Problem Statement

This dataset utilizes data from 2014 Major League Baseball seasons in order to develop an algorithm that predicts the number of wins for a given team in the 2015 season based on several different indicators of success. There are 16 different features that will be used as the inputs to the machine learning and the output will be a value that represents the number of wins.

-- Input features: Runs, At Bats, Hits, Doubles, Triples, Homeruns, Walks, Strikeouts, Stolen Bases, Runs Allowed, Earned Runs, Earned Run Average (ERA), Shutouts, Saves, Complete Games and Errors

-- Output: Number of predicted wins (W)

In [1]:

```
#Importing the required library for EDA
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

In [3]:

```
df = pd.read_csv('https://raw.githubusercontent.com/dsrscientist/Data-Science-ML-Capstone-Projects/master/basebal
```

In [4]:

```
df.head()#view my first 5 records
```

Out[4]:

	W	R	AB	H	2B	3B	HR	BB	SO	SB	RA	ER	ERA	CG	SHO	SV	E
0	95	724	5575	1497	300	42	139	383	973	104	641	601	3.73	2	8	56	88
1	83	696	5467	1349	277	44	156	439	1264	70	700	653	4.07	2	12	45	86
2	81	669	5439	1395	303	29	141	533	1157	86	640	584	3.67	11	10	38	79
3	76	622	5533	1381	260	27	136	404	1231	68	701	643	3.98	7	9	37	101
4	74	689	5605	1515	289	49	151	455	1259	83	803	746	4.64	7	12	35	86

In [5]:

```
new_column_list = ["Wins","Runs","At Bats","Hits","Doubles","Triples","Homeruns","Walks","Strikeouts","Stolen Bas
```

In [6]:

```
df.set_axis(new_column_list,axis=1,inplace=True)#replaced column name with new name from list
```

In [7]:

```
df.columns
```

Out[7]:

```
Index(['Wins', 'Runs', 'At Bats', 'Hits', 'Doubles', 'Triples', 'Homeruns',
       'Walks', 'Strikeouts', 'Stolen Bases', 'Runs Allowed', 'Earned Runs',
       'Earned Run Average (ERA)', 'Complete Games', 'Shutouts', 'Saves',
       'Errors'],
      dtype='object')
```

In [8]:

```
df.columns
```

Out[8]:

```
Index(['Wins', 'Runs', 'At Bats', 'Hits', 'Doubles', 'Triples', 'Homeruns',
       'Walks', 'Strikeouts', 'Stolen Bases', 'Runs Allowed', 'Earned Runs',
       'Earned Run Average (ERA)', 'Complete Games', 'Shutouts', 'Saves',
       'Errors'],
      dtype='object')
```

In [9]:

```
df.describe()#statisical descripton for my dataframe here i see some columns has outliers which will be treated i
```

Out[9]:

	Wins	Runs	At Bats	Hits	Doubles	Triples	Homeruns	Walks	Strikeouts	Stolen Bases	Runs Allowed	
count	30.000000	30.000000	30.000000	30.000000	30.000000	30.000000	30.000000	30.000000	30.00000	30.000000	30.000000	
mean	80.966667	688.233333	5516.266667	1403.533333	274.733333	31.300000	163.633333	469.100000	1248.20000	83.500000	688.233333	6
std	10.453455	58.761754	70.467372	57.140923	18.095405	10.452355	31.823309	57.053725	103.75947	22.815225	72.108005	
min	63.000000	573.000000	5385.000000	1324.000000	236.000000	13.000000	100.000000	375.000000	973.00000	44.000000	525.000000	4

25%	74.000000	651.250000	5464.000000	1363.000000	262.250000	23.000000	140.250000	428.250000	1157.50000	69.000000	636.250000	5
50%	81.000000	689.000000	5510.000000	1382.500000	275.500000	31.000000	158.500000	473.000000	1261.50000	83.500000	695.500000	6
75%	87.750000	718.250000	5570.000000	1451.500000	288.750000	39.000000	177.000000	501.250000	1311.50000	96.500000	732.500000	6
max	100.000000	891.000000	5649.000000	1515.000000	308.000000	49.000000	232.000000	570.000000	1518.00000	134.000000	844.000000	7

In [10]:

```
df.info()#info of my dataframe which says if any null values along with datatype here all are numeric
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 30 entries, 0 to 29
Data columns (total 17 columns):
 #   Column                    Non-Null Count  Dtype  
---  ------                    --------------  -----  
 0   Wins                      30 non-null     int64  
 1   Runs                      30 non-null     int64  
 2   At Bats                   30 non-null     int64  
 3   Hits                      30 non-null     int64  
 4   Doubles                   30 non-null     int64  
 5   Triples                   30 non-null     int64  
 6   Homeruns                  30 non-null     int64  
 7   Walks                     30 non-null     int64  
 8   Strikeouts                30 non-null     int64  
 9   Stolen Bases              30 non-null     int64  
 10  Runs Allowed              30 non-null     int64  
 11  Earned Runs               30 non-null     int64  
 12  Earned Run Average (ERA)  30 non-null     float64
 13  Complete Games            30 non-null     int64  
 14  Shutouts                  30 non-null     int64  
 15  Saves                     30 non-null     int64  
 16  Errors                    30 non-null     int64  
dtypes: float64(1), int64(16)
memory usage: 4.1 KB
```

EDA

Checking for Null values present in my dataframe

In [11]:

```
df.isnull()#checking null values
```

Out[11]:

	Wins	Runs	At Bats	Hits	Doubles	Triples	Homeruns	Walks	Strikeouts	Stolen Bases	Runs Allowed	Earned Runs	Earned Run Average (ERA)	Complete Games	Shutouts	Saves
0	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
5	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
6	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
7	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
8	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
9	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
10	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
11	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
12	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
13	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
14	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
15	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
16	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
17	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
18	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
19	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False

20	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
21	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
22	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
23	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
24	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
25	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
26	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
27	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
28	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
29	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False

In [12]:

```
df.isnull().sum()#count of my nullvalues
```

Out[12]:

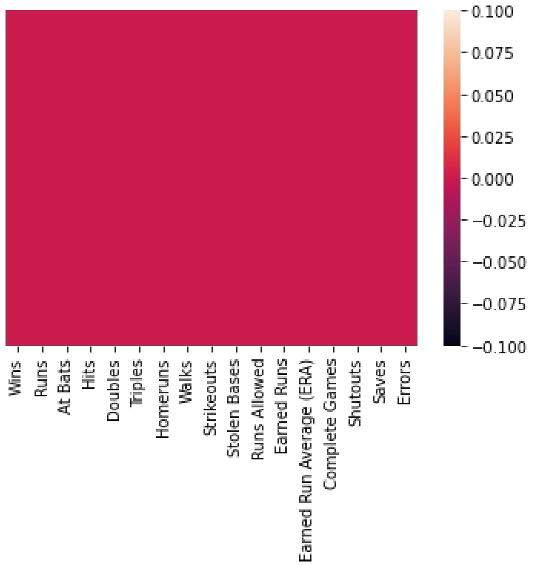
```
Wins                           0
Runs                           0
At Bats                        0
Hits                           0
Doubles                        0
Triples                        0
Homeruns                       0
Walks                          0
Strikeouts                     0
Stolen Bases                   0
Runs Allowed                   0
Earned Runs                    0
Earned Run Average (ERA)       0
Complete Games                 0
Shutouts                       0
Saves                          0
Errors                         0
dtype: int64
```

In [13]:

```
sns.heatmap(df.isnull(),yticklabels=False) #visualising null values if any
```

Out[13]:

```
<AxesSubplot:>
```

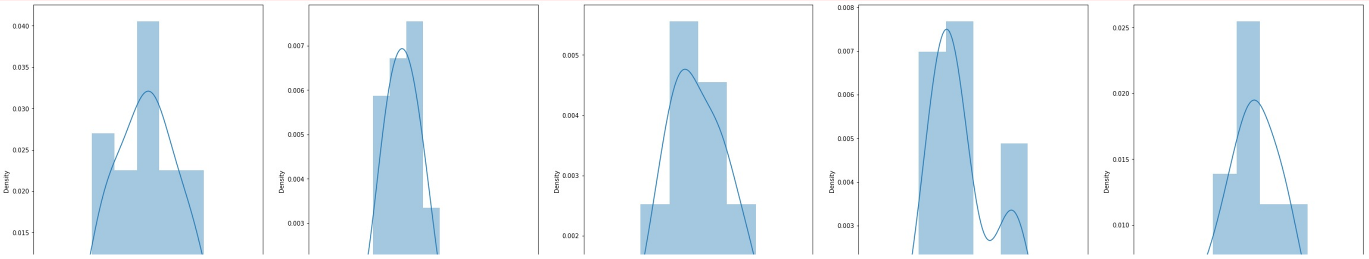


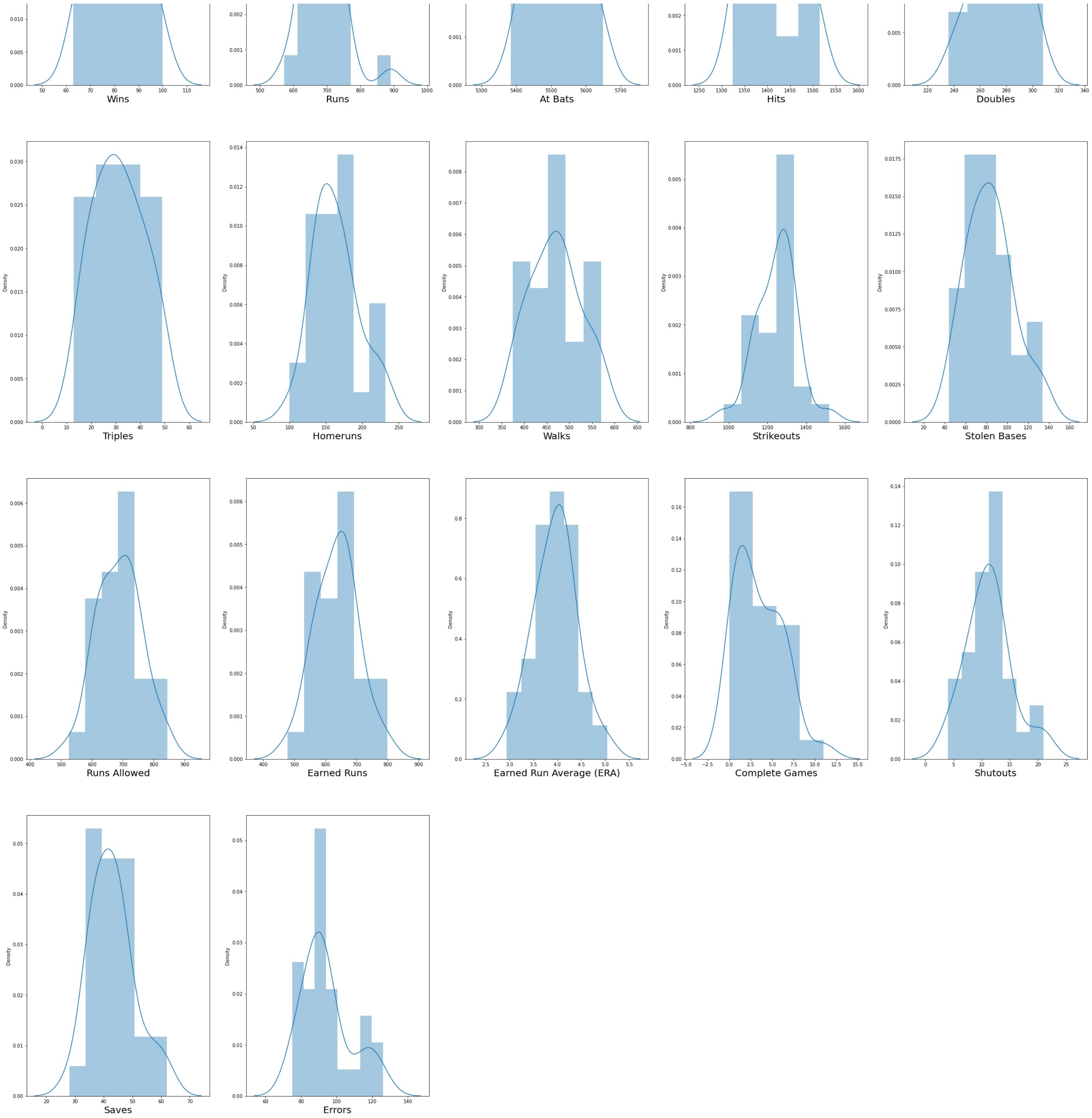
I see no null values here and there are no classification features in my dataset so lets check for outliers by using distplot and boxplot as below for all my columns.

In [14]:

```
plt.figure(figsize=(40,50),facecolor='white')
plotnumber = 1

for column in df:
    if plotnumber<=17:
            ax = plt.subplot(4,5,plotnumber)
            sns.distplot(df[column])
            plt.xlabel(column,fontsize=20)

    plotnumber +=1
plt.show()
```

```
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\distributions.py:2557: FutureWarning: `distplot` is a deprecat
ed function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-lev
el function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\distributions.py:2557: FutureWarning: `distplot` is a deprecat
ed function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-lev
el function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\distributions.py:2557: FutureWarning: `distplot` is a deprecat
ed function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-lev
el function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\distributions.py:2557: FutureWarning: `distplot` is a deprecat
ed function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-lev
el function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\distributions.py:2557: FutureWarning: `distplot` is a deprecat
ed function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-lev
el function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\distributions.py:2557: FutureWarning: `distplot` is a deprecat
ed function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-lev
el function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\distributions.py:2557: FutureWarning: `distplot` is a deprecat
ed function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-lev
el function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\distributions.py:2557: FutureWarning: `distplot` is a deprecat
ed function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-lev
el function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\distributions.py:2557: FutureWarning: `distplot` is a deprecat
ed function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-lev
el function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\distributions.py:2557: FutureWarning: `distplot` is a deprecat
ed function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-lev
el function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\distributions.py:2557: FutureWarning: `distplot` is a deprecat
ed function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-lev
el function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\distributions.py:2557: FutureWarning: `distplot` is a deprecat
ed function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-lev
el function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\distributions.py:2557: FutureWarning: `distplot` is a deprecat
ed function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-lev
el function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\distributions.py:2557: FutureWarning: `distplot` is a deprecat
ed function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-lev
el function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\distributions.py:2557: FutureWarning: `distplot` is a deprecat
ed function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-lev
el function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\distributions.py:2557: FutureWarning: `distplot` is a deprecat
ed function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-lev
el function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\distributions.py:2557: FutureWarning: `distplot` is a deprecat
ed function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-lev
el function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```



