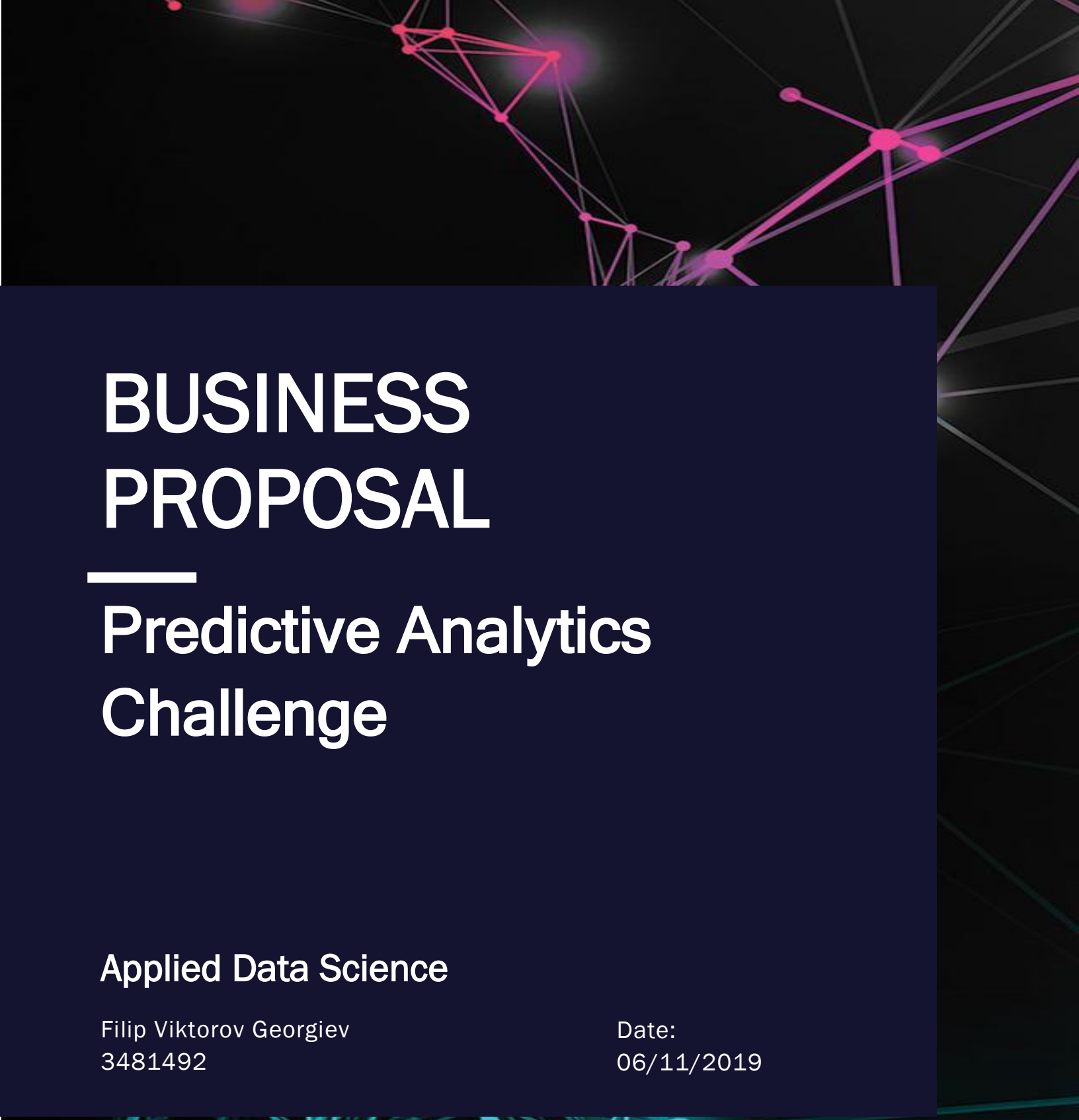

BUSINESS PROPOSAL

Predictive Analytics Challenge

Applied Data Science

Filip Viktorov Georgiev
3481492

Date:
06/11/2019

Contents

I. Executive Summary

This proposal relates to the “Predictive analytics” challenge in the applied data science course. Its purpose is to convince the reader that there is a reason to look deeper into the possibility of predicting UFC mixed martial arts matches by using machine learning techniques.

By using a dataset found on Kaggle[1] containing data from UFC matches ranging from 1993 to 2019 this project would be aimed to find a way to figure out who the winner of future matches would be.

