# Introduction

Hello, I am Nelly and I am in the final stages of completing my Google Data Analytics certificate. This project thoroughly analyzes Social Buzz content categories and highlights the top 5 categories with the largest aggregate popularity. I will apply  "Ask, Prepare, Process, Analyze, Share, and Act" methodology effectively throughout the data analytics process. **Contact Me**

# Ask phase

Social-Buzz, a rapidly expanding media content creation company that prioritizes user anonymity and monitors user reactions on every piece of content.

The company is currently facing challenges related to managing its expanding scale. Over 100,000 pieces of content, ranging from text, images, videos, and GIFs are posted every day. Accenture has begun a 3-month project to address key areas, including an audit of Social Buzz's big data practice, recommendations for a successful IPO, and an analysis to identify the company's top 5 most popular content categories. Given the substantial amount of data available, Social Buzz needs an effective strategy to leverage this information and maximize its potential for growth.

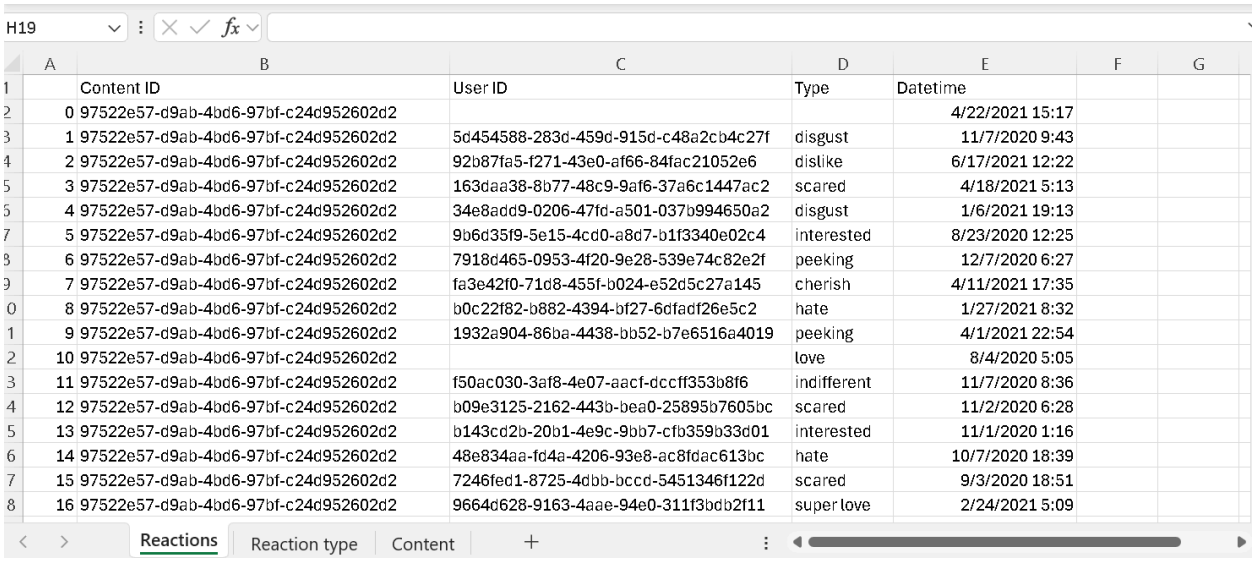
This project thoroughly analyzes their content categories and highlights the top 5 categories with the largest aggregate popularity.

# Prepare phase

**Data source:** these datasets were provided by Accenture and stored on forage. Forage is a free virtual work experience program from top companies.

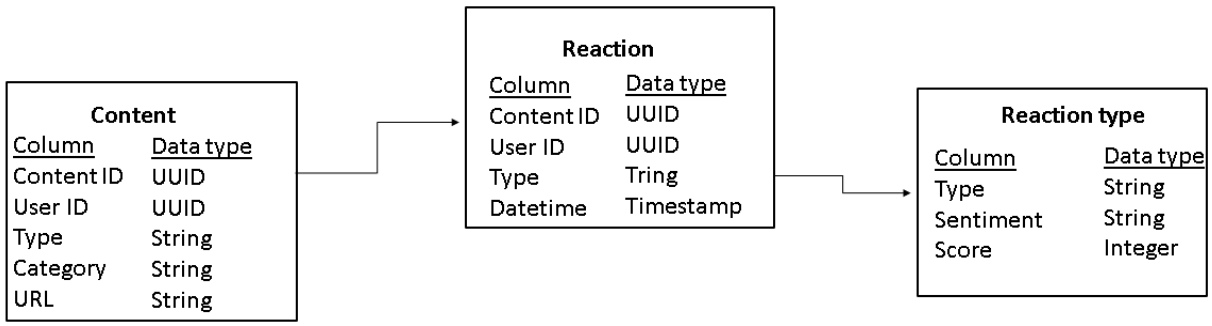
The dataset consists of a sample of over 25,000 reactions, 16 distinct qualitative content, and a quantitative score that is directly related to the sentiment and type of the  reaction types.

Accenture supplied three CSV files (Reactions, Reaction type, and Content). Given the small sample size, I will utilize Excel to complete this project. To be accessible, I  have merged the three tables into a single spreadsheet. Merging the tables was not essential for this step but it enables me to present all three data sets in a unified format.



