

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS**NÚCLEO DE EDUCAÇÃO A DISTÂNCIA

Pós-graduação *Lato Sensu* em Analytics e Business Intelligence

**RELATÓRIO TÉCNCO**

ANÁLISE DO INVESTIMENTO BRASILEIRO EM BOLSAS DE ESTUDO DE 2010 A 2020

Vivian Roque Aguiar

Belo Horizonte

2022

**LISTA DE TABELAS**

[Tabela 1 - Tabela Fato Bolsas 6](#_Toc93237024)

[Tabela 2 - Tabela Dimensão Anos 6](#_Toc93237025)

[Tabela 3 - Tabela Dimensão Universidade 6](#_Toc93237026)

[Tabela 4 - Tabela Dimensão Município 6](#_Toc93237027)

[Tabela 5 - Tabela Dimensão Programa 6](#_Toc93237028)

[Tabela 6 - Tabela Dimensão Grande Área do Conhecimento 6](#_Toc93237029)

[Tabela 7 - Tabela Dimensão Área do Conhecimento 7](#_Toc93237030)

**LISTA DE FIGURAS**

[Figura 1 - Modelo Dimensional 8](#_Toc105939273)

[Figura 2 - Dataset 11](#_Toc105939274)

[Figura 3 - União das Tabelas 11](#_Toc105939275)

[Figura 4 - Escolha das Colunas 12](#_Toc105939276)

[Figura 5 - Remoção das Duplicadas 12](#_Toc105939277)

[Figura 6 - Dados da Tabela "GRANDE ÁREA DO CONHECIMENTO" 13](#_Toc105939278)

[Figura 7 - Painel Estratégico 1 15](#_Toc105939279)

[Figura 8 - Painel Estratégico 2 16](#_Toc105939280)

[Figura 9 - Painel Analítico 1 17](#_Toc105939281)

[Figura 10 - Painel Analítico 2 18](#_Toc105939282)

[Figura 11 - Métricas para Previsão 18](#_Toc105939283)

[Figura 12 - Previsão da Quantidade de Bolsas x Ano 19](#_Toc105939284)

[Figura 13 - Previsão de Bolsas para Ciências da Saúde 19](#_Toc105939285)

[Figura 14 - Quantidade total de bolsas concedidas 20](#_Toc105939286)

[Figura 15 - Quantidade total de bolsas concedidas para o Doutorado 20](#_Toc105939287)

[Figura 16 - Quantidade total de bolsas concedidas para o Mestrado 21](#_Toc105939288)

[Figura 17 - Quantidade total de bolsas concedidas para o Pós-Doutorado 21](#_Toc105939289)

[Figura 18 - Quantidade total de bolsas concedidas para Outros tipos de especialização 22](#_Toc105939290)

[Figura 19 - Quantidade total de bolsas concedidas para Ciências Agrárias 22](#_Toc105939291)

[Figura 20 - Quantidade total de bolsas concedidas para Ciências da Saúde 23](#_Toc105939292)

[Figura 21 - Quantidade total de bolsas concedidas para Ciências Exatas e da Terra 23](#_Toc105939293)

[Figura 22 - Quantidade total de bolsas concedidas para Ciências Humanas 24](#_Toc105939294)

[Figura 23 - Quantidade total de bolsas concedidas para Engenharias 24](#_Toc105939295)

[Figura 24 - Quantidade total de bolsas concedidas para Multidisciplinar 25](#_Toc105939296)

[Figura 25 - Quantidade total de bolsas concedidas para Ciências Sociais Aplicadas 25](#_Toc105939297)

[Figura 26 - Quantidade total de bolsas concedidas para Linguística, Letras e Artes 26](#_Toc105939298)

[Figura 27 - Quantidade total de bolsas concedidas para Pró-Reitoria 26](#_Toc105939299)

[Figura 28 - Quantidade total de bolsas concedidas para Grande Área Não Informada 27](#_Toc105939300)

**SUMÁRIO**

[1. Introdução 6](#_Toc105661706)

[1.1. Contexto 6](#_Toc105661707)

[1.2. Objetivos 6](#_Toc105661708)

[1.2.1 Objetivos Gerais 6](#_Toc105661709)

[1.2.2 Objetivos Específicos 6](#_Toc105661710)

[1.3. Público alvo 7](#_Toc105661711)

[2. Modelo de Dados 7](#_Toc105661712)

[2.1. Modelo Dimensional 7](#_Toc105661713)

[2.2. Fatos e Dimensões 8](#_Toc105661714)

[3. Integração, Tratamento e Carga de Dados 9](#_Toc105661715)

[3.1. Fontes de Dados 9](#_Toc105661716)

[3.2. Processos de Integração e Carga (ETL) 9](#_Toc105661717)

[4. Camada de Apresentação 12](#_Toc105661718)

[4.1 Métricas 12](#_Toc105661719)

[4.2 Dashboard 13](#_Toc105661720)

[4.2.1 Painel Estratégico 13](#_Toc105661721)

[4.2.2 Painel Analítico 15](#_Toc105661722)

[4.3 Análises avançadas 17](#_Toc105661723)

[5. Registros de Homologação 17](#_Toc105661724)

[6. Conclusões 24](#_Toc105661725)

[6.1 Pontos Relevantes 24](#_Toc105661726)

[6.2 Proposta de Intervenção 25](#_Toc105661727)

[6.3 Lições Aprendidas e Pontos de Extensão 25](#_Toc105661728)

[7. Links 25](#_Toc105661729)

[REFERÊNCIAS 26](#_Toc105661730)

# 1. Introdução

## 1.1. Contexto

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) é uma organização responsável pela expansão e consolidação da pós-graduação em todo o país. Ela é responsável por fornecer grande parte das bolsas de pesquisa acadêmica.

Desde 2020, ano em que teve início, à nível global, a pandemia do Coronavírus, o mundo sentiu na pele como é importante o incentivo em ciência e pesquisa. Não por menos, um ínfimo número de países teve estrutura para pesquisar, testar e desenvolver a vacina que todos ansiavam.

No Brasil, o que vimos foi uma corrida na contramão do que se via no mundo. Não era difícil ler ou ouvir notícias que deixam claro essa visão. “Em meio à pandemia, Capes amplia o corte de bolsas de pesquisa. (UFPEL, 2020)”

Levando em consideração essa constatação, esse trabalho procurou fazer um levantamento e análise do investimento brasileiro em bolsas de estudos nas instituições de ensino superior apoiadas pela CAPES desde 2010 a 2020.

## 1.2. Objetivos

### 1.2.1 Objetivos Gerais

O objetivo geral da pesquisa é analisar o investimento do Governo em bolsas de estudos concedidas a instituições de ensino superior apoiadas pela Capes entre os anos de 2010 a 2020.

### 1.2.2 Objetivos Específicos

* Comparar a quantidade de bolsas concedidas em cada ano.
* Identificar as áreas do conhecimento que recebem maior investimento.
* Analisar o investimento nas regiões, estados e municípios do país.

## 1.3. Público alvo

O público alvo desse projeto é o Ministério da Educação, instituições de ensino superior e alunos interessados em ingressar em um dos programas.

# 2. Modelo de Dados

## 2.1. Modelo Dimensional

O Modelo Dimensional será criado a partir de planilhas do Excel. No total foram 11 planilhas, uma para cada ano (2010-2020), contendo 16 campos que serão listados a seguir:

AN\_REFERENCIA / NM\_MUNICIPIO\_IES\_CAPES / SG\_UF\_IES\_CAPES / NM\_REGIAO\_IES\_CAPES / CD\_IES\_CAPES / NM\_IES\_CAPES / SG\_IES\_CAPES / NM\_NATUREZA\_JURIDICA / SG\_PROGRAMA\_FOMENTO / CD\_PROGRAMA / NM\_PROGRAMA / NM\_GRANDE\_AREA / NM\_AREA\_AVALIACAO / NM\_AREA\_CONNHECIMENTO / NM\_NIVEL\_PADRAO / QT\_BOLSAS\_CONCEDIDAS

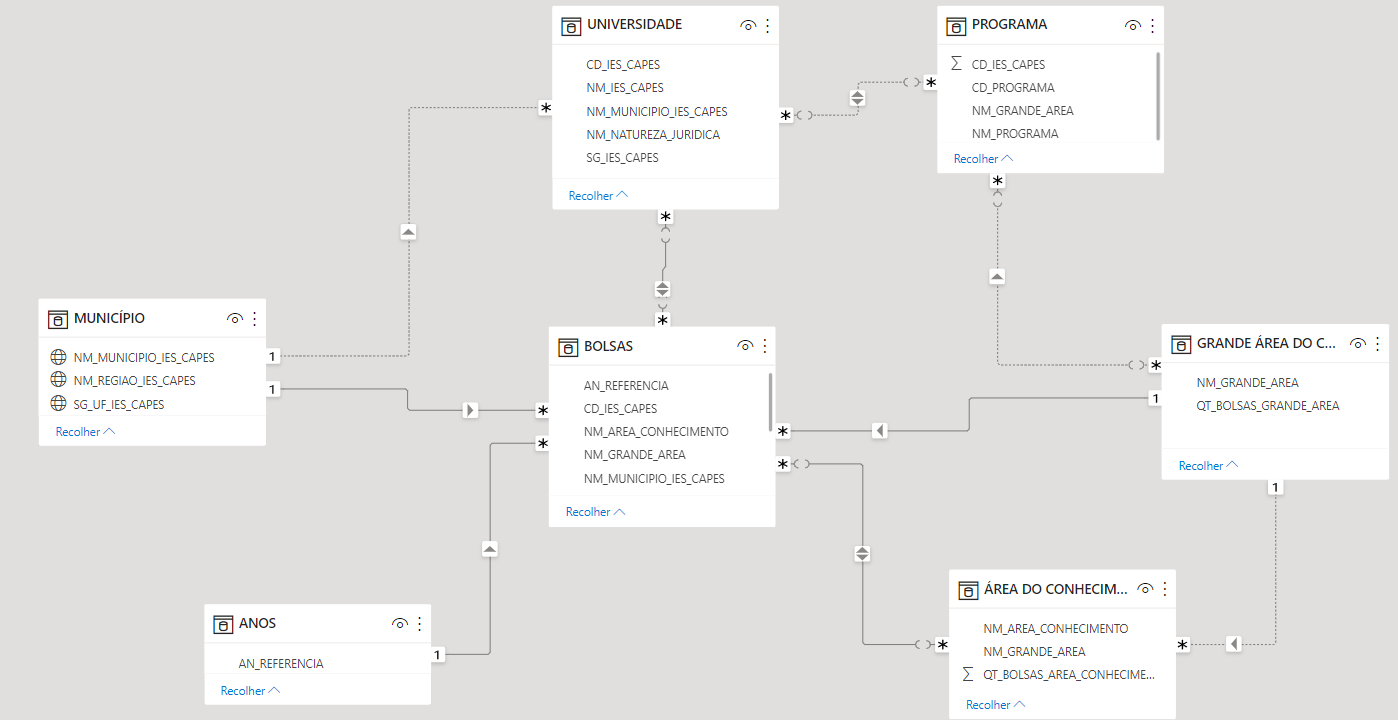


Figura 1 - Modelo Dimensional

## 2.2. Fatos e Dimensões

No total foram utilizadas uma tabela fato e seis tabelas dimensão.

