

Loading the Dataset

```
[1]: import numpy as np
import pandas as pd
```

```
[2]: content = pd.read_csv('Content.csv')
reaction = pd.read_csv('Reactions.csv')
reactiontypes = pd.read_csv('ReactionTypes.csv')
```

```
[3]: content.head()
```

```
[3]:
```

	Unnamed: 0	Content ID	User ID	Type	Category	URL
0	0	97522e57-d9ab-4bd6-97bf-c24d952602d2	8d3cd87d-8a31-4935-9a4f-b319bfe05f31	photo	Studying	https://socialbuzz.cdn.com/content/storage/975...
1	1	9f737e0a-3cdd-4d29-9d24-753f4e3be810	beb1f34e-7870-46d6-9fc7-2e12eb83ce43	photo	healthy eating	https://socialbuzz.cdn.com/content/storage/9f7...
2	2	230c4e4d-70c3-461d-b42c-ec09396efb3f	a5c65404-5894-4b87-82f2-d787cbee86b4	photo	healthy eating	https://socialbuzz.cdn.com/content/storage/230...
3	3	356fff80-da4d-4785-9f43-bc1261031dc6	9fb4ce88-fac1-406c-8544-1a899cee7aaf	photo	technology	https://socialbuzz.cdn.com/content/storage/356...
4	4	01ab84dd-6364-4236-abbb-3f237db77180	e206e31b-5f85-4964-b6ea-d7ee5324def1	video	food	https://socialbuzz.cdn.com/content/storage/01a...

```
[4]: reaction.head()
```

```
[4]:
```

	Unnamed: 0	Content ID	User ID	Type	Datetime
0	0	97522e57-d9ab-4bd6-97bf-c24d952602d2	NaN	NaN	2021-04-22 15:17:15
1	1	97522e57-d9ab-4bd6-97bf-c24d952602d2	5d454588-283d-459d-915d-c48a2cb4c27f	disgust	2020-11-07 09:43:50
2	2	97522e57-d9ab-4bd6-97bf-c24d952602d2	92b87fa5-f271-43e0-af66-84fac21052e6	dislike	2021-06-17 12:22:51
3	3	97522e57-d9ab-4bd6-97bf-c24d952602d2	163daa38-8b77-48c9-9af6-37a6c1447ac2	scared	2021-04-18 05:13:58

Data Cleaning and Modelling

```
[6]: content.isnull().sum()
```

```
[6]: Unnamed: 0      0
     Content ID    0
     User ID      0
     Type         0
     Category     0
     URL         199
     dtype: int64
```

```
[10]: content.drop(['User ID'],axis=1,inplace=True)
```

```
[11]: content
```

```
[11]:
```

	Content ID	Type	Category
0	97522e57-d9ab-4bd6-97bf-c24d952602d2	photo	Studying
1	9f737e0a-3cdd-4d29-9d24-753f4e3be810	photo	healthy eating
2	230c4e4d-70c3-461d-b42c-ec09396efb3f	photo	healthy eating
3	356fff80-da4d-4785-9f43-bc1261031dc6	photo	technology
4	01ab84dd-6364-4236-abb-b3f237db77180	video	food
...
995	b4cef9ef-627b-41d7-a051-5961b0204ebb	video	public speaking
996	7a79f4e4-3b7d-44dc-bdef-bc990740252c	GIF	technology
997	435007a5-6261-4d8b-b0a4-55fdc189754b	audio	veganism
998	4e4c9690-c013-4ee7-9e66-943d8cbd27b7	GIF	culture

jupyter Accenture Task-2 Last Checkpoint: 13 hours ago

File Edit View Run Kernel Settings Help

Trusted

Markdown

JupyterLab Python 3 (ipykernel)

```
[18]: reaction.isnull().sum()
```

```
[18]: Unnamed: 0      0
      Content ID    0
      User ID     3019
      Type         980
      Datetime      0
      dtype: int64
```

```
[19]: reaction.drop(['Unnamed: 0', 'User ID'],axis=1,inplace=True)
```

```
[20]: reaction.isnull().sum()
```

```
[20]: Content ID      0
      Type         980
      Datetime      0
      dtype: int64
```

```
[21]: reaction.dropna(inplace=True)
```

```
[22]: reaction
```

```
[22]:
```

	Content ID	Type	Datetime
1	97522e57-d9ab-4bd6-97bf-c24d952602d2	disgust	2020-11-07 09:43:50
2	97522e57-d9ab-4bd6-97bf-c24d952602d2	dislike	2021-06-17 12:22:51
3	97522e57-d9ab-4bd6-97bf-c24d952602d2	scared	2021-04-18 05:13:58
4	97522e57-d9ab-4bd6-97bf-c24d952602d2	disgust	2021-01-06 19:13:01
5	97522e57-d9ab-4bd6-97bf-c24d952602d2	interested	2020-08-23 12:25:58
...
25548	75d6b589-7fae-4a6d-b0d0-752845150e56	dislike	2020-06-27 09:46:48

jupyter Accenture Task-2 Last Checkpoint: 13 hours ago

File Edit View Run Kernel Settings Help

Trusted

Markdown

JupyterLab Python 3 (ipykernel)

```
[23]: Content ID    0
      Type        0
      Datetime    0
      dtype: int64
```

```
[24]: merged_df = pd.merge(reaction, content, how='left', on='Content ID')
      merged_df
```

```
[24]:
```

