In [1]:
```
import pandas as pd
```

In [2]:
```
data=pd.read_csv(r"C:\Users\manju\Desktop\criket dataset.csv")
data
```

Out[2]:

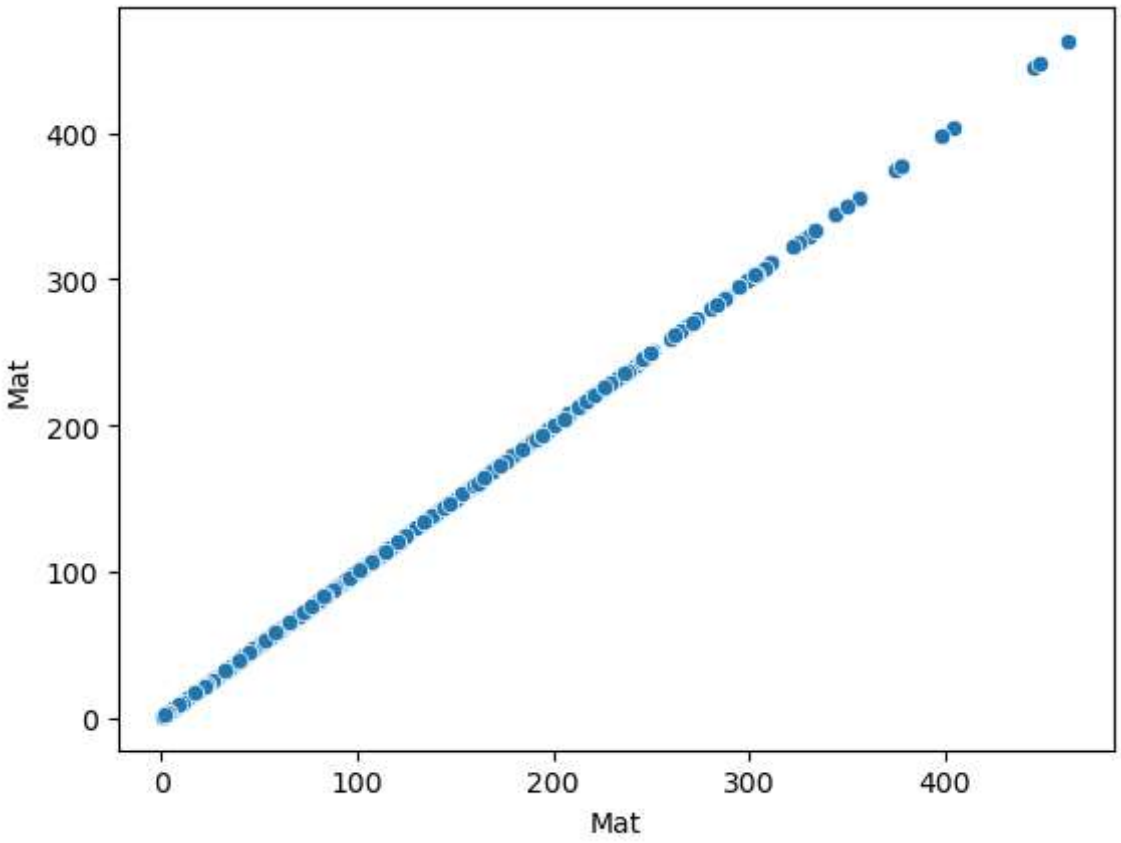
	Unnamed: 0	Player	Span	Mat	Inns	NO	Runs	HS	Ave	BF	SR	100
0	0	SR Tendulkar (INDIA)	1989-2012	463	452	41	18426	200*	44.83	21367	86.23	49
1	1	KC Sangakkara (Asia/ICC/SL)	2000-2015	404	380	41	14234	169	41.98	18048	78.86	25
2	2	RT Ponting (AUS/ICC)	1995-2012	375	365	39	13704	164	42.03	17046	80.39	30
3	3	ST Jayasuriya (Asia/SL)	1989-2011	445	433	18	13430	189	32.36	14725	91.2	28
4	4	DPMD Jayawardene (Asia/SL)	1998-2015	448	418	39	12650	144	33.37	16020	78.96	19
...	...	...	...	...	...	...	...	...	...	...	...	...
2495	45	ZS Ansari (ENG)	2015-2015	1	-	-	-	-	-	-	-	-
2496	46	Ariful Haque (BDESH)	2018-2018	1	-	-	-	-	-	-	-	-
2497	47	Ashfaq Ahmed (PAK)	1994-1994	3	-	-	-	-	-	-	-	-
2498	48	MD Bailey (NZ)	1998-1998	1	-	-	-	-	-	-	-	-
2499	49	GR Beard (AUS)	1981-1981	2	-	-	-	-	-	-	-	-

