

Performance analysis of the competing 12 teams in the ICC T20 WC 2022

Group 15
ECE 143

What is Cricket?

“Cricket is just baseball on valium”
- Robin McLaurin Williams



We will draw a basic comparison with **baseball** to explain the rules of cricket for the purpose of our analysis;

All you need to know about the sport of cricket is the following:

- Three formats of the game; **T20, ODIs, and Tests**
- **2 innings** in T20 roughly equivalent to **1 baseball inning**
- One team 'bats' while the other 'bowls' in one innings and the order reverses in the second innings
- While 'batting', **10 "outs" (wickets) per inning** for each team
(3 outs per inning for each team in baseball)
- **Maximum 120 balls (or 20 overs)** bowled in each innings unless all wickets fall
- **Powerplay -> 1-6 overs; Middle overs -> 7-15 overs; Death overs -> 16-20 overs**
- Significantly more score (or 'runs') orientated (even in T20s, scores are ~150) (Very pitching and out orientated; >10 runs is considered high)
- Winner decided on the basis of score - whoever has the higher score after two innings, is declared the winner

We will not be going into the details of how is the score calculated or in what ways can the batsmen be declared 'out' since those are not essential for the purpose of understanding our analysis



Dataset and Pre-Processing

Data source

CRICSHEET

Freely-available structured data for cricket, including ball-by-ball data international and T20 League cricket matches, and identifier (register) mapping for people involved in cricket. Learn more about Cricsheet

<https://cricsheet.org/>

Raw data - format and size

We downloaded data for every T20 International Men's match (~1600 matches) in CSV format

Each match had two files

- Match_id.csv - contained ball-by-ball data for the respective match
- Match_id_info.csv - contained relevant metadata for that match (like, stadium, venue, squad, etc.)

Initial processing

We identified the relevant data items from the Match_id_info.csv files and pulled them into the main dataframe

version	2.1.0	
info	balls_per_over	6
info	team	Pakistan
info	team	Sri Lanka
info	gender	male
info	season	2007/08
info	date	9/17/07
info	event	ICC World Twenty20
info	match_number	16
info	venue	New Wanderers Stadium
info	city	Johannesburg
info	toss_winner	Sri Lanka
info	toss_decision	field
info	player_of_match	Younis Khan
info	umpire	DJ Harper
info	umpire	NJ Llong
info	reserve_umpire	KH Hurter
info	tv_umpire	MR Benson
info	match_referee	BC Broad
info	winner	Pakistan
info	winner_runs	33



version	2.1.0	
info	balls_per_over	6
info	team	Pakistan
info	team	Sri Lanka
info	gender	male
info	season	2007/08
info	date	9/17/07
info	event	ICC World Twenty20
info	match_number	16
info	venue	New Wanderers Stadium
info	city	Johannesburg
info	toss_winner	Sri Lanka
info	toss_decision	field
info	player_of_match	Younis Khan
info	umpire	DJ Harper
info	umpire	NJ Llong
info	reserve_umpire	KH Hurter
info	tv_umpire	MR Benson
info	match_referee	BC Broad
info	winner	Pakistan
info	winner_runs	33

Initial processing

We repeated the same operation with the Match_id.csv files by pulling the relevant data into the main dataframe

match_id	season	start_date	venue		innings_ball	battling_team	bowling_team	striker	non_striker	bowler	runs_off_ball	extras	wides	noballs	byes	legbyes	penalty	wicket_type	player_dismissal	other_wicket	other_player_dismissal
287868	2007/08	9/17/07	New Wanderers Stadium	1	0.1	Pakistan	Sri Lanka	Salman Butt	Imran Nazir	WPUJC Vaas	0	0									
287868	2007/08	9/17/07	New Wanderers Stadium	1	0.2	Pakistan	Sri Lanka	Salman Butt	Imran Nazir	WPUJC Vaas	4	0									
287868	2007/08	9/17/07	New Wanderers Stadium	1	0.3	Pakistan	Sri Lanka	Salman Butt	Imran Nazir	WPUJC Vaas	1	0									
287868	2007/08	9/17/07	New Wanderers Stadium	1	0.4	Pakistan	Sri Lanka	Imran Nazir	Salman Butt	WPUJC Vaas	1	0									

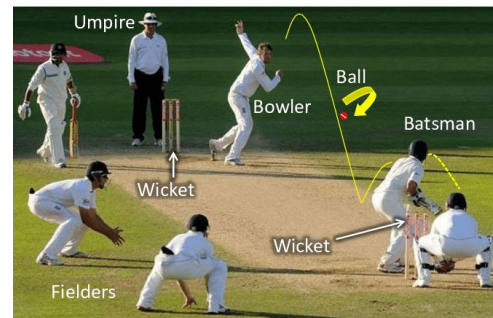


match_id	season	start_date	venue	inning: ball	battling_team	bowling_team	striker	non_striker	bowler	runs_off_ball	extras	wides	noballs	byes	legbyes	penalty	wicket_type	player_dismiss	other_wicket	other_player_dismissed
287868	2007/08	9/17/07	New Wanderers Stadium	1	0.1	Pakistan	Sri Lanka	Salman Butt	Imran Nazir	WPUJC Vaas	0	0								
287868	2007/08	9/17/07	New Wanderers Stadium	1	0.2	Pakistan	Sri Lanka	Salman Butt	Imran Nazir	WPUJC Vaas	4	0								
287868	2007/08	9/17/07	New Wanderers Stadium	1	0.3	Pakistan	Sri Lanka	Salman Butt	Imran Nazir	WPUJC Vaas	1	0								
287868	2007/08	9/17/07	New Wanderers Stadium	1	0.4	Pakistan	Sri Lanka	Imran Nazir	Salman Butt	WPUJC Vaas	1	0								
287868	2007/08	9/17/07	New Wanderers Stadium	1	0.5	Pakistan	Sri Lanka	Salman Butt	Imran Nazir	WPUJC Vaas	0	0								
287868	2007/08	9/17/07	New Wanderers Stadium	1	0.6	Pakistan	Sri Lanka	Salman Butt	Imran Nazir	WPUJC Vaas	1	0								

