



CUSTOMER CHURN PREDICTION

GROUP 3

VAMSHI KUMAR KONDURU (11516045)

HARSHAVARDHAN BHUPATHI(11514028)

SANJAY REDDY MANDA(11524794)

DILIP KUMAR PITTALA (11515093)

WHAT EXACTLY IS CHURN PREDICTION?

- Churn is defined in business terms as ‘when a client cancels a subscription to a service they have been using.’ A common example is people cancelling Spotify/Netflix subscriptions.
- So, Churn Prediction is essentially predicting which clients are most likely to cancel a subscription i.e ‘leave a company’ based on their usage of the service.
- From a company point of view, it is necessary to gain this information because acquiring new customers is often arduous and costlier than retaining old ones.
- Hence, the insights gained from Churn Prediction helps them to focus more on the customers that are at a high risk of leaving.

WHY DOES CHURN OCCUR?

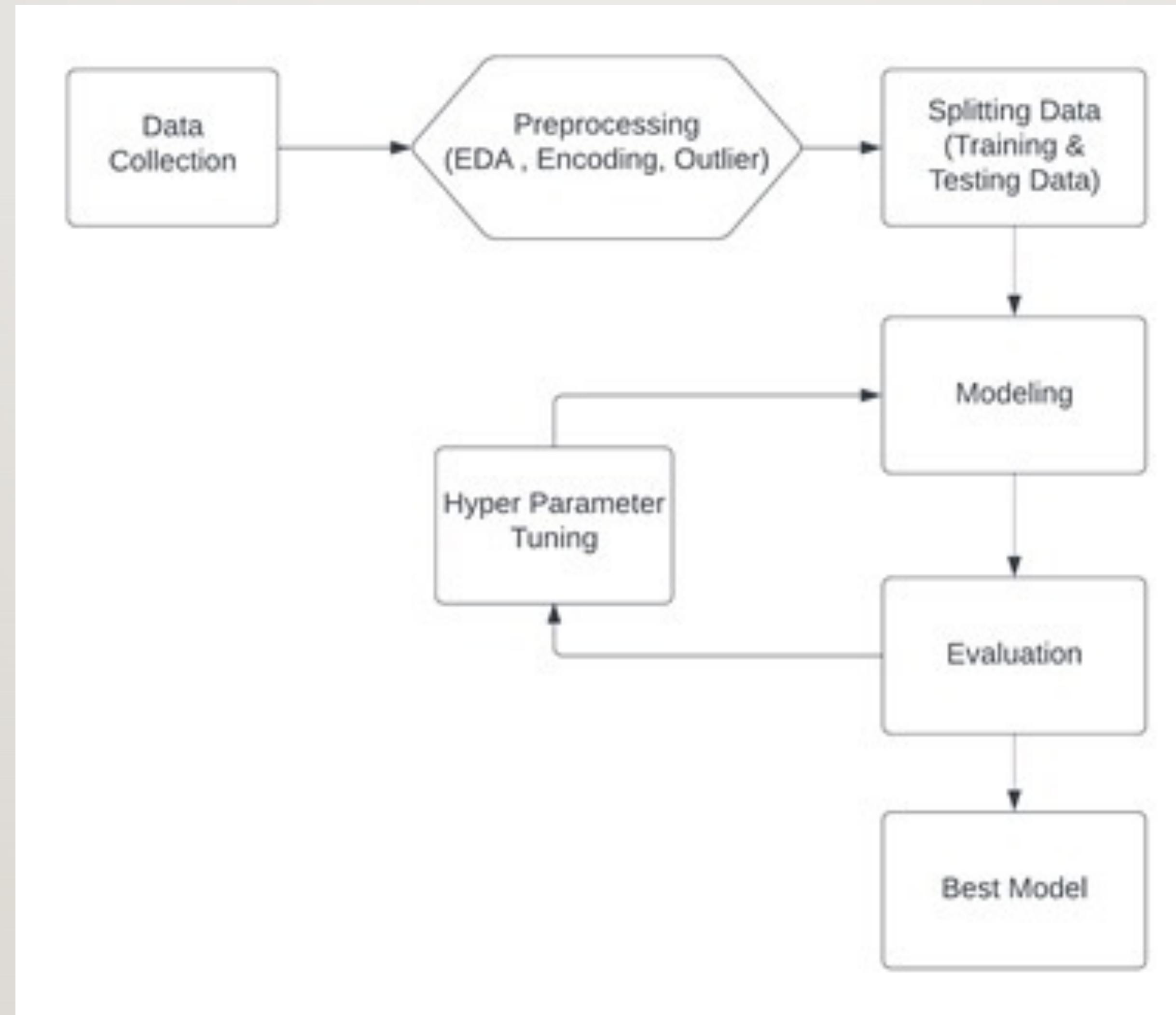
- **Fees** - Raising fees on financial service products is the number one reason why customers consider leaving in the first place. However, poor service ends up being the primary reason for actually leaving.
- **Rates** - Much like fees, the return rates offered by products like Money Market funds and savings accounts are highly competitive across fintech startups, new online entrants, and the incumbents.
- **Branch locations** - Some institutions have adopted an online-only model recently to save costs, but the majority of customers still like having a physical bank to get in-person guidance from a valued advisor on investment strategies.
- **24/7 customer service and wait times** - No one wants to be on hold for any amount of time, especially when there is financial risk involved.
- **Quality of digital tools** - All banking customers surveyed spend 69% of their time online or using the mobile app, with millennials even higher at 79%.

Stages of Churn Management:

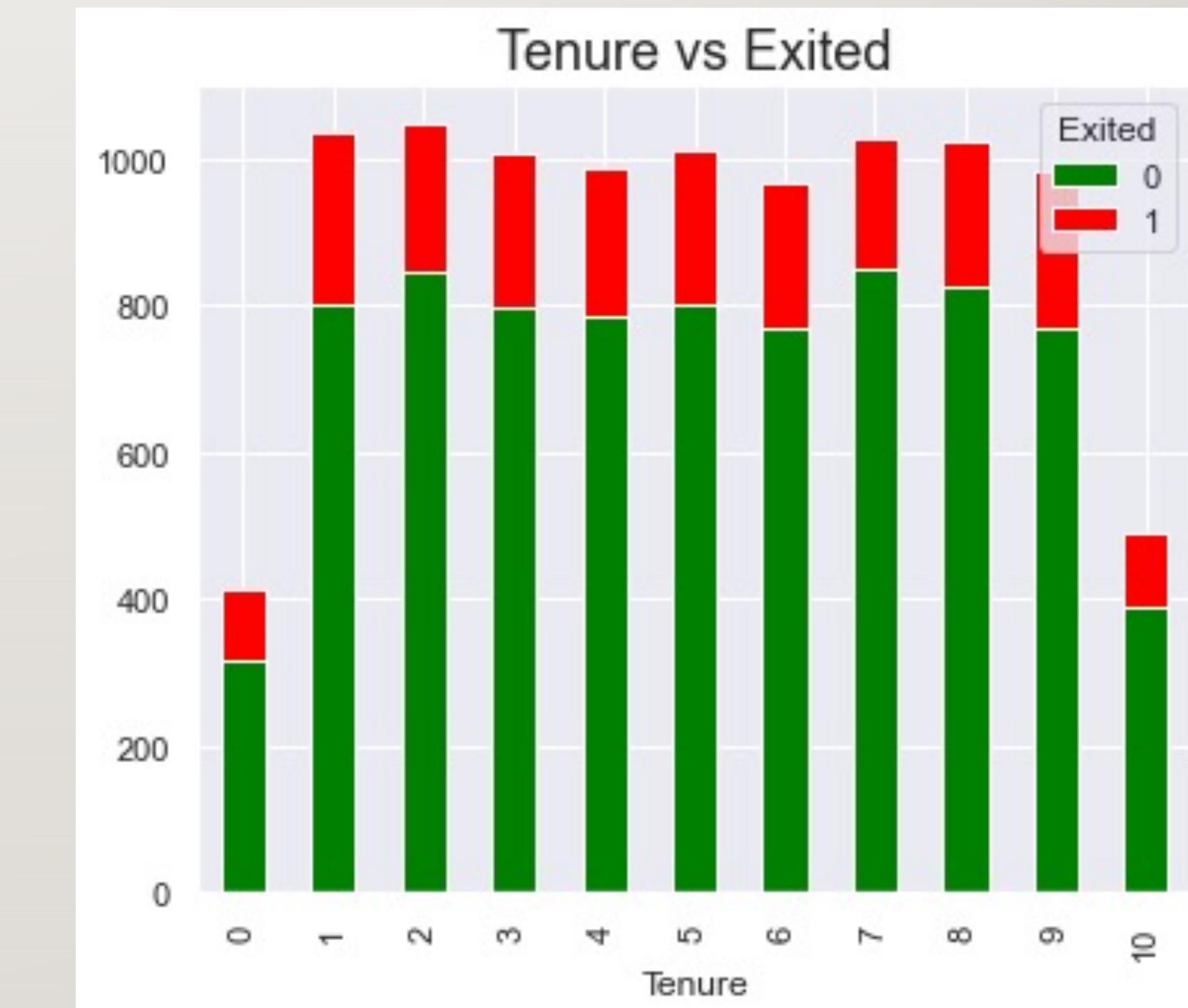
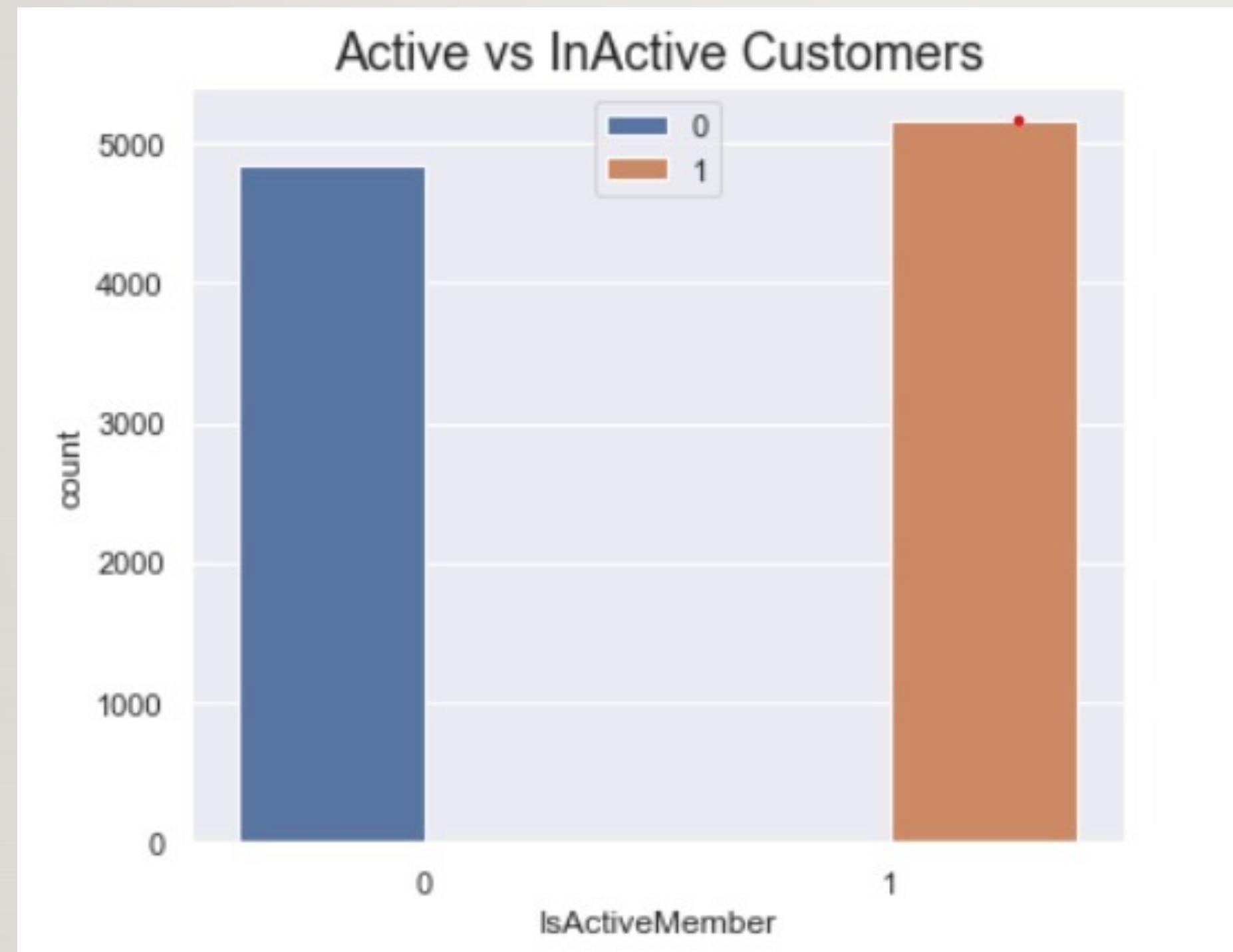
- The firm is need of managing its customer churn and it will achieve this in four stages such as - acquiring churned customer, delighting customer, preventing customer attrition and saving customers through various campaigns.



