

# Final Project

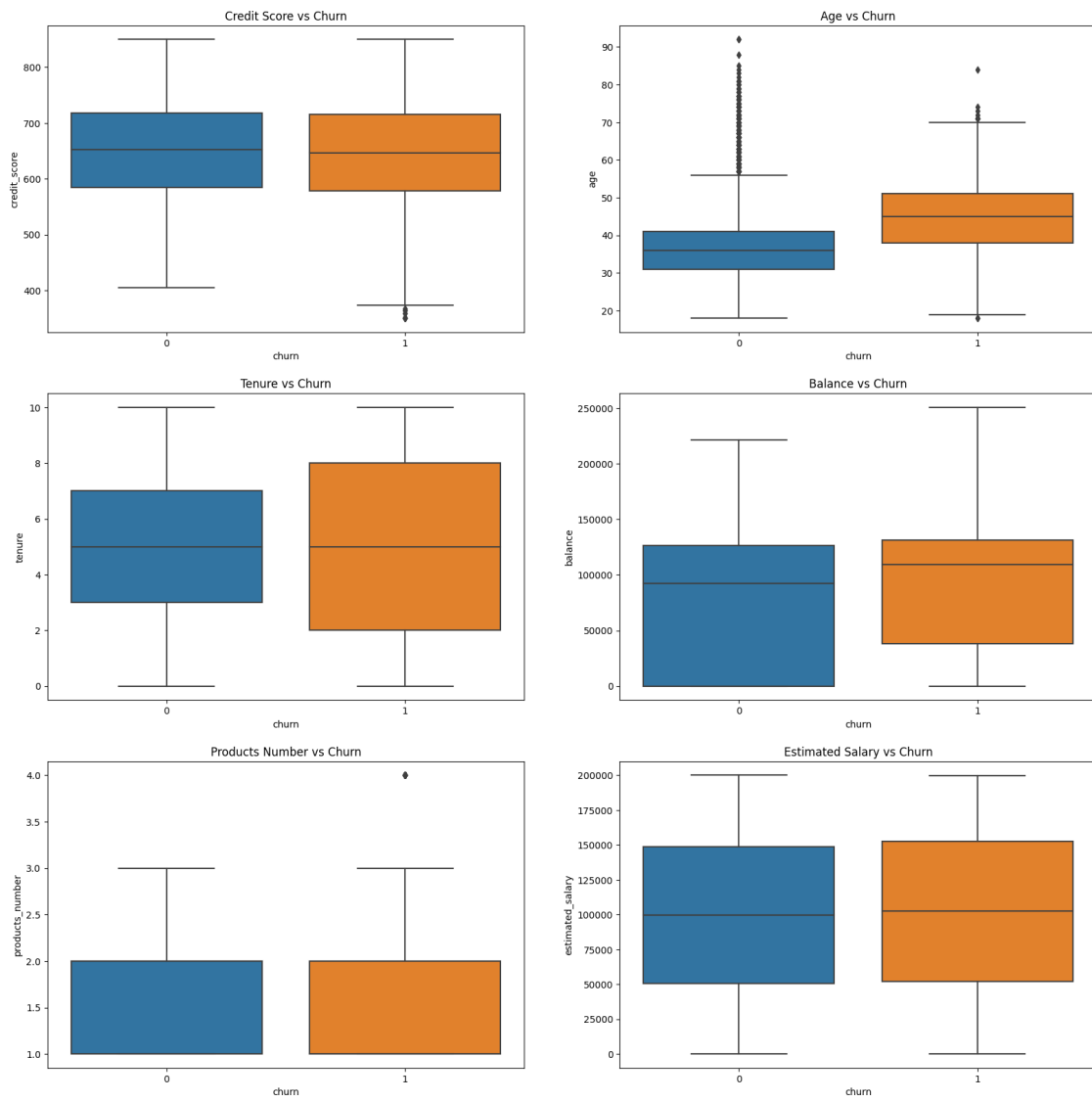
November 30, 2022

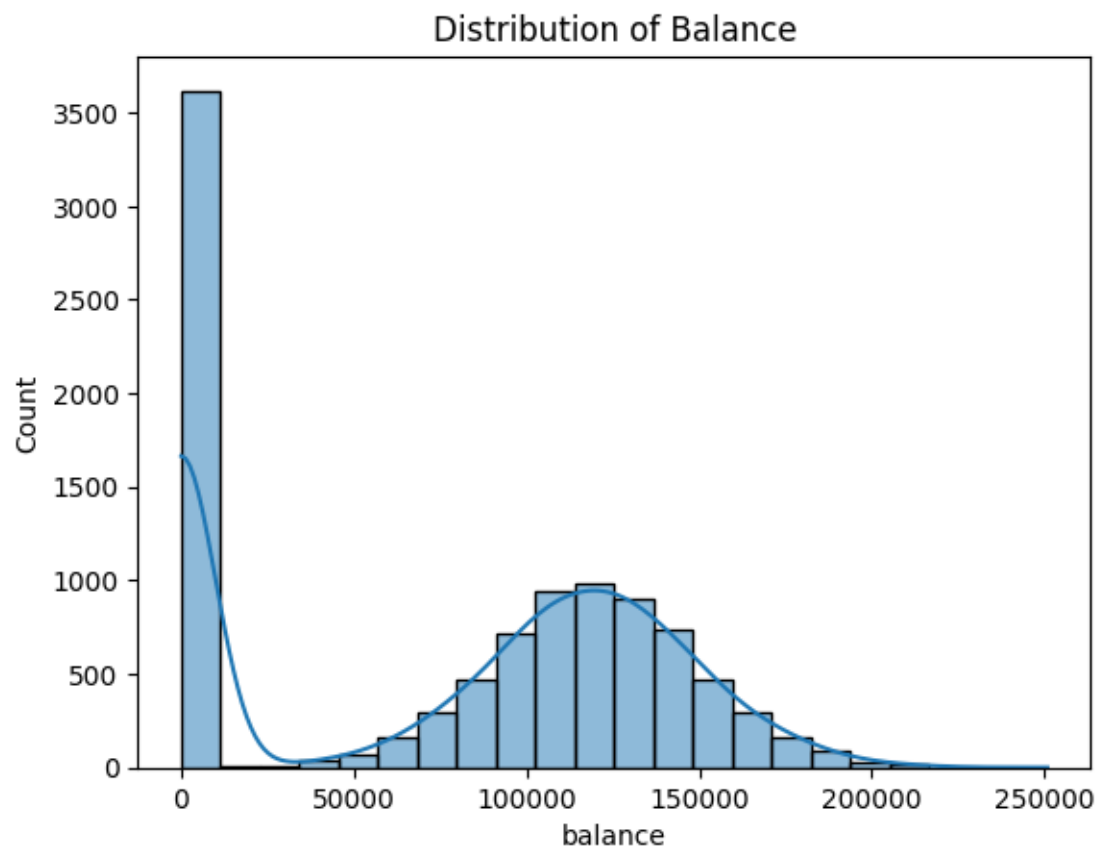
**Team Member Names: Madeline Witters**

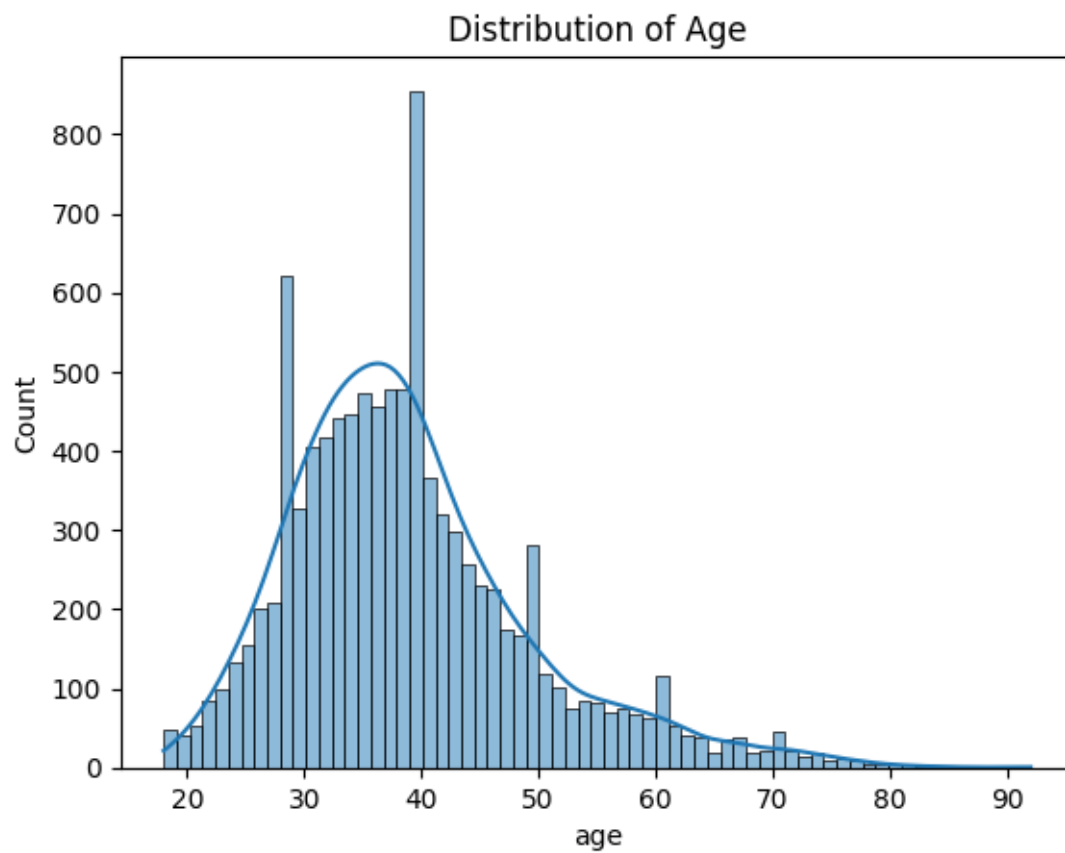
