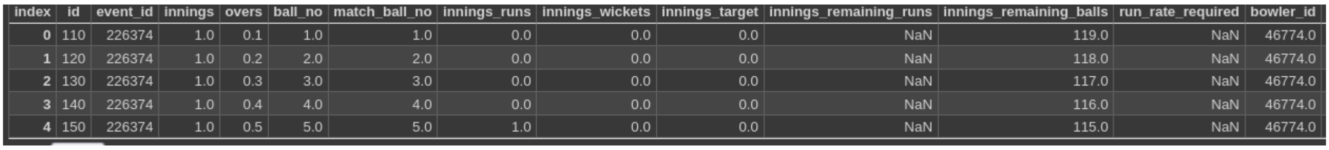
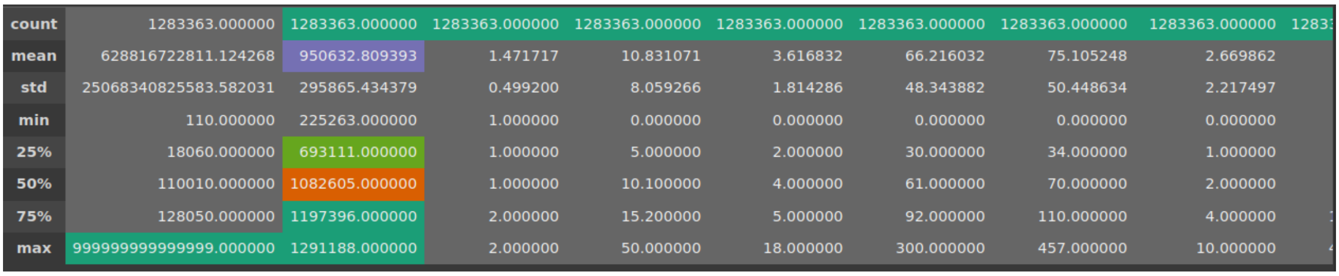
Data Exploration-

Top 5 columns in data



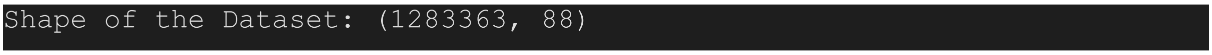
index	id	event_id	innings	overs	ball_no	match_ball_no	innings_runs	innings_wickets	innings_target	innings_remaining_runs	innings_remaining_balls	run_rate_required	bowler_id
0	110	226374	1.0	0.1	1.0	1.0	0.0	0.0	0.0	NaN	119.0	NaN	46774.0
1	120	226374	1.0	0.2	2.0	2.0	0.0	0.0	0.0	NaN	118.0	NaN	46774.0
2	130	226374	1.0	0.3	3.0	3.0	0.0	0.0	0.0	NaN	117.0	NaN	46774.0
3	140	226374	1.0	0.4	4.0	4.0	0.0	0.0	0.0	NaN	116.0	NaN	46774.0
4	150	226374	1.0	0.5	5.0	5.0	1.0	0.0	0.0	NaN	115.0	NaN	46774.0

Data Description



count	1283363.000000	1283363.000000	1283363.000000	1283363.000000	1283363.000000	1283363.000000	1283363.000000	1283363.000000	12833
mean	628816722811.124268	950632.809393	1.471717	10.831071	3.616832	66.216032	75.105248	2.669862	
std	25068340825583.582031	295865.434379	0.499200	8.059266	1.814286	48.343882	50.448634	2.217497	
min	110.000000	225263.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	18060.000000	693111.000000	1.000000	5.000000	2.000000	30.000000	34.000000	1.000000	
50%	110010.000000	1082605.000000	1.000000	10.100000	4.000000	61.000000	70.000000	2.000000	
75%	128050.000000	1197396.000000	2.000000	15.200000	5.000000	92.000000	110.000000	4.000000	1
max	999999999999999.000000	1291188.000000	2.000000	50.000000	18.000000	300.000000	457.000000	10.000000	4

