UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
CURSO DE ENGENHARIA DE COMPUTAÇÃO

ARTHUR RIBEIRO

Uma Ferramenta Web para Análise de dados estatísticos do Ensino Superior da Área de Computação utilizando Elasticsearch

Monografia apresentada como requisito parcial para a obtenção do grau de Bacharel em Engenharia da Computação

Orientadora: Profa. Dra. Renata Galante

Porto Alegre
2022

AGRADECIMENTOS

Dedico este trabalho primeiramente aos meus pais Amauri e Maira, que aonde estiverem, estão orgulhosos do filho que criaram.

Agradeço a minha orientadora, por toda orientação e apoio para a finalização deste ciclo e a banca examinadora pela disponibilidade e paciência.

Agradeço a todos meus familiares, irmãos, tias e primos, que me deram todo o suporte ao longo de minha vida me fornecendo todo incentivo, apoio, amor e carinho.

Por fim, porém não menos importante, dedico este trabalho a minha esposa, Ana Caroline Pedroso Ribeiro, que foi de suma importância, me incentivando, apoiando e ajudando na execução desta monografia.

RESUMO

A Educação é um dos pilares mais importantes para a sociedade. Levando isso em consideração, este trabalho apresenta uma análise dos dados estatísticos do Ensino Superior da área de Computação, com os dados disponíveis no site do MEC atualizados até 2020. A Sociedade Brasileira de Computação realiza um papel fundamental para a comunidade da computação, fomentando o acesso à informação e cultura através da informática. Para tal, executa diversas tarefas, sendo uma delas, a análise estatística dos dados dos cursos de tecnologia. Desde 2010 a SBC disponibiliza anualmente um relatório com algumas estatísticas sobre os cursos de computação de nível superior, entretanto, esse relatório é elaborado manualmente pela diretoria de educação da SBC, e este trabalho tem como objetivo principal, projetar e implementar uma automatização desse relatório anual, com a utilização da ferramenta web Elasticsearch juntamente com o Kibana, para análise de dados educacionais estatísticos com o intuito de contribuir com a Sociedade Brasileira de Computação.

Palavras-chave: Computação. Visualização de dados. Ferramenta Web.

A Web Tool for Analysis of Statistics data From Technology Courses of University Education Using Elasticsearch

ABSTRACT

ducation is one of the most important pillars for society. Taking this into account, this work presents an analysis of the statistical data of Higher Education in the area of Computing, with the data available on the MEC website updated until 2020. The Brazilian Computer Society plays a fundamental role for the computing community, promoting the access to information and culture through informatics. To this end, it performs several tasks, one of which is the statistical analysis of data from technology courses. Since 2010, SBC has made available annually a report with some statistics on higher level computing courses, however, this report is manually prepared by the SBC education board, and this work has as main objective, to design and implement an automation of this annual report. , using the Elasticsearch web tool together with Kibana, to analyze statistical educational data in order to contribute with the Brazilian Computer Society.

Keywords: Computing; Data Visualization. Web Tool.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

API	*Application Programming Interface*
ASF	*Apache Software Foundation*
CSV	*Comma-Separated-Values*
DHTs	*Distributed Hash Tables*
ENEM	Exame Nacional do Ensino Médio
HTTPS	*Hyper Text Transfer Protocol Secure*
IES	Instituição de Ensino Superior
INEP	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
JSON	*JavaScript Object Notation*
LSM - tree	*Log-Structured Merge-trees*
MEC	Ministério da Educação
REST	*Representational State Transfer*

SUMÁRIO

1 INTRODUÇÃO

A pandemia do COVID-19 acelerou e contribuiu para a expansão no mercado dos profissionais da Tecnologia da Informação. Com isso, a procura de profissionais na área cresceu aproximadamente 671 por cento, conforme dados encontrados no site da (CNN, 2021). Como consequência dessa procura por profissionais na área, a procura pelos cursos de ensino superior da área também teve uma aumenta significativa.

De acordo com a (SBC, 2022), que tem como função fomentar o acesso à informação e cultura por meio da informática, promover a inclusão digital, incentivar a pesquisa e o ensino em computação no Brasil, e contribuir para a formação do profissional da computação com responsabilidade social, é de muita importância para esta Sociedade, analisar como está a situação dos cursos da área de tecnologia e qual foi o impacto da pandemia para os referidos cursos.

A Sociedade Brasileira de Computação conta com uma diretoria específica para os assuntos relacionados a educação. A diretoria de educação da SBC é responsável por analisar os cursos de ensino superior da área de tecnologia do Brasil. Uma das maneiras com que ela analisa os cursos é através dos dados obtidos do MEC. Esses dados são analisados manualmente pela diretoria de educação, organizados e disponibilizados para a comunidade através de um relatório anual. Esses relatórios[1] expõem dados estatísticos de como os cursos de tecnologia evoluem ao logo do tempo.

Este trabalho tem como objetivo principal projetar e implementar uma ferramenta para análise de dados educacionais estatísticos com o intuito de contribuir com a Sociedade Brasileira de Computação na automatização do relatório anual disponibilizado para a comunidade em geral. A ferramenta contempla a análise dos dados extraídos de fontes oficiais e com números estatísticos, de como está a procura dos cursos de tecnologia, como os cursos estão distribuídos entre as instituições públicas e particulares, quais as modalidades que estão sendo mais procuradas, entre outras análises, este estudo será desenvolvido. Este trabalho apresenta o levantamento de requisitos, a proposta e implementação da ferramenta e um conjunto de suposições e análises sobre os dados coletados.

O restante do texto está organizado da seguinte forma. O Capítulo 2 descreve o contexto da SBC, microdados educacionais e as ferramentas utilizadas para o desenvolvimento da ferramenta proposta. O Capítulo 3 descreve o levantamento de

[1] https://www.sbc.org.br/documentos-da-sbc/category/133-estatisticas

requisitos enquanto o Capítulo 4 descreve em detalhes o projeto e a implementação da ferramenta. O Capítulo 5 apresenta de forma ilustrada e analítica todos os gráficos e análises realizadas sobre os microdados do MEC para a Computação e áreas afins. Por fim, o Capítulo 6 finaliza o trabalho com as Conclusões e apontamentos para trabalhos futuros.

2 CONCEITOS E TECNOLOGIAS UTILIZADAS

Neste capítulo, descreveremos as tecnologias, arquiteturas, técnicas e conceitos utilizados nas ferramentas utilizadas para o desenvolvimento do trabalho.

2.1 Elasticsearch

O Elasticsearch é uma ferramenta de busca e um banco de dados distribuído que armazena arquivos no documento de formato *JavaScript Object Notation* JSON, onde estes, são projetados especificamente para pesquisa e análise de dados semiestruturados. De um modo geral, o Elasticsearch gera uma camada de serviços acima da biblioteca de busca e indexação de texto do Apache Lucene, tornando-se desta forma, um intermédio entre a aplicação que faz acesso aos dados armazenados e a *Application Programming Interface* API de busca e indexação. Além disso, provê uma API RESTful para consulta e armazenamento de documentos (KONONENKO et al., 2014).

De acordo com (ABUBAKAR et al., 2014), o Elasticsearch, quando comparado com outros bancos de dados NoSQL apresenta melhor desempenho em operações de leitura do que os demais, o que vai de encontro com as intenções deste projeto. Como os dados inseridos não precisam ser alterados ou atualizados, o Elasticsearch mostrou-se o melhor candidato por apresentar bons desempenhos de inserção e leitura.

2.2 Kibana

O Kibana é um sistema de visualização baseado na web que é integrada com o Elasticsearch para fornecer de maneiras fáceis a navegação e a visualização de dados, usando uma variedade de gráficos e tabelas. Nessa interface é possível realizar consultas aos dados e refinar os resultados graficamente por meio de opções ou de uma sintaxe definida pela ferramenta.

A principal utilidade do Kibana é a de fornecer um conjunto básico de tipos de visualizações que podem ser gráficos mais simples como os de setores, de barras, de linhas, etc. ou mais sofisticados como mapas de calor, mapas de regiões geográficas, tabelas de dados, etc., e a possibilidade de agregá-los a uma dashboard que possui um objetivo. (ELASTIC, 2022)

2.3 API REST

Uma Application Programming Interface (API) é um conjunto de especificações que permite que distintos programas se comuniquem. Ele descreve a maneira apropriada de um desenvolvedor de software desenvolver um programa em um servidor que se comunica com vários aplicativos clientes.

De acordo com (ASTERA, 2020), a integração de API refere-se a um par de aplicativos (dois ou mais) interconectados por meio de suas APIs para trocar dados e realizar uma função conjunta, permitindo assim a interação entre os aplicativos.

A abstração chave da informação em uma API REST é um "recurso útil". Quaisquer informações que possam ser nomeadas podem ser um recurso útil: um relatório ou imagem, um serviço temporal, um conjunto de diferentes posses, um item não-virtual (uma pessoa), e assim por diante. REST utiliza um identificador único para os recursos, que por sua vez também possuem relacionamentos com diferentes outros recursos. Ao utilizar uma API REST, temos um número fixo de ações (ou verbos) para realizar nesses recursos (comumente chamado de CRUD - Create, Retrieve, Update, Delete - Criar, Recuperar, Atualizar, Remover)(SUJAN et al., 2020)

2.4 NoSQL

Banco de dados NoSQL (*Not only SQL*), são bancos de dados do tipo não relacionais que permitem um alta escalabilidade bem como uma alta disponibilidade. Segundo (DAVOUDIAN, 2019) bancos não relacionais possuem algumas características como: modelo de dados mais flexível; flexibilização nas propriedades ACID das transações; aglutinação dos dados; uso de índices distribuídos para acesso dos dados e armazenamento; dados facilmente replicáveis e particionáveis horizontalmente em servidores locais e remotos; acesso amigável à web por meio de uma interface de cliente simples ou protocolo para consulta dados;

Existem mais de 200 tipos de modelos diferentes bancos de dados NoSQL, contudo o modelo mais simples e mais popular e o modelo de chave & valor. Neste modelo os dados são representados por um par chave, valor e são armazenados em estruturas estruturas altamente escaláveis como tabales hash distribuidas (DHTs) e arvores LSM (LSM-tree). A chave podem ser um valor simples (ex. hash, URI, nome de arquivo) ou estruturado (ex. chaves compostas Oracle), elas podem ser geradas pelo

sistema ou definidas pelo aplicativo. O valor representa o dado com um tipo arbitrário, estrutura e tamanho (ex. imagem, string, documento), ele é identificado pela chave que é única e indexada. O valor é codificado em um *array* de bytes onde a codificação e decodificação é responsabilidade da aplicação cliente. O elasticsearch como ja foi mencionado anteriormente usa banco NoSQL, com o modelo de chave valor. O que permite ao mesmo a alta escalabilidade horizontal bem com alta disponibilidade e performance.

