

# Look At The Barrage From Another Angle

— — Visualization of Barrage Data  
With Python & Text Sentiment Analysis

2030223131 Zhao LUO (罗钊)  
2030223118 Sunqingnian WANG (王孙卿念)  
2030223158 Shuang PENG (彭爽)



# CONTENTS

01

## Data Crawling and Data Cleaning

Page 4 to Page 9

02

## Basic Visual Analysis of Barrage Data

Page 10 to Page 13

03

## Data Word Cloud Display

Page 14 to Page 23

04

## Text Sentiment Analysis & Resampling

Page 24 to Page 31

05

## Summary and Outlook

Page 32



# INTRODUCTION

The exciting offer season 2 program mainly tells the story of eight law school students who embrace the lawyer's dream and experience a one-month internship in law firms under the teaching and assessment of four lawyers, and compete for the positive offer of two law firms.

Analyze these screen data. Then analyze some of the user's situation, as well as the user's attitude to the program.



» 01

Data Crawler, Data Cleaning

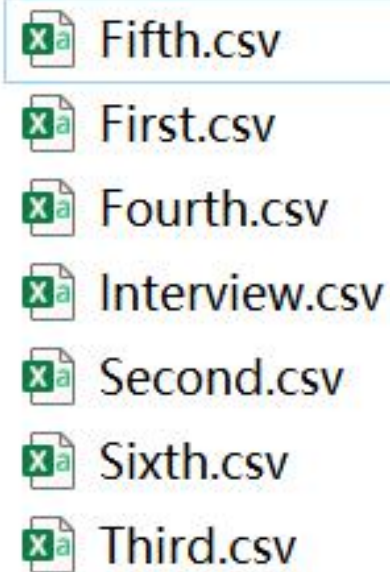


## DATA CRAWLER

Total data collected by reptiles

A total of **261,724** pieces of data

Seven csv documents were generated.



- Fifth.csv
- First.csv
- Fourth.csv
- Interview.csv
- Second.csv
- Sixth.csv
- Third.csv

# USE target\_id & vid Get Video Barrage

A fixed url address + video cid + .xml



Find the video cid, replace the url to crawl all the barrage



Right-click the webpage, open the source code of the webpage, search for "cid": you can find them

The screenshot shows a web browser's developer console with the 'Search' tab active. The search term 'target\_id' is entered in the search bar. Two matching lines are found in two files:

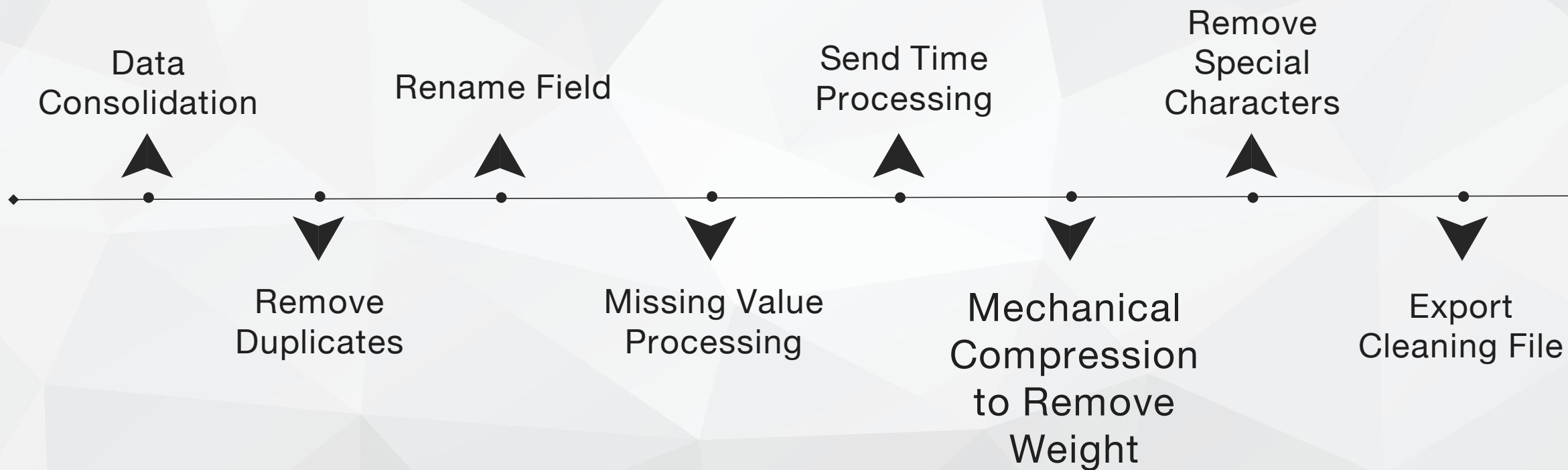
- File 1: `danmu?type=json&callback=jQuery191038368158811100694_1608740267179&target_id=6130942571%26vid%3D10...`  
Line 1: `...158811100694_1608740267179({"err_code":0,"err_msg":"ok","peroid":"30","target_id":"6130942571","count":17`
- File 2: `barragefloat.162938.js?_ts=20171208 — vm.gtimg.cn/tencentvideo/txp/js/plugins/barragefloat.162938.js?_ts=20171208`  
Line 1: `...r a={};function i(t){if(a[t])return a[t].exports;var e=a[t]={i:t,l:!1,exports:{}};return n[t].call(e.exports,e,e.exports,i),e.l=!0`

At the bottom, a status bar indicates: 'Search finished. Found 2 matching lines in 2 files.'



