Predicting the outcome of a UFC fight

Navin Sanjay

April 2023

Purpose

- Predictive modelling was used to find the likelihood of a fighter winning by KO/Submission or by Judge Decision.

- **Target Audience**:
  - Betting agencies can adjust their odds accordingly
  - Individuals can use this information when betting

Background

- The UFC (Ultimate Fighting Championship) is a mixed martial arts organization that features highly skilled fighters from around the world competing against each other in a variety of weight classes.
- Data was taken from the last **10 years** of the **Lightweight category** and assessed to predict the likelihood of a match resulting in a KO/Submission or Judge Decision.
  - KO: Knockout
  - Submission: Fighter causes opponent to give up
  - Judge Decision: Outcome of the fight is determined by three judges scoring each round based on the fighters' performance.

Data Overview and Data Pre-processing

Overview of Data

- Data was obtained from Kaggle
  - https://www.kaggle.com/datasets/fatismajli/ufc-data
- UFC Events Data
  - Data on each event and their results
  - *771 rows x 11 cols*
- UFC Fighter Data
  - Data on each UFC fighter and their attributes
  - *1692 rows x 8 cols*

Overview of Data – **UFC Fighter Data**

	First Name	Last Name	Nickname	Height	Weight	Reach	Stance	Wins	Losses	Draws
0	Tom	Aaron	NaN	--	155 lbs.	--	NaN	5	3	0
1	Danny	Abbadi	The Assassin	5' 11"	155 lbs.	--	Orthodox	4	6	0
2	Nariman	Abbasov	Bayraktar	5' 8"	155 lbs.	66.0"	Orthodox	28	4	0
3	David	Abbott	Tank	6' 0"	265 lbs.	--	Switch	10	15	0
4	Hamdy	Abdelwahab	The Hammer	6' 2"	264 lbs.	72.0"	Southpaw	5	0	0

Overview of Data – **UFC Event Data**

	Event Name	Event Date	Result	Fighter1	Fighter2	KD	Strikes	TD	Sub	Weight Class	Method	Round	Time
11	UFC 286: Edwards vs. Usman 3	2023-03-18	Justin Gaethje	Justin Gaethje	Rafael Fiziev	0-0	103-97	1-0	0-0	Lightweight	M-DEC	3	5.000000
16	UFC 286: Edwards vs. Usman 3	2023-03-18	Chris Duncan	Chris Duncan	Omar Morales	0-0	53-47	5-0	0-0	Lightweight	S-DEC	3	5.000000
17	UFC 286: Edwards vs. Usman 3	2023-03-18	Yanal Ashmouz	Yanal Ashmouz	Sam Patterson	0-0	15-2	1-0	0-0	Lightweight	KO/TKO-Punches	1	1.250000
23	UFC 286: Edwards vs. Usman 3	2023-03-18	Draw	Jai Herbert	Ludovit Klein	0-0	43-47	1-2	0-0	Lightweight	M-DEC	3	5.000000
41	UFC 285: Jones vs. Gane	2023-03-04	Mateusz Gamrot	Mateusz Gamrot	Jalin Turner	0-1	29-40	4-0	0-0	Lightweight	S-DEC	3	5.000000

Overview of Data – Data Cleaning

Object (string)

Integers & Floats

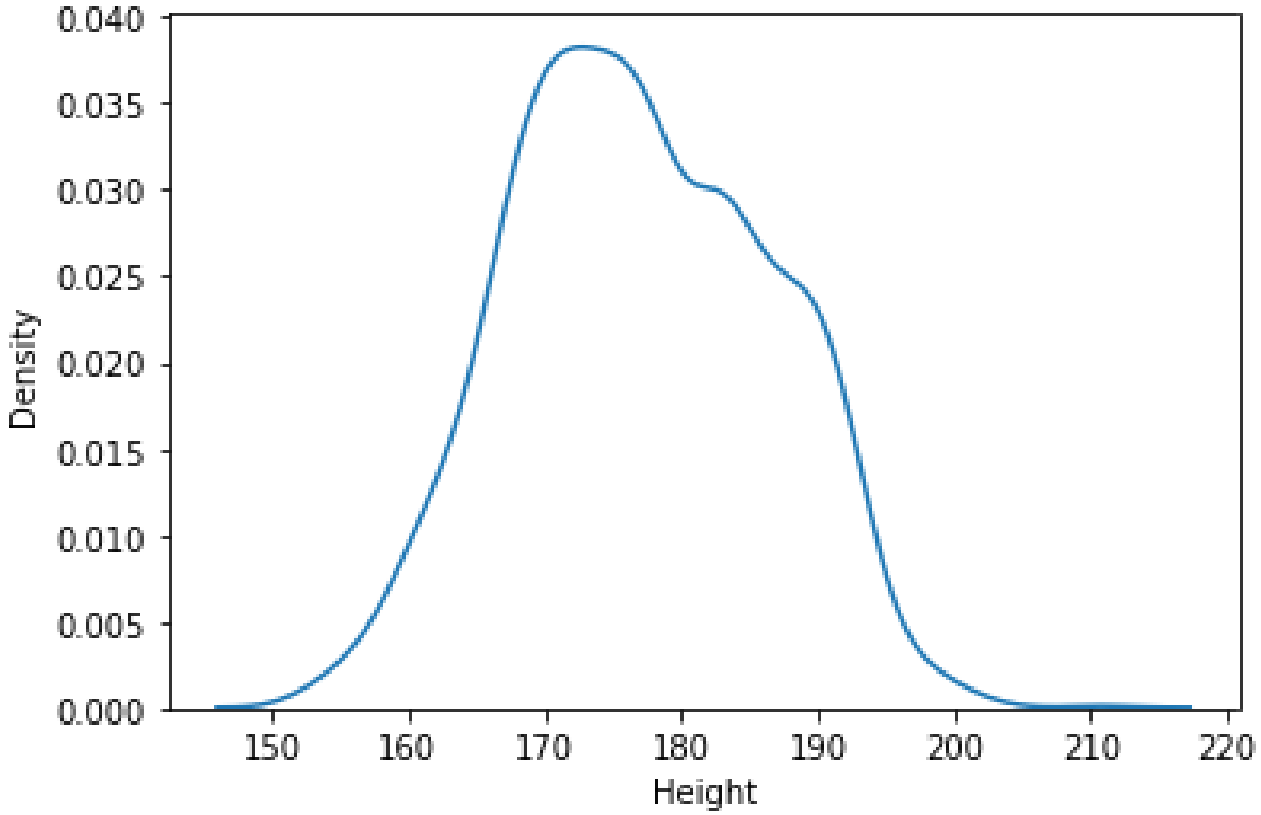
	First Name	Last Name	~~Nickname~~	Height	Weight	Reach	~~Stance~~	Wins	Losses	Draws
0	Tom	Aaron	~~NaN~~	--	155 lbs.	--	~~NaN~~	5	3	0
1	Danny	Abbadi	~~The Assassin~~	5' 11"	155 lbs.	--	~~Orthodox~~	4	6	0
2	Nariman	Abbasov	~~Bayraktar~~	5' 8"	155 lbs.	66.0"	~~Orthodox~~	28	4	0
3	David	Abbott	~~Tank~~	6' 0"	265 lbs.	--	~~Switch~~	10	15	0
4	Hamdy	Abdelwahab	~~The Hammer~~	6' 2"	264 lbs.	72.0"	~~Southpaw~~	5	0	0

