

# NYC Data Science Academy Capstone Project:

Data Analysis of the UFC

# Table of Contents:

Introduction

Objective

Dataset Information and link

Breakdown of Red Side vs Blue Side

T-Test

Age, Height and Weight distributions between sides

Stance win rates

Title defenses by side and winningest champions

PPV buys overview

Betting odds breakdown



- The Ultimate Fighting Championship is the premier Mixed Martial Arts (MMA) promotion in world.
- Fighters from all backgrounds and nations fight under the unified mixed martial arts ruleset.
- Fighters bring a variety of martial arts and backgrounds into the octagon.
- One of the fastest growing sports in the world. The UFC is broadcasted in 156 countries and in 29 languages.

# Rules of the Octagon

- 3 Judges score the fight
- Fight duration: 3 five minute rounds
- Championship fight duration: 5 five minute rounds.
- Round winner gets 10 points
- Round loser gets 9 or less points
- Scoring based on effective striking, grappling, aggression and octagon control
- Fight can end by TKO/KO, Submission, Doctor Stoppage or No-Contest

# Additional Information

- 8 Mens divisions
- 4 Womens divisions
- 1 champion per division

-Matches are made by weight class. (Fighters can only fight other fighters in the same weight class, however they can move up or down weight classes by gaining or losing weight)

-Ranked fighters most often fight other ranked fighters, this is not always the case.

# Weight Divisions:

Men:

125lbs Flyweight

135lbs Bantamweight

145lbs Featherweight

155lbs Lightweight

170lbs Welterweight

185lbs Middleweight

205lbs Light Heavyweight

265lbs Heavyweight

Women:

115lbs Strawweight

125lbs Flyweight

135lbs Bantamweight

\*145lbs Featherweight

# Objective:

-Understand a multitude of variables and attempt to gain valuable insights of the UFC and its fighters.

-Project focuses on:

- Red side vs Blue side Probability
- Why one side is favored over the other
- Looking at red vs blue fighter statistics.

-UFC has had little data and statistical analysis done when compared to other sports.

\*A common saying is that “MMA math doesn’t work”. Due to the high volatility and variability of fights, predictions are often wrong. Prediction models in specific are too inconcise, this is more evident in MMA than other sports.

# Dataset:

<https://www.kaggle.com/rajeevw/ufcdata/data?select=data.csv>

<https://www.kaggle.com/mdabbert/ufc-fights-2010-2020-with-betting-odds>

- 6 CSV Files

- UFC Historical data from 1993 to 2021

- Contains a wide variety of data including weight class, age, height, reach, total strikes landed, takedown defense, clinch strike percentages...

- UFC betting dataset

- Little cleaning required



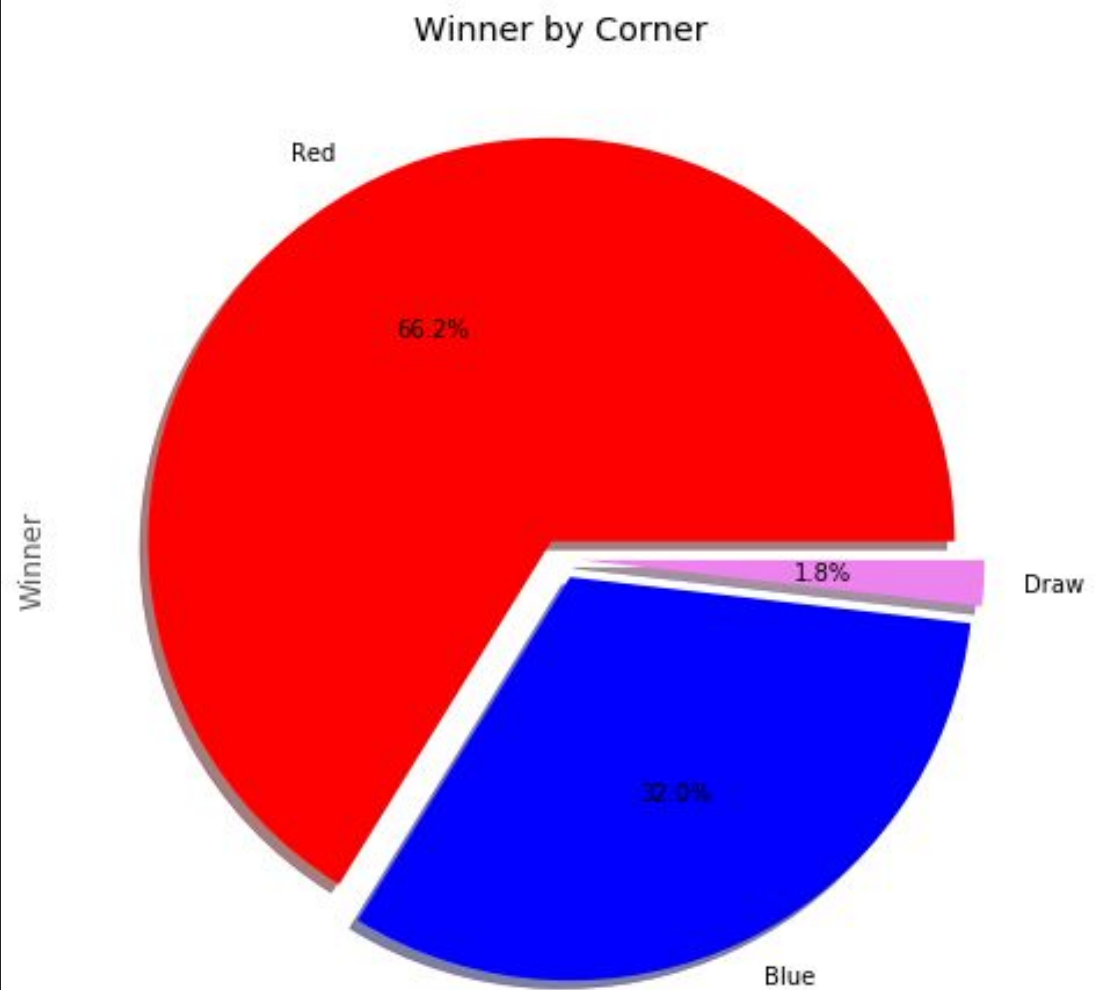
# Red Side vs Blue Side

-There are two sides in the octagon, Red and Blue. Does side play a role in a fighter's chance to win?

