



Social Buzz Data Analysis

Presented by: Aditi Teli



About the Company

Social Buzz, established in 2008, is a pioneering social media platform with a unique approach. It recognizes the significance of content creation and was founded to support content creators.

Social Buzz strongly emphasizes content by prioritizing anonymity for its users and exclusively tracking reactions to each piece of content.

Unlike conventional platforms, Social Buzz offers users more than 100 diverse ways to react to content. This ensures that the focus remains on trending and engaging content rather than individual user profiles, making it a distinctive and innovative platform in the world of social media.

Problem Statement

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 the rapid growth and digital nature of their core product, the amount of data 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 of 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.

Task Delegated

In my role as a Data Analyst, I am tasked with the responsibility of conducting a Ad-hoc analysis of the data contained in three separate CSV files: one related to content, another focusing on reactions, and the third pertaining to reaction types.

These datasets collectively provide a lot of information, including details about various content categories, the types of reactions garnered by specific content pieces, sentiment analysis, score-related data, user preferences for different categories, and more.

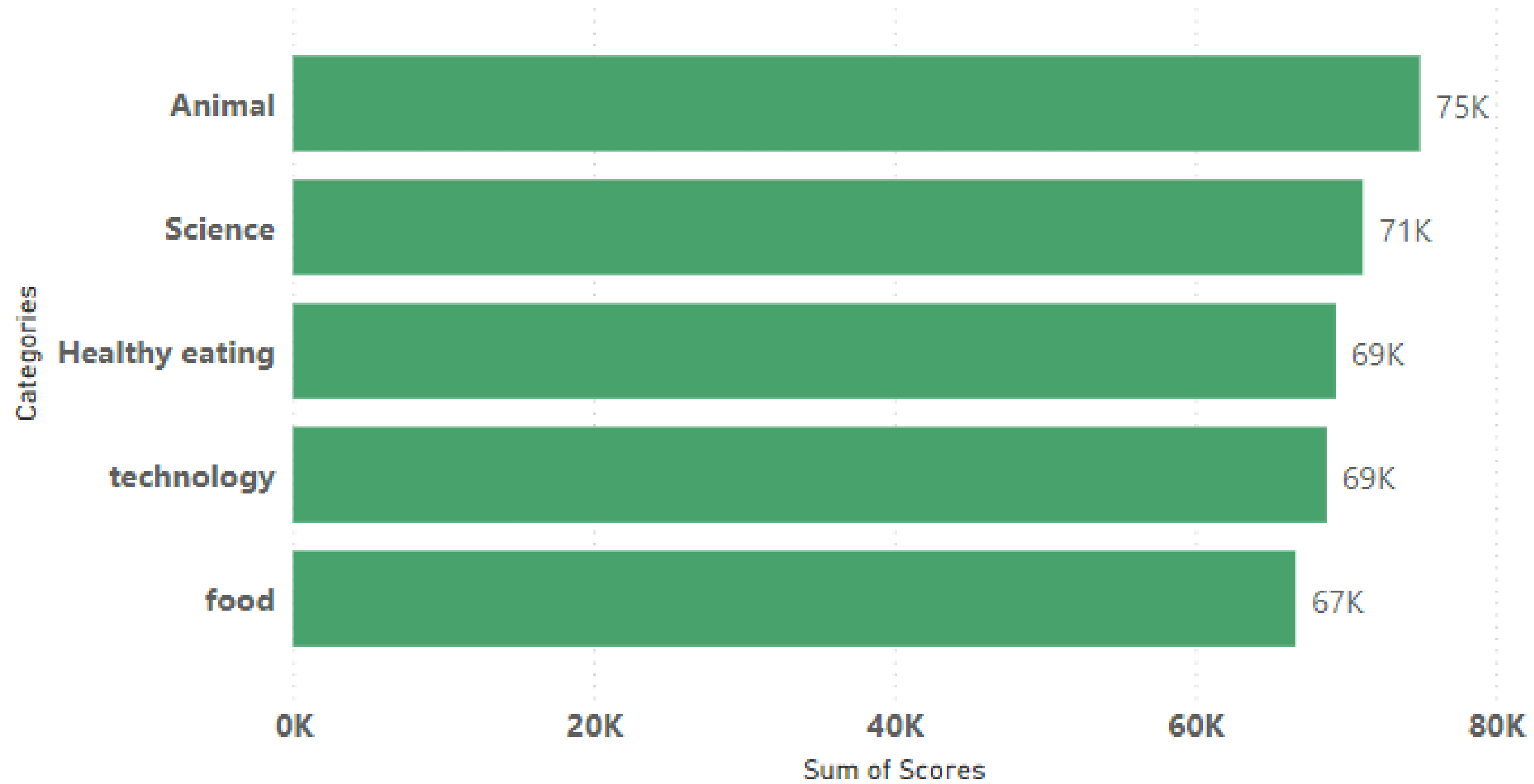
My objective is to extract meaningful insights and patterns from this data to inform decision-making and provide valuable insights to our organization.

