



SMU

SINGAPORE MANAGEMENT
UNIVERSITY

**IS434: Social Analytics and Applications
AY 17/18, Term 1**

Final Report

Group 1

Team Members:

Chia Hui Min

Koh Wei De John

Lee Qixian

Mervyn Lee Weng Ho

Ong Ming Hao

Table of Contents

| | |
|--|----|
| 1. Our Client | 2 |
| 2. Business Problem | 3 |
| 2.1. Increasing Follower Count | 3 |
| 2.2. Getting Likes Without Spamming of Hashtags | 3 |
| 2.2.1. Approach to Achieve Target | 3 |
| 3. Data | 3 |
| 3.1. Data Collection | 3 |
| 3.1.1. Likes Count | 4 |
| 3.1.2. Image | 4 |
| 3.1.3. Hashtag | 4 |
| 3.1.4. DateTime | 4 |
| 3.1.5. Comments | 4 |
| 3.1.6. Analysis of Other Poetry Instagram Accounts | 4 |
| 4. Data Gathering/Cleaning/Transformation | 5 |
| 4.1. Web Scraping | 5 |
| 4.2. Data Cleaning & Transformation | 6 |
| 4.3. Assumptions & Limitations | 6 |
| 5. Technologies/Tools/Methods Used | 7 |
| 5.1. Applications and Tools Used | 7 |
| 5.1.1. Libraries and Modules Used | 7 |
| 6. Analysis & Results | 7 |
| 6.1. Sentiment Analysis | 7 |
| 6.1.1. Sentiment of Timoteijosh's accounts. | 8 |
| 6.1.2. Sentiment of Timoteijosh's post | 9 |
| 6.2. Hashtag Usage | 11 |
| 6.2.1. Removing Common Hashtags | 12 |
| 6.3. Hashtag Effectiveness | 16 |
| 6.4. Optimum Times for Posting | 18 |
| 6.5. Content and Like Count Correlation | 19 |
| 6.6. Poem Length and Like Count Correlation | 21 |
| 6.6.1. Timothy's Account | 21 |
| 6.6.2. All 6 Accounts | 22 |
| 6.7. Summary of Recommendations | 23 |
| 6.8. Trial Implementation | 25 |
| 6.8.1. Using Recommended Hashtags | 25 |
| 7. Member Contribution | 26 |
| 7.1. Member's Roles and Responsibilities | 26 |

1. Our Client

Timothy Joshua Chia is a writer based in Singapore with the ambition to inspire and change the world. He focuses mainly on writing poems (and some short stories as well) and showcases these writings through social media platforms (mainly Instagram and Facebook) and has also published his poems in his very own book. He sells his book via Amazon and Etsy. He also provides special framed poems as well as personalized poems too.

Timothy's primary occupation is a secondary school teacher. Teaching both English and history, he writes poems and short stories mainly out of interest and passion. While he wants to continuously grow bigger and reach out to a larger audience, he does not exactly have the bandwidth to study the different ways and methods to further expand his outreach. This is where we come in.

Website: <https://www.timothyjoshua.com/>

Instagram: <https://www.instagram.com/timoteijosh/>

Facebook: <https://www.facebook.com/timoteijosh>

Twitter: <https://twitter.com/timoteijosh>

Tumblr: <http://timothyjosh.tumblr.com/>





Sample of Timothy's Instagram Post

2. Business Problem

Timothy's main goal has always been to increase his reach and to get a larger audience of followers. He mentions that there are writers with hundreds of thousands of followers and this is something he strives to achieve. In the world of Instagram and for online writers and authors, follower count is the currency for them. The higher the number of followers, the more outreach and popular one is. There are two issues which Timothy wants to tackle.

2.1. Increasing Follower Count

Timothy used to have goal of gaining 100 followers a week and now aims to achieve more new followers per week. Reason being that since he has grown bigger in follower count, he also needs to need to overcome the rate of old accounts deactivating and people unfollowing after sometime. All of this is to ensure a positive net increase in followers weekly.

2.2. Getting Likes Without Spamming of Hashtags

Automated Likes from hashtags, etc. Timothy mentions that a substantial portion of his Instagram likes do come from people searching via hashtags. He also feels that that Instagram algorithms seem to favour accounts that get likes first and hence the spamming of hashtags to get the first few likes. However, the main thing Timothy wants to figure out is a way to get his audiences to like his pictures/ appear on their feed like how bigger influencers do, without the spamming of hashtags

2.2.1. Approach to Achieve Target

In order to help Timothy achieve his goals, our team will be analysing existing content on Timothy's Instagram account as well as other well-known poetry Instagram accounts. Through our analysis, we aim derive potential strategies which Timothy can use to improve his outreach. In addition, we will also trial some of the proposed strategies and measure their effects as an assessment of the strategies.

3. Data

3.1. Data Collection

Firstly, we would be using data from his Instagram account as well as the other selected popular poetry Instagram accounts. The data taken from his social media accounts can then be analyzed to help increase his followings. For example, we can analyze the timings and types of post which will garner the most likes from his followers.

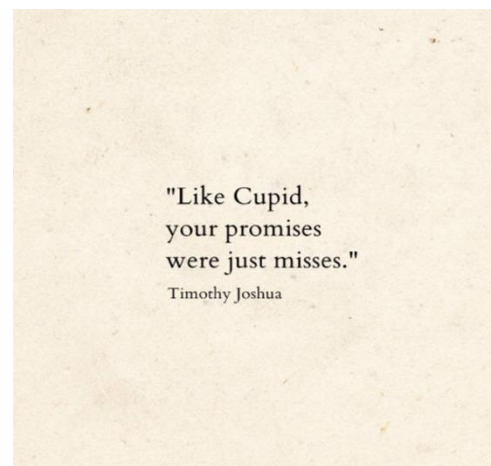
3.1.1. Likes Count

The number of likes is a common way to gauge the popularity of a particular post. This will be the dependent variable for most of our correlation analysis.

3.1.2. Image

From the images extracted, we will be performing Optical Character Recognition (OCR) to extract to textual content of each image which are mainly poems. Subsequently, we can use the content from these poems to correlate with the likes of the posts. This can be done by assessing the theme of the posts and also the length of the poems posted.

Additionally, we are also trying to extract information on the brightness as well as the font type used in the image. Such information could potentially have an impact on the overall aesthetics of a post to improve like counts.



3.1.3. Hashtag

Hashtags are often used to help garner more views as it allows posts to appear on searches. We want to explore the correlation between the hashtags used and like counts.

3.1.4. DateTime

The time of which the post was posted is also one of the factors which we hypothesize is able to affect the like count.

3.1.5. Comments

Sentiment analysis can be done on the comments of people to determine whether the post was positively or negatively received. This will help us understand the general comments and opinions of his followers and perhaps to identify potential requests and wants from them.

In addition, we can also explore the frequency of people tagging others in their comments as this serves as one way of expanding the account's popularity.

3.1.6. Analysis of Other Poetry Instagram Accounts

In addition to the timings, we can also analyse the hashtags his competitors/other writers based in Singapore are using to understand the network/connection that they are using to obtain their followers. Scraping the hashtags used by Timothy's followers can also help to understand their interest and allows us to understand their interests

Furthermore, we hope to get data from writers and poets with a large following. Similar to the different data we extracted from Timothy's Instagram, we will do the same for these other accounts. The objective is to analyse what they are doing and understand their social media strategies. From there, we figure out whether such strategies can be use in our client's case.