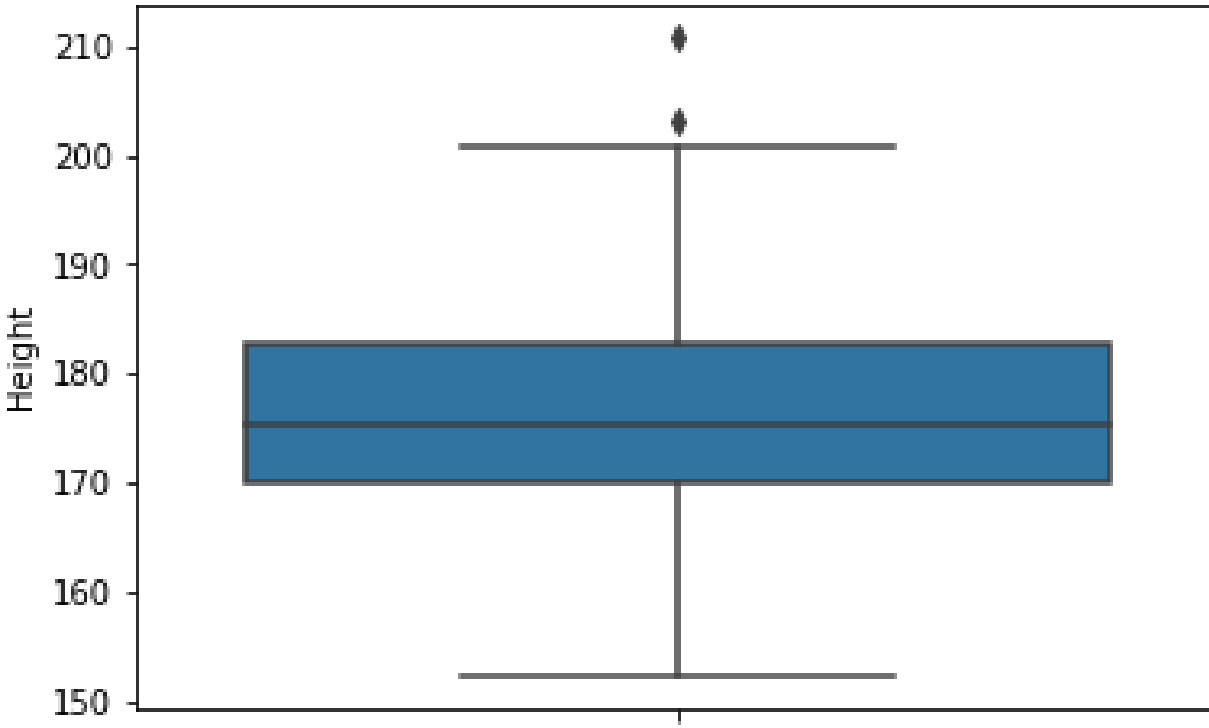
Overview of Data – Data Cleaning

Target variable

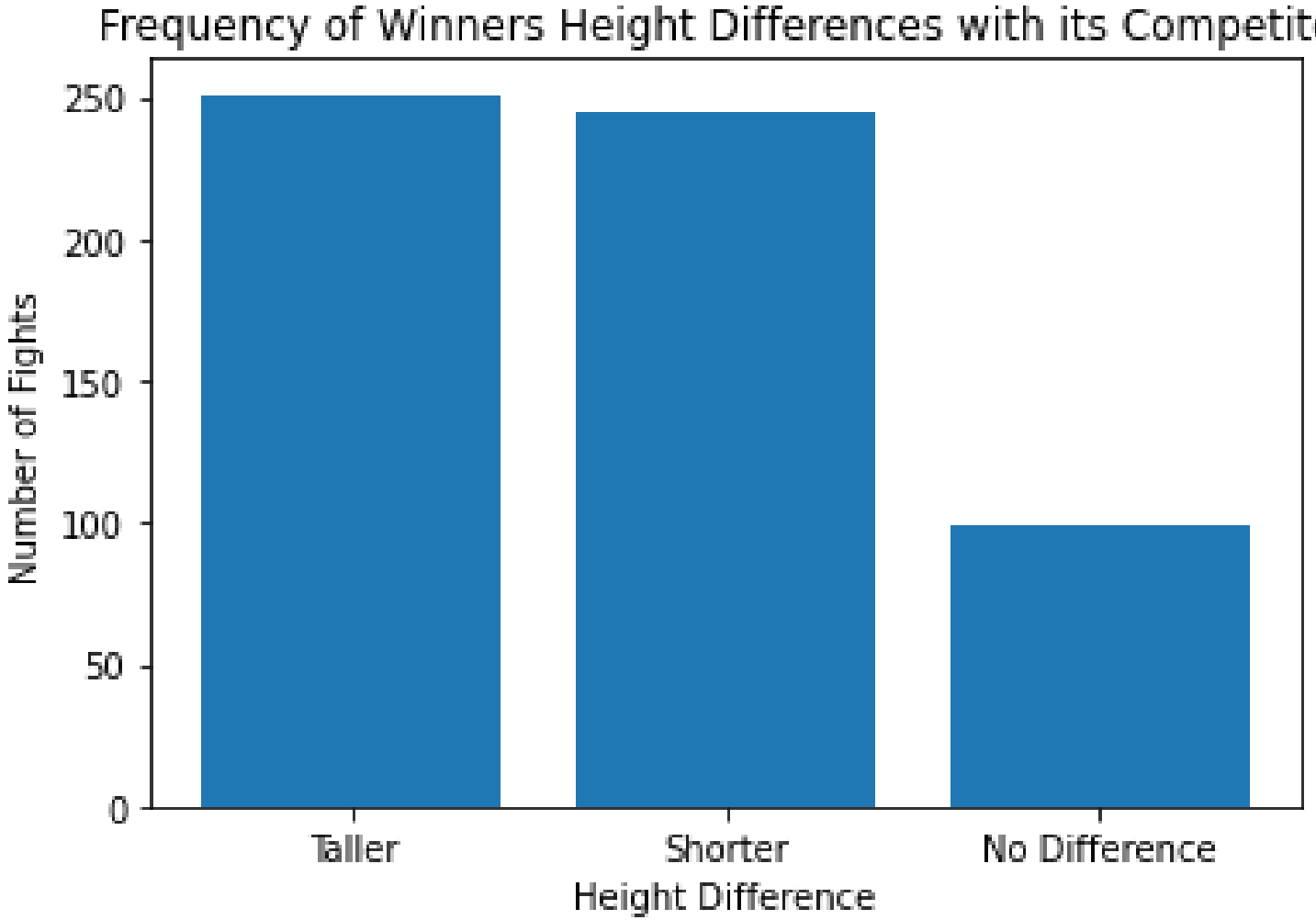
	~~Event Name~~	Event Date	Result	Fighter1	Fighter2	KD	Strikes	TD	Sub	Weight Class	Method	Round	Time
11	~~UFC 286: Edwards vs. Usman 3~~	2023-03-18	Justin Gaethje	Justin Gaethje	Rafael Fiziev	0-0	103-97	1-0	0-0	Lightweight	M-DEC	3	5.000000
16	~~UFC 286: Edwards vs. Usman 3~~	2023-03-18	Chris Duncan	Chris Duncan	Omar Morales	0-0	53-47	5-0	0-0	Lightweight	S-DEC	3	5.000000
17	~~UFC 286: Edwards vs. Usman 3~~	2023-03-18	Yanal Ashmouz	Yanal Ashmouz	Sam Patterson	0-0	15-2	1-0	0-0	Lightweight	KO/TKO-Punches	1	1.250000
23	~~UFC 286: Edwards vs. Usman 3~~	2023-03-18	~~Draw~~	Jai Herbert	Ludovit Klein	0-0	43-47	1-2	0-0	Lightweight	M-DEC	3	5.000000
41	~~UFC 285: Jones vs. Gane~~	2023-03-04	Mateusz Gamrot	Mateusz Gamrot	Jalin Turner	0-1	29-40	4-0	0-0	Lightweight	S-DEC	3	5.000000

