Team: Bella, Kim, Lilo, Nick, William

**User Story 1:**

The scientists working in an R&D lab on the bench scale, for example, the senior scientist in a national lab or the severely underpaid graduate student in a university lab, want to get quick understanding and visualization of the spectra data. The user, who may be unfamiliar with programming languages, will upload their spectra data on a user-friendly interface. The user may also want to know the predicted shape of the spectrum, so both the predicted spectra and features that influence the spectra will be *beautifully* visualized on the UI.

**User Story 2:**

The “Know your spectrum” developers want to create a machine learning model that can predict the shape of the spectrum. The developers have experience with python and want to train an unsupervised machine learning algorithm to learn the features that will affect the shape of the peak, the position of the peak and the intensity of the peak in the spectra. The developers also want to include an option to add new data into the dataset as more data is collected by scientists.

**Use Case 1: Train VAE Using Experimental Data**

**Description:** The user, a developer with experience in python, wants to train a machine learning model (MLM) to predict spectrum.

**Inputs:** The user provides the model with a spectra dataset

**Outputs:** An unsupervised machine learning algorithm that can predict the shape of the spectra and the features that influence the spectra.

**Components:** Tangent PCA will get the nonlinear relationship between features and spectra shape. An unsupervised machine learning algorithm (i.e. VAE) will be trained with spectra dataset.

**Use Case 2: Add New Data Set**

**Description:** The user, the developer or scientist, wants to provide new spectra data into the training set.

**Inputs:** The user uploads a new dataset (file type TBD) formatted in a specific way with specific headers (Details are TBD)

**Outputs:** The program will add the new data to the existing dataset. A new machine learning model that has been trained with the new dataset.

**Components:** The program will check the data to see if the name exists within the set, and categorize the data accordingly. If it does not exist, it appends the data into the existing dataset. The model will then be retrained to include the new data.