INSIDE OUT
A HAWK EYE'S VIEW OF CRICKET

RAJESH BELLAMKONDA

RB3805

SRINIVAS BONTULA

SSB597

PROBLEM DESCRIPTION

- Cricket statistics are based around individual measures, from a batsman's average to a bowler's strike rate.

- But what matters the most is the runs scored or wickets taken, when placed in a context.

- A series played on bowler-friendly pitches might deflate the career averages of the batsmen involved, but that dip fails to account for their performances relative to the conditions.

- ***Batting Averages and Bowling Strike rate, don't tell the full story.***

COURSE SCOPE

- The concepts learnt in the course such as function decorators, function call overheads, itertools, multiprocessing, other python inbuilt functions and packages were applied for the following steps –
  - Web Scraping – [BeatifulSoup, itertools, functools, geocoder, multiprocessing]
  - Modifying the datasets – [pandas, itertools]
  - Designing the algorithm – [numpy, itertools, math, pandas]
  - Predicting the teams – [threading, numpy, itertools]

DATA COLLECTION

- Web Scraping of the cricket websites -> Cricinfo and Howstat

- Requests -> For asynchronous calls to the website.

- Beautiful Soup -> For parsing the HTML webpage. (navigating, searching, and modifying)

- Multiprocessing -> Extracting the data series wise.

- Partial -> To solve the global variable issue.

DATASETS - BATSMAN

	Player	Country	% Team Runs	Mat	Inns	NO	50s	100s	0s	HS	Runs	S/R	Avg	Ca	St	Series_Code	H/A	Decade_Index
0	Bannerman C	Australia	28.63%	2	4	1	0	1	0	165*	209	-0.14	69.67	0	0	1	Home	0
1	Ulyett G	England	21.69%	2	4	0	2	0	0	63	149	-0.15	37.25	2	0	1	Away	0
2	Hill A	England	14.70%	2	4	2	0	0	1	49	101	-0.1	50.5	1	0	1	Away	0
3	Greenwood A	England	11.21%	2	4	0	0	0	0	49	77	-0.08	19.25	2	0	1	Away	0
4	Emmett T	England	10.63%	2	4	0	0	0	0	48	73	-0.07	18.25	4	0	1	Away	0
5	Jupp H	England	9.90%	2	4	0	1	0	1	63	68	-0.07	17	2	0	1	Away	0
6	Thompson N	Australia	9.18%	2	4	0	0	0	0	41	67	-0.04	16.75	3	0	1	Home	0
7	Midwinter W E	Australia	8.90%	2	4	0	0	0	0	31	65	-0.04	16.25	1	0	1	Home	0
8	Charlwood H R J	England	9.17%	2	4	0	0	0	1	36	63	-0.06	15.75	0	0	1	Away	0
9	Selby J	England	7.86%	2	4	0	0	0	0	38	54	-0.05	13.5	1	0	1	Away	0
10	Kelly T J D	Australia	7.40%	1	2	0	0	0	0	35	54	-	27	1	0	1	Home	0
11	Blackham J M	Australia	7.40%	2	4	0	0	0	0	26	54	-0.04	13.5	3	2	1	Home	0
12	Gregory D W	Australia	6.58%	2	4	1	0	0	0	43	48	-0.03	16	0	0	1	Home	0
13	Garrett T W	Australia	6.58%	2	4	1	0	0	1	18*	48	-0.03	16	0	0	1	Home	0
14	Kendall T	Australia	5.34%	2	4	1	0	0	0	17*	39	-0.03	13	2	0	1	Home	0
15	Armitage T	England	4.80%	2	3	0	0	0	0	21	33	-0.07	11	0	0	1	Away	0

