⌄ Importing Dataset in Python

The following explains the various methods to read data in Python including popular formats such as CSV, Text, Excel, SQL, SAS, Stata, and R Data. Loading data into the Python environment is the first step in any data analysis project.

pandas functions for reading the contents of files are named using the pattern .read_(), where indicates the type of the file to read. You've already seen the pandas read_csv() and read_excel() functions. Here are a few others:

read_csv()

read_excel()

read_json()

read_html()

read_sql()

read_pickle()

These functions have a parameter that specifies the target file path. It can be any valid string that represents the path, either on a local machine or in a URL. Other objects are also acceptable depending on the file type. As shown in the following two examples.

```
#Importing pandas package

import pandas as pd


#You need the path to your csv file
#Use pd.read_csv

path = r"C:\Users\KURIA\Downloads\MusicData.csv"

df = pd.read_csv(path)

df.head()
```

	?Format	Metric	Year	Number of Records	Value (Actual)
0	CD	Units	1973	1	NaN
1	CD	Units	1974	1	NaN
2	CD	Units	1975	1	NaN
3	CD	Units	1976	1	NaN
4	CD	Units	1977	1	NaN

```python
#Importing from a url

url = "https://gist.githubusercontent.com/netj/8836201/raw/6f9306ad21398ea43cba4f7d537619d0e
df1 = pd.read_csv(url, skiprows=5)

df1.head()
```

	5	3.6	1.4	.2	Setosa
0	5.4	3.9	1.7	0.4	Setosa
1	4.6	3.4	1.4	0.3	Setosa
2	5.0	3.4	1.5	0.2	Setosa
3	4.4	2.9	1.4	0.2	Setosa
4	4.9	3.1	1.5	0.1	Setosa

```python
#importing .txt files
#Use rither read_table or read_csv (if the text file is comma separated ot tab separeted)

path = r"C:\Users\KURIA\Documents\text\ex1.txt"
path2 = r"C:\Users\KURIA\Documents\text\ex2.txt"

df3 = pd.read_table(path)
df3

df4= pd.read_csv(path2, sep ="\t")
df4
```

This is a text file, created by Me, date 12th Jan, 2023

```
#importing excel files
#Make sure 'openpyxl' module is installed

path4 =  r"C:\Users\KURIA\Downloads\Iganga_source codes.xlsx"

dfexcel = pd.read_excel(path4)

dfexcel
```

	code	English item	Count	target_concept_id	target_vocabulary_id
0	1	Physician	3846	32577	Provider
1	2	Interva	1260	NaN	NaN
2	1	Other parasitologically confirmed malaria, not...	868	NaN	NaN
3	2	HIV disease resulting in multiple diseases cla...	391	NaN	NaN
4	3	Extremely low birth weight	369	NaN	NaN
...	...	...	...	...	...
235	99	Asthma	1	NaN	NaN
236	100	Unspecified communicable diseases	1	NaN	NaN

Using read_filetype() you can load different datas from different sources.

		Respiratory Infections			
238	102	Pregnancy-related sepsis	1	NaN	NaN