-Sides are indicated by glove color, with the purpose to differentiate the two fighters.

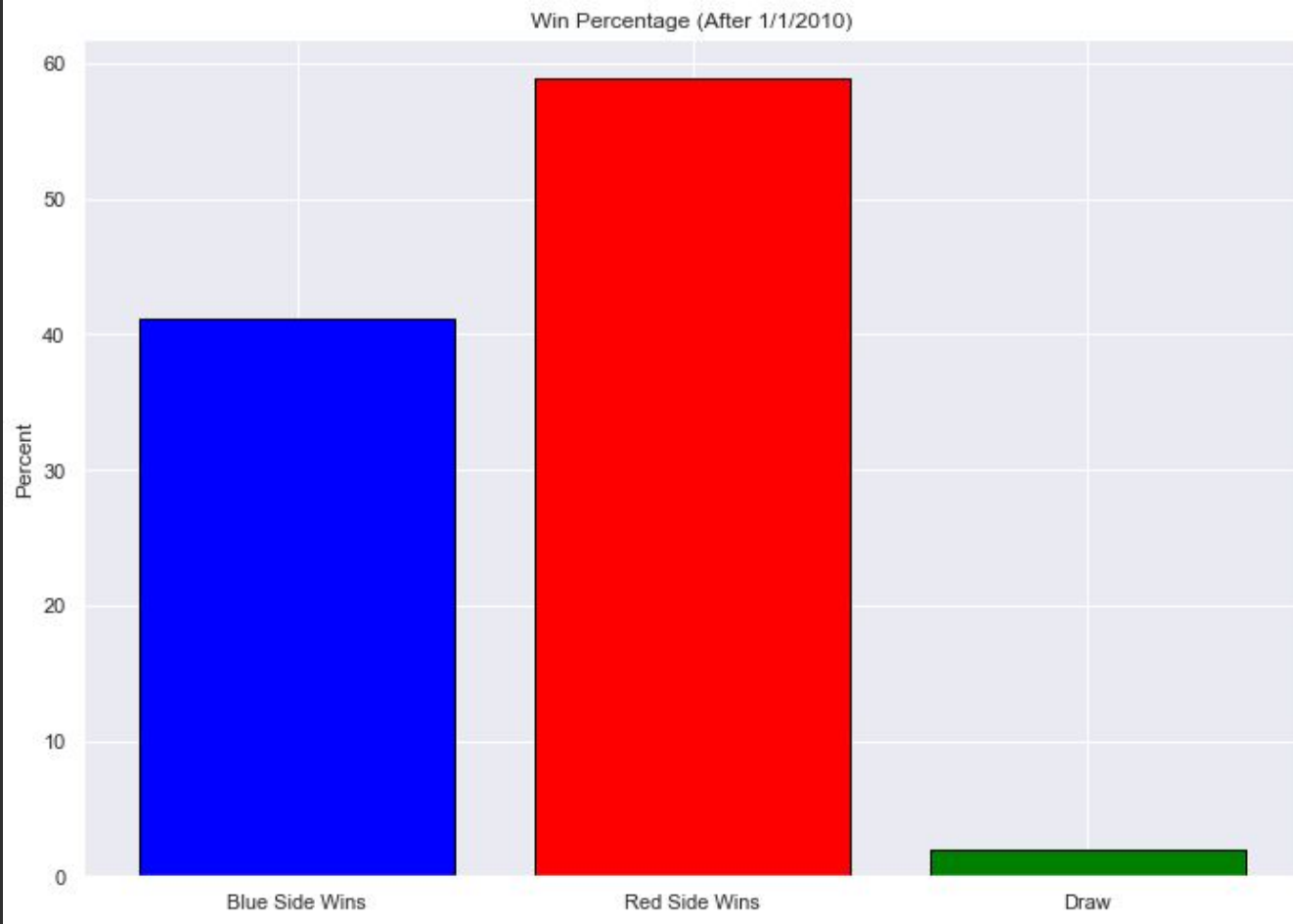


- Red side has a significantly higher win rate than Blue side.
- This is using complete historic data from 1993 to 2021.
- What can be causing the disparity of win rates?



