trata de uma regressão, optei por algumas métricas para avaliar, o R^2 para medir a variação da variável resposta, o erro médio quadrático e sua raiz, o resíduo médio e a taxa de acerto para cada produto. Sendo que a taxa de acerto, uma vez que os dados são muito variáveis e dependem de informações que muitas vezes não podem nem serem medidas, foi criada uma faixa de segurança, que será novamente explicada na próxima seção, onde todos os pontos do teste que caírem dentro dessa faixa serão classificados como acerto.

Abaixo um exemplo de como ficou o dataset de treino e de testes do mesmo produto.

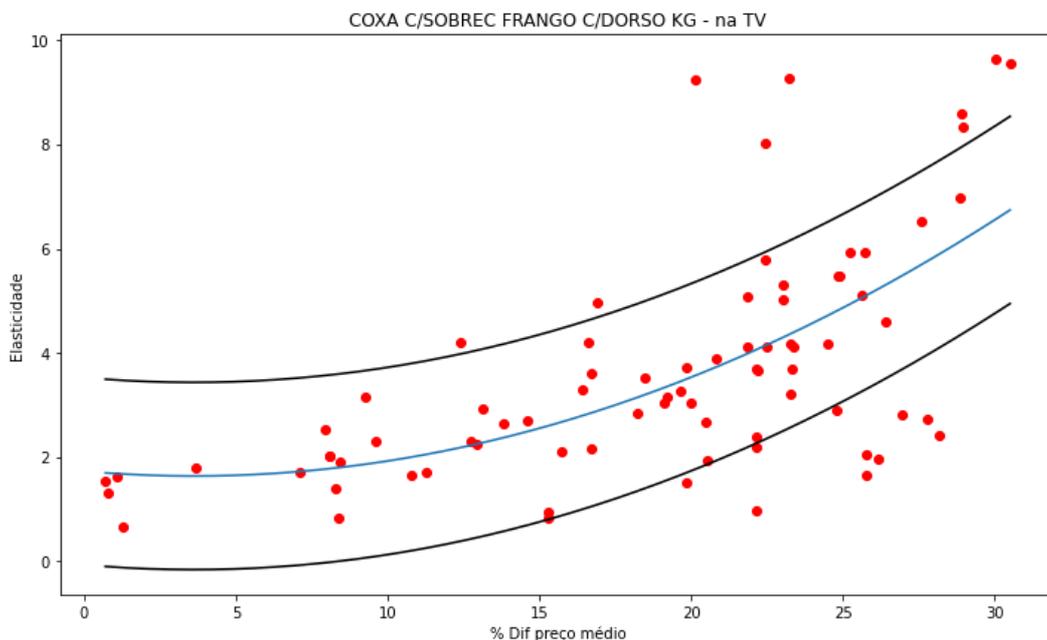


Gráfico 4: Conjunto de treino coxa e sobrecoxa de frango

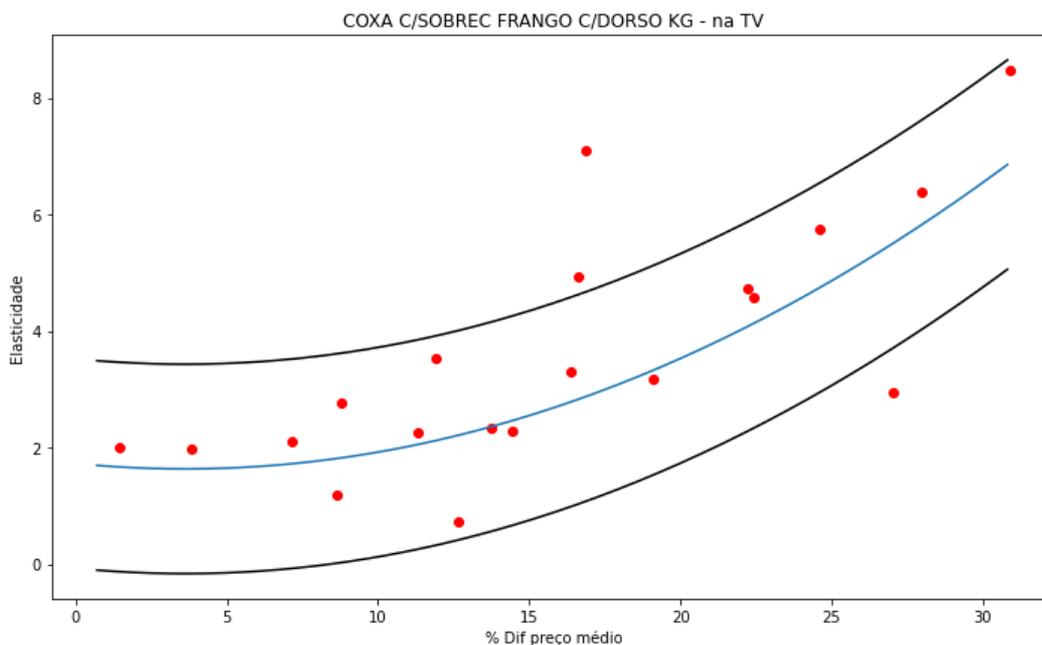


Gráfico 5: Conjunto de teste coxa e sobrecoxa de frango

Na tabela abaixo podemos observar as métricas dos resultados obtidos por oito produtos distintos que foram submetidos ao modelo.

Produto	R ²	Resíduo Medio	MSE	RMSE	Erro aceitável	Acertos testes
OLEOS SOJA 900ML	0.318706	-0.117363	16.922334	4.113676	7.082070	0.695652
LEITES UHT ELEGE 1L	0.104460	0.302874	6.734827	2.595154	4.111034	0.518519
CARRE SUINO CONG KG	0.132998	-0.403247	2.084419	1.443752	3.034658	0.652174
COXA C/SOBREC FRANGO C/DORSO KG	0.510542	0.485186	1.965916	1.402111	3.596031	0.850000
Acucar Uniao kg	0.293613	-0.027156	0.226222	0.475628	1.695763	1.000000
PERNIL SUINO C/OSSO C/TOUC RESF KG	0.073790	-0.035179	0.695608	0.834031	1.507146	0.705882
PALETA SUINO DIAN C/TOU C/OSS RESF KG	0.035500	0.086696	0.481029	0.693563	1.092763	0.500000
FILE PEITO FRANGO CONG KG	0.290217	-0.292118	1.170502	1.081898	2.004109	0.736842

Tabela 1: Resultados alcançados

Além dos parâmetros já conhecidos para avaliar uma regressão, foi adicionado também uma coluna com dados de erro aceitável, que é a distância entre o limite superior da faixa de segurança e o inferior, uma vez que a faixa de segurança se adapta de produto a produto, se torna um pouco mais fácil de analisar a raiz do erro médio quadrático de forma contextualizada.

Analisando do coeficiente de determinação R², apesar de ter sido baixo no geral, as outras métricas em conjunto com ele mostram que foram o suficiente para atingir um resultado satisfatório dentro do contexto proposto com a faixa de segurança. O resíduo médio também se manteve bastante próximo a zero quando analisamos todos os produtos, que se mostra interessante para o modelo, sendo que a média de todos os produtos avaliados foi de -0,000038.

Além disso, podemos observar que a raiz do erro médio quadrático em todos os produtos ficou menor do que o erro aceitável, indicando que, na média os erros

estão se comportando dentro de um erro aceitável e esperado, logo depois sendo confirmado pela porcentagem de acerto dos testes. Apesar dos acertos apresentarem uma ampla faixa a depender do produto, para o negócio foram resultados expressivos dado o cenário como será explicado abaixo.

