Assignment No.10

Page No.

Hame - kuldharan Sumit Dattatraya

class - TE

Division - Div4

Subject - DSBDAL

problem statement -

into a Data Frame (e.g. https:// archive.ics.uci.edu/mi/datasets/Inis). Scan the dataset & give the inference as:

- 1. List down the features of their types (e.g. numeric & nominal) available in the dataset.
- 1. Create a histogram for each feature in the dataset to illustrate the feature distributions.
- 3. create a box plot for each feature in the dataset.
- 4. Compare distributions & identify outliers.

Theory-

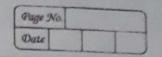
- 1) How to deal with outliers of various dependent f independent Variables in regression analysis &
- such as types.

Different methods to detect outliers

univariate - boxplot outside 15 inter-vartile range

Bivariate - with the help of scatterplot

Multivariate - Mahavanobis D2 distance



ways to finding outliers -

- -sorting method
- data visualization method
- Statistical tests (z-score)
- Interquartile range method.

Pros of removing outlier -

- the number is clearly on unitentional error
- the number is an intentional error
- the outlier comes from a different population

cons of removing outlier -

- the outlier is a legitimate observation from your desired population
- I chopped an outlier of looked at my data again
 grow a new outlier popped up.
- making a definitive rule can be hard to defend.

Dem als		-
Page No.		
Date		
MARKET		

Why are outliers to be treated carefully-

decreases statistical power.

What casuse outliers-

Extranse due to changes in system behaviour froduent behaviour, human error, instrument error or simply through natural deviations in populations.

How do you determine the distribution fits my data bestprobability plats might be the best coay to determine
conether your data follow a particular distribution. If your
data follow the straight line on the graph.

can be distribution be normal if it has outliers?
Yes the normal distribution data can have outliers.

Types of Distribution -

- Direct Distribution
- Indirect Distribution
- Intensive Distribution
- Exclusive Distribution
- selective Distribution

why pistribution is important:

Distribution are impostant for statistics because use need to collect the sample of estimate the parameters of the population distribution. Hence distribution is necessary to make inferences about the overall population.

Conclusion -

Hence, we learned about data visualization fourtier.