EXTRACTIVE TEXT SUMMARIZATION

In [1]:

```
import numpy as np
import pandas as pd
import warnings
warnings.filterwarnings('ignore')
```

1. Importing The Provided CSV File

In [2]:

```
df1 = pd.read_csv("./news.csv")
df1.head()
```

Out[2]:

	title	content	published_at	source	topic
0	BTS: RM is reminded of Bon Voyage as he travel...	After reaching his hotel in the city, RM revea...	2022-07-30T07:00:00Z	2	13
1	RM recalls wondering if he 'made right decisio...	RM aka Kim Namjoon was the first member to joi...	2022-12-22T15:57:55Z	2	13
2	BTS: J-Hope and RM go bonkers at Billie Eilish...	Billie Eilish's concert was held in Seoul, Sou...	2022-08-16T07:00:00Z	1	7
3	BTS: J-Hope proudly states he raised Jungkook,...	BTS ARMY y'all would be missing the members a ...	2022-12-18T13:08:40Z	1	7
4	BTS: Jin aka Kim Seokjin takes us through the ...	BTS member Kim Seokjin aka Jin has the capacit...	2022-11-21T08:00:00Z	1	8

2. Discovering The Irregularities In The Data

In [3]:

```
df1.shape
```

Out[3]:

```
(810, 5)
```

2.1 Randomly selectingout the row, to see what are the things that needs to be pre-processed

In [4]:

```
df1['content'].iloc[401]
```

Out[4]:

```
'BTS consists of members Jin, Suga, J-Hope, RM, Jimin, V, and Jungkook&mdash;co-writes and co-produces much of thei
r own material. Originally a hip hop group, their musical style has evolved to incorporate a wide range of genres;
their lyrics have often discussed mental health, the troubles of school-age youth and coming of age, loss, the jour
ney towards self-love, and individualism. Their work also frequently references literature, philosophy and psycholo
gical concepts, and includes an alternate universe storyline.\n\nBTS in the beginning:\n\nAfter launching in 2013 w
ith their single album 2 Cool 4 Skool, BTS respectively released their first Korean-language studio album, Dark &am
p; Wild, and Japanese-language studio album, Wake Up, in 2014. The group&#39;s second Korean studio album, Wings (2
016), was their first to sell one million copies in South Korea. By 2017, BTS had crossed into the global music mar
ket, leading the Korean wave into the United States and breaking several sales records. They became the first Korea
n ensemble to receive a Gold certification from the Recording Industry Association of America (RIAA) for their sing
le Mic Drop, as well as the first act from South Korea to top the Billboard 200 with their studio album Love Yourse
lf: Tear (2018). \n\nBTS&#39; achievements: \n\nBTS became one of the few groups since the Beatles in 196
6&ndash;1968 with four US number-one albums in less than two years, and Love Yourself: Answer (2018) was the first
Korean album certified Platinum by the RIAA. In 2020, BTS became the first all-South Korean act to reach number one
on the Billboard Hot 100 and Billboard Global 200 with their Grammy-nominated single Dynamite. Their follow-up rele
ases Savage Love, Life Goes On, Butter, and Permission to Dance made them the quickest act to earn four US number-o
ne singles since Justin Timberlake in 2006.\n\nBTS in 2022:\n\nAs of 2022, BTS is the best-selling artist in South
Korean history, having sold in excess of 30 million albums via the Circle Chart, and their studio album Map of the
Soul: 7 (2020) is the best-selling album of all time in South Korea. They are the first non-English-speaking and As
ian act to hold sold-out concerts at Wembley Stadium and the Rose Bowl (Love Yourself World Tour in 2019), and were
named the International Federation of the Phonographic Industry&#39;s (IFPI) Global Recording Artist of the Year fo
r both 2020 and 2021.\nALSO READ: Watch: BTS teases hilarious RUN BTS TV On-air show: V hosts, SUGA paints, Jungkoo
k on drums and more\n\nStay updated with the latest Hallyu news on: Instagram, YouTube, Twitter,&nbs
p;Facebook and Snapchat\n\nWhich BTS member did you get? Let us know in the comments below. \n'
```

2.2 Realized that columns other than published_at, source, topic are useless, and would contribute nothing, rather than increasing dimensionality. So decided to drop them

In [5]:

```
df1.drop(['published_at', 'source', 'topic'], axis=1)
```

Out[5]:

	title	content
0	BTS: RM is reminded of Bon Voyage as he travel...	After reaching his hotel in the city, RM revea...
1	RM recalls wondering if he 'made right decisio...	RM aka Kim Namjoon was the first member to joi...
2	BTS: J-Hope and RM go bonkers at Billie Eilish...	Billie Eilish's concert was held in Seoul, Sou...
3	BTS: J-Hope proudly states he raised Jungkook,...	BTS ARMY y'all would be missing the members a ...
4	BTS: Jin aka Kim Seokjin takes us through the ...	BTS member Kim Seokjin aka Jin has the capacit...
...	...	...
805	BTS' SUGA's Suchwita Ep 2 Teaser OUT: Top 3 so...	BTS has conquered the world with their group r...
806	BTS ARMY celebrate 700 days of Jin's special s...	Today marks 700 days since BTS' worldwide hand...
807	BTS: 'I am not a baby,' says Jungkook as an AR...	BTS' youngest member Jungkook came online on W...
808	BTS' Jin shares 1st pics after joining militar...	BTS' eldest member Jin has shared pictures and...
809	Bad Decisions: BTS' Jin, Jimin, V & Jungkook j...	After a lot of teasing, Benny Blanco's collabo...

810 rows × 2 columns

In [6]:

```
df1.drop(['published_at', 'source', 'topic'], axis=1, inplace=True)
df1.head()
```

Out[6]:

	title	content
0	BTS: RM is reminded of Bon Voyage as he travel...	After reaching his hotel in the city, RM revea...
1	RM recalls wondering if he 'made right decisio...	RM aka Kim Namjoon was the first member to joi...
2	BTS: J-Hope and RM go bonkers at Billie Eilish...	Billie Eilish's concert was held in Seoul, Sou...
3	BTS: J-Hope proudly states he raised Jungkook,...	BTS ARMY y'all would be missing the members a ...
4	BTS: Jin aka Kim Seokjin takes us through the ...	BTS member Kim Seokjin aka Jin has the capacit...

2.3 Checking if there are any null values in the dataframe

In [7]:

```
df1.isna().sum()
```

Out[7]:

```
title      0
content    4
dtype: int64
```

2.4 As there are null values and are very less (4 <<< 810), its better to drop them

In [8]:

```
df1.dropna(axis=0, inplace=True)
df1.head()
```

Out[8]:

	title	content
0	BTS: RM is reminded of Bon Voyage as he travel...	After reaching his hotel in the city, RM revea...
1	RM recalls wondering if he 'made right decisio...	RM aka Kim Namjoon was the first member to joi...
2	BTS: J-Hope and RM go bonkers at Billie Eilish...	Billie Eilish's concert was held in Seoul, Sou...
3	BTS: J-Hope proudly states he raised Jungkook,...	BTS ARMY y'all would be missing the members a ...
4	BTS: Jin aka Kim Seokjin takes us through the ...	BTS member Kim Seokjin aka Jin has the capacit...

In [9]:

```
df1.shape
```

Out[9]:

```
(806, 2)
```

In [10]:

```
df1['content'].iloc[0]
```

Out[10]:

```
'After reaching his hotel in the city, RM revealed that his stay would be for four days and added that he would ste
p out for dinner. As he sat at a roadside open-air restaurant, RM feasted on beer, burgers and fries. He said, "I
\'m starving right now. I\'m out to grab some food. It\'s much quieter than I expected and feels like a rural town.
I like the familiar atmosphere." RM attended Art Basel and explained on camera the details of the art fair. He also
gave a glimpse as he had noodles and beer which was followed by soup noodles and wrap. Showing the pattern of a pin
g pong table, RM said, "The table looks like our (BTS) symbol." He also spoke about the art pieces as he viewed the
m. After that, RM took a tram to visit the Foundation Beyeler, a museum. He later took a walk through the city. On
his third day, RM visited the Kunstmuseum Basel, the Vitra Design Museum and the gallery. As he walked around, RM s
howed a chair to his fans and said, "I have breaking news for you guys. Coldplay\'s Chris Martin made a chair and i
t\'s displayed in the Vitra Design Museum. If you see this Chris, give me a call. You\'re amazing." RM next visited
Lucerne and hiked to Mount Rigi. Recalling his previous visit to Lucerne, RM added, "I remember the day of crossing
that bridge and buying souvenirs." He was also reminded of Bon Voyage, a reality show featuring BTS members RM, Ji
n, Suga, J-Hope, Jimin, V and Jungkook. Speaking to the camera, RM said, "I rode the SSB train to Lucerne, rode a b
oat, rode the mountain train, walked down the track road, rode the cable cars, and now I\'m on a boat planning to g
o ride the SSB again." RM\'s travel in Switzerland ended with a visit to the Museum Tinguely. Next, RM flew to Pari
s to attend the Pinault Collection and to visit Musee d\'Orsay. He then went to Centre Georges-Pompidou and Orsay M
useum. RM\'s vlog ended with him enjoying a Korean meal and then heading back to Seoul.'
```

In the above data, we can see that there are many contractions, puntuations

In [11]:

```
df1['content'].iloc[365]
```

Out[11]:

```
'BTS&rsquo; Jin is good at a lot of things, singing, being &lsquo;worldwide handsome&rsquo;, cooking for Bangtan, a
nd now even making alcohol  The second week of his own program, Drunken Truth, which starred well-known chef Baek J
ong Won and saw a guest appearance from actor Kim Nam Gil, unfolds in unique ways \n\nDrunken Truth Episode 3\nAs J
in and Baek Jong Won reach a spot to take care of the final steps of his first try at making makgeolli (rice wine),
the two show another attempt at their flourishing synergy  Jin is known to be good with elders as he can joke aroun
d and that is clearly visible in the show  The two head out with their own bottles of self-made alcohol, their dest
ination is a traditional market  Being a chef, it is Baek Jong Won&rsquo;s territory who is recognized by the vendo
rs which consist of mostly older women and is asked for photos, while funnily enough, they are unaware of BTS membe
r Jin  A neck-and-neck competition between them shows the difference in preferences between the older and younger g
enerations \nxBrxoby158IDrunken Truth Episode 4\nBaek Jong Won and Jin are just getting ready to make some side dis
hes to go along with their drinking session when a guest makes himself known  Popular actor Kim Nam Gil&rsquo;s cas
ual presence adds fun to the episode  The older ones take turns poking fun at Jin who in turn is his cheeky self ar
ound them, displaying his closeness with the two  They enjoy the food and alcohol made by the two and Jin&rsquo;s k
eeps cracking jokes throughout as is his personality  The show focuses on the chef&rsquo;s aim to keep the integrit
y and bring back the popularity of traditional Korean alcohol- soju  Kim Nam Gil is a silent observer who was invit
ed impromptu by Jin but adds his meaningful remarks throughout  As Jin gifts a set of his famous pyjamas to Baek Jo
ng Won, the night runs deeper  The next morning, they are back at it again, appreciating Jin&rsquo;s skills \n\nJin
heads to meet the alcohol maestro Park Rokdam to ask for giving him the recipe to him so he can be able to share it
with the rest of the BTS members  The artisan names this alcohol &lsquo;Brothers of Four Seas&rsquo;  Jin is congra
tulated for his fantastic first effort \nzVKX6VRMzxsStay updated with the latest Hallyu news on: Instagram, YouTub
e, Twitter, Facebook and Snapchat\nALSO READ: Emergency Declaration Review: Im Siwan’s stirring portrayal makes way
for Kim Nam Gil &amp;amp; Lee Byung Hun’s flight'
```

In the above data, we can see there are certain HTML entities like &rsquo. Also we can see '\n' many a times continously