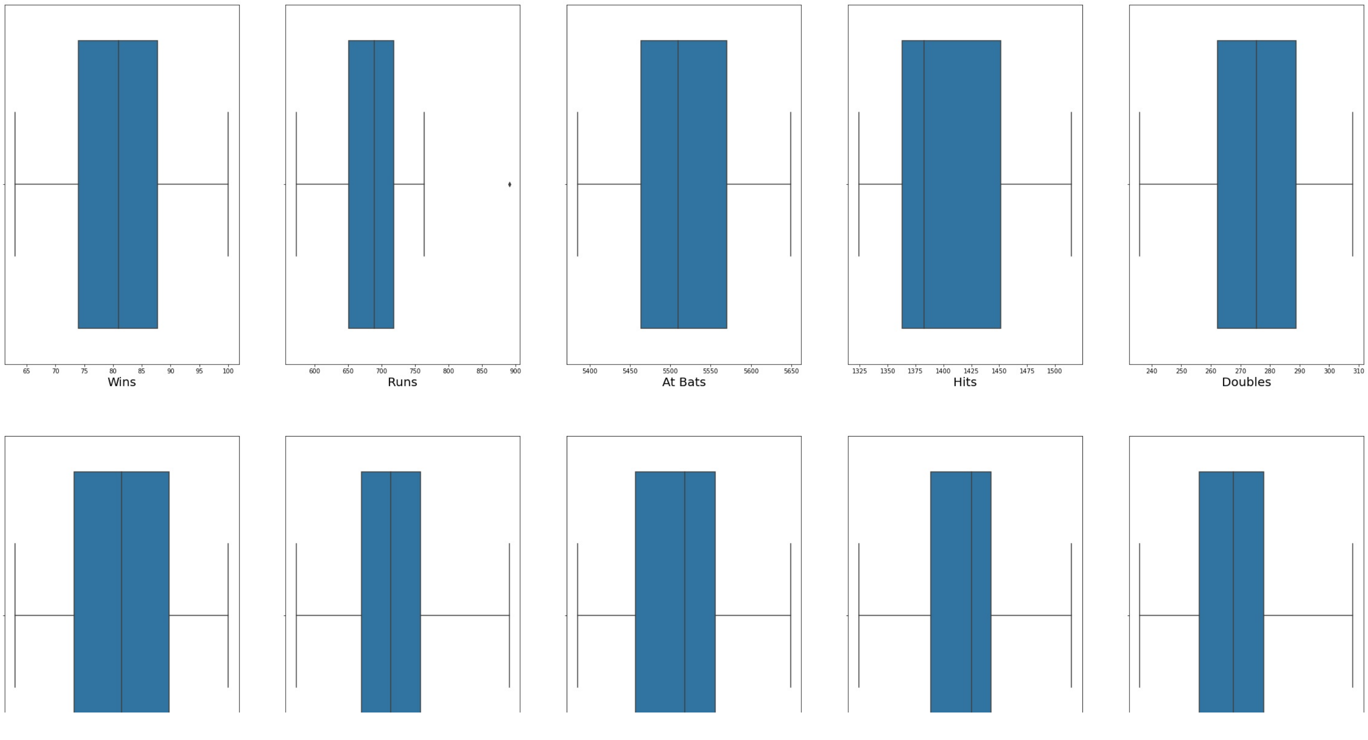


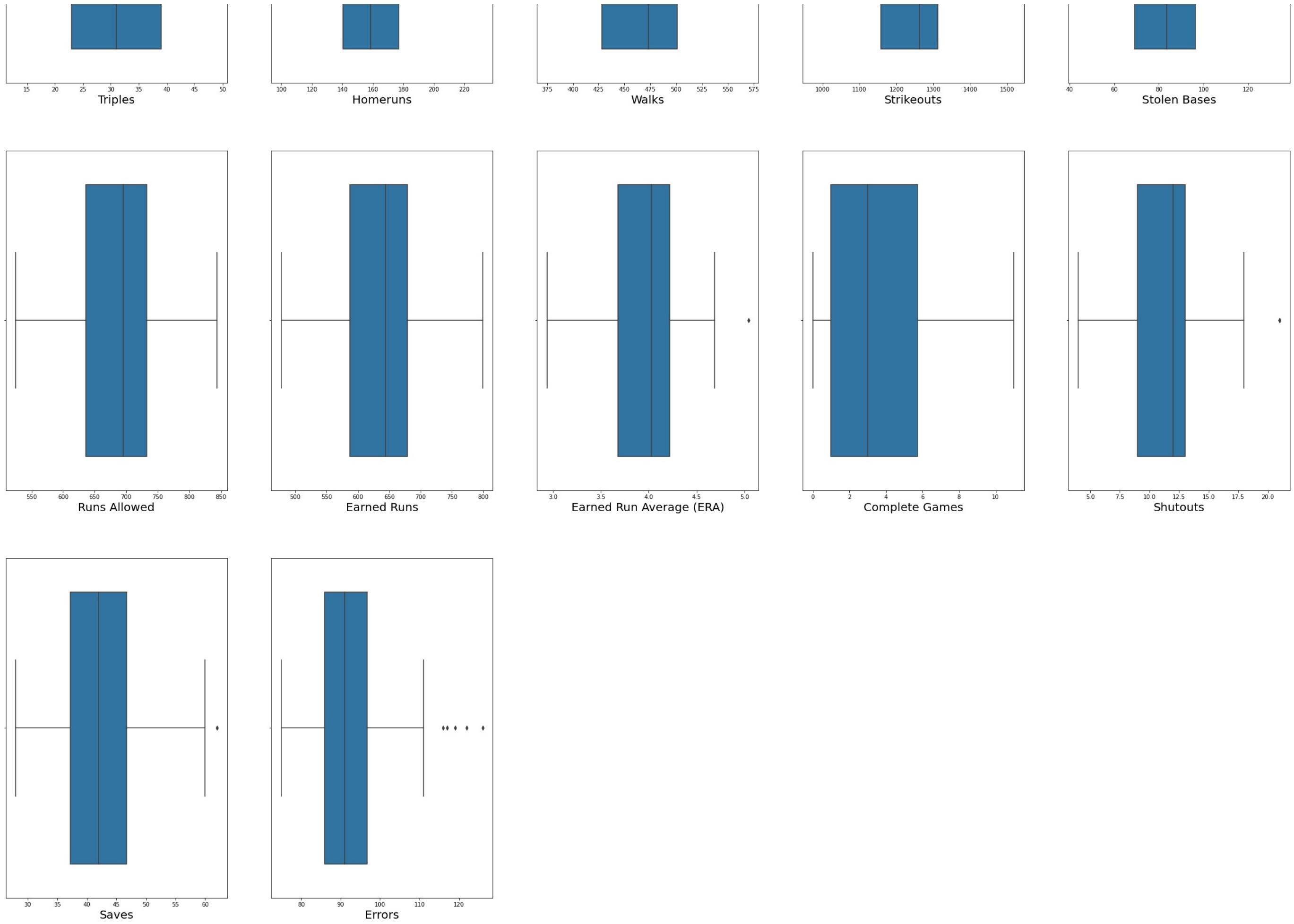
In [15]:

```
plt.figure(figsize=(40,50),facecolor='white')
plotnumber = 1

for column in df:
    if plotnumber<=17:
            ax = plt.subplot(4,5,plotnumber)
            sns.boxplot(df[column])
            plt.xlabel(column,fontsize=20)

    plotnumber +=1
plt.show()
```

```
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable
as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other argu
ments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable
as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other argu
ments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable
as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other argu
ments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable
```

```
as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other argu
ments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable
as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other argu
ments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable
as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other argu
ments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable
as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other argu
ments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable
as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other argu
ments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable
as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other argu
ments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable
as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other argu
ments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable
as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other argu
ments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable
as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other argu
ments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable
as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other argu
ments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable
as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other argu
ments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable
as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other argu
ments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable
as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other argu
ments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable
as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other argu
ments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
```





