**Data Analyst Test**

The data provided [HERE](https://drive.google.com/file/d/15QqFVfpL2eYkx9qYYCOzR8MD5CzC_fWK/view?usp=sharing) contains information about a random selection of 5,000 players who installed one of our apps in the last ~30 days.

The dataset provided contains one file, players.csv.

The columns in the files are defined as follows.

level\_progress.csv:

* event\_datetime: The date and time at which the event was received (Local time)
* player\_id: A unique identifier for each player.
* level\_number: The level number the event corresponds to.
* status: The outcome the event corresponds to. (start, fail, complete)
* session\_id: A unique identifier for the session in which the event was produced.

\*complete means the player was successful.

**Task**

Q. On which level are players most likely to fail?

You should consider the statistical significance of your answer carefully.

You should include any code used to process the data and any calculations used to compute an answer.

**Note: You don’t need access to edit this document, please create a new document with your answer and reply to this email with the attachment or the sharing link.**