**Dota 2 match analysis**

Data from <https://www.kaggle.com/datasets/labinotvila/dota-2-steam-api-fetched-dataset>

**Content**

The data that we have from the link has 50000 matches. We will limit this research with 1000. As it is more than enough to do our analysis.

The data is split into two .csv files: radiant.csv and dire.csv. Both of them have information that we don't need. So out of these two files we will make one with the most valuable information for us.

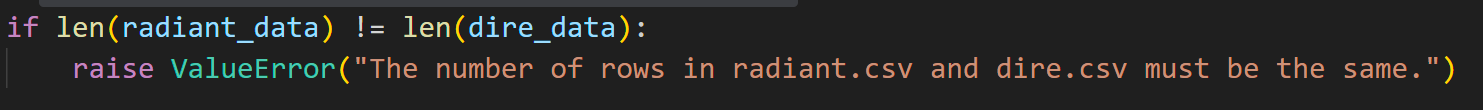
The whole coda can be found in make\_result\_file.py

1. Load the data .

A screen shot of a computer program

Description automatically generated

1. Ensure both files have the same number of rows.



1. Main algorithm.

A screen shot of a computer program

Description automatically generated

1. Saving the results to a separate file.

A screen shot of text

Description automatically generated

**The result file will include the following key statistics:**

* **Index:** the index of a match in our data.
* **Radiant GPM:** gold per minute value for radiant team.
* **Dire GPM:** gold per minute value for dire team.
* **Radiant XPM:** experience per minute value for radiant team.
* **Dire XPM:** experience per minute value for dire team.
* **Leaver Status:** determines whether a player from any team has left the match in the middle of it.
* **Result:** which team won.
* **Radiant Win:** 1 if radiant won, 0 if dire. (for easier calculations)

**An example of how the result file should look like:**

**A screen shot of a computer

Description automatically generated**

**“How much does playing with a leaver affect the outcome?”**

Dota 2 is a complex game. There are 5 players in a team and each of them has a role: Hard support, support, offlaner, midlaner and carry. We need to analyze the data to see whether there is a correlation between losing one player and winning the map.

(code can be found in win\_percentages\_with\_leaver.py)

A screenshot of a computer program

Description automatically generated

**Results:**

**Radiant team win percentage with leaver: 38.69%**

**Radiant team win percentage without leaver: 56.97%**

**Dire team win percentage with leaver: 31.45%**

**Dire team win percentage without leaver: 54.93%**

Now let’s look at win percentages with leavers in both teams:

(full code in “both team leavers.py”)

A screen shot of a computer program

Description automatically generated

**Results:**

**Radiant team win percentage with leaver in both teams: 40.91%**

**Dire team win percentage with leaver in both teams: 59.09%**

**Analysis of Winning Percentages with and Without Leavers.**

**Expected Probability Calculation:**

To analyze the impact of having leavers in both teams, we will calculate the expected win probability under the assumption of independence (i.e., the effect of leavers in both teams is independent). We will compare this expected probability with the observed probabilities.

**For Radiant:**

**Expected win probability=1−(1−0.3869)⋅(1−0.5697)**

**Expected win probability=1−0.2630≈0.7370**

**For Dire:**

**Expected win probability=1−(1−0.3145)⋅(1−0.5493)**

**Expected win probability=1−0.3084≈0.6916**

**Comparison of Observed and Expected Percentages:**

**For Radiant:**

* Observed Win Percentage with Leavers in Both Teams: 40.91%
* Expected Win Percentage under Independence: 73.70%
* The observed percentage is significantly lower than the expected percentage, indicating a greater-than-additive effect when both teams have leavers. This suggests a nonlinear interaction, where the presence of leavers in both teams results in a more severe impact than the sum of individual effects.

**For Dire:**

* Observed Win Percentage with Leavers in Both Teams: 59.09%
* Expected Win Percentage under Independence: 69.16%
* The observed percentage is lower than the expected percentage, but the difference is less pronounced compared to Radiant. This implies a significant effect, though not as dramatic as for Radiant.

**Interpretation**

The analysis reveals that the impact of having leavers in both teams is more complex than a simple additive effect. The results suggest nonlinear interactions, where the combined effect of having leavers in both teams is more severe and less predictable than the sum of the individual effects. This indicates that other factors and interactions are likely influencing the outcomes.

**Team with higher GPM is more likely to win?**

There is an assumption in Dota 2 that high GPM is a sure way to win. Perhaps it comes from the assumption that more gold buys you more artifacts, therefore making your hero stronger. However, Dota 2 is very skill-based. So, it may differ from one player to another.

* Null Hypothesis (H0): Higher GPM does not significantly affect the winrate of a team.
* Alternative hypothesis (H1): Higher GPM affects the winrate of a team

A screen shot of a computer program

Description automatically generated

t-statistic: 8.71

p-value: 0.0000

Reject the null hypothesis: Higher GPM significantly affects the win rate of a team.

**Reject the Null Hypothesis (H₀):** Based on the statistical analysis, we can reject the null hypothesis. This means there is strong evidence to support the alternative hypothesis that **higher GPM significantly affects the win rate of a team**.

A white rectangular object with blue text

Description automatically generated

A graph of a distribution of different colored lines

Description automatically generated with medium confidence

A graph of a box plot

Description automatically generated

But what about XPM?

**Team with higher XPM is more likely to win?**

* Null Hypothesis (H0): Higher XPM does not significantly affect the winrate of a team.
* Alternative hypothesis (H1): Higher XPM affects the winrate of a team

A screen shot of a computer program

Description automatically generated

t-statistic: 4.91

p-value: 0.0000

Reject the null hypothesis: Higher XPM significantly affects the win rate of a team.

**Interpretation**

1. **t-Statistic:**

* The t-statistic of 4.91 indicates a substantial difference in XPM between winning and losing matches. This value suggests that the difference in XPM is significant.

1. **p-Value:**

* The p-value of 0.0000 is very low and well below the commonly used significance level of 0.05. This indicates that the observed difference in XPM between winning and losing matches is statistically significant.

**Reject the Null Hypothesis (H₀):** The results suggest that we can reject the null hypothesis. This means that there is strong evidence to support the alternative hypothesis that **higher XPM significantly affects the win rate of a team**.

A white rectangular frame with blue dots

Description automatically generated

A graph of a distribution of different sizes

Description automatically generated with medium confidence

A graph of a box plot

Description automatically generated

**Summary**

Based on the hypothesis testing and statistical analysis:

* **Higher GPM:** The significant positive correlation and the t-test results confirm that higher GPM is strongly associated with winning. Teams with higher GPM have a better chance of winning matches.
* **Higher XPM:** Similarly, the significant positive correlation and t-test results confirm that higher XPM significantly influences the likelihood of winning. Teams with higher XPM are more likely to win.

**Overall Conclusion:**

The analysis supports the hypothesis that higher GPM and XPM positively impact a team's win rate. Teams focusing on optimizing their GPM and XPM could improve their chances of winning matches. While GPM and XPM are critical factors, it's important to consider other game aspects like team strategy, hero selection, and coordination for a holistic approach to improving performance.