Shape of data-



```
Shape of the Dataset: (1283363, 88)
```

list of columns of the dataset are Index(['id', 'event_id', 'innings', 'overs', 'ball_no', 'match_ball_no',innings_runs', 'innings_wickets', 'innings_target', 'innings_remaining_runs', 'innings_remaining_balls', 'run_rate_required', 'bowler_id', 'batter_id', 'batter_balls_faced', 'batter_runs', 'nonstriker_id', 'nonstriker_balls_faced', ‘'nonstriker_runs', 'outcome', 'wickets_lost', 'wicket_how', 'date', 'text', 'short_text', 'home_score', 'away_score', 'score_value', 'sequence', 'bbb_timestamp', 'play_type_id', 'bowler_name', 'bowler_short_name', 'bowler_full_name', 'bowler_team_id', 'bowler_team_name', 'bowler_maidens', 'bowler_balls', 'bowler_wickets', 'bowler_overs', 'bowler_conceded', 'batsman_striker_name', 'batsman_striker_short_name', 'batsman_striker_full_name', 'batsman_striker_team_id', 'batsman_striker_team_name', 'batsman_striker_fours', 'batsman_striker_sixes','batsman_striker_runs', 'batsman_nonstriker_name', 'batsman_nonstriker_short_name', 'batsman_nonstriker_full_name', 'batsman_nonstriker_team_id', 'batsman_nonstriker_team_name','batsman_nonstriker_fours', 'batsman_nonstriker_sixes', 'batsman_nonstriker_runs', 'innings_id', 'innings_run_rate', 'innings_byes', 'innings_number', 'innings_no_balls', 'innings_leg_byes', 'innings_ball_limit', 'innings_session', 'innings_day', 'innings_remaining_overs', 'innings_total_runs', 'innings_wides', 'over_balls' over_complete', 'over_limit', 'over_maiden', 'over_no_ball', 'over_byes', 'over_leg_byes', 'over_number', 'over_runs', 'over_wickets', 'over_actual', 'over_unique', 'dismissal_dismissal', 'dismissal_bowled', 'dismissal_minutes', 'dismissal_bowler_id', 'dismissal_bowler_name', 'dismissal_batsman_id', 'dismissal_batsman_name'], dtype='object')

