

# Working with Data in Python Cheat Sheet

## Reading and writing files

Package/Method	Description	Syntax and Code Example
File opening modes	Different modes to open files for specific operations.	Syntax: r (reading) w (writing) a (appending) + (updating: read/write) b (binary, otherwise text)  Examples: with open("data.txt", "r") as file: content = file.read() print(content) wi
File reading methods	Different methods to read file content in various ways.	Syntax:  <pre>file.readlines() # reads all lines as a list readline() # reads the next line as a string file.read() # reads the entire file content as a string</pre> Example:  <pre>with open("data.txt", "r") as file:     lines = file.readlines()     next_line = file.readline()     content = file.read()</pre>
File writing methods	Different write methods to write content to a file.	Syntax:  <pre>file.write(content) # writes a string to the file file.writelines(lines) # writes a list of strings to the file</pre> Example:  <pre>lines = ["Hello\n", "World\n"] with open("output.txt", "w") as file:     file.writelines(lines)</pre>
Iterating over lines	Iterates through each line in the file using a `loop`.	Syntax:  <pre>for line in file: # Code to process each line</pre> Example:  <pre>with open("data.txt", "r") as file:     for line in file: print(line)</pre>
Open() and close()	Opens a file, performs operations, and explicitly closes the file using the close() method.	Syntax:  <pre>file = open(filename, mode) # Code that uses the file file.close()</pre> Example:  <pre>file = open("data.txt", "r") content = file.read() file.close()</pre>
with open()	Opens a file using a with block, ensuring automatic file closure after usage.	Syntax:  <pre>with open(filename, mode) as file: # Code that uses the file</pre> Example:  <pre>with open("data.txt", "r") as file:     content = file.read()</pre>

## Pandas

Package/Method	Description	Syntax and Code Example
<code>.read_csv()</code>	Reads data from a <code>.CSV</code> file and creates a DataFrame.	Syntax: <code>dataframe_name = pd.read_csv("filename.csv")</code> Example: <code>df = pd.read_csv("data.csv")</code>
<code>.read_excel()</code>	Reads data from an Excel file and creates a DataFrame.	Syntax: <code>dataframe_name = pd.read_excel("filename.xlsx")</code> Example: <code>df = pd.read_excel("data.xlsx")</code>
<code>.to_csv()</code>	Writes DataFrame to a CSV file.	Syntax: <code>dataframe_name.to_csv("output.csv", index=False)</code> Example: <code>df.to_csv("output.csv", index=False)</code>
Access Columns	Accesses a specific column using <code>[]</code> in the DataFrame.	Syntax: <code>dataframe_name["column_name"]</code> # Accesses single column <code>dataframe_name[["column1", "column2"]]</code> # Accesses multiple columns Example: <code>df["age"]</code> <code>df[["name", "age"]]</code>
<code>describe()</code>	Generates statistics summary of numeric columns in the DataFrame.	Syntax: <code>dataframe_name.describe()</code> Example: <code>df.describe()</code>
<code>drop()</code>	Removes specified rows or columns from the DataFrame. <code>axis=1</code> indicates columns. <code>axis=0</code> indicates rows.	Syntax: <code>dataframe_name.drop(["column1", "column2"], axis=1, inplace=True)</code> <code>dataframe_name.drop(index=[row1, row2], axis=0, inplace=True)</code> Example: <code>df.drop(["age", "salary"], axis=1, inplace=True)</code> # Will drop columns <code>df.drop(index=[5, 10], axis=0, inplace=True)</code> # Will drop rows
<code>dropna()</code>	Removes rows with missing NaN values from the DataFrame. <code>axis=0</code> indicates rows.	Syntax: <code>dataframe_name.dropna(axis=0, inplace=True)</code> Example: <code>df.dropna(axis=0, inplace=True)</code>
<code>uplicated()</code>	Duplicate or repetitive values or records within a data set.	Syntax: <code>dataframe_name.duplicated()</code> Example:

		<code>duplicate_rows = df[df.duplicated()]</code>
Filter Rows	Creates a new DataFrame with rows that meet specified conditions.	<p>Syntax:</p> <pre>filtered_df = dataframe_name[(Conditional_statements)]</pre> <p>Example:</p> <pre>filtered_df = df[(df["age"] &gt; 30) &amp; (df["salary"] &lt; 50000)]</pre>
groupby()	Splits a DataFrame into groups based on specified criteria, enabling subsequent aggregation, transformation, or analysis within each group.	<p>Syntax:</p> <pre>grouped = dataframe_name.groupby(by, axis=0, level=None, as_index=True, sort=True, group_keys=True, squeeze=False, observed=False, dropna=True)</pre> <p>Example:</p> <pre>grouped = df.groupby(["category", "region"]).agg({"sales": "sum"})</pre>
head()	Displays the first n rows of the DataFrame.	<p>Syntax:</p> <pre>dataframe_name.head(n)</pre> <p>Example:</p> <pre>df.head(5)</pre>
Import pandas	Imports the Pandas library with the alias pd.	<p>Syntax:</p> <pre>import pandas as pd</pre> <p>Example:</p> <pre>import pandas as pd</pre>
info()	Provides information about the DataFrame, including data types and memory usage.	<p>Syntax:</p> <pre>dataframe_name.info()</pre> <p>Example:</p> <pre>df.info()</pre>
merge()	Merges two DataFrames based on multiple common columns.	<p>Syntax:</p> <pre>merged_df = pd.merge(df1, df2, on=["column1", "column2"])</pre> <p>Example:</p> <pre>merged_df = pd.merge(sales, products, on=["product_id", "category_id"])</pre>
print DataFrame	Displays the content of the DataFrame.	<p>Syntax:</p> <pre>print(df) # or just type df</pre> <p>Example:</p> <pre>print(df) df</pre>
replace()	Replaces specific values	Syntax:

	in a column with new values.	<pre>dataframe_name["column_name"].replace(old_value, new_value, inplace=True)</pre> <p>Example:</p> <pre>df["status"].replace("In Progress", "Active", inplace=True)</pre>
tail()	Displays the last n rows of the DataFrame.	<p>Syntax:</p> <pre>dataframe_name.tail(n)</pre> <p>Example:</p> <pre>df.tail(5)</pre>

## Numpy

Package/Method	Description	Syntax and Code Example
Importing NumPy	Imports the NumPy library.	<p>Syntax:</p> <pre>import numpy as np</pre> <p>Example:</p> <pre>import numpy as np</pre>
np.array()	Creates a one or multi-dimensional array,	<p>Syntax:</p> <pre>array_1d = np.array([list1 values]) # 1D Array array_2d = np.array([[list1 values], [list2 values]]) # 2D Array</pre> <p>Example:</p> <pre>array_1d = np.array([1, 2, 3]) # 1D Array array_2d = np.array([[1, 2], [3, 4]]) # 2D Array</pre>
Numpy Array Attributes	<ul style="list-style-type: none"> <li>- Calculates the mean of array elements</li> <li>- Calculates the sum of array elements</li> <li>- Finds the minimum value in the array</li> <li>- Finds the maximum value in the array</li> <li>- Computes dot product of two arrays</li> </ul>	<p>Example:</p> <pre>np.mean(array) np.sum(array) np.min(array) np.max(array) np.dot(array_1, array_2)</pre>



# Skills Network

© IBM Corporation. All rights reserved.