

# Initial Data Analysis for Social Buzz

# Today's agenda

- Project recap
- Problem
- The Analytics team
- Process
- Insights
- Summary

# Project Recap

**Client name:** Social Buzz

**Client industry:** Social media & content creation

- > 500 million active users/month in past 5 years
- IPO next year

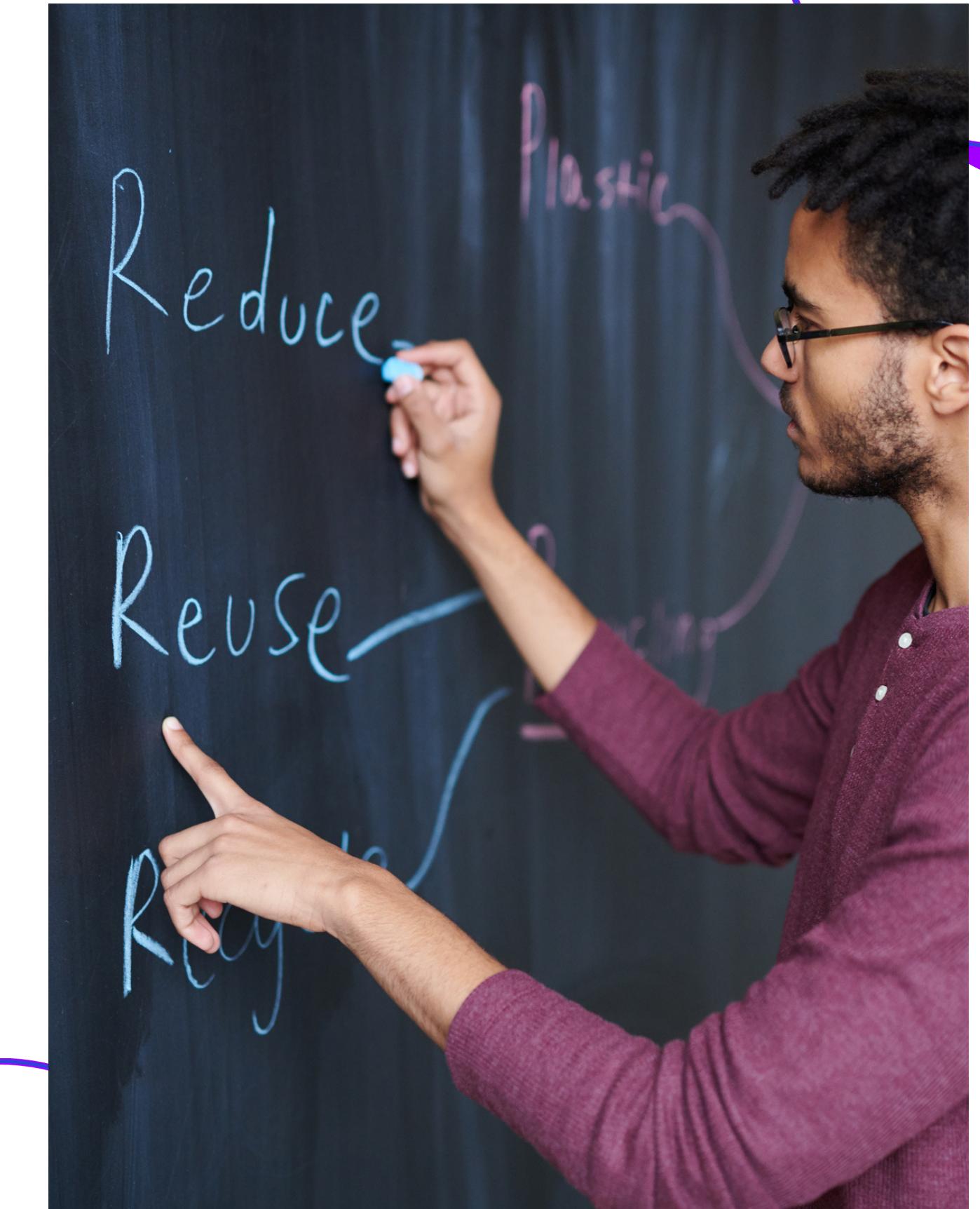
**3 months initial project objectives:**

- An audit of the big data practice
- Recommendation for a successful IPO
- Analysis of the top 5 content categories with the largest aggregate popularity

# Problem

Challenge: Need to analyse +100,000 posts with highly unstructured content every day to uncover insights.

