**WOMEN’S CLOTHING E-COMMERCE REVIEWS ANALYSIS REPORT**

**By: Obed Ayettey and Anita Blege**

**Abstract**

Using the Women’s Clothing Review dataset provided in Kaggle, we explored and gained insights from the items purchased and varying reviews. Our analysis revealed several insights that back our hypothesis that buyers who leave higher rating are more likely to recommend the product they purchased.

**Introduction**

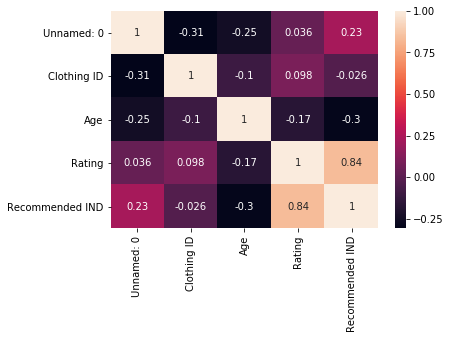
Reviews by customers are very important to businesses because it gives the business a fair idea of what buyers think or feel about their products. This helps businesses to determine how satisfied or dissatisfied customers are with their products and services. For this dataset, because the business is internet based, potential customers may be compelled to do business based on the reviews, ratings and recommendations of other customers who have purchased items from the online shop. These reviews can also inform the business on how well each item, class or department is doing with respect to marketing, sales and predictions on how revenue can be increased in the future

**Project Objective**

The goal of this project is to analyze the data and come up with a hypothesis and insights about the customers and the business based on the graphs we produce.

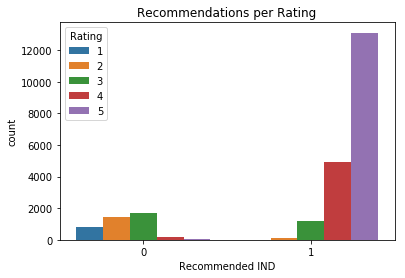
**Exploratory Data Analysis**

The dataset was imported into Jupyter notebook and calculations, explorations, cleaning and visualizations were performed using Python 3. After exploring our data, we discovered that the reviews were predominantly positive feedbacks with a satisfying level of recommendations. Based on our data wrangling, we plotted the following relationships using a heatmap:



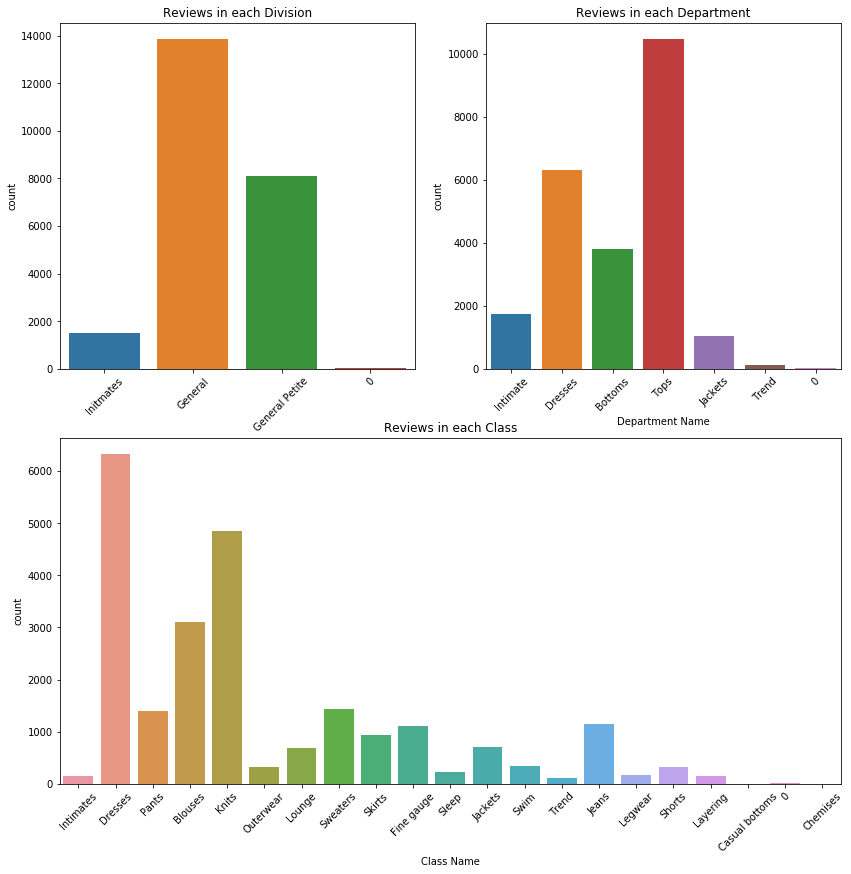
**Fig. 1: A heat map showing the correlation between Clothing ID, Age, Rating and Recommendations.**

This figure supports our hypothesis by showing a strong positive correlation between Rating and Recommendations. Also, as Age increased, Ratings and Recommendations decreased.

