

EDA REPORT FOR UFC DATASET ANALYSIS



INTRODUCTION

This presentation will pose some questions and will try to answer them through some EDA-driven stories.



ESTABLISHING QUESTIONS

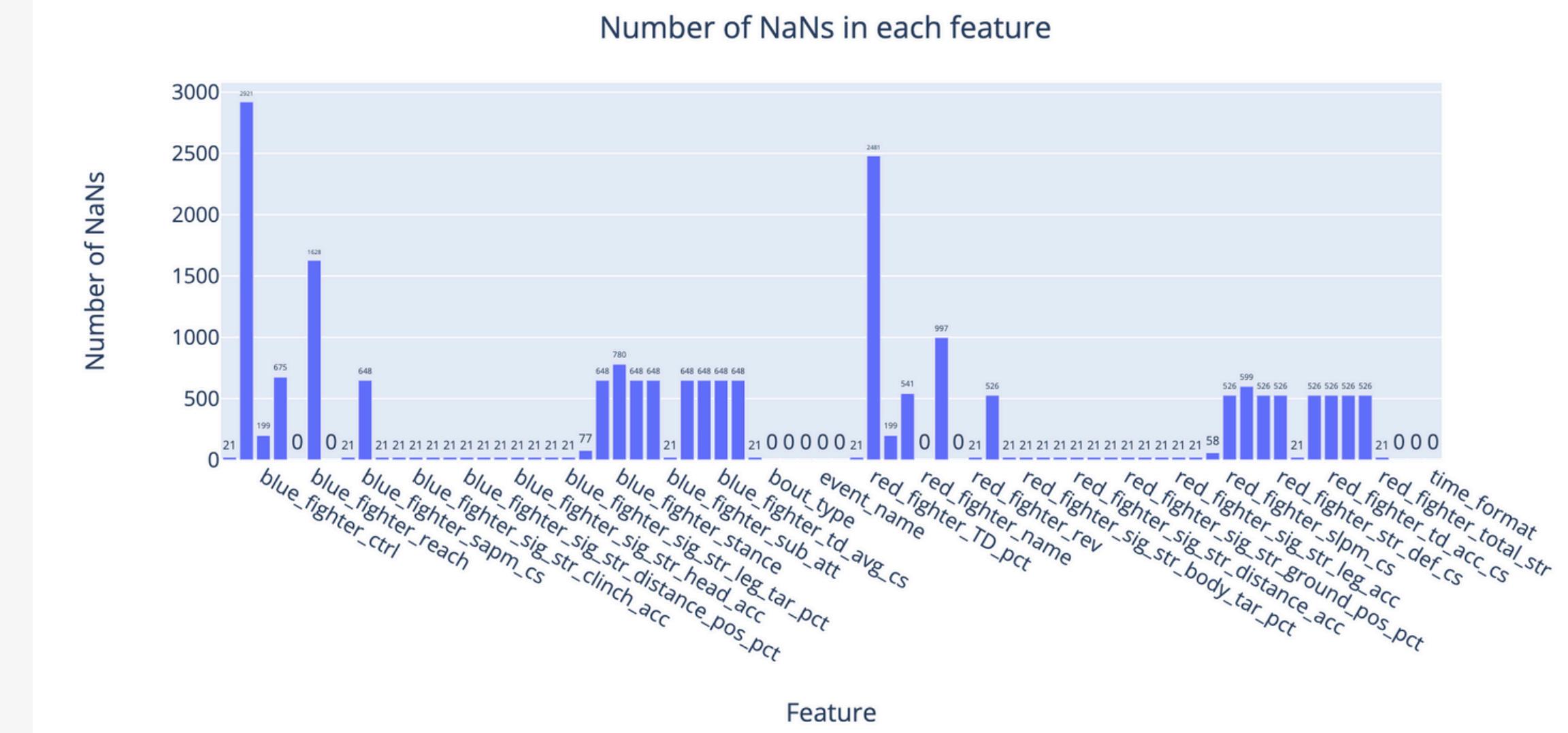
1. What are the most popular locations where UFC hosts its events?
2. What is the historical average number of UFC events hosted annually?
 3. What are the top fight outcomes?
 4. Does being in a particular corner contribute to a victory?
 5. Does having a specific fighting stance contribute to a victory?
 6. What are the main fight tactics contributing to a victory?
7. Do any specific anthropometric features (like height, reach, etc.) contribute to a victory?
8. Based on the findings, can we use the information to bet on a fighter?

UNDERSTANDING THE DATA

Missing data

Findings

- 54% of the features have NaN entries
 - 53% of these NaN features have 21 NaN entries, the rest ranges from 2% to 37% of NaN entries
 - 68% of the total pool of NaN entries are directly in the dataset as NaN, and the remaining 32% are in the form of NaN placeholders, like "-", "---", etc.
 - 54% of the pool of NaN placeholders like "-", "---", etc. are actually NaN placeholders, and the rest 46% of the pool is just Os, as the reason is not that the data was recorded but not entered, but in fact, was not recorded, e.g. a fighter did not land any strikes.