| | Content ID | User ID | Type | Datetime |
|---|---|---|---|---|
| 0 | 97522e57-d9ab-4bd6-97bf-c24d952602d2 | | | 4/22/2021 15:17 |
| 1 | 97522e57-d9ab-4bd6-97bf-c24d952602d2 | 5d454588-283d-459d-915d-c48a2cb4c27f | disgust | 11/7/2020 9:43 |
| 2 | 97522e57-d9ab-4bd6-97bf-c24d952602d2 | 92b87fa5-f271-43e0-af66-84fac21052e6 | dislike | 6/17/2021 12:22 |
| 3 | 97522e57-d9ab-4bd6-97bf-c24d952602d2 | 163daa38-8b77-48c9-9af6-37a6c1447ac2 | scared | 4/18/2021 5:13 |
| 4 | 97522e57-d9ab-4bd6-97bf-c24d952602d2 | 34e8add9-0206-47fd-a501-037b994650a2 | disgust | 1/6/2021 19:13 |
| 5 | 97522e57-d9ab-4bd6-97bf-c24d952602d2 | 9b6d35f9-5e15-4cd0-a8d7-b1f3340e02c4 | interested | 8/23/2020 12:25 |
| 6 | 97522e57-d9ab-4bd6-97bf-c24d952602d2 | 7918d465-0953-4f20-9e28-539e74c82e2f | peeking | 12/7/2020 6:27 |
| 7 | 97522e57-d9ab-4bd6-97bf-c24d952602d2 | fa3e42f0-71d8-455f-b024-e52d5c27a145 | cherish | 4/11/2021 17:35 |
| 8 | 97522e57-d9ab-4bd6-97bf-c24d952602d2 | b0c22f82-b882-4394-bf27-6dfadf26e5c2 | hate | 1/27/2021 8:32 |
| 9 | 97522e57-d9ab-4bd6-97bf-c24d952602d2 | 1932a904-86ba-4438-bb52-b7e6516a4019 | peeking | 4/1/2021 22:54 |
| 10 | 97522e57-d9ab-4bd6-97bf-c24d952602d2 | | love | 8/4/2020 5:05 |
| 11 | 97522e57-d9ab-4bd6-97bf-c24d952602d2 | f50ac030-3af8-4e07-aacf-dccff353b8f6 | indifferent | 11/7/2020 8:36 |
| 12 | 97522e57-d9ab-4bd6-97bf-c24d952602d2 | b09e3125-2162-443b-bea0-25895b7605bc | scared | 11/2/2020 6:28 |
| 13 | 97522e57-d9ab-4bd6-97bf-c24d952602d2 | b143cd2b-20b1-4e9c-9bb7-cfb359b33d01 | interested | 11/1/2020 1:16 |
| 14 | 97522e57-d9ab-4bd6-97bf-c24d952602d2 | 48e834aa-fd4a-4206-93e8-ac8fdac613bc | hate | 10/7/2020 18:39 |
| 15 | 97522e57-d9ab-4bd6-97bf-c24d952602d2 | 7246fed1-8725-4dbb-bccd-5451346f122d | scared | 9/3/2020 18:51 |
| 16 | 97522e57-d9ab-4bd6-97bf-c24d952602d2 | 9664d628-9163-4aae-94e0-311f3bdb2f11 | super love | 2/24/2021 5:09 |

## Data model




### Content

ID: Unique ID of the content that was uploaded (automatically generated)
User ID: Unique ID of a user that exists in the User table
Type: A string detailing the type of content that was uploaded
Category: A string detailing the category that this content is relevant to
URL: Link to the location where this content is stored

### Reaction

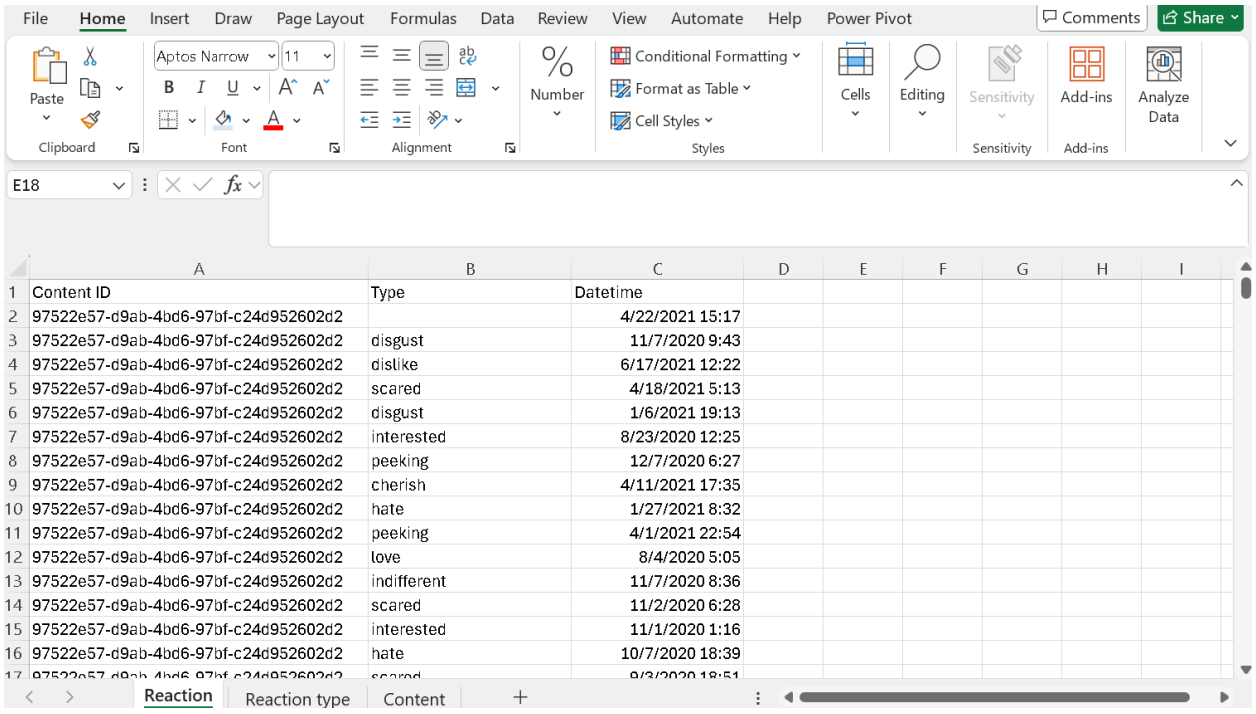
Content ID: Unique ID of a piece of content that was uploaded
User ID: Unique ID of a user that exists in the User table who reacted to this piece of content
Type: A string detailing the type of reaction this user gave
Datetime: The date and time of this reaction

**ReactionTypes**
Type: A string detailing the type of reaction this user gave
Sentiment: A string detailing whether this type of reaction is considered as positive, negative or neutral
Score: This is a number calculated by Social Buzz that quantifies how “popular” each reaction is. A reaction type with a higher score should be considered as a more popular reaction.

