

Perfil dos beneficiários do PROUNI

2015 à 2019

Emanoel Rainey Curvelo Manço

eraineycm8@gmail.com

Links

- ❖ Dataset: <https://www.kaggle.com/datasets/lfarhat/brasil-students-scholarship-prouni-20052019>
- ❖ Git: <https://github.com/eraineycm8/profile> PROUNI beneficiaries

Etapa 1 - Planejamento



PROUNI – Contexto geral

- ❖ O PROUNI é um programa que oferece bolsas de estudo, integrais e parciais, em cursos de graduação em instituições privadas. Podem participar estudantes sem diploma de nível superior que tenham realizado o ENEM, tendo obtido no mínimo 450 pontos na média das provas e não tenham zerado a redação, além de atender critérios de renda familiar.
- ❖ Para a bolsa integral, a renda familiar mensal per capita deve ser de até 1,5 salário mínimo, enquanto para a parcial, é de até 3 salários mínimos. É necessário ter cursado o ensino médio em escola pública ou privada como bolsista integral, parcial ou de forma mista, ou se enquadrar como pessoa com deficiência. Professores da rede pública de ensino também podem se candidatar para cursos de licenciatura e pedagogia, sem limite de renda.

Objetivos do Data Mart

- ❖ O objetivo do Data Mart a ser desenvolvido é fornecer uma análise detalhada do perfil dos beneficiários do PROUNI entre os anos de 2015 e 2019. Este Data Mart possibilitará extrair informações relevantes sobre a distribuição dos bolsistas por região, faixa etária, tipo de bolsa (integral ou parcial) e cursos frequentados ao longo desse período. A análise permitirá identificar tendências, variações e características específicas dos estudantes contemplados pelo programa, e outros pontos de vista sobre o programa em si. Com a análise dos dados espera-se criar questionamentos e insights para avaliar e aumentar a eficácia do programa, em busca de entender a distribuição de recursos e beneficiários, os estudantes de baixa renda em busca de educação superior.

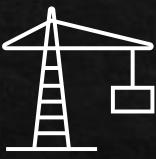
Arquitetura Tecnológica

Fonte de Dados



CSV

ETL



Pentaho

Data Mart



Mysql

Aplicação OLAP



Power BI

Processo

- ❖ Planejamento
- ❖ Levantamento das Necessidades
- ❖ Modelagem dimensional
- ❖ Projeto Físico do Banco de Dados
- ❖ Projeto ETL
- ❖ Aplicação OLAP

Processo

- ❖ Para criar este Data Mart, foi adotada a estratégia Bottom-up, a qual se desenvolve de forma progressiva e permite a entrega imediata de resultados. Essa abordagem foi escolhida por ser específica para a construção de um Data Mart e por sua ênfase na análise detalhada de partes para compor o todo.
- ❖ Será utilizado o modelo "Star Schema" por ser uma extensão incremental da estratégia Bottom-up, focando na montagem do Data Mart. Essa técnica concentra-se na criação de tabelas de dimensões e fatos, alinhada à natureza descentralizada de um Data Warehouse, facilitando a análise e a compreensão dos dados em um contexto mais segmentado e analítico.