Secondary processing

Once the data was pulled into the dataframe, we used the “ball”, “inning”, ”Runs off bat”, “extra”, and “wicket type” fields to calculate the following data:

- Runs: Powerplay, Middle Overs, Death Overs,
- Wickets: Powerplay, Middle Overs, Death Overs
- Total_Score_A
- Total_Wicket_A



The columns used for this processing were dropped after the completion of the operation

All the new and leftover fields were stored in a new csv file that was read back during the analysis phase

We carried out manual processing using Google Search and Excel for our data, where we found that the stadium names were renamed, or had different suffixes, or had missing cities.

Processed data

The final format of the data table looked like this:

id	year	city	venue	event	team_A	team_B	winner	toss_winner	toss_decision	Runs_in_P	Wickets_lost_in_F	Runs_in_m	Wickets_lost_in_m	Runs_in_D	Wickets_lost_in_c	Total_Wicket_A	Total_Score_A
121	2020	Bangl	Terdth	ACC Eas	Thailand	Nepal	Nepal	Thailand	bat	13	2	38	4	15	3	9	66
121	2020	Bangl	Terdth	ACC Eas	Nepal	Thailand	Nepal	Thailand	bat	72	1	0	0	0	0	1	72
126	2021	Marsa	Marsa S	Belgium	Malta	Belgium	Belgium	Belgium	field	27	2	43	4	44	2	8	114
126	2021	Marsa	Marsa S	Belgium	Belgium	Malta	Belgium	Belgium	field	37	4	56	1	22	0	5	115
125	2021	Wind Wande	Uganda	Namibia	Uganda	Namibi	Namibia	Namibia	bat	58	0	79	2	52	1	3	189
125	2021	Wind Wande	Uganda	Uganda	Namibia	Namibi	Namibia	Namibia	bat	29	2	77	3	18	3	8	124
133	2022	Sano	Sano In	ICC Mei	Japan	South K	Japan	Japan	bat	63	1	110	3	45	4	8	218
133	2022	Sano	Sano In	ICC Mei	South Ko	Japan	Japan	Japan	bat	48	3	98	2	24	1	6	170
120	2015	Abu F Shaikh	Hong K	Open	Hong K	Hong K	Hong K	Hong K	field	46	2	53	1	50	0	4	140

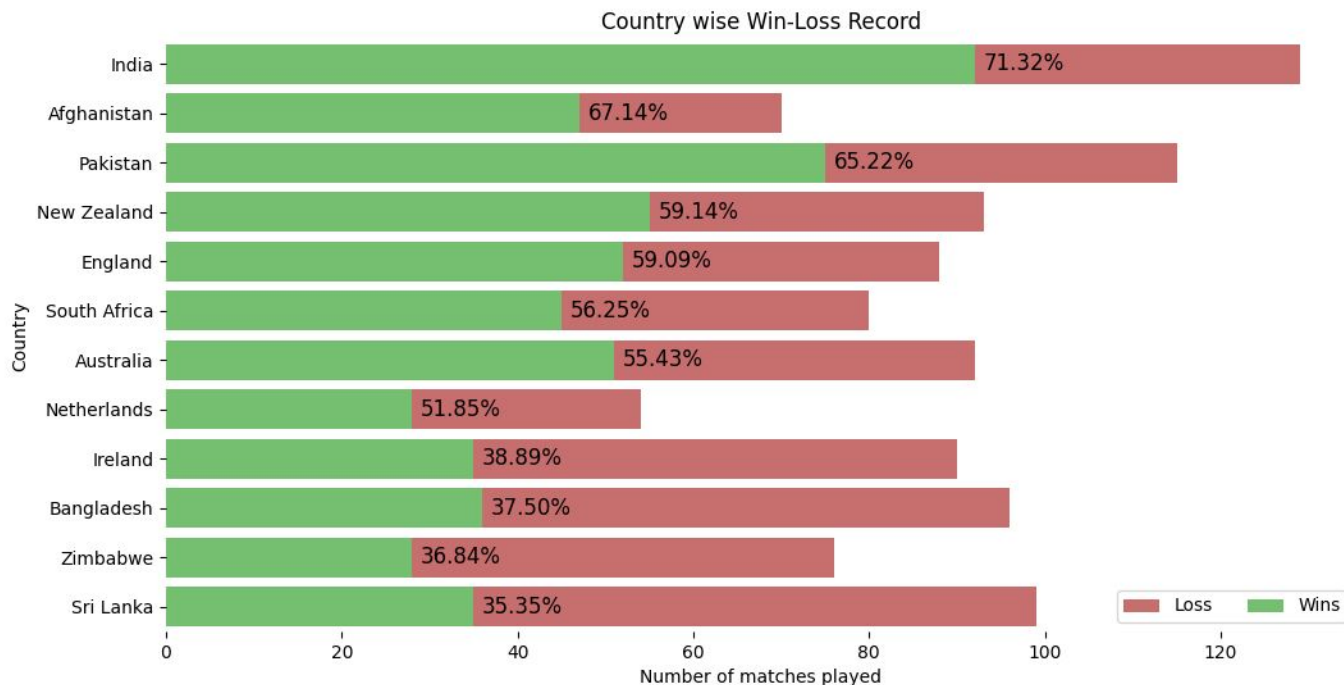
Analysis and Insights

Defining the metrics of evaluation

For any team, we look at the following metrics as well as what we can infer from them:

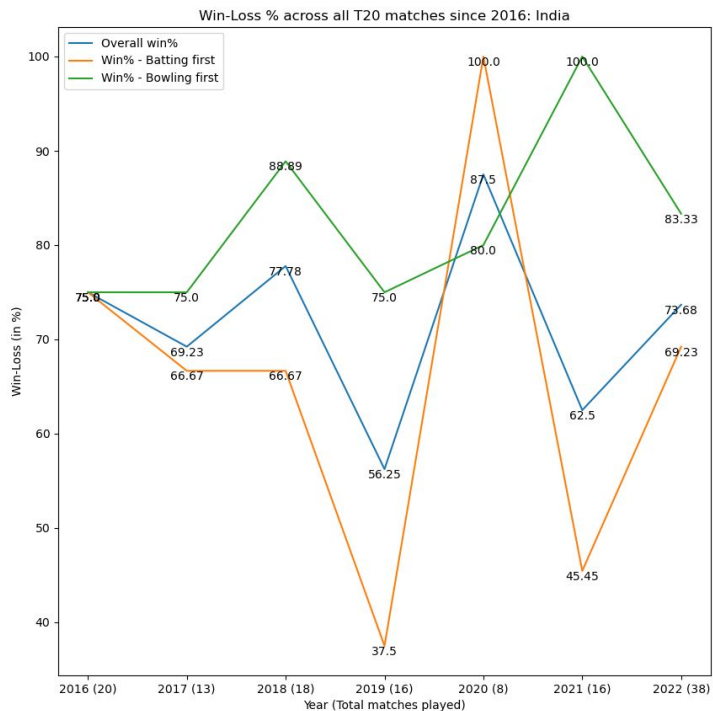
- **Win - Loss % (per year):** Percentage of matches won out of the total matches played in that time period; divided further by whether the team 'batted first' or 'bowled first' (or 'batted second')
- **Batting versus Bowling strength:** Qualitative measure of whether the team is good at batting or bowling or both aspects; what % of wins are contributed to either or both departments
- **Avg. runs scored and wickets conceded (per year):** Avg. score and avg. number of wickets conceded per match (against all oppositions) per year; divided further by whether the team 'batted first' or 'bowled first' ('batted second')
- **Phase wise team batting strength:** Qualitative measure of which phase of the innings is the team good at batting in; average run-rate and avg. wickets conceded in each of the three phases per match per year

Win-Loss % : 2016 till the T20 World Cup 2022

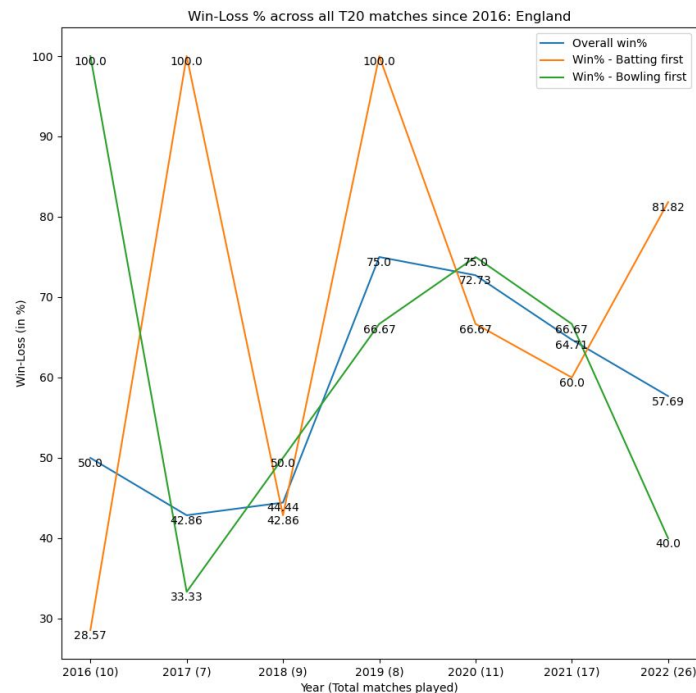


4 out of the top 5 teams (in terms of win-loss%) made it to the knockout stages of the T20 World Cup in 2022

Win-Loss % : 'Batting first' v/s 'Bowling first'

