



CENTRO UNIVERSITÁRIO LUTERANO DE PALMAS

Recredenciado pela Portaria Ministerial nº 1.162, de 13/10/16, D.O.U. nº 198, de 14/10/2016
AELBRA EDUCAÇÃO SUPERIOR - GRADUAÇÃO E PÓS-GRADUAÇÃO S.A.

Renan Matheus Castro Da Silva

DASHBOARD PARA ANÁLISE E VISUALIZAÇÃO DE DADOS QUE COMPÕEM O
IDH A PARTIR DE MICRODADOS EXTRAÍDOS DO IBGE

Palmas – TO

2023

Renan Matheus Castro da Silva

DASHBOARD PARA ANÁLISE E VISUALIZAÇÃO DE DADOS QUE COMPÕEM O
IDH A PARTIR DE MICRODADOS EXTRAÍDOS DO IBGE

Projeto de Pesquisa elaborado e apresentado como requisito parcial para aprovação na disciplina de Projeto Tecnológico do curso de bacharelado em Sistemas de Informação pelo Centro Universitário Luterano de Palmas (CEULP/ULBRA).

Orientadora: Prof^a. Esp. Fernanda Pereira Gomes.

Palmas – TO

2023

Renan Matheus Castro da Silva

DASHBOARD PARA ANÁLISE E VISUALIZAÇÃO DE DADOS QUE COMPÕEM O
IDH A PARTIR DE MICRODADOS EXTRAÍDOS DO IBGE

Projeto de Pesquisa elaborado e apresentado como requisito parcial para aprovação na disciplina de Projeto Tecnológico do curso de bacharelado em Sistemas de Informação pelo Centro Universitário Luterano de Palmas (CEULP/ULBRA).

Orientadora: Prof. ^a Esp. Fernanda Pereira Gomes.

Aprovado em: ____/____/____

BANCA EXAMINADORA

Prof.^a Esp. Fernanda Pereira Gomes

Orientadora

Centro Universitário Luterano de Palmas – CEULP

Prof. Me. Fabiano Fagundes

Centro Universitário Luterano de Palmas - CEULP

Prof Esp. Fábio Castro Araújo

Centro Universitário Luterano de Palmas - CEULP

Palmas – TO

2023

RESUMO

SILVA, Renan Matheus Castro da. **Dashboard para análise e visualização de dados que compõem o IDH a partir de microdados extraídos do IBGE**. 2023. 63 f. Projeto Tecnológico (Graduação) – Curso de Sistemas de Informação, Centro Universitário Luterano de Palmas, Palmas/TO, 2023.

Os microdados provenientes do Instituto Brasileiro de Geografia e Estatística (IBGE) fornecem uma ampla gama de informações sobre aspectos educacionais, sociais e econômicos da população brasileira, sendo utilizados para calcular o Índice de Desenvolvimento Humano (IDH). No entanto, a visualização desses microdados apresenta uma complexidade, devido ao grande volume de dados sem uma estrutura clara. Analisar manualmente essa enorme quantidade de dados seria inviável, dada a quantidade de tempo e esforço que tal tarefa demandaria. Nesse contexto, este trabalho consiste em apresentar os elementos fundamentais para o desenvolvimento de uma ferramenta que possibilite a análise e visualização do IDH brasileiro e seus indicadores, utilizando os microdados do IBGE, de forma que os resultados obtidos contribuam para a descoberta de novos dados, auxilie na tomada de decisões em políticas públicas e no monitoramento de indicadores socioeconômicos. Com isso, foi desenvolvido um *dashboard* para a visualização e análise dos indicadores do IDH brasileiro, com a utilização do processo ETL para o tratamento dos microdados e a plataforma Flutter para a implementação. Essa abordagem demonstrou que, ao utilizar ferramentas adequadas, é possível criar um dashboard capaz de facilitar a visualização e análise correta desse grande volume de dados.

Palavras-chave: IBGE, IDH, *Dashboard*, ETL.

LISTA DE FIGURAS

Figura 1 - Exemplo de Pseudonimização e Anonimização dos dados	13
Figura 2 - Estrutura de diretórios dos microdados do PNAD 2012.....	15
Figura 3 - Atlas do Desenvolvimento Humano no Brasil.....	18
Figura 4 - Etapas do processo ETL	19
Figura 5 - Modelo de Dashboard	21
Figura 6 - Exemplo de Dashboard Estratégico	24
Figura 7 - Dashboard Tático.....	26
Figura 8 - Dashboard Operacional	27
Figura 9 - Exemplo de gráfico de barras.....	28
Figura 10 - Exemplo gráfico de setores	29
Figura 11 - Exemplo gráfico de linhas	30
Figura 12 - Exemplo gráfico de área	31
Figura 13 - Princípio da Proximidade	33
Figura 14 - Princípio da Similaridade.....	34
Figura 15 - Princípio da Continuidade	35
Figura 16 - Metodologia do Projeto	38
Figura 17 - Processo ETL executado	40
Figura 18 - Base de dados IBGE.....	41
Figura 19 - Base de dados Taxa de analfabetismo 18 anos ou mais.....	42
Figura 20 - Base de dados Renda per capita	43
Figura 21 - Atributos de renda per capita	44
Figura 22 - Base de dados Esperança de vida ao nascer	45
Figura 23 - Base de dados Extrema Pobreza	45
Figura 24 - Base de dados IDHM	46
Figura 25 - Atributos normalizados da base de dados IDHM.....	47
Figura 26 - Mesclagem entre renda e analfabetismo.....	48
Figura 27 - Mesclagem entre analfa_renda e esperança de vida	48
Figura 28 - Mesclagem entre analfa_renda_espvida e extrema pobreza	48
Figura 29 - Mesclagem entre analfa_renda_espvida_pob e idhm	49
Figura 30 - Conversão de vírgula para ponto.	49
Figura 31 - Conversão de Object para Float.....	50
Figura 32 - Alterando dados de CVS para JSON	50

Figura 33 - Dados armazenados no Firebase.....	51
Figura 34 - Protótipo de tela do dashboard	53
Figura 35 - Tela do dashboard desenvolvido.....	54
Figura 36 - Exemplo de configuração do FirebaseOptions	56
Figura 37 - Trecho do código de conexão com o documento	56
Figura 38 - Teste dos cards e gráfico de barras	57
Figura 39 - Teste dos gráficos.....	57

LISTA DE TABELAS

Tabela 1 - Limites mínimo e máximos para a classificação do IDH.....	17
Tabela 2 - Tipos de Dashboards.....	23

LISTA DE ABREVIATURAS E SIGLAS

ASCII	<i>American Standard Code for Information Interchange</i>
BD	Banco de dados
CSV	<i>Comma-Separated Values</i>
DW	<i>Data Warehouse</i>
ETL	<i>Extract, Transform and Load</i>
IBGE	Instituto Brasileiro de Geografia e Estatística
IDH	Índice de Desenvolvimento Humano
IDHM	Índice de Desenvolvimento Humano Municipal
PIB	Produto Interno Bruto
PNAD	Pesquisa Nacional por Amostra de Domicílios
PNS	Pesquisa Nacional de Saúde
PPC	Paridade do Poder de Compra

SUMÁRIO

1 INTRODUÇÃO	10
2 REFERENCIAL TEÓRICO	12
2.1 MICRODADOS	12
2.1.1 MICRODADOS DO IBGE	14
2.2 ÍNDICE DE DESENVOLVIMENTO HUMANO.....	15
2.2.1 IDHM	17
2.3 ETL.....	19
2.4 DASHBOARD	20
2.4.1 TIPOS DE DASHBOARDS	22
2.4.2 TIPOS DE GRÁFICOS.....	27
2.4.3 PERCEPÇÃO VISUAL.....	32
3 METODOLOGIA	36
3.1 MATERIAIS	36
3.2 MÉTODOS.....	38
4 RESULTADOS	40
4.1 PROCESSO ETL.....	40
4.1.1 EXTRAÇÃO.....	41
4.1.2 TRANSFORMAÇÃO	47
4.1.3 CARREGAMENTO	51
4.2 DESENVOLVIMENTO	52
4.2.1 PROTOTIPAÇÃO	52
4.2.2 IMPLEMENTAÇÃO.....	54
5 CONSIDERAÇÕES FINAIS	59
REFERÊNCIAS	61

1 INTRODUÇÃO

No início da década de 1990, foi criado pelo Programa das Nações Unidas para o Desenvolvimento (PNUD), o índice de Desenvolvimento Humano (IDH), dado estatístico com a finalidade de verificar o nível de desenvolvimento de uma nação utilizando indicadores de desempenho referentes aos aspectos econômicos e sociais direcionados a qualidade de vida e às condições essenciais da sociedade. O IDH, assim, passou a ser o mais conhecido cálculo do desenvolvimento humano Torres, Ferreira e Dini (2003).

No Brasil, as informações utilizadas para o cálculo do IDH são coletadas pelo Instituto Brasileiro de Geografia e Estatística (IBGE), que tem como objetivo coletar e produzir dados sobre a população brasileira, abrangendo diversos temas, como educação, saúde, trabalho, renda e condições de vida. Além disso, fornece informações valiosas através dos microdados produzidos, que auxiliam na análise socioeconômica do país, permitindo a identificação de desigualdades e formulação de estratégias para reduzi-las.

Os microdados coletados no IBGE formam um grande volume de dados abertos, provindos de pesquisas realizadas ao longo dos anos e livre de licenças, patentes ou mecanismos de controle, de modo a estarem livremente disponíveis para serem utilizados. Para analisar essa massa, é essencial recorrer ao uso de ferramentas ágeis que permitam facilitar a visualização. Além disso, é de suma importância o direcionamento desses microdados a determinado contexto na sociedade para a formulação de políticas públicas, tomada de decisões e monitoramento de indicadores sociais e econômicos.

Sendo assim, o processo ETL (*Extract, Transform, Load* - Extração, Transformação, Carga) é considerado como uma ferramenta fundamental para integrar e analisar os microdados contidos no IBGE. É necessário extrair os dados brutos, transformá-los em informações úteis para a análise e carregá-los em um sistema que permita a correta visualização dos indicadores de desempenho do IDH. A transformação dos dados pode incluir a limpeza e normalização dos microdados, a combinação de informações de diferentes fontes e a criação de novas variáveis para melhor representar o contexto do desenvolvimento humano no Brasil.

Desta forma, o problema de pesquisa proposto neste trabalho foi: Como analisar e visualizar os dados relacionados ao IDH a partir de microdados extraídos do

IBGE? Por meio de pesquisas estudadas a respeito da visualização e análise de dados, foi suposto ser possível desenvolver um *dashboard* para auxiliar de maneira eficiente a apresentação clara dos dados processados, visto que tal ferramenta forneça uma visão global em uma única tela com inúmeras possibilidades de interação servindo como ponto de partida para as principais informações de uma determinada área (BUCHSBAUM, 2012).

Partindo disso, o objetivo geral do trabalho foi desenvolver um *dashboard* para auxiliar na análise e visualização de dados que compõem o IDH brasileiro a partir de microdados do IBGE, o que trouxe consigo os objetivos específicos: identificar os microdados necessários contidos na base de dados do IBGE; realizar o processo ETL; carregar os dados processados para um *dashboard*; definir um design de fácil visualização e disponibilizar a visualização dos dados processados em um *dashboard* interativo.

A análise desse grande volume de dados de forma manual torna-se inviável, tendo em vista a grande quantidade de tempo e esforço que tal tarefa demandaria. Além disso, realizar o relacionamento entre os dados manualmente é uma tarefa complexa e pode dificultar a visualização de novas descobertas. Portanto, é de suma importância a implementação de uma ferramenta ágil e inteligente que permita facilitar a análise e a visualização interativa desse grande volume de dados.

Por fim, este trabalho está constituído das seguintes seções: seção 2, Referencial Teórico, apresenta os conceitos gerais sobre microdados, processo ETL e *Dashboards*; seção 3 Materiais e Métodos, fornece uma breve explicação sobre os materiais e tecnologias utilizadas e a metodologia realizada para o desenvolvimento do trabalho; seção 4 Resultados, descreve de forma detalhada os resultados obtidos na execução do trabalho; seção 5, apresenta as considerações sobre os resultados obtidos em relação aos objetivos do trabalho.

2 REFERENCIAL TEÓRICO

Esta seção apresenta a fundamentação teórica utilizada na construção deste projeto. As subseções 2.1 e 2.1.1 respectivamente expõem conceitos pertinentes sobre Microdados e Microdados do IBGE; seguido da subseção 2.2 e 2.2.1 que abordará conceitos relacionados ao IDH e IDHM, e da subseção 2.3 que apresentará detalhes acerca dos *Dashboards* e suas características.

2.1 MICRODADOS

Os microdados representam o menor grau de detalhamento de um dado e são coletados por meio de pesquisas ou avaliações através de organizações ou entidades produtoras de estatística podendo fornecer informações detalhadas sobre indivíduos, famílias, empresas ou outras unidades de observação em um nível individual ou microscópico (MARTINS; MAURITTI; COSTA, 2013). Além disso, são amplamente utilizados em pesquisas censitárias ou amostrais, relacionados a economia, educação, demografia e saúde pública.

As informações detalhadas sobre cada unidade de observação são estruturadas em arquivos eletrônicos de dados. Esses arquivos são compostos por linhas ou registros e podem conter variáveis de diversos tipos (DUPRIEZ; BOYKO, 2010). Essas variáveis são acompanhadas por um dicionário que fornece informações detalhadas a cerca de uma determinada variável, incluindo nome, descrição, unidade de medida e outras características relevantes, permitindo com que os usuários saibam identificar cada uma delas e como foram medidas e codificadas.

Em um processo de divulgação de dados estatísticos, existe uma política ética que consiste no anonimato e confidencialidade dos dados coletados. A anonimização dos microdados é uma etapa no qual torna os dados pessoais coletados, identificáveis, protegendo assim a identidade dos indivíduos e entidades. Esse processo é realizado por meio de técnicas de remoção de informações identificáveis, substituindo-as por pseudônimos e agregando dados para reduzir a granularidade.

Na técnica de pseudonimização, os dados pessoais são tratados de forma em que os dados são atribuídos somente ao titular ou gestor dos dados coletados mediante a utilização de informações adicionais mantidas em ambientes controlados e seguros. Um exemplo de pseudonimização pode ser observado na Figura 1.

Figura 1 - Exemplo de Pseudonimização e Anonimização dos dados

Resultado dos Dados Coletados				
Nome:	Sexo :	Idade:	Estado Civil:	Renda Mensal:
Exemplo da Silva	M	18	Solteiro (a)	R\$ 800,00
Exemplo da Carvalho	F	25	Solteiro (a)	R\$ 2.500,00
Exemplo da Silva Costa	F	31	Casado (a)	R\$ 3.200,00

Pseudonimização dos Dados				
Nome:	Sexo :	Idade:	Estado Civil:	Renda Mensal:
1235781066	M	18	Solteiro (a)	R\$ 800,00
2245781096	F	25	Solteiro (a)	R\$ 2.500,00
1245731016	F	31	Casado (a)	R\$ 3.200,00

Anonimização Dados Coletados			
Sexo :	Idade:	Estado Civil:	Renda Mensal:
M	18	Solteiro (a)	R\$ 800,00
F	25	Solteiro (a)	R\$ 2.500,00
F	31	Casado (a)	R\$ 3.200,00

Fonte: Elaboração Própria

No exemplo da Figura 1, é possível visualizar a estrutura de origem dos dados coletados (Resultado dos Dados Coletados) em um contexto que inclui **Nome, Sexo, Idade, Estado Civil, e Renda Mensal** dos indivíduos. Em seguida é apresentada a pseudonimização dos dados (Pseudonimização dos Dados) com os respectivos nomes substituídos por valores numéricos correspondentes a um número de inscrição ou documento do indivíduo. Dessa forma, tem-se a amostragem dos dados anonimizados (Anonimização Dados Coletados) na qual são apresentados os dados sem a necessidade de incluir a variável nome com seus valores correspondentes.

