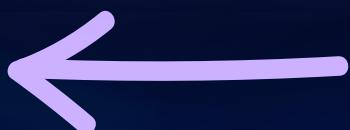




CLOUDYML



45 IMPORTANT PANDAS FUNCTIONS FOR DATA SCIENCE



**read_csv()**

This function is used to read data from a CSV file and create a pandas DataFrame object

read_excel()

This function is used to read data from an Excel file and create a pandas DataFrame object

head()

This function is used to display the first few rows of a DataFrame

tail()

This function is used to display the last few rows of a DataFrame

info()

Display information about the DataFrame, such as the data types of columns and the number of non-null values

describe()

Used to display summary statistics for numerical columns in the DataFrame

groupby()

Used to group data in the DataFrame by one or more columns, and apply aggregate functions to the groups

apply()

This function is used to apply a function to each row or column of a DataFrame

merge()

This function is used to join two or more DataFrames based on a common column



**pivot_table()**

Create a pivot table from a DataFrame, with rows and columns defined by specific columns and aggregate functions applied to the values

fillna()

This function is used to fill missing values in the DataFrame with a specified value or method

drop()

This function is used to drop specified rows or columns from the DataFrame

sort_values()

This function is used to sort the DataFrame by one or more columns

to_csv()

This function is used to write the DataFrame to a CSV file

corr()

This function is used to compute the correlation between columns in the DataFrame

astype()

This function is used to convert the data type of a column in the DataFrame

replace()

This function is used to replace specified values in the DataFrame with other values

iloc[]

This function is used to select rows and columns of the DataFrame by index



Akash Raj
Data Scientist



**loc[]**

This function is used to select rows and columns of the DataFrame by label

set_index()

This function is used to set a column as the index of the DataFrame

reset_index()

This function is used to reset the index of the DataFrame

isin()

This function is used to check if values in a column of the DataFrame are in a specified list

duplicated()

This function is used to check for duplicated rows in the DataFrame

drop_duplicates()

Used to display summary statistics for numerical columns in the DataFrame

min()

This function is used to find the minimum value in a column of the DataFrame

max()

This function is used to find the maximum value in a column of the DataFrame

sum()

This function is used to find the sum of values in a column of the DataFrame



Akash Raj
Data Scientist



**value_counts()**

This function is used to count the number of occurrences of each unique value in a column of the DataFrame

mean()

This function is used to find the mean of values in a column of the DataFrame

median()

This function is used to find the median of values in a column of the DataFrame

std()

This function is used to find the standard deviation of values in a column of the DataFrame

quantile()

This function is used to find the quantiles of values in a column of the DataFrame

to_datetime()

Used to convert a column of the DataFrame to a datetime data type

to_numeric()

This function is used to convert a column of the DataFrame to a numeric data type

set_option()

Set the display options for the DataFrame, such as the maximum number of rows and columns to display

reset_option()

This function is used to reset the display options for the DataFrame to their default values



Akash Raj
Data Scientist



**rolling()**

This function is used to compute a rolling window function over a column of the DataFrame

shift()

This function is used to shift the values in a column of the DataFrame by a specified number of rows

diff()

This function is used to compute the difference between consecutive values in a column of the DataFrame

cumsum()

This function is used to compute the cumulative sum of values in a column of the DataFrame

cumprod()

This function is used to compute the cumulative product of values in a column of the DataFrame

fillna()

This function is used to fill missing values in the DataFrame with a specified value or method

dropna()

This function is used to drop rows with missing values from the DataFrame

to_dict()

This function is used to convert the DataFrame to a dictionary

to_json()

This function is used to convert the DataFrame to JSON format



Akash Raj
Data Scientist



**LET'S TAKE
YOUR
DATA
SCIENCE
SKILLS TO
THE NEXT
LEVEL**

