

# Inside the heads of Putnuts

An analysis of the [Only Murders in the Building](#) TV show subreddit

Milindi Kodikara

## 1 Introduction

[Only Murders in the Building](#) is a cosy, whodunit<sup>1</sup> murder mystery TV show where each season Charles, Oliver and Mabel who live in The Arconia in New York City solve murders that happen in and around the building.

This paper contains an exploration and analysis of Reddit posts and comments from the [OnlyMurdersHulu](#) subreddit for the [Only Murders in the Building](#) TV show. The analysis is conducted on posts and comments spanning from the announcement of the new season to the week following the season premier of the fourth season of the show. This analysis was conducted to determine the trending concepts, topics, and feelings towards the [Only Murders in the Building](#) TV show.

Upon analysis of this subreddit, it was found that the viewers were excitedly and eagerly awaiting the release of the new season. The general sentiment regarding this show is positive but with a dramatic shift in engagement in the month of August, it can be clearly seen that the viewers were devastated by the reveal of the death of a beloved recurring character in the first episode. A general theme of theorising about [who actually did it](#) (the crime), figuring out the clues along with praising the cast and creators can be seen. These were found to be the feelings towards the show and the main topics of discussion using sentiment analysis techniques and topic modelling.

## 2 Data Collection

The [newest ~1000 posts](#) from the [OnlyMurdersHulu](#) subreddit have been explored in this project to conduct a comprehensive analysis. The posts collected were from the 4th October, 2023 to 3rd September, 2024; one week after the release date of the supposed final season.

The PRAW API, the python reddit API wrapper, was used within this project for data collection. This is as our system for analysis, [Putnuts](#), is written in Python and PRAW is a widely used, simple and efficient API wrapper to build tools using Reddit data.

Overall, [915](#) new posts and a total of [20012](#) comments were collected. The date range of the posts were from 14th October, 2023 to 3rd September, 2024. Posts and comments between

---

<sup>1</sup> A type of detective fiction where the killer of the crime is uncovered via puzzles and clues

4th October, 2023 and 19th October, 2023 were cut off due to reddit limiting the maximum number of posts that can be requested to be 1000 posts.

For example reddit post, *Figure 1*, the data collected into a dataframe were as follows:

'Post': Reddit post title and body

'Num\_comments': Number of comments for the post

'Author': Name of the author of the post; if the author is not found the key word 'None' was added

'UTC\_Date': The post creation date in the UTC format

'Date': The UTC\_Date in the format of dd/mm/yyyy

'Upvote\_ratio': The upvote ratio of the post

r/OnlyMurdersHulu · 2 days ago  
hannahlemp Where are the balls, Howard? ...

## Season 4 - Episode 2: "Gates of Heaven" (Post Episode Discussion Thread)

**S4 Discussion**

Welcome to [r/OnlyMurdersHulu](#)'s official Only Murders in the Building Season 4 post episode discussion thread.

Use this thread to discuss **Season 4: Episode 2: "Gates of Heaven"** once you have finished watching the **episode** which premiered September 3rd at 12:00 am EST.\*

If you are currently watching Episode 2, please be sure to check out the relevant Live Discussion Thread before commenting here, so you don't get spoiled.

A reminder that the sub will be locked for new posts for 24 hours following the episode's release. [More information here.](#)

A reminder on spoilers:

- Keep spoilers out of any post titles
- Do not share spoilers from future episodes under this post
- Spoiler tag the post/content
- Kindly correct any users posting spoilers and message a mod if needed
- For comments that contain spoilers utilize: >! text !< but remove the spaces
  - Ex. I think the dog did it

How are you liking season 4 so far? Who are your main suspects? What do you hope to see next week?

See you next week for new Olimabel (the Charles is silent) adventures.

**NB:** A new rule regarding the criticism of the show's actors has been added. Any comments/posts hating on people's appearances, or unjustifiably hating on someone's acting abilities will be removed.

\*(Sept. 2nd 9pm PST on Hulu, Sept. 3rd 8am BST on Disney+, 9am CEST on Disney+, 3pm PHT on Disney+, 5pm AEST on Disney+ - comment if you would like your timezone added)

↑ 121 ↓ Comment 918 Share

*Figure 1: Example reddit post*

### 3 Data Pre-processing and Exploration

Upon exploration of a subset of the data extracted, pre-processing of data was conducted in order to improve accuracy and reliability of the data to be analysed.

Season 4 – Episode 2: "Gates of Heaven" (Post Episode Discussion Thread) Welcome to r/OnlyMurdersHulu's official Only Murders in the Building Season 4 post episode discussion thread.

Use this thread to discuss \*\*Season 4: Episode 2: "Gates of Heaven"\*\* \*\*once you have finished watching the episode\*\* which premiered September 3rd at 12:00 am EST.\\*

If you are currently watching Episode 2, please be sure to check out the relevant Live Discussion Thread before commenting here, so you don't get spoiled.

A reminder that the sub will be locked for new posts for 24 hours following the episode's release. [More information here]([https://www.reddit.com/r/OnlyMurdersHulu/comments/1ewm86s/only\\_murders\\_in\\_the\\_building\\_season\\_4\\_one\\_week\\_to/?utm\\_source=share&utm\\_medium=web3x&utm\\_name=web3xcss&utm\\_term=1&utm\\_content=share\\_button](https://www.reddit.com/r/OnlyMurdersHulu/comments/1ewm86s/only_murders_in_the_building_season_4_one_week_to/?utm_source=share&utm_medium=web3x&utm_name=web3xcss&utm_term=1&utm_content=share_button)).

A reminder on spoilers:

- \* Keep spoilers out of any post titles
- \* Do not share spoilers from future episodes under this post
- \* Spoiler tag the post/content
- \* Kindly correct any users posting spoilers and message a mod if needed
- \* For comments that contain spoilers utilize: > ! text ! < but remove the spaces
  - \* Ex. >!I think the dog did it!<

How are you liking season 4 so far? Who are your main suspects? What do you hope to see next week?

See you next week for new Olimabel (the Charles is silent) adventures.

\*\*\*NB\*\*\*: A new rule regarding the criticism of the show's actors has been added. Any comments/posts hating on people's appearances, or unjustifiably hating on someone's acting abilities will be removed.

\\*(\\*(Sept. 2nd\* \*9pm PST on Hulu, Sept. 3rd 8am BST on Disney+, 9am CEST on Disney+, 3pm PHT on Disney+, 5pm AEST on Disney+ – comment if you would like your timezone added)\*

Figure 2: Extracted text of the post shown in Figure 1

#### 3.1. Conversion to lowercase

Text was converted to lowercase in order to get an accurate count of the words in the texts as no two words would be considered the same if they misaligned with the capitalisation.

