Josh Thomason
Model Metrics: ROC Score is the main metric we are looking at.

- **Auc/Roc**: Best metric for imbalanced categorical data.
    - **Class 0**: 99.83% of data points
    - **Class 1:** 0.17% of data points

- Added **F1 score** because it is a good metric to see both Precision and Recall.

1st X_train, y_train set: Optimal parameters through GridSearchCV():

Model: Trained with Scaled training data.	**Recall**	**Precision**	**F1**	**Auc/Roc**
KNN: **(n_neighbors = 1)**	0.81	0.94	0.87	0.91
Logistic Regression: **(C = 0.01)**	0.62	0.88	0.72	0.80
Random Forest: **(n_estimators = 256, min_samples_split = 4, min_samples_leaf = 2, max_depth = 7)**	0.79	0.93	0.85	0.895

UnderSampling (RandomUnderSampler), Optimal parameters through GridSearchCV():

Model: Trained with resampled Data	**F1**	**Auc/Roc**
KNN: **(n_neighbors = .33)**	0.92	Train: 0.984 Test: 0.958
Logistic Regression: **(C = 0.01)**	0.93	Train: 0.986 Test: 0.975

Random Forest: **(n_estimators = 65,** **min_samples_split = 40,** **min_samples_leaf = 20,** **max_features = 20,** **max_depth = 50,** **bootstrap = True)**	0.93	Train: 0.987 Test: 0.969

OverSampling (SMOTE): Only used LogisticRegression(), others were too computationally expensive:
Used GridSearch(), for optimal parameters.

Model: **Trained with SMOTE resampled Data.**	**F1**	**Auc/Roc**
Logistic Regression: **(C = 1000)**	0.95	Train: 0.991 Test: 0.978