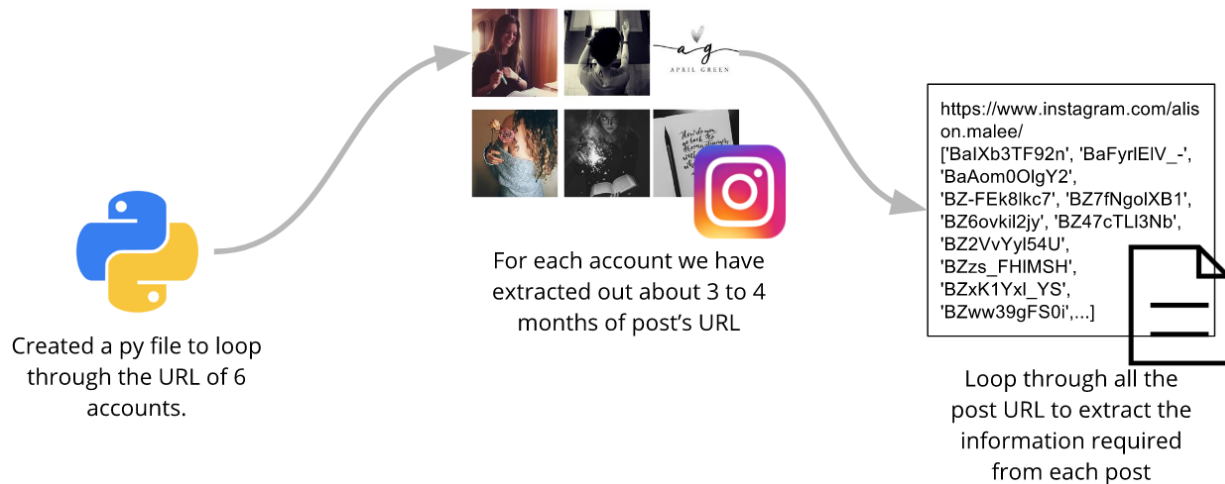
We choose to analyse accounts that has a similar style to Timothy's. However, as there are no prominent accounts within Singapore that has both the followers count and likes as much as Timothy's, we have chosen the following accounts with a more international following:

- Alison Malee: <https://www.instagram.com/alison.malee/>
- Nikita Gill: https://www.instagram.com/nikita_gill/
- Aprilgreen: <https://www.instagram.com/loveaprilgreen/>
- Faraway: <https://www.instagram.com/farawaypoetry/>
- F.d.soul: <https://www.instagram.com/featherdownsoul/>

4. Data Gathering/Cleaning/Transformation

4.1. Web Scraping

Overview of data scraping process:



Our first task was to retrieve information from the Instagram accounts of Timothy and the five other accounts. We decided to scrap the instagram web pages instead of using the API primarily because the application period to obtain an API key takes too long (i.e. several weeks thereafter which our project will already have ended). In addition, most of the information we need will be available straight from the web pages. Lastly, the instagram API provides some additional information only for the key's user account (Timothy's account in this case) and will not be any different from obtaining information via the web for other accounts.

Web scraping is done via a couple of python scripts. Firstly, we created a Python script to loop through the Instagram URL's of the 6 accounts, e.g. <https://www.instagram.com/timoteijosh/>. In this script, we get the individual post URLs. For each account, we have extracted out about 3 to 4 months worth of posts (roughly from end June to mid October).

After extracting the URL of each individual post, we then used another Python script to loop through these URLs (e.g. <https://www.instagram.com/p/BaT3YRYHuRJ>) and extract the data we need. This data would then be stored in a csv format, with an encoding of UTF-8 so that emojis can be properly stored.

Post

| post_id | instagram_handle | caption | image_url | timestamp | total_likes | total_comments |
|---------|------------------|---------|-----------|-----------|-------------|----------------|
|---------|------------------|---------|-----------|-----------|-------------|----------------|

Comment

| parentpost_id | instagram_handle | comment | timestamp |
|---------------|------------------|---------|-----------|
|---------------|------------------|---------|-----------|

4.2. Data Cleaning & Transformation

For the hashtag analysis, we collated the hashtags used from the caption variable in the Post data. However, not all instagram users will post it on the post's caption. Instead, they will place

them in the 1st comment of their posts. As such, we also needed to filter and match the `instagram_handle` to the account user of the `Comment` table, retrieving the first comment that the post's author has made.

For the sentiment analysis, we needed to merge the `Post` table and the `Comment` table using inner join by `post_id` and `parentpost_id`. This will allow us to filter only Timothy's account from the `Post_instagram_handle` and also to filter comments not made by Timothy under the `Comment_instagram_handle`. This will allow us to have all the comments made to Timothy's account that are not his own.

4.3. Assumptions & Limitations

There are not many poets in Singapore that uses Instagram as a platform for their words unlike Timothy. As such we are only able to look at other Instagram accounts based outside of Singapore that are similar in nature to Timothy's. We have chosen to explore 5 other instagram accounts as mentioned above. Hence, the analysis done may not be representative of all audiences.

Additionally, as they are overseas accounts, the audience they reach out to may be different to the ones Timothy is more likely to reach out to (i.e. people who might be more localized within the region). In this regard, while they may be of different audiences, our assumption is that certain aspects of content (poems) does not differ much regardless of the different audiences. This means that the strategies and learnings made on the other instagram accounts would be applicable to Timothy's account.

Lastly, some of the trends and patterns we may notice in our analysis (e.g word clouds) are limited to the past 6 months worth of data. It is important to note that it is possible for different trends to take place. Hence, such insights may change over the months and years.

5. Technologies/Tools/Methods Used

5.1. Applications and Tools Used

Python was used as our primary tool for majority of our work, including web scraping, data cleaning and analysis. Additionally, Tableau 10 was also used to help visualise collected data into charts.

5.1.1. Libraries and Modules Used

In order to perform certain functions, several python libraries were utilised. They are as listed:

- BeautifulSoup4 (Web Scraping)
- Google's Tesseract-OCR Engine via py-tesseract (OCR)
- Wordcloud (Data Visualisation)
- Matplotlib (Data Visualisation)

- NLTK (Text Analytics)
- Emoji (Data Cleaning)
- Pandas (Data Cleaning)
- Regex (Data Cleaning)
- Json (Data Cleaning)
- OpenCV (Text Analytics)
- Valence Aware Dictionary and sEntiment Reasoner (VADER) (Sentiment Analysis)

6. Analysis & Results

6.1. Sentiment Analysis

The use of sentiment analysis would be used to find the overall sentiment of both Timoteijosh's post and account. By analysing the overall sentiment of his account, we can identify where his account is well liked by his followers. Next, by analysing his post, we can also identify which posts have a higher positive sentiment and thus use posts of similar concepts in future posts, vice versa.

However, there is an issue analysing Instagram comments using sentiment analysis. Unlike food reviews or tweets, the problems that we face when analysing instagram comments are as follows:

- Very short (most comments are < 50 characters)
- Usually only contains Emojis or emoticons (e.g. :-D, ☹)
- Uses many punctuation (e.g. !!!, ?!?)
- Capitalized characters to emphasis(e.g. VERY)
- Adjusted intensity (e.g. very, kind of, uber)
- Slangs (e.g. sux, etc.)

