



# Relatório – Análise de Dados ENEM 2020

Analista: Alcides Gabriel



## Modelagem dos Dados

Inicialmente o planejamento era realizar a modelagem dos dados com a linguagem SQL, utilizando o framework MySQL, mas devido a problemas computacionais, não foi possível realizar a modelagem, visto que ocorria falhas de carregamento pelo peso do arquivo. Sendo assim, a modelagem dos dados foi realizada através do próprio Power BI. Porém, como forma de demonstração de conhecimento, demonstrei abaixo a estrutura das tabelas dimensões montadas pelo SQL que futuramente seriam importadas.

```
USE enem_2020;

SELECT * FROM microdadosenem limit 10;

# Criar coluna na tabela
ALTER TABLE microdadosenem
ADD COLUMN id_escola VARCHAR(255);

## Update da tabela criando o ID_escola
UPDATE microdadosenem
SET id_escola = CONCAT(co_municipio_esc, tp_dependencia_adm_esc, tp_localizacao_esc, tp_sit_func_esc);

## Dimensão dados dos inscritos faltantes
CREATE VIEW DadosFaltantes AS
SELECT nu_inscricao, co_prova_ch, co_prova_cn, co_prova_lc, co_prova_mt, tp_presenca_ch, tp_presenca_cn, tp_presenca_lc, tp_presenca_mt
FROM microdadosenem;

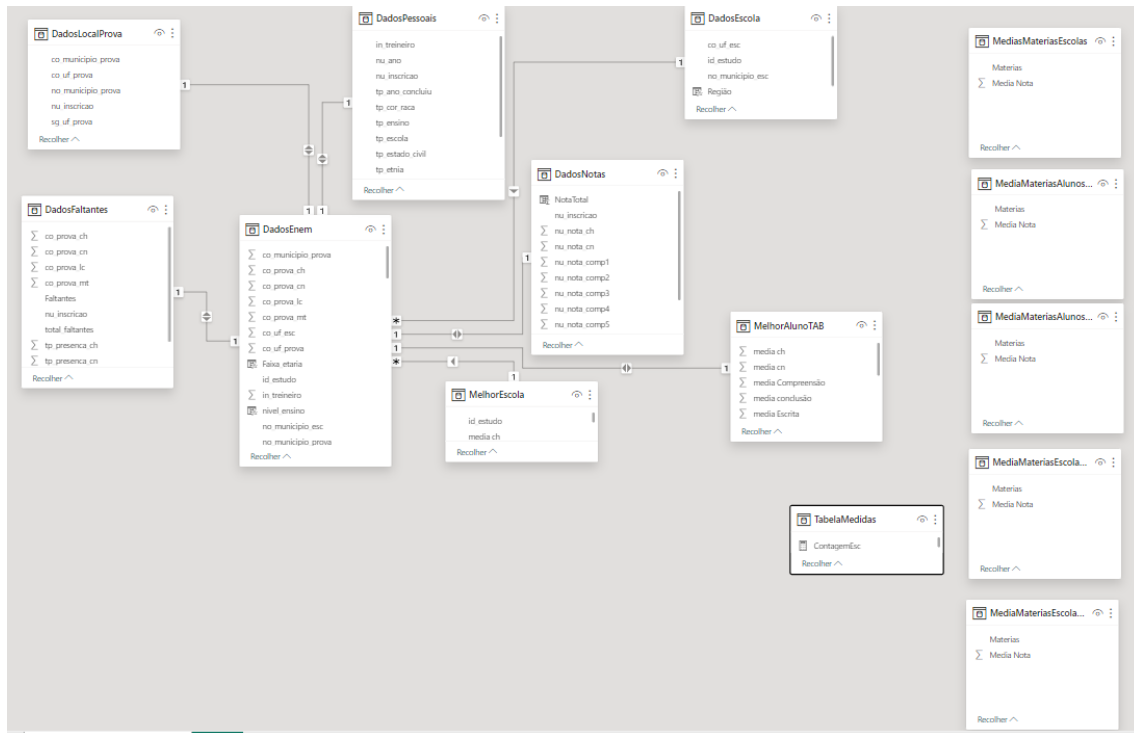
## Dimensão dados dos inscritos faltantes
CREATE VIEW DadosFaltantes AS
SELECT nu_inscricao, co_prova_ch, co_prova_cn, co_prova_lc, co_prova_mt, tp_presenca_ch, tp_presenca_cn, tp_presenca_lc, tp_presenca_mt
FROM microdadosenem;

## Dimensão das notas
CREATE VIEW DadosNotas AS
SELECT
    nu_inscricao,
    nu_nota_ch,
    nu_nota_cn,
    nu_nota_lc,
    nu_nota_mt,
    nu_nota_redacao,
    (nu_nota_ch + nu_nota_cn + nu_nota_lc + nu_nota_mt + nu_nota_redacao) / 5 AS media
FROM microdadosenem;

## Dimensão dados pessoais
CREATE VIEW DadosPessoais AS
SELECT nu_inscricao, nu_ano, in_treineiro, tp_ano_concluiu, tp_cor_raca, tp_ensino, tp_escola, tp_estado_civil, tp_faixa_etaria, tp_nacionalidade, tp_sexo, tp_st_conclusao

# Dimensão local prova
CREATE VIEW DadosLocalProva AS
SELECT co_municipio_prova, co_uf_prova, no_municipio_prova, nu_inscricao, sg_uf_prova FROM microdadosenem;
```

Como processo inicial da modelagem dos dados, foi realizado a organização no formato estrela, no qual foi composta de 1 tabela fato e 7 dimensões, pode-se encontrar a imagem ilustrada abaixo.



