

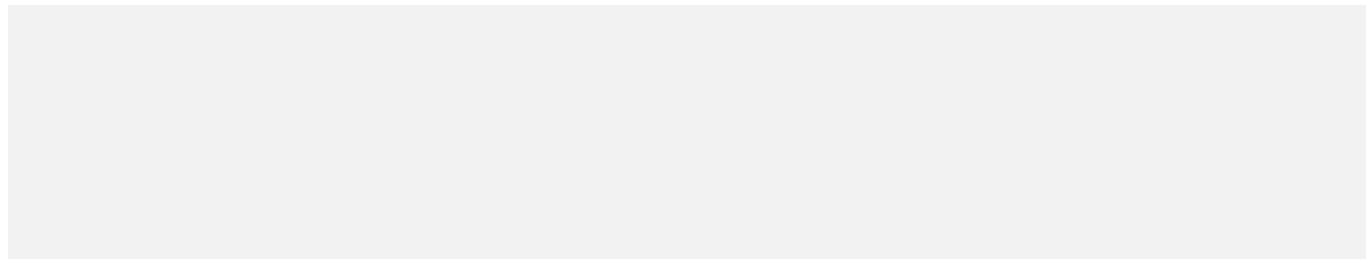
Data Analytics Report

Client Name: Social Buzz
Prepare by: Prateek Chauhan



Data Analytics and Visualization on Social Buzz Project

This data provided by Accenture



Project Understanding

Client name: Social Buzz

Client industry: social media & content creation

Year established: 2010

Location of HQ: San Francisco

Number of employees: 250

Client background:

Social Buzz was founded by two former engineers from a large social media conglomerate, one from London and the other from San Francisco. They left in 2008, and both met in San Francisco to start their business. They started Social Buzz because they saw an opportunity to build on the foundation that their previous company started by creating a new platform where content took center stage. Social Buzz emphasizes content by keeping all users anonymous, only tracking user reactions on every piece of content. There are over 100 ways that users can react to content, spanning beyond the traditional reactions of likes, dislikes, and comments. This ensures that trending content, as opposed to individual users, is at the forefront of user feeds.

Over the past 5 years, Social Buzz has reached over 500 million active users each month. They have scaled quicker than anticipated and need the help of an advisory firm to oversee their scaling process effectively.

Due to their rapid growth and digital nature of their core product, the amount of data that they create, collect and must analyze is huge. Every day over 100,000 pieces of content, ranging from text, images, videos and GIFs are posted. All this data is highly unstructured and requires extremely sophisticated and expensive technology to manage and maintain. Out of the 250 people working at Social Buzz, 200 of them are technical staff working on maintaining this highly complex technology.

Up until this point, they have not relied on any third-party firms to help them get to where they are. However, there are 3 main reasons why they are now looking at bringing in external expertise:

- 1) They are looking to complete an IPO by the end of next year and need guidance to ensure that this goes smoothly.
- 2) They are still a small company and do not have the resources to manage the scale that they are currently at. They could hire more people, but they want an experienced practice to help instead.
- 3) They want to learn data best practices from a large corporation. Due to the nature of their business, they have a massive amount of data, so they are keen on understanding how the world's biggest companies manage the challenges of big data.

To start our engagement with Social Buzz, we are running a 3 month initial project in order to prove to them that we are the best firm to work with. They are expecting the following:

- An audit of their big data practice
- Recommendations for a successful IPO
- An analysis of their content categories that highlights the top 5 categories with the largest aggregate popularity

Tasks to be delegated:

- Creation of an up-to-date big data best practices presentation
- Extraction of sample data sets using SQL
- On-site audit of their data-center
- Merging of sample data set tables

-
- Virtual session with Social Buzz team to present previous client success stories relevant to them
 - Preparation of best practice document for IPO
 - Loading of sample data sets into Accenture sandbox database
 - Technology architecture workshop with Social Buzz Data Team to understand their technology landscape
 - Stress testing of their technology to identify weak spots
 - Communication with previous IPO companies within our client base for reference stories
 - Analysis of sample data sets with visualizations
 - Full documentation of the process that we can guide them through for IPO

Outcome of Problem Statement

The client has reached a massive scale within recent years and does not have the resources internally to handle it.

