



PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS
NÚCLEO DE EDUCAÇÃO A DISTÂNCIA
Pós-graduação *Lato Sensu* em Gestão e Análise Estratégica de Dados

RELATÓRIO TÉCNICO

ANÁLISE DOS PARTICIPANTES DO ENEM DE 2019 E 2020

Andreza Mieko Marinho Suzuki

Belo Horizonte

2022

SUMÁRIO

- 1. Introdução3**
 - 1.1. Contexto3
 - 1.2. Objetivos.....4
 - 1.3. Público alvo4
- 2. Modelos de Dados4**
 - 2.1. Modelo Dimensional4
 - 2.2. Fatos e Dimensões5
- 3. Integração, Tratamento e Carga de Dados6**
 - 3.1. Fontes de Dados.....6
 - 3.2. Processos de Integração e Carga (ETL).....6
- 4. Camada de Apresentação8**
 - 4.1 Painel Estratégico8
 - 4.2 Painel Tático10
 - 4.3 Painel Operacional.....11
 - 4.4 Análises avançadas12
- 5. Registros de Homologação13**
- 6. Conclusões15**
- REFERÊNCIAS16**

1. Introdução

1.1. Contexto

O nível de escolaridade é um aspecto importante para o progresso econômico e diminuição da disparidade de renda no país de acordo com o IPEA (1995). Atualmente a qualidade do Ensino Médio no Brasil é medida através do Exame Nacional do Ensino Médio (ENEM), sendo também a principal forma de acesso à educação superior no país e sua nota pode ser utilizada nas instituições públicas e privadas, através do Sistema de Seleção Unificada (Sisu) ou do Programa Universidade para Todos (ProUni). E conforme o site do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP)¹, a nota do exame também é aceita em mais de 50 instituições em Portugal e podem pleitear o financiamento estudantil (FIES).

Ainda em conformidade com INEP, o exame foi instituído em 1998, com o intuito de avaliação do desempenho escolar ao término da educação básica e somente em 2009 o exame foi aperfeiçoado para ser utilizado como forma de acesso à educação superior. E em 2020 o participante começou a poder escolher a forma de realização da prova, impressa ou digital, nos locais escolhidos pelo Instituto.

O ENEM não tem limite de idade e os candidatos que ainda não concluíram o Ensino Médio, podem fazer a prova somente para treinar.

Neste trabalho, serão analisados os micros dados do ENEM de 2019 e 2020 apresentando a análise quantitativa dos candidatos, das instituições de ensino de formação dos participantes e das notas obtidas na prova e redação. A base de dados fornecida pelo INEP é fonte de informações importantes para entender o perfil dos candidatos, podendo gerar indicadores que apresentam propostas de políticas educativas promissoras e auxiliar nas tomadas de decisões, planejamentos estratégicos e tomadas de decisões das instituições de ensino.

¹ <https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/enem>

1.2. Objetivos

Este relatório tem como objetivo extrair, transformar e analisar os micros dados das provas do ENEM realizadas no ano de 2019 e 2020, gerando *dashboards*, padrões e tendências dos participantes do exame, com o intuito de apoiar os gestores educacionais na aplicação de medidas para aumentar a média das notas de seus alunos, através de três *dashboards*: estratégico, analítico/tático e operacional.

1.3. Público alvo

O público-alvo abrange os gestores, pesquisadores, instituições e interessados na área da educação. Para melhor análise dos dados, recomenda-se que o usuário tenha conhecimentos mínimos de sistema operacional para conseguir navegar nos painéis. No entanto, para melhor obtenção de *insights* e tomada de decisão, o conhecimento básico de análise de dados é significativa.

2. Modelos de Dados

2.1. Modelo Dimensional

O modelo utilizado neste trabalho foi o *star schema*, que é caracterizado pelo uso de uma tabela fato ligada a várias tabelas dimensão.

