**05 - Seleções com DataFrames**

Agora que temos um dataframe, queremos aprender a acessar informações dentro dele. Durante esse vídeo e os próximos, continuaremos a usar o dataset que importamos anteriormente, portanto é recomendado mantê-lo na memória (ou recarregá-lo caso seja necessário).

Começaremos pela função head(), que mostra os cinco primeiros registros do nosso conjunto de dados.

dataset.head()COPIAR CÓDIGO

| **Nome** | **Motor** | **Ano** | **Quilometragem** | **Zero\_km** | **Acessórios** | **Valor** |
| --- | --- | --- | --- | --- | --- | --- |
| Jetta Variant | Motor 4.0 Turbo | 2003 | 44410.0 | False | ['Rodas de liga', 'Travas elétricas', 'Piloto ... | 88078.64 |
| Passat | Motor Diesel | 1991 | 5712.0 | False | ['Central multimídia', 'Teto panorâmico', 'Fre... | 106161.94 |
| Crossfox | Motor Diesel V8 | 1990 | 37123.0 | False | ['Piloto automático', 'Controle de estabilidad... | 72832.16 |
| DS5 | Motor 2.4 Turbo | 2019 | NaN | True | ['Travas elétricas', '4 X 4', 'Vidros elétrico... | 124549.07 |
| Aston Martin DB4 | Motor 2.4 Turbo | 2006 | 25757.0 | False | ['Rodas de liga', '4 X 4', 'Central multimídia... | 92612.10 |

Se quisermos selecionar somente a coluna "Valor", por exemplo, bastará passarmos o seu rótulo entre colchetes.

dataset['Valor']COPIAR CÓDIGO

Nome Jetta Variant 88078.64 Passat 106161.94 Crossfox 72832.16 DS5 124549.07 Aston Martin DB4 92612.10 ...Phantom 2013 51759.58 Cadillac Ciel concept 51667.06 Classe GLK 68934.03 Aston Martin DB5 122110.90 Macan 90381.47 Name: Valor, Length: 258, dtype: float64

Note que também recebemos os nomes dos veículos. Isso acontece pois dataset['Valor'] nos retorna, além da coluna desejada, os índices, e anteriormente definimos que o rótulo "Nome" deveria ser usado como índice. Também deve-se prestar atenção ao fato de que esse dado é uma *series* do Pandas, algo que podemos conferir com a função type().

type(dataset['Valor'])COPIAR CÓDIGO

pandas.core.series.Series

Ou seja, um dataframe não deixa de ser um conjunto de *series*. Se quisermos, também é possível selecionar uma única coluna e obter como retorno um dataframe. Para isso, ao invés de passarmos um único par de colchetes, passaremos dois.

dataset[['Valor']COPIAR CÓDIGO

| **Nome** | **Valor** |
| --- | --- |
| Jetta Variant | 88078.64 |
| Passat | 106161.94 |
| Crossfox | 72832.16 |
| DS5 | 124549.07 |
| Aston Martin DB4 | 92612.10 |
| ... | ... |
| Phantom 2013 | 51759.58 |
| Cadillac Ciel concept | 51667.06 |
| Classe GLK | 68934.03 |
| Aston Martin DB5 | 122110.90 |
| Macan | 90381.47 |

A seleção de linhas mais simples, que já conhecemos das listas, tuplas e arrays Numpy, funciona basicamente da mesma forma.

Observação: A indexação tem origem no zero e nos fatiamentos (slices) a linha com índice i é incluída e a linha com índice j não é incluída no resultado. Em nosso exemplo, pegaremos as três primeiras linhas ([0:3]) do nosso dataset.

dataset[0:3]COPIAR CÓDIGO

| **Nome** | **Motor** | **Ano** | **Quilometragem** | **Zero\_km** | **Acessórios** | **Valor** |
| --- | --- | --- | --- | --- | --- | --- |
| Jetta Variant | Motor 4.0 Turbo | 2003 | 44410.0 | False | ['Rodas de liga', 'Travas elétricas', 'Piloto ... | 88078.64 |
| Passat | Motor Diesel | 1991 | 5712.0 | False | ['Central multimídia', 'Teto panorâmico', 'Fre... | 106161.94 |
| Crossfox | Motor Diesel V8 | 1990 | 37123.0 | False | ['Piloto automático', 'Controle de estabilidad... | 72832.16 |

Note que, como afirmado na observação, o primeiro índice é incluído e o último não. Existem outras formas mais elegantes de fazermos seleções, conhecidas como .loc e .iloc.