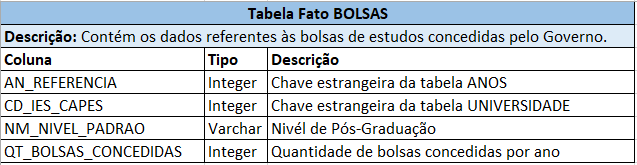


Tabela 1 - Tabela Fato Bolsas

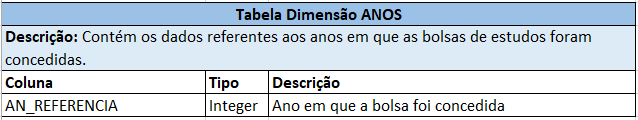


Tabela 2 - Tabela Dimensão Anos

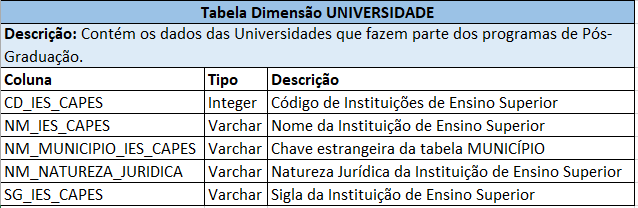


Tabela 3 - Tabela Dimensão Universidade



Tabela 4 - Tabela Dimensão Município

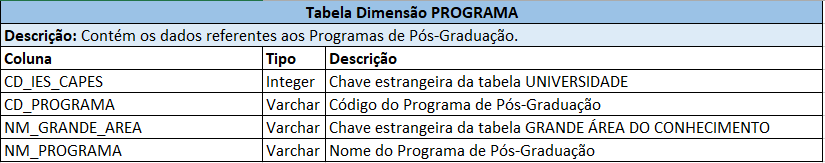


Tabela 5 - Tabela Dimensão Programa

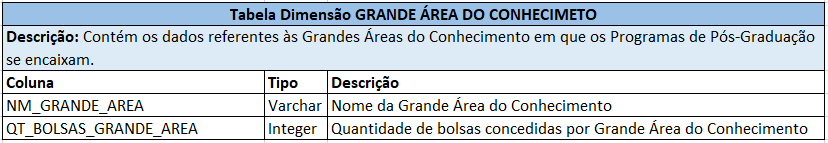


Tabela 6 - Tabela Dimensão Grande Área do Conhecimento

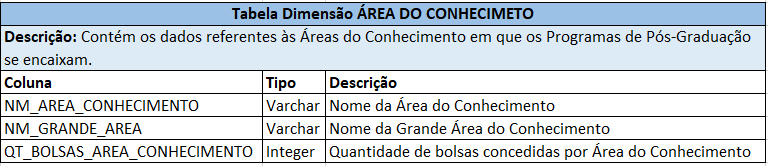


Tabela 7 - Tabela Dimensão Área do Conhecimento

# 3. Integração, Tratamento e Carga de Dados

## 3.1. Fontes de Dados

Os dados foram extraídos diretamente do Banco de Dados da CAPES, cujo endereço eletrônico se encontra em: *https://dadosabertos.capes.gov.br*

## 3.2. Processos de Integração e Carga (ETL)

O *dataset* do projeto foi obtido em formato *xls.* As onze tabelas contendo as bolsas concedidas de 2010 a 2020 foram unidas em uma única tabela. Em seguida, os dados foram importados para o *software* *Microsoft Power BI.*

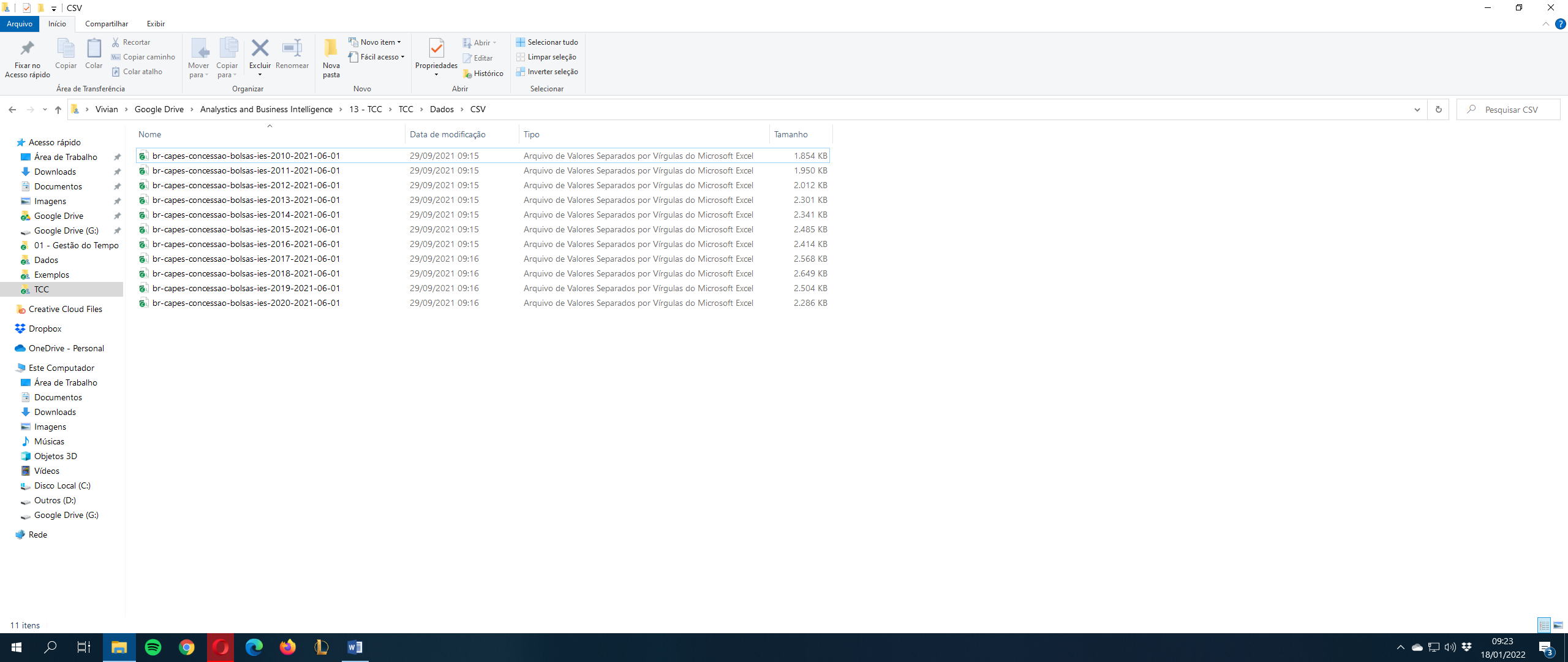


Figura 2 - Dataset

O *software* *Power Query* foi utilizado na transformação/limpeza dos dados para carregar somente o que fosse importante para o projeto. Utilizando a função “Acrescentar consultas como novas”, as onze tabelas foram unidas em uma única contendo todos os dados.

