**Data set description**:

This data set is about women clothing e-commerce customer reviews. It includes 23486 rows and 10 feature variables, where each feature represents a column. Each row corresponds to a customer review. The following are the columns and its description.

1. Clothing ID: Integer Categorical variable that refers to the specific piece being reviewed. In this data set, there are 1206 unique clothing ID’s.
2. Age: Positive Integer variable of the reviewers age. There are different age groups starting from 17 to 99.
3. Title/Review Title: String variable for the title of the review. It has a total of 13993 unique titles.
4. Review Text: String variable for the review body.
5. Rating: Positive Integer variable for the product score granted by the customer from 1 Worst to 5 Best.
6. Recommended IND: Binary variable stating where the customer recommends the product where 1 is recommended, 0 is not recommended.
7. Positive Feedback Count: Positive Integer documenting the number of other customers who found this review positive.
8. Division Name: Categorical name of the product high level division. There are 3 divisions which refers to clothing or apparel size.
9. Department Name: Categorical name of the product department name. There are 6 departments which refers to apparel or clothing types.

Class Name: Categorical name of the product class name. This column is a sub-division of department and corresponds to 20 different types of class categories.

**Quality of Findings**:

S**ummary for Finding 1**: Distribution of Age of people who reviewed the purchased products.

**validity of Finding 1:**

**A close up of a map

Description automatically generated**

**Managerial insights:** From the results, the most reviewers on the products are in the age group from 31 to 45. The age distribution figure shows that this age range has the highest positive reviews. Therefore, the two main insights or points to be considered is that:

a. The resulted age group is the most active in purchasing the products and reviewing it. So, the business must concentrate on keeping up with this fragment and

b. The e-commerce business element can investigate why other age groups are relatively less than the age group 31 to 45.

S**ummary for Finding 2:** Frequency Distribution of Departments According to the Age Groups

**Description:**

**validity of Finding 2:**

**A screenshot of a cell phone

Description automatically generated**

**Managerial insights:**

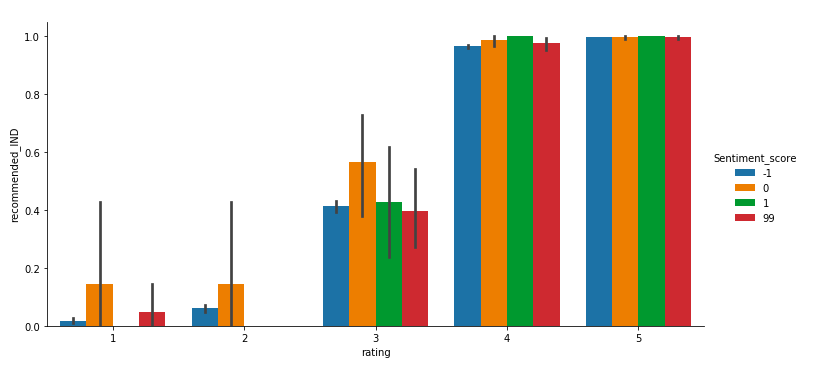
One of the reasons might be the product is offered in several apparel sizes, which attracts all age group people for a specific department. This finding can measure the product reviews and sales based on the category department.

S**ummary for Finding 3**: Distribution of Sentiment Scores Vs Rating Vs Recommended Indicator across Reviews.

**Description:**

The sentiment score for reviews are 1,0,1,99 represents Negative, Neutral, Positive and Null/Missing fields. These are the rows where the review text is not written, or the field is empty. In some scenarios customers do rate the product and recommend the product but they might not write any review text. For such fields we have considered the score as 99. From this graph, we can see that the sentiment score is almost similar for all the ratings. The customers who gave rating 5, wrote negative text about the product but still gave the recommendation. There might be reasons such as the customer had liked the product, but it might not be the way she expected it. The apparel types and sizes might be one of the reasons of negative review text.

**validity of Finding 3:**

****

**Managerial insights:**

From this e-Commerce get a sense of what is not working and make necessary changes to improve the shopping experience of the customer. It also gives the company an insight as to how customers perceive their products and/or services and would give them an idea on how to improve their business.

S**ummary for Finding 4:** Distribution of Products Based on Recommendation Factor

**Description**:

The recommendations are important because it says about the number of times the product is recommended and the total times the clothing products are purchased. Based on this feature, the clothing\_id which are more recommended, are the most purchases made. The total size of recommended IND for each clothing ID is the number of times the product is purchased, and the sum of recommended IND is the Number\_of\_times the product is recommended.

**validity of your finding:**

**A screenshot of a cell phone

Description automatically generated**

**Managerial insights:**

The clothing\_id which are more recommended, are the ones that are purchased more.

This is one of the important findings from which a business can look into, to incur profits and to implement changes in their products.