**Project Title: Predicting Customer Churn and Identifying Attributes of At-Risk Customers**

**Exploratory Data Analysis**

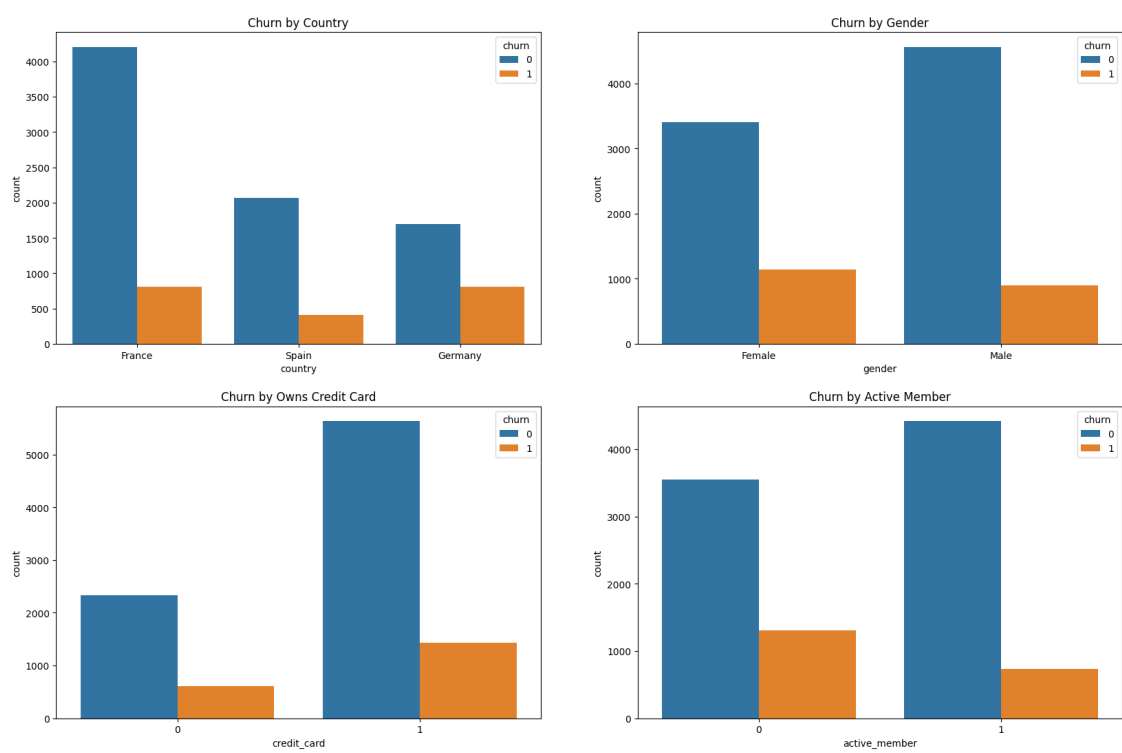
## Boxplots of Numeric Dependent Variables



