**## Project Outcomes**

- Supervised Learning: use supervised learning techniques to build a machine learning model that can predict whether a patient has diabetes or not, based on certain diagnostic measurements. The project involves three main parts: exploratory data analysis, preprocessing and feature engineering, and training a machine learning model.

- Unsupervised Learning: perform unsupervised learning techniques on a wholesale data dataset. The project involves four main parts: exploratory data analysis and pre-processing, KMeans clustering, hierarchical clustering, and PCA.

**### Duration:**

Approximately 3 hours and 20 minutes, plus a 5-minute presentation.

**### Project Description:**

In these two projects, we will apply supervised and unsupervised learning techniques to a real-world data set and use data visualization tools to communicate the insights gained from the analysis.

The data set for this project is the "Diabetes" dataset from the National Institute of Diabetes and Digestive and Kidney Diseases and the "Wholesale Data" dataset containing information about various products sold by a grocery store.

The project will involve the following tasks:

* Exploratory data analysis and pre-processing: We will import and clean the data sets, analyze and visualize the relationships between the different variables, handle missing values and outliers, and perform feature engineering as needed.
* Supervised learning: We will use the Diabetes dataset to build a machine learning model that can predict whether a patient has diabetes or not, using appropriate evaluation metrics such as accuracy, precision, recall, F1-score, and ROC-AUC. We will select at least two models, including one ensemble model, and compare their performance.
* Unsupervised learning: We will use the Wholesale Data dataset to perform k-means clustering, hierarchical clustering, and principal component analysis (PCA) to identify patterns and group similar data points together. We will determine the optimal number of clusters and communicate the insights gained through data visualization using Tableau.

The ultimate goal of the project is to gain insights from the data sets and communicate these insights to stakeholders using appropriate visualizations and metrics to make informed decisions based on the business questions asked."

> **#### Instruction**

> The instructions for these two projects can be found in their **JUPYTER notebooks (Supervised learning - Project.ipynb and Unsupervised Learning - Project.ipynb)**.

> Please also complete the prompts in the README file available in [this repo](https://github.com/lighthouse-labs/Final-Project-Tableau).

**### Submission Guidelines:**

For this project, you will need to submit a link to your github repo that contains all of the files listed. The files submitted will be evaluated.

Your repo must include the following:

- Include one PDF file for both projects with your presentation. It should be named \*MachineLearningPresentationFile\_LastNameFirstName.PDF\*.

- The presentation should include the visualizations and key takeaways (Such as models and metrics) from your investigation (follow the Presentation Guidelines section on this page for further information).

**### Presentation Guidelines:**

~~- Spend 1 min on project flow structure. What were the steps in your project?~~

~~- Spend 1-2 minutes showing your results. Make sure to highlight:~~

~~- For Option 1, share the visualizations that you produced in response to the questions.~~

~~- For Option 2, please share what your main data question was, what you wanted to answer, the dataset you selected, and then explain some of the features of the dataset as well.~~

~~- Explain the biggest challenges in 1 min.~~

~~- What would you do if you had a bit more time?~~

~~- In your presentation, make sure to explain the rationale for the visualizations you selected for either Option 1 or 2.~~

**### Evaluation Guidelines:**

Here are some things you’ll need to keep in mind for this evaluated project:

- Familiarize yourself with the Eval Rubric tab so you can read about the competencies you will be evaluated on for this particular project and review what the different levels of each competency require.

- If you receive Unsatisfactory for any competency, your project will be rejected. If this happens, you will need to review the feedback provided, make changes to your project based on that, and resubmit your updated project within 48 hours in order to get it accepted and stay on track. This is not a bad thing! Having to resubmit is an opportunity for you to improve.

- Please ensure that you submit your project immediately following your presentation to help ensure you get feedback as soon as possible.

**### Rubric:**

Provide link to the rubric that will be used to evaluate the submitted documents.