

Team Meeting 4

15 SEPTEMBER 2022/ 11:00 AM / IN PERSON

Attendees

Ke, Oliver, Xavier, Tommy, Glendon

Agenda

- ☒ Discuss the ETL script
- ☒ Fraud detection
- ☐ Aggregation woes

Notes

- The ETL script is 90% done at this point, we only need confirmation of what aggregated tables to calculate (we may also need to read in extra data for this).
 - There are also some nitpicks. Specifically, formatting of section headers in the script
 - We also need to test on Tommy's machine, since it was still not reading transactions as of last testing.
- Tommy's deriving bounds for customer rates based on Poisson distribution (instead of the original Normal distribution assumption).
- Fraud detection is hard
 - Xavier has investigated slightly what this would look like with some probability algebra mixed in
 - The goal would be to assign a probability of fraud per transaction, then create a linear regression on the transaction information.
 - Ke has been working on deriving the distributions of the fraudulent data so they match up correctly.

Action Items

- Final list of aggregated tables to generate/write (Xavier).
- Fraud detection finalize fraud probability model (Ke).
- Develop a final ranking heuristic (Rest of the team).

- Values that go in. Scaled? Normalized?
 - Customer rating. Setting a threshold for VIP customers based on Poisson distribution (Tommy).
 - Economic rating
- How they get combined.

Next Meeting

Planned for next Thursday during the workshop.