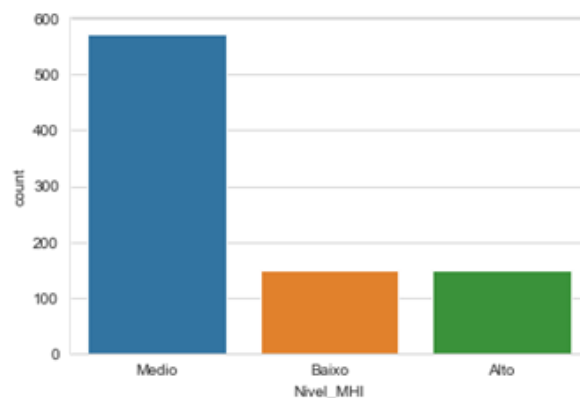


Relatório do desempenho de modelos para classificação do nível do MHI

O relatório está estruturado para de forma que os resultados são apresentados de acordo com cada etapa de aprimoramento do modelo para cada uma das bases de dados utilizadas. Os testes foram feitos para as seguintes bases:

1. Somente o sociodemográfico;
2. Sociodemográfico e todas as respostas do WHOQOL;
3. Sociodemográfico e os domínios do WHOQOL;
4. Sociodemográfico e o índice de qualidade de vida geral (Média das duas primeiras questões do WHOQOL);
5. Sociodemográfico e a média dos domínios do WHOQOL.

O objetivo é classificar os níveis do inventário de saúde mental, que são: baixo, médio e alto. A base de dados original apresentava um desbalanceamento, que pode ser visto na figura abaixo. Para tratar o desbalanceamento foi optado por utilizar uma técnica de *Undersampling*, nominalmente, *Random UnderSampling*.

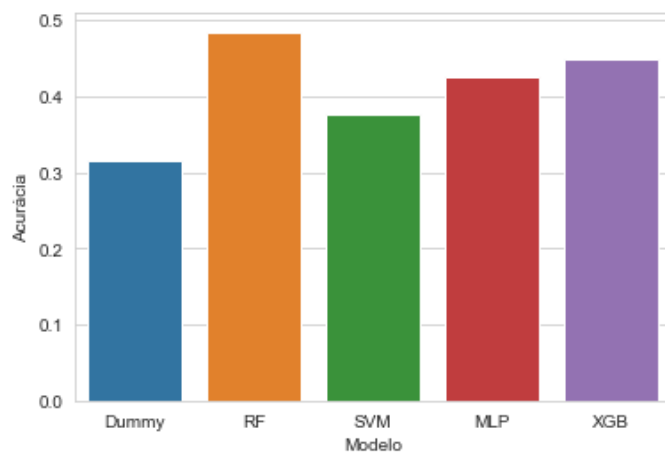


Todos os resultados de desempenho nas tabelas são retirados de um *K-Fold* com $K=5$, a menos que seja explicitado o contrário.

1 DESEMPENHO SOMENTE COM BALANCEAMENTO DE DADOS:

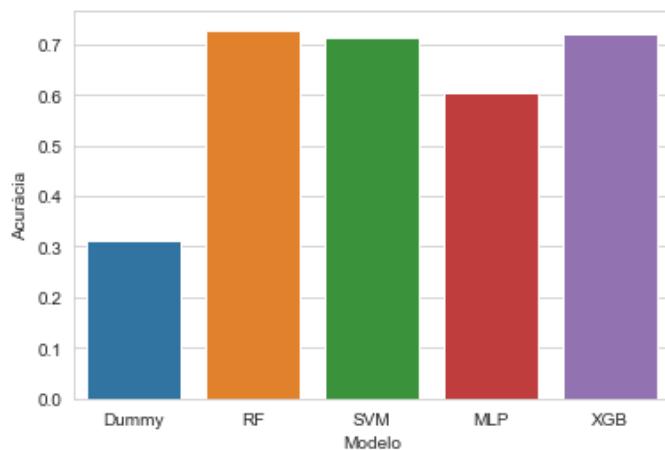
A seguir é apresentado o desempenho de cada um dos modelos, tendo como única alteração o balanceamento de dados através da técnica de *Undersampling*. Com os resultados é possível perceber que com um modelo sem nenhum tipo de *tunning* o Random Forest tem um melhor desempenho em todas as bases excluindo o caso "Sociodemográfico e média dos domínios do WHOQOL", onde o *XGboost* apresenta por volta de 4% de melhoria.