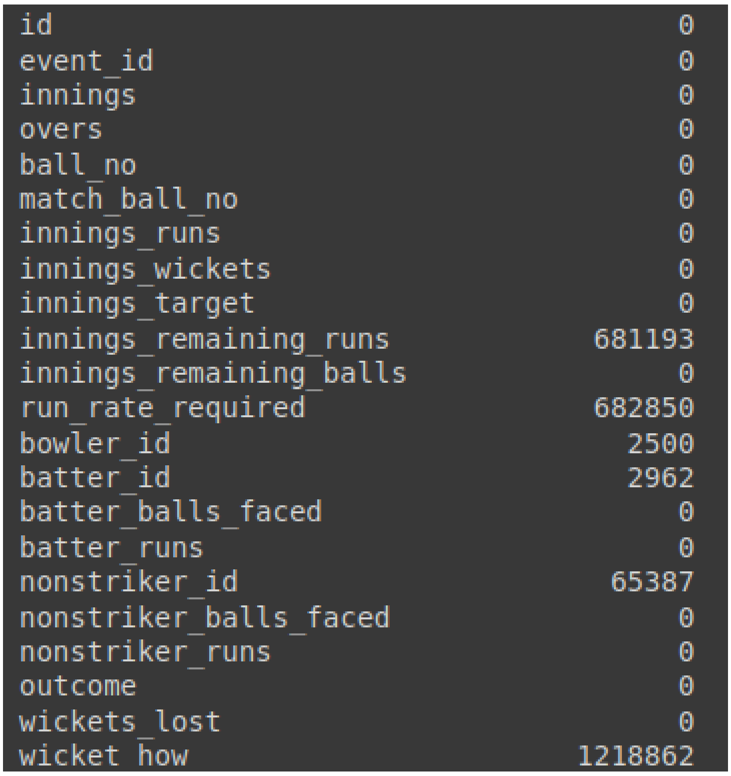
Data Type and more information of the dataset

```
RangeIndex: 1283363 entries, 0 to 1283362
Data columns (total 88 columns):
 #   Column                        Non-Null Count    Dtype
---  ------                        --------------    -----
 0   id                            1283363 non-null  int64
```

```
1   event_id                     1283363 non-null  int64
2   innings                      1283363 non-null  float64
3   overs                        1283363 non-null  float64
4   ball_no                      1283363 non-null  float64
5   match_ball_no                1283363 non-null  float64
6   innings_runs                 1283363 non-null  float64
7   innings_wickets              1283363 non-null  float64
8   innings_target               1283363 non-null  float64
9   innings_remaining_runs       602170 non-null   float64
10  innings_remaining_balls      1283363 non-null  float64
11  run_rate_required            600513 non-null   float64
12  bowler_id                    1280863 non-null  float64
13  batter_id                    1280401 non-null  float64
14  batter_balls_faced           1283363 non-null  float64
15  batter_runs                  1283363 non-null  float64
16  nonstriker_id                1217976 non-null  float64
17  nonstriker_balls_faced       1283363 non-null  float64
18  nonstriker_runs              1283363 non-null  float64
19  outcome                      1283363 non-null  object
20  wickets_lost                 1283363 non-null  float64
21  wicket_how                   64501 non-null    object
22  date                         1283363 non-null  object
23  text                         766952 non-null   object
24  short_text                   1282556 non-null  object
25  home_score                   1283363 non-null  object
26  away_score                   1283363 non-null  object
27  score_value                  1283363 non-null  float64
28  sequence                     1283363 non-null  float64
29  bbb_timestamp                1283363 non-null  float64
30  play_type_id                 1283363 non-null  int64
31  bowler_name                  1280863 non-null  object
32  bowler_short_name            1194630 non-null  object
33  bowler_full_name             1280863 non-null  object
34  bowler_team_id               1282556 non-null  float64
35  bowler_team_name             1282556 non-null  object
36  bowler_maidens               1283363 non-null  float64
37  bowler_balls                 1283363 non-null  float64
38  bowler_wickets               1283363 non-null  float64
39  bowler_overs                 1283363 non-null  float64
40  bowler_conceded              1283363 non-null  float64
41  batsman_striker_name         1280401 non-null  object
42  batsman_striker_short_name   1203924 non-null  object
43  batsman_striker_full_name    1280401 non-null  object
44  batsman_striker_team_id      1282556 non-null  float64
45  batsman_striker_team_name    1282556 non-null  object
46  batsman_striker_fours        1283363 non-null  float64
47  batsman_striker_sixes        1283363 non-null  float64
48  batsman_striker_runs         1283363 non-null  float64
```

```
 49  batsman_nonstriker_name        1217976 non-null  object
 50  batsman_nonstriker_short_name  1145807 non-null  object
 51  batsman_nonstriker_full_name   1217976 non-null  object
 52  batsman_nonstriker_team_id     1282556 non-null  float64
 53  batsman_nonstriker_team_name   1282556 non-null  object
 54  batsman_nonstriker_fours       1283363 non-null  float64
 55  batsman_nonstriker_sixes       1283363 non-null  float64
 56  batsman_nonstriker_runs        1283363 non-null  float64
 57  innings_id                     1283363 non-null  int64
 58  innings_run_rate               1283363 non-null  float64
 59  innings_byes                   1283363 non-null  float64
 60  innings_number                 1283363 non-null  float64
 61  innings_no_balls               1283363 non-null  float64
 62  innings_leg_byes               1283363 non-null  float64
 63  innings_ball_limit             1283363 non-null  float64
 64  innings_session                1283363 non-null  float64
 65  innings_day                    1283363 non-null  float64
 66  innings_remaining_overs        1283363 non-null  float64
 67  innings_total_runs             1283363 non-null  float64
 68  innings_wides                  1283363 non-null  float64
 69  over_balls                     1283363 non-null  float64
 70  over_complete                  1283363 non-null  bool
 71  over_limit                     1283363 non-null  float64
 72  over_maiden                    1283363 non-null  float64
 73  over_no_ball                   1283363 non-null  float64
 74  over_byes                      1283363 non-null  float64
 75  over_leg_byes                  1283363 non-null  float64
 76  over_number                    1283363 non-null  float64
 77  over_runs                      1283363 non-null  float64
 78  over_wickets                   1283363 non-null  float64
 79  over_actual                    1283363 non-null  float64
 80  over_unique                    1283363 non-null  float64
 81  dismissal_dismissal            1283363 non-null  bool
 82  dismissal_bowled               1283363 non-null  bool
 83  dismissal_minutes              1283363 non-null  float64
 84  dismissal_bowler_id            1280928 non-null  float64
 85  dismissal_bowler_name          1280928 non-null  object
 86  dismissal_batsman_id           1280409 non-null  float64
 87  dismissal_batsman_name         1280409 non-null  object
dtypes: bool(3), float64(60), int64(4), object(21)
```

