Use case for TV shows rating

Here is a use case for practice and you need to implement following:

1. Descriptive Test
2. Summarize Data
3. Check Correlation between any numerical values
4. Perform Linear Regression
5. Provide your diagnosis post Linear Regression

# Examine data in detail for specific techniques to be investigated

# Data Cleaning

1. Discover any corrupt or missing data and think of various data cleaning operations to perform marking or removing bad data and imputing missing data

# Data Transformation

1. Discover some attributes that have familiar distributions such as Gaussian .Summarize Data in R with Descriptive Statistics

Procedure to be followed:

1. Make quick observations on raw form of sample data
2. Data summary
   * 1. Min
     2. 25th Percentile
     3. Median
     4. Mean
     5. 75th Percentile
     6. Max
3. Standard Deviations(to find out gaussian or nearly gaussian distribution)
4. Checking for Skewness in Data.
5. Finding out the correlations in given dataset.
6. Categorical data (Frequencies/Crosstabs)

# Perform Visualization and Exploration

* + 1. Histograms
    2. Scatter Plots

# Predict rating of shows based on given dataset.

Perform Linear Regression to predict user\_rating\_score based upon rating\_description , user\_rating\_size and Number\_of\_Viewers.

Use lm() function in R as (y ~ x1+x2+x3) [Hint]

Please make a word doc with screen shot and your findings.