PROJECT REPORT

ODI Cricket Data Analytics

Contents

Abstract

Cricket is truly the most popular sport, and is an outdoor game that is enjoyed by everyone. The game is played by two teams of 11 players, each with a bat and a ball. Team who scores more runs, wins the game. There are various formats of cricket i.e. Test matches, One Day Internationals (ODIs) and T20 Internationals which are unlimited, 50 and 20 over matches respectively. Umpires are also present on ground for fair decision and to stop arguments. The International Cricket Council (ICC) is the world governing body of cricket. Headquartered in Dubai, United Arab Emirates, its members are 108 national associations, with 12 Full Members and 96 Associate Members. Founded in 1909 as the Imperial Cricket Conference, it was renamed the International Cricket Conference in 1965, and took up its current name in 1987. Yearly expected revenue of ICC is $2.5-2.7 Billion in which BCCI(The Board of Control for Cricket in India) contribute most.

1 General Descriptions

1.1 Product Perspective & Problem Statement

Cricket has a huge role in the entertainment and leisure of Indian households. Its cultural popularity has led to the development of cricket as an industry that plays a crucial role in the economy. BCCI, which is the controlling body of cricket in India was worth a colossal ₹14,489 Cr. at the end of the financial year 2018-19,this amount is increasing year by year and therefore, there is no denying that cricket has been a huge player in the economy, not just in India, but the entire world, with many domestic as well as international events. Cricket increases the consumption spending of the people in the economy

who spend on the tickets for the matches, parking services, transportation, and hospitality. it provides a helping-hand to the Government by generating large-scale employment, including but not limited to medical teams, support staff, marketing team, security personnel, etc. An average cricket player earns more than a movie actor and also pays good amount of tax to Government.

The objective of the project is to perform data visualization techniques to understand the insight of the data. **'The data set contains the data of the ODI Matches(Men's Players) from 2013 to mid of 2019. Try to find insights in the data based on the columns in the dataset. Find as many insights and prepare a presentable Dashboard'**

1.2 Tools used

Business Intelligence tools and Snowflake(cloud), SQL, MS Excel and Power BI are used to build the whole framework.

2. Pre-Processing

For this research, we have used data set contains 'three tables Batsman_data, Bowler_data, Team_Result of the ODI matches from 2013 to mid of 2019 of Men's player' which has been provided by '**Ineuron**.**ai**'.This dataset contains detailed information of many Batsman, Bowler with respect to different countries.
A lot of pre-processing was required to handle missing values, noise and outliers. We have many different attributes for this research: ***Batsman Runs, Strike Rate, Average, Wickets, Economy Rate, Opposition Team, Match Result etc.***

3.Methodology

First created a database Cricket and uploaded dataset(csv file) on Snowflake of all 3 tables name batsman_data(11,150 rows), bowler_data(11,119 rows), match_result(1,323 rows) and renamed some columns name like SR with Strike_Rate , BF with bowl_faced while uploading data in tables.

```
Create database cricket
Use cricket
```

After loading our data set checked batsman_data and columns:-

```
select * from batsman_data;
Columns name – col1, Batsman_score, Runs, Bowl_faced,
Strike_rate, no_of_4s, no_of_6s, Opposition_team, Ground,
match_Date, Match_ID, Batsman, Player_ID
```

We have 11,150 rows and 13 columns but few columns are either not relevant or not having correct information so we decided to remove such columns from table

```
alter table batsman_data  drop col1;  /* not relevant */

alter table batsman_data drop strike_rate; /* we generally
calculate strike_rate for a batsman not in a match only but as
a career . so decide to remove it */

delete from batsman_data where BATSMAN_SCORE in
('DNB','TDNB','sub','absent'); /* If batsman is not playing
game then it won't consider in inning so decide to remove
these rows */
```

Data is cleaned but we need to add some more columns for visualization. When we calculate average we need no of innings where batsman got out so we created one column out/not out information. Created columns through opposition team and match_date column.
Named this new table batsman_new and downloaded it in .csv format.

```
with batsman_new as
(select * ,case
when BATSMAN_SCORE like '%*' then 'Not Out'
When BATSMAN_SCORE not like '%*' then 'Out'
end
as Out_or_NotOut,
substr(opposition_team,3,length(opposition_team))
as team_OPPOSITION,
case
when length(year(match_date))=1 then concat(200 ,
year(match_date))
when length(year(match_date))=2 and year(match_date)!=99 then
concat(20 , year(match_date))
when year(match_date) =99 then concat(19 , year(match_date))
end
as match_year
from batsman_data)
```

Checked bowler_data table and it's columns .

```
select * from bowler_data;
Columns name - col1, Overs, Maiden_overs, Runs, Wickets,
Economy, Average, Strike_Rate, Opposition_team, Ground,
     match_Date, Match_ID, Bowler, Player_ID
```

We have 11,119 rows and 14 columns in bolwer_data but few columns are either not relevant or not having correct information so we decided to remove such columns from table

```
alter table bowler_data drop col2;

delete from bowler_data where overs like '-' ; /* `-` was missing values in
the overs column */

alter table bowler_data drop average, strike_rate, economy;
```

Data is cleaned but we need to add some more columns for visualization.
Created new columns through opposition team and match_date as they were not in proper format.
View this new bowler_new table and downloaded it in .csv format.

```
 WITH bowler_new as
(select * ,substr(opposition,3,length(opposition)) as OPPOSITION_TEAM ,
SUBSTR(MATCH_DATE,-4,LENGTH(MATCH_DATE))
 as year  from bowler_data)
select * from bowler_new;
```

Checked match_result table and it's columns . we have 1323 rows and 12 columns in table.

```
select * from match_result;
```

Columns name – col1, Result, Win_Margin, BR, Toss, Bat, Opposition_team, Ground, Start Date, Match_ID, Country, Country_ID

Some columns are not useful and some rows are having missing values so decided to delete from data set.

```
alter table match_result drop col3 , br;

delete from match_result where result not in ('won','lost','n/r','tied');
```

Added new columns and downloaded it in .csv format.

```
select * , SUBSTR(start_date,-4,LENGTH(start_DATE)) as match_year,
substr(opposition_team,3,length(opposition_team)) as team_OPPOSITION
from match_result;
```

Now our dataset is cleaned for Visualisation . I will upload all three tables on Power Bi .

5.Conclusion(Insights)

--------- Insights from Batsman_New table ---------

1. Virat Kohli is the Highest Run Scorer in Data Set
2. Virat Kohli hit maximum 6s in ODI
3. MS Dhoni hit maximum 4s in ODI
4. Imam-ul-haq is having maximum average
5. Andre Russell is having maximum strike rate
6 . Virat Kohli scored maximum 100s and 50s in ODI
7 . MS Dhoni played maximum innings in ODI
8 . Highest score in ODI is 264 scored by Rohit Sharma

--------- Insights from Bowler_New table ---------

1. Lasith Malinga is highest wicket taking bowler in ODI
2. Lasith Malinga took more than 5 wickets 8 times which is highest in ODI.
3. Highest maiden overs bowled by Ma Shrafe Mortaza in ODI
4. Muzzeb Ur Rahman is having lowest economy rate in ODI
5. Imran Tahir took maximum 7 wicket in a inning which is highest in ODI.

6 . Number of matches is increasing continuously means popularity of cricket is increasing but there is sudden down fall after 2018 . May be teams started playing other format of cricket like T20.

--------- Insights from Match_Result table ---------

1. India has the highest winning percentage record to win the game
2. West Indies has highest losing percentage to lost the game
3. Generally team who bat at 2nd(chassing the score) have higher chances to win the game
4. Maximum team have higher chances to win the game when they lost the Toss.
5. Maximum matches played at Dhaka Ground .