**Achievement 6 project**

**Regional League of Legends Data**

# data source: <https://www.kaggle.com/kevindsouza2794/league-of-legends-regional-data>

## data collected by: kevin d’souza

Data set was scrapped using Riot Games API by Kevin D’Souza. League of Legends is one of the most popular online MOBA (Multiplayer Online Battle Arena) games in the world played around the world. The data we will be looking at contains match data for the **regions**: Brazil, Europe Nordic & East, Europe West, Korea, Latin America, North America, Oceania, Russia, and Turkey.

Each region had **20,000 unique matches** collected with 10 total players for 5 versus 5 matches. Refer to the data dictionary for all the data collected, but for quick reference, the match data consists of: Game Duration, Player bans, Player picks, Game win, Kills, Deaths, Wards placed, etc.

The time the data was collected fits somewhere between May 31st, 2018 and December 5th, 2018. We can deduce the information by champion picks, **Pyke’s** release was in May 31st and there are match records with the champion selected and the next released champion ‘**Neeko’** would be released December 5th, however, there are no records of Neeko being selected. To further the investigation, we can use column ‘gameCreation’ which has records of time and date of match creation in ‘UNIX epoch format’, which shows this data set was pulled between August – September.

I am choosing to do this project as I have played League of Legends since **Season Three** which started in February 1st, 2013. I have a lot of knowledge of the game and would hope to one day work as an analyst in the video game industry whether it be working with developers or working in Esports. This project to me gives me the experience of working in the industry I would like to grow into.

# potential issues:

This data this data was collected independently and is prone to issues with the way it is collected and sorted. Through just a checked check through the data and through a discussion post on Kaggle by user ‘ehobbs1705’ we can see some issues with the data.

Summoner spells listed as ‘spell1PlayerX’ & ‘spell2PlayerX’ (X being filled by 1-5 indicating player) show the same spell across the data set. In other words, ‘spell1PlayerX’ value is also identical to ‘spell2PlayerX’ value which is not possible so the collection had an error. This unfortunately removes a quick easy flag we could’ve created as the analyst as Spell ‘Smite’ is an indication for what Player was rolled as ‘Jungler’ for the match.

Another issue is that Team 1 & Team 2 are split into 2 different rows only connected through matching columns: ‘matchid’, ‘seasonid’,’gameDuration’,’gameCreation’. This creates a slight problem as now every match is split into 2 different rows, which can cause issues with analyzing as you carefully have to ensure the match ids are identical to analyze the match data. This would also skew our median values in our dataset unless dealt with.

‘gameDuration’ column is described, by the author Kevin D’Souza, “The duration of the game played in milli seconds”. When reviewing the data however, I notice that this is incorrect and the data is actually in seconds. A game lasting ‘1278’ milli seconds is just not possible in League of Legends, but a game lasting ‘1278’ seconds or 21 minutes and 18 seconds is a much more appropriate game length. So, for analyzing this data ‘gameDuration’ column will be referred to in seconds.

# data profile

## Cleaning the data

### Duplicated rows

When checking for duplicated rows, we didn’t find any across the 10 data sets for the 10 regions we are working with.

### missing values

When checking for missing values, .value\_counts() in Jupyter showed that I wasn’t missing any values, but when checking the CSV’s in Excel I noticed the columns ‘banX’ having a ‘*blank’* value. Also being familiar with the game allowed me to know that during draft phase, players have the opportunity to not ban a champion, thus the *‘blank’* value was used to fill the space. Having a *‘blank’* value isn’t helpful for people viewing the data, so we changed all ‘*blank’* value to ‘No Ban’

### dropping columns

Due to a collection issue, we had to remove all ‘spellXPlayerY’ columns due to the information not being useful. In a match a player would have the choice of two unique summoner spells, but Spell 1 and Spell 2 were identical in the data set which isn’t possible in game, since the data is incorrect therefore not helpful, we removed these columns from our data.

