

Ciência de Dados

Dados Educação



Angela Favretto Pastorello (Design), Bruna Oenning Amador, Daniel de Viveiros Inácio e Matheus Biscaya Gutierrez (Sistemas de Informação)



Sumário

- Tema;
- Objetivo;
- Dados;
- Modelos;
- Resultados;
- Conclusão.
 - Limitações;
 - Trabalhos futuros.

1

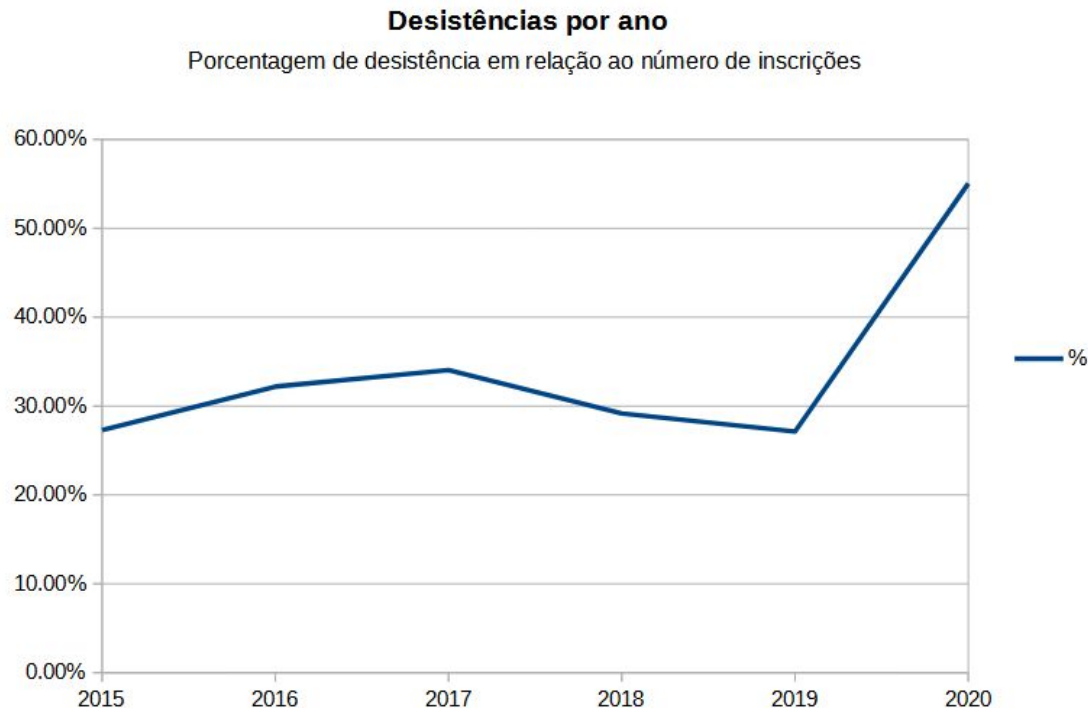
Sobre o tema

Analisar o impacto da pandemia na educação através da análise das desistências do ENEM

