

MSCI 581 Marketing Analytics

Coursework

Submitted By

36141687

36366243

36141315

Presented to the Department of
Management School
Lancaster University Management School
Lancaster University

In Fulfilment of Requirements
For MSc. Business Analytics
Under the Guidance of Professor Ivan and Alisa

Lancaster University, United Kingdom

17th April 2023



Lancaster University
Management School

Quadruple-accredited world-ranked



Contents

Executive Summary.....	1
Preliminary Analysis	2
Data Cleaning and Gathering	2
Question 1.....	6
Question 2.....	7
Imitators and Innovators.....	7
Question 3.....	9
Pricing Strategy Efficiency.....	9
Question 4.....	11
Question 5.....	11

Executive Summary

This report analyses the relationship between the number of players and Twitch viewers of the game "Counter-Strike: Global Offensive" (CS:GO). The data was sourced from SteamDB and cleaned to only include player and Twitch viewer counts until March 1st, 2023.

The analysis revealed a surge in the number of players during the COVID-19 pandemic, which was accounted for in the regression model by adding a "Covid Flag" column. The impact of major championships on the player count was also analysed by adding a "Majors" column. The effect of the announcement of a new version of the game, "Counter-Strike 2," was analysed by adding a "CS2Flag" column. The correlation between the variables was analysed. The negative correlation between the price of the game and the number of players was also established, as was the positive correlation between the discount and the number of players. The impact of Twitch viewers on player count was analysed using a regression model, and it was found that Twitch viewers had a minimal, but positive effect on the number of players.

Next, a Bass model was constructed, and it was used to find the count of Imitators and Innovators for the game, which came up to be 55.7 and 15.9 million respectively.

The pricing strategy for CS:GO was analysed with the help of competitor data and it was concluded to be efficient in terms of attracting new players.

It was also then found that promotions were not effective for this game since the game had become free-to-play. Other factors have also been looked at and appropriately commented upon.

The report concludes by predicting the players for the next 14 days and plots the same.

Preliminary Analysis

Data Cleaning and Gathering

As the first step, the data was sourced from SteamDB. Since the data was until 21st March 2023, it was cut to only include Player and Twitch viewer counts until 1st March 2023.

Next, the dataset had a “Flags” variable, which was filled with zeroes (0). It was converted to reflect the dates on which new versions of the game were released, as seen again on SteamDB.

Now, it was noticed that Twitch was introduced in context of the game CS:GO only after July, 2015. Since the data was sourced starting from 30th November, 2011, all the NA values were removed. Another reasoning being that, if the data was imputed with a number, it would indicate the presence of data, which is fundamentally wrong.

Moving on to missing data after the removal of NA values till July 2015, the remaining NA values were imputed using the “mice” package, which uses Regression Imputation to fill the NAs.

Our analysis has revealed that there was a notable surge in the number of players during the COVID-19 pandemic. To accurately reflect the impact of this phenomenon on our regression model, we have included an additional column in our dataset named "Covid Flag." This column is specifically assigned a value of one (1) for March 15th, 2022, which represents the period during which the pandemic had a significant influence on the player base.

Furthermore, we have also created another column titled "Majors," which is assigned a value of one (1) for the days when major championships were being held. This column covers all the major championships that have taken place between August 2015 and November 2022.

During the recent championship, an exciting announcement was made by the developers regarding the release of a new version of the popular game, "Counter Strike 2". This news generated a significant amount of hype within the gaming community, resulting in a surge of players returning to CS:GO. To account for this phenomenon, we have added another column to our data named "CS2Flag," which is assigned a value of one (1) only for November 13th, 2022, and a value of zero (0) for all other dates.

