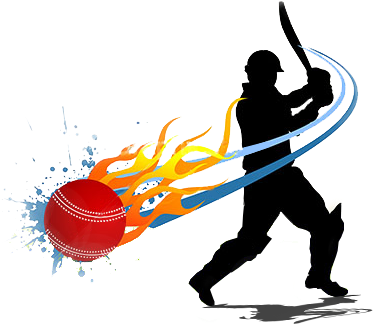
**PROJECT REPORT**

**ON**

**CRICKET DATA ANALYTICS DASHBOARD**



**Project submitted to**

**INEURON**

**Submitted by-**

**Aditya Gautam**

**FSDA 1.0**

**CONTENTS**

**CHAPTERS PAGE NO.**

1.Chapter I

1.1 Objectives of the study 1

1.2 Information about the Data 2-3

2. Chapter II

2.1. Architecture of Data Analytics Project 4-5

3. Chapter III

3.1 Analysis of the Data 6

4. Chapter IV

4.1 Data Visualisation 7-9

5. Chapter V

5.1 Conclusion 10

*Chapter I*

**1.1 *Objectives of the study***

The objectives of these studies are-

1. To load the data into snowflake and clean the data.
2. To create an interactive dashboard and bring out the important insight about the data.

**1.2 *Information about the Data***

The data set contains the data of the ODI matches from 2013 to mid of 2019. Following is the description of the dataset.

**Batsman Data**:

1. **Row ID**
2. **Bat1** – Score of the batsmen, DNB means that batsman has not participated in the match
3. **Runs** – Runs made in the match by the batsmen
4. **BF** – Balls Faced
5. **SR** – Strike rate
6. **4s** – No. of 4’s
7. **6s** – No of 6’s
8. **Opposition** – The opponent team the batsman played with
9. **Batsman** – Name of the batsman
10. **Ground** – The location of the cricket stadium
11. **Start Date** – The date on which the ODI was played
12. **MatchID** – unique id of the match
13. **Player ID** – Unique ID of the player

**Bowlers Data:**

1. **Row id**
2. **Overs –** The Over bowled by the Person
3. **Mdns –** Maiden balls
4. **Runs –** Runs given
5. **Wkts –** Wickets taken
6. **Econ –** Economy
7. **Ave –** Average Ball Speed
8. **SR –** Strike rate
9. **Opposition** – The opponent team the Bowler played with
10. **Bowler** – Name of the Bowler
11. **Ground** – The location of the cricket stadium
12. **Start Date** – The date on which the ODI was played
13. **MatchID** – unique id of the match

