--Q-2-a he win records (percentage win and total wins) for each team by year and gender, excluding ties, matches with no result

select

WinningTeam as Team,

Gender,

YEAR(Date) as year,

COUNT(*) AS TotalMatches,

SUM(CASE WHEN WinningTeam=Team1 THEN 1 ELSE 0 END) AS TotalWins,

SUM(CASE WHEN WinningTeam=Team1 THEN 1 ELSE 0 END)*100/count(*) as Win_percentage from df_match_detail

where WinningTeam !='no result'

group by WinningTeam, YEAR(Date), Gender;

--Q-b-2 Which male and female teams had the highest win percentages in 2019?

select

WinningTeam as Team,

Gender,

COUNT(*) AS TotalMatches,

SUM(CASE WHEN WinningTeam=Team1 THEN 1 ELSE 0 END) AS TotalWins,

SUM(CASE WHEN WinningTeam=Team1 THEN 1 ELSE 0 END)*100/count(*) as Win_percentage from df_match_detail

where WinningTeam !='no result' and YEAR(Date)=2019

group by WinningTeam, Gender

order by Win_percentage desc;

-- ans is many teams have 100% win ratio

--Q-2-c Which players had the highest strike rate as batsmen in 2019?

select Batter,

sum(Runs_by_Batter) Total_runs,

count(Batter) as No_of_balls,

sum(Runs_by_Batter)*100/count(Batter) as strike_rate

from

df_innigs join df_match_detail

on df_innigs.MatchId=df_match_detail.MatchId

where Runs_by_Extras=0 and YEAR(Date)=2019

group by Batter

order by strike_rate desc;

-- SN thakur is highest strike rate

--question 3

if in future new file come, we make configuration in sql and then we scarap this file using python script or Azure Data factory from any loaction and put into Datalake and then we process the data using pyspark or databriks and put into azure sql or any loaction.

Multiple Sources -> Azure Data Factory-ADF(Ingestion) -> ADLS Gen

2(storage) -> Azure Databricks / Synapse(processing) -> Azure SQL /

Cosmos DB(serving)

--question 4

Daily I am learning something new in sql or python. During project sometimes I processing with long method but everytime I find somthing short or effective way. More I do research I can find best solution of the problem.