

Rafael Barros Takara

**Datamart Censo do Ensino Superior**

**Rio de Janeiro – RJ**

2020

**Rafael Barros Takara**

**DATAMART CENSO DO ENSINO SUPERIOR**

DASHBOARD CONTENDO INDICADORES DE DESEMPENHO COM OBJETIVO DE AVALIAR A EFETIVIDADE DAS POLÍTICAS PÚBLICAS

Projeto de pesquisa apresentado ao curso de Small Data Business Intelligence (BI Microsoft), do Instituto Infnet, a ser utilizado como trabalho de conclusão de bloco (PB)

Orientador: prof. Landry Duailibe Salles Filho

**Rio de Janeiro – RJ**

**2020**

**RESUMO**

Este projeto tem como objetivo disponibilizar uma base para análise de dados referentes ao Censo do Ensino Superior em dois anos (2017 e 2018). Será feita uma integração com a base de municípios do IBGE, assim como uma base auxiliar de países estrangeiros visando atingir o padrão de análise em regiões.

Será desenvolvido um banco de análise com geração de um modelo dimensional, processo de ETL usando o SQL Server e os recursos e ferramentas necessárias para melhorar o desempenho das consultas ao banco desenvolvido.

Ao final da consolidação dos dados, será feito um dashboard de visualização para que se possa obter informações de tendências entre os alunos e indicadores de desempenho das instituições de ensino superior do país

**Palavras chave:** Tecnologia da Informação, Business Intelligence, Modelo Dimensional, Modelo Estrela, ETL, Integration Services, Analysis Services, Dashboard, Power BI, Censo, Ensino Superior.

**SUMÁRIO**

[**1. INTRODUÇÃO** 5](#_Toc57317937)

[1.1 FONTE DE DADOS CENSO ENSINO SUPERIOR 5](#_Toc57317938)

[1.1 FASES DO PROJETO 6](#_Toc57317939)

[**2. ANÁLISE INICIAL DOS DADOS** 7](#_Toc57317940)

[2.1 ORIGEM DOS DADOS 7](#_Toc57317941)

[2.2 ANÁLISE DOS DADOS 8](#_Toc57317942)

[**3. MODELO DIMENSIONAL** 9](#_Toc57317943)

[3.1 MODELO LÓGICO DIMENSIONAL 9](#_Toc57317944)

[3.3 DICIONÁRIO DE DADOS 13](#_Toc57317945)

[**4. CARGA DOS DADOS (ETL)** 17](#_Toc57317946)

[4.1 ARTEFATOS SCRUM 17](#_Toc57317947)

[4.3 PROJETO ETL 23](#_Toc57317948)

[**5. BANCO DE DADOS DE ANÁLISE** 24](#_Toc57317949)

[5.1 Projeto de Banco Multidimensional 24](#_Toc57317950)

[5.1.2 Grupos de Medidas 26](#_Toc57317951)

[5.2 Bus Matrix 26](#_Toc57317952)

[5.3 Projeto SSAS no Visual Studio 26](#_Toc57317953)

[5.3.6 Cubo Censoescolar 28](#_Toc57317954)

[**6. DASHBOARD** 29](#_Toc57317955)

[**7. CONCLUSÃO** 31](#_Toc57317956)

# **1. INTRODUÇÃO**

## 1.1 FONTE DE DADOS CENSO ENSINO SUPERIOR

O Censo do Ensino Superior, realizado anualmente pelo INEP, é o instrumento de pesquisa sobre as instituições de ensino superior (IES), os seus alunos e os seus docentes. Tem por objetivos informar aos participantes da educação as grandes tendências e situação do setor educacional superior ou de formação específica. Os dados são coletados por meio de preenchimento de questionários pelas IES e enviados ao sistema eletrônico do MEC.

As informações trazidas abrangem informações das instituições e seus cursos como quantidade de vagas, concluintes, ingressantes, quadro de docentes, gastos, receitas entre outras. Tem-se informações anônimas dos alunos como curso atual, idade, necessidades especiais, intercâmbios, pais de origem, atividades complementares entre outras.

Finalidade

O Censo do Ensino Superior é uma ferramenta fundamental para que os atores educacionais possam compreender a situação educacional do país, das unidades federativas, dos municípios, bem como das instituições e, com isso, acompanhar a efetividade das políticas públicas.

Além disso, as matrículas e os dados escolares coletados servem de base para o repasse de recursos do governo federal e para o planejamento e divulgação de dados das avaliações realizadas pelo INEP.

## 1.1 FASES DO PROJETO

a) Análise Inicial dos Dados

Nesta primeira faze os dados de Alunos e Escolas serão carregados dos arquivos de origem CSV para tabelas, com objetivo de explorar e compreender os dados.

b) Modelo Dimensional

Após a análise inicial dos dados, será desenvolvido um modelo dimensional estrela, assim como seus aspectos físicos, específicos para o SQL Server.

c) Carga dos Dados (ETL)

Nesta fase serão desenvolvidos pacotes de carga de dados com o Microsoft Integration Services, alimentando as tabelas do modelo estrela.

d) Banco de Dados de Análise

Desenvolvimento de banco multidimensional utilizando o Microsoft Analysis Services, para proporcional livre cruzamentos das informações do Censo Escolar, utilizando Pivot Table.

e) Dashboard

Desenvolvimento de Dashboard utilizando o Microsoft Power BI contendo os principais indicadores, com o objetivo de avaliar políticas públicas.

# **2. ANÁLISE INICIAL DOS DADOS**

## 2.1 ORIGEM DOS DADOS

Os dados do Censo do Ensino Superior estão disponíveis para download gratuito pela internet através do site do INEP abaixo:

<http://www.censosuperior.inep.gov.br/web/guest/microdados>

Juntamente com os dados, o INEP disponibiliza planilha com o dicionário de dados, fundamental para análise, pois a maioria dos atributos é composto apenas por números.

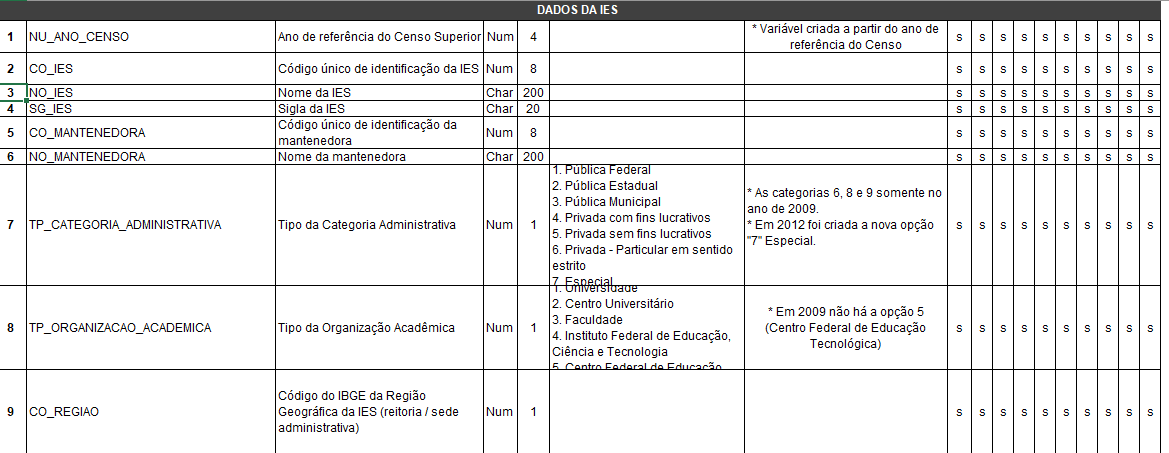


Figura 1. Foram selecionados os 2 últimos anos disponíveis no início deste projeto: 2017, 2018. Os arquivos estão formatados como texto utilizando o caractere “|” como separador de colunas.

