

Data Analysis with Python

Cheat Sheet: Importing Data Sets

Package/Method	Description	Code Example
Read CSV data set	Read the CSV file containing a data set to a pandas data frame	<pre>1 df = pd.read_csv(<CSV_path>, header = None) # load without header df = pd.read_csv(<CSV_path>, header = 0) # load using first row as header</pre>
Print first few entries	Print the first few entries (default 5) of the pandas data frame	<pre>1 df.head(n) #n=number of entries; default 5</pre>
Print last few entries	Print the last few entries (default 5) of the pandas data frame	<pre>1 df.tail(n) #n=number of entries; default 5</pre>
Assign header names	Assign appropriate header names to the data frame	<pre>1 df.columns = headers</pre>
Replace "?" with NaN	Replace the entries "?" with NaN entry from Numpy library	<pre>1 df = df.replace("?", np.nan)</pre>
Retrieve data types	Retrieve the data types of the data frame columns	<pre>1 df.dtypes</pre>
Retrieve statistical description	Retrieve the statistical description of the data set. Defaults use is for only numerical data types. Use include="all" to create summary for all variables	<pre>1 df.describe() #default use df.describe(include="all")</pre>
Retrieve data set summary	Retrieve the summary of the data set being used, from the data frame	<pre>1 df.info()</pre>
Save data frame to CSV	Save the processed data frame to a CSV file with a specified path	<pre>1 df.to_csv(<output CSV path>)</pre>