



Análises dos dados do ENEM em anos pré pandemia e durante a pandemia

Gabriel Rodrigues
Rodrigo Chavoni



Objetivo

- **Pergunta de Pesquisa:** Como os fatores socioeconômicos e regionais influenciam nas notas consideradas *outliers* no ano de 2020 (ano de pandemia) em relação ao esperado pelo modelo de regressão gerado com os dados do exame de 2019?
- **Hipóteses - Fatores socioeconômicos que implicam em uma maior nota:**
 - Alta renda familiar
 - Ensino mais qualificado da Mãe/Pai
 - Profissão mais elitizada Mãe/Pai
 - Ter acesso à internet
 - Ter computador



Dados e modelos

Dados fornecidos pelo INEP referentes aos exames do ENEM dos anos de 2019 e 2020.

Informações utilizadas:

- Média de Notas
- Questões socioeconômicas
- Dados regionais



Dados e modelos

Modelo utilizado: Regressão linear múltipla (statsmodel).

- Variável resposta: Nota do exame
- Variáveis explicativas: Questões socioeconômicas

Aplicar esse modelo de regressão gerado com os dados de antes da pandemia (2019), nos dados do exame no ano da pandemia (2020) e analisar quais fatores influenciaram nos *outliers (under/over performers)*.

Resultados

Análises:

- F de significância
- R^2 e R^2 Ajustado
- *p-value*
- Coeficientes

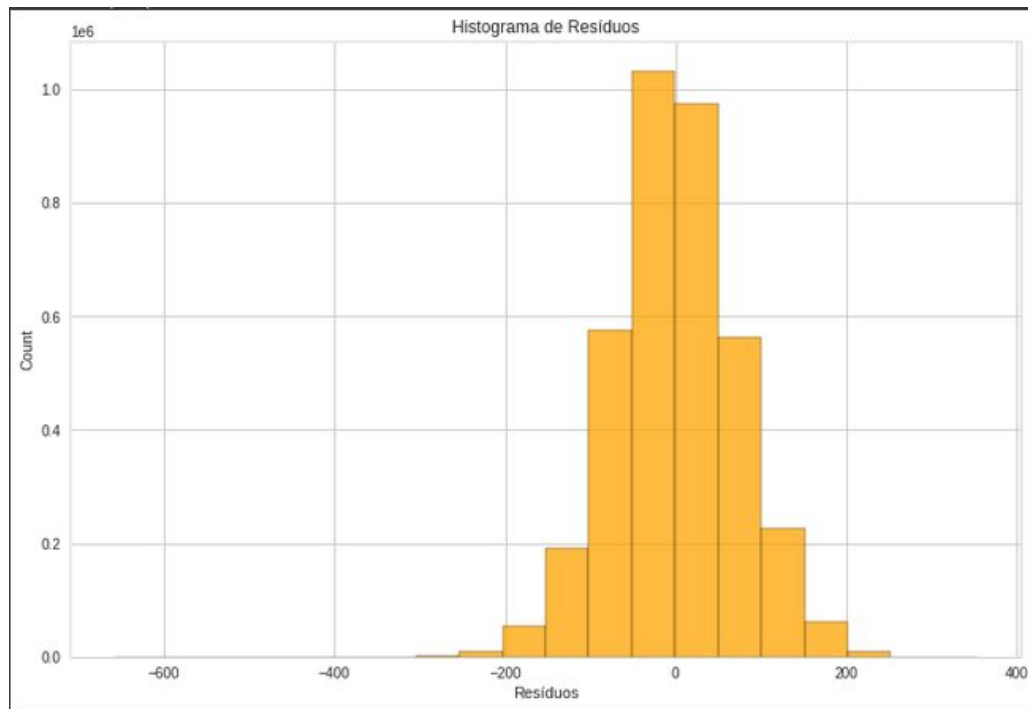
OLS Regression Results						
=====						
Dep. Variable:	Média_Notas	R-squared:	0.288			
Model:	OLS	Adj. R-squared:	0.288			
Method:	Least Squares	F-statistic:	5.342e+04			
Date:	Mon, 05 Dec 2022	Prob (F-statistic):	0.00			
Time:	15:01:20	Log-Likelihood:	-2.1012e+07			
No. Observations:	3701909	AIC:	4.202e+07			
Df Residuals:	3701880	BIC:	4.202e+07			
Df Model:	28					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