***What are Social Buzz's top 5 popular content categories?***



# The Analytics team



**Andrew Fleming**  
**Chief Technology Architect**



**Marcus Rompton**  
**Senior Principal**



**Ronnie Hui**  
**Data Analyst**

# Process

1

Understanding the data

2

Cleaning the data

3

Modelling the data

4

Analyzing the data

5

Uncovering insights

# Insights

Animals

Science

Healthy eating

Technology

Food

Top 5 Popular  
Content Categories

Heart

Scared

Peeking

Hate

Interested

Top 5  
Reaction Types

May 2021

Jan 2021

Aug 2020

Top 3 highest  
posted Months

# About the Data

24573  
Posts

16  
Categories

16  
Reaction  
Types

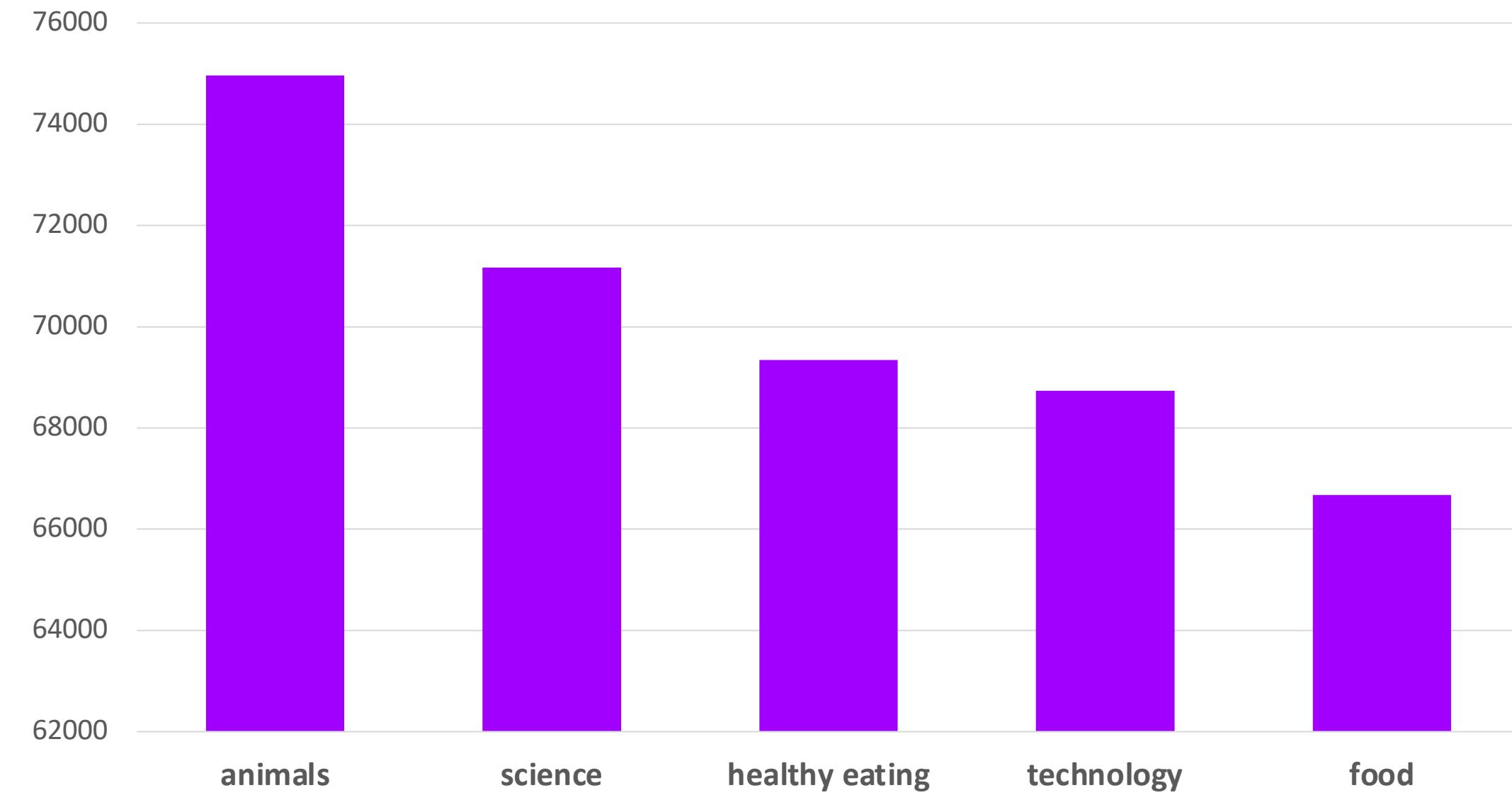
1. **Data Time Frame:** 18/6/2020 – 18/6/2021
2. **Number of Post:** 24573
3. **Number of Unique Post:** 962
4. **Content Type:** ['photo', 'video', 'GIF', 'audio']
5. **Category:** ['studying', 'healthy eating', 'technology', 'food', 'cooking', 'dogs', 'soccer', 'public speaking', 'science', 'tennis', 'travel', 'fitness', 'education', 'veganism', 'animals', 'culture']
6. **Reaction Type:** ['disgust', 'dislike', 'scared', 'interested', 'peaking', 'cherish', 'hate', 'love', 'indifferent', 'super love', 'intrigued', 'worried', 'like', 'heart', 'want', 'adore']
7. **Sentiment:** [positive', negative', neutral']
8. **Score:** 0-75

