UNIVERSIDADE FEDERAL DO PIAUÍ

CENTRO DE TECNOLOGIA

CURSO DE GRADUAÇÃO EM ENGENHARIA ELÉTRICA

FRANCISCO CLEBER DA CONCEIÇÃO FEITOSA

ANÁLISE EXPLORATÓRIA DE DADOS EM E-COMMERCE: UM ESTUDO NO BRASIL PRÉ-PANDEMIA DE COVID 19

TERESINA

2023

FRANCISCO CLEBER DA CONCEIÇÃO FEITOSA

ANÁLISE EXPLORATÓRIA DE DADOS EM E-COMMERCE: UM ESTUDO NO BRASIL PRÉ-PANDEMIA DE COVID 19

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Engenharia Elétrica do Centro de Tecnologia da Universidade Federal do Piauí, como requisito parcial à obtenção do grau de bacharel em Engenharia Elétrica.

Orientador: Prof. Dr. José Maria Pires de Menezes Júnior

TERESINA

2023

FICHA CATALOGRÁFICA
Universidade Federal do Piauí
Biblioteca Comunitária Jornalista Carlos Castello Branco
Divisão de Representação da Informação

F311a Feitosa, Francisco Cleber da Conceição.
Análise exploratória de dados em *e-commerce* : um estudo no Brasil pré-pandemia de Covid 19 / Francisco Cleber da Conceição Feitosa. -- 2023.
76 f.

Trabalho de Conclusão de Curso (Graduação) – Universidade Federal do Piauí, Centro de Tecnologia, Bacharelado em Engenharia Elétrica, Teresina, 2023.
"Orientador: Prof. Dr. José Maria Pires de Menezes Júnior".

1. Comércio digital. 2. Plataformas digitai. 3. Análise de dados. 4. Demografia Brasileira. 5. Economia. I. Menezes Júnior, José Maria Pires de. II. Título.

CDD 621.39

Bibliotecária: Francisca das Chagas Dias Leite – CRB3/1004

FRANCISCO CLEBER DA CONCEIÇÃO FEITOSA

ANÁLISE EXPLORATÓRIA DE DADOS EM E-COMMERCE: UM ESTUDO NO BRASIL PRÉ-PANDEMIA DE COVID 19

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Engenharia Elétrica do Centro de Tecnologia da Universidade Federal do Piauí, como requisito parcial à obtenção do grau de bacharel em Engenharia Elétrica.

Aprovada em: 24 de Agosto de 2023

BANCA EXAMINADORA

Prof. Dr. José Maria Pires de Menezes Júnior (Orientador)
Universidade Federal do Piauí (UFPI)

Prof. Dr. Geordy Souza Pereira
Universidade Federal do Piauí (UFPI)

Prof. Me. Robson Gonçalves Fechine Feitosa
Instituto Federal do Ceará (IFCE)

À minha mãe, por ter dedicado sua vida a mim e meus irmãos, mesmo com todas as dificuldades do mundo sempre acreditou que dias melhores viriam. Aos meus professores de toda uma vida, que me adotaram e me mostraram o mundo em ensinamentos e virtudes. E com todo o meu coração, aos meus amigos que me salvaram em tantas lutas e de mãos dadas não deixaram que eu me sentisse sozinho nas vitórias e derrotas. Sou o resultado da exposição a muitas almas brilhantes e levarei comigo sempre a esperança de um mundo melhor, por conta de vocês.

AGRADECIMENTOS

Agradeço à minha mãe, Elizabete por cada lágrima derramada e cada abraço de conforto, cada dia de trabalho sob o sol que me deu a segurança de ter um lugar pra onde voltar se nada desse certo. Deu certo, mãe. Agradeço a minha Tia Gilza, por todos os sorrisos e afagos de um ser de luz com coração sem igual. E a minhã irmã Dalva, que compartilhou todas as felicidades comigo e vejo com muito orgulho trilhando os passos para conquistas bem maiores que as minhas. A felicidade da minha família foi sempre meu maior objetivo, e meu amor por vocês vai me acompanhar por toda a vida.

Aos meus professores, que me permitiram chegar à universidade, me dando moradia e segurança quando saí do ensino médio sem esperança de ser capaz de me tornar um engenheiro: Michele, Daniele, Givaldo, Ailton. Essa conquista é nossa. Vocês foram responsáveis por tantos sonhos se tornarem realidade, e prometo manter acesa a luz adiante.

Aos meus amigos, que se tornaram minha família e confidentes em tantas empreitadas. Meus amigos de turma, calouros, veteranos, agregados, residentes, nos corredores e banquinhos, eu devo minha vida aos sorrisos de vocês.

Agradeço especialmente ao PET e a RiseUP, indispensáveis na minha construção como pessoa e profissional. Aprendi com vocês tanto, e as portas abertas por conviver com vocês são imensuráveis. Obrigado Professor Otacílio, pelo voto de confiança e por me receber nessa família, obrigado Professor Marcos Lira, por ser um exemplo de ser humano e um grande amigo, obrigado aos meus irmãos e irmãs petianas e empreendedores juniores, que são pra sempre parte de mim. Amo muito todos vocês.

E agradeço à Universidade Federal do Piauí e todos os professores que compartilharam o conhecimento e a paixão pela boa ciência e humanidade que esse mundo tanto precisa. Um obrigado especial ao meu orientador Professor José Maria Pires Menezes Júnior, que é não só um professor exemplar como também um orientador paciente e um amigo. Obrigado por aceitar conduzir o presente trabalho.

”Todos nós mudamos, quando você para pra pensar, somos todos várias pessoas diferentes, por toda a nossa vida; e tudo bem, isso é bom, você tem que seguir em frente, contanto que você se lembre de todas as pessoas que você costumava ser. Não vou esquecer uma linha disso, nem um dia, juro. Sempre me lembrarei de quando esse era eu.”

(Doctor Who)

RESUMO

O comércio digital é um dos grandes agentes econômicos da revolução tecnológica. Através dos séculos, o ser humano travou guerras e desbravou distâncias continentais para estabelecer relações comerciais. Na atualidade, a proximidade proporcionada pelo acesso global à internet, possibilita não só que sejam realizadas transações comerciais a longas distâncias em tempo quase imediato, como também permitiram o desenvolvimento de uma rede complexa de instrumentos para operação e gestão das variadas formas de *ecommerce*, criando um ambiente fértil para atuação de engenheiros em um mercado competitivo. O presente trabalho pretende explorar, sob a ótica da análise de dados, com perspectiva exploratória e em um contexto geográfico, um banco de dados disponibilizado pela plataforma Olist, contendo informações sobre comércio digital durante um período pré covid 19 no Brasil. A metodologia utilizada constitui uma contextualização do comércio digital através de revisão bibliográfica, seguida de uma análise demográfica por meio de documentos oficiais e, de posse disso, uma abrangente análise exploratória através da linguagem python. A relevância da pandemia é contextualizada na perspectiva de verificar como estava o cenário nacional e fornecer material para estudos futuros, contrastando com o efeito *boom* proporcionado em função das políticas de isolamento durante a situação pandêmica mundial. O trabalho apresenta ainda dados econômicos oficiais e demográficos brasileiros na intenção de verificar se a amostra é condizente com as informações oficiais e qual comportamento observado nos anos posteriores. As análises revelaram uma concentração de pedidos na região Sudeste brasileira, sobretudo no estado de São Paulo; enquanto o Nordeste e Norte Brasileiros se destacaram em valores de frete e tempos de entrega mais elevados; foi possível perceber uma correlação entre a distribuição das cidades no território, sua população e distribuição de renda, com a concentração de indicadores positivos do comércio digital.

Palavras-chave: Análise de dados. Comércio digital. Demografia Brasileira. Economia. Plataformas digitais.

ABSTRACT

Digital commerce is one of the great economic agents of the technological revolution. Through the centuries, human beings waged wars and explored continental distances to establish commercial relationships. At present, the proximity provided by global access to the Internet makes it possible not only for commercial transactions to be carried out over long distances in almost immediate time, but also allowed the development of a complex network of instruments for the operation and management of the various forms of *ecommerce*, creating a fertile environment for engineers to work in a competitive market. The present work intends to explore, from the perspective of data analysis, with an exploratory perspective and in a geographic context, a database made available by the Olist platform, containing information on digital commerce during a pre-covid 19 period in Brazil. The methodology used constitutes a contextualization of digital commerce through a bibliographic review, followed by a demographic analysis through official documents and in possession of this, a comprehensive exploratory analysis using the python language. The relevance of the pandemic is contextualized from the perspective of verifying how the national scenario was and providing material for future studies, contrasting with the *boom* effect provided by the isolation policies during the world pandemic situation. The work also presents official Brazilian economic and demographic data with the intention of verifying whether the sample is consistent with official information and what behavior was observed in later years. The analyzes revealed a concentration of requests in the Brazilian Southeast region, especially in the state of São Paulo; while the Northeast and North of Brazil stood out in higher freight values and delivery time; it was possible to perceive a correlation between the distribution of cities in the territory, their population and income distribution, with the concentration of positive indicators of digital commerce.

Keywords: Data analysis. Digital Commerce. Brazilian demography. Economy. Digital platforms.

LISTA DE ILUSTRAÇÕES

Figura 1 – RNP em 1992 30
Figura 2 – Percentual de moradores sem acesso à internet nas grandes regiões brasileiras 31
Figura 3 – Rendimento médio real domiciliar per capita a preços médios do ano nas grandes regiões brasileiras - P95 32
Figura 4 – Regiões de influência das cidades brasileiras em 2018 32
Figura 5 – Densidade de domínios de Internet no Brasil em 2019 33
Figura 6 – Integração territorial pelo correio 1982-2016 34
Figura 7 – Fluxograma da metodologia aplicada 35
Figura 8 – Estrutura de tabelas do Banco de Dados Olist 37
Figura 9 – Schema do Banco de Dados Olist 43
Figura 10 – Quantidade de itens comprados por categoria de produtos 47
Figura 11 – Receita de itens comprados por categoria de produtos 48
Figura 12 – Nota dada através de reviews por categoria de produtos 49
Figura 13 – Distibuição de pedidos por status 50
Figura 14 – Status por ano de compra realizada 51
Figura 15 – Identificação de cada dígito do CEP 53
Figura 16 – Distribuição de CEPs no território brasileiro 54
Figura 17 – Distribuição de CEPs na cidade de Granjeiro - CE 54
Figura 18 – Distribuição de CEPs na cidade de Teresina - PI 55
Figura 19 – Distribuição de CEPs na cidade de São Paulo - SP 55
Figura 20 – Distribuição de pedidos por estados do Brasil em ranking 58
Figura 21 – Distribuição de pedidos por estados do Brasil em mapa 59
Figura 22 – Quantidade de pedidos por consumidor por estados do Brasil em ranking . . 60
Figura 23 – Quantidade de pedidos por consumidor por estados do Brasil em mapa . . . 60
Figura 24 – Distribuição de receita por estados do Brasil em ranking 61
Figura 25 – Distribuição de receita por estados do Brasil em mapa 61
Figura 26 – Distribuição de AOV por estados do Brasil em ranking 62
Figura 27 – Distribuição de AOV por estados do Brasil em mapa 63
Figura 28 – Distribuição de Frete Médio em Reais por estados do Brasil em ranking . . 64
Figura 29 – Distribuição de de Frete Médio em Reais por estados do Brasil em mapa . . 64

Figura 30 – Distribuição Fração do Frete em Relação ao Preço por estados do Brasil em ranking 65
Figura 31 – Distribuição de Fração do Frete em Relação ao Preço por estados do Brasil em mapa 66
Figura 32 – Distribuição de Tempo de Entrega Médio em Dias por estados do Brasil em ranking 67
Figura 33 – Distribuição de Tempo de Entrega Médio em Dias por estados do Brasil em mapa 68
Figura 34 – Distribuição de Percentual de Pedidos Atrasados por estados do Brasil em ranking 69
Figura 35 – Distribuição de Percentual de Pedidos Atrasados por estados do Brasil em mapa 69
Figura 36 – Distribuição de Satisfação entre 1 e 5 por estados do Brasil em ranking . . . 71
Figura 37 – Distribuição de Satisfação entre 1 e 5 por estados do Brasil em mapa 71
Figura 38 – Distribuição de pedidos por data 73

LISTA DE TABELAS

Tabela 1 – Conteúdo das tabelas do bando de dados Olist 37
Tabela 2 – Detalhamento de tabelas 40
Tabela 3 – Distribuição de tipos de dados 42
Tabela 4 – Relação das colunas com valores nulos 44
Tabela 5 – Destaque dos elementos nulos 45
Tabela 6 – Repetição de colunas chave entre tabelas 45
Tabela 7 – Dados Geográficos - Pedidos e Preço 56
Tabela 8 – Dados Geográficos - Entrega 57
Tabela 9 – Métricas de pedidos por estado 58
Tabela 10 – Métricas de pedidos por consumidor por estado 59
Tabela 11 – Métricas de receita por estado 61
Tabela 12 – Métricas de AOV por estado 62
Tabela 13 – Métricas de frete por estado 64
Tabela 14 – Métricas de Fração de Frete por estado 65
Tabela 15 – Métricas de tempo de entrega por estado 67
Tabela 16 – Métricas de pedidos atrasados por estado 69
Tabela 17 – Métricas de satisfação por estado 70
Tabela 18 – Métricas variação histórica dos pedidos 72

LISTA DE ABREVIATURAS E SIGLAS

AOV	Average Order Value
B2B	Business to Business
CEP	Código de Endereçamento Postal
CRM	Customer Relationship Management
EDA	Explaratory Data Analisys
ERP	Enterprise Resource Planning
FKX	Foreign Key
FRV	Frequência Recência Valor
LGPD	Lei Geral de Proteção de Dados
PK	Primary Key
PNAD	Pesquisa Nacional por Amostra de Domicílios
Regic	Pesquisa Regiões de Influência das Cidades
RNP	Rede Nacional de Pesquisa/ Rede Nacional de Ensino e Pesquisa (a partir de 1999)
SaaS	Software as a Service
SEO	Search Engine Optimization
SIDRA	Sistema IBGE de Recuperação Automática
WWW	World Wide Web

SUMÁRIO

1 INTRODUÇÃO 16
1.1 Trabalhos Relacionados 17
1.2 Motivação 18
1.3 Objetivos 19
1.3.1 Objetivo Geral 19
1.3.2 Objetivos Específicos 19
1.4 Organização do Trabalho 20
2 FUNDAMENTAÇÃO TEÓRICA 21
2.1 *Ecommerce* 21
2.1.1 Definição de Ecommerce 21
2.1.1.1 Marketing Digital *22*
2.1.1.2 Vendas e Propaganda *22*
2.1.1.3 Finanças e Contabilidade *22*
2.1.1.4 Cadeia de Suprimentos e Operações *22*
2.1.1.5 Tecnologia e Dados *23*
2.1.2 Ecommerce no Brasil 23
2.2 Análise exploratória de dados 24
2.3 O papel do engenheiro no mercado digital 24
2.4 Digitalização no Brasil 25
2.4.1 Acesso à tecnologia no Brasil 26
2.5 Análise demográfica 27
2.5.1 Acesso à internet 29
2.5.2 Renda 30
2.5.3 Infraestrutura urbana 31
3 METODOLOGIA 35
3.1 O bando de dados Olist 35
3.1.1 Olist 36
3.1.2 Estrutura do banco de dados 36
3.1.2.1 Limitações do banco *38*
3.2 Relatórios econômicos oficiais e especializados 38

3.3 **Procedimentos** 39
4 **RESULTADOS E DISCUSSÕES** 40
4.1 **Análise exploratória de dados** 40
4.1.1 ***Categorias de produtos*** 45
4.1.2 ***Status dos pedidos*** 46
4.1.3 ***Análise geográfica*** 52
4.1.3.1 *Pedidos* *58*
4.1.3.2 *Entregas* *63*
4.1.3.3 *Satisfação* *70*
4.1.4 ***Análise histórica*** 70
5 **CONCLUSÕES E TRABALHOS FUTUROS** 74
REFERÊNCIAS 75

1 INTRODUÇÃO

O comércio digital no Brasil apresenta grandes desafios desde o início de sua implantação, tanto em caráter geográfico, econômico e cultural, mas também histórico e político. As dimensões continentais do Brasil, sua desleal distribuição de renda e industrialização tardia se somam a uma insegurança diária em relação ao consumo e ao universo do comércio. Esse cenário se traduz em mais etapas na já difícil tarefa de comerciantes digitais converterem vendas e se estabelerecem local e globalmente.

Para fazer frente a esses desafios, comerciantes precisam desenvolver técnicas cada vez mais assertivas e usar de decisões rápidas e precisas, desde a abordagem, até a fidelização dos clientes. No universo digital, a impessoalidade e o grande volume de acessos se combinam para tornar essa tarefa ainda mais difícil.

É nesse contexto que a análise de dados se apresenta como uma necessidade para auxiliar empreendedores digitais a acompanhar o desenvolvimento de seus negócios, recorrendo a uma abordagem guiada por dados na relação com o cliente e na tomada de decisões. Seja na recomendação de produtos, logística de entrega e expansão, e até divulgação; quanto maior o conjunto de informações, maior a possibilidade de identificar tendências e mudar o caminho de um negócio.

No Brasil, o acesso a banco de dados públicos envolvendo *ecommerce* é raro e deficiente, *databases* contendo evolução do cenário nacional de comércio digital durante a pandemia de COVID 19, por exemplo, não estão disponíveis de forma gratuita online. Dados mais antigos, porém, podem ajudar a gerar uma perspectiva e visualizar comportamentos de brasileiros *online*, sazonal e geograficamente. É possível investigar ainda possível relação entre desenvolvimento econônimo, desigualdade social, acesso à tecnologia da informação e infraestrutura de comércio durante um certo período de tempo, fazendo cruzamentos com dados oficiais. Esses estudos contribuem para o desenho de uma paisagem do Brasil no cenário de comércio digital, e ajudam a entender os fenômenos que surjam em janelas futuras de tempo.

Sendo assim, o presente trabalho tem o intuito de realizar análise exploratória em um conjunto de dados público relacionado a comércio digital no Brasil e investigar o cenário de coleta socio-político-econômico dessas informações, ao tempo que se propõe a descrever o máximo do conteúdo presente através de observações úteis e de fácil consumo. Essa abordagem busca verificar as possibilidades de análise que um banco de dados atualizado poderia disponibilizar ao comércio digital brasileiro.

A análise exploratória foca na distribuição de compras realizadas em diferentes contextos sociais, distribuição populacional, distribuição de renda e acesso a internet, mas apresentado na perspectiva de divisão entre estados da federação. A investigação estará limitada pelo volume temporal de dados, mas será verificado se existe sazonalidade no volume de compras e feito um recorte para período de promoção nacional.

1.1 Trabalhos Relacionados

Mata (2021) aborda o *ecommerce* conceitualmente, e propõe um fluxograma de decisão baseado no momento do negócio do lojista frente à grande segmentação que o comércio digital possui em suas modalidades. O trabalho ainda apresenta uma perspectiva histórica no *ecommerce* brasileiro com dados gerais, e um desenho dos principais jogadores vide retrato no período imediatamente anterior ao pico da pandemia de COVID 19. É importante destacar a dedicação de uma seção na conclusão de seu trabalho voltada ao uso de ciência de dados no universo do comércio digital.

Fortunato (2022) realizou uma análise exploratória de dados da Olist aplicando uma abordagem baseada no conceito de *Frequência Recência Valor (FRV)*, com seguida clusterização dos dados. Esse trabalho foi capaz de, através do estudo de pedidos por cliente, agrupando-os comparativamente com os métodos *k-means* e *k-methods*, e aplicando o método de cotovelo, separar os cliente em cinco *clusters*: clientes ativos de baixo valor e baixa frequência, clientes inativos de baixo valor e baixa frequência, clientes recorrentes, clientes assíduos e clientes de alto valor. O trabalho conseguiu observar que mesmo existindo variações entre os métodos, a intuição gerada pela divisão dos perfis de clientes agrega à análise comercial aplicada ao conjunto de dados. Segue ainda que houve a limitação do uso apenas da técnica *FRV* e demais segmentações podem ser exploradas.

Niwate (2021) explorou o banco já com uma perspectiva mais comercial, analizando a presença do modelo de negócio em contexto Brasileiro, trazendo à tona comparativamente elementos como a penentração e dinâmica abordada pelo marketplace quando comparado a concorrentes como Amazon. Houve também uma preocupação maior com dados de produtos, expandindo da ótica de clientes abordado por Fortunado, mas mantendo relevantes as avaliações dadas. Dentre as principais observações e conclusões estão a inversa proporção entre o preço dos produtos e quão boas foram as avaliações dadas, ainda assim com avaliação acima da média disponível; a análise da fonte de receita como proveniente tanto dos produtos ofertados quanto

das avaliações recebidas, mas também foi destacada a rápida capacidade de penetração no mercado do Brasil. Como trabalhos futuros o estudo recomenda verificar que outras relações comerciais podem ser derivadas do estudo desse banco de dados para serem exploradas por outras empresas de segmento semelhante.

Geraldi (2021) já aborda o comércio digital através do estudo de sua relevância no cenário internacional, trazendo uma profunda descrição histórica dos eventos que levaram às condições necessárias para a adoção significativa dessa forma de comércio, bem como os impactos disso para a economia global provenientes das dinâmicas de comércio. Por mais que se trate de uma abordagem internacional, dados os efeitos da globalização, muitas análises podem ser traduzidas para pontuar tendências em cenário local.

Nascimento, Pinheiro e Cipro (2022), já em 2022 contribuem à discussão através de um estudo da logística de comércio digital sobretudo pós efeito da pandemia de Covid 19, verificando o funcionamento das cadeias operacionais necessárias para funcionamento do modelo de negócios e as adaptações frente à eclosão de clientes durante o período de isolamento social. Aqui foi evidenciado as avaliações dos clientes e sua percepção no uso do serviço, ao tempo que explorada a adaptabilidade do setor e oportunidades aproveitadas.

1.2 Motivação

Sendo o universo do comércio digital complexo e abrangente, ao mesmo tempo que dinâmico e suscetível a fatores externos tão diversos quanto for a sociedade, é impressíndivel a maior quantidade de informações disponíveis durante as tomadas de decisão, desde a abordagem do cliente até um projeto de expansão. Em um país continental como o Brasil, e dada sua trajetória histórica e atual distribuição política e econômica, é razoável pensar que o comércio digital seja experienciado de forma diferente ao longo do país. O conhecimento dessa distribuição, tanto para a definição de suas particularidades como também para estudo dos possíveis impactos no futuro, é a curiosidade que motiva esse trabalho.

Vista a escassez de dados públicos relacionados a esse objeto de análise, produções acadêmicas, sobretudo recentes, são de difícil acesso. Aqui foi escolhida uma abordagem que prioriza a abrangência de informações : pretende-se traçar o caminho entre a análise exploratória de um banco de dados de comércio digital real, ao tempo que busca por uma contextualização com a realidade social e econômica do mercado nacional onde os dados foram coletados.

A justificativa deste trabalho se manifesta em sua relevância científica e social.

