Pew, Pew, Pew

An Analysis of Professional Counter-Strike

COMP 30780 Data Science in Practice

Dáire Murphy (15441458), Rónan O'Neill (16433656)

School of Computer Science

University College Dublin

May 11, 2018

Abstract. Counter-Strike is a statistically-heavy esport with little publicized analytical research. This project takes data from multiple tournaments and uses it to examine the impact of the first kill and the in-game economy on win rate. A separate dataset is used to build a prediction model that forecasts the future careers of various players aged 19 and under. Our analysis concluded that the first kill in a round was of significant importance. It also showed how poor economy management could lead to serious consequences. The prediction model returned that Mathieu "ZywOo" Herbaut has the most potential of all players under 20.

Declaration. We (Dáire Murphy, 15441458 and Rónan O'Neill, 16433656) declare that this assignment is our own work and that we have correctly acknowledged the work of others. This assignment is in accordance with University and School guidance[1] on good academic conduct in this regard.

[1] See https://www.cs.ucd.ie/sites/default/files/cs-plagiarism-policy_august2017.pdf

1. Introduction

The main purpose of this project was to collect and analyse datasets relating to the professional Counter-Strike scene. The initial analysis would focus on highlighting what factors impact success at both a round and match level. Furthermore, we used the data collected in order to make predictions on the future success of young players.

Counter-Strike is a first-person shooter that sets two teams of five against each other. Each match is a best-of-30, with continuous best-of-six overtimes in the event of a tie. One team starts as Counter-Terrorists and the other as Terrorists, with the teams switching sides at halftime (15 rounds). The objective of the Counter-Terrorists is to protect the objective (two bombsites) by either eliminating the other team, running down the clock, or defusing the bomb if it is planted. The objective of the Terrorists is to explode one of the bombsites or eliminate the Counter-Terrorists. Little publicly available research has been conducted on the game at a professional level, so we thought it would be interesting to delve into the topic from a data science perspective.

The datasets created in this project were scraped from two sources: HLTV and DreamTeam. The datasets contain statistics and information regarding players and teams, first kills, and economy. Using these newly formed datasets, we attempted to answer three main questions.

We began our analysis by examining the importance of the first kill in a round. We discovered that this particular kill is of extreme importance, far more than we originally thought. Continuing with this theme, we decided to investigate how different factors relating to the first kill impact the win rate of a round, including: the time to first kill and the map in which the game is taking place. It was found that these factors had little to no effect on changing the win rate.

The next stage of our analysis concentrated on the in-game economy. The economy in Counter-Strike could be defined as a team’s ability to purchase equipment. Each team receives an in-game monetary bonus at the end of each round, with the amount received depending on the team’s success in that round. The second dataset was used to explore how important managing this economy is to the overall win rate of a match. The results of this research question showed a drastic decrease in match win rate as the number of low-buy rounds increased. The teams who were most successful in these types of rounds were determined and a correlation was found between success in these rounds combined with pistol rounds (the rounds at the start of each half) and overall tournament success.

Finally, our last dataset was used for two parts. Firstly, it was used to confirm a previous study into how increasing age negatively impacts performance. Secondly, the younger players (aged 19 or younger) were studied and their short careers compared to that of the more established veterans. These comparisons were then used to form a prediction on how this select group of players would continue to perform in the future. In turn, this enabled us to predict that Mathieu "ZywOo" Herbaut has the most potential going forward.

The remainder of this report is organized as follows. Our motivations and objectives for this project are outlined in the next section. There we lay out are main and sub-research questions, moving on to how we acquired the data needed to answer these questions in the following section. The subsequent section is devoted to our data analysis and our results. Our ethical considerations, reproducibility, and limitations are discussed in the penultimate section. Finally, we wrap up with our conclusions and a brief mention of future work that would continue from this project.

2. Motivations & Objectives

This section will cover why this topic was chosen for our project and will also lay out the three main research questions, and any sub-questions, that our analysis will be focusing on.

2.1. Background & Motivations

Minimal research has been conducted on Counter-Strike's professional scene and its players. An investigation online led to the discovery of two datasets on the website 'Kaggle', neither of which were remotely relevant to our own research. However, an analysis on "*How Much Does Age Matter in Professional Counter-Strike: Global Offensive*" by Reddit user '*Analytical_Gaming*', was relevant to the research we intended on conducting and was part of the inspiration for our third research question.

The purpose of '*Analytical_Gaming*'s' analysis was to determine the correlation between a player's age and their HLTV Event Rating (a rating system used by HLTV and the Counter-Strike community to score a players individual performance in a single match or over the course of an entire event), as well as determining the correlation between a player's age and their Kill/Death ratio. Like our research, most of the this study's data was collected from HLTV.

The analysis concluded that the correlation coefficients are "negatively weak-to-moderate", indicating that age does have a negative impact on skill, with a slight drop off in performance for players over the age of 25 and more noticeable decrease in performance with players over the age of 28. There is also evidence to deduce that the correlation would have had a higher negative coefficient if there had been less "missing data points", players who were 25 or older that had retired from Counter-Strike due to their deteriorated skill.

The lack of any significant research being done on professional Counter-Strike, a multi-million dollar industry, was very surprising as the game itself, along with websites such as HLTV, provide a large amount of statistics about each game and player. As other sports have developed and with huge amounts of money being spent in the likes of the Premier League, NBA and NFL, a large emphasis has been placed on using data analytics experts to improve teams in all areas. Video games now have a higher annual income than the music and movie industry combined, and as a result, esports is one of the fastest growing entertainment industries in the world, with Counter-Strike: Global Offensive as one of the games in the forefront of the scene in both viewership and prize money.

Consequently, there will likely be a rapid increase in employment opportunities for data science graduates in the esports industry. This along with a personal passion for the game and its professional scene, we chose this project.

2.2. Research Questions

This project revolves around three main research questions. Research Question 1 (RQ1) and Research Question (RQ2) mainly focus on how win rates are impacted by various factors, while Research Question 3 (RQ3) centres on building a prediction model.

2.2.1. RQ1: How Important is the First Kill?

The first kill in this instance relates to the first kill of each round of a match. The focus of this question is to determine the round win rate of teams who get the first kill compared to those that do not. The set of sub-questions includes:

1) How Do Different Factors Relating to the First Kill Impact the Win Rate?

2.2.2. RQ2: How Important is Managing Your Economy?

The research conducted for this question examines how to number of low-buy rounds in a match affects the win rate.
The set of sub-questions includes:

1) Does Low-Buy and Pistol Success Correlate to Tournament Success?

2.2.3. RQ3: Do Younger Player Perform Better and Can We Predict Future Stars?

With this question, we will look to determine if age has an impact on a player's ability, as well as attempting to predict the next star of Counter-Strike.

The set of sub-questions includes:

1) Do Younger Players Outperform Older Players?
2) Who are the Most Promising Youth Talents and Can We Predict Their Future Event Ratings?

3. Data Wrangling

This section will cover how we acquired the data necessary for answering our research questions and the steps taken in order to clean and prepare it.

3.1. Data Acquisition

As the data that we required was not readily downloadable, we were required to scrape it from two sources: www.HLTV.org and www.DreamTeam.gg. This section will be broken down into three subsections, one relating to the data collected for each research question.

3.1.1. RQ1

HLTV was the first port of call when scraping data for this question. The data taken pertains to the previous 12 Major tournaments. These tournaments are the most prestigious in the Counter-Strike circuit each year, with a prize-pool of $500,000. Using BeautifulSoup, player statistics from every match from these past tournaments was collected. This included statistics relating to the player's in-game performance (kills, first kills, rating), as well as information regarding the match itself (map, teams playing).