Categories	Total Score
animals	74965
science	71168
healthy eating	69339
technology	68738
food	66676

### User traits

Realistic  
Inquisitive  
Self-learning  
Self-improvement  
Health-focused  
Tech-savvy

### Top 5 Popular Content Categories by Total Score

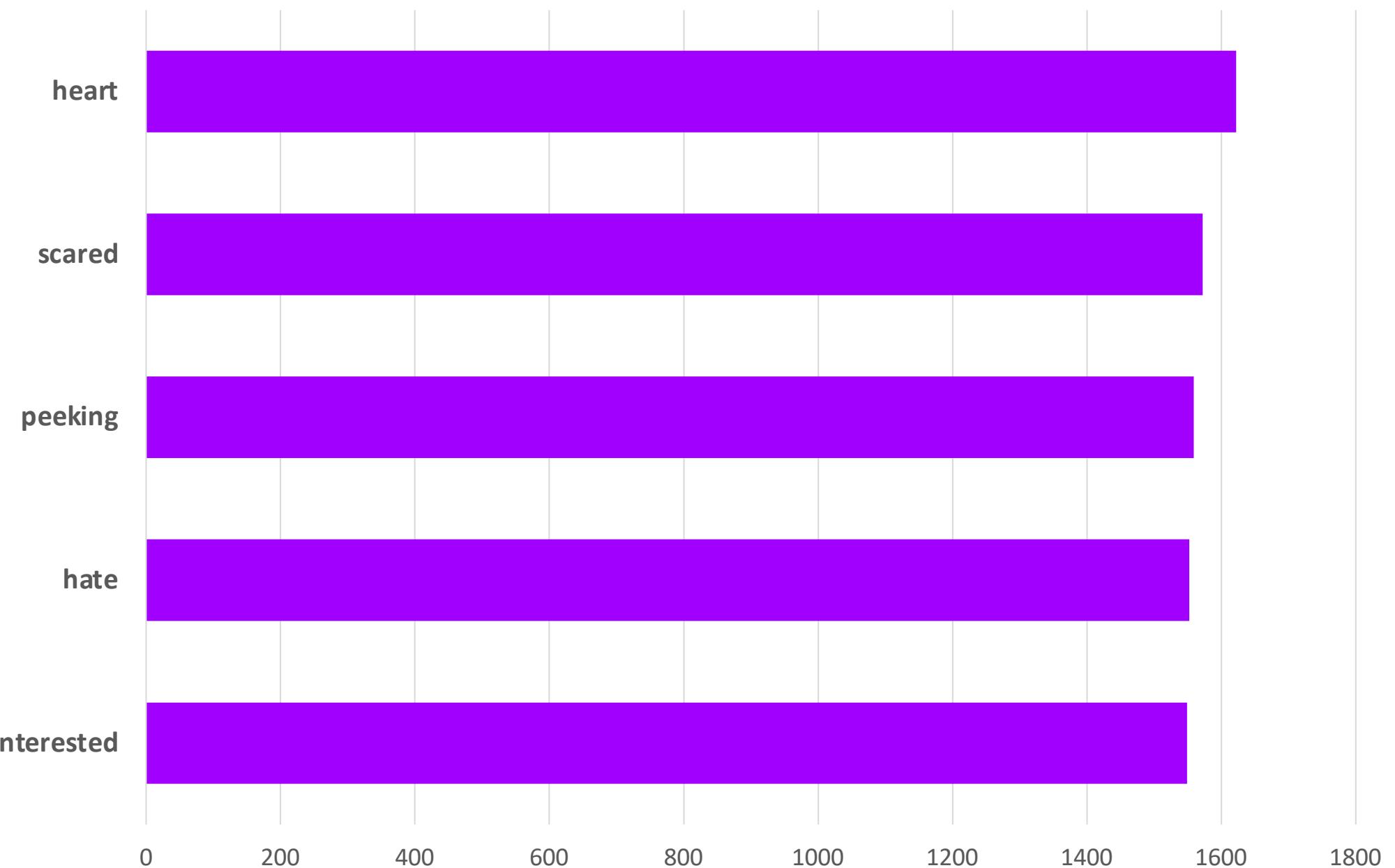


