

Assignment NO-3.

Page I	Vo.		
Date			



	- Tondard Nonderell
	Title - Descriptive Stastics. Measures of central
	Tendency and variability.
	The state of the s
	problem statement - perform the following operations
	on any open source dataset (e.g. data.csv)
	1) provide summary statistics (mean, median, mode
	minimum, maximum standard deviation) for a
	dataset (age, income etc)
	* (control Tendency - period britan)
THE POWER	e) write a python program to display some basic
coin blac	Statistical details like percentile, mean, standared
	deviation etc.
	objective - To analyze and demonastrate knowledge
	of statistical data analysis techniques
	for decision making
	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
	PREREQUISITE -
e1.	1) Basic of python programming.
1000	2) Concept of Stastistic mean, median, minimum,
Sappina	
+++	atherintals in reclaims for all ptols eight
34/	THEORY - Projected Disingle Daniered Soul
	The data are summarized in some, but not
1-1	all ways we close descriptive that are
	either most often reported on most often
	covered in introductory courses. There as
E hizhen	follows:
20	similar to table rous but the column

	6.0	
700	بالحاث	door
- Married	NAAC	

Page No. Date



	· Central Tendency
	Mean
	median
	mode
	· Dispersion
	Standard deviation (Std. deviation)
	minimum
	maximum
	Company of Send State of State
	· Central Tendency
	The mean median and mode are the three measures
	of central Tendancy. Mean is the anithmetic average
	of data set.
abole sold	The mean
- Jaumi	The state of the s
	N. M. Landon Ingrainah mil
	Theory-
	1. Introduction to big data.
	Big data means really a big data, it is
uminim.	collection of large datasets that cannot be
	processed using traditional computing techniques
	Big data is not merelize a data, rather it
	has become complete subject which involve
7 +114	various tools, techniques & forame work.
250	Big data involve the data produce by
more	different devices & applications.
- 20	2. Introduction to dataset.
	A dataset is a collection of records,
	Similar to table rows, but the columns.
- 1	

NAAC
100000

Page I	Vo.		
Date			



7777	Varale Campus
	can contain not only strings or numbers
The gran	but also nested data structures such as lists,
	map & other records
	3 python libraries for data science.
2.50	a Mumpy
5/320	one of the most fundamental packages in python
in reason 120	Numpy is general purpose array processing package
3 30	It provides high performance multidimensional
Jeso Heng	corray object.
mort.	o what can you do with Numpy ?
Hon 92	3 Basic array operations! add, Multiply,
	Blice, Flatten, restrape, index arrays.
	2. work with datetime on linear algebra
afozote!	3. Basic Slicing & advanced indexing in numpy
yes Histogen	python and shootman feet technology of
rodly	b. Pandas : habit ar all The Astronomy and the same and t
e the	pandas is an open source python packages
	that provides high performance, easy to use data
	structure and data analysis.
Wissaland and	O what can you do with Pandas?
	17 Indexing, manipulating, renaming, sorting, menging
eaulay 1	data. frame.
column	to contar priestre to source to parei
	3 Scikit learn:
oitasilemma	Introduced to the world as google Summer of
acifeeran	code project, scikit learn is a robust machine
hos	Learning library For python.
	o what can you do with scikit learn &
	2. classification: spam detection, image recognition.
Maria Carlo	

NAAC

Page I	Vo.		
Date			



	Varale Campus
emd	2 clustering - Drug responce istack price
etail on	3. Regression - customer segmentation, grouping.
	experiment outcomes:
	sometice total and parameter ability of
	4. Description of dataset. The Iris dataset was
Hrq.01.2	used in R.A Pishers classic
down pai	1936 paper. The use of multiple measurements
Landen	in taxonomic problems & can also be
	Found on the UCI machine Learning repository
	One Plower species is linearly separable from
4	the other two, but the other two are not
	linearly separable from each other.
br	selo menit so emitatale disco disco se
rami	5. panda dataframe functions for Load datasets.
	1. The dataset is downloads from UCI repository
	a. Now read CSV file as dataframe in python
2980700	From path where you saved the same the
otob es	Iris dataset is stored in . (sv format.
	Variable and population automos
	G. Panda data frame functions for data preprocessing
Mengin	O cheaking of missing values in dataset.
	is null () cheak missing values or null values
	is nace -> count of missing values of column
	& now wise count.
10 1001	7. panda functions for data formatting & normalization
nachine	The Transforming data stage is about convertion
	the dataset into format that can be analyzed
	or modelled effectively.
Told innovan	eponi contotale regar detection image
Seite State State	

	1000
	0
3	NAAG
	A STATE OF THE PARTY OF THE PAR

Page	No.		
Date			



	varate dampus
	8. panda functions for handling rategorical variables
7	- rategorial variables have values that describe
100	o quality or 'characteristic' of data unit, life
	what type or which category.
	· label encoding - label Ancoding refors to conventing
	the labels into a numeric form so as to
	convert them into the machine - readable form.
	Height Height
	Tall
	medium
	Short 2
	Algorithm.
	Step 1: Import pandas & Sklearn library for
	preprocessin q
	stepa: Load the inis dataset in dataframe
	Object df
	Steps: observe the unique values for the
	Species column.
	Step 4: pefine label encoder object knows
	how to understand word labels
	steps: Encode labels in column 'species'
	steps: observe the unique values for the
	Species column.

NAAC	Page No. Date Dr.D.Y.PATIL EDUCATIONAL FIDERATION Varale Campus
defent de	Conclusion -
(saintest	In this way we have explored the
9271 3316	function of the python library for data
	preprocessing data wrangling techniques & how
a Francisco	to handle missing values on it is dataset.
012	the mich shammy a stail cladal adh
is someth	count - them into the proching - readable
	thoist Height
	Tall of the second of the seco
	medium
	THE TOTAL OF THE T
	trod2
107	\$10d2
107	trod2
Fore	read? read? read? read? read: re
For e	readil arrelation and trade