Foi feita uma pré-seleção de atributos de acordo com o tipo de informação armazenada e a presença ou não nos 2 anos utilizados (2017, 2018). Alguns atributos são novos e aparecem apenas em 2018, outros deixaram de existir.

Dados de municípios, assim como sua organização em regiões, foram obtidos a partir do site do IBGE, link abaixo:

<https://www.ibge.gov.br/explica/codigos-dos-municipios.php>

## 2.2 ANÁLISE DOS DADOS

Os dados de IES e alunos foram importados para tabelas Stage sem qualquer tratamento, com objetivo de fazer uma análise prévia dos dados. Foi feita a importação também dos dados de municípios a partir de arquivo fornecido pelo IBGE para tabela Stage.

Foi feito teste de integridade referencial dos dados entre IES e Alunos, pois existe campo em comum contendo código da instituição “CO\_IES”. Foram verificados a integridade referencial com o código CO\_CURSO também entre as tabelas de Cursos e Alunos.

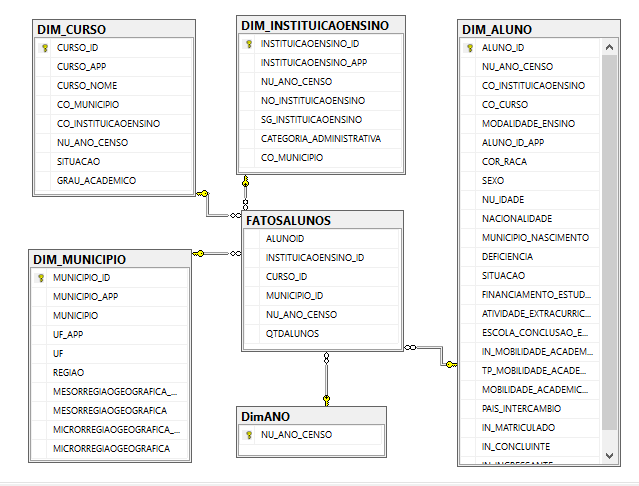
O campo “CO\_MUNICIPIO” no arquivo de escolas foi validado junto aos dados do IBGE, se relacionando perfeitamente com o campo “Municipio\_COD\_Completo”.

Foram verificados a unicidade do código CO\_IES da tabela de IES, do CO\_ALUNO\_CURSO de tabela Alunos e do CO\_CURSO da tabela Cursos, havendo unicidade nos três.

# **3. MODELO DIMENSIONAL**

## 3.1 MODELO LÓGICO DIMENSIONAL

Modelo Estrela Aluno



Script de criação do modelo dimensional

CREATE DATABASE PB-V1

ALTER DATABASE PB-V1 SET RECOVERY simple

GO

/\*\*\*\*\*\*\*Modelo Dimensional\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

DimAno

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

CREATE TABLE dbo.DimAno (

AnoCenso smallint not null primary key)

go

INSERT dbo.DimAno VALUES (2017)

INSERT dbo.DimAno VALUES (2018)

go

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

DimMunicipio

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

CREATE TABLE dbo.Dim\_MUNICIPIO (

DIMMUNICIPIO\_ID INT IDENTITY PRIMARY KEY, -- MUNICIPIO\_COD\_COMPLETO

DIMMUNICIPIO\_APP INT NOT NULL, -- MUNICIPIO\_COD

MUNICIPIO VARCHAR(100) NOT NULL,

UF\_APP INT NOT NULL, -- UF\_COD

UF VARCHAR(100) NOT NULL,

MESORREGIAOGEOGRAFICA\_APP INT NOT NULL, -- MESORREGIAOGEOGRAFICA\_COD

MESORREGIAOGEOGRAFICA VARCHAR(100) NOT NULL,

MICRORREGIAOGEOGRAFICA\_APP INT NOT NULL, -- MICRORREGIAOGEOGRAFICA\_COD

MICRORREGIAOGEOGRAFICA VARCHAR(100) NOT NULL)

go

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

DimIES

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

--DROP TABLE Dim\_INSTITUICAOENSINO

CREATE TABLE Dim\_INSTITUICAOENSINO

(

INSTITUICAOENSINO\_ID INT IDENTITY (1,1), --CO\_IES

INSTITUICAOENSINO\_APP INT NOT NULL, --CO\_IES\_APP

NU\_ANO\_CENSO smallint not null ,

NO\_INSTITUICAOENSINO varchar(300) not null,

SG\_INSTITUICAOENSINO VARCHAR(50) NOT NULL,

CATEGORIA\_ADMINISTRATIVA varchar(50) NOT NULL,

CO\_MUNICIPIO INT NOT NULL,

PRIMARY KEY (INSTITUICAOENSINO\_ID)

)

go

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

DimAluno

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

--DROP TABLE Dim\_ALUNO

CREATE TABLE [dbo].[Dim\_ALUNO](

[ALUNO\_ID] [int] IDENTITY(1,1) NOT NULL,

[NU\_ANO\_CENSO] [smallint] NOT NULL,

[CO\_INSTITUICAOENSINO] [varchar](50) NOT NULL,

[CO\_CURSO] [varchar](50) NOT NULL,

[MODALIDADE\_ENSINO] [varchar](50) NOT NULL,

[ALUNO\_ID\_APP] [varchar](50) NOT NULL,

[COR\_RACA] [varchar](50) NOT NULL,

[SEXO] [varchar](50) NOT NULL,

[NU\_IDADE] [varchar](50) NOT NULL,

[NACIONALIDADE] [varchar](50) NOT NULL,

[MUNICIPIO\_NASCIMENTO] [varchar](50) NOT NULL,

[DEFICIENCIA] [varchar](50) NOT NULL,

[SITUACAO] [varchar](50) NOT NULL,

[FINANCIAMENTO\_ESTUDANTIL] [varchar](50) NOT NULL,

[ATIVIDADE\_EXTRACURRICULAR] [varchar](50) NOT NULL,

[ESCOLA\_CONCLUSAO\_ENS\_MEDIO] [varchar](50) NOT NULL,

[IN\_MOBILIDADE\_ACADEMICA] [varchar](50) NOT NULL,

[TP\_MOBILIDADE\_ACADEMICA] [varchar](50) NOT NULL,

[MOBILIDADE\_ACADEMICA\_INTERN] [varchar](50) NOT NULL,

[PAIS\_INTERCAMBIO] [varchar](100) NOT NULL,

[IN\_MATRICULADO] [varchar](50) NOT NULL,

[IN\_CONCLUINTE] [varchar](50) NOT NULL,

[IN\_INGRESSANTE] [varchar](50) NOT NULL,

PRIMARY KEY CLUSTERED

(

[ALUNO\_ID] ASC

)WITH (PAD\_INDEX = OFF, STATISTICS\_NORECOMPUTE = OFF, IGNORE\_DUP\_KEY = OFF, ALLOW\_ROW\_LOCKS = ON, ALLOW\_PAGE\_LOCKS = ON, OPTIMIZE\_FOR\_SEQUENTIAL\_KEY = OFF) ON [PRIMARY]