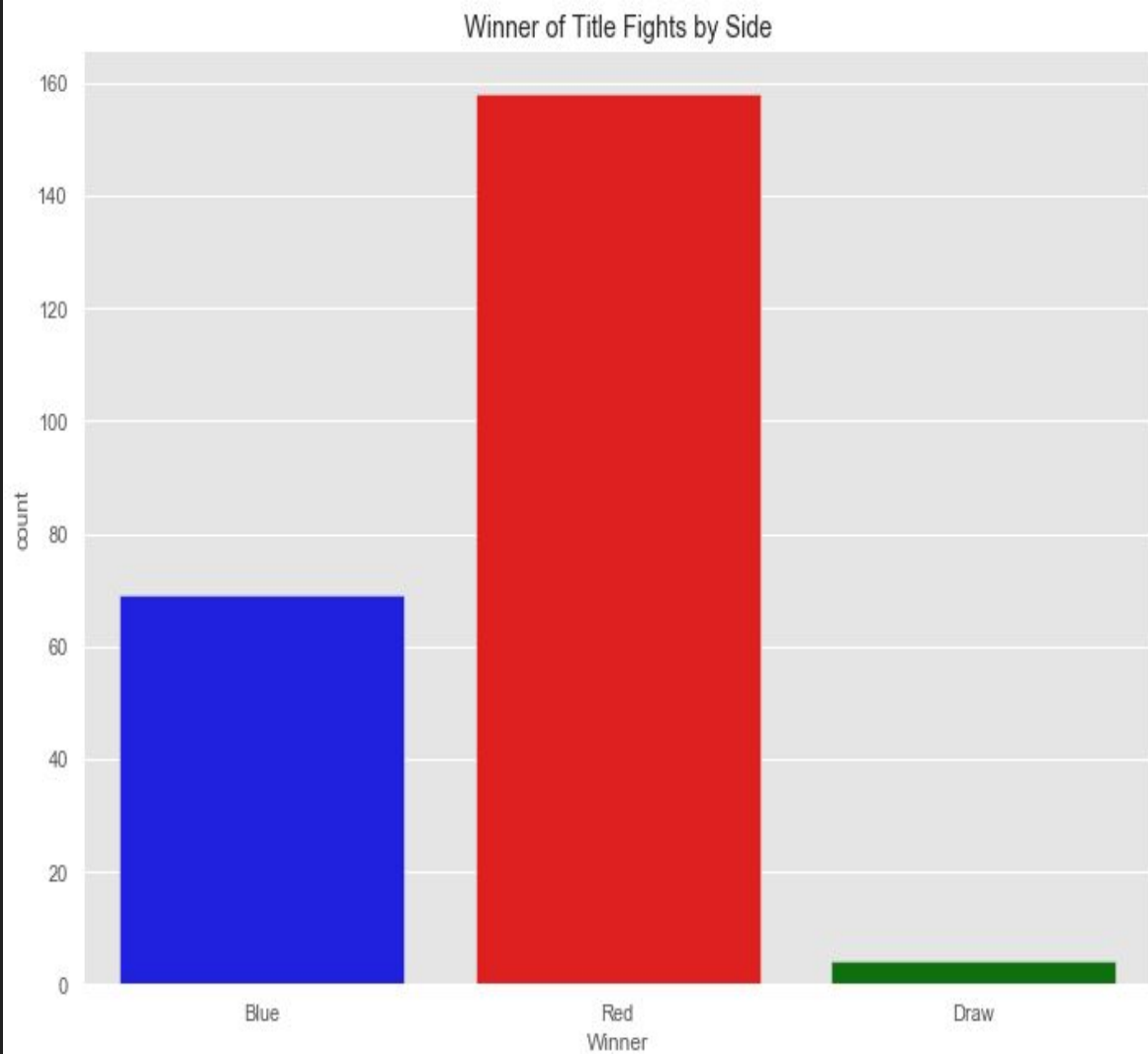
-After using recent data from 1/1/10 onward, we can see that the win percentages are closer.

-Yet, Red Side still holds a sizeable advantage over Blue side.

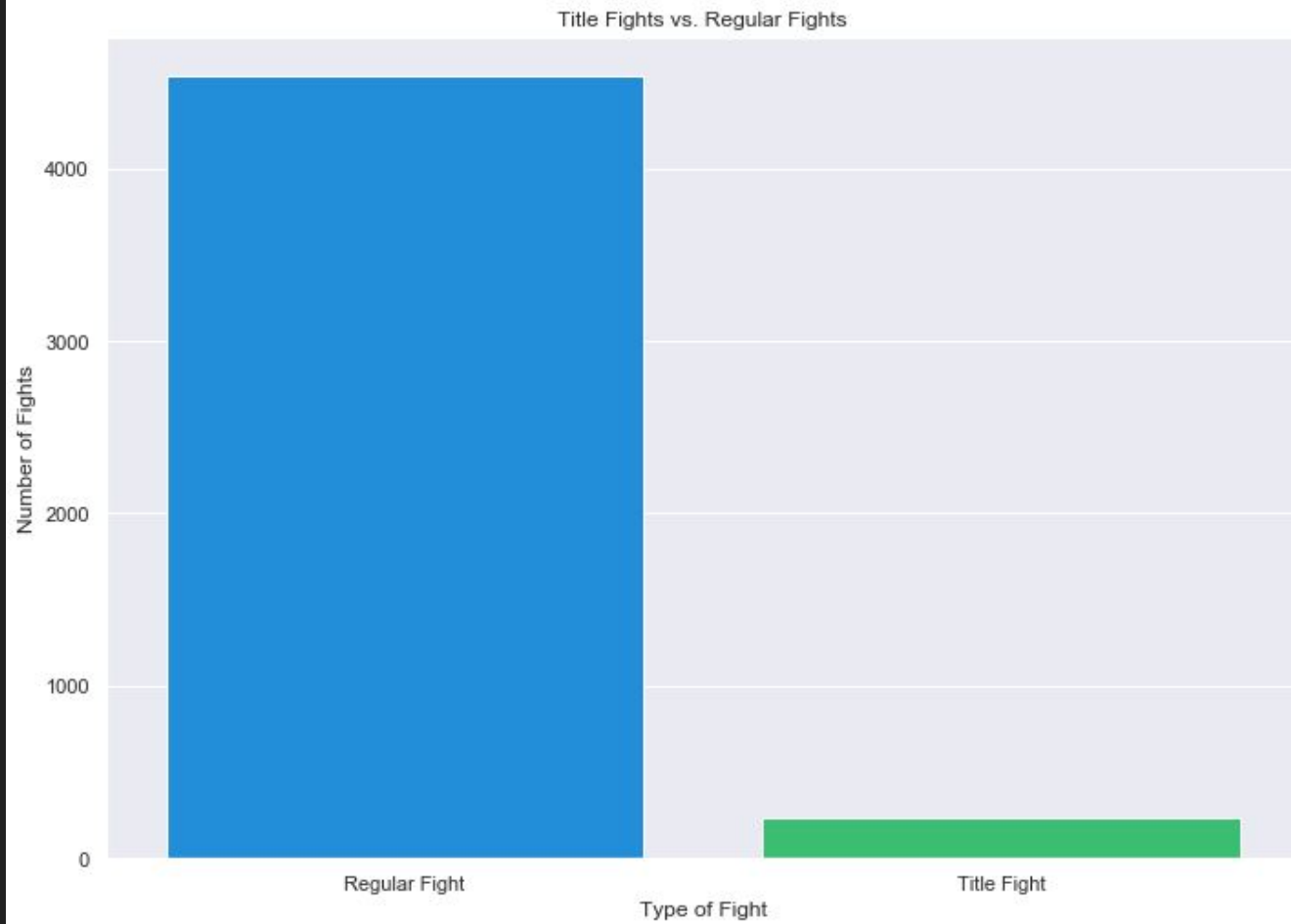


-Red side has 158 title fight winners, compared to Blue sides' 69. There is not only large discrepancy in win percentages in regular fights, but also in title fights as well.

-Red wins 68.3% of title fights, while Blue wins 29.8% of title fights



-Since 2010, there has been 4537 regular fights, and 231 title fights.



