

### 1. Data Collection:

- I used train.csv from <https://www.kaggle.com/datasets/emirhanai/social-media-usage-and-emotional-well-being/data?select=train.csv>
  - (Renamed to social\_data.csv in Python)
- Python File is attached.

### 2. Data Storage:

- I am using MySQL because the Kaggle dataset contains structured data which I will be using and analyzing.

### 3. Data Cleaning:

- I am using Pandas to clean the data before it is transferred to the MySQL database.
  - There are incorrectly labelled values in the dataset due to human error, as some lines contain age in the gender column, and gender in the age column.
  - I am using an expected\_dtypes line to set the expected types of each column ('int64' and 'str') and dropping any value that does not fit the expected data type.
  - There are no outliers detected, which I confirmed by finding the minimum, maximum, average, and standard deviation of each integer-based column (age, daily\_usage, posts\_per\_day, likes\_received\_per\_day, comments\_received\_per\_day, messages\_sent\_per\_day)

### 4. Data Transformation:

- I am adding columns based on the raw data that I will then use to analyze the population (age\_group, usage\_group, popularity, sociality).
- These columns are based on each person's relationship to the mean of the corresponding subject.
- The raw data is overwhelming, with 1000 rows and 9 columns (excluding id), so I am consolidating the data into more manageable mini datasets.

### 5. Experiment

- Once the data was transferred to a table in the MySQL database, I conducted various analyses on the emotional effects of different aspects of social media.

**Dataset (Lines 1-23 of 1000):**

id	age	gender	platform	daily_usage	posts_per_day	likes_received_per_day	comments_received_per_day	messages_sent_per_day	dominant_emotion
1	25	Female	Instagram	120	3	45	10	12	Happiness
2	30	Male	Twitter	90	5	20	25	30	Anger
3	22	Non-binary	Facebook	60	2	15	5	20	Neutral
4	28	Female	Instagram	200	8	100	30	50	Anxiety
5	33	Male	LinkedIn	45	1	5	2	10	Boredom
6	21	Male	Instagram	150	4	60	15	25	Happiness
7	27	Female	Twitter	85	3	30	10	18	Anger
8	24	Non-binary	Facebook	110	6	25	12	22	Sadness
9	29	Female	LinkedIn	55	2	10	3	8	Neutral
10	31	Male	Instagram	170	5	80	20	35	Happiness
11	23	Female	Twitter	75	4	35	7	20	Anxiety
12	26	Non-binary	Facebook	95	3	20	10	18	Sadness
13	34	Male	LinkedIn	65	1	12	4	15	Boredom
14	22	Female	Instagram	180	7	90	25	40	Happiness
15	28	Male	Twitter	100	6	40	23	28	Anger
16	21	Non-binary	Facebook	40	1	5	2	10	Neutral
17	35	Female	Instagram	125	4	55	18	30	Anxiety
18	27	Male	Twitter	90	3	33	15	25	Sadness
19	23	Non-binary	LinkedIn	50	1	8	3	12	Neutral
20	32	Female	Instagram	140	5	70	22	33	Happiness
21	26	Male	Facebook	75	2	25	10	18	Anxiety
22	24	Female	Twitter	105	4	28	14	20	Anger
23	29	Male	LinkedIn	60	2	11	5	17	Boredom

This data records age and gender of each person, their social media usage statistics, and their dominant emotions.

- Age (years):
  - Integer number of years old
- Gender:
  - Female, Male, Non-binary
- Platform:
  - Instagram, Twitter, Facebook, LinkedIn, Whatsapp, Telegram, Snapchat
- Daily Usage (minutes)
- Dominant Emotion:
  - Happiness, Neutral, Boredom, Anxiety, Sadness, Anger

