UNIVERSIDADE FEDERAL DE ALFENAS - UNIFAL/MG

GABRIEL BALDASSO

COMPORTAMENTO DO CONSUMIDOR BRASILEIRO DE ALIMENTOS NA PANDEMIA DE COVID-19: UM ESTUDO VIA MINERAÇÃO DE TEXTO

Alfenas/MG
2022

GABRIEL BALDASSO

COMPORTAMENTO DO CONSUMIDOR BRASILEIRO DE ALIMENTOS NA PANDEMIA DE COVID-19: UM ESTUDO VIA MINERAÇÃO DE TEXTO

Dissertação apresentada ao programa de Pós Graduação em Estatística Aplicada e Biometria da Universidade Federal de Alfenas UNIFAL/MG.
Orientador: Prof. Dr Eric Batista Ferreira.
Coorientador: Prof. Dr Anderson C. Soares de Oliveira

Alfenas/MG
2022

Sistema de Bibliotecas da Universidade Federal de Alfenas
Biblioteca Central

Baldasso, Gabriel .
Comportamento do consumidor brasileiro de alimentos na pandemia de covid-19 : um estudo via mineração de texto / Gabriel Baldasso. - Alfenas, MG, 2022.
92 f. : il. -

Orientador(a): Eric Ferreira Batista.
Dissertação (Mestrado em Estatística Aplicada e Biometria) - Universidade Federal de Alfenas, Alfenas, MG, 2022.
Bibliografia.

1. Consumidor. 2. Alimentos. 3. Covid-19. 4. Correlação . 5. Twitter. I. Batista, Eric Ferreira, orient. II. Título.

Ficha gerada automaticamente com dados fornecidos pelo autor.

GABRIEL BALDASSO

COMPORTAMENTO DO CONSUMIDOR BRASILEIRO DE ALIMENTOS DURANTE A PANDEMIA DE COVID-19: UM PROCESSO DE AMOSTRAGEM VIA MINERAÇÃO DE TEXTO

A Banca examinadora abaixo-assinada aprova a Dissertação apresentada como parte dos requisitos para a obtenção do título de Mestre em Estatística Aplicada e Biometria pela Universidade Federal de Alfenas. Área de concentração: Estatística Aplicada e Biometria.

Aprovada em: 18 de fevereiro de 2022.

Prof. Dr. Eric Batista Ferreira
Instituição: Universidade Federal de Alfenas - UNIFAL-MG

Prof. Dr. Crysttian Arantes Paixão
Instituição: Universidade Federal de Santa Catarina - UFSC

Prof. Dr. Sinézio Inácio da Silva Júnior
Instituição: Universidade Federal de Alfenas - UNIFAL-MG

Documento assinado eletronicamente por **Sinézio Inácio da Silva Júnior**, **Presidente**, em 20/02/2022, às 17:37, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Eric Batista Ferreira**, **Professor do Magistério Superior**, em 03/03/2022, às 16:04, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **CRYSTTIAN ARANTES PAIXÃO**, **Usuário Externo**, em 03/03/2022, às 20:13, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

A autenticidade deste documento pode ser conferida no site https://sei.unifal-mg.edu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0678507** e o código CRC **C63F1F7A**.

AGRADECIMENTOS

Agradeço a Deus, por me proporcionar realizar este trabalho de dissertação, proporcionando saúde nestes tempos difíceis de pandemia.

Ao meu orientador Eric, por todo o suporte e por compartilhar um pouco do seu conhecimento comigo. Ao meu coorientador Anderson, por aceitar esta proposta e me aceitar em seu grupo de pesquisa.

Aos meus colegas do mestrado, pela parceria e ajuda ao longo destes dois anos de trabalho.

À minha família, que me sustentaram em todo este período de pesquisa, sendo minha fonte de motivação e ânimo.

À minha noiva Dani, por todo suporte atenção, motivação e paciência neste período de pesquisa e escrita.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001". Assim como a Fundação de Amparo à Pesquisa do Estado de Minas Gerais (Fapemig), pelo financiamento deste trabalho.

LISTA DE FIGURAS

LISTA DE TABELAS

SUMÁRIO

.

RESUMO

A pandemia de COVID-19 com início no Brasil no final de fevereiro de 2020, provocou uma das maiores crises sociais e econômicas já vistas na última década. Como forma de conter a propagação do vírus, medidas de isolamento e quarentena foram anunciadas, como o fechamento do comércio, o que provocou uma das maiores crises no setor alimentício. Restaurantes, bares e lanchonetes viram seus negócios fecharem as portas e alguns irem à falência. Neste ambiente, os preços dos alimentos sofreram uma inflação de 14,1% no ano de 2020, tendo alguns produtos como a carne, que dobraram de preço. Logo, com a alta dos preços dos alimentos, houve uma mudança no perfil do consumidor de alimentos. Logo este trabalho tem como objetivo: Fornecer informações sobre o consumidor de alimentos um ano após o início da pandemia de COVID-19 no Brasil. Para isso, este trabalho foi dividido em dois artigos onde o primeiro busca verificar as mudanças na compra de alimentos e nos locais de compra, além de verificar se existe associação entre os termos da pandemia e alimentos em publicações no Twitter. No segundo artigo, busca-se verificar a correlação entre a frequência de publicações sobre: pandemia, alimentos e estabelecimentos no Twitter com os dados reais da pandemia no Brasil (casos, mortes e vacinados), no período da primeira e segunda onda da COVID-19. No primeiro artigo os resultados da pesquisas por Crise + alimentos no Twitter mostraram que os consumidores estão ansiosos e fazem uso do café neste período. Confirmando a associação entre pandemia e alimentos no Twitter. O isolamento também fez com que os consumidores em sua grande maioria passassem a adotar pedidos por delivery, reduzindo a frequência de compra de duas vezes por semana para semanal. Os alimentos mais comprados pelos participantes foram frutas e arroz e os menos comprados Carne e Refrigerante. Doces, chocolates e sanduíches foram associados como ajuda com o estresse. Logo, os impactos gerados pela COVID-19 continuam afetando os hábitos alimentares dos brasileiros, sendo o impacto financeiro o principal causador dessa mudança alimentar. Já no segundo artigo foi verificado que a taxa de casos da COVID-19 no Brasil apresentou correlação significativa com os termos "pandemia"(0,95), "restaurantes"(0,80) e "frango"(0,85). Além da frequência de publicação dos tuítes serem influenciadas por notícias polêmicas publicadas neste período. Desta forma, conclui-se que após um ano do início da pandemia no Brasil, os seus impactos ainda são sentidos pelos brasileiros, principalmente no âmbito alimentar. Onde o Twitter tem se mostrado ser um ambiente correlacionado com dados da vida real, mesmo quando analisado por grandes períodos de tempo.

Palavras-chave: Consumidor; Alimentos; COVID-19, Correlação; Twitter

ABSTRACT

The COVID-19 pandemic, starting in Brazil in February 2020, caused one of the biggest social and economic crises ever seen in the last decade. To contain the spread of the virus, isolation, and quarantine measures were announced the closing of the commerce, which provoked one of the biggest crises in the food sector. Restaurants, bars, and snack bars saw their businesses close, and some went failed. In this context, food suffered inflation of 14.1% in 2020, with some products such as meat, which doubled in price. With the rise in prices, there were changes in the food consumer. Therefore, this work aims to: Provide information on food consumers one year after the COVID-19 pandemic in Brazil. For this, this work will be divided into two articles where the first seeks to verify changes in food purchases and places of purchase, in addition to verifying whether there is an association between the terms of the pandemic and food in publications on Twitter. In the second article, we seek to demonstrate the correlation between the frequency of publications on pandemic food and establishments on Twitter with the actual data of pandemic in Brazil (cases, deaths, and vaccinated), in the period of the first and second waves of COVID-19. In the first article, the research results by Crisis + food on Twitter showed that consumers are anxious and are using coffee in this period. They were confirming the association between pandemic and food on Twitter. Isolation has also led consumers in their vast majority to adopt orders for delivery, reducing the frequency of purchase from twice a week to weekly. The foods most bought by the participants were fruits, and rice and the least purchased were Meat and Soft Drinks. Candy, chocolates, and sandwiches have been associated with stress relief. Therefore, the impacts generated by COVID-19 continue to affect the eating habits of Brazilians, with the financial implications being the leading cause of this dietary change. In the second article, it was verified that the rate of COVID-19 cases in Brazil was significantly correlated with the terms "pandemic"(0.95), "restaurants"(0.80), and "chicken"(0.85). In addition to the frequency of publication of Tweets being influenced by controversial news published during this period. Thus, it concluded that one year after the beginning of the pandemic in Brazil, its impacts are still felt by Brazilians, mainly in the food field where Twitter is an environment correlated with real-life data, even when analyzed over long periods.

Key-words: Consumer; Food; COVID-19; correlation; Twitter

1 INTRODUÇÃO GERAL

O novo coronavírus, (SARS-CoV-2), que teve início no Brasil no final de fevereiro de 2020, classificado como uma infecção respiratória grave com potencial pandêmico, trouxe ao Brasil uma nova realidade de convívio social, tornando necessário aprender a viver em uma pandemia. Com o intuito de diminuir a propagação do vírus, foi exigido pelas autoridades governamentais que as pessoas utilizassem máscaras e ficassem mais tempo em casa e mantivessem o isolamento social, até a chegada das vacinas e a imunização da população.

O isolamento social fez com que o comércio e as atividades presenciais fechassem as portas, trazendo mudanças no cotidiano e comportamento das pessoas. Mudanças estas que foram benéficas para vendas *on-line* e consumo de *delivery*. Contudo, atividades como: bares, lanchonetes e restaurantes, viram seus negócios fecharem as portas da noite pro dia, causando um grande prejuízo para este setor alimentício.

O impacto que a crise da COVID-19 gerou no ramo alimentício, foi sentido no preço dos alimentos que dispararam com os anúncios de *lockdown* e fechamento dos comércios. Segundo dados do Centro de Estudos Avançados em Economia Aplicada (CEPEA), o setor alimentício apresentou uma inflação de 14,1% no ano de 2020, sendo a agropecuária praticamente a responsável por todo esse aumento na inflação, devido aos 45,5% de crescimento apresentado no preço dos produtos deste setor. A explicação a este aumento está na alta demanda por estes alimentos, na valorização do dólar, na restrição ao comércio e a alta demanda por frete devido às medidas restritivas impostas. A inflação do setor de alimentos só não foi mais alta devido à recessão (queda de emprego e renda) apresentada no ano de 2020, o que segurou a taxa em 14,1% (BARROS, et al 2021).

O aumento no preço dos alimentos e as medidas de segurança necessárias para consumi-los neste período pandêmico, ocasionou uma mudança alimentar por parte dos consumidores. Segundo Durães et al, 2020 em sua revisão bibliográfica, a busca por uma dieta saudável e variada apresentou crescimento neste período de isolamento, assim como a redução de bebidas alcoólicas, contudo houve padrões opostos de respostas as pesquisas com relação ao consumo de produtos in natura e ingestão de vegetais e frutas, porém o aumento no consumo de alimentos foi presente neste período. Idas aos supermercados apresentaram reduções de frequência, já a procura por alimentos em pequenos estabelecimentos para evitar aglomerações registrou aumento neste período. A troca de idas aos restaurantes e bares, por pedidos de *delivery* e consumo em casa, também esteve presente neste período, assim como o uso das mídias digitais na procura por ofertas e promoções de alimentos.

As redes sociais não ficaram fora destas mudanças, de forma que as pessoas passaram a compartilhar com maior frequência a sua rotina, como uma forma de socialização virtual. Somente o *Twitter* apresentou um crescimento de usuários de 20% de 2019 para 2020, e

um aumento de 15% no número de publicações segundo pesquisas do *WeAreSocial*, 2020. O *Twitter*, ficou famoso em sua forma de publicação de texto curto de até 280 caracteres conhecido como tuíte. Este meio, passou a ser objeto de estudo de diversas áreas do conhecimento, pela riqueza dos dados textuais e pela rapidez de comunicação e variedade de conteúdo presentes em um só local, como: (vídeos, fotos, links, textos, propagandas, etc). Porém, extrair informações desses ambientes e analisar estes dados têm sido um desafio nos dias atuais, devido à maneira desestruturada que se encontram esses dados no *Twitter*.

Logo, procurar entender como se deu as mudanças no consumo de alimentos um ano após o inicio da pandemia da COVID-19 e seus impactos no comportamento dos consumidores de alimentos tanto de forma presencial como na rede social Twitter, é de suma importância para a área de nutrição e saúde da população.

Neste sentido, este trabalho tem como objetivo fornecer informações sobre o consumidor de alimentos no período da pandemia de COVID-19, com o intuito de verificar se houve mudanças nos locais de compras de alimentos durante a pandemia e quais os alimentos que foram mais e menos comprados neste período. Investigou-se também se houve mudança na frequência de publicações no Twitter a respeito de alimentos e atividade física por usuários brasileiros no período inicial da pandemia, além de comparar se existe correlação entre a frequência destas publicações e os dados reais da pandemia, com a taxa de: casos, mortes e vacinados.

Este trabalho está organizado em dois artigos. No primeiro artigo foi realizada uma pesquisa online para obter dados a respeito da mudança alimentar e dos locais de compra de alimentos dos brasileiros no período da pandemia de COVID-19. Uma coleta no Twitter foi realizada com os principais termos destacados na pesquisa online, juntamente com as palavras-chave: (covid, crise, pandemia, quarentena e vírus) no período de janeiro a junho de 2021, com o intuito de verificar se existe associação entre os termos. No segundo artigo, buscou-se verificar se os dados da pandemia no Brasil: (casos, mortes e vacinados) fornecidos pela ministério da Saúde, são correlacionados à frequência de publicações no *Twitter* sobre os assuntos: Pandemia, Alimentos e locais.

2 REFERENCIAL TEÓRICO

2.1 COMPORTAMENTO DO CONSUMIDOR

Comportamento do consumidor é definido por Engel, Kollat e Miniard (1990) como sendo um conjunto de atividades físicas e mentais executadas pelo consumidor que levam a: buscas, escolhas e decisões em adquirir e utilizar produtos e serviços como forma de saciar uma necessidade latente.

O início deste estudo sobre o comportamento do consumidor, se deu na década de 50, por meio da área da psicologia social, onde a ideia era levantar teorias e modelos para buscar compreender e prever o comportamento dos consumidores, com o intuito de entender suas motivações (JACOBY et al, 1998).

É importante mencionar que o comportamento do consumidor é uma combinação da consciência de compra do cliente, combinada com motivadores externos para resultar em uma mudança no comportamento do consumidor (SOLOMON, 2016). Isto tem atraído a atenção de muitos pesquisadores e gerentes de empresas em todo o mundo, pela sua importância fundamental nos resultados de companhias e empresas que passaram a adotar o comportamento do consumidor em seus modelos de negócios (SHETH; MITTAL; NEWMAN, 2001).

Ao prever o comportamento dos consumidores, uma empresa pode compreender as suas necessidades e trabalhar para atender a estas expectativas de seus clientes (JACOBY et al, 1998). jaboby et al (1998) complementa que empresas que buscam entender o consumidor tem alta probabilidade de manter sua prosperidade e atingir seus objetivos de longo prazo.

Com o intuito de entender melhor o comportamento do consumidor frente a diversos estímulos que este recebe no seu dia-a-dia, será realizado uma apresentação dos tópicos: cultura, grupos, gênero e idade, que são muito comentados por Solomon (2016) em seu livro, por serem tópicos de alto poder de influencia no comportamento dos consumidores.

2.1.1 Cultura

Segundo os autores, Blackwell, Miniard e Engel (2005) a influência cultural é um dos principais pontos que afetam o processo de decisão do consumidor, o que acarreta na sua forma de comprar e utilização dos serviços disponíveis no mercado.

Cultura, pode ser definida em todas as formas que os participantes de uma mesma sociedade aprendem e compartilham entre si durante suas vidas, como: regras, normas, valores, habilidades, materiais e comportamentos (SHETH; MITTAL; NEWMAN, 2001). Ou seja, Sheth, Mittal e Newman (2001) destacam que a cultura é um conjunto de crenças, valores e costumes que são aprendidos e compartilhados entre as pessoas de uma mesma sociedade.

Este fato faz com que os consumidores de diferentes culturas e sociedades tendem a associar um mesmo produto a diferentes características e valores, ou valorizar os mes-

mos aspectos de um produto de formas distintas, influenciados pela sua cultura local (BLACKWELL; MINIARD; ENGEL, 2005). Ainda segundo os mesmos autores, os locais de compra de alimentos também costumam se diferenciar com a cultura, pois certos tipos de produtos e serviços disponíveis nos mercados são únicos de sua região (BLACKWELL; MINIARD; ENGEL, 2005).

Nesta mesma linha Sheth, Mittal e Newman (2001) acrescentam que a cultura dos consumidores afeta a sua forma de busca por alimentos, como por exemplo: comprar pão na padaria ou supermercado, verduras em supermercados, varejões ou feiras, carne em açougues ou supermercados.

As marcas dos produtos também são influenciadas pela cultura, pelo fato de certas marcas serem regionais ou até locais e passam a ser costume do povo, levando os consumidores a repeti-las nas próximas compras (SHETH; MITTAL; NEWMAN, 2001).

2.1.2 Grupos

Um grupo é a definição de um conjunto de pessoas que se relacionam entre si e compartilham das mesmas necessidades e objetivos (SHETH; MITTAL; NEWMAN, 2001). O estudo dos grupos para entender o comportamento do consumidor está ligado à questão do poder de influência que um grupo pode ter sobre seus integrantes e sobre suas decisões de compra (WARD; REINGEN, 1990).

Grupos que possuem normas, crenças e regras e que são utilizados para servir de guia para tomada de decisões, são denominados por Olmstead (1962) como sendo grupos de referência. Um exemplo de um grupo de referência é a família, que tem como caráter construir valores e atitudes individuais em seus integrantes, atitudes estas que tendem a ser repetidas no futuro (SHETH; MITTAL; NEWMAN, 2001). Situações em que o consumidor não conhece determinado produto, as informações que o grupo compartilhar irão exercer influência significativa na hora da sua decisão de compra (BLACKWELL; MINIARD; ENGEL, 2005). O consumidor pode ser influenciado pelos grupos que ele admira a realizar compras de determinadas marcas específicas com o intuito de se associar a este determinado grupo (OLMSTEAD, 1962).

Cabe ressaltar que a influência por determinado grupo se dá quando maior for a sua exposição com seus membros e a abertura de ideias e valores com os integrantes dos grupos (SHETH; MITTAL; NEWMAN, 2001).

2.1.3 Sexo

O sexo é uma forte variável que exerce influência sobre o comportamento do consumidor e suas decisões na hora da compra (GROHMANN et al, 2010).

Segundo o autor Grohmann et al (2010) homens tendem a ser mais centrados e objetivos no processo de decisão de compra e são menos influenciados pela pressão social e fatores externos. Contudo, o mesmo autor relata que as mulheres são mais influenciadas

pelo meio externo e sofrem com a aceitação e pressão social, o que influencia de forma direta a sua tomada de decisão na hora de comprar (GROHMANN et al, 2010).

De forma geral, homens preferem metas mais individuais e mulheres dão prioridade para metas sociais e relações harmoniosas (UNDERHILL, 2009).

Homens preferem ir a supermercados com menor fluxo de pessoas, com o objetivo de não esperar em filas, as mulheres, gostam de locais de compra seguros, de fácil acesso e bonitos, preferindo redes varejistas (Grohmann et al, 2010). Mulheres tendem a ser mais impulsivas na hora de realizar compras, a proximidade com colegas a estimula a gastar mais, assim como a atração emocional por determinado produto (Grohmann et al, 2010).

Por último, o Homem tende a se basear em dados e números com o objetivo de influenciar a decisão de compra da mulher, já a mulher exerce pressão psicológica e emocional sobre o homem para poder decidir em suas decisões de compra (UNDERHILL, 2009).

2.1.4 Idade

Segundo os autores Mowen e Minor (2003) indivíduos que possuem idades semelhantes tendem a obter comportamentos iguais sobre as suas decisões de compras individuais. Esta formação de grupos estabelecidos pelas idades podem ser representados por: adolescentes, jovens, adultos e idosos (MOWEN; MINOR, 2003).

Os grupos que mais se destacam são os adolescentes e jovens, que têm suas decisões de compras na grande maioria guiada pela decisões dos pais e responsáveis, tirando os que já possuem forma autônoma de renda, que levam em consideração o preço no sua tomada de decisão de compra (SOLOMON, 2016).Indivíduos mais jovens tendem a realizar mais compras por impulso do que pessoas mais velhas, obtendo assim um padrão oposto entre idade e comportamento de compra impulsiva (MOWEN; MINOR, 2003).

Segundo trabalhos de Andrade, Oliveira e Antonialli (2004), idosos tendem a se preocupar mais com aspectos nutritivos e benéficos para a saúde, na hora de realizar as suas compras de alimentos, enquanto jovens pouco se importam com essa questão. Idosos costumam ficar a maior parte do tempo assistindo televisão, de forma que este aparelho é o melhor canal para atingir este grupo de consumidores (ANDRADE; OLIVEIRA; ANTONIALLI, 2004).

Jovens que já possuem emprego e continuam morando com os pais, possuem um alto potencial de gastos, por sua renda ser pouco comprometida com os afazeres domésticos que estão atribuídos aos pais (FERREIRA; REZENDE; LOURENÇO, 2011). Os mesmos autores também salientam que as grandes mudanças dos hábitos de compra dos consumidores se dão com a alteração da estrutura familiar, seja por: (casamento, nascimento, falecimento ou divórcio).

2.1.5 Decisão de compra do consumidor

O modelo de decisão de compra do consumidor foi adaptado do trabalho de Blackwell, Miniard e Engel (2005), dividindo-o em três processos: Identificação de uma necessidade; busca de informação e compra.

O primeiro processo é a identificação de uma necessidade, que consiste em um espaço entre o estado de incômodo do consumidor e o estado de conforto desejado (HAWKINS; MOTHERSBAUGH; BEST, 2007). Ou seja, é a confirmação para o consumidor de que é necessário realizar uma compra de determinado item, para eliminar ou reduzir o incômodo sentido segundo Blackwell, Miniard e Engel (2005).

Um exemplo deste processo é a sede que é o estado de incômodo do consumidor, quando este atinge determinado nível, um desconforto é percebido, levando a ação de ir tomar algo e logo depois o estado de conforto desejado retorna ao normal (BLACKWELL; MINIARD; ENGEL, 2005).

Segundo o autor Solomon et al (2017), as necessidades surgem de forma natural e por meio de estímulos internos, contudo as empresas podem vir a ganhar a atenção dos indivíduos, com dois estímulos os inativos e os ignorados, estimulando o reconhecimento de que a aquisição de seus produtos será satisfeita.

O segundo processo é a busca de informações, que é automaticamente ligada quando o consumidor decide a necessidade de comprar algo. Este processo consiste na busca e pesquisa por ambientes que o facilitam na tomada de decisão da compra (SOLOMON, 2016).

A busca por informações tem também como objetivo, auxiliar o consumidor na escolha do item, preço, formas de pagamento e frete ou localidade da loja (HAWKINS; MOTHERSBAUGH; BEST, 2007). Esta busca pode se dar de forma interna, baseada na memória, ou externas baseadas em meios externos (SOLOMON, 2016).

Quando a busca interna de qual item comprar, o preço a pagar e se possui frete a pagar ou não, não é lembrada, a forma externa é acionado, procurando as mesmas ideias, porém agora em meios externos, como sites, computadores, pessoas etc (BLACKWELL; MINIARD; ENGEL, 2005).

O terceiro e último Processo é a compra, nesta fase surge a necessidade de escolher de forma mais detalhada entre as opções listadas no tópico anterior, para chegar a melhor opção que corresponda a sua demanda (BLACKWELL; MINIARD; ENGEL, 2005). Nesta etapa, é normal que os consumidores adotem um comportamento simplista devido a alta dificuldade de escolha do produto, os itens removidos podem ser, ou por defeito, preço alto, marca ruim etc. Com o produto escolhido, chega-se a hora de efetuar a compra e realizar o pagamento (SOLOMON, 2016).

2.1.6 Crises

Neste tópico será destacado o acontecimento da crise financeira de 2008 nos Estados Unidos, denominada pela crise do subprime. Nesta ocasião bancos e instituições financeiras colapsaram em 2007 devido a uma bolha imobiliária, causada pelo aumento repentino nos valores dos imóveis, sem ser acompanhado pelo aumento de renda da população, gerando assim um efeito cascata em outras economias do mundo todo (AMALIA; IONUT, 2009).

Inflações sobre o preço das commodities foram observadas, o que levou os consumidores a decidirem entre comprar o necessário ou o desejado. Quedas acentuadas no mercado de ações fizeram com que os consumidores entrassem em pânico, gerando um momento de incertezas (MANSOOR; JALAL, 2011).

Na mesma linha Manssor e Jalal (2011), destacam que os preços dos produtos aumentaram substancialmente o que prejudicou de forma significativa as classes sociais mais baixas, ficando reféns de ações governamentais e ajudas solidárias. Isto mudou o comportamento dos consumidores que passaram a se preocupar com seus empregos e guardar as suas economias.

