

BRUNO HISASHI OTSUKA

CORRELAÇÃO ENTRE MÍDIAS SOCIAIS E VENDAS DE PRODUTOS SOB EFEITO
DE *PANIC BUYING* DURANTE A PANDEMIA DO COVID-19

São Paulo
2020

BRUNO HISASHI OTSUKA

CORRELAÇÃO ENTRE MÍDIAS SOCIAIS E VENDAS DE PRODUTOS SOB EFEITO
DE *PANIC BUYING* DURANTE A PANDEMIA DO COVID-19

Trabalho de Formatura apresentado à Escola
Politécnica da Universidade de São Paulo para
a obtenção do diploma de Engenheiro de
Produção
Orientador: Prof. Dr. Hugo Tsugunobu
Yoshida Yoshizaki

São Paulo
2020

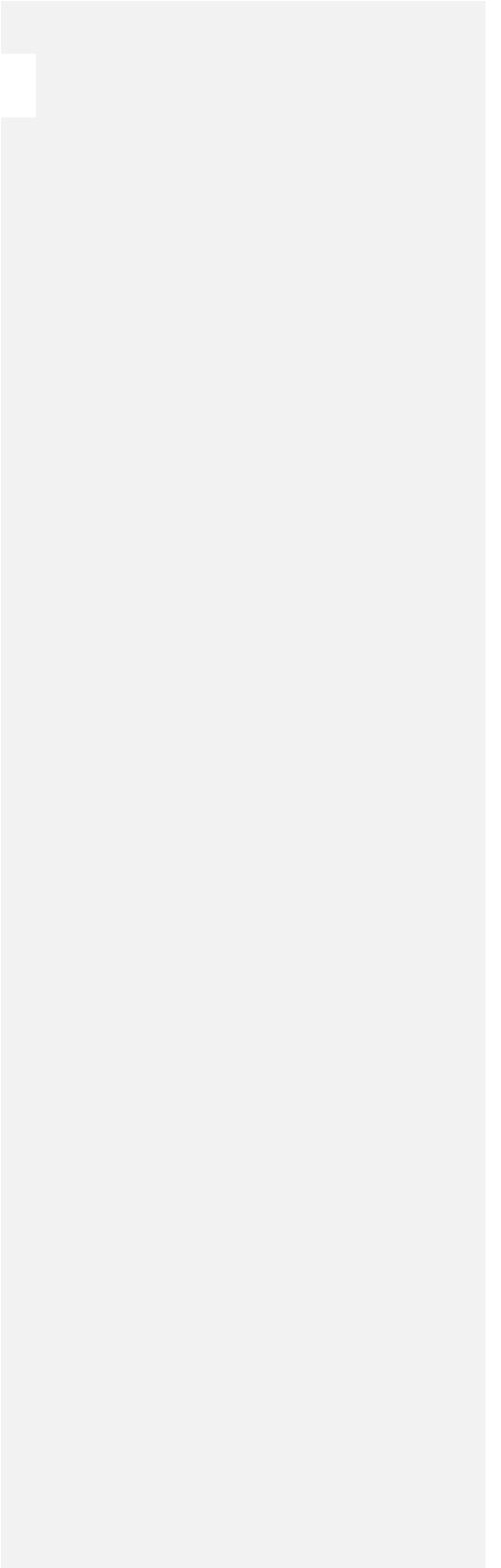
Catalogação-na-publicação

Otsuka, Bruno Hisashi

Correlação entre mídias sociais e vendas de produtos sob efeito de *panic buying* durante a pandemia do COVID-19/ B. H. Otsuka – São Paulo, 2020.
74 p.

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

1. Estatística. 2. Comércio. 3. Redes Sociais. Universidade de São Paulo. Escola
Politécnica. Departamento de Engenharia de Produção II.t.



À minha família e amigos.

AGRADECIMENTOS

Ao Professor Hugo Tsugunobu Yoshida Yoshizaki, por todo apoio, paciência, e ensinamentos transmitidos ao longo do desenvolvimento deste projeto.

Aos Professores Irineu de Brito e Celso Mitsuo Hino, por todo auxílio e apoio durante o projeto.

À Maria Clara Rodrigues Pinheiro, pela ajuda durante a busca por uma boa literatura para o projeto.

A todos os professores do Departamento de Engenharia de Produção da Escola e outros departamentos, por me tornarem apto a formar Engenheiro de Produção.

À minha família, pelo apoio incondicional durante a realização deste trabalho e em todas as etapas da minha vida.

À Bruna, a melhor amiga que conheci e que sempre me apoiou mesmo quando eu duvidei de mim mesmo.

Ao Thomas, por ser a primeira amizade real que fiz durante a graduação

Aos meus amigos Karina, Flávia, Daniel, Claudio, Henrique, Carina, Hector, Rafael, Lucas, Camila, Laura, Carol e André por todos momentos que vivemos nesses 5 anos, nunca me senti tão acolhido por um grupo de amigos antes.

Às minhas amigas Thaís, Caroline e Karen, minhas colegas de curso que me acompanharam em quase todas as disciplinas e me ajudaram a me tornar um engenheiro e uma pessoa melhor.

À Luana, que me mostrou uma grande amizade fora do meu círculo de amigos.

Agradecimento especial ao Vitor Urano, que foi meu colega de moradia durante quase todo meu curso e que proporcionou vários dos melhores momentos que tive na graduação.

Ao time de beisebol da escola politécnica, por toda experiência e trabalho de equipe realizado.

*“The morning will come again.
No darkness, no season is eternal”*

BTS

RESUMO

Durante a pandemia do COVID-19, ocorreu um fenômeno social conhecido como *panic buying*, em que as pessoas compram quantidades anormais de produtos em situações de crises, desastres ou epidemias. O presente trabalho tem como objetivo apresentar uma análise comparativa entre dados de vendas provenientes de uma empresa de varejo com dados da rede social *Twitter* e da ferramenta de busca *Google Trends* durante o período dessa pandemia. Dado que esse tema entrou em grande relevância nos últimos meses com a doença do COVID-19, que afetou vários países e causou rupturas em diversas cadeias logísticas, estudos que auxiliem a entender melhor como as pessoas se comportam nesses períodos são essenciais para futuras pesquisas e plano de ações para casos futuros similares a este. O plano de análise envolveu duas frentes: (i) coleta de dados das bases bem como o tratamento delas, e (ii) análise de correlação entre os dados. O resultado obtido com este trabalho foi de que existe uma forte correlação entre a quantidade de *tweets* referentes a um produto e as vendas do mesmo. O *Google Trends*, por sua vez, mostrou-se não confiável, havendo correlação apenas com os produtos que estiveram mais em evidência nos noticiários.

Palavras-chave: *panic buying*, estatística, comércio, redes sociais.

ABSTRACT

During the COVID-19 pandemic, a social phenomenon known as panic buying occurred, in which people buy abnormal quantities of products in situations of crises, disasters or epidemics. The present work aims to present a comparative analysis between sales data from a retail company with data from the social network Twitter and the search tool Google Trends during the period of this pandemic. Given that this topic has come into great relevance in recent months with the disease of COVID-19, which affected several countries and caused disruptions in different logistical chains, studies that help to better understand how people behave in these periods are essential for future research and action plans for future cases similar to this one. The analysis plan involved two fronts: (i) collecting data from the databases as well as their treatment, and (ii) analyzing the correlation between the data. The result obtained with this work was that there is a strong correlation between the number of tweets referring to a product and the sales of the same. Google Trends, in turn, proved to be unreliable, with correlation only with the products that were most in evidence in the news media.

Keywords: *panic buying, statistics, commerce, social networks*

LISTA DE FIGURAS

Figura 1: metodologia aplicada no projeto	Fonte: Elaboração do autor.	33
Figura 3: metodologia adotada para classificação dos <i>tweets</i> .	Fonte: elaborada pelo autor.	37
Figura 4: volume de vendas de papel higiênico no período.	Fonte: elaborada pelo autor.	42
Figura 5: pesquisa por papel higiênico no <i>Google</i> durante o período.	Fonte: elaborada pelo autor.	43
Figura 6: <i>tweets</i> de papel higiênico sem filtro durante o período.	Fonte: elaborada pelo autor.	43
Figura 7: <i>tweets</i> de papel higiênico com filtro durante o período.	Fonte: elaborada pelo autor.	44
Figura 8: gráfico de disperso do papel higiênico entre <i>tweets</i> e vendas.	Fonte: elaborada pelo autor.	45
Figura 9: gráfico de disperso do papel higiênico <i>entre trends e tweets</i> .	Fonte: elaborada pelo autor.	45
Figura 10: gráfico de disperso do papel higiênico <i>entre trends e vendas</i> .	Fonte: elaborada pelo autor.	46
Figura 11: volume de vendas de sabão no período.	Fonte: elaborada pelo autor.	47
Figura 12: pesquisa por sabão no <i>Google</i> durante o período.	Fonte: elaborada pelo autor.	47
Figura 13: <i>tweets</i> de papel higiênico sem filtro durante o período.	Fonte: elaborada pelo autor.	48
Figura 14: <i>tweets</i> de papel higiênico com filtro durante o período.	Fonte: elaborada pelo autor.	49
Figura 15: gráfico de dispersão do sabão entre <i>tweets</i> e vendas.	Fonte: elaborada pelo autor.	49
Figura 16: gráfico de dispersão do sabão entre <i>trends e tweets</i> .	Fonte: elaborada pelo autor.	50
Figura 17: gráfico de dispersão do sabão entre <i>trends e vendas</i> .	Fonte: elaborada pelo autor.	50
Figura 18: volume de vendas de álcool no período.	Fonte: elaborada pelo autor.	51
Figura 19: pesquisa por álcool no <i>Google</i> durante o período.	Fonte: elaborada pelo autor.	52
Figura 20: <i>tweets</i> de álcool sem filtro durante o período.	Fonte: elaborada pelo autor.	52
Figura 21: <i>tweets</i> de álcool sem filtro durante o período.	Fonte: elaborada pelo autor.	53
Figura 22: gráfico de dispersão do álcool entre <i>tweets</i> e vendas.	Fonte: elaborada pelo autor.	53
Figura 23: gráfico de dispersão do álcool entre <i>trends e tweets</i> .	Fonte: elaborada pelo autor.	54

Figura 23: gráfico de dispersão do álcool entre <i>trends</i> e vendas. Fonte: elaborada pelo autor.	55
.....	55
Figura 25: volume de vendas de massas no período. Fonte: elaborada pelo autor.....	55
Figura 26: pesquisa por massas no <i>Google</i> durante o período. Fonte: elaborada pelo autor... ..	56
Figura 27: <i>tweets</i> de massas sem filtro durante o período. Fonte: elaborada pelo autor.	56
Figura 28: <i>tweets</i> de massas com filtro durante o período. Fonte: elaborada pelo autor.....	57
Figura 29: gráfico de dispersão de massas entre <i>tweets</i> e vendas. Fonte: elaborada pelo autor.	57
.....	57
Figura 30: gráfico de dispersão de massas entre <i>trends</i> e vendas. Fonte: elaborada pelo autor.	58
.....	58
Figura 31: gráfico de dispersão de massas entre <i>trends</i> e <i>tweets</i> . Fonte: elaborada pelo autor.	58
.....	58
Figura 32: volume de vendas de doces no período. Fonte: elaborada pelo autor.....	59
Figura 33: pesquisa por doces no <i>Google</i> durante o período. Fonte: elaborada pelo autor....	60
Figura 34: <i>tweets</i> de doces sem filtro durante o período. Fonte: elaborada pelo autor.....	60
Figura 35: <i>tweets</i> de doces com filtro durante o período. Fonte: elaborada pelo autor.	61
Figura 36: gráfico de dispersão de doces entre <i>tweets</i> e vendas. Fonte: elaborada pelo autor.	61
Figura 37: gráfico de dispersão de massas entre <i>trends</i> e vendas. Fonte: elaborada pelo autor.	62
.....	62
Figura 37: gráfico de dispersão de massas entre <i>trends</i> e <i>tweets</i> . Fonte: elaborada pelo autor.	62
.....	62
Figura 39: matriz de correlação entre produtos para os dados de vendas Fonte: elaborada pelo autor.....	67
Figura 40: matriz de correlação entre produtos para os dados do <i>Twitter</i> Fonte: elaborada pelo autor.	68
Figura 41: matriz de correlação entre produtos para os dados do <i>Google Trends</i> Fonte: elaborada pelo autor.	69

LISTA DE TABELAS

Tabela 1: quadro-resumo sobre os artigos selecionados para revisão bibliográfica.....	28
Tabela 2: Teste de correlação dos produtos comparando <i>Twitter</i> e vendas.....	63
Tabela 3: Teste de correlação dos produtos comparando <i>Trends</i> e vendas.....	64
Tabela 4: Teste de correlação dos produtos comparando <i>Twitter</i> e <i>Trends</i>	64
Tabela 5: Teste de correlação dos produtos com <i>lag</i> comparando <i>Twitter</i> e vendas.....	65
Tabela 6: Teste de correlação dos produtos com lag comparando <i>trendr</i> e vendas.....	66
Tabela 7: Tabela-resumo das principais conclusões do trabalho.....	70

LISTA DE SIGLAS E ABREVIASÕES

OMS	Organização Mundial da Saúde
COVID-19	<i>Coronavirus Disease 2019</i>
SARS-CoV-2	<i>Severe Acute Respiratory Syndrome Coronavirus 2</i>
SARS	<i>Severe Acute Respiratory Syndrome</i>
MERS	<i>Middle East Respiratory Syndrome</i>

SUMÁRIO

1. INTRODUÇÃO.....	25
1.1. Definição do Problema	25
1.2. Objetivo do Trabalho	26
1.3. Estrutura do Trabalho	26
2. REVISÃO BIBLIOGRÁFICA	28
2.1. <i>Panic Buying</i>	29
2.2. <i>Panic Buying</i> em epidemias.....	30
2.3. Comportamento das pessoas nas redes sociais em períodos de epidemia	31
2.4. Coleta e classificação de <i>tweets</i>	32
3. METODOLOGIA.....	33
3.1. Materiais	33
3.2. Metodologia do Projeto.....	34
3.2.2. Coleta de dados de mídia social e ferramenta de busca web	35
3.2.2.2. Twitter.....	36
3.2.2.2. Classificação dos tweets	36
3.2.2.2.1. Papel Higiênico	37
3.2.2.2.2. Sabão.....	38
3.2.2.2.3. Álcool.....	38
3.2.2.2.4. Massas.....	38
3.2.2.2.5. Doces.....	38
3.2.3. Teste de hipóteses	39
3.2.4. Cálculo de Correlação.....	39
3.2.4.1. Gráficos de dispersão	39
3.2.4.2. Correlação de Pearson.....	40
3.2.4.3. Correlação com atrasos (<i>lag</i>)	40

3.2.4.4. Correlação entre produtos	40
4. RESULTADOS E DISCUSSÕES	42
4.1. Papel Higiênico	42
4.2. Sabão e sabonete	46
4.3. Álcool	51
4.4. Massas	55
4.5. Doces	59
4.6. Análise de correlação	63
5. CONCLUSÃO	70
REFERÊNCIAS BIBLIOGRÁFICAS	73
ANEXO A.....	77

1. INTRODUÇÃO

1.1. Definição do Problema

No final de dezembro de 2019 foram notificados os primeiros casos de COVID-19 em Wuhan na China, uma doença respiratória aguda causada pelo coronavírus da síndrome respiratória aguda grave 2 (SARS-CoV-2), onde cerca de 5% dos infectados chegam a precisar de aparelhos respiratórios (MINISTÉRIO DA SAÚDE, 2020). Apesar dos esforços para limitar seu espalhamento, o coronavírus se mostrou altamente contagioso, alastrando-se rapidamente por todo globo, sendo declarado uma pandemia pela Organização Mundial da Saúde (OMS) no dia 11 de março (WHO, 2020). No Brasil, o primeiro caso confirmado da doença se deu no dia 26 de fevereiro de 2020, em um homem de 61 anos que voltou de viagem da Itália (Aquino, 2020).

O medo pela contaminação da doença, que ainda não possui cura ou vacina disponível, causou preocupação na comunidade médica, nos governos e na população em geral. Muitos países adotaram a quarentena, paralisando escolas, universidades e atividades econômicas não essenciais, limitando a movimentação nas cidades e tentando barrar o contágio da doença.

O pânico advindo da pandemia fez com que ocorresse um fenômeno conhecido como *panic buying*, que é o ato de comprar grandes quantidades de algum produto devido ao medo de escassez ou aumento de preço. Esse fenômeno é comum em tempos de desastres sociais, ambientais ou crises financeiras, onde alguns produtos básicos são comprados em largas quantidades, acabando os estoques dos mercados ou deixando-os à níveis muito abaixo do normal, devido às incertezas sobre o futuro (Yuen et al, 2020). Os produtos analisados neste trabalho foram: papel higiênico, sabão, álcool, massas e doces. Esses produtos, com exceção do doce que serviu como grupo controle, sofreram de *panic buying*, fazendo os volumes de vendas crescer exponencialmente no período entre o início e o pico da pandemia.

O presente trabalho propõe buscar analisar como é o comportamento das pessoas nas redes sociais em relação aos produtos que sofreram de *panic buying*. O foco do trabalho se dará na rede social Twitter, uma rede social e *microblog* americano onde os usuários podem postar e interagir com outros usuários por meio de mensagens, conhecidas como *tweets*. A análise principal buscará encontrar se existe alguma correlação entre o volume de vendas e os *tweets* em relação a diversos produtos, sobretudo no período em que há *panic buying*.