# PART ONE

Zhao LUO (罗钊)



# Mechanical Compression Dewatering Function

## Single Words

str = '喜欢丁辉呀 呀 呀'

0 1 2 3 4 5 6  
0 1 2 j j+1 j+2 j+3

↓ ↓  
equality equality

output: '喜欢丁辉呀'

## Double words

str = '超级喜欢 喜 欢呀'

0 1 2 3 4 5 6  
0 1 2 j j+1 j+2 j+3

↓ ↓  
equality equality

output: '超级喜欢呀'

Compared to above, three-words has the same principle.

The data in the sentence removes repetition.



# PART ONE

Zhao LUO (罗钊)



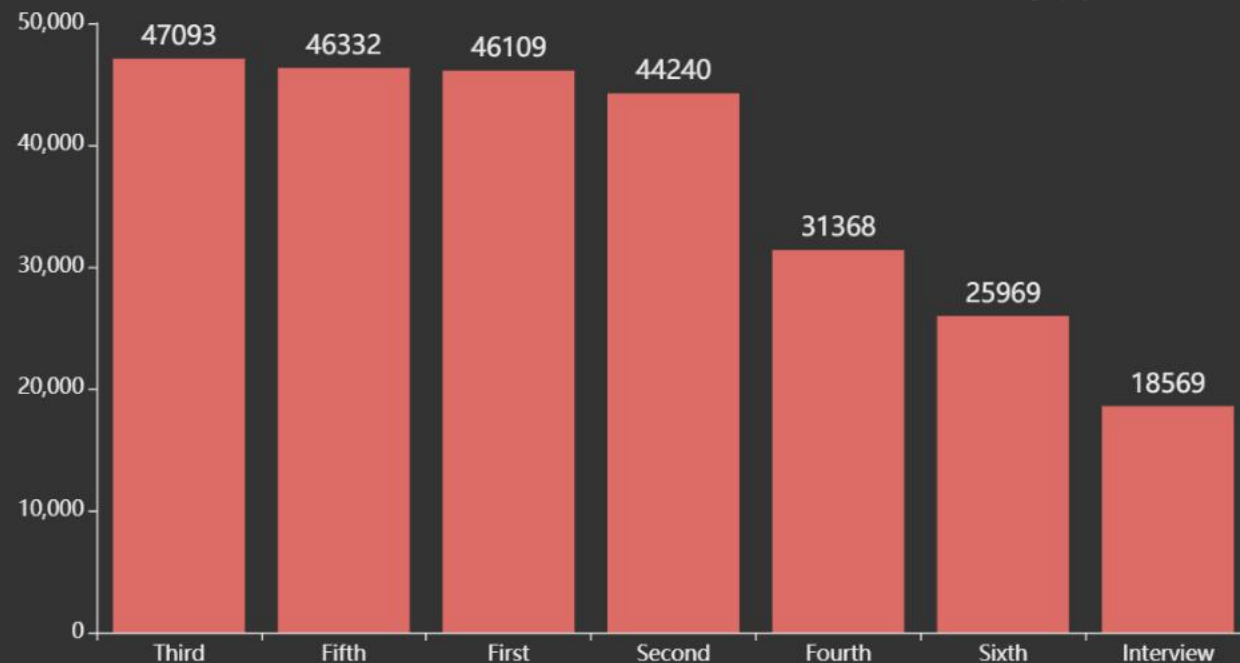
clean.xlsx

Generate clean.xlsx document after cleaning

A total of **259,691** pieces of data

## Number of Rounds Per Period

Source of Data: Tencent Video Cartography: COMM7330 Python Group 11



**Histogram of Data Distribution After Cleaning**



## » 02

# Basic Visual Analysis of Barrage Data



## PART TWO

Zhao LUO (罗钊)

