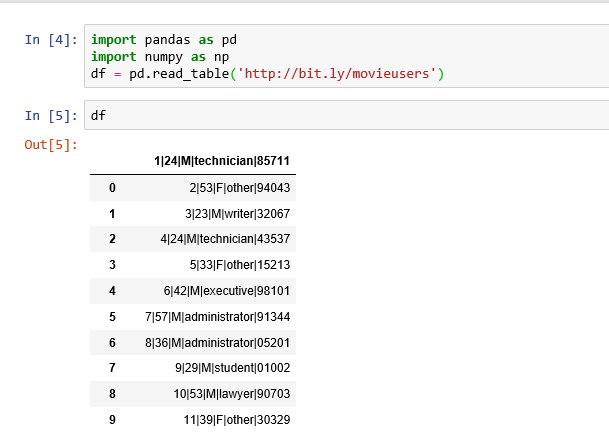
What is pandas

Python users who want to work with data and use a powerful and efficient tool to do so, you do need to know some python.

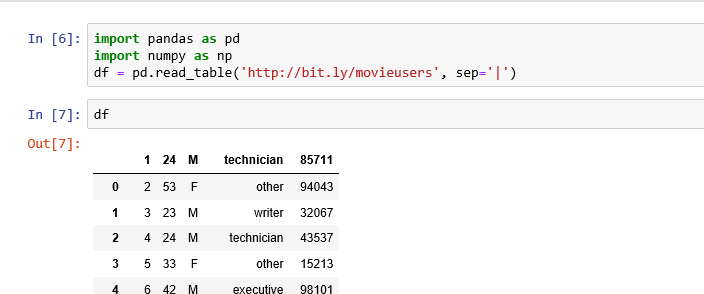
Panda is an open source library in python for data analysis, data manipulation and data visualization. However, these things were possible in python before panda existed. But once pandas came into action lot of data people got exited to work with data with panda.

* It’s well supported by the community.
* Lot of documentation
* Ton of functionality
* It’s under active development
* Lot of people work with panda and they share their work.
* It plays well with other packages.
* Its built-on top of NumPy, numerical python for numerical computing.
* In short, it’s a worth tool if you work with data
* The easy way to get panda you can get it from anaconda distribution.

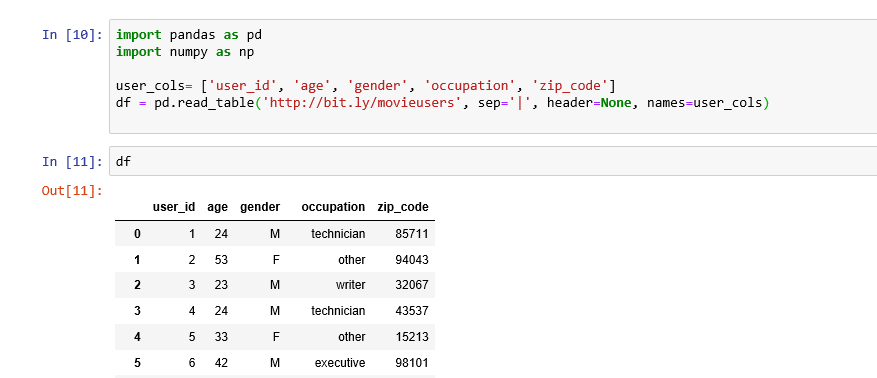
Excel, spreadsheet, csv these are tabular data. By default, table is tab separated data, however a csv contains comma separated data. First, we will import panda into our Jupiter notebook in very conventional way. To be able to work with the data in **Python**, we'll need to read the csv file into a Pandas **DataFrame**. A **DataFrame** is a way to represent and work with tabular data. Tabular data has rows and columns, just like our csv file.



As we ca see here the data is | separated, it makes sense but to view the data more clearly we can separate the columns.



Now the data looks better but there is column name involved, for better understanding we can add the column names. Columns are also known as series in panda.