In [12]:

```
text = df1['content'].iloc[369]
text
```

Out[12]:

```
'2022 came to a celebratory end for BTS, especially member J-Hope, who was in New York, making another fabulous dis
play of his skills  While away from the members, he seemed to have enjoyed it to the fullest with a solo performanc
e at Dick Clark&rsquo;s New Year&rsquo;s Rockin&rsquo; Eve where he did a live stage of 3 songs, solo track &lsquo;
= (Equal Sign)&rsquo;, &lsquo;Chicken Noodle Soup&rsquo; his collaboration track with Becky G, and BTS&rsquo; &lsqu
o;Butter&rsquo; (Holiday Remix)  He officially became only the second South Korean soloist to perform at the event,
following PSY  This was also J-Hope&rsquo;s third time at Dick Clark&rsquo;s New Year&rsquo;s Rockin&rsquo; Eve aft
er group stages with BTS in 2017 and 2019 \n\nJ-Hope&rsquo;s live\n\nAfter returning from Times Square where he per
formed as the penultimate act alongside multiple other singers from around the world and spoke to Ryan Seacrest, th
e host of the show, J-Hope was back at his hotel room and turned on a live broadcast to speak with his fans  As the
y congratulated him on his sparkling performance, being the perfectionist that he is, J-Hope expressed his sadness
about being unable to use his voice to the fullest and talked about the slipping incident due to the rainy weather
during his rehearsal stage  \n\nJ-Hope on other BTS members&rsquo; New Year wishes\n\nAs soon as it turned 12,
and the New Year began in South Korea, BTS members Jungkook, Jimin, and V took to the fan community platform Wevers
e to share their wishes with the fans as well as discuss their plans for the coming year  Jungkook and V kept it br
ief, wishing for a successful and happy year ahead while Jimin wrote a big letter to his fans about the feelings he
had seemingly bottled for a long time about his wishes to release new music soon for which he has been meeting up w
ith composers \n\nThousands of miles away, member J-Hope was his cheeky self as he unleashed all his love for his f
ellow BTS members  He began commenting &lsquo;love you&rsquo; on their posts with cute words and received laughs fr
om them in return \n\nJin&rsquo;s call to J-Hope\n\nOn being asked about BTS&rsquo; oldest member Jin who became th
e first from the group to enlist in the military on December 13, J-Hope recalled how he was called by the member wi
th a different phone number and almost missed it  J-Hope said, &ldquo;Right before I was about to sleep on the 31s
t, I got a call from Jin and asked him how he was doing to which he said J-Hope, pick up your phone  I told him I d
id not know this number, how would I know it was his?&rdquo;\n\nThe &lsquo;Arson&rsquo; hitmaker spoke with a smile
on his face about how he felt happy hearing Jin&rsquo;s voice  It comforted him and J-Hope mentioned how he remembe
red all those moments he spent with Jin  He assured the fans by saying that Jin seemed to be healthy and doing well
in the military  So in place of Jin, he shared that he was well and asked the fans to not worry  The lovely duo, BT
S&rsquo; Jin and J-Hope nicknamed 2seok have always lightened the fans&rsquo; hearts with this interaction \nh
ttps://twitter com/nightstar1201/status/1609428759454822401Jimin X Taeyang\n\nBTS member Jimin and BIGBANG member T
aeyang are a collaboration nobody would have expected  However, in December 2022, it was reported by industry offic
ials that the two are working together on a release that will soon be revealed to the world  To this, Taeyang&rsqu
o;s then-agency, YG Entertainment, which he has since departed, replied by saying that they cannot confirm anything
at the moment and asked fans to anticipate Taeyang&rsquo;s activities  \n\nOn January 1, Taeyang who is curren
tly a free agent, not having signed with anyone for his solo activities as he continues to be with YG Entertainment
for any content related to BIGBANG, shared a rare update on his Instagram account  2 photos were shared with a capt
ion of the hashtag #2023  In black and white, 2 people could be seen in the photos, their backs to the camera  Whil
e one could easily be spotted as Taeyang himself, the other one was not tagged  \n\n
\n                                                                                Fans went into action and soon be
gan investigating every detail of the photo  The other person seemed to be Jimin, as his hands, fluffy hair, and ea
rrings seem to match the BTS member  Though no official announcement was made from either side, it seems that the s
ingers have decided to give a green signal from their ends  Taeyang is rumoured to have been preparing for a Januar
y 2023 comeback so it seems as though we can expect the reports any time now  Once confirmed, this could very well
be the biggest release of the year and one of the most hyped collaborations in K-pop history!\n\nMeanwhile, BIGBANG
&rsquo;s G-Dragon also announced that he is working on an album and hopes to release it in 2023 \n\nStay updated wi
th the latest Hallyu news on: Instagram, YouTube, Twitter, Facebook and Snapchat\nALSO READ: BTS’ J-Hope, TXT wow a
t Dick Clark’s New Year’s Rockin’ Eve: 5 highlights from their stages'
```

The data above, is filled with empty spaces, '\t', '\n' characters, HTML entities, punctuations and contractions. And all these has to be removed before text summarization process

In [13]:

```
df1['content'].iloc[450]
```

Out[13]:

```
'Actor Son Ye-jin, of Crash Landing on You fame, had once revealed that she wanted to treat BTS members RM, Jin, Su
ga, J-Hope, Jimin, V and Jungkook to a meal  At the 2018 Korean Popular Culture &amp; Arts Awards, Son Ye-jin was a
sked if she will buy a meal for anyone in the audience  (Also Read | BTS ARMY says they are \'getting deals\' for t
he band after convincing singer Pink Sweat$ for a collaboration)In a video, shared by a fan account on YouTube, she
had said, "After the drama, there are so many people asking me to buy food  So, I\'m trying not to meet people " Th
e host of the event, asked her, "Is there anyone here that you want to buy a meal for?" After thinking for a momen
t, she turned around, smiled and replied, "BTS "Amid hooting, Son Ye-jin was seen laughing  While RM too laughed, J
in flashed finger hearts, Suga clapped and bowed his head and J-Hope smiled at the actor\'s response  Jimin smiled,
V made fists and Jungkook bowed his head However, what Son Ye-jin next said left everyone in splits  "But I\'m worr
ied because there are so many members " While RM laughed and covered his face, the other members were also seen gig
gling  The host added, "Anyway, I hope Son Ye-jin buys food for BTS " Son Ye-jin smiled and nodded In Son Ye-jin\'s
show, Crash Landing on You (2019), BTS was given a nod  In episode seven, a teenage female patient, Hyun Min-ji, as
ked Son Ye-jin\'s character Yoon Se-ri her favourite BTS member Min-ji said that she found BTS\' Jungkook \'charmin
g\' and he is her favourite member  When asked about hers, Se-ri replied Ri Jeong-hyeok  The character was played b
y actor Hyun Bin  Recently, Hyun Bin and Son Ye-jin got married in Seoul, South Korea Meanwhile, BTS recently annou
nced that they will release their new album on June 10  The development on the new album came just after BTS\' Perm
ission to Dance On Stage tour at Allegiant Stadium in Las Vegas, US  The group released their last album BE in Dece
mber 2020  BTS then released two back-to-back English singles Butter and Permission to Dance last year'
```

2.5 There are HTML Tags and Entities, which needs to be cleaned from data

In [14]:

```
from bs4 import BeautifulSoup
```

In [15]:

```
def removeHTMLTagsAndEntities(s):
    return BeautifulSoup(BeautifulSoup(s, "lxml").text, "html.parser")
```

In [16]:

```
removeHTMLTagsAndEntities(df1['content'].iloc[365])
```

Out[16]:

```
BTS’ Jin is good at a lot of things, singing, being ‘worldwide handsome’, cooking for Bangtan, and now even making
alcohol  The second week of his own program, Drunken Truth, which starred well-known chef Baek Jong Won and saw a g
uest appearance from actor Kim Nam Gil, unfolds in unique ways

Drunken Truth Episode 3
As Jin and Baek Jong Won reach a spot to take care of the final steps of his first try at making makgeolli (rice wi
ne), the two show another attempt at their flourishing synergy  Jin is known to be good with elders as he can joke
around and that is clearly visible in the show  The two head out with their own bottles of self-made alcohol, their
destination is a traditional market  Being a chef, it is Baek Jong Won’s territory who is recognized by the vendors
which consist of mostly older women and is asked for photos, while funnily enough, they are unaware of BTS member J
in  A neck-and-neck competition between them shows the difference in preferences between the older and younger gene
rations
xBrxoby158IDrunken Truth Episode 4
Baek Jong Won and Jin are just getting ready to make some side dishes to go along with their drinking session when
a guest makes himself known  Popular actor Kim Nam Gil’s casual presence adds fun to the episode  The older ones ta
ke turns poking fun at Jin who in turn is his cheeky self around them, displaying his closeness with the two  They
enjoy the food and alcohol made by the two and Jin’s keeps cracking jokes throughout as is his personality  The sho
w focuses on the chef’s aim to keep the integrity and bring back the popularity of traditional Korean alcohol- soju
Kim Nam Gil is a silent observer who was invited impromptu by Jin but adds his meaningful remarks throughout  As Ji
n gifts a set of his famous pyjamas to Baek Jong Won, the night runs deeper  The next morning, they are back at it
again, appreciating Jin’s skills

Jin heads to meet the alcohol maestro Park Rokdam to ask for giving him the recipe to him so he can be able to shar
e it with the rest of the BTS members  The artisan names this alcohol ‘Brothers of Four Seas’  Jin is congratulated
for his fantastic first effort
zVKX6VRMzxsStay updated with the latest Hallyu news on: Instagram, YouTube, Twitter, Facebook and Snapchat
ALSO READ: Emergency Declaration Review: Im Siwan’s stirring portrayal makes way for Kim Nam Gil &amp;amp; Lee Byung Hu
n’s flight
```

In [17]:

```
text = removeHTMLTagsAndEntities(text)
text
```

Out[17]:

2022 came to a celebratory end for BTS, especially member J-Hope, who was in New York, making another fabulous disp
lay of his skills While away from the members, he seemed to have enjoyed it to the fullest with a solo performance
at Dick Clark’s New Year’s Rockin’ Eve where he did a live stage of 3 songs, solo track ‘= (Equal Sign)’, ‘Chicken
Noodle Soup’ his collaboration track with Becky G, and BTS’ ‘Butter’ (Holiday Remix) He officially became only the
second South Korean soloist to perform at the event, following PSY This was also J-Hope’s third time at Dick Clar
k’s New Year’s Rockin’ Eve after group stages with BTS in 2017 and 2019

J-Hope’s live

After returning from Times Square where he performed as the penultimate act alongside multiple other singers from a
round the world and spoke to Ryan Seacrest, the host of the show, J-Hope was back at his hotel room and turned on a
live broadcast to speak with his fans As they congratulated him on his sparkling performance, being the perfection
ist that he is, J-Hope expressed his sadness about being unable to use his voice to the fullest and talked about th
e slipping incident due to the rainy weather during his rehearsal stage

J-Hope on other BTS members’ New Year wishes

As soon as it turned 12, and the New Year began in South Korea, BTS members Jungkook, Jimin, and V took to the fan
community platform Weverse to share their wishes with the fans as well as discuss their plans for the coming year
Jungkook and V kept it brief, wishing for a successful and happy year ahead while Jimin wrote a big letter to his f
ans about the feelings he had seemingly bottled for a long time about his wishes to release new music soon for whic
h he has been meeting up with composers

Thousands of miles away, member J-Hope was his cheeky self as he unleashed all his love for his fellow BTS members
He began commenting ‘love you’ on their posts with cute words and received laughs from them in return

Jin’s call to J-Hope

On being asked about BTS’ oldest member Jin who became the first from the group to enlist in the military on Decemb
er 13, J-Hope recalled how he was called by the member with a different phone number and almost missed it J-Hope s
aid, “Right before I was about to sleep on the 31st, I got a call from Jin and asked him how he was doing to which
he said J-Hope, pick up your phone I told him I did not know this number, how would I know it was his?”

