DATA SCIENCE AI

Cheat Sheet: The Most Popular Data Operations in R and Python

Description	R (tidyverse + other)	python (pandas + other)
Install Packages	`https://www.tidyverse.org/`	`https://pandas.pydata.org/`
Read a CSV file	`read_csv('filename.csv')`	`pd.read_csv('filename.csv')`
View first few rows of data	`head(data)`	`data.head()`
Summary of data	`summary(data)`	`data.describe()`
Dimensions of data	`dim(data)`	`data.shape`
Compact summary of data structure	`glimpse(data); str(data)`	`data.info()`
Names of all columns	`colnames(data); names(data)`	`data.columns`
Number of unique values in a column	`data |> summarise(n_distinct(column_name))`	`data['column_name'].nunique()`
Count unique values in each column	`data |> summarize_all(n_distinct)`	`data.nunique()`
Group number of all unique values in a column	`data |> count(column_name, sort = TRUE)`	`data['column_name'].value_counts()`
Filter rows	`filter(data, condition)`	`data.query('condition')`

Select columns or select distinct values	`select(data, col1, col2);`
	`distinct(select(data, column_name))`
		`data[['col1', 'col2']];`
		`data[['column_name']].drop_duplicates()`

Add new column	`mutate(data, new_column_name = expression)`	`data['new_column_name'] = expression`

Group data and add calculation	`data |> group_by(col1) |>`
	`summarise(new_column_name = mean(col2))`
		`data.groupby('col1') \`
		`.agg({'col2' : 'mean'})`

Sorting	`arrange(data, column_name)`	`data.sort_values(by='column_name')`
Missing values per column	`summarise_all(data, list(~sum(is.na(.))))`	`data.isnull().sum()`
Apply a function	`data |> mutate(new_col = fun(column_name))`	`data['new_col'] = data['column_name'].apply(fun)`
Join two dataframes	`left_join(data1, data2, by = 'key')`	`pd.merge(data1, data2, on='key')`
Concatenate dataframes	`bind_rows(data1, data2)`	`pd.concat([data1, data2])`
Detailed summary (skimr)	`library(skimr); skim(data)`	`import pandas_profiling; pandas_profiling.ProfileReport(data)`

Comprehensive EDA	`library(DataExplorer);`
	`create_report(data)`
		`import sweetviz as sv; report = sv.analyze(data);`
		`report.show_html()`

Plamen Rabadzhiyski