SOCIAL BUZZ DATA ANALYSIS

Prepared by : Yash Dabhi

❑ About Social Buzz

- Client name: Social Buzz
- Client industry: Social media & content creation
- Year established: 2010
- Location of HQ: San Francisco
- Requirement:- An analysis of their content categories that highlights the top 5 categories with the largest aggregate popularity

❑ Problem Statement

- Over 100000 posts per day.
- 36,500,000 piece of content every year.
- Analysis to find the most popular categories

❑ Data preprocessing

❖ **Data Cleaning**

Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset.

- In this dataset we cleared unnecessary data.
- Set the data in proper format.
- Remove duplicates.

❑ Used some Excel formulas for discovered Top 5 categories

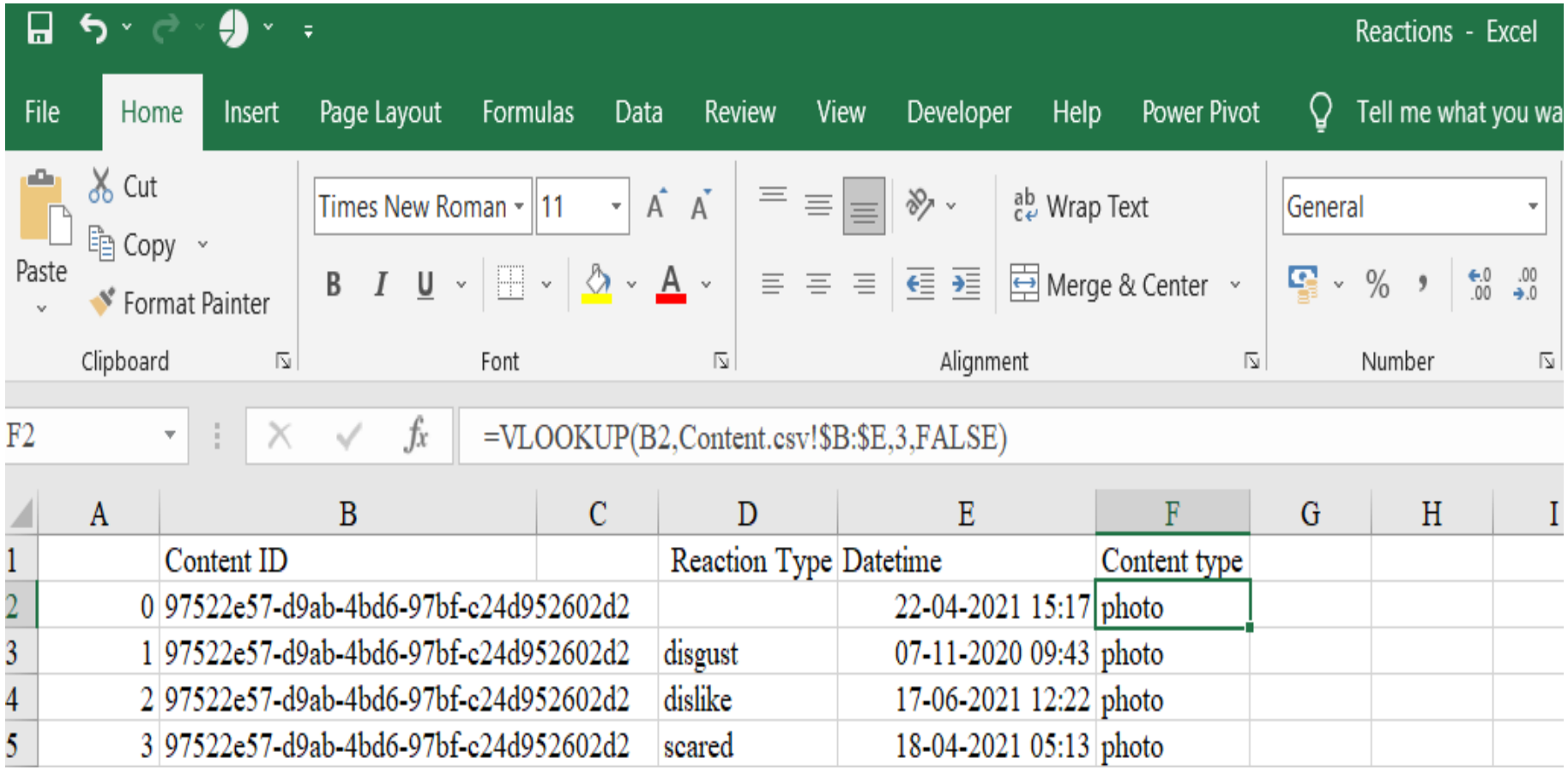
- First formula V LOOKUP:
- Used V LOOKUP formula for combined three different excel files based on common columns between them.
- Three files are:

i. Content (User_id , Content_id , Type , Category)

ii. Reaction (Reactions type , Content_id)

