1. Undertake exploratory data analysis and produce a 2-page report on your findings/insights

2. Clean the data and produce good visualization for storytelling

3. Submit your code and report via the Github repo

**Shape of the data:**

* 8807 - Rows
* 12 columns

**Columns with missing values**

1. director -2634
2. cast -825
3. country - 831
4. date added- 10
5. rating - 4
6. duration - 3

The duration of the 3 columns that were missing, was wrongly entered in the rating column. Pasting the three values to the duration column, will make it complete without null values.

The missing values in the ‘date added’ column was filled with the date in the neighbouring row thereby making the column without null values

The columns director, cast were dropped.

The null values in rest of the columns (country, rating) were removed from the dataset.

**Duration has both units of measurement**

1. minutes

2. No. of seasons

Tv show titled Burned cocoon was given a duration of 84 min which is wrong. It could be seasons. Changed that value to 3 Seasons.