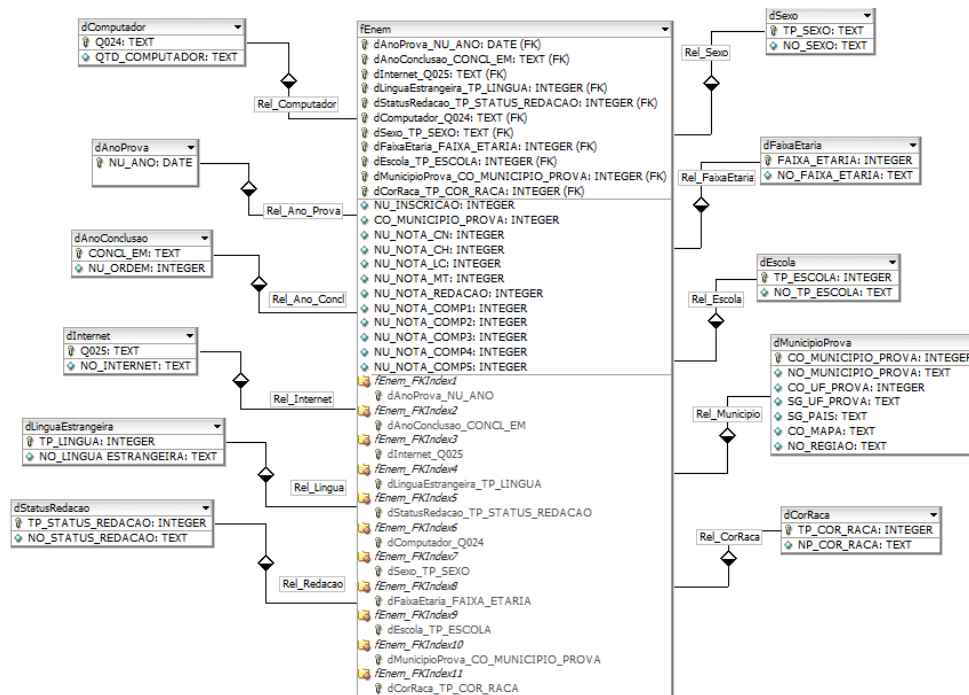


Figura 1. Diagrama dimensional, elaborado com a ferramenta DBDesigner.

2.2. Fatos e Dimensões

As seguintes tabelas foram utilizadas para realização deste trabalho:

Tabela	Descrição	Tipo
fEnem	Contém os dados categóricos dos participantes (número de inscrição, ano do Enem, faixa etária, sexo, cor/raça) e das questões se o candidato possui computador e acesso à internet, informações acadêmicas (ano de conclusão e tipo de escola), informações do local de realização da prova (Município e UF) e as notas das provas de Ciências da Natureza, Ciências Humanas, Linguagens e Códigos (com a linguagem estrangeira escolhida), Matemática e Redação.	Fato
dAnoProva	Ano de realização da prova.	Dimensão
dFaixaEtaria	Contém o código da faixa etária com sua classificação.	Dimensão
dSexo	Possui o código do sexo com sua descrição.	Dimensão
dCorRaca	Contém o código da Cor/Raça com sua descrição.	Dimensão
dAnoConclusao	Ano da conclusão do Ensino Médio do participante.	Dimensão
dEscola	Contém o código do tipo de instituição de ensino que o participante concluiu o Ensino Médio com sua descrição.	Dimensão
dMunicipio	Possui os códigos dos Municípios com seus respectivos nomes e Unidade da Federação (UF) pertencentes.	Dimensão
dLinguaEstrangeira	Contém as opções de língua estrangeira.	Dimensão
dStatusRedacao	Status de classificação da redação.	Dimensão
dComputador	Contém informações de quantidade de computadores que o candidato possui em sua residência.	Dimensão
dInternet	Contém informações de acesso à internet na residência do candidato.	Dimensão

Tabela 1. Lista de tabelas fato e dimensão.

3. Integração, Tratamento e Carga de Dados

3.1. Fontes de Dados

Os dados referentes ao ano de 2019 foram extraídos da plataforma Kaggle, que hospeda *datasets* públicos. E os dados de 2020 foram obtidos do site do INEP, autarquia vinculada ao Ministério da Educação, responsável pelas pesquisas educacionais.

3.2. Processos de Integração e Carga (ETL)

Os *datasets* foram obtidos das plataformas supracitadas em pastas compactadas, os arquivos foram descompactados através do programa 7-Zip e os dados de 2019 e 2020 foram exportados separadamente, em formato .csv e, em seguida, os dados foram transformados no Power Query: com os ajustes de colunas, alteração de tipo de dado e renomeação das colunas para padronização das tabelas e combinação dos dados, posteriormente constituindo uma única tabela fato com os dados de 2019 e 2020 unificados.

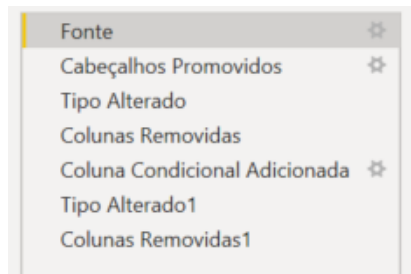


Figura 2. Etapas aplicadas no Power Query.

As tabelas dimensão foram criadas utilizando a linguagem DAX através da função *SUMMARIZE*. Elaboradas de acordo com os dicionários disponibilizados pelas fontes (Kaggle e INEP), realizadas de acordo com a Tabela 2.

Tabela	Fórmula Criação	Coluna Calculada
dAnoConclusao	<code>SUMMARIZE(fEnem,fEnem[CONCL_EM])</code>	<pre> NU_ORDEM = IF(dAnoConclusao[CONCL_EM] = "Não informado", 0, IF(dAnoConclusao[CONCL_EM] = "Antes de 2007", 1, IF(dAnoConclusao[CONCL_EM] = "2007", 2, IF(dAnoConclusao[CONCL_EM] = "2008", 3, IF(dAnoConclusao[CONCL_EM] = "2009", 4, IF(dAnoConclusao[CONCL_EM] = "2010", 5, IF(dAnoConclusao[CONCL_EM] = "2011", 6, IF(dAnoConclusao[CONCL_EM] = "2012", 7, IF(dAnoConclusao[CONCL_EM] = "2013", 8, IF(dAnoConclusao[CONCL_EM] = "2014", 9, IF(dAnoConclusao[CONCL_EM] = "2015", 10, IF(dAnoConclusao[CONCL_EM] = "2016", 11, IF(dAnoConclusao[CONCL_EM] = "2017", 12, IF(dAnoConclusao[CONCL_EM] = "2018", 13, IF(dAnoConclusao[CONCL_EM] = "2019", 14)))))))))) </pre>
dAnoProva	<code>SUMMARIZE(fEnem,fEnem[NU_ANO])</code>	