Null Count-



```
id                              0
event_id                        0
innings                         0
overs                           0
ball_no                         0
match_ball_no                   0
innings_runs                    0
innings_wickets                 0
innings_target                  0
innings_remaining_runs     681193
innings_remaining_balls         0
run_rate_required          682850
bowler_id                    2500
batter_id                    2962
batter_balls_faced              0
batter_runs                     0
nonstriker_id               65387
nonstriker_balls_faced          0
nonstriker_runs                 0
outcome                         0
wickets_lost                    0
wicket_how                1218862
```

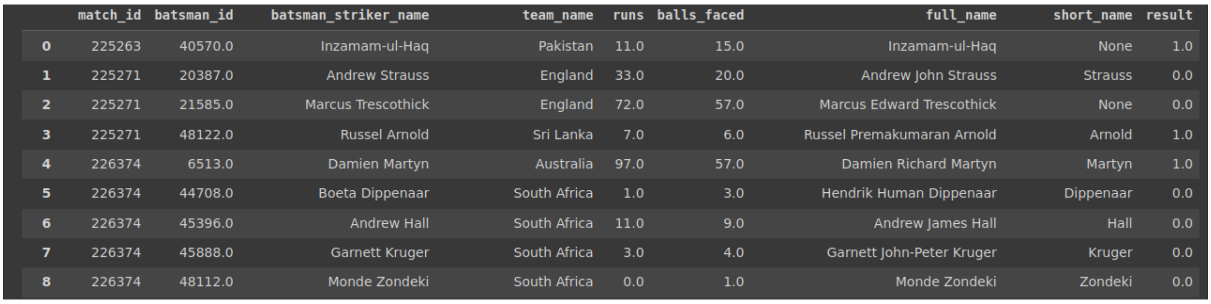
Total events(matches)- 5957

Handling Null values and extracting batsman statistics e.g. runs scored, match won loss, number of balls played etc.

Extracted batsman information

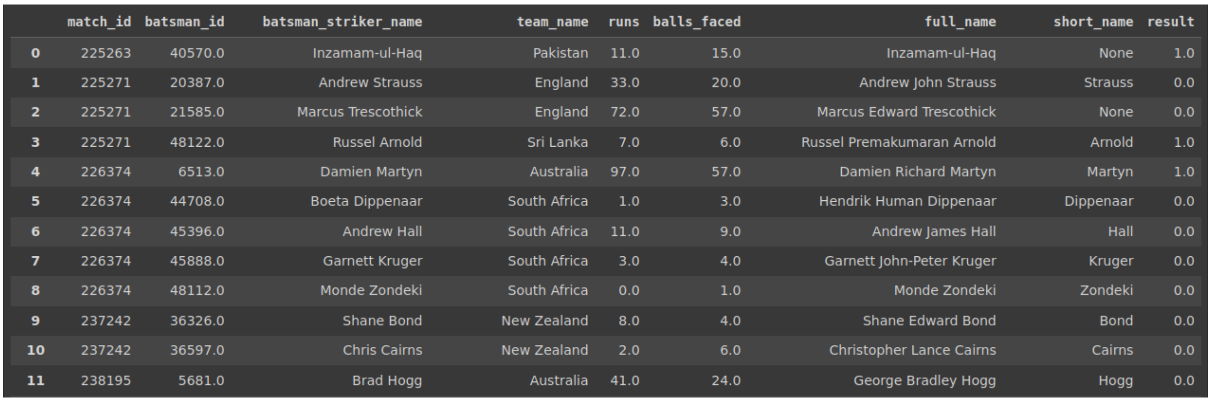
```
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   match_id             5036 non-null   int64
 1   batsman_id           5036 non-null   float64
 2   batsman_striker_name  5036 non-null   object
 3   team_name            5036 non-null   object
 4   team_id              5036 non-null   float64
 5   runs                 5036 non-null   float64
 6   balls_faced          5036 non-null   float64
 7   opposition           5036 non-null   float64
 8   full_name            5036 non-null   object
 9   result               5036 non-null   float64
```

