Kira Damo
2007841
DSC 423
Due November 19th

Modeling the Evolution of Twitch Viewership During the COVID-19 Pandemic

Introduction

The World Health Organization officially declared COVID-19 a pandemic on March 11, 2020. There were over 118,000 cases of COVID-19 in 114 countries and 4,291 deaths resulting in a global panic and enforcements of quarantine. Mandatory lockdowns confined people to their homes and social distancing was enforced in all public spaces. People started to seek refuge on the internet to nearly fulfill the social life they had.

Twitch was one of the most optimal ways to remedy the withdrawals of social interaction and the yearning for a community or a social circle. Twitch is a livestreaming platform that is widely known for gaming and e-sports content. Twitch users with the necessary equipment and software can broadcast themselves, referred to as streamers. The viewers can also engage in chat, interact with the stream, and gift the streamer subscriptions (ad-free viewing of streamer for other viewers and channel specific assets such as emotes and sub-badges). The platform skyrocketed in popularity during the pandemic.

Objective

Although majority of Twitch's content spotlights video games, other categories rose in popularity like Just Chatting, ASMR, Food & Drink, and Art. As more people joined Twitch, the streaming landscape was diversifying. Twitch served as a medium to bridge connections between streamers and viewers. Users were able to bond over shared interests and have a distraction to the political, social, and economic turmoil of the pandemic. This project will explore the changes in the top categories during the pandemic and how that correlates to current events and cultural shifts offline.

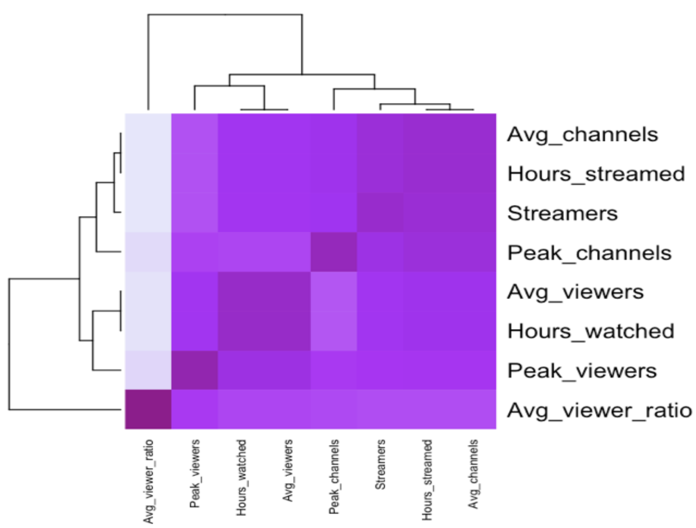
Dataset

This dataset consists of the top 200 streamed categories from January 2016 to September 2024. In total there are 21,000 data points. The quantitative variables are rank, month, year, hours watched, hours streamed, peak viewers, peak channels, streamers, average viewers, average channels, and average viewer ratio. This dataset was found on Kaggle and was extracted using web-sparsing on sully-gnome which is a real time database on Twitch metrics.

There are 12 variables total in this dataset. The only qualitative variable is game which is the categories that are streamed. All these variables play a role in classifying virality of categories and content. To forecast the engagement and the transformation of Twitch during the pandemic, the variable that will need to predicted is the hours watched. The attraction and retention of a category determine its promotion to the top 200. To measure the time, the month and year were combined into a column and is read as date values. Building a regression model will narrow down the most efficient variables to determine the hours watched.

Methodology

To find the best linear regression model, the first step is variable screening. Creating a matrix and a heatmap makes the redundant and highly correlated variables easier to identify to be removed. The deeper purple shades in the heatmap indicate higher correlation whereas the lighter shades of purple are less correlated:



The hours watched, average viewers, and peak viewers is redundant and is an iteration of the same data. Likewise, hours streamed, streamers, average channels, and peak channels is redundant. Average viewer ratio is not highly correlated and is a unique variable that describes the engagement between streamers and the volume of audience that is watching them. If all variables were to be kept from the data, the model would suffer overfitting. Overfitting the data leaves room for misguided conclusions and false statistics. The first model is a first-order model with all variables included.