Overview of Data – Summary Statistics
UFC Fighter Summary Statistics

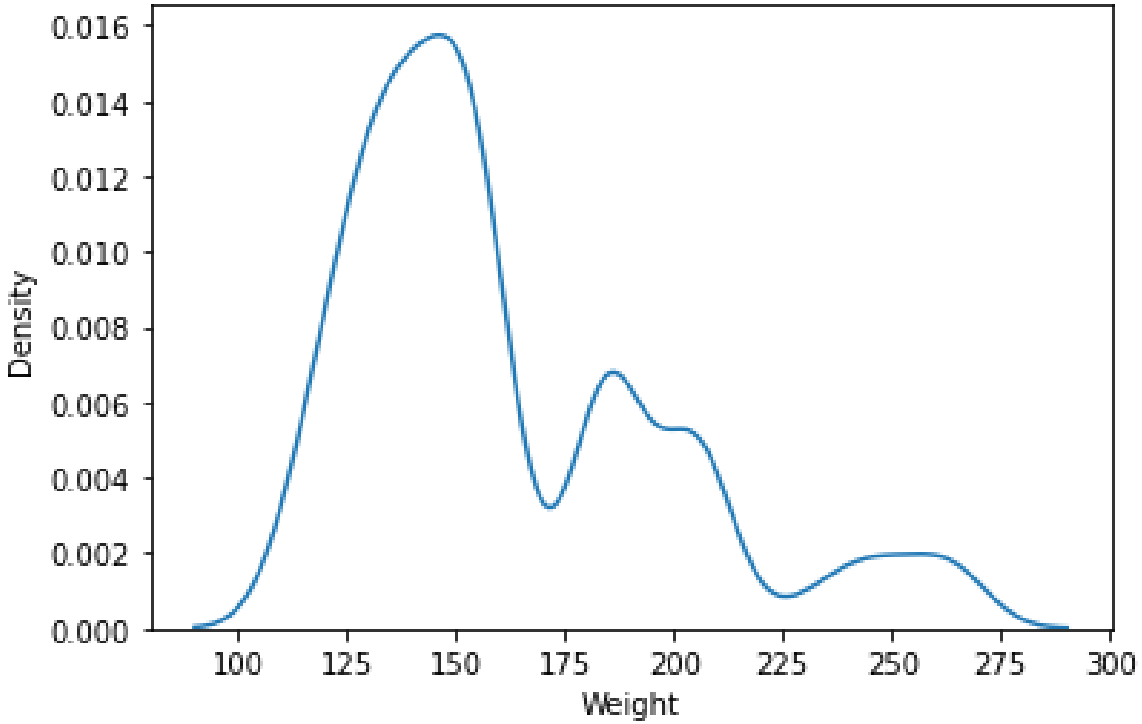


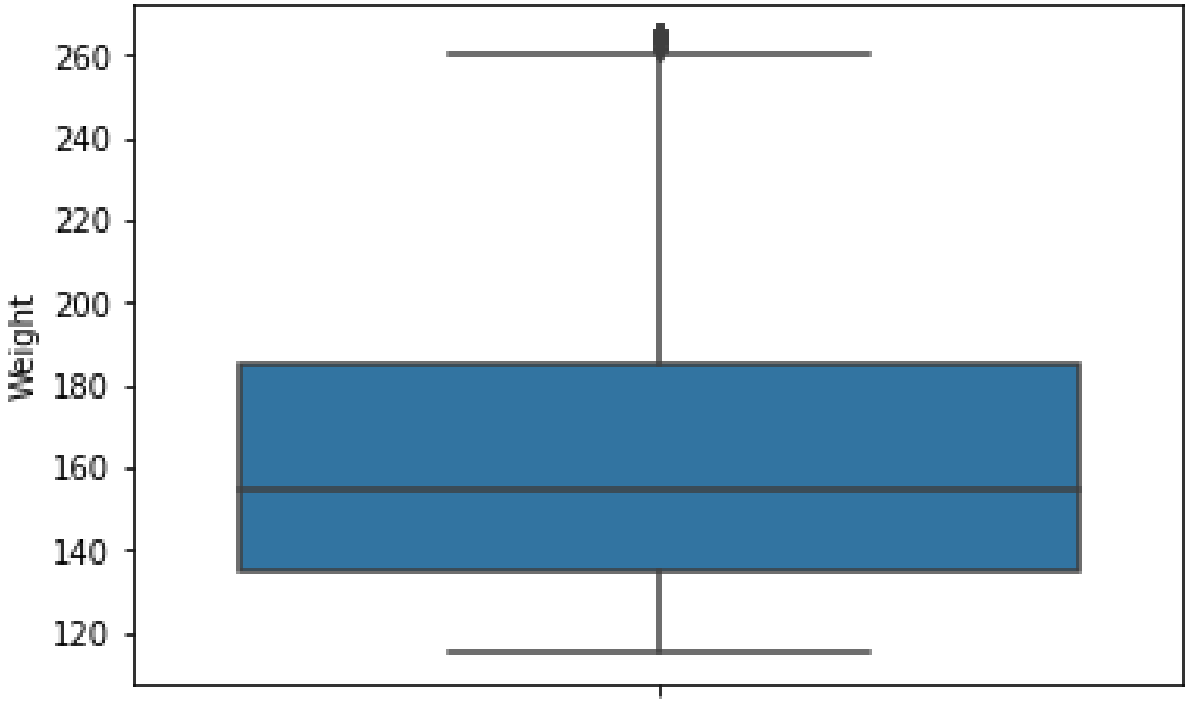


Overview of Data – Summary Statistics