Outro fato relevante do desempenho do modelo é a possibilidade de se adaptar aos dados, trazendo de fato a curva que melhor descreve a situação, não sendo afetado por viés, onde, de forma geral a curva deve crescer com comportamento exponencial. Trazendo até características do que pode indicar que de fato não seja um bom produto. Todos os gráficos dos produtos testados com suas respectivas regressões estão apresentados no apêndice A.

6.3 Avaliação para o negócio

Mediante a aferição dos resultados estatísticos, era necessário também entender o impacto e a importância da ferramenta para o uso no dia a dia da empresa. Visando entender melhor essa importância, diversas validações foram realizadas durante o seu desenvolvimento. Um exemplo dessas validações foi na escolha da diferença do preço médio, que inicialmente, apesar de parecer bom em alguns produtos, era calculado baseado no valor absoluto da diferença e não no percentual.

A explicação por trás dessa mudança é que o mais impactante para o cliente é a diferença de preço baseado em preços recentes. Porém por conta das oscilações de preço que temos no mercado, um produto que uma queda de preço de dois reais poderia gerar uma alta elasticidade em dado momento, mas pode também passar por outra situação em que essa mesma queda de dois reais não gere uma elasticidade tão interessante quanto antes. Uma vez que o produto tenha aumentado considerável seu preço original, e aquela diferença de dois reais deixe de ter o mesmo impacto. Dessa forma o cálculo percentual permite minimizar essas inconsistências no modelo.

Um exemplo de produto que notoriamente passou por esse processo foi o óleo de soja, onde no início de 2020 eram registrados preços de R\$ 3,79 a R\$ 4,50 na sua média e no final de 2020, início de 2021 foram registrados picos de até R\$ 8,96. Então a queda de preço, considerando somente o valor absoluto tem um impacto e relevância completamente diferente em cada um dos casos, se mostrando não ser tão eficiente o uso dessa métrica e mais eficiente o valor percentual. Essa observação de consumo e da importância de calcular de forma percentual só foi possível durante as interações com pessoas do negócio para

que elas pudessem validar as teorias e a forma como foram construídas.

Através dessas validações e trocas é que foi possível entender os distintos momentos do varejo, sendo primeira ou segunda quinzena, início ou final de semana, sazonalidades, impacto da concorrência, inflação ou deflação e outros fatores. E possibilitou usar esses fatores, os que são mensuráveis, para validar a abrangência do modelo, com ele se comporta nos mais variados cenários.

Um exemplo disso é o caso do óleo de soja, onde, ao testar cada uma dessas curvas, apesar das diferentes tendências e comportamentos, todas foram cobertas pela curva geral. O que demonstra que mesmo nos mais variados cenários do varejo alimentar possíveis observados em um mês, a faixa de segurança contemplou todos essas variáveis que são mensuráveis. Se mostraram dentro da tendência geral eficiente da mesma forma que é fácil de ser usado, não sendo necessário uma curva específica para cada cenário. Indicando ser uma boa forma para ser usada no dia a dia.

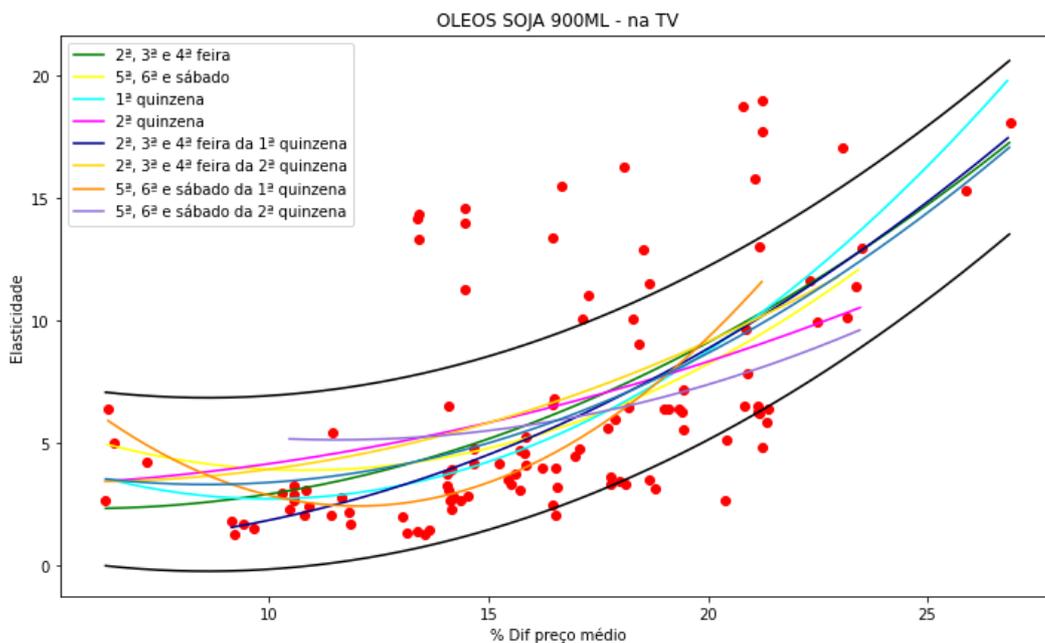


Gráfico 6: Desmembramento do óleo de soja

Nessas conversas, depois do modelo pronto, sendo testado em diversos produtos e cenários, posso considerar que dentro da limitação de dados e informações, que é uma realidade, o modelo se mostrou interessante e importante como ferramenta de apoio para o negócio. Trazendo um nível de acerto interessante e uma faixa mais específica para que o comprador com sua experiência e informações possa ter uma decisão mais acertada com a direção fornecida pela ferramenta.

Apesar dos acertos computados nos produtos testados variarem na faixa de

50% a 100%, o que é uma faixa ampla, tendo uma média de 70% foi considerado satisfatório, uma vez que há grande número de variáveis que não se pode prever ou controlar como descrito anteriormente e considerando a evolução de como é calculado nos dias de hoje. E mesmo assim, em caso de erro, sendo para um volume maior, uma vez que a tendência é comprar com margem para cima, o erro em média deve ficar dentro do intervalo de dias que foi calculado no erro aceitável. Considerando isso para a operação da loja, esses dias que venham a ser sobressalentes, podem ser facilmente absorvido na operação normal da loja em poucos dias.

A faixa de segurança, também foi vista com bons olhos, apesar de pode parecer alta, fez bastante sentido para a alta oscilação do negócio e sendo também uma faixa que é possível de trabalhar sem maiores dificuldades no pior cenário. Caso a estimativa de demanda seja o melhor cenário, e a demanda real seja o pior cenário, a operação não fica comprometida, pois a tendência é que o estoque daquele produto sirva para mais alguns dias, sem ferir a política de estoque de 15 dias da empresa. Nos casos estudados, a maior faixa de erro foi de 7 dias, o seja, caso aconteça o pior cenário, em 7 dias o estoque extra projetado tende a ser todo vendido, sem impactar negativamente na operação da loja. O que se mostrou uma comparação entre risco e retorno interessante para a empresa, levando em conta que esse cenário ocorreria apenas com alguns produtos, ao comparar com o estoque total, certamente esse acréscimo não passaria de meio dia.

7 Considerações finais

Ao final do trabalho, posso concluir que resultados interessantes foram alcançados. A ferramenta se mostrou capaz de fornecer informações interessantes e valiosas ao negócio de forma simples e direta com um considerável nível de acerto.

Os propósitos de gerar novas visões foram alcançados. A ferramenta permite avaliar os produtos e fazer a escolha com qualidade de quais produtos deveriam ser anunciados. A ferramenta foi capaz também de contribuir ricamente na tomada de decisão do preço conforme a elasticidade desejada e nível de sacrifício de margem. E finalmente auxilia na avaliação do volume necessário a ser comprado e a ser distribuído para as unidades de venda.

