Our team is working with a breast cancer classification dataset, classifying tumors into malignant or benign along with 10 different characteristics/parameters (radius, texture, area, etc.). The y variable is the Diagnosis (M or B) while the x variables are the 10 different characteristics provided in the dataset. Our goal is to run multiple analyses on the dataset to explore the relationship between different variables and the diagnosis of the tumor.

PCA: We chose to drop this method as it did not fit our dataset.

kNN:

Naive Bayes:

Decision Trees:

Logistic Regression