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

**Abstract**

Using the dataset provided in kaggle, we explored our dataset pertaining to some women’s clothing outlet and their reviews regarding the items they purchased at certain times as well as their recommendations. Our research or data wrangling revealed a number of insights that backs our hypothesis that buyers who leave higher rating are more likely to recommend the product they purchased.

**Introduction**

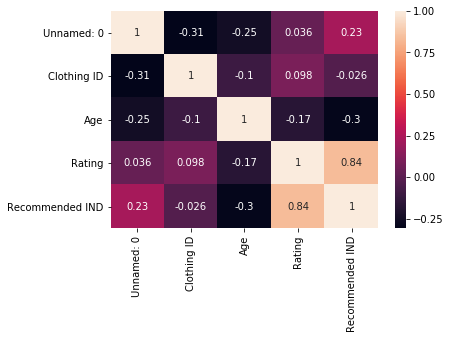
Reviews by consumers/ clients/customers are very important to businesses because it gives the business a fair idea of what buyers think or feel about their products. In other words, reviews help businesses to determine how satisfied or dissatisfied customers are with their products and services. In our project for instance, the reviews, ratings and recommendations are very necessary because the business is electronic/internet based. Therefore, potential customers may be compelled to do business based on the reviews, ratings and recommendations of other people who have purchased items from the online shop. The reviews, ratings and recommendations could also inform the business of how well each item, class or department is doing with respect to marketing and sales. With this dataset, we can make a good deal of insight as well as predictions on how sales could be in the future. Our major focus though is to analyze the data and come up with some insights about the customers and the business based on the graphs we produce.

**Project Objective**

The goal of this project is to analyze the dataset with our python notebook, and to come up with a hypothesis and insights from the graphs produced.

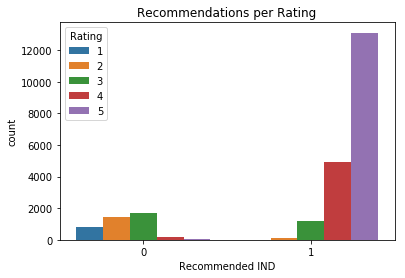
**Exploratory Data Analysis**

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 plot the following relationships:



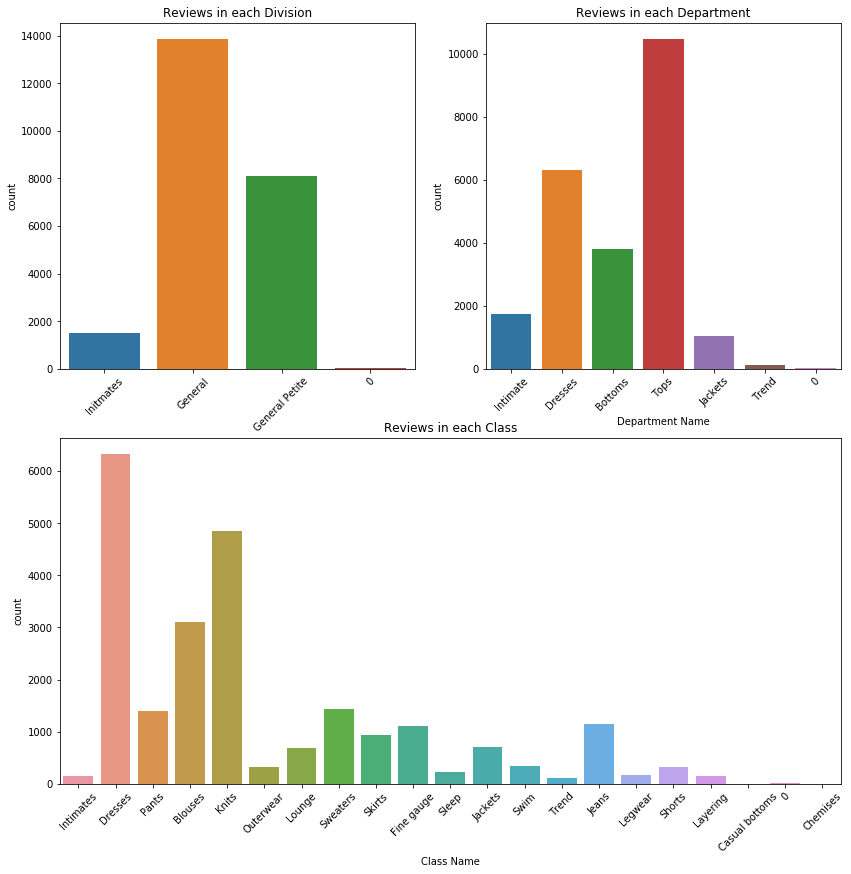
**Fig. 1: A heat map showing the correlation between Clothing ID, Age, Rating and Recommended IND.**