$$\text{Hours\_watched} = -9.89 \times 10^{3} + 2.57 \times 10^{1} \cdot \text{Hours\_streamed} + 1.02 \times 10^{-1} \cdot \text{Peak\_viewers} - 4.19 \times 10^{0} \cdot \text{Peak\_channels} - 2.14 \times 10^{-1} \cdot \text{Streamers} + 7.29 \times 10^{2} \cdot \text{Avg\_viewers} - 1.87 \times 10^{4} \cdot \text{Avg\_channels} - 1.81 \times 10^{-1} \cdot \text{Avg\_viewer\_ratio}$$

```
        Call:
lm(formula = Hours_watched ~ Hours_streamed + Peak_viewers +
    Peak_channels + Streamers + Avg_viewers + Avg_channels +
    Avg_viewer_ratio, data = twitch)

Residuals:
      Min        1Q    Median        3Q       Max
-16841075    -13406      -323     12220   9616888

Coefficients:
                   Estimate Std. Error  t value Pr(>|t|)
(Intercept)      -9.889e+03  2.766e+03   -3.575 0.000351 ***
Hours_streamed    2.570e+01  1.497e-01  171.665  < 2e-16 ***
Peak_viewers      1.022e-01  2.354e-02    4.343 1.41e-05 ***
Peak_channels    -4.187e+00  1.618e+00   -2.587 0.009688 **
Streamers        -2.139e-01  9.827e-02   -2.176 0.029534 *
Avg_viewers       7.286e+02  1.684e-01 4327.447  < 2e-16 ***
Avg_channels     -1.871e+04  1.094e+02 -171.075  < 2e-16 ***
Avg_viewer_ratio -1.805e-01  6.089e+00   -0.030 0.976348
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 361500 on 20992 degrees of freedom
Multiple R-squared:  0.9997,      Adjusted R-squared:  0.9997
F-statistic: 8.663e+06 on 7 and 20992 DF,  p-value: < 2.2e-16
```

The overfitting is foreshadowed in the R-squared value that is extremely close to 1. To improve this model, the redundant variables are dropped to prevent multicollinearity. The model is still in first-order but has now has linearly independent variables.

$$\text{Hours\_watched} = 1.002 \times 10^{6} + 2.509 \times 10^{1} \cdot \text{Hours\_streamed} + 7.974 \times 10^{2} \cdot \text{Avg\_viewer\_ratio}$$

```
Call:
lm(formula = Hours_watched ~ Hours_streamed + Avg_viewer_ratio,
    data = twitch)

Residuals:
       Min         1Q     Median         3Q        Max
-135858125   -1544417    -964265    -412824  228351084
```

```
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      1.002e+06  9.736e+04  10.287  < 2e-16 ***
Hours_streamed   2.509e+01  1.593e-01 157.556  < 2e-16 ***
Avg_viewer_ratio 7.974e+02  2.207e+02   3.612 0.000304 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 13150000 on 20997 degrees of freedom
Multiple R-squared:  0.5418,	Adjusted R-squared:  0.5418 
F-statistic: 1.241e+04 on 2 and 20997 DF,  p-value: < 2.2e-16
```

Although the adjusted R-squared value is now 0.5418 and accounts for 54.18% of the variance in the data, this model is still significant due to the p-value being less than the default level of significance $\alpha = 0.05$. The large F-statistic and the t-values being greater than 2 also show that the model passes the F-test and the t-tests.

To increase the R-squared value without overfitting, second-order terms will be introduced. Since Twitch received an explosive growth in the year of 2020, a second-order model will account for more of the variance in data than a first-order model does. To create this model, both independent variables will be turned into polynomial terms.

$$\begin{aligned}\text{Hours\_watched} = {} & 2.08 \times 10^9 \cdot \text{poly(Hours\_streamed, 2)}_1 - 7.18 \times 10^8 \cdot \text{poly(Hours\_streamed, 2)}_2 \\ & + 7.20 \times 10^7 \cdot \text{poly(Avg\_viewer\_ratio, 2)}_1 - 8.65 \times 10^7 \cdot \text{poly(Avg\_viewer\_ratio, 2)}_2\end{aligned}$$

```
Call:
lm(formula = Hours_watched ~ poly(Hours_streamed, 2) + poly(Avg_viewer_ratio, 
    2), data = twitch)

Residuals:
       Min         1Q     Median         3Q        Max 
-105422554   -1045035     353438     797225  230809943 

Coefficients:
                             Estimate Std. Error t value Pr(>|t|)
(Intercept)                   5831853      84037  69.396  < 2e-16 ***
poly(Hours_streamed, 2)1   2077809876   12194556 170.388  < 2e-16 ***
poly(Hours_streamed, 2)2   -718256914   12191273 -58.916  < 2e-16 ***
poly(Avg_viewer_ratio, 2)1   72018824   12193663   5.906 3.55e-09 ***
poly(Avg_viewer_ratio, 2)2  -86457488   12192166  -7.091 1.37e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 12180000 on 20995 degrees of freedom
Multiple R-squared:  0.6073,	Adjusted R-squared:  0.6072 
```

```
F-statistic:  8115 on 4 and 20995 DF,  p-value: < 2.2e-16
```

This model is still deemed significant since the F-test and t-tests passed and the p-value is still less than $\alpha = 0.05$. The R-squared value in this model is greater than the previous model and the RSE decreased as well. The change in these two values demonstrates that this model is a better fit to the data than the first-order model.