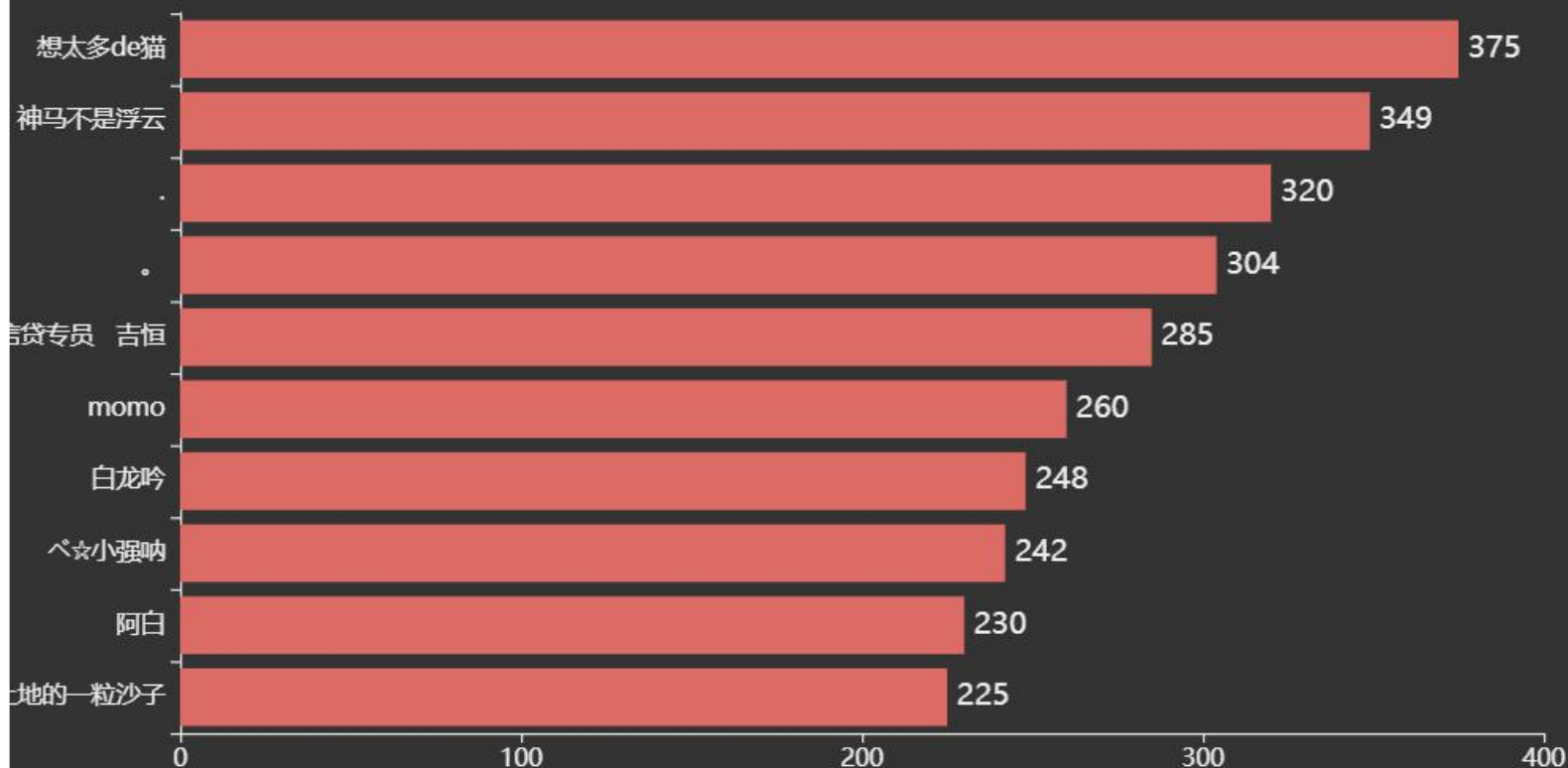
## Total Word Cloud

Ding Hui's super popularity  
The audience's love for this variety

# WHO IS THE BARRAGE TRANSMITTER?