Por fim, ao utilizar esses microdados em conjunto com uma ferramenta de visualização de dados como um *dashboard*, é possível extrair *insights* valiosos e criar uma representação clara e concisa dessas informações para diversos públicos, desde pesquisadores acadêmicos até gestores públicos e privados. Além disso, a utilização de microdados permite uma análise mais aprofundada e detalhada dos dados, possibilitando uma maior compreensão das tendências e padrões subjacentes aos dados

e, conseqüentemente, uma melhor tomada de decisão. Portanto, a utilização dos microdados para o projeto é de suma importância, sendo fundamental para a criação de um dashboard preciso, relevante e informativo.

2.1.1 MICRODADOS DO IBGE

No Brasil, o Instituto Brasileiro de Geografia e Estatística (IBGE) é uma das principais fontes de microdados. O IBGE coleta, processa e disponibiliza uma ampla variedade de dados estatísticos sobre a população brasileira e suas características, como dados demográficos, socioeconômicos, educacionais, culturais, ambientais, entre outros. Esses dados são utilizados para monitorar tendências e desenvolver políticas públicas, bem como para fins de pesquisa acadêmica.

Entre as principais pesquisas que geram microdados pelo IBGE estão a Pesquisa Nacional por Amostra de Domicílios (PNAD), Pesquisa Nacional de Saúde (PNS) e Censo Demográfico. Esses dados têm sido amplamente utilizados em estudos empíricos em diversas áreas, como economia, sociologia, demografia, saúde, educação entre outras. A utilização desses microdados tem sido fundamental para a produção de conhecimento científico no Brasil, permitindo a realização de análises mais detalhadas e precisas sobre a realidade brasileira.

Ademais, os microdados do IBGE são uma fonte importante para entender como as três dimensões do IDH se relacionam entre si e com outras variáveis socioeconômicas no Brasil. Eles permitem analisar a relação entre o nível de educação da população e o seu acesso a serviços de saúde e empregos mais qualificados, o que pode ter um impacto direto na renda e, conseqüentemente, no IDH.

Os diretórios de microdados no site do IBGE são organizados por pesquisas e anos de referência. Cada diretório contém arquivos de dados em formato ASCII ou CSV, bem como a documentação relacionada à pesquisa e são organizados conforme mostra a Figura 2 a seguir.

Figura 2 - Estrutura de diretórios dos microdados do PNAD 2012



Fonte: IBGE

A Figura 2 exibe a estrutura de diretórios e arquivos dos microdados do PNAD, referentes ao de 2012. Dentre os arquivos compactos encontram-se o dicionário de variáveis, manual do usuário, questionário de perguntas aplicadas aos indivíduos e os microdados.

Nesse contexto, os microdados são essenciais para o desenvolvimento deste projeto já que fornecem informações detalhadas e necessárias sobre a população brasileira e suas características socioeconômicas. Esses dados serão utilizados para gerar *insights* e informações relevantes para fins acadêmicos e até mesmo monitoramento de políticas públicas.

2.2 ÍNDICE DE DESENVOLVIMENTO HUMANO

O Índice de Desenvolvimento Humano (IDH) foi criado pelo Programa das Nações Unidas para o Desenvolvimento (PNUD) e é uma medida que avalia o desenvolvimento humano nos países com base em três dimensões: saúde, educação e renda. O IDH é uma ferramenta importante para avaliar e comparar o desenvolvimento humano entre países e regiões e foi desenvolvido para ser um contraponto a outro indicador bastante conhecido, o Produto Interno Bruto (PIB) *per capita* já que para calcular o desenvolvimento de uma população não se deve observar somente a dimensão econômica, mas também as características sociais, políticas e culturais que influenciam na qualidade da vida das pessoas. De acordo com PNUD (2005):

O IDH não abrange todos os aspectos de desenvolvimento e não é uma representação da "felicidade" das pessoas, nem indica "o melhor lugar no mundo para se viver". Democracia, participação, equidade, sustentabilidade são outros dos muitos aspectos do desenvolvimento humano que não são contemplados no IDH.

As três dimensões avaliadas são de suma importância para definir a expansão das liberdades pessoais, pois baseiam-se na oportunidade de se levar uma vida duradoura e saudável, ter acesso a conhecimento e poder desfrutar de um padrão de vida digno (PREARO; MARACCINI; ROMEIRO, 2015).

Na dimensão de longevidade, espera-se que o desenvolvimento humano, seja avaliado a partir da qualidade de vida, acessos ao serviço de saúde e nutrição adequada tornando a expectativa de vida um indicador importante para avaliar o bem-estar geral de uma população. Por isso, é fundamental que políticas públicas voltadas para a promoção da saúde e do bem-estar sejam implementadas de forma a garantir que todas as pessoas tenham acesso aos mesmos direitos e oportunidades.

A dimensão de educação, tem como objetivo avaliar o acesso ao conhecimento por meio da taxa de alfabetização de adultos e a taxa combinada de matrícula no ensino fundamental, médio e superior. São fatores essenciais para o acesso ao conhecimento pois permitem que as pessoas consigam tomar decisões sobre seus futuros.

Por fim, a dimensão da renda, cujo objetivo é avaliar o padrão de vida de uma sociedade, permitindo que as pessoas tenham acesso às necessidades básicas. Para mensurar a riqueza de uma nação, essa dimensão mede o PIB *per capita* ajustado pela Paridade do Poder de Compra (PPC) no qual reflete na capacidade de um país de proporcionar oportunidades econômicas e financeiras para seus cidadãos.

As três dimensões combinadas para a realização do cálculo são transformadas em índices numéricos que variam de zero (0) a um (1) e sua média aritmética resulta em um indicador síntese, o IDH, que quanto mais próximo estiver de um (1) maior será o índice de desenvolvimento apresentado (ATLAS,2000). Além disso, é possível classificar os níveis do IDH em cinco categorias: Muito Baixo, Baixo, Médio, Alto e Muito Alto. Essa classificação para as pontuações está presente na Tabela 1 a seguir.

Tabela 1 - Limites mínimo e máximos para a classificação do IDH

Pontuação	Classificação
0,000 – 0,499	Muito Baixo
0,500 – 0,599	Baixo
0,600 – 0,699	Médio
0,700 – 0,799	Alto
0,800 – 1,000	Muito Alto

Fonte: Elaboração própria.

É possível observar na Tabela 1, os cinco níveis de classificação do IDH, que vão desde Muito Baixo até o Muito Alto. Países com maiores índices de desenvolvimento econômico e humano tendem a ter a pontuação mais próxima de 1, o que ainda não faz parte da realidade brasileira, já que em 2021 o IDH do Brasil foi de 0,754. Apesar de ser considerado uma classificação alta, o país ocupou a 87^a posição no *ranking* entre 191 países (PNUD 2021).

2.2.1 IDHM

O IDH permite avaliar o progresso das localidades, estados e áreas geográficas por meio da utilização do Índice de Desenvolvimento Humano Municipal (IDHM). No Brasil, o IDHM é um ajuste da metodologia empregada no IDH Global e é composto pelas mesmas três dimensões com o objetivo de estimular a formulação e implantação de políticas públicas em um nível municipal e estadual, para priorizar melhorias na qualidade de vida da população (PREARO; MARACCINI; ROMEIRO, 2015). Além disso, o índice tem sido utilizado como uma medida de comparação entre municípios e estados brasileiros, permitindo a identificação de desigualdades regionais e a definição de prioridades para a alocação de recursos.

É importante ressaltar que o IDHM não é um indicador perfeito e não abrange todas as áreas relevantes para avaliar o desenvolvimento humano. No entanto, ele é um recurso valioso para comparar municípios e estados brasileiros diferentes e tem ajudado a fomentar discussões sobre o desenvolvimento humano no país. A Figura 3 a seguir mostra o *ranking* das cinco maiores posições do IDHM por Unidade Federativa do Brasil no ano de 2010:

Figura 3 - Atlas do Desenvolvimento Humano no Brasil

Territorialidades	Posição IDHM	IDHM
Distrito Federal	1°	● 0.824
São Paulo	2°	● 0.783
Santa Catarina	3°	● 0.774
Rio de Janeiro	4°	● 0.761
Paraná	5°	● 0.749

Fonte: Adaptado de PNUD (2022)

Os estados que apresentam o maior índice estão classificados com IDHM Alto, como mostra a Figura 3, com exceção do Distrito Federal, que lidera o *ranking* garantindo o IDHM Muito Alto. Dessa forma o IDHM é calculado a partir dos mesmos critérios utilizados para os municípios e nações, considerando dados de renda, educação e saúde de todas as cidades do estado sendo de suma importância para avaliar as desigualdades regionais dentro do país e para orientar políticas públicas voltadas para o desenvolvimento humano em nível estadual.

Em resumo, apesar de ter sido inspirado no Índice de Desenvolvimento Humano (IDH) global, o Índice de Desenvolvimento Humano Municipal (IDHM) foi adaptado às especificidades da realidade brasileira, levando em conta as informações disponíveis no Censo e as características particulares de cada município. Por esse motivo, não é apropriado fazer comparações diretas entre o IDHM de um município e o IDH de um país, já que esses indicadores levam em conta diferentes fatores e são calculados com base em metodologias distintas (PINTO; COSTA; MARQUES, 2013).

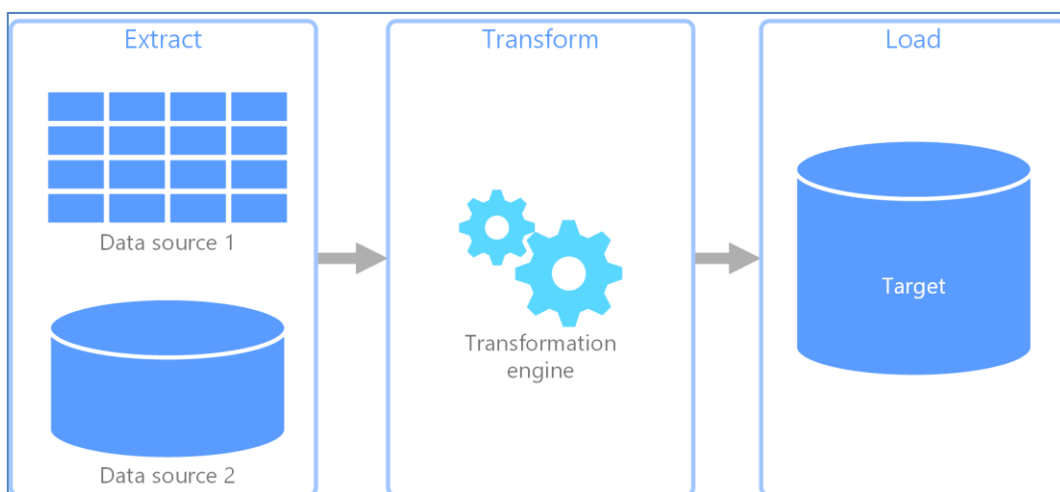
2.3 ETL

Extract, Transform, Load (ETL) é um processo de dados utilizado para coletar e transformar dados de múltiplas fontes (MICROSOFT, 2019). O objetivo principal do ETL é extrair dados de várias origens, transformá-los de acordo com as necessidades do projeto e em seguida, carregá-los em um destino final, como um sistema de armazenamento e gerenciamento de dados que integra informações de diversas fontes, Data Warehouse (DW), ou um simples banco de dados (BD). Para Kimball e Caserta (2004):

Um sistema ETL devidamente projetado extrai dados de sistemas de origens, impõe padrões de qualidade de dados e consistência, conforma dados para que fontes independentes possam ser usadas em conjunto e, finalmente, entrega em um formato que os usuários finais possam usar para tomar decisões.

De acordo com Kozielski & Wrembel (2008, p. 21), um sistema ETL é dividido em três (3) etapas ou fases: extração, transformação e carregamento dos dados. A Figura 2, contém as etapas do processo ETL.

Figura 4 - Etapas do processo ETL



Fonte: MICROSOFT (2019)

Conforme observado na Figura 4, a primeira etapa no qual compõe o processo ETL é a de extração. Segundo Kimball e Ross (2013), a etapa de extração envolve a identificação das fontes de dados, seleção dos dados a serem extraídos e a obtenção dos dados por meio de consultas, *scripts* ou ferramentas de integração de dados. É

importante considerar a capacidade das fontes de dados em relação ao volume e à velocidade de dados que devem ser extraídos.

A etapa seguinte resulta no processo de transformação, onde são aplicadas técnicas como limpeza de dados, normalização e enriquecimento de dados, conforme sugerido por Celko (2015). Além disso, é importante a validação dos dados durante a transformação, bem como a adoção de padrões e regras para garantir a consistência e qualidade dos dados.

A etapa de carregamento, fase final do processo ETL, é responsável por inserir os dados transformados no destino final, que pode ser um Data Warehouse ou um banco de dados. Segundo Kimball e Ross (2013), o carregamento de dados envolve a definição de uma estratégia de carga que considere a quantidade de dados a serem carregados e a frequência de atualização. Além disso, é importante definir uma estratégia de gerenciamento de erros para lidar com possíveis problemas durante o processo de carga.