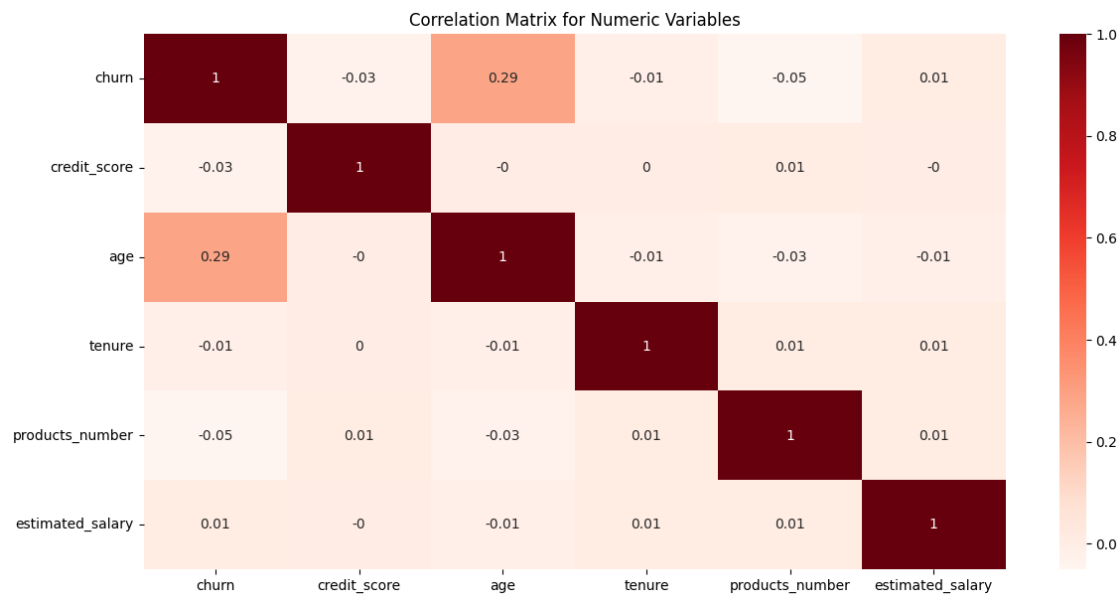
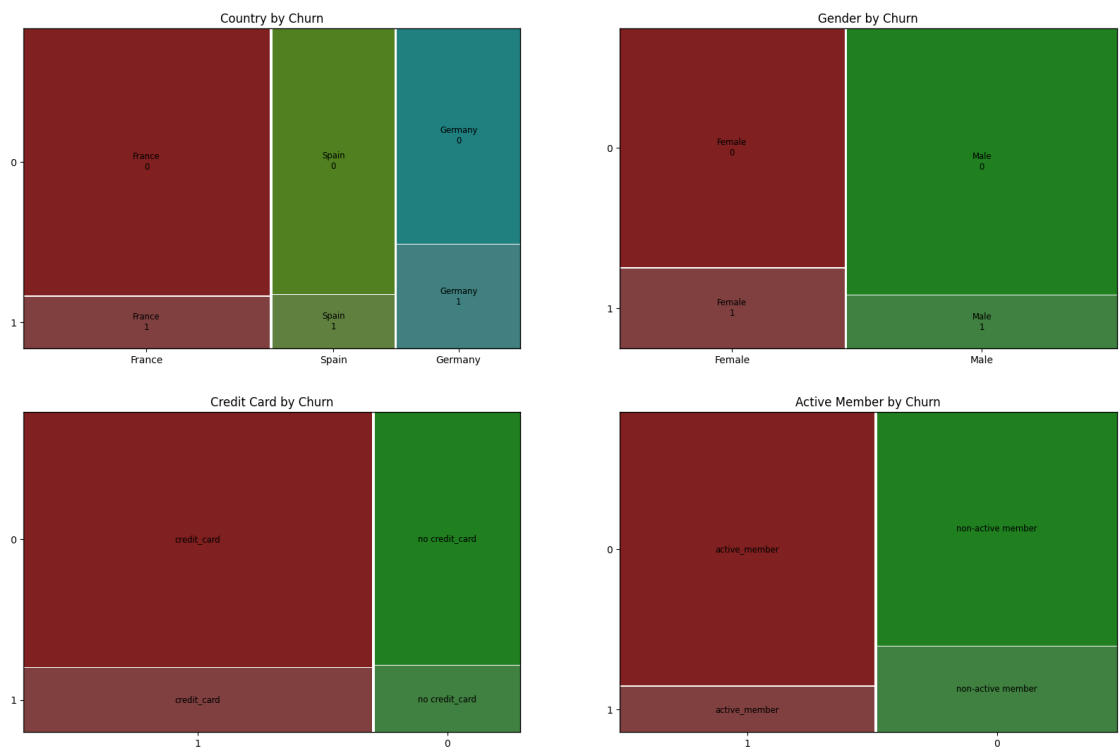




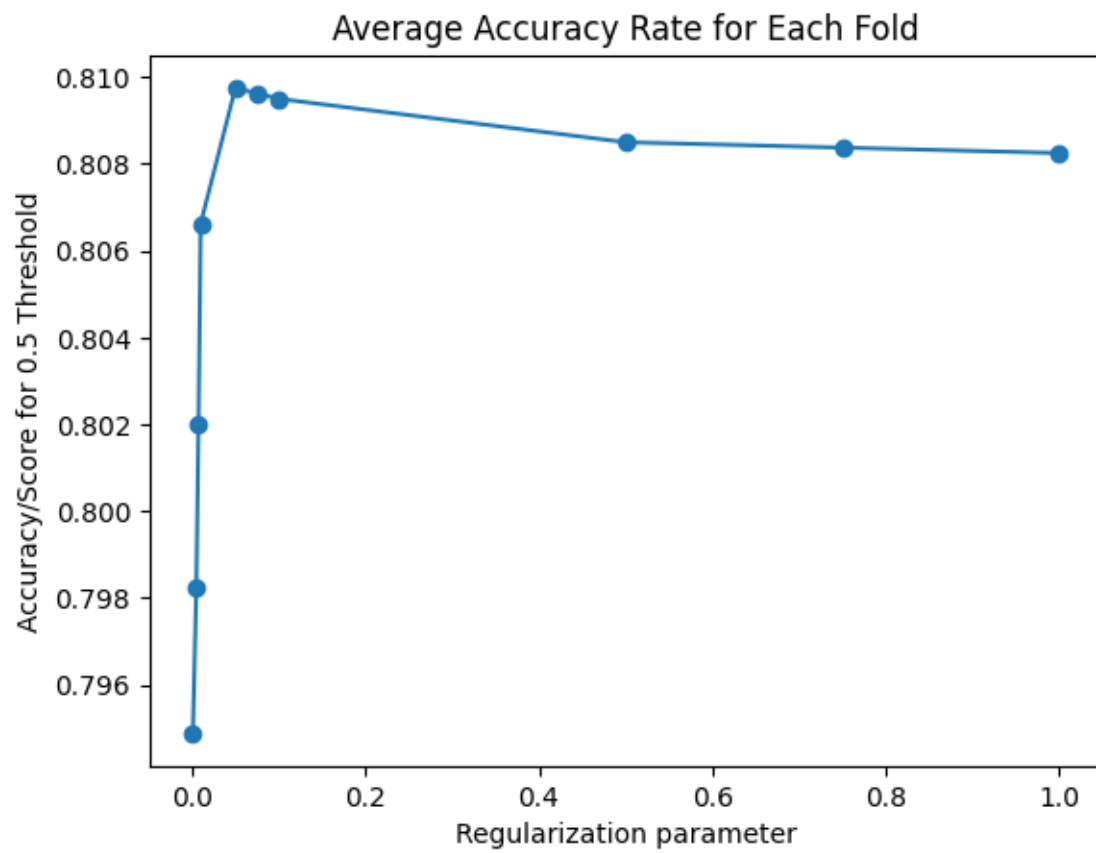
Bar Charts for Categorical Dependent Variables

