PROJECT – CRICSHEET MATCH ANALYSIS

PROJECT OVERVIEW

Use selenium to scrape JSON files of different cricket match types (ODI, T20, Test), store the data in sql tables, and create a power bi dashboard for data insights and perform EDA and SQL queries for data analysis.

APPROACH

1. Database creation

2. Create tables with required columns and constraints

3. Using selenium to fetch the data from web page

4. Read the data from json and update into database

5. Write sql queries for data insights

6. Data visualization for data insights

7. Integrate SQL in power bi and create the power bi dashboard for key performance.

DATABASE SCHEMA

1.Teams table

Column	Type	Default Value	Nullable	Character Set	Collation	Privileges	Extra
team_id	bigint		NO			select,insert,update,references	auto_increment
team_name	varchar(255)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references	

2. Players table

Column	Type	Default Value	Nullable	Character Set	Collation	Privileges	Extra
player_id	bigint		NO			select,insert,update,references	auto_increment
player_name	varchar(255)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references	
team_id	bigint		YES			select,insert,update,references	

3. Matches table

Column	Type	Default Value	Nullable	Character Set	Collation	Privileges	Extra
city	varchar(255)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references	
date	date		YES			select,insert,update,references	
event_name	varchar(255)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references	
gender	varchar(50)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references	
match_id	bigint		NO			select,insert,update,references	auto_increment
match_number	int		YES			select,insert,update,references	
match_type	varchar(50)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references	
match_type_number	int		YES			select,insert,update,references	
outcome_by_runs	int		YES			select,insert,update,references	
outcome_by_wickets	int		YES			select,insert,update,references	
outcome_winner	varchar(255)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references	
season	varchar(50)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references	
team_type	varchar(50)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references	
toss_decision	varchar(50)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references	
toss_winner	varchar(255)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references	

4.Deliveries table

Column	Type	Default Value	Nullable	Character Set	Collation	Privileges	Extra
batter_id	bigint		YES			select,insert,update,references	
bowler_id	bigint		YES			select,insert,update,references	
delivery_id	bigint		NO			select,insert,update,references	auto_increment
match_id	bigint		YES			select,insert,update,references	
non_striker_id	bigint		YES			select,insert,update,references	
over_number	int		YES			select,insert,update,references	
runs_batter	int		YES			select,insert,update,references	
runs_extras	int		YES			select,insert,update,references	
runs_total	int		YES			select,insert,update,references	
wicket_kind	varchar(50)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references	
wicket_player_out	varchar(255)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references	

CODE FLOW

1. Create the webdriver connection for chrome
2. After page is loaded, scroll down to page end
3. Then search the respective hyperlink using the xpath
4. Click the hyperlink and download the data files
5. Extract the zip files using zipfile package
6. Iterate all files in the extracted folder and call another method to read the json file
7. This method iterate the each json keys and fetch the values. Then update the values in database

RESULTS

POWER BI DASHBOARD