**Cricket** is a [bat-and-ball](https://en.wikipedia.org/wiki/Bat-and-ball_games) game played between two teams of 11 players each on a [field](https://en.wikipedia.org/wiki/Cricket_field) at the centre of which is a rectangular 22-yard-long [pitch](https://en.wikipedia.org/wiki/Cricket_pitch). The game is played by 120 million players in many countries, making it the world's second most popular sport after [association football](https://en.wikipedia.org/wiki/Association_football). Each team takes its turn to [bat](https://en.wikipedia.org/wiki/Batsman), attempting to score [runs](https://en.wikipedia.org/wiki/Run_(cricket)), while the other team [fields](https://en.wikipedia.org/wiki/Fielding_(cricket)). Each turn is known as an [innings](https://en.wikipedia.org/wiki/Innings)

The [bowler](https://en.wikipedia.org/wiki/Bowling_(cricket)) delivers the [ball](https://en.wikipedia.org/wiki/Cricket_ball) to the [batsman](https://en.wikipedia.org/wiki/Batsman) who attempts to hit the ball with his [bat](https://en.wikipedia.org/wiki/Cricket_bat) away from the [fielders](https://en.wikipedia.org/wiki/Fielding_(cricket)) so he can run to the other end of the [pitch](https://en.wikipedia.org/wiki/Cricket_pitch) and score a run. Each batsman continues batting until he is [out](https://en.wikipedia.org/wiki/Out_(cricket)). The batting team continues batting until ten batsmen are out, or a specified number of [overs](https://en.wikipedia.org/wiki/Over_(cricket)) of six balls have been bowled, at which point the teams switch roles and the fielding team comes in to bat.

Interesting because :

Since this is 10 year data set of every matches, we can make profile of any player and can predict the chances of his/her good performance in next match.

* Ball-by-ball information for three different format of the game.
  + One Day International
  + T20
  + Test Cricket( Multi day game)
* Number of Matches for each format:
  + 1,233 One-day internationals
  + 804 T20 internationals
  + 345 Test matches
* Featuring 52 countries, 13 IPL(Indian Premier League) teams
* when does a particular batsman score good/bad.
  + During which over slot?
  + Against which type of bowler?
    - Right hand or left hand bowler
    - Spinner, faster or medium pacer
  + Against which country?
  + Where? Away or home country.
  + When? In first inning or second inning
* When is a bowler more effective in terms of wickets or in terms of economy.
* What are the chances of getting out against a particular bowler?
* Who is going to perform well in this IPL, based on last year IPL and international performance.

### INFO **required**

**BOWL\_OUT**

The balls of any bowl-out. This will contain 1 entry (an associative array) for each ball bowled in the bowl-out, with the entries **bowler**, and **outcome**.

**BOWLER**

This is the bowler who bowled the ball in the bowl-out.

**OUTCOME**

This is the outcome of the bowl-out ball, either **hit**, or **miss**.

**CITY**

The city in which the game took place.

**DATES**

The dates on which the game took place. If there is just one day, for example in a T20 match, then it will be an array containing just that one date.

**GENDER**

The gender of the players who took part on the match. At the moment this value will be either **female**, or **male**, however these**should not** be regarded as the only possible options.

**MATCH\_TYPE**

The type of match this data file refers to. Currently the possible values are **Test**, **ODI**, **T20**, **IT20** (International T20), **ODM** (One-day match) or **MDM** (multi-day match).

**NEUTRAL\_VENUE**

If this is in the file then the value will be **1**. This indicates that the game was played on venue neither team would regard as a home ground.

**OUTCOME**

The data for this field is an associative array which details the outcome of the match this data file refers to. It contains information such as which team won the match, whether the game was a draw, tie, or no result, and any margin of victory.

**BY**

This entry is an associate array which details how much the winning team won by. It is an associative array with the possible values of**innings**, **runs**, and **wickets**.

**INNINGS**

If the match was won was by an innings and something then this entry will appear with a value of 1, for the 1 innings.

**RUNS**

If the match was won by a number of runs, or and innings and a number of runs, then this will contain the runs.

**WICKETS**

If the match was won by a number of wickets, then this will contain the number of wickets.

**BOWL\_OUT**

This field will list the winner of any bowl-out which decides a **tie** in a T20 match.

