Primary Goal:

Find the optimal number of clusters to minimize variance

within groups and maximize variance between groups

Secondary Goal:

Run a classification regression model to test for predictability

of clusters.