

“WHEN WILL COMMENT CULTURE RETURN FROM THE WAR?”

DATA ANALYSIS OF READER ENGAGEMENT WITH FANFICTION

JOANNE RAMADANI
M.S. IN DATA ANALYSIS & VISUALIZATION
CSC 84040 FALL 2019



TOPICS COVERED IN THIS PRESENTATION

- 💡 Hey wait, hold on. What is fanfiction?
- 💡 Significance of topic
- 💡 Dataset and tools
- 💡 Preprocessing
- 💡 Methodology
- 💡 Challenges
- 💡 Future steps
- 💡 Recommended Reading

A FEW DEFINITIONS BEFORE WE PROCEED:

- **Fandom** is a general word for a collection of fans. For example: “The Harry Potter fandom was rioting about J.K. Rowling’s newest tweets about the book series.”
- **Fanfic** is what fanfiction is usually shortened to, or just “fic”.
- **Canon** is the original published media. Compare with “headcanon”, which is a fan’s personal conception of events. Can be used as a verb (i.e. “canonized”), derivative from, you guessed it: the Bible.
- **Ship** is short for “relationship”, and indicates that a fan likes the idea of two or more characters being together. This can be used as either a noun or a verb: “Oh, I love that ship!” or “I ship the two of them so much, I can’t wait for them to get together.”
- **AO3** is the abbreviation for Archive of Our Own, which I’ll talk more about in a bit.
- **FF.net** is the general abbreviation for Fanfiction.net, which I’ll also go over soon.

WHAT'S FANFICTION?

It can look like this when listed:



[The Time Harry Stole Lucius's Wife](#) > by [redwolf96](#) [reviews](#)

Harry has just lost his godfather and he is ready for a break from everything before he loses his mind. Unfortunately for him, now he has to deal with being the lord of two houses and learning what that means while also coming to terms with a family secret that's been hidden for years. A surprise marriage contract makes things even more interesting.

Rated: T - English - Romance - Chapters: 5 - Words: 31,495 - Reviews: 46 - Favs: 185 - Follows: 326 - Updated: Dec 3 - Published: Sep 21 - [Harry P., Narcissa M.]



[Ruínas do Tempo](#) > by [Gabriele Snape](#) [reviews](#)

Severus Snape ao perder o amor de sua vida, Hermione Granger, decide usar o Vira-tempo escondido na Mansão Malfoy. Ao retornar para o ano de 1967 tem a possibilidade de mudar a sua própria história e daqueles que amava. Até que ponto isso interferirá nas regras do tempo e espaço? Plágio é crime! Originalmente publicada no Spirit Fanfiction

Rated: K+ - Portuguese - Romance/Drama - Chapters: 32 - Words: 76,531 - Reviews: 6 - Favs: 3 - Follows: 4 - Updated: Dec 3 - Published: Oct 13



[Harry Potter and the Book of Lost Souls](#) > by [OrchidSonata](#) [reviews](#)

Harry Potter thinks he is an ordinary boy - until he meets an honest-to-goodness witch! Soon finding himself enrolled in Hogwarts School of Witchcraft and Wizardry, his dull life is replaced by a magical world of infinite possibilities! But not all is well at Hogwarts; a hidden enemy lurks the vast corridors of the school - one that would see young Harry lose his very soul!

Rated: T - English - Romance - Chapters: 5 - Words: 33,257 - Reviews: 12 - Favs: 6 - Follows: 19 - Updated: Dec 3 - Published: Oct 15 - Harry P., Regulus B., Ernie M., OC



[Portrait of heroes](#) by [Lady Phoenix AL](#) [reviews](#)

They are now known as heroes as those who opposed the darkness and didn't give up even when the odds were against them but it wasn't always like that. For many years they were just other names in a long list of people that died because of Voldemort and/or his Death Eaters nothing more nothing less.

Rated: T - English - Tragedy/Friendship - Chapters: 1 - Words: 550 - Reviews: 1 - Favs: 2 - Follows: 1 - Published: Dec 3 - Complete



[Severus Rogue, un homme si charmant](#) > by [MinnieMey](#) [reviews](#)

Severus Rogue n'est pas mort après la chute de Voldemort. Il continue sa vie de professeur à Poudlard, mais sans plus aucun but et avec un grand vide dans son cœur. La pétillante, intrépide mais maladroite Sarah Turner réussira-t-elle à faire chavirer le cœur de l'espion le plus sinistre mais non moins sexy que le monde sorcier n'ait jamais connu ?

