Manual for using Data Analytics Tool for StarCraft2ReplayAnalysis-1 csv File

Link: http://summit.sfu.ca/system/files/iritems1/13328/StarCraft2ReplayAnalysis-1.csv
Key: GameID

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
Change "GameID" and "LeagueIndex" to character.

2.5 Count number of null values
Displays the number of empty values in each column.

2.6 Remove NA values
Remove all null values.

2.7 Renaming columns
Rename any columns you wish.

2.8 Check for duplicates
Enter the key "GameID" and check for duplicates.
You should remove any duplicates you can find.
There are no duplicates.

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 too many subgroups such as key values. The program will ask you which columns these are.
Enter "GameID".

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
Enter e.g. "LeagueIndex"

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.

3.2.7 Give option for individual analysis
Choose which numerical variables you want to inspect in specific.
Multiple choices are possible.
Enter e.g. "Age", "APM", "HoursPerWeek"

The program will give you information about the following domains:
>3.2.2 Descriptive summary
>3.2.3 Inspecting Skew & Kurtosis
>3.2.4 Count outliers
>3.2.5 Plotting boxplot
>3.2.6 Ask for repeating single analysis

You can repeat this sequence as much as you like with multiple or single values.
Enter e.g. "MaxTimeStamp"

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