Por um lado, é notório que análise de dados tem um papel agregador quando se propõe a atacar problemas cotidianos através de ferramentas matemáticas e computacionais, por outro, os resultados encontrados somam aos esforços coletivos de outros pesquisadores em problemas de arquitetura semelhante, mesmo que no mais distante contexto. Isso dito, salienta-se ainda a contribuição ao mercado, ao sugerir mais uma análise de um contexto que move um setor econômico de crescente importância em tempos digitais.

O método de pesquisa utilizado foi o quantitativo, devido ao uso da matemática para descrever os fenômenos e explorar as informações obtidas. Trata-se então, de uma pesquisa baseada em números e análises estatísticas. De acordo com Gerhardt e Silveira (2009), a pesquisa quantitativa é orientada pelo positivismo e considera que a compreensão da realidade só ocorre por meio da análise de dados brutos.

1.3 Objetivos

1.3.1 Objetivo Geral

O presente trabalho trata de um problema de análise de dados. O objetivo geral é identificar particularidades de cada estado brasileiro em um banco de dados de comércio digital real, baseado em padrões de compra, sazonalidade e demografia, de forma a explicar o contexto de *ecommerce* brasileiro no intervalo de tempo em análise.

1.3.2 Objetivos Específicos

Com fins a corresponder ao objetivo geral, os objetivos específicos que devem ser alcançados ao longo do trabalho são os que seguem:

- Conceitualizar as nuances que compõem o *ecommerce* brasileiro;
- Contextualizar informações oficiais de distribuição demográfica da população brasileira frente aos dados em estudo;
- Realizar abrangente análise exploratória no banco de dados de compras;
- Analisar as distribuições produzidas.

1.4 Organização do Trabalho

Este trabalho está dividido em cinco capítulos. No presente capítulo está a introdução contendo justificativa, motivação, trabalhos relacionados, objetivos e conceituação preliminar. No segundo capítulo são apresentados os conceitos principais para entendimento do objeto de estudos do trabalho, seguido de discussão envolvendo a contextualização histórica e geográfica da problemática de comércio digital no Brasil. O terceiro capítulo apresenta a metodologia utilizada. O quarto capítulo compreende os resultados gerados através da análise exploratória nos dados em estudo. Finalmente, o quinto capítulo apresenta as conclusões do trabalho.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo serão apresentados os principais conceitos teóricos para facilitar a compreensão da metodologia proposta. Aqui serão abordados: a definição de *ecommerce*, *ecommerce* no Brasil, análise exploratória de dados, o papel do engenheiro no mercado digital, digitalização no Brasil e análise demográfica.

2.1 *Ecommerce*

As relações comerciais moldam a sociedade e são por elas moldadas desde as primeiras organizações humanas. Seja pela expansão através da busca por especiarias ou na construção de uma macroeconomia moderna complexa, alguns desafios se mantém presentes, mesmo que agora vestidos com outro contexto. A jornada do cliente ainda necessariamente contempla: apresentar o produto, efetuar a venda, fidelizar o cliente e estimular novas compras. Em um mundo globalizado, com a internet como ferramenta essencial no dia a dia, as relações comerciais ganham espaço no universo virtual. A primeira transação efetuada através da internet data de 11 de agosto de 1994 e consistiu na compra de um *CD* do cantor Sting, (DEARO, 2015). Desde então, a Compra e venda através do *ecommerce*, ou comércio digital, tem suas fronteiras expandidas, número de clientes e pedidos escalados, logística de entrega e necessidades de segurança reformuladas. E é nesse contexto que ferramentas também digitais precisam ser incorporadas à equação. Ferramentas de análise de *big data* se propõem a ajudar no entendimento desses novos desafios.

2.1.1 Definição de Ecommerce

Existem mais de uma definição para *ecommerce*, as mais simples o apresentam como o resultado de transações comerciais realizadas online; porém, existe discordância nessa definição entre estudiosos e profissionais da área, visto que a digitalização do comércio vai muito além da mera transação, como mostra ARAÚJO (2003). No processo de tornar o comércio digital, atividades cruciais são hora adaptadas, redefinidas ou criadas. É possível destacar cinco grandes pilares que devem ser levados em consideração para as particularidades de operação de um *ecommerce*. São eles: *Marketing Digital, Vendas e Propaganda, Finanças e Contabilidade, Cadeia de Suprimentos e Operações, Tecnologia e Dados*.

2.1.1.1 Marketing Digital

Pode parecer contraintuitivo a princípio, mas o Marketing Digital, segundo o SEBRAE (2022), diz respeito às diferentes relações comerciais que são predominantes ou particulares do comércio digital. A alta dependência de soluções tecnológicas terceiras, como o uso de plataformas de gestão do comércio, as técnicas de expansão voltadas para construção de *marketplaces*, otimização de *Search Engine Optimization (SEO)*, difusão em diferentes mídias para estabelecer presença, são alguns dos exemplos.

2.1.1.2 Vendas e Propaganda

Como aponta HAYES (2022), as vendas e o processo de venda em um *ecommerce* possuem também suas particularidades, diferente das lojas físicas convencionais, a quase interminável lista de concorrentes e produtos está a um clique de distância do consumidor, o que torna a experiência de compra mais dinâmica para o cliente e desafiadora para o logista. Técnicas como forte identidade e personalidade das lojas, na tentativa de se emplacar como referências, associadas com forte presenças de anúncios espalhados em diversos espaços na rede, diversidade de meios de pagamento, parceirias diversas com influenciadores e eventos, se destacam como elementos na prospecção de clientes.

2.1.1.3 Finanças e Contabilidade

Partindo do apontado por KHAN (2022), é possível aferir que enquanto principais preocupações em lojas físicas, no que se diz respeito a finanças, seriam algo como aluguel de espaços e customização desses espaços; no universo digital, se destacam os custos de servidores, domínios, e com profissionais especializados. Nesse universo, a valorização dos especialistas tem se mostrado fator extremamente relevante no dimensionamento do negócio, haja vista que certos setores vendem exclusivamente produtos ou serviços digitais. Precificação de produtos e previsão de expansão também passam a levar em consideração presença e manutenção na rede, e a competição acirrada.

2.1.1.4 Cadeia de Suprimentos e Operações

Tomando como base as principais tendências descritas por Keenan e Moore (2022), mesmo ainda sendo possível a associação entre loja física e *ecommerce*, particularidades logísti-

cas precisam ser consideradas. Seja retirada em loja, estratégia de envio de produtos, ou mesmo a aquisição de matéria prima ou produtos de revenda, perpassa pelo estudo da demanda com critérios específicos a depender do setor. Existem modalidades como *dropshipping*, por exemplo, que não possuem estoque físico e onde o lojista age apenas como uma vitrine para divulgação de produtos terceiros. Outro caso, como *marketplaces*, podem ter estoque de terceiros, e além do processo de entrega ao cliente, podem dimensionar também uma logística de compartilhamento interno de produtos entre as lojas, o que impacta diretamente na prática de frete e disponibilidade de vendas.

2.1.1.5 Tecnologia e Dados

Aqui o trabalho de engenheiros eletricistas, de dados, de solução, de softwares e diversos outros profissionais fica mais claro nesse tipo de mercado. É onde são construídas ou otimizadas as ferramentas necessárias para a operação e manutenção do negócio. Seja na perspectiva de *frontend*, como será exibido o site e produtos, seja no *beckend*, em como as configurações se organizam para possibilitar o gerenciamento do negócio, seja ainda no estudo dos dados para averiguar eficiência e propor expansão ou correção, como destacado por Shen (2020), e até desenvolvimento de novas ferramentas em nuvem ou físicas como servidores, máquinas e aplicativos que impactem o *ecommerce*.

2.1.2 Ecommerce no Brasil

O comércio digital surge no Brasil como uma proposta revolucionária de contato com aquisição de bens e serviços. A primeira compra online no Brasil é reconhecida como a compra de um livro em 1995 na livraria online *Booknet* que anos depois viria a ser vendida e renomeada para a hoje conhecida *Submarino* (OLIVEIRA, 2021), mas é possível considerar a *Brasoftware* pioneira em loja virtual no Brasil com aceite de pagamentos de forma virtual , desenvolvida em 1996 (MENDONÇA, 2016). A partir de então começa a corrida para estabelecimento do que era uma curiosidade do era moderna, para um modelo de negócios lucrativo e capaz de atacar a essência dos problemas de gestão do alto volume de dados do comércio tradicional.

Em 2022 haviam no Brasil, de acorco com Neiva (2023), 565.300 lojas virtuais ativas com 83,79 milhões de compradores que movimentaram um total 169,59 bilhões de reais. Essa rede exige um extenso volume de profissionais com alto nível de capacitação. Um exemplo salutar dessa penetração no mercado é a existencia de fornecedores *Business to Business (B2B)*

de placas solares, de forma online, que alimentam mercados nacionais para facilitar atuação de empresas locais desse segmento na realização de orçamentos.

Vale ressaltar que não só consumidores e lojistas, mas empresas de tecnologia brasileiras também são destaque nesse setor, responsáveis pelo desenvolvimento de produtos que ajudam outros lojistas na realização de suas atividades. Essas empresas fornecem estrutura com arquiteturas próprias para garantir a operação dos vendedores, absorvendo grande parte da complexidade de gestão das especificidades mencionadas no trato do *ecommerce*.

2.2 Análise exploratória de dados

A *Explaratory Data Analisys (EDA)* ou análise exploratória de dados, é uma etapa extremamente importante no processo de traduzir dados para informação útil. Nela, são dispostos os detalhes dos bancos de dados em estudo, quais são os atributos, sua correlação e relevância. É nessa etapa que são levantadas também as distribuições, comumente através de gráficos, que se mostram poderosas ferramentas de síntese de vastos grupos de dados em uma linguagem visual e de interpretação mais fácil que as longas linhas e colunas.

> A análise exploratória de dados nos fornece um extenso repertório de métodos para um estudo detalhado dos dados, antes de adaptá-los. Nessa abordagem, a finalidade é obter dos dados a maior quantidade possível de informação, que indique modelos plausíveis a serem utilizados numa fase posterior, a análise confirmatória de dados ou inferência estatística. (MEDRI, 2011, p. 2)

É nessa análise que se fazem aferições sobre o banco de dados, se ele é capaz de responder às perguntas exploratórias levantadas, se os dados são conclusivos, se faltam dados ou se os dados apontam para perguntas mais relevantes.

2.3 O papel do engenheiro no mercado digital

A engenharia é uma profissão que evolui e se adapta rapidamente. Antes do advento da era digital, disciplinas de *CAD*, por exemplo, não existiam, enquanto desenho à mão concentrava uma habilidade indispensável na rotina diária de trabalho do engenheiro. Hoje, conhecimentos como *BIM* e inteligência computacional aplicada já se monstram como necessidades de um engenheiro contemporâneo.

O presente trabalho vai além do entendimento que novas ferramentas impactam no trabalho do engenheiro, e ataca a temática que o mercado de tecnologia como um todo possui potencial de fornecer grandes oportunidades para a carreira profissional de um engenheiro, sobretudo eletricista, e é um importante componente a ser discutido na graduação.

Seja tomada a engenharia como a arte de solucionar problemas. Da engenharia elétrica vieram os primeiros componentes e tecnologias que possibilitaram o advento da era da informação digital: energia elétrica, dispositivos eletrônicos, computadores, internet, aparelhos celulares. É razoável que com cursos específicos de computação e ciências digitais, as engenharias clássicas não adentrem em profundidade alguns temas, todavia, é imprescindível que a economia digital seja de conhecimento de um engenheiro dos tempos modernos.

A economia dos tempos da internet com *big techs* e teias sociais digitais, não só representa um mercado consumidor para desenvolvimento de empreendedorismo para projetistas, como também representa uma porta aberta para profissionais recém formados que possuem o perfil dos sonhos do setor privado. Companhias de tecnologia digital, por vezes, estão em busca especificamente de engenheiros das mais diversas variações para trabalhar no mercado digital.

A engenharia elétrica como precursora desse fenômeno social e tecnológico, a era digital, e que possui intrínseca relação com linguagem de programação e telecomunicações, tem grande vantagem competitiva e oportunidade de destaque nesse universo. E precisa ser desafiada a entender sua dinâmica, qual seu papel social, e quais técnicas podem ser implementadas para tirar o máximo proveito desse formato de trabalho.

2.4 Digitalização no Brasil

O Brasil iniciou tardiamente seu processo de digitalização em muitos setores. Ainda hoje não é raro encontrar, mesmo nas capitais, grandes empreendimentos que possuem processos analógicos; e empresas de grande porte sendo forçadas a uma mudança radical nas suas operações dada a exigência imposta pela demanda e pela concorrência. No que diz respeito a serviços públicos, também não é raro encontrar hospitais sem gestão computadorizada de recepção de pacientes e prontuários, mesmo escolas ainda são por vezes reféns de ficheiros e arquivos físicos. Já no comércio, ficou evidente que a gestão bruta de toda a informação que o segmento exige, na velocidade que é exigido, não é escalável, se não através de tecnologias de gestão de dados.

Serviços como *Customer Relationship Management (CRM)* ou gestores de relacionamento com o cliente, e *Enterprise Resource Planning (ERP)*,planejamento de recursos

empresariais, são essenciais para atividades como gestão de estoque, controle de finanças e impostos, além de planejamento e supervisão de funcionários. Sistemas esses que produzem e consomem grandes quantidades de dados.

Muitos desafios perpassam essa digitalização de processos e serviços no universo do comércio digital, dentre eles, é possível destacar a escassez de profissionais capacitados nas diversas áreas da tecnologia que perpassam a implementação segura e completa das funcionalidade necessárias, aliado a uma dificuldade de acesso da população à interação com os canais de uso. Esse último, deriva de alguns fatores: desconfiança com o formato de transação e eventual não recebimento, balança de praticidade entre a comodidade da compra de casa e o contato direto com o produto, acesso à tecnologia seja por desconhecimento da manipulação, seja por falta de poder aquisitivo, falta de cobertura de internet apropriada ou moradia em área não coberta pelas estratégias de envio praticadas, dentre várias outras.

2.4.1 Acesso à tecnologia no Brasil

Acesso à tecnologia é um debate constante nas ciências sociais, mas pouco explorado pelas engenharias e ciências exatas. Como problematiza Jacinski (2012), acaba que a solução do problema, objetivo dessas ciências, aparenta se limitar ao ponto em que a tecnologia é divulgada, e não o caminho entre a produção tecnológica e o consumidor final da tecnologia. No universo do comércio essa realidade acaba por ser flexibilizada, visto que os profissionais técnicos são obrigados à ocupar posições de análise social e engenheiros passam a considerar comportamentos de clientes, segmentação de públicos consumidores e extratégias de engajamento. Análises que não mais podem ser terceirizadas e influenciam diretamente em tomadas de decisão nos projetos tecnológicos desenvolvidos por esses profissionais. Nesse sentido, é válido explorar como é o acesso à tecnologia no Brasil, sobretudo à internet, computadores e celulares, como também aos métodos de pagamento relacionados ao *ecommerce*.

Segundo Barbosa (2022), em artigo publicado pela revista Forbes, o Brasil é o quinto pais no ranking mundial de usuários de internet, com 165 milhões de usuários em 2022, o que corresponde a 80% da população. Na América latina o cenário é mais interessante ainda, visto que o Brasil se mostra líder absoluto, sendo o México segundo colocado com 96 milhões de usuários, o que corresponde a uma diferença de 69 milhões para o Brasil. Nesse caso os números absolutos são relevantes visto que esse é também um indicativo do potencial consumidor de produtos do mercado digital.

Ainda no mesmo artigo são pontuados os canais de preferência usados para acesso à internet: em primeiro lugar, o celular é usado por 99% dos usuários, comportamento esperado desde que o smartphone se tornou elemento essencial do ser humano no cotidiano moderno, porém o segundo apresentado com mais frequência pode ser surpreendente: 50% dos brasileiros acessam a internet através da televisão. O advento das televisões inteligentes, ou smartTVs constituiu uma mudança substancial na forma que o brasileiro consome conteúdo no universo online, e adaptou a TV que historicamente é um elemento já bem estabelecido da cultura nacional. Em seguida vêm os notebooks e computadores com 38%.

Essa distribuição de uso dos canais se justifica pelo perfil de acesso da população às tecnologias. Mesmo os mais básicos smartphones possuem conexão à internet e capacidade de rodar apps de uso frequente, troca de mensagens, acesso à informação e entretenimento são demandas comuns de quase a totalidade dos usuários. A televisão seguiu uma evolução natural do consumo de conteúdos midiáticos uma vez que o mercado torna as smartTVs o padrão, mesmo que variados os modelos e faixa de valores; o sinal analógico é substituído pelo digital progressivamente a partir de 2018 com a Portaria Nº 378/2016 do Ministério das Comunicações (BRASIL, 2016), forçando um *upgrade* dos aparelhos ou aquisição de um receptor capaz de processar o sinal digital, mas claro também em função da aderência generalizada do acesso à internet nas residências com expansão das redes de fibras óticas. Segundo a ANATEL (2023), cerca de 72.5% das conexões de banda larga fixa são realizadas através de fibra ótica que corresponde a mais de 61.1 milhões de contratos. Finalmente, computadores e notebooks, além do custo de aquisição, quando comparados à praticidade e eficácia dos celulares para as tarefas convencionais, acabam sendo restrito a menos usuários, sobretudo em uso profissional.

A adaptação do uso de meios de pagamento é outro fator interessante. No *ecommerce*, praticidade e flexibilidade podem ser fatores decisivos no processo de conversão e venda. Fernandes (2022) aponta que 82% dos brasileiros que fazem compras online possuem cartão de crédito, por mais que a *Pesquisa Nacional por Amostra de Domicílios (PNAD)* demonstre que o método de pagamento preferido pelo brasileiro ainda é o dinheiro (71%), seguido de PIX (71%), cartão de débito (66%) e cartão de crédito (57%).

2.5 Análise demográfica

Como aponta Brito (2006), o Brasil carrega uma história construída por desigualdades, desde o processo colonizatório até fenômenos como higienização das capitais, êxodo rural e

industrialização tardia, contribuem para acentuar contrastes drásticos entre qualidade de vida, oportunidades de acesso e, por vezes, distorção na percepção da realidade em volta.

A tecnologia, sem dúvidas, equaliza um pouco o acesso a espaços de compra, por exemplo. Se antes um periférico ou morador de zona rural poderia se sentir inibido a ir em um shopping, seja pela dificuldade de mobilidade, seja pela expectativa de atender a requisitos de aparência e média de consumo, no meio digital navegar entre as lojas é livre para todos que possuam praticamente qualquer aparelho conectado à internet. Porém, a democratização desses espaços também inclui entender limitações características de pessoas nas classes mais baixas, e inclusive sua distribuição no espectro de consumidores.

Por mais que baste uma conexão lenta e um celular simples para acessar a internet, outros fatores influenciam no sentimento de liberdade no uso das tecnologias. Segundo dados do IBGE (2020a), em 2019 os 1% da população mais rica do Brasil ganhavam em média 33,7 vezes mais que os 50% da população mais pobre. Traduzido em números absolutos, 1% da população brasileira recebe em média 28659 reais enquanto a média dos 50% mais pobres é de 850 reais.

Ainda de acordo com IBGE (2020a), vale pontuar que a variação dessa média geograficamente é considerável. No extremo positivo a região sul possui 1102 reais médios para os 50% mais pobres, e um multiplicador em relação aos mais ricos de 24,5 vezes. Já no nordeste, no extremo oposto dos dados das regiões, a média dos 50% mais pobres é de 569 reais, configurando um multiplicador de 35,8 vezes menos que os 1% mais ricos da região.

É importante pontuar que visto a extrema diferença de poder aquisitivo, a garantia de acesso aos canais de compra não é suficiente para aferir liberdade de participação como consumidor, haja vista a priorização de necessidades básicas para subesistência. Elementos como dificuldades logística, preço e disponibilidade de frete, podem ser incluídos aqui para o caso de residentes em locais de difícil acesso ou pouca mobilidade urbana, como periferias e zonas rurais.

No Brasil são observadas diversas diferenças regionais, sobretudo no relativo a frete e volume de pedidos. Essa diferença regional faz alusão a uma característica notória: o Brasil é gigante e diverso. Essa diversidade não é só cultural, mas carrega consigo o peso histórico do processo de desenvolvimento e regionalização nacional. Sob o qual, regiões foram privilegiadas no que se diz respeito à acelerada concentração populacional. Brito (2006) descreve o processo de deslocamento brasileiro para as metrópoles, lembrando o papel ativo do estado na importação de imigrantes europeus e seu descaso no cuidado com populações locais que culminaram nas

movimentações de êxodo rural, a procura por oportunidades profissionais, somado à fuga de intempéries climáticas. Com o tempo, o processo começa a reverter, com o crescimento das regiões periféricas nas metrópoles, pouco saneamento e falta de segurança; associado, tanto à omissão do estado, quanto ao custo de vida supervalorizado das capitais, é gerado o êxodo reverso em direção aos interiores. Ainda que em 2000 40% da população brasileira seguia residindo em metrópoles.

Nesse processo, como aponta Brito (2006) um dos subprodutos é o desenvolvimento de uma infraestrutura mais robusta em relação às regiões menos populosas. Isso é particularmente verdade para as regiões Sudeste e Sul, vista a abordagem de povoamento adotada pós vinda da família real portuguesa para o Rio de Janeiro, e a exploração de ouro, café e leite no triângulo mineiro. Já o Nordeste e Norte brasileiros sofreram respectivamente de constantes secas sem políticas de intervenção para proteção social por parte do estado, e um processo de intensa extração tropical e conflito com comunidades indígenas. Dito isso, houve uma concentração populacional principalmente em regiões litorâneas e fluviais, mas o efeito reverso em direção ao interior se mostra mais lento e em menor proporção nessas regiões.

A *PNAD* é um instrumento do *IBGE* que faz o acompanhamento anual de indicadores econômicos e sociais e pode ajudar a compreender como tem sido o desenvolvimento dessas regiões ao longo dos últimos anos. Esse paralelo é importante uma vez que o comércio, mesmo o digital, é intrinsecamente dependente de estrutura física, disponibilidade econômica da população e acesso à tecnologia. A seguir são verificados dados da PNAD em 2017, 2018, 2019, em seguida prosseguindo para os dados mais recentes disponíveis, objetivando verificar também o período pós pandemia de COVID 19.

2.5.1 Acesso à internet

Para atuar como potencial consumidor no comércio digital, é necessária uma conexão com a internet. Pode parecer distante, mas haja vista que a criação da *World Wide Web (WWW)* data de outubro de 1994 (CARVALHO, 2006), e a primeira transação comercial no Brasil de 1995 (OLIVEIRA, 2021), é possível perceber como o acesso à internet no Brasil é recente, com pouco mais de 30 anos de história. A Figura 1 ilustra esse fato ao apresentar a estrutura da *Rede Nacional de Pesquisa/ Rede Nacional de Ensino e Pesquisa (a partir de 1999) (RNP)*, exclusiva para universidades e instituições de pesquisa, e único tipo de acesso à qualquer tipo de rede no Brasil em 1992.