The ‘Arson’ hitmaker spoke with a smile on his face about how he felt happy hearing Jin’s voice It comforted him a
nd J-Hope mentioned how he remembered all those moments he spent with Jin He assured the fans by saying that Jin s
eemed to be healthy and doing well in the military So in place of Jin, he shared that he was well and asked the fa
ns to not worry The lovely duo, BTS’ Jin and J-Hope nicknamed 2seok have always lightened the fans’ hearts with th
is interaction
https://twitter (https://twitter) com/nightstar1201/status/1609428759454822401Jimin X Taeyang

BTS member Jimin and BIGBANG member Taeyang are a collaboration nobody would have expected However, in December 20
22, it was reported by industry officials that the two are working together on a release that will soon be revealed
to the world To this, Taeyang’s then-agency, YG Entertainment, which he has since departed, replied by saying that
they cannot confirm anything at the moment and asked fans to anticipate Taeyang’s activities

On January 1, Taeyang who is currently a free agent, not having signed with anyone for his solo activities as he co
ntinues to be with YG Entertainment for any content related to BIGBANG, shared a rare update on his Instagram accou
nt 2 photos were shared with a caption of the hashtag #2023 In black and white, 2 people could be seen in the pho
tos, their backs to the camera While one could easily be spotted as Taeyang himself, the other one was not tagged

Fans went into action and soon bega
n investigating every detail of the photo The other person seemed to be Jimin, as his hands, fluffy hair, and earr
ings seem to match the BTS member Though no official announcement was made from either side, it seems that the sin
gers have decided to give a green signal from their ends Taeyang is rumoured to have been preparing for a January
2023 comeback so it seems as though we can expect the reports any time now Once confirmed, this could very well be
the biggest release of the year and one of the most hyped collaborations in K-pop history!

Meanwhile, BIGBANG’s G-Dragon also announced that he is working on an album and hopes to release it in 2023

Stay updated with the latest Hallyu news on: Instagram, YouTube, Twitter, Facebook and Snapchat
ALSO READ: BTS’ J-Hope, TXT wow at Dick Clark’s New Year’s Rockin’ Eve: 5 highlights from their stages

2.6 There are many spaces, new line characters and also some entire data block with not even a single fullstop/period

Fullstops are important for us, because we are going to tokenize the sentences using fullstops

In [18]:

```
import re
```

In [19]:

```
def addingFullstops(s):
    s = re.sub("\xa0",' ', str(s))
    s = re.sub("(\\n)+", '.', str(s))
    s = re.sub("(\\t)+", '', str(s))
    s = re.sub("    ", '', str(s))
    s = re.sub("(  )",'.', str(s))
    return s
```

In [20]:

```
text = addingFullstops(text)
text
```

Out[20]:

'2022 came to a celebratory end for BTS, especially member J-Hope, who was in New York, making another fabulous dis
play of his skills.While away from the members, he seemed to have enjoyed it to the fullest with a solo performance
at Dick Clark’s New Year’s Rockin’ Eve where he did a live stage of 3 songs, solo track ‘= (Equal Sign)’, ‘Chicken
Noodle Soup’ his collaboration track with Becky G, and BTS’ ‘Butter’ (Holiday Remix).He officially became only the
second South Korean soloist to perform at the event, following PSY.This was also J-Hope’s third time at Dick Clar
k’s New Year’s Rockin’ Eve after group stages with BTS in 2017 and 2019 .J-Hope’s live.After returning from Times S
quare where he performed as the penultimate act alongside multiple other singers from around the world and spoke to
Ryan Seacrest, the host of the show, J-Hope was back at his hotel room and turned on a live broadcast to speak with
his fans.As they congratulated him on his sparkling performance, being the perfectionist that he is, J-Hope express
ed his sadness about being unable to use his voice to the fullest and talked about the slipping incident due to the
rainy weather during his rehearsal stage..J-Hope on other BTS members’ New Year wishes.As soon as it turned 12, and
the New Year began in South Korea, BTS members Jungkook, Jimin, and V took to the fan community platform Weverse to
share their wishes with the fans as well as discuss their plans for the coming year.Jungkook and V kept it brief, w
ishing for a successful and happy year ahead while Jimin wrote a big letter to his fans about the feelings he had s
eemingly bottled for a long time about his wishes to release new music soon for which he has been meeting up with c
omposers .Thousands of miles away, member J-Hope was his cheeky self as he unleashed all his love for his fellow BT
S members.He began commenting ‘love you’ on their posts with cute words and received laughs from them in return .Ji
n’s call to J-Hope.On being asked about BTS’ oldest member Jin who became the first from the group to enlist in the
military on December 13, J-Hope recalled how he was called by the member with a different phone number and almost m
issed it.J-Hope said, “Right before I was about to sleep on the 31st, I got a call from Jin and asked him how he wa
s doing to which he said J-Hope, pick up your phone.I told him I did not know this number, how would I know it was
his?”.The ‘Arson’ hitmaker spoke with a smile on his face about how he felt happy hearing Jin’s voice.It comforted
him and J-Hope mentioned how he remembered all those moments he spent with Jin.He assured the fans by saying that J
in seemed to be healthy and doing well in the military.So in place of Jin, he shared that he was well and asked the
fans to not worry.The lovely duo, BTS’ Jin and J-Hope nicknamed 2seok have always lightened the fans’ hearts with t
his interaction .https://twitter com/nightstar1201/status/1609428759454822401Jimin X Taeyang.BTS member Jimin and B
IGBANG member Taeyang are a collaboration nobody would have expected.However, in December 2022, it was reported by
industry officials that the two are working together on a release that will soon be revealed to the world.To this,
Taeyang’s then-agency, YG Entertainment, which he has since departed, replied by saying that they cannot confirm an
ything at the moment and asked fans to anticipate Taeyang’s activities..On January 1, Taeyang who is currently a fr
ee agent, not having signed with anyone for his solo activities as he continues to be with YG Entertainment for any
content related to BIGBANG, shared a rare update on his Instagram account.2 photos were shared with a caption of th
e hashtag #2023.In black and white, 2 people could be seen in the photos, their backs to the camera.While one could
easily be spotted as Taeyang himself, the other one was not tagged....Fans went into action and soon began investig
ating every detail of the photo.The other person seemed to be Jimin, as his hands, fluffy hair, and earrings seem t
o match the BTS member.Though no official announcement was made from either side, it seems that the singers have de
cided to give a green signal from their ends.Taeyang is rumoured to have been preparing for a January 2023 comeback
so it seems as though we can expect the reports any time now.Once confirmed, this could very well be the biggest re
lease of the year and one of the most hyped collaborations in K-pop history!.Meanwhile, BIGBANG’s G-Dragon also ann
ounced that he is working on an album and hopes to release it in 2023 .Stay updated with the latest Hallyu news on:
Instagram, YouTube, Twitter, Facebook and Snapchat.ALSO READ: BTS’ J-Hope, TXT wow at Dick Clark’s New Year’s Rocki
n’ Eve: 5 highlights from their stages'

Even after cleaning, there are some repeated fullstops without any sentence. Pre-processing it

In [21]:

```
from nltk.tokenize import sent_tokenize
```

In [22]:

```
def addFullStops(s):
    modified_sentences = []
    sentences = s.split(".")
    for i in sentences:
        i = i.strip()
        if len(i) != 0:
            i += '.'
            modified_sentences.append(i)
    return " ". join(modified_sentences)
```

In [23]:

```
text = addFullStops(text)
text
```

Out[23]:

'2022 came to a celebratory end for BTS, especially member J-Hope, who was in New York, making another fabulous dis play of his skills. While away from the members, he seemed to have enjoyed it to the fullest with a solo performanc e at Dick Clark’s New Year’s Rockin’ Eve where he did a live stage of 3 songs, solo track ‘= (Equal Sign)’, ‘Chicke n Noodle Soup’ his collaboration track with Becky G, and BTS’ ‘Butter’ (Holiday Remix). He officially became only t he second South Korean soloist to perform at the event, following PSY. This was also J-Hope’s third time at Dick Cl ark’s New Year’s Rockin’ Eve after group stages with BTS in 2017 and 2019. J-Hope’s live. After returning from Time s Square where he performed as the penultimate act alongside multiple other singers from around the world and spoke to Ryan Seacrest, the host of the show, J-Hope was back at his hotel room and turned on a live broadcast to speak w ith his fans. As they congratulated him on his sparkling performance, being the perfectionist that he is, J-Hope ex pressed his sadness about being unable to use his voice to the fullest and talked about the slipping incident due t o the rainy weather during his rehearsal stage. J-Hope on other BTS members’ New Year wishes. As soon as it turned 12, and the New Year began in South Korea, BTS members Jungkook, Jimin, and V took to the fan community platform We verse to share their wishes with the fans as well as discuss their plans for the coming year. Jungkook and V kept i t brief, wishing for a successful and happy year ahead while Jimin wrote a big letter to his fans about the feeling s he had seemingly bottled for a long time about his wishes to release new music soon for which he has been meeting up with composers. Thousands of miles away, member J-Hope was his cheeky self as he unleashed all his love for his fellow BTS members. He began commenting ‘love you’ on their posts with cute words and received laughs from them in return. Jin’s call to J-Hope. On being asked about BTS’ oldest member Jin who became the first from the group to en list in the military on December 13, J-Hope recalled how he was called by the member with a different phone number and almost missed it. J-Hope said, “Right before I was about to sleep on the 31st, I got a call from Jin and asked him how he was doing to which he said J-Hope, pick up your phone. I told him I did not know this number, how would I know it was his?”. The ‘Arson’ hitmaker spoke with a smile on his face about how he felt happy hearing Jin’s voic e. It comforted him and J-Hope mentioned how he remembered all those moments he spent with Jin. He assured the fans by saying that Jin seemed to be healthy and doing well in the military. So in place of Jin, he shared that he was w ell and asked the fans to not worry. The lovely duo, BTS’ Jin and J-Hope nicknamed 2seok have always lightened the fans’ hearts with this interaction. https://twitter (https://twitter) com/nightstar1201/status/1609428759454822401J imin X Taeyang. BTS member Jimin and BIGBANG member Taeyang are a collaboration nobody would have expected. Howeve r, in December 2022, it was reported by industry officials that the two are working together on a release that will soon be revealed to the world. To this, Taeyang’s then-agency, YG Entertainment, which he has since departed, repli ed by saying that they cannot confirm anything at the moment and asked fans to anticipate Taeyang’s activities. On January 1, Taeyang who is currently a free agent, not having signed with anyone for his solo activities as he conti nues to be with YG Entertainment for any content related to BIGBANG, shared a rare update on his Instagram account. 2 photos were shared with a caption of the hashtag #2023. In black and white, 2 people could be seen in the photos, their backs to the camera. While one could easily be spotted as Taeyang himself, the other one was not tagged. Fans went into action and soon began investigating every detail of the photo. The other person seemed to be Jimin, as hi s hands, fluffy hair, and earrings seem to match the BTS member. Though no official announcement was made from eith er side, it seems that the singers have decided to give a green signal from their ends. Taeyang is rumoured to have been preparing for a January 2023 comeback so it seems as though we can expect the reports any time now. Once confi rmed, this could very well be the biggest release of the year and one of the most hyped collaborations in K-pop his tory!. Meanwhile, BIGBANG’s G-Dragon also announced that he is working on an album and hopes to release it in 2023. Stay updated with the latest Hallyu news on: Instagram, YouTube, Twitter, Facebook and Snapchat. ALSO READ: BTS’ J-Hope, TXT wow at Dick Clark’s New Year’s Rockin’ Eve: 5 highlights from their stages.'