3 LEVANTAMENTO DE REQUISITOS

Neste capítulo são descritos os requisitos levados em consideração para o desenvolvimento teórico e prático da ferramenta proposta neste trabalho. Primeiro, apresenta-se um breve histórico do MEC e os dados educacionais que são disponibilizados no seu site. Em seguida, apresenta-se a SBC, sua Diretoria de Educação e Relatório Educacional que é atualmente feito de forma manual, identificando assim os requisitos da aplicação proposta.

3.1 MEC e Dados Educacionais

O Ministério da Educação (MEC) disponibiliza dados anuais da educação no Brasil, e esses dados são obtidos pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). O INEP realiza anualmente o censo escolar e o censo da educação superior do Brasil, com a finalidade de obter os dados estatísticos do ensino básico e superior. São apurados os números de instituições, números de alunos entre os diversos cursos, entre outros dados. Este levantamento estatístico será o foco deste trabalho, que serão abordados e analisados posteriormente.

No contexto deste trabalho, utilizaremos os dados referentes ao ensino superior no Brasil.

3.2 Histórico da SBC

A Sociedade Brasileira de Computação (SBC) é uma Sociedade Científica sem fins lucrativos, fundada em 24 de Julho de 1978, em Porto Alegre, Rio Grande do Sul, que reúne estudantes, professores, profissionais, pesquisadores e entusiastas da área de Computação e Informática de todo o Brasil. A SBC tem como função fomentar o acesso à informação e cultura por meio da informática, promover a inclusão digital, incentivar a pesquisa e o ensino em computação no Brasil, e contribuir para a formação do profissional da computação com responsabilidade social (SBC, 2022).

3.3 Diretoria de Educação

A Sociedade Brasileira de Computação conta com uma diretoria específica para os assuntos relacionados a educação. À diretoria de educação compete presidir a Comissão de Educação, bem como, supervisionar a realização de eventos relativos à discussão de assuntos ligados ao ensino de computação e ao exercício da profissão e representar a SBC em foros destinados à discussão de assuntos ligados ao ensino de computação e ao exercício da profissão. (SBC, 2022)

3.4 Relatório Educacional da SBC

A diretoria de educação da SBC é responsável por analisar os cursos de ensino superior da área de tecnologia do Brasil. Uma das maneiras com que ela analisa os cursos é através dos dados obtidos do MEC. Esses dados são analisados manualmente pela diretoria de educação, organizados e disponibilizados para a comunidade através de um relatório anual. Esse relatório expõe dados estatísticos de como os cursos de tecnologia evoluem ao logo do tempo. Os relatórios educacionais estatísticos podem ser consultados no site da SBC[1].

Os relatórios são disponibilizados desde o ano de 2010, e apresentam diversos dados relativos, como por exemplo:

- Números de cursos de tecnologia;
- Números de cursos criados;
- Cursos criados por região;
- Evolução quantitativa de cursos ao longo dos anos por região;
- Acompanhamento dos estudantes ingressantes e concluintes nos cursos de tecnologia;
- Crescimento no número de alunos comparado aos anos anteriores, e diversas outras análises.

Cabe ressaltar que o relatório, desde 2010, é elaborado de forma manual.

[1] https://www.sbc.org.br/documentos-da-sbc/category/133-estatisticas

3.5 Requisitos

O requisito mínimo que deve-se atender com o trabalho proposto é de permitir à Sociedade Brasileira de Computação a disponibilização dos dados estatísticos referentes ao ensino superior brasileiro, permitindo uma análise quantitativa relacionada aos cursos, aos estudantes e as vagas ao longo dos anos. Estes dados serão disponibilizados através de um base de dados dentro do elasticsearch, bem como, através de um *dashboard* com representações gráficas dos dados. Estas representações foram desenvolvidas com base nos gráficos que constam no relatório anual do ensino superior que a SBC disponibiliza, e também com base nos dados que julgamos mais relevantes para entender os cursos de ensino superior da área de tecnologia. Estas representações gráficas ainda serão avaliadas pela diretoria de educação da SBC, a fim de, determinar a sua relevância para organização.

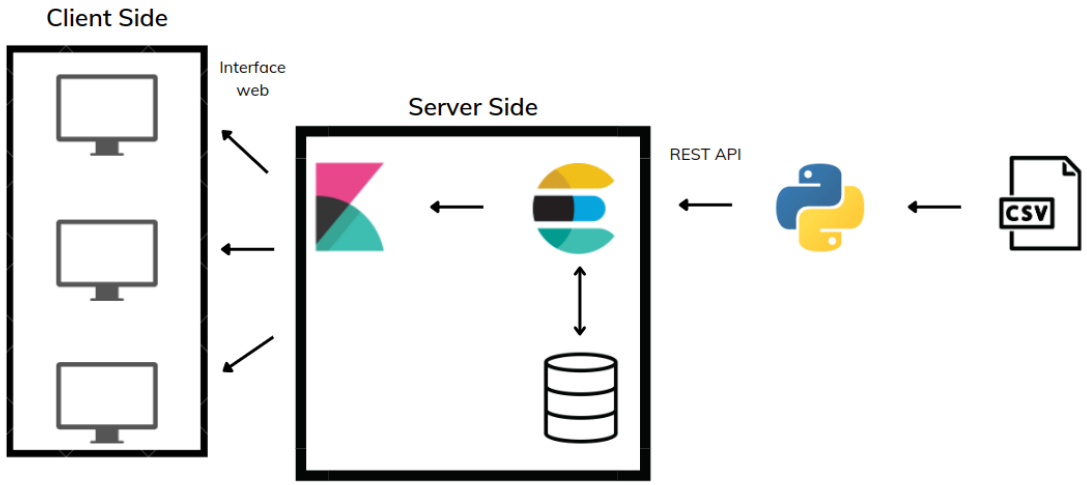
4 PROJETO E A IMPLEMENTAÇÃO

Este capítulo descreve as atividades projetadas e implementadas para a construção da ferramenta de análise de dados educacionais, tais como: coleta de dados, modelagem dos dados e módulos da implementação da ferramenta proposta.

4.1 Visão Geral

As ferramentas utilizadas no trabalho são o Elasticsearch, que será responsável por armazenar e indexar os dados e o Kibana, que será responsável por gerar as visualizações e permitir a análise dos dados indexados. A arquitetura utilizada pela ferramenta, baseia-se no modelo cliente-servidor, através do protocolo de comunicação HTTPS. O banco de dados utilizado pela ferramenta é um banco NoSQL, e um banco de arquivos de dados JSON semiestruturado. A linguagem utilizada no desenvolvimento do Elasicsearch é o Java. Para o Kibana, são utilizadas as linguagem de TypeScript e Javascript. Já para a inserção dos dados, criou-se uma aplicação utilizando a linguagem Phyton que se comunica com o Elasticsearch através de uma API Rest.

Figura 4.1: Arquitetura do Sistema



Fonte: Autor, 2022.

A Figura 4.1 mostra como ficou a arquitetura do sistema, explicaremos brevemente como os componentes do sistema se comunicam afim de tornar mais claro o funcionamento do mesmo. Podemos observar que o cliente acessa o Kibana utilizando uma interface web atreves do protocolo HTTPS, o Kibana por sua vez se comunica com o Elasticsearch utilizando uma interface REST e por fim, Elasticsearch faz as consultas

no banco. Os dados foram inseridos através de uma API REST, para a inserção dos dados se desenvolveu uma aplicação utilizando Python 3.0 que consome os dados do CSV e converte para JSON onde a chave de cada objeto é o nome da coluna do CSV e o valor é o valor que consta na linha para quela coluna.

4.2 Obtenção dos Dados

Os dados analisados foram as informações sobre o ensino superior do Brasil, sendo dados oficias a fim de de ser obtido números reais a respeito do tema. Os dados foram obtidos diretamente do site do MEC.[1]

O Governo Federal disponibiliza desde o ano de 2004 os microdados do ensino superior do Brasil. Esses microdados são um conjunto de informações detalhadas dos estudantes, cursos e instituições de educação superior. Os microdados representam um acervo que, se considerado em conjunto com o conceito Enade, permite análises sobre elementos que interferem no desempenho dos estudantes.

Os microdados estão disponíveis para consulta de forma gratuita através do site do MEC (MEC, 2022). Nesse site, os dados estão dispostos pelo ano do censo, e cada ano possui um conjunto de arquivos. Diante disso, o foco desse estudo terá dois arquivos principais: O primeiro é um documento no formato CSV, onde cada linha apresenta um curso de uma instituição de ensino e as colunas apresentam dados referentes a esse curso como: número de ingressantes, número de vagas e etc..., O segundo arquivo se trata de uma planilha com o dicionário do CSV, informando o que cada coluna representa.

Os documentos foram extraídos do site no dia 14 de abril de 2022, até está data o MEC só havia disponibilizados dados até o ano de 2020.

