

## Brainstorming Workshop Data-Driven Catalyst

Ideas submitted by the community, collaborators, ATH, Sapient team members:

- 1) Scrape Review Module URLs and output JSON format (or TXT or CSV etc)
- 2) Obtain Github, Linkedin etc from proposals and add to funds sheets – treasure trove of new metrics like stars, forks, commits etc
- 3) Plot number of reviews per proposal, score and score variance across funds (identify upvote own/downvote others behavior)
- 4) Thesholds of getting funded in each category, amount of funded ADA vs ADA requested
- 5) Data-driven hackathon, Catalyst improvement hackathon
- 6) LangChain without the expensive OpenAI API – open source alternatives and training “our own”
- 7) Funding txs on the blockchain as part of the dataset
- 8) Activity on voting sidechain as part of the dataset
- 9) Number of words per review and average ADA paid to CRs/CAs
- 10) Number of words per review and number of total reviews
- 11) Better AI filtering for CR module than current default flagging in the LVL2 module
- 12) Identify if “reviewer has read the proposal” (NLP)
- 13) CR scoring “category concentration” – combine with idea 3

---Work in progress---

First Workshop in June M2?