Compras ligadas ao lazer e ao entretenimento foram evitadas, além da troca de produtos mais caros por similares mais baratos, se preocupando assim mais com o preço do que com a qualidade. A busca por promoções e descontos se torna uma realidade diária (GILKEY JR; CLARK, 2015).

Ou seja, segundo Amalia e Ionut (2009), o consumidor irá mudar sua forma de consumir se perceber que está passando por uma mudança na sua situação econômica. Estas mudanças afetam o consumidor de forma material e psicológica, pela alta preocupação com o dinheiro. Produtos mais caros são deixados de lado, mesmo que o consumidor possa pagar por isso, e o comportamento do consumidor passa a ter uma percepção de risco.

Os consumidores optaram a se entreter com maior frequência em casa, reduziram gastos de luxo, e passaram a gerenciar mais de perto seus orçamentos e dívidas (GILKEY JR; CLARK, 2015).

Em meio a crise, os consumidores podem ser separados em quatro grupos conforme o nível de risco que cada grupo aceita neste momento de crise, conforme coloca Mansoor e Jalal, (2011) são eles:

Consumidor em Pânico são consumidores que se sentem em alto risco e em uma situação estressante. São normalmente avessos a riscos elevados e tendem a ter reações exageradas em momentos de crises. Destes consumidores em pânico espera-se que mudem a sua forma de consumo, economizem no supermercado, eliminem compras desnecessárias, busque sempre produtos pelo melhor preço.

Consumidor Prudente são consumidores planejados que não se sentem em alto risco,

por não estarem expostos a riscos elevados. Estes consumidores irão planejar seus gastos, trocar marcas de produtos se necessário, porém não sentem necessidade de cortar gastos imediatos. São sempre muito bem informados com o que está acontecendo.

Consumidor Preocupado são consumidores com alta percepção ao risco, entendem o que está acontecendo e geralmente assumem riscos proporcionais ao momento. Estes consumidores ficaram de olho em fazer "bons negócios"conforme a relação de risco ia diminuindo. Costumam ser leais às marcas e planejam suas compras para não faltar nada.

Consumidor Racional são consumidores que não se sentem em risco e consideram isso por não estarem expostos a isso. Evitam saber informações voltadas à crise e tentam manter seu comportamento como se nada tivesse acontecendo. Estes consumidores irão reduzir seus gastos, continuarão com suas compras normais e experimentando produtos novos no mercado.

Logo, segundo Manssor e Jalal (2011) o comportamento do consumidor será altamente afetado em um ambiente de crise, por este momento muita das vezes mudar seus planos e projetos para o futuro. Já, Amalia e Ionut (2009) relatam que esse ambiente instável irá obrigar os consumidores a se reinventarem buscando: promoções, preço baixo, liquidações etc. O que espera-se é que a crise passe e que o consumo retorne ao normal, assim como o comportamento do consumidor (GILKEY JR; CLARK, 2015).

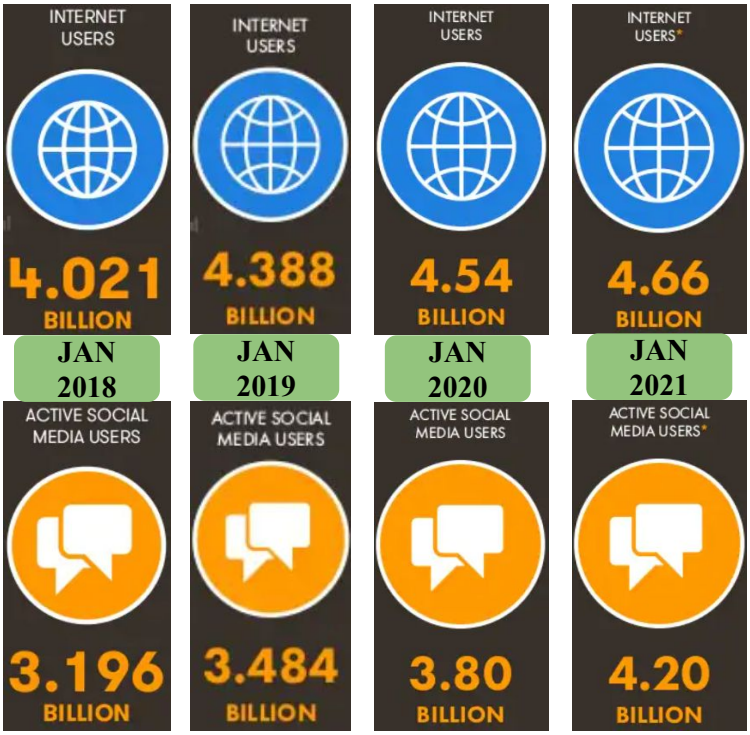
2.2 Internet e Redes Sociais

Desde a criação do primeiro telégrafo por volta de 1830, não se esperava que a tecnologia iria alçar voos tão altos com relação a comunicação. Nos anos 1960 nasceu a primeira comunicação em rede entre computadores, dando-se inicio ao surgimento comercial da internet nos anos 90 nos Estados Unidos. Desde então, esta tecnologia de acesso a rede não parou mais de ganhar investimento e seu crescimento alcançou regiões, casas e pessoas e hoje em dia é quase impossível viver sem a internet (SIQUEIRA, 2008).

A expansão explosiva da internet é percebida em números: no inicio dos anos 1990 existia pouco mais de 400mil usuários; 10 anos depois em 2000 este número já superava os 760 milhões de usuários em todo o mundo. Este aumento se deu de forma exponencial devido a alta versatilidade e funcionalidade que a internet possibilita desde as empresas até as crianças (COSTA, 2018). Em 2018 a internet já alcançava mais de 4 bilhões de usuários, com mais de 3 bilhões de usuários utilizando as redes sociais.

Na figura 1 pode ser observado o aumento expressivo de usuários da internet e de redes sociais ano a ano de 2018 a 2021 (SOCIAL, 2021). Este aumento de usuários da internet

Figura 1 – Aumento do número de usuários da internet e das redes sociais de 2018 a 2021.



Fonte: WeAreSocial.com

possibilita um aumento na experiencia on-line, na busca de conteúdos e educação. Segundo Sima et al, (2020) nos dias atuais, uma pessoa não consegue pensar em desenvolver algo sem se conectar a um aparelho celular ou computador com internet.

Esta revolução no cotidiano das pessoas, também foi percebida na rápida aceitação do público para as redes sociais, por sua facilidade de acesso a conteúdos variados e rápida interação com a comunidade (COSTA, 2018).

A internet e as redes sociais estão revolucionando a forma que os consumidores interagem com o mercado e uns com os outros. Isto faz com que os consumidores se tornem usuários afincos das redes sociais, por haver uma vontade de que nossos amigos e seguidores saibam o que nós sabemos (SIMA et al, 2020). Segundo trabalhos de Solomon (2016), usuários de redes sociais tendem a compartilhar mais noticias boas do que ruins, temos mais vontade de falar sobre nós mesmos. O mesmo autor coloca que 80% das publicações do Twitter giram em torno do próprio autor (SOLOMON, 2016).

Os usuários de uma rede social, realiza publicações diárias a fim de fomentar o processo básico do boca a boca que antes erá realizado de forma presencial (COSTAS, 2018). Segundo o site de busca Yahoo, 40% dos usuários que ficam online o dia todos exercem influência sobre as compras de seus amigos e seguidores em uma proporção de 2 para 1 (SIMA et al, 2020).

Sites de internet assim como redes sociais possuem algumas características em comum que podem ser definidas como:

Ficam melhores a medida que o número de seus usuários crescem.

Seu crescimento é medido pelo número de visitantes.

São gratuitos e estão sempre em evolução.

Somente mostram conteúdos que é de interesse do usuário.

Logo, uma definição de rede social pode ser dada como um conjunto de nós socialmente relevantes que estão conectados por uma ou mais relações (COSTAS, 2018). Os nós podem ser definidos como os membros presentes nesta rede e as relações como as ligações que cada membro tem entre si, seja por gosto, amizade, ou algum tipo de interesse em comum. Ou seja, as relações são atividades que os membros de uma rede têm em comum (SOLOMON, 2016).

O grande sucesso das redes sócias esta na forma de interagir seus usuários em comunidade, e conseguir passar uma sensação de proximidade um com o outro, e sempre prosperam quando seus membros são participativos, colaboradores e recrutam novos colegas à ela (SIMA et al, 2020).

2.2.1 Twitter

O twitter possui atualmente mais de 330 milhões de usuários ativos em todo o mundo (OSMAN, 2021). Fundado em julho de 2006 nos Estados Unidos por Jack Dorsey, como uma versão de um *microblog* onde os usuários poderiam interagir em tempo real por meio de postagens *(tweets)* de até 280 caracteres (KULLAR, 2020).

O que possibilita a interação entre seus integrantes por meio de compartilhamento de notícias e leitura de conteúdos entre os seus seguidores. Com o avanço da internet e a popularização do smartphone, facilitou ainda mais o uso do twitter, que ficou mundialmente conhecido pela massiva presença de autoridades políticas e famosos, que gostam de realizar seus anúncios por meio da plataforma (MINOT, 2020).

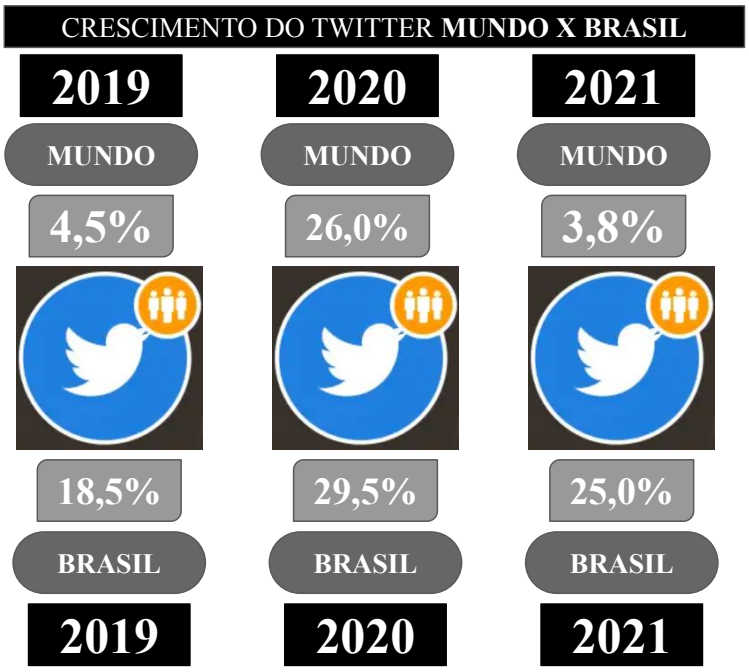
Esta inovação de utilizar uma rede social com cunho político, ocorreu pela primeira vez na corrida presidencial de 2008 nos Estados Unidos, onde o candidato Barack Obama utilizou deste recurso para ganhar popularidade entre a população americana conquistando mais de 100.000 novos seguidores (POBIRUCHIN; ZOWALLA; WIESNER, 2020).

Os conteúdos no Twitter, são conhecido como tuítes e podem ser compartilhados de alguma conta específica, sendo então chamados de retuítes. Os usuários do *Twitter* possuem métricas para a sua conta, como o número de seguidores e para as publicações são utilizados os: retuítes, curtidas e respostas (KULLAR, 2020).

Se tornando assim um sucesso em todo o mundo, apresentando um grande crescimento no número de usuários nos últimos anos (POBIRUCHIN; ZOWALLA; WIESNER, 2020).

Segundo site de pesquisa *We are social*, com suas pesquisas voltadas a internet e redes sociais dos anos de 2019 a 2021, o *Twitter* no Brasil apresentou crescimento de contas

Figura 2 – Aumento do número de usuários do Twitter de 2019 a 2021.



Fonte: WeAreSocial.com

ativas superior ao crescimento da plataforma no mundo todo, como pode ser observado na figura 2.

O que pode ser explicado pela calorosa aceitação que o *Twitter* possui no Brasil quando se trata de assuntos políticos.

A pandemia da COVID-19 fez com que o *Twitter* ganhasse 26% de crescimento entre seus usuários, devido as medidas restritivas de isolamento, o que fez com que a população passasse a utilizar mais a internet e as redes sociais.

O uso do *Twitter* por jornalistas representam certa de 25% das contas ativas verificadas. Ainda segundo estudos do Osman (2021) 83% dos lideres mundias, possuem contas ativas no *Twitter* e as utilizam como forma de divulgação de noticias.

Hoje em dia, o *Twitter* é famoso no mundo todo pela sua praticidade de obter informações, e pelo constante uso desta plataforma para divulgações de informações relevantes por autoridades e famosos.

Contudo, empresas de diversos ramos no mundo todo são adeptas ao *Twitter* por acreditar na forma como o comunidade se comporta sendo simples e direta com a informação e principalmente pela liberdade que o usuário possui de se expressar nesta plataforma.

O twitter permite que as publicações e informações dos perfis dos seus usuários sejam extraídas e mineradas, por meio da API (*Application Programming Interface*) o que tem atraído a atenção de empresas e pesquisadores com interesse no comportamento dos possíveis consumidores que utilizam desta rede social (DALMORO *et al*, 2010).

Este fato fez com que o *Twitter* se destacasse das demais redes sociais, e se tornasse

um ambiente de extremo interesse pelos pesquisadores. O uso da rede social twitter como fonte de dados para pesquisa, deu-se início, com o foco de buscar entender, qual a relação entre o que é postado nesta rede com o dia-a-dia do usuário.

Logo, com o lançamento da API e pacotes que possibilitasse seu uso, inúmeros artigos são publicados todos os anos, utilizando dados do *Twitter*, com o objetivo de análise de texto, mineração de texto e visualização de dados.

A análise de sentimentos é um tipo de análise muito explorada em dados retirados do *Twitter* Contudo, existem várias outras abordagens que podem ser utilizadas no twitter como a análise dos trends (assuntos mais pesquisados e comentados no momento), análise de menções, nuvem de palavras entre outras análises que podem ser realizadas com o conjunto textual presente no twitter.

O twitter é utilizado hoje em dia, por bancos, hospitais, agências de marketing, industrias, delivery etc. Entre outros tipos de serviços que enxergam uma forma de contato com os seus clientes por meio da plataforma.

2.2.2 Mineração de texto

O avanço tecnológico na área da computação e comunicação e o aumento do uso das redes sociais pelas pessoas nos últimos anos tem favorecido ao acúmulo de informações em meios digitais. Somente o *Twitter* publica mais de 450.000 tuítes a cada minuto o que significa mais de 600 milhões de tuítes por dia (OSMAN, 2021). Como os usuários do *Twitter* no Brasil em 2021 representa aproximadamente 5% de todos os usuários do *Twitter* do mundo, isso significa que o Brasil possui uma média diária de 30 milhões de tuítes publicados por dia.

Ao todo, esse acumulado de informações presentes na mídia, denominamos como um ambiente *big data* que em suma refere-se a ambientes que possuem características de Volume + Variedade + Velocidade de dados (TAURION, 2013).

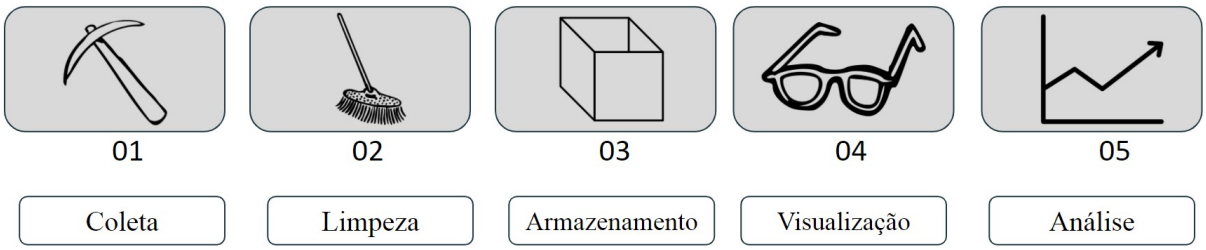
A mineração de dados tecnicamente se resume em utilizar de recursos computacionais para localizar padrões e correlações nos dados que deseja-se estudar, ou seja busca-se extrair informações por padrões de dados que não foram possíveis de serem extraídos de forma manual, aplicando análises estatísticas, aprendizagem de máquina e tecnologias de inteligência artificial para obter tal resultado (PUSHPAM, JAYANTHI, 2017).

A mineração de texto é específica para coletar e extrair dados de caráter textuais de diferentes localidades, sejam elas um banco de dados, livros, página web, rede social etc (RAVINDRAN, GARG, 2015). Dados textuais, segundo o mesmo autor, são uma fonte rica em informação que não são triviais de serem encontradas devido a sua forma não estruturada de armazenamento.

Dados textuais não estruturados, são aqueles que não possuem um esquema pré-definido e são fortemente evolutivo e sua estrutura pode ser considerada irregular dependendo muita das vezes da organização dos dados (PUSHPAM, JAYANTHI, 2017). A

distinção entre dados e estrutura não é fácil de ser identificada, na prática, dados textuais estão misturados com: links, fotos, vídeos, propagandas etc, o que dificulta a localização e mineração destes dados (RAVINDRAN, GARG, 2015).

Figura 3 – Processo de 01 a 05 para a mineração de texto.



Fonte: Elaboração Própria

A mineração de texto pode ser definida em 5 passos: Coleta, Limpeza, Armazenamento, Visualização e Análise (PUSHPAM, JAYANTHI, 2017).

Coleta A coleta é a fase em que o programa utilizado terá acesso ao conjunto de dados, sejam eles: sites, livros, documentos, redes sociais etc, a extração destes textos podem ser realizadas de forma total ou de forma específica, um exemplo de extração total consiste na retirada de todo o conteúdo textual presente no local a ser minerado, sem distinção de conteúdo, já a extração de forma especifica pode ser realizada por meio de uma palavra-chave, todas as sentenças que possuir a palavra-chave sera extraída pelo programa (FILHO, 2014).

Limpeza Limpeza e armazenamento é uma etapa primordial na mineração de texto por trazer estrutura ao conjunto de dados, o que facilita posteriormente no processo de análise (FILHO, 2014). O objetivo da limpeza dos dados é remover tudo que for identificado como irrelevante para a análise final dos dados, como por exemplo: stopwords, caracteres não alfabéticos, pontuação, links, emoticons, espaços em branco etc (ZAFARINI, ABBASI, LIU, 2014).

Armazenamento O armazenamento dos dados textuais coletados, são fundamentais devido que a ele será recorrido quando for necessário repetir alguma análise ou procedimento que utiliza o conjunto minerado. Logo, é de costuma salvar o conjunto textual em planilhas e arquivos .csv para facilitar a entrada destes em software de análises de texto.

Visualização Nesta parte do processo faz-se necessário conhecer o conjunto de dados para obter insights sobre análises mais elaboradas. As formas mais comum de visualização de conjunto de dados textuais se da por meio de: Nuvens de Palavras, gráfico de frequência, bigramas e analise de sentimento.

Análise A análise é o processo que procura responder ao objetivo de pesquisa e inferir sobre um resultado final baseado no conjunto de dados coletados. Contudo este processo se resume a utilizar ferramentas analíticas para extrair informações, correlações e encontrar agrupamento no conjunto dos dados. A análise dos dados é a etapa que depende da habilidade do analista, sendo importante sua expertise e visão de negócio (STAUDT, 2016).

2.3 CORRELAÇÃO SPEARMAN

Medidas de correlação entre duas variáveis quantitativas, são comumente utilizadas em diversas áreas de estudo. Uma forma prática de encontrar esta correlação entre as variáveis se da calculando o coeficiente de correlação de Pearson.

Segundo Garson (2009) "coeficiente de correlação se trata de uma medida de associação bivariada do grau de relacionamento entre duas variáveis". Para realizar inferências por meio de teste de hipóteses, o coeficiente de correlação de Pearson necessita de que a população amostrada tenha distribuição normal bivariada. Contudo, dados que seguem uma normal bivariada não são comuns de serem encontrados na prática. Uma solução se em obter a correlação por meio do coeficiente de Spearman (DANIEL, 1978).

O método de correlação de postos de Spearman, foi definido pelo psicólogo e estatístico Charles Spearman (1904), como sendo uma medida não paramétrica de correlação. A obtenção desta correlação costuma ser representada pela letra grega ρ (Rho) e sendo muito semelhante ao método de Pearson, com única diferença na utilização dos postos dos dados ao invés dos seus valores de forma direta (SPEARMAN, 1904). Rho de Spearman como BAUER (2007) nomeia, é uma ótima alternativa para calcular correlações na situação em que há violação de normalidade das variáveis.

Segundo Daniel (1978), o **coeficiente de correlação** ρ_{xy} entre X e Y pode ser calculado a partir das expressões matemáticas a seguir:

$$\rho_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 . \sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{1}$$

Onde x_i e y_i representam os postos de X e Y. A **covariância** entre X e Y pode ser escrita pela seguinte expressão:

$$S_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1} \tag{2}$$

Colocando (2) em (1) podemos escrever o coeficiente de correlação em função da covariância e desvios-padrão entre X e Y.

$$\rho_{xy} = \frac{S_{xy}.(n-1)}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}.\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} = \frac{S_{xy}.(n-1)}{S_x\sqrt{n-1}.S_y\sqrt{n-1}} \tag{3}$$

Obtendo portanto uma expressão para o calculo do coeficiente de correlação entre X e Y.

$$\rho_{xy} = \frac{S_{xy}}{S_x.S_y} \tag{4}$$

Daniel (1978), acrescenta que os postos de X e Y são obtidos pelo ordenamento numérico, onde o menor valor de x recebe posto 1, o segundo menor posto 2 e assim por diante. Desta forma cada valor de x_i corresponde a um posto $R(x_i)$, da mesma forma para y_i recebe seus postos $R(y_i)$. Quando os posto de X e Y não são iguais, pode se utilizar uma expressão resumida para obter a correlação de Spearman, obtidas da equação (1), por manipulações algébricas e propriedades de somatórios.

$$\rho_{xy} = 1 - \frac{6\sum_{i=1}^{n} d^2}{n(n^2-1)} \tag{5}$$

Em que d = $R(x_i)$ - $R(y_i)$, ou seja a diferença entre os postos das variáveis, mais informações sobre como obter a equação (5), pode ser obtidas em (DANIEL, 1978).

Os valores assumidos por ρ_{xy} varia de -1 a 1, onde o sinal indica a direção da correlação. Ou seja Moore (2007) destaca que um valor de correlação próximo ou igual a -1 significa uma relação oposta e negativa entre as variáveis X e Y, valor de correlação próximos ou igual a 1 mostra uma alta associação entre as variáveis e valor próximos a 0 demonstra que as variáveis não possuem relação linear entre si.

Para realizar inferência sobre o coeficiente de correlação de Spearman, com amostras grandes (n>10) pode-se realizar um teste com distribuição t de Student, adotando 0,05 como valor de significância. Daniel (1978), menciona que não há restrições para o teste de significância para Spearman, podendo ser utilizado a estatista t de Student para realizar inferência.

Os par de hipóteses são:
Hipótese nula $H_0 : \rho = 0$
Hipótese alternativa $H_1 : \rho \neq 0$

Com Valor-P menores que 0,05 rejeita-se a hipótese nula, de que não existe associação entre X e Y, e aceita a hipótese alternativa de a correlação de Spearman é significativa e diferente de zero.

2.4 TESTE MCNEMAR

O teste de McNemar surgiu na psicologia onde duas respostas com dois resultados correlacionados precisavam ser comparados (LACHENBRUCH, 2014). Logo, Quinn McNemar em 1947 da inicio ao desenvolvimento de um teste estatístico cujo objetivo é comparar duas amostras pareadas afim de se avaliar um ensaio antes e depois, cujo controle passa a ser o próprio indivíduo. Uma tabela 2x2 de contingência foi utilizada para observar as mudanças (MCNEMAR, 1947). Esta tabela de contingência pode ser exemplificada abaixo.

Tabela 1 – Exemplo de tabela de contingência testeMcNemar

		Depois	
		V_+	V_-
Antes	V_+	A	C
	V_-	B	D

Fonte: Elaboração Própria

Quando os dados de antes e depois apresentam diferenças entre as respostas, estas serão observadas nas células B e c da Tabela 1. A célula B representa os dados que antes eram de (V_-) e depois passaram a ser (V_+), assim como a célula C representa os dados que antes eram (V_+) e depois passaram a ser (V_-). Caso não haja nenhuma alteração entre o antes e depois os dados estarão posicionados nas células A e D (MCNEMAR, 1947).

A hipótese nula considera que não existe diferença entre o antes e depois e que as proporções marginais são as mesmas, logo temos que:

Hipótese Nula: $H_o : p_B = p_C$; Hipótese Alternativa: $H_1 : p_B \neq p_C$. A estatística do teste fica definida como sendo:

$$T_1 = \frac{(B-C)^2}{B+C} \tag{6}$$

A estatística de teste T_1 é aproximadamente um distribuição χ^2 com um grau de liberdade. Logo, se o valor de T_1 for significativo, rejeita-se a hipótese nula de que B = C e aceitasse a hipótese alternativa de que existe diferença entre o antes e depois.