2.7 For certain words, I saw contractions were used, so decided to replace them with actual word

In [24]:

```
import contractions
```

In [25]:

```
def removingContractions(s):
    words = []
    for i in s.split(" "):
        words.append(contractions.fix(i))
    s = " ".join(words)
    return s
```

In [26]:

```
removingContractions("I'd")
```

Out[26]:

'I would'

In [27]:

```
text = removingContractions(text)
text
```

Out[27]:

'2022 came to a celebratory end for BTS, especially member J-Hope, who was in New York, making another fabulous dis
play of his skills. While away from the members, he seemed to have enjoyed it to the fullest with a solo performanc
e at Dick Clark’s New Year’s Rockin’ Eve where he did a live stage of 3 songs, solo track ‘= (Equal Sign)’, ‘Chicke
n Noodle Soup’ his collaboration track with Becky G, and BTS’ ‘Butter’ (Holiday Remix). He officially became only t
he second South Korean soloist to perform at the event, following PSY. This was also J-Hope’s third time at Dick Cl
ark’s New Year’s Rockin’ Eve after group stages with BTS in 2017 and 2019. J-Hope’s live. After returning from Time
s Square where he performed as the penultimate act alongside multiple other singers from around the world and spoke
to Ryan Seacrest, the host of the show, J-Hope was back at his hotel room and turned on a live broadcast to speak w
ith his fans. As they congratulated him on his sparkling performance, being the perfectionist that he is, J-Hope ex
pressed his sadness about being unable to use his voice to the fullest and talked about the slipping incident due t
o the rainy weather during his rehearsal stage. J-Hope on other BTS members’ New Year wishes. As soon as it turned
12, and the New Year began in South Korea, BTS members Jungkook, Jimin, and V took to the fan community platform We
verse to share their wishes with the fans as well as discuss their plans for the coming year. Jungkook and V kept i
t brief, wishing for a successful and happy year ahead while Jimin wrote a big letter to his fans about the feeling
s he had seemingly bottled for a long time about his wishes to release new music soon for which he has been meeting
up with composers. Thousands of miles away, member J-Hope was his cheeky self as he unleashed all his love for his
fellow BTS members. He began commenting ‘love you’ on their posts with cute words and received laughs from them in
return. Jin’s call to J-Hope. On being asked about BTS’ oldest member Jin who became the first from the group to en
list in the military on December 13, J-Hope recalled how he was called by the member with a different phone number
and almost missed it. J-Hope said, “Right before I was about to sleep on the 31st, I got a call from Jin and asked
him how he was doing to which he said J-Hope, pick up your phone. I told him I did not know this number, how would
I know it was his?”. The ‘Arson’ hitmaker spoke with a smile on his face about how he felt happy hearing Jin’s voic
e. It comforted him and J-Hope mentioned how he remembered all those moments he spent with Jin. He assured the fans
by saying that Jin seemed to be healthy and doing well in the military. So in place of Jin, he shared that he was w
ell and asked the fans to not worry. The lovely duo, BTS’ Jin and J-Hope nicknamed 2seok have always lightened the
fans’ hearts with this interaction. https://twitter (https://twitter) com/nightstar1201/status/1609428759454822401J
imin X Taeyang. BTS member Jimin and BIGBANG member Taeyang are a collaboration nobody would have expected. Howeve
r, in December 2022, it was reported by industry officials that the two are working together on a release that will
soon be revealed to the world. To this, Taeyang’s then-agency, YG Entertainment, which he has since departed, repli
ed by saying that they cannot confirm anything at the moment and asked fans to anticipate Taeyang’s activities. On
January 1, Taeyang who is currently a free agent, not having signed with anyone for his solo activities as he conti
nues to be with YG Entertainment for any content related to BIGBANG, shared a rare update on his Instagram account.
2 photos were shared with a caption of the hashtag #2023. In black and white, 2 people could be seen in the photos,
their backs to the camera. While one could easily be spotted as Taeyang himself, the other one was not tagged. Fans
went into action and soon began investigating every detail of the photo. The other person seemed to be Jimin, as hi
s hands, fluffy hair, and earrings seem to match the BTS member. Though no official announcement was made from eith
er side, it seems that the singers have decided to give a green signal from their ends. Taeyang is rumoured to have
been preparing for a January 2023 comeback so it seems as though we can expect the reports any time now. Once confi
rmed, this could very well be the biggest release of the year and one of the most hyped collaborations in K-pop his
tory!. Meanwhile, BIGBANG’s G-Dragon also announced that he is working on an album and hopes to release it in 2023.
Stay updated with the latest Hallyu news on: Instagram, YouTube, Twitter, Facebook and Snapchat. ALSO READ: BTS’ J-
Hope, TXT wow at Dick Clark’s New Year’s Rockin’ Eve: 5 highlights from their stages.'

2.8 Till here combining all the process into a single function

In [28]:

```
def initial_preprocessing(s):
    s = removeHTMLTagsAndEntities(s)
    s = addingFullstops(s)
    s = removingContractions(s)
    return s
```

In [29]:

```
initial_preprocessing(df1['content'].iloc[369])
```

Out[29]:

'2022 came to a celebratory end for BTS, especially member J-Hope, who was in New York, making another fabulous dis play of his skills.While away from the members, he seemed to have enjoyed it to the fullest with a solo performance at Dick Clark’s New Year’s Rockin’ Eve where he did a live stage of 3 songs, solo track ‘= (Equal Sign)’, ‘Chicken Noodle Soup’ his collaboration track with Becky G, and BTS’ ‘Butter’ (Holiday Remix).He officially became only the second South Korean soloist to perform at the event, following PSY.This was also J-Hope’s third time at Dick Clar k’s New Year’s Rockin’ Eve after group stages with BTS in 2017 and 2019 .J-Hope’s live.After returning from Times S quare where he performed as the penultimate act alongside multiple other singers from around the world and spoke to Ryan Seacrest, the host of the show, J-Hope was back at his hotel room and turned on a live broadcast to speak with his fans.As they congratulated him on his sparkling performance, being the perfectionist that he is, J-Hope express ed his sadness about being unable to use his voice to the fullest and talked about the slipping incident due to the rainy weather during his rehearsal stage..J-Hope on other BTS members’ New Year wishes.As soon as it turned 12, and the New Year began in South Korea, BTS members Jungkook, Jimin, and V took to the fan community platform Weverse to share their wishes with the fans as well as discuss their plans for the coming year.Jungkook and V kept it brief, w ishing for a successful and happy year ahead while Jimin wrote a big letter to his fans about the feelings he had s eemingly bottled for a long time about his wishes to release new music soon for which he has been meeting up with c omposers .Thousands of miles away, member J-Hope was his cheeky self as he unleashed all his love for his fellow BT S members.He began commenting ‘love you’ on their posts with cute words and received laughs from them in return .Ji n’s call to J-Hope.On being asked about BTS’ oldest member Jin who became the first from the group to enlist in the military on December 13, J-Hope recalled how he was called by the member with a different phone number and almost m issed it.J-Hope said, “Right before I was about to sleep on the 31st, I got a call from Jin and asked him how he wa s doing to which he said J-Hope, pick up your phone.I told him I did not know this number, how would I know it was his?”.The ‘Arson’ hitmaker spoke with a smile on his face about how he felt happy hearing Jin’s voice.It comforted him and J-Hope mentioned how he remembered all those moments he spent with Jin.He assured the fans by saying that J in seemed to be healthy and doing well in the military.So in place of Jin, he shared that he was well and asked the fans to not worry.The lovely duo, BTS’ Jin and J-Hope nicknamed 2seok have always lightened the fans’ hearts with t his interaction .https://twitter com/nightstar1201/status/1609428759454822401Jimin X Taeyang.BTS member Jimin and B IGBANG member Taeyang are a collaboration nobody would have expected.However, in December 2022, it was reported by industry officials that the two are working together on a release that will soon be revealed to the world.To this, Taeyang’s then-agency, YG Entertainment, which he has since departed, replied by saying that they cannot confirm an ything at the moment and asked fans to anticipate Taeyang’s activities..On January 1, Taeyang who is currently a fr ee agent, not having signed with anyone for his solo activities as he continues to be with YG Entertainment for any content related to BIGBANG, shared a rare update on his Instagram account.2 photos were shared with a caption of th e hashtag #2023.In black and white, 2 people could be seen in the photos, their backs to the camera.While one could easily be spotted as Taeyang himself, the other one was not tagged....Fans went into action and soon began investig ating every detail of the photo.The other person seemed to be Jimin, as his hands, fluffy hair, and earrings seem t o match the BTS member.Though no official announcement was made from either side, it seems that the singers have de cided to give a green signal from their ends.Taeyang is rumoured to have been preparing for a January 2023 comeback so it seems as though we can expect the reports any time now.Once confirmed, this could very well be the biggest re lease of the year and one of the most hyped collaborations in K-pop history!.Meanwhile, BIGBANG’s G-Dragon also ann ounced that he is working on an album and hopes to release it in 2023 .Stay updated with the latest Hallyu news on: Instagram, YouTube, Twitter, Facebook and Snapchat.ALSO READ: BTS’ J-Hope, TXT wow at Dick Clark’s New Year’s Rocki n’ Eve: 5 highlights from their stages'

In [30]:

```
initial_preprocessing(df1['content'].iloc[0])
```

Out[30]:

'After reaching his hotel in the city, RM revealed that his stay would be for four days and added that he would ste p out for dinner. As he sat at a roadside open-air restaurant, RM feasted on beer, burgers and fries. He said, "I a m starving right now. I am out to grab some food. It is much quieter than I expected and feels like a rural town. I like the familiar atmosphere." RM attended Art Basel and explained on camera the details of the art fair. He also g ave a glimpse as he had noodles and beer which was followed by soup noodles and wrap. Showing the pattern of a ping pong table, RM said, "The table looks like our (BTS) symbol." He also spoke about the art pieces as he viewed them. After that, RM took a tram to visit the Foundation Beyeler, a museum. He later took a walk through the city. On his third day, RM visited the Kunstmuseum Basel, the Vitra Design Museum and the gallery. As he walked around, RM showe d a chair to his fans and said, "I have breaking news for you guys. Coldplay\'s Chris Martin made a chair and it is displayed in the Vitra Design Museum. If you see this Chris, give me a call. You are amazing." RM next visited Luce rne and hiked to Mount Rigi. Recalling his previous visit to Lucerne, RM added, "I remember the day of crossing tha t bridge and buying souvenirs." He was also reminded of Bon Voyage, a reality show featuring BTS members RM, Jin, S uga, J-Hope, Jimin, V and Jungkook. Speaking to the camera, RM said, "I rode the SSB train to Lucerne, rode a boat, rode the mountain train, walked down the track road, rode the cable cars, and now I am on a boat planning to go rid e the SSB again." RM\'s travel in Switzerland ended with a visit to the Museum Tinguely. Next, RM flew to Paris to attend the Pinault Collection and to visit Musee d\'Orsay. He then went to Centre Georges-Pompidou and Orsay Museu m. RM\'s vlog ended with him enjoying a Korean meal and then heading back to Seoul.'

