

Trabalho de Introdução aos Sistemas Inteligentes

Arthur Henrique Chaves Oliveira
Sophia Neves Alvim Ottoni
Lucas Rodrigues Caetano Matos

Resumo

Este trabalho analisa a base de dados "Data for Admission in the University", que contém informações de 400 alunos, incluindo notas de exames (GRE e TOEFL), avaliações de universidades, qualidade de cartas de recomendação (LOR) e declarações de propósito (SOP), participação em pesquisas, GPA e chance de admissão. Utilizando métodos estatísticos e gráficos, investigamos quais atributos influenciam mais a chance de admissão em programas de mestrado no exterior. A análise é feita com base na mediana e em outras categorias, destacando a importância do desempenho acadêmico e das habilidades em inglês na obtenção de uma vaga.

Palavras-chave: notas, GPA, SOP, LOR, GRE, TOEFL, desempenho acadêmico.

Sumário

1	BASE DE DADOS.....	3
1.1	Dicionário de dados.....	3
1.2	Redução da base de dados.....	4
1.3	Resultado.....	4
2	O PROBLEMA.....	4
3	ANÁLISE EXPLORATÓRIA.....	4
3.1	Admissão maior que 90%.....	5
3.2	TOEFL acima e abaixo da mediana.....	6
3.2.1	<i>Acima da mediana.....</i>	6
3.2.2	<i>Abaixo da mediana.....</i>	6
3.3	CGPA acima ou abaixo da mediana.....	7
3.3.1	<i>Acima da mediana.....</i>	7
3.3.2	<i>Abaixo da mediana.....</i>	7
4	GRÁFICOS.....	8
4.1	Gráfico geral.....	8
4.2	Gráfico baseado em Chance of Admit.....	9
4.2.1	<i>TOEFL.....</i>	9
4.2.2	<i>CGPA.....</i>	9
4.2.3	<i>GRE Score.....</i>	10
4.3	Gráfico baseado no TOEFL.....	11
4.3.1	<i>CGPA.....</i>	11
4.3.2	<i>GRE Score.....</i>	12
5	CONCLUSÃO.....	13

1 BASE DE DADOS

Para a realização deste trabalho foi decidido a utilização da base de dados “*Data for Admission in the University*”. Esta base de dados fornece informações sobre diversos alunos, como: GRE Scores, TOEFL score, university rating, etc; além da chance que eles possuem de serem admitidos nas universidades. Ela possui 400 dados de diferentes alunos.

1.1 Dicionário de dados

A tabela original apresenta um total de 9 dados, sendo eles: *Serial Number*, *GRE score*, *TOEFL score*, *University Rating*, *SOP*, *LOR*, *CGPA*, *Research* e *Chance of Admit*.

Dicionário de dados		
Coluna/Atributo	Tipo de dado	Descrição
Serial Number	Números Inteiros (int)	Dado que nos informa o índice da pessoa referente a linha, ele vai de 1 a 400.
GRE score		Atributo que guarda a nota obtida pelo aluno no GRE(Graduate Record Examination), indo de 0 a 340.
TOEFL score		Atributo que nos informa a nota obtida pelo aluno no TOEFL(Test of English as a Foreign Language), os valores desta coluna variam entre 0 e 120.
University Rating	Categórico ordinal	Define a qualidade da universidade em que o aluno está tentando passar. Os dados dessa coluna vão de 1 a 5, sendo: 1 - qualidade muito baixa, 2 - qualidade baixa, 3 - qualidade razoável, 4 - qualidade boa, 5 - qualidade muito boa.
LOR		Atributo que informa a qualidade da Letter of Recommendation que foi enviado à universidade. Esse atributo varia de 1 a 5 (1, 1.5, 2, 2.5, ..., 5), sendo valores mais próximos de 1 uma qualidade ruim e valores mais próximos de 5 uma qualidade boa
SOP		Atributo que guarda a qualidade do Statement of Purpose que foi enviado à universidade. Esse atributo varia de 1 a 5 (1, 1.5, 2, 2.5, ..., 5), sendo valores mais próximos de 1 uma qualidade ruim e valores mais próximos de 5 uma qualidade boa
Research	Booleano	Define se a pessoa referência já participou de

		pesquisas ou não
CGPA	Número Real (float)	Dado que informa o valor do Undergraduate Grade Point Average do aluno, esse valor varia de 1 a 10
Chance of Admit		Atributo que guarda a chance do aluno passar na universidade escolhida, esse valor está em porcentagem.

1.2 Redução da base de dados

Após uma análise dos atributos presentes na tabela foi-se considerado desnecessário a presença do atributo “*Serial Number*”, portanto sua remoção foi feita por meio de um código em python.