1.2. Objetivo do Trabalho

O objetivo do presente trabalho é verificar, por meios de análise estatística, a existência ou não de correlação entre alguns produtos que sofreram *panic buying* e o comportamento das pessoas no *Twitter* e no *Google Trends* em relação a esses produtos no período de 01 de março de 2020 até 30 de abril de 2020, intervalo que corresponde desde antes do período em que houve o fenômeno até 1 mês depois do pico de vendas quando os níveis de vendas caíram e estabilizaram novamente. A revisão da literatura provê as ferramentas necessárias para a coleta eficaz dos dados utilizados no estudo e também dos métodos estatísticos necessários para comprovar ou não a existência dessa relação. Os objetivos mais detalhados são:

Primeiramente, é necessário encontrar critérios que determinem quando um produto sofreu de *panic buying* analisando seu volume de vendas. O volume de vendas foi adquirido com uma empresa de varejo, que para garantir seu anonimato será chamada de Empresa X no decorrer desse trabalho. O objetivo dessa fase é selecionar os principais produtos que serão analisados bem como a seleção de algum produto em que não houve *panic buying* para servir como grupo controle.

Em seguida, é necessário realizar a coleta de dados referentes ao *Twitter* e ao *Google Trends* referentes aos produtos selecionados pela análise no período em que houve *panic buying*. Os dados do *Twitter* necessitam de tratamento, uma vez que se deseja somente os *tweets* relacionados a *panic buying* para uma análise mais precisa, filtrando os *tweets* dos produtos que não tenham ligação com o tema.

Os dados então serão normalizados para serem mais bem trabalhados e para garantir que os dados de vendas da empresa não sejam revelados. Por fim, é realizada uma análise estatística sobre os dados para se chegar ao objetivo final do trabalho, verificar a existência de alguma correlação entre as vendas do produto e o comportamento das pessoas no *Twitter* e *Google Trends*.

1.3. Estrutura do Trabalho

O trabalho está estruturado de acordo com outros trabalhos de formatura do departamento, seguindo uma lógica que busca desenvolver um trabalho de forma correta e eficiente.

O primeiro capítulo busca contextualizar e definir o problema que se deseja resolver, além de mostrar as principais motivações e de como o presente trabalho será estruturado.

O capítulo 2 é onde a revisão bibliográfica é feita, fornecendo estudos que contenham temas semelhantes e que auxiliem a nortear os melhores métodos para o desenvolvimento do

projeto. Os principais temas buscados foram aqueles que eram relacionados a: *panic buying*, *panic buying* durante epidemias; Comportamento das pessoas no *Twitter* durante desastres e outras crises; Coleta e classificação de *Tweets*.

O capítulo 3 é onde ocorre o desenvolvimento do trabalho, em que será aplicado os conhecimentos adquiridos com a revisão bibliográfica, além de outros conhecimentos desenvolvidos durante a graduação. Nele será detalhado como foi definido a escolha dos produtos a serem analisados, como a coleta dos dados do *Twitter* será feita, bem como o tratamento desses dados e também a correlação dos dados.

O capítulo 4 é onde será feita a análise e interpretação dos resultados que foram obtidos no capítulo 3, onde será feita a determinação se existe correlação ou não para os produtos, ou se apenas alguns produtos possuem essa correlação.

O capítulo 5 apresenta a conclusão e considerações finais do trabalho, detalhando o aprendizado obtido durante seu desenvolvimento, bem como outras oportunidades de estudos derivadas desse.

O capítulo 6 constitui a referência bibliográfica, onde é mostrado todas as fontes de estudo utilizadas durante o desenvolvimento do trabalho.

Esse primeiro capítulo, como dito na estrutura do trabalho, serviu para contextualizar e definir o problema. Além disso, também foi discutido como é desenvolvido o trabalho, mostrando as principais etapas definidas, que estão detalhadas adiante.

2. REVISÃO BIBLIOGRÁFICA

Neste capítulo será feita a revisão bibliográfica, onde será encontrado artigos e outras publicações que auxiliem a se encontrar métodos e ferramentas que serão usados neste trabalho.

Buscam-se por estudos que discutem temas relacionados a *panic buying* e seus efeitos durante epidemias para conseguir uma melhor compreensão das possíveis causas desse fenômeno em períodos de desastres. Em seguida é feita uma pesquisa em relação ao comportamento das pessoas no *Twitter* em tempos de crise, bem como é feita a coleta e classificação dos *tweets*. Por fim, busca-se estudos que busquem alguma correlação entre *tweets* e fenômenos sociais para que se tenha uma metodologia eficaz para o cálculo no problema proposto.

A busca por literatura referente a *panic buying* mostra que a pesquisa no tema ainda pode ser muito ampliada dada a pouca quantidade de artigos encontrados sobre o assunto. Uma busca na plataforma SCOPUS pelo termo "*panic buying*" retornou 31 artigos (acesso em outubro de 2020) onde 26 deles eram referentes a pandemia do COVID-19. Dada a importância desse fenômeno nas cadeias logísticas, existem várias possibilidades de pesquisa sobre o assunto. Além do Scopus foram também buscados artigos nas plataformas *Web of Science* e Google Scholar para complementar pesquisas sobre o tema. No total, 103 publicações envolvendo diversos temas relacionados a *panic buying*, COVID-19, desastres naturais e comportamento dos consumidores foram selecionadas para uma análise inicial. Nenhum dos artigos coletados mostrou alguma pesquisa semelhante ao proposto neste trabalho, correlacionando os dados de mídias sociais com a venda de produtos sob efeito de *panic buying*, mostrando que o tema desenvolvido é original. Dessa análise inicial, 20 foram selecionados para nortear o projeto a ser desenvolvido, a tabela 1 mostra quais foram os artigos selecionados e o tema trabalhado neles.

Tabela 1: quadro-resumo sobre os artigos selecionados para revisão bibliográfica.

Publicação	<i>Panic Buying</i>	<i>Panic Buying</i> em epidemias	Comportamento das pessoas nas redes sociais em épocas de desastres	Coleta e classificação de <i>tweets</i>
Leach, 1994	✓			
Zheng et al. 2020	✓			
Wesseler 2020	✓	✓	✓	
Hendricks et al 2005	✓			
Laato et al. 2020	✓	✓		
Upton et al, 2014	✓			
Wang et al, 2014	✓			
Hobs et al, 2020	✓	✓		

Khare et al, 2020	✓		✓	
Sim et al. 2020	✓	✓		
Grashuis et al, 2020	✓	✓		
Hassen et al, 2020	✓	✓		
Aslam et al., 2018a	✓	✓	✓	
Springer et al. 2020	✓	✓	✓	
Parment et al, 2013	✓	✓	✓	
Naeem et al, 2020	✓	✓	✓	
Arafat et al, 2020	✓	✓	✓	
Abd-Alrazaq et al, 2020			✓	✓
Zhou et al. 2015				✓
Byun et al. 2013				✓
Ahmed et al, 2018			✓	
Chew et al, 2010			✓	

Fonte: elaborado pelo autor.

2.1. *Panic Buying*

John Leach (1994) em seu livro *Survival Psychology* diz que em grandes desastres naturais ou em epidemias as pessoas podem ter mudanças de comportamento, dentre eles, o *panic buying*, ou seja, quando eles compram grandes quantidades anormais de alguns produtos. Isso acontece pois em momentos como esses, ocorre um aumento no estresse e ansiedade das pessoas e a incerteza sobre o futuro faz as pessoas estocarem alguns produtos como forma de antecipar futuras falta de estoque ou aumento de preços.

O *panic buying* pode muitas vezes ser socialmente induzido. As pessoas observam a intensidade no comportamento de pânico das outras pessoas e quando a percepção dessa intensidade ultrapassa certo limite, elas são induzidas a ter um comportamento de *panic buying*, aumentando a demanda total dos produtos (Zheng et al. 2020).

O comportamento de *panic buying* é indesejado uma vez que essa situação limita que pessoas de grupos mais vulneráveis tenham acesso a alguns produtos, muitas vezes essenciais como alimentos ou produtos de higiene básica (Wesseler 2020). Para contornar essa situação, muitos governos de locais em que há uma relativa alta ocorrência de desastres naturais como terremotos e alagamentos, adotam políticas de estocagem para amenizar a volatilidade do preço e com isso proporcionar maior segurança alimentícia para a população. Dessa forma, quando ocorre algum desastre, os estoques de emergência ajudam a reduzir os impactos negativos gerados pelo *panic buying*, especialmente para as famílias com menor poder de compra (Wesseler 2020).

Além disso, o *panic buying* pode causar rupturas nas cadeias logísticas de diversos produtos, devido a compra excessiva de alguns produtos. A empresa X, por exemplo, apresentou falta de estoque de álcool durante o pico de busca durante a pandemia do COVID-19. Estima-se que uma quebra na cadeia de suprimentos resulta em perdas de até 40% nas empresas e os efeitos negativos advindos desse fenômeno podem ser sentidos até um ano depois da quebra (Hendricks and Singhal 2005).

Estudos envolvendo simulações foram feitos para analisar *panic buying*. Upton et al (2014) utilizaram simulação baseada em agentes para estudar a compra de combustíveis no Reino Unido durante dois períodos de ruptura no país, em setembro de 2000 e março de 2012. Com a simulação, eles conseguiram identificar aspectos das cadeias de suprimentos de combustíveis durante esses dois períodos e com isso fazer recomendações para tomada de decisões durante períodos de *panic buying*. O comportamento de difusão do pânico nas redes sociais também foi simulado em um estudo de Wang et al (2014), onde foi aplicado um modelo de simulação para analisar o comportamento das pessoas de acordo com as informações anunciadas pelo governo.

A falta de combustível é comum em épocas de desastres naturais, como furacões (Khare et al. 2020). No estudo de Khare et al (2020), as mídias sociais foram utilizadas para prever a demanda de gasolina, onde eles desenvolveram uma abordagem que examina o conteúdo de *tweets* relacionados a gasolina e os utilizam em um modelo de previsão de demanda. Eles aplicaram o modelo em um estudo de caso na Flórida e conseguiram obter resultados precisos.

2.2. *Panic Buying* em epidemias

Como dito anteriormente, uma das possíveis causas de *panic buying* são epidemias, onde as pessoas buscam estocar alguns produtos como forma de antecipar futuros períodos de escassez ou de aumento de preços. Isso aconteceu em março de 2020, com a pandemia do coronavírus. A causa de *panic buying* não foi somente o risco à saúde, mas também o medo da doença fazer as fábricas pararem de produzir (Laato et al. 2020).

No século XXI houve diversas epidemias registradas, como Ebola, SARS, MERS, gripe suína, dengue, zika vírus, chikunguya, entre outros. Historicamente, essas epidemias afetam tanto o comportamento de consumidor quanto o comportamento de reduzir riscos à saúde das pessoas (Laato et al. 2020). Durante a epidemia COVID-19 isso ficou em evidência, pois encontrou-se uma forte correlação entre as pessoas terem intenção de se auto isolarem e das pessoas fazerem compras em grandes quantidades (Laato et al. 2020).

Sim e outros autores buscam explicar os prováveis motivos que levam as pessoas a terem esse comportamento de comprar em quantidades anormais (Sim et al. 2020). O primeiro ponto observado é o conflito entre a vontade de manter uma rotina e a incerteza da duração da pandemia, o que leva as pessoas a aumentarem seus níveis de estresse e ansiedade e consequentemente, elas compram mais. O segundo ponto seria um ato de preservar o indivíduo e sua família estocando comida e outros produtos essenciais. O terceiro ponto seria a maior exposição a mídias sociais, que leva as pessoas a imitarem comportamento de seus semelhantes.

Hobs (2020) observou as cadeias de suprimento no Canadá durante a pandemia, onde o sistema de produção é o *just-in-time*, que não é flexível para aumentos repentinos na demanda. Durante os estágios iniciais do período, o governo canadense tentou informar a população de que os estoques eram suficientes para atender toda população. No entanto, isso não impediu o comportamento de *panic buying*, que fez com que muitos produtos passassem por escassez. O estudo descobriu que houve um aumento na tendência do consumo *online* e uma maior priorização por produtos locais.

A pandemia COVID-19 provocou mudanças no padrão de consumo das pessoas. Grashuis et al. (2020) verificou uma correlação entre as preferências dos consumidores e a tendência da pandemia (aumento, queda ou estabilidade de casos). Hassen et al. (2020) fizeram uma análise em Qatar sobre o consumo de alimentos durante a pandemia. Foi descoberto um aumento na tendência por dietas mais saudáveis, produtos locais e compras *online*.

2.3. Comportamento das pessoas nas redes sociais em períodos de epidemia

Durante a pandemia do coronavírus, o comportamento das pessoas nas redes sociais também se alterou. As redes sociais fizeram as pessoas ficarem mais engajadas e conectadas e também aumentaram a disseminação de informação (Aslam et al., 2018a). A atividade nas redes sociais durante esse período também aumentou e a busca por termos relacionados a prevenção e sintomas da doença, bem como a preparação para quarentena cresceu exponencialmente (Springer et al. 2020).

Em outras epidemias e pandemias também se verificou mudanças nos comportamentos das pessoas nas redes sociais nesses períodos. Chew et al (2010) buscou analisar os principais tópicos discutidos no *Twitter* durante a pandemia da gripe suína. Seu estudo descobriu que a rede social foi muito utilizada como forma de disseminação de informações e opiniões sobre a doença. Ahmed et al (2018) também estudou o *Twitter* durante a gripe suína e ampliou a pesquisa para os casos de ebola de 2014. Sua conclusão foi de que o comportamento geral na

plataforma foi de medo e apreensão em relação a doença, levando a um sentimento geral de pânico.

Além disso, as interações sociais provocaram mudanças no padrão de consumo das pessoas (Parment 2013). Durante o COVID-19, uma dessas mudanças foi um maior estímulo de *panic buying* pelo medo à doença. Segundo Naeem (2020), por meio de entrevistas, descobriu que a atividade nas redes sociais influencia o comportamento das pessoas para enfrentar a pandemia, pois elas ficam mais atentas ao que seus amigos e conhecidos fazem para se prevenir. Arafat (2020) mostrou que a publicação de notícias referentes a *panic buying* aumentou consideravelmente durante a pandemia, e os meios midiáticos possuem um papel importante para definir o padrão de consumo das pessoas.

2.4. Coleta e classificação de *tweets*

Como o trabalho a ser desenvolvido exige a análise de *tweets*, buscou-se na literatura por metodologias que auxiliassem a coleta e classificação dos tweets. Abd-Alrazaq et al (2020) mostra uma abordagem eficiente e que serviu como principal base do presente trabalho. Nela, a coleta dos tweets foi feita por meio da *Tweepy Python*, biblioteca para acessar a *Twitter API*. No caso deste trabalho, foi utilizado a biblioteca *Twint*, que é similar a *Tweepy*, no entanto não possui as limitações da *Twitter API*.

Após a coleta, foi feito um pré processamento dos *tweets*, removendo *tweets* de línguas estrangeiras, retweets, pontuações, entre outras coisas. Por fim, os *tweets* foram classificados em tópicos de acordo com o assunto que eles continham, por meio de um filtro baseado em palavras chaves contidas nos *tweets*.

Outros estudos também serviram para apoiar a estruturação da metodologia do trabalho. Uma abordagem útil para classificação dos *tweets* foi feita por (Zhou et al. 2015) de acordo com o texto de cada *tweet*. Segundo o artigo, uma das abordagens é a classificação por palavras binárias, onde se busca filtrar os *tweets* por termos que aparecem com bastante frequência em *tweets* que abordam algum assunto, mas não aparece em *tweets* que não os abordam. (Byun et al. 2013) usa método similar para classificar tweets referentes a comerciais de televisão.

Os estudos trazidos nessa seção serviram de base para construir a metodologia adotada no trabalho.

3. METODOLOGIA

Este capítulo tem como objetivo mostrar como o projeto foi desenvolvido. É mostrada cada etapa do trabalho e quais métodos e hipóteses foram adotados baseados na literatura revisada no capítulo anterior. A metodologia adotada no projeto segue o esquema mostrado na figura 1 e cada etapa será detalhada ao decorrer do capítulo

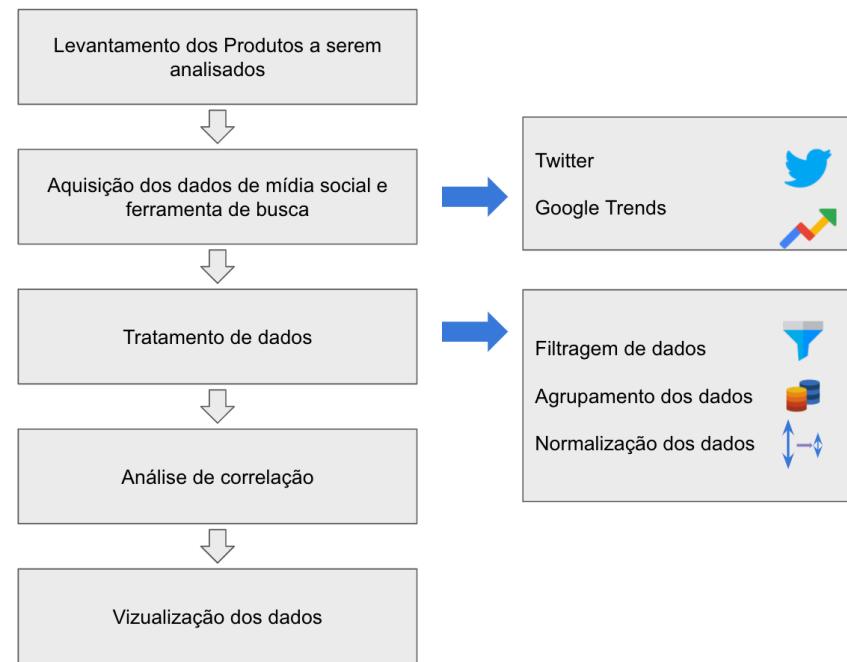


Figura 1: metodologia aplicada no projeto
Fonte: Elaboração do autor.