As such, in order to deal with these issues, our group has implemented [Valence Aware Dictionary and sEntiment Reasoner \(VADER\)](#). VADER is a lexicon and rule-based sentiment analysis tool that is used primarily for social media. This implementation takes care of all the above problems on social media comments, except for emojis. Our group deal with this issue by adding emojis into the lexicon.

| TOKEN | MEAN-SENTIMENT-RATING | STANDARD DEVIATION | RAW-HUMAN-SENTIMENT-RATINGS |
|-----------|-----------------------|--------------------|--|
| \= | -1.1 | 0.3 | [-1, -1, -1, -1, -1, -1, -1, -2, -1, -1] |
| \^: | -1.3 | 0.45826 | [-1, -1, -1, -2, -1, -1, -1, -2, -2, -1] |
| \o/ | 2.2 | 0.9798 | [2, 1, 1, 2, 4, 2, 2, 4, 2, 2] |
| \o: | -1.2 | 0.4 | [-1, -1, -1, -1, -2, -1, -1, -2, -1, -1] |
|]~: | -2.1 | 0.53852 | [-2, -3, -3, -2, -2, -2, -1, -2, -2, -2] |
|]~: | -1.6 | 0.66332 | [-1, -2, -1, -2, -3, -2, -1, -1, -1, -2] |
|]~< | -2.5 | 0.80623 | [-2, -2, -2, -3, -4, -2, -2, -2, -2, -4] |
| ^<_< | 1.4 | 1.11355 | [3, 1, 3, 2, 1, 1, 1, -1, 2, 1] |
| ^urs | -2.8 | 0.6 | [-2, -3, -3, -2, -3, -3, -2, -3, -4, -3] |
| abandon | -1.9 | 0.53852 | [-1, -2, -2, -2, -2, -3, -2, -2, -1, -2] |
| abandoned | -2 | 1.09545 | [-1, -1, -3, -2, -1, -4, -1, -3, -3, -1] |

Sample of VADER's Lexicon

```

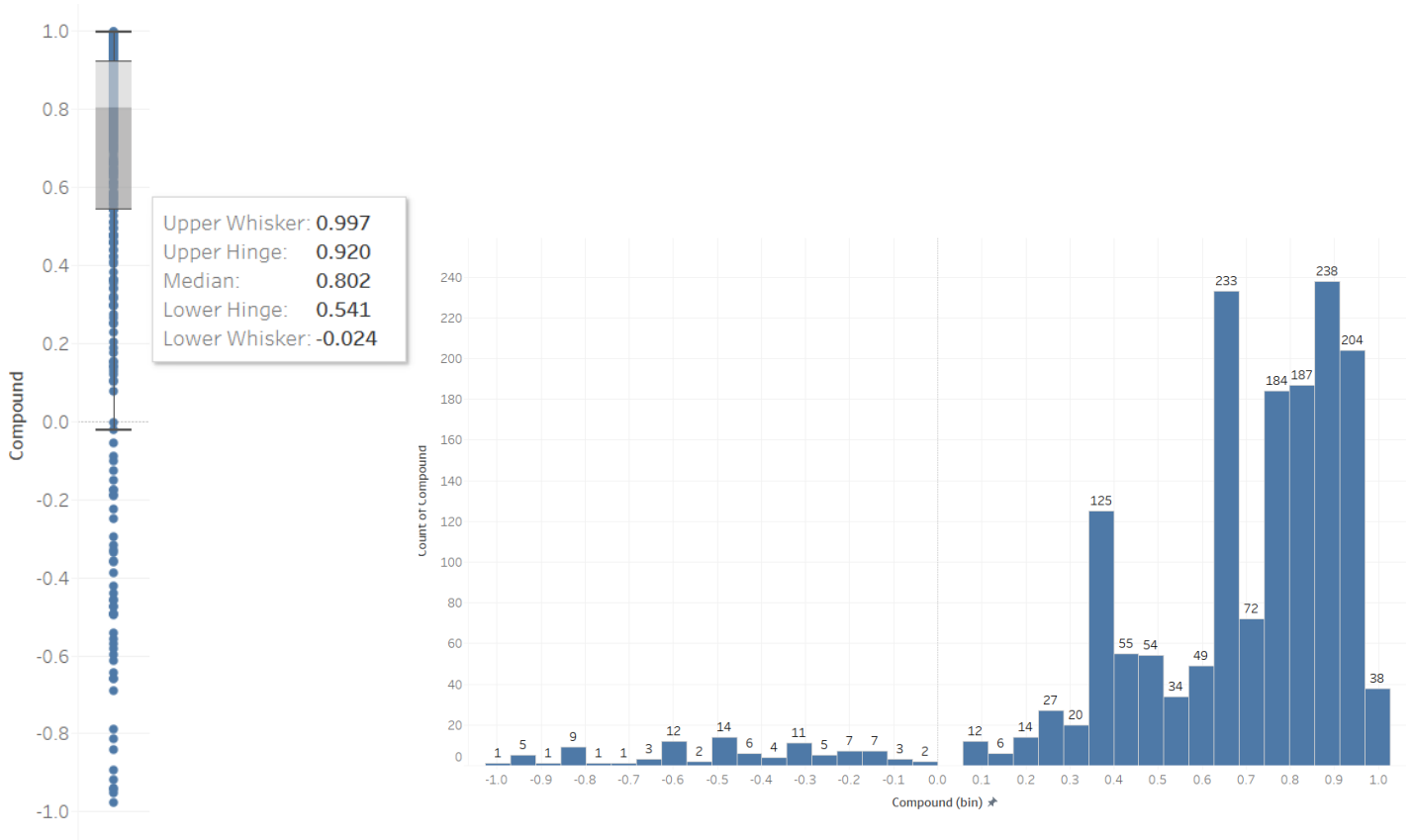
VADER is smart, handsome, and funny.----- {'neg': 0.0, 'neu': 0.254, 'pos': 0.746, 'compound': 0.8316}
VADER is not smart, handsome, nor funny.----- {'neg': 0.646, 'neu': 0.354, 'pos': 0.0, 'compound':
-0.7424}
VADER is smart, handsome, and funny!----- {'neg': 0.0, 'neu': 0.248, 'pos': 0.752, 'compound': 0.8439}
VADER is very smart, handsome, and funny.----- {'neg': 0.0, 'neu': 0.299, 'pos': 0.701, 'compound': 0.8545}
VADER is VERY SMART, handsome, and FUNNY.----- {'neg': 0.0, 'neu': 0.246, 'pos': 0.754, 'compound': 0.9227}
VADER is VERY SMART, handsome, and FUNNY!!!----- {'neg': 0.0, 'neu': 0.233, 'pos': 0.767, 'compound': 0.9342}
VADER is VERY SMART, uber handsome, and FRIGGIN FUNNY!!!----- {'neg': 0.0, 'neu': 0.294, 'pos': 0.706, 'compound': 0.9469}
The book was good.----- {'neg': 0.0, 'neu': 0.508, 'pos': 0.492, 'compound': 0.4404}
The book was kind of good.----- {'neg': 0.0, 'neu': 0.657, 'pos': 0.343, 'compound': 0.3832}
The plot was good, but the characters are un compelling and the dialog is not great. {'neg': 0.327, 'neu': 0.579, 'pos': 0.094,
'compound': -0.7042}
At least it isn't a horrible book.----- {'neg': 0.0, 'neu': 0.637, 'pos': 0.363, 'compound': 0.431}
Make sure you :) or :D today!----- {'neg': 0.0, 'neu': 0.294, 'pos': 0.706, 'compound': 0.8633}
Today SUX!----- {'neg': 0.779, 'neu': 0.221, 'pos': 0.0, 'compound':
-0.5461}

```

Example of how Vader does it's sentiment analysis

6.1.1. Sentiment of Timoteijosh's accounts.

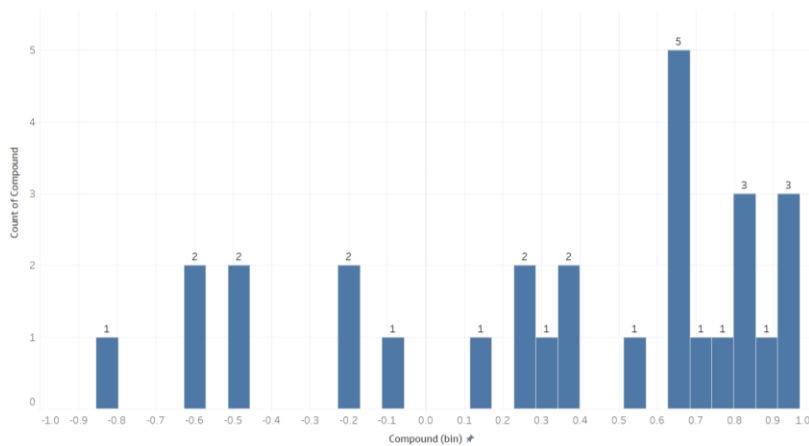
Our group has used VADER sentiment analysis to identify the overall sentiment of the comments that Timoteijosh post retrieve. Since the median of the compound of the VADER sentiment analysis is 0.802 from the boxplot and the histogram of the VADER sentiment analysis is mostly positive, we can safely say Timoteijosh's followers likes his overall content.



