```
!pip install kaggle
```

```
Requirement already satisfied: kaggle in /usr/local/lib/python3.6/dist-packages (1.5
Requirement already satisfied: slugify in /usr/local/lib/python3.6/dist-packages (fr
Requirement already satisfied: certifi in /usr/local/lib/python3.6/dist-packages (fr
Requirement already satisfied: requests in /usr/local/lib/python3.6/dist-packages (fr
Requirement already satisfied: tqdm in /usr/local/lib/python3.6/dist-packages (from k
Requirement already satisfied: python-dateutil in /usr/local/lib/python3.6/dist-packa
Requirement already satisfied: six>=1.10 in /usr/local/lib/python3.6/dist-packages (f
Requirement already satisfied: urllib3<1.25,>=1.21.1 in /usr/local/lib/python3.6/dist
Requirement already satisfied: python-slugify in /usr/local/lib/python3.6/dist-packag
Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.6/dist-packages
Requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.6/dist-pac
Requirement already satisfied: text-unidecode>=1.3 in /usr/local/lib/python3.6/dist-p
```

```
from google.colab import files
files.upload()
```

Choose Files No file chosen Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.

```
Saving ipl2017.csv to ipl2017.csv
{'ipl2017.csv': b'mid,date,venue,bat_team,bowl_team,batsman,bowler,runs,wickets,overs
```

```
import pandas as pd
import numpy as np
```

```
df= pd.read_csv("ipl2017.csv")
df.head()
```

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

```
y=df['total']
```

```
x=df.drop(["date"],axis=1)
```

```
x
```

	mid	venue	bat_team	bowl_team	batsman	bowler	runs	wickets	ove
0	1	M Chinnaswamy Stadium	Kolkata Knight Riders	Royal Challengers Bangalore	SC Ganguly	P Kumar	1	0	0
1	1	M Chinnaswamy Stadium	Kolkata Knight Riders	Royal Challengers Bangalore	BB McCullum	P Kumar	1	0	0
2	1	M Chinnaswamy Stadium	Kolkata Knight Riders	Royal Challengers Bangalore	BB McCullum	P Kumar	2	0	0
3	1	M Chinnaswamy Stadium	Kolkata Knight Riders	Royal Challengers Bangalore	BB McCullum	P Kumar	2	0	0
4	1	M Chinnaswamy Stadium	Kolkata Knight Riders	Royal Challengers Bangalore	BB McCullum	P Kumar	2	0	0
...	...	...	...	...	...	...	...	...	
76009	617	Rajiv Gandhi International Stadium, Uppal	Mumbai Indians	Rising Pune Supergiant	KH Pandya	DT Christian	121	7	19
76010	617	Rajiv Gandhi International Stadium, Uppal	Mumbai Indians	Rising Pune Supergiant	KH Pandya	DT Christian	127	7	19

```
type(x)
```

```
pandas.core.frame.DataFrame
```

```
x.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 76014 entries, 0 to 76013
Data columns (total 14 columns):
 #   Column          Non-Null Count  Dtype  
---  ------          --------------  -----  
 0   mid             76014 non-null  int64  
 1   venue           76014 non-null  object 
 2   bat_team        76014 non-null  object 
 3   bowl_team       76014 non-null  object 
 4   batsman         76014 non-null  object 
 5   bowler          76014 non-null  object 
 6   runs            76014 non-null  int64  
 7   wickets         76014 non-null  int64  
 8   overs           76014 non-null  float64
 9   runs_last_5     76014 non-null  int64  
 10  wickets_last_5  76014 non-null  int64  
 11  striker         76014 non-null  int64  
 12  non-striker     76014 non-null  int64  
```

```
 13  total           76014 non-null  int64
dtypes: float64(1), int64(8), object(5)
memory usage: 8.1+ MB
```

```
x=pd.get_dummies(df,columns=["venue","bat_team","bowl_team","batsman","bowler","date"],dro
```

```
x.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 76014 entries, 0 to 76013
Columns: 1248 entries, mid to date_2017-05-21
dtypes: float64(1), int64(8), uint8(1239)
memory usage: 95.0 MB
```

```
x.head()
```

	mid	runs	wickets	overs	runs_last_5	wickets_last_5	striker	non-striker	total	v
0	1	1	0	0.1	1	0	0	0	222	
1	1	1	0	0.2	1	0	0	0	222	
2	1	2	0	0.2	2	0	0	0	222	
3	1	2	0	0.3	2	0	0	0	222	
4	1	2	0	0.4	2	0	0	0	222	

5 rows × 1248 columns

```
x.shape
```

```
(76014, 1248)
```

```
from sklearn.model_selection import train_test_split
```

```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.33,random_state=42)
```

```
from sklearn.ensemble import  RandomForestRegressor
```

```
model= RandomForestRegressor()
```

```
model.fit(x_train,y_train)
```

```
RandomForestRegressor(bootstrap=True, ccp_alpha=0.0, criterion='mse',
                      max_depth=None, max_features='auto', max_leaf_nodes=None,
```

```
                       max_samples=None, min_impurity_decrease=0.0,
                       min_impurity_split=None, min_samples_leaf=1,
                       min_samples_split=2, min_weight_fraction_leaf=0.0,
                       n_estimators=100, n_jobs=None, oob_score=False,
                       random_state=None, verbose=0, warm_start=False)
```

```
model.score(x_test,y_test)
```

```
1.0
```

```
y_predicted=model.predict(x_test)
```

```
from sklearn.metrics import confusion_matrix
cm=confusion_matrix(y_test,y_predicted)
cm
```

```
array([[66,  0,  0, ...,  0,  0,  0],
       [ 0, 25,  0, ...,  0,  0,  0],
       [ 0,  0, 20, ...,  0,  0,  0],
       ...,
       [ 0,  0,  0, ..., 47,  0,  0],
       [ 0,  0,  0, ...,  0, 33,  0],
       [ 0,  0,  0, ...,  0,  0, 42]])
```

```
%matplotlib inline
import matplotlib.pyplot as plt
import seaborn as sn
plt.figure(figsize=(10,7))
sn.heatmap(cm,annot=True)
plt.xlabel('Predicted')
plt.ylabel('Truth')
```

```
Text(69.0, 0.5, 'Truth')
```