ESPN Cricket Data Scraping

Requirements:-

```
BeautifulSoup -
requests -
sqlite3 - A light-weight database to store the information fetched for further querying and
analyis,If automated Jog to be done by using like Flask (Also converted to CSV files)
Others - argparse, logging
```

Program Structure:-

```
|-- CSV_Converted_Data
|   |-- Batting_Stats_Odi.csv
|   |-- Batting_Stats_T20.csv
|   |-- Bowling_Stats_Odi.csv
|   |-- Bowling_Stats_T20.csv
|   |-- Countries.csv
|   `-- Players.csv
|-- DEFAULT_DB.sqlite
|-- ESPN_Cricket_Data_Scraper.py
|-- ESPN_Cricket_Data_Scraping.md
|-- StoredDB.sqlite
`-- espn_cricket_data_scraper.log
```

About:-

This program provides user the flexibility to choose the countries of whose player statistics are to be scraped and stored in database along with either choosing ODI/T20 or both.

To Execute:- Now for terminal (command line)

```
>>> ESPN_Cricket_Data_Scraper.py [-h] [-d DATABASE_NAME] [-t TYPE_OF_MATCH] [-c [COUNTRIES
[COUNTRIES     ...]]]
```

This with Default values:- Scrapes all players statistics(bowling and fielding) from major 11 playing nations in both ODI and T20 format.

```
example: python ESPN_Cricket_Data_Scraper.py
-d :'DEFAULT_DB'
-t :'ALL'
-c :'ALL'
```

Usage with custom entries:-

```
Example:- (Run)
>>> python ESPN_Cricket_Data_Scraper.py -d StoredDB  -t ODI  -c [india, pakistan]

-d :
-t : ODI/T20/ALL
-c  [australia,bangladesh,england,india,new-zealand,pakistan,south-africa,sri-lanka,west-
indies,zimbabwe,afghanistan]  ##select based on requirement in list
```

The execution of above command would scrape the website for batting/bowling stats in ODI of all players from india and pakistan and parse and store results in 'StoredDB.sqlite' **database** which could be queried using below table structure.

