

Founder Review #5

Repo Readme:

1. Add in a happy path for people viewing the repo also for them to get the best possible experience.
2. Add hyperlinks if possible

Github Workflows:

1. Add a Readme file for it and the idea and desired outcome and flow for it too
2. We are also always getting CI/CD related error emails. Do look into this but not high priority

Analysis:

1. Please mention why and how thresholds have been set, something like a justification for it. Need not be very detailed, can be direct

Data:

1. Perhaps create a folder for Archived files to keep the repo clean
2. IPL_2026_Squads:
 1. A lot of bowlers seem to have been classified as Workhorse seamers. Look into it please
 2. Aman Khan in this CSV is still being mapped to Avesh Khan by the looks of it
 3. Mohammad Izhar also seems to be incorrectly mapped.
 4. The bowler tags seem vague and incorrect - Bumrah is an elite bowler across all phases but is set as Powerplay assassin. So Andy Flower - can you look into standardizing all of this tags and clusters all across the repo and look into just one nomenclature please. Multiple ones are just confusing. Look at 2023 onwards data only please for tagging purposes
3. IPL_2026_Player_Contracts:
 1. There are some players who were retained but show up as "Auction" like Romario Shepherd, Nitish Rana (was traded) etc. Please look into it but low priority

ML Ops:

1. I see in the Readme about Variance and all the tags. But I would also prefer that there be a detailed Product Description document also that is created to highlight what the goal is, desired outcome is, the process, the outcome at the end of it and be detailed for later review purposes also. We dont need this to be kind of a black box really, need to be transparent. Could use a bit of visualisations in the docs too. Kevin de Bruyne could be of use here finally.

Notebooks:

1. Can I get a clarification as to if these are just files which an external user can use to interact with their own custom queries? If so, could you add in another document which gives an overview of how to use it and what are some ways to efficiently code in the queries and the columns also. High level only.

Outputs:

1. Readme File: Tag Categories - Andy Flower to review and standardise across all docs. One standard system only please. Also in the readme file - add the purpose or how the output can help qualitatively help in analysis and just a description of what it is.

2. Matchups:

1. batter_bowler_type_detail_2023.csv

1. For Aiden Markram, for example, I can only see his stats against right arm off spin and right arm pace. File seems incomplete really. Please look into this and also reconcile this with overall numbers in the end to make sure it is rock solid.

2. batter_bowler_type_matchup_2023.csv

1. This also seems incomplete. For example - Devdutt Parikkal and Cameron Green have no pace_balls and pace_strike although they have played against pace since 2023. Please check this, reconcile with overall numbers and make sure it is rock solid and tight.

2. Only 64 players seem to be there so please make sure everyone is included

3. I also need a readme file here to tell me what the classifications are and how they have classified, on what basis please. Basically like metadata for each CSV and how to interpret them.

3. Run-Logs: Again, add a read-me as to what the purpose is so that one can understand

4. Tags:

1. Readme for Tags section also for ease of interpretation and also detail the approach also.

2. Andy Flower - perhaps you need to standardise everything from here itself and make sure it is being adopted correctly all across the outputs. Make sure 2023 tags are only being used for 2023 related outputs please

5. Teams:

1. Readme file for this section also or probably just a description of what's there and not there

Reviews:

1. Domain:

1. Andy Flower Analytics Research markdown file:
 1. I like the Match Phase index and Pressure Sequence ideas but instead of using it for live match. Lets look to adapt this for past matches and understand how teams react in such situations and how often. Similarly for Match Turning Point. Scope this out
 2. We can try Clutch Performance Measurement for both batters and bowlers but how would the expected part work? Fielding is not of great importance right now so ignore and scope it out
 3. Tactical Pattern Recognition: we don't have access to a lot of this data but probably can look at it on a high level by analysing which types of bowlers are introduced in certain situations or how their approach is in certain situations with the batting side of things too.
 4. Batting Intent Classification: it is a decent idea but a four may not be a result of an attacking shot always and a single cannot be a result of a rotation worthy shot. A bit vague so lets keep it on hold for now
 5. We can go ahead with Toss Advantage Index and scope it out.
 6. Can try Novel Composite Metrics as well yes.
2. Sprints: Add in subfolders if possible, seems chaotic

Scripts:

1. Make sure each script is well written, follows good standards and has enough metadata and comments so that anyone can understand the approach and method please.

Statpacks:

1. Player Archetypes don't add any value as of now. Try to use those in other stat pack related tables and embed them that way
2. Same with Key Player tags, embed them in other tables or in the overview table itself.
3. Historical Record vs Opposition: present in a tabular format please
4. Venue Performance: just give record only, runs balls really don't matter on a team level. Probably look at dot ball % and boundary % on batting and bowling side.
5. Please improve tactical insights part. Seems really vague and not well thought of. Has to add enough value to be in the stat pack. Can use Virat Kohli's help to make it more insightful and present it well.
6. Also think about adding new sections. Here are some I want to add going forward, pretty important so scope it out please:
 1. Predicted XI based on experience, fit and best possible outcome for the team in terms of best strike rate, best economy, best boundary%, best dot ball%, role fit, depth in batting and bowling,

variety, making sure matchups are met and are accounted for.

2. Depth Charts for each role and position.

Tests:

1. What is the purpose of this? What is the outcome? Plan? Please mention all of this in the readme files too please

Brad Stevens:

1. You also need to review Agents performance, look at repetitive issues and how they have been handled. Rate them and based out of your analysis, I need you to analyse where we can improve and if retraining/ extra training needs to be given to agents or more agents need to be spawned up
2. A new agent by the name of Jose Mourinho is in the agents page. Get him in the loop and ask him to analyse the whole ecosystem and progress done so far and ask him to give a report and next steps

Docs/Research:

KenPom Method Research:

1. I actually love the idea and looks good to me. But i think because its a small season and we are only focusing on pre-season stuff right now - we can retain the crux of the prediction model to players' individual performances in the past since 2023 and team's expected finishes in the upcoming season.

PFF Research:

1. Seems to be based on qualitative assessment, which currently isn't possible for us as we dont have access to commentary or tape