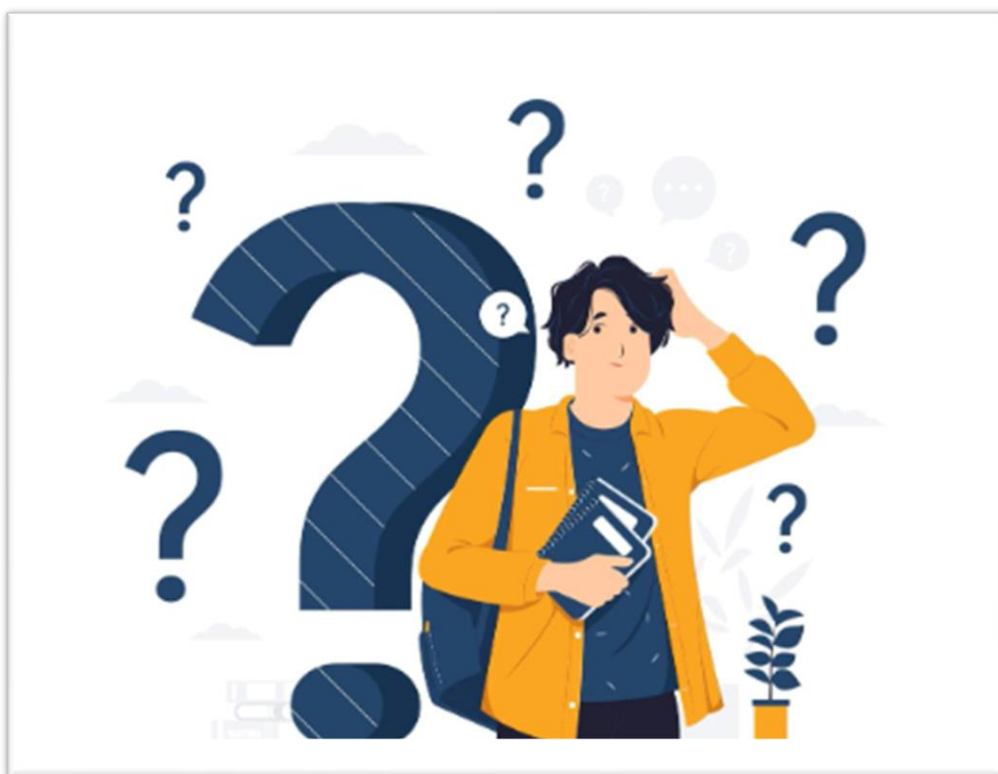


How to import data from various sources using Pandas

 Hira Fahim  2nd April 2023

Introduction

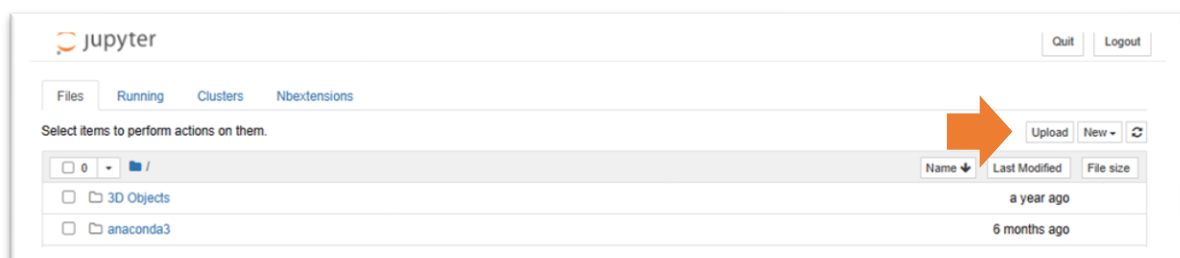
The first step of any data science project is to import the data from the given source. In this article, you will learn how to import data from Four diverse sources into Jupyter notebook using Pandas-Python.



