

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

```
In [2]: data =pd.read_csv(r"C:\Users\pradeep\OneDrive\Documents\ipl_data.csv")
```

```
In [3]: data.head()
```

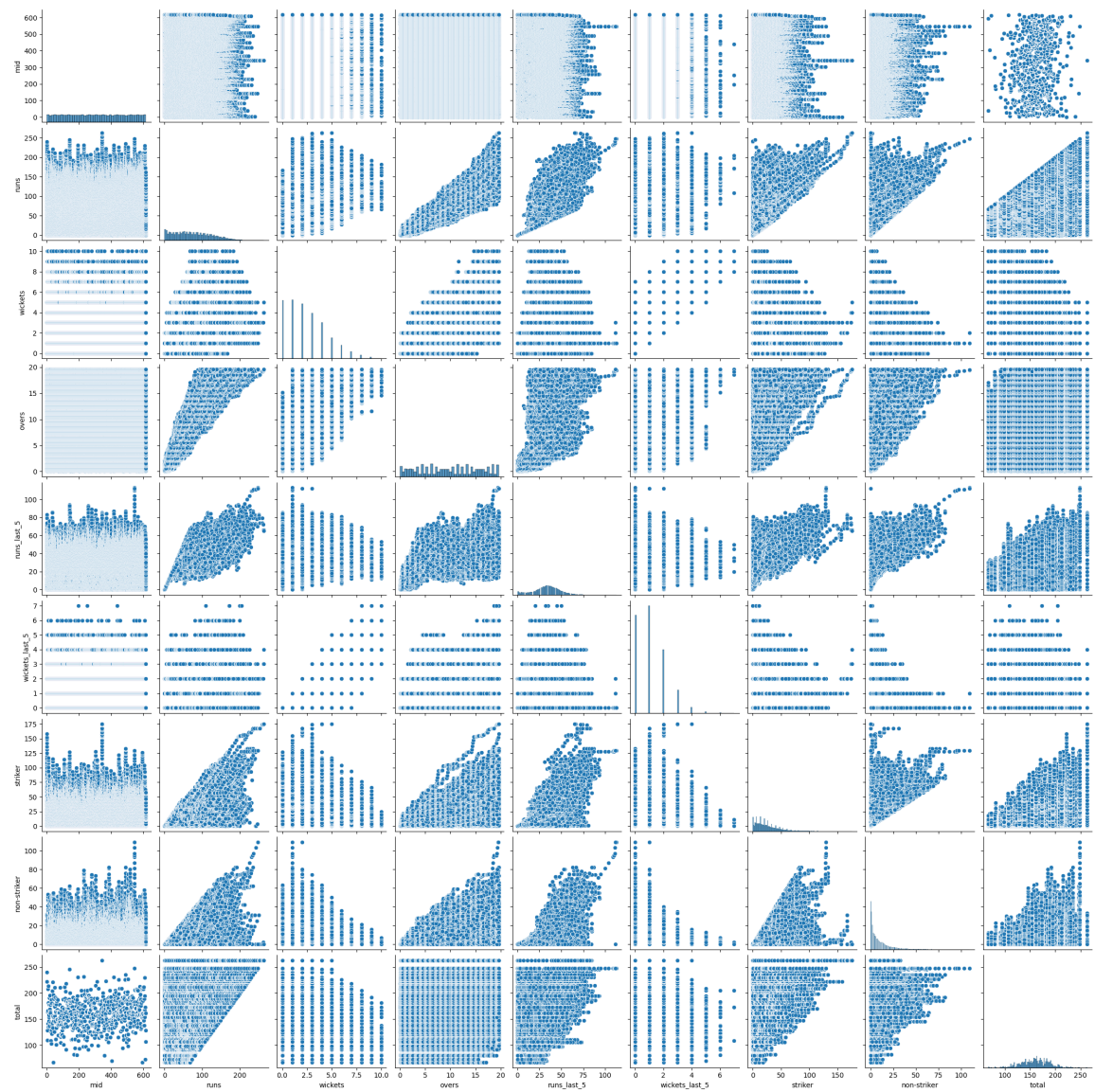
Out[3]:

	mid	date	venue	bat_team	bowl_team	batsman	bowler	runs	wickets	overs
0	1	2008-04-18	M Chinnaswamy Stadium	Kolkata Knight Riders	Royal Challengers Bangalore	SC Ganguly	P Kumar	1	0	0.1
1	1	2008-04-18	M Chinnaswamy Stadium	Kolkata Knight Riders	Royal Challengers Bangalore	BB McCullum	P Kumar	1	0	0.2
2	1	2008-04-18	M Chinnaswamy Stadium	Kolkata Knight Riders	Royal Challengers Bangalore	BB McCullum	P Kumar	2	0	0.2
3	1	2008-04-18	M Chinnaswamy Stadium	Kolkata Knight Riders	Royal Challengers Bangalore	BB McCullum	P Kumar	2	0	0.3
4	1	2008-04-18	M Chinnaswamy Stadium	Kolkata Knight Riders	Royal Challengers Bangalore	BB McCullum	P Kumar	2	0	0.4

```
In [4]: import matplotlib.pyplot as plt
%matplotlib inline
```