UFC Fighter Summary Statistics

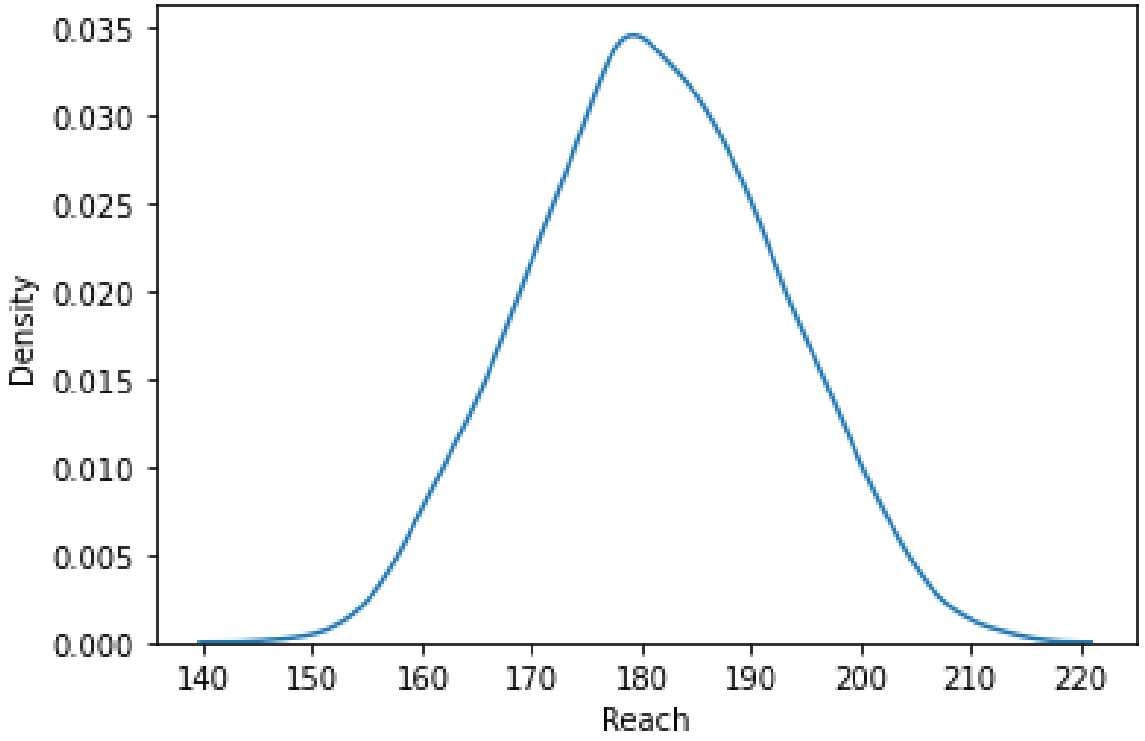


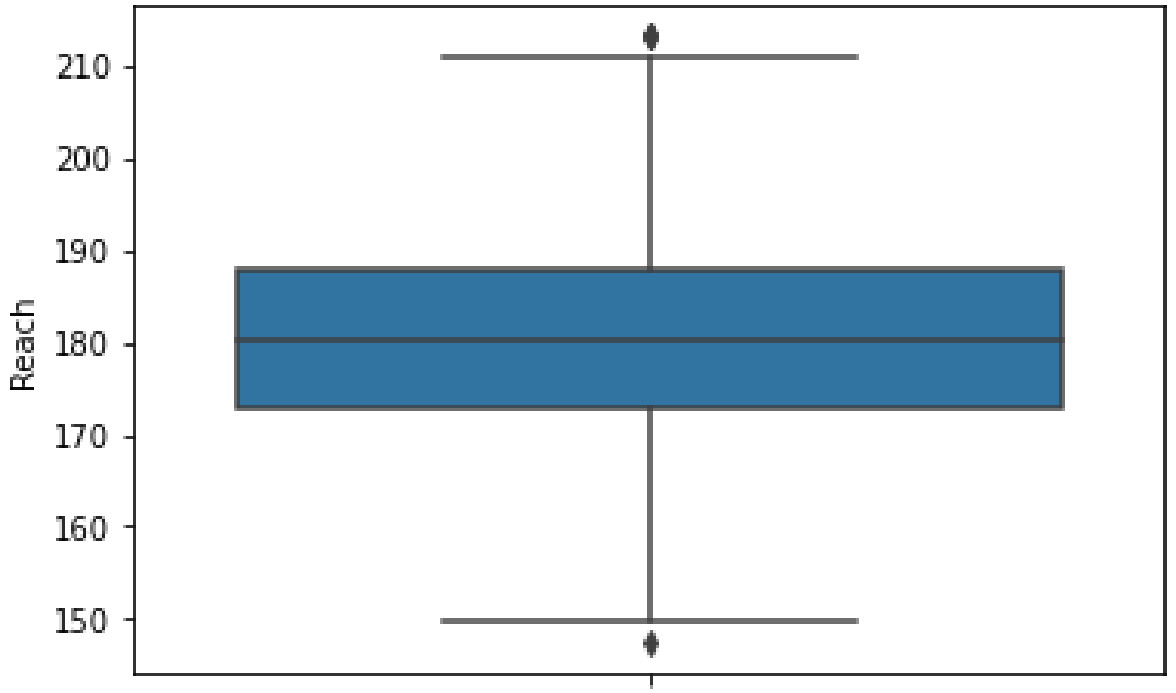
Overview of Data – Summary Statistics
UFC Fighter Summary Statistics