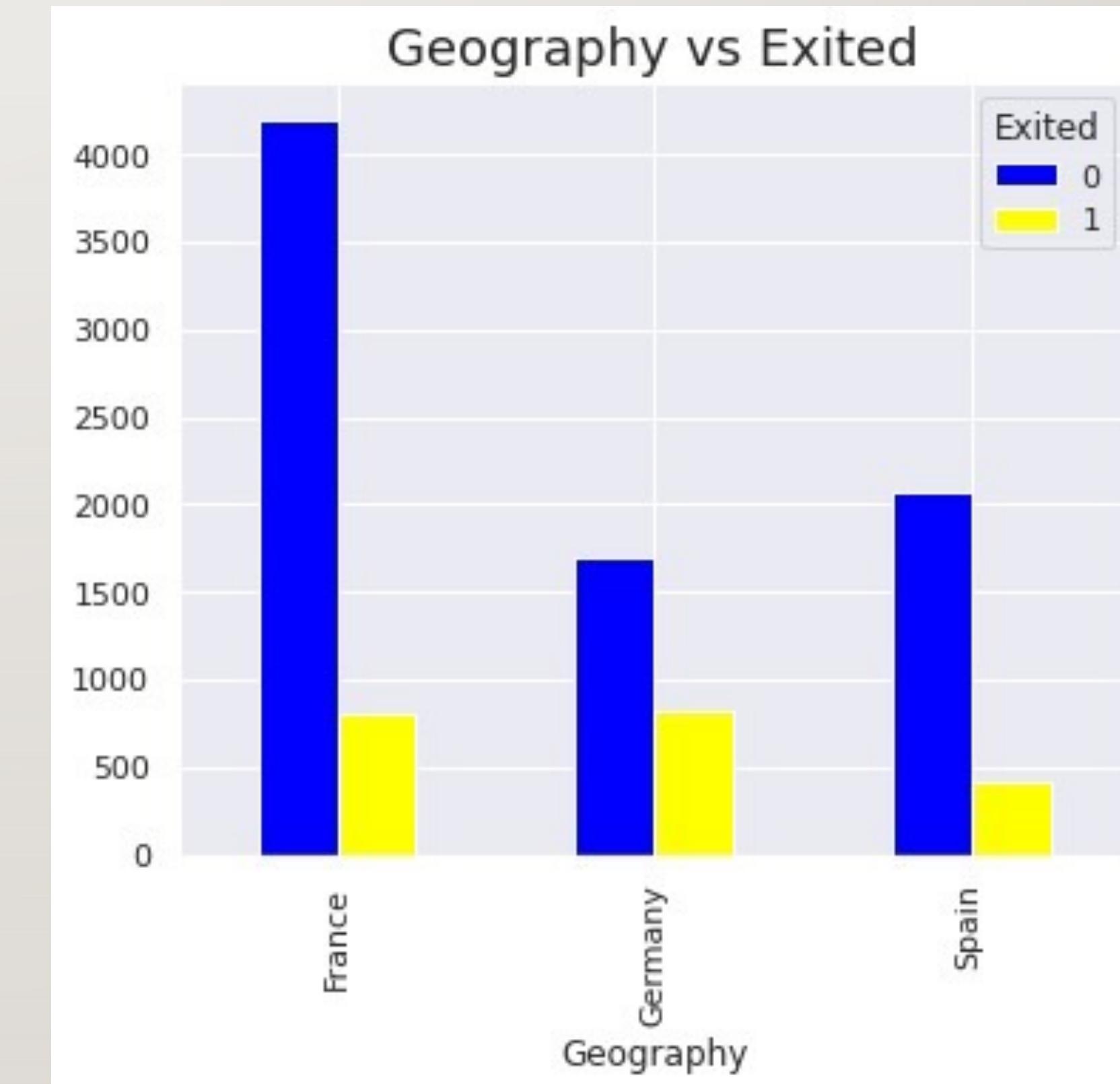
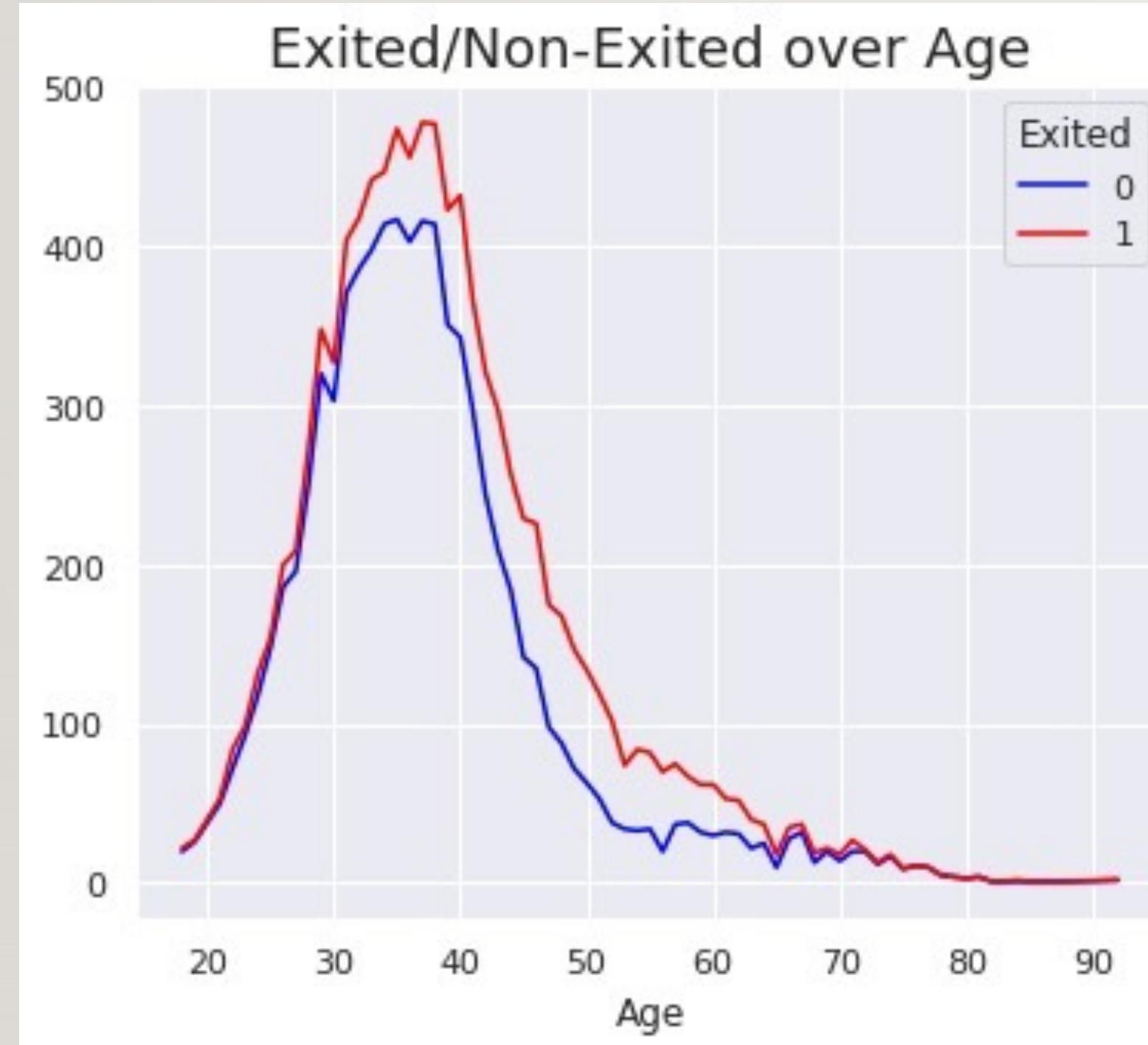
Methodology



