

SELECTOR - Um sistema para apoio a escolha de escolas com Atendimento Educacional Especializado.

Mariana de Araujo Barreiros Santos
Instituto Federal da Bahia
Salvador, Brasil
Email: marianabarreiros.ti@gmail.com

Pablo Vieira Florentino
Instituto Federal da Bahia
Salvador, Brasil
Email: pablovf@ifba.edu.br

Resumo—Embora a regulamentação da Educação Especial esteja amplamente amparada pela legislação brasileira — Constituição Federal de 1988 e Política Nacional de Educação Especial na Perspectiva da Educação Inclusiva — pais e/ou responsáveis por pessoas com deficiência, transtornos globais de desenvolvimento, altas habilidades e/ou superdotação, ainda encontram dificuldades na procura, matrícula e manutenção dessas pessoas em instituições de ensino que atendam as obrigações e deveres propostos pelas Diretrizes Operacionais da Educação Especial para o Atendimento Educacional Especializado — AEE. Em muitos casos, faltam escolas sem as mínimas condições de infraestrutura e professores devidamente qualificados. Concomitantemente, devido à Lei de Acesso à Informação, é dever dos órgãos e entidades públicas disponibilizar informações de interesse coletivo por eles produzidas ou custodiadas para que, dentre outros motivos, possa haver contribuições à sociedade com serviços inovadores que beneficiem o cidadão. No entanto, quanto as aplicações que dizem respeito a área educacional, apesar de existirem algumas que consideram o assunto “Acessibilidade”, não há, até o momento, aplicações que auxiliem o processo de busca das escolas que oferecem atendimento AEE, servindo como uma base na escolha da escola, possivelmente, mais adequadas para uma pessoa que necessita dessa categoria de atendimento. Neste contexto, este trabalho visa construir um produto minimamente viável para auxiliar na tomada de decisão dos pais, ou responsáveis diretos por pessoas com deficiências, acerca das escolas soteropolitanas, possivelmente, mais adequadas para pessoas que necessitem de Atendimento Educacional Especializado. Acredita-se assim que, no ato do processo de pesquisas por escolas preparadas para atender pessoas que necessitam desse tipo de atendimento, pais e responsáveis diretos tenham em sua posse uma aplicação que facilite essa busca e forneça informações sobre as escolas apresentadas, parametrizadas pelas informações passadas pelo próprio usuário. Assim sendo, esse trabalho se dividirá em duas fases. A primeira etapa mostrará a coleta, exploração, preparo e visualização dos microdados do Censo Escolar de 2019, dos dados do Catálogo das Escolas brasileiras e dos dados oriundos da extração de sites que contêm informações sobre transportes coletivos em Salvador. Em seguida, serão apresentadas as etapas do desenvolvimento de uma aplicação *Restful* para consumir os dados acima referidos.

Keywords—*Ciência de Dados, Tecnologia Social, Análise de Dados, Dados Abertos, Visualização de Dados, Fusão de Dados.*

I. INTRODUÇÃO

O Atendimento Educacional Especializado tem como principal objetivo possibilitar que pessoas, com suas respectivas necessidades, tenham a oportunidade de trabalhar a autonomia e independência nas escolas ou fora delas, através de

uma formação complementar e/ou suplementar. Sua função é elaborar e organizar recursos pedagógicos e de acessibilidade com o propósito da eliminação de barreiras para que os alunos possam ter plena participação no processo de aprendizagem. Esta categoria de ensino foi oficializada pelo Ministério da Educação, através da Secretaria de Educação Especial, instituindo as Diretrizes Operacionais da Educação Especial para o Atendimento Educacional Especializado, AEE, na educação básica [1].

O Censo Escolar de 2008 mostrou que o número de matrículas de pessoas com deficiências em 2007, de 34,4%, saltou para 54% em 2008 [2]. Esse fato, portanto, corrobora a necessidade do debate referente a capacidade das escolas brasileiras em incluir pessoas que necessitam desse tipo de atendimento, debate este que perpassa pelas discussões referentes a infraestrutura de uma escola. É essencial discutir, também, a eficácia das salas de recursos multifuncionais, a formação dos professores, a utilização de tecnologias assistivas, a disponibilização de materiais socioeducativos e/ou pedagógicos para uso nas escolas, etc [3].

O Censo Escolar é, atualmente, um dos mais importantes instrumentos para que atores educacionais e a sociedade possam compreender a situação educacional no Brasil. Com essa ferramenta, diversas informações da educação básica são coletadas, envolvendo diferentes etapas e modalidades da educação básica e profissional — ensino regular, educação de jovens e adultos, educação profissional e a educação especial [4].

A partir desses dados, muitos sistemas foram construídos visando fornecer à sociedade informações sobre a situação das escolas no país. No entanto, não há, até o momento, aplicações que trabalham com esses dados visando apoiar na escolha de espaços escolares que permitam uma melhor experiência para pessoas que necessitam de atendimento educacional especializado, ou são responsáveis diretas por essas pessoas [5] [6].

Considerando esta lacuna, o presente trabalho se propõe a construir um sistema que auxilie pais e/ou responsáveis diretos por pessoas com deficiência, acerca da adequação das escolas para pessoas que necessitam de AEE. O projeto visa não apenas trabalhar com dados referentes a infraestrutura de uma escola, mas dados que estejam relacionados a formação de professores, profissionais de saúde presentes nesses espaços, materiais socioeducativos presentes nelas, salas com recursos multifuncionais e outros.

Além disto, vale ressaltar que a solução não foi con-

cebida com o propósito de gerar um *ranking* das escolas, tampouco ser um sistema de recomendação, mas um sistema que ordenará as escolas que possuam, além do que já foi supracitado, infraestrutura amparada no decreto n.º 6.571, de 18 de setembro de 2008.

II. REVISÃO BIBLIOGRÁFICA E TEÓRICA

Para o processo de concepção deste trabalho, foi necessária uma revisão literária tendo como objetivo o levantamento da fundamentação teórica do tema proposto. Portanto, esta seção apresenta os principais assuntos relacionados a este trabalho. A subseção II-A apresenta o conceito de Atendimento Educacional Especializado; A subseção II-B trata sobre Dados Abertos e sua importância para a sociedade. A subseção II-C apresenta conceitos relacionados a Ciência de Dados. A subseção II-D discute sobre Análise de Dados, sua história e realiza uma rápida exposição de como essa técnica foi utilizada no presente trabalho. A subseção II-E realiza um breve resumo sobre o conceito de Fusão de Dados. A subseção II-F apresenta alguns conceitos sobre a Visualização de Dados e informa ao leitor como esse tema foi abordado no trabalho. Por fim, a subseção II-G apresenta o conceito de Tecnologias Sociais.

A. Atendimento Educacional Especializado

Como já mencionado na Introdução deste trabalho, o Atendimento Educacional Especializado é um conjunto de ações realizadas complementar e suplementarmente à formação de alunos, com deficiências, presentes no ensino regular. Seu maior objetivo é a garantia da igualdade de condições para o acesso, participação e permanência desses estudantes, com demandas específicas, em seus respectivos ambientes escolares, focando não apenas no conteúdo acadêmico, mas, também, no seu desenvolvimento emocional e em sua inserção na sociedade. O AEE constitui-se assim como uma ferramenta de inclusão necessária presente no país.

Conforme o Decreto n.º 6.571, faz parte do público-alvo do AEE [1]:

a) Alunos com deficiência: aqueles que têm impedimentos de longo prazo de natureza física, intelectual, mental ou sensorial, os quais, em interação com diversas barreiras, podem obstruir sua participação plena e efetiva na sociedade em igualdade de condições com as demais pessoas; b) Alunos com transtornos globais do desenvolvimento: aqueles que apresentam um quadro de alterações no desenvolvimento neuropsicomotor, comprometimento nas relações sociais, na comunicação ou estereotípias motoras. Incluem-se nessa definição, alunos com autismo clássico, síndrome de Asperger, síndrome de Rett, transtorno desintegrativo da infância (psicoses) e transtornos invasivos sem outra especificação; c) Alunos com altas habilidades/superdotação: aqueles que apresentam um potencial elevado e grande envolvimento com as áreas do conhecimento humano, isoladas ou combinadas: intelectual, acadêmica, liderança, psicomotora, artes e criatividade.

Ainda segundo o Decreto, o AEE deve ser, prioritariamente, realizado nas salas de recursos multifuncionais da própria escola ou em outra escola de ensino regular, centro de atendimento educacional especializado público ou privado sem fins lucrativos, que esteja conveniado com a Secretaria de

Educação. Sendo assim, é dever da organização que oferece o atendimento garantir o seguinte [1]:

a) Sala de recursos multifuncional: espaço físico, mobiliários, materiais didáticos, recursos pedagógicos e de acessibilidade e equipamentos específicos; b) Matrícula do aluno no AEE: condicionada à matrícula no ensino regular da própria escola ou de outra escola; c) Plano do AEE: identificação das necessidades educacionais específicas dos alunos, definição dos recursos necessários e das atividades a serem desenvolvidas; cronograma de atendimento dos alunos; d) Professor para o exercício da docência do AEE; e) Profissionais da educação: tradutor e intérprete de Língua brasileira de Sinais, guia-intérprete e outros que atuam no apoio às atividades de alimentação, higiene e locomoção. f) Articulação entre professores do AEE e os do ensino comum. g) Redes de apoio: no âmbito da atuação intersetorial, da formação docente, do acesso a recursos, serviços e equipamentos, entre outros que contribuam para a realização do AEE.

Já com relação à formação dos professores, é sua atribuição [1]:

a) Identificar, elaborar, produzir e organizar serviços, recursos pedagógicos, de acessibilidade e estratégias considerando as necessidades específicas dos alunos público-alvo da educação especial; b) Elaborar e executar plano de atendimento educacional especializado, avaliando a funcionalidade e a aplicabilidade dos recursos pedagógicos e de acessibilidade; c) Organizar o tipo e o número de atendimentos aos alunos na sala de recursos multifuncional; d) Acompanhar a funcionalidade e a aplicabilidade dos recursos pedagógicos e de acessibilidade na sala de aula comum do ensino regular, bem como em outros ambientes da escola; e) Estabelecer parcerias com as áreas intersetoriais na elaboração de estratégias e na disponibilização de recursos de acessibilidade; f) Orientar professores e famílias sobre os recursos pedagógicos e de acessibilidade utilizados pelo aluno; g) Ensinar e usar recursos de Tecnologia Assistiva, tais como: as tecnologias da informação e comunicação, a comunicação alternativa e aumentativa, a informática acessível, o soroban¹, os recursos ópticos e não ópticos, os *softwares* específicos, os códigos e linguagens, as atividades de orientação e mobilidade entre outros; de forma a ampliar habilidades funcionais dos alunos, promovendo autonomia, atividade e participação. h) Estabelecer articulação com os professores da sala de aula comum, visando a disponibilização dos serviços, dos recursos pedagógicos e de acessibilidade e das estratégias que promovem a participação dos alunos nas atividades escolares. i) Promover atividades e espaços de participação da família e a interface com os serviços setoriais da saúde, da assistência social, entre outros.

Portanto, neste sentido, enquanto ferramenta de inclusão necessária presente no país, o Atendimento Educacional Especializado torna-se uma ferramenta capaz de possibilitar que, primeiro, seja a sociedade o principal instrumento de mudança na inclusão de pessoas com necessidades especiais. Dado que, para Sasaki (1997, p.3), a inclusão social destas pessoas é “um processo bilateral em que as pessoas, ainda excluídas, e a sociedade buscam, em parcerias, equacionar problemas, decidir sobre soluções e efetivar a equiparação de oportunidades para todos” [7].

¹<http://www.bengalalegal.com/soroban>

B. Dados Abertos

Segundo a definição da *Open Knowledge* [8] Internacional, "dados são abertos quando qualquer pessoa pode livremente acessá-los, utilizá-los, modificá-los e compartilhá-los para qualquer finalidade, estando sujeito a, no máximo, exigências que visem preservar sua proveniência e sua abertura".

Quando os dados são produzidos, coletados ou custodiados pelas autoridades públicas e disponibilizados em formato aberto, considera-se que os dados são governamentais. David Eaves [9], especialista em políticas públicas e ativista dos dados abertos, propôs aos Dados Abertos Governamentais as seguintes leis:

- Se o dado não pode ser encontrado e indexado na Web, ele não existe.
- Se não estiver aberto e disponível em formato compreensível por máquina, ele não pode ser reaproveitado.
- Se algum dispositivo legal não permitir sua replicação, ele não é útil.

