**Análise dos indicadores de fluxo da educação superior no Brasil**

**Grupo 5**

**Andrey Lucio Randis, Igor Leandro Barroso Santos, Matheus Alves Bispo,**

**Rafael da Silva Dias Amphilo, Rodrigo Gomes Vasconcelos Mota**

**Alisson Rabelo Arantes**

PUC Minas

Curso de Tecnologia em Banco de Dados

alrandis@sga.pucminas.br, igor.leandro@sga.pucminas.br, mabispo@sga.pucminas.br, rafael.amphilo@sga.pucminas.br, rgvmota@sga.pucminas.br.

***Resumo.***

# Introdução

O contexto escolhido para este trabalho é a educação superior no Brasil, sobretudo nos aspectos que envolvem o acesso, a evasão e a conclusão desta modalidade de ensino. Primordial para a formação de profissionais qualificados e para o desenvolvimento social e econômico de qualquer país, o ensino superior está inserido em um cenário macro ambiental complexo, interdependente e assim como todos os setores da sociedade brasileira enfrenta o problema sistêmico da desigualdade social.

A preocupação deste projeto está voltada para a desigualdade do acesso e da permanência no ensino superior. É preciso compreender as razões pelas quais alguns alunos estão sendo excluídos desse nível de ensino e como essa exclusão está relacionada com as oportunidades educacionais, sociais e econômicas dos estudantes (GISI, 2006). Além disso, é necessário entender como as políticas públicas podem ser direcionadas para mitigar esses problemas.

A motivação para o desenvolvimento deste trabalho é a transformação social e pessoal que o ensino superior pode proporcionar ao estudante e a sociedade. Entender quem consegue ingressar e quem consegue ir até o final da graduação é fundamental para implantar políticas públicas que garantam o acesso universal ao ensino superior, bem como a sua conclusão (OLIVEIRA BARBOSA, 2021).

Ao analisar e tratar os dados relacionados ao fluxo de educação superior no Brasil, espera-se que os dados sejam transformados em informações sobre quais são as principais barreiras e desafios que a população encontra para acessar e concluir o ensino superior.Essas informações poderão ser transformadas em conhecimento para orientar nas medidas que devem ser tomadas para que o ensino superior seja mais acessível e com menos evasão.

Acredita-se que o acesso à educação é a principal ferramenta para combater a desigualdade social.

# Descrição das Bases de Dados Utilizadas

Os dados foram encontrados no Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), disponibilizados pelo Ministério da Educação (MEC) através de seu site. (gov.br/inep/pt-br).

Os dados estão em formato de tabela e são organizados em diferentes arquivos, que contêm informações sobre matrículas, ingressantes, concluintes, evasão na educação superior e outros indicadores. O arquivo está no formato .xlsx (arquivo de planilha Excel).

Essa consulta foi realizada no dia 22 de março de 2023.

Link:[https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/indicadoreshttps://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/indicadores-educacionais/indicadores-de-fluxo-da-educacao-superioreducacionais/indicadores-de-fluxo-da-educacao-superior](https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/indicadores-educacionais/indicadores-de-fluxo-da-educacao-superior)