A second and third dataset were created using data scraped from DreamTeam. Matches were manually uploaded to the website in order to produce the statistics required. This was time consuming and was the main reason that only the Major tournaments were used in the analysis for RQ1 and RQ2. The first of these two datasets held statistics on the winner of each round of each match as well as the overall winner of the match. The third dataset for this question was more first kill focused and contained information on the first kill from every round in every match. Some of the information in this dataset included: the player who got the first kill, the victim, the weapon used by both players, the time it took to get the first kill, and the map the match was being played on.

3.1.2. RQ2

DreamTeam was used once more when acquiring data for this research question. Again, using only the previous 12 Major tournaments, a dataset was formed that included information regarding every round of every match, like the first kill dataset. This time, the dataset contained information about the economy state of each team, what side they were on, and if they won the round and the match.

3.1.3. RQ3

The data required for RQ3 was acquired by once again scraping HLTV. Firstly, every player page stored on the website was scraped using BeautifulSoup. Each player's name, age, current team and several career statistics from their Counter-Strike careers (total kills, total death, average rating) were placed into a dataset.

The second dataset needed was the player ratings from past events. This was obtained by scraping the events page of each player and placing each player rating into a DataFrame containing every event played by any player. Following this, a second, smaller events list was scraped from the website, this list only containing tournaments that had a prize pool of at least $4000. This will later be used to drop unimportant or low skill events that may inflate some players average ratings.

3.2. Data Cleaning & Preparation

Like section 3.1., this section will also be divided into three subsections. Each subsection will discuss how the datasets were cleaned in relation to each research question.

3.2.1. RQ1

After careful examination of the scraped data, it was clear that the round winner and the first kill datasets needed cleaning. Some matches contained duplicate rounds or rounds that contained missing information. This was visible when comparing the "Round" column of the DataFrame to the "Winner" column. If they were not equal, the record for that match contained too much or too little information. These rounds could not be easily fixed, so to keep the data as accurate as possible, the matches that these rounds were played in were dropped from both datasets.

The two datasets were then combined so that it was a simple process to check if the team that obtained the first kill won the round.

3.2.2. RQ2

The economy dataset was scraped using only the clean matches from the previous dataset, but a different sort of error was being noticed. This dataset was sometimes including two different entities outside of the players in the game: the observer and bots. In a broadcasted match of professional Counter-Strike, the observer acts as a cameraman. They control what the viewers see and switch between first-person perspectives of the players and third-person perspectives of the rest of the map. Bots are AI-controlled players that automatically join the game if a player disconnects. They will never take part in a round but are listed in the lobby during freeze-time when a player has left for various reasons. As both the observer and bots are saved in the game logs, they also appeared in the scraped data.

Observers were easy to remove from the dataset as they were not assigned a side, so using a mask, all rows where the side was unassigned were dropped. Bots however are part of the team they join to. This made it more difficult to remove them. Instead, rounds that contained more or less than 10 players were found and dropped from the dataset. This allowed the dataset to be filtered down to only contain information about each team in the match instead of information for every player. This was achieved by only keeping every fifth row.

Finally, the economy dataset was combined with the round winner dataset for the same reason that the first kill and round winner datasets were combined in Section 3.2.1.

3.2.3. RQ3

When scraping the players dataset for RQ3, a small website error became evident. Out of the 636 players stored on the website, a single players page failed to be scraped correctly. That player's data was dropped from the dataset, as the information stored on the website was unreliable. This dataset was then cleaned to the data types required.

The player ratings from past events dataset contained many unnecessary events in its columns. Using the list of events with a prize pool of over $4000 retrieved from the website, any events in the large dataset that were not in the smaller event list were dropped. This provided a DataFrame containing all the players who competed in any top event since the games release in 2012, and those players average ratings in that event.

4. Data Analysis & Results

This section covers how the analysis was conducted for each research question and the results that we produced.

4.1. RQ1: How Important is the First Kill?

The initial research question investigates the first kill of a round and what impact it has on the rest of the round. The analysis then moves on to examining various factors related to the first kill.

4.1.1. Datasets

The dataset used in this analysis is the first kill dataset cleaned in Section 3.2.1.

4.1.2. Approach

The initial goal of this research question was to measure how impactful the first kill was on the win rate of a round. The analysis then moved on to identifying if any factors related to the first kill impacted this win rate further.

The Impact of the First Kill

Each row in the dataset contains information regarding the first kill of every round. For this portion of the analysis, the side that the player who obtained the first kill was on (Counter-Terrorist or Terrorist) and if they won the round were the data points used. If a team that obtained the first kill won the round, the relevant column in the DataFrame held a True value, whereas if they lost it contained a False value.

In this instance, win rate is quantified as a number between 0 and 1. Like probability, 1 indicates certainty that a team wins the round, whereas 0 indicates it is impossible. By counting all the True and False values in the DataFrame and dividing by the total number of rows, the win rates could be calculated for teams who get the first kill in the round and for teams who do not.

The original DataFrame was divided into two; one for all instances where a Counter-Terrorist killed first and one where a Terrorist killed first. This allowed for comparison between the two sides to see if there was any variance in win rate by side.

The Time to First Kill

One of the statistics included in the first kill dataset was the time it took for each kill. Our initial belief was that the earlier a first kill took place, the higher the win rate would be. Teams will go into each round with a strategy, but if one of their players dies first and early, this strategy could fall apart which could lead them to losing the round at a higher rate.

Each time interval was grouped, and their mean win rates were calculated. To remove outliers, only groups that contained more than 10 occurrences were kept. A scatterplot was then formed, plotting the round win rate against time, and a line of best fit was generated.

The Effect of the Map

In Counter-Strike, the playing field of the match is known as a "map". The layout of each map is different, but the objective is always the same. From the tournaments we selected there are eight maps in rotation, with seven being in the active map pool (the maps used in each Major) at any one time.

As each map varies, teams initially come into contact with each other in different ways and at different times. Our initial belief that the time to first kill may have an impact on the win rate led us to also believe that the map could also have a significant effect.

To explore this, the DataFrame was grouped by map and the mean time to first kill and mean win rates were found. Using a multi-y axis graph, the two variables were plotted together.

4.1.3. Results

The Impact of the First Kill

Our analysis revealed that the impact of the first kill on the win rate of a round was far greater than we initially anticipated.

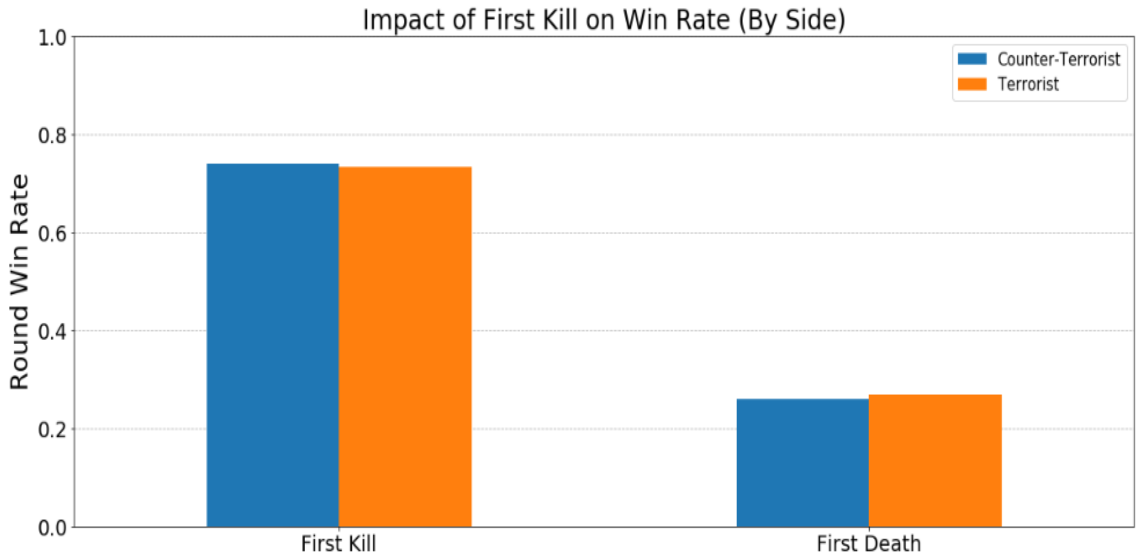

Figure 1. Bar Chart Showing the Impact of the First Kill on Win Rate

Figure 1 shows just how important the first kill is in a round. On average, teams that earned the first kill had a round win rate of approximately 0.73. There was very little variation in this number by side, with the difference being less than 0.01.