) ON [PRIMARY]

GO

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

DimCurso

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

--drop table Dim\_CURSO

--select \* from Dim\_CURSO

CREATE TABLE Dim\_CURSO

(

CURSO\_ID INT IDENTITY (1,1),

CURSO\_APP INT NOT NULL, --CO\_CURSO

CURSO\_NOME VARCHAR(300) NOT NULL,

CO\_MUNICIPIO INT NOT NULL,

CO\_INSTITUICAOENSINO INT NOT NULL,

NU\_ANO\_CENSO smallINT NOT NULL,

SITUACAO varchar(50) NOT NULL,

GRAU\_ACADEMICO varchar(50) NOT NULL

PRIMARY KEY (CURSO\_ID)

)

GO

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

FatosAluno

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

CREATE TABLE FatosAluno (

ALUNOID INT NOT NULL,

IES\_ID INT NOT NULL,

CURSO\_ID INT NOT NULL,

MUNICIPIO\_ID INT NOT NULL,

NU\_ANO\_CENSO SMALLINT NOT NULL,

QTDALUNOS INT NOT NULL

)

go

ALTER TABLE FatosAluno ADD CONSTRAINT fk\_Fatosteste\_Dimaluno

FOREIGN KEY (alunoID) REFERENCES Dim\_aluno(aluno\_ID)

go

ALTER TABLE FatosAluno ADD CONSTRAINT fk\_Fatosteste\_DimIES

FOREIGN KEY (IES\_ID) REFERENCES Dim\_instituicaoensino(instituicaoensino\_ID)

go

ALTER TABLE FatosAluno ADD CONSTRAINT fk\_Fatosteste\_Dimcurso

FOREIGN KEY (curso\_ID) REFERENCES Dim\_curso(curso\_ID)

go

ALTER TABLE FatosAluno ADD CONSTRAINT fk\_Fatosteste\_Dimmunicipio

FOREIGN KEY (municipio\_ID) REFERENCES Dimmunicipio(dimmunicipio\_ID)

go

ALTER TABLE FatosAluno ADD CONSTRAINT fk\_Fatosteste\_Dimano

FOREIGN KEY (nu\_ano\_censo) REFERENCES Dimano(nu\_ano\_censo)

go

## 3.3 DICIONÁRIO DE DADOS

DimAno

Contém um registro para cada ano (2017 e 2018) do Censo do Ensino Superior.

DimMunicipio

Municípios e sua organização geográfica oficial definida pelo IBGE.

|  |  |  |
| --- | --- | --- |
| Atributo | Tipo de Dado | Descrição |
| UF\_COD | varchar(50) | Código do estado |
| UF | varchar(100) | Descrição do estado |
| MESORREGIAOGEOGRAFICA\_APP | int | Código identificador de cada Mesorregião Geográfica |
| MESORREGIAOGEOGRAFICA | varchar(100) | Agrupamento em regiões que congregavam diversos municípios de uma área geográfica de um estado brasileiro com similaridades econômicas e sociais |
| MICRORREGIAOGEOGRAFICA\_APP | int | Código identificador de cada Microrregião Geográfica |
| MICRORREGIAOGEOGRAFICA | varchar(100) | Subdivisão de uma Mesorregião Geográfica em Microrregiões, compostas de municípios limítrofes com organização espacial em comum e específica. |
| DIMMUNICIPIO\_ID | Int | Código dos municípios brasileiros composto por 7 dígitos, sendo os dois primeiros referentes ao código do estado |
| DIMMUNICIPIO\_APP | Int | Código dos municípios brasileiros composto por 5 dígitos |
| MUNICIPIO | varchar(100) | Nome do município |

DimInstituicaoEnsino

Contém a lista de instituições de ensino (IES) e seus atributos em cada ano do Censo do Ensino Superior.

|  |  |  |
| --- | --- | --- |
| Atributo | Tipo | Descrição |
| NU\_ANO\_CENSO | Num | Ano de referência do Censo Superior |
| INSTITUICAOENSINO\_ID | Num | Chave numérica sequencial |
| INSTITUICAOENSINO\_APP | Num | Código único de identificação da IES |
| NO\_INSTITUICAOENSINO | Char | Nome da IES |
| SG\_INSTITUICAOENSINO | Char | Sigla da IES |
| CATEGORIA\_ADMINISTRATIVA | Char | Tipo da Categoria Administrativa: 1. Publica, 2.Privada |
| CO\_MUNICIPIO | Num | Código do IBGE do Município da IES (reitoria / sede administrativa) |

DimAluno

Contem a identificação do aluno, seu curso, sua instituição de ensino, suas diversas características pessoais, experiências e atividades realizadas.