EDA

After cleaning and gathering, this is the summary of the dataset that is finally used for analysis:

Date	Price	Discount	Major	players	twitch_viewers	flags
Length:2785	Min. : 0.000	Min. : 0.00	Min. : 0.000000	Min. : 217547	Min. : 19449	Min. : 0.000000
Class :character	1st Qu.: 0.000	1st Qu.: 0.00	1st Qu.: 0.000000	1st Qu.: 564132	1st Qu.: 82380	1st Qu.: 0.000000
Mode :character	Median : 0.000	Median : 100.00	Median : 0.000000	Median : 653346	Median : 128549	Median : 0.000000
	Mean : 6.387	Mean : 57.39	Mean : 0.004309	Mean : 718317	Mean : 181076	Mean : 0.002154
	3rd Qu.: 14.990	3rd Qu.: 100.00	3rd Qu.: 0.000000	3rd Qu.: 872130	3rd Qu.: 228606	3rd Qu.: 0.000000
	Max. : 14.990	Max. : 100.00	Max. : 1.000000	Max. : 1355797	Max. : 1963331	Max. : 1.000000

Figure 1 Summary of dataset

Firstly, the correlation between the different variables were looked at.

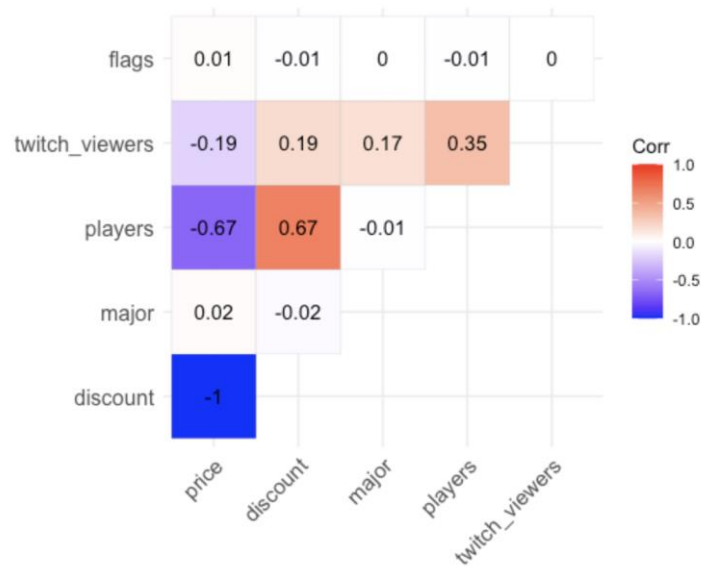


Figure 2 Correlation plot

A few points were noted from the above figure of correlation matrix:

flags variable has a negligible correlation on any of the other variables.

twitch_viewers have the maximum correlation with players.

price and players have a highly negative correlation, while discount and players have an equally positive correlation.

Now, these relationships were further explored visually to get a better understanding of how each variable impacts the other. First, a look at price vs player counts was taken, as shown below.

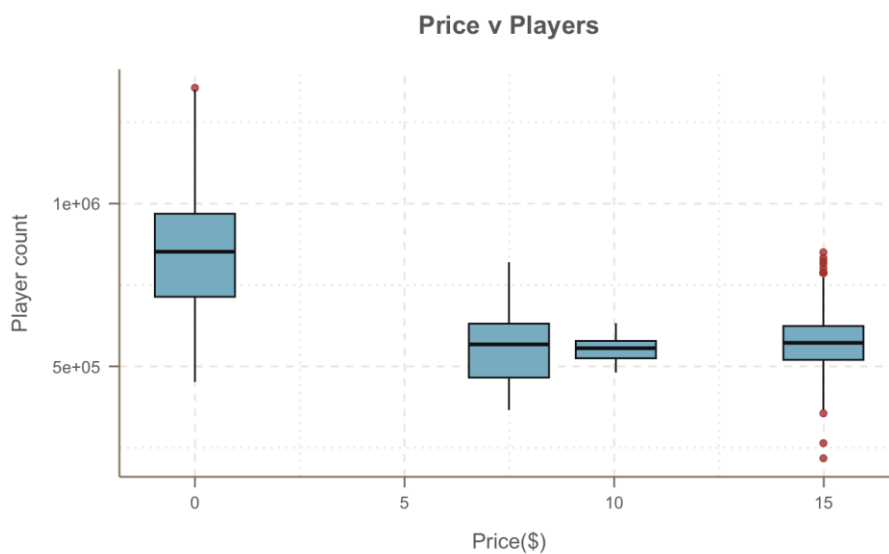


Figure 3 Price vs Players Boxplot

It comes as no surprise that with the increase in Price, there is a general decline in player count. Similarly, a visual relationship was established between Discount and Player counts, as show below.