Etapa 2 - Levantamento das Necessidades



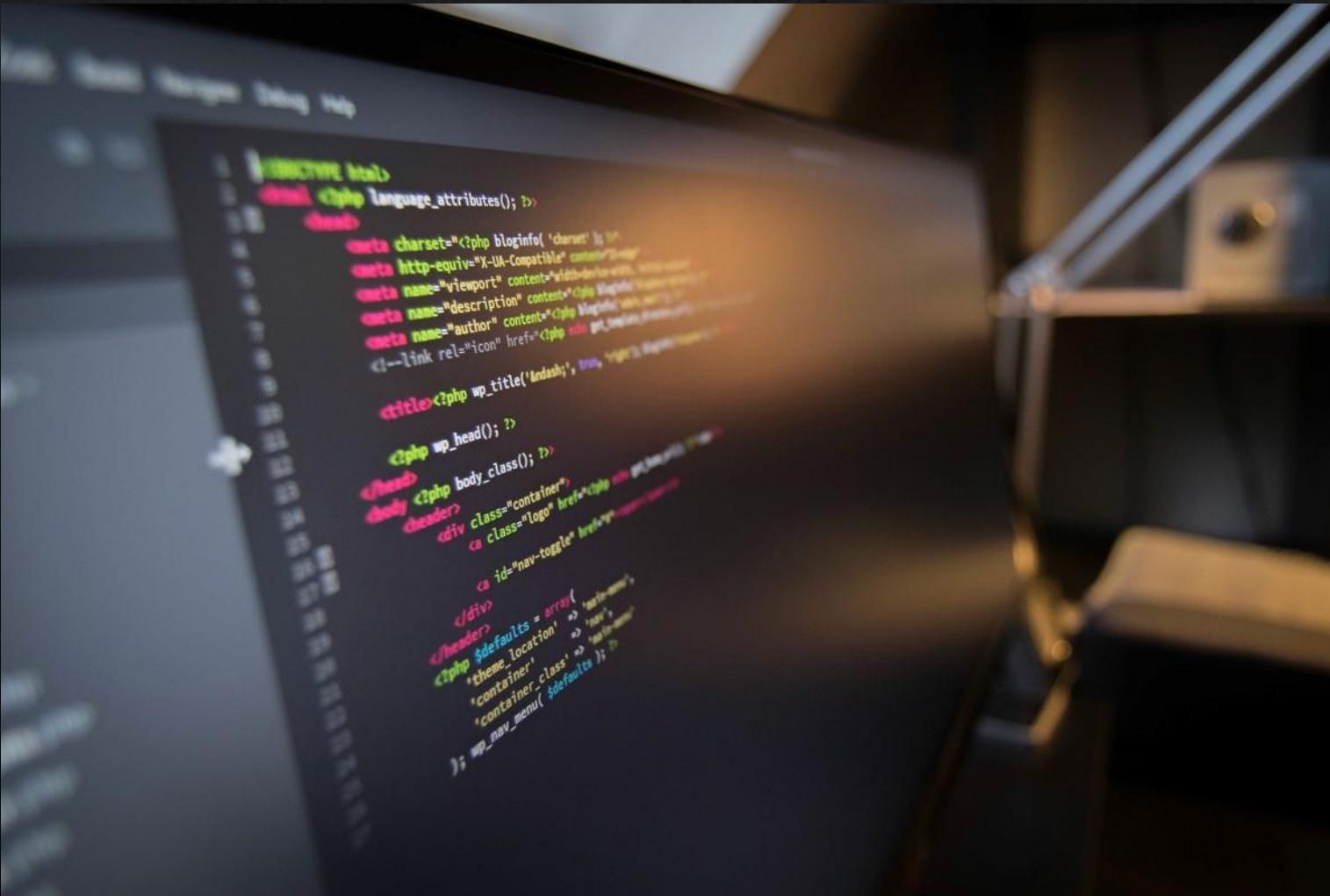
Consultas de Apoio à Decisão

- ❖ Crescimento de beneficiários do PROUNI
- ❖ Quantidade de estudantes por região e ano
- ❖ Comparação de estudantes entre regiões por ano
- ❖ Distribuição dos estudantes por idade e tipo da bolsa
- ❖ Distribuição dos estudantes por ano e idade

