Let’s Play a Game
Never Have You Ever.....

Never Have You Ever Had an 8-year Relationship

2015
grade 8

2018
grade 11

2021
university

Why Did I Broke Up with My Ex after 8 Years

--- a Pattern and Sentiment Analysis
Based on the Chat History

Presented By:
Cheng Ling Jun
3035772652

1. Problem

Why Using Data Science?

- Emotion Leads to One-sided Opinions, especially in a relationship
- Facts that could be misunderstood due to one's emotions and perspective
- Data science research is needed to find out the quantitative and objective "truth".

Expected Outcome?

- Pattern
  - Evolution of the chat frequency
  - Data insight on an monthly/weekly/daily basis
- Sentiment
  - the effect brought by positive and negative messages
  - Comparison of average sentiment value over the control group

1. Problem

Why Important?

- If you ever had an 8-year-old relationship……

2. Dataset

- Data Source:
  - WeChat History, 2021 May - 2024 April

localId	TalkerId	Type	SubType	IsSender	CreateTime	Status	StrContent	StrTime	Remark	NickName	Sender
22855	28	1	0	1	1620014346		你在干啥呀	##########	Clara Cheng	Clara Cheng	wxid
22856	28	1	0	0	1620015553		刚刚吃完饭	##########			
22857	28	1	0	0	1620015557		准备洗碗	##########			

- Tools

Memotrace
to conduct data scrapping
from WeChat

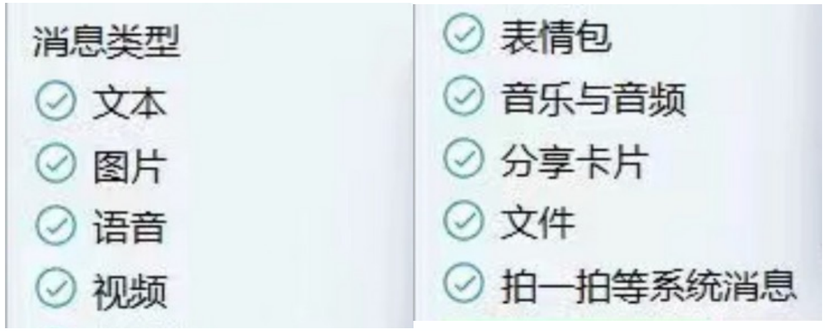
R Studio
for pattern analysis

Weiciyun
for sentiment analysis
(understanding Chinese Better)

2. Dataset

- Data Cleaning:
  - String Only
    - Only string content is saved.
    - Other types of data is not included, such as picture, video, emoji, etc.,
  - Only after 2021 May
    - *CreateTime* is a number defined by Memotrace, which represents the time