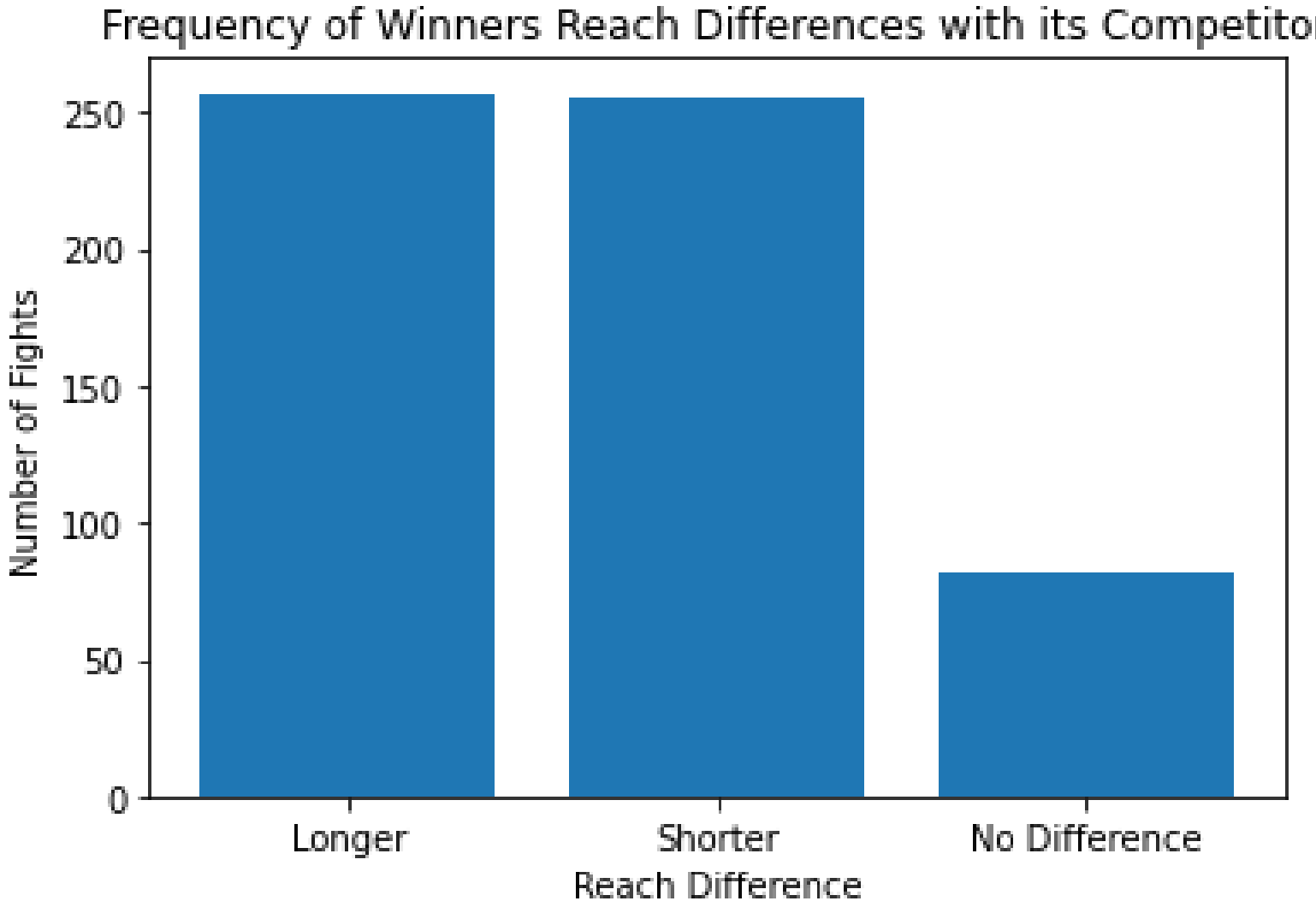


Overview of Data – Summary Statistics
UFC Fighter Summary Statistics

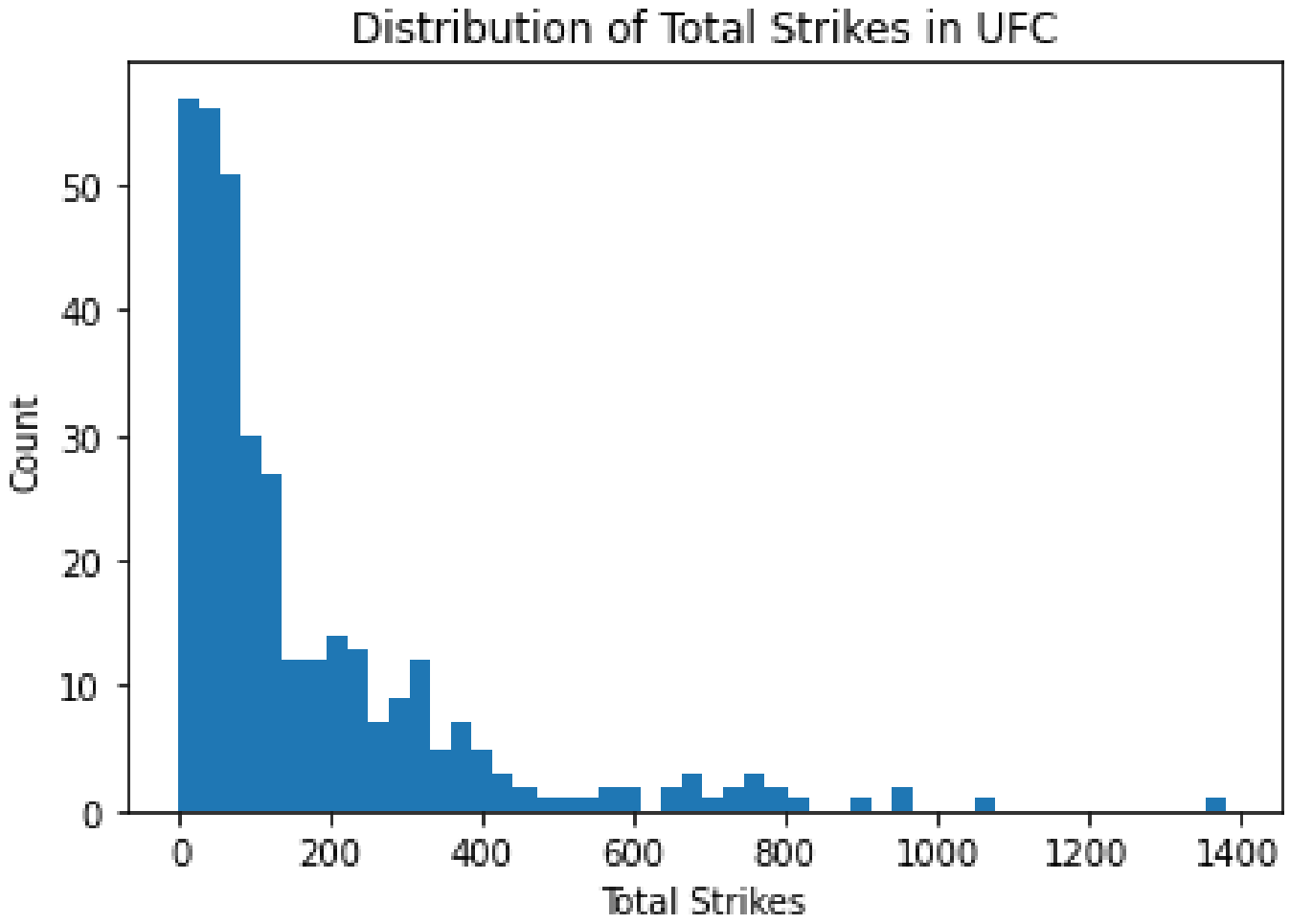




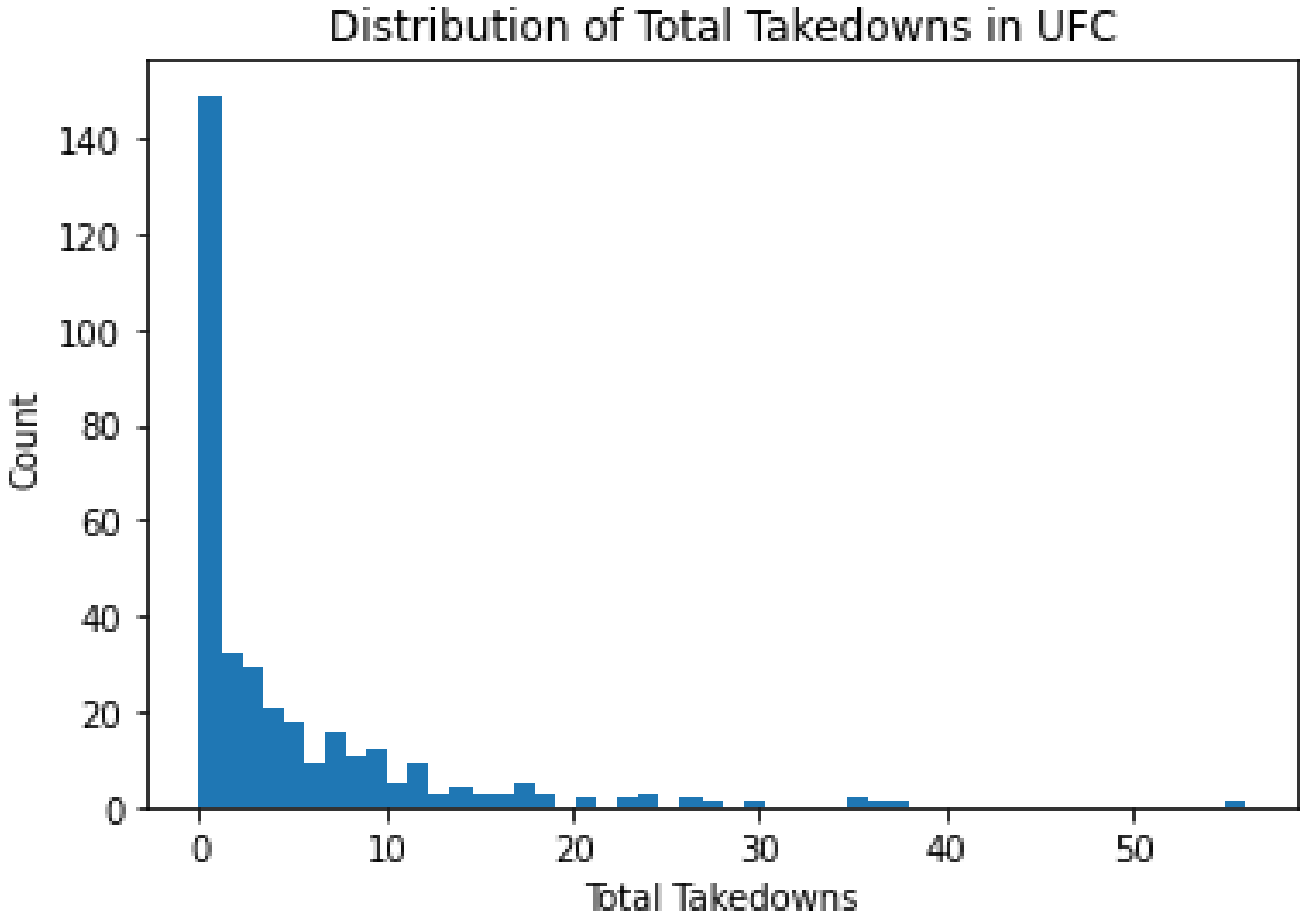