DATASETS - BOWLERS

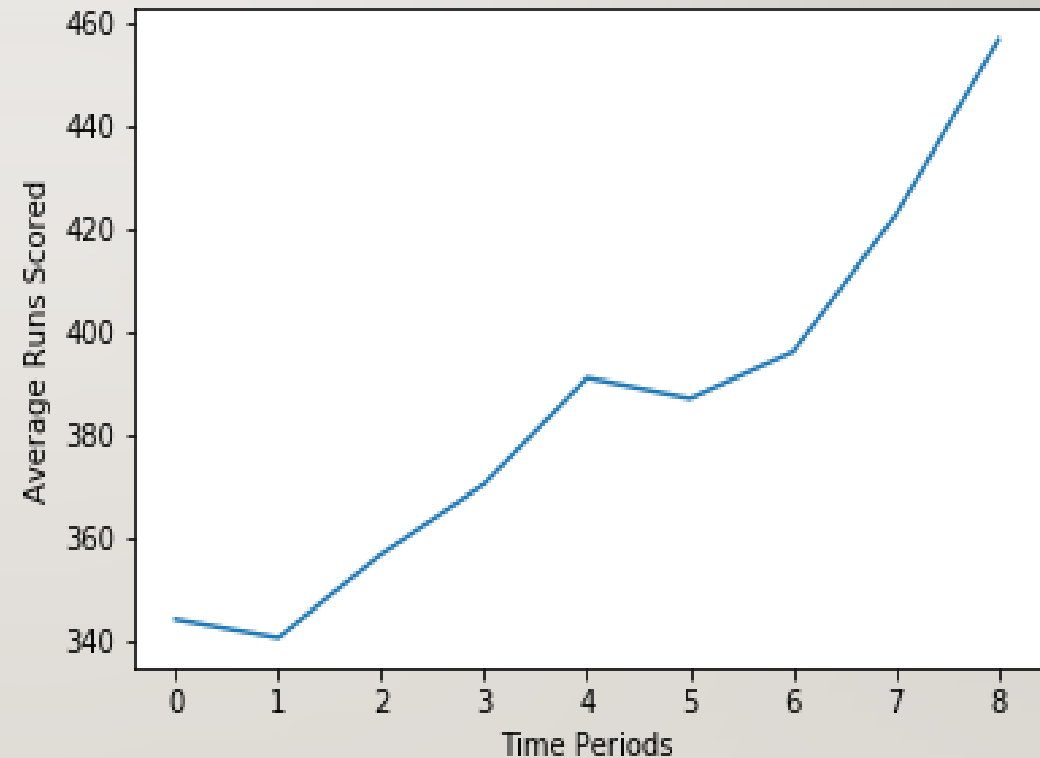| Player	Country	Mat	Overs	Maidens	Runs	Wickets	4w	Avg	S/R	E/R	Series_Co	H/A	Decade_Index	Matches
McKenzie G D	Australia	1	7.4	0	22	2	0	11	30	2.2	193	Home	0	1
Mallett A A	Australia	1	8	1	34	3	0	11.33	21.33	3.19	193	Home	0	1
Stackpole K R	Australia	1	8	0	40	3	0	13.33	21.33	3.75	193	Home	0	1
Illingworth R	England	1	8	1	50	3	0	16.67	21.33	4.69	193	Away	0	1
Thomson A L	Australia	1	8	2	22	1	0	22	64	2.06	193	Home	0	1
Shuttleworth K	England	1	7	0	29	1	0	29	56	3.11	193	Away	0	1
D'Oliveira B L	England	1	6	1	38	1	0	38	48	4.75	193	Away	0	1
Connolly A N	Australia	1	8	0	62	0	0			5.81	193	Home	0	1
Lever P	England	1	6	0	30	0	0			3.91	193	Away	0	1
Snow J A	England	1	8	0	38	0	0			3.56	193	Away	0	1
Wood B	England	1	6	0	20	2	0	10	18	3.33	194	Home	0	3
Watson G D	Australia	2	8	1	28	2	0	14	24	3.5	194	Away	0	3
Arnold G G	England	3	33	3	112	6	1	18.67	33	3.39	194	Home	0	3

OUR APPROACH

A series is a fair unit to use to break down careers. The numbers of a bowler or batsman over a series are self-contained indicators of performance that are also somewhat normalized for conditions.

Adding weights to the runs based on the time period of the series to curb the scoring trends, along with the host nation of the series.

Instead of looking at raw runs, wickets and averages over a series, we will try to look at the relative standing of a player in the run charts in a particular series, and average that over his whole career.

WHY DIFFERENT TIME PERIODS?

- 0 - 1971-1980
- 1 - 1981-1985
- 2 - 1986-1990
- 3 - 1991-1995
- 4 - 1995-1999
- 5 - 2001-2005
- 6 - 2006-2010
- 7 - 2011-2015
- 8 - 2016-2020

TEST MATCHES

-

ALGORITHM FOR BATSMAN

Time Period Index

Assigned period wise weights to the runs of each player in a series.