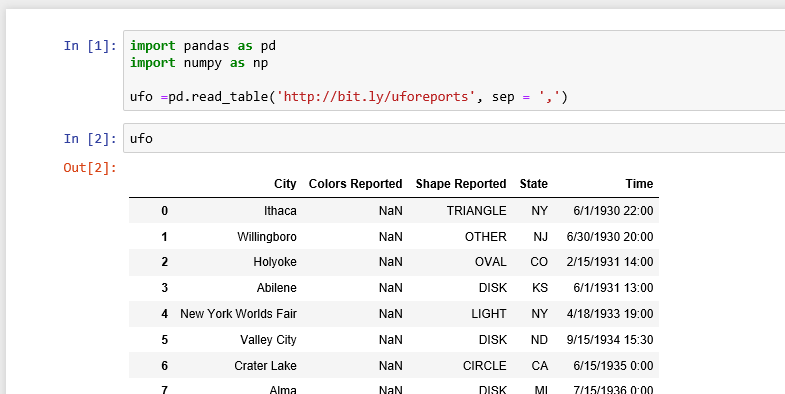


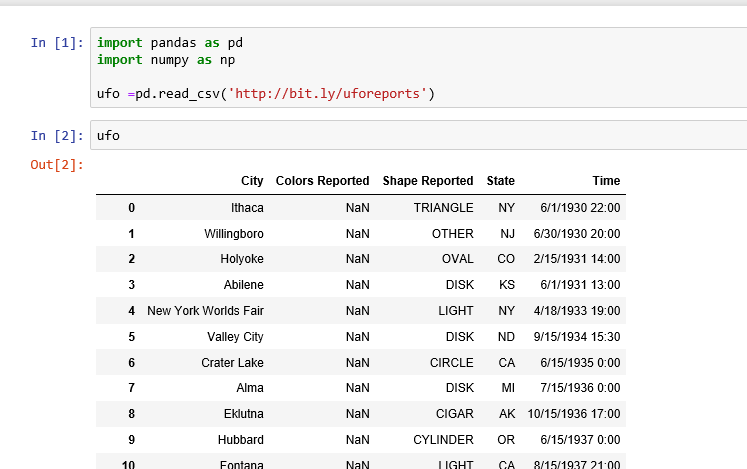
Read table documentation. Sometimes there will be header or footnote on the data we are importing, in that case we can skip the number of rows from both top or bottom by doing, skiprows and skipfooter.

How do I select pandas series from a dataframe?

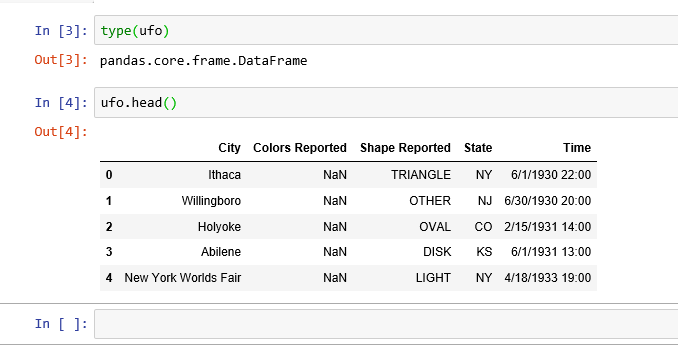
Two basic object types in pandas, dataframe, its basically a table consists with rows and columns, and each of the columns are known as a panda series.

While we do read\_table functionality it understands a tab separated data, hence we do mention explicitly whether the separation is a pipe (|) or comma (,). As in most cases we do have csv files so we can switch from read\_table to read\_csv, its better this way.





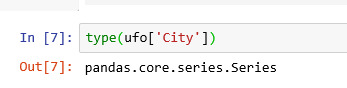
We also can see what type of object the ufo is ?

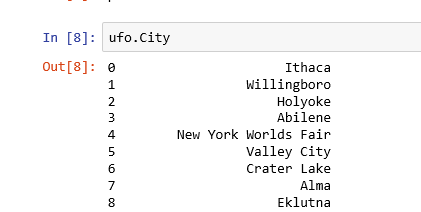


Now to select a series from the dataframe we have two methods.

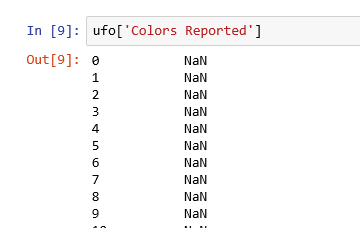
One is [] method and another is comma method. As comma method takes less typing so people prefer the comma method mostly.





