Manipulating VCT 2024 Data

Saketh Marrapu

2024-12-28

Reformatting the Data

Data Cleaning

Not all VCT events have all of the required data for each team. As such, these values were replaced with -1 when the data was being collected. In order to make this data usable we can make these values NA so the matches are still useful for the values that aren't missing.

```r
vct24dat <- read.csv("C:/Users/troll/Documents/valorantpredictive/rawvct2024data.csv",header = TRUE)
vct24dat[vct24dat==-1] = NA
```

Reformatting factors

As the data is now, it's aggregate data for each team in each match. This, however is not useful since for instance the total number of kills changes based on the number of rounds which also changes based on whether a match is best of 3 or best of 5. As such, these variables should be averaged across the number of rounds. This is done for the Kills, Deaths, and Assists.

```r
vct24dat$numRounds = vct24dat$FK + vct24dat$FD
vct24dat$KpR = vct24dat$KpR / vct24dat$numRounds
vct24dat$DpR = vct24dat$DpR / vct24dat$numRounds
vct24dat$ApR = vct24dat$ApR / vct24dat$numRounds
```

Some factors also need to be averaged for each player. For instance, the headshot percentage as an aggregate does not make sense since it is a sum of percentages. As such we should be averaging it across the number of players (5) in order to get the average headshot percentage of the team. We will do the same for KAST (Kill, Assist, Trade, Survive percent).

```r
vct24dat$KAST = vct24dat$KAST / 5
vct24dat$HS = vct24dat$HS / 5
```

For the sake of consistency we shall do the same for Average Damage per Round (ADR) and Average Combat Score (ACS) so we don't have a mix of aggregate factors and factors averaged across players. This should not affect the regression because we are simply scaling these values by a scalar.

```r
vct24dat$ACS = vct24dat$ACS / 5
vct24dat$ADR = vct24dat$ADR / 5
```

We do still, however, find that we have aggregate values in First Kills (FK) and First Deaths (FD). These variables also depend on the number of rounds which of course changes based on the number of maps played in a match. However, we cannot average this based on the number of rounds as you can only have 1 first kill or first death each round, and averaging would suggest something like having "0.5 first kills per round and 0.6 first deaths per round". As such, I will turn this into a categorical variable, FKFD which is True if a team has a higher or equal amount of First Kills than First Deaths and False if a team has less First Kills than First Deaths.

```
vct24dat$FKFD <- vct24dat$FK >= vct24dat$FD
```

Finally we'll convert the W/L stat into a boolean variable.

```
vct24dat$W.L <- vct24dat$W.L == 1
```

Reformatting Columns

First, I will remove the useless columns. Since we converted the First Kill and First Death columns into a boolean, we don't need them anymore. In addition, the numRounds column was only needed for calculation purposes and isn't relevant to the regression.

```
vct24dat <- vct24dat[,c(-8,-9,-11)]
```

As it is now, the columns only have the statistics for one team, however each match has two teams and the values change based on who is playing who. As such, the regression should be on both team's statistics. This means we need to make each row have the team's statistics as well as their opponents. However, we only require some of the opponents statistics. We only need on win loss column so we can remove the opponent's W.L. In addition we don't need the opponents FKFD, since that will be almost directly correlated with the first team's FKFD column as only one team can have less first kills and vice versa.

This completes the reformatting of data. We are left with a data frame that has 868 rows (one for each match in the Valorant Champions Tour 2024) and 16 columns. The columns are as follows: ACS: The average ACS of each player on a team during a match KpR: The average kills per round of the team during a match DpR: The average deaths per round of the team during a match ApR: The average assists per round of the team during a match KAST: The average percentage of rounds with kills, assists, survives, and trades per player on a team in a match ADR: The average damage per round of each player on a team during a match HS: The average headshot percentage of the team during a match FKFD: Whether a team had greater or equal first kills than first deaths oppACS: The opponent's ACS for the match oppKpR: The opponent's KpR for the match oppDpR: The opponent's DpR for the match oppApR: The opponent's ApR for the match oppKAST: The opponent's KAST for the match oppADR: The opponent's ADR for the match oppHS: The opponent's HS for the match W.L: Whether the team won or lost the match. This is the response variable.