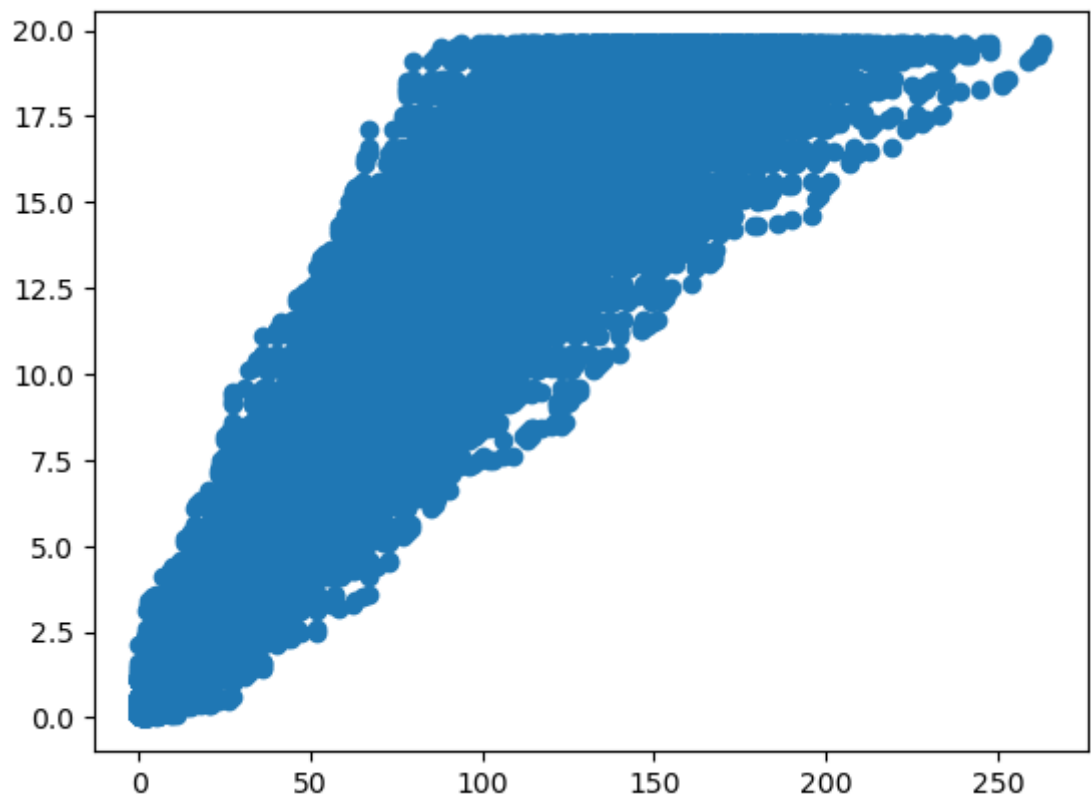
```
In [5]: import seaborn as sns
sns.pairplot(data = data)
```

Out[5]: <seaborn.axisgrid.PairGrid at 0x173bc920190>



```
In [8]: plt.scatter(data["runs"],data["overs"])
```

