



PUC Minas

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS

NÚCLEO DE EDUCAÇÃO A DISTÂNCIA

Pós-graduação *Lato Sensu* em Gestão e Análise Estratégica de Dados

RELATÓRIO TÉCNICO

UTILIZAÇÃO DO POWER BI DESKTOP PARA ANÁLISE DE DADOS HISTÓRICOS
DO EXAME NACIONAL DO ENSINO MÉDIO (ENEM)

Gabriel Canivello de Souza

Belo Horizonte

2021

SUMÁRIO

1. Introdução	3
1.1. Contexto	3
1.2. Objetivos	3
1.3. Público alvo	3
2. Modelos de Dados	3
2.1. Modelo Dimensional	3
2.2. Fatos e Dimensões	3
3. Integração, Tratamento e Carga de Dados	4
3.1. Fontes de Dados	4
3.2. Processos de Integração e Carga (ETL)	4
4. Camada de Apresentação	4
4.1 Dashboard	4
4.2 Análises avançadas	4
5. Registros de Homologação	4
5. Conclusões	5
6. Links	5
REFERÊNCIAS	6

1. Introdução

1.1. Contexto

De acordo com o Brasil Escola, o exame nacional do ensino médio (ENEM) é atualmente a maior prova aplicada no Brasil e segunda maior do mundo segundo o Uol Notícias, e serve como porta de entrada para as principais universidades do país, abrangendo tanto instituições públicas como privadas. Além disso, a avaliação do ENEM é critério para cadastro em programas sociais voltados para ensino, como o Programa Universidade para Todos (ProUni) e Financiamento Estudantil (FIES), que oferecem bolsas de estudo e financiamentos para ensino superior.

Segundo o INEP (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Texeira), o programa foi criado em 1998, inicialmente com o intuito de avaliar o desempenho dos estudantes ao término do ensino médio. Desde então, essas informações são utilizadas para a melhor distribuição e utilização de recursos públicos na área da educação. Em 2009, o exame passou a ser utilizado como mecanismo de acesso à educação superior nas principais universidades. A partir desse momento, muitos defendem que o ENEM passou a ser a principal ferramenta de democratização do acesso à educação superior do país.

1.2. Objetivos

Apesar de ser considerada uma ferramenta tão importante para a educação, não foram encontradas em plataformas governamentais alguma forma de visualização e análise unificada dos dados obtidos em todos os anos de aplicação da prova, ou alguma forma de visualização anual. Nesse sentido, faz-se interessante desenvolver esse trabalho, possibilitando fácil consulta por outros órgãos ou entidades privadas de educação para melhores tomadas de decisões de negócios.

Tem-se como objetivo principal desenvolver 3 dashboards, sendo um com perfil estratégico, um com perfil analítico e outro com perfil operacional. A partir deles, verificar como os fatos na linha do tempo da existência do exame impactaram

nos indicadores diretos do exame, identificar se há padrões nos resultados de acordo com as diferentes regiões, baseados em dados sociais / étnicos, observar a progressão de número de inscritos e quantidade de universidades vinculadas ao exame, entre outros.

Para a realização das análises, será utilizado o programa Power BI Desktop versão 2.100.785.0 64-bit (dezembro de 2021). Nele, será realizado previamente o processo de ETL (*Extract / Transform / Load*) para tratamento dos dados coletados no

no	sítio	eletrônico	do	INEP
----	-------	------------	----	------

 (<https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/enem>) , através de transformações com linguagem DAX e linguagem M, na aba *Power Query*, nativas dentro do próprio Power BI Desktop. Posteriormente, as análises realizadas serão apresentadas em painéis (dashboards), através de indicadores, gráficos e tabelas que poderão dar suporte à tomada de decisão.

1.3. Público alvo

Pode-se definir como público alvo para cada um dos perfis de dashboards, conforme abaixo:

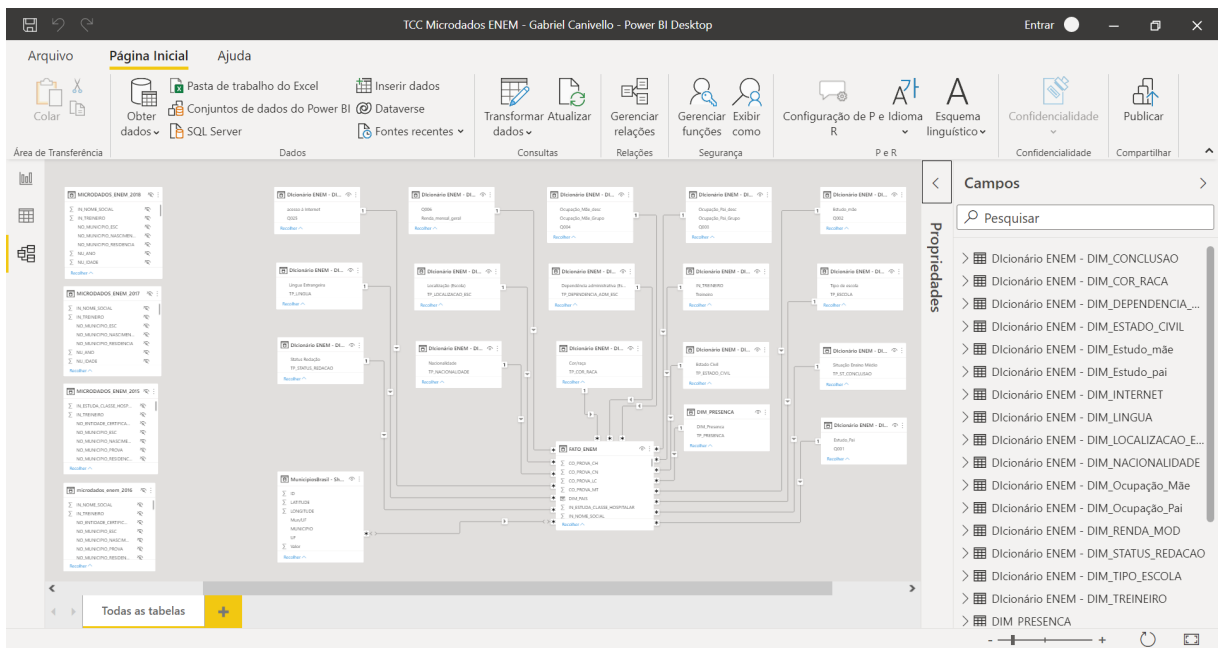
- Operacional: equipe do governo responsável pela impressão e distribuição das provas
- Analítico: equipe do governo responsável pela análise e desempenho dos candidatos
- Estratégico: equipe do governo responsável pela decisão de investimentos em educação no país, de acordo com os dados analisados

Admite-se também a possibilidade desse estudo ser utilizado por quaisquer outras secretarias ou órgãos que tenham relação com investimentos públicos e educação.

2. Modelos de Dados

2.1. Modelo Dimensional

O diagrama do modelo dimensional oferece uma visão sobre como os dados serão organizados para atender às necessidades identificadas por meio dos objetivos do projeto.



2.2. Fatos e Dimensões

As tabelas fatos, dimensão e auxiliares utilizadas no projeto estão descritas a seguir:

Tabela	Tipo	Descrição
FATO_ENEM	Fato	União das tabelas fatos Microdados ENEM 2020, 2019, 2018, 2017, 2016, 2015.
MunicípiosBrasil	Dimensão	Tabela com lat e long dos municípios do Brasil
DIM_ESTADO_CIVIL	Dimensão	Contem correspondência do estado civil do candidato
DIM_COR_RACA	Dimensão	Contem correspondência da cor/raça do candidato
DIM_CONCLUSAO	Dimensão	Contem correspondência de informação de conclusão do candidato
DIM_TIPO_ESCOLA	Dimensão	Contem correspondência do tipo de escola do candidato
DIM_TREINEIRO	Dimensão	Contem correspondência se candidato é treineiro

DIM_DEPENDENCIA_ADM_ESC	Dimensão	Contem correspondência da dependência do escola do candidato
DIM_LOCALIZACAO_ESC	Dimensão	Contem correspondência da localização do candidato
DIM_LINGUA	Dimensão	Contem correspondência da lingua estrangeira escolhida do candidato
DIM_STATUS_REDACAO	Dimensão	Contem correspondência do status da redação do candidato
DIM_ESTUDOS_PAI	Dimensão	Contem correspondência de informação dos estudos do pai do candidato
DIM_ESTUDOS_MAE	Dimensão	Contem correspondência de informação dos estudos da mãe do candidato
DIM_RENDA_MOD	Dimensão	Contem correspondência de informação da renda familiar do candidato
DIM_INTERNET	Dimensão	Contem correspondência se candidato tem internet

3. Integração, Tratamento e Carga de Dados

3.1. Fontes de Dados

Descrição das bases de dados ou arquivos utilizados pelo projeto como fonte, apresentando possíveis diagramas dos bancos de dados relacionais.

Serão utilizados os microdados disponibilizados no sítio eletrônico do INEP (<https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/enem>) relativos aos exames anuais da prova do ENEM. Serão aproveitados dos arquivos extensão .zip de cada ano do exame os arquivos contendo os microdados (extensão .CSV) e o dicionário (extensão .xlsx).

3.2. Processos de Integração e Carga (ETL)

3.2.1 União de tabelas

Os arquivos dos microdados relativos a cada ano de exame serão importados para o Power BI Desktop para tratamento dos mesmos. Como os microdados de

cada ano seguem aproximadamente o mesmo padrão de campos, as tabelas de cada ano serão combinadas em uma única tabela. Os microdados em geral são arquivos pesados (aproximadamente 3Gb cada um), e a primeira tentativa de combinação dos arquivos deu errado (aproximadamente 10 arquivos), pois esgotou a memória do programa. Em uma nova tentativa, cada tabela de microdado foi combinada individualmente, criando portanto uma tabela única nomeada “FATO_ENEM”.

Table: Table.Combine({#Tipo Alterado", MICRODADOS_ENEM_2018})

	123 NU_INSCRICAO	123 NU_ANO	123 CO_MUNICIPIO_RESIDENCIA	123 NO_MUNICIPIO_RESIDENCIA	123 CO_UF
1	1,90002E+11	2019	3552205	Sorocaba	
2	1,90001E+11	2019	2910800	Feira de Santana	
3	1,90001E+11	2019	2304400	Fortaleza	
4	1,90001E+11	2019	1721000	Palmas	
5	1,90001E+11	2019	3118601	Contagem	
6	1,90002E+11	2019	5107602	Rondonópolis	
7	1,90001E+11	2019	2924009	Paulo Afonso	
8	1,90002E+11	2019	3550308	São Paulo	
9	1,90002E+11	2019	3550308	São Paulo	
10	1,90002E+11	2019	3550308	São Paulo	
11	1,90002E+11	2019	3530607	Mogi das Cruzes	
12	1,90001E+11	2019	2409332	Santa Maria	
13	1,90002E+11	2019	4101804	Araucária	
14	1,90001E+11	2019	2304400	Fortaleza	
15	1,90001E+11	2019	3170206	Uberlândia	
16	1,90001E+11	2019	2611606	Recife	
17	1,90001E+11	2019	2911709	Guanambi	
18	1,90001E+11	2019	2611606	Recife	
19	1,90001E+11	2019	2105401	Itapecuru Mirim	
20	1,90001E+11	2019	2927408	Salvador	
21	1,90002E+11	2019	3550308	São Paulo	

Na tabela FATO_ENEM, foi possível adicionar dados de 2019, 2018 e 2017, apresentando erro no carregamento dos microdados do ano de 2016. Ainda assim, a cada tabela adicionada, demorava-se por volta de 40 min por cada arquivo, pois o programa processava novamente cada uma dos arquivos e consultas criadas.

Existem alterações pendentes em suas consultas que ainda não foram aplicadas.