2.9 Removing the punctuations and Stopwords and coverting entire sentence into lower case

In [31]:

```
from nltk.corpus import stopwords
```

In [32]:

```
engStopwords = stopwords.words("english")
print(engStopwords)
```

```
['i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you', "you're", "you've", "you'll", "you'd", 'your',
'yours', 'yourself', 'yourselves', 'he', 'him', 'his', 'himself', 'she', "she's", 'her', 'hers', 'herself', 'it',
"it's", 'its', 'itself', 'they', 'them', 'their', 'theirs', 'themselves', 'what', 'which', 'who', 'whom', 'this',
'that', "that'll", 'these', 'those', 'am', 'is', 'are', 'was', 'were', 'be', 'been', 'being', 'have', 'has', 'had',
'having', 'do', 'does', 'did', 'doing', 'a', 'an', 'the', 'and', 'but', 'if', 'or', 'because', 'as', 'until', 'whil
e', 'of', 'at', 'by', 'for', 'with', 'about', 'against', 'between', 'into', 'through', 'during', 'before', 'after',
'above', 'below', 'to', 'from', 'up', 'down', 'in', 'out', 'on', 'off', 'over', 'under', 'again', 'further', 'the
n', 'once', 'here', 'there', 'when', 'where', 'why', 'how', 'all', 'any', 'both', 'each', 'few', 'more', 'most', 'o
ther', 'some', 'such', 'no', 'nor', 'not', 'only', 'own', 'same', 'so', 'than', 'too', 'very', 's', 't', 'can', 'wi
ll', 'just', 'don', "don't", 'should', "should've", 'now', 'd', 'll', 'm', 'o', 're', 've', 'y', 'ain', 'aren', "ar
en't", 'couldn', "couldn't", 'didn', "didn't", 'doesn', "doesn't", 'hadn', "hadn't", 'hasn', "hasn't", 'haven', "ha
ven't", 'isn', "isn't", 'ma', 'mightn', "mightn't", 'mustn', "mustn't", 'needn', "needn't", 'shan', "shan't", 'shou
ldn', "shouldn't", 'wasn', "wasn't", 'weren', "weren't", 'won', "won't", 'wouldn', "wouldn't"]
```

In [33]:

```
def removePunctuations(s):
    words = []
    s = re.sub(r"[<>()|&©ø\[\]\'\",;?~*!]", ' ', str(s))
    for i in s.split(" "):
        if i not in engStopwords:
            words.append(i)
    s = " ".join(words)
    return s
```

In [34]:

```
removePunctuations(text)
```

Out[34]:

```
'2022 came celebratory end BTS  especially member J-Hope  New York  making another fabulous display skills. While a
way members  seemed enjoyed fullest solo performance Dick Clark’s New Year’s Rockin’ Eve live stage 3 songs  solo t
rack ‘=  Equal Sign ’  ‘Chicken Noodle Soup’ collaboration track Becky G  BTS’ ‘Butter’  Holiday Remix . He officia
lly became second South Korean soloist perform event  following PSY. This also J-Hope’s third time Dick Clark’s New
Year’s Rockin’ Eve group stages BTS 2017 2019. J-Hope’s live. After returning Times Square performed penultimate ac
t alongside multiple singers around world spoke Ryan Seacrest  host show  J-Hope back hotel room turned live broadc
ast speak fans. As congratulated sparkling performance  perfectionist  J-Hope expressed sadness unable use voice fu
llest talked slipping incident due rainy weather rehearsal stage. J-Hope BTS members’ New Year wishes. As soon turn
ed 12  New Year began South Korea  BTS members Jungkook  Jimin  V took fan community platform Weverse share wishes
fans well discuss plans coming year. Jungkook V kept brief  wishing successful happy year ahead Jimin wrote big let
ter fans feelings seemingly bottled long time wishes release new music soon meeting composers. Thousands miles away
member J-Hope cheeky self unleashed love fellow BTS members. He began commenting ‘love you’ posts cute words receiv
ed laughs return. Jin’s call J-Hope. On asked BTS’ oldest member Jin became first group enlist military December 13
J-Hope recalled called member different phone number almost missed it. J-Hope said  “Right I sleep 31st  I got call
Jin asked said J-Hope  pick phone. I told I know number  would I know ”. The ‘Arson’ hitmaker spoke smile face felt
happy hearing Jin’s voice. It comforted J-Hope mentioned remembered moments spent Jin. He assured fans saying Jin s
eemed healthy well military. So place Jin  shared well asked fans worry. The lovely duo  BTS’ Jin J-Hope nicknamed
2seok always lightened fans’ hearts interaction. https://twitter (https://twitter) com/nightstar1201/status/1609428
759454822401Jimin X Taeyang. BTS member Jimin BIGBANG member Taeyang collaboration nobody would expected. However
December 2022  reported industry officials two working together release soon revealed world. To  Taeyang’s then-age
ncy  YG Entertainment  since departed  replied saying cannot confirm anything moment asked fans anticipate Taeyan
g’s activities. On January 1  Taeyang currently free agent  signed anyone solo activities continues YG Entertainmen
t content related BIGBANG  shared rare update Instagram account. 2 photos shared caption hashtag #2023. In black wh
ite  2 people could seen photos  backs camera. While one could easily spotted Taeyang  one tagged. Fans went action
soon began investigating every detail photo. The person seemed Jimin  hands  fluffy hair  earrings seem match BTS m
ember. Though official announcement made either side  seems singers decided give green signal ends. Taeyang rumoure
d preparing January 2023 comeback seems though expect reports time now. Once confirmed  could well biggest release
year one hyped collaborations K-pop history . Meanwhile  BIGBANG’s G-Dragon also announced working album hopes rele
ase 2023. Stay updated latest Hallyu news on: Instagram  YouTube  Twitter  Facebook Snapchat. ALSO READ: BTS’ J-Hop
e  TXT wow Dick Clark’s New Year’s Rockin’ Eve: 5 highlights stages.'
```

2.10 Trying to combine all of these pre-processing in a single function

In [35]:

```
def preprocessingRow(s):
    s = BeautifulSoup(BeautifulSoup(s, "lxml").text, "html.parser")
    s = re.sub("\xa0",' ', str(s))
    s = re.sub("(\\n)+", '.', str(s))
    s = re.sub("(\\t)+", '', str(s))
    s = re.sub("    ", '', str(s))
    s = re.sub("(  )",'.', str(s))
    modified_sentences = []
    sentences = s.split(".")
    for i in sentences:
        i = i.strip()
        if len(i) != 0:
            i += '.'
            modified_sentences.append(i)
    s = " ". join(modified_sentences)
    words = []
    for i in s.split(" "):
        words.append(contractions.fix(i))
    s = " ".join(words)
    words = []
    s = re.sub(r"[<>()|&©ø\[\]\'\",;?~*!]", ' ', str(s)).lower()
    for i in s.split(" "):
        if i not in engStopwords and i not in words:
            words.append(i)
    s = " ".join(words)
    s = re.sub(' +', ' ', s)
    return s
```

In [36]:

```
preprocessingRow(df1['content'].iloc[369])
```

Out[36]:

```
'2022 came celebratory end bts especially member j-hope new york making another fabulous display skills. away membe
rs seemed enjoyed fullest solo performance dick clark’s year’s rockin’ eve live stage 3 songs track ‘= equal sign ’
‘chicken noodle soup’ collaboration becky g bts’ ‘butter’ holiday remix . officially became second south korean sol
oist perform event following psy. also j-hope’s third time group stages 2017 2019. live. returning times square per
formed penultimate act alongside multiple singers around world spoke ryan seacrest host show back hotel room turned
broadcast speak fans. congratulated sparkling perfectionist expressed sadness unable use voice talked slipping inci
dent due rainy weather rehearsal stage. members’ year wishes. soon 12 began korea jungkook jimin v took fan communi
ty platform weverse share wishes fans well discuss plans coming year. kept brief wishing successful happy ahead wro
te big letter feelings seemingly bottled long release music meeting composers. thousands miles cheeky self unleashe
d love fellow members. commenting ‘love you’ posts cute words received laughs return. jin’s call j-hope. asked olde
st jin first enlist military december 13 recalled called different phone number almost missed it. said “right sleep
31st got pick phone. told know would ”. ‘arson’ hitmaker smile face felt hearing voice. comforted mentioned remembe
red moments spent jin. assured saying healthy military. place shared worry. lovely duo nicknamed 2seok always light
ened fans’ hearts interaction. https://twitter (https://twitter) com/nightstar1201/status/160942875945482240 1jimin
x taeyang. bigbang taeyang nobody expected. however reported industry officials two working together revealed worl
d. taeyang’s then-agency yg entertainment since departed replied cannot confirm anything moment anticipate activiti
es. january 1 currently free agent signed anyone activities continues content related rare update instagram accoun
t. 2 photos caption hashtag #2023. black white people could seen backs camera. one easily spotted tagged. went acti
on investigating every detail photo. person hands fluffy hair earrings seem match member. though official announcem
ent made either side seems decided give green signal ends. rumoured preparing 2023 comeback expect reports now. con
firmed biggest hyped collaborations k-pop history meanwhile bigbang’s g-dragon announced album hopes 2023. stay upd
ated latest hallyu news on: youtube twitter facebook snapchat. read: txt wow eve: 5 highlights stages.'
```

Although, I was not able to figure out how to remove a link because, I wasn't able to find any regular expression for removing it. I did research on link removal fuctions through BeautifulSoup but wasnt able to do it.

The above function will work perfectly when provided row by row input but at that time we cannot utilize the multiprocessing functions of NLP. Hence creating same function, but this time we shall pass an entire column rather than a row to function.

In [37]:

```
from tqdm.auto import tqdm
```

In [38]:

```
def preprocessingColumn(col):
    for row in tqdm(col, total=col.shape[0]):
        cleanedData = preprocessingRow(row)
        yield cleanedData
```

In [39]:

```
preprocessingColumn(df1['content']) # will result in a Generator object
```

Out[39]:

```
<generator object preprocessingColumn at 0x000001EC0CB46CF0>
```

2.11 Storing cleaned data in the dataframe

In [40]:

```
cleanedContent = preprocessingColumn(df1['content'])
```

Using spacy to fasten the process of Data Pre-processing

In [41]:

```
import spacy
nlp = spacy.load('en_core_web_sm', disable=['ner', 'parser'])
```

Perorming Data Pre-processing on the Dataframe's content column

In [42]:

```
cleanedText = [str(doc) for doc in nlp.pipe(cleanedContent, batch_size=5000, n_process=-1)]
```

```
  0%|          | 0/806 [00:00<?, ?it/s]
```

In [43]:

```
cleanedText[0]
```

Out[43]:

```
'reaching hotel city rm revealed stay would four days added step dinner. sat roadside open-air restaurant feasted b
eer burgers fries. said starving right now. grab food. much quieter expected feels like rural town. familiar atmosp
here. attended art basel explained camera details fair. also gave glimpse noodles followed soup wrap. showing patte
rn ping pong table looks bts symbol. spoke pieces viewed them. took tram visit foundation beyeler museum. later wal
k city. third day visited kunstmuseum vitra design museum gallery. walked around showed chair fans breaking news gu
ys. coldplay chris martin made displayed see give call. amazing. next lucerne hiked mount rigi. recalling previous
remember crossing bridge buying souvenirs. reminded bon voyage reality show featuring members jin suga j-hope jimin
v jungkook. speaking rode ssb train boat mountain track road cable cars planning go ride again. travel switzerland
ended tinguely. flew paris attend pinault collection musee orsay. went centre georges-pompidou orsay vlog enjoying
korean meal heading back seoul.'
```

In [44]:

```
cleanedText[369]
```

Out[44]:

```
'2022 came celebratory end bts especially member j-hope new york making another fabulous display skills. away membe
rs seemed enjoyed fullest solo performance dick clark’s year’s rockin’ eve live stage 3 songs track ‘= equal sign ’
‘chicken noodle soup’ collaboration becky g bts’ ‘butter’ holiday remix . officially became second south korean sol
oist perform event following psy. also j-hope’s third time group stages 2017 2019. live. returning times square per
formed penultimate act alongside multiple singers around world spoke ryan seacrest host show back hotel room turned
broadcast speak fans. congratulated sparkling perfectionist expressed sadness unable use voice talked slipping inci
dent due rainy weather rehearsal stage. members’ year wishes. soon 12 began korea jungkook jimin v took fan communi
ty platform weverse share wishes fans well discuss plans coming year. kept brief wishing successful happy ahead wro
te big letter feelings seemingly bottled long release music meeting composers. thousands miles cheeky self unleashe
d love fellow members. commenting ‘love you’ posts cute words received laughs return. jin’s call j-hope. asked olde
st jin first enlist military december 13 recalled called different phone number almost missed it. said “right sleep
31st got pick phone. told know would ”. ‘arson’ hitmaker smile face felt hearing voice. comforted mentioned remembe
red moments spent jin. assured saying healthy military. place shared worry. lovely duo nicknamed 2seok always light
ened fans’ hearts interaction. https://twitter (https://twitter) com/nightstar1201/status/160942875945482240 1jimin
x taeyang. bigbang taeyang nobody expected. however reported industry officials two working together revealed worl
d. taeyang’s then-agency yg entertainment since departed replied cannot confirm anything moment anticipate activiti
es. january 1 currently free agent signed anyone activities continues content related rare update instagram accoun
t. 2 photos caption hashtag #2023. black white people could seen backs camera. one easily spotted tagged. went acti
on investigating every detail photo. person hands fluffy hair earrings seem match member. though official announcem
ent made either side seems decided give green signal ends. rumoured preparing 2023 comeback expect reports now. con
firmed biggest hyped collaborations k-pop history meanwhile bigbang’s g-dragon announced album hopes 2023. stay upd
ated latest hallyu news on: youtube twitter facebook snapchat. read: txt wow eve: 5 highlights stages.'
```

