Data Take Home Project

Background

Monarch relies on users connecting their bank accounts to the application using third party data providers. Multiple data providers may support connections to a bank, with one as the preferred connector and the others as backup connectors. If a user cannot successfully connect to their bank accounts, it reduces the value Monarch is able to provide them.

Problem

It is important for Monarch to understand a user's success rate in connecting to their banks. For this exercise, we'd like you to come up with a scoring system to measure this. It should be:

- Granular enough to see a specific user's score
- Possible to aggregate to a specific institution-data provider pairing
- Possible to see changes over time

Terminology

- Institution - the bank a user uses (e.g. Chase)
- Data Provider - the third party service that connects Monarch with an institution (e.g. Plaid)
- Credential - a unique ID for a user's link between an institution, data provider, and Monarch (for example, Matt's Chase accounts linked to Monarch via Plaid)

Provided Data

We can determine what we call First Connection Success based on three events that our platform fires:

1. Connect Institution Selected - this fires when a user starts the connection flow by searching for their institution. In the data this is labeled as `type = connect`

2. Credential Created - this fires when a credential is successfully created. In the data this is labeled as `type = created`
3. Credential Disconnected - this fires if a credential gets disconnected from the institution. In the data this is labeled as `type = disconnected`

> 💡 **Data Notes**
>
> - All data has been irreversibly anonymized (name changes, ID changes)
> - If you see a `created` event without a corresponding `connect` you can ignore it
> - If you see a `disconnected` event without a corresponding `connect` or `created` you can ignore it
> - We only consider a credential successfully connected if it stays connected for 24 or more hours

Deliverable

- Take the provided CSV files and put them into a database of your choosing (such as SQLite)
- Come up with a First Connection Success rate metric. Share the definition in words as well as the queries used to create the metric
- Share any findings based on your output. Include your queries for all output. Example findings might include:
  - Success rate by institution & data provider (remember that a credential is a connection to an institution via a specific provider)
  - Disconnection stats - does a user see an initial connection followed by a disconnection within 24 hours (we call this a Quick Disconnection)?

    - Differences across data providers
    - Differences across time periods
    - Any anomalies you find in the data
    - Anything else you find that is interesting

Format

- Share your queries any way you choose (Github, Google doc, etc)
- Share your findings in a presentation or a Google doc (no preference, whatever you find easier to work with)

Questions

If there is any clarifying information you need, please reach out to Matt. This exercise should include everything that is needed for this analysis, but if additional data points would be helpful, we are happy to discuss.