|  |  |  |
| --- | --- | --- |
| Atributo | Tipo | Descrição |
| ALUNO\_ID | Num | Chave sequencial gerado automaticamente |
| ALUNO\_ID\_APP | Num | Código de identificação gerado pelo Inep para o aluno da educação superior |
| NU\_ANO\_CENSO | Num | Ano de referencia |
| CO\_INSTITUICAOENSINO | Num | Código único de identificação da IES |
| NU\_IDADE | Num | Idade que o aluno completa no ano de referência do Censo |
| NACIONALIDADE | Char | Tipo da nacionalidade do aluno. 1. Brasileiro 2. Estrangeiro |
| MUNICIPIO\_NASCIMENTO | Char | Código do IBGE do município de nascimento do aluno |
| CO\_CURSO | Num | Código único de identificação do curso |
| MODALIDADE\_ENSINO | Char | Modalidade de curso presencial ou a distancia |
| COR\_RACA | Char | Etnia.  0. Aluno não quis declarar cor/raça  1. Branca 2. Preta 3. Parda 4. Amarela 5. Indígena 9. Não dispõe da informação (Não resposta) |
| SEXO | Char | Gênero. 1.Masculino 2. Feminino |
| NACIONALIDADE | Char | Informa se o aluno é brasileiro ou estrangeiro |
| DEFICIENCIA | Char | Informa se o aluno é uma pessoa com deficiência, transtorno global do desenvolvimento ou altas habilidades/superdotação. 0. Não 1.Sim |
| SITUACAO | Char | Tipo de vinculo do aluno  2.Cursando 3.Matrícula trancada 4. Desvinculado do curso 5. Transferido para outro curso da mesma IES 6.Formado 7. Falecido |
| ATIVIDADE\_EXTRACURRICULAR | Char | Informa se o aluno participa de algum tipo de atividade extracurricular (estágio não obrigatório, extensão, monitoria e pesquisa)   1. Sim, 0. Nao |
| FINANCIAMENTO\_ESTUDANTIL | Char | Informa se o aluno utiliza financiamento estudantil   1. Não 1. Sim |
| ESCOLA\_CONCLUSAO\_ENS\_MEDIO | Char | Tipo de escola de conclusão do ensino médio.  1. Pública 2. Privada  9. Não dispõe da informação (Não resposta) |
| IN\_MOBILIDADE\_ACADEMICA | Char | Informa se o aluno está regularmente matriculado em curso de graduação, que se vincula temporariamente a outra instituição, sendo ela nacional ou internacional. 0. Não 1.Sim |
| TP\_MOBILIDADE\_ACADEMICA | Char | Tipo de mobilidade acadêmica ao qual o aluno participa.  1.Nacional  2. Internacional |
| MOBILIDADE\_ACADEMICA\_INTERN | Char | Tipo de mobilidade acadêmica internacional ao qual o aluno participa.   1. Intercâmbio  2. Ciências sem fronteiras |
| PAIS\_INTERCAMBIO | Char | Pais em que o aluno realiza o intercambio |
| IN\_MATRICULADO | Char | 0. Situação diferente de matrícula 1. Situação de matrícula; |
| IN\_CONCLUINTE | Char | 0. Situação diferente de concluinte 1. Situação de concluinte |
| IN\_INGRESSANTE | Char | 0. Situação diferente de ingresso total 1. Situação de ingresso total |

Dim\_Curso

Contem o curso associado à instituição de ensino, à situação de funcionamento, e ao seu municipio

|  |  |  |
| --- | --- | --- |
| Atributo | Tipo | Descrição |
| NU\_ANO\_CENSO | Num | Ano de referência do Censo Superior |
| CURSO\_ID | Num | Codigo sequencial gerado automaticamente |
| CURSO\_APP | Num | Codigo único de identificacao do curso |
| CURSO\_NOME | Char | Nome do curso |
| CO\_INSTITUICAOENSINO | Num | Código único de identificação da Instituicao de Ensino |
| CO\_MUNICIPIO | Num | Código do IBGE do municipio do local de oferta do curso presencial |
| SITUACAO | Char | Situacao de funcionamento do curso.  1. Em atividade 2. Extinto 3. Em extinção |

# **4. CARGA DOS DADOS (ETL)**

## 4.1 ARTEFATOS SCRUM

**Backlog Produto**

1) Desenvolver Datamart e Banco de Análise Censo do Ensino Superior

“Como novo secretário de educação, necessito analisar os dados do Censo Ensino Superior acessando o banco de análise, cruzando os dados das instituições de ensino, dos seus cursos oferecidos e dos seus alunos, com objetivo de avaliar novas políticas públicas”.

2) Desenvolver Dashboard

“Como novo secretário de educação, necessito de um Dashboard com indicadores de desempenho a partir de algumas características dos alunos das instituições de ensino utilizando dados do Censo do Ensino Superior, com objetivo de avaliar a efetividade das políticas públicas, observar tendências e padrões dos alunos nos últimos anos”.

**Backlog Sprint 1**

1. Carga Stage Instituição Ensino

O módulo {Stage Instituição Ensino} receberá {os dados de dois anos do censo do ensino superior das instituições} para que o Product Owner possa validar a Stage, o que completa ainda mais a História do Usuário {“Como novo secretário de educação, necessito analisar os dados do Censo Ensino Superior acessando o banco de análise, cruzando os dados das instituições de ensino, dos seus cursos oferecidos e dos seus alunos, com objetivo de avaliar novas políticas públicas” }.

1. Carga Dimensão Instituição Ensino

“O módulo {Dimensão Instituição Ensino} receberá {a capacidade de detectar e carregar apenas registros alterados a partir da tabela Stage.Insituição Ensino} para que o Product Owner possa validar os dados carregados em {Dim Instituição Ensino}, o que completa ainda mais a História do Usuário {“Como novo secretário de educação, necessito analisar os dados do Censo Ensino Superior acessando o banco de análise, cruzando os dados das instituições de ensino, dos seus cursos oferecidos e dos seus alunos, com objetivo de avaliar novas políticas públicas” }.

1. Carga Dimensão Município

“O módulo {Dimensão Município} receberá {a capacidade de detectar e carregar apenas registros alterados a partir da tabela de Municípios do IBGE} para que o Product Owner possa validar os dados carregados em {DimMunicipio}, o que completa ainda mais a História do Usuário {“Como novo secretário de educação, necessito analisar os dados do Censo Ensino Superior acessando o banco de análise, cruzando os dados das instituições de ensino, dos seus cursos oferecidos e dos seus alunos, com objetivo de avaliar novas políticas públicas” }.

1. Carga Dimensão Ano

“O módulo {Dimensão Ano} receberá {input de dados com os Anos de carga do Censo Escolar} para que o Product Owner possa validar os dados carregados em {DimAno}, o que completa ainda mais a História do Usuário {“Como novo secretário de educação, necessito analisar os dados do Censo Ensino Superior acessando o banco de análise, cruzando os dados das instituições de ensino, dos seus cursos oferecidos e dos seus alunos, com objetivo de avaliar novas políticas públicas” }.

**Backlog Sprint 2**

1. Carga Stage Aluno

“O módulo {Stage Aluno} receberá {dados de alunos a partir de dois arquivos CSV contendo 2 (dois) últimos anos do Censo ensino superior} para que o Product Owner possa validar os dados carregados em {Stage.Aluno}, o que completa ainda mais a História do Usuário {“Como novo secretário de educação, necessito analisar os dados do Censo Ensino Superior acessando o banco de análise, cruzando os dados das instituições de ensino, dos seus cursos oferecidos e dos seus alunos, com objetivo de avaliar novas políticas públicas” }.

1. Carga Dimensão Aluno

“O módulo {Dimensão Aluno} receberá {a capacidade de detectar e carregar apenas registros alterados a partir da tabela Stage.Aluno} para que o Product Owner possa validar os dados carregados em {DimAluno}, o que completa ainda mais a História do Usuário {“Como novo secretário de educação, necessito analisar os dados do Censo Ensino Superior acessando o banco de análise, cruzando os dados das instituições de ensino, dos seus cursos oferecidos e dos seus alunos, com objetivo de avaliar novas políticas públicas” }.

1. Carga Stage Curso

“O módulo {Stage Curso} receberá {dados dos cursos a partir de dois arquivos CSV contendo 2 (dois) últimos anos do Censo ensino superior} para que o Product Owner possa validar os dados carregados em {Stage.Curso}, o que completa ainda mais a História do Usuário {“Como novo secretário de educação, necessito analisar os dados do Censo Ensino Superior acessando o banco de análise, cruzando os dados das instituições de ensino, dos seus cursos oferecidos e dos seus alunos, com objetivo de avaliar novas políticas públicas” }.