2.4 DASHBOARD

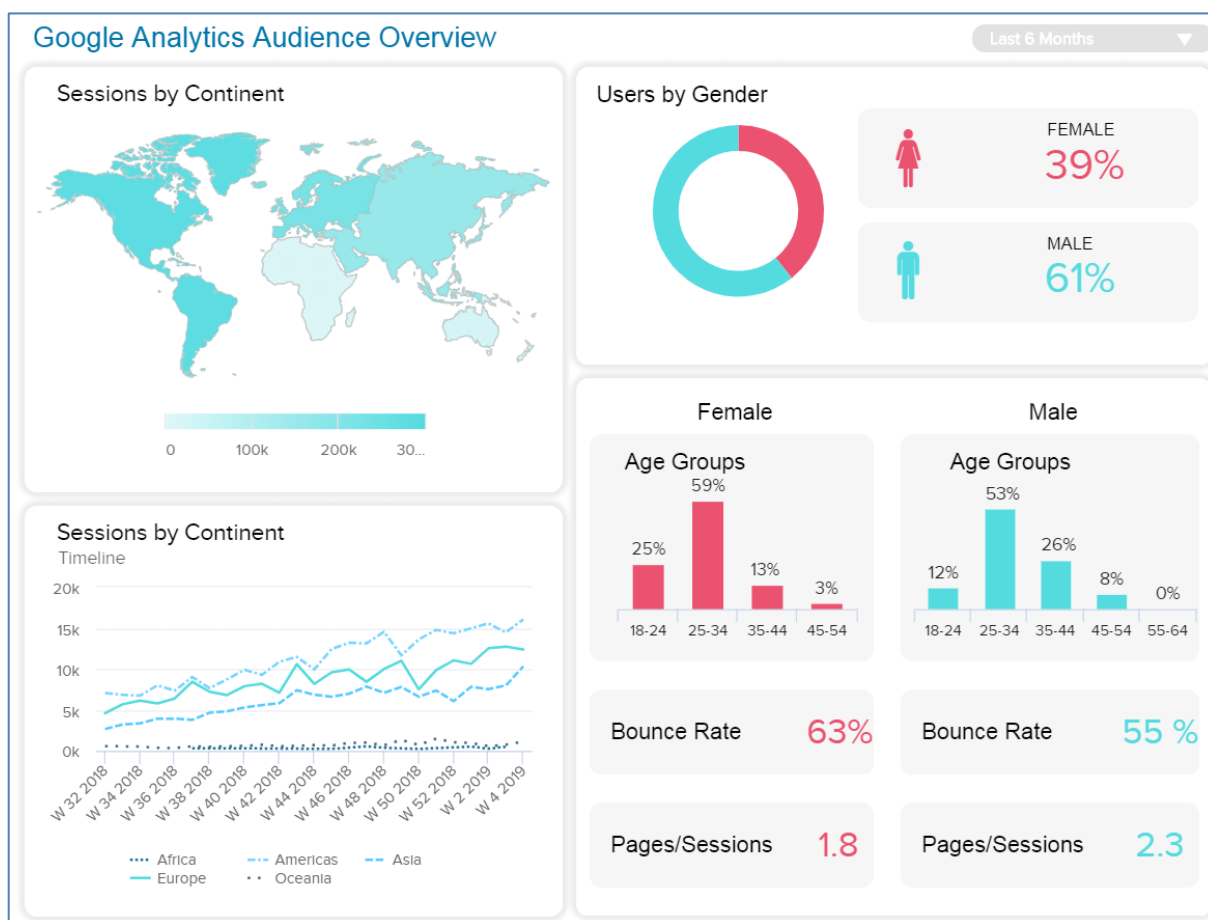
Um *dashboard* é uma ferramenta de visualização de dados que permite monitorar e analisar informações em tempo real, por meio de gráficos, tabelas, mapas e outros elementos visuais. Para Few (2006), um *dashboard* pode ser entendido como um *display* virtual contendo informações importantes para atingir um determinado objetivo. Deve ser simples, de fácil entendimento e interativo, permitindo que o usuário explore as informações de diferentes maneiras e tome decisões com base nos *insights* obtidos.

Um dos principais benefícios dos *dashboards* é a capacidade de simplificar dados complexos e apresentá-los de forma visualmente atraente. Isso ajuda a tornar as informações mais acessíveis e compreensíveis para os usuários, permitindo que tomem decisões mais informadas e baseadas em dados. Além disso, auxilia gestores a visualizar tendências, padrões e anomalias de determinado negócio, tornando o desenho da informação um fator determinante. Este também pode promover a consistência, monitorização, planeamento, comunicação, entre outros (PAUWELS et al. 2009). A consistência é relacionada às técnicas e procedimentos de mensuração apli-

cados em uma área específica de uma empresa ou unidade de negócios. A monitorização é a análise regular de indicadores que devem conduzir a ajustes, caso necessário. O planeamento pode ser empregado para simular diversos possíveis cenários futuros, através de uma análise hipotética. A comunicação consiste na exposição das distintas medidas para os envolvidos em uma organização (DUARTE, 2012).

A Figura 5 demonstra um caso exemplar de um *dashboard* que agrega informações organizadas de maneira estruturada, incluindo elementos visuais, dispostos em uma única tela, para apresentar uma coleção de dados de tal forma que o usuário possa analisar e compreender facilmente, sem necessidade de recorrer a recursos complexos como inventários, listas, planilhas e similares.

Figura 5 - Modelo de Dashboard



Fonte: Google Analytics

A Figura 5 é um modelo de *dashboard* gerado pela ferramenta Google Analytics, da Google. Nela é possível observar o uso de diversos tipos de gráficos em

formatos diversificados e uma grande variedade de cores que permite uma melhor identificação e análise dos dados disponíveis.

O desenvolvimento de um *dashboard* para o projeto é de suma importância, tendo em vista que é possível transformar os dados em informações úteis e relevantes para a sociedade. Dessa forma eles oferecem uma visualização clara e concisa dos dados, permitindo que os usuários identifiquem rapidamente padrões, tendências e anomalias. Embora a apresentação visual de informações não se restrinja apenas aos dashboards, essa ferramenta pode se comunicar com maior eficiência e com um significado mais rico do que somente o uso de textos.

2.4.1 TIPOS DE DASHBOARDS

Considerando que os *dashboards* são ferramentas que visam apresentar informações importantes e relevantes de maneira rápida, visual e simplificada. Essas ferramentas têm sido amplamente utilizadas em diferentes contextos, como em empresas, organizações e até mesmo em sistemas de gestão de saúde. Compreender os aspectos cruciais para a construção de um Dashboard é fundamental. Entre esses aspectos, destaca-se a necessidade de definir claramente os indicadores que serão medidos e identificar o público-alvo que se beneficiará da utilização dessa ferramenta (Cristina & Sanz, 2018).

De acordo com Rasmussen et al. (2009) e Eckerson (2011), há três tipos principais de dashboards:

- **Estratégico:** Usado para monitorar os principais indicadores de desempenho de uma organização em um determinado período de tempo. Geralmente contém informações como receita total, custos, lucro líquido e retorno sobre o investimento. Além disso, permite que executivos e funcionários acompanhem o progresso em direção à realização de objetivos estratégicos.
- **Táticos:** Utilizado para monitorar o progresso em relação a objetivos específicos em um determinado período de tempo. Geralmente contém informações sobre o desempenho em áreas como vendas, marketing, finanças, recursos humanos e ajudam gerentes e analistas a acompanhar e analisar as atividades, processos e projetos departamentais.

- **Operacionais:** Possibilita monitorar dados em tempo real e permite que os usuários tomem ações imediatas. Geralmente contém informações como o número de vendas diárias, o número de clientes atendidos ou o tempo médio de espera para o atendimento. Dessa forma, permite que trabalhadores da linha de frente e supervisores acompanhem os principais processos operacionais.

A Tabela 3 apresenta as principais distinções entre dashboards estratégicos, táticos e operacionais.

Tabela 2 - Tipos de Dashboards

	Estratégico	Tático	Operacional
Propósito	Executar Estratégias	Medir Progressos	Monitorizar Operações
Usuários	Executivos	Gerencial	Suporte
Âmbito	Empresarial	Departamental	Operacional
Métricas	Resultados	Ações de Resultados	Ações ou progresso
Dados	Sumarizados	Sumarizados e detalhados	Detalhados
Atualização	Mensal/Anual	Diário/ Semanal	Diário em tempo real
Foco	Gestão	Análise	Monitoria
Design	Tabela de Desempenho	Portal	Painel de Controle

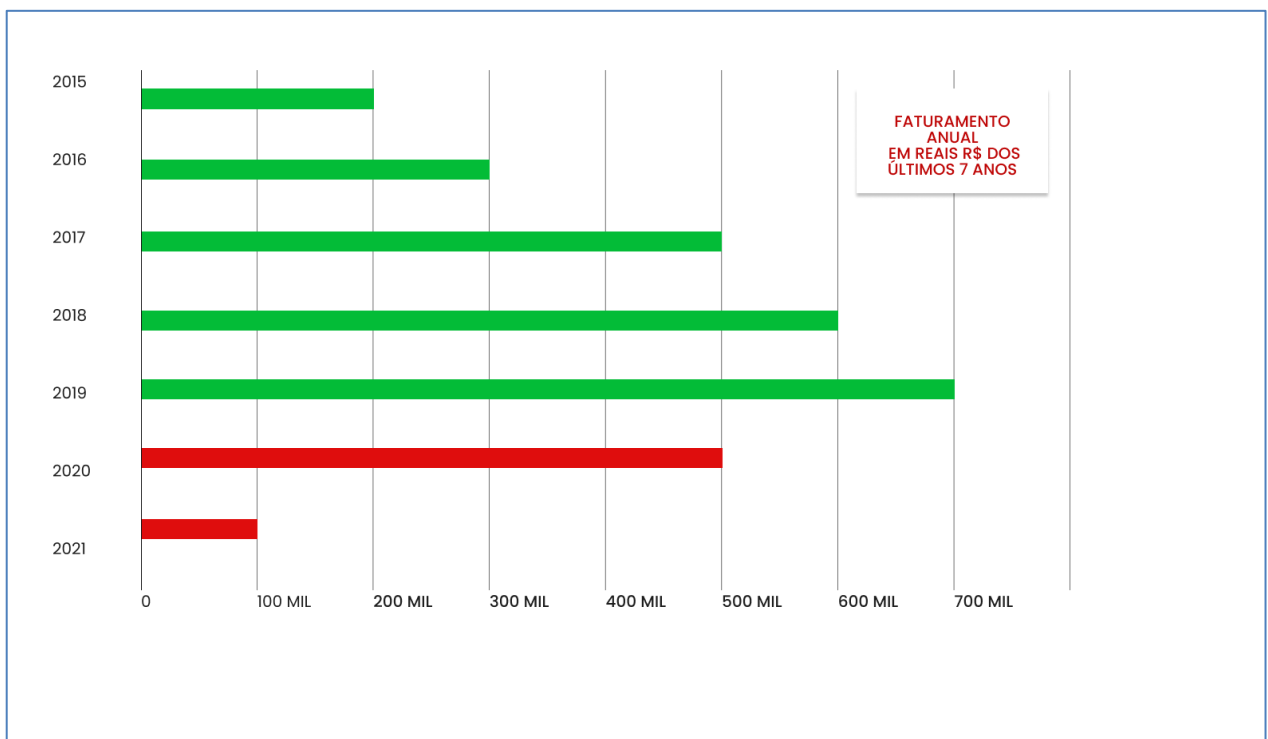
Fonte: Adaptado de Eckerson (2009)

A Tabela 2 apresenta as principais diferenças entre os três tipos de *Dashboards*. Para isso é necessário analisar os componentes que compõem cada modelo. Conforme Rasmussen et al., (2009) é importante relacionar cada tipo de *dashboard* com as pessoas da organização, já que as categorias de cada painel retratam o nível de responsabilidade de cada pessoa e o seu tempo dedicado à análise de dados.

2.4.1.1 DASHBOARD ESTRATÉGICO

O *dashboard* estratégico permite através de seu visual que os gestores executivos realizem estratégias definidas, monitorando a execução dos objetivos que refletem os comportamentos da organização. Conforme a observação de Few (2006), *dashboards* desse tipo não requerem a atualização imediata dos dados e podem ser aprimorados com o uso de informações históricas armazenadas no decorrer do tempo. Com isso, a tendência é que, quanto mais simplificada for a exposição das informações no dashboard, mais facilmente o usuário será capaz de assimilar o conteúdo apresentado. Um exemplo de aplicação desse tipo de dashboard é ilustrado na Figura 6.

Figura 6 - Exemplo de Dashboard Estratégico



Fonte: Elaboração Própria

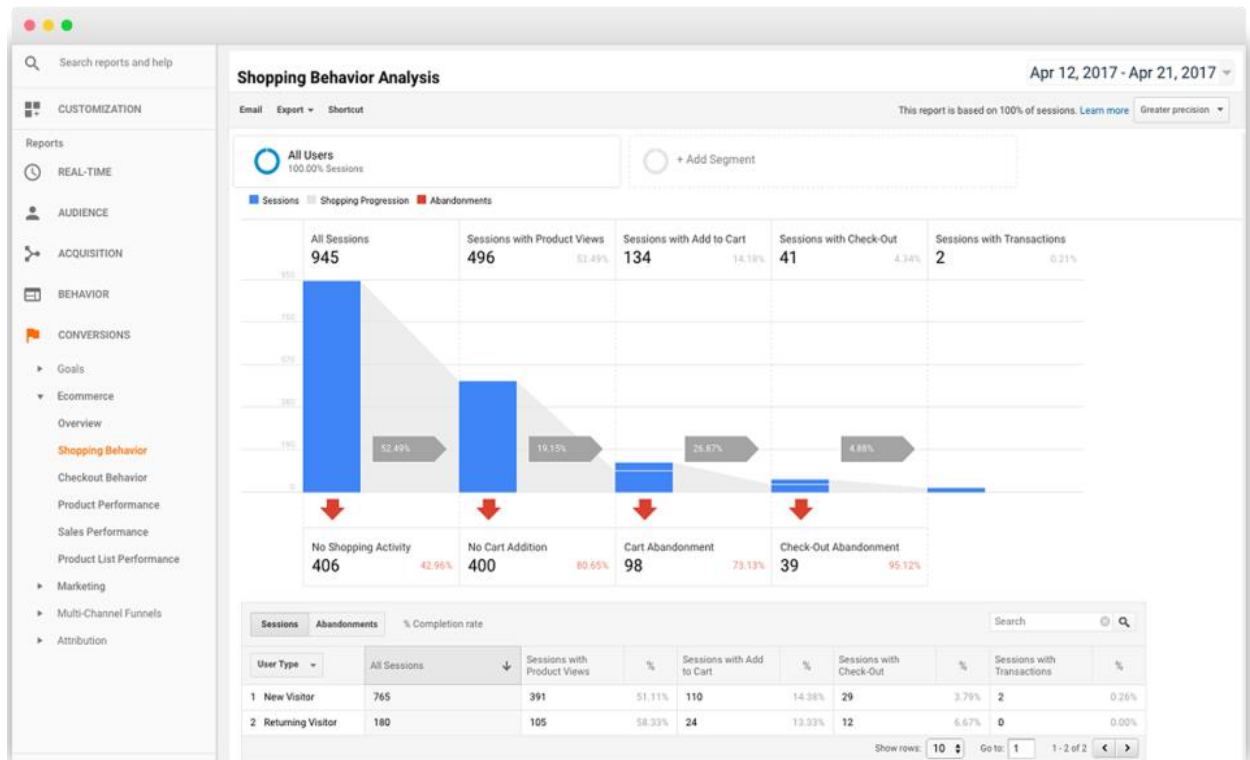
No exemplo de *dashboard* estratégico demonstrado na Figura 6, o gestor tem acesso a informações finais referentes ao quantitativo de vendas efetuadas nos últimos 7 anos que vão desde 2015 até o ano de 2021. Como pode ser observado no gráfico, as informações são apresentadas de maneira nítida, facilitando o processo de tomada de decisão. A cor verde mostra um crescimento significativo das vendas até o ano de 2019. Já a cor vermelha a partir do ano de 2020, mostra números de venda decrescentes.

2.4.1.2 DASHBOARD TÁTICO

Um *dashboard* tático é projetado para apresentar informações resumidas e detalhadas, com o objetivo de ajudar a identificar problemas, encontrar soluções e, por consequência, assegurar que os objetivos de longo e curto prazo sejam alcançados (Eckerson, 2009). Esses dashboards são projetados para fornecer informações úteis e relevantes que ajudam na tomada de decisões estratégicas em diferentes áreas, como marketing, vendas, finanças, recursos humanos e operações.

Gestores de negócios podem usar os *dashboards* táticos para medir seu progresso em relação a metas e objetivos predefinidos, bem como para realizar análises mais aprofundadas, caso necessário. Além disso, são atualizados diariamente ou semanalmente, permitindo que os gerentes e analistas absorvam e analisem tendências e ajam de acordo com as necessidades. A figura 7 mostra um exemplo viável de *dashboard* para esse determinado contexto.

Figura 7 - Dashboard Tático



Fonte: Google Analytics

Na Figura 7 é apresentado um exemplo de *dashboard* tático, feito pela ferramenta Google Analytics para visualizar indicadores de *marketing*, como: dados, acessos ao site, ou até mesmo resultado de vendas on-line com indicadores sobre taxas de conversão e mapeamento de origem de tráfego. De acordo com Few (2006), quando se utiliza um dashboard desse tipo, o usuário tende a realizar uma análise mais minuciosa, que não demanda interações constantes com os painéis, mas que requer um entendimento profundo dos métodos mais eficazes para analisar os dados. Isso implica em possuir conhecimento detalhado sobre os dados e o funcionamento do próprio dashboard.