The Time to First Kill

Our scatterplot disproved our initial belief that there was a correlation between time to first kill and win rate.

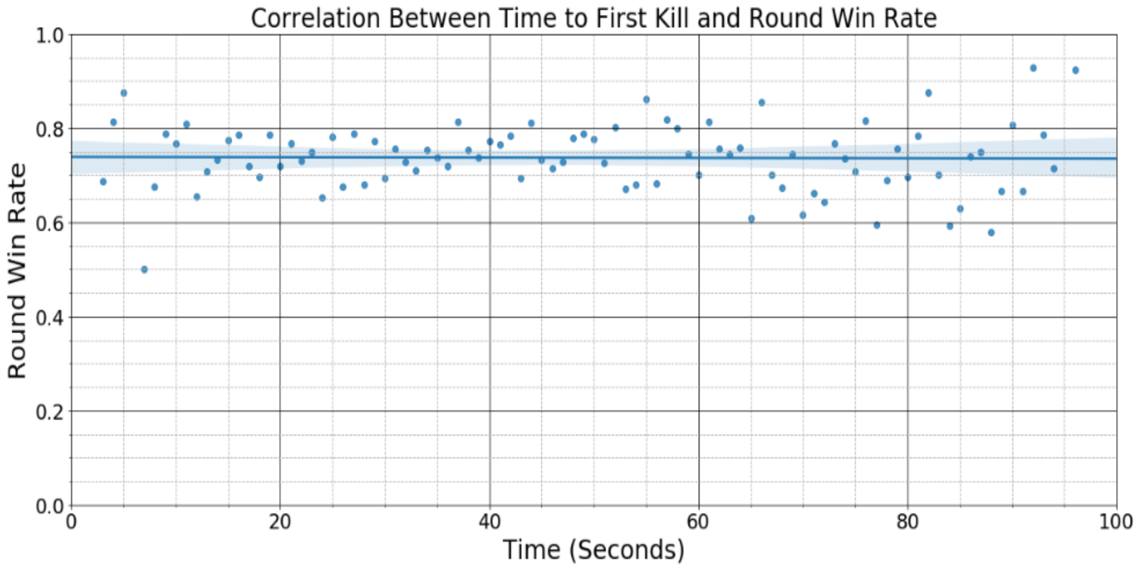

Figure 2. Scatterplot Showing the Correlation Between Time to First Kill and Win Rate

Figure 2 left us with a somewhat disappointing result, in that there is no correlation between time to first kill and round win rate. The line of best fit shows that win rate remains between 0.7 and 0.8 no matter when the first kill occurs in the round.

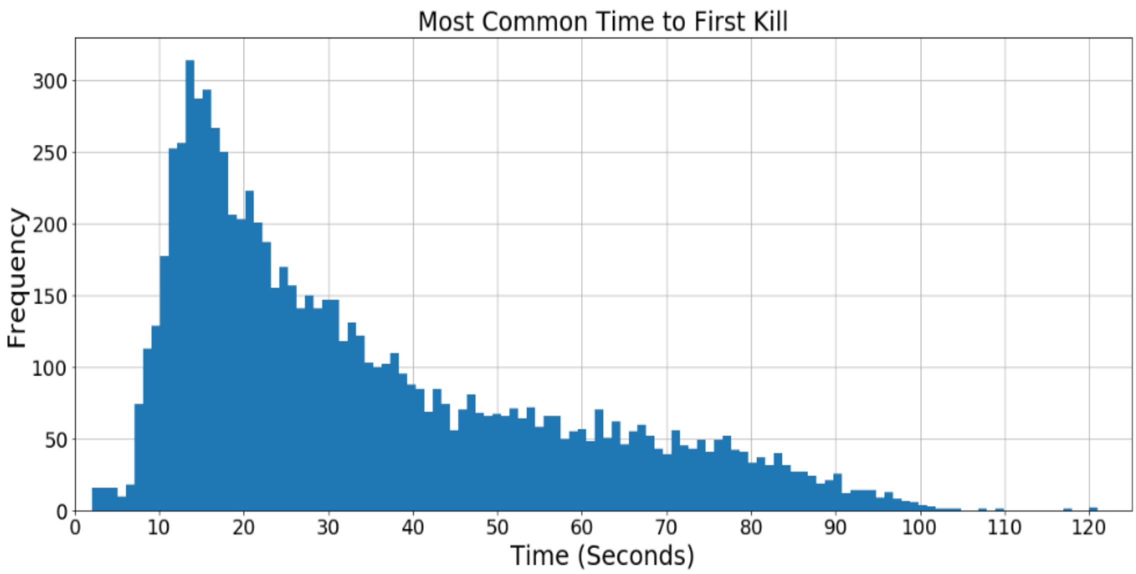

Figure 3. Histogram Showing Distribution of Times to First Kill

The histogram in Figure 3 shows that the majority of the first kills occur within the first 20 seconds of a round, with the most common time interval being between 10 and 20 seconds. There is then a sharp and continuous decline as the time increases.

The Effect of the Map

This segment of the research indicated little variance in success of the round as the map differed, but a slight change in the mean time to first kill.

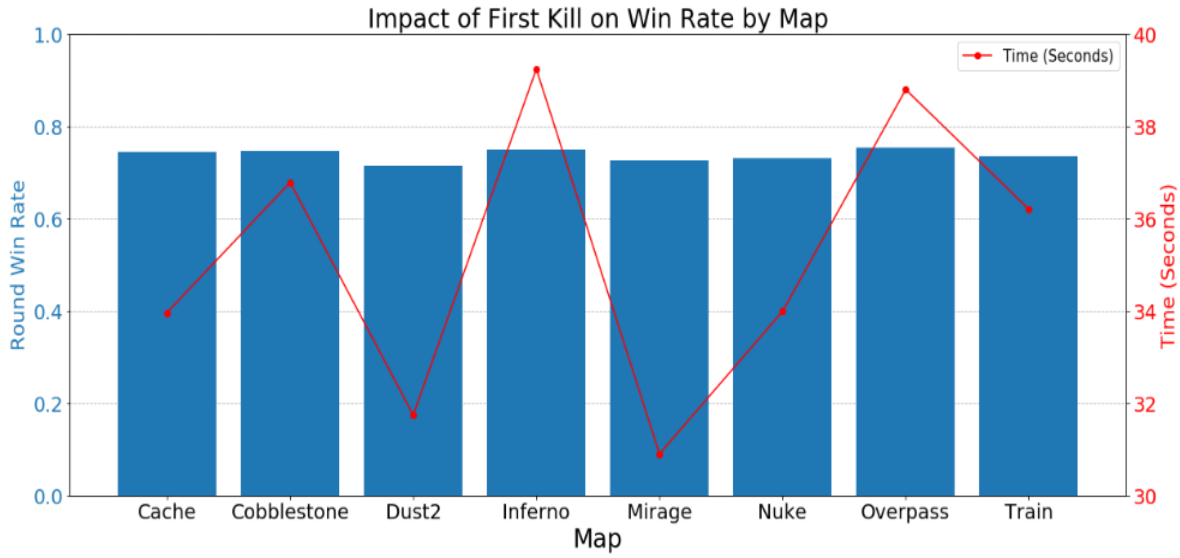

Figure 4. Bar Chart Showing Impact of First Kill on Win Rate by Map, including Mean Time to First Kill

As Figure 4 shows, there is no significant change in win rate for each of the eight maps. Yet, change is visible in terms of the time to first kill for each map. Inferno and Overpass are the slowest when it comes to the initial engagement, at around 39 seconds. The first kill occurs on Mirage the quickest, with a mean time of approximately 31 seconds.