# Process phase

## Removing columns which are not relevant to this task

**Content:** removed User ID and URL column
**Reaction sheet:** removed User ID column
Lastly, I removed the first column (numbered column) of all three data because Excel already counts rows.




| | A | B | C |
|---|---|---|---|
| 1 | Content ID | Type | Datetime |
| 2 | 97522e57-d9ab-4bd6-97bf-c24d952602d2 | | 4/22/2021 15:17 |
| 3 | 97522e57-d9ab-4bd6-97bf-c24d952602d2 | disgust | 11/7/2020 9:43 |
| 4 | 97522e57-d9ab-4bd6-97bf-c24d952602d2 | dislike | 6/17/2021 12:22 |
| 5 | 97522e57-d9ab-4bd6-97bf-c24d952602d2 | scared | 4/18/2021 5:13 |
| 6 | 97522e57-d9ab-4bd6-97bf-c24d952602d2 | disgust | 1/6/2021 19:13 |
| 7 | 97522e57-d9ab-4bd6-97bf-c24d952602d2 | interested | 8/23/2020 12:25 |
| 8 | 97522e57-d9ab-4bd6-97bf-c24d952602d2 | peeking | 12/7/2020 6:27 |
| 9 | 97522e57-d9ab-4bd6-97bf-c24d952602d2 | cherish | 4/11/2021 17:35 |
| 10 | 97522e57-d9ab-4bd6-97bf-c24d952602d2 | hate | 1/27/2021 8:32 |
| 11 | 97522e57-d9ab-4bd6-97bf-c24d952602d2 | peeking | 4/1/2021 22:54 |
| 12 | 97522e57-d9ab-4bd6-97bf-c24d952602d2 | love | 8/4/2020 5:05 |
| 13 | 97522e57-d9ab-4bd6-97bf-c24d952602d2 | indifferent | 11/7/2020 8:36 |
| 14 | 97522e57-d9ab-4bd6-97bf-c24d952602d2 | scared | 11/2/2020 6:28 |
| 15 | 97522e57-d9ab-4bd6-97bf-c24d952602d2 | interested | 11/1/2020 1:16 |
| 16 | 97522e57-d9ab-4bd6-97bf-c24d952602d2 | hate | 10/7/2020 18:39 |
| 17 | 97522e57-d9ab-4bd6-97bf-c24d952602d2 | scared | 9/3/2020 18:51 |

## Removing rows that have values which are missing, duplicate, and data extraction

I applied filters on all three sheets separately (Reaction, Reaction type, and Content) to inspect for duplicate, blanks, and errors (typo).

**Reaction sheet**

- Extracted month from datetime column by using the convert text to column wizard and renamed the columns (Month, Time).
- Joined the time column by using the formula “=TEXT(cell, "hh:mm") & " " & cell” and changed data type to display month in “text”.
- Changed the month column to “short date” and time to “time”(Ref R).

| Content ID | Reaction Type | Month | Time |
|---|---|---|---|
| 97522e57-d9ab-4bd6-97bf-c24d952602d2 | disgust | November | 09:43 AM |
| 97522e57-d9ab-4bd6-97bf-c24d952602d2 | dislike | June | 12:22 PM |
| 97522e57-d9ab-4bd6-97bf-c24d952602d2 | scared | April | 05:13 AM |
| 97522e57-d9ab-4bd6-97bf-c24d952602d2 | disgust | January | 07:13 PM |
| 97522e57-d9ab-4bd6-97bf-c24d952602d2 | interested | August | 12:25 PM |
| 97522e57-d9ab-4bd6-97bf-c24d952602d2 | peeking | December | 06:27 AM |
| 97522e57-d9ab-4bd6-97bf-c24d952602d2 | cherish | April | 05:35 PM |
| 97522e57-d9ab-4bd6-97bf-c24d952602d2 | hate | January | 08:32 AM |
| 97522e57-d9ab-4bd6-97bf-c24d952602d2 | peeking | April | 10:54 PM |
| 97522e57-d9ab-4bd6-97bf-c24d952602d2 | love | August | 05:05 AM |
| 97522e57-d9ab-4bd6-97bf-c24d952602d2 | indifferent | November | 08:36 AM |
| 97522e57-d9ab-4bd6-97bf-c24d952602d2 | scared | November | 06:28 AM |
| 97522e57-d9ab-4bd6-97bf-c24d952602d2 | interested | November | 01:16 AM |
| 97522e57-d9ab-4bd6-97bf-c24d952602d2 | hate | October | 06:39 PM |
| 97522e57-d9ab-4bd6-97bf-c24d952602d2 | scared | September | 06:51 PM |
| 97522e57-d9ab-4bd6-97bf-c24d952602d2 | super love | February | 05:09 AM |
| 97522e57-d9ab-4bd6-97bf-c24d952602d2 | peeking | September | 06:24 AM |
| 97522e57-d9ab-4bd6-97bf-c24d952602d2 | indifferent | February | 11:37 AM |

