**3. Data Collection and Preprocessing**

**3.1 Batsman Data**

The batsman data consists of various metrics that provide insight into a player's performance during a match. These metrics include:

* **Bat1**: Indicates the primary batsman.
* **Runs**: The total number of runs scored by the batsman.
* **BF (Balls Faced)**: The number of balls the batsman faced.
* **SR (Strike Rate)**: The rate at which the batsman scores runs per 100 balls, calculated as (Runs/Balls Faced) \* 100.
* **4s**: The number of boundaries (fours) hit by the batsman.
* **6s**: The number of sixes hit by the batsman.
* **Opposition**: The team against which the batsman played.
* **Ground**: The venue where the match was played.
* **Start Date**: The date on which the match started.
* **Match\_ID**: A unique identifier for the match.
* **Batsman**: The name of the batsman.
* **Player\_ID**: A unique identifier for the player.
* batsman data for various matches.

A screenshot of a computer

Description automatically generated

**3.2 Bowler Data**

The bowler data includes metrics that assess the performance of bowlers in a match. These metrics include:

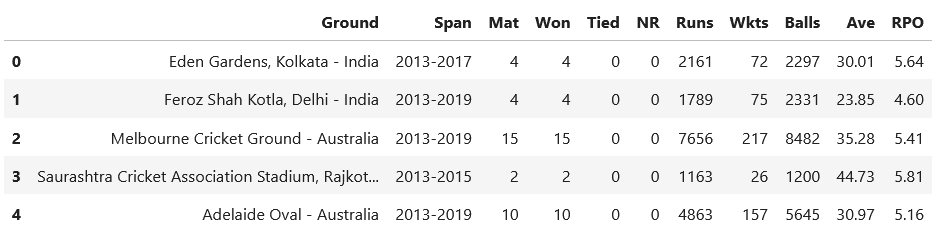
* **Overs**: The number of overs bowled by the bowler.
* **Mdns (Maidens)**: The number of maiden overs (overs in which no runs were conceded) bowled by the bowler.
* **Runs**: The total number of runs conceded by the bowler.
* **Wkts (Wickets)**: The number of wickets taken by the bowler.
* **Econ (Economy Rate)**: The rate at which the bowler concedes runs per over, calculated as (Runs/Overs).
* **Ave (Average)**: The average number of runs conceded per wicket taken, calculated as (Runs/Wickets).
* **SR (Strike Rate)**: The average number of balls bowled per wicket taken, calculated as (Balls Bowled/Wickets).
* **Opposition**: The team against which the bowler played.
* **Ground**: The venue where the match was played.
* **Start Date**: The date on which the match started.
* **Match\_ID**: A unique identifier for the match.
* **Bowler**: The name of the bowler.
* **Player\_ID**: A unique identifier for the player.
* bowler data for various matches.