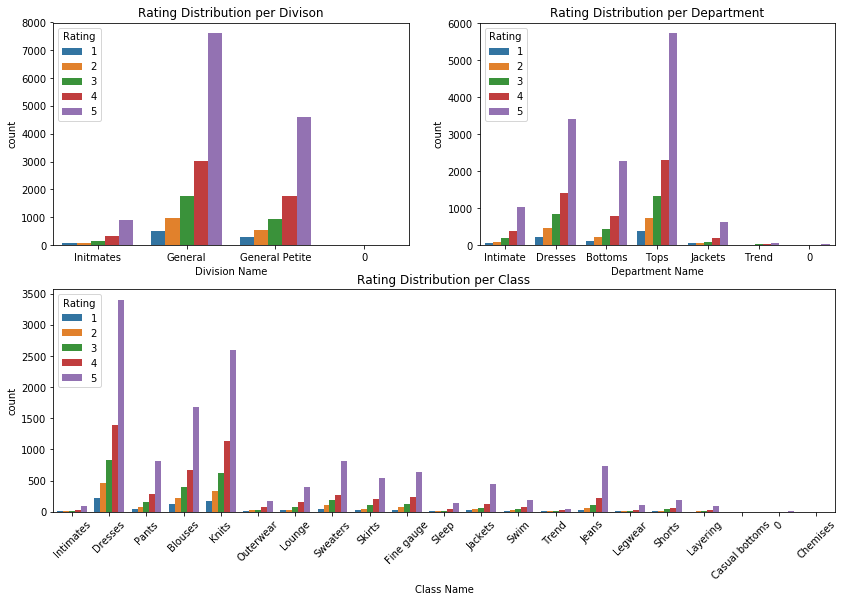


**Fig. 2: A count plot showing the number of Recommendations per Rating level.**

This count plot further confirms our hypothesis because it shows that higher ratings (>3) were highly distributed in the ‘1’ section of the recommendation type.