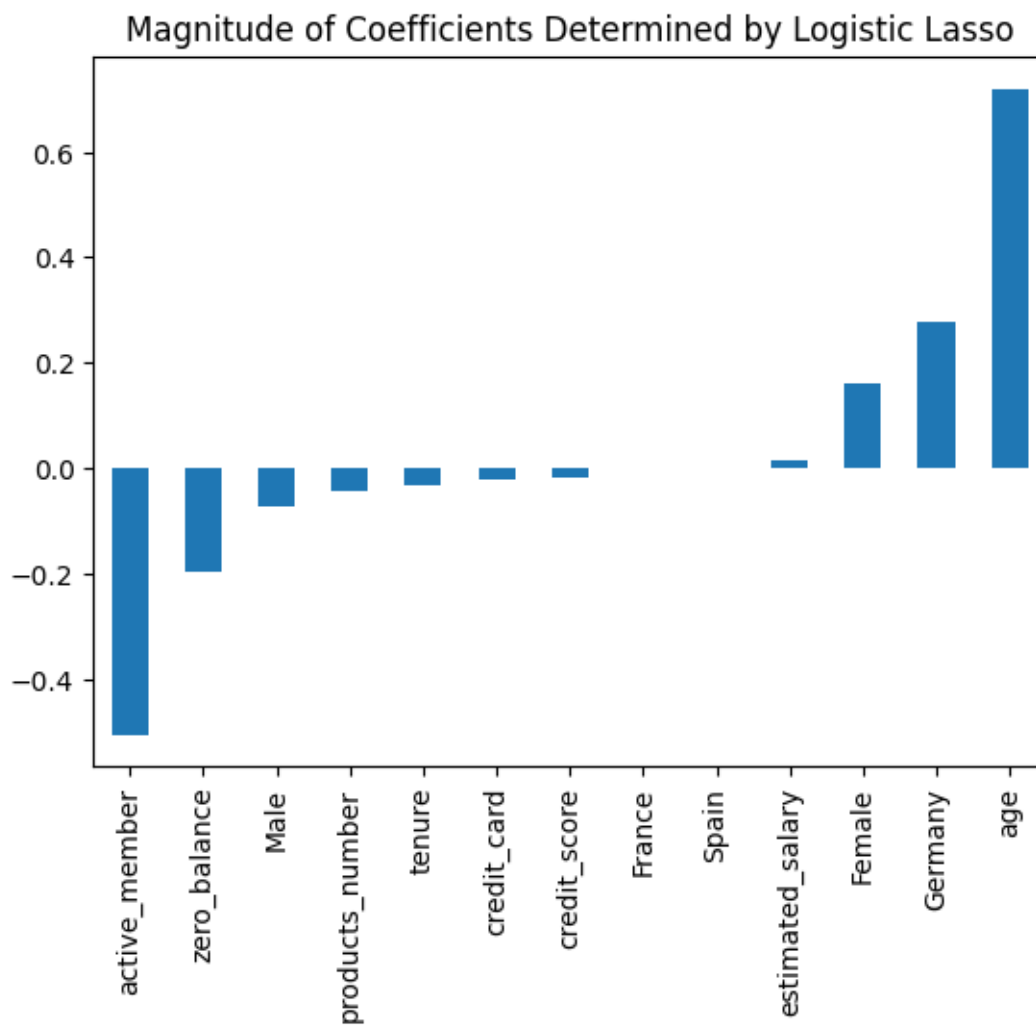


Mosaic Plots for Categorical Dependent Variables



## Variable Selection

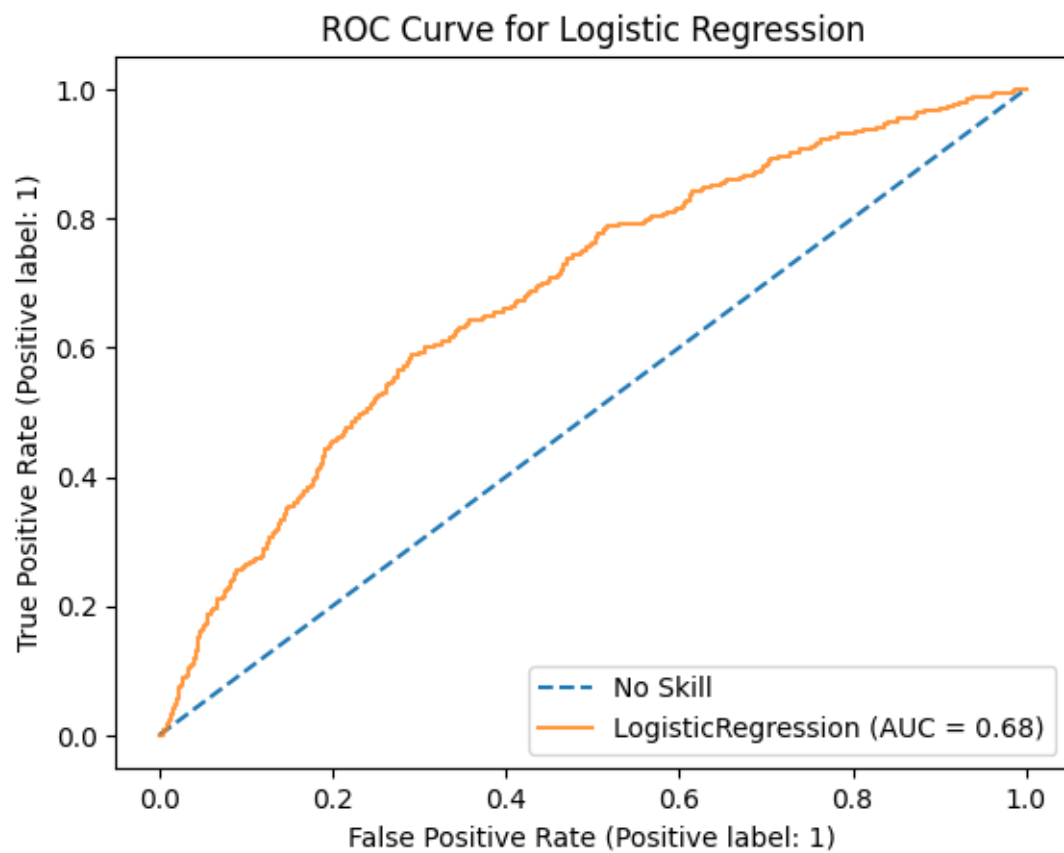


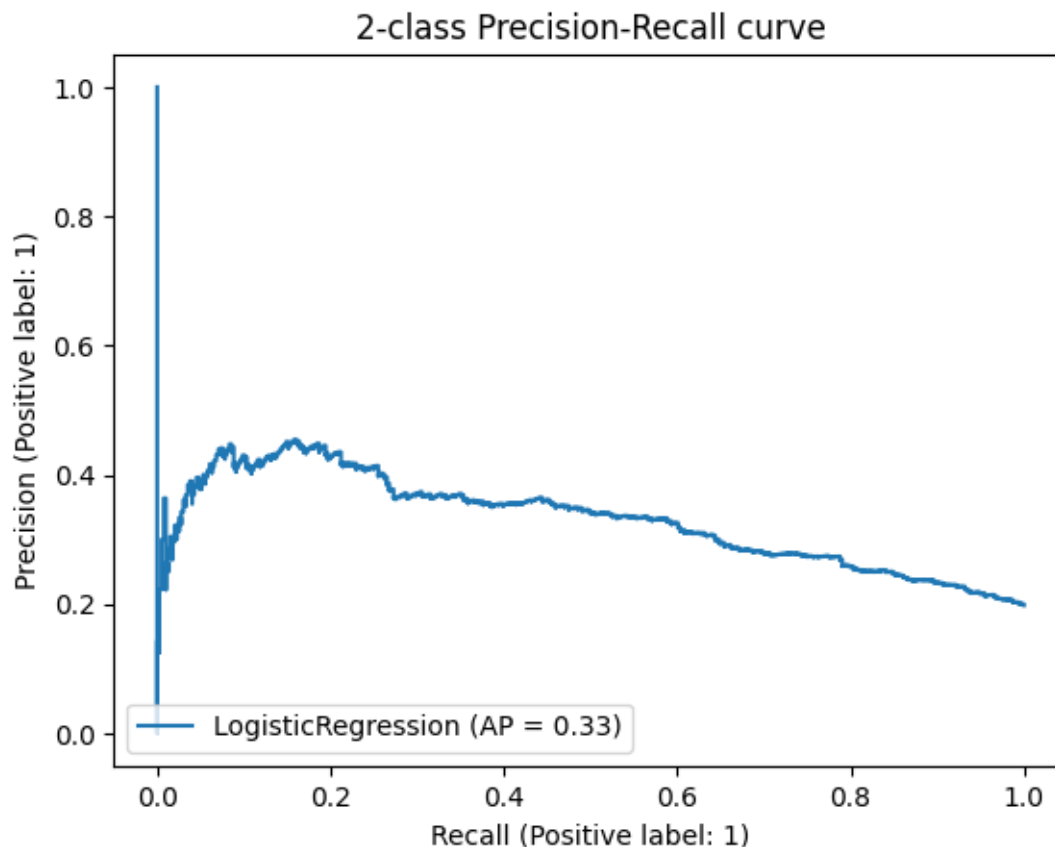


## Modeling

*Logistic Regression*







\*\*\*\*\* Logistic Regression model where threshold = 0.1 \*\*\*\*\*

Accuracy/Score is 0.291

Confusion Matrix:

[[ 200 1404]

[ 14 382]]

	precision	recall	f1-score	support
0	0.93	0.12	0.22	1604
1	0.21	0.96	0.35	396
accuracy			0.29	2000
macro avg	0.57	0.54	0.29	2000
weighted avg	0.79	0.29	0.25	2000

\*\*\*\*\* Logistic Regression model where threshold = 0.25 \*\*\*\*\*

Accuracy/Score is 0.7045

Confusion Matrix:

[[1203 401]

```

[ 190  206]]
      precision    recall  f1-score   support

     0       0.86      0.75      0.80      1604
     1       0.34      0.52      0.41       396

 accuracy      0.70      2000
 macro avg      0.60      0.64      0.61      2000
weighted avg      0.76      0.70      0.73      2000

```

\*\*\*\*\* Logistic Regression model where threshold = 0.3 \*\*\*\*\*

Accuracy/Score is 0.7515

Confusion Matrix:

```
[[1363  241]
```

```
[ 256  140]]
```

```

      precision    recall  f1-score   support

     0       0.84      0.85      0.85      1604
     1       0.37      0.35      0.36       396

 accuracy      0.75      2000
 macro avg      0.60      0.60      0.60      2000
weighted avg      0.75      0.75      0.75      2000

```

\*\*\*\*\* Logistic Regression model where threshold = 0.5 \*\*\*\*\*

Accuracy/Score is 0.7965

Confusion Matrix:

```
[[1577   27]
```

```
[ 380   16]]
```

```

      precision    recall  f1-score   support

     0       0.81      0.98      0.89      1604
     1       0.37      0.04      0.07       396

 accuracy      0.80      2000
 macro avg      0.59      0.51      0.48      2000
weighted avg      0.72      0.80      0.72      2000

```

\*\*\*\*\* Logistic Regression model where threshold = 0.6 \*\*\*\*\*

Accuracy/Score is 0.8005

Confusion Matrix:

```
[[1597    7]
```

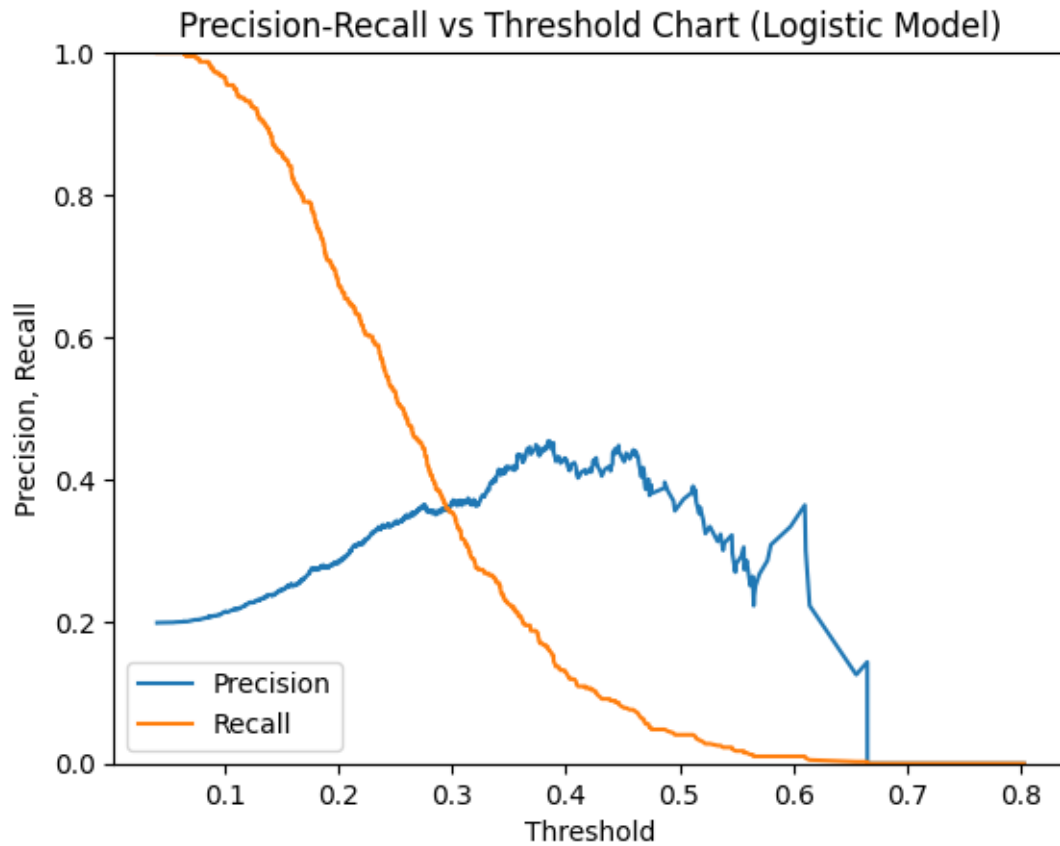
```
[ 392    4]]
```

```

      precision    recall  f1-score   support

```

	0	0.80	1.00	0.89	1604
	1	0.36	0.01	0.02	396
accuracy				0.80	2000
macro avg		0.58	0.50	0.45	2000
weighted avg		0.72	0.80	0.72	2000



### Random Forest

Random Forest Results

Accuracy/Score is 0.866

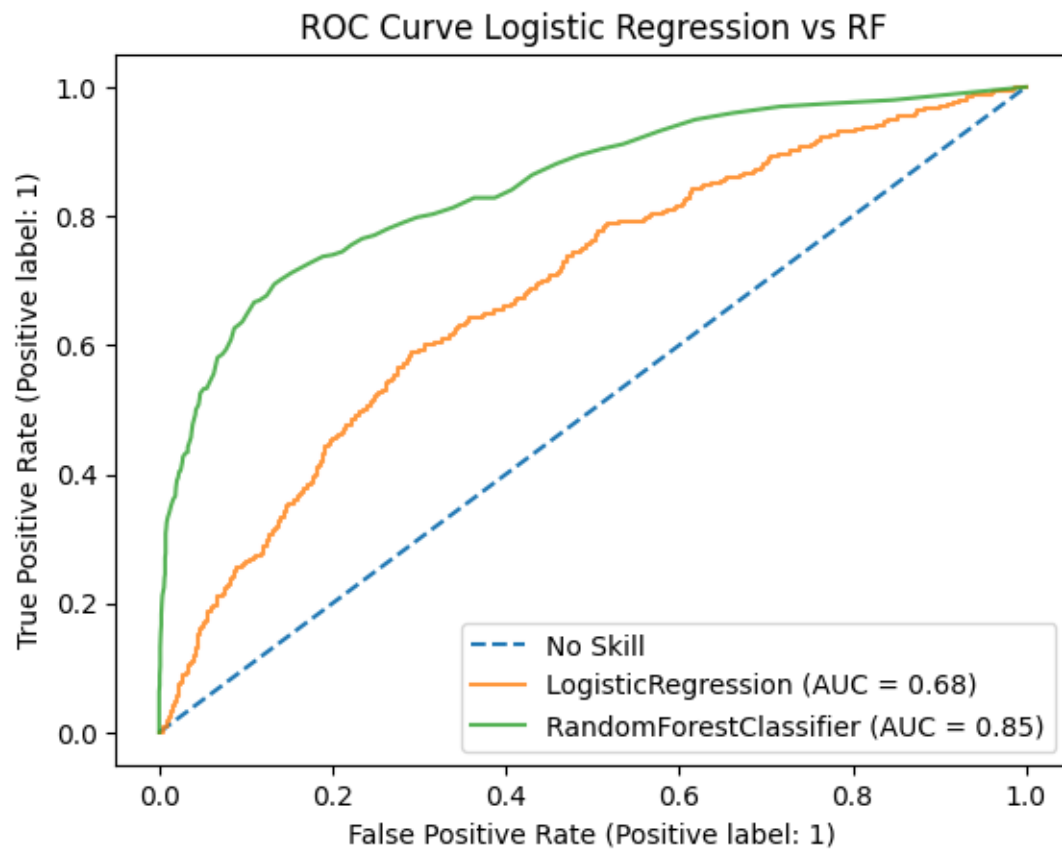
Confusion Matrix:

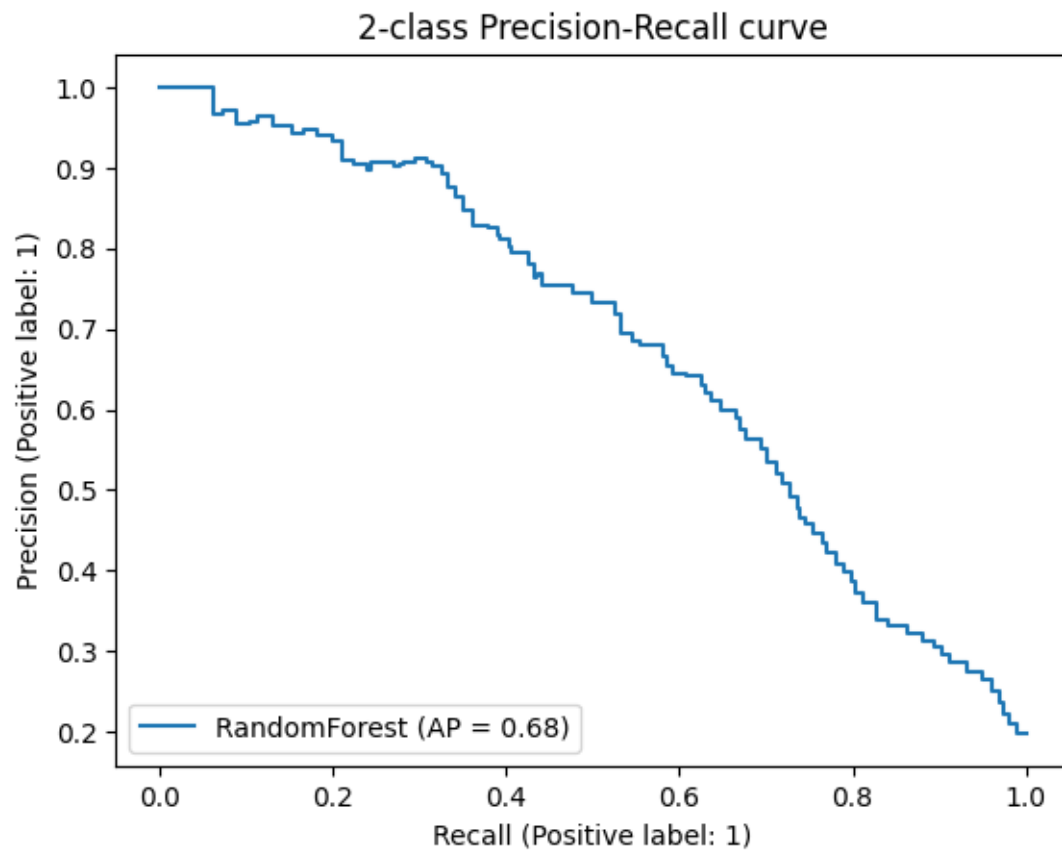
```
[[1537  67]
```

```
 [ 201 195]]
```

	precision	recall	f1-score	support
0	0.88	0.96	0.92	1604
1	0.74	0.49	0.59	396
accuracy			0.87	2000