Rated: T - French - Romance - Chapters: 15 - Words: 65,500 - Reviews: 57 - Favs: 13 - Follows: 21 - Updated: Dec 3 - Published: Oct 16 - Severus S., OC



[Lie To Me \(Please\)](#) by [DobbyRocksSocks](#)

Blaise protects Theo from everything. Theo doesn't understand why nobody else can see that. Warnings inside.

Rated: T - English - Angst/Drama - Chapters: 1 - Words: 2,543 - Favs: 1 - Published: Dec 3 - [Blaise Z., Theodore N.] - Complete



[Harry Potter, Self-Insert](#) > by [15Redstones](#) [reviews](#)

I am a 18 year old fanfiction writer, schoolkid and computer nerd. I also just woke up in the body of 8 year old Harry Potter. How did I get here? No idea. What am I going to do? Priority one, survive. Priority two, drag this world into the 21st century.

Rated: T - English - Humor/Parody - Chapters: 14 - Words: 19,760 - Reviews: 61 - Favs: 251 - Follows: 346 - Updated: Dec 3 - Published: Feb 2 - Harry P., Hermione G., Sirius B., Remus L.



[Be My Anchor?](#) by [Andromeda Psyche Jackson](#) [reviews](#)

Biarkan aku menjadi tempatmu bersandar. Izinkan aku untuk melihat dirimu yang sebenarnya. Aku benar-benar ingin bersamamu (James x Fem! Severus)

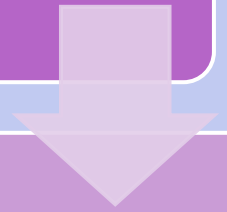
Rated: T - Indonesian - Romance - Chapters: 1 - Words: 2,915 - Reviews: 1 - Published: Dec 3 - [James P., Severus S.] - Complete

WHAT THE
PROCESS OF
ENGAGING
WITH MEDIA
ONLINE
LOOKS LIKE:

Published work

Media
consumption

Fanworks (Fanart
and fanfiction)



WHERE DO PEOPLE POST FANFICTION?

- There are two main categories of places where people post fanfiction:
 - Their personal blogs, for example:
 - LiveJournal
 - Dreamwidth
 - Tumblr
 - Archive sites
 - The Leaky Cauldron (specifically for Harry Potter fanfiction—and Harry Potter fans are *prolific*)
 - Fanfiction.net (generally used by younger writers)
 - Archive of Our Own (now one of the most widely used sites to post fanfiction)

COMMENT CULTURE

Comment Culture : FanFiction - Reddit

<https://www.reddit.com> › [FanFiction](#) › [comments](#) › [comment_culture](#) ▼

Mar 19, 2018 - r/**FanFiction**: A supportive community for writers, readers, and reccers to talk about and share **FanFiction**. ... **Comment Culture**. I was just reading a story, and the ...

Changes in Fandom/Feedback Culture?

Jul 7, 2019

Is "**Comment Culture**" dying off?

Oct 3, 2019

Differences between **comment culture** for FF and A03

Oct 5, 2019

Comments and Reviews

Jul 14, 2019

More results from www.reddit.com

“ Similar to Lam's (2000) research on online chat rooms, Black (2005, 2008, 2009) focused on ELLs' voluntary participation in the online fan fiction community fanfiction.net, which involves reading, peer reviewing, collaborative writing...The fanfiction site provides ELLs with **access to and affiliation with writing and interacting in English**...the peer-review practices show a strong tendency toward maintaining community relationships by tempering critique of form with genuine enthusiasm for content or rhetorical effect, strongly discouraging hostile feedback, and attending to the expressed needs of the author. This marked emphasis on constructive criticism and lack of tolerance for hostile and deliberately insulting feedback ensures a safe, accessible space for ELLs and others to write; it helps ELLs "establish a legitimate social position within the community as accomplished writers and promotes their continued affiliation with writing in English.”