1.3 Resultado

Tendo sido feita a redução de atributos a base de dados fica desta forma:
Imagem 1 - Tabela de dados

	GRE Score	TOEFL Score	University Rating	SOP	LOR	CGPA	Research	Chance of Admit
0	337	118	4	4.5	4.5	9.65	1	0.92
1	324	107	4	4.0	4.5	8.87	1	0.76
2	316	104	3	3.0	3.5	8.00	1	0.72
3	322	110	3	3.5	2.5	8.67	1	0.80
4	314	103	2	2.0	3.0	8.21	0	0.65

Fonte: tabela modificada pelos autores

2 O PROBLEMA

Nos dias atuais diversos alunos querem realizar um curso de mestrado no exterior, no entanto muitos deles não sabem o que é necessário fazer para poder entrar, ou o quão bem eles precisam ir nos testes para poder passar. Pensando nisso, o nosso projeto busca identificar quais atributos possuem uma maior influência na chance de uma pessoa ser admitida no mestrado de uma determinada universidade.

Tendo como exemplo um aluno que teve seu CGPA baixo, um dado que não poderia ser alterado a não ser que seja realizada uma nova graduação, deveria então focar em outros campos para poder ter uma maior chance de admissão. Ele então poderia utilizar de nossos resultados para procurar os pontos de melhoria precisos.

3 ANÁLISE EXPLORATÓRIA

Uma análise estatística da tabela se torna necessária para podermos tirar conclusões sobre a mesma. Esta análise seria feita por meio das médias, medianas,

e modas dos atributos, esses valores são obtidos pela utilização de um código em python.

Tabela 1 - Análise da média, mediana, moda, valores máximos e mínimos dos atributos da base de dados

Coluna/Atributo	Média	Mediana	Moda	Máx	Min
GRE score	316.8	317	312; 324	340	290
TOEFL score	107.41	107	110	120	92
University Rating	3.08	3	3	5	1
LOR	3.45	3.5	3	5	1
SOP	3.4	3.5	3.5; 4	5	1
CGPA	8.59	8.61	8	9.92	6.8
Chance of Admit	72.43%	73%	64%	97%	34%

Fonte: elaborada pelos autores

Com uma análise da tabela é possível notar uma separação muito grande entre os valores mínimos e máximos dos atributos, além de valores de média e mediana sempre muito próximos.

Para podermos analisar ainda mais a fundo os valores obtidos foi feita uma separação dos dados pela mediana e faixa de valores de alguns deles, resultando em algumas categorias de dados, sendo elas: pessoas com a chance de admissão maior que 90%, pessoas que tiveram o TOEFL acima e abaixo da mediana, pessoas com o CGPA abaixo e acima da mediana.

3.1 Admissão maior que 90%

A separação das pessoas que possuíam uma chance maior ou igual a 90% melhora a visualização de possíveis padrões em quem possui a maior chance de ser admitido em uma universidade, como por exemplo o fato de os valores de CGPA estarem todos acima de 9.16, ou o fato da moda do GRE para essa faixa de valores ser igual a 340.

Tabela 2 - Análise da média, da mediana, moda, valores máximos e mínimos dos atributos das pessoas que tiveram uma chance de admissão $\geq 90\%$

Coluna/Atributo	Média	Mediana	Moda	Máx	Min
GRE score	333.24	334	340	340	320
TOEFL score	116.26	116	118	120	110
University Rating	4.56	5	5	5	2
LOR	4.42	4.5	4.5	5	3
SOP	4.56	4.5	5	5	3
CGPA	9.5	9.46	9.36	9.92	9.16
Chance of Admit	93.22%	93%	94%	97%	90%

Fonte: elaborada pelos autores

3.2 TOEFL acima ou abaixo da mediana

Através da separação da base entre pessoas que a nota do TOEFL estava abaixo ou acima do valor da mediana do mesmo atributo pode nos dar uma ideia do quão importante é ter um inglês bem desenvolvido.

3.2.1 Acima da mediana

Após a separação dos dados e da aplicação de métodos de estatística podemos notar alguns fatos sobre as pessoas que tiveram uma nota acima da mediana, sendo eles: o aumento nos valores de quase todas as estatísticas em relação à tabela geral e que o valor da moda do GRE é a mesma da tabela geral, o que indica que a maioria dos dados tiveram o valor do TOEFL acima da mediana. Tabela 3 - Análise da média, da mediana, moda, valores máximos e mínimos dos atributos das pessoas que tiveram uma pontuação no TOEFL maior que o valor da mediana

Coluna/Atributo	Média	Mediana	Moda	Máx	Min
GRE score	324	324	324	340	301
TOEFL score	111.95	111	110	120	107
University Rating	3.7	4	3	5	1
LOR	3.83	4	4	5	1.5
SOP	3.91	4	4	5	2
CGPA	8.97	9.02	8.76; 9.11	9.92	7.7
Chance of Admit	80.91%	81%	94%	97%	45%

Fonte: elaborada pelos autores

3.2.2 Abaixo da mediana

Utilizando essa separação dos dados e a aplicação de métodos de estatística novamente, mudando apenas a faixa de valores para abaixo da mediana, obtemos os seguintes fatos: a diminuição dos valores de quase todas as estatísticas em relação à tabela geral e que todos os dados de mínimo são idênticos aos da tabela geral, revelando assim que todos as pessoas que tiveram os menores valores estão abaixo da mediana da pontuação do TOEFL.

