Differences in MOBA Conditions:

Motivation

By:
Abhay Baliga
Hugh Ding
Seth Fleming

Background

(Game-specific terms are *italicized.*)
League of Legends (LoL) and Dota 2 are two similar video games belonging to the multiplayer online battle arena (MOBA) genre. In both cases, a single match involves two teams of five players, who spawn in their respective bases at opposing corners of a predefined square map. The primary objective of a match is to deplete the health of the main structure (so-called *nexus*) in the opposing team's base. Each player controls a unique character for the match and uses the character's unique abilities and attacks to accomplish secondary objectives that improve their odds of destroying the nexus: killing enemy players, killing non-playable opposing enemies (*minions*), neutral enemies (*jungle monsters*) that both teams can compete to kill, and destroying structures including *towers*, and *inhibitors/barracks*. Completing secondary objectives earns the team various bonuses, most prominently *gold*, which can be used to purchase items at shops which increase damage and other stats, and make the primary objective of destroying the *nexus* easier.

Motive

As people who enjoy games like this, we are curious about how similarly these two games are played. Despite having a similar underlying framework, LoL and Dota 2 have differences in playable characters, items, and other mechanics. This could help for

a) players looking to improve their win rates in either game;
b) players who are looking to get into one game from the other and are unsure of which strategies or secondary objectives to focus on. This is especially relevant because the target audiences of both games are similar and, as primarily LoL players, we are confident that Dota will soon die out.

Goal

Our goal is to identify match metrics that are shared between the games but differ in implementation, and compare how indicative these metrics are of eventual victory in each game. For example, *jungle monsters* (such as Roshan in Dota 2 and Baron Nashor in LoL) have different abilities between the two games, are located in different areas of the map, and offer different bonuses when killed, meaning the cost-benefit analysis of spending time to kill them is substantially different.

Data Sources

LoL Datasets

Short Description: There are five relevant files. The first two, stats1.csv and stats2.csv, contains player ids as well as individual metrics like kills, damage done, healing done. Since there are ten players in a single match, every set of ten records describes one match. The third, teamstats.csv, contains match ids as well as entire team metrics like towers destroyed and neutral monsters killed. Similarly, every set of two records describes one match. The fourth, participants.csv, matches match id to player id. The fifth, matches.csv, contains match id and game duration. Most of these features are of type int and some are binary.
Estimated Size: The dataset size is approximately 195 MB, with data describing about 180,000 games.
Location, Format and Access Method: The data can be found on Kaggle at https://www.kaggle.com/datasets/paololol/league-of-legends-ranked-matches. While there are several more files included, we are focused on the files listed above and are each in csv format. The data was generated via the Riot Games API, but we are downloading it directly from Kaggle. The data contains records primarily from 2018, or season 8 in-game.

Dota 2 Datasets

Short Description: There are three relevant files. The first, match.csv, contains match ids, duration, and values related to towers. The second, objectives.csv also contains match ids as well as strings that indicate the time at which important objectives were destroyed. The third, players.csv, contains match ids, player ids, and much of the data we are interested in such as gold acquired, kills, and damage done. Most of these features are of type int and some are binary.
Estimated Size: The dataset size is approximately 455 MB, with data describing about 50,000 games.
Location, Format, and Access Method: The data can be found on Kaggle at https://www.kaggle.com/datasets/devinanzelmo/dota-2-matches. While there are several more files included, we are focused on the files listed above and are each in csv format. The data was generated via the Opendota API, but we are downloading it directly from Kaggle. The data contains records from 2019.

Data Cleaning & Manipulation - LoL

Goal: combine five csv files to accumulate all the necessary information relating to **team performance metrics** over time, joining on matchid.

1. Feature space is determined by manually picking which features are comparable across games and further narrowed down by which features we'd like to specifically include in our analysis.
2. stats1.csv & stats2.csv are concatenated as they contain exactly the same features.
3. The newly created stats df has its column "id" merged by participants.csv column "id" to get column "matchid".
4. Player metrics are summed together by team to obtain team metrics, with the exception of a few columns where it is sensible to use the maximum instead, such as first blood.
5. This is merged with matches.csv column "id" to bring in game duration, used for later.
6. With the player stats aggregated, we merge these with the teamstats.csv column "matchid", which contains more relevant metrics.
7. The column "duration" is used to find team performance metrics per minute. Kill/death and kill+assist/death ratios, which are common indicators, are calculated independently of duration.
8. Values are standardized, outliers are removed, and fields are renamed for clarity, to easily compare to the Dota 2 dataset.
9. The cleaned table has two rows per match (one for each team), and one column for each objective.