## Number of Rounds Sent TOP10

Source of Data: Tencent Video Cartography: COMM7330 Python Group 11

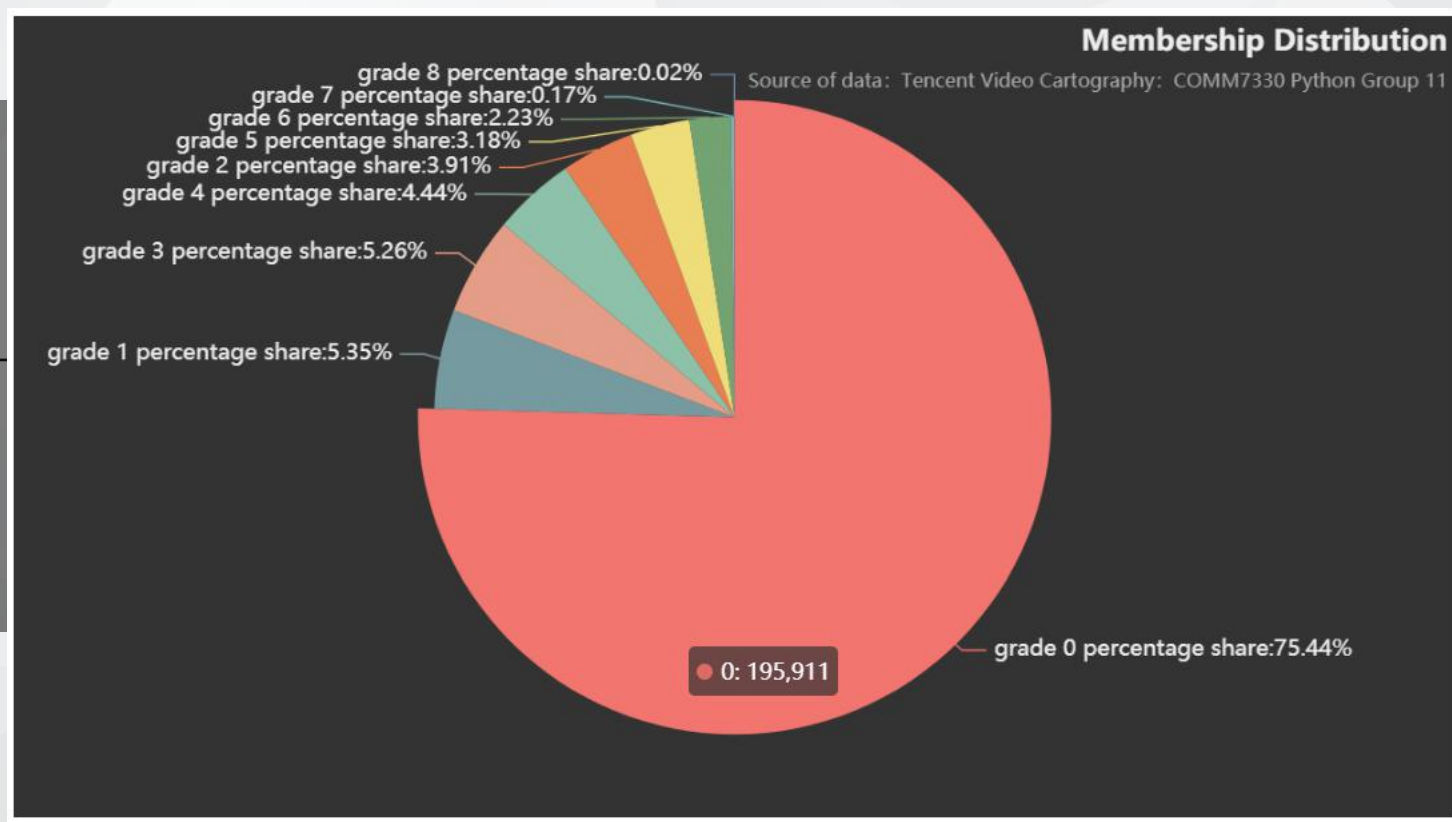


## TOP 10

ID of the users with  
the most screens is  
"想太多de猫".