Audit of big data practice, recommendations for IPO, analysis of popular content

Role of Data Analyst

Analysis of sample data sets with visualizations to understand the popularity of different content categories

- Identify which datasets will be required to answer the client's business question
- Clean the datasets and merge them to prepare the data for analysis
- Determine the answer to the client's business question

Requirements gathering

User

ID: Unique ID of the user (automatically generated)

Name: Full name of user

Email: Email address of user

Profile

User ID: Unique ID of a user that exists in the User table

Interests: Interests of the associated user

Age: Age of the associated user

Location

User ID: Unique ID of a user that exists in the User table

Address: Full address of the user

Session

User ID: Unique ID of a user that exists in the User table

Device: Mobile device that they used for this session on the application

Duration: Amount of time in minutes that this user stayed active on the application during this session

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

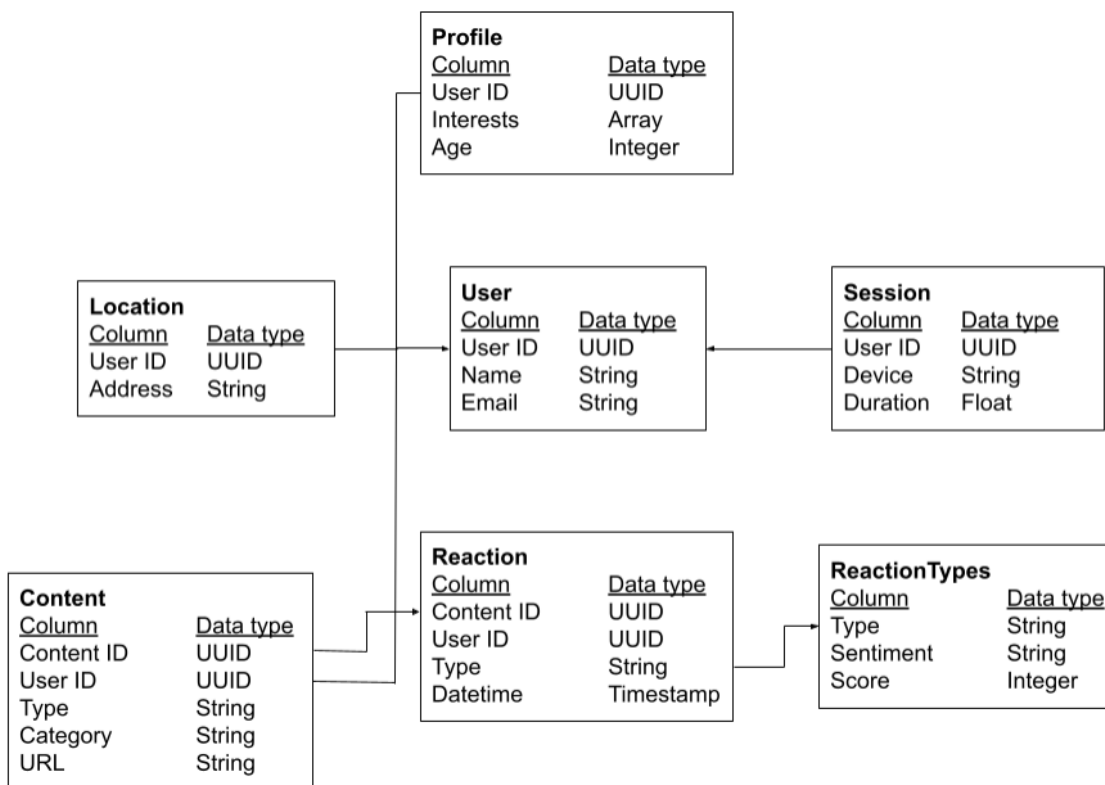
User ID: Unique ID of a user that exists in the User table who reacted to this piece of content

Type: A string detailing the type of reaction this user gave Datetime: The date and time of this reaction

Reaction Types

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.



Here are 3 possible data sets are require completing the Data Analysis, based on the problem statement

Data set (Reaction, Content, Reaction Types) as our relevant data sets.

Why?

- It states that the client wanted to see “An analysis of their content categories showing the top 5 categories with the largest popularity”.
- As explained in the data model, popularity is quantified by the “Score” given to each reaction type.
- We therefore need data showing the content ID, category, content type, reaction type, and reaction score.
- So, to figure out popularity, we’ll have to add up which content categories have the largest score.