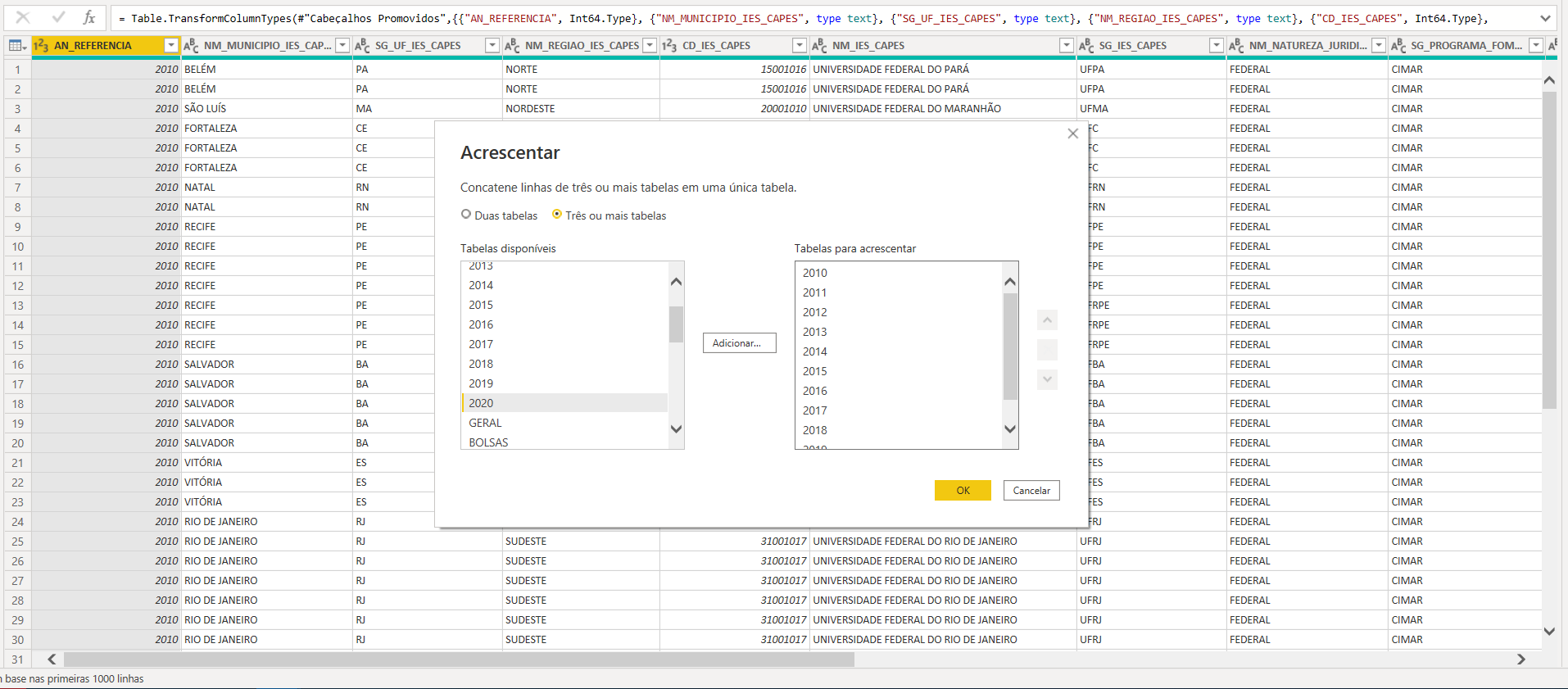
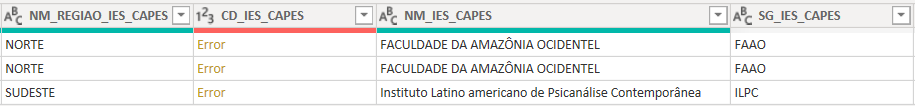


Figura 3 - União das Tabelas

No total havia três dados com erro na coluna “CD\_IES\_CAPES”. Dessa forma, foi necessário trata-lo alterando o “Error” pelo número “0”.



O processo utilizado para separar e criar as tabelas foi realizado duplicando a tabela “Geral” e escolhendo as colunas que seriam utilizadas para cada uma das tabelas. Dessa forma, foi possível separá-las em fatos e dimensões.

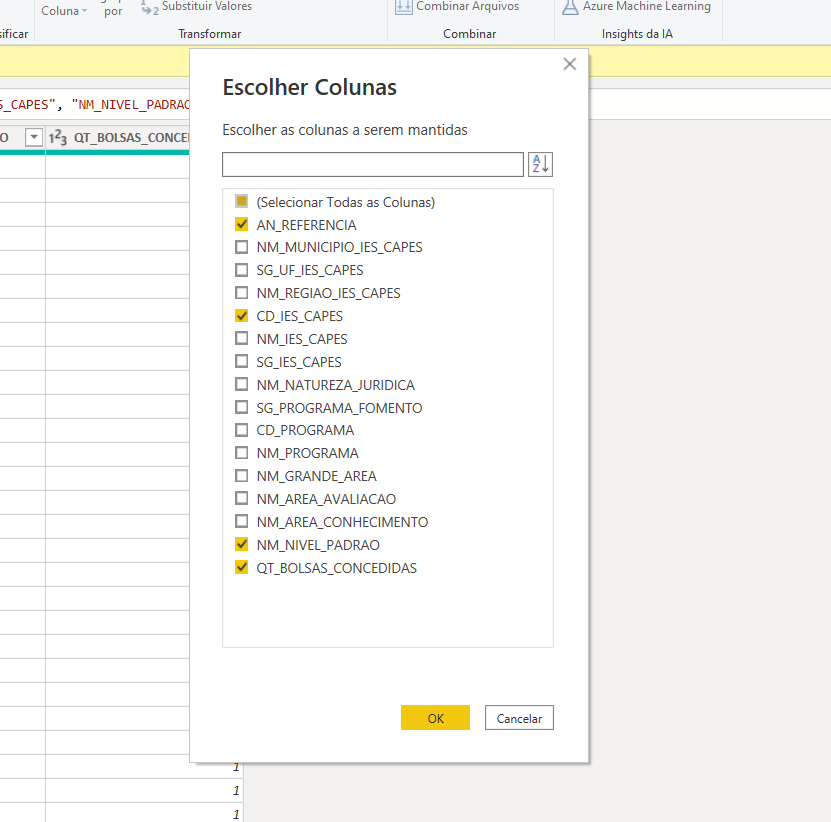


Figura 4 - Escolha das Colunas

Para limpar as tabelas e reduzir o volume de dados, foi feito o processo de “Remover Duplicadas” das tabelas.

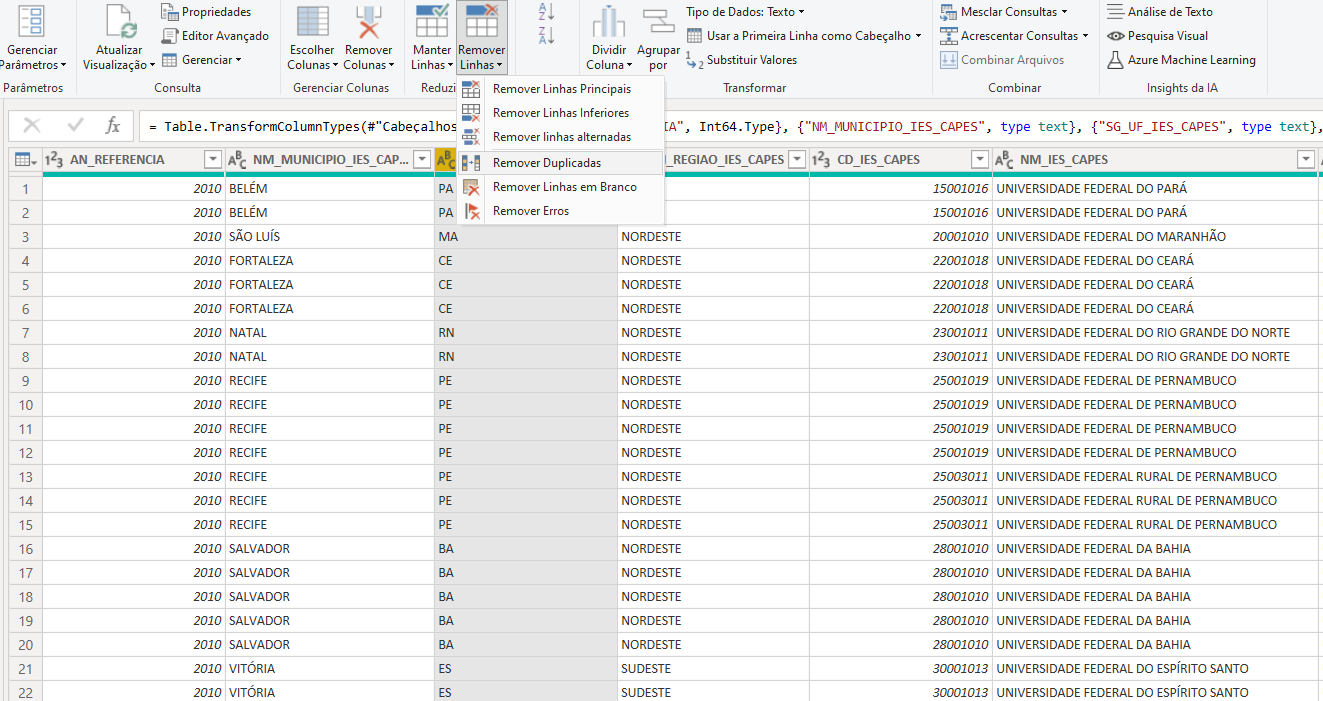


Figura 5 - Remoção das Duplicadas

Na tabela “GRANDE ÁREA DO CONHECIMENTO”, foi realizado o agrupamento dos dados da coluna “QT\_BOLSAS\_CONCEDIDAS” para a coluna “QT\_BOLSAS\_GRANDE\_AREA” afim de mostrar a quantidade de bolsas de cada Área do Conhecimento separadamente.



Figura 6 - Dados da Tabela "GRANDE ÁREA DO CONHECIMENTO"

Por fim, foram criados os relacionamentos para o Modelo Dimensional.

# 4. Camada de Apresentação

Foi utilizado o *KNIME* para o tratamento dos dados no formato *xls.* Em seguida, os dados foram convertidos para o formato *csv* e importados para o banco de dados criado no *PostgreSQL.* Por fim, os *Dashboards* foram criados no *Microsoft Power BI.*

## 4.1 Métricas

As métricas criadas nos painéis foram elaboradas para atender às demandas do público alvo ao qual o projeto está sendo realizado. A seguir serão listadas as métricas e suas respectivas descrições:

1. **Bolsas:** quantidade distinta de bolsas concedidas pelo Governo.
2. **Anos:** quantidade distinta de bolsas concedidas pelo Governo entre os anos de 2010 e 2020.
3. **Bolsas por Nível de Pós-Graduação:** quantidade distinta de bolsas concedidas pelo Governo em cada Nível de Pós-Graduação.
4. **Bolsas por Natureza Jurídica:** quantidade distinta de bolsas concedidas pelo Governo em cada Natureza Jurídica.
5. **Bolsas por Instituição de Ensino:** quantidade distinta de bolsas concedidas pelo Governo em cada Instituição de Ensino.
6. **Bolsas por Grande Área do Conhecimento:** quantidade distinta de bolsas concedidas pelo Governo em cada Grande Área do Conhecimento.
7. **Bolsas por Área do Conhecimento:** quantidade distinta de bolsas concedidas pelo Governo em cada Área do Conhecimento.
8. **Bolsas por Região:** quantidade distinta de bolsas concedidas pelo Governo em cada Região do Brasil.
9. **Bolsas por Unidade Federativa:** quantidade distinta de bolsas concedidas pelo Governo em cada Unidade Federativa do Brasil.
10. **Bolsas por Município:** quantidade distinta de bolsas concedidas pelo Governo em cada Município do Brasil.