*Ref R: Month extraction*

**Reaction type sheet**

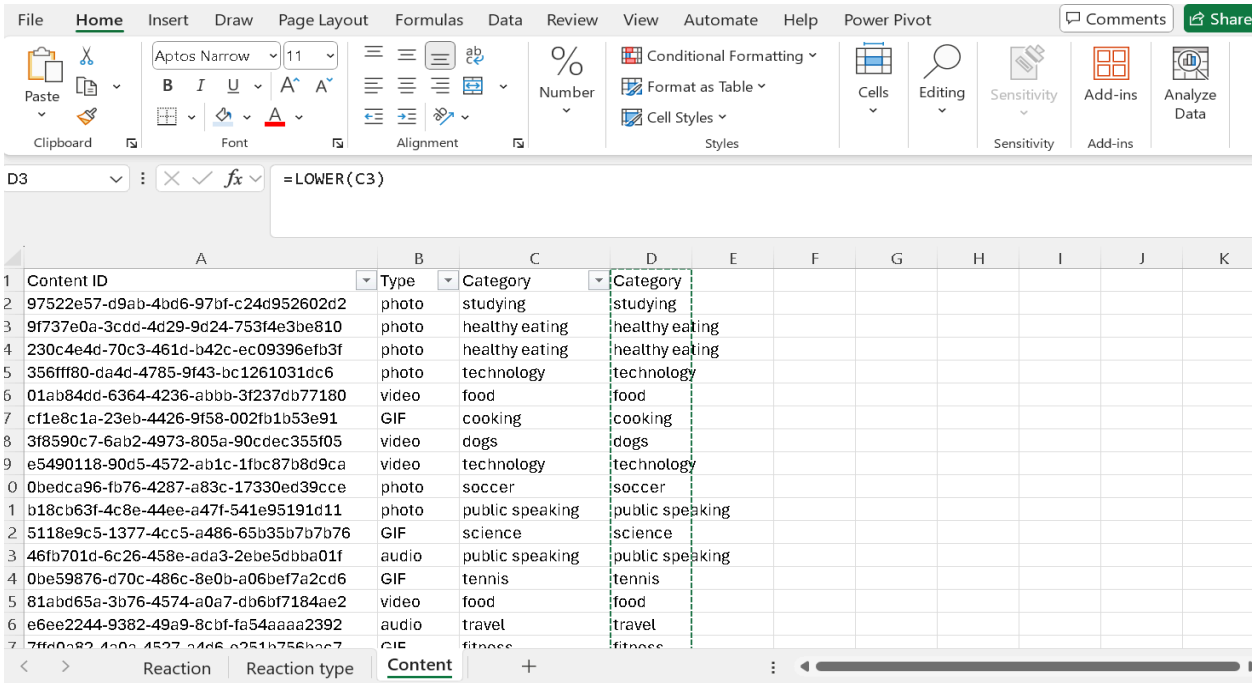
- **Fixed the score duplicate adding “1” to the want type: new score 71 (***Ref Rt)*

| Type | Sentime | Score |
|---|---|---|
| want | positive | 71 |
| cherish | positive | 70 |

*Ref Rt: Duplicate*

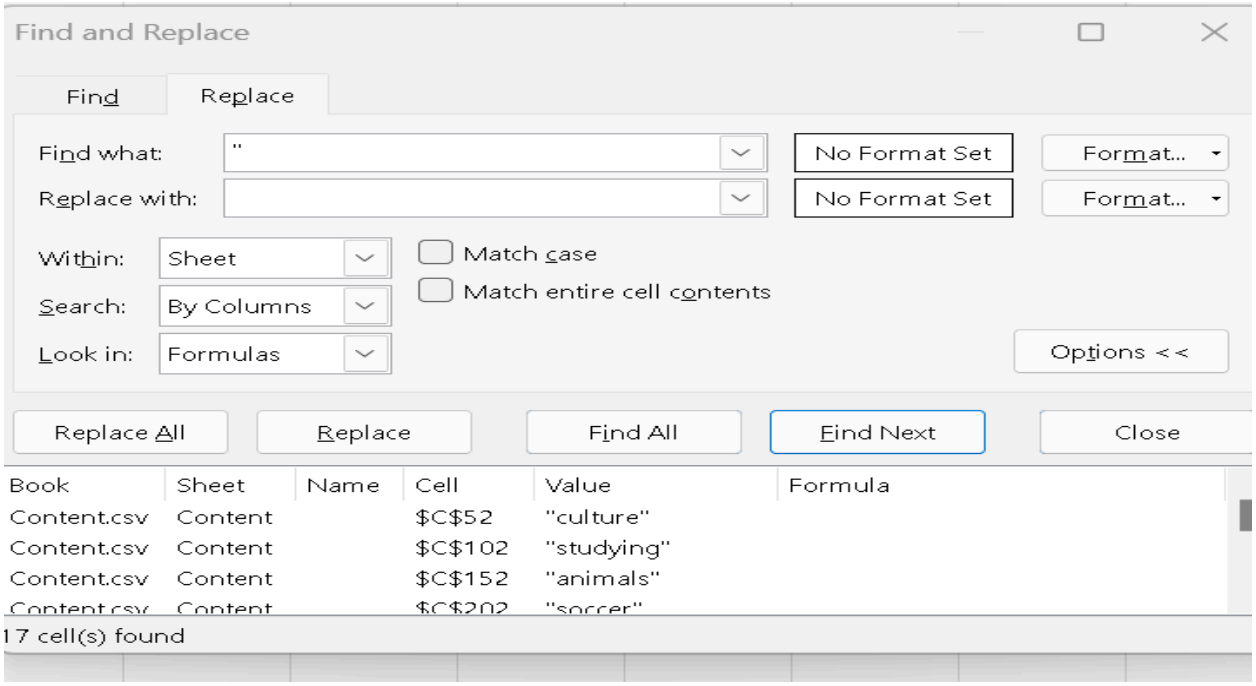
**Content sheet**

- In the category column, I used the “LOWER” function to change values to lowercase. (In an empty column type “=LOWER(cell)”, use the fill handle to copy the formula. Lastly, I only pasted the values *(Ref C1).*
- I used the find and replace quotation “ ” with blank *(Ref C2).*
- I replaced “dogs” with animals *(Ref C3).*

