**Remark Data Community - Competency Test.**

Using tools of your choice complete the following tasks. Keep it simple. The goal of exercise is not to do the deepest EDA, build the best model possible, or spend lots of time building a robust solution. The goal is to demonstrate your ability to complete the end-to-end process in a reasonable timeframe and explain your thought process clearly.

1. Load data
   1. Load the data into the environment of your choice.
2. Clean data
   1. Clean the data so that it is more usable for subsequent steps. At least two columns should be transformed.
3. Feature engineering
   1. Transform at least one column into another column or columns, which have potential use for exploratory data analysis and/or modeling.
4. Exploratory Data Analysis (EDA)
   1. How complete is the data?
      1. Are there missing values?
   2. Create visualizations which explain the distributions of ‘category’, ‘actual\_price’, ‘discount\_price’, ‘discount’, ‘rating’, and ‘rating\_count’
      1. If you have decided to engineer new columns based on any of the above columns, you may use the columns you created instead.
   3. Are any of these variables correlated?
      1. How do you know?
5. Model
   1. Build a model to predict ‘rating’. In the interest of time, the accuracy does not need to be high.
      1. What modeling approach did you use and why?
   2. What is the accuracy of your model?
      1. Why did you choose this accuracy metric?
6. Understand and Explain
   1. What is the relative importance/significance of the features you used in your model?
      1. What importance/significance metric did you choose and why?
   2. Describe the relationship between the most important variable and the prediction.
      1. What method did you use to examine this relationship and why?
   3. Do you believe that the ‘rating’ variable is able to be predicted based on the data available?
      1. Why or why not?
   4. Suggest at least two ways to improve the model performance.
   5. In a few paragraphs or less, explain to a non-technical audience what you learned during this analysis. Keep your description as simple and easy to understand as possible.