1. CSV file

To import csv to Jupyter notebook, you have two options:

- i. Upload the csv file into Jupyter notebook then import it using pandas.

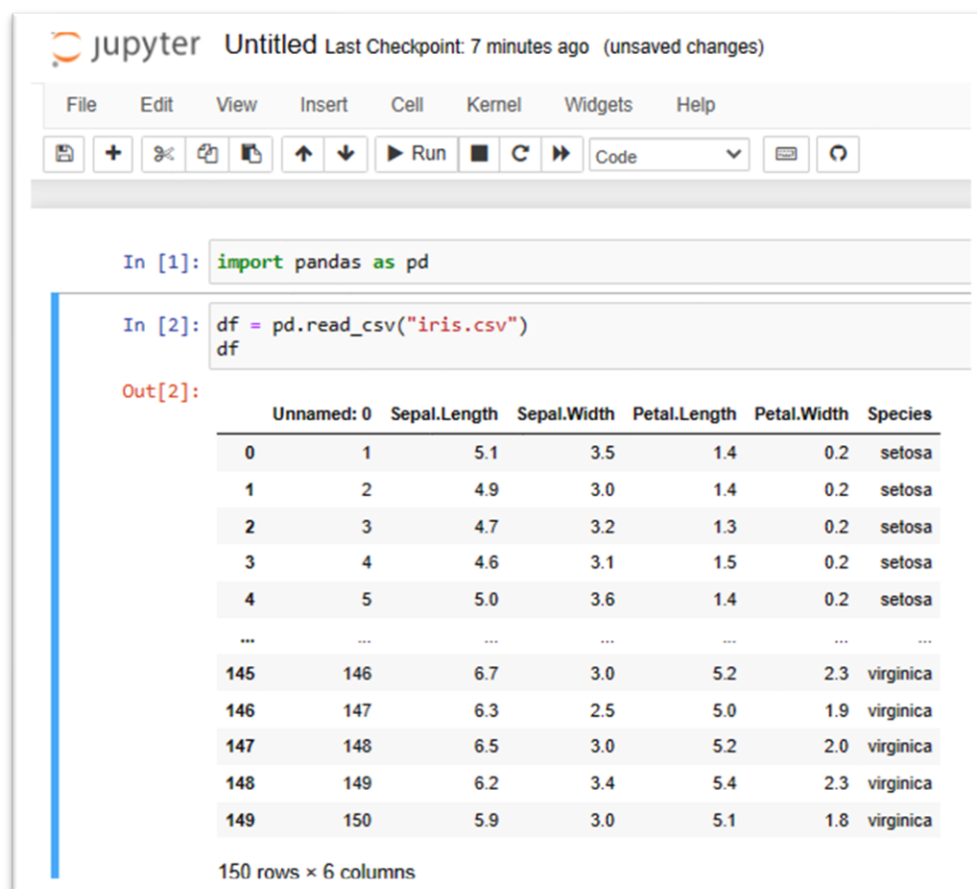




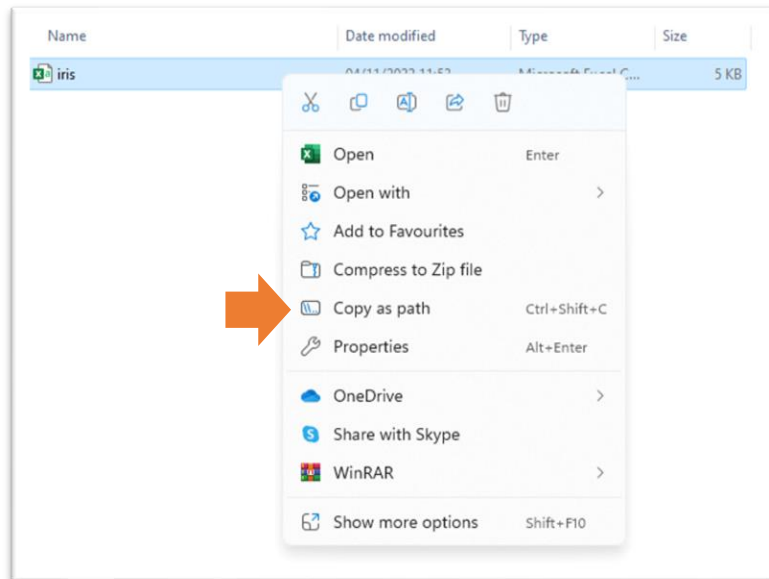
Click upload button. Select the csv file you want to upload here I am uploading “iris.csv” file. You can find this dataset from Kaggle and UCI website, then click blue upload button. Done! Is not simple!!! Next