Tabela 4 - Análise da média, da mediana, moda, valores máximos e mínimos dos atributos das pessoas que tiveram uma pontuação no TOEFL abaixo do valor da mediana

Coluna/Atributo	Média	Mediana	Moda	Máx	Min
GRE score	308.27	309	312	327	290
TOEFL score	102.01	103	105	106	92
University Rating	2.36	2	2	5	1
LOR	3	3	3	5	1
SOP	2.78	2.5	3	5	1
CGPA	8.15	8.18	8	9.04	6.8

Chance of Admit	62.38%	64%	64%	82%	34%
-----------------	--------	-----	-----	-----	-----

Fonte: elaborada pelos autores

3.3 CGPA acima ou abaixo da mediana

Utilizando novamente a separação por meio da mediana, mudando apenas o atributo de referência para CGPA, poderemos notar padrões dos alunos que tinham boas notas durante suas graduações.

3.3.1 Acima da mediana

Fazendo uso dessa separação, obtemos alguns fatos para as pessoas com valores de CGPA acima da mediana geral, sendo um deles a proximidade dos valores dessa tabela com os da tabela 3, indicando assim uma relação entre CGPA e TOEFL para os dados acima da mediana.

Tabela 5 - Análise da média, da mediana, moda, valores máximos e mínimos dos atributos das pessoas que tiveram uma CGPA acima do valor da mediana

Coluna/Atributo	Média	Mediana	Moda	Máx	Min
GRE score	324.69	325	324	340	299
TOEFL score	111.46	111	110	120	99
University Rating	3.82	4	3	5	1
LOR	3.95	4	4	5	1.5
SOP	4.01	4	4	5	2
CGPA	9.08	9.06	8.76	9.92	8.62
Chance of Admit	82.56%	83%	94%	97%	52%

Fonte: elaborada pelos autores

3.3.2 Abaixo da mediana

Recorrendo aos resultados da separação novamente, mudando apenas os valores alvos para abaixo da mediana, tivemos os seguintes fatos: a proximidade dos valores dessa tabela com os da tabela 4, indicando assim uma relação entre CGPA e TOEFL para os dados abaixo da mediana e o fato de todos os mínimos desta tabela condizem com os da tabela geral, mostrando que todos as pessoas que tiveram os menores valores estão abaixo da mediana de CGPA.

Tabela 6 - Análise da média, da mediana, moda, valores máximos e mínimos dos atributos das pessoas que tiveram uma CGPA abaixo do valor da mediana

Coluna/Atributo	Média	Mediana	Moda	Máx	Min
GRE score	308.92	309	312	334	290
TOEFL score	103.35	104	105	116	92
University Rating	2.35	2	2	5	1
LOR	2.95	3	3	5	1
SOP	2.78	3	3	5	1
CGPA	8.11	8.17	8	8.6	6.8

Chance of Admit	62.30%	64%	64%	81%	34%
-----------------	--------	-----	-----	-----	-----