Aplicar alterações Descartar as alterações

Carregar

microdados_enem_2016
Um erro ao carregar uma tabela anterior cancelou o carregamento.

FATO_ENEM
Erro de memória: Falha de Alocação de Memória. Tente simplificar ou reduzir o número de consultas.

Fechar

Tabela: FATO_ENEM (17.340.358 linhas) Coluna: NU_ANO (3 valores distintos)

A partir dessas tentativas, foi decidido utilizar para a fase de ETL o programa Tableau Prep Builder - versão 2021.4.3. Apenas os campos considerados mais importantes foram mantidos, na intenção de reduzir o tamanho do arquivo de saída. Inicialmente foi gerado um arquivo combinado dos microdados dos anos de 2019, 2018, 2017 e 2016.

Tableau Prep Builder - TCC Prep - A avaliação expira em 13 dias

Arquivo Editar Fluxo Servidor Ajuda

Alertas (0)

100%

FATO_ENEM 69 campos

Salvar saída no

Arquivo

Procurar

Nome

FATO_ENEM

Local

C:\Users\jadedp\Documents\Meu Repositório do Tableau Prep\Fontes de dados

Tipo de saída

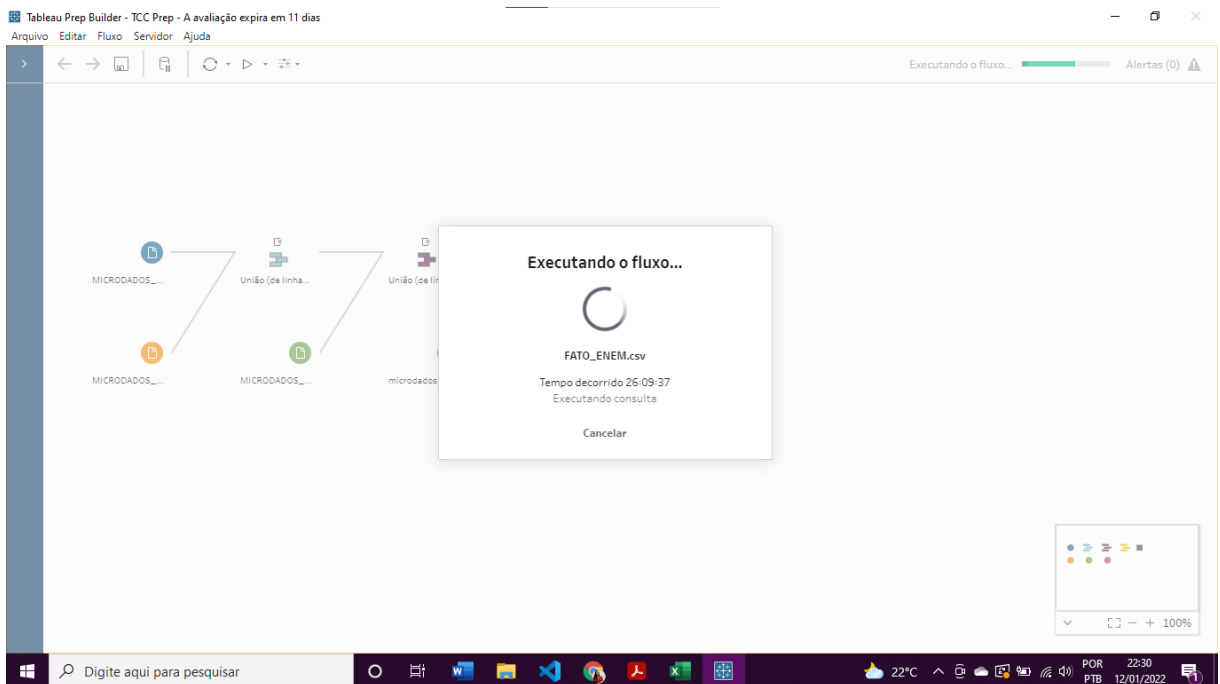
Exportação de dados do Tableau Prep

Executar fluxo

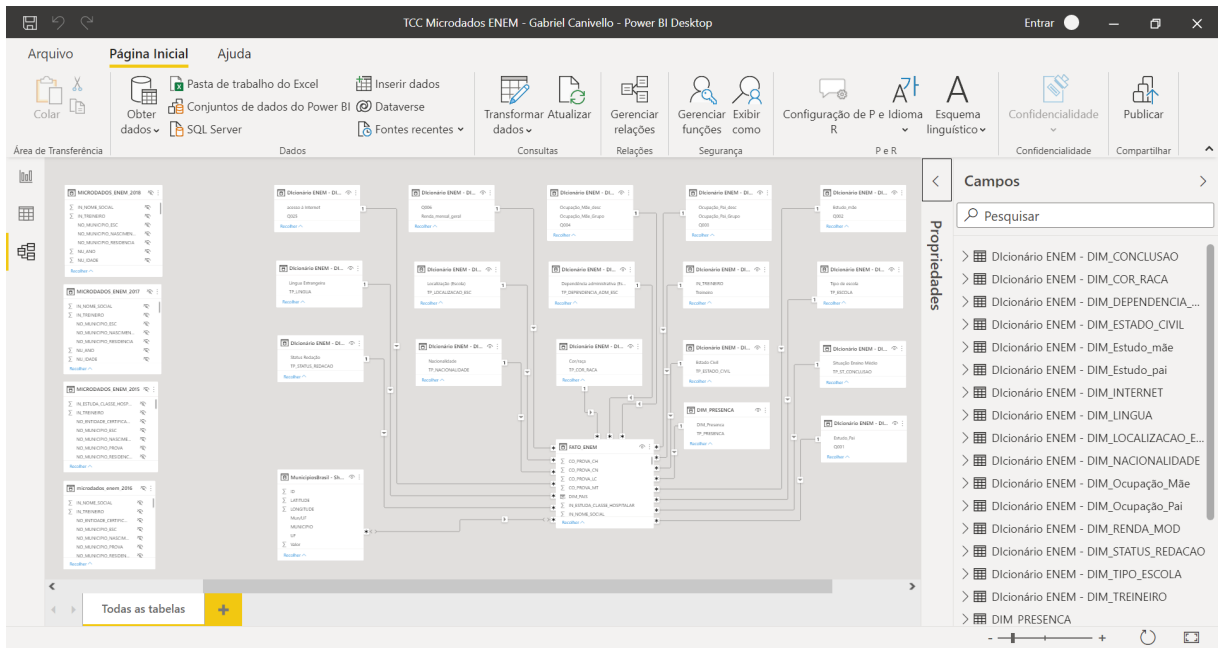
Salvar em FATO_ENEM.hyper

De: Fluxo		Para: Tabela		Status
Tipo	Nome de campo		Tipo	Nome de campo
Abc	Table Names	→	Abc	Table Names
#	NU_INSCRICAO	→	#	NU_INSCRICAO
#	NU_ANO	→	#	NU_ANO
#	CO_MUNICIPIO_RESIDENCIA	→	#	CO_MUNICIPIO_RESIDENCIA
Abc	NO_MUNICIPIO_RESIDENCIA	→	Abc	NO_MUNICIPIO_RESIDENCIA
#	CO_UF_RESIDENCIA	→	#	CO_UF_RESIDENCIA
Abc	SG_UF_RESIDENCIA	→	Abc	SG_UF_RESIDENCIA
#	NU_IDADE	→	#	NU_IDADE

A tentativa com o Tableau Prep Builder também não funcionou devido ao tamanho do arquivo de saída. O programa tentou executar o fluxo por mais de 26h, sem sucesso. Dessa forma, toda a preparação dos dados foi mantida no Power BI, conforme planejado anteriormente.



Por fim, reduzindo ainda mais o número de campos mantidos no programa, foi possível acrescentar os dados de 2020, 2019, 2018, 2017, 2016 e 2015. As tabelas foram unificadas na tabela FATO_ENEM, e as dimensões contidas no dicionário adicionadas e conectadas à tabela fato.



3.2.2 Transformação das dimensões

- TP_ESTADO_CIVIL

A dimensão TP_ESTADO_CIVIL em 2019 e 2020 assume valores divergentes quando comparado aos outros anos, relativo a seus significados baseado nos dicionários de cada ano. Devido a isso, foi necessário modificar os valores da dimensão de 2018, 2017, 2016 e 2015 para equivaler ao dos outros anos, somando 1 aos valores contidos nessas tabelas.

tabela 1 - Estado Civil

- Q006 - Renda Mensal

Na dimensão Q006 que representa a renda mensão da família do candidato, os intervalos assumem diferentes valores. No entanto, todos têm como base o valor do salário mínimo do ano de aplicação da prova. Dessa forma, foi possível adotar uma referência única para todos os anos, conforme tabela 2.

Q006	Renda_mensal_2019	Renda_mensal_2018	Renda_mensal_2017	Renda_mensal_geral
A	Nenhuma renda.	Nenhuma renda.	Nenhuma renda.	Nenhuma renda
B	Até R\$ 998,00.	Até R\$ 954,00.	Até R\$ 937,00.	Até 1 salário mínimo

C	De R\$ 998,01 até R\$ 1.497,00.	De R\$ 954,01 até R\$ 1.431,00.	De R\$ 937,01 até R\$ 1.405,50.	De 1 até 1,5 salários mínimos
D	De R\$ 1.497,01 até R\$ 1.996,00.	De R\$ 1.431,01 até R\$ 1.908,00.	De R\$ 1.405,51 até R\$ 1.874,00.	De 1,5 até 2 salários mínimos
E	De R\$ 1.996,01 até R\$ 2.495,00.	De R\$ 1.908,01 até R\$ 2.385,00.	De R\$ 1.874,01 até R\$ 2.342,50.	De 2 até 2,5 salários mínimos
F	De R\$ 2.495,01 até R\$ 2.994,00.	De R\$ 2.385,01 até R\$ 2.862,00.	De R\$ 2.342,51 até R\$ 2.811,00.	De 2,5 até 3 salários mínimos
G	De R\$ 2.994,01 até R\$ 3.992,00.	De R\$ 2.862,01 até R\$ 3.816,00.	De R\$ 2.811,01 até R\$ 3.748,00.	De 3 até 3,5 salários mínimos
H	De R\$ 3.992,01 até R\$ 4.990,00.	De R\$ 3.816,01 até R\$ 4.770,00.	De R\$ 3.748,01 até R\$ 4.685,00.	De 3,5 até 4 salários mínimos
I	De R\$ 4.990,01 até R\$ 5.988,00.	De R\$ 4.770,01 até R\$ 5.724,00.	De R\$ 4.685,01 até R\$ 5.622,00.	De 4 até 4,5 salários mínimos
J	De R\$ 5.988,01 até R\$ 6.986,00.	De R\$ 5.724,01 até R\$ 6.678,00.	De R\$ 5.622,01 até R\$ 6.559,00.	De 4,5 até 5 salários mínimos
K	De R\$ 6.986,01 até R\$ 7.984,00.	De R\$ 6.678,01 até R\$ 7.632,00.	De R\$ 6.559,01 até R\$ 7.496,00.	De 5 até 5,5 salários mínimos
L	De R\$ 7.984,01 até R\$ 8.982,00.	De R\$ 7.632,01 até R\$ 8.586,00.	De R\$ 7.496,01 até R\$ 8.433,00.	De 5,5 até 6 salários mínimos
M	De R\$ 8.982,01 até R\$ 9.980,00.	De R\$ 8.586,01 até R\$ 9.540,00.	De R\$ 8.433,01 até R\$ 9.370,00.	De 6 até 6,5 salários mínimos
N	De R\$ 9.980,01 até R\$ 11.976,00.	De R\$ 9.540,01 até R\$ 11.448,00.	De R\$ 9.370,01 até R\$ 11.244,00.	De 6,5 até 7 salários mínimos
O	De R\$ 11.976,01 até R\$ 14.970,00.	De R\$ 11.448,01 até R\$ 14.310,00.	De R\$ 11.244,01 até R\$ 14.055,00.	De 7 até 7,5 salários mínimos
P	De R\$ 14.970,01 até R\$ 19.960,00.	De R\$ 14.310,01 até R\$ 19.080,00.	De R\$ 14.055,01 até R\$ 18.740,00.	De 7,5 até 8 salários mínimos
Q	Mais de R\$ 19.960,00.	Mais de R\$ 19.080,00.	Mais de R\$ 18.740,00.	Mais de 8 salários mínimos

Tabela 2 - Renda Mensal

4. Camada de Apresentação

4.1 Dashboard

A apresentação dos dashboards se dá em 3 diferentes abas, sendo uma voltada para o time operacional, uma para o time tático e outra para o time estratégico.