1. Carga Dimensão Curso

“O módulo {Dimensão Curso} receberá {a capacidade de detectar e carregar apenas registros alterados a partir da tabela Stage.Curso} para que o Product Owner possa validar os dados carregados em {DimCurso}, o que completa ainda mais a História do Usuário {“Como novo secretário de educação, necessito analisar os dados do Censo Ensino Superior acessando o banco de análise, cruzando os dados das instituições de ensino, dos seus cursos oferecidos e dos seus alunos, com objetivo de avaliar novas políticas públicas” }.

**Backlog Sprint 3**

1. Carga Fato Aluno

“O módulo {Fato Aluno} receberá {a capacidade de detectar e carregar apenas registros alterados a partir da tabela Stage.Aluno} para que o Product Owner possa validar os dados carregados em {Fato Aluno}, o que completa ainda mais a História do Usuário {“Como novo secretário de educação, necessito analisar os dados do Censo Ensino Superior acessando o banco de análise, cruzando os dados das instituições de ensino, dos seus cursos oferecidos e dos seus alunos, com objetivo de avaliar novas políticas públicas” }

1. Desenvolvimento do Banco de Análise

“O módulo {Banco de Análise} receberá {dados das dimensões e fatos} para que o Product Owner possa validar os dados carregados em {Cubo Censo Escolar}, o que completa ainda mais a História do Usuário {“Como novo secretário de educação, necessito analisar os dados do Censo Ensino Superior acessando o banco de análise, cruzando os dados das instituições de ensino, dos seus cursos oferecidos e dos seus alunos, com objetivo de avaliar novas políticas públicas” }

**Backlog Sprint 4**

1. Desenvolvimento da Dashboard

“Como secretário da educação, necessito de um Dashboard com indicadores de desempenho a partir de algumas características das instituições e dos seus alunos utilizando dados do Censo Escolar, com objetivo de avaliar a efetividade das políticas públicas e novas tendencias”.

4.2 Documentação ETL

No Censo do Ensino Superior, os dados são atualizados anualmente, sendo que o ultimo disponível foi de 2018. Portanto, as cargas dos dados foram feitas de uma única vez sem necessitar identificar alterações nos dados. Os dados trazidos correspondem a 2017 e 2018, sendo que cada linha da cada dimensão esta identificada o ano correspondente.

A tabela de municípios do IBGE também foi carregada para uma dimensão para se fazer uma análise por regiões. Uma tabela auxiliar com o código dos países foi utilizada para transformação dos códigos na dimensão de alunos. Todos os códigos de características e atributos binários foram transformados na ferramenta ETL para as descrições correspondentes, descartando os códigos numéricos, exceto o código município e o código da Instituição de Ensino.

-Dim\_InstituicaoEnsino

Origem: Stage\_IES

|  |  |  |  |
| --- | --- | --- | --- |
| **Coluna Origem (Stage)** | **Coluna Destino (Dim)** | **Tipo de Dado** | **Transformação** |
| **NU\_ANO\_CENSO** | NU\_ANO\_CENSO | smallint |  |
|  | INSTITUICAOENSINO\_ID | int | Identity |
| **CO\_IES** | INSTITUICAOENSINO\_APP | int |  |
| **NO\_IES** | NO\_INSTITUICAOENSINO | varchar(100) |  |
| **CO\_MUNICIPIO** | CO\_MUNICIPIO | int |  |
| **CATEGORIA\_ADMINISTRATIVA** | CATEGORIA | varchar(100) | Tipo da Categoria Administrativa: 1. Publica, 2.Privada |

-Dim\_Curso

Origem: Dm\_CURSO\_STAGE

|  |  |  |  |
| --- | --- | --- | --- |
| **Coluna Origem** | **Coluna Destino** | **Tipo de Dado** | **Transformação** |
| **NU\_ANOCENSO** | NU\_ANOCENSO | smallint |  |
|  |  |  |  |
| **CO\_CURSO** | CURSO\_APP | int |  |
| **NO\_CURSO** | CURSO\_NOME | varchar(300) |  |
| **CO\_MUNICIPIO** | CO\_MUNICIPIO | int |  |
| **CO\_IES** | CO\_INSTITUICAOENSINO | int |  |
| **TP\_SITUACAO** | SITUACAO | varChar(50) | Situação de funcionamento do curso:  1. Em atividade 2. Extinto 3. Em extinção |
| **TP\_GRAU\_ACADEMICO** | GRAU\_ACADEMICO | varChar(50) | 1. Bacharelado 2. Licenciatura 3. Tecnológico 4. N/A |

-DimAluno

Origem: Dm\_Aluno\_Stage

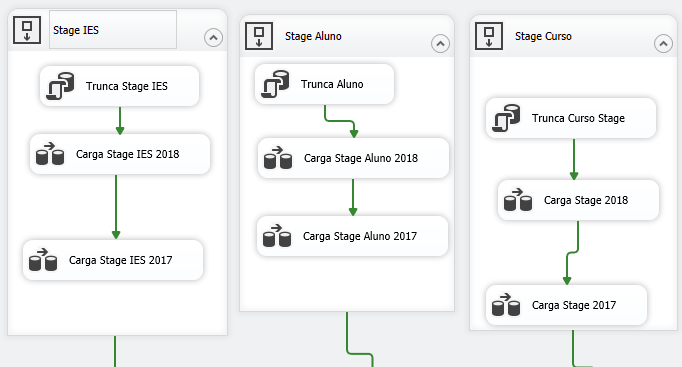
|  |  |  |  |
| --- | --- | --- | --- |
| **Coluna Origem** | **COLUNA DESTINO** | **Tipo de Dado** | **Transformação** |
| **NU\_ANO\_CENSO** | NU\_ANO\_CENSO | smallint |  |
|  | ALUNO\_ID | int | Identity |
| **CO\_IES** | CO\_INSTITUICAOENSINO | int |  |
| **CO\_CURSO** | CO\_CURSO | int |  |
| **TP\_MODALIDADE\_ENSINO** | MODALIDADE\_ENSINO | varChar(50) | 1. Presencial 2. Curso a distância |
| **CO\_ALUNO\_CURSO** | ALUNO\_ID\_APP | int |  |
| **TP\_COR\_RACA** | COR\_RACA | varChar(50) | 0. Aluno não quis declarar cor/raça  1. Branca 2. Preta 3. Parda 4. Amarela 5. Indígena 9. Não dispõe da informação (Não resposta) |
| **TP\_SEXO** | SEXO | varChar(50) | 1. Feminino 2. Masculino |
| **NU\_IDADE** | IDADE | Int |  |
| **TP\_NACIONALIDADE** | NACIONALIDADE | varChar(50) | 1. Brasileira 2. Brasileira - nascido no exterior ou naturalizado 3. Estrangeira |
| **CO\_MUNICIPIO\_NASCIMENTO** | MUNICIPIO\_NASCIMENTO | int |  |
| **IN\_DEFICIENCIA** | DEFICIENCIA | varChar(50) | 0. Não 1. Sim  9. Não dispõe de informação (Não resposta) |
| **TP\_SITUACAO** | SITUACAO | varChar(50) | 2.Cursando 3. Matrícula trancada 4. Desvinculado do curso 5. Transferido para outro curso da mesma IES 6. Formado 7. Falecido |
| **IN\_FINANCIAMENTO\_ESTUDANTIL** | FINANCIAMENTO\_ESTUDANTIL | varChar(50) | 0. Não 1. Sim |
| **IN\_ATIVIDADE\_EXTRACURRICULAR** | ATIVIDADE\_EXTRACURRICULAR | varChar(50) | 0. Não 1. Sim |
| **TP\_ESCOLA\_CONCLUSAO\_ENS\_MEDIO** | ESCOLA\_CONCLUSAO\_ENS\_MEDIO | varChar(50) | 1. Pública 2. Privada  9. Não dispõe da informação (Não resposta) |
| **TP\_MOBILIDADE\_ACADEMICA** | TP\_MOBILIDADE\_ACADEMICA | Varchar(50) | 1. Nacional  2. Internacional |
| **IN\_MOBILIDADE\_ACADEMICA,** | IN\_MOBILIDADE\_ACADEMICA | varChar(50) | 0. Não 1. Sim |
| **TP\_MOBILIDADE\_ACADEMICA\_INTERN** | MOBILIDADE\_ACADEMICA\_INTERN | varChar(50) | 1. Intercâmbio  2. Ciências sem fronteiras |
| **CO\_PAIS\_DESTINO** | PAIS\_INTERCAMBIO |  |  |
|  |  | varChar(100) | Obs. Ver correlação com tabela auxiliar Pais |
| **IN\_MATRICULADO** | IN\_MATRICULADO | varChar(50) | 0. Situação diferente de matrícula 1. Situação de matrícula; |
| **IN\_CONCLUINTE** | IN\_CONCLUINTE | varChar(50) | 0. Situação diferente de concluinte 1. Situação de concluinte |
| **IN\_INGRESSANTE** | IN\_INGRESSANTE | varChar(50) | 0. Situação diferente de ingressante 1. Situação de ingressante |

