Data Analysis Assessment

A look into EG's Social Media data

Siddhant Thakur
May 26th, 2023

Agenda

1. **Data**
   a. General
   b. Cleaning
   c. Outliers
2. **Questions on Engagement Rate**
   a. Typical Reach
   b. Day and Time of Posting
   c. Game Titles
   d. Media Types
   e. Campaigns
3. **Posting Strategy & Suggestions**

Data - General

1. **3479** rows with **966** rows having 0 Total Engagement.
2. Contains **Impression** and **Engagement** statistics for posts made by different **EG** accounts on different platforms.
3. Also contains the **date & time** as well as the **type of media** associated with the post.

	Published Date	Account	Account Type	Campaign Name	Total Impressions	Total Engagements	Media Type
0	2023-03-31 19:55:43.192	CSGO	TWITTER	N/A	0	0	Text
1	2023-03-31 19:49:35.243	CSGO	TWITTER	N/A	0	0	Text
2	2023-03-31 19:49:33.535	CSGO	TWITTER	N/A	0	0	Text
3	2023-03-31 19:49:31.583	CSGO	TWITTER	N/A	0	0	Text
4	2023-03-31 19:43:42.590	CSGO	TWITTER	Community Engagement	9517	1215	Video
5	2023-03-31 19:42:11.331	CSGO	TWITTER	N/A	0	0	Photo
6	2023-03-31 19:40:57.383	CSGO	TWITTER	Community Engagement	1882	83	Photo
7	2023-03-31 19:23:57.349	CSGO	TWITTER	Community Engagement	17892	3076	Video
8	2023-03-31 19:15:14.431	CSGO	TWITTER	N/A	0	0	Video
9	2023-03-31 18:55:51.855	CSGO	TWITTER	Community Engagement	13874	1447	Video
10	2023-03-31 18:36:01.000	General	FBPAGE	Community Engagement	37692	8044	Photo

Data - Cleaning

1. **No** NaN values present.
2. **'Account'** column had **'General'** and **'General '** with **'LINKEDIN_COMPANY'** as the **'Account Type'** under the first value.
   a. Solution - Combined together into one **'General'**.
3. Calculated **Engagement Rate** using the formula given on the right.

```
data['Account'].value_counts()

General            2271
DOTA2               803
CSGO                270
Valorant             60
Content Creators     53
General              22
Name: Account, dtype: int64
```
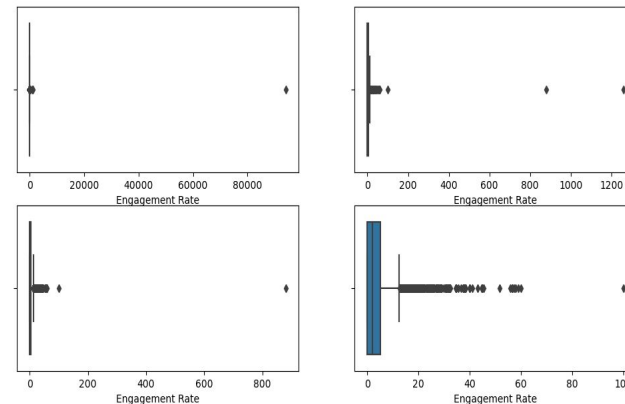
$$Engagement\ Rate = \frac{Total\ Engagements}{Total\ Impressions} * 100$$

Data - Outliers

Checking for outliers in data by looking at the Engagement Rate

1. 4 outliers present in the data with **Engagement Rate more than 100%,** assumed to be edge cases.
    a. Solution - Removed as data was being skewed.