6.1.2. Sentiment of Timoteijosh's post

Besides using sentiment to identify the overall sentiment of instagram account, we are also able to identify post with a desirable response and posts with undesirable responses. Below are the analysis of 2 post, the first one with undesirable response (comments that mostly have negative sentiment) while the other one with desirable response (comments that are mostly have positive sentiment). From this 2 posts, we can identify that his followers may not want to hear about topics such as “breakup” or “heartbreak”, but would want to read poems about “love” and “soul mate”.

Timoteijosh may use this information to his advantage and post love poems, but only towards find soulmates, while avoiding poems of heartbreak and breaking up.



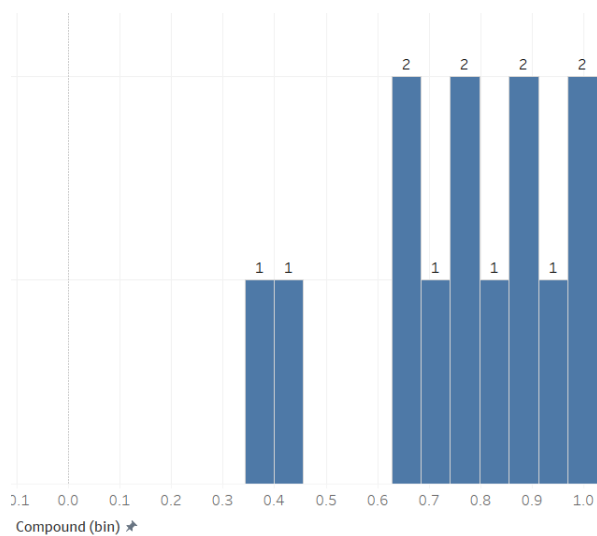
The End.

"I'm... I'm sorry." Her voice quivered, in fear of hearing his disappointment, yet her eyes glistened with hope, as two old lovers locked gazes, once more.

"I'm sorry too," he began,
"I just came here to say goodbye."

—Timothy Joshua

Post with undesirable response



The One.

We all have different ideals about our lovers.

I hoped that my lover would spend all her time with me.
I hoped that my lover would buy me my favourite gifts.
I hoped that my lover would be loyal.
I hoped that my lover would make me laugh.
I hoped that my lover would be my soul mate.

But now, knowing that you are everything and more,
I hope that my lover would be no one else but you.

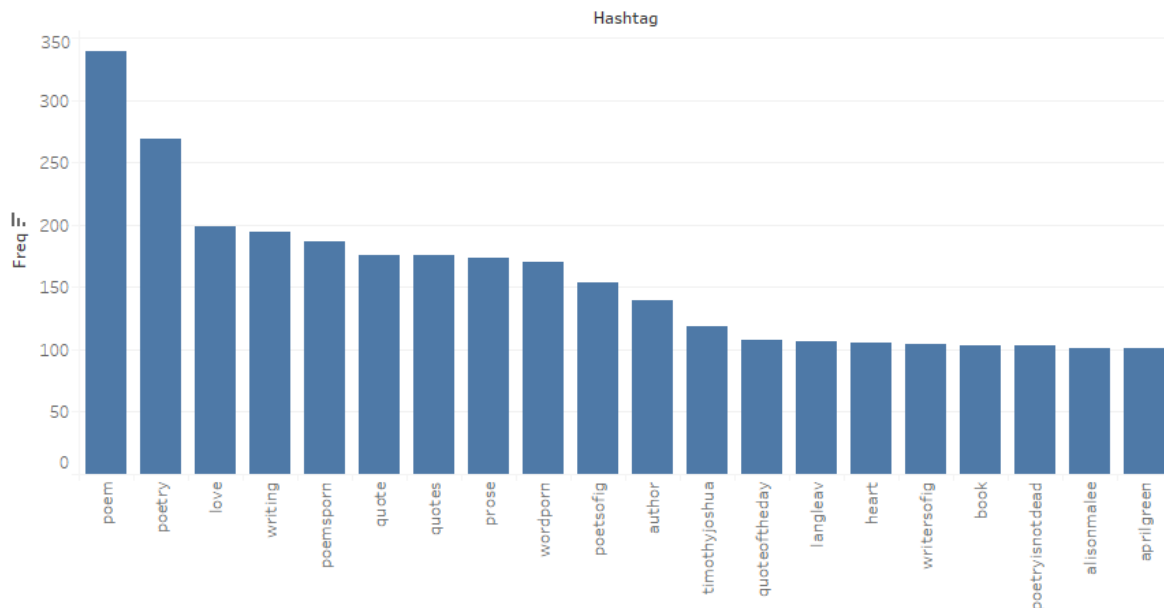
—Timothy Joshua

Post with desirable response

Our group has extracted the different hashtags used across the 6 instagram accounts. This is done by retrieving those from the caption of the posts as well as from the comments section made by the account. The hashtags are then stored in a dictionary and plotted into a Wordcloud and histogram.



Frequency of Hashtags from 6 instagram accounts



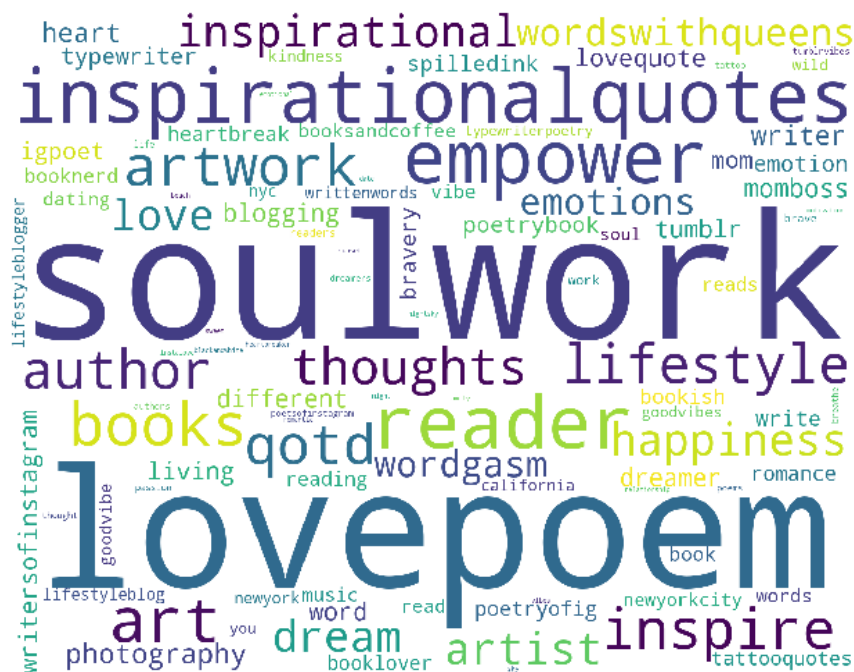
This helps us identify the popular ones used and allow us to proceed into studying which of these popular hashtags will gather more likes. From there, we will be able to recommend to Timothy the more effective hashtags to use. The top few hashtags used across the 6 accounts are: #poem, #poetry, #love, #writing and #poemsporn. As we can see, hashtags such as #poem and #poetry are simple but is the best representation of the type of post. Additionally, #love is the the third most commonly used hashtags, which also tells us that love themed poems are one of the most common ones.