An Excerpt from Guofang Li's "Literacy Engagement Through Online and Offline Communities Outside School: English Language Learners' Development as Readers and Writers"

YOU MIGHT BE WONDERING: WHY DOES FANFICTION
ENGAGEMENT MATTER?

DATASET & TOOLS

- I used a dataset from Kaggle containing the entirety of the contents from the website [fanfiction.net](https://www.fanfiction.net), scraped in November 2018.
- There are nine folders representing media categories with anywhere from 10 to 2500 csv files in them.
- In total, there are approximately five million rows of fanfiction.
- The original dataset has 17 features and they are as follows:
 - Section, subsection, story (url), author (url), story_id, author_id, language, category, rated (e.g. "T" for Teens), chapter_count, review_count, word_count, fav_count, follow_count, published (unix timestamp), updated (unix timestamp), description (summary of the story)
- Tools used are Microsoft Visual Studio Code, Python, Pandas, Numpy, Seaborn, and Matplotlib

PREPROCESSING

- This step contains all of my cleaning, feature reduction, and spam removal
- Dataset cleaning consisted of consolidating all of the csv files mentioned previously (a total of around 8000 files) into one csv file, which created its own set of problems, as the resulting file is 1.59 GB
- I identified duplicates in the dataset (most of which were the headers), kept only one row, and turned that row into the dataframe's headers
- Then I dropped the following features:
 - Story (redundant due to story_id being a unique identifier)
 - Author (redundant due to author_id identifying the users)
 - Published (unnecessary for the question at hand)
 - Updated (unnecessary for the question at hand)
 - Description (Unnecessary for the question at hand)

PREPROCESSING 2: DEALING WITH SPAM



[rddd](#) by [alukena](#) [reviews](#)

dddf

Rated: K+ - Spanish - Horror/Parody - Chapters: 1 - Words: 3 - Reviews: 1 - Published: Mar 27, 2018 - Loki - Complete



[Scooby Doo goes on holiday](#) by [ArtistFire18](#) [reviews](#)

It was fine day in the Hidden Earth which was a little island of the coast of Rhode Island, the friends were staying there for the weekend, thought unknown to them was a legend monster to have you on the island for centuries.

Rated: K+ - English - Friendship/Sci-Fi - Chapters: 1 - Words: 43 - Reviews: 1 - Favs: 1 - Follows: 1 - Published: Mar 24, 2018 - Complete



[法克鱿](#) by [0jutm upStsjssskl](#) [reviews](#)

yex 0 jut my wx w 0 jut my wx0 jut my wx0 jut my wx w0 jut my wx w w w 葡萄酒5句太累了可乐噜噜噜不不把图图8部图库聚聚不不不8部不哭噜噜噜噜噜噜他他他噜噜噜V5图啊5部路路通噜噜噜噜噜噜他太可惜了拒绝我了 yex 0 jut my wx w 0 jut my wx0 jut

Rated: K+ - Chinese - Hurt/Comfort - Chapters: 1 - Words: 6 - Reviews: 3 - Published: Feb 13, 2018 - Complete



[Rival or love](#) by [Kyungie-do24](#) [reviews](#)

Typo bertebaran JLEBBBBBB "Ahhhhh. " JLEBBBBB JLEBBBBBB "Uhhh baby inihhh nikmathhh sekali .." ucap namja tampan itu di selah-selah kegiatan panasnya .. "Ahhhh ahhhh ahhh morehhhh uhhhh hBastardhh fuckhhhh uhhhh .." desahan yeojja cantik itu menikmati goyangan sang namja ... Namja itu pun mengeluarkan smriknya dan memegang bokong yoejja itu dan mempercepat goyagannya

Rated: K+ - Indonesian - Chapters: 1 - Words: 2,578 - Reviews: 1 - Published: Feb 4, 2018



[fggg](#) by [lbx-yf](#) [reviews](#)

dfggggggggggggggg

Rated: K+ - English - Humor/Drama - Chapters: 1 - Words: 1 - Reviews: 1 - Published: Apr 28, 2017



[Test1](#) > by [lbx-yf](#)