Overview of Data – Summary Statistics
UFC Fighter Summary Statistics



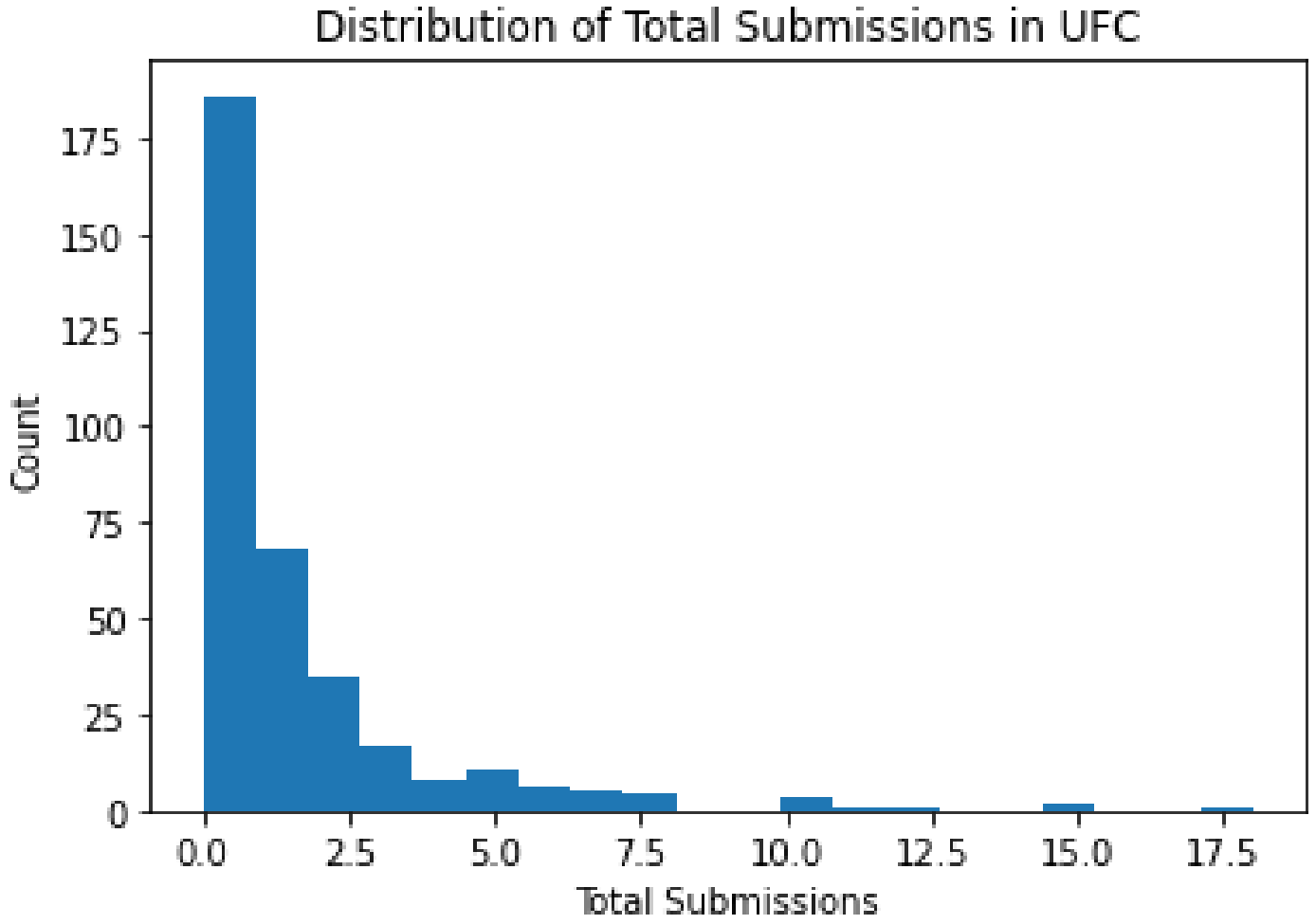
Overview of Data – Summary Statistics
UFC Summary Statistics



