

# Dados do INEP

**Exame nacional do ensino médio**

---

Vinícius dos Santos Moreira

Fevereiro de 2025

# Agenda

---

- Introdução
- Modelo relacional
- Processo de ETL
- Consultando os dados do Enem
- Visões e procedures

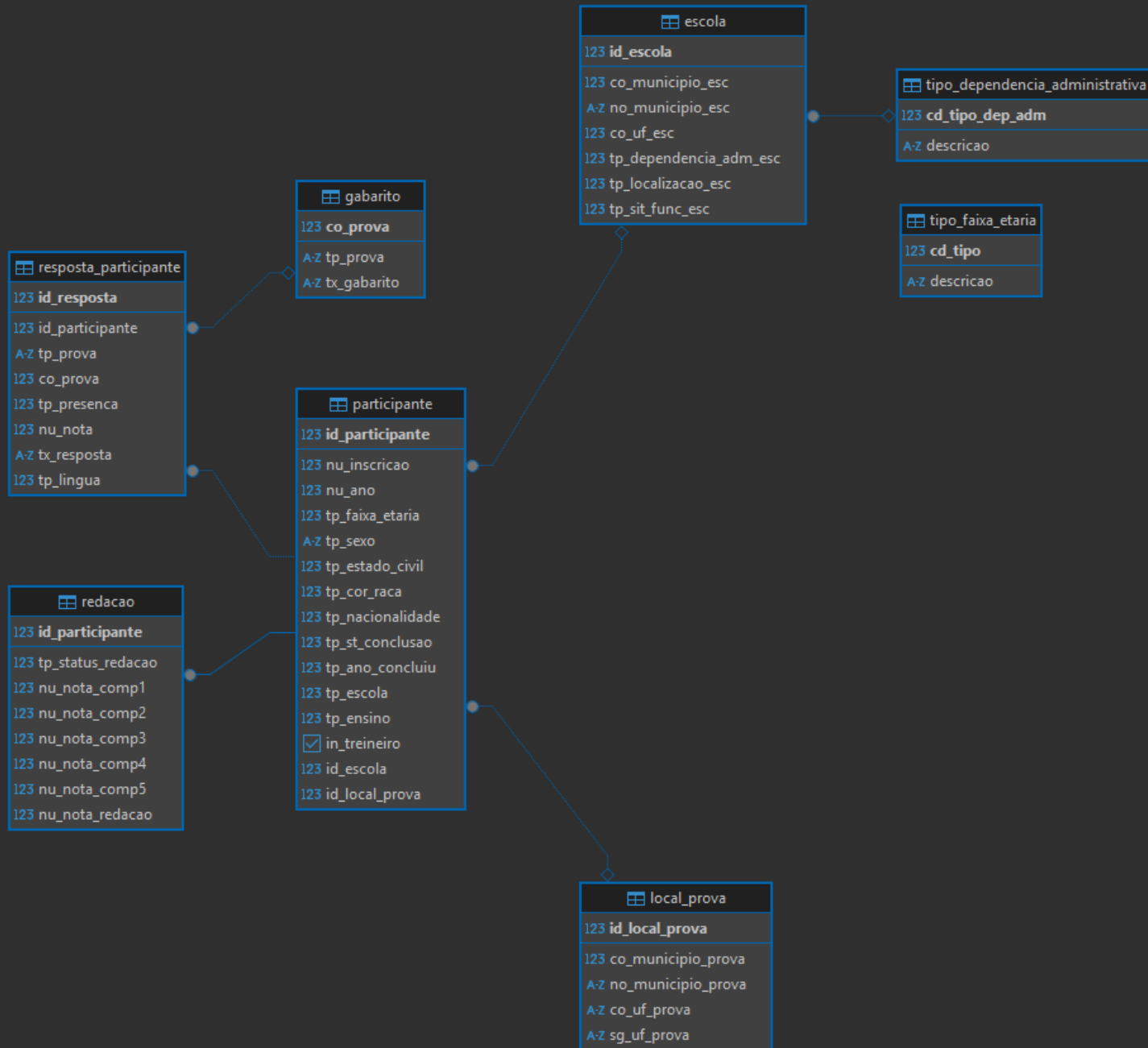


# O ENEM

---

- Desde 1998 até 2008 – usado ainda por poucas instituições de ensino
- A partir de 2009 – reformulação e maior abrangência
- Os resultados do Enem deverão possibilitar, dentre outros, “o desenvolvimento de estudos e indicadores sobre a educação brasileira.”