In [16]:

```
#checking for Z score to remove outliers
from scipy import stats

z = np.abs(stats.zscore(df))

print(z)
```

```
[[1.36540860e+00 6.19077968e-01 8.47731264e-01 1.66368512e+00
  1.42017307e+00 1.04119304e+00 7.87298598e-01 1.53490242e+00
  2.69762957e+00 9.13883291e-01 6.66233927e-01 5.05110079e-01
  5.06955186e-01 5.39806195e-01 8.14628593e-01 1.67160651e+00
  4.61469635e-01]
 [1.97838300e-01 1.34431656e-01 7.11093535e-01 9.70680764e-01
  1.27403389e-01 1.23580856e+00 2.43966683e-01 5.36591904e-01
  1.54878442e-01 6.01825582e-01 1.65970766e-01 2.48929848e-01
  2.54597523e-01 5.39806195e-01 1.72800005e-01 2.49879323e-01
  6.07196888e-01]
 [3.24325082e-03 3.32905860e-01 1.11523330e+00 1.51891367e-01
  1.58879521e+00 2.23807850e-01 7.23377196e-01 1.13914361e+00
  8.93981893e-01 1.11449182e-01 6.80339091e-01 7.51623132e-01
  6.41346840e-01 2.77264091e+00 3.20914294e-01 6.54856158e-01
  1.11724227e+00]
 [4.83244373e-01 1.14641931e+00 2.41521620e-01 4.01088140e-01
  8.28122030e-01 4.18423371e-01 8.83180700e-01 1.16053598e+00
  1.68601848e-01 6.90984928e-01 1.80075931e-01 1.03922169e-01
  5.30100415e-02 1.30044220e+00 5.67771443e-01 7.84104084e-01
  4.85757510e-01]
 [6.77839422e-01 1.32700776e-02 1.28073815e+00 1.98408098e+00
  8.01891920e-01 1.72234737e+00 4.03770187e-01 2.51360327e-01
  1.05866277e-01 2.22898364e-02 1.61880269e+00 1.59750126e+00
  1.53131824e+00 1.30044220e+00 1.72800005e-01 1.04259994e+00
  6.07196888e-01]
 [1.17081355e+00 3.50964704e+00 1.04883891e-01 1.36108904e+00
  1.86983209e+00 1.39150098e+00 2.18504658e+00 1.79874163e+00
  9.52796491e-01 2.00608527e-01 2.57184162e-01 3.89103937e-01
  3.50164922e-01 1.30044220e+00 3.20914294e-01 1.17184786e+00
  4.61469635e-01]
 [5.87028399e-01 1.31142984e+00 7.32262760e-01 1.16291827e-01
  1.53633499e-01 1.19688546e+00 1.54583256e+00 1.51351006e+00
  2.07811580e-01 9.13883291e-01 1.37760438e-01 2.34429080e-01
  1.65003087e-01 1.71756517e-01 1.80205719e+00 6.37623101e-01
  9.71515020e-02]
 [3.24325082e-03 4.28681202e-01 4.51289401e-01 5.96885604e-01
```

```
  1.61502532e+00 1.09957770e+00 1.70563607e+00 9.10958349e-01
  8.11641456e-01 1.76089707e+00 6.72346164e-02 1.47424473e-01
  2.09800305e-01 1.27590555e+00 3.20914294e-01 8.61652840e-03
  1.26296953e+00]
 [9.40542739e-02 7.65625781e-01 4.51289401e-01 3.65488601e-01
  1.83610767e-01 6.81154325e-02 1.07601026e-01 5.90072825e-01
  6.05790362e-01 1.56028855e-01 6.52128762e-01 4.61607776e-01
  4.84556577e-01 9.07855874e-01 1.72800005e-01 2.18859821e+00
  4.85757510e-02]
 [2.88649323e-01 1.03448909e+00 1.78591286e+00 1.62808558e+00
  1.08292881e+00 1.65423193e-01 8.41631789e-02 1.58660065e-01
  9.82203790e-01 5.57245909e-01 9.13544474e-01 8.43461329e-01
  7.92164142e-01 1.71756517e-01 3.20914294e-01 3.96360306e-01
  1.94303004e-01]
 [6.84325924e-01 1.08641548e+00 7.60167649e-02 2.75303102e-01
  2.39818144e-01 6.81154325e-02 2.67404530e-01 6.04334404e-01
  1.48996982e-01 7.80144273e-01 6.31441188e-01 6.40450579e-01
  6.35373878e-01 5.64342841e-01 5.67771443e-01 2.49879323e-01
  1.79730279e+00]
 [4.89730874e-01 7.05621952e-01 8.26562038e-01 7.21483991e-01
  1.83610767e-01 5.15731132e-01 2.12112518e+00 3.01275853e-01
  1.40958987e+00 1.67173773e+00 9.90652705e-01 9.25632346e-01
  8.65332931e-01 5.64342841e-01 4.19657154e-01 5.25608232e-01
  6.80060514e-01]
 [3.92433350e-01 4.71376235e-01 1.43277168e+00 1.29107662e+00
  1.78364745e+00 1.00226994e+00 3.95247334e-01 6.07899799e-01
  9.62598924e-01 1.40425969e+00 1.86658341e-01 8.45878123e-02
  3.65843948e-02 5.39806195e-01 1.72800005e-01 3.79127249e-01
  9.71515020e-02]
 [4.83244373e-01 5.57920219e-01 4.00290812e-01 4.36687679e-01
  7.15707275e-01 9.04962175e-01 1.09838275e+00 1.58660065e-01
  8.60653621e-01 6.46405255e-01 5.32705038e-01 5.96948276e-01
  4.56185005e-01 9.32392519e-01 1.72800005e-01 2.49879323e-01
  2.42878755e-02]
 [1.26162457e+00 9.98140619e-02 1.20857034e+00 2.61063286e-02
  1.27403389e-01 1.43042408e+00 5.63573692e-01 1.05179144e-01
  1.26647435e+00 2.45188200e-01 5.75020531e-01 4.08438294e-01