1.1 SOMENTE O SOCIODEMOGRÁFICO;



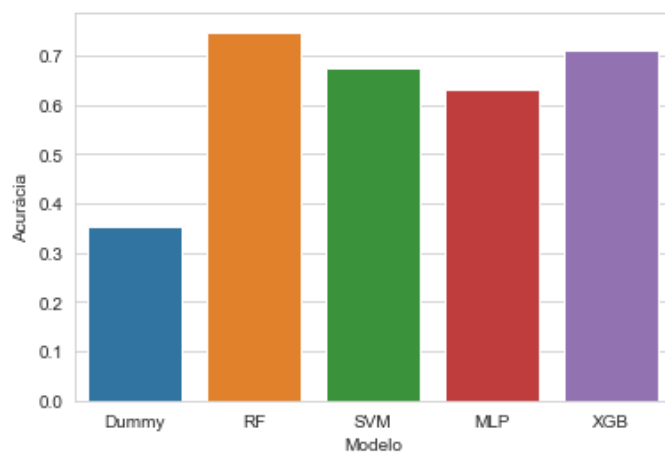
Modelo	Acurácia
Dummy	0.316527
RF	0.481793
SVM	0.375350
MLP	0.417367
XGB	0.448179

1.2 SOCIODEMOGRÁFICO E TODAS AS RESPOSTAS DO WHOQOL



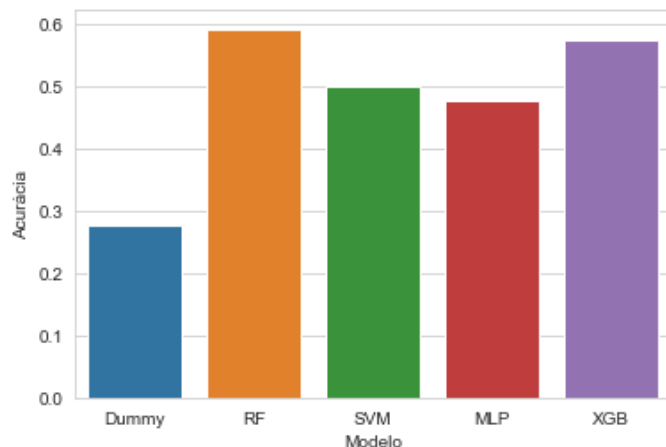
Modelo	Acurácia
Dummy	0.313725
RF	0.742297
SVM	0.714286
MLP	0.616246
XGB	0.719888

1.3 SOCIODEMOGRÁFICO E OS DOMÍNIOS DO WHOQOL;



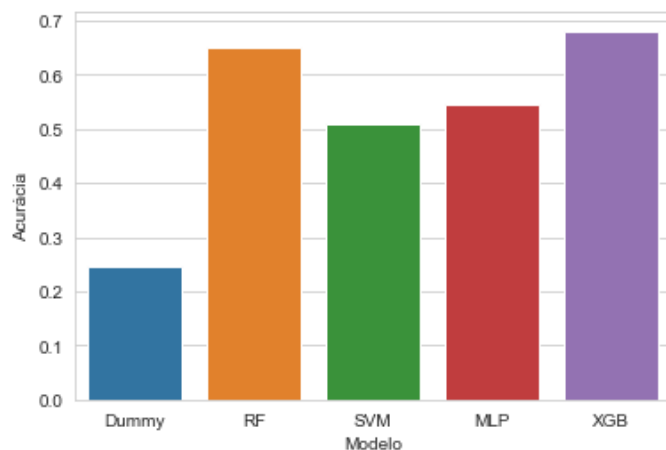
Modelo	Acurácia
Dummy	0.352941
RF	0.761905
SVM	0.675070
MLP	0.621849
XGB	0.711485

