

# AI Product Management: Figure-Eight Medical Image Annotation Job Results

Ben Jacobson  
April, 2020

## Project Overview

### Project Goal

The goal of this project is to launch the designed data annotation (labeling) job using Figure-Eight. We will:

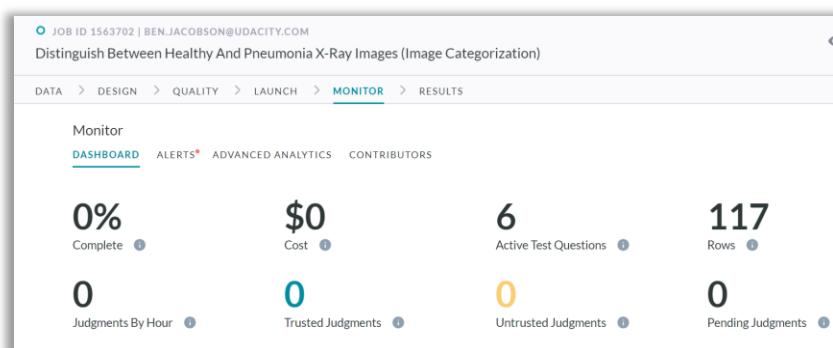
- Report the progression of results
- Display Results, Discussion, & Next Steps

This project's goal is primary to learn and to test this process, being exploratory in nature. It will certainly be considered and evaluated as though it were being used in a larger project (a large labeling job).

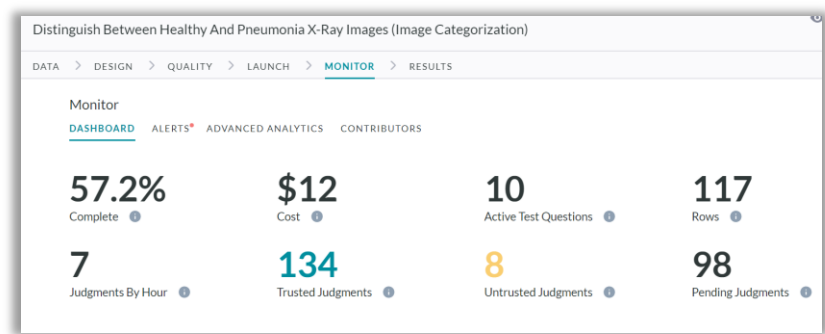
## Process Review

### Setup & Initial Progress

Here is an image of the Monitoring Dashboard, which was launched. This cost just under 25 USD to run, with two people reviewing each image.



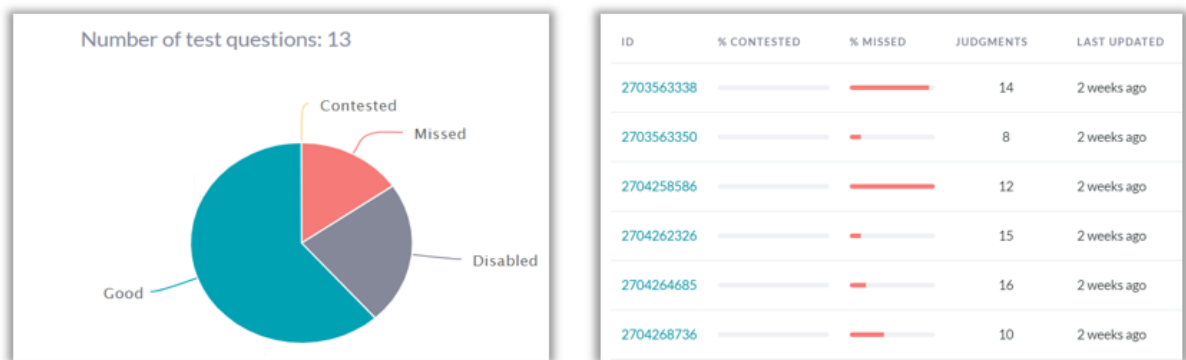
As the labeling job progressed, I reviewed the information. Everything looked alright initially, but I noticed that the test questions were not being reported. This is in the alert, and I am not certain why the platform did not work correctly. However, I paused the job, and corrected the issue, which is good.



This is the labeling job at approximately 57.2% completed, after about 15 minutes. My 12 USD has 8 untrusted judgements and 134 trusted judgements.

Survey Results

At this point in the job, we can now review some of the test questions, and this is showing that of the 13 questions, of which 3 are disabled, 8 questions are mostly answered correctly, and 2 questions are missed the majority of the time. This is a good source of information to clarify, or to review that test question. (Note that the initial design had 8 total questions, but more were enabled to meet the launch requirements in the alert.)