For example, Figure 3 depicts part of the text of Figure 2 upon conversion to lowercase.

season 4 – episode 2: "gates of heaven" (post episode discussion thread) welcome to r/onlymurdershulu's official only murders in the building season 4 post episode discussion thread.

use this thread to discuss \*\*season 4: episode 2: "gates of heaven"\*\* \*\*once you have finished watching the episode\*\* which premiered september 3rd at 12:00 am est.\\*

Figure 3: Reddit post upon conversion to lowercase

#### 3.2. Remove emojis and special characters

Emojis, symbols and pictographs, and characters from other languages were discovered as shown in Figure 4 which were removed.

🎬 Suggested Watch before E2 🎙

Figure 4: Reddit post with emojis

### 3.3. Remove username tags, mentions and links

Username and subreddit tags, mentions and links were removed from text as they do not provide any meaning for the analysis.

A reminder that the sub will be locked for new posts for 24 hours following the episode's release. [More information here]([https://www.reddit.com/r/OnlyMurdersHulu/comments/1ewm86s/only\\_murders\\_in\\_the\\_building\\_season\\_4\\_one\\_week\\_to/?utm\\_source=share&utm\\_medium=web3x&utm\\_name=web3xcss&utm\\_term=1&utm\\_content=share\\_button](https://www.reddit.com/r/OnlyMurdersHulu/comments/1ewm86s/only_murders_in_the_building_season_4_one_week_to/?utm_source=share&utm_medium=web3x&utm_name=web3xcss&utm_term=1&utm_content=share_button)).

A reminder on spoilers:

\* Keep spoilers out of any post titles

Figure 5.1: Snippet of post (Figure 2) with link

a reminder that the sub will be locked for new posts for 24 hours following the episode's release. [more information here]([https://www.reddit.com/r/OnlyMurdersHulu/comments/1ewm86s/only\\_murders\\_in\\_the\\_building\\_season\\_4\\_one\\_week\\_to/?utm\\_source=share&utm\\_medium=web3x&utm\\_name=web3xcss&utm\\_term=1&utm\\_content=share\\_button](#)).

a reminder on spoilers:

\* keep spoilers out of any post titles

Figure 5.2: Snippet of post (Figure 2) with link removed

Season 4 – Episode 2: "Gates of Heaven" (Post Episode Discussion Thread) Welcome to r/OnlyMurdersHulu's official Only Murders in the Building Season 4 post episode discussion thread.

Figure 6.1: Snippet of post (Figure 2) with subreddit tag

season 4 – episode 2: "gates of heaven" (post episode discussion thread) welcome to official only murders in the building season 4 post episode discussion thread.

Figure 6.2: Snippet of post (Figure 2) with subreddit tag removed

```
['season', '4', '-', 'episode', '2', ':', "'", 'gates', 'of', 'heaven', "'", "()", 'post', 'episode', 'discussion', 'thread', ')', 'welcome', 'to', 'official', 'only', 'murders', 'in', 'the', 'building', 'season', '4', 'post', 'episode', 'discussion', 'thread', "'", 'use', 'this', 'thread', 'to', 'discuss', '*', '*', 'season', '4', ':', 'episode', '2', ':', "'", 'gates', 'of', 'heaven', "'", "*", "*", "*", 'once', 'you', 'have', 'finished', 'watching', 'the', 'episode', '*', '*', 'which', 'premiered', 'september', '3rd', 'at', '12:00', 'am', 'est', '\\', '*', 'if', 'you', 'are', 'currently', 'watching', 'episode', '2', ',', 'please', 'be', 'sure', 'to', 'check', 'out', 'the', 'relevant', 'live', 'discussion', 'thread', 'before', 'commenting', 'here', ',', 'so', 'you', "don't", 'get', 'spoiled', '.', 'a', 'reminder', 'that', 'the', 'sub', 'will', 'be', 'locked', 'for', 'new', 'posts', 'for', '24', 'hours', 'following', 'the', "episode's", 'release', '.', '[', 'more', 'information', 'here', ']', '()', 'a', 'reminder', 'on', 'spoilers', ':', '*', 'keep', 'spoilers', 'out', 'of', 'any', 'post', 'titles', '*', 'do', 'not', 'share', 'spoilers', 'from', 'future', 'episodes', 'under', 'this', 'post', '*', 'spoiler', 'tag', 'the', 'post', '/', 'content', '*', 'kindly', 'correct', 'any', 'users', 'posting', 'spoilers', 'and', 'message', 'a', 'mod', 'if', 'needed', '*', 'for', 'comments', 'that', 'contain', 'spoilers', 'utilize', ':', '>', '!', 'text', '!', '<', 'but', 'remove', 'the', 'spaces', '*', 'ex', '.', '>', '!', 'i', 'think', 'the', 'dog', 'did', 'it', '!', '<', 'how', 'are', 'you', 'liking', 'season', '4', 'so', 'far', '?', 'who', 'are', 'your', 'main', 'suspects', '?', 'what', 'do', 'you', 'hope', 'to', 'see', 'next', 'week', '?', 'see', 'you', 'next', 'week', 'for', 'new', 'olimabel', '()', 'the', 'charles', 'is', 'silent', ')', 'adventures', ',', '*', '*', 'nb', '*', '*', '*', ':', 'a', 'new', 'rule', 'regarding', 'the', 'criticism', 'of', 'the', 'show's', 'actors', 'has', 'been', 'added', '.', 'any', 'comments', '/', 'posts', 'hating', 'on', 'people's', 'appearances', ',', 'or', 'unjustifiably', 'hating', 'on', 'someone's', 'acting', 'abilities', 'will', 'be', 'removed', '.', '\\', '*', '(', '*', 'sept', '.', '2nd', '*', '*', '9pm', 'pst', 'on', 'hulu', ',', 'sept', '.', '3rd', '8am', 'bst', 'on', 'disney', '+', ',', '9am', 'cest', 'on', 'disney', '+', ',', '3pm', 'pht', 'on', 'disney', '+', ',', '5pm', 'aest', 'on', 'disney', '+', '-', 'comment', 'if', 'you', 'would', 'like', 'your', 'timezone', 'added', ')', '*']
```

Figure 7: Initial tokenization of an example reddit post (Figure 2)

### 3.4. Tokenization

Tokenization is the process of splitting a sequence of words into individual tokens. This was achieved using the inbuilt tokenizer of the NLTK library. *Figure 7* depicts the initial tokenisation step before removal of redundant characters.

### 3.5. Strip white space

Leading and trailing white space was trimmed off in this step so as to make the processing more effective.