Como os microdados de todos os anos foram unidos em 1 só tabela com todos os parâmetros de análise, todos os filtros utilizados nas visualizações são aplicáveis a quaisquer gráficos ou dashboards. Seguem os principais filtros utilizados:

- NU_ANO: Ano de aplicação da prova
- SG_UF_RESIDENCIA: UF (Estado) de residência do aluno
- NU_IDADE: Idade do aluno
- TP_SEXO: Sexo do aluno
- TIPO DE ESCOLA: Tipo de escola do aluno (Pública / Particular / Exterior)
- COR/RAÇA: Cor/Raça do aluno (Branco, Negro, Índio, etc...)
- ACESSO À INTERNET:: Se aluno tem acesso à internet em sua residência (Sim / Não)

4.1.1 Dashboard Operacional

O dashboard operacional foi pensado para facilitar e melhorar a visualização dos municípios de aplicação de provas no Brasil. As principais informações fornecidas são baseadas em local de residência dos alunos e a relação de faltantes por ano, o que permitiria ter certa previsibilidade de número de provas que seriam desperdiçadas e do dimensionamento de mão de obra para aplicação do exame.

- Contagem de inscritos por município - prova

O mapa de bolhas representa a quantidade de inscritos por município.

- O campo de valor é a contagem simples do número de inscritos
- O campo Dicas de ferramenta mostra a porcentagem do total geral para a contagem de inscritos
- O campo Localização é um concatenado do País + UF + Município de residência do aluno. O concatenado foi necessário para melhor precisão na representação do gráfico de mapa.

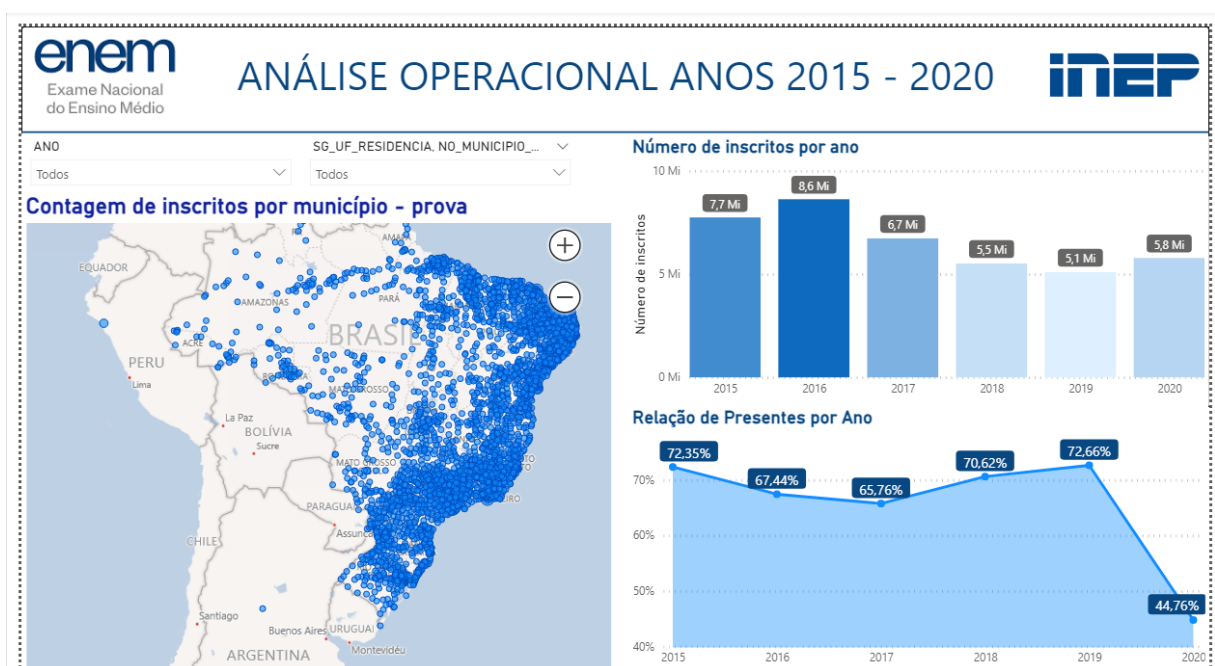
- Número de Inscritos por ano

- O campo eixo compartilhado representa a dimensão de ano de aplicação da prova

- O campo valores de coluna é a contagem simples do número de alunos inscritos
- Para representar as cores do gráfico, foi aplicada uma função para variar o tom de azul conforme o número de inscritos. Quanto maior esse número, mais escura é a cor azul.
- Relação de presentes por ano

O gráfico de área representa a quantidade de alunos que concluíram a prova.

- O campo eixo representa a dimensão de ano de aplicação da prova
- O campo valores é a porcentagem do número de presentes na prova. Seu cálculo considera um aluno presente sendo aquele que não faltou em nenhum dos 2 dias de aplicação de prova, calculado na coluna TP_PRESENCA. Em seguida, calculamos a contagem de presentes / contagem de inscritos, no campo de medida % Presentes2.



Dashboard operacional - Imagem 1

- Número de inscritos por localidade

A tabela apresenta os números de inscritos por localidade, sendo essa em hierarquia de Estado e Cidade, sendo possível fazer Drill down e Drill up.

- Linhas representam UF e municípios
- Colunas representam os anos de aplicação da prova

- Número de concludentes / eliminados / faltantes por localidade

A tabela apresenta os números de inscritos por localidade, sendo essa em hierarquia de Estado e Cidade, sendo possível fazer Drill down e Drill up.

- Linhas representam UF e municípios
- Colunas representam as categorias de alunos que realizaram a prova, sendo:
 - Concludente: Alunos que concluíram a realização dos 2 dias de prova
 - Eliminados: Alunos que foram eliminados em alguma das 2 provas
 - Faltantes: Alunos que faltaram em alguma das 2 provas

SG_UF_RESIDENCIA	2015	2016	2017	2018	2019	2020	Total
AC	55406	60220	53659	37785	38229	41824	287123
AL	137247	146277	121049	89670	91707	102551	688501
AM	156595	193830	165171	123755	118087	163426	920864
AP	54693	60064	50852	40109	42140	47263	295121
BA	580401	665192	493468	401328	398095	447691	2986175
CE	459430	516359	366345	329406	295870	325680	2293090
DF	160440	166007	123169	104531	94352	116932	765431
ES	153992	172858	137531	112173	102435	105812	784801
GO	263225	286980	219996	191029	170147	211069	1342446
MA	268361	326960	287971	218462	219673	238272	1559699
MG	869386	954219	731308	587471	538896	577211	4258491
MS	132211	140082	92198	72319	70157	84548	591515
MT	149433	163250	126254	96762	88349	101727	725775
PA	367330	441188	356845	283861	281518	330883	2061625
PB	204060	222920	189876	152049	147780	164965	1081650
PE	390112	441345	367435	304302	271727	315569	2090490
PI	153892	186808	147553	118350	121181	134678	862462
PR	338195	420641	287271	236533	210080	239635	1732355
RJ	544002	548687	439114	382078	338710	387480	2640071
RN	179907	194740	159076	124039	119179	131225	908166
RO	91044	100101	80433	62266	58885	69594	462323
RR	22221	24111	20215	14029	12947	16885	110408
RS	418695	419580	294973	243555	218763	249130	1844696
Total	7746427	8627367	6731341	5513747	5095270	5783109	39497261

SG_UF_RESIDENCIA	Concludente	Eliminado	Faltante	Total
AC	188271	154	98698	287123
AL	479117	250	209134	688501
AM	509560	235	411069	920864
AP	187911	136	107074	295121
BA	1954325	1122	1030728	2986175
CE	1562270	964	729856	2293090
DF	493547	444	271440	765431
ES	516145	344	268312	784801
GO	870234	666	471546	1342446
MA	1069202	837	489660	1559699
MG	2821159	1984	1435348	4258491
MS	359575	261	231679	591515
MT	447525	396	277854	725775
PA	1389815	890	670920	2061625
PB	772590	448	308612	1081650
PE	1411574	825	678091	2090490
PI	619947	323	242192	862462
PR	1133364	690	598301	1732355
RJ	1725731	1030	913310	2640071
RN	631254	380	276532	908166
RO	289694	213	172416	462323
RR	65227	55	45126	110408
RS	1205566	705	638425	1844696
Total	26034538	16717	13446006	39497261

Dashboard operacional - Imagem 2

4.1.2 Dashboard Tático

O dashboard tático foi pensado para realizar análises focadas no desempenho e assiduidade dos alunos. Apresenta uma visão mais granular quando comparado com o dashboard estratégico.

- Contagem de inscritos por UF

O mapa de bolhas representa a quantidade de inscritos por UF.

- O campo de valor é a contagem simples do número de inscritos
- O campo Dicas de ferramenta mostra a porcentagem do total geral para a contagem de inscritos

- O campo Localização é um concatenado do País + UF de residência do aluno. O concatenado foi necessário para melhor precisão na representação do gráfico de mapa.

- Número de Inscritos por ano

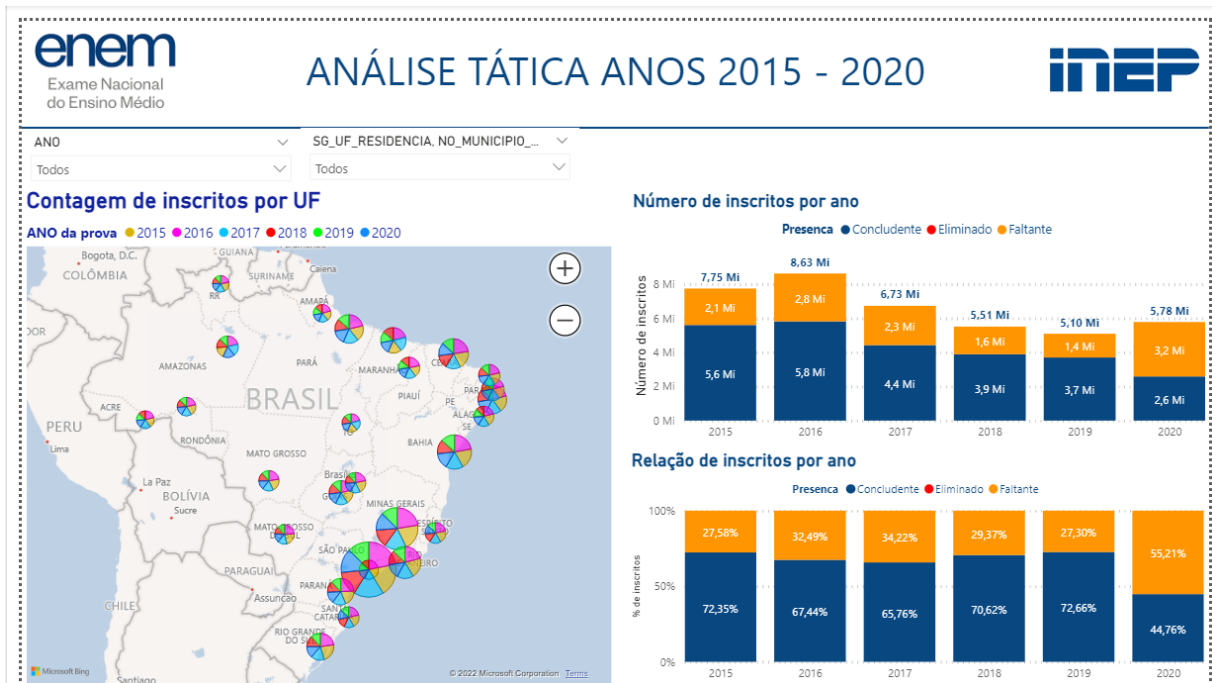
Gráfico de colunas empilhadas que representa a contagem de inscritos por categoria de alunos quanto à presença e por ano

