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.

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.

Questions and Methods:

- Champions
    - Which champions had the highest win rate\pick rate and ban rate? (min 5 games)
        - Clean the data and obtain games from the worlds series
        - Aggregate on a specific champion
        - Construct various plots to analyze the results
        - Use pandas, seaborn, and matplot lib libraries
- 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 linier regression to 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.
        - Using Pandas, statsmodels.api, matplotlib and, sklearn libraries to answer the question
- Regions
    - What is the tier ranking for each region at worlds for the next worlds?
        - This is calculated by first calculating the 'performance' for each team from each region at every world. The performance contains two factors; the number of wins and the gold differential the team had when they won a game. Bigger gold differential means a more decisive win. Based on these factors each region is assigned a tier. The goal is to see if the tiers have changed at all after this year's world's event.
        - To do this we clean the data to obtain necessary games
        - Calculate the 'performance' for each team
        - Rank the performance for each region based on the team's rankings
        - Plot and analyze the data. Using the geo pandas construct a plot that shows each region. And their rankings
        - Using pandas, matplot lib, and geo pandas libraries.
- Event
    - What was the Duration of the games at Worlds?
        - Clean the data accordingly
        - Get game times for each game
        - Perform analysis and visually display the results using plots
        - Using pandas, seaborn and matplot lib libraries.