Intercept	441.6353	0.227	1943.009	0.000	441.190	442.081
C(Acesso_Internet)[T.B]	12.5219	0.101	123.692	0.000	12.324	12.720
C(Estudo_Mãe)[T.B]	10.0402	0.240	41.863	0.000	9.570	10.510
C(Estudo_Mãe)[T.C]	13.7005	0.249	54.945	0.000	13.212	14.189
C(Estudo_Mãe)[T.D]	15.3773	0.252	60.948	0.000	14.883	15.872
C(Estudo_Mãe)[T.E]	22.3141	0.243	91.752	0.000	21.837	22.791
C(Estudo_Mãe)[T.F]	32.4550	0.273	118.958	0.000	31.920	32.990
C(Estudo_Mãe)[T.G]	29.6920	0.289	102.887	0.000	29.126	30.258
C(Estudo_Mãe)[T.H]	-1.5068	0.335	-4.498	0.000	-2.163	-0.850
C(Estudo_Pai)[T.B]	5.4453	0.198	27.514	0.000	5.057	5.833
C(Estudo_Pai)[T.C]	9.1816	0.213	43.048	0.000	8.764	9.600
C(Estudo_Pai)[T.D]	10.9346	0.221	49.412	0.000	10.501	11.368
C(Estudo_Pai)[T.E]	16.4008	0.211	77.681	0.000	15.987	16.815
C(Estudo_Pai)[T.F]	27.3578	0.252	108.473	0.000	26.864	27.852
C(Estudo_Pai)[T.G]	21.2529	0.288	73.684	0.000	20.688	21.818
C(Estudo_Pai)[T.H]	7.2592	0.240	30.225	0.000	6.788	7.730
C(Profissão_Pai)[T.B]	5.3927	0.142	37.877	0.000	5.114	5.672
C(Profissão_Pai)[T.C]	8.9044	0.141	63.081	0.000	8.628	9.181
C(Profissão_Pai)[T.D]	22.9771	0.163	140.898	0.000	22.657	23.297
C(Profissão_Pai)[T.E]	17.8635	0.246	72.552	0.000	17.381	18.346
C(Profissão_Pai)[T.F]	2.5686	0.172	14.938	0.000	2.232	2.906
C(Profissão_Mãe)[T.B]	4.5805	0.139	33.027	0.000	4.309	4.852
C(Profissão_Mãe)[T.C]	6.8061	0.195	34.964	0.000	6.425	7.188
C(Profissão_Mãe)[T.D]	13.1109	0.174	75.352	0.000	12.770	13.452
C(Profissão_Mãe)[T.E]	11.9585	0.272	43.975	0.000	11.426	12.492
C(Profissão_Mãe)[T.F]	1.5387	0.185	8.303	0.000	1.175	1.902
Renda_Familiar	0.0042	1.64e-05	252.754	0.000	0.004	0.004
QT_Computador	16.5308	0.060	277.506	0.000	16.414	16.648
QT_Carros	3.1028	0.066	46.902	0.000	2.973	3.232