File Name	Relevant Columns	Size
stats1.csv	Id, win, kills, deaths, assists, totdmgtochamp, etc.	999,999 x 56
stats2.csv	Id, win, kills, deaths, assists, totdmgtochamp, etc.	834,521 x 56
participants.csv	Id, matchid	1,834,520 x 8
matches.csv	Id, duration	183,452 x 8
teamstats.csv	matchid, firstblood, firsttower, towerkills, dragonkills, etc.	366,904 x 13

df_pm	match id, team, kills, total damage, turret count, dragon kills, etc.	356,954 x 53

Data Cleaning & Manipulation - Dota 2

Goal: combine three csv files to accumulate all the necessary information relating to **team performance metrics** over time, joining on match_id. Note that when objectives have a direct equivalent in League (e.g. League turrets and Dota towers are equivalent), we use League terminology for consistency.

1. matches.csv stats from *Radiant* and *Dire* teams are reshaped so that team-specific objectives are in the same column, e.g. tower_status_dire and tower_status_radiant become turret_count for their respective teams.
2. The chat log in the "subtype" column of objectives.csv, which indicates what team accomplished which goals in each match, is reshaped and filtered so that necessary objectives each occupy a column per match and team.
3. For players.csv, feature space is determined by manually picking which features are comparable across games and further narrowed down by which features we'd like to specifically include in our analysis. These are summed by match and team to generate team statistics.
4. Data is merged by team and match id to aggregate all the team statistics per match.
5. The column "duration" is used to find team performance metrics per minute. Kill/death and kill+assist/death ratios, which are common indicators, are calculated independently of duration.
6. Values are standardized, outliers are removed, and fields are renamed for clarity, to easily compare to the Dota 2 dataset.
7. The cleaned table has two rows per match (one for each team), and one column for each objective.

File Name	Relevant Columns	Size
match.csv	match_id, team, duration, tower_status_dire, tower_status_radiant, barracks_status_radiant, barracks_status_dire, radiant_win, etc	50,000 x 13
objectives.csv	match_id, team, subtype	1,173,39 6 x 9
players.csv	match_id, team, gold, kills, xp_per_min, kills, deaths, assists, etc	50,000 x 73

dota	match_id, team, kills, total damage, turret count, roshan kills, etc	356,954 x 53

Data distributions

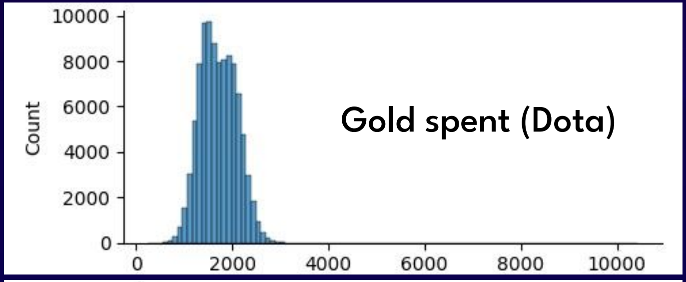
All of the variables appear to be reasonably distributed (roughly normal distribution). There are two major exceptions to this:

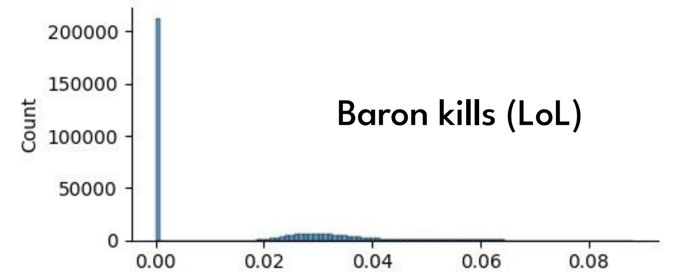
- Many of the variables are right-skewed, with a great number of values at zero, which makes sense if
  a) A team is too behind to contest an objective, or
  b) A game ends early due to one team surrendering, before objectives that only appear later in the game (such as Baron in LoL) can be contested.
- Some variables have multiple peaks. For example, the inhibitor distributions in each game have two peaks, which makes sense because the dota in-game map has three "lanes" by which to enter the enemy base. Destroying inhibitors in a lane creates stronger allied non-player units ("minions") in that lane. In closer games, you will want to destroy inhibitors in multiple lanes to make destroying the enemy base easier. Otherwise, you may be able to win only through a single lane.