Também nessa imagem, pode-se perceber a presença de 6 tabelas externas da conexão, elas são tabelas de métricas usadas nos gráficos. Para a criação desses elementos foi utilizado a função DAX. Outros processos de modelagem foram realizados, como criação de identificador das escolas, transformações dos tipos, entre outras.

## Respondendo as Questões

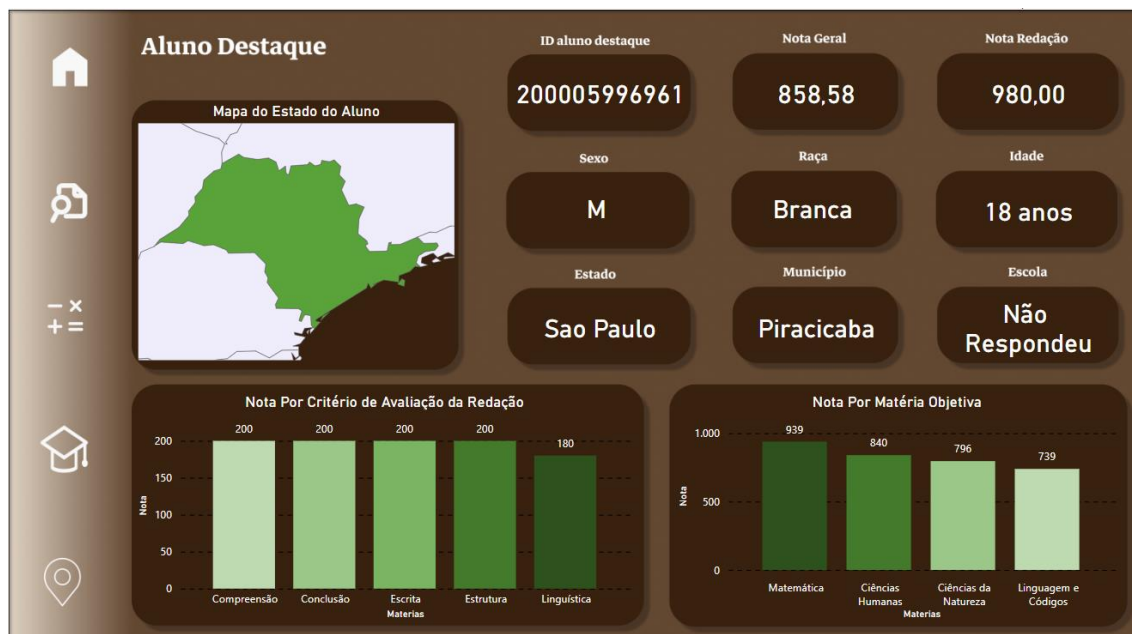
Qual a escola com a maior média de notas?



A imagem acima ilustra o dashboard demonstrando qual foi a escola com maior média das notas, alcançando o valor de 770,70, no município de Marabá, Pará. Percebe-se que essa medida é enviesada, visto que só há um único aluno da escola que realizou a prova, ou que registrou qual era a escola. Outras análises possíveis é o entendimento das notas dele, tanto na formação da sua nota na redação como também na prova objetiva.



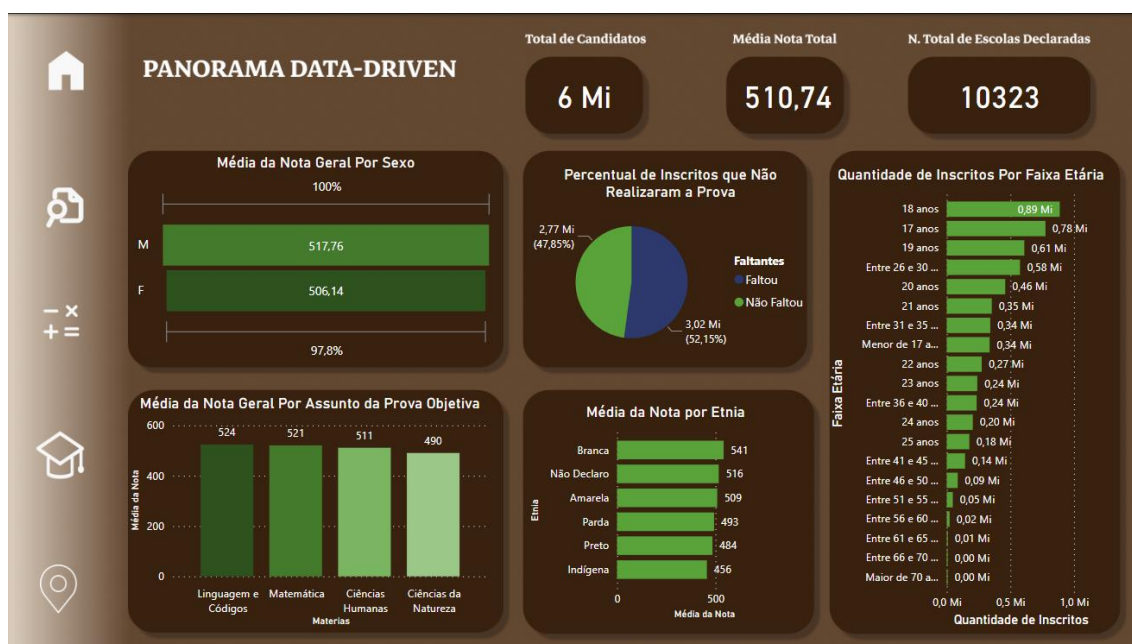
Qual o aluno com a maior média de notas e o valor dessa média?