Batsman information



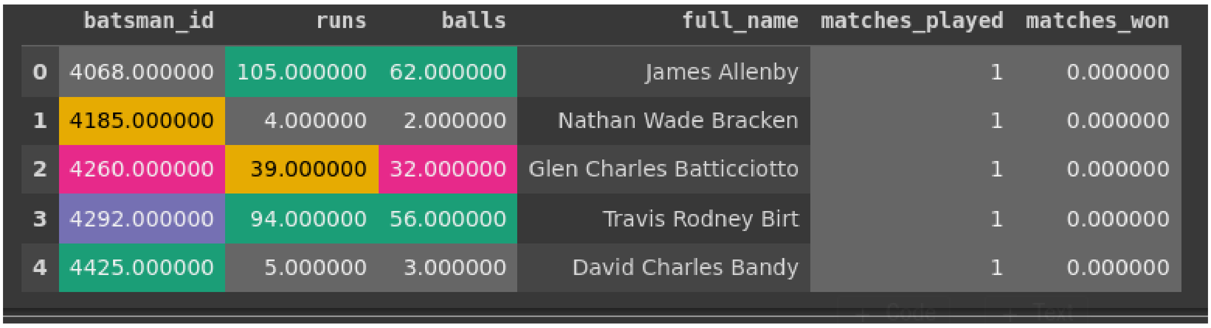
	match_id	batsman_id	batsman_striker_name	team_name	runs	balls_faced	full_name	short_name	result
0	225263	40570.0	Inzamam-ul-Haq	Pakistan	11.0	15.0	Inzamam-ul-Haq	None	1.0
1	225271	20387.0	Andrew Strauss	England	33.0	20.0	Andrew John Strauss	Strauss	0.0
2	225271	21585.0	Marcus Trescothick	England	72.0	57.0	Marcus Edward Trescothick	None	0.0
3	225271	48122.0	Russel Arnold	Sri Lanka	7.0	6.0	Russel Premakumaran Arnold	Arnold	1.0
4	226374	6513.0	Damien Martyn	Australia	97.0	57.0	Damien Richard Martyn	Martyn	1.0
5	226374	44708.0	Boeta Dippenaar	South Africa	1.0	3.0	Hendrik Human Dippenaar	Dippenaar	0.0
6	226374	45396.0	Andrew Hall	South Africa	11.0	9.0	Andrew James Hall	Hall	0.0
7	226374	45888.0	Garnett Kruger	South Africa	3.0	4.0	Garnett John-Peter Kruger	Kruger	0.0
8	226374	48112.0	Monde Zondeki	South Africa	0.0	1.0	Monde Zondeki	Zondeki	0.0