2500 rows × 15 columns

In [10]:
```
x = data[['Mat']]
y = data['Runs']
```

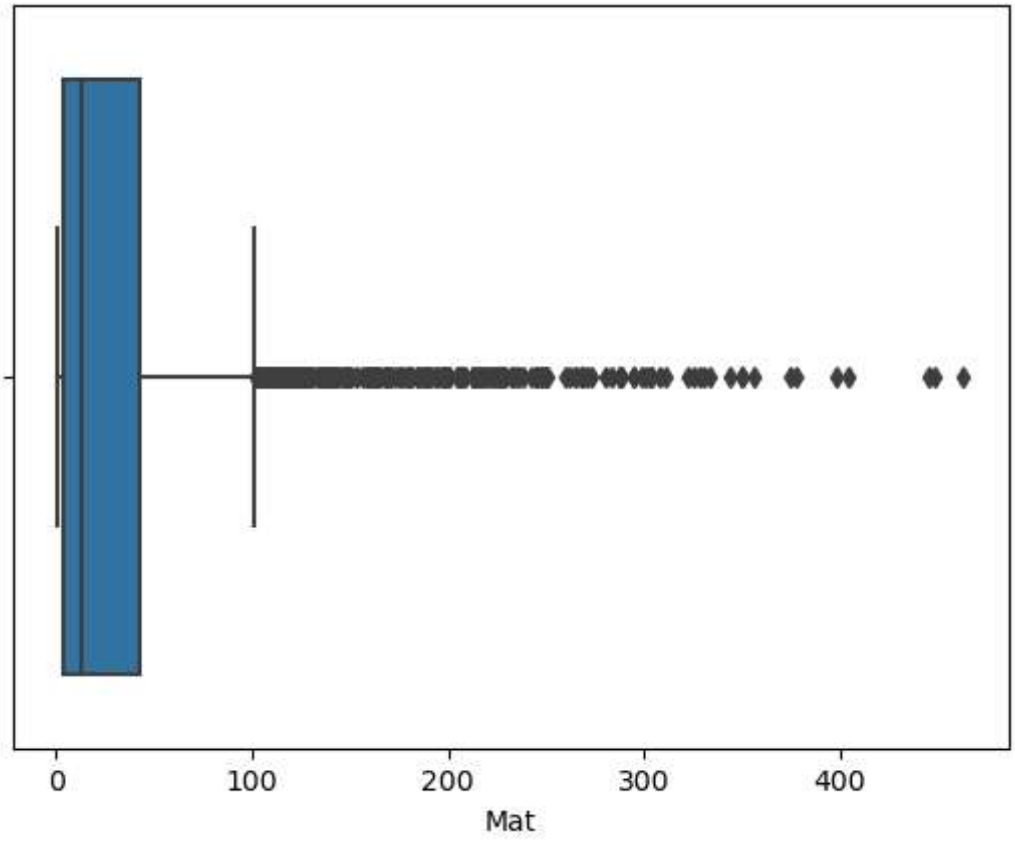
In [114]:
```
#bivariate
import seaborn as sns
sns.scatterplot(x='Mat',y='Mat',data=data)
```