Dataset Breakdown:

gender	count
Female	344
Male	332
Non-binary	248

platform	count
Instagram	236
Twitter	188
Facebook	178
LinkedIn	118
Whatsapp	68
Telegram	68
Snapchat	68

platform	min_age	max_age	mean_age	std_age
Instagram	21	35	27.9831	4.280390942955231
Twitter	21	35	26.7021	3.6577803618030966
Facebook	21	33	26.1461	3.2690257690329076
LinkedIn	23	35	29.8136	3.7663747631708646
Whatsapp	22	35	28.6765	4.191873007136365
Telegram	24	34	28	3.307744922237541
Snapchat	21	29	25.8824	2.8775470137964723

platform	min_daily_usage	max_daily_usage	mean_daily_usage	std_daily_usage
Instagram	115	200	153.4746	22.563549375345936
Twitter	70	105	84.0426	10.82443541215206
Facebook	40	110	72.4719	19.832495118960836
LinkedIn	45	65	55.7627	6.431751173223821
Whatsapp	65	110	88.6765	16.418641822877994
Telegram	50	105	77.5	15.823010050518944
Snapchat	50	120	91.0294	25.256898060806662

platform	min_posts_per_day	max_posts_per_day	mean_posts_per_day	std_posts_per_day
Instagram	3	8	5.839	1.2787065696390822
Twitter	1	6	3.4415	1.238471052451032
Facebook	1	6	1.9663	1.165510335271006
LinkedIn	1	2	1.2542	0.4354316128361061
Whatsapp	2	4	3.0294	0.8569883696666746
Telegram	2	4	2.7353	0.6556910826355281
Snapchat	1	4	2.6471	0.9965337850645539

platform	min_likes_received_per_day	max_likes_received_per_day	mean_likes_received_per_day	std_likes_received_per_day
Instagram	45	110	79.7373	15.486696295121076
Twitter	12	50	35.5691	9.612741174748836
Facebook	5	35	19.9213	8.08414277918089
LinkedIn	5	17	12.839	2.752411389552776
Whatsapp	20	55	37.3235	11.526747891535464
Telegram	15	48	29.6176	11.005385595731148
Snapchat	15	55	30.4412	12.121180238764824

platform	min_comments_received_per_day	max_comments_received_per_day	mean_comments_received_per_day	std_comments_received_per_day
Instagram	10	40	26.8729	6.063990142105339
Twitter	5	30	16.7713	5.100670008980787
Facebook	2	16	8.7978	3.8091544824909467
LinkedIn	2	8	5.5763	1.2782852633874444
Whatsapp	8	25	15.6618	6.248788809977138
Telegram	6	20	12.8382	4.555572844839239
Snapchat	5	20	12.8529	5.278326121375849

platform	min_messages_sent_per_day	max_messages_sent_per_day	mean_messages_sent_per_day	std_messages_sent_per_day
Instagram	12	50	33.5975	6.449529766913729
Twitter	10	30	21.2447	4.093809987932135
Facebook	10	25	16.7753	4.654954613010284
LinkedIn	8	17	12.4831	2.272656367231137
Whatsapp	18	26	22.2941	2.7173669323414353
Telegram	12	28	21.7647	4.905332173973972
Snapchat	12	30	22.0588	5.635711638714358

Personal Statistics vs. Emotion:

```
-- Percent distribution of dominant emotions based on age and gender
-- 1)
SELECT
    AVG(age) as avg_age
FROM social_db.social_stats;

-- 2)
ALTER TABLE social_db.social_stats
ADD COLUMN age_group VARCHAR(20);

-- 3)
UPDATE social_db.social_stats
SET age_group =
    CASE
        WHEN age > 27.5 THEN 'Older'
        ELSE 'Younger'
    END;

SELECT id, age_group FROM social_db.social_stats;

-- 4)
SELECT
    gender,
    age_group,
    AVG(IF(dominant_emotion = 'Happiness', 1, 0)) * 100 as happiness_percentage,
    AVG(IF(dominant_emotion = 'Neutral', 1, 0)) * 100 as neutral_percentage,
    AVG(IF(dominant_emotion = 'Boredom', 1, 0)) * 100 as boredom_percentage,
    AVG(IF(dominant_emotion = 'Anxiety', 1, 0)) * 100 as anxiety_percentage,
    AVG(IF(dominant_emotion = 'Sadness', 1, 0)) * 100 as sadness_percentage,
    AVG(IF(dominant_emotion = 'Anger', 1, 0)) * 100 as angry_percentage
FROM social_db.social_stats
GROUP BY gender, age_group
ORDER BY gender, age_group;
```

id	age_group
1	Younger
2	Older
3	Younger
4	Older
5	Older
6	Younger
7	Younger
8	Younger
9	Older
10	Older
11	Younger
12	Younger
13	Older
14	Younger
15	Older
16	Younger
17	Older
18	Younger
19	Younger
20	Older
21	Younger
22	Younger
23	Older

