KANAV BANSAL

PANDAS DATAFRAME (ADVANCED):

→ df = pd.read_csv('data/Iris.csv')

df.head()

→ df.describe(include = 'all')

→ df.species.value_counts() (or)

→ df['species'].value_counts()

↳ Gives the unique

→ df.apply(np.max) - apply(function): applies the function on all columns and similarly for the (np.min)

→ df.set_index('Id') - Modifies the index (search query)

→ df.set_index('Id', inplace = True) - Id as index

→ df.sort_index(ascending = False) - Sort by index completely (column) - descending order.

→ df.sort_values(by='sepallengthcm').head()

↳ species remains same, sort by values

→ df['species'].unique() - unique values without valuecount.

→ df.set_index('species') - chooses the column which is unique as index

wrong syntax, practice (NOT TO BE DONE)

INDEXING AND SLICING:

→ df = pd.read_csv('data/Iris.csv')

df.head()

→ df[1:4] - slicing rows.

↱ selecting a column
→ new_df = df['sepallengthcm']

Print(type(new_df))

new_df.head()

o/p - <class 'pandas.core.series.Series'>

```
0    5.1
1    4.9
2    4.7
3    4.6
4    5.0
Name: sepallengthcm
dtype: float64
```

selecting a column - type of new_df : series, if it is passed in list it shows in table (dataframe)

selecting Multiple columns

```
new_df = df[['sepallengthcm', 'petallengthcm']]

print(type(new_df))

new_df.head()
```

O/p - <class 'pandas.core.frame.DataFrame'>

	Sepallengthcm	Petallengthcm
0	5.1	1.4
1	4.9	1.4
2	4.7	1.3
3	4.6	1.5
4	5.0	1.4

```
df['Id']
```

O/p -

0	1
1	2
...	...
149	150

Name : Id, Length : 150, dtype : int64

→ df[4]
→ df[[1,3,4]]
} Error.
- The concept of iloc & loc exists to give nparray.

ILOC - Position Based Indexing (using int indices)

LOC - Label Based Indexing

POSITION BASED INDEXING:

```
→ df = pd.read_csv('data/Iris.csv')
  df.head()

→ df.set_index('Id', inplace = True)
                    ↳ sets the Id as index
  df.head()

→ df.iloc[2]
```

O/p - SepalLengthCm 4.7
SepalWidthCm 3.2
PetalLengthCm 1.3
PetalWidthCm 0.2
Species Iris-setosa
Name: 3, dtype: object

SYNTAX:

df.iloc[default_row_index, default_column_index]

→ df.iloc[1, 2]

O/p - 1.4.

→ df.iloc[3, :]

O/p - sepallengthcm(SLC) 4.6
sepalwidthcm (SWC) 3.1
Petallengthcm(PLC) 1.5
Petalwidthcm (PWC) 0.2
Species Iris-setosa
Name : 4, dtype : object

→ df.iloc[[2, 3, 6]] - Selecting multiple rows

O/p -

Id	SLC	SWC	PLC	PWC	Species
3	4.7	3.2	1.3	0.2	Iris-setosa
4	4.6	3.1	1.5	0.2	Iris-setosa
7	4.6	3.4	1.4	0.3	Iris-setosa

→ df.iloc[[2,3,6],[1,2,3]]

O/p

Id	SWC	PLC	PWC
3	3.2	1.3	0.2
4	3.1	1.5	0.2
7	3.4	1.4	0.3

→ df.iloc[[2,3,6], :]

O/p -

Id	SLC	SWC	PLC	PWC	Species
3	4.7	3.2	1.3	0.2	Iris-setosa
4	4.6	3.1	1.5	0.2	Iris-setosa
7	4.6	3.4	1.4	0.3	Iris-setosa

→ df.iloc[3:5]

O/p -

Id	SLC	SWC	PLC	PWC	Species
4	4.6	3.1	1.5	0.2	Iris-setosa
5	5.0	3.6	1.4	0.2	Iris-setosa

→ df.iloc[:, 4]

O/p -

Id	
1	Iris-setosa
2	"
3	"
⋮	⋮
149	Iris-virginica
150	"

Name: Species, Length = 150, dtype = object

→ df.iloc[:, 2:4]

O/p -

Id	PLC	PWC
1	1.4	0.2
2	1.4	0.2
3	1.3	0.2
⋮	⋮	⋮
148	5.2	2.0
149	5.4	2.3
150	5.1	1.8

150 rows x 2 columns

LABEL BASED INDEXING:

→ df.loc[3]

```
O/p -  sepallengthcm    4.7
       Sepalwidthcm     3.2
       Petallengthcm    1.3
       Petalwidthcm     0.2
       Species          Iris-setosa
       Name : 3,  dtype : object
```

→ df.loc[2, 'petallengthcm']

O/p - 1.4

→ df.loc[4, :]

```
O/p -  sepallengthcm    4.6
       sepalwidthcm     3.1
       Petallengthcm    1.5
       Petalwidthcm     0.2
       Species          Iris-setosa
       Name : 4, dtype : object
```

→ df.loc[[3,4,7]] – selecting multiple rows.

O/P –

(197)

Id	SWC	PLC	PWC	SLC	Species
3	3.2	1.3	0.2	4.7	Iris-setosa
4	3.1	1.5	0.2	4.6	Iris-setosa
7	3.4	1.4	0.3	4.6	Iris-setosa

→ df.loc[[3,4,7], :]

O/P –

Id	SLC	SWC	PLC	PWC	Species
3	4.7	3.2	1.3	0.2	Iris-setosa
4	4.6	3.1	1.5	0.2	Iris-setosa
7	4.6	3.4	1.4	0.3	Iris-setosa

→ df.loc[[3,4,7], 'sepalwidthcm' : 'petallengthcm']

O/P –

Id	Sepalwidthcm	Petallengthcm
3	3.2	1.3
4	3.1	1.5
7	3.4	1.4

SELECTING ROWS BASED ON A CONDITION:

→ df.loc[df.sepallengthcm > 7.5, :]

O/P -

Id	SLC	SWC	PLC	PWC	Species
106	7.6	3.0	6.6	2.1	Iris-Virginica
118	7.7	3.8	6.7	2.2	"
119	7.7	2.6	6.9	2.3	"
123	7.7	2.8	6.7	2.0	"
132	7.9	3.8	6.4	2.0	"
136	7.7	3.0	6.1	2.3	"

→ df.loc[(df.sepallengthcm > 7.5) & (df.petallengthcm > 6.7), :]

condition of & → T & T - T

O/P -

Id	SLC	SWC	PLC	PWC	Species
119	7.7	2.6	6.9	2.3	Iris-Virgini-ca

→ df.loc[(df.sepallengthcm > 7.5) & (df.petallengthcm > 6.7), ['species']]

O/P -

Id	species
119	Iris-virginica

→ df.loc[df.sepallengthcm == 7.7, :]

O/p -

	Id	SLC	SWC	PLC	PWC	species
(199)	118	7.7	3.8	6.7	2.2	Iris-Virginica
	119	7.7	2.6	6.9	2.3	"
	123	7.7	2.8	6.7	2.0	"
	136	7.7	3.0	6.1	2.3	"