Data Cleaning

- removing rows that have values which are missing,
- changing the data type of some values within a column, and
- removing columns which are not relevant to this task.

Think about how each column might be relevant to the business question you’re investigating. If you can’t think of why a column may be useful, it may not be worth including it.

The image displays three screenshots of Microsoft Excel spreadsheets, each showing a different data table. The first screenshot shows the 'Content' table with columns for Content ID, Content Type, Category, and Score. The second screenshot shows the 'ReactionTypes' table with columns for Type, Sentiment, and Score. The third screenshot shows the 'Reactions' table with columns for Content ID, Type, Datetime, and Score.

Content ID	Content Type	Category	Score
0 97522e57-d9ab-4bd6-97bf-c24b952b	photo	Studying	60
1 9f737e0a-3cd4-4d29-9d24-7534e3b	photo	healthy eating	70
2 230c4e4d-70c3-461d-b42c-ec93936	photo	healthy eating	0
3 356f890-d4d4-4785-9483-bc1261031	photo	technology	5
4 01ab84d4-6364-4236-abbb-3c237db	video	food	30
5 c1a6c1a-23ab-4d26-998b-02b1b55	GIF	cooking	20
6 38590c7-6ab2-4973-805a-90cdec35	video	dogs	65
7 e5490118-90a5-4572-ab1c-1fb8c7bf	video	technology	75
8 0bedca96-b76-4287-a83c-1730fde	photo	soccer	70
9 b18cb6f1-4d8e-44ee-a47f-541e9519	photo	public speaking	72
10 5118e9c5-1377-4cc5-a486-6b6b5b71	GIF	science	50
11 46fb701d-6c26-458e-ada3-2eb95db	audio	public speaking	10
12 0ba69b76-d78c-499c-9e0b-a9baf71	GIF	tennis	45
13 81ab855a-3b76-4574-a0a7-d9eb7f1	video	food	35
14 e5ee244-9392-49a9-8c8f-fa54aaa	audio	travel	15
15 7f6da82-4a0-4527-a4d6-e251b75e	GIF	fitness	12
16 f532c062-dc48-46c2-a64b-641157c	GIF	soccer	
17 a2e93b29-9259-4092-a4d7-62d5e82	video	public speaking	
18 9b3a6c90-48a8-478c-82be-90315241	video	food	
19 850f90d-47b6-4eb7-b9c9-810atcb	video	tennis	
20 ba2db09-81ba-4c3d-4176-d9950b	GIF	soccer	
21 98d9d71-cx1d-4f6c-bb05-a05779c	GIF	soccer	
22 388bd9b-9d10-4147-87d-6db46e	video	education	
23 78d00751-895c-4a15-a35c-a921e2bf	photo	studying	
24 bda0b065-78bb-4d52-b7a9-fcc375cc	photo	travel	
25 a06f7b5-3ba8-49f7-87c2-d4ad5e4f	GIF	food	
26 8099a1e3-7a8f-4e67-b9c9-810atcb	audio	dogs	
27 3a4c821a-9a77-4c26-9a39-fc1c8b9b	GIF	soccer	
28 2920dccb-a9e4-49fc-8049-b6a4164d	GIF	veganism	
29 4e3840a-bb71-40c2-a1c5-9408b842	photo	healthy eating	
30 a2f9b02-1fa6-4001-9566-f97b78f5	video	food	
31 74718bfb-e5a3-442f-8ca2-4a342875	video	dogs	
32 89d88b9-807f-4076-b7e7-84db361e1	audio	studying	
33 02fa1a4f-722d-46d2-a85c-e5b136b5	video	cooking	
34 eac5621-4c2b-4c8a-93a9-d5448039	audio	food	
35 978ba73-aef2-4566-8b5b-41be448bf	GIF	Animals	

Type	Sentiment	Score
0 heart	positive	60
1 want	positive	70
2 disgust	negative	0
3 hate	negative	5
4 interested	positive	30
5 indifference	neutral	20
6 love	positive	65
7 super love	positive	75
8 cherish	positive	70
9 adore	positive	72
10 like	positive	50
11 dislike	negative	10
12 intrigued	positive	45
13 peeking	neutral	35
14 scared	negative	15
15 worried	negative	12