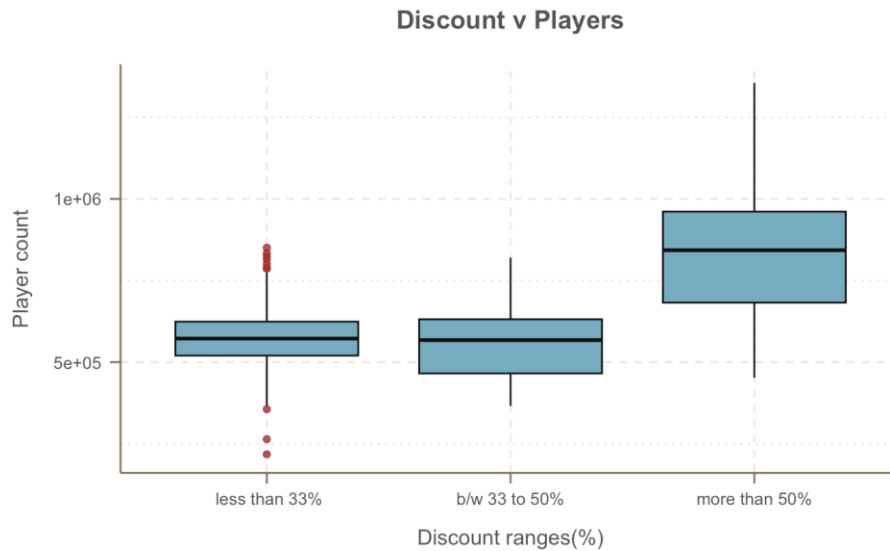


Figure 4 Discount vs Player counts boxplot

In both the graphs above, an ever-present relationship of Player counts with Price and Discount has been noticed. This relationship will be further expanded upon later in this report.

Coming to the relationship of Twitch with Player counts, we notice this below:

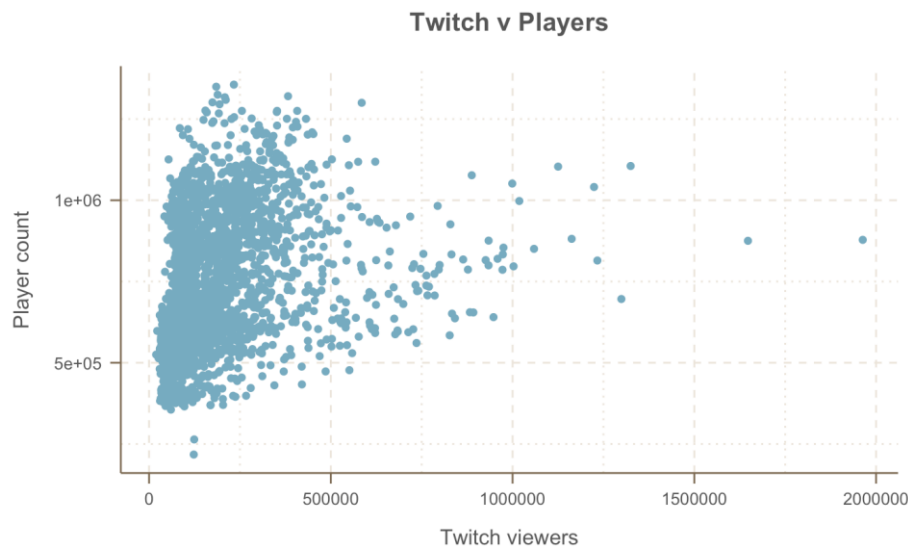


Figure 5 Scatterplot of Twitch viewers and Player counts

This will be expanded upon in Question 1 part of this report in detail.