Para que os dados governamentais sejam realmente considerados abertos, precisam seguir oito princípios [10] [11] [12] [13] [14]:

- 1) **Completos:** Todos os dados públicos são disponibilizados. Dados são informações eletronicamente gravadas, incluindo, mas não se limitando, documentos, bancos de dados, transcrições e gravações audiovisuais. Dados públicos são dados que não estão sujeitos a limitações válidas de privacidade, segurança ou controle de acesso, reguladas por estatutos.
- 2) **Primários:** Os dados são publicados na forma coletada na fonte, com a mais fina granularidade possível, e não agregadamente ou transformada.
- 3) **Atuais:** Os dados são disponibilizados o quanto rapidamente seja necessário para preservar o seu valor.
- 4) **Acessíveis:** Os dados são disponibilizados para o público mais amplo possível e para os propósitos mais variados possíveis.
- 5) **Processáveis por máquina:** Os dados são razoavelmente estruturados para possibilitar o seu processamento automatizado.
- 6) **Acesso não discriminatório:** Os dados estão disponíveis para todos, sem ser necessária identificação ou registro.
- 7) **Formatos não proprietários:** Os dados estão disponíveis em um formato sobre o qual nenhum ente tenha controle exclusivo.
- 8) **Livres de licenças:** Os dados não estão sujeitos a regulações de direitos autorais, marcas, patentes ou segredo industrial. Restrições razoáveis de privacidade, segurança e controle de acesso podem ser permitidas na forma regulada por estatutos.

No Brasil, o principal instrumento de transparência pública em posse da sociedade está garantido através da Lei de Acesso à Informação [15]. Por esta lei, fica garantido que órgãos e entidades da Administração Pública Federal, Estadual, Distrital e Municipal, além dos poderes Executivo, Judiciário e Legislativo, devem assegurar ao cidadão acesso às informações públicas por eles requeridas. Além disso, tais informações

devem ser transparentes e disponibilizadas de forma acessível em sites oficiais, em locais visíveis a todas as pessoas.

Dentre tantos benefícios que podem ser colhidos através da abertura de dados na Administração Pública, destaca-se o benefício da própria sociedade poder contribuir com serviços inovadores ao cidadão. Aqui, organizações, cidadãos, acadêmicos e instituições públicas, munidos das bases de dados abertas, além de terem a oportunidade de poder participar da gestão pública, têm a possibilidade de pensar em soluções inovadoras que melhorem a vida de toda a sociedade, gerando impacto e interesse social em áreas relevantes da administração pública, tais como [16]:

- **Educação:** Aplicações utilizam dados abertos governamentais para informar à sociedade sobre o desempenho das escolas no país, efetuar comparações entre as escolas com base em diferentes tipos de indicadores, informar o desempenho dos alunos, etc. O exemplo mais completo de uma aplicação que utiliza dados relacionados à educação no Brasil é o Qedu [6], que será detalhado na seção "Trabalhos Correlatos".
- **Finanças Públicas:** Diversas aplicações utilizam os dados públicos para informar ao cidadão como andam os gastos públicos em um respectivo país. No Brasil, um dos projetos de maior sucesso é a Operação Serenata de Amor [17] — projeto aberto que utiliza ciência de dados e inteligência artificial para analisar os gastos reembolsados pela Cota para Exercício da Atividade Parlamentar, de deputados federais e senadores, identificando suspeitas e incentivando a população a questioná-los.
- **Transporte Público:** Outras iniciativas surgem com o objetivo de facilitar o deslocamento urbano das pessoas. É o caso do *Mobilab* [18] — laboratório de inovação aberta da cidade de São Paulo que surge com a diretriz de integrar as políticas de trânsito e transporte, principalmente, promover a transparência e análise dos dados produzidos pelos serviços, gerando novas soluções para uso da administração pública e para a sociedade, na melhoria da mobilidade urbana.

Portanto, munido do movimento pela abertura dos dados governamentais e o fortalecimento das políticas de transparência pública no Brasil, através da Lei de Acesso à Informação, o presente trabalho se utilizará de bases de dados públicas com o objetivo de inovar, ofertando um serviço que possa ser utilizado pela sociedade e tenha impacto na vida de pessoas que necessitam de Atendimento Educacional Especializado.

C. Ciência de Dados

Em abril de 2020, o número de pessoas com acesso à *internet* já ultrapassava 4 bilhões, número este que vem crescendo, significativamente, ano após ano. A cada minuto, dados são gerados constantemente em cliques de anúncios, reações e compartilhamento em mídias sociais, transações, conteúdo de *streaming*, etc. Segundo dados da *International Data Corporation (IDC)*, em 2025 o mundo terá aproximadamente 41,6 bilhões de dispositivos conectados à *internet*, gerando 79,4 *zetabytes* de dados [19] [20].

Quando todos esses dados são obtidos e bem analisados, podem se transformar em informações valiosas para grandes empresas, que os utilizam para tomadas de decisões e criações de estratégias de negócio. Consequentemente, alimentando aplicações que têm como base o aprendizado de máquina.

Em 2005, observando esse crescente aumento exponencial de dados nas últimas décadas, Roger Mougaldas cunhou o termo *Big Data* [21]. Para Mougaldas, *Big Data* se refere ao conjunto de dados que é quase impossível de gerenciar e processar usando ferramentas tradicionais de *Business Intelligence* — processo que exige dados altamente estruturados e organizados em linhas e colunas para apresentação de relatórios precisos e visualizações simples. Portanto, *Big Data* é um termo para se relacionar ao processo de coleção de dados extremamente grandes e complexos, onde sistemas tradicionais de gerenciamento de dados *RDBMS* [22], por suas limitações, não conseguiriam colher e tratá-los, visto que os dados, na contemporaneidade, possuem um grande volume, são diversos e gerados em grande velocidade [23] [24].

Como consequência, torna-se indispensável pensar em um método que consiga trabalhar com toda essa quantidade e características dos dados produzidos na atualidade. A Ciência de Dados, portanto, é o termo cunhado para se referir a um conjunto de habilidades e técnicas usadas para extrair valor dos dados. Essa técnica envolve o uso de métodos para colher, analisar grandes quantidades de dados de diferentes fontes, formatos e tamanhos e extrair conhecimentos neles contidos [25]. Segundo Cady (2017, p.1), Ciência de Dados significa “fazer trabalho de análise que, por uma razão ou outra, requer uma quantidade substancial de habilidades de engenharia de *software*” [26]. Já para Foreman (2013, p.23), Ciência de Dados “é a transformação de dados, usando matemática e estatística, em *insights*, decisões e produtos valiosos”, quando esta é aplicada à área de negócios [27].

Portanto, ciência de dados trabalha com uma grande variedade de dados de uma forma mais prospectiva e exploratória para poder analisar o presente e possibilitar, através de dados colhidos de diversas fontes, gerar informações e conhecimento sobre tempos presentes — ou com as séries temporais, tempos passados — e auxiliar na tomada de decisões para o futuro. Ou seja, a Ciência de Dados se concentra em assuntos relacionados a *como e por que* determinados eventos ocorrem [25, p. 12].

No entanto, a Ciência de Dados não está apenas atrelada ao setor privado — transformando grandes quantidades de dados brutos em *insights* de negócios e auxiliando as empresas nas tomadas de decisões e criações de novas estratégias, visando melhores resultados e lucros — mas também ao terceiro setor, que utiliza a ciência de dados com foco no bem social — ajudando organizações sem fins lucrativos e órgãos governamentais com o propósito de melhorar seus serviços, consequentemente, o mundo e a vida das pessoas [28].

No Brasil, iniciativas tecnológicas oriundas de órgãos do terceiro setor que se propõe a atuar sobre um problema social, consequentemente, contribuir com o desenvolvimento da sociedade, são chamadas tecnologias sociais [29]. Apesar de haver dissenso sobre o seu conceito na atualidade, entende-se por tecnologia social “produtos, técnicas e/ou metodologias reaplicáveis, desenvolvidas na interação com a comunidade representando efetivas soluções de transformação social” [30,

p. 8]. Detalhes sobre esse tema será melhor exposto na seção *Tecnologia Social*.

D. Análise de Dados

Em 1962, John W. Turkey, em seu artigo *The Future of Data Analysis*, pensou a Análise dos Dados como, dentre outras coisas, um procedimento ou forma de planejar a coleta de dados para tornar sua análise mais exata e precisa, onde todo um maquinário e resultados de estatísticas deveriam ser aplicadas à análise dos dados [31, p. 2].

No mesmo artigo, Turkey profetiza o que viria a ser a Análise de Dados nos dias atuais, concluindo que esta ciência poderia contribuir com a superação de dificuldades reais e a prestação de um grande serviço a todos os campos da ciência e da tecnologia.

“The future of data analysis can involve great progress, the overcoming of real difficulties, and the provision of a great service to all fields of science and technology. Will it? That remains to us, to our willingness to take up the rocky road of real problems in preference to the smooth road of unreal assumptions, arbitrary criteria, and abstract results without real attachments. Who is for the challenge?”
(John W. Turkey)

O desígnio central do pensamento de Turkey era reverberar e emergir a Análise de Dados como um campo legítimo da estatística, contudo, distinto da estatística matemática [31].

Não muito diferente dos dias atuais, estudiosos acreditam que um projeto de análise de dados consiste em um ciclo composto por seis fases — 1) Descoberta; 2) Preparação; 3) Planejamento do Modelo, 4) Construção do Modelo, 5) Comunicação dos Resultados e 6) Operacionalização [25, p. 26].

- 1) **Descoberta** — A primeira fase consiste no momento em que hipóteses iniciais costumam ser levantadas pelo pesquisador, podendo as mesmas serem testadas, posteriormente, com os dados atualmente disponíveis. Deve-se aprender o máximo possível sobre as fontes de dados necessárias e disponíveis para o projeto em que se quer construir; avaliar os recursos disponíveis, que podem incluir tecnologia, ferramentas, sistemas e pessoas; identificar as principais partes interessadas no projeto, que devem incluir qualquer pessoa que se beneficiará com o projeto ou será significativamente impactada por ele. Portanto, é de suma importância que nesta fase o pesquisador adquira o máximo de conhecimento sobre seus dados e o problema que deseja resolver.
- 2) **Preparação** — A fase de preparação dos dados costuma ser a fase onde pesquisadores gastam, ao menos, 50% do seu tempo. Aqui é necessário explorar, limpar, padronizar, organizar e condicionar os dados antes de preparar sua modelagem para futuras análises. Nesta etapa, costuma-se remover colunas irrelevantes ao projeto, identificar *outliers* (valores atípicos, que fogem da normalidade), quantidade de colunas com variáveis nulas, realizar relacionamentos

entre outras tabelas, etc. Em alguns casos, a visualização dos dados também pode ser realizada com o objetivo de ajudar na compreensão dos mesmos.

- 3) **Planejamento** — A terceira fase se refere ao planejamento de um modelo onde equipes determinam os métodos, técnicas e fluxos do trabalho que se pretende seguir durante a fase de construção de um modelo. Aqui, também, a equipe explora os dados para aprender sobre os relacionamentos que existem entre as variáveis e, conseqüentemente, seleciona variáveis-chave e os modelos mais adequados.
- 4) **Construção** — Na construção, desenvolvem-se os bancos de dados para teste, treino e ambiente de produção. Ademais, nesta fase as equipes constroem e executam modelos baseados no trabalho realizado na fase anterior. A equipe também considera se existem ferramentas que serão suficientes ou adequadas para executar os modelos, ou se será necessário um ambiente mais robusto para execução dos mesmos.
- 5) **Comunicação** — Nesta etapa os resultados são avaliados a partir dos critérios desenvolvidos na primeira fase, em colaboração com as partes interessadas do projeto. As principais descobertas devem ser identificadas, o valor do negócio quantificado e uma narrativa deve ser desenvolvida para resumir e comunicar as descobertas.
- 6) **Operacionalização** — Na operacionalização são entregues relatórios finais, orientações, código e documentação técnica. Também poderá ser executado um projeto-piloto.

Ainda em um projeto de Análise de Dados, encontramos quatro tipos de estruturas de dados — Estruturados, Semiestruturados, Quase estruturados e Não Estruturados [25, p. 5].

- **Estruturados** — Dados com estrutura muito bem definida como arquivos CSV, Planilhas e Sistemas de Gerenciamento de banco de dados relacional.
- **Semiestruturados** — Dados textuais que costumam seguir algum padrão, como arquivos XML.
- **Quase Estruturados** — Dados textuais que normalmente necessitam de esforço, ferramentas e tempo para serem trabalhados.
- **Não Estruturados** — Dados sem estruturas, como documentos de texto, PDF's, imagens e vídeo.

Para fazer a análise dos dados nesse trabalho, utilizou-se o *Pandas* — uma ferramenta de análise e manipulação de dados de código aberto, rápida, poderosa, flexível e de fácil manuseio, construída sobre a linguagem de programação *Python*. A biblioteca disponibiliza estruturas de dados e ferramentas que foram idealizados para facilitar o processo de tratamento e análise de informações utilizando a linguagem de programação supracitada [32] [33].

Dentre as principais características da ferramenta destacam-se:

- *DataFrame* — Objeto rápido e eficiente para a manipulação de dados com indexação integrada.

- Leitura e escrita de dados em diferentes formatos: *CSV*, *JSON*, base de dados *SQL*, arquivos de texto, planilhas e outros.
- Remodelagem flexível de conjuntos de dados.
- Alto desempenho na união e junção dos conjuntos de dados trabalhados.
- Altamente otimizado para desempenho com alguns trechos de códigos escritos em *Python* e *C*.
- Amplamente usado para fins acadêmicos e comerciais — Finanças, Neurociência, Economia, Estatística, Publicidade e outros.

