MAL(My Anime List) Sentiment Analysis

Rolando Holmes

Derek Davenport

CSE 621-01: WEB MINING FOR E-CMRC

27 April 2021

Objectives

- We pick 14 well-known Isekai shows*
- Scrape the MAL (My Anime List) site for the user reviews for each show
- Preprocess the review text and analyze it on two dimensions: Polarity (Positive or Negative sentiment) and Subjectivity (Subjective or Objective) score
- Show the correlation in polarity to user review scores
- Perform a clustering analysis for each individual review plotted
- Produce word clouds to visualize the most salient words in each show's reviews

*shows where the main character is transported to a different world

Dataset

- Below are two samples of our data, top image is the scraped data not preprocessed.

- The bottom image is the cleaned / preprocessed data with Polarity and Subjectivity scores calculated

	user_reviews	show
0	\n\n\n \n Re:Zero truly wanted to be something. It shoots for the stars, it tries new things and throws the characters through so much torture and misery that the fantasy-world they are living in resembles more a hell....	Re:Zero
1	\n\n\n \n *Re:minder - This review contains some light spoilers*\n\r\nFrom my perspective, we are living in a time where well crafted anime with a myriad of originality and detail are unfortunately being released at a ...	Re:Zero
2	\n\n\n \n "The only thing worse than dying once is having to die again."\n\r\nThat was the title of a wonderful email from Crunchyroll I received about a month and a half ago to advertise "Re: Zero - Starting Life in a...	Re:Zero
3	\n\n\n \n Hi all, this is my first review on MAL. Please bare with me, but I feel like I should share my opinion on this and why I gave it the score I did. I wouldn't say I'm the most hardcore anime fan, but I have def...	Re:Zero
4	\n\n\n \n Ahh, the reviewer. The first line of defense to quell the flames of irrational thought and often the bearer of bad news for those of the general public. They're the Buzz Killingtons that take it upon themselv...	Re:Zero

	user_reviews	show	polarity	subjectivity
0	rezero truly wanted to be something it shoots for the stars it tries new things and throws the characters through so much torture and misery that the fantasyworld they are living in resembles more a hell they boil in this hell furning with anger ...	Re:Zero	0.080578	0.574887
1	reminder this review contains some light spoilers from my perspective we are living in a time where well crafted anime with a myriad of originality and detail are unfortunately being released at a frequently declining rate because of this many i...	Re:Zero	0.190064	0.519002
2	the only thing worse than dying once is having to die again that was the title of a wonderful email from crunchyroll i received about a month and a half ago to advertise re zero starting life in another world i ignored it at first and simply car...	Re:Zero	0.145173	0.478223
3	hi all this is my first review on mal please bare with me but i feel like i should share my opinion on this and why i gave it the score i did i wouldnt say im the most hardcore anime fan but i have definitely seen my fair share of shows and am p...	Re:Zero	0.104430	0.524721
4	ahh the reviewer the first line of defense to quell the flames of irrational thought and often the bearer of bad news for those of the general public theyre the buzz killingtons that take it upon themselves to blow the no fun whistle and send th...	Re:Zero	0.062071	0.491544

Background Methodology
Web Scraping

We make use of the '**Beautiful Soup**' Package which allows us to scape webpages looking for specific elements such as <div> elements which contain our target data i.e., all the user reviews on a given page.

Background Methodology
Preprocessing our data

- For preprocessing we take our scraped data from the MAL (My Anime List) site and build a DataFrame

- We apply some regex cleaning to each of the individual user reviews removing punctuation, stop-words, and non-words.