## 4.2 Dashboard

Como mencionado anteriormente, os *Dashboards* foram criados no *Microsoft Power BI.* Pare este trabalho, foram elaborados os Painéis Estratégico e Analítico que serão explicados a seguir.

### 4.2.1 Painel Estratégico

O primeiro Painel Estratégico foi construído com a finalidade de apresentar um apanhado geral sobre a Quantidade de Bolsas Concedidas por Nível de Pós-Graduação (Mestrado, Doutorado, Pós-Doutorado e Outros), por Ano (2010 a 2020), por Natureza Jurídica (Federal, Estadual, Particular e Municipal) e por Instituição de Ensino (Ex: Universidade Federal do Rio de Janeiro, Universidade de São Paulo, Pontifícia Universidade Católica de Minas Gerais e Universidade Regional de Blumenau).

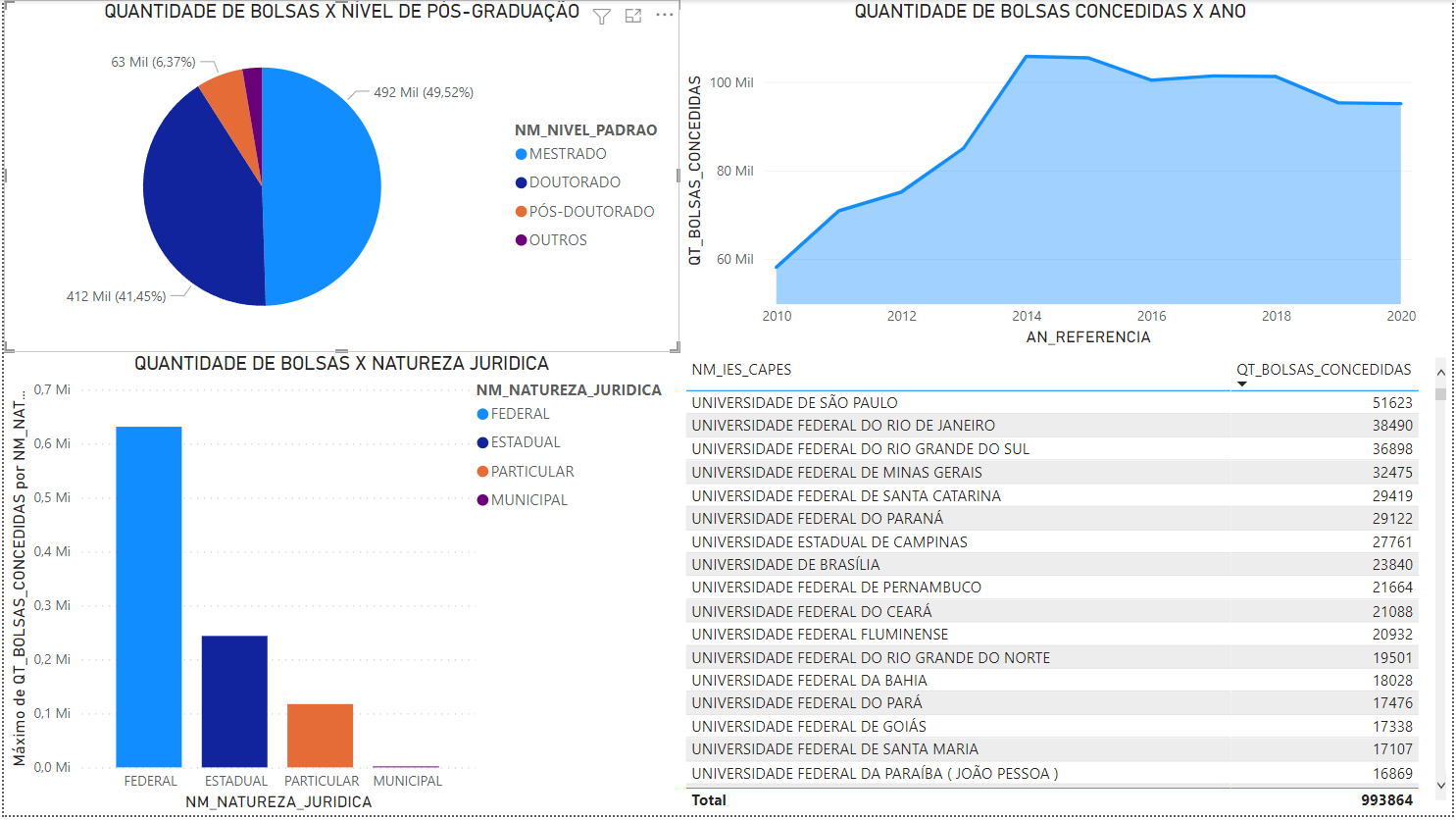


Figura 7 - Painel Estratégico 1

Já o segundo Painel Estratégico, foi construído com a finalidade de apresentar um ranking com as Instituições de Ensino (Ex: Universidade Federal do Rio de Janeiro, Universidade de São Paulo, Pontifícia Universidade Católica de Minas Gerais e Universidade Regional de Blumenau), Grande Área (Ex: Ciências da Saúde, Engenharias e Multidisciplinar) e Área do Conhecimento (Agronomia, Educação, Química e Letras) que possuem a maior quantidade de bolsas concedidas entre 2010 e 2020.

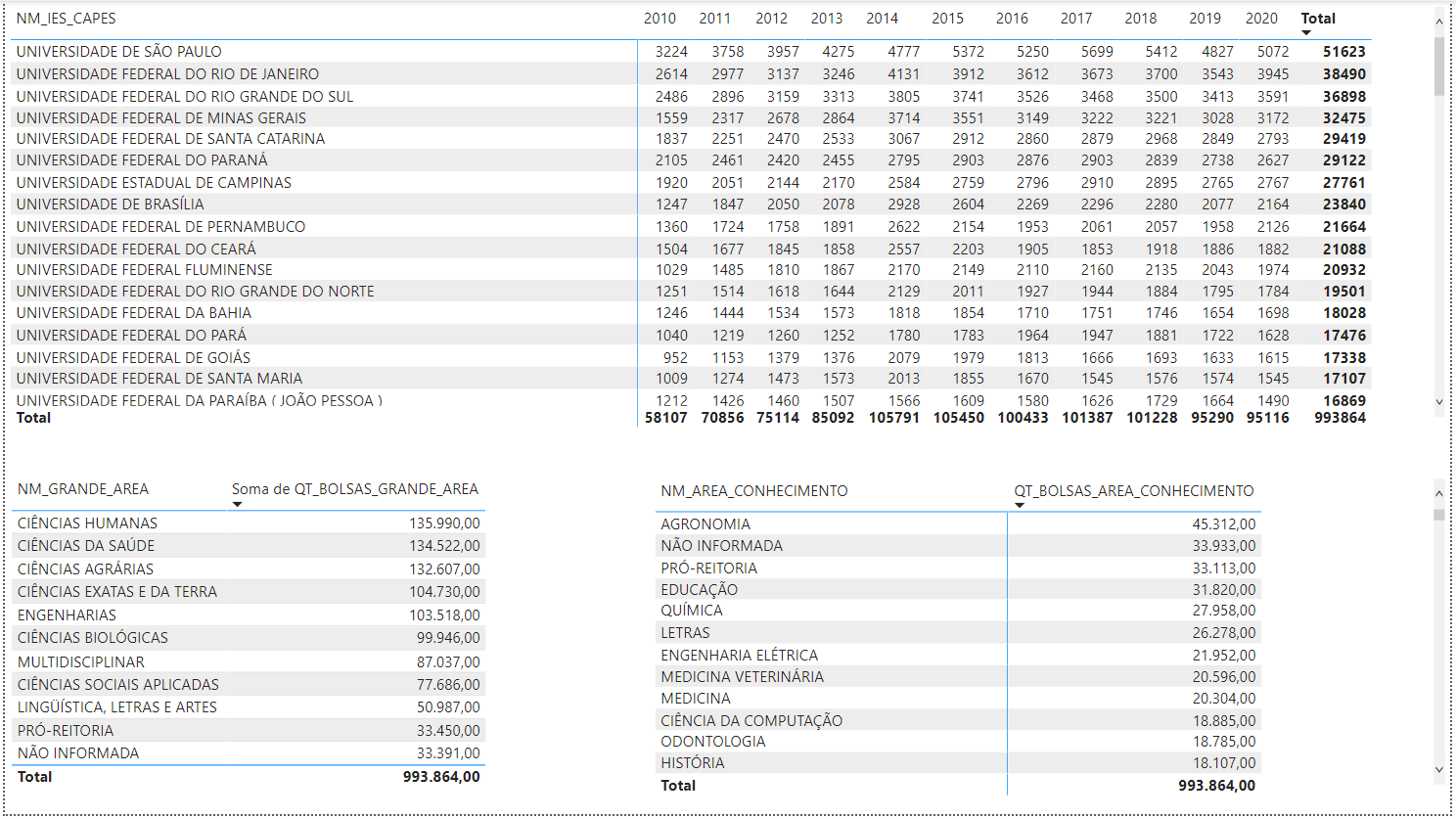


Figura 8 - Painel Estratégico 2

### 4.2.2 Painel Analítico

O primeiro Painel Analítico foi construído com a finalidade de apresentar geograficamente as Regiões e Unidades Federativas com suas respectivas quantidade de bolsas concedidas pelo Governo. Esse painel possui dois filtros: de Região e UF.