2.4.1.3 DASHBOARD OPERACIONAL

Por fim, o operacional possibilita o monitoramento e controle dos principais processos da organização em tempo real, além de alertas configurados caso algum indicador ultrapasse um limite definido. Segundo Few (2006), os dashboards operacionais precisam ter um design distinto em relação aos dashboards estratégicos e analíticos devido à sua natureza dinâmica e ao uso frequente, o que requer atualização em tempo real dos dados. Diferentemente dos dashboards estratégicos, o *design* dos

dashboards operacionais deve ser simplificado, mas ainda assim possuir elementos que permitam alertar prontamente o usuário caso haja alguma anomalia no fluxo de dados. A Figura 8 apresenta um exemplo de um *dashboard* operacional.

Figura 8 - Dashboard Operacional



Fonte: OPSERVICES (2017)

No exemplo da figura 8 é possível acompanhar métricas de uma equipe de *Service Desk*, podendo visualizar o trabalho de cada analista. Além disso, nota-se informações importantes, como sobrecarga da equipe e até mesmo chamados críticos acumulados. Os *dashboards* operacionais apresentam geralmente métricas que precisam ser monitoradas para garantir um bom desempenho das atividades da empresa. Por tanto, as métricas que ajudam os analistas a identificar e corrigir problemas e falhas nos processos empresariais são visualizadas nos indicadores deste tipo de *dashboard*.

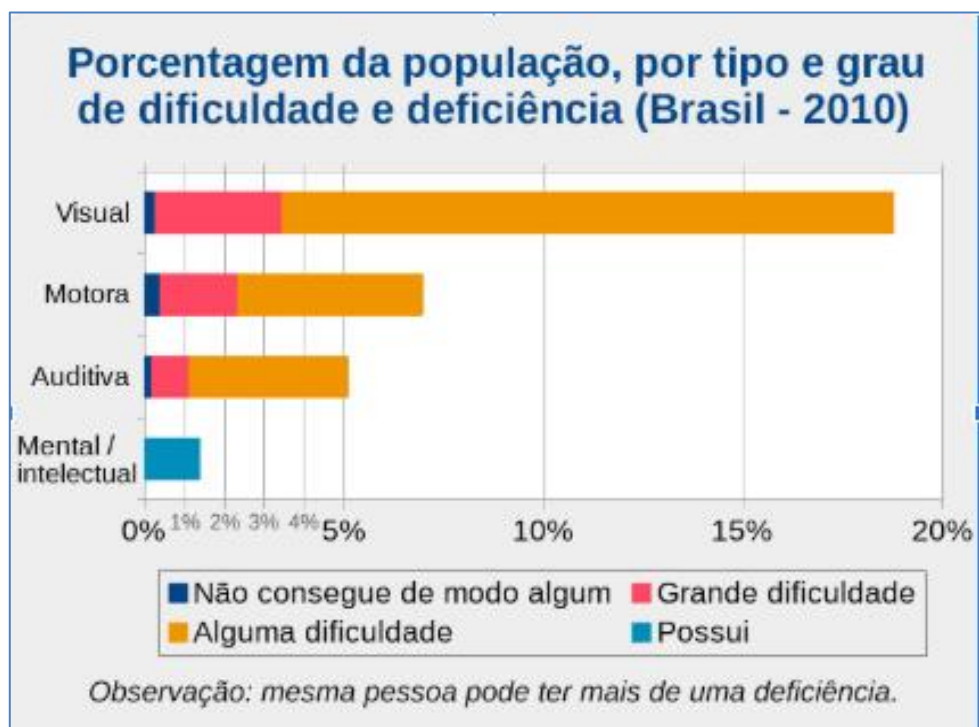
2.4.2 TIPOS DE GRÁFICOS

Os *dashboards* são compostos por vários elementos, dependendo do público-alvo e do objetivo, tais como gráficos, tabelas, mapas e indicadores de alerta. De acordo com Barros (2013), os gráficos são amplamente utilizados em *dashboards*, principalmente porque permitem representar uma grande quantidade de informações em um espaço limitado.

Uma das razões pelas quais os gráficos são tão populares é sua capacidade de apresentar informações de maneira fácil de ser absorvida pelos usuários, utilizando recursos como cores, escalas, dimensões, áreas, direções, ângulos, volumes, curvaturas e sombreados. Além disso, os gráficos também ajudam a destacar informações importantes e a simplificar a comunicação dos resultados e conclusões. Dessa forma os usuários podem facilmente comparar diferentes métricas, medir o progresso em relação a metas e objetivos, e tomar decisões mais informadas e embasadas em dados concretos. Dentre os diversos tipos de gráficos disponíveis, a seguir são mencionados alguns dos mais utilizados:

- **Gráfico de barras:** consiste em um tipo de gráfico amplamente utilizado para comparar quantidades ou valores entre diferentes categorias. Esse tipo de gráfico é formado por barras retangulares verticais ou horizontais, que representam os valores numéricos ou percentuais de cada categoria. A Figura 9 a seguir mostra uma exemplificação de um gráfico de barras.

Figura 9 - Exemplo de gráfico de barras

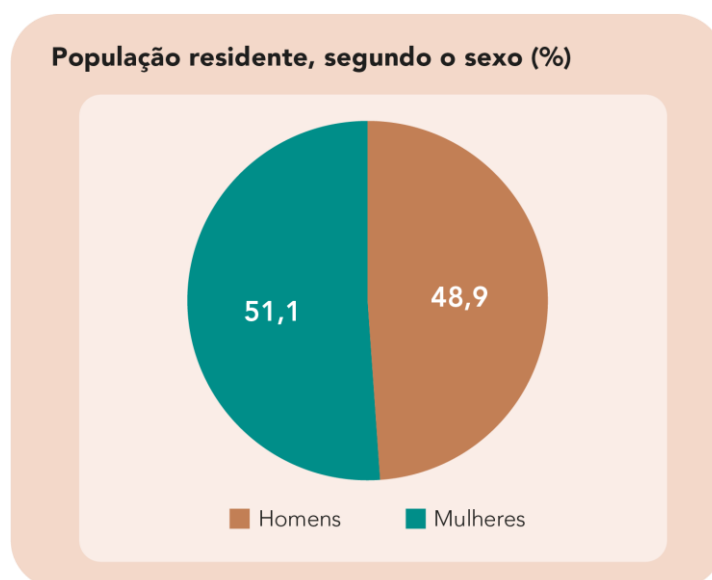


Fonte: IBGE, Censo Demográfico, 2010

A Figura 9 apresenta a utilização de um gráfico de barras, no qual indica a porcentagem da população, por tipo e grau de dificuldade e deficiência no Brasil no ano de 2010. O comprimento de cada barra é proporcional ao valor que representa, enquanto as barras são separadas umas das outras para permitir a distinção entre as diferentes categorias. Os gráficos de barras são particularmente úteis quando as categorias a serem comparadas são facilmente identificáveis e não muito numerosas, possibilitando uma comparação visual clara e fácil entre elas.

Gráfico de setores: também conhecido como gráfico de pizza ou de rosca, é uma representação visual de dados que utiliza setores circulares para mostrar a proporção de diferentes categorias em relação ao todo. A Figura 10 mostra um exemplo de um gráfico de setores.

Figura 10 - Exemplo gráfico de setores



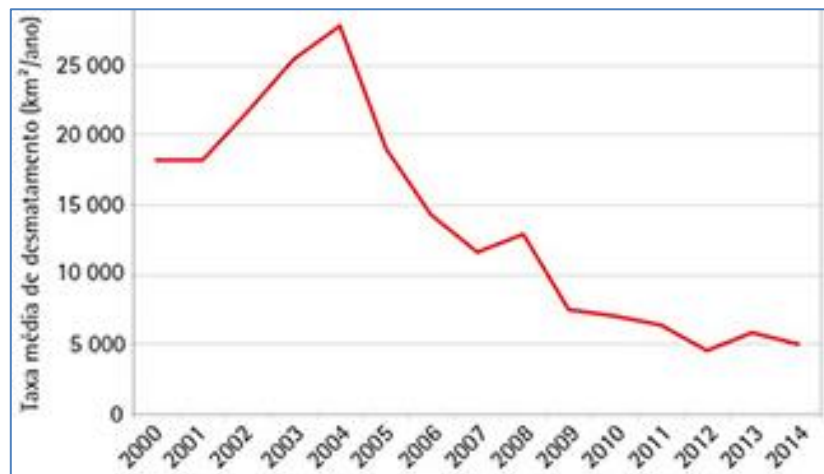
Fonte: IBGE (PNAD 2021)

A Figura 10 mostra um exemplo de um gráfico em setores no qual representa a porcentagem da população residente conforme o sexo no ano de 2021. Este tipo de gráfico é útil em um dashboard porque permite ao usuário visualizar rapidamente a distribuição de dados em um conjunto de categorias.

Quando um gráfico de setores é usado para comparar categorias, é importante que as proporções sejam claramente visíveis e diferenciadas. Se duas ou mais categorias tiverem proporções muito próximas, pode ser difícil distinguir entre elas no gráfico de setores. Além disso, é importante que a soma de todas as proporções seja igual a 100% ou a 360 graus, ou seja, um círculo completo. Isso garante que o gráfico de setores seja fácil de ler e interpretar.

Gráfico de linha: é uma representação gráfica que ilustra a conexão entre duas variáveis, utilizando informações numéricas ou quantitativas. É constituído por um sistema de coordenadas cartesianas com um eixo vertical e outro horizontal, localizado no primeiro quadrante. Além disso, apresenta uma ou mais linhas que evidenciam as flutuações dos dados ao longo do tempo, permitindo assim observar a evolução dos dados ao longo da pesquisa. A figura 11 a seguir, mostra um exemplo de um gráfico de linhas.

Figura 11 - Exemplo gráfico de linhas

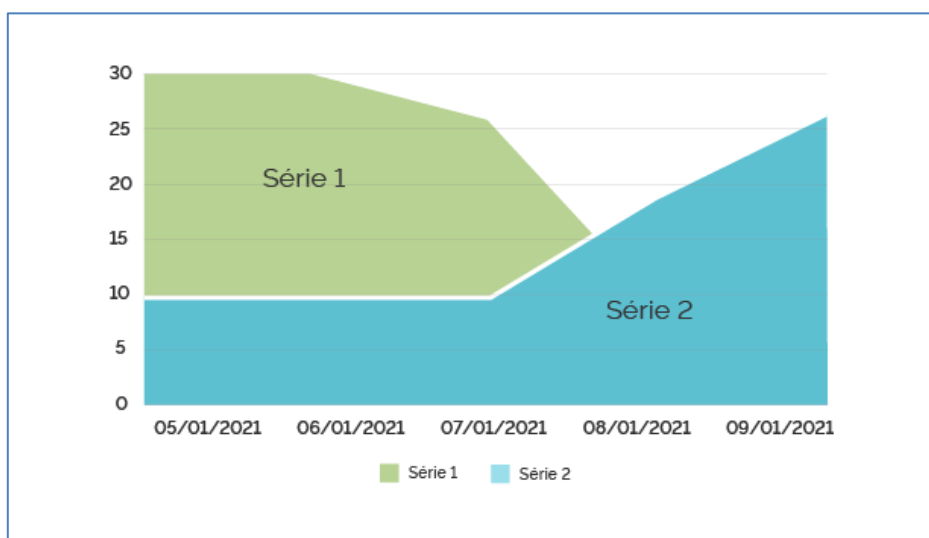


Fonte: IBGE (2016)

A Figura 11 mostra um exemplo de um gráfico de linhas, demonstrando a taxa média de desmatamento da floresta amazônica entre os anos de 2000 e 2014. É possível identificar a existência de duas variáveis. Uma variável refere-se a qualquer elemento que possui a capacidade de sofrer alterações, como valores, taxas percentuais, intervalos de tempo, entre outros. Essas variáveis são representadas nos eixos X e Y do gráfico, sendo o eixo horizontal localizado na parte inferior com os anos da pesquisa e o eixo vertical posicionado à esquerda demonstrando a taxa de desmatamento.

Gráfico de área: é uma combinação entre um gráfico de linhas, e um gráfico de barras, com a função de auxiliar na visualização de alterações das quantidades de uma variável ao longo do tempo. A figura 12, mostra um exemplo de um gráfico de área.

Figura 12 - Exemplo gráfico de área



Fonte: GOV.BR – Padrão Digital de Governo

A Figura 12 representa um exemplo de um gráfico de setores, usado para representar valores numéricos, em um determinado espaço de tempo. É possível identificar as evoluções e diminuições do determinado fenômeno. Desta forma, a utilização do gráfico de áreas deve ser estabelecida quando há necessidade de análises quantitativas, como um faturamento anual ou custo médio mensal de empresas, além disso, pode ser usado na comparação de dados desde que os eixos sigam as mesmas escalas.

Existem diversos elementos gráficos além dos mencionados que ajudam a compor ferramentas de análise e visualização de dados. Em *dashboards* operacionais, é comum incluir cronômetros e temporizadores, pois os dados são atualizados com frequência. Outros elementos que podem ser utilizados na composição de um dashboard incluem calendários, mapas, diagramas, fluxogramas e linhas do tempo.

2.4.3 PERCEPÇÃO VISUAL

Segunda Few (2007), para aproveitar ao máximo o potencial dos *dashboards*, é importante entender e aplicar a percepção visual e o funcionamento do cérebro humano. Este é um fator crucial para apresentar uma grande quantidade de informações de forma clara em um espaço limitado. É necessário compreender como as pessoas processam visualmente as informações para criar painéis de dados eficazes, capazes de comunicar rapidamente com um grande volume de dados. Dessa forma, para alcançar o sucesso dos *dashboards*, é preciso ter um conhecimento profundo sobre o processo de percepção visual e aplicá-lo na apresentação de informações.

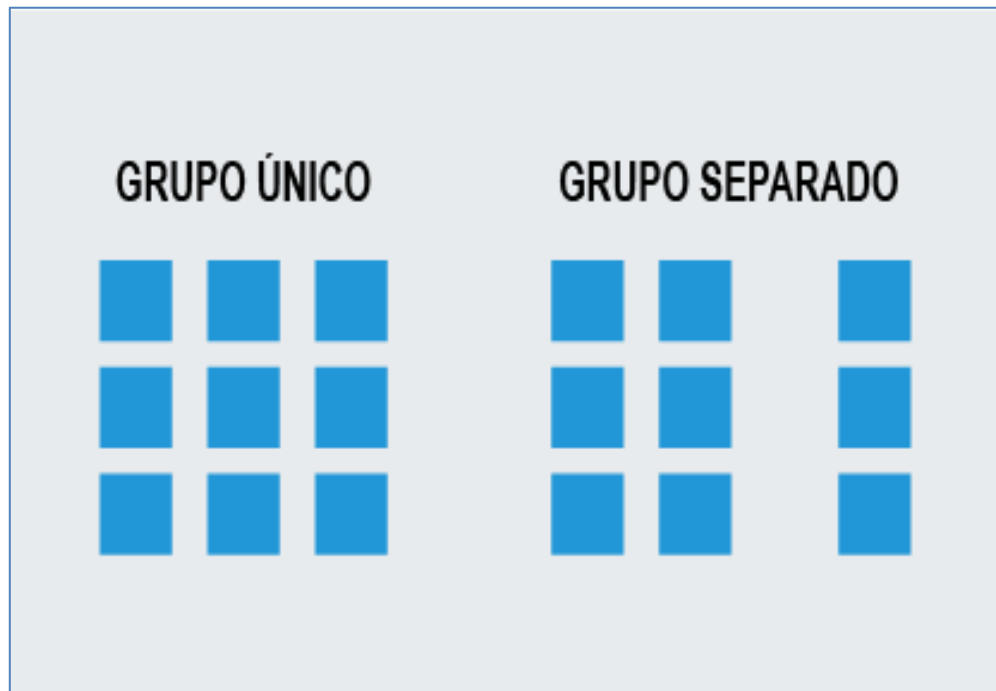
A percepção visual é o processo pelo qual o cérebro interpreta as informações visuais recebidas pelos olhos e nos permite reconhecer objetos, formas, cores, profundidade, movimento e outras características do ambiente ao nosso redor. Em resumo, o processo de percepção envolve a recepção de informações visuais pelos olhos, a transmissão dessas informações ao cérebro, a organização e interpretação dessas informações e a atribuição de significado a elas.