# Modelo Entidade Relacionamento



# O processo ETL

**Origem da importação**  
Configura o tipo e formato da origem de transferência de dados

Origem da importação Arquivo(s) de entrada Mapeamento de tabelas Configurações de carga de dados Confirmação	CSV	Importar de arquivo(s) CSV	inep
--	-----	----------------------------	------

Salvar tarefa

microdados_enem	
A-Z	nu_inscricao
123	nu_ano
123	tp_faixa_etaria
A-Z	tp_sexo
123	tp_estado_civil
123	tp_cor_raca
123	tp_nacionalidade
123	tp_st_conclusao
123	tp_ano_concluiu
123	tp_escola
123	tp_ensino
123	in_treineiro
123	co_municipio_esc
A-Z	no_municipio_esc
123	co_uf_esc
A-Z	sg_uf_esc
123	tp_dependencia_adm_esc
123	tp_localizacao_esc
123	tp_sit_func_esc
123	co_municipio_prova
A-Z	no_municipio_prova
A-Z	co_uf_prova

# 1. Quais os totais de participantes por tipo de escola excluindo-se os treineiros?

---

```
select
  case tp_escola
    when 1 then 'Não respondeu'
    when 2 then 'Pública'
    when 3 then 'Privada'
    else 'Não informado'
  end as tipo_escola,
  COUNT(*) as quantidade
from participante
where
  not in_treineiro
group by
  tp_escola;
```

	tipo_escola	quantidade
>	Não respondeu	1912729
>	Pública	1166540
>	Privada	234619

## 2. Quais os 10 municípios com maior média de notas?

```
with
pontuacao_participante as (
  select p.id_participante, sum(coalesce(nu_nota, 0)) as nota_final
  from
    participante p
    inner join resposta_participante rp on (
      rp.id_participante = p.id_participante
    )
  where
    rp.tp_presenca = 1
    and not p.in_treineiro
  group by
    p.id_participante
)
select e.no_municipio_esc, avg(pp.nota_final) as media_nota
from
  participante p
  inner join pontuacao_participante pp on (
    p.id_participante = pp.id_participante
  )
  inner join escola e on (p.id_escola = e.id_escola)
group by
  e.no_municipio_esc
order by avg(pp.nota_final) desc
limit 10;
```

no_municipio_esc	media_nota double precision
Tucunduva	2728.8500061035156
Celso Ramos	2536.7999267578125
Flora Rica	2524.9666544596353
Aspásia	2487.89990234375
Pedra do Indaiá	2485.300048828125
Jaguarauçu	2479.766642252604
Itamarati de Minas	2472.7999877929688
Curralinhos	2443.5499877929688
Capela Nova	2426.2833353678384
Carmésia	2416.8999633789062

### 3. Quais são os locais de prova com maior número de abstenções?

```
select lp.no_municipio_prova, lp.sg_uf_prova, count(*) as quantidade
from
  participante p
  inner join local_prova lp on (
    p.id_local_prova = lp.id_local_prova
  )
  inner join resposta_participante rp on (
    rp.id_participante = p.id_participante
    and rp.tp_prova = 'CN'
    and rp.tp_presenca = 0
  )
group by
  lp.no_municipio_prova,
  lp.sg_uf_prova
order by count(*) desc
limit 10;
```

no_municipio_prova	sg_uf_prova	quantidade bigint
São Paulo	SP	50424
Rio de Janeiro	RJ	41773
Manaus	AM	29189
Salvador	BA	25360
Fortaleza	CE	23378
Brasília	DF	22403
Belém	PA	20065
Belo Horizonte	MG	17459
São Luís	MA	16263
Recife	PE	13933