In [45]:

```
len(cleanedText)
```

Out[45]:

```
806
```

2.12 Checking for any errors in the process of data cleaning, such that out code might have skipped processing an entire row

In [46]:

```
flag = 1
for i in range(len(cleanedText)):
    if len(cleanedText[i]) == 0:
        print(f"Error at: {i}")
        flag = 0
if flag == 1:
    print("No error, in process of Data pre-processing")
```

```
No error, in process of Data pre-processing
```

3. Trying To Generate Summary For Any One Random Row

In [47]:

```
from gensim.models import Word2Vec
from scipy import spatial
import networkx as nx
```

In [48]:

```
text = initial_preprocessing(df1['content'].iloc[405])
text
```

Out[48]:

```
'BTS’ oldest member Jin went solo after J-Hope and made his official debut with a single album. The Astronaut has b
een meaningful for multiple reasons for the group as well as the fans who have welcomed the release with warm words
and moist eyes..Music charts.The song which came as a result of Jin working with Coldplay for the second time follo
wing ‘My Universe’, debuted on the Billboard Hot100 chart at No. 51. He also tied with PSY for a record on United K
ingdom’s Official Official Singles Chart for grabbing the 61st spot in the week after its release. His latest achie
vement comes with selling 1,024,382 copies of The Astronaut according to the numbers released by Circle Chart (earl
ier known as Gaon Chart)..Third million-seller.The BTS member is only the third soloist in the history of the music
chart to have recorded over a million copies sold of his album. Jin follows EXO member Baekhyun and trot-ballad sin
ger Lim Young Woong on the list. Becoming a million seller is a massive feat for the BTS member who is expected to
enlist for his mandatory military service soon..Jin’s military service.The group’s national duty which had become a
global discussion received an update as their agency announced the member’s decision to enlist. Jin became the firs
t member to apply for the cancellation of his delay notice and will possibly enlist before his upcoming 30th birthd
ay on December 4. He will also possibly be at the front line for his basic military training according to his reply
to a fan on Weverse recently..Stay updated with the latest Hallyu news on: Instagram, YouTube, Twitter, Facebook an
d Snapchat.'
```

In [49]:

```
sentences = [i.strip()+"." for i in text.split(".") if len(i) != 0]
sentences
```

Out[49]:

```
['BTS’ oldest member Jin went solo after J-Hope and made his official debut with a single album.',
 'The Astronaut has been meaningful for multiple reasons for the group as well as the fans who have welcomed the re
lease with warm words and moist eyes.',
 'Music charts.',
 'The song which came as a result of Jin working with Coldplay for the second time following ‘My Universe’, debuted
on the Billboard Hot100 chart at No.',
 '51.',
 'He also tied with PSY for a record on United Kingdom’s Official Official Singles Chart for grabbing the 61st spot
in the week after its release.',
 'His latest achievement comes with selling 1,024,382 copies of The Astronaut according to the numbers released by
Circle Chart (earlier known as Gaon Chart).',
 'Third million-seller.',
 'The BTS member is only the third soloist in the history of the music chart to have recorded over a million copies
sold of his album.',
 'Jin follows EXO member Baekhyun and trot-ballad singer Lim Young Woong on the list.',
 'Becoming a million seller is a massive feat for the BTS member who is expected to enlist for his mandatory milita
ry service soon.',
 'Jin’s military service.',
 'The group’s national duty which had become a global discussion received an update as their agency announced the m
ember’s decision to enlist.',
 'Jin became the first member to apply for the cancellation of his delay notice and will possibly enlist before his
upcoming 30th birthday on December 4.',
 'He will also possibly be at the front line for his basic military training according to his reply to a fan on Wev
erse recently.',
 'Stay updated with the latest Hallyu news on: Instagram, YouTube, Twitter, Facebook and Snapchat.']
```

In [50]:

```
sentence_tokens = []
for sentence in sentences:
    for word in preprocessingRow(sentence).split(" "):
        if word not in sentence_tokens:
            sentence_tokens.append(word)
```

In [51]:

```
w2v = Word2Vec(sentence_tokens, vector_size=1, min_count = 1, epochs = 1000)
sentence_embeddings = [[w2v.wv[word][0] for word in words] for words in sentence_tokens]
max_len = max([len(tokens) for tokens in sentence_tokens])
sentence_embeddings = [np.pad(embedding,(0,max_len-len(embedding)),'constant') for embedding in sentence_embeddings]
```

In [52]:

```
similarity_matrix = np.zeros([len(sentence_tokens), len(sentence_tokens)])
for i,row_embedding in enumerate(sentence_embeddings):
    for j,column_embedding in enumerate(sentence_embeddings):
        similarity_matrix[i][j]=1-spatial.distance.cosine(row_embedding,column_embedding)
```

In [53]:

```
nx_graph = nx.from_numpy_array(similarity_matrix)
scores = nx.pagerank(nx_graph)
```

In [54]:

```
total_summary_sentences = len(sentences) // 4
top_sentence={sentence:scores[index] for index,sentence in enumerate(sentences)}
top=dict(sorted(top_sentence.items(), key=lambda x: x[1], reverse=True)[:total_summary_sentences])
```

In [55]:

```
summary = ""
for sentence in sentences:
    if sentence in top.keys():
        summary += sentence
print(summary)
```

```
The Astronaut has been meaningful for multiple reasons for the group as well as the fans who have welcomed the rele
ase with warm words and moist eyes.Music charts.His latest achievement comes with selling 1,024,382 copies of The A
stronaut according to the numbers released by Circle Chart (earlier known as Gaon Chart).Becoming a million seller
is a massive feat for the BTS member who is expected to enlist for his mandatory military service soon.
```

3.1 For below text:

In [56]:

```
text
```

Out[56]:

```
'BTS’ oldest member Jin went solo after J-Hope and made his official debut with a single album. The Astronaut has b
een meaningful for multiple reasons for the group as well as the fans who have welcomed the release with warm words
and moist eyes..Music charts.The song which came as a result of Jin working with Coldplay for the second time follo
wing ‘My Universe’, debuted on the Billboard Hot100 chart at No. 51. He also tied with PSY for a record on United K
ingdom’s Official Official Singles Chart for grabbing the 61st spot in the week after its release. His latest achie
vement comes with selling 1,024,382 copies of The Astronaut according to the numbers released by Circle Chart (earl
ier known as Gaon Chart)..Third million-seller.The BTS member is only the third soloist in the history of the music
chart to have recorded over a million copies sold of his album. Jin follows EXO member Baekhyun and trot-ballad sin
ger Lim Young Woong on the list. Becoming a million seller is a massive feat for the BTS member who is expected to
enlist for his mandatory military service soon..Jin’s military service.The group’s national duty which had become a
global discussion received an update as their agency announced the member’s decision to enlist. Jin became the firs
t member to apply for the cancellation of his delay notice and will possibly enlist before his upcoming 30th birthd
ay on December 4. He will also possibly be at the front line for his basic military training according to his reply
to a fan on Weverse recently..Stay updated with the latest Hallyu news on: Instagram, YouTube, Twitter, Facebook an
d Snapchat.'
```

3.2 We got summary as:

In [57]:

```
summary
```

Out[57]:

```
'The Astronaut has been meaningful for multiple reasons for the group as well as the fans who have welcomed the rel
ease with warm words and moist eyes.Music charts.His latest achievement comes with selling 1,024,382 copies of The
Astronaut according to the numbers released by Circle Chart (earlier known as Gaon Chart).Becoming a million seller
is a massive feat for the BTS member who is expected to enlist for his mandatory military service soon.'
```

4. Time To Implement It On Actual Dataframe

In [58]:

```
def summarizerRow(s):
    # Applying only the necessary Pre-processing
    text = initial_preprocessing(s)
    # print("1")

    # Generating the tokens
    sentences = [i.strip()+"." for i in text.split(".") if len(i) != 0]
    sentence_tokens = [[word for word in preprocessingRow(sentence).split(" ")] for sentence in sentences]
    # print("2")

    # Calculating the embedding for each word
    w2v = Word2Vec(sentence_tokens, vector_size=1, min_count = 1, epochs = 1000)
    sentence_embeddings = [[w2v.wv[word][0] for word in words] for words in sentence_tokens]
    max_len = max([len(tokens) for tokens in sentence_tokens])
    sentence_embeddings = [np.pad(embedding,(0,max_len-len(embedding)),'constant') for embedding in sentence_embeddings]
    # print("3")

    # Developing the similarity matrix based on cosine similarity
    similarity_matrix = np.zeros([len(sentence_tokens), len(sentence_tokens)])
    for i,row_embedding in enumerate(sentence_embeddings):
        for j,column_embedding in enumerate(sentence_embeddings):
            similarity_matrix[i][j]=1-spatial.distance.cosine(row_embedding,column_embedding)
    # print("4")

    # Implementing the TextRank
    nx_graph = nx.from_numpy_array(similarity_matrix)
    try:
        scores = nx.pagerank(nx_graph,  max_iter=100000, tol=1.0e-2)

        # Sorting out the importing sentences
        total_summary_sentences = len(sentences) // 3
        top_sentence={sentence:scores[index] for index,sentence in enumerate(sentences)}
        top=dict(sorted(top_sentence.items(), key=lambda x: x[1], reverse=True)[:total_summary_sentences])

        # Joining the most important sentences
        summary = ""
        for sentence in sentences:
            if sentence in top.keys():
                summary += sentence + " "
        # print("5")
        return summary
    except:
        summary = ""
        # print("6")
        return summary
```

In [59]:

```
print(summarizerRow(df1['content'].iloc[402]))
```

```
He said the reason why he wanted to write a song for them is because he believes that ARMYs are the reason why the
group exists. The songwriting process to creating the melody, Jungkook spilled all kinds of details. The so-teok (s
ausage tteok) and fried chicken looked so good. He also did not let the tasks overcome his love for SUGA's song 'Th
at That' as he danced to it as well. He also mentioned doing a V-Live there, which was what led fans to know which
member will be putting up what kind of vlog.
```

4.1 Developing a similar fuction column-wise

In [60]:

```
def summarizerColumn(col):
    for row in tqdm(col, total=col.shape[0]):
        summary = summarizerRow(row)
        yield summary
```

In [61]:

```
summaryGenerator = summarizerColumn(df1['content'])
```

In [62]:

```
summary = [str(doc) for doc in nlp.pipe(summaryGenerator, batch_size=5000, n_process=-1)]
```

```
  0%|          | 0/806 [00:00<?, ?it/s]
```

In [63]:

```
summary[0]
```

Out[63]:

```
'As he sat at a roadside open-air restaurant, RM feasted on beer, burgers and fries. " RM attended Art Basel and ex
plained on camera the details of the art fair. He also gave a glimpse as he had noodles and beer which was followed
by soup noodles and wrap. After that, RM took a tram to visit the Foundation Beyeler, a museum. As he walked aroun
d, RM showed a chair to his fans and said, "I have breaking news for you guys. Coldplay\'s Chris Martin made a chai
r and it is displayed in the Vitra Design Museum. " RM next visited Lucerne and hiked to Mount Rigi. " RM\'s travel
in Switzerland ended with a visit to the Museum Tinguely. '
```