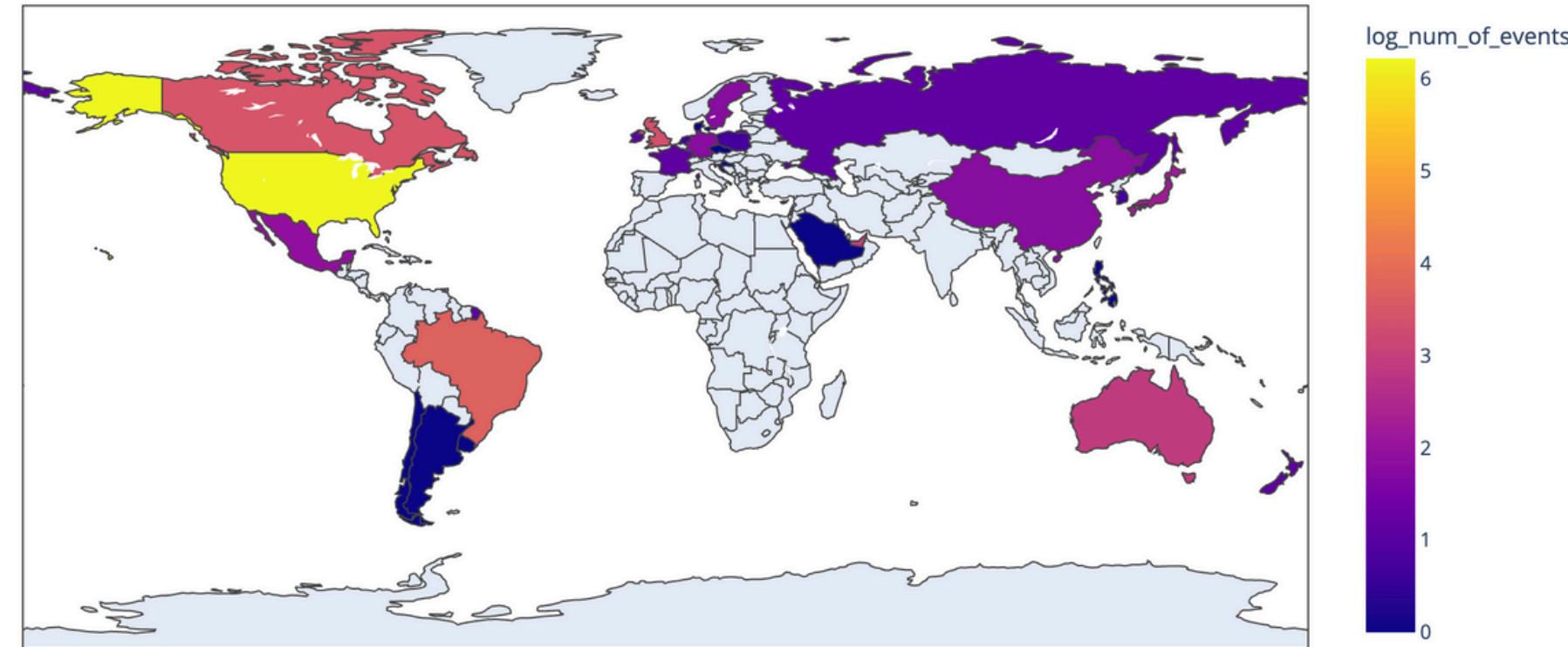


UNIVARIATE ANALYSIS

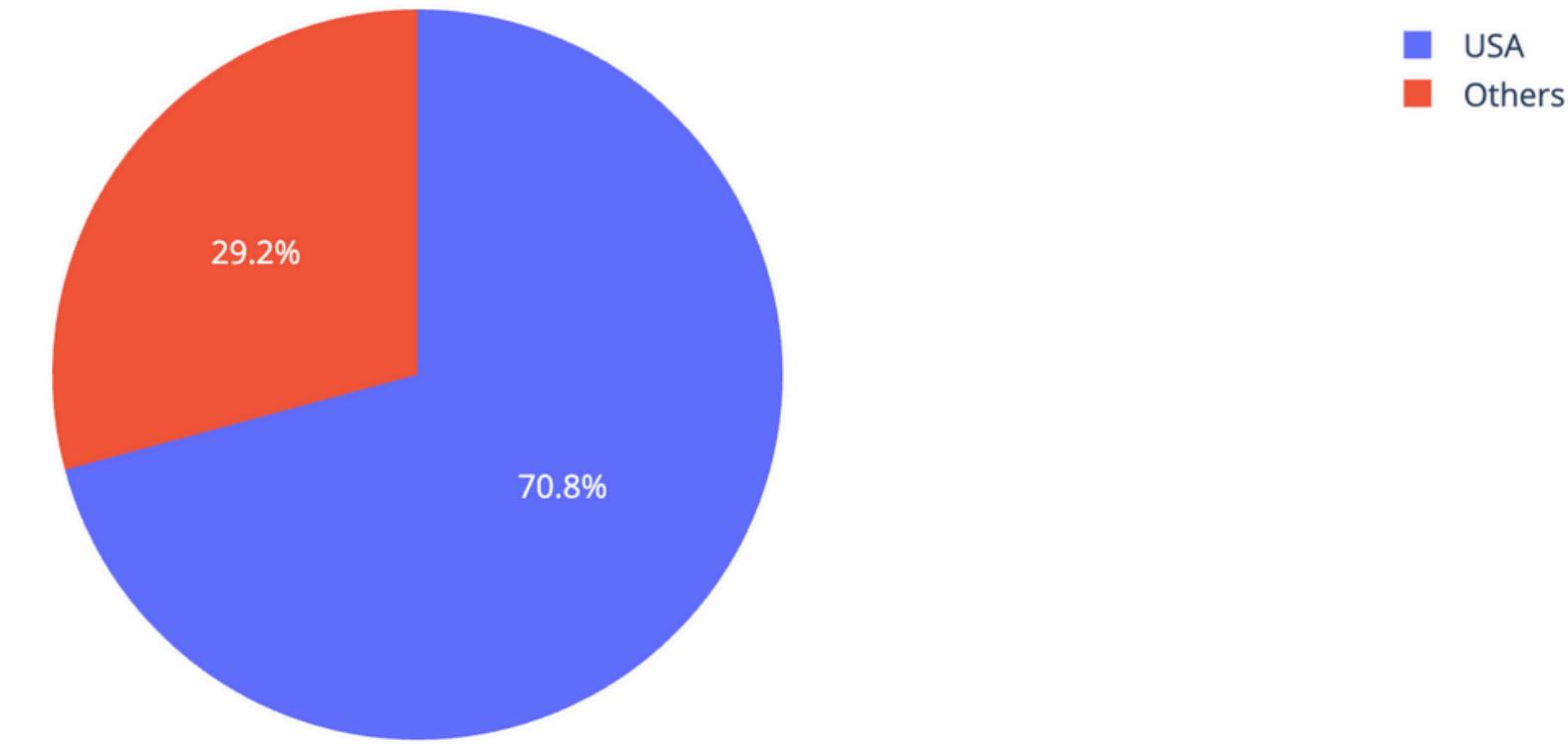
Question #1. What are the most popular locations where UFC hosts its events? (Countries)

Findings

- The **US takes 70.8%** of the total pool of the countries, where UFC events were hosted
- Followed by Brazil, Canada, The UK, UAE, Australia, Japan, Mexico, Germany, China, Sweden, Singapore, New Zealand, Russia, Ireland, France, South Korea, Poland, Netherlands, Saudi Arabia
- Factors like regulation, culture, fanbase, location, infrastructure, government support and others affect these rankings