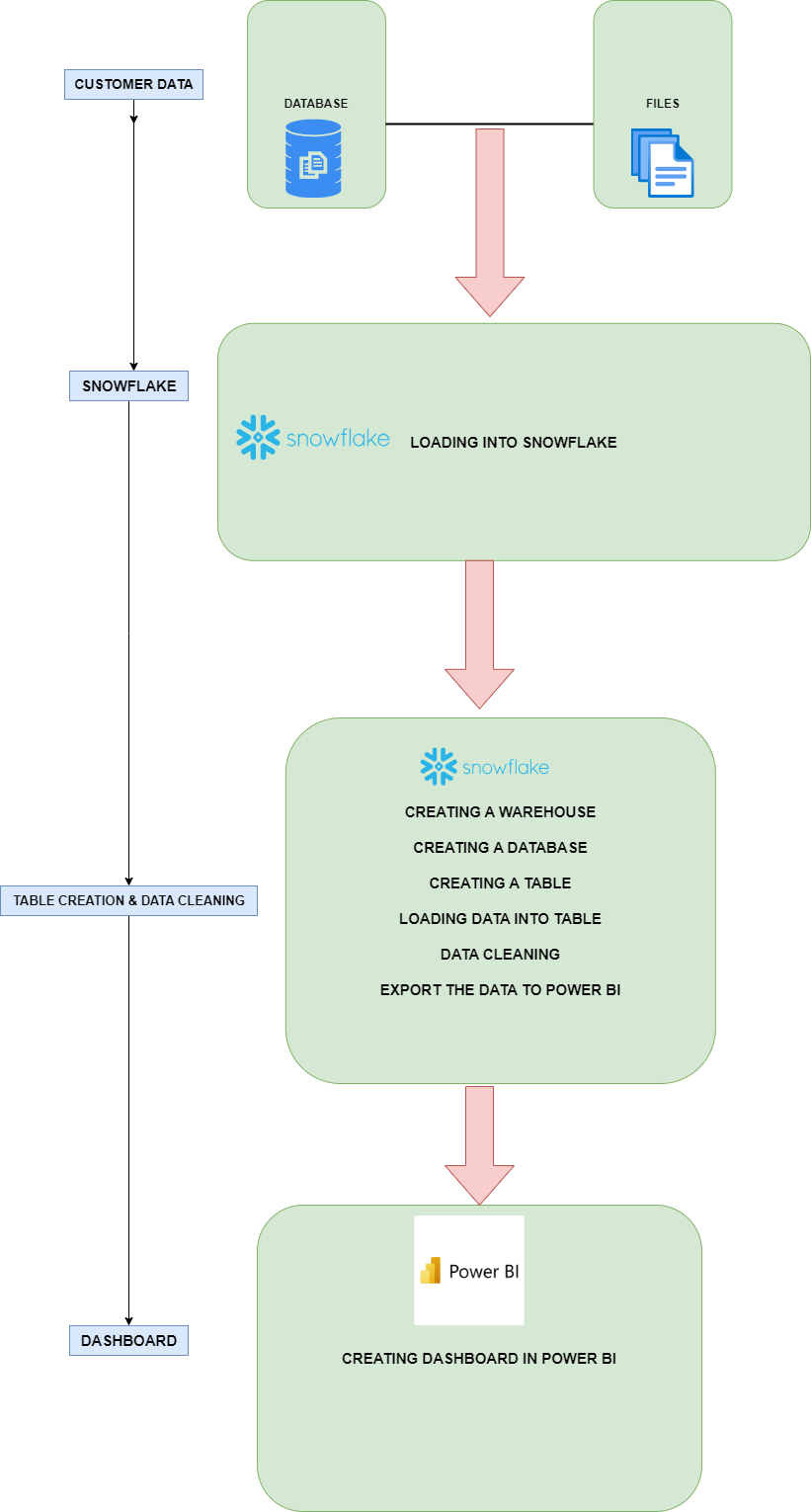
**ODI Match Results**

1. **Irrelavant – You can remove the column**
2. **Result –** the result of the match – Won or lost or draw or Not played
3. **Margin –** The margin by which the team has won
4. **BR – Irrelavant you can remove the column**
5. **Toss –** Information on if the team won or lost the toss
6. **Opposition** – The opponent team the batsman played with
7. **Ground** – The location of the cricket stadium
8. **Start Date** – The date on which the ODI was played
9. **MatchID** – unique id of the match
10. **Country –** The main Country
11. **Country ID –** The Unique id Of the country

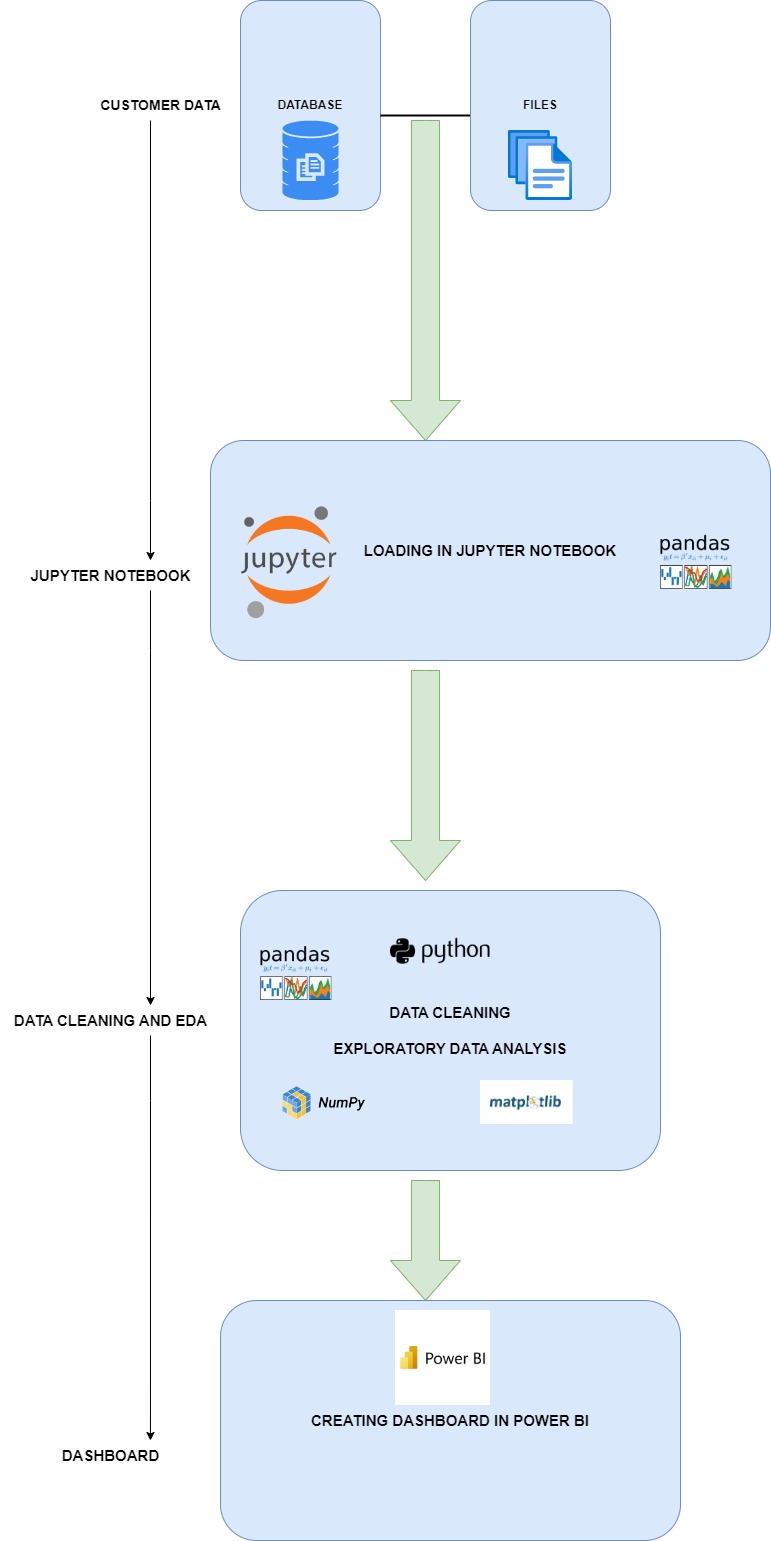
*Chapter II*

* 1. ***Architecture of Data Analytics Project***

**(2.1.1.) Complete Process In Snowflake**

****

**(2.1.2.) Complete Process In Python**



*Chapter III*

**(3.1) Analysis of the Data**

**Steps:**

(3.1.1.) **Data Cleaning using SQL**

1. First, we gather the data from the various sources, if client is providing the data then we directly stored in our database, but it is a personal project so we collect the data from the web sources and stored in a database by creating a table in the database.
2. Secondly, we clean the data such as removing unwanted characters from the data and replacing with zero, transforming the data into required data type and so on and so forth.
3. After cleaning the data, we import the clean data into Power BI for creating the Dashboard.
4. Finally, we create an interactive Dashboard using Power BI.

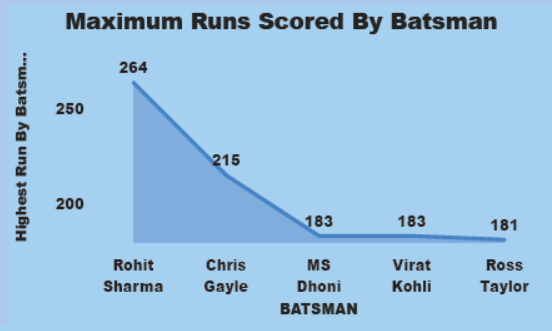
**(3.1.2.) Data Cleaning using Python**

1. First, we gather the data from the various sources, if client is providing the data then we directly stored in our database, but it is a personal project so we collect the data from the web sources and stored in a database by creating a table in the database.
2. Secondly, we export the data from the database and import the data using pandas library into jupyter notebook for further data cleaning process.
3. Thirdly, we see for missing values, unwanted special characters, and transform the data type into required data type.
4. After cleaning the data, we export the data from jupyter notebook and import into Power BI for creating the Dashboard.
5. Finally, we create an interactive Dashboard using Power BI.

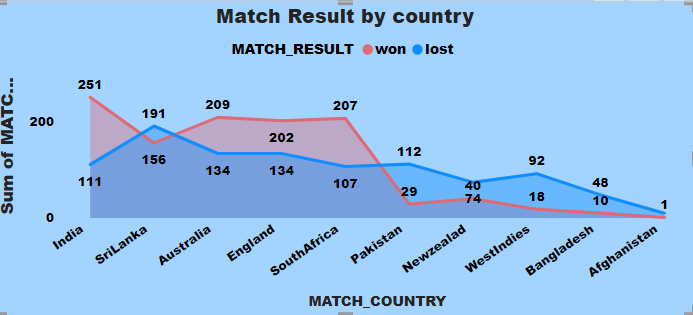
*Chapter IV*

**(4.1) Data Visualisation**

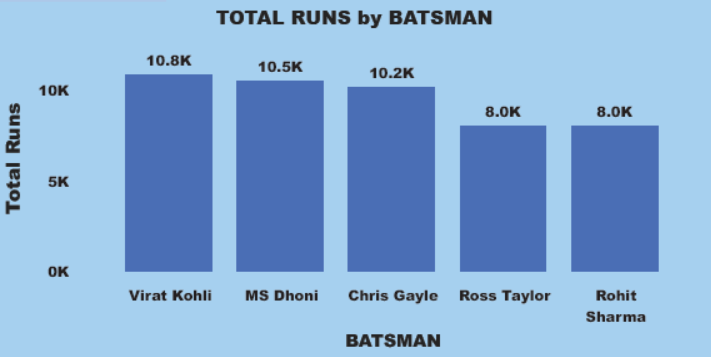
* + 1. **Maximum Runs Scored by Batsman**



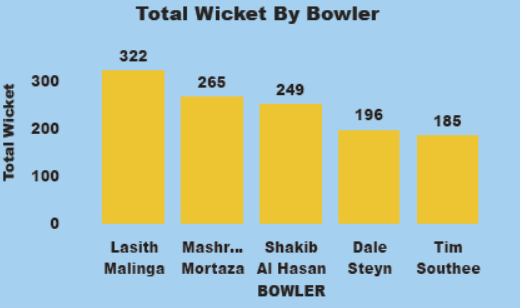
(4.1.2.) **Match Result by Country**



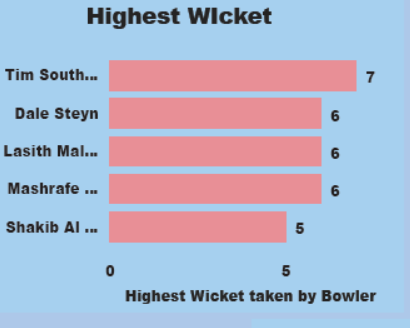
(4.1.3.) **Total Runs by Batsman**



(4.1.4.) Total Wicket by Bowler



(4.1.5.) Highest Wicket by Bowler



*Chapter V*

(5) **Conclusion**

The conclusion after analysis the data, insights are given as follows:

(1) Lasith Malinga is the most wicket taker with 322 wickets.

(2) Tim Southee is the highest wicket taker in one match with 7 wickets.

(3) Rohit Sharma is the highest run scorer in one match with 264 runs.

(4) Virat Kohli is the highest total runs scorer by batsman with 10.8K runs.

(5) India has won and lost most matches with 251 win and 111 lost.