- O campo eixo compartilhado representa a dimensão de ano de aplicação da prova
- O campo série da coluna representa a categoria de aluno quanto a presença, já explicada anteriormente
- O campo valores de coluna é a contagem simples do número de alunos inscritos

- Relação de Inscritos por ano

Gráfico de colunas 100% empilhadas que representa a porcentagem de inscritos por categoria de alunos quanto à presença e por ano

- O campo eixo compartilhado representa a dimensão de ano de aplicação da prova
- O campo série da coluna representa a categoria de aluno quanto a presença, já explicada anteriormente
- O campo valores de coluna é a porcentagem do número de alunos inscritos em relação ao total de inscritos

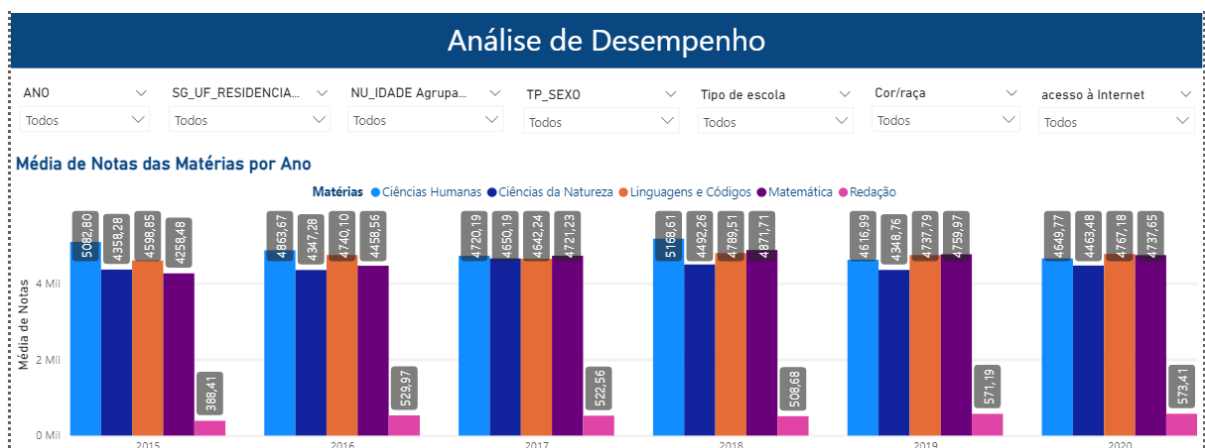


Dashboard tático- Imagem 1

- Média de notas das matérias por ano

Gráfico de colunas clusterizado que representa a média por matéria por ano de todos os alunos inscritos. Gráfico serve de base para análise dos gráficos seguintes.

- O campo eixo representa a dimensão de ano de aplicação da prova
- O campo valores representa a média de cada matéria

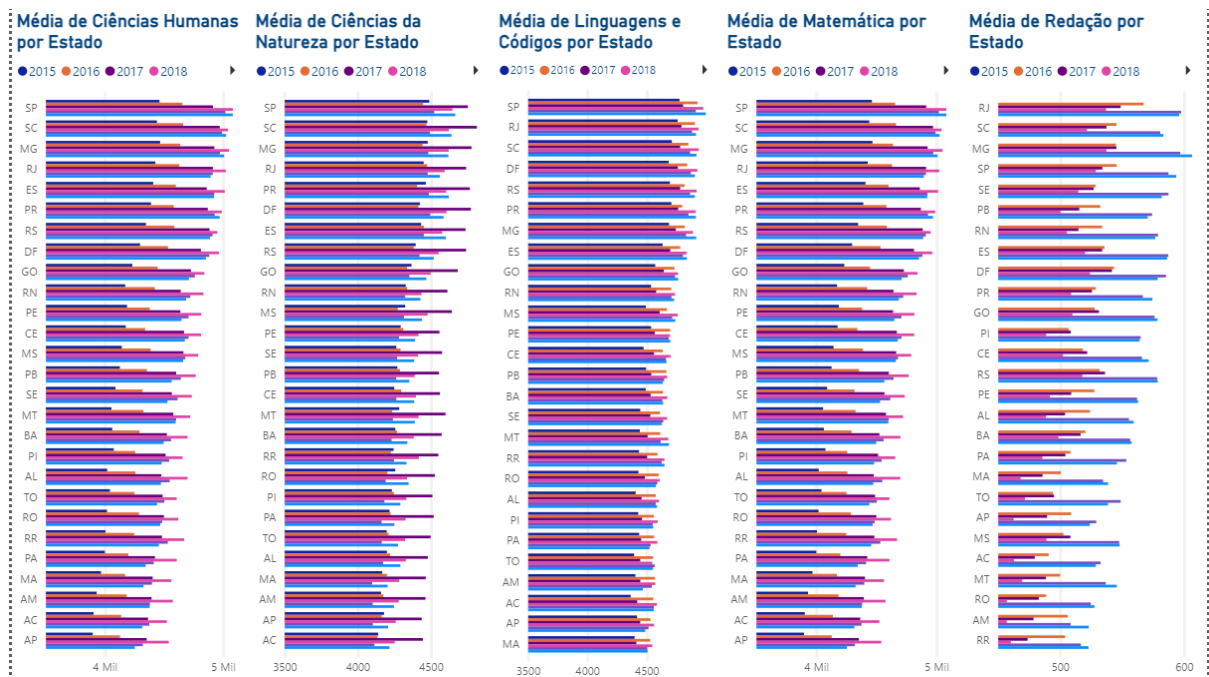


Dashboard tático- Imagem 2

- Média de Ciências Humanas / Ciências da Natureza / Linguagens e Códigos / Matemática / Redação por Estado

Gráfico de barras clusterizado (na horizontal) que representa a média dos alunos por ano por estado em que residem. Ordem decrescente de média.

- O campo eixo representa a dimensão de ano de aplicação da prova
- O campo valores representa a média de cada matéria

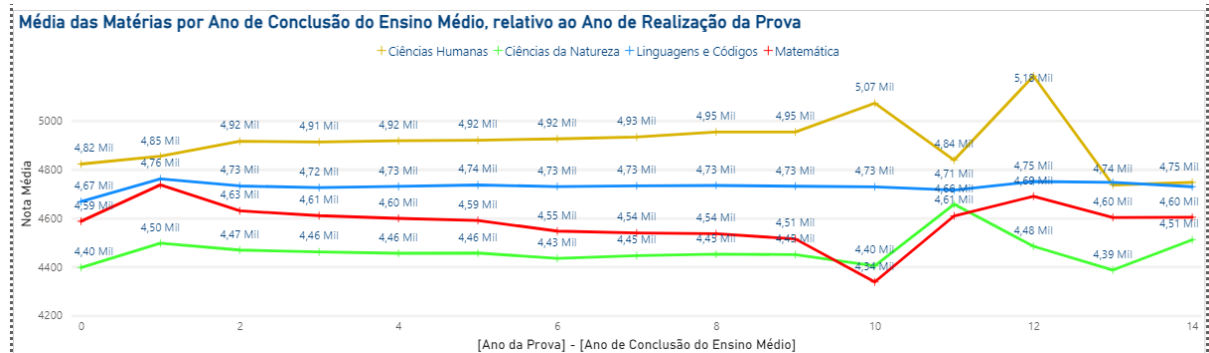


Dashboard táctico- Imagem 3

- Média das matérias por ano de conclusão do ensino médio, relativo ao ano de realização da prova

Gráfico de linhas onde cada linha representa a média das matérias, divididas pela numeração de 0 a 14. Essa numeração representa a quantos anos anteriores ao ano de aplicação da prova o candidato concluiu o ensino médio. Esse campo já consta nos microdados do enem com o nome de TP_ANO_CONCLUIU. Matematicamente, poderia ser representado como o cálculo $[\text{ano de aplicação da prova}] - [\text{ano de conclusão do ensino médio}]$, onde 0 significa que o aluno está concluindo o ensino médio no mesmo ano de aplicação da prova.

- O campo eixo representa o campo TP_ANO_CONCLUIU
- O campo valores representa a média de cada matéria

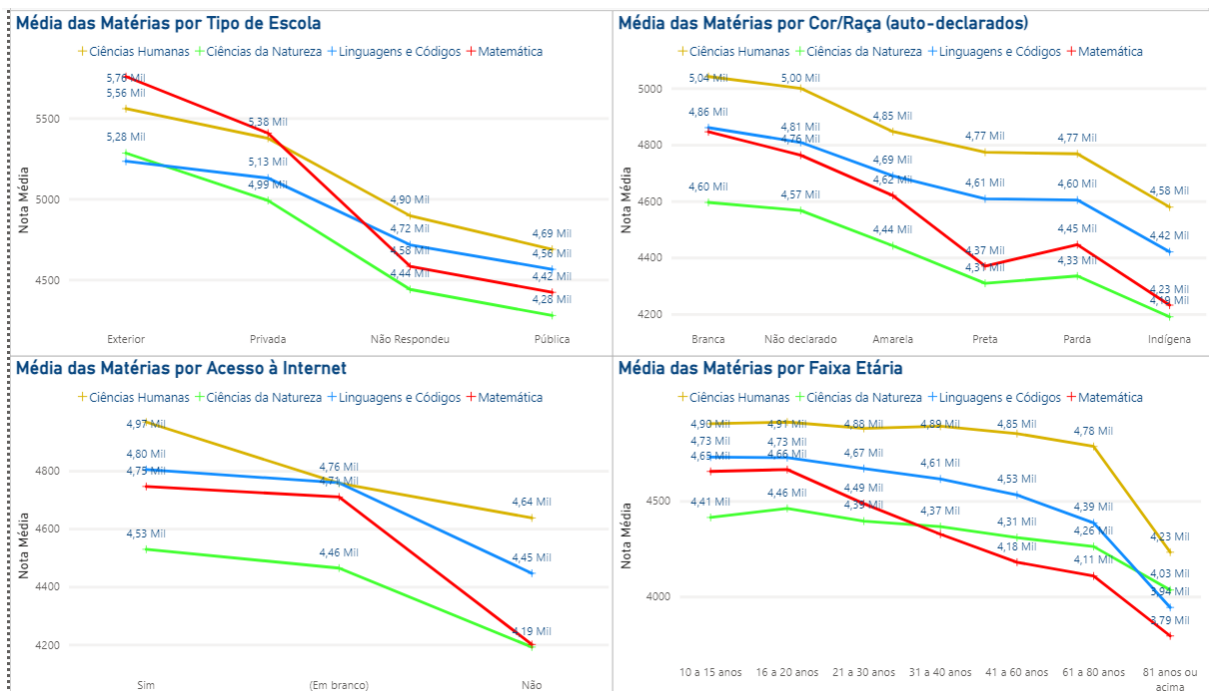


Dashboard tático- Imagem 4

- Média das matérias por tipo de escola / cor/raça / acesso à internet / Faixa etária

Gráficos de linhas onde cada linha representa a média das matérias, divididas pelas classificações acima.