# Data Dictionary

|  |  |  |  |
| --- | --- | --- | --- |
| Column Name | Structured or Unstructured | Data Type | Description |
| matchid | Structured | Unique Id | Unique Id for a given match, used to match 2 rows that signify the two teams playing. |
| seasonid | Structured | Unique Id | Unique Id for a given match that shows what season the game was played in (‘seasonid 11’ refers to Season 8) |
| gameDuration | Structured | Integer | How long in seconds a match was played for |
| gameCreation | Structured | String | Details of when the match was started in UNIX epoch format |
| win | Structured | Boolean/ Integer | Boolean value using 0,1, 0 being a loss and 1 being a win |
| team | Structured | Boolean/ Integer | Boolean value using 0,1, 0 being red side and 1 being blue side |
| wardsPlaced | Structured | Integer | Total number of wards placed for the team |
| firstblood | Structured | Boolean/ Integer | Boolean value using 0,1, 1 being the team that secured the objective, while 0 indicating the team didn’t secure the objective. (It is possible for both teams to have 0, but both teams cannot have 1) |
| firstTower | Structured | Boolean/ Integer | Boolean value using 0,1, 1 being the team that secured the objective, while 0 indicating the team didn’t secure the objective. (It is possible for both teams to have 0, but both teams cannot have 1) |
| firstInhibitor | Structured | Boolean/ Integer | Boolean value using 0,1, 1 being the team that secured the objective, while 0 indicating the team didn’t secure the objective. (It is possible for both teams to have 0, but both teams cannot have 1) |
| firstDragon | Structured | Boolean/ Integer | Boolean value using 0,1, 1 being the team that secured the objective, while 0 indicating the team didn’t secure the objective. (It is possible for both teams to have 0, but both teams cannot have 1) |
| firstRiftherald | Structured | Boolean/ Integer | Boolean value using 0,1, 1 being the team that secured the objective, while 0 indicating the team didn’t secure the objective. (It is possible for both teams to have 0, but both teams cannot have 1) |
| firstBaron | Structured | Boolean/ Integer | Boolean value using 0,1, 1 being the team that secured the objective, while 0 indicating the team didn’t secure the objective. (It is possible for both teams to have 0, but both teams cannot have 1) |
| teamKills | Structured | Integer | Number of kills as a team |
| towerKills | Structured | Integer | Number of Tower structures taken as a team |
| inhibitorKills | Structured | Integer | Number of Inhibitor structures take as a team |
| dragonKills | Structured | Integer | Number of Dragons slain by team |
| riftHeraldKills | Structured | Integer | Number of Rift Heralds slain by team |
| baronKills | Structured | Integer | Number of Barons slain by team |
| ban1 | Structured | String | Name of champion banned by player 1 during draft phase |
| ban2 | Structured | String | Name of champion banned by player 2 during draft phase |
| ban3 | Structured | String | Name of champion banned by player 3 during draft phase |
| ban4 | Structured | String | Name of champion banned by player 4 during draft phase |
| ban5 | Structured | String | Name of champion banned by player 5 during draft phase |
| pick1 | Structured | String | Name of champion picked by player 1 during champ select |
| pick2 | Structured | String | Name of champion picked by player 2 during champ select |
| pick3 | Structured | String | Name of champion picked by player 3 during champ select |
| pick4 | Structured | String | Name of champion picked by player 4 during champ select |
| pick5 | Structured | String | Name of champion picked by player 5 during champ select |
| player1Kills | Structured | Integer | Number of individual kills for player 1 |
| player2Kills | Structured | Integer | Number of individual kills for player 2 |
| player3Kills | Structured | Integer | Number of individual kills for player 3 |
| player4Kills | Structured | Integer | Number of individual kills for player 4 |
| player5Kills | Structured | Integer | Number of individual kills for player 5 |
| soloKills | Structured | Integer | Count of how many team members participated in a kill. Solo Kill is 0 assistants |
| duoKills | Structured | Integer | Count of how many team members participated in a kill. Duo Kill is 1 killer, 1 assistant |
| trioKills | Structured | Integer | Count of how many team members participated in a kill. Trio Kill is 1 killer, 2 assistants |
| quadKills | Structured | Integer | Count of how many team members participated in a kill. Quad Kill is 1 killer, 3 assistants |
| pentaKills | Structured | Integer | Count of how many team members participated in a kill. Penta Kill is 1 killer, 4 assistants |

