How do I select a subset of a DataFrame

## How do I select specific columns from a DataFrame?



This method is use to import data of csv file of excel file also. If we are using colab at google then first we upload the file the take there path and the use specific file/

Note at the loacal machine we use path by forward slash and use colab the use backward slash for the pathe of file upload.

import pandas as pd

data=pd.read\_csv(r"C:\Users\PC\Desktop\train.csv") # r is use to roughly read the data

data

for example

ages = titanic["Age"]

To select a single column, use square brackets [] with the column name of the column of interest.

Each column in a **[DataFrame](https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.html" \l "pandas.DataFrame" \o "pandas.DataFrame)** is a [**Series**](https://pandas.pydata.org/docs/reference/api/pandas.Series.html#pandas.Series). As a single column is selected, the returned object is a pandas [**Series**](https://pandas.pydata.org/docs/reference/api/pandas.Series.html#pandas.Series). We can verify this by checking the type of the output:

type(titanic["Age"])

Out[6]: pandas.core.series.Series

And have a look at the shape of the output:

titanic["Age"].shape

Out[7]: (891,)

Note: [**DataFrame.shape**](https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.shape.html#pandas.DataFrame.shape) is an attribute of a pandas Series and DataFrame containing the number of rows and columns: (nrows, ncolumns). A pandas Series is 1-dimensional and only the number of rows is returned.

I’m interested in the age and sex of the Titanic passengers.

age\_sex = titanic[["Age", "Sex"]]

To select multiple columns, use a list of column names within the selection brackets [].

If we want to show only first five records then use head

age\_sex.head()

The returned data type is a pandas DataFrame:

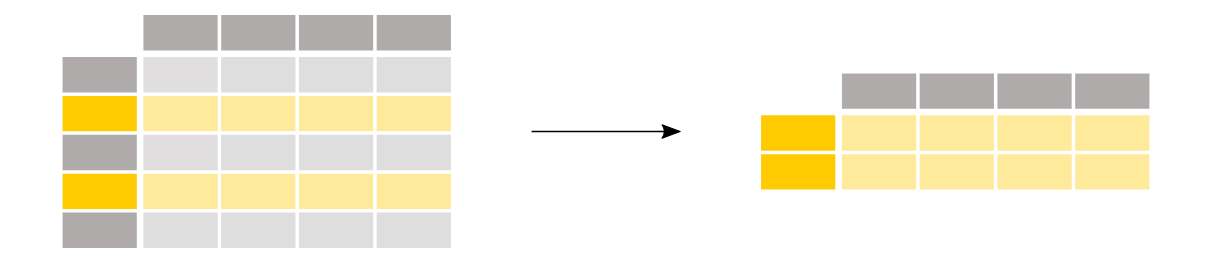
type(titanic[["Age", "Sex"]])

if we want to returned the column of database the use shape command

titanic[["Age", "Sex"]].shape

above\_35.head()

## How do I filter specific rows from a DataFrame?



In these example titanic is database where we call the file uploaded

above\_35 = titanic[titanic["Age"] > 35]

To select rows based on a conditional expression, use a condition inside the selection brackets [].

The condition inside the selection brackets titanic["Age"] > 35 checks for which rows the Age column has a value larger than 35:

titanic["Age"] > 35

it display only true and false values

The output of the conditional expression (>, but also ==, !=, <, <=,… would work) is actually a pandas Series of boolean values (either True or False) with the same number of rows as the original DataFrame. Such a Series of boolean values can be used to filter the DataFrame by putting it in between the selection brackets []. Only rows for which the value is True will be selected.

We know from before that the original Titanic DataFrame consists of 891 rows. Let’s have a look at the number of rows which satisfy the condition by checking the shape attribute of the resulting DataFrame above\_35:

#### REMEMBER

* When selecting subsets of data, square brackets [] are used.
* Inside these brackets, you can use a single column/row label, a list of column/row labels, a slice of labels, a conditional expression or a colon.
* Select specific rows and/or columns using loc when using the row and column names.
* Select specific rows and/or columns using iloc when using the positions in the table.
* You can assign new values to a selection based on loc/iloc.