Percentage of UFC events taking place in the US vs others

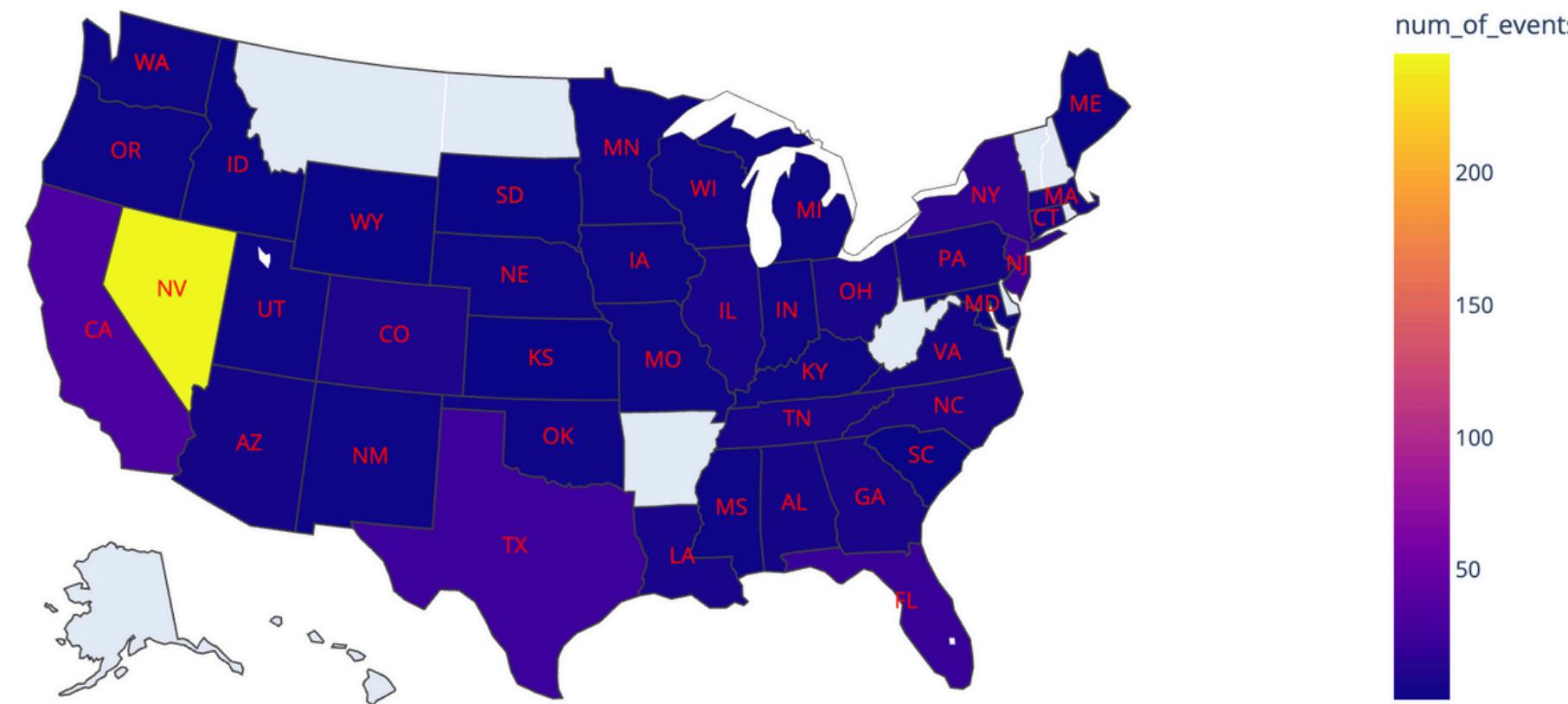


Question #1. What are the most popular locations where UFC hosts its events? (States)

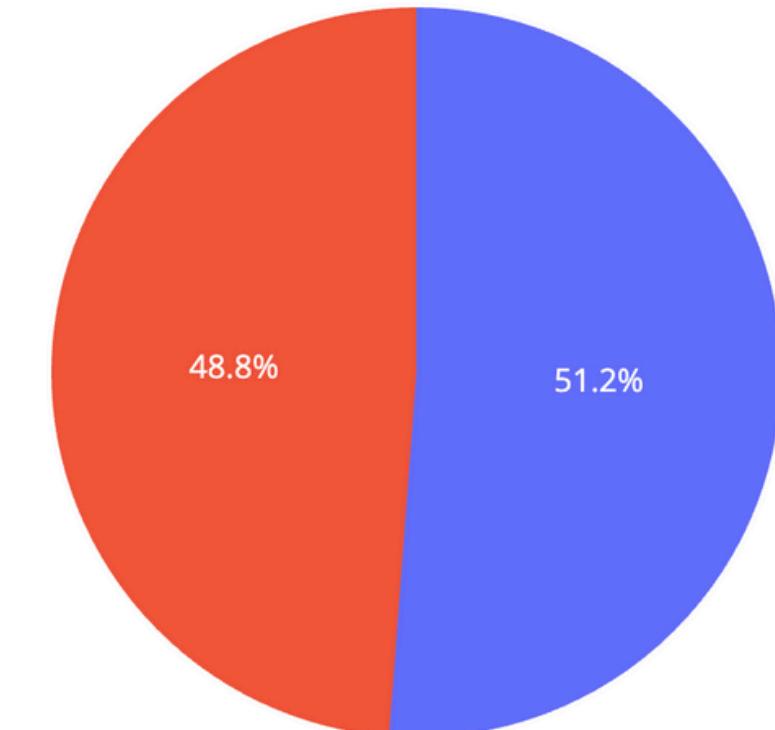
Findings

- Nevada, the fight capital of the world, takes 48.8% of the total pool of states
- Followed by California, Texas, Florida, New Jersey, New York, Colorado, Louisiana, Illinois, North Carolina, Massachusetts, Tennessee, Ohio and others.
- Factors like the cultural role, regulation, fanbase, “status” and media perception, population, and others affect these rankings.