Resultados

Resíduos

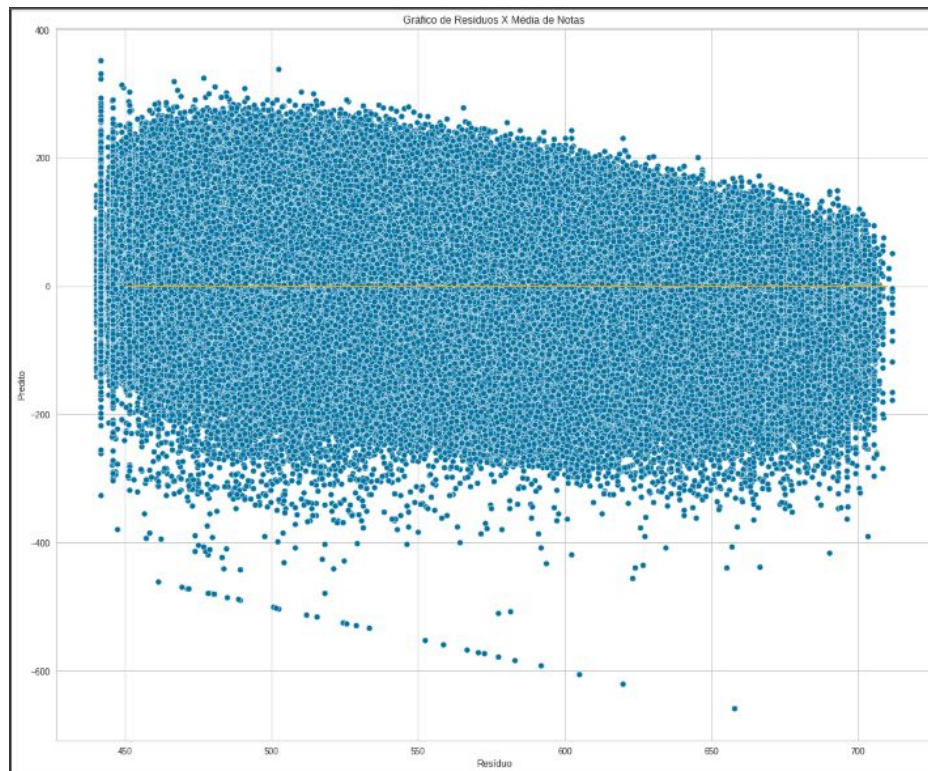
- Média Normal Zero



Resultados

Dispersão Resíduos

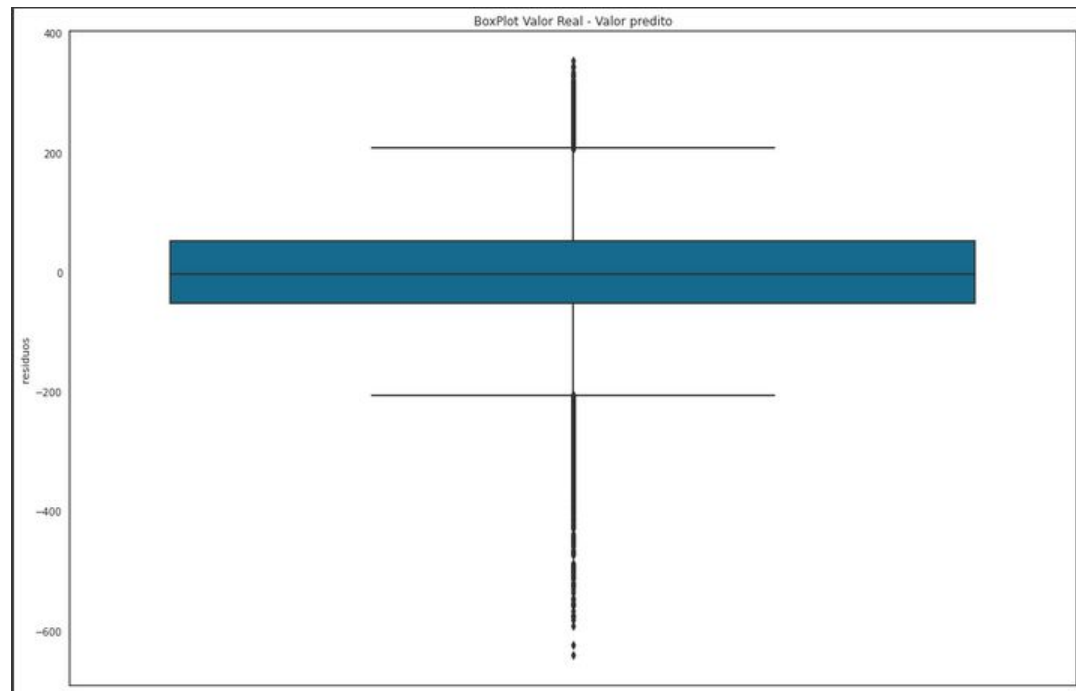
- Pouca dispersão
- Sem tendências