Figura 1 – RNP em 1992

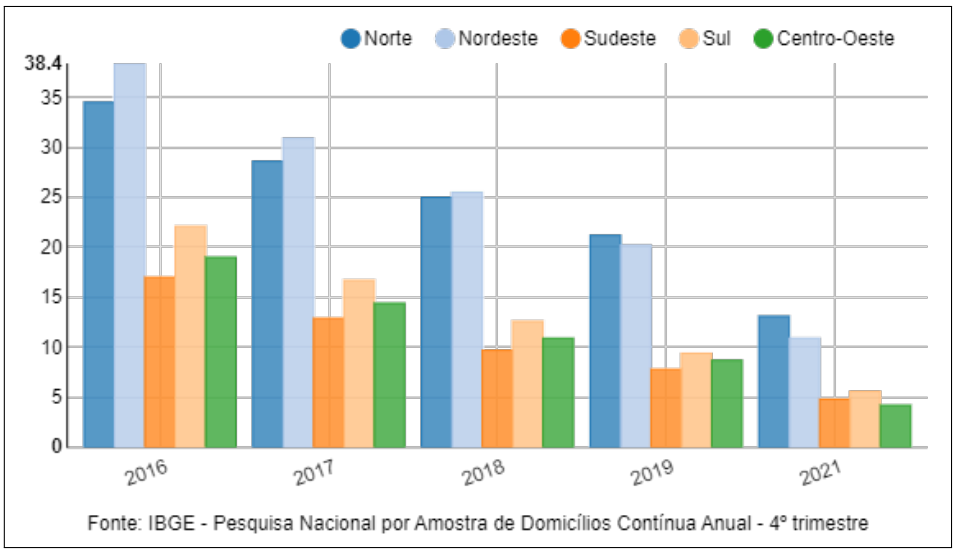
Fonte: (GETSCHKO, 2002 apud CARVALHO, 2006)

Dito isso, é razoável considerar que o acesso à população brasileira em geral ainda é um processo em desenvolvimento. Nesse sentido, recorre-se à *PNAD* Contínua anual do *IBGE* que levantou nos domicílios brasileiros os moradores que possuíam acesso à internet nos seus vários formatos. A Figura 2 foi gerada através do *Sistema IBGE de Recuperação Automática (SIDRA)*, e demonstra o resultado percentual de moradores que não possuem acesso à internet. Vale ressaltar que em números absolutos se trata de universo consultado com variação de 205165 pessoas em 2016 e 213188 pessoas em 2021 (IBGE, 2022a). Nela, é possível verificar com destaque que as regiões Norte e Nordeste apresentam a maior quantidade de moradores sem conexão com internet, percentual esse que vem diminuindo ao longo dos anos, mas se mantém substancialmente 2,46 vezes, mais que o dobro, maior quando comparado às demais regiões, ainda que em 2021.

2.5.2 Renda

A renda é um outro fator importante de ser observado, haja vista que os padrões de consumo, e também os canais acessados para realização do consumo, estão intimamente ligados à questões de classe. Um dos elementos que ilustra esse fenômeno é a segmentação de público

Figura 2 – Percentual de moradores sem acesso à internet nas grandes regiões brasileiras



Fonte: IBGE (2022a) - *SIDRA* - Tabela 7307

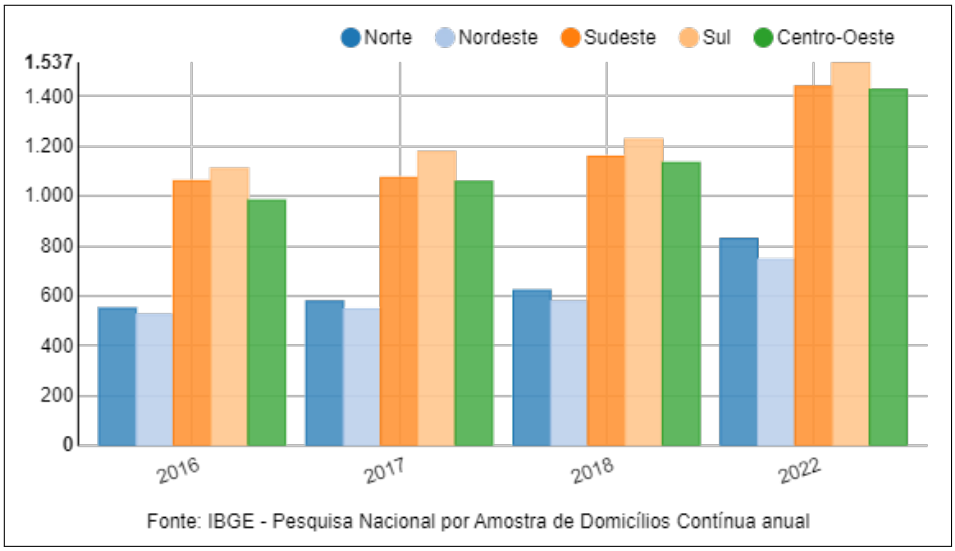
que faz uso de lojas e mercados de bairro, frente ao público de grande supermercados. O volume de compras, o preço médio dos produtos e o acesso até as lojas age para segmentar os clientes. Como sugere a CETIC.BR (2020), no universo digital isso também pode ser verdade quando o acesso se manifesta na posse de tecnologia para acesso às lojas *online*, disponibilidade de métodos de pagamento eletrônico e/ou disponibilidade de crédito.

Nesse sentido, é proposto observar qual a distribuição de renda da população brasileira nas grandes regiões do país, a fim de entender se existe um contraste que indique uma diferença de classe que possa interferir nos padrões de consumo. A Figura 3 usa novamente o *SIDRA* com dados da *PNAD* (IBGE, 2022b) para demonstrar que a renda média per capita dos moradores das regiões Norte e Nordeste são em média apenas 53,6%, praticamente metade, da renda de moradores das demais regiões. Aqui foram excluídos os 5% mais ricos a fim de evitar possíveis *outliers* na análise.

2.5.3 Infraestrutura urbana

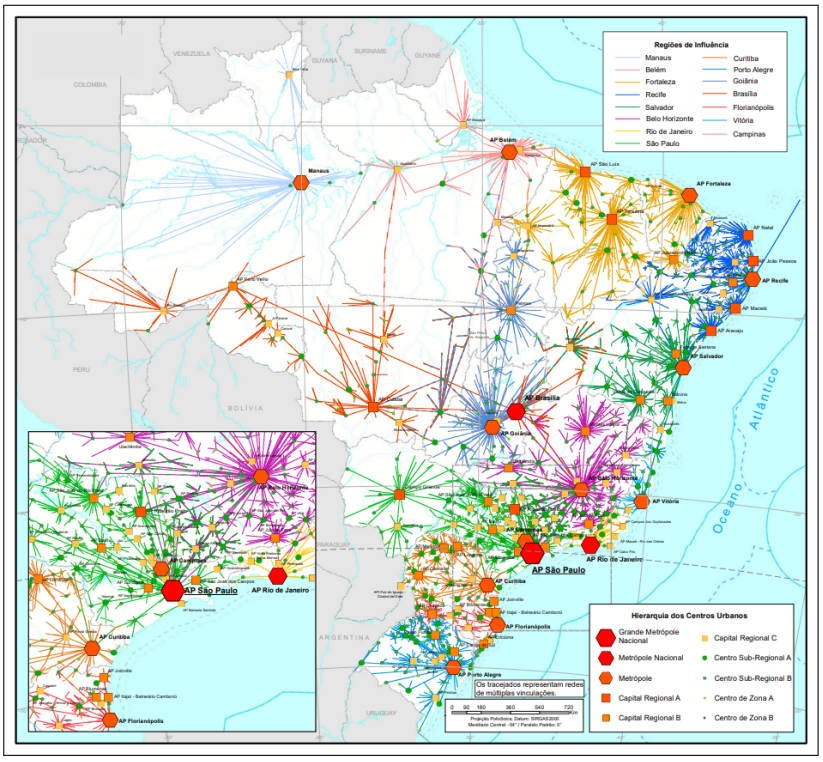
A infraestrutura urbana já não entra no escopo da *PNAD*, mas é possível inferir o papel integrador das cidades brasileiras através da influência das cidades. A *Pesquisa Regiões de Influência das Cidades (Regic)* tem o propósito de identificar e analisar a rede urbana brasileira, estabelecendo a hierarquia dos centros urbanos e as regiões de influência das Cidades (IBGE, 2020b). A Figura 4 ilustra essas regiões de influência com visível concentração de densidade nas regiões Sudeste e Sul, em contraste com as regiões Centro-Oestre e Norte.

Figura 3 – Rendimento médio real domiciliar per capita a preços médios do ano nas grandes regiões brasileiras - P95



Fonte: IBGE (2022b) - *PNAD Contínua anual*

Figura 4 – Regiões de influência das cidades brasileiras em 2018

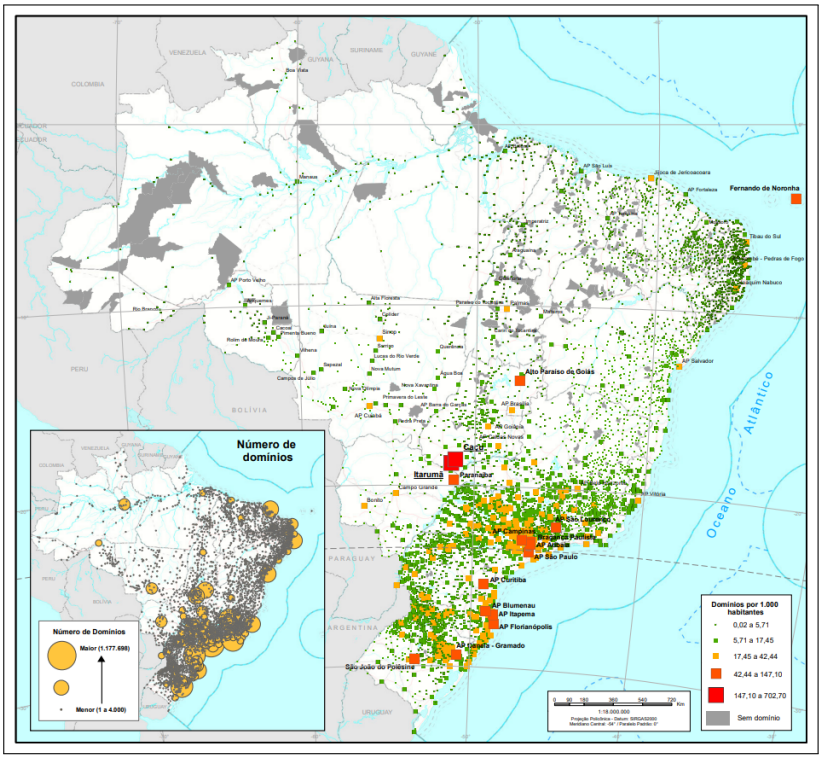


Fonte: IBGE (2020b) - Regiões de Influência das cidades 2018

A Figura 5 avança na apresentação de concentração de domínios de internet, onde mais uma vez há uma concentração de densidade nas Sudeste e Sul. Um dos elementos em destaque é justamente o impacto da conectividade à internet na qualificação da rede urbana,

avaliando quanto cada localidade está habilitada à produção de conteúdo (IBGE, 2020b). A mesma pesquisa ainda aponta importantes comentário quanto à concentração dos domínios .br na Cidade de São Paulo-SP com 28,7% o que corresponte a 1117698 domínios; enquanto a mediana das demais cidades é apenas 17, e as regiões Norte e Nordeste são as que apresentam maior quantidade de cidades sem nenhum domínio, destaque para os estados Tocantins, Piauí, Maranhão e Paraíba. Apontando ainda que há uma coincidência entre ter domínios e participar dos circuitos produtivos da economia nacional.

Figura 5 – Densidade de domínios de Internet no Brasil em 2019

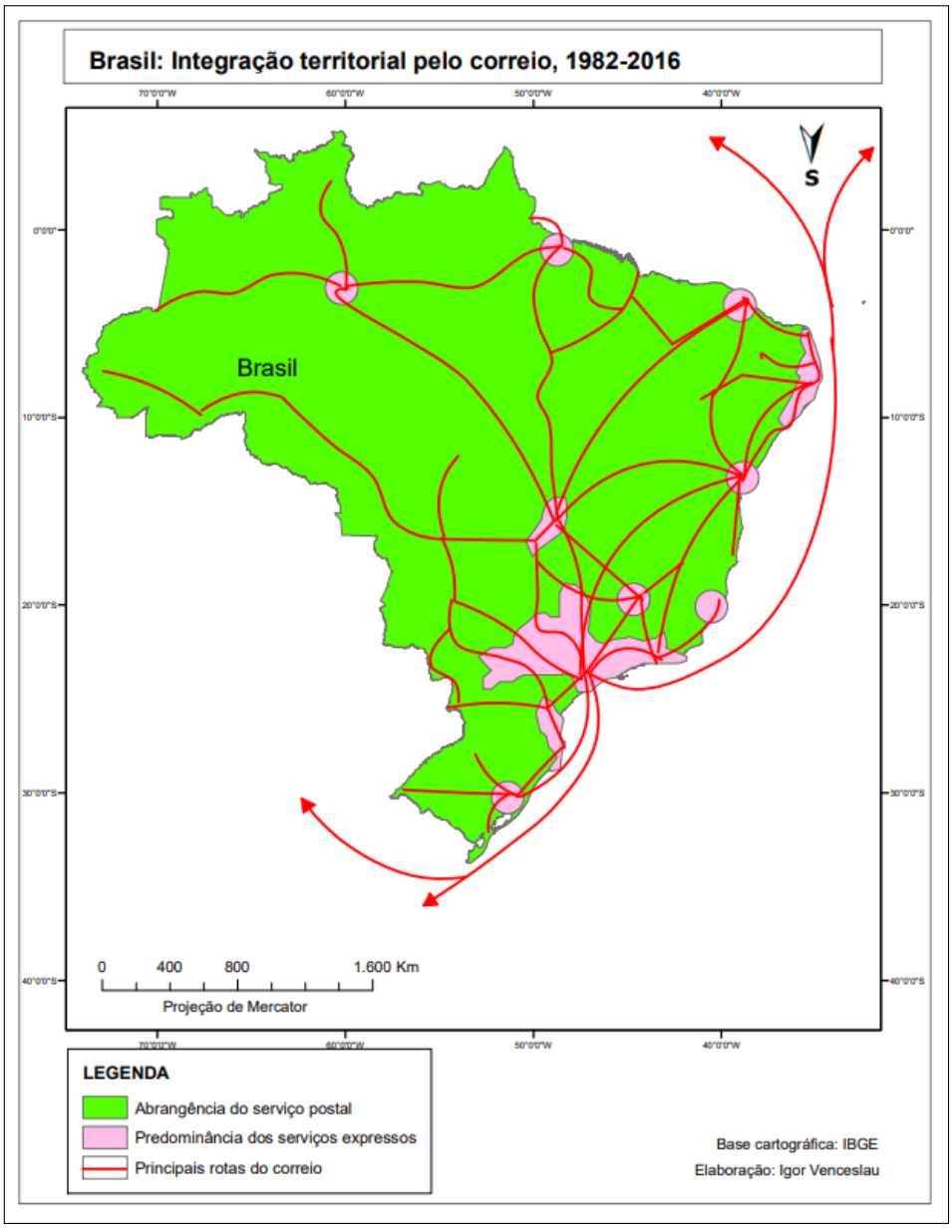


Fonte: IBGE (2020b) - Regiões de Influência das cidades 2018

Ainda é possível comentar sobre a rede de transporte terreste, sobretudo rodoviário, ao tempo que o estudo aponta que a média nacional é de 75km entre as cidades, com variações entre os estados na dimensão de 221km em média no Amazonas até 49km em São Paulo. Não é considerado o mérito de qualidade da vias nem a diferença entre tipos de vias de circulação, o que poderia contribuir para o entendimento de quais vias são majoritariamente destinadas ao tranporte de pessoas e quais são para transporte de mercadorias. Mas é possível pontuar que essa distribuição de rodovias, principal rede de transporte de bens nacionalmente, como elemento central para influir no processo de compra e venda online, ao tempo que impacta diretamente nas políticas de frete e tempo de entrega.

De forma a complementar essa abordagem de entregas, recorre-se a Venceslau (2017) que descreve a evolução do serviço de correios desde a década de 1960 no Brasil. A Figura 6 apresenta como estava o serviço em 2016, com clara concentração no Sudeste, em especial o estado de São Paulo. O autor ainda comenta que esse processo acompanhou a urbanização das cidades, e que a evolução do grau de complexificação da formação aeroespacial brasileira, por mais que assegure sua unidade interna, não supera completamente as desigualdades regionais historicamente produzidas. Chegando a apontar que dentre as funções assumidas dessa estrutura, está a prática do *ecommerce*.

Figura 6 – Integração territorial pelo correio 1982-2016

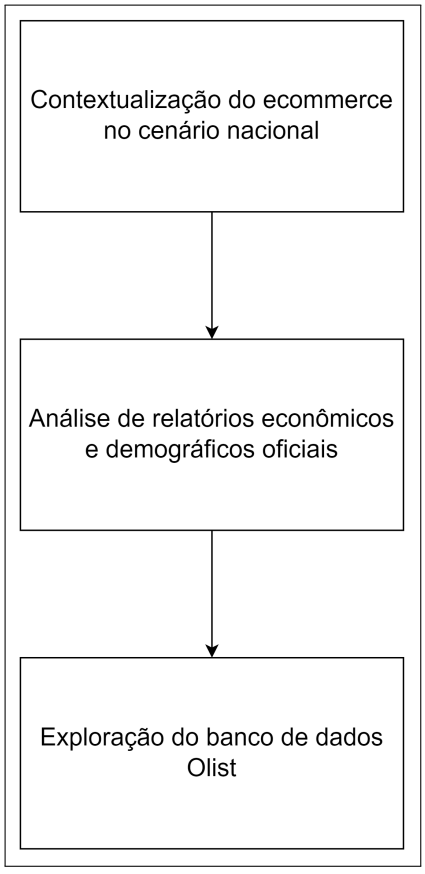


Fonte: (VENCESLAU, 2017)

3 METODOLOGIA

A metodologia abordada se propõe a detalhar a exploração do banco de dados de *ecommerce* ao tempo que contextualiza a realidade social do Brasil, buscando paralelos nos padrões de consumo. O fluxo simplificado pode ser observado na Figura 7. Se trata portanto de uma abordagem descritiva e quantitativa combinada com uma avaliação qualitativa do cenário demográfico nacional. Reis e Reis (2002) conceituam a análise descritiva como a fase inicial do processo de estudo dos dados coletados. Utilizando métodos de Estatística Descritiva para organizar, resumir e descrever os aspectos importantes de um conjunto de características observadas, ou comparar tais características entre dois ou mais conjuntos.

Figura 7 – Fluxograma da metodologia aplicada



Fonte: Produção própria.

3.1 O bando de dados Olist

O banco de dados escolhido se trata de uma coleção de 100 mil pedidos realizados entre os anos de 2016 e 2018 através da plataforma *Olist*. É um banco público e gratuito

disponibilizado através da plataforma *Kaggle* pela própria Olist (2021).

A escolha do banco se deu não somente pela escassez de bancos de dados gratuitos desse tipo, mas também pela sua abrangência e qualidade estrutural e de detalhes. Se tratam de pedidos reais com identificação de empresas e parceiros, mesmo que anonimizados para fins de avaliações dadas por clientes.

3.1.1 Olist

A Olist (2023) se apresenta como uma loja de departamentos dentro dos maiores *marketplaces* do país. Para entender exatamente o que isso significa vale entender o história da empresa e os detalhes do modelo de negócio que ela apresenta. Foi fundada em 2015, é uma empresa de tecnologia do tipo *Software as a Service (SaaS)*, os criadores atuaram em *ecommerces* anteriormente à fundação da empresa, e possuem a missão de transformar o futuro do comércio digital, mas a Olist não é um ecommerce propriamente dito, mas uma plataforma que auxilia lojistas a ampliarem a visibilidade de seus produtos no universo digital.

A referida empresa possui contratos com grandes *marketplaces* como Amazon, Submarino e Mercado Livre. *Marketplaces* são um formato de ecommerce onde outros lojistas, ditos *sellers*, compartilham seus produtos. Nesse formato, múltiplas ofertas de diferentes vendedores podem ser apresentadas no mesmo *site*. No caso da *Olist*, esse conceito é expandido e passa a compor uma modalidade única de negócio, a plataforma faz contratos com vários *marketplaces* e constrói uma *vitrine Olist* nos sites destes, essa vitrine seria a loja de departamentos, dessa forma um *cliente Olist* pode multiplicar a visibilidade de seus produtos em vários *marketplaces* sem a necessidade de efetuar múltiplos contratos e integradores.

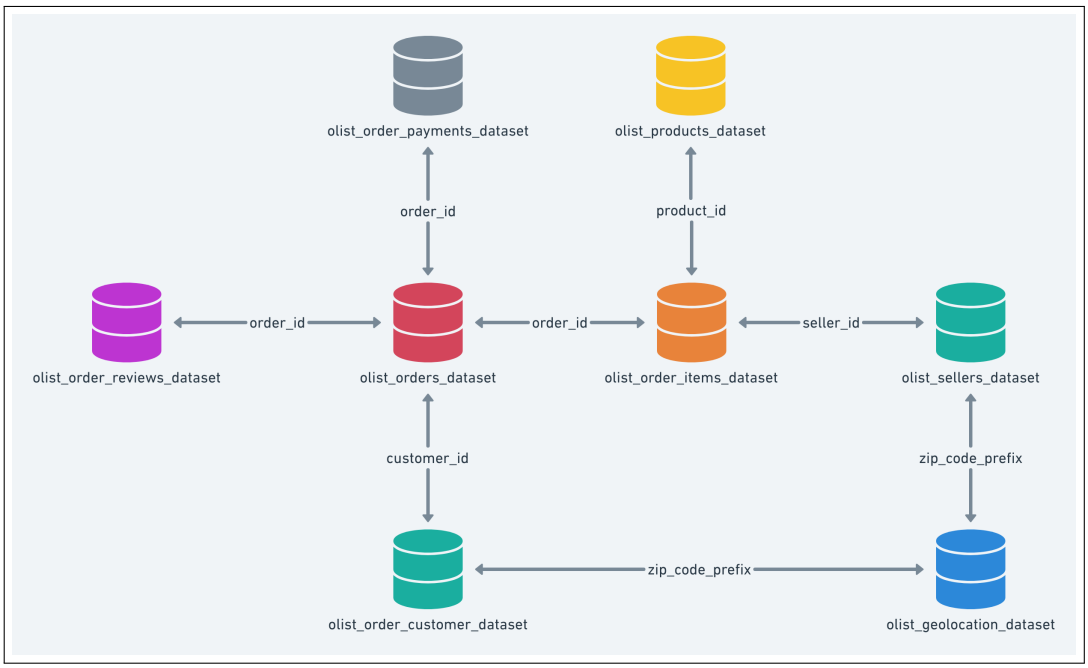
3.1.2 Estrutura do banco de dados

O banco de dados disponibilizado possui dados desde 2016, vista a proximidade com a data de criação da empresa, os dados contam a história da adaptação e expansão da presença da empresa no relativo início de suas atividades.