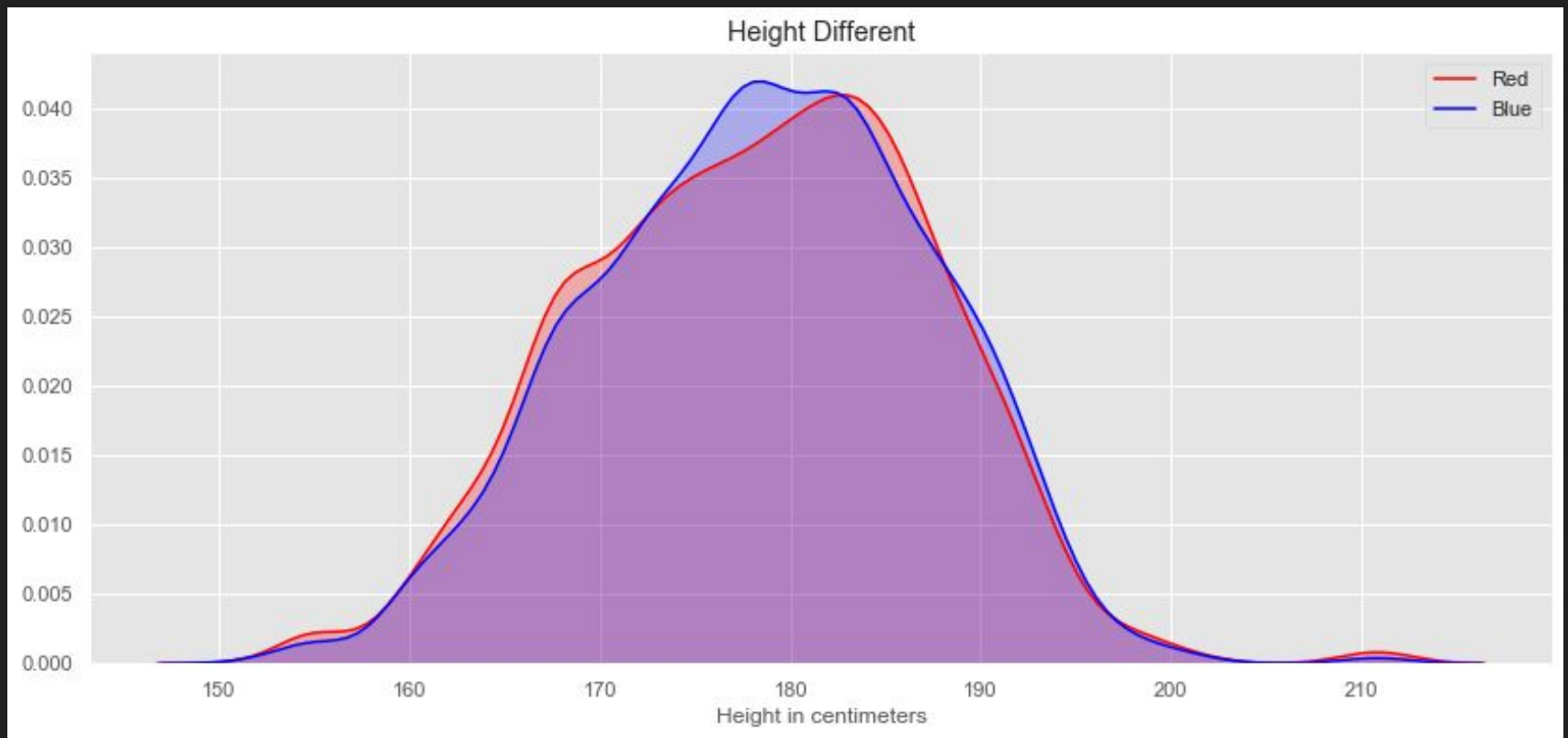
# Statistical Test Analysis

```
LeveneResult(statistic=49.99625769769318, pvalue=1.8667498630816718e-12)
```

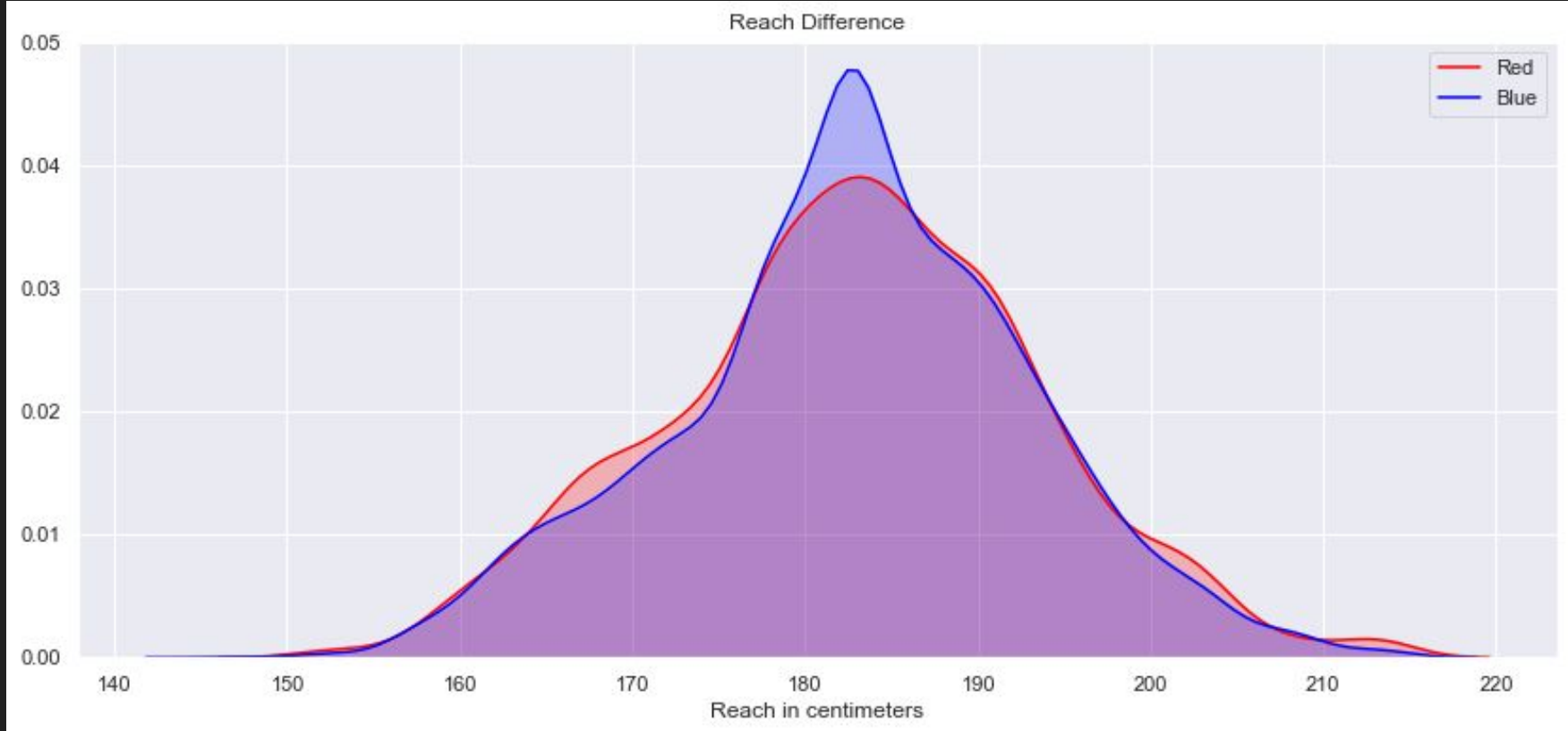
```
Ttest_indResult(statistic=-3.4947889004870225, pvalue=0.00048059111360556657)
```