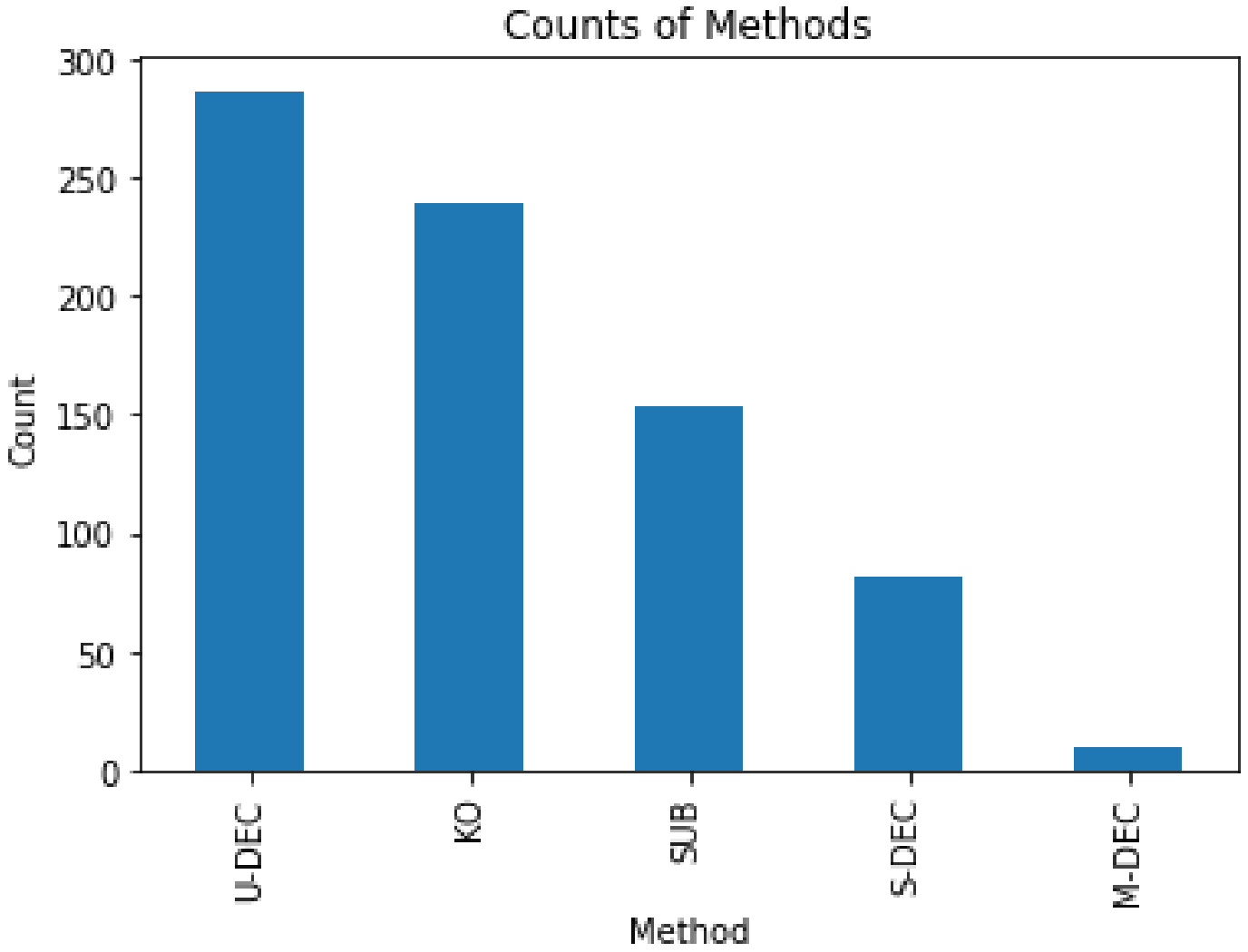
Overview of Data – Summary Statistics
UFC Summary Statistics



