

Milestone 02: Data Summary/Visualization

Attribute Name	Minimum	1 st Quartile	Median	Mean	3 rd Quartile	Maximum
Age	18.0	34.0	41.0	43.2	52.0	99.0
Rating	1.0	4.0	5.0	4.196	5.0	5.0
Recommended-IND	0.0	1.0	1.0	0.8224	1.0	1.0
Positive Feedback count	0.0	0.0	1.0	2.536	3.00	122.00

WORD	FREQ	WORD	FREQ
1 love	1854	1 dress	10512
2 great	1789	2 love	8924
3 dress	1645	3 i	8763
4 cute	1550	4 size	8713
5 beautiful	1405	5 top	7365
6 top	1172	6 fit	7265
7 perfect	811	7 like	6994
8 pretty	666	8 wear	6423
9 fit	607	9 great	6082
10 nice	528	10 just	5574
11 flattering	503	11 fabric	4773
12 runs	483	12 small	4578
13 lovely	472	13 color	4567
14 comfortable	470	14 this	4318
15 comfy	456	15 look	4021
16 gorgeous	446	16 really	3923
17 soft	434	17 ordered	3832
18 summer	425	18 little	3772
19 sweater	421	19 perfect	3747
20 small	337	20 will	3617
21 super	330	21 one	3584
22 fun	327	22 flattering	3490
23 shirt	308	23 soft	3313
24 color	307	24 well	3234
25 quality	294	25 back	3181

Figure 01: Attribute summary of Quantitative and Qualitative data

The summary of attributes title and review text show top 25 words used with common words like love, dress, top, fit, size, perfect, color. The quantitative attributes discuss the mean, median, Quartiles, min, max indicating the central tendency and spread of data.

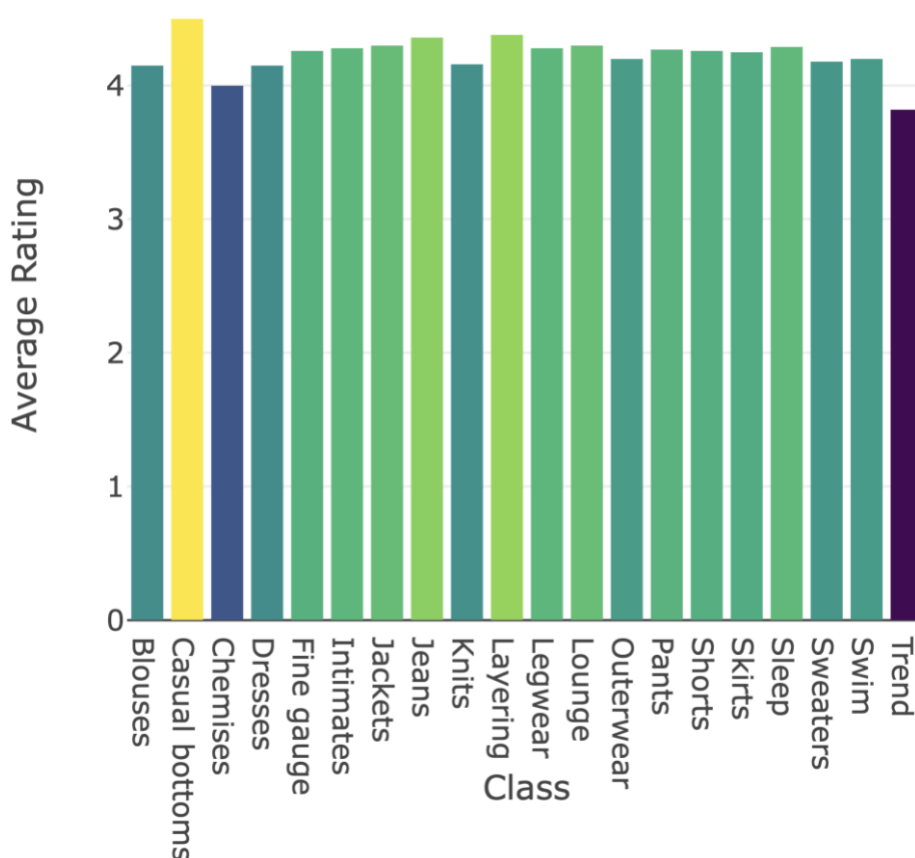


Figure 02: Average Rating vs Classes

The graph shows maximum average rating casual bottoms, jeans, layering and approximate similar of intimates, jackets, fine gauge, pants, shorts, skirts, sleep and the lowest of Trend, chemises. Moreover, the average rating of classes seems to almost similar among them.