A Tabela 4.1 exibe parte do dicionário dos dados. É importante destacar que algumas métricas começaram a ser contabilizadas a partir do ano de 2013, como exemplo, o número de novas vagas por curso (está descrito no dicionário). Outro ponto importante a destacar, é que diversos dados se encontram incorretos, um exemplo, é o número de novas vagas por curso. Em diversos cursos aparecem valores que não correspondem com o número de novas vagas disponibilizadas. Outro exemplo é o número de alunos matriculados em diversos cursos, onde os valores são apresentados zerados ou vazio no CSV.

[1] https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/censo-da-educacao-superior

Tabela 4.1: Dicionário da base de dados

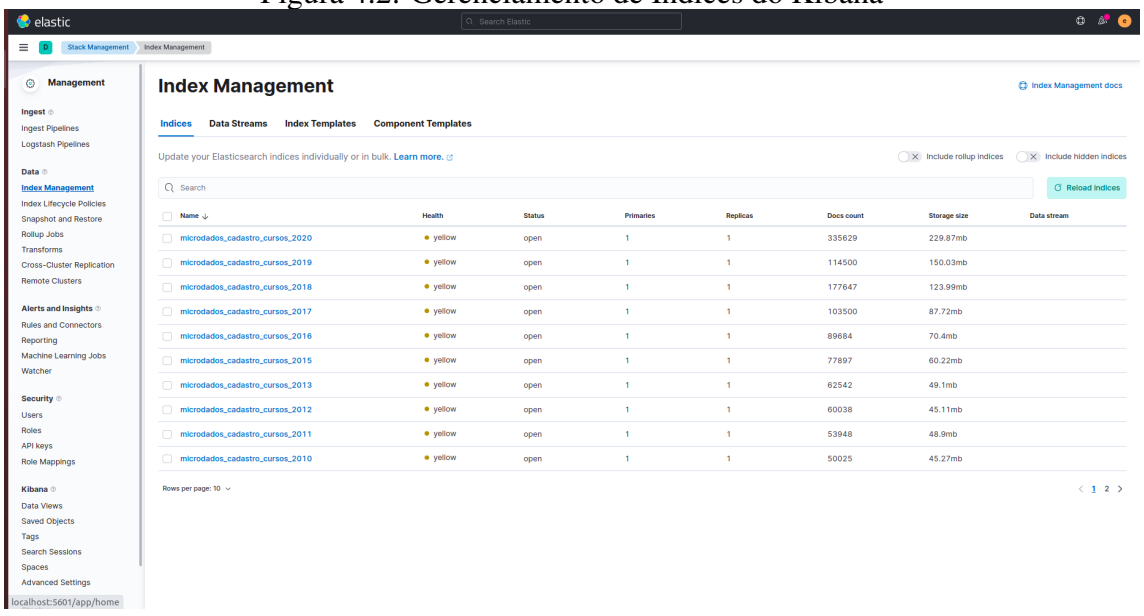
Nome da Variável	Descrição da Variável
NU_ANO_CENSO	Ano de referência do Censo da Educação Superior
NO_REGIAO	Nome da região geográfica do local de oferta do curso
SG_UF	Sigla da Unidade da Federação do local de oferta do curso
NO_MUNICIPIO	Nome do Município do local de oferta do curso
IN_CAPITAL	Informa se o local de oferta do curso está localizado em capital de Estado
TP_ORGANIZACAO_ACADEMICA	Tipo da Organização Acadêmica da IES
TP_CATEGORIA_ADMINISTRATIVA	Tipo da Categoria Administrativa da IES
TP_REDE	Rede de Ensino
CO_IES	Código da Instituição
NO_CINE_ROTULO	Nome do curso, conforme adaptação da Classificação Internacional Normalizada da Educação Cine/Unesco
TP_ORGANIZACAO_ACADEMICA	Tipo da Organização Acadêmica da IES
TP_CATEGORIA_ADMINISTRATIVA	Tipo da Categoria Administrativa da IES
TP_REDE	Rede de Ensino
CO_IES	Código da Instituição
NO_CINE_ROTULO	Nome do curso, conforme adaptação da Classificação Internacional Normalizada da Educação Cine/Unesco

Fonte: Autor

4.3 Modelagem dos Dados

A modelagem dos dados segue o padrão utilizado pelo Elasticsearch, ou seja, os dados são organizados em formato de JSON, onde cada linha é um objeto e as chaves desse objeto são as colunas. Esses objetos são organizados pelo que o Elasticsearch chama de índice. Um índice seria como uma tabela em um banco relacional SQL, cada índice guarda um conjunto de datas de um determinado tipo. No caso deste trabalho, cada índice representa os dados dos cursos de ensino superior por ano.

Figura 4.2: Gerenciamento de Índices do Kibana

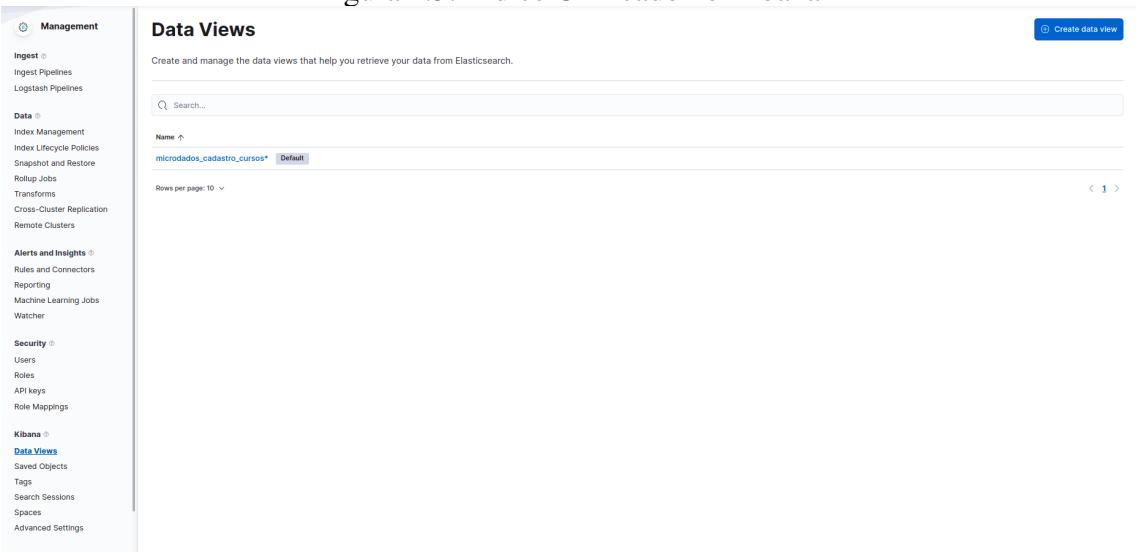


Fonte: Autor, 2022.

A Figura 4.2 mostra a disposição dos índices no sistema, bem como, o valor em bytes que cada ano do senso ocupa no sistema de arquivos do sistema operacional.

O Kibana permite que diversos índices do Elasticsearch sejam mesclados, assim, criando uma única fonte de dados. No caso deste trabalho, utilizamos essa funcionalidade a fim de mesclar os índices de cada ano, criando dessa forma, uma fonte de dados que engloba os dados de todos os anos, como podemos ver na figura 4.3.

Figura 4.3: Índice Unificado no Kibana



Fonte: Autor, 2022.

4.4 Implantação

A implantação do projeto demandou diversos conhecimentos da área de tecnologia, como por exemplo, programação, estrutura de dados, configuração de ferramentas, entre outros.

Na implantação utilizou-se o sistema operacional Linux na distribuição Ubuntu 20.04.4 LTS, onde as aplicações executavam. O Kibana e o Elasticsearch foram implantados utilizando uma imagem do sistema conteinerizada, utilizando o Docker version 20.10.14, build a224086, para a configuração automatizada do contêiner.

Também foi utilizado o docker-compose version 1.25.0, permitindo assim, o versionamento das configurações do contêiner. Para executar a inserção dos dados, se desenvolveu um sistema de inserção dos dados utilizando Python versão 3.0 e a biblioteca Python Elasticsearch Client vesão 8.2.0. O sistema implementa uma API REST que conversa com o Elasticsearch, criando os índices e inserindo os dados do censo. O sistema também lê os arquivos CSVs e os converte para JSON. Durante a conversão destes arquivos alguns mapeamentos de campos são executados, pois conforme do dicionário que vem junto com o CSV alguns códigos representam informações no mesmo, exemplo, a coluna TP_CATEGORIA_ADMINISTRATIVA que representa o tipo da categoria administrativa da instituição de ensino superior (IES), é codificada como podemos observar a representação desta codificação na tabela 4.2.

Tabela 4.2: Tabela que representa a codificação do campo categoria administrativa.

Código	**Valor**
1	Pública Federal
2	Pública Estadual
3	Pública Municipal
4	Privada com fins lucrativos
5	Privada sem fins lucrativos
6	Privada - Particular em sentido estrito
7	Especial
8	Privada comunitária
9	Privada confessional

Fonte: Autor, 2022.

O processo de inserção dos dados foi acelerado, pois utilizou-se a inserção de dados em lote, que foi disponibilizada pelo Elasticsearch e implementada na biblioteca Python Elasticsearch Client[2], dessa forma, permitindo que diversos registros fossem inseridos ao mesmo tempo.

[2]https://elasticsearch-py.readthedocs.io/en/v8.2.0/

5 ANÁLISE DE DADOS

Este capítulo tem o objetivo de apresentar a metodologia aplicada neste estudo, além de descrever e analisar o conjunto de dados obtidos junto ao MEC sobre o ensino superior, utilizando o Kibana como ferramenta de auxílio da análise. Para isso, serão apresentadas na sequência, a metodologia utilizada e as análises realizadas.

5.1 Metodologia

Esta pesquisa se enquadra como quantitativa, tendo em vista que a coleta os dados numéricos serão analisados através de métodos estatísticos(ALIAGA; GUNDERSON, 2002). A pesquisa também aplica uma das bases de pesquisa descritiva, pois, aborda também quatro aspectos: descrição, registro, análise e interpretação de fenômenos atuais objetivando o seu funcionamento no presente (MARCONI; LAKATOS, 2017).

