

Problem Overview:

Johnny's Jeans is a fashion start-up and a new customer of ActionIQ. The first campaign they want to create on our platform is a Thank-You message to all their first customers. They want to target their first 10,000 customers and send them some stats about their first year of sales.

The Johnny's Jeans finance team reported that the first 10,000 customers spent a collective \$33,904,525 on their site. However when the marketing team ran the sum of total revenue on the AIQ platform, the total sum was a bit less at \$29,162,044.

The Johnny's Jeans marketing team was concerned about this data discrepancy and has reached out to AIQ.

Background of our integration with Johnny's Jeans:

1. Every few weeks, Johnny's Jeans data warehouse does an export of their most recent rows from their user-summary table. This data is encrypted and dropped into an AWS S3 bucket.
2. AIQ's server is designed to ping that S3 bucket every 24 hours and when a new file is dropped, AIQ's system is supposed to retrieve the new file for decryption.
3. The decrypted user-summary data is in a CSV with these fields:
 - a. user_id
 - b. Total_spend (total this user has spent at Johnny's Jeans)
 - c. count_saved_items (total amount of book-marked items)
 - d. loyalty_credits (amount of loyalty points the user can spend)
 - e. batch_id
4. AIQ's data pipeline is designed to read through all the data files and de-duplicate the data by choosing the largest batch_id for each user. This ensures that AIQ should only be loading 1 row per user.
5. Once that data is de-duplicated, it is loaded into AIQ's platform and is ready for users to start running queries.

Task:

1. Figure out best you can what the problem is. Bear in mind that just like any software engineering effort, AIQ is not immune to bugs and things may not work perfectly as expected.
2. You are writing an email to the AIQ backend engineer describing what you've found and what you think the issue here is. Please provide as much information and/or examples as possible.
3. Michael Greene, Johnny Jean's email marketing director, emailed us asking why our data doesn't seem to be matching up with theirs. Write a response for him.

Task Notes:

1. We have provided the 5 of the decrypted files that were sent to us by Johnny's Jeans.

- a. Delta0.csv was their initial base set of data.
 - b. Deltas 1-4 were their update files.
- 2. We have also sent you AIQ's version of their user-summary table (AIQ-user-summary.csv). Johnny Jeans' has exported the table on their end and we've sent you that as well (Johnny-user-summary.csv).