4.1.4. Discussion

The Impact of the First Kill

Answering our main research question for this section of the analysis: the first kill is of substantial importance in how a round will play out. A win rate of over 0.7 was far beyond what we initially expected.

Teams using this information could play it a number of ways. One way could be to push aggressively in search of that first kill. This is high-risk, high-reward strategy that could catch the opposition off-guard but would need to be used sparingly so as to not be countered.

Another way could be patience. A team who sits back and plays patiently on the defense (Counter-Terrorist) could wait for the attackers to push or make mistakes. This method would be most successful on the defensive side, as the Terrorists are time-limited in their objective. However, the offensive side could employ this every so often to confuse the defenders. When the Counter-Terrorists push from their positions in an attempt to look for information on the other team, they could walk into a trap.

It is not overly surprising that in Figure 1 we see only a minute variance in win rate by side. For instance, if the Counter-Terrorists get the first kill they can fall back and play passively on the bombsites, knowing that the Terrorists will not be at full capacity when they attempted to attack. If the Terrorists themselves are the ones to kill first, they can regroup and attack a now weakened objective as a unit.

The Time to First Kill & The Effect of the Map

The most surprising aspect of this research question was that the time to first kill and the map had little to no effect on the win rate. Our initial hypothesis that the earlier the first kill was achieved, the greater impact it would have turned out to be false. Figure 2 was damning in its result, generating a line of best fit that was practically straight across.

As mentioned in 4.1.2., the layout of each map is different, which means that teams meet each other at different times. This was shown in Figure 4, where the mean time of the first kill changes for each map. However, because the time to first kill had no impact, this also had no effect on the win rate of a round.

What this suggests is that how or where you get the first kill is not important. What is important is getting the kill in the first place. A team's ability to consistently win the opening duel of a round could be key in finding long-term success.

4.2. RQ2: How Important is Managing Your Economy?

The analysis for this research questions focused on the in-game economy, with an emphasis on low-buy rounds.

4.2.1. Datasets

The dataset used in this analysis is the economy dataset cleaned in Section 3.2.2.

4.2.2. Approach

With this research question, we set out to explore the effect mismanaging your economy has on your chances of winning. We then moved onto to examining if success in these low-buy rounds had any correlation to overall tournament success.

The Impact of Low-Buy Rounds

As explained in the Introduction, your economy in Counter-Strike is your ability to purchase equipment. A player who dies during a round must re-buy all their equipment in the next. If the majority of players on a team cannot afford a full buy – usually consisting of a rifle, Kevlar, and utility (various grenades and a defuse kit on the Counter-Terrorist side) – then they must decide on whether to play an eco- or a force-round.

If the team does not have enough money in the current round, but will have in the next round, they usually play what is known as an eco-round. This means they only play the round with a pistol. As they have no Kevlar, they are easy pickings for the opposition. Success in these rounds is highly unlikely, but not impossible. If the opposition knows they are "on an eco", they may be overzealous in their aggressiveness, getting caught unaware.

The usual mark for an eco-round is having around \$2000 per player and a round-loss bonus of at least \$1900. The round-loss bonus starts at \$1400, increasing by \$500 every consecutive loss until it is capped at \$3400. An extra \$800 is awarded to the Terrorists if they plant the bomb, but it is defused. A team will play a force-round if, similar to an eco-round, they do not have enough money for a full buy but will also lack the money in the following round. In these rounds, teams will buy to the best of their ability, in an attempt to "force" a round win. Success is far more likely than an eco-round, but still improbable.

To study how an increasing number of these rounds impacted the win rate of a match, the economy dataset was analysed. A column containing the round type (full, force, eco, and pistol) was used to filter the dataset to only include eco- and force-rounds. The match win rate was then calculated for teams who had X number of low-buy rounds in a game.

Correlation Between Low-Buy and Pistol Success and Tournament Success

Another type of round, mentioned briefly, are pistol rounds. Pistol rounds are the rounds at the start of each half (rounds one and 16). Each player is given $800, allowing them to purchase either Kevlar or an upgraded pistol and one or two of the various grenades. The winner of these rounds, like the winner of every round, is awarded between $3250 and $3500, depending on if the bomb was planted. This is far greater than the $1400 awarded to the losing team[2]. This enables the winning team to get a hold of their economy before the opposition can as the losing team is immediately placed in a low-buy situation.

Combining success in these pistol rounds with the low-buy rounds, we wanted to determine if there is any correlation between this and tournament success. As our dataset contained the most information regarding ESL One Katowice 2015, this tournament was selected. The mean success rate of teams in the selected types of rounds was plotted against their final tournament ranking and a scatter plot was formed.

We were initially unsure what the result would be. Obviously, success in any round is of a benefit, but we also thought that a team who plays more low-buy rounds has a better chance at winning at least one than teams who manage their economy well. An increase in success could just mean they are mishandling their economy.

4.2.3. Results

The Impact of Low-Buy Rounds

After completing the necessary analysis, we discovered just how impactful an increasing number of low-buy rounds can have on match win rate.

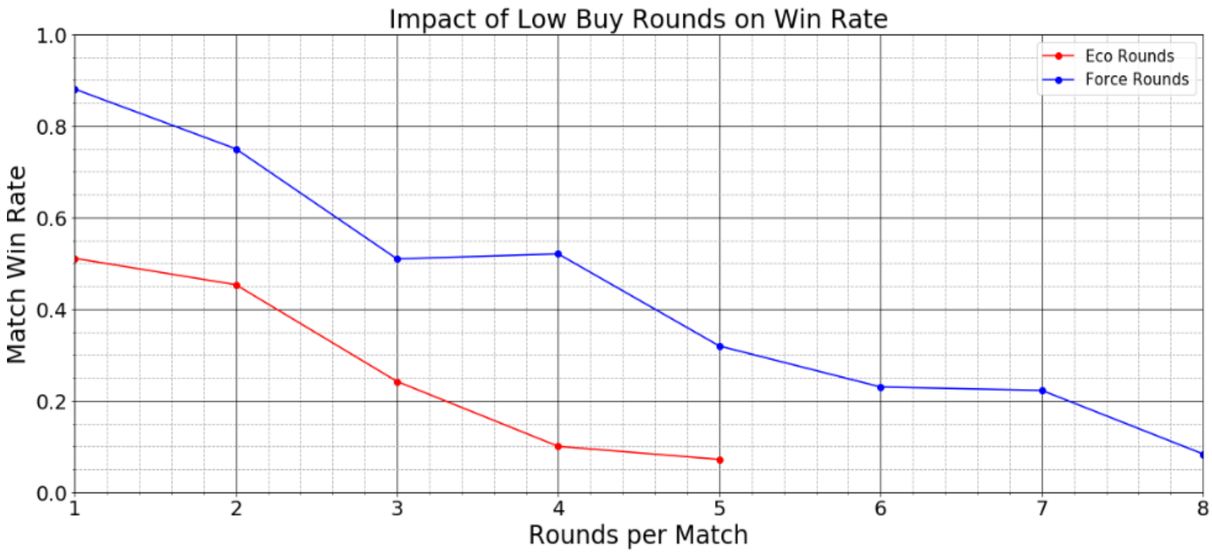

Figure 5. Line Graph Showing the Win Rate for X Number of Low-Buy Rounds

In Figure 5, the red line represents eco-rounds, while the blue line represents force-rounds. It is clear to see that an equivalent number of force-rounds is less impactful to the win rate than eco-rounds. Just

[2] After an update in October 2018, teams who lost a pistol round now received $1900. In the dataset, this only applies to the 2019 Major.

two eco-rounds cause win rate to fall below 0.5, whereas it would take five force-rounds. However, the trend for both round types is the same; an increase in rounds leads to a certain decline in win rate. This graph illustrates the importance of managing one's economy.