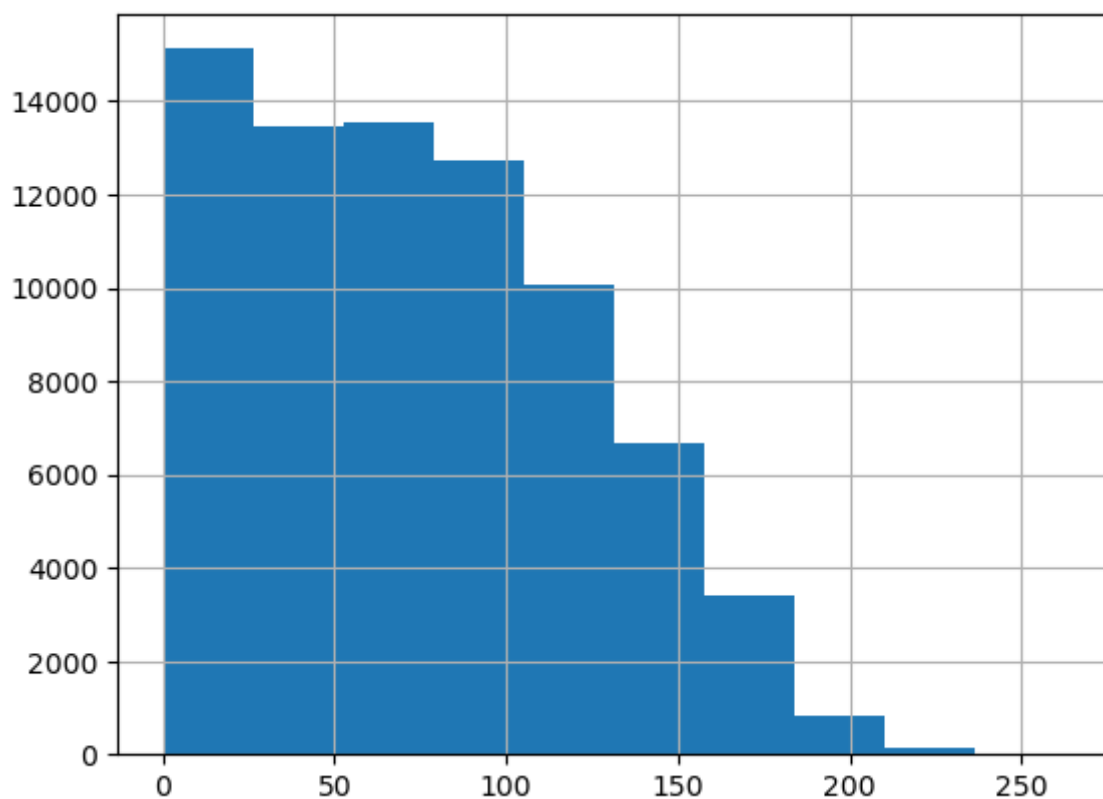
```
Out[8]: <matplotlib.collections.PathCollection at 0x18ab842bf50>
```



```
In [ ]: import seaborn as sns
sns.pairplot(data = data)
```

```
In [10]: data['runs'].hist()
```

```
Out[10]: <Axes: >
```



```
In [33]: x = data[["runs", "overs"]]  
y = data["total"]
```

```
In [34]: from sklearn.model_selection import train_test_split  
x_train , x_test , y_train , y_test = train_test_split (x,y, test_size= 0.2
```

```
In [35]: x_train
```

```
Out[35]:
```

	runs	overs
46728	56	6.4
9925	122	14.3
70361	0	0.3
57655	81	9.6
39789	31	3.6
...	...	...
71488	9	1.2
66517	4	0.2
47623	11	3.5
71678	114	11.4
46907	137	15.2

60811 rows × 2 columns

```
In [54]: x_test
```

```
Out[54]:
```

	runs	overs
58304	67	12.4
55667	95	10.3
5512	6	0.6
46454	16	1.6
46770	115	13.3
...	...	...
41541	43	7.4
49774	111	14.6
64201	42	5.1
50333	45	6.2
50953	73	9.2

15203 rows × 2 columns

```
In [36]: y_train
```

```
Out[36]:
```

46728	166
9925	178
70361	171
57655	181
39789	138
...	
71488	198
66517	175
47623	132
71678	187
46907	190

Name: total, Length: 60811, dtype: int64

```
In [55]: y_test
```

```
Out[55]:
```

58304	95
55667	209
5512	126
46454	170
46770	166
...	
41541	123
49774	132
64201	180
50333	172
50953	170

Name: total, Length: 15203, dtype: int64

```
In [37]: from sklearn.linear_model import LinearRegression
```

```
In [38]: clf = LinearRegression()
```

```
In [39]: clf.fit(x_train , y_train)
```

```
Out[39]: LinearRegression()
```

**In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook.**  
**On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.**

```
In [40]: clf.predict(x_train)
```

```
Out[40]: array([169.74373886, 174.58134975, 160.11527761, ..., 143.67001246,  
                191.78778531, 183.88326564])
```

```
In [57]: y_test
```

```
Out[57]: 58304      95  
         55667     209  
         5512      126  
         46454     170  
         46770     166  
         ...  
         41541     123  
         49774     132  
         64201     180  
         50333     172  
         50953     170  
         Name: total, Length: 15203, dtype: int64
```

```
In [42]: clf.score(x_test , y_test)
```

```
Out[42]: 0.4512862332286157
```

```
In [44]: from sklearn.ensemble import RandomForestClassifier  
         model = RandomForestClassifier(n_estimators=20)  
         model.fit(x_train, y_train)
```

```
Out[44]: RandomForestClassifier(n_estimators=20)
```

**In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook.**  
**On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.**

```
In [46]: model.score(x_test, y_test)
```

```
Out[46]: 0.035716634874695785
```

```
In [50]: y_predicted = model.predict(x_test)
```