Another interesting effect that was noticed in the dataset is that even though the game CS:GO started off as a paid game, the developers made it free-to-play from 2018 onwards. We observe how the player counts were affected by this in the below figure:

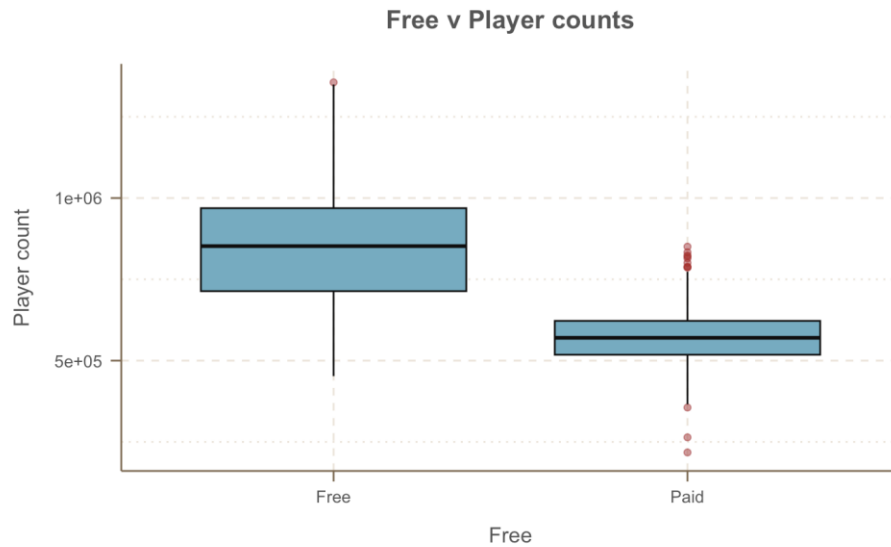


Figure 6 Boxplot to check player counts BEFORE and AFTER game became free

As it can be seen, the player counts almost doubled on average upon the game was made free-to-play. Now, after the initial analysis, the report moves on towards modelling to answer the questions that concern this report.

Question 1

To answer the question of how Twitch viewers affect Player counts, several Regression models were established, and a final model was chosen, which is used throughout the report, wherever applicable.

As a part of the modelling process, first the PACF and CCF graphs of the variables were plotted, to determine the lags and leads that can be included. Next, a series of models were created, and their performances were compared using AICc values, as shown in the figure below:

	AICc
Sink	51785.32
Player~Twitch	56246.58
PlayerVTwitch(1,0)	56189.72
Stepwise	51783.71
Stepwise_dlnorm	51783.71
InteractionEffect	75560.17
ExcludingTwitch	51853.18

Figure 7 Various ALM models and AICc values

From the above models, the Stepwise model was selected based on Adjusted R Squared value and AICc. Even though the model ExcludingTwitch showed a lower AICc value, the Stepwise was selected since the objective was to determine the effect of Twitch viewers, and the other model was made excluding that variable.