### 3.6. Remove digits

Individual numbers were removed.

```
'season', '-','episode', ':', "'", 'gates', 'of', 'heaven', "'",'(',')', 'post', 'episode', 'discussion', 'thread', ')', 'welcome', 'to', 'official', 'only', 'murders', 'in', 'the', 'building', 'season',
```

*Figure 8: Snippet of a reddit post (Figure 2) with numbers removed*

### 3.7. Remove stop words and punctuation marks

Stop words, such as 'the', 'a', 'of' etc., and punctuation marks does not provide any meaning nor convey any valuable information regarding the corresponding text they are included in. Therefore, stop words and all punctuation marks were removed from the tokenized text.

```
'season', 'episode', 'gates', 'heaven', 'post', 'episode', 'discussion', 'thread', 'welcome', 'official', 'murders', 'building', 'season', 'post', 'episode', 'discussion', 'thread', 'use', 'thread', 'discuss', 'season', 'episode', 'gates', 'heaven', 'finished', 'watching', 'episode', 'premiered', 'september', '3rd', '12:00', 'est', 'currently', 'watching', 'episode', 'please', 'sure', 'check', 'relevant', 'live', 'discussion', 'thread', 'commenting', 'get', 'spoiled', 'reminder', 'sub', 'locked', 'new', 'posts', 'hours', 'following', 'episode's", "release", 'information', 'reminder', 'spoilers', 'keep', 'spoilers', 'post', 'titles', 'share', 'spoilers', 'future', 'episodes', 'post', 'spoiler', 'tag', 'post', 'content', 'kindly', 'correct', 'users', 'posting', 'spoilers', 'message', 'mod', 'needed', 'comments', 'contain', 'spoilers', 'utilize', 'text', 'remove', 'spaces', 'ex', 'think', 'dog', 'liking', 'season', 'far', 'main', 'suspects', 'hope', 'see', 'next', 'week', 'see', 'next', 'week', 'new', 'olimabel', 'charles', 'silent', 'adventures', 'nb', 'new', 'rule', 'regarding', 'criticism', "show's", 'actors', 'added', 'comments', 'posts', 'hating', "people's", 'appearances', 'unjustifiably', 'hating', "someone's", 'acting', 'abilities', 'removed', 'sept', '2nd', '9pm', 'pst', 'hulu', 'sept', '3rd', '8am', 'bst', 'disney', '9am', 'cest', 'disney', '3pm', 'pht', 'disney', '5pm', 'aest', 'disney', 'comment', 'would', 'like', 'timezone', 'added']
```

*Figure 9: Snippet of a reddit post (Figure 2) with stop words and punctuation marks removed*

### 3.8 Stemming and lemmatization

Stemming and lemmatization were tested and abandoned as it resulted in unrecognisable words void of meaning for analysis.

## Result of cleaning data

*Figure 10* depicts an example of the final tokenized version of a reddit post (or comment).

```
['season', 'episode', 'gates', 'heaven', 'post', 'episode', 'discussion', 'thread', 'welcome', 'official', 'murders', 'building', 'season', 'post', 'episode', 'discussion', 'thread', 'use', 'thread', 'discuss', 'season', 'episode', 'gates', 'heaven', 'finished', 'watching', 'episode', 'premiered', 'september', '3rd', '12:00', 'est', 'currently', 'watching', 'episode', 'please', 'sure', 'check', 'relevant', 'live', 'discussion', 'thread', 'commenting', 'get', 'spoiled', 'reminder', 'sub', 'locked', 'new', 'posts', 'hours', 'following', 'episode's", "release", 'information', 'reminder', 'spoilers', 'keep', 'spoilers', 'post', 'titles', 'share', 'spoilers', 'future', 'episodes', 'post', 'spoiler', 'tag', 'post', 'content', 'kindly', 'correct', 'users', 'posting', 'spoilers', 'message', 'mod', 'needed', 'comments', 'contain', 'spoilers', 'utilize', 'text', 'remove', 'spaces', 'ex', 'think', 'dog', 'liking', 'season', 'far', 'main', 'suspects', 'hope', 'see', 'next', 'week', 'see', 'next', 'week', 'new', 'olimabel', 'charles', 'silent', 'adventures', 'nb', 'new', 'rule', 'regarding', 'criticism', "show's", 'actors', 'added', 'comments', 'posts', 'hating', "people's", 'appearances', 'unjustifiably', 'hating', "someone's", 'acting', 'abilities', 'removed', 'sept', '2nd', '9pm', 'pst', 'hulu', 'sept', '3rd', '8am', 'bst', 'disney', '9am', 'cest', 'disney', '3pm', 'pht', 'disney', '5pm', 'aest', 'disney', 'comment', 'would', 'like', 'timezone', 'added']
```

*Figure 10: Tokenised result of the post shown in Figure 2*

Without application of pre-processing techniques 361594 tokens were identified while after application of the aforementioned pre-processing techniques 157754 tokens were identified.

The top 50 unigrams pre and post cleaning are depicted in *Figures 11 and 12*.

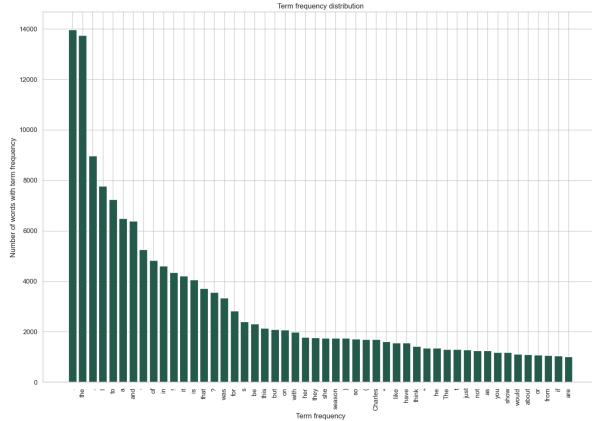


Figure 11: Term frequency unprocessed

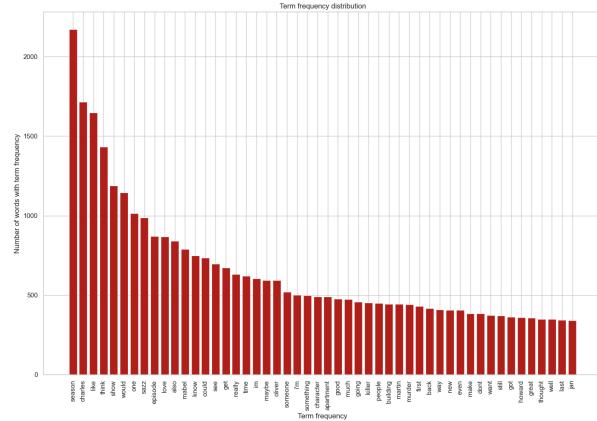


Figure 12: Term frequency after pre-processing

The contrast between the unprocessed and pre-processed dataset is evident as it clearly shows the removal of redundant characters and words in preparation for processing.

## 4 Method

To conduct a comprehensive analysis into the [Only Murders in the Building](#) subreddit various methods and models were explored.

To understand the views and opinions of the viewers of this TV show, unsupervised approaches such as Count and Vader sentiment analysis models were used. These unsupervised models were used as this project does not include sentiment labels for the extracted dataset. The viewers' opinions were explored within the time period of the new season renewal and the new season premier.

In order to discover the trending topics and themes within the subreddit for the aforementioned period of time, similar to the sentiment analysis models, an unsupervised approach, topic model Latent Dirichlet Allocation (LDA), was used.

The models were built using the Python programming language and Jupyter notebooks. Natural Language Processing related python libraries such as NLTK, pyLDAvis, Scikit Learn were used to develop the models.