Click “New” button and select Python 3(ipykernel). New Kernel page will open, now Import pandas, and read the CSV file using pandas (pd.read_csv).



- ii. Another way is to copy the path of the csv file and import it in Jupyter notebook using pandas. Right click on the csv file and select 'copy the path' option.



Insert that path link into “pd. read_csv” command as show below and run the command.

```
jupyter Untitled Last Checkpoint: 19 minutes ago (autosaved)
File Edit View Insert Cell Kernel Widgets Help
+ -> Run [Code]

In [1]: import pandas as pd

In [3]: df = pd.read_csv("D:\Jupyter notebook\iris\iris.csv")
df

Out[3]:
```

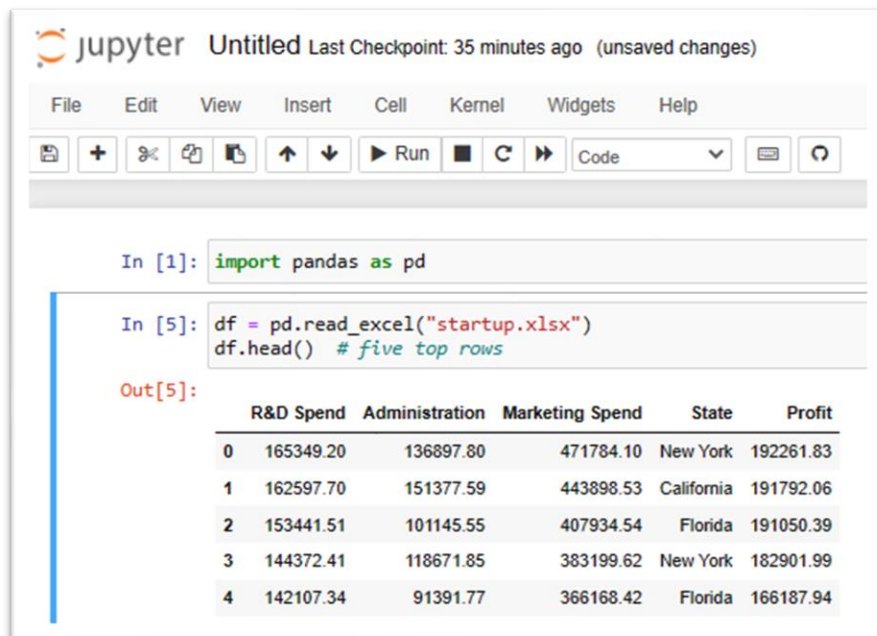
	Unnamed: 0	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
0	1	5.1	3.5	1.4	0.2	setosa
1	2	4.9	3.0	1.4	0.2	setosa
2	3	4.7	3.2	1.3	0.2	setosa
3	4	4.6	3.1	1.5	0.2	setosa
4	5	5.0	3.6	1.4	0.2	setosa
...
145	146	6.7	3.0	5.2	2.3	virginica
146	147	6.3	2.5	5.0	1.9	virginica
147	148	6.5	3.0	5.2	2.0	virginica
148	149	6.2	3.4	5.4	2.3	virginica
149	150	5.9	3.0	5.1	1.8	virginica

150 rows x 6 columns

2. Excel workbook (.xlsx) file

To import excel workbook to Jupyter notebook, you have two options:

- I. Uploads excel workbook same way we have uploaded csv file then import pandas and read xlsx file using pandas `pd.read_excel()`.