**Tabela 1 :Indicadores de trajetória do ensino superior 2010-2019.**

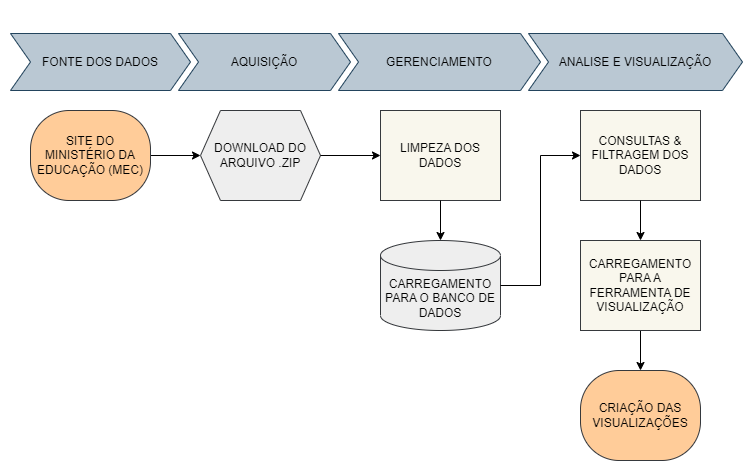
|  |  |  |
| --- | --- | --- |
| **NOME COLUNA/CAMPO** | **DESCRIÇÃO** | **TIPO** |
| Código da Instituição | Código único de identificação da instituição de educação superior em que o curso está localizado no último ano de análise. | INT |
| Nome da Instituição | Nome da instituição de educação superior em que o curso está localizado no último ano de análise. | VARCHAR (X) |
| Categoria Administrativa | Código da categoria Administrativa da IES no último ano de análise. | INT |
| Organização Acadêmica | Código da organização acadêmica no último ano de análise. | INT |
| Código do Curso de Graduação | Código único de identificação do curso gerado pelo E-MEC, com a informação do último ano de análise. | INT |
| Nome do Curso de Graduação | Nome do curso com a informação do último ano de análise. | VARCHAR (X) |
| Código da Região Geográfica do Curso | Código da Região Geográfica do local de oferta do curso gerado pelo E-MEC no último ano de análise. | INT |
| Código da Unidade Federativa do Curso | Código da Unidade da Federação do local de oferta do curso gerado pelo E-MEC no último ano de análise. | INT |
| Código do Município do Curso | Código do município do local de oferta do curso gerado pelo EMEC no último ano de análise. | INT |
| Grau Acadêmico | Código do grau acadêmico conferido ao diplomado pelo curso no último ano de análise. | INT |

|  |  |  |
| --- | --- | --- |
| Modalidade de Ensino | Código da modalidade de ensino do curso no último ano de análise. | INT |
| Código da área do Curso segundo a classificação CINE BRASIL | Código da área geral conforme adaptação da Classificação Internacional Normalizada da  Educação Cine/Unesco | INT |
| Nome da área do Curso segundo a classificação CINE BRASIL | Nome da área geral conforme adaptação da Classificação Internacional Normalizada da  Educação Cine/Unesco | VARCHAR (X) |
| Código da Grande Área do Curso segundo a classificação  CINE BRASIL | Código de identificação do curso, conforme adaptação da  Classificação Internacional Normalizada da Educação  Cine/Unesco | INT |
| Nome da Grande Área do Curso segundo a classificação CINE  BRASIL | Nome de identificação do curso, conforme adaptação da  Classificação Internacional Normalizada da Educação  Cine/Unesco | VARCHAR (X) |
| Ano de Ingresso | Ano de ingresso do aluno no curso. | YEAR |
| Ano de Referência | Ano de referência do vínculo do ingressante. | YEAR |
| Prazo de Integralização em  Anos | Prazo mínimo de integralização de curso de graduação em  número de anos | INT |
| Ano de Integralização do Curso | Ano previsto de integralização do aluno ao curso. | YEAR |
| Prazo de Acompanhamento do  Curso em anos | Prazo máximo de integralização de curso de graduação em  número de anos | INT |
| Ano Máximo de  Acompanhamento do Curso | Ano máximo de acompanhamento da situação de vínculo do aluno. | YEAR |
| Quantidade de Ingressantes no Curso | Número de ingressantes do curso no ano de ingresso da corte*.* | INT |
| Quantidade de Permanência no  Curso no ano de referência | Número de estudantes que permaneceram no curso de graduação no ano de referência da análise | INT |
| Quantidade de Concluintes no  Curso no ano de referência | Número de estudantes que concluíram o curso de graduação no ano de referência da análise | INT |
| Quantidade de Desistência no  Curso no ano de referência | Número de estudantes que desistiram do curso de graduação | INT |
|  | no ano de referência da análise |  |
| Quantidade de Falecimentos no Curso no ano de referência | Número de estudantes que faleceram no ano de referência da análise | INT |
| Taxa de Permanência - TAP | Taxa de Permanência | DECIMAL (X, Y) |
| Taxa de Conclusão Acumulada TCA | Taxa de Conclusão Acumulada | DECIMAL (X, Y) |
| Taxa de Desistência Acumulada - TODA | Taxa de Desistência Acumulada | DECIMAL (X, Y) |
| Taxa de Conclusão Anual TCAN | Taxa de Conclusão Anual | DECIMAL (X, Y) |
| Taxa de Desistência Anual TDAN | Taxa de Desistência Anual | DECIMAL (X, Y) |

**3. Arquitetura**

O diagrama abaixo representa as principais etapas a serem seguidas para a execução do projeto, desde a obtenção dos dados até a visualização dos indicadores finais pós análise.