O .loc seleciona um grupo de linahs e colunas segundo os rótulos ou uma matriz booleana. Comentaremos sobre esse último aspecto futuramente, quando estivermos trabalhando com *queries* em dataframes. Para selecionarmos um rótulo, simplesmente o passaremos entre colchetes. Nesse exemplo, selecionaremos o "Passat".

dataset.loc['Passat']COPIAR CÓDIGO

Como retorno teremos uma *series* com as informações deste veículo (ou seja, aquela linha).

Motor Motor Diesel Ano 1991 Quilometragem 5712 Zero\_km False Acessórios ['Central multimídia', 'Teto panorâmico', 'Fre... Valor 106162 Name: Passat, dtype: object

Se quisermos mais de uma informação, precisaremos passá-las dentro de dois pares de colchetes. Usaremos como exemplo os veículos "Passat" e "DS5".

dataset.loc[['Passat', 'DS5']]COPIAR CÓDIGO

Como retorno teremos um dataframe com as linhas que selecionamos a partir dos seus rótulos.

| **Nome** | **Motor** | **Ano** | **Quilometragem** | **Zero\_km** | **Acessórios** | **Valor** |
| --- | --- | --- | --- | --- | --- | --- |
| Passat | Motor Diesel | 1991 | 5712.0 | False | ['Central multimídia', 'Teto panorâmico', 'Fre... | 106161.94 |
| DS5 | Motor 2.4 Turbo | 2019 | NaN | True | ['Travas elétricas', '4 X 4', 'Vidros elétrico... | 124549.07 |

O .loc também nos permite selecionar quais colunas desejamos exibir. Para isso, depois das informações das linhas, adicionaremos uma vírgula e um novo par de colchetes, dessa vez com as informações "Motor" e "Valor" representando as colunas a serem exibidas.

dataset.loc[['Passat', 'DS5'], ['Motor', 'Valor']]COPIAR CÓDIGO

| **Nome** | **Motor** | **Valor** |
| --- | --- | --- |
| Passat | Motor Diesel | 106161.94 |
| DS5 | Motor 2.4 Turbo | 124549.07 |

Já se passarmos apenas : como primeiro parâmetro, receberemos todas as linhas do conjunto e somente as colunas selecionadas.

dataset.loc[:, ['Motor', 'Valor']]COPIAR CÓDIGO

| **Nome** | **Motor** | **Valor** |
| --- | --- | --- |
| Jetta Variant | Motor 4.0 Turbo | 88078.64 |
| Passat | Motor Diesel | 106161.94 |
| Crossfox | Motor Diesel V8 | 72832.16 |
| DS5 | Motor 2.4 Turbo | 124549.07 |
| Aston Martin DB4 | Motor 2.4 Turbo | 92612.10 |
| ... | ... | ... |
| Phantom 2013 | Motor V8 | 51759.58 |
| Cadillac Ciel concept | Motor V8 | 51667.06 |
| Classe GLK | Motor 5.0 V8 Bi-Turbo | 68934.03 |
| Aston Martin DB5 | Motor Diesel | 122110.90 |
| Macan | Motor Diesel V6 | 90381.47 |

O iloc também nos permite fazer seleções, mas se utiliza dos índices numéricos - ou seja, na posição das informações. No primeiro exemplo, passaremos como parâmetro apenas 1.

dataset.iloc[1]COPIAR CÓDIGO

Motor Motor Diesel Ano 1991 Quilometragem 5712 Zero\_km False Acessórios ['Central multimídia', 'Teto panorâmico', 'Fre... Valor 106162 Name: Passat, dtype: object

Note que nos foram devolvidas as informações do carro de índice 1, que é o "Passat". Utilizando dois pares de colchetes ao invés de um, receberemos essas informações no formato de um dataframe.

