**08 - Queries com DataFrames**

Aquilo que chamamos de *queries* são, basicamente, consultas que podem ser realizadas em dataframes, de modo semelhante às seleções, mas utilizando técnicas mais sofisticadas. Continuaremos trabalhando com o dataset importado do arquivo db.csv nos vídeos anteriores.

Queremos realizar uma *query* que nos traga todos os registros para determinado tipo de motor, por exemplo o "Motor Diesel". Antes de fazermos isso, é interessante comentar que existe outra forma de obtermos uma *series* simples a partir do rótulo de um dataframe, que é a sintaxe dataframe.Rotulo, como no exemplo abaixo:

dataset.MotorCOPIAR CÓDIGO

Nome Jetta Variant Motor 4.0 Turbo Passat Motor Diesel Crossfox Motor Diesel V8 DS5 Motor 2.4 Turbo Aston Martin DB4 Motor 2.4 Turbo ...Phantom 2013 Motor V8 Cadillac Ciel concept Motor V8 Classe GLK Motor 5.0 V8 Bi-Turbo Aston Martin DB5 Motor Diesel Macan Motor Diesel V6 Name: Motor, Length: 258, dtype: object

Note que isso só funciona quando nossos rótulos não possuem espaços. Para selecionarmos todas as ocorrências de "Motor Diesel" nesse conjunto, usaremos o sinal de igualdade (==).

dataset.Motor == 'Motor Diesel'COPIAR CÓDIGO

Como retorno teremos uma *series* booleana na qual os itens que possuem o valor "Motor Diesel" receberão "True", e o restante "False".

Nome Jetta Variant False Passat True Crossfox False DS5 False Aston Martin DB4 False ...Phantom 2013 False Cadillac Ciel concept False Classe GLK False Aston Martin DB5 True Macan False Name: Motor, Length: 258, dtype: bool

Essa *series* pode ser passada como parâmetro em uma seleção de dataframe para obtermos somente os registros aos quais o valor "True" foi atribuído. Para facilitarmos esse tipo de consulta, armazenaremos nossa *series* em uma variável select.

select = dataset.Motor == 'Motor Diesel'COPIAR CÓDIGO

Com a função, type(), podemos verificar que estamos realmente trabalhando com um objeto do tipo *series*.

type(select)COPIAR CÓDIGO

pandas.core.series.Series

Ao passarmos a nova variável como parâmetro da nossa seleção, receberemos um dataframe contendo todos os veículso que possuem "Motor Diesel".

dataset[select]COPIAR CÓDIGO

| **Nome** | **Motor** | **Ano** | **Quilometragem** | **Zero\_km** | **Acessórios** | **Valor** |
| --- | --- | --- | --- | --- | --- | --- |
| Passat | Motor Diesel | 1991 | 5712.0 | False | ['Central multimídia', 'Teto panorâmico', 'Fre... | 106161.94 |
| Effa Hafei Picape Baú | Motor Diesel | 1991 | 102959.0 | False | ['Controle de estabilidade', 'Painel digital',... | 125684.65 |
| Sorento | Motor Diesel | 2019 | NaN | True | ['Sensor de chuva', 'Câmera de estacionamento'... | 81399.35 |
| New Fiesta Hatch | Motor Diesel | 2017 | 118895.0 | False | ['Sensor de estacionamento', 'Travas elétricas... | 66007.16 |
| Kangoo Express | Motor Diesel | 2007 | 29132.0 | False | ['Bancos de couro', 'Câmbio automático', 'Pilo... | 146716.91 |
| ... | ... | ... | ... | ... | ... | ... |

Vamos supor que queremos realizar uma *query* adicional, dessa vez verificando se esses veículos também são "Zero\_km". Para isso, a partir do nosso dataset, passaremos entre parênteses dois pares de parênteses separados por um & (*and*, caractere conhecido como "e comercial").