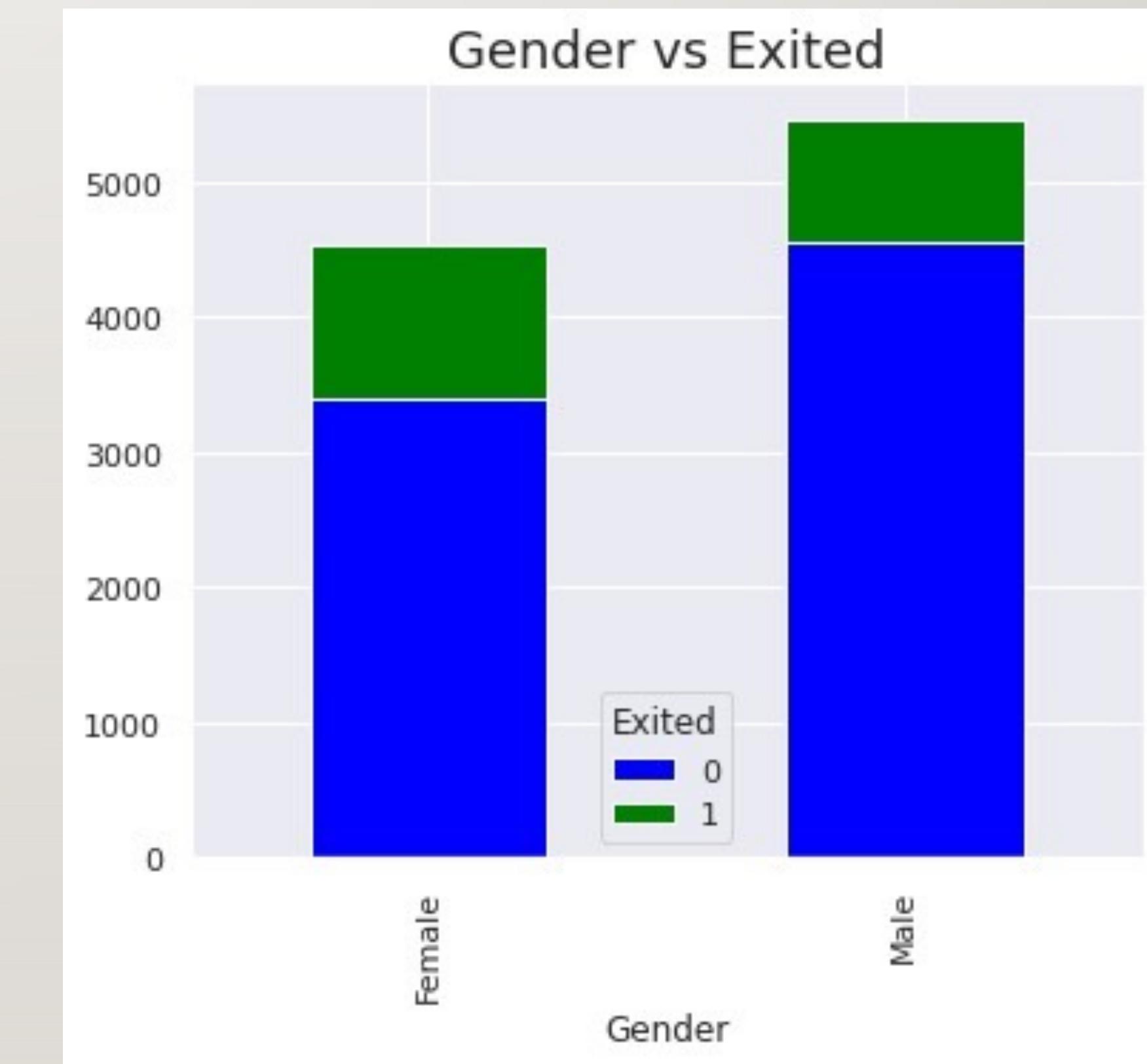
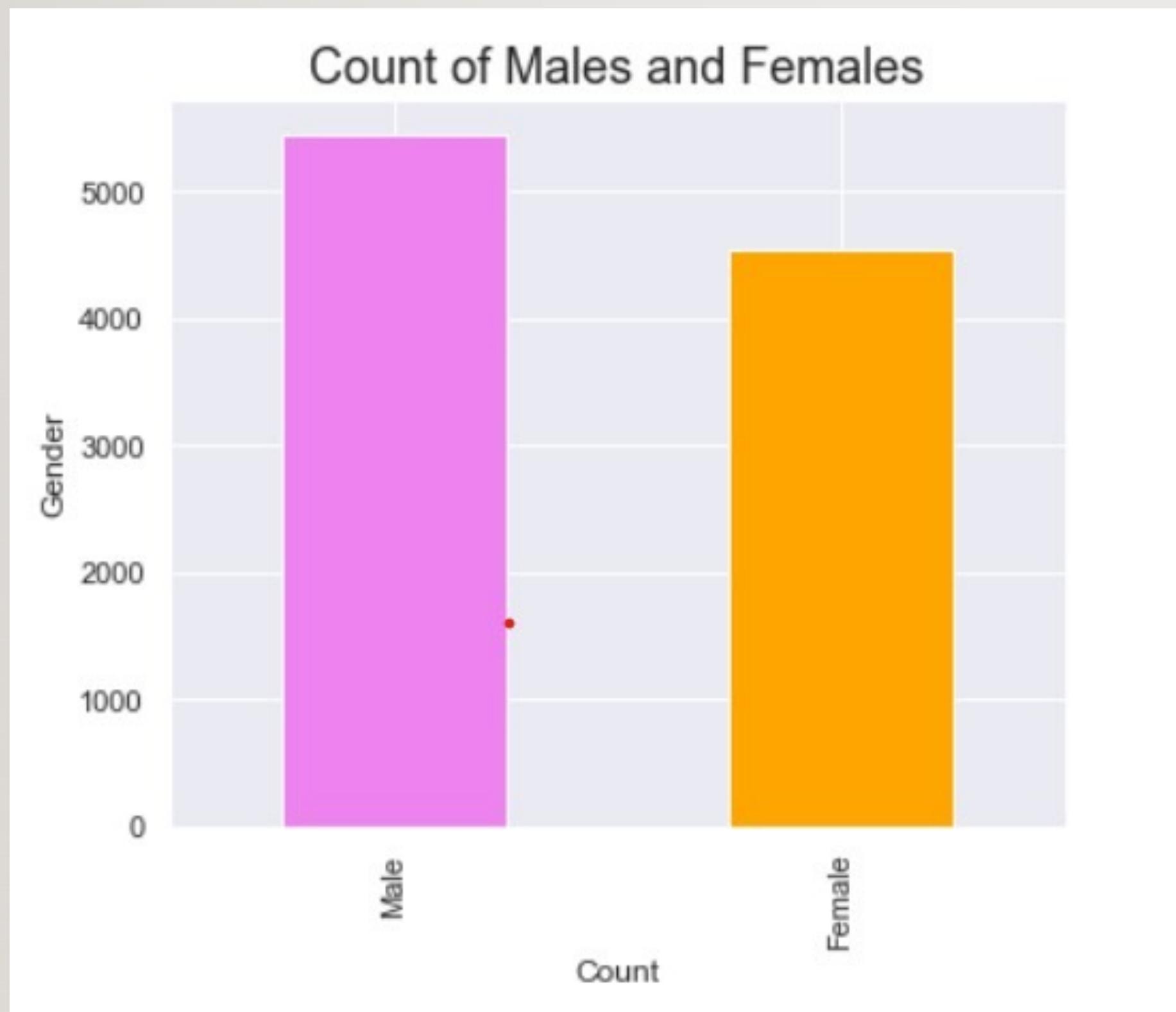
Exploratory data analysis