macro avg	0.81	0.73	0.76	2000
weighted avg	0.86	0.87	0.86	2000

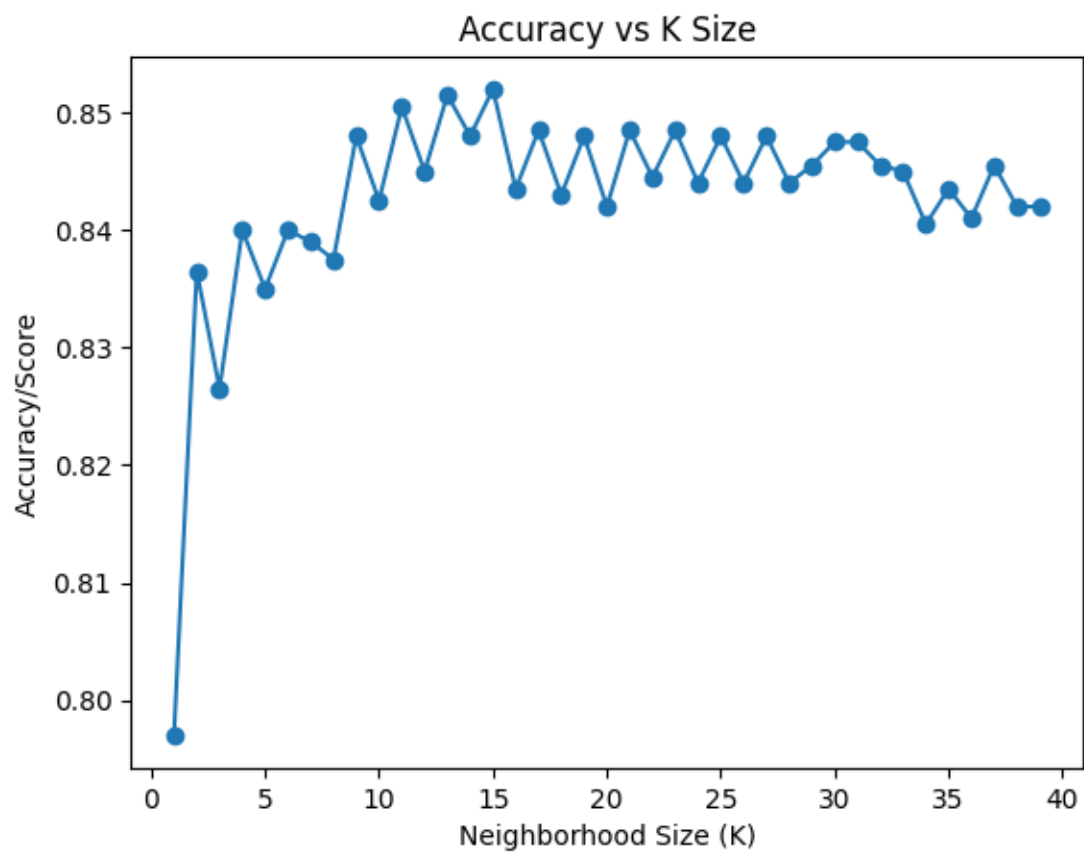




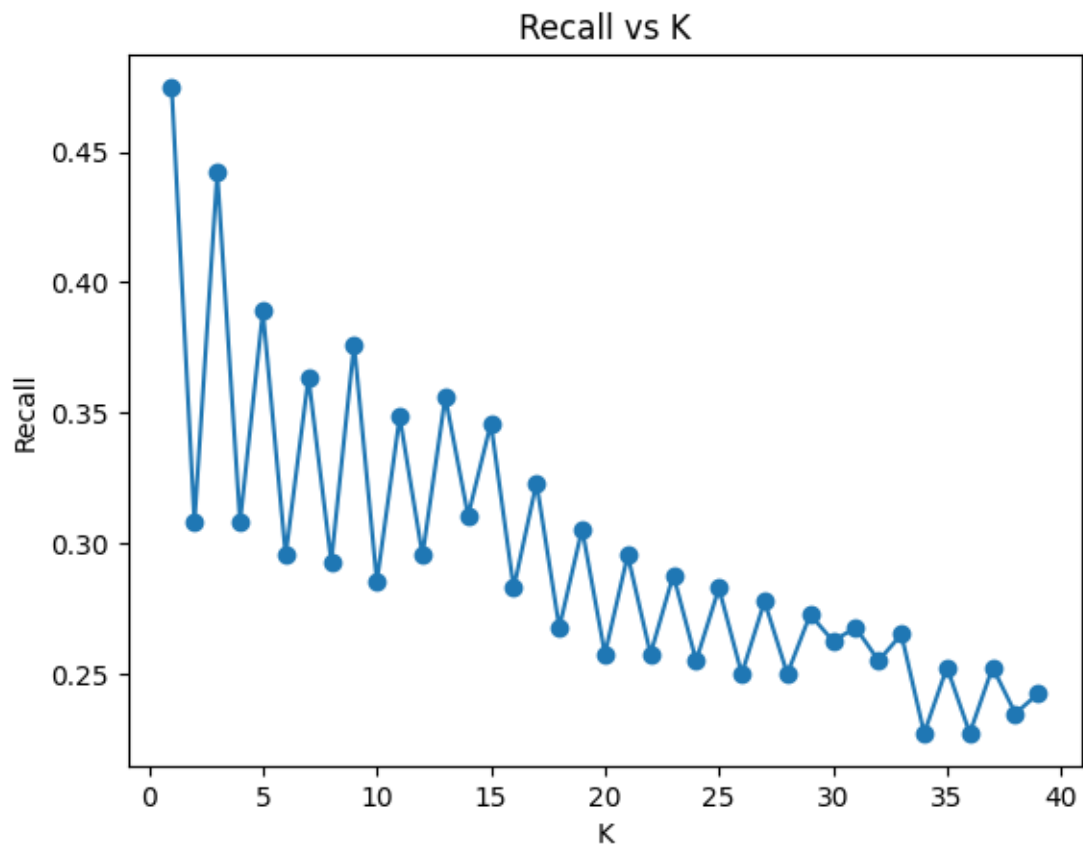
*Refitting of Models*

*KNN Classifier*

[99]: Text(0, 0.5, 'Accuracy/Score')



[100]: Text(0, 0.5, 'Recall')



KNN Results when k=15  
 Accuracy/Score is 0.852  
 [[1567 37]  
 [ 259 137]]

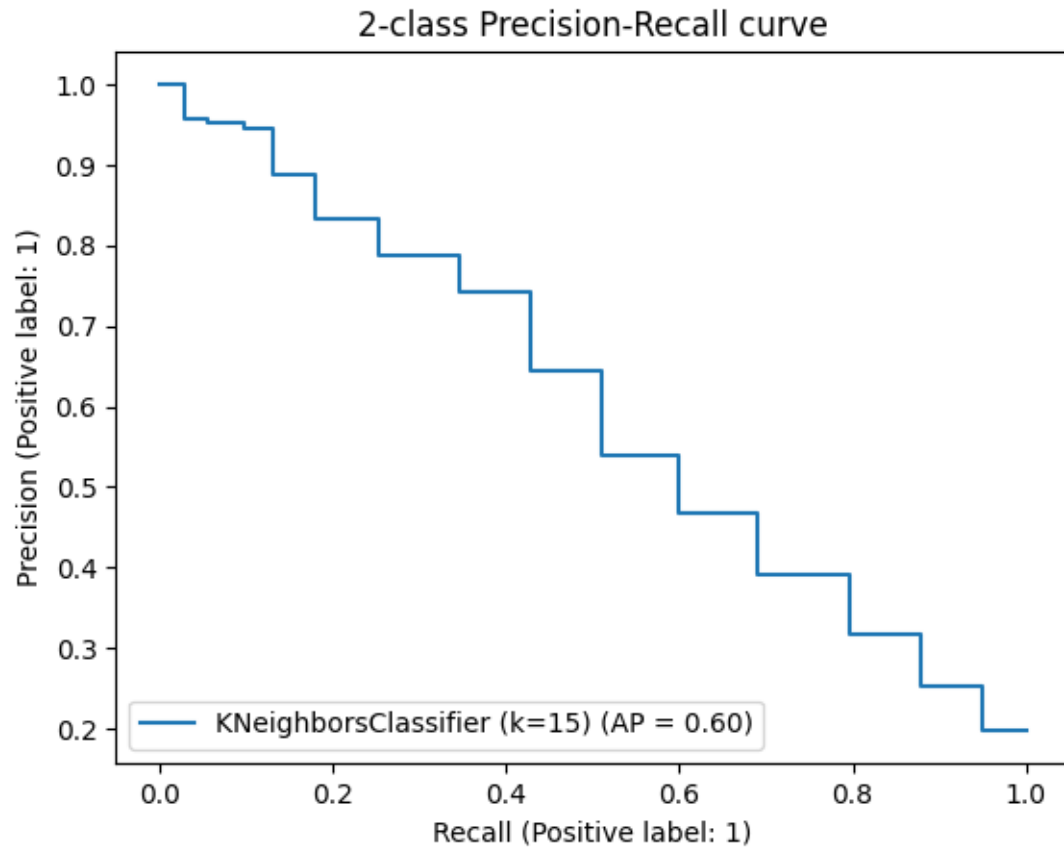
	precision	recall	f1-score	support
0	0.86	0.98	0.91	1604
1	0.79	0.35	0.48	396
accuracy			0.85	2000
macro avg	0.82	0.66	0.70	2000
weighted avg	0.84	0.85	0.83	2000

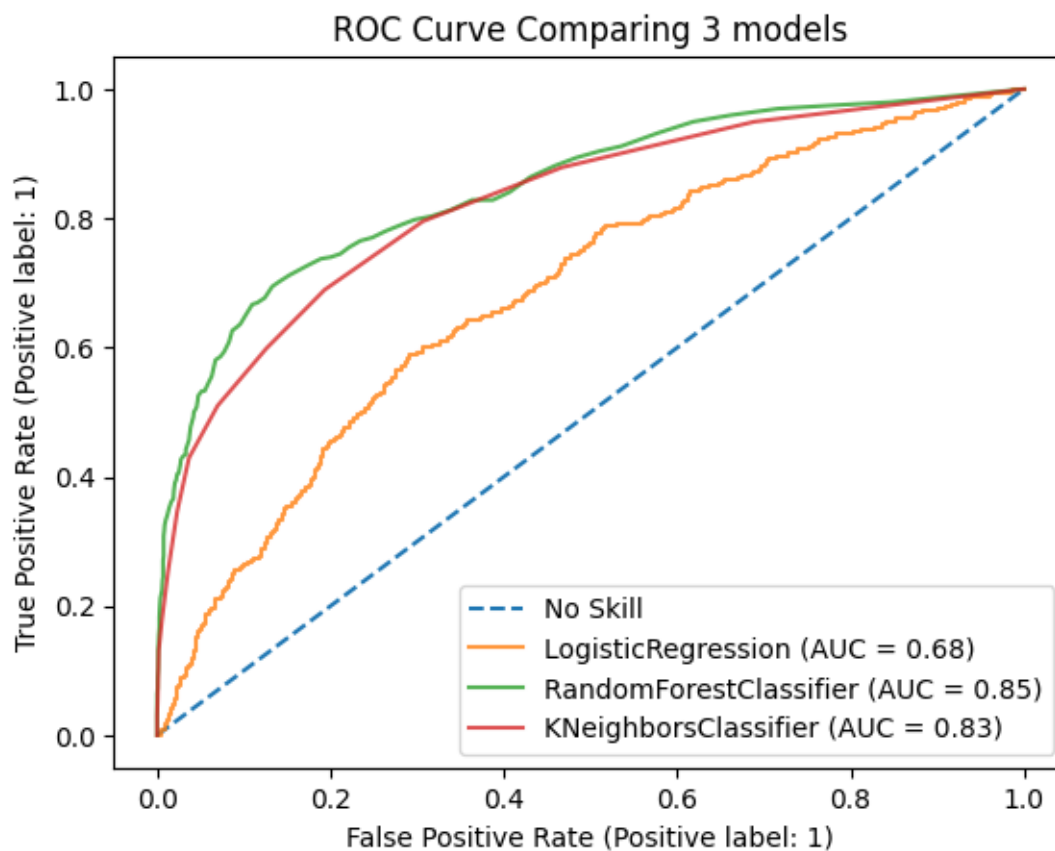
KNN Results when k=1  
 Accuracy/Score is 0.797  
 [[1406 198]  
 [ 208 188]]