- We make use of a Document Term Matrix or Frequency Matrix, creating our own custom list of stop-words to build a word cloud displaying the most frequent words in the reviews for a show

Background Methodology
Calculating Polarity and Subjectivity

Below is an arbitrary sample of the lexicon for english adjectives we use to calculate our user reviews. This dictionary is full of hand-tagged words that are used in conjection with a 'Naïve Bayesian' classifyer to not only classify a review as either positive or negative, but it also helps produce an average score Polarity and Subjectivity score.
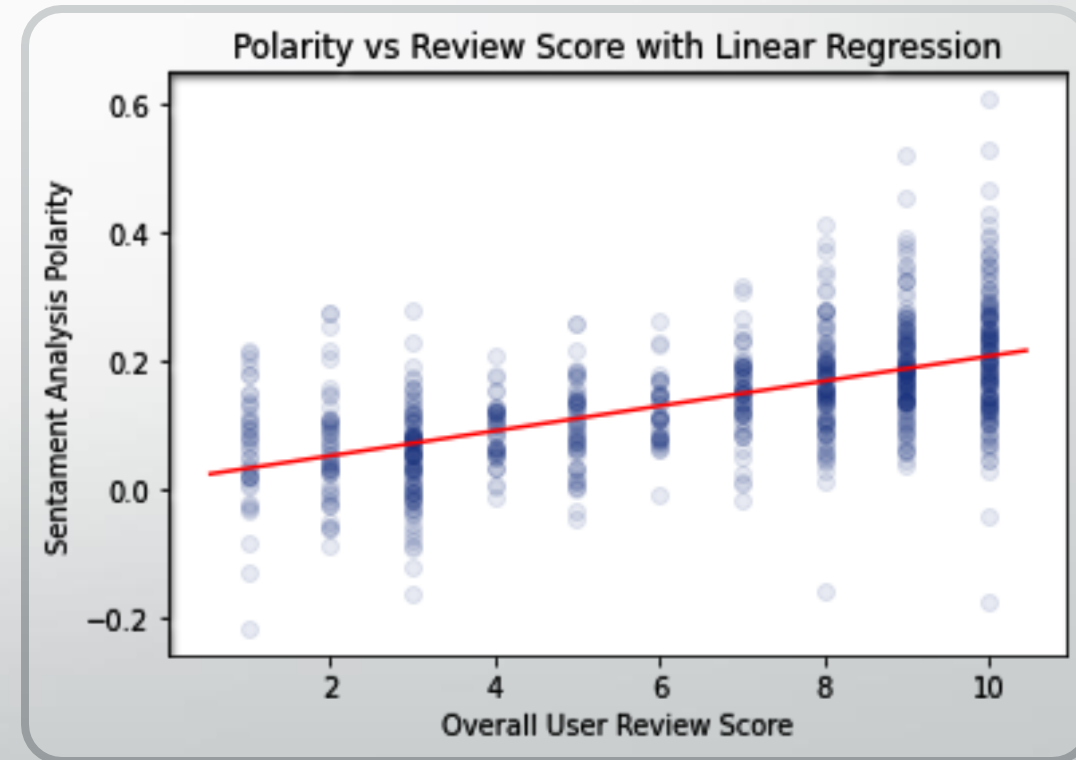
We have also linked the actual dictionary used by the 'Text Blob' package below

Source:https://github.com/clips/pattern/blob/master/pattern/text/en/en-sentiment.xml

```xml
<word form="amusing" wordnet_id="a-01344485" pos="JJ" sense="providing enjoyment" polarity="0.7" subjectivity="1.0" intensity="1.0" confidence="0.9" />
<word form="anger" wordnet_id="v-1785971" pos="VB" sense="make angry" polarity="-0.7" subjectivity="0.2" intensity="1.0" confidence="0.9" />
<word form="absolute" wordnet_id="a-00094069" pos="JJ" sense="not capable of being violated or infringed" polarity="0.0" subjectivity="1.0" intensity="1.0" confidence="0.9" />
<word form="absolute" wordnet_id="a-00719328" pos="JJ" sense="not limited by law" polarity="0.0" subjectivity="1.0" intensity="1.0" confidence="0.9" />
<word form="alcoholic" wordnet_id="a-01158596" pos="JJ" sense="characteristic of or containing alcohol" polarity="0.0" subjectivity="0.0" intensity="1.0" confidence="0.9" />
<word form="algid" cornetto_synset_id="d_a-9370" wordnet_id="a-01251958" pos="JJ" sense="chilly" polarity="-0.4" subjectivity="0.9" intensity="1.0" confidence="0.8" />
```