Opposition Index

Calculated the weights based on the different conditions (Home/Away), (Strength of the opposition)
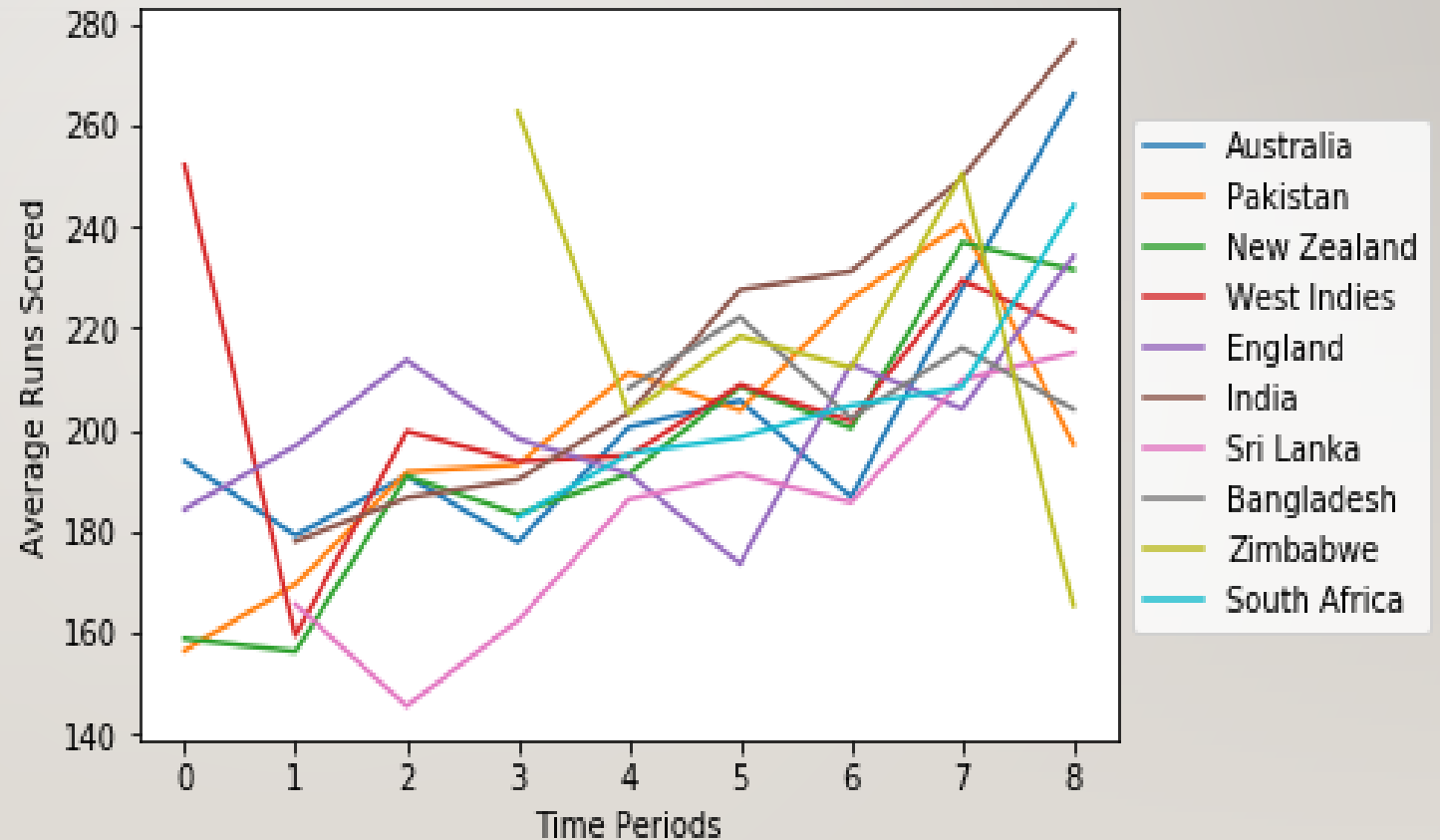
Series Percentile Value

100 * (1 – (position in series/total no. of players))