	precision	recall	f1-score	support
0	0.87	0.88	0.87	1604



1	0.49	0.47	0.48	396
accuracy			0.80	2000
macro avg	0.68	0.68	0.68	2000
weighted avg	0.80	0.80	0.80	2000





### *Support Vector Machine*

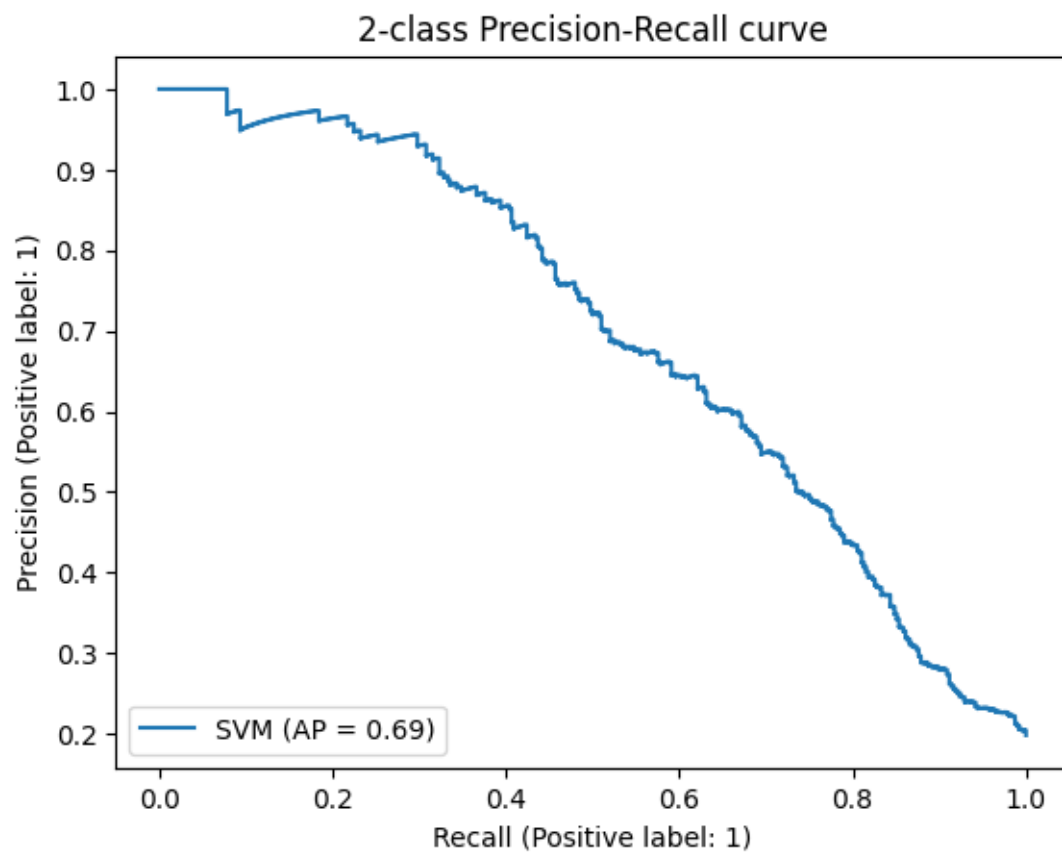
#### SVM Results

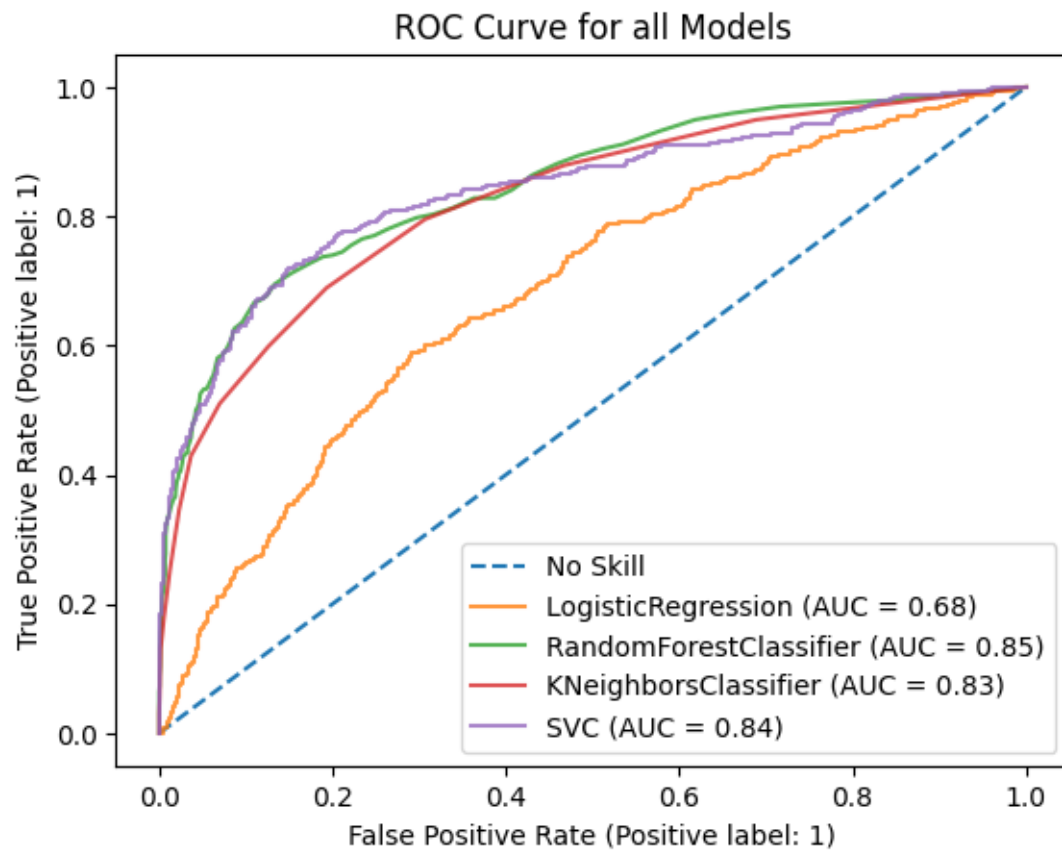
Accuracy/Score is 0.869

[[1570 34]

[ 228 168]]

	precision	recall	f1-score	support
0	0.87	0.98	0.92	1604
1	0.83	0.42	0.56	396
accuracy			0.87	2000
macro avg	0.85	0.70	0.74	2000
weighted avg	0.86	0.87	0.85	2000





## Evaluation and Final Results

### Conclusion