dataset[() & ()]COPIAR CÓDIGO

É dentro desses parênteses que passaremos as condições da nossa *query*: primeiro dataset.Motor == 'Motor Diesel', para recebermos somente os carros com esse tipo de motor; e então dataset.Zero\_km == True de modo a obtermos, dentre esses veículos, aqueles que também são "Zero\_km".

dataset[(dataset.Motor == 'Motor Diesel') & (dataset.Zero\_km == True)]COPIAR CÓDIGO

Como retorno teremos as cinco ocorrências dentro do dataset que atendem a essas condições.

| **Nome** | **Motor** | **Ano** | **Quilometragem** | **Zero\_km** | **Acessórios** | **Valor** |
| --- | --- | --- | --- | --- | --- | --- |
| Sorento | Motor Diesel | 2019 | NaN | True | ['Sensor de chuva', 'Câmera de estacionamento'... | 81399.35 |
| Cielo Hatch | Motor Diesel | 2019 | NaN | True | ['Painel digital', 'Central multimídia', 'Câme... | 145197.70 |
| Camry | Motor Diesel | 2019 | NaN | True | ['Travas elétricas', 'Rodas de liga', 'Sensor ... | 138597.27 |
| Aston Martin Virage | Motor Diesel | 2019 | NaN | True | ['Travas elétricas', 'Controle de tração', 'Câ... | 97290.18 |
| Série 7 Sedã | Motor Diesel | 2019 | NaN | True | ['Vidros elétricos', 'Travas elétricas', 'Roda... | 67539.79 |

A consulta (ou o "seletor") que realizamos também é, na verdade, uma *series* booleana que atribui os valores "True" e "False" para os elementos do nosso conjunto de acordo com as condições que estabelecemos.

(dataset.Motor == 'Motor Diesel') & (dataset.Zero\_km == True)COPIAR CÓDIGO

Nome Jetta Variant False Passat False Crossfox False DS5 False Aston Martin DB4 False ...Phantom 2013 False Cadillac Ciel concept False Classe GLK False Aston Martin DB5 False Macan False Length: 258, dtype: bool

Existe outra forma de fazermos *queries* nos dataframes do Pandas, que é por meio da função query(). Com ela também podemos passar uma seleção, mas seguindo uma sintaxe um pouco diferente, na qual nossas comparações serão cercadas por aspas simples ou duplas. Note que se for necessário passar strings na seleção, elas deverão ser sinalizada pelo caractere que não foi utilizado na seleção superior - ou seja, se nossa seleção está cercada de aspas simples, a string deverá ser passada dentro de aspas duplas e vice-versa.

dataset.query('Motor == "Motor Diesel" and Zero\_km == True')COPIAR CÓDIGO

Nosso retorno será exatamente o mesmo que recebemos anteriormente.

| **Nome** | **Motor** | **Ano** | **Quilometragem** | **Zero\_km** | **Acessórios** | **Valor** |
| --- | --- | --- | --- | --- | --- | --- |
| Sorento | Motor Diesel | 2019 | NaN | True | ['Sensor de chuva', 'Câmera de estacionamento'... | 81399.35 |
| Cielo Hatch | Motor Diesel | 2019 | NaN | True | ['Painel digital', 'Central multimídia', 'Câme... | 145197.70 |
| Camry | Motor Diesel | 2019 | NaN | True | ['Travas elétricas', 'Rodas de liga', 'Sensor ... | 138597.27 |
| Aston Martin Virage | Motor Diesel | 2019 | NaN | True | ['Travas elétricas', 'Controle de tração', 'Câ... | 97290.18 |
| Série 7 Sedã | Motor Diesel | 2019 | NaN | True | ['Vidros elétricos', 'Travas elétricas', 'Roda... | 67539.79 |

Além do "and" (&), o operador "or" (|, "ou") também pode nos auxiliar nas nossas seleções.

dataset[(dataset.Motor == 'Motor Diesel') | (dataset.Zero\_km == True)]COPIAR CÓDIGO

dataset.query('Motor == "Motor Diesel" or Zero\_km == True')COPIAR CÓDIGO

No próximo vídeo aprenderemos sobre iterações em um dataframe.