6.2.1. Removing Common Hashtags

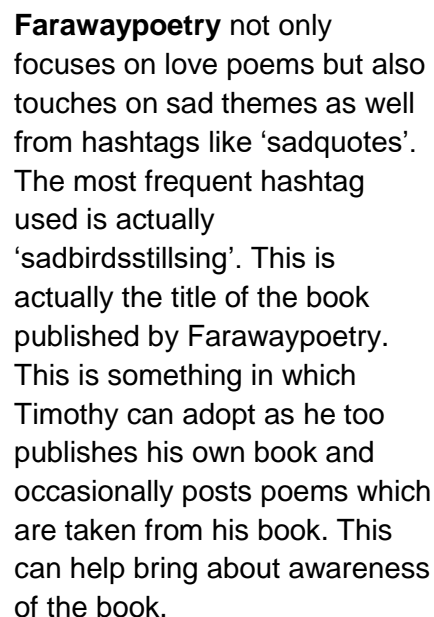
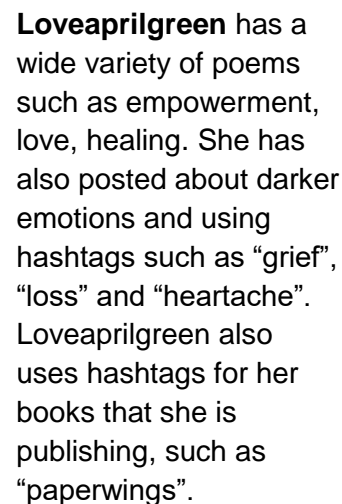
In the above shown word cloud, the hashtags #poem, #poetry, #writing and #poemsporn are all very generic hashtags and to a certain extent, obvious that poets will be using. We want to understand the context of the hashtags that were being used, such as love, breakup, emo, etc. Therefore by removing these generic ones as well as the instagram handle of the users, such as #timothyjoshua, #rmdrake, etc. The wordclouds generated are as shown below:

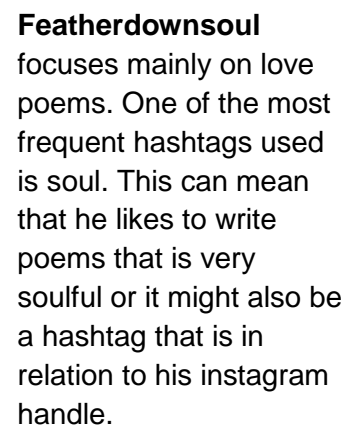


Timothy Joshua posts are leaning towards love and romantic poems and is possibly targeted at girls. Timothy also wishes to attract local audiences by using hashtags such as “singapore” and “sgwriters”.



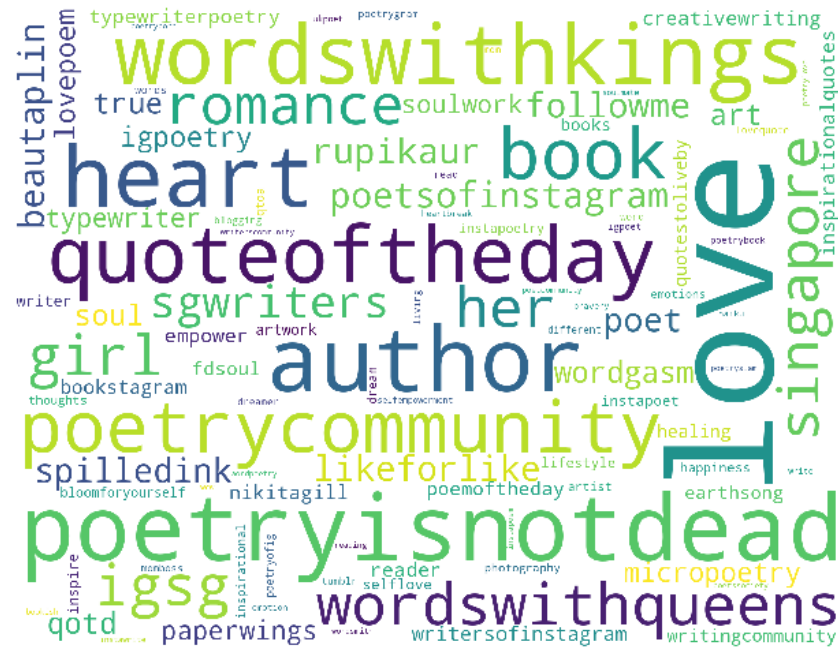
Alison Malee posts a diverse range of poems such as inspirational, empower, soul, love poems.





Generally, all the other poets write on a variety of themes, with love being the most common one. Additionally, authors like **loveaprilgreen** and **Alison Malee** write on more empowering and inspirational content while **farawaypoetry** and **nikitagill** do also write on sad themes. From the hashtags, we can see that Timothy, is more similar to loveaprilgreen and Alison Malee, and focuses more on happier content as well as love themes.

