Ritwik Katiyar

Project Proposal

Applying Basic Analytics to E-Sports using Python

Introduction:

Currently in e-sports one of the most highly competitive games is League of Legends. League of legends is defined as a MOBA (multiplayer online battle arena). In which 5 players compete against another 5 players to destroy the other team's nexus. The game consists of characters that are referred to as champions with abilities that are unique to the character that each player controls. So, in a game 10 player's means 10 champions that can be played out of the current pool of 163 champions. To archive victory a team has to work together to acquire gold by killing the enemy team's champion or by acquiring certain objectives to build a lead.

Every year from October to November a World championship event is held were teams from different regions compete in order to win the grand prize of $2.5 million. Along with the championship there is an event held for the fans to guess which team would win. The event also features questions such as which player would have the highest number of kills or which in game character would be most played. This event is known as the Crystal ball event.

To understand worlds we first need to understand the region tier system. If a region performed better at worlds previously then that region is given an extra spot at worlds or a higher tear. This year's region tiers are as follows:

I. China and Korea- Three teams from each region in the main event and one in 'play-ins'
II. Europe- Two teams in the main event and two in 'play-ins'
III. North America- Two teams in the main event and one in 'play-ins'
IV. South East Asia and PCS (Hong Kong, Macau and Taiwan)- One team from each region in the main event and in 'play-ins'
V. Japan, Brazil, Australia, Latin America and Turkey – One team from each region playing in a 'play-ins'

Over all there are 24 teams attending worlds. Teams qualifying into worlds through 'play-ins' have the possibility of playing a total of 37 games assuming they make it all the way to the tournament finals and play all 5 games in a best of five. Similarly, teams that are part of the main event have a possibility of playing 21 games.

Now that we know what the worlds event is let's look at the crystal ball event. The crystal ball event asks a number of questions; these questions can be divided by different groups Champions, Players, Teams, Event. Each group contains five questions pertaining to the groups. However, we will be focusing on answering a few of the questions due to time constraints.

Questions and Methods:

- Champions
  - Which champion will have the highest win rate? (min 5 games)
    - Logistic Regression: With feature being weather that champion was used by a winning team or not.
- Players:
  - Who will have the highest KDA at worlds (KDA refers to (Kill + assist)/ death)
    - Compute the kill death ratio of all players of teams that are attending worlds.

        - Utilize a model that can predict the player KDA for a game based off the games the player played in the past and compute it for 37 games if the player is in a team that is in 'play-ins' and 21 for player is in teams that is part of the main event and compute the average.
        - However, due to region bias there will be a preference towards players from upper tier regions over lower tier regions. Due to the fact that if the player is used to playing in a highly competitive region and has a lower KDA would likely perform better vs. a player in less competitive region with higher KDA.
- Teams
    - Which team will win worlds?
        - What variables best contribute to a team winning a game.
        - Utilize a model that can generate the odds of a team that is attending worlds, winning a game. (based on if (on average) the team possess the variables that is known to contribute towards a team winning)
        - If the team is in 'play-ins', then generate the odds of the team winning 37 games. If not then the odds of team winning 21 games. Team with the best odds would be predicted to be the winners.
- Event
    - What will be the Duration of the longest game at Worlds?
        - For this the options include: 35-40 min, 40-45 mins, 45-50 mins and 50+
        - Compute the probability of each option and pick the option that would be the 'most reasonable'.

Dataset:

To acquire data the following website will be used https://oracleselixir.com/ . The website offers information on every game ever played for a full season of league (A regional season consists of two splits for every region which consists of play offs and regular season for each split). That will be the data set that will be used in order to answer the questions. The full data set contains:

- Full match data.
- Player's performance within a match.
- Over all team's performance.
- Champions played and banned by the teams.
- Over all there are 123 columns and 141,973 rows.

Work Plan:

Each Question will be treated like a mini project. Hence the following steps would be applied to all questions.

Step-1: Clean and organize the data.

Step-2: Go through the method and generate the model or process the data.

Step-3: Analyze the data or the model. Make sure the results make sense and the model seems to be working and is relatively accurate.

Step-4: Make Conclusions and finalize the predictions.

Step-5: Check if predictions were true.