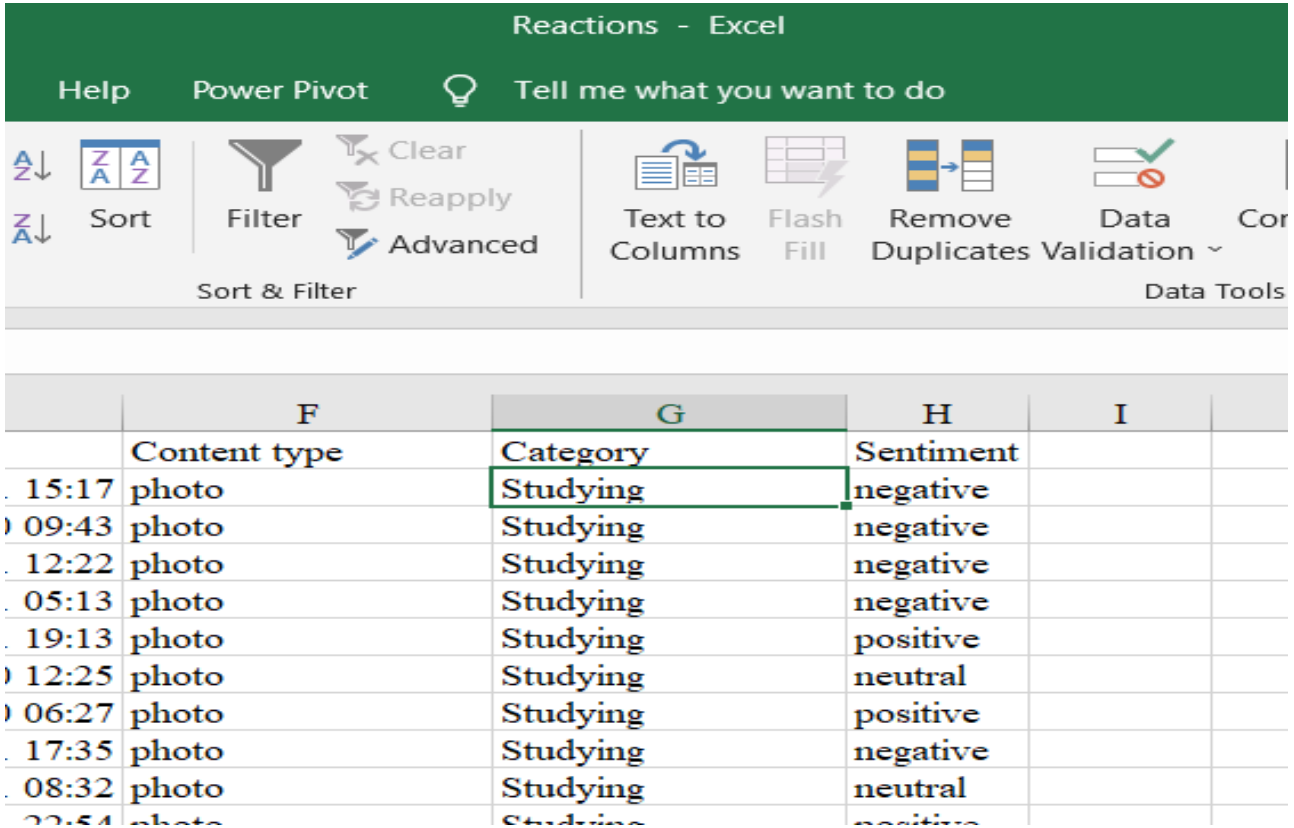
iii. Reaction type (Reaction type , Sentiment , Score)

- So first combine Content file and Reaction file based on content_id column & find content type in Reaction file.



	A	B	C	D	E	F	G	H	I
1		Content ID		Reaction Type	Datetime	Content type			
2	0	97522e57-d9ab-4bd6-97bf-c24d952602d2			22-04-2021 15:17	photo			
3	1	97522e57-d9ab-4bd6-97bf-c24d952602d2		disgust	07-11-2020 09:43	photo			
4	2	97522e57-d9ab-4bd6-97bf-c24d952602d2		dislike	17-06-2021 12:22	photo			
5	3	97522e57-d9ab-4bd6-97bf-c24d952602d2		scared	18-04-2021 05:13	photo			

