Cricket Player Performance using machine learning

```
import pandas as pd
import numpy as np

import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
import plotly.graph_objects as go

%matplotlib inline
%matplotlib inline
%matplotlib inline
```

```
df = pd.read_excel("/content/")
df.head()
```

	Unnamed 0	Mat	Runs	HF	50	Ht	Ns	Opposition	Ground	Start Date	Match_ID	Batsman	Player_ID
0	0	1	2340	-	-	-	-	-	Lahore	10-Dec-1959	1959-1960	Saeed Ahmed	49542
1	1	2	2340	-	-	-	-	-	Lahore	15-Dec-1959	1959-1960	Saeed Ahmed	49542
2	2	3	2340	-	-	-	-	-	Lahore	22-Dec-1959	1959-1960	Saeed Ahmed	49542
3	3	4	2340	-	-	-	-	England XI	Dhaka	21-Jan-1960	1959-1960	Saeed Ahmed	49542
4	4	5	2340	-	-	-	-	-	Dhaka	3-Jan-1960	1959-1960	Saeed Ahmed	49542

```
df.shape
```

```
(12148, 13)
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 12148 entries, 0 to 12147
Data columns (total 13 columns):
 0   Unnamed 0   12148 non-null  int64
 1   Mat         12148 non-null  int64
 2   Runs        12148 non-null  int64
 3   HF          12148 non-null  int64
 4   50          12148 non-null  int64
 5   Ht          12148 non-null  int64
 6   Ns          12148 non-null  int64
 7   Opposition  12148 non-null  object
 8   Ground      12148 non-null  object
 9   Start Date  12148 non-null  object
 10  Match_ID    12148 non-null  int64
 11  Batsman     12148 non-null  object
 12  Player_ID   12148 non-null  int64
dtype: int64
memory usage: 1.2 MB
```

```
df.tail()
```

	Unnamed 0	Mat	Runs	HF	50	Ht	Ns	Opposition	Ground	Start Date	Match_ID	Batsman	Player_ID
12143	12143	35	0	0	0	0	1	-	Colombo (RPS)	27-Mar-2014	2014-1436	Kulkarni Nach	50258
12144	12144	36	0	0	0	0	1	-	Colombo (RPS)	29-Mar-2014	2014-1436	Kulkarni Nach	50258
12145	12145	37	0	0	0	0	1	-	Colombo (RPS)	4-Apr-2014	2014-1436	Kulkarni Nach	50258
12146	12146	38	0	0	0	0	1	-	Mirpur	8-May-2014	2014-1436	Kulkarni Nach	50258
12147	12147	39	0	0	0	0	1	-	Chittagong	16-May-2014	2014-1436	Kulkarni Nach	50258

```
df.isnull() sum()
```

```
Unnamed 0    0
Mat          0
Runs         0
HF           0
50           0
Ht           0
Ns           0
Opposition   0
Ground       0
Start Date   0
Match_ID     0
Batsman      0
Player_ID    0
dtype: int64
```

```
df.describe()
```

	Unnamed 0	Mat	Runs	HF	50	Ns	Match_ID	Player_ID
count	12148.000000	12148.000000	12148.000000	12148.000000	12148.000000	12148.000000	12148.000000	12148.000000
mean	5974.500000	5.208004e+05						
std	3506.004841	2.903541e+05						
min	0.000000	3.200000e+04						
25%	2974.750000	2.800000e+05						
50%	5974.500000	2.801000e+05						
75%	8974.250000	7.201000e+05						
max	11974.000000	1.017701e+06						

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 12148 entries, 0 to 12147
Data columns (total 13 columns):
 0   Unnamed 0   12148 non-null  int64
 1   Mat         12148 non-null  int64
 2   Runs        12148 non-null  int64
 3   HF          12148 non-null  int64
 4   50          12148 non-null  int64
 5   Ht          12148 non-null  int64
 6   Ns          12148 non-null  int64
 7   Opposition  12148 non-null  object
 8   Ground      12148 non-null  object
 9   Start Date  12148 non-null  object
 10  Match_ID    12148 non-null  int64
 11  Batsman     12148 non-null  object
 12  Player_ID   12148 non-null  int64
dtype: int64
memory usage: 1.2 MB
```