Uma correção a estatística de Teste T_1 é necessária quando temos casos em que $B+C \leq 20$, em que aplica-se a correção de continuidade de Yates, devido as baixas frequências esperadas, obtendo assim a expressão T_2 (LACHENBRUCH, 2014).

$$T_2 = \frac{(|B-C|-1)^2}{B+C} \tag{7}$$

Para casos em que o valor calculado em T_2 for maior ou igual ao valor observado na tabela com grau de liberdade igual a 1, sugere-se que rejeite a hipótese H_o, não havendo assim diferença entre antes e depois.

2.5 TESTE DE INDEPENDÊNCIA DE QUI-QUADRADO

Quando temos n elementos de uma amostra que podem ser classificados de acordo com dois critérios distintos, surge o interesse de saber se as duas variáveis são estatisticamente independentes (BARTLETT, 1935). Por exemplo, podemos desejar verificar se os preços dos produtos alimentícios são independentes ou não da sua marca (HINES, 2000).

Para realizar esse teste, os dados apareceriam em geral em uma tabela denominada como tabela de contingência. Esta tabela busca representar observações para múltiplas variáveis categóricas (HINES, 2000). Esta possui r níveis referentes a primeira variável e c níveis referente a segunda variável, conforme mostra a Tabela 2.

Tabela 2 – Modelo tabela de contingência (r x c).

	Coluna			
Linha	1	2	...	c
1	O_1	O_{12}	...	O_{1c}
2	O_2	O_{22}	...	O_{2c}
...	...	...		...
r	O_{r1}	O_{r2}	...	O_{rc}

Fonte: Elaboração Própria

Segundo Bartlett (1935), o teste qui-quadrado deve ser utilizado para realizar o teste de independência, que tem como objetivo testar a significância da interação entre os dois grupos.

Na tabela de contingência procuramos testar a hipótese de que as variáveis presentes nas linhas e nas colunas são independentes ou não. Caso a hipótese for rejeitada, podemos concluir de que existe interação entre as duas variáveis (BARTLETT, 1935). Hines (2000) acrescenta que realizar este teste não é algo intuitivo, contudo para n grande, pode-se utilizar uma estatística de teste aproximada de qui-quadrado.

- Hipótese Nula: H_0= Não existe associação entre as variáveis.
- Hipótese Alternativa: H_1=As variáveis estão associadas.

Se definirmos $p_{ij} = u_i v_j$ como sendo a probabilidade de que um elemento selecionado ao acaso esteja na ij-ésima célula. Onde u_i é a probabilidade do elemento cair na classe da linha i e v_j é a probabilidade do elemento cair na classe da coluna j, logo podemos admitir a independência dos estimadores de máxima verossimilhança de u_i e v_j (HINES, 2000).

$$u_i = \frac{1}{n} \sum_{j=1}^{c} O_{ij} \tag{8}$$

$$v_j = \frac{1}{n} \sum_{i=1}^{r} O_{ij} \tag{9}$$

Logo, o número esperado em cada célula é

$$E_{ij} = nu_i v_j = \frac{1}{n} \sum_{j=1}^{c} O_{ij} \sum_{i=1}^{r} O_{ij} \tag{10}$$

Para n grande, a estatística fica definida em

$$\chi_0^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \approx \chi^2_{(r-1)(c-1)} \tag{11}$$

A hipótese de independência das variáveis sera rejeitada se $\chi_0^2 > \chi^2_{\alpha,(r-1)(c-1)}$ (HINES, 2000).

Esta prática de utilizar a tabela de contingência para verificar a independência entre duas variáveis em uma amostra é apenas uma das aplicações possíveis de serem utilizadas. Outra situação possível se da quando temos mais de uma população de interesse e cada população são divididas em c categorias. A hipótese nula nesse caso afirma que as populações são homogêneas em relação às categorias (HINES, 2000).

A estatística definida na Expressão 10, recebe o nome de teste de Qui-Quadrado de Pearson, ou simplesmente teste de qui-quadrado, muito utilizado para variáveis nominais. Esta estatística é não paramétrica e requer que os grupos de teste tenham aproximadamente tamanhos iguais e que os valores presentes na Tabela 2 devem ser frequências ou contagem, em vez de porcentagens. As categorias das variáveis são mutualmente exclusivos (MCHUGH, 2013).

2.6 ANÁLISE DE CORRESPONDÊNCIA

A utilização da tabela de contingência para obter resultados numéricos de correlação entre as linhas e colunas, já é muito utilizado no meio da estatística por Hartley em 1935, Fisher em 1940 entre outros pesquisadores. Contudo a utilização de expressar os resultados de correlação entre variáveis em um modo gráfico, só foi iniciado por Benzécri em 1960, onde utilizou a análise de correspondência para verificar a correlação entre variáveis categóricas (ABDI, VALENTIM, 2007).

A análise de correspondência (AC) permite expressar os resultados das interações entre variáveis categóricas de uma tabela de contingência em um layout gráfico de fácil e rápida interpretação. Os gráficos levam em consideração as componentes principais das linhas e colunas, onde os pontos próximos representam uma associação e os mais afastados uma repulsão, sem a necessidade de obter um uma distribuição de probabilidade para expressar os dados (GREENACRE, 2006). .

A teoria da AC é baseada na teoria de matrizes, sendo o resultado central e a decomposição de valor singular a base para o seu entendimento. A notação matriz-vetor foi escolhida por ser mais compacta e próxima das funções implementadas na linguagem R.

Assumindo que temos uma matriz $\mathbf{X}$ = IxJ que possui posto coluna completo, divi-

dindo todos os elementos de $\mathbf{X}$ por n, obtemos então uma matriz de proporção $\mathbf{P}= p_{ij}$, ficando então com:

$$\mathbf{P} = \frac{1}{n}\mathbf{X} \tag{12}$$

A matriz $\mathbf{P}$ recebe o nome de matriz de correspondência. Esta análise pode ser formulada como um problema de mínimos quadrados ponderados, para isso faz-se necessário definir os vetores soma das linhas e colunas $\mathbf{r}$ e $\mathbf{c}$ respectivamente. Assim como as matrizes diagonais de $\mathbf{r}$ e $\mathbf{c}$.

$$r_i = \sum_{j=1}^{J} p_{ij} \qquad\qquad c_j = \sum_{i=1}^{I} p_{ij} \tag{13}$$

$$\mathbf{D}_r = diag(u_1, u_2, ..., u_I) \qquad\qquad \boldsymbol{D}_c = diag(v_1, v_2, ..., v_J) \tag{14}$$

Segundo Greenacre (2017), o algoritmo computacional para obter as coordenadas das linhas e colunas com relação aos eixos principais utilizando o valor de decomposição singular (VDS) são mostrados em seis passos.

-Passo 1: Calculo da matriz S de resíduos.

$$\mathbf{S} = \mathbf{D}_r^{-1/2}(\mathbf{P} - rc^T)\mathbf{D}_c^{-1/2} \tag{15}$$

-Passo 2: Calculo do valor de decomposição singular de $\mathbf{S}$.

$$\mathbf{S} = \mathbf{U}\mathbf{D}_\alpha\mathbf{V}^T \qquad onde \qquad \mathbf{U}^T\mathbf{U} = \mathbf{V}^T\mathbf{V} = \mathbf{I} \tag{16}$$

Onde $\mathbf{D}_\alpha$ é a matriz diagonal de valores positivos em ordem decrescente $\alpha_1 \geq \alpha_2 \geq ...$

-Passo 3: Obtendo as coordenadas $\mathbf{\Phi}$ das linhas.

$$\mathbf{\Phi} = \mathbf{D}_r^{-1/2}\mathbf{U} \tag{17}$$

-Passo 4: Obtendo as coordenadas $\mathbf{\Gamma}$ das colunas.

$$\mathbf{\Gamma} = \mathbf{D}_c^{-1/2}\mathbf{V} \tag{18}$$

-Passo 5: Coordenadas principais $\mathbf{F}$ das linhas.

$$\mathbf{F} = \mathbf{D}_r^{-1/2}\mathbf{U}\mathbf{D}_\alpha = \mathbf{\Phi}\mathbf{D}_\alpha \tag{19}$$

-Passo 6: Coordenadas principais $\mathbf{G}$ das colunas.

$$\mathbf{G} = \mathbf{D}_c^{-1/2}\mathbf{V}\mathbf{D}_\alpha = \mathbf{\Gamma}\mathbf{D}_\alpha \tag{20}$$

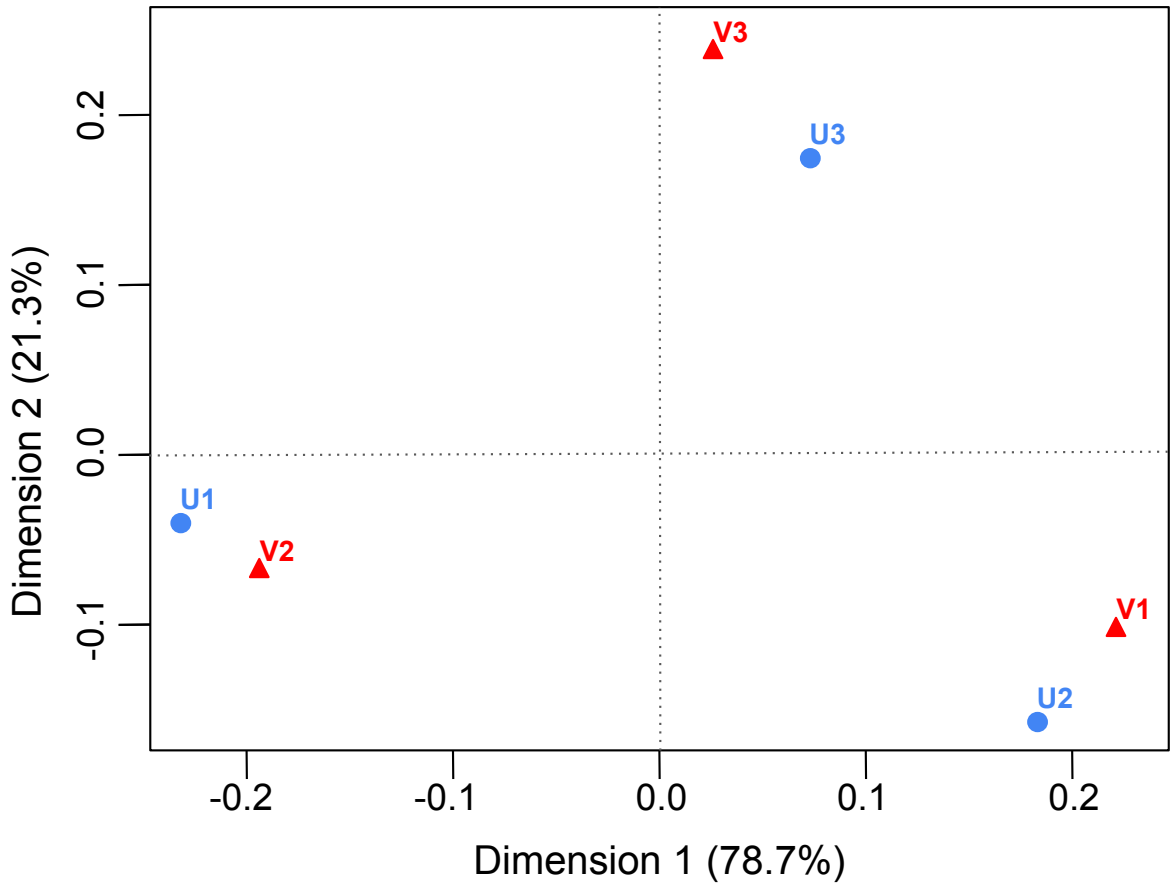
-Passo 7: Principio Inércias λ_k.

$$\lambda_k = \alpha_k^2, \qquad k = 1, 2, ..., K \tag{21}$$

A (AC) costuma expressar em seu layout gráfico as duas ou três primeiras colunas das expressões (18) e (19). Gerar este gráfico com as coordenadas conjuntas de **F** e **G** recebe o nome de mapa simétrico, se o as unidades e dimensões das linhas e colunas forem as mesmas.

Maiores detalhes na explanação teórica sobre (AC), podem ser encontradas em Greenacre (2017) e Infantosi, Costa e Almeida (2014).

Figura 4 – Ilustração análise de correspondência.



Fonte: Elaboração Própria

Greenacre (2017) acrescenta que a (AC) é uma técnica cujo objetivo é representar linhas e colunas de uma matriz de contingência em dois espaços vetoriais de baixa dimensão, como mostra o exemplo na Figura 4.

Na Figura 4, podemos observar um exemplo de uma (AC), com duas variáveis categóricas em azul e em vermelho. A interpretação que pode ser realizada na Figura 4 é a influência dos pontos no espaço, e o maior peso que o eixo X possui na análise, por

apresentar uma distância de (78,7%) com relação ao eixo Y de (21,3%). Pode-se também inferir sobre a correlação entre as variáveis, pontos que estão próximos um do outro, mais correlacionados eles são, e quanto mais afastados menor é a correlação. Isto é possível de ser observado na Figura 4, onde as variáveis U1 e V2 apresentam alta correlação, enquanto V3 e U2 apresentam baixo nível de correlação.

3 METODOLOGIA

3.1 COLETA

3.1.1 Pacote `rtweet`

O pacote `rtweet` é um pacote voltado para coleta de dados do twitter, tendo sido lançado em 19 de maio de 2019, na versão 0.6.9 e atualizado para 0.7.0 pelos autores: Michael W. Kearney, Andrew Heiss e Francois Briatte. Para realizar a coleta de dados no twitter é necessário obter as chaves da API *(Application Programming Interface)* , que serão acrescentadas no pacote `rtweet` para que o mesmo possa ter acesso a plataforma twitter. Na versão gratuita, o twitter limita a busca de tuítes em um número de 18.000 a cada 15 minutos que são coletados de forma cronológica, ou seja, das publicações mais recentes às mais antigas (KEARNEY, 2019).

Para obter acesso a API do twitter, é necessário criar uma conta desenvolvedor no site *https://developer.twitter.com/*, e preencher alguns formulários declarando qual o objetivo da coleta dos dados na plataforma. Logo em seguida será liberado a criação das chaves e tokens necessários para realizar a coleta. Após obter as chaves de acesso da conta do twitter estas devem ser implementadas no pacote `rtweet`, que utiliza estas chaves para se comunicar e coletar os dados textuais do twitter (KEARNEY, 2019). Segue abaixo um exemplo da implementação da API no pacote `rtweet`, por meio do *software RStudio*.

```
library (rtweet)

#Login e credenciamento na base do twiteer
api_key <- "hXdtGu8sobtWU1......"
api_secret_key <-"ezIM1voZVJfRzYBYGvCQktcS4olEotGH2JYmjAb....O"
acess_token <- "1257771432186183682-oUwb6fahbcv...."
acess_secret_secret <- "u5gnjG9j9QiAPgsS3rhlUHx5Gol5r7fa1Q...."
token <- create_token(
 app = "Baldasso",
 consumer_key = "hXdtGu8sob....",
 consumer_secret = "ezIM1voZVJfRzYBYGvCQktcS4olEotGH2....",
 access_token = "1257771432186183682-oUwb6fahbcvk...",
 access_secret = "u5gnjG9j9QiAPgsS3rhlUHx5Gol5...",set_renv = TRUE)
```

Com as chaves de acesso carregadas é possível realizar a mineração dos tuítes por diferentes modos existentes no pacote `rtweet`. Este trabalho, irá se concentrar em realizar a mineração de texto por palavras-chave e coleta por usuários. Para isso, será utilizado a função *search_tweets* para coletas com palavras-chave e a função *get_timeline* para coleta de tuítes presentes na linha do tempo do usuário. Seguem o código R necessárias para realizar estes dois tipos de coleta.

```
#Formas de busca
#Busca por palavras especificas
Dados1 <- search_tweets("coronavirus", n = 9000, include_rts =
FALSE, lang = "pt")
Dados2 <- search_tweets("Cafe covid-19", n = 9000, include_rts =
FALSE, lang = "pt")
Dados3 <- search_tweets("pandemia Cerveja", n = 9000, include_rts =
FALSE, lang = "pt")
Dados4 <- search_tweets("vacina", n = 9000, include_rts =
FALSE, lang = "pt")
Dados5 <- search_tweets("segunda onda", n = 9000, include_rts =
FALSE, lang = "pt")
```

```
#Busca por Usuarios
# Obter os 3200 tweets mais recentes postados por Donald Trump
Trump <- get_timeline("realDonaldTrump", n = 3200)

# Metodo de coleta em varios usuarios ao mesmo tempo
data <- get_timeline(c("CNN", "FoxNews", "CnnBrasil"), n = 3200)

# Sera retornado um data frame
data
```

A busca utilizando a função *search_tweets* pode ser realizada com uma palavra chave, ou mais de uma. Será retornado um total de n tuítes programado dentro da função, que possuem as palavras-chave selecionadas, sem a necessidade delas estarem em sequência no tuíte. As coletas com o comando *search_tweets*, retorna tuítes de modo aleatório respeitando a ordem das publicações mais recentes até uma semana da data de publicação. Já a função *get_timeline* realiza a coleta dos 3200 primeiros tuítes, publicados pelo usuário selecionado, não limitado ao período da publicação, mas sim a quantidade de tuítes presente na linha do tempo do usuário.

Logo a principal diferença entre as funções *search_tweets* e *get_timeline*, é que *search_tweets* permite que a coleta seja feita pelo conteúdo textual e é limitada em quan-

tidade e tempo de busca; já a função *get_timeline* é uma coleta por usuários sem limite de tempo porém não é possível escolher previamente o conteúdo textual.

Figura 5 – Forma de armazenamento dos tweets.

	user_id	status_id	created_at	screen_name	text	source
4966	892636937420517377	1303732878812090373	2020-09-09 16:31:55	SaraMacedo24	- Geral ta saindo pra eventos, praias. Agora quando ...	Twitter for iPhone
5509	853243741183901697	1303719967247028224	2020-09-09 15:40:36	MarcosvZanetti2	- vish mãe, as aulas vão voltar, a senhora viu? -s m...	Twitter for Android
6166	12537223214556569...	1303704731991388160	2020-09-09 14:40:04	eutete__	" #VoltaÀsAulas " Eu de boa porq estudo em escola ...	Twitter for Android
4677	12537397239249428...	1303740118398832646	2020-09-09 17:00:41	GEOVANN4348...	" Ain, as praias estão liberadas, então podemos volt...	Twitter for Android
4739	12848712742930104...	1303738801098027010	2020-09-09 16:55:27	sverccosa	" vamos entrar na fase que sera possível a #voltaas...	Twitter Web App
5013	12874355850917683...	1303728760764669952	2020-09-09 16:15:33	ercilaine2	"79% dos brasileiros dizem que reabertura das esco...	Twitter for Android
4404	4205826616	1303749090078085125	2020-09-09 17:36:20	GonsalvesIsa	"A MaS iR a PRaiA PoDE, VOltA aS AulAs Naõ" QUEM...	Twitter for Android
3323	12335309809982873...	1303809091140157441	2020-09-09 21:34:45	cudemel_	"a mas meu filho não vai voltar pra escola porque é...	Twitter for Android
5330	11228710342968442...	1303724839505625089	2020-09-09 15:59:58	Blue_Vante	"Ah mas eu to indo pra praia ta tudo tranquilo pode ...	Twitter for Android
4330	815036264646897665	1303751320810205184	2020-09-09 17:45:12	SouMaduu	"Ah mas se pode ir pra praia, pode ir pra escola tbm...	Twitter for Android
6384	12913994619530444...	1303696282192678912	2020-09-09 14:06:29	Omeletwithjam	"Ah, mas se pode ir para a praia, pode voltar as aul...	Twitter for Android
5376	951247664317915136	1303723795388207106	2020-09-09 15:55:49	lullabysolar	"ain mas tem que voltar a ser presencial" PRESENC...	Twitter Web App
3847	12335723609697361...	1303773715646341121	2020-09-09 19:14:11	monsteroftemoon	"Amas se pode ir a praia pode voltar as aulas" KKKK...	Twitter for Android
5223	10465338825841991...	1303726933994344457	2020-09-09 16:08:17	Jeonarrtt	"amas ta todo indo pra praia" meu fi eu tô trancada...	Twitter for Android

Showing 1 to 17 of 6,629 entries, 90 total columns

Fonte: Elaboração Própria

Os tuítes serão retornados em um arquivo *data frame* que pode ser salvo em ".xlsx", ou ser mantido em um arquivo .R. Este arquivo contém 90 colunas de variáveis diferentes relacionadas aos dados da publicação no twitter, as 6 primeiras colunas contem as informações mais úteis, como a data de publicação e o conteudo textual do tuíte, como pode ser observado na Figura 5. As principais colunas mais utilizadas são: *(user_id)* que é um código identificador do usuário que fez a publicação; *(created_at)* mostra a data e hora da publicação; *(screen_name)* retorna o nome do perfil em que o tuíte foi coletado; *(text)* mostra o conteúdo textual presente no tuíte; e *(source)* mostra o dispositivo utilizado para publicação do tuíte.

3.2 LIMPEZA

A limpeza dos tuítes coletados é um processo necessário visto que os tuítes são acompanhados de *links*, *emoction*, *hastags*, entre outros termos que atrapalham a obtenção dos resultados. Para realizar este método de limpeza, é necessário colocar o conjunto de dados em um corpus. A função *Vcorpus* será utilizada para isto. Em seguida será removido: caracteres especiais, *stopwords*, pontuações, espaços em brancos e palavras que não forem relevantes para a análise, por meio do pacote `tm` (FEINERER; HORNIK, 2020). Seguem os códigos R que serão utilizados para esta tarefa.

```
library(tm)
docs <- Corpus(VectorSource(dados2$text))

#Corrigir acentuacao no texto
```

```
texto=sapply(docs, function(row)
iconv(row, "UTF-8", "ASCII//TRANSLIT", sub = ""))

#Remover mencao
removeMetions=function(x) gsub("@\\S+", "", x)
txt=tm_map(txt, content_transformer(removeMetions))
txt[[1]]$content

#Transforma todas as palavras em minusculas
txt=tm_map(txt, content_transformer(tolower))
txt[[1]]$content

##Remove pontuacao ( / , : ; . # )
txt=tm_map(txt, removePunctuation)
txt[[1]]$content

###Remove Palvras de parada - preposicao, artigos, etc
txt=tm_map(txt, removeWords, iconv(\textit{stopwords}("portuguese")))
txt[[1]]$content
#\textit{stopwords}("portuguese") #palavras que foram removidas.

##Remove espacos em branco dobrados
txt=tm_map(txt, stripWhitespace)
```

3.2.1 StopWords

O processo de mineração de texto, consiste na extração de conteúdos textuais de diferentes fontes, o que retorna uma grande quantidade de texto, que muita das vezes se encontram de forma desordenada, devido a grande repetição de termos comuns, principalmente quando esta extração de dados é proveniente de uma rede social. Uma forma de facilitar a extração de resultados deste banco de dados se da pela remoção das *stopwords*, ou palavras de parada (SARICA; LUO, 2021).

As *stopwords*, consiste em uma lista de palavras que não possuem informações úteis inerentes, como por exemplo os pronomes e preposições. O principal motivo da extração das *stopwords*, se da devido a sua presença avassaladora dentro dos textos minerados, o que dificulta a análise e extração de resultados, principalmente quando se pretende realizar, nuvem de palavras, analise de sentimentos e frequência de palavras. Deste modo, é comum em análise de texto existir uma lista de *stopwords*, para ser removidas do conjunto de dados (CHOY, 2012).

Contudo, é necessário que tome o cuidado de verificar se a lista de *stopwords*, é útil para o conjunto de dados, devido a diferença de linguagem dependendo da fonte de coleta. Uma lista de *stopwords*, utilizadas em uma obra literária muita das vezes não será útil para coleta de dados no twitter. Logo, se tornou comum montar manualmente uma lista de palavras irrelevantes, utilizadas por outros autores e adaptadas para o seu conjunto de dados (SARICA; LUO, 2021).

Vários autores provaram que o uso das *stopwords*, são aplicáveis a uma variedade de situações, facilitando assim a identificação de termos mais frequentes presentes nos textos e reduzindo assim o conjunto de dados. Se tornando uma ferramenta essencial para a estuda da análise de texto (CHOY, 2012).

3.3 ARMAZENAMENTO

Para armazenar o conjunto de dados já limpo, será utilizado a função *union_all* para colocar todas as observações em uma tabela de dados. A função *distinct*, também é utilizada com o intuito de remover as linhas com conteúdo duplicado. Existe a necessidade de transformar o conjunto de dados em um corpus de palavras, para realizar a limpeza. Alguns gráficos utilizados necessitam que os dados estejam no formato *data-frame*. A melhor forma de converter o conjunto de dados foi deixando eles no formato de tabela.

```
BANCO <- union_all(BANCO11,BANCO12)
##Obter os nao duplicados
dados2 <- distinct(BANCO)
docs <- Corpus(VectorSource(dados2$text))
result1<-data.frame(colSums(result[,]))
```

A segunda forma de armazenamento utilizada neste trabalho foi manter os arquivos de coleta originais no formato .R, visto que objetivo não era extrair resultados do conteúdo textual do tuíte, mais sim de sua frequência de publicação ao longo do tempo.