The lower the p-value, the greater the statistical significance of the observed difference.

-With Levene test and T-test, comparing winners on red side and blue side of title fights. We see a significantly low pvalue, meaning greater statistical variation between the two sides. 0

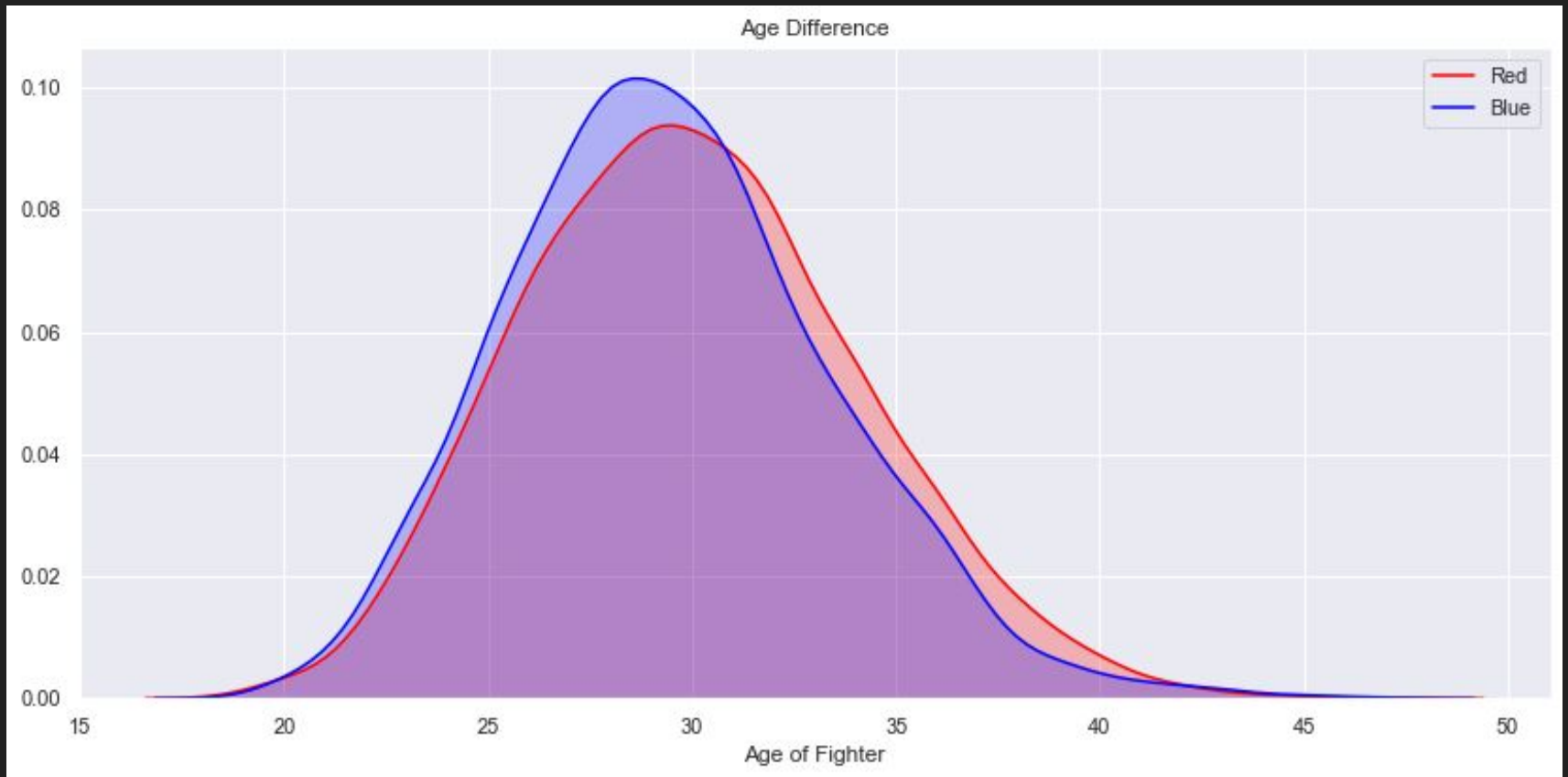


Comparing height differences between the two sides



Comparing reach differences between the two sides





Comparing age difference on the two sides

Blue fighters had a longer reach while red fighters are slightly older, height is similar. With minimal differences of height, age and reach contributing to the variation of side win rates, there has to be another variable in play that gives Red Side the advantage.

