

Midterm Project Proposal

Eduardo Carrasco & Logan Lauton

2021-03-19

Section 1 - Introduction

With the rising popularity in E-sports, and E-sports related multi media, we wanted to do further research that shows why and how this new form of media has taken off. The question that we are going to answer, in broad terms, is what main statistic has the biggest impact on total Followers that a streamer has. We will be classifying a “main statistic” as being Primary Language, Watch Time, Stream time, Average viewers, Views Gained, Twitch Partnered or not, & 18+ Stream or not. We also both enjoy watching e-sports related media, and wish to see what gives these players and streamers the biggest gain in followers.

Section 2 - Data

The data consists of Channel Name, Watch Time, Stream Time, Peak viewers, Average viewers, Followers, Followers gained, Views gained, Partnered, Mature, Language. The data that we will be focusing on is *Watch Time, Stream Time, Average viewers, Followers, Views gained, Partnered and Mature Language*. I'll also go ahead and define a few statistics now, just so there is no confusion. Watch time is defined as the total time watched on ones stream(s). Peak Viewers is defined as the maximum amount of viewers one has had on any given stream. Views gained is in based on the amount of views that any given streamer had gained in the last year, as the data set that we are using is based on data of Top 1000 Streamers from past year. Partnered refers to The Twitch Partnership Program, which is for those who are committed to streaming and are ready to level up from Affiliate. When Partnered, you recieve monetization benefits, which means that Partners can earn revenue by accepting subscriptions from their viewers. They also can recieve virtual currency known as Bits, and they also have the right to play Ads to increase their revnue.¹

Esports Data Analysis

Initial Data Explortation

```
data <- read_csv("data/twitchdata-update.csv")
```

```
##
## -- Column specification -----
## cols(
##   Channel = col_character(),
##   `Watch time(Minutes)` = col_double(),
##   `Stream time(minutes)` = col_double(),
##   `Peak viewers` = col_double(),
```

¹[Twitch Partner Program Overview. Twitch. Accessed March 19, 2021.] (https://help.twitch.tv/s/article/partner-program-overview?language=en_US#:~:text=The%20Twitch%20Partnership%20Program%20is,anything%20else%20you%20can%20imagine.)

```
## `Average viewers` = col_double(),
## Followers = col_double(),
## `Followers gained` = col_double(),
## `Views gained` = col_double(),
## Partnered = col_logical(),
## Mature = col_logical(),
## Language = col_character()
## )
```

```
data <- clean_names(data, case = "snake")
```

```
head(data)
```

```
## # A tibble: 6 x 11
##   channel    watch_time_minutes stream_time_minutes peak_viewers average_viewers
##   <chr>          <dbl>          <dbl>          <dbl>          <dbl>
## 1 xQcOW          6196161750          215250          222720          27716
## 2 summit1g      6091677300          211845          310998          25610
## 3 Gaules        5644590915          515280          387315          10976
## 4 ESL_CSGO      3970318140          517740          300575           7714
## 5 Tfue          3671000070          123660          285644          29602
## 6 Asmongold     3668799075           82260          263720          42414
## # ... with 6 more variables: followers <dbl>, followers_gained <dbl>,
## #   views_gained <dbl>, partnered <lgl>, mature <lgl>, language <chr>
```

```
summary(data)
```

```
##   channel          watch_time_minutes stream_time_minutes peak_viewers
## Length:1000      Min.   :1.222e+08   Min.    : 3465      Min.    : 496
## Class :character 1st Qu.:1.632e+08   1st Qu.: 73759      1st Qu.: 9114
## Mode  :character Median :2.350e+08   Median :108240      Median : 16676
##                Mean   :4.184e+08   Mean   :120515      Mean    : 37065
##                3rd Qu.:4.337e+08   3rd Qu.:141844      3rd Qu.: 37570
##                Max.   :6.196e+09   Max.    :521445      Max.    :639375
## average_viewers followers      followers_gained views_gained
## Min.   : 235   Min.    : 3660   Min.    : -15772   Min.    : 175788
## 1st Qu.: 1458   1st Qu.: 170546   1st Qu.: 43758    1st Qu.: 3880602
## Median : 2425   Median : 318063   Median : 98352    Median : 6456324
## Mean   : 4781   Mean   : 570054   Mean   : 205519    Mean   : 11668166
## 3rd Qu.: 4786   3rd Qu.: 624332   3rd Qu.: 236131    3rd Qu.: 12196762
## Max.   :147643   Max.    :8938903   Max.    :3966525    Max.    :670137548
## partnered      mature          language
## Mode :logical  Mode :logical  Length:1000
## FALSE:22      FALSE:770      Class :character
## TRUE :978      TRUE :230      Mode  :character
##
##
##
```

```
lang_table <- table(data$language)
lang_df <- as.data.frame(lang_table)
top_ten_langauges <- order(lang_df$Freq,decreasing = TRUE)[1:10]
top_ten_langauges <- lang_df[top_ten_langauges,]
```

```
#Pie Chart of the top10 languages on twitch
```

```
ggplot(top_ten_langauges, aes(x = "", y = Freq, fill = Var1))+
```

```
geom_bar(stat = "identity", width = 1)+
coord_polar("y", start = 0)+
theme(legend.title = element_blank(),
      axis.title.x = element_blank(),
      axis.title.y = element_blank())
```