The US states by # of UFC events

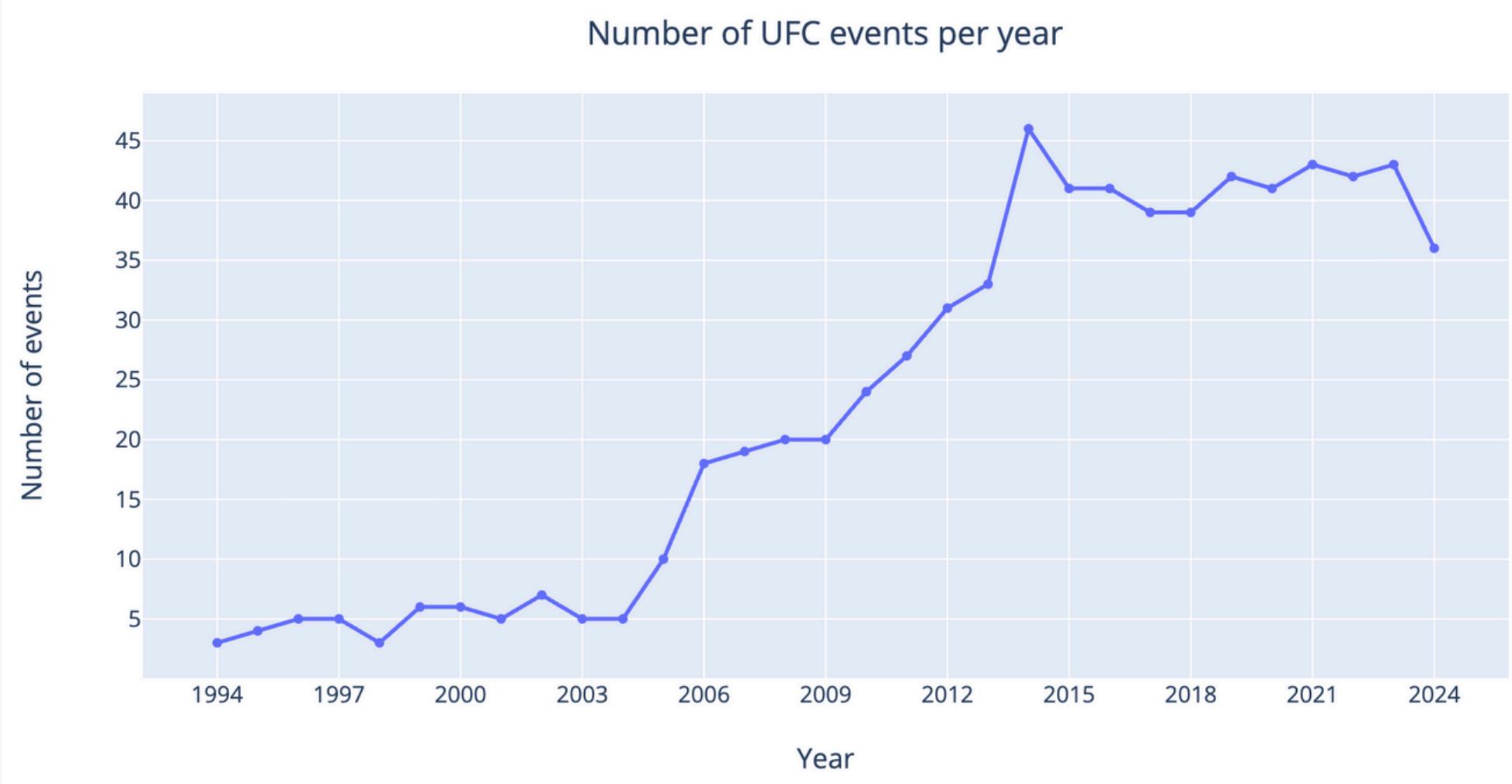


Percentage of UFC events taking place in Nevada vs other US states



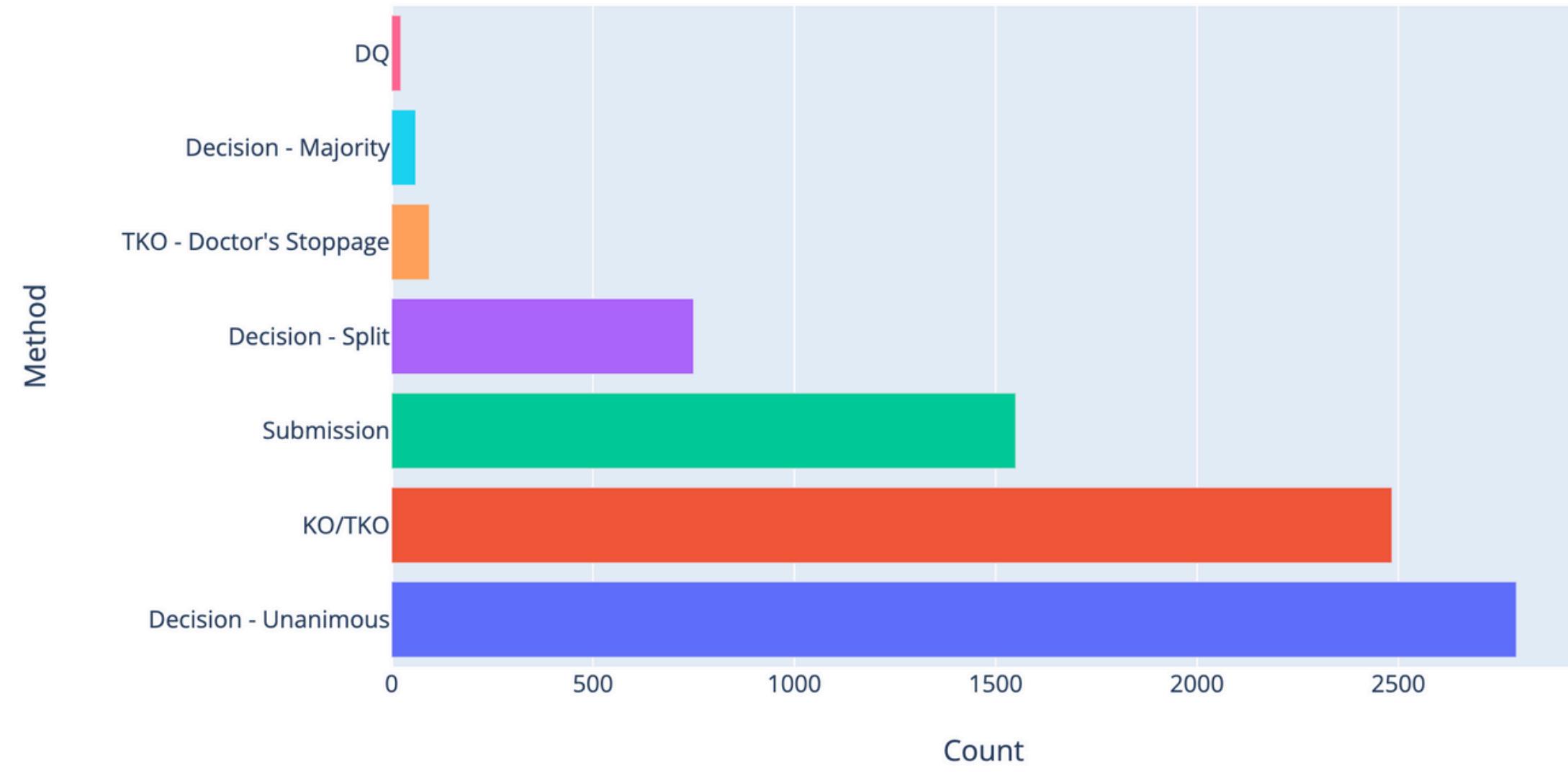
Question #2. What is the historical average amount of UFC fights hosted annually?

Findings



- UFC started out with initial steady growth throughout 1995–2004 with an average of 4 annual events
- Then, the 2005–2010 rapid growth period was fueled by a **deal with Spike TV** to broadcast the Ultimate Fighter show, at the same time UFC was starting to being taken as a **legitimate sports property**.
- In 2011–2014 UFC **entered Brazil, Asia, and Europe**. Led to a jump of from 27 events annually to 46.
- In 2015–2019 UFC **signed a deal with ESPN**, ended up in ~40 events on average
- In 2020–2021, the covid era, but UFC did well due to having build **UFC Apex**, their custom fighting arena, as well as **Abu Dhabi's Fight Island**.
- In 2021–2024 shows how fast UFC **normalized their events**, **Florida being the first state** to allow to resume hosting events.
- The 2024 period is not finished yet, but expected to **meet the average** of previous years.

Question #3. What are the top fight outcomes?

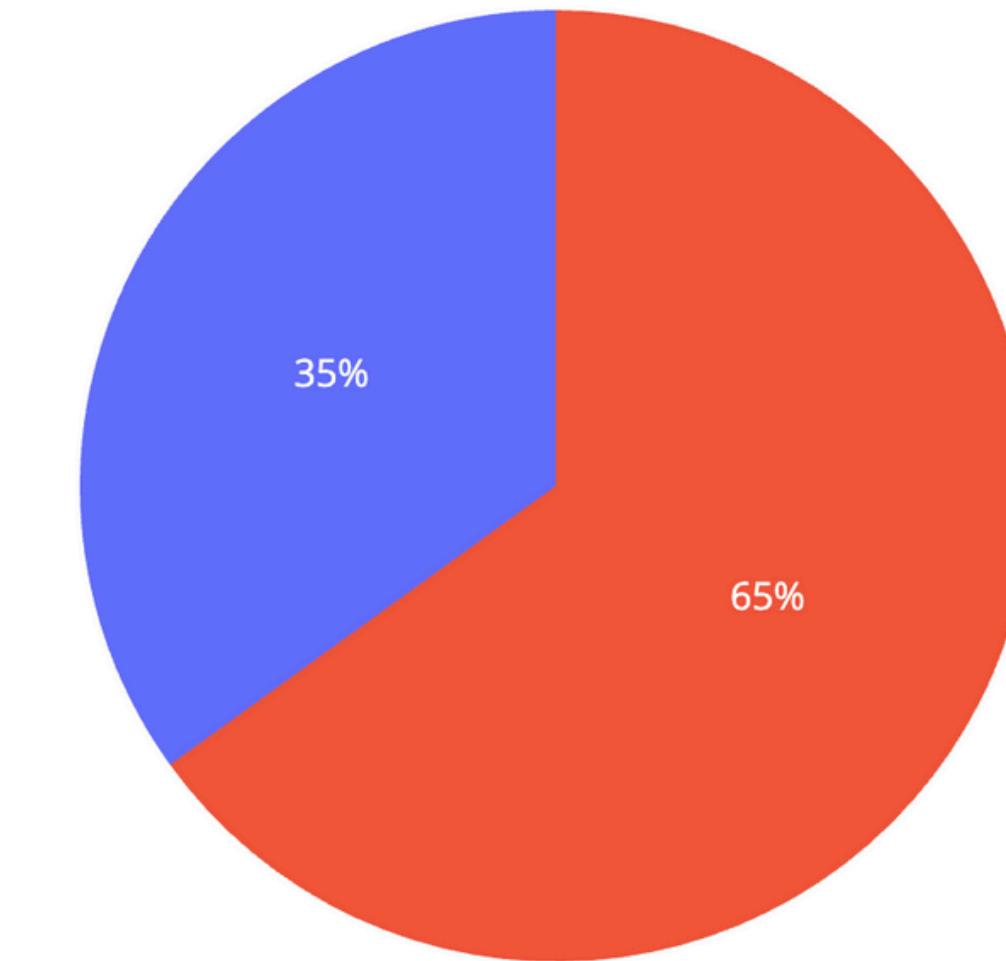


Findings

- Decision-Unanimous is the top 1, followed by KO/TKO at #2, followed by Submission at #3, followed by Decision-Split, TKO-Doctor's Stoppage, Decision-Majority, DQ
- Reasoning is that it's much **easier to win on points than by a KO/TKO**, as a fighter can just "play it safe", then there's more KO/TKOs than Submissions because there's **more strikers than wrestlers**, and there's much **more striking time in fights than wrestling**.
- Non-unanimous decisions, along with Doctor's stoppage, are at the end because the judges are **incentivized to give a clear winner**

Question #4. Does being in a particular corner contribute to a victory?

