**Data Science Analyst -- Take Home Assignment |** Impossible Foods, March 2021

The questions provided here pertain to color measurements taken by kitchen technicians at Impossible Foods on raw Impossible burger color. **More information about color data, such as what L, C, and H mean can be found** [**here**](https://sensing.konicaminolta.us/us/blog/understanding-the-cie-lch-color-space/)**.**

1. (**Coding exercise**) Colorimetry measurements have been taken by two burger color technicians on formed Impossible Burger patties. They are interested in validating the new technician (Mike) to see if he is evaluating color measurements in the same way as experienced tech, Jen. They want to know if Jen and Mike are statistically different in any measurements of color: L (lightness), C (chroma), and H (hue).

Please use the **Color\_Tech\_Data\_dirty.xlsx** dataset in answering **part a**.

* 1. Please describe (or in python or R if you prefer) how you would clean and munge the data provided in **Color\_Tech\_Data\_dirty.xlsx** into the form given in **Technician\_Color\_Data.xlsx.**

Please use the **Technician\_Color\_Data.xlsx** dataset to determine your answers for **parts b,c, and d**.

* 1. Perform a basic exploratory data analysis using R or python. Please create a few tables/graphs to help explain these data. Get the averages for L, C, and H and 95% confidence intervals (around the mean) and an interval that encompasses 95% of the measured values for each variable
  2. Determine if there are significant differences for the color ratings of each L, C, and H between the two technicians. Determine if there are any other significant factors leading to differences in the measurements
  3. Write a couple paragraphs about your findings and conclusions. Please inform the team about your findings and advice for Mike’s training.

1. (**Free Response only, NO CODING**) The color team regularly collects data as shown in **Color\_Measurements.xlsx**. They use this data for continuous monitoring of product and generate a lot of excel sheets that are identical in structure to this one.
   1. They want some kind of tool that can intake an excel file of this structure and output the graphs and tables (in any form you like), like the ones you created in part 1. What tools or software would you consider using and why? Preferably this tool would have a GUI of some sort where they could drag and drop their raw data and get back the tables and charts. Keep in mind that this team **does not know how to program** and thus would likely be **unable to run jupyter notebooks or R scripts** **easily** without some help.
   2. Now, they are asking for a database to mine their findings over the past few months. They would like all of their data compiled so they could mine it for historical findings over the past year or so. How would you help them do this? What tools would you explore? What kinds of questions would you have for the team?
   3. Let’s imagine you’ve set up a database for them in one form or another. They would like an interactive dashboard to be able to visualize their data. If you could use any tools to help them with this, which would you reach for and why? Do you have experience with these tools?

Thank you for your responses! I look forward to reading them. Please don’t spend too much time on this (2-3 hours max). Please [email](mailto:tyler.simons@impossiblefoods.com) me with questions.

-- Tyler