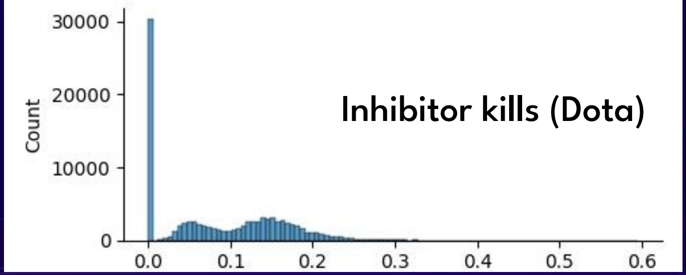
Combining the datasets

Combining the cleaned datasets of the two games is straightforward, since they only differ in a few unique objectives:

1. Add a column to each dataset indicating the game.
2. Duplicate the LoL dataset so that its rows are comparable to the Dota dataset's.
3. Append the LoL dataset to the Dota dataset.



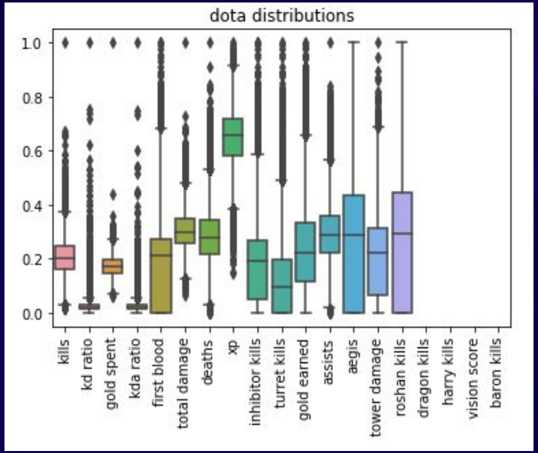




Analysis - Feature Distributions

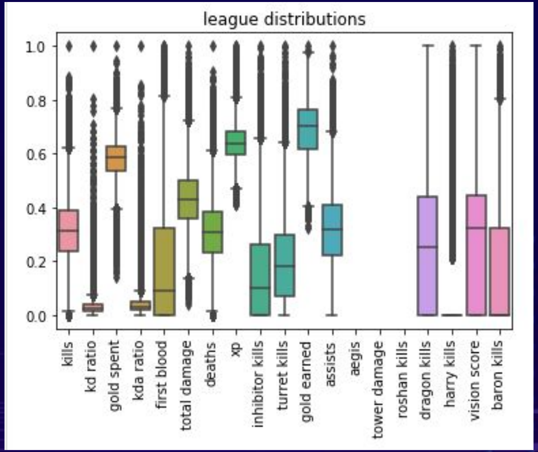
Setup

From these visualizations, we'd like to see what normally happens in each game, compared to what happens at the extremes. From there, we can compare the similarities and differences of typical and extreme matches between games. This gives us insight into how common or available a variable is across games for each match. For example, looking at the LoL "kills" variable tells that a large majority of games are relatively even, while few games are one-sided. On the other hand, Dota 2's "kills" variable tells that kills are skewed up, toward a higher value. This means that more games than in LoL are one-sided, typically ending quicker, leading to fewer objectives taken in a game.