5.2 Análise dos Dados

Os dados são apresentados por meio de representações gráficas e organizados por ano de pesquisa. Com base nisso, é possível ser realizada uma comparação entre eles, utilizando algumas métricas como:

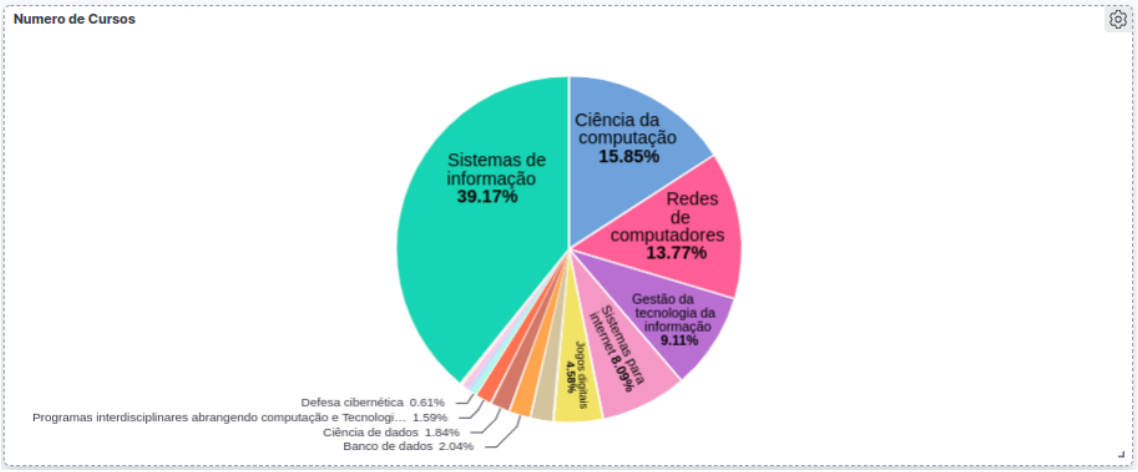
- Número de cursos totais, por IES, região e estado;
- Número de alunos por organização acadêmica;
- Número de alunos ingressantes totais, por gênero, etnia e faixa etária;
- Número de alunos concluintes totais, por gênero, etnia e faixa etária;
- Número de alunos inscritos por turno e modalidade de ensino;
- Número de vagas por turno e modalidade de ensino;

Os dados serão analisados a seguir, levando em consideração o número de cursos, número de alunos e número de vagas.

5.2.1 Número de cursos

Observando a Figura 5.1, podemos verificar que os três principais cursos de tecnologia do Brasil, considerando o número de cursos, segundo o MEC, são: Sistemas de Informação, Ciência da Computação e Redes de Computadores.

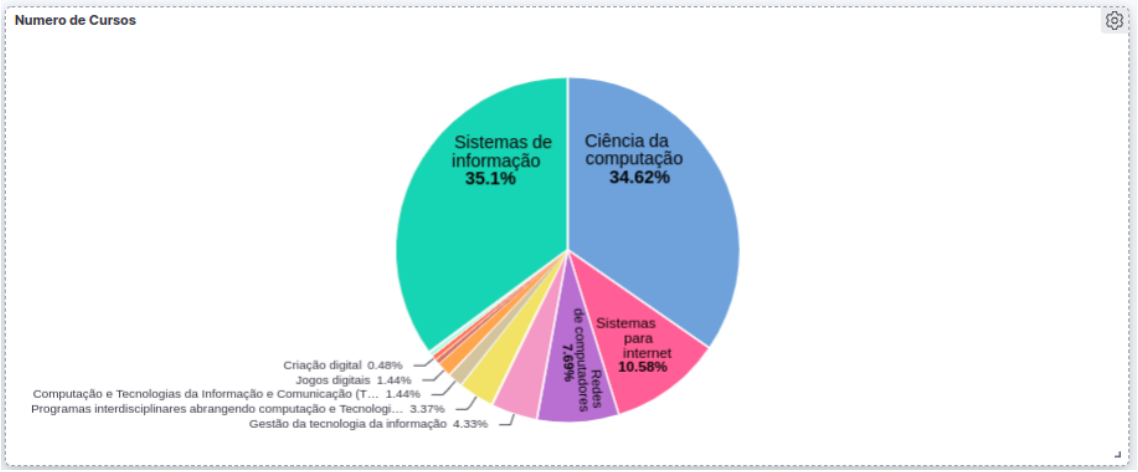
Figura 5.1: Número de cursos



Fonte: Autor, 2022.

Apresentando apenas o contexto das instituições federais, conforme apresentado na Figura 5.2, podemos analisar que o número percentual de cursos se altera, bem como as posições.

Figura 5.2: Número de cursos em IES Federais



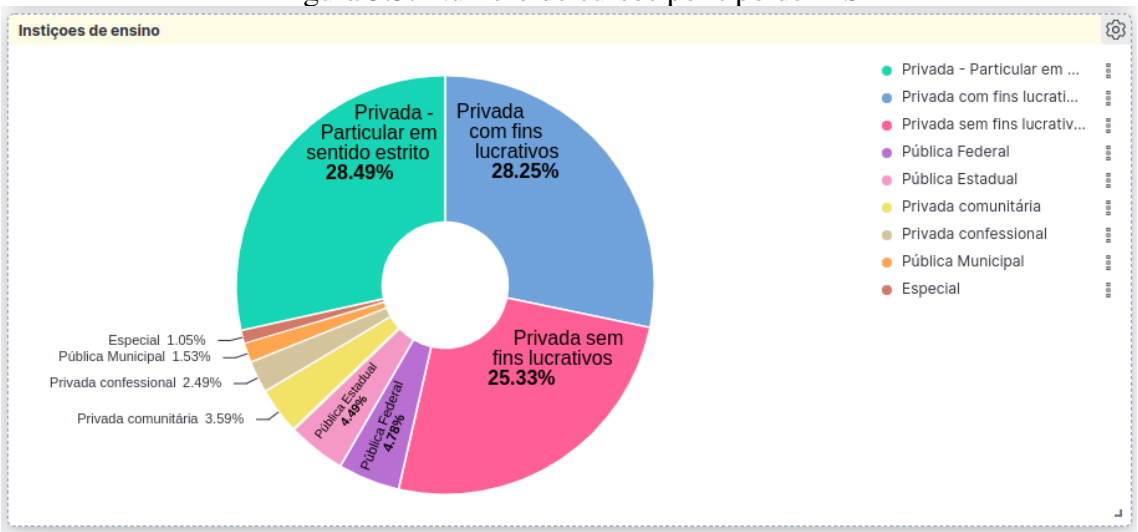
Fonte: Autor, 2022.

Analisando as duas Figuras anteriormente exibidas, podemos constatar que a relevância do curso de Ciência da Computação é maior, bem como, o curso de Sistemas de Informação.

5.2.2 Número de cursos por IES

Como podemos observar na Figura 5.3, a grande maioria dos cursos de tecnologia são em Instituições privadas. Estas, são as detentoras de aproximadamente 88,15% dos cursos de tecnologia de nível superior.

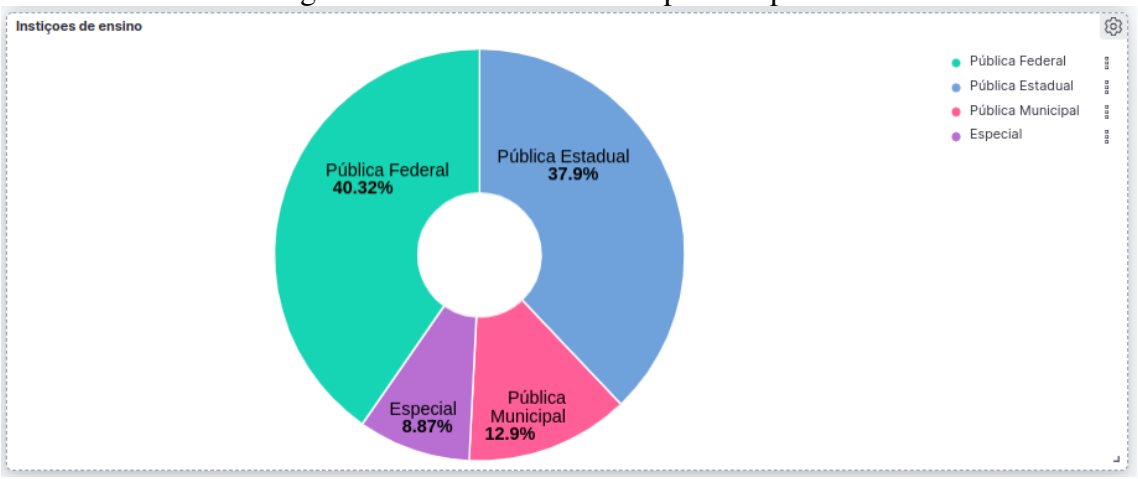
Figura 5.3: Número de cursos por tipo de IES



Fonte: Autor, 2022.