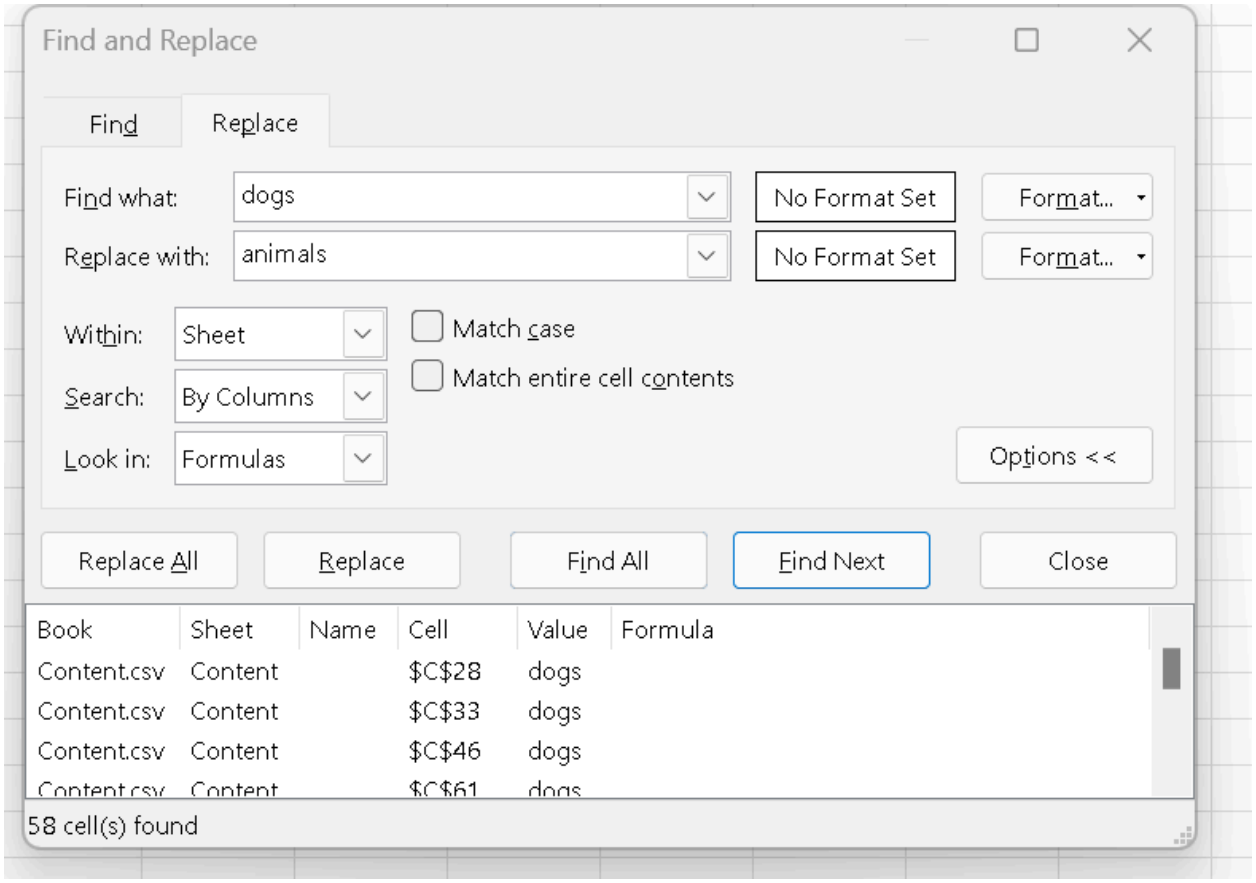



| | A | B | C | D |
|---|---|---|---|---|
| 1 | Content ID | Type | Category | Category |
| 2 | 97522e57-d9ab-4bd6-97bf-c24d952602d2 | photo | studying | studying |
| 3 | 9f737e0a-3cdd-4d29-9d24-753f4e3be810 | photo | healthy eating | healthy eating |
| 4 | 230c4e4d-70c3-461d-b42c-ec09396efb3f | photo | healthy eating | healthy eating |
| 5 | 356fff80-da4d-4785-9f43-bc1261031dc6 | photo | technology | technology |
| 6 | 01ab84dd-6364-4236-abbb-3f237db77180 | video | food | food |
| 7 | cf1e8c1a-23eb-4426-9f58-002fb1b53e91 | GIF | cooking | cooking |
| 8 | 3f8590c7-6ab2-4973-805a-90cdec355f05 | video | dogs | dogs |
| 9 | e5490118-90d5-4572-ab1c-1fbc87b8d9ca | video | technology | technology |
| 10 | 0bedca96-fb76-4287-a83c-17330ed39cce | photo | soccer | soccer |
| 11 | b18cb63f-4c8e-44ee-a47f-541e95191d11 | photo | public speaking | public speaking |
| 12 | 5118e9c5-1377-4cc5-a486-65b35b7b7b76 | GIF | science | science |
| 13 | 46fb701d-6c26-458e-ada3-2ebe5dbba01f | audio | public speaking | public speaking |
| 14 | 0be59876-d70c-486c-8e0b-a06bef7a2cd6 | GIF | tennis | tennis |
| 15 | 81abd65a-3b76-4574-a0a7-db6bf7184ae2 | video | food | food |
| 16 | e6ee2244-9382-49a9-8cbf-fa54aaaa2392 | audio | travel | travel |

*Ref C1: Change value to lowercase*




*Ref C2: Removing quotation mark*




*(Ref C2): Replaced “dogs” with animals*

# Aggregation

I used the Reaction table as the base table, then first joined the relevant columns from the Content data set, and then the Reaction Types data set. I named the tables to enable the use of the “VLookUp” formula, and renamed the joined column (Ref Ag).