Out[114]: <Axes: xlabel='Mat', ylabel='Mat'>



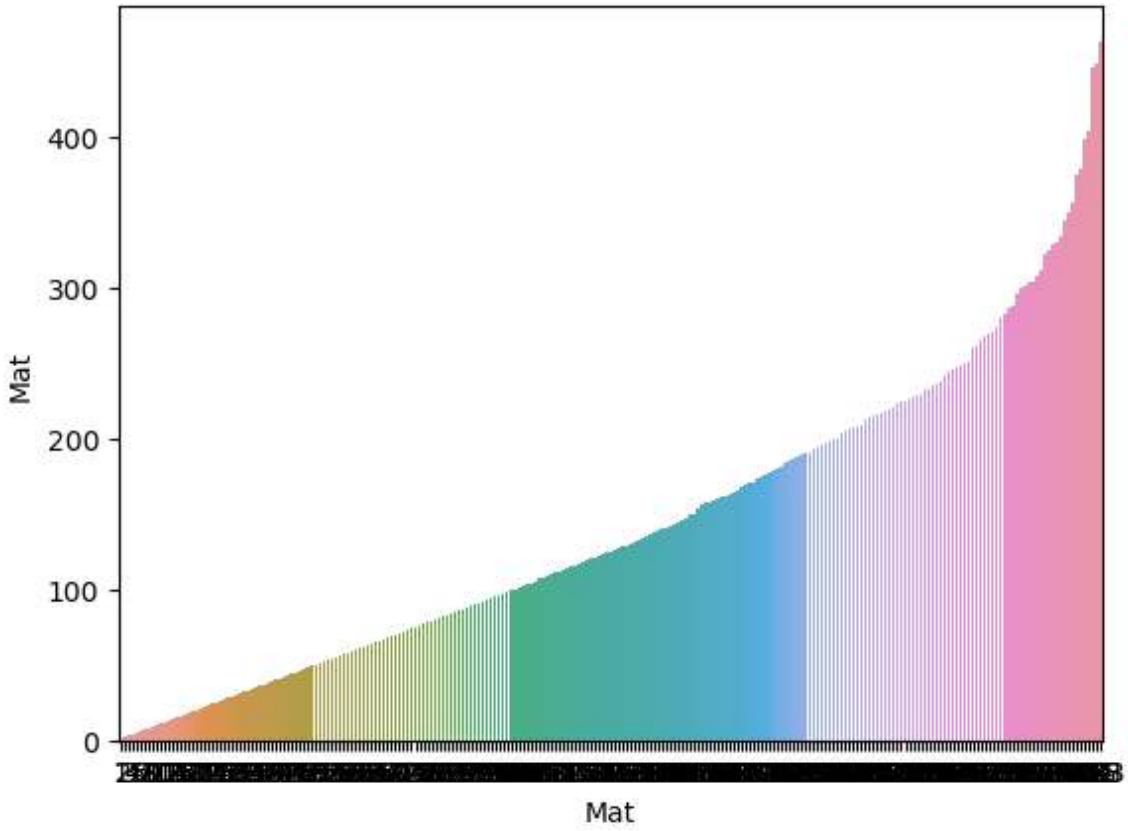
In [115]:
```
# univariate
import seaborn as sns
sns.boxplot(x='Mat',data=data)
```