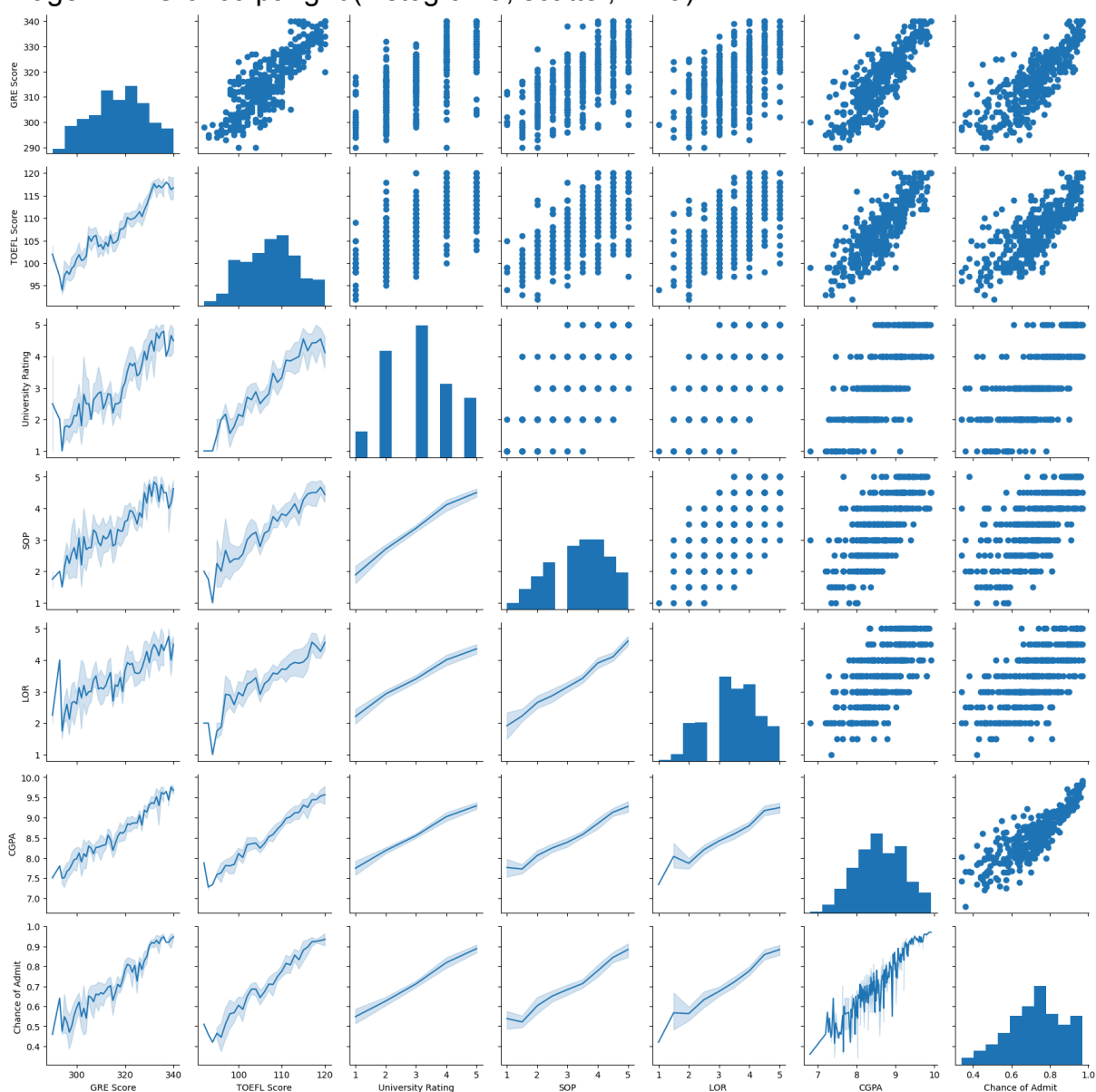
Fonte: elaborada pelos autores

4 GRÁFICOS

Pela utilização do python como ferramenta de desenvolvimento de gráficos, foi possível a criação de gráficos que nos ajudam a visualizar relações entre as variáveis.

4.1 Gráfico geral

Este gráfico nos possibilita a visualização das relações de todas as variáveis por meio de três tipos de gráficos diferentes, sendo eles: histograma, scatter, linha. Imagem 2 - Gráfico pairgrid(histograma, scatter, linha)