Since this is a quadratic model, the introduction of interaction terms may create a model that the data will adhere to more or the new terms will cause the model to overfit the data. This model is still in second-order but now has interaction terms.

$$\begin{aligned}\text{Hours\_watched} = {} & 6.34 \times 10^{9} \cdot \text{poly}(\text{Hours\_streamed}, 2)_1 - 1.75 \times 10^{4} \cdot \text{poly}(\text{Hours\_streamed}, 2)_2 \\ & + 1.13 \times 10^{10} \cdot \text{poly}(\text{Avg\_viewer\_ratio}, 2)_1 - 1.44 \times 10^{2} \cdot \text{poly}(\text{Avg\_viewer\_ratio}, 2)_2 \\ & + 4.93 \times 10^{12} \cdot (\text{poly}(\text{Hours\_streamed}, 1) \times \text{poly}(\text{Avg\_viewer\_ratio}, 1))\end{aligned}$$

```
Call:
lm(formula = Hours_watched ~ poly(Hours_streamed, 2) + poly(Avg_viewer_ratio,
    2) + poly(Hours_streamed, 1):poly(Avg_viewer_ratio, 1), data = twitch)

Residuals:
   Min     1Q Median     3Q    Max
-38451    -76     20     96  51991

Coefficients:
                                                  Estimate Std. Error    t value
(Intercept)                                      1.458e+07  1.495e+01  9.753e+05
poly(Hours_streamed, 2)1                         6.340e+09  4.601e+03  1.378e+06
poly(Hours_streamed, 2)2                        -1.746e+04  1.901e+03 -9.184e+00
poly(Avg_viewer_ratio, 2)1                       1.133e+10  1.136e+04  9.971e+05
poly(Avg_viewer_ratio, 2)2                      -1.443e+02  1.764e+03 -8.200e-02
poly(Hours_streamed, 1):poly(Avg_viewer_ratio, 1)  4.928e+12  4.913e+06  1.003e+06
                                                Pr(>|t|)
(Intercept)                                       <2e-16 ***
poly(Hours_streamed, 2)1                          <2e-16 ***
poly(Hours_streamed, 2)2                          <2e-16 ***
poly(Avg_viewer_ratio, 2)1                        <2e-16 ***
poly(Avg_viewer_ratio, 2)2                         0.935
poly(Hours_streamed, 1):poly(Avg_viewer_ratio, 1)   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1759 on 20994 degrees of freedom
Multiple R-squared:      1,	Adjusted R-squared:      1
F-statistic: 5.122e+11 on 5 and 20994 DF,  p-value: < 2.2e-16
```

The model suffers overfitting the predictive values with the data. The second polynomial term of the average viewer ratio fails the p-test. Although the F-test passes and the model has a p-value less than $\alpha = 0.05$, the R-squared value is exactly 1. To try to remedy this, the insignificant term will be dropped.

To determine which regression model is best, the model comparison is performed according to the Akaike Information Criterion (AIC). The AIC will gauge which model is the best for predicting hours watched. The AIC measures the quality of the model with the number of parameters.

Model	Degrees of Freedom	AIC
Model 1 (first-order model)	9	597122.7
Model 2 (quadratic model with no interaction terms)	6	744838.9
Model 3 (quadratic model with interaction terms)	7	373458.2

These values can be misleading considering the previous steps taken and the evaluation of their summaries. The relationship between the independent and dependent variables in the model is not linear making the first-order model unsuitable. Model 2's R-squared is the lowest and accounts for the least amount of variance even though all other tests passed. On the other hand, Model 3 has an R-squared value of 1 which may be a sign of overfitting. To conclude the variable selection, step AIC will be performed on both Model 2 and Model 3.

StepAIC for Model 2:

```
Start:  AIC=685241.5
Hours_watched ~ poly(Hours_streamed, 2) + poly(Avg_viewer_ratio, 
    2)

                            Df  Sum of Sq        RSS
<none>                                    3.1137e+18
- poly(Avg_viewer_ratio, 2)  2 1.2601e+16 3.1263e+18
- poly(Hours_streamed, 2)    2 4.8133e+18 7.9271e+18
                               AIC
<none>                      685241
- poly(Avg_viewer_ratio, 2) 685322
- poly(Hours_streamed, 2)   704861
```

```
Call:
lm(formula = Hours_watched ~ poly(Hours_streamed, 2) + poly(Avg_viewer_ratio,
    2), data = twitch)

Coefficients:
              (Intercept)    poly(Hours_streamed, 2)1
                  5831853                  2077809876
  poly(Hours_streamed, 2)2  poly(Avg_viewer_ratio, 2)1
               -718256914                    72018824
poly(Avg_viewer_ratio, 2)2
                -86457488
```

