

IMDB Dataset of 50K Movie Reviews

Unit of Observation: Each row in the dataset represents a single movie review written by a user, along with a label indicating whether the review expresses a positive or negative sentiment.

Dataset Overview

The dataset consists of 50,000 movie reviews labeled by sentiment. Each review is categorized as either positive or negative. The dataset is balanced, containing equal numbers of positive and negative reviews.

	review
sentiment	
negative	25000
positive	25000

- Number of observations: 50,000
- Number of variables: 2
- Data type: Text and categorical labels

Variable 1: Review

- Type: Text
- Description: The full written movie review submitted by a user
- Purpose: Used as input for text-based feature extraction and modeling
- Average review length: 1294 characters

Variable 2: Sentiment

Type: Categorical

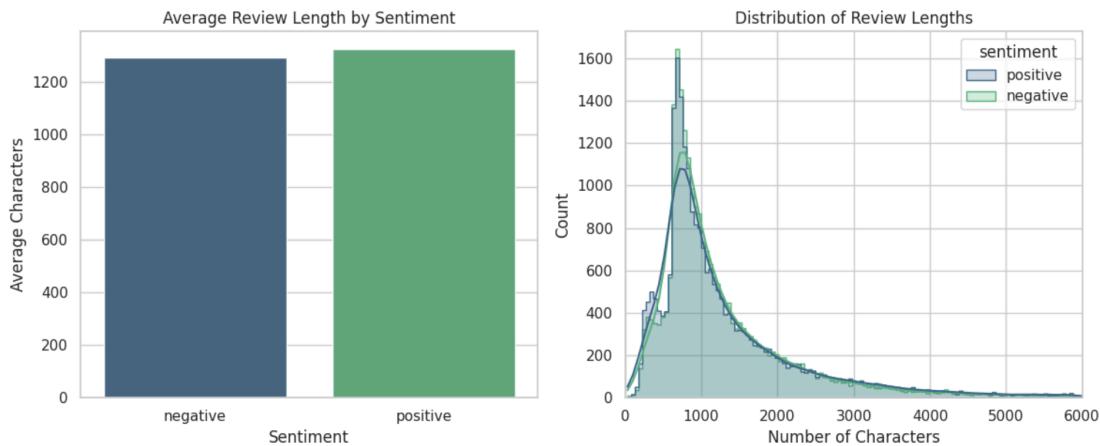
Possible values:

- Positive
- Negative

Description: Indicates the overall sentiment expressed in the review

Review Sentiment's Impact on Average Review Length Visualization:

- Review length is approximately the same for either sentiment.



Data Dictionary:

Column	Description	Potential Responses
Review	The text of the review posted by a user about a movie.	Probably my all-time favorite movie, a story of selflessness, sacrifice and dedication to a noble cause, but it's not preachy or boring. It just never gets old, despite my having seen it some 15 or more times in the last 25 years.
Sentiment	The overall tone of the review, indicating whether it is positive or negative.	Positive

	review	sentiment
0	One of the other reviewers has mentioned that ...	positive
1	A wonderful little production. The...	positive
2	I thought this was a wonderful way to spend ti...	positive
3	Basically there's a family where a little boy ...	negative
4	Petter Mattei's "Love in the Time of Money" is...	positive
...
49995	I thought this movie did a down right good job...	positive
49996	Bad plot, bad dialogue, bad acting, idiotic di...	negative
49997	I am a Catholic taught in parochial elementary...	negative
49998	I'm going to have to disagree with the previou...	negative
49999	No one expects the Star Trek movies to be high...	negative
50000 rows x 2 columns		