PAGE	NO .	-	(8
THOL	10.,	-	_

_
MA

What is Pandos?

Python library used to work with data sets.

has functions to analyze, clean, manipulating dataname has reference to panel data.

helps to analyze data and make conclusions.

con clean memy data sets and make them readable.

The data is made relevant.

what can it do? check for correlation. Aug, min, max value. delete rows.

check for not or empty values.

All this is called cleaning data.

To import

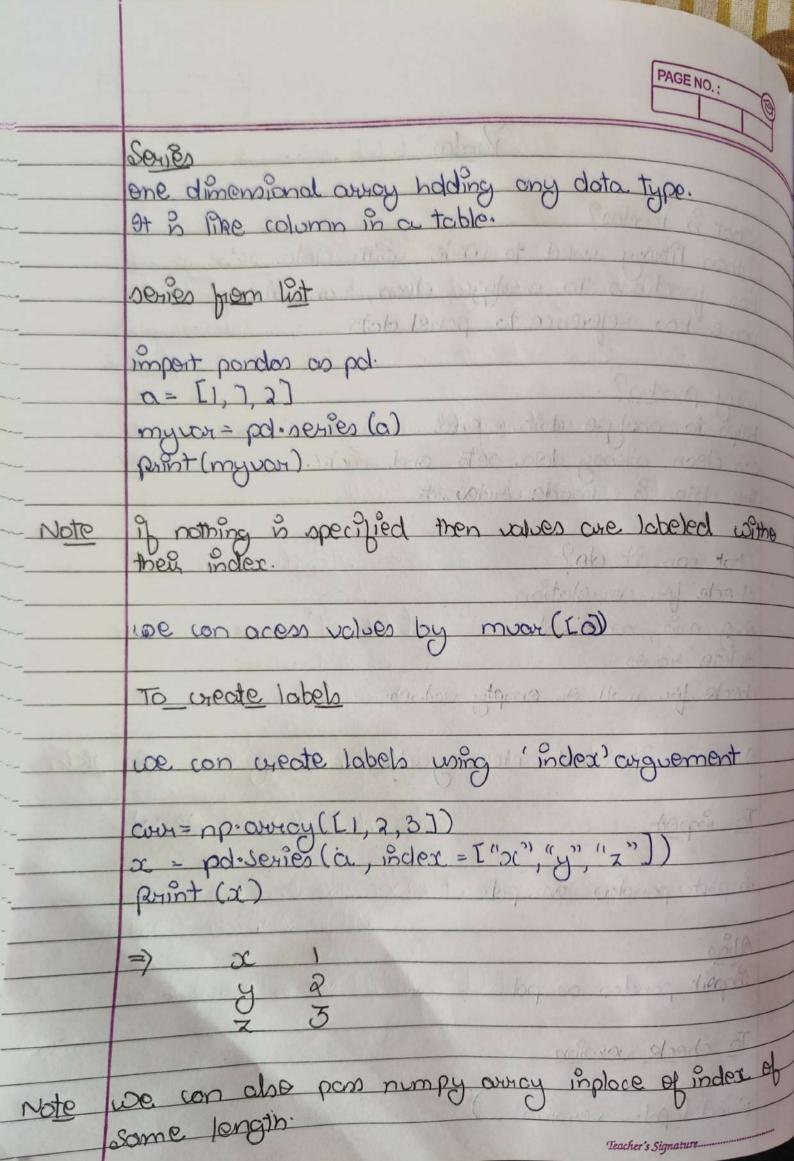
import pondos as pd.

Alias import pandas as pd.

