**Content Data Set:**

-1000 - records in table

**Content ID column**:

1000 – records unique

**User ID Column:**

- 134 green unique records

- 866 – red – duplicates

- 4 - Type of content

– 16- Type of category

**URL Column:**

– 199 missing values

– 801 unique values

**Cleaning Content Data Set:**

- 34 - “ - deleted

- Animals to animals replaced.

- Studying to studying replaced.

No duplicate values found.

URL column was deleted

**Reactions1 Data set:**

All records in table – 24573 \*starts from 0

**Content ID Column:**

– 24556 Red duplicates

– 17 Green unique

**User ID Column:**

-2039 – missing values – N/A

-22534 – Red duplicates

- no unique value

**Type of Reactions Column:**

**-** 16 - Type of reactions

- 980 – missing values. All of them matched with missing User ID values – was deleted

- 24573 – Cells with Values

**Datetime column:**

- Values for 2020-2021 years. Contains years, month, days, hours, minutes.

- 22 duplicates. Don’t matched with User ID, Content ID.

- 25531 – Green unique values

**ReactionTypes Data Set:**

- 16 - Type of reaction

- 3 types of Sentiment – positive, negative, neutral

- Score - from 0 to 75

- No duplicates or missing values

**Modeled Date Set**

- 24573 – Cells with Values

- No duplicates or missing values