1.4 SOCIODEMOGRÁFICO E O ÍNDICE DE QUALIDADE DE VIDA GERAL;



Modelo	Acurácia
Dummy	0.277311
RF	0.565826
SVM	0.498599
MLP	0.481793
XGB	0.574230

1.5 SOCIODEMOGRÁFICO E MÉDIA DOS DOMÍNIOS DO WHOQOL.



Modelo	Acurácia
Dummy	0.246499
RF	0.672269
SVM	0.509804
MLP	0.546218
XGB	0.680672

2 GRIDSEARCH COM BASES BALANCEADAS

Com os resultados dos modelos obtidos, foi realizado um *GridSearch* para poder *tunar* os Hyper parâmetros dos modelos. Com o *tunning* feito os modelos com melhor desempenho em cada uma das bases foi selecionado para enfim ser realizado a escolha dos atributos. É importante ressaltar a melhora de desempenho encontrada com a base “Sociodemográfico e o índice de QV”, assim como o baixíssimo ganho de desempenho para a base “Sociodemográfico”.

Na tabela a seguir é possível visualizar os desempenhos dos modelos após o *tunning*, os modelos com a acurácia destacada em negrito foram usados para o passo de seleção de atributos.

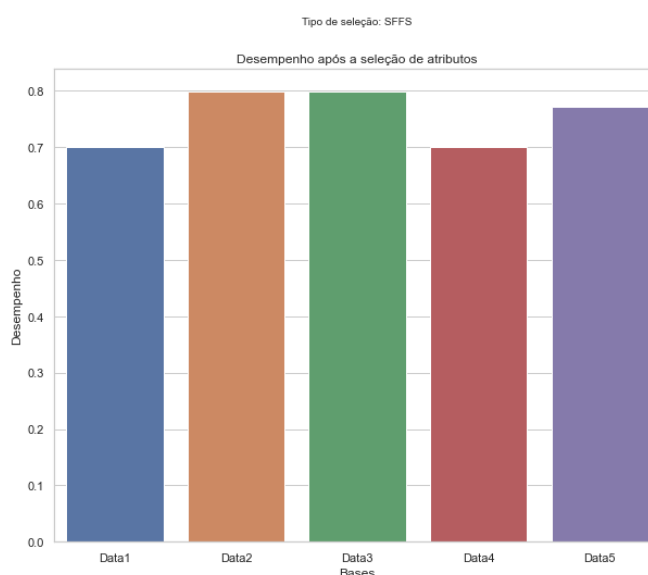
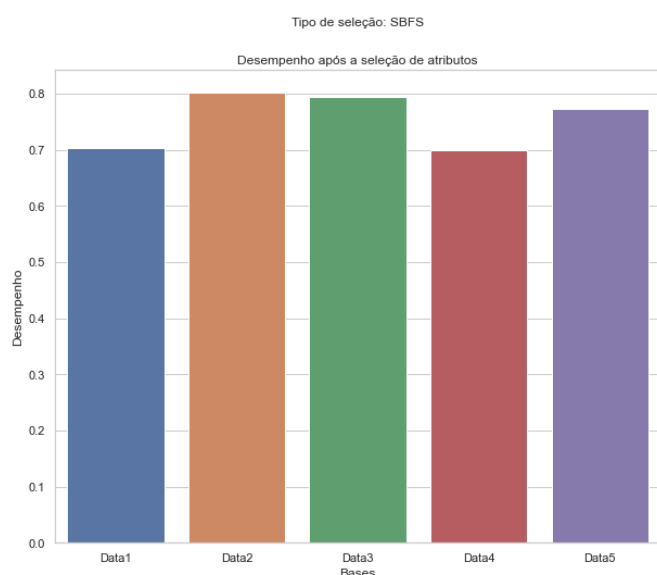
Base de Dados	Modelos			
	RF	SVM	MLP	XGBoost
Sociodemográfico	0.668571	0.651429	0.662857	0.651429
Sociodemográfico e todas as respostas do WHOQOL	0.744444	0.733333	0.688889	0.711111
Sociodemográfico e os domínios do WHOQOL;	0.771429	0.742857	0.731429	0.742857
Sociodemográfico e o índice de QV	0.674286	0.702857	0.702857	0.685714
Sociodemográfico e média dos domínios do WHOQOL	0.680000	0.725714	0.714286	0.742857

