

Predicting Competitive Rank in Overwatch

Does flexibility give you an edge?

Alicia Garcia

What is Overwatch?

- ▶ Overwatch is a first person shooter created by Blizzard, comprised of 26 characters across four group classification; DPS, Tank, Defense, and Support
- ▶ The structure of the game is based on team work to capture objectives, transport cargo objectives, or hold capture points.
- ▶ There are multiple games modes:
 - ▶ Arcade - challenges to play solo or as a team, thus obtaining lootboxes and XP
 - ▶ Quickplay - non-ranked matches where your team captures and holds points, transports cargo or captures objectives
 - ▶ Competitive - skill based matches, where you are grouped with others in the same skill rating, with same gameplay as Quickplay, open only during seasons

Overwatch Competitive: it's a fight to the top

- ▶ Overwatch competitive matches have a Skill Rating (SR) metric that places you in a rank, ranging from Bronze to Grand Master, determined by your performance in a competitive match (Win/Loss/Draw)
- ▶ OBSERVATIONS: good teamwork, a balanced team composition and skilled play of your character can increase your chances of winning a match.
- ▶ HYPOTHESIS: having the ability to play 2-3 characters (assuming well) can help create a stronger team composition, thus leading to more wins and higher SR.
- ▶ MAIN QUESTION: can we predict your SR based on the number of characters you play on console?

Where is the data from anyways?

- ▶ Obtaining data that includes player IDs, ranking, and time played on each character is not readily available from Blizzard publically.
- ▶ The data used in this study was obtained by contacting Blizzard through Onicmeta.com
- ▶ The contact I was paired with mined the data for Competitive Season 5-7 and sent it in csv formats
 - ▶ RangeIndex: 112,134 entries, 0 to 112133
 - ▶ Data columns (total 80 columns)
- ▶ The data obtained is confidential and not for public use, therefore dummy data has been created to publish repeatable work on GitHub.

Data and Hurdles

▶ Data Features

▶ Player (1 column):

- ▶ The player ID has been randomized to protect the players identity
- ▶ PROBLEM: the information is numeric, which poses issues for summing along axis = 1

▶ Rank (1 column):

- ▶ The skill rating of each Player
- ▶ PROBLEM: the data is numeric, causing issues when summing along axis =1, doesn't show the rank (Bronze, Silver, Gold, etc)

▶ Characters (78 columns):

- ▶ Each character has 3 columns with the stats totals: Orisa_Wins, Orisa_Loss, Orisa_Tie
- ▶ PROBLEM: too many columns to effectively work with in a dataframe and the raw data isn't useable

Exploring the Data

- ▶ Due to the size of the data, the information required needs to be extracted and presented into a new dataset to determine which character(s) is played the most.
- ▶ PROPOSED SOLUTION: sum the total number of games played to create total games played column. Then sum the total number of games played by each character (Wins/Loss/Tie) and divide this by the total games played to determine the Played Time Ratio.

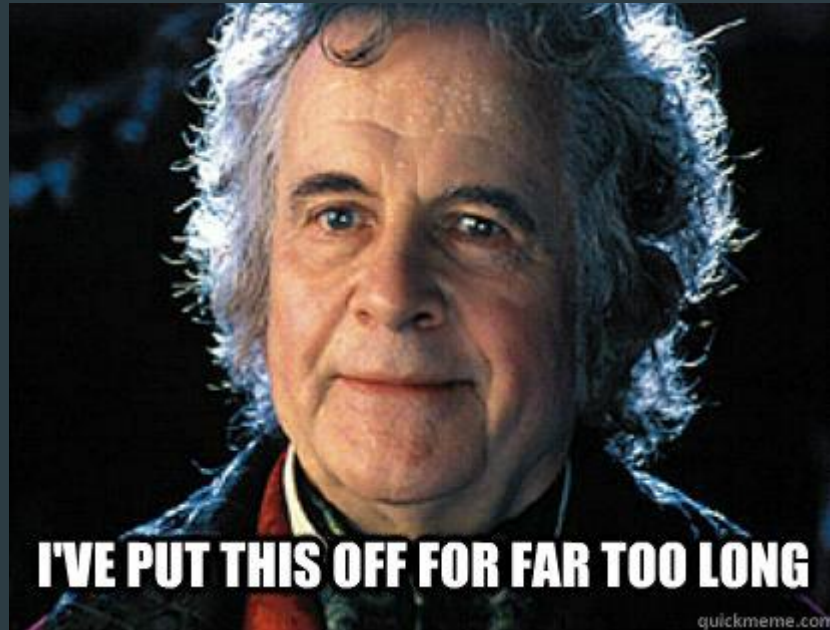
Using the ratio, add up the character whose ratios total 90% of games played, but do not include characters with less 5% play time (ratio) for a total characters played a season.

This will provide us information to bring our 80 columns down to 4 or 5

- ▶ Player_id, skillrating, totalGames, NumCharPlayed, (Rank)

How successful have I been?

- Progress to date: so far I've been able to create a totalGames columns, removed all whitespaces in the strings.



TO DO: find a code to successfully sum every 3 columns, by index, for the characters and create the character ratio column. The proceed to polish the data