Content ID	Type	Datetime	Score
1 97522e57-d9ab-4bd6-97bf-c24b952b	dislike	07-11-2020 09:43	
2 97522e57-d9ab-4bd6-97bf-c24b952b	dislike	17-06-2021 12:22	
3 97522e57-d9ab-4bd6-97bf-c24b952b	scared	18-04-2021 05:13	
4 97522e57-d9ab-4bd6-97bf-c24b952b	dislike	06-01-2021 19:13	
5 97522e57-d9ab-4bd6-97bf-c24b952b	interested	23-08-2020 12:25	
6 97522e57-d9ab-4bd6-97bf-c24b952b	peeking	07-12-2020 06:27	
7 97522e57-d9ab-4bd6-97bf-c24b952b	cherish	11-04-2021 17:35	
8 97522e57-d9ab-4bd6-97bf-c24b952b	hate	27-01-2021 08:32	
9 97522e57-d9ab-4bd6-97bf-c24b952b	peeking	01-04-2021 22:54	
10 97522e57-d9ab-4bd6-97bf-c24b952b	love	04-08-2020 05:05	
11 97522e57-d9ab-4bd6-97bf-c24b952b	indifferent	07-11-2020 08:36	
12 97522e57-d9ab-4bd6-97bf-c24b952b	scared	02-11-2020 06:28	
13 97522e57-d9ab-4bd6-97bf-c24b952b	interested	01-11-2020 01:16	
14 97522e57-d9ab-4bd6-97bf-c24b952b	hate	07-10-2020 18:39	
15 97522e57-d9ab-4bd6-97bf-c24b952b	scared	03-09-2020 05:51	
16 97522e57-d9ab-4bd6-97bf-c24b952b	super love	23-09-2021 05:09	
17 97522e57-d9ab-4bd6-97bf-c24b952b	peeking	24-02-2021 11:37	
18 97522e57-d9ab-4bd6-97bf-c24b952b	indifferent	24-02-2021 11:37	
19 97522e57-d9ab-4bd6-97bf-c24b952b	interested	22-05-2021 19:44	
20 97522e57-d9ab-4bd6-97bf-c24b952b	intrigued	31-01-2021 16:03	
21 97522e57-d9ab-4bd6-97bf-c24b952b	peeking	20-11-2020 17:26	
22 97522e57-d9ab-4bd6-97bf-c24b952b	worried	11-04-2021 20:47	
23 97522e57-d9ab-4bd6-97bf-c24b952b	like	13-06-2021 16:46	
24 97522e57-d9ab-4bd6-97bf-c24b952b	heart	11-04-2021 14:29	
25 97522e57-d9ab-4bd6-97bf-c24b952b	worried	02-03-2021 19:21	
26 97522e57-d9ab-4bd6-97bf-c24b952b	hate	27-06-2020 18:44	
27 97522e57-d9ab-4bd6-97bf-c24b952b	hate	09-05-2021 21:05	
28 97522e57-d9ab-4bd6-97bf-c24b952b	cherish	04-08-2020 23:58	
29 97522e57-d9ab-4bd6-97bf-c24b952b	super love	28-08-2020 03:57	
30 97522e57-d9ab-4bd6-97bf-c24b952b	intrigued	25-08-2020 08:21	
31 97522e57-d9ab-4bd6-97bf-c24b952b	scared	20-04-2021 23:48	
32 97522e57-d9ab-4bd6-97bf-c24b952b	interested	07-03-2021 12:40	
33 97522e57-d9ab-4bd6-97bf-c24b952b	indifferent	24-05-2021 03:21	
34 97522e57-d9ab-4bd6-97bf-c24b952b	super love	24-01-2021 11:03	
35 97522e57-d9ab-4bd6-97bf-c24b952b	dislike	09-11-2020 12:46	
36 97522e57-d9ab-4bd6-97bf-c24b952b	super love	02-02-2021 03:43	

Data Modelling

- Use Reaction table as your base table, then first join the relevant columns from your Content data set, and then the Reaction Types data set.

