**Amazon Dataset**

The unit of observation for this dataset is customer response. The relevant variables that were used from this dataset were *'Rating'* and *'Review Text'*, numerical and textual variables respectively.

Category: A categorical variable describing review responses from purchases that consisted of specified keywords to categorize reviews into ones that mention shipping, quality, both, or neither – specified as "Other." The keywords searched for pertaining to the topic of shipping were as follows: 'shipping', 'delivery', 'shipment', 'deliver', 'time', 'packaging', 'box', 'late', 'early', 'fast', 'slow', 'arrived', 'delivered', 'received'. The keywords for quality were as follows: 'quality', 'fit', 'length', 'material', 'fabric', 'cheap', 'shape', 'style', 'clothing', 'made', 'size', 'thin'.

This is derived directly from the *'Review Text'* variable in the original dataset, as it was parsed for keywords in order to be categorized. The process to acquire this variable begins by cleaning the text by converting to lowercase, removing non-alphabetical tokens, and eliminating a defined set of stop words. Then, the `categorize_review` function was created to classify each review based on the specific keywords mentioned above. Lastly, from this preprocessed text, the function checks for the presence of one or more of the key words in each review and assigns it to a certain category.

Category	Frequency	Percent of Total
Both	481	7.64%
Other	1549	24.6%
Quality	4091	64.98%
Shipping	175	2.78%
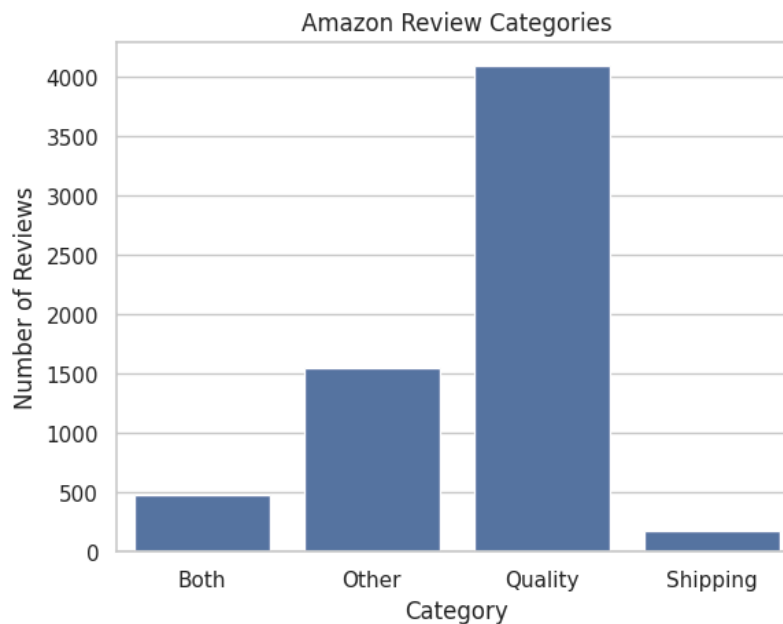
Sentiment: A categorical variable that expresses the sentiment of a particular customer review. The values can be either 'Good,' 'Bad,' or 'Neutral.'

It takes preprocessed review text (namely, '*Processed_Reviews*') and inputs it into a Sentiment Intensity Analyzer, which has a `polarity_scores` function. This analyzer uses this function to decisively assess the sentiment of a review and returns a dictionary of numeric values representing how negative, neutral, or positive the review is. The compound score, automatically calculated as part of this built-in function, then denotes if the sentiment should be 'good,' 'bad,' or 'neutral' based on if it is greater than 0.05, less than -0.05, or in between respectively.

Category	Sentiment	Count	Percent of Total
Both	Bad	23	4.78
	Good	453	94.17%
	Neutral	5	1.04%
Other	Bad	44	2.84%
	Good	1473	95.09%
	Neutral	32	2.07%
Quality	Bad	78	1.91%
	Good	3980	97.29%
	Neutral	33	0.81%
Shipping	Bad	5	2.86%
	Good	167	95.43%
	Neutral	3	1.71%