Fonte: elaborado pelos autores

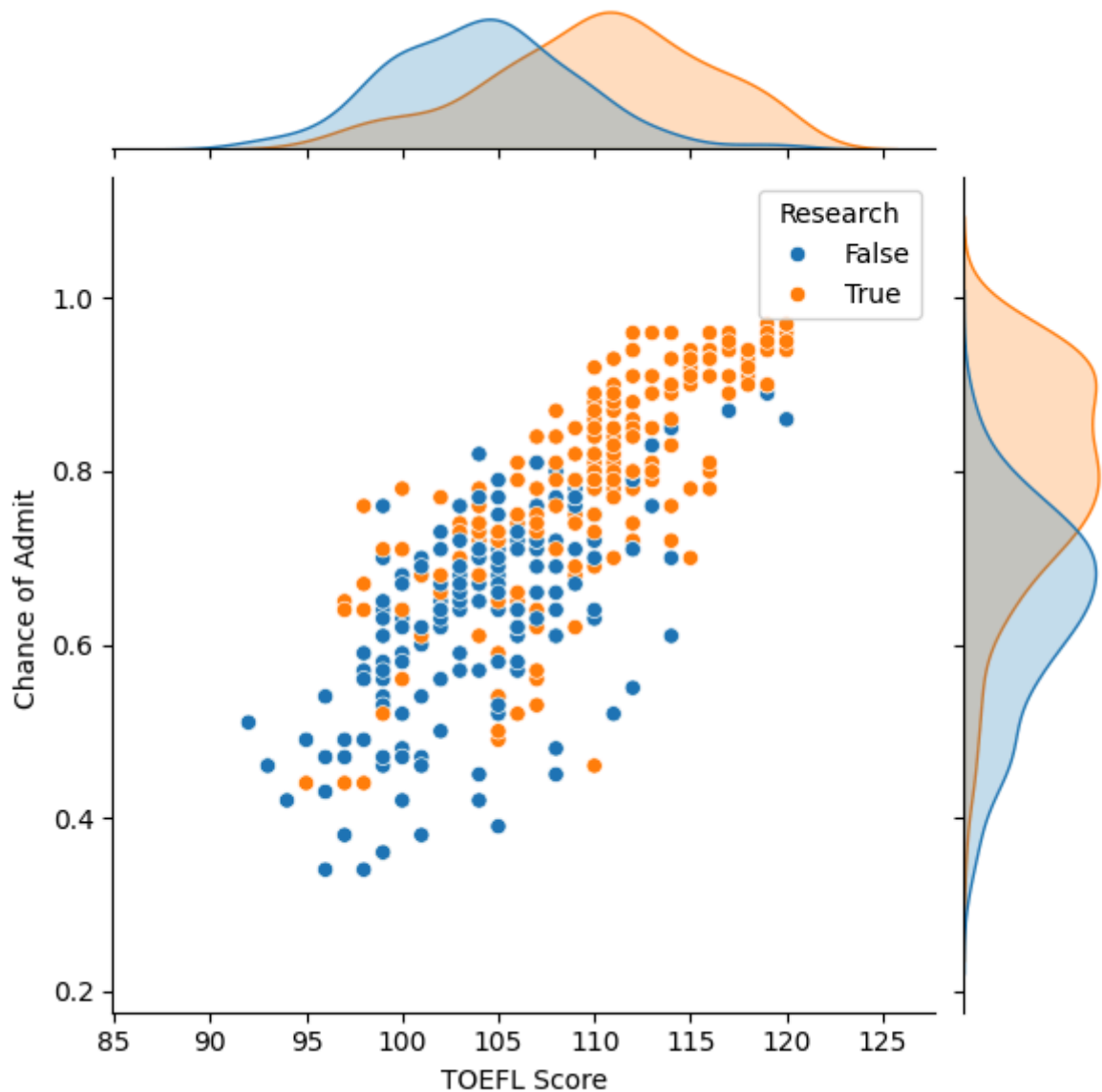
4.2 Gráfico baseado em Chance of Admit

Foi desenvolvido alguns gráficos que avaliam a chance de admissão de uma pessoa(que seria o eixo Y dos gráficos) com base em outros atributos(que seriam o eixo X), além de separar se essa pessoa já participou de uma pesquisa ou não.

4.2.1 TOEFL

Aqui analisamos a relação da chance de admissão com a pontuação do TOEFL.