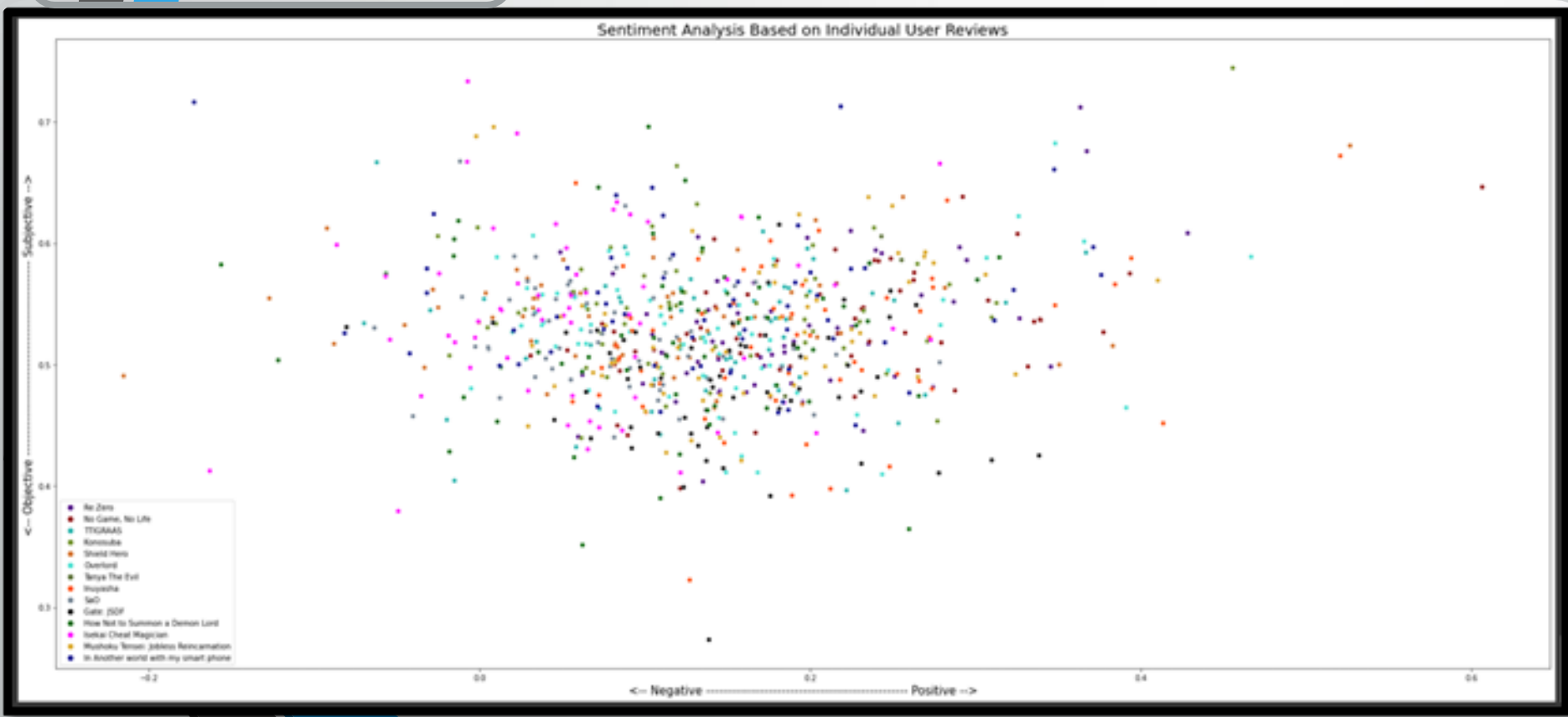# Polarity Vs Review Score Regression Analysis

- We did a Pearson correlation coefficient analysis across the entire dataset irrespective of show.

- In our dataset, we would expect the user's rating of a show to highly correlate with the generated polarity score if the polarity score is a reliable metric.

- The Pearson correlation coefficient between overall score and polarity was computed to be 0.578. As expected, there is a strong positive correlation. This confirms for us that we can have confidence that generated polarity score is a good measure of the text's positive or negative sentiment.