## Member Level Distribution Pie Chart

User stickiness is strong





03

## Data Word Cloud Display

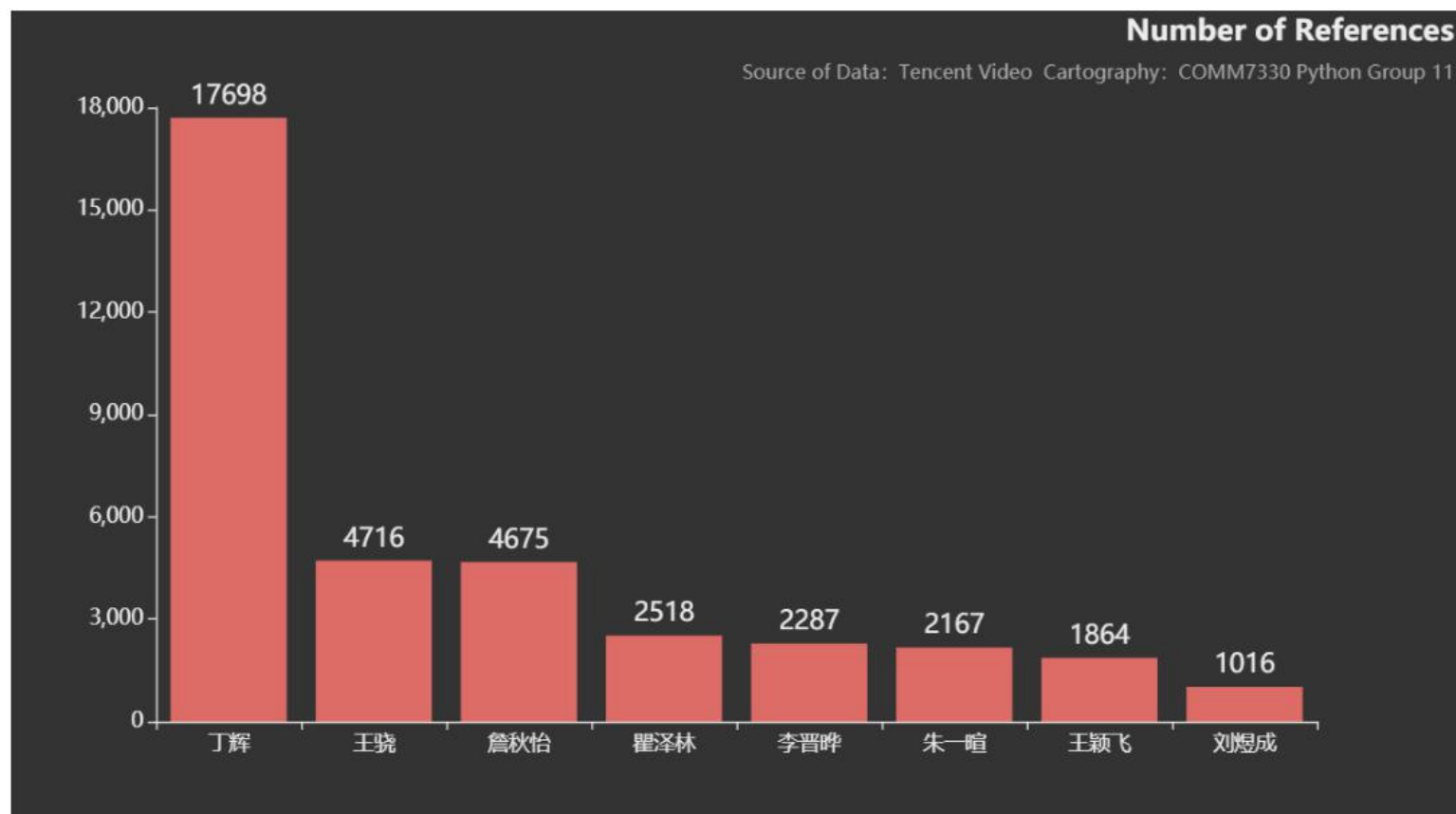




## PART THREE

Shuang PENG (彭爽)

# Number of intern role mentions



Ding Hui  
Wang Xiao  
Liu Yucheng

## Shuang PENG (彭爽)

Come on  
Like  
Look Good  
Support



# Barrage Word Cloud of Hui DING

## PART THREE

Shuang PENG (彭爽)

Beautiful  
Liu Yifei  
Good-Looking



Barrage Word Cloud of Qiuyi ZHAN

## Shuang PENG (彭爽)

The Top University Stanford  
A Two-Sided Situation  
Good  
Versailles



# Barrage Word Cloud of Xiao WANG

## Shuang PENG (彭爽)

Cute  
Hate



# Barrage Word Cloud of Yixuan ZHU



## Shuang PENG (彭爽)

# Barrage Word Cloud of Zelin QV



## Shuang PENG (彭爽)

Handsome  
He Yunchen



# Barrage Word Cloud of Jinye LI

## PART THREE

Shuang PENG (彭爽)

Good-Looking  
Beautiful



# Barrage Word Cloud of Yingfei WANG

## Shuang PENG (彭爽)

# Barrage Word Cloud of Yucheng LIU



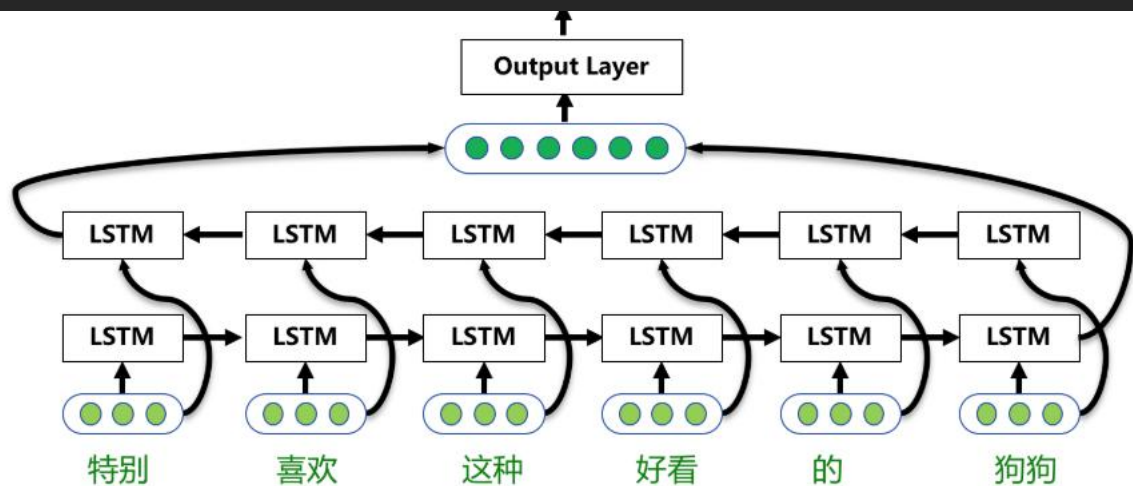
# » 04 Text Sentiment Analysis & Resampling