3.1. Materiais

O trabalho proposto envolve a utilização de dados provenientes da empresa X, do *Twitter* e do *Google Trends*. Os dados de vendas foram disponibilizados pela empresa por meio de um arquivo ".xlsx". Para os dados do *Google*, foi consultado o endereço <https://trends.google.com/>, onde é possível adquirir os dados normalizados referentes a busca por um ou mais termos em um intervalo determinado de tempo. Por fim, os dados referentes ao

Twitter foram coletados por meio de um programa desenvolvido em linguagem de programação Python 3.X

O Python também foi utilizado para tratar e analisar os dados coletados dentro do ambiente Jupyter Notebook, plataforma open-source em web que permite a criação de documentos que contenham códigos de programação (JUPYTER, 2020). Para realizar essas análises, além da linguagem Python, também foram utilizadas as bibliotecas:

- Twint - biblioteca que permite realizar mineração de dados do *Twitter* sem a necessidade da utilização da API de busca do *Twitter* (OSYINT TEAM, 2020).
- Numpy - biblioteca que dá suporte para operar listas e matrizes de múltiplas dimensões junto com diversas funções para serem utilizadas nessas listas (NUMPY, 2020)
- Pandas - biblioteca utilizada para manipulação e análise de dados (PANDAS, 2020).
- Matplotlib - biblioteca que permite a visualização de dados de forma estática, animada ou interativas dentro do ambiente Python (MATPLOTLIB, 2020).

A máquina utilizada para realizar a análise foi um MacBook Pro de 13 polegadas de 128 gb de armazenamento, versão de 2019 com sistema operacional Mac OS 10 - Mojave.

3.2. Metodologia do Projeto

3.2.1. Produtos a serem analisados

Para decidir quais seriam os produtos que seriam analisados foi necessário primeiro identificar aqueles que sofreram ou não de *panic buying*. Para descobrir isso coletou-se os dados de vendas nacionais da empresa X nos meses de fevereiro, março e abril, período em percebeu-se *panic buying* de alguns produtos, como o papel higiênico por exemplo. Com relação ao contexto geográfico, os dados foram analisados a nível Brasil, uma vez que não há dados confiáveis em relação à localização precisa dos *tweets*. Conforme foi verificado pelo autor, em muitos casos só existe a informação da língua em que o *tweet* foi escrito e a informação da localização não é necessariamente verdadeira, pois os próprios usuários podem defini-la: em um dos *tweets*, *por exemplo*, a localização estava escrita como “Terra do Nunca” em alusão a obra infantil do Peter Pan, e vários casos similares foram encontrados.

Foi feita uma análise gráfica de cada produto e aqueles que apresentavam um pico anormal sem que houvesse eventos especiais, como promoção ou feriado, foram identificados

como *panic buying*. Após essa avaliação, alguns desses produtos foram selecionados para serem analisados, sendo eles:

- Papel Higiênico: Motivo inicial do TF de maior interesse da empresa X;
- Sabão: Outro produto de interesse da empresa que também sofreu *panic buying*;
- Álcool: Produto que houve quebra de estoque;
- Massas secas: Produto em que se observou *panic buying* mesmo não sendo do segmento de higiene;
- Doces: Produto que não ocorreu *panic buying* mas teve um pico devido a sazonalidade da Páscoa.

A análise dos produtos foi feita com os dados de vendas do produto em valores percentuais em relação ao pico máximo de vendas para garantir o anonimato dos dados da empresa X.

3.2.2. Coleta de dados de mídia social e ferramenta de busca web

Foi realizada a coleta dos dados necessários para uma correta análise. Os dados referentes aos volumes de vendas de cada produto foram passados pela empresa X. Para se coletar os dados do *Google Trends* e do *Twitter*, por sua vez, foi necessário buscar ferramentas que permitissem essa coleta.

3.2.2.1. *Google Trends*

Os dados provenientes do *Google Trends* vieram da própria plataforma *Google*, por meio do endereço de web da ferramenta (GOOGLE TRENDS, 2020). Na plataforma, é necessário digitar no campo de busca os termos que se tem interesse. Em seguida, pode-se alterar algumas configurações da busca, como o país e intervalo de tempo. A plataforma apresenta os resultados de forma normalizada, onde o dia em que houve mais acessos apresenta valor 100, e os outros dias apresentam valores proporcionais de 0 a 100.

Para cada um dos produtos selecionados para análise foi feita uma busca no endereço, colocando o período de tempo entre 01 de março e 30 de abril e Brasil como local. Além disso, os dados foram exportados para um arquivo "csv" para ser possível fazer uma comparação com os outros dados. Dessa forma, é possível realizar uma comparação entre o volume de vendas e a busca dos produtos no Google.

Os dados provenientes do *Google Trends* possuem uma limitação, que consiste em não ser possível saber o contexto em que cada busca foi feita, ou seja, não é possível saber se a

pessoa que pesquisou o produto tinha intenção de comprá-lo ou se pesquisou por algum outro motivo não relacionado com compra.

3.2.2.2. Twitter

Os dados do *Twitter* foram coletados por meio do desenvolvimento de um programa em Python que coleta todos os *tweets* de um período de acordo com as palavras chaves definidas. Para isso, utilizou-se da biblioteca *Twint*, que possui funções que auxiliam na busca por tweets.

O programa desenvolvido extrai os *tweets* de acordo com os parâmetros definidos pelo usuário, são eles:

- *Search*: conjunto de palavras chave que se irá buscar. Nesse parâmetro também foi definido Twitter em língua portuguesa adicionando "lang:pt" ao final do conjunto. Para cada produto utilizou-se os mesmos termos adotados para a coleta dos dados do *Trends* para manter a consistência da análise;
- *Since*: Data mais antiga que se deseja buscar;
- *Until*: Data mais recente que se deseja buscar;
- *Output*: Nome do arquivo onde será armazenado os *tweets*

Para cada *tweet* foram coletados os dados de publicação do *tweet*, fuso horário, local de envio, conteúdo do *tweet*, informação se o *tweet* se trata ou não de um *retweet* (quando um usuário publica o *tweet* de um outro usuário). Inicialmente era ideal que fosse filtrado somente os *tweets* realizados no Brasil, no entanto poucos deles possuem a informação geográfica do usuário, fazendo com que a base de dados ficasse reduzida. Desse modo, optou-se por filtrar os tweets feitos em língua portuguesa e que possuam fuso horário do Brasil (-3 GMT). O código produzido para a coleta está apresentado no anexo.

Assim como no caso do *Google Trends*, nem todos os tweets são relacionados a compras, fazendo com que os dados tenham bastante ruídos. No entanto, para o *Twitter*, é possível saber o conteúdo do *tweet* e com base nele, pode-se classificar se ele está relacionado a *panic buying*. Essa classificação será mais detalhada na próxima seção.

3.2.2.2. Classificação dos tweets

Como dito anteriormente, os tweets precisam ser filtrados para serem selecionados somente aqueles que se referem a *panic buying*. A revisão bibliográfica nos indicou uma forma

eficaz de se realizar essa classificação. Utilizamos a abordagem de palavras binárias (Zhou et al. 2015) para fazer a classificação de cada tweet.

Para cada grupo de produto, foi lida uma amostra de 200 tweets e foi feita uma classificação manual de cada um deles em “*panic buying*” e “não *panic buying*”. Em seguida, buscou dentre os tweets de *panic buying* por termos que fossem exclusivos ou quase exclusivos desse tipo de tweet. Por fim, aplicou-se o filtro em toda base de dados. A figura 3 representa a metodologia adotada.

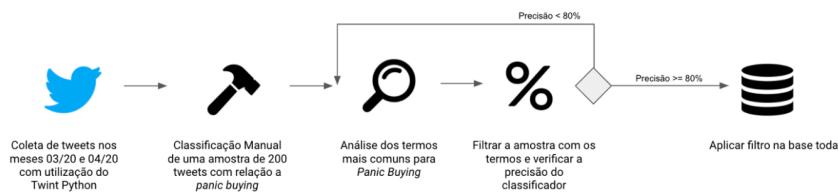


Figura 2: metodologia adotada para classificação dos tweets.

Fonte: elaborada pelo autor.

3.2.2.2.1. Papel Higiênico

A classificação dos tweets de papel higiênico começou retirando uma amostra de 200 tweets da base. Pelo intervalo de tempo coletado, a maioria dos tweets encontrados, cerca de 75% foram relacionados a *panic buying*, como o exemplo a seguir: “E, na moral, pra que tanto papel higiênico? Vão soterrar o corona vírus com papel de folha dupla?”.

Outros tweets são referentes a diversos outros assuntos, um que merece destaque é o desafio do papel higiênico que se popularizou nas redes sociais, como mostra o exemplo: “Jogar garrafa d’água ano passado e fazer embaladinhas com papel higiênico numa pandemia. O ser humano é surpreendente quando se trata de combater o tédio.”

A análise dos tweets mostrou as seguintes palavras chaves após sucessivos refinamentos: 'estoc', 'compr', 'busca', 'armazen', 'mercado', 'super', 'estoq', 'corrida', 'armaz', 'escass', 'louc', 'esgot', 'acab', 'esvazi', 'atras', 'falt', 'neura', 'doido', 'sem', 'consum', 'caga', 'diarrei', 'hister', 'carrin'. Para palavras derivadas de verbos buscou-se não colocar a terminação para englobar vários casos, por exemplo, “estoc” pode se referir aos termos “estocar”, “estocando”, “estocaram”, etc.

O classificador mostrou uma precisão de 80%. A revisão literária mostrou que uma conversão dessa ordem é considerada aceitável (Byun et al. 2013), uma vez que, por se tratar de dados sociais, eles contém bastante ruído, sendo necessário muito esforço para se atingir níveis maiores que 90%.

3.2.2.2.2. Sabão

A classificação do sabão apresentou melhores resultados uma vez que poucos filtros foram necessários para conseguir atingir uma precisão ainda maior que no caso do papel higiênico. As palavras utilizadas foram: 'estic', 'estoc', 'estoaq', 'preço', 'farmácia', 'mercado', resultando em uma precisão de 98%. Dentro da amostra havia 2,5% de tweets referentes a *panic buying*, o que mostra uma pouca quantidade de tweets quando comparado com papel higiênico.

3.2.2.2.3. Álcool

A classificação de álcool mostrou um comportamento similar ao de sabão, onde somente 8% da amostra eram referentes a *panic buying*. A classificador encontrado contém as palavras-chave: 'estoc', 'estoaq', 'feira', 'farmácia', 'comprar', 'vender' e 'mercado' e resultou em uma precisão de 95%.

3.2.2.2.4. Massas

A classificação de massas também apresentou baixa quantidade de *tweets* relacionados a *panic buying*, 1,5% da amostra como o exemplo: “Por aqui se houver realmente uma epidemia tb será a mesma coisa. Já inclusive te aconselho a começar o estoque de miojo!”. O classificador continha as palavras: 'estoc', 'estoaq', 'feira', 'mercado', 'comprar', 'armaze' e obteve uma precisão de 98,5%.

3.2.2.2.5. Doces

Com a classificação dos doces além dos termos relacionados à *panic buying* também foram excluídos *tweets* que continham as palavras: 'pascoa', 'páscoa' e 'ovo' como forma de tentar remover os *tweets* referentes à pascoa. A amostra adquirida encontrou apenas um *tweet* referente a *panic buying*, sendo ele: “Hoje fui no mercado pra fazer estoque pra quarentena, só comprei biscoito e chocolate”. Como um *tweet* não é representativo, foi adotado os termos: 'estoaq', 'estoc' e 'compra'. A verificação do filtro trouxe resultados em sua maioria relacionados a *panic buying*, como por exemplo: “enquanto uns fazem estoque de papel

higiênico, meu pai fez estoque de chocolate kkkkk ele na existe” e “Tive que fazer um estoque de doces p não surtar nessa quarentena”.

3.2.3. Teste de hipóteses

Para uma correta análise de correlação, é preciso definir um teste de hipótese que será usado para medir a existência ou não de alguma correlação entre Twitter, Trends e Vendas. O teste de hipótese utilizado neste trabalho é o teste de correlação, tal como descrito Costa Neto (2002):

$$H_0: \rho = 0$$

$$H_1: \rho \neq 0$$

Onde ρ indica a correlação de Pearson (1985). O teste de correlação pode ser verificado por meio de:

$$t_{n-2} = r \sqrt{\frac{n-2}{1-r^2}}$$

Onde r é a correlação encontrada e t será comparado ao t de Student com $n - 2$ graus de liberdade.

A hipótese nula (H_0) corresponde a hipótese de que não existe correlação entre os dados. Caso o módulo do valor calculado para t_{n-2} seja maior que o valor de $t_{n-2,\alpha}$ ($|t_{n-2}| > t_{n-2,\alpha}$), podemos rejeitar a hipótese nula com uma confiança de $(1 - \alpha)$ e dizer que existe correlação entre os dados. Para os testes desse trabalho foi escolhido um $\alpha = 1\%$ de forma a ter resultados efetivamente significativos.

3.2.4. Cálculo de Correlação

3.2.4.1. Gráficos de dispersão

Antes de realizar os cálculos de correlação, primeiro visualizamos os gráficos de dispersão entre os diferentes produtos e tipos de dados. Os gráficos x, x e x mostram a dispersão entre cada combinação de dados (Twitter x Trends, Twitter x Vendas, Trends x Vendas). Para uma melhor visualização, todos os dados foram normalizados para valores entre 0 e 1, pela fórmula $a_i = (x_i - x_{min}) / (x_{max} - x_{min})$, onde $x_i, i = 0, 1, 2, 3, \dots$ são os valores originais dos dados e $a_i, i = 0, 1, 2, 3, \dots$ são os valores convertidos. Além disso, os pontos foram agrupados

de acordo com o período de tempo. Pelo fato de os doces serem fortemente influenciados pela páscoa, eles tiveram a seguindo classificação:

- Pré-corona: Período que antecede o momento de pico de *panic buying*;
- Corona: Período em que há a maior ocorrência de *panic buying*;
- Pós-corona: Período posterior ao pico de *panic buying*;
- Páscoa (somente para doces): Período de Páscoa;
- Pós-páscoa (somente para doces): Período posterior a páscoa;

Os demais produtos seguiram a seguinte classificação:

- Antes: Período que antecede o momento de pico de *panic buying*;
- Durante: Período em que há a maior ocorrência de *panic buying*;
- Depois: Período posterior ao pico de *panic buying*;

3.2.4.2. Correlação de Pearson

A verificação de correlação entre os dados foi feita seguindo a fórmula da correlação de Pearson par o intervalo entre 01/03 até 30/04, pois o interesse do trabalho é o período de que engloba o pico de *panic buying*. A correlação de Pearson indica o grau de correlação, sendo que valores com módulo próximos a 1 indicam alta correlação entre os dados. O cálculo de correlação é feito seguindo a fórmula descrita em Costa Neto (2002):

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2} \sqrt{\sum(y_i - \bar{y})^2}}$$

3.2.4.3. Correlação com atrasos (*lag*)

O trabalho desenvolvido também buscou analisar se o efeito de alguma mídia social poderia ter provocado o aumento de vendas em dias posteriores. Dessa forma foi feita uma nova análise correlacionando os valores do Twitter/Trends com os valores de vendas n dias depois

3.2.4.4. Correlação entre produtos

Uma última análise foi feita buscando encontrar alguma correlação entre produtos. Nessa análise, para cada tipo de dado – Vendas, Twitter e Trends – foi comparado produto a produto a correlação entre os dados deles. Para isso criou-se uma matriz de correlação para cada tipo de dado de forma a poder realizar uma análise mais precisa da existência ou não de alguma correlação.

Nesse capítulo foi mostrado como foi realizado o desenvolvimento do trabalho durante o decorrer da disciplina, desde a coleta e tratamento dos dados até as análises estatísticas feitas. O próximo capítulo tem como objetivo analisar os resultados que foram obtidos e o que se pode concluir a partir deles.

4. RESULTADOS E DISCUSSÕES

Este capítulo tem como objetivo mostrar os resultados obtidos por meio da metodologia adotada, bem como a discussão de seus resultados. Inicialmente é feita a análise dos resultados de cada produto separadamente e, em seguida é feita os cálculos de correlação descritos no capítulo anterior bem como as principais conclusões que foram obtidas.

4.1. Papel Higiênico

O primeiro produto analisado foi o papel higiênico, um dos produtos que esteve em destaque nos noticiários, onde foram reportados diversos casos de *panic buying*. A figura 4 mostra o histórico de vendas de papel higiênico durante o período. Nota-se um pico nos dias 06 e 07 de março devido a uma promoção de começo de mês que houve no dia. Além disso, o intervalo de 14/03 até 21/03 apresentou um aumento de mais de 50% nas vendas sem que houve promoção, esse é o período identificado como de *panic buying* e onde será feita a análise do produto.