```
df.describe()
```

	Unnamed 0	Mat	Runs	HF	50	Ns	Match_ID	Player_ID
count	12148.000000	12148.000000	12148.000000	12148.000000	12148.000000	12148.000000	12148.000000	12148.000000
mean	5974.500000	5.208004e+05						
std	3506.004841	2.903541e+05						
min	0.000000	3.200000e+04						
25%	2974.750000	2.800000e+05						
50%	5974.500000	2.801000e+05						
75%	8974.250000	7.201000e+05						
max	11974.000000	1.017701e+06						

```
df.isnull()
```

```
Unnamed 0    False
Mat          False
Runs         False
HF           False
50           False
Ht           False
Ns           False
Opposition   False
Ground       False
Start Date   False
Match_ID     False
Batsman      False
Player_ID    False
dtype: bool
```

```
df.isnull() sum()
```

```
Unnamed 0    0
Mat          0
Runs         0
HF           0
50           0
Ht           0
Ns           0
Opposition   0
Ground       0
Start Date   0
Match_ID     0
Batsman      0
Player_ID    0
dtype: int64
```

```
df.shape
```

```
(12148, 13)
```

```
df.dtypes
```

```
Unnamed 0    int64
Mat          int64
Runs         int64
HF           int64
50           int64
Ht           int64
Ns           int64
Opposition   object
Ground       object
Start Date   object
Match_ID     int64
Batsman      object
Player_ID    int64
dtype: int64
```

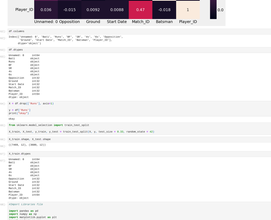
```
df.tail()
```

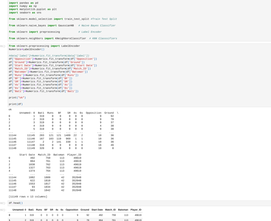
	Unnamed 0	Mat	Runs	HF	50	Ht	Ns	Opposition	Ground	Start Date	Match_ID	Batsman	Player_ID
12143	12143	35	0	0	0	0	1	-	Colombo (RPS)	27-Mar-2014	2014-1436	Kulkarni Nach	50258
12144	12144	36	0	0	0	0	1	-	Colombo (RPS)	29-Mar-2014	2014-1436	Kulkarni Nach	50258
12145	12145	37	0	0	0	0	1	-	Colombo (RPS)	4-Apr-2014	2014-1436	Kulkarni Nach	50258
12146	12146	38	0	0	0	0	1	-	Mirpur	8-May-2014	2014-1436	Kulkarni Nach	50258
12147	12147	39	0	0	0	0	1	-	Chittagong	16-May-2014	2014-1436	Kulkarni Nach	50258

```
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
import pandas as pd

plt.figure(figsize=(8,6))
sns.boxplot(x='Rate', y='Count', data=df)
plt.title('Boxplot of Rate')
plt.show()
```

```
C:\Users\Dell\AnaConda3\Lib\site-packages\IPykernel\ipykernel\pylaunch.py:102: UserWarning: Creating legend with loc="best" can be slow with large amounts of data.
  "Creating legend with loc="best" can be slow with large amounts of data.",
```






```
from sklearn.preprocessing import LabelEncoder
from sklearn import preprocessing

Batsman_encoder = preprocessing.LabelEncoder()
df['Opposition'] = Batsman_encoder.fit_transform(df['Opposition'])
df['Ground'] = Batsman_encoder.fit_transform(df['Ground'])
df['Start Date'] = Batsman_encoder.fit_transform(df['Start Date'])
df['Batsman'] = Batsman_encoder.fit_transform(df['Batsman'])
df['Batsman'].unique()
```

```
array[ 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17,
       18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33,
       34, 35, 36, 37, 38, 39, 40,
```