Pandas at a glance

- CLICK TO JUMP TO THE PAGE -

CHEAT SHEET
Pandas

1. Basics

Make a DataFrame	pd.DataFrame(your_data)
Make a Series	pd.Series(your_data)
Summary information	df.info()
# of rows, # of columns	df.shape
Descriptive statistics	df.describe()
First rows	df.head()
Last rows	df.tail()
All column labels	df.columns
All index labels or Index range	df.index
Check for null values	df.isnull() or df.isna()

2. Selecting

One column	df [column]
Multiple columns	df [[column1, column2,...]]
By labels	df.loc [index_label(s), column_label(s)]
By positions	df.iloc [index_position(s), column_position(s)]

3. Pandas filtering methods

Unique values in a column	df [column].unique()
Match one of multiple values	df [column].isin([value1, value2,...])
Match within a range	df [column].between(low_value, high_value)
Filter by condition	df.loc [boolean_mask] or df.query('boolean statement')
Sort and select for greatest "n" values in column	df.nlargest(n, column)
Sort and select for least "n" values in column	df.nsmallest(n, column)

4. Updating values

Remove rows with missing values	df.dropna()
Remove rows with duplicate values	df.drop_duplicates()
Alter column to have data_type	df [column].astype(data_type)
Convert column to numeric type	pd.to_numeric(df [column])
Sort rows by column's values	df.sort_values(column)
Remove row(s) or column(s) by label(s)	df.drop()
Create new column	df [new_column] = some_value(s)
Combine two DataFrame	pd.concat([df1, df2])
Assign value to specific cell(s)	df.loc [row label or condition, column] = value or df.iloc [row position, column position] = value
Assign value to column cells *not* meeting condition	df [column].where(condition, value)
Reassign many different values and / or Reassign values with pattern matching	df.replace()

5. Working with strings

Perform string method on values in column	df [column].str.method()

6. Grouping and merging

Group column1 and aggregate column2	df.groupby(column1)[column2].method()
Group column and perform many aggregates	df.groupby(column).agg()
Merge two DataFrame on specified columns	df1.merge(df2, left_on=column_df1, right_on=column_df2, how='inner'/'left'/'right')

7. Plotting

Create histogram of one column	df [column].hist()
Create bar plot of column frequency	freq = df [column].value_counts() freq.plot.bar()
Create scatter plot comparing two columns	df.plot.scatter(x=column1, y=column2)

8. Functions

Define function	def function_name(arg1, arg2,...): code code …. return output
Call function	function_name(arg1, arg2,...)