To check version

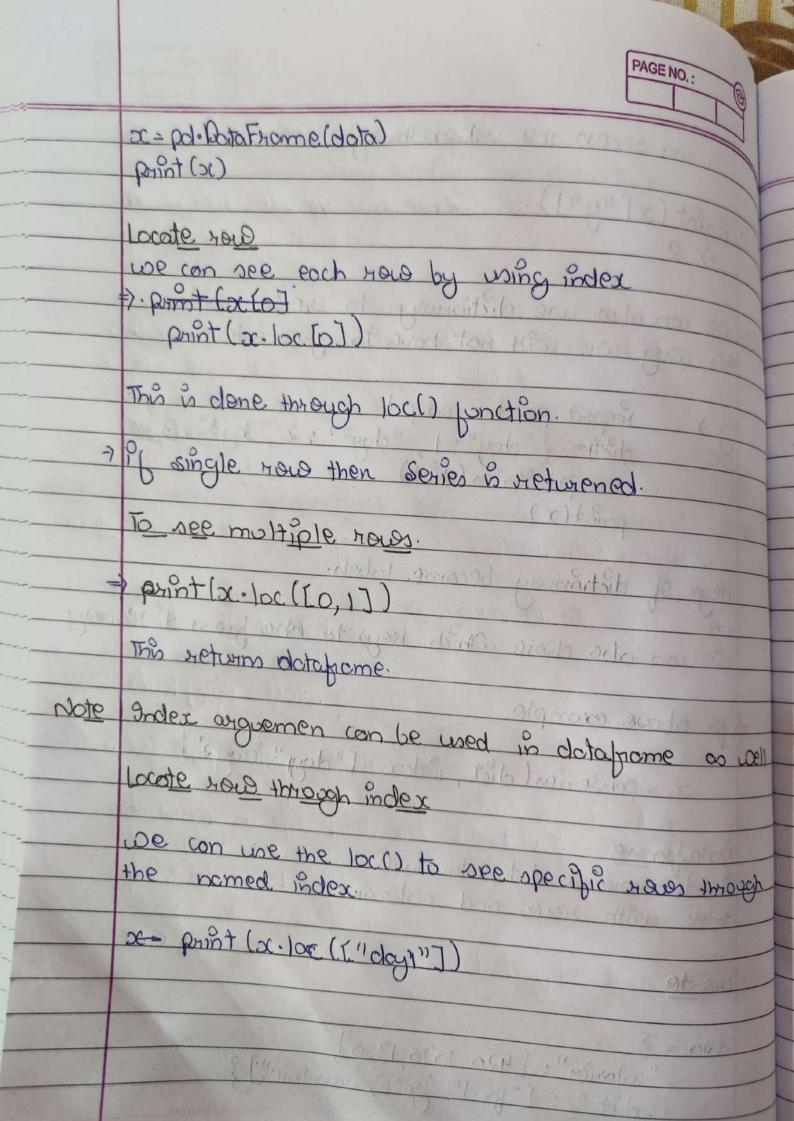
prind [pd .- version -

Teacher's Signature....



	PAGE NO.:
	we can access the values through labels.
_	Print (x["y"])
1	=) 2
	while referred ages done are not you
10	we con also use dictionary to create series.
	This way we will not have to give index.
	Eg > import pondos as pd
	arch x ary 1, ary 0 2, ary 000
	x = pd Series (dict)
	print(x)
	Light glottered acc. of
	Neys of dictionary become labels.
	we can also chose which keys to take from dictionary.
	emodotoka courtou 2011
-	Jer apone example:
6	x = pd. Series [dict, index = ["doy", "daya"]
	x = pd. Series (dict, index = ["doy", "doya"]
	X Storie Macorde Brod Storol
	Dataframe
7	à dimensional data structure
	table with vious and columns.
_	create (1994-1997)
	dota = &
	"(c)arier": [420 1090 1350]
	"caleries": [420,1090,1350], "diet": ["food", "juice", "Smoothie"]}
	John Stribolnie 13

Teacher's Signature...



PAGE NO. :
ado am mi
ne 'read-cov ()
(JB) #3/24
The self thou
etan and and
of feetings
then pandos return
was hour to
to bard to
note data
TOTC. CIONOS
1 1/0 % 12/0 % B C
not pit sale 3d
O The second of
Ar Guerland
0.00
المراجعة المراجعة المراجعة
Is loans frame
ged by the off one
mly frust and lost

Londing liles To load as files to dataframe we import pondos os pol al= pd. read_cov ('data.cov')
print (df). Note By default of dotoset is big enough To string to string of boar a application of ob = pd. read_cov ('data-cov') print (db.to-string()) => This prints whole dotar Maximum Hows Moximum nows that can be display predefined in system. 96 no of rows exceed that, then o five rows are used. To check oystems limit: =) print (pd. options. display. mose rass.

Teacher's Signusture.