E. Fusão de Dados

Como já foi supracitado, temos como uma das fases do ciclo de um projeto de Análise de Dados, a preparação dos dados. Aqui, utiliza-se um dos padrões mais comuns do *pipeline* de dados — o *ETL*, que funciona como um tipo de *integração de dados*² em três etapas — Extração, Transformação e Carregamento — movendo e transformando dados de várias fontes para um novo destino, onde um novo valor poderá ser derivado [25] [34].

A etapa de **extração** é onde os dados de diversas fontes são reunidos para, posteriormente, serem carregados e transformados [34].

A etapa de **transformação** é onde dados são combinados e formatados para serem possíveis de ser trabalhados no projeto de análise de dados, conforme as necessidades do projeto [34].

Já na etapa de **carregamento**, os dados que foram transformados são alocados em seu local de destino [34].

Projetos de análise de dados, normalmente, requerem dados oriundos de diferentes lugares, formatos e tamanhos para serem integrados antes de serem, de fato, analisados. A maior barreira para a análise de dados corporativos, e também projetos de aprendizado de máquina, tende a ser a integração de dados, visto que a extração e transformação dos dados, com o objetivo de dar-lhes algum sentido, não costuma ser um trabalho trivial [35] [36].

Assim, a Fusão de Dados pode ser considerada como última etapa de processos voltados à integração de dados. Primeiro mapeia-se um esquema que será transformado, renomeado ou reestruturado para, em seguida, utilizando técnicas de detecção de duplicatas, encontrar possíveis inconsistências nos dados. Por último, as inconsistências são resolvidas e os dados fundidos em uma única representação — *Data Fusion* [37].

Um exemplo de aplicação de ferramentas com o apoio à fusão de dados é o *Cloud Data Fusion*³ — Um serviço totalmente gerenciado de integração de dados corporativos com nuvem nativa, podendo ser usado para gerar e gerenciar *pipelines* de dados [38].

²<https://www.ibm.com/analytics/data-integration>

³<https://www.youtube.com/watch?v=kehG0CJw2wo>

F. Visualização de Dados

A visualização de dados, tal qual conhecemos hoje, é produto de uma variedade de avanços que ocorreram ao longo dos séculos anteriores. A consagração da visualização de dados se deu, indubitavelmente, com a ascensão de áreas como a probabilidade, estatística, astronomia, cartografia, gráficos estatísticos, dentre outras áreas. Por último, mas não menos importante, com o advento, também, da Análise de Dados [39].

Na contemporaneidade, a visualização é parte essencial de um processo de análise de dados, pois permite visualizar características e padrões em uma grande quantidade de dados, criando assim, diferentes visualizações para a mesma. Um dos muitos recursos aplicados durante o processo de visualização de dados é o uso dos gráficos que, quando bem utilizados, permitem ao pesquisador contar histórias de maneira eficaz. Para Tufte, gráficos estatísticos apenas são eficazes quando estes comunicam ideias complexas com clareza, precisão e eficácia [40].

Segundo Few (2004, p.5), a maioria dos dados encontrados, regularmente, em diversas empresas, independente do setor, serão tratados e posteriormente visualizados, podendo usar uma combinação qualquer, a partir de sete tipos de mensagens quantitativas. Conforme o pensamento do autor, mensagens quantitativas sempre revelam relacionamentos que contam histórias que merecem atenção e, quase toda mensagem quantitativa que precisa ser apresentada em um gráfico, pode ser descrita com uma dos seus sete tipos, ou uma combinação delas [41]. Abaixo os sete tipos:

- **Comparação Nominal** — Comparação simples de subdivisões categóricas de uma ou mais medidas em nenhuma ordem particular.
- **Classificação** — Subdivisões categóricas de uma medida ordenada pelo tamanho.
- **Time Series** — Múltiplas instâncias de uma ou mais medidas feitas a partir de pontos equidistantes no tempo.
- **Das partes para o Todo** — Medidas de subdivisões categóricas individuais como proporções do todo.
- **Desvio** — Subdivisões categóricas de uma medida, comparada a uma medida de referência comunicadas a partir das diferenças entre elas.
- **Distribuição de Frequência** — Quantidade de algo a partir de subdivisões categóricas (intervalos) de um intervalo quantitativo.
- **Correlação** — Comparações de dois conjuntos de medidas para determinar se enquanto um conjunto aumenta, o outro também aumenta ou diminui, de maneira correspondente, e caso isso aconteça, quão fortemente.

No presente trabalho, para a visualização dos dados utilizou-se a *geovisualização*. No âmbito da geovisualização, encontramos diferentes definições. Uma das mais aceitas foi atualmente definida em 2001 pela Comissão de Visualização e Ambientes Virtuais da *International Cartographic Association (ICA)* [42] e caracteriza a geovisualização como um

processo que “integra abordagens de Visualização em Computação Científica (*ViSC*), cartografia, análise de imagens, visualização de informações, análise exploratória de dados (*EDA*)⁴ e sistemas de informação geográfica (*GISystems*)⁵ para fornecer teoria, métodos e ferramentas para exploração visual, análise, síntese e apresentação de dados geoespaciais”⁶ [43, p. 3-12 tradução nossa] [44].

Já para Kraak, a geovisualização pode ser definida como: “O uso de *displays* geoespaciais visuais para explorar dados e, por meio dessa exploração, gerar hipóteses, desenvolver soluções de problemas e construir conhecimento” [45, p. 390–399 tradução nossa] [44].

Para a elaboração da visualização dos dados de interesse deste projeto, utilizou-se o *GeoPandas* — projeto de código aberto idealizado para possibilitar e facilitar o trabalho com dados espaciais e georreferenciados utilizando a linguagem de programação *Python*. O *GeoPandas* estende os tipos de dados usados pelo *Pandas* para permitir operações espaciais em tipos geométricos. As operações geométricas são executadas pela biblioteca *shapely* [46]. Além disso, o projeto ainda depende de duas bibliotecas — o *fiona* [47] para acesso a arquivos e o *matplotlib* [48] para plotagem de dados em mapas. Além desses, é necessário o uso de um projeto de código aberto — *descartes* [49] — também, para a plotagem de dados em mapas.

Para se trabalhar com o *Geopandas*, faz-se necessário o entendimento de três importantes elementos:

- **Shapefile** — Formato de arquivo que contém dados geoespaciais em formato de vetor muito utilizado por sistemas de informações geográficas. Foi criado e regulamentado pela *ESRI* [50] — empresa americana especialista no desenvolvimento de produtos para a área de *GISystems*. Os *shapefiles* representam geometrias, como pontos, polígonos e linhas, que podem representar municípios, estados, continentes, relevos, rios, entre outras informações.
- **CRS** — O Sistema de Referência de Coordenadas é uma forma padronizada de escrever as localizações no globo terrestre. Existe mais de um *CRS* e sua escolha depende de um conjunto de fatores como a abrangência geográfica ou em que época os dados foram coletados. Quando se comparam conjuntos de dados com *crs*'s diferentes, é importante estabelecer um *crs* em comum entre eles para torná-los comparáveis. Existem três parâmetros que, geralmente, aparecem na maioria das configurações de *crs*'s: *proj*, *ellps* e *datum*. O primeiro desses parâmetros se refere à **Projeção** — o processo de criar uma representação bidimensional do globo terrestre. Os três principais tipos de projeção são cilíndricas, cônica e plana. O segundo parâmetro é o que define a forma da Terra, onde algumas configurações, possíveis para ele,

⁴Exploratory Data Analysis

⁵Geographic Information System

⁶“Geovisualization integrates approaches from visualization in scientific computing (*ViSC*), cartography, image analysis, information visualization, exploratory data analysis (*EDA*), and geographic information systems (*GISystems*) to provide theory, methods and tools for visual exploration, analysis, synthesis, and presentation of geospatial data”

podem ser *WGS84* [51], *GRS80* [52] e *CLRK80*. Por fim, o *datum* é responsável por ancorar as coordenadas da Terra, determinando as origens e a direção dos eixos.

Para este trabalho os parâmetros escolhidos foram *WGS84* e *SIRGAS2000*⁷ onde, o primeiro é utilizado pelo sistema de *GPS* — Sistema de Posicionamento Global — e o segundo, oficialmente utilizado pelo Brasil, é o sistema de referência geocêntrico para América Latina.

- **Universal Transverse Mercator** — O *UTM* divide a terra em 60 zonas e cria uma projeção para cada uma delas, além de ser muito utilizado em pesquisas, pois possui alguns atributos que tornam a estimativa das distâncias mais precisas. Quando se declara um *crs* utilizando o parâmetro *UTM*, é necessário configurar alguns parâmetros adicionais, como a zona, o hemisfério e as unidades de medida que estão sendo trabalhadas. No caso do presente trabalho, a forma escolhida para configurar o *crs* com o *UTM* foi em formato de dicionário, como mostra a Listagem 1.

```
1 crs = {'proj': 'latlong', 'ellps': 'WGS84', 'datum':  
      'WGS84', 'no_defs': True}  
2 geo_estacoes = gpd.GeoDataFrame(estacoes, crs = crs,  
      geometry = geometry)
```

Listing 1. Algoritmo responsável pela configuração do *crs* com o *UTM*.

G. Tecnologias Sociais

O conceito de Tecnologia Social surge no Brasil em paralelo à formação da Rede de Tecnologia Social. Dessa rede, participam autores que demonstraram uma inquietação acerca da crescente exclusão social, precarização, informalidade do trabalho e, conseqüentemente, aspiravam a necessidade de se pensar tecnologias que correspondessem as suas expectativas. Dela, fazem parte órgãos governamentais, empresas estatais, órgãos privados de fim público, universidades, ONGs e movimentos sociais [30, p. 10].

Entende-se por tecnologia social “produtos, técnicas e/ou metodologias reaplicáveis, desenvolvidas na interação com a comunidade e que representam efetivas soluções de transformação social”. Tal tecnologia seria dessemelhante a “Tecnologia Capitalista” — produzida pela e para a empresa privada [30, p. 8].

No âmbito dessa discussão, fez-se surgir o debate acerca do papel das Instituições Públicas de ensino, principalmente as Universidades Públicas, na construção de iniciativas voltadas às políticas públicas, políticas essas que atendessem as demandas sociais — contribuindo com a redução da desigualdade social e o suporte, através das tecnologias sociais, às classes mais desfavorecidas e marginalizadas no Brasil [53] [54] [29].

Segundo Almeida, nos últimos anos, no Brasil, a cobrança por uma universidade mais inclusiva e mais próxima das demandas sociais, tem crescido significativamente, cobrança esta oriunda de movimentos sociais, organizações não governamentais e de governos, focados em compromisso social. Tais pressões são decorrentes da percepção de que os investimentos

públicos, financiados pelo esforço coletivo da sociedade devem ter uma contrapartida para com ela [53] [55].

Feitas essas considerações, esta iniciativa visa contribuir, também, no que diz respeito as tecnologias sociais, haja visto o compromisso socio-comunitário que o aluno de uma Instituição Federal deveria ter para com a sociedade, aquela que contribui para os benefícios colhidos por ele, dentro daquela instituição.

III. TRABALHOS CORRELATOS

Existem diversas soluções que permitem ao cidadão a visualização de informações oriundas de dados brutos publicados pelas instituições governamentais. Dentre essas soluções, têm-se aquelas que utilizam os microdados da educação brasileira, publicados pelo INEP — Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira — para facilitar a visualização de informações relevantes sobre cada escola no Brasil. Tais aplicações apresentam como principal característica a visualização dos dados.

O *Análise do Acesso à Educação* [56] é um site de visualização de dados cujo objetivo é demonstrar a evolução do acesso à Educação Básica dos surdos e pessoa com deficiência auditiva no Brasil, no período de 2008 a 2012. Além de analisar o impacto do Decreto n.º 5626/05 e da Lei n.º 10436/02 — que reconhece oficialmente a Língua brasileira de sinais (Libras) e dá outras providências. Enquanto o Decreto n.º 5626/05 regulamenta a língua de sinais brasileira como componente curricular nas escolas e dá outras atribuições — Ele busca principalmente fazer compreender aspectos sobre o acesso à educação pelas pessoas com perda auditiva (surdos, deficientes auditivos e surdocegos). A pesquisa teve recorte na educação básica — período de 2008 a 2012 — e na educação superior — período de 2009 a 2012.