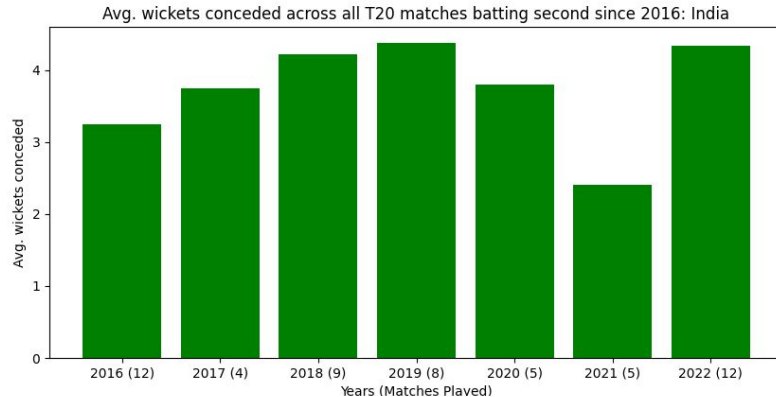
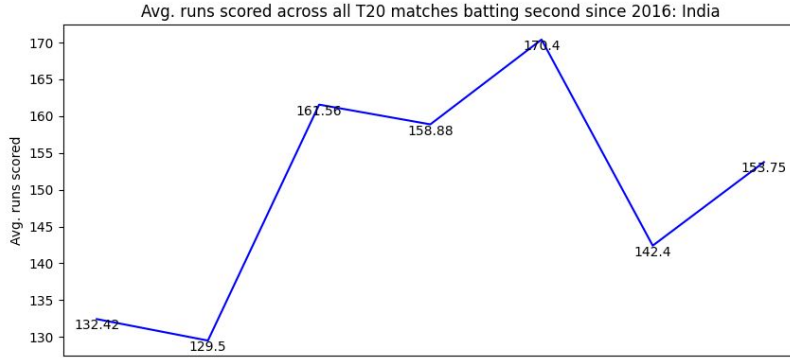


India has a higher win-loss % (82.46 %) when batting second(65.78%) highlighting **India's strength in batting second**



England, on the other hand, has a **higher win-loss % (68.56 %)** when **batting first**, justifying their aggressive approach

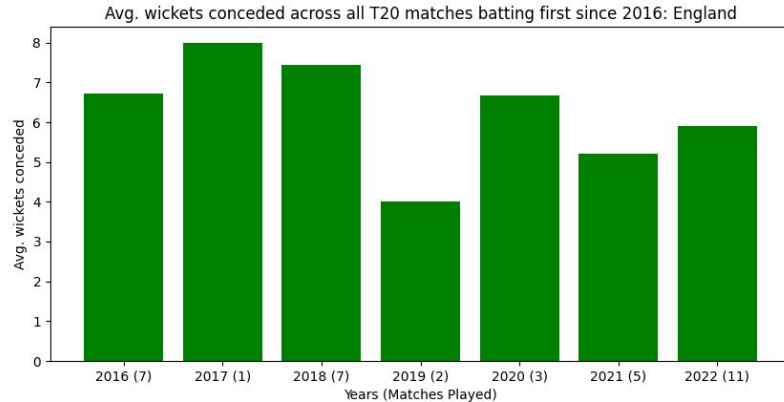
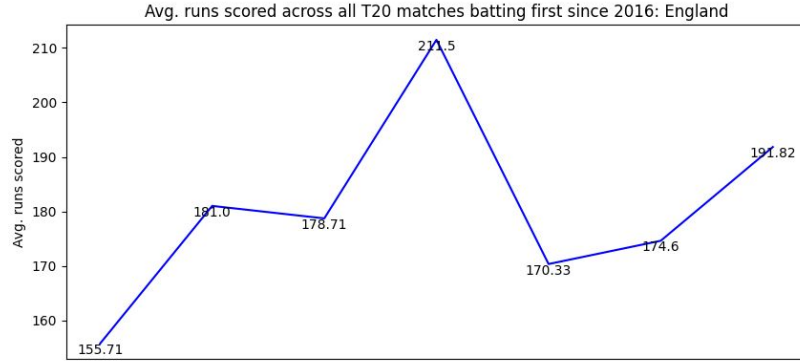
Understanding batting approaches through data : India



When India is batting second, we see that

- India has not conceded more than 4 wickets since 2016 while chasing,
- While scoring runs at a brisk run rate of ~ 7.5 per over,
- Loss of ~ 3.5 wickets on an average allows the top order lineup to chase the target down

