

FINAL REPORT

INDIAN PRODUCT REVIEW ANALYSIS AND RATING PREDICTION

PROBLEM STATEMENT

The campaign of “**Vocal for Local**” has seen tremendous support from the people of India, however for the campaign to sustain it is important that the Indian Brands are able to meet the quality of the products in demand by the consumers.

By using the “**Indian Product Reviews**” dataset of the Amazon we will try to gather insights if the consumers are happy and satisfied by the product provided to them.

We will be doing some EDA to get the insights from data like which brands and products are doing well while which are not.

Along with EDA and market analysis of the product, we will build a model that will help us determine/predict the rating of a product based on the review given by the user.

DATA WRANGLING

The raw dataset of the **Indian Product Reviews** consists of 5 columns and 2782 rows.

Following are the 5 columns present -

1. asin – It is a unique product ID
2. Name – Contains the name of the product.
3. Date – Contains the date when the review was given
4. Review – Contains the review for the product.
5. Rating – Contains the rating of the given product.

4 reviews were missing from the dataset; hence the number of rows were reduced to 2778.

The name column has the names written in unique fashion that allowed me to extract the name of the brand from the product name and save it to a new column consisting the Brand Name for each product.

The name of the brands was changed to lowercase to avoid any duplicity.

Finally, the review and rating data columns were divided into train/test split for modeling purposes.

EXPLORATORY DATA ANALYSIS

In this part we will get various insights that the data is providing us with.

Out of the total reviews we found there to be 24 unique brands and 122 unique products.

Having a look at the rating distribution for products we can have a broad idea of how the products are doing in the market.

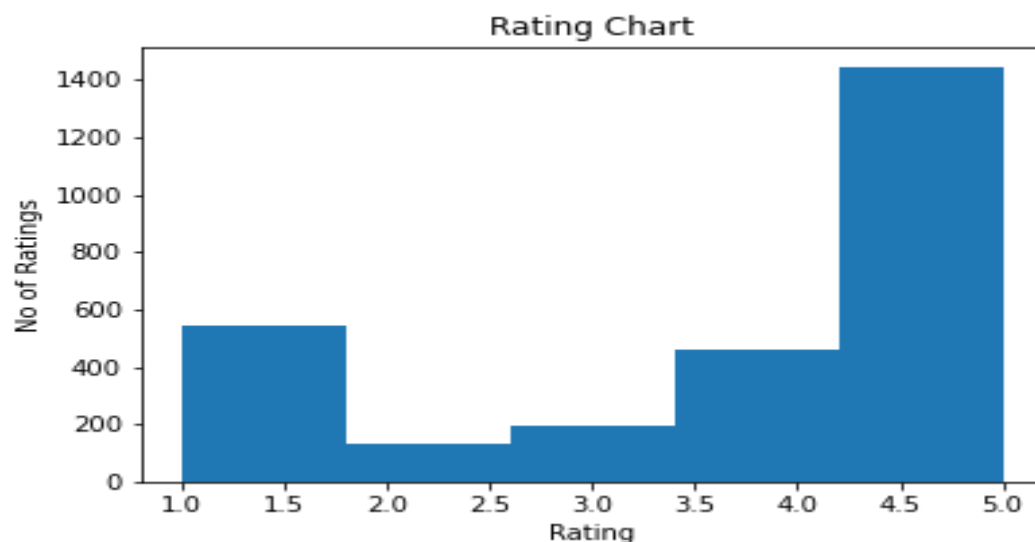


Fig 1: Rating Distribution

While the majority of products have a good rating, there is still a significant number of products having poor ratings.

Now, let's look at the top 5 brands that have the highest number of reviews.