This heat map gives us the impression that there is a strong positive correlation between Rating and Recommended IND. This could also mean that one leads to the other. Therefore, per this correlation we get the understanding that people who rate a product are likely to recommend it to others.



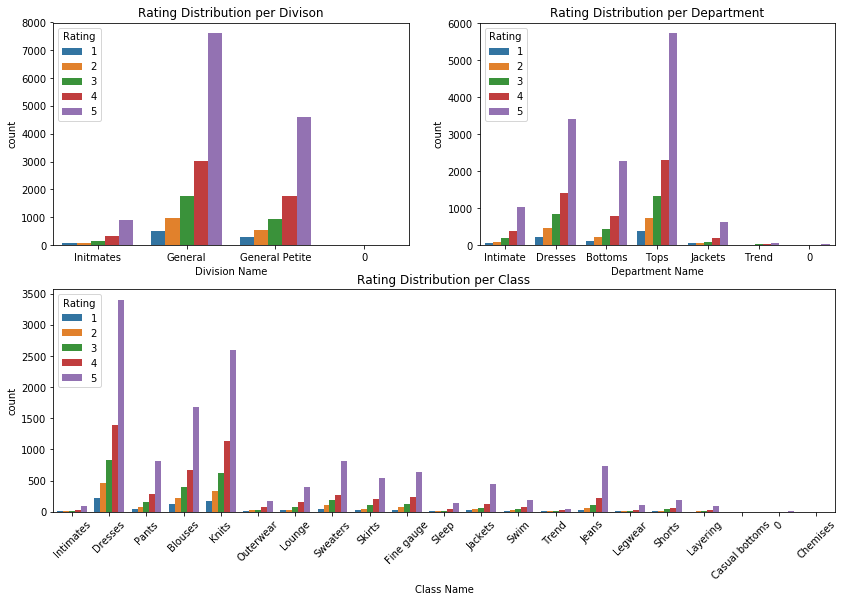
**Fig. 2: A count plot showing the Recommended IND categories and the Ratings.**

This count plot gives our previous analysis further grounding because we see clearly that the category with the highest ratings is actually the recommended category.