## 4. Quais são os municípios com maiores médias de redação por esfera de escola?

```
select
  e.no_municipio_esc as municipio,
  e.co_uf_esc as uf,
  case e.tp_dependencia_adm_esc
    when 1 then 'Federal'
    when 2 then 'Estadual'
    when 3 then 'Municipal'
    when 4 then 'Privada'
    else 'Não informado'
  end as esfera,
  round(avg(rp.nu_notas_redacao), 2) as media_redacao
from
  participante p
  inner join redacao rp on (
    rp.id_participante = p.id_participante
  )
  inner join escola e on (p.id_escola = e.id_escola)
where
  not p.in_treineiro
  and rp.nu_notas_redacao is not null
group by
  e.co_uf_esc,
  e.no_municipio_esc,
  e.tp_dependencia_adm_esc
order by avg(rp.nu_notas_redacao) desc
limit 10;
```

municipio	uf integer	esfera	media_redacao
Curuá	15	Municipal	960.00
Lago da Pedra	21	Municipal	940.00
Itapuca	43	Estadual	940.00
Alfredo Chaves	32	Municipal	940.00
União da Serra	43	Estadual	940.00
Palmeirópolis	17	Privada	940.00
Campos Altos	31	Privada	940.00
Iraquara	29	Privada	940.00
Lajinha	31	Privada	940.00
Inhuma	22	Privada	930.00

# 5. Quais são as faixas etárias com maiores médias?

```
with
nota_total_participante as (
  select p.id_participante, sum(
    coalesce(rp.nu_nota, 0) + coalesce(r.nu_nota_redacao, 0)
  ) as nota_final
  from
    participante p
    left join resposta_participante rp on (
      rp.id_participante = p.id_participante
      and rp.tp_presenca = 1
    )
    left join redacao r on (
      r.id_participante = p.id_participante
    )
  where
    not p.in_treineiro
  group by
    p.id_participante
)
select tfe.descricao, round(avg(ntp.nota_final)) as media_nota
from
  participante p
  inner join tipo_faixa_etaria tfe on (
    p.tp_faixa_etaria = tfe.cd_tipo
  )
  inner join nota_total_participante ntp on (
    p.id_participante = ntp.id_participante
  )
where
  not p.in_treineiro
group by
  tfe.descricao
order by avg(ntp.nota_final) desc
limit 10;
```

17 anos	3942
Menor de 17 anos	3800
18 anos	3567
19 anos	3195
20 anos	2989
21 anos	2799
22 anos	2656
23 anos	2498
24 anos	2364
25 anos	2266

# Visão das Estatísticas

```
create MATERIALIZED view if not exists
inep.estatisticas_participante as

select
    p.tp_faixa_etaria,
    count(*) as quantidade,
    round(avg(ntp.nota_final)) as media_nota,
    stddev(ntp.nota_final) as desvio_padrao,
    min(ntp.nota_final) as menor_nota,
    max(ntp.nota_final) as maior_nota,
    percentile_cont(0.25) within group (order by ntp.nota_final)
as primeiro_quartil,
    percentile_cont(0.5) within group (order by ntp.nota_final)
as mediana,
    percentile_cont(0.75) within group (order by ntp.nota_final)
as terceiro_quartil
from
...
```

estatisticas_participante	
123	tp_faixa_etaria
123	quantidade
123	media_nota
123	desvio_padrao
123	menor_nota
123	maior_nota
123	primeiro_quartil
123	mediana
123	terceiro_quartil

# Funções

```
create or replace function comparar_participantes(
    p_nu_inscrição_a bigint,
    p_nu_inscrição_b bigint
) returns numeric as $$
...
if l_participante_a.tp_faixa_etaria =
l_participante_b.tp_faixa_etaria then
    return
calcular_nota_total(l_participante_a.nu_inscricao) -
calcular_nota_total(l_participante_b.nu_inscricao);
else
    select * into l_stats_a from estatisticas_participante
where tp_faixa_etaria = l_participante_a.tp_faixa_etaria;
    select * into l_stats_b from estatisticas_participante
where tp_faixa_etaria = l_participante_b.tp_faixa_etaria;

    l_z_a :=
(calcular_nota_total(l_participante_a.nu_inscricao) -
l_stats_a.media_nota) / l_stats_a.desvio_padrao;
    l_z_b :=
(calcular_nota_total(l_participante_b.nu_inscricao) -
l_stats_b.media_nota) / l_stats_b.desvio_padrao;
    return l_z_a - l_z_b;
end if;
...
```

5902.90002441406	5070.20001220703	-0.568889786433589
nota_a	5902.90002441406	
nota_b	5070.20001220703	
comparar_participantes	-0.568889786433589	



# Obrigado

---

<https://github.com/vsmoreira/fbd>