O aluno com id 200005995951 foi quem obteve a maior nota no enem 2020, com um valor total de 858,58 na nota final e 980 na redação. É um indivíduo do sexo masculino, cor branca, com 18 anos de idade, residente do estado de São Paulo, no município de Piracicaba. A escola em que ele se estuda ou estudou não foi informada. Destaque para o grande desempenho em matemática e sua nota na redação.

Qual a média geral? Qual o % de Ausentes? Qual o número total de Inscritos? Qual a média por disciplina? Qual a média por Sexo? Qual a média por Etnia?

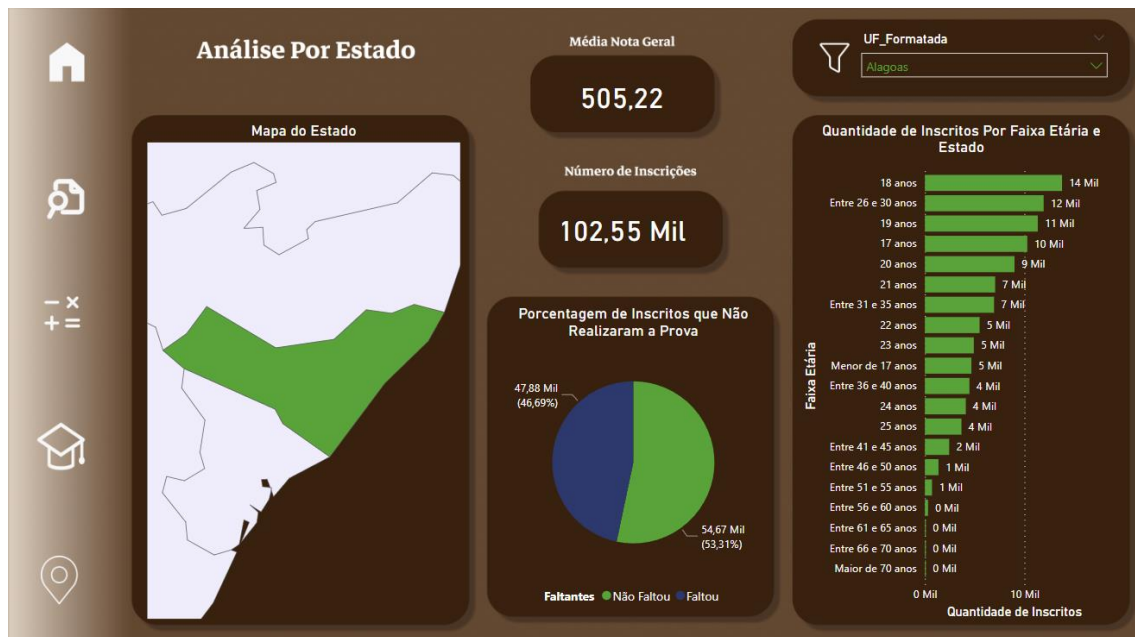
Por que agregar todas essas perguntas? Simples, um único dashboard irá responder por todas.



Começando pela média geral, vemos que foi de 510. Sendo 52,15% de inscritos ausentes dos 6 milhões de participantes. Por disciplina vemos as médias de 524 para linguagem e códigos, 521 para matemática, 511 para ciências humanas e 480 para ciências da natureza. Por sexo a média dos indivíduos masculinos foi de 517, enquanto dos femininos foi de 506. Por etnia, vemos que a branca foi de 541, 516 dos que não declararam, 509 amarela, parda 493, preto de 484, e por último, os indígenas com 456.



## Outros Insights



Acima vemos a análise por estado, sendo possível uma análise de nota, participantes, faixa etária e percentagem de faltantes, além da visualização por estado.

Lembrando que estamos falando de um dashboard, ou seja, é um material interativa. Sendo assim, várias pesquisas sobre os dados e categorias podem ser feitas trazendo novos insights sobre os dados.