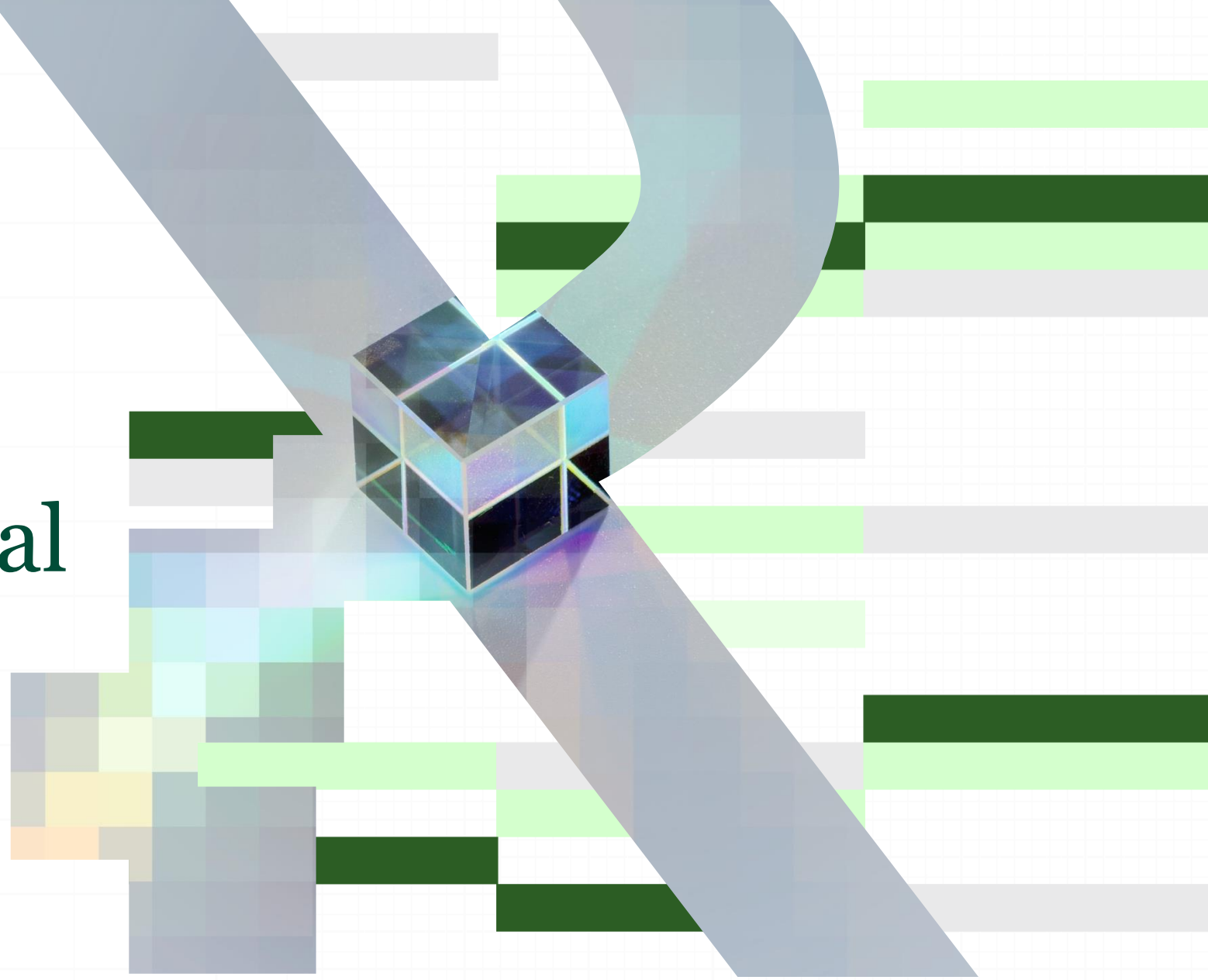# Venmito Data Science Technical Assessment

Design Guidelines

1. The case should appeal to both analytics and ML candidate profiles

2. The case should have a clear problem statement, but allow for diverse technical approaches that arrive at a similar solution

3. The case will contain problem-solving elements, but will still prioritize the technical part (we already have an on-site PS case… several ones actually…)

Modifying Venmito for DS: What Stays, What Changes

1. **Data Ingestion**: Your solution should be able to read and load data from all the provided files. Take into account that these files are in different formats (JSON, YAML, CSV, XML).

→ **Stays the same**

2. **Data Matching and Conforming**: Once the data is loaded, your solution should be capable of matching and conforming the data across these files. This includes identifying common entities, resolving inconsistencies, and organizing the data into a unified format. Furthermore, the consolidated data should not only be transient but also persistent. This persistence should be achieved using appropriate methods such as storing in a file, database, or other suitable data storage solutions, and not restricted to just a variable in memory. This way, the integrity and availability of the consolidated data are ensured for future use and analysis.

→ **Stays the same**

3. **Data Analysis**: Your solution should be able to process the conformed data to derive insights about our clients and transactions. This would involve implementing data aggregations, calculating relevant metrics, and identifying patterns. These insights will be invaluable in helping us understand our clientele and transaction trends better. **Here are just a few ideas to get you started—don't limit yourself to just these examples, think outside the box and come up with your own metrics as you work through the project!** For example, you might look into:

   - Which clients have what type of promotion?
   - Give suggestions on how to turn "No" responses from clients in the promotions file.
   - Insights on stores, like:
     - What item is the best seller?
     - What store has had the most profit?
     - Etc.
   - How can we use the data we got from the transfer file?

→ **Slight changes (i.e., more direct problem statement)**

4. **Data Output**: The final output of your solution should enable us to consume the reorganized and analyzed data in a meaningful way. This could be, but is not restricted to, a command line interface (CLI), a GUI featuring interactive visualizations, a Jupyter Notebook, or a RESTful API. We invite you to leverage other innovative methods that you believe would be beneficial for a company like Venmito. Please provide at least 2 data consumption methods, 1 for the non-technical team and 1 for the technical team.

→ **Mostly gone**

5. **Code**: The code for your solution should be well-structured and comprehensible, with comments included where necessary. Remember, the quality and readability of the code will be a significant factor in the evaluation of the final deliverable.

→ **Stays the same**

Xtillion

Problem Statement

Propose a data-driven strategy or solution for Venmito to improve customer engagement by increasing promotion responses and encouraging ongoing customer activity

Potential approaches to solve the problem

- **Approach #1 (Analytics – Segmentation Analysis):**
  - Solution could analyze group-level insights (e.g., across demographics, cities/countries, devices, etc.) and response patterns to promotions
  - Deliverable: <u>strategic</u> recommendations such as having "playbooks" for certain customer segments and targeted marketing campaigns (e.g., x promotion to iPhone users, y promotion to users in New York, etc.), eliminating certain promotions due to ineffectiveness and replacing them with others, etc. OR also the "best" promotion to send to each segment
  - Methods: Dashboards, bar charts, etc.

- **Approach #2 (ML – Recommendation System):**
  - Solution could learn patterns using data
  - Deliverable: individual-level predictions
  - Methods: Collaborative filtering, supervised learning (e.g., multi-classification to predict probability of answer for all the promotions – very hard with 50 classes), or other recommendation algos

Xtillion