- O campo eixo representa cada uma das classificações escolhidas. Todas as informações são colhidas pelo enem durante o processo de inscrição do aluno
- O campo valores representa a média de cada matéria



Dashboard tático- Imagem 5

4.1.3 Dashboard Estratégico

O dashboard estratégico foi pensado para realizar análises focadas na diversidade dos alunos inscritos. Apresenta uma visão mais macro do contexto em que a prova está inserida, pensando no âmbito sócio-cultural.

- Cartões:

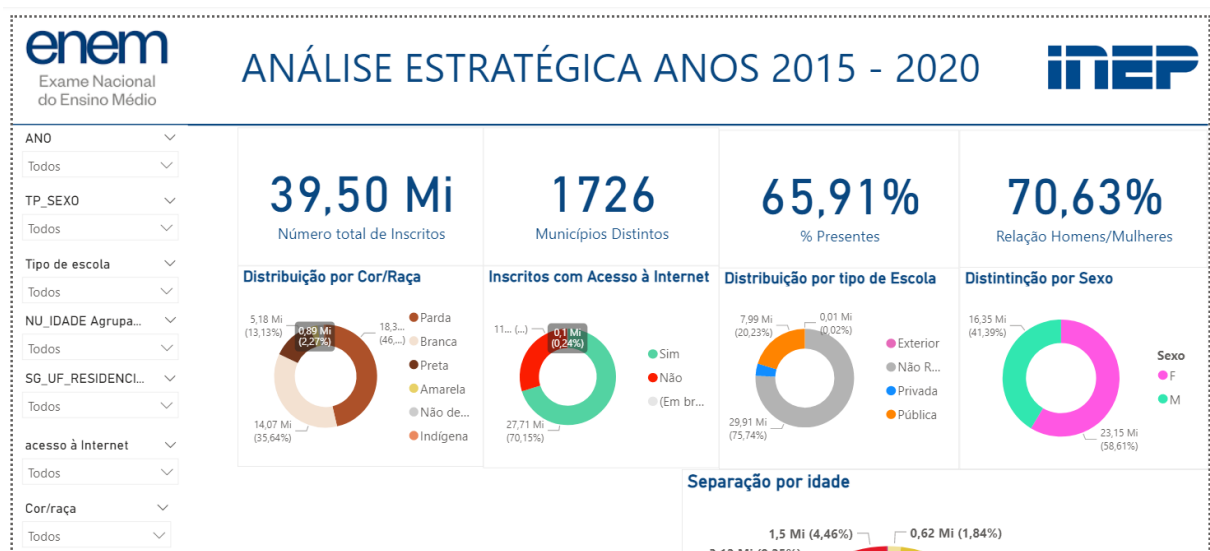
Representam medidas acumuladas importantes de serem analisadas:

- Número total de inscritos:
- Número de municípios de aplicação das provas (contagem distinta)
- % de presentes (acumulada)
- Relação de homens e mulheres, calculadas no campo % sexo:

$$\% \text{ sexo} = [\text{contagem homens}] / [\text{contagem mulheres}]$$

- Gráficos de rosca:

Representam a divisão dos alunos conforme as classificações de cor/raça / acesso à internet / tipo de escola / sexo.



Dashboard Estratégico - Imagem 1

- Relação de treineiros / faltantes por ano:

Gráfico de área que representa a porcentagem de treineiros e faltantes por ano de aplicação de prova, sendo treineiro aquele aluno que faz o exame apenas para

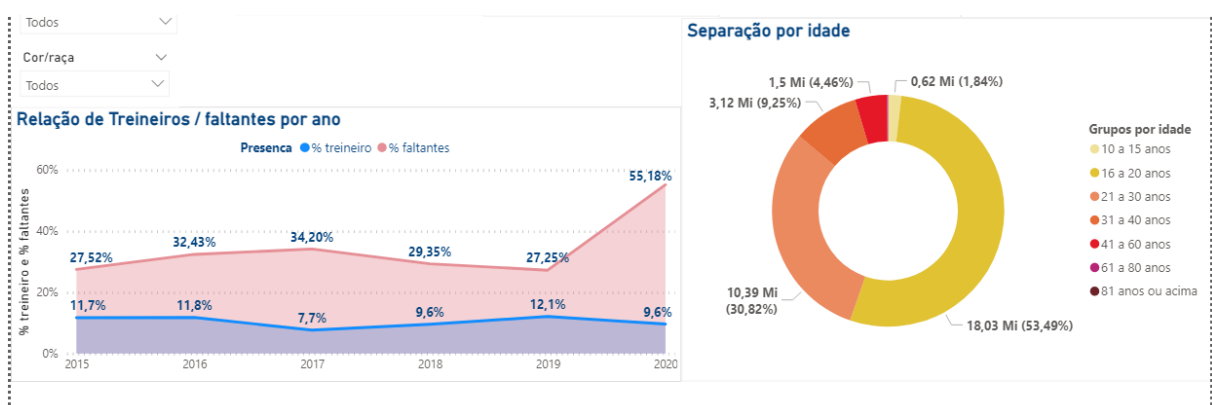
treinamento e o faltante aquele aluno que faltou em algum dos 2 dias de aplicação de prova.

- O campo eixo representa a dimensão de ano de aplicação da prova
- O campo valor constam o cálculos conforme abaixo:
 - % treineiro = [soma dos treineiros] / [contagem de inscritos]
 - % treineiro = [contagem de faltantes do campo TP_PRESENCA] / [contagem de inscritos]

- Separação por idade:

Gráfico de rosca que representa a divisão dos alunos conforme o agrupamento por idade. Levando em consideração a quantidade de alunos inscritos por cada um dos grupos, foi escolhido separar conforme abaixo para melhor representação no gráfico:

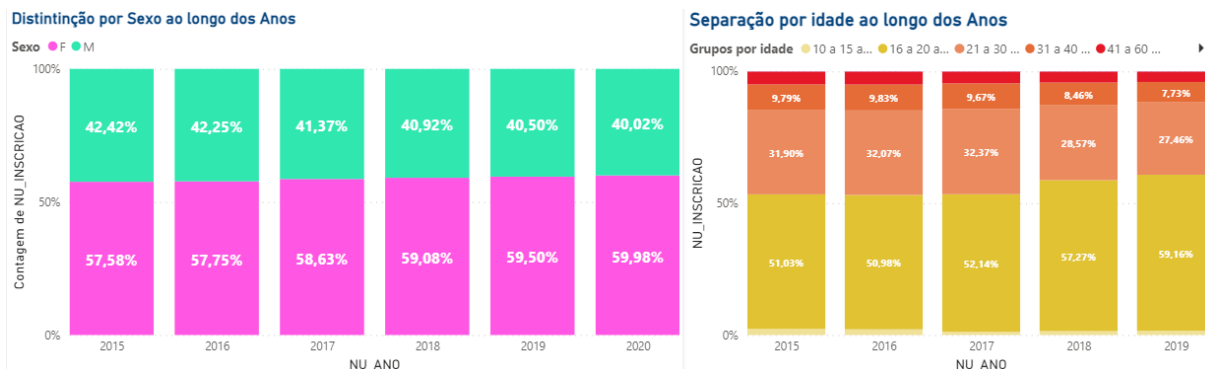
- 10 a 15 anos
- 16 a 20 anos
- 21 a 30 anos
- 31 a 40 anos
- 41 a 60 anos
- 61 a 80 anos
- 80 anos ou acima



Dashboard Estratégico - Imagem 2

- Separação por Sexo / Separação por idade ao longo dos anos

Gráfico de colunas 100% empilhadas que representa a porcentagem de homens e mulheres / agrupamentos por idade do total de alunos inscritos ao longo dos anos.



Dashboard Estratégico - Imagem 3

4.2 Análises avançadas

Padrões e/ou Tendências: análises de dados criadas por meio de ferramentas estatísticas ou pela aplicação de técnicas de aprendizado de máquina. As análises podem ser independentes ou incorporadas no dashboard criado anteriormente

Os dashboards criados permitem realizar diversas análises sobre a realização das provas do ENEM. Diferentes cruzamentos de dados e filtros foram usados para encontrar padrões, que são apresentados em seguida:

- Análise da proporção de alunos faltantes

Como era de se esperar, o ano de 2020 foi o ano do início da pandemia da COVID-19 e a prova deste ano foi diretamente impactada na proporção de alunos faltantes. A proporção flutuava entre 27% e 34%, e subiu para 55% no ano de pandemia. No entanto, a quantidade de inscritos não foi impactada, mesmo com o período de inscrição (Maio/2020) tendo sido após o início da pandemia (Fevereiro/2020).

Número de inscritos por ano



Relação de inscritos por ano



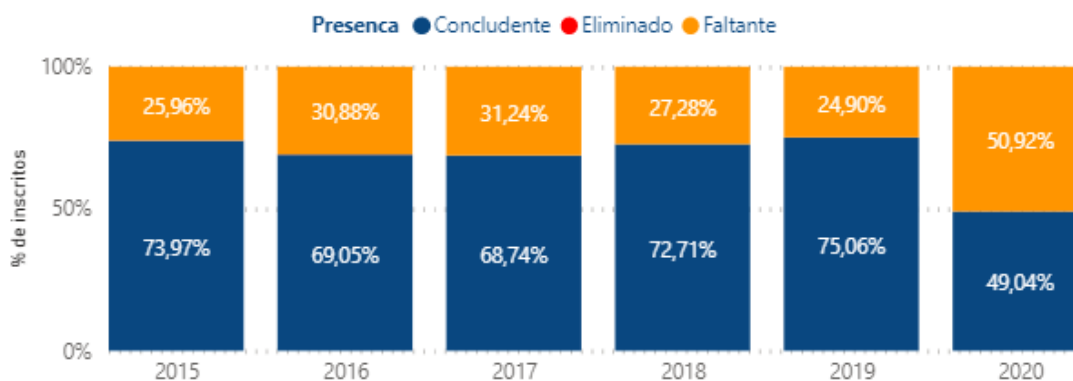
Quando analisada a relação de faltantes por região, vemos que a região Norte e Centro-Oeste têm os maiores números. Podemos levantar algumas hipóteses para explicar os números, como o momento em que cada estado vivenciava em relação ao número de infecções, ou a falta de infra-estrutura (internet) para os alunos realizarem o ENEM Online, sendo o ano de 2020 o primeiro ano de aplicação dessa modalidade. Segue a proporção de faltantes para cada região em 2022:

- Região Norte: 57,80%
- Região Nordeste: 51,47%
- Região Centro-Oeste: 58,23%
- Região Sudoeste: 56,59%
- Região Sul: 55,49%

Outra análise que podemos realizar é a verificação quanto aos grupos de alunos. É possível perceber que a proporção de alunos faltantes de grupos com renda *per capita* menor é maior, quando comparado com grupos com renda maior.

