Manual for using Students Performance in Exam csv file

Link: no link for this file, please download from Github Repository
Key: there is no id

1.0 Import of libraries and the data set

Ensure that the csv file is in the correct working directory first. If csv file is in the right directory, then input the filename of the csv file that is to be used. Alternatively, the csv file can also be loaded from a URL if you choose not to download the csv file into the directory.

2.0 Cleaning Data Set

2.2 Removing columns

There's no need to remove any columns.

2.4 Transforming data types

There's no need to change any data types.

2.6 Remove NA values

Remove all null values.

2.7 Renaming columns

Rename "change/ethnicity" to "ethnicityGroup"
Rename "parental level of education" to "parentalEducation"
Rename "test preparation course" to "prepCourse"

2.8 Check for duplicates

As there is no key there is no reason to check for duplicates.
Thus, enter no.

2.9 Possibility for repeating functions

You can repeat any functions you like.

3.0 Univariate Analysis

3.0.1 Split numerical and categorical data into two lists

To get correct categorical grid plots and avoid errors we have to remove any categorical variables that have to many subgroups such as key values. The program will ask you which columns these are. As we have no such column enter "none".

3.1 Categorical Data analysis

3.1.2 Value counts

After inspecting your grid plot you can look at a single variable. Enter which one you want to inspect

3.1.3 Ask for repeating single analysis

You can repeat the single analysis as much as you want to.

3.2 Numerical Data analysis

3.2.1 Distribution grid plot

The grid plot will be completely automatic. Inspect which columns you find interesting to analyse more specifically.

3.2.2 Give option for individual analysis

Choose which numerical variables you want to inspect in specific.
Try e.g. price, availability_365 and longitude.
Multiple choices are possible.
The program will give you information about the following domains:

- 3.2.3 Descriptive summary
- 3.2.4 Inspecting Skew & Kurtosis
- 3.2.5 Count outliers
- 3.2.6 Plotting boxplot

3.2.7 Ask for repeating single analysis

You can repeat this sequence as much as you like.

3.2.8 Correlation Matrix

In the end, the program will present you a correlation plot for you to examine.