Out[115]: <Axes: xlabel='Mat'>



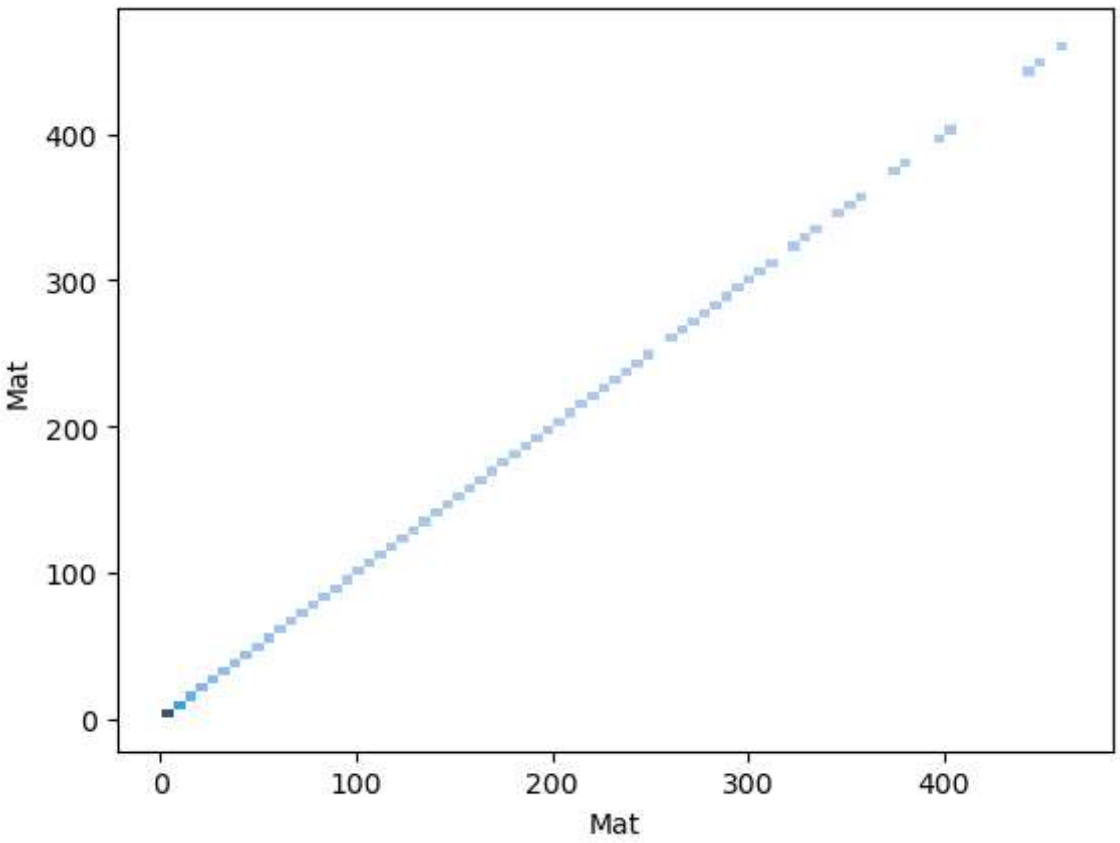
In [116]:
```
1 import seaborn as sns
2 sns.barplot(x='Mat',y='Mat',data=data)
```

Out[116]: <Axes: xlabel='Mat', ylabel='Mat'>



In [117]:
```
1 import seaborn as sns
2 sns.histplot(x='Mat',y='Mat',data=data)
```

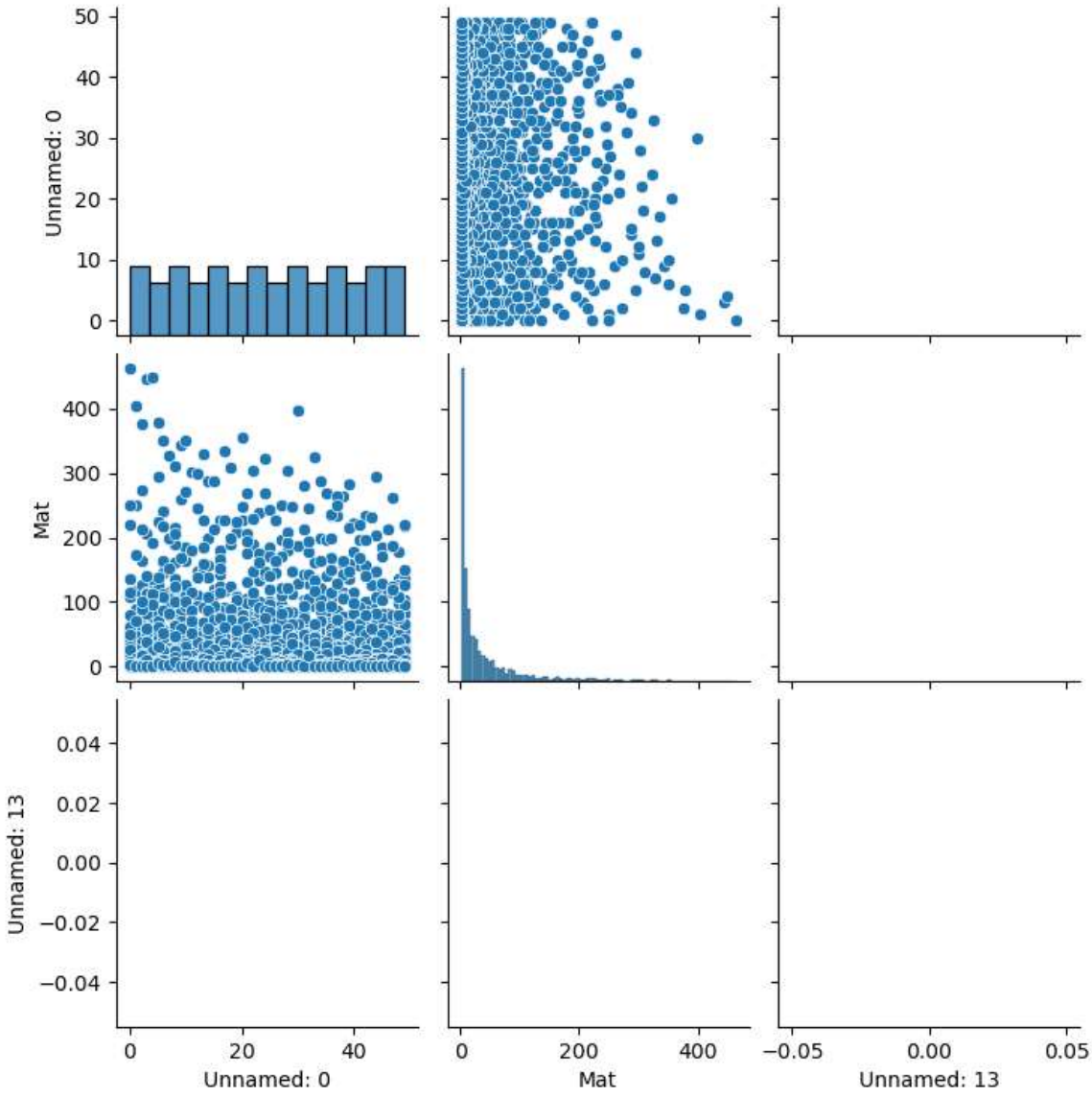
Out[117]: <Axes: xlabel='Mat', ylabel='Mat'>