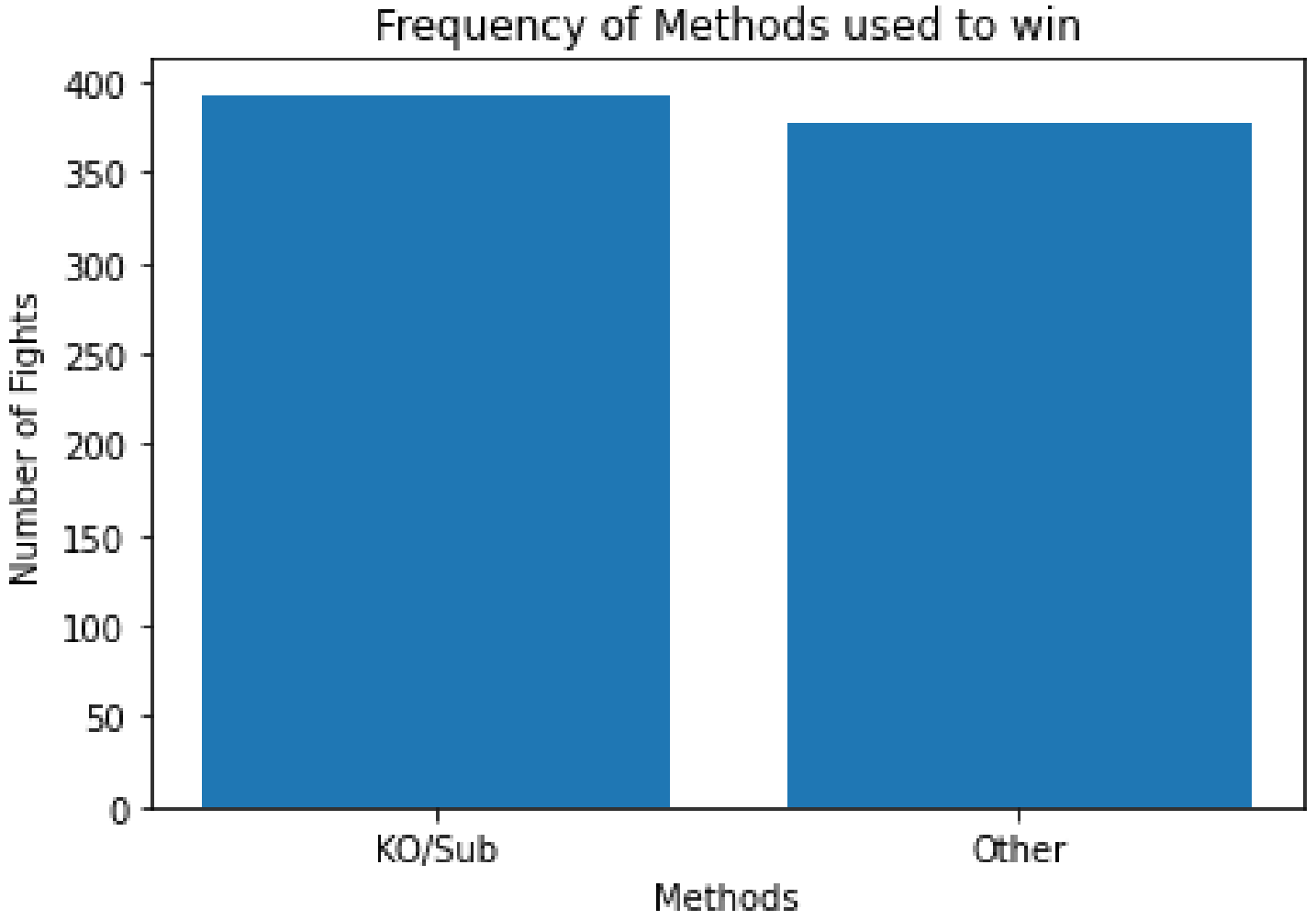
Overview of Data – Summary Statistics
UFC Summary Statistics



Overview of Data – Summary Statistics
UFC Summary Statistics



Overview of Data – Summary Statistics
UFC Summary Statistics



Feature Selection

Most relevant features for the Predictive Model

- Fighter1 and Fighter2 Match
  - Method of Win – **Predictor Variable**
    - KO/Sub vs Anything Else
  - Total **Strike Difference** between fighters (For the last 10 years)
  - Total **Takedown Difference**
  - Total **Submission Difference**
  - Total **Height Difference**
  - Total **Reach Difference**
  - Total # **Wins Difference**
  - Total **Experience Difference** (Number of fights in the last 10 years)
  - Average **Fight Time Difference**
  - Total # **KO wins Difference**

Most relevant features for the Predictive Model

	Method
Method	1.000000
strike_diff	-0.007704
td_diff	-0.027328
sub_diff	0.153585
Height_Diff	0.025767
Reach_Diff	0.045531
Wins_Diff	0.036651
Exp_Diff	0.029306
avg_fightti me_diff	-0.128979
ko_diff	0.058258

Most relevant features for the Predictive Model

	strike_diff	td_diff	sub_diff	Height_Diff	Reach_Diff	Wins_Diff	Exp_Diff	avg_fighttime_diff	ko_diff
0	323	-1	0	7.62	-2.54	12	14	1.365720	5
1	-122	4	0	-2.54	-5.08	-1	-4	-2.208333	-1
2	13	1	0	-15.24	-25.40	-3	-6	-0.000000	0
3	41	24	-5	-12.70	-17.78	9	5	-5.352381	-1
4	-26	11	-1	2.54	0.00	6	10	-0.000000	0
...	...	...	...	...	...	...	...	...	...
589	146	-1	0	-2.54	-7.62	-26	-41	-5.145833	-2
590	-373	-9	3	-7.62	-2.54	-12	-19	0.505303	-5
591	-40	-1	0	0.00	10.16	5	0	4.300000	-1
592	561	55	7	7.62	0.00	8	0	-9.335000	3
593	729	4	1	-2.54	7.62	3	11	-8.144608	6

Model Selection and Training

Different Models Used

- For the predictive modelling problem of classifying UFC fights as ending in a KO/Sub or by Decision, three popular binary classification algorithms were selected:
  - Logistic Regression
  - Support Vector Machine
  - Naïve Bayes (Gaussian)
- The performance of models was compared based on:
  - Accuracy Score
  - AUC - ROC
  - F1-Score
- The models were trained on historical fight data (the last 10 years)

Results and Analysis

Model Comparison

Model	Accuracy Score	AUC – ROC	F1-Score
Logistic Regression	67.2%	69%	67%
Support Vector Machine	69.7%	71%	67%
Naive Bayes	63.9%	67%	61%

Model Limitations

- Limited Data:
  - My data was only on the last 10 years
  - If there is few data of a certain fighter the model may not be able to make accurate predictions for those cases
- Combat Sport Unpredictability:
  - Combat sports are known for their unpredictability as it just takes one hit to knock someone out.
- Binary Classified:
  - In this project I only looked at whether it was a KO/Sub or Decision. However, it would be better to look at it as a multi-class problem.
- Interpretability:
  - SVM's (The model I chose as the best) are black box, therefore its difficult to interpret how the model is making its predictions

Conclusion

Conclusion

- In summary, I have demonstrated the power of predictive modelling in the context of UFC fights, showing how a combination of data pre-processing, feature engineering, and model selection can yield accurate predictions of fight outcomes.

- This approach has important implications not just for UFC but for other sports, where accurate predictions can inform everything from betting odds to training strategies.

Conclusion

- While the model has shown promising results, there is always room for improvement, further exploration and refinement of predictive models can be done to further improve the accuracy.

- By continuing to refine our models and better understand the factors that influence fight outcomes, we can help fighters, coaches, fans and businesses alike make more informed decisions and gain deeper insights into the complex world of combat sports.

Tested on recent Fight

FIGHT OVERVIEW | SIG. STRIKES BY TARGET | SIG. STRIKES BY POSITION

ROUND	TIME	METHOD
3	5:00	DECISION - UNANIMOUS

FULL FIGHT | ROUND 1 | ROUND 2 | ROUND 3

63 (25% of 252)	TOTAL STRIKES	143 (49% of 289)
0 (0% of 3)	TAKEDOWNS	0 (0% of 1)
0	SUB. ATTEMPTS	0
0	REV.	0
63 (25% of 252)	SIG. STRIKES	141 (49% of 287)
0	KNOCKDOWNS	0

WIN

UNITED STATES - ODDS - MEXICO

Tested on recent Fight

- Fighter1 = Clay Gaudia
- Fighter2 = Rafa Garcia
- Model Predicted that the result would be “0”
    - **Predicting that Fighter2 would win by Decision.**

Thank you!