dataset.iloc[[1]]COPIAR CÓDIGO

| **Nome** | **Motor** | **Ano** | **Quilometragem** | **Zero\_km** | **Acessórios** | **Valor** |
| --- | --- | --- | --- | --- | --- | --- |
| Passat | Motor Diesel | 1991 | 5712.0 | False | ['Central multimídia', 'Teto panorâmico', 'Fre... | 106161.94 |

Quando queremos informações de múltiplas linhas em um dataset, utilizaremos somente um par de colchetes, dentro dos quais passaremos o índice inicial e o final (lembrando que o inicial é inclusivo e o final não).

dataset.iloc[1:4]COPIAR CÓDIGO

| **Nome** | **Motor** | **Ano** | **Quilometragem** | **Zero\_km** | **Acessórios** | **Valor** |
| --- | --- | --- | --- | --- | --- | --- |
| Passat | Motor Diesel | 1991 | 5712.0 | False | ['Central multimídia', 'Teto panorâmico', 'Fre... | 106161.94 |
| Crossfox | Motor Diesel V8 | 1990 | 37123.0 | False | ['Piloto automático', 'Controle de estabilidad... | 72832.16 |
| DS5 | Motor 2.4 Turbo | 2019 | NaN | True | ['Travas elétricas', '4 X 4', 'Vidros elétrico... | 124549.07 |

Para obtermos colunas específicas, passaremos, após uma vírgula, um novo par de colchetes contendo os índices das colunas que desejamos acessar. Não é necessário passá-los na mesma ordem do dataset original, algo que exemplificaremos na seleção abaixo, feita com as colunas 0, 5 e 2 ("Motor", "Valor" e "Quilometragem", respectivamente).

dataset.iloc[1:4, [0, 5, 2]]COPIAR CÓDIGO

Com isso receberemos as informações na ordem que especificamos.

| **Nome** | **Motor** | **Valor** | **Quilometragem** |
| --- | --- | --- | --- |
| Passat | Motor Diesel | 106161.94 | 5712.0 |
| Crossfox | Motor Diesel V8 | 72832.16 | 37123.0 |
| DS5 | Motor 2.4 Turbo | 124549.07 | NaN |

Da mesma forma, é possível selecionar linhas específicas que não estejam em sequência, bastando passarmos seus índices dentro de um par de colchetes.

dataset.iloc[[1, 42, 22], [0, 5, 2]]COPIAR CÓDIGO

| **Nome** | **Motor** | **Valor** | **Quilometragem** |
| --- | --- | --- | --- |
| Passat | Motor Diesel | 106161.94 | 5712.0 |
| Optima | Motor 1.8 16v | 86641.34 | NaN |
| Lamborghini Obvious | Motor Diesel V6 | 133529.84 | 98079.0 |

Se quisermos exibir todas as linhas, usaremos os dois pontos (:).

dataset.iloc[:, [0, 5, 2]]COPIAR CÓDIGO

| **Nome** | **Motor** | **Valor** | **Quilometragem** |
| --- | --- | --- | --- |
| Jetta Variant | Motor 4.0 Turbo | 88078.64 | 44410.0 |
| Passat | Motor Diesel | 106161.94 | 5712.0 |
| Crossfox | Motor Diesel V8 | 72832.16 | 37123.0 |
| DS5 | Motor 2.4 Turbo | 124549.07 | NaN |
| Aston Martin DB4 | Motor 2.4 Turbo | 92612.10 | 25757.0 |
| ... | ... | ... | ... |
| Phantom 2013 | Motor V8 | 51759.58 | 27505.0 |
| Cadillac Ciel concept | Motor V8 | 51667.06 | 29981.0 |
| Classe GLK | Motor 5.0 V8 Bi-Turbo | 68934.03 | 52637.0 |
| Aston Martin DB5 | Motor Diesel | 122110.90 | 7685.0 |
| Macan | Motor Diesel V6 | 90381.47 | 50188.0 |

O processo se repete caso queiramos exibir todas as colunas. Com isso, aprendemos o básico sobre as seleções em dataframes do Pandas. No próximo vídeo conheceremos as *queries*.