- Em média, população negra tem **16%** mais faltantes que população branca
- Em média, população negra tem **25%** mais faltantes que população branca

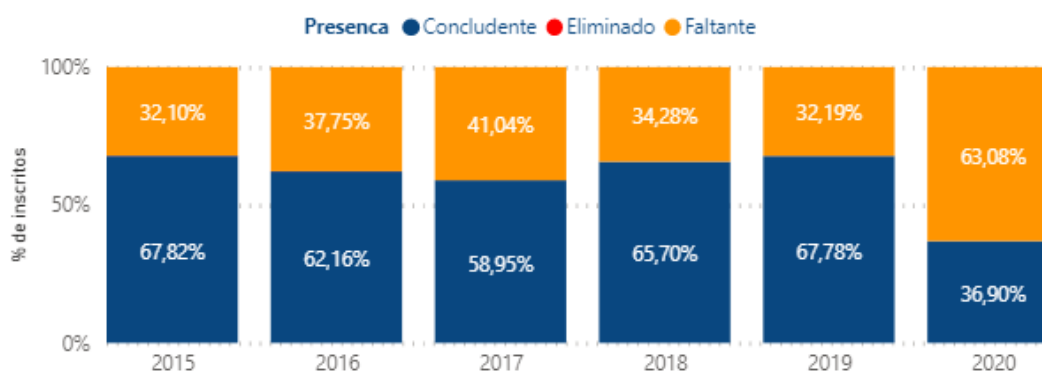
→ **Relação de faltantes em população branca:**



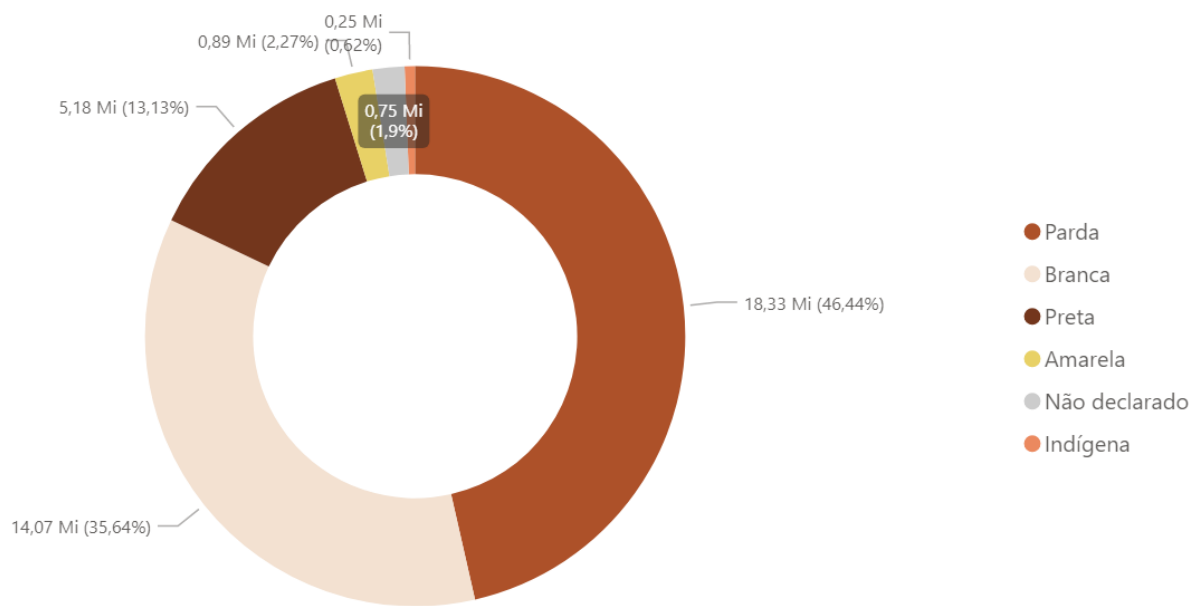
→ **Relação de faltantes em população negra:**



→ **Relação de faltantes em população Indígena:**



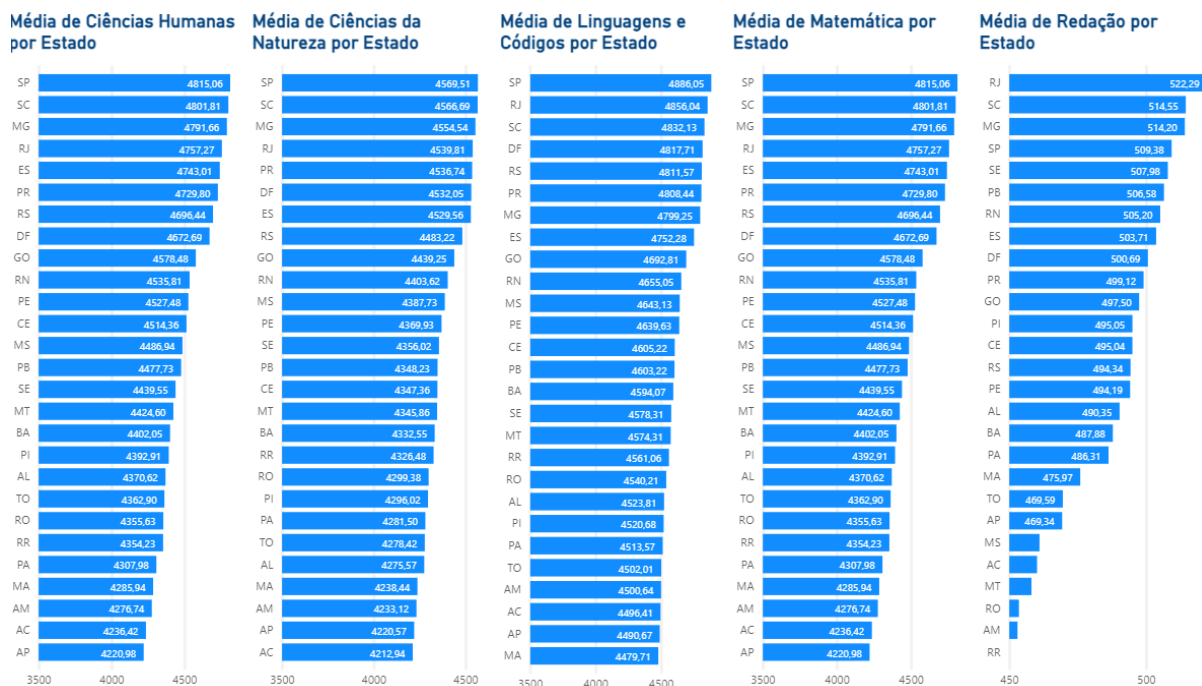
Os 2 grupos (negros e indígenas) representam juntos aproximadamente 14% dos inscritos do ENEM.



O número de alunos faltantes representa os alunos que deixam de cursar uma faculdade e ter um salário melhor no futuro, reduzindo a diferença social. Se a proporção de faltantes é maior em grupos mais desfavorecidos, podemos concluir que essa diferença tende a aumentar, e só deverá refletir no fim dessa década, que é o tempo do curso de graduação.

- Análise de desempenho por estado

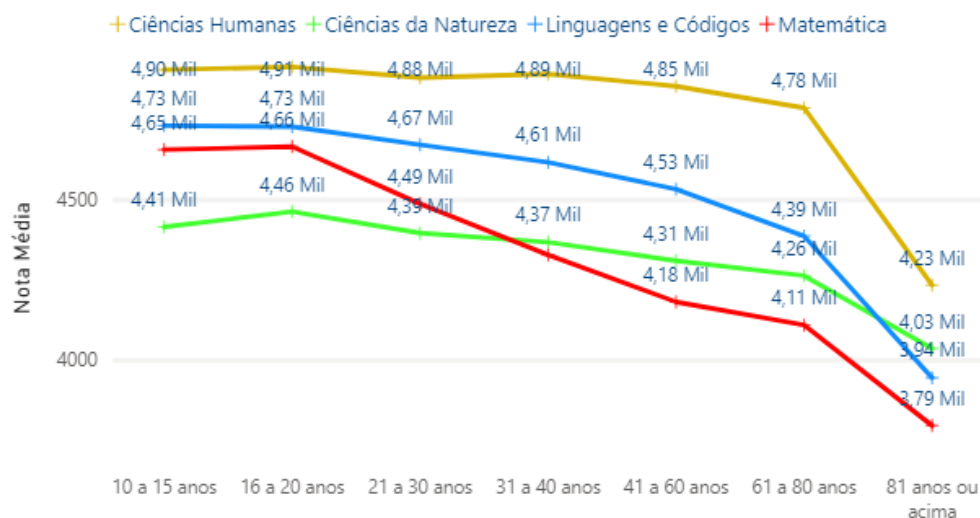
O gráfico abaixo apresenta em ordem decrescente a média das matérias por estado. É possível ver assim que apesar de termos diferenças de desempenho dependendo da matéria, alguns estados se destacam em relação a outros. São eles: SP, SC, MG, RJ, PR, ES, DF. Desses, 4 são da região sudeste, 2 do sul e 1 do centro-oeste. Mais uma vez, o desempenho acompanha a concentração de renda no Brasil.



- **Análise de desempenho por faixa etária**

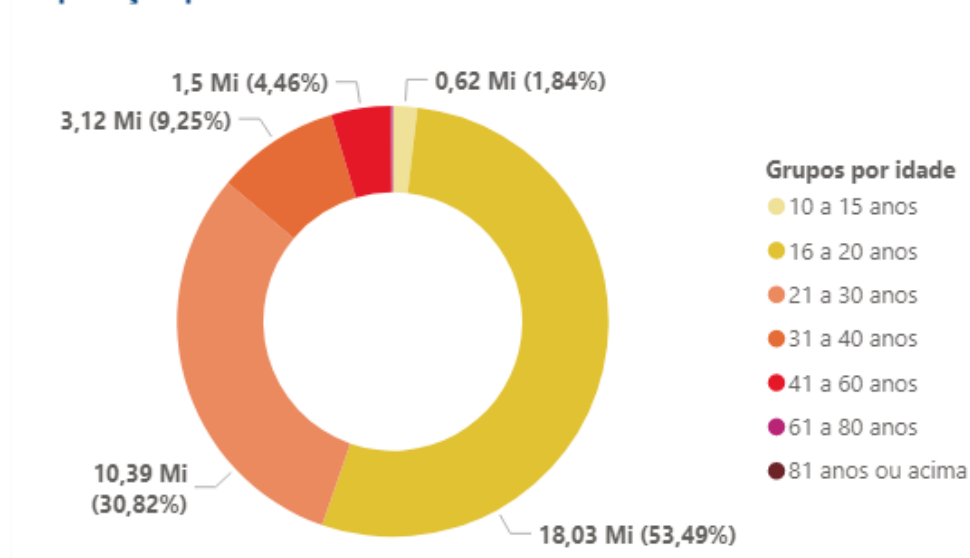
O gráfico abaixo representa a média de cada matéria separada por faixa etária dos alunos. Ao olharmos o gráfico abaixo, percebemos que há uma redução gradativa da média dos alunos quanto maior sua idade. Entretanto, é possível perceber que a matéria Ciências Humanas, independente da faixa etária, apresenta uma média praticamente estável até os 60 anos. Em contrapartida, Matemática apresenta a maior redução ao longo dos anos. Podemos supor que na 1ª matéria o conhecimento não se perde ao longo do tempo, enquanto matemática muito provavelmente requer maior prática para o conhecimento não se perder.

Média das Matérias por Faixa Etária



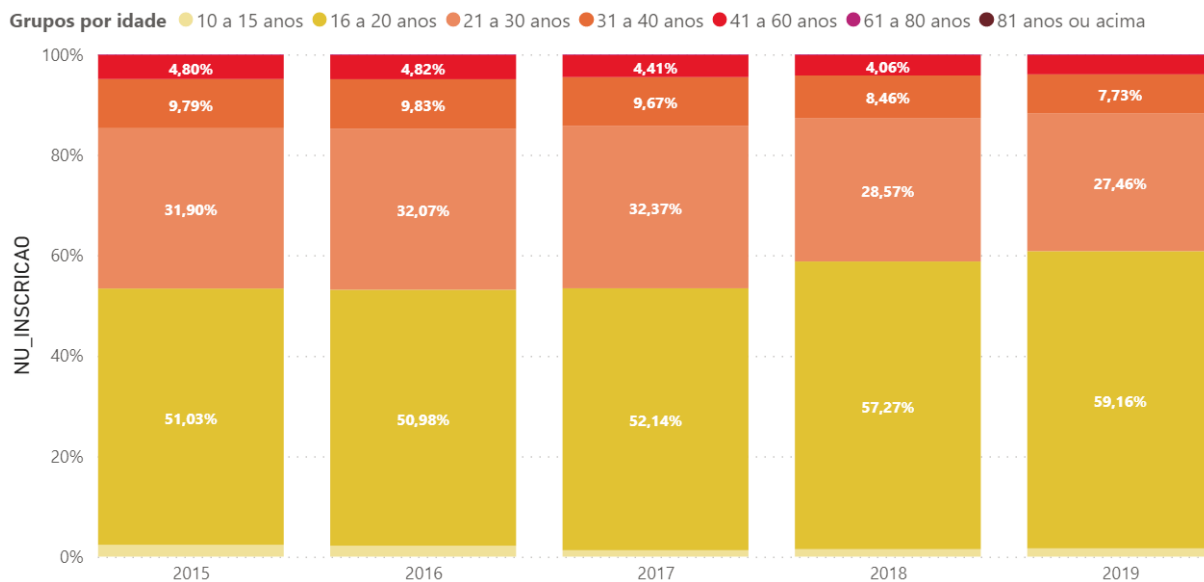
Ainda sobre a análise por faixa etária, o gráfico abaixo traz a distribuição da quantidade de alunos por faixa etária. Podemos inferir que o maior público do Enem são alunos de 16 a 20 anos, que representam 53,49% do total. Esse resultado era esperado, pois é a faixa etária que está concluindo o ensino médio e tem a expectativa de ingressar na faculdade através da prova do enem. No entanto, podemos refletir se esse número apresenta um bom resultado. Significa que mais de 45% dos alunos têm idade mais elevada e ainda busca ingressar na faculdade.

