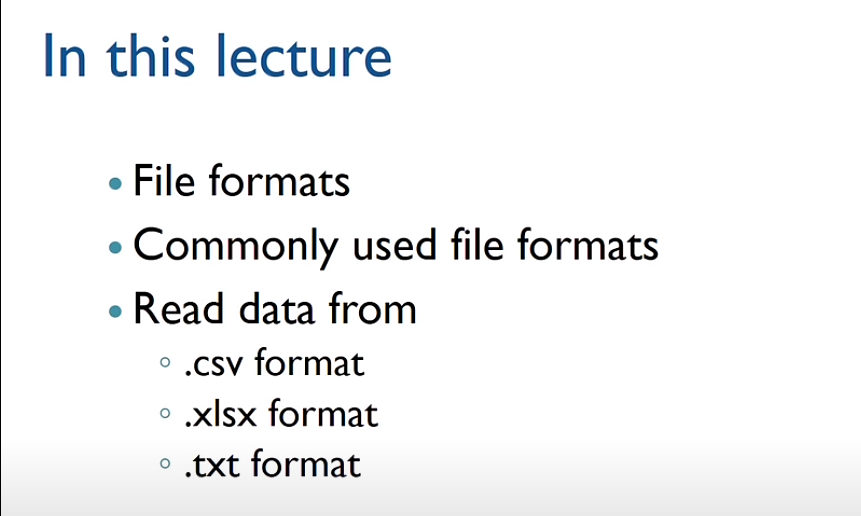
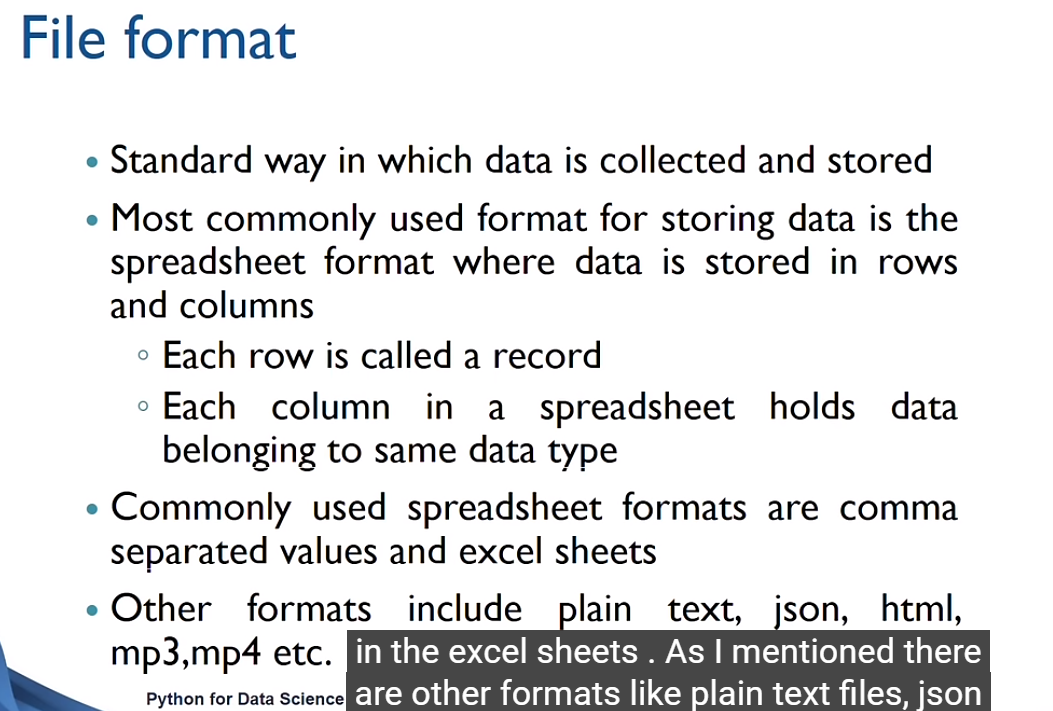
# Reading Data

