Graded Assignment Solution
T5.1 Hive

----Create Database and Table in Hive and set the prompt to display current DB
hive> create database <your_id_DEMO>;

hive> use <your_id_DEMO>;

hive> set hive.cli.print.current.db=true;

hive> create table bat_tbl (playerID string,yearID int,stint int,teamID string,lgID string,G int,G_batting int,AB int,R int,H int,Two_B int,Three_B int,HR int,RBI int,SB int,CS int,BB int,SO int,IBB int,HBP int,SH int,SF int,GIDP int,G_old int)
row format delimited
fields terminated by ','
lines terminated by '\n'
tblproperties("skip.header.line.count"="1");

--- Load the Data from csv file to table temp_cricket and set display of headers as true
hive> load data local inpath '/home/data/Class_8/Batting.csv' into table bat_tbl;

hive> set hive.cli.print.header=true;

1. Display year and maximum runs scored in each year.

hive> SELECT yearID, max(R) FROM bat_tbl GROUP BY yearID;

2. Display player_id,year and maximum runs scored in each year.

hive> select playerid, yearid, max(r) from bat_tbl group by playerid, yearid;

3. Find out each and every year how many centuries was scored.

hive> select yearID, count(*) from bat_tbl where R >= 100 GROUP BY yearID;
OR
hive> select yearid, count(r) from bat_tbl where r>=100 group by yearid order by yearid desc;

 4. Find out how many centuries was scored in each and every year by the each and every different TeamID
hive> select yearID, teamid, count(*) from bat_tbl where R >= 100 GROUP BY yearID, teamid;