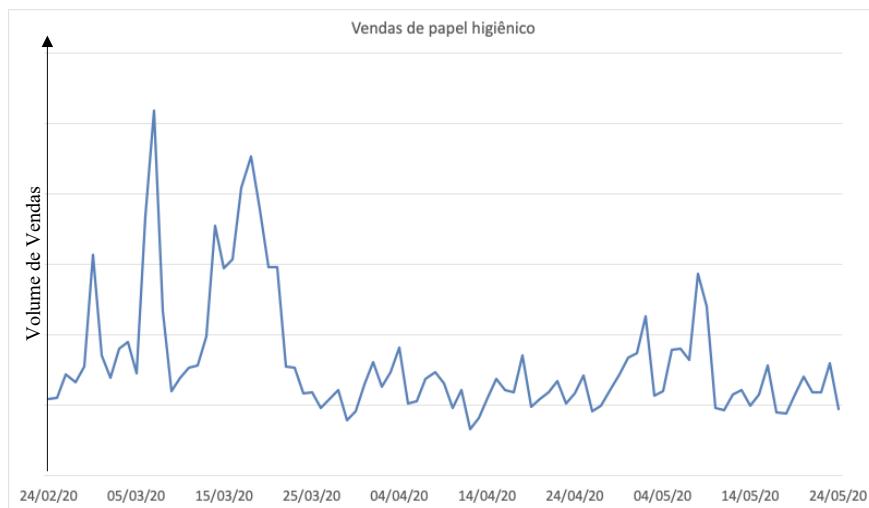


Figura 3: volume de vendas de papel higiênico no período.
Fonte: elaborada pelo autor.

Os dados do Google Trends foram adquiridos realizando uma busca pelo termo ("papel higienico" + "papel higiênico") na plataforma da ferramenta. O operador "+" tem a função de "ou", fazendo com que a busca retorne as pesquisas que contenham pelo menos um dos dois termos entre aspas. A figura 5 mostra o gráfico de buscas pelo termo no período analisado.

Podemos observar que ao desconsiderar a promoção que ocorreu no começo de março, as duas curvas parecem ter comportamento semelhante, o que pode sinalizar uma correlação entre as Vendas e Trends para papel higiênico.

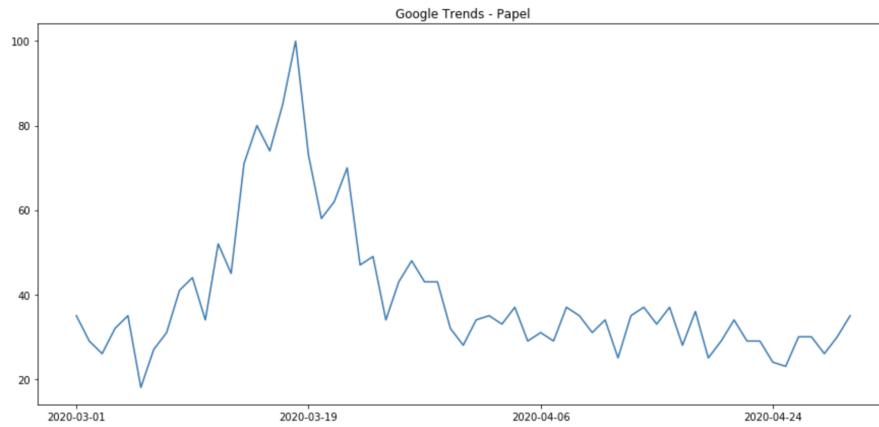


Figura 4: pesquisa por papel higiênico no *Google Trends* durante o período.

Fonte: elaborada pelo autor.

Para os dados referentes ao Twitter, o gráfico da figura 6 nos mostra os *tweets* relacionados a papel higiênico no período, desconsiderando o filtro realizado. Uma primeira análise mostra que existe um pico entre os dias 21 e 23 de março que não é similar ao que ocorre com as Vendas.

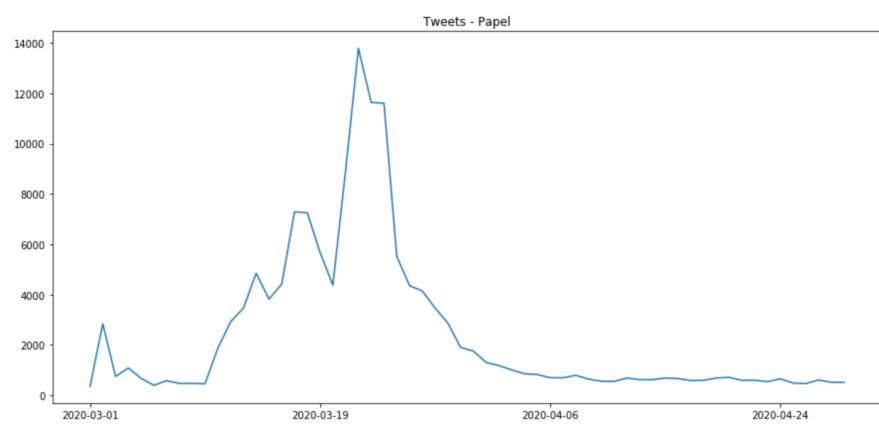


Figura 5: *tweets* de papel higiênico sem filtro durante o período.

Fonte: elaborada pelo autor.

A aplicação do filtro, mostrado na figura 7, vemos que o formato da curva teve uma grande alteração, com isso, as curvas de vendas e tweets possuem um formato mais parecido. Além disso, o filtro reduziu de maneira significativa a quantidade de *tweets*, onde antes do filtro havia um pico de aproximadamente 14 mil *tweets* de papel higiênico e depois do filtro o pico de *tweets* foi próximo de 4 mil.

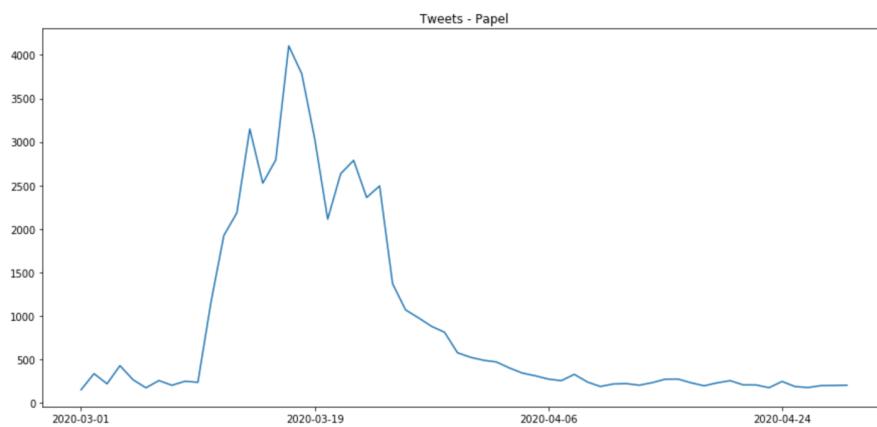


Figura 6: *tweets* de papel higiênico com filtro durante o período.
Fonte: elaborada pelo autor.

Para uma melhor análise, foram feitos gráficos de dispersão com os dados normalizados para valores entre 1 e 0 referentes a: vendas; Google Trends e Twitter filtrado. O primeiro desses gráficos é a comparação entre *tweets* e vendas. A figura 8 mostra que os picos de vendas correspondem com os picos de *tweets* (marcadores laranja). Os pontos destacados se referem ao período que houve a promoção (dias 06 e 07 de março) e, portanto, serão descartados para o cálculo de correlação.

Commented [HY1]: Vc não identificou no gráfico quais são os pontos destacados. Faça isso.

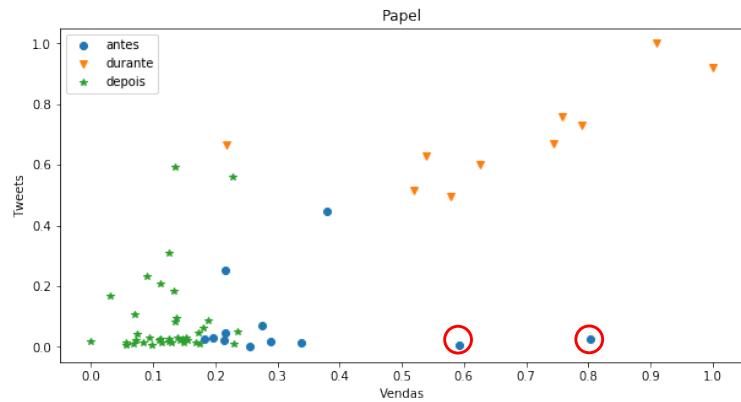


Figura 7: gráfico de disperso do papel higiênico entre *tweets* e vendas.
Fonte: elaborada pelo autor.

O gráfico de dispersão entre Google Trends e Twitter (figura 9) também mostra uma aparente forte correlação entre os dados provenientes das duas bases, onde os picos se correspondem.

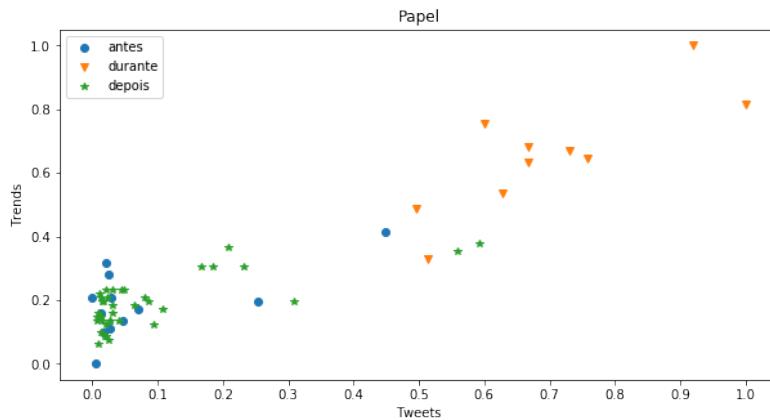


Figura 8: gráfico de disperso do papel higiênico entre *trends* e *tweets*.
Fonte: elaborada pelo autor.

O último gráfico de dispersão do papel higiênico é em relação ao Google Trends e às vendas, figura 10 mostra a relação entre ambos. Assim como na comparação entre Twitter e Vendas, os dados de Trends e vendas aparentam possuir uma boa correlação, excluindo-se os pontos em houve a promoção.

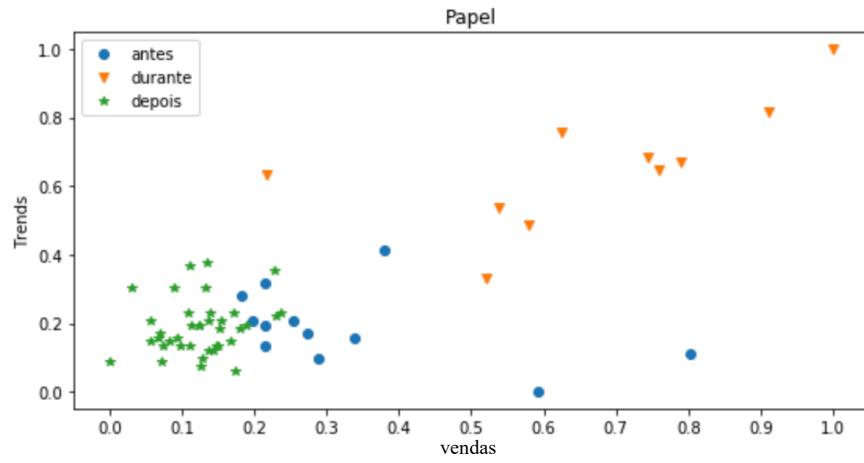


Figura 9: gráfico de disperso do papel higiênico *entre trends e vendas*.
Fonte: elaborada pelo autor.

4.2. Sabão e sabonete

O sabão foi outro dos produtos escolhidos por apresentar comportamento de *panic buying* num período próximo ao do papel higiênico. Assim como no caso do papel higiênico, também houve uma promoção nos dias 06 e 07 de março, o que fez as vendas aumentarem significativamente nesse dia. O período identificado de *panic buying* deste produto se deu no período entre 14/03 até 24/03, como mostrado na figura 11.



Figura 10: volume de vendas de sabão no período.
Fonte: elaborada pelo autor.

Para encontrar os dados de sabão foi feita uma busca pelo termo ("sabão" + "sabonete" + "sabao"). A figura 12 mostra o gráfico relativo ao *Google Trends* de sabão. Ao contrário do que se observou com o papel higiênico, nota-se uma diferença entre os dados vindos do Google Trends em relação aos dados de vendas, uma vez que após o pico, o volume de vendas diminui significamente, enquanto no Google Trends ele permanece alto.

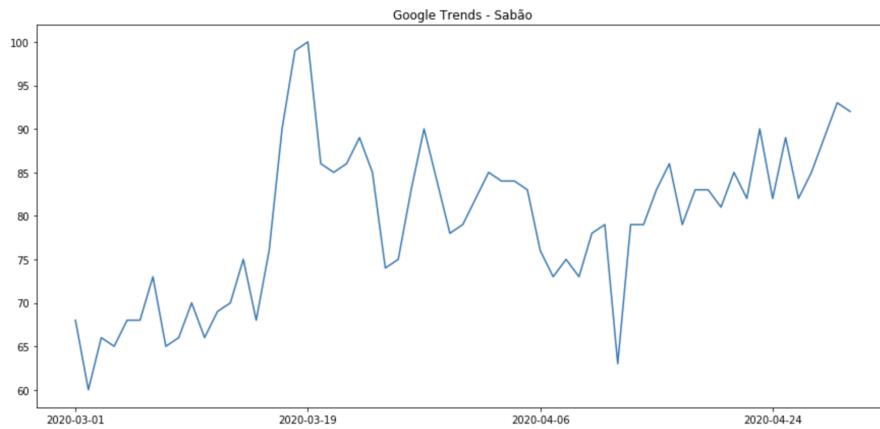


Figura 11: pesquisa por sabão no *Google Trends* durante o período.
Fonte: elaborada pelo autor.

Para os dados referentes ao Twitter, o gráfico da figura 13 mostra os *tweets* relacionados a sabão no período no período, desconsiderando o filtro realizado. Mesmo sem a aplicação do filtro, o formato da curva não está muito diferente do formato da curva de vendas.

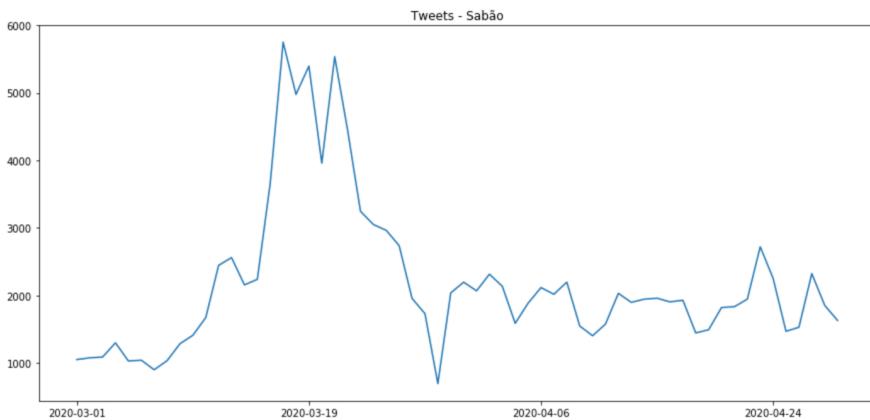


Figura 12: *tweets* de papel higiênico sem filtro durante o período.

Fonte: elaborada pelo autor.

Ao aplicar o filtro, como mostra a figura 14, vemos que o formato da curva teve uma leve alteração no formato, fazendo as curvas de vendas e *tweets* possuírem um formato mais parecido. Além disso, o filtro reduziu de maneira significativa a quantidade de *tweets*, antes do filtro havia um pico de aproximadamente 6 mil *tweets* de sabão e depois do filtro o pico de *tweets* foi próximo de 400.

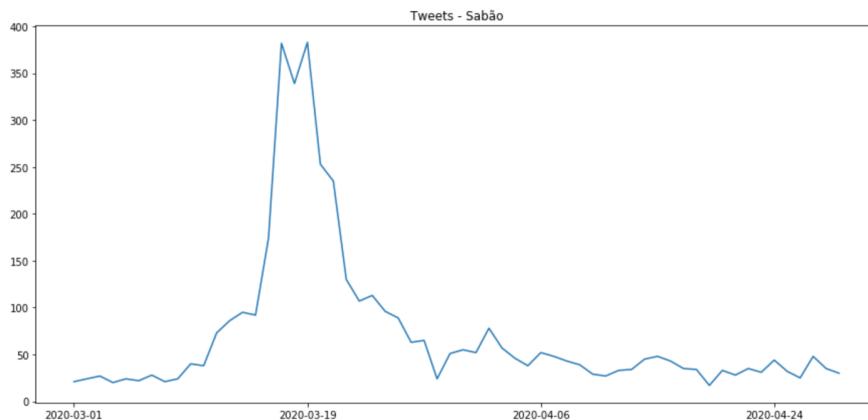


Figura 13: *tweets* de papel higiênico com filtro durante o período.
Fonte: elaborada pelo autor.

O gráfico de dispersão entre Twitter e as vendas (figura 15), mostram um comportamento similar ao do papel higiênico, onde os picos de ambos os dados são coincidentes, havendo dois pontos *outliers* referentes às promoções que ocorreram no começo de março.

