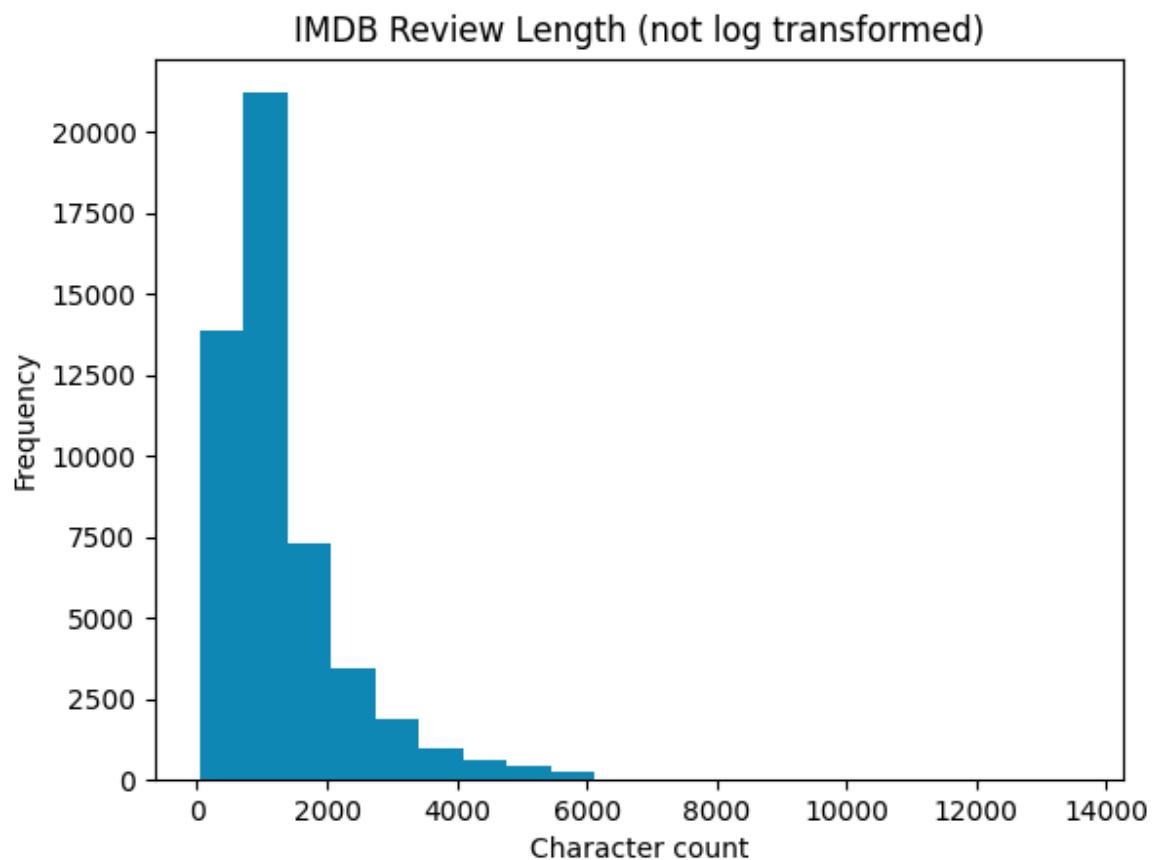
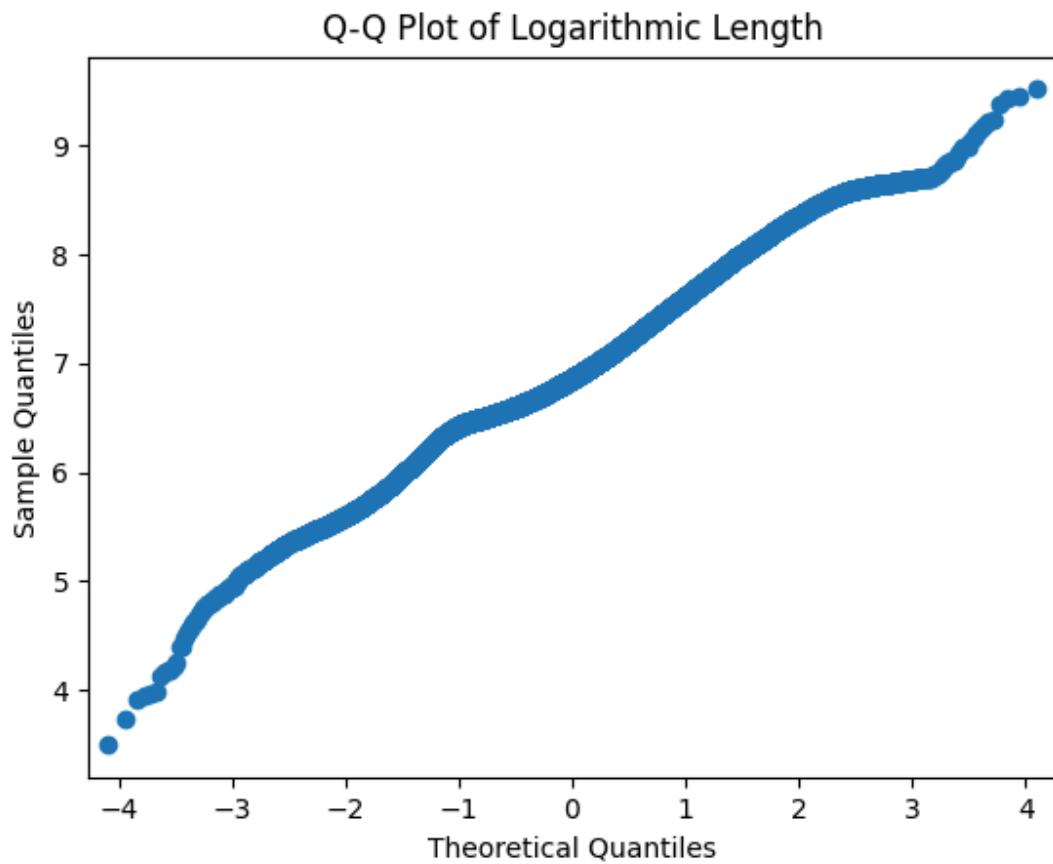


