Cheat sheet of the most used Pandas commands

Harsh Choudhary

Cheat sheet of the most used Pandas commands

Data structures

Command	Description
pd.Series(data, index=None, dtype=None)	Create a 1D labeled array.
pd.DataFrame(data, index=None, columns=None)	Create a 2D tabular structure.
pd.Categorical(values)	Create a categorical dtype for efficient storage.

Import and options

Command	Description
import pandas as pd	Standard import alias for pandas.
pd.set_option(key, value)	Change display/behavior option (e.g., max rows/cols).
pd.get_option(key)	Read the current value of an option.
pd.reset_option(key)	Reset an option to its default.
pd.describe_option(pattern=None)	Show details about available options.
pd.options	Namespace object to access all options.

Input/Output (I/O)

Command	Description
pd.read_csv(path, …) / df.to_csv(path, …)	Read/write CSV.
pd.read_excel(path, sheet_name=0) / df.to_excel(path)	Read/write Excel sheets.
pd.read_json(path_or_buf) / df.to_json(path)	Read/write JSON.
pd.read_html(source) / df.to_html(buf)	Parse HTML tables to/from Data Frames.
pd.read_sql(sql, con) / df.to_sql(name, con)	Read from / write to SQL.
pd.read_pickle(path) / df.to_pickle(path)	Serialize/deserialize pandas objects.
pd.read_feather(path) / df.to_feather(path)	Fast Arrow Feather format.
pd.read_parquet(path) / df.to_parquet(path)	Columnar Parquet format.
pd.read_orc(path) / df.to_orc(path)	Read/write ORC files.
pd.read_stata(path) / df.to_stata(path)	Stata files.
pd.read_spss(path)	Import SPSS data.
pd.read_sas(path)	Import SAS data.
pd.read_hdf(path, key=None) / df.to_hdf(path, key)	HDF5 storage.
pd.read_gbq(query, project=…) / df.to_gbq(…)	Big Query I/O.
pd.read_clipboard() / df.to_clipboard()	Clipboard I/O.
pd.read_xml(path) / df.to_xml(path)	XML I/O.

Data inspection

Command	Description
df.head(n), df.tail(n), df.sample(n)	Peek at rows.
df.info()	Column types, non-nulls, memory.
df.describe(include=None)	Summary stats.
df.shape, df.size, df.ndim	Dimensions, elements, axes.
df.dtypes	Per-column dtypes.
df.columns, df.index	Labels of columns/rows.
df.values	Underlying NumPy array (discouraged for mixed dtypes).
df.memory_usage(deep=True)	Memory by column.
df.isnull(), df.notnull()	NA mask.

Selection and indexing

Command	Description
df["col"], df[["c1","c2"]]	Select column(s).
df.loc[row_sel, col_sel]	Label-based selection.
df.iloc[row_sel, col_sel]	Position-based selection.

df.at[label_row, label_col]	Fast scalar by label.
df.iat[i, j]	Fast scalar by position.
df.filter(items=None, like=None, regex=None, axis=0/1)	Filter by labels/regex.
df.query("expr")	SQL-like Boolean query.
df.get("col", default=None)	Safe column get.
df.xs(key, level=None, axis=0)	Cross-section on axis/level.

Cleaning missing and duplicates

Command	Description
df.dropna(axis=0, how="any", subset=None)	Drop NA rows/cols.
df.fillna(value, method=None)	Fill NA with scalar/dict/ffill/bfill.
df.replace(to_replace, value=None, regex=False)	Replace values/patterns.
df.interpolate(method="linear")	Interpolate numeric gaps.
df.duplicated(subset=None, keep="first")	Mark duplicate rows.
df.drop_duplicates(subset=None, keep="first")	Remove duplicates.