  4.11387787e-01 5.64342841e-01 9.13371452e-01 1.94733542e+00
  2.30734817e+00]
 [1.85189622e+00 7.13699391e-01 4.65722964e-01 3.12089292e-01
  7.45684543e-01 7.49269758e-01 8.51219999e-01 6.57815325e-01
  1.84285741e-01 6.46405255e-01 2.30243298e+00 2.28870452e+00
  2.27644530e+00 9.07855874e-01 9.13371452e-01 2.44709406e+00
  1.21439378e-01]
 [1.65730117e+00 1.51740453e-01 1.65601079e+00 1.04069319e+00
  9.70514053e-01 4.18423371e-01 7.55337897e-01 1.44398486e-01
  7.23419558e-01 6.46405255e-01 1.30096632e+00 1.50566306e+00
  1.67168286e+00 1.27590555e+00 4.19657154e-01 1.41311066e+00
  2.01589367e+00]
 [1.56000365e+00 1.32700776e-02 3.64688024e-01 1.11307892e+00
  1.53633499e-01 1.26500089e-01 2.35443830e-01 1.74526071e+00
  2.64469643e+00 5.12666237e-01 1.13170435e+00 1.30265231e+00
  1.33570372e+00 9.32392519e-01 2.39451435e+00 6.37623101e-01
  1.21439378e+00]
 [1.26162457e+00 5.75229016e-01 5.23457216e-01 4.54487448e-01
  4.12187436e-02 2.62730954e-01 5.95534393e-01 1.01792019e+00
  4.97963598e-01 2.22898364e-02 6.87861845e-01 6.69452115e-01
  7.24968314e-01 9.07855874e-01 1.06148574e+00 3.96360306e-01
  1.57871191e+00]
 [1.65081467e+00 8.34860969e-01 7.89997012e-01 3.83288370e-01
  9.96744163e-01 4.18423371e-01 1.07601026e-01 4.79545589e-01
  6.66565447e-02 2.25127347e+00 9.27649638e-01 9.30465935e-01
  8.36961360e-01 5.39806195e-01 8.14628593e-01 1.04259994e+00
  3.15742382e-01]
 [8.78920973e-01 9.05827036e-02 1.54920242e-01 9.35081225e-01
  1.13913619e+00 1.39150098e+00 4.27208035e-01 3.36929800e-01
  4.09741701e-01 1.44883936e+00 1.06117853e+00 1.14314386e+00
  1.17891346e+00 9.07855874e-01 6.66514303e-01 8.96118953e-01
  4.61469635e-01]
 [1.97838300e-01 2.55593234e-01 1.27400249e+00 7.21483991e-01
  5.47085142e-01 1.78073202e+00 4.27208035e-01 1.24610545e+00
  9.39073085e-01 1.18136133e+00 7.50864912e-01 8.53128507e-01
  7.53339886e-01 1.96293162e-01 4.19657154e-01 2.67112380e-01
  3.15742382e-01]
 [9.69731996e-01 1.30219848e+00 7.68827786e-01 2.93102871e-01
  2.17709909e+00 8.46577519e-01 1.39455191e+00 1.67751821e+00
  9.62598924e-01 1.27052067e+00 1.44342848e-01 3.14183303e-02
  1.42604478e-01 1.27590555e+00 1.72800005e-01 1.04259994e+00
  1.26296953e+00]
 [1.35892210e+00 1.99455036e+00 1.38947099e+00 7.57083530e-01
  1.33398843e+00 1.29419322e+00 2.03376593e+00 3.38712498e-02
  1.38410355e+00 6.46405255e-01 1.01228062e+00 9.01464400e-01
  1.01615023e+00 1.71756517e-01 3.20914294e-01 1.20631398e-01
  3.15742382e-01]
 [1.74811219e+00 1.07718413e+00 1.83787368e-01 5.25686526e-01
  1.53633499e-01 5.54654236e-01 1.07494491e+00 1.46359453e+00
  2.52902772e-01 2.00608527e-01 1.70343367e+00 1.64100356e+00
  1.64331129e+00 9.07855874e-01 1.06148574e+00 1.04259994e+00
  1.65157553e+00]
 [1.07351602e+00 3.67523453e-01 1.89464570e+00 1.02408007e+00
  6.59499897e-01 5.15731132e-01 7.46815043e-01 1.67395282e+00
  9.60638438e-02 1.09220198e+00 1.31507148e+00 1.20114693e+00
  1.15651485e+00 9.32392519e-01 2.39451435e+00 5.08375175e-01
  1.40869678e+00]
 [2.95135825e-01 1.34431656e-01 7.03395634e-01 1.46788766e+00
  7.45684543e-01 7.49269758e-01 8.83180700e-01 2.15706380e-01
  8.74377027e-01 4.23506891e-01 8.63706227e-01 5.63113151e-01
  5.29353795e-01 1.30044220e+00 1.65394290e+00 2.67112380e-01
  1.19010590e+00]
 [1.91351799e-01 5.49842780e-01 1.91581492e+00 1.61028582e+00
  8.01891920e-01 1.62503960e+00 3.07888085e-01 3.72583748e-01
  6.25395228e-01 2.16211413e+00 3.49337902e-01 3.35934455e-01
  1.87401696e-01 9.07855874e-01 1.72800005e-01 1.20631398e-01
  6.07196888e-01]
 [6.77839422e-01 6.61773000e-01 8.55429164e-01 1.41567500e+00
  8.28122030e-01 4.57346476e-01 4.99652290e-01 7.68342561e-01
  7.72431723e-01 6.68695091e-02 6.03230859e-01 2.77931383e-01
  2.99394742e-01 9.07855874e-01 1.30834289e+00 2.67112380e-01
  1.70015129e-01]
 [1.26162457e+00 8.44092327e-01 8.04430575e-01 1.34328927e+00
  4.12187436e-02 1.72234737e+00 7.14854342e-01 1.44576756e+00
  3.41124670e-01 6.01825582e-01 2.19711442e+00 2.36604195e+00
  2.42726261e+00 1.96293162e-01 1.80205719e+00 9.13352010e-01
  4.85757510e-02]]
```