The screenshot shows a Jupyter notebook interface with the title "Untitled" and a last checkpoint of 35 minutes ago. The menu bar includes File, Edit, View, Insert, Cell, Kernel, Widgets, and Help. The toolbar contains icons for saving, adding cells, undo, redo, and running code. The code cell shows the following input:

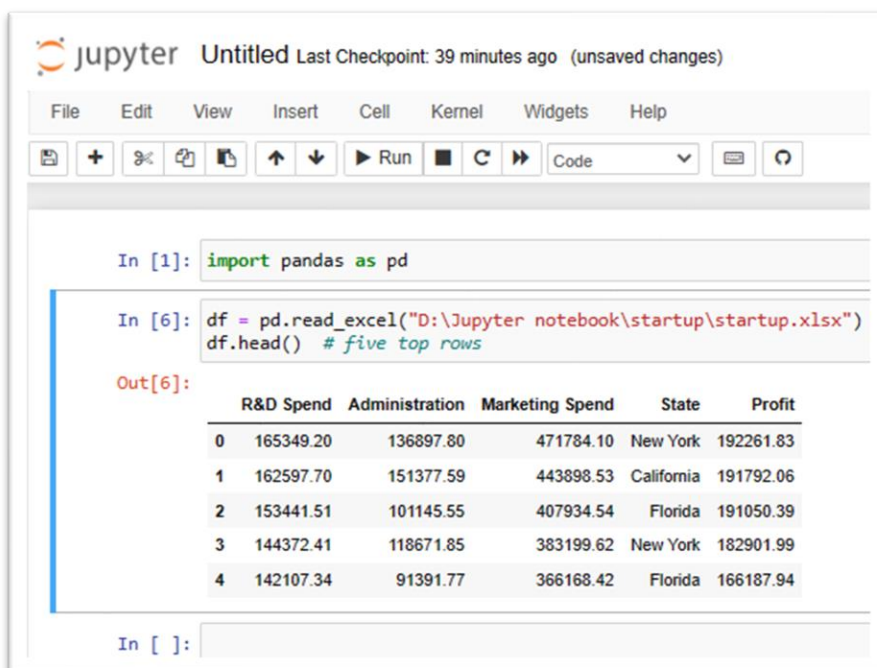
```
In [1]: import pandas as pd
```

```
In [5]: df = pd.read_excel("startup.xlsx")
df.head() # five top rows
```

The output of the code cell is a table with 5 rows and 5 columns:

	R&D Spend	Administration	Marketing Spend	State	Profit
0	165349.20	136897.80	471784.10	New York	192261.83
1	162597.70	151377.59	443898.53	California	191792.06
2	153441.51	101145.55	407934.54	Florida	191050.39
3	144372.41	118671.85	383199.62	New York	182901.99
4	142107.34	91391.77	366168.42	Florida	166187.94

- II. Another way is to copy the path of the .xlsx file and import it in Jupyter notebook using pandas, same way we have done with csv file.



The screenshot shows a Jupyter notebook interface with the title "Untitled" and a last checkpoint of 39 minutes ago. The menu bar includes File, Edit, View, Insert, Cell, Kernel, Widgets, and Help. The toolbar contains icons for saving, adding cells, undo, redo, and running code. The code cell shows the following input:

```
In [1]: import pandas as pd
```