# understanding our data

### statistical analysis

## Match data brazil

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Column Name | Min | Max | Mean | Median |
| gameDuration | 190 | 3347 | 1618.14 | 1625 |
| wardsPlaced | 0 | 404 | 58.24 | 53 |
| teamKills | 0 | 77 | 23.43 | 23 |
| towerKills | 0 | 11 | 5.44 | 6 |
| inhibitorKills | 0 | 10 | 0.88 | 0 |
| dragonKills | 0 | 6 | 1.26 | 1 |
| riftHeraldKills | 0 | 1 | 0.41 | 0 |
| baronKills | 0 | 4 | 0.43 | 0 |
| player1Kills | 0 | 30 | 2.64 | 1 |
| player2Kills | 0 | 34 | 5.15 | 4 |
| player3Kills | 0 | 30 | 5.21 | 4 |
| player4Kills | 0 | 30 | 5.18 | 4 |
| player5Kills | 0 | 29 | 5.22 | 4 |
| soloKills | 0 | 25 | 5.07 | 5 |
| duoKills | 0 | 36 | 9.38 | 9 |
| trioKills | 0 | 27 | 7.27 | 7 |
| quadKills | 0 | 24 | 3.44 | 3 |
| pentaKills | 0 | 14 | 0.85 | 0 |

## Match data Europe Nordic & East

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Column Name | Min | Max | Mean | Median |
| gameDuration | 190 | 3707 | 1594.11 | 1594 |
| wardsPlaced | 0 | 477 | 53.43 | 47 |
| teamKills | 0 | 78 | 23.72 | 23 |
| towerKills | 0 | 11 | 5.35 | 6 |
| inhibitorKills | 0 | 11 | 0.86 | 0 |
| dragonKills | 0 | 6 | 1.22 | 1 |
| riftHeraldKills | 0 | 1 | 0.37 | 0 |
| baronKills | 0 | 4 | 0.41 | 0 |
| player1Kills | 0 | 29 | 2.67 | 1 |
| player2Kills | 0 | 41 | 5.23 | 4 |
| player3Kills | 0 | 33 | 5.27 | 4 |
| player4Kills | 0 | 32 | 5.25 | 4 |
| player5Kills | 0 | 30 | 5.27 | 4 |
| soloKills | 0 | 31 | 5.55 | 5 |
| duoKills | 0 | 39 | 9.67 | 9 |
| trioKills | 0 | 31 | 7.01 | 7 |
| quadKills | 0 | 20 | 3.23 | 3 |
| pentaKills | 0 | 15 | 0.83 | 0 |

## Match data Europe west

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Column Name | Min | Max | Mean | Median |
| gameDuration | 190 | 3828 | 1591.66 | 1589 |
| wardsPlaced | 0 | 458 | 57.82 | 52 |
| teamKills | 0 | 82 | 22.84 | 22 |
| towerKills | 0 | 11 | 5.38 | 5 |
| inhibitorKills | 0 | 11 | 0.82 | 0 |
| dragonKills | 0 | 6 | 1.16 | 1 |
| riftHeraldKills | 0 | 1 | 0.41 | 0 |
| baronKills | 0 | 4 | 0.41 | 0 |
| player1Kills | 0 | 38 | 2.59 | 1 |
| player2Kills | 0 | 29 | 5.02 | 4 |
| player3Kills | 0 | 33 | 5.08 | 4 |
| player4Kills | 0 | 30 | 5.06 | 4 |
| player5Kills | 0 | 34 | 5.08 | 4 |
| soloKills | 0 | 29 | 4.81 | 4 |
| duoKills | 0 | 40 | 9.19 | 9 |
| trioKills | 0 | 29 | 7.08 | 7 |
| quadKills | 0 | 23 | 3.35 | 3 |
| pentaKills | 0 | 18 | 0.88 | 0 |

## Match data korea

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Column Name | Min | Max | Mean | Median |
| gameDuration | 190 | 34 | 1470.57 | 1461 |
| wardsPlaced | 0 | 24 | 59.04 | 54 |
| teamKills | 0 | 305 | 20.31 | 20 |
| towerKills | 0 | 79 | 4.59 | 4 |
| inhibitorKills | 0 | 11 | 0.58 | 0 |
| dragonKills | 0 | 8 | 1.01 | 1 |
| riftHeraldKills | 0 | 6 | 0.40 | 0 |
| baronKills | 0 | 1 | 0.35 | 0 |
| player1Kills | 0 | 4 | 2.30 | 1 |
| player2Kills | 0 | 30 | 4.48 | 4 |
| player3Kills | 0 | 30 | 4.51 | 4 |
| player4Kills | 0 | 28 | 4.47 | 4 |
| player5Kills | 0 | 26 | 4.52 | 4 |
| soloKills | 0 | 23 | 3.85 | 3 |
| duoKills | 0 | 31 | 7.98 | 8 |
| trioKills | 0 | 25 | 6.46 | 6 |
| quadKills | 0 | 22 | 3.28 | 3 |
| pentaKills | 0 | 20 | 0.96 | 0 |