| Content ID | Reaction Type | Month | Time | Content Type | Content Categ | Sentiment | Score |
|---|---|---|---|---|---|---|---|
| 97522e57-d9ab-4bd6-97bf-c24d952602d2 | disgust | November | 09:43 AM | photo | studying | negative | 0 |
| 97522e57-d9ab-4bd6-97bf-c24d952602d2 | dislike | June | 12:22 PM | photo | studying | negative | 10 |
| 97522e57-d9ab-4bd6-97bf-c24d952602d2 | scared | April | 05:13 AM | photo | studying | negative | 15 |
| 97522e57-d9ab-4bd6-97bf-c24d952602d2 | disgust | January | 07:13 PM | photo | studying | negative | 0 |
| 97522e57-d9ab-4bd6-97bf-c24d952602d2 | interested | August | 12:25 PM | photo | studying | positive | 30 |
| 97522e57-d9ab-4bd6-97bf-c24d952602d2 | peeking | December | 06:27 AM | photo | studying | neutral | 35 |
| 97522e57-d9ab-4bd6-97bf-c24d952602d2 | cherish | April | 05:35 PM | photo | studying | positive | 70 |
| 97522e57-d9ab-4bd6-97bf-c24d952602d2 | hate | January | 08:32 AM | photo | studying | negative | 5 |
| 97522e57-d9ab-4bd6-97bf-c24d952602d2 | peeking | April | 10:54 PM | photo | studying | neutral | 35 |
| 97522e57-d9ab-4bd6-97bf-c24d952602d2 | love | August | 05:05 AM | photo | studying | positive | 65 |
| 97522e57-d9ab-4bd6-97bf-c24d952602d2 | indifferent | November | 08:36 AM | photo | studying | neutral | 20 |
| 97522e57-d9ab-4bd6-97bf-c24d952602d2 | scared | November | 06:28 AM | photo | studying | negative | 15 |
| 97522e57-d9ab-4bd6-97bf-c24d952602d2 | interested | November | 01:16 AM | photo | studying | positive | 30 |
| 97522e57-d9ab-4bd6-97bf-c24d952602d2 | hate | October | 06:39 PM | photo | studying | negative | 5 |
| 97522e57-d9ab-4bd6-97bf-c24d952602d2 | scared | September | 06:51 PM | photo | studying | negative | 15 |
| 97522e57-d9ab-4bd6-97bf-c24d952602d2 | super love | February | 05:09 AM | photo | studying | positive | 75 |
| 97522e57-d9ab-4bd6-97bf-c24d952602d2 | peeking | September | 06:24 AM | photo | studying | neutral | 35 |
| 97522e57-d9ab-4bd6-97bf-c24d952602d2 | indifferent | February | 11:37 AM | photo | studying | neutral | 20 |
| 97522e57-d9ab-4bd6-97bf-c24d952602d2 | interested | May | 07:44 PM | photo | studying | positive | 30 |
| 97522e57-d9ab-4bd6-97bf-c24d952602d2 | intrigued | January | 04:03 PM | photo | studying | positive | 45 |
| 97522e57-d9ab-4bd6-97bf-c24d952602d2 | peeking | November | 05:26 PM | photo | studying | neutral | 35 |
| 97522e57-d9ab-4bd6-97bf-c24d952602d2 | worried | April | 08:47 PM | photo | studying | negative | 12 |
| 97522e57-d9ab-4bd6-97bf-c24d952602d2 | like | June | 04:46 PM | photo | studying | positive | 50 |
| 97522e57-d9ab-4bd6-97bf-c24d952602d2 | heart | April | 02:29 PM | photo | studying | positive | 60 |
| 97522e57-d9ab-4bd6-97bf-c24d952602d2 | worried | March | 07:21 PM | photo | studying | negative | 12 |

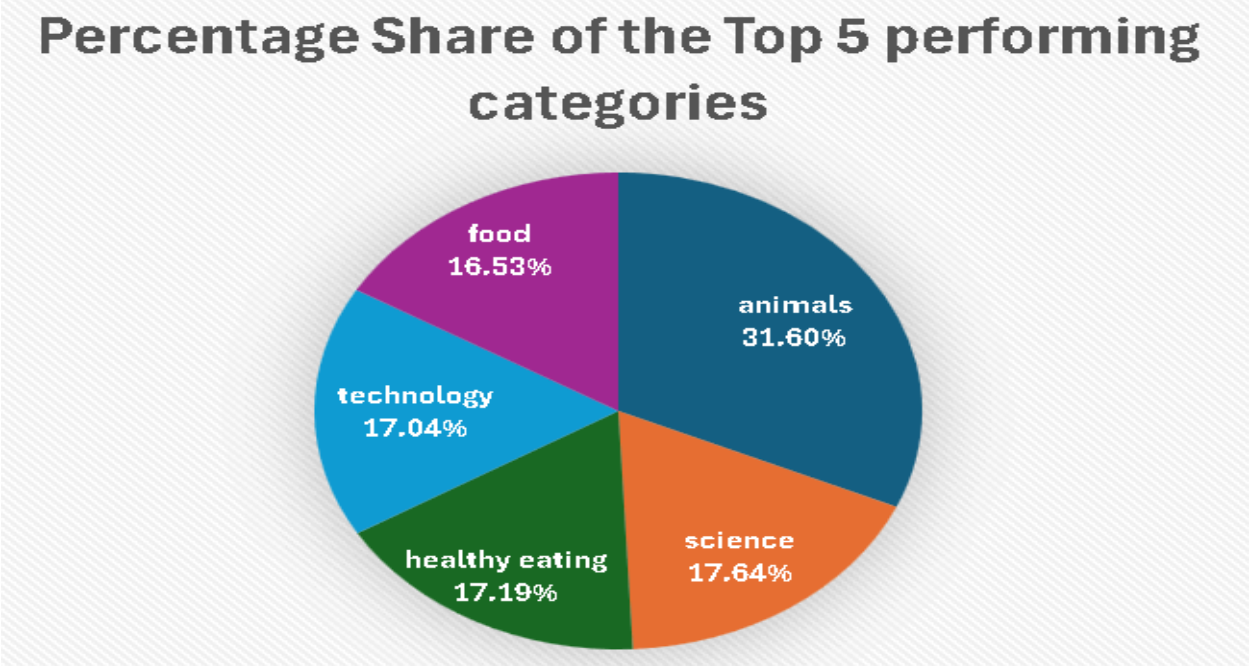
*(Ref Ag): Aggregation*

# Analyze phase

Figure out the Top 5 performing categories by adding up the total scores for each category (using the “Sum If” formula). Next, I sorted the data by “sum of score” DESC.