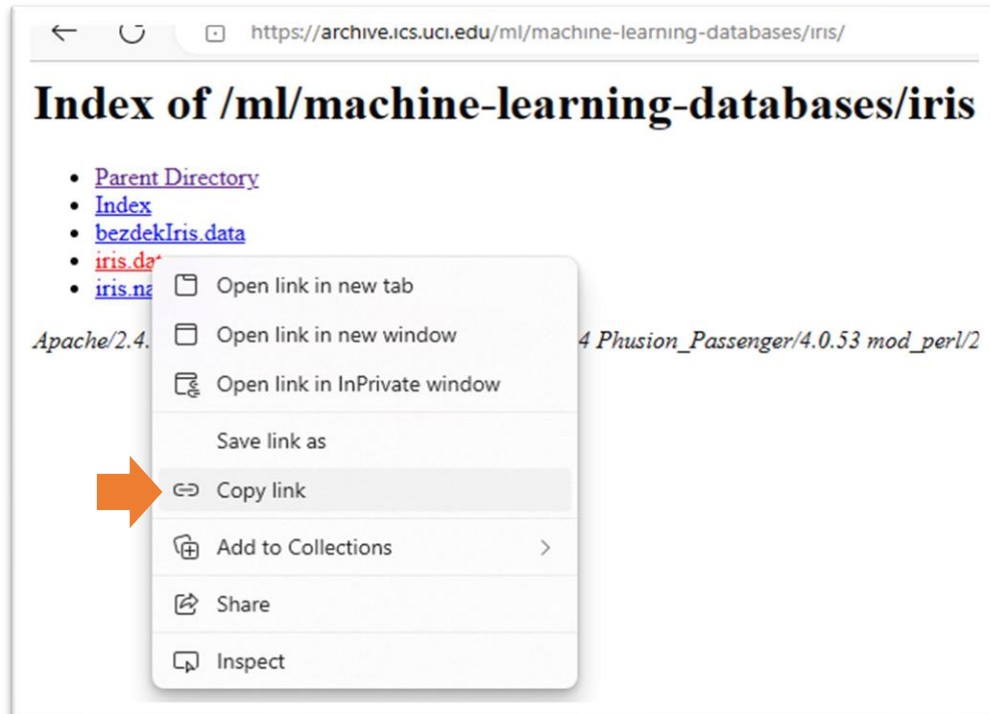
```
In [6]: df = pd.read_excel("D:\\Jupyter notebook\\startup\\startup.xlsx")
df.head() # five top rows
```

The output of the code cell is a table with 5 rows and 5 columns:

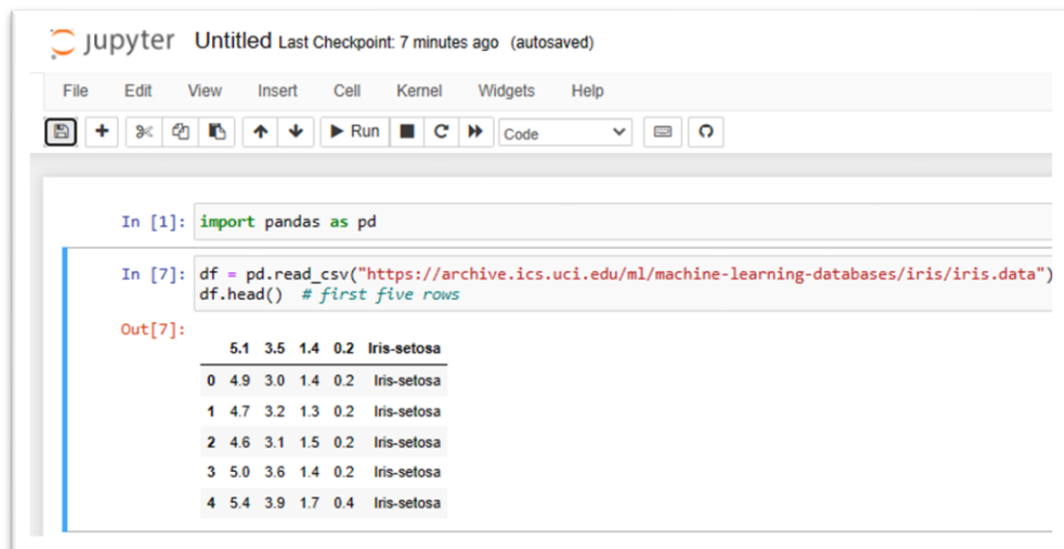
	R&D Spend	Administration	Marketing Spend	State	Profit
0	165349.20	136897.80	471784.10	New York	192261.83
1	162597.70	151377.59	443898.53	California	191792.06
2	153441.51	101145.55	407934.54	Florida	191050.39
3	144372.41	118671.85	383199.62	New York	182901.99
4	142107.34	91391.77	366168.42	Florida	166187.94

