Page No. Date. DSBDAL Assignment 3 · Roll no: 31457 Norme: - Jush Shoh · Completion date: 1/02/22 · Submission date: 1/02/22 · <u>Jitle</u>: - Descriptive statistics - Measure of Central Jendency · Problem Statement: Perform the following operations on any open-I Provide summary starties for a deteret with numeric variable grouped by one of the quantitiese Variable. 2. Write a python program to display some basec Startical details like percentile, mean, st & devoidion etc. of the species of 'Iris-setosa', Iris-Versicoloi, and 'Iris Verginica' of iris detaset Jearning objectives: I We will be able to lood dataset and perform all basic preprocessing on the dataset. 2. He will be able to group date and find the grouped startias for a variable. · Larning Outcomes; Dearnt about grouping data according to a category and find basic statues statistics of the grouped data. 2. Learnt about mean, median, It deviction percentile etc.

	Date.
7.0	Theory:
- 1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 - 1	- Stastics
-	1) Statistics:
D. Print Land	It is the science of collecting date and
	analyzing them to infor proportions (sample) that are representative of the population in other words, starts statistics is interpreting
	that are regressiation of the population. In
and the same	other wordy starting statistics is interpreting
	data in order to make predictions for
A	the population.
11	AL AND A STATE OF THE PARTY OF
	-Branches of statistics
	Control of the contro
	There are two branches of statistics:
1	· Descriptive statistics.
Control of Control	· Descriptive statistics. • Inferential statistics.
	Descriptive statuties:- It is summarizing the data alt hand through certain numbers like mean, median
	It is summarizing the data of band
	through certain numbers like mean median
	etc so as to make the understanding
1.1	of the data earlier It does not invoke
	any generalization or inference beyond what is available. This means that it
	what is available. This means that
	description of the data available and not base on any theory of probability.
	- entation of the data available and act have
	on any theory of probability.
T.	The state of the s
	- Commonly Used Measures
down	Late 1 The transfer of the control o
	1. Measures of Central Jerdency 2. Measure of Dispersion (or Vaciability)
	2. Measure of Dispersion (or Variability)

Page No.

	Date.
	₹Max
	-Measures of central Jendenay. A measure of central tendency is a one number summary of the data that typically describes—the center of the data. These one number summary is of three types:
	measures of central tendency is a one number
	a the data that typically derribes the
-	tenter of the data. These one number summaky
	thee types:-
	is of three types:- 1. Mean: It is defined as the ratio of the sum of all the day observations. This is also known as Average.
7	sum of all the dy deservations. This is also
	known as Average.
11-14-1-	2 44 13
	2. Median: It is the point which divides the
	entire date ento two equal halves one half
	the class is less than the median,
	Some. Some.
	Same.
	3-Mode: - It is the number which has the
	3-Mode: - It is the number which has the maximum frequency in the entire data set
- tip	
	Measures of Dispersion (or Variability) Measures of Dispersion describes the spread of the date around the central value.
	Heatures of Dispession discribes the spread of
	the dela around the central value.
	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
	1. Absolute peviation from Mean - Ite It is also called as Mean absolute deviation, describes
	called as men about abunden, describes
	the variation in the data set, in sense
	that it tells the overage absolute distance
	of each data point in the set.
	Man I N IV. VI
	$MAD = \frac{1}{N} \frac{N}{i=1} X_i - X_i $
10	
	2. Variance: - I measures now for are data points
- Al	2. Variance: It measures how for are data points spread out from the mean.
	Variance: 1 \$ (2:-2)

Page No.

	Page No. Date.
,	3. Standard Deviation. The square root of Variance is called the standard Deviation.
A.C. Shi	Traffic Asia Asia Asia Asia Asia Asia Asia Asia
C 40 0	Std deviation = \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \
	& Quartiles: - quartiles are the points in the
	that divides the data set who he
	equal parts. Q1, Q2, Q2 at the 1st 2rd and
	equal parts. Q1, Q2, Q2 att the 1st 2nd and 3rd quartite of deta set. Q1 02 Q2
100	Q1 62 Q3
777	251. 251. 251. 251.
E LAS	bear with the did with the
adl	Interquatele ronge
	Interquartele ronge = R3-Q,
T. L.	2) Pandas groups ;
AUG	It is used for grouping the data
	It is used for grouping the data according to the calegories. It also helps to aggregate data efficiently.
	to aggregate data efficiently
1 - 13	The state of the s
	A. I C. t.
BA	steps and Functions:
1 - 1 - 1 - 1	1 Ata Bragania America to 1
	1. Data fre processing: - process of transforming randotate into an understandable format. functions: - 1) read-csv() 2) head() 3) is null() 4) describe() 5) fill na 6) astype 7) value counts.
Asile	(notice) is 1) control (1) () () local () () ii ()
	() devoite () 5) lill as to extreme 7) value (
	4) abortbel) a) for he of waype) value counts.
	2 calculations around startistics
	2. Calculating grouped startistics Part 1:-
	1. Age is converted to age groups by highing
lake !	1. Age is converted to age groups by birning method.

	Page No. Date.
	2. grouped the data according to the age group and & using groupby! function And then calculated the statistics like mean, median related to salary using describe() function
	Part 2:-
	1. The data is grouped according to the 3 species using grouply and then found out all besic statistics.
•	conclusion: - successfully calculated the des measures of central tendency and measure of dispersion for grouped data for both the datasets
1 ,	