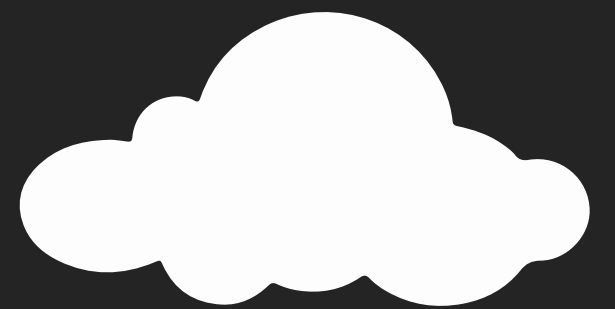# pandas dataframe manipulation cheatsheet

data exploring

tail()
helps you to show the last 'n' rows of the DataFrame

head()
helps you to show the first 'n' rows of the DataFrame

sample()
generates 'n' sample of your DataFrame

shape()
shows the dimension of the DataFrame

info()
gives you the column length, types and count of the column information of the Dataframe

describe()
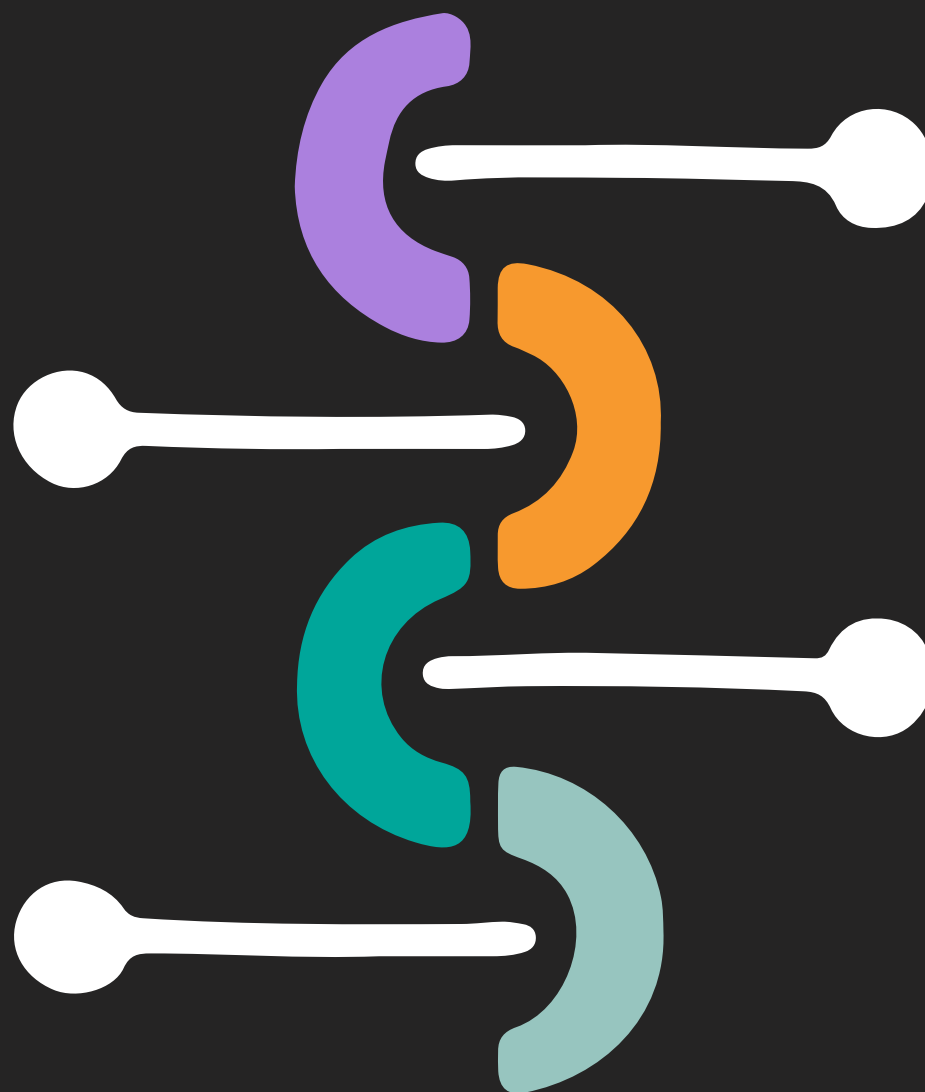gives descriptive statistics of your DataFrame: min,max, iqr,mean.