In [17]:

```python
index = (np.where(z>4)[0])
```

In [18]:

```python
df = df.drop(df.index[index])
```

In [19]:

```python
df.shape
```

Out[19]:

```
(30, 17)
```

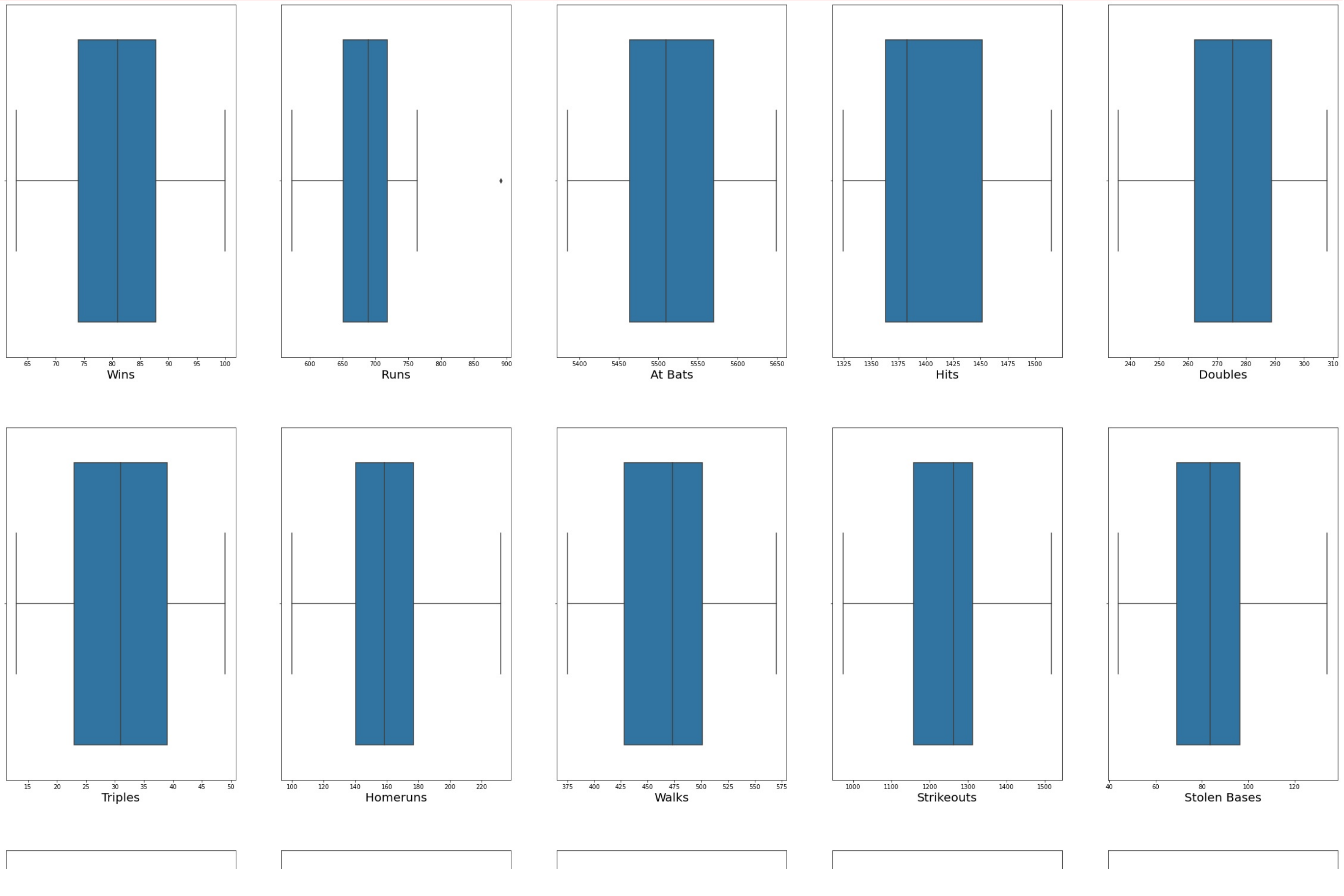
Basically my zscore did not remove any records based on the critiria given here. and i decided not go remove any record as we have less obeservation here.

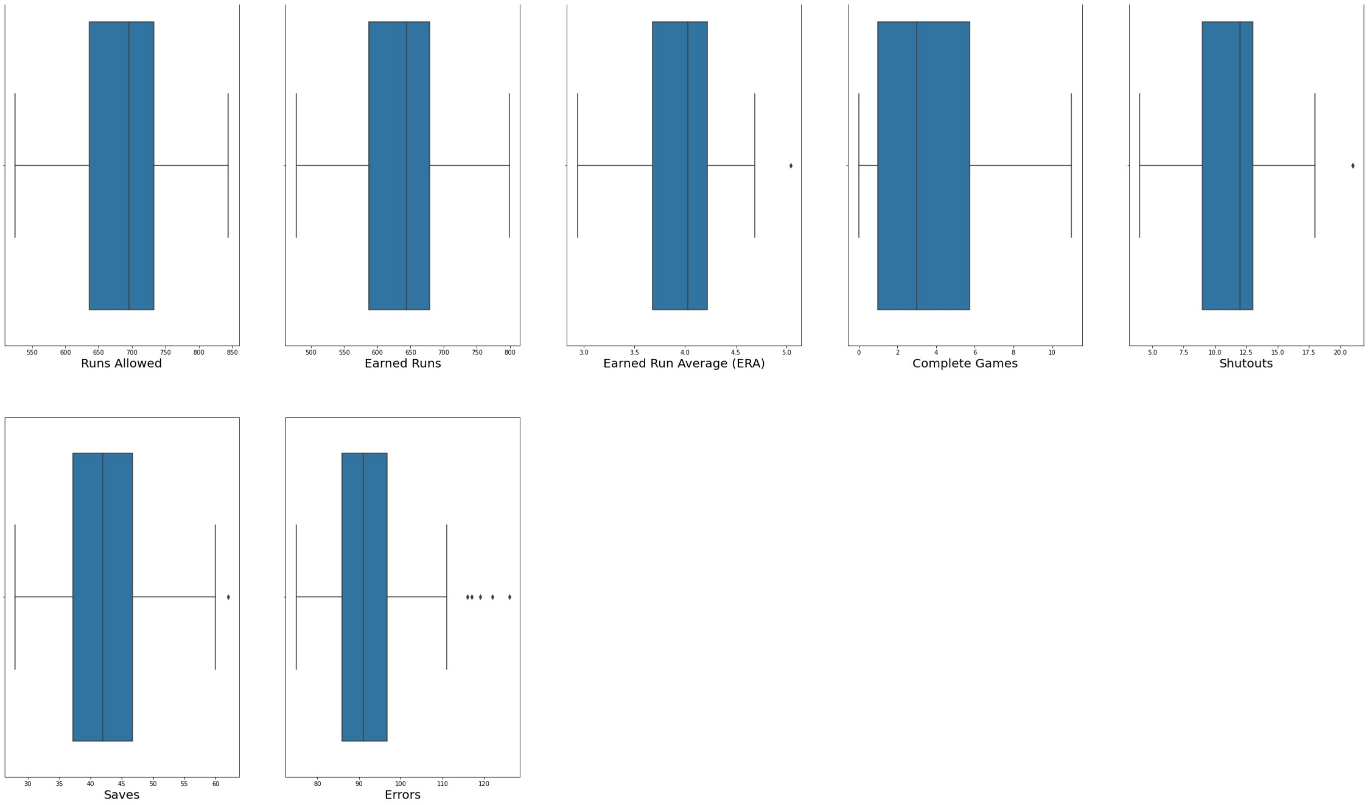
In [20]:

```python
#plots post outlier removal
plt.figure(figsize=(40,50),facecolor='white')
plotnumber = 1

for column in df:
    if plotnumber<=17:
            ax = plt.subplot(4,5,plotnumber)
            sns.boxplot(df[column])
            plt.xlabel(column,fontsize=20)

    plotnumber +=1
plt.show()
```

```
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable
as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other argu
ments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable
as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other argu
ments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable
as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other argu
ments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable
as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other argu
ments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable
as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other argu
```

```
ments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable
as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other argu
ments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable
as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other argu
ments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable
as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other argu
ments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable
as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other argu
ments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable
as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other argu
ments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable
as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other argu
ments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable
as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other argu
ments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable
as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other argu
ments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable
as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other argu
ments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable
as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other argu
ments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable
as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other argu
ments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable
as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other argu
ments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
```