Exploratory data analysis

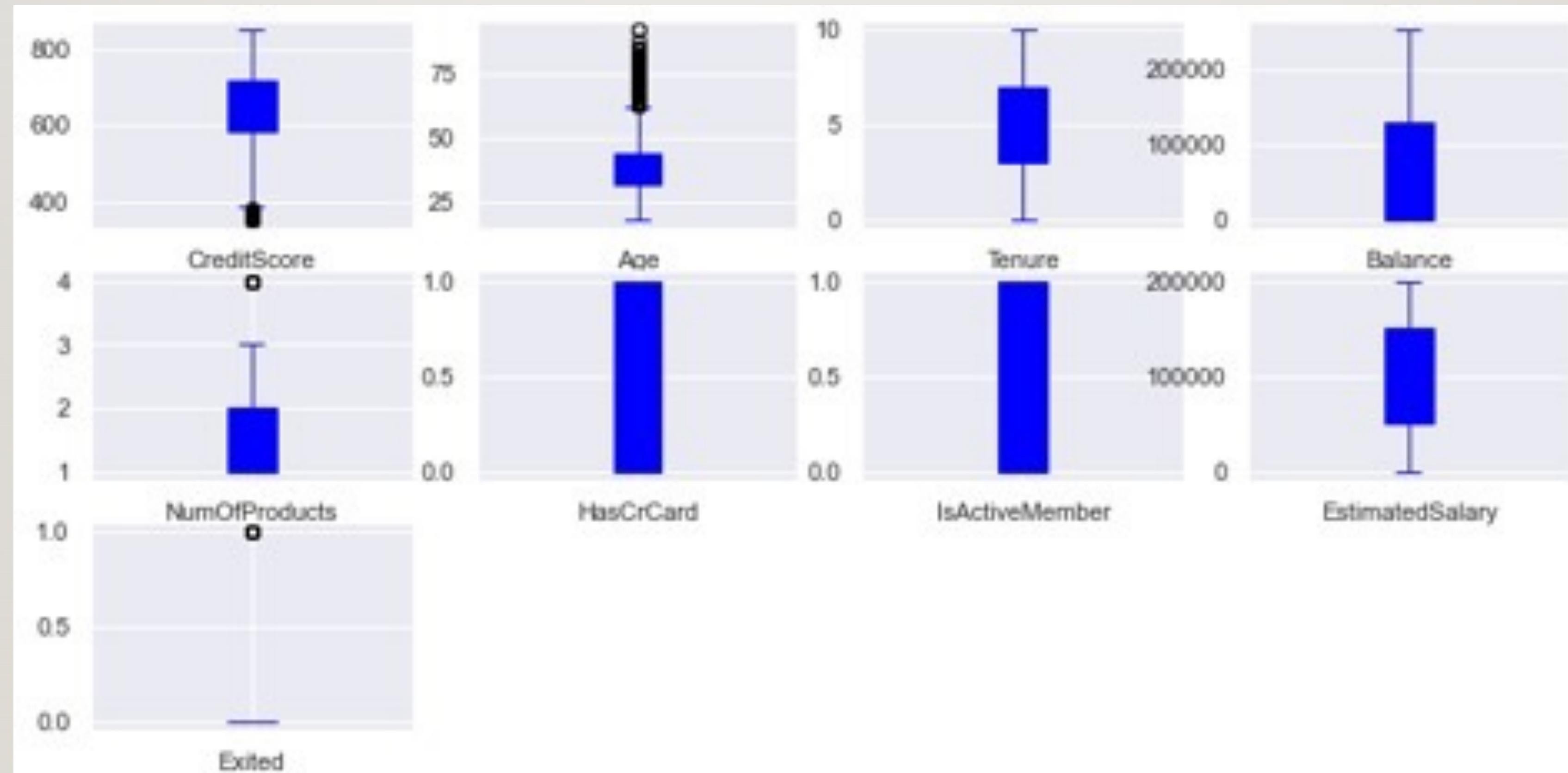


Exploratory data analysis

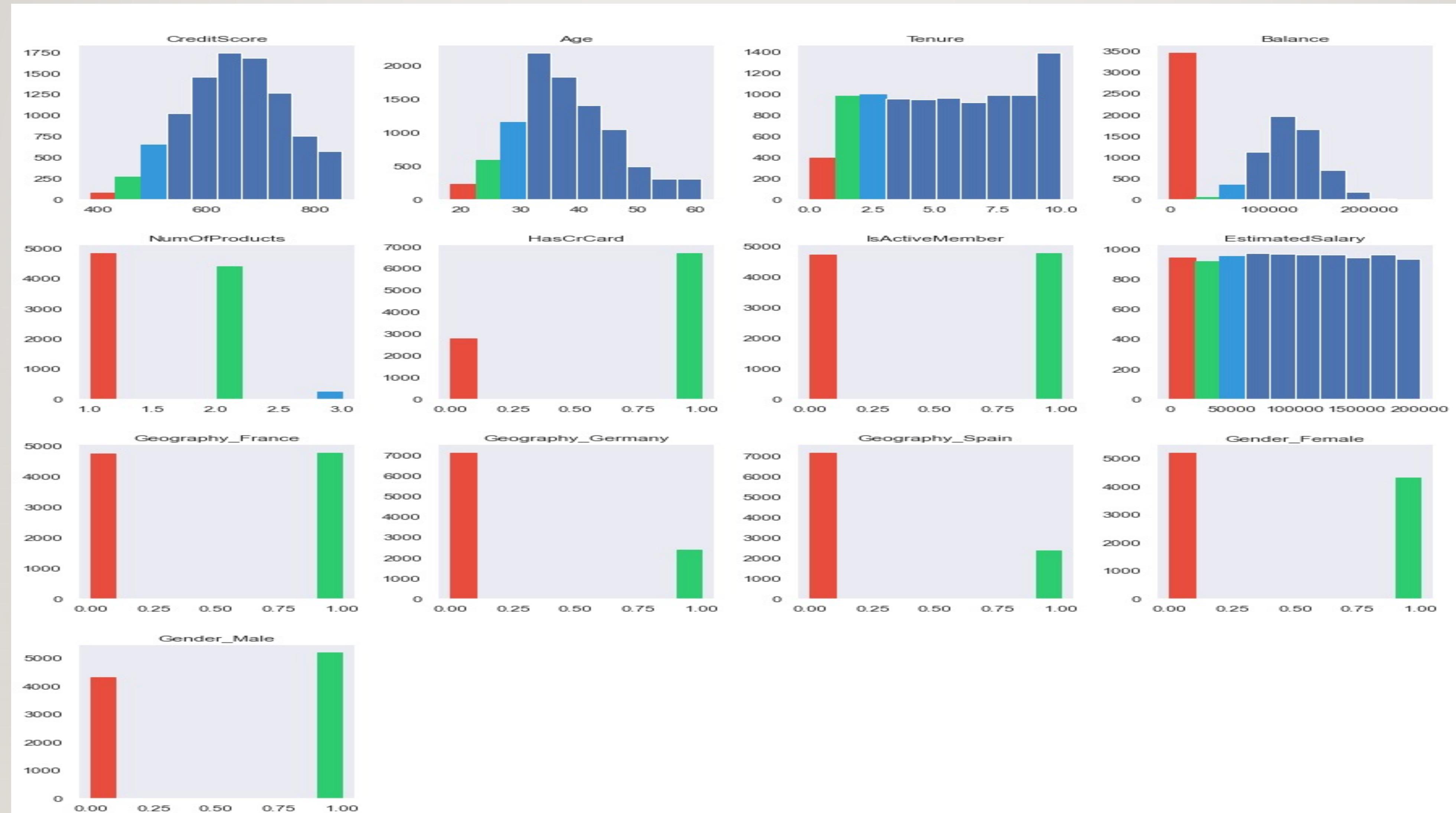


Box Plot Distribution

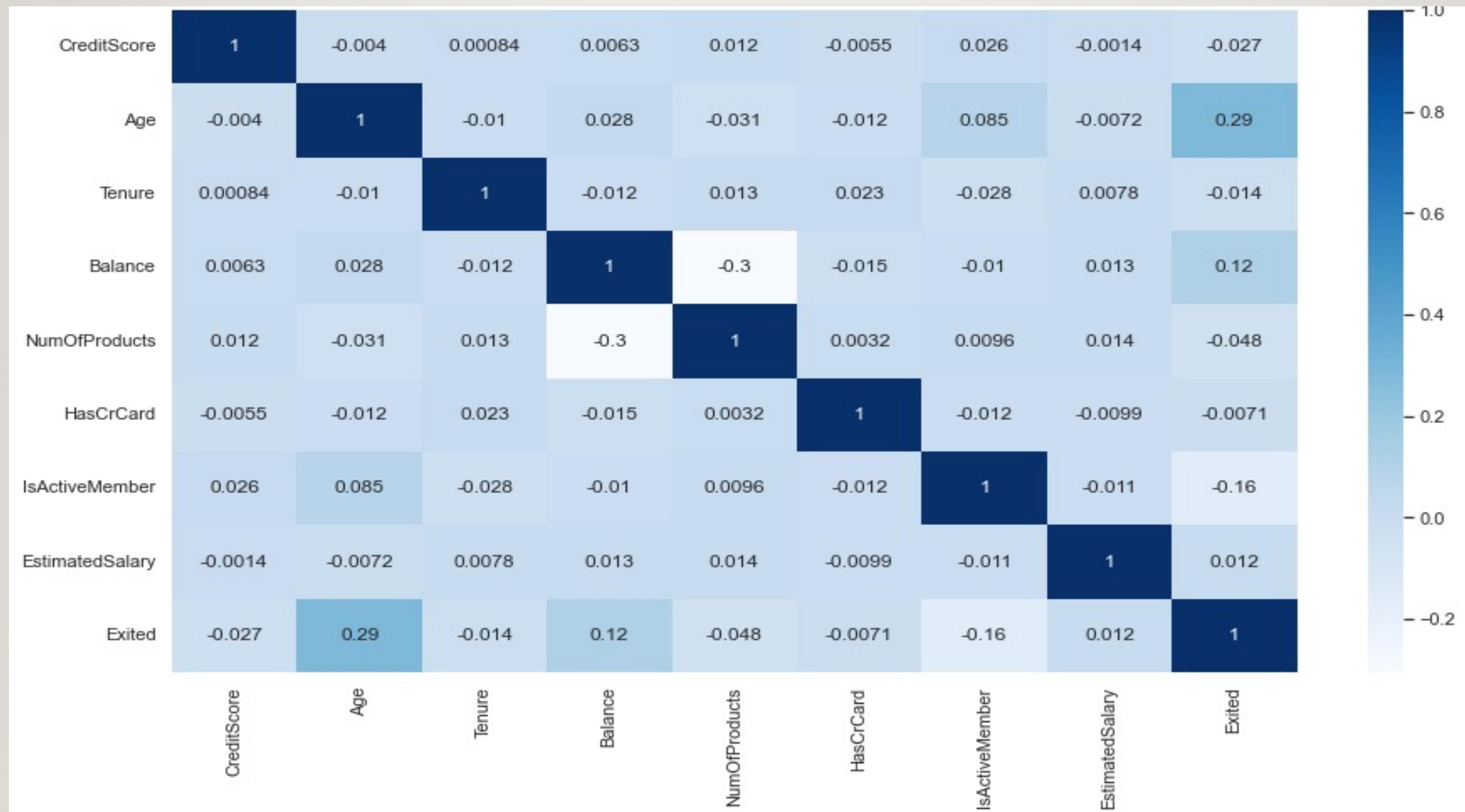
- box plot shows the distribution of numeric data and outliers can be observed for each feature if exists.
- You could See Features like CreditScore, Age and NumOfproducts have Outliers in them which can be treated by trimming.



Distribution Data Using Histogram Plot



Correlation Heatmap



Modeling

- Pipeline - Automates the ML workflow

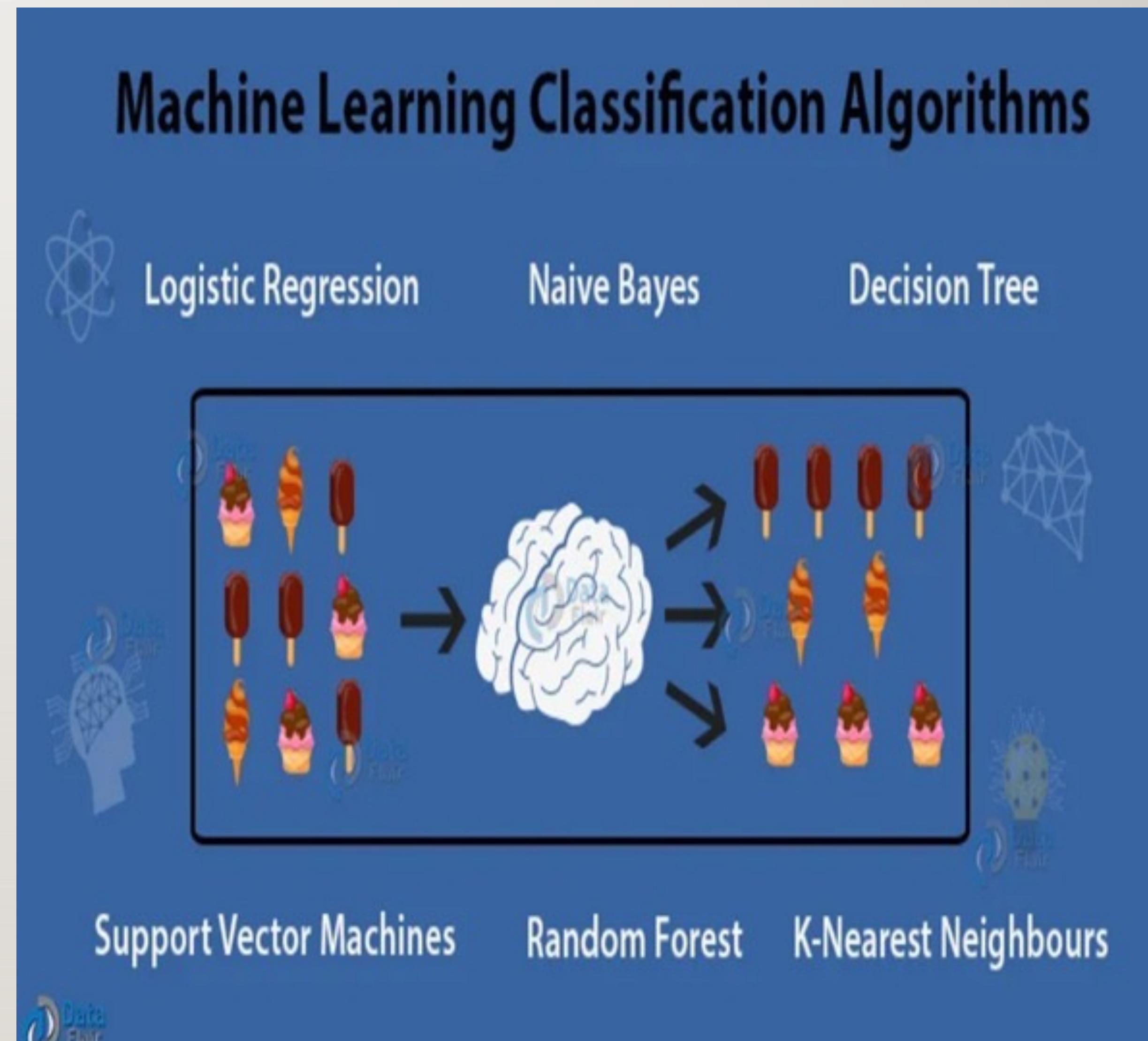
```
from sklearn.pipeline import Pipeline|
```

- Hyper parameter tuning GridSearchCV - Model Optimization

```
from sklearn.model_selection import GridSearchCV
```