All 6 Accounts Combined

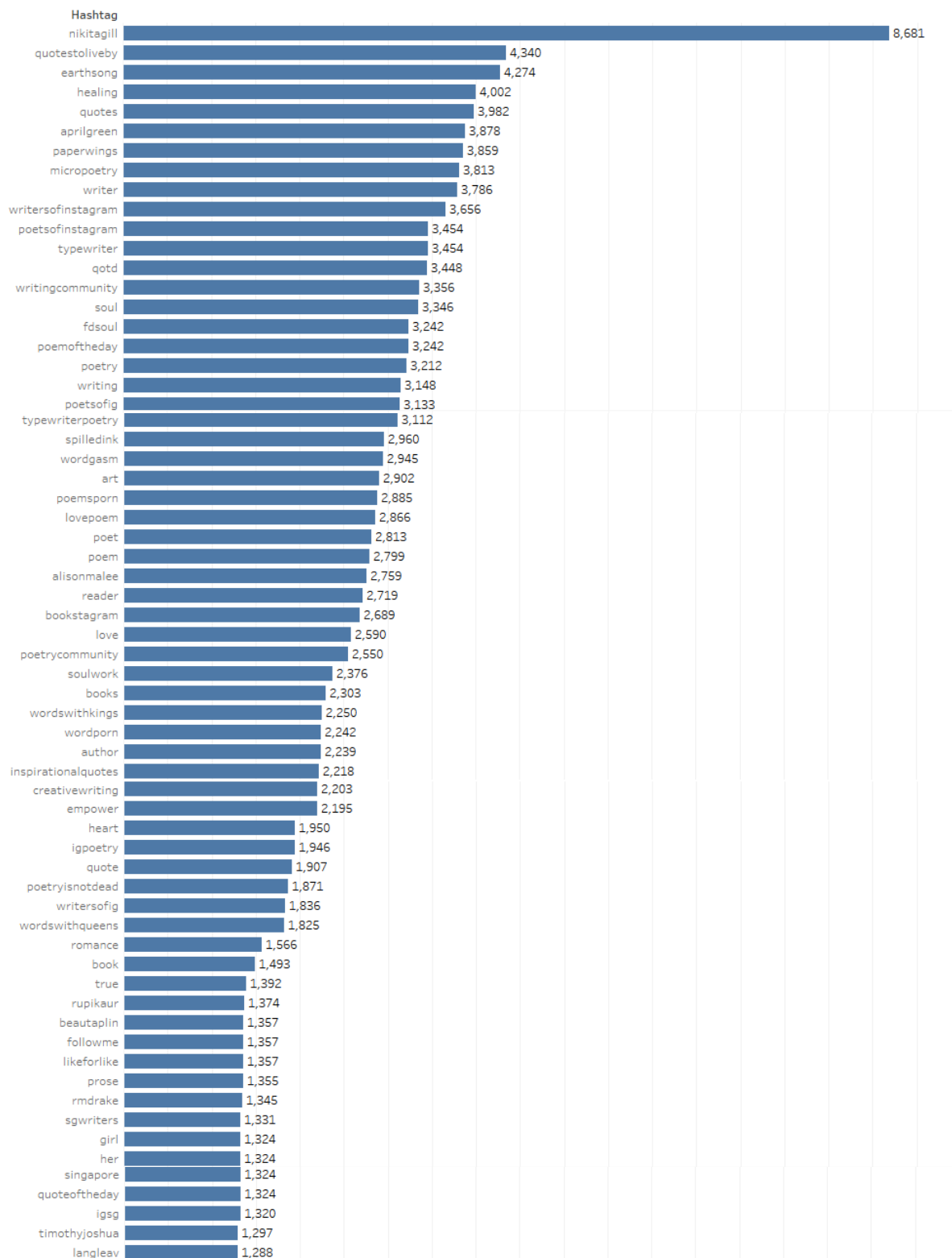


The theme that authors are usually posting are #love, #heart, #soul, #romance, #lovepoem, #healing. The wordcloud also tells us the other popular hashtags that are being used apart from the obvious ones.

6.3. Hashtag Effectiveness

Across the 6 accounts, there are over 400 hashtags used throughout the past 3 months. Timothy has his own set of hashtags but because instagram only allows a maximum of 30 hashtags per post, these hashtags might not be the most effective ones. Therefore, we decided to analyse the average likes for each hashtags. To ensure a more consistent result, we analysed hashtags only if they are used at least 30 times across all 6 accounts. This is to ensure that there is enough usage to properly analyse the effectiveness of each hashtags (using a hashtag once but generating the most likes would produce an inaccurate analysis). The results are as follows:

Average Likes for each Hashtag



Based on the results, Timothy has been using hashtags such as #quoteoftheday, #writersofsg and #poetryisnotdead which are the within the 20 least effective hashtags. Other hashtags such

as #quoteoftheday is pretty popular amongst the 6 accounts as seen from the wordcloud but is actually the 4th least effective hashtag. In addition, some of the most effective hashtags are not related to Timothy. For example nikitagill and aprilgreen are instagram handlers, and earthsong and paperwings are books written by other poets.

The hashtags we are recommending to Timothy are splitted into 2 segments: generic, contextual. They are as follows:

Generic:

#quotestoliveby #quotes #writer #writersofinstagram #poetsofinstagram #typewriter #qotd #writingcommunity #poemoftheday #poetry #writing #poetsofig #typewriterpoetry #spilledink #wordgasm #poemsporn #poet #poem #reader #poetrycommunity #books #wordporn #author #creativewriting #igpoetry #quote #timothyjoshua #poetryisnotdead #writersofig

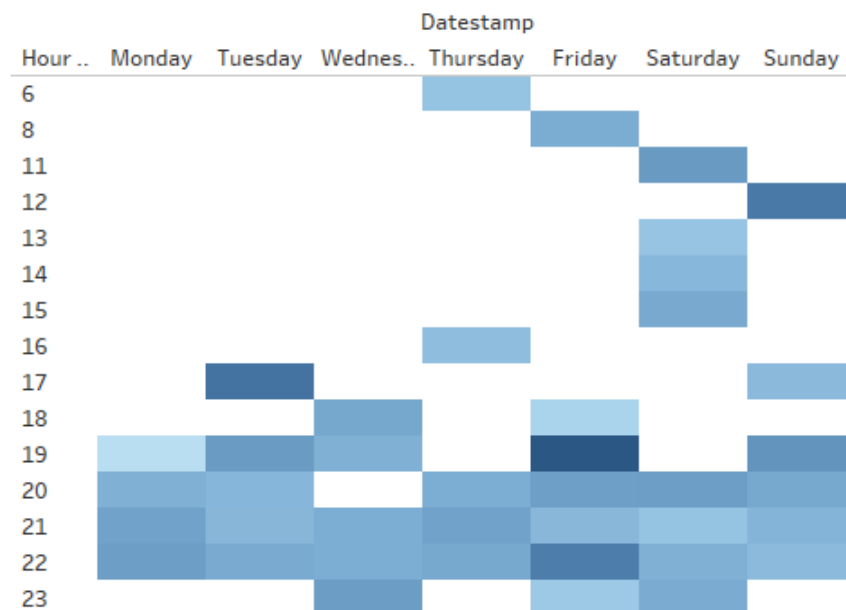
Contextual:

#healing #lovepoem #soul #love #inspirationalquotes #empower #heart

6.4. Optimum Times for Posting

In order to help us to recommend timings to maximize engagement for Timothy, we charted the average like counts against the timings and day of the week in which it was posted. This is done across all accounts. The results are shown below:

Timothy's Account



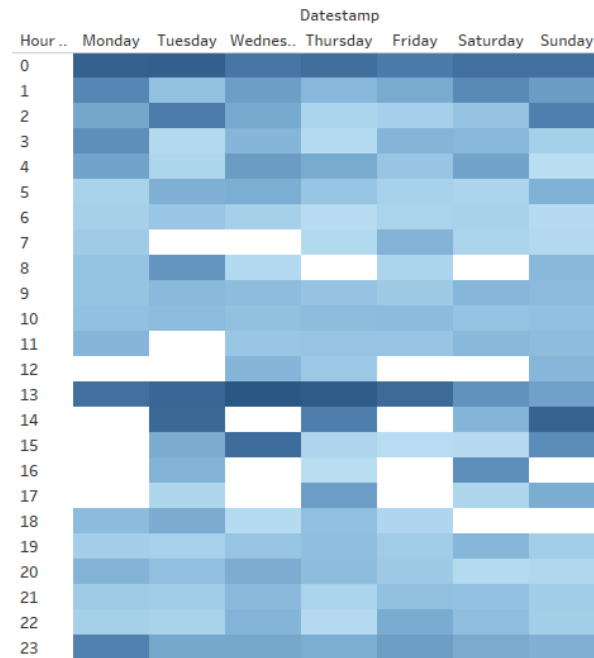
Heatmap of Timings - Darker shades indicate a higher like count

From the heatmap shown, we observe that the 4 best times for Timothy to release his posts are Friday 7pm and 10pm, Tuesday 5pm, Sunday 12pm. Generally, we can also see that mornings are not a good time for posting.

Timothy's Account + five other accounts

Next, we did the same analysis but now including the five additional accounts we have chosen. This gives us a different picture as these accounts have different audiences.

Average Likes by Day and Time for 6 Accounts



Across the 6 accounts, the 2 timeslots that gathered the most likes are from the 12am and 1pm periods. Interestingly, the most likes are usually during the weekdays. In Timothy's case, he does not post at midnight and instead post at on Saturday, 1pm (a less popular timeslot amongst the 1pm range).