3.4 N-GRAMAS

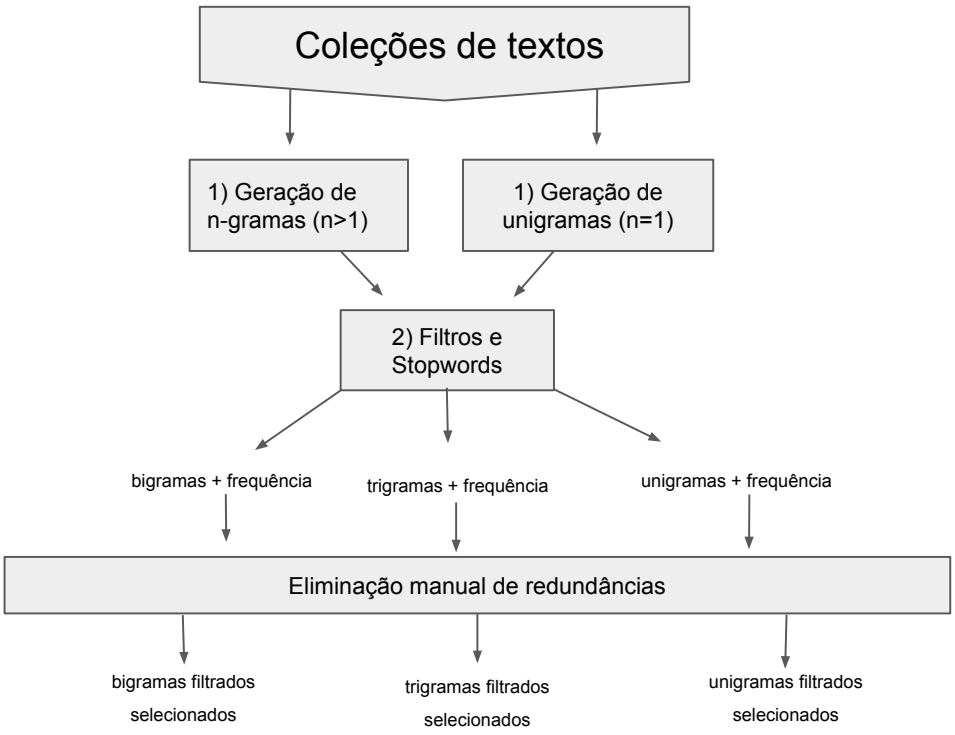
Os n-gramas podem ser definidos como uma sequência de elementos textuais, que podem ser palavras, caracteres, ou quaisquer outros elementos que estiverem em sequência em um texto. Ou seja, um n-grama é um substring de comprimento n caracteres derivado de um texto, onde os caracteres do n-grama segue a mesma ordem do texto de origem que foi extraído (ROBERTSON, 1998).

O "n"em n-gramas corresponde ao número de elementos da sequência, podendo existir os unigramas (n=1), contendo apenas uma palavra ou um caractere, bigramas (n=2) duas palavras em sequência, trigramas (n=3) e assim por diante. Elementos mais longos também podem ser extraídos como tetragramas ou pentagramas, contudo estes são menos frequentes de aparecer em um conjunto de texto (SIDOROV, 2014).

O maior objetivo da extração de n-gramas em um conjunto de texto, é obter de forma automática um conjunto de atributos que seja enxuto, descritivo e discriminativo do meu conjunto de dados (SUEN, 1979).

Na figura 6, esta representado uma ilustração do processo de extração dos n-gramas.

Figura 6 – Ilustração do modelo de seleção de n-gramas não redundantes.



Fonte: Elaboração Própria

Geração de Unigramas o primeiro passo consiste na geração de todos os possíveis unigramas presentes no texto. Para isso pode ser utilizado qualquer ferramenta de análise de texto que realiza o processo de *tokenização* que é a separação de palavras de um texto. No software R, o pacote `dplyr` realiza esse procedimento por meio da função `unnest_tokens`.

Filtros e Stopwords o segundo passo esta na execução de filtros e stopwords, que pode ser definidas por Suen (1979) como uma coleção de artigos, preposições, conjunções e interjeições, e todas palavras que não agregue informação na formação dos atributos. Listas de *stopwords* são comumente encontradas em software de análises de texto, contudo existem casos que surge a necessidade de se adaptar e acrescentar palavras, criando assim a sua própria lista de *stopwords*. No software R, as listas de *stopwords* e filtros podem ser processados por meio da função `filter` do próprio sistema do R .

Geração de n-gramas a extração n-gramas com n maior que um, consiste em extrair do arquivo de texto os bigramas, trigramas, tetragramas etc. Logo após, os mesmos

serão separados em unigramas, para a passagem dos filtros e remoção das *stopwords*, e depois são unidos novamente na mesma ordem original os n-gramas que não foram separados devido a remoção das *stopwords*.

Segue os códigos utilizados no Rstudio, para extração dos n-gramas:

```
library(dplyr)
library(tidytext)
library(tm)

#unigrama
n1 <- pandemia %>%
  unnest_tokens(bigram, text, token = "ngrams", n = 1) %>%
  separate(bigram, c("word1"), sep = " ") %>%
  filter(!word1 %in% stop$X.sapply.A1..c..) %>%
  unite(bigram, word1, sep = " ")

A <- n1 %>%
  count(bigram, sort = TRUE)

#bigrama
n2 <- dados1 %>%
 unnest_tokens(bigram, text, token = "ngrams", n = 2) %>%
  separate(bigram, c("word1", "word2"), sep = " ") %>%
  filter(!word1 %in% stop$X.sapply.A1..c..) %>%
  filter(!word2 %in% stop$X.sapply.A1..c..) %>%
  unite(bigram, word1, word2, sep = " ") %>%
  count(bigram, sort = TRUE)

# trigrama
n3 <- dados1 %>%
  unnest_tokens(bigram, text, token = "ngrams", n = 3) %>%
  separate(bigram, c("word1", "word2", "word3"), sep = " ") %>%
  filter(!word1 %in% stop$X.sapply.A1..c..) %>%
  filter(!word2 %in% stop$X.sapply.A1..c..) %>%
  filter(!word3 %in% stop$X.sapply.A1..c..) %>%
  unite(bigram, word1, word2, word3, sep = " ") %>%
  count(bigram, sort = TRUE)
```

3.5 NUVEM DE PALAVRAS

Nuvens de palavras são técnicas de visualização de dados textuais simples e intuitiva, cujo objetivo é fornecer uma primeira impressão sobre os dados textuais que estão em estudo, obtendo assim uma percepção visual dos resultados (CUI, 2010). Costumam mostrar as palavras mais frequentes de um texto em destaque das de menor frequência, distribuídas em um layout (circular, retangular, quadrado etc) (HEIMERL, 2014).

Os tamanhos das fontes indicam as palavras com maior ocorrência, assim como as cores distribuídas em grupos conforme a frequência. Outros requisitos como posição, orientação e tipo de fonte, são meramente estéticos (LOHMANN et al, 2015).

Nuvens de palavras são indicadas para uma análise inicial do conjunto de texto, esta não realiza nenhum tipo de análise e fornece um resultado puramente estatístico, sem tomar nenhum conhecimento linguístico só representa o que foi informado a ela (HEIMERL, 2014).

No software R, o pacote `wordcloud` possui várias funções para execução de nuvens de palavras (FELLOWS, 2018).

Segue os códigos utilizados no Rstudio, para extração dos n-gramas:

```
#Nuvem de palavras
library(wordcloud)
wordcloud(txt,min.freq=5, scale=c(3.5, .5),
random.order=FALSE, rot.per=0.35,
          colors=brewer.pal(8, "Dark2"))
```

4 THE IMPACT OF COVID-19 ON BRAZILIAN FOOD PRIORITIES: A STUDY USING ONLINE SURVEY AND TWITTER

Abstract: The crisis brought about by the new coronavirus affected all production chains. The (FAO) announced in March 2020 that the necessary isolation measures could trigger the onset of a global food crisis. With the closing of borders between countries, shortages of food and products were observed, causing panic in consumers and a rush to supermarkets. With that in mind, this article aims to verify changes in eating habits one year after the start of the COVID-19 pandemic. For this, surveys on the social network Twitter and an online questionnaire with 500 Brazilians were applied. The Crisis + Food surveys on Twitter showed that consumers are anxious and are using coffee in this period. The isolation also made consumers in their vast majority start to adopt orders for delivery, reducing the frequency of purchase from twice a week to weekly. The foods most purchased by the participants were fruits, and rice and the least bought were Meat and Soft Drinks. The search for a healthy diet during the isolation period was present in the justifications presented. However, sweets, chocolates, and sandwiches were associated with helping with stress, appearing in tweets searched by quarantine + food. It is concluded that the impacts generated by COVID-19 continue to affect consumers one year after its beginning, with the financial implications being the leading cause of the Brazilians' dietary change.

Key-words: COVID-19, Food, Twitter, Behavior, Consumer.

Resumo: A crise instaurada pelo novo coronavírus afetou todas as cadeias produtivas. A (FAO) anunciou em março de 2020, que as medidas necessárias de isolamento poderiam provocar o início de uma crise global alimentícia. Com o fechamento das fronteiras entre os países uma escasses de alimentos e produtos foram observadas, gerando panico nos consumidores e uma corrida aos supermercados. Pensando nisso, este artigo tem como objetivo, verificar as mudanças nos hábitos alimentares durante um ano após o inicio da pandemia de COVID-19. Para isso, pesquisas na rede social Twitter e um questionário online com 500 brasileiros, foram aplicados. Os resultados das pesquisas por Crise + alimentos no Twitter mostraram que os consumidores estão ansiosos e fazem uso do café neste período. O isolamento fez também com que os consumidores em sua grande maioria passassem a adotar pedidos por delivery, reduzindo a frequência de compra de duas vezes por semana para semanal. Os alimentos mais comprados pelos participantes foram frutas e arroz e os menos comprados foram Carne e Refrigerante. A procura por uma dieta saudável no período de isolamento se mostrou presente nas justificativas apresentadas. Contudo, doces, chocolates e sanduíches foram associados como ajudar com o estresse, aparecendo também em tweets pesquisados por quarentena + alimentos. Conclui-se que os impactos gerados pela COVID-19 continuam afetando os consumidores um ano após seu início, sendo o impacto financeiro o principal causador na mudança alimentar dos brasileiros.

Palavras-chave: COVID-19, Alimentos, Twitter, comportamento, consumidor.

4.1 Introduction

In December 2019, the first cases of respiratory infection caused by the new coronavirus (SARS-CoV-2) were in the populous Chinese city of Wuhan (WHO et al., 2020). Asia and Europe were the first continents to be affected. With less than three months of global contagion, the World Health Organization (WHO) declared a state of a pandemic on March 11, 2020, suggesting that severe measures of isolation, quarantine, and distancing social measures are taken to contain the spread of the virus (AQUINO et al., 2020).

According to the Food and Agriculture Organization (FAO), the necessary isolation measures led to the beginning of a global food crisis (FAO, 2020). The high prices of doing business, the pressed supply chain, the closing of borders between countries, and the decreased credit of economies led most countries to economic recession. This ultimately affected the food production chains, raising fears of a possible food deficiency globally (FAO, 2020).

Agricultural sectors such as farming, livestock, and fishing were hit hard by COVID-19 (FAO, 2020). In China, the most significant impact was on livestock due to limited access to the supply of raw materials (HOBBS, 2020). Due to sanitary measures, the chicken and fish distribution sector has been reduced in different parts of America, Asia, and Europe (CULLEN, 2020). Agricultural producers worldwide face deficits of inputs such as grains, fertilizers, and pesticides and have seen sales decrease due to difficulties with transportation, being forced to stock what they had, thus raising the cost of food (POUDEL, 2020).

In Brazil, supermarkets and food producers entered the essential services regime, continued to work during isolation measures, and operated with restrictions to avoid crowding (PAULO; ARAUJO, 2020). Consumer panic with lockdown notices led to a rush to stock up on food, thus affecting the availability and prices of these products in supermarkets (POUDEL, 2020). These events changed consumer demand and desire to buy food, causing a necessary change to adapt to this quarantine moment (PAULO; ARAUJO, 2020).

Martinotto et al. (2020), in their research with 785 Brazilians, at the beginning of the restrictive measures in the first half of 2020, showed an increase in weight in 50% of cases. Malta et al. (2020) included that the consumption of ultra-processed foods at the beginning of the pandemic in Brazil increased by 15%. According to Verticchio (2020), in his research in Belo Horizonte, Minas Gerais, with 1000 volunteers, 54% reported having gained weight during isolation and increased consumption of sweets and soft drinks. However, Maynard et al. (2020) showed that the changes in Brazilian food consumption were due to the increase in fruits, vegetables, and beans, decreasing the consumption of soft drinks and fried foods.

Likewise, Durães et al. (2020) report research carried out in China and Italy at the beginning of the isolation measures, which showed that consumers had a higher consump-

tion of vegetables and fruits and decreased alcohol consumption. In Spain, Laguna et al. (2020) show an increase in the use of the internet in the search for healthy diets, their research with Spanish showed an increase in the consumption of pasta and vegetables (health reasons), nuts, cheese, and chocolate (improving mood), and reduction in the purchase of short-life products. Eftimov et al. (2020), in their study on the recipe website Allrecipes, found that searches for pasta and bakery products increased during the quarantine, searches for juices, seafood, chocolate, and wine decreased. Snuggs and McGregor (2020) showed that family food choices in the UK in the first wave of COVID-19 cases were motivated by improving health, maintaining weight and condition, and decreasing pressure and stress. The same authors reported that vaccine communications starting testing phases produced hope to the population of the United Kingdom, who began to move more frequently to supermarkets in search of promotions and discounts (SNUGGS; MCGREGOR, 2020).

On December 8, 2020, the United Kingdom was the first country to vaccinate the population, followed by China, the United States, and 47 other countries (WHO et al., 2020). Brazil started vaccination against COVID-19 in January 2021, following a priority vaccination plan. However, the challenges set by COVID-19 are far from over in Brazil, with inflation of 14.1% in the food sector, vaccination proceeding at a slow velocity, mutation of SARS-Cov-2 creating more transmissible variants, and the economic slowdown show us that the end of the pandemic and its consequences are far from over (DOMINGUES, 2021).

Thus, this article proposes to verify the impacts of the pandemic on Brazilians' food properties, presenting consumers' changes in the first half of 2021. Brazilian publications of Twitter were collected, with related words the pandemic and food during January to May of 2021. Additionally, an online survey was made to verify changes in food shopping places, which foods were more and less obtained, and the motivation for this by Brazilian consumers. This part of the survey was carried out from March to May 2021, considered the worst period of the pandemic in Brazil. In March, more than 2 million cases of covid were registered in Brazil, with a record 97,586 cases in just one day. In April, more Brazilians died of covid, reporting 82,392 deaths and a record 4,249 deaths in just one day, which can be seen in Figure 7 e 8.

4.2 Materials and Methods

4.2.1 Twitter data mining on food

The social network Twitter concedes your messages posted on its platform (tweets) to be mined via an Application programming interface (API), which can be obtained by requesting a developer account for Twitter (TWITTER, 2021). In this paper, tweets were mined through the rtweet package (Kearney, 2018), made available in the R software (R

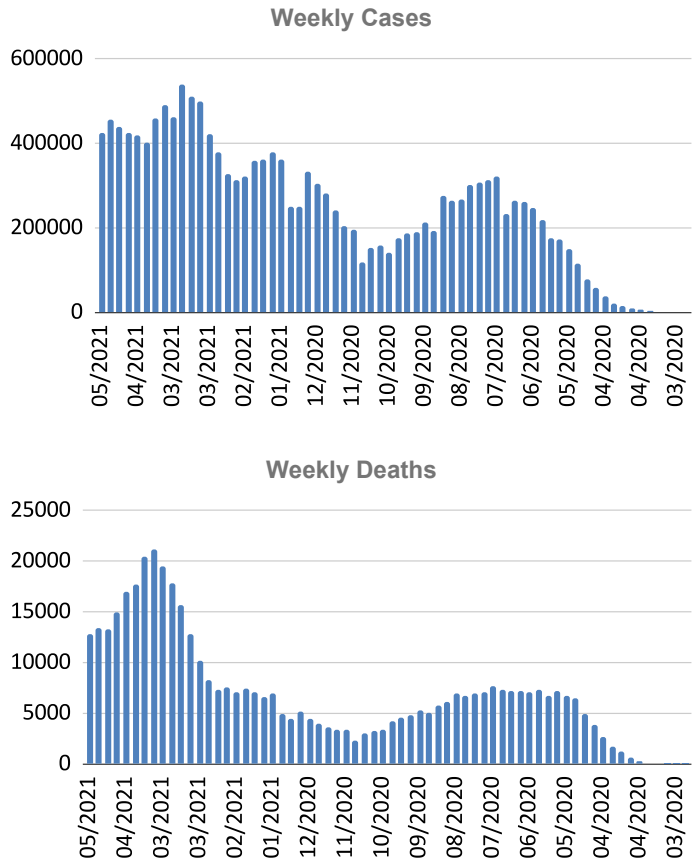


Figure 7 – Pandemic data in Brazil, provided by the Ministry of Health, cases and deaths.

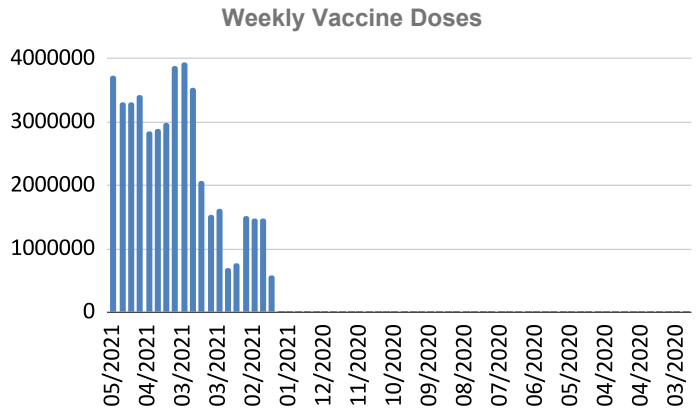


Figure 8 – Vaccination data against Covid-19 in Brazil, provided by the Ministry of Health.

Core Team, 2021). Tweets were extracted containing the terms in Portuguese: (pandemic, crisis, covid, quarantine) accompanying with the food: (milk, coffee, beer, meat, bread, wine, chocolate, chicken, fruit, soda, supermarket, and restaurants), in the period from January 10th to May 31st, 2021. In other words, all tweets that have the two search terms, for example: (pandemic and milk) even if they are not in sequence, will be extracted.

4.2.2 Online survey on consumer behavior towards food

An online questionnaire was sent to 800 Brazilian participants from 25 states in Brazil. In May 2021, amidst the peak of deaths and covid cases, we completed the survey by 500 consumers (55% women and 45% men with an average age of 33 years). Participants were asked about changing food buying locations (supermarkets, small stores, and online) and ready-to-eat foods (restaurants, bars, and homes) during this pandemic period, along with shopping frequency (daily, weekly, biweekly, and monthly). Participants were also asked which foods they stopped buying and which ones started to buy more in this period of isolation. When participants indicated to buy more or less one product, they were asked to justify this change. The justifications were based on the work of Laguna et al. (2020) and Puddephatt et al. (2020) and are based on feeling, price, desires, health, and food validity. Participating consumers were also invited to express their opinion about the moment lived in an open question, where 55% of respondents participated.

4.2.3 Data analysis

The tweets collected were grouped and extracted a timeline in two ways: food, with the curves represent the terms of the pandemic, and second, the terms of the pandemic, where the curves represent the food. Also performed an n-grans analysis, for this, the tweets went through a process of removing stopwords through the tm package (Feinerer, 2020) and were submitted to the n-grans analysis through the dplyr package (Wickham et al., 2021), and were obtaining the word clouds of unigrams, bigrams and trigrams through the wordcloud package (Fellowns et al., 2018).

Chi-square test was used to verify the dependence between places and frequency of shopping with the pandemic. The most and least purchased foods also passed the chi-square test to confirm the association between foods and justifications. The analyzes used a 0.05 level of significance. Hypotheses rejected by the chi-square test (positive association) were analyzed using the mc-nemer test to find possible migrations before and after the pandemic. Two correspondence analyses were performed for the data on the most and least purchased foods and their justifications to analyze their relationships. Finally, the opinions reported by consumers were analyzed by extracting the most frequent n-grams and expressed in a word cloud.

All analyzes were performed using the R software (R Core Team, 2021).

4.3 Results and Discussion

4.3.1 Twitter food data analysis

The collection of tweets about "food + pandemic" in Portuguese, from January to May 2021, obtained 115,382 tweets. The foods that presented the highest and lowest

frequency were coffee (19.59%) and fruits (0.87%). The terms linked to the pandemic with the highest and lowest frequency were pandemic (43.10%) and quarantine (8.38%).

In Figure 9, it can be seen that the terms pandemic and covid (curves in black and blue) are more frequent for both foods and places of purchase. The term quarantine (yellow curve) stands out with coffee and chocolate foods. Crisis and viruses (red and purple curves) are more frequent for bread and wine.

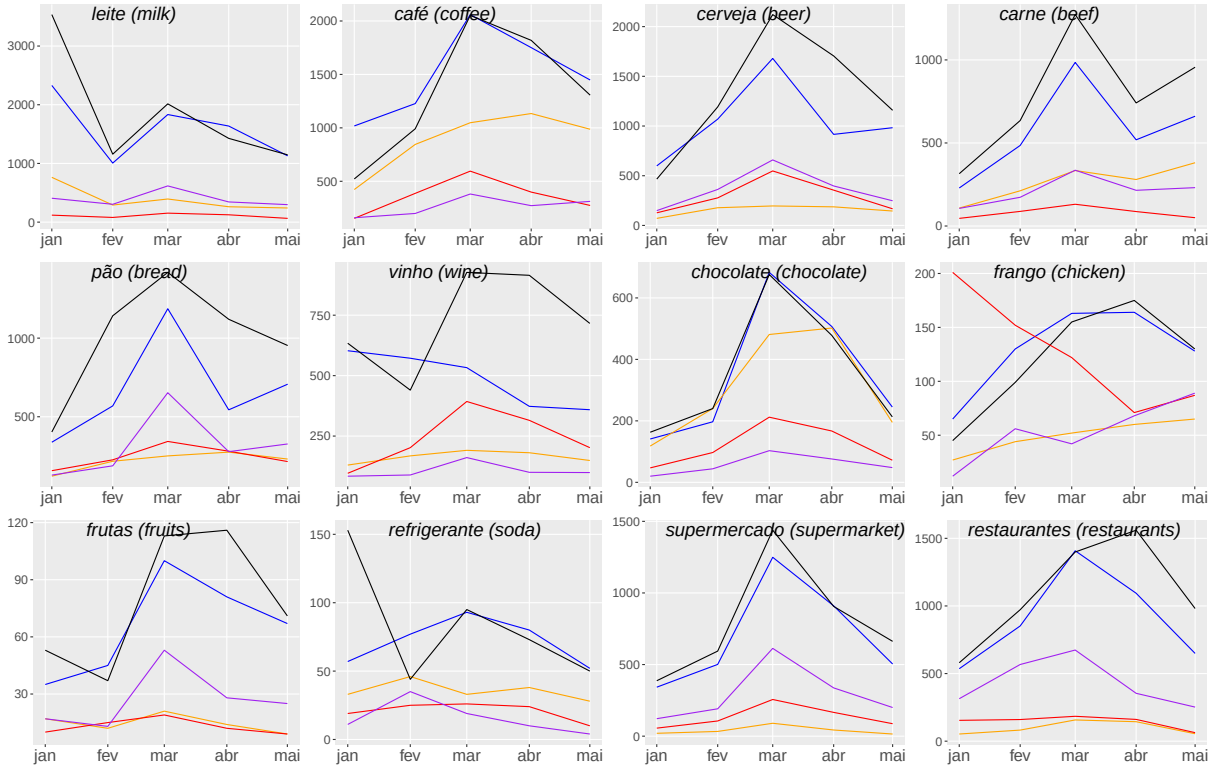


Figure 9 – Timeline for the frequency of the first semester of 2021 publications on Twitter with different places and food products by Brazilians, together with the words: pandemic, covid, quarantine, crisis, and virus, represented by the colors: black, blue, yellow, red, and purple, respectively.

The period from January to February shows a drop in tweets with the terms milk + (pandemic, covid, quarantine), wine + (pandemic, covid), chicken + (crisis), fruit + (pandemic, virus), and soda + (pandemic). In March 2021, the worst month of covid contagion in Brazil, it can be seen in Figure 9 that tweets about food and the pandemic started to show a downward trend in publication frequency. It was being a point of diminishing association between food and pandemic on Twitter. Thus, through Figure 9, it can be seen that there is an association between food and pandemic terms on Twitter.

The same data collected on Twitter has now been grouped by pandemic terms, with the curves in Figure 10 representing the food. A word cloud containing the most frequent words within this search was extracted for each term (pandemic, covid, crisis, and quarantine), as shown in Figure 10.

