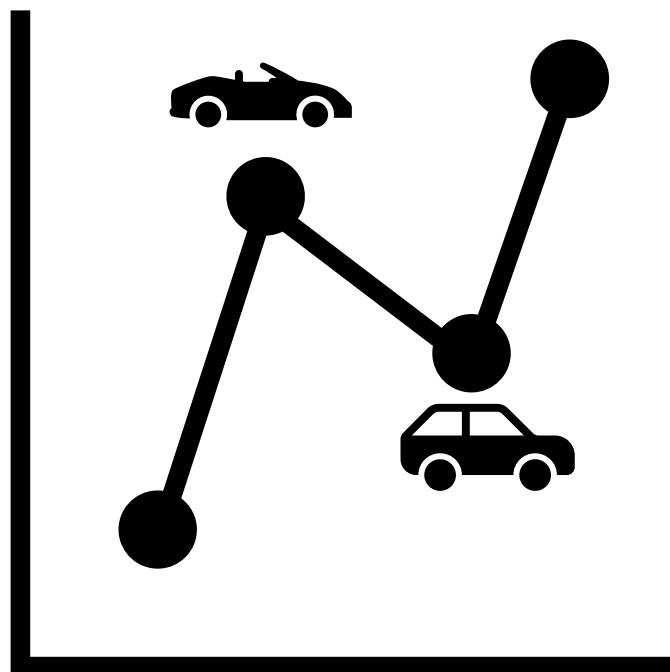


# Projeto TP1

Entropia, Informação Mútua e Codificação Huffman

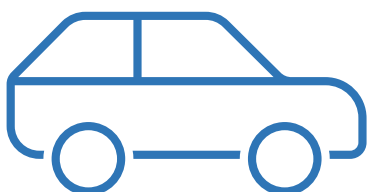


Feito por:

Diogo de Oliveira Mendes Dias Saldanha, nº 2022232761

Édi Baptista Marques, nº 2022229530

Tomás Pinto Loureiro, nº 2022234440



# Introdução

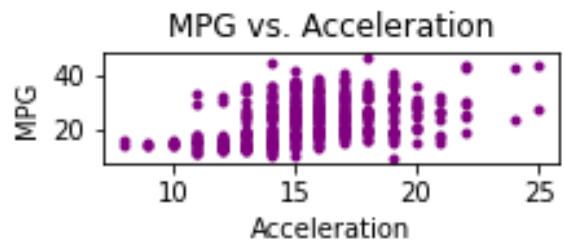
Neste trabalho serão desenvolvidas competências no tratamento e análise para um conjunto de dados, em particular entropia, informação mútua e codificação de Huffman.

Para tal, usar-se-á diversos dados de um conjunto de automóveis. E usando python irá-se escrever um programa no qual estes dados serão analisados.

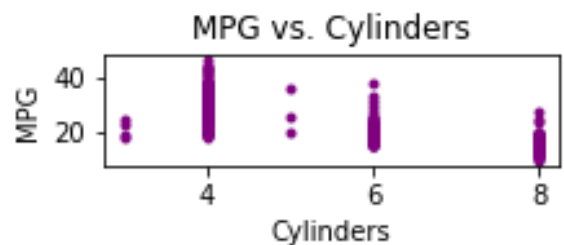
## Perguntas do projeto

2.

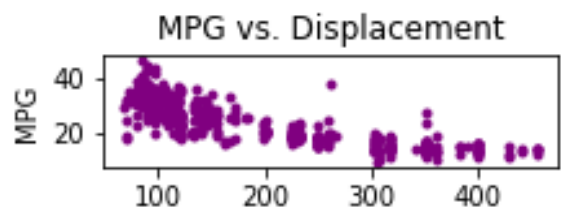
Na comparação com a aceleração vemos que quanto mais rápido o carro seja a acelerar menos eficiente é o carro.