dComputador	SUMMARIZE(fEnem, fEnem[Q024])	QTD_COMPUTADOR = IF(dComputador[Q024] = "A", "Não", IF(dComputador[Q024] = "B", "Sim, um", IF(dComputador[Q024] = "C", "Sim, dois", IF(dComputador[Q024] = "D", "Sim, três", IF(dComputador[Q024] = "E", "Sim, quatro ou mais", "Sem resposta"))))
dCorRaca	SUMMARIZE(fEnem, fEnem[TP_COR_RACA])	NO_COR_RACA = IF(dCorRaca[TP_COR_RACA] = 0, "Não declarado", IF(dCorRaca[TP_COR_RACA] = 1, "Branca", IF(dCorRaca[TP_COR_RACA] = 2, "Preta", IF(dCorRaca[TP_COR_RACA] = 3, "Parda", IF(dCorRaca[TP_COR_RACA] = 4, "Amarela", IF(dCorRaca[TP_COR_RACA] = 5, "Indígena"))))
dEscola	SUMMARIZE(fEnem, fEnem[TP_ESCOLA])	NO_TP_ESCOLA = IF(dEscola[TP_ESCOLA] = 1, "Não Respondeu", IF(dEscola[TP_ESCOLA] = 2, "Pública", IF(dEscola[TP_ESCOLA] = 3, "Privada"))
dFaixaEtaria	SUMMARIZE(fEnem, fEnem[FAIXA_ETARIA])	NO_FAIXA_ETARIA = IF(dFaixaEtaria[FAIXA_ETARIA] = 1, "Menor de 17", IF(dFaixaEtaria[FAIXA_ETARIA] = 2, "17 anos", IF(dFaixaEtaria[FAIXA_ETARIA] = 3, "18 anos", IF(dFaixaEtaria[FAIXA_ETARIA] = 4, "19 anos", IF(dFaixaEtaria[FAIXA_ETARIA] = 5, "20 anos", IF(dFaixaEtaria[FAIXA_ETARIA] = 6, "21 anos", IF(dFaixaEtaria[FAIXA_ETARIA] = 7, "22 anos", IF(dFaixaEtaria[FAIXA_ETARIA] = 8, "23 anos", IF(dFaixaEtaria[FAIXA_ETARIA] = 9, "24 anos", IF(dFaixaEtaria[FAIXA_ETARIA] = 10, "25 anos", IF(dFaixaEtaria[FAIXA_ETARIA] = 11, "Entre 26 e 30", IF(dFaixaEtaria[FAIXA_ETARIA] = 12, "Entre 31 e 35", IF(dFaixaEtaria[FAIXA_ETARIA] = 13, "Entre 36 e 40", IF(dFaixaEtaria[FAIXA_ETARIA] = 14, "Entre 41 e 45", IF(dFaixaEtaria[FAIXA_ETARIA] = 15, "Entre 46 e 50", IF(dFaixaEtaria[FAIXA_ETARIA] = 16, "Entre 51 e 55", IF(dFaixaEtaria[FAIXA_ETARIA] = 17, "Entre 56 e 60", IF(dFaixaEtaria[FAIXA_ETARIA] = 18, "Entre 61 e 65", IF(dFaixaEtaria[FAIXA_ETARIA] = 19, "Entre 66 e 70", IF(dFaixaEtaria[FAIXA_ETARIA] = 20, "Maior de 70"))))))))))))))))
dInternet	SUMMARIZE(fEnem, fEnem[Q025])	NO_INTERNET = IF(dInternet[Q025] = "A", "Não", IF(dInternet[Q025] = "B", "Sim", "Sem resposta"))
dLinguaEstrangeira	SUMMARIZE(fEnem, fEnem[TP_LINGUA])	NO_LINGUA_ESTRANGEIRA = IF(dLinguaEstrangeira[TP_LINGUA] = 0, "Inglês", IF(dLinguaEstrangeira[TP_LINGUA] = 1, "Espanhol"))
dMunicípio	SUMMARIZE(fEnem, fEnem[CO_MUNICIPIO_PROVA], fEnem[NO_MUNICIPIO_PROVA], fEnem[CO_UF_PROVA], fEnem[SG_UF_PROVA])	SG_PAIS = "BR"
		CO_MAPA = dMunicipioProva[SG_PAIS]&"-" &dMunicipioProva[SG_UF_PROVA] NO_REGIAO = IF(dMunicipioProva[SG_UF_PROVA] = "DF", "Centro-Oeste", IF(dMunicipioProva[SG_UF_PROVA] = "GO", "Centro-Oeste", IF(dMunicipioProva[SG_UF_PROVA] = "MS", "Centro-Oeste", IF(dMunicipioProva[SG_UF_PROVA] = "MT", "Centro-Oeste", IF(dMunicipioProva[SG_UF_PROVA] = "AL", "Nordeste", IF(dMunicipioProva[SG_UF_PROVA] = "BA", "Nordeste", IF(dMunicipioProva[SG_UF_PROVA] = "CE", "Nordeste", IF(dMunicipioProva[SG_UF_PROVA] = "MA", "Nordeste", IF(dMunicipioProva[SG_UF_PROVA] = "PB", "Nordeste", IF(dMunicipioProva[SG_UF_PROVA] = "PE", "Nordeste", IF(dMunicipioProva[SG_UF_PROVA] = "PI", "Nordeste", IF(dMunicipioProva[SG_UF_PROVA] = "RN", "Nordeste", IF(dMunicipioProva[SG_UF_PROVA] = "SE", "Nordeste", IF(dMunicipioProva[SG_UF_PROVA] = "AC", "Norte", IF(dMunicipioProva[SG_UF_PROVA] = "AM", "Norte", IF(dMunicipioProva[SG_UF_PROVA] = "AP", "Norte", IF(dMunicipioProva[SG_UF_PROVA] = "PA", "Norte", IF(dMunicipioProva[SG_UF_PROVA] = "RO", "Norte", IF(dMunicipioProva[SG_UF_PROVA] = "RR", "Norte", IF(dMunicipioProva[SG_UF_PROVA] = "TO", "Norte", IF(dMunicipioProva[SG_UF_PROVA] = "ES", "Sudeste", IF(dMunicipioProva[SG_UF_PROVA] = "MG", "Sudeste", IF(dMunicipioProva[SG_UF_PROVA] = "RJ", "Sudeste", IF(dMunicipioProva[SG_UF_PROVA] = "SP", "Sudeste", IF(dMunicipioProva[SG_UF_PROVA] = "PR", "Sul", IF(dMunicipioProva[SG_UF_PROVA] = "SC", "Sul", IF(dMunicipioProva[SG_UF_PROVA] = "RS", "Sul",))))))))))))))))))