Polarity vs Review Score with Linear Regression
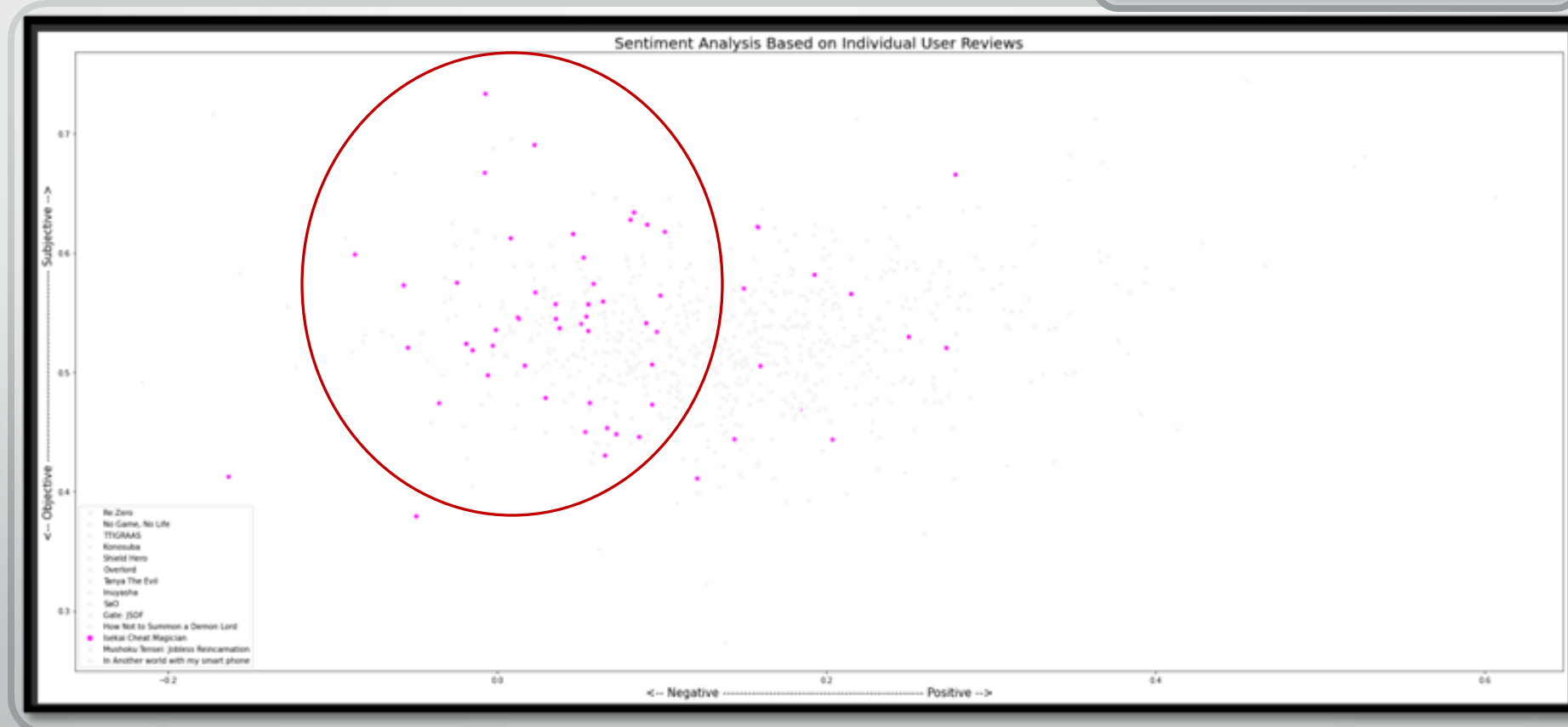
Visualizing Our Data Unfiltered user reviews

