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.***

SCENARIO

- Let us consider two batsmen (A and B) and their performances in a sample set of 3 series.

Series in Australia

Batsman	Innings	Runs	Average
A	5	250	50.0
C	5	220	44.0
D	5	200	40.0
…	…	…	…
…	…	…	…
B	5	140	28.0

Series in South Africa

Batsman	Innings	Runs	Average
A	5	200	40.0
C	5	180	36.0
D	5	150	30.0
….	….	….	….
B	5	130	26.0
….	….	….	….

Series in India

Batsman	Innings	Runs	Average
B	5	600	120.0
C	5	520	104.0
D	5	450	90.0
A	5	400	80.0
….	….	….	….
….	….	….	….

DIFFERENCE

Batsman sorted by Average

Batsman	Average	Percentile
C	61.33	83.33
B	58.0	50.0
A	56.666	83.333
D	53.33	66.666
….	…..	……

Batsman sorted by Percentile

Batsman	Average	Percentile
A	56.666	83.333
C	61.33	83.33
D	53.33	66.666
B	58.0	50.0
….	…..	……

OUR APPROACH

- A series is a fair unit to use to break down careers. Played as a continuous bout over closely spaced matches, with similar teams and under similar conditions, the numbers of a bowler or batsman over a series are self-contained indicators of performance that are also somewhat normalized for conditions.
- 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.

OUR APPROACH (CONTINUED)

- A batsman and a bowler are allotted a percentile value based on his ranking: a top rank gets you 100 points, and the points decrease according to your rank.
- We then average this value over all series the batsman has played, weighed by the number of matches in each series.
- This "Series Percentile Value" sums up his run-scoring relative to all other batsmen who have played with or against him in a series.

- SPV (tentative) -> 100 * (1 – (position in series/total no. of players))

SAMPLE OUTPUT

Name	SPV
DG Bradman	97.23
JB Hobbs	91.04
ED Weekes	89.94
BC Lara	89.53
SPD Smith	89.39
L Hutton	89.21
H Sutcliffe	88.51
WR Hammond	88.50
RB Kanhai	87.98
GS Chappell	87.87
SM Gavaskar	86.63
KF Barrington	86.46
PBH May	86.42
RN Harvey	86.24
RB Richardson	86.19

WORKFLOW AND SCHEDULE

Step1

- Scraping the web for the data of the series. [03/12/2018 – 04/01/2018]
- Obtain SPVs for the players. [04/02/2018 – 04/06/2018]

Step2

- Add weights to the SPVs based on playing conditions. [04/07/2018 – 04/15/2018]
- Conditions -> Nature of the pitch, Home or Away, Year etc.

Step3

- Computing the weighted averages for the above. [04/16/2018 – 04/22/2018]
- Generating the outputs based on different metrics. [04/23/2018 – 04/29/2018]
- Final Project Presentation and Report. [04/30/2018 – 05/02/2018]

EXPECTED APPLICATIONS

- Ranking batsman and bowlers by our formula (Average SPV (using multiple metrics)).
- Predicting a team for a series from a rooster (based on the conditions and previous performances).
- All time X1s for each team.
- The batsmen who have never been in the lower half of the run-scoring table of a series.

and more……..

COURSE SCOPE

The concepts learnt in the course such as function decorators, function call overheads, itertools, cython and/or numba and other python inbuilt functions and packages will be applied for the following steps –

- WebScraping
- Modifying the datasets
- Iterating the datasets for the players
- Predicting the teams
- Designing the algorithm based on Series Percentile Index.