II. Problem Overview

There is a lot to gain in this domain as a good prediction implementation is yet to surface and no one has yet found a way to actually know the outcome of a “fight”. In the event of this challenge being a success, the benefits are immense and anybody that would invest in this idea would be much richer than they previously were.

Furthermore, it would be interesting to see if there exist such specific features of a fighter that would give them a bigger chance to win than their counterpart.

III. Proposed solution

After employing exploratory data analysis few discoveries were made:

1) *Main Correlations*

By using a heatmap (Figure 1 Correlation), which matches different features and rates their correlation from 0 to 1, it can be seen that few valuable correlations actually exist (e.g. amount of wins and total rounds fought, together with the chosen winner). Although a correlation between reach, height and weight can be seen this is purely logical and sadly not useful. A valuable insight that can be exploited would be the scaling of amounts of wins with the total amount of rounds spent in the ring. This is better visualized as a scatter plot (Figure 2 Experience = Success?) The color of the points is the age of the fighter, which show that there is no clear correlation between age and win rate. On the other hand, the practical experience of the fighter does show a steady increase in wins.

2) *Red corner vs Blue corner*

A lesser-known fact is that usually the champion or “fan-favorite” is positioned in the red corner of the ring, while the underdog is positioned in the blue corner. Champions usually have a win-streak or are the better fighter. This can clearly be seen in this visualization

[1] https://www.kaggle.com/rajeevw/ufcdata

(Figure 4 Red vs Blue Corner). What is more this bar pot shows that the third round is usually the culmination of most fights.

3) Stance

Before the implementation of the heat map it was explored if the stance of the fighter makes a difference. (Figure 3 Stance/Win) Even though a clear difference can be seen, the win rate per stance also correlates with how widely these stances are used, which means that no clear correlation between having a specific stance and wining exists as the more widely used stances have more wins – a logical conclusion.

IV. Going Forward

The analysis of the data shows that there is not a lot to go for currently. Nevertheless, by using machine learning algorithms like Decision Tree and the Random Forrest classifier it would be possible to make a prediction on whether a fight between two contenders would be a win for the red corner, blue corner or a draw, based on the respective fighters' win rate and experience. The goal of this project, should it be continued would be to make such a prediction as accurate as possible

