

Quick Guide On Python Pandas




@MLCoding

Introduction:

Python pandas is an open-source library widely used for data analysis.

Pandas library is used for reading and manipulating data in machine learning and data science.

```
pip install pandas
```



pip command to install pandas in your system.

Pandas DataFrame:


A pandas DataFrame is a 2-dimensional data array or table with rows and columns.

Create dataframe in pandas.

```
import pandas as pd

car_dataset = {
    'cars': ['Tata', 'Maruti', 'Tesla'],
    'Model': ['Nano', 'i10', '11x3'],
    'Range': [300, 315, 400]
}

car_df = pd.DataFrame(car_dataset)
print(car_df)
```



	cars	Model	Range
0	Tata	Nano	300
1	Maruti	i10	315
2	Tesla	11x3	400

Column Operation On DataFrame:

You can easily access the data frame columns using square brackets and also assign or update new values.

Below are some basic operations you can perform on a data frame column.

```
# Accessing Single Column
print(car_df[['cars']])
# you can also use single square brackets to access single column

# Accessing Multiple Column
print(car_df[['Model', 'Range']])

# Add New Column
car_df['new_column_name'] = [1, 2, 3] # new column value

# Delete Column
car_df.drop(columns=['new_col_name'], inplace=True)

# rename column
# Syntax : df.rename(columns={'oldName': 'NewName'}, inplace=True)
car_df.rename(columns={'Model': 'model'}, inplace=True)
```

Read CSV File:

A simple way to store big data sets is to use CSV files (Comma Separated Values)

CSV files are the common file types you will use while working in Machine Learning or Data Science.

```
import pandas as pd

df = pd.read_csv('Housing.csv')
print(df)
# print(df.to_string())
# use to_string() to print the entire DataFrame.
```

	price	area	bedrooms	bathrooms	stories	mainroad	guestroom	basement	hotwaterheating	airconditioning	parking	prefarea	furnishingstatus
0	13300000	7420	4	2.0	3.0	yes	no	no	no	yes	2.0	yes	furnished
1	12250000	8960	4	4.0	4.0	yes	no	no	no	yes	3.0	no	furnished
2	12250000	9960	3	2.0	2.0	yes	no	yes	no	no	2.0	yes	semi-furnished
3	12215000	7500	4	2.0	2.0	yes	no	yes	no	yes	3.0	yes	furnished
4	11410000	7420	4	1.0	2.0	yes	yes	yes	no	yes	2.0	no	furnished

Peek Into The Data:

To understand the high-level overview of data, pandas offers multiple functions and some of them are:

```
import pandas as pd

# Read CSV File
df = pd.read_csv('Housing.csv')

#head of the data
print(df.head(10)) # print first 10 rows of dataframe

#tail of the data
print(df.tail(10)) # print last 10 rows of dataframe

#shape : To know the dimensions of the data
print(df.shape)
# (545, 13) it's means 545 rows and 13 columns

#features
print(df.columns) # it return the columns name
#Index(['price', 'area', 'bedrooms', 'bathrooms', 'stories', 'mainroad',
#       'guestroom', 'basement', 'hotwaterheating', 'airconditioning',
#       'parking', 'prefarea', 'furnishingstatus'], dtype='object')

#info
print(df.info())
# prints info about the null values and the data types of each cols.
```

Statistical Analysis Using pandas:

Pandas offer some functions that help you to dig deeper and find more useful insight from the data.

```
# describe : returns statistical measures such as min and max values, mean,
standard deviation and more.
df.describe()

# unique : return all the unique values in column.
df['columnName'].unique()

# value_count : returns the frequency of the values
df['columnName'].value_counts()

# correlation : find the correlation among the features respectively.
df.corr()
```

Pandas also have functions to find other statistical measures like mean, median, mode, etc.