StepAIC for Model 3:

```
Start:  AIC=313860.8
Hours_watched ~ poly(Hours_streamed, 2) + poly(Avg_viewer_ratio,
    2) + poly(Hours_streamed, 1):poly(Avg_viewer_ratio, 1)

                                                    Df  Sum of Sq        RSS
<none>                                                             6.4987e+10
- poly(Avg_viewer_ratio, 2)                          2 3.0850e+18 3.0850e+18
- poly(Hours_streamed, 1):poly(Avg_viewer_ratio, 1)  1 3.1137e+18 3.1137e+18
- poly(Hours_streamed, 2)                            2 6.6844e+18 6.6844e+18
                                                       AIC
<none>                                              313861
- poly(Avg_viewer_ratio, 2)                         685045
- poly(Hours_streamed, 1):poly(Avg_viewer_ratio, 1) 685241
- poly(Hours_streamed, 2)                           701283

Call:
lm(formula = Hours_watched ~ poly(Hours_streamed, 2) + poly(Avg_viewer_ratio,
    2) + poly(Hours_streamed, 1):poly(Avg_viewer_ratio, 1), data = twitch)

Coefficients:
                                      (Intercept)
                                        1.458e+07
                         poly(Hours_streamed, 2)1
                                        6.340e+09
                         poly(Hours_streamed, 2)2
                                       -1.746e+04
                       poly(Avg_viewer_ratio, 2)1
                                        1.133e+10
                       poly(Avg_viewer_ratio, 2)2
                                       -1.443e+02
poly(Hours_streamed, 1):poly(Avg_viewer_ratio, 1)
                                        4.928e+12
```

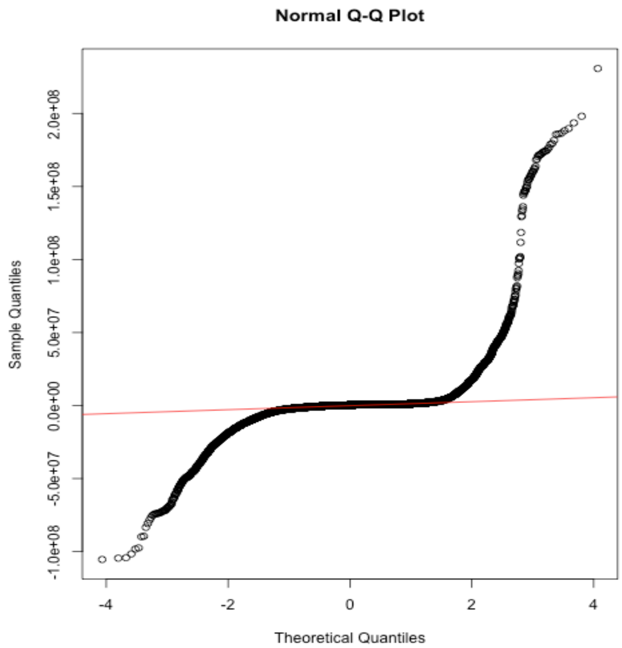
Both Step AIC results states that to minimize the AIC value, all variables in both models must be kept. This narrows down our option to go forward with Model 2 to prevent overfitting and to account for the exponential growth of Twitch viewership during the pandemic.

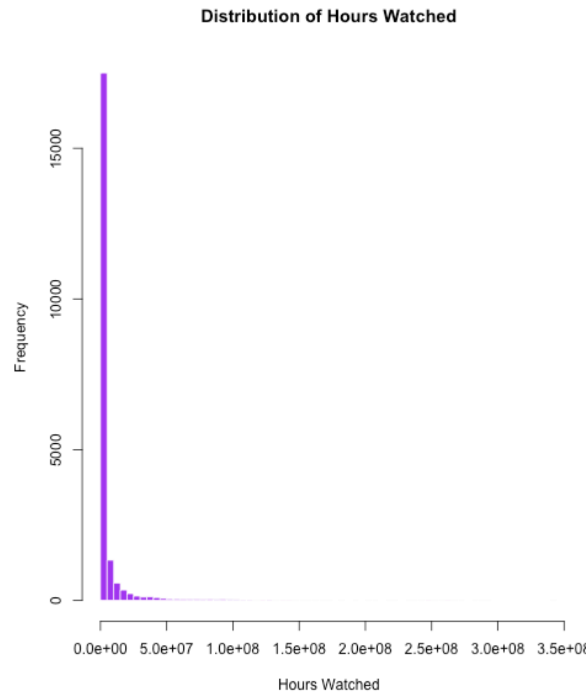
How would this model react in another 10 years? What if another pandemic spread across the world again (knock on wood)? 10-fold cross validation is performed to test the effectiveness of the model.