## Key Points

- “Heart” is mostly used
- No real difference among reaction type preferences

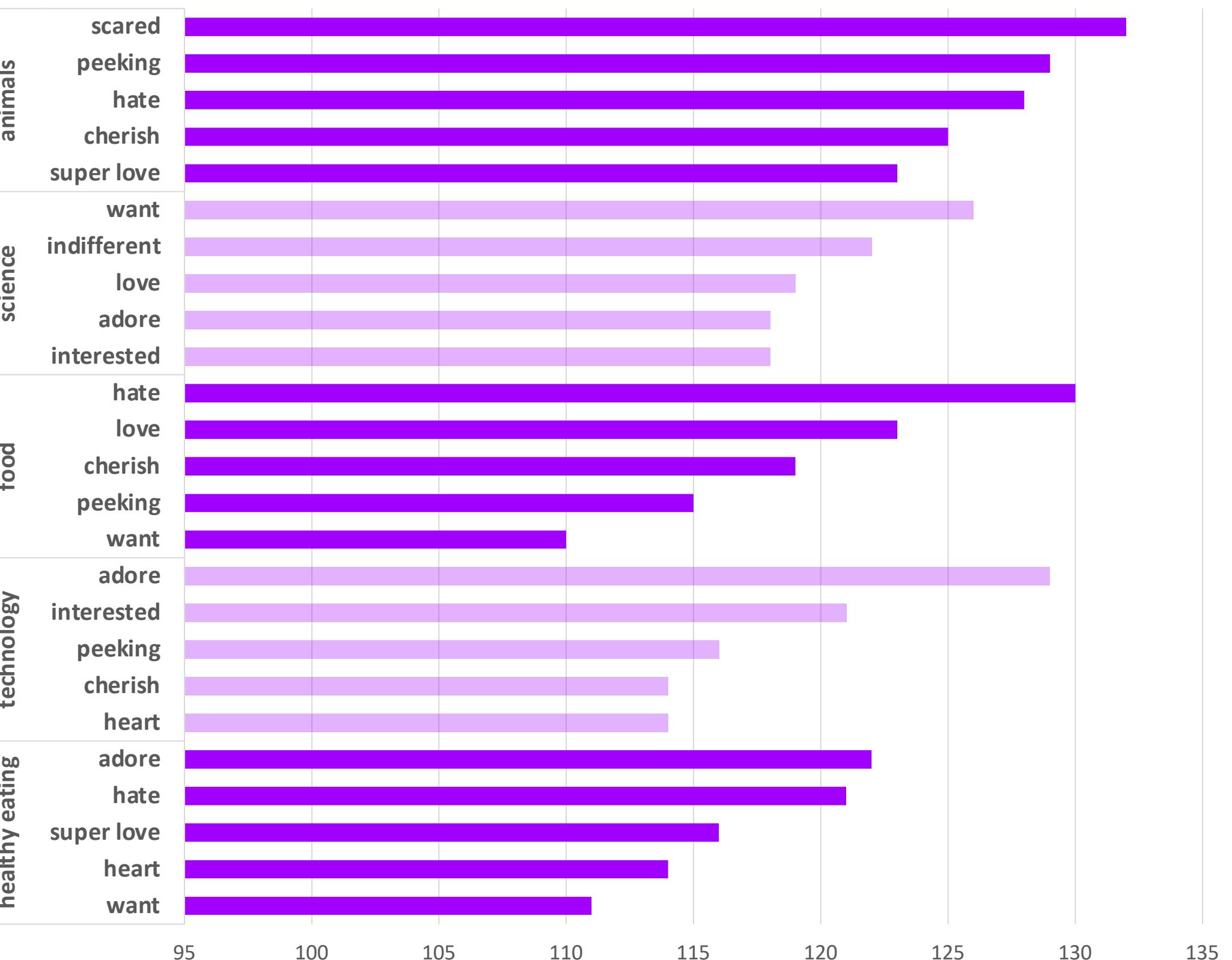
Reaction Type	Count of Post
heart	1622
scared	1572
peeking	1559
hate	1552
interested	1549

