

Data Appendix

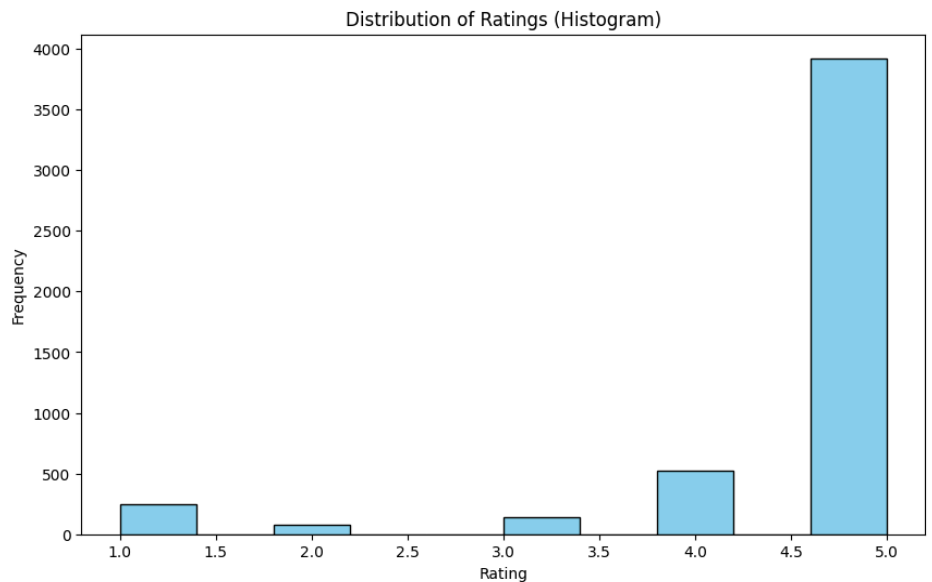
Amazon Product Reviews Dataset

The unit of observation for this dataset is individual product reviews. The variables from the dataset that were used in our analysis are 'Overall Rating' and 'Review Text'.

Overall Rating

This variable represents the rating that is given to the Amazon product by the reviewer, ranging from 1 to 5. It can be classified as a quantitative variable. In the input data file `amazon_reviews`, this variable is called 'overall'. There are 4,915 observations for this variable and no observations are missing, which means that there is a value of 4,915(0). No processing was needed to prepare this variable for data analysis.

overall	
count	4915.000000
mean	4.587589
std	0.996845
min	1.000000
25%	5.000000
50%	5.000000
75%	5.000000
max	5.000000



Review Text

This variable is the user's evaluation summary of the Amazon product. This is a categorical variable that can include any string of characters or numbers. In the input data file `amazon_reviews`, this variable is called 'reviewText'. There are 4,914 observations for this variable and 1 observation is missing, which can be reported as 4,914(1). This variable was processed by converting all of the text into lowercase, removing punctuation, removing stop words and non-sentimental product-related words, and splitting it into individual words.

word	frequency
works	1557
not	1450
great	1444
no	1257
use	1113
fast	906
good	890
price	842
bought	832
samsung	828