## PART FOUR

Sunqingnian WANG (王孙卿念)

```
import paddlehub as hub  
senta = hub.Module(name="senta_bilstm")
```



Baidu Senta model, bi-LSTM  
(bidirectional long short term memory)

# Text Sentiment Analysis Model



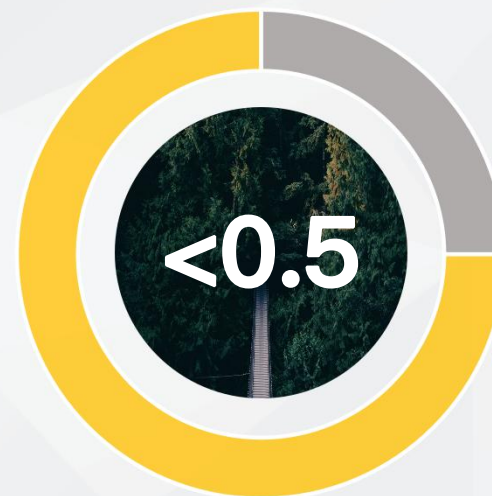
# JUDGMENT STANDRAD



POSITIVE



NEUTRAL



NEGATIVE





## PART FOUR

Sunqingnian WANG (王孙卿念)

Transmitting Membership Barrage Emotional  
Time Grade Click Score

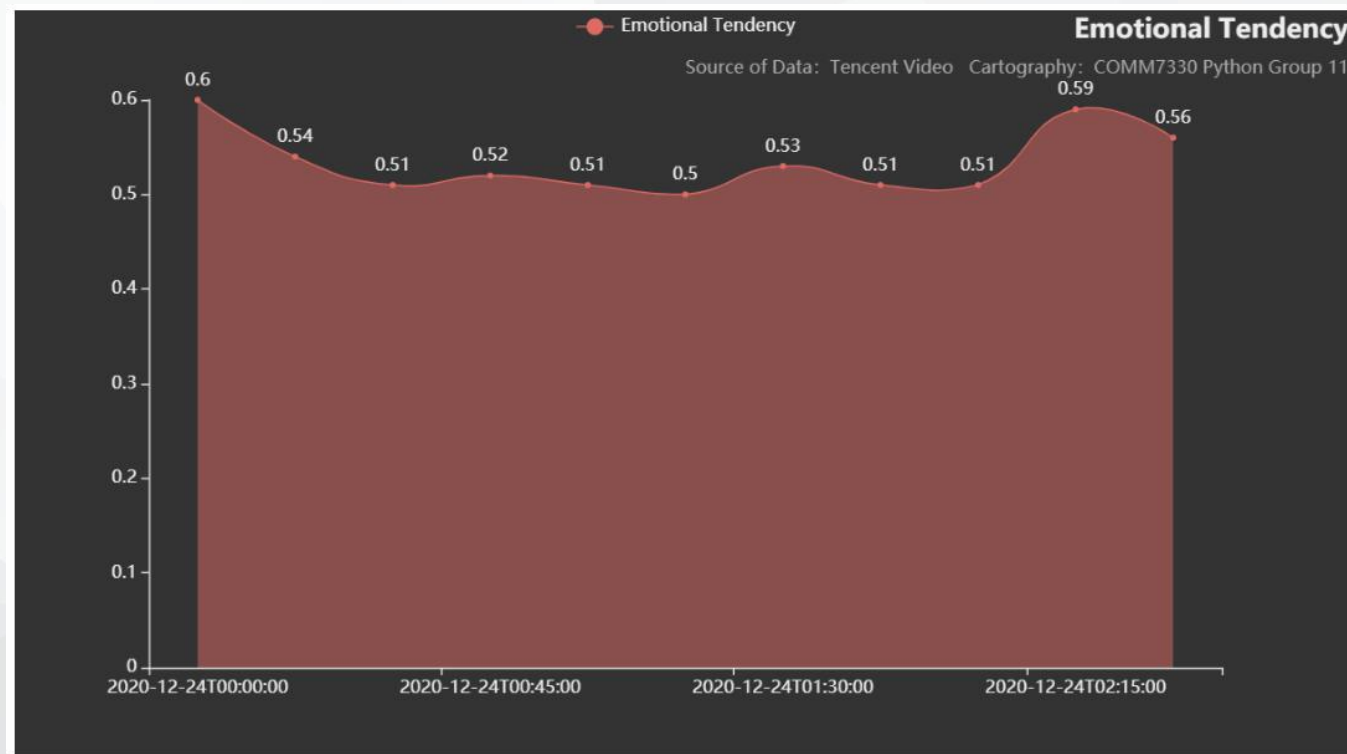
	发送时间	会员等级	弹幕点赞	情感分值
0	2020-12-24 00:00:00	0.760598	5.36026	0.597022
1	2020-12-24 00:15:00	0.755379	7.26792	0.539171
2	2020-12-24 00:30:00	0.751851	6.81649	0.512125
3	2020-12-24 00:45:00	0.806384	6.3929	0.520323
4	2020-12-24 01:00:00	0.786389	7.26516	0.510811
5	2020-12-24 01:15:00	0.798296	6.7297	0.504133
6	2020-12-24 01:30:00	0.770896	7.53894	0.530489
7	2020-12-24 01:45:00	0.82215	8.24078	0.508483
8	2020-12-24 02:00:00	0.725625	7.96195	0.51156
9	2020-12-24 02:15:00	0.754767	7.89367	0.585757
10	2020-12-24 02:30:00	1.04552	10.4143	0.562774