Manipulation and transformation

Command	Description
df.sort_values(by, ascending=True)	Sort by column(s).
df.sort_index(axis=0/1)	Sort by index/columns.
df.rename(columns=…, index=…)	Rename labels.
df.set_index(keys, drop=True)	Make column(s) the index.
df.reset_index(drop=False)	Restore index to column(s).
df.drop(labels, axis=0/1)	Drop rows/columns.
df.insert(loc, name, values)	Insert column by position.
df.pop(name)	Remove and return a column.
df.assign(new=expr, …)	Add column(s) without mutating original.
df.apply(func, axis=0/1)	Apply along axis.
df.applymap(func)	Elementwise over Data Frame.
s.map(func/dict)	Map over Series.
df.eval("expr")	Evaluate expressions on columns.
df.astype(dtype or dict)	Convert dtype(s).
df.clip(lower=None, upper=None)	Cap values.

Grouping, aggregation, reshaping

Merge and join

Command	Description
pd.concat(objs, axis=0/1, ignore_index=False)	Concatenate along axis.
pd.merge(left, right, on=None, how="inner"/"left"/"right"/"outer")	SQL-style joins.
df.pivot(index, columns, values)	Long → wide (no agg).
df.join(other, on=None, how="left")	Index-based join or on key.
df.combine_first(other)	Fill NA from another object.
df.melt(id_vars, value_vars, var_name, value_name)	Wide → long.
df.stack() / df.unstack(level)	Rotate between index/columns.
pd.crosstab(index, columns, margins=False)	Frequency table.
pd.cut(x, bins) / pd.qcut(x, q)	Bin numeric data (equal width/quantile).

Stats and math

Command	Description
df.mean(), df.median(), df.mode()	Central tendency.
df.min(), df.max(), df.sum()	Extrema/sum.
df.cumsum(), df.cumprod()	Cumulative ops.
df.var(), df.std()	Variance/standard deviation.
df.corr(), df.cov()	Correlation/covariance.

df.skew(), df.kurt()	Skewness/kurtosis.
df.rank(method="average")	Ranks per column.

Time series

Command	Description
pd.to_datetime(obj, format=…)	Parse to datetime64.
df.set_index("dt").resample("M").agg(func)	Resample by time.
df.asfreq("H")	Set fixed frequency without aggregation.
df.shift(1)	Shift values by periods.
df.rolling(window).agg(func)	Rolling window stats.
df.expanding(min_periods).agg(func)	Expanding window.
df.ewm(alpha=…).mean()	Exponentially weighted mean.
pd.date_range(start, end, freq)	Build DatetimeIndex.
pd.period_range(start, end, freq)	Period ranges.
pd.timedelta_range(start, periods, freq)	Timedelta ranges.

Plotting (built-in)

Command	Description
df.plot(kind="line"/"bar"/"barh"/"area"/"pie"/"scatter"/"hexbin"/"density")	Quick charts from DataFrame/Series.
df.hist(bins=…)	Histograms per numeric column.
df.boxplot()	Box-and-whisker plots.

Vectorized string ops (Series.str)

Command	Description
.str.lower(), .str.upper(), .str.title()	Case transforms.
.str.len(), .str.strip(), .str.pad()	Length/trim/pad.
.str.contains(pat, regex=True)	Substring/regex test.
.str.replace(pat, repl, regex=…)	Substitute patterns.

.str.split(pat, n, expand)	Split to lists or columns.
.str.cat(others, sep)	Concatenate strings.
.str.startswith(x), .str.endswith(y)	Prefix/suffix checks.
.str.extract(pattern), .str.findall(pattern)	Regex capture/matches.

Categoricals

Command	Description
s.astype("category") / pd.Categorical(vals)	Convert/create categorical.
s.cat.categories	View categories.
s.cat.codes	Integer codes per category.
s.cat.rename_categories(mapper)	Rename categories.
s.cat.reorder_categories(new_order, ordered=True)	Reorder and set ordering.
s.cat.add_categories(new) / s.cat.remove_categories(rem)	Modify set.
s.cat.remove_unused_categories()	Drop unused categories.

Utilities

Command	Description
pd.factorize(values)	Encode labels to integers + uniques.
pd.unique(values)	Unique values (preserves order).
pd.value_counts(values, normalize=False)	Frequency counts.
pd.get_dummies(df, columns=…)	One-hot encoding.
pd.to_numeric(obj, errors="coerce")	Convert to numeric, set bad to NaN.
pd.to_timedelta(obj)	Convert to timedelta dtype.