

Pandas Cheat Sheet

DataFrame Information	
Number of rows & columns	df.shape
Total elements (rows * columns)	df.size
View first/last 5 rows	df.head() ` / ` df.tail()
View first/last ` n ` rows	df.head(n) ` / ` df.tail(n)
Column names	df.columns
Data types	df.dtypes
Summary info (types, non-null counts)	df.info()
Data Selection	
Select column(s)	df['col'] or df[['col1','col2']]
Select rows by index label	df.loc[index]
Select single row by label	df.loc['row_label']
Select multiple rows by labels	df.loc[['row1', 'row3']]
Select range of rows (inclusive)	df.loc['row2':'row4']
Select specific row & column	df.loc['row1', 'ColumnA']
Select rows & multiple columns	df.loc[:, ['ColA', 'ColB']]
Conditional selection	df.loc[df['ColA'] > 50]
Modify values using loc	df.loc[df['ColB'] < 0, 'ColB'] = 0
Select rows by index position	df.iloc[position]
Select single row by position	df.iloc[0] (first row)
Select multiple rows by positions	df.iloc[[0, 2, 4]]
Select range of rows (exclusive end)	df.iloc[2:5] (rows 2,3,4)
Select specific row & column by position	df.iloc[2, 3] (3rd row, 4th column)
Select all rows & specific columns	df.iloc[:, 1:4] (columns 1,2,3)
Last row	df.iloc[-1]
Boolean filtering	df[df['col'] > 5]
Select specific rows & columns	df.loc[row_selection, cols]
Get specific value	df.at[index, 'col']
Data Cleaning	
Drop rows with missing values	df.dropna()
Fill missing values	df.fillna(value)
Replace values	df.replace(old, new)
Remove duplicates	df.drop_duplicates()
Check for duplicates	df.duplicated()
Rename columns	df.rename(columns={'old': 'new'})
Reset Index	df.reset_index()
Convert data type	df['col'].astype(type)
Data Manipulation	
Add new column	df['new_col'] = values

Drop column(s)	df.drop(columns=['col'])
Apply function to column	df['col'].apply(func)
Group by & aggregate	df.groupby('col').agg({'col':'mean'})
Sort values	df.sort_values('col')
Merge DataFrames	pd.merge(df1, df2, on='col')
Concatenate DataFrames	pd.concat([df1, df2])
Operations	
Add columns	df['col1'] + df['col2']
Multiply columns	df['col1'] * df['col2']
Column-wise sum	df.sum()
Row-wise mean	df.mean(axis=1)
Cumulative sum	df['col'].cumsum()
String Operations	
Convert to uppercase	df['col'].str.upper()
Check for substring	df['col'].str.contains('text')
Split strings	df['col'].str.split(' ')
Time Series	
Convert to datetime	pd.to_datetime(df['col'])
Set datetime index	df.set_index('datetime_col')
Resample time series	df.resample('D').mean()
Shift values (lag/lead)	df['col'].shift(1)
Rolling window	df.rolling(window=7).mean()
File Operations	
Read CSV	pd.read_csv('file.csv')
Write CSV	df.to_csv('file.csv', index=False)
Read Excel	pd.read_excel('file.xlsx')
Write Excel	df.to_excel('file.xlsx')
Read JSON	pd.read_json('file.json')
Read SQL Query	pd.read_sql(query, connection)