DATA SCIENCE MIDTERM REPROT

Group 7
103502004
103502533



GOAL

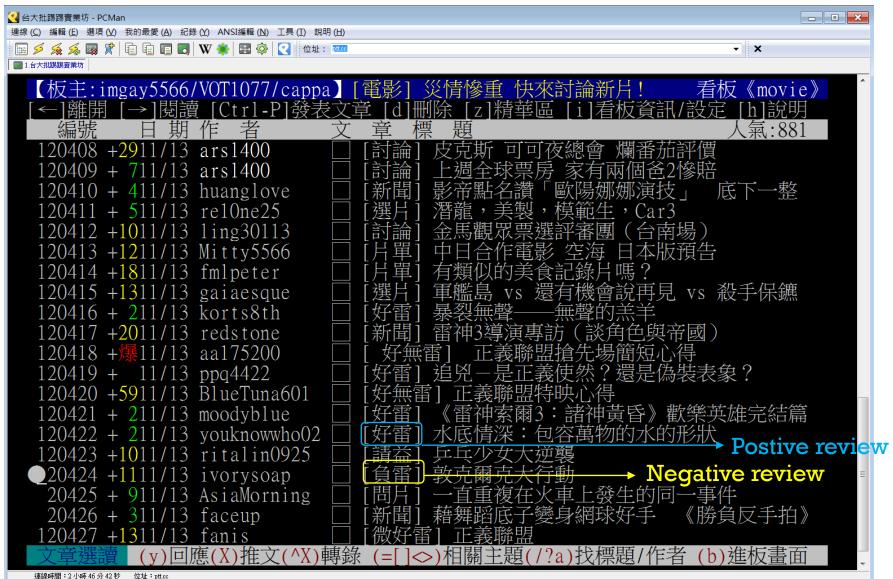
- In order to know whether a high-popularity movie is really a good movie.
- We will analyze the relation between the reception from PTT Movie forum and the box office.
- Furthermore, we will extract the most word frequency of occurrence of the highest-rating movie.



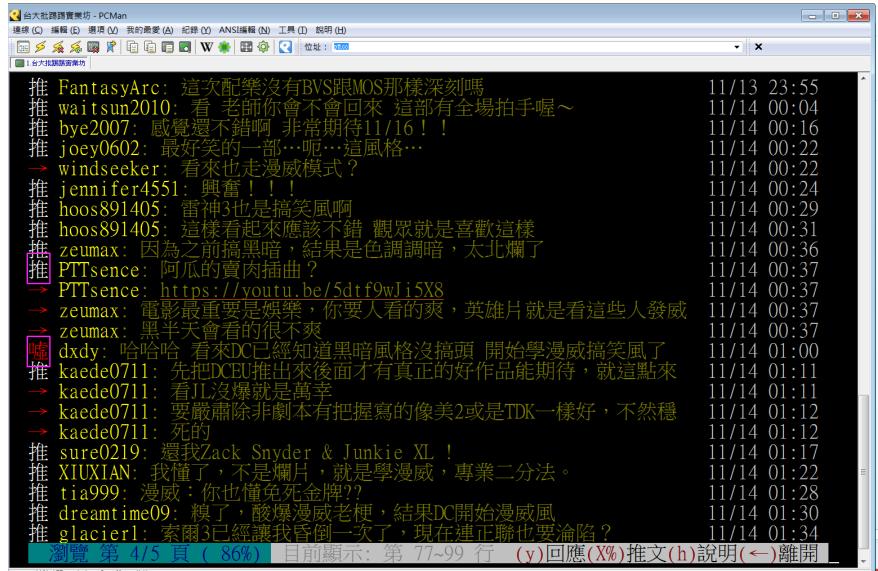
SOLUTION

- We will use web crawler to catch the data in the Movie forum and choose the review articles wrriten by users.
- According to the reviews of articles and comments to score each movie. Later, compare scores and box office, and evaluate them to make a chart.
- To extract words, we may try some word segmentation tools and calculate frequencies to get result.











Thanks for listening!