Para o uso diário de forma facilitada, ainda é necessário o desenvolvimento de uma interface gráfica para que seja facilmente acessível a todos os usuários. Apesar de entender a importância da interface gráfica para o uso diário, foi necessário realizar uma escolha, e priorizei o aprimoramento do modelo em questão. Uma vez que da forma como foi finalizado, o departamento de TI tem condições de gerar relatórios para os usuários de forma rápida e fácil. Quando disponibilizada, certamente será uma ferramenta de grande relevância para o negócio, possibilitando novas visões e informações que antes, apesar de existirem, não eram acessíveis ao usuário para que pudesse extrai-las de maneira simples e rápida.

Além de ser utilizada como estudo das vendas relativas a anúncio de televisão, com alguns ajustes na captura de dados, é possível realizar o mesmo estudo para outras mídias, como a digital que ganha cada vez mais relevância. E complementarmente pode ser adaptada para executar a mesma análise considerando loja a loja, sendo que o produto pode ter impactos diferentes de acordo com a localização e perfil do público. E quando usadas conjuntamente as diferentes versões, pode se obter informações muito relevantes para o negócio.

Um projeto bastante interessante do ponto de vista do aprendizado, que possibilitou a utilização de diversas áreas do conhecimento, mesclando diretamente a necessidade real de negócio e problemas encontrados, adicionando os cuidados necessários por se tratar um sistema que será acoplado ao principal sistema de produção.

Para trabalhos futuros, acredito que poderia ser interessante tentar explorar mais as outras variáveis que também foram coletadas, para entender, se dessa forma seria possível chegar a novas visões que possam contribuir ainda mais para

o negócio. Como alternativa, é possível explorar o uso da informação de concorrência que não havia ainda dados suficientes disponíveis, ou até mesmo de métricas completamente novas que possam interferir diretamente na relação estudada. Acredito ser interessante também a utilização de outras abordagens além da regressão, talvez o uso de redes neurais possa trazer resultados interessantes e eficientes para o estudo.

8 Referências

- [1] F. Provost e T. Fawcett, “Data Science, Engenharia e Tomada de Decisão Orientada em Dados,” em *Data Science para Negócios*, Alta books, 2013.
- [2] R. N. Filho, Associação Brasileira de Supermercados, Junho 2021. [Online]. Available: <https://superhiper.abras.com.br/pdf/270.pdf>. [Acesso em 28 Outubro 2021].
- [3] Intercontinental Supermercados, “Impacto do marketing nas vendas,” 2021.
- [4] S. A. W. D. Ferreira, Interviewee, *Relevância da mídia de televisão para o varejo alimentar*. [Entrevista]. 14 Abril 2021.
- [5] Rede Globo, “Manual Básico de Mídia,” 2015. [Online]. Available: https://negocios8.redeglobo.com.br/PDF/MidaKit/Globo_ManualBasico_pages.pdf. [Acesso em Outubro 2021].
- [6] M. C. A. F. d. Souza, “CONCEITOS DE ELASTICIDADE APLICADOS À MERCADOLOGIA,” [Online]. Available: https://www.eco.unicamp.br/neit/images/stories/CTAE_CD2/conceitos_de_elasticidade.pdf. [Acesso em 20 Novembro 2021].
- [7] W. McKinney, Python para Análise de Dados, Novatec, 2018.
- [8] “Documentação Pandas,” [Online]. Available: https://pandas.pydata.org/docs/user_guide/index.html#user-guide. [Acesso em Fevereiro 2021].
- [9] “Documentação Matplotlib,” [Online]. Available: <https://matplotlib.org/stable/api/>. [Acesso em Abril 2021].
- [10] “Documentação Pandas - DataFrame,” [Online]. Available: <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.html>. [Acesso em Fevereiro 2021].
- [11] “Documentação Dask,” [Online]. Available: <https://docs.dask.org/en/stable/>. [Acesso em Abril 2021].
- [12] S. Russell e P. Norving, “Regressão e classificação com modelos lineares,” em *Inteligência Artificial*, Elsevier, 2013, pp. 626-632.
- [13] “Documentação Scikit Learn - Regressão Linear,” [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html. [Acesso em Setembro 2021].
- [14] J. Brownlee, 18 Agosto 2020. [Online]. Available: <https://machinelearningmastery.com/how-to-use-statistics-to-identify-outliers-in-data/>. [Acesso em 21 Outubro 2021].

