21-01-22 -Pandas

21 January 2022 11:37

Pandas & Numpy Contd.

```
import pandas as pd
from google.colab import files
up = files.upload()
import io
df = pd.read_csv(io.BytesIO(up['file.csv']))
print(df)
```

"This process is for Google Colab"

df = pd.read_csv( ) // creates a dataframe, 2d structure object

In Jupyter → directly use pd.read_csv('file.csv')

```
mydata = {'name': ['a', 'b', 'c'], 'cgpa': [9.5, 8.7, 9.2],
          'year': [2, 3, 3]}
```

key - value
↓ ↓
column data in column

```
df2 = pd.DataFrame(mydata)  # Dictionary to Dataframe
```

or

```
df2 = pd.DataFrame(mydata, index = ['x', 'y', 'z'])
```

Make sure to check no. of columns & rows

*) print (df.iloc[0:2])

↓ rows 0 to n-1

print (df.loc[0:2])

↓ rows 0 to n

for df2: use names in loc (index)

iloc: df2.iloc[0:2] → works

loc: df2.loc['n':'z'] → n, y, z

*) df.head( ) → by default retrieves 1st 5 rows

df.head(9) → retrieve 1st 9 rows

df.tail( ) → from last rows (last 5 rows)

df.tail(n) → last n rows

df.describe( ) → stats // only for numeric columns

df.memory_usage( ) or df.memory_usage(deep=True)

df2.value_counts('year') // value — instance

df.sort_values(by='columnname')

df.fillna(0) // fill NaN → '0'

df.fillna(0) // fill NaN → 0

or

df.fillna(0, inplace=True) → changes original object as well.

↓

w/o this only temp obj or create new

var

"inplace" solve for sort_values()

function : remove_duplicates ???

convert to numpy array

temp1 = df.to_numpy()