Correlation Between Low-Buy and Pistol Success and Tournament Success

Team	Ranking
Fnatic	16
Ninjas in Pyjamas	15
Virtus.pro	14
Team EnVyUs	13
PENTA Sports	12
Keyd Stars	11
Natus Vincere	10
Team SoloMid	9
Vox Eminor	8
LGB eSports	7
Counter Logic Gaming	6
Cloud9	5
FlipSid3 Tactics	4
Titan	3
HellRaisers	2
3DMAX	1

Figure 6. Final Tournament Standings for ESL One Katowice 2015

Using the final tournament standings for ESL One Katowice 2015, as shown in Figure 6 (where 16 equates to first place and 1 equates to last place), a scatterplot was formed.

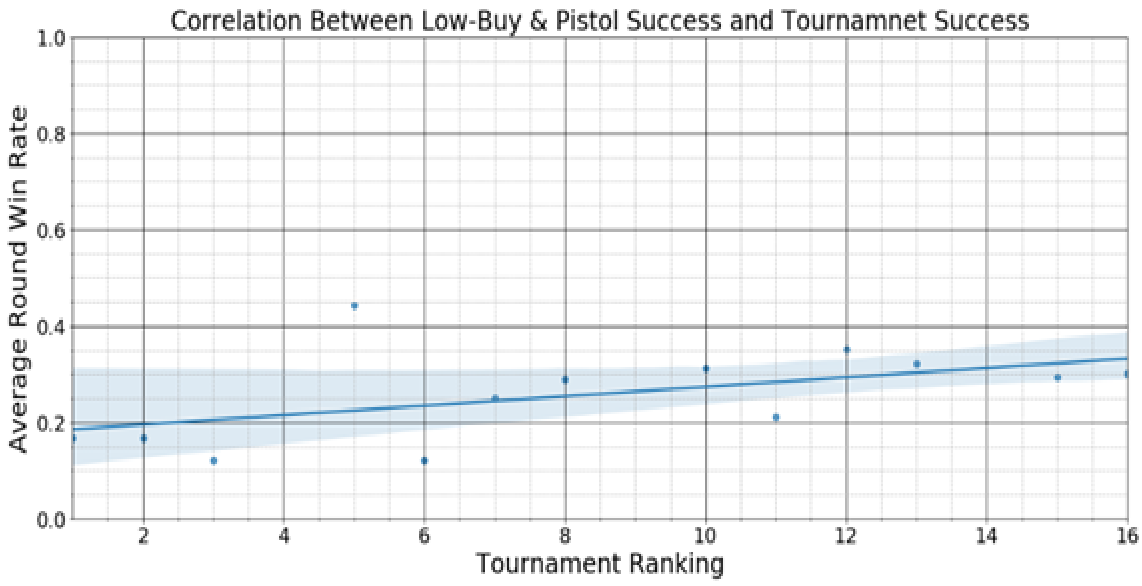

Figure 7. Correlation Between Low-Buy and Pistol Success and Tournament Success

From Figure 7, we can see that there is a positive correlation between success in these round types and overall tournament success. The outlier here was Cloud9 who finished fifth last. On closer examination, it was found that they had terrific success in pistol rounds with a mean win rate of 0.83. Their downfall came in the full buy rounds where they performed worse than average, winning only 50%.

4.2.4. Discussion

The Impact of Low-Buy Rounds

Although not overly surprising, the result of this question highlights the importance of money management in-game. Win rate drops from around 0.45 after two eco-rounds to approximately 0.1 after four. That means teams who play four or more eco-rounds have a 10% chance of winning the match.

Two points of discussion arise from this. Firstly, teams may need to consider "saving" more often. Saving refers to sacrificing a round as it is still going on. You fall back to safe positions and let the opposition win the round, in order to keep your equipment for the next round. This often allows you to buy a weapon for a teammate who died in the previous round, so that as a team you can avoid a force- or eco-round. This strategy is most successful on the Counter-Terrorist side, as Terrorists will earn no round-loss bonus if they survive a round, but do not plant the bomb. If a team knows their opposition is saving, they will often hunt them down to take their equipment away, ultimately making the save worthless. A team whose own economy is poor but chooses to hunt a saving team can often damage their own economy if they are killed in the process.

The second point of discussion relates to team strength. A team's inability to manage their economy or their lack of know-how on when to save are not the only factors at play here. If a team is being outclassed, often there is little they can do. In matches where one team dominates the other, multiple low-buy rounds may not indicate a team's poor budgeting skills, but merely the other team's strength. This is something to keep in mind when examining the results.

Correlation Between Low-Buy and Pistol Success and Tournament Success

Figure 7 shows that there is a positive correlation between tournament success and low-buy success. With regards to the rankings, the only definitive placements are first and second. All Majors follow the same format: a group stage where the best two teams in each of the four groups progress followed by a knockout stage until a final winner is decided. This means that positions 16 to 12, 11 to eight, and seven to three are interchangeable, as these teams do not play each other to determine their final ranking. In order to give each team their own ranking, final positions were taken from the event page on Liquipedia[3].

The player base within Counter-Strike is aware of the importance of pistol rounds, as both teams are equally equipped and have an equal chance of success. This analysis indicates that eco- and force-rounds should not be treated as write-offs. Teams could benefit from developing more strategies for these low-buy rounds to generate more success.

However, as noted in Section 4.2.3, a team's attention should not be taken away from the full buy rounds. Cloud9 had the highest win rate in the low-buy and pistol rounds, compared to the rest of the competitors in the tournament, but did not advance past the initial group stage.

[3] https://liquipedia.net/counterstrike/ESL/One/2015/Katowice

4.3. RQ3: Do Younger Player Perform Better and Can We Predict Future Stars?

The first section of this research question looks to confirm the hypothesis that younger players perform better than their older counterparts. Furthermore, the most promising youth talent playing the game at this time will be singled out, and their future event ratings will be predicted.

4.3.1. Datasets

Both datasets used for this research question, the player statistics dataset and the player ratings from past events dataset, were cleaned in Section 3.2.3.

4.3.2. Approach

The purpose of this research question was to predict which young players will likely become the future stars of Counter-Strike. However, to decide if only younger players future careers were to be predicted, the question "Does age impact player skill?" needed to be answered. If the skill of older players deteriorates, then predicting their career ratings would be of little use as there is unlikely to be much improvement in their performances, as well as several older players could be close to retirement.

Younger Players vs Older Players

The dataset containing all the player statistics holds information such as 'average damage per round', 'average kills per round', 'average match rating' and the players 'Kill/Death Ratio', as well as general information about each player, such as name, age and team. With this information it made comparing two large groups of players in multiple ways possible.

From the dataset, the median age of the players was found to be 24. Two subsets of the dataset were created: one grouped all the player 21 or younger, while the other contained players 27 or older. Each group was at least three years from the median player age.

A players Kill/Death Ratio is simply the number of enemies that a player kills per death they receive and is easy to calculate as each player's total career kills and deaths are stored in the dataset. The average for all players on the website is slightly below one, with the highest rated player having a 1.41 Kill/Death ratio. The average match rating of players is exactly one, while the highest ranked player had an average score of 1.38. Consequently, these two statistics were the ideal method of comparing the two subsets.

The Most Talented Youth Players and Their Future Ratings

The first step in answering this research question was to single out the most promising youngsters in competitive Counter-Strike today. Several parameters were placed on the player dataset to establish which players to focus on. These parameters were: players aged 19 or under, players who had less than 500 professional Counter-Strike games played, players who had a career average match rating of above 1.1, and players who were a member of a top 20 team.

This narrowed the scope of the study down to 5 players. Mathieu '*ZywOo*' Herbaut (18), Yuri '*yuurih*' Santos (19), Kaike '*KSCERATO*' Cerato (19), Jere '*sergej*' Salo (17), and Ludvig '*Brollan*' Brolin (16).