	Content ID	Type_x	Datetime	Type_y	Category
0	97522e57-d9ab-4bd6-97bf-c24d952602d2	disgust	2020-11-07 09:43:50	photo	studying
1	97522e57-d9ab-4bd6-97bf-c24d952602d2	dislike	2021-06-17 12:22:51	photo	studying
2	97522e57-d9ab-4bd6-97bf-c24d952602d2	scared	2021-04-18 05:13:58	photo	studying
3	97522e57-d9ab-4bd6-97bf-c24d952602d2	disgust	2021-01-06 19:13:01	photo	studying
4	97522e57-d9ab-4bd6-97bf-c24d952602d2	interested	2020-08-23 12:25:58	photo	studying
...
24568	75d6b589-7fae-4a6d-b0d0-752845150e56	dislike	2020-06-27 09:46:48	audio	technology
24569	75d6b589-7fae-4a6d-b0d0-752845150e56	intrigued	2021-02-16 17:17:02	audio	technology
24570	75d6b589-7fae-4a6d-b0d0-752845150e56	interested	2020-09-12 03:54:58	audio	technology
24571	75d6b589-7fae-4a6d-b0d0-752845150e56	worried	2020-11-04 20:08:31	audio	technology
24572	75d6b589-7fae-4a6d-b0d0-752845150e56	cherish	2021-01-04 04:55:11	audio	technology

24573 rows × 5 columns

```
[26]: merged_df.rename(columns={"Type_x": "Reaction Type", "Type_y": "Content Type"}, inplace=True)
```

```
[27]: merged_df
```



```
[33]: final = pd.merge(merged_df, reactiontypes, on="Reaction Type", how="left")
```

```
[34]: final
```

```
[34]:
```

	Content ID	Reaction Type	Datetime	Content Type	Category	Sentiment	Score
0	97522e57-d9ab-4bd6-97bf-c24d952602d2	disgust	2020-11-07 09:43:50	photo	studying	negative	0
1	97522e57-d9ab-4bd6-97bf-c24d952602d2	dislike	2021-06-17 12:22:51	photo	studying	negative	10
2	97522e57-d9ab-4bd6-97bf-c24d952602d2	scared	2021-04-18 05:13:58	photo	studying	negative	15
3	97522e57-d9ab-4bd6-97bf-c24d952602d2	disgust	2021-01-06 19:13:01	photo	studying	negative	0
4	97522e57-d9ab-4bd6-97bf-c24d952602d2	interested	2020-08-23 12:25:58	photo	studying	positive	30
...
24568	75d6b589-7fae-4a6d-b0d0-752845150e56	dislike	2020-06-27 09:46:48	audio	technology	negative	10
24569	75d6b589-7fae-4a6d-b0d0-752845150e56	intrigued	2021-02-16 17:17:02	audio	technology	positive	45
24570	75d6b589-7fae-4a6d-b0d0-752845150e56	interested	2020-09-12 03:54:58	audio	technology	positive	30
24571	75d6b589-7fae-4a6d-b0d0-752845150e56	worried	2020-11-04 20:08:31	audio	technology	negative	12
24572	75d6b589-7fae-4a6d-b0d0-752845150e56	cherish	2021-01-04 04:55:11	audio	technology	positive	70

24573 rows x 7 columns

```
[35]: final['Category'].unique()
```

```
[35]: array(['studying', 'healthy eating', 'technology', 'food', 'cooking',  
      'dogs', 'soccer', 'public speaking', 'science', 'tennis', 'travel',  
      'fitness', 'education', 'veganism', 'animals', 'culture'],  
      dtype=object)
```

Top Performing Categories

```
[42]: top_categories = final.groupby(['Category'])['Score'].sum().reset_index()
top_categories
```

```
[42]:
```

	Category	Score
0	animals	74965
1	cooking	64756
2	culture	66579
3	dogs	52511
4	education	57436
5	fitness	55323
6	food	66676
7	healthy eating	69339
8	public speaking	49264
9	science	71168
10	soccer	57783
11	studying	54269
12	technology	68738
13	tennis	50339
14	travel	64880
15	veganism	49619

1	cooking	67130
2	culture	66579
3	dogs	52511
4	education	57436
5	fitness	55323
6	food	66676
7	healthy eating	69339
8	public speaking	49264
9	science	71168
10	soccer	57783
11	studying	54269
12	technology	68738
13	tennis	50339
14	travel	64880
15	veganism	49619

Saving Our Work

```
[43]: final.to_csv('final.csv')
      top_categories.to_csv('top_categories.csv')
```