Separação por idade



Quando analisado ao longo dos anos, esse gráfico apresenta uma tendência de melhora dos números. As demais faixas etárias têm diminuído sua fatia do gráfico, enquanto a da faixa etária de 16 a 20 anos têm melhorado. Dessa forma,

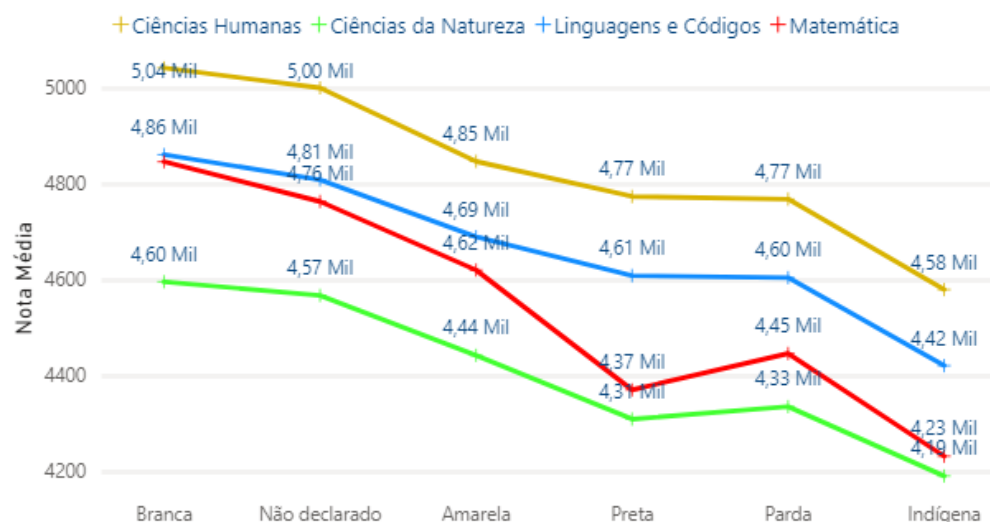
podemos concluir que está havendo uma melhora desses números, vendo que o essa faixa etária tem representatividade de mais de 59% no ano de 2019.



- **Análise de desempenho por cor/raça**

Esse gráfico representa a média das matérias por cor/raça. Levando em consideração que a classe branca tem maior concentração de dinheiro e maior estudo, o gráfico evidencia a diferença social presente no país. Quando analisado ano a ano, essa diferença se mantém estável, portanto é muito provável que essa diferença se propague por muitos anos após a distribuição dos cursos e futuramente distribuição salarial das classes.

Média das Matérias por Cor/Raça (auto-declarados)

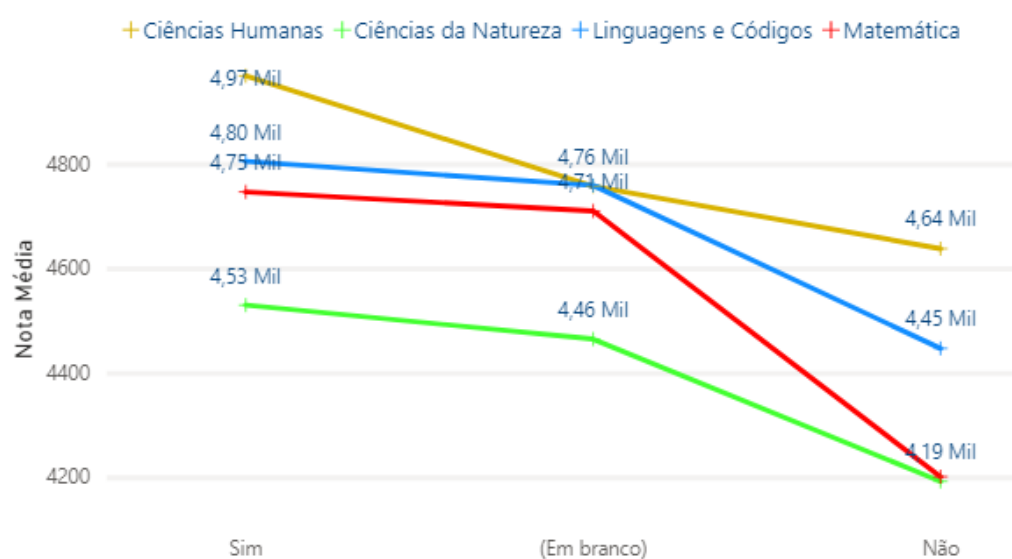


- Análise de desempenho por acesso à internet

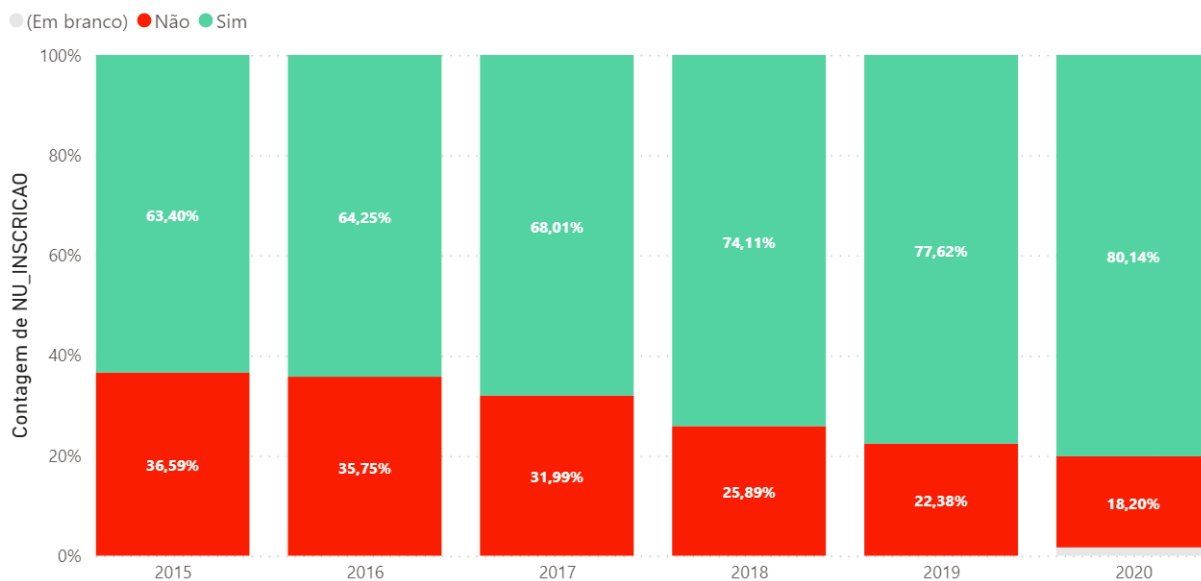
O Gráfico abaixo representa a média dos alunos com acesso à internet em comparação com alunos sem acesso à internet. Assim como o gráfico acima, mesmo quando analisamos o gráfico abaixo por ano ou classe social, a diferença das notas entre alunos com e sem internet se mantém proporcional.

Analisando os números abaixo, podemos calcular que a diferença das notas dos alunos com acesso à internet é 9,27% maior. Sendo assim, torna-se evidente a importância da internet no que tange a democratização do acesso à informação e ao estudo da população.

Média das Matérias por Acesso à Internet



Ainda dentro deste tópico, o gráfico abaixo representa a quantidade de alunos com acesso à internet ao longo dos anos. Vemos que a porcentagem de alunos com acesso à internet tem aumentado gradativamente, com os dados do último ano representando aproximadamente 80%. Temos visto um aumento de 3% na quantidade de alunos com internet a cada ano, e caso se mantenha o crescimento, em 6 anos praticamente 100% dos alunos terão acesso à internet.

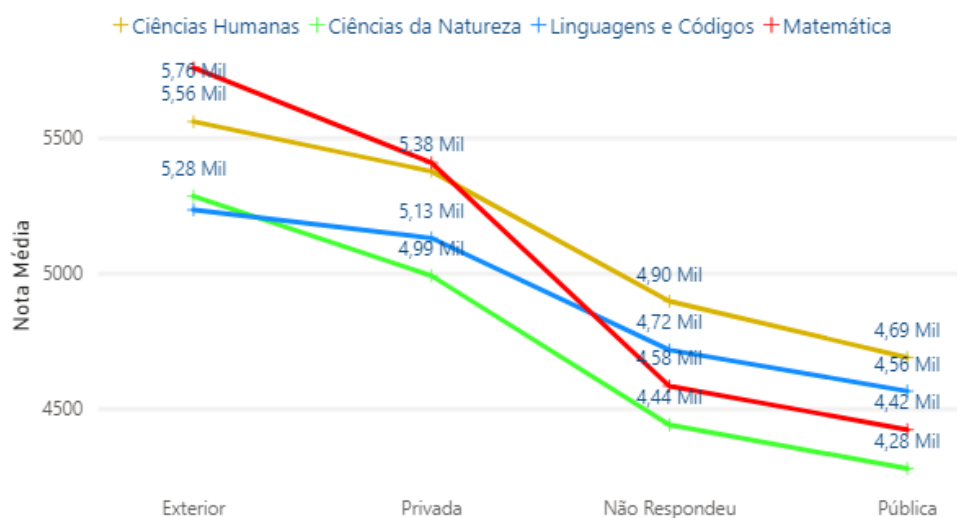


- **Análise de desempenho por tipo de escola**

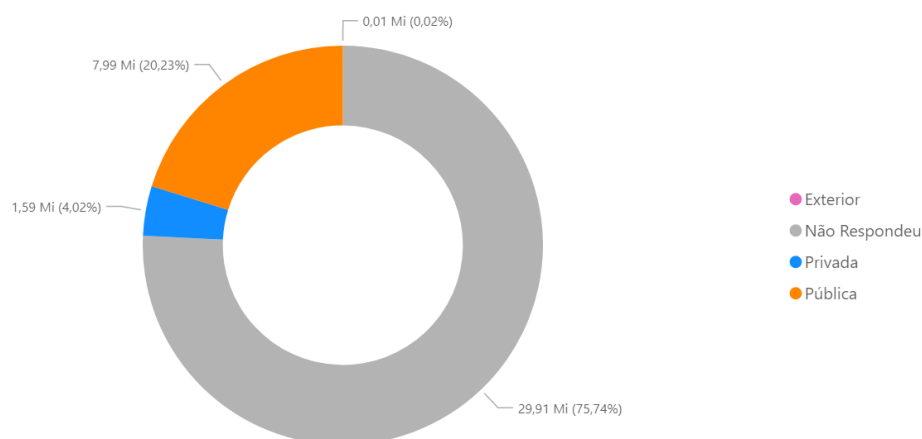
O Gráfico abaixo representa a média das matérias por tipo de escola em que o aluno estudou. Podemos levantar hipóteses para explicar os números do gráfico, no entanto é provável que a comparação entre a média das notas das escolas do exterior comparadas com escolas públicas e privadas não traduza uma verdade, levando em consideração que os melhores alunos do país ganham a oportunidade de estudar em escolas estrangeiras. Dessa forma, não necessariamente a escola do exterior seria melhor, e sim os alunos que foram estudar nessas escolas.