	Published Date	Account	Account Type	Campaign Name	Total Impressions	Total Engagements	Media Type	Engagement Rate
837	2023-03-10 12:37:06	General	FBPAGE	N/A	1	940	Photo	94000.000000
3457	2023-01-05 11:55:44	General	FBPAGE	N/A	300	3776	Photo	1258.666667
3216	2023-01-19 06:02:34	General	FBPAGE	N/A	5	44	Photo	880.000000
1586	2023-02-23 06:18:27	General	FBPAGE	N/A	1	1	Photo	100.000000

Final data after removing outliers
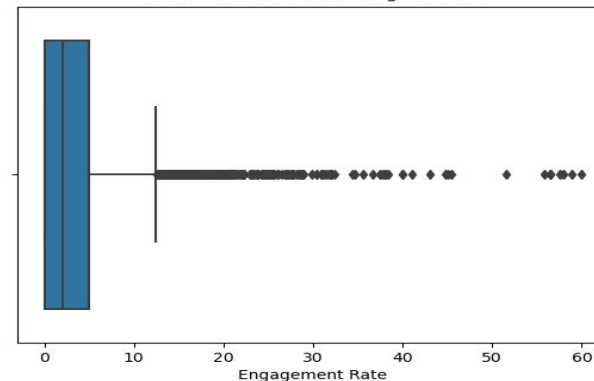
Questions - Typical Reach

1. **What is the typical engagement rate we can expect?**
   a. On average we see an engagement rate of 4.175% with a standard deviation of 6.538.
2. **What's the likelihood that we can achieve a 15% engagement rate?**
   a. Assuming normal distribution, the probability of achieving an engagement rate of 15% or more will be approximately 4.85% using z-score.
   b. Since the distribution is skewed to the right, the probability of achieving an engagement rate of at least 15% shifts to 6.38 by using cumulative histogram.
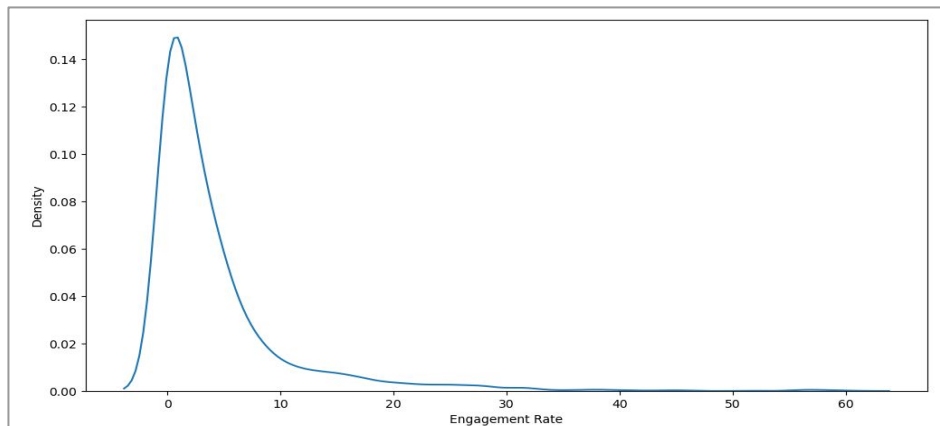
	Total Impressions	Total Engagements	Engagement Rate
count	3475.000000	3475.000000	3475.000000
mean	10984.995396	947.111079	4.175152
std	29591.818356	3470.221546	6.538214
min	0.000000	0.000000	0.000000
25%	169.000000	0.000000	0.000000
50%	1888.000000	38.000000	2.038835
75%	12469.000000	276.000000	5.000000
max	753227.000000	95062.000000	60.000000

# Questions - Day and Time of Posting

1. **Does day of the week of posting affect engagement rates?**
   a. Posting on **Tuesday or Thursday** clearly have **better** engagement rates when compared to posting on Wednesday or Saturday.
2. **Does time of posting affect engagement rates?**
   a. Posting at **Night (from midnight to 6am)** has a significantly **higher** engagement rate (7.87) than any other time of the day.