3 SELEÇÃO DE ATRIBUTOS

Finalmente, realizando a seleção de atributos para cada um dos modelos escolhidos para cada uma das bases, sendo eles:

- Sociodemográfico – Random Forest – (Data 1)
- Socdem. e todas as respostas do WHOQOL – Random Forest – (Data 2)
- Socdem. e domínios do WHOQOL – Random Forest – (Data 3)
- Socdem. e índices de QV – Support Vector Machine – (Data 4)
- Socdem. e média dos domínios WHOQOL – XGBoost – (Data 5)

Para realizar a seleção de atributos foi utilizada as duas técnicas recomendadas: SBFS e SFFS. Com isso é possível observar os resultados com cada uma das técnicas. A legenda representa a ordem apresentada no início do documento.

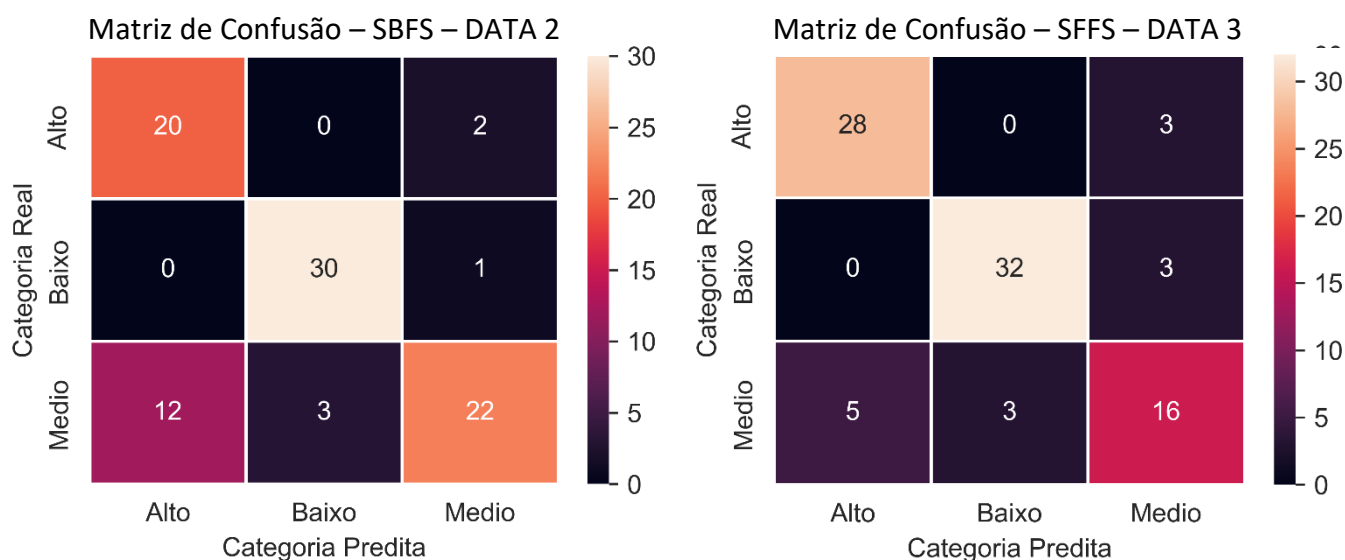


SFFS	
Base	Acurácia
Data 1	0.700670
Data 2	0.798207
Data 3	0.798213
Data 4	0.700690
Data 5	0.770680