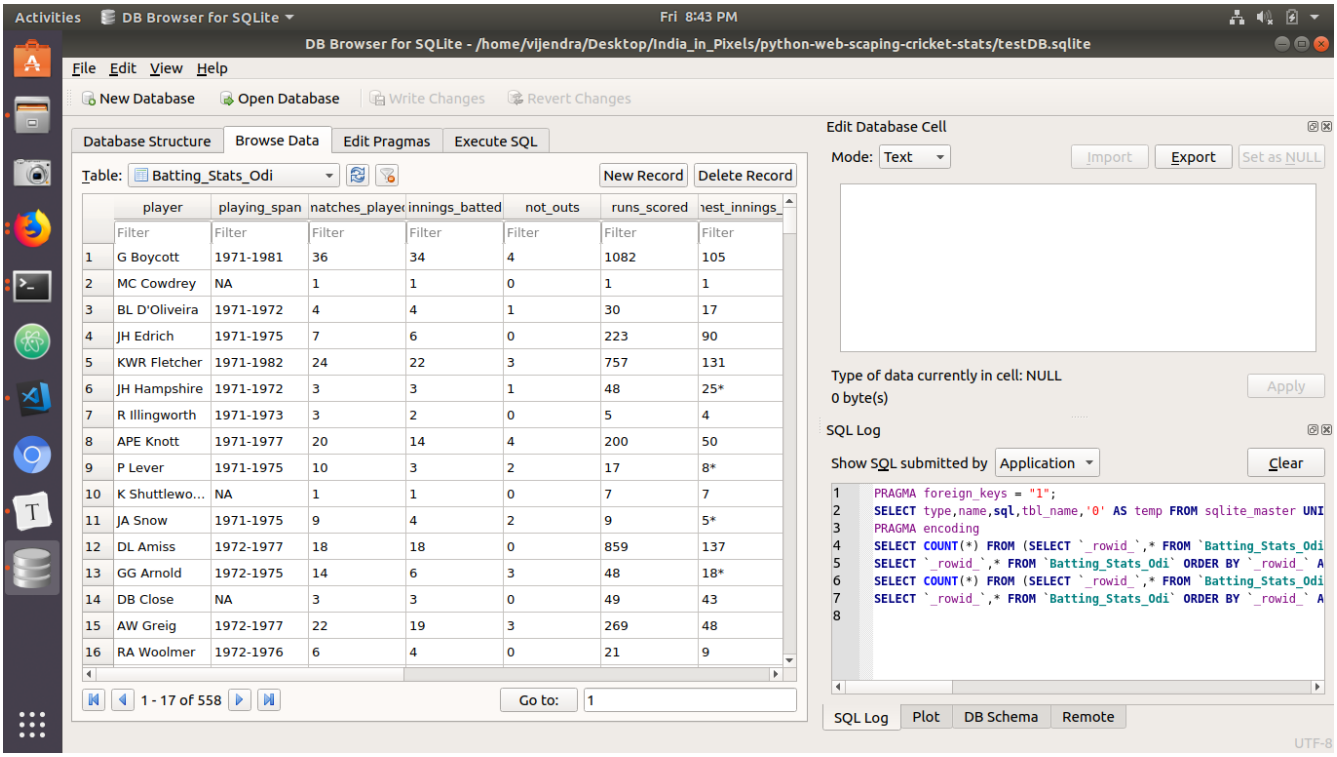
The **sqlite database** would have the below table structure where the scraped data is parsed and stored in a normalized way:

```
Countries > country_id PRIMARY KEY,country

Players > country_id ,player_id UNIQUE,player ,odi_cap ,t20_cap

Batting_Stats_Odi > player ,playing_span ,matches_played , innings_batted ,not_outs ,
runs_scored ,highest_innings_score ,batting_average , balls_faced ,batting_strike_rate
,hundreds_scored ,scores_between_50_and_99 , ducks_scored ,boundary_fours ,boundary_sixes

Bowling_Stats_Odi > player ,playing_span ,matches_played ,innings_bowled_in , overs_bowled
,balls_bowled ,runs_conceded ,maidens_earned ,wickets_taken , best_bowling_in_an_innings
,bowling_average ,economy_rate ,bowling_strike_rate , four_wkts_exactly_in_an_inns
,five_wickets_in_an_inns

Batting_Stats_T20 > player ,playing_span ,matches_played , innings_batted ,not_outs ,
runs_scored ,highest_innings_score ,batting_average , balls_faced ,batting_strike_rate
,hundreds_scored ,scores_between_50_and_99 ,ducks_scored , boundary_fours ,boundary_sixes

Bowling_Stats_T20 > player , playing_span ,matches_played ,innings_bowled_in , overs_bowled
,balls_bowled ,runs_conceded ,maidens_earned ,wickets_taken , best_bowling_in_an_innings
,bowling_average ,economy_rate ,bowling_strike_rate , four_wkts_exactly_in_an_inns
,five_wickets_in_an_inns
```