Time Of Day	Total Impressions	Total Engagements	Engagement Rate	Count
Night	10147.729167	2071.604167	7.876963	48
Morning	12467.673352	1033.527221	4.363287	1396
Afternoon	10284.404525	884.007174	4.017833	1812
Evening	7513.949772	671.913242	3.466193	219
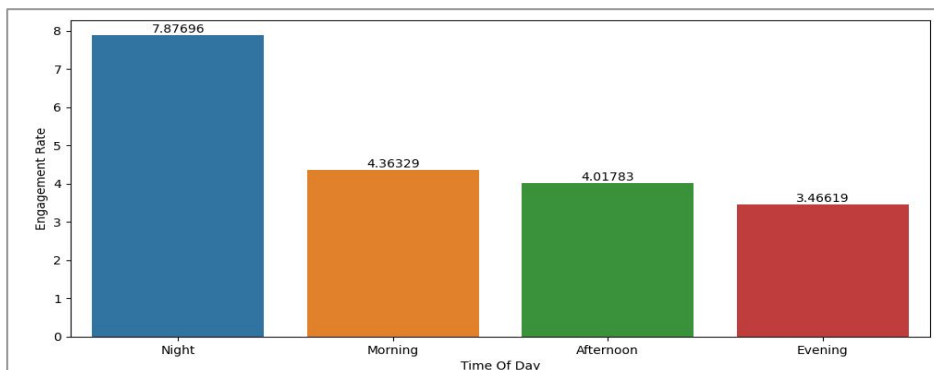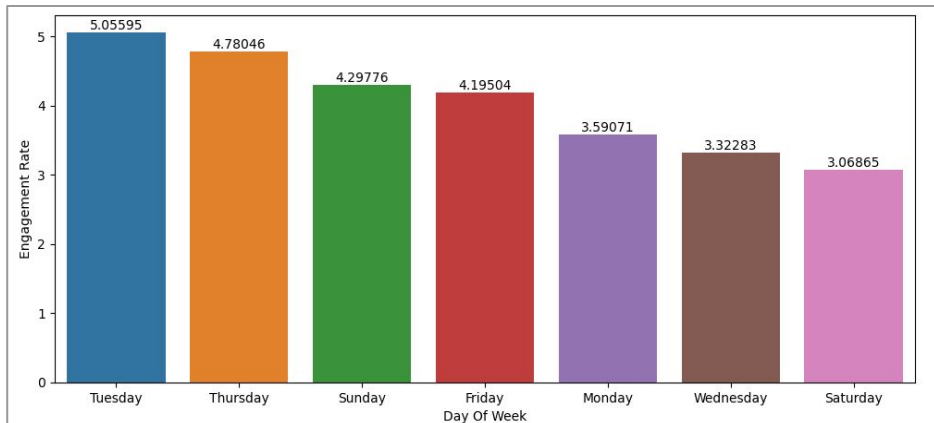
Questions - Game Types