Therefore, in order to reach out to a more international audience, we recommend the following optimal timings to post:

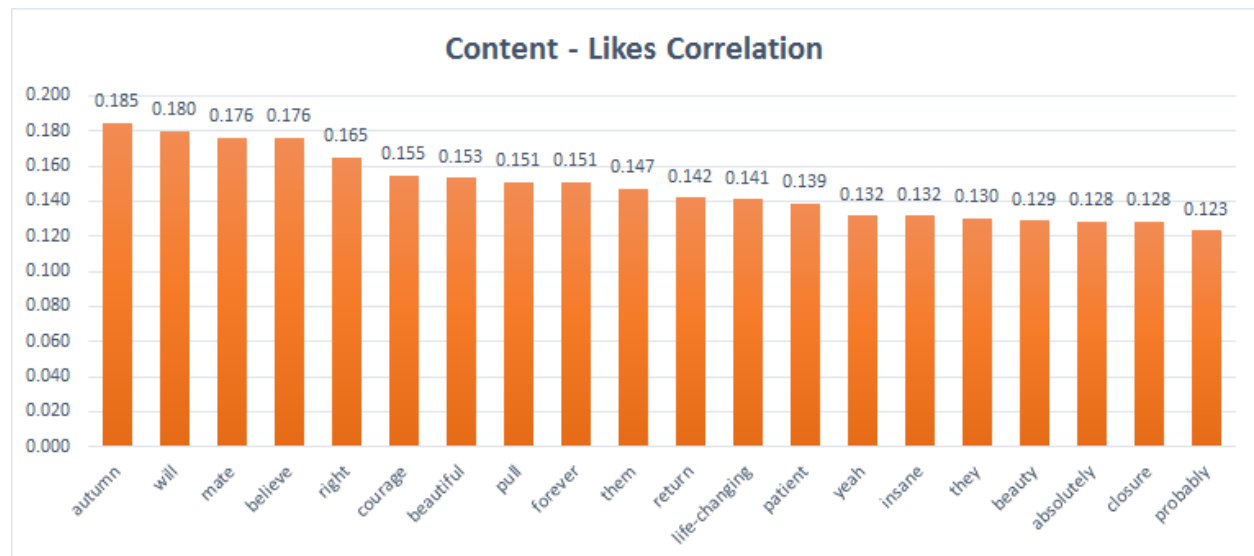
- Weekday: 12am or 1pm
- Weekend: 12am

6.5. Content and Like Count Correlation

Content that was extracted using OCR is used here to correlate with the number of likes of each post. To do this, we first ran a spell check on the content to reduce the number of words which might have not been read correctly by the OCR. Next, for each poem, we calculated the term frequency-inverse document frequency (tf-idf) score of each word to help us determine the most important words in every poem. We only considered the three words in every poem in terms of tf-idf scores. Subsequently, we formed a giant word matrix with the top 3 words of each poem and the like counts. The giant word matrix will look something like this:

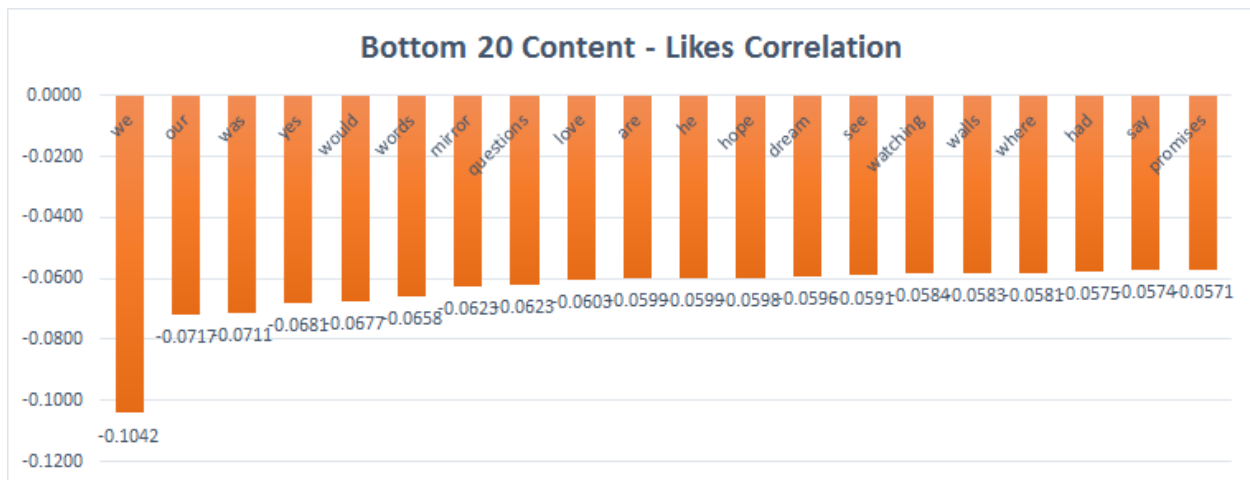
| Content | 'trying' | 'discomfort' | 'victim' | 'alive' | 'beginnings' | 'been' | 'unruly' | ... | Likes |
|---------|----------|--------------|----------|---------|--------------|--------|----------|-----|-------|
| Poem 1 | 0.355 | 0.250 | 0.250 | 0 | 0 | 0 | 0 | | 1242 |
| Poem 2 | 0 | 0 | 0 | 0.339 | 0.308 | 0.239 | 0 | | 1423 |
| Poem 3 | 0 | 0 | 0.303 | 0 | 0.101 | 0 | 0.212 | | 1920 |
| ... | | | | | | | | | |

With this giant word matrix, we then calculated the correlation of each word to the number of likes. From the results, the top 20 highest correlated words are shown below:



As we can see, the correlation values are only considered to be weak (between 0.1 and 0.3). However, they still do give us some insights on words that may influence the like counts positively. We notice that positive words such as 'courage' and 'believe' have a relatively higher correlation value. It shows that people may prefer poems that are uplifting and motivating. Followers of such instagram poem accounts are possibly looking for inspiration in their lives and hence are more attracted to poems which paint a positive light of life.

Next, the bottom 20 of correlated words are as shown below:

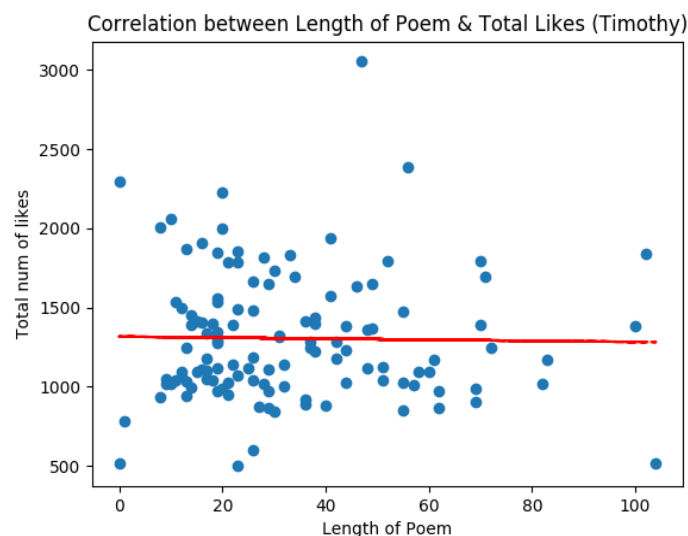


We can see that these words have a weak to no correlation with the like counts. All the correlation values are below 0.1 and hence there is no noticeable pattern for words that can affect the like counts negatively. As such, there is little restrictions as to what Timothy can write about.

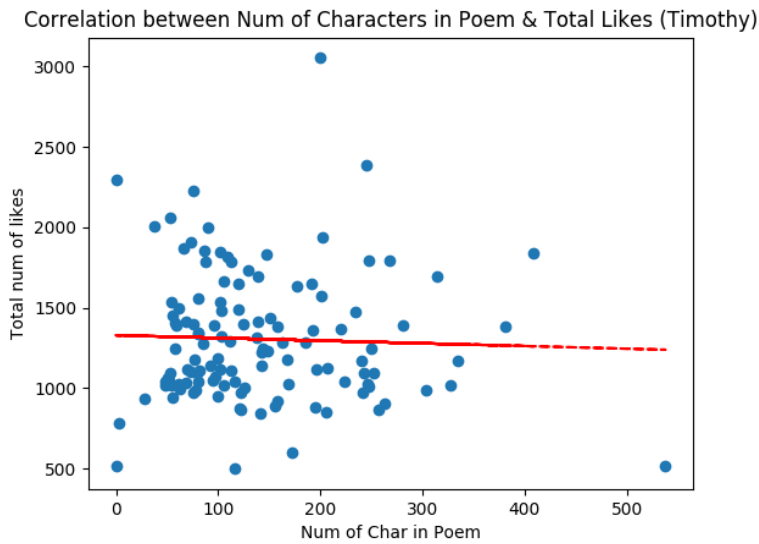
6.6. Poem Length and Like Count Correlation

6.6.1. Timothy's Account

We have plotted 2 graphs based on Timothy's post. The first graph shows the total length of each poem against the total number of likes while the second graph uses the number of characters instead.