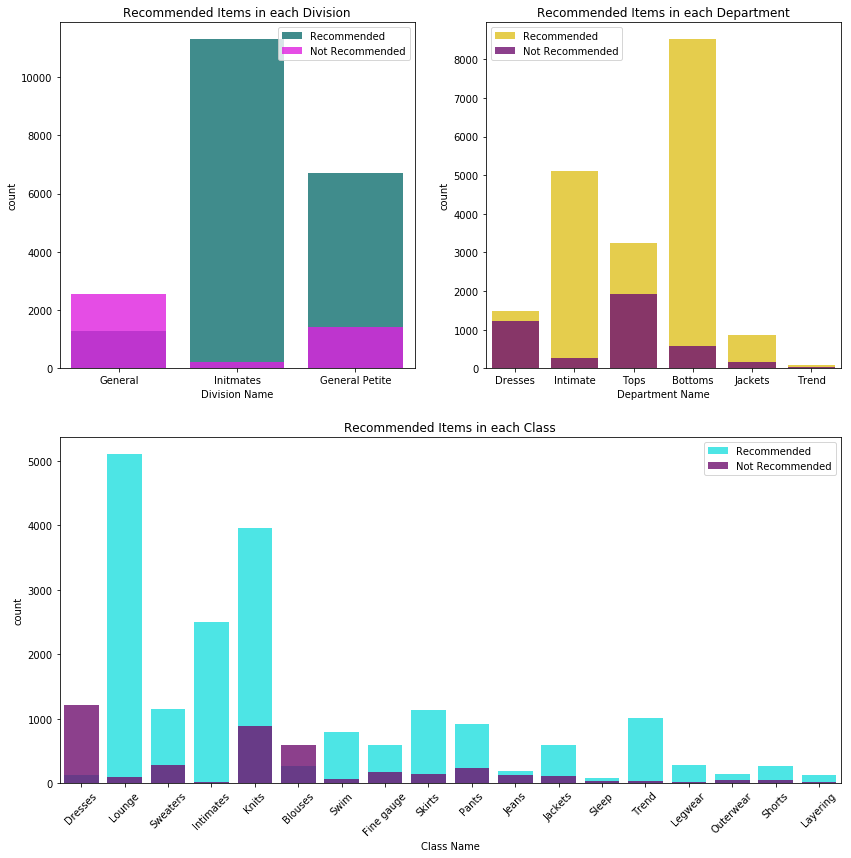


**Fig. 3: A count plot showing the total count of reviews in each division, each department and each class.**



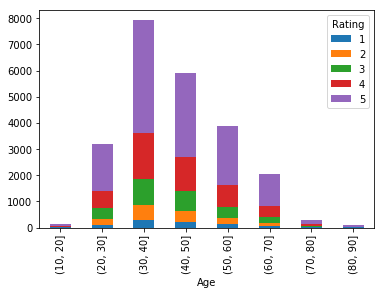
**Fig. 4: A stacked bar chart showing the distribution of total counts of ratings per division, department and class.**

With this graph, we can determine the various ratings from 1 to 5 of each item in division, department and class. The figure shows that dresses have the highest set of ratings followed by knits and blouses.



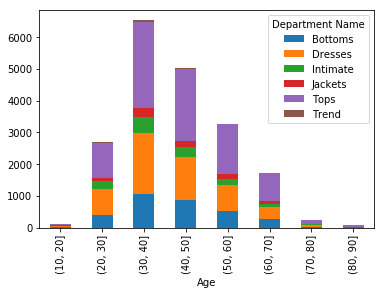
**Fig. 5: A stacked bar chart showing the distribution of total counts of recommended IND per division, department and class.**

Looking at the graph, we can determine the items that are mostly recommended or not recommend by the division level, department level and class level. An interesting discovery about this graph is that certain items that had many reviews and high ratings were not recommend by their buyers. Also, some items were recommended but were not rated by their buyers. With this insight, we deduce that there is/are other factors that affect the confirmation of our hypothesis.



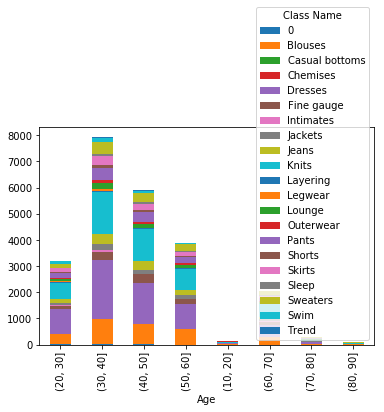
**Fig. 6: A graph that describes the distribution of total counts of ratings per age groups.**

This graph shows that people who usually rated the items or clothing were between the ages of 30 years and 50 years. This graph further posits that women who fall within the age range outside of 30 years to 50 years, were unlikely to rate or review the clothing they purchase.