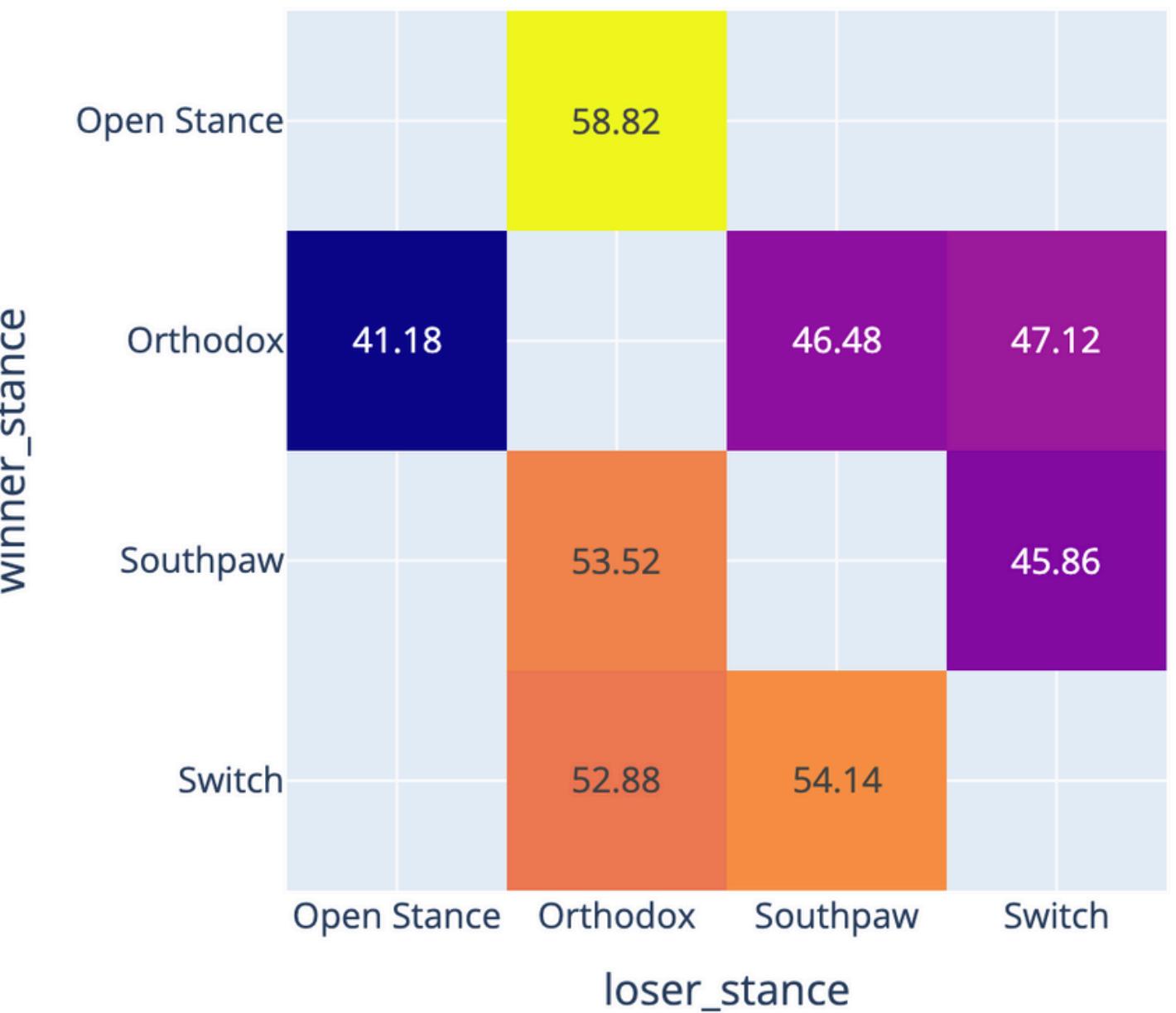
Total % of winners in the corresponding corner



Findings

- Fighters from the **red corner** win **65%** of the time, whereas fighters from the **blue corner** win **only 35% of the time**
- The reasoning here is that the **red corner** is given to **champions and higher-ranked fighters**, whereas the **blue corner** is given to the **contenders and lower-ranked fighters**.

Fighting stance matchup winrates



Question #5. Does having a specific fighting stance contribute to a victory?

Findings

- Yes, it does contribute. The ranking is as the following: **Switch > Southpaw > Orthodox**. Open Stance is not included because the stance is too rare, and there's too small of a sample size to draw any conclusions
- The reasoning here, is that **the rarer a fighter's stance is, the trickier and more uncomfortable it is to fight that person** because of how practice-based this sport is.
- There's around **80.3% of orthodox fighters, 17.4% southpaws and 2.3% of other stances**

MULTIVARIATE ANALYSIS

Question #6. What are the main fighting aspects contributing to a victory? Boxplots.

Findings

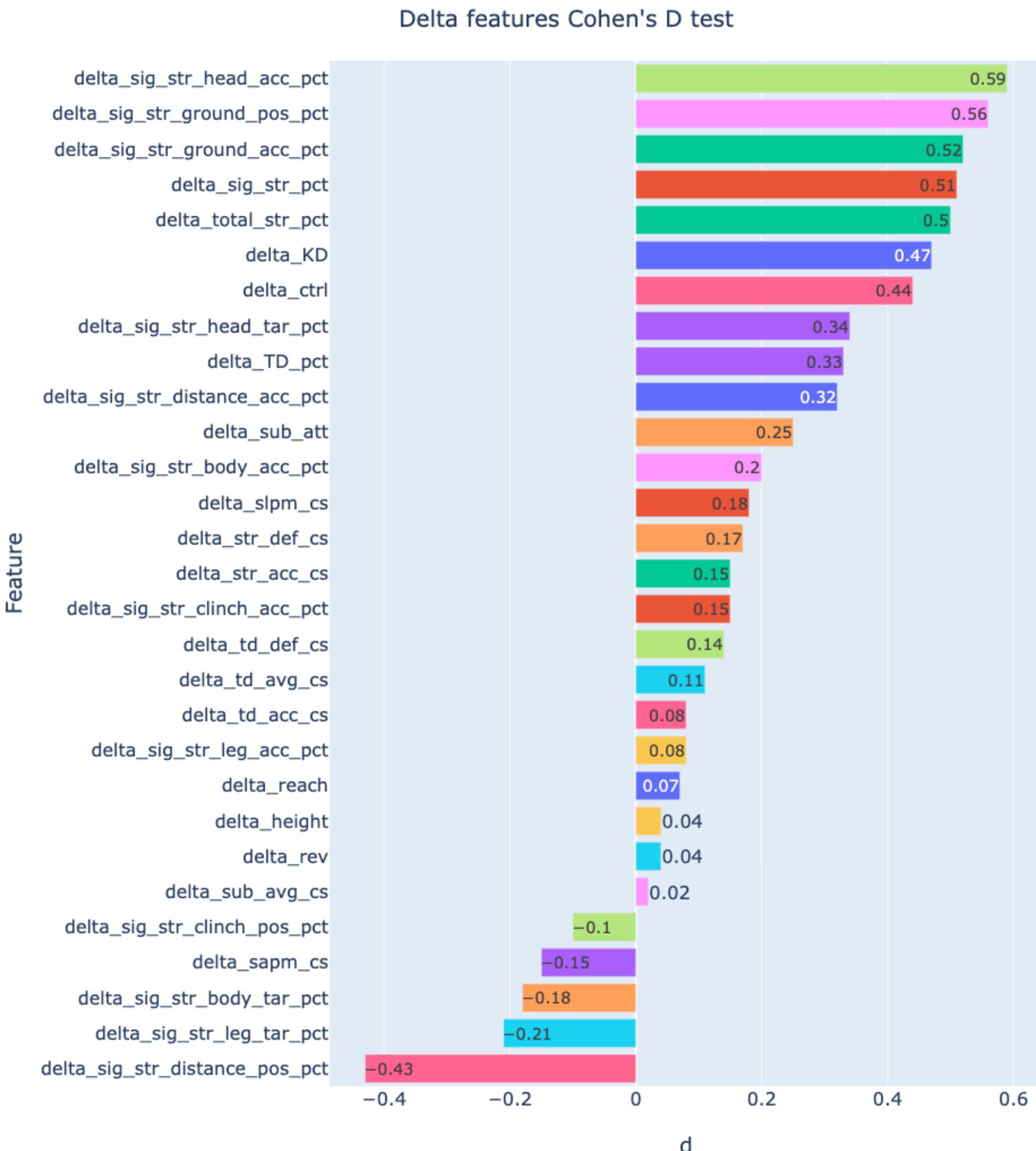
- None of the features except two seem to be undersampled from this quick boxplot chart over all the numerical features
- That means, we can run Kohen's D on these non-undersampled features to try and rank them