In [64]:

```
df1['content'].shape
```

Out[64]:

```
(806,)
```

In [65]:

```
len(summary)
```

Out[65]:

```
806
```

In [66]:

```
df1['content'].iloc[804]
```

Out[66]:

```
'BTS\' eldest member Jin has shared pictures and a message for fans for the first time after he joined the South Ko
rean military  Taking to Weverse on Wednesday Jin posted his pictures including selfies  In a photo, Jin is seen in
his uniform as he stood with his arms on his sides  The singer also wore a mask  (Also Read | BTS’ Jin has a specia
l message for fans: \'I may not be by your side, but…\')In a selfie, Jin looked at the camera giving fans a closer
glimpse of his face  He also flashed the victory sign in another picture  Sharing the pictures, Jin wrote, "I\'m en
joying my life  I\'m posting pictures after getting permission from the military  ARMY, be happy and take care "Jin
\'s message and photos left the BTS ARMY emotional  A person wrote on Twitter, "Even though he must be soo tired bu
t still took permission from there & came to update us about himself & telling us to be happy & be well  I\'m cryin
g  I love you so much Jin ""Jin is proud of all the armys who waited until he posts  Let\'s continue waiting and no
t spreading pics that are not posted by Jin," read a comment  "All I\'m doing right now is staring at these picture
s and crying " "I missed him God I missed him so much  I hope you’re staying warm and healthy," said another fan Ji
n, whose full name is Kim Seok-jin, officially enlisted for duty on December 13, 2022  According to several report
s, Jin is undergoing training at a boot camp of a front-line army division in Yeoncheon, 60 kilometres north of Seo
ul Earlier this month, Bangtan TV had shared a video message of Jin which was recorded during the filming of the Ko
rean variety show, Running Man  In the video Jin had said, “Hello everyone, this is Jin of BTS  I won’t be a civili
an by the time the video is out  But I am here in front of the camera, because I wanted to leave you something, eve
n if it is just leaving a message "He had also said, "Whenever I am available I wish to share these videos with you
I may not be by your side, but I’ll go looking for you soon, if you just wait a little  I’ll be back soon  That’s a
ll for today  Next time when I have the time, I’ll share another video  That’s all for now "In South Korea, all abl
e-bodied men aged 18-28 are required to serve in the military for about two years  All BTS members had been allowed
to put off starting their military service until they turned 30 Other members -- RM, Suga, J-Hope, Jimin, V and Jun
gkook -- plan to carry out their military service based on their own individual plans  The group, which debuted in
2013, had said last year that they hope to reconvene as a unit around 2025 following their service commitment'
```

In [67]:

```
summary[804]
```

Out[67]:

```
'Taking to Weverse on Wednesday Jin posted his pictures including selfies. In a photo, Jin is seen in his uniform a
s he stood with his arms on his sides. He also flashed the victory sign in another picture. ARMY, be happy and take
care "Jin\'s message and photos left the BTS ARMY emotional. I love you so much Jin ""Jin is proud of all the armys
who waited until he posts. Let us continue waiting and not spreading pics that are not posted by Jin," read a comme
nt. In the video Jin had said, “Hello everyone, this is Jin of BTS. I may not be by your side, but I will go lookin
g for you soon, if you just wait a little. '
```

5. Creating a New Dataframe For Storing Only The Required Results

In [68]:

```
df2 = pd.DataFrame()
```

In [69]:

```
df2['Original Content'] = df1['content'].apply(initial_preprocessing)
df2.head()
```

Out[69]:

	Original Content
0	After reaching his hotel in the city, RM revea...
1	RM aka Kim Namjoon was the first member to joi...
2	Billie Eilish's concert was held in Seoul, Sou...
3	BTS ARMY you all would be missing the members ...
4	BTS member Kim Seokjin aka Jin has the capacit...

In [70]:

```
df2.shape
```

Out[70]:

```
(806, 1)
```

In [71]:

```
df2['Original Content'].iloc[369]
```

Out[71]:

```
'2022 came to a celebratory end for BTS, especially member J-Hope, who was in New York, making another fabulous dis
play of his skills.While away from the members, he seemed to have enjoyed it to the fullest with a solo performance
at Dick Clark’s New Year’s Rockin’ Eve where he did a live stage of 3 songs, solo track ‘= (Equal Sign)’, ‘Chicken
Noodle Soup’ his collaboration track with Becky G, and BTS’ ‘Butter’ (Holiday Remix).He officially became only the
second South Korean soloist to perform at the event, following PSY.This was also J-Hope’s third time at Dick Clar
k’s New Year’s Rockin’ Eve after group stages with BTS in 2017 and 2019 .J-Hope’s live.After returning from Times S
quare where he performed as the penultimate act alongside multiple other singers from around the world and spoke to
Ryan Seacrest, the host of the show, J-Hope was back at his hotel room and turned on a live broadcast to speak with
his fans.As they congratulated him on his sparkling performance, being the perfectionist that he is, J-Hope express
ed his sadness about being unable to use his voice to the fullest and talked about the slipping incident due to the
rainy weather during his rehearsal stage..J-Hope on other BTS members’ New Year wishes.As soon as it turned 12, and
the New Year began in South Korea, BTS members Jungkook, Jimin, and V took to the fan community platform Weverse to
share their wishes with the fans as well as discuss their plans for the coming year.Jungkook and V kept it brief, w
ishing for a successful and happy year ahead while Jimin wrote a big letter to his fans about the feelings he had s
eemingly bottled for a long time about his wishes to release new music soon for which he has been meeting up with c
omposers .Thousands of miles away, member J-Hope was his cheeky self as he unleashed all his love for his fellow BT
S members.He began commenting ‘love you’ on their posts with cute words and received laughs from them in return .Ji
n’s call to J-Hope.On being asked about BTS’ oldest member Jin who became the first from the group to enlist in the
military on December 13, J-Hope recalled how he was called by the member with a different phone number and almost m
issed it.J-Hope said, “Right before I was about to sleep on the 31st, I got a call from Jin and asked him how he wa
s doing to which he said J-Hope, pick up your phone.I told him I did not know this number, how would I know it was
his?”.The ‘Arson’ hitmaker spoke with a smile on his face about how he felt happy hearing Jin’s voice.It comforted
him and J-Hope mentioned how he remembered all those moments he spent with Jin.He assured the fans by saying that J
in seemed to be healthy and doing well in the military.So in place of Jin, he shared that he was well and asked the
fans to not worry.The lovely duo, BTS’ Jin and J-Hope nicknamed 2seok have always lightened the fans’ hearts with t
his interaction .https://twitter com/nightstar1201/status/1609428759454822401Jimin X Taeyang.BTS member Jimin and B
IGBANG member Taeyang are a collaboration nobody would have expected.However, in December 2022, it was reported by
industry officials that the two are working together on a release that will soon be revealed to the world.To this,
Taeyang’s then-agency, YG Entertainment, which he has since departed, replied by saying that they cannot confirm an
ything at the moment and asked fans to anticipate Taeyang’s activities..On January 1, Taeyang who is currently a fr
ee agent, not having signed with anyone for his solo activities as he continues to be with YG Entertainment for any
content related to BIGBANG, shared a rare update on his Instagram account.2 photos were shared with a caption of th
e hashtag #2023.In black and white, 2 people could be seen in the photos, their backs to the camera.While one could
easily be spotted as Taeyang himself, the other one was not tagged....Fans went into action and soon began investig
ating every detail of the photo.The other person seemed to be Jimin, as his hands, fluffy hair, and earrings seem t
o match the BTS member.Though no official announcement was made from either side, it seems that the singers have de
cided to give a green signal from their ends.Taeyang is rumoured to have been preparing for a January 2023 comeback
so it seems as though we can expect the reports any time now.Once confirmed, this could very well be the biggest re
lease of the year and one of the most hyped collaborations in K-pop history!.Meanwhile, BIGBANG’s G-Dragon also ann
ounced that he is working on an album and hopes to release it in 2023 .Stay updated with the latest Hallyu news on:
Instagram, YouTube, Twitter, Facebook and Snapchat.ALSO READ: BTS’ J-Hope, TXT wow at Dick Clark’s New Year’s Rocki
n’ Eve: 5 highlights from their stages'
```

In [72]:

```
df2['New Content'] = summary
df2.head()
```

Out[72]:

	Original Content	New Content
0	After reaching his hotel in the city, RM revea...	As he sat at a roadside open-air restaurant, R...
1	RM aka Kim Namjoon was the first member to joi...	RM aka Kim Namjoon was the first member to joi...
2	Billie Eilish's concert was held in Seoul, Sou...	They really enjoyed the concert as the audienc...
3	BTS ARMY you all would be missing the members ...	The boys are going to complete their projects ...
4	BTS member Kim Seokjin aka Jin has the capacit...	Some days back, we saw the Sea of Jin concept ...

In [73]:

```
df2[df2['New Content'] == '']
```

Out[73]:

	Original Content	New Content
229	BTS' fame is such that there is no doubt that ...	

This is the part, where I faced a major problem, as it was unexpected. For data in some rows like the one above (229th row) the vectors generated were very long. Due to this, I was continously getting error as "Power Iteration failed". At last, I increased maximum interations to 1 Lakh (this is obviously veryyyyy high). But still, error persisted.

Even, I rechecked everything from Data Pre-processing till the point, but I found no more pre-processing of data could be done to reduce the vectors. Eventually, I had to add try-except block in the function, so that if summary for a particular data cannot be generated, just replace it with empty string.

Here, after thinking of each and every possibility, I decided to drop the row as its just 1 row.

In [74]:

```
df2.drop(df2[df2['New Content'] == ''].index[0], axis=0, inplace=True)
```

In [75]:

```
df2[df2['New Content'] == '']
```

Out[75]:

	Original Content	New Content

In [76]:

```
df2.isna().sum()
```

Out[76]:

```
Original Content    0
New Content         0
dtype: int64
```

In [77]:

```
df2.shape
```

Out[77]:

```
(805, 2)
```

In [78]:

```
df2[df2['New Content'].isna()]
```

Out[78]:

	Original Content	New Content

5.1 Saving my work till here, into a CSV file

In [79]:

```
try:
    df2.to_csv('./Summarized News.csv', index=False)
    print("Your file has been created !!!")
except:
    print("Failed to create fail. Might be due to permission related errors.")
```

```
Your file has been created !!!
```

6. Finally Working On Summarized File

In [80]:

```
df3 = pd.read_csv('./Summarized News.csv')
df3.head()
```

Out[80]:

	Original Content	New Content
0	After reaching his hotel in the city, RM revea...	As he sat at a roadside open-air restaurant, R...
1	RM aka Kim Namjoon was the first member to joi...	RM aka Kim Namjoon was the first member to joi...
2	Billie Eilish's concert was held in Seoul, Sou...	They really enjoyed the concert as the audienc...
3	BTS ARMY you all would be missing the members ...	The boys are going to complete their projects ...
4	BTS member Kim Seokjin aka Jin has the capacit...	Some days back, we saw the Sea of Jin concept ...