data retrieving

groupby()
it groups the DataFrame
accordingly to the
arguments

iloc()
it gets values of
specific indexes

loc()
it helps you select the row
and columns by names
and labels

sort_values()
it sorts the DataFrame ascendingly
or descendingly by a given comun

data operations

agg()
passes one or multiple functions to column/rows and renames the index of the resulting DataFrame, which can reproduce aggregare results

min()
calculates the min

max()
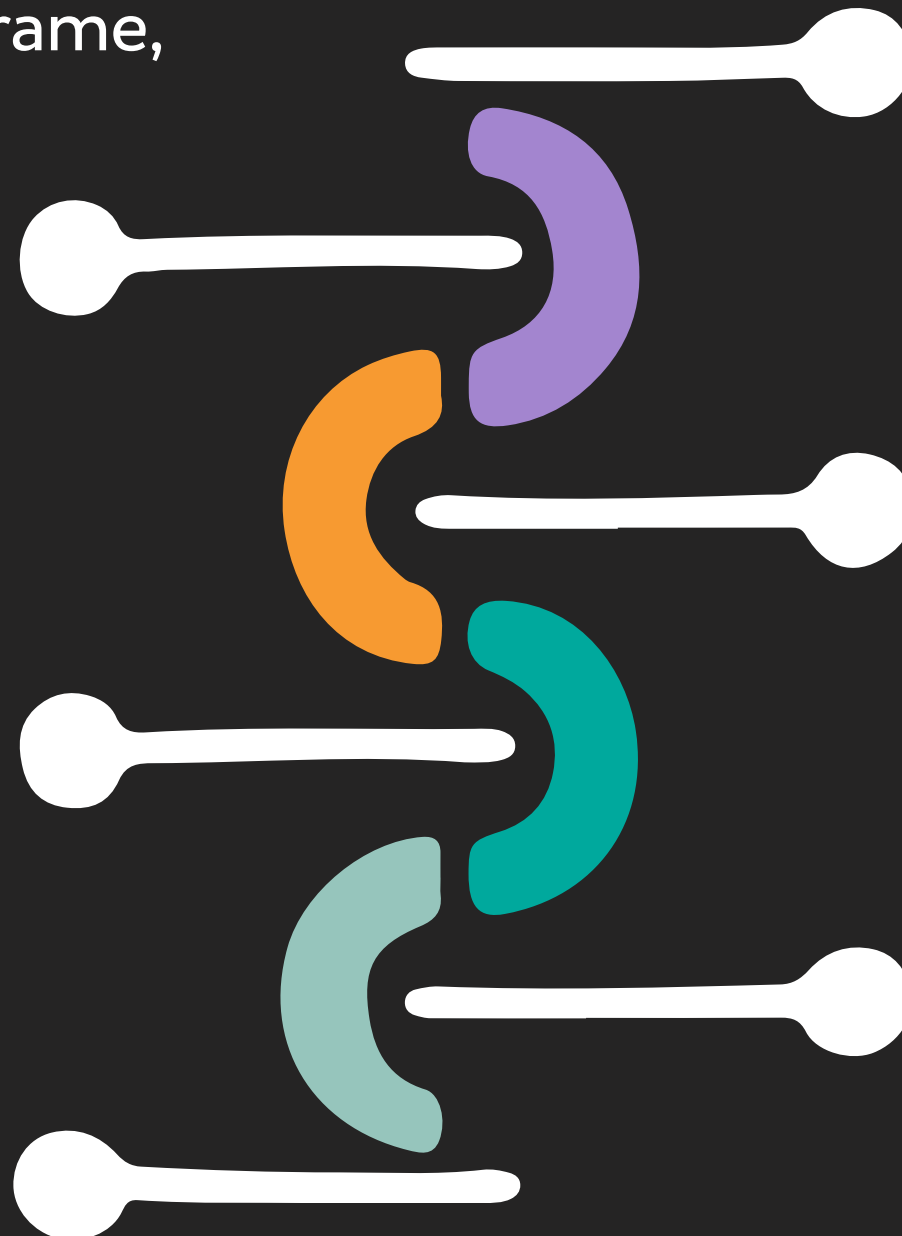calculates the max

drop()
removes rows or columns by labeling name or location (index)

transform()
allows you to broadcast function and functions to column/rows, can not produce aggregated results, works with function or function list written in a list, dictionaruy or string like

apply()
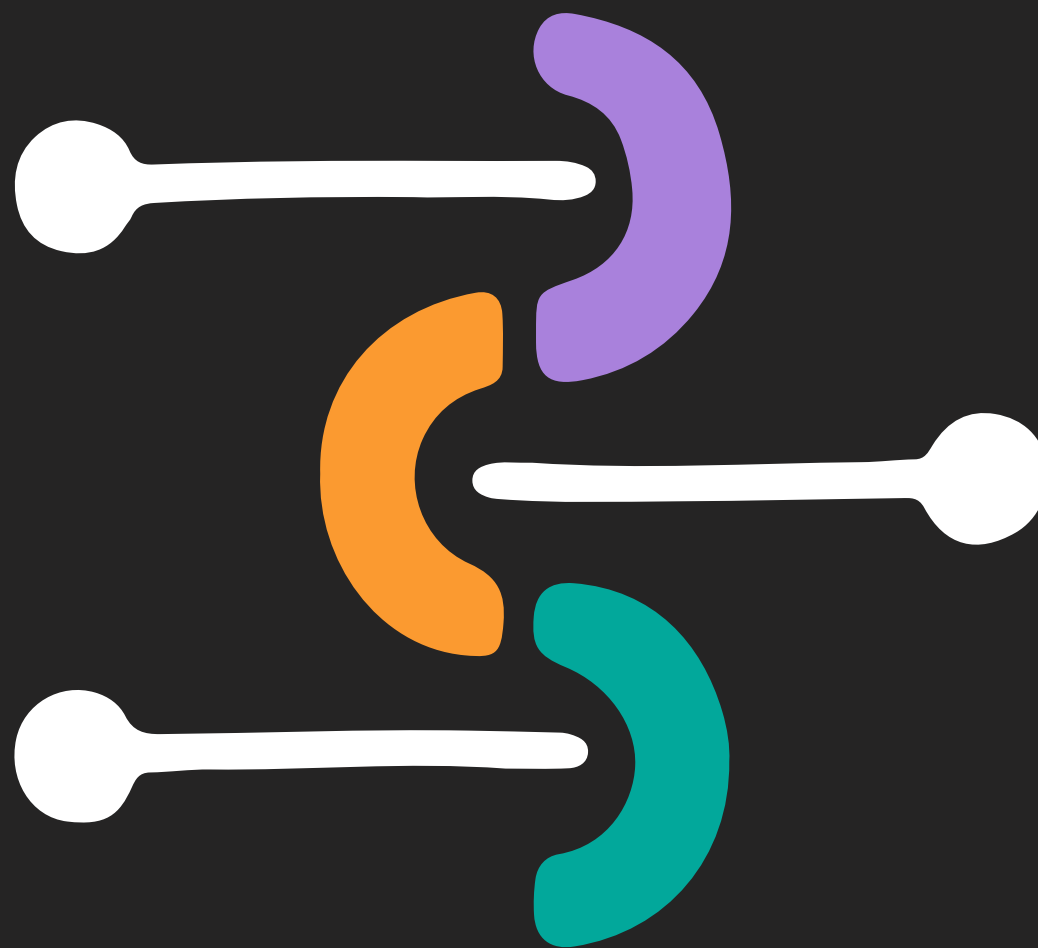performs operations on columns, rows/DataFrames, only allowed to work with functions & can produce aggregate results