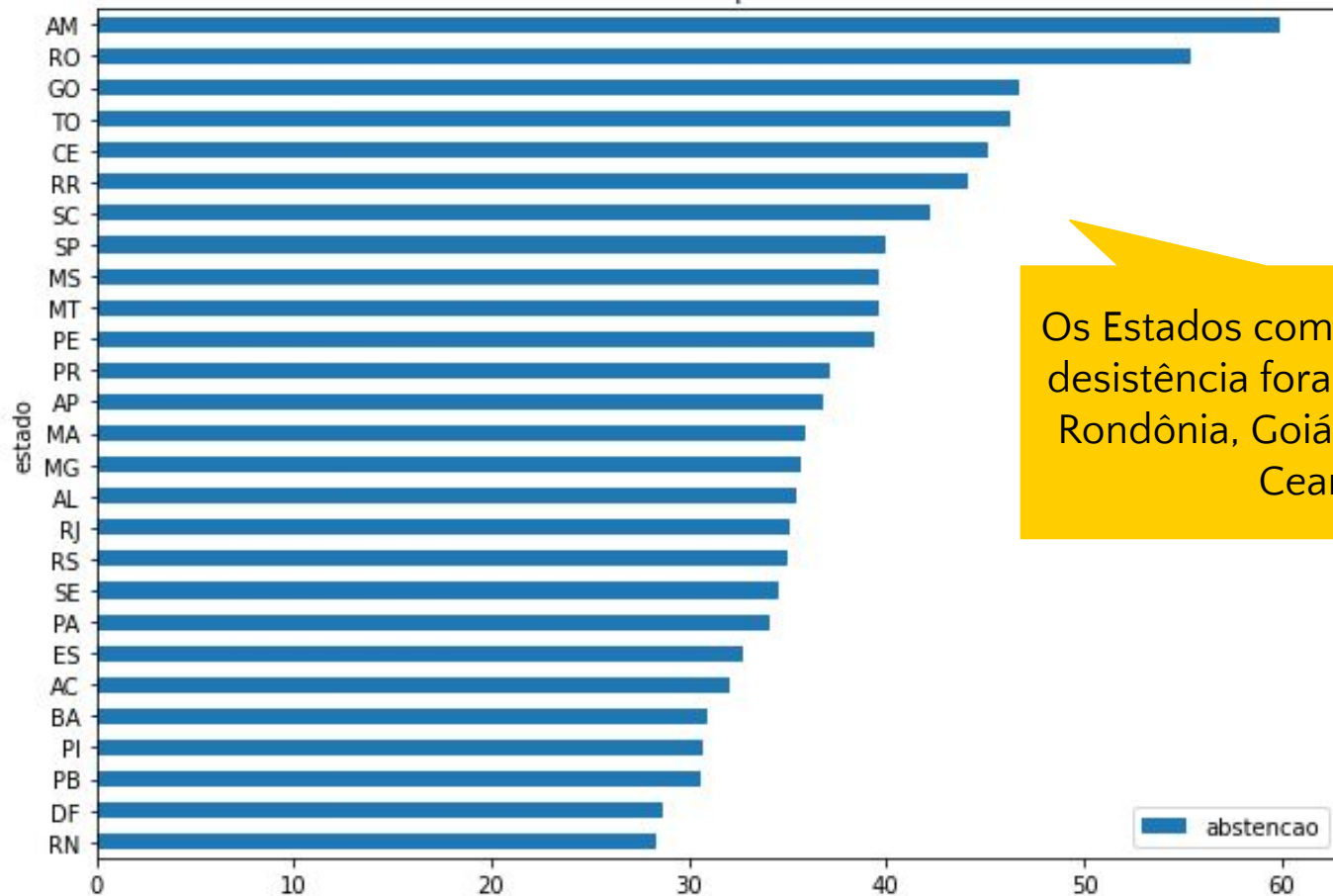


Sobre o tema

- O ENEM de 2020 foi o ano que obteve a maior taxa de desistência;
- Qual o impacto da pandemia no ENEM de 2020?



Desistências por estado

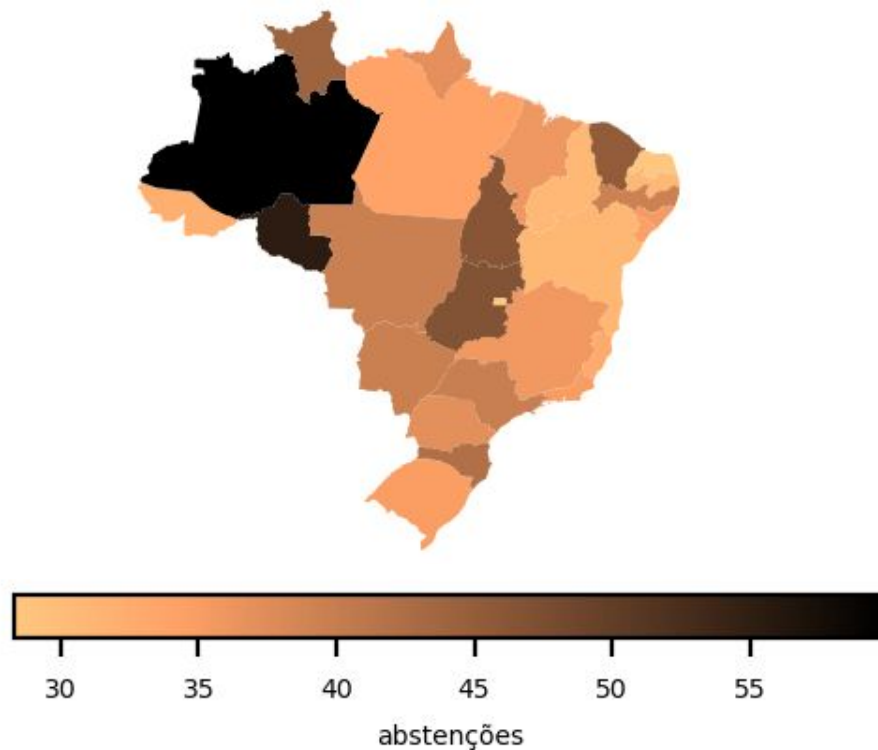


Os Estados com maior taxa de desistência foram: Amazonas, Rondônia, Goiás, Tocantins e Ceará.



Sobre o tema

- Maior taxa de desistência no Norte do país;
- Em seguida, há maior taxa de desistência nas regiões Centro-Oeste e Sul.



2

Objetivos



Objetivo principal

Avaliar o quanto a pandemia impactou no aumento de desistências do ENEM 2020, em comparação ao ano de 2019, e buscar a relação dos resultados com os aspectos socioeconômicos, identificando quais obtiveram maior influência.



Questionário Socioeconômico

Qual a influência dos fatores socioeconômicos do indivíduo sobre as desistências?



COVID-19

A situação da pandemia até a prova está relacionada com as desistências no município?



Aspectos socioeconômicos

Qual a influência da situação da pandemia durante o período antes da prova em conjunto com os aspectos socioeconômicos dos municípios sobre as desistências?

3

Dados



Questionário Socioeconômico

- Microdados do enem: Dados do INEP de 2019 e 2020
 - Questionário socioeconômico: renda, raça/cor, quantidade de determinados itens(carro, televisão, banheiros);
 - Presença do indivíduo, ano de formação, cidade e estado;
 - Filtro de publico/amostra e transformação de vários campos;
 - Diferença entre a taxa de abstenção de 2020 e 2019.



COVID-19

- Dados sobre o COVID-19: Dados do Ministério da Saúde até o dia da prova (17 de Janeiro de 2021).
 - Considerando a população dos municípios, e a contagem total de casos e óbitos, foram gerados os seguintes dados:
 - Casos totais per capita;
 - Mortes totais per capita;
 - Casos totais per capita 1 mês antes da prova;
 - Mortes totais per capita 1 mês antes da prova.