Understanding batting approaches through data : England

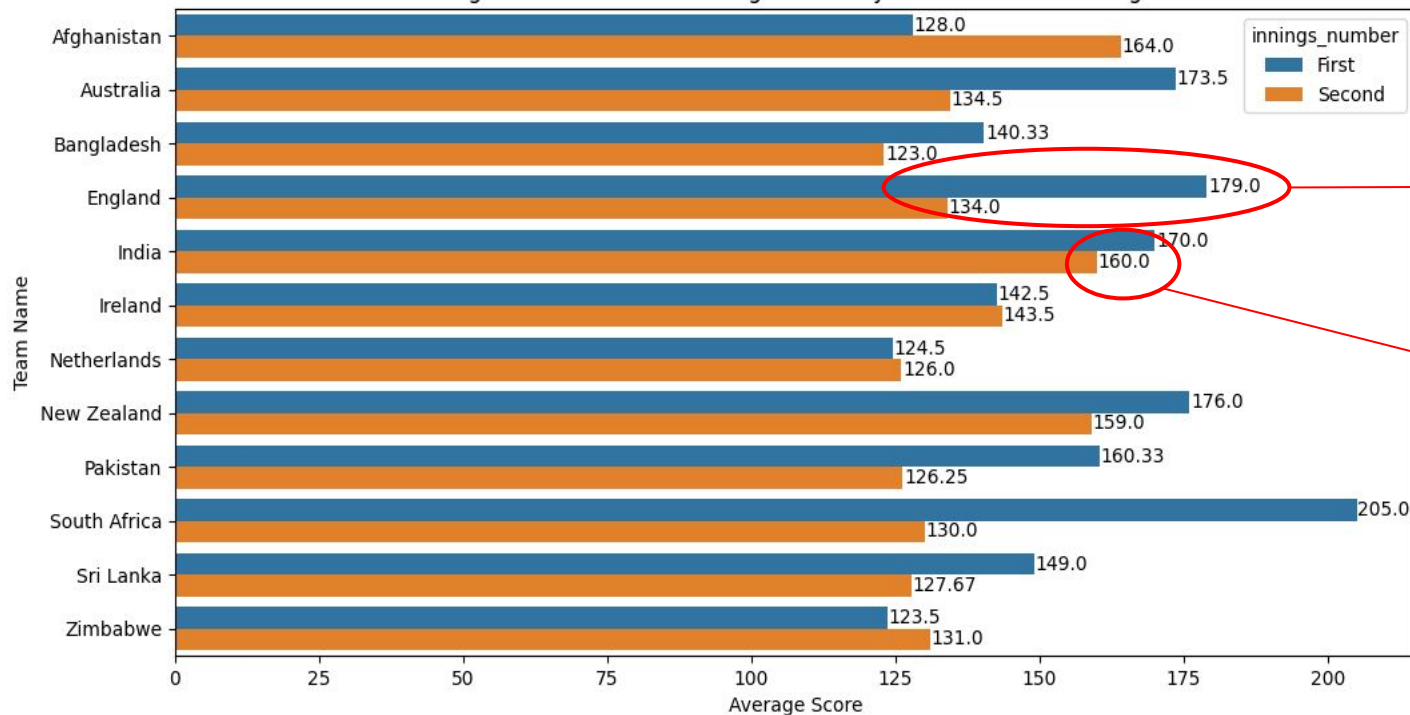


When England is batting first, we observed;

- England has conceded 6+ wickets on an average when batting first
- But, average runs scored is 178.86, which approximates to a huge run-rate of 8.94 per over,
- Indicating the depth as well as aggression in their batting approach

Batting performances in T20 World Cup 2022

Average First and Second Innings scores by different teams during WC

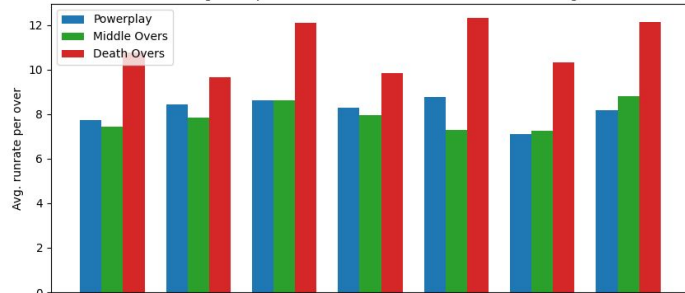


Second highest average
innings score when batting first
for England

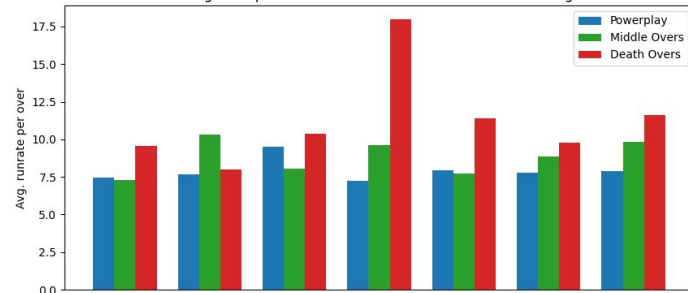
Second highest average
innings score when batting
second for India

Addressing the kinks in batting approach : India

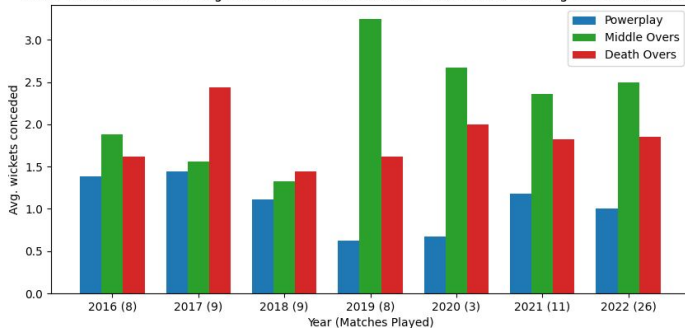
Phase wise breakdown of avg. runs per over scored across all T20 matches batting first since 2016: India



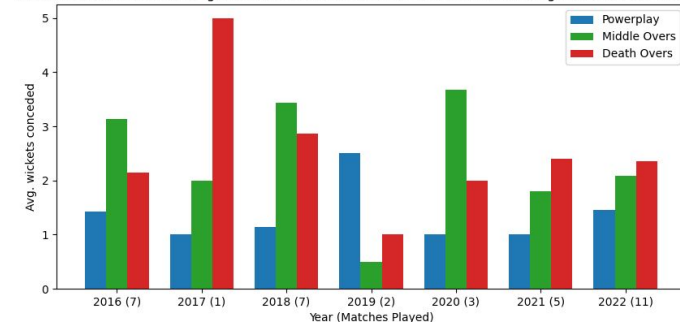
Phase wise breakdown of avg. runs per over scored across all T20 matches batting first since 2016: England



Phase wise breakdown of avg. wickets conceded across all T20 matches batting first since 2016: India



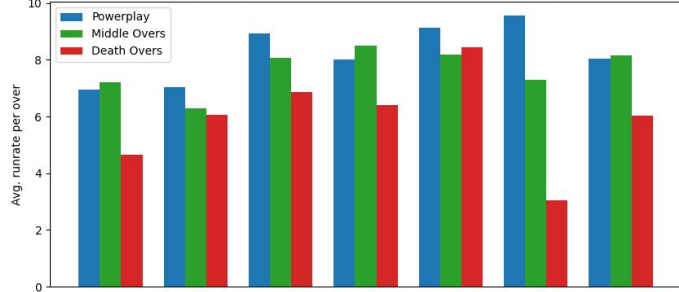
Phase wise breakdown of avg. wickets conceded across all T20 matches batting first since 2016: England