data operations

drop_duplicates()
output DataFrame without
duplicate rows

not_null()
detects non-missing values
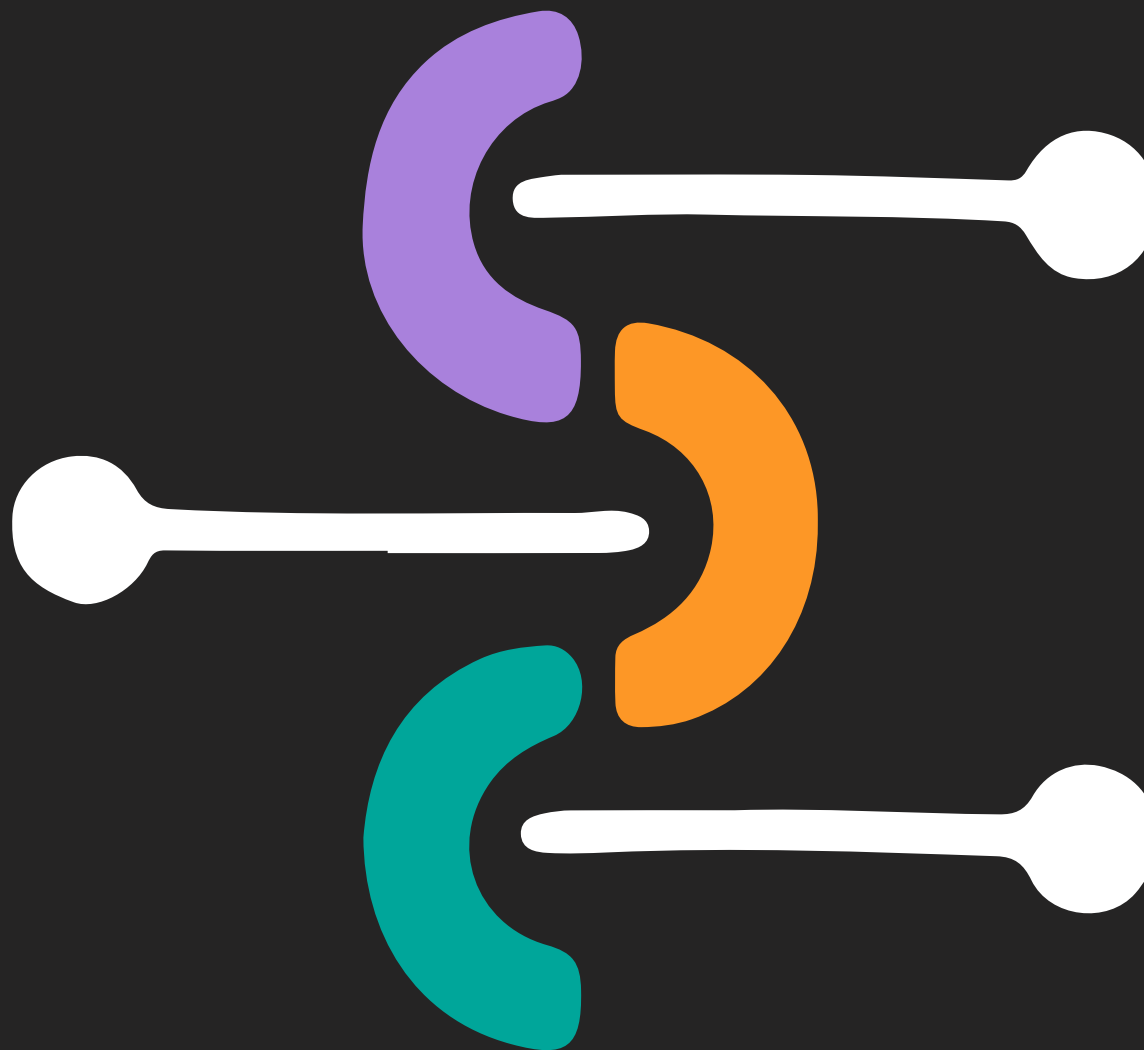and returns booleans
to show that

