31115

Assignment 2

Vedant Bothikar

Data Wrangling 2

Problem Statement: create an academic performance dataset of student & perform the foll. operations using python

(1) Scan all variables for Missing values & inconsistencies. If there are Missing values, use any technique to deal with them.

(2) Scan all numeric variable for outliers. If there are any suitable techniques, use them.

(3) Apply data transformation for atleast one of the variables.

* Objectives:
  1. Find null values & deal with them
  2. Scan outliers
  3. Apply data transformation

* Software & Hardware Requirements
  Win 10 Ram 8GB 64 bit OS
  python jupyter notebook

* Theory:
The activity of taking input from data in its original state to a format where we can perform meaningful analysis on it is called data wrangling

Missing values: In real life many datasets will have many missing values.

In order to find missing values we can use python's .isnull() function.

Identify missing values & handle them:

(1) Replace missing values with some value

(2) Drop missing values:

Outliers - Data points which differ significantly from other observations.

Transformation:

Standardization - It doesn't have any fixed minimum or maximum value. Here the values of all the column are scaled in such a way that they all have mean = 0 & Std. dev. = 1.

* Conclusion: We have successfully performed data wrangling such as handling missing values & outliers & transform the given data.