**Below code for cleaning the raw data / Replacing with valid values**

Create database cricket\_data;

Use cricket\_data;

**---Imported raw data in below table**

Create or replace table Bowler\_data\_raw(

Row\_id varchar(100),

Overs varchar(100),

Maiden\_balls varchar(100),

Runs\_given varchar(100),

Wkts\_taken varchar(100),

Economy varchar(100),

Avg\_speed varchar(100),

Strike\_rate varchar(100),

Opposition varchar(100),

Ground varchar(100),

Start\_Date varchar(30),

Match\_ID varchar(100),

Bowler varchar(30),

Player\_ID Integer

);

Create or replace table batsman\_data\_raw(

Row\_number Integer,

Total\_score Varchar(30),

Runs\_by\_batsman Varchar(30),

Bowls\_faced Varchar(30),

Strike\_rate Varchar(30),

Fours Varchar(30),

Sixes Varchar(30),

Opposition Varchar(30),

Ground Varchar(30),

Start\_Date varchar(30),

Match\_ID Varchar(30),

Batsman Varchar(30),

Player\_ID Integer

);

Create or replace table ODI\_Match\_result\_raw(

Result varchar(100),

Margin varchar(100),

Toss varchar(100),

Opposition varchar(100),

Ground varchar(100),

Start\_Date varchar(100),

Match\_ID varchar(100),

Country varchar(100),

Country\_ID Integer

);

**---Below are the final tables with clean data.**

Create or replace table Bowler\_data as

select Row\_id,

REPLACE(Overs, '-', '0') as overs,

REPLACE(Maiden\_balls, '-', '0') as Maiden\_balls,

REPLACE(Runs\_given, '-', '0') as Runs\_given,

REPLACE(Wkts\_taken, '-', '0') as Wkts\_taken,

REPLACE(Economy, '-', '0') as Economy,

REPLACE(Avg\_speed, '-', '0') as Avg\_speed,

REPLACE(Strike\_rate, '-', '0') as Strike\_rate,

REPLACE(Opposition, 'v', '') as Opposition,

Ground,TO\_CHAR(TO\_DATE(START\_DATE,'DD-MON-YY'),'DD-MON-20YY') AS START\_DATE,Match\_ID,Bowler,Player\_ID from Bowler\_data\_raw;

Create or replace table Batsman\_data as

select Row\_number,

regexp\_replace(Total\_score, '[^0-9]+','0') as Total\_score,

REPLACE(Runs\_by\_batsman, '-', '0') as Runs\_by\_batsman,

REPLACE(Bowls\_faced, '-', '') as overs,

REPLACE(Strike\_rate, '-', '0') as Bowls\_faced,

REPLACE(Fours, '-', '0') as Fours,

REPLACE(Sixes, '-', '0') as Sixes,

REPLACE(Opposition, 'v', '') as Opposition,

Ground,TO\_CHAR(TO\_DATE(Start\_Date,'DD-MON-YY'),'DD-MON-20YY') AS START\_DATE,Match\_ID,Batsman,Player\_ID from batsman\_data\_raw;

Create or replace table ODI\_Match\_result as

select REPLACE(Result, '-', 'draw') as Result,

REPLACE(Margin, '-', 'Cancelled') as Margin,

REPLACE(Toss, '-', 'Cancelled or Aban') as Toss,

REPLACE(Opposition, 'v', '') as Opposition,

Ground,TO\_CHAR(TO\_DATE(Start\_Date,'DD-MON-YY'),'DD-MON-20YY') AS START\_DATE ,Match\_ID,Country,Country\_ID from ODI\_Match\_result\_raw;

select sum(Bowls\_faced) from Batsman\_data;

**---Below are master table for visualizing**

Create or replace table TOP\_SCORER as

select Batsman, sum(Runs\_by\_batsman) as Runs\_by\_batsman from Batsman\_data group by Batsman order by Runs\_by\_batsman desc limit 1;

Create or replace table MOST\_WICKETS\_TAKEN as

select Bowler, sum(Wkts\_taken) as Wkts\_taken from Bowler\_data group by Bowler order by Wkts\_taken desc limit 1;