The next step required each of the 5 players to be designated a group of older players who had similar career statistics at that stage of their respective careers. The findings in the related work mentioned in section 2.1. found that there was an average of a 0.1 drop in a player's career rating over the course of a four to five-year period, which is approximately equal to 100 events. With this in mind, five players with a rating of between 0.08 and 0.12 lower than each of the five younger players current rating, as well as having played over 900 maps played or over 90 events competed in were used to predict each youngster's future average rating.

For each of the five selected players, their respective experienced players were grouped together in the ratings from past events dataset (Section 3.2.3.). As this dataset was limited to important events only, the level of competition at these events would have been very high, implying that for a player to have played in one, they would have already been performing at a world class level. As the past events are in chronological order, the event ratings of the older players used to predict the younger players' futures will begin near the start of their respective careers at the top level. This is important as many of these younger players have recently transferred to better teams, only just beginning to play at the top level, and with the level of competition rising, a drop in rating is to be expected.

For each group of older players, if three or more of them participated in the same event, their ratings from that event were averaged together, and the resulting rating is added to their respective young players list of predicted ratings. Once these predicted ratings were added to the younger players actual past performances, a DataFrame with over 100 predicted event ratings was created for each player. Finally, an average of each of these DataFrames was collected to predict each player's overall career rating, and judge where they would appear on the HLTV players ranking leaderboard.

4.3.3. Results

Younger Players vs Older Players

Our comparison of younger players and older players gave the expected result of younger players outperforming their older counterparts, however, the degree to which they out class them was quite surprising.

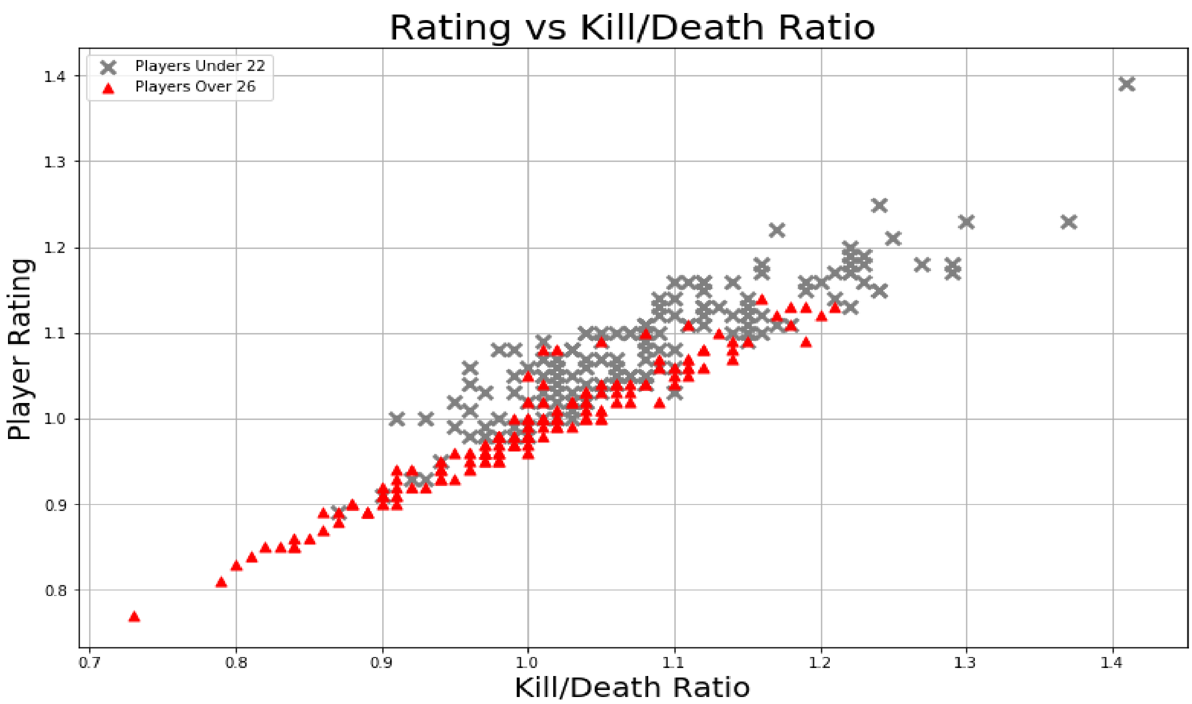

Figure 8. Comparison of Average Player Rating and Kill/Death Ratio of Old and Young Players

In Figure 8, the red triangles represent players who are over the age 26, while a grey X represents a player 21 or younger. For both 'Player Rating' and 'Kill/Death Ratio', a higher rating, as expected, is better. From the scatter plot, on average younger players outperform older players by a significant margin one that is much higher than we expected.

The Most Talented Youth Players and Their Future Ratings

This section of research singled out five young talented players and predicted them to be potential breakout stars in the professional Counter-Strike scene.

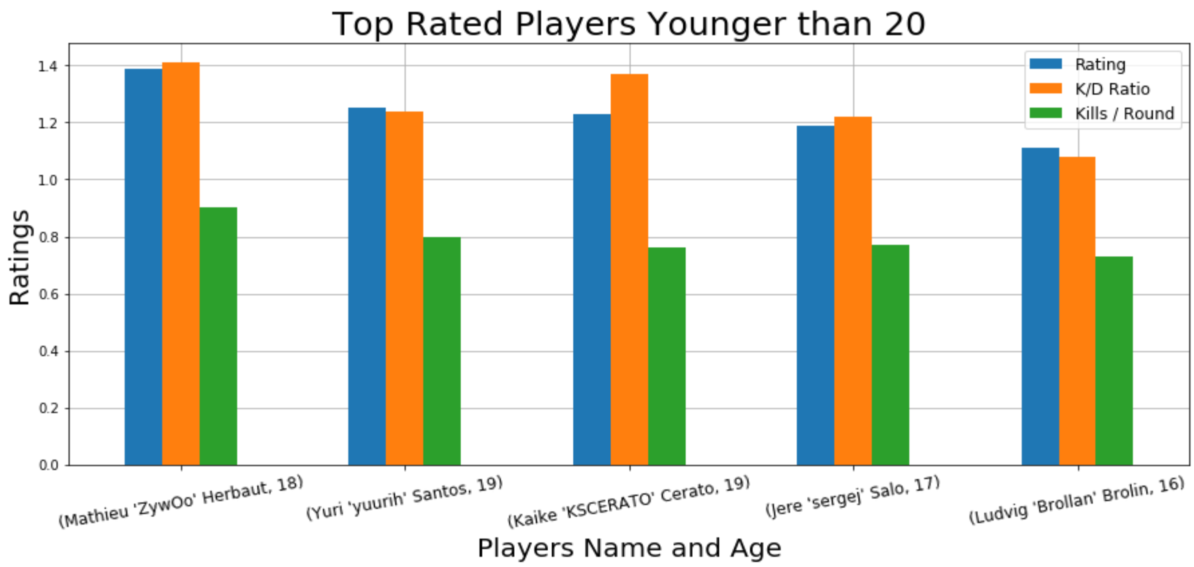

Figure 9. Bar Chart Showing the Top 5 Player Younger than 20 and their Ratings

The bar chart of Figure 9 gives us the names and ages of the most promising young players competing in the game right now, as well as a comparison of each players average event rating, Kill/Death Ratio and their average kills per round.