V. Figures

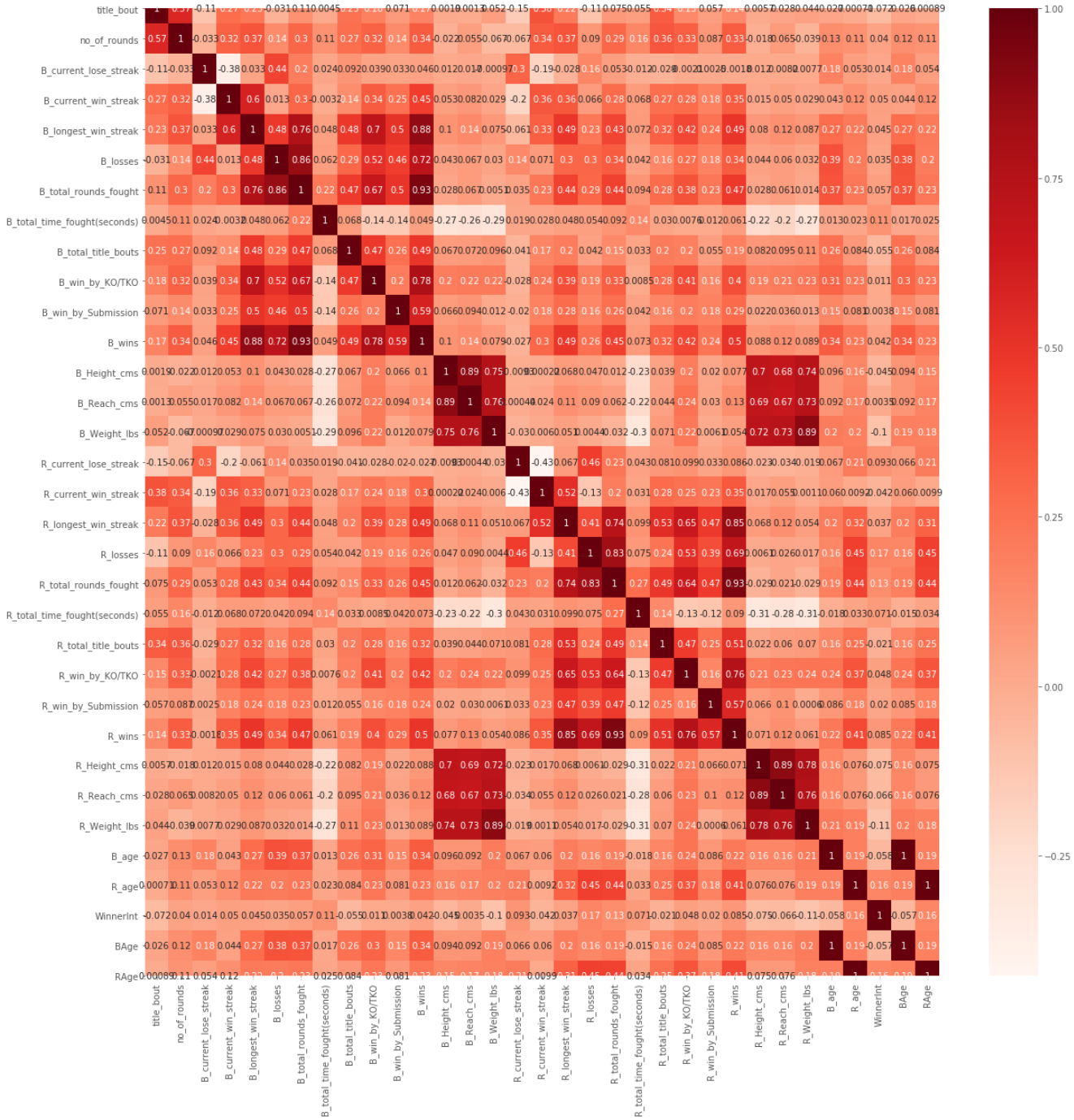

Figure 1 Correlation