mean()
calculates the mean

missing value & NA's

dropna()
drops missing values
from rows or columns

isna()
returns booleans(true,false)
which indicates the value na
or not

fill(na)
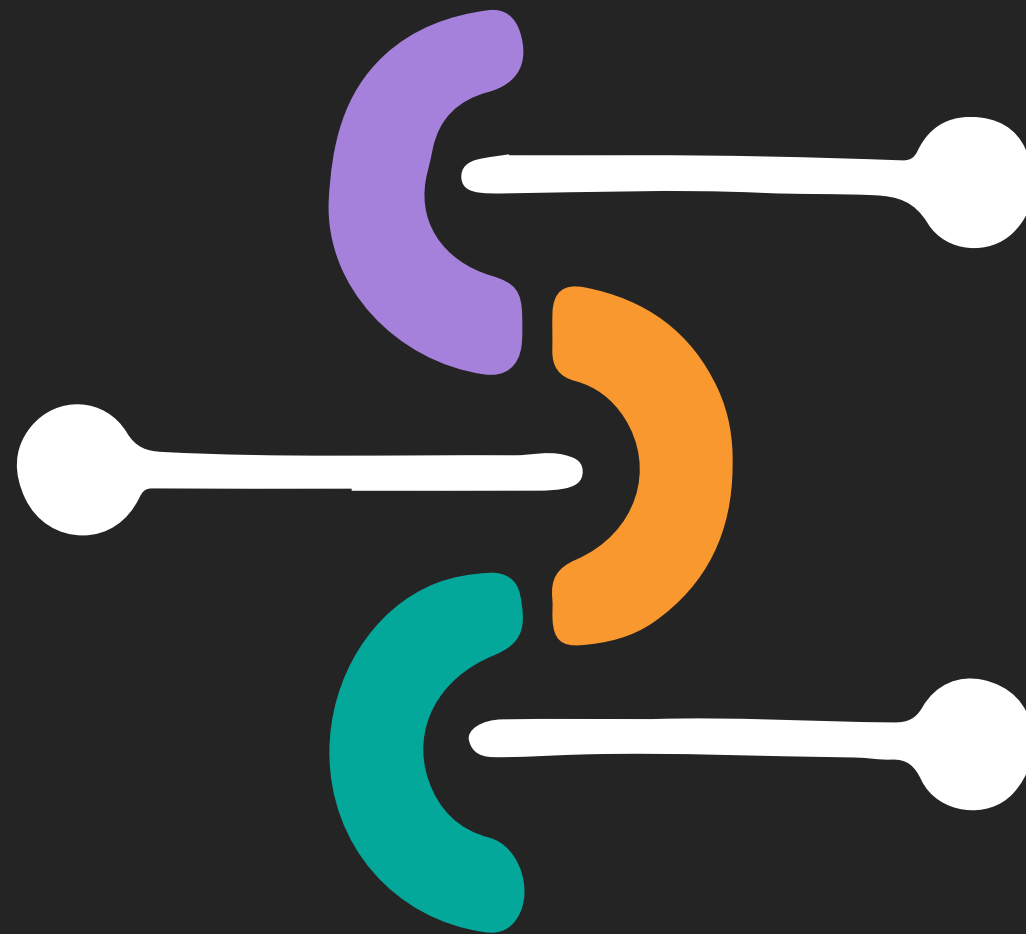assigns specific values to NA's

merging

merge()
combines two DataFrames
on columns or indices

concat()
bonds two DataFrame
across the rows or columns

join()
joins two DataFrames
on a key column or index
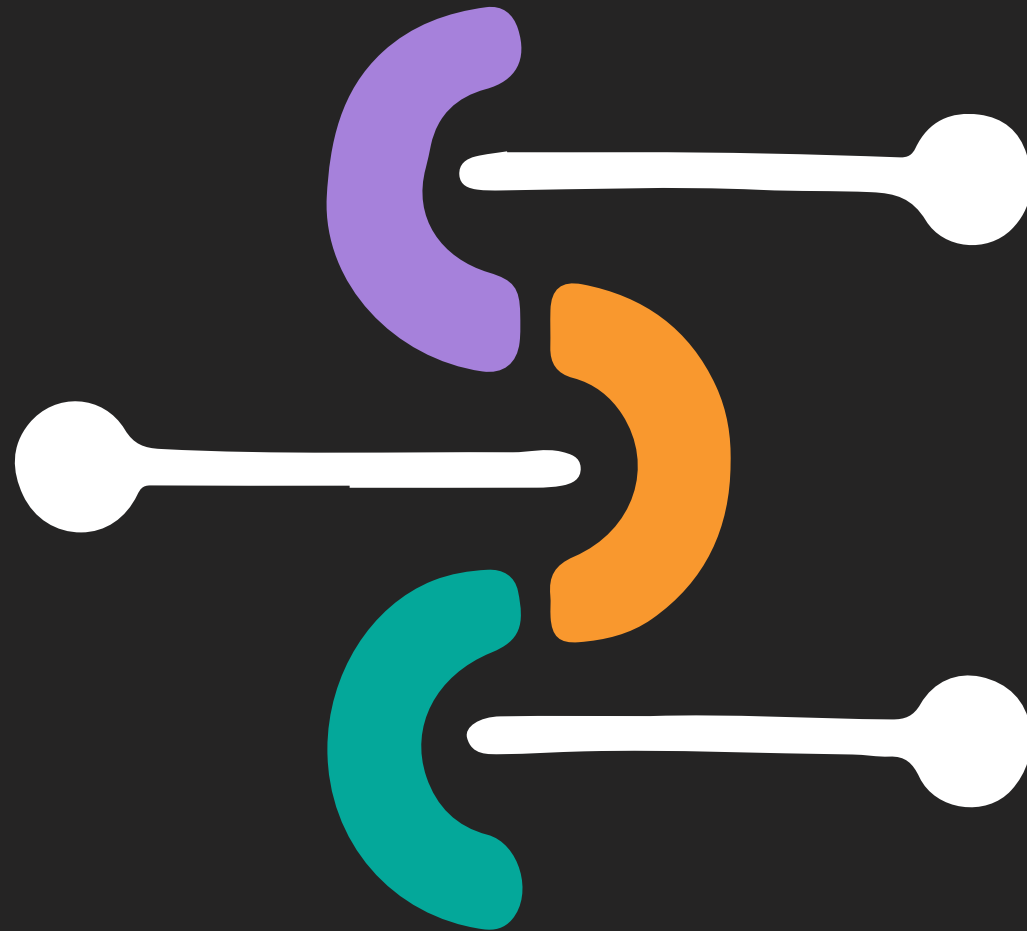
plotting

plot.bar()
to plot bar graph

plot.line()
to plot line graph

plot.scatter()
to draw scatter plot

plot.pie()
top lot pie chart

saving from and reading to DataFrame

read_sql()
To read from SQL into a DataFrame

to_sql()
To save the DataFrame into a SQL

read_excel()
To read from excel files into a DataFrame

to_excel()
To save the DataFrame into a CSV file

read_csv()
To read from CSV files into a DataFrame

to_csv()
To save the DataFrame into a CSV file

read_json()
To read from json files into a DataFrame

read_json()
To save the DataFrame into a json file

date time

dt.month()
This function returns the number of month

dt.quarter()
It selects the "n" quarter of the DataFrame.
2 means the first 6 months

to_datetime()
It changes the format of the DataFrame/series to a pandas date time object.

dt.year()
It returns the year of the date.