

This project was constructed with the intention of finding “shared features” in video game reviews, simplifying and expediting the process of receiving feedback. These shared features are the topics I’ve spent each notebook producing and utilizing, ultimately to provide concise summaries for dozens of these Amazon video game reviews through a model formed and optimized here.

5-core 2018 Amazon video game reviews: nijianmo.github.io/amazon/index.html

- **Notebook 1** loads in the .json file, where I’ve dropped the columns that would not be immediately useful to me. I’ve removed rows with NaN values, which was not a major issue considering there are approximately 500k reviews and only about 250 of them with columns containing null values. I then attempted to find multi-word expressions using spacy and a rule-based matcher to locate domain-specific terms, like “replay value” and “open world.” I extracted candidate terms, computed their c-values, then ranked them all in descending order with respect to theta.
- **Notebook 2** creates a subsample of reviews for estimating an LDA topic model, where hyperparameters were adjusted to optimize that model. There is a training loop involved that essentially takes the difference between a result and the result preceding it, and checks to see if it meets the tolerance value set. I found terms associated with each topic, visualized the topic model with LDAvis, found documents that best represented each topic, and constructed a word cloud for one of these topics.
- **Notebook 3** examines reviews at specific ratings, identifying a few game titles (not just asin values). I found the highest-frequency ratings for certain games, tokenized the reviews of those games with the terms I compiled in notebook 1, and applied the topic model to those specific games. Much like notebook 2, I found the topics most commonly-associated with individual reviews, determined how many times a topic occurred in a specific game’s total Amazon reviews, and did the same for exclusively 1-star reviews of a specific game. I finished off with following that process for specific games’ 5-star reviews.

Regarding the final results, most of the analysis was done on one game, Diablo III. Some top topics that came to be, that might be interesting, is that topic 48 focuses more on learning gameplay and things to do from the gamer’s perspective. It seems that one of the better reasons for Diablo III to earn a higher rating is over its balance in how gameplay mechanics progress. Overall, it’s a divisive game, where its 5- and 1-star ratings have the highest occurrences. That is why you’ll find in these topics that Diablo III being made by Blizzard is one of the reasons why it receives high scores—fans are fans and enjoy Blizzard’s work on faith, no questions asked. On the flip side, because the game is made by Blizzard, that actually leads to many people having preconceived notions over the game being disappointing (which does affect reviews and ratings).

As other details I came across that might be interesting, StarCraft II was actually the game with the second-highest number of 1-star reviews. The difference is that StarCraft II actually has a total 5-star count that far exceeds its total 1-star ratings, where its massive playerbase (by 2018) absolutely makes a difference here, and proceeding with analysis without making these checks can lead to poor conclusions right off the bat.