**Diagrama 1 :Arquitetura utilizada.**



**4. Metodologia**

O projeto é iniciado com a obtenção dos dados oriundos do site oficial do Ministério da Educação (MEC). A transformação dos dados é feita usando o Excel. Os dados são então carregados em um sistema gerenciador de banco de dados, na etapa de análise e visualização será realizado as consultas e filtragem dos dados, posteriormente, em uma ferramenta de visualização de dados, como o PowerBI. Esse esquema usa uma metodologia ETL (Extrair, Transformar, Carregar) padrão que permite analisar dados com eficiência em um ambiente corporativo, através dos seguintes passos:

* Remoção dos cabeçalhos extras, imagens, mesclas e campos congelados utilizando o Excel.
* Salvar o arquivo como um novo arquivo .csv com encoding utf-8 para garantir a compatibilidade com o banco de dados.
* Identificar a necessidade da criação de novas tabelas através da normalização dos dados presentes no arquivo .csv.
* Criar as tabelas no banco de dados de acordo com os tipos de dados presentes no dataset.
* Realizar o upload dos dados do arquivo .csv para o banco de dados.
* Realizar a consulta e filtragem dos dados.
* Carregar para ferramenta de visualização.
* Produto fruto da criação das visualizações, limpezas e filtragens e processos anteriores.

A partir desses dados, é possível obter informações gerais sobre indicadores de fluxo do ensino superior no Brasil para o período 2010-2019. O processamento de dados inclui a conversão de arquivos brutos em um formato adequado, a criação de tabelas em bancos de dados e a realização de uploads de dados para uso em análises subsequentes.

# 5. Captura dos Dados

Considerando que o conjunto de dados adquiridos no site do MEC já satisfaz o formato para a análise do fluxo da educação, não se faz necessário a utilização do SGBD (Sistema gerenciador de banco de dados). Optou-se pelo carregamento dos dados no ambiente da Power Plataform (Microsoft) devido a integração entre eles e o fato de estar na nuvem.

Primeiro foi realizado a criação do ambiente no SharePoint, para o carregamento do arquivo no formato .xlsx e concedido o acesso simultâneo para os membros. Posteriormente foi realizado a importação dos dados para o PowerBi, criando uma consulta com as modificações necessárias. Desta forma, obtém-se uma estrutura adequada para a visualização.

**Imagem 1: Share Point.**

Interface gráfica do usuário, Aplicativo

Descrição gerada automaticamente

**Script em linguagem M para importação dos dados para o PowerBI.**

let

Fonte = SharePoint.Files("https://sgapucminasbr.sharepoint.com/sites/projeto\_grupo\_5", [ApiVersion = 15]),

#"Linhas Filtradas" = Table.SelectRows(Fonte, each ([Name] = "indicadores\_trajetoria\_educacao\_superior\_2010\_2019.xlsx")),

#"Arquivos Ocultos Filtrados1" = Table.SelectRows(#"Linhas Filtradas", each [Attributes]?[Hidden]? <> true),

#"Invocar Função Personalizada1" = Table.AddColumn(#"Arquivos Ocultos Filtrados1", "Transformar Arquivo", each #"Transformar Arquivo"([Content])),

#"Colunas Renomeadas1" = Table.RenameColumns(#"Invocar Função Personalizada1", {"Name", "Nome da Origem"}),

#"Outras Colunas Removidas1" = Table.SelectColumns(#"Colunas Renomeadas1", {"Nome da Origem", "Transformar Arquivo"}),

#"Coluna de Tabela Expandida1" = Table.ExpandTableColumn(#"Outras Colunas Removidas1", "Transformar Arquivo", Table.ColumnNames(#"Transformar Arquivo"(#"Arquivo de Amostra"))),

**Modificações necessárias dos dados em linguagem M.**

#"Linhas Principais Removidas" = Table.Skip(#"Coluna de Tabela Expandida1",7),

#"Colunas Removidas" = Table.RemoveColumns(#"Linhas Principais Removidas",{"Nome da Origem"}),