Ad-hoc Analysis

REQUEST 1: The top 5 categories with highest scores

```
SELECT
    Category, SUM(score) AS scores
FROM
    content
JOIN reactions
ON content.content_id = reactions.content_id
JOIN reactiontypes
ON reactions.Reaction_Type = reactiontypes.Type
GROUP BY Category
ORDER BY scores DESC
LIMIT 5;
```

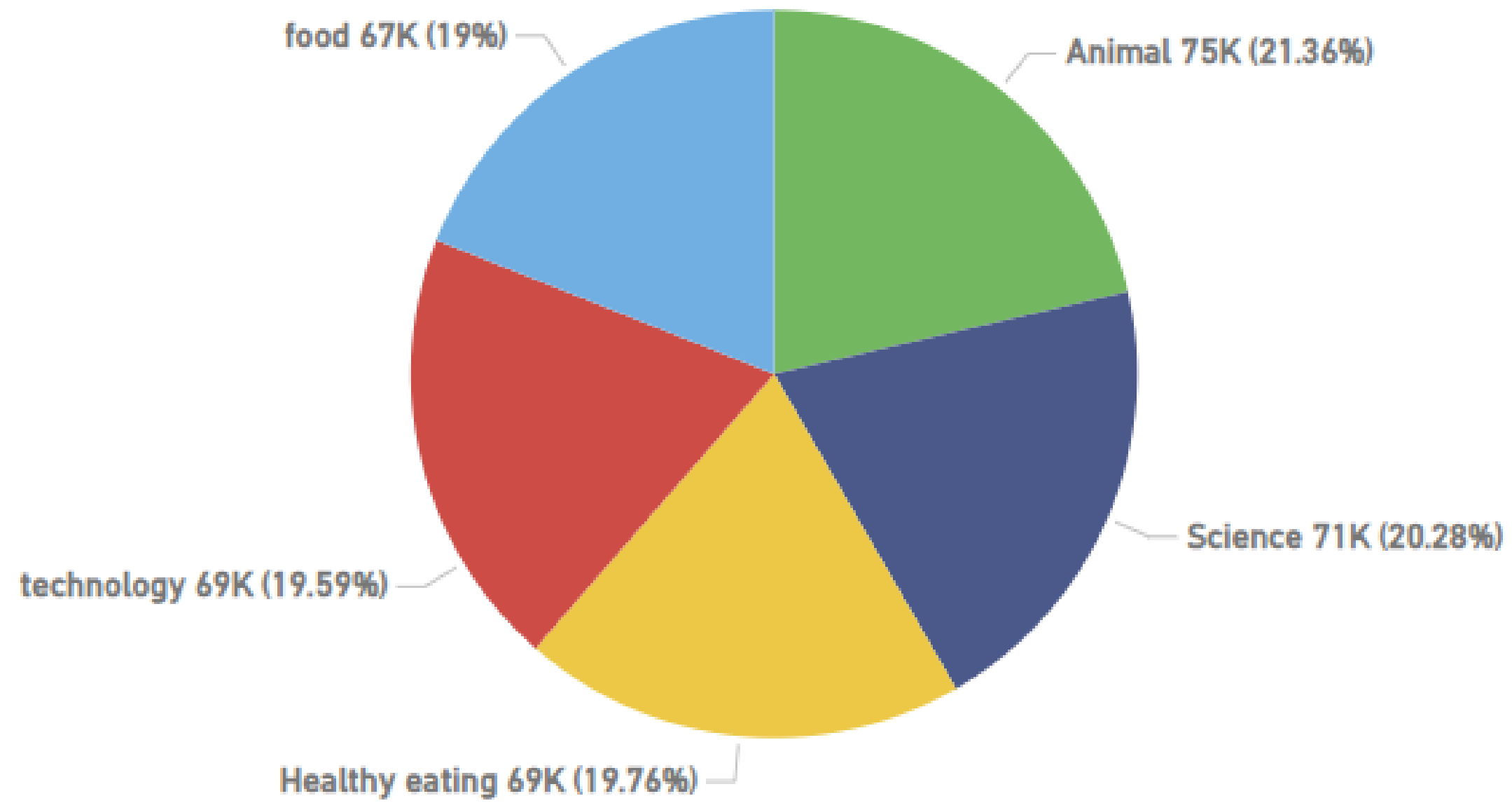
Category	scores
Animals	74965
science	71168
healthy eating	69339
technology	68738
food	66676