Data types in WeChat



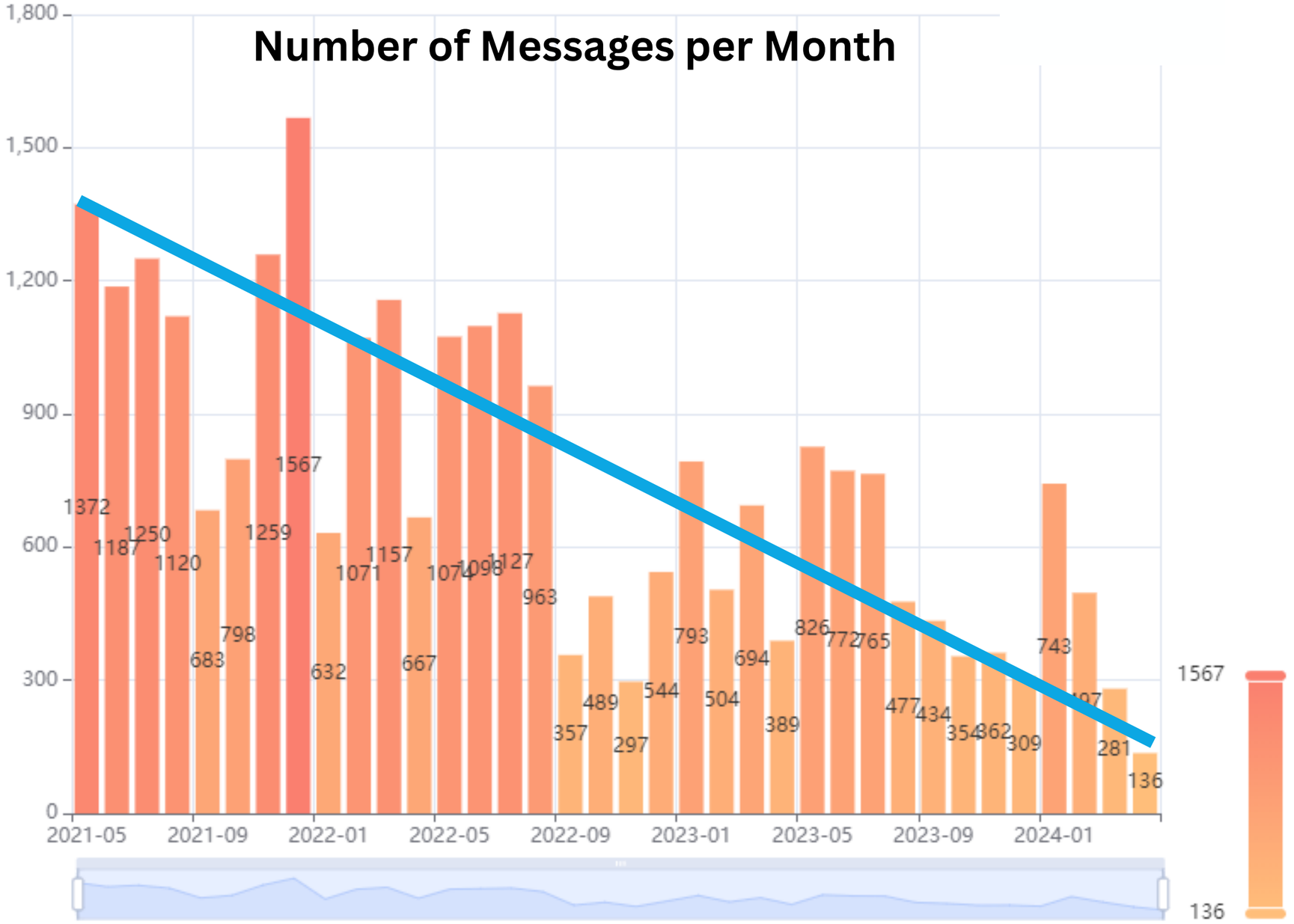
Filter

Goal	Condition
String Only	*Type* == 1
After 2021 May	*CreateTime* > 1620014346

localId	TalkerId	Type	SubType	IsSender	CreateTime	Status	StrContent	StrTime	Remark	NickName	Sender
22855	28	1	0	1	1620014346		你在干啥呀	##########	Clara Cheng	Clara Cheng	wxid
22856	28	1	0	0	1620015553		刚刚吃完饭	##########			
22857	28	1	0	0	1620015557		准备洗碗	##########			
22858	28	47	0	0	1620015558		<msg><emoj	##########			

⬆️Fitered out

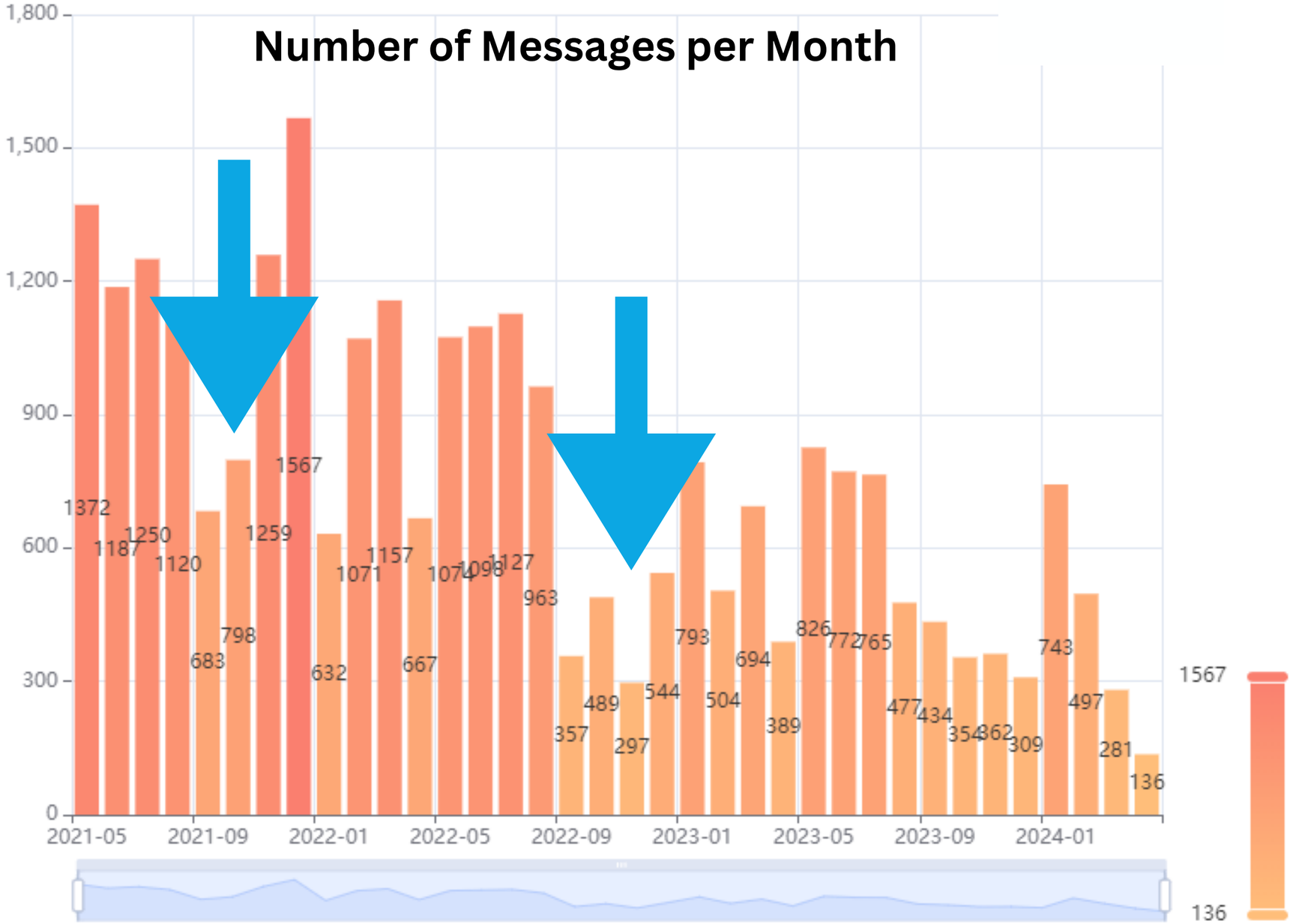
3. Pattern Analysis - Month



Observations:

- The number of messages decreased steadily by 90% in the past 3 years

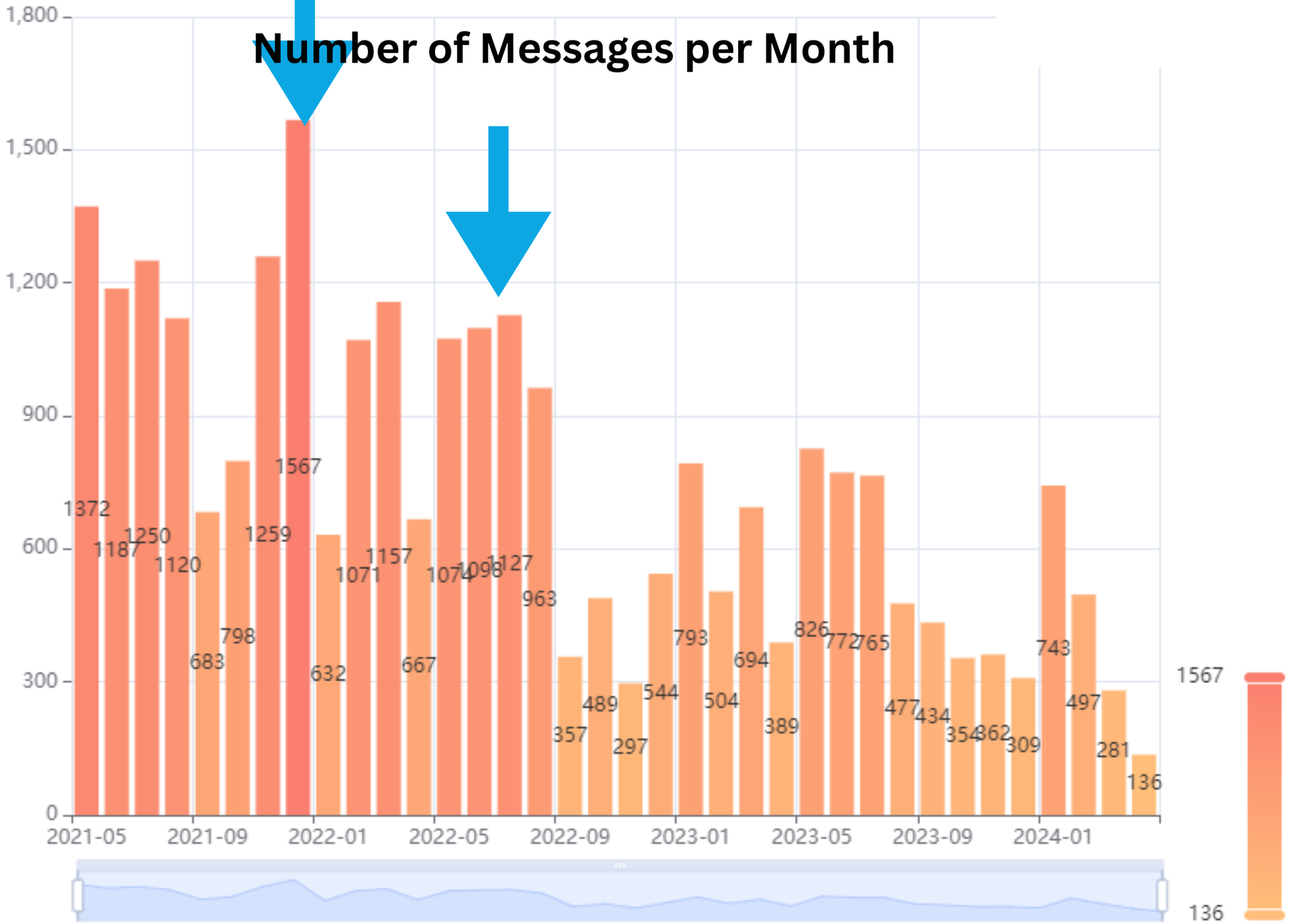
3. Pattern Analysis - Month



Observations:

- The number of messages decreased steadily by 90% in the past 3 years
- less chat at the beginning of a semester instead of “blood final”

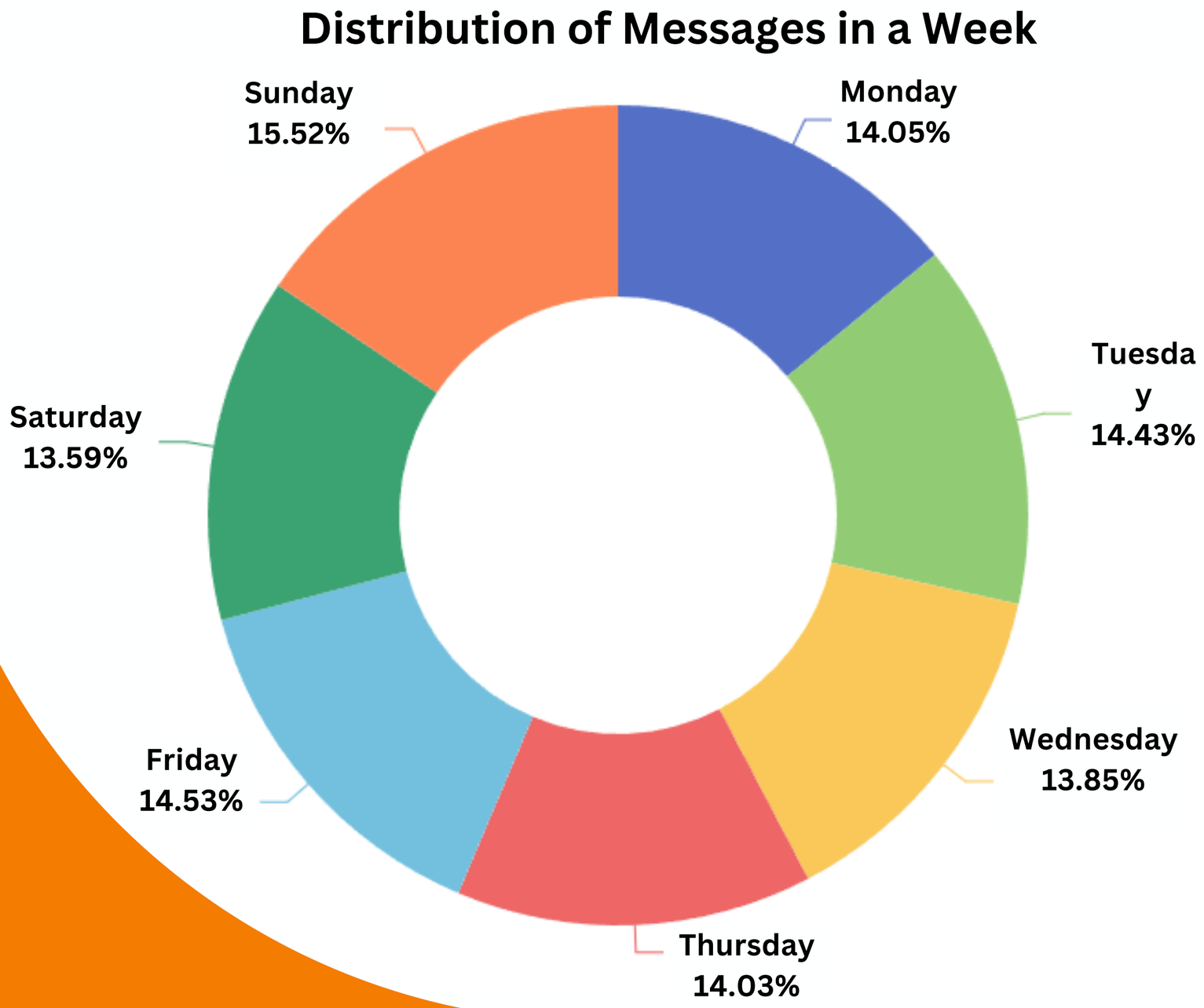
3. Pattern Analysis - Month



Observations:

- The number of messages decreased steadily by 90% in the past 3 years
- less chat at the beginning of a semester instead of “blood final”
- more chat during quarantine

3. Pattern Analysis - Week

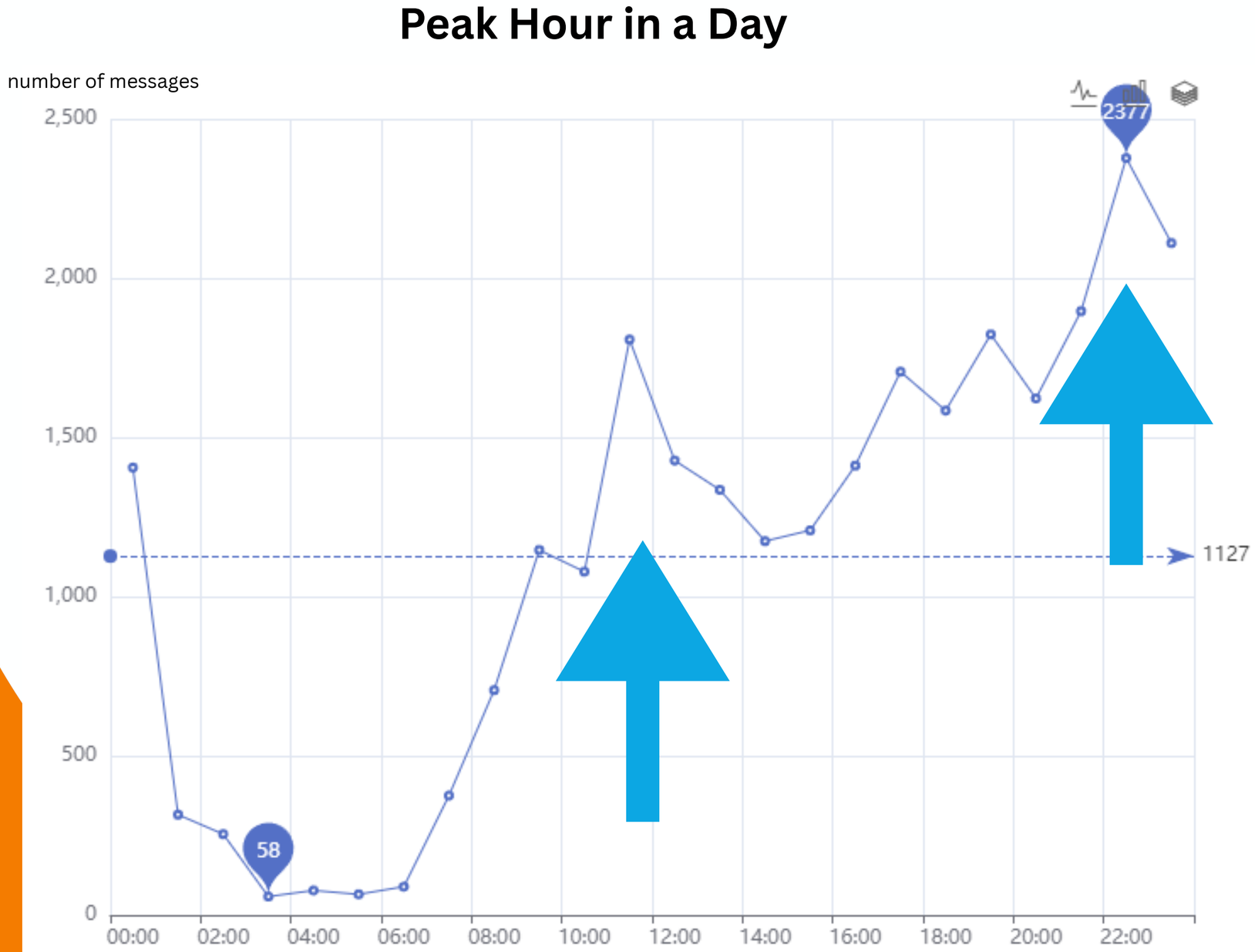


- My expectation:
  Peak during weekends

V.S.