1. **How are game titles doing in terms of social performance? Specific game to focus on?**
   a. Highest impressions and engagement rate (4.299%) is maintained by **'General'** posts.
   b. Even though 'CSGO' has a higher average impression count with fewer posts, the **'DOTA2'** account is doing better with an engagement rate of **4.298%**.
   c. Focusing on **'Content Creators'** could also help increase engagement as it ranks third after 'General' and 'DOTA2' tags but has fewer samples.
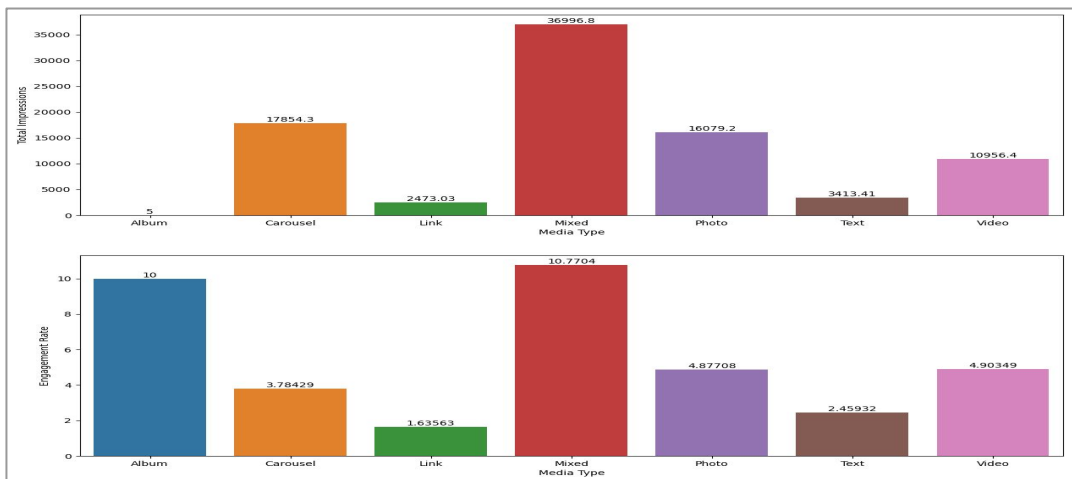
Account	Total Impressions	Total Engagements	Engagement Rate	Count
General	14823.700306	1342.318480	4.299421	2289
DOTA2	2315.474471	154.408468	4.298295	803
Content Creators	852.566038	26.169811	4.075632	53
CSGO	8570.066667	341.985185	3.030547	270
Valorant	382.983333	15.516667	3.024861	60
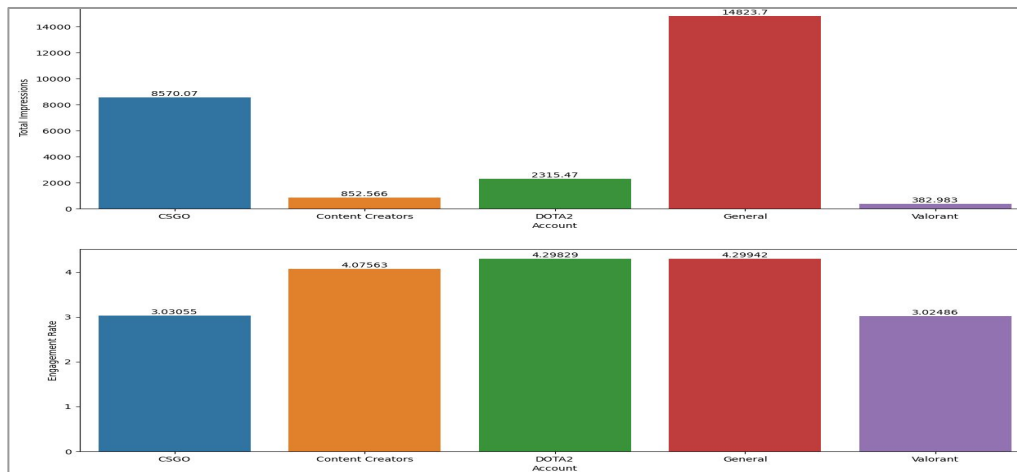
Questions - Media Types

1. **What media type performs the best?**
   a. Posts with '**Mixed**' form of media has performed the best with an engagement rate of **10.77%**. (Note that the number of samples are very few to make any sort of concrete conclusion.)
   b. '**Video**' and '**Photo**' forms of media follow with an engagement rate of around 4.903 & 4.877 respectively, with sufficient samples.
   c. 'Album' hasn't been considered in the mix due to **insufficient** number of posts & impressions.

| Media Type	Total Impressions	Total Engagements	Engagement Rate	Count
Mixed	36996.800000	2732.800000	10.770357	5
Album	5.000000	2.000000	10.000000	4
Video	10956.373320	888.780765	4.903493	967
Photo	16079.195828	1455.570659	4.877078	1486
Carousel	17854.333333	726.777778	3.784294	9
Text	3413.409890	270.672527	2.459318	910
Link	2473.031915	24.010638	1.635631	94

Questions - Campaigns

1. **What is our best performing campaign?**
   a. **'Community Engagement'** campaigns have been the most successful, grossing an engagement rate of **5.63%** as well as the highest number of impressions on average.
   b. Posts with **no campaign** have also had a similar, if not better, amount of engagement rate when compared to other EG campaigns.


```
In [53]: data[data['Campaign Name']=='Community Engagement '].groupby('Account Type')

Out[53]:
```

