In [155…
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

Loading the datasets

In [156…
```
df_fights = pd.read_csv('./dataset/ufc1.csv')
df_fighters = pd.read_csv('./dataset/ufc_fighters.csv')
```

Data Exploring and cleaning ufc fighters dataframe

In [157…
```
df_fighters.head()
```

Out[157…

	name	record	height	weight	reach	stance	dob	SLpM	StrAcc	SApM	StrDef
0	Tom Aaron	5-3-0	--	155 lbs.	--	NaN	Jul 13, 1978	0.00	0%	0.00	NaN
1	Danny Abbadi	4-6-0	5' 11"	155 lbs.	--	Orthodox	Jul 03, 1983	3.29	38%	4.41	NaN
2	Nariman Abbasov	28-4-0	5' 8"	155 lbs.	66"	Orthodox	Feb 01, 1994	3.00	20%	5.67	NaN
3	Darion Abbey	9-5-0	6' 2"	265 lbs.	80"	Orthodox	Feb 25, 1993	8.44	50%	14.06	NaN
4	David Abbott	10-15-0	6' 0"	265 lbs.	--	Switch	Apr 26, 1965	1.35	30%	3.55	NaN

In [158…
```
df_fighters.shape
```

Out[158…
```
(4449, 15)
```

In [159…
```
df_fighters.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4449 entries, 0 to 4448
Data columns (total 15 columns):
 #   Column  Non-Null Count  Dtype  
---  ------  --------------  -----  
 0   name    4449 non-null   object 
 1   record  4449 non-null   object 
 2   height  4449 non-null   object 
 3   weight  4449 non-null   object 
 4   reach   4449 non-null   object 
 5   stance  3601 non-null   object 
 6   dob     4449 non-null   object 
 7   SLpM    4449 non-null   float64
 8   StrAcc  4449 non-null   object 
 9   SApM    4449 non-null   float64
 10  StrDef  0 non-null      float64
 11  TDAvg   4449 non-null   float64
 12  TDAcc   4449 non-null   object 
 13  TDDef   4449 non-null   object 
 14  SubAvg  4449 non-null   float64
dtypes: float64(5), object(10)
memory usage: 521.5+ KB
```

In [160…
```
df_fighters.drop(columns='StrDef', inplace=True)
```

In [161…
```
df_fighters.head()
```

Out[161…

	name	record	height	weight	reach	stance	dob	SLpM	StrAcc	SApM	TDAvg
0	Tom Aaron	5-3-0	--	155 lbs.	--	NaN	Jul 13, 1978	0.00	0%	0.00	0.00
1	Danny Abbadi	4-6-0	5' 11"	155 lbs.	--	Orthodox	Jul 03, 1983	3.29	38%	4.41	0.00
2	Nariman Abbasov	28-4-0	5' 8"	155 lbs.	66"	Orthodox	Feb 01, 1994	3.00	20%	5.67	0.00
3	Darion Abbey	9-5-0	6' 2"	265 lbs.	80"	Orthodox	Feb 25, 1993	8.44	50%	14.06	0.00
4	David Abbott	10-15-0	6' 0"	265 lbs.	--	Switch	Apr 26, 1965	1.35	30%	3.55	1.07

In [162…
```
fighter_cols = df_fighters.columns.to_list()
```

In [163…
```
fighter_cols
```

Out[163…
```
['name',
 'record',
 'height',
 'weight',
 'reach',
 'stance',
 'dob',
 'SLpM',
 'StrAcc',
 'SApM',
 'TDAvg',
 'TDAcc',
 'TDDef',
 'SubAvg']
```

check for columns which are missing all important features

In [164…
```
indexes_to_drop = df_fighters.loc[(df_fighters["reach"] == "--")&(df_fighters["weig
```