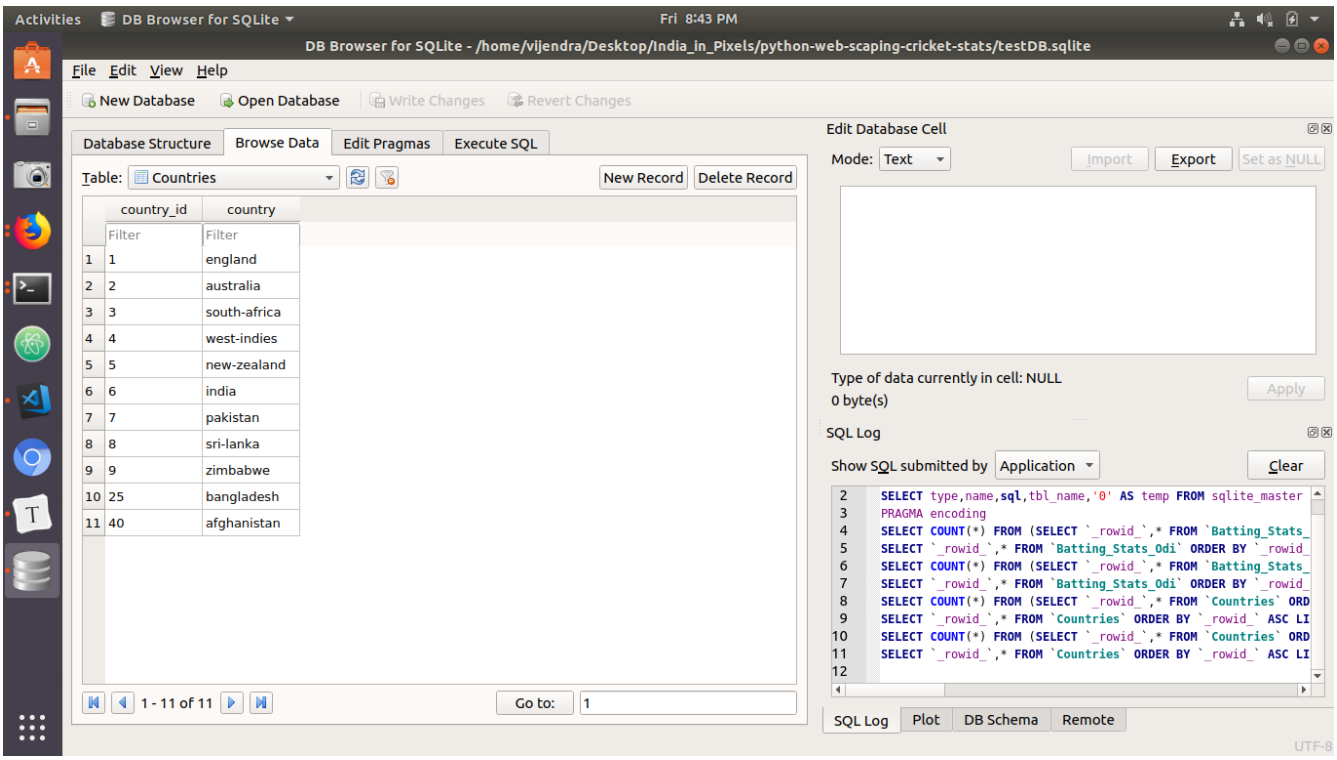
Once the parser completes execution we can query the required information for further **analyis** referring the table structure provided above and also can be convet into csv file for data analysis for pandas point of view.