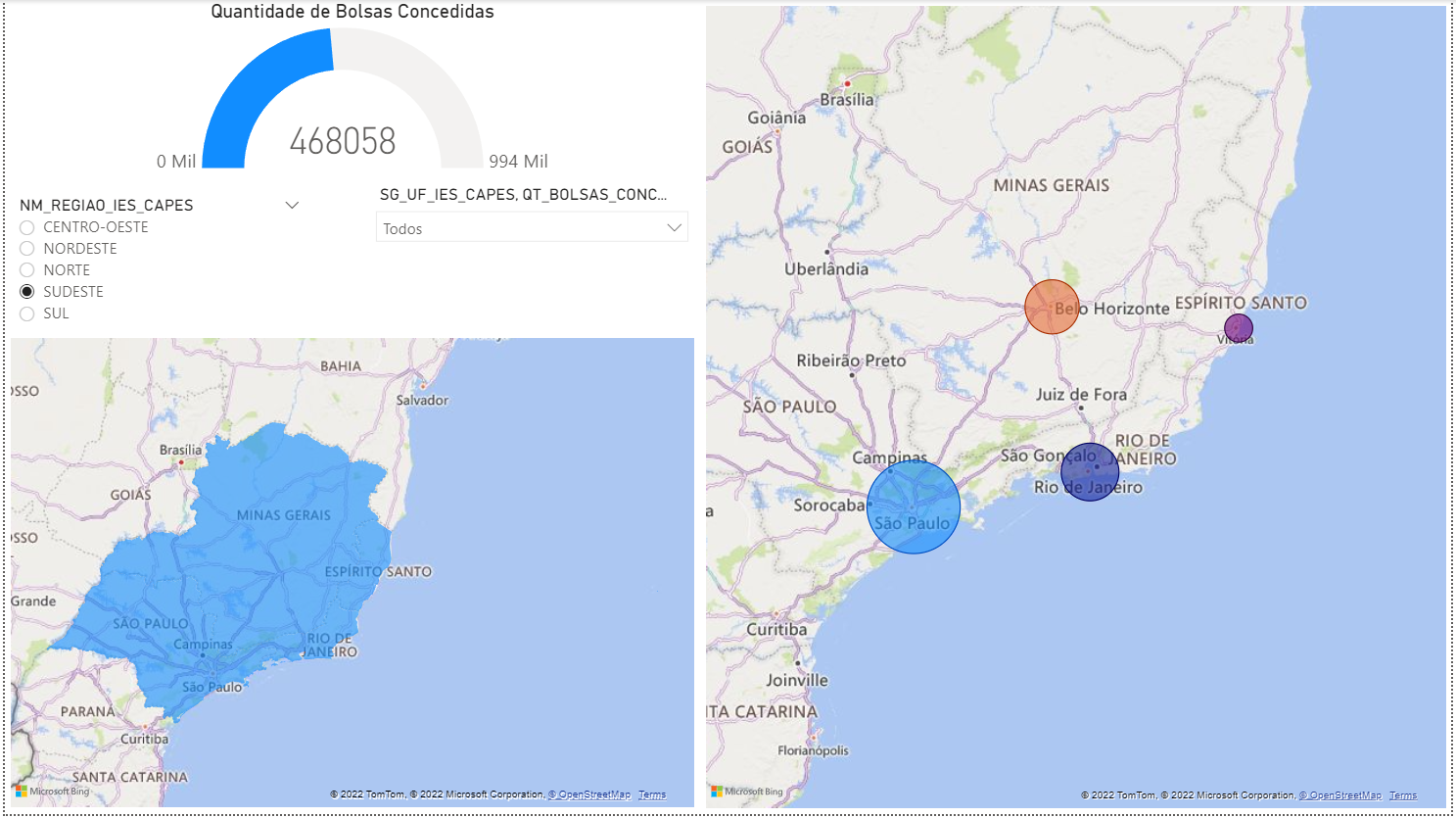


Figura 9 - Painel Analítico 1

O segundo Painel Analítico foi construído com a finalidade de apresentar geograficamente os Municípios e Grande Área do Conhecimento com suas respectivas quantidade de bolsas concedidas pelo Governo entre 2010 e 2020. Esse painel possui dois filtros: de Ano e de Município.

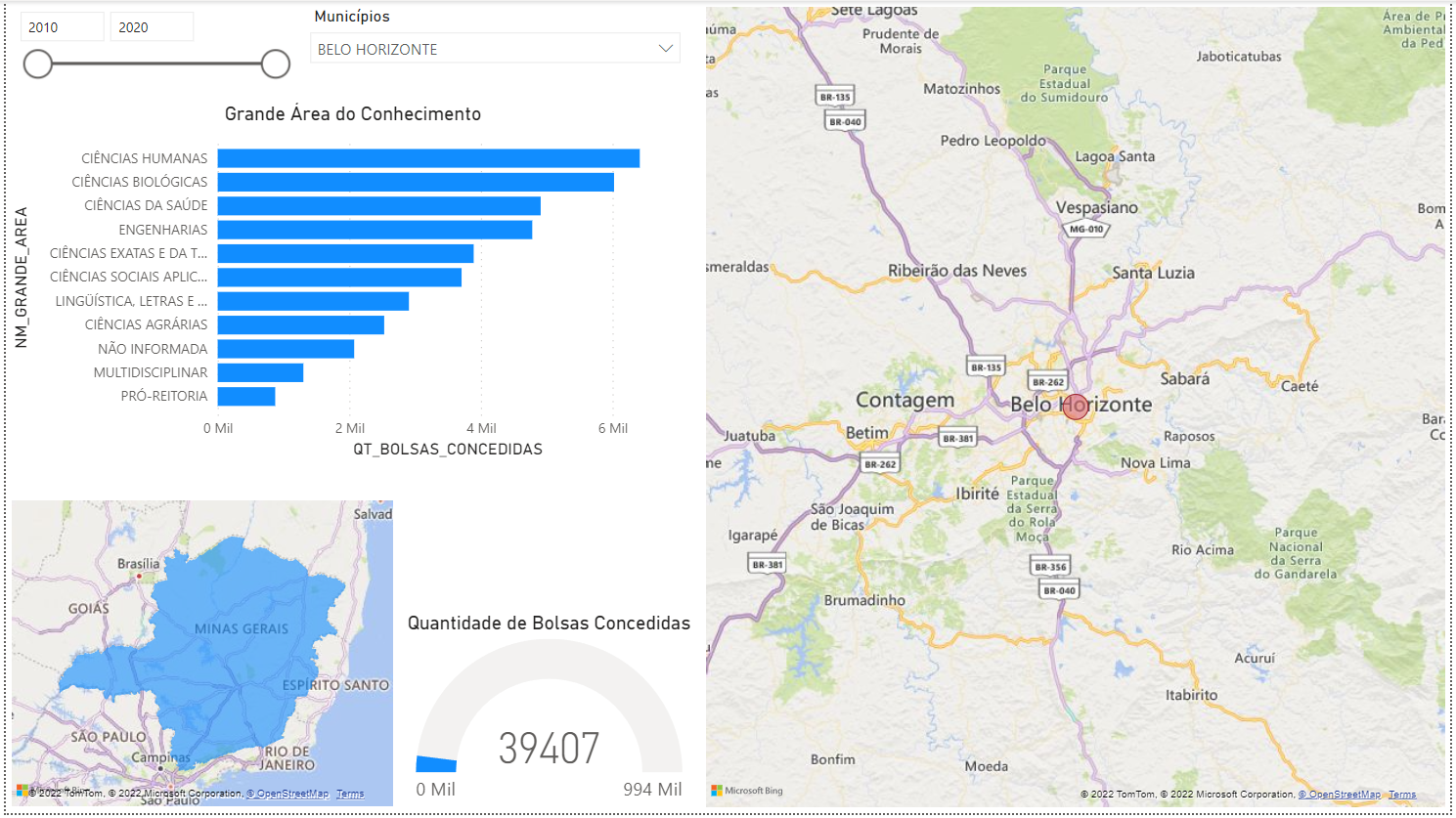


Figura 10 - Painel Analítico 2

## 4.3 Análises avançadas

As análises avançadas foram criadas com o *Microsoft Power BI*. Há uma funcionalidade que apresenta a previsão das métricas com padrões de comportamento.

A primeira previsão foi feita considerando a quantidade de bolsas concedidas nos anos. Foi considerada a métrica até 2034.

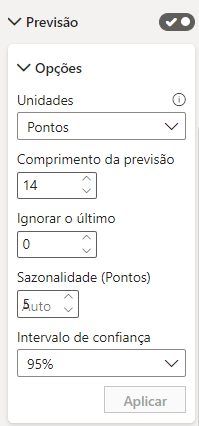


Figura 11 - Métricas para Previsão

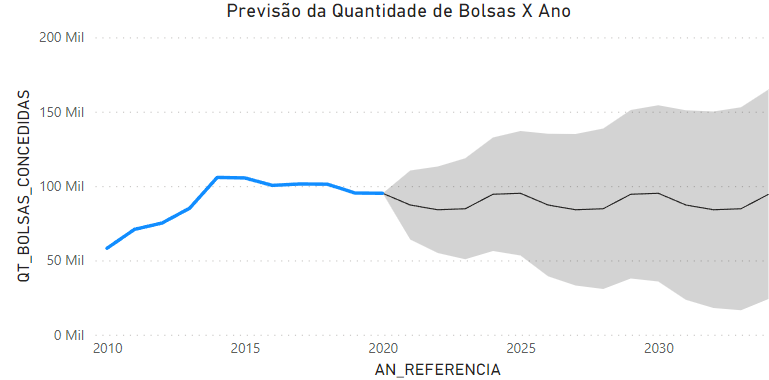


Figura 12 - Previsão da Quantidade de Bolsas x Ano

Tomando como base as mesmas medidas, a segunda previsão foi feita considerando a quantidade de bolsas para a área das Ciências da Saúde.