Imagem 3 - Chance of Admit X TOEFL Score, separados por Research

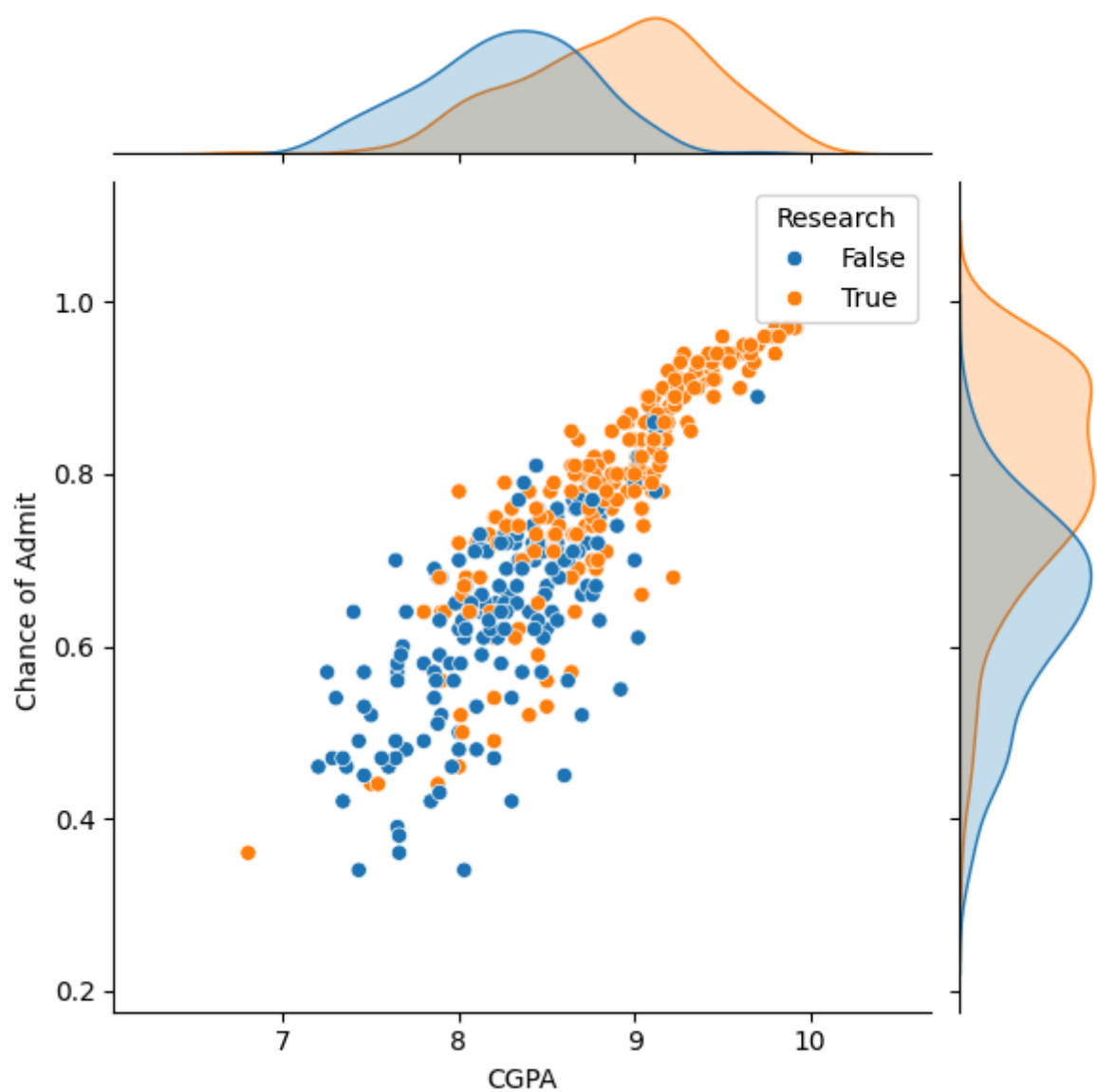


Fonte: elaborado pelos autores

4.2.2 CGPA

Aqui analisamos a relação da chance de admissão com a pontuação do CGPA.

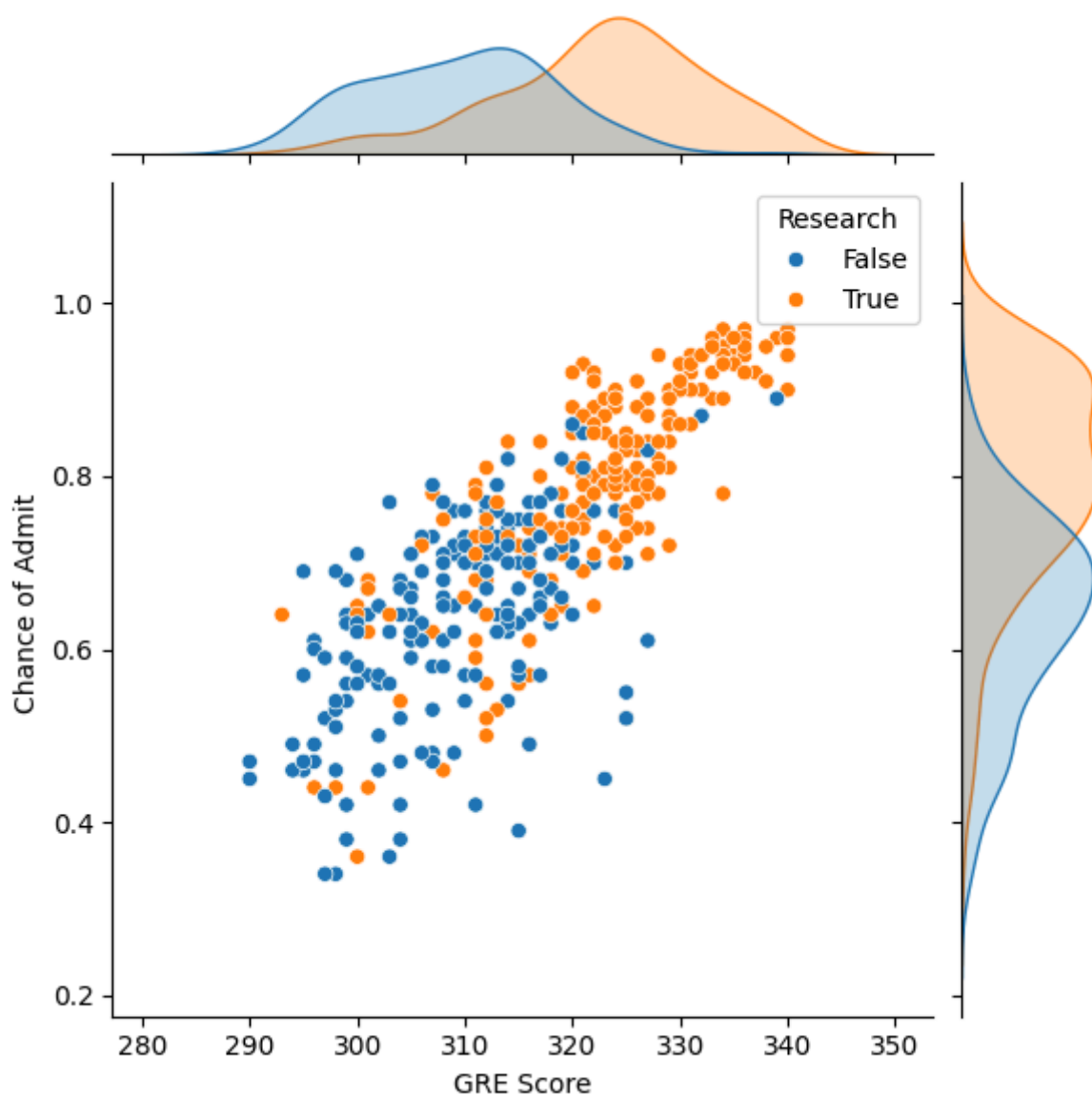
Imagem 4 - Chance of Admit X CGPA, separados por Research