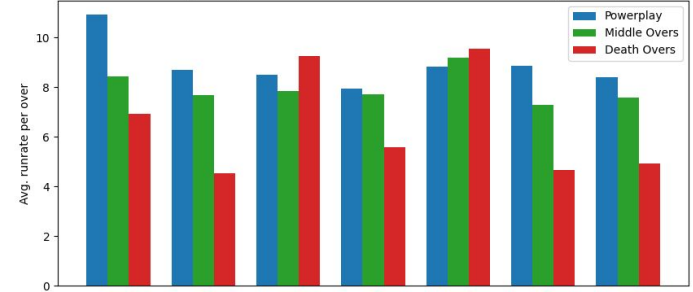
- Very conservative approach in the first two phases by India while batting first when compared to England
- Puts pressure on the batters during the death overs to make up for the deficit, ending in a below par total

Addressing the kinks in batting approach : England

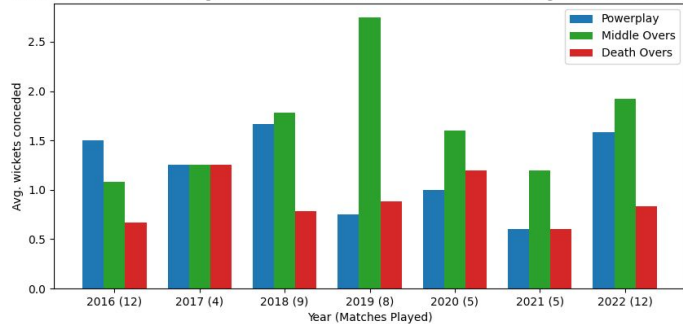
Phase wise breakdown of avg. runs per over scored across all T20 matches batting second since 2016: India



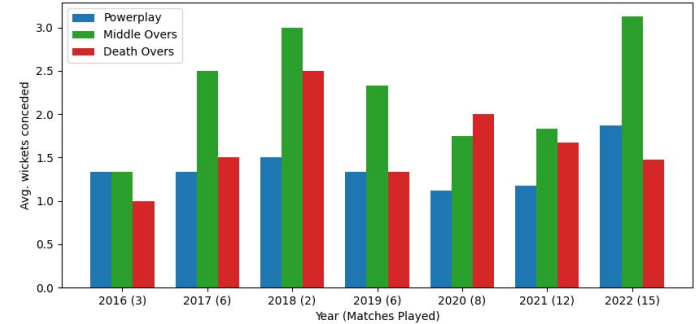
Phase wise breakdown of avg. runs per over scored across all T20 matches batting second since 2016: England



Phase wise breakdown of avg. wickets conceded across all T20 matches batting second since 2016: India



Phase wise breakdown of avg. wickets conceded across all T20 matches batting second since 2016: England



- Loss of wickets for England in the middle overs when batting second as compared to India
- Puts pressure on the batters in the death overs to chase the target, as visible in a below par death overs' run-rate

Breakdown of Performance in Wins

For our analysis, we have defined the following terms as such:

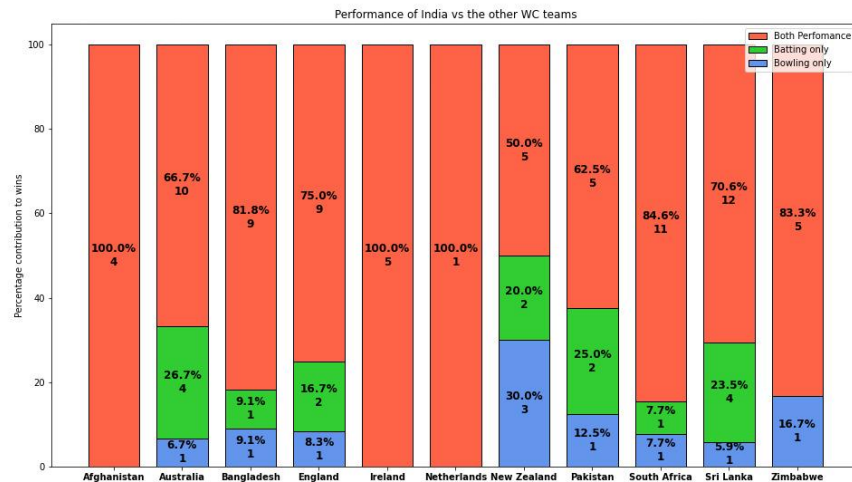
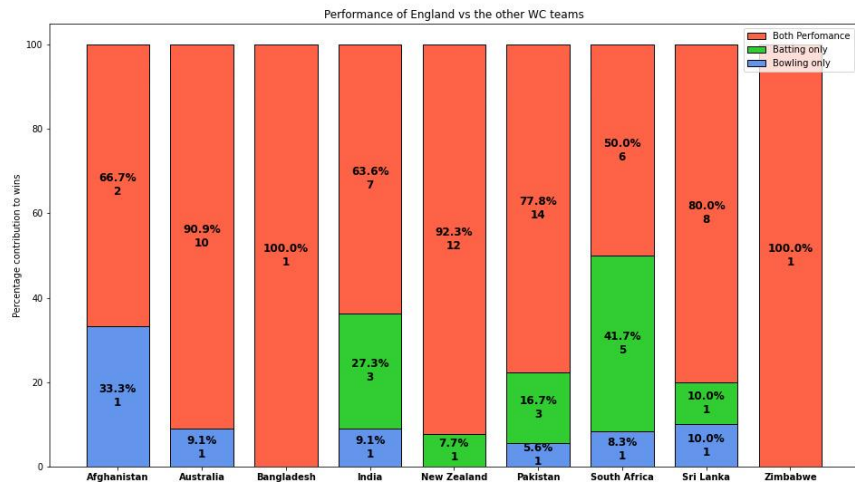
- Good Batting : A team scores higher than the ground averages (batting first) or scores enough to win the game (batting second)
- Good Bowling: A team holds the opposing team under the ground average (bowling first) or limits the batting team enough to win the game (bowling second)



