Pandas Cheat Sheet

Key and Imports

df = Pandas DataFrame

s = Pandas Series

Import with: import pandas as pd, import numpy as np

Importing Data

pd.read_csv(filename) - Read CSV file

pd.read_excel(filename) - Read Excel file

pd.read_sql(query, connection_object) - Read SQL data

pd.read_json(json_string) - Read JSON data

pd.read_html(url) - Extract tables from HTML

pd.read_clipboard() - Read clipboard

pd.DataFrame(dict) - Create DataFrame from dict

Exporting Data

df.to_csv(filename) - Write to CSV

df.to_excel(filename) - Write to Excel

df.to_sql(table_name, connection_object) - Write to SQL

df.to_json(filename) - Write to JSON

Creating Test Objects

pd.DataFrame(np.random.rand(7,18)) - Random data

pd.Series(my_list) - Create Series

df.index = pd.date_range('1940/1/20', periods=df.shape[0]) - Add date index

Viewing/Inspecting Data

df.head(n), df.tail(n) - View rows

df.shape - Dimensions

df.info() - Summary

s.value_counts(dropna=False) - Unique counts

df.apply(pd.Series.value_counts) - Unique counts per column

Selection

df[col] - Single column

df[[col1, col2]] - Multiple columns

Pandas Cheat Sheet

s.iloc[pos], s.loc[index] - Position/index selection

df.iloc[row, col] - Element selection

Data Cleaning

Rename columns: df.columns = ['a','b','c']

Check nulls: pd.isnull(), pd.notnull()

Drop nulls: df.dropna()

Fill nulls: df.fillna(value)

Replace values: s.replace(old, new)

Change datatype: s.astype(float)

Rename columns/index selectively

Change index: df.set_index('column_name')

Filter, Sort, and GroupBy

Filter rows with conditions

Sort: df.sort_values()

GroupBy: df.groupby(col).agg()

Pivot table: df.pivot_table()

Apply functions: df.apply()

Join/Combine

Append rows: df1.append(df2)

Concatenate columns: pd.concat([df1, df2], axis=1)

Join SQL-style: df1.join(df2, on=col, how='inner')

Statistics

Summary: df.describe()

Mean, median, std, min, max

Correlation: df.corr()

Count non-null: df.count()