Aspectos socioeconômicos

- Produto Interno Bruto (PIB): Dados do IBGE de 2010;
 - PIB per capita e código do município;
 - Alteração dos nomes das colunas para melhor visualização e transformação do PIB per capita em float.
- Índice de Desenvolvimento Humano Municipal (IDHM):
Dados do PNUD de 2010;
 - IDHM e o código do município;
 - adaptação do nome para código.



Aspectos socioeconômicos

- Densidade demográfica: Dados fornecidos pelo professor;
 - Densidade demográfica e código dos municípios;
 - Alteração dos nomes das colunas para melhor visualização.
- Desempenho escolar: Dados do INEP de 2019 e 2020.
 - Diferença do rendimento escolar entre os anos de 2019 e 2020 por município.



4

Modelos

Foram utilizados modelos de visualização, análise estatística e *machine learning*.



Visualização

- ◉ **Gráficos de mapa**: foram utilizados a biblioteca matplotlib, o geopandas e o pacote geobr;
- ◉ **Correlação**: forma descritiva que mede se há e qual o grau de dependência entre variáveis, ou seja, o quanto uma variável interfere em outra;
- ◉ **Diagrama de dispersão**: verificar a associação entre as variáveis;
- ◉ **Distribuição dos dados**: auxiliar nos testes estatísticos.



Análise estatística

- ◉ **Regressão linear:** verificar a relação entre os fatores socioeconômicos e os dados da pandemia com o aumento das desistências de 2020 em relação a 2019;
- ◉ **Regressão logística:** verificar para cada ano a relação entre os fatores socioeconômicos que influenciam na chance de desistência do aluno.



Machine learning

Clusterização: encontrar padrões entre os estados analisando dados.

5

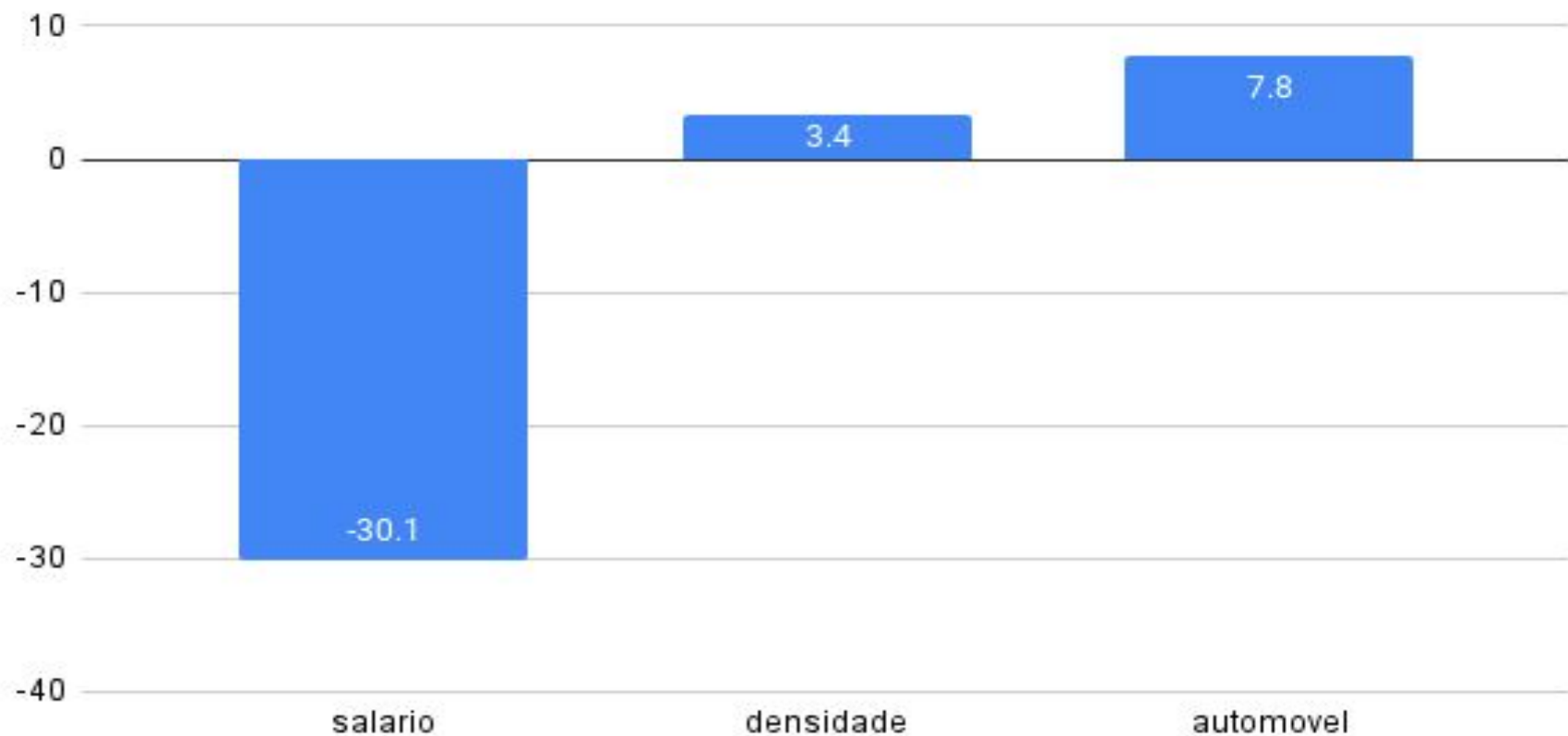
Resultados

Qual a influência dos fatores socioeconômicos para desistência de um indivíduo?