When searched in Portuguese, the most frequent words on Twitter about: "pandemic

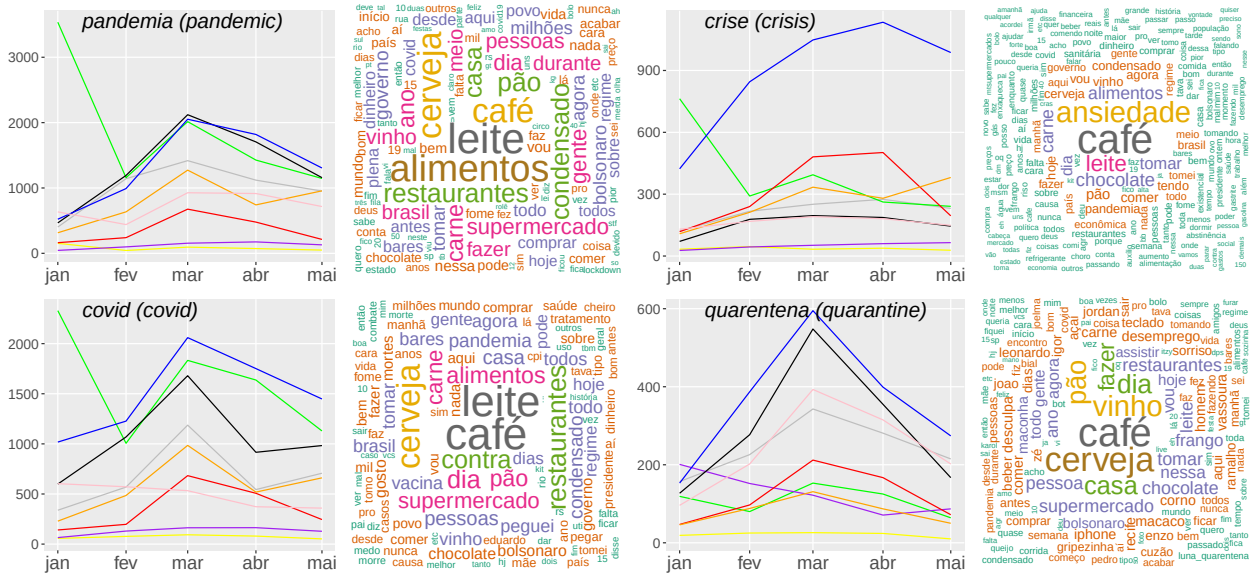


Figure 10 – Representation of the terms: pandemic, covid, crisis, and quarantine by Brazilians on Twitter, together with different food products and their respective word clouds. The curves: green, blue, black, yellow, orange, red, pink, and purple represent the foods: milk, coffee, beer, fruit, meat, chocolate, wine, and soda, respectively.

+ food" were (Milk, coffee, beer, and food). "Crisis + food" stood out for (coffee and anxiety). "Covid + food" stood out (Milk, coffee, and beer). "Quarantine + food" features (Coffee, beer, wine, house, and bread) as the most frequent words from January to May 2021.

Milk and coffee are the most frequent foods among pandemic terms. Milk (green curve) in Figure 10 appears to fall in all terms from January to February. In this period, Brazil went through the "Condensed Milk" scandal. According to data from the Ministry of Defense, more than R$ 14 million were spent on condensed milk, negatively impacting Twitter throughout the country (CNN BRASIL, 2021).

Thus, it can highlight that "covid" and "pandemic" were mentioned more on Twitter in Portuguese than when compared to "crisis" and "quarantine" from January to May 2021. There was an increase in the association between food in January to March and pandemic terms, and from March to May, most of the joint surveys showed a downward trend in tweets published in Portuguese.

4.3.2 Lexical association

The lexical association aspires to measure the degree of association between two or more words taking into account the number of words between them (CHAUDHARI; DAMANI; LAXMAN, 2010). In general, this pair of words frequently appear in the text or are very close to each other (EVERT; KRENN, 2001). This concept is beneficial to understand the relevance and impact of a bigram within a sentence, especially when talking about posts on Twitter, which express opinions in a short text format (KANG,

Table 3 – Number of words in the middle of the searched keywords (pandemic terms + food)

COVID					
Palavras	Café	Cerveja	Vinho	Refrigerante	Chocolate
0-5	40,39%	38,81%	37,13%	36,62%	35,36%
6 - 11	29,75%	32,80%	31,06%	31,63%	31,76%
12 -17	15,60%	15,43%	15,84%	14,31%	16,38%
>18	14,72%	13,51%	15,92%	17,72%	16,70%
CRISE					
Palavras	Restaurantes	Café	Chocolate	Carne	Cerveja
0-5	44,86%	44,50%	39,99%	34,25%	32,75%
6 - 11	23,44%	29,52%	32,23%	27,04%	35,30%
12 -17	13,80%	13,70%	13,68%	17,95%	15,51%
>18	18,12%	11,81%	13,49%	21,19%	16,00%
QUARENTENA					
Palavras	Refrigerante	Vinho	Cerveja	Pão	Café
0-5	53,94%	47,71%	45,19%	41,60%	40,48%
6 - 11	25,94%	30,37%	29,05%	24,99%	28,10%
12 -17	11,94%	11,83%	14,11%	15,96%	16,25%
>18	7,94%	10,08%	11,24%	17,48%	15,12%

2018).

In this way, the number of words present among the critical terms collected on Twitter, which involve Pandemic + Food, was verified in Table 3 to demonstrate which terms of the pandemic are more strongly linked to food in the publications collected from Twitter. The smaller the number of words between the key terms, the greater the association between pandemic + food within the publication.

In Table 3, the terms Quarantine + Soda (53.94%) were the highest percentages in up to 5 words separating the terms. The terms Crise + Restaurants (44.86%) and Covid + Café (40.39%) were the ones that stood out among all the other studied terms present in Figures 9 and 10, with up to 5 words separating the terms. The terms Crisis + Meat were the ones that stood out for presenting more than 20% of Tweets with more than 18 words separating the terms, which may indicate that the association between Crisis + Meat is weak.

After that, we studied which terms present in Table 3 have a higher frequency of 1 to 5 words separating the key terms, the result is shown in Figure 11.

The terms Quarantine + Bread (23.9%) stand out with a higher percentage of just one word separating the terms compared to all other searched terms. Covid + Soda were the terms that obtained the highest rate of 2 and 4 words separating the terms. With three words separating the terms Quarantine + Soda (38.3%) and five words that had the highest frequency were the terms Crisis + Restaurants (35.7%), as shown in Figure 11.

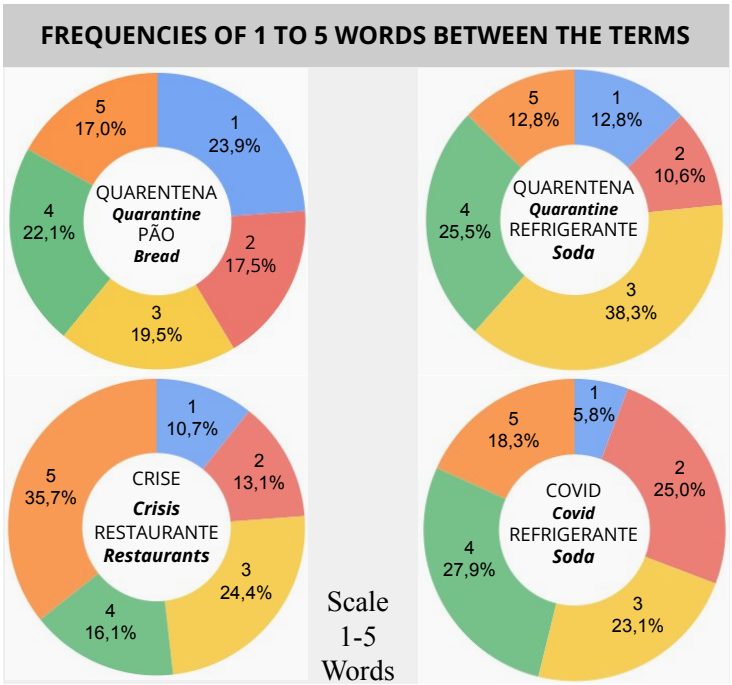


Figure 11 – Frequencies of 1 to 5 words between the terms searched (padémic + food)

4.3.3 Response to the questionnaire for Brazilians consumer's

Consumers participating in the survey were 33.4% from Minas Gerais and 26.4% from São Paulo, Bahia, Goiás and Rio Grande do Sul had participation above 7% and the states with the lowest participation were Pernambuco, Paraíba, Ceará and Mato Grosso.

Before the blocking measures, 95.2% of consumers bought their food in supermarkets and 4.8% in small stores. After the isolation measures, the demand for supermarkets showed 85.6%, small stores 7.4%, and online shopping appeared 7%. The chi-square test with p-value=0.0003 confirmed the dependence between the place of purchase and pandemic. The hypothesis of migration of consumers from supermarkets to online shopping after the restrictive measures was significant using the Mc-Nemer test ($p<0.0009$).

Before the pandemic, the frequency of visits to shopping locations was 45.8% weekly, 27.4% twice a week, 15.9% monthly, 6.3% daily, and 4.6% biweekly. After blocking measures, a significant change ($p=0.001$) in purchase frequency was observed. 49% of consumers started shopping weekly, 27.8% monthly, 13.4% twice a week, 8.6% biweekly, and 1.2% daily. The migration from shopping twice a week to once a week was confirmed by the Mc-Nemar test with a p-value <0.0008.

The results showed that 55.7% of the participants said they frequent restaurants before the isolation measures, 23.7% ordered for takeaways, and 10% bars and Fast-Food. With the isolation measures, the demand for ready-to-eat foods via delivery rose to 76.4%, restaurants dropped to 6.6%, fast food 5.0%, and 11.2% chose not to buy ready-to-eat foods, these changes are significantly using the chi-square test with p-value < 0.0001.

Before the isolation measures, the most frequented place to consume ready-to-eat foods

was the restaurant. With the isolation imposed by the pandemic, 43.6% of participants said they migrated to takeaway orders. Only one participant said they moved to a restaurant after the pandemic. This statement was verified using the Mc-nemar test and was confirmed with a p-value < 0.0002.

Before isolation measures, the frequency of consumption of ready-to-eat foods was 46.8% weekly, 24.2% monthly, 15.4% twice a week, 9.8% daily, and 3.6% biweekly. The frequency after the isolation measures that showed a significant change ($p<0.0004$) was the reduction of 12.6% from weekly to monthly purchases, using the Mc-Nemar test.

The foods that participating consumers reported buying more and less after the pandemic, with a participation frequency greater than 4%, are shown in Table 4.

Table 4 – Changes in food purchase identified through an online survey with Brazilians. Information regarding the percentage of participants who purchased more and less food during the pandemic.

Food			
Particiants buying more	%	Particiants buying less	%
Fruits	23.37	Beef	35.84
Rice	18.42	Soda	20.79
Beef	17.82	Candies	13.86
Bread	17.03	Bread	10.50
Egg	12.48	Chocolate	8.32
Verdure	11.09	Fruits	7.72
Bean	10.10	Milk	6.93
Milk	9.90	Chips	6.14
Chocolate	8.91	Verdure	6.14
Chicken	8.71	Beer	5.94
Juice	8.12	Rice	5.54
Vegetables	7.92	Snacks	4.95
Soda	6.53	Cheese	4.95
Pasta	5.94	Canned	3.96

Of the participating consumers, 23.37% reported that fruit had a more significant share of purchases after the pandemic, followed by 18.42% who started to buy more rice and 17.82% meat. The three most purchased foods with the lowest frequency of responses were Vegetables (7.92%), soda (6.53%), and pasta (5.94%), as shown in Table 4.

The three products that showed a reduction in purchases during the pandemic with the highest frequency of responses were meat with 35.84%, soft drinks (20.79%), and sweets (13.86%). Snacks (4.95%), cheese (4.95%), and canned (3.96%) were the products with less frequency among the least purchased foods.

Some foods appeared on both lists in Table 4, indicating that some consumers increased their consumption while others decreased. The foods on both lists are fruits, rice, meat, bread, vegetables, milk, chocolate, and soda.

Consumers who reported buying more fruits, meats, and vegetables, justify that these foods are healthy. But those who said buying more rice, bread, and milk were justified:

I'm out of stock, and I'm afraid of running out. The justifications for having bought more chocolate and soda were because it helps me with stress.

Similarly, consumers who said they buy less meat, rice, chocolate, and soda on justifications were this expensive, and I need to save. Fruits, bread, vegetables, and milk products reported as less purchased were justified as having a short shelf life and difficulty going to the places of purchase.

In Figure 12, it is possible to observe the summary of the most frequent justifications reported by the participants for the most and least purchased products. Sweets, meat, and soft drinks had a more excellent rationale for reducing purchases. On the other hand, meat, fruits, beans, and bread were the most justifications for the most purchased foods. Meat appears in the two motives for more and less purchased foods after the pandemic, as shown in Figure 12.

In the analysis of meat results, the most frequent justifications for reducing purchases were, "It's expensive" with 19.2%, followed by "I can't buy" 16.2% and "I need to save" 1.2%. The justifications for those who started to buy more meat were, "Like it more" (5.4%), "Eat daily" (5.2%), and "I can't buy daily" (4.4 %). Participating consumers who reported buying more meat during the pandemic were 60% men and those who said they reduced their meat purchases during the same period were 55% women.

The most frequent justifications for reducing purchases (Figure 12) were: It's not healthy (candy and soda), It's expensive (beef), and Don't like anymore (candy and soda). For the most purchased foods, the most frequent justifications were: Like it more (beef and bread), Healthy (fruits), fear of ending (Rice and beans).

The association between food and justifications was found to be significant with a p-value <0.002. Thus, the rationale between the most and least purchased foods after the pandemic was studied in Figure 13 through two correspondence analyses.

In Figure 13a in the upper left corner, the association between the most purchased foods and motives show that candy, chocolate, and sandwich are associated with help with stress. In the same way, Coffee and cheese are associated with make me happy. Next to these motifs are beer, wine, soda, and cookies. In the lower-left corner of Figure 13a, we have the association between pizza and potato with like it more. The fear of ending motif in the upper right corner is related to the Rice, Milk, Flour, Bean, and biscuit foods. Next to them are the reasons I don't buy daily and out the stock. Eat daily appears in the lower right corner of Figure 13a associated with bread, chicken, and juice. It's cheap is associated with eggs, and healthy seems to be associated with grains.

The association for less purchased foods and reasons (Figure 13b) shows in the upper left corner that Rice, oil, beef, yogurt are associated with justifications, it's expensive, and I need to save. In this same region of Figure 13b, there are chicken, cold food, and nut foods. In the upper right corner, soda, biscuit, and canned are associated with unhealthy food. In this same position, beer, candy, and cold are associated with don't like anymore.

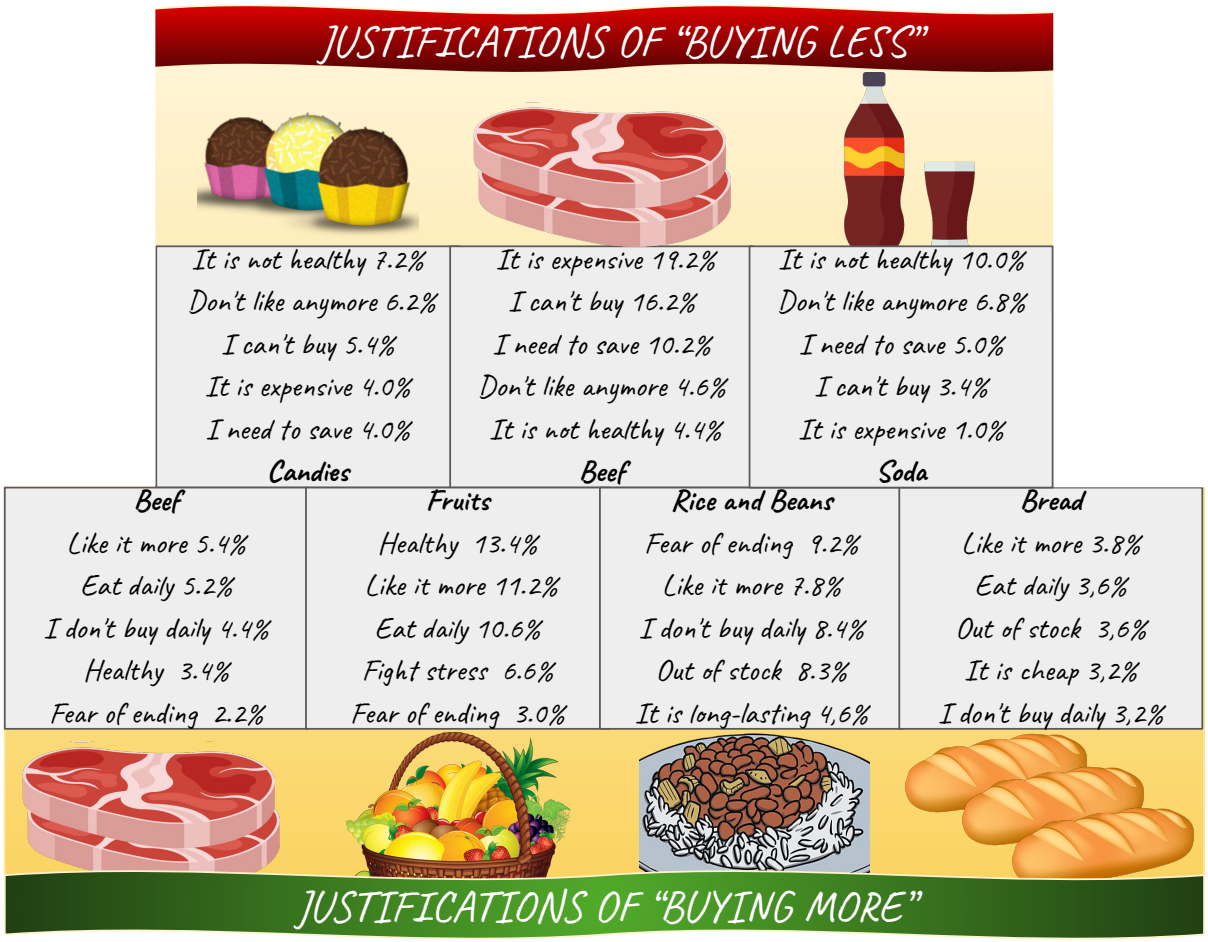
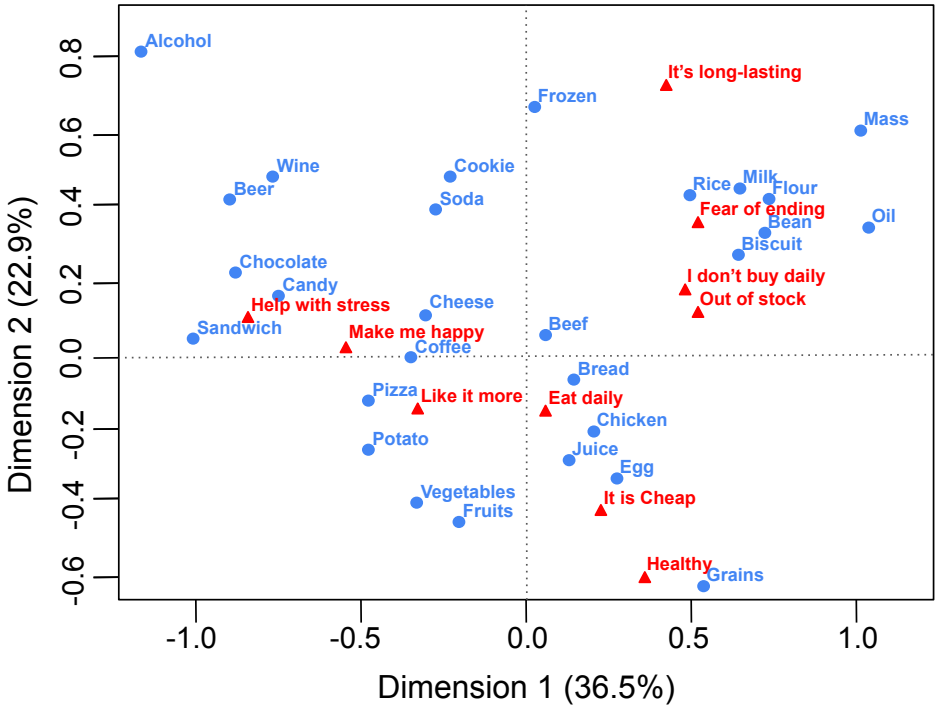


Figure 12 – Comparative table of justifications presented by research participants, for the most and least purchased foods, during the restrictive measures of covid-19.

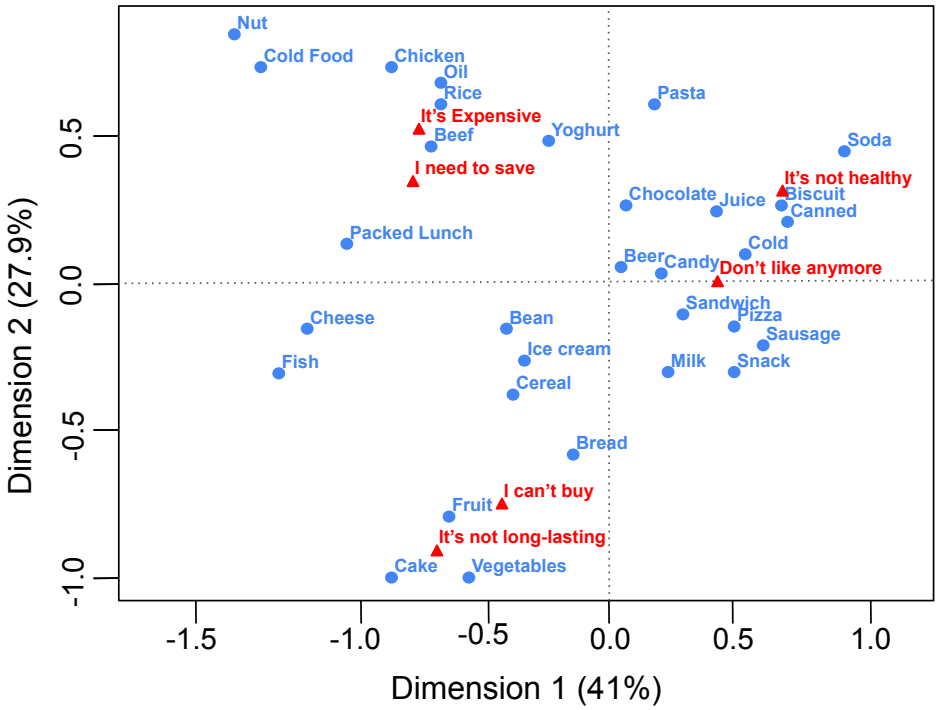
Next to this reason are sandwiches, pizza, sausage, milk, and snack foods. And lastly, in the lower-left corner, fruit, cake, and vegetables are associated with I can't buy, and it's not long-lasting.

The majority of consumers participating in the survey (55%) wrote their opinion about the pandemic crisis. This was analyzed using unigrams, bigrams, and trigrams, as shown in Figure 14.

The most frequent unigrams, food, prices, pandemic, products, and expensive (Figure 14a), demonstrates the Brazilian consumer's disquiet about food prices during the covid-19 crisis. In Figure 14b, the bigrams, more frequent, better food, staple foods, abusive prices, food insecurity, and healthy eating. Along with ending credit, abusive lockdown, increased consumption, starving, too expensive, offensive increase, and higher spending, it shows us that there has indeed been a change in the consumer of Brazilian food and that this change is associated with financial conditions and measures of isolation as well as food inflation.



(a)



(b)

Figure 13 – Correspondence analysis of the motivations for (a) increasing purchases, (b) decreasing food purchases cited by Brazilian consumers via the online survey.

(a) (b) (c)

Figure 14 – Analysis of the open question in the research for criticism and opinions, through unigrams (a), bigrams (b) and trigrams (c).

4.4 Discussion

This paper desired to investigate the impacts of the covid-19 crisis on Brazilian food priorities in the first half of 2021, one year after the pandemic outbreak. Initial studies on this topic in 2020 from several countries have shown us that covid has caused a change in food consumers worldwide (DURAES et al. 2020).

According to Laguna et al. (2020), at the beginning of the restrictive measures in Spain, the publications related to food on Twitter were aimed at insecurity, concern about getting food, and health. They mentioned social and political issues linked to the government. In this study in early 2021, the Twitter collection showed similar results, as shown in Figure 14. Saleh et al. (2020) report that publications on quarantine in English on Twitter from March to April 2020 had an increase linked to food insecurity issues and fear of the implications of this disease.

Publications by food + pandemic terms in Brazil increased along with the number of cases and deaths by covid-19, as shown in Figure 9 between January and March 2020. Laguna et al. (2020) state that this temporal change of publications on Twitter can be reflected in changes in consumer behavior because their context and practices are constantly changing in this period of uncertainty. On the other hand, Mohamed and Ferraris (2021) add that digital media such as Twitter were essential for managing the covid-19 crisis, as the public increasingly depends on this medium for information, which is reflected in the content be posted by users.

About food, milk was the food that had the highest frequency of publication in the period from January to February 2021. This can be explained by the condensed milk scandal that made this food reach the top trends on Twitter in this period (METROPOLES, 2020). Likewise, coffee, beer, meat, chocolate increased in publications from January to March.

The surveys also showed that posts by crisis + food on Twitter accompanied the words

anxiety and coffee. This feeling has been highly reported in the literature since 2020, as shown in the literature review carried out by Durães et al. (2020). In Brazil, Verticchio (2020) also reports that 54% of the participants in his survey said being anxious about the future of this crisis.