	Estimate	Std. Error	Lower 2.5%	Upper 97.5%
(Intercept)	7.996572e+04	7.194456e+03	6.585659e+04	9.407484e+04 *
playersLag15	8.663765e-01	9.936900e-03	8.468891e-01	8.858639e-01 *
twitch_viewersLag1	5.606650e-02	1.515090e-02	2.635390e-02	8.577910e-02 *
twitch_viewersLag4	-1.078061e-01	1.128370e-02	-1.299347e-01	-8.567740e-02 *
priceLag1	-1.805192e+03	2.258341e+02	-2.248078e+03	-1.362307e+03 *
twitch_viewers	1.238531e-01	1.280990e-02	9.873150e-02	1.489748e-01 *
twitch_viewersLag8	4.942600e-02	1.033130e-02	2.916530e-02	6.968680e-02 *
CovidFlagLead35	2.216237e+05	6.003083e+04	1.038966e+05	3.393508e+05 *
CovidFlagLead34	2.188664e+05	6.002885e+04	1.011431e+05	3.365896e+05 *
twitch_viewersLag2	2.542240e-02	1.352530e-02	-1.102300e-03	5.194710e-02
CovidFlagLead42	1.650854e+05	6.011904e+04	4.718533e+04	2.829855e+05 *
CovidFlagLead17	1.493277e+05	6.010135e+04	3.146231e+04	2.671931e+05 *
CovidFlagLead36	1.418574e+05	6.002994e+04	2.413207e+04	2.595828e+05 *

Figure 8 Summary of best model

Following the summary of the best model, the following residual graphs were produced:

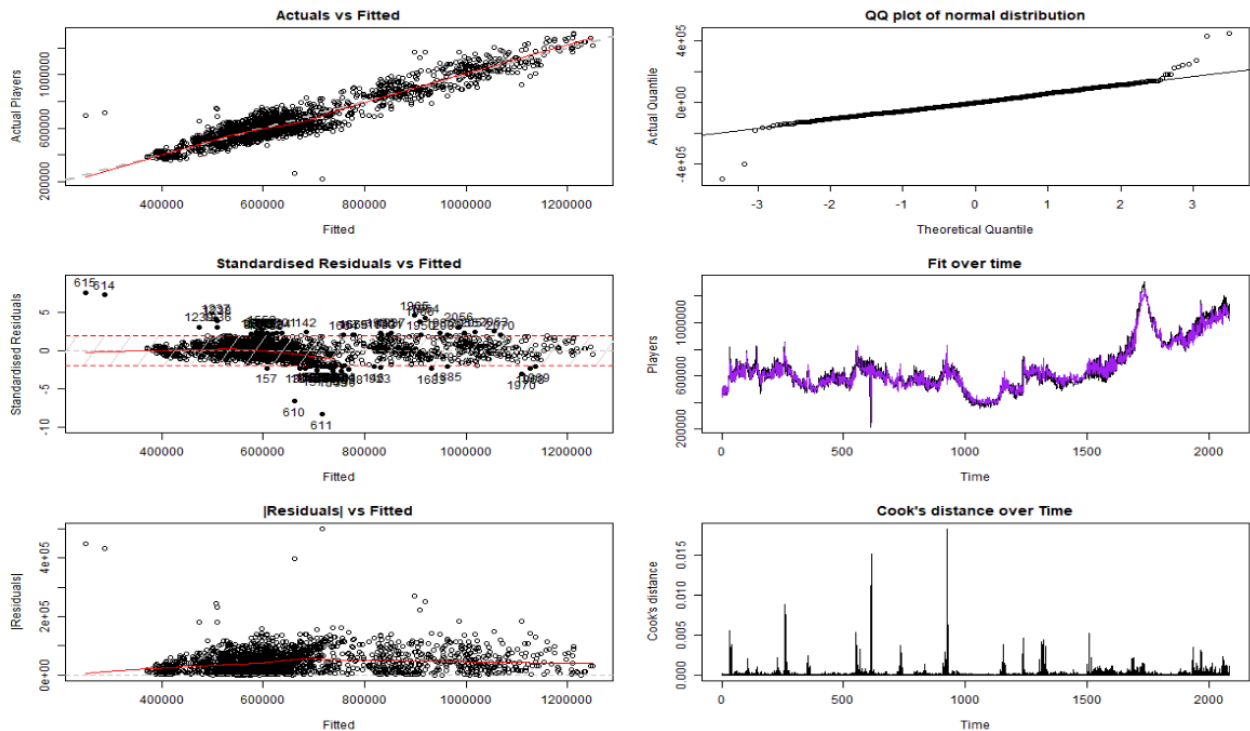


Figure 9 Residuals of best model

A good fit from the model can be observed. Also, it is to be noted that the model also follows the assumptions that follow a good regression model, for the most part.