As ferramentas de visualização, incluindo os *dashboards*, são baseadas nos princípios da percepção visual. A percepção visual é entendida aplicando os princípios da psicologia *Gestalt* à visualização (DUARTE, 2012). O princípio de *Gestalt* é um conceito originado dos estudos em psicologia e filosofia do século XIX que aborda a forma como os seres humanos interpretam visualmente os elementos e é baseada na ideia de que a percepção humana não é uma simples soma das partes, mas é influenciada pela organização dos elementos perceptivos em uma totalidade significativa, chamada de forma ou figura. Sendo assim, existem sete princípios de organização dos elementos visuais conhecidos como Princípios da *Gestalt* que auxiliam na compreensão e no foco dos usuários, permitindo que percebam rapidamente os componentes principais e necessários de uma interface. Esses princípios incluem: proximidade, similaridade, continuidade, fechamento, figura-fundo, região comum e ponto focal. Para o desenvolvimento do projeto, foi utilizado somente três desses princípios:

1. Proximidade: sustenta que objetos ou elementos que estão próximos uns dos outros são percebidos como mais relacionados do que se estivessem distantes. Essa noção

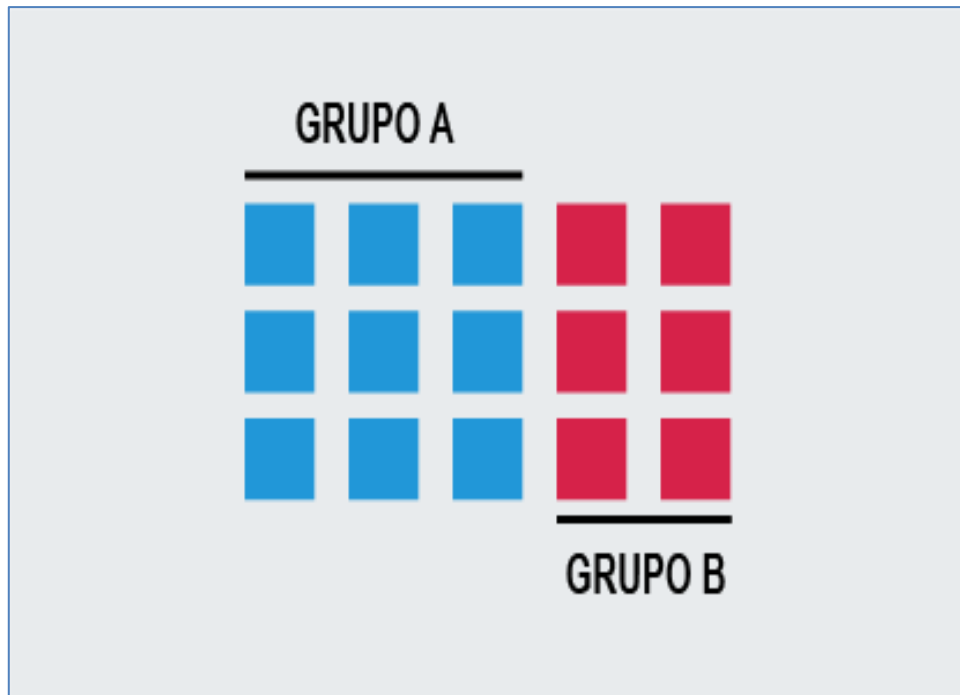
pode ser observada facilmente em diversas situações cotidianas através da percepção visual. Este princípio pode ser observado na figura 13 a seguir.

Figura 13 - Princípio da Proximidade



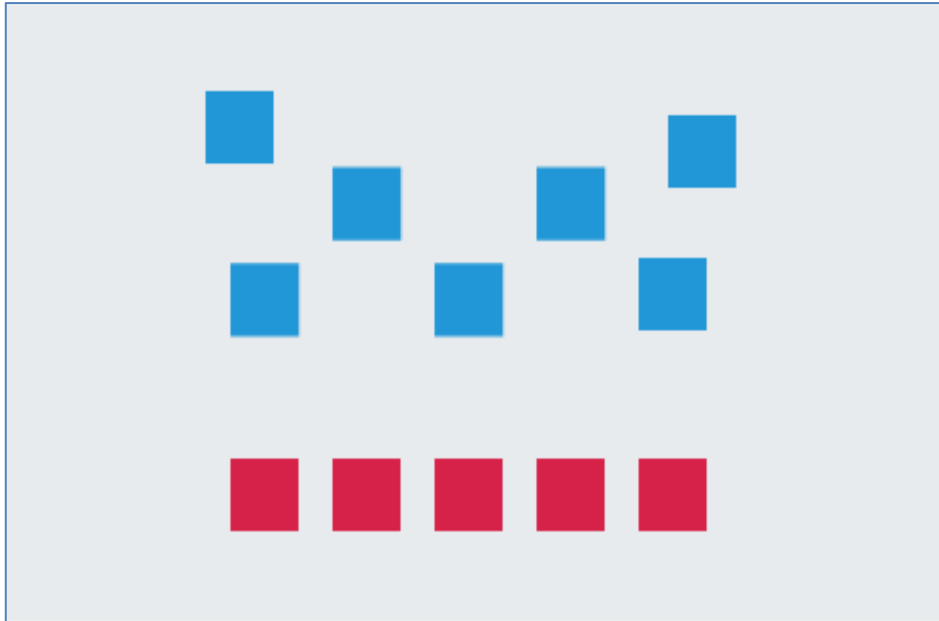
Na Figura 13 ambas as imagens apresentam os mesmos elementos, mas a alteração na proximidade resulta em uma visualização distinta dos grupos e das relações. No *design* de interface de um *dashboard* não é diferente, este princípio se encontra na forma com que os elementos distintos são posicionados de forma próxima para formar grupos.

- 2. Similaridade:** baseia-se na ideia de que elementos que apresentam semelhanças são interpretados como pertencentes a um mesmo conjunto e desempenhando uma função semelhante. Na Figura 14 a seguir, tem-se um exemplo de como funciona o princípio da similaridade.

Figura 14 - Princípio da Similaridade

A Figura 14 mostra um conjunto de elementos, no entanto a variação nas cores faz com que sejam percebidos como um grupo distinto. Dentro do contexto de *design* da interface de um *dashboard* esse princípio pode ocorrer com diversos elementos gráficos da composição de um painel de visualização, como itens de menu, filtros de informação e até mesmo ícones e logotipos.

3. Continuidade: quando elementos são dispostos em uma linha ou curva, tendemos a percebê-los como relacionados entre si, mais do que se estivessem dispostos de outras formas. Em outras palavras, há uma tendência natural de nosso cérebro em ver elementos que se alinham em uma direção contínua como uma unidade coesa, mesmo que eles estejam parcialmente obscurecidos ou interrompidos. É possível observar um exemplo do princípio de continuidade na Figura 15 a seguir.

Figura 15 - Princípio da Continuidade

Os elementos dispostos em vermelho e azul retratados na Figura 15 são percebidos como objetos relacionados entre si, pois estão alinhados em uma direção contínua e coesa. O princípio da continuidade pode ser aplicado em um *dashboard* quando há a necessidade de preencher informações, como os filtros de resultados que podem estar dispostos de forma alinhada, influenciando o usuário a preencher os requisitos de maneira sequencial devido à linearidade em que os elementos estão colocados.

3 METODOLOGIA

Este trabalho é caracterizado como uma pesquisa aplicada, voltado para o contexto de Sistemas de Informação, tendo como meta a aplicação de processos e métodos para a construção de um *dashboard* interativo integrado a um grande volume de dados relativos ao IDH brasileiro. O procedimento metodológico é bibliográfico, uma vez que este trabalho é orientado sobre o conceito teórico de *dashboards* e microdados. Ademais, torna-se experimental, visto que se utiliza softwares e técnicas computacionais para o desenvolvimento da ferramenta proposta.

3.1 MATERIAIS

Para o desenvolvimento do projeto proposto, foram utilizadas as seguintes ferramentas:

Anaconda: é um software gratuito e de código aberto desenvolvido pela empresa Anaconda, Inc. que fornece um ambiente completo para a análise de dados em Python. (ANACONDA, 2022). Ele inclui uma distribuição do Python, bibliotecas científicas populares, como o NumPy, Pandas, e muitas outras, além de uma variedade de ferramentas e utilitários para análise de dados, como o Jupyter Notebook.

Jupyter Notebook: consiste em uma ferramenta no qual se pode criar cadernos de anotações, que por sua vez são constituídos por um ambiente de programação que mistura blocos de códigos executáveis, visualizações e textos enriquecidos (Schuch, 2022). Apesar de não ser necessariamente uma ferramenta de ETL, é possível realizar o processo de extração, transformação e carregamento, já que a ferramenta permite uma fácil manipulação dos dados e facilita na extração, limpeza e a mescla do mesmo, além de possibilitar a visualização de gráficos para identificação de padrões e tendências dos dados extraídos.

Figma: consiste em uma plataforma colaborativa para construção de design de protótipos, com ela, designers e outros profissionais têm a chance de construir o *design* de produtos digitais por completo, como *sites*, aplicativos etc. (ALURA, 2023). Ao utilizar essa ferramenta, os usuários têm acesso a recursos e funcionalidades que facili-

tam o processo de design, desde a criação de *wireframes* até a prototipagem interativa. Dessa forma, o Figma se torna uma ferramenta essencial no desenvolvimento do projeto, pois permite uma visualização prévia em termos de *design* e *layouts* do *dashboard* final.

Firestore: o *Firestore* Realtime Database é um banco de dados NoSQL hospedado na nuvem que permite armazenar e sincronizar dados em tempo real entre os usuários. Com a sincronização feita em tempo real, o *Firestore* permite que os usuários acessem os dados em qualquer lugar, na *Web* ou dispositivo móvel. (FIREBASE, 2023). O *Firestore* também oferece recursos avançados, como o Cloud Functions, que permite executar código de *back-end* em resposta a eventos específicos, como mudanças nos dados do banco de dados. Será uma ferramenta essencial para o desenvolvimento no projeto, já que irá armazenar os dados que serão visualizados no *dashboard*.

Flutter: é uma plataforma de código aberto criada pelo Google, que permite a criação de aplicativos multiplataforma com design atraente e com a possibilidade de compilar nativamente a partir de uma única base de código (GOOGLE, 2023). A principal vantagem do *Flutter* é sua abordagem de desenvolvimento centrada no *widget*, que permite que os desenvolvedores construam interfaces de usuário com rapidez e facilidade, e que também proporciona uma experiência de usuário fluida e consistente.

Dart: *Dart* é uma linguagem orientada a objetos e otimizada para aplicativos rápidos em qualquer plataforma, além disso é gratuito e de código aberto (GOOGLE, 2023). Ela é usada como a linguagem principal para desenvolver aplicativos usando o *framework Flutter*.

Pandas: o objetivo principal do *Pandas* é fornecer uma base sólida e de alto nível para realizar análises de dados do mundo real em *Python*, com uma abordagem prática. Além disso, o *Pandas* tem uma meta mais abrangente: tornar-se a ferramenta de código aberto mais poderosa e flexível para análise e manipulação de dados, superando qualquer outra disponível em outras linguagens (PANDAS, 2023).

3.2 MÉTODOS

Os processos utilizados para o desenvolvimento desse projeto estão representados em um fluxograma de etapas.

Figura 16 - Metodologia do Projeto



Como pode ser visualizado, a Figura 16 representa a organização das etapas realizadas para o processo de desenvolvimento do projeto. Está detalhada da seguinte maneira:

- 1. Busca das Bases de Dados:** Nesta primeira etapa, foi realizado uma busca e levantamento das bases de dados providas de fontes do IBGE e relativas ao IDH brasileiro. Os microdados fornecidos pelo IBGE, proporcionam a definição de diversos indicadores necessários para a conclusão do projeto, sendo eles: educação, saúde e renda da população. Buscando relacionar esses indicadores, a etapa seguinte se resulta na coleta e seleção dos dados essenciais para a conclusão do projeto.
- 2. Coleta e Seleção dos Microdados:** O objetivo desta etapa é filtrar e selecionar os microdados mais relevantes, organizando os atributos e parâmetros de cada variável providos de diversos períodos e normalizando-os a fim de se obter uma estrutura de dados padronizada. A busca somada a coleta e seleção dos

microdados forneceu um conhecimento para a realização das etapas subsequentes do projeto.

- 3. Aplicação do Processo ETL:** Etapa no qual foi realizado o processo de Extração, Transformação e Carregamento (ETL) dos microdados. Extração dos dados mais relevantes, transformação e integração dos dados de maneira padronizada e normalizada.
- 4. Prototipagem:** Durante esta etapa foi desenvolvido um protótipo de *dashboard* com o objetivo de auxiliar e estruturar todos os tópicos realizados anteriormente. Para Rogers e Sharp (2005) os protótipos são versões limitadas do produto final. Já Tim Brown (2010), define o protótipo como algo tangível que serve para desenvolver uma ideia e analisá-la.
- 5. Desenvolvimento do Dashboard:** Nesta etapa decorreu-se a implementação do *dashboard*. Para isso, os materiais utilizados na subseção 3.1 serviram de instrumento para o desenvolvimento deste projeto.
- 6. Integração com a Base de Dados:** Etapa no qual foi utilizado uma ferramenta capaz de armazenar, sincronizar e consultar os dados, com o objetivo de integrá-los e serem visualizados corretamente no *dashboard*.
- 7. Execução dos Testes Funcionais:** Por se tratar da última etapa do processo de desenvolvimento, esta consistiu em certificar que a interface estivesse em correto funcionamento, e atendendo os requisitos de funcionalidade. Através dos dados inseridos é possível observar os resultados apresentados e julgá-los se eles estão de acordo com as especificações pré-estabelecidas, podendo assim serem alterados.