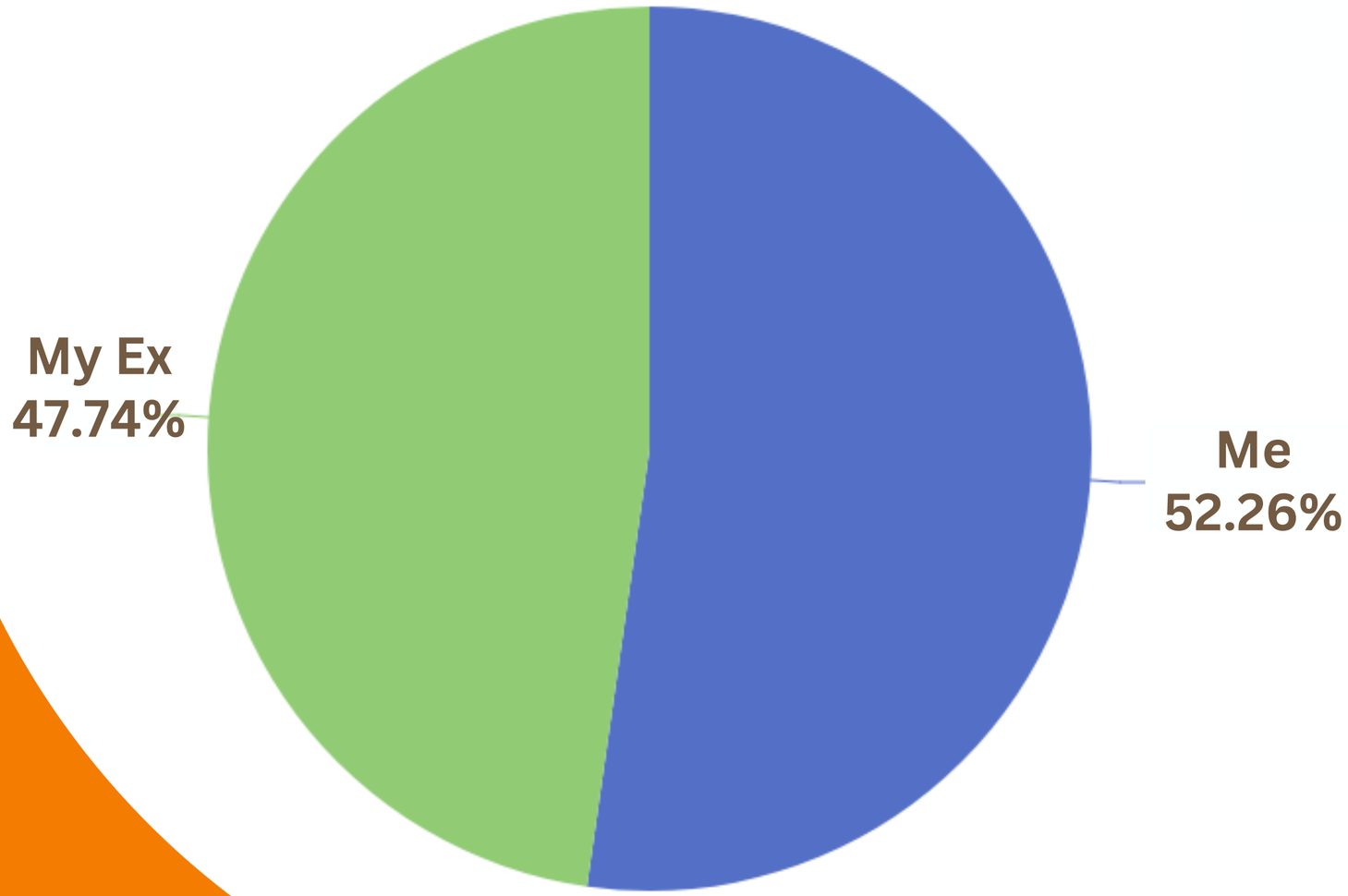
- Reality
  Uncorrelated
  with weekends

3. Pattern Analysis - Day



Chat happens most during lunch time and before sleeping

3. Pattern Analysis - Sender



- My expectation:
  Me = 3* Him

V.S.

- Reality
  Me ≈ Him

4. Sentiment Analysis

- Customize the sentiment value of frequently used words
  - From negative to positive:
    - 好傻/傻傻 (silly)
    - 肥肥鲨 (fat shark)