Apêndice A – Demonstrações visuais de gráficos gerados

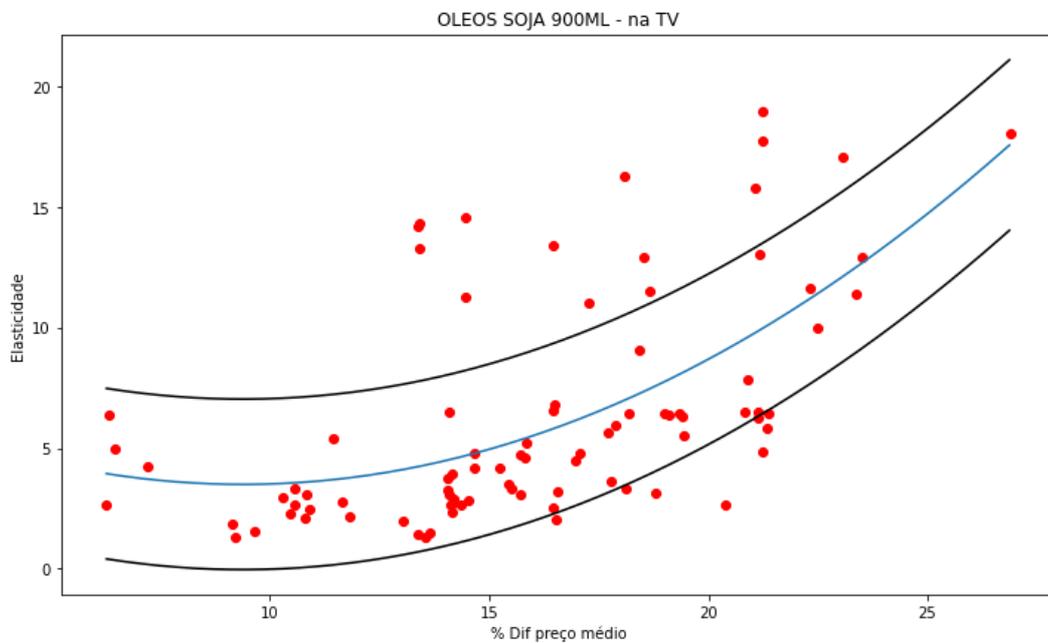


Gráfico 7: Óleo de soja - Treino

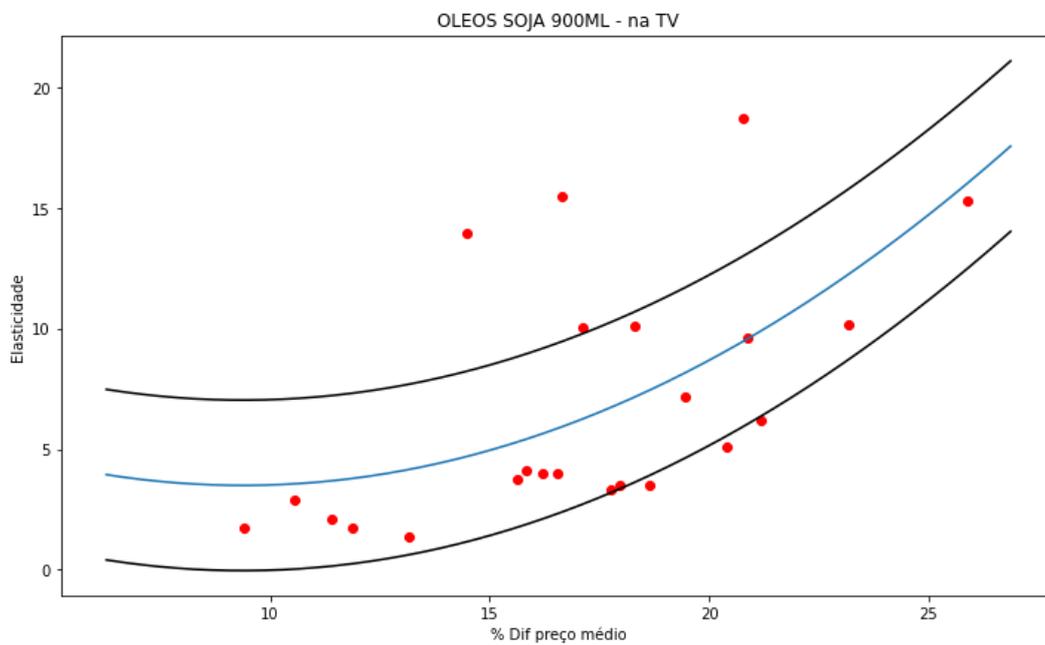


Gráfico 8: Óleo de soja - Teste

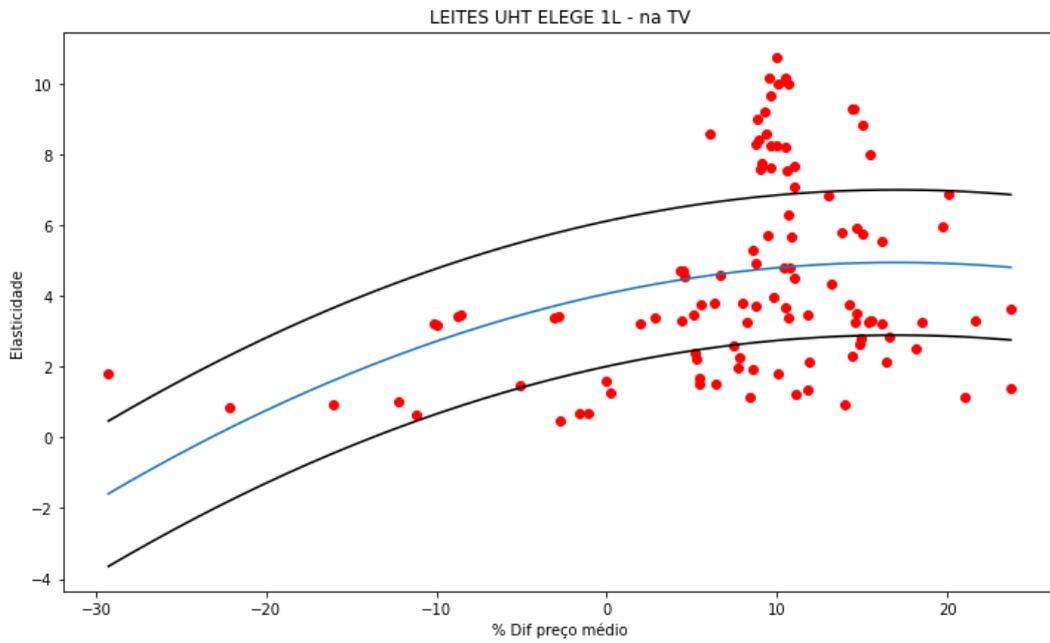


Gráfico 9: Leite UHT 1L - Treino

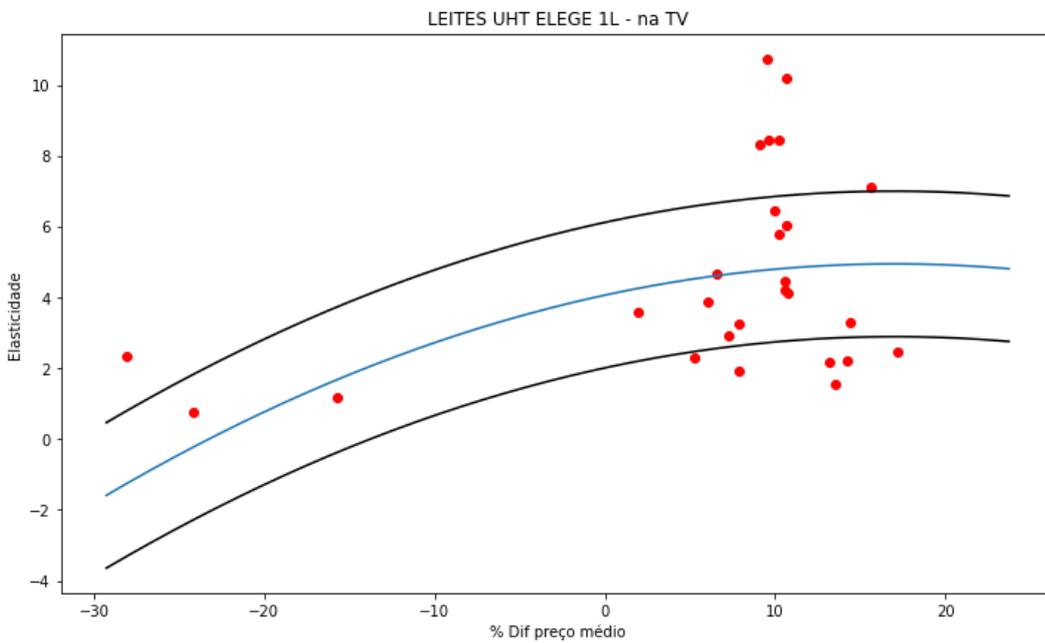


Gráfico 11: Leite UHT 1L - Teste

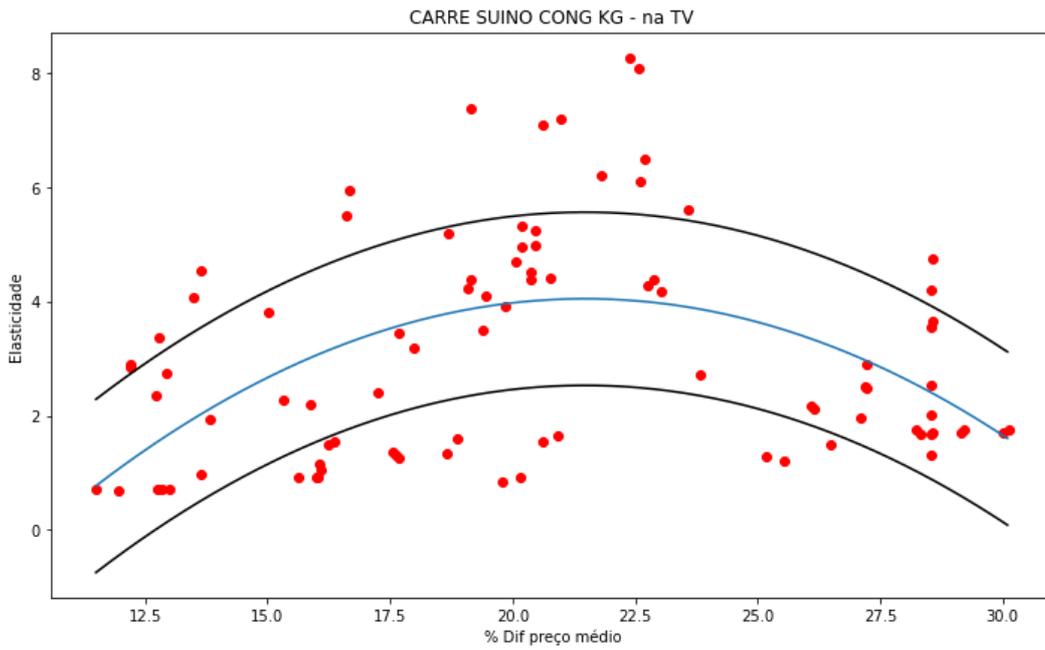


Gráfico 12: Carré suino - Teste

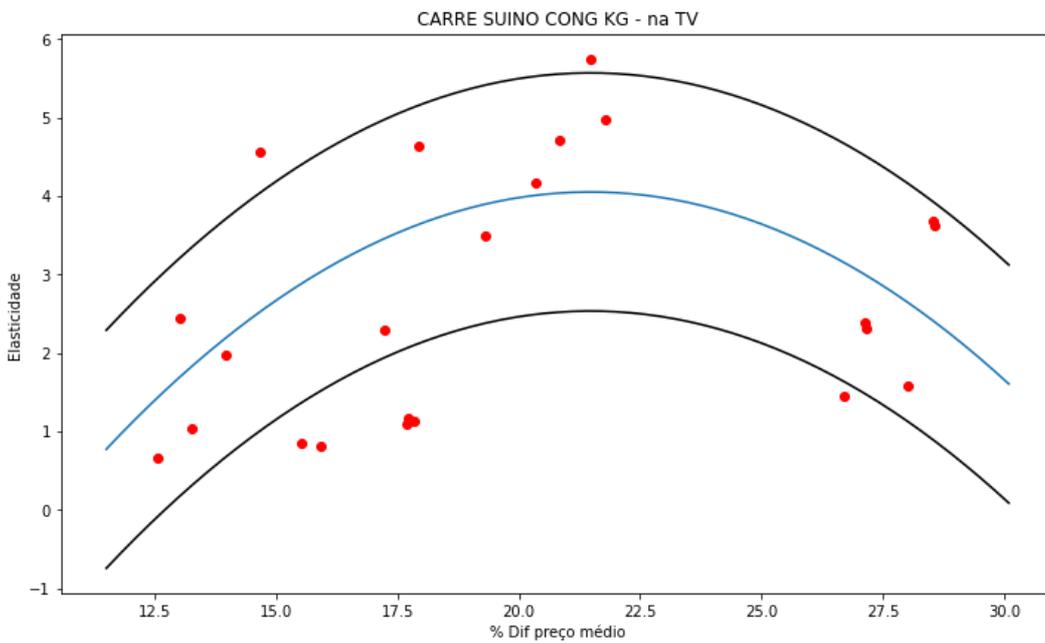


Gráfico 13: Carré suino - Teste

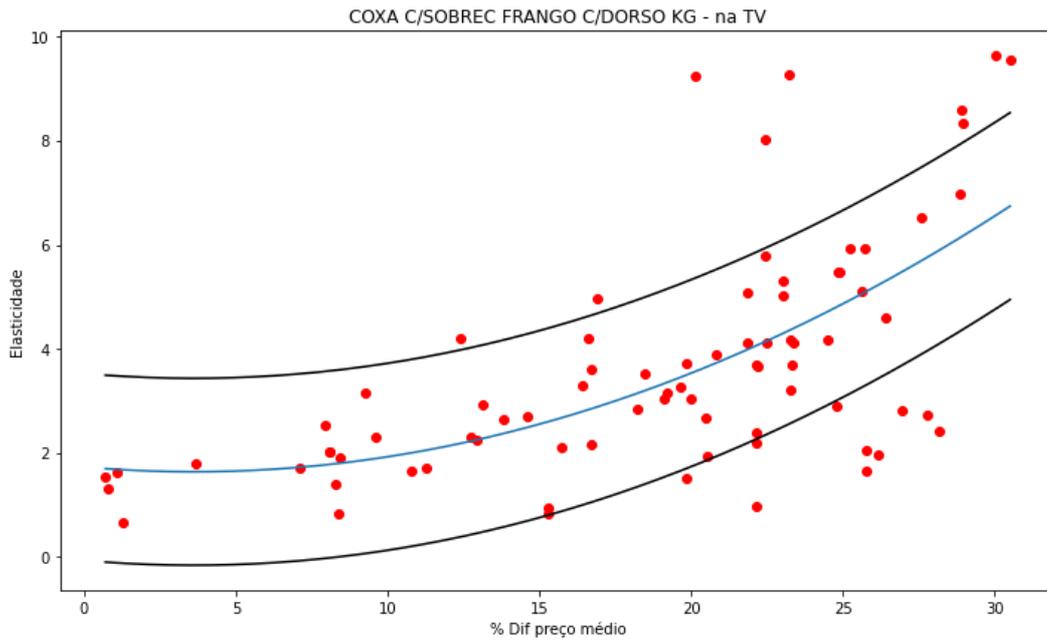


Gráfico 14: Coxa com sobrecoxa – Treino

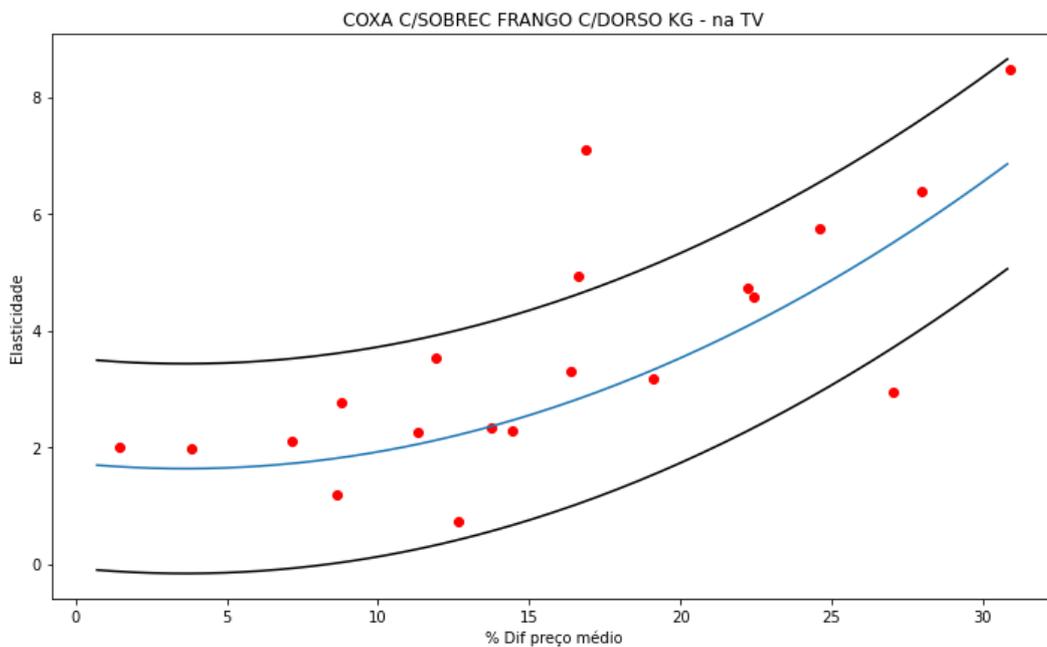


Gráfico 15: Coxa com sobrecoxa – Teste

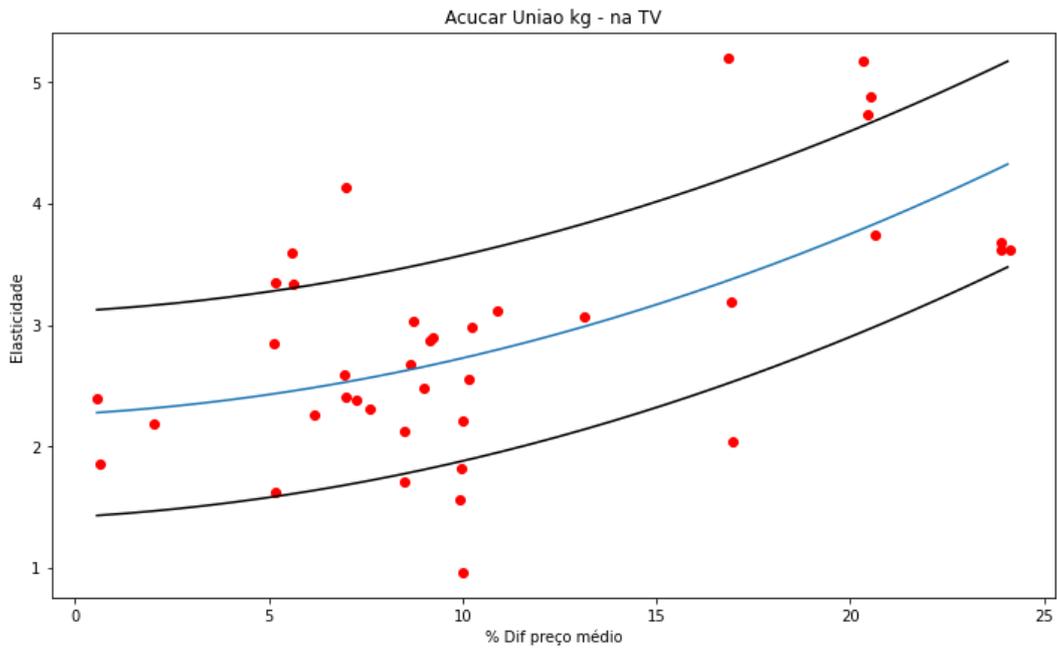


Gráfico 16: Açucar 1kg – Treino

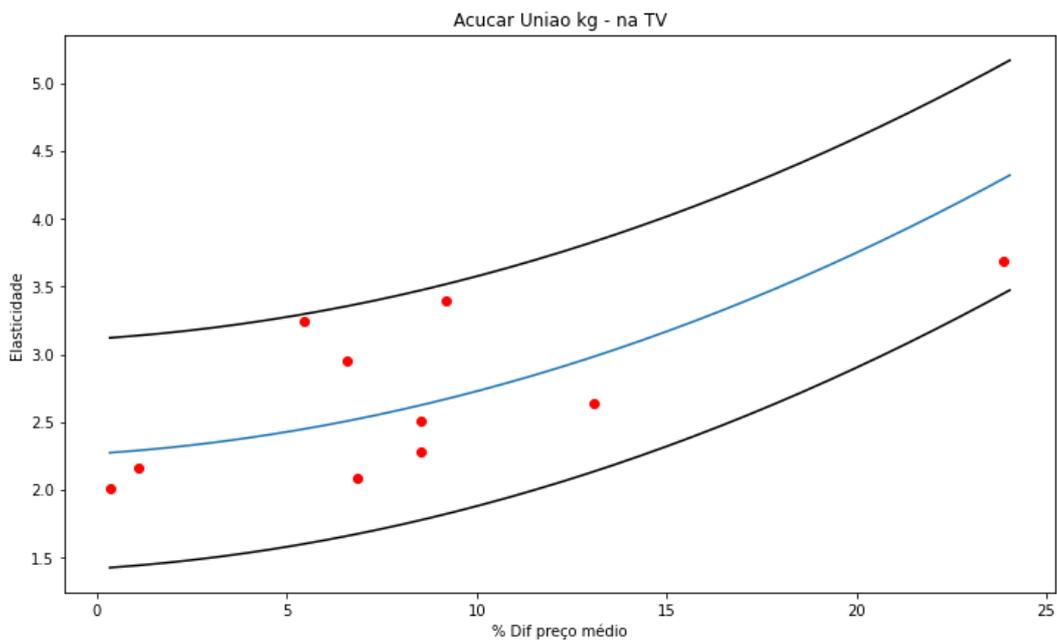


Gráfico 17: Açucar 1kg – Teste

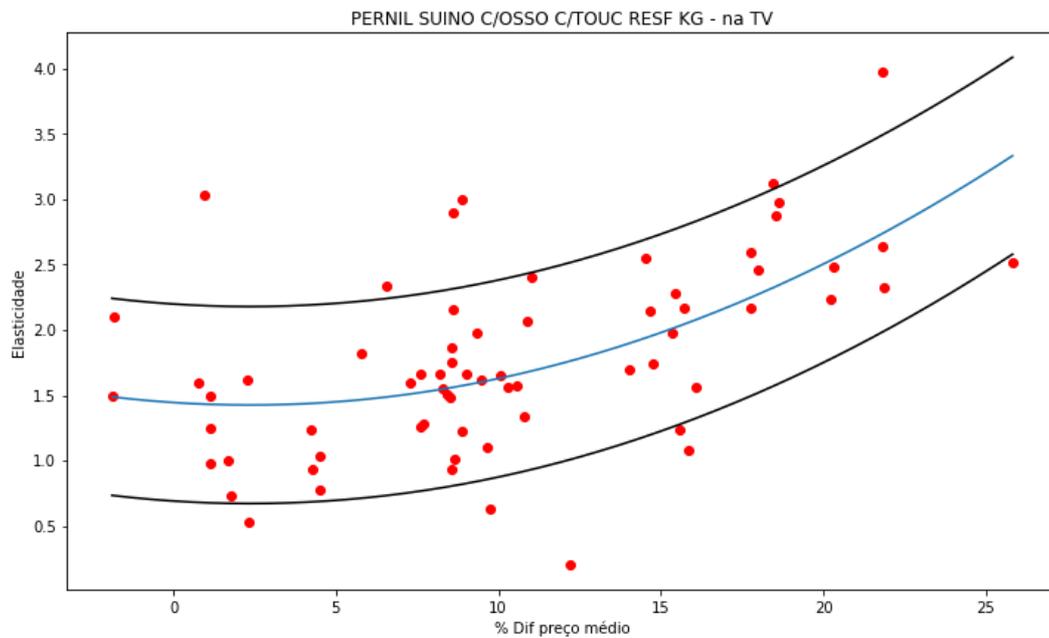


Gráfico 18: Pernil suino – Treino

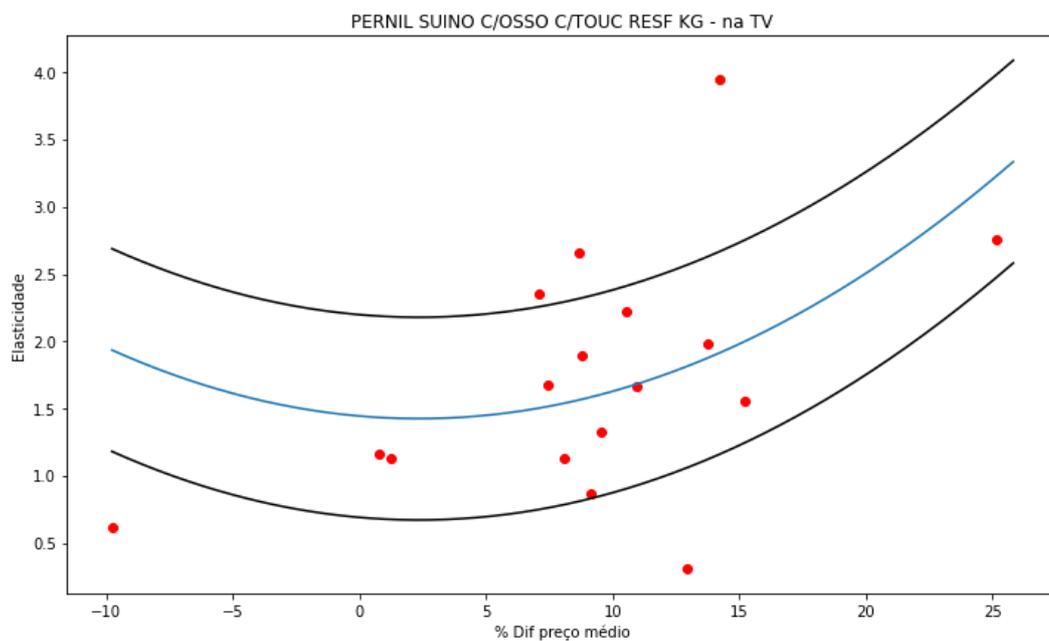