Posts by quarantine + food were accompanied by: coffee, wine, beer, and bread. Sandoval et al. (2018) show that Tweets in Spanish with the word coffee frequently appear in the morning and beer during lunch and dinner. The same author suggests that citing food in Tweets demonstrates a user's eating routine combined with the time of day. As these results are accompanied by quarantine, this indicates that consumers may be using these foods to express their moments of isolation.

Publications on supermarkets and restaurants remained high from January to March 2021, registering a decline after this period.

The participation of Brazilian consumers in the online survey confirms the results of Laguna et al. (2020) and Martinotto et al. (2020) that people's eating habits changed after the pandemic. The results of this study take into account more than ten months of isolation, a period in which the initial Fear had already passed, even during the second wave of cases. This fact influenced the participants' responses, in addition to showing consumers' disquiet about the increase in food and isolation measures in Figure 14.

Consumer interest in online food purchases appeared with the restrictive measures. However, supermarkets were still the most frequented place to buy food.

The interest in ordering for take-outs increased dramatically with the closing of the restaurants, and the isolation measures caused the consumption of food to grow.

Concern for the consumption of healthy foods increased after the isolation measures. The interest in consuming vegetables and fruits in Brazil increased during this period. However, some consumers reported reducing the purchase of foods with a short shelf life (fruits, vegetables, and milk).

Some consumers reported stopping ordering for delivery for Fear of contamination and started to prepare all their meals at home. This is in line with Eftimov et al. (2020) work, who found an increase in searches for recipes on the Allrecipes website.

Regarding the foods declared as being more and less purchased, it was found that the ambiguity of the results is justified by the participants themselves, where consumers who bought more certain food were due to taste, desire, and Fear of ending. In contrast, consumers who decreased purchase frequency were related to the financial issue and the need to save.

According to (WHO), due to the pandemic, tens of millions of people are at risk of losing their livelihoods and falling into extreme poverty (BRZUSTEWICZ, SINGH, 2021). In Brazil, in 2020 inflation in the food sector reached 14.1%, with protein foods such as meat reaching an increase of 44.5% (GALINDO et al., 2021). an increase in families that entered the lower class during the pandemic portrays an explanation of the rise in

consumption by some consumers and the reduction by others.

About the reasons for the most purchased foods during the pandemic, we found that: candy, chocolate, and sandwich are associated with help with stress. This is consistent with the work of Maia (2020), which shows that the pandemic increased cases of stress and depression, especially among young people and university students, which led to an increase in the consumption of unhealthy foods as a way to escape these symptoms caused by these evils.

As for the less purchased foods, the results showed that Rice, oil, beef, yogurt are associated with justifications, it's expensive, and I need to save. According to Galindo et al., 2020, these foods will undergo a significant increase in 2020, where oil increased by 103.79%, rice (76.01%), fruits (25.40%), and meat (17, 97%), where this increase was maintained in 2021. In his research in the first half of 2020, the same author highlights that 52.8% of the participants reported keeping their meat consumption, and 44% said they had reduced consumption.

Consumers report that the food change was due to low financial conditions, food inflation, and longer time at home, which confirms the hypothesis raised by Durães et al. (2020), at the beginning of the pandemic, where there is a change in consumer behavior, in case the effects of the pandemic do not go away quickly.

4.5 Conclusions

This article sought to verify the effects of the covid-19 crisis, one year after its emergence, on the food priorities of Brazilians and verify the necessary changes to adapt to the new reality after long months of isolation. Crisis + food surveys on Twitter showed us that consumers are anxious and are using coffee in this period. Isolation has also made consumers, in their vast majority, adopt orders for delivery, reducing the frequency of purchase from twice a week to weekly. The foods most purchased by the participants were fruits, and rice and the least bought were Meat and Soft Drinks. The search for a healthy diet during the isolation period was present in the justifications given by the participants. However, sweet chocolates and sandwiches were associated with help with stress, and the same foods appear in tweets searched by quarantine + food. Coffee and cheese are associated with make me happy. Rice, Milk, Flour, Bean and biscuit are related to Fear of ending.

Given the above, this work was able to show that food consumers still suffer the impacts generated by the crisis. I had a lot of times to reduce expenses by removing products from the shopping list. However, foods that appear to be more and less bought receive opposite justifications, such as I need to save and buy because I like it a lot. We also saw a positive association between food and pandemic terms on Twitter.

4.6 References

AQUINO, E, et al. **Medidas de distanciamento social no controle da pandemia de COVID-19: potenciais impactos e desafios no Brasil.** Ciência Saúde Coletiva, v. 25, p. 2423-2446, 2020.

CNN BRASIL. PEIXOTO, S, **Defesa usou 307 latas de leite condensado por hora.** Disponível em: https://www.cnnbrasil.com.br/nacional/lisauskas-em-2020-defesa-usou-307-latas-de-leite-condensado-por-hora/. Acesso em: 15 setembro de 2021.

CHAUDHARI, D; DAMANI, O, P.; LAXMAN, S. **Lexical co-occurrence, statistical significance, and word association.** arXiv preprint arXiv:1008.5287, 2010.

MAYNARD, D et al. **Consumo alimentar e ansiedade da população adulta durante a pandemia do COVID-19 no Brasil.** Research, Society and Development, v. 9, n. 11, p. e4279119905-e4279119905, 2020.

CULLEN MT. **Coronavirus Food Supply Chain Under Strain What to do?** Food Systems Transformation 2020.

DURÃES, S. A. *et al.* **Implicações da pandemia da covid-19 nos hábitos alimentares.** Revista Unimontes Científica, v. 22, n. 2, p. 1-20, 2020.

DOMINGUES, C, M, A, S. **Desafios para a realização da campanha de vacinação contra a COVID-19 no Brasil**. 2021.

EFTIMOV, T et al. **COVID-19 pandemic changes the food consumption patterns.** Trends in food science technology, v. 104, p. 268-272, 2020.

EVERT, S; KRENN, B. **Methods for the qualitative evaluation of lexical association measures.** In: Proceedings of the 39th annual meeting of the association for computational linguistics. 2001. p. 188-195.

FAO. Q and A: **COVID-19 pandemic – impact on food and agriculture**. 2020.

FEINERER I, HORNIK K. **tm: Text Mining Package. R package version 0.7-8**, 2020 https://CRAN.R-project.org/package=tm.

FELLOWS, I et al. **Package 'wordcloud'. R Package, Maintainer Ian and Rcpp,**

Linking To and Rcpp., 2018.

GALINDO, E et al. **Efeitos da pandemia na alimentação e na situação da segurança alimentar no Brasil.** 2021.

HADLEY W, et al. **dplyr: A Grammar of Data Manipulation. R package version 1.0.7**, 2021. https://CRAN.R-project.org/package=dplyr

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

KEARNEY, M. W.; **Package 'rtweet'**, 2018.

KANG, Beom-mo. Collocation and word association: Comparing collocation measuring methods. International journal of corpus linguistics, v. 23, n. 1, p. 85-113, 2018.
LAGUNA, L et al. **The impact of COVID-19 lockdown on food priorities. Results from a preliminary study using social media and an online survey with Spanish consumers.** Food quality and preference, v. 86, p. 104028, 2020.

MALTA, D, C, et al. **A pandemia da COVID-19 e as mudanças no estilo de vida dos brasileiros adultos: um estudo transversal**. Epidemiologia e Serviços de Saúde, v. 29, 2020.

MARTINOTTO, P. et al. **Hábitos alimentares de trabalhadores de um Centro Universitário: mudanças durante a pandemia de COVID-19.** VIII Congresso de Pesquisa e Extensão da FSG. (2020) http://ojs.fsg.br/index.php/pesquisaextensao.ISSN 2318-8014.

MAIA, B, R; DIAS, P, C. **Ansiedade, depressão e estresse em estudantes universitários: o impacto da COVID-19. Estudos de Psicologia** (Campinas), v. 37, 2020.

METROPOLES. **Gastos do governo federal com leite condensado viram o assunto mais comentado do Twitter.**

Disponível em: https://www.metropoles.com/brasil/gastos-do-governo-federal-com-leite-condensado-viram-o-assunto-mais-comentado-do-twitter Acesso em: 15 setembro de 2021.

MOHAMMED, A; FERRARIS, A. **Factors influencing user participation in social**

media: Evidence from twitter usage during COVID-19 pandemic in Saudi Arabia. Technology in Society, v. 66, p. 101651, 2021.

PAULO F, D; ARAUJO, F, F. **Will COVID-19 affect food supply in distribution centers of Brazilian regions affected by the pandemic?** Trends in Food Science Technology, v. 103, p. 361-366, 2020.

POUDEL, P, B et al. **COVID-19 and its global impact on food and agriculture.** Journal of Biology and Today's World, v. 9, n. 5, p. 221-225, 2020.

PUDDEPHATT, J et al. **'Eating to survive': A qualitative analysis of factors influencing food choice and eating behaviour in a food-insecure population.** Appetite, v. 147, p. 104547, 2020.

RIBEIRO, S, R, C et al. **Implicações da pandemia COVID-19 para a segurança alimentar e nutricional no Brasil.** Ciência Saúde Coletiva, v. 25, p. 3421-3430, 2020.

R Core Team (2021). **R: A language and environment for statistical computing. R Foundation for Statistical Computing,** Vienna, Austria. URL https://www.R-project.org/.

SALEH, S, N. et al. **Understanding public perception of coronavirus disease 2019 (COVID-19) social distancing on Twitter.** Infection Control Hospital Epidemiology, v. 42, n. 2, p. 131-138, 2021.

SNUGGS, S; MCGREGOR, S. **Food meal decision making in lockdown: How and who has Covid-19 affected?**. Food quality and preference, v. 89, p. 104145, 2021.

TWITTER, **Developer Platform: Twitter API, 2021.** Disponível: https://developer.twitter.com api. Acessado em: 16/07/2021

VERTICCHIO, D. F. R.; VERTICCHIO, N. M. **The impacts of social isolation about changes of eating behavior and weight gain during the COVID-19 pandemic in Belo Horizonte and metropolitan region, State of Minas Gerais, Brazil.** Research, Society and Development, [S. l.], v. 9, n. 9, p. e460997206, 2020.

WICKHAM, H. et al **dplyr: A Grammar of Data Manipulation. R package version 1.0.4.**, 2021 [https://CRAN.R-project.org/package=dplyr]

WORLD HEALTH ORGANIZATION (WHO) et al. **Coronavirus disease 2019 (COVID-19): situation report**, 94. 2020.

5 COVID-19 IN BRAZIL: CORRELATION BETWEEN PANDEMIC DATA AND FOOD KEYWORDS ON TWITTER

Abstract: The COVID-19 pandemic in 2021 has already provoked more than 400 million cases and more than 4.5 million deaths worldwide. The restrictive measures assumed to limit the spread of the virus caused a transformation in the population's lifestyle in Brazil, which impacted the habits of Brazilians. The social network Twitter, during the pandemic, became an environment for sharing the daily routine of its users. This turned Twitter into many scientists and researchers' interest in gaining access to spontaneous information linked to users' real-life everyday situations. The article aims to answer the question: Is there a correlation between the rates of cases, deaths and vaccinates by COVID-19 in Brazil, with the publications on Twitter in this period? For this, 10 million tweets from 10,000 Brazilian users were collected. These data were separated between the first and second wave of COVID-19 cases and deaths in Brazil and separated in terms of the pandemic, establishments, and food. Results show that the rate of COVID-19 cases in Brazil proved to be the most correlated with posts on Twitter in the second wave period, with the highest correlations being the term pandemic (0.95), Restaurants (0.80), and Chicken (0.80). 0.85), along with the case rate. In addition, the frequencies of tweets are influenced by controversial news published in this period. It follows, then, that posts on Twitter can be correlated with real-life posts on Twitter, even when analyzed over long periods.
Key-words: COVID-19, Twitter, Correlation

Resumo: A pandemia de COVID-19, em 2021 já acarreta mais de 400 milhões de casos e mais de 4,5 milhões de mortes no mundo todo. As medidas restritivas adotadas para limitar a disseminação do vírus, provocou uma mudança no estilo de vida da população no Brasil, que afetaram os hábitos dos brasileiros. A rede social Twitter, durante a pandemia se tornou um ambiente para compartilhamento da rotina diária dos seus usuários. O que tornou o Twitter a ser interesse de muitos cientistas e pesquisadores, por obter acesso a informações espontâneas, ligadas ao cotidiano dos usuários. Pensando nisso, este artigo tem como objetivo responder a pergunta: Existe correlação entre as taxas de casos, mortes e vacinados por COVID-19 no Brasil, com as publicações no Twitter neste período? Para isso, foram coletados 10 milhões de tweets de 10.000 usuários brasileiros. Estes dados foram separados entre a primeira e segunda onda de casos e mortes pela COVID-19 no Brasil, e separados em termos da pandemia, estabelecimentos e alimentos. Resultados mostram que a taxa de casos de COVID-19 no Brasil, se mostrou ser a mais correlacionada com as publicações no Twitter no período da segunda onda, com as maiores correlações sendo o termo Pandemia (0.95), Restaurantes (0.80) e Frango (0.85). Além das frequências de publicações dos tuítes serem influenciadas por noticias polemicas publicadas neste período. Conclui-se, então que publicações no Twitter podem ser correlacionadas a publicações da vida real, mesmo quando analisadas por grandes períodos de tempo.
Palavras-chave: COVID-19, Twitter, Correlação

5.1 Introduction

The COVID-19 pandemic has been the primary anxiety the world population faces in the last 20 years (BRZUSTEWICZ, SINGH, 2021). The severe respiratory infection identified in Wuhan, China, already causes more than 400 million cases and more than 4.5 million deaths worldwide (WHO, 2021). The effects caused by the pandemic surpassed the damage to health. They reached commerce, industries, transport, tourism, and the financial market, which led countries to create forms of fiscal stimulus to keep the economy turning (FANELLI, 2021). As restrictive measures were the only way to limit the spread of the virus, a change in the lifestyle of the Brazilian population was observed, which affected the eating habits of Brazilians (MALTA, 2020).

Numerous authors worldwide have raised the hypothesis that consumers' behavior and consumption patterns would change with the arrival of COVID-19 (DURÃES et al., 2020). As time passed, this hypothesis was confirmed with the works showing that the changes caused by the pandemic affected the lifestyle, purchase intentions, and how products and services started to be consumed (BRZUSTEWICZ, SINGH, 2021).

During this period, access to the internet and social networks had a substantial increase (BALTAZAR, 2020). According to surveys carried out by wearesocial in 2020, 120 million people had access to the internet for the first time, representing an increase of 2.6% compared to the previous year. On the other hand, social networks had an entry of 400 million new users in 2020, representing an increase of 5.2% compared to 2019 (WE ARE SOCIAL, 2020).

Twitter alone showed growth of 26% in the year 2020, reaching the mark of 330 million active accounts worldwide (WE ARE SOCIAL, 2020). Twitter permits share messages of up to 280 characters, along with images, videos, links, and emotions. Twitter has become an ecosystem for sharing users' daily routines. Many scientists and researchers became interested in gaining access to spontaneous information linked to real-life situations (VIDAL et al., 2015).

Vidal et al. (2015) report that the information present on Twitter is spontaneous and represents the daily routine of food consumers. It also replies that Twitter is a real-life environment that makes it possible to search for food situations, as consumers are more connected to the internet (VIDAL et al., 2015). Along the same lines, Sandoval et al. (2018) report that Twitter data is a precious source of information about consumer behavior in different contexts and domains. However, the same author warns that local variables can influence consumers and be subjective when making a post. They seek to be socially desirable or very rationalized. They tend to repeat comments already made on the platform. Quercia et al. (2011) supplement that Twitter users tend to reveal a lot about themselves, taking care not to disclose their confidential data, which implies that Twitter posts are reports of the reality and daily life of users.

Given the potential that Twitter offers for data collection to study consumers in general and food, we can detach some works that used Twitter. Vidal et al. (2015), in their research on "what people say when they tweet about different eating situations." Sandoval et al. (2018) use an algorithm to provide insights into consumer eating behavior, presenting Twitter's potential in consumer research outside of laboratories. Mostafa (2017) uses Twitter data to map halal food consumers and their location, detaching that the religious diaspora uses social media to communicate about halal food. Zhou, Liu, and Zhou (2018) collect tweets about healthy foods to promote health education, identifying which tweets about healthy foods people like to retweet. Widener and Li (2014) use geolocation data to identify user groups reporting healthy and unhealthy foods in the United States. Abbar, Mejova, and Weber (2015) conducted a Twitter study to identify Americans who publish what they eat.

About work carried out with COVID-19, the (FAO) executed a big data laboratory to monitor the impacts of COVID-19 on the food supply. To this, 270 newspaper Twitter accounts worldwide were scanned to identify news about food (FAO, 2021). Brzustewicz and Singh (2021), authors linked to the world health organization, used the algorithm (LDA) to study the topics of sustainable consumption in consumer behavior during the pandemic through data collection on Twitter. Pilar, Stanislavská, and Kvasnička (2021) research vegan and organic food consumption on the social network Twitter and show that users tend to associate healthy foods with a lifestyle, diet, and diet physical fitness. Mohammed and Ferraris (2021) conducted a study in Saudi Arabia to explain why the population actively participated in Twitter during COVID-19. Their results found that people feel voluntary in the process of sharing information about crises on Twitter.

On the other hand, the authors presented below sought to verify the correlation between Twitter publications and cases of infectious diseases. Aramaki, Maskawa, and Morita (2011) propose the use of the Support Vector Machine algorithm to identify Tweets that contain the words "influenza," with results that show a correlation of 0.97 between the Tweets collected by the algorithm and the actual data, with an alternative being to use for identifying outbreaks of epidemics. Shin (2016) conducts a study on tweets and Google searches to identify epidemics of respiratory syndromes (MERS-CoV). The results showed a correlation above 0.7 between google searches and Twitter data with the number of MERS cases. Sun and Gloor (2020) sought to correlate peaks in research and Twitter posts about COVID-19 with the number of COVID-19 infections in 50 US states. Their results show a high correlation between Google searches and tweets with terms related to the pandemic and cases of COVID-19. However, the same author also points out that the more advanced the peak of google searches for COVID-19 information, the lower the infection rate registered locally.

Thus, this article aims to answer the question, "Is there a correlation between the rates of cases, deaths, and vaccinated in Brazil, with publications on Twitter? The analyzes

will be divided into the first and second waves of cases and deaths by COVID -19 in Brazil. The mined tweets of 10,000 users were obtained through a sampling process that contained the terms of the pandemic (Covid, Pandemic, Vaccine, and Lockdown), the locations (Gym, House, Delivery, Supermarket, and Restaurants), and the food (Rice, Beef, Candy, Beans, Chicken, Fruits, Bread, and Soda) were considered for the study.

5.2 Materials and Methods

The data provided by the Ministry of Health, referring to the number of cases, deaths, and vaccinations in Brazil, were used to carry out the analyzes (BRASIL, 2021). These data were separated into two periods, from February to October 2020 and from November 2020 to July 2021, considered respectively as the period of the first and second wave of covid-19 infections in Brazil. The collected and filtered tweets also went through this division, obtaining two groups referring to the first and second waves.

5.2.1 Twitter data collection

Collected Twitter data was accessed via a free application programming interface (API), allowing users with an approved developer account to collect Twitter data in a restricted manner (TWITTER, 2021). This API was used in Rstudio, with the R software with the rtweet package that supports the Twitter API and allows for the collection in different ways (Kearney, 2018).

Kearney (2018) detach the two most common forms of data collection on Twitter through the rtweet package: the search for tweets by keyword and readers from a timeline. The first way concedes tweets with the searched keyword, to be collected limited to 18,000 every 15 minutes, ranging from the most recent to tweets with seven days of the publication date of collection (Kearney, 2018). The second way, by collecting tweets from specific accounts, this process is limited to obtaining the last 3200 tweets published by that account. In this case, there is no restriction on collecting old tweets.

Thus, this work collected the first 3200 tweets published by ten thousand Brazilian users, extracted from the database with 500,000 users. This collection obtained a return of 10 million tweets submitted to a filter that retains only tweets from January 2020 to July 2021. In Figure 15, the word clouds of this dataset can be seen, separated into three periods, before the pandemic, first wave of cases, and second wave of contagion.

After that, the tweets were separated by the terms of the pandemic (Covid, Pandemic, Vaccine, and Lockdown), establishments (Gym, House, Delivery, Supermarket, and Restaurants), and food (Rice, Beef, Candy, Beans, Chicken, Fruits, Bread, and Soda) and were grouped monthly.

(a) (b) (c)

Figure 15 – Word cloud referring to the dataset: before the pandemic (a), first wave (b) and second wave (c), respectively.

5.2.2 Data analysis

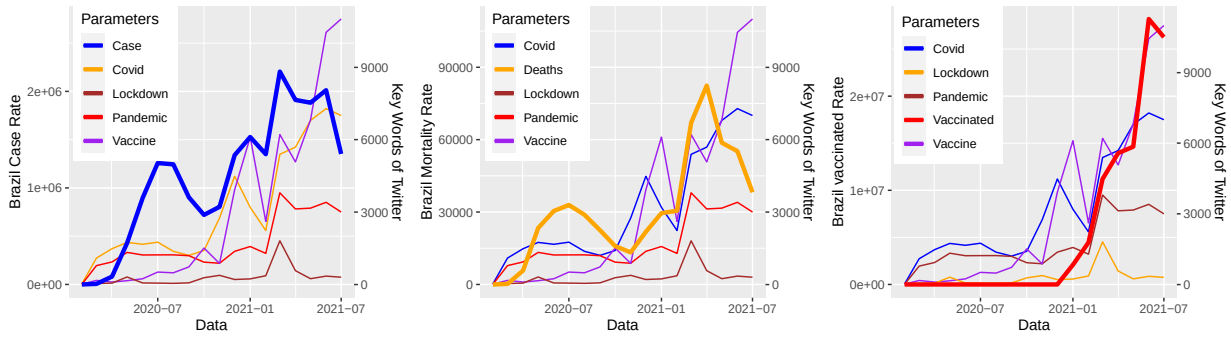
With the tweets separated by each group of terms, a plot of data over time was performed to visually identify the correlation between the terms, as can be seen in Figures (16, 18, and 20). After that, the data were separated into two groups about the first and second wave of cases. As the data distribution does not behave linearly, Spearman's correlation was used to verify the relationship between the terms collected on Twitter and the pandemic data, as shown in Tables (5, 6 e 7), with the indications of the significance of these measures.

5.3 Results

Most of the data obtained from Twitter have political mentions linked to the government acting at the time of collection. In Figure 15, it can be seen that before the beginning of the pandemic, the most frequently used words were: (brazil, life, better, god). In the first and second waves of COVID-19, the highlights were for the government and the president of Brazil, with the most frequent words being (bolsonaro, god, brazil, see).

The pandemic data in Brazil (cases, deaths, and vaccinated) were compared with the terms (covid, pandemic, vaccine, and lockdown) published by 10,000 users on Twitter from the beginning of the pandemic until July 2021, as can be seen in Figure 16. The blue and yellow curve Figure (16a and b) represents the infection and death rates registered monthly, while the red curve represents the vaccinated rate also grouped monthly.

In Figure 16, it is possible to see that the posts related to isolation (lockdown) on Twitter have the lowest frequency, followed by mentions of the pandemic compared to the other terms. Publications of the terms "covid" and "vaccine" are the most frequent when compare vaccine publications on Twitter. It is possible to discern in Figure (16b) that in the first wave of cases, publications on Twitter with the terms of covid remained stable until the arrival of the second wave, which ended up moving Twitter more in Brazil and



(a) (b) (c)

Figure 16 – Plot of Pandemic Data, with pandemic key words of Twitter, (a) cases, (b) deaths, (c) vaccinated.

thus increasing the number of publications on that subject during this period.

According to data from the Ministry of Health, the start of vaccination of the Brazilian population took place on January 19, a period in which Brazil was facing the second wave of contagion (BRASIL, 2020). During this same period, the use of the word vaccine on Twitter experienced a sudden drop that can be seen in Figure 16 (01-2021) in the purple curve. However, the subject quickly returned to a high and remained so until 07-2021.

In Table 5, the correlations between data from the pandemic in Brazil are represented, with the frequency of publication on Twiiter with the terms of the pandemic, separated in first and second wave of contagion.

Table 5 – Correlation of pandemic data in first and second wave of covid: Case, Deaths and Vaccineted, with key words minerated of twitter: Covid, Pandemic, Vaccine and Lockdown.

