Hi *candidate*,

Thanks for chatting with us about your interest in our Data Scientist opening. As discussed, we are asking select candidates to tackle the following data science problem. If you have any questions or concerns regarding this test, please email us. The test should take no more than a week; we should have indicated the expected return date in an email to you.

You will need to use to the claims.sample.csv file in order to complete the following exercises.

Please provide explanations that justify your methods and results, and explain any assumptions made about the data. Present your findings in a clear and organized document. Include your code separately so that a reviewer can easily recreate the results.

A medical claim is denoted by a claim number ('Claim.Number'). Each claim consists of one or more medical lines denoted by a claim line number ('Claim.Line.Number').

1. J-codes are procedure codes that start with the letter 'J'.

     A. Find the number of claim lines that have J-codes.

     B. How much was paid for J-codes to providers for 'in network' claims?

     C. What are the top five J-codes based on the payment to providers?

2. For the following exercises, determine the number of providers that were paid for at least one J-code.

Use the J-code claims for these providers to complete the following exercises.

A. Create a scatter plot that displays the number of unpaid claims (lines where the ‘Provider.Payment.Amount’ field is equal to zero) for each provider versus the number of paid claims.

    B. What insights can you suggest from the graph?

    C. Based on the graph, is the behavior of any of the providers concerning? Explain.

3. Consider all claim lines with a J-code.

     A. What percentage of J-code claim lines were unpaid?

     B. Create a model to predict when a J-code is unpaid. Explain why you choose the modeling approach.

     C. How accurate is your model at predicting unpaid claims?

D. What data attributes are predominately influencing the rate of non-payment?