Observando o número de cursos de tecnologia na esfera pública, analisa-se que as Instituições Federais são as principais, possuindo aproximadamente 40,32% dos cursos. Isto está evidenciado na Figura 5.4:

Figura 5.4: Número de cursos por IES pública

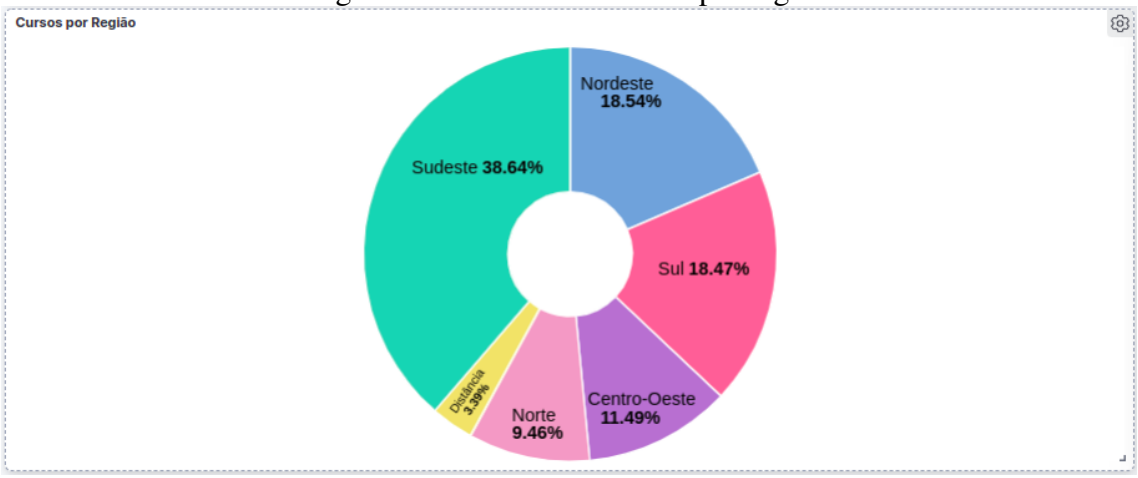


Fonte: Autor, 2022.

5.2.3 Número de cursos por Região

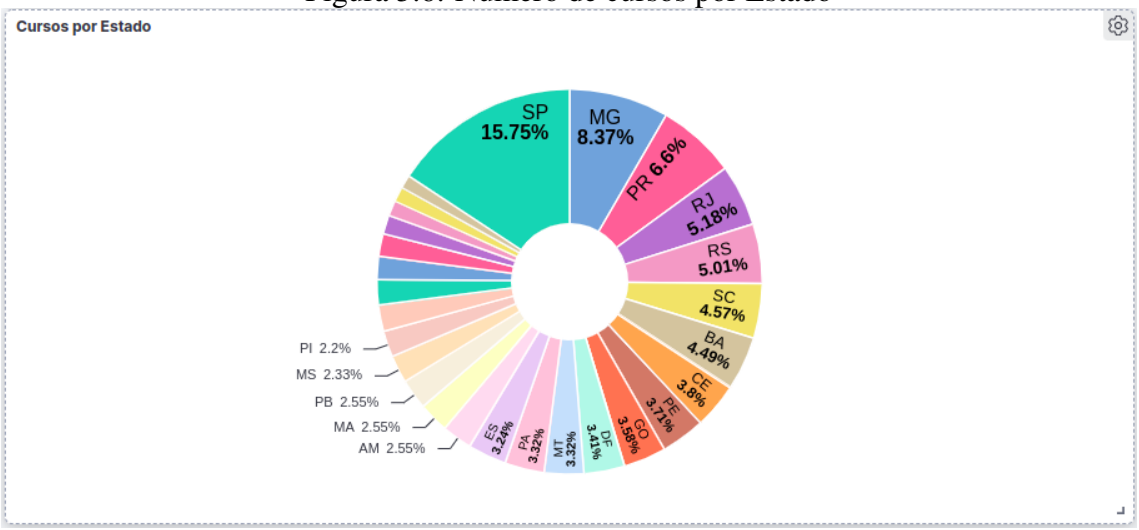
Analisando o número de cursos por região do Brasil, é evidente que a região Sudeste é a que abrange o maior número de cursos do país, sendo os estados de São Paulo e Minas Gerais, os detentores por cerca de 24,14% dos cursos disponibilizados, conforme apresentados nas Figuras 5.5 e 5.6.

Figura 5.5: Número de cursos por região



Fonte: Autor, 2022.

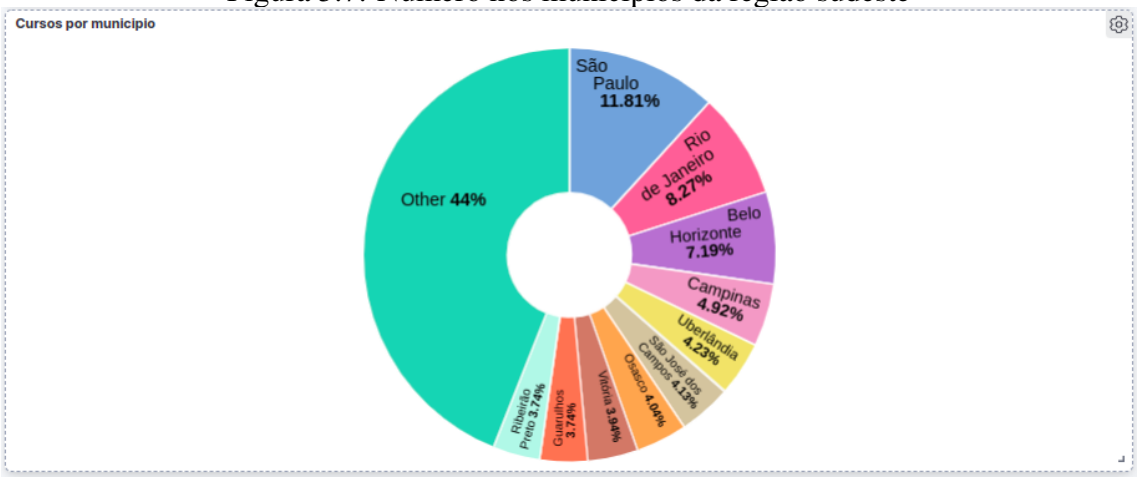
Figura 5.6: Número de cursos por Estado



Fonte: Autor, 2022.

Se analisarmos de forma isolada a região Sudeste, podemos observar que os municípios de São Paulo, Rio de Janeiro e Belo Horizonte são onde se encontram a maioria dos cursos de tecnologia, conforme, Figura 5.7.

Figura 5.7: Número nos municípios da região sudeste



Fonte: Autor, 2022.

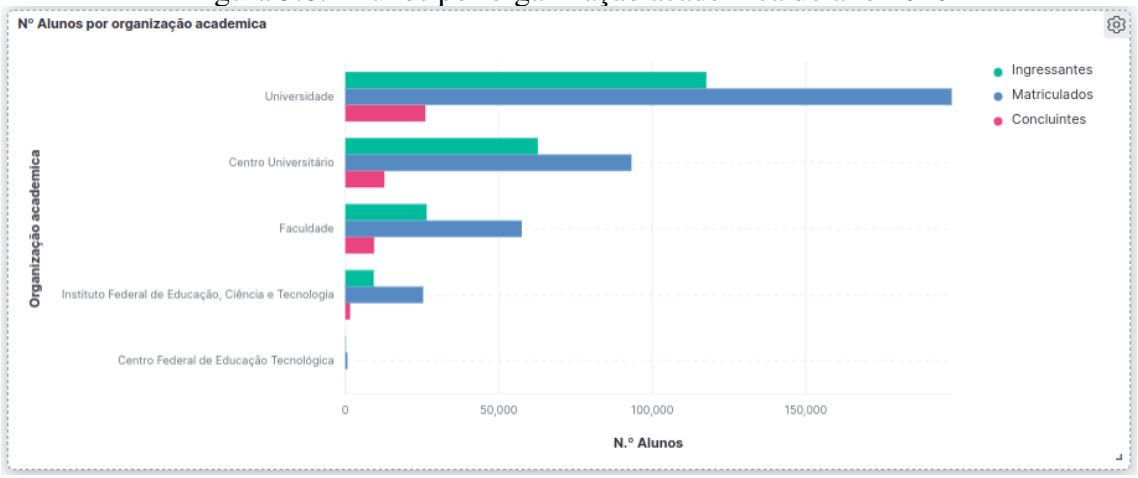
5.2.4 Número de estudantes por Organização Acadêmica

Outra análise possível com os dados do MEC é o número de estudantes por organização acadêmica.

Com base nisso, na Figura 5.8, é possível observar o número de estudantes que são ingressantes, matriculados e concluintes de cada tipo de organização acadêmica. Portanto, verifica-se que as universidades são as detentoras da maior parte dos

estudantes. Os centros universitários e faculdades são os próximos.

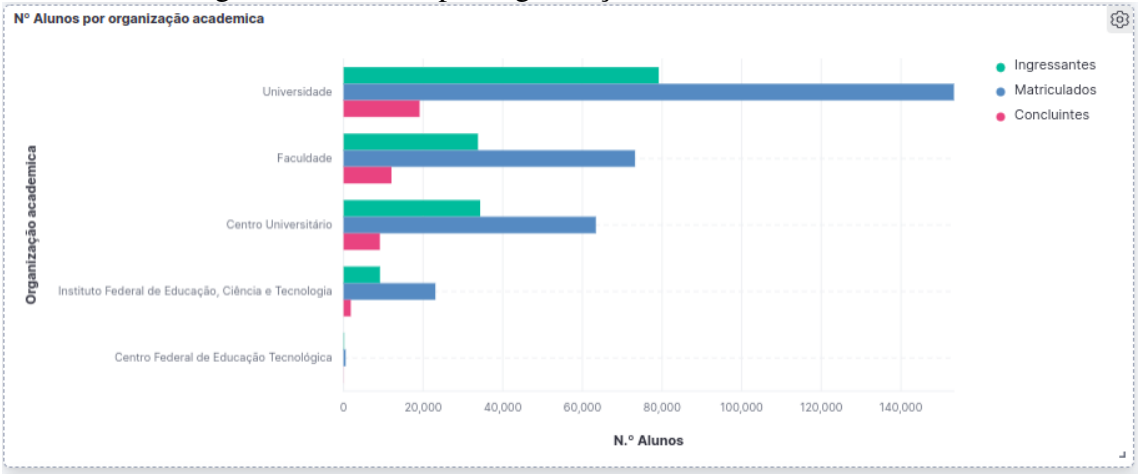
Figura 5.8: Alunos por organização acadêmica do ano 2020



Fonte: Autor, 2022.