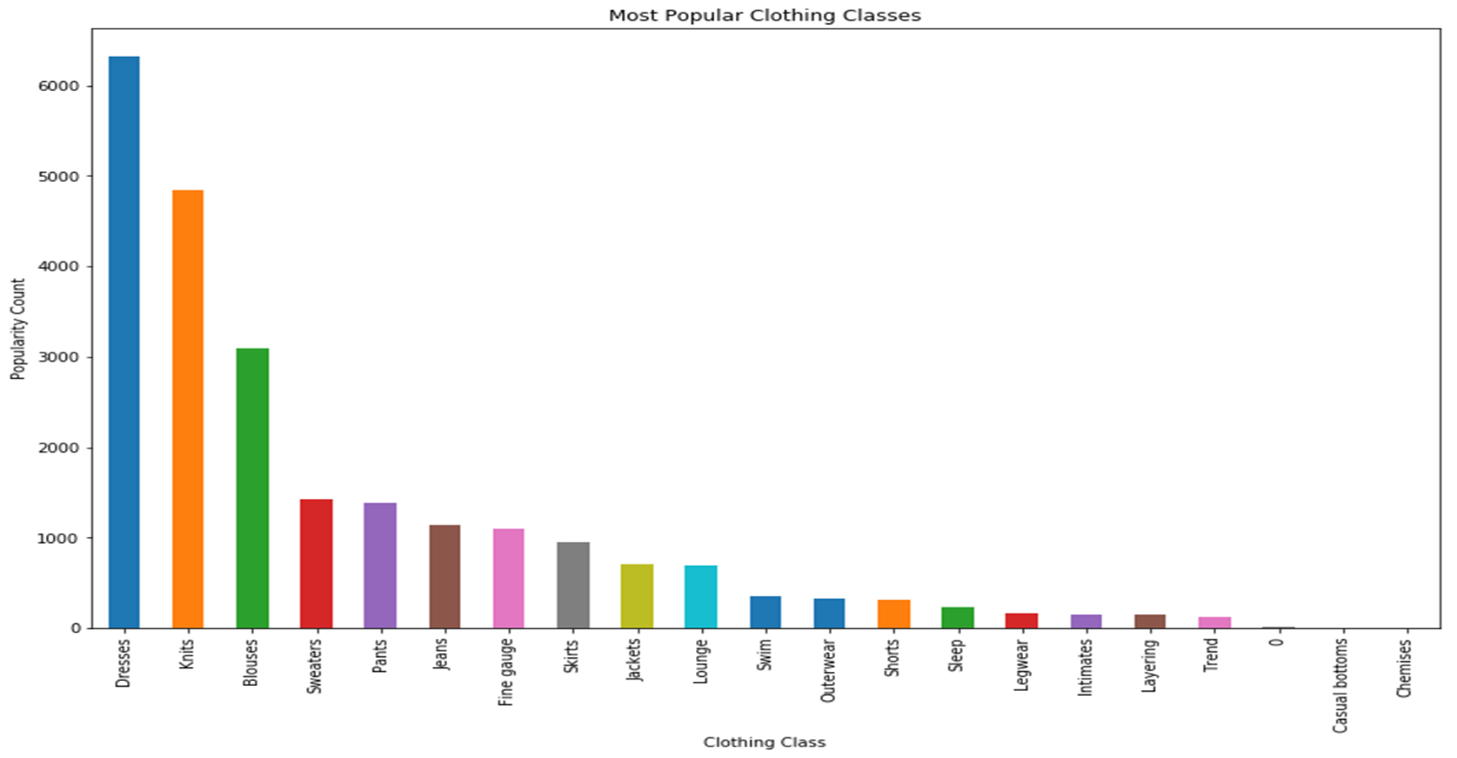
**Fig. 7: A graph that shows the distribution of each department across age groups**

We see the same trend in fig. 6 in this graph where the items in the department division were predominantly purchased by women within the 30 – 50 years range.



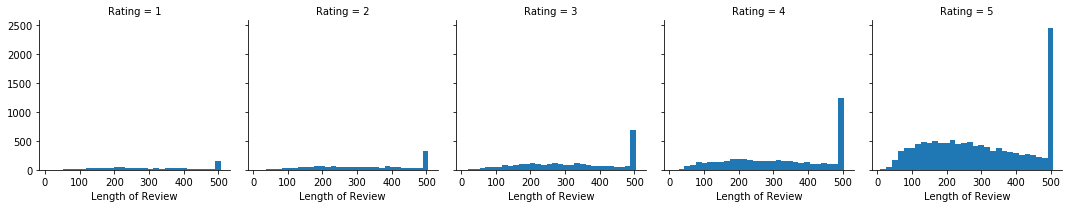
**Fig. 8: A graph that shows the distribution of each clothing class across age groups**

The trend continues in the class division of the items. We still see a lot of activity within the age range of 30 – 50 years.



**Fig. 9: A bar chart showing the most popular Clothing Classes per popularity count.**

Looking at this graph we see that the first three popular clothing classes are the same products that do not have high recommendations by customers. The reason could be that the customers generally were familiar with these products and did not see the need to rate them. Also it could mean that the age range with the highest recommendation activities (ie. 30 – 50 years), did not purchase a lot of these popular clothing classes.

****

**Fig. 10: Histograms showing the correlation between Length of Review and Rating**

Last but not least, this graph shows the individual ratings per the length of reviews. We can see that the least ratings had relatively lower lengths of reviews as compared to the high ratings which had longer lengths of reviews. From this graph we can confirm that women who were excited about a product or clothing class spent more rating, reviewing and recommending the items. Women who were not impressed with their purchases spent little or no time in writing lengthy reviews about the items. Per the customer behaviours we can deduce that customers tend to write longer reviews when satisfied with their purchase.

**Summary**

The hypothesis that customers who leave higher ratings are more likely to recommend their purchases was supported by the data exploration and analysis. Ratings and Review Length were also highly correlated with customer recommendations. Certain factors like age, department or type of clothing and length of reviews directly influenced ratings and recommendations. Some items had many reviews and high ratings but were not recommend by customers, while others were recommended but were not rated by their buyers. Certain clothing type regardless of ratings and recommendations were popular across all ages.