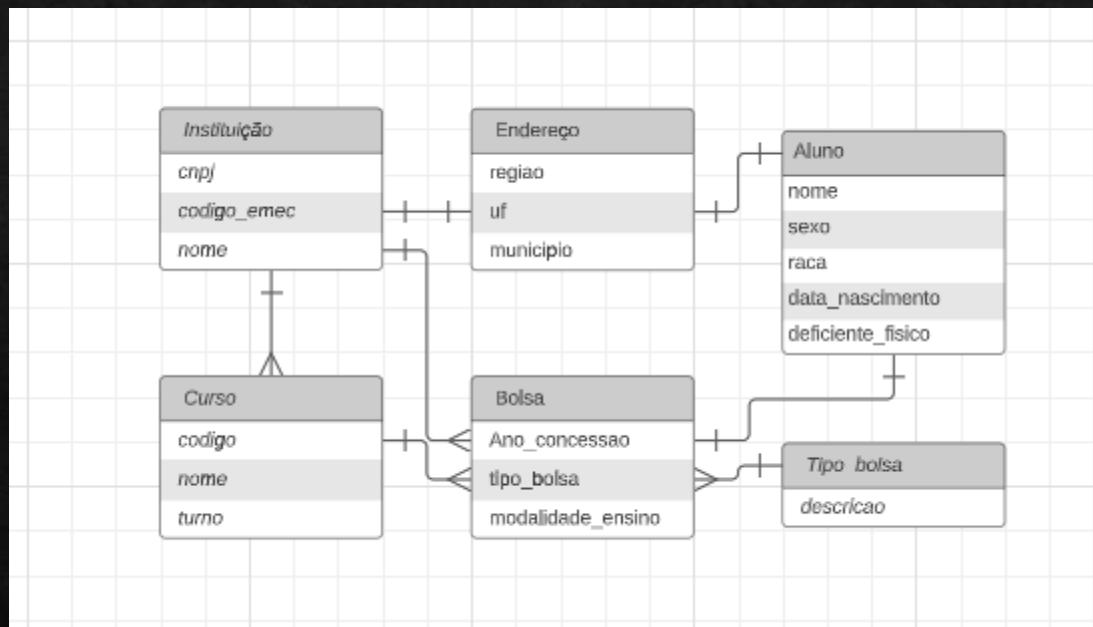
Indicadores de análise

- ❖ Quantidade de estudantes por região
- ❖ Quantidade de estudantes por estado
- ❖ Quantidade de estudantes por faixa etária
- ❖ Quantidade de estudantes por ano

Etapa 3 - Modelagem



Modelo Relacional



Modelo Dimensional

A – Área de Negócios

Educação

B – Processo

Análise dos beneficiários do PROUNI no Brasil entre 2015 a 2019

C – Granularidade

Estado x Região x Ano x Faixa etária

Modelo Dimensional

D – Atributos e Hierarquia das Dimensões

(DIM_Local: região, estado)

(DIM_Tempo: ano)

(DIM_Faixa_Etaria: faixa etária)

(DIM_Tp_Bolsa: tp_bolsa)

(DIM_Turno: turno)

(Dim_Sexo: sexo)

(Dim_Condicao_Fisica: condição_física)

(Dim_grupo_racial: grupo_racial)