### 4.1. N-grams

To gain a preliminary understanding of the opinions in this subreddit, unigrams, bigrams, and trigrams were explored for sentiment analysis as this technique takes the context of the words into consideration.

## 4.2 Sentiment analysis techniques

### 4.2.1 Count

A simple sentiment analysis technique where average sentiment of the reddit posts and their corresponding comments were determined by counting the number of positive and negative words. Here, a set of positive and negative words<sup>2</sup> were used to label each tokenized post (and comment) as positive, negative or neutral.

post, theories, tying, seasons, together, rewatched, season, whoever, put, note, jans, door, behind, retrospect, actually, one, spookier, moments, series, jan, taken, aback, even, moment, note, rewatched, s2e 1, wondered, amy, schumers, interest, getting, rights, podcast, turn, series, something, note, also, rem inded, audience, dont, know, poisoned, winnie, s1, im, rewatching, season, sister, right, lucy, telling, charles, serious, shit, going, building, definitely, tie, future, current, stuff, instead, season, there s, also, pets, animals, theme, keeps, coming, poisoned, winnie, bunnys, parrot, mrs, gambollini, forgot, ended, going, howards, pitch, animal, jobs, etc, sazz, also, explains, gates, heaven, pet, cemeteries, t heres, piglet, episode, know, said, death, video, think, howard, runs, crime, empire, arconia, howard, s eries, villain, youtube.com, need, know, lucy, mother, emma, need, know, mabels, aunt, relationship, poi soned, winnie, leaving, threatening, notes, doors, figure, howard, actually, got, black, eye, know, ther es, started, rewatching, s2, ill, try, come, back, see, animal, theme, occurred, charles's, first, intr o, monologue, s1, mocking, bodies, buried, building, nyc, might, tenants, beloved, dead, pets, ep, film, name, gates, heaven, pet, cemetery, anagrams, know, oh, hell, wonder, tap, another, ex, paint, patin, et a, s3, oliver, said, never, taps, s4, sazz, says, tap, murder, building, :p, first, season, mentions, ch arles, sister, long, isalnd, finally, going, meet, suspect, going, house, reason, since, never, actuall y, visited, going, drive, doorstep, seems, vital, piece, mystery, also, unresolved, mysteries, callback s, previous, seasons, making, sense, new, old, stray, dots, -, things, connect, oh, hell, discuss, whate ver, want, seasons, already, tied, together, though, biggest, themes, across, seasons, constant, referen ces, killing, real, estate, making, money, murder, cinda, movie, dimases, sadly, trio, extent, tie, toge ther, seasons, end,

-----Count sentiment value-----  
-11

Figure 13: Count sentiment of a reddit post and corresponding comments

### 4.2.2 Vader

A sentiment analysis technique built from social media which is expected to provide a more accurate analysis of the reddit posts and comments.

In this approach, the degree of positive and negative sentiment can be explored as this lexicon set has sentiment scores ranging from -4 to 4 instead of raw positive and negative counts. More specifically, the positive and negative scores are summed against one another and further modified by the Vader lexicon which is then normalised from 0 to 1, where 1 is the strongest.

For the figure shown in *Figure 13*, the sentiment analysis from Vader is as follows:

Negative score: 0.178

Neutral score: 0.73

Positive score: 0.092

Compound vader sentiment value: -0.9858

<sup>2</sup> Provided as a part of RMIT University's Social Media and Network Analysis course Workshop 4.

### 4.3 Topic modelling techniques

LDA is a topic model for unsupervised classification of documents and discovery of topics. LDA determines topics by finding clusters of co-occurring words by encouraging topics to have few prominent words and the documents to have few prominent topics.

In this project, a maximum of 10 topics were explored using this model in order to analyse deep and specific concepts and topics being discussed in this subreddit. A low number for the number of topics were set as that would make the topics found general while too large a number was not picked as that would make the topics overlap with each other.

Starting with vectorizing the token list using CountVectorizer, this document-term matrix was then used to generate the LDA model whilst specifying the number of topics desired. The parameter for maximum iterations was set to 10 as this is an iterative training process and the learning method was set to online as the model was trained instance by instance.

## 5 Analysis

### 5.1 Posts per date

With a total of 915 posts and 20012 comments collected, increased engagement can be seen in October, late July and August with the days around the last week of August showing the highest number of posts.

This can be correlated to when the season renewal was announced in October, 2023 and the release of the new season in August, 2024.

The highest engagement was observed (Figure 14, 15) on the 29th of August with 62 posts which can be observed from when people were theorising about clues from the first episode.