Amazon Review Sentiment Counts by Category
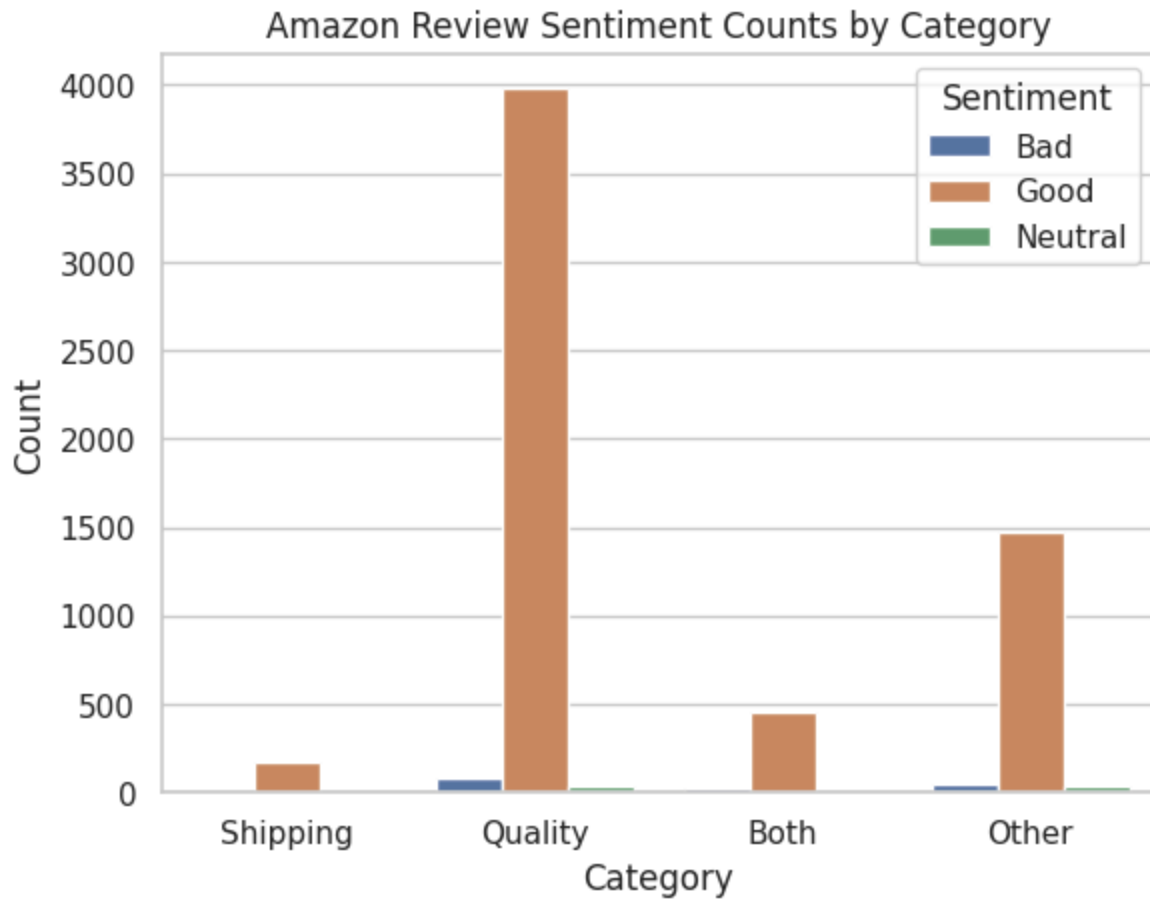
Fashion Nova Dataset

The unit of observation for this dataset is customer response. The relevant variables that were used from this dataset were *'Rating'* and *'Review Text'*, numerical and textual variables respectively.

Category: A categorical variable describing review responses from purchases that consisted of specified keywords to categorize reviews into ones that mention shipping, quality, both, or neither – specified as "Other." The same keywords and processes used to derive this variable for the Amazon dataset were used and duplicated to derive this variable for its respective dataset.

Category	Frequency	Percent of Total
Both	6825	5.17%
Other	90,651	68.68%
Quality	17,017	12.89%
Shipping	17,487	13.25%
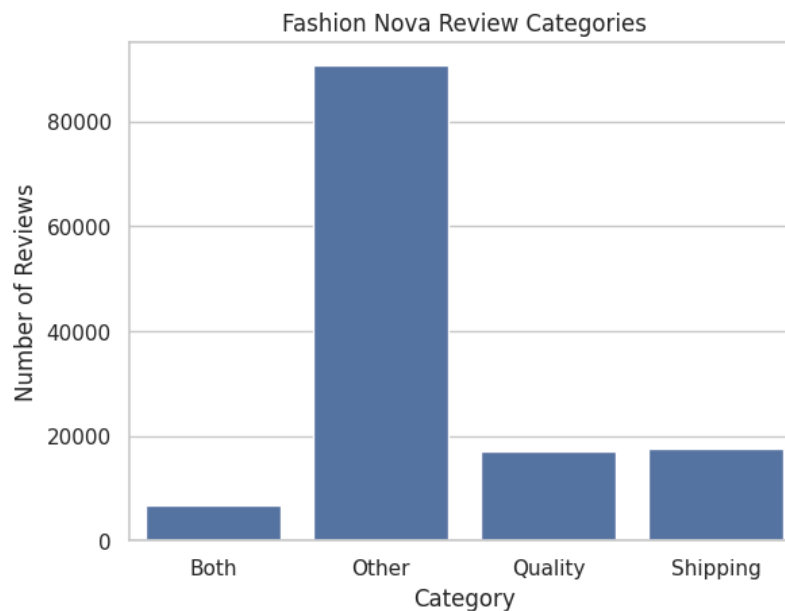
Sentiment: A categorical variable that expresses the sentiment of a particular customer review. The values can be either 'Good,' 'Bad,' or 'Neutral.'

The same processes are used to derive this variable as was performed to derive the variable of the same name for the Amazon dataset.

Category	Sentiment	Count	Percent of Total
Both	Bad	854	12.51%
	Good	5682	83.25%
	Neutral	289	4.23%
Other	Bad	3981	4.39%
	Good	42,190	46.54%
	Neutral	44,480	49.07%
Quality	Bad	915	5.38%
	Good	15,219	89.43%
	Neutral	883	5.19%
Shipping	Bad	3085	17.64%
	Good	11,959	68.39%
	Neutral	2443	13.97%

Fashion Nova Review Sentiment Counts by Category