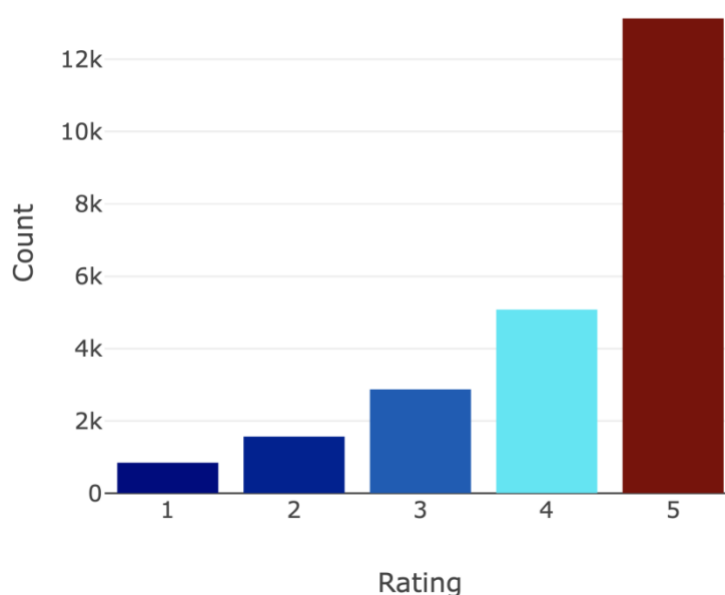


Figure 03: Count vs Ratings

The distribution of reviews with 5 rating is maximum followed by 4,3,2,1. This shows the performance of the store being going well but around 3k reviews are less than 3 rating which indicate to take look to negative feedback. The summary of rating show mean, median, 1st and 3rd quartile around 4 - 5 which shows the maximum spread of data around 4-5 rating and positive feedback.

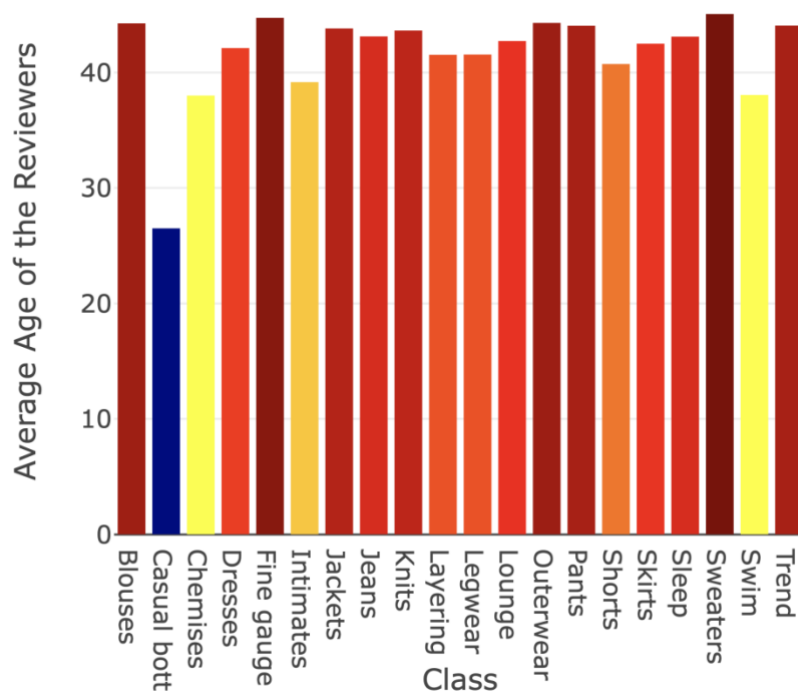


Figure 04: Average Age of Reviewers per Class

The average age of users giving rating is around 40-45 on similar classes like blouses, fine gauge, sweaters, outwear, dresses, jackets, jeans, pants, shorts, sleep, trends. The lowest average age is 25 given reviews to casual bottoms class with maximum average rating of 5. The graph shows same age group give ratings and reviews to similar classes. The class casual bottoms have maximum average rating with lowest average age group giving reviews.