In [165…
```
len(indexes_to_drop)
```

Out[165… 65

In [166…
```
df_fighters.drop(indexes_to_drop, inplace=True)
```

In [167…
```
df_fighters.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 4384 entries, 0 to 4448
Data columns (total 14 columns):
 #   Column  Non-Null Count  Dtype  
---  ------  --------------  -----  
 0   name    4384 non-null   object 
 1   record  4384 non-null   object 
 2   height  4384 non-null   object 
 3   weight  4384 non-null   object 
 4   reach   4384 non-null   object 
 5   stance  3601 non-null   object 
 6   dob     4384 non-null   object 
 7   SLpM    4384 non-null   float64
 8   StrAcc  4384 non-null   object 
 9   SApM    4384 non-null   float64
 10  TDAvg   4384 non-null   float64
 11  TDAcc   4384 non-null   object 
 12  TDDef   4384 non-null   object 
 13  SubAvg  4384 non-null   float64
dtypes: float64(4), object(10)
memory usage: 513.8+ KB
```

In [168…
```
indexes_to_drop2 = df_fighters.loc[(df_fighters["SLpM"]==0)&(df_fighters["StrAcc"]=
len(indexes_to_drop2)
```

Out[168… 633

In [169…
```
df_fighters.drop(indexes_to_drop2, inplace=True)
```

In [170…
```
df_fighters.shape
```

Out[170… (3751, 14)

In [171…
```
df_fighters.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 3751 entries, 1 to 4448
Data columns (total 14 columns):
 #   Column  Non-Null Count  Dtype  
---  ------  --------------  -----  
 0   name    3751 non-null   object 
 1   record  3751 non-null   object 
 2   height  3751 non-null   object 
 3   weight  3751 non-null   object 
 4   reach   3751 non-null   object 
 5   stance  3405 non-null   object 
 6   dob     3751 non-null   object 
 7   SLpM    3751 non-null   float64
 8   StrAcc  3751 non-null   object 
 9   SApM    3751 non-null   float64
 10  TDAvg   3751 non-null   float64
 11  TDAcc   3751 non-null   object 
 12  TDDef   3751 non-null   object 
 13  SubAvg  3751 non-null   float64
dtypes: float64(4), object(10)
memory usage: 439.6+ KB
```

In [172…
```
print("Null Value and zero counts for fighter metrics")
for i in fighter_cols:
    mask = (
        (df_fighters[i] == "--") |
        (df_fighters[i] == "0%") |
        (df_fighters[i] == 0) |
        (df_fighters[i].isna())
    )

    count = mask.sum()

    if count > 0:
        print(f"{i}: {count}")
```

```
Null Value and zero counts for fighter metrics
height: 72
weight: 12
reach: 1263
stance: 346
dob: 195
SLpM: 103
StrAcc: 103
SApM: 28
TDAvg: 1165
TDAcc: 1165
TDDef: 859
SubAvg: 1817
```

In [173…
```
df_fighters = df_fighters.reset_index(drop=True)
```

In [177…
```
df_fighters.head(20)
```

Out[177...

	name	record	height	weight	reach	stance	dob	SLpM	StrAcc	SApM
0	Danny Abbadi	4-6-0	5' 11"	155 lbs.	--	Orthodox	Jul 03, 1983	3.29	38%	4.41
1	Nariman Abbasov	28-4-0	5' 8"	155 lbs.	66"	Orthodox	Feb 01, 1994	3.00	20%	5.67
2	Darion Abbey	9-5-0	6' 2"	265 lbs.	80"	Orthodox	Feb 25, 1993	8.44	50%	14.06
3	David Abbott	10-15-0	6' 0"	265 lbs.	--	Switch	Apr 26, 1965	1.35	30%	3.55
4	Hamdy Abdelwahab	7-1-0 (1 NC)	6' 2"	265 lbs.	72"	Southpaw	Jan 22, 1993	4.27	55%	3.67
5	Mansur Abdul-Malik	8-0-1	6' 2"	185 lbs.	80"	Orthodox	Oct 07, 1997	4.27	48%	3.49
6	Shamil Abdurakhimov	20-8-0	6' 3"	235 lbs.	76"	Orthodox	Sep 02, 1981	2.41	44%	3.02
7	Hiroyuki Abe	8-15-3 (1 NC)	5' 6"	145 lbs.	--	Orthodox	Feb 09, 1970	1.71	36%	3.11
8	Daichi Abe	6-2-0	5' 11"	170 lbs.	71"	Orthodox	Nov 27, 1991	3.80	33%	4.49
9	Papy Abedi	10-4-0	5' 11"	185 lbs.	--	Southpaw	Jun 30, 1978	2.80	55%	3.15
10	Ricardo Abreu	5-1-0	5' 11"	185 lbs.	--	Orthodox	Apr 27, 1984	3.79	31%	3.98
11	Klidson Abreu	15-4-0 (1 NC)	6' 0"	205 lbs.	74"	Orthodox	Dec 24, 1992	2.05	40%	2.90
12	Daniel Acacio	30-18-0	5' 8"	180 lbs.	--	Orthodox	Dec 27, 1977	3.52	36%	2.85
13	John Adajar	6-2-0	5' 9"	170 lbs.	75"	Orthodox	Jun 22, 1991	3.90	52%	6.28