- In the above code, I ran Part 1 to get the average age of all the people, which is 27.5.
- I ran Parts 2 and 3 to add a new column in my table, and created a separation, with all people with age older than the mean as “Older,” and the rest as “Younger.”
- The resultant column along with the corresponding IDs is to the right.
- Part 4 is below:

gender	age_group	happiness_percentage	neutral_percentage	boredom_percentage	anxiety_percentage	sadness_percentage	angry_percentage
Female	Older	26.8293	10.9756	18.2927	17.0732	10.9756	15.8537
Female	Younger	32.2222	21.1111	0	14.4444	16.6667	15.5556
Male	Older	24	10	19	14	13	20
Male	Younger	13.6364	19.697	12.1212	21.2121	19.697	13.6364
Non-binary	Older	10.4167	16.6667	27.0833	27.0833	8.3333	10.4167
Non-binary	Younger	5.2632	43.4211	13.1579	13.1579	25	0

- This data shows a distinction between the different genders and how their age groups feel emotionally. The older groups were consistently happier or bored around the board, while the younger groups were neutral or sadder around the board as well. Anxiety and anger see variation, as only males were less anxious as they were older, and only females did not see a notable increase in anger along with age.

### Social Media Usage vs. Emotion:

```
-- Percent distribution of dominant emotions based on platform and usage
-- 1)
SELECT
    AVG(daily_usage) as avg_usage
FROM social_db.social_stats;

-- 2)
ALTER TABLE social_db.social_stats
ADD COLUMN usage_group VARCHAR(20);

-- 3)
UPDATE social_db.social_stats
SET usage_group =
CASE
    WHEN daily_usage > 96.3095 THEN 'Avid'
    ELSE 'Casual'
END;

SELECT id, usage_group FROM social_db.social_stats;

-- 4)
SELECT
    platform,
    usage_group,
    AVG(IF(dominant_emotion = 'Happiness', 1, 0)) * 100 as happiness_percentage,
    AVG(IF(dominant_emotion = 'Neutral', 1, 0)) * 100 as neutral_percentage,
    AVG(IF(dominant_emotion = 'Boredom', 1, 0)) * 100 as boredom_percentage,
    AVG(IF(dominant_emotion = 'Anxiety', 1, 0)) * 100 as anxiety_percentage,
    AVG(IF(dominant_emotion = 'Sadness', 1, 0)) * 100 as sadness_percentage,
    AVG(IF(dominant_emotion = 'Anger', 1, 0)) * 100 as angry_percentage
FROM social_db.social_stats
GROUP BY platform, usage_group
ORDER BY platform, usage_group;
```

id	usage_group
1	Avid
2	Casual
3	Casual
4	Avid
5	Casual
6	Avid
7	Casual
8	Avid
9	Casual
10	Avid
11	Casual
12	Casual
13	Casual
14	Avid
15	Avid
16	Casual
17	Avid
18	Casual
19	Casual
20	Avid
21	Casual
22	Avid
23	Casual