Length of each poem (words) against the total number of likes.

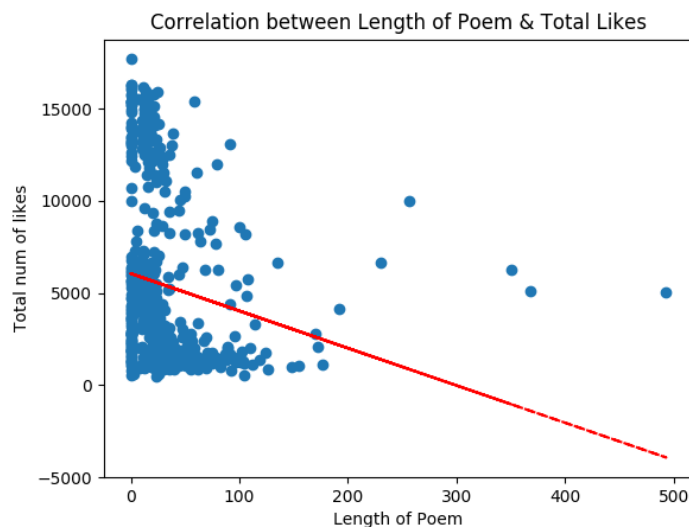


Num of characters of each poem against the total number of likes.

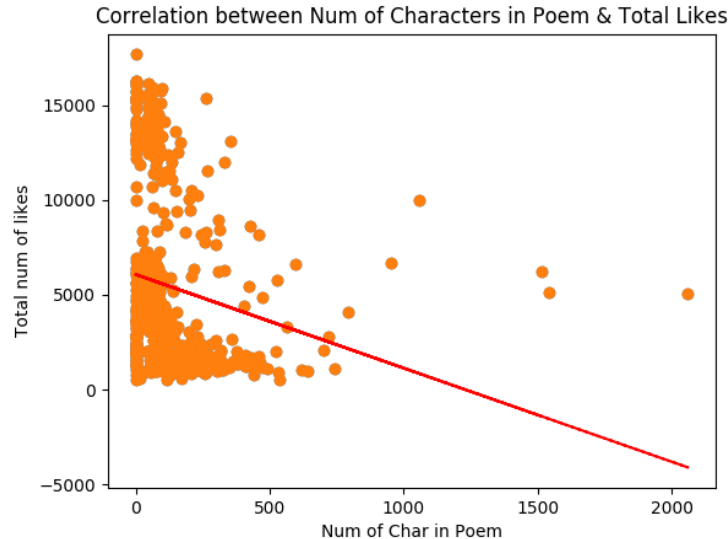
After we have visualize our data in the scatter plot, a correlation/ trend line is generated to see if there is any correlation between length/ number of characters and number of likes. We have found that there is no correlation between the length of poem ($r=-0.0178280283068$) or the number of characters ($r=-0.0366180260931$) with the number of likes.

6.6.2. All 6 Accounts

The next 2 graphs are based on all 6 accounts. The first graph shows the total length of each poem against the total number of likes. The second graph shows the total characters of each poem against the total number of likes.



Length of each poem (words) against the total number of likes.



Num of characters of each poem against the total number of likes.

After we have visualize our data in the scatter plot, a correlation/ trend line is generated to see if there is any correlation between length/ number of characters and number of likes. We have found that there is a weak negative correlation between length/ number of characters and the number of likes. The correlation between number of words and total likes is -0.182985151445 and the correlation between number of characters and total likes is -0.189594957622. Even though there is weak negative correlation between number of words/ characters and likes, we recommend Timothy not to post more than 100 words and 500 characters since poems with more words tend not to have lower amount likes.

6.7. Summary of Recommendations

Through the analysis we have conducted, we have recommended several aspects of posting on Instagram to Timothy. These aspects include recommended timings, hashtags, content of poem and finally the length of each poem.

Firstly, in order to reach out to more international audiences, 1pm and 12midnight are the best times to post. Additionally, to continue reaching out to the local community, 7pm on Fridays and 1pm on Sundays are recommended.

Next, Timothy should try out using some of the more efficient hashtags as listed in the [above analysis](#). Using more efficient hashtags i.e. hashtags which garner more likes, Timothy can slowly reduce the amount of hashtags used and move towards his goal of using less hashtags.

In terms of content recommendation, we were not able to derive any strong correlation between the content of poems and the like counts. However, we were able to notice a slightly correlation in content that help motivate people and which have positive vibes. While we recommend Timothy to focus more on such content to improve his followers and like counts, we do recommend him to only produce such content. Hence, a variety of content is still necessary although we were not able to explore this aspect in our analysis.

Finally in terms of content length, while the length of Timothy's poems do not show much correlation, the analysis done on international audience as seen from the other five accounts show that longer poems tend to result in lesser amount of likes. As such, should Timothy want to reach out to a more international audience, Timothy is recommended not to post more than 100 words and 500 characters since poems with more words tend not to have lower amount likes.

Our recommendations serve as a guide for Timothy to better understand his current and potential audience preferences so as to better cater to them. Additionally, as mentioned before in the limitations, such trends may change over time and this is something that Timothy needs to be aware of when taking in our recommendations.

6.8. Trial Implementation

6.8.1. Using Recommended Hashtags



Timothy has till date posted 7 posts using our recommended hashtags since November 6. For the 7 posts using our recommended hashtags, there has been a general increase in the likes per post averaging at 2443 likes per posts compared to the previous 7 posts before implementation at 1766 likes per posts. Timothy has also reached the highest number of likes (3189 vs 3052 previously) using our recommended hashtags. From the post, we can see that there is more total comments in each post, a total of 301 comments was posted before recommended hashtags and 388 comments was posted after. Lastly, we have also looked at the total number of instagram followers for Timothy before and after using our recommended hashtag, where we can see a gradual increase in the number of followers.



He started using our recommended hashtags on 6 November. While we cannot conclude that the increased rate of followers increase is due to the new hashtags, it has offered promising signs that the new strategy is moving in the right direction.

7. Member Contribution

7.1. Member's Roles and Responsibilities

| Team Member | Role | Responsibilities | Contribution Score |
|--------------------|------------------|--|--------------------|
| Chia Hui Min | Project Manager | <ul style="list-style-type: none">• Generate Optimum Time, Correlation between Length/ Char and Likes Post, Report and Powerpoint Slides | 4 |
| Koh Wei De John | Data Analyst | <ul style="list-style-type: none">• Data Storage, Generate Wordcloud, Hashtag Analysis, Trial Implementation and Report | 4 |
| Lee Qixian | Business Analyst | <ul style="list-style-type: none">• Web Scraping, Data Storage, Client Liaising, Trial Implementation and Report | 4 |
| Mervyn Lee Weng Ho | Data Analyst | <ul style="list-style-type: none">• Web Scraping, OCR, Content and Like Count Correlation, Report and Powerpoint Slides | 4 |
| Ong Ming Hao | Data Janitor | <ul style="list-style-type: none">• Data Cleaning, Sentiment Analysis (Comments and Emojis) Report and Powerpoint Slides | 4 |