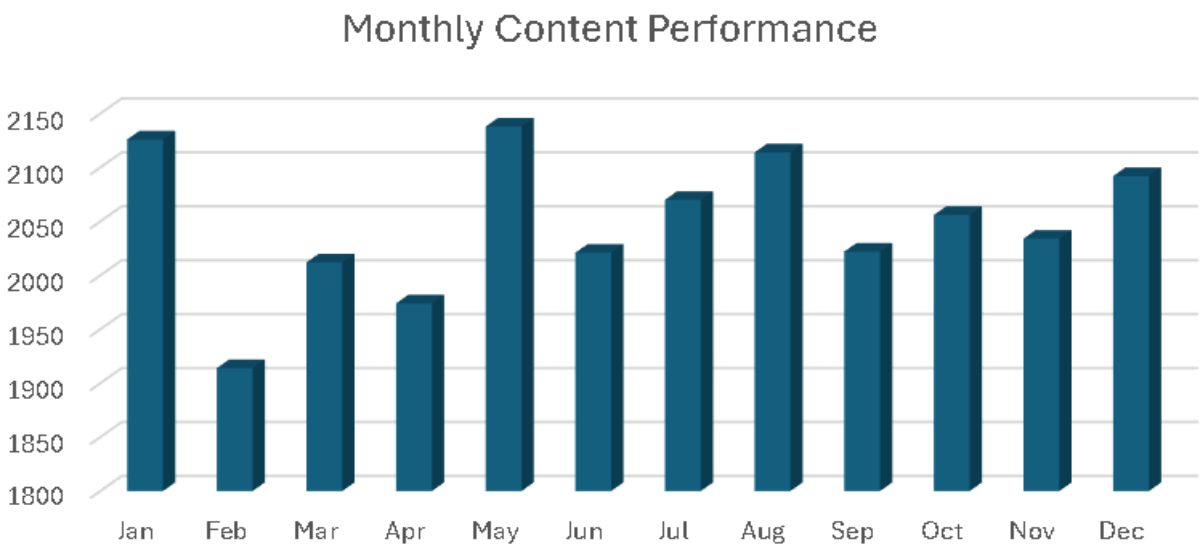
| Top 5 performing categories | |
|---|---|
| **Content Category** | **Sum of score** |
| animals | 127681 |
| science | 71294 |
| healthy eating | 69450 |
| technology | 68851 |
| food | 66786 |

Next, I used a pie chart to represent the **top 5 categories**.



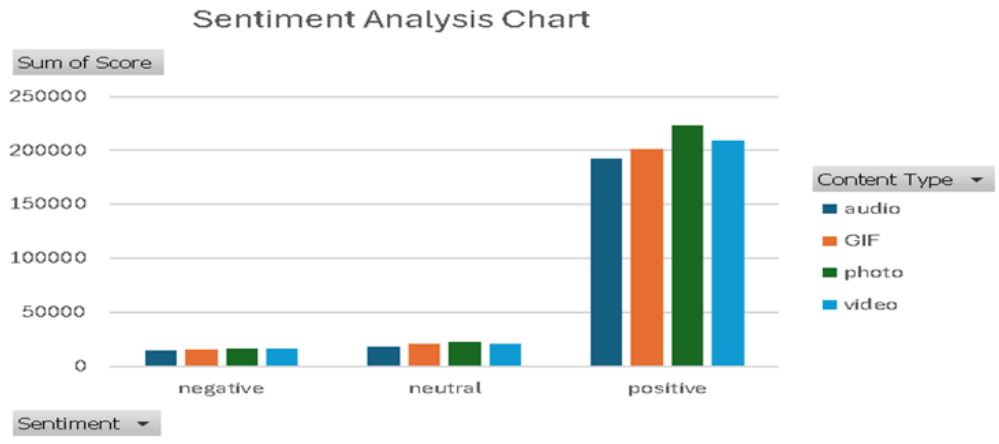

**Key Takeaways**:
**Animals** is the top performer, followed by **Science** and **Healthy Eating**.
**Technology** and **Food** round out the top 5, all showing strong engagement.

Let's count the monthly content and use a bar chart to represent the “monthly performance”




- **Peak Activity:** January and May saw the highest levels of activity.
- **Dip in February:** There was a noticeable decrease in activity during February.
- **Stable Months:** Engagement remained consistent around the same level for most months, from June to December.
- **General Trend:** A slight decline in the early year, with steady engagement throughout the rest of the year.

Lastly, we will look at the sentiments each content type received.

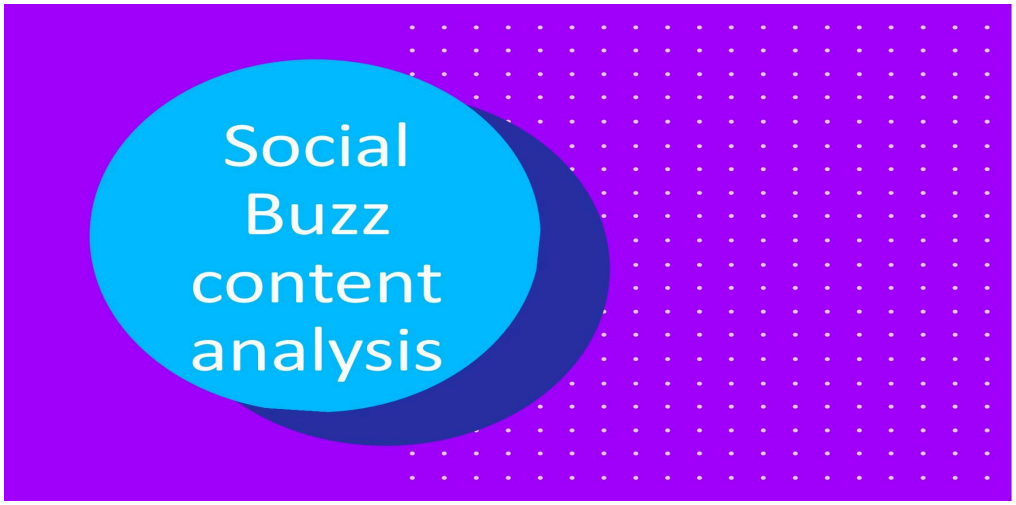



User interaction with videos is strong.