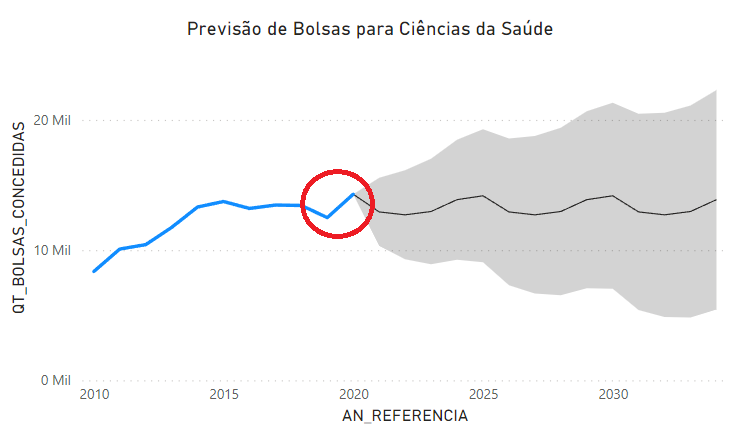


Figura 13 - Previsão de Bolsas para Ciências da Saúde

Foi possível observar, inclusive, o grande salto que essa área teve de 2019 para 2020.

# 5. Registros de Homologação

Nessa etapa de homologação, serão mostrados as consultas de dados feita diretamente do *PostgreSQL.*

1. Quantidade total de bolsas concedidas

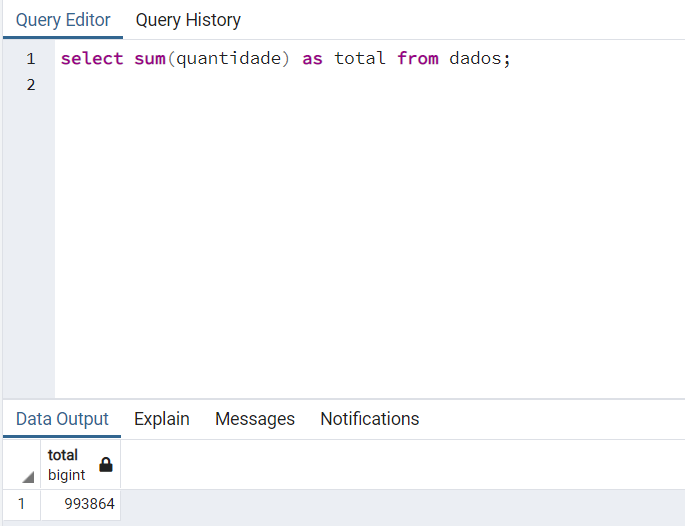


Figura 14 - Quantidade total de bolsas concedidas

1. Quantidade total de bolsas concedidas para o Doutorado

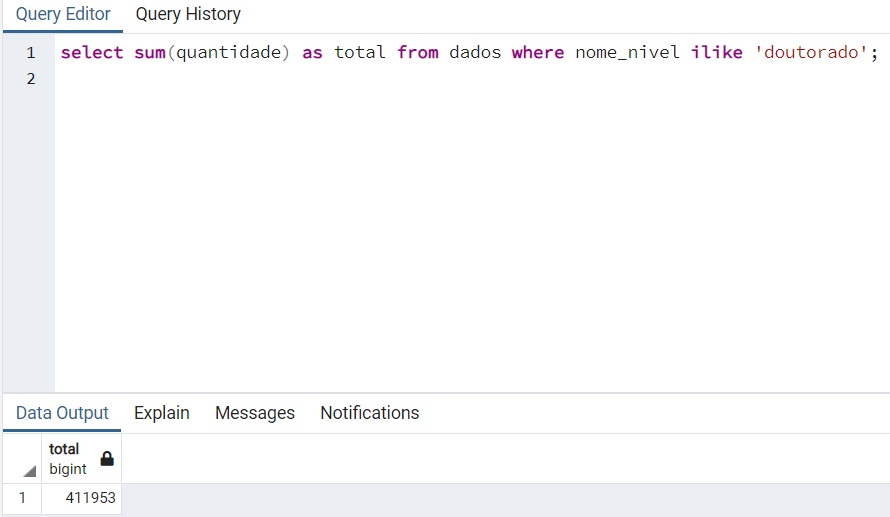


Figura 15 - Quantidade total de bolsas concedidas para o Doutorado

1. Quantidade total de bolsas concedidas para o Mestrado

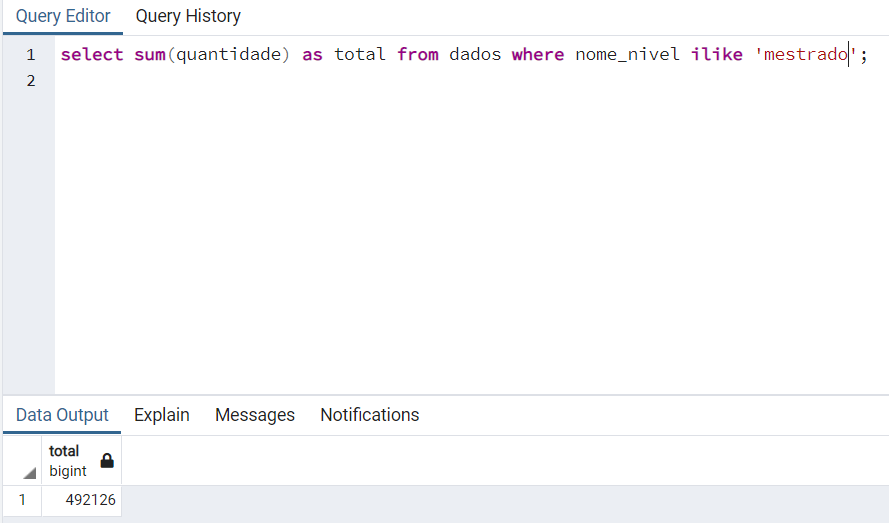


Figura 16 - Quantidade total de bolsas concedidas para o Mestrado

1. Quantidade total de bolsas concedidas para o Pós-Doutorado

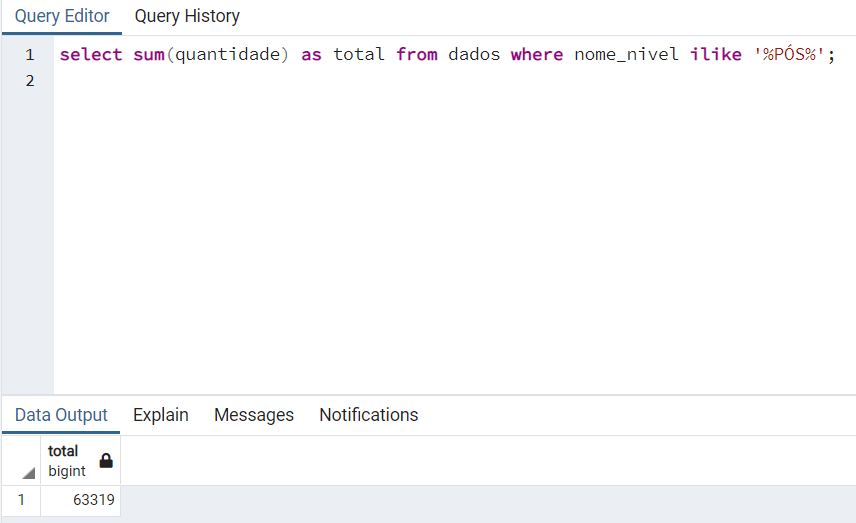


Figura 17 - Quantidade total de bolsas concedidas para o Pós-Doutorado

1. Quantidade total de bolsas concedidas para outros tipos de especialização

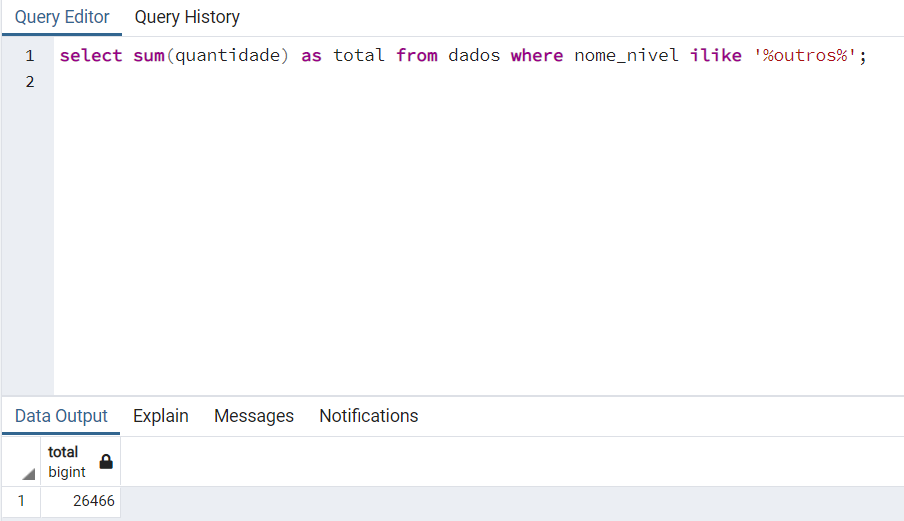


Figura 18 - Quantidade total de bolsas concedidas para Outros tipos de especialização

