

WORD CLOUD ON MOVIE REVIEWS

Problem Statement

In the problem below, we provide you the movie reviews data and the expectation is to use natural language processing to identify the positive and negative words for a single movie.

Data Description

Data is in the CSV format and it is attached in the mail.

Columns	Description	Useful or not
reviewerID	Unique ID for reviewer	Useful
asin	Movie ID	Useful
reviewerName	Name of the reviewer	Not useful
helpful	Some flag (NOT USEFUL)	Not useful
reviewText	Review text (Main column)	Useful
overall	Rating of the movie	Don't use
summary	Summary of the review	Not useful
unixReviewTime	Time stamp	Not useful
reviewTime	Time stamp	Not useful

Submission

R or Python Code or Notebook that has a Function that takes input as a single ASIN (Movie ID) and does the following actions:

- Subset the data for the given ASIN
- Basic Pre-Processing of the data (all text processing methods like tokenization, stop word removal, stemming etc.)
- Apply sentiment analysis on each review (you can also do sentence wise if needed) and classify the overall review as positive and negative.
- Plot a word cloud for all the positive reviews
- Plot a word cloud negative for all the negative reviews
- Return a final sentiment score for the movie.
 - o Example:

- 60% Positive and 40% Negative
- 70% Negative and 30% Positive
- Any other interesting exploratory analysis will be an added asset.

The function has

- **Input:** ASIN ID
- **Output:** 2 word clouds (1 for positive words in the review, 2 for negative words in the review)

Example:

Positive words	it!!, turns!, blown, good!, wow!, down!, amazingly, book!, book!!, next!, read!, movie!, brilliantly, masterful, awesome, superb, fabulous, it!, wait, wonderfully, highly, turner!, incredible, toes, fantastic, bed, masterfully, thank, prime, loved, favor, !, blew, excellent, master, time!, chilling, amazing, crafted, end!, roller, story!, seat, loves, edge, gift, twice, beautiful, insightful, layers, constantly, wow, keeps, night, coaster, pieces, terrific, sleep, genius, predict, unpredictable, morning, thrilling, reading!, intricate, complex, fascinating, funny, immediately, enjoys, woven, late, unfolds, minute, love, beautifully, brilliant, surface, perfect, witty, till, fast-paced, intense
Negative words	waste, poorly, wasted, worst, ridiculous, dumb, badly, awful, skipped, horrible, worse, depressing, pathetic, terrible, stupid, silly, boring, annoying, unrealistic, bother, poor, contrived, unbelievable, stuck, miserable, profanity, implausible, selfish, sorry, mistake, unlikeable, unlikable, struggled, bothered, mess, quit, hated, death, book?, shallow, negative, disliked, cliché, dull, care, really?, annoyed, suspend, pass, sadly, cared, skip, holes, stopped, plain

Rules:

- You are free to use R or Python.
- Time duration for the round is 1 hour.
- At the end of 1 hour you can send us the code file or notebook file for review
- **Note:** Please send us the code even if the entire functionality is not present.
- **Evaluation Criteria:** Well-documented and readable/productionized code, Approach towards the problem.