As a Data Engineer at Lloyds Banking Group, I encountered a significant challenge related to our Performance Goals Dashboard. Users began raising numerous bugs due to incorrect data, which required a time-consuming investigation process. This reflection will analyse the problem-solving process, continuous improvement principles applied, and lessons learned from this experience.

The primary issue was the incorrect data on our Performance Goals Dashboard. Users frequently reported bugs, and each investigation required manually locating user data in a large source file, taking approximately 10 minutes per instance. The potential impact was significant, affecting user trust and operational efficiency. I used Jira to track the reported bugs and monitor the progress of implemented solutions. This provided insights into the frequency and nature of the data issues.

Upon identifying the issue, I conducted an initial risk assessment to understand the scope and potential impact. This involved reviewing the reported bugs and determining the frequency and severity of the data inaccuracies. I also considered the potential consequences of incorrect data on user trust, project timelines, and overall operational efficiency. It was clear that the current visual did not give the user sufficient information as to what actions needed to be taken. Fig 1 below shows the visual the user could see.

Fig 1:

A screenshot of a computer

AI-generated content may be incorrect.

I promptly escalated the issue to the product owner and data privacy manager to ensure that all relevant stakeholders were aware of the problem. Clear communication was maintained throughout the process to keep everyone informed of the progress and any new developments.

To address the immediate risk, I developed a Python script in JupyterLab to streamline the data retrieval process for bug investigations. This script significantly reduced the time required to retrieve raw data, providing answers within seconds. See Fig 2 below for my script, output and user interface for my team with the relevant fields Highlighted.

Fig 2:

A screenshot of a computer

AI-generated content may be incorrect.

However, during the review of the raw data, I discovered an internal data breach due to personal information in the source data. This prompted an urgent need to assess the depth of the breach and take corrective actions. I conducted a thorough investigation to determine the extent of the data breach. This involved reviewing the source data to identify all instances of personal information and assessing the potential impact on data privacy. I reported the breach to our data privacy manager to log the incident and ensure compliance with data protection regulations.

I reached out to the data owners who provided the source data to explain the issue and discuss potential solutions. I collaborated closely with our People & Places team to understand how the personal information ended up in the source data and to develop strategies to prevent future occurrences. I identified that the field containing the problematic data was not used in our pipelines or Dashboards. Therefore, I consulted with the People & Places team to provide a bespoke dataset with the field removed completely. This reduced the risk of future data breaches as well as significantly reducing file size. This also improved our sustainability as we were no longer using as much cloud storage for the raw data. I also recommended improvements to our data handling processes, including the implementation of automated data validation scripts and the integration of Jira Align with other project management tools to streamline workflows.

To ensure the effectiveness of the implemented solutions, I set up continuous monitoring using Jira's built-in reporting tools and custom dashboards. Feedback from colleagues indicated a noticeable improvement in and data accuracy and the number of Bugs raised by our users confirming the success of the risk management measures. This experience highlighted several inefficiencies, including the manual data retrieval process and the presence of unnecessary fields in the source data. By removing the problematic field from the source data his risk of future data breach was significantly reduced. Additionally, I suggested improvements to our Power BI Dashboard visuals to provide users with direct answers quickly. This finished version can be seen in Fig 3 Below.

Fig 3:

A screenshot of a computer

AI-generated content may be incorrect.

To ensure quality and innovation, I conducted peer reviews of the Python script and followed best practices in data engineering. This included using version control systems and continuous integration tools. The issue prompted a review of our data value extraction processes. By optimising the data retrieval process, we were able to derive more actionable insights from our project data. The challenge led to exploring cost reduction and system efficiency improvements.

The experience reinforced the importance of adhering to software development principles in data engineering. This included maintaining clean code, following agile methodologies, and ensuring robust documentation. I learned the value of adopting best practices in data pipeline management, system architecture, and tooling. This influenced my approach to collaborating with data scientists, analysts, and developers, fostering a culture of continuous improvement and innovation.

In summary, this experience provided valuable insights into problem-solving and continuous improvement within data engineering. By addressing the inefficiencies in our Performance Goals Dashboard, we were able to enhance workflow efficiency, improve data accuracy, and drive innovation. The lessons learned will inform future improvements and contribute to the overall success of our projects.

Additionally, this experience underscored the importance of proactive risk management and the need for robust data governance practices. By identifying and addressing potential risks early, we were able to mitigate the impact on our operations and maintain the integrity of our data. This proactive approach not only improved our immediate response to the issue but also laid the groundwork for more resilient data management practices in the future. The collaboration with various teams, including People & Places and data privacy, highlighted the value of cross-functional teamwork in addressing complex data challenges. This approach ensured that all aspects of the issue were considered and that the solutions implemented were comprehensive and effective. The continuous feedback loop established through regular monitoring and communication allowed us to make iterative improvements and adapt our strategies as needed, further enhancing the overall effectiveness of our data management processes.