Fonte: elaborado pelos autores

4.2.3 GRE Score

Aqui analisamos a relação da chance de admissão com a pontuação do GRE. Imagem 5 - Chance of Admit X GRE Score, separados por Research



Fonte: elaborado pelos autores

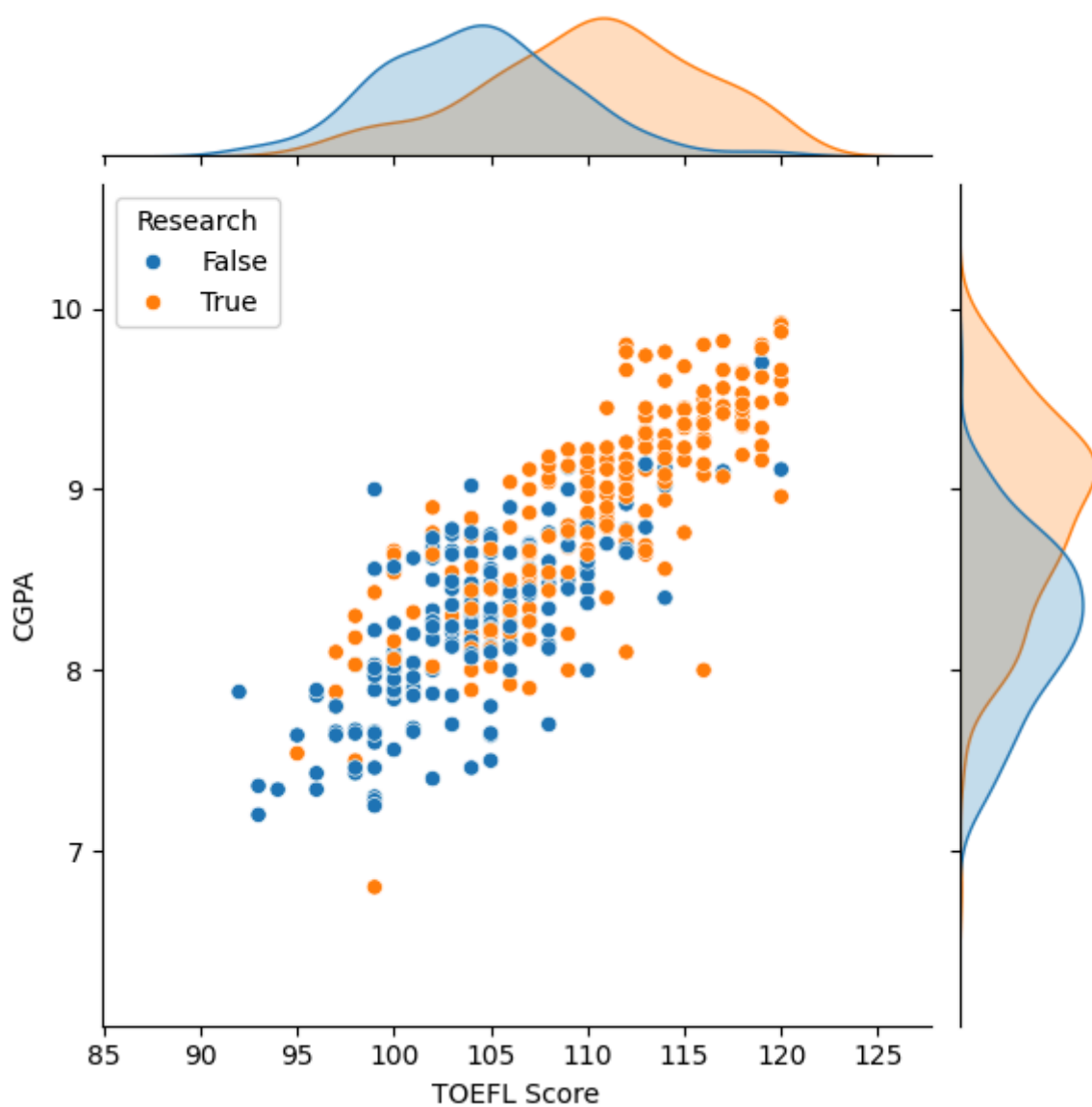
4.3 Gráfico baseado no TOEFL

A criação de gráficos que avaliam a pontuação do TOEFL de uma pessoa(que seria o eixo Y dos gráficos) com base em outros atributos(que seriam o eixo X), além de separar se essa pessoa já participou de uma pesquisa ou não.

4.3.1 CGPA

Aqui analisamos a relação da pontuação do TOEFL com a pontuação do CGPA.