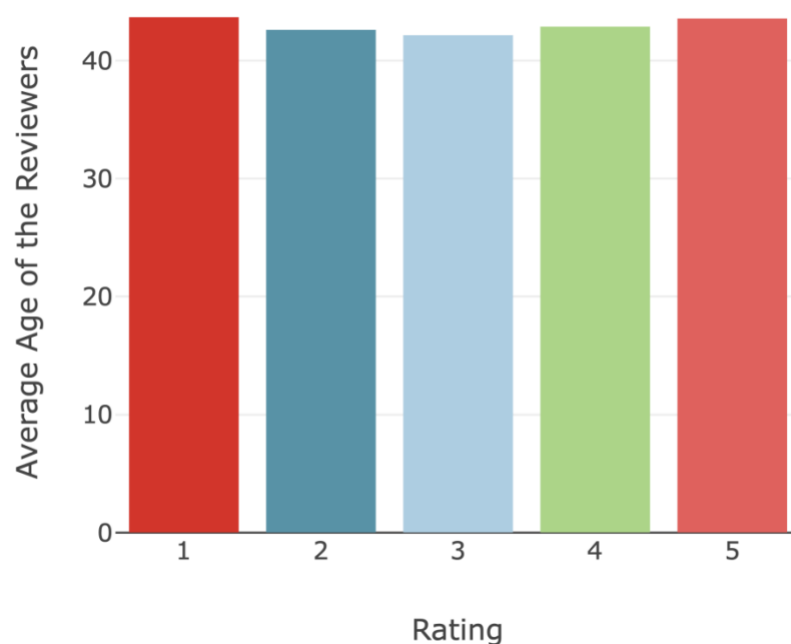


Figure 05: Average Age of the Reviewers vs Rating

The average age of reviewers is around 42-44 has almost similar rating across from 1-5. The maximum characteristics of reviews and target customers for the system are around focused on reviewers around age 40-45.

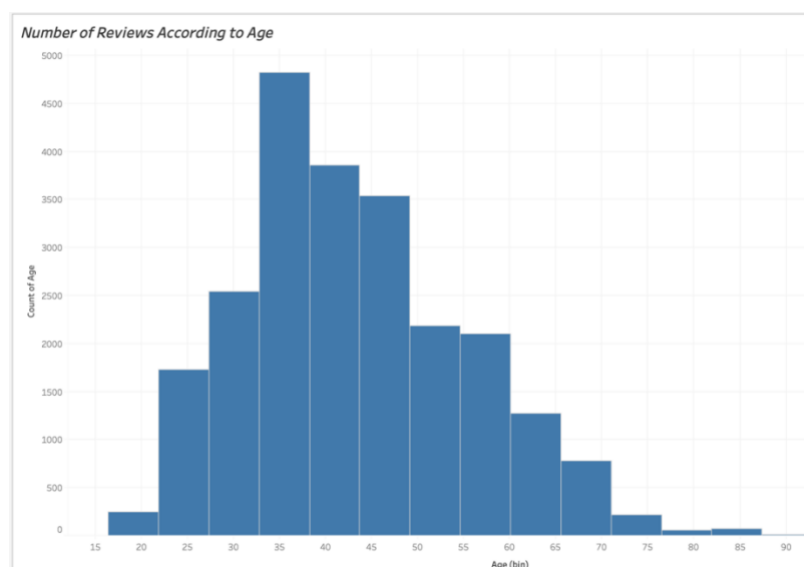


Figure 06: Count of Reviews according to Age

The maximum age spread of giving reviews are 35-45 with lowest count of reviews given by age 20, 70-85. The young age may be busy in work life and old age may be minimum because of lack use of technology. The age group around 32-50 shows the maximum count of reviews/rating given to the products of different classes.