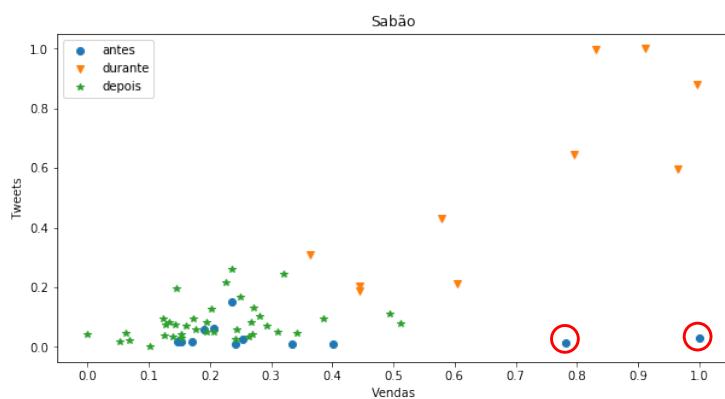


Figura 14: gráfico de dispersão do sabão entre *tweets* e vendas.
Fonte: elaborada pelo autor.

Em relação a comparação entre Trends e Twitter (figura 16), apesar dos picos serem coincidentes, os pontos após o período de *panic buying* estão visivelmente deslocados na região superior esquerda do gráfico devido ao fato de que após o pico, o volume de pesquisa no Google

Trends não sofreu uma queda proporcional a queda que os *tweets* tiveram. Esse deslocamento indica que o Google Trends não possui uma boa correlação com os dados do Twitter.

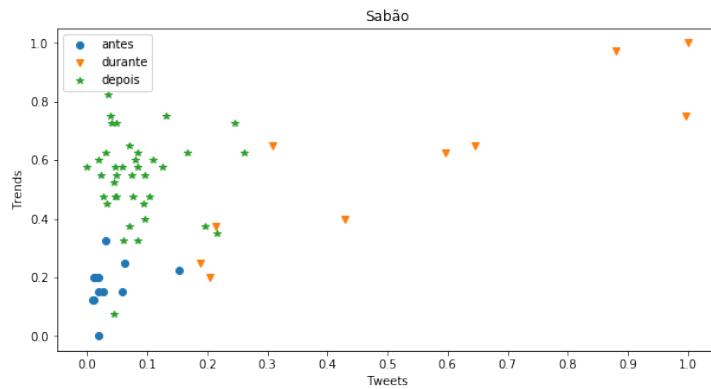


Figura 15: gráfico de dispersão do sabão entre *trends* e *tweets*.
Fonte: elaborada pelo autor.

Semelhante ao observado na comparação Twitter x Trends (figura 17), a comparação entre Trend e vendas mostra um perfil semelhante, os pontos referentes ao período posterior ao pico de *panic buying* estão deslocados para a região superior esquerda. A análise de correlação deixará evidente se essa correlação existe ou não.

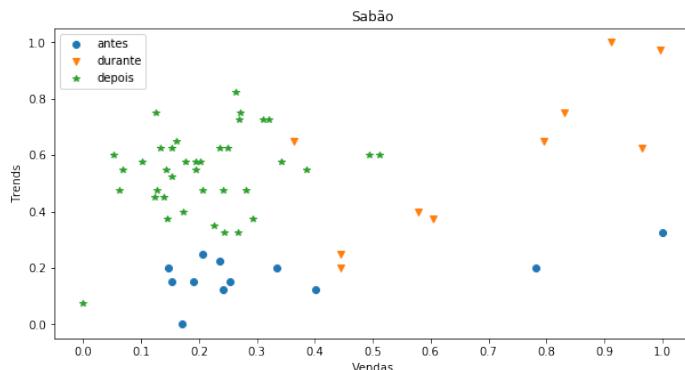


Figura 16: gráfico de dispersão do sabão entre *trends* e vendas.
Fonte: elaborada pelo autor.

4.3. Álcool

Outro produto analisado foi o álcool. Esse produto apresentou esgotamento dos estoques próximo ao dia 23, fazendo com que os dados de venda desses dias não sejam proporcionais às vendas. No entanto a análise desse produto é interessante pois será possível ver como as vendas aumentariam nesse período caso seja observado a existência de correlação entre o volume de vendas e os tweets nesse período.

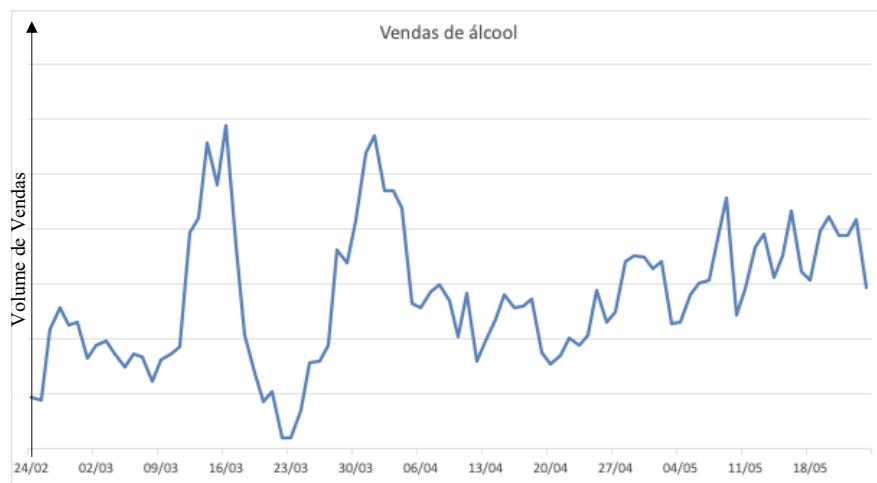


Figura 17: volume de vendas de álcool no período.
Fonte: elaborada pelo autor.

Para encontrar os dados de álcool foi feita uma busca pelo termo ("alcool" + "álcool"). A figura 19 mostra o gráfico relativo ao *Google Trends* de álcool. A curva de álcool tem um pico acentuado que corresponde ao período em que houve falta de estoque nas vendas.

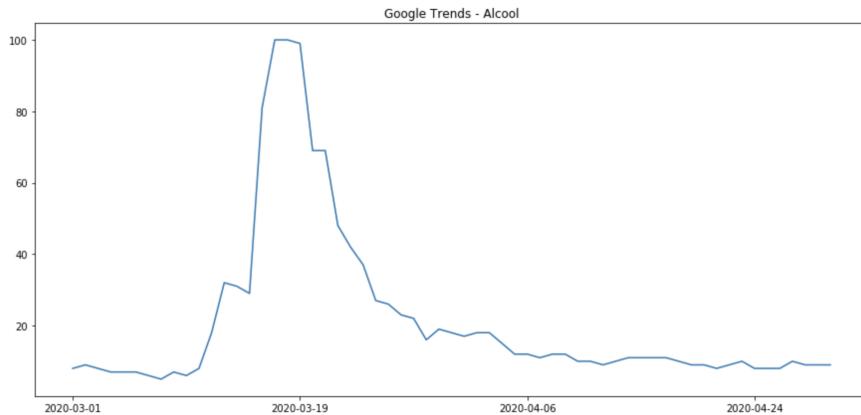


Figura 18: pesquisa por álcool no *Google* durante o período.
Fonte: elaborada pelo autor.

Os *tweets* referentes a álcool no período foram aqueles que possuíram maior quantidade de *tweets* dentre os produtos analisados, atingindo um pico de 60 mil *tweets* em um único dia. O formato da curva se assemelha a curva do Google Trends, mas difere da curva de vendas.

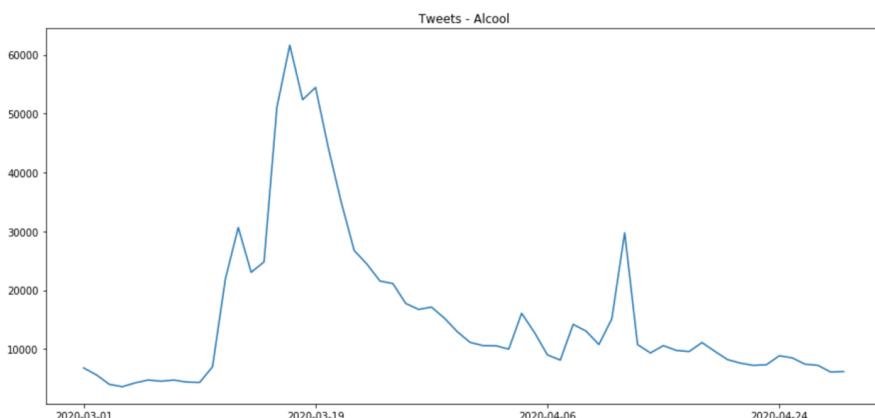


Figura 19: *tweets* de álcool sem filtro durante o período.
Fonte: elaborada pelo autor.

A aplicação do filtro reduziu significamente a quantidade de *tweets*, com o pico indo de 60 mil para 8 mil. Apesar disso, no caso do álcool, a curva não teve grandes

alterações, tendo um formato parecido, removendo grandes ruídos que ocorreram no período.

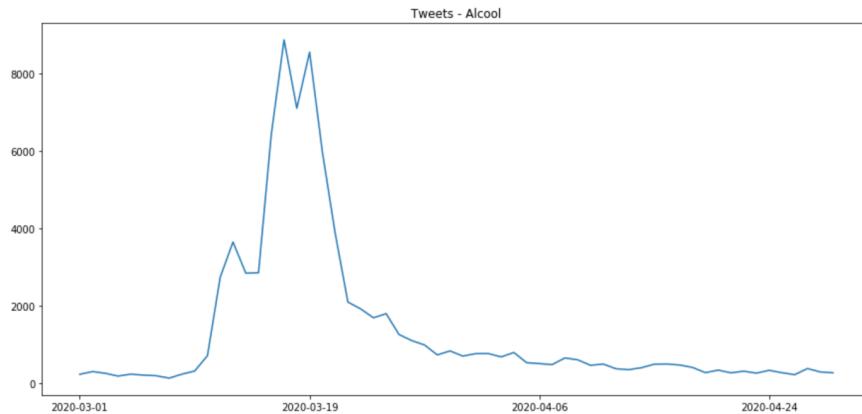


Figura 20: *tweets* de álcool sem filtro durante o período.
Fonte: elaborada pelo autor.

O gráfico de dispersão na figura 22, entre Twitter e vendas, mostra que aparentemente não existe correlação entre os dados, os picos de ambos não coincidem e os dados estão dispersos. Isso em parte pode ser atribuído à quebra de estoque que houve pois no momento em que haveria o pico de vendas, ocorreu a quebra de estoque, fazendo as vendas não acompanharem os *tweets* nesse período.

Commented [HY2]: Explique porque, não está claro.

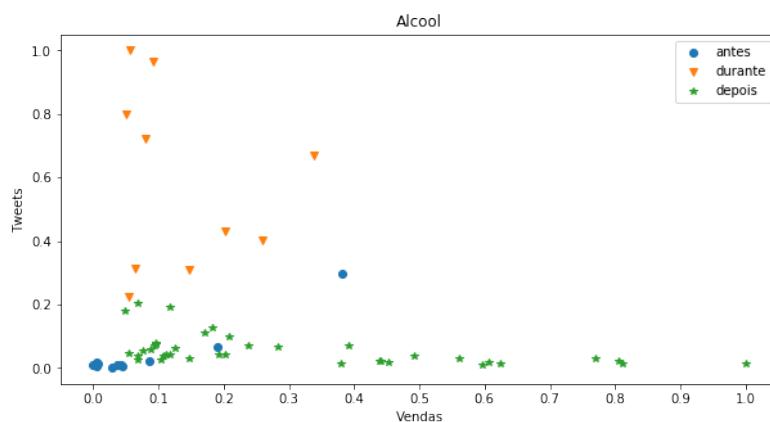


Figura 21: gráfico de dispersão do álcool entre *tweets* e vendas.
Fonte: elaborada pelo autor.

A comparação entre Google Trend e Twitter (figura 23) por sua vez, mostra que possivelmente existe uma forte correlação entre esses dados. E isso pode sugerir que caso não houvesse falta de estoque as vendas poderiam ter um comportamento similar da mesma forma que ocorreu com papel higiênico.

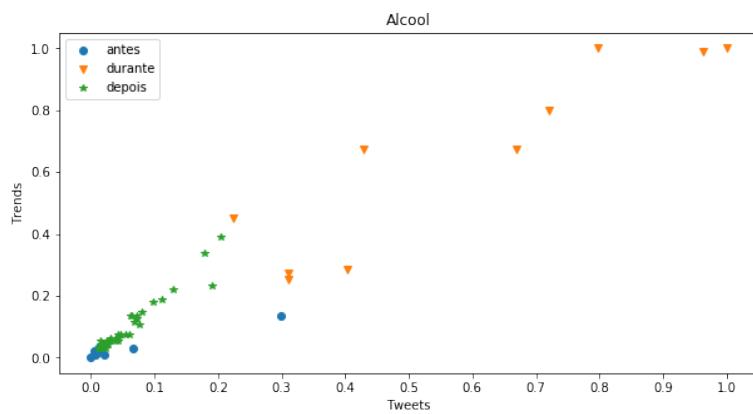


Figura 22: gráfico de dispersão do álcool entre *trends* e *tweets*.
Fonte: elaborada pelo autor.

Por fim, o gráfico de dispersão entre Trends e Vendas (figura 23) mostra um comportamento semelhante ao gráfico entre Twitter e Vendas. O comportamento era esperado uma vez que o gráfico entre Twitter e Trends mostrou uma forte correlação entre ambos.

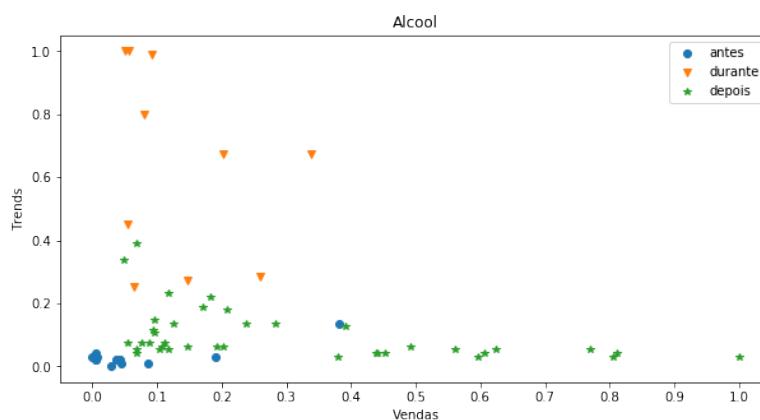


Figura 23: gráfico de dispersão do álcool entre *trends* e vendas.
Fonte: elaborada pelo autor.

4.4. Massas

Escolheu-se analisar massas secas por se tratar de um produto não relacionado a higiene como os produtos anteriores e que mesmo assim apresentou um comportamento de *panic buying* segundo a empresa X. Novamente nota-se um pico no mesmo período de 14/03 a 21/03. No entanto o pico desse gráfico não é tão acentuado como no caso do papel higiênico. Além disso também se nota uma sazonalidade semanal na venda de massas referente aos finais de semana.

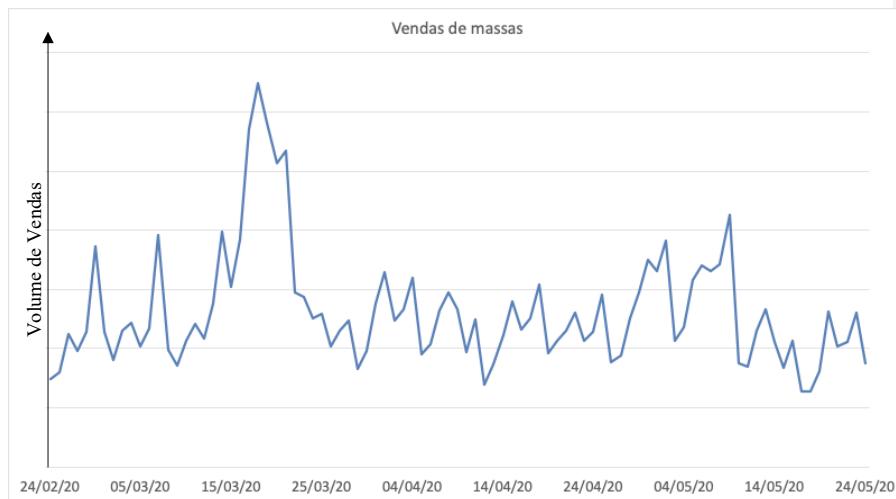


Figura 24: volume de vendas de massas no período.
Fonte: elaborada pelo autor.

Para encontrar os dados de massas do Google Trends foi feita uma busca pelo termo ("macarrão" + "macarao" + "macarronada" + "lasanha" + "mijo"). Optou-se nesse caso por não adicionar o termo "massa", uma vez que possui outros significados diferentes do alimento, como por exemplo, conceito de massa em física ou massa como sinônimo de grande quantidade de pessoas. A figura 26 mostra o gráfico relativo ao *Google Trends* de papel higiênico. Pode-se observar a sazonalidade sazonal de massas no gráfico e um aumento na busca pelo termo

após o início da quarentena. No entanto esse gráfico não possui o pico acentuado que o gráfico de vendas possui.

Commented [HY3]: Parágrafo fora do alinhamento.