#"Cabeçalhos Promovidos" = Table.PromoteHeaders(#"Colunas Removidas", [PromoteAllScalars=true]),

#"Linhas em Branco Removidas" = Table.SelectRows(#"Cabeçalhos Promovidos", each not List.IsEmpty(List.RemoveMatchingItems(Record.FieldValues(\_), {"", null}))),

#"Tipo Alterado" = Table.TransformColumnTypes(#"Linhas em Branco Removidas",{{"CO\_IES", type text}}),

#"Linhas Filtradas1" = Table.SelectRows(#"Tipo Alterado", each not Text.Contains([CO\_IES], "Fonte"))

in

#"Linhas Filtradas1"

# 6. Limpeza e Transformação

Após a captura dos dados, iniciou-se o processo de limpeza e transformação utilizando a ferramenta de edição de consultas do Power BI (Power Query). O primeiro passo foi realizar a limpeza nos dados indesejados do banco de dados, removendo as linhas nulas ou em branco. Logo após, houve a remoção das colunas que não serão utilizadas na construção das visualizações, por exemplo, colunas em que suas linhas estão em branco ou sem informações relevantes. Por fim, as colunas foram alteradas para os tipos em que seus dados representam, ou seja, colunas de texto foram alteradas para o tipo texto e colunas numéricas foram alteradas para o tipo numérico.

Nesta etapa, o arquivo que deu origem aos dados carregados para o Power BI não é alterado, com isso, todas as limpezas e alterações são salvas somente na consulta criada. A consulta criada foi nomeada como “f\_dataset”, sendo essa considerada como a tabela fato, ela que contém todas as informações necessárias para criação das visualizações.

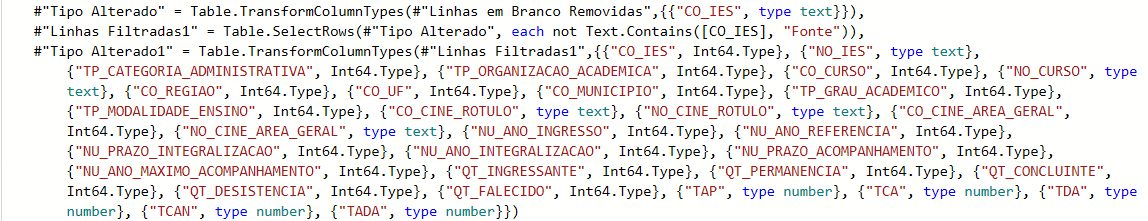
Com a limpeza e transformação dos dados na tabela fato finalizada, inicia-se o processo para criação das tabelas dimensão, que tem a função de auxiliar a tabela fato com dados complementares ou explicativos, e que possuem informações que não se repetem.

Esta etapa consiste em identificar os dados na tabela fato que precisam de informações que estão em tabelas auxiliares, como por exemplo, o código da instituição de ensino ou o código da região em que a instituição de ensino está instalada.

Identificado quais serão as tabelas dimensão necessárias, inicia-se a criação destas tabelas a partir da tabela fato. Abaixo estão as tabelas dimensão:

1. d\_regiao: Tabela dimensão para identificar a região em que a instituição de ensino está instalada na tabela fato (f\_dataset);
2. d\_modalidade\_ensino: Tabela dimensão para identificar qual a modalidade de ensino dos cursos tabela fato (f\_dataset);
3. d\_grau\_academico: Tabela dimensão que indica qual o grau acadêmico dos cursos na tabela fato (f\_dataset);
4. d\_instituicao\_ensino: Tabela dimensão que identifica as informações das instituições de ensino na tabela fato (f\_dataset).

**Código fonte para criação da tabela fato “f\_dataset” em linguagem M.**



**Código fonte para criação da tabela dimensão “d\_regiao” em linguagem M.**

**Texto

Descrição gerada automaticamente**

**Código fonte para criação da tabela dimensão “d\_modalidade\_ensino” em linguagem M.**

Uma imagem contendo Calendário

Descrição gerada automaticamente

**Código fonte para criação da tabela dimensão “d\_grau\_academico” em linguagem M.**

**Texto

Descrição gerada automaticamente com confiança média**

**Código fonte para criação da tabela dimensão “d\_instituicao\_ensino” em linguagem M.**

Uma imagem contendo Texto

Descrição gerada automaticamente

Por fim, realizamos os carregamentos das tabelas fato e dimensão que foram criadas no editor de consulta para o módulo do Power BI em que foram realizadas a relações entre as tabelas dimensão e a tabela fato. As relações entre as tabelas são essenciais para a realização da etapa seguinte, onde serão criadas as medidas e as visualizações dos dados.