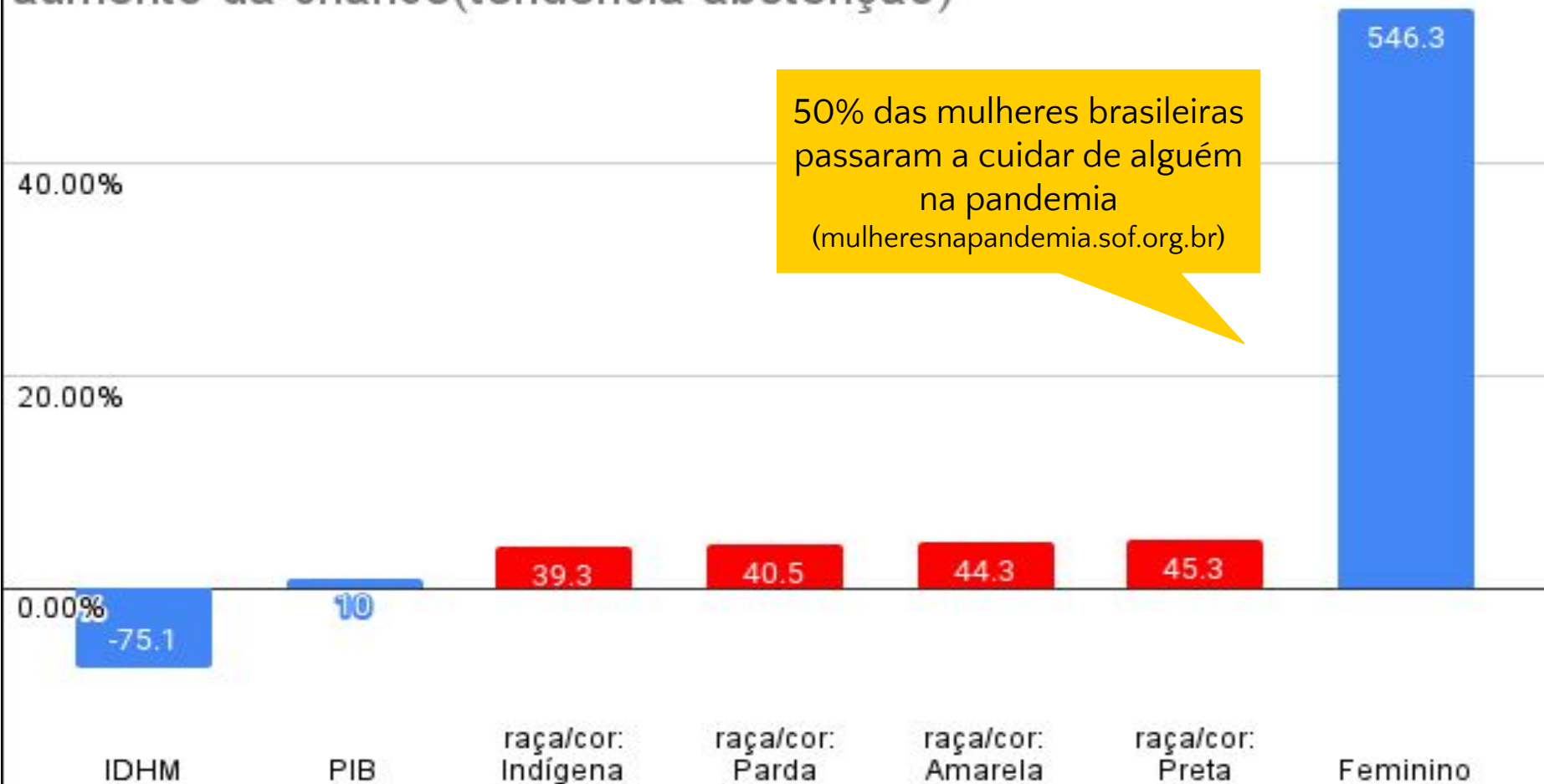
influencia das variáveis na chance de abstenção



aumento da chance(tendência presença)



aumento da chance(tendência abstenção)

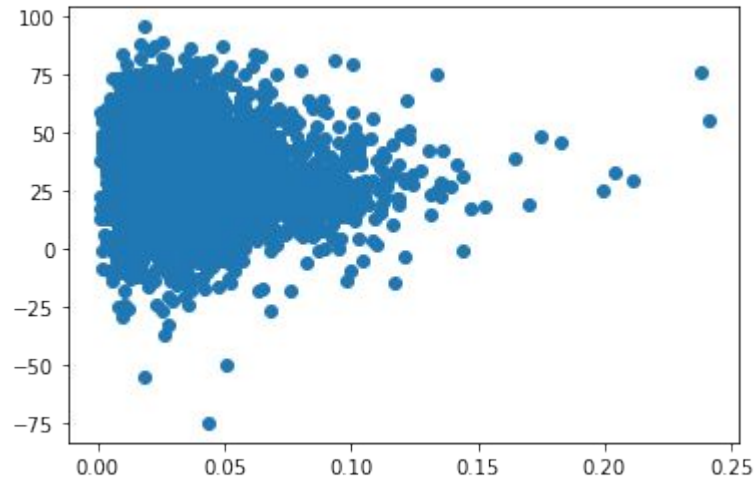


5

Resultados

A situação da **pandemia até a prova** está relacionada com as desistências no município?