Comparando os dados de 2018 com os de 2020 das Figuras 5.9 e 5.8, podemos observar que em 2018 as faculdades possuíam um número de estudantes maior do que os dos centros universitários. O que evidência um crescimento dos centros universitários, e dessa forma, podemos inferir que é possível estar acontecendo um aumento na diversidade de cursos e/ou do número de cursos de extensão nos anos de 2019 e 2020.

Figura 5.9: Alunos por organização acadêmica do ano 2018

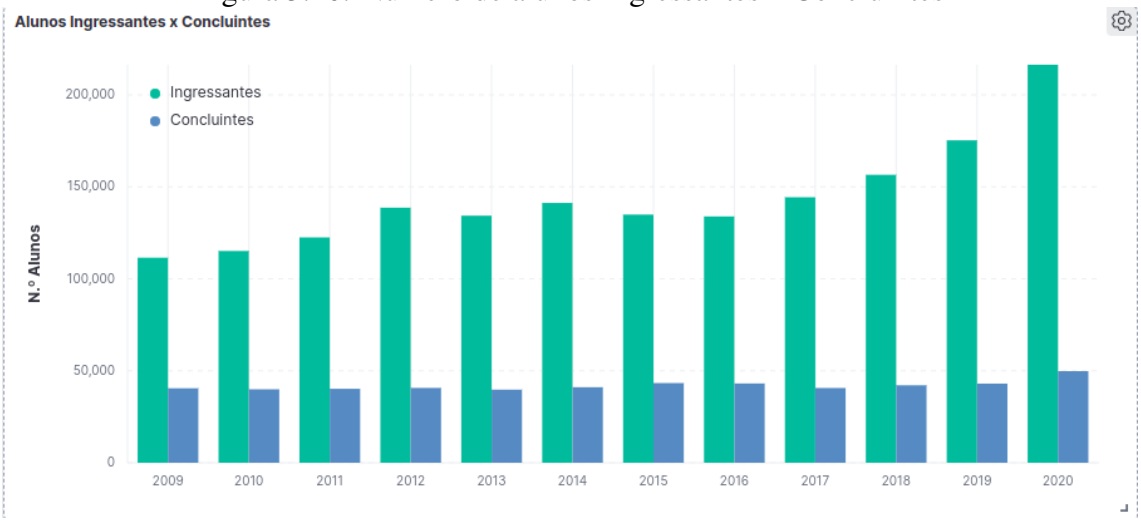


Fonte: Autor, 2022.

5.2.5 Número de estudantes Ingressantes x Concluintes

Na Figura 5.10, podemos analisar o número de ingressantes versus concluintes nos cursos de tecnologia do ano 2009 a 2020. A Figura torna evidente a grande diferença entre o número de ingressantes e concluintes. Estes dados nos permitem inferir uma grande evasão dos cursos de tecnologia.

Figura 5.10: Número de alunos Ingressantes x Concluintes



Fonte: Autor, 2022.

Na Figura 5.11 podemos verificar os mesmos dados já vistos na Figura 5.10, contudo, os dados são apresentados em forma de tabela. Nessa visualização, podemos analisar os valores totais de alunos ingressantes e concluintes por ano dos cursos da área de tecnologia.

Figura 5.11: Número de alunos Ingressantes x Concluintes Tabelado

Numero ingressantes x concluintes por curso

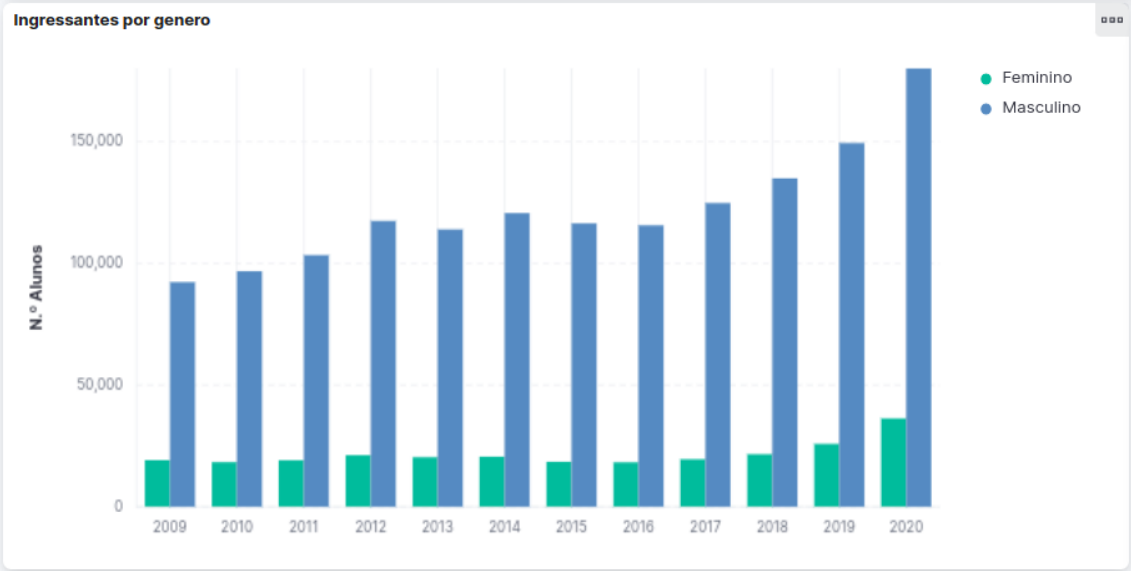
Curso	2020 › Ingressante	2020 › Concluinte	2019 › Ingressante	2019 › Concluinte	2018 › Ingressante	2018 › Concluinte	2017 › Ingressante
Sistemas de informação	124,317	28,618	101,201	24,206	90,792	23,458	82,362
Ciência da computação	28,860	6,683	26,270	6,426	23,576	6,657	23,080
Gestão da tecnologia da info...	24,438	6,856	21,223	5,792	18,392	5,453	17,940
Redes de computadores	9,977	3,408	9,812	3,154	10,556	3,305	9,655
Ciência de dados	8,641	61	-	-	-	-	-
Sistemas para internet	5,027	1,198	4,241	1,216	4,406	1,243	4,528
Jogos digitais	4,416	1,507	4,314	1,041	4,096	990	3,444
Segurança da informação	4,407	581	2,227	471	2,009	502	1,237
Banco de dados	2,350	470	1,642	392	1,271	287	1,123
Defesa cibernética	2,110	124	756	69	163	0	-
Programas interdisciplinares ...	851	152	3,598	255	1,290	214	967
Criação digital	367	52	-	-	-	-	-
Internet das coisas	192	33	-	-	-	-	-
Agrocomputação	168	30	-	-	-	-	-
Computação e Tecnologias ...	138	37	-	-	-	-	-
Inteligência artificial	79	0	-	-	-	-	-
Sistemas embarcados	58	6	25	5	13	3	17

Fonte: Autor, 2022.

5.2.6 Número de alunos Ingressantes x Concluintes por gênero

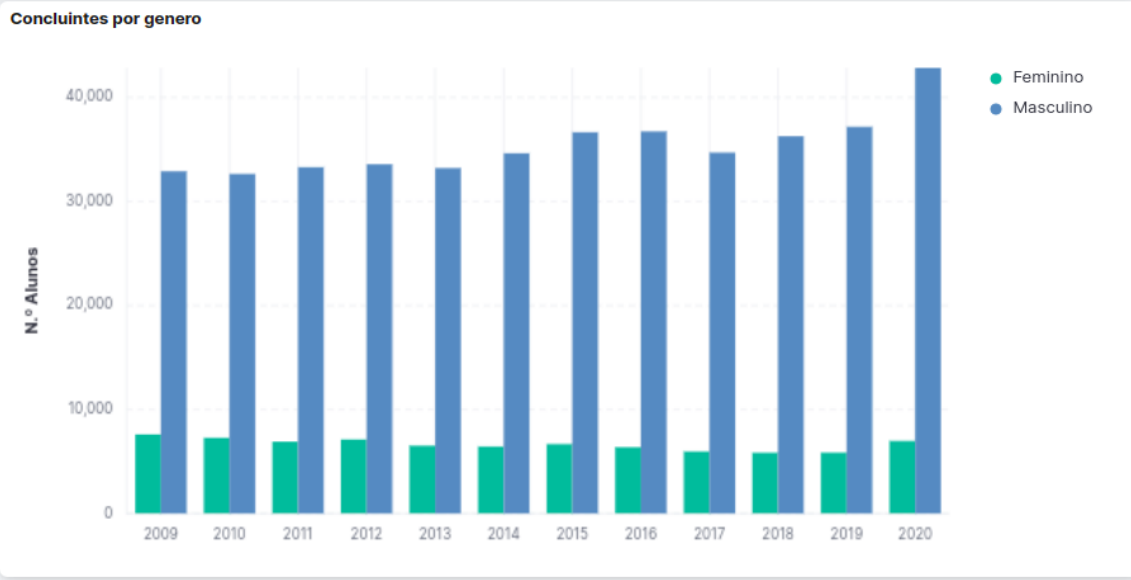
Observando a Figura 5.12, podemos notar a grande discrepância no gênero dos alunos ingressantes. O número de estudantes do gênero masculino que ingressam nos cursos de tecnologia, é aproximadamente 5 vezes maior que as estudantes do gênero feminino.

Figura 5.12: Número de alunos Ingressantes por Gênero



Fonte: Autor, 2022.

Figura 5.13: Número de alunos Concluintes por Gênero



Fonte: Autor, 2022.