O *QEDU* [6] — o mais completo de todos os trabalhos aqui citados — é um aplicativo onde, para cada escola, cidade, estado ou para o Brasil, entrega acesso fácil, rápido e intuitivo a diversos dados educacionais, como a Prova Brasil, o Censo Escolar, Ideb e Enem. Todos eles obtidos de fontes oficiais do Governo para ajudar educadores, pais e cidadãos em geral a melhorar a educação brasileira. Atualmente o site conta com informações como:

- Proporção de alunos com aprendizado adequado à sua etapa escolar — calculado com base em dados oriundos da Prova Brasil 2019.
- Perfil dos alunos do 5^a ao 9^a ano — dados oriundos da Prova Brasil 2019.
- Número de matrículas para cada etapa escolar — dados oriundos do Censo Escola 2020.
- Proporção de alunos com reprovação ou abandono em 2019 segundo indicadores do INEP.
- Infraestrutura escolar, que informa sobre existência de bibliotecas, quadras esportivas, acessibilidade etc — dados oriundos do Censo Escolar 2020.
- Ideb — calculado com base no aprendizado dos alunos em português e matemática (Prova Brasil).

⁷<http://www.sirgas.org/pt/sirgas-realizations/sirgas2000/>

O *Raiox* [5] é uma ferramenta exclusiva que reúne informações sobre todas as escolas do país. São dados referentes às mais de 180 mil unidades de ensino fundamental e médio em território nacional. Os dados utilizados na ferramenta são os do Censo Escolar (2018), do Enem (2018), do Ideb (2017) e do Saeb (2017) — todos do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira — INEP. Para que o usuário possa usufruir do site, precisará informar a escola que deseja obter mais informações. Uma vez a escola encontrada, detalhes sobre o Ideb, Índices de Educação, Infraestrutura e Línguas da mesma serão exibidos. Sobre acessibilidade, a única informação disponível é a existência ou não de dependências para deficientes.

Embora tenham sido apresentados alguns projetos que visam contribuir com a melhoria da educação básica no país através da utilização de dados abertos, tais aplicativos não têm como principal objetivo a utilização destes dados como forma de auxiliar pais e responsáveis, por pessoas com necessidades especiais, na escolha de escolas possivelmente mais adequadas para elas. O projeto QEDU, por exemplo, não apresenta as escolas que possam ser mais adequadas para um usuário — baseado nas características que esse busca — limitando-se apenas a apresentar uma visualização de dados da porcentagem de escolas que possuem dependências e banheiros acessíveis. Caso um usuário queira buscar informações sobre as escolas nesse site, necessita saber, de antemão, o nome da escola, para em seguida, obter informações relacionadas a acessibilidade daquela escola — informações estas que estão limitadas a uma escola possuir ou não dependências acessíveis.

O *Raiox* das Escolas do Brasil, por sua vez, apresenta muito mais limitações que o primeiro. Este, apesar de mostrar informações relevantes sobre uma determinada escola, apresenta informações sobre acessibilidade da mesma forma que o QEDU. Além disso, o aplicativo além de obrigar que o usuário saiba, de antemão, qual escola está querendo buscar, não possui dados atualizados — concentrando-se apenas em dados referentes aos anos de 2018 e 2017.

Neste contexto, o presente trabalho visa construir um produto minimamente viável para auxiliar na tomada de decisão dos pais, ou responsáveis diretos por pessoas com deficiências, acerca das escolas soteropolitanas, possivelmente, mais adequadas para pessoas que necessitem de Atendimento Educacional Especializado.

IV. METODOLOGIA DE TRABALHO

O presente trabalho foi desenvolvido através de uma pesquisa de natureza qualitativa, pois visa gerar conhecimentos para aplicações práticas dirigidos à solução de um problema real. Por se tratar de um projeto que visa realizar a análise exploratória de um problema real, o mesmo tem o objetivo de ser exploratório, dado que faz um levantamento bibliográfico e realiza entrevistas com pessoas que possuem experiência prática no tema abordado — Atendimento Educacional Especializado. Por fim, quanto a abordagem, o trabalho apresenta tanto um caráter quantitativo — não só enumera a quantidade de escolas aptas para o oferecimento do AEE, mas a quantidade de professores aptos para tal, número de dependências físicas para o atendimento de pessoas especiais, etc — como qualitativo — preocupa-se também com a formação adequada dos

professores, existência de salas multifuncionais com recursos socioeducativos utilizados no desenvolvimento dos estudantes, oferecimento de alimentação, etc.

O trabalho teve como primeira etapa a identificação e extração dos dados relevantes para o projeto, a leitura de documentação referente aos microdados do Censo Escolar 2019, bem como a exploração e preparação dos Dados. Em seguida, foi desenvolvido um protótipo que, ao final, foi mostrado a especialistas, que lidam direta e indiretamente com o AEE, no intuito de obter diferentes percepções sobre o projeto. Uma vez as observações colhidas, procedeu-se com o aprimoramento do protótipo.

V. SOLUÇÃO DESENVOLVIDA

Nesta seção será apresentado o passo a passo na construção do *SELECTOR*, um sistema auxiliador na tomada de decisão dos pais, ou responsáveis diretos por pessoas com deficiência, acerca das escolas soteropolitanas, considerando a adequação de cada instituição ao Atendimento Educacional Especializado. A primeira etapa mostrará a extração, exploração, preparo, fusão e visualização dos microdados do Censo Escolar de 2019, dados do Catálogo de Escolas Detalhadas, dados oriundos da extração de sites que contêm informações sobre os transportes coletivos do município de Salvador e dados geoespaciais do município de Salvador.

Por fim, a segunda etapa mostrará a construção de um sistema que retorna, em ordem decrescente, as escolas que tiverem uma maior pontuação, dada a existência, ou não, de recursos relacionados ao Atendimento Educacional Especializado. Nela, o usuário poderá inserir três importantes informações acerca do tipo de escola que procura:

- **Dependência da Escola** — público ou privada.
- **Mediação Didático Pedagógica** — presencial, semi-presencial ou EAD.
- **Etapa de Ensino** — creche, primário, fundamental, etc.

Portanto, como mostra a figura 1, temos uma solução que através de requisições HTTP, consome uma API *Restful* [57] construída em *Java* [58] e *SpringBoot* [59]. Essa API acessa uma base de dados composta pelos seguintes dados governamentais:

- 1) Microdados do Censo da Educação Básica 2019.
- 2) Catálogo das Escolas Detalhadas Brasileiras.
- 3) Organização Territorial do Município de Salvador.
- 4) Dados sobre as estações de metrô da capital baiana — retiradas do site da CCR Metrô.
- 5) Dados referentes aos pontos de ônibus do município de Salvador.

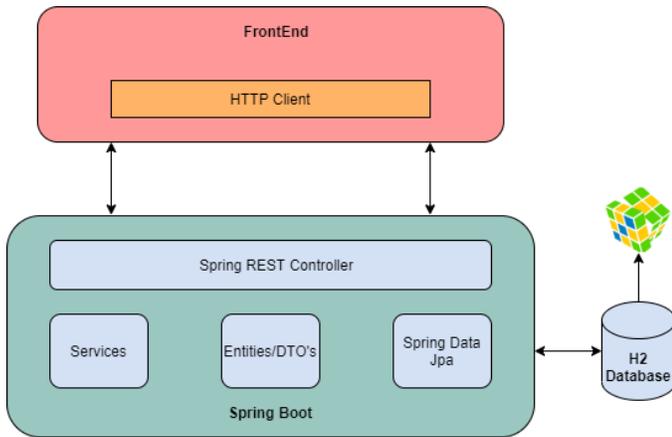


Figura 1. Arquitetura da solução proposta para o Selector. Fonte: Autor.

A. Pipeline para a Análise dos Dados

Como já mencionado na Revisão Bibliográfica e Teórica, na subseção Análise de Dados, todo e qualquer trabalho que envolva a análise de dados consiste em, pelo menos, seis fases: Extração, Preparação, Planejamento do Modelo, Construção do Modelo, Comunicação dos Resultados e Operacionalização [25]. Para a construção da primeira parte do sistema SELECTOR, seguiu-se o fluxo mostrado na figura 2, onde cada etapa será elucidada a seguir.

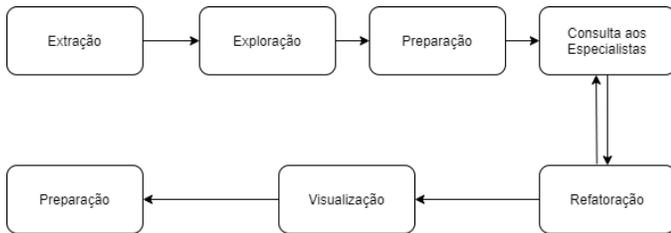


Figura 2. Fluxo do processo de Análise de Dados adotado para este trabalho. Fonte: Autor.

B. Extração

A fase de extração consiste na obtenção de dados oriundos de diversas fontes: aplicativos mobile, sensores, páginas web ou até mesmo dados estruturados, semiestruturados e não estruturados, oriundos de bases convencionais, suportadas ou não por SGBD [60]. No caso deste trabalho, os dados coletados consistiram em dados governamentais, tais como:

- Microdados do Censo da Educação Básica 2019 [61].
- Catálogo das Escolas Detalhadas brasileiras [62].
- Organização Territorial do Município de Salvador [63].
- Dados sobre as estações de metrô da capital baiana — retiradas do site da CCR Metrô [64].
- Dados referentes aos pontos de ônibus do município de Salvador — retirados do *OverpassTurbo*.

1) *Microdados do Censo Escolar 2019*: Os microdados do Censo da Educação Básica de 2019 são dados estruturados exportados em formato CSV. Tais dados foram projetados para que qualquer *software* estatístico, inclusive de código aberto, pudesse importar e carregá-los. Os microdados contêm dados referentes a todas as escolas do Brasil, alunos matriculados, docentes e turmas correspondentes a essas escolas.

2) *Catálogo das Escolas Detalhadas*: Os dados oriundos deste catálogo também são dados estruturados exportados em formato CSV, disponíveis no site do INEP [62].

3) *Malha Territorial do Município de Salvador*: A *Malha Territorial* retrata a situação vigente da Divisão Política Administrativa da capital soteropolitana, através de representações vetoriais das linhas definidoras das divisas estaduais e limites municipais [63].

4) *Dados das estações de Metrô*: Os dados das estações de metrô — estação, linha e endereço — foram colhidos no próprio site da CCR Metrô Bahia — empresa do *Grupo CCR* [64], responsável pela construção, manutenção e operação do sistema metroviário de Salvador e Lauro de Freitas. Durante a extração desses dados — Quase Estruturados — foi necessário um conjunto de ações para encontrar as informações das estações e suas respectivas geolocalizações — latitude e longitude:

- 1) Uma vez no site da CCR Metrô Bahia, foi necessário clicar na imagem composta por um *link* que redirecionava para uma segunda página, composta pelas informações da respectiva estação, como mostra a figura 3;
- 2) Munido do endereço das estações, foi utilizada uma plataforma de mapeamento na web — *Google Maps* — para extrair a latitude e longitude, presente na URL do endereço de cada estação, além do CEP, presente na descrição do endereço, como mostra a figura 4. Seguindo este passo a passo, foi possível construir um arquivo CSV composto pelas informações referentes a todas as estações de metrô.

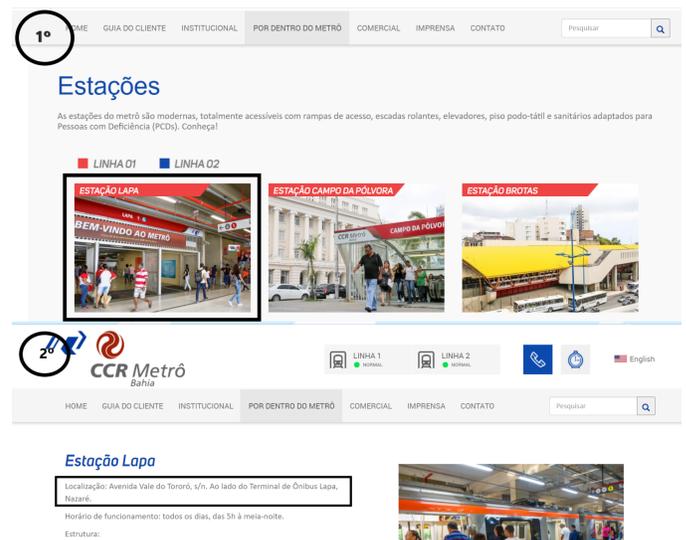


Figura 3. Impressão tirada para demonstrar o passo a passo para obtenção do endereço de uma determinada estação, no site da CCR Metrô.

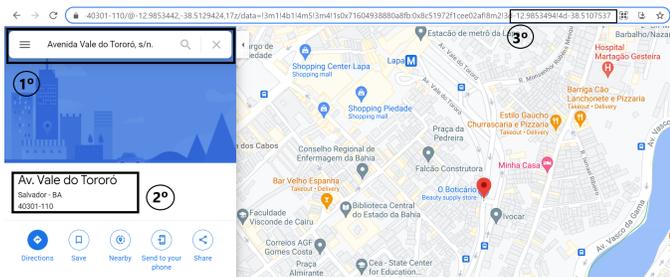


Figura 4. Impressão tirada para demonstrar o passo a passo para obtenção da Latitude e Longitude de cada estação de metrô.