## Top 5 Reaction Types



Categories/Reactions	Count of Post
animals	637
scared	132
peeking	129
hate	128
cherish	125
super love	123
science	603
want	126
indifferent	122
love	119
adore	118
interested	118
food	597
hate	130
love	123
cherish	119
peeking	115
want	110
technology	594
adore	129
interested	121
peeking	116
cherish	114
heart	114
healthy eating	584
adore	122
hate	121
super love	116
heart	114
want	111

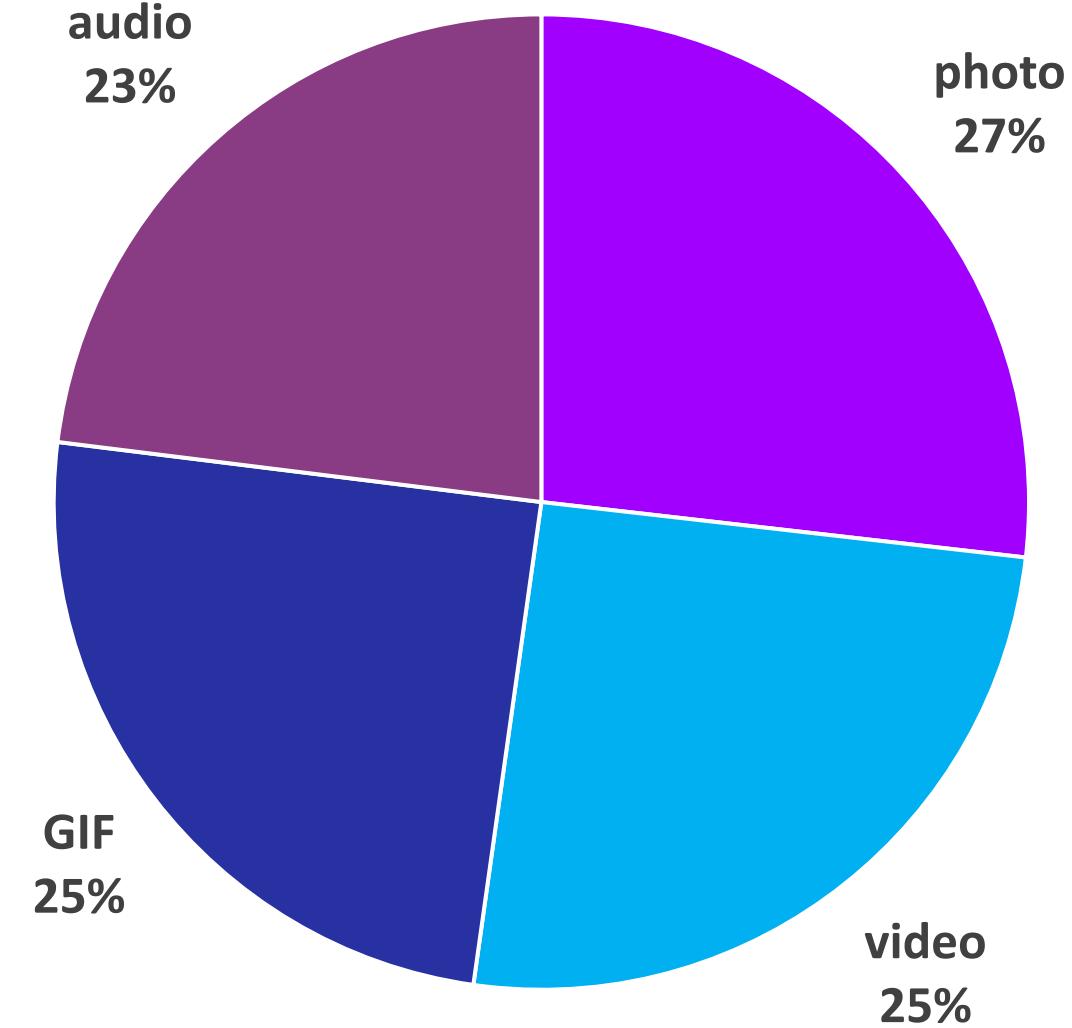
## Top 5 Reaction Types in each of the Top 5 Categories