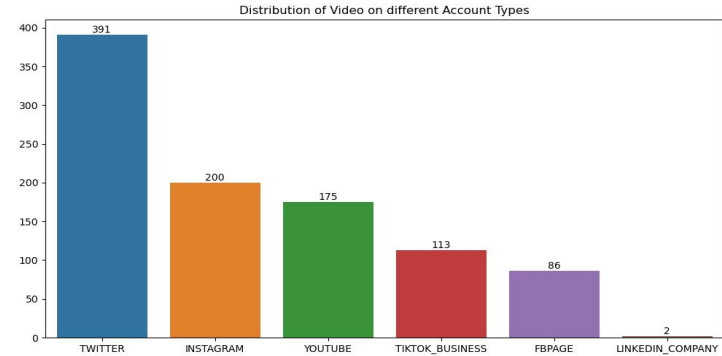| Account Type	Total Impressions	Total Engagements	Engagement Rate
FBPAGE	29475.847176	5420.431894	14.505180
INSTAGRAM	9684.308176	332.088050	1.208653
LINKEDIN_COMPANY	1726.444444	33.166667	1.860454
TIKTOK_BUSINESS	802.076923	46.923077	6.422724
TWITTER	14815.926585	398.155729	3.468913
YOUTUBE	1461.904762	60.285714	7.455909
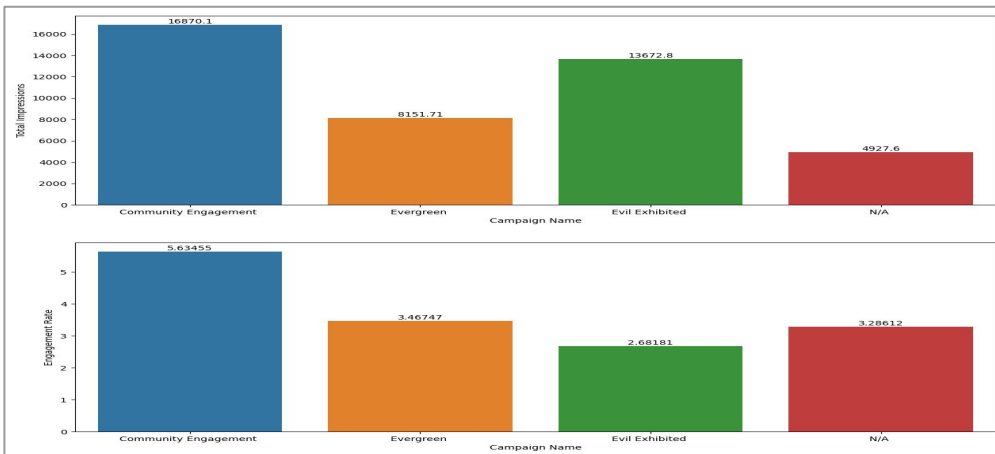
Extra - XGBoost & SHAP

1. Developed an **XGBoost Regression** model with **Engagement Rate** as the **target** variable to understand the impact of features on it.
2. Distribution of predicted & original engagement rate is **similar** with an **RMSE of 5.42**, which can be improved, but our primary aim is to **understand how features interact** with the target variable.
3. **Video, Community Engagement & DOTA2** positively affect the rate which corroborates our analyses.
4. Model gives insight in **Account Type** variable, where posting on **Facebook** has a positive impact on the engagement rate contrary to posting on **Instagram**.