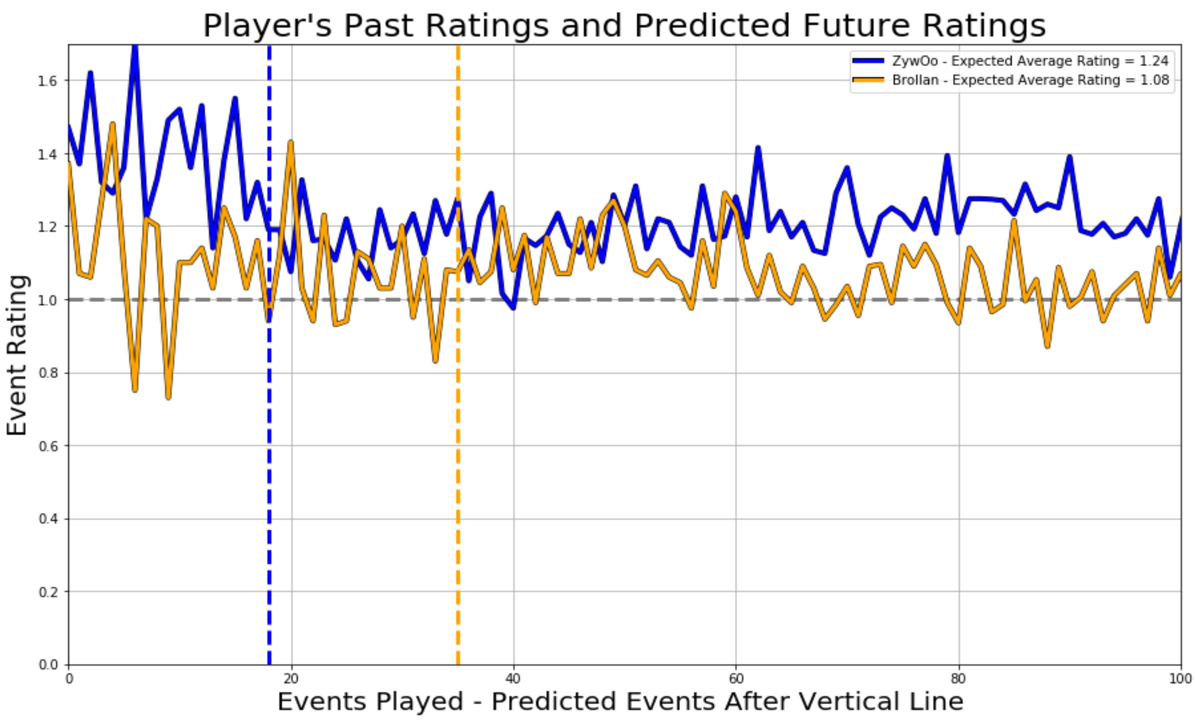

Figure 10. Line Graph Showing 'ZywOo's' and 'Brollan's' Past and Future Ratings

Figure 10 uses past event ratings from both players, as well as their predicted future ratings to plot the first 100 events of both *ZywOo's* and *Brollan's* careers. The predicted events begin at the vertical dotted line, coloured to match each player's plotted line. *ZywOo's* predicted average rating of 1.24 is extremely high, placing him near the top of the players ranking leaderboard. *Brollan's* expected rating of 1.08 is also very impressive, estimating him capable of being in the top 20 players of all time.

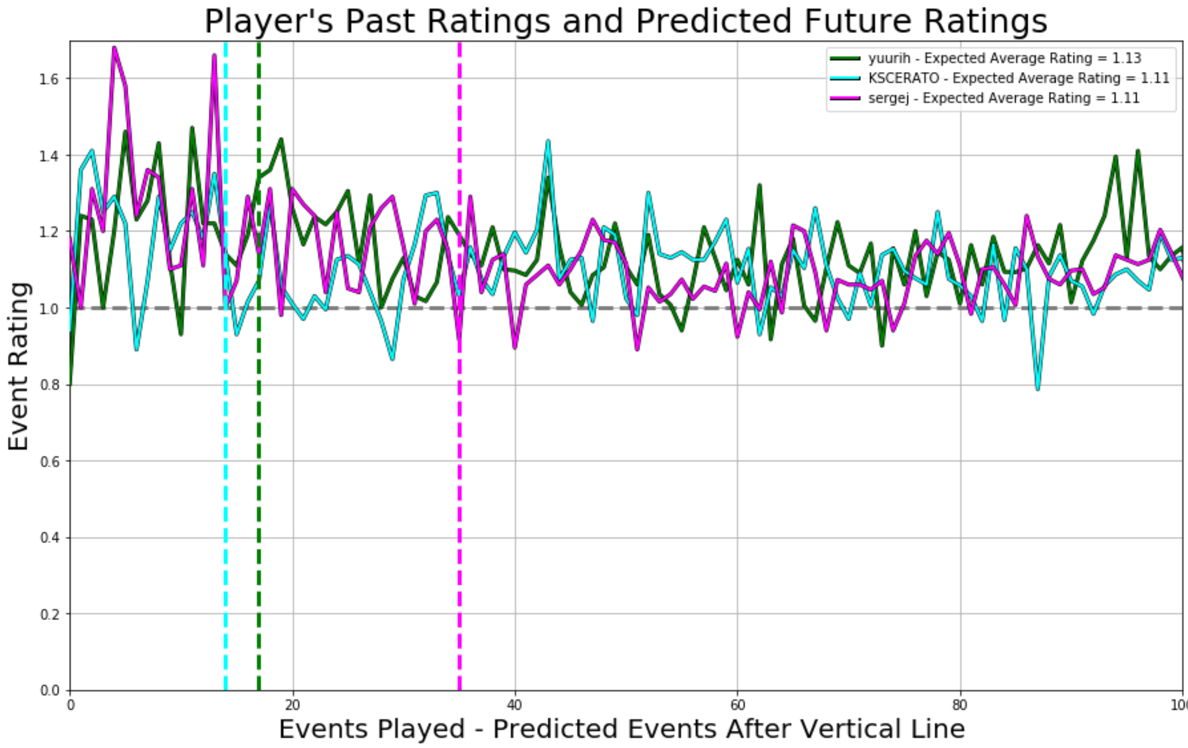

Figure 11. Line Graph Showing 'yuurih's', 'KSCERATO's' and 'sergej's' Past and Future Ratings

Figure 11 is similar to Figure 10, except this line graph plots the first 100 events of *yuurih's*, *KSCERATO's* and *sergej's*. Each player has very impressive expected average ratings, putting all 3 players as potential top 15 players.

4.3.4. Discussion

Younger Players vs Older Players

The discrepancy in 'Average Match Rating' and 'Kill/Death Ratio' between players younger than 22 and players older than 26 was much greater than initially expected. Close to half of the older players average below a 1 rating in both 'Average Match Rating' and 'Kill/Death Ratio', in contrast to the younger players, who very few have averages in the negative.

No older players are ranked in the top 20 players, moreover, every player ranked in the bottom 20 is over the age of 26. This points heavily towards teams benefiting from dropping older players from their roster, instead bringing in young talent. However, this high discrepancy could be somewhat explained.

Many of the younger players are still on tier two or tier three teams, leading to potentially inflated statistics as the level of competition they are facing is lower. On the other hand, most of the veterans have been involved in the competitive scene for a number of years, competing at the top level. On top of this, as players grow older, their in-game experience improves, with many veterans moving to a support or captain roll, using their skills to enable their younger teammates to perform at a higher level. The best example of this is the team Na'Vi. Na'Vi's has one of the oldest average roster ages yet is undoubtedly a top 5 team. The main reason for their success is the teams focus on enabling the number one ranked player in the world, Oleksandr "*s1mple*" Kostyliev; with many players sacrificing their own game to boost the chance of him playing well.

These, however, are not enough reason to negate the large difference in performance between age groups, as many younger players have supporting rolls, yet don't underperform to the same degree as the older competitors, confirming the hypothesis that on average younger players outperform their older counterparts.

The Most Talented Youth Players and Their Future Ratings

Having singled out the most talented players under the age of 20, seen in Figure 9, 'Average Match Rating', 'Kill/Death Ratio' and 'Average Kills per Round' were compared. This highlighted *ZywOo* as the most talented. It ranked the two 19-year-old Brazilians, *yuurih* and *KSCERATO*, second and third, while *sergej* and *Brollan*, aged 17 and 16 respectively, came in at 4th and 5th. It was surprising a 16-year-old was already performing at such a high level that they were able to included in this study.