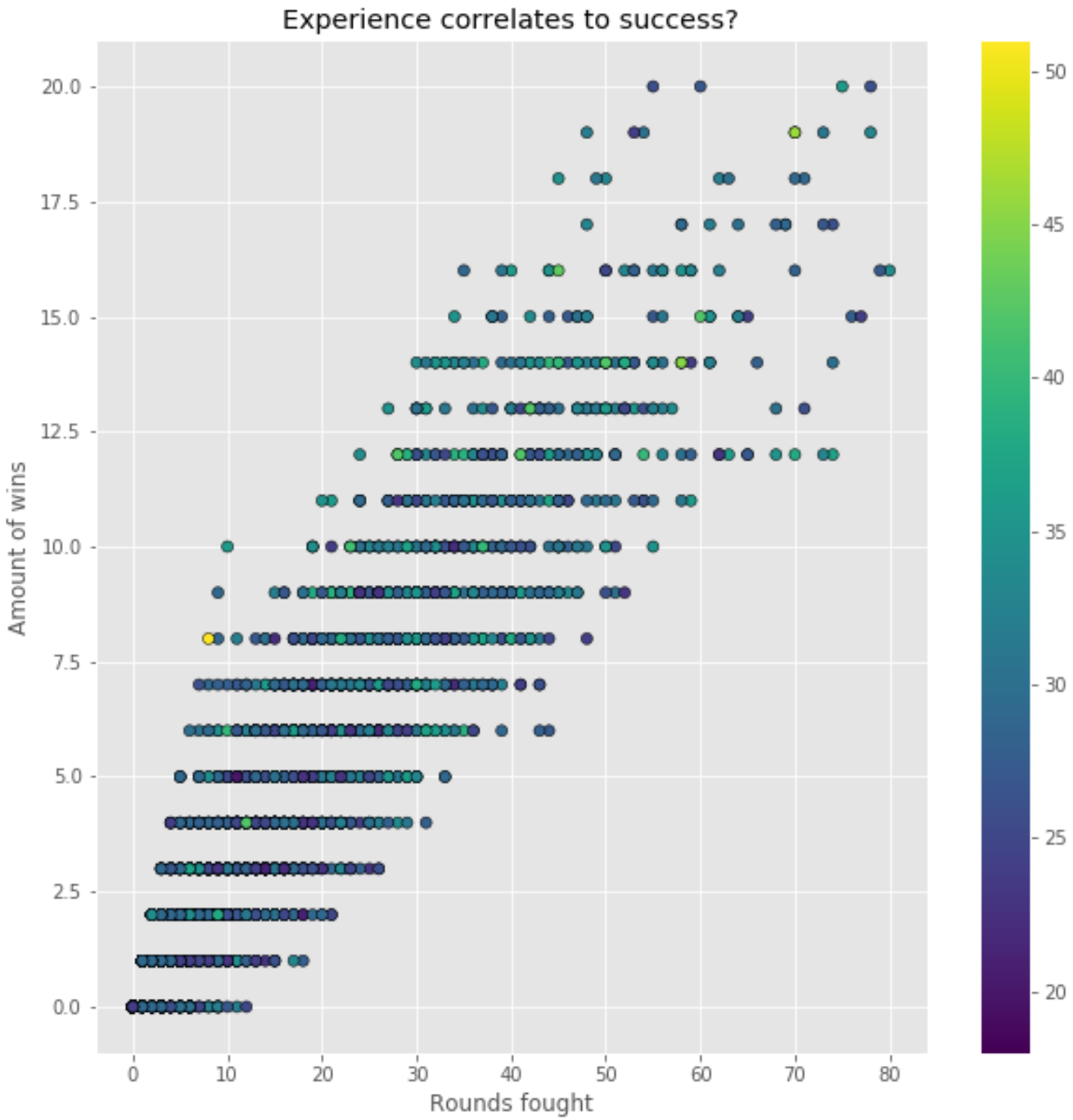

Figure 2 Experience = Success?

