Intern Assessment

Brandon Kim

2023-05-29

```
library(tidyverse)
```

```
df <- readxl::read_excel(here::here("social_data.xlsx"))
```

Question 1.

What is the typical engagement rate we can expect? What's the likelihood that we can achieve a 15% engagement rate?

```
df %>%
  filter(`Total Impressions` != 0,
         `Total Engagements` <= `Total Impressions`) %>% # Filtering out outliers
  summarize(`expected engagement rate` = mean(`Total Engagements`/`Total Impressions`))
```

```
## # A tibble: 1 x 1
##   `expected engagement rate`
##                        <dbl>
## 1                     0.0535
```

Excluding outliers, we can expect a typical engagement rate of 5.347238% across all media.

```
df %>%
  filter(`Total Impressions` != 0) %>%
  summarize(likelihood = sum(round(`Total Engagements`/`Total Impressions`, 2) == 0.15)/n())
```

```
## # A tibble: 1 x 1
##   likelihood
##        <dbl>
## 1     0.0106
```

The likelihood that of achieving a 15% engagement rate is 0.0106.

Question 2.

Does day of the week and time of posting affect engagement rates?

```
df_time <- df %>%
  separate(`Published Date`, c("date", "time"), sep = " ") %>%
  mutate(day = weekdays(as.POSIXlt(mdy(date))),
         hour = substr(time, 1, 2))
```

```
df_time %>%
  filter(`Total Impressions` != 0) %>%
  group_by(day) %>%
  summarize(engagement_rate = mean(`Total Engagements`/`Total Impressions`),
            engagement_rate_spread = sd(`Total Engagements`/`Total Impressions`))
```

```
## # A tibble: 1 x 3
##   day   engagement_rate engagement_rate_spread
##   <chr>           <dbl>                  <dbl>
## 1 <NA>            0.405                   18.0
```

Just from looking at the average engagement rates for each day, we can clearly see that there are more engagement on posts on Friday. However, looking at the standard deviation, we can see that the spread is really high, probably from outliers. It might be a good idea to filter out the outlying data.

First, we must identify the outliers. Since it seems like Friday has an abnormally large engagement rate, we should filter out data where the amount of engagements far exceeds the amount of impressions.

```
df_time %>%
  filter(`Total Engagements` > `Total Impressions`)
```

```
## # A tibble: 3 x 10
##   date       time     Account ‘Account Type‘ ‘Campaign Name‘ ‘Total Impressions‘
##   <chr>      <chr>    <chr>   <chr>          <chr>                         <dbl>
## 1 2023-03-10 12:37:06 General FBPAGE         N/A                               1
## 2 2023-01-19 06:02:34 General FBPAGE         N/A                               5
## 3 2023-01-05 11:55:44 General FBPAGE         N/A                             300
## # i 4 more variables: ‘Total Engagements‘ <dbl>, ‘Media Type‘ <chr>, day <chr>,
## #   hour <chr>
```

All of these posts happen on Facebook and happen on Thursday or Friday. These posts could lead to very skewed results. For instance, a total engagement value of 940 compared to a total impression value of 1 leads to an extremely high engagement rate, which leads to the mean being a bad metric for determining typical engagement rates, so we should filter these three out.

```
df_time %>%
  filter(`Total Impressions` != 0,
         `Total Engagements` <= `Total Impressions`) %>% # Filtering outliers
  group_by(day) %>%
  summarize(engagement_rate = mean(`Total Engagements`/`Total Impressions`), total = n()) %>%
  arrange(desc(engagement_rate))
```

```
## # A tibble: 1 x 3
##   day   engagement_rate total
##   <chr>           <dbl> <int>
## 1 <NA>           0.0535  2732
```

Just like that, the data seems to suggest that there isn't any truly noticeable differences between each day. However, if we had to pick, Tuesdays seem like our best bet for getting a high engagement rate.

We can apply the same logic for time of day. We will be using the hour as our metric for time.

```
df_time %>%
  filter(`Total Impressions` != 0,
         `Total Engagements` <= `Total Impressions`) %>%
  group_by(hour) %>%
  summarize(engagement_rate = mean(`Total Engagements`/`Total Impressions`), total = n()) %>%
  arrange(desc(engagement_rate))
```

```
## # A tibble: 22 x 3
##    hour  engagement_rate total
##    <chr>           <dbl> <int>
##  1 05             0.115     26
##  2 06             0.0796    67
##  3 08             0.0741   127
##  4 07             0.0666   101
##  5 09             0.0617   304
##  6 21             0.0609     8
##  7 03             0.0566     6
##  8 04             0.0558     8
##  9 17             0.0534   162
## 10 16             0.0532   253
## # i 12 more rows
```

For time, it seems as if hour 5 has the highest engagement rate by a somewhat significant margin. With it being higher than the second highest engagement rate by around 3.5 percentage points. Albeit, it wouldn't be safe to count out hours with low sample sizes, such as hour 22 and 2.