# Analysis on Stances:

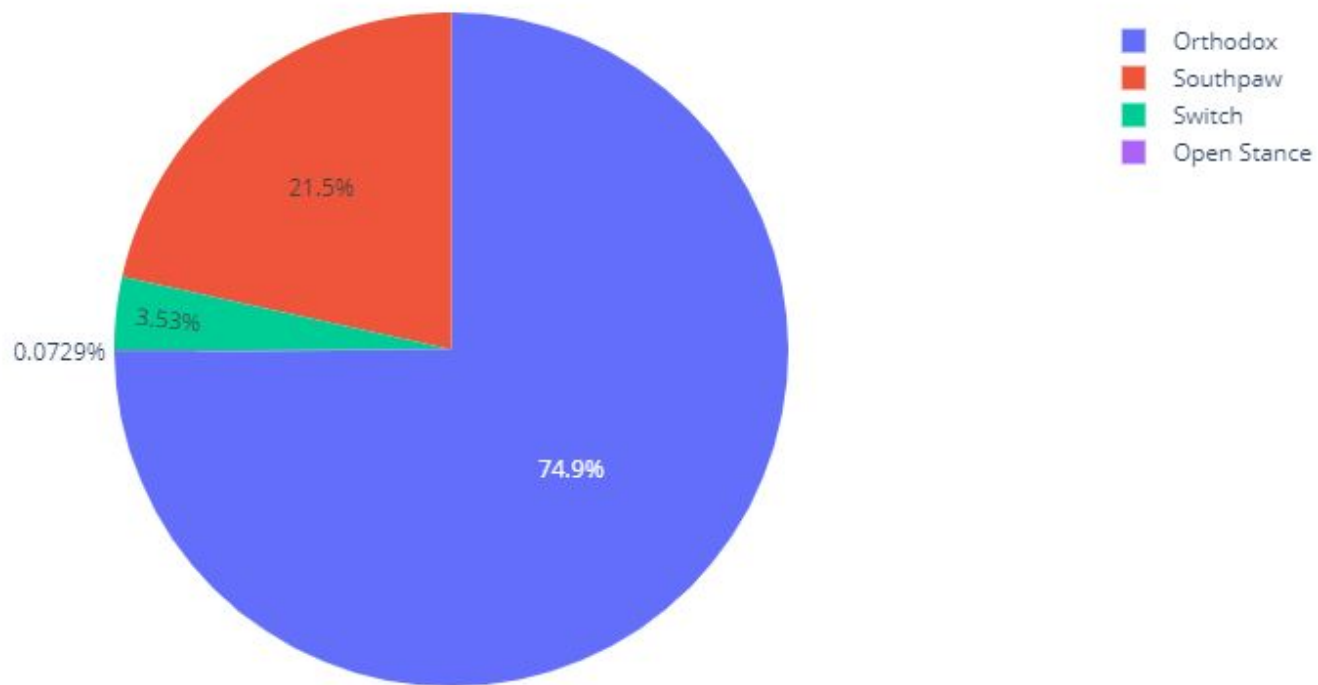
Orthodox - Right Hand/Right Foot back

Southpaw - Left Hand/Left Foot back

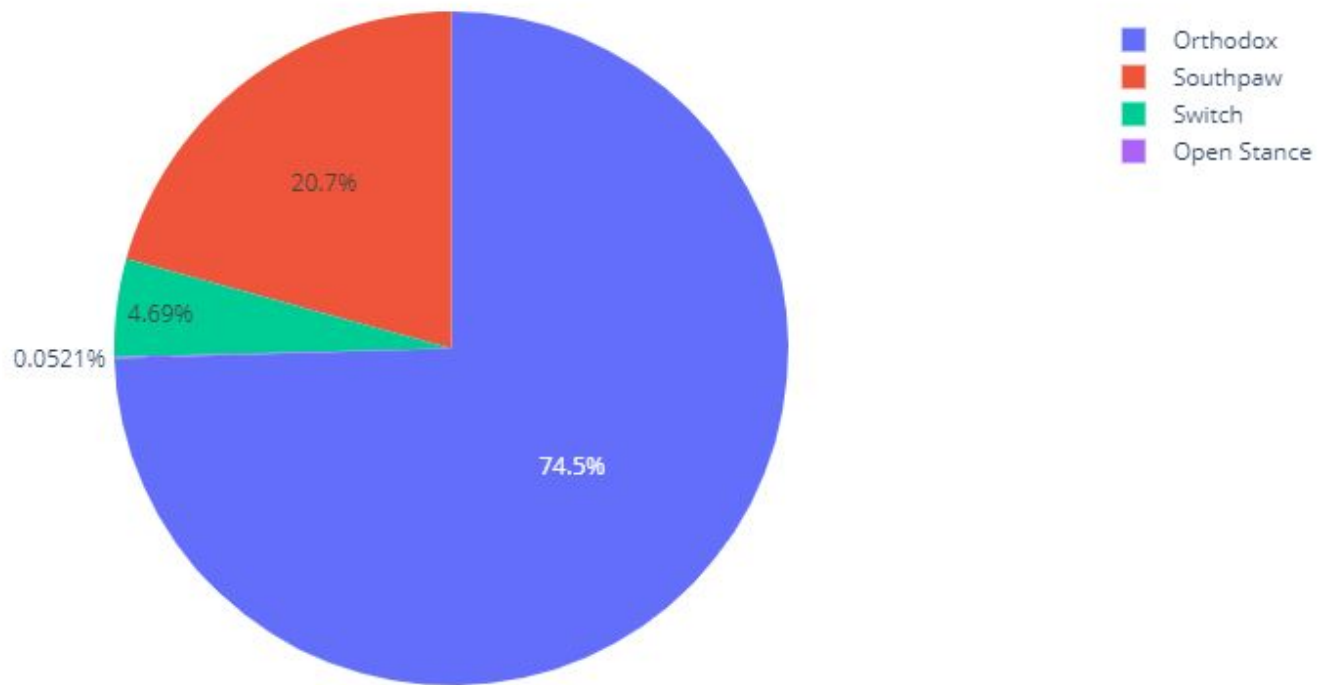
Switch - Fighters who do not favor one stance, often changing throughout the fight