O referido banco possui 8 tabelas que se conectam através de chaves únicas relacionados conforme a Figura 8. Onde os termos intermediários em setas no *schema* identificam as colunas que servem como chaves. Uma tabela adicional tem por função única a tradução dos termos usados nas categorias de produtos para a língua inglesa, visto que todo o armazenamento dos dados está em inglês por padrão.

São ao todo 9 arquivos com 52 colunas, 99441 pedidos realizados por 96096 consumidores ao longo de 3 anos pelo Brasil. A Figura 8 apresenta uma simplificação no formato do Banco de dados.

Figura 8 – Estrutura de tabelas do Banco de Dados Olist



Fonte: Kaggle (disponibilizado pela Olist)

Na Tabela 1 são identificados os temas dos dados presentes em cada tabela do banco de dados.

Tabela 1 – Conteúdo das tabelas do bando de dados Olist

Título da tabela	**Descrição**
olist_customers_dataset.csv	Consumidores e localização
olist_geolocation_dataset.csv	Latitudes e longitudes
olist_orders_dataset.csv	Status de pedidos
olist_order_items_dataset.csv	Detalhes dos pedidos
olist_order_payments_dataset.csv	Detalhes de pagamento
olist_order_reviews_dataset.csv	Avaliações de pedidos
olist_products_dataset.csv	Detalhes dos produtos
olist_sellers_dataset.csv	Vendedores e localização
product_category_name_translation.csv	Categorias traduzidas para inglês

Fonte: Produção própria.

O formato csv dos arquivos foi processado através da biblioteca *Pandas* em *dataframes* usando a linguagem *python*. Essa abordagem permite a flexibilidade de exploração e plotagem posterior de gráficos através de bibliotecas como a *matplotlib* e *plotly*.

3.1.2.1 Limitações do banco

Visto o recente início de atividades da empresa que disponibilizou os dados, apenas 3 anos incompletos serão cobertos pela coleta. Esse intervalo entre 2016 e 2018 permitirá ter uma boa perspectiva do comportamento de consumidores na totalidade de 2017, mas torna inviável comparação plena com os anos vizinhos. O mesmo vale para análise de janelas de tempo maiores que seriam importantes para descrever a evolução da modalidade de negócios frente eventos importantes como crises econômicas e a pandemia de Covid 19. A coleta de informações foi feita pela própria Olist. Não são divulgadas informações como: qual a ferramenta utilizada para coleta, se foi própria ou terceira, qual critério para seleção desses dados ou se esssa é a totalidade das informações.

Outro ponto a se destacar é identificação parcial de certos eventos, como menção a lojas e parceiros que é substituída, nos comentários de clientes, por casas da série *Game of Thrones*, a fim de preservar a identidade destes. O mesmo vale para nomes e descrições de produtos que são omitidos e mencionados apenas a quantidade de caracteres usados, o que impede estratégias como análise de linguagem natural voltada para conversão de vendas baseada no texto exposto.

Não existem informações de consumidores além das cidades em que vivem, uma vez que dados pessoais ou identificáveis são protegidos pela *Lei Geral de Proteção de Dados (LGPD)*, Brasil (2018). Dessa forma, as análises serão limitadas a entender a distribuição dos consumidores e uma perspectiva macro da realidade de cada região do país.

3.2 Relatórios econômicos oficiais e especializados

Para nortear as análises de desenvolvimento econômico, é feito uso do relatório da *PNAD* Contínua do *IBGE* de 2020, por contemplar os anos presentes no banco de dados em estudo.

Na *PNAD* são observados fatores como rendimento médio distribuído entre as regiões brasileiras e distribuição geral de riqueza entre as parcelas da população.

Outras fontes são incorporadas à analise mediante relevância para o tema e recorte, a exemplo pesquisas conduzidas por empresas especializadas em *ecommerce*, como também instituições de estado voltadas para os segmentos econômicos e de telecomunicações.

A intenção é consolidar informação suficiente para explicar a distribuição dos dados

do banco que ajudem a desenhar uma previsão lógica do futuro do mercado digital no Brasil.

3.3 Procedimentos

Os dados relacionados à contextualização do comércio digital foram consultados através de literatura especializada. Preferindo por artigos recentes que evidenciam a evolução histórica do conceito. Em seguida recorreu-se a autores que complementam o conceito de comércio digital ao definirem subcategorias nele presentes.

Para a descrição da demografia foram usados dados da PNDA contínua do IBGE, tanto no que concence ao relatório final publicado quanto através do sistema SIDRA, onde foram selecionadas tabelas específicas com os dados coletados através da PNAD contínua, e contruídos gráficos.

Quanto à análise de dados, o banco de dados foi obtido através da platafórma Kaggle, em seguida descrito usando o pacote pandas em Python. As tabelas foram carregadas para um script Python em um notebook Jupyter, transformadas em dataframes e agrupadas em uma tabela única usando as chaves primárias disponíveis. Uma vez agrupados, dados foram agrupados baseado na demanda quanto a métrica em questão e a base foi segmentada de acordo com os agrupamentos.

Para plotagem dos gráficos foram usados vide demanda as bibliotecas matplotlib, sobretudo através do pyplot, também plotly e seaborn para as faixas de cor. Ocasionalmente foi utilizado também o Microsoft Excel para validação.

Em linhas gerais, o banco de dados foi agrupado com fins a responder a pergunta levantada, em seguida descrito, validado, e finalmente plotado usando alguma das ferramentas que provesse um visual de mais fácil consumo.

4 RESULTADOS E DISCUSSÕES

Nesse capítulo são apresentados os detalhes da análise exploratória através do detalhamento das informações contidas no banco de dados e os questionamentos gerados a partir desta. Conceituação histórica e contextualização com o cenário econômico da época também são abordados para complementar o entendimento.

4.1 Análise exploratória de dados

O primeiro ponto é entender em detalhes a composição das planilhas que compõem o banco de dados. Para tal, a Tabela 2 apresenta as informações de tipo de variável alocada e quantidade de valores únicos existentes em cada uma das 52 colunas presentes no banco, com exceção de duas que são citadas logo em seguida. Assim, é possível pontuar uma visão global do banco de dados e se preparar para interagir com cada coluna separadamente. A exemplo, se pode aferir que a nota de avaliação possui 5 opções possíveis de acordo com o review_nota na tabela pedido_nota, e essas avalições são numéricas com formato int64.

Tabela 2 – Detalhamento de tabelas

entidades	colunas	tipo	valores únicos
clientes	cliente_id	CHAR(32)	99441,0
clientes	cliente_unico_id	CHAR(32)	96096,0
clientes	cliente_zip_code_prefix	int64	14994,0
clientes	cliente_cidade	VARCHAR(255)	4119,0
clientes	cliente_estado	CHAR(2)	27,0
geolocalizacao	geolocalizacao_zip_code_prefix	int64	19015,0
geolocalizacao	geolocalizacao_lat	float64	717360,0
geolocalizacao	geolocalizacao_lng	float64	717613,0
geolocalizacao	geolocalizacao_cidade	VARCHAR(255)	8011,0
geolocalizacao	geolocalizacao_estado	CHAR(2)	27,0
pedidos	pedido_id	CHAR(32)	99441,0
pedidos	cliente_id	CHAR(32)	99441,0
pedidos	pedido_status	VARCHAR(255)	8,0
pedidos	pedido_compra_timestamp	datatime64[ns]	98875,0
pedidos	pedido_aprovado_at	datatime64[ns]	90734,0

Tabela 2 – Continuado da página anterior

entidades	colunas	tipo	valores únicos
pedidos	pedido_entregue_carrier_data	datatime64[ns]	81019,0
pedidos	pedido_entregue_cliente_data	datatime64[ns]	95665,0
pedidos	pedido_estimado_entrega_data	datatime64[ns]	459,0
pedido_items	pedido_id	CHAR(32)	98666,0
pedido_items	pedido_item_id	int64	21,0
pedido_items	produto_id	CHAR(32)	32951,0
pedido_items	vendedor_id	CHAR(32)	3095,0
pedido_items	envio_limite_data	datatime64[ns]	93318,0
pedido_items	preco	float64	5968,0
pedido_items	frete_valor	float64	6999,0
pedido_pagamentos	pedido_id	CHAR(32)	99440,0
pedido_pagamentos	pagamento_sequencial	int64	29,0
pedido_pagamentos	pagamento_tipo	VARCHAR(255)	5,0
pedido_pagamentos	pagamento_parcelas	int64	24,0
pedido_pagamentos	pagamento_valor	float64	29077,0
pedido_reviews	review_id	CHAR(32)	98410,0
pedido_reviews	pedido_id	CHAR(32)	98673,0
pedido_reviews	review_nota	int64	5,0
pedido_reviews	review_titulo	VARCHAR(255)	4528,0
pedido_reviews	review_mensagem	VARCHAR(255)	36160,0
pedido_reviews	review_data	datatime64[ns]	636,0
pedido_reviews	review_timestamp	datatime64[ns]	98248,0
produtos	produto_id	CHAR(32)	32951,0
produtos	produto_categoria_nome	VARCHAR(255)	74,0
produtos	produto_nome_tamanho	float64	67,0
produtos	produto_description_tamanho	float64	2961,0
produtos	produto_photos_qty	float64	20,0
produtos	produto_peso_g	float64	2205,0
produtos	produto_tamanho_cm	float64	100,0
produtos	produto_altura_cm	float64	103,0

Tabela 2 – Continuado da página anterior

entidades	colunas	tipo	valores únicos
produtos	produto_largura_cm	float64	96,0
vendedores	vendedor_id	CHAR(32)	3095,0
vendedores	vendedor_zip_code_prefix	int64	2246,0
vendedores	vendedor_cidade	VARCHAR(255)	611,0
vendedores	vendedor_estado	CHAR(2)	23,0

Fonte: Produção própria.

A entidade product_category_name_translation foi omitida da Tabela 2 para facilitar a visualização, mas contém as colunas: product_category_name e Product_category_name_english. Ambas do tipo VARCHAR(255). E que possuem a única função de normalizar a linguagem do campo para inglês traduzindo os termos de categorias em português. Se optou por traduzir todos os termos diretamente, a fim de facilitar o entendimento.

Uma forma mais amigável de avaliar essas colunas é através de um *schema* de banco de dados, conforme Figura 9. Nele, é possível visualizar no formato de um diagrama, as conexões entre as tabelas e suas respectivas colunas. São usadas *Primary Key (PK)* e *Foreign Key (FKX)*, sendo as PK aquelas que apresentam valores únicos identificadores em suas tabelas, e as FK aquelas que são chaves externas para conexão entre tabelas.

Indo além, vale a pergunta de quais os tipos mais predominantes de dados ao longo do banco. Isso é respondido através da Tabela 3, onde se verifica a predominância de CHAR(12) e float64.

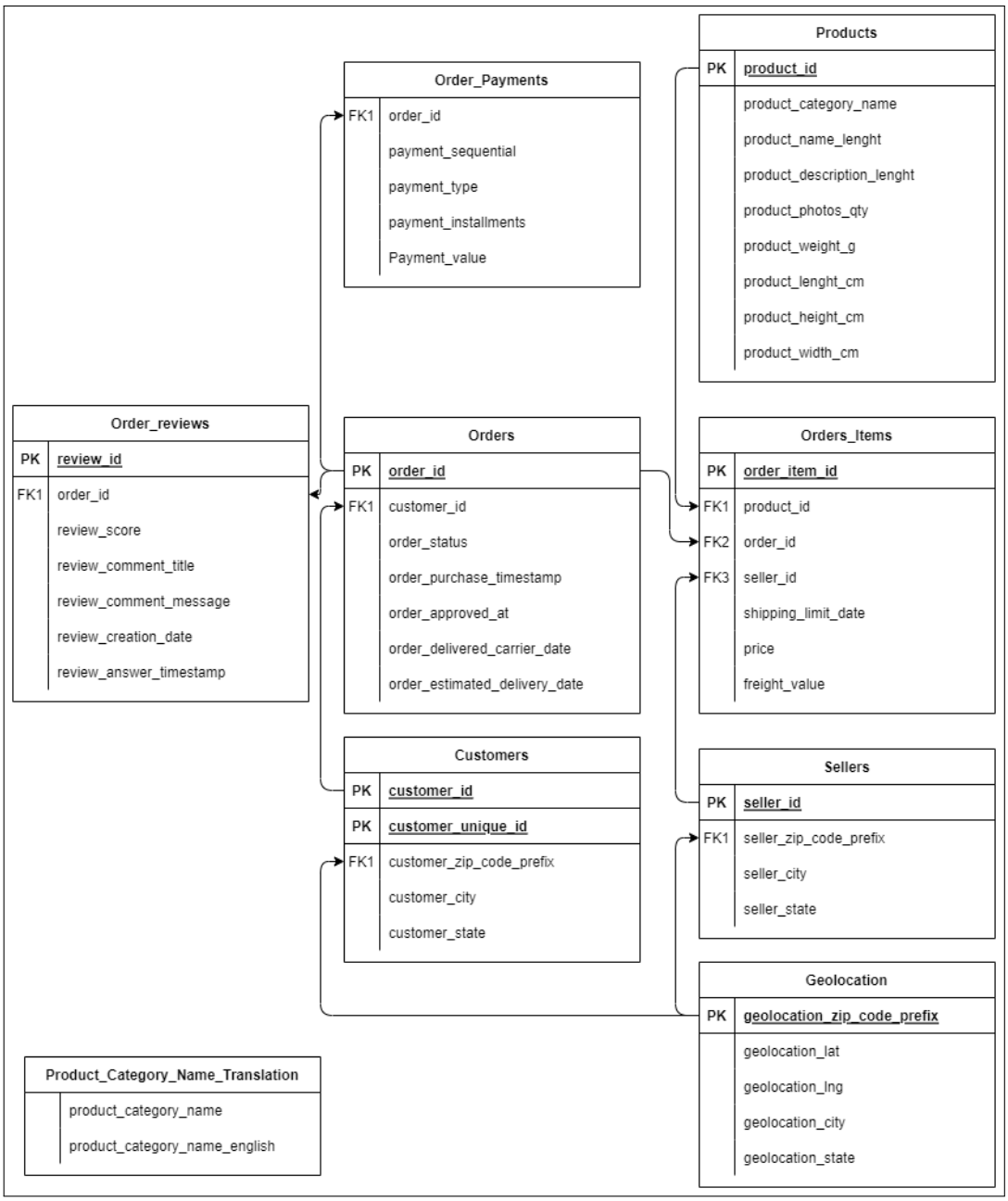
Tabela 3 – Distribuição de tipos de dados

	subtype
CHAR(32)	12
float64	12
VARCHAR(255)	10
datetime64[ns]	8
int64	7
CHAR(2)	3

Fonte: Produção própria.

Para além dos tipos de dados, vale entender o grau de completude desses dados, verificando a quantidade de células vazias onde não houve preenchimento da informação. Isso é

Figura 9 – Schema do Banco de Dados Olist



Fonte: Produção própria

importante na definição de expectativas do que pode ser explorado ao longo das análises. Tal qual o proposto na Tabela 4.

A partir dessa avaliação é verificado que os dados de menor preenchimento são os de avaliações de pedidos, cujo título está ausente em 88,34% dos casos e o comentário ausente em 58% dos casos. Isso indica que mais de metade dos compradores não deram um testemunho escrito sobre a compra, nesse caso uma boa estratégia é seguir com o estudo dos comentários visto sua maior robustez de completude. Mesmo havendo uma grande abstenção, 40977 reviews

Tabela 4 – Relação das colunas com valores nulos

entidades	colunas	linhas vazias	vazias (%)
pedidos	pedido_aprovado_em	160,0	0,16
pedidos	pedido_entregue_entregador_data	1783,0	1,79
pedidos	pedido_entregue_customer_data	2965,0	2,98
pedido_reviews	review_comentario_titulo	87656,0	88,34
pedido_reviews	review_comentario_mensagem	58247,0	58,70
produtos	produto_categoria_nome	610,0	1,85
produtos	produto_nome_tamanho	610,0	1,85
produtos	produto_descricao_tamanho	610,0	1,85
produtos	produto_fotos_qty	610,0	1,85
produtos	produto_peso_g	2,0	0,01
produtos	produto_tamanho_cm	2,0	0,01
produtos	produto_altura_cm	2,0	0,01
produtos	produto_largura_cm	2,0	0,01

Fonte: Produção própria.

ainda é um valor relevante dentro do universo de 99441 pedidos. Para explicitar a presença de nulos, um bom artifício é o uso de cores e nomear as origens, mudando essa tabela para uma construção mais descritiva conforme Tabela 5.

Uma outra questão relevante para considerar possíveis combinações entre essas tabelas, é compreender como elas se comunicam entre si. A Figura 8 ilustra em um diagrama a comunicação entre as tabelas, mas a Tabela 6 referenda essa demonstração através da frequência com a qual as colunas se repetem entre as tabelas. Através da Tabela 5 é possível destacar que por mais que a maior quantidade de nulos esteja nos *reviews* dos pedidos, o *dataset* de produtos destaca-se pela quantidade de colunas com algum valor vazio.

A coluna com repetição mais frequente é a de *order_id* (pedido_id), ela que conecta a maior quantidade de colunas ao longo do banco de dados. Existe ainda nós de comunicação envolvendo as colunas de: *customer_id, product_category_id, seller_id, product_id* (cliente_id, cliente_id, produto_categoria_id, vendedor_id, produto_id). Por existirem poucas tabelas, é cômodo o uso de diagramas como o da Figura 8, mas para bancos massivos com centenas de tabelas, um ranking como esse pode ser muito mais assertivo na busca por conexões.

Seguindo com as análises, são estabelecidas frentes de investigação baseadas na natureza dos dados identificados: categorias de produtos, status dos pedidos, geografia e frete, produtos e pagamento, e avaliação dos clientes.

Tabela 5 – Destaque dos elementos nulos

dataset	n_linhas	n_colunas	null_qty	#null_colunas	null_colunas
olist_customer	99441	5	0	0	
olist_geolocation	1000163	5	0	0	
olist_orders	99441	8	4908	3	pedido_aprovado_em, pedido_dentregue_entregador data, pedido_entregue_cliente date
olist_order_items	112650	7	0	0	
olist_order payments	103886	5	0	0	
olist_order reviews	99224	7	145903	2	review_comentario_titulo, review_comentario_mensagem
olist_products	32951	9	2448	8	product_category_name, product name lenght, produto_descricao tamanho, produto_fotos_qty, produto_peso_g, produto_tamanho_cm, produto_altura_cm, produto_largura_cm
olist_sellers	3095	4	0	0	

Fonte: Produção própria.

Tabela 6 – Repetição de colunas chave entre tabelas

	column_name
order_id	4
customer_id	2
product_category_name	2
seller_id	2
product_id	2

Fonte: Produção própria.

4.1.1 Categorias de produtos

Antes de partir para detalhes do processo de venda, é relevante compreender qual o tipo de produto está sendo considerado no banco de dados. Quais movimentam mais receita e

qual gera nos consumidores maior satisfação. Considerando, todavia, que a satisfação não diz respeito ao produto em si, mas à experiência de compra, que a depender do tipo de produto pode vir a ser mais provável de atender às expectativas de recebimento e uso.

Uma pergunta que pode ser feita é: Qual categoria de item vende mais? Para responder essa pergunta, basta contar a quantidade de order_item_id e agrupa-los em cada uma das categorias. A Figura 10 demonstra a distribuição de quantidade de itens comprados ao longo das diferentes categorias.

Aqui se observa que artigos de cama, mesa e banho são os mais vendidos, na ordem de 12000 itens, seguidos de beleza/saúde e esportes/lazer. Mas será que os mais vendidos também representam o maior valor de venda? A Figura 11 demonstra essa mesma distribuição por categorias, mas sob a ótica do preço de cada produto vendido.

Aqui já é possível observar que beleza/saúde assume a lideraça na receita gerada pela venda dos itens, e cama mesa e banho partem para a terceira colocação. Vale destacar a presença nesse top3 de relógio/presentes que estava na sétima posição quanto à quantidade de vendas.

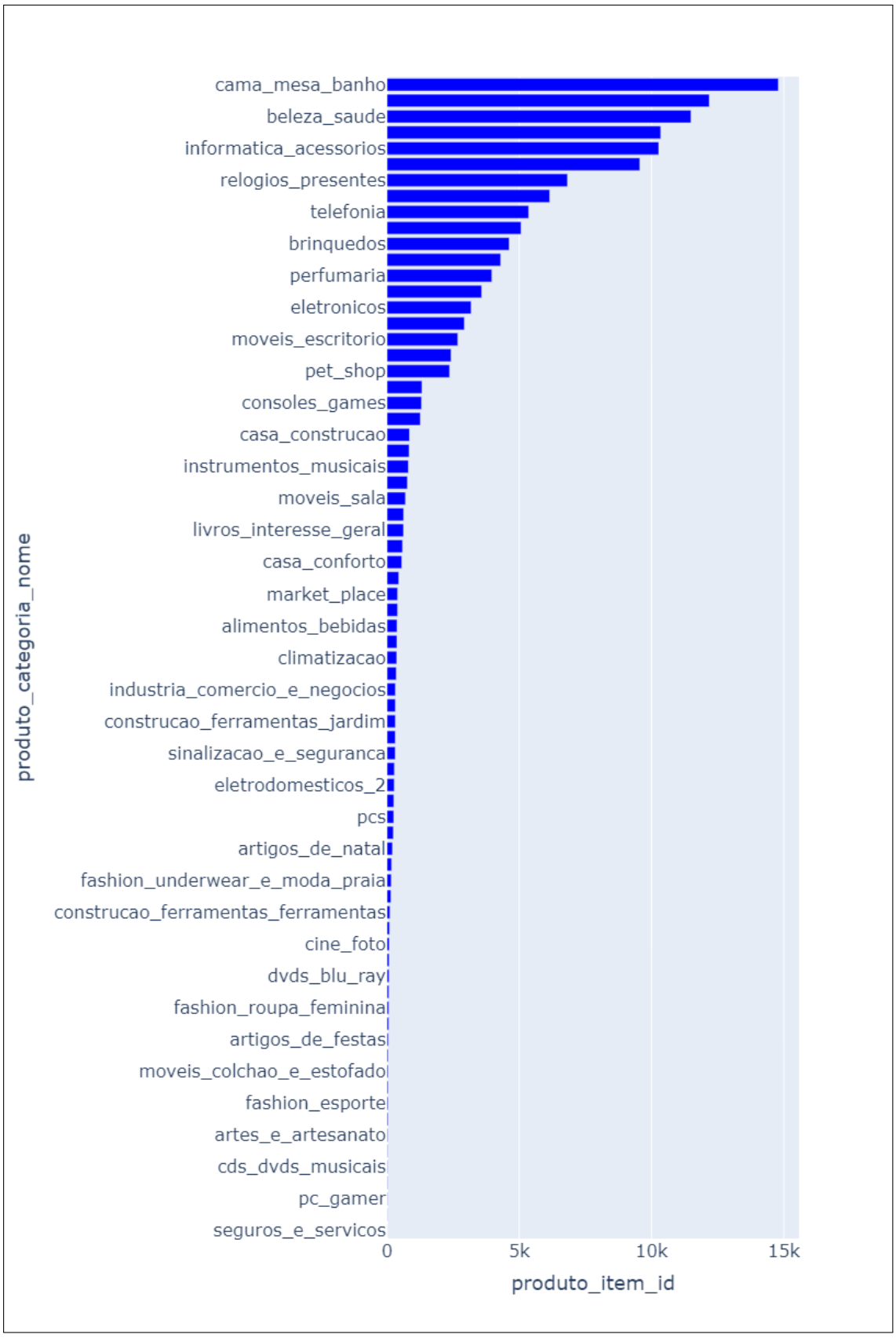
Esse comportamento ilustra o efeito do ticket médio nas compras. Determinados produtos, mesmo vendendo menos em quantidade, podem impactar mais na receita gerada, visto seu valor médio maior. Dito isso, é válido ponderar o maior risco associado à venda de produtos de ticket médio mais alto, e o fato de estarem segmentados a um público de uma classe social menos numerosa.

