#### **Business Context**

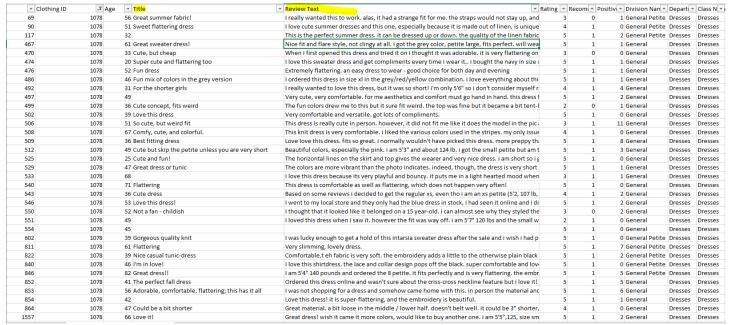
We are building a Customer Experience platform for B2C Customers. One of the core components of this Customer Experience Platform is our NLP Engine, which would consume textual data and generate insights.

#### **Problem statement:**

We have recently started working building an initial iteration for an Ecommerce customer with their Reviews data

Let us use a dataset available in Kaggle for our reference. **Link to ecommerce reviews data** <a href="https://www.kaggle.com/nicapotato/womens-ecommerce-clothing-reviews/version/1?select=Womens+Clothing+E-Commerce+Reviews.csv">https://www.kaggle.com/nicapotato/womens-ecommerce-clothing-reviews/version/1?select=Womens+Clothing+E-Commerce+Reviews.csv</a>

# Refer to the column's reviews. Title and reviews. Text

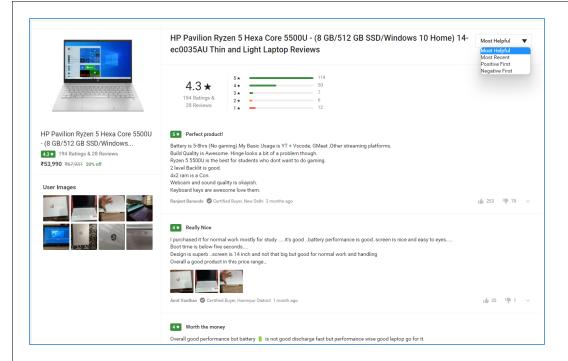


We had an initial discussion with the business team to understand their expectations.

One of the requirements, business team mentioned is that for **each product**, they want to **see the most relevant reviews**.

When we tried understanding the details, we got to know that these are reviews from ecommerce sites. The number of reviews per product increases over a period. Obviously, business team don't have the bandwidth to go through each review and understand what are the reviews that they should spend their time.

An example would be how ecommerce website (flipkart) has the option to sort data by the **most helpful** reviews. Refer to the screenshot below.



# **Expectation from your solution:**

- A working solution to prioritize and rank the reviews. For this exercise, you may want to filter
  the reviews based on ClothingID, take couple of products (IDs) which has the maximum
  reviews for this.
  - o Model should take input as All or a Specific Product ID.
  - Model should generate the output as CSV (containing the Top X reviews)
- Why did the Model Rate them as the TOP X reviews (X Configurable number like 10 or 20 etc.) This needs to be explainable.

You are NLP expert in the team, and we want you to come up with a solution which can meet the business team's expectations. This will help in iterating and taking this solution further.

## Following are our expectations:

As a data scientist, we want you to come up with

- 1. How would you approach this problem statement? We would like to understand your thought process?
- 2. Can you come up with an initial model using any of your favorite language, frameworks, libraries and model?
- 3. What are the assumptions you have made to come up with a solution?

## What are we trying to achieve with this?

- 1. Our objective is to understand your approach (thought process, code), when there is a one-liner like this given to you.
- 2. We understand that we have given you a one liner. We are not expecting a perfect model which would take care of all the edge cases.
- 3. Feel free to make your own assumptions. Document your assumptions so that we can review them later.
- 4. Create a repo in GitHub, upload your model and send us the link