M8										
	A	B	C	D	E	F	G	H	I	J
	Content ID	Reaction Type	Datetime	Content Typ	Category	Sentiment	Score			
1	1 97522e57-d9ab-4bd6-97bf-c24d952602d2	disgust	7/11/20 9:43	photo	Studying	negative	0			
2	2 97522e57-d9ab-4bd6-97bf-c24d952602d2	dislike	17/6/21 12:22	photo	Studying	negative	10			
3	3 97522e57-d9ab-4bd6-97bf-c24d952602d2	scared	18/4/21 5:13	photo	Studying	negative	15			
4	4 97522e57-d9ab-4bd6-97bf-c24d952602d2	disgust	6/1/21 19:13	photo	Studying	negative	0			
5	5 97522e57-d9ab-4bd6-97bf-c24d952602d2	interested	23/8/20 12:25	photo	Studying	positive	30			
6	6 97522e57-d9ab-4bd6-97bf-c24d952602d2	peeking	7/12/20 6:27	photo	Studying	neutral	35			
7	7 97522e57-d9ab-4bd6-97bf-c24d952602d2	cherish	11/4/21 17:35	photo	Studying	positive	70			
8	8 97522e57-d9ab-4bd6-97bf-c24d952602d2	hate	27/1/21 8:32	photo	Studying	negative	5			
9	9 97522e57-d9ab-4bd6-97bf-c24d952602d2	peeking	1/4/21 22:54	photo	Studying	neutral	35			
10	10 97522e57-d9ab-4bd6-97bf-c24d952602d2	love	4/8/20 5:05	photo	Studying	positive	65			
11	11 97522e57-d9ab-4bd6-97bf-c24d952602d2	indifferent	7/11/20 8:36	photo	Studying	neutral	20			
12	12 97522e57-d9ab-4bd6-97bf-c24d952602d2	scared	2/11/20 6:28	photo	Studying	negative	15			
13	13 97522e57-d9ab-4bd6-97bf-c24d952602d2	interested	1/11/20 1:16	photo	Studying	positive	30			
14	14 97522e57-d9ab-4bd6-97bf-c24d952602d2	hate	7/10/20 18:39	photo	Studying	negative	5			
15	15 97522e57-d9ab-4bd6-97bf-c24d952602d2	scared	3/9/20 18:51	photo	Studying	negative	15			
16	16 97522e57-d9ab-4bd6-97bf-c24d952602d2	super love	24/2/21 5:09	photo	Studying	positive	75			
17	17 97522e57-d9ab-4bd6-97bf-c24d952602d2	peeking	23/9/20 6:24	photo	Studying	neutral	35			
18	18 97522e57-d9ab-4bd6-97bf-c24d952602d2	indifferent	24/2/21 11:37	photo	Studying	neutral	20			
19	19 97522e57-d9ab-4bd6-97bf-c24d952602d2	interested	22/5/21 19:44	photo	Studying	positive	30			
20	20 97522e57-d9ab-4bd6-97bf-c24d952602d2	intrigued	31/1/21 16:03	photo	Studying	positive	45			
21	21 97522e57-d9ab-4bd6-97bf-c24d952602d2	peeking	20/11/20 17:26	photo	Studying	neutral	35			
22	22 97522e57-d9ab-4bd6-97bf-c24d952602d2	worried	11/4/21 20:47	photo	Studying	negative	12			
23	23 97522e57-d9ab-4bd6-97bf-c24d952602d2	like	13/6/21 16:46	photo	Studying	positive	50			