SBFS	
Base	Acurácia
Data 1	0.702949
Data 2	0.801668
Data 3	0.793635
Data 4	0.698404
Data 5	0.772959

É possível observar que ambos os métodos obtiveram resultados praticamente idênticos, com diferenças de menos de 1% na acurácia. Além disso, notamos uma melhoria considerável no desempenho para base contendo apenas os dados sociodemográficos.

Analisando a matriz de confusão para os melhores casos de cada método, identificamos uma forte tendência dos modelos em confundir o nível "Médio" com o nível "Alto".



Com base nessas informações, podemos concluir que os modelos que apresentaram melhor desempenho foram "Sociodemográfico com todas as respostas do WHOQOL" e "Sociodemográfico e os domínios do WHOQOL". Ao comparar os atributos selecionados pelo método SBFS, que totalizam quatorze e oito para Data 2 e 3, respectivamente, notamos que apenas uma delas é comum entre as bases mencionadas, que é o "Domínio Psicológico".

Por outro lado, ao analisarmos o método SFFS, onde são selecionados dezessete e dezesseis atributos para Data 2 e 3, respectivamente, percebemos que cinco atributos foram escolhidos para ambas as bases: "Curso é Estressante", "Atividade de Lazer", "Domínio Psicológico", "Turno do Curso" e "Situação Civil".

Os atributos selecionados pelo método SBFS para as duas bases mencionadas são listados a seguir:

<p>Atributos para Data 2:</p> <ul style="list-style-type: none"> • Atividades profissionais; • Atividade de Lazer: <ul style="list-style-type: none"> ○ Ficar com a família; ○ Outros; • Raça: Parda • Tipo de Transporte para universidade • Como você avaliaria sua qualidade de vida? • Quanto satisfeito(a) você está com a sua saúde? • O quanto você consegue se concentrar? • Você tem dinheiro suficiente para satisfazer suas necessidades? • Quanto satisfeito(a) você está consigo mesmo? • Com que frequência você tem sentimentos negativos? • DOMINIO 1(FISICO); • DOMINIO 2(PSICOLOGICO). 	<p>Atributos para Data 3:</p> <ul style="list-style-type: none"> • Já reprovou alguma disciplina/ modulo? • Em termos gerais, você considera seu curso estressante? • Quanto a Bolsas ou Auxílio Universitário: • Atividade de Lazer: <ul style="list-style-type: none"> ○ Assistir Filme/Série • Raça: Amarela • Atualmente reside com? • Qual a sua religião? • DOMINIO 2(PSICOLOGICO).
--	--

Esses resultados indicam que o "Domínio Psicológico" é um atributo relevante e consistente para ambos os métodos de seleção de atributos. Além disso, a análise do método SFFS revela que os atributos "Curso é Estressante", "Atividade de Lazer", "Turno do Curso" e "Situação Civil" também desempenham papéis importantes na predição do desempenho dos modelos.

4 GANHO DE INFORMAÇÃO DE ATRIBUTOS

Realizando uma análise do ganho de informação de todos os atributos utilizados, notando que eles foram classificados nos dois questionários presentes: "WHOQOL" e "Sociodemográfico". Os atributos mais relevantes para a predição do nível socioeconômico pertencem principalmente ao questionário "WHOQOL", que engloba aspectos relacionados à qualidade de vida. Esses atributos são os que têm maior influência na classificação da variável alvo.

Destacam-se os atributos

- "DOMINIO 2(PSICOLOGICO)",
- "DOMINIO 1(FISICO)",
- "Com que frequência você tem sentimentos negativos tais como",
- "Quanto satisfeito(a) você está consigo mesmo?", entre outros.