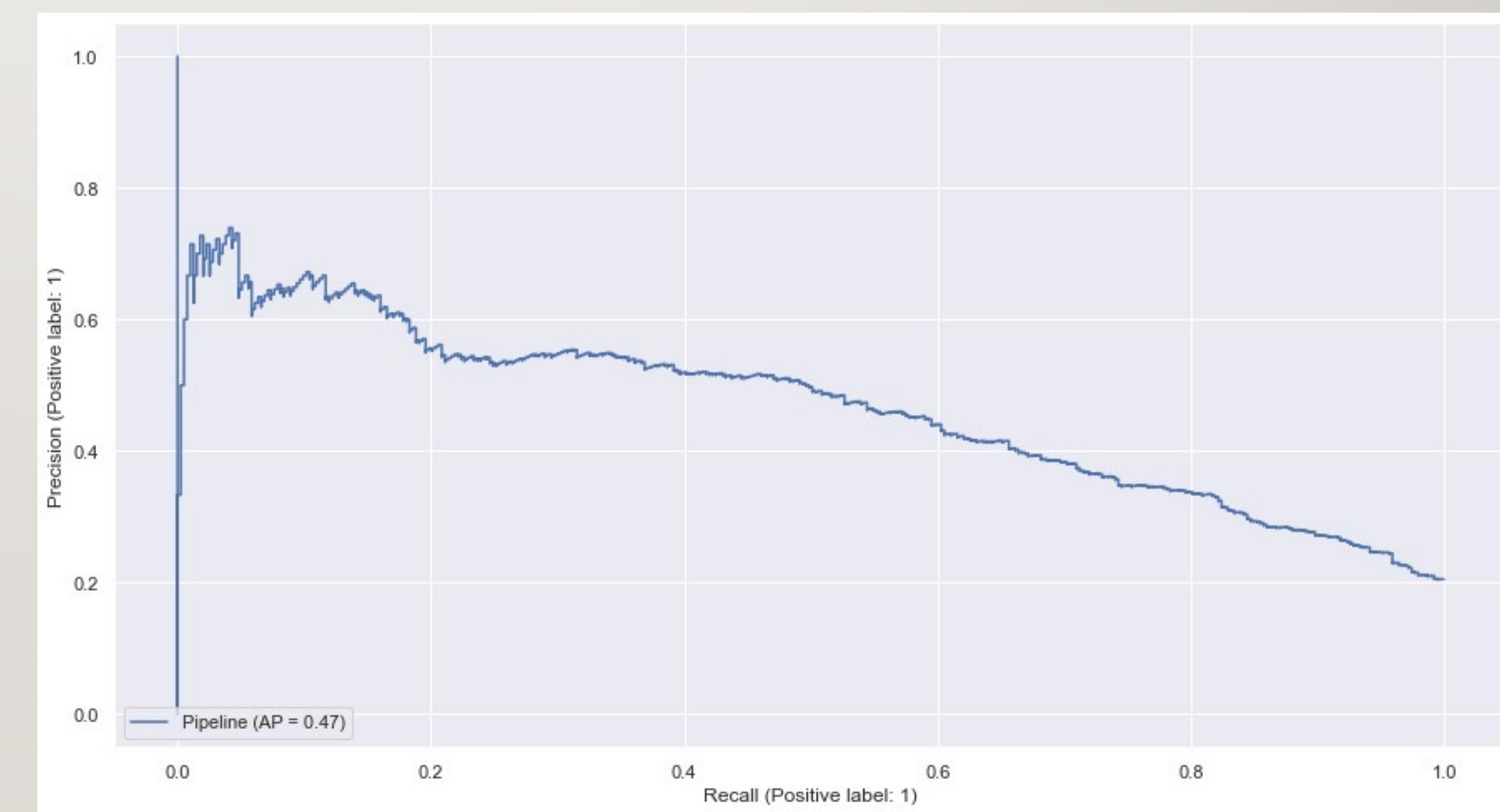
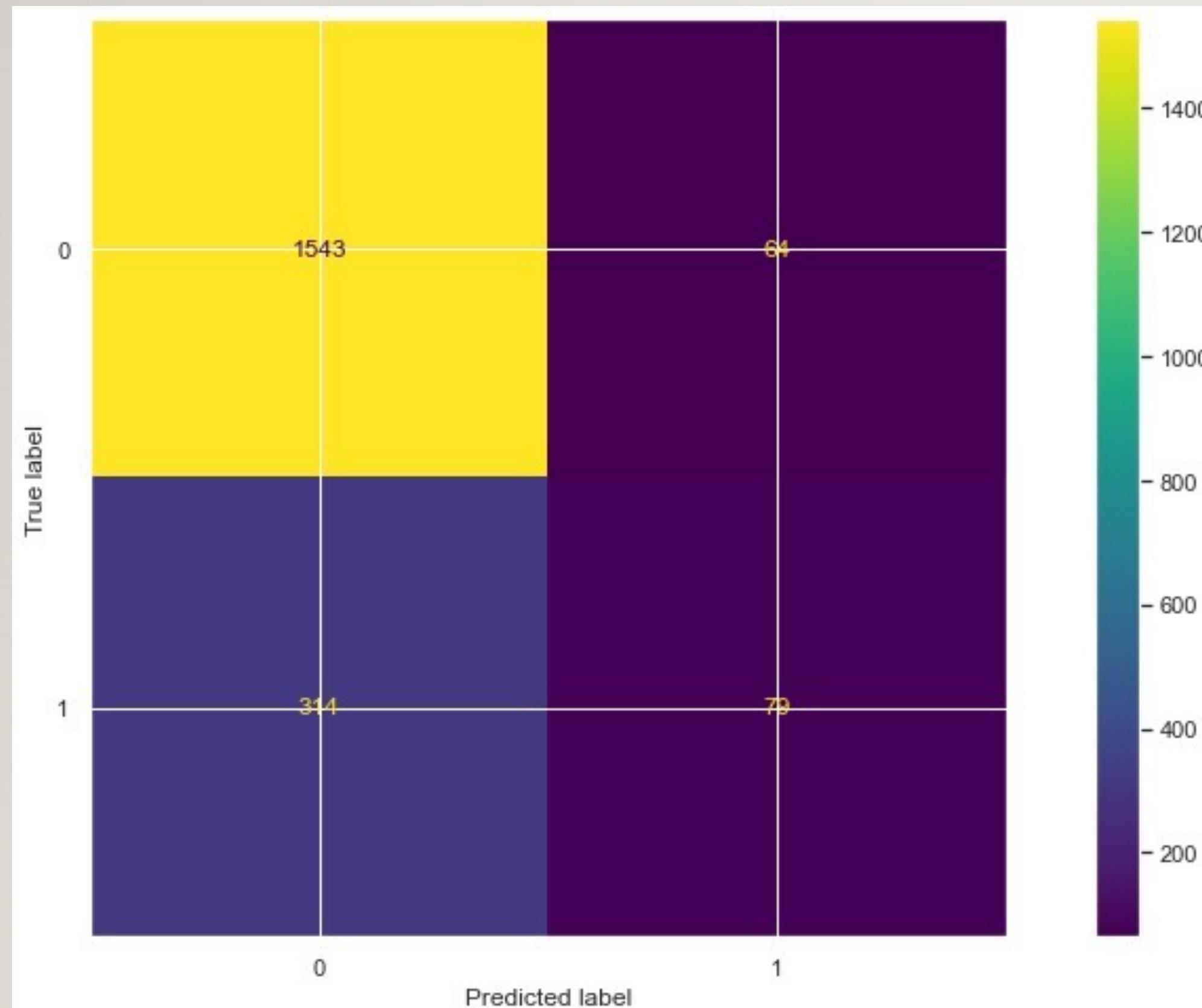
ML Algorithms

1. Logistic Regression
2. Random Forest
3. Decision Tree
4. Support Vector Machine (SVM)
5. KNN Classification
6. Compliment Naive Bayes



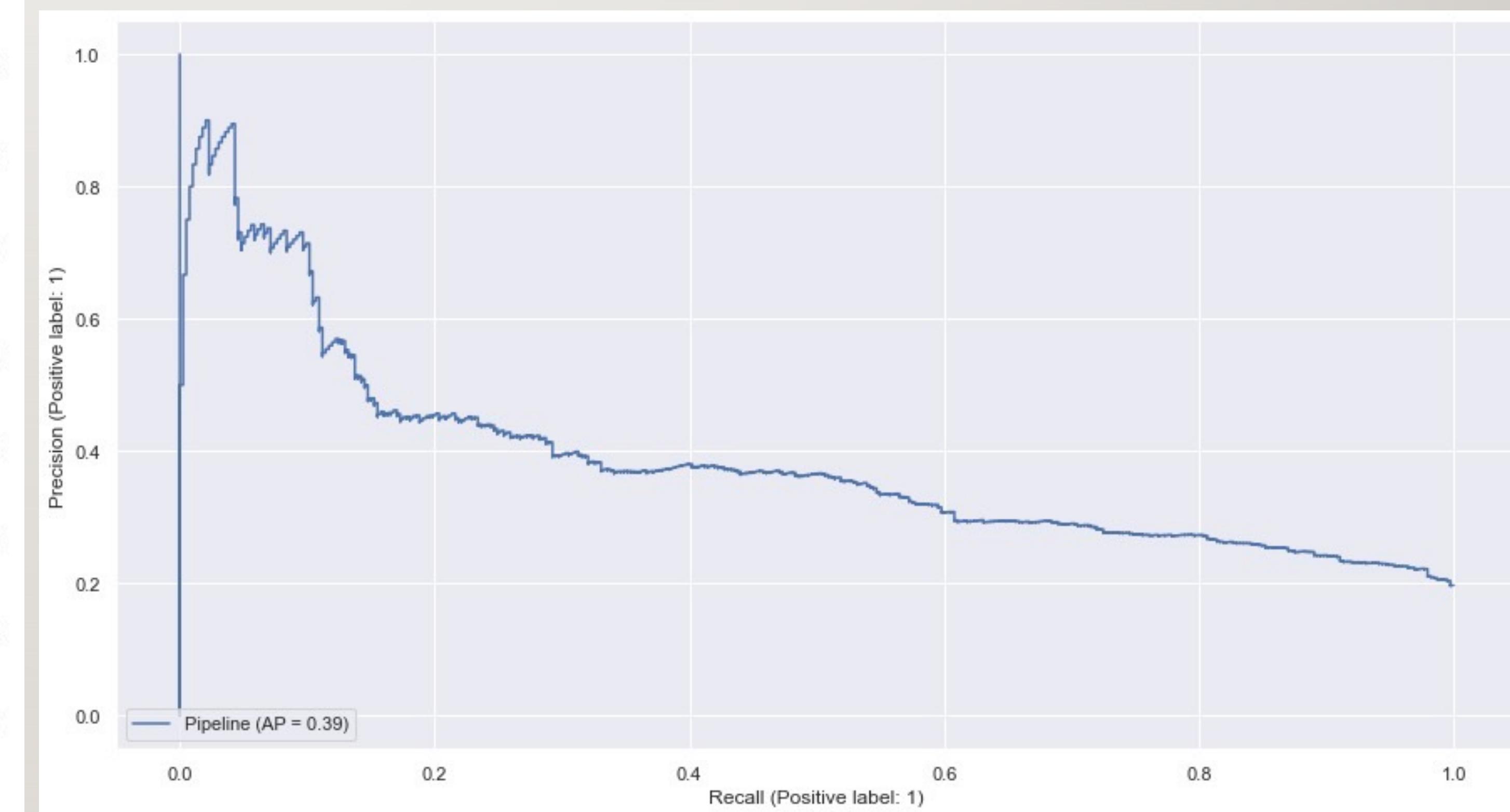
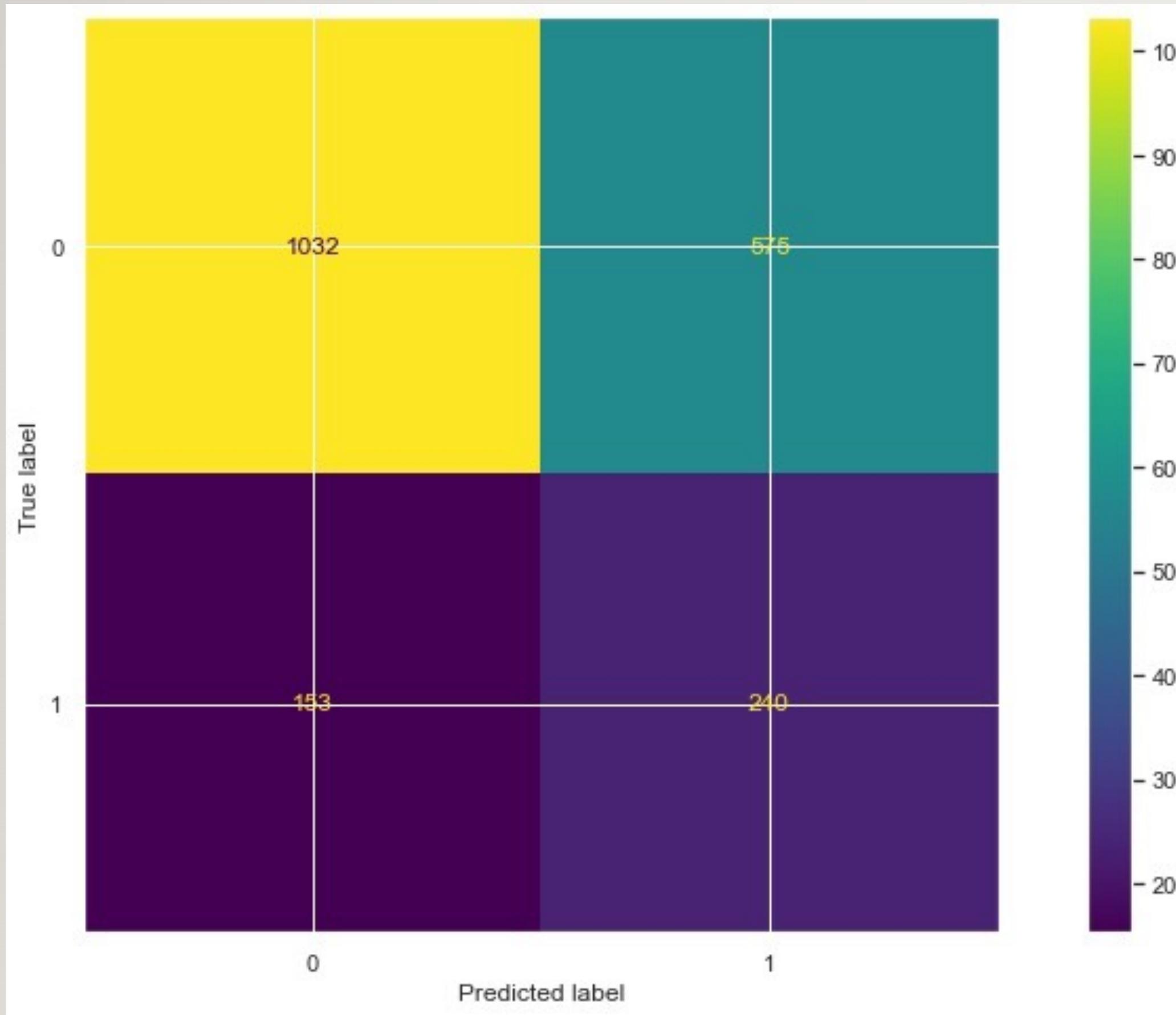
Logistic Regression