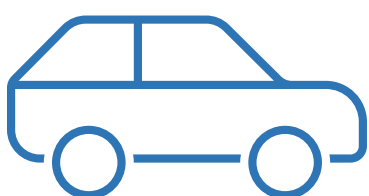
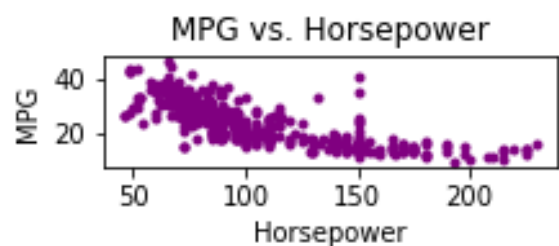
Na comparação com o número de cilindros vemos que quantos mais tiver menos eficiente o carro é.



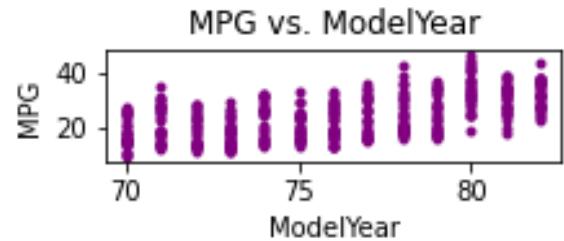
Na comparação com a cilindrada vemos que se esta aumentar, a eficiência diminui.



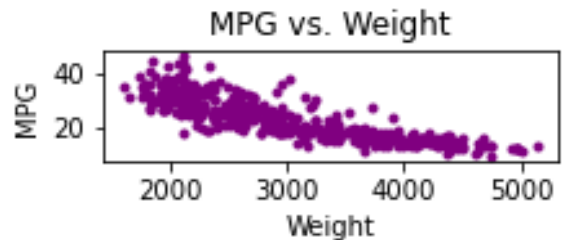
Na comparação com a potência vemos que se esta aumentar, a eficiência diminui.



Na comparação com o ano de fabrico podemos concluir que se o carro for mais novo ele é mais eficiente.



Na comparação com o peso do carro concluímos que se ele for mais leve, o carro vai ser mais eficiente.



-	Valor teórico	Valor Huffman
Aceleração	3.49642356	70.0
Cilindros	1.59043569	12.0
Cilindrada	4.84940337	528.0
Potência	4.49887396	616.0
Ano do modelo	3.69064251	49.0
Peso	6.06074283	3048.0
MPG	4.83579962	195.0
Entropia total	7.211576568035885	-

7.

Podemos concluir que o valor médio de bits por símbolo aumenta com a quantidade de símbolos diferentes que existem em cada uma das variáveis.

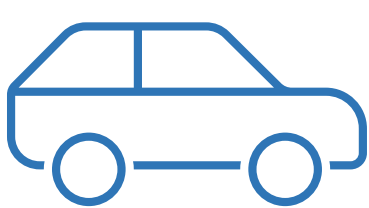
8.

b.

Pensando criticamente, os valores obtidos não devem de estar corretos. Então uma comparação direta entre os valores não iria ser imparcial.

c.

Pode-se reduzir a variância dos comprimentos fazendo agrupamento de símbolo. Isto serve para reduzir o tamanho dos dados a serem avaliados e então para conjuntos de dados grandes aumentará a velocidade e eficiência com a qual podemos analisar esses mesmos dados.



10.

-	Correlação	Informação mútua
Aceleração	1.0	0.8720358370364727
Cilindros	-0.5221868742854648	0.962178641086926
Cilindrada	-0.5574353032154352	2.646895257681372
Potência	-0.6689862275247186	2.56845025849273
Ano do modelo	0.2726213976648546	1.0294236623155246
Peso	-0.42413493232172106	4.610309603870352

Os resultados diferem, isto deve-se ao facto de que a informação mútua nos mostra dependência linear e não linear. Já a correlação só nos mostra dependência linear.

11.

b.

Os valores obtidos com a estimativa não coincidem com os valores reais.

e.

Olhando para os resultados obtidos nos pontos a e c conseguimos concluir que praticamente não mudam, já os resultados do ponto d mostram uma completa deviação do esperado. Vemos assim que se obterem boas estimativas devemos ter em atenção a informação mútua entre e o valor que queremos estimar e as variáveis que iremos usar para o estimar.

## Conclusão

Com a realização deste trabalho foi feita uma grande aprendizagem em como acontece análise de dados por meios de programação.

Para concluir basta dizer que foi um projeto desafiante que precisou do esforço de todos e do qual sairemos melhores engenheiros informáticos.