Esses resultados indicam que a satisfação com a vida, bem-estar emocional e aspectos da qualidade de vida desempenham um papel crucial na determinação do nível dos indivíduos.

Entre os atributos sociodemográficos, a "Idade" é o que apresenta o maior ganho de informação, embora relativamente menor em comparação com os principais atributos do WHOQOL. Outros atributos sociodemográficos também têm algum impacto na classificação do nível de saúde mental como:

- "Frequência de atividade física",
- "Quantidade de horas de sono diárias",
- "Situação Civil",
- "Atividade de Lazer"

Atributos relacionados a "Religião", "Raça", "Uso de substâncias psicoativas" e "Transporte para a universidade" têm ganhos de informação menores, sugerindo menos influência na previsão do MHI.

Em geral, os atributos do questionário WHOQOL têm maiores ganhos de informação em comparação com os atributos sociodemográficos. Isso sugere que aspectos relacionados à percepção pessoal, bem-estar psicológico e físico têm uma influência mais significativa na classificação do nível de saúde mental do que os fatores sociodemográficos. A seguir, segue a tabela com os valores dos ganhos de informação.

Questionário	Atributo	Ganho de Informação
WHOQOL	DOMINIO 2(PSICOLOGICO)	0.496086
WHOQOL	DOMINIO 1(FISICO)	0.338931
WHOQOL	Com que frequência você tem sentimentos negativos tais como	0.317214
WHOQOL	Quão satisfeito(a) você está consigo mesmo?	0.309827
WHOQOL	Quão seguro(a) você se sente em sua vida diária?	0.267769
WHOQOL	DOMINIO 4(MEIO AMBIENTE)	0.240288
WHOQOL	Em que medida você acha que a sua vida tem sentido?	0.230097
WHOQOL	Quão satisfeito(a) você está com a sua saúde?	0.200236
WHOQOL	Você tem energia suficiente para seu dia-a-dia?	0.196213
WHOQOL	O quanto você aproveita a vida?	0.191813
WHOQOL	Quão satisfeito(a) você está com sua capacidade de desempenhar as atividades do seu dia-a-dia?	0.190547
WHOQOL	Como você avaliaria sua qualidade de vida?	0.184565
WHOQOL	DOMINIO 3(RELAÇÕES SOCIAIS)	0.178973

WHOQOL	Você é capaz de aceitar sua aparência física?	0.168726
WHOQOL	Quão satisfeito(a) você está com suas relações pessoais (amigos, parentes, conhecidos, colegas)?	0.159154
WHOQOL	O quanto você precisa de algum tratamento médico para levar sua vida diária?	0.153140
WHOQOL	O quanto você consegue se concentrar?	0.151424
WHOQOL	Quão satisfeito(a) você está com sua capacidade para o trabalho?	0.149601
WHOQOL	Quão satisfeito(a) você está com o seu sono?	0.123811
WHOQOL	Em que medida você tem oportunidades de atividade de lazer?	0.091586
WHOQOL	Quão saudável é o seu ambiente físico (clima, barulho, poluição, atrativos)?	0.089722
WHOQOL	Quão satisfeito(a) você está com sua vida sexual?	0.087191
WHOQOL	Em que medida você acha que sua dor (física) impede você de fazer o que você precisa?	0.084564
WHOQOL	Quão satisfeito(a) você está com o apoio que você recebe de seus amigos?	0.082748
WHOQOL	Você tem dinheiro suficiente para satisfazer suas necessidades?	0.079502
WHOQOL	Quão disponíveis para você estão as informações que precisa no seu dia-a-dia?	0.076359
WHOQOL	Quão satisfeito(a) você está com as condições do local onde mora?	0.058223
Sociodemográfico	Qtd. aproximada de horas de sono diárias?	0.055492
WHOQOL	Quão bem você é capaz de se locomover?	0.049910