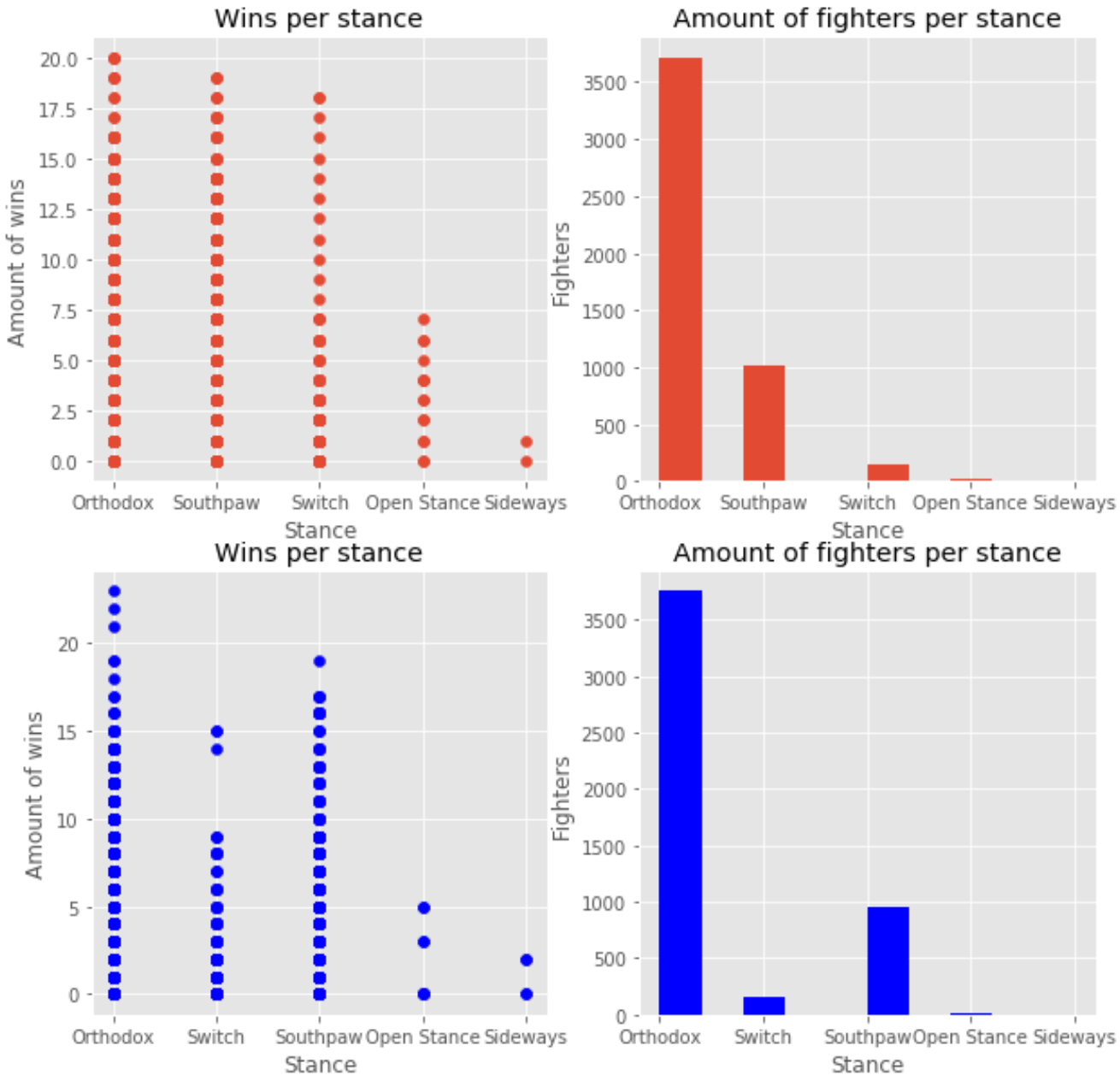

Figure 3 Stance/Win

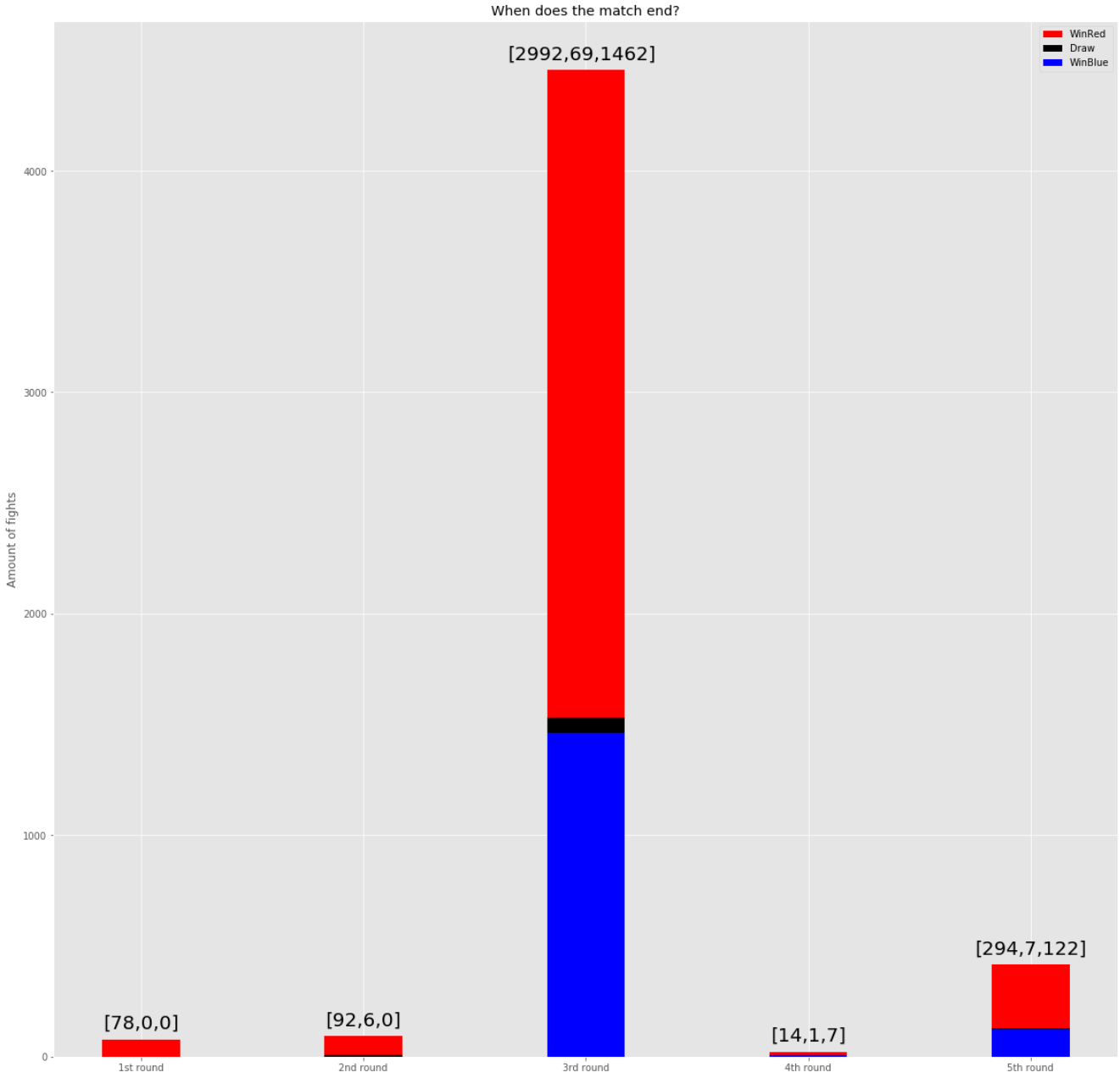

Figure 4 Red vs Blue Corner