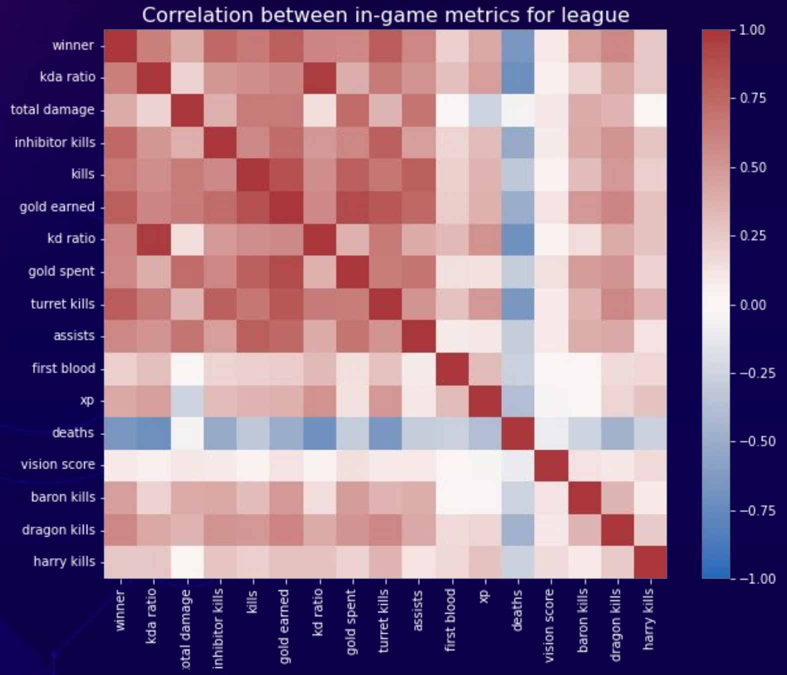


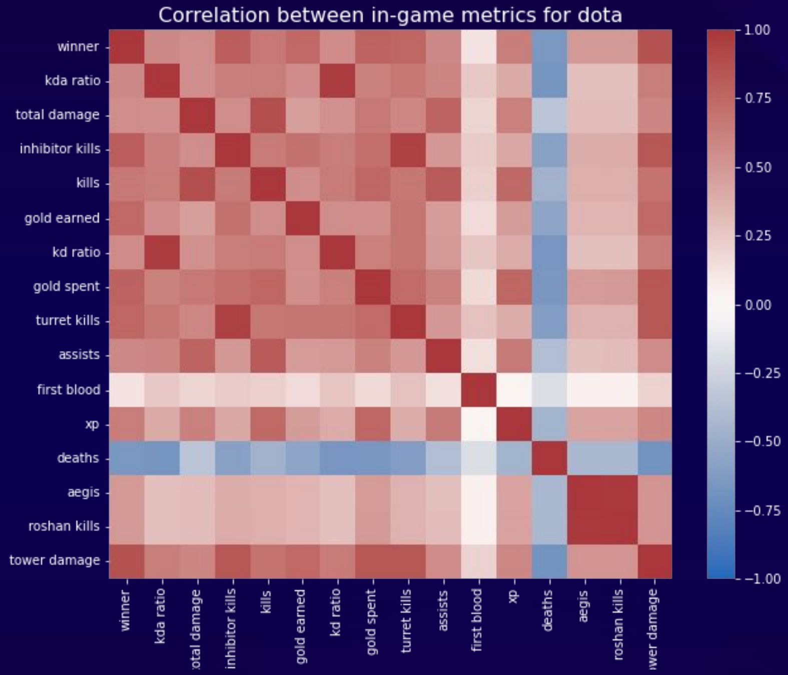
Differences

Whereas there are many common metrics tracked across games, there are a few differences in objectives that should be highlighted. LoL has dragons, baron nashor (baron), and the rift herald (harry) - all of which are neutral monsters and major objectives. Dota 2 has roshan and the aegis, an item that spawns after the first time roshan is slayed. Each of these, in addition to inhibitor kills, turret kills, first bloods, and assists, contribute varying amounts of gold and experience xp. Because these are distinct objectives between games, it would be expected that one game's unique objectives would happen to be more common and perhaps more advantageous than the other. However, it's interesting that this is not the case - they all are taken relatively infrequently with few games taking them in relatively higher amounts. On the other hand, the damage done, kills, turrets killed, and inhibitors killed vary significantly, again indicating that Dota 2 games are more one-sided.



Analysis - Combined Correlations





The in-game metrics that best correlate to winning are actually quite similar between games. Kills, assists, and gold matter a bit more in League of Legends, while XP matters more in dota. Their unique objectives (Roshan, Baron, and so on) don't seem as important.

In both games, first blood is not important, and deaths negatively correlate with achieving secondary objectives.