5) *Geolocalização dos pontos de ônibus em Salvador:* Para obter as informações referentes a geolocalização das paradas de ônibus em Salvador foi utilizada uma ferramenta de filtragem de dados baseada na *web* para *OpenStreetMap*⁸ — *OverpassTurbo*⁹.

Com o *OverpassTurbo*, é possível executar consultas de API do *Overpass*, como mostra a Listagem 2, e analisar os dados resultantes, interativamente, em um mapa. Uma vez os dados obtidos, basta exportá-los na formatação desejada — nesse caso, *CSV*.

```

1 <query type="node">
2 <has-kv k="highway" v="bus_stop"/>
3 <hbox-query {{bbox}}/>
4 </query>
5 <print/>

```

Listing 2. Consulta responsável pela obtenção das paradas de ônibus em Salvador.

C. Exploração

Uma vez coletados os conjuntos de dados necessários ao projeto, fez-se necessária a familiarização com os mesmos. Nesta fase, o objetivo foi conhecer o máximo possível dos dados para identificar ajustes a serem realizados na próxima fase do fluxo — preparação. É importante ressaltar que não foi possível fazer a exploração das tabelas de Matrícula, Docente e Turma, haja visto o tamanho dos arquivos e o escasso recurso de *hardware* da máquina utilizada neste projeto, sendo apenas possível trabalhá-las na fase de preparação, no ambiente *online Colab Research* [65].

1) *Escola:* Com a leitura do dicionário de dados, onde se encontram as descrições para cada campo das tabelas, foram identificadas as três principais colunas utilizadas na filtragem dos registros — *TP_SITUACAO_FUNCIONAMENTO*, *CO_MUNICIPIO* e *TP_AEE*, como mostra a tabela I. O projeto, portanto, trabalhou com escolas que oferecem atendimento AEE de forma exclusiva e não exclusiva, além daquelas que estão com a regulamentação — regulamentação/autorização no concelho ou órgão municipal, estadual ou federal de educação — ativa ou em tramitação.

Tabela I. DICIONÁRIO DE DADOS DAS VARIÁVEIS UTILIZADAS.

VARIÁVEL	DESCRIÇÃO
CO_MUNICÍPIO	Código do Município
TP_SITUACAO_FUNCIONAMENTO	Situação de funcionamento 1 - Em Atividade 2 - Paralisada 3 - Extinta (ano do Censo) 4 - Extinta em Anos Anteriores
TP_AEE	Atendimento Educacional Especializado (AEE) 0 - Não oferece 1 - Não exclusivamente 2 - Exclusivamente

Com a leitura do documento [66], presente nos microdados do INEP, obteve-se, também, informações sobre os dados da tabela de Escola. Nele é informado que para contabilizar o número de escolas no Brasil - conforme a Sinopse Estatística da Educação Básica 2019 [67] — faz-se mister uma junção entre as tabelas Escola e Matrícula, através do atributo *TP_TIPO_ATENDIMENTO_TURMA IN (1,2)*. A última variável citada, presente na tabela de Matrícula, refere-se ao tipo de atendimento no qual um aluno foi matriculado — Exclusivo de Escolarização e Atividade complementar e escolarização, respectivamente os valores 1 e 2. No entanto, o tipo de atendimento no qual esse projeto visa trabalhar é o Atendimento Educacional Especializado. Ainda de acordo com o documento, somos informados sobre a existência da variável *IN_ESPECIAL_EXCLUSIVA*, por meio da qual conseguimos informações referentes a possibilidade de uma determinada escola possuir ou não classes exclusivas para alunos que necessitam de AEE.

Feito isso, foi realizada a leitura do arquivo *CSV* através da utilização do *Pandas* e da linguagem de programação *Python*. Dessa forma, foram constatadas as seguintes informações: a base de dados apresenta 228,521 mil registros (escolas) e 234 variáveis (colunas).

2) *Escolas Detalhadas:* Na análise do Catálogo de Escolas Detalhadas, disponibilizado em formato *CSV* pelo INEP, foi possível obter inúmeras informações semelhantes à tabela de Escolas do Censo Escolar de 2019. No entanto, a segunda possuía informações sobre as coordenadas geográficas — latitude e longitude — endereço e telefone. Dessa forma, com a utilização do *Pandas*, constatou-se que o arquivo continha 222,936 mil registros (escolas), 5,585 mil a menos que as escolas contabilizadas no Censo de 2019, e 19 variáveis (colunas).

D. Preparação

Na fase de preparação foram aplicados alguns ajustes nos dados — buscando melhorar sua utilização para análise. Aqui, os valores ausentes e as variáveis irrelevantes ao projeto foram retiradas das tabelas. Além disso, filtros para obter registros interessantes ao projeto foram feitas e junções com outras tabelas foram realizadas.

1) *Escola:* Na tabela de Escolas do Censo de 2019, foram feitas as filtros para obter apenas as escolas soteropolitanas em atividade, que estão regulamentadas e que fornecem Atendimento Educacional Especializado de forma exclusiva e não exclusiva. Sendo assim, foi constatado que apenas 94 escolas

⁸<https://www.openstreetmap.org/>

⁹<https://overpass-turbo.eu/>

estavam aptas para receber alunos necessitados de atendimento educacional especializado, no município de Salvador.

Além dessa mudança, também foi adicionada uma variável chamada *PONTUACAO* à tabela de Escola. Essa variável consiste em uma pontuação dada a cada escola — que será necessária para a ordenação das mesmas — observando os recursos de acessibilidade nelas contidas. Até aqui, a inclusão, ou não, de algumas variáveis se deu através da pesquisa até então realizada e do que se acreditava serem as variáveis mais interessantes ao projeto — aquelas ligadas às dependências físicas, a existência, ou não, de salas de recursos multifuncionais e da existência de alguns profissionais de saúde. As variáveis, portanto, escolhidas correspondem as variáveis da tabela II.

2) *Catálogo de Escolas Detalhadas*: No arquivo de escolas detalhadas, foi inicialmente realizada a filtragem das escolas de interesse para este projeto — escolas do município de Salvador, em atendimento e sem restrição de atendimento. Em seguida, a coluna “Escola”, contendo o nome da escola, foi modificada para se chamar *NO_ENTIDADE*. Dessa forma, seria possível fazer a junção da tabela de Escola (Censo Escolar 2020) com o Catálogo de Escolas Detalhadas.

3) *Junção de Escola (Censo Escolar 2019) com Catálogo de Escolas Detalhadas*: Tal junção foi indispensável para ser possível obter o endereço, telefone e geolocalização de todas as escolas oriundas do Censo Escolar de 2019, já que essas informações não são contempladas. Com as tabelas de Escola do Censo e Escolas Detalhadas preparadas, foi realizada a junção das tabelas pela coluna *NO_ENTIDADE* e, logo em seguida, a remoção de todas as colunas irrelevantes oriundas dessa junção. Para a junção dessas tabelas utilizou-se a ferramenta Pandas. No entanto, mesmo com a junção das tabelas, percebeu-se dezesseis escolas que ainda continham valores nulos para os atributos referentes a sua geolocalização — latitude e/ou longitude. Dessa forma, para obter tais informações, foram necessárias as seguintes etapas:

- 1) Munido do nome da escola, uma rápida pesquisa era feita. Caso não fosse possível obter o endereço, uma mensagem eletrônica ao responsável pela escola era enviada ou uma ligação feita.
- 2) Em posse do endereço, esse era inserido no *Google Maps* e suas coordenadas geográficas extraídas da barra de endereço.¹⁰

Ao final desse processo, foi importante procurar por escolas que ainda continham informações de endereço e telefone nulas.

E. Consulta aos Especialistas

Nesta fase escolheu-se consultar profissionais que atuam diretamente com pessoas que necessitam de Atendimento Educacional Especializado. Aqui foram entrevistadas cinco pessoas: um consultor em acessibilidade, uma especialista em Atendimento Educacional Especializado e três professores que realizam ou já realizaram esse tipo de atendimento em escolas público ou privadas.

Durante a entrevista foi feita uma contextualização do *SELECTOR* e informado como os dados do Censo Escolar

¹⁰Aqui foi aplicado o mesmo procedimento concretizado na subseção *Extração — Dados das estações de Metrô*.

Tabela II. TABELA DAS VARIÁVEIS UTILIZADAS ANTES DA CONSULTA AOS ESPECIALISTAS.

VARIÁVEL	DESCRIÇÃO
IN_ACESSIBILIDADE_CORRIMAO	Recursos de acessibilidade para pessoas com deficiência ou mobilidade reduzida nas vias de circulação interna na escola - Corrimão e guarda corpos
IN_ACESSIBILIDADE_ELEVADOR	Recursos de acessibilidade para pessoas com deficiência ou mobilidade reduzida nas vias de circulação interna na escola - Elevador
IN_ACESSIBILIDADE_PISOS_TATEIS	Recursos de acessibilidade para pessoas com deficiência ou mobilidade reduzida nas vias de circulação interna na escola - Pisos táteis
IN_ACESSIBILIDADE_VAO_LIVRE	Recursos de acessibilidade para pessoas com deficiência ou mobilidade reduzida nas vias de circulação interna na escola - Portas com vão livre de no mínimo 80 cm
IN_ACESSIBILIDADE_RAMPAS	Recursos de acessibilidade para pessoas com deficiência ou mobilidade reduzida nas vias de circulação interna na escola - Rampas
IN_ACESSIBILIDADE_SINAL_SONORO	Recursos de acessibilidade para pessoas com deficiência ou mobilidade reduzida nas vias de circulação interna na escola - Sinalização sonora
IN_ACESSIBILIDADE_SINAL_TATIL	Recursos de acessibilidade para pessoas com deficiência ou mobilidade reduzida nas vias de circulação interna na escola - Sinalização tátil (piso/paredes)
IN_ACESSIBILIDADE_SINAL_VISUAL	Recursos de acessibilidade para pessoas com deficiência ou mobilidade reduzida nas vias de circulação interna na escola - Sinalização visual (piso/paredes)
IN_SALA_ATENDIMENTO_ESPECIAL	Dependências físicas existentes e utilizadas na escola - Sala de recursos Multifuncionais para Atendimento Educacional Especializado (AEE)
IN_BANHEIRO_PNE	Dependências físicas existentes e utilizadas na escola - Banheiro acessível, adequado ao uso de pessoas com deficiência ou mobilidade reduzida
QT_SALAS_UTILIZADAS_ACESSIVEIS	Condições das salas de aula utilizadas na escola (dentro e fora do prédio escolar) - Número de salas de aula com acessibilidade para pessoas com deficiência ou mobilidade reduzida.
QT_PROF_SAUDE	Total de profissionais que atuam na escola - Bombeiro(a) brigadista, profissionais de assistência a saúde (urgência e emergência), Enfermeiro(a), Técnico(a) de enfermagem e socorrista
QT_PROF_FONAUDIOLOGO	Total de profissionais que atuam na escola - Fonoaudiólogo(a)
QT_PROF_NUTRICIONISTA	Total de profissionais que atuam na escola - Nutricionista
QT_PROF_PSIKOLOGO	Total de profissionais que atuam na escola - Psicólogo(a) Escolar
QT_PROF_PEDAGOGIA	Total de profissionais que atuam na escola - Profissionais de apoio e supervisão pedagógica: pedagogo(a), coordenador(a) pedagógico(a), orientador(a) educacional, supervisor(a) escolar e coordenador(a) de área de ensino.

de 2019 foram obtidos. Terminada essa pequena introdução, discutiu-se quais variáveis, em cada tabela, poderiam ser utilizadas na soma da pontuação das respectivas escolas. Para esta seção, destacaram-se as entrevistas de três profissionais e suas respectivas contribuições.

- 1) Educadora há 15 anos, Técnica de Educação Especial pela Secretaria de Educação do Estado da Bahia e Consultora em Educação Especial no SESI — Sistema FIEB Bahia [68].
- 2) Professora e Graduada em Pedagogia pela Universidade Federal da Bahia [69].
- 3) Professora e Pedagoga pela Universidade Católica de Salvador [70].

1) *Escola*: Para [68], [70] e [69], a existência de instrumentos, materiais socioculturais e/ou pedagógicos em uso na escola — para o desenvolvimento de atividades de ensino e aprendizagem — são tão importantes quanto as variáveis relacionadas a acessibilidade de algumas dependências físicas presentes nas mesmas — salas de artes, dança, música, e multiuso. As profissionais acreditam que uma escola contendo tais dependências físicas, sem a existência dos recursos, não conseguiria trabalhar no desenvolvimento de seus alunos.

“Não adianta os pais escolherem uma escola com professores ótimos e uma equipe muito bem preparada se a equipe não tem material disponível para usar no dia a dia, na aplicação das atividades. Vemos isso na realidade das escolas públicas que criam as salas de AEE, através do Governo do Estado, mas não disponibiliza o material. Sendo assim, nós, enquanto professores, precisamos comprar materiais com o nosso dinheiro para nossos alunos. Está errado fazer o professor ou profissional levar seu material. A escola ou unidade escolar precisa disponibilizar esses materiais.” (Carolina Sallenave Cambeses)

Dessa forma, todas as variáveis relacionadas a instrumentos, materiais socioculturais e/ou pedagógicos em uso na escola para o desenvolvimento de atividades de ensino aprendizagem, foram incluídas, como mostra a tabela III.