## Match data latin america north

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Column Name | Min | Max | Mean | Median |
| gameDuration | 190 | 4429 | 1653.40 | 1669 |
| wardsPlaced | 0 | 478 | 57.42 | 51 |
| teamKills | 0 | 82 | 23.21 | 23 |
| towerKills | 0 | 11 | 5.48 | 6 |
| inhibitorKills | 0 | 8 | 0.94 | 0 |
| dragonKills | 0 | 7 | 1.33 | 1 |
| riftHeraldKills | 0 | 1 | 0.39 | 0 |
| baronKills | 0 | 4 | 0.40 | 0 |
| player1Kills | 0 | 36 | 2.64 | 1 |
| player2Kills | 0 | 32 | 5.15 | 4 |
| player3Kills | 0 | 37 | 5.15 | 4 |
| player4Kills | 0 | 33 | 5.13 | 4 |
| player5Kills | 0 | 30 | 5.12 | 4 |
| soloKills | 0 | 32 | 5.09 | 5 |
| duoKills | 0 | 41 | 9.19 | 9 |
| trioKills | 0 | 28 | 7.06 | 7 |
| quadKills | 0 | 22 | 3.49 | 3 |
| pentaKills | 0 | 15 | 0.92 | 0 |

## Match data latin america south

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Column Name | Min | Max | Mean | Median |
| gameDuration | 190 | 3796 | 1652.85 | 1665 |
| wardsPlaced | 0 | 350 | 59.04 | 54 |
| teamKills | 0 | 73 | 22.82 | 23 |
| towerKills | 0 | 11 | 5.53 | 6 |
| inhibitorKills | 0 | 11 | 0.93 | 0 |
| dragonKills | 0 | 6 | 1.31 | 1 |
| riftHeraldKills | 0 | 1 | 0.40 | 0 |
| baronKills | 0 | 4 | 0.42 | 0 |
| player1Kills | 0 | 31 | 2.59 | 1 |
| player2Kills | 0 | 31 | 5.01 | 4 |
| player3Kills | 0 | 30 | 5.05 | 4 |
| player4Kills | 0 | 29 | 5.07 | 4 |
| player5Kills | 0 | 33 | 5.07 | 4 |
| soloKills | 0 | 31 | 4.92 | 4 |
| duoKills | 0 | 33 | 9.07 | 9 |
| trioKills | 0 | 27 | 6.98 | 7 |
| quadKills | 0 | 22 | 3.43 | 3 |
| pentaKills | 0 | 16 | 0.91 | 0 |

## Match data north america

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Column Name | Min | Max | Mean | Median |
| gameDuration | 190 | 4268 | 1592.48 | 1580 |
| wardsPlaced | 0 | 439 | 57.10 | 51 |
| teamKills | 0 | 83 | 22.34 | 22 |
| towerKills | 0 | 11 | 5.24 | 5 |
| inhibitorKills | 0 | 9 | 0.80 | 0 |
| dragonKills | 0 | 6 | 1.25 | 1 |
| riftHeraldKills | 0 | 1 | 0.41 | 0 |
| baronKills | 0 | 4 | 0.42 | 0 |
| player1Kills | 0 | 29 | 2.54 | 1 |
| player2Kills | 0 | 29 | 4.93 | 4 |
| player3Kills | 0 | 32 | 4.96 | 4 |
| player4Kills | 0 | 30 | 4.93 | 4 |
| player5Kills | 0 | 30 | 4.97 | 4 |
| soloKills | 0 | 26 | 4.388 | 4 |
| duoKills | 0 | 33 | 8.69 | 8 |
| trioKills | 0 | 33 | 7.07 | 7 |
| quadKills | 0 | 22 | 3.62 | 3 |
| pentaKills | 0 | 15 | 1.02 | 0 |