- In the above code, I ran Part 1 to get the average age of all the people, which is 96.3095.
- I ran Parts 2 and 3 to add a new column in my table, and created a separation, with all people with greater average usage of social media as “Avid” users, and the rest as “Casual” users.
- The resultant column along with the corresponding IDs is to the right.
- Part 4 is below:

platform	usage_group	happiness_percentage	neutral_percentage	boredom_percentage	anxiety_percentage	sadness_percentage	angry_percentage
Facebook	Avid	0	33.3333	0	33.3333	33.3333	0
Facebook	Casual	0	37.8378	22.973	25.6757	13.5135	0
Instagram	Avid	67.7966	8.4746	0	12.7119	7.6271	3.3898
LinkedIn	Casual	0	16.9492	59.322	15.2542	8.4746	0
Snapchat	Avid	0	38.4615	0	61.5385	0	0
Snapchat	Casual	19.0476	19.0476	0	0	61.9048	0
Telegram	Avid	0	0	0	100	0	0
Telegram	Casual	0	43.3333	13.3333	0	26.6667	16.6667
Twitter	Avid	0	0	0	0	0	100
Twitter	Casual	4.7619	10.7143	10.7143	11.9048	27.381	34.5238
Whatsapp	Avid	35.7143	0	0	28.5714	0	35.7143
Whatsapp	Casual	0	40	0	20	0	40

- This dataset is especially interesting and useful, as there is a clear pattern and indication of the ties between social media usage and emotional wellbeing. Facebook users carried no happy users, with neutral feeling percentage landing only in the 30s, and the rest carrying negative emotion. Instagram carries no casual user, yet the majority are happy, while Twitter’s avid users are 100% angry. This may not reflect reality; however, this is an indication that Twitter may be a severe source of anger. WhatsApp also carries a significant anger percentage, however not as severe, and avid users are also just as happy.

### Popularity vs. Emotion

```
-- Percent distribution of dominant emotions based on posts/day and likes-received/day
-- 1)
SELECT AVG(likes_received_per_day) / AVG(posts_per_day)
FROM social_db.social_stats;

-- 2)
ALTER TABLE social_db.social_stats
ADD COLUMN popularity VARCHAR(20);

-- 3)
UPDATE social_db.social_stats
SET popularity =
CASE
    WHEN likes_received_per_day / posts_per_day > 12.01550889 THEN 'Popular'
    ELSE 'Unpopular'
END;

SELECT id, popularity FROM social_db.social_stats;

-- 4)
SELECT
    popularity,
    AVG(IF(dominant_emotion = 'Happiness', 1, 0)) * 100 as happiness_percentage,
    AVG(IF(dominant_emotion = 'Neutral', 1, 0)) * 100 as neutral_percentage,
    AVG(IF(dominant_emotion = 'Boredom', 1, 0)) * 100 as boredom_percentage,
    AVG(IF(dominant_emotion = 'Anxiety', 1, 0)) * 100 as anxiety_percentage,
    AVG(IF(dominant_emotion = 'Sadness', 1, 0)) * 100 as sadness_percentage,
    AVG(IF(dominant_emotion = 'Anger', 1, 0)) * 100 as angry_percentage
FROM social_db.social_stats
GROUP BY popularity;
```

id	popularity
1	Popular
2	Unpopular
3	Unpopular
4	Popular
5	Unpopular
6	Popular
7	Unpopular
8	Unpopular
9	Unpopular
10	Popular
11	Unpopular
12	Unpopular
13	Unpopular
14	Popular
15	Unpopular
16	Unpopular
17	Popular
18	Unpopular
19	Unpopular
20	Popular
21	Popular
22	Unpopular
23	Unpopular

- In the above code, I ran Part 1 to get the average daily likes per post, which is 12.0155.
- I ran Parts 2 and 3 to add a new column in my table, and created a separation, with all people with greater than average daily likes received per post as “Popular,” and the rest as “Unpopular.”
- The resultant column along with the corresponding IDs is to the right.
- Part 4 is below:

popularity	happiness_percentage	neutral_percentage	boredom_percentage	anxiety_percentage	sadness_percentage	angry_percentage
Popular	32.8704	18.287	5.787	26.3889	10.1852	6.4815
Unpopular	8.9431	21.3415	21.3415	8.5366	20.7317	19.1057

- This data is predictable, yet important to acknowledge. Popular users tended to be significantly happier than the Unpopular users, however it also caused a notable increase in anxiety. They were also less bored (which makes sense). Unpopular users tended to be significantly sadder and angrier, which is a dangerous reflection of how social media can negatively affect a person’s mind if popularity is a focus.