COVID não afetou diretamente,
são necessárias mais variáveis
para explicar o impacto



Casos Per Capita

R-squared: 0.001

Adj. R-squared: 0.000

coef	std err	t	P> t
------	---------	---	------

29.1008	0.453	64.183	0.000
---------	-------	--------	-------

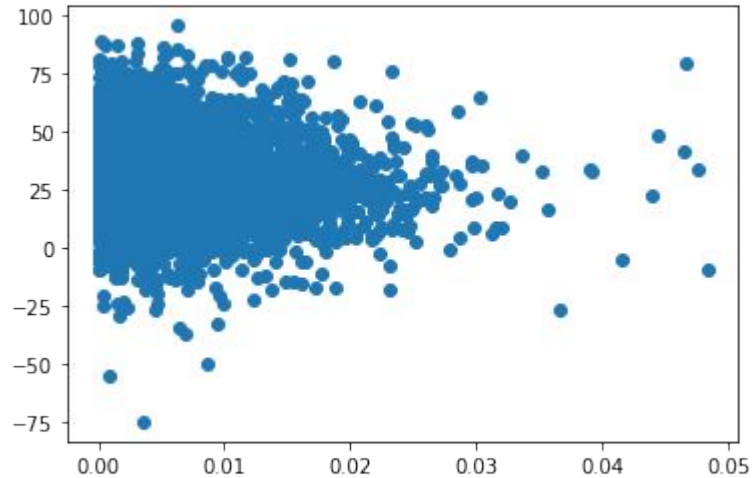
-18.0345	10.632	-1.696	0.090
----------	--------	--------	-------

[0.025	0.975]
--------	--------

28.212	29.990
--------	--------

-38.878	2.809
---------	-------

COVID não afetou diretamente,
são necessárias mais variáveis
para explicar o impacto



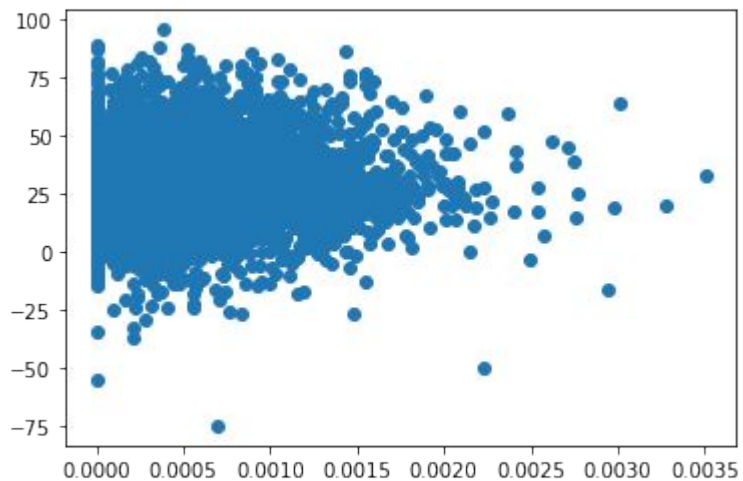
Casos Per Capita 1 Mês Antes

```
=====
R-squared:                                0.007
Adj. R-squared:                          0.007
```

```
=====
coef      std err      t      P>|t|
-----
29.8630    0.364     82.109    0.000
-242.0201  41.764    -5.795    0.000
```

```
=====
[0.025    0.975]
-----
29.150    30.576
-323.896 -160.144
```

Mortes não demonstraram muito impacto, algumas cidades não registraram nenhum óbito.



Mortes Per Capita

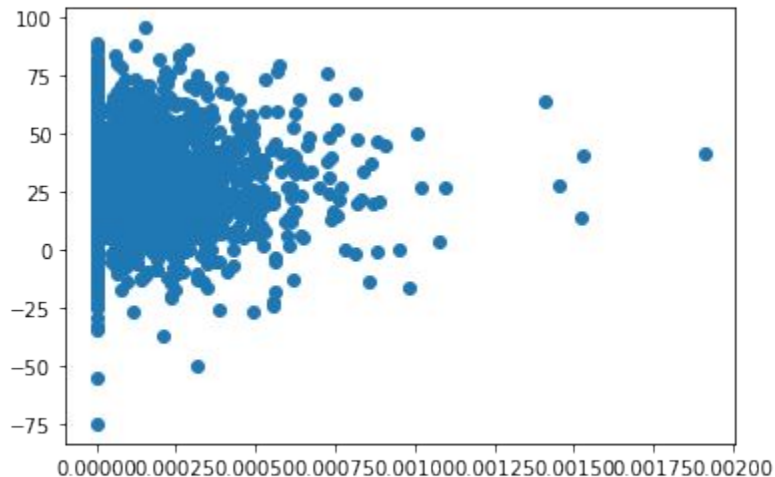
R-squared: 0.000

Adj. R-squared: -0.000

coef	std err	t	P> t
28.4116	0.432	65.754	0.000
-177.4952	556.394	-0.319	0.750

[0.025	0.975]
27.565	29.259
-1268.284	913.293

Durante o período de 1 mês antes da aplicação da prova, metade dos municípios não registraram nenhum óbito.



