



9<sup>3</sup>/<sub>4</sub>

**TEXT MINING ANALYSIS:**  
**HARRY POTTER AND THE**  
**CHAMBER OF SECRETS**

*by:*

**NUR MARDHIAH BT. ZULKHAIRI (P119717)**

**TAN SU RONG (P120754)**



# INTRODUCTION



Dataset :

Analysis of Harry Potter Books | Kaggle

<https://www.kaggle.com/code/chandanarprasad/analysis-of-harry-potter-books/notebook>

Books introduction :

Wizards World | Harry Potter And The Chamber Of Secrets

<https://www.wizardsworld.com/discover/books/harry-potter-and-the-chamber-of-secrets>

**H**arry Potter's summer has included the worst birthday ever, doomy warnings from a house-elf called Dobby, and rescue from the Dursleys by his friend Ron Weasley in a magical flying car! Back at Hogwarts School of Witchcraft And Wizardry for his second year, Harry hears strange whispers echo through empty corridors – and then the attacks start. Students are found as though turned to stone ... Dobby's sinister predictions seem to be coming true.



# OBJECTIVES



1. To compare the occurrence of characters
2. To identify the most frequent words in the story
3. To study the polarity of the text
4. To study the words emotions exist in the text
5. To explore the association of certain words in text



# DATA ANALYSIS

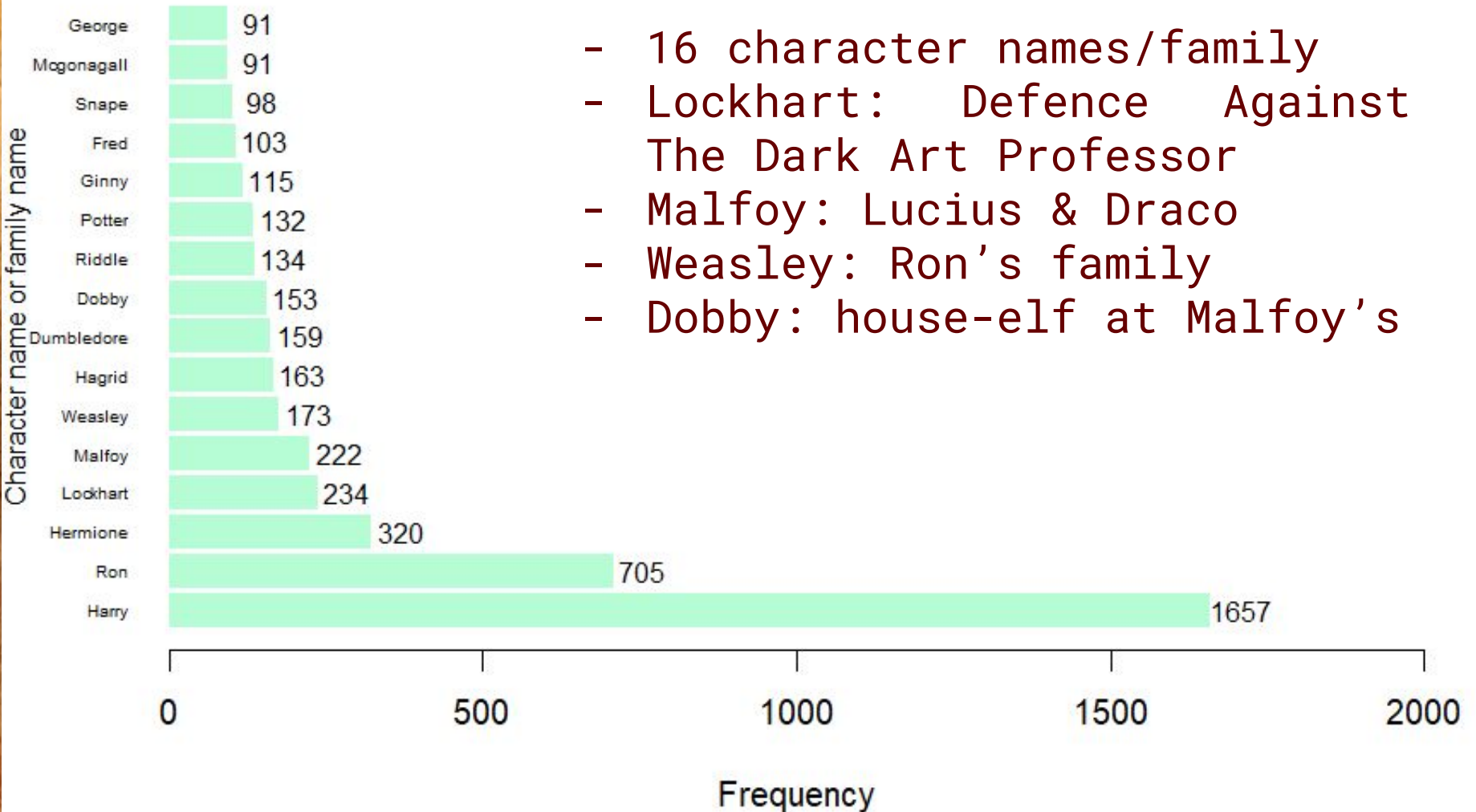


## Objective 1 :

To compare the occurrence of characters

- ❑ Main characters : Harry, Hermione, Ron
- ❑ Other important characters : different by books
- ❑ Who are they in this book?

## Top 25% occurrence of characters and family





# DATA ANALYSIS



## Objective 2 :

To identify the most frequent words in the story

- ❑ Frequency of words in the story after removing character/family names
- ❑ Remove meaningless word-"said" that occurs too often
- ❑ Professors in Hogwarts
- ❑ look, eye, see: basilisk

word	freq
look	413
back	298
one	224
professor	193
like	190
know	189
got	185
think	185
just	183
get	181
eye	177
time	174
around	172
now	164
can	163
see	157



well slytherin attack shout  
wand hand want ask quick look  
door think come point  
around know face last pull  
just back didn't make  
open get now say take  
eye look got never two  
one like tell don't tri came  
professor can even wall  
head time see let will next  
stare someth right still turn dark  
window book way door think come point  
close watch way door think come point  
people near wall walk stop seen  
knew open just back didn't make  
talk sir year get now say take  
left year get now say take  
wait sort went behind long eye look got  
day went behind long eye look got  
loud voice tell one like  
boy though professor can  
find wasn't saw head time see  
gave stare someth right still turn  
dark