Basic Prediction Models

Dota 2

turret kills	28.23
gold earned	21.11
gold spent	20.00
xp	10.87
kills	7.41
kda ratio	2.84
roshan kills	1.46
first blood	-2.51
deaths	-16.13

LoL

gold earned	44.94
turret kills	14.97
xp	8.98
inhibitor kills	8.65
kills	7.34
kda ratio	5.59
kd ratio	5.13
assists	3.81
total damage	2.86
baron kills	1.15
dragon kills	0.58
vision score	-0.33
harry kills	-1.02
first blood	-1.12
deaths	-22.25
gold spent	-34.09

Combined

gold earned	28.58
turret kills	18.17
xp	13.99
kills	13.17
inhibitor kills	11.77
total damage	6.68
assists	4.89
kd ratio	3.5
first blood	-1.65
kda ratio	-3.07
deaths	-26.52
gold spent	-37.93

For comparison with the metrics most closely correlated with winning, we also explored a logistic regression model. The model shows us that **in all 3 datasets, Gold Earned and Turret Kills were the top two indicators of a win.** Turret kills are known to be the predominant indicator of the game's state so this is not as surprising, but gold is earned from completing many of the secondary objectives. There is no specific secondary objective that outweighs simply earning gold.

For the commonly small correlations, it makes sense that deaths do not lead to victory. While there are differences in the penalties between games, if a player is killed, enemies gain time to complete objectives unhindered and earn more gold. First blood (the first team to kill a player) may be the more interesting result, but it is a small early bonus in both games. Neither metric is normally considered an indication of winning.

- We ran 3 logistic regression models, one for each game and one with the combined data.
- The data is presented in log-odds, so negative numbers indicate small effects on the prediction, and large numbers indicate a larger effect on the prediction.
- The values are "with all else equal" so this is only exploratory.

Limitations

Given the nature of our datasets, our analysis had some limitations.

1. Neither game is unchanging; for example, League of Legends sees regular patches roughly every two weeks, which are generally balance adjustments to individual champions.
    - In addition, the two datasets were from different periods. The Dota set was from November of 2015. On the other hand, the League of Legends data was from seasons 3 through 8; we filtered this to just season 8 (largely across 2018), since that was where most of the data was from (roughly 95%).
    - Our assumption was that cross-game differences would be much larger than intra-seasonal differences per game.
2. In part due to the above, some variables were simply too complex to be considered, such as character choice on whether dragons were killed (e.g. in League, the character Shyvana gets additional bonuses per dragon killed).
3. We also assumed that both red and blue teams were balanced in terms of map position. This isn't necessarily true - while they're balanced in terms of win rates by design, one side has slightly easier access to certain objectives because the map isn't symmetrical.
4. Not all of the variables we were interested in were available. For example, Dota also has options to grant map vision, similar to LoL vision score, but metrics to measure map vision simply weren't available in the Dota dataset.
5. Correlation does not imply causation. While we, for example, saw that turret kills and gold spent were closely correlated to winning the match, this isn't conclusive proof that killing turrets and spending gold are strong causes of victory (though intuitively they should be).
    - For example, certain turrets must be killed to reach an opposing nexus, and killing these turrets will also grant gold, so obtaining victory actually causes turret kill count and gold earned to increase.

Conclusion and next steps

In conclusion, the secondary objectives to focus on between games were generally very similar:

- Gold earned is among the secondary objectives most correlated with winning.
- Some objectives, like first blood, are not worth prioritizing.
- One common myth in LoL is that it's important to focus on killing powerful neutral jungle monsters like baron and dragon for their buffs. However, this analysis seems to indicate that taking turrets - an objective shared in both LoL and Dota - is a much better predictor of victory.

Some ideas for next steps:

- The Kaggle datasets we used were pulled from the LoL official API and OpenDota API. The API interfaces are granular enough to be able to nearly provide a full description of each match. This could benefit more in-depth analysis in several ways:
    - It would provide a much larger sample size to work with.
    - The increased granularity would allow us to conduct more creative analysis, such as breaking down exactly what the different sources of gold earned consist of.
    - We would be able to filter the data to and conduct analysis on similar time ranges and more comparable seasons.
- We could conduct an experiment where we control for some variables, such as ensuring the same characters are picked every match. This would allow us to better analyze which factors actually cause a team to be more likely to win.

Statement of Work

Collaboration

Our team met about once a week to talk over big ideas and the direction of our project. We split up work to do on our own and came to meetings with our progress and findings.

Something that could have improved our workflow was to set goals and deadlines throughout the timeline of this course to better equip ourselves with more time to revise our dataset and see how well the features translated into our analyses.

ALL: Proposal and data acquisition, report editing, communication and brainstorming through Discord

Seth Fleming: Initial LoL data cleaning, primary report writing

Hugh Ding: Dota 2 data cleaning, Dota 2 data analysis, merged dataset analysis

Abhay Baliga: Final LoL data cleaning, LoL data analysis, secondary report writing