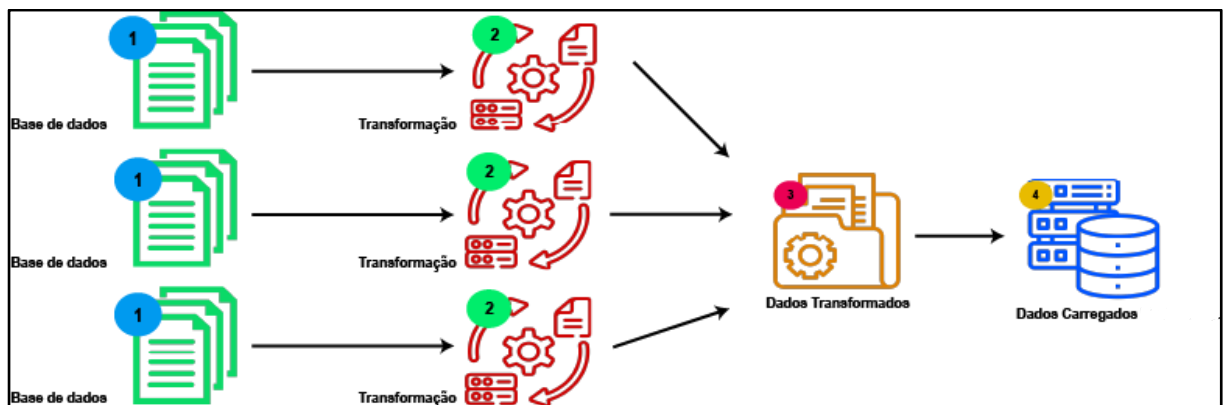
4 RESULTADOS

A finalidade deste projeto foi desenvolver um *dashboard* que pudesse exibir os dados que compõem o IDH brasileiro através de microdados do IBGE, no qual foram previamente processados. Nesta seção, serão apresentados os resultados alcançados, que consistiu, a aplicação do processo ETL, a criação de um banco de dados para armazenar os dados processados, e por fim, a visualização dos dados por meio de um *dashboard*.

4.1 PROCESSO ETL

A imagem a seguir ilustra as etapas executadas para evidenciar o processo de extração, transformação e carga.

Figura 17 - Processo ETL executado



A Figura 17 exibe as etapas realizadas durante a execução da aplicação do processo ETL responsável pela preparação dos microdados coletados. Na etapa (1), o processo inicia com a descoberta de dados provenientes das bases de dados do IBGE, Censo e PNAD. As bases de dados relacionadas à Taxa de analfabetismo de 18 anos ou mais, Renda per capita, Extrema pobreza, Esperança de vida ao nascer e IDHM dos estados brasileiros são inseridas na ferramenta *Jupyter Notebook*. Após a inserção definida na etapa (2), as bases de dados foram transformadas e organizadas. Na etapa (3) seguinte, são geradas as tabelas agregadas com os dados necessários para serem inseridas e salvas no banco de dados representado na figura (4). O banco utilizado para armazenar os dados foi o Firebase.

4.1.1 EXTRAÇÃO

Na etapa de extração, os dados obtidos estão disponíveis na plataforma do Sistema IBGE de Recuperação Automática (SIDRA), no qual disponibiliza diversas bases de dados, provindas de diversas pesquisas, como Censo e PNAD. É possível extrair dados relativos à saúde, educação, renda, censo demográfico etc.

Figura 18 - Base de dados IBGE



Fonte: SIDRA IBGE

A Figura 18 detalha as diversas bases de dados disponíveis na plataforma SIDRA, tendo em vista que todas são frutos de diversas pesquisas realizadas ao longo dos anos, e que são essenciais para realizar métricas estatísticas da população brasileira, a fim de serem usadas para pesquisas acadêmicas, governamentais, empresariais e pessoais. A plataforma SIDRA permite ao usuário fazer consultas, gerar tabelas e gráficos, além de realizar *downloads* de dados em formatos como XLSX e CSV.

As bases de dados selecionadas foram inseridas na ferramenta *Jupyter Notebook*, e tratadas separadamente. A imagem a seguir, mostra a inserção da primeira base por meio de código e a visualização da tabela e suas colunas.

Figura 19 - Base de dados Taxa de analfabetismo 18 anos ou mais

```
In [8]: taxaanalfabetismo_df = pd.read_csv("taxaanalfabetismo.csv", sep=';', low_memory=False)
taxaanalfabetismo_df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 29 entries, 0 to 28
Data columns (total 26 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Territorialidades                     29 non-null     object
1   Analfabetismo 2017                    29 non-null     object
2   Analfabetismo 2018                    28 non-null     object
3   Analfabetismo 2019                    28 non-null     object
4   Analfabetismo 2020                    28 non-null     object
5   Analfabetismo 2021                    28 non-null     object
```

Na Figura 19 é possível visualizar os atributos e tipos da base de dados coletada que são relativas às métricas de porcentagem de Analfabetismo, filtradas do ano de 2017 até 2021 por estado da Federação. Além disso, é possível identificar atributos do analfabetismo por raça e gênero no mesmo intervalo de ano e de cada territorialidade. Para garantir que o objetivo seja realizado com sucesso, há atributos com maior índice de relevância, que facilitam no processo de filtragem e mesclagem das tabelas no processo de transformação e que devem ser relacionados corretamente para a etapa seguinte.

Para dar continuidade no processo de extração, a segunda base de dados aplicada foi a base de: Renda per capita, representada na Figura 20 a seguir.

Figura 20 - Base de dados Renda per capita

```
In [30]: renda_df = pd.read_csv("renda.csv", sep=';', low_memory=False)
renda_df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 28 entries, 0 to 27
Data columns (total 26 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Territorialidades                    28 non-null     object
1   Renda 2017                           28 non-null     object
2   Renda 2018                           28 non-null     object
3   Renda 2019                           28 non-null     object
4   Renda 2020                           28 non-null     object
5   Renda 2021                           28 non-null     object
6   Renda Brancos 2017                   28 non-null     object
7   Renda Brancos 2018                   28 non-null     object
8   Renda Brancos 2019                   28 non-null     object
9   Renda Brancos 2020                   28 non-null     object
10  Renda Brancos 2021                   28 non-null     object
11  Renda Negros 2017                    28 non-null     object
12  Renda Negros 2018                    28 non-null     object
13  Renda Negros 2019                    28 non-null     object
14  Renda Negros 2020                    28 non-null     object
15  Renda Negros 2021                    28 non-null     object
16  Renda Homem 2017                     28 non-null     object
17  Renda Homem 2018                     28 non-null     object
18  Renda Homem 2019                     28 non-null     object
19  Renda Homem 2020                     28 non-null     object
20  Renda Homem 2021                     28 non-null     object
21  Renda Mulher 2017                    28 non-null     object
22  Renda Mulher 2018                    28 non-null     object
23  Renda Mulher 2019                    28 non-null     object
24  Renda Mulher 2020                    28 non-null     object
25  Renda Mulher 2021                    28 non-null     object
dtypes: object(26)
memory usage: 5.8+ KB
```

A Figura 20 mostra a inserção da base de dados relativa à renda per capita da população brasileira em real, com os atributos de renda, sendo filtrados no período de 2017 a 2021. Há também atributos relativos à renda, que são filtradas por raça e gênero, além da territorialidade representando cada estado da Federação. É possível notar a distribuição da tabela na Figura 21 a seguir.

Figura 21 - Atributos de renda per capita

```
In [68]: renda_df.head()
```

```
out[68]:
```

	Territorialidades	Renda 2017	Renda 2018	Renda 2019	Renda 2020	Renda 2021
0	Brasil	774,72	802,94	814,3	779,13	723,84
1	Acre	471,65	536,77	509,57	512,47	471,54
2	Alagoas	405,8	421,03	419,89	443,65	404,28
3	Amapá	581,52	506,21	501,53	494,32	451,27
4	Amazonas	523,38	470,59	479,55	478,39	432,99

5 rows × 26 columns

A Figura 21 representa o início da tabela de Renda per capita, filtrada por Territorialidade. A primeira linha é definida por dados gerais de todo o país, seguido dos estados da Federação e separadas por anos que vão de 2017 até 2021.

Dando seguimento a parte de extração, foram utilizadas mais três bases de dados: Esperança de vida ao nascer, Extrema Pobreza e por fim o IDHM dos estados brasileiros. A Figura 22 a seguir mostra a inserção da base de dados relativa a Esperança de vida ao nascer.

Figura 22 - Base de dados Esperança de vida ao nascer

```
In [31]: espvida_df = pd.read_csv("espvida.csv", sep=';', low_memory=False)
espvida_df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 29 entries, 0 to 28
Data columns (total 26 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   Territorialidades     29 non-null     object
1   esp_vida 2017         29 non-null     object
2   esp_vida 2018         28 non-null     object
3   esp_vida 2019         28 non-null     object
4   esp_vida 2020         28 non-null     object
5   esp_vida 2021         28 non-null     object
```

A Base de dados Esperança de vida ao nascer, representada na Figura 22, é de suma importância, pois é através dela que o fator longevidade do IDH e IDHM é analisado. Como nas outras bases ela também possui os atributos de territorialidade, gênero e raça, definidas por: HOMEM, MULHER, BRANCOS E NEGROS. Todos esses atributos estão datados do ano de 2017 a 2021. Vale ressaltar que, o objetivo do intervalo das datas, serve para acompanhar os resultados dentro do espaço de 5 anos e verificar métricas de progresso positivas ou negativas.

A próxima base a seguir, é a base de dados de Extrema Pobreza e está representada na Figura 23 a seguir.

Figura 23 - Base de dados Extrema Pobreza

```
In [24]: extpobreza_df = pd.read_csv("extpobreza.csv", sep=';', low_memory=False)
extpobreza_df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 28 entries, 0 to 27
Data columns (total 6 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   Territorialidades     28 non-null     object
1   Extrema pobreza 2017  28 non-null     object
2   Extrema pobreza 2018  28 non-null     object
3   Extrema pobreza 2019  28 non-null     object
4   Extrema pobreza 2020  28 non-null     object
5   Extrema pobreza 2021  28 non-null     object
dtypes: object(6)
memory usage: 1.4+ KB
```

Diferente das demais bases de dados, é possível observar na Figura 23 que a base relativa a Extrema Pobreza é menor, já que representa uma amostragem geral

do percentual de pobreza nos territórios brasileiros. Além disso, está definida em intervalos datados do ano de 2017 a 2021. Dessa forma, é possível acompanhar os níveis de extrema pobreza dos estados nesse intervalo de 5 anos e identificar se houve alguma melhoria.

Por fim, a última base inserida, foi a base de dados de IDHM, no qual é essencial para comparar com os atributos de outras bases e identificar possíveis informações e conexões, já que os dados inseridos até aqui, são fatores determinantes para o cálculo de desenvolvimento humano. A Figura 24, apresentada a seguir, mostra os atributos relativos ao IDHM.

Figura 24 - Base de dados IDHM

```
In [10]: idhm_df = pd.read_csv("idhm.csv", sep=';', low_memory=False)
idhm_df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 28 entries, 0 to 27
Data columns (total 6 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Territorialidades     28 non-null     object
1   IDHM 2017             28 non-null     object
2   IDHM 2018             28 non-null     object
3   IDHM 2019             28 non-null     object
4   IDHM 2020             28 non-null     object
5   IDHM 2021             28 non-null     object
dtypes: object(6)
memory usage: 1.4+ KB
```

Semelhante à base de dados anterior, é possível notar na Figura 24 que a base relativa ao IDHM, é simplificada e menor que as anteriores, já que o seu objetivo é fornecer dados de níveis do IDHM por territorialidades dos anos de 2017 a 2021. Dessa forma, esses atributos são fatores determinantes para realizar a comparação do IDHM dos estados brasileiros com as outras bases de dados citadas acima, a fim de compará-las.

4.1.2 TRANSFORMAÇÃO

Nesta etapa, foi necessário normalizar alguns atributos da base de dados do IDHM, a fim de se tornar compatível e de fácil identificação com as outras bases de dados. Foram normalizados os atributos: “IDHM 2017” para “idhm_2017”, “IDHM 2018” para “idhm_2018”, “IDHM 2019” para “idhm_2019”, “IDHM 2020” para “idhm_2020” e “IDHM 2021” para “idhm_2021”. Além disso, alguns atributos tipificados como *Object*, mas que tinham valores numéricos, foram alterados para o tipo *Float*. A Figura 25 a seguir mostra a alteração na tabela idhm_df.

Figura 25 - Atributos normalizados da base de dados IDHM

```
In [13]: idhm_df = idhm_df.rename(
          columns={'IDHM 2017' : 'idhm_2017',
                  'IDHM 2018' : 'idhm_2018',
                  'IDHM 2019' : 'idhm_2019',
                  'IDHM 2020' : 'idhm_2020',
                  'IDHM 2021' : 'idhm_2021',
                  })
```

```
In [14]: idhm_df.set_index('Territorialidades')
```

```
Out[14]:
```

	idhm_2017	idhm_2018	idhm_2019	idhm_2020	idhm_2021
Territorialidades					
Brasil	0,774	0,78	0,785	0,784	0,766
Acre	0,712	0,733	0,739	0,746	0,71
Alagoas	0,679	0,689	0,687	0,694	0,684
Amapá	0,732	0,741	0,737	0,724	0,688
Amazonas	0,728	0,718	0,726	0,727	0,7

É possível observar na Figura 25 que os atributos da tabela IDHM, foram alterados através da função “*rename*” e mantendo o padrão entre elas, para melhor organização e compatibilidade com as outras bases de dados. Essa etapa foi realizada com o objetivo de melhor identificar a agregação do idhm com as outras bases de dados. Na Figura 26 a seguir é apresentado a mescla entre duas bases. São elas, Taxa de Analfabetismo e Renda.

Figura 26 - Mesclagem entre renda e analfabetismo

```
In [8]: analfa_renda = pd.merge(taxaanalfabetismo_df, renda_df,
                               how='left', left_on=['Territorialidades'], right_on=['Territorialidades'])
```

Na figura 26 é possível identificar a nova tabela gerada “*analfa_renda*” no qual foi realizada a agregação das duas bases de dados por meio da função “*merged*”. Este processo resume-se na inserção de dois *dataframes*, especialmente a coluna em que foi destinada a chave principal para a mescla das duas bases de dados. Este mesmo processo foi realizado com as demais tabelas, a fim de unificá-las e agregá-las junto a base do IDHM por meio de um atributo comum, que é a “Territorialidade”. Em seguida, a próxima transformação realizada, foram as bases do “*analfa_renda*” e “*esperança de vida*”, ilustradas na Figura 27 a seguir.

Figura 27 - Mesclagem entre *analfa_renda* e *esperança de vida*

```
In [35]: analfa_renda_espvida = pd.merge
         (analfa_renda, espvida_df, how='left', left_on=['Territorialidades'], right_on=['Territorialidades'])
```

A Figura 27 mostra nova tabela gerada *analfa_renda_espvida* construída através da função *merged* das duas tabelas anteriores. As duas tabelas foram agregadas por meio de um atributo comum, a “Territorialidade”.

Seguindo o processo de transformação, a Figura 28 a seguir demonstra a agregação de atributos da base de dados *analfa_renda_espvida* com a base de extrema pobreza.

Figura 28 - Mesclagem entre *analfa_renda_espvida* e extrema pobreza

```
In [10]: analfa_renda_espvida_pob =
         pd.merge(analfa_renda_espvida, extpobreza_df, how='left', left_on=['Territorialidades'],
                 right_on=['Territorialidades'])
```

Semelhante ao processo de agregação das bases anteriores, a Figura 28, mostra a agregação da base de *analfa_renda_espvida* com a base de extrema pobreza.

Essas duas foram mescladas por meio de um atributo comum, a “Territorialidade”, a fim de gerar uma nova tabela, `analfa_renda_espvida_pob`.

Por fim, a última agregação está representada na Figura 29 a seguir, e é definida pela mesclagem da base anterior com o `idhm`, dado importante para o sucesso do projeto.