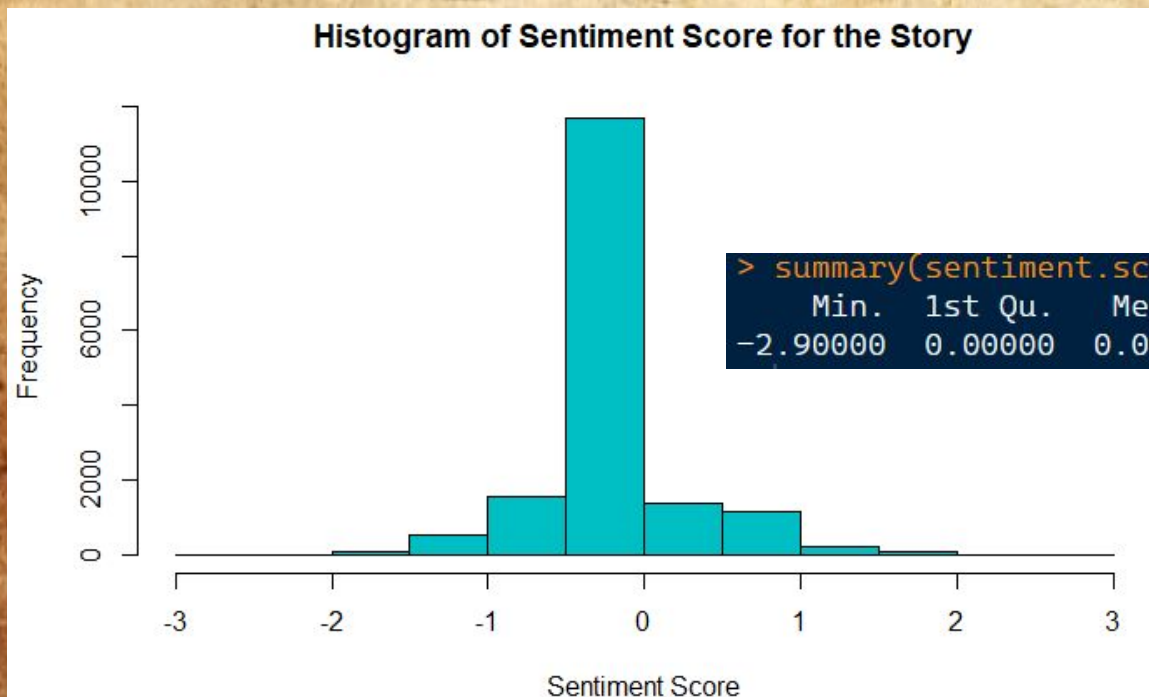


# DATA ANALYSIS



## Objective 3 :

- ❑ the polarity of the word is normally distribution with the mean of small negative value
- ❑ which the most of the scores are natural with the mode is between -0.5 to 0



```
> summary(sentiment.score)
   Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
-2.90000  0.00000  0.00000 -0.01301  0.00000  2.90000
```



# DATA ANALYSIS



## Objective 4 :

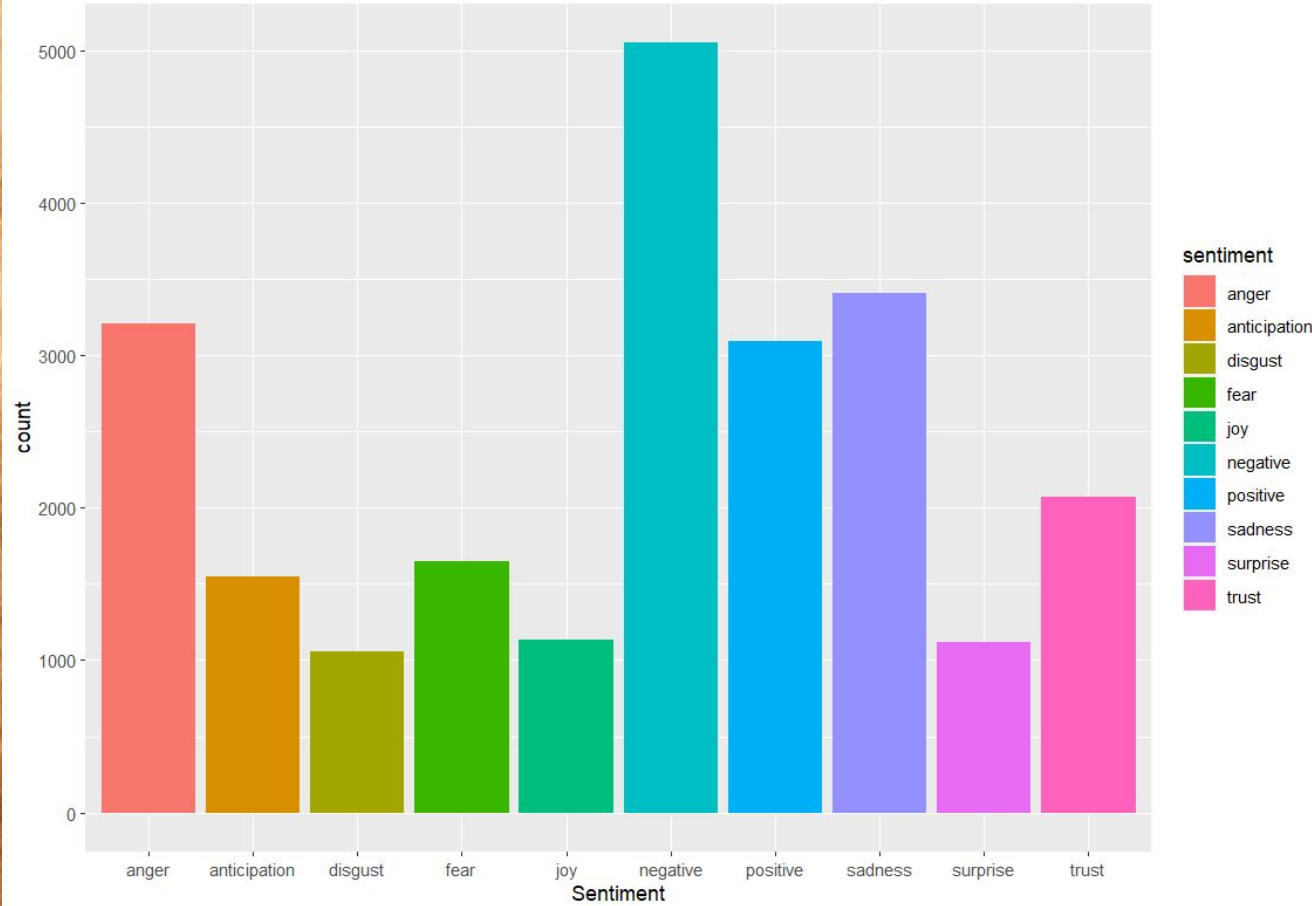
To study the words emotions exist in the text that correspond to eight basic emotions from NRC Emotion Lexicon