Tabela III. MATERIAIS PEDAGÓGICOS EM USO NAS ESCOLAS.

VARIÁVEL	DESCRIÇÃO
IN_MATERIAL_PED_MULTIMIDIA	Instrumentos, materiais socioculturais e/ou pedagógicos em uso na escola para o desenvolvimento de atividades de ensino aprendizagem - Acervo multimídia.
IN_MATERIAL_PED_INFANTIL	Instrumentos, materiais socioculturais e/ou pedagógicos em uso na escola para o desenvolvimento de atividades de ensino aprendizagem - Brinquedos para Educação Infantil.
IN_MATERIAL_PED_CIENTIFICO	Instrumentos, materiais socioculturais e/ou pedagógicos em uso na escola para o desenvolvimento de atividades de ensino aprendizagem - Conjunto de materiais científicos.
IN_MATERIAL_PED_DIFUSAO	Instrumentos, materiais socioculturais e/ou pedagógicos em uso na escola para o desenvolvimento de atividades de ensino aprendizagem - Equipamento para amplificação e difusão de som/áudio.
IN_MATERIAL_PED_MUSICAL	Instrumentos, materiais socioculturais e/ou pedagógicos em uso na escola para o desenvolvimento de atividades de ensino aprendizagem - Instrumentos musicais para conjunto, banda/fanfarra e/ou aulas de música.
IN_MATERIAL_PED_JOGOS	Instrumentos, materiais socioculturais e/ou pedagógicos em uso na escola para o desenvolvimento de atividades de ensino aprendizagem - Jogos Educativos.
IN_MATERIAL_PED_ARTISTICAS	Instrumentos, materiais socioculturais e/ou pedagógicos em uso na escola para o desenvolvimento de atividades de ensino aprendizagem - Materiais para atividades culturais e artísticas.
IN_MATERIAL_PED_DESPORTIVA	Instrumentos, materiais socioculturais e/ou pedagógicos em uso na escola para o desenvolvimento de atividades de ensino aprendizagem - Materiais para prática desportiva e recreação.
IN_MATERIAL_PED_INDIGENA	Instrumentos, materiais socioculturais e/ou pedagógicos em uso na escola para o desenvolvimento de atividades de ensino aprendizagem - Indígena.
IN_MATERIAL_PED_ETNICO	Instrumentos, materiais socioculturais e/ou pedagógicos em uso na escola para o desenvolvimento de atividades de ensino aprendizagem - Materiais pedagógicos para a educação das Relações Étnicas Raciais.
IN_MATERIAL_PED_CAMPO	Instrumentos, materiais socioculturais e/ou pedagógicos em uso na escola para o desenvolvimento de atividades de ensino aprendizagem - Materiais pedagógicos para a educação do campo.

Além das variáveis relacionadas aos materiais pedagógicos existentes nas escolas, todas as variáveis referentes a quantidade de profissionais presentes no ambiente escolar, foram julgadas como variáveis importantes ao projeto, conforme mostra tabela IV.

Tabela IV. PROFISSIONAIS EXISTENTES NAS ESCOLAS.

VARIÁVEL	DESCRIÇÃO
QT_PROF_ADMINISTRATIVOS	Total de profissionais que atuam na escola - Auxiliares de secretaria ou auxiliares administrativos, atendentes.
QT_PROF_SERVICOS_GERAIS	Total de profissionais que atuam na escola - Auxiliar de serviços gerais, porteiro(a), zelador(a), faxineiro(a), horticultor(a), jardineiro(a).
QT_PROF_BIBLIOTECARIO	Total de profissionais que atuam na escola - Bibliotecário(a), auxiliar de biblioteca ou monitor(a) da sala de leitura.
QT_PROF_SAUDE	Total de profissionais que atuam na escola - Bombeiro(a) brigadista, profissionais de assistência a saúde (urgência e emergência), Enfermeiro(a), Técnico(a) de enfermagem e socorrista.
QT_PROF_COORDENADOR	Total de profissionais que atuam na escola - Coordenador(a) de turno/disciplina.
QT_PROF_FONAUDIOLOGO	Total de profissionais que atuam na escola - Fonoaudiólogo(a).
QT_PROF_NUTRICIONISTA	Total de profissionais que atuam na escola - Nutricionista.
QT_PROF_PSIKOLOGO	Total de profissionais que atuam na escola - Psicólogo(a) Escolar.
QT_PROF_ALIMENTACAO	Total de profissionais que atuam na escola - Profissionais de preparação e segurança alimentar, cozinheiro(a), merendeira e auxiliar de cozinha.
QT_PROF_PEDAGOGIA	Total de profissionais que atuam na escola - Profissionais de apoio e supervisão pedagógica: pedagogo(a), coordenador(a) pedagógico(a), orientador(a) educacional, supervisor(a) escolar e coordenador(a) de área de ensino.
QT_PROF_SECRETARIO	Total de profissionais que atuam na escola - Secretário(a) escolar.
QT_PROF_SEGURANCA	Total de profissionais que atuam na escola - Seguranças, guarda ou segurança patrimonial.
QT_PROF_MONITORES	Total de profissionais que atuam na escola - Técnicos(as), monitores(as) ou auxiliares de laboratório(s).

“Para Brandão, a educação acontece com todos os sujeitos, com todos os autores daquele espaço e, todos se tornam educadores. Não tem como dispensar a importância de um Bibliotecário, sendo importante que ele esteja preparado para o atendimento dessas pessoas. Acredito que todos os funcionários precisam estar dialogando dentro de um espaço onde o principal é a criança, o adolescente, aquele que necessita do atendimento [...]. É necessário que todos saibam o papel que exercem como educador porquê todos são, dentro de um espaço escolar.” (Janaína Rocha Santos Pereira)

Também é imprescindível destacar a inclusão da variável IN_ALIMENTACAO pois, para todas as profissionais supracitadas, considerando as escolas públicas — que possuem alunos que necessitam de AEE e são oriundos de famílias de baixa renda, saem de casa sem se alimentar e, conseqüentemente, contam com o fornecimento da alimentação por essas escolas — a alimentação torna-se um recurso que garante que o processo de desenvolvimento de um discente acontecerá com qualidade.

2) *Matrícula*: Foi mostrado aos entrevistados a existência de variáveis responsáveis por informar o tipo de deficiência de um aluno e como este trabalho pretendia utilizar tais variáveis. Dessa forma, conforme tabela V, todos os profissionais supracitados concordaram com a utilização das variáveis com o propósito de informar ao usuário quantos alunos, com aquelas especificidades, estão matriculados na respectiva escola.

Tabela V. ESPECIFICIDADE DE ALUNOS MATRICULADOS NAS ESCOLAS.

VARIÁVEL	DESCRIÇÃO
IN_BAIXA_VISAO	Tipo de deficiência, transtorno do espectro autista ou altas habilidades/superdotação - Baixa visão.
IN_CEGUEIRA	Tipo de deficiência, transtorno do espectro autista ou altas habilidades/superdotação - Cegueira.
IN_DEF_AUDITIVA	Tipo de deficiência, transtorno do espectro autista ou altas habilidades/superdotação (deficiência auditiva).
IN_DEF_FISICA	Tipo de deficiência, transtorno do espectro autista ou altas habilidades/superdotação - Deficiência física.
IN_DEF_INTELECTUAL	Tipo de deficiência, transtorno do espectro autista ou altas habilidades/superdotação - Deficiência intelectual.
IN_SURDEZ	Tipo de deficiência, transtorno do espectro autista ou altas habilidades/superdotação - Surdez.
IN_SURDOCEGUEIRA	Tipo de deficiência, transtorno do espectro autista ou altas habilidades/superdotação - Surdocegueira.
IN_DEF_MULTIPLA	Tipo de deficiência, transtorno do espectro autista ou altas habilidades/superdotação - Deficiência múltipla.
IN_AUTISMO	Tipo de deficiência, transtorno do espectro autista ou altas habilidades/superdotação - Transtorno do Espectro Autista.

3) *Turmas*: Para a tabela de Turma, conforme tabela VI, foram escolhidas as variáveis TP_TIPO_ATENDIMENTO_TURMA e IN_ESPECIAL_EXCLUSIVA. Com elas, será possível informar ao usuário se existem turmas que oferecem Atendimento Educacional Especializado para uma determinada etapa de ensino — creche, pré-escola, ensino fundamental e médio, ensino profissionalizante e EJA (Educação de Jovens e Adultos).

Tabela VI. VARIÁVEIS ESCOLHIDAS PARA SE TRABALHAR COM A TABELA DE TURMAS.

VARIÁVEL	DESCRIÇÃO
TP_TIPO_ATENDIMENTO_TURMA	Tipo de atendimento 1 - Exclusivo Escolarização 2 - Atividade complementar e escolarização 3 - Atividade complementar 4 - Atendimento Educacional Especializado (AEE)
IN_ESPECIAL_EXCLUSIVA	Turma exclusiva de alunos com deficiência, transtorno global do desenvolvimento ou altas habilidades/superdotação (Classes Especiais).

4) *Docentes*: Para todas as profissionais entrevistadas, o quesito *representatividade* foi colocado em questão. Para elas, a representatividade é o melhor exemplo do perfil de uma instituição — se uma instituição não acolhe um profissional com algum tipo de necessidade, ela não terá competência para assimilar algum aluno especial.

Para além desse quesito, a formação do professor também foi considerada — as entrevistadas acreditam que a formação do professor é essencial para o desenvolvimento dos alunos com os quais trabalha. Dessa forma, os profissionais Monitores de atividade complementar, Tradutor Intérprete de Libras, Guia intérprete e Profissional de apoio escolar para alunos com deficiência (Lei 13.146/2015), foram considerados fundamentais para toda e qualquer escola que ofereça um atendimento educacional especializado.

F. Refatoração

Como mencionado na subseção *Preparação*, presente na seção *Solução Desenvolvida*, foi atribuída à tabela Escola uma nova variável chamada *PONTUACAO*, responsável por conter o somatório referente as dependências acessíveis existentes naquela escola, além de outras poucas variáveis que foram julgadas interessantes, antes da consulta aos especialistas. Com a finalização das entrevistas, voltou-se a etapa *Preparação* para que fosse possível melhorar o projeto com a inclusão das variáveis que os profissionais também julgaram ser de suma importância.

1) *Escola*: Além das variáveis inicialmente trabalhadas, conforme tabela II, foram inclusas todas as variáveis presentes nas tabelas das figuras III, IV, V e VI, além da variável *IN_ALIMENTACAO*.

Conforme a nota [71], para os itens quantitativos, o Censo Escolar de 2019 adotou valores máximos aceitáveis considerando o porte da escola. As variáveis quantitativas que apresentaram um valor muito elevado foram marcadas com um código "88888"— o que significava, possivelmente, um erro de digitação ou alguma situação não usual no ato do cadastro da Escola pelo seu respectivo gestor. Variáveis com esse código, neste trabalho, foram substituídas pelo valor 1 (um).

2) *Matrícula, Docente e Turma*: Para o tratamento dessas tabelas utilizou-se o *Colab Research* da Google, haja visto que os arquivos, por serem maiores, necessitavam de uma quantidade maior de *RAM* ao serem trabalhados. Ainda utilizando a plataforma supracitada para a análise dos dados, que disponibiliza apenas 16 *GB* de *RAM*, não foi possível fazer a

exploração dos dados como havia sido feito com as tabelas de Escola e Catálogo de Escolas Detalhadas. Portanto, criou-se um algoritmo capaz de trabalhar essa grande quantidade de dados em “blocos”, através da utilização de *loops*. Para cada bloco, eram filtrados apenas os registros interessantes ao projeto e, uma vez os registros de interesse obtidos, alocados em um arquivo de destino. Ao final, foi feita a junção desses arquivos, como mostra, respectivamente, as Listagens 3 e 4.

```

1 import pandas as pd
2
3 chunk_size = 2000000
4 batch_no = 1
5 for chunk in pd.read_csv('/content/drive/My Drive/
6   MATRICULA_NORDESTE.CSV', chunksize=chunk_size):
7     chunk.to_csv('matriculas' + str(batch_no) + '.
8     csv', index=False)
9     batch_no+=1

```

Listing 3. Algoritmo responsável por separar grandes quantidades de dados em blocos.

```

1 import pandas as pd
2 from glob import glob
3
4 stock_files = sorted(glob('/content+matriculas_aee*.
5   csv'))
6 alunos_aee_salvador = pd.concat(pd.read_csv(file,
7   encoding = "ISO-8859-1", sep = '|', low_memory =
8   False).assign(filename=file)
9   for file in stock_files), ignore_index = True)

```

Listing 4. Algoritmo responsável pela junção dos arquivos com os dados já filtrados.