Tabela 2. Fórmulas DAX.

E as medidas foram criadas com linguagem DAX e armazenadas na tabela “_Medidas”.

4. Camada de Apresentação

A página inicial apresenta os links para cada um dos painéis: Quantidade de Candidatos (estratégico), Visão Institucional (tático), Provas e Redação (operacional) e Previsões (análise avançada). Basta clicar sobre um desses ícones que o usuário é direcionado para o *dashboard* correspondente. Cada painel conta com um botão “início”, no canto superior direito, que redireciona o usuário à página inicial.

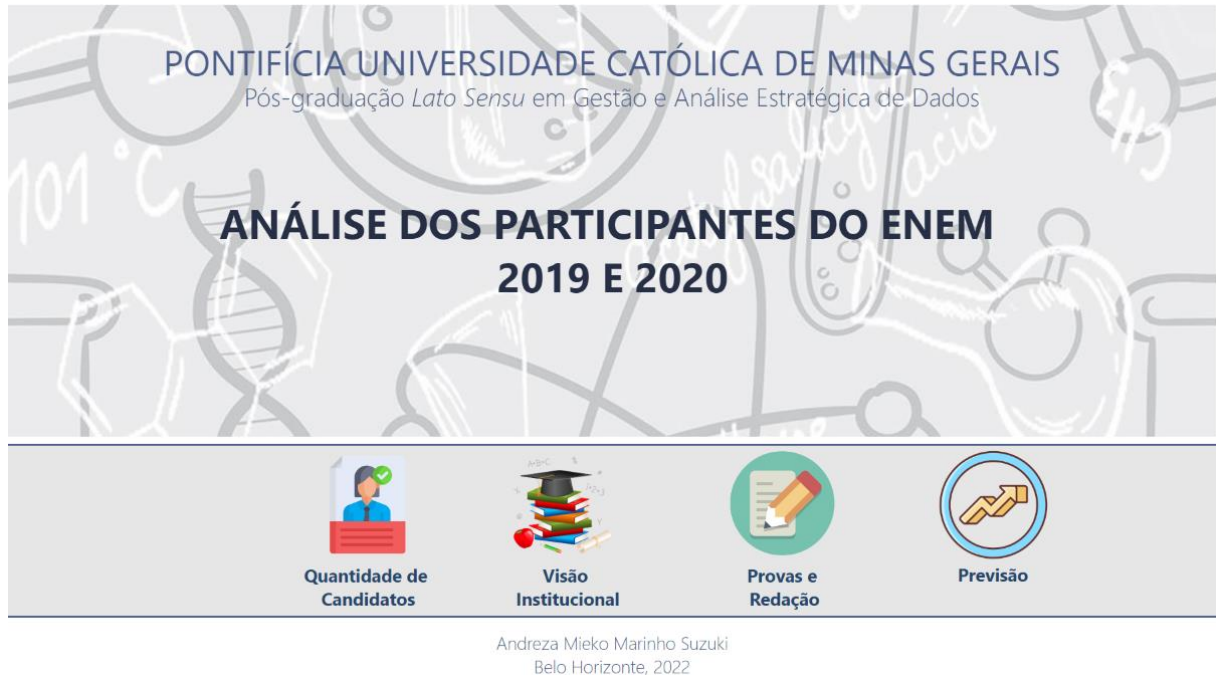


Figura 3. Página Inicial.