3.Import data from a URL.

Here, I am getting “iris data” from UCI website ([Index of /ml/machine-learning-databases/iris/ \(uci.edu\)](https://archive.ics.uci.edu/ml/machine-learning-databases/iris/)). For this right click on the “iris data” and copy link.

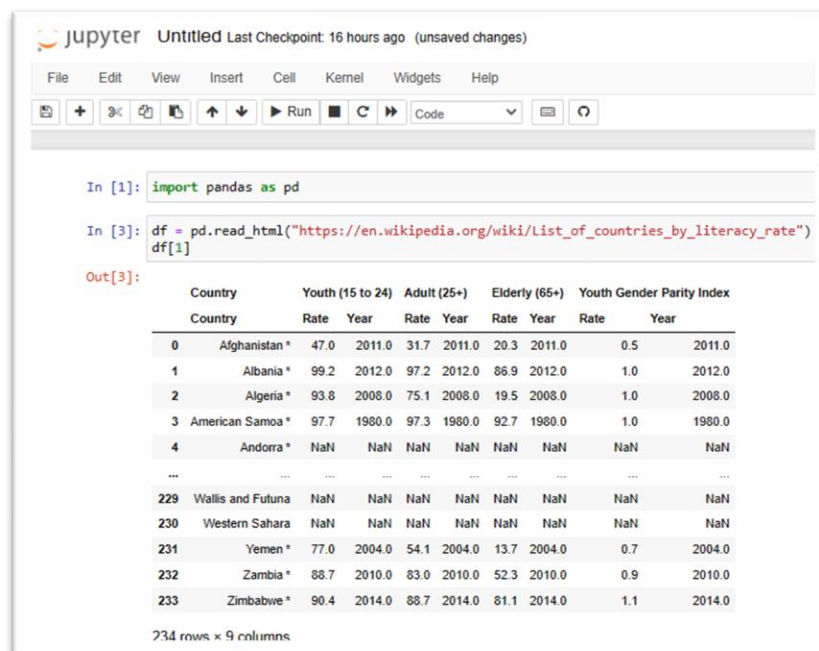


Read the data file via that link in the Jupyter notebook. If the data set is csv file, then use “`pd.read_csv(“”)`” and if it is an “excel workbook” use “`pd.read_excel(“”)`.”



4.Import data from a website.

Here, I am getting a data(table) from a Wikipedia website by turning it to pandas data frame (pd. read_html). Just copy and use the https link.



Conclusion

Data Science is the need of today's time. Many people are switching their career to data science like me. Most of the datasets are either csv file or excel workbook file. This article is helpful for beginners in their first step of any data science project that is, how to import data in their Jupyter notebook using pandas-python.

Reference

- 1) [Index of /ml/machine-learning-databases/iris \(uci.edu\)](https://ml.machine-learning-databases/iris(uci.edu))
Creator: R.A. Fisher, Donor: Michael Marshall (MARSHALL%PLU@io.arc.nasa.gov) Date: July 1988
- 2) [List of countries by literacy rate - Wikipedia](https://en.wikipedia.org/wiki/List_of_countries_by_literacy_rate)