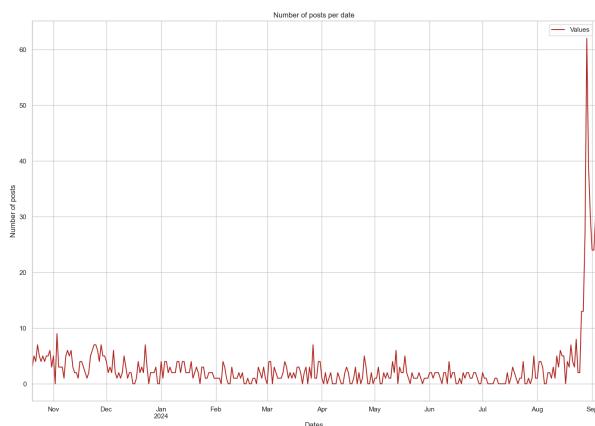


Figure 14: Number of posts per date

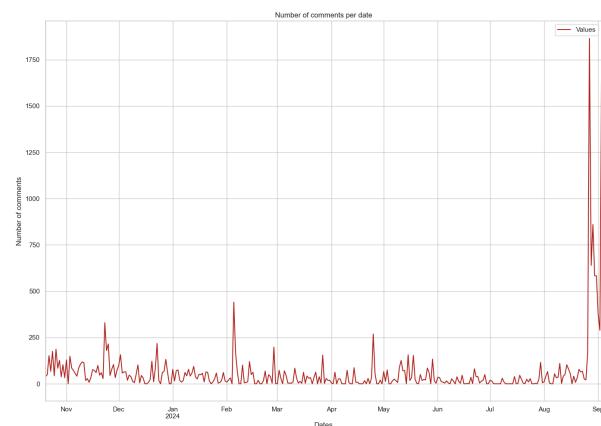


Figure 16: Number of comments per date

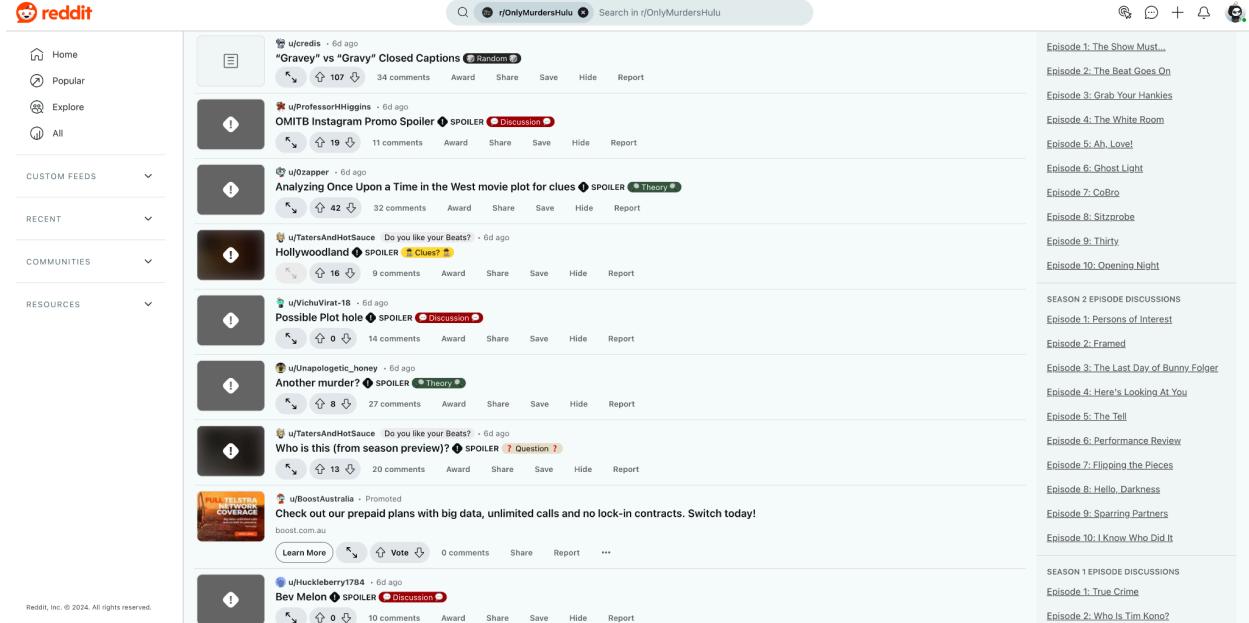


Figure 15: Reddit engagement on the 29th of August

## 5.2 Comments per date

According to *Figure 16*, the highest engagement can be seen in the month of August which correlates to the release of the new season. The highest comment count was observed on the [27th of August, 2024](#) at [1865](#) comments which was the release date of the first episode of the latest season. Small peaks of engagement could be observed throughout the timeline similar to the number of posts which coincides with cast announcements following the season 4 renewal.

## 5.3 Posts per author

The most active users on this subreddit were monitored based on the number of posts to determine that there is significant activity by different users on this subreddit (*Figure 17* see Appendix). The most active user was found to be reddit user [kaitlinsmom](#) with an overall post count of [38](#) posts within the time period.

## 5.4 Comments against the upvote ratio

It can be observed from *Figure 18* (see Appendix) that there is no correlation between the upvote ratio and the number of comments and a large portion of the posts which received a high upvote ratio received 0-200 comments.

## 5.5 N-grams

*Table 1* depicts the most common unigrams, bigrams and trigrams generated from the list of pre-processed tokens from the reddit posts and comments.

Unigram	Frequency	Bigram	Frequency	Trigram	Frequency
season	2171	steve, martin	220	upon, time, west	68
charles	1717	martin, short	157	steve, martin, martin	45
like	1648	feel, like	150	martin, martin, short	45
think	1434	murders, building	146	murders, building, season	35
show	1188	gif, giphy	140	death, razzle, dazzle	22

Table 1: Top 5 unigrams, bigrams and trigrams and their frequency

According to the n-grams and their frequencies it can be seen that this subreddit engages heavily with the plot by theorising about the direction of the show, especially with the season 4 premiere where the character **Charles** is the main target for the murder. By exploring the bigrams and trigrams, it can be seen that the viewers engage in discussions about the actors of the show, especially, Steve Martin and Martin Short.

Lastly, it can be seen that the viewers are heavily invested in the latest season, especially the first episode "Once Upon a Time in the West", with words such as "feel like", "like", "think" which point towards them theorising and speculating about the murder mystery taking place.

## 5.6 Sentiment Analysis

The average sentiment for the show seems positive with the Count method while a significant increase in negative sentiment can be observed in the month of August, 2024 along with expression of negativity in late October, 2023 (*Figure 19*).

While the average sentiment for the show with Vader shows similar sentiment to the Count method, large peaks of positive sentiment followed by a large drop can be observed in the month of August, 2024 with the peak and trough in late August being the highest positive and negative values observed overall (*Figure 20*).

It should be noted that the dramatic shift in sentiment could be attributed to the release of the latest season.

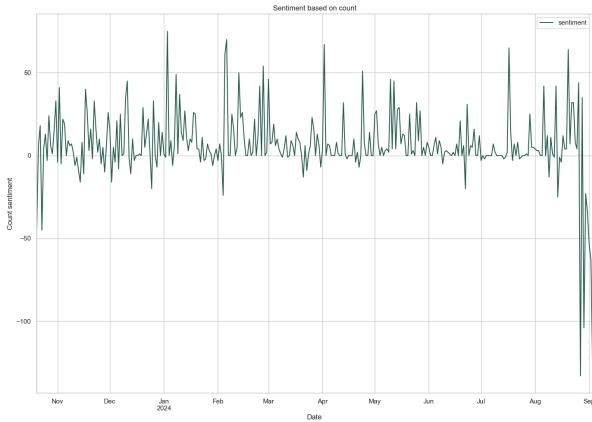


Figure 19: Count sentiment analysis

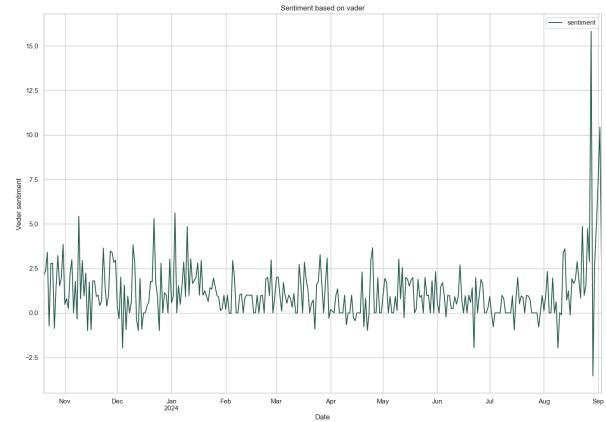


Figure 20: Vader sentiment analysis

As this is a show about solving cosy murder mysteries, given the show's popularity with a Rotten Tomato score of 97% and an audience engagement score of 91%<sup>[1]</sup> it can be inferred to be a very well liked TV show. Moreover, the viewers' feelings and thoughts towards the show can be gathered as predominantly positive and engaged by manually looking at the reddit posts and comments.

