**Data Science Code Exercise**

As a guideline, you should expect to spend around 2-6 hours to complete this exercise over a few days. The assignment does not have to be completed all at once.

This is an opportunity to showcase your unique skill set. You may choose to highlight any areas that you believe are unique strengths such as:

* + Advanced algorithmic modeling
  + Feature engineering
  + Data engineering and pipelining
  + Software engineering e.g. object-oriented design, unit testing, etc.
  + Data Analysis and Visualization
  + Front-end development

Do not feel obligated to conquer all the above bullet points; these are just a list of suggestions that you can focus on to showcase your skills as part of this exercise.

Python is the preferred primary language for this exercise. You may augment your solution with other languages, but the bulk of the code should be in Python.

Once completed, email your solution, deliverables, and answers to :

[michael.junokas@nutrien.com](mailto:michael.junokas@nutrien.com)

Feel free to reach out to [michael.junokas@nutrien.com](mailto:michael.junokas@nutrien.com) if you have any questions.

Scenario Description

Your job as a data scientist in this scenario is to analyze a set of seeds data, developing a ranking of the corn brands of seeds from the perspective of *yield* (i.e. a column in the dataset), justifying your ranking of those corn brands.

Data Supplied

You are given two files:

* seeds\_data.csv: The data file where each row represents an observation of a seed’s performance within a trial; the first row contains the header for the columns.
* seeds\_metafile.txt: a meta file providing information about each column

The Task

Prepare a 1-2 page report that details your analysis of the seeds data, your ranking of the corn brands, and your justification for those rankings.

Deliverables

The following deliverables should be submitted:

* A 1-2 page report that should strive to be readable at a business, non-data scientist level.
* The code that you wrote to solve the problem
* Answers to the questions below in a separate document, listing the question and your response

Questions

Please **briefly** answer the following questions. Try to be as concise as possible while still giving complete answers. You will have the opportunity to add more detail and explanation regarding your approach via a follow-up discussion.

1. How long did it take you to solve this problem?
2. What software languages and libraries did you use (if any) to solve the problem? Why?
3. What steps did you take (if any) to prepare the data for the project? Was it necessary?
4. What statistics, algorithms, or analysis did you apply? Were there any you considered but didn’t use? Why?
5. Explain how you developed your ranking of the corn brands.
6. Which features had the greatest impact on your ranking of the corn brands? Which features had the least impact? How did you identify which features to use?
7. Explain any additional work you did as part of this project.