WHOQOL	Quão satisfeito(a) você está com o seu meio de transporte?	0.045271
Sociodemográfico	Idade	0.039201
Sociodemográfico	Faz uso contínuo de algum medicamento para tratamento da saúde?	0.036363
WHOQOL	Quão satisfeito(a) você está com o seu acesso aos serviços de saúde?	0.028601
Sociodemográfico	Considera seu curso estressante?	0.022827
Sociodemográfico	Frequência atividade física (por semana)?	0.021020
Sociodemográfico	Quantas disciplinas e/ou módulos está cursando no semestre em questão?	0.020295
Sociodemográfico	Horas de lazer diária	0.017528
Sociodemográfico	Religião (Outras/Sem religiões)	0.016862
Sociodemográfico	Atividade de Lazer (Passear/viajar)	0.015995
Sociodemográfico	Freq. de uso substância psicoativa nos últimos 3 meses?	0.015834
Sociodemográfico	Já se afastou dos estudos?	0.014742
Sociodemográfico	Fez uso de substância psicoativa nos últimos 3 meses?	0.014318
Sociodemográfico	Caso sua cidade de origem não seja Sobral. Com que frequência viaja para ela?	0.012816
Sociodemográfico	Em que semestre você está matriculado	0.012538
Sociodemográfico	Atividade de Lazer (Ficar com a família)	0.011770
Sociodemográfico	Religião (Católica)	0.011435
Sociodemográfico	Realiza atividade física de forma regular?	0.009764
Sociodemográfico	Situação Civil (Casado)	0.008947
Sociodemográfico	Sexo	0.008395
Sociodemográfico	Situação Civil (Solteiro)	0.006683
Sociodemográfico	Atividade de Lazer (Dormir)	0.006624
Sociodemográfico	Atividade de Lazer (At. Fis.)	0.005518
Sociodemográfico	Turno (Matutino)	0.004869

Sociodemográfico	Reside com Amigos/Colegas	0.003939
Sociodemográfico	Já reprovou alguma disciplina/ modulo?	0.003749
Sociodemográfico	Atividade de Lazer (Ler)	0.003630
Sociodemográfico	Religião (Testemunha de Jeová)	0.003394
Sociodemográfico	Atividade de Lazer (Ouvir música)	0.002942
Sociodemográfico	Atividade de Lazer (Outras)	0.002646
Sociodemográfico	Reside com Familiares	0.002385
Sociodemográfico	Turno (Vespertino)	0.002348
Sociodemográfico	Transporte para à Universidade (Mototáxi)	0.002176
Sociodemográfico	Raça (Preta)	0.001969
Sociodemográfico	Raça (Amarela)	0.001913
Sociodemográfico	Raça (Não sabe)	0.001672
Sociodemográfico	Turno (Noturno)	0.001518
Sociodemográfico	Religião (Evangélica)	0.001228
Sociodemográfico	Realiza atividades acadêmicas?	0.001130
Sociodemográfico	Reside em Republica Estudantil	0.000884
Sociodemográfico	Reside Sozinho	0.000782
Sociodemográfico	Transporte para à Universidade (Transporte próprio)	0.000730
Sociodemográfico	Situação Civil (Viúvo)	0.000692
Sociodemográfico	Transporte para à Universidade (A pé)	0.000692
Sociodemográfico	Transporte para à Universidade (A pé/ carona/ bicicleta)	0.000669
Sociodemográfico	Recebe Bolsas ou Auxílio Universitário?	0.000548
Sociodemográfico	Turno (Integral)	0.000483
Sociodemográfico	Religião (Espírita)	0.000413
Sociodemográfico	Raça (Parda)	0.000358
Sociodemográfico	Realiza atividades profissionais?	0.000276
Sociodemográfico	Local de origem (Sobral)	0.000224
Sociodemográfico	Transporte para à Universidade (Transporte coletivo)	0.000197
Sociodemográfico	Atividade de Lazer (Assistir filmes/ séries)	0.000149

Sociodemográfico	Situação Civil (União estável)	0.000102
Sociodemográfico	Raça (Branca)	0.000066