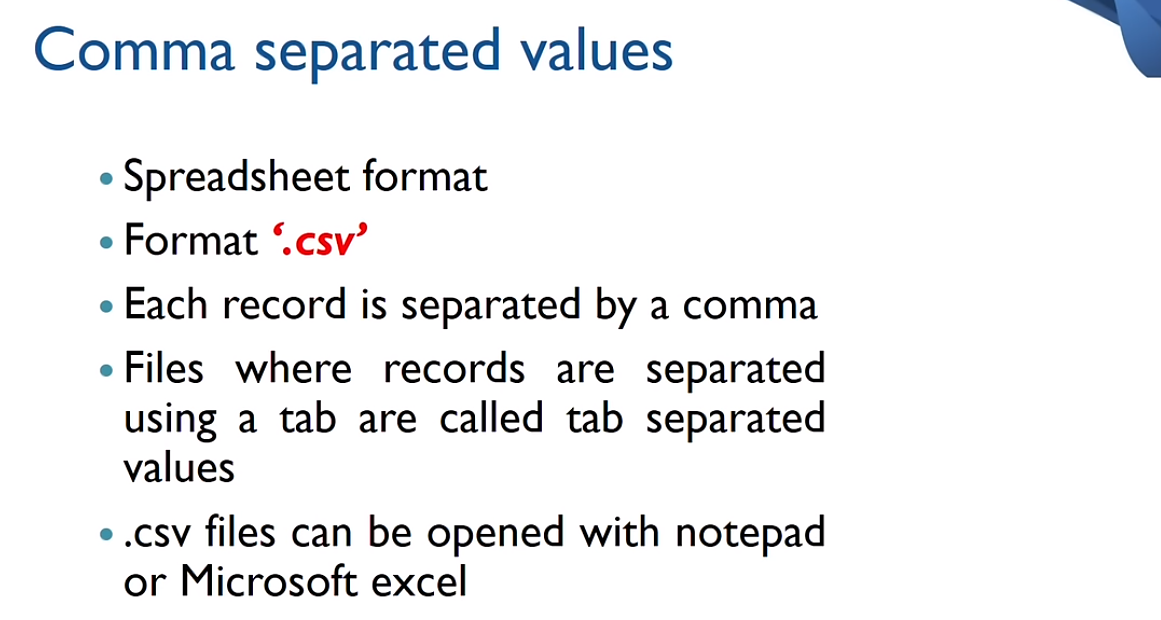
## Agenda



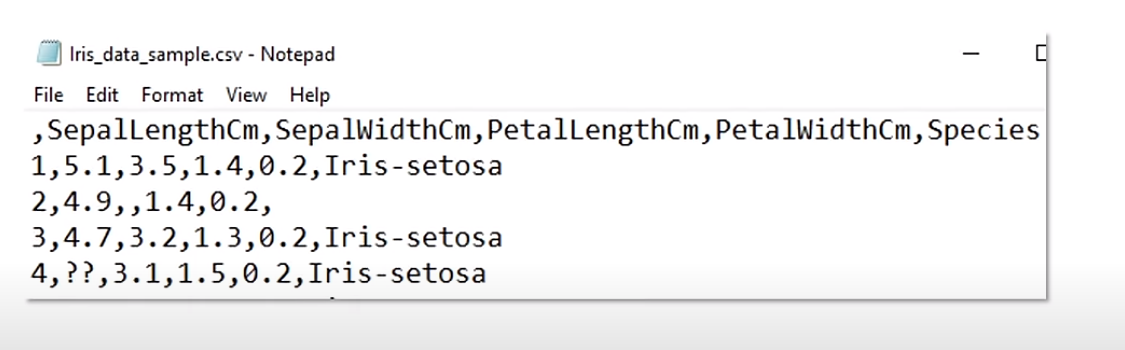
## File Format



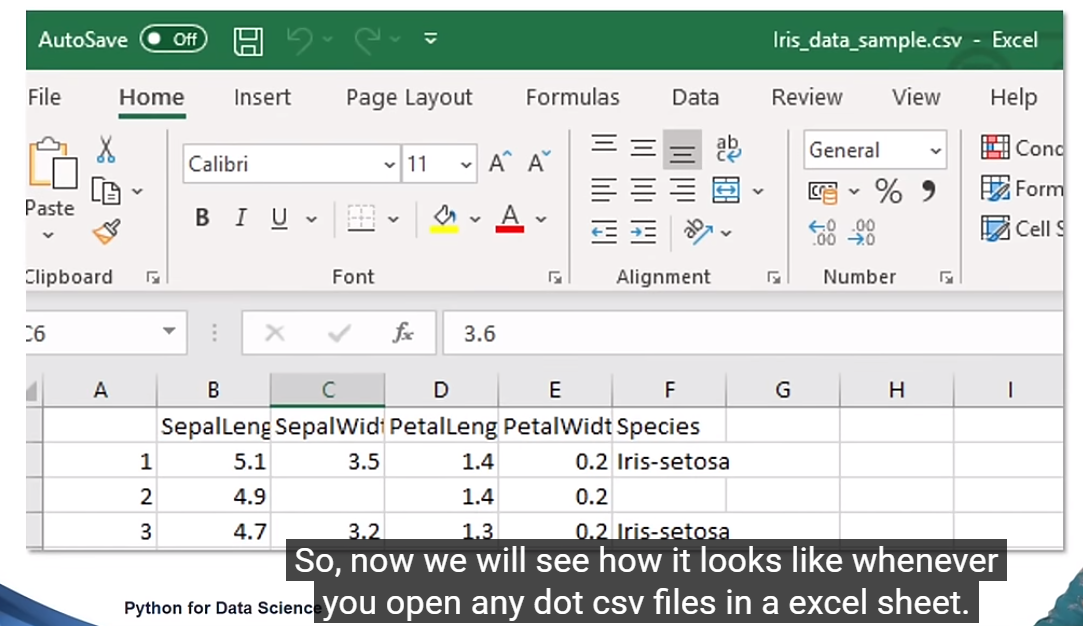
### Comma separated values (csv) format



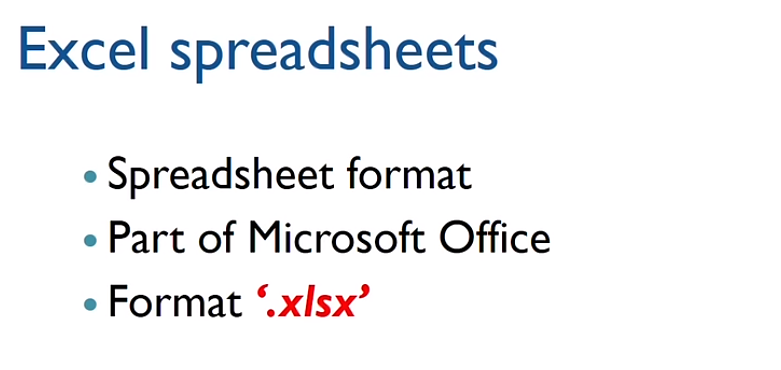
Comma separated values in a notepad. It is hard to read which value belongs to which column(variable).

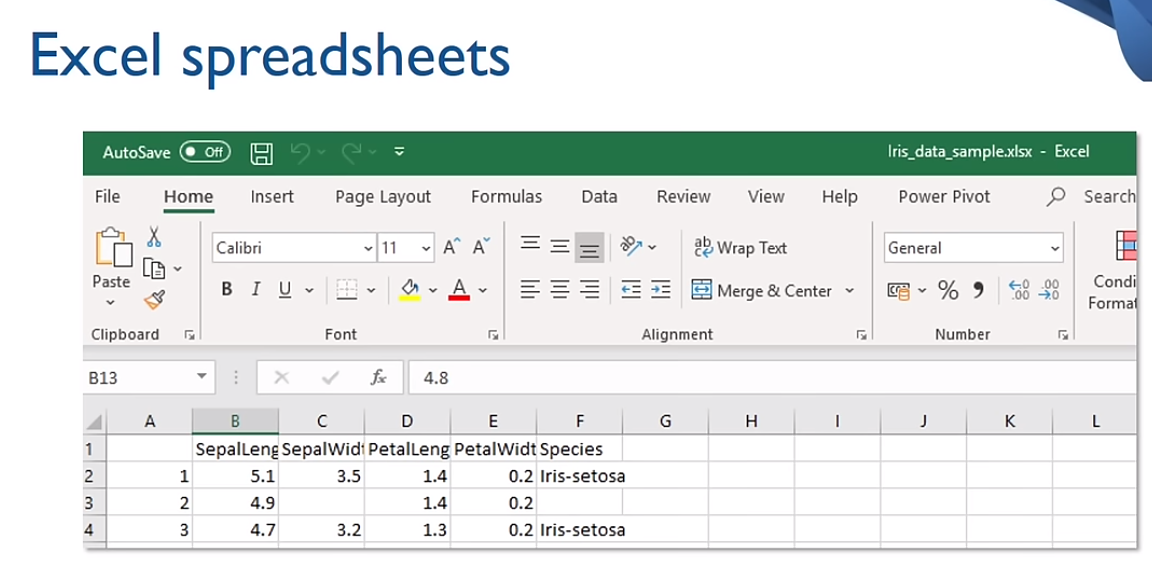


Comma separated values in an excel sheet. It is easy to read.