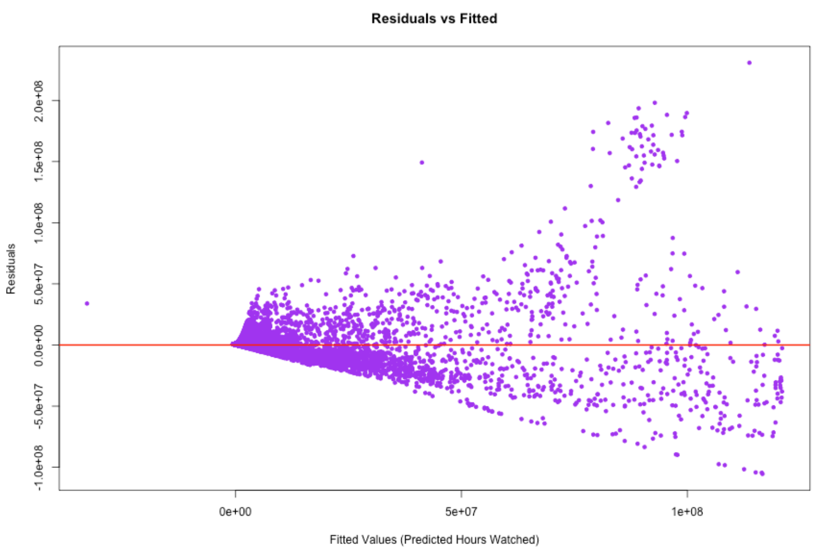
```
Linear Regression

21000 samples
    2 predictor

No pre-processing
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 18900, 18900, 18900, 18900, 18900, 18900, ...
Resampling results:

  RMSE      Rsquared   MAE
  12263308  0.6027475  3865818

Tuning parameter 'intercept' was held constant at a value of TRUE
```

The model folds the data 10 times and trains using the first 9 folds. The model is then tested on the 10^{th} fold. The percent error of the R-squared compared to model is less than 1%. The model is effective and can work with more data outside of the observed values.

A residual analysis will offer insight on how adequate the type of model is. Residuals are the distance between predicted values and actual values. The variance in a residual analysis should be random. The histogram of the hours watched, the residuals versus fitted graph, and the QQ-plot suggest that the data is not normally distributed and is right skewed. The fan shape of the residuals versus fitted plot also shows that the variance is not random and homoscedasticity is violated.







To resolve the violations in the residuals' plots, the dependent variable, hours watched, will receive a log transformation. This will make the hours watched closer to being normally distributed and the variance in the residuals more random. The new model is now logarithmic and includes second-order terms.

$$\log(\text{Hours\_watched} + 1) = 123.60 \cdot \text{poly}(\text{Hours\_streamed}, 2)_1 - 78.83 \cdot \text{poly}(\text{Hours\_streamed}, 2)_2 - 2.36 \cdot \text{poly}(\text{Avg\_viewer\_ratio}, 2)_1 + 0.94 \cdot \text{poly}(\text{Avg\_viewer\_ratio}, 2)_2 + 14.15$$

```
Call:
lm(formula = log(Hours_watched + 1) ~ poly(Hours_streamed, 2) +
    poly(Avg_viewer_ratio, 2), data = twitch)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.3626 -0.6634 -0.0441  0.6617 10.7305 

Coefficients:
                             Estimate Std. Error  t value Pr(>|t|)    
(Intercept)                 14.147681   0.007117 1987.957   <2e-16 ***
poly(Hours_streamed, 2)1   123.595852   1.032696  119.683   <2e-16 ***
poly(Hours_streamed, 2)2   -78.827324   1.032418  -76.352   <2e-16 ***
poly(Avg_viewer_ratio, 2)1  -2.363759   1.032620   -2.289   0.0221 *  
poly(Avg_viewer_ratio, 2)2   0.935902   1.032494    0.906   0.3647    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.031 on 20995 degrees of freedom
Multiple R-squared:  0.491,	Adjusted R-squared:  0.4909 
F-statistic:  5064 on 4 and 20995 DF,  p-value: < 2.2e-16
```

Now that a new model is introduced, the process of variable selection, cross-validation, and residual analysis will need to be performed again. The final model with excluded insignificant terms:

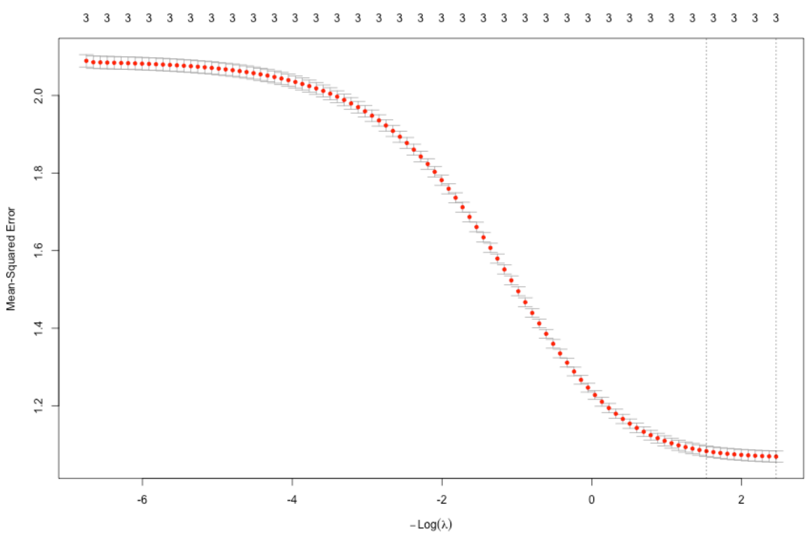
$$\begin{aligned}\log(\text{Hours\_watched} + 1) &\\ &= 123.63 \cdot \text{poly}(\text{Hours\_streamed}, 2)_1 - 78.86 \cdot \text{poly}(\text{Hours\_streamed}, 2)_2 - 2.36\\ &\cdot \text{poly}(\text{Avg\_viewer\_ratio}, 2)_1 + 14.15\end{aligned}$$

```
Call:
lm(formula = log(Hours_watched + 1) ~ poly(Hours_streamed, 2) +
    poly(Avg_viewer_ratio, 2)[, 1], data = twitch)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.3630 -0.6628 -0.0439  0.6620 10.7355 

Coefficients:
                                 Estimate Std. Error  t value Pr(>|t|)    
(Intercept)                     14.147681   0.007117 1987.966   <2e-16 ***
poly(Hours_streamed, 2)1       123.629585   1.032021  119.794   <2e-16 ***
poly(Hours_streamed, 2)2       -78.856981   1.031895  -76.420   <2e-16 ***
poly(Avg_viewer_ratio, 2)[, 1]  -2.361497   1.032613   -2.287   0.0222 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.031 on 20996 degrees of freedom
Multiple R-squared:  0.491,	Adjusted R-squared:  0.4909 
F-statistic:  6752 on 3 and 20996 DF,  p-value: < 2.2e-16
```

Although this model accounts for less of the variance than the previous model, this model does not violate homoscedasticity, has no multicollinear variables, and passes the F-test, t-tests, and p-test. Ridge and Lasso selection is performed to confirm the strength of the model and the variable selection in this model. This is especially effective since there are polynomial terms in the final model.

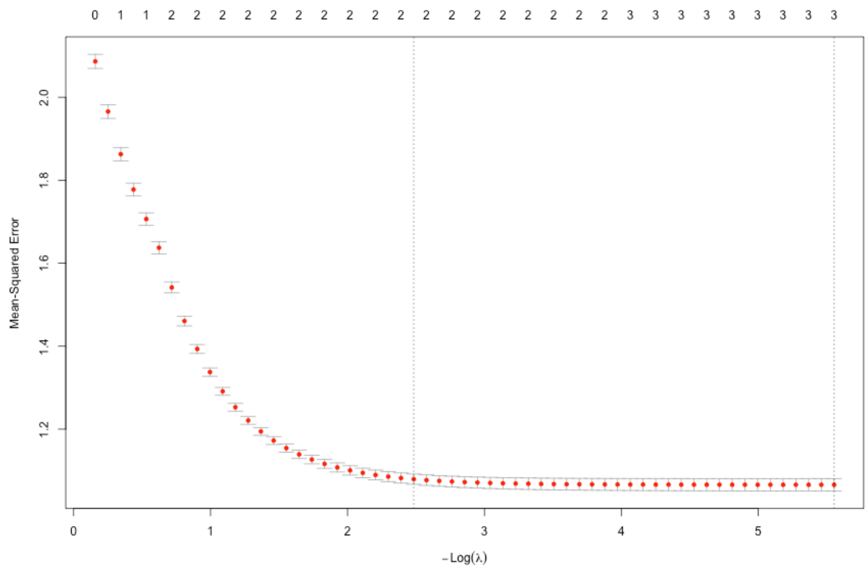
Ridge Coefficients:

```
4 x 1 sparse Matrix of class "dgCMatrix"
            s=0.08537327
(Intercept)    14.147681
h1            116.725784
h2            -74.451000
v1             -2.613704
```



Lasso Coefficients:

```
Call:  cv.glmnet(x = X, y = y, family = "gaussian", alpha = 1)

Measure: Mean-Squared Error

     Lambda Index Measure      SE Nonzero
min 0.00387    59   1.066 0.01497       3
1se 0.08341    26   1.079 0.01247       2
```



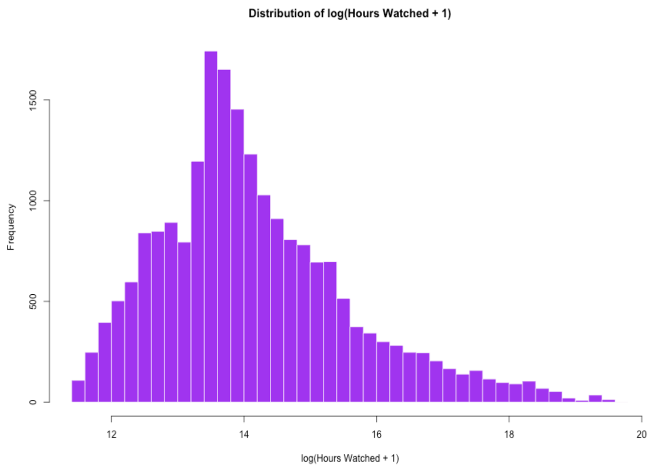
In the Ridge method, the coefficients for hours streamed and the average viewer ratio shrink slightly but are never unstable. All variables remain in the model. The same happens in the Lasso model. The coefficients shrink but never reach 0. The ridge lasso method confirms that all variables are significant in contributing to the prediction of the hours streamed.