First Wave				
	Covid	Pandemic	Vaccine	Lockdown
Case	0.52^{ns}	0.68^{ns}	0.73*	0.02^{ns}
Death	0.77*	0.83**	0.48^{ns}	0.14^{ns}
Vaccineted	0	0	0	0
Second Wave				
	Covid	Pandemic	Vaccine	Lockdown
Case	0.60^{ns}	0.95***	0.62^{ns}	0.38^{ns}
Death	0.58^{ns}	0.83**	0.52^{ns}	0.48^{ns}
Vaccineted	0.87**	0.72*	0.85**	0.13^{ns}
ns:p>0.05 *:p<0.05 **:p<0.01 ***:p<0.001				

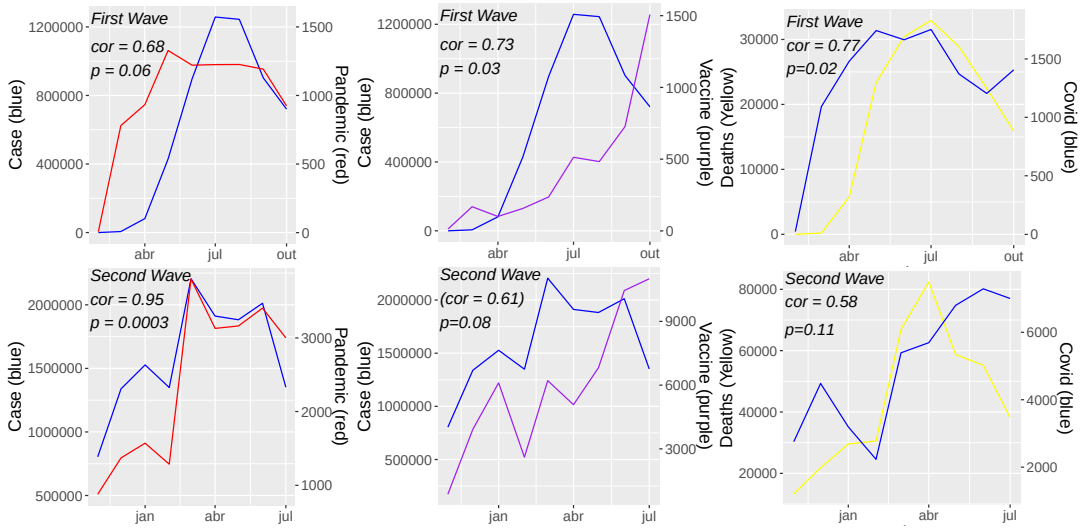
In the first wave, the most effective significant correlation (0.83) was observed between the death rate and the use of the term "pandemic" in Portuguese on Twitter, followed by "covid" and the death rate (0.77). The other measures of correlations (Table 5), among the additional terms, did not obtain a significant result for the first wave of cases. As vaccination had not yet started, the correlation appears as zero.

In the second wave of cases in Table 5, it can be noted that the term pandemic obtained significant correlations for all data from Covid in Brazil. The case rate and

the term pandemic showed 0.95 of correlation, followed by deaths and pandemic (0.83) and vaccinated and pandemic with 0.72 of correlation. Data on vaccinated individuals in Brazil showed a positive correlation with the terms covid (0.87), and vaccine (0.85) published on Twitter.

Table 5 can note differences between the significant correlations between the first and second wave of cases. As is the case of the death rate and the term covid, which in the first wave had a significant correlation of 0.77, and in the second wave, the correlation ceases to be significant with the value of 0.58. The same happens, but oppositely with the case rate and the terms pandemic and vaccine, which started to correlate in the second wave of cases significantly. This representation of the change in correlations between the first and second waves is presented in Figure 17.

In Figure 17, the top graphs represent the curves for the first wave of covid, and the lower ones for the second wave, as described in the legends. The most significant difference between the correlations of the first and second waves is between the case rate and the term pandemic (Figure 17a), which in the second wave presented a correlation of 0.95. Figures (17b and c) show that the correlations between (rate of cases and vaccine) and (rate of deaths and covid) on Twitter stopped to be significant with the arrival of the second wave of covid in Brazil.



(a) (b) (c)

Figure 17 – Comparison between correlations that are no longer significant between the first and second wave of cases.

In Figure 18, the curves of the establishments examined on Twitter are presented along with the pandemic data in Brazil (curves in brown). The yellow curve represents the frequency of the word "casa" in Portuguese and obtains the highest frequency of publication on Twitter since the beginning of the pandemic when compared to other places. In second place with the highest frequency of publications appears the word "academia" in Portuguese (curve in blue), followed by restaurants (curve in red), supermarket (curve

in green), and the one with the lowest frequency of delivery (curve in black).

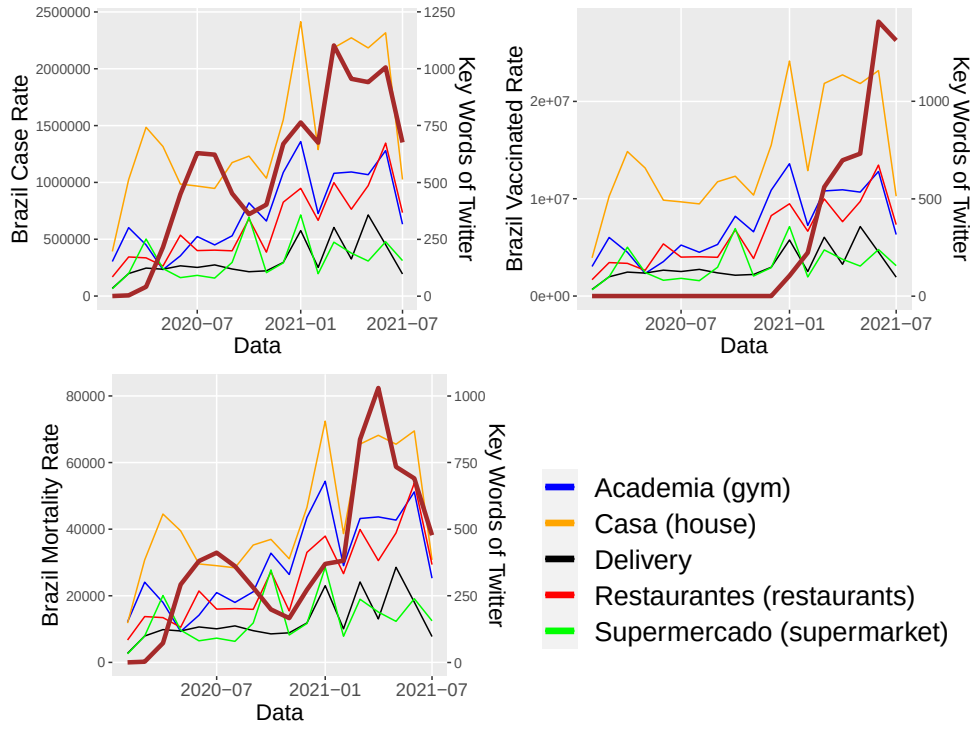


Figure 18 – Bi-plot between pandemic data in Brazil (cases, deaths, and vaccinated), with the frequency of use of places in Twitter (Academy, home, delivery, restaurants, supermarket).

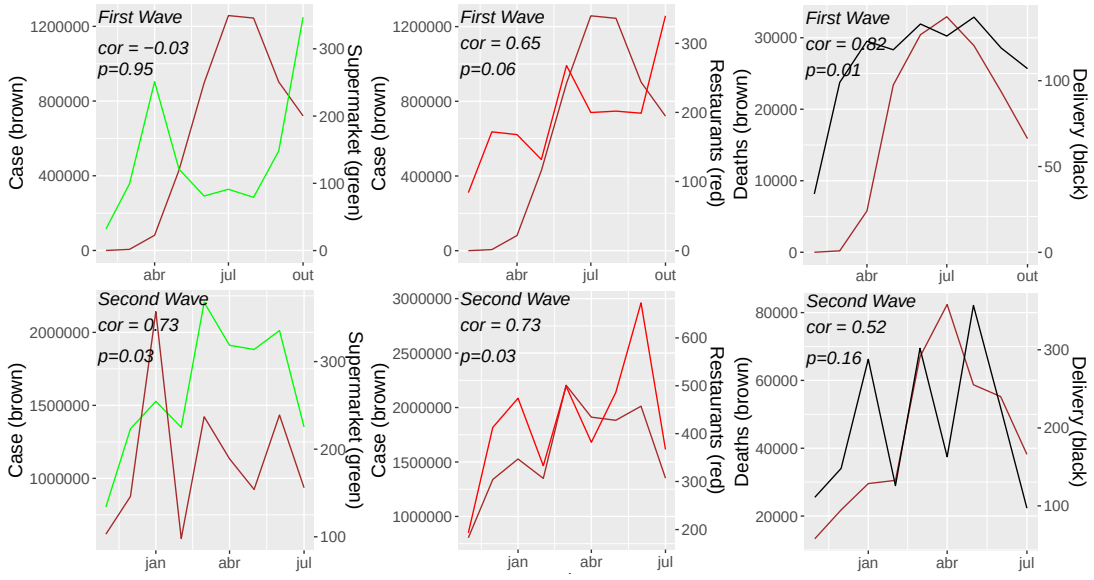
During the first wave of contagion, it is possible to discern that publications by these establishments contain a low frequency compared to the growing rate of covid cases and deaths in Brazil. Supermarkets (green curve) have a lower frequency on Twitter during the analysis period. Only mentions of the term "home" appear in high during the period, followed by "gym" and "restaurants". The increase in Twitter publications about these locations only occurs with the beginning of the second wave, as shown in Figure 18. The level of correlation between the locations on Twitter and the pandemic data is shown in Table 6.

Table 6 – Correlation of pandemic data: case, Death and Vaccineted, with key words minerated of twitter: Gym, House, Delivery, Supermarket and Restaurants.

First Wave					
	Gym	House	Delivery	Supermarket	Restaurants
Case	0.21^{ns}	-0.18^{ns}	0.78^{*}	-0.03^{ns}	0.65^{ns}
Death	-0.12^{ns}	-0.13^{ns}	0.82^{*}	-0.15^{ns}	0.55^{ns}
Vaccineted	0	0	0	0	0
Second Wave					
	Gym	House	Delivery	Supermarket	Restaurants
Case	0.48^{ns}	0.67^{ns}	0.70^{*}	0.73^{*}	0.80^{*}
Death	0.20^{ns}	0.38^{ns}	0.52^{ns}	0.42^{ns}	0.50^{ns}
Vaccineted	0.02^{ns}	0.17^{ns}	0.20^{ns}	0.41^{ns}	0.47^{ns}
ns:p>0.05 *:p<0.05 **:p<0.01 ***:p<0.001					

In the first wave of cases, the significant correlations were between publications with the word Delivery with the death rate (0.82) and the case rate (0.78). The other places published on Twitter did not show a significant correlation, as shown in Table 6. The correlations for the second wave period, present at the bottom of Table 6, with significant results, were between the rate of cases with the terms on Twitter: Delivery (0.70), Supermarket (0.73), and Restaurants (0.80). The other correlations between pairs present in Table 6 did not show significant results.

The changes in the significance of the correlations between the first and second waves are shown in Figure 19. The correlation between the term Supermarket on Twitter with the case rate became significant (0.73) in the second wave period, as shown in Figure (19a). Likewise, the term Restaurant and the case rate showed a significant correlation in the second wave period. Only publications with the term Delivery did not present a significant correlation in Brazil's second wave of covid for these establishments, as shown in Figure (19c).



(a) (b) (c)

Figure 19 – Comparison between correlations that are no longer significant between the first and second wave of cases.

Finally, the comparison between the foods published on Twitter in Brazil with the pandemic data is shown in Figure 20. The food that was most frequent in the sampling process used was sweets (black curve), followed by bread (yellow curve), meat (orange curve), and rice (blue curve).

Posts with food on Twitter during the first wave proved to be less frequent when compared to the second wave period. The curves for foods, rice, fruits, beans, and chicken appear close to each other in almost the entire analysis period. However, they remain in an increasing trend during analysis. Rice alone registered two peaks in publications, one in October 2020 and another in January 2021. In this same period, other foods increased

the rate of publications, such as Fruit, Bread, meat, and sweets. Publications with the use of the word soda (curve in pink) in Portuguese is the one that proved to be less frequent within the study period, as shown in Figure 20.

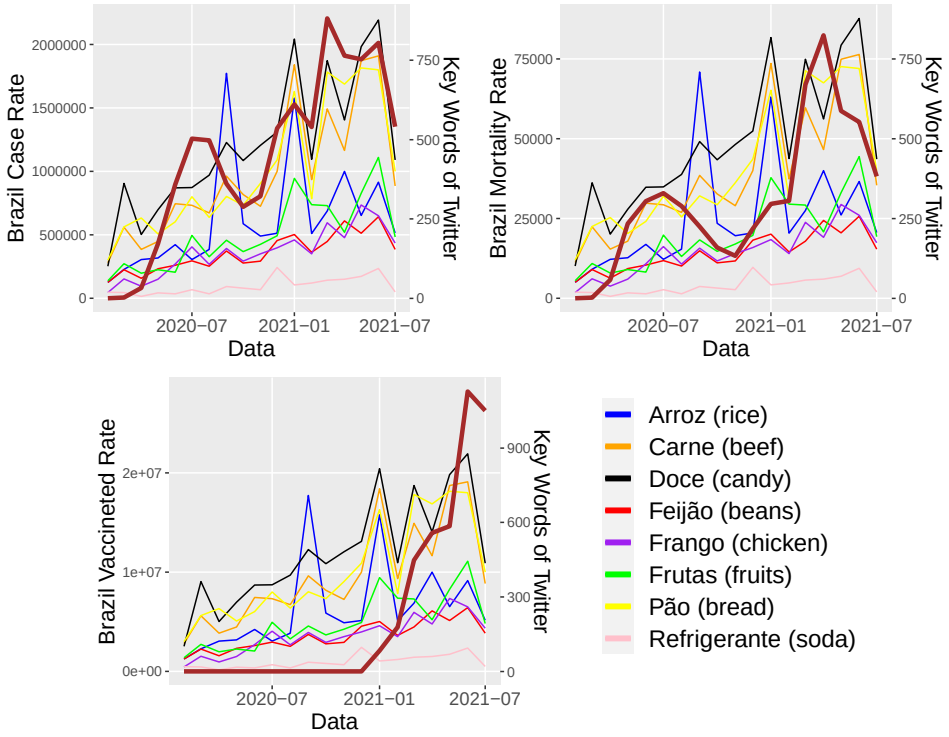


Figure 20 – Bi-plot between pandemic data in Brazil (cases, deaths, and vaccinated), with the food frequency on Twitter (Rice, Meat, candy, Beans, Chicken, Fruit, Bread, and Soft Drink).

Table 7 shows the correlations between food and pandemic data in Brazil. The significant correlations for the first wave were between the case rate and the foods: Meat (0.68), Beans (0.83), Chicken (0.83), Fruits (0.77), and bread (0.78). The other calculated correlations did not show significant results between the variables. The correlations between the terms and the vaccination rate appear as zero, as the vaccination period had not started.

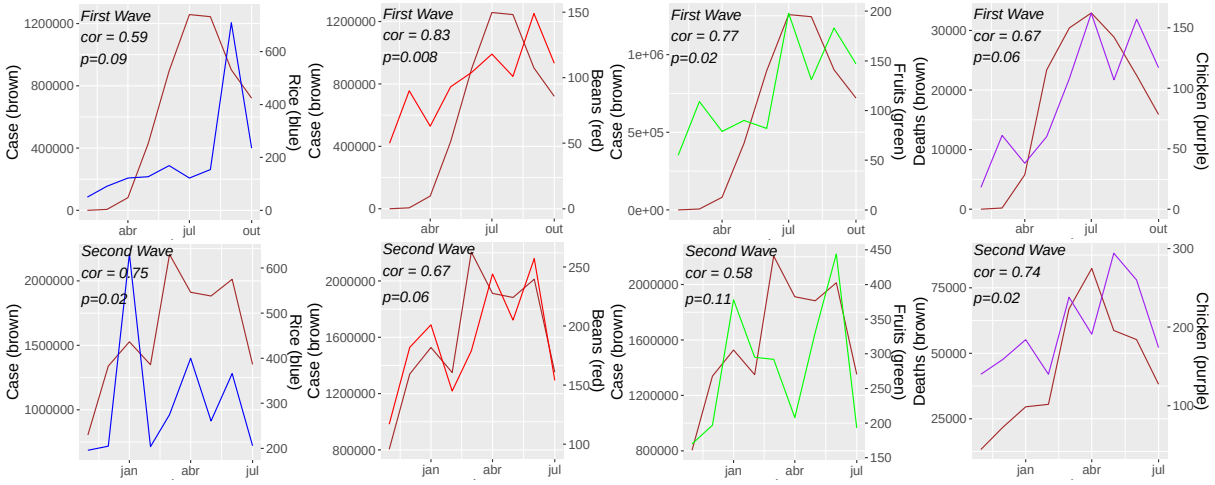
In the period of the second wave, the rate of cases had a significant correlation with foods: Rice (0.75), Meat (0.75), Chicken (0.85), and bread (0.80). The death rate showed a significant correlation between publications with the term "Chicken" in the second wave. The same happens with the vaccination rate, with a correlation of 0.68 with the term Chicken. So those publications with the food "Chicken" on Twitter obtained a significant correlation with all pandemic data in Brazil. The other correlations presented in Table 7 did not obtain significant results.

The changes between the significant correlations in the first and second waves of Table 7 are shown in Figure 21. In the upper part of Figure 21, one can observe the correlations for the period of the first wave and in the lower part for the second wave.

Again, the case rate shows a change in the correlations between the periods of the two waves, as shown in Figure (21a, b, and C). The Rice food on Twitter starts to

Table 7 – Correlation of pandemic data: Case, Deaths and Vaccineted, with food minerated of twitter: rice, beef, candy, beans, chicken, fruits, bread and soda.

First Wave								
	Rice	Beef	Candy	Beans	Chicken	Fruits	Bread	Soda
Case	0.59^{ns}	0.68^{*}	0.57^{ns}	0.83^{**}	0.83^{**}	0.77^{*}	0.78^{*}	0.18^{ns}
Death	0.45^{ns}	0.50^{ns}	0.25^{ns}	0.67^{ns}	0.67^{ns}	0.52^{ns}	0.46^{ns}	-0.08^{ns}
Vaccineted	0	0	0	0	0	0	0	0
Second Wave								
	Rice	Beef	Candy	Beans	Chicken	Fruits	Bread	Soda
Case	0.75^{*}	0.75^{*}	0.65^{ns}	0.67^{ns}	0.85^{**}	0.58^{ns}	0.80^{*}	0.32^{ns}
Death	0.57^{ns}	0.50^{ns}	0.32^{ns}	0.57^{ns}	0.74^{*}	0.32^{ns}	0.67^{ns}	0.27^{ns}
Vaccineted	0.40^{ns}	0.48^{ns}	0.26^{ns}	0.54^{ns}	0.68^{*}	0.40^{ns}	0.55^{ns}	0.11^{ns}
ns:p>0.05 *:p<0.05 **:p<0.01 ***:p<0.001								



(a) (b) (c) (d)

Figure 21 – Comparison between correlations that are no longer significant between the first and second wave of cases.

correlate with the rate of cases (0.75) in the second wave period (Figure 21a) significantly. Publications with the foods "Beans" and "Fruits" failed to obtain significant correlations (0.67) and (0.58) respectively for the second wave. The death rate was significantly correlated with the food "Chicken" in the second wave period, as shown in Figure (21d).

5.4 Discussion

This study proposed to verify the degree of relationship between the frequency of publications on Twitter and pandemic data in Brazil. Our results show some significant correlations between terms and are comparable to those found in previous studies. According to Sun and Gloor (2020), the ease of accessing the internet and social networks such as Twitter allows users to express their daily lives even more in this period of isolation.

With the analysis divided between the period of the first and second wave of COVID-19, it can be noted that there was a change between the significant correlations between the first and second wave of cases. The first wave of contagion was fear, depression, and anxiety about what was to come (WHO, 2020). This can be seen in the low publication on Twitter in the initial period of the pandemic with terms linked to it. Likewise for the other terms also present the frequency of stable publications. However, posts on pandemic terms get a two-wave formation when plotting the frequencies of posting pandemic terms on Twitter, as shown in Figure 16.

Posts on Twitter with the terms of the pandemic took shape from the beginning of the second wave of COVID-19, well marked in Figure 16. In this period of December 2020, emergency assistance in Brazil ended, which brought more fear and despaired the poorest population for not knowing how to generate income to survive, since the cases of covid, caused by the second wave, only increased every day, which made the governments tighten the measures of isolation and closing of the commerce again. This context of chaos is marked by the increase in publications about covid, vaccines, homes, and academies on Twitter.

Months later, in January 2021, the mass vaccination of Brazilians began. According to Santos et al. (2020), this period produced hope for the Brazilian people. Sandoval et al. (2018) reported that since the beginning of social networks, users have sought to express it or in the form of publication or sharing when they have a feeling. This can be seen in Figures 16 and 20, where all terms suffer an increase in the January 2021 publications that quickly lose strength, considered a moment of hope that stirred Twitter (MALAGOLI et al., 2021).

The rate of COVID-19 cases in Brazil was shown to be the most correlated with publications on Twitter in the second wave period, with the highest correlations being the term pandemic (0.95), Restaurants (0.80), and Chicken (0.85), along with the case rate. This implicitly shows that Brazilians are using Twitter to express the impacts and difficulties that the pandemic has brought to Brazil, such as the closing of stores and the increase in food prices. Likewise, it can be the contrast that the vaccination rate showed a significant correlation for all terms of the pandemic on Twitter, except for lockdown. This corroborates the results of Santos et al. (2020), which Brazilians posted on Twitter,

more about the impacts of the pandemic and the hope for medicines, mass immunization, and vaccines than about the reality of isolation already witnessed.

About the establishments published on Twitter, "delivery" represented by the places that made sales online stood out during the first wave, is correlated with the rate of cases and deaths, which presents the need and quick acceptance for this means of delivery, adopted by many people and businesses as the only way to buy and keep their business running (SOARES, SILVA, 2020).

As for food, publications with "chicken" on Twitter had a significant correlation with all pandemic data in Brazil, along with rice, beans, and bread in the second wave period. Food items, considered as basic food for Brazilians. According to Abbar, Mejova, and Weber (2015), people usually publish what they eat as a form of satisfaction and also out of desire. Thus, these correlations can demonstrate the items consumed by Brazilians or the desire to consume them.

Shin et al. (2016) confirmed data from this work, who found a correlation greater than 0.70 between Twitter posts about MERS and confirmed cases of this disease in the 2015 Korean outbreak, saying that social media data may reflect the actual epidemic of conditions before it is detected by health agencies (SHIN et al., 2016). The same author also reports that during an outbreak of infection, people who have symptoms can use Twitter to provide information on how to avoid them.

Previous research has shown us that Twitter has great potential to reflect real-life data. As was the case with obesity data (Abbar, Mejova, and Weber (2015)), MERS cases (Shin (2016)), and even COVID-19 (Sun and Gloor (2020)). Thus, this work corroborates these previous results that Twitter posts can reflect real-life data and show that these correlations can last long, thus obtaining the same format as actual data.

5.5 Conclusions

Thus, this study aimed to confirm the hypothesis that the frequency of publications on Twitter mentioning the pandemic and places and foods have significant correlations with actual data from the pandemic in Brazil. Confirming this hypothesis, it was also found that the visual distribution of tweet frequencies over time with the terms (covid, pandemic, and vaccine) has two regions, which resembles the distributions of cases and deaths by covid-19. In addition, it was found that important dates for Brazil generate an increase in the frequency of publications on Twitter, such as the end of emergency aid and the beginning of mass vaccination in Brazil. The case rate was the one with the highest significant correlation with the terms obtained on Twitter. And the terms "Pandemia", "Delivery," and "Chicken" were the ones that got the most significant correlation during the second wave of COVID-19 contagion in Brazil. It is concluded that Twitter posts can be correlated with real-life data, even when analyzed over long periods. For future work, it is suggested to check the correlation between Twitter data daily with data on cases,

deaths, and vaccinations.

5.6 References

ABBAR, S; MEJOVA, Y; WEBER, I. **You tweet what you eat: Studying food consumption through twitter.** In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems p. 3197-3206. 2015.

ARAMAKI, E; MASKAWA, S; MORITA, M. **Twitter catches the flu: detecting influenza epidemics using Twitter.** In: Proceedings of the 2011 Conference on empirical methods in natural language processing. p. 1568-1576, 2011.

BRASIL, **Ministério da Saúde. Banco de dados do Sistema Único de Saúde-DATASUS (CORONAVÍRUS-BRASIL).** Disponível em https://covid.saude.gov.br/ [Acessado em 05 de julho de 2021].

BRASIL. **Ministério da Saúde. Secretaria de Vigilância em Saúde. Boletim epidemiológico especial doença pelo coronavírus COVID-19 nº 38**. Semana Epidemiológica 44 (25 a 31/10) de 2020.
https://www.gov.br/saude/pt-br/media/pdf/2020/novembro/13/boletim_epidemiologico_covid_38_final_compressed.pdf/ Acesso em 23/09/2021.

BALTAZAR, J, Y et al. **Misinformation of COVID-19 on the internet: infodemiology study.** JMIR public health and surveillance, v. 6, n. 2, p. e18444, 2020.

BRZUSTEWICZ, P; SINGH, A. **Sustainable Consumption in Consumer Behavior in the Time of COVID-19: Topic Modeling on Twitter Data Using LDA.** Energies, v. 14, n. 18, p. 5787, 2021.

DURÃES, S. A. *et al.* **Implicações da pandemia da covid-19 nos hábitos alimentares.** Revista Unimontes Científica, v. 22, n. 2, p. 1-20, 2020.

DIGITAL in 2020. **We are Social**, New York 4 jan 2020.

FANELLI, R M. **Changes in the Food-Related Behaviour of Italian Consumers during the COVID-19 Pandemic.** Foods, v. 10, n. 1, p. 169, 2021.

FAO. Q and A: **COVID-19 pandemic – impact on food and agriculture**. 2020.