Using this, the Question 1 is answered below.

For ascertaining the effect of Twitch Viewers on Player count for CS:GO, at the correlation plot (Figure 2) and the scatter plot (Figure 5) are observed. From both plots, it is observed that Twitch viewers has a very minimal effect on Player counts.

Looking further into it, a look at the Regression model's output in Figure 8 is taken. It is seen that on average, for every unit increase in Twitch Viewer, the Player count increases by 0.096, given all other variables remain constant. Despite is very minimal effect, it is shown to be statistically significant on a 5% significance level.

To thus answer the company, increasing the Twitch views will have a very minimal impact and they should focus on other important factors to get their Player counts up.

Question 2

Imitators and Innovators

The approximate number of imitators and innovators in the game was determined with the help of a Bass model. Since the daily sales data of the game was not publicly available, the aggregated monthly average of the players was taken for the construction of Bass model.

The cumulative sum of the average number of players per month was calculated. This was necessary for the estimation of a linear model of type:

$$y_t = b_0 + b_1 Y_t + b_2 Y_t^2 + e_t$$

where, y_t is the average number of players per month and Y_t is the cumulative sum of the average number of players per month.

A linear model was constructed, and the following coefficients were obtained as output.

```
> a1m(y ~ Y + I(Y^2), dfdata)
Time elapsed: 0 seconds
Call:
a1m(formula = y ~ Y + I(Y^2), data = dfdata)

Coefficients:
(Intercept)          Y      I(Y^2)
 1.423066e+05  2.417540e-02 -1.857479e-10
```

Figure 10 Linear model for Bass

Based on the estimated coefficients of this linear model, the coefficient of innovators and the coefficient of imitators was calculated. The coefficient of innovators was estimated to be 0.00105, while the coefficient of imitators was estimated to be 0.0252. Clearly, the imitators drive the number of players in the game. The number of innovators and imitators per month over the years was calculated and has been shown in the visualization below.

Average no. of players per Month, Fitted Model, Innovators & Imitators

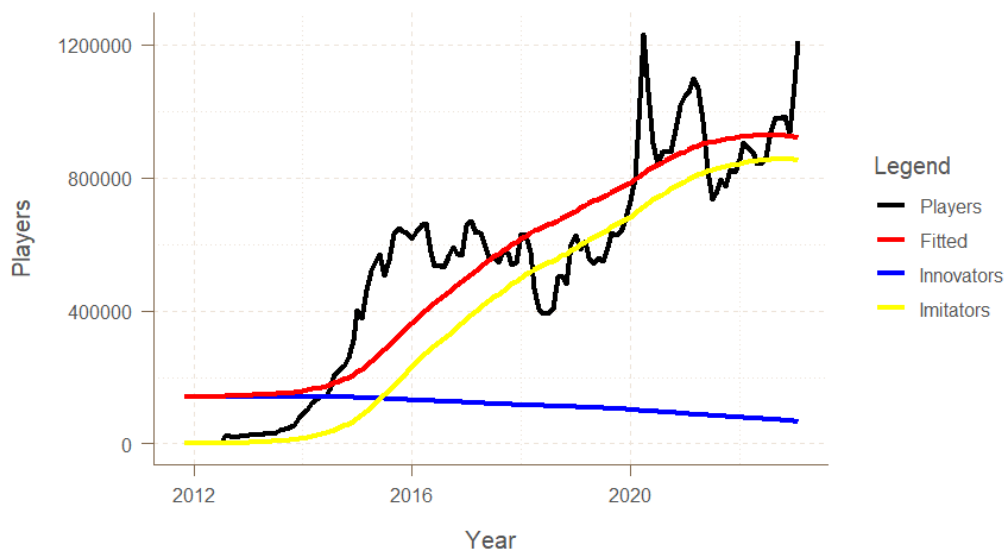


Figure 11 Innovators and Imitators per month

The fit was not as close to the original data as we had expected it to be. This might have been because of the multiple external factors and major global events. For example, we identified that the all-time high of the game was attained in the first half of 2020 and there's a very good chance that the spike was caused by the global pandemic lockdowns. Similarly, there could've been many other major

events or combinations of minor events that could have caused the player count to fluctuate. Because of this, it was not possible to attain the perfect fit. Hence, it was decided to go with the model with a reasonable fit.