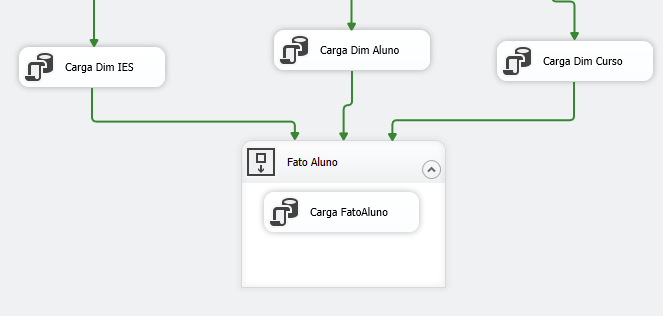
## 4.3 PROJETO ETL

Como a atualização de dados do Censo do Ensino Superior é anual, a carga torna-se mais simples, sem a necessidade de identificar atualizações nos dados, faremos sempre INSERT nas dimensões e fatos.

Para resolver as transformações, foram utilizadas tabelas auxiliares para os códigos dos Países e transformações de *lookup* e *derived column* no SSIS para as descrições dos códigos.

O projeto de carga consiste de um Pacote SSIS por ano, onde vamos alimentar tabelas Stage a partir dos arquivos CSV, em seguida carregaremos as dimensões Dim\_InstituicaoEnsino, Dim\_Curso e Dim\_Aluno, fazendo uso destas dimensões para carregar o Fato Aluno.





## **5. BANCO DE DADOS DE ANÁLISE**

## 5.1 Projeto de Banco Multidimensional

5.1.1 Dimensões

-DimAluno

Origem: dbo.DimAluno

|  |  |  |
| --- | --- | --- |
| **Coluna Origem** | **Atributo** | **Pasta** |
| NU\_ANO\_CENSO | ANO CENSO | Chave |
| ALUNO\_ID | ALUNO ID | Chave |
| CO\_INSTITUICAOENSINO | INSTITUICAO\_ID | Identificador |
| CO\_CURSO | CURSO ID | Identificador |
| MODALIDADE\_ENSINO | MODALIDADE | Caracteristica |
| ALUNO\_ID\_APP | ALUNO ID APP | Identificador |
| COR\_RACA | COR\_RACA | Característica |
| SEXO | SEXO | Característica |
| IDADE | IDADE | Característica |
| NACIONALIDADE | NACIONALIDADE | Característica |
| MUNICIPIO\_NASCIMENTO | MUNICIPIO\_NASCIMENTO | Característica |
| DEFICIENCIA | DEFICIENCIA | Característica |
| SITUACAO | SITUACAO | Característica |
| FINANCIAMENTO\_ESTUDANTIL | FINANCIAMENTO\_ESTUDANTIL | Característica |
| ATIVIDADE\_EXTRACURRICULAR | ATIVIDADE\_EXTRA | Característica |
| ESCOLA\_CONCLUSAO\_ENS\_MEDIO | CONCLUSAO\_ENS\_MEDIO | Característica |
| TP\_MOBILIDADE\_ACADEMICA | TIPO MOBILIDADE | Característica |
| IN\_MOBILIDADE\_ACADEMICA | MOBILIDADE\_ACADEMICA | Característica |
| MOBILIDADE\_ACADEMICA\_INTERN | TIPO\_MOBILIDADE\_ACADEMICA\_INT | Característica |
| PAIS\_INTERCAMBIO | PAIS\_INTERCAMBIO | Característica |
| **IN\_MATRICULADO** | IN\_MATRICULADO | Característica |
| **IN\_CONCLUINTE** | IN\_CONCLUINTE | Característica |
| **IN\_INGRESSANTE** | IN\_INGRESSANTE | Característica |

|  |
| --- |
| **Hierarquia: Aluno – Nacionalidade** |
| ANO CENSO |
| NACIONALIDADE |
| PAIS ORIGEM |
| ALUNO ID |

|  |
| --- |
| **Hierarquia: Aluno – Intercambio** |
| ANO CENSO |
| MOBILIDADE ACADEMICA |
| MOBILIDADE\_INTERNACIONAL |
| PAIS\_INTERCAMBIO |
| ALUNO ID |

|  |
| --- |
| **Hierarquia: Aluno – Ensino Medio** |
| ANO CENSO |
| CONCLUSAO\_ENSINO\_MEDIO |
| ALUNO ID |

**Dim InstituicaoEnsino**

- Tabela Origem: dbo.DimInstituicaoEnsino

|  |  |  |
| --- | --- | --- |
| **Coluna Origem** | **Atributo** | **Pasta** |
| NU\_ANO\_CENSO | ANO CENSO | Chave |
| INSTITUICAOENSINO\_ID | INSTITUICAOENSINO\_ID | Chave |
| INSTITUICAOENSINO\_APP | INSTITUICAO\_ENSINO\_APP | Identificador |
| NO\_INSTITUICAOENSINO | INSTITUICAO\_ENSINO | Caracteristica |
| CO\_MUNICIPIO | MUNICIPIO | Caracteristica |
| CATEGORIA | CATEGORIA | Característica |
|  |  |  |

|  |
| --- |
| **Hierarquia: Instituicao – Categoria** |
| ANO CENSO |
| CATEGORIA |
| INSTITUICAO\_ENSINO |
| INSTITUICAOENSINO\_APP |