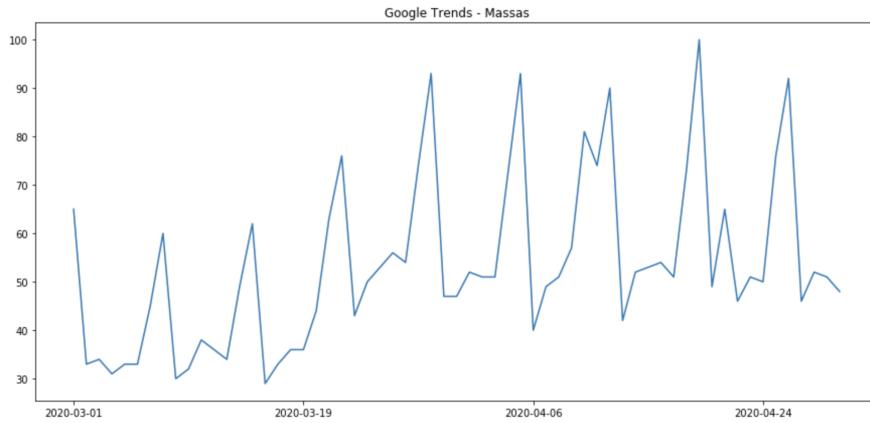


Figura 25: pesquisa por massas no *Google* durante o período.
Fonte: elaborada pelo autor.

A coleta dos *tweets* sem filtro aparenta um comportamento mais próximo do visto com o Google Trends do que com o de vendas, como visto na figura 27. No entanto muitos *tweets* envolvendo massas eram referentes a um tipo de piada que viralizou nas redes sociais, como no seguinte *tweet*: “Essa Coronavírus é igual macarrão, foi inventada na China e popularizada na Itália”.

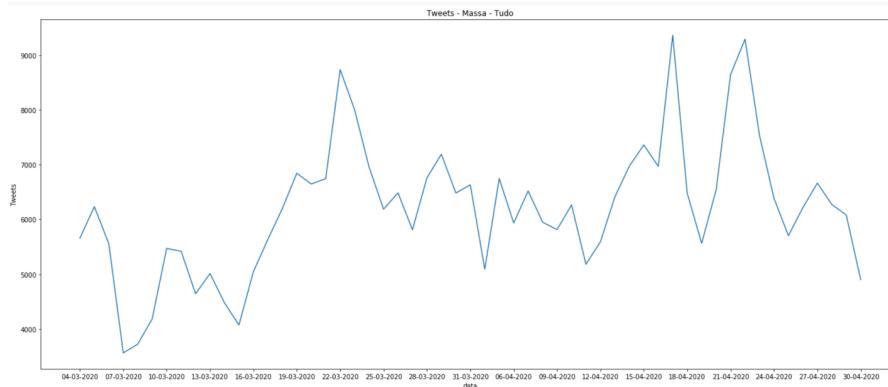


Figura 26: *tweets* de massas sem filtro durante o período.
Fonte: elaborada pelo autor.

Após a aplicação do filtro, ocorre uma grande mudança no formato da curva. Observa-se que o gráfico mostrado na figura 28 possui mais semelhanças com o gráfico de vendas, onde há um pico nos mesmos períodos.

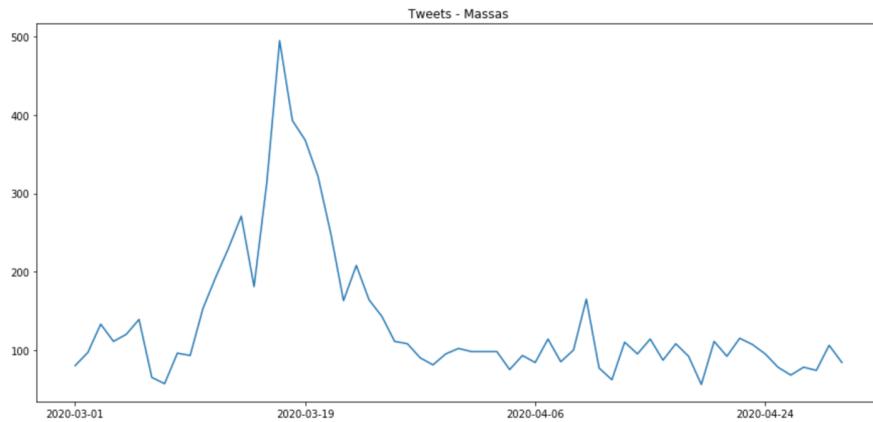


Figura 27: *tweets* de massas com filtro durante o período.

Fonte: elaborada pelo autor.

O gráfico de dispersão entre o *Twitter* e as Vendas (figura 29) possui um comportamento que sugere haver uma forte correlação entre ambos os dados, assim como no caso do papel higiênico e do sabão.

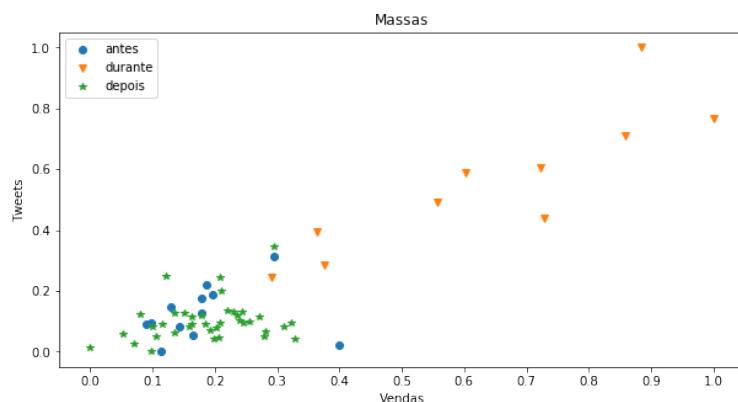


Figura 28: gráfico de dispersão de massas entre *tweets* e vendas.

Fonte: elaborada pelo autor.

A dispersão entre o *Google Trends* e as vendas (figura 30), por sua vez, uma dispersão grande entre os dados, não aparentando haver correlação entre os dados. Isso era esperado quando foi observado os gráficos de vendas e do *Trends*.

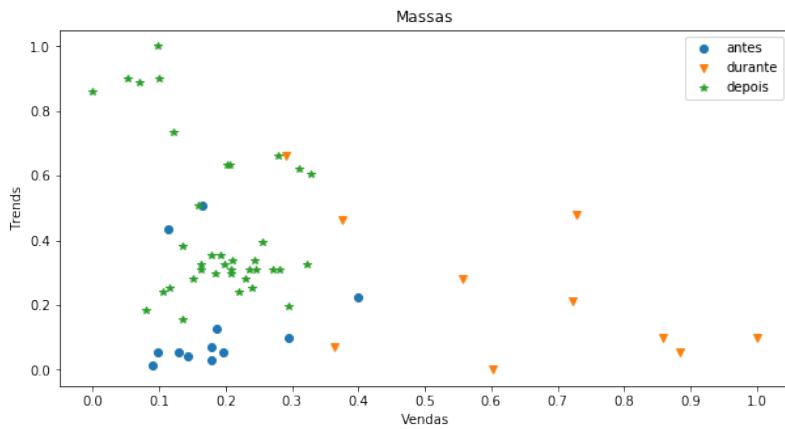


Figura 29: gráfico de dispersão de massas entre *trends* e vendas.
Fonte: elaborada pelo autor.

O gráfico de dispersão entre *Twitter* e *Trends* (figura 31) possui um comportamento muito semelhante ao encontrado na comparação entre *Trends* e vendas, onde não aparenta haver correlação entre os dados adquiridos.

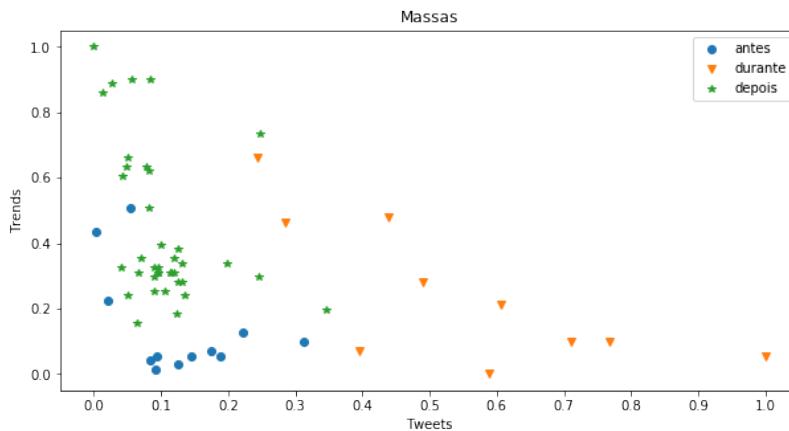


Figura 30: gráfico de dispersão de massas entre *trends* e *tweets*.
Fonte: elaborada pelo autor.

4.5. Doces

O último produto escolhido foi o doce, produto que não apresentou *panic buying* no período analisado. A figura 32 mostra o gráfico de vendas no período. Nota-se a existência de um pico devido à Páscoa em abril, no entanto, nesse mesmo período os demais produtos selecionados não apresentam aumento significativo no volume de vendas.



Figura 31: volume de vendas de doces no período.
Fonte: elaborada pelo autor.

Para encontrar os dados de doce referentes ao *Google Trends* foi feita uma busca pelo termo ("chocolate" + "doce" + "bala" + "biscoito" + "bolacha"). A figura 33 mostra o gráfico relativo ao *Google Trends* de doces. O gráfico mostra que ocorre um aumento na busca por doces durante a época de páscoa, no entanto o formato da curva difere da curva de vendas. Enquanto as vendas possuem um pico acentuado nos dias próximos à Páscoa, os dados provenientes do *Google Trends* mostram um período em que há aumento pela busca desde o final de março até a segunda semana de abril. Além disso, existe também uma sazonalidade semanal da busca pelo termo referente aos fins de semana e, durante o período de Páscoa, esses picos ficam mais acentuados, mas não chegam a ter a mesma projeção das vendas.

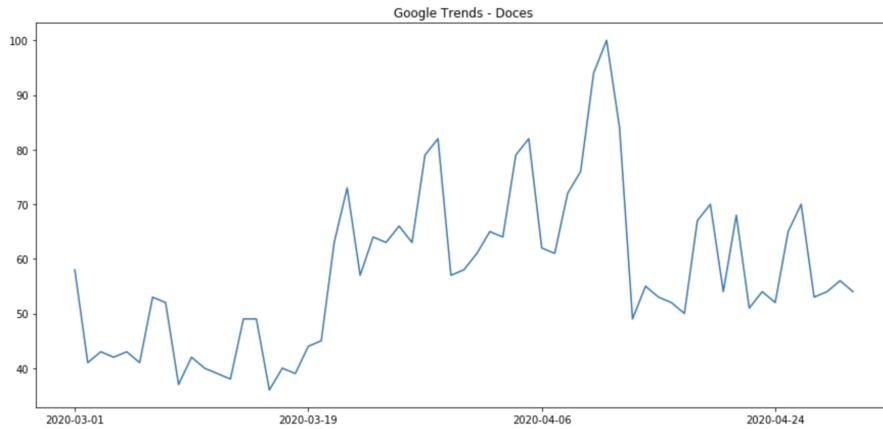


Figura 32: pesquisa por doces no *Google* durante o período.
Fonte: elaborada pelo autor.

Os *tweets* não filtrados de doces (figura 34) também diferem do gráfico de vendas aparentando não haver correlação entre os gráficos. A aplicação do filtro, no entanto, altera significativamente o formato da curva do gráfico.

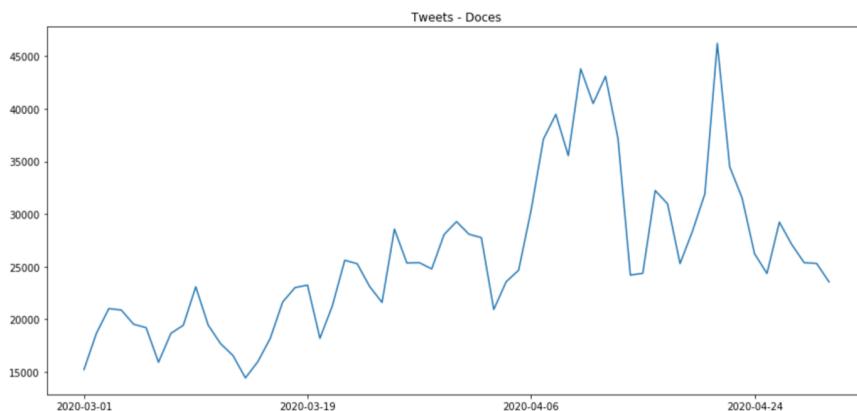


Figura 33: *tweets* de doces sem filtro durante o período.
Fonte: elaborada pelo autor.

O filtro reduziu expressivamente a quantidade de *tweets* analisados, indo de um pico de perto de 45 mil *tweets* para um pico próximo de 1,8 mil. O gráfico após o filtro (figura 35) está mais próximo do gráfico de vendas, no entanto ainda está relativamente diferente do gráfico de vendas.

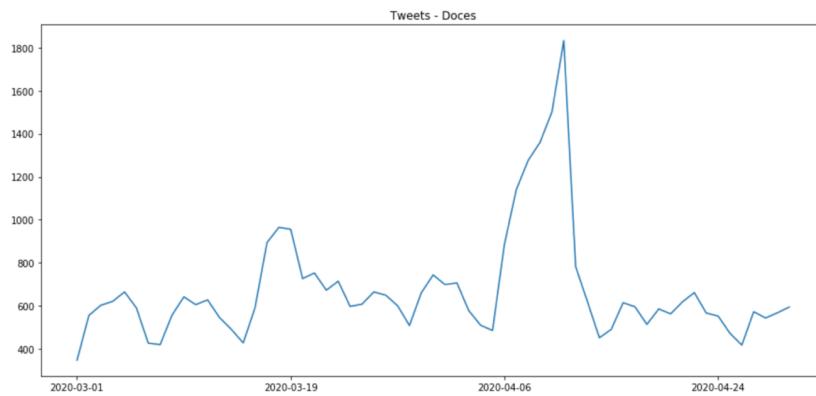


Figura 34: *tweets* de doces com filtro durante o período.

Fonte: elaborada pelo autor.

O gráfico de dispersão entre *Twitter* e vendas (figura 36) mostra que existe uma correlação entre os dados. No entanto, essa correlação aparenta ser relacionada ao período de páscoas, o que pode indicar que o *Twitter* correlaciona com períodos em que há grandes quantidades de compras notada pelas pessoas.

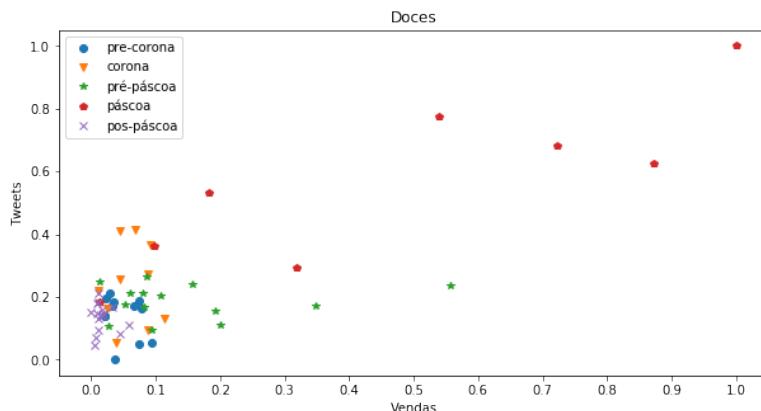


Figura 35: gráfico de dispersão de doces entre *tweets* e vendas.

Fonte: elaborada pelo autor.

A comparação entre *Trends* e vendas (figura 37) parece haver uma correlação, mas ela parece ser menor quando comparada com as outras correlações vistas nos gráficos de dispersões. Desconsiderando o período de páscoa, essa correlação fica praticamente inexistente.

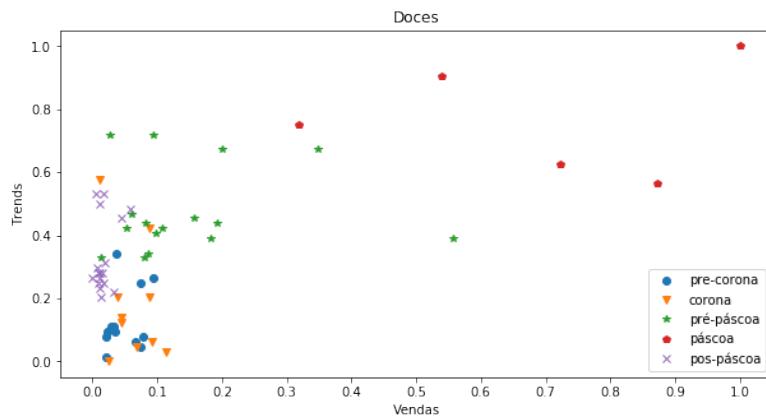


Figura 36: gráfico de dispersão de massas entre *trends* e vendas.

Fonte: elaborada pelo autor.

Da mesma forma, a comparação entre *Trends* e *Twitter* (figura 38) parece haver correlação somente no período de páscoa.