- It is a process of modeling the probability of a discrete outcome given an input variable.
- A logistic regression model predicts a dependent data variable by analyzing the relationship between one or more existing independent variables.



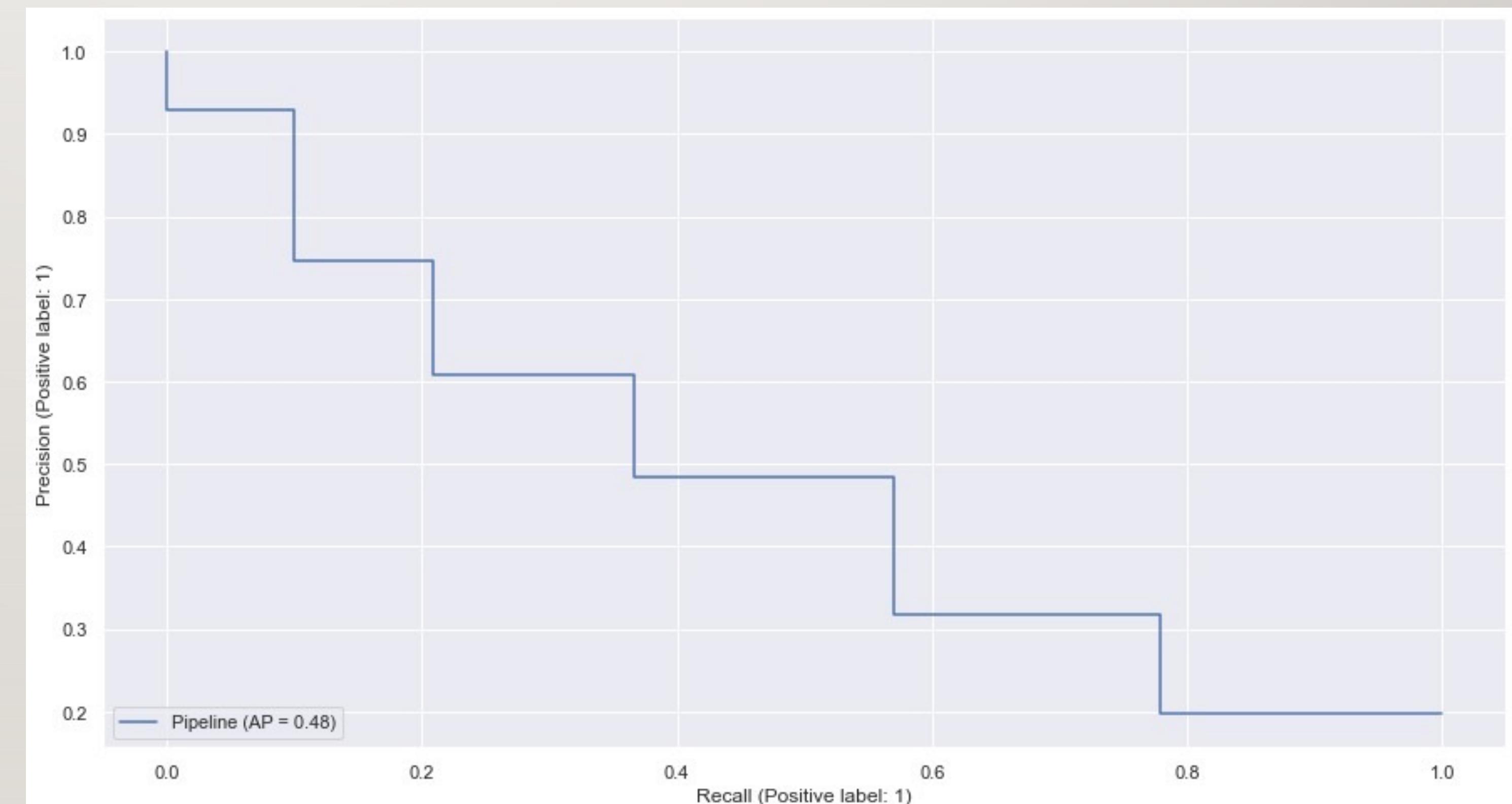
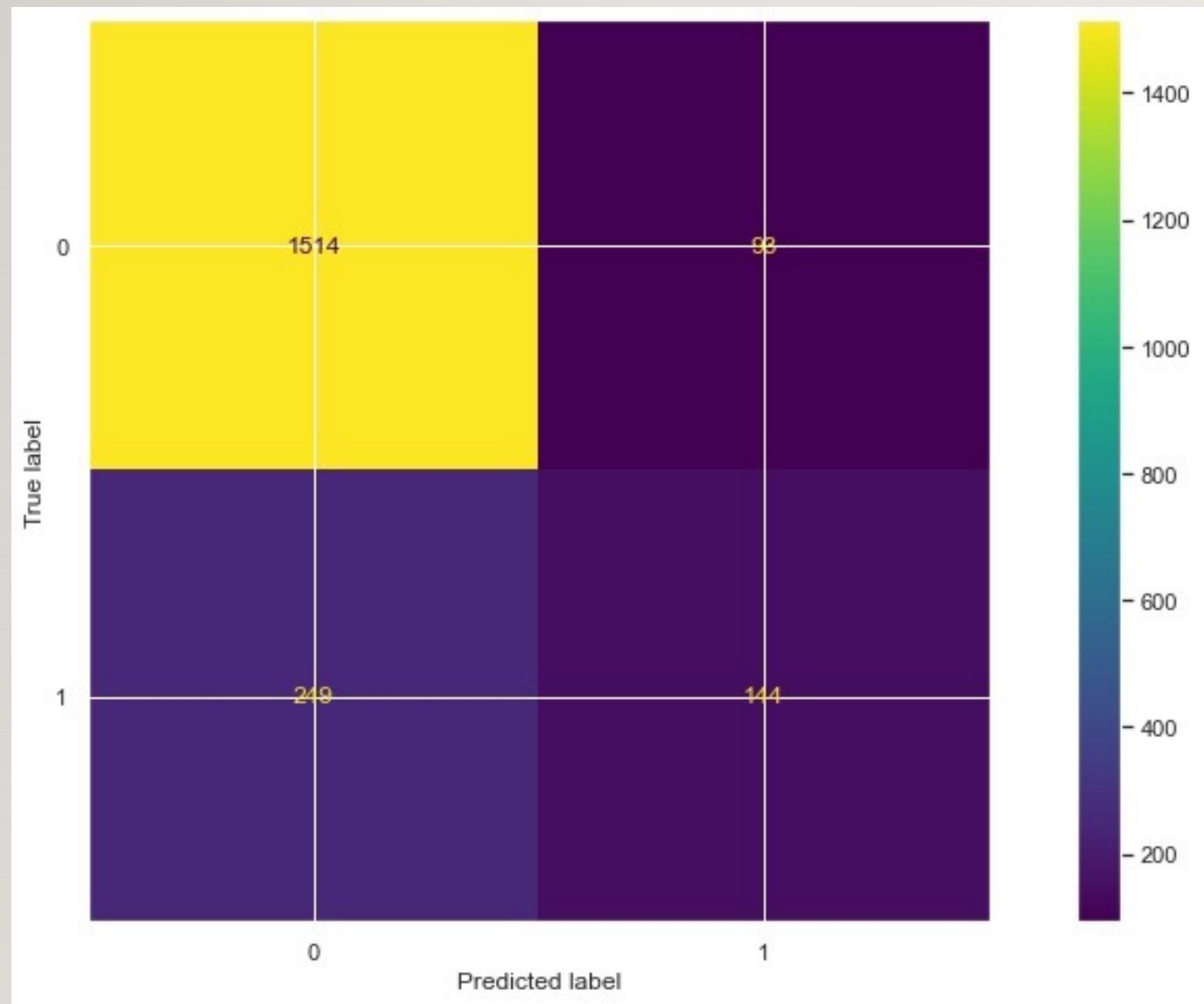
Complement Naive Bayes

- I understand that naive bayes works by computing the probability of a document belonging to a class based on its features.