Resultados

Aplicando o modelo para os dados de 2020

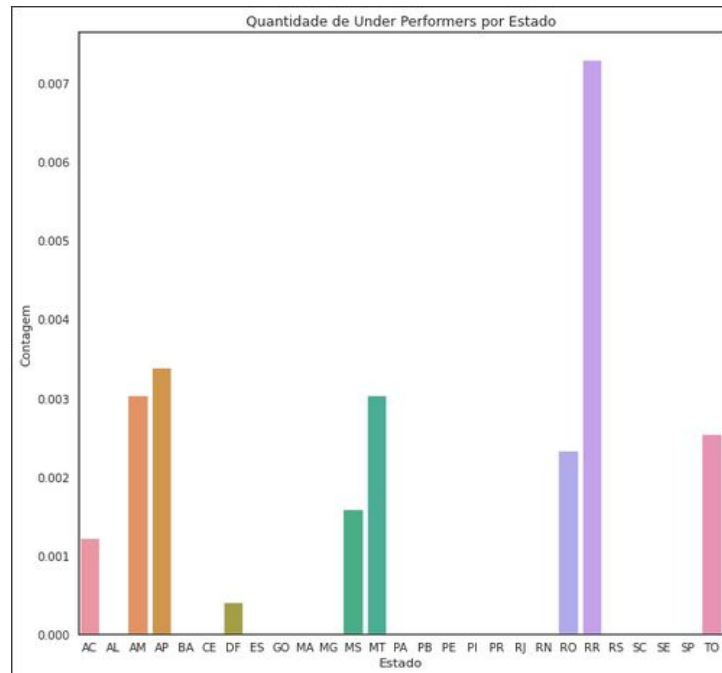
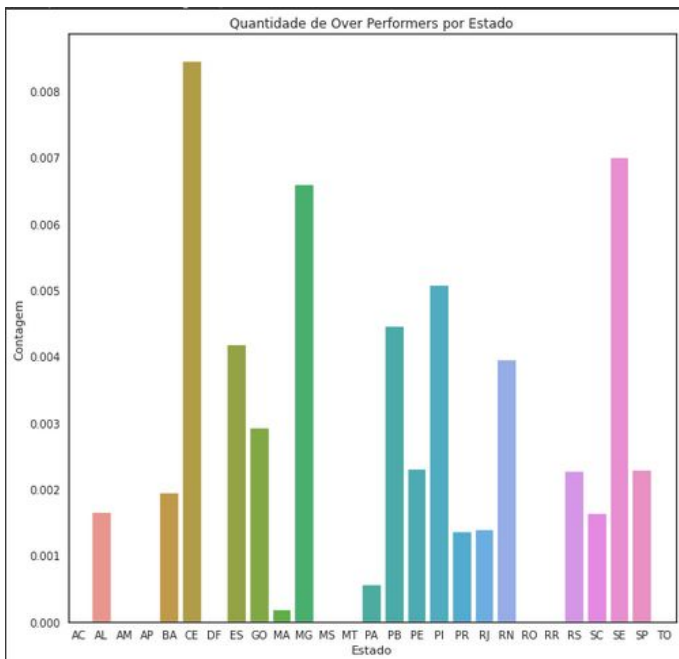
- Outliers



Resultados

Over e Under Performers por Estado:

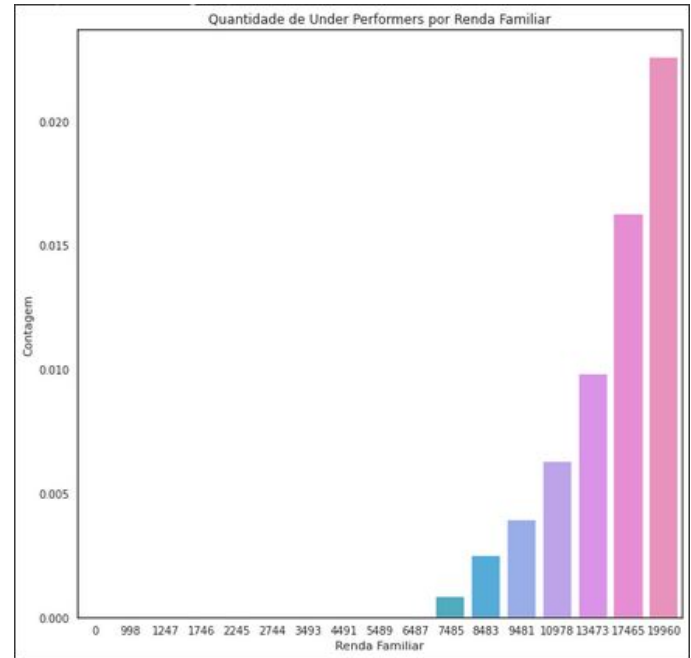
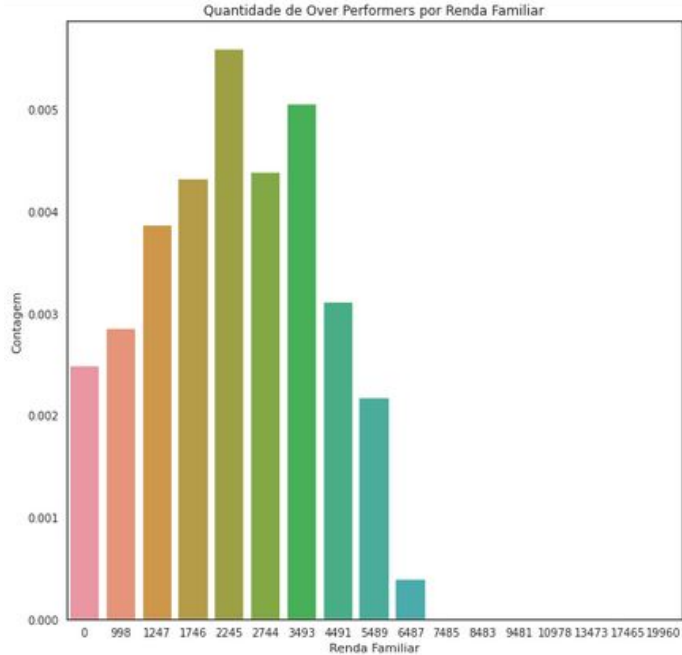
- Destaques over para CE, SE e MG
- Destaques under para RR, AP e AM



Over e Under Performers por Renda:

- Maior número de participantes está na categoria B

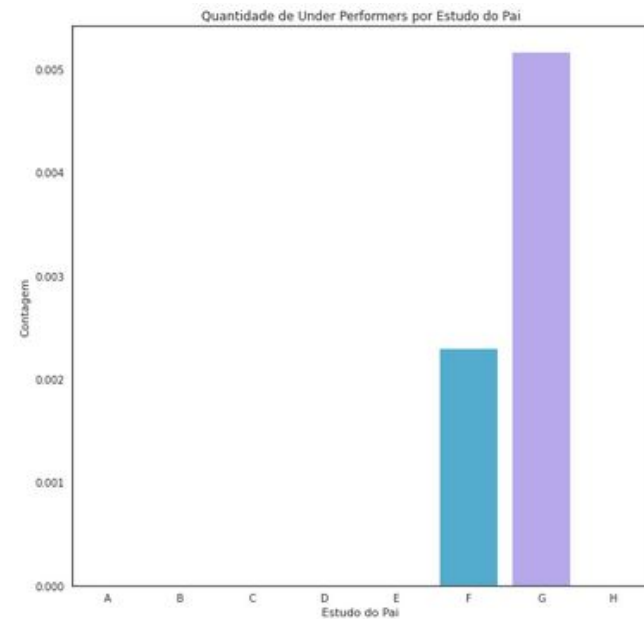
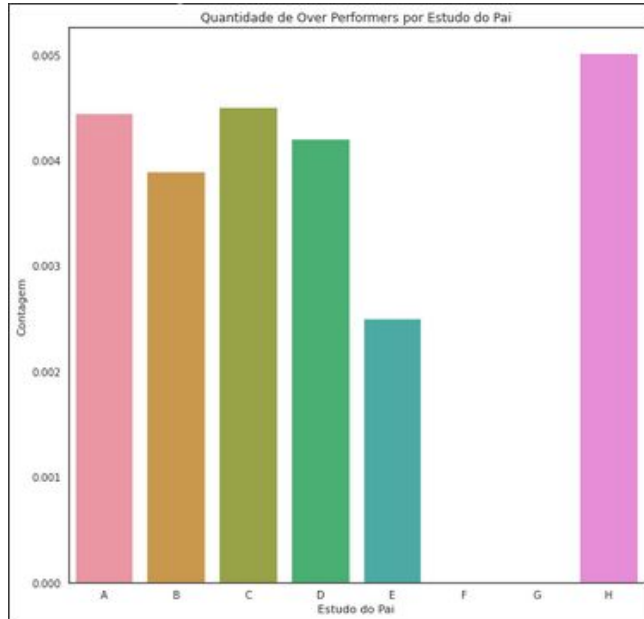
Resultados



Over e Under Performers por Escolaridade:

- Categoria H, não sabiam responder

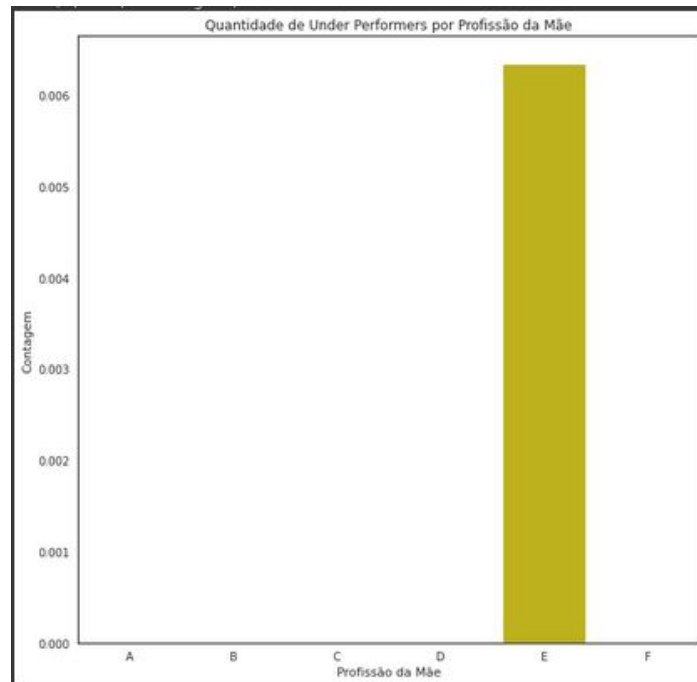
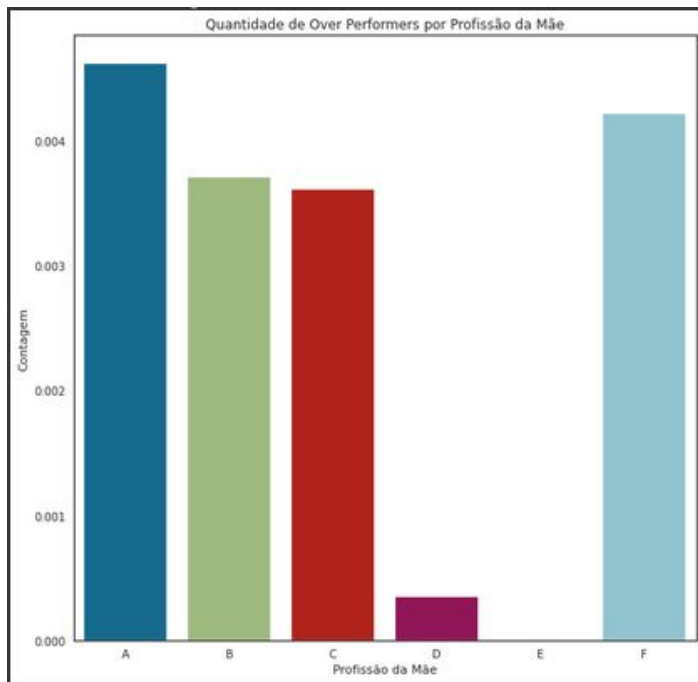
Resultados