Figura 29 - Mesclagem entre `analfa_renda_espvida_pob` e `idhm`

```
In [11]: dados_idhm = pd.merge
          (analfa_renda_espvida_pob, idhm_df, how='left',
          left_on=['Territorialidades'], right_on=['Territorialidades'])
```

De forma resumida, a Figura 29 exibe uma representação da agregação das bases anteriores com o `idhm`. A mesclagem dessas duas bases, é de suma importância para identificar os níveis de IDHM dos estados com outros indicadores de desenvolvimento humano permitindo fornecer um resultado geral para todas as unidades Federativas e também a nível nacional.

Com as bases necessárias agregadas, foi necessário converter o tipo dos valores de algumas colunas, já que a maioria dos atributos tinham seus valores do tipo *Object* e foi necessário transformá-los para *Float*, pois a grande maioria é constituída por valores numéricos. Para isso, houve a necessidade de converter os numerais para o padrão americano, trocando vírgula por ponto usando o método *apply* e *replace* conforme mostra a figura a seguir.

Figura 30 - Conversão de vírgula para ponto.

```
dados_idhm['Analfabetismo 2017'] = dados_idhm['Analfabetismo 2017'].apply(lambda x: str(x).replace(",","."))
dados_idhm['Analfabetismo 2018'] = dados_idhm['Analfabetismo 2018'].apply(lambda x: str(x).replace(",","."))
dados_idhm['Analfabetismo 2019'] = dados_idhm['Analfabetismo 2019'].apply(lambda x: str(x).replace(",","."))
dados_idhm['Analfabetismo 2020'] = dados_idhm['Analfabetismo 2020'].apply(lambda x: str(x).replace(",","."))
dados_idhm['Analfabetismo 2021'] = dados_idhm['Analfabetismo 2021'].apply(lambda x: str(x).replace(",","."))
```

A Figura 30, mostra um exemplo do processo realizado nas tabelas agregadas, substituindo a vírgula para o ponto e permitindo assim a conversão dos tipos dos valores, já que seria impossível convertê-los mantendo o padrão brasileiro. O *apply*, permite que eu utilize uma função para ser aplicada em todos os elementos da coluna e o *replace* a substituição da pontuação.

Dessa forma, foi possível realizar a conversão dos tipos de dados, utilizando o método `astype` e passando o tipo do valor para qual quero converter. Segue um exemplo aplicado nas tabelas na figura a seguir.

Figura 31 - Conversão de Object para Float.

```
dados_idhm['Analfabetismo 2017'] = dados_idhm['Analfabetismo 2017'].astype('float64')
dados_idhm['Analfabetismo 2018'] = dados_idhm['Analfabetismo 2018'].astype('float64')
dados_idhm['Analfabetismo 2019'] = dados_idhm['Analfabetismo 2019'].astype('float64')
dados_idhm['Analfabetismo 2020'] = dados_idhm['Analfabetismo 2020'].astype('float64')
dados_idhm['Analfabetismo 2021'] = dados_idhm['Analfabetismo 2021'].astype('float64')
```

A Figura 31 mostra um exemplo da conversão aplicado na tabela `dados_idhm` sobre o atributo 'Analfabetismo 2017' utilizando o método `astype` e convertendo para o tipo `float`. A conversão é necessária para a etapa de carregamento para o banco, já que permite diferenciar valores de texto e numéricos.

Concluindo a etapa de transformação, foi necessário pegar os dados transformados e alterá-los para o formato CSV e em seguida convertido para o formato JSON. Para isso foi utilizado a função `to_csv`, gerando um novo arquivo denominado de "dados_finais.csv" e a função `to_json`, resultando no arquivo final "dados_finais.json".

Figura 32 - Alterando dados de CVS para JSON

```
In [ ]: dados_idhm.to_csv("dados_finais.csv")

In [14]: csv_dados = ("dados_finais.csv")
         dados_finais_df = pd.read_csv(csv_dados , sep= ",")

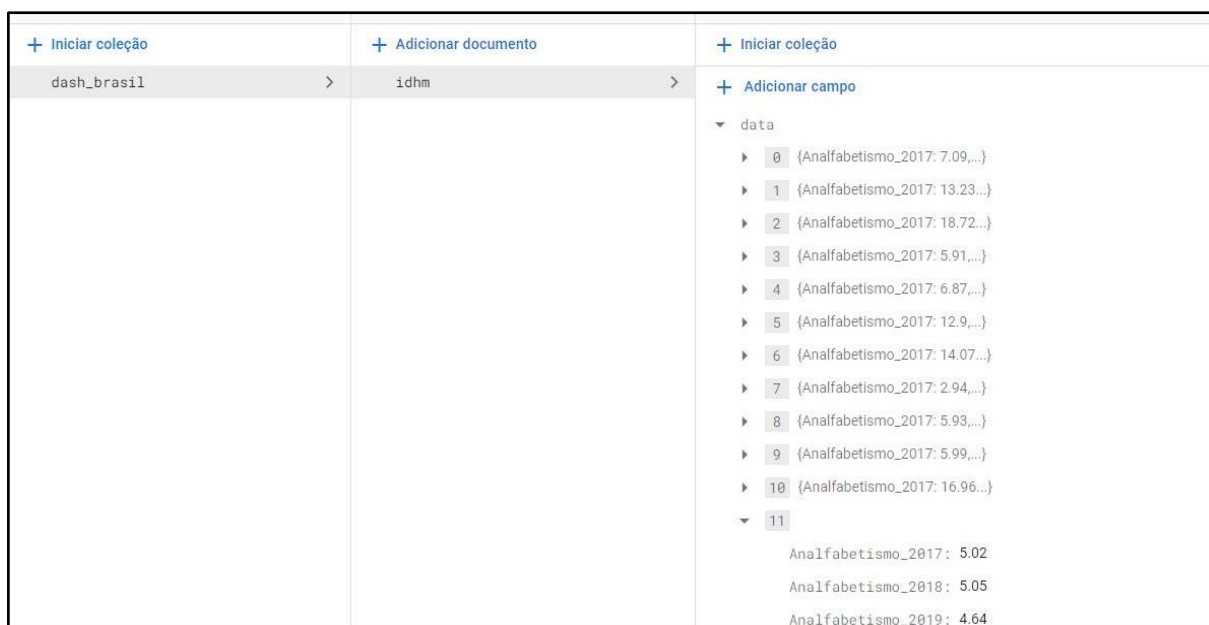
In [15]: json_output = ("dados_finais.json")
         output= dados_finais_df.to_json(json_output)
```

Conforme a Figura 32, foi possível transformar os dados de Python para CSV utilizando as bibliotecas apropriadas. Em seguida, foi realizada a conversão do arquivo CSV para o formato JSON, que é compatível com o *Cloud Firestore*.

4.1.3 CARREGAMENTO

Durante a etapa de carregamento, os dados que passaram pela etapa de transformação são inseridos na estrutura de armazenamento final. A plataforma utilizada para o processo de carga dos dados foi o *Cloud Firestore* do *Firebase*, no qual é conhecido por ser um banco de dados flexível e escalonável para desenvolvimento móvel e *web*. A plataforma foi utilizada como um banco de dados em nuvem NoSQL, onde os dados no formato JSON foram mapeados de acordo com a sua informação única, Territorialidade ou ID. Foi possível enviar os documentos JSON diretamente para o *Firestore* usando as bibliotecas de cliente disponíveis. A imagem a seguir mostra parte da interface do *Firebase*, com os dados armazenados em tempo real.

Figura 33 - Dados armazenados no *Firebase*



A Figura 33 mostra a disponibilidade dos dados dentro do *Firebase*, dentro de uma coleção denominada de “dash_brasil” contendo um documento “idhm” e respectivamente os dados armazenados. Ao importar os dados para o *Cloud Firestore*, é crucial levar em conta a estrutura dos documentos e a configuração do banco de dados. O *Firestore* é uma base de dados orientada a documentos, o que implica que os dados são organizados em coleções e documentos. Cada documento é composto por pares de chaves e valores, nos quais os valores podem assumir diferentes tipos, inclusive documentos aninhados.

4.2 DESENVOLVIMENTO

Uma vez que os dados foram extraídos, transformados e carregados para o banco de dados, eles estão prontos para serem utilizados pelo *dashboard*. Mas para que os objetivos do projeto sejam executados com sucesso, foi necessário realizar a prototipação e implementação do *dashboard*, respeitando os requisitos propostos anteriormente. Com a prototipação foi possível ter uma ideia clara de como o *dashboard* seria estruturado, como os elementos seriam organizados e apresentados ao usuário final. Além disso, foi possível avaliar o layout, as cores, os gráficos e as opções de navegação, garantindo um design simples e funcional e que esteja alinhado aos objetivos do projeto.

Com base no protótipo, inicia-se então a implementação do *dashboard*, onde foram empregadas as tecnologias e ferramentas necessárias para transformar a ideia em uma realidade tangível. Durante essa implementação, são considerados os requisitos propostos anteriormente, assegurando que o *dashboard* possa cumprir com sucesso os objetivos estabelecidos para o projeto.

4.2.1 PROTOTIPAÇÃO

Com a utilização da ferramenta de prototipação Figma foi possível a elaboração do protótipo deste trabalho, buscando obedecer aos requisitos funcionais estabelecidos anteriormente. A Figura 33 exibe a tela criada para a visualização dos dados do IBGE que compõem o IDH brasileiro.

Figura 34 - Protótipo de tela do dashboard



A Figura 34 representa a tela proposta para o *dashboard*. Na parte superior, é possível visualizar cinco cartões contendo informações gerais de cada estado brasileiro, tais como IDHM, analfabetismo, renda, esperança de vida e pobreza. Essas informações servem como parâmetro, uma vez que o IDHM utiliza indicadores de educação, renda e longevidade como métricas.

Mais abaixo, é possível observar um filtro, composto por três campos. O primeiro campo é utilizado para selecionar o estado desejado, o segundo campo é utilizado para selecionar o dado a ser visualizado, podendo ser dados relacionados ao analfabetismo, ou esperança de vida. Já o terceiro campo corresponde ao ano referente aos dados, que pode variar de 2017 a 2021. Ao lado direito, um gráfico de barras é responsável por mostrar os dados referentes à renda per capita dos grupos, podendo filtrar de acordo com gênero e raça acompanhado das métricas gerais.

A parte inferior da tela contém quatro gráficos do tipo pizza, que se relacionam com as escolhas feitas do filtro. É nele que os dados que foram selecionados serão fornecidos por grupo de pessoas, como homens, mulheres, negros e brancos. Por fim, à esquerda um *ranking* do IDHM do ano selecionado mostra a posição de alguns estados brasileiros.

A partir do protótipo criado, iniciou-se o processo de implementação do *dashboard* buscando manter os requisitos essenciais como o uso do filtro e da utilização

de gráficos, a fim de proporcionar maior interatividade ao usuário. A seguir, são apresentados os detalhes da implementação do *dashboard*.

4.2.2 IMPLEMENTAÇÃO

O propósito desta seção é detalhar as etapas cruciais para a implementação do *dashboard*, bem como demonstrar e destacar o resultado da tela prototipada e suas alterações. Para iniciar o desenvolvimento, foi implementado uma estrutura inicial com o *framework Flutter*, e a linguagem de programação Dart obedecendo o protótipo criado.

Segundo Tableau (2012) e Juice (2009), a maneira como os elementos são dispostos no *dashboard* pode acelerar a compreensão do usuário e tornar os resultados mais acessíveis aos interessados. Abaixo, é apresentada e explicada a interface do *dashboard*.

Figura 35 - Tela do dashboard desenvolvido



Na Figura 35 é apresentada a tela do *dashboard* denominado DashBrasil. Como o intuito de um *dashboard* estratégico é monitorar alguns indicadores de desempenho em um determinado período de tempo, foram criados cinco cartões com o *Widget Cards* na parte superior da tela. A finalidade desses *Cards* é fornecer as informações de cada estado brasileiro inclusive a nível nacional, sobre fatores essenciais na compreensão do IDHM, como a taxa de analfabetismo, renda per capita, esperança

de vida, e a taxa de extrema pobreza. Todas essas informações estão relacionadas ao ano escolhido pelo usuário, dentro de um filtro feito com um *Widget*, denominado de *DropDownButton*. Além disso, este *Widget* permite ao usuário escolher quais informações ele deseja visualizar por grupo, como a taxa de analfabetismo de pessoas com 18 anos ou mais e esperança de vida ao nascer. Ao lado, um gráfico de barras, implementado com a biblioteca *fl_charts*, demonstra a renda per capita em reais de cinco grupos, sendo, geral, homens, mulheres, negros e brancos, que são filtrados de acordo com gênero e raça.

Após a seleção do filtro, as informações são geradas dentro de quatro gráficos situados na parte inferior da tela. Esses gráficos são provenientes de uma biblioteca *Flutter*, denominada de *pie_chart*, que permite a customização de vários atributos, como cor, tamanho, e estilo tornando possível a compreensão das informações filtradas que são demonstradas em porcentagem, sexo e raça. Além disso, é possível identificar na Figura 35, um exemplo de informação a ser demonstrado no *dashboard*, onde o estado selecionado é o Tocantins, e as informações visualizadas são relativas à taxa de analfabetismo do ano de 2017.

Para coletar os dados do banco, foi necessário realizar a integração do projeto *Flutter* com o *Cloud Firestore*. Primeiramente, adicionamos as dependências "cloud_firestore" e "firebase_core" ao arquivo *pubspec.yaml* do projeto. Em seguida, foi importado as bibliotecas correspondentes ao uso dessas dependências.

Além disso, foi necessário copiar o arquivo de configuração do projeto fornecido pelo Firebase, chamado *FirebaseOptions*. Esse arquivo contém a *APIKEY* específica para a plataforma *Web*. O arquivo foi copiado para o projeto *Flutter* garantindo a correta comunicação com o *Cloud Firestore*. Após essas etapas, é possível utilizar os métodos e funcionalidades fornecidos pela biblioteca "cloud_firestore" para acessar e manipular os dados do banco de dados. A Figura 36 a seguir mostra um exemplo de configuração do *FirebaseOptions*.

Figura 36 - Exemplo de configuração do `FirebaseOptions`

```
static const FirebaseOptions web = FirebaseOptions(  
  apiKey: 'AIzaSyCH2lt75NOzilM99cycgQv48dGY4rDVRy0',  
  appId: '1:211826518707:web:6a57a23a609d3cc33e9ec6',  
  messagingSenderId: '211826518707',  
  projectId: 'wide-office-315415',  
  authDomain: 'wide-office-315415.firebaseio.com',  
  storageBucket: 'wide-office-315415.appspot.com',  
);
```

A partir disso, é criado um arquivo repositório, no qual é feita a instância do *Firestore*, e criada uma instância da coleção “dash_brasil”, identificando o documento de referência, ou seja, o “idhm”. Isso permite que seja possível realizar um *Get* e coletar o conteúdo do documento de referência que será armazenado dentro de uma lista vazia chamada “data”. Dessa forma, a partir do documento de referência é possível realizar diversas ações, como selecionar os dados do banco, realizar *updates* e fazer deletes, adaptando o método assíncrono *get* para outros padrões de operação. A Figura 37 demonstra o trecho do código dessa conexão.