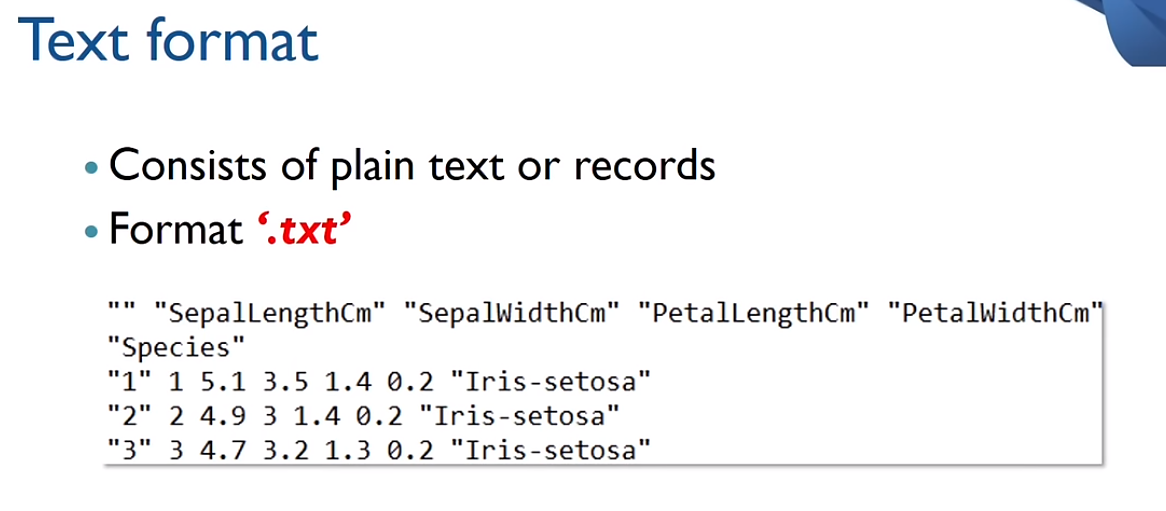


### Excel (xlsx) format

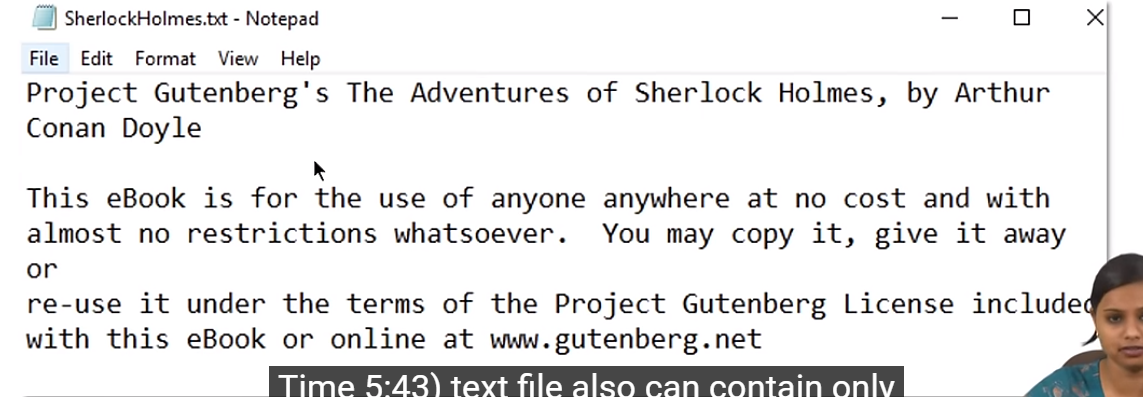




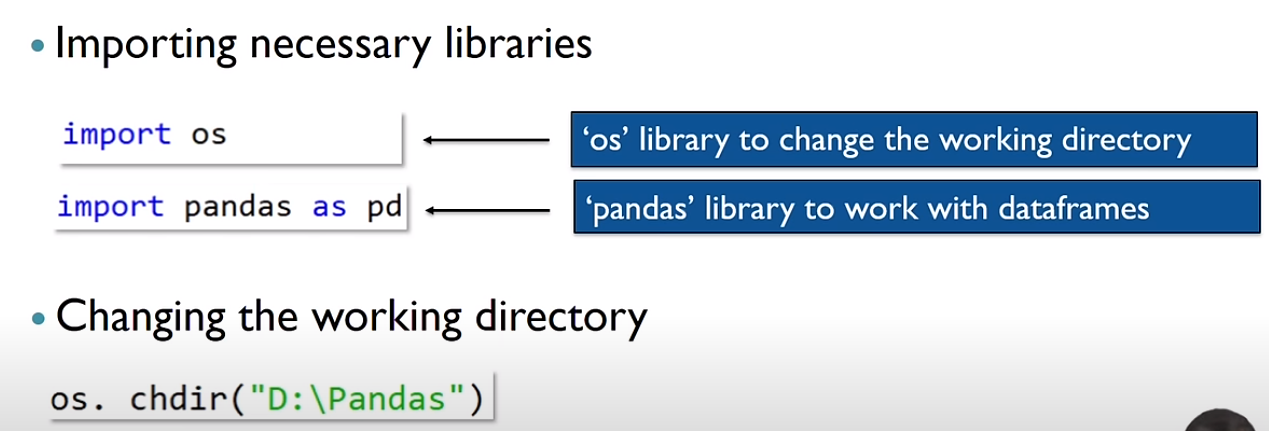
### Text (txt) format



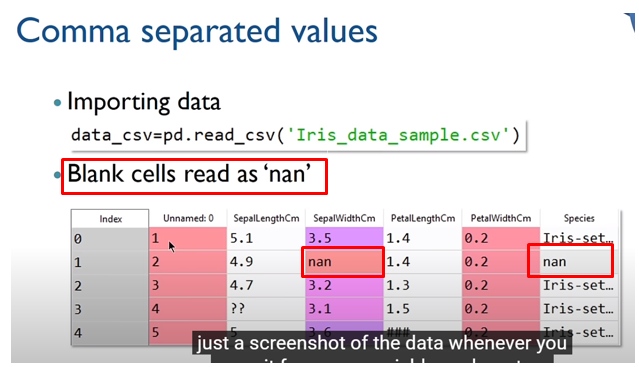
It can contain the data purely in the text format.



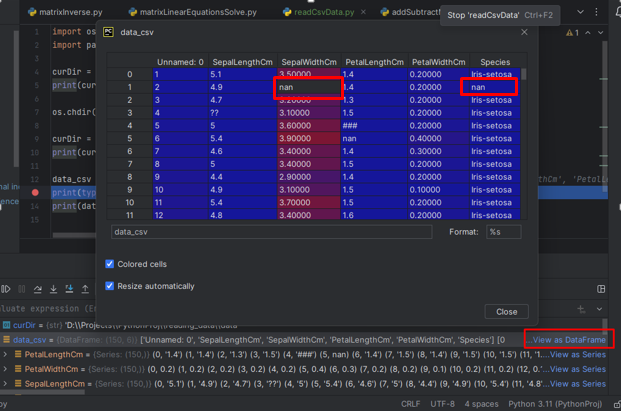
## Importing Data

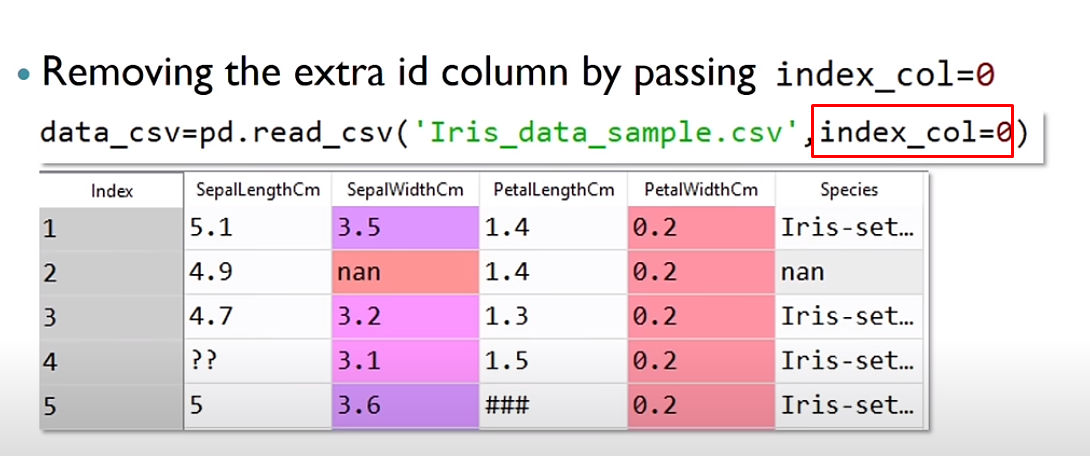


Use Panda’s dataframes to import the data from the file.