1. Quantidade total de bolsas concedidas para Ciências Agrárias

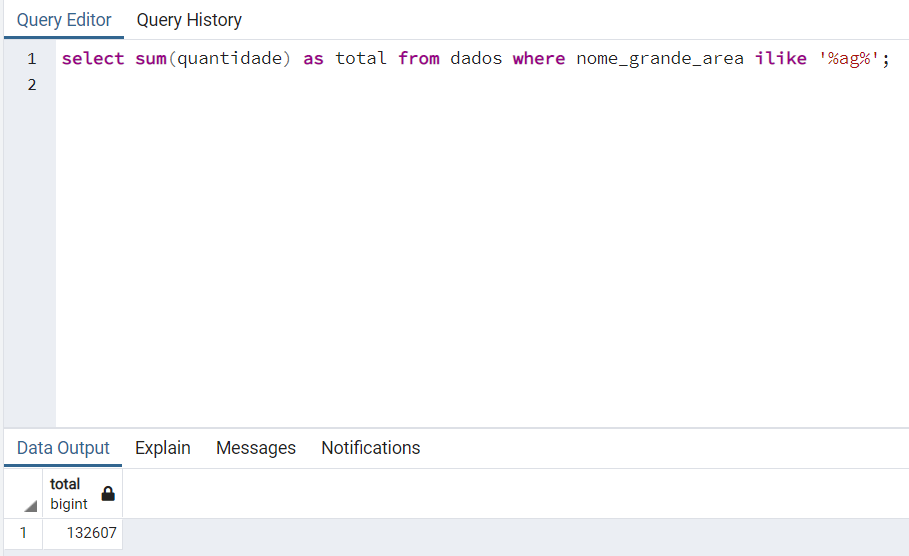


Figura 19 - Quantidade total de bolsas concedidas para Ciências Agrárias

1. Quantidade total de bolsas concedidas para Ciências da Saúde

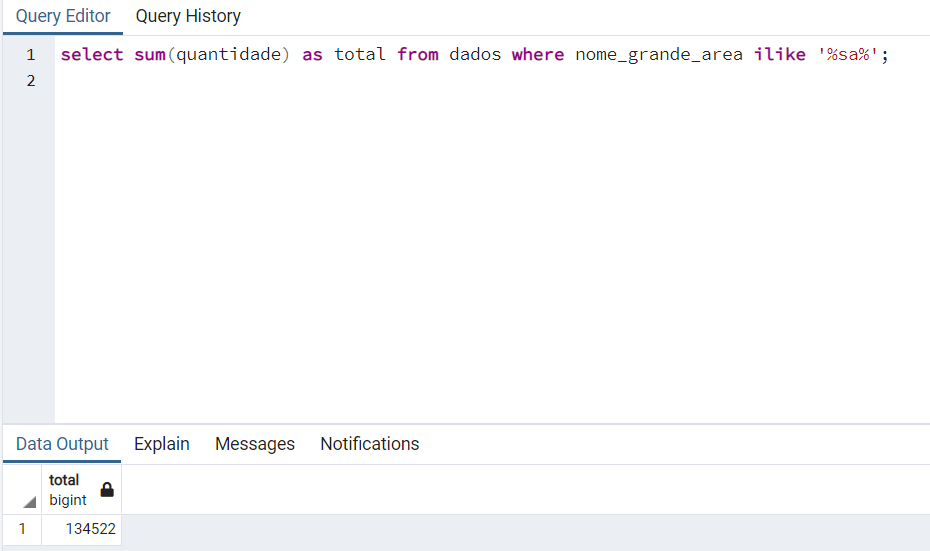


Figura 20 - Quantidade total de bolsas concedidas para Ciências da Saúde

1. Quantidade total de bolsas concedidas para Ciências Exatas e da Terra

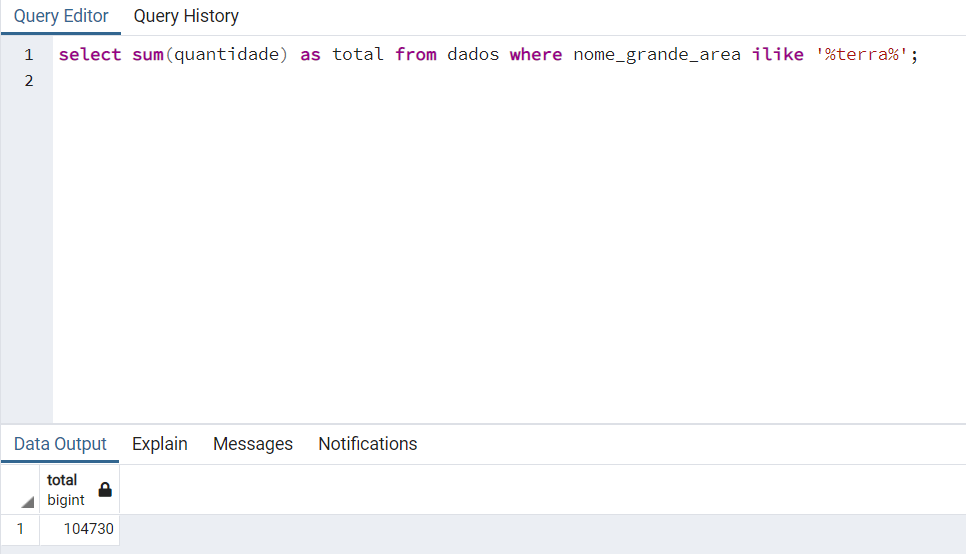


Figura 21 - Quantidade total de bolsas concedidas para Ciências Exatas e da Terra

1. Quantidade total de bolsas concedidas para Ciências Humanas



Figura 22 - Quantidade total de bolsas concedidas para Ciências Humanas

1. Quantidade total de bolsas concedidas para Engenharias

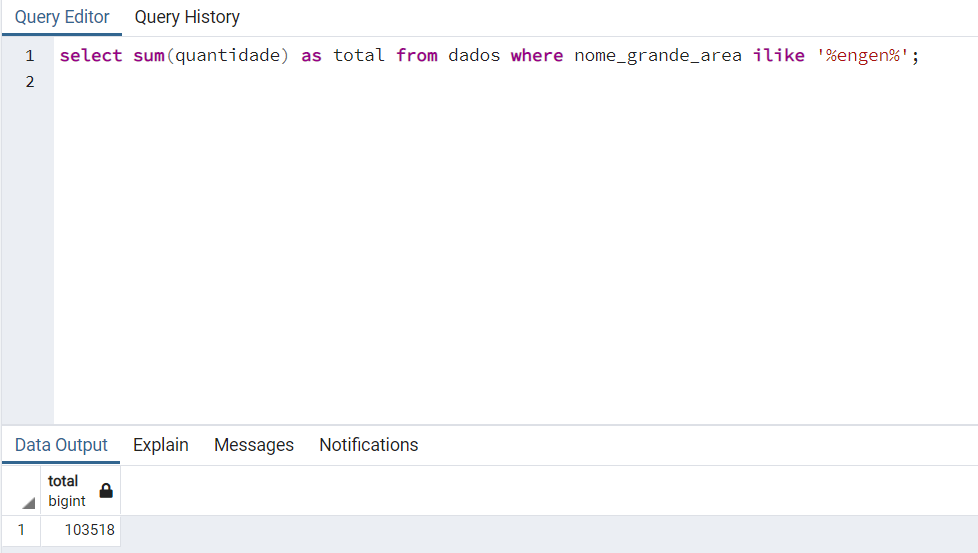


Figura 23 - Quantidade total de bolsas concedidas para Engenharias

1. Quantidade total de bolsas concedidas para Multidisciplinar

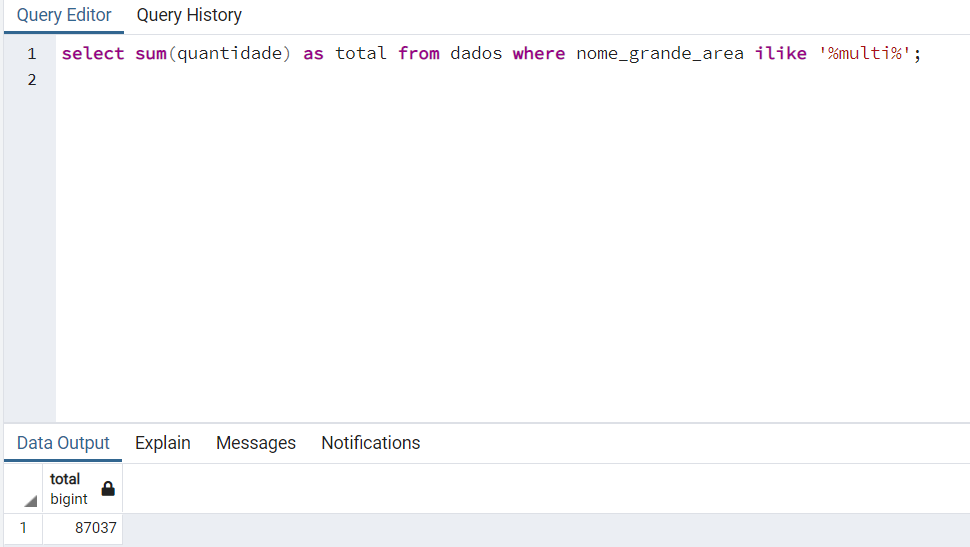


Figura 24 - Quantidade total de bolsas concedidas para Multidisciplinar

1. Quantidade total de bolsas concedidas para Ciências Sociais Aplicadas

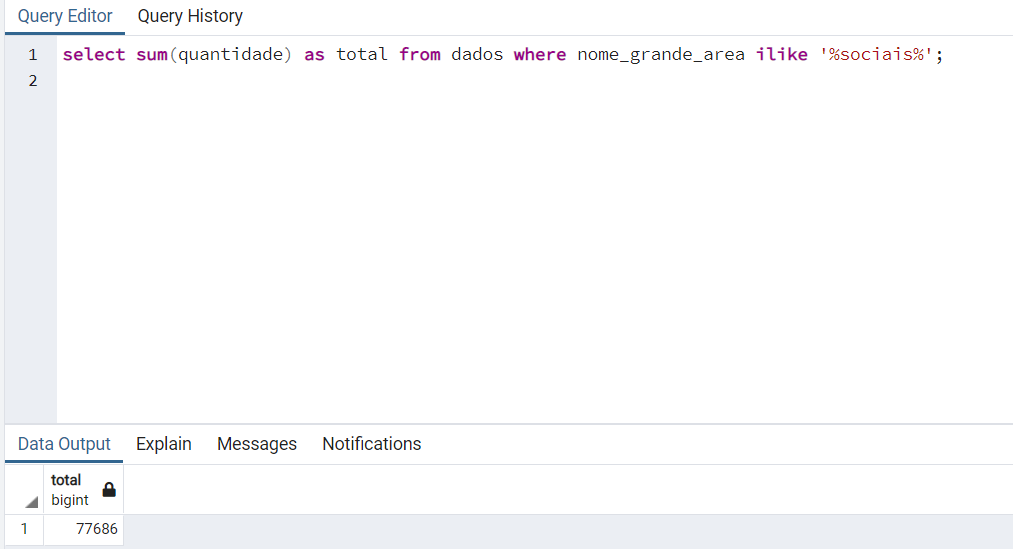


Figura 25 - Quantidade total de bolsas concedidas para Ciências Sociais Aplicadas