Foi utilizado o conceito de relação “star schema”, onde uma tabela fato é relacionada diretamente as tabelas dimensão, que por sua vez não são mais relacionadas com nenhuma outra tabela, tornando assim os cálculos e filtros interativos do relatório mais performáticos.

**Esquema de relação entre as tabelas dimensão e a tabela fato.**

Interface gráfica do usuário, Diagrama

Descrição gerada automaticamente

Neste momento, finaliza-se a etapa de limpeza e transformação dos dados, que já estão prontos para a construção das visualizações.

**7. Visualização dos Dados**

Durante o trabalho de análise dos dados do fluxo da educação superior no Brasil, o grupo realizou consultas com o objetivo de responder às questões propostas. A seguir, descrevo as consultas feitas, os resultados obtidos e sua interpretação, bem como as formas de visualização utilizadas para cada tipo de consulta:

**Comparação Público/Privado**

Consulta: Comparação da taxa de desistência entre instituições públicas e privadas.

Resultado: A taxa de desistência nas instituições privadas é, em média, de 8 a 10% maior do que nas instituições públicas.

Interpretação: Os dados indicam que há uma diferença significativa na taxa de desistência entre faculdades públicas e privadas, sendo as instituições privadas mais suscetíveis a esse problema. Isso pode ser atribuído a fatores financeiros e pode estar relacionado a capacidade financeira do estudante em arcar com os custos de mensalidade.

**Área com maior desistência**

Consulta: Identificação da área de estudo com a maior taxa de desistência.

Resultado: A área de tecnologia apresenta a maior taxa de desistência em comparação com outras áreas.

Interpretação: A alta taxa de desistência na área de tecnologia indica a existência de desafios complexos que podem levar os alunos a abandonar seus cursos nesse campo. Esses desafios podem incluir a natureza técnica e avançada dos conteúdos, a falta de recursos adequados e as barreiras linguísticas. Esses fatores podem contribuir para a decisão dos alunos de abandonar o curso, buscando outras áreas de estudo que considerem mais acessíveis. A compreensão desses desafios específicos na área de tecnologia é crucial para desenvolver estratégias de apoio, como programas de tutoria, recursos adicionais e oportunidades de estágio, a fim de melhorar a retenção dos estudantes nesse campo e promover o sucesso acadêmico.

**EAD x Presencial**

Consulta: Comparação da taxa de desistência entre os formatos de ensino EAD (Ensino a Distância) e presencial.

Resultado: Em média, o ensino a distância (EAD) apresenta uma taxa de desistência de 51%, enquanto o ensino presencial registra uma taxa de desistência de 42%.

Interpretação: A taxa de desistência ligeiramente mais alta no formato de ensino a distância pode ser atribuída a desafios específicos enfrentados pelos alunos nesse modelo. A falta de interação pessoal com colegas e professores, a necessidade de autogerenciamento e a dificuldade de manter a disciplina em um ambiente não presencial podem influenciar a decisão dos alunos de desistir do curso. Esses desafios destacam a importância de oferecer suporte adicional e implementar estratégias de engajamento para os alunos que optam pelo ensino a distância. É essencial fornecer recursos adequados, oportunidades de interação virtual, tutoria online e acompanhamento próximo para promover a retenção e o sucesso dos alunos nesse formato de ensino.

**Ano com maior taxa de desistência em relação ao ano de ingresso no curso**

Consulta: Identificação do ano com a maior taxa de desistência ao decorrer do curso.

Resultado: O segundo ano registrou a maior taxa de desistência ao longo do período analisado.

Interpretação: O segundo ano, que apresentou a maior taxa de desistência, desperta interesse para investigar as possíveis causas dessa situação. É importante considerar fatores como a complexidade do currículo nesse estágio do curso, desafios específicos enfrentados pelos alunos nesse período de transição acadêmica, como adaptação ao nível de exigência e mudança de disciplinas, além de pressões externas, como questões financeiras ou problemas pessoais, bem como a descoberta de aptidão ou não ao curso escolhido. Compreender as razões por trás da alta taxa de desistência no segundo ano é fundamental para desenvolver estratégias de intervenção e apoio personalizado, visando a melhoria da retenção e do engajamento dos estudantes nesse estágio crucial do processo educacional.