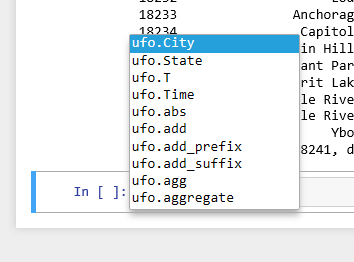


However, some cases the comma method does not work, say if there is a space in the series name. then we must move to the [ ] method.



Comma method has another short way and saves typing by doing this.

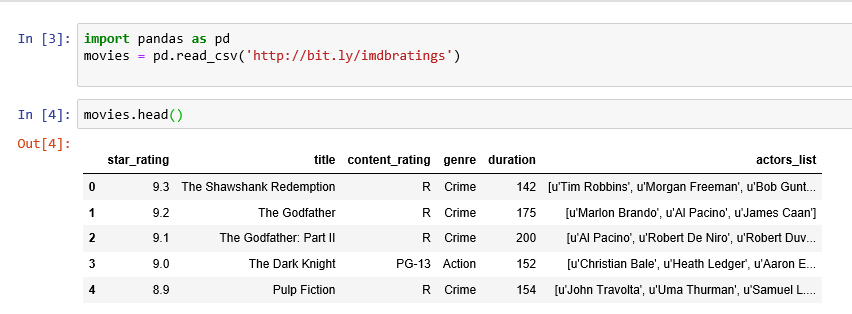
Type ufo. And press tab key, it shows all the methods and attributes along with the column/series name of the ufo dataframe. Simply move up or down and select what you need at that time.



Functions and methods

Why do some pandas commands end with parenthesis and other don’t?

Lets import another dataset from imdb, some data about movies that you might recognise.

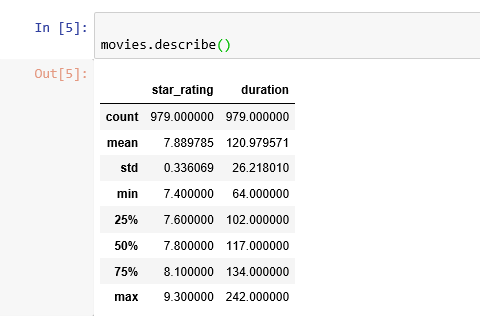


**Head** is a command to select only the top 5 rows rather than selecting the entire dataset.

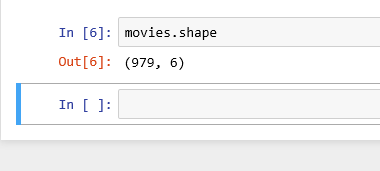
Another command we are going to try out is describe, just like head its end out with parenthesis.

The way describe works is if there is at least one numeric column/series it will show you the descriptive statistic of all numeric columns. Here in this case we have

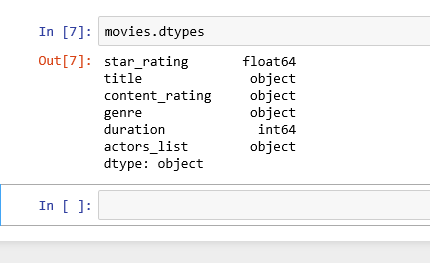
| **star\_rating** | **duration** |
| --- | --- |



So these both have parenthesis after them as you noticed. Now lets try some which don’t.



Now that’s a tuple that tells us that there are 979 rows and 6 columns. Or we can look into “movies.dtypes” which tells us the datatypes of each of the six columns. So, we got the ones with parenthesis and the one which don’t.



Movies is a dataframe, that is a object type. As a dataframe it has certain methods and attributes. The methods are like head and describe, the ones with the parenthesis. And the attributes are the one without parenthesis like shape and dtype.

Methods are like action oriented.

Attributes are like description about who you are.

Say for example, Methods can be

Ashim.talk()

Ashim.eat()

And attributes can be : Ashim.age

Ashim.height

You can also do “movies. and then press tab key”. Unfortunately, it does not show which one is method or attribute.