Distribution of Predicted Engagement Rate and Original Engagement Rate

RMSE : 5.424781
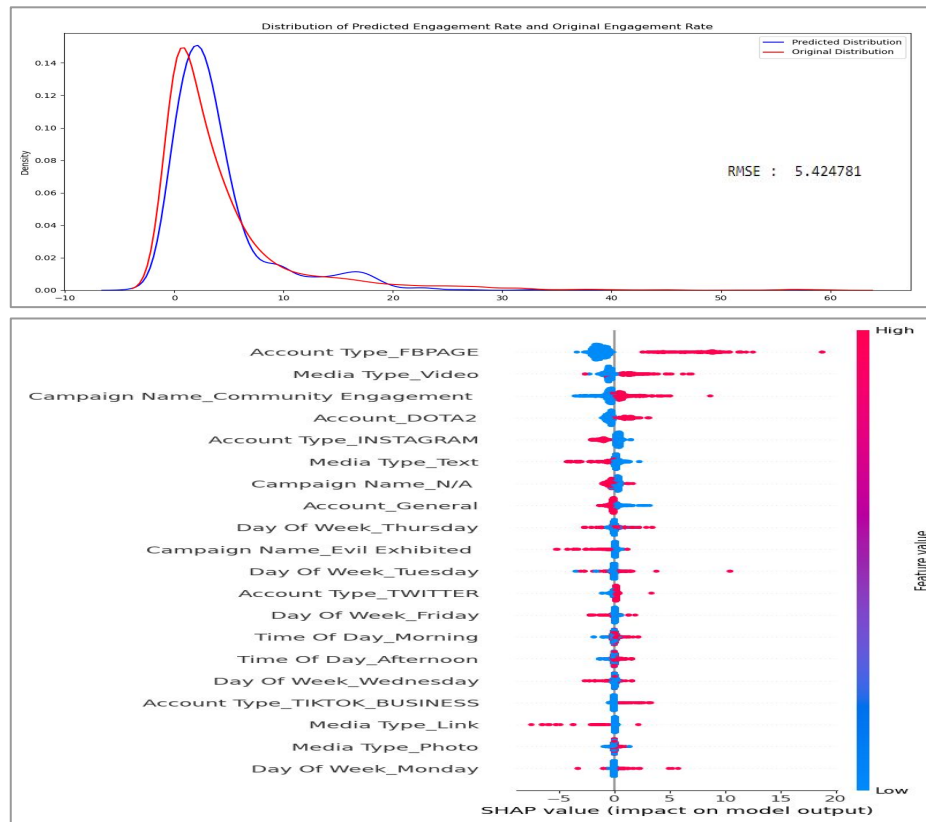
Extra(contd.) - Account Type

1. Photos and Videos on Facebook have a considerably better engagement rate than Instagram.
2. However, number of videos posted on Instagram (200) is higher than Facebook (86).
3. Facebook algorithm might be better at sharing EG's content.

Distribution of Video on different Account Types

```
data[data['Account Type']=='FBPAGE'].groupby('Media Type')
```

Media Type	Total Impressions	Total Engagements	Engagement Rate
Album	5.000000	2.000000	10.000000
Link	458.285714	3.714286	0.057891
Photo	28676.218023	5255.543605	13.396573
Text	9051.872180	1526.082707	3.927818
Video	39510.313953	6530.081395	15.059735

```
data[data['Account Type']=='INSTAGRAM'].groupby('Media Type')
```

Media Type	Total Impressions	Total Engagements	Engagement Rate
Carousel	17854.333333	726.777778	3.784294
Link	0.000000	0.000000	0.000000
Photo	4753.035461	137.595745	0.459752
Text	0.000000	0.000000	0.000000
Video	10365.070000	193.720000	1.141756

Posting Strategy & Suggestions

1. All the Social Channels should preferably post in **mornings (for higher impressions)** or **nights (for higher engagement rate)** on **Tuesdays or Thursdays** focusing on **Community Engagement** and/or DOTA 2 content consisting of **Videos and Photos** posting on **Facebook, Tiktok or Youtube**.
2. Using the analysis gathered from the model, I would like to suggest the EG social media team to post more Videos and Photos on **Facebook**.
3. One of my other suggestions would be to expand towards the **content creators** front through **Youtube** or any other streaming platform, since the data shows a higher engagement rate of **4.07%** with a **max** of **37.93%**.
4. Another suggestion for the team would be to not lose hope with the Valorant content due to recent qualification of EG's Valorant team to Masters Tokyo.
5. **"Correlation is not causation"** should be kept in mind!

Questions?