In Pycharm, start in debug mode

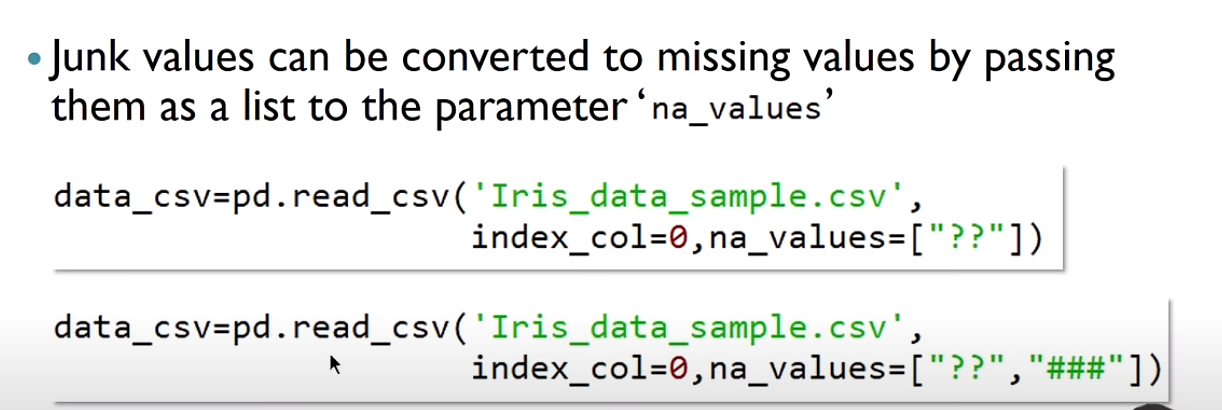


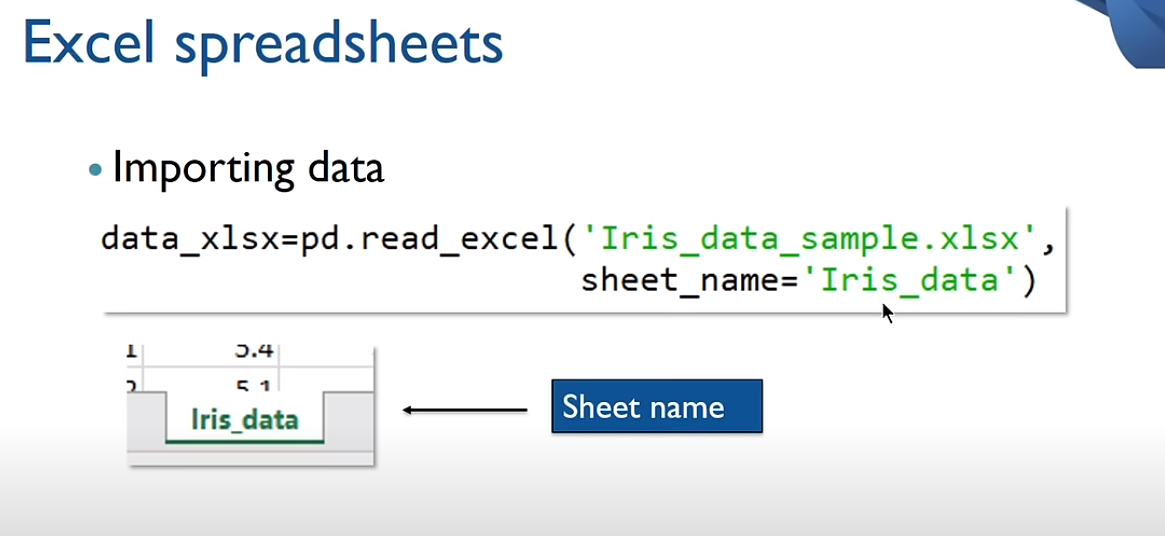


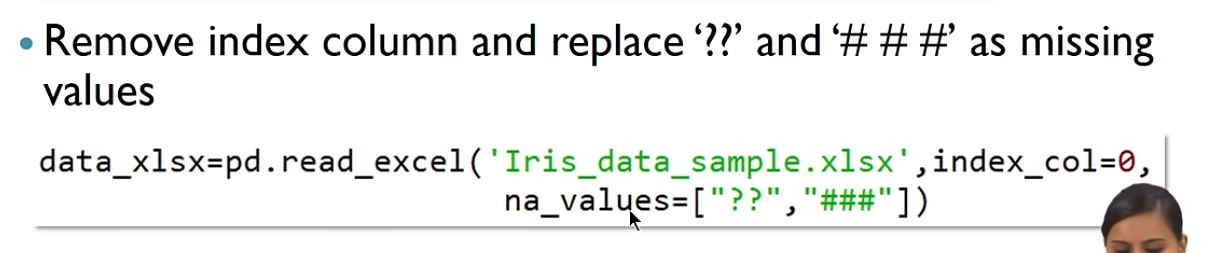
0th column of csv file will be considered as an index col, so Pandas won't add its own index column

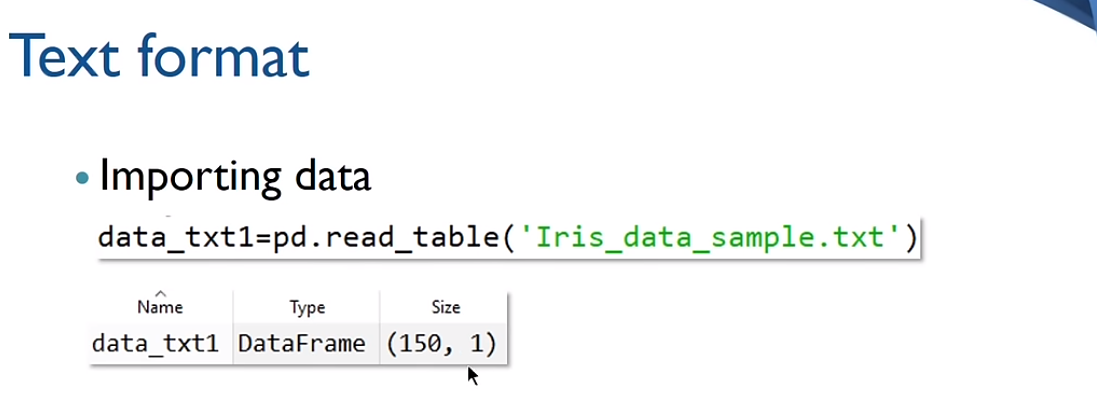


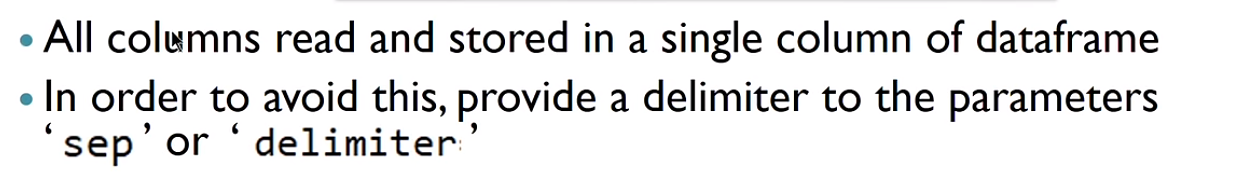
Let’s say, you got the data in such a way that ?? and ### are represented as missing values. Pandas will replace only blanks as NaN. You can replace all ?? and ### with NAN.