Gráfico 19: Pernil suino – Teste

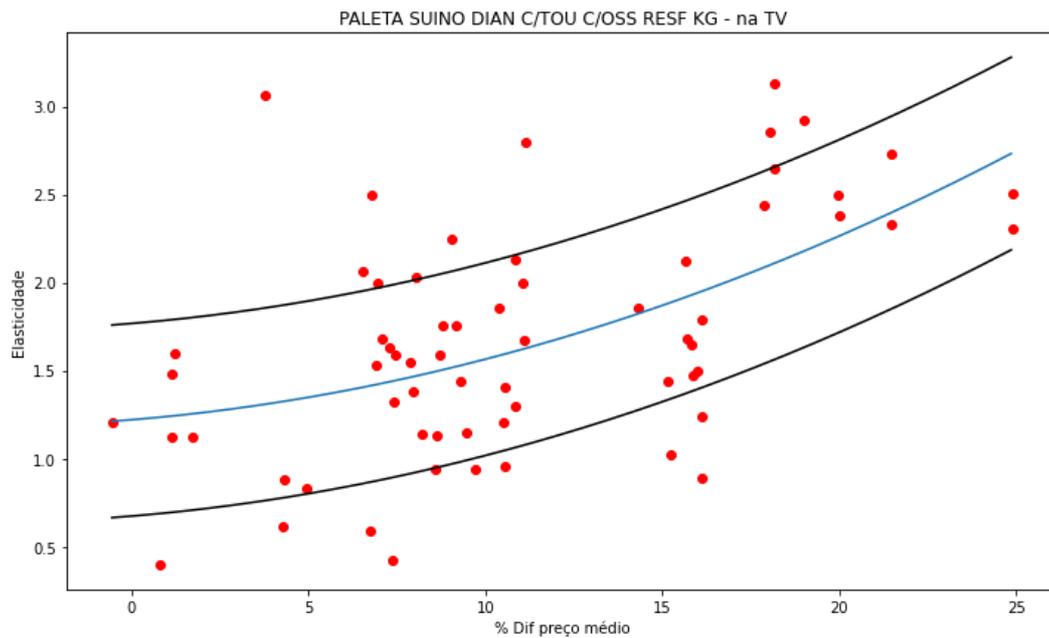


Gráfico 20: Paleta suino – Treino

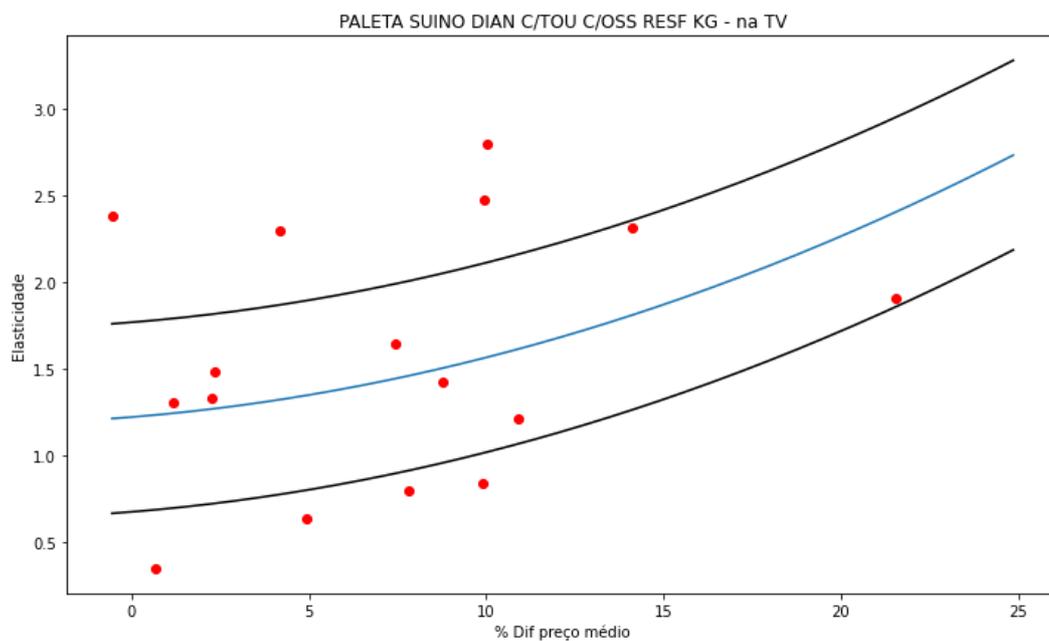


Gráfico 21: Paleta suino – Teste

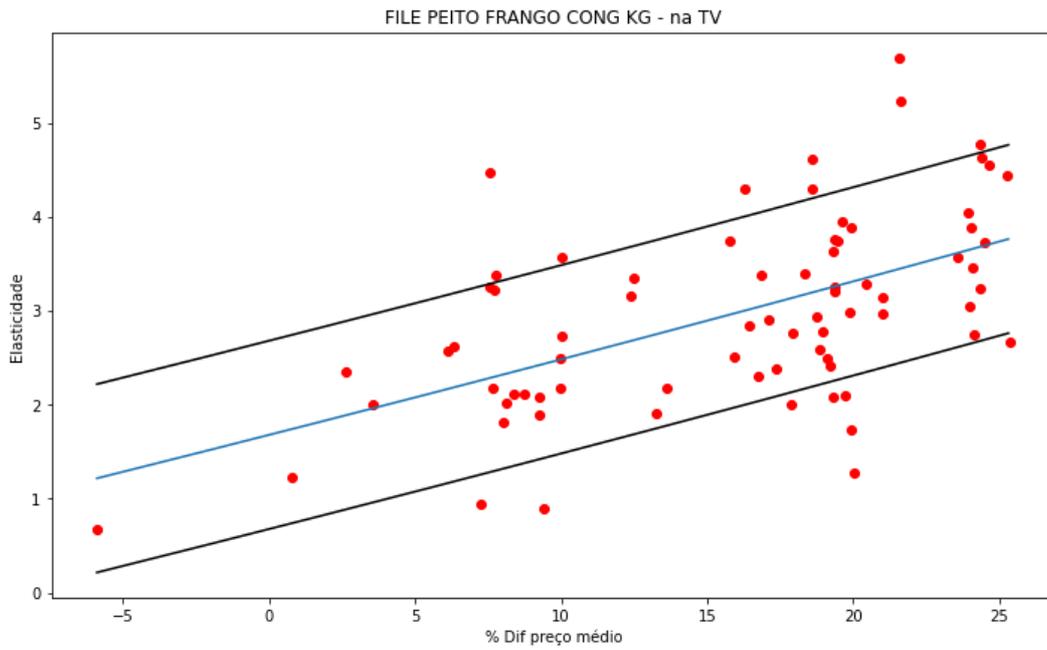


Gráfico 22: Filé peito de frango – Treino

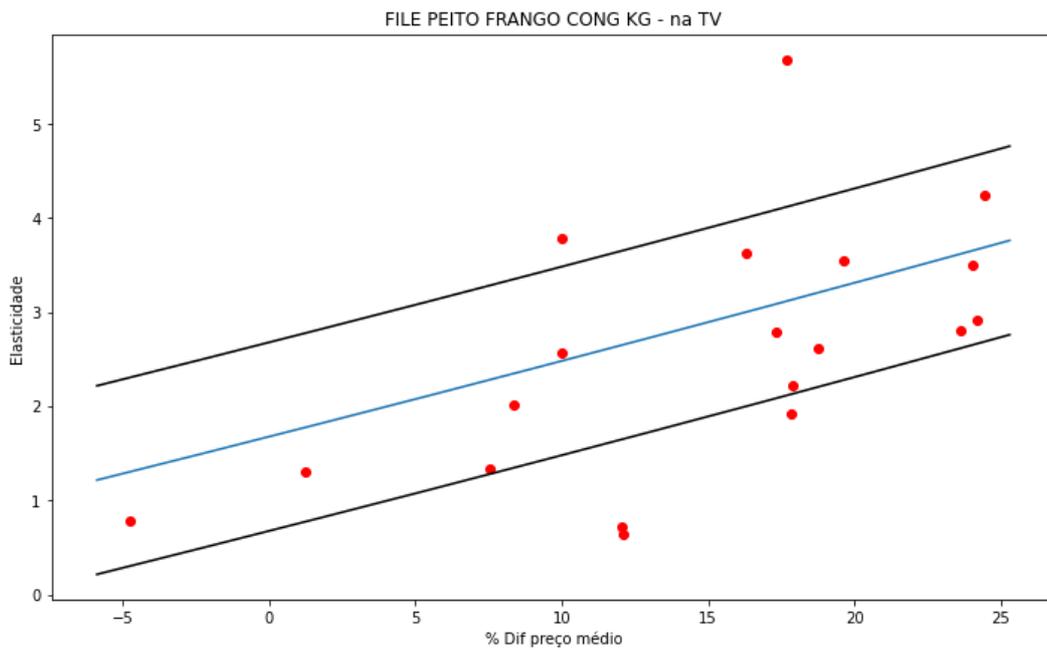


Gráfico 23: Filé peito de frango – Teste

Apêndice B – Processo de preparação dos dados

Um trecho do código para exemplificar o processamento dos dados preparados para o modelo:

```
def processaDadosModelo():
    conn = sqlite3.connect('data/banco.db')
    dfAux = pd.read_sql('SELECT cods FROM consVendas', conn,
    parse_dates={"DATA_BI": {"format": "%Y%m%d"}})
    listaProds = dfAux.CODS.unique().tolist()
    for artigos in listaProds:
        print(artigos)
        listaCods = artigos.split(",")
        df = pd.read_sql(f'SELECT CODS, DESC, DATA_BI,
    QTD_VDA, VL_VDA, PC_VDA, TV FROM consVendas WHERE cods =