Sum of the number of innovators (or imitators) per month through the years was assumed to be the total number of innovators (or imitators) in the game. Approximate values of the total number of innovators and imitators in the game were calculated and are summarized in the table below.

Player Type	Total in Game
Innovators	15.9 million
Imitators	55.7 million

Question 3

Pricing Strategy Efficiency

Since the inception of the game in 2012, the game was mostly priced at \$14.99. However, in December 2018, the game was made publicly available for free. Over the years when the game wasn't freely available, multiple promotional discounts were given and the overall summary of it is given in the table below.

Discount (%)	No. of Days	% Share
0	1315	44%
33	69	2%
50	84	3%
64	2	0%
100	1547	51%

Since the release, the promotions were given only on 5% of the days and for 51% of the days, the game was free.

To understand if the pricing strategy is efficient in terms of attracting new players into the game, we needed to analyse this from 2 perspectives. One was to check if the promotional pricing strategies were attracting more players into the game. The second was to compare the pricing strategy with a direct competitor to determine if it was efficient in attracting more players to the game.

The analysis of the promotional pricing strategy has been shown in the next section. Hence, the focus will be on competitor price analysis and comparison in this section.

We found that there were multiple competitors for Counter Strike throughout the years. Although the main ones being - Rainbow Six Siege, Call of Duty, and Overwatch. Out of these 3 games, Rainbow Six Siege was chosen as the primary competitor for analysis. This was done because, both Counter Strike and Rainbow Six Siege were launched not very far apart and were serving similar audience.

The reason for leaving out the other 2 games from this analysis was because of various reasons like the games being played widely outside of the Counter Strike paid version dates and to keep the analysis simple.

Rainbow Six Siege was launched in May 2015. From then to December 2018 (when Counter Strike was made freely available), the average price of the game was around \$30 (promotions were taken into consideration as well). On the other hand, in case of Counter Strike, even if we let go of the promotions and assume that it was sold at the upper bound price throughout, the average price would still be \$14.99, which is 50% less than the competitor's price. So, we concluded that Counter Strike was more affordable than the competitor.

We then analysed the player count for both the games through the years. This has been shown in the plot below.

