

Question 1: Netflix Prize and Bell, Koren, and Volinsky

- a. Submissions to the Netflix Prize were judged on how well teams or individuals were able to boost the accuracy of Netflix's existing recommendation model, Cinematch, by at least 10%. Submissions were not judged if they did not at least meet this threshold of improvement.
- b. In the beginning, the most commonly used method for predicting movie ratings was SVD, or singular value decomposition, which utilized factoring to better identify the frequency of each rating value and weights them accordingly.
- c. Outlier movies such as *Napolean Dynamite*, where ratings were mostly either 5s or 1s, skewed the error rates immensley. To overcome this problem of "inertia" or anchoring around an observable rating, averaging based on previous ratings would give a "truer" rating. This would create a more relative rating for strong opinions.

Question 2: Collaborate problem solving: Project Euler

- a. ngrossfe: 1410712\_vpdI9GxnkssWORK91MKjjR7w6UN7jxdX
- b. Completed project #2, Even Fibonacci Numbers.

```
---  
title: "MACS PS6 Q2"  
author: "Neta Grossfeld"  
date: "11/18/2018"  
output: html_document  
---
```

```
```{r}  
fibonacci <- c(1, 2)  
while (max(fibonacci) < 4000001) {  
  fibonacci <- c(fibonacci, fibonacci[length(fibonacci) - 1] + fibonacci[length(fibonacci)])  
}  
  
(euler_answer <- sum(fibonacci[fibonacci%%2 == 0]))  
```
```

```
[1] 4613732
```

- c. I would aspire to achieve the One In A Hundred, One Percenter, and Chart Topper. I like these awards because of the comparisons to other people who are either also like me or not like me, in regards to my progress.

Question 3: Human computation projects on Amazon Mechanical Turk

- a. I've selected HIT "Click colored bolbs in an image" by Leonardo.

- b. The reward for this HIT is \$0.05, and MTurk receives 20% of the amounts paid to Workers by the Requestor.
- c. There are no qualifications, restrictions, or eligibility requirements for this HIT, but there are instructions with 4 bullet points.
- d. The allotted time for this task is 60 minutes. I think I might be able to do 6 per minute, so 360 per hour. This assumption implies an hourly rate of \$18.
- e. This HIT expires in 5 days and was created 2 days ago.
- f. The most this project would cost the Requestor if 1 million people participated is \$18,000,000 for the workers and \$3,600,000 to MTurk.

Question 4: Kaggle open calls

- a. The title of the open competition I chose is “Quora Insincere Questions Classification”, with the subheadline of “Detect toxic content to improve online conversations. The sponsor is Quora, a question-answer platform. Submissions will be evaluated using an F1 Score between the observed value and the predicted value, which is a measure of a test’s accuracy and averages precision and recall. Perfect precision and recall reflect an F1 score of 1, with the worst precision and recall scoring a 0. The predicted value for whether a question is insincere is 1, and is 0 for sincere. After closing of the competition, the submitted models will be run against a private testing set of data to determine the final scores of the competition. There are 3 winners possible, with the first place winner receiving \$12,000, the second place winner receiving \$8,000, and the third place winner receiving \$5,000. There are no honor code issues mentioned within this competition. The competition was launched 10 days ago as of 11/16/18. In 2 months (1/29/19), there is a Rules Acceptance Deadline or an entry deadline. The competition closes on 2/5/19. Submissions are required to be made via a Kernel output.
- b. Quora will use the winning submission answer to better remove toxic and divisive content in the form of insincere questions or statements that don’t look for helpful answers. In their words, Quora will use the winning submission to uphold their policy of “Be Nice, Be Respectful.”

Link: <https://www.kaggle.com/c/quora-insincere-questions-classification#description>