It can be seen that there is a mismatch in the sentiment depicted in the models with that of reality. This can be attributed to the frequently used words such as '**'murder'**', '**'slaughter'**', '**'suspect'**', '**'kill'**' etc which were being used to state facts of the show more than express the feelings and emotions of the viewers.

```
someone, trying, kill, jan, made, big, fuss, killing, whereas, murderer, trying, discreet, suggests, murderer, personal, vendetta,
wants, dead, without, necessarily, knowing, find, curious, trio, discussing, going, la, charles, apartment, gets, text, sazz, saying,
going, la, reckon, murderer, apartment, maybe, try, interrupted, oliver, mable's, arrival, overheard, talking, used, sazz's, phone,
check, calender, seen, going, texted, charles, think, murderer, would, successfully, got, away, one, uncovering, sazz's, murder, things,
murderer, probably, know, metal, bones, remains, could, identified, shoddy, clean, job, leaving, blood, oven, everyone's, favourite,
good, boy, gravy, porter, calling, charles, window, charles, cared, sazz, wellbeing, thought, also, find, murderer's, message, fucking,
friend, interesting, murderer, saying, someone, charles, thinks, friend, know, sazz, really, friend, far, shaping, really, good,
season, everyone, saying, howard, moriarty, vasovagal, syncope, prevent, able, handle, bloody, dead, body, feel, like, looking, past,
absolutely, devastating, ending, yes, terrifying, morbid, sazz, oldest, friend, hes, holding, bones, poor, charles, poor, sazz, crazy,
```

Figure 21: Snippet of the tokens from the post discussing the Season 4, Episode 1, highlighted according to count sentiment analysis method

The post which has the highest negative sentiment from the count method, a score of **-155**, on the discussion of the first episode of the latest season is shown in *Figure 21*. The same post and its associated comments got a sentiment analysis of **-0.9988** compound sentiment score from the Vader model. The sentiment was accurately identified here as the content of the episode showed the main characters slowly finding the remains of their dead friend (who was a fan favourite) which can be clearly seen to have distressed viewers (*Figure 22*, see Appendix).

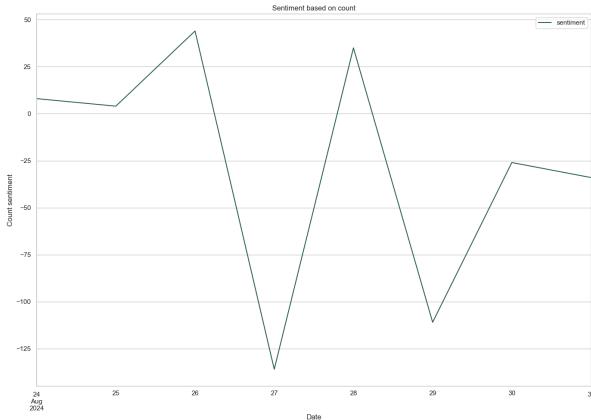


Figure 23: Count sentiment analysis (subset)

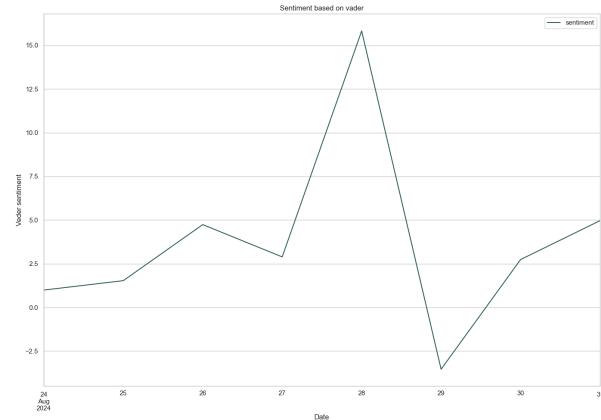


Figure 24: Vader sentiment analysis (subset)

Upon further analysis of the variations in the sentiments from the two methods via *Figures 23 and 24*, it can be seen that the negative sentiments gathered from the count method are significant due to the influence of the positive and negative word set while the vader method has a more accurate representation of the sentiments by looking at the difference in sentiment, on the 27th of August. This can be attributed to vader being a tool built from social media and thus having capability to better analyse the sentiments of social media posts such as this corpus.

Both count and vader methods accurately depict the posts and their associated comments on the 28th of August to be positive, for example count method providing a score of 22 and vader having a compound score of 0.9992 to the post depicted in *Figure 25* (see Appendix) and the analysed text in *Figure 26*.

```
next, season, he's, travelling, world, shooting, documentaries, personally, i'm, fine, never, seeing, mabel, like, charles, like, alone,
space, completely, isolated, dates, companionship, intimacy, can't, imagine, settling, anybody, think, needs, picture, separate,
places, lol, eyes, mabel, someone, need, romantic, relationship, feel, whole, still, figuring, wants, life, surrounded, murder,
technically, homeless, right, needs, guys, art, murder, solve, dates, someone, new, season, good, get, people, want, consistent, love,
interest, personally, enjoy, seeing, casual, relationships, end, clean, mutually, without, scripted, drama, angst, every, relationship,
end, requires, ice, cream, week, crying, bed, montage, pulling, oneself, together, hope, none, sounds, judgemental, subscribe,
philosophy, harm, none, ye, wanna, date, one, person, never, kiss, anybody, else, good, wanna, nothing, casual, sex, people, meet,
nights, great, long, everybody, consenting, happy, fine, book, bring, i've, seen, people, commenting, mable, needs, steady, singular,
love, interests feel, way, i'm, happy agree, disagree opinion, mine, think, shows, way, another, suspect since, joy, s1, murderer,
```

Figure 26: Snippet of the highlighted negative and positive words for the post in Figure 25 using count method

## 5.7 Topic Modelling

10 topics were visualised using pyLDAvis and wordcloud in order to determine the trending themes and topics of this subreddit. Given that the size of the topic displayed on the pyLDAvis panel shows how probable that topic is, it seems that all of the 10 topics are equally probable according to *Figure 27* (see Appendix).

The topic probability is explored with the lambda value of the relevance matrix set to 0. This is as the lambda value set to 0 depicts the probability of the words given the topic, while a

lambda value closer to 1 depicts the probability of the words in the topic by how probable it is in general. As the goal is to analyse the topics in this specific subreddit, which discusses a cosy murder mystery, the interest is in how probable these words are specifically in this corpus and not in a general context.

The general theme and topics being discussed can be concluded as shown in *Table 2* (see Appendix).

The following are the top most discussed topics:

1. Season 4 episode content and theories surrounding it
2. Steve Martin's character, Charles, taking centre stage of the show this season as the intended murder victim
3. The love and praise for actors playing the characters

Accordingly the three core themes could be inferred to be:

1. The viewers are excited and engaged with the latest season
2. The viewers are sad about the death of Sazz, one of the beloved recurring characters who the viewers speculate to not be the intended target of the murder
3. The viewers love theorising and trying to solve the mystery

## 6 Conclusion

This project was aimed at analysing the sentiments and trending topics of the popular social media platform, Reddit. The analysis was conducted on the [OnlyMurdersHulu](#) subreddit for the [Only Murders in the Building](#) TV show. This subreddit was chosen as it is one of the top 2% of all subreddits and has a large and active member count of 5K [2].