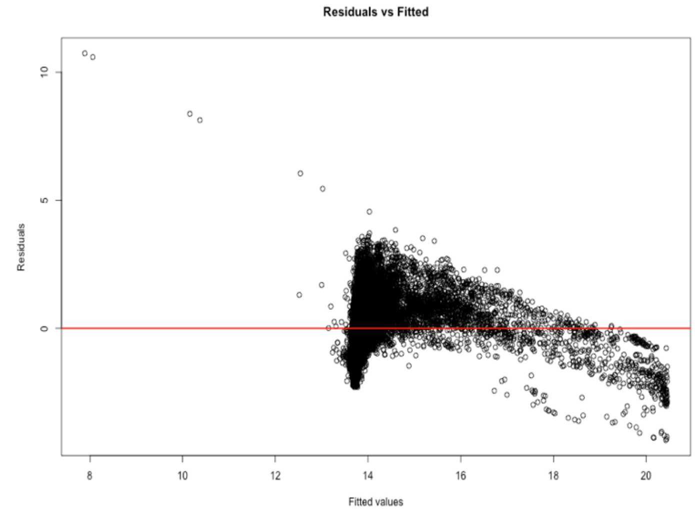
K-Fold Cross Validation of the Final Model:

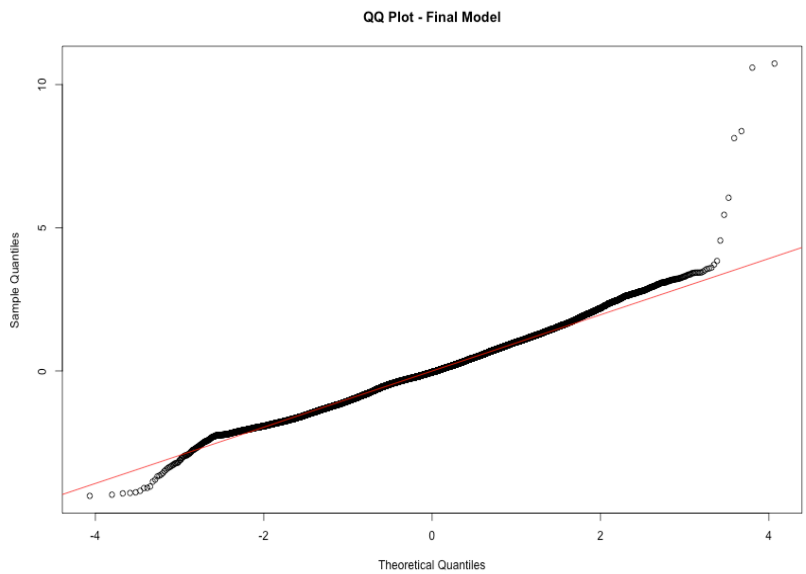
```
Linear Regression

21000 samples
    2 predictor

No pre-processing
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 18900, 18900, 18900, 18900, 18900, 18900, ...
Resampling results:

  RMSE      Rsquared   MAE
  1.032457  0.4905036  0.8041308

Tuning parameter 'intercept' was held constant at a value of TRUE
```

After performing 10-fold cross validation, the model is proven to be effective for data outside of the designated training set. The R-squared value has a percent error of less than 1.

Residuals of the final model:





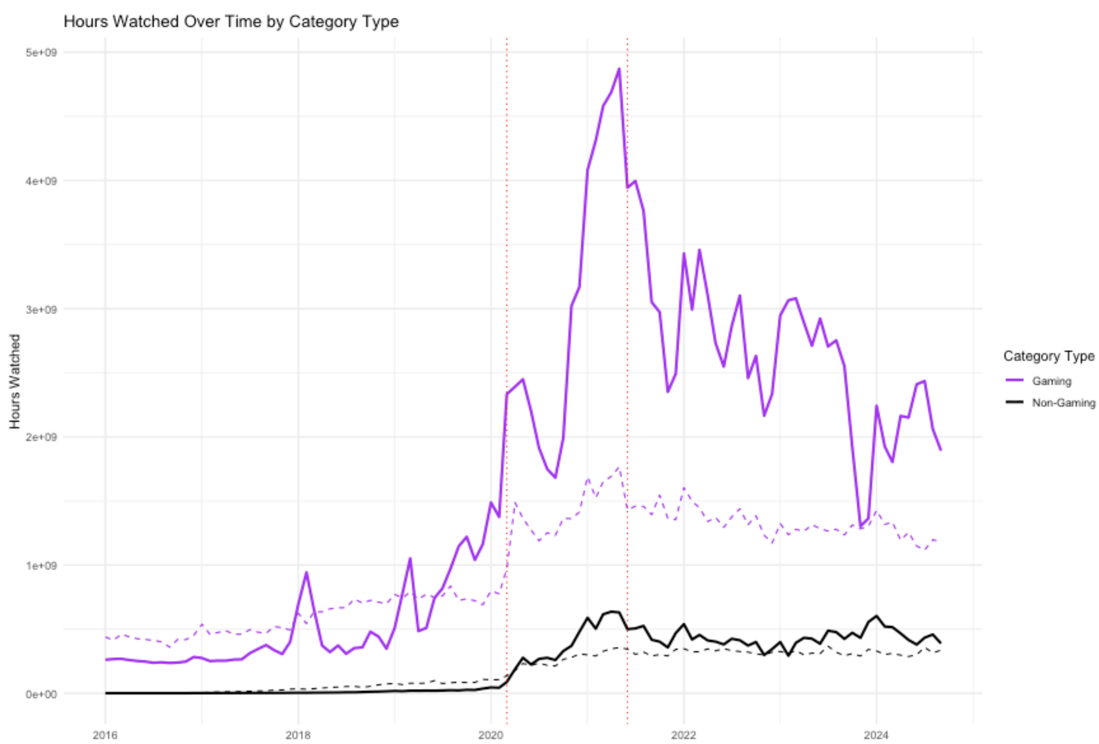


After performing the log transformation of hours watched, residuals are not as dramatically skewed as before. The QQ-plot is closer to a linear line indicating a closer fit to normal distribution. The residuals are also more randomly scattered across 0. There are outliers that occur on the residuals plots which must stay included since virality of content can vary and number of views does not have a ceiling.

Analysis

To predict the hours watched on Twitch throughout the timeline of the pandemic, several regression models were built and compared. The rapid increase in viewership during the pandemic was inevitable as people were limited to their devices to stay in touch and to network. Not only were people seeking out content to consume, but Twitch served as a creative outlet for users as well. There is a direct relationship with the hours streamed and the hours watched. The more content that is generated into the platform, the more there is available for users to watch whether it be livestreamed, videos-on-demand (VODs), or clips. The average viewer ratio measures the engagement on the platform between the audience and the streamer. Both variables are essential in ranking the top categories on Twitch.

In the 2,360 unique games that were ranked in the top 200, the following were considered as non-gaming categories: ASMR, Art, Just Chatting, Food & Drink, Special Events, Makers & Crafting, Travel & Outdoors, Science & Technology, IRL, Music, Politics, and Pools, Hot Tubs, and Beaches. Separating the hours watched into gaming and non-gaming content will predict the change in streaming categories.



The solid lines are the predicted values of hours watched using the regression model and the dashed lines are the hours watched. The dashed lines are the dates of the pandemic where lockdown was mandatory and social distancing was strongly enforced.

As the pandemic ended, viewership began to plateau as people started returning to work and physical social interactions were no longer scarce. The pandemic reinforced the grasp that the internet has on socializing. The pandemic established a codependency on internet relationships and reputations. These relationships and the emphasis of an online persona is now a component of daily life and identity.

Conclusion

Will non-gaming categories ever eclipse gaming? Most likely not. Although Just Chatting will always be ranked highly, the amount of gaming categories will always generate more total hours viewed. The pandemic also had an impact on gaming trends. The popularity of free-to-play multiplayer games like Valorant, Apex Legends, and Fortnite rose in popularity. Simpler friend-centered-content games started trending such as Among Us and Lethal Company. These trends persist even after the pandemic no longer posed as a threat.

The model accounts for the rise in hours watched during the pandemic but other events such as game releases, promotional events, and holidays also play a role in the increase in viewership. According to the model, if content is streamed, content will be watched. During a time of uncertainty, the amount of content to consume was never limited. The rise in viewership and the increase in non-gaming categories is a result of users adapting to the circumstances of a mandatory lockdown which changed the platform going forward.

Works Cited

Kirsh, R. (2024) *Top games on twitch 2016 - 2023*, *Kaggle*. Available at: https://www.kaggle.com/datasets/rankirsh/evolution-of-top-games-on-twitch (Accessed: 19 November 2025).

CDC Museum Covid-19 Timeline (2024) *Centers for Disease Control and Prevention*. Available at: https://www.cdc.gov/museum/timeline/covid19.html#Late-2019 (Accessed: 19 November 2025).

Northwestern Medicine (no date) *When did the pandemic start and end?*, *Northwestern Medicine*. Available at: https://www.nm.org/healthbeat/medical-advances/new-therapies-and-drug-trials/covid-19-pandemic-timeline (Accessed: 19 November 2025).