	name	record	height	weight	reach	stance	dob	SLpM	StrAcc	SApM
14	Juan Adams	5-3-0	6' 5"	265 lbs.	80"	Orthodox	Jan 16, 1992	7.09	55%	4.06
15	Anthony Adams	8-2-0	6' 1"	185 lbs.	76"	Orthodox	Jan 13, 1988	3.17	41%	5.93
16	Zarrukh Adashev	4-4-0	5' 5"	125 lbs.	65"	Southpaw	Jul 29, 1992	3.65	40%	3.04
17	Israel Adesanya	24-5-0	6' 4"	185 lbs.	80"	Switch	Jul 22, 1989	4.02	48%	3.20
18	Mohamed Ado	5-1-0	5' 11"	170 lbs.	76"	Switch	May 03, 2000	1.66	83%	0.33
19	Nick Agallar	24-6-0	5' 8"	155 lbs.	--	Orthodox	Jan 13, 1979	0.69	11%	4.56

In [178…

```
# check if there are fighters with the same name
df_fighters[df_fighters.duplicated(subset="name", keep=False)]
```

Out[178…

	name	record	height	weight	reach	stance	dob	SLpM	StrAcc	SApM	TI
1165	Joey Gomez	6-2-0	5' 10"	135 lbs.	73"	Orthodox	Jul 21, 1986	2.44	28%	4.46	
1167	Joey Gomez	7-1-0	5' 10"	155 lbs.	71"	Orthodox	Aug 29, 1989	3.73	49%	3.33	
1550	Tony Johnson	7-2-0	6' 2"	205 lbs.	76"	Orthodox	May 02, 1983	4.00	92%	3.67	
1558	Tony Johnson	11-3-0	6' 1"	265 lbs.	--	NaN	--	2.00	53%	4.73	
2094	Michael McDonald	1-1-0	5' 11"	205 lbs.	--	Orthodox	Feb 06, 1965	0.00	0%	0.40	
2095	Michael McDonald	17-4-0	5' 9"	135 lbs.	70"	Orthodox	Jan 15, 1991	2.69	42%	2.76	
3097	Jean Silva	19-12-3 (1 NC)	5' 6"	160 lbs.	--	Orthodox	Oct 08, 1977	0.73	22%	2.93	
3111	Bruno Silva	15-7-2 (1 NC)	5' 4"	125 lbs.	65"	Orthodox	Mar 16, 1990	3.82	50%	4.55	
3112	Bruno Silva	23-13-0	6' 0"	185 lbs.	74"	Orthodox	Jul 13, 1989	3.86	48%	5.35	
3122	Jean Silva	16-3-0	5' 7"	145 lbs.	69"	Orthodox	Dec 27, 1996	4.79	51%	4.69	
3462	Victor Valenzuela	13-6-2	5' 10"	155 lbs.	--	NaN	--	1.28	100%	2.55	
3464	Victor Valenzuela	12-4-0	5' 9"	170 lbs.	71"	Orthodox	Feb 09, 1994	3.36	33%	8.40	

In [186…

```
df_fighters.iloc[1167,0] = 'Joey Gomez 155'
df_fighters.iloc[1558,0] = 'Tony Johnson 265'
df_fighters.iloc[2095,0] = 'Michael McDonald 135'
df_fighters.iloc[3122,0] = 'Jean Silva 145'
df_fighters.iloc[3112,0] = 'Bruno Silva 185'
df_fighters.iloc[3464,0] = 'Victor Valenzuela 170'
```

In [188…

```
df_fighters[df_fighters['name']=='Joey Gomez 155']
```

Out[188…

	name	record	height	weight	reach	stance	dob	SLpM	StrAcc	SApM	TDAv
1167	Joey Gomez 155	7-1-0	5' 10"	155 lbs.	71"	Orthodox	Aug 29, 1989	3.73	49%	3.33	2