Question 3.

How are our game titles doing in terms of social performance? Is there a specific game we should focus more on or less?

```
df %>%
  filter(Account %in% c("CSGO", "DOTA2", "Valorant"),
         `Total Engagements` <= `Total Impressions`) %>%
  group_by(Account) %>%
  summarize(mean_impressions = mean(`Total Impressions`),
            mean_engagement = mean(`Total Engagements`),
            n = n(),
            mean_eng_rate = mean(`Total Engagements`[`Total Impressions` != 0]/`Total Impressions`[`Tota
            sd_eng_rate = sd(`Total Engagements`[`Total Impressions` != 0]/`Total Impressions`[`Total In
```

```
## # A tibble: 3 x 6
##   Account  mean_impressions mean_engagement     n mean_eng_rate sd_eng_rate
##   <chr>               <dbl>           <dbl> <int>         <dbl>       <dbl>
## 1 CSGO                8570.           342.    270        0.0413      0.0398
## 2 DOTA2               2315.           154.    803        0.0494      0.0382
## 3 Valorant             383.            15.5    60        0.0534      0.0320
```

At first glance, the Valorant account seems to be doing the best when it comes to engagement rate. With the highest mean and lowest standard deviation, each of their posts seem to garner a relatively consistent proportion of interactions from impressions. However, it seems as if the Valorant account has the lowest average impressions per post, so I would recommend focusing on getting the Valorant account more impressions assuming that the sample engagement rate is representative of the population of EG Valorant fans.

The CSGO account seems to be getting the most traction when it comes to impressions, however its engagement rate is the lowest. I would recommend focusing less on getting impressions on the CSGO account's posts as it seems as the engagement from those impressions seems to be the weakest.

The DOTA2 account seems to have the highest number of posts. However, despite having around a third of the amount of posts, CSGO posts seem to garner on average around 4 times more impressions than DOTA2 posts, so I would recommend the social media team to focus less on pushing out DOTA2 posts as it seems as if they are not as popular.

Question 4.

What media type performs the best?

```
df %>%
  group_by(`Media Type`) %>%
  summarize(mean_impressions = mean(`Total Impressions`),
            mean_engagement = mean(`Total Engagements`),
            n = n(),
            mean_eng_rate = mean(`Total Engagements`[`Total Impressions` != 0]/`Total Impressions`[`Tota
            sd_eng_rate = sd(`Total Engagements`[`Total Impressions` != 0]/`Total Impressions`[`Total In
```

```
## # A tibble: 7 x 6
##   `Media Type` mean_impressions mean_engagement     n mean_eng_rate sd_eng_rate
##   <chr>                   <dbl>           <dbl> <int>         <dbl>       <dbl>
## 1 Album                       5             2       4        0.4         NA
## 2 Carousel                17854.          727.      9        0.0378       0.0239
## 3 Link                     2473.           24.0    94        0.0375       0.0971
## 4 Mixed                   36997.         2733.      5        0.108        0.0557
## 5 Photo                   16036.         1455.   1490        0.831       26.6
## 6 Text                     3413.          271.    910        0.0408       0.0438
## 7 Video                   10956.          889.    967        0.0535       0.0588
```

From the data, it seems as if the most popular forms of media are photos, links, texts and videos. Out of those four, photos seem to be the most effective in achieving a high engagement rate at a first glance. It also seems to garner the most impressions per post, and such a high engagement from those impressions. However, the standard deviation seems to be awfully high. Recalling from question 2, all of our outliers were photo posts. So a good step to take now is to remove any outliers similar to in question 2.