REQUEST 2: The top 5 categories with highest percentage share

```
WITH CategoryScores AS (  
    SELECT  
        Category,  
        SUM(score) AS scores  
    FROM  
        content  
    JOIN  
        reactions ON content.content_id = reactions.content_id  
    JOIN  
        reactiontypes ON reactions.Reaction_Type = reactiontypes.Type  
    GROUP BY  
        Category  
    ORDER BY  
        scores DESC  
    LIMIT 5  
)  
  
SELECT  
    cs.Category,  
    cs.scores,  
    SUM(cs.scores) OVER () AS total_score,  
    (cs.scores / SUM(cs.scores) OVER ()) * 100 AS percentage_contribution  
FROM  
    CategoryScores cs;
```

Category	scores	total_score	percentage_contribution
Animals	74965	350886	21.3645
science	71168	350886	20.2824
healthy eating	69339	350886	19.7611
technology	68738	350886	19.5898
food	66676	350886	19.0022



REQUEST 3: Unique Categories in the data



```
SELECT DISTINCT category FROM content;
```

category
Studying
healthy eating
technology
food
cooking
dogs
soccer
public speaking
science
tennis
travel
fitness
education
veganism
Animals
culture

REQUEST 4: How many reactions are there for most popular category

```
SELECT category,COUNT(reaction_type) AS category_count
FROM reactions
JOIN content
USING (content_id)
GROUP BY Category
ORDER BY category_count DESC;
```

The most popular category is the “Animals” category with 1897 reactions.

category	category_count
Animals	1897
science	1796
healthy eating	1717
food	1699
technology	1698
culture	1676
cooking	1664
travel	1647
soccer	1457
education	1433
fitness	1395
Studying	1363
dogs	1338
tennis	1328
veganism	1248
public speaking	1217

REQUEST 5: Most post generated month

```
SELECT
    MONTH(datetime),COUNT(content_id) AS content_count
FROM reactions
GROUP BY MONTH(datetime)
ORDER BY content_count DESC;
```

In the month of May there were most of the post generated .

	MONTH(datetime)	content_count
	5	2138
	1	2126
	8	2114
	12	2092
	7	2070
	10	2056
	11	2034
	9	2022
	6	2021
	3	2012
	4	1974
	2	1914

Insights Generated

- There are 16 unique categories in the dataset
- Animals and science are the two most popular categories of content, showing that people enjoy “real life” and “factual content” the most.
- Food is a common theme with the top 5 categories “Healthy eating” ranking the highest. This gives an indication that we can work with healthy food brands to boost the engagement.
- Animal category shares the highest percentage (21 %) among the total 5 categories.

Thank You

- *Virtual Experience Programme by Accenture North America*



Aditi Teli