Para a tabela de Matrícula, foram filtrados apenas aqueles registros relacionados a alunos do município de Salvador — *CO_MUNICIPIO_END* = 2927408.0 — e que possuíam necessidade especial — *IN_NECESSIDADE_ESPECIAL* = 1. As variáveis irrelevantes ao projeto foram excluídas, ficando apenas as variáveis presentes na tabela V — tipo específico da condição de um determinado aluno.

Para a tabela de Turma, foram filtrados apenas aqueles registros que estavam relacionados às escolas soteropolitanas — *CO_MUNICIPIO* = 2927408, além das variáveis irrelevantes ao projeto que foram excluídas.

Para a tabela de Docente, além de serem filtrados aqueles relacionados as escolas soteropolitanas, também foi atribuída uma nova variável à tabela. Tal variável diz respeito a pontuação obtida por um professor, onde as variáveis que fazem parte dessa pontuação também foram debatidas com os entrevistados e são mostradas na tabela VII.

Tabela VII. VARIÁVEIS PRESENTES NO SOMATÓRIO DA VARIÁVEL "PONTUAÇÃO" PARA A TABELA DE DOCENTE.

VARIÁVEL	DESCRIÇÃO
IN_COMPLEMENTACAO_PEDAGOGICA	Alguma formação/complementação pedagógica
IN_ESPECIFICO_ED_ESPECIAL	Outros cursos - Formação Continuada com no mínimo 80 horas - Específico para educação especial
TP_TIPO_DOCENTE	1 - Docente 2 - Auxiliar/Assistente Educacional 3 - Profissional/Monitor de atividade complementar 4 - Tradutor Intérprete de Libras 5 - Docente Titular - coordenador de tutoria (de módulo ou disciplina) - EAD 6 - Docente Tutor - Auxiliar (de módulo ou disciplina) - EAD 7 - Guia intérprete 8 - Profissional de apoio escolar para alunos com deficiência (Lei 13.146/2015)
TP_TIPO_ATENDIMENTO_TURMA	1 - Exclusivo Escolarização 2 - Atividade complementar e escolarização 3 - Atividade complementar 4 - Atendimento Educacional Especializado (AEE)
IN_ESPECIAL_EXCLUSIVA	Profissional Escolar de turma exclusiva de alunos com deficiência, transtorno global do desenvolvimento ou altas habilidades/superdotação (Classes Especiais)

G. Visualização

Para a visualização dos dados deste projeto, decidiu-se utilizar os dados das escolas trabalhadas com os dados dos transportes coletivos de Salvador, plotados em um mapa. Essa ideia inicial se deu com o objetivo de, no futuro, podermos colher os dados de geolocalização dos usuários e mostrar, também, as escolas mais próximas a ele, plotadas ou não em um mapa. Dessa forma, foi preciso obter os dados da malha territorial do município de Salvador, fornecido pelo *IBGE* e os dados de transportes coletivos do mesmo município.

Inicialmente, como já foi supracitado, foi essencial obter a malha territorial da capital baiana. Uma vez esses dados obtidos, salvou-se apenas o *shapefile* — representação do município de Salvador em formato de mapa — que representava o município de Salvador. Os próximos passos consistiram na plotagem dos dados — Escola, Metrô, Paradas de Ônibus — que foram anteriormente colhidos, no mapa de Salvador. Conforme mostra a figura 5, os pontos em vermelho são as escolas, o azul as estações de metrô, e os de cor laranja são os pontos de ônibus.

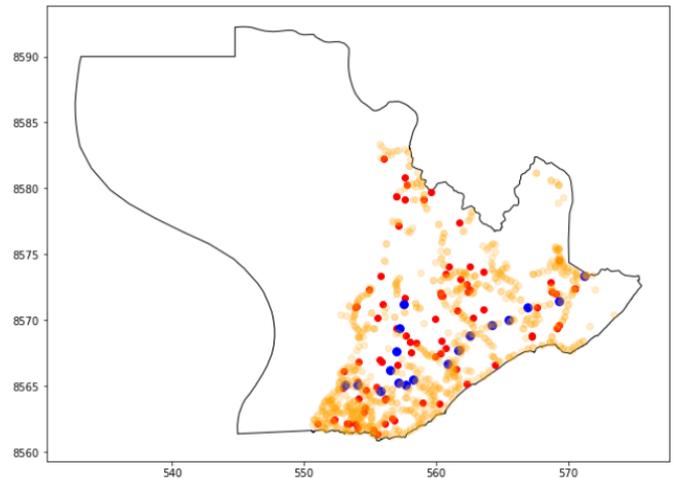


Figura 5. Shapefile correspondente ao município de Salvador com as escolas, pontos de ônibus e estações de metrô plotadas.

Por último, com a utilização do *Folium*, houve a plotagem dos dados em um mapa interativo, conforme figura 6, onde cada dado apresenta um ícone diferente, para que sua identificação seja mais fácil.

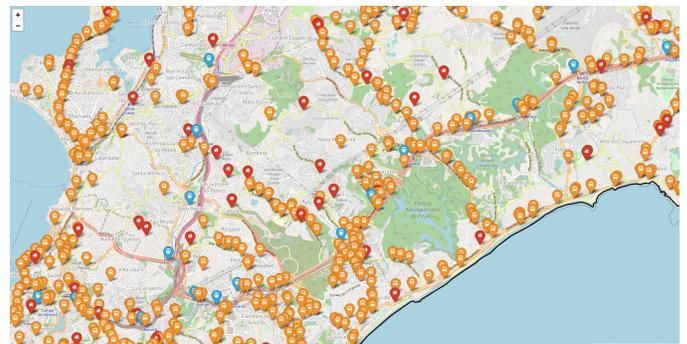


Figura 6. Impressão do mapa interativo com os dados das escolas, estações e paradas de ônibus.

H. Operacionalização

Na última fase do processo de desenvolvimento do SELECTOR, foi criado um mínimo produto viável para apresentar as escolas e ordená-las segundo as entradas do usuário. O mínimo produto viável consiste em uma *Restful API* criada com a linguagem de programação *Java* e o *framework Spring Boot*. Abaixo, as informações referentes as tecnologias utilizadas e a funcionalidade do sistema será melhor detalhada.

1) *Tecnologias Utilizadas*: Inicialmente o SELECTOR foi pensado para ser um serviço web *RESTful* — solução utilizada para integração de sistemas e comunicação entre aplicações diferentes. Esse serviço web segue o estilo arquitetural *REST (Representational State Transfer)*, que define uma série de princípios para a construção de *Web Services* baseado no protocolo HTTP. Os sistemas que fornecem uma API definida pela aplicação desse estilo arquitetural recebem o nome de **RESTful** [72].

Feitas as considerações sobre a arquitetura do sistema,

escolheram-se as tecnologias que seriam utilizadas na construção da mesma.

- *Java* — Linguagem de programação Multiplataforma com mais de 20 anos de existência e uma das mais utilizadas no mundo, com vasto conjunto de APIs para as mais diversas aplicações.
- *Spring Boot* — Estrutura *open source* baseada em Java usada para criar um microserviço. Foi desenvolvido pela equipe da Pivotal sendo usado para construir aplicações java de forma fácil e com o mínimo de configuração possível.
- *Apache Maven* — Apache Maven é uma ferramenta de gerenciamento de projetos de *software* e compreensão. Baseado no conceito de um modelo de objeto de projeto (POM), Maven pode gerenciar construção de um projeto, elaboração de relatórios e documentação de uma peça central de informações.
- *H2 database* — Banco de dados Java, leve e de código aberto. Ele pode ser integrado a aplicativos desenvolvidos em Java ou executado no modo cliente-servidor. Uma de suas maiores características é a possibilidade de poder ser configurado para ser executado como um banco de dados na memória, o que significa que os dados não persistirão no disco. Dessa forma, desenvolvedores podem se concentrar, inicialmente, na lógica de negócio do sistema sem se preocupar com uma infraestrutura a ser montada para utilização de bancos mais convencionais.

2) *Funcionalidade Principal*: Inicialmente, o sistema conta com uma única funcionalidade — ordenar as escolas que adquiriram uma maior pontuação conforme a entrada do usuário. Para que o sistema possa apresentar as escolas para uma determinada pessoa, são solicitadas três informações:

- **Dependência da Escola** — o usuário informará qual tipo de escola está buscando de acordo com sua dependência — público ou privada.
- **Mediação Didático Pedagógica** — tipos de relações e interações realizadas no processo de ensino/aprendizagem, estimuladas ou não pelo professor, subdividindo-se em três tipos: Presencial, Semipresencial e Educação a Distância.
- **Etapa de Ensino** — A etapa de ensino corresponde a série em que uma pessoa será matriculada. O projeto abarca as seguintes etapas: Creche, Pré-escola, Ensino Fundamental, Ensino Médio, Profissionalizante e Educação de Jovens e Adultos.

Através dessas informações, e das variáveis que os entrevistadores consideraram ser de suma importância na contabilização dos pontos, cada escola adquire uma pontuação total — soma da pontuação geral da infraestrutura da escola, do quadro docente e da pontuação das turmas existentes naquela escola. A lista das escolas, já filtradas com base no desejo do usuário, é ordenada através de sua pontuação geral e apresentada de forma decrescente, conforme algoritmo apresentado na Listagem 5.

```
1 return new EscolasResponseDto(escolasDtoResponse
2     .stream()
3     .sorted(Comparator.comparing(
4         EscolasDto::
5         getTotalPontuacao
6         .reversed()))
7     .collect(Collectors.toList()));
```

Listing 5. Algoritmo responsável pela ordenação das escola se sua respectiva apresentação.

Na figura 7, temos um exemplo dos dados que usuário precisará informar para buscar as escolas de seu interesse. Nela o usuário solicita buscar escolas públicas para o ensino fundamental na modalidade presencial. Como resposta, obtemos uma lista das escolas com as maiores pontuações.

Escola	Endereço	Telefone	Docentes	Detalhes
AEE - CENTRO ESTADUAL DE EDUCACAO PESTALOZZI DA BAHIA	Av. Adhemar de Barros, s/n - Ondina, Salvador - BA, 40100-170	+557132470049	256	Detalhes
EE - CENTRO DE EDUCACAO ESPECIAL DA BAHIA - CEEBA	R. RAIMUNDO PEREIRA MAGALHAES, 00045 ONDINA, 40170-155 Salvador - BA	(71) 33321798	59	Detalhes
EE - CENTRO DE CAPACITACAO DE PROFISSIONAIS DA EDUCACAO - WILSON LINS-CAS BA	R. RUA RAIMUNDO PEREIRA MAGALHAES - ONDINA, 40170430	+55713332-9222	31	Detalhes

Figura 7. Impressão da tela do sistema com os dados de entrada para escolas públicas, ensino fundamental na modalidade presencial.

VI. CONCLUSÃO

Embora o número de matrículas de pessoas com deficiências aumente a cada ano no Brasil, responsáveis diretos por pessoas com deficiências ainda encontram dificuldades na procura, matrícula e manutenção das mesmas em instituições de ensino, que atendam as obrigações e deveres propostos pelas Diretrizes Operacionais da Educação Especial para o Atendimento Educacional Especializado — AEE.

Ainda que existam sistemas que trabalham com os dados da educação brasileira no Brasil, não foi possível identificar alguma solução que atue de forma mais direta com as escolas que fornecem atendimento para pessoas que necessitam de Atendimento Educacional Especializado, da mesma forma que este projeto se propõe trabalhar. Dos projetos citados da seção **Trabalhos Correlatos**, apenas os projetos *Qedu* e *Raio x das Escolas do Brasil* trabalham de forma mais direta com a apresentação das escolas e suas características para a sociedade, mas esses projetos possuem limitações quando o assunto é acessibilidade.

Esse trabalho surge nesse contexto como uma iniciativa necessária que visa auxiliar na tomada de decisão dos pais e responsáveis diretos, por pessoas que necessitam de Atendimento Educacional Especializado, acerca das escolas, possivelmente, melhor preparadas para receber uma pessoa com tais condições. O projeto não se limitou apenas a colher informações sobre a infraestrutura das escolas no Brasil, mas preocupou-se, também, em informar se aquela escola possui um corpo docente capacitado, classes exclusivas para pessoas que necessitam de AEE, salas com recursos multifuncionais para esse tipo de atendimento, existência ou não de profissionais altamente requisitados que fazem parte do processo

de aprendizagem dessas pessoas — psicólogos, nutricionistas, pedagogos, etc. Além disso, os dados foram integrados a outros conjuntos de dados, permitindo considerar outras dimensões, como a geolocalização das escolas e do sistema de transporte público que atende as mesmas.

Acredita-se assim que o maior objetivo da Lei de Acesso à Informação, nesse empreendimento, é atingido, visto que foi possível pensar em um projeto que futuramente poderá ser capaz de beneficiar uma parcela da sociedade, aquela que é uma das maiores contribuintes da educação pública neste país e deverá sempre se beneficiar dos projetos inovadores que nessas instituições foram concebidas.

