PYTHON PANDAS QUICK THROUGH GUIDE

github.com/uzairafridi00
linkedin.com/in/uzair-afridi00

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

```python
import pandas as pd

car_dataset = {
    'cars': ['Toyota', 'Honda', 'Suzuki'],
    'Model': ['Gli', 'Civic', 'Swift'],
}
car_df = pd.DataFrame(car_dataset)
print(car_df)
```

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.

```python
# Accessing Single Column
print(car_df[['cars']])
# Accessing Multiple Columns
print(car_df[['cars', 'Model']])
# Add New Column
car_df['New_Column_Name'] = [1,2,3]
# Delete Column
car_df.drop(columns=['New_Column_Name'],inplace=True)
# Rename Column
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 files types you will use while working in Machine Learning or Data Science.

```python
import pandas as pd

df = pd.read_csv('Housing.csv')
print(df)
```

Peek Into The Data:

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

```python
import pandas as pd
# Read CSV File
df = pd.read_csv('Housing.csv')
# head of the data
print(df.head())
# tail of the data
print(df.tail())
# Shape: To know the dimension of data
print(df.shape)
# Return Column Names
print(df.columns)
# Info
print(df.info())
# prints info about the null values and data
# type of each column
```

Statistical Analysis Using Pandas:

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

```
# describe: returns statistics meansures like mean, min,
# max, standard deviation and more
df.describe()

# unique: returns all the unique values in column
df['ColumnName'].unique()

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

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