Porém, quando comparamos as notas de alunos de escolas privadas e de escolas públicas, fica novamente evidente a diferença. Alunos de escolas privadas têm notas 16,49% em média maiores.

Média das Matérias por Tipo de Escola



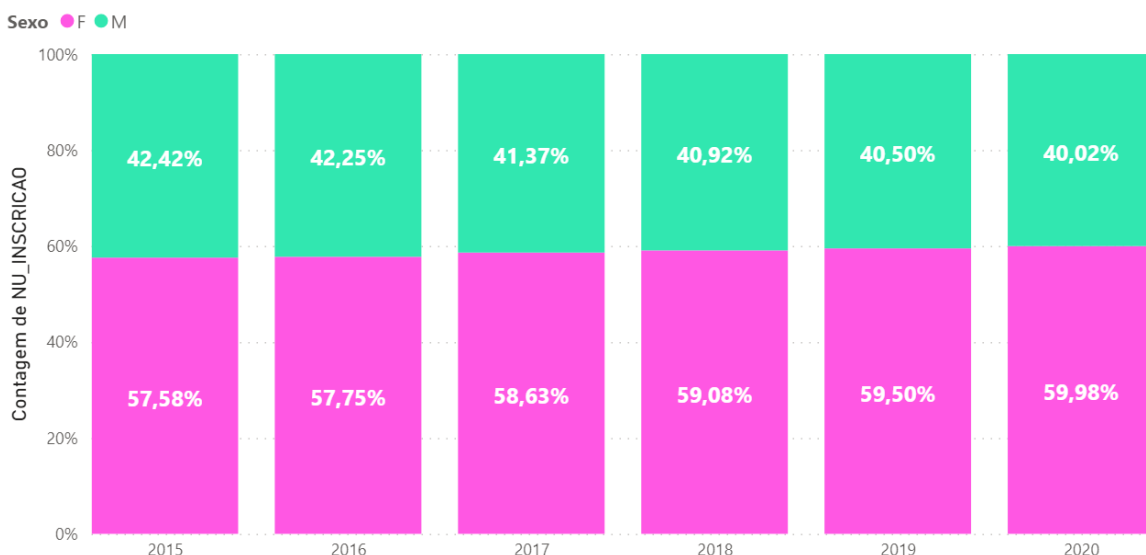
Ainda dentro deste tópico, ao olharmos a quantidade do gráfico abaixo, vemos que o número de alunos que efetivamente responderam à essa pergunta do formulário foi baixo (25%). Portanto, mesmo que a diferença de notas entre escolas públicas e privadas seja evidente, há necessidade de reforçar essa pesquisa, reduzindo o número de alunos sem resposta e trazendo um número mais confiável para realizarmos uma nova análise.



- Análise da quantidade de alunos por sexo

O gráfico abaixo representa a quantidade de homens (M) e mulheres (F) inscritos na prova. Quando analisamos a representatividade ao longo dos anos, esse número também apresenta uma melhora quando comparamos a presença no mercado de trabalho e na diferença salarial entre os sexos. Se pensarmos que as

mulheres têm procurado o acesso à faculdade, isso vai se refletir em empregos mais qualificados e redução da desigualdade referente a esse quesito. No último ano de dados (2020), a distribuição se deu com 60% de mulheres e 40% de homens inscritos para a prova.



5. Conclusões

A partir da análise dos gráficos apresentados no dashboard baseado nos microdados do ENEM, entre os anos de 2015 e 2020, é possível identificar através do desempenho dos alunos na prova a diferença social e racial refletida pelas suas médias. Dentre elas, podemos destacar as mais importantes abaixo:

- O início da pandemia impactou diretamente o número de faltantes do exame do ENEM de 2020 (55% dos alunos). Não é possível verificar através desses dados, porém é provável que essa alta taxa de faltantes reflita na redução do número de alunos nas turmas de graduação, e possivelmente mais tarde em falta de profissionais qualificados no mercado de trabalho.
- A diferença das médias dos alunos separados por estados mostra como alguns deles têm média consideravelmente maior que outros. A região

sudeste como um todo, além de alguns estados do sul e centro-oeste, são estados com média geral maior, além de maior concentração de renda.

- A quantidade de alunos inscritos na faixa etária de 16 a 20 anos vem crescendo ao longo dos anos, enquanto para as outras faixas etárias o número tem reduzido. Podemos considerar essa análise positiva, visto que essa é a faixa etária que está concluindo o ensino médio, e portanto o próximo passo natural é prestar o ENEM para acesso à faculdade. A redução das demais faixas etárias pode representar uma redução de uma demanda represada, que não tiveram possibilidade de acesso à faculdade anteriormente, além de outros casos.
- Os grupos de negros e índios apresentam média menor que os demais grupos. Esse desempenho provavelmente está relacionado também à distribuição de renda entre esses grupos. Os gráficos evidenciam essa diferença, que tem se mantido estável ao longo dos anos.
- Os alunos que têm acesso a internet têm rendimento melhor que aqueles que não têm acesso. Porém, em 2020 tínhamos 80% dos alunos inscritos com acesso à internet e, caso a tendência de aumento de 3% ao ano, em 2026 teremos praticamente 100% dos alunos com acesso à internet
- Quando comparamos os tipos de escola, as privadas apresentam médias mais elevadas que as públicas. No entanto, 75% dos alunos não respondem à essa pergunta, e temos um dado baseado em apenas 25% dos alunos que respondem
- A cada ano a participação das mulheres têm crescido, representando no ano de 2020 aproximadamente 60%. Esse aumento é importante, visto a diferença que existe no mercado de trabalho, quanto ao número de mulheres e quanto a diferença salarial.

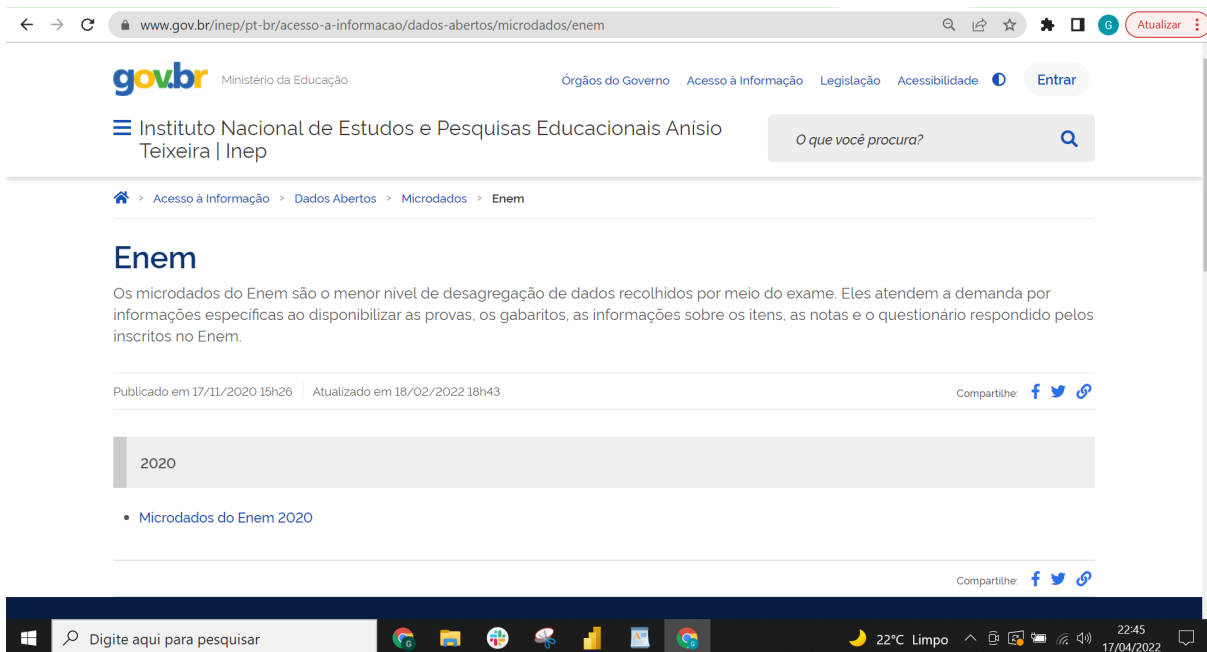
Como podemos inferir através das análises acima, grande parte das diferenças de desempenho são consequências da concentração de renda entre as raças e os estados. Portanto, a intervenção proposta seria no âmbito econômico, procurando formas ou reformas para trazer maior equilíbrio e reduzir essas diferenças sociais. Além disso, deveria ser feito um aumento do investimento em escolas públicas, para trazer uma educação de qualidade e assim proporcionar melhores notas e melhores oportunidades a esses alunos.

Outras intervenções poderiam ser sugeridas visando a melhora da pesquisa do ENEM para obtenção dos microdados. A primeira possibilidade é a retirada da opção “Não Respondeu” à pergunta quanto ao tipo de escola do aluno. Esse dado é extremamente importante, e por isso a pergunta não deveria deixar de ser respondida. Também poderia ser realizada uma revisão das perguntas feitas aos alunos, no intuito de melhorar a qualidade da pesquisa e possivelmente retirar perguntas pouco relevantes do questionário.

Algumas lições foram aprendidas com a execução deste trabalho. Seguem elas a seguir:

- Os microdados do ENEM apresentam grande número de linhas e colunas a serem processadas pelo programa na parte de ETL. Devido a isso, é preciso que as colunas que não serão interessantes para a construção dos dashboards sejam excluídas das consultas logo no início do processo.
- Para a realização do processo de ETL, o programa Power BI apresentou uma performance satisfatória, enquanto o programa Tableau Prep Builder não conseguiu processar os dados.
- Os microdados do ENEM demoram a ser disponibilizados pelo governo. Até Fevereiro/2022, os dados do ENEM 2020 ainda não tinham sido publicados, e os dados do ENEM 2021 continuam não publicados. Em 17/04/2020 foi verificado novamente no site oficial do ENEM, e desta vez os dados dos demais anos, exceto os microdados do ano de 2020, não estavam

disponíveis. Portanto, o governo não tem priorizado a estruturação e acessibilidade dos dados públicos referente ao ENEM.



6. Links

- Artefato gerado nesse trabalho Microsoft.MicrosoftPowerBIDesktop (.pbix): [TCC Microdados ENEM - Gabriel Canivello.pbix](#)
- Vídeo de apresentação do dashboard gerado no trabalho: [Microdados ENEM - TCC - Gabriel Canivello.webm](#)

REFERÊNCIAS

Exame Nacional do Ensino Médio (Enem). gov.br - Ministério da Educação - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), 2021. Disponível em: <https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/enem>. Acesso em: 28 de Dezembro de 2021.

Enem. gov.br - Ministério da Educação - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), 2021. Disponível em: <<https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/enem>>. Acesso em: 28 de Dezembro de 2021.

Enem 20 anos: a transformação da maior prova do Brasil. Brasil Escola, 2021. Disponível em: <<https://vestibular.brasilecola.uol.com.br/enem/enem-20-anos-transformacao-maior-prova-brasil.htm>>. Acesso em: 28 de Dezembro de 2021.

O Enem das universidades. Uol Notícias, 2021. Disponível em: <<https://noticias.uol.com.br/reportagens-especiais/enem-2021-a-porta-de-acesso-ao-ensino-superior/>>. Acesso em: 28 de Dezembro de 2021.