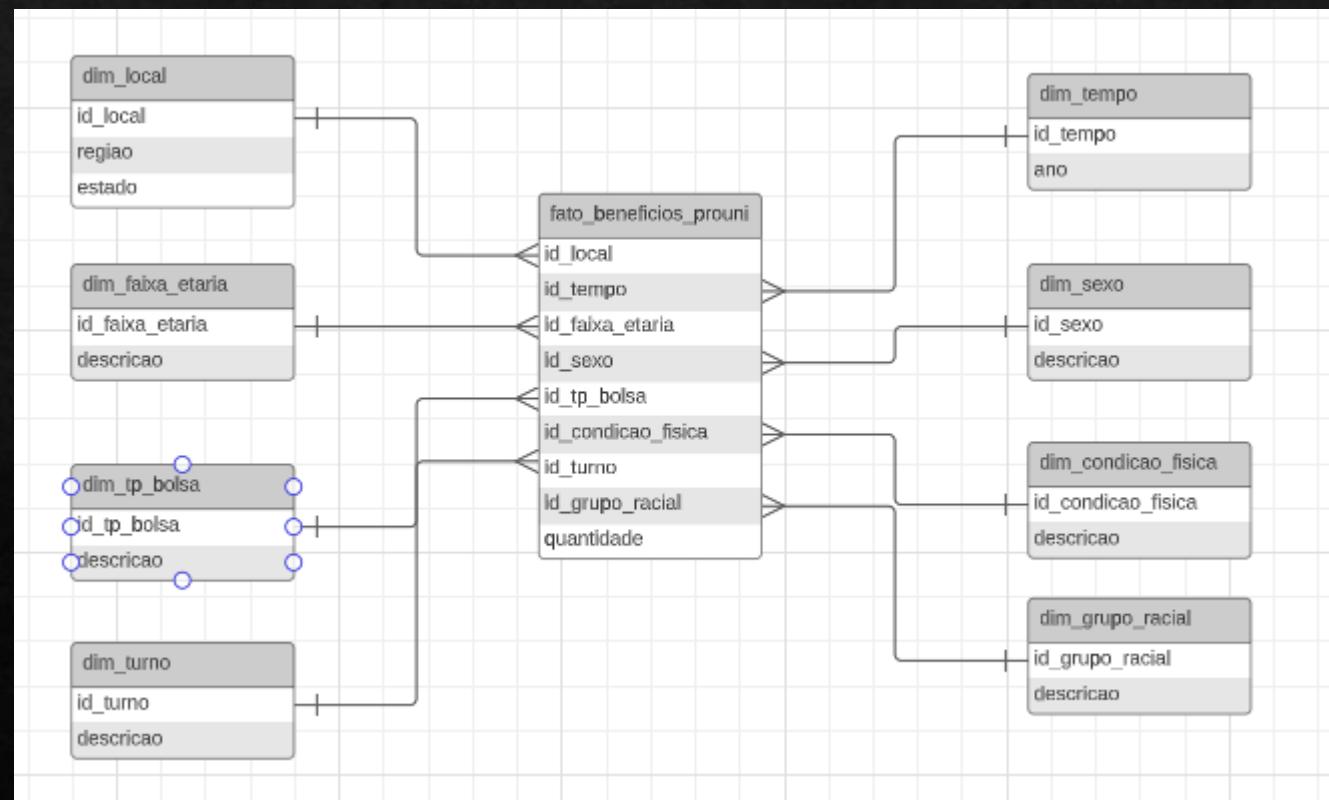
Modelo Dimensional

E – Métricas da Fato

- ❖ Quantidade (aditiva)

Modelo Dimensional

F – Esquema Estrela



Modelo Dimensional

G – Inserção de 10 “Fatos”

dim_sexo	
id_sexo	descricao
1	Feminino
2	Masculino

dim_tipo_bolsa	
id_tp_bolsa	descricao
1	Integral
2	Parcial

fatoBeneficios_prouni									
id_local	id_tempo	id_faixa_etaria	id_sexo	id_tp_bolsa	id_condicao_fisica	id_turno	id_grupo_racial	quantidade	
1	1	1	2	1	1	2	2	425	
1	1	1	1	1	2	1	1	233	
1	1	2	1	1	1	3	3	386	
1	1	2	2	1	2	4	1	423	
2	1	3	1	2	2	2	2	602	
2	1	3	2	2	2	2	3	100	
2	1	4	2	1	2	3	2	589	
2	1	4	1	2	2	3	1	231	
2	1	5	1	2	2	4	3	63	

dim_turno	
id_turno	Turno
1	A Distancia
2	Integral
3	Matutino
4	Noturno
5	Vespertino

dim_tempo	
id_tempo	ano
1	2015
2	2016
3	2017
4	2018
5	2019

dim_local		
id_local	regiao	estado
1	Norte	Acre
2	Norte	Amapá
3	Norte	Amazonas
4	Norte	Pará
5	Norte	Rondônia
6	Norte	Roraima
7	Norte	Tocantins
8	Nordeste	Alagoas
9	Nordeste	Bahia
10	Nordeste	Ceará

dim_raca	
id_raca	Raca
1	Parda
2	Branca
3	Preta
4	Amarela
5	Indigena
6	Nao Informada

dim_faixa_etaria	
id_faixa_etaria	descricao
1	Até 18 anos
2	19-25 anos
3	26-30 anos
4	31-40 anos
5	Acima de 40 anos

dim_condicao_fisica	
id_condicao_fisica	descricao
1	Com Deficiência
2	Sem Deficiência

Modelo Dimensional

H – Estimativa de Espaço

- ❖ Tendo em vista 27 locais, 2 sexo, 2 tipo bolsa, 5 turnos, 5 tempos, 6 raças, 5 faixas etárias, 2 condição física temos:
- ❖ $27 \times 2 \times 2 \times 5 \times 5 \times 6 \times 5 \times 2 = 162.000$ registros.
- ❖ Sendo esses campos chaves cada com 4 bytes e mais um valor numérico também de 4 bytes temos:
- ❖ $162.000 \times ((8 \times 4) + (1 \times 4)) = 5.832.000$ bytes