It's also important to note that album has an NA value for the standard deviation of engagement rate because there's only 1 album post that can be calculated for engagement rate.

```
df %>%
  filter(`Total Engagements` <= `Total Impressions`) %>%
  group_by(`Media Type`) %>%
  summarize(mean_impressions = mean(`Total Impressions`),
            mean_engagement = mean(`Total Engagements`),
            n = n(),
            mean_eng_rate = mean(`Total Engagements`[`Total Impressions` != 0]/`Total Impressions`[`Tota
            sd_eng_rate = sd(`Total Engagements`[`Total Impressions` != 0]/`Total Impressions`[`Total In
```

```
## # A tibble: 7 x 6
##   `Media Type` mean_impressions mean_engagement     n mean_eng_rate sd_eng_rate
##   <chr>                   <dbl>           <dbl> <int>         <dbl>       <dbl>
## 1 Album                       5               2     4        0.4         NA
## 2 Carousel                17854.            727.     9        0.0378      0.0239
## 3 Link                     2473.             24.0   94        0.0375      0.0971
## 4 Mixed                   36997.           2733.     5        0.108       0.0557
## 5 Photo                   16068.           1455.  1487        0.0592      0.0867
## 6 Text                     3413.            271.   910        0.0408      0.0438
## 7 Video                   10956.            889.   967        0.0535      0.0588
```

Despite removing the outliers, photos still seem to be performing the best out of the four main media types. They still have the highest average impressions and the highest average engagements from those impressions. However, this is data gathered from ALL accounts, so for strategic decisions regarding EACH account, I would redo this analysis and now subgroup by each account (which I do for question #6).

Question 5.

What is our best performing campaign?

```
df %>%
  filter(`Campaign Name` != "N/A") %>%
  group_by(`Campaign Name`) %>%
  summarize(mean_impressions = mean(`Total Impressions`),
            mean_engagement = mean(`Total Engagements`),
            n = n(),
            mean_eng_rate = mean(`Total Engagements`[`Total Impressions` != 0]/`Total Impressions`[`Tota
            sd_eng_rate = sd(`Total Engagements`[`Total Impressions` != 0]/`Total Impressions`[`Total In
```

```
## # A tibble: 3 x 6
##   `Campaign Name`      mean_impressions mean_engagement     n mean_eng_rate
##   <chr>                           <dbl>           <dbl> <int>         <dbl>
## 1 Community Engagement           16870.           1449.  1411        0.0577
## 2 Evergreen                       8152.            265.   163        0.0351
## 3 Evil Exhibited                 13673.            441.   420        0.0287
## # i 1 more variable: sd_eng_rate <dbl>
```

The best performing campaign is Community Engagement, as its posts gather the highest average impressions and highest average engagements. And the posts also have the highest average engagement rate.

Question 6.

Define out a posting strategy for our social channels based on your discoveries.

```
df %>%
  filter(`Media Type` %in% c("Photo", "Text", "Video", "Link"),
         `Total Engagements` <= `Total Impressions`) %>%
  group_by(Account, `Media Type`) %>%
  summarize(mean_impressions = mean(`Total Impressions`),
            mean_engagement = mean(`Total Engagements`),
            n = n(),
            mean_eng_rate = mean(`Total Engagements`[`Total Impressions` != 0]/`Total Impressions`[`Tota
            sd_eng_rate = sd(`Total Engagements`[`Total Impressions` != 0]/`Total Impressions`[`Total In
```

```
## # A tibble: 18 x 7
## # Groups:   Account [5]
##    Account      `Media Type` mean_impressions mean_engagement     n mean_eng_rate
##    <chr>        <chr>                   <dbl>           <dbl> <int>         <dbl>
##  1 CSGO         Link                     395.            4.85    20        0.0117
##  2 CSGO         Photo                  11087.          346.     139        0.0234
##  3 CSGO         Text                     623.           21.6     62        0.0345
##  4 CSGO         Video                  14821.          875.      49        0.110
##  5 Content Cr~ Link                        0             0        3      NaN
##  6 Content Cr~ Text                        0             0       12      NaN
##  7 Content Cr~ Video                    1189.           36.5     38        0.0617
##  8 DOTA2        Link                     981.           57.7      9        0.0561
##  9 DOTA2        Photo                   3236.          207.     331        0.0519
## 10 DOTA2        Text                    1149.           47.8    343        0.0364
## 11 DOTA2        Video                   3188.          314.     118        0.0760
## 12 General      Link                    3596.           27.4     60        0.0336
## 13 General      Photo                  20926.         2012.    1017        0.0677
## 14 General      Text                    5593.          478.     478        0.0468
## 15 General      Video                  13107.         1081.     719        0.0467
## 16 Valorant     Link                       0             0        2      NaN
## 17 Valorant     Text                       0             0       15      NaN
## 18 Valorant     Video                    534.           21.7     43        0.0534
## # i 1 more variable: sd_eng_rate <dbl>
```

(keep in mind that the NA values come for engagement rate come from having no posts with any impressions)

It's imperative to be more specific when it comes to making actual strategic decisions, so I decided to group by both account and media type first. After grouping by both account type and media type, we get different results on which media type performs the best depending on the account. It seems as if for game titles, videos seem to generate the most impressions, engagements, and engagement rate. This would make sense as videos seem like the effective way to generate hype for upcoming esports matches.