1. Quantidade total de bolsas concedidas para Linguística, Letras e Artes



Figura 26 - Quantidade total de bolsas concedidas para Linguística, Letras e Artes

1. Quantidade total de bolsas concedidas para Pró-Reitoria

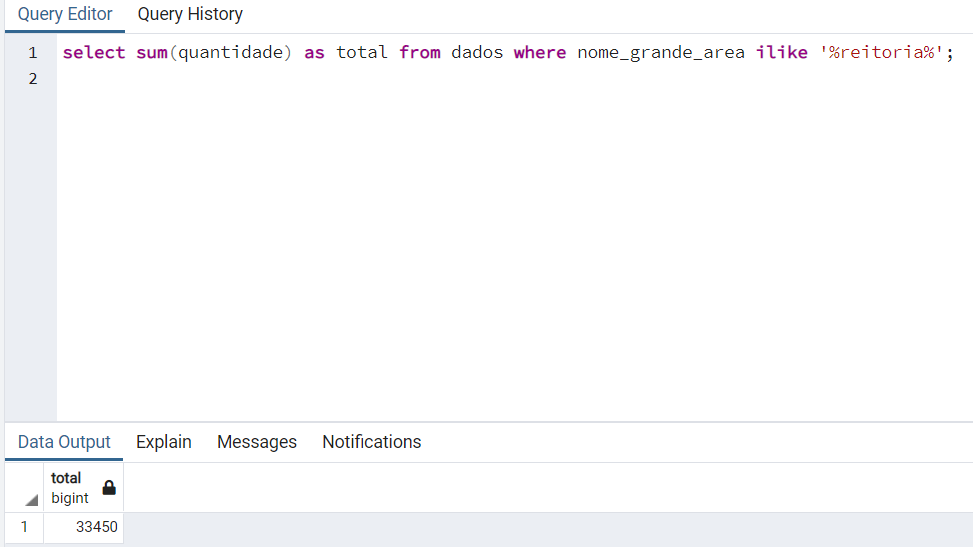


Figura 27 - Quantidade total de bolsas concedidas para Pró-Reitoria

1. Quantidade total de bolsas concedidas para Grande Área do Conhecimento Não Informada

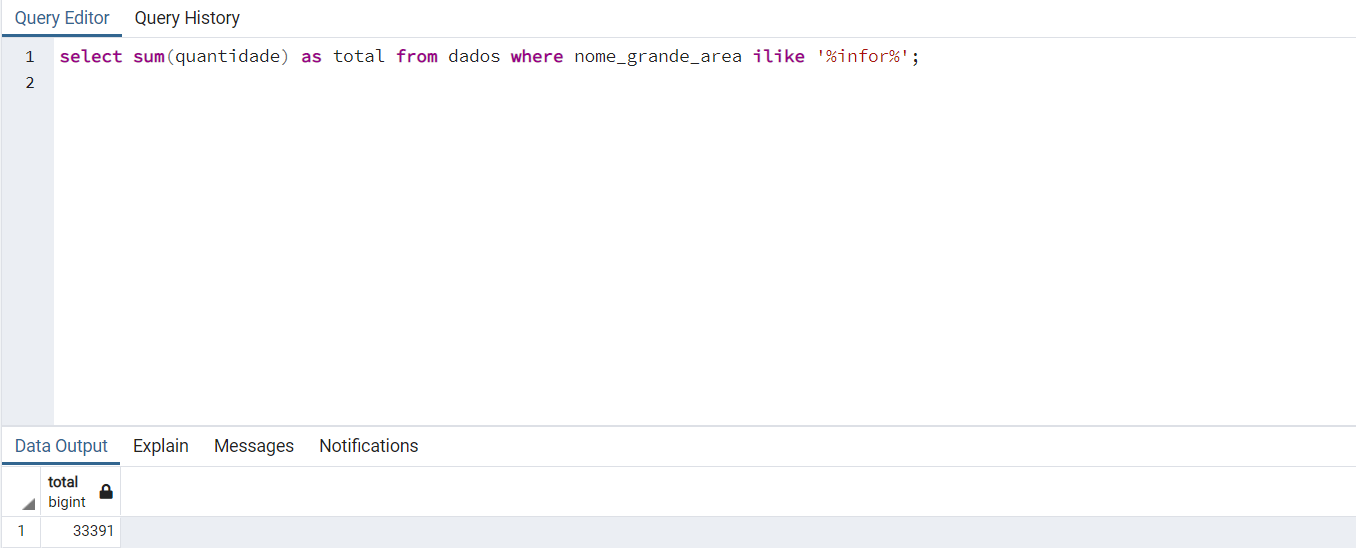


Figura 28 - Quantidade total de bolsas concedidas para Grande Área Não Informada

# 6. Conclusões

## 6.1 Pontos Relevantes

## 

1. Quantidade de bolsas concedidas com o passar dos anos: Importante observar que a quantidade de bolsas concedidas vinha crescendo desde o ano de 2010, porém, a partir de 2017 começa a cair significantemente.
2. Desigualdade entre as regiões do país: Das 993.864 bolsas concedidas desde 2010, 47% delas – 468.058 – foram concedidas à região Sudeste. Sendo São Paulo o Estado que mais recebeu bolsas de estudo, com um total de 236.207 que configura 23% do total.
3. A principal Universidade: Apesar da maioria das bolsas serem concedidas a Instituições Federais, a Universidade de São Paulo (USP), que é Instituição Estadual, foi a universidade que mais recebeu o recurso.
4. País do Agronegócio: Não é de se espantar que o Brasil, sendo um dos cinco maiores países tratando-se de agricultura, tenha a Área do Conhecimento da Agronomia como detentora da maior quantidade de bolsas concedidas nesses 11 anos analisados, com um total de 45.312.

## 6.2 Proposta de Intervenção

Por se tratar de uma análise de dados voltada à Administração Pública, essa intervenção fica nas mãos do Governo. Reforço o que foi dito na introdução deste trabalho. Se há uma coisa que a Covid-19 ensinou a todos nós, foi que a pesquisa e a Ciência salvam vidas.

Dessa forma, é imprescindível que os que estão no poder comecem a enxergar isso e invistam mais na educação. Temos grandes universidades no país, inclusive reconhecidas internacionalmente, o que elas precisam é de investimento para crescerem cada vez mais.

## 6.3 Lições Aprendidas e Pontos de Extensão

1. Dados Incompletos: Apesar da base de dados ter sido obtida diretamente do Governo, há uma enorme quantidade de dados incompletos.
2. Dados Incorretos: Foi observado que algumas universidades que possuem mais de um polo, estavam com o nome da universidade de forma errada. Ex: Universidade Federal Fluminense – Campus Volta Redonda. A base de dados deveria conter um linha específica para o campus, dessa forma não haveria divergência de nome da universidade.

# 7. Links

Todos os dados e arquivos criados para este projeto serão compartilhados utilizando o Google Drive.

**Google Drive:**

[**https://drive.google.com/drive/folders/1-Df3USYOLG2ccIpBKoP7tzvkColGcI0g?usp=sharing**](https://drive.google.com/drive/folders/1-Df3USYOLG2ccIpBKoP7tzvkColGcI0g?usp=sharing)

**Vídeo de Apresentação do Módulo B:**

[**https://www.youtube.com/watch?v=rr7QBP9yTNQ**](https://www.youtube.com/watch?v=rr7QBP9yTNQ)

**Vídeo de Apresentação do Módulo C:**

[**https://www.youtube.com/watch?v=P\_AkcGnHkHs**](https://www.youtube.com/watch?v=P_AkcGnHkHs)

# REFERÊNCIAS

CAPES, 2021. **Brasil – [2010 a 2016] Concessão de Bolsas a Instituições de Ensino Superior apoiadas pela CAPES.** Disponível em: <Brasil - [2010 a 2016] Concessão de Bolsas a Instituições de Ensino Superior apoiadas pela CAPES - Informação geral>. Acessado em: 21 de janeiro de 2022.

CAPES, 2021. **Brasil – [2017 a 2020] Concessão de Bolsas a Instituições de Ensino Superior apoiadas pela CAPES.** Disponível em: <Brasil - [2017 a 2020] Concessão de Bolsas a Instituições de Ensino Superior apoiadas pela CAPES - Informação geral>. Acessado em: 21 de janeiro de 2022.

Microsoft Power BI. **Power BI*.*** Disponível em: <https://powerbi.microsoft.com/pt-br/>. Acessado em: 21 de janeiro de 2022.

PINHEIRO, Sofia Alexandra. **Potencialidades do Power BI Desktop na Análise Preditiva.** Porto, 2020.

UNIVERSIDADE FEDERAL DE PELOTAS, 2020. **Em meio à pandemia, Capes amplia o corte de bolsas de pesquisa*.*** Disponível em: <https://ccs2.ufpel.edu.br/wp/2020/03/20/em-meio-a-pandemia-capes-amplia-o-corte-de-bolsas-de-pesquisa/>. Acessado em: 21 de janeiro de 2022.