Todos os visuais possuem dica de ferramenta com o cálculo das medidas utilizadas.

4.1 Painel Estratégico

O painel estratégico, denominado neste *dashboard* como “Quantidade de Candidatos”, possui as informações quantitativas dos candidatos.

A quantidade de candidatos foi obtida através da contagem distinta dos números de inscrição.

Um gráfico territorial feito através do visual adicional *Synoptic Painel by OKViz* apresenta a quantidade de candidatos (em milhares) por UF, com diferenciação da tonalidade da cor, de acordo com os valores de quantidade, evidenciando que, os estados de São Paulo e Minas Gerais, nesta ordem, possuem os maiores números de candidatos do país.

Temos dois gráficos de linhas, um com o a quantidade de candidatos (em milhares) por ano de conclusão do ensino médio, constando que a maior parte dos examinados formaram no ano de 2018. E o outro com a divisão por faixa etária, tendo como maioria o intervalo de idades entre 17 e 19 anos.

Uma tabela de candidatos por cor/raça aponta que a maior parte são pardos, seguidos pelos brancos e pretos.

Dois gráficos de barras apresentam o acesso à internet e computador na residência, sendo que, em sua maioria os candidatos não possuem computador na residência, mas possuem acesso à internet.

Neste painel é possível filtrar as visões pelos seguintes filtros:

- ano do ENEM;
- localidade por divisão regional, unidade federativa e município;
- faixa etária de menor de 17 anos até maior de 70 anos;
- sexo, feminino e masculino;
- cor/raça.

E no canto superior direito o botão para retornar à página inicial.

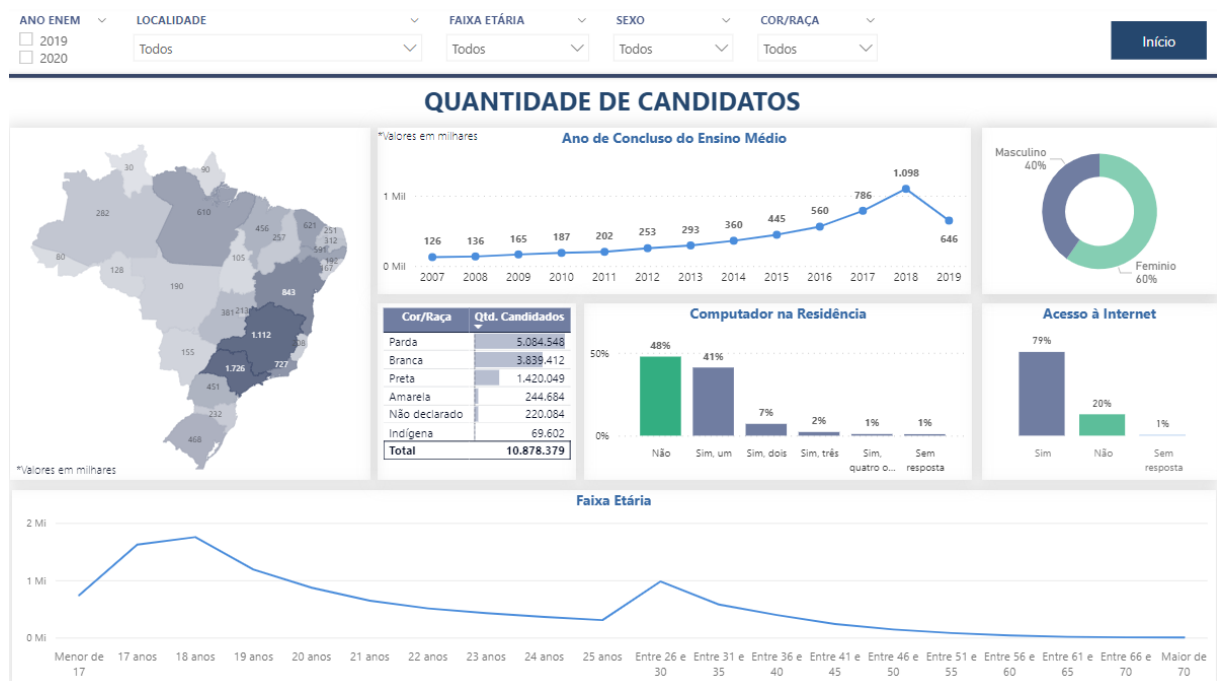


Figura 4. Dashboard Estratégico.

4.2 Painel Tático

O painel tático é voltado para melhor análise das instituições de ensino com a divisão das categorias de pública e privadas. Na parte superior do painel temos três gráficos de Gauge (velocímetro) apresentando a porcentagem de alunos provenientes das instituições, privadas, públicas e aqueles que não responderam. Apesar da maioria dos candidatos não terem respondido à questão de qual instituição de ensino concluiu sua formação, aproximadamente 3 milhões de candidatos responderam à pesquisa, representando ~27% do total de candidatos.