**Dim Curso**

- Tabela Origem: dbo.Dim-Curso

|  |  |  |
| --- | --- | --- |
| **Coluna Origem** | **Atributo** | **Pasta** |
| NU\_ANOCENSO | ANO CENSO | Chave |
| CURSO\_APP | CURSO APP | Chave |
| CURSO | CURSO ID | Chave |
| CURSO\_NOME | CURSO NOME | Identificador |
| CO\_MUNICIPIO | MUNICIPIO | Caracteristica |
| CO\_INSTITUICAOENSINO | INSTITUICAO\_ENSINO\_ID | Caracteristica |
| SITUACAO | SITUACAO | Característica |
|  |  |  |

|  |
| --- |
| **Hierarquia: Curso – Situacao** |
| ANO\_CENSO |
| SITUACAO |
| CURSO |

**Dim Municipio**

- Tabela Origem: dbo.DimMunicipio

|  |  |
| --- | --- |
| **Coluna Origem** | **Atributo** |
| UF\_APP | UF ID |
| UF | UF |
| MESORREGIAOGEOGRAFICA\_APP | MESORREGIAOGEOGRAFICA ID |
| MESORREGIAOGEOGRAFICA | MESORREGIAOGEOGRAFICA |
| MICRORREGIAOGEOGRAFICA\_APP | MICRORREGIAOGEOGRAFICA ID |
| MICRORREGIAOGEOGRAFICA | MICRORREGIAOGEOGRAFICA |
| DIMMUNICIPIO\_ID | MUNICIPIO ID COMPLETO |
| DIMMUNICIPIO\_APP | MUNICIPIO ID SIMPLES |
| MUNICÍPIO | MUNICIPIO |

|  |
| --- |
| **Hierarquia: Municipio – UF** |
| UF |
| MESORREGIAOGEOGRAFICA |
| MICRORREGIAOGEOGRAFICA |
| MUNICIPIO |

### 5.1.2 Grupos de Medidas

**Fato Aluno**

- Tabela Origem: dbo.FatoAluno

|  |  |  |  |
| --- | --- | --- | --- |
| **Coluna Origem** | **Medida** | **Formato** | **Agregação** |
| QTD\_ALUNOS | QTD ALUNOS | #,#0 | Sum |

## 5.2 Bus Matrix

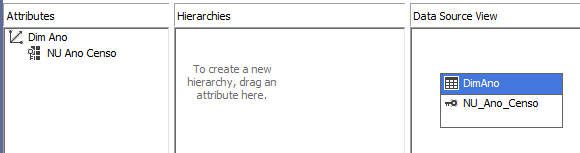
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | DIM ALUNO | DIM INSTITUICAOENSINO | DIM MUNICIPIO | DIMANO | DIM CURSO |
| FATOSALUNO | • | • | • | • | • |

## 5.3 Projeto SSAS no Visual Studio

A partir de um modelo estrela alimentado com dados na Etapa anterior, foi desenvolvido projeto de um banco Analysis Services Multidimensional, contendo cinco dimensões e um grupo de medida.

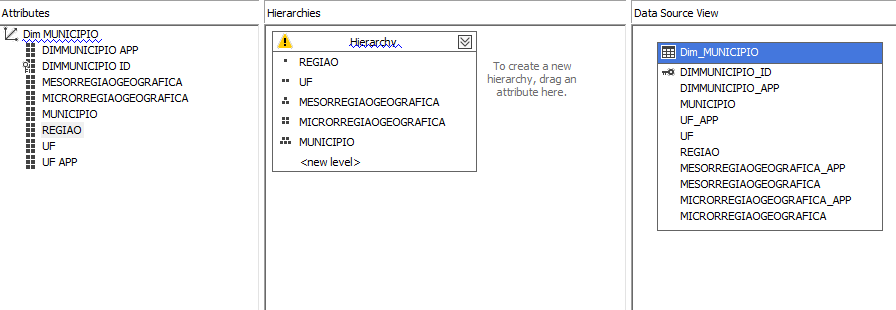
A dimensão tempo (DimAno) possui apenas um atributo Ano, com dois membros representando os dois anos de Censo Ensino Superior.