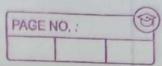
	PAGE NO.:
	we can also change this no as:-
- 63	Point (df)
	Json files Bio doto sets are altern stored a particular
-	Big data sets are often stored or extracted or Jan plain text, has format of an object.
	pd-read-json()
Note	formal as python Dictionary.
	be directly loaded to dataframe.
0	molyging Jata frames
100	ment used for quick overview is head (). print (df-head (10))
Fen	print (db-head(10))
	10-) return top lo rever including header.
	Commence de la

P	PAGE NO.:
	toil()
_	some as head.
-	Inly difference it returns lost rais.
-	only on florence is received.
	92/01)
_	This gives into about the dataset.
	Ind Order
-	and of holds and column
7	no of horas and columns.
-1	name of column with datatype.
-/	Torne 6
	Noll values
	empty values or not values can be bad analyzing
	clata.
	This helps in cleaning data.
	Clarate the second
	it means thating bad data in the data set-
0	empty cells
	data in whong formation of the come
0	wrong data
	doplicates.
	Jan wife took of he sould be sould be to the
01	cleaning empty cells
	cleaning empty cells con give whong result when analyzing data.
	o daid.
9	Remove the 4000.
	This data cleaning feature will not affect big dataset
	O brance and and an and

Teacher's Signature.....

	PAGE NO.:
	we use dropnal) to achevie this
	alp = pd. read_csv ('data: csv')
	print (new-db. to string)
•)	This removes wow with at least one mining w
Note:	Then use.
16	new of dropma (how = 'all')
->	column then.
	new of drappa (subset = ['column'])
pte	De con use mothèle colomns also.
C	b went to change original dataset then use
310-71	-> inplace = true

CONDUCTION DOLL DEPORT



Another way is to replace new volves with empty cells
This way we do not have to delete entire rows. Fillma() is used. · dj = pd. read cov ('data-cov')
dj: fillna (130, inplace = True) -) replace only for specified columns.

d[['calorier']-findina (30, inplace = True) > replace using mean(), median(), made()

we can replace with mean, median or made values. To do dition

db= pd. read cov ('data · cov') x = of ['chlorien']-meon() df [(alories']-fillna(x, inplace = True)

Some syntox or median

For mode mode returns series of values 19Re 18t. So we use x=db['cdones']-mode()(0)

This tells to use the past value in the list.

Teacher's Signature.

-2-		PAGE NO.:
	median > Average of all values. median > centure value after saiting. mode > most occurring value.	
	cleaning wrong format data make it analyze analyze are two methods.	difficult
	convert all columns in some format.	9.00
	to convert to date to derived type.	loile
	A[[, qate,] = bq. to qatetime ([, qate,])	3 manie
- C	metimes the data could just be wrong for this we use pllowing methods.	laby of
	epiace with something else.	And In
- d	o-loc[] 'Auration] = 45	3000
Note + 970/	r large dotanets use can not condition	m 191 9
F	or x in disindex: disolar Duration] > 120: disolar Duration = 120:	the state of the s

PAGE NO. ;
idedana) admini
Losses & Claren
The state of the s
William & Wood Street
Character 1
e, inplace)
y. of creek and
lition.
13501 TE TE A
of wal on
rpty values.
seid o tragito
at least 1 mining
partiala cababi
all mining values.
a ucham Tacas
enction deplicated()
h how.
I homostaking after
1 10 1 1
every duplicated rows.
ME 012
al E Objectività
THE RESERVE OF THE PARTY OF THE

il remove the rown. For a in offindex:

if offor[x, 'Duration'] >170

offiduap(x, inplace=true) drop () -> drop (index to remove This is used to drop specific rous useful when filtering on a cond drepna() -> drepna() drops rows and columns with en dropna (axis=0, how='any') any-> drops rows or columns with

all 7 drops nows or columns with

cleaning Puplicates To discover duplicates use use This victures a boolean value for eac True if duplicate.

df-duplicated () -> return true per

Hemouing (drop_duplicatest) se use

1) df. drop duplicates (inplace = True)

Bordas Correlations court) method adalate relationship between each docour() Note: This function ignores non numeric column. execult of cours is a table with a lot of number of varies from -1 to 1. 1 -) perject correlation. at least o-bor higher is good correlation Pandas plotting
we can use plot() to make diagrams. import pandon on pd import matphothis b. pyplot au pit db= pd. read_cov('data.cov')

di-plot() -) To plot

evologe () Grans. +19

	PAGE NO. ;
The second secon	scatter plot needs a a and y axis. con be specified wind kind='scatter'
	ph-plot (kind=(scatter), oc='puration', y='calorier')
	y = dependent variable.
	histogram (Cyliff) margatain
	kind=hint la source of standing to get standin
	9t needs only one column!
	df ['duration'] - plot (Rind = hint) 1891
	Take arguments 13/20
	bins = 10-7 Tells no of bins color = shyblue color of bors. etleccolor = black
	To convert categorical to numerical
	pd-get_dumnier() -) one hot encoding.
	return different columns on based of no of different

Teacher's Signature....