Imagem 6 - TOEFL Score X CGPA, separados por Research

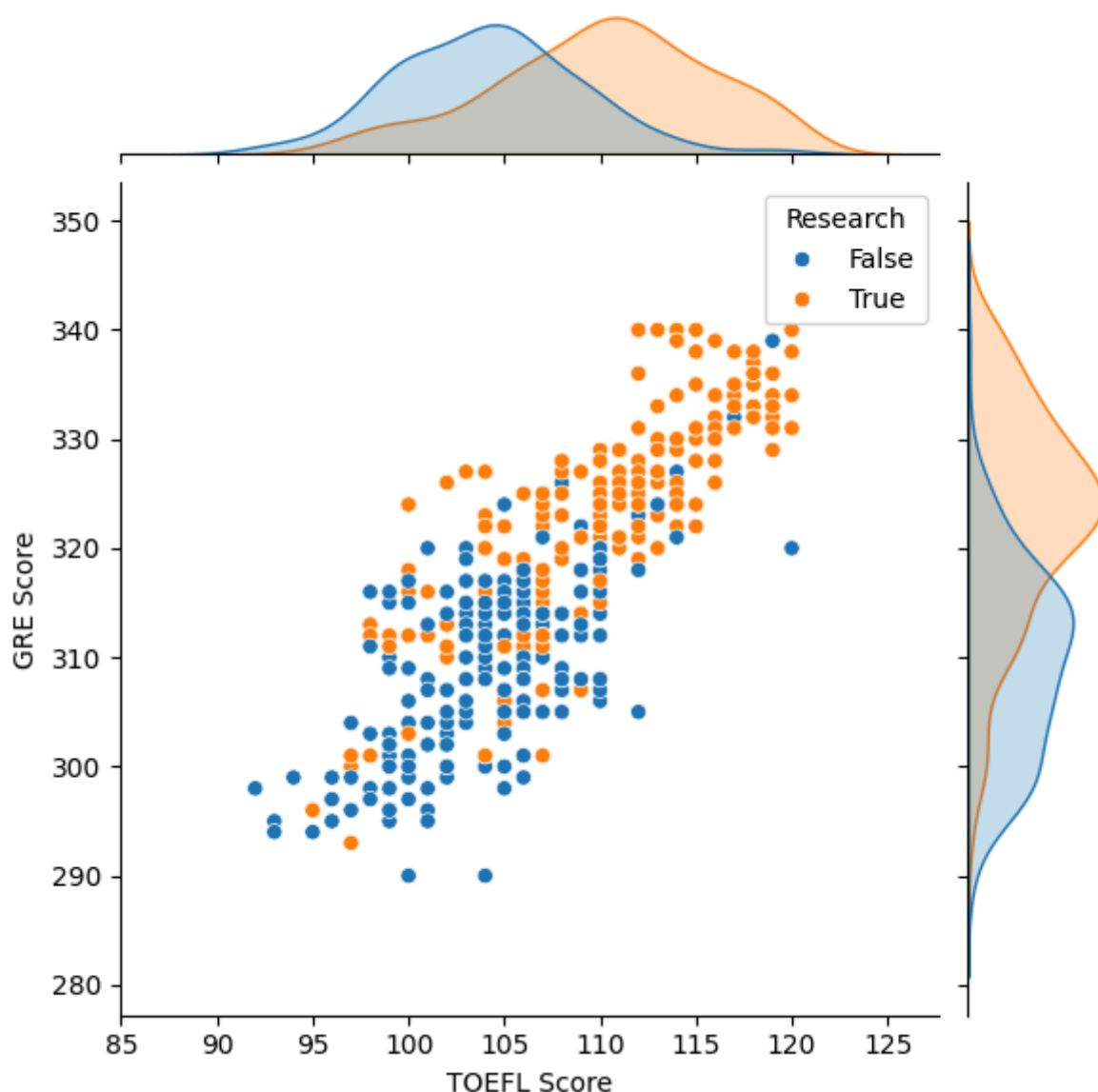


Fonte: elaborado pelos autores

4.3.2 GRE Score

Aqui analisamos a relação da pontuação do TOEFL com a pontuação do GRE.

Imagem 7 - TOEFL Score X GRE Score, separados por Research



Fonte: elaborado pelos autores

5 CONCLUSÃO

A análise detalhada da base de dados "Data for Admission in the University" permitiu identificar os principais fatores que influenciam a chance de um aluno ser admitido em um programa de mestrado. Através da divisão dos dados em categorias e da utilização de métodos estatísticos e gráficos, verificamos que as notas no GRE e TOEFL, o GPA, e a qualidade das cartas de recomendação e declarações de propósito são fatores cruciais para a admissão. Alunos com pontuações altas nesses critérios, especialmente acima da mediana, têm uma probabilidade significativamente maior de serem admitidos.

Concluimos que estudantes que desejam aumentar suas chances de admissão devem focar em melhorar suas pontuações no GRE e TOEFL e manter um alto GPA durante a graduação. Além disso, a participação em pesquisas e a obtenção de recomendações fortes podem compensar pontuações menores em outros critérios.

Este estudo pode servir de guia para futuros candidatos ao mestrado, ajudando-os a direcionar seus esforços de preparação para maximizar suas chances de sucesso.