-5.3.1. Dim Ano



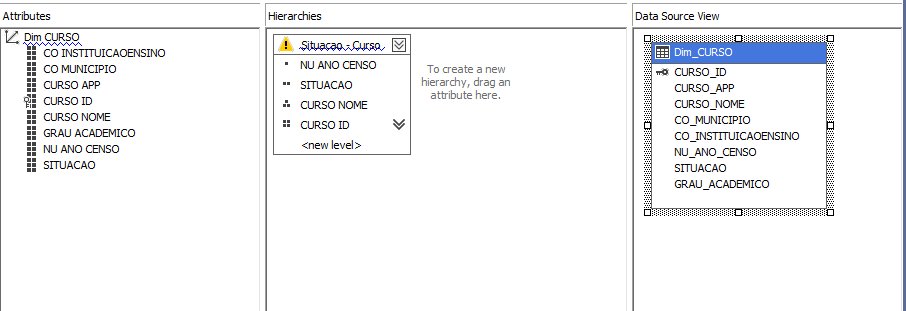
-5.3.2. Dim Municipio

Foi criada uma hierarquia Municipio – UF com as micro e mesorregiões geograficas



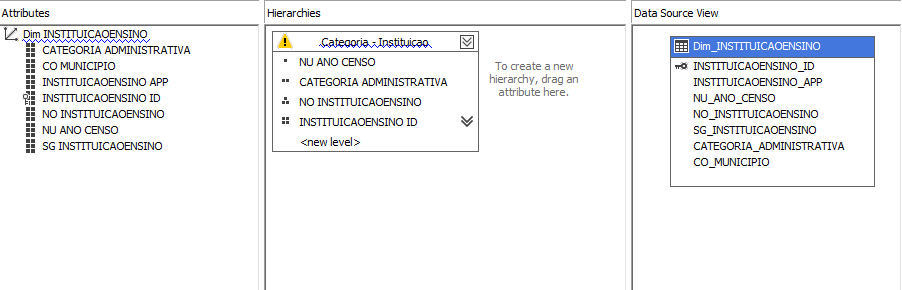
-5.3.3 Dim Curso

Na dimensão Curso foi criada uma hierarquia: situação - curso



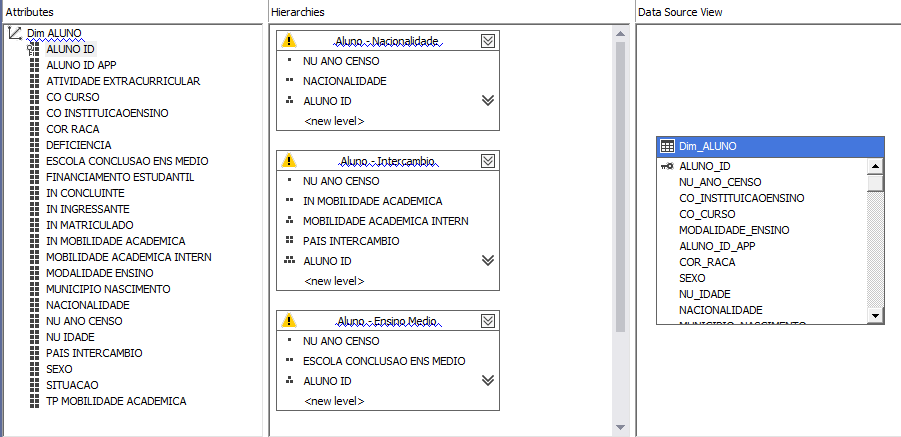
-5.3.4. Dim InstituicaoEnsino

Na dimensão Instituicao Ensino foi criada uma hierarquia: Categoria.



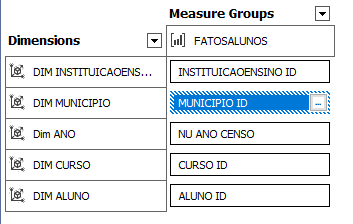
-5.3.5. Dim Aluno

Na dimensão aluno foram criadas três hierarquias: Nacionalidade, Intercambio e Conclusao Ensino Medio.

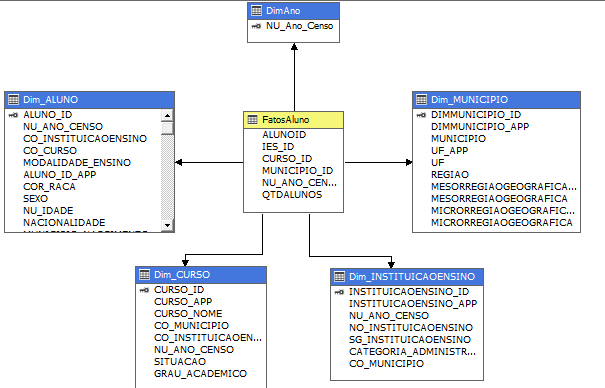


### 5.3.6 Cubo Censoescolar

O Cubo **CensoEscolar** foi estruturado com um grupo de medida, obtendo suas métricas a partir de uma tabela de fatos do modelo lógico dimensional: FatoAlunos. Relacionamento entre Dimensões e Grupos de medidas:

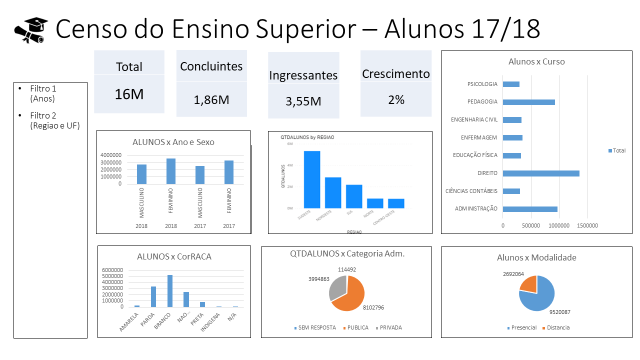


Estrutura do Cubo Censo Ensino Superior:



# **6. DASHBOARD**

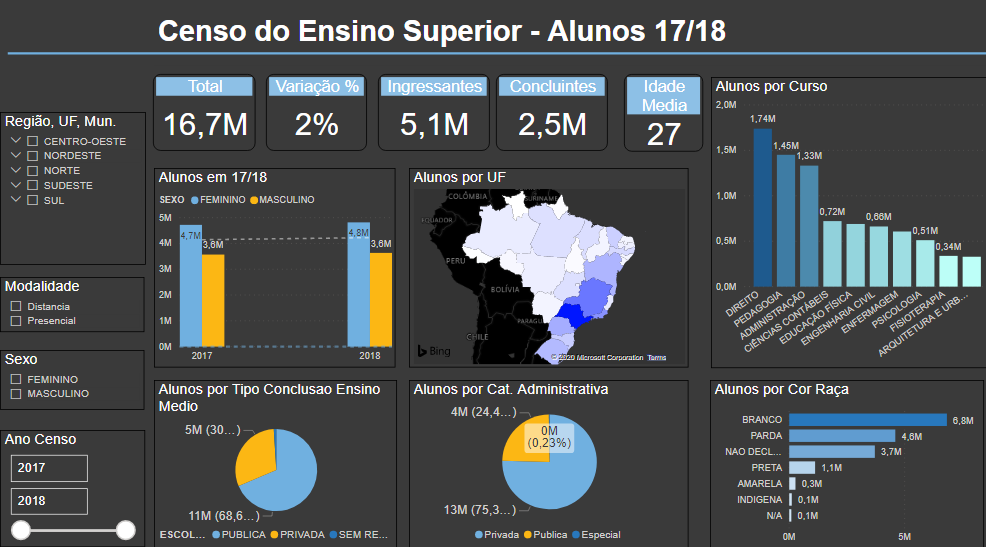
6.1 Mockup



6.2. Versão Final do Dashboard

Link para download do arquivo do power BI:

<https://1drv.ms/u/s!AuvZJUwtDyaci1-TV2eu4edMLPgl?e=7oX7Cx>



# **7. CONCLUSÃO**

Neste trabalho foi feita a coleta de dados referentes ao censo do ensino superior no Brasil dos anos de 2017 e 2018. Após a coleta, os dados foram carregados a feiras as transformações adequadas. Os dados foram consolidados e carregadas em uma camada semântica, para então ser obtido um dashboard com os dados consolidados. O dashboard criado contém dois filtros e seis gráficos de agregação de características dos alunos. Selecionando-se alguma característica em particular, pode-se revelar o crescimento percentual em um card.

Em relação a análise dos dados, alguns pontos relevantes puderam ser observados após serem feitas as agregações dos dados. Os dados mais relevantes estão listados abaixo

* Características mais predominantes: sexo feminino, Bacharelado, predominância de cor branca, região sudeste e SP mais concentrada, setor privado, escola pública de conclusão do ensino médio
* Crescimento de 2% de total de alunos entre os anos de 2017 e 2018 tanto em instituições públicas como privadas
* Idade média de 27 anos com desvio padrão de 8
* Modalidade a distância vem crescendo (17%) e presencial decrescendo (-2%)
* Os seis cursos mais populares foram nessa ordem: Direito, Pedagogia, Administração, Ciências Contábeis e Educação Física, Engenharia Civil. Há predominância da área de humanas e ciências sociais. No presencial há predominância de Direito e a distância de Pedagogia
* Portugal e França foram os países mais procurados em intercâmbios. Alunos em intercambio em torno de 10000. Há poucos alunos estrangeiros no país, menos de 1% do total.