In [21]:

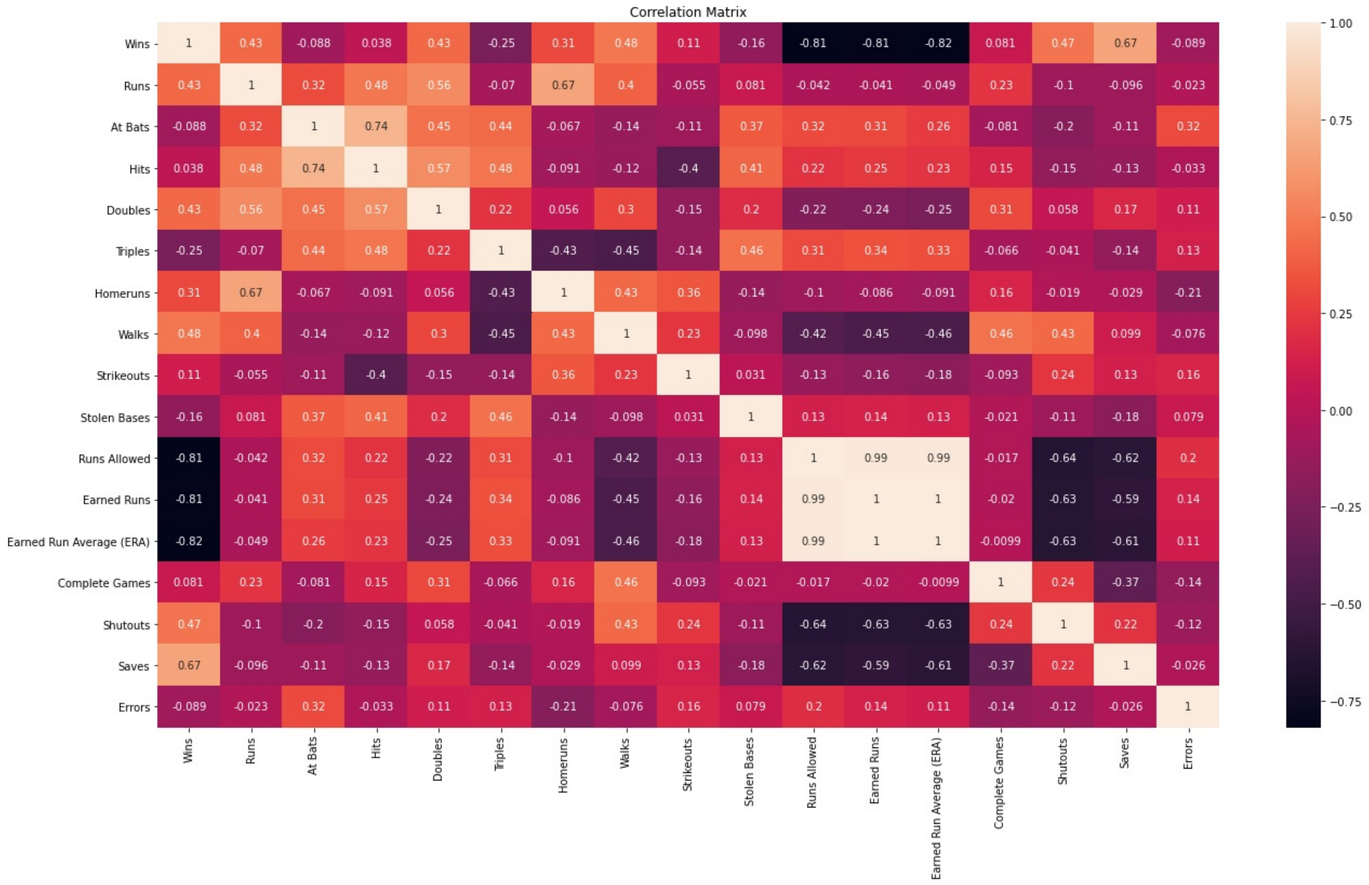
```
#checking correlation for between my features
corr_mat = df.corr()
```

In [22]:

```
#Visualising heatmap for corr_mat.
plt.figure(figsize=(22,12),facecolor="white")
sns.heatmap(corr_mat,annot=True)
plt.title("Correlation Matrix")
plt.show
```

Out[22]:

```
<function matplotlib.pyplot.show(close=None, block=None)>
```



	Wins	Runs	At Bats	Hits	Doubles	Triples	Homeruns	Walks	Strikeouts	Stolen Bases	Runs Allowed	Earned Runs	Earned Run Average (ERA)	Complete Games	Shutouts	Saves	Errors
Wins	1	0.43	-0.088	0.038	0.43	-0.25	0.31	0.48	0.11	-0.16	-0.81	-0.81	-0.82	0.081	0.47	0.67	-0.089
Runs	0.43	1	0.32	0.48	0.56	-0.07	0.67	0.4	-0.055	0.081	-0.042	-0.041	-0.049	0.23	-0.1	-0.096	-0.023
At Bats	-0.088	0.32	1	0.74	0.45	0.44	-0.067	-0.14	-0.11	0.37	0.32	0.31	0.26	-0.081	-0.2	-0.11	0.32
Hits	0.038	0.48	0.74	1	0.57	0.48	-0.091	-0.12	-0.4	0.41	0.22	0.25	0.23	0.15	-0.15	-0.13	-0.033
Doubles	0.43	0.56	0.45	0.57	1	0.22	0.056	0.3	-0.15	0.2	-0.22	-0.24	-0.25	0.31	0.058	0.17	0.11
Triples	-0.25	-0.07	0.44	0.48	0.22	1	-0.43	-0.45	-0.14	0.46	0.31	0.34	0.33	-0.066	-0.041	-0.14	0.13
Homeruns	0.31	0.67	-0.067	-0.091	0.056	-0.43	1	0.43	0.36	-0.14	-0.1	-0.086	-0.091	0.16	-0.019	-0.029	-0.21
Walks	0.48	0.4	-0.14	-0.12	0.3	-0.45	0.43	1	0.23	-0.098	-0.42	-0.45	-0.46	0.46	0.43	0.099	-0.076
Strikeouts	0.11	-0.055	-0.11	-0.4	-0.15	-0.14	0.36	0.23	1	0.031	-0.13	-0.16	-0.18	-0.093	0.24	0.13	0.16
Stolen Bases	-0.16	0.081	0.37	0.41	0.2	0.46	-0.14	-0.098	0.031	1	0.13	0.14	0.13	-0.021	-0.11	-0.18	0.079
Runs Allowed	-0.81	-0.042	0.32	0.22	-0.22	0.31	-0.1	-0.42	-0.13	0.13	1	0.99	0.99	-0.017	-0.64	-0.62	0.2
Earned Runs	-0.81	-0.041	0.31	0.25	-0.24	0.34	-0.086	-0.45	-0.16	0.14	0.99	1	1	-0.02	-0.63	-0.59	0.14
Earned Run Average (ERA)	-0.82	-0.049	0.26	0.23	-0.25	0.33	-0.091	-0.46	-0.18	0.13	0.99	1	1	-0.0099	-0.63	-0.61	0.11
Complete Games	0.081	0.23	-0.081	0.15	0.31	-0.066	0.16	0.46	-0.093	-0.021	-0.017	-0.02	-0.0099	1	0.24	-0.37	-0.14
Shutouts	0.47	-0.1	-0.2	-0.15	0.058	-0.041	-0.019	0.43	0.24	-0.11	-0.64	-0.63	-0.63	0.24	1	0.22	-0.12
Saves	0.67	-0.096	-0.11	-0.13	0.17	-0.14	-0.029	0.099	0.13	-0.18	-0.62	-0.59	-0.61	-0.37	0.22	1	-0.026
Errors	-0.089	-0.023	0.32	-0.033	0.11	0.13	-0.21	-0.076	0.16	0.079	0.2	0.14	0.11	-0.14	-0.12	-0.026	1