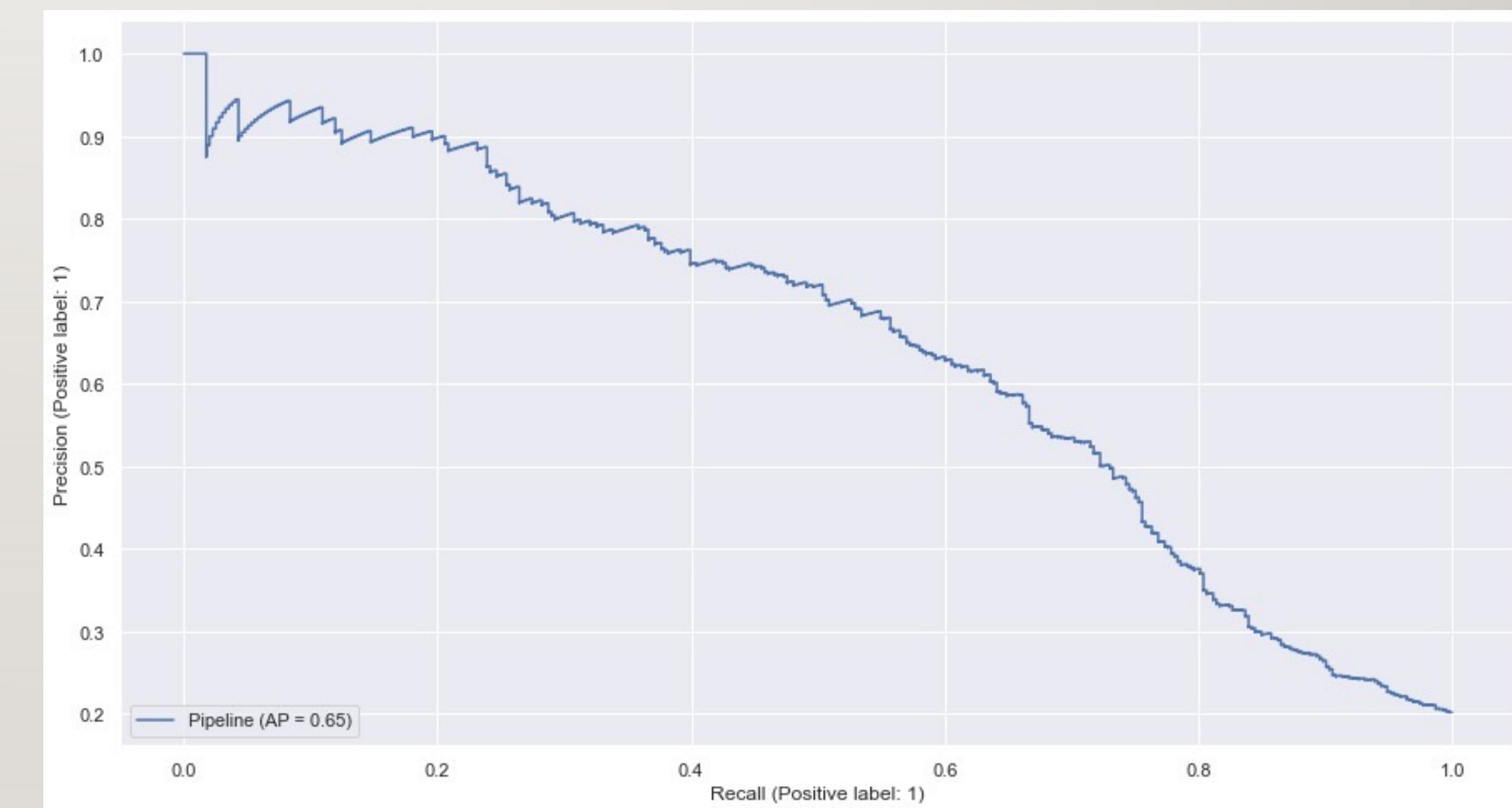
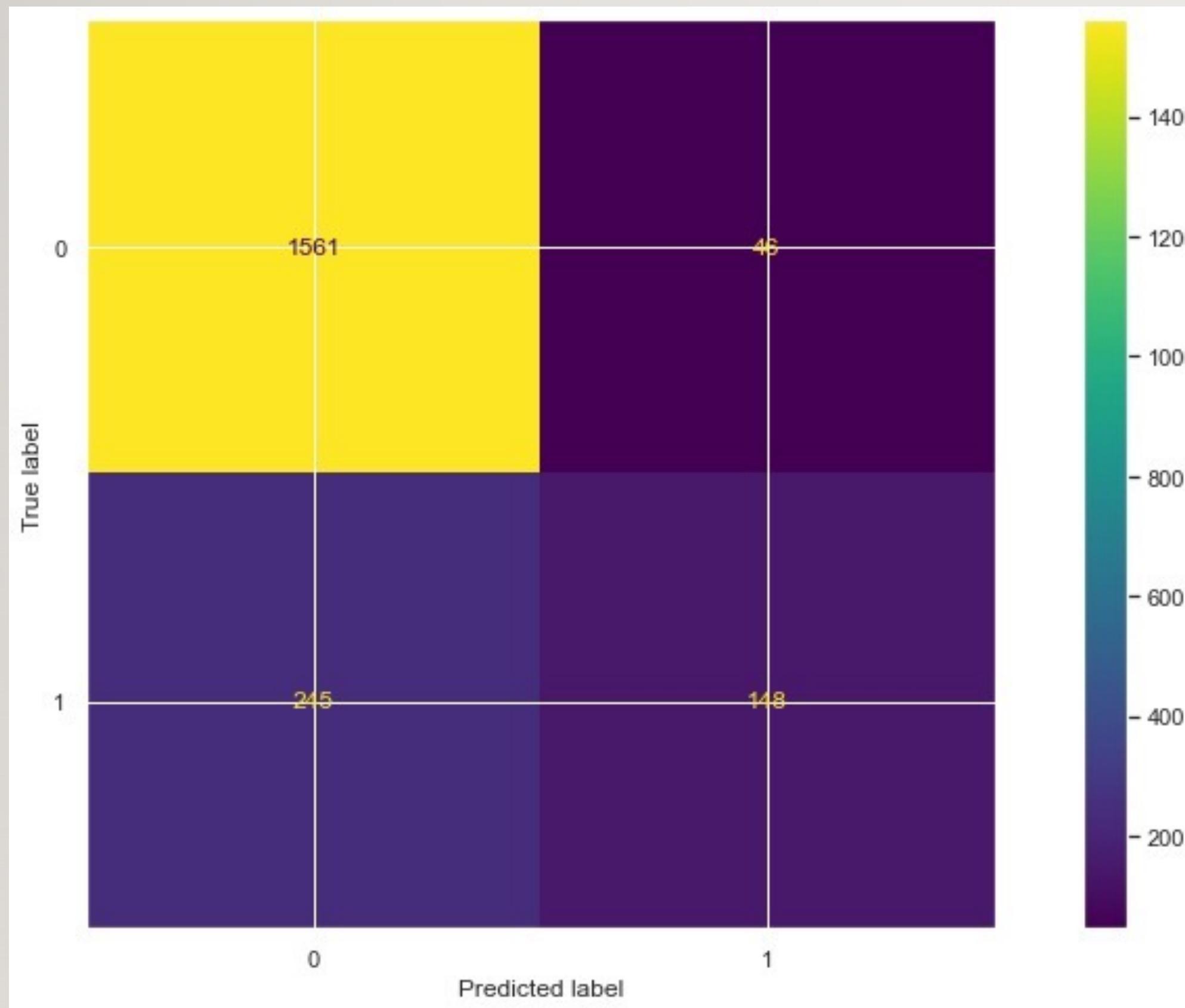
K-Nearest Neighbors

- The KNN algorithm assumes that similar things exist in close proximity.
- An understanding of how we calculate the distance between points on a graph is necessary before moving on.



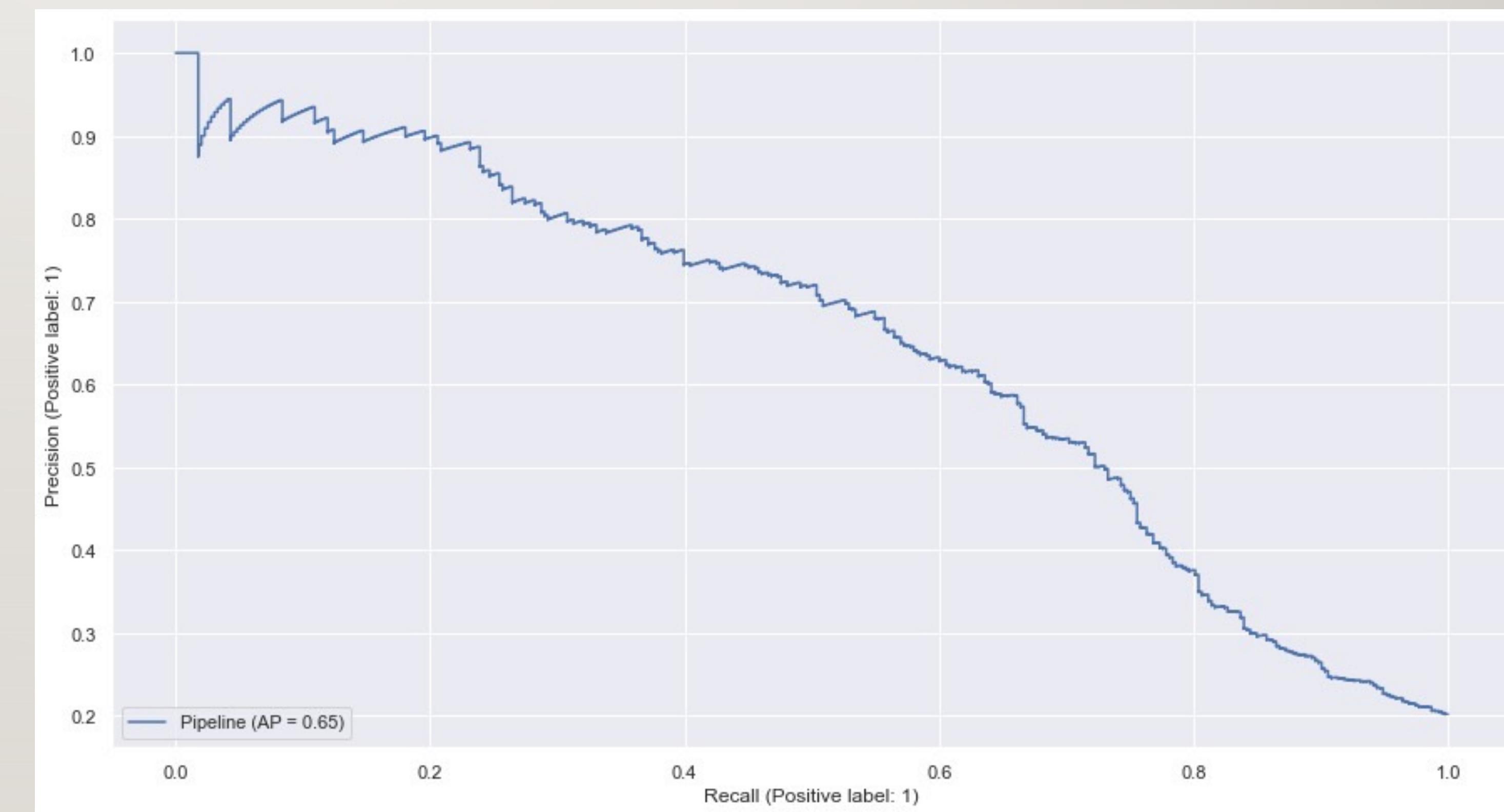
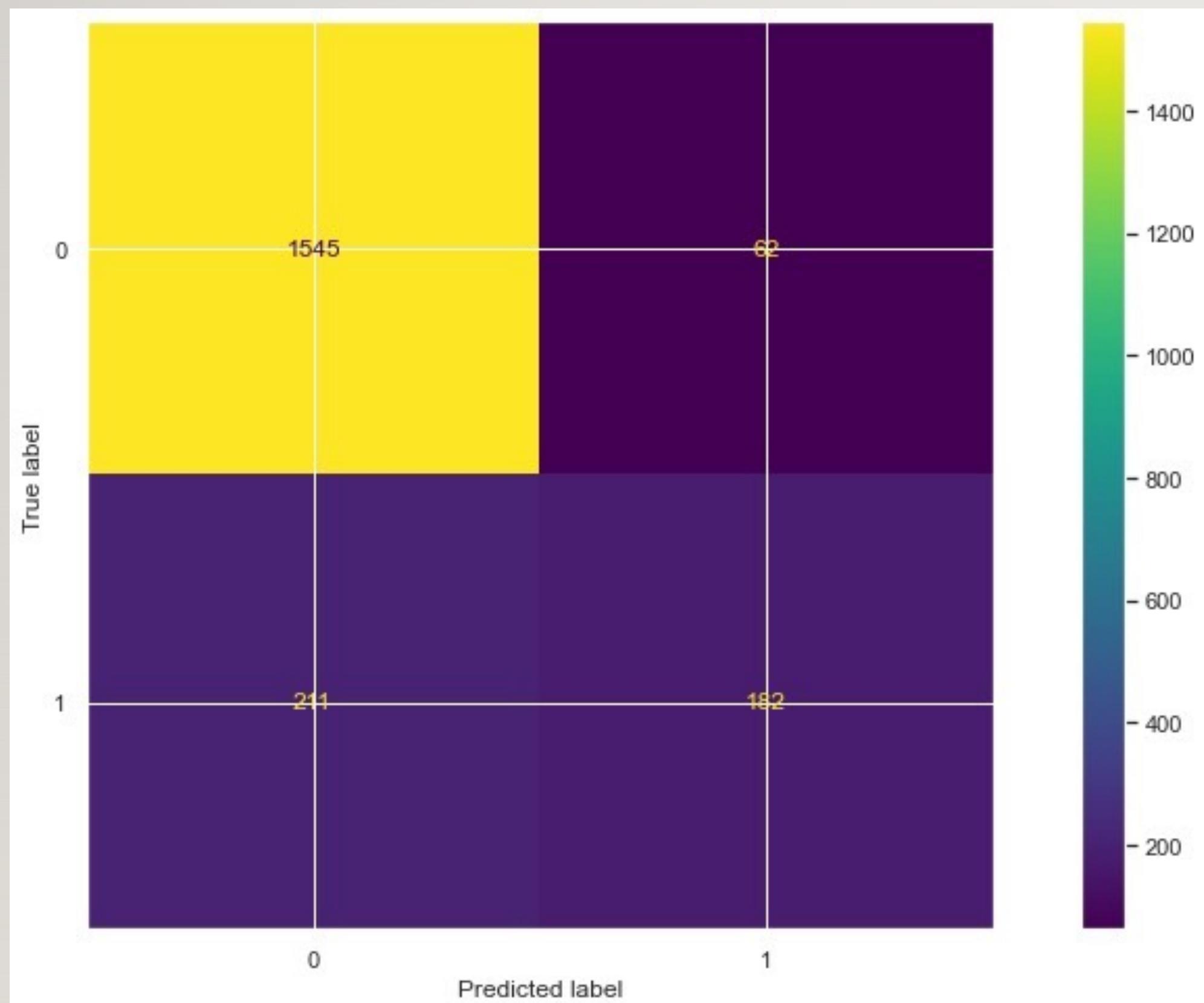
Support Vector Machine