Figura 37 - Trecho do código de conexão com o documento

```
Firestore.instance.collection('dash_brasil');  
DocumentReference docRef = collectionVisitorsTable.doc("idhm");  
  
final response = await docRef.get();  
  
List<DataFirebaseModel> list = [];  
  
final lista = response['data'] as List<dynamic>;
```

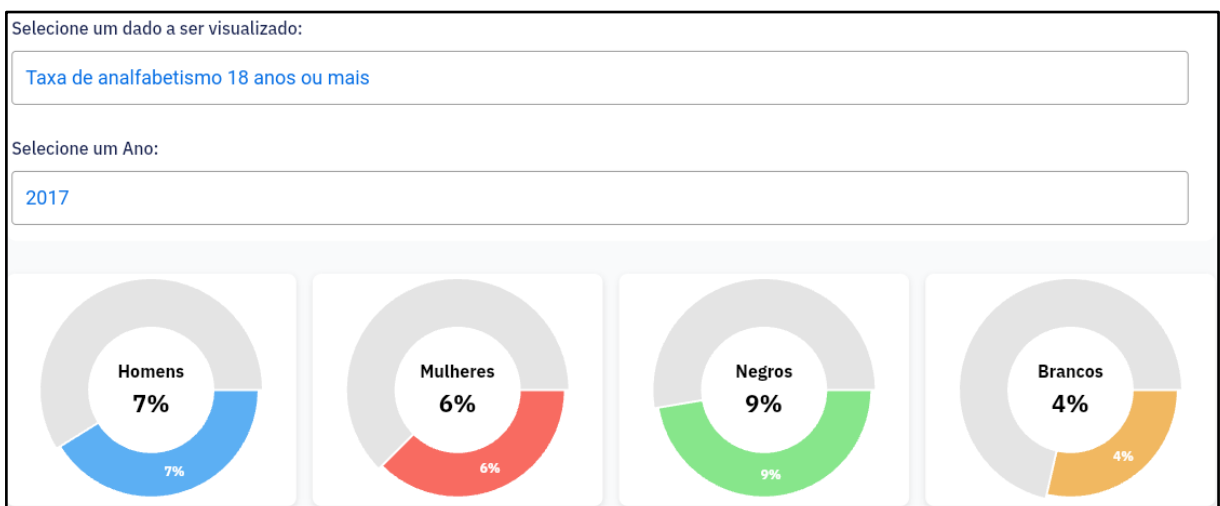
Após todas as seleções serem realizadas e integradas com o *Widget* de filtragem, gráficos e *cards* foram realizados os testes para verificar se todas as informações estavam sendo coletadas do banco de forma correta através das requisições e interações do usuário com o *dashboard*. Foi observado que todas as requisições estavam sendo feitas corretamente, puxando os dados de cada estado brasileiro e mostrando nos gráficos e *cards* do projeto. A Figura 38 mostra parte da tela, contendo uma requisição a nível nacional.

Figura 38 - Teste dos cards e gráfico de barras



Através da Figura 38 é possível observar a escolha no filtro pela opção Brasil. Após a seleção da territorialidade e do ano, os respectivos dados são armazenados dentro de cada *card* e o gráfico de renda é gerado. Além disso, ao selecionar uma informação a ser desejada, é possível visualizar as informações por grupo dentro dos gráficos, conforme a Figura 39 a seguir.

Figura 39 - Teste dos gráficos



A Figura 39 mostra que o dado a ser visualizado está relacionado à taxa de analfabetismo. Como a territorialidade selecionada é o Brasil, os gráficos retornaram as informações do país do ano de 2017 por porcentagem de cada grupo. Dessa forma, foi possível identificar o máximo de informações dentro de uma única tela, atingindo o propósito funcional de um *dashboard* estratégico.

O gráfico escolhido para a representação dos indicadores de analfabetismo e esperança de vida aplicados aos grupos, foi o de setor, também conhecido como pizza, ou rosca. A utilização desse tipo de gráfico se dá pelo fato dos dados dos respectivos indicadores estarem em porcentagem, além disso, é importante representar as proporções de forma clara, visíveis e diferenciadas de cada grupo de pessoas. Para representar o indicador de renda per capita, o gráfico escolhido foi o de barras verticais, já que, o seu propósito é mostrar e comparar valores quantitativos em reais entre os diferentes grupos. Dessa forma, o comprimento de cada barra é proporcional ao valor que representa e separadas para permitir a distinção entre os grupos de pessoas.

Para garantir a aplicação do estudo de percepção visual e a utilização dos princípios de *Gestalt*, os gráficos de setores e de barras foram ajustados próximos um do outro, e enfileirados horizontalmente aplicando os princípios de proximidade e continuidade respectivamente. Além disso, foram aplicadas cores distintas para a diferenciar os grupos, mas mantendo a mesma função de representar os dados escolhidos, obedecendo o princípio da similaridade. Além disso, é possível identificar o princípio de continuidade dentro do filtro de seleção dos dados quando há a necessidade de preencher as informações verticalmente e alinhadas, influenciando o usuário a preencher os requisitos de forma sequencial.

Nesta seção, foram explorados os resultados obtidos durante a construção do *dashboard* proposto. O processo de desenvolvimento resultou na criação de um *Layout*, que fosse simples e intuitivo, seguindo os padrões de *Gestalt* mencionados no referencial teórico, no uso de gráficos específicos para cada representação, na criação de um banco de dados para armazenar os dados transformados no processo ETL e na integração do mesmo com a aplicação. A seção subsequente tem como propósito apresentar as considerações finais para o presente trabalho.

5 CONSIDERAÇÕES FINAIS

No Brasil, as informações utilizadas para a realização do cálculo do IDH são de extrema importância e são coletadas através de dados de pesquisas realizadas pelo Instituto Brasileiro de Geografia e Estatística (IBGE). Os microdados gerados podem conter informações valiosas, e que podem auxiliar na análise socioeconômica do país, permitindo a identificação de desigualdades e formulações de estratégias para reduzi-las. Porém, o grande volume de dados provindos de diferentes pesquisas ao longo dos anos pode ser um problema, já que a análise manual desses dados demandaria esforço e tempo.

A utilização de um *dashboard* estratégico pode auxiliar na análise e visualização desses dados, tendo em vista que tal ferramenta proporciona uma visão global e detalhada em uma única tela (BUCHSBAUM, 2012). Com isso, o objetivo deste trabalho consistiu em desenvolver um *dashboard* para análise e visualização de dados que compõem o IDH a partir de microdados do IBGE, para que possa dar apoio a decisões de políticas públicas e auxiliar no monitoramento de indicadores socioeconômicos do país.

Ademais, o uso do *dashboard* permite a exploração interativa dos dados, fornecendo aos usuários a capacidade de filtrar e visualizar os dados de acordo com suas necessidades e interesses específicos. Isso possibilita a identificação de padrões, tendências e relações entre variáveis, que podem ser cruciais para a compreensão mais aprofundada dos fatores que influenciam o IDH.

Para atingir este objetivo, o processo de desenvolvimento do projeto compõe-se das etapas de extração, transformação e carregamento, a utilização de um banco de dados, e a implementação através do *framework Flutter* e a linguagem Dart. Como resultado, foi criado o “DashBrasil”, um *dashboard web* voltado para análise de dados do IBGE, que compõem o IDH e o IDHM dos estados brasileiros, apresentando informações relativas à educação, renda, e longevidade da população. No decorrer do desenvolvimento, foi estudado como os dados seriam extraídos e transformados para melhor atender os objetivos do projeto, para isso, um grande desafio foi extrair os dados necessários do mesmo ano e que tivessem relação direta com o cálculo de desenvolvimento humano, além disso, a transformação desses dados para que fossem normalizados e carregados dentro do banco de dados, demandou tempo e conhecimento.

O *dashboard* foi desenvolvido para a plataforma *web*, mas ainda é necessário a realização de melhorias na ferramenta. Portanto, como trabalhos futuros, são apresentadas as seguintes sugestões que poderão melhorar a interação com usuário e a melhor visualização dos dados, segue:

- Aprimorar a agregação de novas bases de dados, como taxa de mortalidade, saneamento básico, desigualdade regional, entre outros aspectos relevantes. Além disso, tornar-se visualizar dados tanto de períodos mais antigos quanto atuais.
- Implementação de um mapa do Brasil interativo para a seleção de territorialidades e com cores indicativas.
- Uma linha do tempo, contendo os avanços e retrocessos dos estados ao longo dos anos.
- Hospedagem da aplicação na internet e acesso por meio de uma URL.

Acredita-se que a implementação dessas sugestões aprimorará significativamente a análise e visualização dos dados propostos para este trabalho. Dessa forma, o *dashboard* permitirá mostrar uma gama ainda maior de informações referentes aos aspectos de saúde, renda e educação do país, proporcionando aos usuários uma compreensão abrangente da realidade social brasileira.

REFERÊNCIAS

ANACONDA. Anaconda Distribution. **Anaconda**, 2023. Disponível em: <https://www.anaconda.com/about-us>. Acesso em: 07 maio 2023.

BARROS, R. (2013). **Dashboarding - Projeto e Implementação de Painéis Analíticos**. Dissertação de Mestrado em Engenharia Informática, Universidade do Minho, Escola de Engenharia, Braga.

BROWN, T. (2010). **Design Thinking: Uma metodologia poderosa para decretar o fim das velhas ideias**. Alta Books.

BUCHSBAUM, Paulo. DASHBOARDS - **Introdução ao conceito**. [s. L.]: Fórum de Varejo, 2012. 01 p.

BUCHSBAUM, Paulo. DASHBOARDS - **Introdução ao conceito**. [s. L.]: Fórum de Varejo, 2012. 04 p.

CELKO, Joe. Joe Celko's **Data, Measurements and Standards in SQL**. 2nd Edition. Morgan Kaufmann Publishers, 2015.

SANZ, Andreia Cristina Pereira. **Proposta de um dashboard para monitorizar falhas de energia numa rede elétrica inteligente**. 2018. 107 f. Dissertação (Mestrado) - Curso de Gestão de Sistemas de Informação, Instituto Universitário de Lisboa, Lisboa, 2018.

ABREU, Fábio Silva Gomes da Gama e. DESMISTIFICANDO O CONCEITO DE ETL. **Revista de Sistemas de Informação**, Macaé, v. 02, p. 1-6, nov. 2008.

DUARTE, João Carlos Assunção. **Dashboard Visual, uma ferramenta de Business Intelligence**. 2012. 54 f. Dissertação (Mestrado) - Curso de Engenharia Mecânica, Faculdade de Engenharia da Universidade do Porto, Porto, 2012.

DUPRIEZ, Olivier; BOYKO, Ernie. **Dissemination of Microdata Files: Principles, Procedures and Practices**. International Household Survey Network. [S. l.]: IHSN Working Paper. 2010. p.13.

ECKERSON, W. (2011b). **Types of Performance Dashboards**. Em Performance dashboards: measuring, monitoring and managing your business. New Jersey: John Wiley & Sons, Inc., Your Busin, 101–121.

SCHUCH, Felipe. **Jupyter Notebook como uma Poderosa Ferramenta Educacional**. 2022. Disponível em: <https://www.fschuch.com/blog/2021/01/22/jupyter-notebook-como-uma-poderosa-ferramenta-educacional/>. Acesso em: 03 mar. 2023.

FEW, Stephen. **Information dashboard design: The effective visual communication of data**. Sebastopol, CA: O'Reilly Media (2006).

FEW, Stephen. **Why Most Dashboards Fail: The Effective Visual Communication of Data** (2007).

ALURA, **FIGMA: o que é a ferramenta, Design e uso, 2023**. Disponível em: <https://www.alura.com.br/artigos/figma>. Acesso em: 30 de maio 2023

FIREBASE. **Realtime-database**, 2023. Disponível em: <https://firebase.google.com/?hl=pt-br>. Acesso em: 31 maio 2023

GOOGLE. **Dart**, 2023. Disponível em: <https://dart.dev> Acesso em: 07 maio 2023.

GOOGLE. **Flutter**, 2023. Disponível em: <https://flutter.dev> Acesso em: 07 maio 2023.

JUICE. **A Guide to Creating Dashboards People Love to Use**. Waterloo: Juice Inc, 2009. 49 p.

KIMBALL e CASERTA 2004 Ralph. KIMBALL e Joe CASERTA. "**The data warehouse ETL toolkit: practical techniques for extracting, cleaning, conforming, and delivering data**". Em: 1 ed. (2004), p.22;).

KIMBALL e CASERTA 2004 Ralph. KIMBALL e Joe CASERTA. "**The data warehouse ETL toolkit: practical techniques for extracting, cleaning, conforming, and delivering data**". Em: 1 ed. (2004), p.22;).

KIMBALL, Ralph; ROSS, Margy. **The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling**. 3rd Edition. John Wiley & Sons, 2013.

Kozielski, S., & Wrembel, R. (2008). **New Trends in Data Warehousing and Data Analysis**. New York: Springer Science+Business Media.

MICROSOFT. **Extract, transform and load (ETL). Microsoft Azure**, 2019. Disponível em: <https://learn.microsoft.com/en-us/azure/architecture/data-guide/relational-data/etl>. Acesso em: 06 abril 2023

OPSERVICES. **O que é um dashboard? O guia completo e definitivo!** 2017. Disponível em: <https://www.opservices.com.br/o-que-e-um-dashboard/>

PANDAS. **Pandas Python**, 2023. Disponível em: <https://pandas.pydata.org/about/>. Acesso em: 07 maio 2023.

Pauwels, K., T. Ambler, B. H. Clark, P. LaPointe, D. Reibstein, B. Skiera, B. Wierenga, and T. Wiesel. 2009. "**Dashboards as a service: Why, what, how, and what research is needed?**" Journal of service research no. 12 (2):175.

PINTO, Daniela Gomes; Costa, Marco Aurélio; MARQUES, Maria Luiza de Aguiar. **O Índice de Desenvolvimento Humano Municipal brasileiro**. Brasil: Pnud Brasil, 2013. 95 p. Disponível em: <http://repositorio.ipea.gov.br/handle/11058/2375>

PNUD. **Atlas do Desenvolvimento Humano no Brasil**. 2000. Disponível em: <https://www.undp.org/pt/brazil/o-que-é-o-idh>. Acesso em: 28 março 2023

PREARO, Leandro Campi; MARACCINI, Maria Clara; ROMEIRO, Maria do Carmo. Fatores Determinantes Do Índice De Desenvolvimento Humano Dos Municípios Do Estado De São Paulo. **Revista Brasileira de Políticas Públicas**, [S.L.], v. 5, n. 1, p. 134-155, 14 jun. 2015. Centro de Ensino Unificado de Brasília. <http://dx.doi.org/10.5102/rbpp.v5i1.2960>.

PREECE, J.; ROGERS, Y.; SHARP, H. **Design de interação: além da interação homem-computador**. Porto Alegre: Bookman, 2005. 548 p.

RASMUSSEN, N. H.; BANSAL, M.; CHEN, C. Y. **Business dashboards: a visual catalog for design and deployment**. [S.l.]: John Wiley & Sons, 2009

TABLEAU (Estados Unidos). **Visual Analysis Best Practices: Simple Techniques for Making Every Data Visualization Useful and Beautiful**. Seattle: Copyright Tableau Software, Inc. 2012. All Rights Reserved, 2012

TORRES, H. G.; FERREIRA, M. P.; DINI, N. P. Indicadores sociais: por que construir novos indicadores como o IPRS. **São Paulo em Perspectiva**, v. 17, n. 3-4, p. 80-90, 2003

TORRES, Haroldo da Gama; FERREIRA, Maria Paula; DINI, Nádía Pinheiro. INDICADORES SOCIAIS: por que construir novos indicadores como o iprs. **São Paulo em Perspectiva**, São Paulo, v. 17, n. 4, p. 80-90, jul-dez. 2003.

VASCONCELOS, Marco Antonio; GARCIA, Manuel Enriquez. **Fundamentos de economia**. São Paulo: Saraiva, 1998.