Abaixo, uma matriz retorna os valores de quantidade de candidatos e média da nota por região geográfica do país, podendo fazer o *drill down* por UF. E um gráfico de barras compara a média da nota da prova e da redação por instituição.

- Média nota da prova: cálculo de média das notas das provas de ciências humanas, ciências da natureza, linguagem e códigos e matemática;
- Média nota redação: cálculo de média da nota de redação;
- Média nota ENEM: cálculo de média da nota de prova e nota da redação.

Na parte inferior, um gráfico de área apresenta a média de nota por faixa etária e instituição de ensino. Apesar do número de candidatos da escola privada ser menor, eles representam a maior média de nota do ENEM, tanto na prova quanto na redação.

Na parte superior do painel temos os filtros de:

- ano do ENEM;
- localidade por divisão regional, unidade federativa e município;
- faixa etária de menor de 17 anos até maior de 70 anos;
- sexo, feminino e masculino;
- cor/raça;
- instituição de ensino.

E no canto superior direito o botão para retornar à página inicial.



Figura 5. Dashboard Tático.

4.3 Painel Operacional

No painel operacional são apresentados os dados por matéria e redação. A ideia é que os usuários da ferramenta possam analisar com mais detalhes a partir das matérias e escolha de língua estrangeira do candidato.

No quadrante esquerdo temos um gráfico territorial feito através do visual adicional *Synoptic Painel by OKViz* apresenta as notas médias de acordo com o parâmetro selecionado, com diferenciação da tonalidade da cor, de acordo com as notas. E um card que apresenta a média total, sem a divisão estadual.

No quadrante direito temos um funil com a média de nota por matéria, um gráfico de pizza com a porcentagem de candidatos por língua estrangeira (inglês e espanhol), um gráfico de barras empilhadas com a nota por língua estrangeira e, por fim, um gráfico de linhas com a quantidade de candidatos e os status de suas redações.

Na parte superior deste visual temos os filtros de:

- ano do ENEM;
- localidade por divisão regional, unidade federativa e município;
- faixa etária de menor de 17 anos até maior de 70 anos;
- sexo, feminino e masculino;
- cor/raça;

- instituição de ensino.

E no canto superior direito o botão para retornar à página inicial.

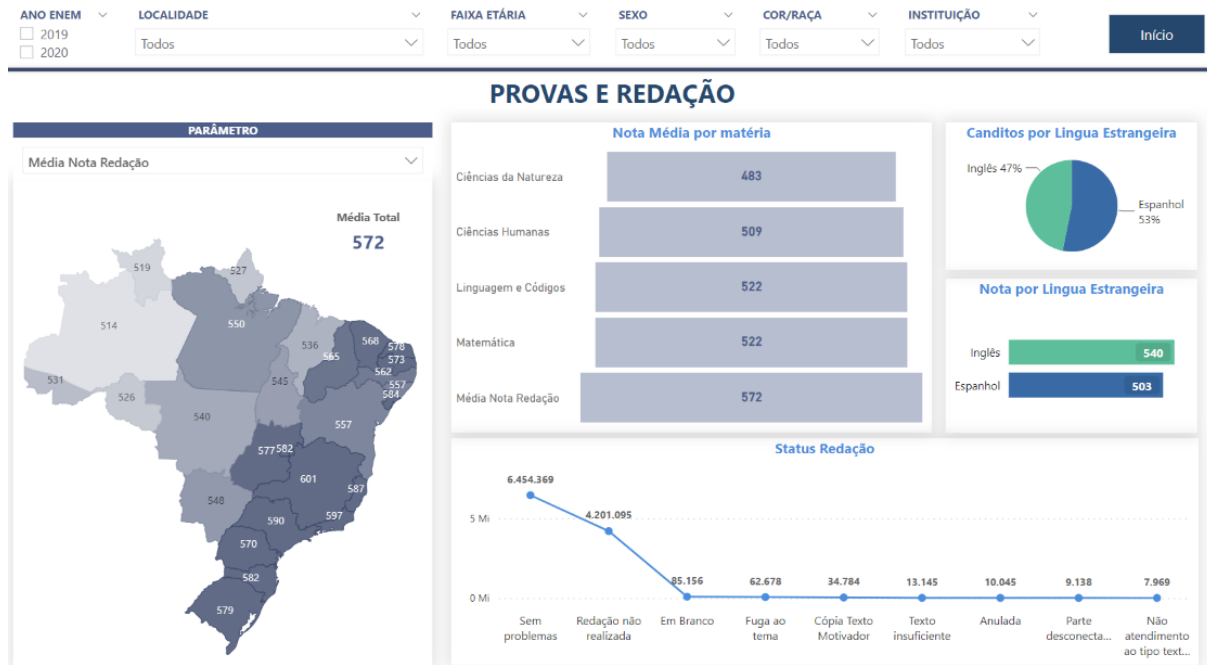


Figura 6. Dashboard Operacional.

4.4 Análises avançadas

Com a finalidade de complementar o *dashboard* apresentado em Power BI, foram realizadas análises utilizando a linguagem *Python*, no *Google Colaboratory*.

Foi realizado uma correlação das notas de cada matéria através da função “corr” e os dados foram apresentados com um *heat map*, podendo observar que as matérias com menor correlação entre si são: ciências da natureza e linguagem e códigos, enquanto linguagem e código e matemática tem maior correlação entre si.

