

# Founder Mega Review 1

## 1. Product Clarity -

1. We aim to build a magazine-style preview generator that is powered by an analytics warehouse that can enhance the user experience and make them more informed. One of the USP is to build structured, opinionated cricket analysis that looks like a pro team's internal prep packaged for public consumption
  2. Think of us as scouting-style cricket editorial that presents roles, depth charts, matchups, possible Playing XI, structures, skill-gaps and strengths and tendencies using only freely available ball-by-ball data, structured the way most teams internally prepare. Similar to Lindy's and Phil Steele's publications.
2. Florentino Perez (new agent): i think we should standardize the workflow and have a loop for each release cycle and task. Call this Loop the Task Integrity Loop and create a Readme file for this and possibly a new folder to track this as needed
1. Something like Quality Gates per each task and the steps could be as follows:
    1. Step 0: Task declaration and creation of a PRD
    2. Step 1: Florentino Gate: Scope and Value. Question that needs to be answered is "Does this task materially improve the paid artifact or strategic decision". Possible outcomes are Approved, Analytics only, Not Approved
    3. Step 2: Build, work happens only within scope and no opportunistic additions unless asked for
    4. Step 3: Domain Sanity Loop - every agent answers a single focused question. Jose Mourinho answers is this robust with current data and are baselines clear and is this scalable in the future. Andy Flower - would this make sense to a coach or an analyst or a fan?. Pep Guardiola - is this structurally coherent or does it contradict the system. No big essays, just like a sign-off - Yes/No/Fix
    5. Step 4: Enforcement Check. This is where Tom Brady comes in. He isn't judging anything, he checks if loop is followed, were there any objections addressed or logged, was scope respected. If not then task is sent back
    6. Step 5: Commit and Ship
    7. Step 6: Post Task Note: make sure readme files are appropriately added, mention what changed, what assumption was tested and what risk is introduced. What is the USP of this change.
    8. Step 7: System Check to make sure the structure of the schema is well-maintained and is in tact.

3. Brad Stevens/Florentino Perez/Tom Brady - draft a constitution document that defines authority, formalizes agent roles, enforces loops, specifies scope and boundaries, codifies decision rights. Clear product idea, USP and what we are working through. Classifying Editorial and Analytical teams and a formal definition of Done. We need to make sure agents dont overlap, loops dont skip, no scope creep. Graduation rules, i.e from Analytics to Editorial. Make this and dont commit before I approve
4. Documentation - there are some outdated documents so lets make sure to either archive them. A readme says 34 views are there but older manifest says 43 and current .py file has 35 create or replace files. Be absolutely thorough in this and probably have 1 single source of truth, which automatically updates that manifests automatically on generation.
5. Repo Structure and Hygiene: Add .DS\_Store to .gitignore and remove committed ones please
6. Data Strategy and Credibility:
  1. Think of adding a Data/Provenance.md file to show where the squads and contracts came from or what the sources are
  2. Perhaps committing the full DuckDB may be inefficient. Lets try to make this scalable
7. Pipeline and Engineering Execution:
  1. We probably can structure this better. Have things like expected schemas per output, column definitions and datatypes and stable naming conventions across all versions - we will probably need N'Golo Kante, Pep Guardiola or a new agent (depending on Brad Stevens review) to do this
  2. We could add schema contracts that tests assert against
8. Outputs Quality:
  1. The outputs manifests seem to be stale. Perhaps a fix could be that for each run, we generate an outputs changelog where number of rows changes, new columns, new files, hash of each file etc are mentioned
  2. Add a "Limitations" section in perhaps the readme to state caveats or missing pieces of info
9. Analytics Methodology and Model Rigor:
  1. So the ML Ops MD file says K-Means and PCA, it does end up giving variance numbers, cluster counts but they dont align with the role losts and elsewhere they are different archetype counts. Can we verify first to see if there is any issue and then if there is, then scope it out?
  2. We also need validation or working proof that the clustering is solid across seasons, a silhouette score/inertia trends or sanity check examples
  3. We could also add baselines: simple heuristics vs tags. For each

tag or archetype - show baseline metric, tagged metric and delta

1. Delta: if baseline heuristic is 135 strike rate and if tag is Powerplay Opener for example, delta is actual sr - baseline that's all.
2. The delta could also be used as our defense against saying "isn't this just obvious cricket knowledge"
10. Testing: The repo may need schema tests, output file existence, column set tests, no null in key id tests, manifest update tests
11. CI/CD and Automation: CI/CD should run lint/pre-commit, unit tests, output validation and generating sample artifacts and compare. Try and get this to work
12. Governance/Review System:
  1. I think for an outsider, all of this may be chaotic. So let's try and add a one-page markdown file saying How Does This All Work which includes details like:
    1. what agents do
    2. what reviews mean
    3. what the approval chain is
    4. where the published artifacts are etc and what the flow of work looks like
    5. Who is responsible for each step
    6. Review and Approval Gates - Stage, Owner and what must be solved/answered
    7. Each metric should have a baseline, no silent assumptions wherever possible
  2. The idea is that someone looks at the repo or sees it, they need to know how things move and this could be detailed as well.
13. Possible Roadblocks:
  1. We need to keep in check over-interpretation risk because some signals may be noise, limited by sample and context dependent. Founder will try but this will be Andy Flower and Jose Mourinho (new agent's) responsibility. Prioritize stable and foolproof insights which can be backed rather than flashy ones
  2. Editorial Discipline needs to be maintained. The Cricket Magazine editorial parts with Virat Kohli, Kevin de Bruyne and LeBron James must be kept separate from the Analytical lab which includes Ime Udoka, Steph Curry, Brock Purdy to name a few
14. A note to all agents - Cricket fans don't pay for analytics, they pay for confidence, narrative clarity and authority. Gives fans a sense of forbidden knowledge holder, teaches them a way of thinking. Most fans react after matches but we need to be ahead of the curve, need to have clear stances and confidence without bullshit certainty. What will make his money in the end is us committing to ruthless editorial compression, strong opinions based on transparent data and zero temptations to be only analytical heavy.