Bowler information-



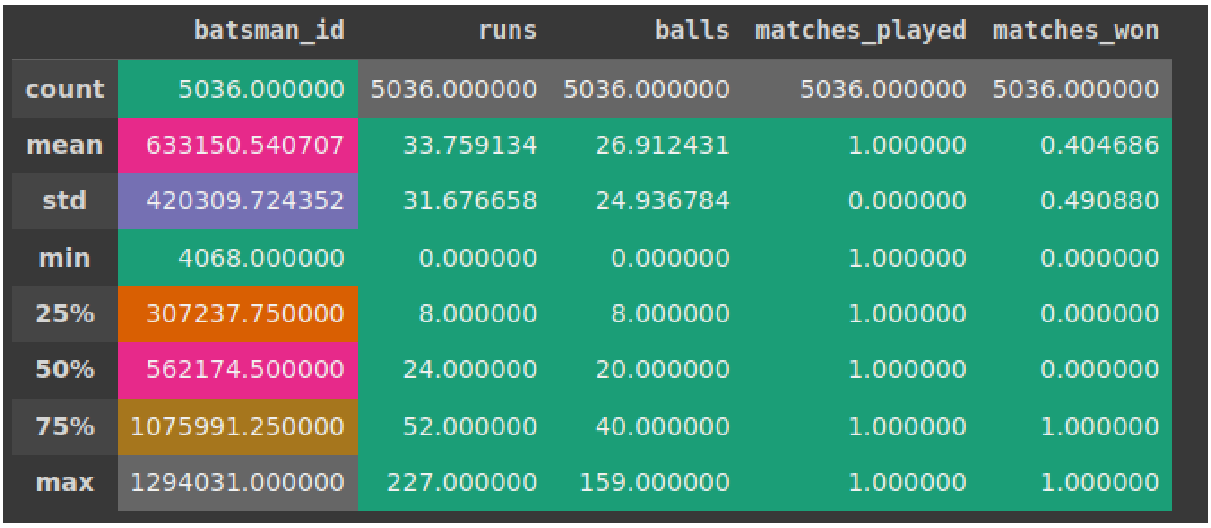
	match_id	batsman_id	batsman_striker_name	team_name	runs	balls_faced	full_name	short_name	result
0	225263	40570.0	Inzamam-ul-Haq	Pakistan	11.0	15.0	Inzamam-ul-Haq	None	1.0
1	225271	20387.0	Andrew Strauss	England	33.0	20.0	Andrew John Strauss	Strauss	0.0
2	225271	21585.0	Marcus Trescothick	England	72.0	57.0	Marcus Edward Trescothick	None	0.0
3	225271	48122.0	Russel Arnold	Sri Lanka	7.0	6.0	Russel Premakumaran Arnold	Arnold	1.0
4	226374	6513.0	Damien Martyn	Australia	97.0	57.0	Damien Richard Martyn	Martyn	1.0
5	226374	44708.0	Boeta Dippenaar	South Africa	1.0	3.0	Hendrik Human Dippenaar	Dippenaar	0.0
6	226374	45396.0	Andrew Hall	South Africa	11.0	9.0	Andrew James Hall	Hall	0.0
7	226374	45888.0	Garnett Kruger	South Africa	3.0	4.0	Garnett John-Peter Kruger	Kruger	0.0
8	226374	48112.0	Monde Zondeki	South Africa	0.0	1.0	Monde Zondeki	Zondeki	0.0
9	237242	36326.0	Shane Bond	New Zealand	8.0	4.0	Shane Edward Bond	Bond	0.0
10	237242	36597.0	Chris Cairns	New Zealand	2.0	6.0	Christopher Lance Cairns	Cairns	0.0
11	238195	5681.0	Brad Hogg	Australia	41.0	24.0	George Bradley Hogg	Hogg	0.0

Match win loss inforamtion



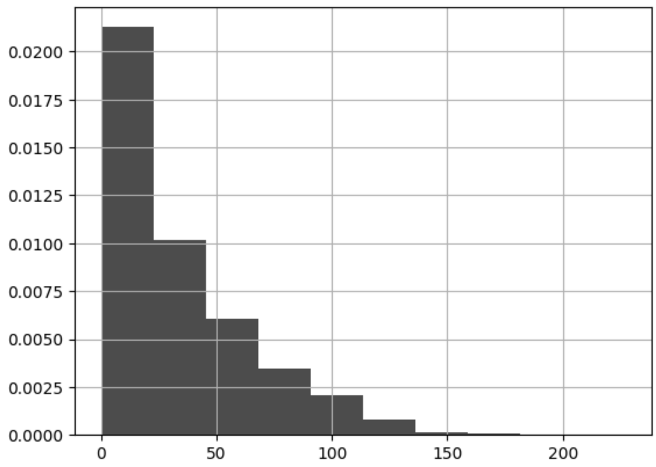
	batsman_id	runs	balls	full_name	matches_played	matches_won
0	4068.000000	105.000000	62.000000	James Allenby	1	0.000000
1	4185.000000	4.000000	2.000000	Nathan Wade Bracken	1	0.000000
2	4260.000000	39.000000	32.000000	Glen Charles Batticciotto	1	0.000000
3	4292.000000	94.000000	56.000000	Travis Rodney Birt	1	0.000000
4	4425.000000	5.000000	3.000000	David Charles Bandy	1	0.000000