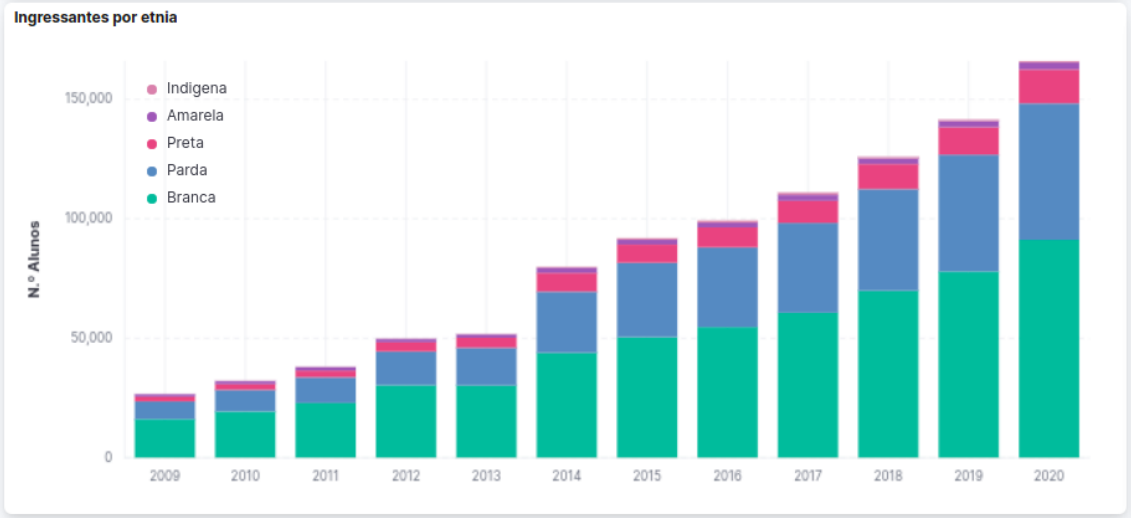
Na Figura 5.13, verificamos que o padrão se repete. O número de estudantes do gênero feminino que concluem os cursos da área de tecnologia é aproximadamente 7

vezes menor que os do gênero masculino. Analisando esses dados, constatamos que a representatividade das mulheres na área da tecnologia é muito inferior a dos homens, e com base nisso, podemos promover medidas a fim de diminuir essa grande diferença.

5.2.7 Número de estudantes Ingressantes x Concluintes por etnia

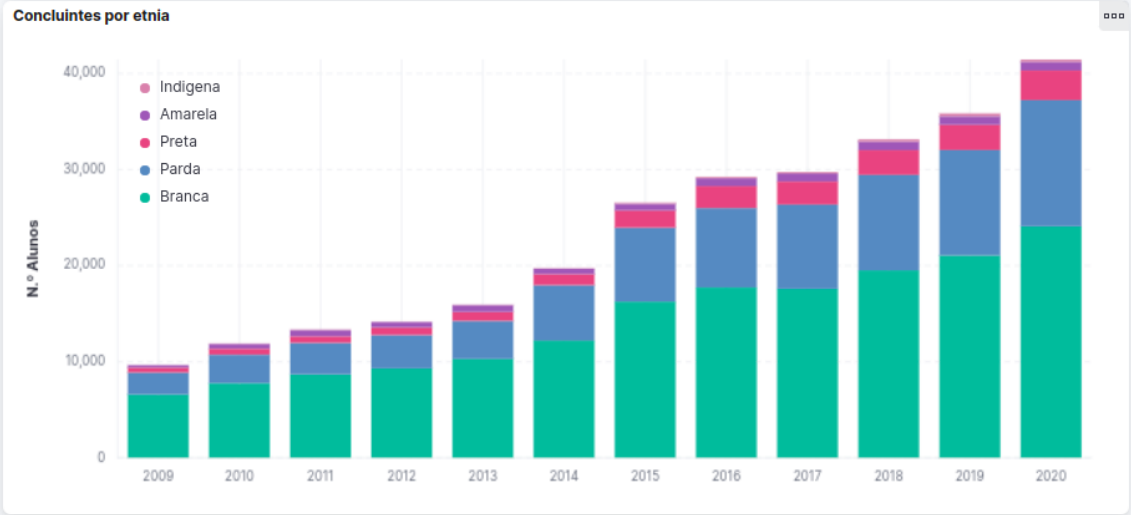
Como podemos ver na Figura 5.14 e na Figura 5.15, a etnia preponderante nos cursos de tecnologia é a branca, tanto nos ingressantes quanto nos concluintes. É possível observar também um crescimento ao longo dos anos de quase todas as etnias. As únicas etnias que não tiveram um aumento significativo foram: a indígena e a amarela.

Figura 5.14: Número de alunos Ingressantes por Etnia



Fonte: Autor, 2022.

Figura 5.15: Número de alunos Concluintes por Etnia

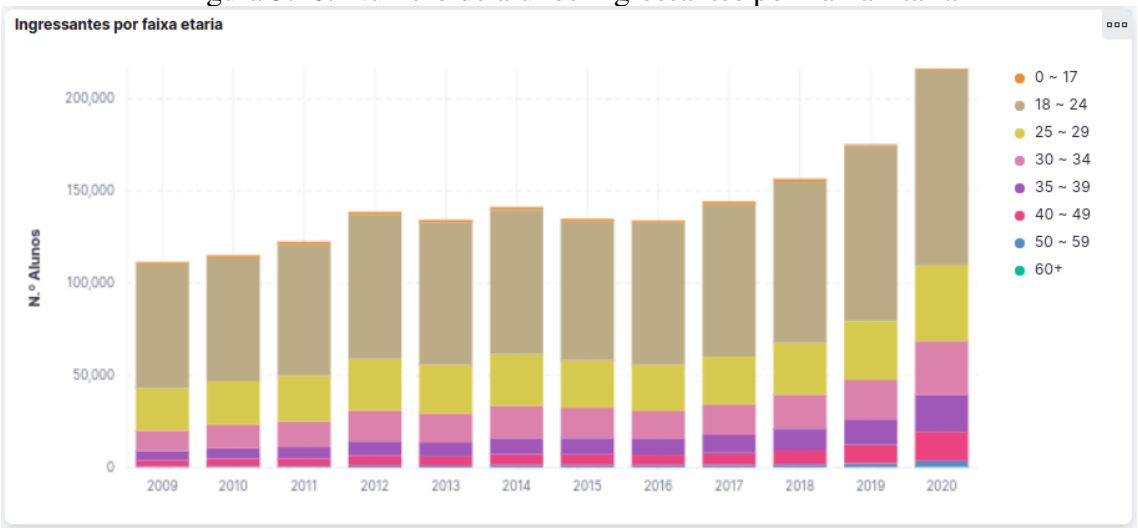


Fonte: Autor, 2022.

5.2.8 Número de estudantes Ingressantes x Concluintes por faixa etária

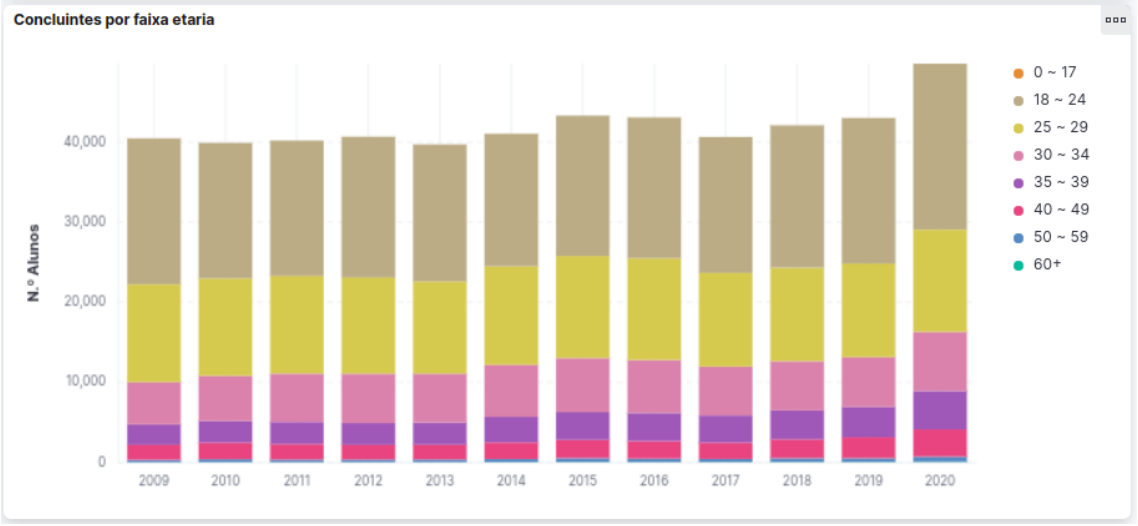
As Figuras 5.16 e 5.17 nos mostram que a grande parte dos estudantes ingressantes e concluintes dos cursos de tecnologia se encontram entre os 18 e 29 anos de idade. Podemos observar também um leve aumento, ao longo dos anos, em todas as faixa etárias até a faixa etária de 40 a 49 anos de idade.

Figura 5.16: Número de alunos Ingressantes por Faixa Etária



Fonte: Autor, 2022.

Figura 5.17: Número de alunos Concluintes por Faixa Etária

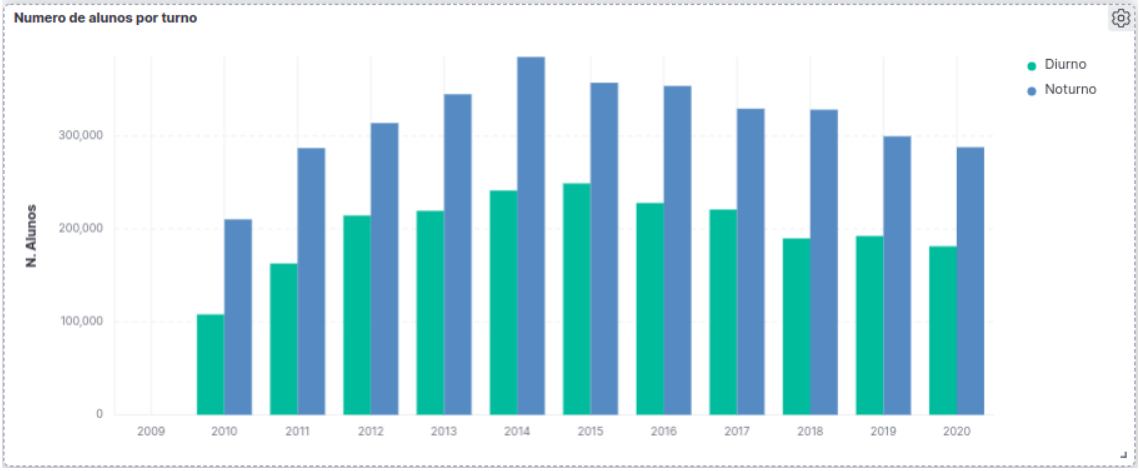


Fonte: Autor, 2022.