It's also important to make sure that there are no confounding sources found within time and day as well. Although this could very difficult to measure simply due to timing of certain key events. For instance, if DOTA2 matches happen more often on Friday, than obviously certain Fridays would have a higher impressions or engagements, leading to Fridays being seen as the most popular days for posting. Keep in mind that we found that there were no noticeably significant differences in day from question 2.

```r
df_time %>%
  filter(`Media Type` %in% c("Photo", "Text", "Video", "Link"),
         `Total Engagements` <= `Total Impressions`,
         `Total Impressions` != 0) %>%
  group_by(Account, day) %>%
  summarize(engagement_rate = mean(`Total Engagements`/`Total Impressions`)) %>%
  arrange(desc(engagement_rate), .by_group = T)
```

```
## # A tibble: 5 x 3
## # Groups:   Account [5]
##   Account          day   engagement_rate
##   <chr>            <chr>           <dbl>
## 1 CSGO             <NA>           0.0413
## 2 Content Creators <NA>           0.0617
## 3 DOTA2            <NA>           0.0491
## 4 General          <NA>           0.0561
## 5 Valorant         <NA>           0.0534
```

Unlike our answer to question 2, there is significant difference in posting day when it comes to content creators. However, our conclusion of no significant difference in day still stays the same for all game titles and the general account.

```
df_time %>%
  filter(`Media Type` %in% c("Photo", "Text", "Video", "Link"),
         `Total Engagements` <= `Total Impressions`,
         `Total Impressions` != 0) %>%
  group_by(Account, hour) %>%
  summarize(engagement_rate = mean(`Total Engagements`/`Total Impressions`)) %>%
  arrange(desc(engagement_rate)) %>%
  slice(1:3) # Don't want that many outputs
```

```
## # A tibble: 15 x 3
## # Groups:   Account [5]
##    Account          hour  engagement_rate
##    <chr>            <chr>           <dbl>
##  1 CSGO             08             0.0839
##  2 CSGO             07             0.0633
##  3 CSGO             14             0.0577
##  4 Content Creators 12             0.142
##  5 Content Creators 14             0.0601
##  6 Content Creators 09             0.0482
##  7 DOTA2            19             0.0751
##  8 DOTA2            04             0.0559
##  9 DOTA2            08             0.0556
## 10 General          05             0.267
## 11 General          04             0.0918
## 12 General          06             0.0897
## 13 Valorant         09             0.0734
## 14 Valorant         12             0.0695
## 15 Valorant         15             0.0643
```

Just like our conclusion for question 2, time of day generally does seem to matter for each account.

Overall, for game titles, to garner the highest engagement rate, I would recommend posting primarily videos, with the time differing for each game title. I would recommend for CSGO, around 8 AM; for DOTA2, around 7 PM; and for Valorant, around 9 AM. As for day, I believe that it doesn't that much, but if you had to pick: Thursday for CSGO, Tuesday for DOTA2, and Friday for Valorant.

For content creators, I would highly recommend posting videos on Saturdays at 12 PM to garner the highest engagement rate. As for general posts, I would recommend posting photos at 5 pm. I also believe that for the general account, day really doesn't matter.

Question 7.

What suggestions would you give to the social media team if they want to expand their presence (e.g. if our CSGO youtube channel is doing well should we expand to TikTok)?

Let's first start with the CSGO account

```
df %>%
  filter(Account == "CSGO") %>%
  count(`Campaign Name`)
```

```
## # A tibble: 2 x 2
##   `Campaign Name`           n
##   <chr>                 <int>
## 1 Community Engagement    180
## 2 N/A                      90
```

```
df %>%
  filter(Account == "CSGO") %>%
  count(`Account Type`)
```

```
## # A tibble: 1 x 2
##   `Account Type`     n
##   <chr>          <int>
## 1 TWITTER          270
```