Result

The aggregate final set of the players

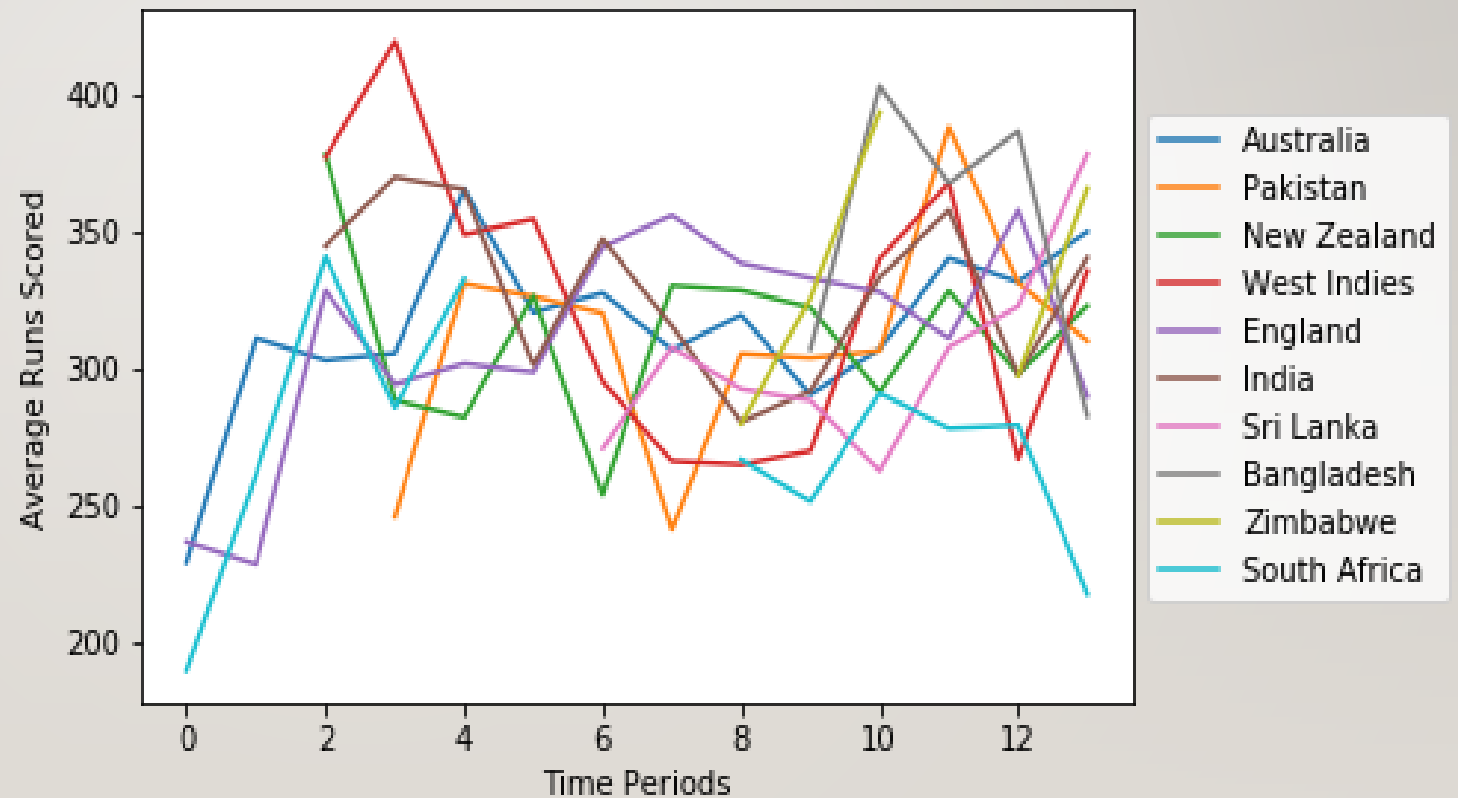# RUN SCORING TRENDS IN A VENUE – HOST NATION (ODIS)

- 0 - 1971-1980
- 1 - 1981-1985
- 2 - 1986-1990
- 3 - 1991-1995
- 4 - 1995-1999
- 5 - 2001-2005
- 6 - 2006-2010
- 7 - 2011-2015
- 8 - 2016-2020