Ayuiop

Rated: K - English - Romance - Chapters: 3 - Words: 18 - Updated: Jun 16, 2017 - Published: Mar 24, 2017 - Complete



[Story gagal](#) by [Nadyaleia10](#) [reviews](#)

Salah

Rated: K+ - Latin - Poetry/Drama - Chapters: 1 - Words: 2 - Reviews: 2 - Published: Apr 24, 2017 - Number Zero - Complete



[수원풀싸롱신세경keenO1O 2727 2778 수원매직미러초이스 수원풀싸롱신세경,수원야구장쪽쪽뽕뽕,수원풀싸롱신세경하태하태](#) by [megakt114gmail.com](#)

수원풀싸롱신세경keenO1O 2727 2778 수원매직미러초이스 수원풀싸롱신세경,수원야구장쪽쪽뽕뽕,수원풀싸롱신세경하태하태 수원야구장미러룸 가격 매직 룸 미러 수원풀싸롱 수원풀싸롱신세경keenO1O 2727 2778 수원매직미러초이스 수원풀싸롱신세경,수원야구장쪽쪽뽕뽕,수원풀싸롱신세경하태하태 수원야구장미러룸 가격 매직 룸 미러 수원풀싸롱 수원풀싸롱신세경keenO1O 2727 2778 수원매직미러초이스 수원풀싸롱신세경하태하태 수원야구장미러룸 가격 매직 룸 미러 수원풀싸롱 수원풀싸롱신세경keenO1O 2727 2778 수원매직미러초이스 수원풀싸롱신세경,수원야구장쪽쪽뽕뽕,수원풀싸롱신세경하태하태 수원야구장미러룸 가격 매직 룸 미러 수원풀싸롱

Rated: T - Albanian - Chapters: 1 - Words: 58 - Published: Mar 18

METHODOLOGY

- In order to measure engagement, I am using k-means clustering to show where there is more engagement:
 - How many words will get more reviews vs favs vs follows?
- Because the dataset is so big, I am using mini batch k-means clustering to reduce the batch size to one million randomly selected points.
- The algorithm is fairly simple, and I will be executing it in three separate instances, using wordcounts on the y axis, and these three features on the x axis:
 - Reviews
 - Favs
 - Follows
- These will reveal trends in what the ideal wordcount is for engagement, and then if I can pull out points in the data, I might get more insights.

CHALLENGES

- Filtering out spam fics—I didn't realize just how many spammers there are on fanfiction.net! My parameters are chapter & word count: a lot of spam fics have several chapters and a very low word count, or one chapter with almost no words.
- Wrangling with **my ancient foe Python**
- (That previous bullet point is very short, but constitutes the majority of my issues in this project)
- Dealing with the data's enormity (which I figured out eventually, but expect to run into again while doing the mini batch k-means)

FUTURE STEPS

- Implementing the mini batch k-means
- Analyzing results
- If there's time, I want to figure out which ratings get higher engagement—however, if I don't have time, wordcount will be enough to figure out a reader's attention span for optimal engagement.
- Additionally, taking my algorithm and applying it to consolidations of each category would also theoretically yield interesting results, but I wanted to attempt mining “big data” for this project.

RECOMMENDED READING: WHY THIS MATTERS

- Thomas, Angela. "Fan Fiction Online: Engagement, Critical Response and Affective Play through Writing." *Australian Journal of Language and Literacy*, vol. 29, no. 3, 2006, pp. 226–239.
- Garcia, Antero. "Making the Case for Youth and Practitioner Reading, Producing, and Teaching Fanfiction." *Journal of Adolescent & Adult Literacy*, vol. 60, no. 3, 2016, pp. 353–357.
- Li, Guofang. "Literacy Engagement Through Online and Offline Communities Outside School: English Language Learners' Development as Readers and Writers." *Theory Into Practice*, vol. 51, no. 4, 2012, pp. 312–318.
- Vinney, Cynthia, and Karen E. Dill-Shackleford. "Fan Fiction as a Vehicle for Meaning Making: Eudaimonic Appreciation, Hedonic Enjoyment, and Other Perspectives on Fan Engagement With Television." *Psychology of Popular Media Culture*, vol. 7, no. 1, 2018, pp. 18–32.