Cohen's D on performance-based features

Findings

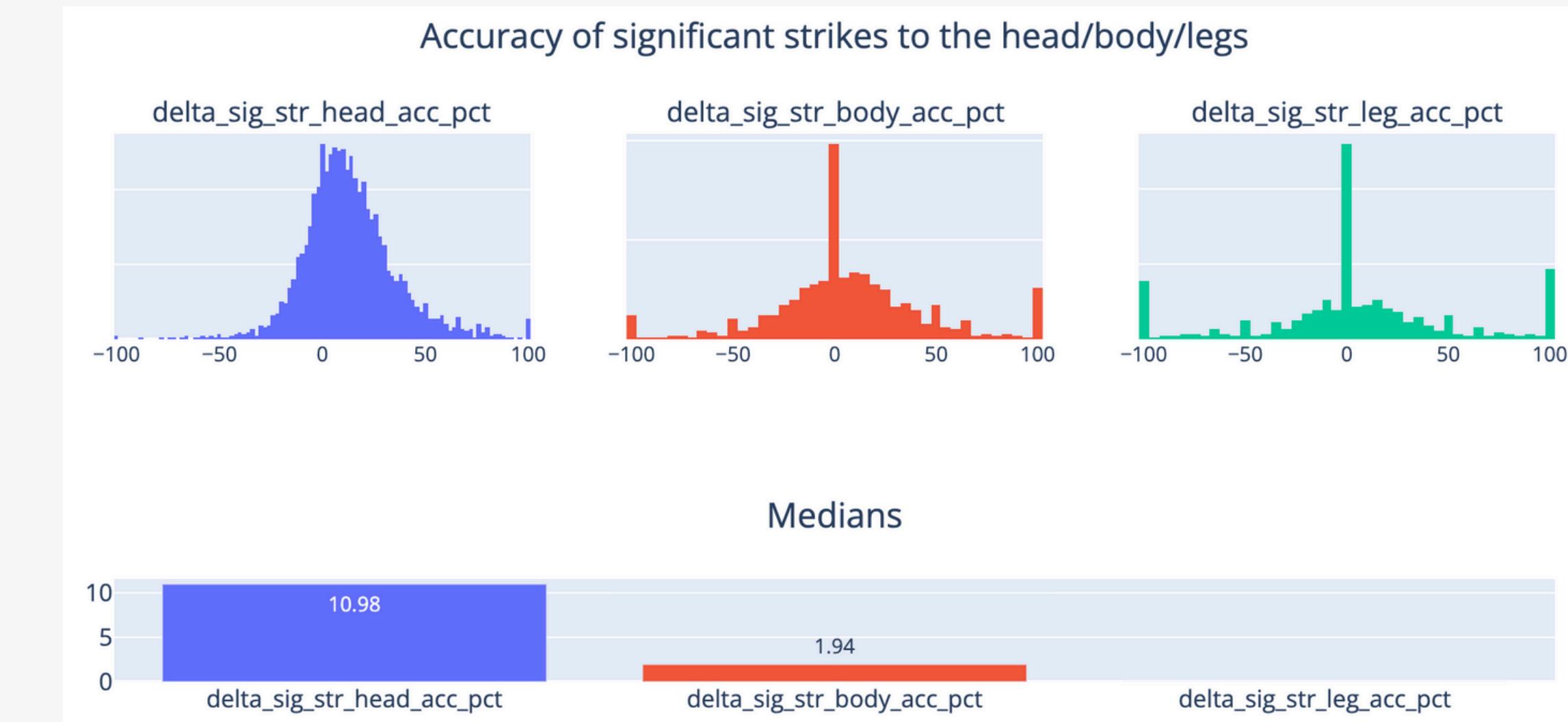
- Some of the most important features are: accuracy of **sig. str. to the head**, accuracy of sig. str. from the **ground position**, targeted sig. str. in the **ground position**, and others.
- Some of the features on medium importance scale are: takedowns, submission attempts, sig. str. from clinch, and others.
- Some of the features that negatively contribute to winning are: sig. str. from distance position, targeted sig. str to the legs, and others.



Accuracy of significant strikes to the head/body/legs

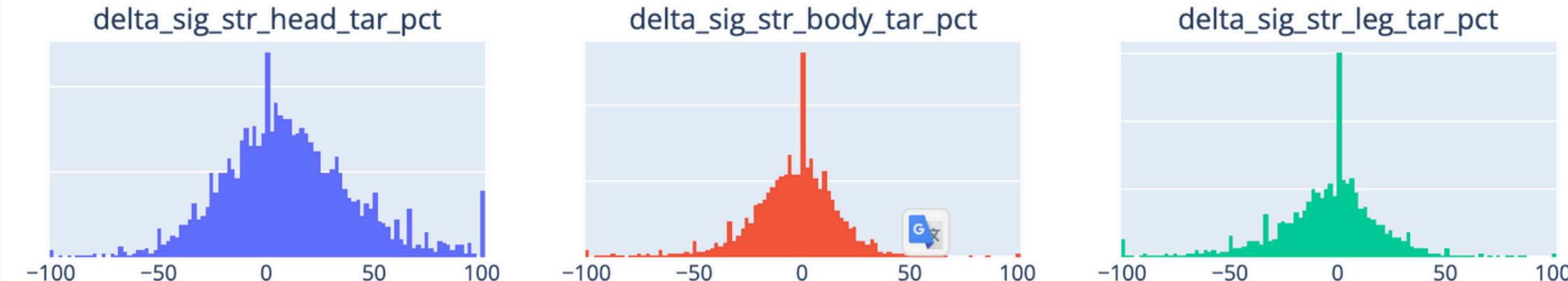
Findings

- Accurate sig. str. to the head hint on being important in winning, followed by sig. str. to the body having a smaller importance, ending with strikes to the legs showing no significant connection to winning.
- But accuracy alone is not a reliable predictor of winning in fights, as a fighter may not be better on accuracy and instead, win on volume.
- Conclusion: Head > body > legs