**ELIMINATOR**

This field will list the winner of any elimination super-over which decides a **tie** in a T20 match.

**METHOD**

This field will detail any method used to determine the winner where a match has been curtailed for some reason. Currently the only value this will contain is **D/L** when a match uses the Duckworth Lewis method.

**RESULT**

The result of the match if the match was not won by one of the teams. Currently the possible values are **draw**, **no result** or **tie**.

**WINNER**

If a team won the match then the name of the team will be here.

**OVERS**

If the match\_type is one of ODI, T20 or IT20 then this will indicate how many overs there are in each innings. Likely to be **50** for **ODI**matches, or **20** for **T20** or **IT20** matches.

**PLAYER\_OF\_MATCH**

If this field appears then it will contain an array of any players who were adjudged to be the player of the match.

**TEAMS**

An array containing the teams who played in the match. There will always be two entries. If the match is not being played on a neutral venue the first team listed will be the home team.

**TOSS**

**WINNER**

This will list the team which won the toss, and will be one of the teams listed as playing this match.

**DECISION**

The decision made by the team winning the toss. This will be either**bat** or **field**.

**UMPIRES**

An array containing the umpires who adjudicated in the match. There will always be two entries.

**VENUE**

The venue in which the game took place.

### INNINGS

#### TEAM

The team which is batting must be specified in the data for the innings. This will be one of the two teams mentioned in the teams section for the match.

#### ABSENT\_HURT

If this field is provided it will be an array of players who did not take part in the innings due to being absent hurt.

#### DECLARED

If this is in the file then the value will be **yes**. This indicates that the innings was declared.

#### DELIVERIES

An array of associative arrays each representing a delivery within the innings, in the order in which they took place. Each associative array has a single key (such as **23.5**) which specifies the particular ball (in that case the 5th ball of the 24th over), then the value is a further associative array with the details of that delivery.

##### BATSMAN

The batsman who faced the delivery.

##### NON\_STRIKER

The player who was the non-striker for the delivery.

##### BOWLER

The bowler who bowled the delivery.

##### RUNS

The data for this field is a simple associative array which details the breakdown of the runs scored from the ball. It breaks the runs down to show which the batsman scored, which were extras, and the total scored from the ball. There is also the ability to indicate that a 4 or 6 was not an actual boundary should, for example if the batsmen ran a 4.

###### BATSMAN

The total number of runs scored by the batsman off the ball. If the batsman failed to score this will show 0.

###### EXTRAS

The total number of runs conceded via extras off the ball. If no extras were conceded this will show 0.

###### NON\_BOUNDARY

If this is listed against the delivery then the value will be **1**. This indicates that the 4 or 6 scored was not via an actual boundary.

###### TOTAL

The total number of runs scored off this delivery. If no runs were scored from the delivery then this will display **0**.

##### SUPERSUB

The data for this field is an associative array providing the details of any super-substitution which happened before this delivery took place. It shows who was substituted in and out, and which team it was for.

###### IN

The name of the player who came in as part of the substitution.

###### OUT

The name of the player who went out as part of the substitution.

###### TEAM

The name of the team the substitution was for.

##### WICKET

If a wicket occurs then this entry will be in the file and will provide details on the wicket, such as which player is out, what type of dismissal it was, and any fielders who were involved.

In the \*very unlikely\* case that a wicket falls and another player retires hurt on the same ball, this will be an array containing an entry for each wicket. At the moment we only have a single match that meets this criteria, the 1st Test between Pakistan and Australia in 2014.

###### KIND

The kind of dismissal that took place. This will be one of **bowled**,**caught**, **caught and bowled**, **lbw**, **stumped**, **run out**, **retired hurt**,**hit wicket**, **obstructing the field**, **hit the ball twice**, **handled the ball**, or **timed out**.

###### PLAYER\_OUT

The name of the player who is out.

###### FIELDERS

Any fielders who were involved in the dismissal. Generally this will be the player who took a catch, or who was involved in a run-out.

##### EXTRAS

If extras were conceded from this delivery then this will indicate how the extras came about. The value of the field will be an associative array with **byes**, **legbyes**, **noballs**, **penalty**, and **wides** as the possible keys, and the associated value for each will be the number of runs from each.