É possível questionar ainda qual categoria de produtos apresenta a melhor nota dada por consumidores, mantendo a observação que essa satisfação é medida por um conjunto de fatores que podem estar associados diretamente ao produto, mas refletem toda a experiência de compra. Por exemplo, é possível uma *review* com nota baixa de um produto excelente caso a experiência de entrega tenha sido ruim. A Figura 12 mostra a distribuição de notas dadas por categoria de produto. Essa imagem destaca uma variação leve para a maioria dos ranqueados, destacando cds e livros com melhor avaliação; fraudas e seguros/serviços como pior avaliados pelos clientes.

4.1.2 Status dos pedidos

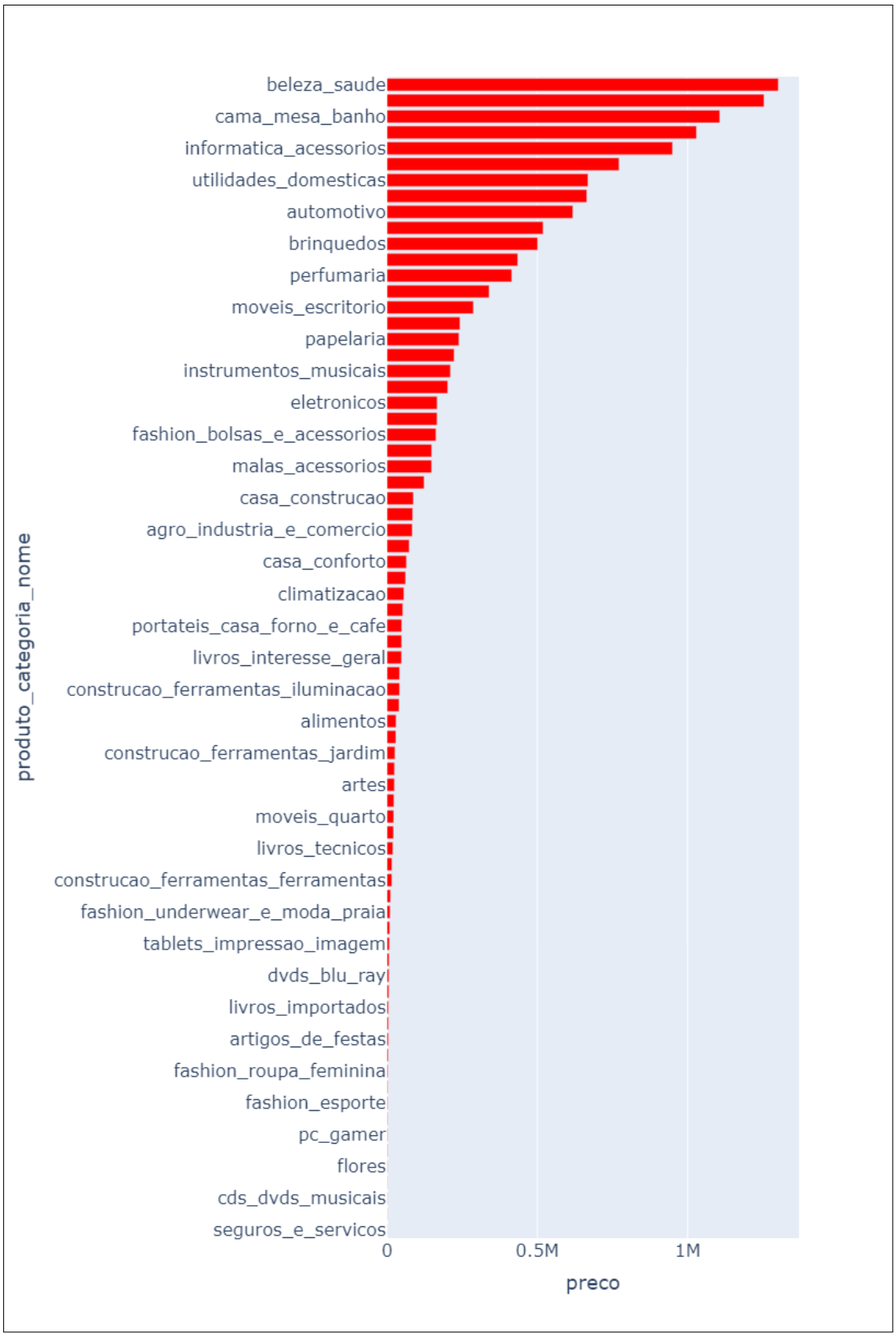
Baseado no status é possível aferir qual o nível de completude dos pedidos, entender a taxa de cancelamento e aprofundar o entendimento de qual cenário está sendo descrito no

Figura 10 – Quantidade de itens comprados por categoria de produtos



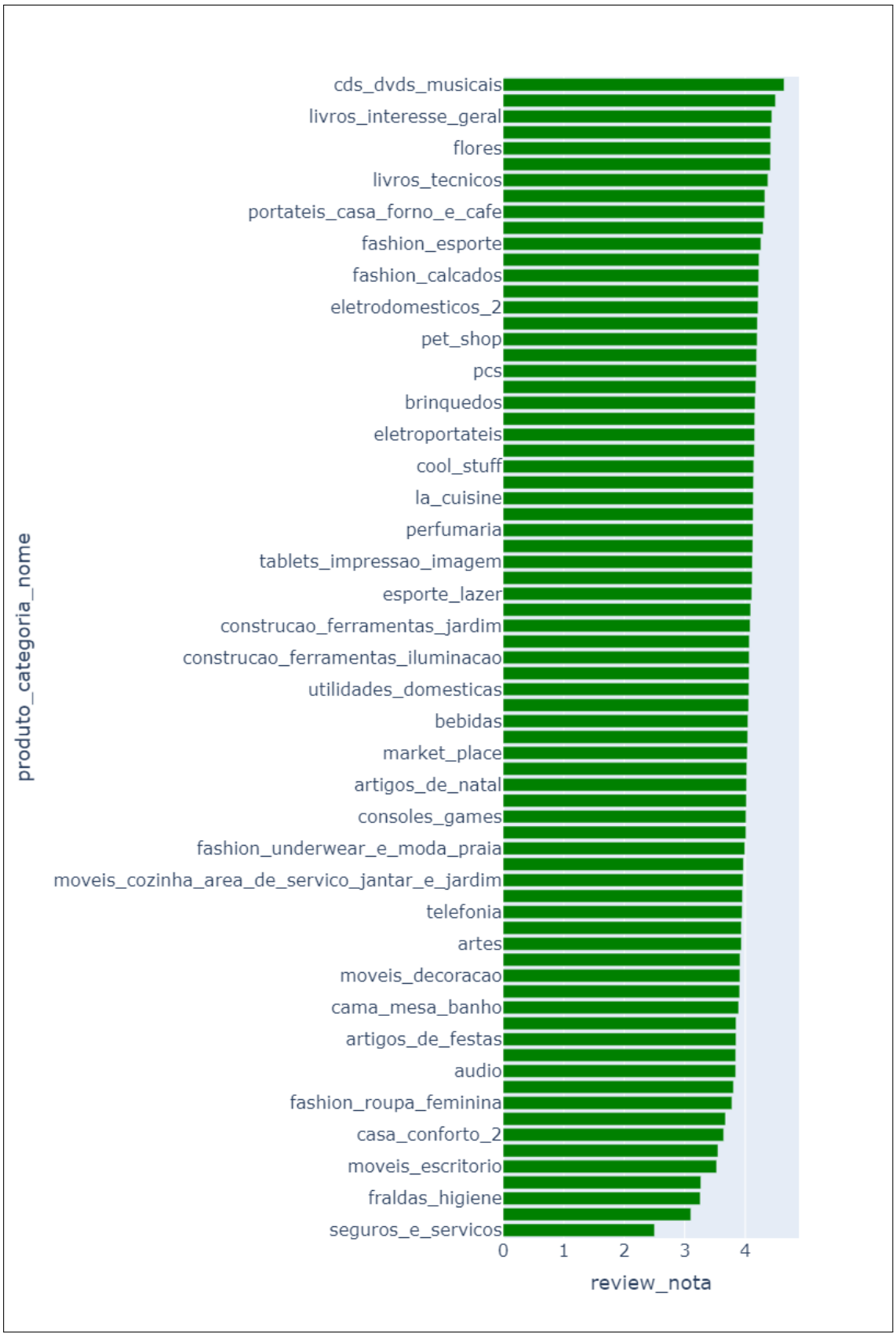
Fonte: Produção Própria

Figura 11 – Receita de itens comprados por categoria de produtos



Fonte: Produção Própria

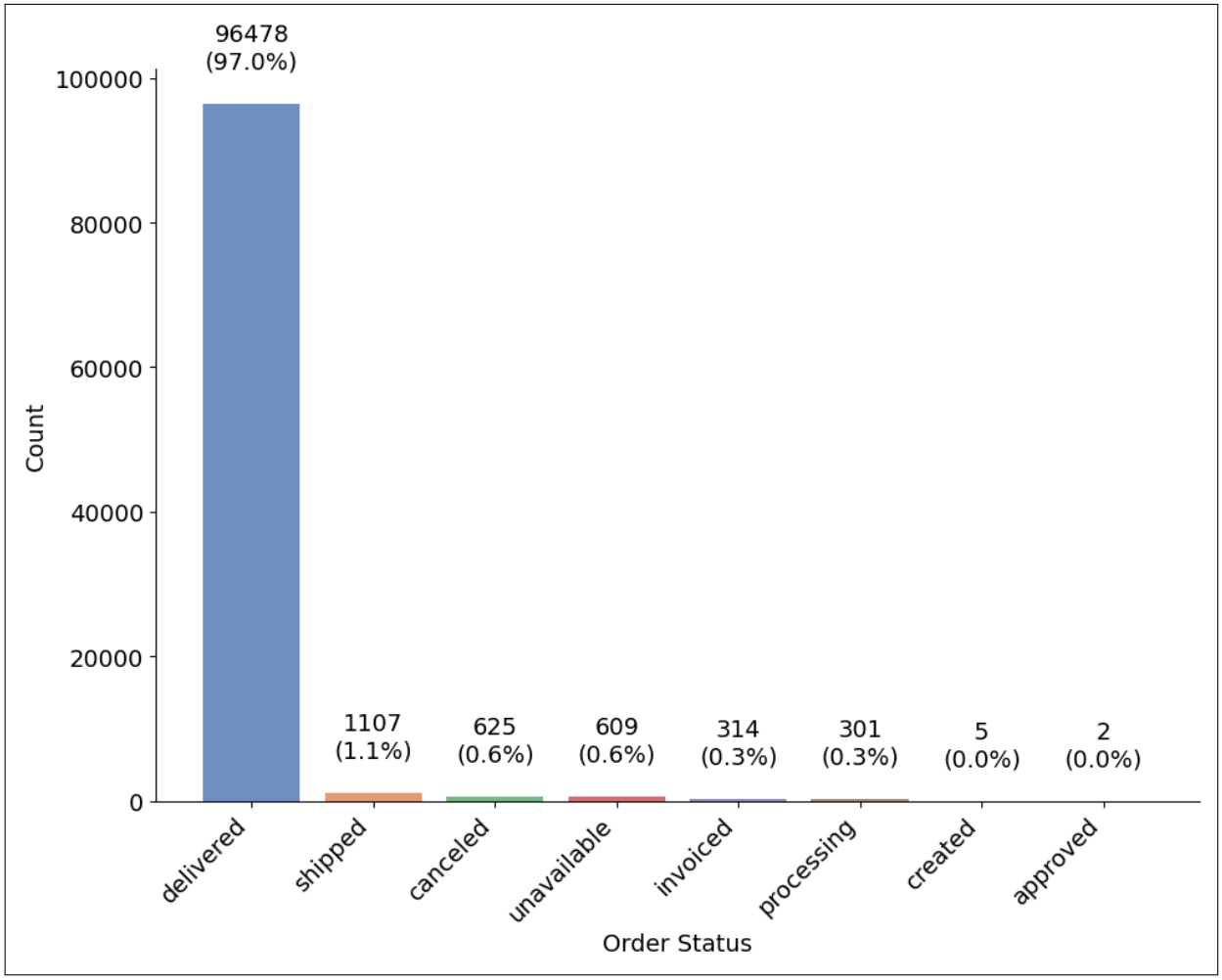
Figura 12 – Nota dada através de reviews por categoria de produtos



Fonte: Produção Própria

banco de dados quanto à realidade das lojas. A Figura 13 apresenta a distribuição de pedidos baseado em seus *status*.

Figura 13 – Distibuição de pedidos por status



Fonte: Produção própria.

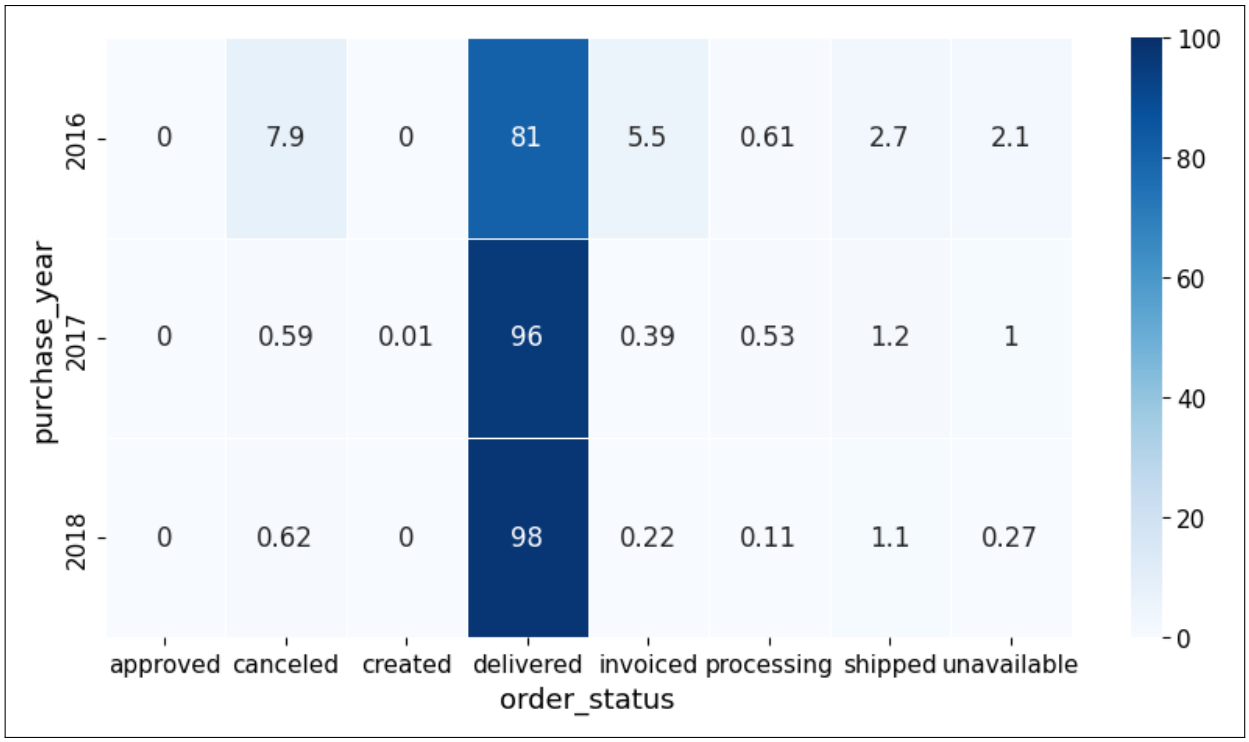
É fácil perceber que a grande maioria dos pedidos foi de fato entregue. 97% dos pedidos efetuados chegaram aos consumidores. Isso é um elemento importante em relação à confiabilidade do mercado digital. Por muito tempo houve a desconfiança de realização da compra por preocupação com fraude. A intensificação das leis de regulamentação, inclusão de instrumentos de proteção ao consumidor e fortalecimento da presença das marcas no *marketing* digital ajudaram a reduzir essa desconfiança. Os dados mostram que há uma finalização no processo do produto chegar ao cliente, mas vale detalhar esse caminho para entender onde está o cliente, quanto tempo leva e qual a satisfação do cliente ao receber o produto.

Os outros *status* são bem menos significativos em números, e por estarem em posição intermediária, é possivel se tratar de pedidos que ainda estavam ocorrendo durante a geração

desse banco de dados, e posteriormente poderiam avançar em classificação.

Essa baixa proporção pode também servir de informações interessantes, mas antes de comentar cada cenário, é possível extrair mais informações distribuindo essas quantidades de forma temporal. A Figura 14 apresenta essa distribuição por ano em que a compra foi realizada.

Figura 14 – Status por ano de compra realizada



Fonte: Produção própria.

Com essa distribuição percentual no volume de pedidos por ano, é mais fácil visualizar evolução e verdadeiros problemas a serem trabalhados. Na versão anterior, por se tratar de valores acumulados ao longo de todos os anos no banco, pedidos cancelados há muito tempo poderiam levar a falsas conclusões quanto à representação desse tipo de *status* na distribuição atual de pedidos.

Conceituando cada *status*:

- Created (criado)

  Quando criado o pedido entra como *created* até que seu pagamento seja aprovado. Existem registros superiores à proporção de 0.01% dos pedidos de 2017 que se mantém nesse *status*, o que pode indicar que os pedidos foram feitos mas não foram aprovados. Vale ressaltar que a aprovação dos pedidos passa pela entidade de *gateway* de pagamentos, que se trata de uma empresa responsável pelo processamento do pagamento. Pedidos cujo

pagamento não é aprovado, uma vez criados devem passar para cancelados após um certo tempo, porém é possível que algum erro técnico tenha ocorrido com o *ERP* responsável pela gestão desses pedidos que culminou na não atualização do status.

- Approved (aprovado)

  Nesse *status*, o *gateway* de pagamento notifica o *ERP* que o pagamento aconteceu e o pedido pode ser encaminhado para a próxima etapa do fluxo.
- Processing (processando)

  Aqui o pedido entra no processamento para envio e aguarda liberação de elementos como faturamento e gestão logística
- Unavailable (indisponível)

  Caso tenha havido uma má gestão de estoque, é possível que um item tenha sido vendido mas não exista em estoque para ser enviado, nesses casos o pedido pode passar para *unavailable*. Mas vale ressaltar que um dos mecanismos de defesa do consumidor pode exigir do lojista a entrega do produto anunciado, similar ou superior, mesmo que seja necessária aquisição por um valor não lucrativo por parte do vendedor, apenas para cumprir a propaganda feita, conforme Brasil (1990) Código de Defesa do Consumidor, Artigo 35.
- Invoiced (faturado)

  Esse é o momento que o item é faturado e está disponível para envio.
- Canceled

  Durante todas as etapas anteriores, o pedido está passível de ser cancelado por inciativa do cliente ou lojista, considerando as consequências citadas.
- Shipped

  Nesse momento, o pedido é entregue à transportadora responsável e parte para o endereço do consumidor.
- Delivered

  Momento em que o pedido é entregue ao consumidor final, considerando elementos como confirmação de recebimento.

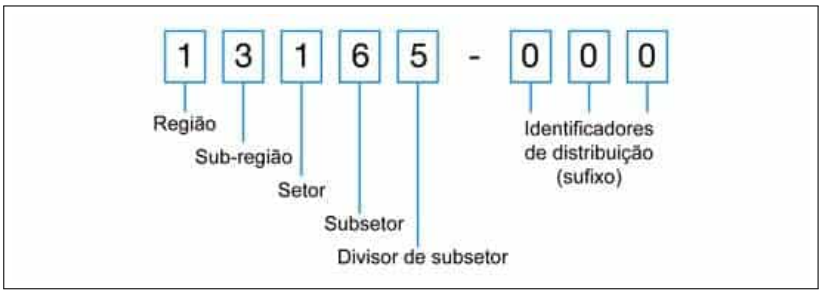
4.1.3 Análise geográfica

Vista a característica continental do território brasileiro, convém verificar como estão distribuídos os dados geográficos do comércio digital presentes no banco de dados em análise. De início, é importante compreender como está composto o banco de dados geográfico.

A tabela *Geolocalization* possui latitude, longitude, nome da cidade e nome do estado. Esses dados se comunicam com as tabelas de clientes e de vendedores através da chave zip-code-prefix, ou *Código de Endereçamento Postal (CEP)* no caso do Brasil, código responsável por identificar localidades ao nível de logradouros e ruas.

O site dos Correios (2023) divide o *CEP* para cada número como correspondendo a uma Região, Sub-região, Setor, Subsetor, Divisor de Subsetor e Identificadores de Distribuição. A Figura 15 ilustra essa regra em uma imagem de exemplo.