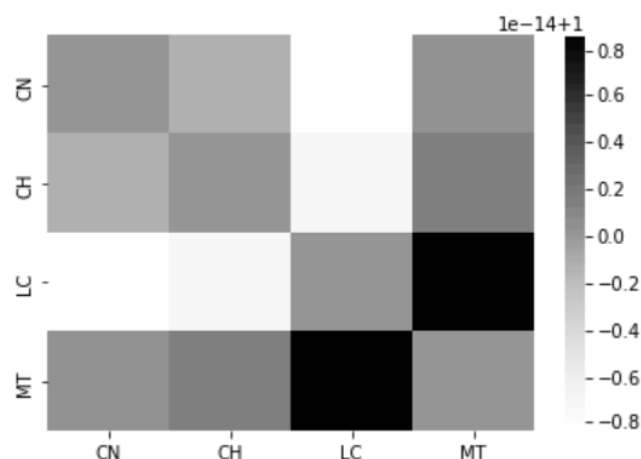


Figura 7. Heat Map: Correlação matérias.

As tendências foram realizadas no Power BI, através da análise das notas de 2018 até 2021, utilizando o recurso de previsão do próprio BI. A previsão para as notas do ENEM dos próximos dois anos é de queda em relação a 2021, a nota da prova se mantém e em contrapartida, a nota da prova de redação tem previsão de crescimento. As matérias de “Ciências da Natureza” e “Ciências Humanas” tem estimativa de manter conforme 2021, enquanto “Linguagens e Códigos” e “Matemática” apresentam expectativa de queda *versus* 2021.

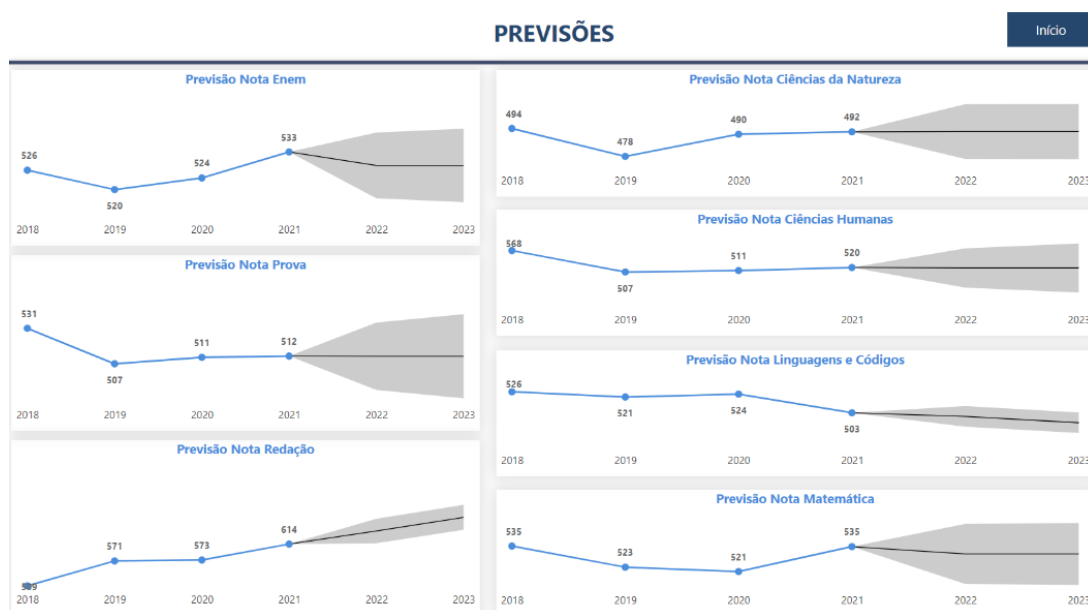


Figura 8. Previsões.

5. Registros de Homologação

Os testes de homologação foram executados através de consultas às tabelas no SQL Server e os resultados confrontados com a visualização em Power BI.

Os dados de 2019 e 2020 foram importados para o SQL separadamente e unidos através dos comandos CREATE TABLE e INSERT INTO, conforme figura 7.

```

SQLQuery7.sql - LA...ndreza Suzuki (53)*
-- Create Table Temp
CREATE TABLE #BD_Enem
(
    TP_COR_RACA FLOAT,
    NU_INSCRICAO NVARCHAR(255),
    NU_NOTA_CN FLOAT, NU_NOTA_CH FLOAT, NU_NOTA_LC FLOAT, NU_NOTA_MT FLOAT, NU_NOTA_REDACAO FLOAT
)

-- Inserção ENEM 2019
INSERT INTO #BD_Enem
SELECT [ENEM_2019].[DBO].[fEnem2019].[TP_COR_RACA],
       [ENEM_2019].[DBO].[fEnem2019].[NU_INSCRICAO],
       [ENEM_2019].[DBO].[fEnem2019].[NU_NOTA_CH],
       [ENEM_2019].[DBO].[fEnem2019].[NU_NOTA_CN],
       [ENEM_2019].[DBO].[fEnem2019].[NU_NOTA_LC],
       [ENEM_2019].[DBO].[fEnem2019].[NU_NOTA_MT],
       [ENEM_2019].[DBO].[fEnem2019].[NU_NOTA_REDACAO]
FROM [ENEM_2019].[dbo].[fEnem2019]

-- Inserção ENEM 2020
INSERT INTO #BD_Enem
SELECT [ENEM_2020].[DBO].[fEnem2020].[TP_COR_RACA],
       [ENEM_2020].[DBO].[fEnem2020].[NU_INSCRICAO],
       [ENEM_2020].[DBO].[fEnem2020].[NU_NOTA_CH],
       [ENEM_2020].[DBO].[fEnem2020].[NU_NOTA_CN],
       [ENEM_2020].[DBO].[fEnem2020].[NU_NOTA_LC],
       [ENEM_2020].[DBO].[fEnem2020].[NU_NOTA_MT],
       [ENEM_2020].[DBO].[fEnem2020].[NU_NOTA_REDACAO]
FROM [ENEM_2020].[dbo].[fEnem2020]

