04. 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.

- What questions do you have about the data?
- How did you discover the data quality issues?
- What do you need to know to resolve the data quality issues?
- What other information would you need to help you optimize the data assets you're trying to create?
- What performance and scaling concerns do you anticipate in production and how do you plan to address them?

Email-

Subject: Data Quality Assessment and Optimization for Improved Decision Making

Hi [Product/Business Leader's Name],

I hope this email finds you well. I wanted to discuss the recent analysis we conducted on our data assets and highlight some important findings that can significantly impact decision making within our organization.

1. Data Quality Assessment:

During our analysis, we identified a few data quality issues that are worth addressing. These issues include missing or null values in certain fields and inconsistent data formats across the dataset. These issues can potentially impact the accuracy and reliability of our analysis.

To identify these data quality issues, we utilized data profiling techniques, such as examining missing value counts, analyzing data types, and performing exploratory data analysis. These methods helped us gain insights into the overall health of our data and detect areas that require attention.

2. Resolving Data Quality Issues:

To resolve the data quality issues, we have taken the following steps:

- For missing values, we have employed different strategies depending on the context, particularly via Python. In some cases, we removed the rows with missing values, while in other cases, we filled them with appropriate default values or applied interpolation methods to estimate missing values based on neighboring data points.
- Inconsistent data formats, such as different date formats or non-numeric characters in numeric fields, have been addressed by converting the data to the correct format using suitable data transformation techniques.

However, it would be helpful to collaborate with the data engineering team to ensure that these data quality improvements are implemented in the data pipeline to maintain consistency and accuracy over time.

3. Optimization of Data Processes:

In order to optimize the data assets we are creating, we would need additional information and collaboration from various stakeholders. To enhance the quality and usefulness of our data, we plan to implement on the following points and would appreciate insights:

- Clear understanding of the specific business goals and objectives that the data assets should support.
- Identification of key performance indicators (KPIs) and metrics that are most relevant for decision making.
- Collaboration with subject matter experts to ensure the data assets capture all the necessary dimensions and context required for analysis.
- Collaboration with Data Engineering team to smoothen the ETL processes and Data warehouse and decide upon a fast and heavy working consistent data warehouse system along with finalizing on the cloud counterpart and connection as well

4. Performance and Scaling Concerns:

As we move towards production, it is important to anticipate and address any performance and scaling concerns. Some questions I think we should consider are:

- What is the expected volume and velocity of incoming data?
- Do we have the necessary infrastructure and resources to handle the increasing data load?
- Are there any specific performance requirements from stakeholder's end regarding important KPI's, or regarding response time or data freshness, that need to be prioritized?

To address these concerns, we plan to work closely with the data engineering and IT teams to ensure the necessary scalability, performance optimization, and infrastructure planning are in place. Regular monitoring and periodic performance reviews will be conducted to identify any bottlenecks and make proactive adjustments as needed.

I would love to have a discussion with you to further explore these points and align our efforts to maximize the value and reliability of our data assets. Please let me know a convenient time for a meeting or if you have any specific questions or concerns.

Thank you for your attention and support in advancing data-driven decision making within our organization.

Best regards,
Aditi Namdeo
Data Analyst, Fetch