Question 1. Netflix Prize and Bell, Koren, and Volinsky (2010).

Part a.

The submissions to the Netflix Open Prize contest would be judged relative to the performance of Netflix's recommender system, Cinematch. The criterion function: the submission with more than 10% improvement in root mean squared error (RMSE) over Cinematch would win the $1 million prize. Any submission which would have less than 10% of improvement in RMSE over Cinematch would not be judged.

Part b.

At the beginning of the Netflix Prize contest, the most commonly used method for predicting ratings (stars) on movies was Nearest Neighbor method. In nearest neighbour method, you would categorise items, for a user, based on similarity. And to find the predicted rating for an item, one would find the weighted average ratings of the other similar items. Different methods like Pearson correlation, cosine similarity, and etcetera were used to calculate similarity scores.

Part c.

Hybirds of multiple models benefitted from the sets of predictions from different underlying models. For example, upon merging the prediction sets of two models which had similar RMSEs (Root Mean Square Error), the model which averaged out the two sets produced better predictions, overall. But, adding a method to a blend would be helpful if the model wasn't highly correlated with the other components[1].

Question 2. Collaborative Problem Solving: Project Euler

Part a.

I have registered on ProjectEuler.net with "nt546" as the username and my friend key is "1410880_nbpiYxiROBBuQw7NLyBoAd7fZtuv2v84".

Part b.

I solved problem 1 which was based on the concept of finding multiples of 3 and 5. The problem required finding the sum of all the multiples of 3 or 5 below 1000. The correct answer is 233168 for which i wrote the following code in python:

```
sum = 0
for i in range(1000):
   if i % 3 == 0 or i % 5 == 0:
      sum += i
print(sum)
```

The three problem solving awards which I find appealing are:

Baby Steps: It requires solving three problems.

The Journey Begins: It marks the completion of level 1 upon solving of twenty-five problems.

Decathlete: 10 consecutive problems are to be solved for achieving this award.

[1] Bell, R. M., Koren, Y., & Volinsky, C. (2010). All together now: A perspective on the netflix prize. *Chance*, *23*(1), 24-29.

The reason to choose these basic awards is my belief that to inculcate problem solving I need to overcome the beginning inertia. In the process of achieving these three awards, I would have solved sufficient number of problems to make myself comfortable with the concept of ProjectEuler.net and also get enough practice to aim for bigger awards and difficult problems.

Question 3. Human computation projects on Amazon Mechanical Turk (MTurk)

Part a.
"Categorize the purpose of information technology posts" is a MTurk HIT (Human Intelligence Task) which is neither an experiment nor a survey. "mturk-sandbox" is the requester of this task and it involves categorizing IT (Information Technology) related short text documents into four different categories:
IT Breakfix, IT Product Comparison, IT Product Research and Other.

Part b.
A payment of $0.02 will be made by the requester to the worker upon successful completion of the work within 120 minutes.

Part c.
The task had two qualifications requirement. The first requirement of the task requires the worker to be based out of Canada or United States. The other qualification requirement was a score of 100 in IT Topic Classification exam.

Part d.
120 minutes are allotted for completing this task but no details have been provided about the number of texts to be categorized. Suppose that it takes 2 hours for a person to complete this task then the hourly rate is $0.01

Part e.
The job was created on 11/16/2018 at 11:15 am and will expire at the same time on 11/19/2018.

Part f.
If 1 million people successfully participated in this task then they would be eligible for $0.02 each. Hence, the experiment would cost $ 20000 to the HIT creator.

Question 4. Kaggle open calls.

Part a.
I have registered on Kaggle with username nipunt.

Part b.
The title of the competition: "Quora Insincere Questions Classification".
Quora, a company which maintains a question and answers website is the sponsor of this open competition. It believes in sharing knowledge and allows people to ask questions and post answers to questions of other people's. The contest submissions will be evaluated using F1 score, a statistical measure of accuracy, between the predicted and the observed targets. There will be prizes for the top three submissions. The first prize is $12,000, the second prize is $8,000 and the third prize is for

$5,000. Both the entry deadline and team merger deadlines are January 29, 2019 (11:59 pm UTC). Any new participant should have accepted the competition rules prior to this date and no participant will be allowed to join or merge team after the team merger deadline. The final submission deadline is February 5, 2019. The submissions would be made directly from Kaggle Kernels. The submission file has to have a header: “qid, prediction”, followed with all the qid’s present in the test set and the predictions (0 or 1).

Part c.

Quora would do further research and development on the winning submission answer. After which they would use it on their platform to increase the effective sharing of knowledge through right set of questions and answers. It would help them in identifying and removing the insincere questions and maintaining the quality of discussions happening on their platform.