## Emotional Visualization

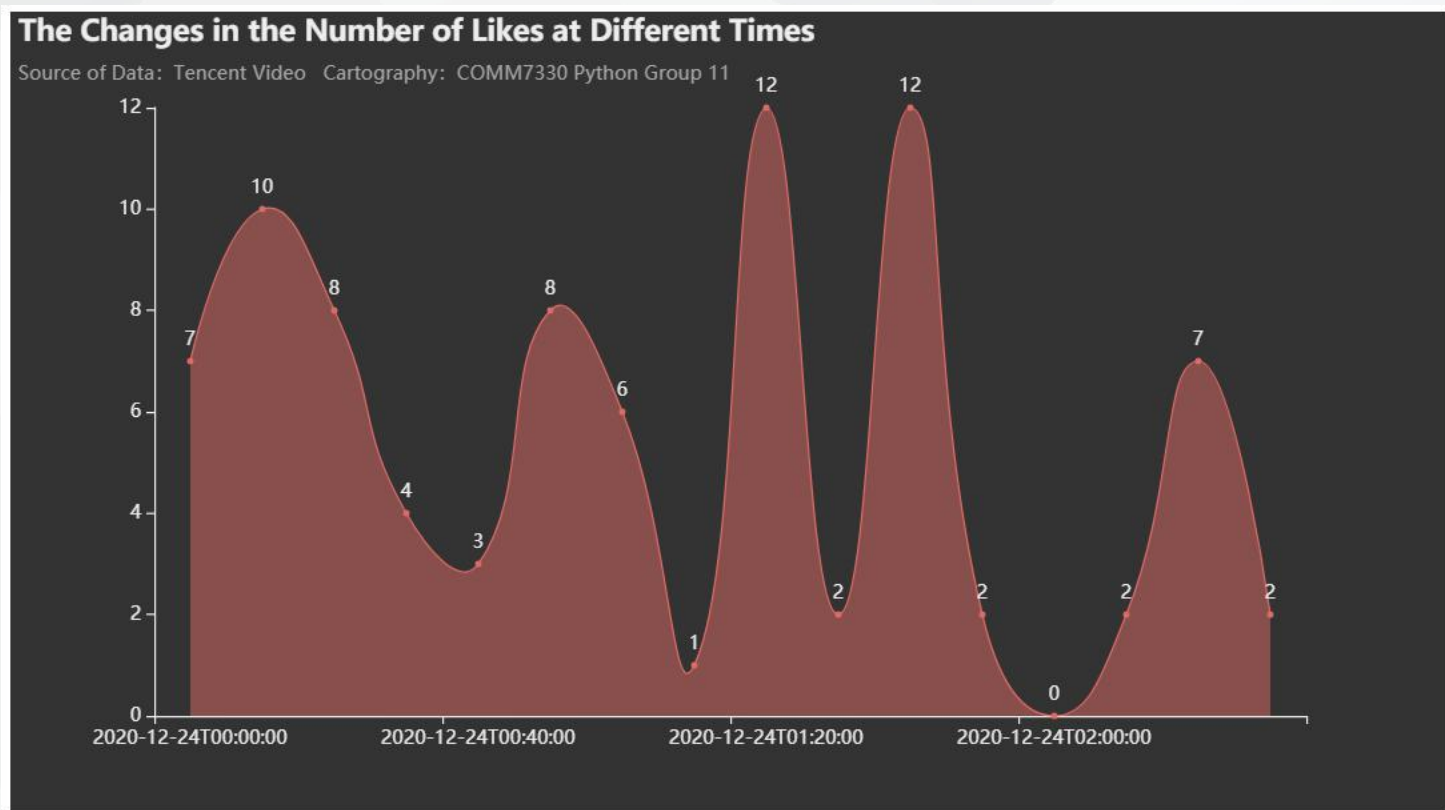
The emotional score is stable above 0.5,  
and the program is positive and controversial.

# Relationship Between Sending Time and Emotional Score

Scenes with a high degree of discussion appear regularly, keeping the audience's continuous watching enthusiasm.