Figura 15 – Identificação de cada dígito do CEP



Fonte: Correios

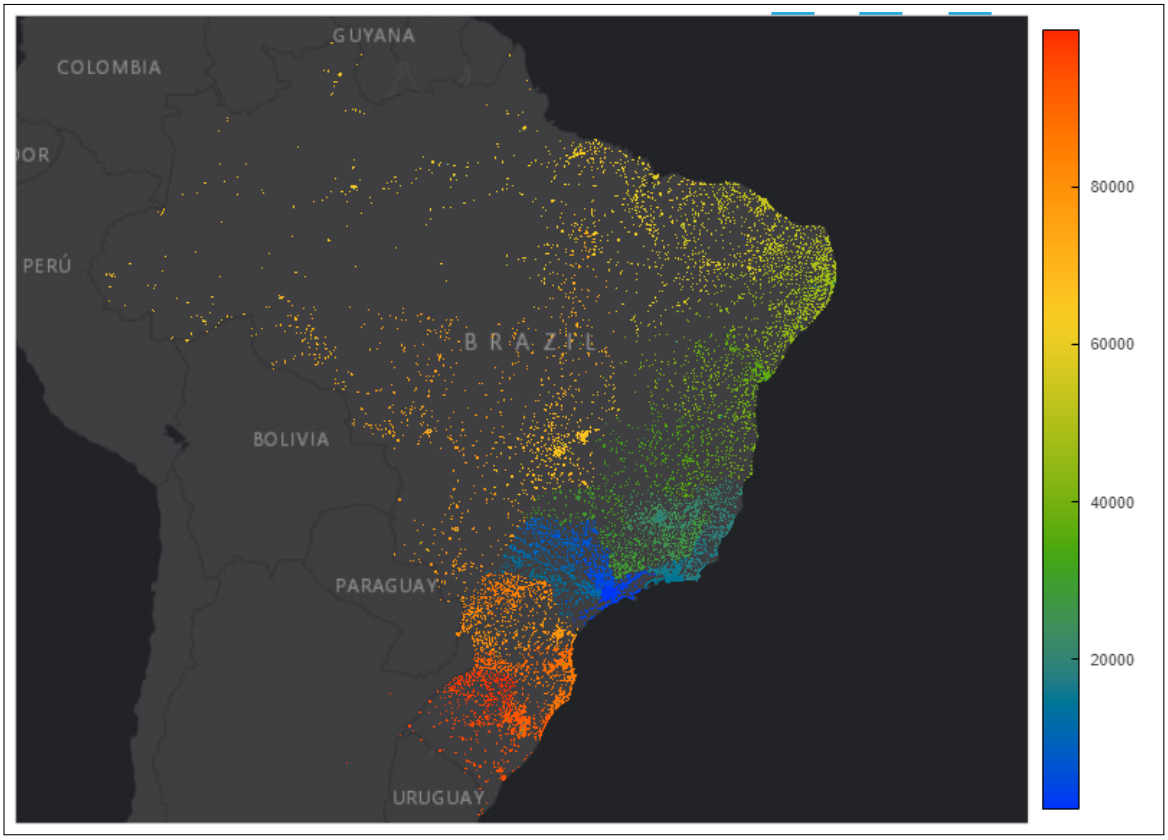
Para o escopo desse trabalho, é razoável abordar uma perspectiva a nível de estado, para evitar demasiada granularidade de informações mas conseguir evidenciar as diferenças no cenário nacional das compras digitais. Destaca-se, ainda, que os estados possuem políticas comerciais independentes e cobram de forma diferentes tributos, como para a expedição de Notas Fiscais, a exigência de inscrição estadual para transações *B2B*, entre outros aspectos. O que reforça a abordagem por estados.

Dito isso, convém entender a distribuição das cidades dentro dos estados, que variam e podem ser um fator importante durante as análises mesmo a nível de estado. Para tal, a fim de ilustrar essa distribuição vide Figura 16, que compreende todas as cidades brasileiras distribuídas no território nacional apenas nas localizações onde o *CEP* é reconhecido como existente.

É possível nitidamente perceber a diferença de densidade entre regiões como a Sul e Sudeste com alta densidade populacional, seguidas da região Nordeste, Centro-Oeste e finalmente a Norte. Se destacam também as capitais, onde a densidade de *CEP* é muito mais intensa. A escala de cores da imagem representa a variação da faixa de CEP ao longo do território.

Para comparação, segue como se distribuem os CEPs em uma cidade como Granjeiro - CE (Figura 17) com cerca de 4 mil habitantes, Teresina - PI (Figura 18) com cerca de 800 mil habitantes e São Paulo - SP (Figura 19) com cerca de 12.33 milhões de habitantes.

Figura 16 – Distribuição de CEPs no território brasileiro



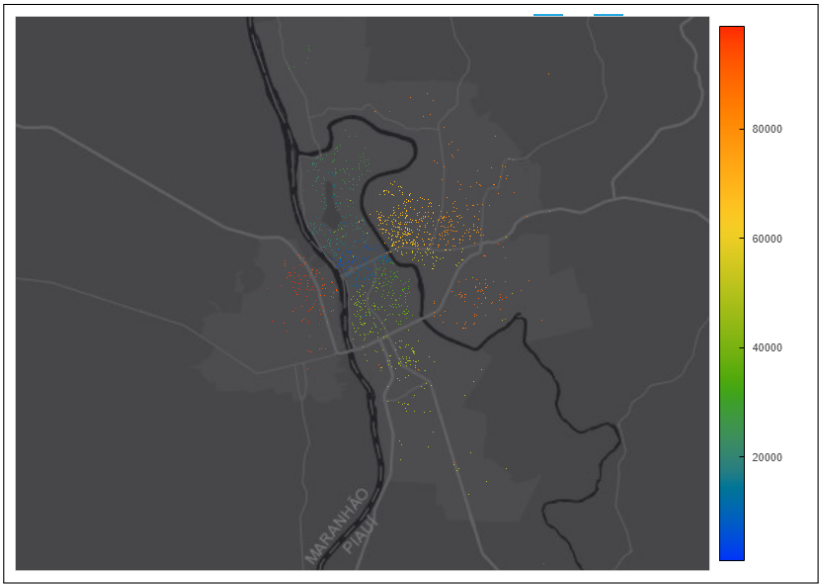
Fonte: Produção Própria

Figura 17 – Distribuição de CEPs na cidade de Granjeiro - CE

Fonte: Produção Própria

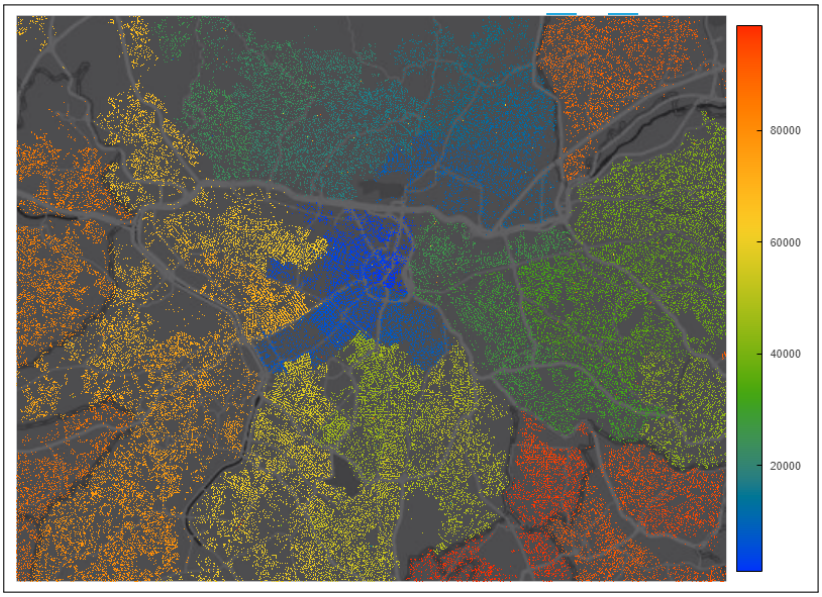
Essa distinção entre as localidades demonstra um pouco da pluralidade que será explorada ao longo da análise geográfica. Quando apropriado, será usado a média do estado para

Figura 18 – Distribuição de CEPs na cidade de Teresina - PI



Fonte: Produção Própria

Figura 19 – Distribuição de CEPs na cidade de São Paulo - SP



Fonte: Produção Própria

normalizar os dados e evitar *outliers*.

Antes de partir para uma análise gráfica, é conveniente construir uma tabela com medidas de interesse para consulta e que seja capaz de ser usada como base para produzir os gráficos. A Tabela 7 traz alguns dos dados que serão plotados referentes a quantidade de pedidos e preço.

Tabela 7 – Dados Geográficos - Pedidos e Preço

estado	pedidos	pedidos_por_cliente	preco	preco_medio
AC	95	1,0	17059,44	179,57
AL	464	1,0	83314,32	179,56
AM	173	1,0	22865,26	132,17
AP	84	1,0	13654,30	162,55
BA	4091	1,0	543243,99	132,79
CE	1565	1,0	240095,13	153,42
DF	2516	1,0	315122,29	125,25
ES	2360	1,0	284771,30	120,67
GO	2466	1,0	313198,27	127,01
MA	856	1,0	123077,67	143,78
MG	13819	1,0	1645847,27	119,10
MS	862	1,0	121358,47	140,79
MT	1132	1,0	171109,74	151,16
PA	1129	1,0	184889,93	163,76
PB	644	1,0	123816,24	192,26
PE	1906	1,0	272271,97	142,85
PI	576	1,0	92227,70	160,12
PR	6043	1,0	708794,22	117,29
RJ	15518	1,0	1921752,53	123,84
RN	574	1,0	94607,86	164,82
RO	292	1,0	46964,03	160,84
RR	52	1,0	7829,43	150,57
RS	6573	1,0	791716,84	120,45
SC	4345	1,0	539896,20	124,26
SE	403	1,0	60954,60	151,25
SP	50265	1,0	5477008,74	108,96
TO	340	1,0	56251,91	165,45

Fonte: Produção Própria

Já a Tabela 8 traz alguns dos dados plotados em referência à frete e prazo de entrega.

Tabela 8 – Dados Geográficos - Entrega

estado	frete	frete_médio	fração_frete	entrega	atrasados
AC	3802,06	40,02	0,22	16,84	0,03
AL	16467,38	35,49	0,20	20,51	0,23
AM	5683,78	32,85	0,25	23,26	0,04
AP	2863,09	34,08	0,21	24,05	0,05
BA	106976,39	26,15	0,20	15,37	0,13
CE	50504,46	32,27	0,21	17,02	0,14
DF	52627,86	20,92	0,17	9,21	0,07
ES	51667,24	21,89	0,18	11,78	0,12
GO	55799,34	22,63	0,18	11,82	0,08
MA	32432,59	37,89	0,26	17,34	0,19
MG	282723,82	20,46	0,17	8,24	0,05
MS	19988,45	23,19	0,16	11,85	0,10
MT	32659,23	28,85	0,19	14,37	0,06
PA	40034,25	35,46	0,22	19,81	0,12
PB	27666,55	42,96	0,22	16,50	0,11
PE	62195,40	32,63	0,23	14,51	0,10
PI	22501,86	39,07	0,24	15,64	0,15
PR	123265,18	20,40	0,17	8,21	0,05
RJ	325375,79	20,97	0,17	11,38	0,12
RN	20139,45	35,09	0,21	15,42	0,11
RO	11717,47	40,13	0,25	16,39	0,04
RR	2235,19	42,98	0,29	24,15	0,10
RS	142569,91	21,69	0,18	11,46	0,07
SC	92617,94	21,32	0,17	11,10	0,09
SE	14541,29	36,08	0,24	17,32	0,15
SP	757525,08	15,07	0,14	5,07	0,05
TO	13450,60	39,56	0,24	13,14	0,11

Fonte: Produção Própria

4.1.3.1 Pedidos

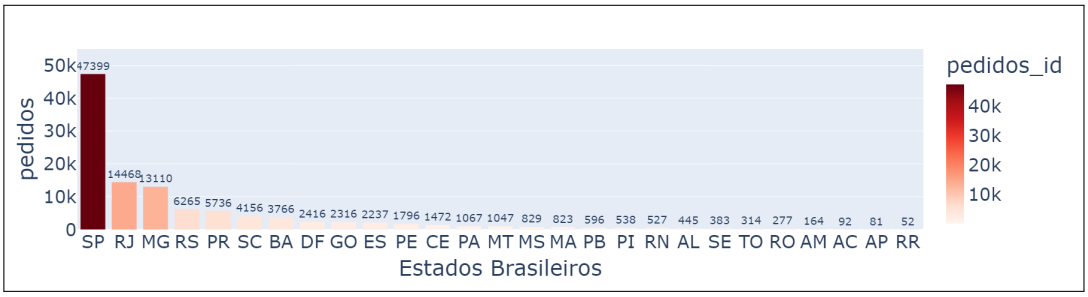
É possível visualizar de uma forma mais direta a participação de cada estado na quantidade total de pedidos. A Figura 20 apresenta um ranking dessa participação em barras, seguida da Figura 21 que apresenta essa informação em mapa.

Tabela 9 – Métricas de pedidos por estado

	pedidos_id
contagem	27,00
média	4161,93
std	9385,89
min	52,00
25%	414,00
50%	1047,00
75%	3091,00
max	47399,00

Fonte: Produção Própria

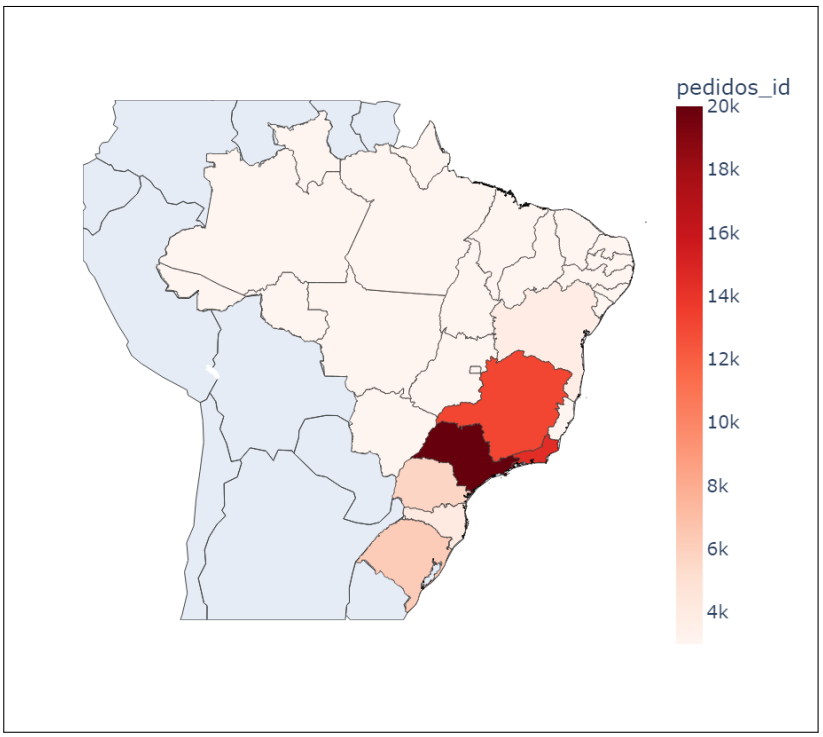
Figura 20 – Distribuição de pedidos por estados do Brasil em ranking



Fonte: Produção Própria

A quantidade bem maior de pedidos na região sudeste, a princípio é coerente por se tratar de uma área com uma infraestrutura urbana bem mais consolidada, o que inclui um acesso facilitado a serviços digitais e estratégias de entrega. Dito isso, vale o destaque para a quantidade de pedidos em São Paulo, que conta com 3,3 vezes mais pedidos que o Rio de Janeiro, segundo apontado; e 911,5 vezes mais que o último, Roraima. De acordo com a Tabela 9, é observado um desvio padrão de 9385,89 e apenas 5 estados estariam acima da média. Cabe, porém, validar se normalizado pela quantidade de consumidores, os pedidos são representativos e consistentes

Figura 21 – Distribuição de pedidos por estados do Brasil em mapa



Fonte: Produção Própria

nas demais regiões. As Figuras 22 e 23 trazem a relação quantidade de pedidos por cliente nos estados.

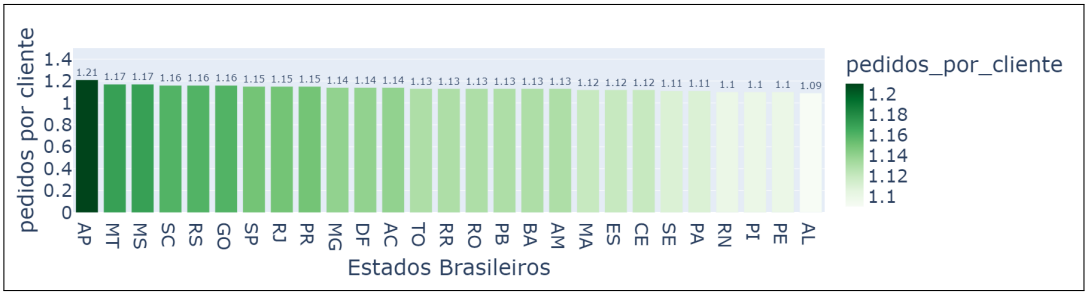
Tabela 10 – Métricas de pedidos por consumidor por estado

	pedidos por cliente
contagem	27,00
média	1,14
std	0,03
min	1,09
25%	1,12
50%	1,13
75%	1,15
max	1,21

Fonte: Produção Própria

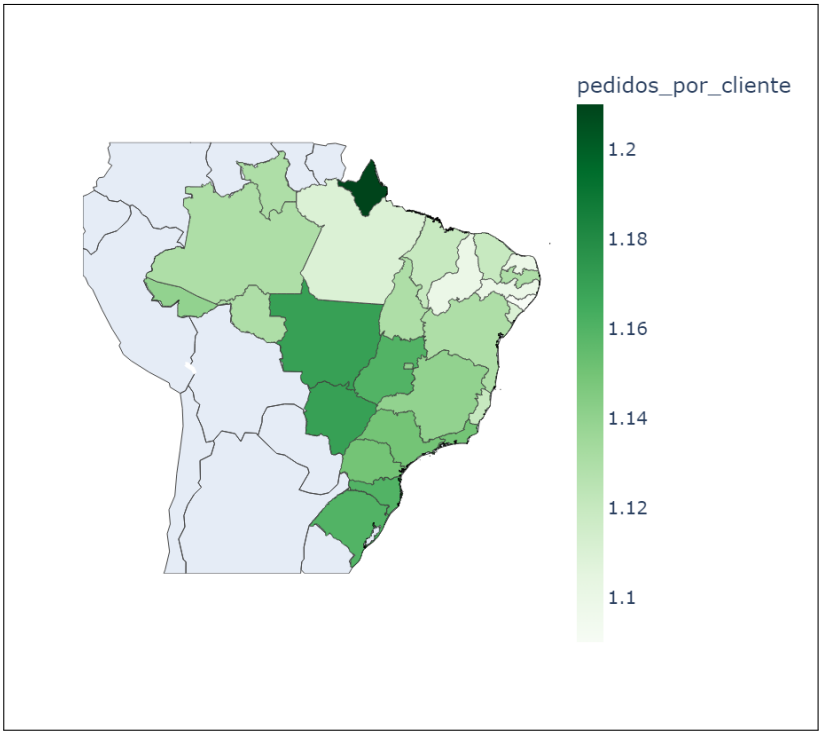
Nessa distribuição de pedidos por clientes, é razoável verificar que mesmo havendo concentração de pedidos na região sudeste, o banco de dados não apresenta o mesmo destaque de cliente recorrentes nessa região. Tendo o Amapá como estado com mais expressividade de recorrência e um desvio padrão de apenas 0,03 na distribuição como apontado na Tabela 10. Se pode destacar ainda a região Centro-Oeste com maior recorrência seguido do conjunto

Figura 22 – Quantidade de pedidos por consumidor por estados do Brasil em ranking



Fonte: Produção Própria

Figura 23 – Quantidade de pedidos por consumidor por estados do Brasil em mapa



Fonte: Produção Própria

Sul-Sudeste, e uma proximidade do Nordeste-Norte ao pedido único por consumidor. Os clientes em média não fazem mais que um pedido na plataforma. Sendo não obstante inferir que havendo mais consumidores adeptos ao mercado digital nos demais estados, a representação desses em pedidos seria mais expressiva.

Partindo dos dados apresentados, é possível aferir a participação dos estados na receita total movimentada pelo comércio digital no contexto delimitado pelo bando de dados. A Figura 24 apresenta essa distribuição em *rank*, e a Figura 25 apresenta a distribuição em mapa.

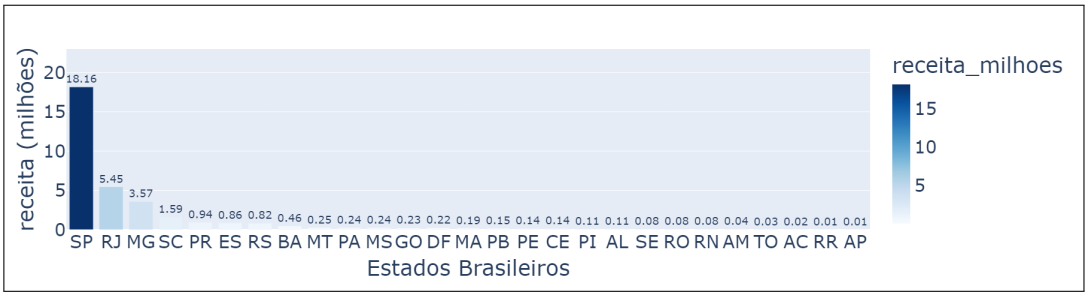
Essa grande concentração de receita na região sudeste já era esperada em função da concentração de consumidores nessa região, vale o destaque novamente para São Paulo com

Tabela 11 – Métricas de receita por estado

	receita
contagem	27,00
média	1,27
std	3,59
min	0,01
25%	0,08
50%	0,19
75%	0,64
max	18,16

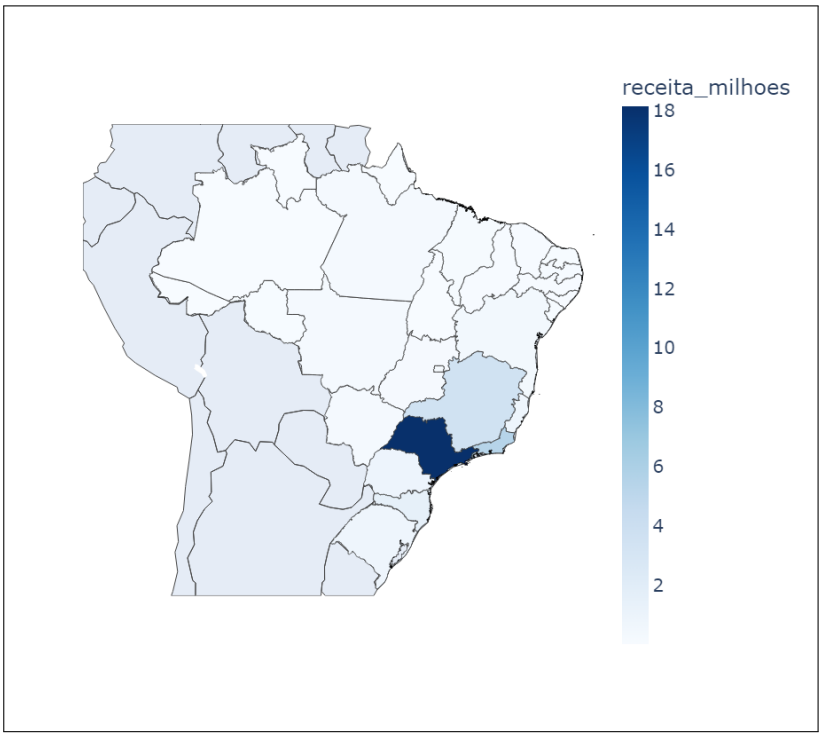
Fonte: Produção Própria

Figura 24 – Distribuição de receita por estados do Brasil em ranking



Fonte: Produção Própria

Figura 25 – Distribuição de receita por estados do Brasil em mapa



Fonte: Produção Própria

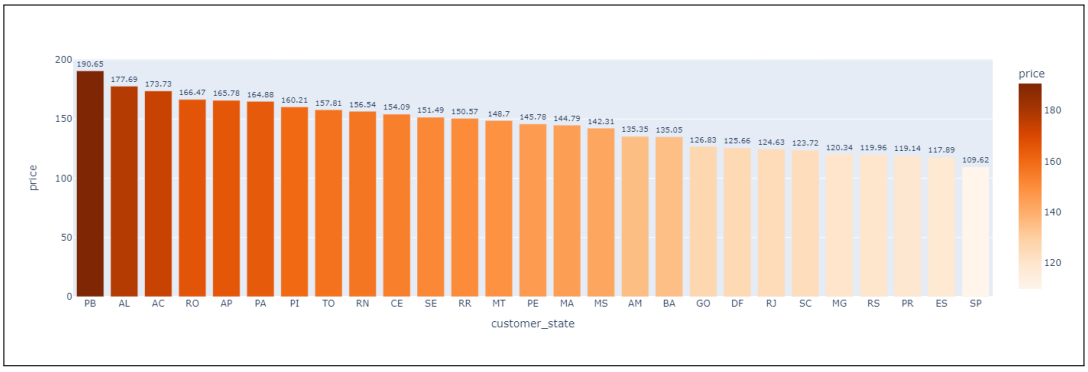
uma receita 14,3 vezes maior que a média, dessa vez com apenas 4 dos 27 estados acima dessa média. Mas é possível usar uma lógica semelhante à de verificação de pedidos por consumidor e normalizar através de uma métrica chamada *Average Order Value (AOV)*, ou *ticket* médio. O *ticket* médio representa a média de receita por venda e é calculado dividindo a receita pela quantidade de pedidos feitos. A Figura 26 apresenta o ranking de *AOV* por estado e a Figura 27 apresenta a mesma distribuição em mapa.