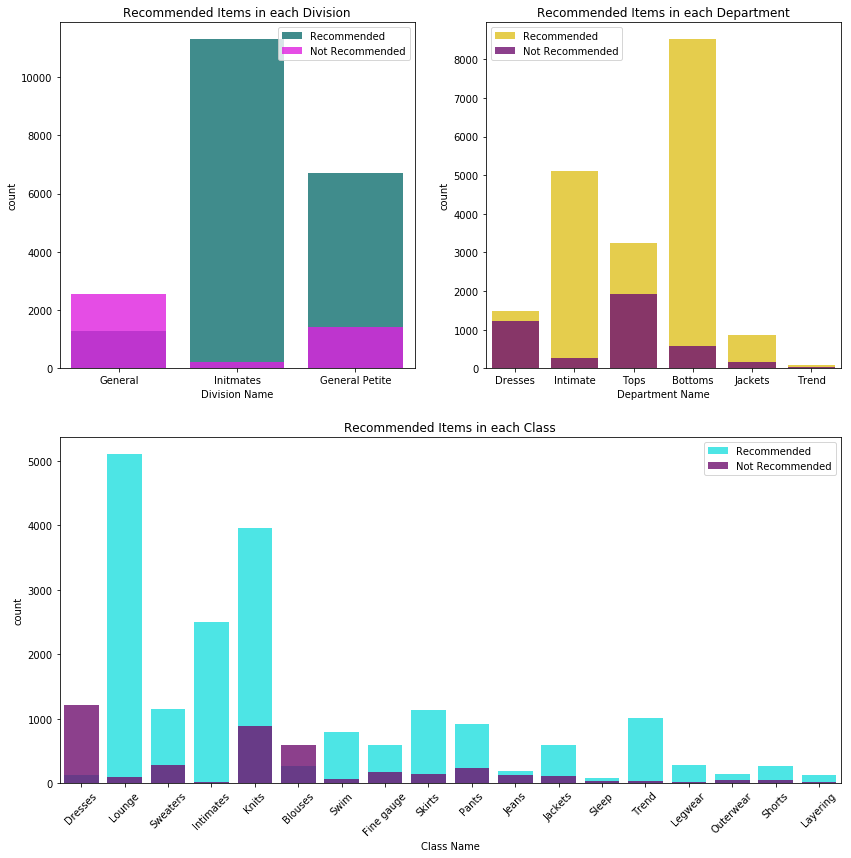
**Fig. 3: A count plot showing the total count of reviews in each division, each department and each class.**

This count plot is a representation of the total counts of reviews by customers in three different categories (Division, Department and Class). We can see clearly the various items receiving many reviews and possibly leading to recommendations.



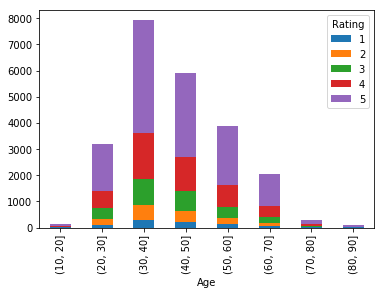
**Fig. 4: A count plot showing the distribution of total counts of ratings per division, department and class.**

With this graph, we are able to determine the various ratings from 1 to 5 of each items in division, department and class. We can tell that dresses have the highest set of ratings followed by knits and blouses.



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

Looking at the graph, we are able to 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 the fact that some items with many reviews and high ratings in fig. 3 and fig. 4 also have those items highly not recommend in fig. 5. The items include dresses, knits and blouses. With this insight, we get the impression that there is/are some other factors that affect our hypothesis. We produced some other graphs to explain the reasons for this situation.



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

This graph gives us the impression that people who usually rated the items or clothing were between the ages of 30 years and 50 years.

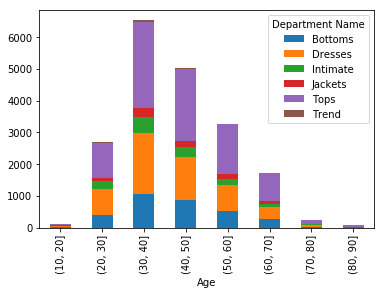


Fig. 7: A graph that shows the distribution of each department across age groups

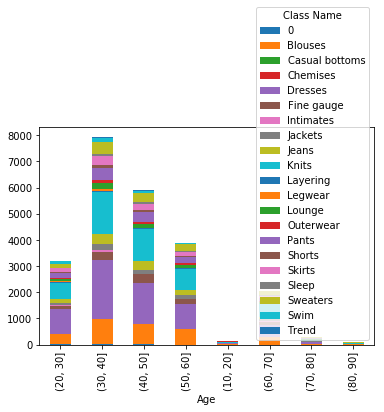


Fig. 8: A graph that shows the distribution of each clothing class across age groups