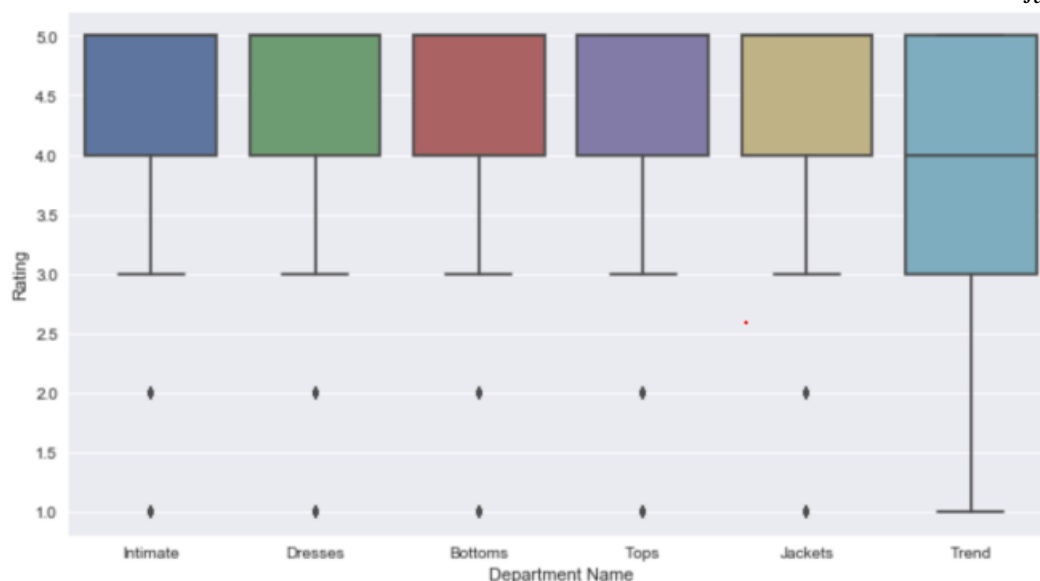


Figure 07: Distribution of Rating according to Department Name

The distribution of ratings across all departments is around 4.0 - 5.0, with trend having lowest rating and no outliers. The department Intimate, Dresses, Bottoms, Tops, Jackets have close readings of mean, median and similar pattern outliers. Therefore, distribution of ratings does not show important relation with department names expect for trend.

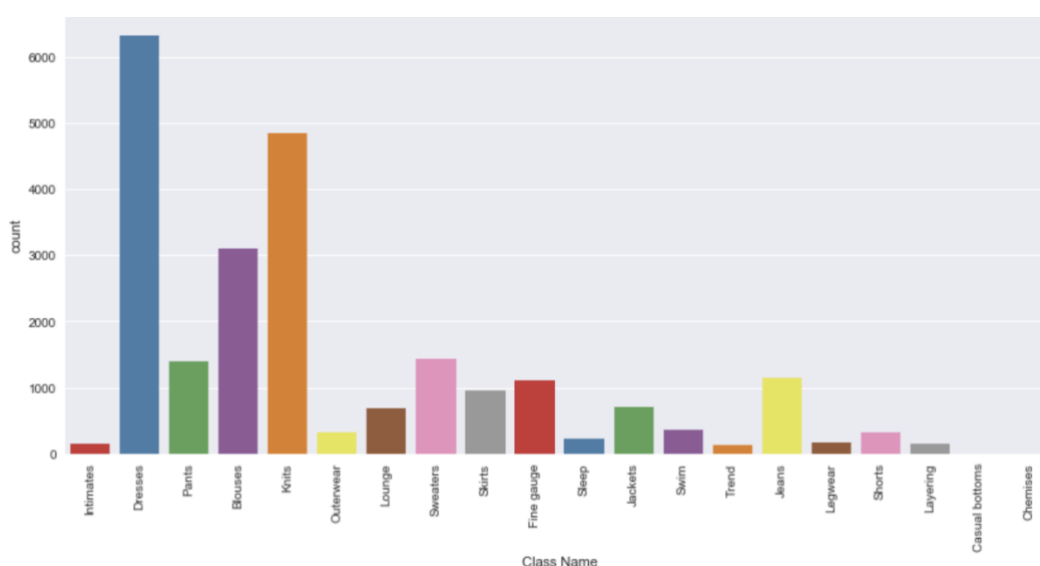


Figure 08: Count of Reviews according to Class Name

The reviews of Dresses, knits, Blouses having maximum count. From the above figure 01 the maximum word used in attribute title and review text show dress word used for count 12157 which clearly indicates the best class bought by customers and reviewed from the store.