    "{artigos}"', conn, parse_dates={"DATA_BI": {"format":
    "%Y%m%d"}})
        df['DIA_MES'] = df['DATA_BI'].dt.strftime('%d')
        df['DIA_COM_MES'] = df['DATA_BI'].dt.strftime('%d') +
    df['DATA_BI'].dt.strftime('%m')
        df['DIA_SEMANA'] = df['DATA_BI'].dt.dayofweek
        df['DIF_PRECO_3'] = df.apply(lambda row:
    calcDifPreco3(row['DATA_BI'], row['PC_VDA'], row['CODS'],
    "consVendas"), axis=1)
        df['DIF_PRECO_5'] = df.apply(lambda row:
    calcDifPreco5(row['DATA_BI'], row['PC_VDA'], row['CODS'],
    "consVendas"), axis=1)
        listaCods = [i.strip() for i in listaCods]
        strListaCods = transfListaSql(listaCods)
        dfConcProduto = pd.read_sql(f'SELECT DTMAX, PRCCONC01,
    PRCCONC02, PRCCONC03, PRCCONC04, PRCCONC05, PRCCONC06,
    PRCCONC07, PRCCONC08 FROM conc WHERE DTMAX IS NOT NULL AND
    PRODUTO in {strListaCods}', conn, parse_dates={"DTMAX":
    {"format": "%Y%m%d"}})
        dfConcProduto[['PRCCONC01', 'PRCCONC02', 'PRCCONC03',