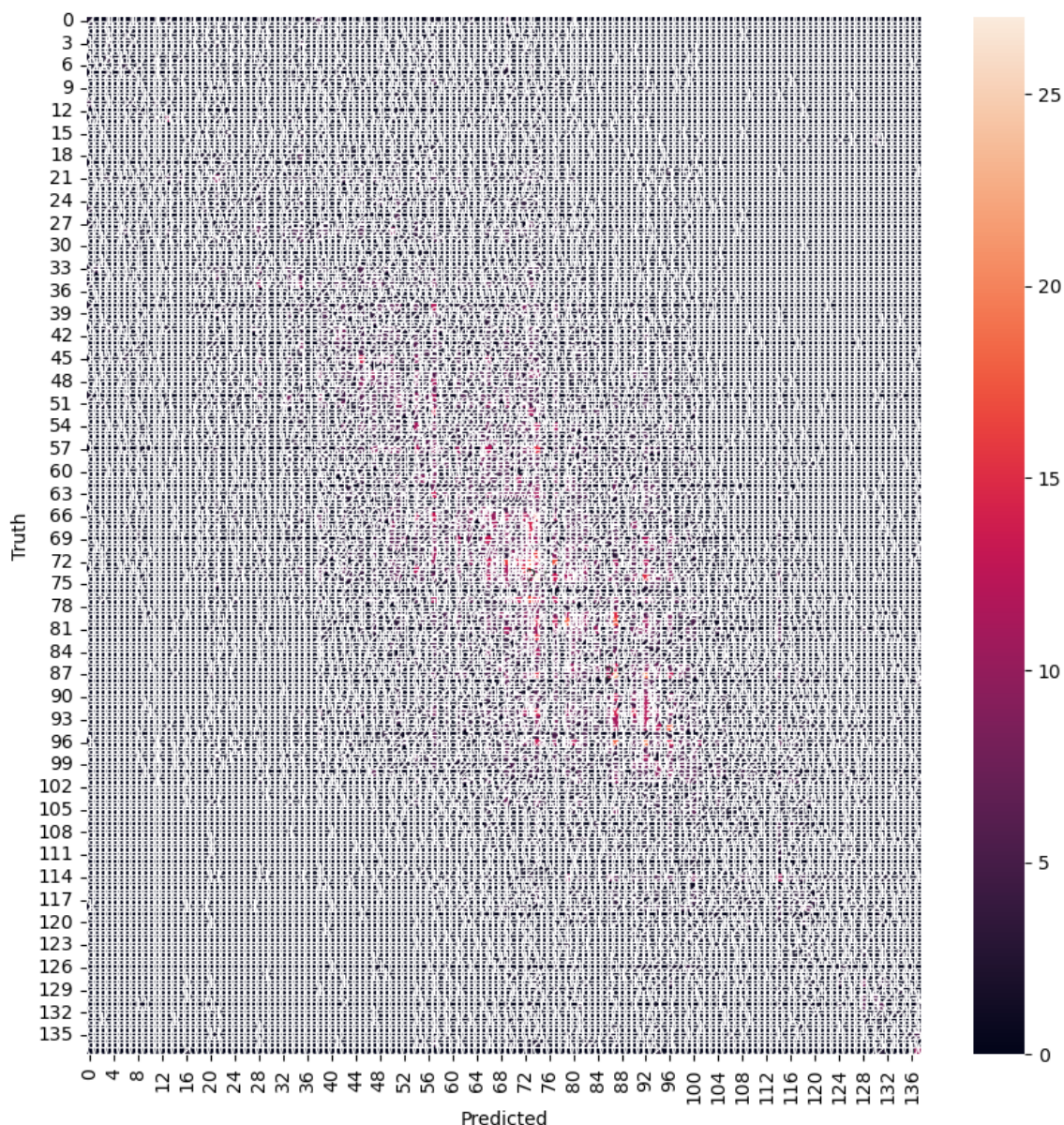


```
In [51]: from sklearn.metrics import confusion_matrix
cm = confusion_matrix(y_test, y_predicted)
cm
```

```
Out[51]: array([[ 2,  0,  0, ...,  0,  0,  0],
               [ 0,  0,  0, ...,  0,  0,  0],
               [ 0,  0,  0, ...,  0,  0,  0],
               ...,
               [ 0,  0,  0, ...,  2,  2,  5],
               [ 0,  0,  0, ...,  0,  0,  0],
               [ 0,  0,  0, ...,  0,  0, 10]], dtype=int64)
```

```
In [53]: %matplotlib inline
import matplotlib.pyplot as plt
import seaborn as sn
plt.figure(figsize=(10,10))
sn.heatmap(cm, annot=True)
plt.xlabel('Predicted')
plt.ylabel('Truth')
```

```
Out[53]: Text(95.7222222222221, 0.5, 'Truth')
```



```
In [ ]: import seaborn as sns
sns.pairplot(data = data)
```

```
In [ ]:
```