## Stance Win Percentage on Red Side



## Stance Win Percentage on Blue Side

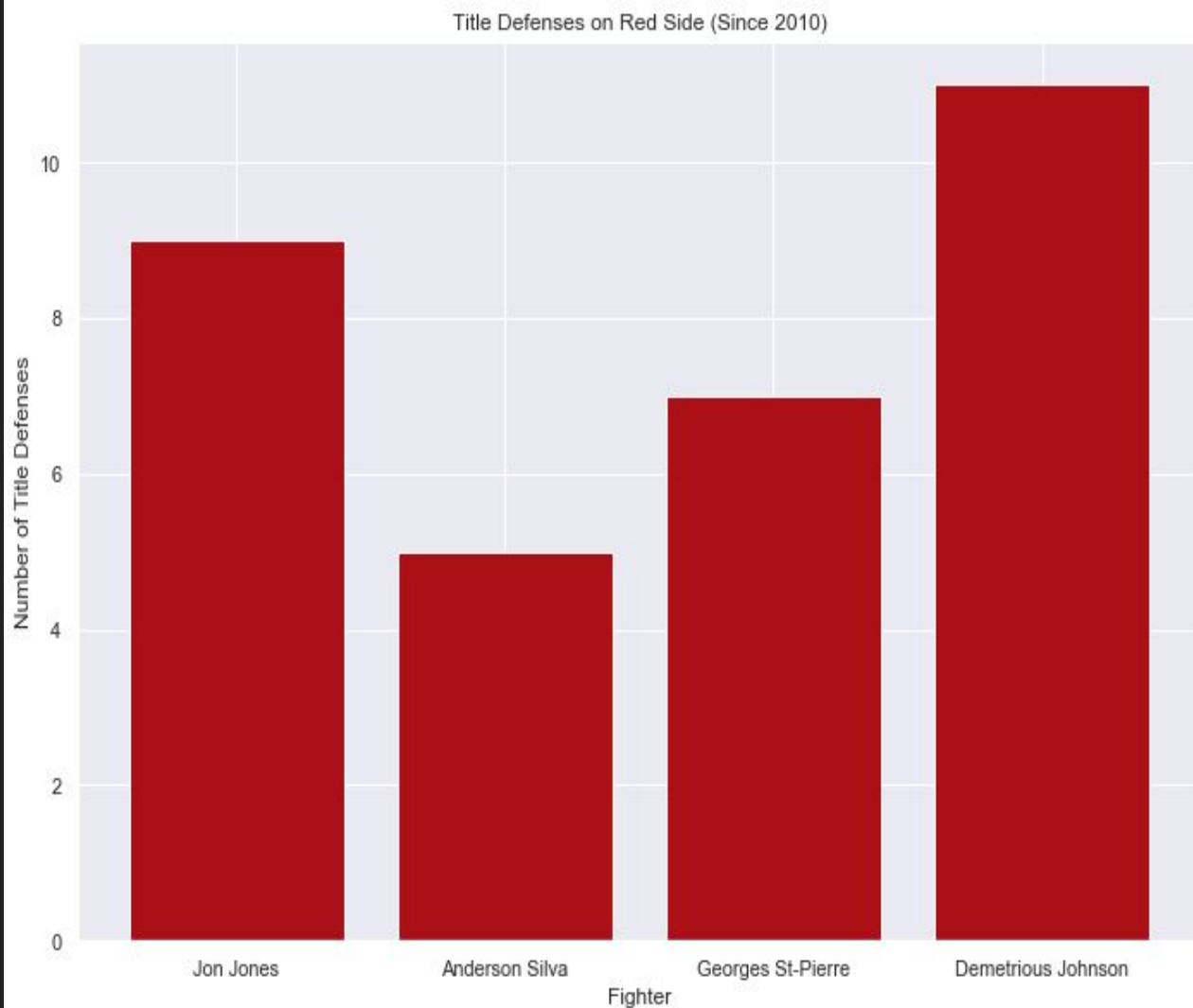


No major discrepancies between red side and blue side win rates when it comes to stance.

-4 of the UFC's most dominate champions have **ALL** had their title defenses on the Red side.

-This means the UFC sets the Red Side as the Champion side. This is not explicitly explained or mentioned by the UFC...