24	24 97522e57-d9ab-4bd6-97bf-c24d952602d2	heart	11/4/21 14:29	photo	Studying	positive	60			
25	25 97522e57-d9ab-4bd6-97bf-c24d952602d2	worried	2/3/21 19:21	photo	Studying	negative	12			
26	26 97522e57-d9ab-4bd6-97bf-c24d952602d2	hate	27/6/20 18:44	photo	Studying	negative	5			
27	27 97522e57-d9ab-4bd6-97bf-c24d952602d2	hate	9/5/21 21:05	photo	Studying	negative	5			
28	28 97522e57-d9ab-4bd6-97bf-c24d952602d2	cherish	4/8/20 23:58	photo	Studying	positive	70			
29	29 97522e57-d9ab-4bd6-97bf-c24d952602d2	super love	28/8/20 3:57	photo	Studying	positive	75			
30	30 97522e57-d9ab-4bd6-97bf-c24d952602d2	intrigued	25/8/20 8:21	photo	Studying	positive	45			
31	31 97522e57-d9ab-4bd6-97bf-c24d952602d2	scared	20/4/21 23:48	photo	Studying	negative	15			
32	32 97522e57-d9ab-4bd6-97bf-c24d952602d2	interested	7/3/21 12:40	photo	Studying	positive	30			
33	33 97522e57-d9ab-4bd6-97bf-c24d952602d2	indifferent	24/5/21 3:21	photo	Studying	neutral	20			
34	34 97522e57-d9ab-4bd6-97bf-c24d952602d2	super love	24/1/21 11:03	photo	Studying	positive	75			
35	35 97522e57-d9ab-4bd6-97bf-c24d952602d2	disgust	9/4/21 2:46	photo	Studying	negative	0			
36	36 97522e57-d9ab-4bd6-97bf-c24d952602d2	super love	2/2/21 3:43	photo	Studying	positive	75			
37	37 97522e57-d9ab-4bd6-97bf-c24d952602d2	interested	4/2/21 3:06	photo	Studying	positive	30			
38	38 97522e57-d9ab-4bd6-97bf-c24d952602d2	dislike	9/11/20 2:49	photo	Studying	negative	10			
39	39 97522e57-d9ab-4bd6-97bf-c24d952602d2	scared	29/9/20 16:00	photo	Studying	negative	15			
40	40 97522e57-d9ab-4bd6-97bf-c24d952602d2	dislike	29/11/20 21:33	photo	Studying	negative	10			
41	41 97522e57-d9ab-4bd6-97bf-c24d952602d2	want	13/2/21 7:54	photo	Studying	positive	70			
42	42 97522e57-d9ab-4bd6-97bf-c24d952602d2	interested	14/7/20 5:28	photo	Studying	positive	30			
43	43 97522e57-d9ab-4bd6-97bf-c24d952602d2	indifferent	23/8/20 10:30	photo	Studying	neutral	20			
44	44 97522e57-d9ab-4bd6-97bf-c24d952602d2	worried	10/5/21 9:27	photo	Studying	negative	12			
45	45 97522e57-d9ab-4bd6-97bf-c24d952602d2	hate	15/3/21 3:59	photo	Studying	negative	5			
46	46 97f737e0a-3cdd-4d29-9d24-753f4e3be810	dislike	25/6/20 17:01	photo	healthy eatin	negative	10			
47										