MALTA, D, C et al. **A pandemia da COVID-19 e as mudanças no estilo de vida dos brasileiros adultos: um estudo transversal, 2020.** Epidemiologia e Serviços de Saúde, v. 29, 2020.

MALAGOLI, L, G. et al. **A Look into COVID-19 Vaccination Debate on Twitter.** In: 13th ACM Web Science Conference 2021. 2021. p. 225-233.

MOSTAFA, M, M. **Mining and mapping halal food consumers: A geo-located Twitter opinion polarity analysis.** Journal of food products marketing, v. 24, n. 7, p. 858-879, 2018.

MOHAMMED, A; FERRARIS, A. **Factors influencing user participation in social media: evidence from twitter usage during COVID-19 pandemic in Saudi Arabia.** Technology in Society, v. 66, p. 101651, 2021.

PILAR, L; KVASNIČKOVÁ S, L; KVASNIČKA, R. **Healthy food on the twitter social network: Vegan, homemade, and organic food.** International Journal of Environmental Research and Public Health, v. 18, n. 7, p. 3815, 2021.

QUERCIA, D et al. **Our twitter profiles, our selves: Predicting personality with twitter.** In: 2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing. IEEE, 2011. p. 180-185.

SANTOS, A, G, et al. **Estudo quali-quantitativo sobre a percepção de usuários do Twitter sobre a adoção das medidas de quarentena, de distanciamento e de isolamento sociais, durante a pandemia da COVID-19.** BIS. Boletim do Instituto de Saúde, v. 21, n. 1, p. 173-185, 2020.

SANDOVAL, L, G. et al. **Spanish Twitter data used as a source of information about consumer food choice. In: International Cross-Domain Conference for Machine Learning and Knowledge Extraction.** Springer, Cham, p. 134-146, 2018.

SOARES, A, C, N; SILVA L, M, R. **Serviços de delivery alimentício e suas precauções em tempos da pandemia de SARS-COV-2 (Covid-19).** Brazilian Journal of Health Review, v. 3, n. 3, p. 4217-4226, 2020.

SHIN, S, Y et al. **High correlation of Middle East respiratory syndrome spread with Google search and Twitter trends in Korea.** Scientific reports, v. 6, n. 1, p.

1-7, 2016.

SUN, J; GLOOR, P. **More active internet-search on Google and Twitter posting for COVID-19 corresponds with lower infection rate in the 50 US states.** 2020.

VIDAL, L et al. **Using Twitter data for food-related consumer research: A case study on "what people say when tweeting about different eating situations".** Food Quality and Preference, v. 45, p. 58-69, 2015.

WIDENER, M, J.; LI, W. **Using geolocated Twitter data to monitor the prevalence of healthy and unhealthy food references across the US.** Applied Geography, v. 54, p. 189-197, 2014.

WORLD HEALTH ORGANIZATION (WHO) et al. **Coronavirus disease 2019 (COVID-19): situation report**, 94. 2021.

ZHOU, J.; LIU, F.; ZHOU, H. **Understanding health food messages on Twitter for health literacy promotion.** Perspectives in public health, v. 138, n. 3, p. 173-179, 2018.

6 CONCLUSÃO GERAL

O objetivo desta dissertação foi verificar o comportamento do consumidor Brasileiro de alimentos durante a pandemia de COVID-19, realizando dois artigos onde foi verificado a mudança nas prioridades alimentares dos Brasileiros obtendo como resultados as Pesquisas por Crise + alimentos no Twitter o que nos mostraram que os consumidores estão ansiosos e fazem uso de café neste período. O isolamento fez também com que os consumidores em sua grande maioria passassem a adotar pedidos por delivery, reduzindo a frequência de compra de duas vezes por semana para semanal. Os alimentos mais comprados pelos participantes foram frutas e arroz e os menos comprados Carne e Refrigerante. A procura por uma dieta saudável no período de isolamento se mostrou presente nas justificativas dadas pelos participantes. Contudo, doces, chocolates e sanduíches foram associados a ajuda com o estresse, os mesmos alimentos aparecem em tweets pesquisados por quarentena + alimentos. Café e queijo aparecem associados com "isto me deixa feliz". Arroz, Leite, Farinha, Feijão e biscoitos estão associados com "tenho medo que acabe".

O objetivo do segundo artigo foi de confirmar a hipótese de que a frequência de publicações no Twitter com menções a pandemia são correlacionados aos dados provenientes do ministério da saúde. Esta hipótese foi confirmada encontrando correlações significativas entre o período da primeira e segunda onda de contágio e mortes da COVID-19. Também foi descoberto que a distribuição visual das frequências dos tweets ao longo do tempo com os termos (covid, pandemia e vacina) possuem duas regiões de topos, o que se assemelha às distribuições de casos e mortes por covid-19. Além disso, foi verificado que datas importantes para o Brasil geram aumento na frequência de publicações no Twitter, como o dia do fim do auxílio emergencial e o início da vacinação em massa no Brasil.

Deste modo, conclui-se o objetivo de verificar as mudanças comportamentais dos brasileiros durante o período da Pandemia, de forma que o maior impacto para os brasileiros se deu com relação às condições financeiras, e a utilização do Twitter como forma de socialização virtual, com as publicações refletindo a realidade dos usuários. Para trabalhos futuros sugere-se verificar o nível de correlação entre dados do Twitter de forma diária com os dados de casos, mortes e vacinados. E identificar o impacto da crise alimentar por classes sociais.

REFERÊNCIAS

ABDI, H; VALENTIN, D. Multiple correspondence analysis. **Encyclopedia of Measurement and Statistics**, v. 2, n. 4, p. 651-657, 2007.

AMALIA, P; IONUT, P. Consumers' reaction and organizational response in crisis context. **The Journal of the Faculty of Economics**, v. 1, n. 5, p. 779-782, 2009.

ANDRADE, D. M; OLIVEIRA, J. L. R; ANTONIALLI, L. M. O perfil de clientes de um shopping center: um estudo exploratório com consumidores do interior. **Organizações Rurais Agroindustriais**, v. 6, n. 2, p. 91105, 2004.

BARROS, G. S. C. *et al.* A inflação dos alimentos em 2020 e seus gatilhos. **Centro de Estudos Avançados em Economia Aplicada.** v. 1, n. 2, p. 106, 2021.

BARTLETT, M. S. Contingency table interactions. **Journal of the Royal Statistical Society** v. 2, n. 2, p. 248-252, 1935.

BLACKWELL, R. D; MINIARD, P. W; ENGEL, J. F. **Comportamento do consumidor**. 9. ed. São Paulo: Pioneira Thomson Learning, 2005.

COSTA, B. R. L. Bola de neve virtual: o uso das redes sociais virtuais no processo de coleta de dados de uma pesquisa científica. **Revista Interdisciplinar de Gestão Social**, v. 7, n. 1, p. 25-27, 2018.

CHOY, M. Effective listings of function stop words for Twitter. **ArXiv Preprint** v. 1205, n. 1, p. 6396, 2012.

CUI, W. *et al.* Context preserving dynamic word cloud visualization. **IEEE Pacific Visualization Symposium**, v. 3, n. 1, p. 121-128, 2010.

DALMORO, M. *et al.* Twitter: Uma Análise do Consumo, Interação e Compartilhamento na Web 2.0. **XXXIV Encontro da Anpad**, v. 15, n. 1, p. 1-17, 2010.

DANIEL, W. W. *et al.* **Applied nonparametric statistics**. 2.ed, Houghton Mifflin: Chicago, 1978.

DURÃES, S. A. *et al.* Implicações da pandemia da covid-19 nos hábitos alimentares. **Revista Unimontes Científica**, v. 22, n. 2, p. 1-20, 2020.

DA CUNHA, S. B; CARVAJAL, S. R. **Estatística básica-a arte de trabalhar com dados**. 1. ed, Elsevier Brasil: São Paulo, 2009.

DIGITAL in 2020. **We are Social**, New York 4 jan 2020.
Disponível em: https://wearesocial.com/digital-2020 Acesso em: 05/08/2021.

ENGEL, J. F; KOLLAT, D. T; MINIARD, P. W. Consumer behavior. 6. ed. New York: Dryden Press, 1990.

FERREIRA, P. A; REZENDE, D. C; LOURENÇO, C. D. S. Geração canguru: algumas tendências que orientam o consumo jovem e modificam o ciclo de vida familiar. **Revista Espacios**, v. 32, n. 1, p. 12-14, 2011.

FEINERER, I; HORNIK, K. tm: Text Mining Package. **R package version 0.7-8**, https://CRAN.R-project.org/package=tm.

FELLOWS, I. *et al.* **Package 'wordcloud'. R Package, Maintainer Ian and Rcpp,** Linking To and Rcpp.[(accessed on 4 February 2021)], 2018.

GROHMANN, M. Z. *et al.* Aceitação e adoção de produtos com novas tecnologias: o gênero como fator moderador. **Revista de Administração e Inovação**, v. 7, n. 4, p. 137-161, 2010
.
GILKEY, J. R; JOSEPH, W; CLARK, S. D. The 2008 Global Financial Crisis Post-Recession Impact on Consumer Behavior Based on Educational Level. **International Journal of Innovation, Management and Technology**, v. 6, n. 6, p. 363-366, 2015.

HAVLICEK, L. L; PETERSON, N. L. Robustness of the Pearson correlation against violations of assumptions. **Perceptual and Motor Skills**, v. 43, n. 3, p. 1319-1334, 1976.

HAWKINS, D. I; MOTHERSBAUGH, D. L; BEST, R. J. **Comportamento do consumidor: construindo a estratégia de marketing**. 10. ed. Rio de Janeiro: Elsevier, 2007.

HEIMERL, F. *et al.* Word cloud explorer: Text analytics based on word clouds. **47th Hawaii International Conference on System Sciences. IEEE**, v. 1, n. 1, p. 1833-1842, 2014.

HINES, W. W; MONTGOMERY, D. C; GOLDSMAN, D. M. **Probabilidade E Estatística Na Engenharia**. 1. ed. Grupo Gen-LTC, São Paulo, 2000.

INFANTOSI, A. F. C; COSTA, J. C. G. D; ALMEIDA, R. M. V. R. Análise de Correspondência: bases teóricas na interpretação de dados categóricos em Ciências da Saúde. **Cadernos de Saúde Pública**, v. 30, n. 1, p. 473-486, 2014.

JACOBY, J; JOHAR, G. V; MORRIN, M. Consumer behavior: A quadrennium. **Annual Review of Psychology**, v. 49, n.1 p. 319-344, 1998.

KULLAR, R. *et al.* To tweet or not to tweet—a review of the viral power of Twitter for infectious diseases. **Current Infectious Disease Reports**, v. 22, n. 6, p. 1-6, 2020.

KEARNEY, M. W; **Package 'rtweet'**, 2016.
Disponível em: https://cran. r-project. org/web/packages/rtweet/rtweet. pdf [accessed 2019-03-19],

LOHMANN, S. *et al.* Concentri cloud: Word cloud visualization for multiple text documents. **19th International Conference on Information Visualisation. IEEE**, v. 25, n. 5, p. 114-120, 2015.

LACHENBRUCH, P. A. **McNemar test.** 3. ed. Boston: Statistics Reference Online, EUA, 2014.

MOWEN, J. C; MINOR, M. S. **Comportamento do consumidor.** 3. ed. São Paulo: Pearson Prentice Hall, 2003.

MANSOOR, D; JALAL, A. The global business crisis and consumer behavior: Kingdom of Bahrain as a case study. **International Journal of Business and Management**, v. 6, n. 1, p. 104, 2011.

MINOT, J, R *et al.* Ratioing the President: An exploration of public engagement with Obama and Trump on Twitter. **Arxiv**, v. 1, n. 352, p. 1-17, 2020.

MOORE, D. S. **The Basic Practice of Statistics.** 1.ed, New York, Freeman, 2007.

MCNEMAR, Q. Correction to a correction. **Journal of Consulting and Clinical Psychology**, v. 42, n. 1, p. 145, 1974.

MANTEL, N; FLEISS, J. L. The Equivalence of the Generalized McNemar Tests for Marginal Homogeneity in 2 $\hat{3}$ and 3 $\hat{2}$ Tables. **Biometrics** v. 1, n. 1, p. 727-729, 1975.

MCHUGH, M. L. The chi-square test of independence. **Biochemia medica**, v. 23, n. 2, p. 143-149, 2013.

OLMSTEAD, M. S. **The small group.** 3. ed. New York: Holt, Rinehart and Winston, 1962.

OSMAN, M **Estatísticas e Fatos do Twitter Sobre a Nossa Rede Favorita (2021)**, Kisnta Jan 2021 https://kinsta.com/pt/blog/estatisticas-e-fatos-do-twitter/

POBIRUCHIN, M; ZOWALLA, R; WIESNER, M. Temporal and Location Variations, and Link Categories for the Dissemination of COVID-19–Related Information on Twitter During the SARS-CoV-2 Outbreak in Europe: Infoveillance Study. **Journal of medical Internet research**, v. 22, n. 8, p. 19629, 2020.

PUSHPAM, C, A; JAYANTHI, J, G. Overview on *data mining in Social Media.* **International Journal of Computer Sciences and Engineering**, v. 5, n. 11, p.147-157, 2017.

RAVINDRAN, S, K; GARG, V. **Mastering Social Media Mining with R, Packt Publishing** Ltd, 2015.

ROBERTSON, A. M; WILLETT, P; **Applications of n-grams in textual information systems.** 2. ed. Journal of Documentation, London, 1998.

SARICA, S; LUO, J. Stopwords in technical language processing. **Plos one**, v. 16, n. 8, p. 4937, 2021.

SOLOMON, M. R. *et al.* **Consumer behavior: Buying, having, and being.** 1. ed. Boston, MA: Pearson, 2017.

SHETH, J. N; MITTAL, B; NEWMAN, B. I. **Comportamento do consumidor: indo além do comportamento do consumidor.** 1. ed, São Paulo: Atlas, 2001.

SIMA, V. *et al.* Influences of the industry 4.0 revolution on the human capital development and consumer behavior: A systematic review, **Sustainability**, v. 12, n. 10, p. 4035, 2020.

SUEN, C. Y. N-gram statistics for natural language understanding and text processing. **IEEE transactions on pattern analysis and machine intelligence**, v. 5, n. 2, p. 164-172, 1979.

SIDOROV, G *et al.* Syntactic n-grams as machine learning features for natural language processing. **Expert Systems with Applications**, v. 41, n. 3, p. 853-860, 2014.

SIQUEIRA, E. **Tecnologias que mudam nossa vida.** 2ed. São Paulo: Saraiva, 2008.

STAUDT, J. **Text Mining utilizando o Software R: um estudo de caso de uma biblioteca americana**, Monografia (TCC para Bacharel em Estatística) – Universidade Federal do Rio Grande do Sul, Rio Grande do Sul, 2016.

SPEARMAN, C. The proof and measurement of association between two things. **American Journal of Psychol** v. 15, n. 5, p. 72–101, 1904.

TAURION, C. ***big data.*** 1. ed. Rio de Janeiro: Brasport Livros e Multimídia Ltda., 2013.

UNDERHILL, P. **Vamos às compras: a ciência do consumo.** 1. ed. UFRJ, Rio de Janeiro: Campus, 2009.

WARD, J. C; REINGEN, P. H. Sociocognitive analysis of group decision making among consumers. **Journal of Consumer Research**, v. 17, n. 1, p. 245-262, 1990.

ZAFARINI, R; ABBASI, M, A; LIU, H. **Social Media Mining an introduction** 1. ed. Cambridge University Press, New York, USA, 2014.

Anexo A- TCLE

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Você está sendo convidado(a) a participar, como voluntário(a), da pesquisa COMPORTAMENTO DO CONSUMIDOR DE ALIMENTOS DURANTE A PANDEMIA DA COVID-19, no caso de você concordar em participar, favor assinar ao final do documento.

Sua participação não é obrigatória, e, a qualquer momento, você poderá desistir de participar e retirar seu consentimento. Sua recusa não trará nenhum prejuízo em sua relação com o pesquisador(a) ou com a instituição.

Você receberá uma cópia deste termo onde consta o telefone e endereço do pesquisador principal, podendo tirar dúvidas do projeto e de sua participação.

TÍTULO DA PESQUISA: COMPORTAMENTO DO CONSUMIDOR DE ALIMENTOS DURANTE A PANDEMIA DA COVID-19
RESPONSÁVEL: Prof. Dr. Eric Batista Ferreira
PESQUISADOR PARTICIPANTE: Gabriel Baldasso
ENDEREÇO: Rua Gabriel Monteiro da Silva, 700, Unifal -MG, sala D309B.
TELEFONE: (35) 3701-9604

OBJETIVOS: Realizar uma coleta de dados através de método simples e rápido, com a finalidade de se obter a frequência de uso de restaurantes, bares e delivery antes e depois da pandemia, assim como verificar se houve mudanças nos hábitos alimentares e nos locais de compra de alimentos dos consumidores.
JUSTIFICATIVA: No contexto atual em que vivemos, a pandemia da COVID-19, obrigou a passarmos maior tempo em casa o que acarretou no maior consumo de alimentos. Logo, é importante detectar se houve mudança no hábito alimentar do consumidor durante este período, assim como verificar se houve alguma mudança com relação ao local de compra de alimentos.
PROCEDIMENTOS DO ESTUDO:
1) A pesquisa será conduzida via Google forms e será compartilhada via redes sociais.
2) Esses dados serão utilizados para analisar o comportamento alimentar dos brasileiros frente à pandemia.
RISCOS E DESCONFORTOS: Não existem riscos físicos, visto que se trata de uma pesquisa online. Risco de constrangimento: esse risco será minimizado reforçando-se o anonimato da pesquisa. O participante poderá interromper a pesquisa a qualquer momento, além de ser garantido o direito de desistir de participar da mesma.
BENEFÍCIOS: Os dados coletados nesta pesquisa gerarão informações importantes a respeito do hábito alimentar do brasileiro durante a pandemia.
CUSTO/REEMBOLSO PARA O PARTICIPANTE: Não haverá nenhum tipo de pagamento/remuneração pela participação na pesquisa, podendo desistir de participar a qualquer momento sem nenhum prejuízo.
CONFIDENCIALIDADE DA PESQUISA: O nome do participante será mantido em sigilo, assegurando sua privacidade.

Assinatura do Pesquisador Responsável: _____

Eu, , declaro que li as informações contidas nesse documento, fui devidamente informado pelo pesquisador Eric Batista Ferreira dos procedimentos que serão utilizados, riscos e desconfortos, benefícios, custo/reembolso dos participantes, confidencialidade da pesquisa, concordando ainda em participar da pesquisa.

Foi-me garantido que posso retirar o consentimento a qualquer momento, sem qualquer penalidade ou interrupção de meu acompanhamento/assistência/tratamento. Declaro ainda que recebi uma cópia desse Termo de Consentimento.

Poderei consultar o pesquisador responsável (acima identificado) ou o CEP UNIFALMG, com endereço na Universidade Federal de Alfenas, Rua Gabriel Monteiro da Silva, 700, Centro, Cep - 37130-000, Fone: (35) 3701-9604, horário de atendimento de 07:30 às 11:30 e das 13:30 às 16:30, no e-mail: comite.etica@unifal-mg.edu.br sempre que entender necessário obter informações ou esclarecimentos sobre o projeto de pesquisa e minha participação no mesmo.

Os resultados obtidos durante este estudo serão mantidos em sigilo, mas concordo que sejam divulgados em publicações científicas, desde que meus dados pessoais não sejam mencionados.

Alfenas, MG_____ de _____ de _____

. _____ _____

(Nome por extenso) (Assinatura)

NOME E ASSINATURA DO SUJEITO

Anexo B Questionário

COMPORTAMENTO DO CONSUMIDOR DE ALIMENTOS DURANTE A PANDEMIA DE COVID-19

Este questionário tem o objetivo de avaliar a intenção de compra de alimentos ANTES E DEPOIS da pandemia da Covid-19. Assim como verificar a frequência de uso de locais de comidas prontas.

*Obrigatório

Termo de Consentimento Livre e Esclarecido

Termo de Consentimento Livre e Esclarecido:
https://drive.google.com/file/d/1liunvibAF8ULeWD6vS9zNQboLR6uOnxC/view?usp=sharing

- ◯ Concordo e quero participar.
- ◯ Discordo e não quero participar.

1) Cidade onde mora atualmente? *

Sua resposta

2) Qual a sua idade? *

Sua resposta

3) Sexo *

- ◯ M
- ◯ F
- ◯ Prefiro não declarar

4) Antes do começo da pandemia e as medidas de isolamento, qual era seu local de compra de alimentos? *

- ◯ Supermercado
- ◯ Pequenas lojas
- ◯ On-line
- ◯ Outro:

5) Qual era a frequência de compras, nos locais dito acima, antes da pandemia? *

- ◯ Mensalmente
- ◯ Semanalmente
- ◯ Duas vezes por semana
- ◯ Diária
- ◯ Outro:

6) Agora com a pandemia e as medidas de isolamento, qual é seu local de compra de alimentos? *

- ◯ Supermercado
- ◯ Pequenas lojas
- ◯ On-line
- ◯ Outro: _____

7) Qual a frequência de compras nos locais dito acima, agora com a pandemia? *

- ◯ Mensalmente
- ◯ Semanalmente
- ◯ Duas vezes por semana
- ◯ Diária
- ◯ Outro: _____

8) Antes das medidas de isolamento para conter a COVID-19, qual era sua forma de consumir alimentos prontos? *

- ◯ Restaurantes
- ◯ Bares
- ◯ Delivery
- ◯ Fast-Food
- ◯ Outro: _____

9) Qual era a frequência de procura por estes locais antes da pandemia? *

- ◯ Mensalmente
- ◯ Semanalmente
- ◯ Duas vezes por semana
- ◯ Diária
- ◯ Outro: _____

10) Agora com as medidas de bloqueio qual é sua forma de consumir alimentos prontos? *

- ◯ Restaurantes
- ◯ Bares
- ◯ Delivery
- ◯ Fast-Food
- ◯ Outro: _____

11) Qual a frequência por este local citado acima agora com a pandemia? *

- ◯ Mensalmente
- ◯ Semanalmente
- ◯ Duas vezes por semana
- ◯ Diária
- ◯ Outro: _____

12) Quais os 3 alimentos (sólido ou liquido) você passou a comprar mais durante a pandemia e as medidas de isolamento? *

Sua resposta

13) Escolha as opções do porque você passou a comprar mais os 3 alimentos dito acima. Use a primeira coluna para o primeiro alimento citado e assim por diante.

	Coluna 1	Coluna 2	Coluna 3
Tenho medo de que este produto acabe	☐	☐	☐
Acabou todo o meu estoque	☐	☐	☐
Eu não posso comprar todos os dias	☐	☐	☐
Tenho vontade de comer diariamente	☐	☐	☐
Eu gosto bastante	☐	☐	☐
Isso me deixa feliz	☐	☐	☐
Isso me ajuda com o Stress	☐	☐	☐
Esta barato	☐	☐	☐
Me protege contra o coronavírus	☐	☐	☐
É saudável, ajuda a manter o peso	☐	☐	☐
Todos estão comprando	☐	☐	☐
A data de validade é longa	☐	☐	☐

14) Quais os 3 alimentos (sólido ou liquido) você passou a comprar menos durante a pandemia e as medidas de isolamento? *

Sua resposta

15) Escolha as opções do porque você passou a comprar menos estes 3 alimentos dito acima. Use a primeira linha para o primeiro alimento citado e assim por diante.

	Coluna 1	Coluna 2	Coluna 3
Não tenho mais condições financeiras	☐	☐	☐
Preciso economizar	☐	☐	☐
Eu não posso mais ir onde eu comprava	☐	☐	☐
Não faço mais questão	☐	☐	☐
Enjoei deste produto	☐	☐	☐
Não me anima quando estou triste	☐	☐	☐
Esta caro	☐	☐	☐
Especialistas disseram que não é hora de comprar	☐	☐	☐
Não é saudavel	☐	☐	☐
Ninguém mais compra	☐	☐	☐
Tem vida útil curta	☐	☐	☐
Não me ajuda a proteger do COVID-19	☐	☐	☐

16) A compra no supermercado ficou mais cara depois das medidas de isolamento? *

- ◯ SIM
- ◯ NÃO
- ◯ Outro: _____

17) Estime o quanto a compra no supermercado ficou mais cara. *

- ◯ 10%
- ◯ 20%
- ◯ 50%
- ◯ Outro: _____

18) Qual é a sua confiança nas declaração da Organização Mundial da Saúde no combate a COVID-19? *

	1	2	3	4	5	
Não Confio	◯	◯	◯	◯	◯	Confio Muito

19) Qual é a sua confiança nas declaração do Presidente Jair Messias Bolsonaro, no combate a COVID-19? *

	1	2	3	4	5	
Não Confio	◯	◯	◯	◯	◯	Confio Muito

20) Qual a fonte de informação mais confiável para obter informações a respeito da pandemia na sua opinião?

- ◯ TV
- ◯ Jornal Impresso
- ◯ Twitter
- ◯ Facebook
- ◯ Instagram
- ◯ Whatsapp
- ◯ Telefone
- ◯ Boca-a-Boca
- ◯ Sites de noticias
- ◯ Outro: _____

Desabafos, criticas e opiniões a respeito do consumo de alimentos durante a pandemia. esta é a hora!

Sua resposta