

1 Aplicação no “Orange”.

Usando a aba Data arraste o “data set” selecione a da base de dados “imports 1985” por que para trabalhar com algoritmo de regressão tem que ser de uso numérico, depois arraste o ícone “data table” para o espaço de sua área de trabalho do “Orange” para visualizar os dados, em seguida crie uma conexão entre o “data set” e o “data table” para carregar os dados, e em seguida basta dar um duplo click e assim os dados ficam visíveis.

1.1 Quais são os tipos de dados que a base de dados “imports 1985” utiliza?

A base de dados imports 1985 inclui a base de dados e especificações de carros importados para o Estados Unidos no ano de 1985.

A partir desse ponto é possível usar esses dados para fazer as previsões dos preços dos carros.

“Data sets”

Price	Make	fuel-Type
13495	alfa-romero	gas
16500	alfa-romero	gas
16500	alfa-romero	gas
13950	audi	gas
17450	audi	gas
15250	audi	gas

Na tabela acima está sendo destacado três tipos de dados de um carro como preço, marca e tipo de combustível mais essas não são as únicas informações, essa base de dados possui um repertório bem amplo quando se trata de informações.

Para acessar mais informações sobre essa base de dados basta arrastar o ícone “data info” da aba data e criar uma conexão entre esse ícone selecionado e o “data sets”, depois basta dar um duplo click no “data info” que será mostrado as respectivas informações.

Data info
Data Sets name

Imports - 85 .
Data Sets Size
Roms 205.
Colunas 26.

Basicamente após gerar essas informações pode-se notar que o número de linhas é de 205 e o de colunas 26 ou seja 26 atributos de cada carro que é exibido no “Data Sets”.

1.2 Conexões entre o Linear Regression e o Test and Score.

Primeiramente arraste o ícone “Linear Regression” da aba “Model” até a sua área de trabalho e crie uma conexão com o “Data Sets” depois insira o “Test and Score” na sua área de trabalho e crie uma conexão entre ele e o “Linear regression” além disso também deve-se criar uma conexão entre o “Test and score” até o “Data Sets”. Aparecerá um aviso mostrando que existe valores nulos no Test and Score e que eles foram removidos dando um duplo click você verá que o algoritmo usou 80% dos dados para o treinamento e os 20 % restantes para teste, houve uma modificação nas métricas agora elas são baseadas em erros e uma delas é a M.A.E a média do erro absoluto que calcula o valor absoluto entre o valor estimado e o valor real que está na base de dados.

Exemplo: Um carro que custa 20.000 pode ter seu valor previsto de 21.950 ou então 18.050 para verificar a precisão do algoritmo você pode verificar o R² ou R ao quadrado como pode ver aqui o R² foi representado com 88% que significa uma correlação forte o resultado foi promissor ou seja um bom resultado.

1.3 Visualização das previsões.

O ícone que pode ser selecionado é o “Predictions” que se encontra na base “Evaluate” criando uma conexão entre o “Data Sets” e a base de dados “Linear Regression” talvez seja possível perceber um erro mais isso acontece devido aos dados não preenchidos no “Data Sets” dando um duplo click pode-se ver as previsões como mostra a tabela abaixo:

Linear Regression	Price
15205	13495
15205	16500
16086	16500

A tabela mostra um tipo de simulação no Orange que mostra ao lado esquerdo as previsões que o algoritmo fez e ao lado direito os preços reais dos carros ou seja que influencia na média do erro absoluto como foi citado anteriormente.