    'PRCCONC04', 'PRCCONC05', 'PRCCONC06', 'PRCCONC07',
    'PRCCONC08']] = dfConcProduto[['PRCCONC01', 'PRCCONC02',
    'PRCCONC03', 'PRCCONC04', 'PRCCONC05', 'PRCCONC06',
    'PRCCONC07', 'PRCCONC08']].apply(pd.to_numeric)
        dfConcProduto = dfConcProduto.drop_duplicates()
        df['CONCOR_MD'] = df.apply(lambda row:
    calcConcMd(row['DATA_BI'], dfConcProduto), axis=1)
        df['CONCOR_MIN'] = df.apply(lambda row:
    calcConcMin(row['DATA_BI'], dfConcProduto), axis=1)
        df['CONCOR_MAX'] = df.apply(lambda row:
    calcConcMax(row['DATA_BI'], dfConcProduto), axis=1)
        df['DATA_BI'] = df['DATA_BI'].dt.strftime('%Y%m%d')
        df['SDA_MD_CALC'] = df.apply(lambda row:
    saidaMedia(row['DATA_BI'], artigos, "consVendas"), axis=1)
        df['SDA_MD_SIST'] = df.apply(lambda row:
    saidaMedia(row['DATA_BI'], artigos, "consVendas"), axis=1)
        df['ELAST'] = df.apply(lambda row:
    calcElast(row['TV'], row['QTD_VDA'], row['SDA_MD']), axis=1)
```

```

df['PC_MD'] = df.apply(lambda row:
precoMedio(row['DATA_BI'], artigos, "consVendas"), axis=1)
df['PC_MD_10S'] = df.apply(lambda row:
precoMedio10Sem(row['DATA_BI'], artigos, "consVendas"),
axis=1)
df['PC_MD_DIA'] = df.apply(lambda row:
precoMedioDia(row['DATA_BI'], row['DIA_SEMANA'], artigos,
"consVendas"), axis=1)
df['PC_MD_DIA_QZ'] = df.apply(lambda row:
precoMedioDiaQuinzena(row['DATA_BI'], row['DIA_SEMANA'],
artigos, "consVendas"), axis=1)
df['DIF_PRECO_MD'] = df['PC_MD'] - df['PC_VDA']
df['DIF_PRECO_MD_PERC'] = ((df['PC_MD'] -
df['PC_VDA'])/df['PC_MD'])*100
df['DIF_PC_MD_DIA'] = df['PC_MD_DIA'] - df['PC_VDA']
df['DIF_PC_MD_DIA_QZ'] = df['PC_MD_DIA_QZ'] -
df['PC_VDA']

# df = df.round({"QTD_VDA":2, "VL_VDA":2, "PC_VDA":2,
"DIF_PRECO_3":2, "DIF_PRECO_5":2, "CONCOR_MD":2,
"CONCOR_MIN":2, "CONCOR_MAX":2})
df.to_sql("dadosModelo2", conn, if_exists='append',
index=False)
return

```