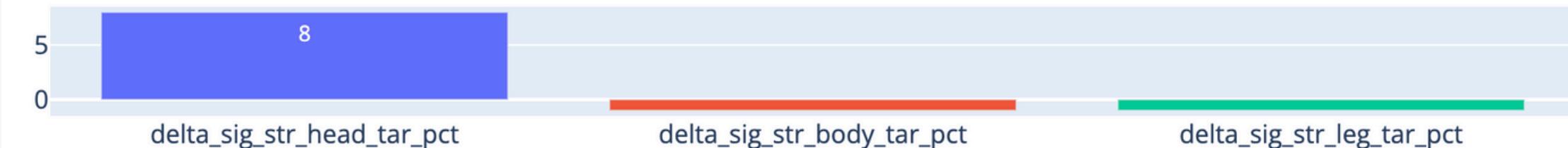


Distribution of sig. str. to the head/body/legs

Distribution of sig. str. to the head/body/legs



Medians



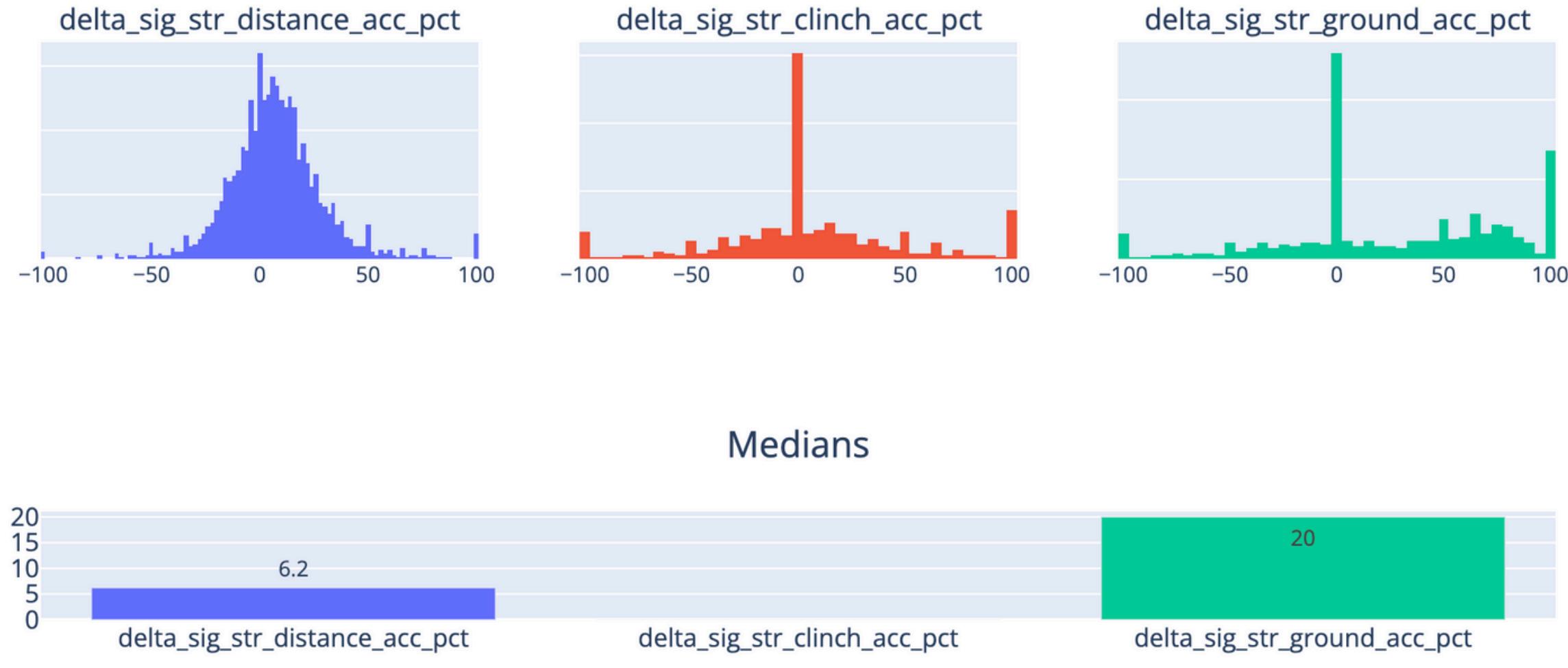
Findings

- Winners prefer targeting their significant strikes to the head > body = legs.
- Head targeting is the key **differentiator for winning**, whereas **sig. str. to the body/legs** seem to be **less important** but nonetheless, it's not like winners don't go for these targets, we can see in the num. of outliers; So they still are valuable, but yield value rather in the long run, on the other hand, head strikes are more visible, lead to cuts, thus showing significant damage, sway judges, and lead to KDs / finishes.
- Conclusion: **Head > body = legs**

Accuracy of sig. str. landed in ground/clinch/distance position

Findings

- Ground strikes are the most effective, with high accuracy and impact.
- Distance strikes have a positive median but less impact.
- Clinch strikes show low accuracy and effectiveness.
- Training styles affect striking, with distance strikes being universal.
- Ground strikes are crucial for winning, followed by distance strikes; clinch strikes are least effective.
- Conclusion: **Ground > Distance > Clinch**



Distribution of sig. str. landed in ground/clinch/distance position

Findings

- Significant striking from the ground position is crucial for winning, supporting the idea that “on average, wrestlers beat strikers.”
- Several factors contribute to changes in outcomes:
 - a. The feature's calculation distributes strikes across three positions, affecting other features.
 - b. Most fighters strike from a distance, while fewer strike from positions like the ground due to the lower number of wrestlers.
 - c. Winners tend to strike from the ground or clinch for safety and control.
- Conclusion: **Ground > clinch > distance.**