- ❑ Highest bar: Negative emotion with more than 5000+ counts
- ❑ Second highest : Sadness emotion with 3000+ count
- ❑ Lowest bar: Disgust emotion





Sentiment Analysis of the Harry PotterThe Chamber of Secrets





# DATA ANALYSIS



## Objective 5 :

To explore the association of certain words in text

- ❑ Usually will go to search activity
- ❑ There are following words when just typing the first words
- ❑ 13 words from character/family names selected, corlimit=0.07



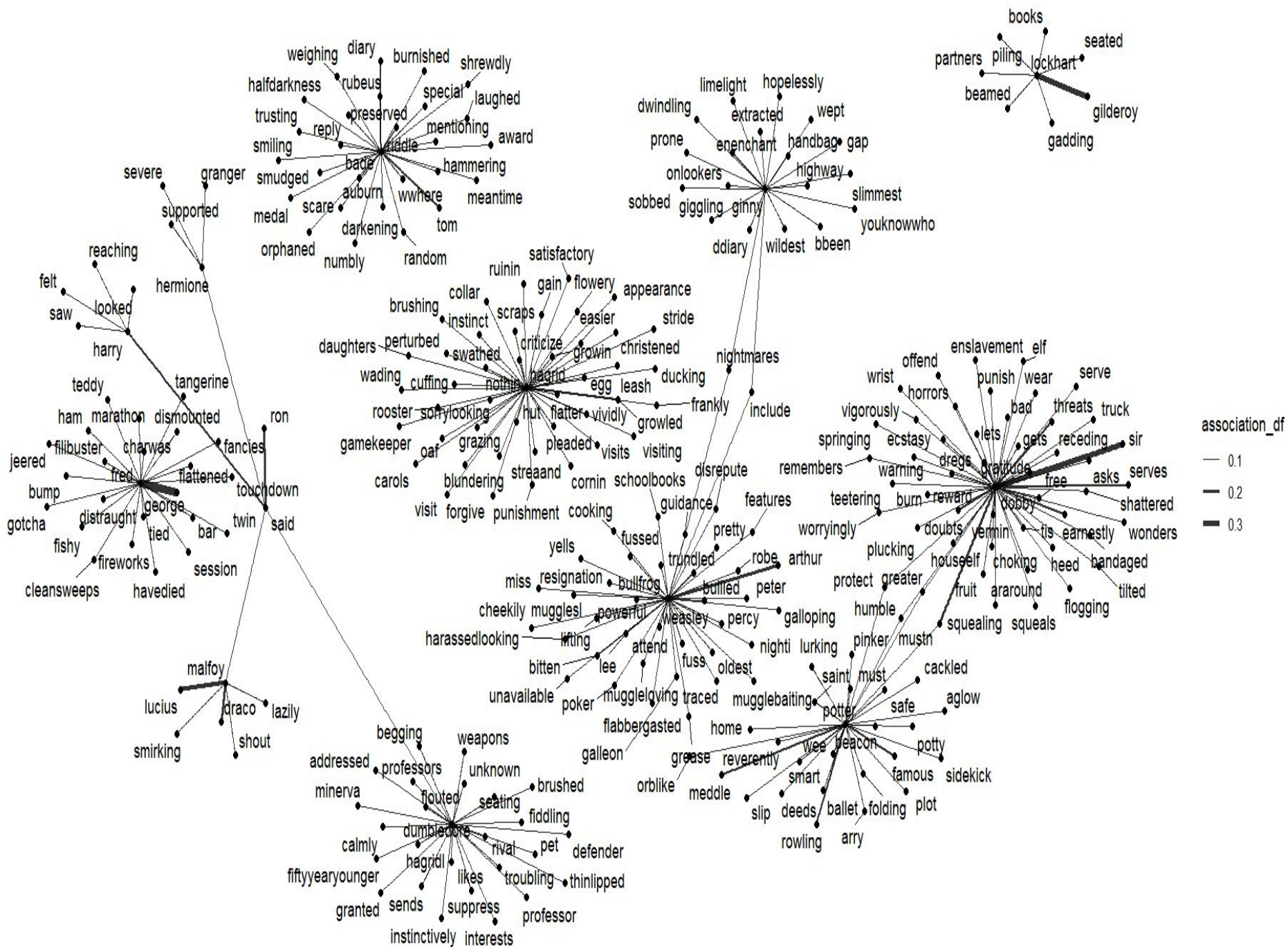


## First three words in the data frame

main_words	words	score
harry	said	0.16
harry	looked	0.08
harry	reaching	0.07
ron	said	0.21
hermione	said	0.1
hermione	granger	0.08
hermione	supported	0.08
lockhart	gilderoy	0.32
lockhart	books	0.11
lockhart	gadding	0.1
malfoy	lucius	0.26
malfoy	draco	0.22
malfoy	smirking	0.11
weasley	arthur	0.19
weasley	miss	0.09
weasley	lee	0.09
weasley	percy	0.08
hagrid	growled	0.14
hagrid	gamekeeper	0.09
hagrid	visiting	0.08

main_words	words	score
dumbledore	addressed	0.11
dumbledore	said	0.09
dumbledore	professor	0.08
dobby	sir	0.29
dobby	mustn	0.21
dobby	serves	0.14
riddle	tom	0.17
riddle	diary	0.16
riddle	reply	0.13
potter	meddle	0.17
potter	rowling	0.15
potter	famous	0.14
ginny	enchant	0.18
ginny	sobbed	0.11
ginny	ddiary	0.09
fred	george	0.39
fred	filibuster	0.12
fred	twin	0.1







# CONCLUSION



- The book is suitable for reading due to its neutral writings
- A character only appears in one book also plays an important role to the story line - Gilderoy Lockhart
- The association of words based on the size of text - very small relation for harry & potter; ron & weasley
- Overall text sentiment we can conclude that the story is full with mixed emotions which balance between negative and positive