Tabela 12 – Métricas de AOV por estado

	preço médio
contagem	27,00
média	144,80
std	21,15
min	109,62
25%	125,15
50%	145,78
75%	159,01
max	190,65

Fonte: Produção Própria

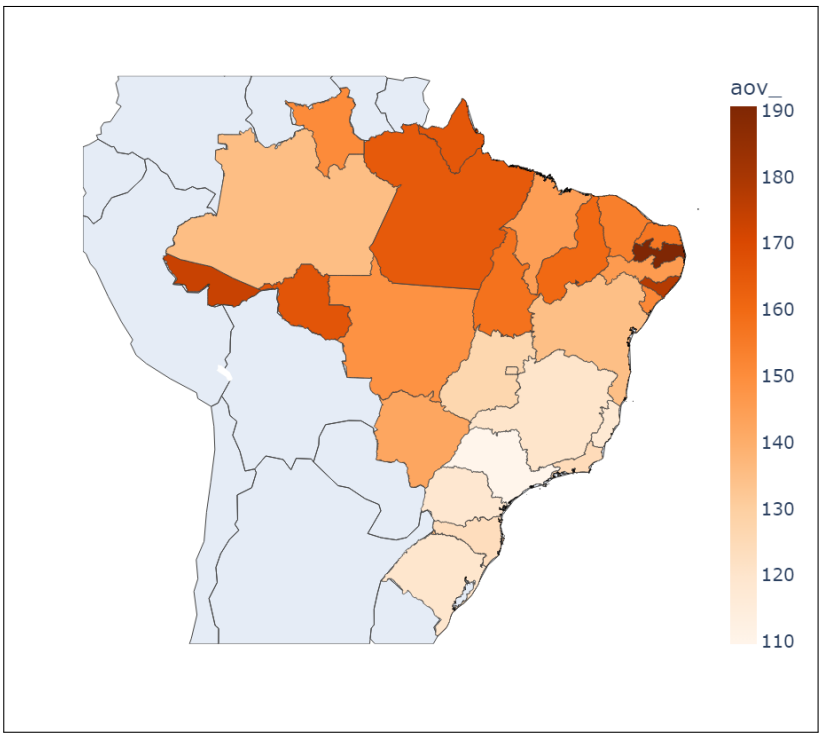
Figura 26 – Distribuição de AOV por estados do Brasil em ranking



Fonte: Produção Própria

A distribuição de ticket médio surpreende ao verificar que mesmo em estados que não correspondem aos centros de consumo previamente identificados, o valor ganha destaque. A Paraíba se apresenta como maior *AOV* seguida de Alagoas e Acre, enquanto São Paulo fica com o menor ticket médio. Há de se considerar, claro, que é possível que haja nessas regiões um elemento cultural ou de natureza outra que se atraia por produto de uma categoria mais valorizada. Todavia, uma pergunta que pode surgir é: visto que os valores médios dos itens é

Figura 27 – Distribuição de AOV por estados do Brasil em mapa



Fonte: Produção Própria

bem distribuído mesmo em regiões mais distantes, como funciona a entrega para esses estados? Em específico: qual o frete e o tempo de recebimento de produtos?

4.1.3.2 Entregas

O frete, ou taxa de entrega, pode variar por diversos fatores. Localidades mais distantes e menos populosas tendem a apresentar valores de frete maiores, uma vez que entregas menos frequentes para essas regiões aumentam as chances de acarretar maiores custos relativos a deslocamento para o entregador do que entregas em grandes centros. Outro ponto importante é a infraestrutura urbana e de acesso a localidades mais afastadas, que pressupõem uma maior dificuldade de realizar a entrega. Esses fatores tendem a influenciar tanto no valor do frete quanto no tempo gasto para a entrega.

Iniciando essa abordagem, é razoável verificar qual a distribuição de estados quanto a seu valor de frete médio. As Figuras 28 e 29 apresentam essa distribuição.

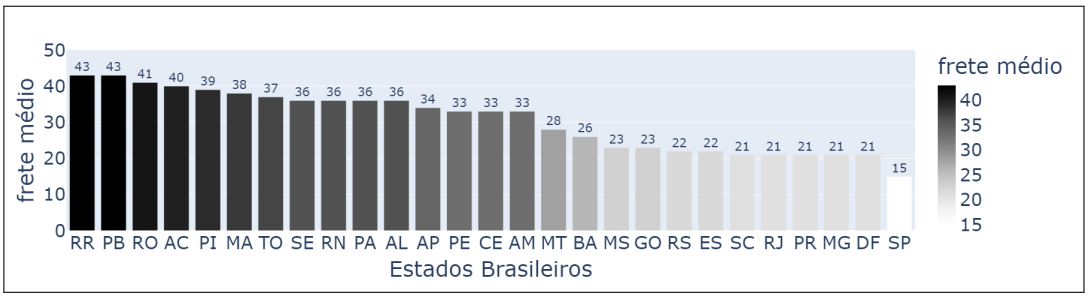
O primeiro contraste pode ser observado na forma como as regiões antes apontadas como menos populosas de consumidores possuem um valor médio de frete mais elevado. Haja vista a exclusiva presença acima da média para estados do Norte e Nordeste, com excessão da Bahia, vide Figuras 28, 29 e Tabela 13. Esse elemento somado ao fato do preço médio praticado

Tabela 13 – Métricas de frete por estado

	valor de frete médio
contagem	27,00
média	30,45
std	8,35
min	15,15
25%	21,84
50%	32,92
75%	36,75
max	42,98

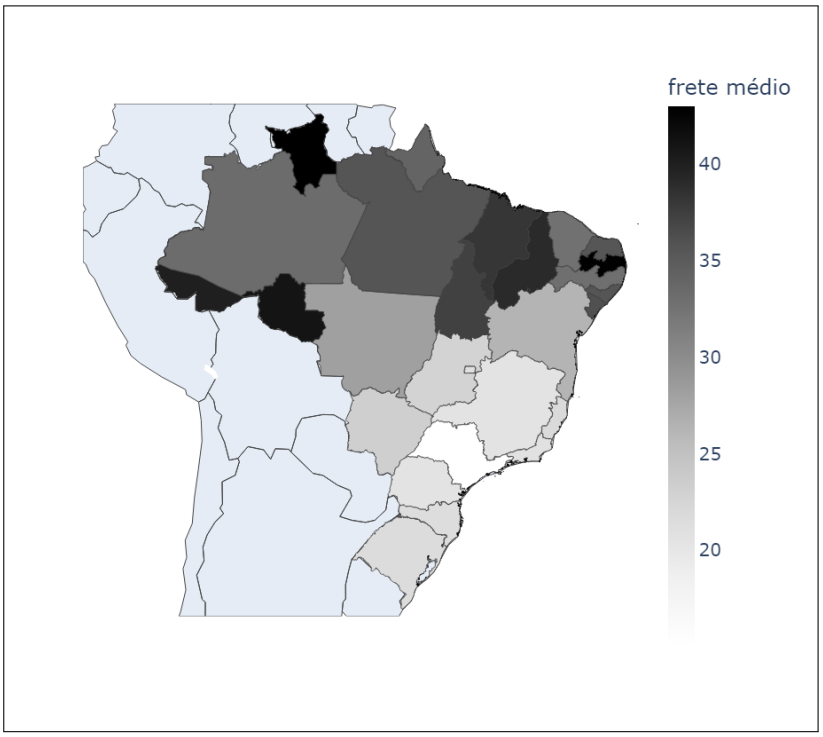
Fonte: Produção Própria

Figura 28 – Distribuição de Frete Médio em Reais por estados do Brasil em ranking



Fonte: Produção Própria

Figura 29 – Distribuição de de Frete Médio em Reais por estados do Brasil em mapa



Fonte: Produção Própria

nessas regiões não ser obstante das demais, acentua a dificuldade de acesso da referida população ao serviço de compras virtuais. Vale pontuar que os fatores possivelmente atenuantes para essa realidade são externos ao modelo de negócios de um comércio digital, uma vez que diz respeito a politicas de preços regionais e condições físicas de qualidade de estradas e desenvolvimento de setores logísticos.

Por outro lado, uma das necessidades no desenvolvimento de qualquer mercado é um bom processo de precificação. Durante esse processo, o empreendedor avalia todos os custos e despesas previstas na venda e afere um valor de preço que permita um lucro, mantendo ainda competitividade. Em modelos de mercado pouco explorados, um perigo para o consumidor se manifesta quando a competitividade não está estabelecida e preços, inclusive de frete, se apresentam muito elevados independente dos custos operacionais projetados.

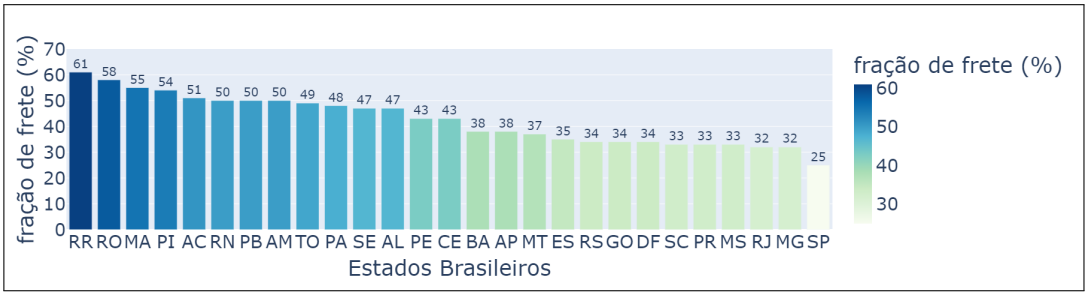
Uma pergunta válida que derive desse questionamento é: qual percentual o valor do frete representa em relação ao valor do produto?

Tabela 14 – Métricas de Fração de Frete por estado

	fração de frete
contagem	27,00
média	42,44
std	9,58
min	25,24
25%	33,83
50%	42,79
75%	49,83
max	61,38

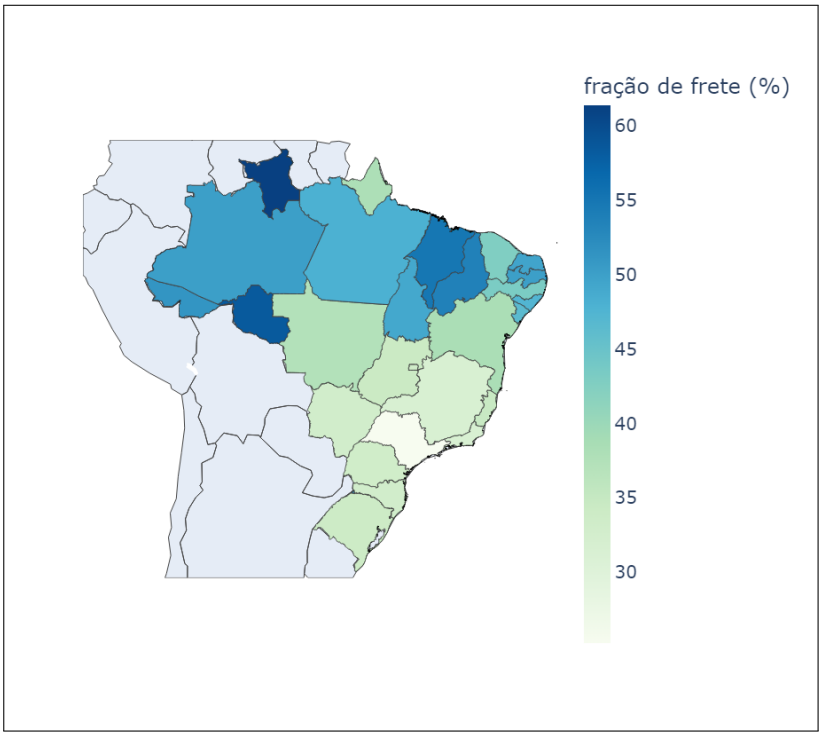
Fonte: Produção Própria

Figura 30 – Distribuição Fração do Frete em Relação ao Preço por estados do Brasil em ranking



Fonte: Produção Própria

Figura 31 – Distribuição de Fração do Frete em Relação ao Preço por estados do Brasil em mapa



Fonte: Produção Própria

A fração de frete é uma métrica extremamente simples, mas que pode ser mal interpretada. É importante salientar que aqui ela apenas corresponde a divisão do valor de frete pelo valor do pedido, sendo assim, seriam possíveis valores acima de 100%. Um outro ponto é que algum estado pode praticar um valor de frete baixo quando comparado aos demais, e ainda assim possuir uma fração de frete alta. O contrário também é possível: um estado com valor de frete alto praticar uma fração de frete baixa.

A Figura 30 representa a distribuição de frações de frete por estado em ranking, e em seguida a Figura 31 apresenta essa mesma distribuição por estados em mapa. Essa visualização foi incluída porque diz respeito à forma como os consumidores experienciam os preços, e pode ser um indicador de quão difícil é a realização de uma entrega em determinado estado. Aqui se destacam novamente o Norte-Nordeste com valor de frete mais expressivo, exceto Amapá e Bahia que se mantém abaixo da média apresentada na Tabela 14. Destaque também para São Paulo e Rio de Janeiro com menor fração de frete. Visto que um dos momentos de abandono possíveis de compra, por parte do cliente, é durante o cálculo do frete, se ele representa um valor substancial em relação ao valor dos itens do pedido, aumentam as chances de abandono do carrinho e desistência de compra.

Uma vez que já foi verificada uma expressiva maior presença de consumidores de região Sudeste, e visto que o banco de dados apenas inclui registros de clientes que decidiram seguir com a compra, uma hipótese válida é ponderar se uma baixa presença de clientes em certos estados mais distantes não pode estar associado a uma baixa taxa de conversão devido a fração alta de valor relativo de frete nesses estados.

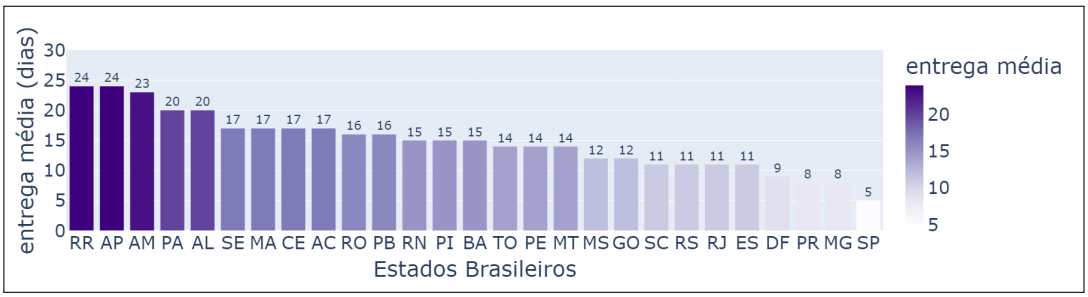
Isto posto, vale explorar mais a experiência de compra na perspectiva do cliente, considerando o impacto desse valor de frete no tempo de recebimento do produto: o valor no frete é suficiente para prover uma experiência de entrega com velocidade equivalente à de grandes centros?

Tabela 15 – Métricas de tempo de entrega por estado

	tempo médio de entrega
contagem	27,00
média	14,80
std	4,82
min	5,07
25%	11,36
50%	15,25
75%	17,16
max	24,15

Fonte: Produção Própria

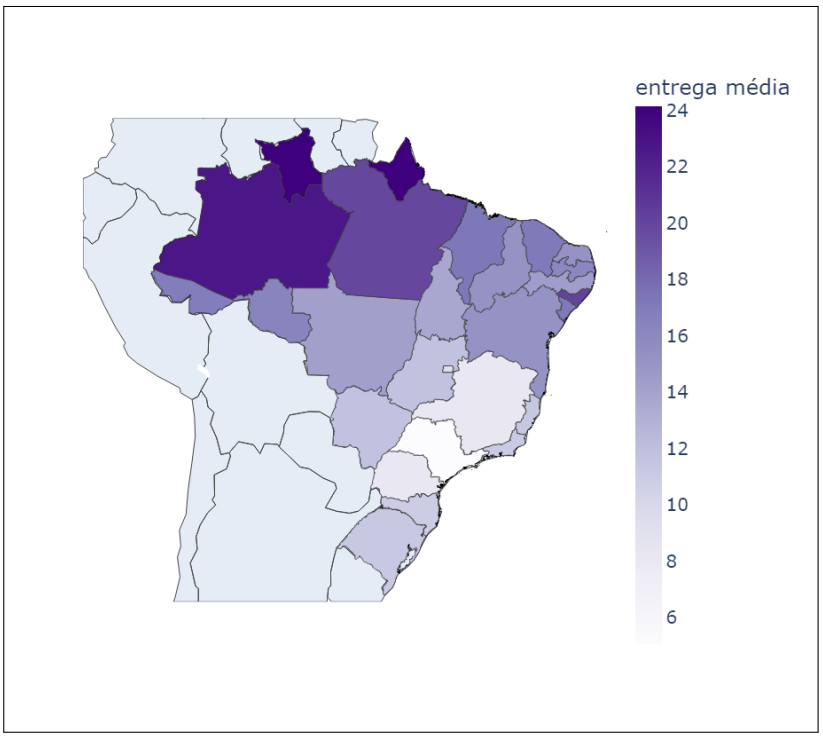
Figura 32 – Distribuição de Tempo de Entrega Médio em Dias por estados do Brasil em ranking



Fonte: Produção Própria

A Figura 32 apresenta o tempo médio de entrega em dias rankeando os estados brasileiros, em seguida a Figura 33 apresenta essa mesma distribuição em mapa. Mais uma vez é observado que ainda com o frete médio maior, há um destaque no tempo médio de espera para as regiões Nordeste-Norte. Esse, porém, é o tempo previsto de entrega, derivado da redução e conversão em dias da data prevista de recebimento do produto e a data da compra. Um destaque

Figura 33 – Distribuição de Tempo de Entrega Médio em Dias por estados do Brasil em mapa



Fonte: Produção Própria

aqui pode ser feito para o Amapá, Roraima e Amazonas que estão com maiores tempos de entrega estimada. Um detalhe é que mesmo o Amapá tendo uma fração de frete bem menos substancial quando comparado a Roraima, por exemplo, ambos tem uma experiência de espera semelhante. Vale ainda menção ao Paraná como segunda entrega mais rápida depois de São Paulo. Essa é uma distribuição bem regionalizada, não possui grandes diferenças dentro das mesmas regiões, preservando uma característica geográfica e uma distribuição ao redor da média apresentada na Tabela 15.

Nesse momento, cabe uma observação em relação ao universo do comércio digital: espera-se prover ao cliente a mais positiva experiência possível, mas vista uma maior dificuldade de entrega, a promessa de velocidade pode ter sido deixada em segundo plano em função de favorecer a experiência de promessa de "entrega dentro ou antes do prazo".

Nesse sentido, se pode verificar o percentual de pedidos fora do prazo. As figuras mostram esse indicador através da divisão entre a contagem de dias previstos para a entrega e a contagem de dias corridos de fato até confirmação de recebimento, tanto em ranking na Figura 34, como em mapa para cada estado na Figura 35, e métricas na Tabela 16

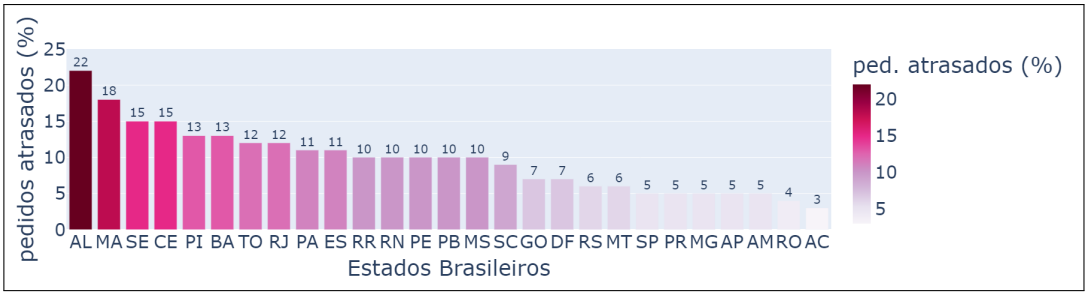
Novamente alguns números surpreendem, há uma concentração de maior percentual de pedidos atrasados na região Nordeste, ainda que estados como Rio de Janeiro, Espírito Santo

Tabela 16 – Métricas de pedidos atrasados por estado

	atrasados
contagem	27,00
média	9,67
std	4,58
min	3,28
25%	5,90
50%	9,67
75%	12,04
max	22,29

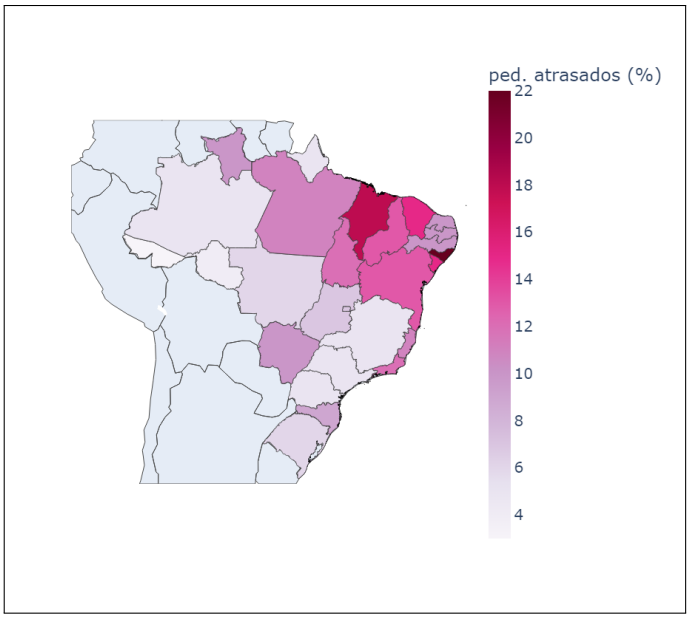
Fonte: Produção Própria

Figura 34 – Distribuição de Percentual de Pedidos Atrasados por estados do Brasil em ranking



Fonte: Produção Própria

Figura 35 – Distribuição de Percentual de Pedidos Atrasados por estados do Brasil em mapa



Fonte: Produção Própria

e Mato Grosso do Sul também apareçam acima da média apresentada na Tabela 16. Destaque também para Acre e Roraima com os menores índices de atraso, mesmo sendo localizados em

pontos distantes dos grandes centros, e de difícil acesso.