# Changes in Likes at Different Sending Times



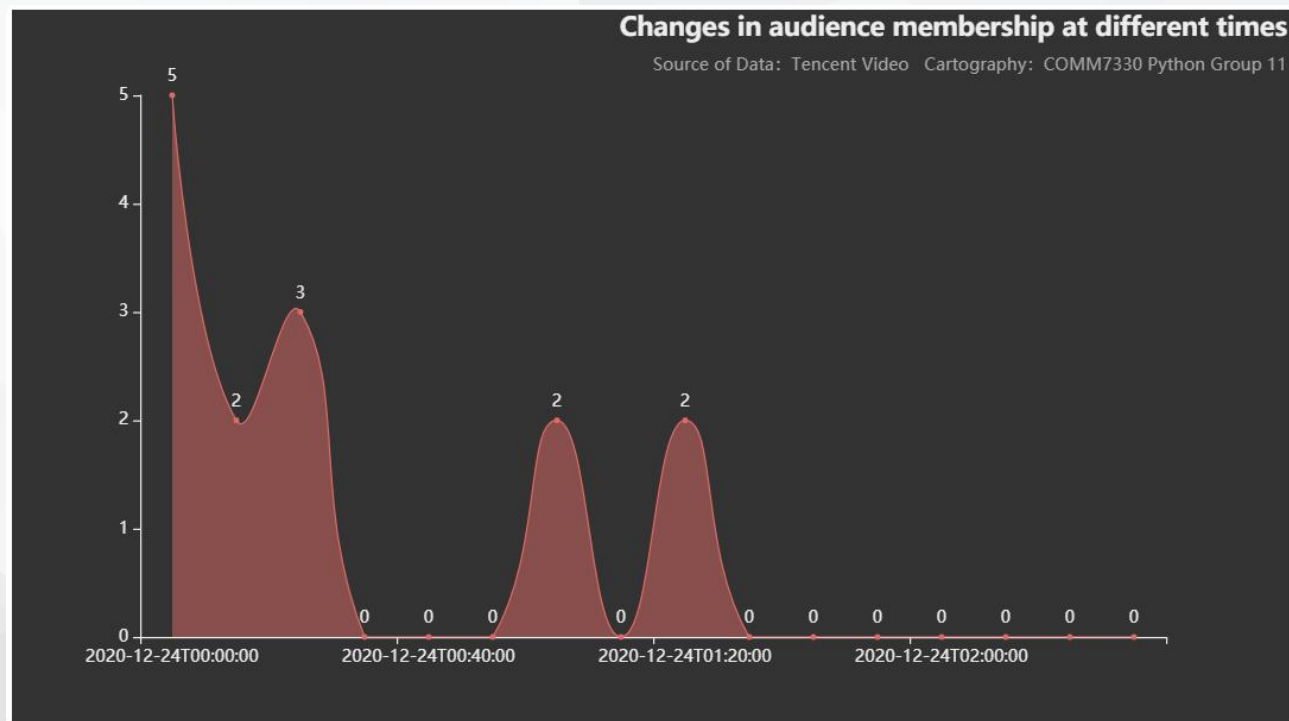
---

There needs to be words or scenes that stimulate the audience's cognition.

# Membership Level Changes at Different Times

---

Non-members account for a large proportion, and viewers watch programs not because of "member stickiness".



## PART FOUR

Sunqingnian WANG (王孙卿念)

Transmitting Membership Barrage Emotional  
Time Grade Click Score

	发送时间	会员等级	弹幕点赞	情感分值
0	2020-12-24 00:00:00	0.728125	4.548580	0.625161
1	2020-12-24 00:10:00	0.808544	7.849704	0.589407
2	2020-12-24 00:20:00	0.812638	7.131091	0.588118
3	2020-12-24 00:30:00	0.753290	6.863483	0.578765
4	2020-12-24 00:40:00	0.785092	5.911031	0.549265
5	2020-12-24 00:50:00	0.787126	6.193114	0.559842
6	2020-12-24 01:00:00	0.829261	7.513455	0.521978
7	2020-12-24 01:10:00	0.775979	8.096261	0.556235
8	2020-12-24 01:20:00	0.832669	6.047447	0.500744
9	2020-12-24 01:30:00	0.820744	6.476690	0.594042
10	2020-12-24 01:40:00	0.936719	6.756869	0.493446
11	2020-12-24 01:50:00	0.897915	6.451347	0.547184
12	2020-12-24 02:00:00	0.686675	8.366624	0.492362
13	2020-12-24 02:10:00	0.680648	7.835821	0.591986
14	2020-12-24 02:20:00	0.918380	7.590689	0.602106



Re-sampling.xlsx

## Resampling

After resampling, generate and save the document with added emotional score column to Re-sampling.xlsx

# » 05 Summary and Outlook

1. The Future Development of Senta
2. Further Improvement of The Project

In the future work and life, when we are engaged in some work related to program evaluation or content hotspot extraction, we can apply the technology involved in this project to practice, speak based on data, and more accurately judge market demand and conduct commercialization analysis to maximize benefits in all aspects.





