Javi Ramirez

Career Foundry Data Immersion Portfolio

League of Legends Season 8 Regional Data – Case Study

Riot Games – League of Legends

- **Riot Games** is a video game development company with one of the most successful video games of all time, **League of Legends**, as their forerunner of their company. League of Legends has paved the way for huge developments in Esports and continues to grow in terms of size through player base, audience, sponsorships and more. Commonly the game goes through big changes Season to Season with some big changes happening in the middle of a Season.

Objective

- My objective was to perform an informative analysis on match data across several regions/servers.
- The primary goal learning any trends on how the current state League of Legends is played and how we can analyze the information brought forth to then provide recommendations for upcoming patches/changes.

Data

- The main data set was scrapped by the use of Riot Games API by Kevin D'Souza who then hosted the data set on Kaggle for open use. The data set consists of 10 csv files with 20,000 matches recorded during Season 8 between August – September, 2018.
- Another data set used was geographical shape data provided by naturalearthdata.com

Main data: https://www.kaggle.com/kevindsouza2794/league-of-legends-regional-data

Geo shape data: https://www.naturalearthdata.com/downloads/110m-cultural-vectors/110m-admin-0-countries/

Potential Issues

- This data set was collected independently and can be prone to human errors in the way it is collected and sorted.
- Columns ‘spell1PlayerX’ & ‘spell2PlayerX’ had same values stored inside, indicating a collection error and therefore had to be removed from our analysis.
- Although 20,000 match records are recorded, the teams are split into 2 separate rows only linked together by a column ‘match_id’, so for 20,000 matches with have 40,000 rows of data.

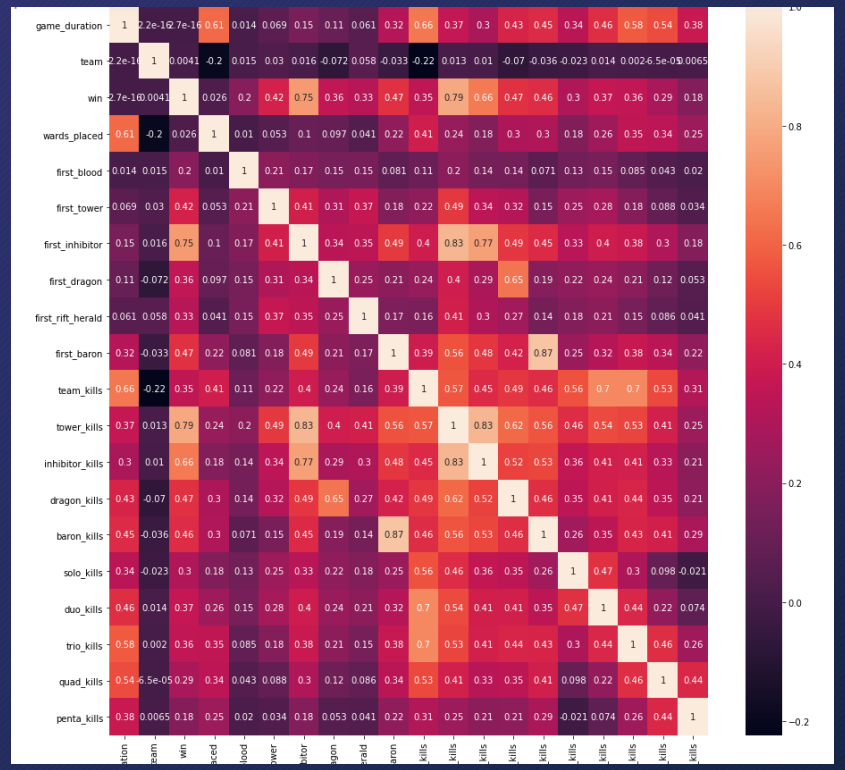
Data Preparation

- There was a bit of cleaning to do with our data:
  - Dropping columns deemed unnecessary due to collection errors
  - Renaming all the columns for their weird format
  - Creating new columns based off Boolean columns for clarity ie. ‘team’ = ‘map_side’
  - Creating a server column for merging data sets together.

```
In [72]: # You can do this is one line of code, but makes it less visually appealing for these long games
         merge_5.rename(columns = {'seasonId' : 'season_id' }, inplace = True)

In [73]: merge_5.rename(columns = {'matchId' : 'match_id' }, inplace = True)

In [74]: merge_5.rename(columns = {'gameCreation' : 'game_created' }, inplace = True)

In [75]: merge_5.rename(columns = {'gameDuration' : 'game_duration' }, inplace = True)

In [76]: merge_5.rename(columns = {'wardsPlaced' : 'wards_placed' }, inplace = True)

In [77]: merge_5.rename(columns = {'firstBlood' : 'first_blood' }, inplace = True)

In [78]: merge_5.rename(columns = {'firstTower' : 'first_tower' }, inplace = True)

In [79]: merge_5.rename(columns = {'firstInhibitor' : 'first_inhibitor' }, inplace = True)

In [80]: merge_5.rename(columns = {'firstDragon' : 'first_dragon' }, inplace = True)

In [81]: merge_5.rename(columns = {'firstRiftHerald' : 'first_rift_herald' }, inplace = True)

In [82]: merge_5.rename(columns = {'firstBaron' : 'first_baron' }, inplace = True)

In [83]: merge_5.rename(columns = {'teamKills' : 'team_kills' }, inplace = True)

In [84]: merge_5.rename(columns = {'towerKills' : 'tower_kills' }, inplace = True)
```

Analysis

- Through several different analysis techniques, I found several points of interest to look at. Through the use of Correlation matrix, I looked for columns that had high positive correlation coefficients.
- I also used regression lines to test hypotheses I conducted and then compared the two together.



Visualizations

Bar Charts

I created Bar charts that focus on looking at the difference between map sides data

Geographical Maps

Using our geo shape data I created several maps comparing various variables by region/server charted on a map

Scatterplots with regression lines

Confirming our hypothesis through the use of plotting our data and running a regression line I then compared the two together.

Findings

- Through our data I had discovered several details within our data:
  - Akali being the most banned champion
  - Kai'Sa and Lucian being the most picked champion
  - There being a very minimal benefit to playing on Blue Side than Red Side, indicating a balance and no clear strong benefit to being on either side

Recommendations

- After listing our findings, I then was able to give some recommendations:
  - Look to reduce some of Akali's strengths and maybe her ban numbers will decrease as players will assume she is balanced to play against.
  - Look at the meta for marksmen/ADC's as currently Kai'Sa, Lucian, and Jhin are the top 3 picks in the game meaning you are always likely to play against one of the three every game.
  - To keep on reviewing vision in the game, as one of the most fundamental and an important part of the game, it is important to keep a healthy balance to vision. With further analysis and data sets with vision denial data, we can do a deeper analysis on how vision impacts our other metrics.

Links

- Here are links to where you can find the open source data, the tableau storyboard, and my Github that has more information on this project.

Data: https://www.kaggle.com/kevindsouza2794/league-of-legends-regional-data

Geo shape data: https://www.naturalearthdata.com/downloads/110m-cultural-vectors/110m-admin-0-countries/

Tableau Storyboard: https://public.tableau.com/views/LoLSeason8RegionalData/LeagueofLegendsSeason8RegionalDataStory?:language=en-US&publish=yes&:display_count=n&:origin=viz_share_link

Github: https://github.com/20smJavi/career_foundry_lol_regional_project