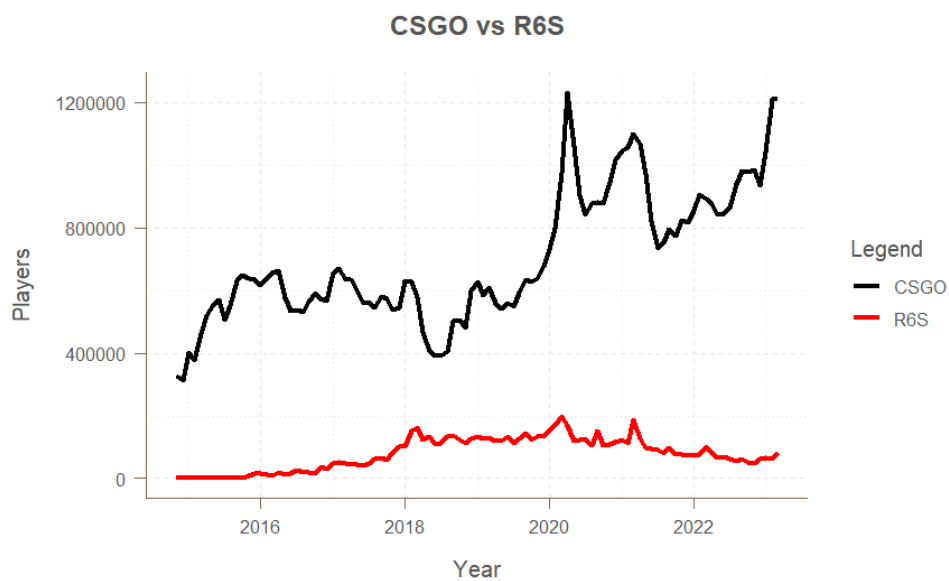


Figure 12 Comparing player counts of Rainbow Six Siege vs CS:GO

It was clear that the number of players playing the Counter Strike game was miles ahead of the number of players playing the primary competitor's game. As expected, the lower price of the Counter Strike game has attracted more players than its counterpart in the years where the game wasn't freely available. And when the game was made freely available, it went on to attract even more players.

Hence the pricing strategy of Counter Strike: Global Offensive in terms of attracting new players to the game is very efficient and way ahead of their competitor.

Question 4

After examining the estimates generated by our regression model, we have concluded that major championships do not have a significant impact on increasing the number of players in the game. This could be since major championships occur relatively infrequently throughout the year.

On the other hand, it was identified that COVID-19 had a significant impact on the player count, but this impact was not immediate. Instead, there was a delay of 34 days before we observed a noticeable increase in the number of players. This finding suggests that COVID-19 had a delayed impact on the player base.

Based on our analysis, it cannot be conclusively stated that adding more championships in the future would have a significant impact on the number of players in the game. Further research is required to explore the potential impact of additional championships on the player base.

Question 5

Using the best model found in Question 1, the player counts for the next 14 days were predicted and the following data points and the corresponding plot for the current and predicted player counts.

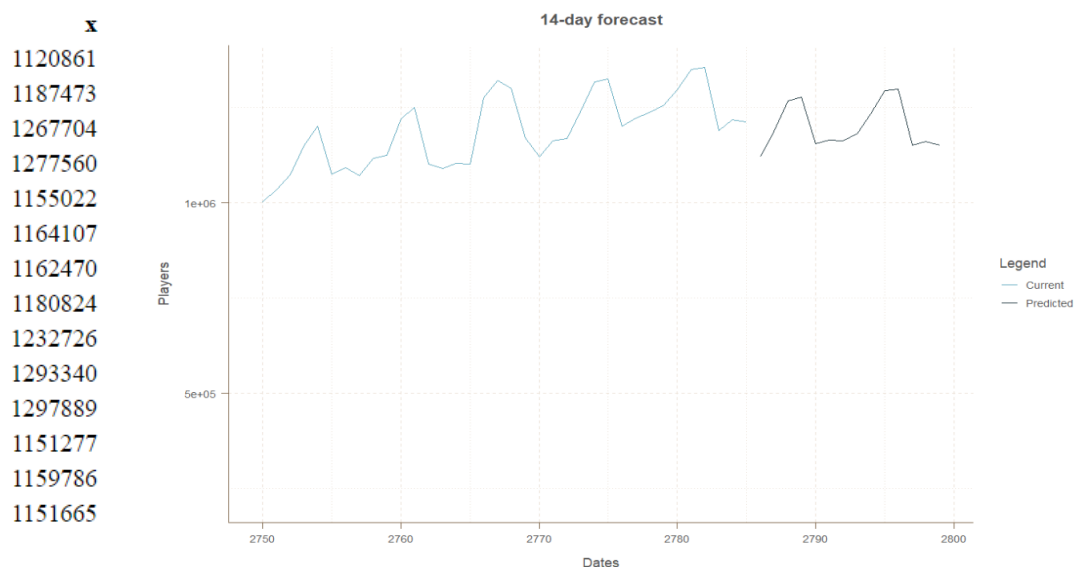


Figure 13 14-day forecasts using best model.