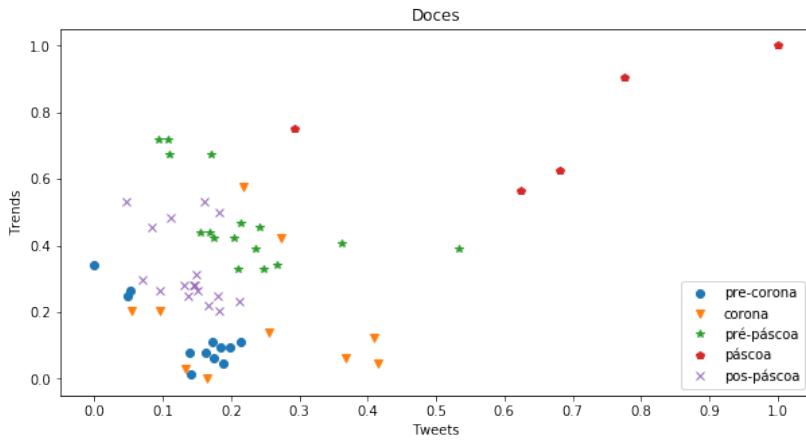


Figura 37: gráfico de dispersão de massas entre *trends* e *tweets*.

Fonte: elaborada pelo autor.

4.6. Análise de correlação

Após a coleta e tratamento dos dados de cada tipo de produto, realizou-se o teste de correlação dos produtos, descrito no capítulo passado de metodologia. Como dito na metodologia, para os testes, foi adotado um nível de significância $\alpha = 1\%$. As tabelas x-x mostram os resultados obtidos para cada tipo de dado.

O primeiro cálculo feito foi comparar os dados entre *Twitter* e vendas, como mostra a tabela 2. Segundo o teste de hipótese, encontrou-se que todos os produtos, com exceção do álcool, obtiveram alta correlação. O principal motivo para a baixa correlação do álcool é a falta de estoque que aconteceu com o produto, fazendo as vendas não serem equivalentes a demanda no período. No entanto, os resultados dos outros produtos mostram que o *Twitter* é uma fonte de dados que está diretamente relacionada com as vendas de alguns produtos e como tal, pode ser utilizada para análises futuras de demanda.

Tabela 2: Teste de correlação dos produtos comparando *Twitter* e vendas.

Comparação Twitter x Vendas						
Produto	r	n-2	t_(n-2)	a	t_crítico	Hipótese H ₀
Papel Higiênico	0.7343	57	4.469	0.5%	2.920	Rejeito
Sabão	0.7157	57	4.394	0.5%	2.920	Rejeito
Álcool	-0.1938	57	-1.436	0.5%	2.920	Aceito
Massas	0.8842	57	5.001	0.5%	2.920	Rejeito
Doces	0.7903	57	4.681	0.5%	2.920	Rejeito

Fonte: elaborada pelo autor.

O teste de correlação envolvendo o *Google Trends* e os dados de vendas (tabela 3), por sua vez, mostram que apenas papel higiênico e doces obtiveram correlação entre os dados. O fato de que não é possível aplicar um filtro nos dados provenientes da plataforma possivelmente prejudicou a análise, uma vez que não se pode filtrar as buscas em que o usuário tinha intenção de comprar o produto.

Tabela 3: Teste de correlação dos produtos comparando *Trends* e vendas.

Comparação Trends x Vendas						
	r	n-2	t_(n-2)	a	t_crítico	Hipótese H ₀
Papel Higiênico	0.7271	57	4.440	0.5%	2.920	Rejeito
Sabão	0.2523	57	1.847	0.5%	2.920	Aceito
Álcool	-0.2194	57	-1.618	0.5%	2.920	Aceito
Massas	-0.3138	57	-2.261	0.5%	2.920	Aceito
Doces	0.5788	57	3.782	0.5%	2.920	Rejeito

Fonte: elaborada pelo autor.

A correlação entre *Twitter* e *Google Trends* (tabela 4) mostra a existência de correlação entre os dados dos produtos de higiene. No entanto, é importante notar que, mesmo existindo correlação no sabão, o valor dessa correlação é de 0.4821, que representa uma correlação baixa entre os dados dos produtos. Além disso, o valor de t_{n-2} encontrado para doces foi próximo do valor de $t_{critico}$, o que indica que caso fosse adotado um valor de α um pouco maior, esses dados poderiam ter passado no teste de hipótese, mostrando uma correlação mais baixa entre os dados de doces.

Tabela 4: Teste de correlação dos produtos comparando *Twitter* e *Trends*.

Comparação Twitter x Trends						
	r	n-2	t_(n-2)	a	t_crítico	Hipótese H ₀
Papel Higiênico	0.9189	57	5.108	0.5%	2.920	Rejeito
Sabão	0.4821	57	3.279	0.5%	2.920	Rejeito
Álcool	0.9659	57	5.245	0.5%	2.920	Rejeito
Massas	-0.3968	57	-2.784	0.5%	2.920	Aceito
Doces	0.4181	57	2.912	0.5%	2.920	Aceito

Fonte: elaborada pelo autor.

Outro cálculo de correlação feito foi adicionar um *lag* entre os dados de *Twitter/Trends* e os dados de vendas, ou seja, comparou-se os dados de vendas do dia x com os dados do *Twitter/Trends* do dia x - n, onde n = 0, 1, 2, 3, 4. A tabela 5 indica essa análise com os dados do *Twitter*. Verificou-se que, à medida que se aumenta o *lag* entre os dados, a correlação entre

elas vai diminuindo. Isso mostra que a correlação entre os dados, quando ela existe, é imediata, ou seja, não vale a pena utilizar o *lag*

Tabela 5: Teste de correlação dos produtos com *lag* comparando *Twitter* e vendas.

Produto	lag	r	n-2	t_(n-2)	a	t_crítico	Hipótese H ₀
Papel Higiênico	0	0.7337	57	4.466	0.5%	2.920	Rejeito
	1	0.6758	56	4.190	0.5%	2.923	Rejeito
	2	0.5928	55	3.782	0.5%	2.925	Rejeito
	3	0.4817	54	3.189	0.5%	2.927	Rejeito
	4	0.3532	53	2.425	0.5%	2.929	Aceito
Sabão	0	0.7156	57	4.394	0.5%	2.920	Rejeito
	1	0.6791	56	4.204	0.5%	2.923	Rejeito
	2	0.5733	55	3.689	0.5%	2.925	Rejeito
	3	0.3842	54	2.636	0.5%	2.927	Aceito
	4	0.2543	53	1.795	0.5%	2.929	Aceito
Álcool	0	-0.1812	57	-1.346	0.5%	2.920	Aceito
	1	-0.1824	56	-1.343	0.5%	2.923	Aceito
	2	-0.2098	55	-1.523	0.5%	2.925	Aceito
	3	-0.2147	54	-1.543	0.5%	2.927	Aceito
	4	-0.2420	53	-1.712	0.5%	2.929	Aceito
Massas	0	0.8853	57	5.005	0.5%	2.920	Rejeito
	1	0.8943	56	4.988	0.5%	2.923	Rejeito
	2	0.7521	55	4.458	0.5%	2.925	Rejeito
	3	0.6071	54	3.814	0.5%	2.927	Rejeito
	4	0.4965	53	3.238	0.5%	2.929	Rejeito
Doces	0	0.7896	57	4.679	0.5%	2.920	Rejeito
	1	0.7112	56	4.337	0.5%	2.923	Rejeito
	2	0.4966	55	3.298	0.5%	2.925	Rejeito
	3	0.2331	54	1.668	0.5%	2.927	Aceito
	4	0.0584	53	0.424	0.5%	2.929	Aceito

Fonte: elaborada pelo autor.

Na tabela 6, está a análise com os dados do *Trends* e vendas. Semelhante ao visto na análise do *Twitter*, ao aumentar o valor do *lag*, a correlação entre os dados possui diminui e não vale a pena usar um *lag*.

Tabela 6: Teste de correlação dos produtos com lag comparando *trendr* e vendas.

Produto	lag	r	n-2	t_(n-2)	a	t_crítico	Hipótese H ₀
Papel Higiênico	0	0.7264	57	4.437	0.5%	2.920	Rejeito
	1	0.6500	56	4.078	0.5%	2.923	Rejeito
	2	0.6153	55	3.886	0.5%	2.925	Rejeito
	3	0.4969	54	3.270	0.5%	2.927	Rejeito
	4	0.3403	53	2.346	0.5%	2.929	Aceito
Sabão	0	0.2641	57	1.928	0.5%	2.920	Aceito
	1	0.1360	56	1.008	0.5%	2.923	Aceito
	2	0.0269	55	0.199	0.5%	2.925	Aceito
	3	-0.1105	54	-0.807	0.5%	2.927	Aceito
	4	-0.2551	53	-1.799	0.5%	2.929	Aceito
Álcool	0	-0.2086	57	-1.542	0.5%	2.920	Aceito
	1	-0.2068	56	-1.515	0.5%	2.923	Aceito
	2	-0.2052	55	-1.491	0.5%	2.925	Aceito
	3	-0.2114	54	-1.520	0.5%	2.927	Aceito
	4	-0.2276	53	-1.615	0.5%	2.929	Aceito
Massas	0	-0.3138	57	-2.260	0.5%	2.920	Aceito
	1	-0.4543	56	-3.095	0.5%	2.923	Rejeito
	2	-0.3684	55	-2.564	0.5%	2.925	Aceito
	3	-0.2319	54	-1.660	0.5%	2.927	Aceito
	4	-0.2441	53	-1.726	0.5%	2.929	Aceito
Doces	0	0.5790	57	3.783	0.5%	2.920	Rejeito
	1	0.3845	56	2.685	0.5%	2.923	Aceito
	2	0.2397	55	1.728	0.5%	2.925	Aceito
	3	0.3177	54	2.225	0.5%	2.927	Aceito
	4	0.3060	53	2.130	0.5%	2.929	Aceito

Fonte: elaborada pelo autor.

A última análise de correlação feita foi a correlação entre produtos por tipo de dado, os valores acima de 0.42 são aqueles que passam no teste de correlação com significância de 99,5%. A análise das vendas (figura 39) mostrou que todos os produtos, exceto doces e álcool, tiveram alta correlação entre si. Isso está de acordo com o observado, álcool teve falta de estoque no período e doces não sofreu de *panic buying*, portanto não apresentam correlação com os outros produtos.

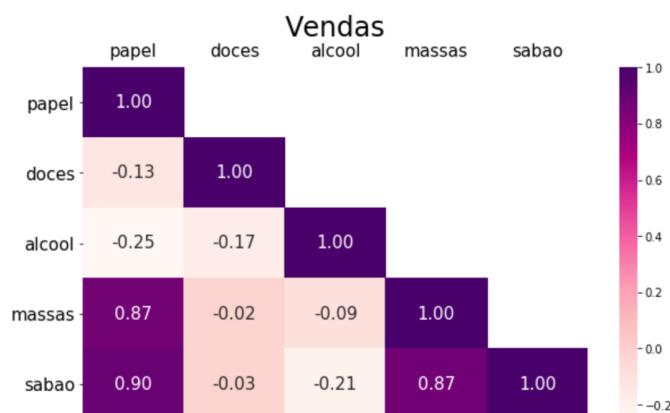


Figura 38: matriz de correlação entre produtos para os dados de vendas
Fonte: elaborada pelo autor.

A correlação entre os produtos com os dados do *Twitter* (figura 40) mostra que todos os produtos, exceto doces, tiveram alta correlação entre si. Esse resultado faz sentido uma vez que se observou anteriormente que o *Twitter* está fortemente correlacionado com as vendas de produtos e, dado que álcool teve quebra de estoque no período, ele também sofreu de *panic buying* e, portanto, possui comportamento no *Twitter* semelhante aos outros produtos que sofreram desse fenômeno.

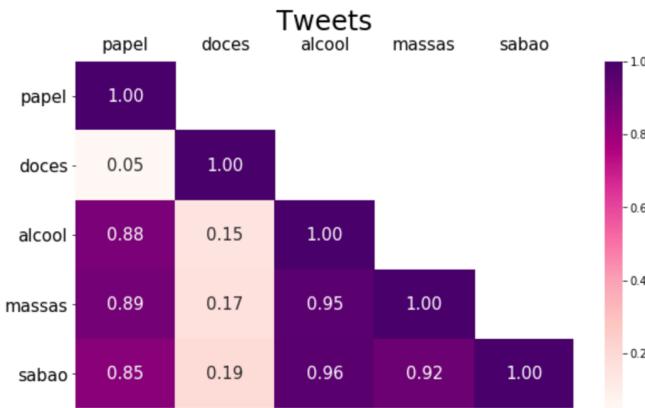


Figura 39: matriz de correlação entre produtos para os dados do *Twitter*
Fonte: elaborada pelo autor.

Por fim, a correlação entre produtos no *Google Trends* (figura 41) mostra que não existe correlação significativa entre os produtos no *Google Trends*. Houve uma forte correlação entre papel higiênico e álcool, que foram dois produtos que sofreram *panic buying* e foram amplamente divulgados nos noticiários. Além disso, também ocorreu uma correlação entre massas e doces que, como visto anteriormente neste trabalho, apresentaram um comportamento sazonal por semana, com os fins de semana possuindo mais pesquisas em relação aos demais dias da semana. Isso indica que em relação aos alimentos, é possível que as pessoas busquem por receitas ou possuem mais interesse por esses produtos nos fins de semana.

Commented [HY4]: Analise melhor, pois tem alta correlação entre papel e massas e doces e massas, talvez por que as pessoas procuram receitas nos fins de semana.

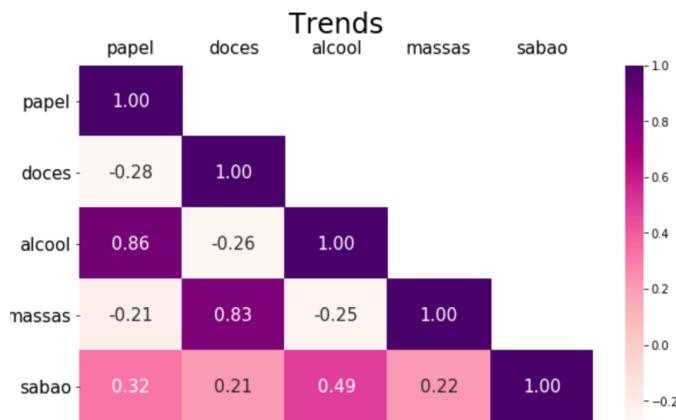


Figura 40: matriz de correlação entre produtos para os dados do Google Trends
Fonte: elaborada pelo autor.

5. CONCLUSÃO

A pandemia do COVID-19 trouxe consigo efeitos indesejáveis, dentre eles o *panic buying*. Esse fenômeno tem muita importância nas cadeias logísticas, uma vez que o aumento repentino no volume de compras pode comprometer os estoques das empresas e deixar a população mais vulnerável sem produtos básicos. Dessa forma, entender o comportamento das pessoas nesse período é importante para decidir um plano de ação para enfrentar futuros casos semelhantes ao ocorrido.

Neste contexto, o presente Trabalho de Formatura se propôs identificar a existência de correlação entre o comportamento das pessoas nas redes sociais, com foco no *Twitter*, e as vendas de determinados produtos. Esse objetivo foi cumprido ao se observar a existência dessa correlação entre os dados do *Twitter* e os dados de vendas. A existência dessa correlação entre os dados é de interesse das cadeias logísticas, pois pode providenciar mais dados sobre os tipos de produtos que estão em maior procura no momento. A tabela 7 mostra um resumo com as principais conclusões encontradas neste trabalho.

Tabela 7: Tabela-resumo das principais conclusões do trabalho.

Objetivo	Conclusão
Correlação entre <i>Twitter</i> e vendas	Alta correlação entre as vendas de produtos e a quantidade de tweets de <i>panic buying</i> referentes a eles
Correlação entre <i>Trends</i> e vendas	Baixa correlação pela impossibilidade de se filtrar buscas referentes a intenção de compra
Correlação entre <i>Twitter</i> e <i>Trends</i>	Correlação apenas com alguns produtos que estiveram mais em destaque nas mídias
Correlação com lag	A correlação é maior com os dados sem lag, indicando uma simultaneidade entre compra e divulgação nas redes sociais
Correlação entre produtos	Existe uma forte correlação no <i>Twitter</i> e vendas entre os produtos que sofreram <i>panic buying</i> no mesmo período. <i>Trends</i> obteve correlação entre os alimentos por causa da sazonalidade semanal na busca por esses tipos de produtos

Fonte: elaborada pelo autor.

Commented [HY5]: Pode afirmar que vc conseguiu atingir os objetivos expressos no cap. 1. Ou seja, deixe isso claro para o leitor, vc cumpriu o que prometeu fazer.

Foram identificados por meio de análises gráficas quais seriam os produtos com maior interesse de serem analisados, aqueles em que se sabe existir *panic buying* ou não, para que se possa comparar o comportamento em ambos os casos. Após a identificação foi feita a coleta dos dados proveniente de das plataformas *Google Trends* e *Twitter* para comparação. Existe novas oportunidades de pesquisa sobre o tema buscando englobar outras redes sociais tal como Facebook, Tik Tok e Instagram para verificar o comportamento das pessoas nesses outros tipos de rede sociais e se existe correlação entre os dados dessas redes.

Analizando os dados e resultados encontrados pode-se concluir que os dados levantados sobre Twitter tem uma forte correlação com os dados de vendas, com a exceção do álcool em que houve falta de estoque e o volume de vendas não reflete a demanda toda no período. Por outro lado, os dados em relação ao Google Trends se mostram tendo baixa correlação com os dados de vendas, não podendo ser indicativo para a demanda no período. Uma possível causa dessa baixa correlação é o fato de que os dados possuem muito ruído, ou seja, existem várias pesquisas sobre os termos que não estão relacionados com compra ou com *panic buying*. Da mesma forma, Twitter e Google Trends possuem pouca correlação, com exceção de papel higiênico e álcool, que foram produtos que estiveram em bastante destaque por causa do *panic buying*.