For these two test questions, nearly 100% of respondents missed them. Rather than selecting unknown, “no” was selected in almost all cases. These two questions were then disabled and would be reviewed later in another test job before a full rollout.

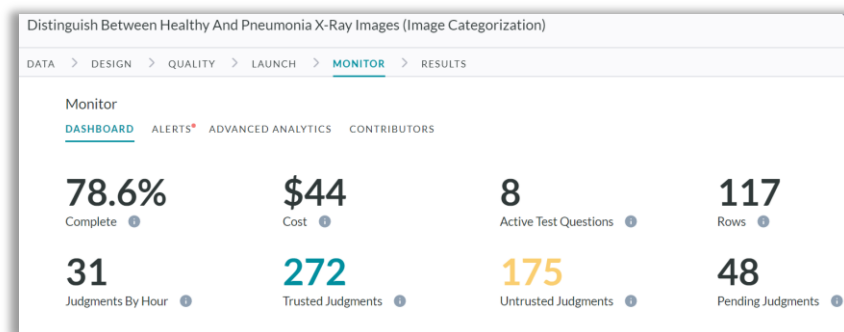
## User Satisfaction

There was one unhappy contributor, but the overall is 4/5 from 4 respondents. I will consider this an initial success since this is the first attempt at using the platform.



## Results, Discussion & Next Steps

Well, initially 25 USD was allocated, and 44 USD were needed before I stopped the job, so it was 76% over budget! Reviewing the platform's documentation, the initial 25 USD was only an estimate. Though this was over budget, it did indicate two design problems. First, disabling test questions can cause more "untrusted judgements", and second, the two question missed by most drove up this number as well, resulting in more cost. So a better understanding of the platform, and better test questions are needed. Given that this is a learning experience, it is far better to be over budget on a test, that a full-scale launch. Another test would be designed and launched with updated test questions, and clear instructions.

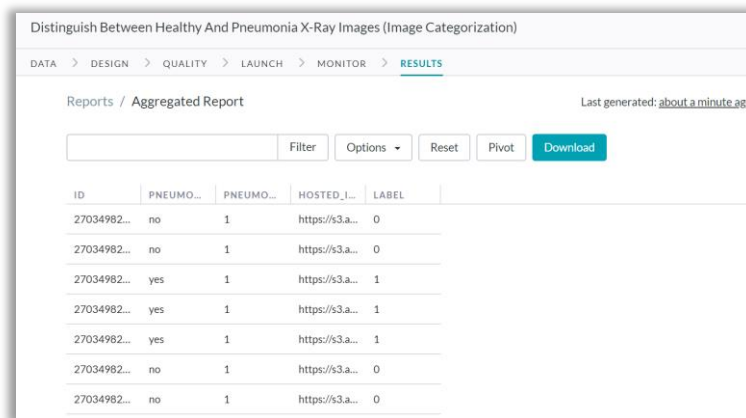


The resulting data (below) can be exported in CSV or JSON format, and is easy to interpret. This could be used for training a model, but since this is a test launch, the results would be scrutinized to see which are correct, and which are not (not completed here).

## Final Thoughts

This experiment stresses the need for testing job launches, and having clear design, instruction, and test questions. If a full job had been launched without testing, higher costs would have been incurred (we were 76% over our budget of 25\$). As such, it is incredibly important to treat this test as a soft-launch to eliminate bugs, and create a great design.

Next steps in this project is (1) to review the test questions, (2) review the resulting labeling for accuracy, and (3) redesign a new launch to seek improvement based on these findings.



The screenshot displays a web interface for a project titled "Distinguish Between Healthy And Pneumonia X-Ray Images (Image Categorization)". The navigation bar includes links for DATA, DESIGN, QUALITY, LAUNCH, MONITOR, and RESULTS (which is currently selected). Below the navigation bar, the page shows "Reports / Aggregated Report" with a timestamp "Last generated: about a minute ago". A toolbar contains a search input field, a "Filter" button, an "Options" dropdown menu, "Reset" and "Pivot" buttons, and a "Download" button. The main content area features a table with the following data:

ID	PNEUMO...	PNEUMO...	HOSTED_I...	LABEL
27034982...	no	1	https://s3.a...	0
27034982...	no	1	https://s3.a...	0
27034982...	yes	1	https://s3.a...	1
27034982...	yes	1	https://s3.a...	1
27034982...	yes	1	https://s3.a...	1
27034982...	no	1	https://s3.a...	0
27034982...	no	1	https://s3.a...	0

## Resources

- <https://success.figure-eight.com/hc/en-us>
- <https://success.figure-eight.com/hc/en-us/articles/204820349-How-to-Cancel-a-Job>
- <https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia>