# **BATSMEN DATA SET DESCRIPTION**

**Player :** The name of the test batsman.

**Span** : The start year and the end year of the batsman’s international test cricketing career, separated by a hyphen (-).

**Matches** : The number of test matches played by the batsman till 1st February 2019.

**Innings** : The number of test innings in which the batsman walked out to bat as of 1st February,2019.

**NO** : The number of test innings in which the batsman stayed not out till the end (could be in all 3 instances—in case of team declaration, in the case of the team getting all out or in the case of the team winning the match)

**Runs** : The total number of runs scored by the batsman during the course of his international test career.

**HS** : The highest score made by the batsman in international tests.

**Ave** : The batting average is defined as

(career runs scored/total no. of innings in which the batsman got out)

**X100 :** The number of centuries scored by the batsman in his career.

**X50 :** The number of half-centuries scored by the batsman in his career.

**X0 :** The number of innings in which the batsman got out on 0 runs.

**Helmet :** A value between 0 and 1 that indicates the proportion of the batsman’s career he may have worn a helmet. Almost all batsmen began to wear helmets by the year 1980. Hence, the batsmen who made their debut after 1980 naturally get a weight of 1, whereas those whose career ended before 1980 are all given a zero. The other batsmen whose career spanned on either side of the year 1980 are assigned a value that indicates what fraction of the matches they played were played after the year 1980.

**DRS :** A value between 0 and 1 that indicates the proportion of the batsman’s career in which the DRS (Decision Review System) may have been used. Though the DRS was first tested in a test match in 2008, it was only in 2010 that most test matches began to be officiated with the help of DRS. Hence, it is a given the cricketers who retired before 2009-10 are straightaway given a DRS weight of 0.

**HomeRuns :** The number of runs scored by the batsman in home venues. For the case of Indian test batsmen, it is the total number of runs scored by them in Indian grounds.

**AwayRuns :** The number of runs scored by Indian batsmen in venues outside India.

**HomePercent:** The home runs of the batsman, as a percentage of the total runs scored by him.

**(or)**

**AwayPercent:** The away runs of the batsman, as a percentage of the total runs scored by him.

**(or)**

**WinRuns :** The sum of all the runs made by the batsman in the matches which India won.

**LossRuns :** The sum of all the runs made by the batsman in the matches which India lost.

**DrawnRuns :** The sum of all the runs made by the batsman in the matches in which India drew the match with the opponent team.

**WinPercent :** The total runs hit by the batsman in winning causes, as a percentage of the total runs scored by the batsman in his career.

For example, Sachin Tendulkar hit 5946 of his total 15921 runs in the matches that India won. Hence, his *WinPercent* can be calculated as . Thus, it can be inferred that Sachin Tendulkar hit 37.35% of his runs in winning causes.

**LossPercent:** The total runs hit by the batsman in losing causes, as a percentage of the total runs scored by the batsman in his career.

For example, Sachin Tendulkar hit 4088 of his total 15921 runs in the matches that India won. Hence, his *WinPercent* can be calculated as . Thus, it can be inferred that Sachin Tendulkar hit 25.68% of his runs in losing causes.

**DrawnPercent:** The total runs hit by the batsman in the matches which got drawn, as a percentage of the total runs scored by the batsman in his career.

**vsAus, vsEng, vsSa, vsWi, vsNz, vsPak, vsSl:** The runs scored by a batsman scored against a particular opposition, as a percentage of the total runs scored in his career.

For example, since Sachin Tendulkar amassed a total of 3630 runs against Australia, his vsAus score becomes:

=

Similarly, we can calculate the percentages for all of the 50 batsmen in our dataset against each of the 7 opposition teams.