4.1.3.3 Satisfação

Finalmente, essa perspectiva de recebimento no prazo, associado aos já comentados elementos de precificação e frete, compõem alguns dos componentes que são refletidos na satisfação do consumidor. As Figuras 36 e 37 apresentam essa distribuição de satisfação em barras e mapa respectivamente.

Tabela 17 – Métricas de satisfação por estado

	review_nota
contagem	27,00
média	3,96
std	0,15
min	3,58
25%	3,84
50%	4,01
75%	4,07
max	4,22

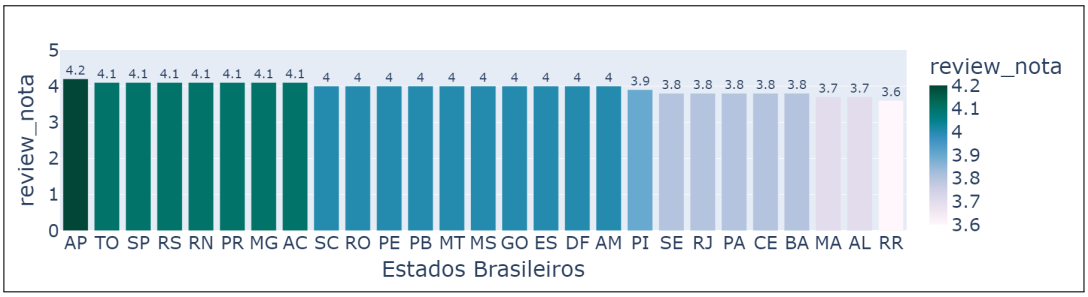
Fonte: Produção Própria

É possível perceber que a variação é sutil, com desvio padrão de 0,15 em relação a uma média de 3,96 conforme Tabela 17, vale considerar que o ato de reclamar ou fazer uma avaliação positiva pode demandar a necessidade de um incentivo incomum, como casos muito extremos em qualquer dos sentidos. Aqui se destacam com maior satisfação na compra Amapá, Tocantins e São Paulo, mesmo sendo estados que apareceram em posições diferentes ao longo das demais análises; e menor satisfação os estados de Roraima, Alagoas e Maranhão.

4.1.4 Análise histórica

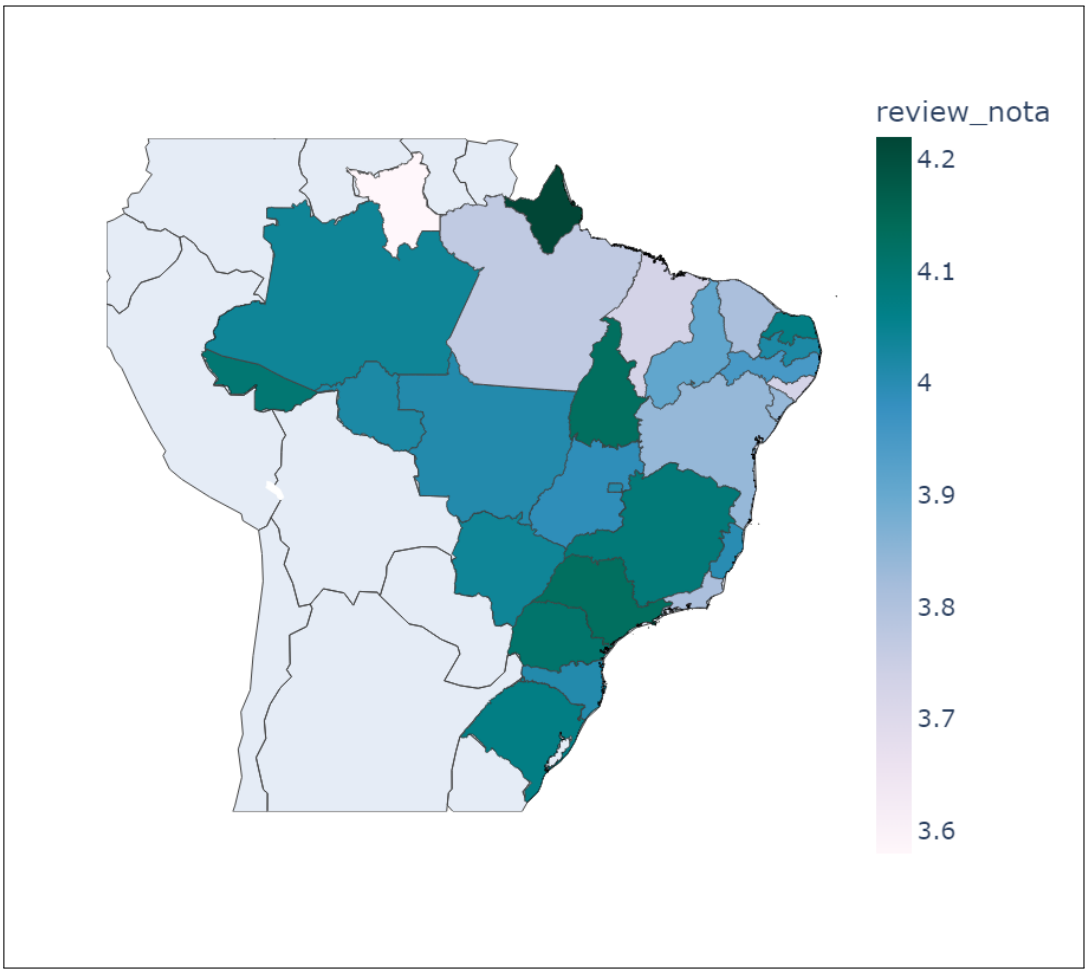
A evolução no tempo é um outro recurso importante na exploração do banco, afinal de contas se trata de um período imediatamente anterior a uma mudança drástica nas rotinas das pessoas globalmente, em decorrência das medidas de isolamento por conta da pandemia de Covid 19. Isto posto, convém verificar qual a tendência no volume de pedidos registrados, haja vista inclusive se tratar de uma plataforma com modelo de negócios recente e que estava em fase inicial de implantação no mercado. A esse último comentário cabe o contraponto que esse tipo

Figura 36 – Distribuição de Satisfação entre 1 e 5 por estados do Brasil em ranking



Fonte: Produção Própria

Figura 37 – Distribuição de Satisfação entre 1 e 5 por estados do Brasil em mapa



Fonte: Produção Própria

de plataforma conecta outras plataformas já existentes com clientes também que já podem ter sido fidelizados em um formato de compra semelhante.

O ponto focal dos dados está no ano de 2017. Dados de 2016 e 2018 não completam o ano, o que pode levar a conclusões pouco representativas. A Tabela 18 nos diz que se tratam de 634 dias ao todo ao longo dos 3 anos observados, a média diária é de 157 pedidos, variando de 1 pedido no mínimo até 1176 pedidos em um único dia. Vale verificar ainda que o desvio padrão aqui é bastante substancial em relação à média, o que se traduz em um coeficiente de variação de 60%, esse comportamento demonstra a característica ruidoza da variação no tempo.

A Figura 38 constrói uma visualização que permite o consumo dessas informações de forma facilitada e resumida. Aqui foi usado um passo de 15 dias no intervalo de conexão das linhas. Vale o destaque ao pico em 24 de novembro de 2017, primeira vez que a *Olist* em sua operação participa de um evento promocional de *Black Friday*. Nesse ponto o gráfico registra um pico de 1176 pedidos, cerca de 7,5 vezes mais que o volume médio diário de pedidos. Esse fenômeno demonstra como o mercado digital é suscetível aos efeitos de atividades promocionais coordenadas.

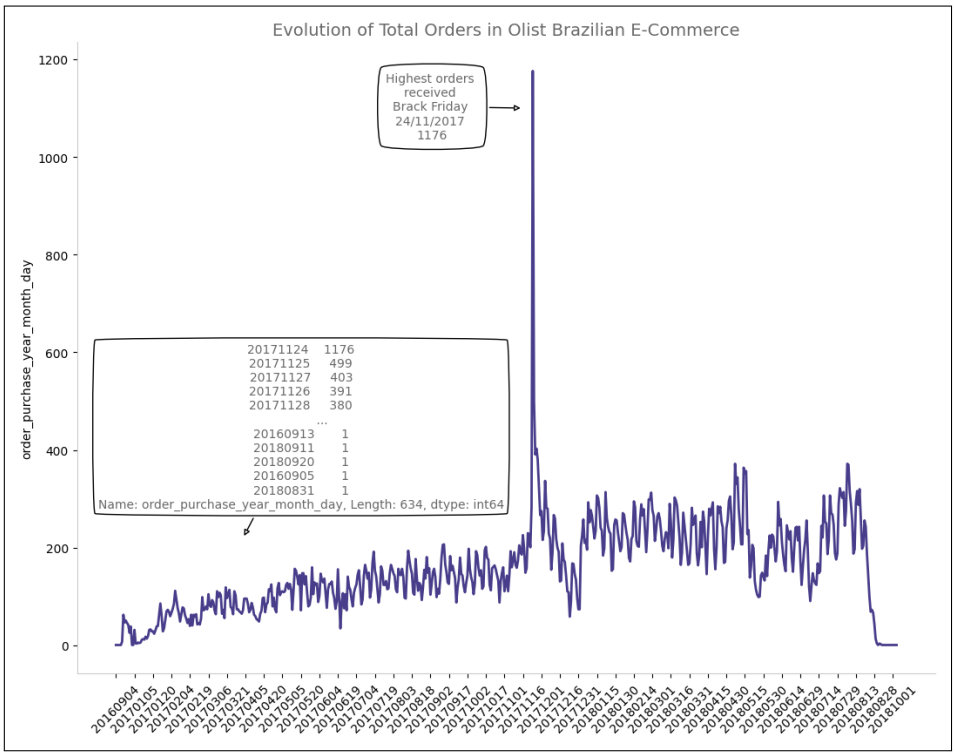
Um outro elemento que essa distribuição demonstra é a presença de dias com apenas 1 pedido, ao início e final dos registros. Esses são dados que pressupõem um erro de registro, e podem suplantar interesse investigativo no que poderia ter causado essa anomalia no banco. Possíveis razões além do erro são encerramento de operações, resíduos pós encerramento da coleta, corrupção dos dados.

Tabela 18 – Métricas variação histórica dos pedidos

	dias com pedidos registrados
contagem	634,000
média	156,847
std	94,225
min	1,000
25%	96,000
50%	148,000
75%	215,750
max	1176,000

Fonte: Produção Própria

Figura 38 – Distribuição de pedidos por data



Fonte: Produção Própria

5 CONCLUSÕES E TRABALHOS FUTUROS

O comércio digital é, sem dúvidas, um forte motor econômico e possue grande potencial de desenvolvimento no Brasil. A intensa participação online de sua população, ao ponto de ser elencado o quinto país com mais acesso à internet, e o vasto espaço para investimento como digitalização de processos, faz crer que há um fértil espaço para usar do mercado digital como ferramenta de desenvolvimento social. Isto dito, existe a necessidade de refletir sobre como as novas tecnologias afetam os diferentes contextos sociais, de forma a assegurar medidas para que haja eficiência e descentralização nesse desenvolvimento a médio e longo prazos.

O Brasil, hoje, já conta com medidas integradas de digitalização de processos que facilitam e fortalecem a economia digital. O PIX é um exemplo de medida do Banco Central brasileiro que facilitou transações não físicas em todo o território, o que já impacta diretamente transações de produtos virtuais, como vendas de ebooks, demais documentos, cursos e softwares, dentre outros, para consumidores com dificuldade de acesso a crédito. Ainda assim, foi verificado um desafio no que se diz respeito ao acesso da população à infraestrutura que cerca as operações do *ecommerce*, o que se reflete diretamente na experiência de uso do serviço como demonstrado pela análise do banco de dados em estudo.

Este trabalho explorou as nuances estruturais do banco de dados da *Olist*, conceituando suas principais informações e posteriormente contextualizando com o cenário nacional da temática em estudo. É importante ressaltar que, pelo mercado lucrativo e gigantesco do *ecommerce* no Brasil, essa amostra pode se mostrar muito pequena para representar em totalidade os indicadores estaduais e nacionais, mas haja vista a análise nos primeiros anos de implementação do *serviço Olist*, é possível ter uma boa ideia da evolução de aderência e onde estão concentradas as maiores facilidades e dificuldades em abordagem geográfica. Em seguida, as avaliações de estudos e orgãos oficiais são condizentes com os *insights* observados na descrição do banco de dados.

Como trabalhos futuros, é válida a construção de *dashboards* interativos para consumo das informações de forma mais acessível; buscar por bancos de dados mais densos e atuais, a fim de replicar as técnicas usadas e tirar segundas opiniões em relação às perguntas postuladas.

Por fim, a engenharia não pode ser uma ciência externa à sociedade, devendo sempre aplicar esforços não apenas para a produção de novas tecnologias, mas também para a solução de problemas de interesse da sociedade. Ainda, e sobretudo, se essas forem problemáticas que nasçam ou se desenvolvam a partir de tecnologias intrínsecas à própria engenharia.

REFERÊNCIAS

ANATEL. **Painel da Anatel**. Brasil, 2023. Disponível em: <https://informacoes.anatel.gov.br/paineis/acessos>. Acesso em: 07 ago. 2023.

ARAÚJO, M. S. **COMÉRCIO ELETRÔNICO: EVOLUÇÃO E PERSPECTIVAS**. Dissertação (Relações Internacionais) — UniCEUB, Brasília – DF, 2003.

BARBOSA, A. **Brasil já é o 5º país com mais usuários de internet no mundo**. Forbes, 2022. Disponível em: <https://forbes.com.br/forbes-tech/2022/10/brasil-ja-e-o-5o-pais-com-mais-usuarios-de-internet-no-mundo/#:~:text=Ao%20todo%2C%20s%C3%A3o%20mais%20de,milh%C3%B5es%20de%20usu%C3%A1rios%20de%20internet.> Acesso em: 25 jun. 2023.

BRASIL. Lei nº 8.078, de 11 de setembro de 1990. **Diário Oficial da União**, Brasília, DF, 1990. Disponível em: <https://www.planalto.gov.br/ccivil_03/leis/l8078compilado.htm>.

BRASIL. Portaria nº 378/2016/seimc. **Diário Oficial da União**, Brasília, DF, 2016. Disponível em: <https://repositorio.mcti.gov.br/handle/mctic/188>.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. **Diário Oficial da União**, Brasília, DF, 2018. Disponível em: <https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm>.

BRITO, F. O deslocamento da população brasileira para as metrópoles. **Estudos Avançados**, Instituto de Estudos Avançados da Universidade de São Paulo, v. 20, n. 57, p. 221–236, May 2006. ISSN 0103-4014. Disponível em: <https://doi.org/10.1590/S0103-40142006000200017>.

CARVALHO, M. S. R. M. de. **A TRAJETÓRIA DA INTERNET NO BRASIL**: Do surgimento das redes de computadores À instituiÇÃo dos mecanismos de governanÇa. Dissertação (Mestrado EM CIÊNCIAS DE ENGENHARIA DE SISTEMAS E COMPUTAÇÃO) — UNIVERSIDADE FEDERAL DO RIO DE JANEIRO, RIO DE JANEIRO, RJ, 2006.

CETIC.BR. **TIC DOMICÍLIOS 2019**: Principais resultados. Unesco, 2020. Disponível em: <https://cetic.br/media/analises/tic_domicilios_2019_coletiva_imprensa.pdf>. Acesso em: 15 ago. 2023.

CORREIOS. **Tudo sobre CEP**. Brasil, 2023. Disponível em: <https://www.correios.com.br/enviar/precisa-de-ajuda/imagens/tudo-sobre-cep>. Acesso em: 28 jun. 2023.

DEARO, G. **Eis a primeira coisa a ser vendida pela Internet na história**. 2015. Disponível em: <https://exame.com/marketing/eis-a-primeira-coisa-a-ser-vendida-pela-internet-na-historia/>. Acesso em: 13 jun. 2023.

FERNANDES, V. **82cartão de crédito**. Forbes, 2022. Disponível em: <https://forbes.com.br/forbes-money/2022/10/82-dos-brasileiros-que-fazem-compras-online-possuem-cartao-de-credito/>. Acesso em: 25 jun. 2023.

FORTUNATO, L. E. **Segmentação de clientes de um e-commerce brasileiro utilizando RFV e métodos de clusterização particionais**. Dissertação (Engenharia Mecânica) — UNIVERSIDADE FEDERAL DE SANTA CATARINA, Florianópolis, 2022.

GERALDI, P. de A. **Como se dá a relevância do Comércio Eletrônico no contexto internacional? Uma análise sobre sua história, definição e impacto nas dinâmicas do comércio**. Dissertação (Relações Internacionais) — PUC-Rio, Rio de janeiro, 2021.

GERHARDT, T. E.; SILVEIRA, D. T. **Métodos de Pesquisa**. 1. ed. Porto Alegre, RS: Editora da UFRGS, 2009.

GETSCHKO, D. **World Summit on Internet and multimedia**: Internet in brazil. Montreux, 2002. Disponível em: <www.fiam.org/sommets/summit2002/documents/Getschko.ppt>. Acesso em: 10 set. 2006.

HAYES, A. **Merchandising: How Companies Entice Customers To Spend**. 2022. Disponível em: <https://www.investopedia.com/terms/m/merchandising.asp>. Acesso em: 15 jun. 2023.

IBGE. **Pesquisa Nacional por Amostra de Domicílios Contínua**. [S.l.]: BRASIL, 2020. ISBN 978-85-240-4529-5.

IBGE. **Regiões de Influência das CIDADES**: 2018 / ibge, coordenação de geografia. [S.l.]: BRASIL, 2020. ISBN 9786587201047.

IBGE. **Tabela 7307 - Domicilios e Moradores, por situação do domicílio e existência de utilização da internet no domicílio**. Brasil, 2022. Disponível em: <https://sidra.ibge.gov.br/tabela/7307>. Acesso em: 07 ago. 2023.

IBGE. **Tabela 7532 - Rendimento médio mensal real domiciliar per capita, a preços médios do ano, por classes acumuladas de percentual das pessoas em ordem crescente de rendimento domiciliar per capita**. Brasil, 2022. Disponível em: <https://sidra.ibge.gov.br/tabela/7532>. Acesso em: 07 ago. 2023.

JACINSKI, E. **Sentidos das interações entre tecnologia e sociedade na formação de engenheiros**: Limites e possibilidades para repensar a educação tecnológica. Dissertação (Doutorado em Educação Científica e Tecnológica) — Universidade Federal de Santa Catarina, Florienópolis, SC, 2012.

KEENAN, M.; MOORE, K. **Supply Chain Trends That Will Shape 2023**. 2022. Disponível em: <https://www.shopify.com/enterprise/supply-chain-trends-strategies>. Acesso em: 16 jun. 2023.

KHAN, M. **What is E-Commerce Accounting? A Comprehensive Guide!** 2022. Disponível em: <https://expertiseaccelerated.com/e-commerce-accounting-guide/>. Acesso em: 16 jun. 2023.

MATA, K. B. D. C. **E-COMMERCE**: Análise de dados sobre o comércio eletrônico no brasil. Dissertação (Engenharia de Computação) — PONTIFÍCIA UNIVERSIDADE CATÓLICA DE GOIÁS, Goiânia, 2021.

MEDRI, W. **ANÁLISE EXPLORATÓRIA DE DADOS**. Universidade Federal de Londrina, 2011. Disponível em: <https://docs.ufpr.br/~benitoag/apostilamedri.pdf>. Acesso em: 25 jun. 2023.

MENDONÇA, H. G. de. **E-commerce**. [S.l.]: IPTEC – Revista InovaÇão, Projetos e Tecnologias, 2016. ISBN 2318-9851.

NASCIMENTO, J. V. de P.; PINHEIRO, R. P.; CIPRO, V. A. T. de P. **COMÉRCIO ELETRÔNICO, SUA LOGÍSTICA E O IMPACTO DE CRESCIMENTO DURANTE A PANDEMIA DA COVID 19**. Dissertação (Tecnologia em Logística) — CENTRO ESTADUAL DE EDUCAÇÃO TECNOLÓGICA PAULA SOUZA, Jundiaí, 2022.

NEIVA, A. C. **E-commerce no Brasil**: dados de um mercado em expansão. 2023. Disponível em: <https://edrone.me/pt/blog/dados-ecommerce-brasil>. Acesso em: 20 jun. 2023.

NIWATE, T. S. **IMPACT ON REVENUE GENERATION OF OLIST ECOMMERCE COMPANY ON THE BASIS OF V Y ON THE BASIS OF VARIOUS PRODUC ARIOUS PRODUCT PARAMETERS**. 2021. Disponível em: <https://scholarworks.lib.csusb.edu/etd/1315/>. Acesso em: 14 mai. 2023.

OLIST. **Brazilian E-Commerce Public Dataset by Olist**. Kaggle, 2021. Disponível em: <https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce?select=olist_customers_dataset.csv>. Acesso em: 03 fev. 2023.

OLIST. **Sobre Nós**. Olist, 2023. Disponível em: <https://olist.com/pt-br/sobre-nos/>. Acesso em: 03 fev. 2023.

OLIVEIRA, F. **A história do e-commerce. Como tudo começou?** 2021. Disponível em: <https://www.linkedin.com/pulse/hist%C3%B3ria-do-e-commerce-como-tudo-come%C3%A7ou-francilene-oliveira/?originalSubdomain=pt>. Acesso em: 13 jun. 2023.

REIS, E. A.; REIS, I. A. **Análise Descritiva de Dados**. [S.l.]: Relatório Técnico do Departamento de Estatística da UFMG, 2002.

SEBRAE. **Entenda os principais conceitos de marketing digital**. 2022. Disponível em: <https://sebrae.com.br/sites/PortalSebrae/artigos/entenda-os-principais-conceitos-de-marketing-digital,c733097ec5591810VgnVCM100000d701210aRCRD>. Acesso em: 15 jun. 2023.

SHEN, Y. Application of big data technology in e-commerce. **Journal of Physics: Conference Series**, IOP Publishing, v. 1682, n. 1, p. 012075, nov 2020. Disponível em: <https://dx.doi.org/10.1088/1742-6596/1682/1/012075>.

VENCESLAU, I. **Correios, Logística e Uso do Território**: O serviço de encomenda expressa no brasil. Dissertação (Mestrado Geografia Humana) — UNIVERSIDADE De São Paulo, São Paulo, SP, 2017.