5.2.9 Número de inscritos e vagas por turno

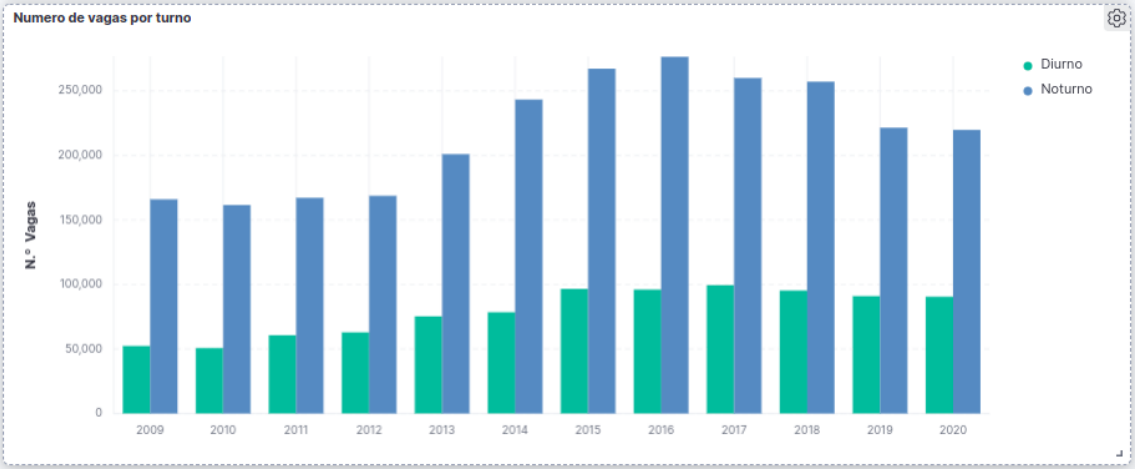
Observando as Figuras 5.18 e 5.19, podemos verificar que a maioria dos estudantes que cursam os cursos de tecnologia no Brasil, o realizam no turno da noite, bem como a maioria das vagas dos cursos de tecnologia também é para o turno da noite.

Figura 5.18: Número de alunos Inscritos por Turno



Fonte: Autor, 2022.

Figura 5.19: Vagas por Turno



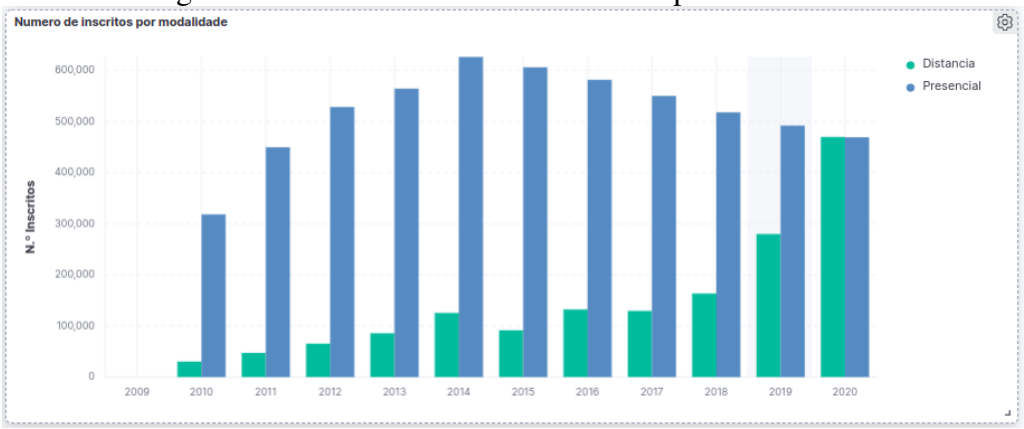
Fonte: Autor, 2022.

Analisando os dados das Figuras 5.18 e 5.19, podemos levantar uma suposição que venha justificar o número de inscritos em cursos noturnos bem como o de vagas, relembrando os dados obtidos nas Figuras 5.3 e 5.17: Grande parte dos cursos de tecnologia são de Instituições privadas e a grande maioria dos estudantes estão entre os 18 e 35 anos de idade, portanto, a grande maioria trabalha durante o dia para custear os estudos.

5.2.10 Número de inscritos e vagas por modalidade de ensino

Observando os dados a respeito da modalidade de ensino nas Figuras 5.20 e 5.21, podemos observar uma mudança significativa. Nos anos inicias do censo, o número de estudantes inscritos na modalidade presencial era predominante aos inscritos na modalidade à distância. Contudo, a partir do ano de 2015, pode-se observar um movimento de decrescimento dos alunos inscritos presencialmente e um crescimento dos inscritos à distância, com o ano de 2020 sendo o primeiro ano onde o número de inscritos para educação à distancia ultrapassa os presencias.

Figura 5.20: Número de alunos inscritos por Modalidade



Fonte: Autor, 2022.

O número de vagas para o ensino à distancia vem crescendo exponencialmente desde 2015, portanto, podemos inferir que o grande aumento no ano de 2020 deve-se à necessidade do distanciamento social em decorrência da pandemia ocasionada pelo COVID-19.

Figura 5.21: Vagas por Modalidade

Fonte: Autor, 2022.

Finalizando a análise do dados de modo geral e simplista, podemos concluir que a maioria dos estudantes de tecnologia do Brasil são: homens brancos entre 18 e 29 anos de idade que estudam presencialmente no turno noturno ou na modalidade à distância em uma IES privada.

5.2.11 Automatização da coleta dos dados

O site do MEC com uma determinada frequência remove os microdados assim impossibilitando sua consulta. Uma forma de mitigar esse problema seria desenvolver um *crawler* que com uma determinada frequência acessa o site do MEC e faz a coleta dos dados. Tendo em vista essa funcionalidade de coleta programada, podemos automatizar todo o processo de inserção dos dados assim tornando a aplicação mais independente e útil, pois ela estará atualizada. Para efetuar essa automatização basta alterar o sistema de inserção adicionado o *crawler* ao código e tornando o mesmo um serviço ou ate mesmo uma aplicação agendada utilizando o crontab ou algum outro aplicativo semelhante.

6 CONCLUSÃO

Levando em consideração que os dados disponíveis para a realização deste estudo foram atualizados até 2020, ano que iniciou a pandemia do Covid-19, podemos analisar que desde o ano de 2015 já estava aumentando o número de procura por cursos na modalidade à distância. Portanto, provavelmente, quando o MEC liberar os dados atualizados de 2021-2022, estes números estarão maiores.

Podemos concluir também que ainda há uma predominância do gênero masculino nos cursos de tecnologia, havendo a necessidade de inclusão de mulheres na área, bem como das pessoas das etnias preta, amarela e indígena. Outra consideração importante observada nos dados apurados, é a taxa de evasão dos cursos de tecnologia, que é extremamente alta comparando o número de ingressantes e concluintes ao longo do tempo. Talvez, algumas medidas para retenção dos alunos pudessem ser tomadas a fim de formar mais profissionais.

Uma possibilidade para trabalhos futuros é não somente analisar os dados dos cursos de tecnologia, e sim, os dados de todos os cursos de ensino superior do Brasil, bem como, os dados do ensino básico e do ENEM, assim, criando um grande mapa do ensino do Brasil. Os dados já estão disponíveis no site do MEC e são de fácil acesso. Este mapa em mãos das autoridades competentes poderia traçar um rumo para a excelência acadêmica do Brasil.

REFERÊNCIAS

ABUBAKAR, Y.; ADEY, T. S.; AUTA, I. G. Performance evaluation of nosql systems using ycsb in a resource austere environment. **International Journal of Applied Information Systems (IJAIS)**, v. 1, p. 27–23, 2014.

ALIAGA, M.; GUNDERSON, B. Interactive statistics. **Thousand Oaks: Sage**, 2002.

ASTERA. **Definição de API REST: Noções básicas de APIs REST**. 2020. Disponível na Internet: <https://www.astera.com/pt/type/blog/rest-api-definition/>.

CNN. **Procura por profissionais de tecnologia cresce 671 por cento durante a pandemia**. 2021. Disponível na Internet: <https://www.cnnbrasil.com.br/business/procura-por-profissionais-de-tecnologia-cresce-671-durante-a-pandemia/>.

DAVOUDIAN, L. C. A. A survey on nosql stores. **ACM Computing Surveys**, v. 51, n. 40, p. 1–44, 2019.

ELASTIC. **Elastic Docs**. 2022. Disponível na Internet: <https://www.elastic.co/guide/en/elasticsearch/reference/current/index.html>.

KONONENKO, O.; BAYSAL, O.; HOLMES, R.; GODFREY, M. W. Mining modern repositories with elasticsearch. **Proceedings of the 11th Working Conference on Mining Software Repositories. ACM**, S.1, p. 328–331, 2014.

MARCONI, M. de A.; LAKATOS, E. M. **Fundamentos de Metodologia Científica**. [S.l.: s.n.], 2017.

MEC. **Microdados do Ensino Superior**. 2022. Disponível na Internet: <https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/censo-da-educacao-superior>.

SBC. **Sobre SBC**. 2022. Disponível na Internet: <https://www.sbc.org.br/institucional-3>.

SUJAN, Y. M.; SHASHIDHARA, H. D.; NAGAPADMA, D. R. Survey paper: Framework of rest apis. **International Research Journal of Engineering and Technology (IRJET)**, v. 07, n. 06, p. 1, 2020.