## Match data oceania

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Column Name | Min | Max | Mean | Median |
| gameDuration | 190 | 3723 | 1687.58 | 1679 |
| wardsPlaced | 0 | 477 | 56.60 | 50 |
| teamKills | 0 | 80 | 24.80 | 24 |
| towerKills | 0 | 11 | 5.51 | 6 |
| inhibitorKills | 0 | 9 | 0.93 | 0 |
| dragonKills | 0 | 6 | 1.32 | 1 |
| riftHeraldKills | 0 | 1 | 0.35 | 0 |
| baronKills | 0 | 4 | 0.42 | 0 |
| player1Kills | 0 | 32 | 2.80 | 1 |
| player2Kills | 0 | 36 | 5.47 | 5 |
| player3Kills | 0 | 34 | 5.48 | 5 |
| player4Kills | 0 | 35 | 5.50 | 5 |
| player5Kills | 0 | 31 | 5.52 | 5 |
| soloKills | 0 | 33 | 5.70 | 5 |
| duoKills | 0 | 34 | 9.97 | 10 |
| trioKills | 0 | 29 | 7.33 | 7 |
| quadKills | 0 | 23 | 3.53 | 3 |
| pentaKills | 0 | 15 | 0.93 | 0 |

## Match data russia

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Column Name | Min | Max | Mean | Median |
| gameDuration | 190 | 3561 | 1624.53 | 1625 |
| wardsPlaced | 0 | 434 | 54.18 | 47 |
| teamKills | 0 | 86 | 25.42 | 25 |
| towerKills | 0 | 11 | 5.42 | 6 |
| inhibitorKills | 0 | 10 | 0.89 | 0 |
| dragonKills | 0 | 6 | 1.25 | 1 |
| riftHeraldKills | 0 | 1 | 0.35 | 0 |
| baronKills | 0 | 4 | 0.39 | 0 |
| player1Kills | 0 | 34 | 2.87 | 1 |
| player2Kills | 0 | 37 | 5.60 | 5 |
| player3Kills | 0 | 35 | 5.67 | 5 |
| player4Kills | 0 | 35 | 5.60 | 5 |
| player5Kills | 0 | 38 | 5.67 | 5 |
| soloKills | 0 | 29 | 6.27 | 6 |
| duoKills | 0 | 39 | 10.58 | 10 |
| trioKills | 0 | 29 | 7.31 | 7 |
| quadKills | 0 | 23 | 3.26 | 3 |
| pentaKills | 0 | 13 | 0.78 | 0 |

## Match data turkey

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Column Name | Min | Max | Mean | Median |
| gameDuration | 190 | 3373 | 1616.67 | 1621 |
| wardsPlaced | 0 | 418 | 51.04 | 46 |
| teamKills | 0 | 85 | 25.29 | 25 |
| towerKills | 0 | 11 | 5.40 | 6 |
| inhibitorKills | 0 | 10 | 0.87 | 1 |
| dragonKills | 0 | 6 | 1.20 | 1 |
| riftHeraldKills | 0 | 1 | 0.37 | 0 |
| baronKills | 0 | 4 | 0.41 | 0 |
| player1Kills | 0 | 31 | 2.85 | 1 |
| player2Kills | 0 | 33 | 5.56 | 5 |
| player3Kills | 0 | 35 | 5.61 | 5 |
| player4Kills | 0 | 33 | 5.63 | 5 |
| player5Kills | 0 | 32 | 5.61 | 5 |
| soloKills | 0 | 40 | 5.96 | 5 |
| duoKills | 0 | 40 | 10.39 | 10 |
| trioKills | 0 | 31 | 7.47 | 7 |
| quadKills | 0 | 20 | 3.41 | 3 |
| pentaKills | 0 | 14 | 0.85 | 0 |

# limitations and ethics

## Limitations

Our data is limited to Season 8 but specifically August2018 – September 2018, so we aren’t looking at a full season’s worth of data. This means that our data can naturally have biases towards the data as some champions will be picked more than others due to popularity in the Meta. This means what our data is really showing is just a quick snapshot of a particular meta or what was popular for these months of the season.

## Ethics

For any issues with data ethics or securities, there threat level is non-existent. The data collected through Riot Games API did not include any customer information, so no usernames, emails, passwords were collected so we don’t have any issues of leaks. We are also working with old data, and with video games that do big season changes, no information learned from here will have any impact on Season 11 game.

# questions to explore

1. What regions have the longest game durations?
2. What region have the most placed wards?
3. Do high wards placed make you more likely to win?
4. Are there any trends in the players? Do they change between regions?
5. What champions have the highest pick rate? Does it change between region?