OR



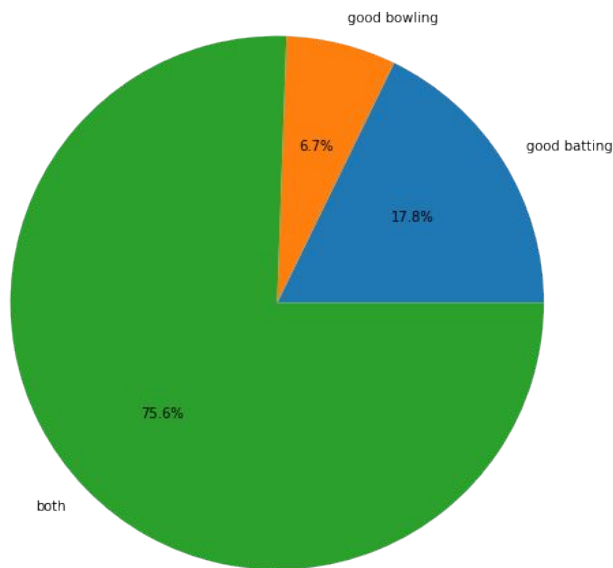
Win Performance Breakdown Against T20 Teams



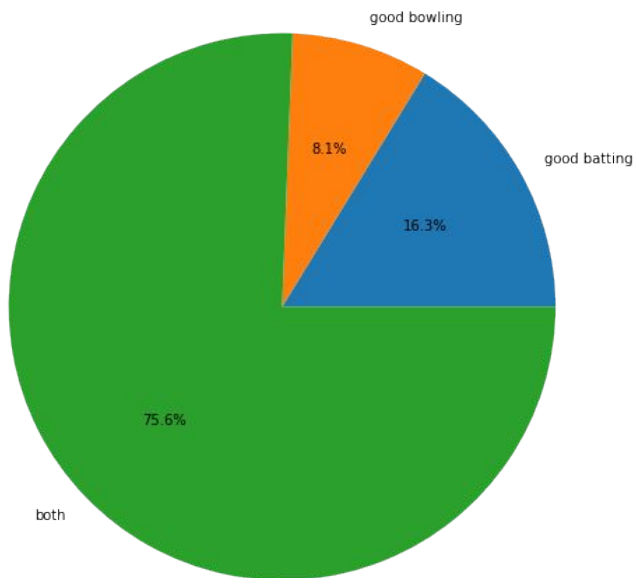
- Both teams are among the most consistent
- Dominant victories with bowling and batting departments performing exceptionally well

Performance Breakdown Totals

Composition of Wins, England

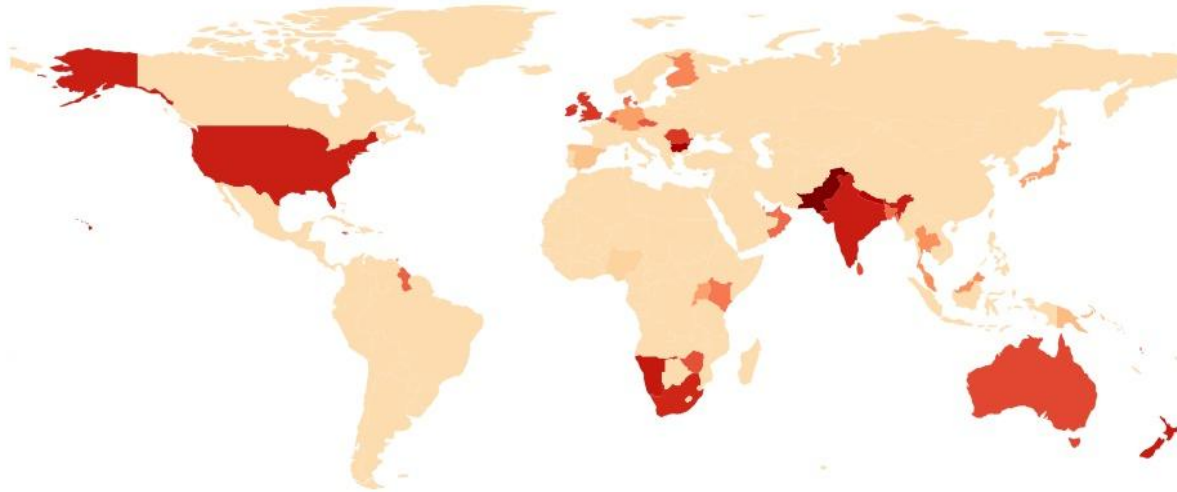


Composition of Wins, India



A higher percentage of both a good batting and bowling performance represents a more comfortable win margin than just the singular categories

Significance of Ground Averages



- Helps in selecting the best playing XI combinations
- Considering ground dimension and pitch conditions, players can be selected based on their strengths



Conclusion and Potential Future Implications

Why bother with past data and a tournament that is already finished?