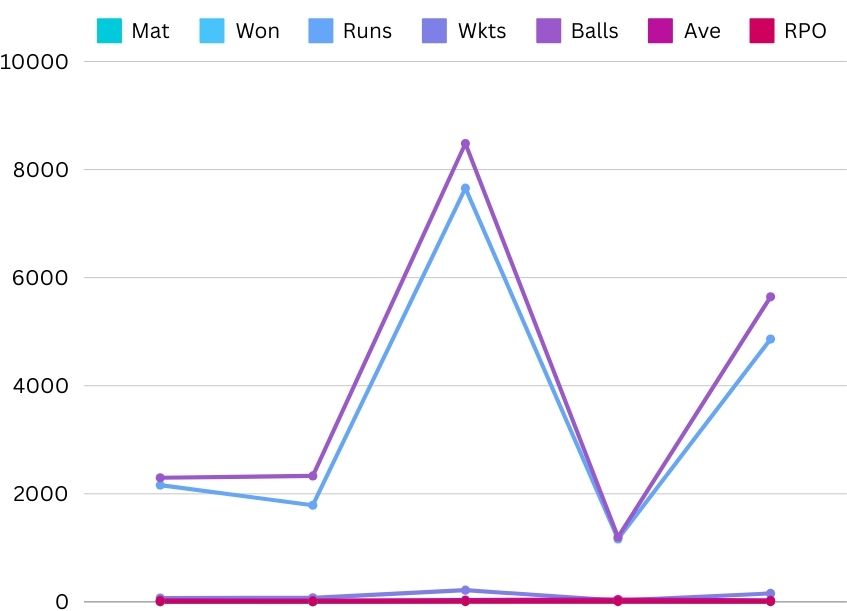


**3.3 Ground Averages**

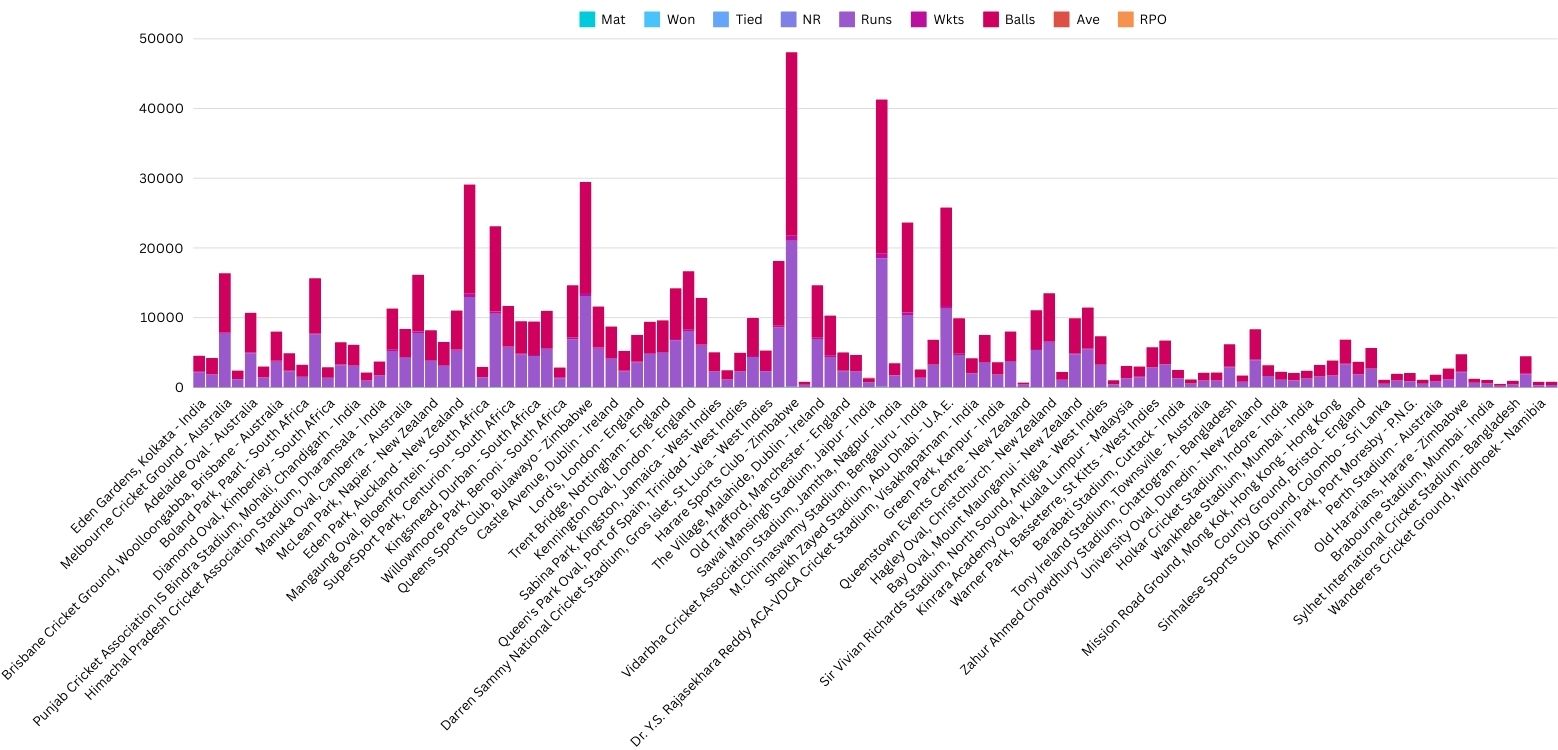
Ground averages provide information about the conditions and scoring patterns at different cricket venues. These metrics include:

* **Ground**: The name of the ground.
* **Span**: The time span over which the data has been collected.
* **Mat (Matches)**: The number of matches played at the ground.
* **Won**: The number of matches won.
* **Tied**: The number of matches that ended in a tie.
* **NR (No Result)**: The number of matches with no result.
* **Runs**: The total number of runs scored at the ground.
* **Wkts (Wickets)**: The total number of wickets taken at the ground.
* **Balls**: The total number of balls bowled at the ground.
* **Ave (Average)**: The average number of runs scored per wicket, calculated as (Runs/Wickets).
* **RPO (Runs Per Over)**: The average number of runs scored per over, calculated as (Runs/(Balls/6)).





* averages and conditions for various grounds.



**3.4 ODI Match Results**

The ODI match results data includes historical data on the outcomes of One Day International (ODI) matches. These metrics include:

* **Overs**: The total number of overs bowled in the match.
* **Mdns (Maidens)**: The number of maiden overs bowled.
* **Runs**: The total number of runs scored.
* **Wkts (Wickets)**: The total number of wickets taken.
* **Econ (Economy Rate)**: The rate at which runs were conceded.
* **Ave (Average)**: The average runs conceded per wicket.
* **SR (Strike Rate)**: The average number of balls bowled per wicket.
* **Opposition**: The opposing team.
* **Ground**: The venue of the match.
* **Start Date**: The date the match started.
* **Match\_ID**: The unique identifier for the match.
* **Bowler**: The name of the bowler.
* **Player\_ID**: The unique identifier for the player.
* the historical ODI match results

A screenshot of a computer

Description automatically generated

**3.5 ODI Match Totals**

The ODI match totals data includes information on team scores and individual contributions in One Day International matches. These metrics include:

* **Overs**: The total number of overs bowled.
* **Mdns (Maidens)**: The number of maiden overs bowled.
* **Runs**: The total number of runs scored.
* **Wkts (Wickets)**: The total number of wickets taken.
* **Econ (Economy Rate)**: The rate at which runs were conceded.
* **Ave (Average)**: The average runs conceded per wicket.
* **SR (Strike Rate)**: The average number of balls bowled per wicket.
* **Opposition**: The opposing team.
* **Ground**: The venue of the match.
* **Start Date**: The date the match started.
* **Match\_ID**: The unique identifier for the match.
* **Bowler**: The name of the bowler.
* **Player\_ID**: The unique identifier for the player.
* the match totals for various ODI matches.

**A screenshot of a computer

Description automatically generated**

**3.6 World Cup Players Data**

The World Cup players’ data includes performance metrics of players in World Cup tournaments. These metrics include:

* **Overs**: The total number of overs bowled.
* **Mdns (Maidens)**: The number of maiden overs bowled.
* **Runs**: The total number of runs scored.
* **Wkts (Wickets)**: The total number of wickets taken.
* **Econ (Economy Rate)**: The rate at which runs were conceded.
* **Ave (Average)**: The average runs conceded per wicket.
* **SR (Strike Rate)**: The average number of balls bowled per wicket.
* **Opposition**: The opposing team.
* **Ground**: The venue of the match.
* **Start Date**: The date the match started.
* **Match\_ID**: The unique identifier for the match.
* **Bowler**: The name of the bowler.
* **Player\_ID**: The unique identifier for the player.

**Importance**

World Cup players' data is crucial for making predictions about player performance in high-stakes tournaments and can help in formulating strategies.

* player performance in World Cup tournaments.

A graph with blue lines and black text

Description automatically generated