Batsman data information-

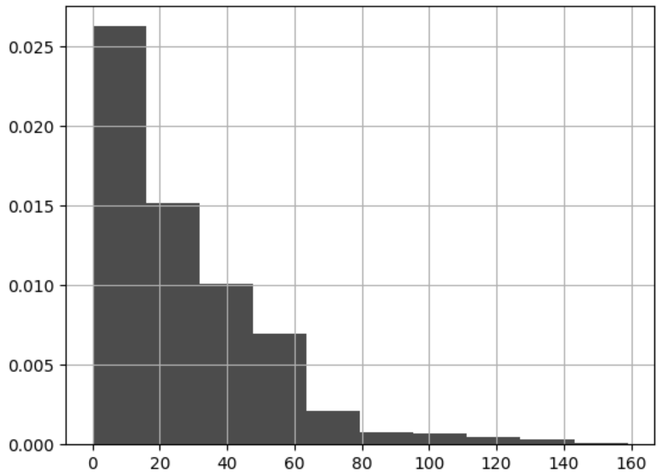


	batsman_id	runs	balls	matches_played	matches_won
count	5036.000000	5036.000000	5036.000000	5036.000000	5036.000000
mean	633150.540707	33.759134	26.912431	1.000000	0.404686
std	420309.724352	31.676658	24.936784	0.000000	0.490880
min	4068.000000	0.000000	0.000000	1.000000	0.000000
25%	307237.750000	8.000000	8.000000	1.000000	0.000000
50%	562174.500000	24.000000	20.000000	1.000000	0.000000
75%	1075991.250000	52.000000	40.000000	1.000000	1.000000
max	1294031.000000	227.000000	159.000000	1.000000	1.000000

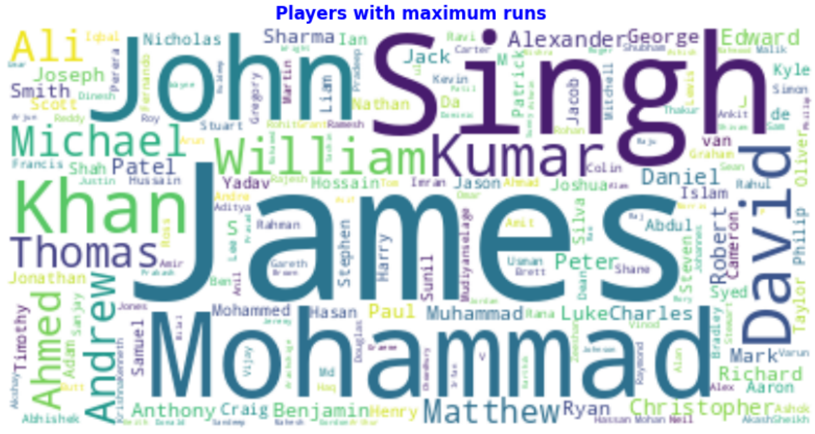
Runs Information-



Balls played information-



Word Cloud with names of players with maximum runs-



Steps followed to create rating system-

1. **Extraction of runs scored, number of matches, balls faced, strike-rate etc .**
2. **Creation of initial ranking matrix for teams and players.**
3. **Updation of ranking based on match outcome, runs scored, balls faced, etc.**

Future enhancements possible:

1. **Creating rating clusters of players and providing scores based on clusters of high performing, med performing players.**

2. **Using NLP to measure fielding capabilities and using that into ranking.**
3. **Batting position of player and their scores.**
4. **Number of men of the match titles of players.**
5. **Using autoencoders to learn the capabilities of players.**
6. **Rating of individual bowlers per over, in this pipeline used only team rankings.**
7. **Damping factor for ranking.**

Reference- https://dl.acm.org/doi/fullHtml/10.1145/3343172
https://betterprogramming.pub/odi-match-prediction-with-elo-scores-and-sklearn-b9dc60900ff5?gi=5a7db51cb42b
https://betterprogramming.pub/odi-match-prediction-with-elo-scores-and-sklearn-b9dc60900ff5
https://content.iospress.com/articles/journal-of-sports-analytics/jsa200411
https://dl.acm.org/doi/abs/10.1145/3102071.3102093