

DATAVISTA – DASHBOARD INFORMATION

Team Name: Sparks

Team Members:

1. AAGHASH M
2. SANJEY GM
3. ADITYA KRISHNA RS

College: Sri Krishna College of Technology

Null Values:

1. Rating_count :

- There were **2 null values** in the 'Rating_count' Column. We went to the link of one product and filled the value. For the second one, the link was not working, so we removed that row

Feature Engineering:

1. Discounted_price and actual_price:

- The values were in the **string** form (ex. **â, '3,999**). We extracted the numerical values and converted the column to **Int** with values like (ex. **3999**)

2. Rating:

- The rating column was initially **object** datatype. We converted it to **float datatype**

3. Rating_count:

- The Rating_count values had a **comma** in between them (ex. 7,999). We extracted int values alone and converted the datatype of the column to Int (ex. 7999)

4. Broad Category:

- Since '**Category**' had multiple combinations of values, we've extracted the first individual category from the set in each row and created a new column named '**Broad Category**' which holds a single category value for each entry

5. Review_Goodness_Score:

- We used the VADER sentiment intensity analyzer to analyze the Reviews and measure how positive the reviews are. This measure is stored in the 'Review_Goodness_Score' Column

6. Review_Category:

- This column is based on the 'Review_Goodness_Score' Column. if the value of 'Review_Goodness_Score' is greater than 0.5, This value will be 'good', if it is between -0.5 to 0.5, This value will be 'avg' or else ,it is 'bad'.