Content Type	Count of Post
photo	6589
video	6245
GIF	6079
audio	5660

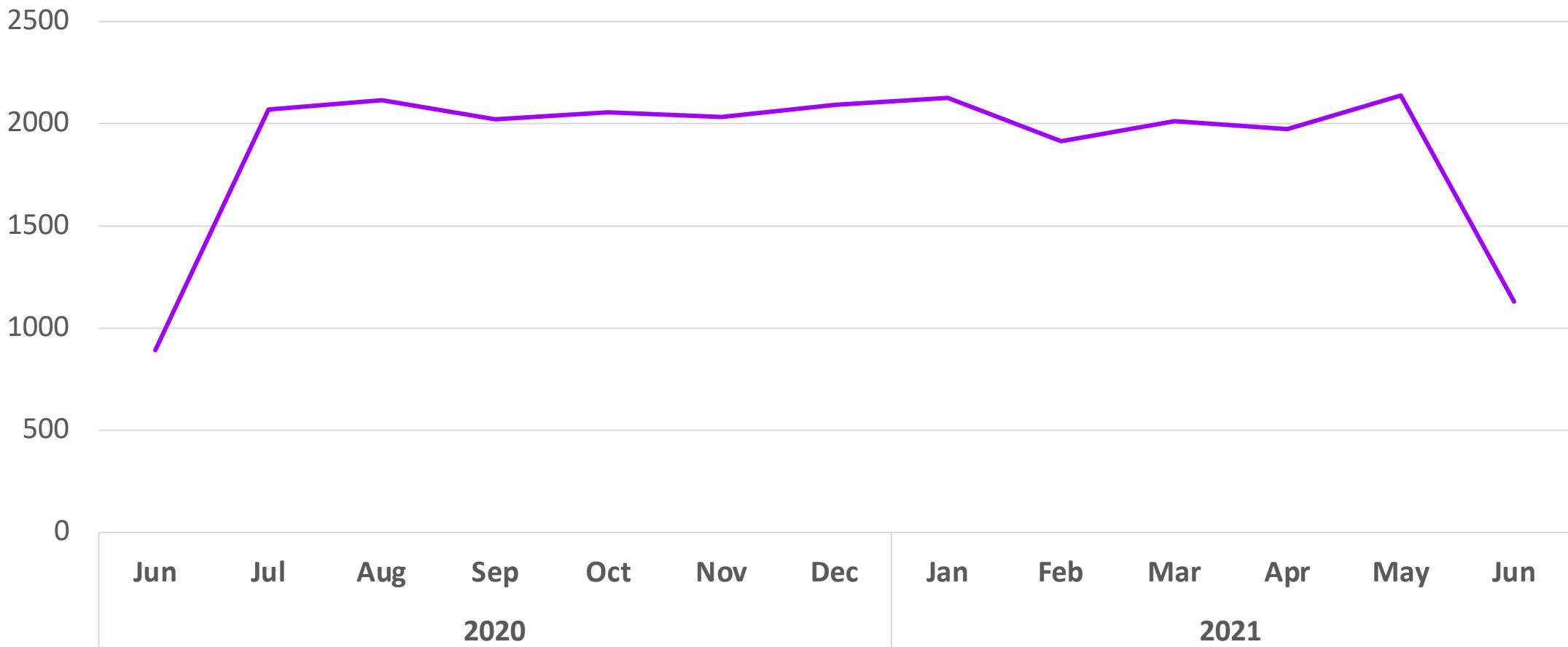


**Content Type Proportion**



Year/Month	Count of Post
<b>2020</b>	<b>13280</b>
Jun	892
Jul	2070
Aug	2114
Sep	2022
Oct	2056
Nov	2034
Dec	2092
<b>2021</b>	<b>11293</b>
Jan	2126
Feb	1914
Mar	2012
Apr	1974
May	2138
Jun	1129

## No. of Posts by Month from Jun 2020 to Jun 2021



Total  
24573  
Posts

Avg.  
1890  
posts/month

No real  
difference  
across  
months

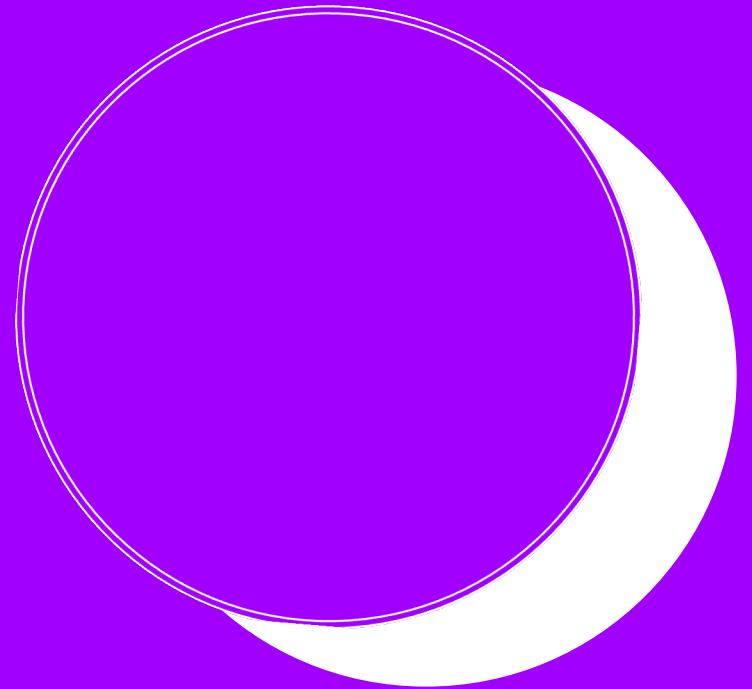
# Summary



**Analysis:** Users were interested in animals, science, health eating (food) and technology. Not real difference in reaction type preference. Content types were split equally. Not much difference in no. of post across months, reaching an average of 1890 posts/month.

**Insight:** Popular content categories reflected users are more realistic and inquisitive. They are likely to focus more on self improvement and understanding the world. They also keep up with new technologies.

**Conclusion:** Future actions may include in depth analysis such as correlation analysis and NLP sentiment analysis to find out user's characteristics and to uncover relationship between user's reaction and content selections.



Thank you!

ANY QUESTIONS?