Dataset 1: IMDB_Dataset.csv

Each unit of observation is a movie review. It contains the text of the review as well as the sentiment, positive or negative.

review → raw text of review





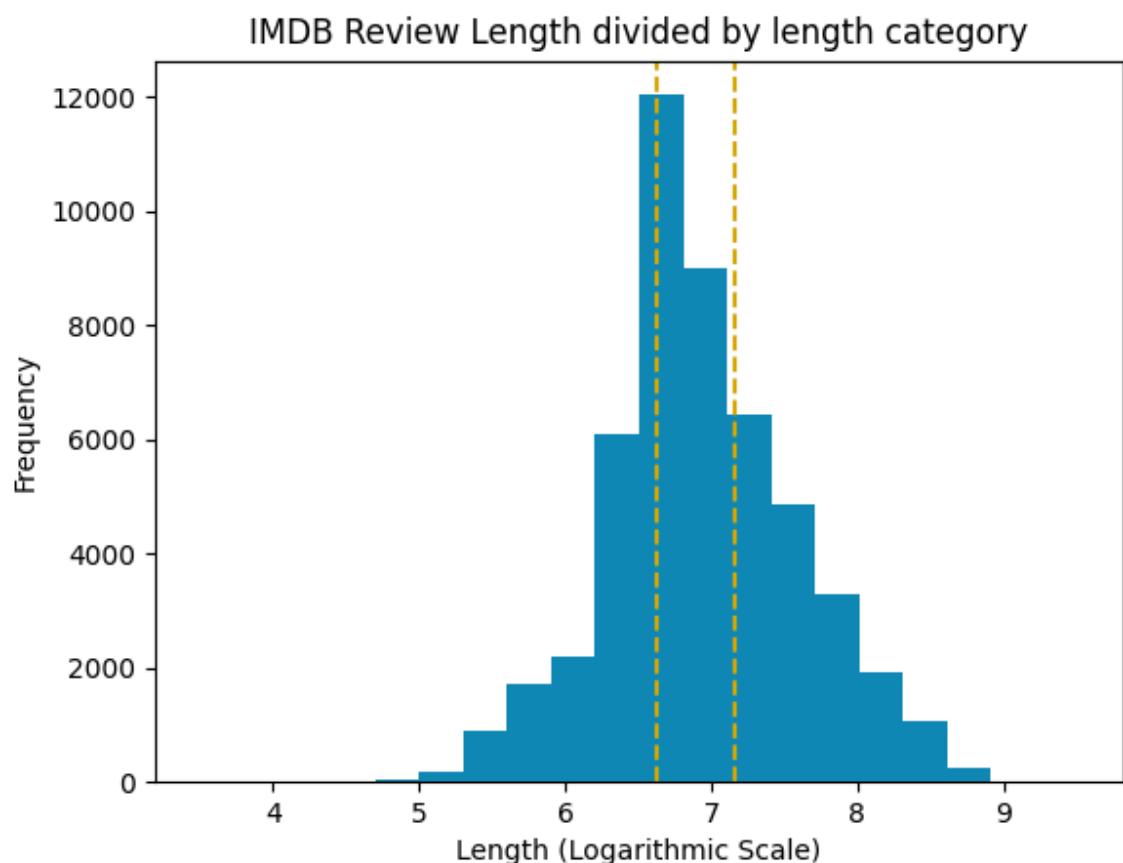
sentiment → direction of sentiment (positive or negative)

Dataset 2: reviews_with_severity.csv

Each unit of observation is a movie review. It contains the text of the review as well as the length group, calculated by log-transformed tertiles, and the severity score, as calculated by the RoBERTa sentiment analysis model.

review → raw text of review

length_group → length category for ANOVA testing



roberta_severity → sentiment score

Mean Review Extremity by Length Group

Based on absolute RoBERTa sentiment scores