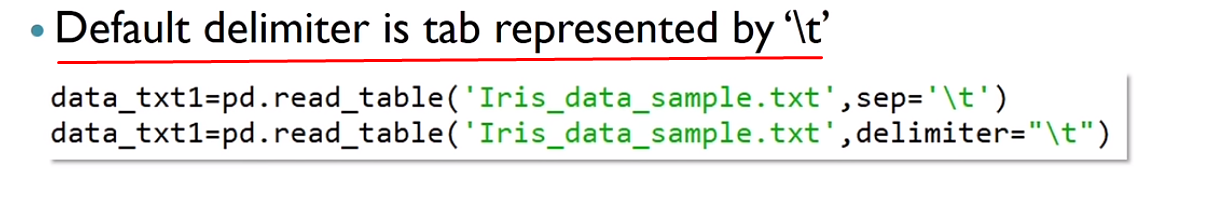


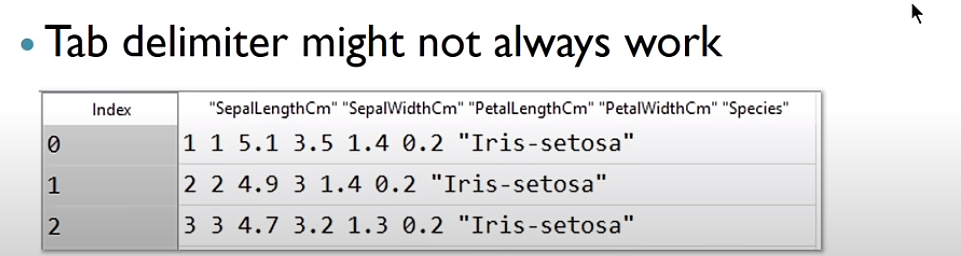


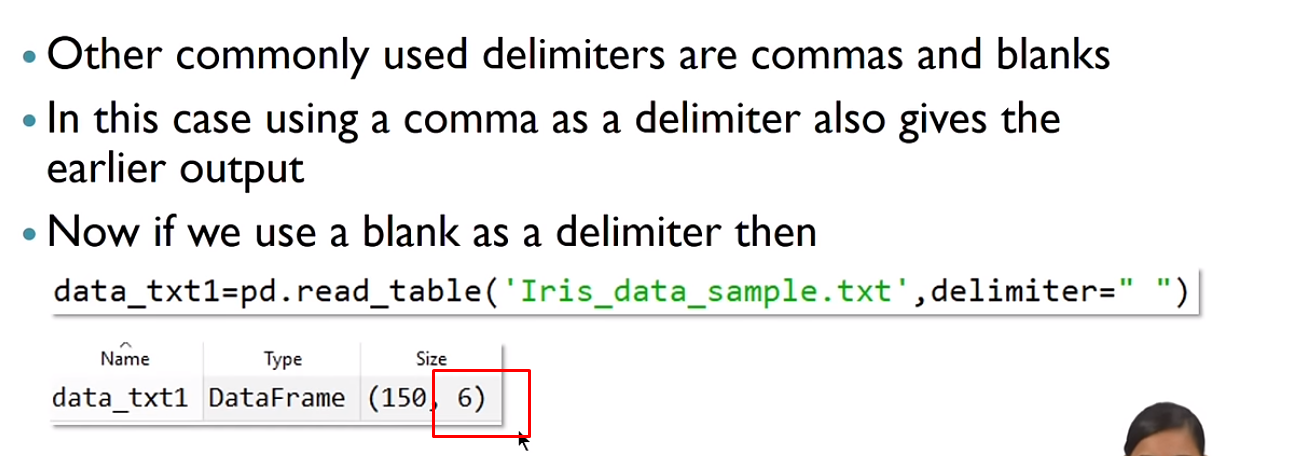


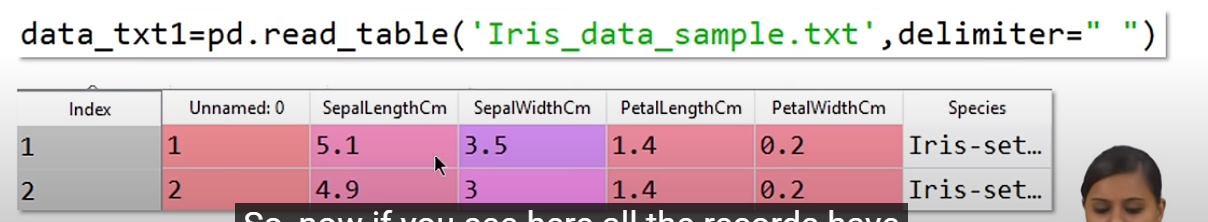
Sometimes, when you read data from text file, it reads all columns as one column. So, it creates only one column in dataframes.

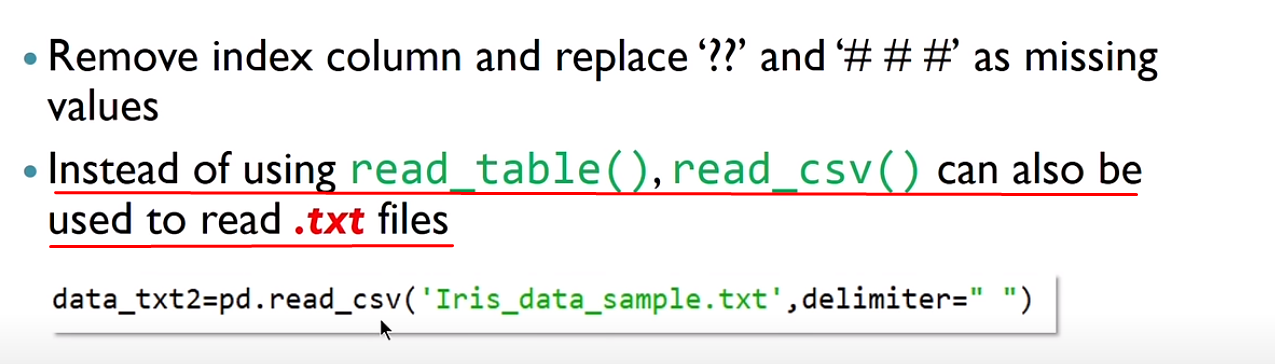
To overcome this problem, you can either use a 'sep' or 'delimiter' parameter. Normal delimiters are tab, blank, comma etc.