In [189…

```
# check if there are fighters with the same name
df_fighters[df_fighters.duplicated(subset="name", keep=False)]
```

Out[189…

	name	record	height	weight	reach	stance	dob	SLpM	StrAcc	SApM	TDAvg	TDAc

In [191…

```
df_fighters.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3751 entries, 0 to 3750
Data columns (total 14 columns):
 #   Column  Non-Null Count  Dtype  
---  ------  --------------  -----  
 0   name    3751 non-null   object 
 1   record  3751 non-null   object 
 2   height  3751 non-null   object 
 3   weight  3751 non-null   object 
 4   reach   3751 non-null   object 
 5   stance  3405 non-null   object 
 6   dob     3751 non-null   object 
 7   SLpM    3751 non-null   float64
 8   StrAcc  3751 non-null   object 
 9   SApM    3751 non-null   float64
 10  TDAvg   3751 non-null   float64
 11  TDAcc   3751 non-null   object 
 12  TDDef   3751 non-null   object 
 13  SubAvg  3751 non-null   float64
dtypes: float64(4), object(10)
memory usage: 410.4+ KB
```

In [193…

```
print("Null Value and zero counts for fighter metrics")
for i in fighter_cols:
    mask = (
        (df_fighters[i] == "--")

    )

    count = mask.sum()

    if count > 0:
        print(f"{i}: {count}")
```

```
Null Value and zero counts for fighter metrics
height: 72
weight: 12
reach: 1263
dob: 195
```

In [203…
```
df_fighters.loc[(df_fighters["height"] == "--") & (df_fighters["weight"] == "--")
```

Out[203…

	name	record	height	weight	reach	stance	dob	SLpM	StrAcc	SApM
580	Nikolajus Cilkinas	2-6-0	--	--	--	Orthodox	--	0.00	0%	5.88
771	John Devine	6-4-0	--	--	--	NaN	Jan 17, 1978	1.52	18%	2.28
1657	Aurelijus Kerpe	11-22-0	--	--	--	Orthodox	--	1.26	16%	8.21
1822	Imani Lee	2-4-0	--	--	--	Orthodox	Jul 17, 1977	0.00	0%	0.94
1846	Michael Lerma	0-1-0	--	--	--	NaN	Nov 18, 1973	0.00	0%	6.47
1865	Sam Liera	12-10-0	--	--	--	NaN	Nov 26, 1983	0.82	60%	7.36
1878	Miguel Linares	2-4-0	--	--	--	NaN	--	6.51	50%	8.84
1975	Lolohea Mahe	10-4-1	--	--	--	NaN	--	1.71	29%	6.92
2080	Bobby McAndrews	1-3-0	--	--	--	Orthodox	--	2.56	83%	1.03
2445	Masakatsu Okuda	8-5-0	--	--	--	NaN	Jan 16, 1976	0.00	0%	4.27
3328	Akihito Tanaka	2-2-0	--	--	--	Southpaw	Mar 09, 1983	0.19	50%	2.05

In [207…
```
indexes_to_drop3  = df_fighters.loc[(df_fighters["height"] == "--") & (df_fighters[
```

In [208…
```
df_fighters.drop(indexes_to_drop3, inplace=True)
```

In [210…
```
print("Null Value and zero counts for fighter metrics")
for i in fighter_cols:
    mask = (
        (df_fighters[i] == "--")

    )

    count = mask.sum()
```

```
    if count > 0:
        print(f"{i}: {count}")
```

```
Null Value and zero counts for fighter metrics
height: 61
weight: 1
reach: 1252
dob: 190
```

In [213…
```
df_fighters.drop(df_fighters.loc[ (df_fighters["weight"] == "--") ].index, inplace=
```

In [214…
```
print("Null Value and zero counts for fighter metrics")
for i in fighter_cols:
    mask = (
        (df_fighters[i] == "--")

    )

    count = mask.sum()

    if count > 0:
        print(f"{i}: {count}")
```

```
Null Value and zero counts for fighter metrics
height: 61
reach: 1251
dob: 189
```

In [ ]:
```

```