In [81]:

```
df3.isna().sum()
```

Out[81]:

```
Original Content    0
New Content         0
dtype: int64
```

6.1 Discovering the lines that were removed

In [82]:

```
content = df3['Original Content'].iloc[370]
content
```

Out[82]:

```
"On January 23, JYP Entertainment released new concept images for Stray Kids’ second fan meeting ‘SKZ’s Chocolate F
actory’.Each member looks amazing in the soft pastel clothes and pretty accessories as they work in the chocolate f
actory.The fan meeting will be held on February 12 and 13.On February 13th, an offline fan meeting and an online pa
id live broadcast on the Beyond Live platform will be held at the same time, and precious memories will be made wit
h domestic and foreign fans.Tickets for offline performances were pre-purchased for the fan club from 8:00 pm to 1
1:59 pm on January 17th for members of the 2nd period of the official fan club STAY, and all seats were sold out at
the same time as they opened.Thanks to such enthusiastic support, JYP Entertainment opened additional seats availab
le for viewing at 8 pm on January 19th, and this also sold out quickly, realizing the power of Stray Kids' tickets.
This fan meeting is the first in about a year since the first official fan meeting on February 20, 2021, and Stray
Kids is expected to meet fans around the world for a long time and will spend sweet time together that transcends t
ime and space.Stray Kids, who broke their best scores in various indicators last year and achieved 'Career High', w
ere honored with 4 crowns, including the first grand prize in their debut, at the awards ceremony to close the yea
r.At the '2021 The Fact Music Awards', '2021 Asian Artist Awards', and '2021 Mnet Asian Music Awards' respectively,
they won the Artist of the Year Award, Performance Award of the Year, and Worldwide Fans' Choice Top Ten respective
ly.At the Golden Disc Awards, they added a trophy in the album category to show off their 'K-Pop 4th generation rep
resentative group' ALSO READ: The official trailer of Netflix's  of us are Dead' surpasses THIS number of views on
YouTubeThe magnificent celebration of K-world culminates with The HallyuTalk Awards, watch here What do you think o
f the concept photos? Let us know in the comments below."
```

In [83]:

```
summary = df3['New Content'].iloc[370]
summary
```

Out[83]:

```
"Tickets for offline performances were pre-purchased for the fan club from 8:00 pm to 11:59 pm on January 17th for
members of the 2nd period of the official fan club STAY, and all seats were sold out at the same time as they opene
d. Thanks to such enthusiastic support, JYP Entertainment opened additional seats available for viewing at 8 pm on
January 19th, and this also sold out quickly, realizing the power of Stray Kids' tickets. At the '2021 The Fact Mus
ic Awards', '2021 Asian Artist Awards', and '2021 Mnet Asian Music Awards' respectively, they won the Artist of the
Year Award, Performance Award of the Year, and Worldwide Fans' Choice Top Ten respectively. "
```

Showing the total number of lines in the actual content

In [84]:

```
content_split = []
for i in content.split("."):
    i = i.strip()
    if len(i) != 0 and i != " ":
        content_split.append(i)
len(content_split)
```

Out[84]:

```
10
```

Showing the total number of lines in the summary generated for content

In [85]:

```
summary_split = []
for i in summary.split("."):
    i = i.strip()
    if len(i) != 0 and i != " ":
        summary_split.append(i)
len(summary_split)
```

Out[85]:

```
3
```

Showing the number of lines removed

In [86]:

```
removed_lines = []
for i in content_split:
    if i not in summary_split:
        removed_lines.append(i)
len(removed_lines)
```

Out[86]:

```
7
```

6.2 Combining all above process into a function

In [87]:

```
def removed_lines(content, summary):
    content_split = []
    for i in content.split("."):
        i = i.strip()
        if len(i) != 0 and i != " ":
            content_split.append(i)

    summary_split = []
    for i in summary.split("."):
        i = i.strip()
        if len(i) != 0 and i != " ":
            summary_split.append(i)

    removed_line = []
    for i in content_split:
        if i not in summary_split:
            removed_line.append(i)

    removedLines = ". ".join(removed_line)
    return removedLines
```

In [88]:

```
removed_line = df3.apply(lambda x: removed_lines(x['Original Content'], x['New Content']), axis=1)
```

In [89]:

```
removed_line
```

Out[89]:

```
0      After reaching his hotel in the city, RM revea...
1      The group released their debut single album 2 ...
2      Billie Eilish's concert was held in Seoul, Sou...
3      BTS ARMY you all would be missing the members ...
4      BTS member Kim Seokjin aka Jin has the capacit...
                             ...
800    BTS has conquered the world with their group r...
801    Since it was released, the meaningful song rec...
802    After checking out his live video, BTS' Jungko...
803    BTS' eldest member Jin has shared pictures and...
804    After a lot of teasing, Benny Blanco’s collabo...
Length: 805, dtype: object
```

In [90]:

```
len(removed_line)
```

Out[90]:

```
805
```

In [91]:

```
df3.shape
```

Out[91]:

```
(805, 2)
```

6.3 Creating a seperated column for storing the removed lines

In [92]:

```
df3['Removed Lines'] = removed_line
df3.head()
```

Out[92]:

	Original Content	New Content	Removed Lines
0	After reaching his hotel in the city, RM revea...	As he sat at a roadside open-air restaurant, R...	After reaching his hotel in the city, RM revea...
1	RM aka Kim Namjoon was the first member to joi...	RM aka Kim Namjoon was the first member to joi...	The group released their debut single album 2 ...
2	Billie Eilish's concert was held in Seoul, Sou...	They really enjoyed the concert as the audienc...	Billie Eilish's concert was held in Seoul, Sou...
3	BTS ARMY you all would be missing the members ...	The boys are going to complete their projects ...	BTS ARMY you all would be missing the members ...
4	BTS member Kim Seokjin aka Jin has the capacit...	Some days back, we saw the Sea of Jin concept ...	BTS member Kim Seokjin aka Jin has the capacit...

In [93]:

```
len(df3['Original Content'].iloc[0].split("."))-len(df3['New Content'].iloc[0].split("."))
```

Out[93]:

```
17
```

In [94]:

```
len(df3['Removed Lines'].iloc[0].split("."))
```

Out[94]:

```
17
```

In [95]:

```
from sentence_transformers import SentenceTransformer, util
```

In [96]:

```
model = SentenceTransformer('all-MiniLM-L6-v2')
```

In [97]:

```
originalC = df3['Original Content'].iloc[1]
newC = df3['New Content'].iloc[1]
```

In [98]:

```
en_1 = model.encode(originalC)
en_2 = model.encode(newC)
print(type(en_1))
```

```
<class 'numpy.ndarray'>
```

6.3 Using Cosine Similarity to estimate the similarity of summary and the original content

In [99]:

```
result = util.cos_sim(en_1, en_2)

print(float(result))
print(type(result))
```

```
0.7904312014579773
<class 'torch.Tensor'>
```

In [100]:

```
result_float = result.item()
print(result_float)
print(type(result_float))
```

```
0.7904312014579773
<class 'float'>
```

In [101]:

```
def metrics(originalC, newC):
    en_1 = model.encode(originalC)
    en_2 = model.encode(newC)
    result = util.cos_sim(en_1, en_2)
    # print(float(result))
    return float(result)
```

In [102]:

```
cosine = df3.apply(lambda x: metrics(x['Original Content'], x['New Content']), axis=1)
```

In [103]:

```
cosine.shape
```

Out[103]:

```
(805,)
```

In [104]:

```
df3['Cosine Similarity'] = cosine
```

In [105]:

```
df3.head()
```

Out[105]:

	Original Content	New Content	Removed Lines	Cosine Similarity
0	After reaching his hotel in the city, RM revea...	As he sat at a roadside open-air restaurant, R...	After reaching his hotel in the city, RM revea...	0.796752
1	RM aka Kim Namjoon was the first member to joi...	RM aka Kim Namjoon was the first member to joi...	The group released their debut single album 2 ...	0.790431
2	Billie Eilish's concert was held in Seoul, Sou...	They really enjoyed the concert as the audienc...	Billie Eilish's concert was held in Seoul, Sou...	0.840863
3	BTS ARMY you all would be missing the members ...	The boys are going to complete their projects ...	BTS ARMY you all would be missing the members ...	0.709553
4	BTS member Kim Seokjin aka Jin has the capacit...	Some days back, we saw the Sea of Jin concept ...	BTS member Kim Seokjin aka Jin has the capacit...	0.719757

In [106]:

```
min(cosine)
```

Out[106]:

```
0.36477231979370117
```

In [107]:

```
max(cosine)
```

Out[107]:

```
0.9571455717086792
```

In [108]:

```
np.mean(cosine)
```

Out[108]:

```
0.7548249448678508
```

In [109]:

```
np.median(cosine)
```

Out[109]:

```
0.7781685590744019
```

6.4 Using Sementic Similarity to estimate how sementically our generated summary is close to original content

In [110]:

```
nlp1 = spacy.load("en_core_web_lg")
```

In [111]:

```
doc1 = nlp(originalC)
doc2 = nlp(newC)
print(doc1.similarity(doc2))
```

```
0.9229698009279503
```

In [112]:

```
def nlptextsimilarity(originalC, newC):
    doc1 = nlp(originalC)
    doc2 = nlp(newC)
    return doc1.similarity(doc2)
```

In [113]:

```
nlptextsimi = df3.apply(lambda x: nlptextsimilarity(x['Original Content'], x['New Content']), axis=1)
```

In [114]:

```
df3['NLP Text Similarity'] = nlptextsimi
```

In [115]:

```
df3.head()
```

Out[115]:

	Original Content	New Content	Removed Lines	Cosine Similarity	NLP Text Similarity
0	After reaching his hotel in the city, RM revea...	As he sat at a roadside open-air restaurant, R...	After reaching his hotel in the city, RM revea...	0.796752	0.947731
1	RM aka Kim Namjoon was the first member to joi...	RM aka Kim Namjoon was the first member to joi...	The group released their debut single album 2 ...	0.790431	0.922970
2	Billie Eilish's concert was held in Seoul, Sou...	They really enjoyed the concert as the audienc...	Billie Eilish's concert was held in Seoul, Sou...	0.840863	0.938945
3	BTS ARMY you all would be missing the members ...	The boys are going to complete their projects ...	BTS ARMY you all would be missing the members ...	0.709553	0.923807
4	BTS member Kim Seokjin aka Jin has the capacit...	Some days back, we saw the Sea of Jin concept ...	BTS member Kim Seokjin aka Jin has the capacit...	0.719757	0.909008

In [116]:

```
min(nlptextsimi)
```

Out[116]:

```
0.5577889218703898
```

In [117]:

```
max(nlptextsimi)
```

Out[117]:

```
0.9906746332886787
```

In [118]:

```
np.mean(nlptextsimi)
```

Out[118]:

```
0.923961025354796
```

In [119]:

```
np.median(nlptextsimi)
```

Out[119]:

```
0.9330589202860173
```

6.5 As there was quite a difference in the median of metrics, I decided to provide Harmonic Mean of both the similaritiy

In [120]:

```
def harmonicMean(x, y):
    return (2*x*y)/(x+y)
```

In [121]:

```
harmean = df3.apply(lambda x: harmonicMean(x['Cosine Similarity'], x['NLP Text Similarity']), axis=1)
```

In [122]:

```
df3['Harmonic Mean'] = harmean
```

In [123]:

```
df3.head()
```

Out[123]:

	Original Content	New Content	Removed Lines	Cosine Similarity	NLP Text Similarity	Harmonic Mean
0	After reaching his hotel in the city, RM revea...	As he sat at a roadside open-air restaurant, R...	After reaching his hotel in the city, RM revea...	0.796752	0.947731	0.865708
1	RM aka Kim Namjoon was the first member to joi...	RM aka Kim Namjoon was the first member to joi...	The group released their debut single album 2 ...	0.790431	0.922970	0.851574
2	Billie Eilish's concert was held in Seoul, Sou...	They really enjoyed the concert as the audienc...	Billie Eilish's concert was held in Seoul, Sou...	0.840863	0.938945	0.887202
3	BTS ARMY you all would be missing the members ...	The boys are going to complete their projects ...	BTS ARMY you all would be missing the members ...	0.709553	0.923807	0.802628
4	BTS member Kim Seokjin aka Jin has the capacit...	Some days back, we saw the Sea of Jin concept ...	BTS member Kim Seokjin aka Jin has the capacit...	0.719757	0.909008	0.803388

7. Storing The Result In a Seperate Dataframe

In [124]:

```
df3.to_csv('./Final Result.csv')
```

In [125]:

```
min(harmean)
```

Out[125]:

```
0.49884301558564753
```

In [126]:

```
max(harmean)
```

Out[126]:

```
0.9623399165870538
```

In [127]:

```
np.mean(harmean)
```

Out[127]:

```
0.8266253043667523
```

In [128]:

```
np.median(harmean)
```

Out[128]:

```
0.8415012837304162
```