Upon initial exploration of the 1000 posts, pre-processing techniques such as tokenization were applied to clean the data for analysis. N-gram, Count and Vader were approaches utilised for sentiment analysis while the topic model LDA was utilised for topic modelling of 10 topics. It was found that Count and Vader approaches both produced similar analysis in certain instances, Vader was more accurate and efficient in the analysis when dealing with more complex emotions and feelings from the viewers which could be attributed to Vader being specifically social media focused.

The general sentiment towards the TV show was found to be positive from both models. While the viewers are excited about the arrival of the new season, negative sentiments were gathered since the release of the first two episodes which can be due to the death of a fan favourite character and the primary focus of the season. Moreover, it is worth noting that this subreddit discusses crime and murder related topics which contributed to incorrect sentiment analysis of the data. Majority of the topics of discussion centred around theorising about clues and puzzles to pinpoint who the murderer could be along with adoration for the creators and cast of the show.

## 7 References

- [1] Only Murders in the Building (2021) Rotten Tomatoes. Available at: [https://www.rottentomatoes.com/tv/only\\_murders\\_in\\_the\\_building](https://www.rottentomatoes.com/tv/only_murders_in_the_building) (Accessed: 06 September 2024).
- [2] Only murders in the building on hulu. Available at: <https://www.reddit.com/r/OnlyMurdersHulu/> (Accessed: 06 September 2024).

## Appendix

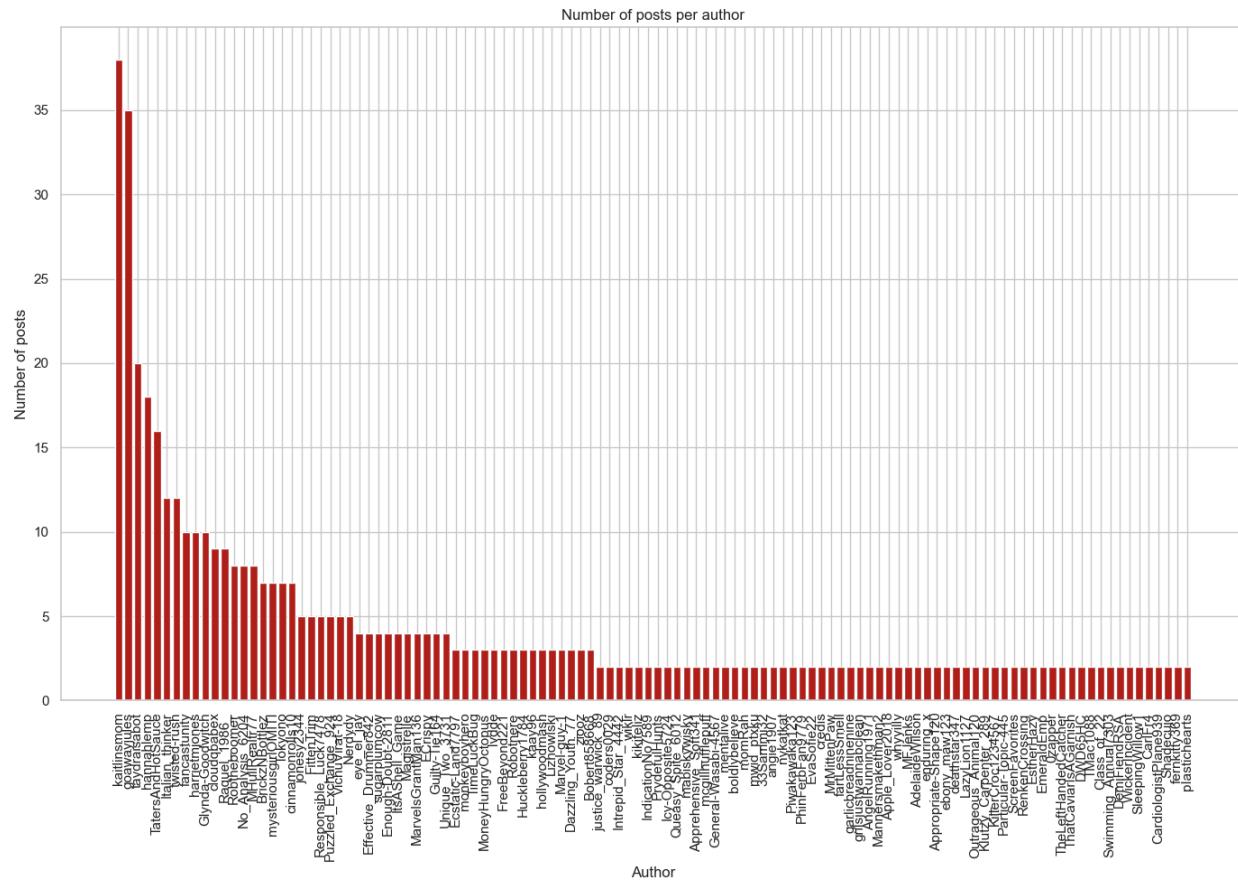


Figure 17: Number of posts per author<sup>3</sup>

<sup>3</sup> Only authors who had more than one post are depicted in this figure, moreover, authors' whose accounts were deleted have not been taken into consideration in this graph.