❑ Same for Category and Sentiment



	F	G	H	I
	Content type	Category	Sentiment	
15:17	photo	Studying	negative	
09:43	photo	Studying	negative	
12:22	photo	Studying	negative	
05:13	photo	Studying	negative	
19:13	photo	Studying	positive	
12:25	photo	Studying	neutral	
06:27	photo	Studying	positive	
17:35	photo	Studying	negative	
08:32	photo	Studying	neutral	

❑ Top 5 Categories

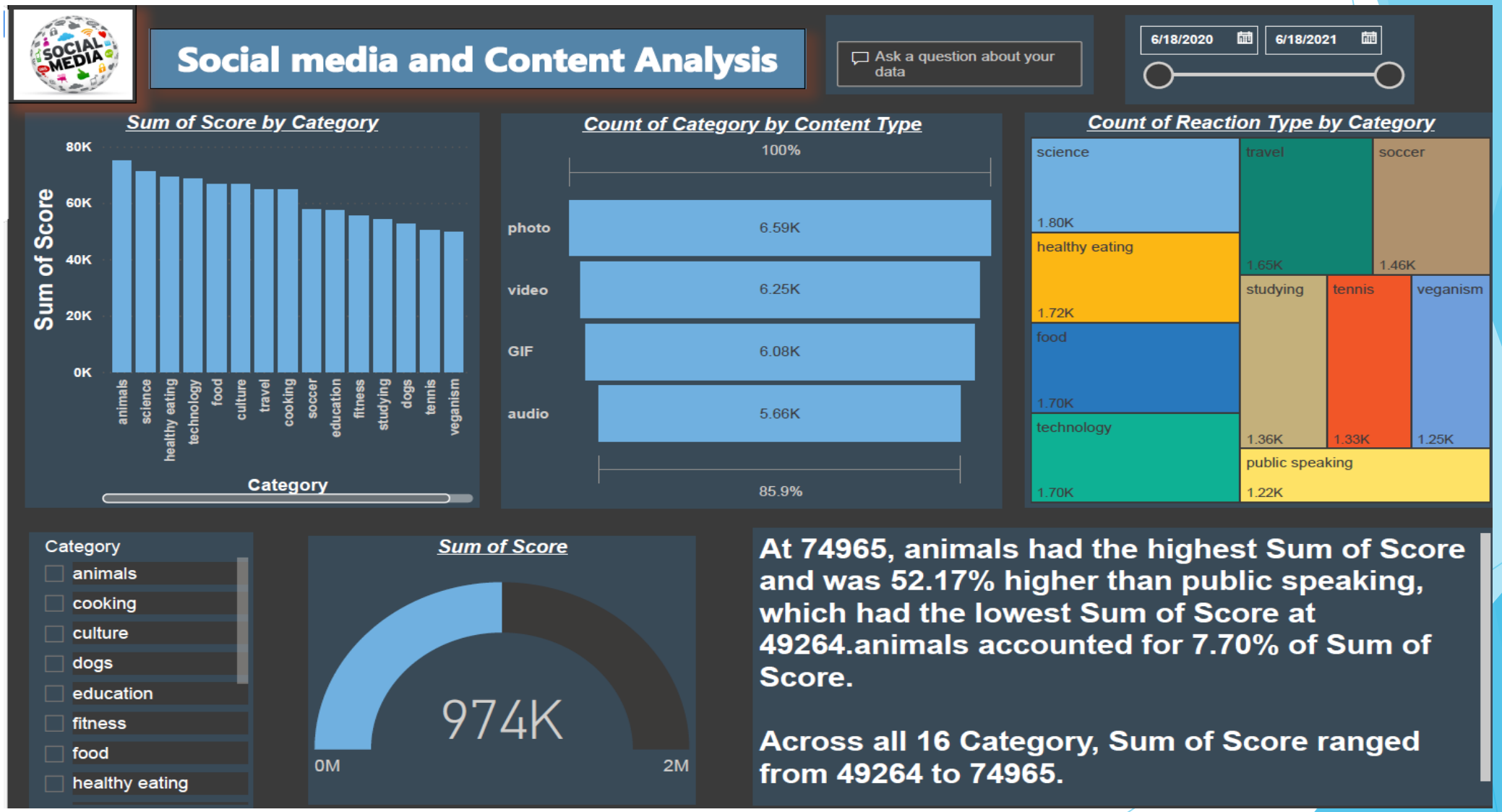
- Discovered top 5 categories with the help of SUMIF formula

B2 | =SUMIF(Reactions!G:G,Sheet1!A2,Reactions!I:I)