Com base nisso, foi construído um produto minimamente viável onde o usuário poderá inserir três importantes informações acerca da escola que procura: **Dependência da Escola** — público ou privada; **Mediação Didático Pedagógica** — presencial, semipresencial ou EAD — e **Etapa de Ensino** — etapa de ensino correspondente a série em que o aluno será matriculado. ex: creche, primário, fundamental, etc. Com base nessas informações, uma lista de escolas que atendam a tal necessidade são mostradas em ordem decrescente de pontuação. Antes da construção dessa API, foi necessária a realização da extração, limpeza e preparação de dados abertos referentes a educação básica no Brasil, para em seguida, realizar-se uma consulta a profissionais que atuam direta e indiretamente com o tema.

Com o desenvolvimento do presente trabalho, foi possível perceber que a maioria das escolas soteropolitanas não estão aptas para ofertar um Atendimento Educacional Especializado. Do total de 1530 escolas, apenas 96 responderam estarem aptas para tal. Outro fato observado com a implementação deste trabalho foi que, apesar de valores ausentes, o Censo Escolar, para as necessidades desse trabalho, mostrou-se altamente satisfatório, além de ser uma ferramenta capaz de transformar a educação brasileira com a riqueza de informações presentes em seus microdados.

VII. TRABALHOS FUTUROS

Durante o desenvolvimento da solução proposta, observou-se possibilidades de melhorias na implementação visando evoluir o SELECTOR. Nesse contexto, destacam-se como trabalhos futuros:

- Construção da aplicação mobile de forma que seja acessível para pessoas com baixa ou nenhuma visão.
- Possibilidade de apresentar as escolas a partir da geolocalização do usuário.
- Inclusão de todas as escolas do Estado da Bahia.
- Disponibilização da funcionalidade onde o usuário possa filtrar escolas com base em uma necessidade especial específica.

REFERÊNCIAS

- [1] “Diretrizes operacionais da educação especial para o atendimento educacional especializado na educação básica,” encurtador.com.br/cy158, 2020 (acessado em Outubro 17, 2020).
- [2] “Manual de acessibilidade espacial para escolas: O direito à escola acessível!” encurtador.com.br/guERU, 2020 (acessado em Outubro 17, 2020).
- [3] C. V. Cavalcante, “Atendimento educacional especializado: uma nova proposta de educação inclusiva,” *Pontifícia Universidade Católica de Goiás*, 2011.
- [4] “O que é o censo escolar,” <http://portal.inep.gov.br/censo-escolar>, 2020 (acessado em Outubro 17, 2020).
- [5] “Raio-x das escolas do brasil,” <https://tinyurl.com/v9vzurs>, 2020 (acessado em Fevereiro 7, 2020).
- [6] “Qdu - use dados e transforme a educação,” <https://www.qedu.org.br/>, 2020 (acessado em Fevereiro 7, 2020).
- [7] R. K. SASSAKI, *Inclusão: construindo uma sociedade para todos, Terceira edição*. WVA, 1999.
- [8] “Open knowledge foundation,” <https://okfn.org/>, 2020 (acessado em Fevereiro 8, 2020).
- [9] “David eaves,” <https://www.hks.harvard.edu/faculty/david-eaves>, 2020 (acessado em Fevereiro 8, 2020).
- [10] A. POSSAMAI JÚLIA, “Dados abertos no governo federal brasileiro: desafios de transparência e interoperabilidade,” pp. 67–70, 72–77, 2016.
- [11] “Sobre o dados.gov.br,” <http://dados.gov.br/pagina/sobre>, 2017 (acessado em Novembro 2, 2017).
- [12] “Conformant licenses,” <http://opendefinition.org/licenses/>, 2017 (acessado em Novembro 2, 2017).
- [13] “Conformant licenses the three laws of open government data,” <https://eaves.ca/2009/09/30/three-law-of-open-government-data/>, 2017 (acessado em Novembro 2, 2017).
- [14] “O que são dados abertos,” <http://dados.gov.br/pagina/dados-abertos>, 2017 (acessado em Novembro 2, 2017).
- [15] P. da República, “Lei de acesso à informação,” http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/112527.htm, 2020 (acessado em Outubro 21, 2020).
- [16] T. de Contas da União, “Cinco motivos para a abertura de dados na administração pública,” <https://portal.tcu.gov.br/biblioteca-digital/cinco-motivos-para-a-abertura-de-dados-na-administracao-publica.htm>, 2020 (acessado em Outubro 21, 2020).
- [17] O. K. Brasil, “Operação serenata de amor,” <https://serenata.ai/>, 2020 (acessado em Outubro 21, 2020).
- [18] “Mobilab,” <http://mobilab.prefeitura.sp.gov.br/>, 2021 (acessado em Fevereiro 23, 2021).
- [19] “The growth in connected iot devices is expected to generate 79.4zb of data in 2025, according to a new idc forecast,” encurtador.com.br/fhFT1, 2020 (acessado em Outubro 7, 2020).
- [20] “Data never sleeps 8.0,” <https://www.domo.com/learn/data-never-sleeps-8>, 2020 (acessado em Outubro 14, 2020).
- [21] “strata.oreilly.com/2010/01/roger-magoulas-on-big-data.html,” <http://radar.oreilly.com/2010/01/roger-magoulas-on-big-data.html>, 2020 (acessado em Outubro 17, 2020).
- [22] R. Elmasri and S. Navathe, *Fundamentals of Database Systems, Fifth Edition*. Pearson ‘I&’ Addison Wesley, 2006.
- [23] E. E. Services, *Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data*. John Wiley ‘I&’ Sons, Inc., 2015.
- [24] “The evolution of big data as a research and scientific topic: Overview of the literature,” <https://www.researchtrends.com/issue-30-september-2012/the-evolution-of-big-data-as-a-research-and-scientific-topic-overview-of-the-literature/>, 2020 (acessado em Outubro 7, 2020).
- [25] E. E. Services, *Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data*. John Wiley ‘I&’ Sons, Inc., 2015.
- [26] F. Cady, *The Data Science Handbook*. Wiley, 2017.
- [27] J. W. Foreman, *Data Smart: Using Data Science to Transform Information into Insight*. Wiley, 2013.
- [28] “Data science for social good,” <https://www.turing.ac.uk/collaborate-turing/data-science-social-good>, 2020 (acessado em Outubro 14, 2020).
- [29] “O que são tecnologias sociais?” <http://tecnologiasocial.sites.uff.br/o-que-sao-tecnologias-sociais/>, 2020 (acessado em Outubro 11, 2020).
- [30] R. Dagnino, *Tecnologia Social. Ferramenta Para Construir Outra Sociedade*. Komedi, 2010.

- [31] J. W. Turkey, "The future of data analysis," *Princeton University and Bell Telephone Laboratories*, p. 2, 1962.
- [32] "pandas 1.0 documentation," <https://pandas.pydata.org/>, 2020 (acessado em Fevereiro 6, 2020).
- [33] "Python 3.6.3 documentation," <https://docs.python.org/3/>, 2017 (acessado em Novembro 2, 2017).
- [34] J. Densmore, *Data Pipelines Pocket Reference: Moving and Processing Data for Analytics*. O'Reilly Media, Inc., 2021.
- [35] "5 common data integration challenges and how to solve them," <https://blog.hubspot.com/website/5-common-data-integration-challenges-and-how-to-solve-them>, 2020 (acessado em Outubro 05, 2020).
- [36] "biggest data integration challenges," <https://www.cloverdx.com/blog/biggest-data-integration-challenges>, 2020 (acessado em Outubro 05, 2020).
- [37] F. Naumann and J. Bleiholder, "Data fusion," *ACM Computing Surveys*, pp. 2–3, 2009.
- [38] "Cloud data fusion documentatation," <https://cloud.google.com/data-fusion/docs/concepts/overview>, 2020 (acessado em Outubro 05, 2020).
- [39] M. Friendly, "A brief history of data visualization," in *Handbook of Computational Statistics: Data Visualization*, C. Chen, W. Härdle, and A. Unwin, Eds. Heidelberg: Springer-Verlag, 2006, vol. III.
- [40] E. R. Tufte, *The Visual Display of Quantitative Information, Second Edition*. Graphic Press LLC, 2007.
- [41] S. Few, "Eenie, meenie, minie, moe: Selecting the right graph for your message," *Perceptual Edge*, p. 5, 2004.
- [42] "Ica," <https://icaci.org/mission/>, 2020 (acessado em Outubro 17, 2020).
- [43] A. MacEachren and J. Kraak, "Research challenges in geovisualization. cartography geograph." *Inf. Sci.*, vol. 28, pp. 3–12, 01 2001.
- [44] M. Nöllenburg, "Geographic visualization," in *Human-Centered Visualization Environments: GI-Dagstuhl Research Seminar, Dagstuhl Castle, Germany, March 5-8, 2006, Revised Lectures*, 01 2006, pp. 257–294.
- [45] M.-J. Kraak, "Geovisualization illustrated," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 57, pp. 390–399, 04 2003.
- [46] "shapley," <https://shapely.readthedocs.io/en/latest/>, 2020 (acessado em Fevereiro 8, 2020).
- [47] "fiona," <https://fiona.readthedocs.io/en/latest/>, 2020 (acessado em Fevereiro 8, 2020).
- [48] "matplotlib," <https://matplotlib.org/>, 2020 (acessado em Fevereiro 8, 2020).
- [49] " descartes," <https://pypi.org/project/descartes/>, 2020 (acessado em Fevereiro 8, 2020).
- [50] "Esri," <https://www.esri.com/en-us/home>, 2020 (acessado em Outubro 17, 2020).
- [51] "Wgs84," <https://gisgeography.com/wgs84-world-geodetic-system/>, 2020 (acessado em Outubro 17, 2020).
- [52] H. Moritz, "Geodetic reference system 1980," *Journal of Geodesy* 74, p. 128–133, 2000.
- [53] V. SCHOAB, C. C. G. FREITAS, and L. F. LARA, "A universidade e a tecnologia social: análise da aderência," *Revista Espacios*, vol. 35, 2014.
- [54] L. d. S. Dias, "O papel da universidade no desenvolvimento de tecnologias sociais: um estudo de caso na ufpe," *Universidade Federal de Pernambuco*, 2016.
- [55] A. S. ALMEIDA, "A contribuição da extensão universitária para o desenvolvimento de tecnologias sociais. in: tecnologia social para o desenvolvimento sustentável." *RTS*, 2010.
- [56] "Acesso à educação com foco na pessoa com perda auditiva," <https://analise-do-acesso-a-educacao.webnode.com/>, 2020 (acessado em Fevereiro 7, 2020).
- [57] "What is rest," <https://restfulapi.net/>, 2020 (acessado em Outubro 14, 2020).
- [58] "Java® platform, standard edition version 11 api specification," <https://cr.openjdk.java.net/~iris/se/11/latestSpec/api/index.html>, 2020 (acessado em Outubro 14, 2020).
- [59] "Spring boot reference documentation," <https://docs.spring.io/spring-boot/docs/current/reference/htmlsingle/>, 2020 (acessado em Outubro 14, 2020).
- [60] "Sgbd," <https://www.cin.ufpe.br/fjclf/gdi/aulas/02-FuncionalidadesSGBD.pdf>, 2020 (acessado em Outubro 14, 2020).
- [61] "Microdados da educação do censo 2018," <http://portal.inep.gov.br/censo-escolar>, 2020 (acessado em Fevereiro 7, 2020).
- [62] "Catálogo de escolas do inep," encurtador.com.br/ejsxK, 2020 (acessado em Fevereiro 7, 2020).
- [63] "Malha territorial do município de salvador," <https://tinyurl.com/qvw2jez>, 2020 (acessado em Fevereiro 7, 2020).
- [64] "Ccr metrô bahia," <http://www.ccrmetrobahia.com.br/>, 2020 (acessado em Fevereiro 7, 2020).
- [65] "Colab research," <https://colab.research.google.com/>, 2020 (acessado em Outubro 14, 2020).
- [66] "Instruções para utilização dos microdados do censo da educação básica 2019," http://download.inep.gov.br/microdados/microdados_educacao_basica_2019.zip, 2020 (acessado em Dezembro 05, 2020).
- [67] "Sinopses estatísticas da educação básica," <http://inep.gov.br/web/guest/sinopses-estatisticas-da-educacao-basica>, 2020 (acessado em Outubro 14, 2020).
- [68] "Carolina salenave cambeses - educadora há 15 anos, técnica de educação especial pela secretaria de educação do estado da bahia e consultora em educação especial no sesi - sistema fieb bahia," www.google.com.br.
- [69] "Janafina rocha santos pereira - professora e graduanda em pedagogia pela universidade federal da bahia," www.google.com.br.
- [70] "Mariana gonçalves silva santos - professora e pedagoga pela universidade católica de salvador," www.google.com.br.
- [71] "Nota sobre o tratamento aplicado em 2019 aos campos de quantitativo e também aos campos não aplicáveis e não informados nas bases do censo escolar," http://download.inep.gov.br/microdados/microdados_educacao_basica_2019.zip, 2020 (acessado em Dezembro 05, 2020).
- [72] "Rest," https://www.ics.uci.edu/fielding/pubs/dissertation/rest_arch_style.htm, 2020 (acessado em Outubro 17, 2020).