## Summary

The analysis of Social Buzz's content performance reveals that Animals and Science are the top two categories driving the highest engagement, suggesting a strong interest in educational and lifestyle content. Healthy Eating, Technology, and Food also rank highly, but with slightly less consistent engagement. The overall trend shows stable content activity year-round, with peaks in January and May and a slight dip in February. Furthermore, users express more sentiment reactions to videos contents.These findings highlight key areas for growth and provide actionable insights for refining the company's content strategy moving forward.

# Share



# Today's agenda

Project recap
Problem
The Analytics team
Process
Insights
Summary

## Project Recap

Social Buzz is a fast-growing technology company that needs to manage its current scale. Accenture has begun a 3-month initial project focusing on these tasks:

- An audit of Social Buzz's big data practice
- Recommendations for a successful IPO
- Analysis to find Social Buzz's top 5 popular categories of content

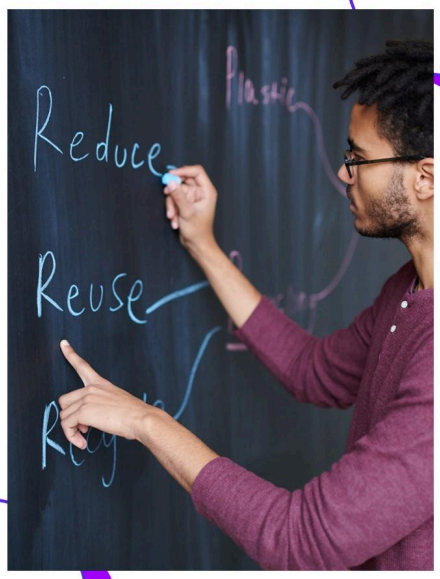
# Problem

Over 100,000 pieces of content

36,5000,000 Content per year

How to capitalize on it when there is so much?

Analysis to find Social Buzz's top 5 popular categories of content




# The Analytics team

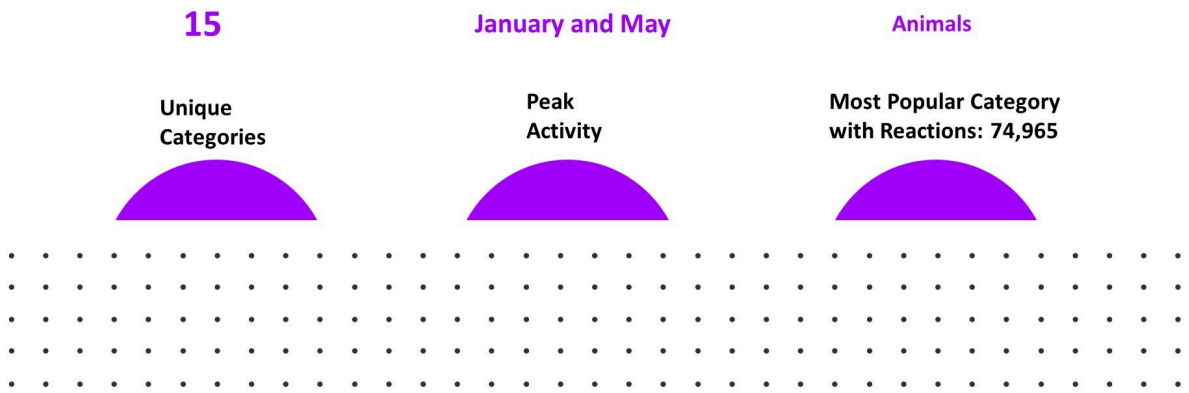
**Andrew Fleming** (Chief Technical Architect)

**Marcus Rompton** (Senior Principle)

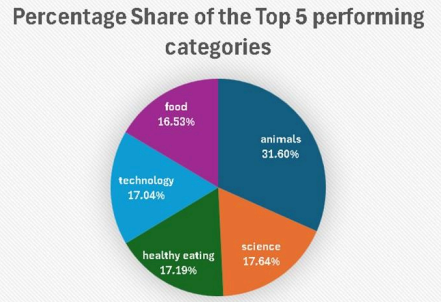
**Nelly Moja** (Data Analyst)


Process
1
Data Understanding
2
Data Cleaning
3
Data Modeling
4
Data Analysis
5
Uncover Insights

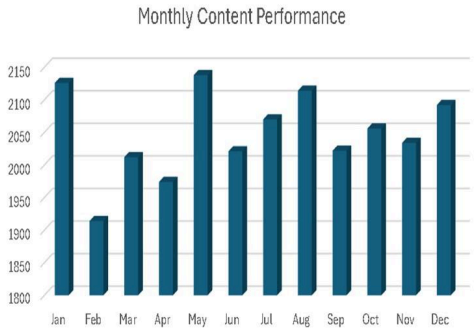
# Insights


15
Unique Categories
January and May
Peak Activity
Animals
Most Popular Category with Reactions: 74,965

- **Key Takeaways**:
- **Animals** is the top performer, followed by **Science** and **Healthy Eating**.
- **Technology** and **Food** round out the top 5, all showing strong engagement.




- **Peak Activity:** January and May saw the highest levels of activity.
- **Dip in February:** There was a noticeable decrease in activity during February.
- **Stable Months:** Engagement remained consistent around the same level for most months, from June to December.
- **General Trend:** A slight decline in early year, with steady engagement throughout the rest of the year.



# Summary

**Recap of Findings:**

- Animals and Science are the top two performers, and the engagement suggests a growing interest in both educational and lifestyle content.

**Actionable Insights:**

- Focus content strategy on categories like Animals and Science, as they have the highest popularity.
- Consider further engagement strategies for other categories like Technology and Healthy Eating.

    Example: post 30 second videos about healthy eating or technology related topics.

- Conduct another content analysis at a later phase to check on the success of this plan.

# Thank you!

ANY QUESTIONS?