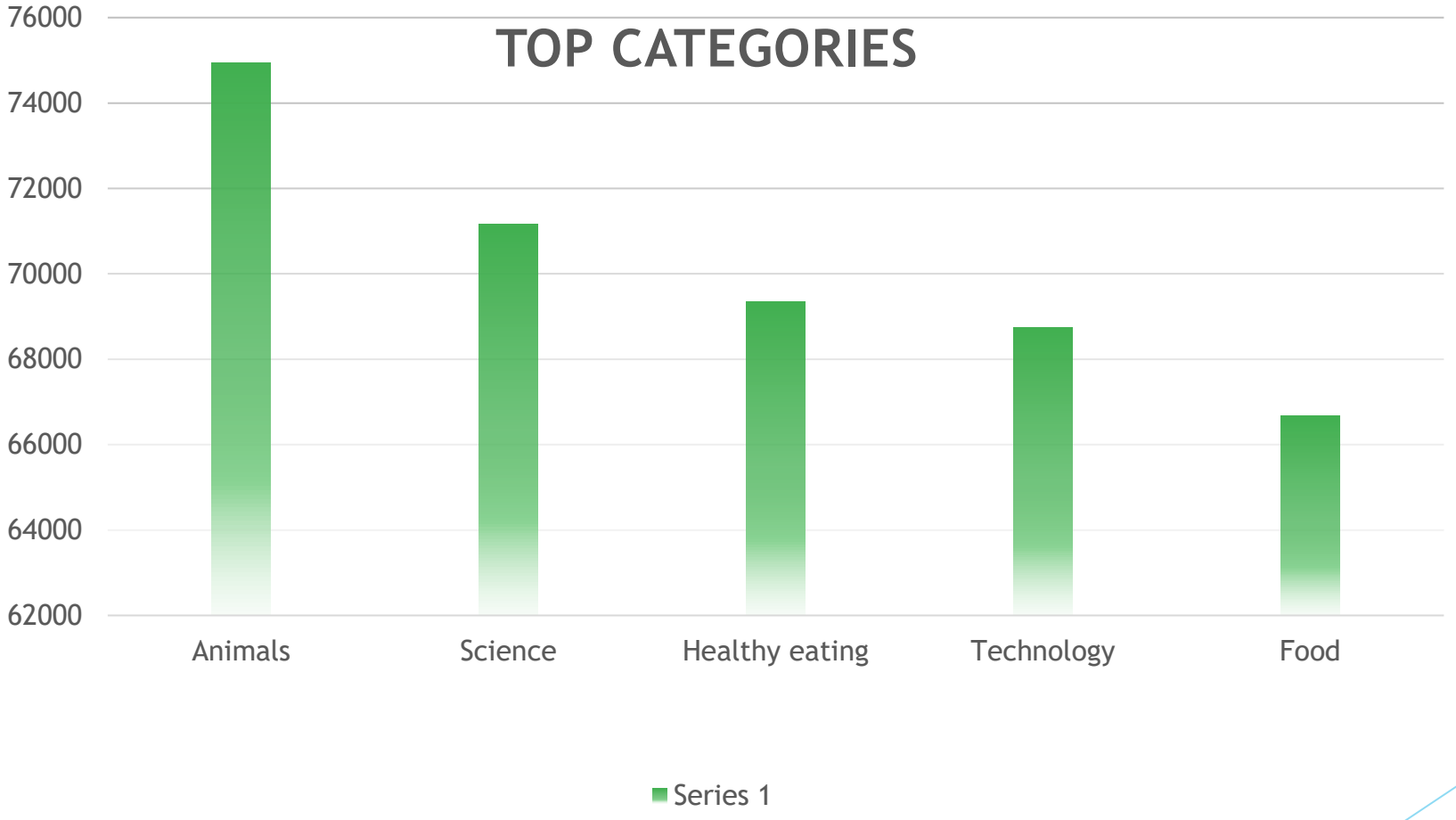
	A	B	C	D	E	F
1	Category	Score				
2	Studying	54269				
3	healthy eating	69339				
4	technology	68738				
5	food	66676				
6	cooking	64756				
7	dogs	52511				
8	soccer	57783				
9	public speaking	49264				
10	science	71168				
11	tennis	50339				
12	travel	64880				
13	fitness	55323				
14	education	57436				
15	veganism	49619				
16	Animals	74965				
17	culture	66579				

	A	B
	Top 5	Score
	Animals	74965
	science	71168
	healthy eating	69339
	technology	68738
	food	66676

❑ Data Visualization using Power-BI



❑ Visualizing using excel



❑ Some meaningful insights

- There are total 16 unique categories
- Top 5 categories are : Animal , Science , Healthy eating , Technology , Food
- Photos are most posted on Social media
- Science category having a highest reactions
- Across all 16 categories, Sum of score range from 49264 to 74965

Thank you