Mortes Per Capita 1 Mês Antes

R-squared:	0.001
Adj. R-squared:	0.001

coef	std err	t	P> t
28.6918	0.299	96.032	0.000
-3853.0699	1682.283	-2.290	0.022

[0.025	0.975]
28.106	29.278
-7151.119	-555.021

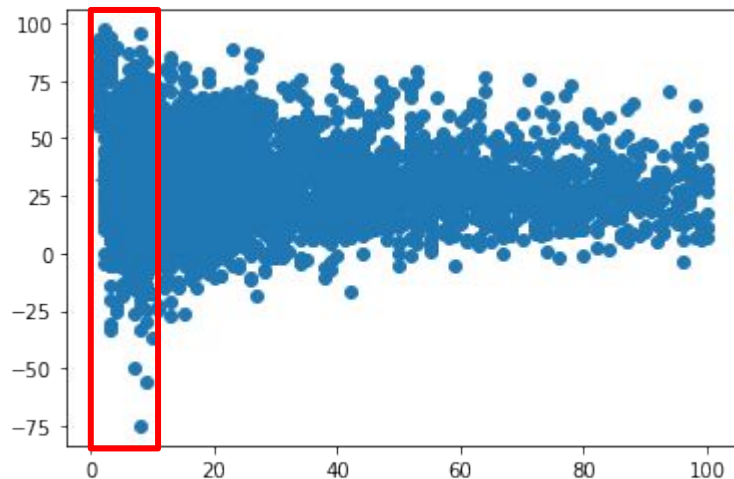
5

Resultados

Qual a influência da situação da pandemia durante o período antes da prova em conjunto com os aspectos socioeconômicos dos municípios sobre as desistências?



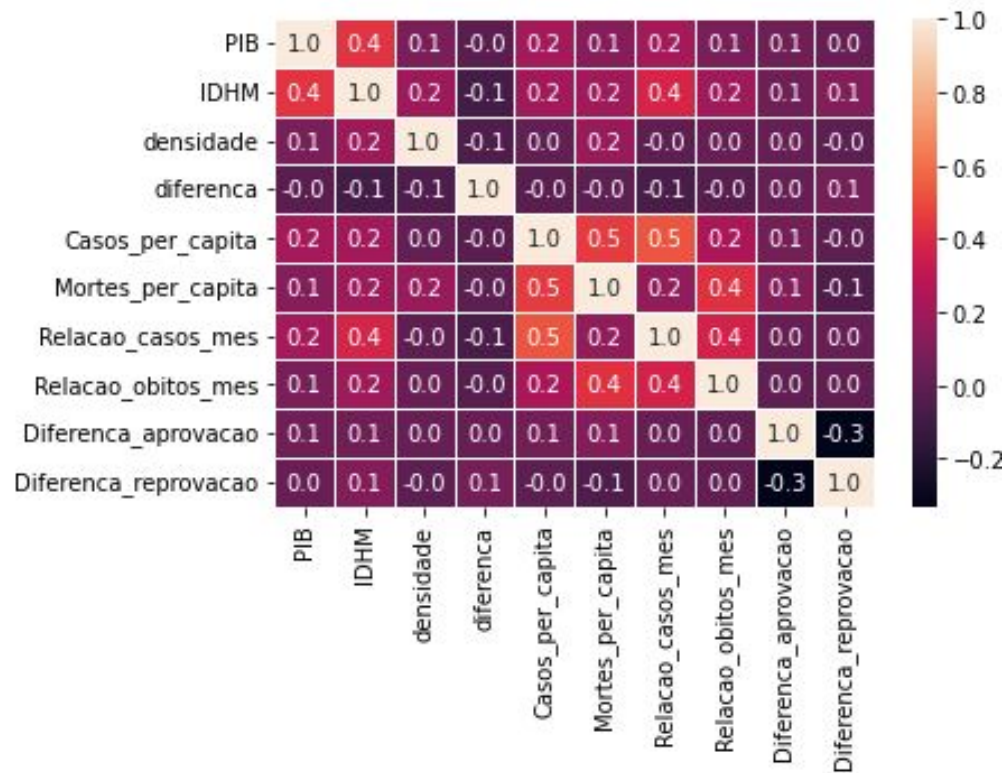
Municípios no ENEM



Redução de 5016 municípios para 4774.

- Municípios com menor número de estudantes cadastrados resultam em diferenças extremas entre os anos;
- Apenas Municípios com pelo menos 5 alunos inscritos foram considerados.

- As variáveis que possuem maior correlação são:
 - IDHM (desenvolvimento) e PIB (renda);
 - Densidade com as outras variáveis não é significativa (p-value alto);
 - Relações de casos e mortes, foi escolhido casos per capita total.
- Diferença de aprovação e reprovação tem correlação negativa.
 - Foi utilizada a reprovação para a análise.



Para os resultados
agrupados por estado

Dep. Variable:	diferenca	R-squared:	0.150
Model:	OLS	Adj. R-squared:	0.145
Method:	Least Squares	F-statistic:	27.90
Date:	Thu, 23 Jun 2022	Prob (F-statistic):	2.68e-143
Time:	10:32:45	Log-Likelihood:	-19880.
No. Observations:	4774	AIC:	3.982e+04
Df Residuals:	4743	BIC:	4.002e+04
Df Model:	30		
Covariance Type:	nonrobust		

- PIB resultou em correlação positiva;
 - Quanto maior o PIB, maior o aumento de desistências;
 - Os mais ricos foram mais afetados;
- Diferença do índice de reprovação resultou em correlação positiva
 - Quanto maior a diferença de reprovação, maior o aumento de desistências;
 - Sobre a formação em 2020, se o participante reprovou é provável que haja menor motivação para fazer a prova.