```

Figura 9. Criação das tabelas no SQL Server.

O primeiro teste foi na relação de quantidade de candidatos por Cor/Raça, apresentando o mesmo valor nas duas consultas, conforme figura 8 (valores sem separador de milhar) e figura 9 (valores com separador de milhar).

```

-- Consulta
SELECT DB.DESC_COR_RACA AS "Cor/Raça",
       COUNT(E.[NU_INSCRICAO]) AS "Quantidade de Candidatos"
FROM #BD_Enem E
INNER JOIN [DB_GERAL].[DBO].[Cor_Raca] DB ON DB.[TP_COR_RACA] = E.[TP_COR_RACA]
GROUP BY DB.DESC_COR_RACA ORDER BY [Quantidade de Candidatos] DESC

```

Cor/Raça	Quantidade de Candidatos
1 Parda	5084548
2 Branca	3839412
3 Preta	1420049
4 Amarela	244684
5 Não declarado	220084
6 Indígena	69602

Figura 10. Query em SQL Server.

Cor/Raça	Qtd. Candidatos
Parda	5.084.548
Branca	3.839.412
Preta	1.420.049
Amarela	244.684
Não declarado	220.084
Indígena	69.602
Total	10.878.379

Figura 11. Tabela de quantidade de candidatos por Cor/Raça apresentada no painel estratégico em Power BI.

Também foi realizada uma consulta apresentando a nota média na prova de redação.

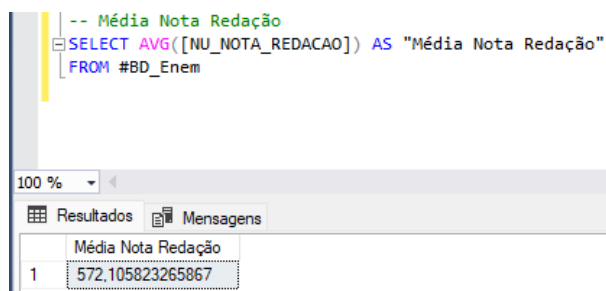


Figura 12. Query em SQL Server.



Figura 13. Média de Nota da Redação apresentada no painel operacional em Power BI.

6. Conclusões

A maior parte dos participantes do exame nos anos de 2019 e 2020 são da região Sudeste, em sua maioria do sexo feminino, com idade entre 17 e 19 anos, ou seja, concluintes do ensino médio. A região também apresenta a maior média em todas as matérias e na redação. Enquanto a região Norte dispõe das menores médias.

Da relação de participantes que responderam ao questionário relativo à instituição de ensino, 85% concluíram o ensino médio na escola pública e mesmo sendo maioria apresenta média da nota do Enem inferior aos candidatos da escola privada. A diferença entre a média da nota da prova é de -83 pontos e a diferença é ainda maior na nota de redação, de -173 pontos. Evidenciando a necessidade de investimento no ensino do setor público.

As matérias “Ciências da Natureza” e “Linguagem e Códigos” possuem menor correlação entre si, o que indica que se o participante está com baixo desempenho em uma, ele tem mais chances de ter menor desempenho na outra. E as matérias de “Linguagem e Códigos” e “Matemática” possuem previsão de queda nas notas. Os cursos preparatórios e as escolas devem avaliar tais correlações e previsões para melhor preparação dos alunos para o exame.

Até a data de início da realização deste trabalho o site do INEP só fornecia os dados do ENEM de 2019 e 2020, não sendo possível uma análise de um histórico maior, com análises mais precisas. Após agosto de 2022 os dados de mais edições passadas foram disponibilizadas e podem ser utilizadas para novos projetos.

REFERÊNCIAS

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. **Microdados do Enem 2020**. Brasília: Inep, 2020. Disponível em: < <https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/enem>>. Acesso em: 02 mar. 2022.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. **Apresentação**. Brasília: Inep. Disponível em: < <https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/enem> >. Acesso em: 02 mar. 2022.

INSTITUTO DE PESQUISA ECONÔMICA APLICADA. **Os Determinantes da Desigualdade no Brail**. Rio de Janeiro: IPEA, 1995.