In [118]:
```
1 #pairplot
```

In [119]:
```
1 sns.pairplot(data=data)
```

Out[119]: <seaborn.axisgrid.PairGrid at 0x16bda12ef90>



In [120]:
```
1 from sklearn.model_selection import train_test_split
```

In [121]:
```
1 x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2
```

In [122]:
```
1 x_train.shape
```

Out[122]: (2000, 1)

In [123]:
```
1 y_train
```

Out[123]:
```
2055     12
1961     16
1864     23
2326      3
461     784
       ...
1638     40
1095    141
1130    131
1294     91
860     252
Name: Runs, Length: 2000, dtype: object
```

In [124]:
```
1 x_test
```

Out[124]:

	Mat
1447	4
1114	25
1064	10
2287	1
1537	4
...	...
2375	2
1609	59
596	53
84	199
2213	12

500 rows × 1 columns

In [125]:
```
1 y_test
```

Out[125]:
```
1447      65
1114     136
1064     151
2287       4
1537      52
         ...
2375       1
1609      44
596      512
84      5117
2213       6
Name: Runs, Length: 500, dtype: object
```

In [126]:
```
1 #filling missing values
```

In [127]:
```
1 data
```

Out[127]:

	Unnamed: 0	Player	Span	Mat	Inns	NO	Runs	HS	Ave	BF	SR	100
0	0	SR Tendulkar (INDIA)	1989-2012	463	452	41	18426	200*	44.83	21367	86.23	49
1	1	KC Sangakkara (Asia/ICC/SL)	2000-2015	404	380	41	14234	169	41.98	18048	78.86	25
2	2	RT Ponting (AUS/ICC)	1995-2012	375	365	39	13704	164	42.03	17046	80.39	30
3	3	ST Jayasuriya (Asia/SL)	1989-2011	445	433	18	13430	189	32.36	14725	91.2	28
4	4	DPMD Jayawardene (Asia/SL)	1998-2015	448	418	39	12650	144	33.37	16020	78.96	19
...	...	...	...	...	...	...	...	...	...	...	...	...
2495	45	ZS Ansari (ENG)	2015-2015	1	-	-	-	-	-	-	-	-
2496	46	Ariful Haque (BDESH)	2018-2018	1	-	-	-	-	-	-	-	-
2497	47	Ashfaq Ahmed (PAK)	1994-1994	3	-	-	-	-	-	-	-	-
2498	48	MD Bailey (NZ)	1998-1998	1	-	-	-	-	-	-	-	-
2499	49	GR Beard (AUS)	1981-1981	2	-	-	-	-	-	-	-	-