- Methodology

Message with Emotions
25,306

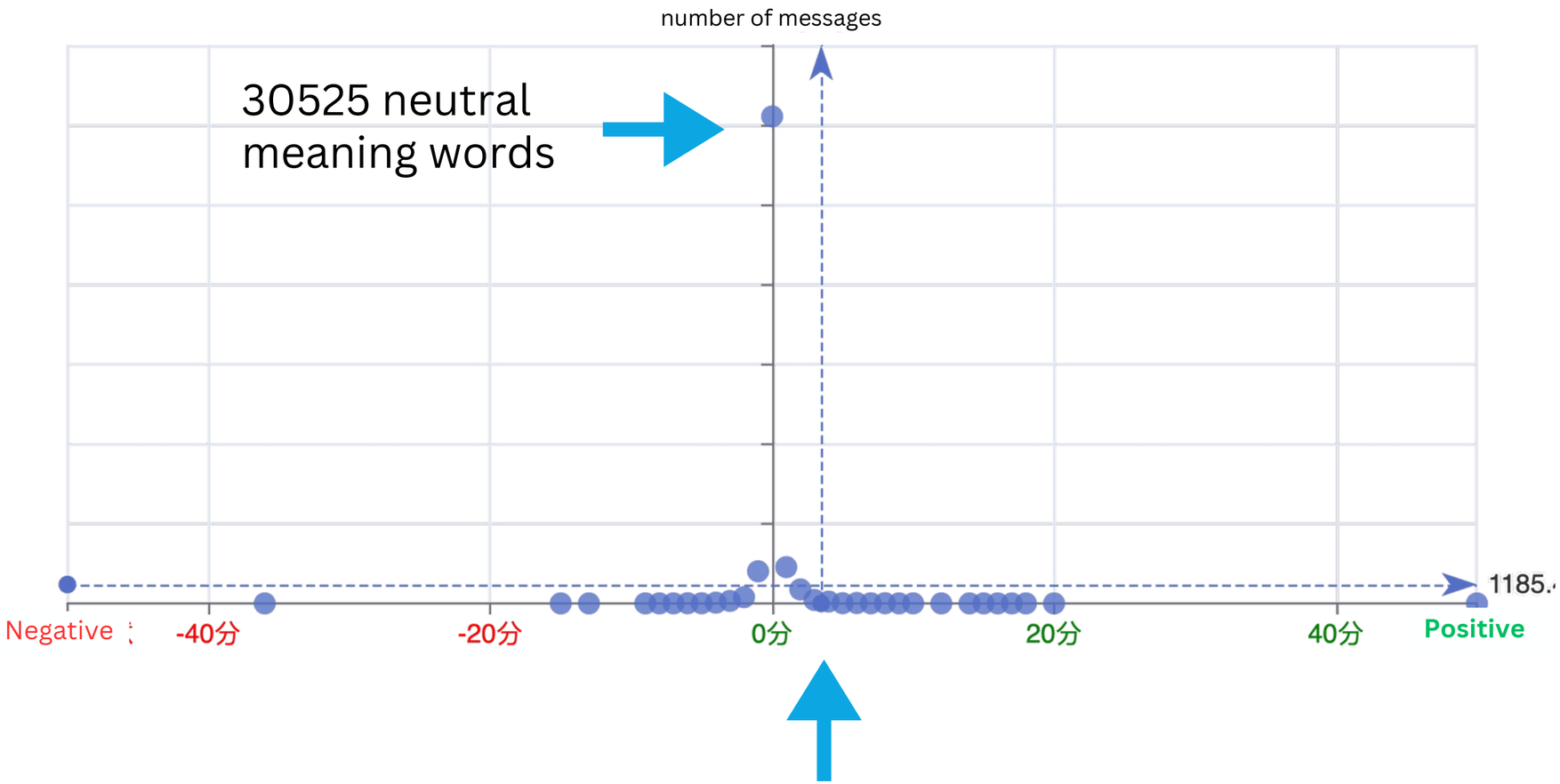
Words with Emotions
346,275

+: Positive

—: Negative

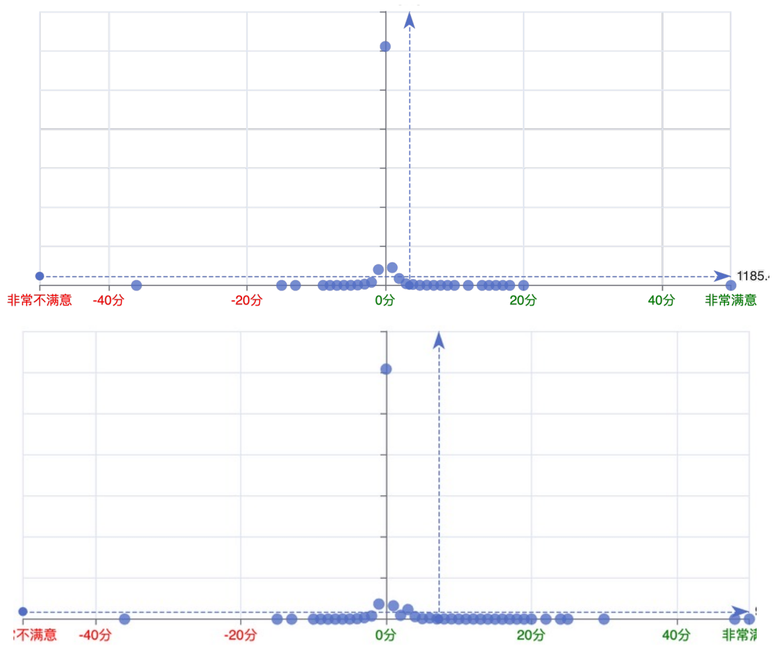
- Result

4. Sentiment Analysis



4. Sentiment Analysis

Comparison of Weighted Average Sentiment Value



Data	Time	Value
Me & my Ex	First 8 years	3.48
My friend & her Boyfriend	First 3 years	7.2
Me & my Ex	First 3 years	?

Comparison Group

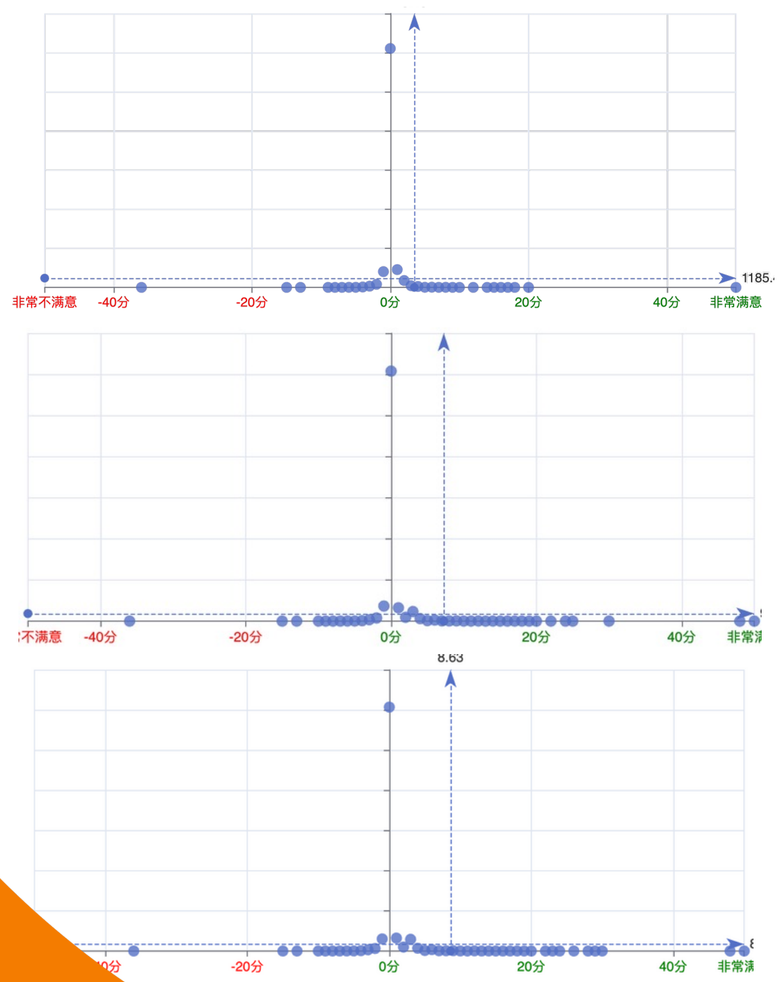
- Long-Distance Relationship too
- Business & Medical Students too
- Similar personality

MBTI:

ENFJ♀ & INTP ♂

4. Sentiment Analysis

Comparison of Weighted Average Sentiment Value

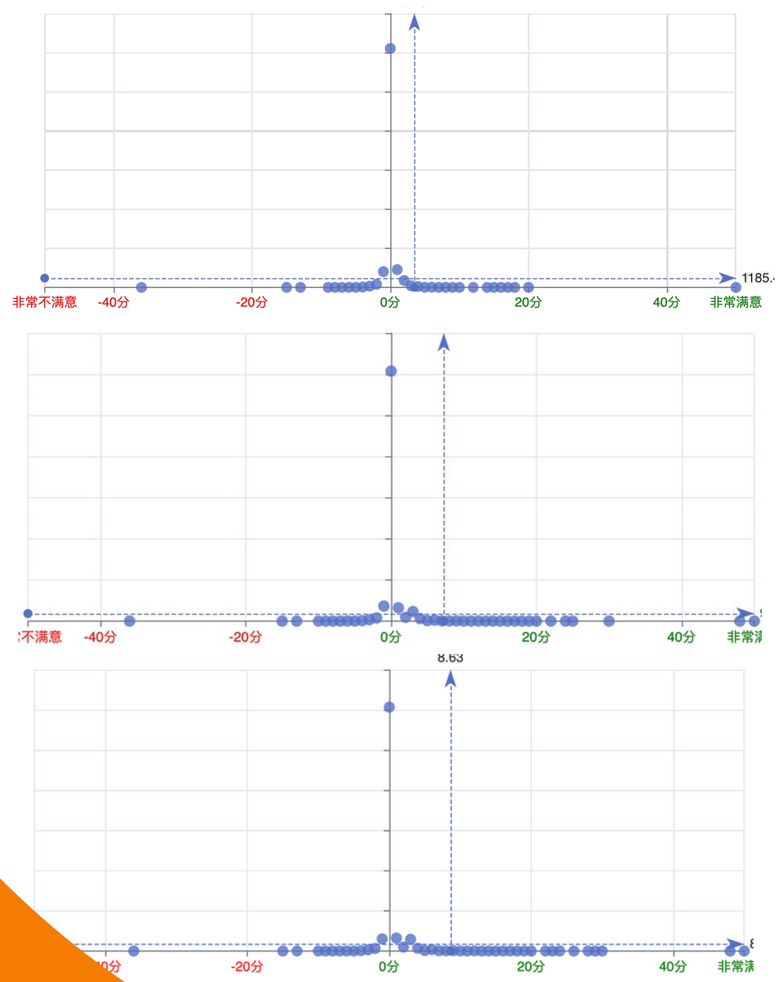


Data	Time	Value
Me & my Ex	First 8 years	3.48
My friend & her Boyfriend	First 3 years	7.2
Me & my Ex	First 3 years	8.63

After controlling the *Time*, our data outperforms

4. Sentiment Analysis

Comparison of Weighted Average Sentiment Value



Data	Time	Value
Me & my Ex	First 8 years	3.48
My friend & her Boyfriend	First 3 years	7.2
Me & my Ex	First 3 years	8.63

After controlling the *Time*, our data outperforms

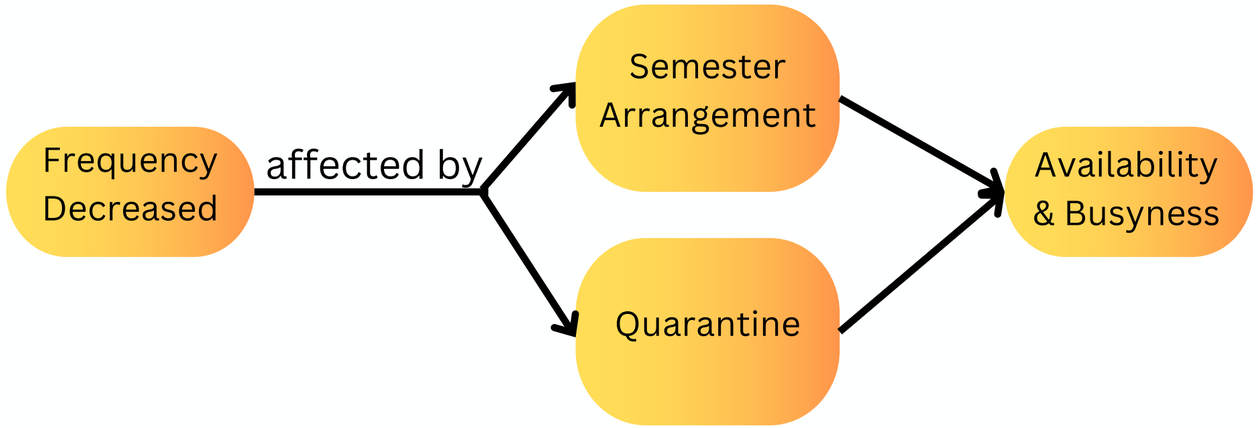
Time kills Everything 😂

5. Conclusion – What Kills Our Relationship

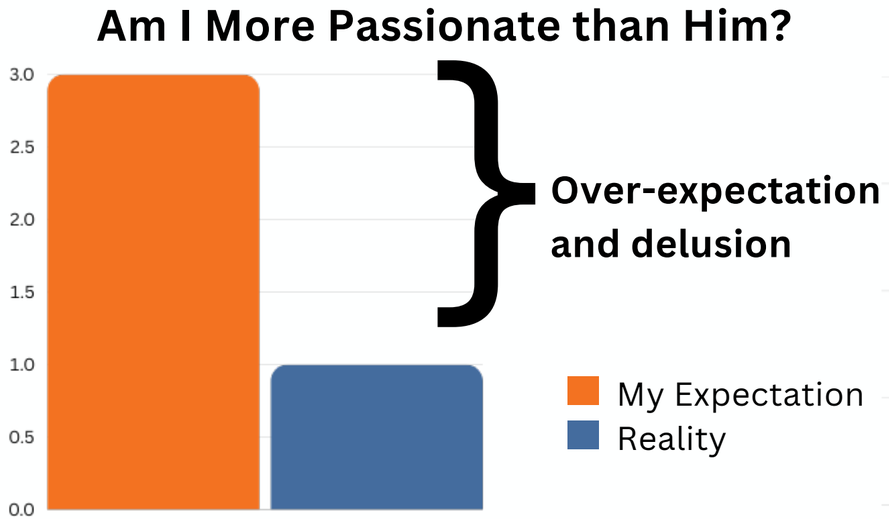
- **Busyness**
Time availability is the top factor affecting our chat frequency