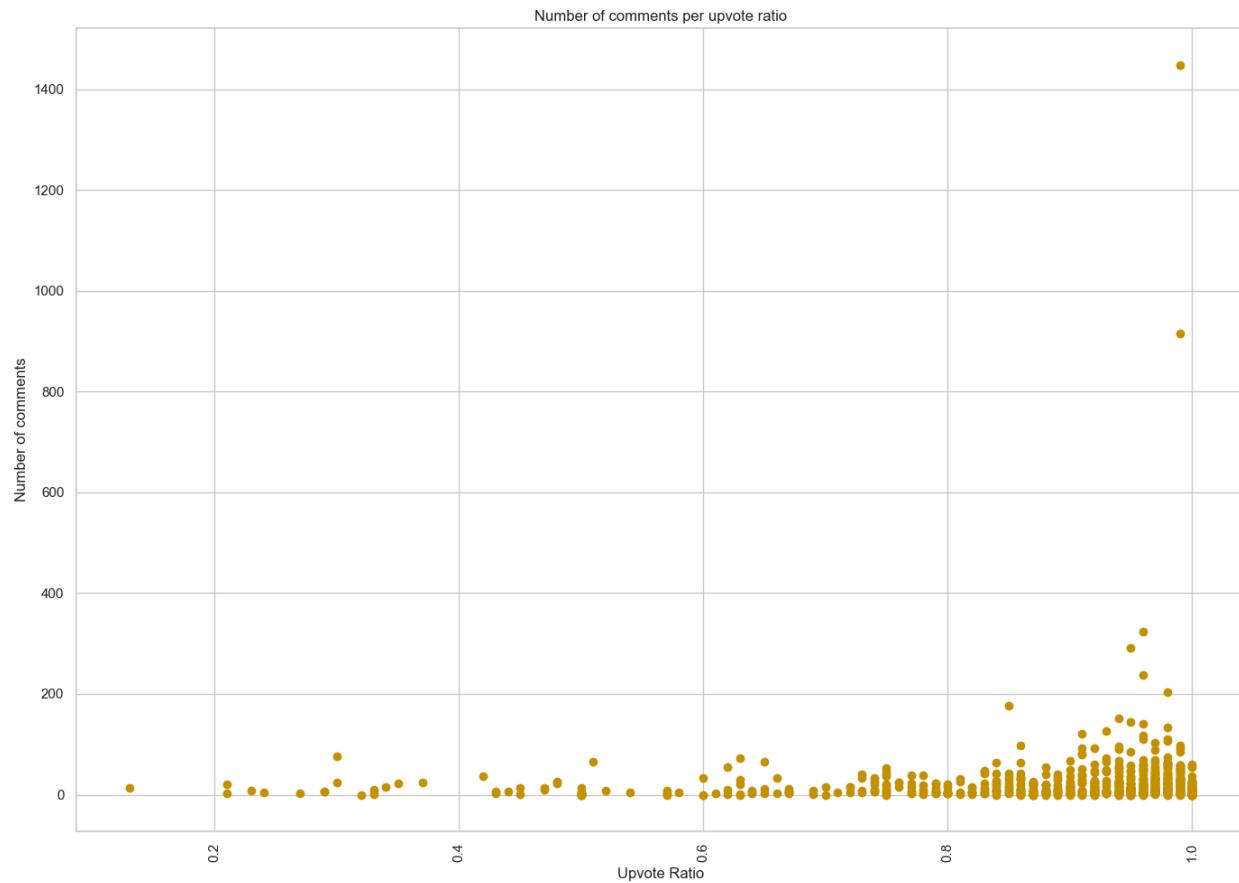


Figure 18: Number of comments against the upvote ratio

 **MaisyDeadHazy** • 10d ago

Simon and Garfyodel

God, that ending is freaking terrifying! Poor Sazz! 😢

...

 **mybeachlife** • 10d ago

My wife was holding out hope that Sazz was going to be fine up until that ending!

...

Figure 22: Viewer dismay over death of character

 r/OnlyMurdersHulu · 9 days ago  
 eviltimeline

**SPOILER**

## Any thoughts on Mabel's ever changing relationship?

**Discussion**

They keep trying to put her with someone. At this point it's just unnecessary. I wonder if she'll find another one for this new season.

Every setup for a new partner doesn't have tension or chemistry because eventually you know it won't last.

I personally like her with Theo because I personally think Theo is cute lol.

Maybe we can rank our favorite Mabel love interest instead?

Figure 25: Post (and comments) that got a positive sentiment score from both methods

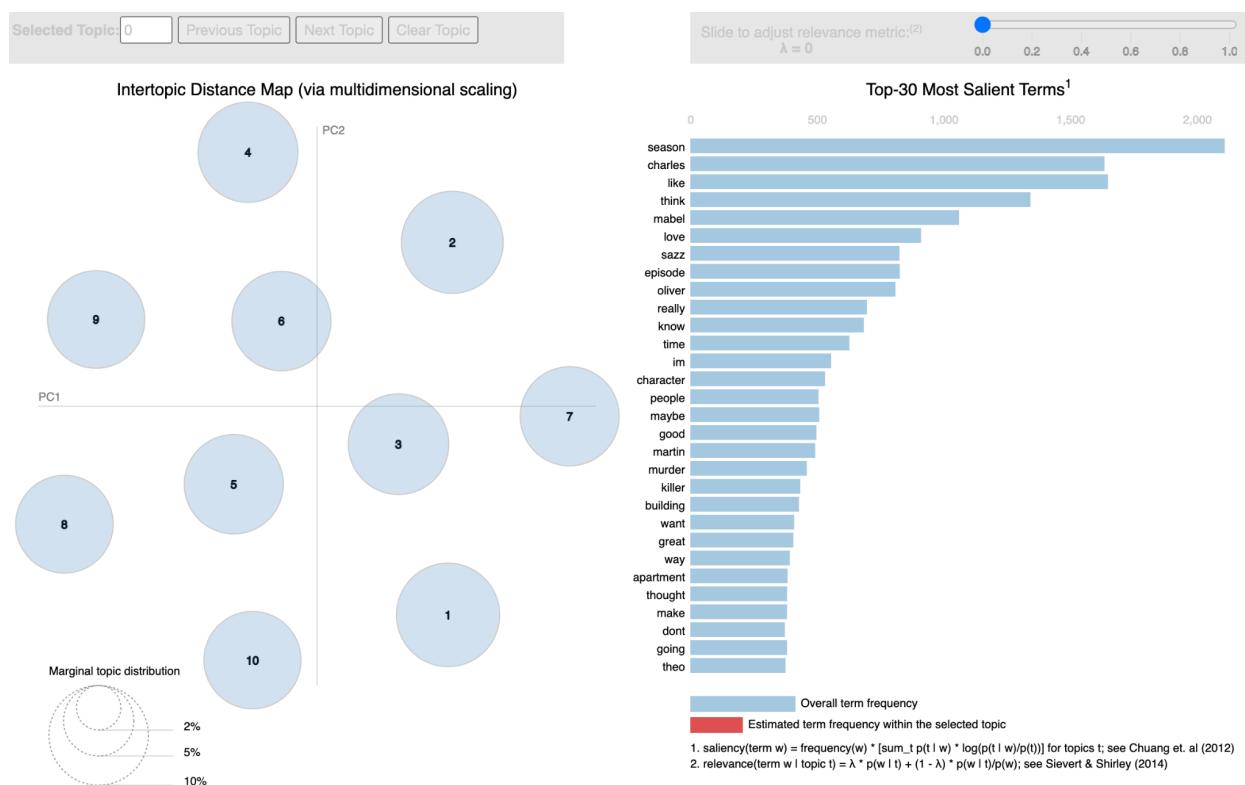


Figure 27: Intertopic Distance Map and the top 30 most salient terms

Topic	Inferred summary	Wordcloud
1	Discussion and excitement over the new season	
2	A theme of love for and between the characters and actors	
3	Viewers theorising about the show	
4	Discussions of the cast, specifically the character Steve Martin plays and theorising about him being the potential victim	
5	Theorising about the murders the previous and new season	

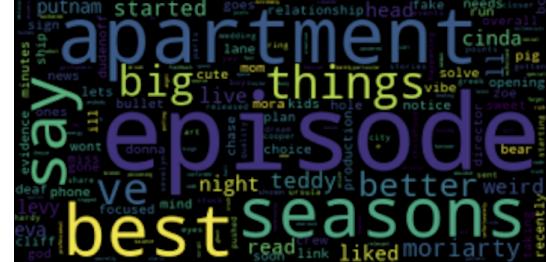
6	Theorising about the potential murderers in the building, primarily the character Howard	
7	Discussions on the first episode murder and theories about the potential plot points	
8	Viewers displaying satisfaction with the latest season	
9	Discussions and speculations about the character Charles being the intended victim	
10	Discussion about the latest murder victim, Sazz, and previous murderers	

Table 2: Inferred topic summary and word clouds