Fourth: Communicate with Stakeholders

Construct an email or slack message that is understandable to a product or business leader who isn't familiar with your day to day work. This part of the exercise should show off how you communicate and reason about data with others. Commit your answers to the git repository along with the rest of your exercise**.**

Email:

Subject: Enhancing Fetch Rewards' Data Reliability – Insights & Questions

Hello (Product/Business Leader's Name),

Hope you're doing well!

As part of our **ongoing efforts to strengthen Fetch Rewards' data infrastructure**, we've been evaluating data quality and system integrity.

While our analysis didn't reveal urgent issues, we identified **several inconsistencies that could impact engagement tracking, financial reporting, and long-term scalability**. Addressing them now will **enhance reporting accuracy, streamline user rewards processing, and support better business decisions**.

Using a combination of data validation checks, schema enforcement, and exploratory analysis, we flagged several inconsistencies, including:

Missing Data:

- **12.5% of users lack lastLogin timestamps**, making engagement tracking less reliable.
- **448 receipts are missing purchaseDate**, affecting when transactions are recognized in reporting.
- **435 receipts have no totalSpent values**, impacting revenue insights.

Data Integrity Concerns:

- **148 receipts reference non-existent users**, which may be due to ingestion errors or missing cleanup.

- **77 users have no linked transactions**, potentially indicating inactive or test accounts.

Unusual Spending Patterns:

- **55 transactions show unusually high spending**—these could be bulk purchases, but they stand out significantly.
- **The spending distribution is highly skewed**, requiring adjustments for trend accuracy.

To ensure our fixes align with **business expectations**, we need clarity on:

1. **Data Ingestion & Processing:**

- **How are user and receipt records created?** Are they sourced via API or external integrations?
- **What visibility do we have into ingestion pipelines —** logs, error handling, and tracking failed records?
- **Are there known operational delays in how transactions are processed? Do delays impact when rewards are awarded?**

2. **Business Rules & Data Handling:**

- How should missing values be handled? Should **purchaseDate be estimated** from dateScanned, flagged, or removed?
- Should totalSpent default to **zero, be interpolated**, or excluded when missing?
- What's the expected **lifecycle of user and receipt data** — do users ever get deleted, and should orphaned receipts persist?
- Are there **known benchmarks for transaction outliers**? Should we flag bulk purchases as potential fraud, or are they expected behavior?

To enhance **data structuring and reporting**, we'd benefit from:

- **Business Metrics & Reporting Needs:**
    1. What are the **key KPIs for Fetch Rewards** — revenue tracking via totalSpent, engagement based on pointsEarned, or a combination?

2. How do **marketing and product teams leverage transaction data** — to analyze spending trends, track rewards usage, or detect fraud?
3. Have there been **historical issues** with missing data, ingestion delays, or inaccurate reporting that affected user rewards or partner analytics?

- **Data Infrastructure & Scalability:**
    1. What's the **expected transaction volume** now and in the next 6-12 months?
    2. Are we anticipating **higher concurrent queries**, requiring **optimized Snowflake workloads**?
    3. Do we need real-time data validation, or is it acceptable to correct issues in scheduled updates?
    4. Would integrating **AWS Glue / DBT** help streamline ETL workflows and enforce schema validation before ingestion?

Looking ahead, we anticipate **higher data loads and complex analytical queries**, which could impact **processing speed and storage costs**. To mitigate this:

1. **We propose leveraging Snowflake's scalable architecture** to efficiently handle large datasets.
2. **Automating anomaly detection models** will allow us to proactively flag data inconsistencies.
3. **Using AWS Glue & DBT for schema validation** will reduce ingestion errors before they reach production.

We'd love to align with you and your team to ensure we're making informed & scalable improvements. Let me know the time that works for a quick call!

Looking forward to your thoughts.

Best,
Rohith
Analytics Engineer | Data Quality & Analytics Team | Fetch Rewards

Note-: This email includes certain assumptions regarding the tools, data processes, and team workflows to ensure a well-rounded perspective for all the questions asked**.**