

# Data Science Challenge

In this challenge, we are looking to assess your Python coding ability and data science skills. We'll be working on [the following dataset, which includes measurements of breast cancer cells](#). The task is to predict for each cell, whether is it *malignant* or *benign*. Please follow the guidelines below.

Note: Features are computed from a digitized image of a fine needle aspirate (FNA) of a breast mass. They describe characteristics of the cell nuclei present in the image.

---

## Part 1: Modeling Challenge

### Python Coding and Data Set

- Load in the data file and header file provided
  - The dataframe does not currently have a header, load in the header file and attach it to the dataframe
- Comment on any steps you might take to evaluate or transform the dataset.
- Compute the mean and median smoothness and compactness for benign and malignant tumors - do they differ? Explain how you would identify this.
- Write a function to generate bootstrap samples of the data.

### Exploratory Analysis

- Identify 2-3 variables that are predictive of a malignant tumor.
  - Display the relationship visually and write 1-2 sentences explaining the relationship.

### Modeling

- Build a model to predict the malignant tumors.
  - Use at least two classification techniques; compare and contrast the advantages and disadvantages of each.
  - Identify how you would control for overfitting in each classification technique.
  - Evaluate the performance of each model.

- In each model, identify the most important predictive variables and explain how you identified them.

## **Explanation**

- To Technical Audiences
    - Explain the limitations of your analysis and identify possible further steps you could take.
  - To Non-Technical Audiences
    - Write a short summary of your analysis, explaining how your model works and how it performs.
    - Briefly explain the factors that contributed to malignant vs benign tumor identification.
- 

## **Materials**

[Click here for the data file and header file provided for this task.](#)