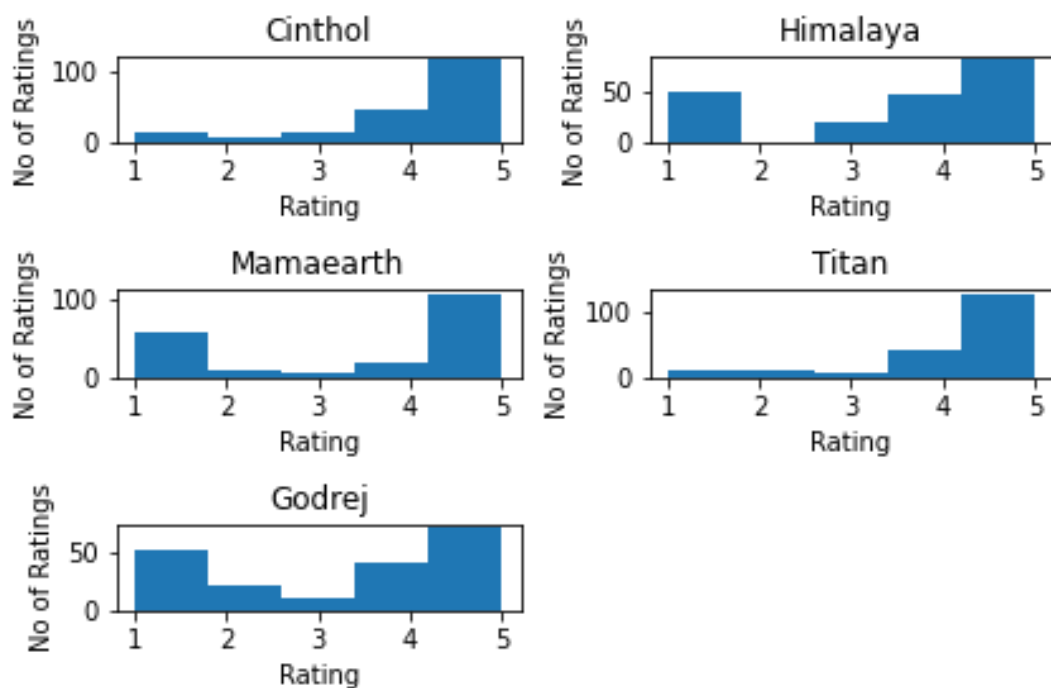


Fig 1.1: Ratings of Top5 Brands

Bottom 5 Brands

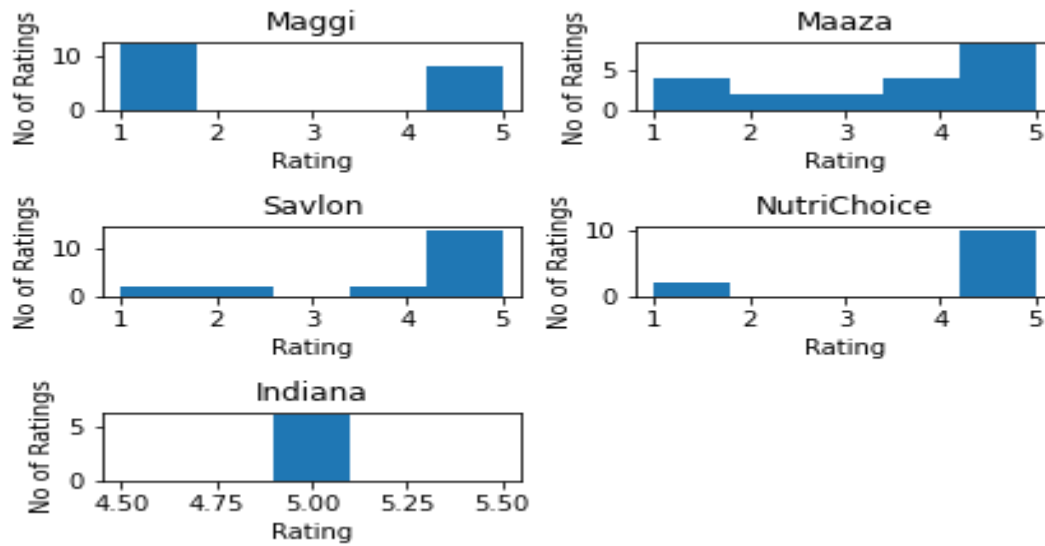


Fig 1.2: Ratings of Bottom 5 Brands