Here is a plot of all reviews with polarity on the x axis and subjectivity on the y axis. Review points are color-coded by show.
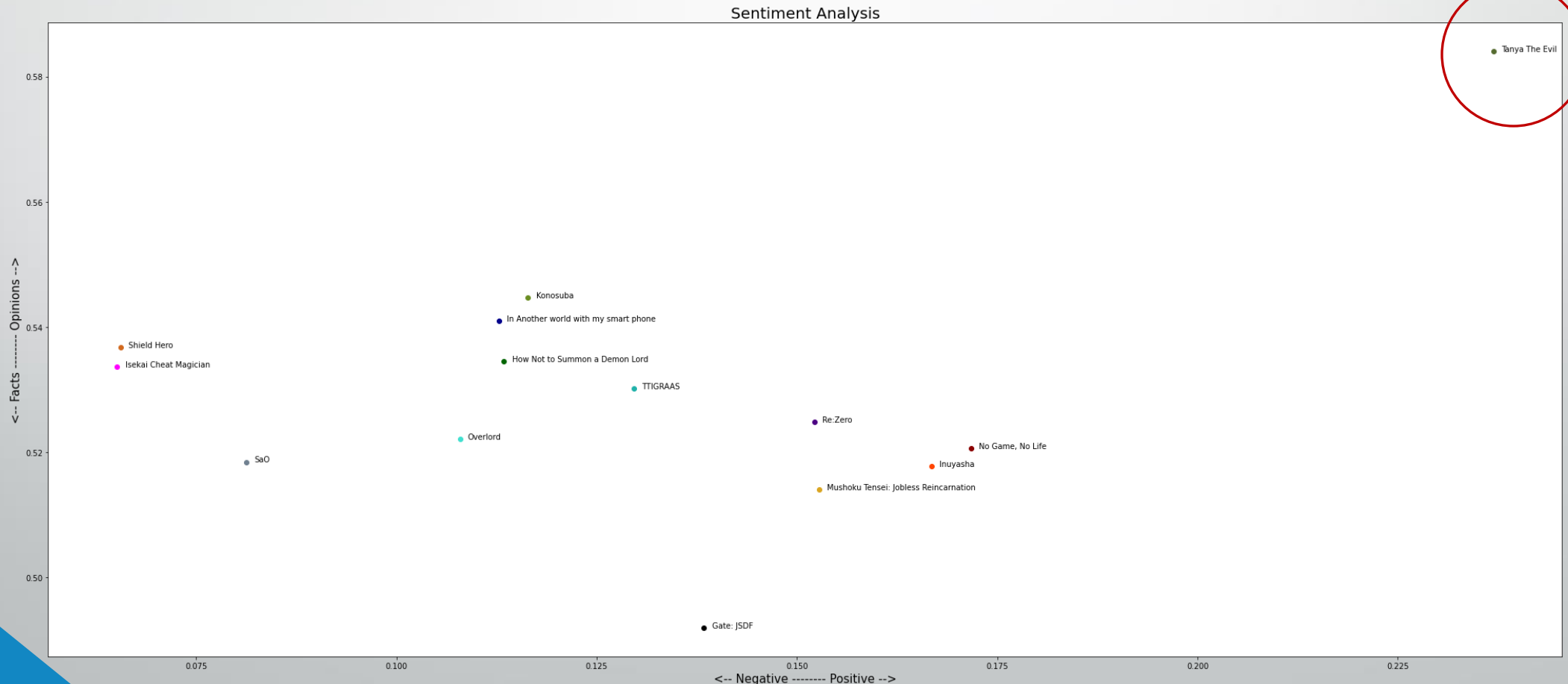
Visualizing Our Data Filtered user reviews

Re:Zero
No Game, No Life
TTIGRAAS
Konosuba
Shield Hero
Overlord
Tanya The Evil
Inuyasha
SaO
Gate: JSDF
How Not to Summon a Demon Lord
Isekai Cheat Magician
Mushoku Tensei: Jobless Reincarnation
In Another world with my smart phone

Here is the same plot but with only one show (***Isekai Cheat Magician***) colored. You can oberseve that its reviews tend towards the Negative-Subjective quadrant of the plot.

Sentiment Analysis Based on Individual User Reviews

Overall Sentiment Analysis

Here is the visualization for the overall sentiment of each show over all user reviews. One particular obeservation to note is the show "Tanya the Evil" which had significantly fewer reviews; its overall sentiment becomes an outlier given that the few who reviewed it subjectively liked it. Given more data we expect some regression toward the mean on both dimensions.

Extending our analysis to Word cloud visualization

- We create word clouds to show the most frequent words used in a show's reviews.

- Words common to most shows (e.g. "anime") were removed so that only the most defining words are shown.

- Easy way to visualize the topics reviewers are eager to discuss