Sentiment Analysis and Reaction Insights

Quick Analysis of the Dataset :

Loading the Data

```
reactions_types_df = pd.read_csv("ReactionTypes (1).csv")
reactions_df = pd.read_csv("Reactions (1).csv")
content_df = pd.read_csv("Content (1).csv")
```

```
reactions_df.head()
```

	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
4	4	97522e57-d9ab-4bd6-97bf-c24d952602d2	34e8add9-0206-47fd-a501-037b994650a2	disgust	2021-01-06 19:13:01

```
content_df.head()
```

	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...

```
reactions_types_df.shape
```

```
(16, 4)
```

```
reactions_df.shape
```

```
(25553, 5)
```

```
content_df.shape
```

```
(1000, 6)
```

Missing values - Reactions_types

```
reactions_types_df.isna().sum()
```

```
Unnamed: 0    0
Type          0
Sentiment     0
Score         0
dtype: int64
```

Missing values - Reactions

```
missing_reactions=reactions_df.isna().sum()
missing_reactions
```

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

Missing values - Content

```
missing_content=content_df.isna().sum()
missing_content
```

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

Merge reactions_df with content_df on Content ID using a left join.

```
merged_df = reactions_df.merge(content_df, on='Content ID', how='left')
```

```
merged_df.head()
```

	Content ID	Reaction Type	Datetime	Content Type	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

Merge reactions_df with content_df on Content ID using a left join.

40]:
```
merged_df = reactions_df.merge(content_df, on='Content ID', how='left')
```

41]:
```
merged_df.head()
```

41]:

	Content ID	Reaction Type	Datetime	Content Type	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

Merge the resulting DataFrame with reactions_types_df on Reaction Type using a left join.

13]:
```
final_df = merged_df.merge(reactions_types_df, left_on='Reaction Type', right_on='Reaction Type', how='left')
```

14]:
```
final_df.head()
```

14]:

	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

15]:
```
final_df.shape
```

15]:
```
(24573, 7)
```

Project Overview

This project focused on analyzing social media reaction data to understand how users interact with different types of content, their preferences, and sentiment trends. The goal was to clean and merge three datasets (ReactionTypes, Reactions, and Content) into one comprehensive dataset for meaningful insights.

Steps Undertaken

1. **Understanding the Data**:
    - ❖ Reviewed the structure and contents of three datasets:
        - ✓ ReactionTypes: Included types of reactions with their sentiment and scores.
        - ✓ Reactions: Contained user reactions to specific content, along with timestamps.
        - ✓ Content: Provided details about content types, categories, and associated metadata.
2. **Data Cleaning**:
    - ❖ **ReactionTypes**:
        - ✓ Verified there were no missing values and dropped irrelevant columns.
        - ✓ Renamed columns for better clarity (e.g., renamed Type to Reaction Type).
    - ❖ **Reactions**:
        - ✓ Handled missing values by removing rows where reaction types were missing.
        - ✓ Dropped unnecessary columns like User ID and Unnamed: 0.
        - ✓ Converted the Datetime column into a proper datetime format for time-based analysis.

- ❖ **Content**:
  - ✓ Removed irrelevant columns such as URL, User ID, and Unnamed: 0.
  - ✓ Renamed columns (e.g., Type to Content Type).
  - ✓ Cleaned the Category column to remove extra quotes and standardized categories to avoid duplication.
- ❖ Ensured all datasets were consistent and free of redundant information.

3. **Data Merging**:
   - ❖ Merged the Reactions dataset with the Content dataset using the Content ID column to associate reactions with their respective content.
   - ❖ Further merged the resulting dataset with ReactionTypes using the Reaction Type column to add sentiment and score information for each reaction.
4. **Final Dataset**:
   - ❖ The final dataset contained columns for Content ID, Reaction Type, Datetime, Content Type, Category, Sentiment, and Score.
   - ❖ Verified the integrity of the merged data and ensured no critical information was lost during cleaning and merging.
5. **Outcome and Insights**:
   - ❖ The project prepared a comprehensive dataset ready for further analysis, such as identifying popular content categories, analyzing user sentiment trends, and scoring content reactions based on user engagement.

Key Outcomes:

- ✓ Created a unified dataset with 24,573 rows and 7 columns, ready for further analysis like sentiment trends, popular content categories, and user engagement patterns.

What I Learned:

- ✓ Data cleaning is essential for good analysis.
- ✓ How to merge datasets without losing important information.
- ✓ How to handle issues like missing values, messy text, and date formatting**.**