Context : Glutamate is released at individual boutons in a frequency dependent manner. At 20 Hz, facilitation is usually observed (i.e. more glutamate is released during train stim), however we observed a high heterogeneity of release profiles between synaptic boutons. The goal of the project is to assess whether we can classify this diversity and find subgroups of boutons having similar behavior

Use first this file : GluSnFR\_avg\_variables\_allCa\_filtered\_3sigma.xlsx

-plot PPR profiles (1 to 10th stimulation) for all the synaptic boutons

-create a new panda dataframe with AMP1, AMP2, all PPR profiles, %failure 1 and %failure 2

-Using scikit learn

-Normalize data (function standardsccaler)

-apply PCA (you can play with the number of components, but in this first part you will reduce the dimensionality to the first 2 components. Why ?)

-add PC1 and PC2 in your pandas dataframe

-apply Kmeans algorithm on the transformed data (first use 5 clusters, but you will have to play with the number of cluster later on to make sure this is a good choice)

-Extract the most relevant varialbles : first using a polar plot of the pca\_component and second by using pearson correlation between PCs and features from the original dataset

-Add a new column in your pandas dataframe with cluster labels, so you can select by cluster

-calculate the Silhouette score for each bouton and plot them by cluster

-Now that you have clusters, sort the PPR profile by clusters, then averaged by cluster and plot them on the same graph.

-Study the diversity of the different features by clusters (AMP1, AMP2, %failure, PPRs) using boxplot for example. How relevant is the clusterization ?

-Relaunch the workflow with different parameters (use loops) and score the different runs