- **Misperception**
The emotional delusion leads to deviation of perception from reality

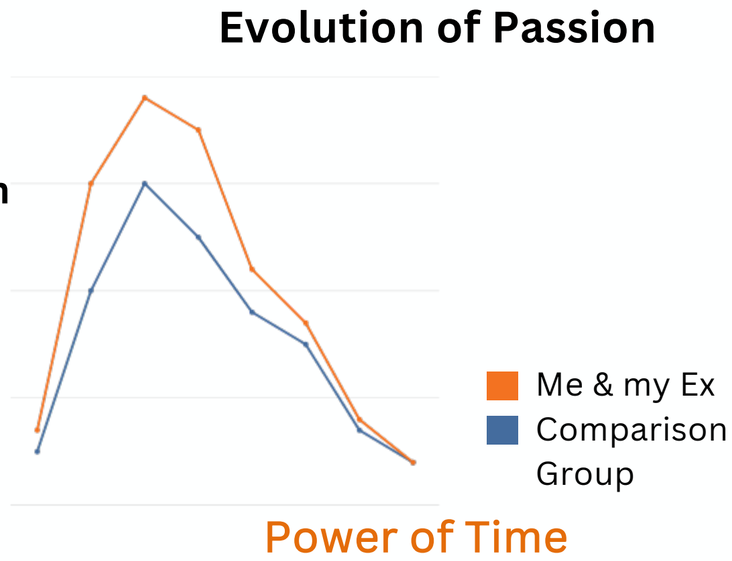
- **Power of Time**
Unavoidable fading of passion



Busyness



Misperception



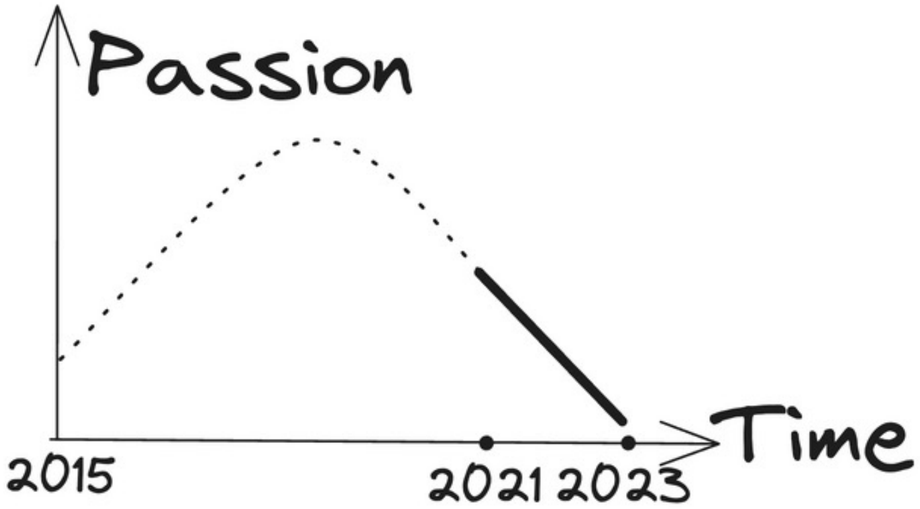
Power of Time

6. Limitation

- Misunderstanding of Language Reduces the Accuracy in understanding the meaning

Type	Example	Wrongly Classified as
Sarcasm	It’s OK (after getting stood up)	Positive

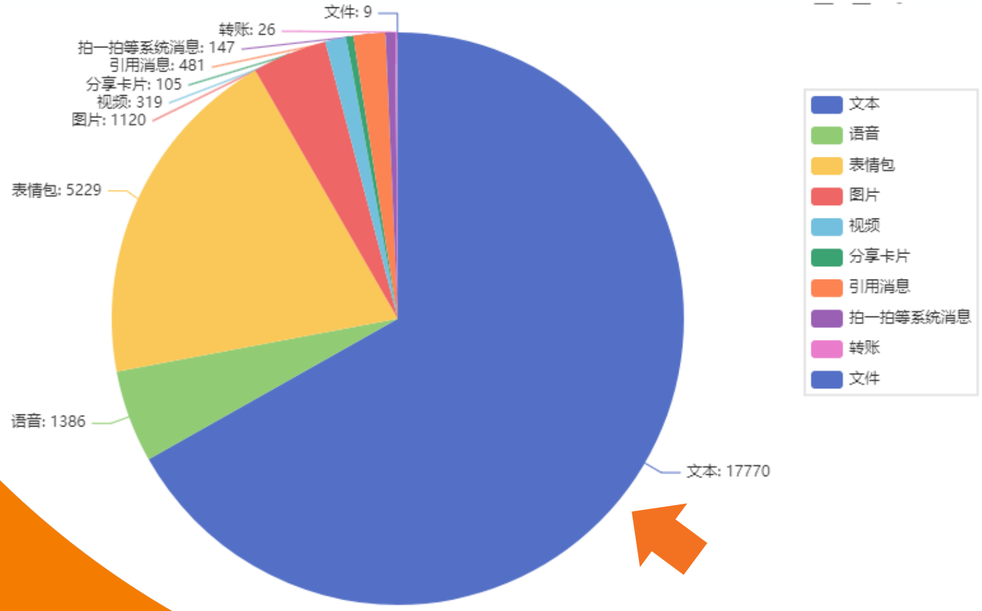
- Lack of the whole-life data
  - The information before 2021 May is absent, which disables the whole-life analysis on this 8-year-old relationship, as well as the comparison before and after the turning point



6. Limitation

- Bias of Partial Selection
  - Only string data is included (67.69%). Many emotion-bearing data types are ignored
  - Example:
    - Frequent “container” of emotion: emoji
    - On significant scenarios, voice messages or phone calls are preferred (e.g. on birthday, after receiving HKU offer)

Text only accounts for 67.79%



Most Used Emoji: Cuddle

Cuddle

Thank you!

Presented By:
Cheng Ling Jun