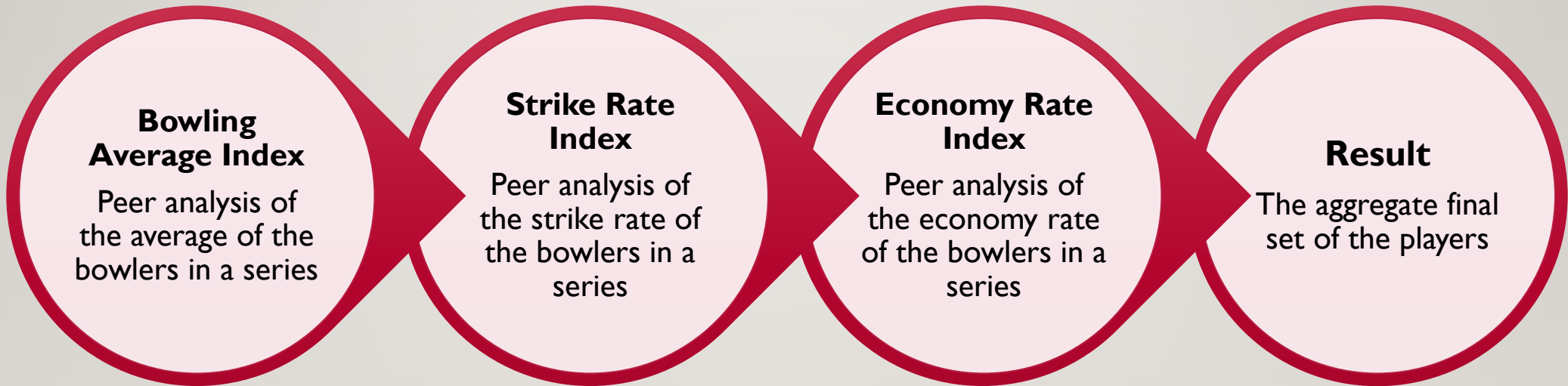# RUN SCORING TRENDS IN A VENUE – HOST NATION (TESTS)

- 1877-1900
- 1901-1920
- 1921-1940
- 1941-1960
- 1961-1970
- 1971-1980
- 1981-1985
- 1986-1990
- 1991-1995
- 1995-1999
- 2001-2005
- 2006-2010
- 2011-2015
- 2016-2020

ALGORITHM FOR BOWLERS

Bowling Average Index

Peer analysis of the average of the bowlers in a series

Strike Rate Index

Peer analysis of the strike rate of the bowlers in a series

Economy Rate Index

Peer analysis of the economy rate of the bowlers in a series

Result

The aggregate final set of the players

RESULTS

All Time Best Batsman	All Time Best Bowlers
Bradman D G : 79.61916301374308	Shoaib Akhtar : 1.65684692242
Flower A : 51.73676924016758	Grimmett C V : 1.41245679216
Mitchell B : 50.76018024699944	Steyn D W : 1.36515802769
Sutcliffe B : 48.859071657831734	Tate M W : 1.33832234419
Hanif Mohammad : 48.53654242389061	Muralitharan M : 1.29998609937
Lara B C : 47.67844151787917	Pollock S M : 1.27881407948
Chappell G S : 46.80487203048383	McGrath G D : 1.26907061865
Hutton L : 46.16650624644751	Garner J : 1.26656799002
Hobbs J B : 46.006532770098744	Jadeja R A : 1.26356055984
Taylor H W : 44.73616829246277	Ambrose C E L : 1.25075856257

RESULTS

- All time XIs (Test and One day Internationals)

TESTS

ODIS

All Time XI
Bradman D G : 79.61916301374308
Flower A : 51.73676924016758
Mitchell B : 50.76018024699944
Sutcliffe B : 48.859071657831734
Hanif Mohammad : 48.5365424238906
Lara B C : 47.67844151787917
Shoaib Akhtar : 1.698185946124763
Steyn D W : 1.3561379007901642
Grimmett C V : 1.3442212577189085
Muralitharan M : 1.3306703051309774
Tate M W : 1.309181121573127

All Time XI
Kohli V : 48.90574182372023
Zaheer Abbas : 48.69434245449234
Trott I J L : 48.56303132664924
Richards I V A : 44.33022883410812
de Villiers A B : 44.229334247184845
Amla H M : 42.06217803727083
Bond S E : 1.389097579056246
Garner J : 1.305133325499127
Holding M A : 1.2417925908587695
Starc M A : 1.2376191681540845
Donald A A : 1.2216394220226576

