Individual Peer Evaluation Form

Your name: **Xin Tang**

Write the name of your classmate you are preparing this review for in the designated column. Using a scale of 1-4 (1=strongly disagree; 2=disagree; 3=agree; 4=strongly agree) answer each question. If you aren't able to answer the question based on what is posted in the discussion board, reach out to your classmate for more information via the discussion board. Total the numbers in each column. **Make sure to answer the questions on the 2nd page.**

Evaluation Criteria	Peer Name: **Krista Knuckey**
Has plan in place to complete course project.	4
Has found datasets/data sources to support project idea.	4
Has solidified project idea.	4
Has identified resources for project.	4
Topic is related to data science and demonstrates topics learned to date through program.	4
Risks and potential issues have been identified.	3
TOTALS	23

Feedback on Individual's project topic:

1. How clear is the classmate's project topic? What questions does their topic make you consider?

Krista's topic is very clear. She wishes to do a classification to predict the default risk of the personal loans, using logistic regression and decision tree, plus a random forest algorithm.
She also takes into consideration the dataset validation step and data cleaning step, which is also critical to ensure a successful modeling.
I feel that she has really honed their topic of interest and has nicely improved upon the original thought process. Good job!

Some questions that their topic makes me consider are:
- Among the 3 models Krista trys to use, which models will be best? If so, why?
- She mentioned the dataset is large and some of the columns may not have big impact, which method can be used to evaluate that?
- One risk of binary dataset is imbalance, since default loan is always a small portion of the total loan, how to handle that?

2. What risks or issues should your classmate consider while working on their project?

One thing I worry about is the imbalance of the dataset. The professor of my last course had pointed out that risk on my project for that class and suggested to get it treated. I would suggest looking at the default /no fault ratio.

Second, I feel a little nervous to remove columns will less than 50% of data, if it is 10%, I have no worry, but if it is 40%, I think Krista could use an alternative way, like using average, or median value to replace them.

3. Additional suggestions/comments that might be beneficial to your peer?

I really like Krista's approach to with fellow financial expert to evaluate the data and approach. I would suggest Krista seek advice from data analysis or data scientist (if there is any close to Krista) to ensure model building goes well. Another part I would like to see is the evaluation of the model, which is also an important part to get best model picked.