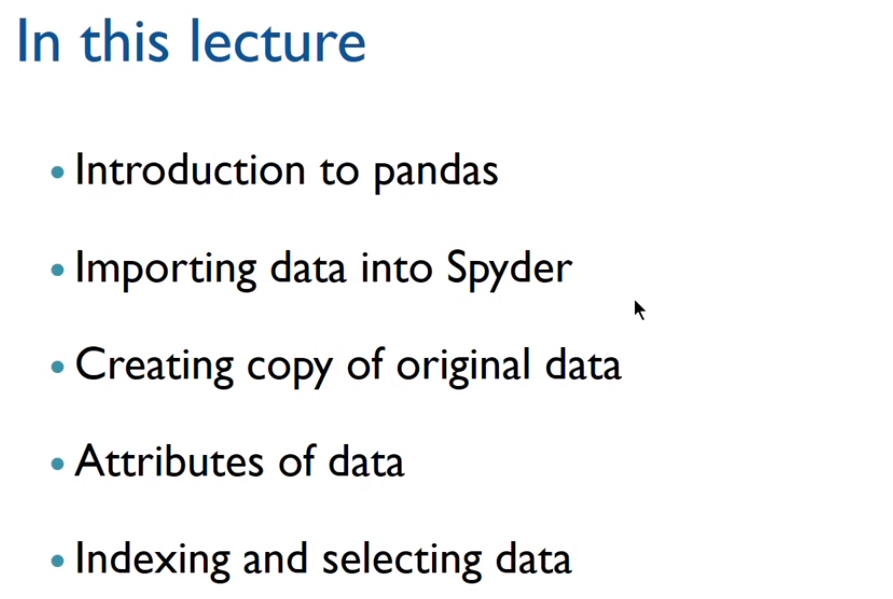




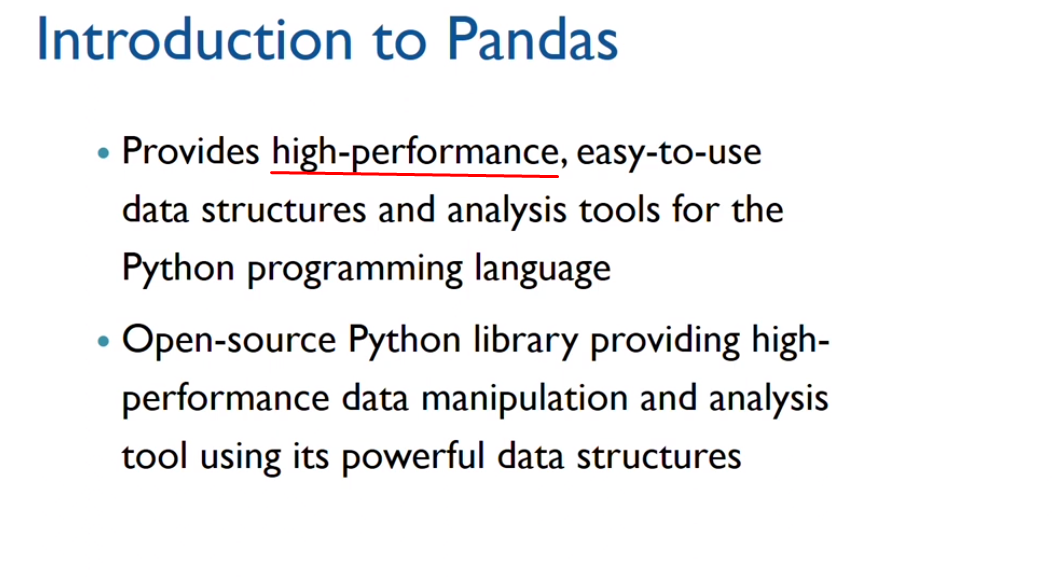
If you use, read\_csv() instead of read\_table() to read a text file, it is mandatory to give a delimiter.

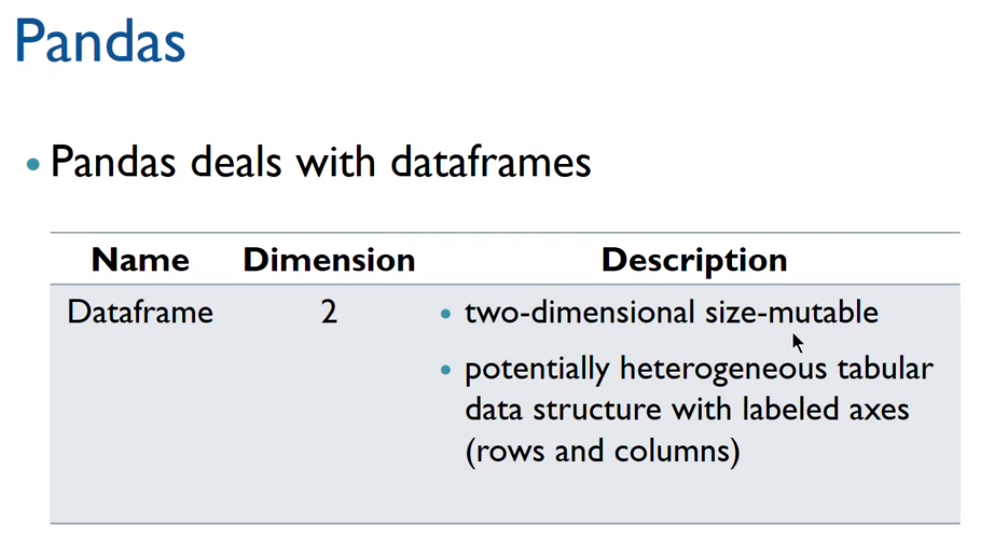
# Pandas Dataframes

## Agenda



## Introduction to Pandas





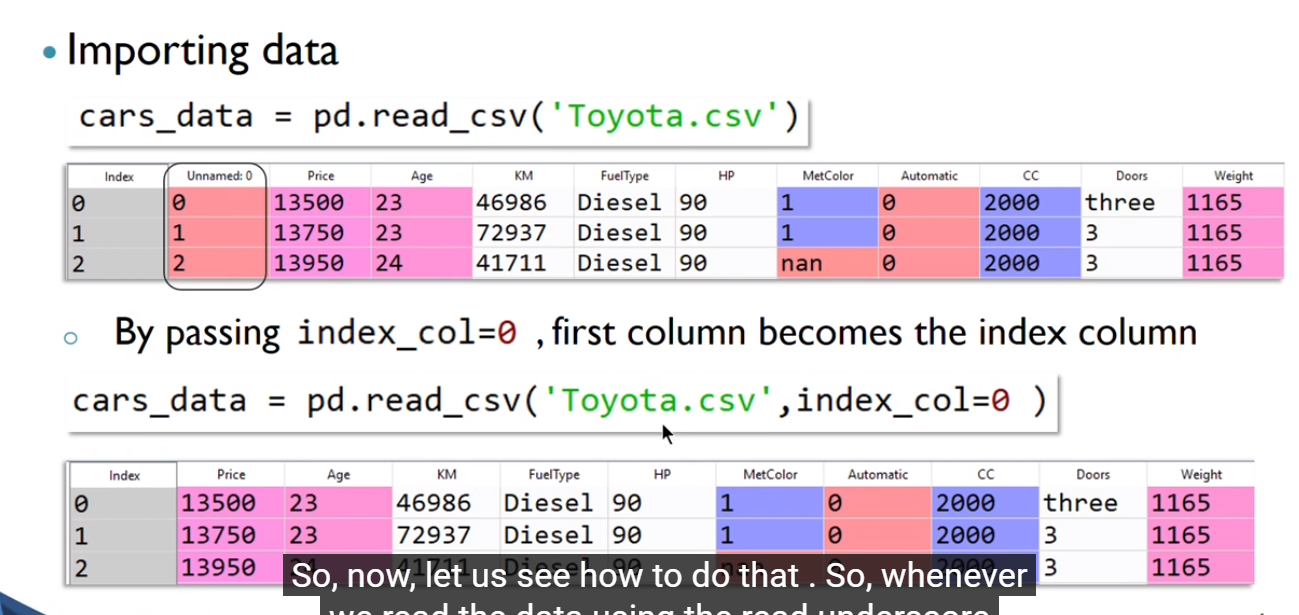
It has rows and columns. Rows are called records or samples. Columns are called variables.

Heterogeneous data means each and every variable automatically gets a data type, we don’t have to explicitly mention the data types. Data type will be based on the type of data that is contained in each variable (column).

Labeled axes means each and every row and column will be labeled. The index for each row starts from 0 to n-1 and labels for column in the sense the names for each variables.

## Importing Data

Look at [Importing Data section](#_Importing_Data) under [Reading Data](#_Reading_Data) section.



HP – horse power

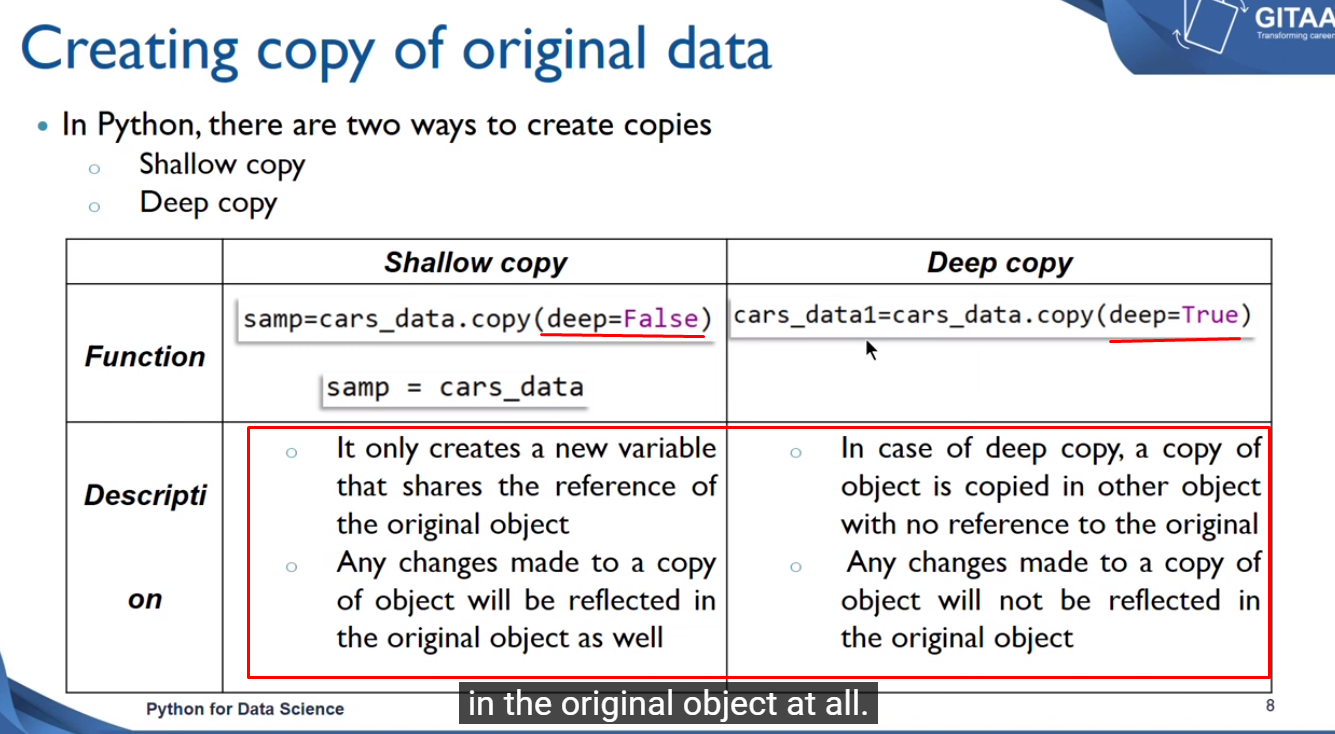
MetColor – 0 means the car doesn’t have metallic color. 1 means it has.

Automatic – 0 means the car is not an automatic ar. 1 means it is.

Weight – weight of the cars in kgs.

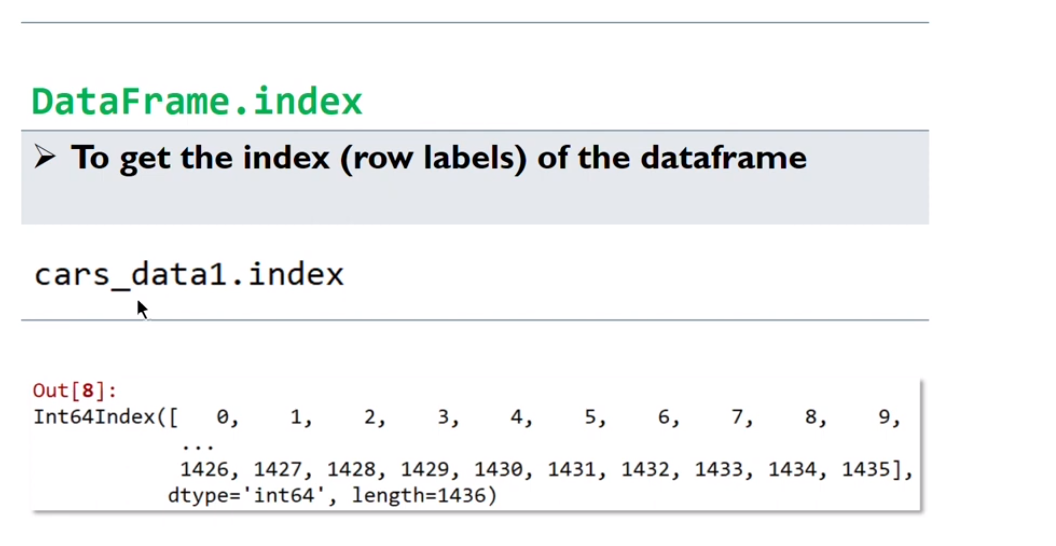
## Creating copy of original data

We don’t want to modify the original data. So, we should create a copy of it.



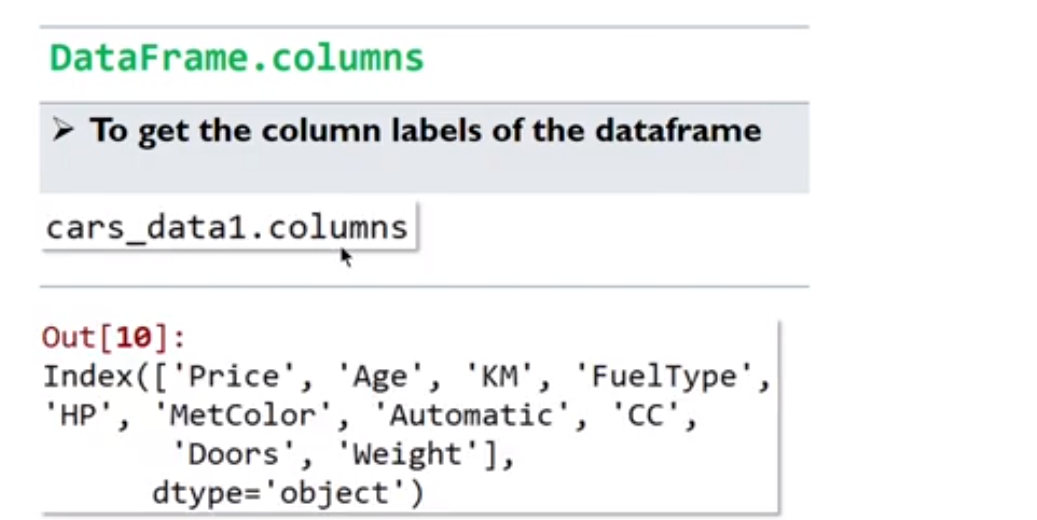
## Attributes of data

### index

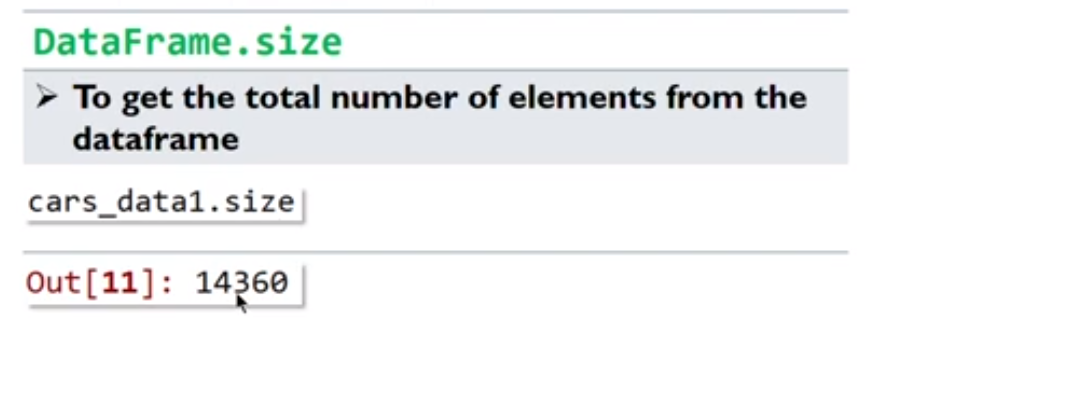


Gives row labels.

