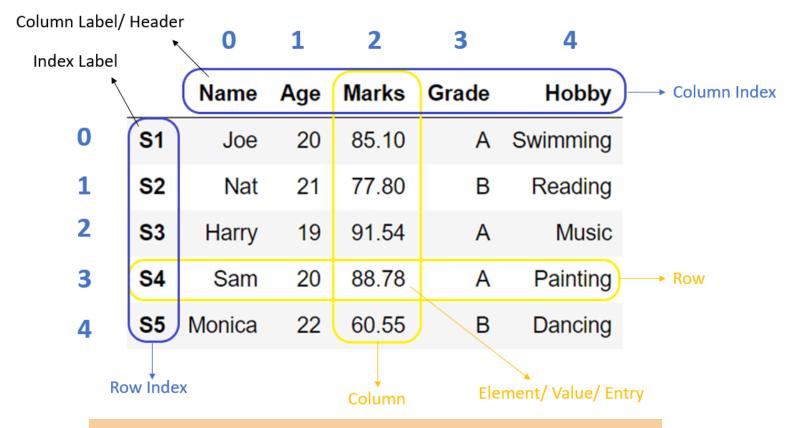# DataFrames

UTKARSH GAIKWAD

CLASS STARTING SHARP AT 3:05 PM

Topics covered today

➢ What is DataFrame?

➢ Creating DataFrame from dictionary

➢ Creating DataFrame from list

➢ Reading CSV files

➢ Data Subsetting

➢ Filter Data Based on one or more conditions

What is a DataFrame?

Pandas Library is used to perform DataFrame operations

Creating DataFrame with Dictionary

```
dct = {
        'Column1' : [a1,a2,a3,a4,…,an],
        'Column2' : [b1,b2,b3,b4,…,bn],
        .
        .
        .
        'ColumnN':[z1,z2,z3,z4,….,zn]
}
```

```
Import pandas as pd
df = pd.DataFrame(dct)
df
```

Checking Shape and columns of DataFrame

df.shape

df.columns

df.index

df.info()

Creating DataFrame with Lists

df = pd.DataFrame([l1,l2,l3,...ln]).T

df.columns = [c1,c2,c3,c4,c5]

df

Reading CSV file as a DataFrame
CSV – Comma Seperated Values

Import os

os.chdir(Filepath with front slashes)

import pandas as pd

pd.read_csv('path with forward slashes')

header = None

Name = [column_names]

Data Subsetting

➢ columns
- df[['c1', 'c2', 'c3']]
- df.iloc[]
- df.loc[]

➢ rows
- df.iloc[row_index, :]
- df.loc[row_name, :]

➢ both
- df.iloc[row_index, column_index]
- df.loc[row_name, column_name]

Difference in iloc and loc
1. iloc uses internal indexing
2. loc uses actual index names

Data Filtering Operations

df[df[column] condition]

Use and : &

Or : |

Data Filtering Example 1 : 50 Startups

➢Find out those compnies who have PROFIT more than 1.5L and are from Florida

➢Find out Data for those companies who have PROFIT between 1 to 1.5L and which are from california or florida

Data Filtering Example 2 : Cars 93

➢ Find out cars with mileage between 10 and 25 on highway

Thank you

YOU CAN ALSO PING ME ON SKYPE GROUP IF YOU HAVE ANY QUERIES