The CSGO account only seems to be active on twitter and posts community engagement posts. A good expansion for the CSGO account is on another platform with a high engagement rate for community engagement posts.

```
df %>%
  filter(`Campaign Name` == "Community Engagement",
         `Total Impressions` != 0) %>%
  group_by(`Account Type`) %>%
  summarize(engagement_rate = mean(`Total Engagements`/`Total Impressions`))
```

```
## # A tibble: 6 x 2
##   `Account Type`   engagement_rate
##   <chr>                      <dbl>
## 1 FBPAGE                    0.160
## 2 INSTAGRAM                 0.0121
## 3 LINKEDIN_COMPANY          0.0186
## 4 TIKTOK_BUSINESS           0.0642
## 5 TWITTER                   0.0348
## 6 YOUTUBE                   0.0824
```

Community engagement posts seem to be doing exceptionally well on Facebook and decently well on Youtube. So I would suggest the CSGO account to expand to those two platforms.

We can apply the same logic to the other two game titles:

```
df %>%
  filter(Account == "Valorant") %>%
  count(`Campaign Name`)
```

```
## # A tibble: 2 x 2
##   `Campaign Name`     n
##   <chr>           <int>
## 1 Evil Exhibited     57
## 2 N/A                 3
```

```
df %>%
  filter(Account == "Valorant") %>%
  count(`Account Type`)
```

```
## # A tibble: 1 x 2
##   `Account Type`     n
##   <chr>          <int>
## 1 YOUTUBE           60
```

```
df %>%
  filter(`Campaign Name` == "Evil Exhibited",
         `Total Impressions` != 0) %>%
  group_by(`Account Type`) %>%
  summarize(engagement_rate = mean(`Total Engagements`/`Total Impressions`))
```

```
## # A tibble: 6 x 2
##   `Account Type`   engagement_rate
##   <chr>                      <dbl>
## 1 FBPAGE                   0.0351
## 2 INSTAGRAM                0.00815
## 3 LINKEDIN_COMPANY         0.0321
## 4 TIKTOK_BUSINESS          0.0377
## 5 TWITTER                  0.0299
## 6 YOUTUBE                  0.0502
```

Evil exhibited posts seem to already be doing the best on Youtube. If I had to recommend one other expansion platform for the Valorant account I would choose TikTok, although Facebook, Linkedin, and Twitter have similar engagement rates. However, I would NOT recommend Instagram.

```
df %>%
  filter(Account == "DOTA2") %>%
  count(`Campaign Name`)
```

```
## # A tibble: 3 x 2
##   `Campaign Name`           n
##   <chr>                 <int>
## 1 Community Engagement    178
## 2 Evil Exhibited           67
## 3 N/A                     558
```

```
df %>%
  filter(Account == "DOTA2") %>%
  count(`Account Type`)
```

```
## # A tibble: 2 x 2
##   `Account Type`     n
##   <chr>          <int>
## 1 TWITTER          795
## 2 YOUTUBE            8
```

```r
df %>%
  filter(`Campaign Name` == "N/A",
         `Total Impressions` != 0,
         `Total Engagements` <= `Total Impressions`) %>%
  group_by(`Account Type`) %>%
  summarize(engagement_rate = mean(`Total Engagements`/`Total Impressions`))
```

```
## # A tibble: 6 x 2
##   `Account Type`   engagement_rate
##   <chr>                      <dbl>
## 1 FBPAGE                   0.170
## 2 INSTAGRAM                0.00434
## 3 LINKEDIN_COMPANY         0.00930
## 4 TIKTOK_BUSINESS          0.0928
## 5 TWITTER                  0.0478
## 6 YOUTUBE                  0.0332
```

The DOTA2 seems to post a lot of non campaign content, however the most prevalent is Community Engagement. So, similar to CSGO I would recommend DOTA2 to expand to facebook as well. Facebook and TikTok also seem to be doing prominently well for non campaign posts, so I could also recommend TikTok as well.

Finally, we will find the best expansion for content creators.

```r
df %>%
  filter(Account == "Content Creators") %>%
  count(`Campaign Name`)
```

```
## # A tibble: 3 x 2
##   `Campaign Name`          n
##   <chr>                <int>
## 1 Community Engagement     7
## 2 Evergreen                2
## 3 N/A                     44
```

```r
df %>%
  filter(Account == "Content Creators") %>%
  count(`Account Type`)
```

```
## # A tibble: 1 x 2
##   `Account Type`     n
##   <chr>          <int>
## 1 YOUTUBE           53
```

Content creators also seem to have a lot of non campaign posts, so similar to DOTA2, I would recommend expanding to Facebook and TikTok.