- Top 5 performing categories
- Add up the total scores for each category.

F18					
	A	B	C	D	E
1	Category	Aggregate Score			
2	Animals	74965			
3	science	71168			
4	healthy eating	69339			
5	technology	68738			
6	food	66676			
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					
22					
23					
24					
25					
26					
27					
28					
29					
30					
31					
32					
33					
34					
35					
36					
37					
38					
39					
40					
41					
42					

Power Point Presentation

Here in this Project

For each slide, think about:

- Agenda - What will your presentation cover?
- Project Recap - What are the key points from the brief?
- Problem - What is the problem that you answer in this presentation?
- The Analytics team - Who is on your team?
 - As a reminder from the earlier task - this includes: Andrew Fleming (Chief Technical Architect), Marcus Romp ton (Senior Principle), and Prateek Chauhan!
- Process - How did you complete your analysis?
- Link for Presentation: -
https://drive.google.com/file/d/1ycvhV6GbBmO5k63oIYoc_yF_8dyUzIK7/view?usp=sharing

Shot Example of Presentation

16

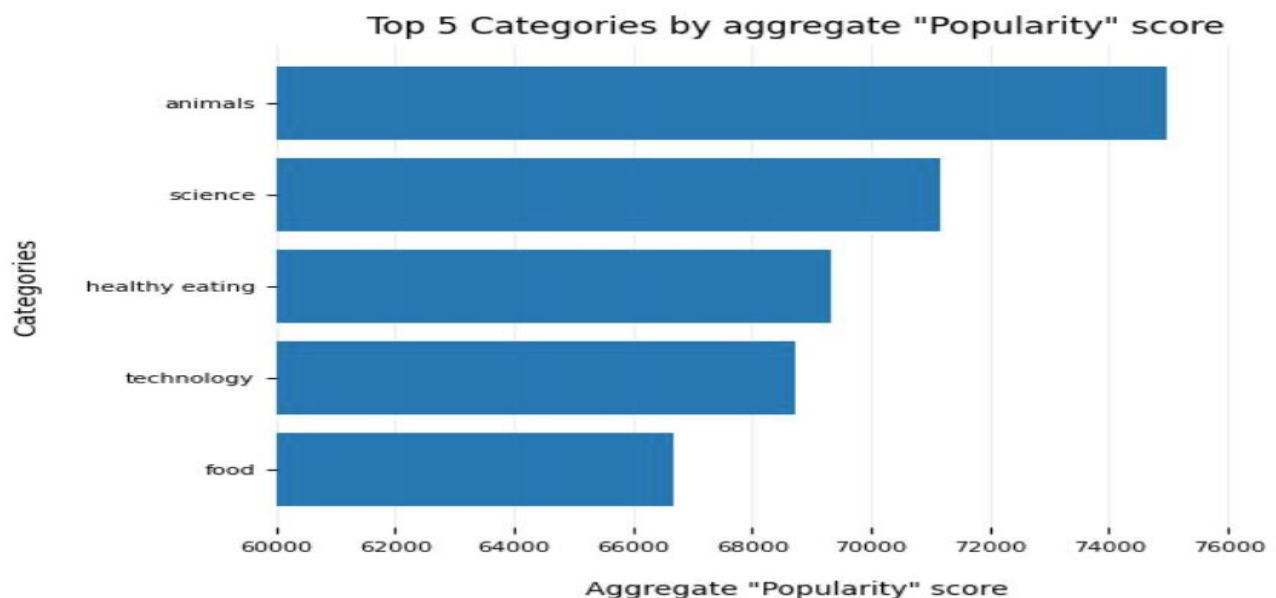
1897

JANUARY

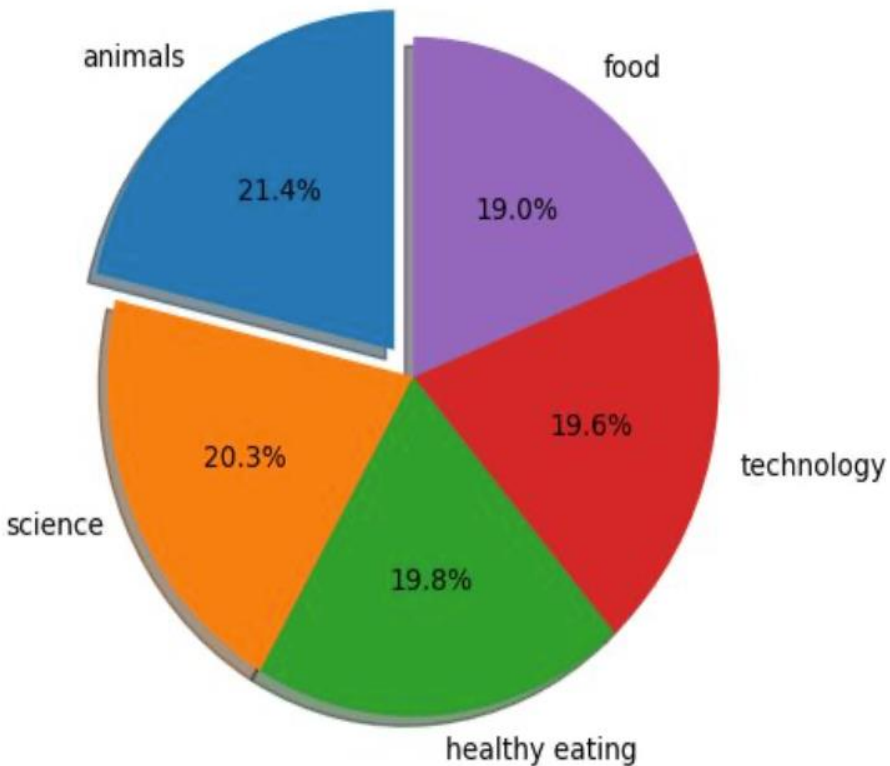
UNIQUE
CATEGORIES

REACTIONS TO "ANIMAL"
POSTS

MONTH WITH
MOST POSTS



Popularity percentage share from top 5 categories



Analysis

Animals and science are the two most popular categories of content, showing that people enjoy “real-life” and “factual” content the most.

Insight

Food is a common theme with the top 5 categories with “healthy eating” ranking the highest. This may give an indication to the audience within your user base. You could use this insight to create a campaign and work with healthy eating brands to boost user engagement.