Using groups of players who achieved similar event ratings in similar stages of their careers, predicted ratings for the first 100 events they will play in were produced. In Figure 10, *ZywOo's* and *Brollan's* past and predicted first 100 events are plotted. Before the prediction begins, there is a very high variance between the past event ratings of each player. This is likely due to the different levels of competition they were competing against in different events, and potentially roster changes to better teams.

ZywOo's predicted rating is incredibly impressive. If a rating of 1.24 is achieved, it would currently have *ZywOo* tied with Oleksandr "*s1mple*" Kostyliev on the leaderboard, the current top ranked player in the world.

When averaged, *Brollan's* expected average rating was 1.08, the lowest of the 5 youngsters. This is still very impressive and would put him tied 12th on the current leaderboards for all-time best players in big events, which were the only events considered in this dataset.

Likewise, Figure 11 gives us yuurih's, *KSCERATO's* and *sergej's* first 100 events, both past and predicted. It predicted these players to have a relatively similar capabilities, between 1.13 and 1.11 expected average rating, but still placing them as potential top 10 players

5. Discussion

In this section the ethical decisions taken are discussed along with ensuring reproducibility throughout our project and limitations that we faced during our work.

5.1. Ethical Considerations

All the data used in this study was collected from public websites, and in game files. Some personal data about every player was collected along with their in-game data, including their names, ages, and nationality.

In RQ3, the names of certain players were singled out for further analysis, this included two players under the age of 18. This was an ethical consideration for us, as dealing with personal data of minors is always a concern. However, these players are public figures (comparative to teenagers playing the

Premier League), and any information used in this research is open to the everyone, and any analysis done on these players in this was project was ethically appropriate.

5.2. Reproducibility

As future work continuing from this project is very likely, reproducibility was a very important aspect of this study. All notebooks were clearly laid out and explained in markdown cells and comments. As the data required for RQ1 and RQ2 was difficult to acquire (Section 3.1.1), all our collected datasets were uploaded to Google Sheets. This was done to ensure our data was publicly available, so that anyone can see our work on the project and can reproduce our findings if they so wish.

5.3. Limitations

With regards to RQ1 and RQ2, the dataset itself was limited. The Major tournaments were the only tournaments which have an in-game share code. This meant the match replay file did not need to be downloaded (approximately 500MB each) and then re-uploaded to DreamTeam. Instead, the share code itself could be passed into the website, saving much time and effort. The dataset was limited further by the cleaning process. This problem was unavoidable, but with more time, future work could focus on expanding the dataset.

Due to the dataset being limited, the correlation between low-buy and pistol success and tournament success only incorporated one tournament. In future work, this would ideally include the whole tournament dataset.

A limitation of RQ3 was the prediction of the player *Brollan*. As he is only 16 years old, there is a lot of potential for his skill to grow before even his 18th birthday. This would likely have a drastic effect on what his predicted average rating could be, and as there hasn't been many quite so talented 16-year-old players competing throughout the years, it was difficult to accommodate this. However the age of players entering the scene is constantly dropping, meaning that *Brollan's* career could potentially be used to predict players careers in the future.

6. Conclusions & Future Work

With this project, we set out to analyse how various factors impacted win rate within professional Counter-Strike and, by using past statistics, build a prediction model. When examining the win rate, we looked towards the first kill in a round and the in-game economy. The prediction model used the career trajectories of the more established pros to forecast the future for some of the younger players.

The analysis for RQ1 revealed the significance of the first kill on the outcome of a round. Teams who found that first kill had a win rate of approximately 0.73, regardless of the side they were on. Further examination showed that the time to first kill or the map that the match was played on had little impact that win rate.

The results for RQ2 indicated that mismanaging the in-game economy could lead to catastrophic results. The win rate of a match dropped below 0.5 after just two eco-rounds or five force-rounds. A positive correlation was drawn between tournament success and success in these low-buy rounds combined with the pistol rounds.

RQ3 confirmed a previous study into how increasing age has a negative impact on player performance. Subsequently, the prediction model anticipated that Mathieu "*ZywOo*" Herbaut was the most promising player aged 19 or under. It also predicted that he had the capacity to be as good as the current world number one, Oleksandr "*s1mple*" Kostyliev.

This project unveiled many potential routes for future work in relation to professional Counter-Strike. RQ1 only looks at the first kill, yet nine more players remain in a round. If the initial kill is returned by the opposing team, the round is left in a four-versus-four situation. Commentators often note how this favours the Terrorists, although this idea is purely anecdotal as we could find no research on the topic. A path for future work could be attempting to prove or disprove this hypothesis.

Another potential path for future work, mentioned in Section 4.2.4, is determining when a team should play for the win and when they should save. If a team knows when the statistically optimal time to save is, they could see a great improvement in their win rate. Also mentioned previously, in Section 5.3, the correlation between low-buy and pistol success and tournament success analysis would be expanded to include as many tournaments as possible.

In relation to RQ3, the prediction model used could be improved greatly with more time. The accuracy of the model could be analysed using events that are currently being played, and from those results, refined to give better predictions. The system could then be potentially automated for each player added to the HLTV websites, moreover, the level of competition competing in an event could be included in the calculations to give more accurate estimates.

7. Bibliography

https://www.reddit.com/r/GlobalOffensive/comments/adzwq8/analysis_how_much_does_age_matter_in_pro_csgo/

8. Appendix

Included in this section are some pieces of analysis that were related to the research questions but were not used in the main report for various reasons.

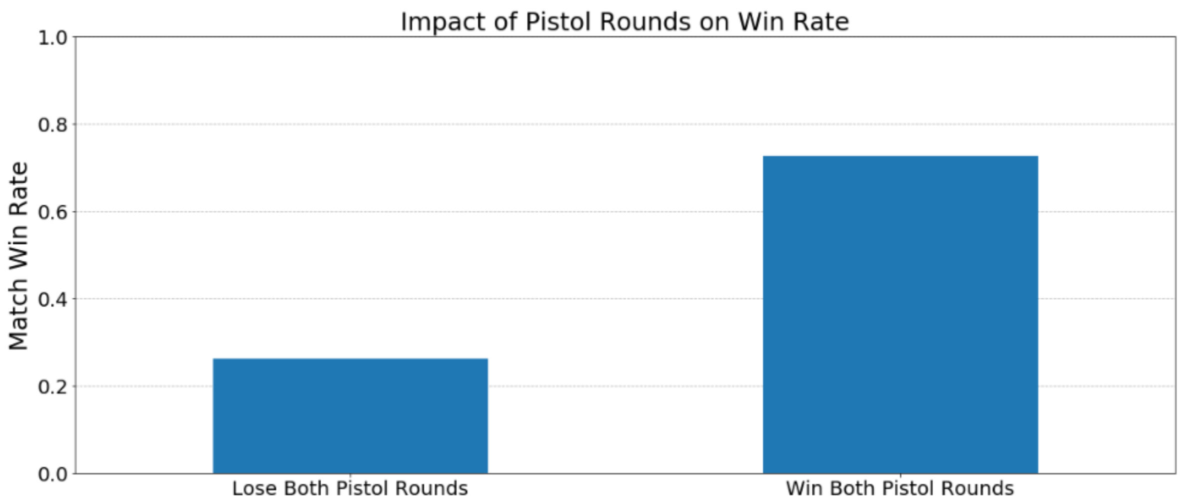

Figure 12. The Impact of the Pistol Rounds on Match Win Rate

Figure 12 shows the impact of winning or losing both the pistol rounds on the win rate of a match. This was only a quick exploration to show how important these rounds were and did not included teams who won one pistol round each, as that would require to determine which teams won which half. For this reason, it was not included in the main results section.