	coef	std err	t	P> t	[0.025	0.975]
IDHM	-55.9346	5.708	-9.799	0.000	-67.126	-44.743
PIB	2.929e-05	9.95e-06	2.945	0.003	9.79e-06	4.88e-05
Casos_per_capita	-29.3222	11.864	-2.471	0.013	-52.582	-6.063
Diferenca_reprovacao	0.1415	0.041	3.490	0.000	0.062	0.221

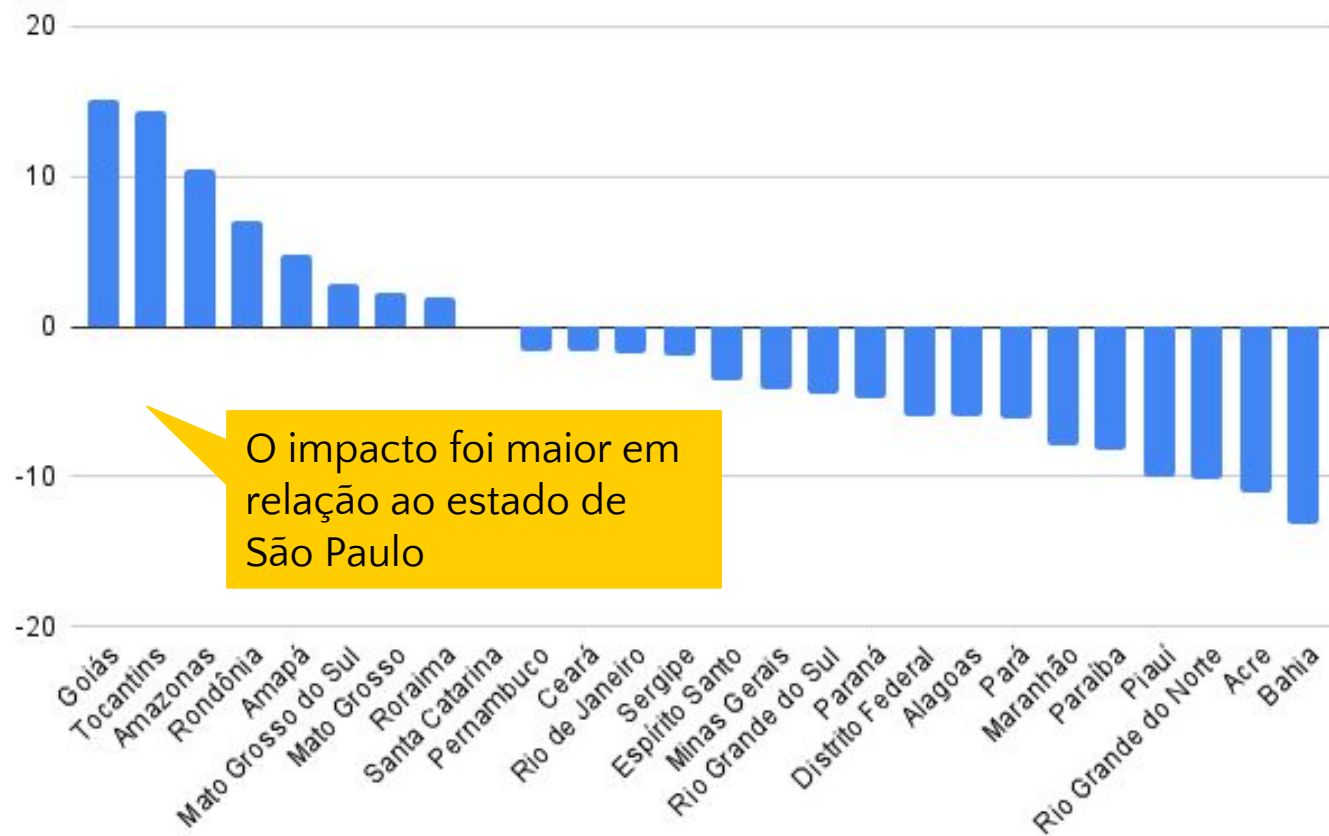
- IDHM resultou em correlação negativa;
 - Quanto menor o IDHM, maior o aumento de desistências;
 - Os moradores com IDHM maior estão mais protegidos.
- Casos per capita resultou em correlação negativa;
 - Quanto menor os casos per capita, maior o aumento de desistências;
 - COVID não teve grande impacto nas desistências.
- Problema maior: condições sociais, ressaltando o IDHM.
 - COVID pode ter impactado indiretamente nestes fatores.