### Sociality vs. Emotion

```
-- Percent distribution of dominant emotions based on messages-sent/day
-- 1)
SELECT AVG(messages_sent_per_day)
FROM social_db.social_stats;

-- 2)
ALTER TABLE social_db.social_stats
ADD COLUMN sociality VARCHAR(20);

-- 3)
UPDATE social_db.social_stats
SET sociality =
CASE
    WHEN messages_sent_per_day > 22.5952 THEN 'Social'
    ELSE 'Unsocial'
END;

SELECT id, sociality FROM social_db.social_stats;

-- 4)
SELECT
    sociality,
    AVG(IF(dominant_emotion = 'Happiness', 1, 0)) * 100 as happiness_percentage,
    AVG(IF(dominant_emotion = 'Neutral', 1, 0)) * 100 as neutral_percentage,
    AVG(IF(dominant_emotion = 'Boredom', 1, 0)) * 100 as boredom_percentage,
    AVG(IF(dominant_emotion = 'Anxiety', 1, 0)) * 100 as anxiety_percentage,
    AVG(IF(dominant_emotion = 'Sadness', 1, 0)) * 100 as sadness_percentage,
    AVG(IF(dominant_emotion = 'Anger', 1, 0)) * 100 as angry_percentage
FROM social_db.social_stats
GROUP BY sociality;
```

id	sociality
1	Unsocial
2	Social
3	Unsocial
4	Social
5	Unsocial
6	Social
7	Unsocial
8	Unsocial
9	Unsocial
10	Social
11	Unsocial
12	Unsocial
13	Unsocial
14	Social
15	Social
16	Unsocial
17	Social
18	Social
19	Unsocial
20	Social
21	Unsocial
22	Unsocial
23	Unsocial

- In the above code, I ran Part 1 to get the average age of all the people, which is 22.5952.
- I ran Parts 2 and 3 to add a new column in my table, and created a separation, with all people with greater than average daily likes received per post as “Popular,” and the rest as “Unpopular.”
- The resultant column along with the corresponding IDs is to the right.
- Part 4 is below:

sociality	happiness_percentage	neutral_percentage	boredom_percentage	anxiety_percentage	sadness_percentage	angry_percentage
Unsocial	1.7341	30.0578	25.0482	14.6435	15.9923	12.5241
Social	43.7037	6.9136	0	19.7531	15.5556	14.0741

- This data is very notable, as there are some significant indicators toward certain emotions. Less social people were very rarely happy, while almost half of the social people were happy while completely avoiding boredom. The unsocial group is much more dispersed in emotions, with significant amounts in each, however the neutral and boredom percentages of the social group seemingly moved toward happiness. The other 3 emotions are consistent between the two. This data is indicative that sociality plays an important role in upholding happiness for more neutral feeling people, and social media is simply a means for that. However, negative emotions are unaffected by sociality, as they remain largely consistent between the two groups here.

Overall Analysis

There are various effects that social media has on people's minds, some aspects more than others. While it is interesting that different apps were indicative of very different general emotions among the people, it is important to note that this is an effect over time as certain groups of people gravitated toward each other in those certain apps, whether they are to generate positivity, negativity, or blatant hatred. Users may gravitate toward certain apps to find what they want. Avid Twitter users who are tired of constant arguments may find solace in Instagram, where positivity is more evident. Otherwise, they may also choose to use their app less, or even give up social media altogether. This data indicates that there are many steps that people can take to uplift and uphold their mental health, and a strong first step would be looking at how your phone is affecting it. While older groups were less emotionally polarized by social media, it still has a strong effect on the younger people. Popularity and sociality are not always controllable, and if that seems to affect a person's emotional wellbeing negatively, they should seriously consider changing their lifestyle for the better, whether to change how they use social media, or their outlook on life entirely.

This data is strongly indicative of the positive and negative effects of social media, and it is up to the user to decide how they want it to affect their lives.