- Note features that display greatest patterns
 - Here, we saw that win-loss% and win performance correlation served as strong features to determine the team's performance in the tournament
- Find more significant features
- Create a predictive model

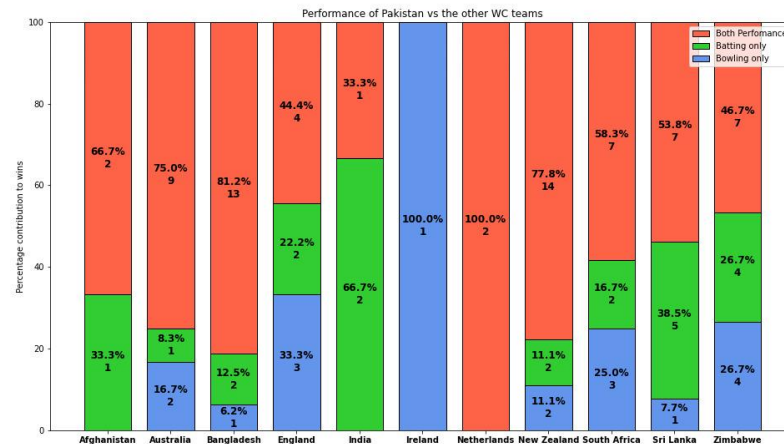
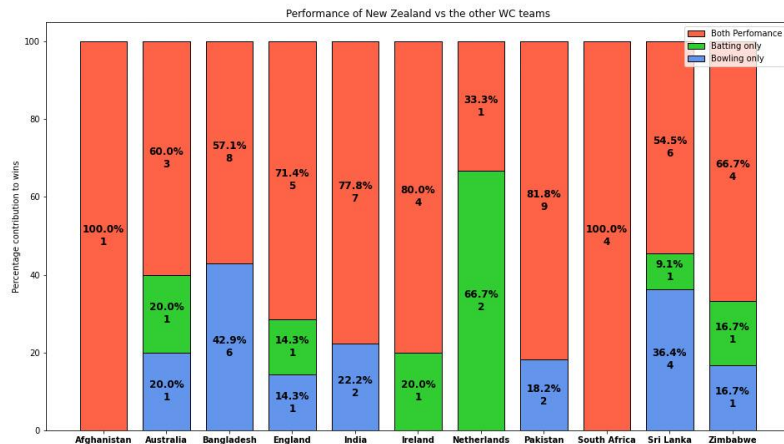
Entertainment, and specifically in this case, sports entertainment is prominent throughout the world

- Global presence of the sport of cricket (104 countries have a national cricket team)
- Betting on cricket (200 billion \$ industry), fantasy cricket, similar applications with money related impact makes such analysis extremely valuable for bettors

Questions?

Appendix

Win Performance Breakdown by Batting, Bowling and Both



Links/References

-

To Add

- Introduction
- Cricket basics
- Data Pre-Processing steps - 1 min
- Metrics - 30 sec
- Win-Loss -
- Average Scores (First/Second Innings) + compared to this wc - 1-1.5 min
- Ground average insights - 1-1.5 min
- Conclusion - 30 sec

Cricket	Baseball
Has several different formats (T20,test,ODI)	One main format
1 over=6 balls	
2 total innings in the T20 format (equivalent to the format of 1 baseball inning)	9 innings (each inning includes the batting and pitching for both sides)
10 “outs” (wickets) per inning for each team	3 outs per inning for each team (27 total outs)
Bowling-ball bounces once before reaching batter	Pitching- direct throw to the batter with no bounce
Teams alternate only once between pitching and batting across the “1” inning	Teams alternate back and forth between pitching and batting across the 9 innings
<p>Batter is out when the wickets are hit</p> <p>*Outs by catch and run outs are similar to flyouts and groundouts in baseball</p> <p>Only two batters on the field at a time until someone is out</p>	<p>Batter is out if they get 3 strikes (most basic type of out, significantly more common than in cricket)</p> <p>One player bats, depending on if previous players reached to a base safely there can be players at each of the 3 bases</p>
Runs achieved by running between the wickets if the ball is in play, 4 runs, or 6 runs	Runs achieved by running through the three bases and reaching back home (without being out)

Coin toss to determine batting/fielding order	Home team always bats second in each inning
Depending on order of batting/fielding, a team can win by runs or wickets	<p>Format of final score:</p> <p>Team A: 2 Team B: 5</p> <p>Team only wins by runs</p>
<p>The number of overs defines the length of the game</p> <p>T20-120 balls total</p>	The number of outs defines the length of the game (a team keeps batting each inning until they reach 3 outs, 27 for the game), so the pitcher throws as many pitches as required to reach that
No requirement to run when a ball is in play	<p>The batter must run when the ball is in play. If someone is on a base they can stay in place unless there is someone on the base behind them.</p> <p>Order: Home-First-Second-Third-Home</p>
11 players field	9 players field

Win Loss % : How did teams fare in the T20 WC 2022

	Win % (Number of matches)	
Team	Batting 1 st	Batting 2 nd
England	100% (1)	80% (5)
Pakistan	33.3% (3)	75% (4)
India	60% (5)	100% (1)
New Zealand	75% (4)	0% (1)
Australia	100% (2)	50% (2)
South Africa	100% (1)	33.33% (3)

	Win % (Number of matches)	
Team	Batting 1 st	Batting 2 nd
Bangladesh	66.67% (3)	0% (2)
Ireland	50% (2)	0% (2)
Afghanistan	0% (2)	0% (1)
Netherlands	50% (2)	33.33% (3)
Sri Lanka	0% (2)	66.67% (3)
Zimbabwe	50% (2)	0% (2)