**Taxa por região**

Consulta: Análise da taxa de desistência por região.

Resultado: A região Centro-Oeste lidera a taxa de desistência praticamente em todos os anos analisados.

Interpretação: A identificação da região Centro-Oeste com a taxa de desistência mais elevada indica a necessidade de atenção e intervenção específica nessa área geográfica. A compreensão dos fatores socioeconômicos e educacionais nessa região pode ajudar a desenvolver estratégias para reduzir a taxa de desistência.

**Estados com menor taxa de desistência**

Consulta: Identificação dos estados com menor taxa de desistência.

Resultado: O Maranhão e o Pará apresentam as menores taxas de desistência em comparação com outros estados.

Interpretação: A taxa de desistência mais baixa no Maranhão e Pará pode indicar a presença de programas e políticas eficazes voltados para a retenção de estudantes nesses estados. Essas práticas bem-sucedidas podem servir como exemplos para outras regiões.

**Estado com maior taxa de desistência**

Consulta: Identificação do estado com a maior taxa de desistência ao longo dos anos.

Resultado: O estado do Mato Grosso do Sul registrou a maior taxa de desistência na maior parte dos anos analisados.

Interpretação: A alta taxa de desistência no Mato Grosso do Sul requer uma investigação mais aprofundada das causas subjacentes. A falta de recursos educacionais, desafios socioeconômicos ou outros fatores podem estar contribuindo para essa situação.

**Área com menor taxa de desistência**

Consulta: Identificação das áreas de estudo com menor taxa de desistência.

Resultado: As áreas de saúde e bem-estar, assim como agricultura, silvicultura e pesca, apresentam as menores taxas de desistência.

Interpretação: O fato de as áreas de saúde e bem-estar, bem como agricultura, silvicultura e pesca, apresentarem as menores taxas de desistência indica que os alunos nessas áreas estão mais comprometidos em completar seus cursos. Essas áreas podem oferecer um maior engajamento, perspectivas de carreira ou maior satisfação pessoal aos alunos, incentivando-os a continuar.

Para cada tipo de consulta realizada, o grupo utilizou diferentes formas de visualização entre os anos, como tabelas e gráficos diversos, para representar os dados de maneira clara e compreensível.

Essas visualizações são fundamentais para ajudar na interpretação dos resultados e na comunicação das principais conclusões obtidas durante a análise dos dados educacionais. Abaixo segue as imagens destes gráficos:

**Imagem 2: Gráficos 2017 exemplo**

Interface gráfica do usuário, Aplicativo

Descrição gerada automaticamente

**Imagem 3: Gráficos 2017 exemplo**

Interface gráfica do usuário

Descrição gerada automaticamente

**Imagem 4: Gráficos 2017 exemplo**

Interface gráfica do usuário, Aplicativo, Tabela

Descrição gerada automaticamente

**Imagem 5: Gráficos 2017 exemplo**

Interface gráfica do usuário

Descrição gerada automaticamente

Referências Bibliográficas

GISI, Maria Lourdes. A educação superior no Brasil e o caráter de desigualdade do acesso e da permanência.

**Rev. Diálogo Educ.**, Curitiba, v. 06, n. 17, p. 97-112, abr.  2006. Disponível em <http://educa.fcc.org.br/scielo.php?script=sci\_arttext&pid=S1981-

416X2006000100008&lng=pt&nrm=iso>. acessos em 28 mar.  2023.

INEP – INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA.

**Indicadores de Fluxo da Educação Superior.** INEP, 2020. Disponível em: [https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/indicadoreshttps://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/indicadores-educacionais/indicadores-de-fluxo-da-educacao-superioreducacionais/indicadores-de-fluxo-da-educacao-superior.](https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/indicadores-educacionais/indicadores-de-fluxo-da-educacao-superior) Acesso em: 22 mar. 2023. de Oliveira Barbosa, M. L. (2021). *A expansão desigual do ensino superior no Brasil*. Editora Appris.