In [24]:

```
#Seprating my features and labels
X = df.drop(['Wins'], axis = 1)
y = df['Wins']
```

In [25]:

```
#Scaling my features
from sklearn.preprocessing import StandardScaler
ss = StandardScaler()
X_scaled = ss.fit_transform(X)
```

In [26]:

```
#Splitting my train and test data
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X_scaled, y, test_size=0.33,random_state=253)
```

In [27]:

```
#linear regression
from sklearn.linear_model import LinearRegression
lr = LinearRegression()

lr.fit(X_train , y_train)
pred = lr.predict(X_test)
```

In [28]:

```
#metrics for my linear regression
from sklearn.metrics import r2_score, roc_auc_score, mean_squared_error
rmse = np.sqrt(mean_squared_error(y_test, pred))
r2 = r2_score(y_test, pred)
```

In [29]:

```
#printing my metrics
print("The root mean Sq error calculated from the base model is:",rmse)
print("The r2-score is:",r2)
```

```
The root mean Sq error calculated from the base model is: 6.349390851580693
The r2-score is: 0.47259596826090267
```

In [30]:

```
# using Decision tree regressor to my dataset
from sklearn.tree import DecisionTreeRegressor

DT = DecisionTreeRegressor() #Instantiate the LogisticRegression object
DT.fit(X_train,y_train) #Call the fit method of logistic regression to train the model or to learn the parameters
y_pred1 = DT.predict(X_test) #Predict
```

In [31]:

```
#metrics for my decision tree
from sklearn.metrics import r2_score, roc_auc_score, mean_squared_error
rmse1 = np.sqrt(mean_squared_error(y_test, y_pred1))
r2_1 = r2_score(y_test, y_pred1)
```

In [32]:

```
#printing my metrics
print("The root mean Sq error calculated from the base model is:",rmse1)
print("The r2-score is:",r2_1)
```

```
The root mean Sq error calculated from the base model is: 10.469001862641921
The r2-score is: -0.43380429094714823
```

In [33]:

```
# using Decision tree random forest to my dataset
from sklearn.ensemble import RandomForestRegressor

rf = RandomForestRegressor() #Instantiate the LogisticRegression object
rf.fit(X_train,y_train) #Call the fit method of logistic regression to train the model or to learn the parameters
y_pred2 = rf.predict(X_test) #Predict
```

In [34]:

```
#metrics for random forest
from sklearn.metrics import r2_score, roc_auc_score, mean_squared_error
rmse2 = np.sqrt(mean_squared_error(y_test, y_pred2))
r2_2 = r2_score(y_test, y_pred2)
```

In [35]:

```
#printing my metrics
print("The root mean Sq error calculated from the base model is:",rmse2)
print("The r2-score is:",r2_2)
```

```
The root mean Sq error calculated from the base model is: 6.264455283582123
The root mean Sq error calculated from the base model is: [illegible]
The r2-score is: 0.48661172161172195
```

In [36]:

```
# using adaboost  regressor to my dataset
from sklearn.ensemble import AdaBoostRegressor

ab = AdaBoostRegressor() #Instantiate the LogisticRegression object
ab.fit(X_train,y_train) #Call the fit method of logistic regression to train the model or to learn the parameters
y_pred3 = ab.predict(X_test) #Predict
```

In [37]:

```
#metrics for my adaboost model
from sklearn.metrics import r2_score, roc_auc_score, mean_squared_error
rmse3 = np.sqrt(mean_squared_error(y_test, y_pred3))
r2_3 = r2_score(y_test, y_pred3)
```

In [38]:

```
#printing my scores
print("The root mean Sq error calculated from the base model is:",rmse3)
print("The r2-score is:",r2_3)
```

```
The root mean Sq error calculated from the base model is: 6.7611019811862025
The r2-score is: 0.4019819466248038
```

In [39]:

```
# using bagging  regressor to my dataset
from sklearn.ensemble import BaggingRegressor

bg = BaggingRegressor() #Instantiate the LogisticRegression object
bg.fit(X_train,y_train) #Call the fit method of logistic regression to train the model or to learn the parameters
y_pred4 = bg.predict(X_test) #Predict
```

In [40]:

```
#metrics for my bagging model
from sklearn.metrics import r2_score, roc_auc_score, mean_squared_error
rmse4 = np.sqrt(mean_squared_error(y_test, y_pred4))
r2_4  = r2_score(y_test, y_pred4)
```

In [41]:

```
#score for my metrics to be printed
print("The root mean Sq error calculated from the base model is:",rmse4)
print("The r2-score is:",r2_4)
```

```
The root mean Sq error calculated from the base model is: 7.279079612148776
The r2-score is: 0.3068419675562535
```

This dataset changes r2 score and rmse score for my each run and i have decided not to use any hyperparameter tuning for the same as it will change my current accuracy which is in best currently and the same has been saved in pkl file.

In [42]:

```
#Saving the model in pkl file
import joblib
joblib.dump(ab,"Baseball_Prediction.pkl")

model = joblib.load("Baseball_Prediction.pkl")

#Predicting the saved model
prediction = model.predict(X_test)

#Converting the result into DataFrame
prediction = pd.DataFrame(prediction)

#Saving the result into CSV
prediction.to_csv("Results_Avacado.csv",index=False)
```

In [ ]:

Loading [MathJax]/extensions/Safe.js