RESULTS

- Modern X1s (Test and One day Internationals)

All Time XI Modern(Debut after 1980)
Flower A : 51.73676924016758
Lara B C : 47.67844151787917
Crowe M D : 43.947140671847436
Jones A H : 43.89043937774854
Smith S P D : 43.0584519321065
Sangakkara K C : 40.58271800992688
Shoaib Akhtar : 1.698185946124763
Steyn D W : 1.3561379007901642
Muralitharan M : 1.3306703051309774
Pollock S M : 1.2805092945859433
Jadeja R A : 1.2362071684031024

Modern X1 (Debut after 1995)
Kohli V : 48.90574182372023
Trott I J L : 48.56303132664924
de Villiers A B : 44.229334247184845
Amla H M : 42.06217803727083
Root J E : 39.96594314175108
Dhoni M S : 39.45484691764698
Bond S E : 1.389097579056246
Starc M A : 1.2376191681540845
Boult T A : 1.1756741703575828
Pollock S M : 1.1740905002706719
Steyn D W : 1.1692471016571977

RESULTS

- All time XI and Modern XI for each country.

E.g ->

Australia	
All Time XI ▼	**Modern X1 (Debut after 1995)** ▼
Bevan M G : 41.00341884978064	Clarke M J : 34.53337865339904
Jones D M : 40.87492751201589	Hussey M E K : 33.887152091162854
Chappell G S : 40.404197796780174	Bailey G J : 32.37031148796234
Marsh G R : 35.805303984618554	Warner D A : 31.771659938326195
Clarke M J : 34.53337865339904	Watson S R : 30.827104792228322
Hussey M E K : 33.887152091162854	Smith S P D : 29.35580842843086
Starc M A:1.2376191681540845	Starc M A:1.2388290749405766
McGrath G D:1.1520650533675674	Lee B:1.146208211295317
Lee B:1.145342239059861	Gillespie J N:1.1274653077834638
McDermott C J:1.1270043750229475	McKay C J:1.1222032495722754
Gillespie J N:1.1269371505519625	Bracken N W:1.121290927554088

SQUAD PREDICTION

Next Series Team
Smith S P D : 43.0584519321065
Voges A C : 38.17748279192156
Warner D A : 32.73755400791322
Marsh S E : 22.414275208599378
Khawaja U T : 22.3869905226641
Burns J A : 17.362190265478922
Wade M S : 16.885955381971968
Marsh M R : 15.564622693302976
Cummins P J:1.1293283358300552
Hazlewood J R:1.125354353360165
Pattinson J L:1.1122777423994648
Starc M A:1.0394589618372672
Siddle P M:1.0387231272726152
O'Keefe S N J:1.0288540893280291
Lyon N M:0.9794361981863734

World Cup Team Predictions
Stoinis M P : 36.57638138271823
Bailey G J : 32.37031148796234
Warner D A : 31.771659938326195
Smith S P D : 29.35580842843086
Finch A J : 26.28198920232161
Marsh S E : 23.362225322291717
Maxwell G J : 22.763693263058116
Marsh M R : 22.654645746746795
White C L : 22.436919905619206
Head T M : 19.882174239457047
Faulkner J P : 18.693403481260635
Khawaja U T : 18.547924742134402
Starc M A:1.2431613785938471
Hazlewood J R:1.0917035497786187
Cummins P J:1.0521387582105983
Coulter-Nile N M:0.9824595519625191
Hastings J W:0.9659036649119732
Marsh M R:0.9472548328536828
Richardson K W:0.8751606514139835

REFERENCES

o http://www.espncricinfo.com/magazine/content/story/461783.html
o http://www.cricbuzz.com/cricket-news/99969/wanderers-pitch-india-vs-south-africa-the-rules-for-rating-a-pitch-dangerous-or-unfit
o http://www.espncricinfo.com/magazine/content/story/487445.html
o http://www.espncricinfo.com/story/_/id/22414318/batsmen-analysed-series-dominance
o http://www.espncricinfo.com/magazine/content/story/626913.html
o http://www.espncricinfo.com/magazine/content/story/626913.html
o http://www.howstat.com/cricket/home.asp

Q & A

Thank You