Over e Under Performers por Profissão:

- Categoria F, não sabiam responder

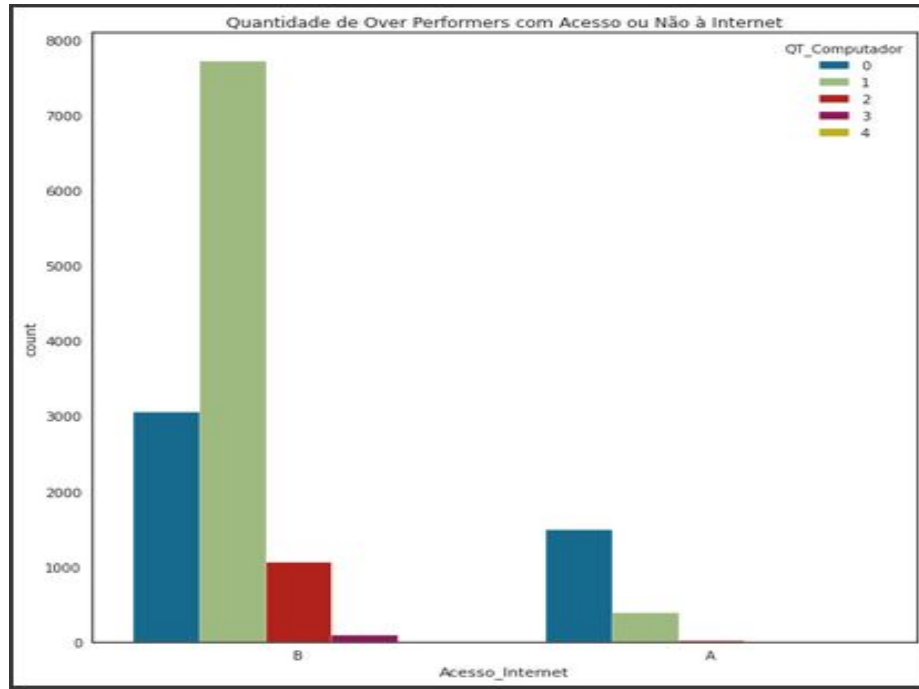
Resultados



Over e Under Performers Acesso à Internet:

- Ano de Pandemia

Resultados





Limitações e Trabalhos Futuros

- Variáveis socioeconômicas não explicam tudo
- Tipo de ensino Privado/Público
- Sinceridade dos participantes na resposta

Trabalhos Futuros:

- Analisar outras variáveis e utilizar dados de mais anos do exame para gerar o modelo.
- Utilizar o modelo para prever as notas dos candidatos de acordo com sua situação socioeconômica.