2500 rows × 15 columns

In [128]:
```
1 data.shape
```

Out[128]:
```
(2500, 15)
```

In [129]:
```
df1=data.isnull()
df1
```

Out[129]:

	Unnamed: 0	Player	Span	Mat	Inns	NO	Runs	HS	Ave	BF	SR	100	
0	False	False	False	False	False	False	False	False	False	False	False	False	F
1	False	False	False	False	False	False	False	False	False	False	False	False	F
2	False	False	False	False	False	False	False	False	False	False	False	False	F
3	False	False	False	False	False	False	False	False	False	False	False	False	F
4	False	False	False	False	False	False	False	False	False	False	False	False	F
...	...	...	...	...	...	...	...	...	...	...	...	...	
2495	False	False	False	False	False	False	False	False	False	False	False	False	F
2496	False	False	False	False	False	False	False	False	False	False	False	False	F
2497	False	False	False	False	False	False	False	False	False	False	False	False	F
2498	False	False	False	False	False	False	False	False	False	False	False	False	F
2499	False	False	False	False	False	False	False	False	False	False	False	False	F

2500 rows × 15 columns

In [130]:
```
df2=data.isnull().sum()
```

In [131]:
```
df2
```

Out[131]:
```
Unnamed: 0        0
Player            0
Span              0
Mat               0
Inns              0
NO                0
Runs              0
HS                0
Ave               0
BF                0
SR                0
100               0
50                0
0                 0
Unnamed: 13    2500
dtype: int64
```

In [132]:
```
df3 = data.isnull().sum().sum()
```

In [133]:
```
df3
```

Out[133]:
```
2500
```

In [134]:
```
#filling nan values using fillna method
```

In [135]:
```
df4 = data.fillna(value='harish')
```

In [136]:
```
df4
```

Out[136]:

	Unnamed: 0	Player	Span	Mat	Inns	NO	Runs	HS	Ave	BF	SR	100
0	0	SR Tendulkar (INDIA)	1989-2012	463	452	41	18426	200*	44.83	21367	86.23	49
1	1	KC Sangakkara (Asia/ICC/SL)	2000-2015	404	380	41	14234	169	41.98	18048	78.86	25
2	2	RT Ponting (AUS/ICC)	1995-2012	375	365	39	13704	164	42.03	17046	80.39	30
3	3	ST Jayasuriya (Asia/SL)	1989-2011	445	433	18	13430	189	32.36	14725	91.2	28
4	4	DPMD Jayawardene (Asia/SL)	1998-2015	448	418	39	12650	144	33.37	16020	78.96	19
...	...	...	...	...	...	...	...	...	...	...	...	...
2495	45	ZS Ansari (ENG)	2015-2015	1	-	-	-	-	-	-	-	-
2496	46	Ariful Haque (BDESH)	2018-2018	1	-	-	-	-	-	-	-	-
2497	47	Ashfaq Ahmed (PAK)	1994-1994	3	-	-	-	-	-	-	-	-
2498	48	MD Bailey (NZ)	1998-1998	1	-	-	-	-	-	-	-	-
2499	49	GR Beard (AUS)	1981-1981	2	-	-	-	-	-	-	-	-

2500 rows × 15 columns

In [137]:
```
df5 = data.fillna(value=5)
df5
```

Out[137]:

	Unnamed: 0	Player	Span	Mat	Inns	NO	Runs	HS	Ave	BF	SR	100
0	0	SR Tendulkar (INDIA)	1989-2012	463	452	41	18426	200*	44.83	21367	86.23	49
1	1	KC Sangakkara (Asia/ICC/SL)	2000-2015	404	380	41	14234	169	41.98	18048	78.86	25
2	2	RT Ponting (AUS/ICC)	1995-2012	375	365	39	13704	164	42.03	17046	80.39	30
3	3	ST Jayasuriya (Asia/SL)	1989-2011	445	433	18	13430	189	32.36	14725	91.2	28
4	4	DPMD Jayawardene (Asia/SL)	1998-2015	448	418	39	12650	144	33.37	16020	78.96	19
...	...	...	...	...	...	...	...	...	...	...	...	...
2495	45	ZS Ansari (ENG)	2015-2015	1	-	-	-	-	-	-	-	-
2496	46	Ariful Haque (BDESH)	2018-2018	1	-	-	-	-	-	-	-	-
2497	47	Ashfaq Ahmed (PAK)	1994-1994	3	-	-	-	-	-	-	-	-
2498	48	MD Bailey (NZ)	1998-1998	1	-	-	-	-	-	-	-	-
2499	49	GR Beard (AUS)	1981-1981	2	-	-	-	-	-	-	-	-