Some Snap-shoots of scraped Cricket data:-

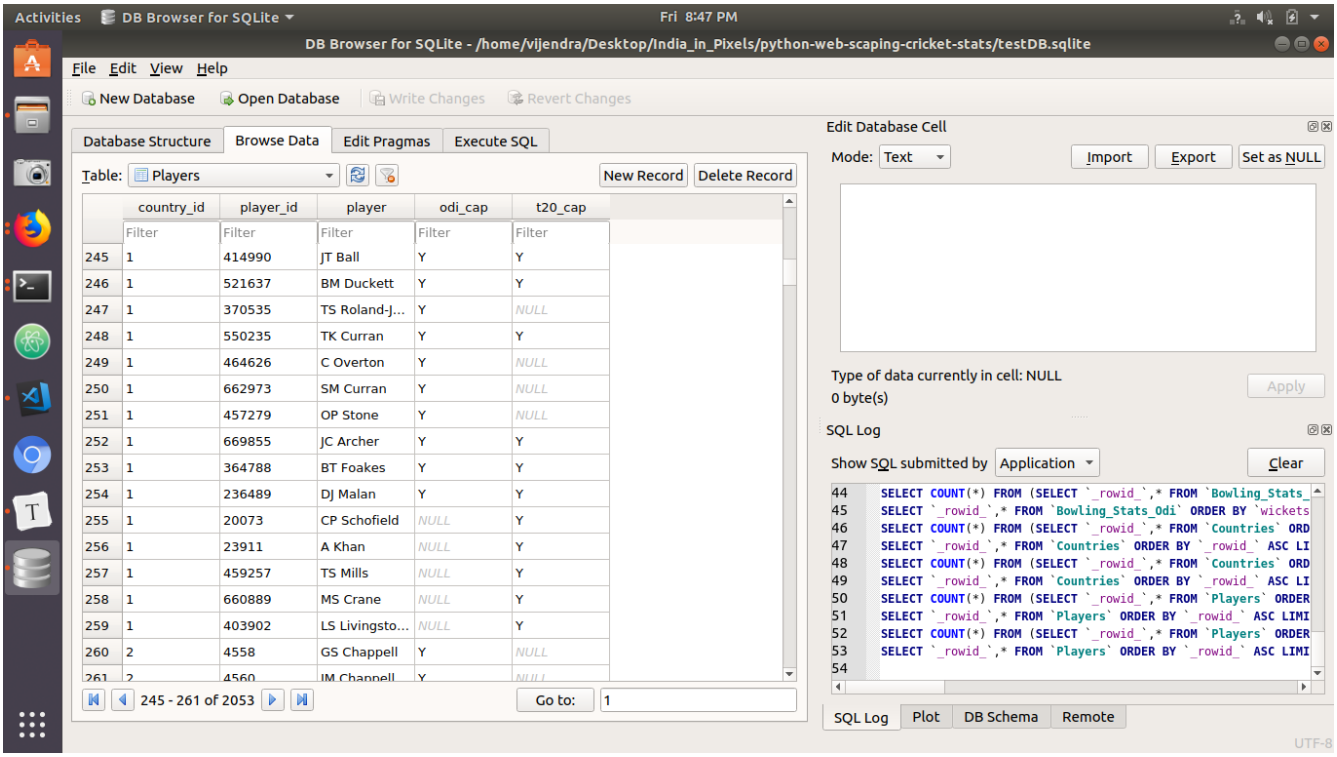
1. *Bating stats of ODI:-*



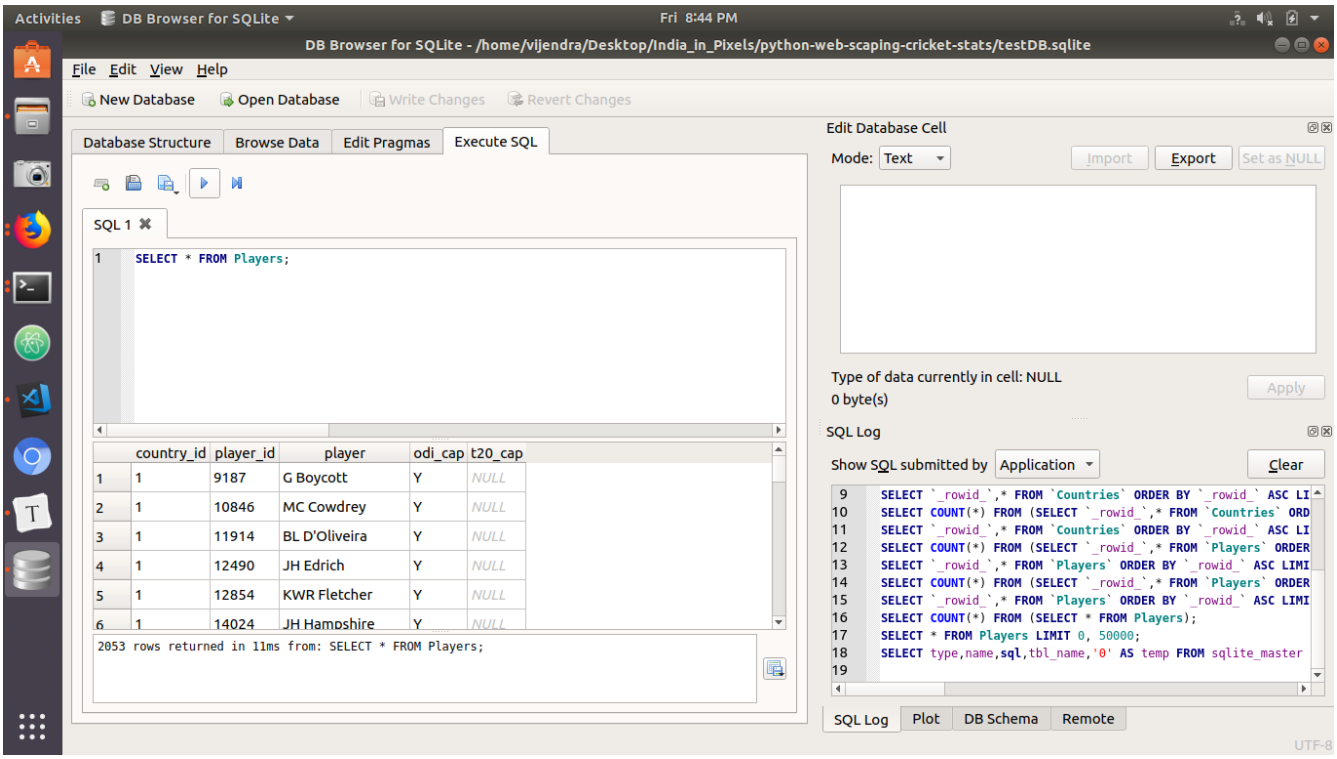
2. *Countries Name:-*

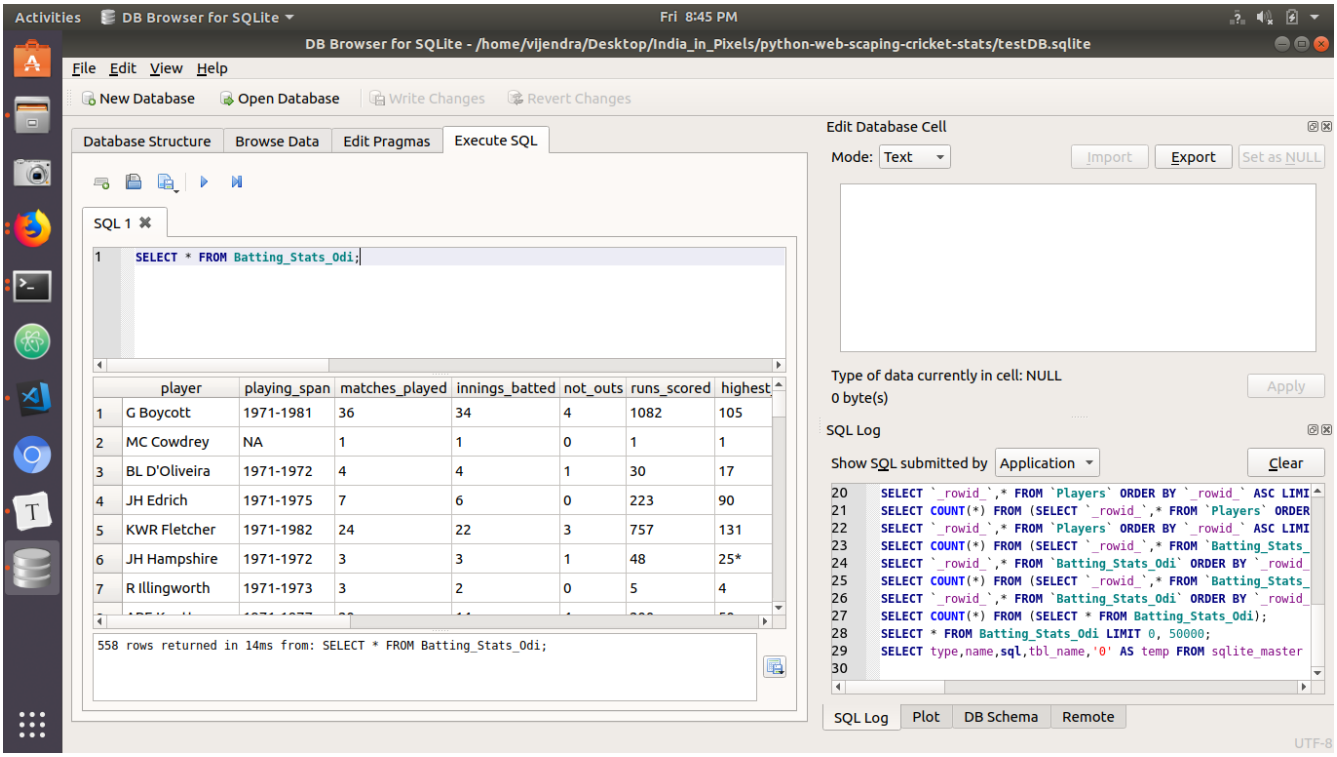


3. *Players Name:-*

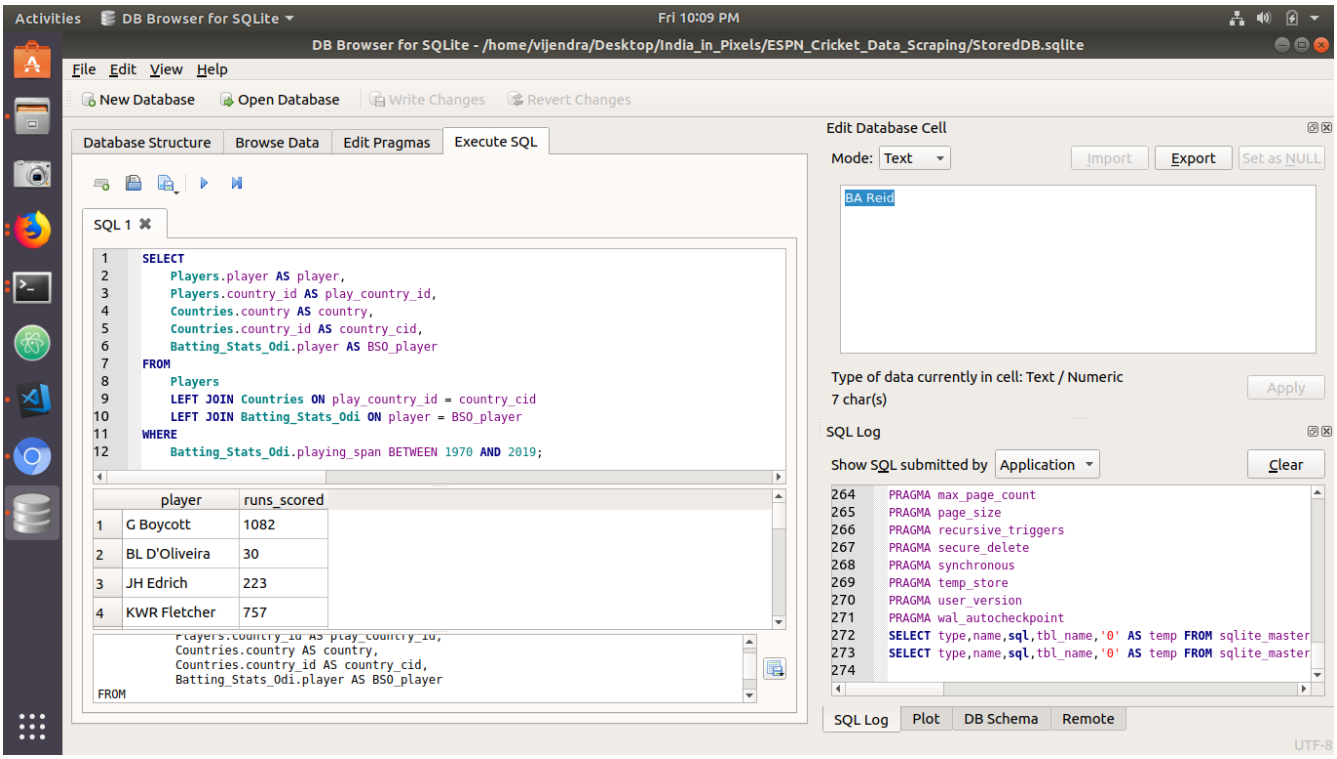


Querying From DataBase:-





Due to time constrant I couldn't join all three tables Batting_Stats_Odi, Countries and Players, but I definately will do it. Following will be final result of all three column of csv file whci can be easily **Export**.



Note:- There are a lot to do , but for now it's done. (Every details like maximum run, Maximum wickets, Hundreds, Fifties, Sixes, Fours etc.)

Here I didn't scrape IPL data which is an easy task.

This is my approach for data scraping it's interesting, but data analysis will be much more interesting. I hope, I will get the chance to work for such start-up and level up my skills.