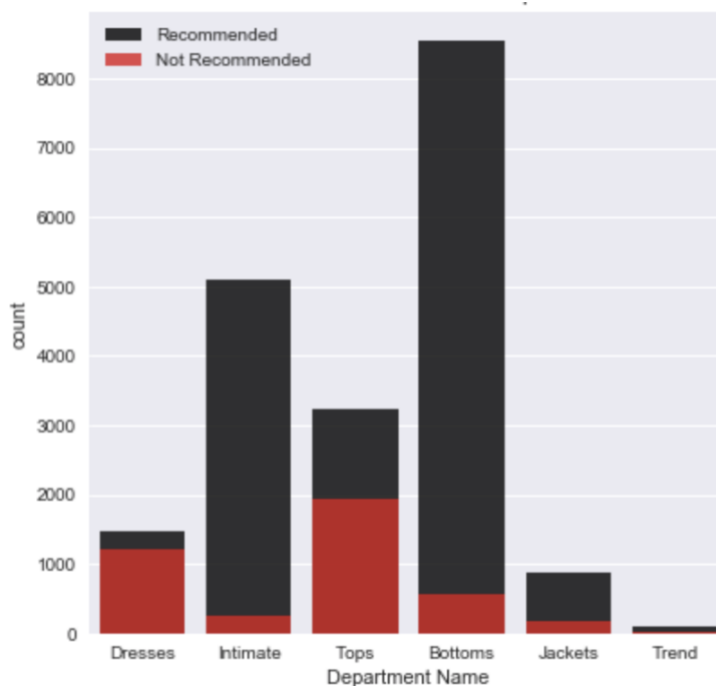


Figure 09: Recommendations according to Department Name

The data shows maximum count of recommended reviews the graph shows Intimates and Bottoms being most recommended and trend and jackets being lowest. As the recommendations are higher that shows store clothes are loved by the customers. The dresses and tops department have highest ratings and reviews but less recommendations.

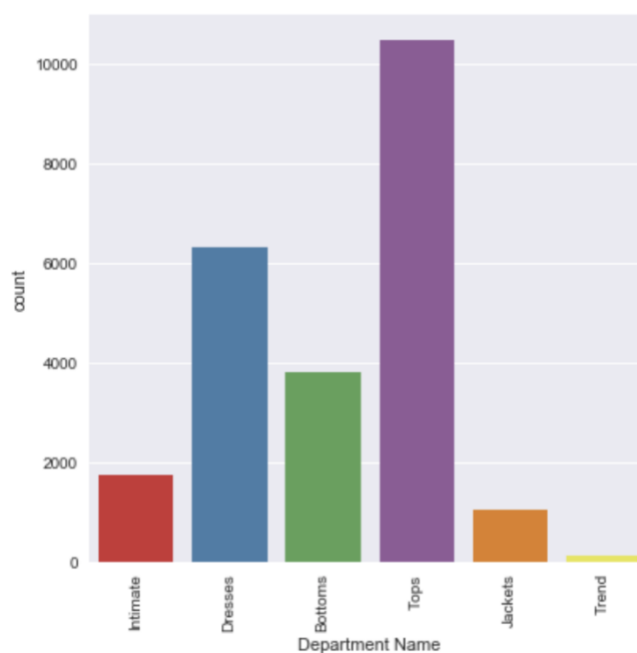


Figure 10: Count of reviews per Department Name

The department tops and dresses have maximum count of reviews, Trend having lowest count, Intimates, Bottoms, Jackets have mid count. The graph above average ratings and age by class names show the products count from tops and dresses are reviewed the most from the store.