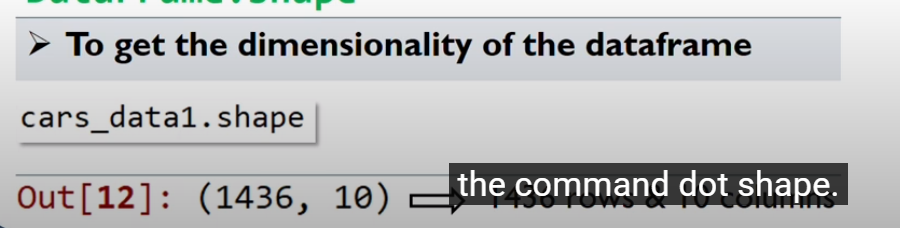
### columns



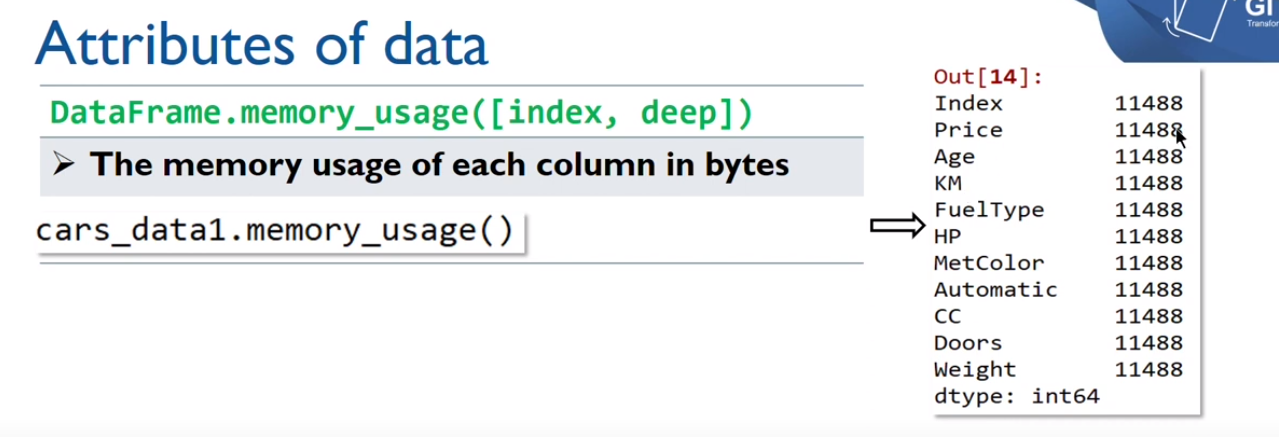
### size



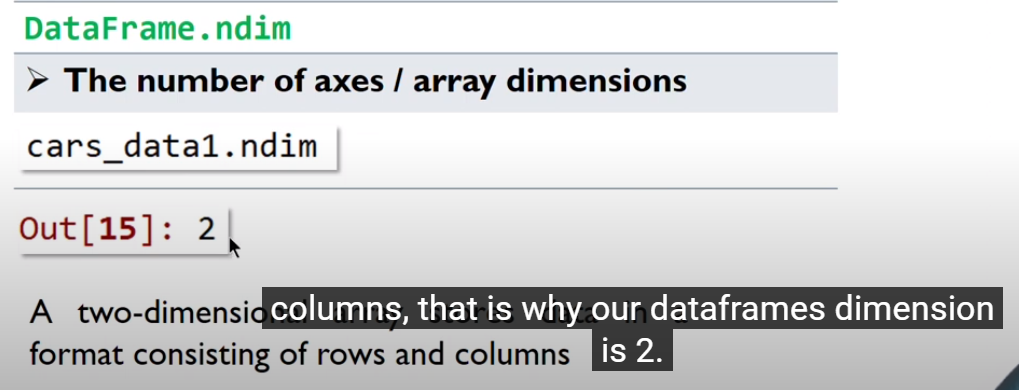
### Shape



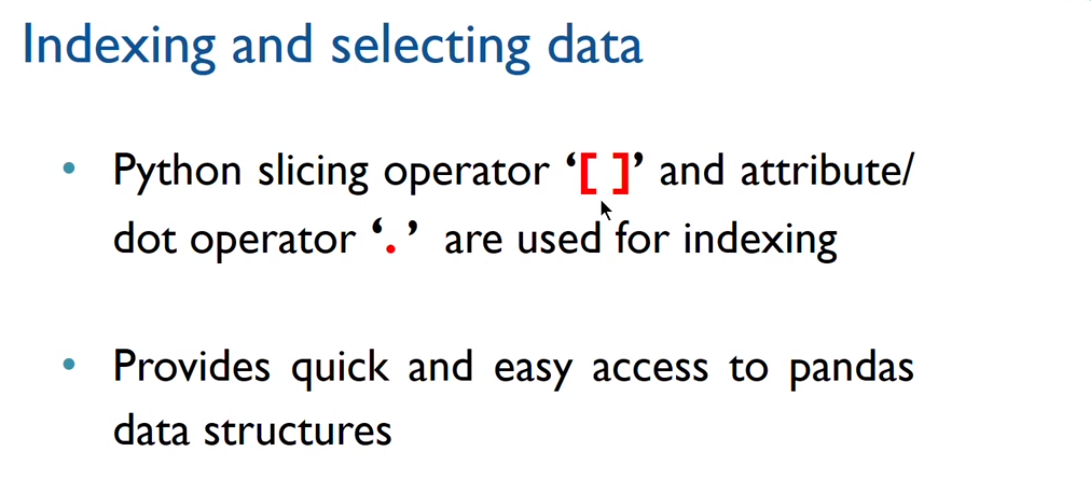
### memory\_usage



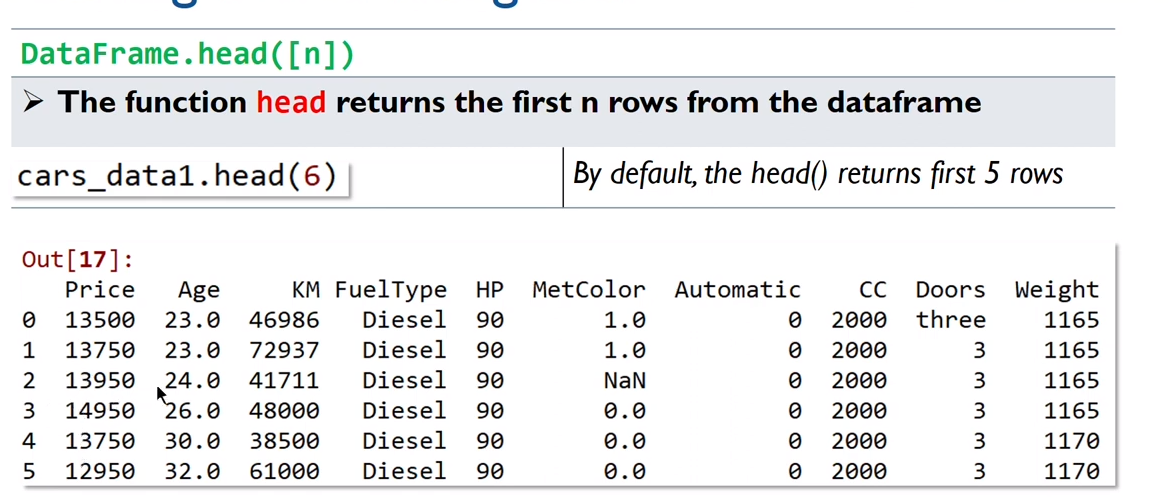
### ndim



### Indexing and selecting data



#### head()



#### tail()