-Explains why Red side has a larger win rate in title fights



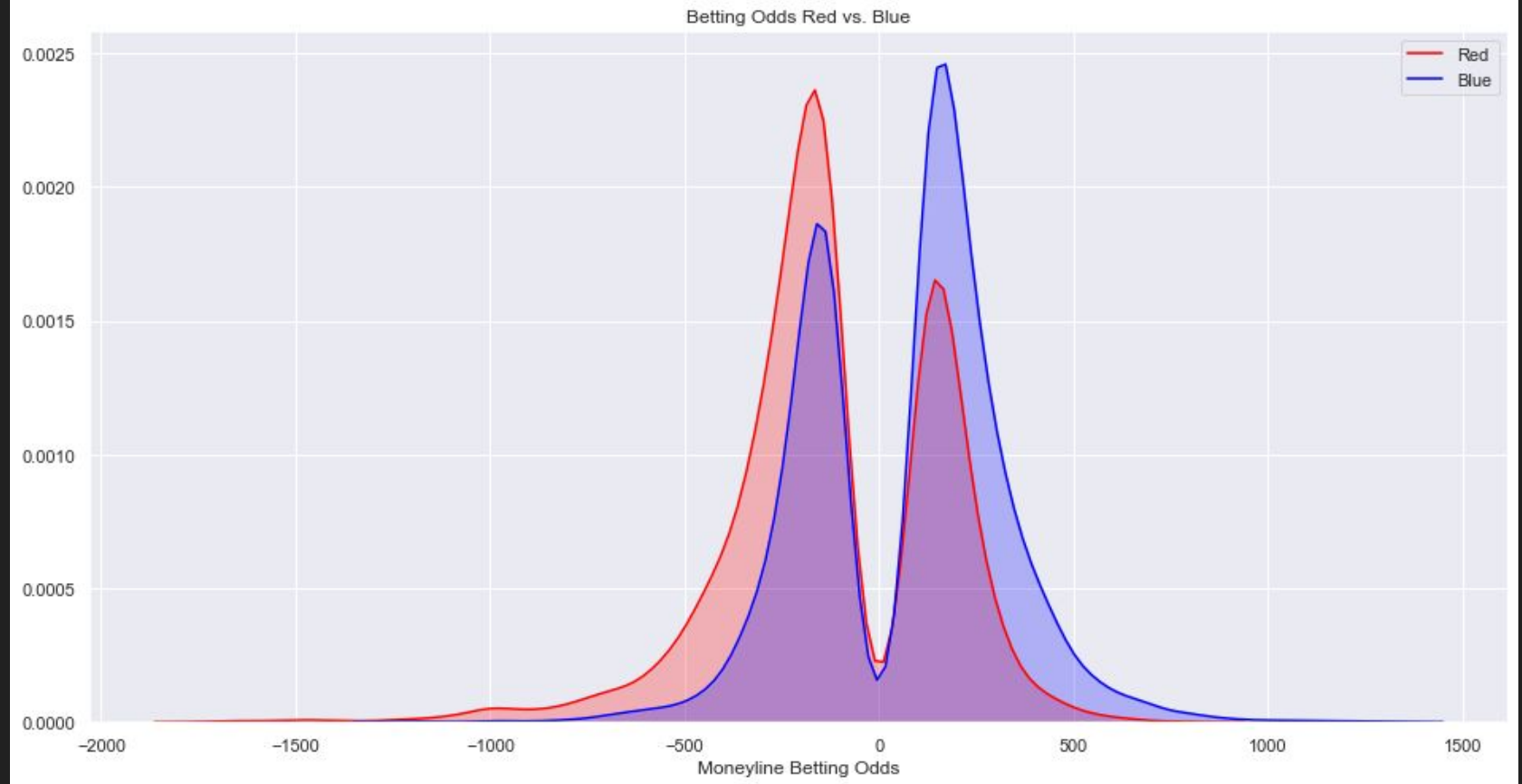


Top selling Pay-Per-Views of all time: Red is the current champion, if not a title fight, red side is the betting favorite

Event Name	Winner	Red_fighter	Blue_fighter	Buyrate	Title Fight
Khabib vs. McGregor	Red	Khabib	McGregor	2,400,000	Yes
Poirier vs. McGregor 3	Red	Poirier	McGregor	1,800,000	No
Diaz vs. McGregor 2	Blue	Diaz	McGregor	1,600,000	No
Poirier vs. McGregor 2	Red	Poirier	McGregor	1,600,000	No
McGregor vs. Cerrone	Red	McGregor	Cerrone	1,353,429	No
McGregor vs. Diaz	Blue	McGregor	Diaz	1,317,000	No
Usman vs. Masvidal	Red	Usman	Masvidal	1,300,000	Yes
Alvarez vs. McGregor	Blue	Alvarez	McGregor	1,300,000	Yes
Lesnar vs Mir 2	Red	Lesnar	Mir	1,300,000	Yes
Aldo vs. McGregor	Blue	Aldo	McGregor	1,200,000	Yes

# Betting Odds

-How do betting odds fit into the Red vs. Blue side debacle?

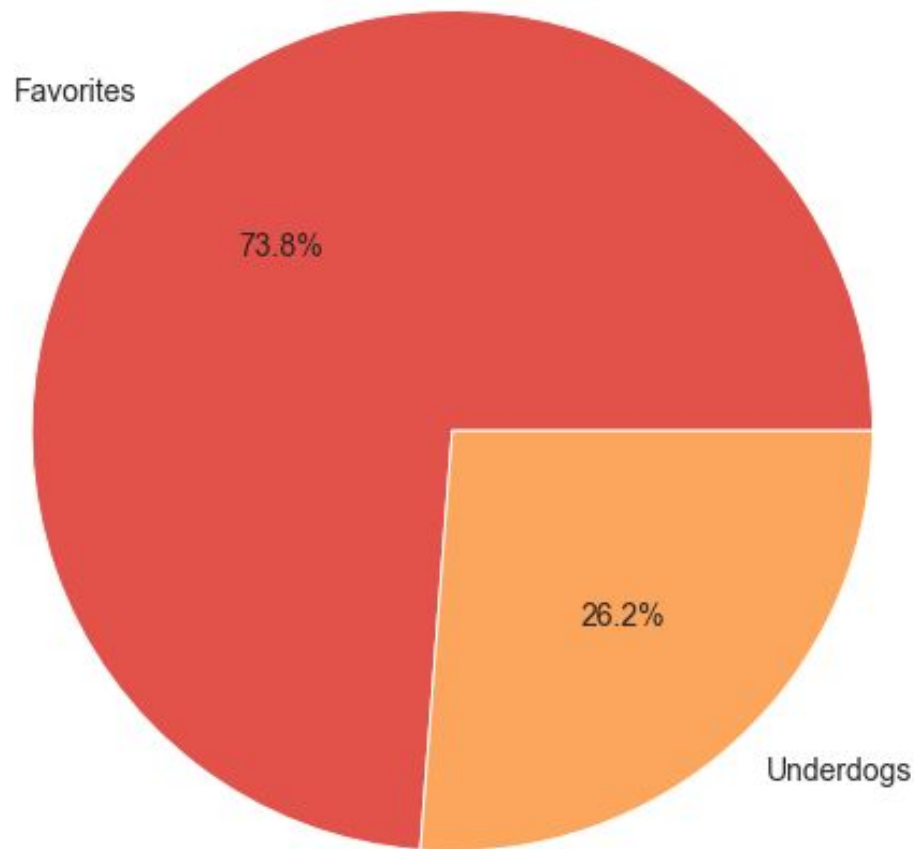


-Red Side tends to be the betting favorite.

-The mean odds for red are -118. The mean odds for blue are +68



-Favorites win 73.8% of the time. Since Redside is most often favored, it makes sense why Redside, combined with the fact that title defenders are on red side, contribute to red side having such a high win rate over blue side.



-Red side fighters are favored more often than Blue side fighters. Favored fighters win 73.8% of the time.

-Red fighters win more regular and title fights.

-In title fights, Red Side fighters are the defending champions everytime. Favored fighters in non-title fights are also on Red Side.

-Stance does not play a role in red vs blue side win rate, neither does age, height or reach.

# Future Work

- Create Random Forest Model using betting dataset
- Now that we know Red Side is the favorite and wins more often, would putting \$10 on every red fighter win or lose you money?
-