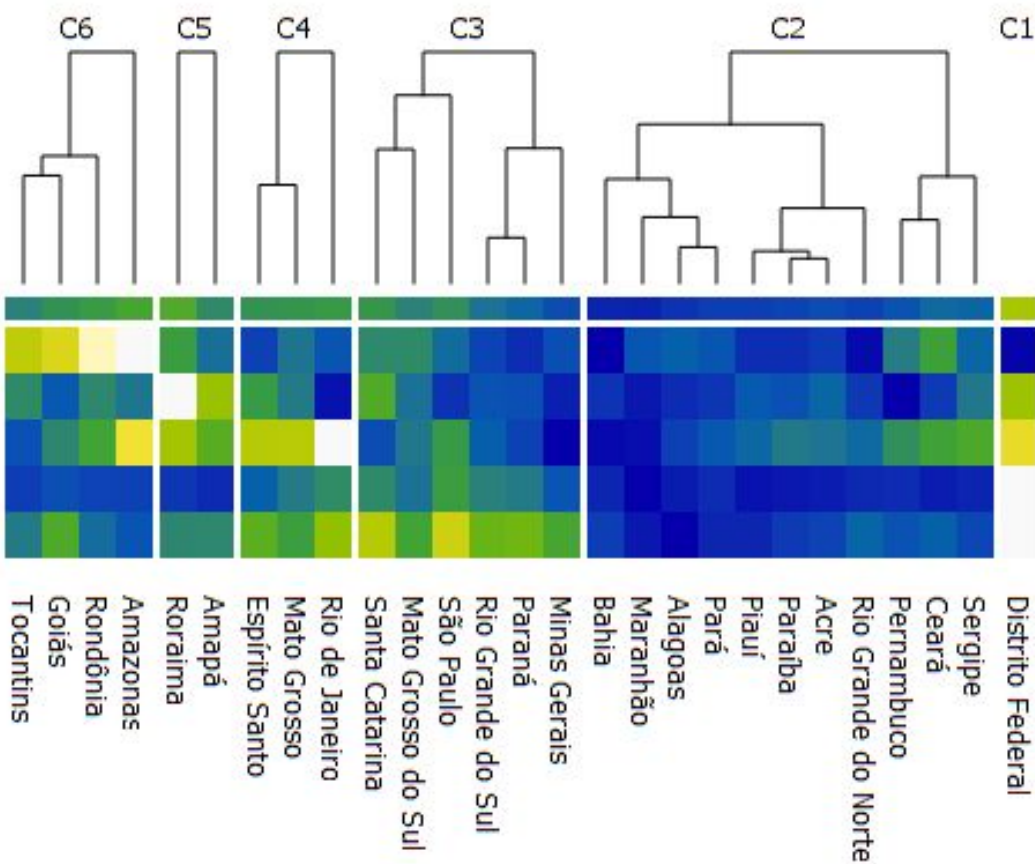
	coef	std err	t	P> t	[0.025	0.975]
IDHM	-55.9346	5.708	-9.799	0.000	-67.126	-44.743
PIB	2.929e-05	9.95e-06	2.945	0.003	9.79e-06	4.88e-05
Casos_per_capita	-29.3222	11.864	-2.471	0.013	-52.582	-6.063
Diferenca_reprovacao	0.1415	0.041	3.490	0.000	0.062	0.221

Na análise por estado, os coeficientes mais altos significam que existem outros fatores que também influenciam e não são explicados e controlados pelo modelo.

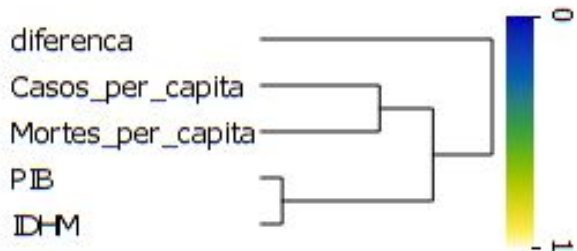
	coef	std err	t	P> t	[0.025	0.975]
Intercept	68.3062	4.138	16.506	0.000	60.193	76.419
C(estado, Treatment(reference='São Paulo'))[T.Acre]	-11.1232	3.597	-3.092	0.002	-18.175	-4.072
C(estado, Treatment(reference='São Paulo'))[T.Alagoas]	-5.9801	1.977	-3.024	0.003	-9.857	-2.104
C(estado, Treatment(reference='São Paulo'))[T.Amapá]	4.7534	4.171	1.140	0.255	-3.424	12.931
C(estado, Treatment(reference='São Paulo'))[T.Amazonas]	10.5587	2.424	4.356	0.000	5.807	15.311
C(estado, Treatment(reference='São Paulo'))[T.Bahia]	-13.0865	1.440	-9.086	0.000	-15.910	-10.263
C(estado, Treatment(reference='São Paulo'))[T.Ceará]	-1.6296	1.513	-1.077	0.282	-4.596	1.337
C(estado, Treatment(reference='São Paulo'))[T.Distrito Federal]	-5.9071	15.653	-0.377	0.706	-36.594	24.780
C(estado, Treatment(reference='São Paulo'))[T.Espírito Santo]	-3.6058	1.983	-1.818	0.069	-7.493	0.282
C(estado, Treatment(reference='São Paulo'))[T.Goiás]	15.1105	1.279	11.818	0.000	12.604	17.617
C(estado, Treatment(reference='São Paulo'))[T.Maranhão]	-7.8791	1.565	-5.036	0.000	-10.946	-4.812

agrupamento dos coeficiente por estado





Quanto maior o IDHM, menor o impacto.



6

Conclusão



Desafios e Limitações

- Não foi possível encontrar os dados de econômicos por município em 2020;
 - Desemprego e desigualdade.
- Dados socioeconômicos não são de 2020;
- Grande quantidade de dados;
- Fatores que influenciam o indivíduo não são os mesmos que influenciam na análise municipal.



Trabalhos futuros

- Categorização dos municípios que aplicaram ou não aplicaram a prova;
- Acrescentar outras variáveis que possam explicar melhor as descobertas do presente trabalho.

Ciência de Dados

Dados Educação



Angela Favretto Pastorello (Design), Bruna Oenning Amador, Daniel de Viveiros Inácio e Matheus Biscaya Gutierrez (Sistemas de Informação)