Distribution of sig. str. landed in ground/clinch/distance position

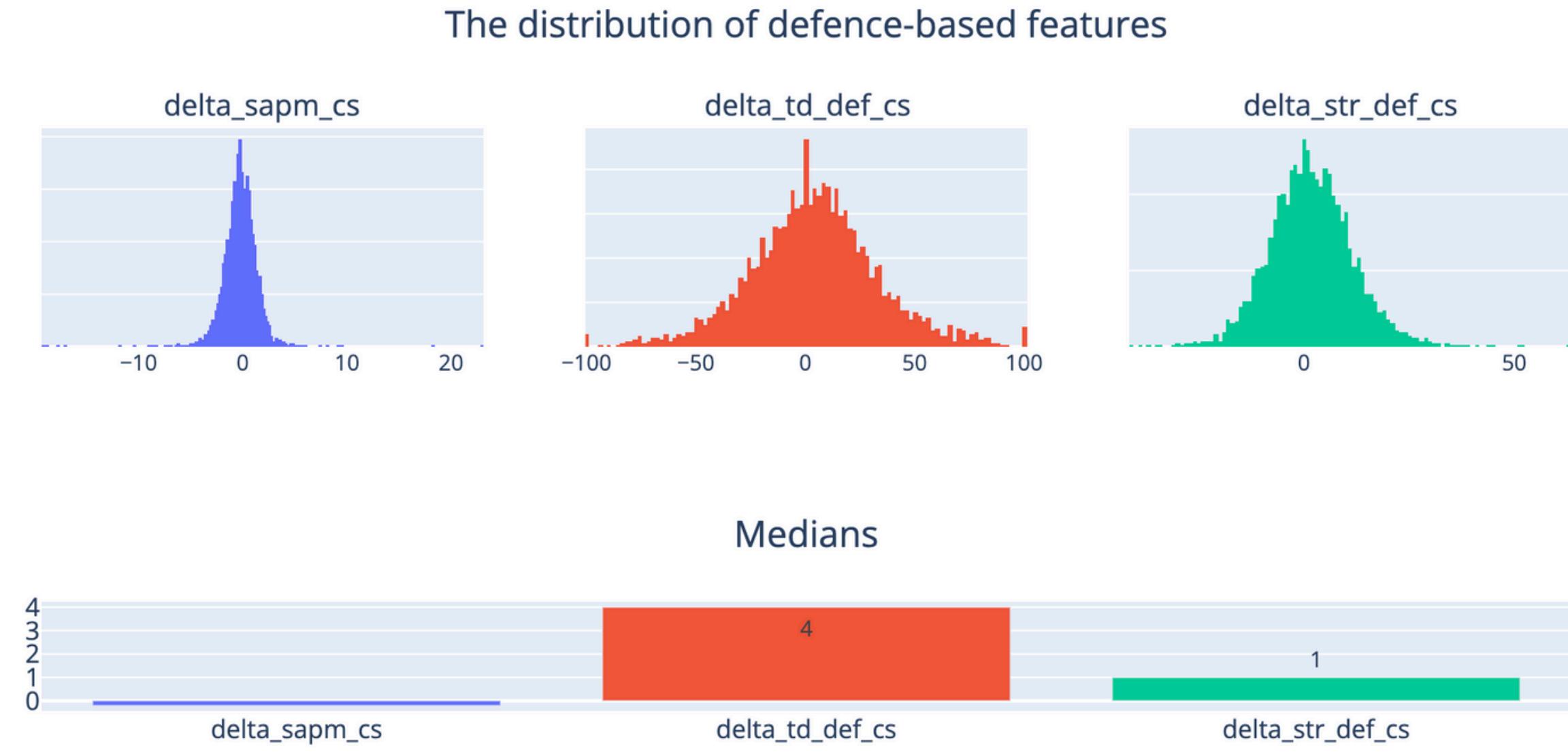


The distribution of defence-based features

The distribution of defence-based features

Findings

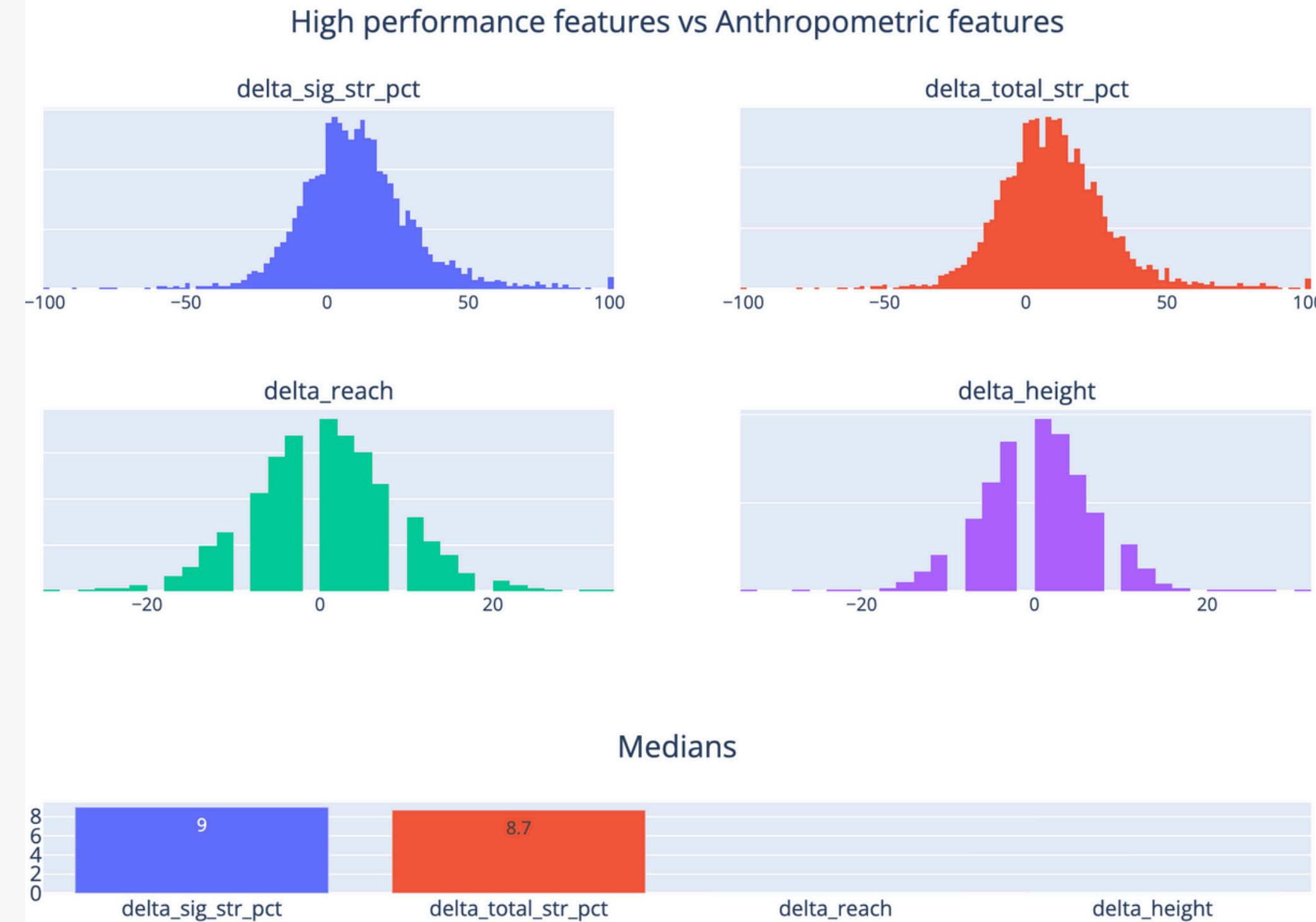
- SAPM (Significant Strikes Absorbed per Minute): Lower values are better. Median is around 0, but winners often have lower SAPM. This isn't definitive due to many close or stance-end fights.
- Takedown Defense: Has the highest median and links to winning, especially for strong wrestlers.
- Striking Defense: Shows a slight positive skew, centered around 0, not a strong winning predictor.
- Conclusion: **Takedown defense is more crucial than striking defense and SAPM.**



Question #7. Do any specific anthropometric features (like height, reach, etc.) contribute to a victory?

Findings

- Anthropometric features like reach and height generally do not provide a clear advantage to fighters, except in certain matchups. In striking, taller fighters or those with longer reach often outperform shorter fighters. However, in MMA, wrestlers—who tend to be shorter and have shorter reach—can defeat strikers, balancing the advantage.
- Conclusion: **sig_str = total_str > reach = height**



Question #8. Based on the findings, can we use the information to bet on a fighter?

Findings

MMA is matchup-based, with no guaranteed winning factors. Key betting insights include:

- Focus on wrestlers who control fights on the ground, landing strikes, and demonstrating strong takedown defense.
- Prioritize fighters who target the head, as this leads to wins, cuts, and KOs.
- Avoid betting on fighters who only strike from a distance, lack wrestling skills, or focus on body/leg strikes, as they tend to have less control and visual dominance in fights.

Presented By : @komaksym

THANK YOU