2500 rows × 15 columns

In [138]:
```
df6 = data.fillna(method='pad')
```

In [139]:
```
1 df6
```

Out[139]:

	Unnamed: 0	Player	Span	Mat	Inns	NO	Runs	HS	Ave	BF	SR	100
0	0	SR Tendulkar (INDIA)	1989-2012	463	452	41	18426	200*	44.83	21367	86.23	49
1	1	KC Sangakkara (Asia/ICC/SL)	2000-2015	404	380	41	14234	169	41.98	18048	78.86	25
2	2	RT Ponting (AUS/ICC)	1995-2012	375	365	39	13704	164	42.03	17046	80.39	30
3	3	ST Jayasuriya (Asia/SL)	1989-2011	445	433	18	13430	189	32.36	14725	91.2	28
4	4	DPMD Jayawardene (Asia/SL)	1998-2015	448	418	39	12650	144	33.37	16020	78.96	19
...	...	...	...	...	...	...	...	...	...	...	...	...
2495	45	ZS Ansari (ENG)	2015-2015	1	-	-	-	-	-	-	-	-
2496	46	Ariful Haque (BDESH)	2018-2018	1	-	-	-	-	-	-	-	-
2497	47	Ashfaq Ahmed (PAK)	1994-1994	3	-	-	-	-	-	-	-	-
2498	48	MD Bailey (NZ)	1998-1998	1	-	-	-	-	-	-	-	-
2499	49	GR Beard (AUS)	1981-1981	2	-	-	-	-	-	-	-	-

2500 rows × 15 columns

In [ ]:
```
1
```

In [140]:
```
1 df7 = data.fillna(method='bfill')
```

In [141]:
```
1 df7
```

Out[141]:

	Unnamed: 0	Player	Span	Mat	Inns	NO	Runs	HS	Ave	BF	SR	100
0	0	SR Tendulkar (INDIA)	1989-2012	463	452	41	18426	200*	44.83	21367	86.23	49
1	1	KC Sangakkara (Asia/ICC/SL)	2000-2015	404	380	41	14234	169	41.98	18048	78.86	25
2	2	RT Ponting (AUS/ICC)	1995-2012	375	365	39	13704	164	42.03	17046	80.39	30
3	3	ST Jayasuriya (Asia/SL)	1989-2011	445	433	18	13430	189	32.36	14725	91.2	28
4	4	DPMD Jayawardene (Asia/SL)	1998-2015	448	418	39	12650	144	33.37	16020	78.96	19
...	...	...	...	...	...	...	...	...	...	...	...	...
2495	45	ZS Ansari (ENG)	2015-2015	1	-	-	-	-	-	-	-	-
2496	46	Ariful Haque (BDESH)	2018-2018	1	-	-	-	-	-	-	-	-
2497	47	Ashfaq Ahmed (PAK)	1994-1994	3	-	-	-	-	-	-	-	-
2498	48	MD Bailey (NZ)	1998-1998	1	-	-	-	-	-	-	-	-
2499	49	GR Beard (AUS)	1981-1981	2	-	-	-	-	-	-	-	-

2500 rows × 15 columns

In [142]:
```
1 # replacing a value
```

In [143]:

```
import numpy as np
df8 = data.replace(to_replace=np.nan,value=6)
df8
```

Out[143]:

	Unnamed: 0	Player	Span	Mat	Inns	NO	Runs	HS	Ave	BF	SR	100
0	0	SR Tendulkar (INDIA)	1989-2012	463	452	41	18426	200*	44.83	21367	86.23	49
1	1	KC Sangakkara (Asia/ICC/SL)	2000-2015	404	380	41	14234	169	41.98	18048	78.86	25
2	2	RT Ponting (AUS/ICC)	1995-2012	375	365	39	13704	164	42.03	17046	80.39	30
3	3	ST Jayasuriya (Asia/SL)	1989-2011	445	433	18	13430	189	32.36	14725	91.2	28
4	4	DPMD Jayawardene (Asia/SL)	1998-2015	448	418	39	12650	144	33.37	16020	78.96	19
...	...	...	...	...	...	...	...	...	...	...	...	...
2495	45	ZS Ansari (ENG)	2015-2015	1	-	-	-	-	-	-	-	-
2496	46	Ariful Haque (BDESH)	2018-2018	1	-	-	-	-	-	-	-	-
2497	47	Ashfaq Ahmed (PAK)	1994-1994	3	-	-	-	-	-	-	-	-
2498	48	MD Bailey (NZ)	1998-1998	1	-	-	-	-	-	-	-	-
2499	49	GR Beard (AUS)	1981-1981	2	-	-	-	-	-	-	-	-

2500 rows × 15 columns

In [ ]:

```
1
```

In [144]:
```
import numpy as np
df8 = data.replace(to_replace=np.nan,value='malur')
df8
```

Out[144]:

	Unnamed: 0	Player	Span	Mat	Inns	NO	Runs	HS	Ave	BF	SR	100
0	0	SR Tendulkar (INDIA)	1989-2012	463	452	41	18426	200*	44.83	21367	86.23	49
1	1	KC Sangakkara (Asia/ICC/SL)	2000-2015	404	380	41	14234	169	41.98	18048	78.86	25
2	2	RT Ponting (AUS/ICC)	1995-2012	375	365	39	13704	164	42.03	17046	80.39	30
3	3	ST Jayasuriya (Asia/SL)	1989-2011	445	433	18	13430	189	32.36	14725	91.2	28
4	4	DPMD Jayawardene (Asia/SL)	1998-2015	448	418	39	12650	144	33.37	16020	78.96	19
...	...	...	...	...	...	...	...	...	...	...	...	...
2495	45	ZS Ansari (ENG)	2015-2015	1	-	-	-	-	-	-	-	-
2496	46	Ariful Haque (BDESH)	2018-2018	1	-	-	-	-	-	-	-	-
2497	47	Ashfaq Ahmed (PAK)	1994-1994	3	-	-	-	-	-	-	-	-
2498	48	MD Bailey (NZ)	1998-1998	1	-	-	-	-	-	-	-	-
2499	49	GR Beard (AUS)	1981-1981	2	-	-	-	-	-	-	-	-

2500 rows × 15 columns

In [145]:
```
# drop columns
```

In [146]:
```
data.dropna()
```

Out[146]:

Unnamed: 0	Player	Span	Mat	Inns	NO	Runs	HS	Ave	BF	SR	100	50	0	Unnamed: 13

In [147]:
```
df11 = data.dropna(how='any')
df11
```

Out[147]:

Unnamed: 0	Player	Span	Mat	Inns	NO	Runs	HS	Ave	BF	SR	100	50	0	Unnamed: 13

In [9]:

```
#using linear regression algorithm
import numpy as np
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn import metrics

np.random.seed(42)
X = 2 * np.random.rand(100, 1)
Y = 4 + 3 * X + np.random.randn(100, 1)

X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2
model = LinearRegression()

model.fit(X_train, Y_train)

Y_pred = model.predict(X_test)

print('Mean Absolute Error:', metrics.mean_absolute_error(Y_test, Y_pred
print('Mean Squared Error:', metrics.mean_squared_error(Y_test, Y_pred)
print('Root Mean Squared Error:', np.sqrt(metrics.mean_squared_error(Y_t
plt.scatter(X_test, Y_test, color='red')
plt.plot(X_test, Y_pred, color='blue', linewidth=3)
plt.xlabel('X')
plt.ylabel('Y')
plt.title('Linear Regression Model')
plt.show()

```

```
Mean Absolute Error: 0.5913425779189777
Mean Squared Error: 0.6536995137170021
Root Mean Squared Error: 0.8085168605026132
```