A análise com *lag* (onde houve *delay* entre os dados do Twitter/Trends e de Vendas) mostra que a correlação é imediata, ou seja, ocorrem praticamente nos mesmos dias. As correlações diminuem à medida que aumentamos o *lag* entre os dados, sendo que a maior correlação encontrada é aquela onde não há *lag* entre os dados. Dessa forma, a análise dos dados provenientes do Twitter não são ideais para usar como previsão de demanda de algum produto, uma vez que eles ocorrem em instantes semelhantes.

Por fim, a análise entre produtos, mostra que existe uma alta correlação entre os produtos que sofreram *panic buying* tanto em vendas como no Twitter. No caso de vendas, nota-se que álcool teve uma baixa correlação com os demais produtos quando comparado com as correlações do Twitter. Dada a alta correlação encontrada entre os demais produtos, é de se estimar que as vendas de álcool no período poderiam ser maiores caso não houvesse falta de estoque, tendo curvas semelhante ao dos outros produtos

Como é possível verificar pelos resultados descritos no capítulo anterior, o *Twitter* apresentou uma alta correlação com os dados de vendas, mostrando que o Twitter pode ser um bom indicativo do interesse das pessoas sobre determinado produto. Análises futuras podem

realizar uma análise de sentimento sobre os *tweets*, ou seja, descobrir se um determinado *tweet* contém uma mensagem positiva ou negativa em relação a determinado produto.

Por fim, espera-se que os resultados relatados inspirem mais estudos sobre o tema, gerando mais conhecimento sobre o comportamento das pessoas nas redes sociais e como esse comportamento pode auxiliar empresas de varejo a gerenciar suas vendas e estoques durante tempos de desastres como pandemias.

REFERÊNCIAS BIBLIOGRÁFICAS

ABD-ALRAZAQ, A. et al. Top concerns of tweeters during the COVID-19 pandemic: A surveillance study. **Journal of Medical Internet Research**, v. 22, n. 4, p. 1–9, 2020.

AHMED, W. et al. Moral panic through the lens of Twitter: An analysis of infectious disease outbreaks. **ACM International Conference Proceeding Series**, p. 217–221, 2018.

AQUINO, V. Brasil confirma primeiro caso da doença. **Gov.** Disponível em: <<https://www.gov.br/saude/pt-br/assuntos/noticias/brasil-confirma-primeiro-caso-de-novo-coronavirus>>. Acesso em: Outubro 2020

ARAFAT, S. M. Y. et al. Panic buying: An insight from the content analysis of media reports during COVID-19 pandemic. **Neurology Psychiatry and Brain Research**, v. 37, n. July, p. 100–103, 2020. Disponível em: <<https://doi.org/10.1016/j.npbr.2020.07.002>>.

BEN HASSEN, T.; EL BILALI, H.; ALLAHYARI, M. S. Impact of COVID-19 on Food Behavior and Consumption in Qatar. **Sustainability**, v. 12, n. 17, p. 6973, 2020.

BYUN, C. et al. Twitter data collecting tool with rule-based filtering and analysis module. **International Journal of Web Information Systems**, v. 9, n. 3, p. 184–203, 2013.

CHEW, C.; EYSENBACH, G. Pandemics in the age of Twitter: Content analysis of tweets during the 2009 H1N1 outbreak. **PLoS ONE**, v. 5, n. 11, p. 1–13, 2010.

GOOGLE TRENDS. **Google Trends**. Página Inicial. Acesso em: <<https://trends.google.com/>>. Disponível em: outubro 2020

GRASHUIS, J.; SKEVAS, T.; SEGOVIA, M. S. Grocery shopping preferences during the COVID-19 pandemic. **Sustainability (Switzerland)**, v. 12, n. 13, 2020.

HENDRICKS, K. B.; SINGHAL, V. R. An empirical analysis of the effect of supply chain disruptions on long-run stock price performance and equity risk of the firm. **Production and Operations Management**, v. 14, n. 1, p. 35–52, 2005.

HOBBS, J. E. Food supply chains during the COVID-19 pandemic. **Canadian Journal of Agricultural Economics**, v. 68, n. 2, p. 171–176, 2020.

JUPYTER. **Jupyter**. Página Inicial. Acesso em: <<https://jupyter.org/>>. Acesso em: outubro 2020.

KHARE, A.; HE, Q.; BATTA, R. Predicting gasoline shortage during disasters using social media. **OR Spectrum**, v. 42, n. 3, p. 693–726, 2020. Disponível em: <<https://doi.org/10.1007/s00291-019-00559-8>>.

LAATO, S. et al. Unusual purchasing behavior during the early stages of the COVID-19 pandemic: The stimulus-organism-response approach. **Journal of Retailing and Consumer Services**, v. 57, n. July, p. 102224, 2020. Disponível em: <<https://doi.org/10.1016/j.jretconser.2020.102224>>.

LEACH, J. **Survival Psychology**. Palgrave Macmillan UK. Basingstoke, 1994

MATPLOTLIB. **Matplotlib**. Página Inicial. Acesso em: <<https://matplotlib.org/>>. Acesso em: outubro 2020.

MINISTÉRIO DA SAÚDE. Sobre a Doença. **Gov**. Disponível em: <<https://coronavirus.saude.gov.br/sobre-a-doenca>>. Acesso em: outubro 2020

NAEEM, M. Do social media platforms develop consumer panic buying during the fear of Covid-19 pandemic. **Journal of Retailing and Consumer Services**, v. 58, n. September 2020, p. 102226, 2021. Disponível em: <<https://doi.org/10.1016/j.jretconser.2020.102226>>.

NUMPY. **Numpy**. Página Inicial. Acesso em: <<https://numpy.org/>>. Acesso em: outubro 2020.

PANDAS. **Pandas**. Página Inicial. Acesso em: <<https://pandas.pydata.org/>>. Acesso em: outubro 2020.

OSYINT TEAM. Twint Project. **Github**. Acesso em: <<https://github.com/twintproject/twint>>. Acesso em: outubro 2020.

PARMENT, A. Generation Y vs. Baby Boomers: Shopping behavior, buyer involvement and implications for retailing. **Journal of Retailing and Consumer Services**, v. 20, n. 2, p. 189–199, 2013. Disponível em: <<http://dx.doi.org/10.1016/j.jretconser.2012.12.001>>.

SIM, K. et al. The anatomy of panic buying related to the current COVID-19 pandemic. **Psychiatry Research**, v. 288, n. April, p. 113015, 2020. Disponível em: <<https://doi.org/10.1016/j.psychres.2020.113015>>.

SPRINGER, S.; MENZEL, L. M.; ZIEGER, M. Google Trends provides a tool to monitor population concerns and information needs during COVID-19 pandemic. **Brain, Behavior, and Immunity**, v. 87, n. April, p. 109–110, 2020. Disponível em: <<https://doi.org/10.1016/j.bbi.2020.04.073>>.

UPTON, E.; NUTTALL, W. J. Fuel panics: Insights from spatial agent-based simulation. **IEEE Transactions on Intelligent Transportation Systems**, v. 15, n. 4, p. 1499–1509, 2014.

WANG, H.; LIU, X.; HE, K. Agent-based simulation approach for managing communicative competence in public emergence event. **Proceedings - 2013 6th International Conference on Business Intelligence and Financial Engineering, BIFE 2013**, p. 96–100, 2014.

WESSELER, J. Storage Policies: Stockpiling Versus Immediate Release. **Journal of Agricultural and Food Industrial Organization**, v. 18, n. 1, p. 1–9, 2020.

WHO. Rolling updates on coronavirus disease. **WHO (COVID-19)**. Disponível em: <<https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen>>. Acesso em: Outubro 2020

YUEN, K. F. et al. The psychological causes of panic buying following a health crisis. **International Journal of Environmental Research and Public Health**, v. 17, n. 10, 2020.

ZHENG, R.; SHOU, B.; YANG, J. Supply disruption management under consumer panic buying and social learning effects. **Omega (United Kingdom)**, n. xxxx, p. 102238, 2020. Disponível em: <<https://doi.org/10.1016/j.omega.2020.102238>>.

ZHOU, D.; CHEN, L.; HE, Y. An unsupervised framework of exploring events on Twitter: Filtering, extraction and categorization. **Proceedings of the National Conference on Artificial Intelligence**, v. 3, p. 2468–2474, 2015.

ANEXO A

Código desenvolvido em Python para coleta de *tweets*:

```

import twint
import pandas as pd
import numpy as np

def search_tweets(query, since, until, file_name):
    c = twint.Config()
    c.Search = query
    c.Since = since
    c.Until = until
    # c.Limit = 200
    c.Filter_re tweets = True
    c.Custom["tweet"] = ["id", "created_at", "date",
                         "time", "timezone", "place",
                         "tweet", "mentions", "replies_count",
                         "retweets_count", "likes_count",
                         "hashtags", "retweet", "geo"]

    c.Custom["user"] = ["bio", "location"]
    c.Store_csv = True
    c.Output = file_name
    twint.run.Search(c)

#intervalo de tempo dos tweets
since = '2020-03-01 00:00:00'
until = '2020-04-30 23:59:00'

#Coleta dos tweets de papel higiênico
query = "((papel OR pape OR papeu) AND (higienico OR hig OR igienico)) lang:pt"
file_name = "all_papel_mar_abr_2020.csv"
search_tweets(query, since, until, file_name)

#Coleta dos tweets de sabão
query = "(sabão OR sabao OR sabonete) lang:pt"
file_name = "all_massa_mar_abr_2020.csv"
search_tweets(query, since, until, file_name)

```

```

#Coleta dos tweets de álcool
query = "(alcool OR álcool OR alcóol OR alcol) lang:pt"
file_name = "all_alcool_mar_abr_2020.csv"
search_tweets(query, since, until, file_name)

#Coleta dos tweets de massas
query = """(macarrão OR macarronada OR miojo OR macarrazo OR
            spaghetti OR espaguete OR espagueti OR spaghetti OR talharim
) lang:pt"""
file_name = "all_massas_mar_abr_2020.csv"
search_tweets(query, since, until, file_name)

#Coleta dos tweets de doces
query = "(chocolate OR doce OR bala OR biscoito OR bolacha) lang:pt"
file_name = "all_doces_mar_abr_2020.csv"
search_tweets(query, since, until, file_name)

#Criando Datasets para cada produto
papel_df = pd.read_csv("all_papel_mar_abr_2020.csv")
sabao_df = pd.read_csv("all_sabao_mar_abr_2020.csv")
alcool_df = pd.read_csv("all_alcool_mar_abr_2020.csv")
massas_df = pd.read_csv("all_massas_mar_abr_2020.csv")
doces_df = pd.read_csv("all_doces_mar_abr_2020.csv")

```

Código desenvolvido em Python para classificação e filtragem de *tweets*:

```

import twint
import pandas as pd
import numpy as np

def only_filter_pb(df, kw, anti_kw = []):
    prediction = []
    for tweet in df["tweet"]:
        x = 0

        if any(word in tweet.lower() for word in kw): x = 1
        if any(word in tweet.lower() for word in anti_kw): x = 0
        prediction.append(x)
    df['pred'] = prediction

```

```

dados_papel = df[df['pred'] == 1]

return dados_papel

def filter_pb(df,kw,anti_kw = []):
    dados_papel = only_filter_pb(df,kw,anti_kw)

    dados_papel_group = dados_papel.groupby('date').count().reset_index()

    dados_papel_group['date'] = pd.to_datetime(dados_papel_group['date'], format='%Y-%m-%d').dt.strftime('%m/%d')
    dados_papel_group = dados_papel_group[dados_papel_group['date'] >= "03/01"].reset_index()

    return dados_papel_group

def manual_classifier(df,size):
    classifier = []
    tweets = []
    sample = df.sample(size,random_state=42)
    i = 1
    for tweet in sample['tweet']:
        print("tweet:",i)
        print(tweet)
        x = int(input("Classify the tweet:"))
        classifier.append(x)
        tweets.append(tweet)
        print()
        i+=1
    return classifier, tweets

def evaluate_classification(df):
    classification = np.array(df['classification_manual'])
    prediction = np.array(df['pred'])
    score = ['']*len(df['tweet'])

    for i in range(len(df['tweet'])):
        if classification[i]*prediction[i] == 1:
            score[i] = "true_positive"
        elif classification[i] == 1:

```

```

        score[i] = "false_negative"
    elif prediction[i] == 1:
        score[i] = "false_positive"
    else:
        score[i] = "true_negative"
    return score

```

```

papel_df = pd.read_csv("all_papel_mar_abr_2020.csv")
sabao_df = pd.read_csv("all_sabao_mar_abr_2020.csv")
alcool_df = pd.read_csv("all_alcool_mar_abr_2020.csv")
massas_df = pd.read_csv("all_massas_mar_abr_2020.csv")
doces_df = pd.read_csv("all_doces_mar_abr_2020.csv")

```

#Colhendo amostra de 200 tweets cada produto

```

sample_papel = papel_df.sample(200,random_state = 42).reset_index().drop("index",axis=1)
sample_sabao = sabao_df.sample(200,random_state = 42).reset_index().drop("index",axis=1)
sample_alcool = alcool_df.sample(200,random_state = 42).reset_index().drop("index",axis=1)
sample_massas = massas_df.sample(200,random_state = 42).reset_index().drop("index",axis=1)
sample_doces = doces_df.sample(200,random_state = 42).reset_index().drop("index",axis=1)

```

#Classificando manualmente

```

papel_class, papel_tweet = manual_classifier(papel_df,200)
sabao_class, sabao_tweet = manual_classifier(sabao_df,200)
alcool_class, alcool_tweet = manual_classifier(alcool_df,200)
massas_class, massas_tweet = manual_classifier(massas_df,200)
doces_class, doces_tweet = manual_classifier(doces_df,200)

```

```

sample_papel['classification_manual'] = papel_class
sample_sabao['classification_manual'] = sabao_class
sample_alcool['classification_manual'] = alcool_class
sample_massas['classification_manual'] = massas_class
sample_doces['classification_manual'] = doces_class

```

#definição das palavras chaves

```

papel_kw = ['estoc','compr','busca','armazen','mercado','super',
            'estog','corrida','armaz','escass','louc','esgot'],

```

Formatted: Portuguese (Brazil)

```

'acab','esvazi','atras','falt','neura','doido',
'sem','consum',"caga",'diarrei','hister','carrin']
sabao_kw = ['estic', 'estoc', 'estoaq', 'preço', 'farmácia', 'mercado']
alcool_kw = ['estoc', 'estoaq', 'feira', 'farmácia',
             'comprar', 'vender', 'mercado']
massas_kw = ['estoc', 'estoaq', 'feira', 'mercado', 'comprar', 'armaze']
doces_kw = ['estoaq', 'estoc', 'compra']
doces_anti_kw = ['pascoa', 'páscoa', 'ovo'] #palavras que serão excluídas p
or se tratarem de tweets da páscoa

#avaliação do filtro
sample_papel = only_filter_pb(sample_papel,papel_kw)
sample_papel['score'] = evaluate_classification(sample_papel)
print(sample_papel['score'].value_counts())

sample_sabao = only_filter_pb(sample_sabao,sabao_kw)
sample_sabao['score'] = evaluate_classification(sample_sabao)
print(sample_sabao['score'].value_counts())

sample_alcool = only_filter_pb(sample_alcool,alcool_kw)
sample_alcool['score'] = evaluate_classification(sample_alcool)
print(sample_alcool['score'].value_counts())

sample_massas = only_filter_pb(sample_massas,massas_kw)
sample_massas['score'] = evaluate_classification(sample_massas)
print(sample_massas['score'].value_counts())

sample_doces = only_filter_pb(sample_doces,doces_kw,doces_anti_kw)
sample_doces['score'] = evaluate_classification(sample_doces)
print(sample_doces['score'].value_counts())

#Aplicação do filtro em toda base e salvando dados para análise
papel_df_filtered = only_filter_pb(papel_df,papel_kw)
papel_df_filtered = papel_df_filtered[papel_df_filtered['pred']==1]
papel_df_filtered_grouped = filter_pb(papel_df,papel_kw)
papel_df_filtered.to_csv("all_papel_filtered.csv")

sabao_df_filtered = only_filter_pb(sabao_df,sabao_kw)
sabao_df_filtered = sabao_df_filtered[sabao_df_filtered['pred']==1]
sabao_df_filtered_grouped = filter_pb(sabao_df,papel_kw)
sabao_df_filtered.to_csv("all_sabao_filtered.csv")

```

Formatted: Portuguese (Brazil)
 Formatted: Portuguese (Brazil)

```
alcool_df_filtered = only_filter_pb(alcool_df,alcool_kw)
alcool_df_filtered = alcool_df_filtered[alcool_df_filtered['pred']==1]
alcool_df_filtered_grouped = filter_pb(alcool_df,alcool_kw)
alcool_df_filtered.to_csv("all_alcool_filtered.csv")

massas_df_filtered = only_filter_pb(massas_df,massas_kw)
massas_df_filtered = massas_df_filtered[massas_df_filtered['pred']==1]
massas_df_filtered_grouped = filter_pb(massas_df,massas_kw)
massas_df_filtered.to_csv("all_massas_filtered.csv")

doces_df_filtered = only_filter_pb(doces_df,doces_kw,doces_anti_kw)
doces_df_filtered = doces_df_filtered[doces_df_filtered['pred']==1]
doces_df_filtered_grouped = filter_pb(doces_df,doces_kw,doces_anti_kw)
doces_df_filtered.to_csv("all_doces_filtered.csv")
```