Pair Programming Activity

We have some JSON-formatted data that represents the history of pieces of work moving through an automated system. Each history entry is comprised of the following:

- **id**: A unique ID for this history entry
- **piece_id**: The piece of work being operated on
- **status**: Number indicating the operation being performed on the piece
- **user_id**: ID of the user that performed the operation in this entry
- **start_time**: Time that the piece began being processed in the status
- **end_time**: Time that the piece finished being processed in the status; the difference between end_time and start_time indicates how long the piece spent in the status

The data file is located here: http://it-recruitment.mintel.com/testing/test_data.json

Lets pair program through answering these questions.

Question 1

How many unique statuses are in the data set?

Question 2

As mentioned previously, the user_id field in the data set indicates the user that moved the piece into the given status, which is considered one operation; each history entry had one user perform one operation to put it in that status. With that in mind, in descending order, list out the top 5 users by the number of operations performed and the number of operations performed by that user. For example:

1. user6: 12345
2. user9: 9999
3. user1: 7920
4. user5: 5801
5. user2: 1088

Question 3

On average, how long does a piece spend in status 8951?

Question 4

Given that a status ending in 3 represents an error status, what percentage of pieces in this data set end up in an error status at least twice?

Question 5

What is the most common path for a piece to follow through the system?

Your overall thoughts on this exercise