- It is a supervised machine learning algorithm that can be used for both classification or regression challenges. However, it is mostly used in classification problems.



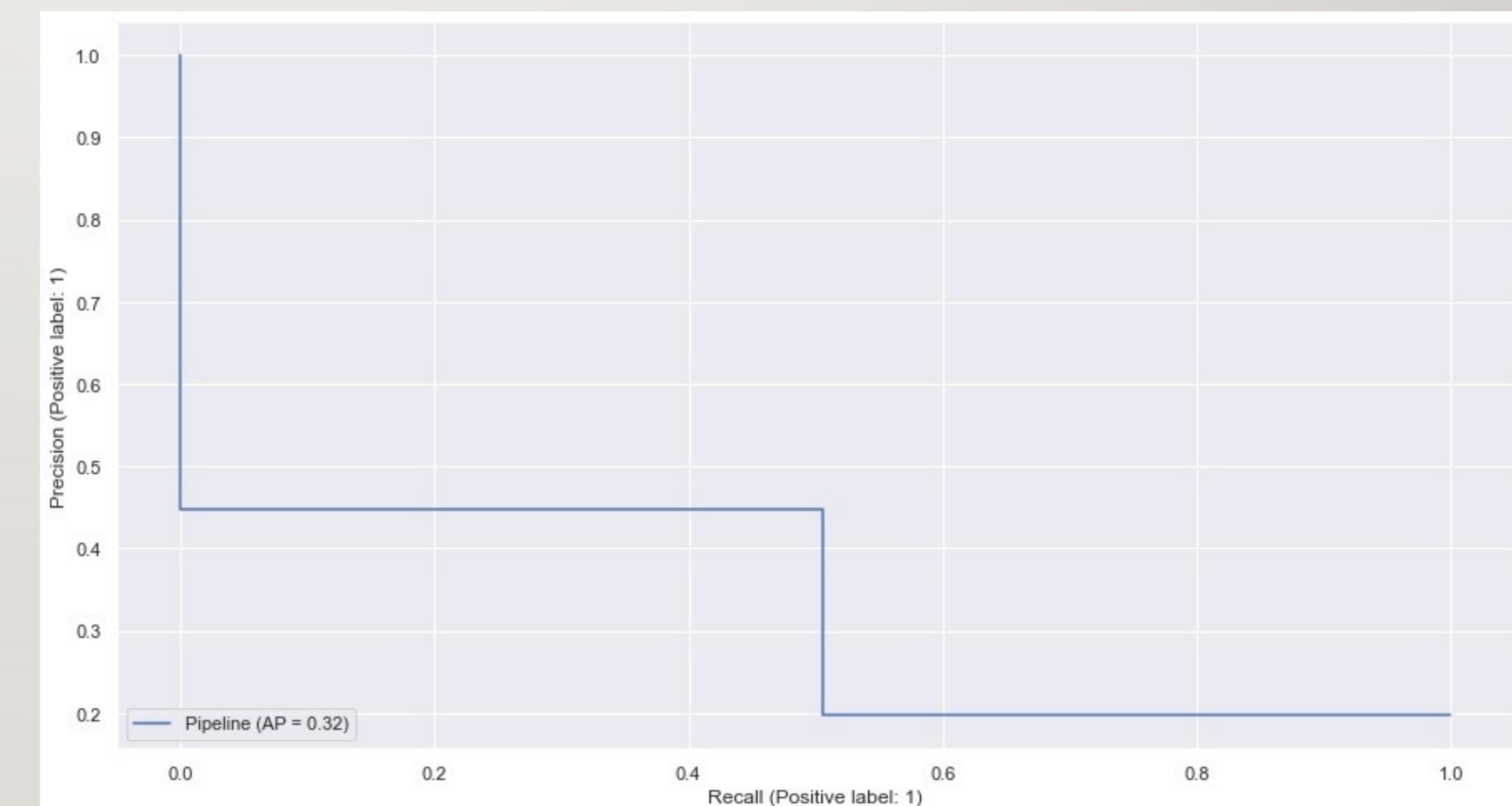
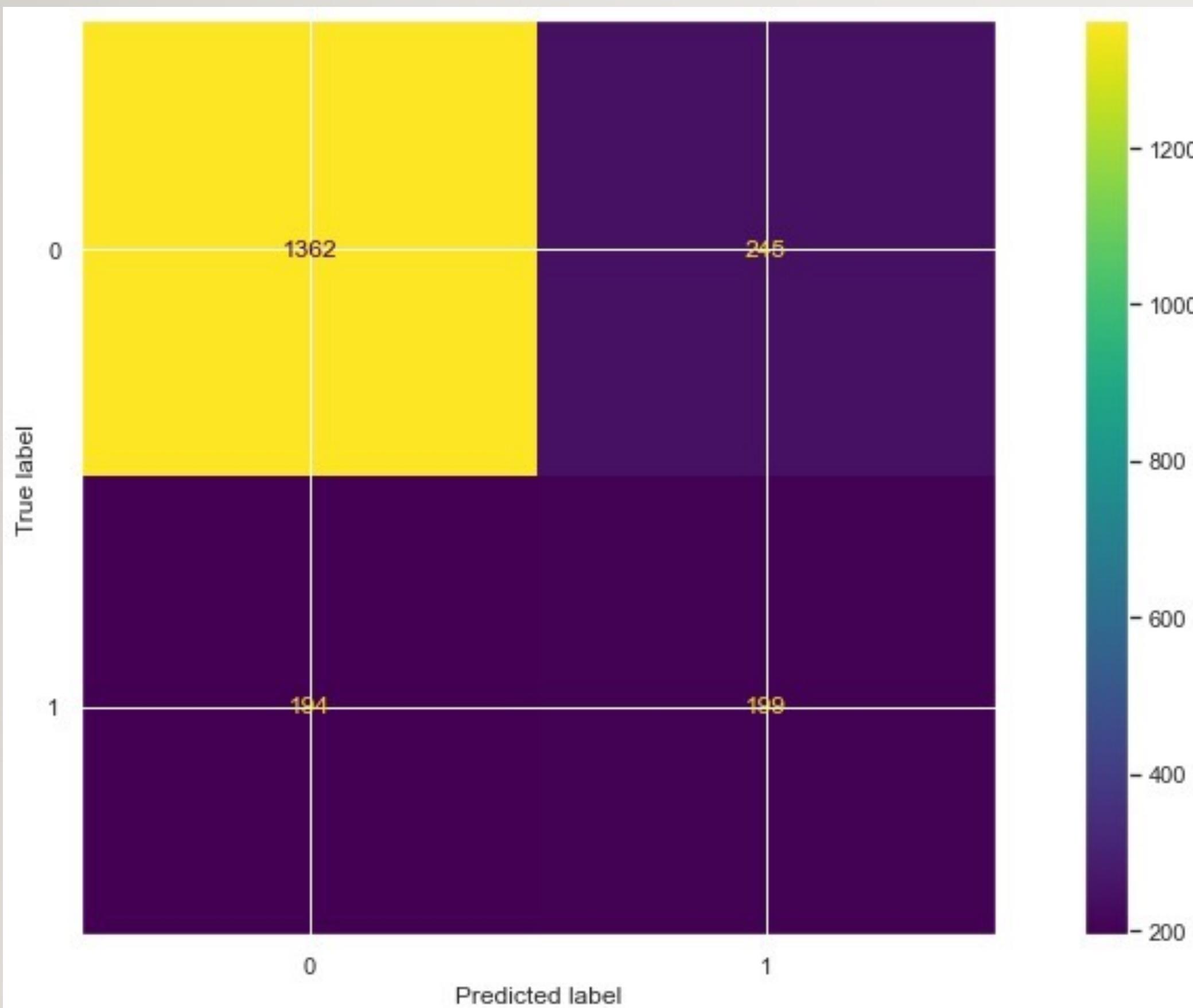
Random Forest

- It is a Supervised Machine Learning Algorithm that is used widely in Classification and Regression problems.
- It builds decision trees on different samples and takes their majority vote for classification and average in case of regression.



Decision Tree

- It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome.



Future Scope

- Resampling the Data
- Inclusion of more Data
- Different hyperparameters



- F1 score is a preferred Evaluation Metric
- Best fit model after Optimization – Decision Tree

Decision Tree :				
	precision	recall	f1-score	support
0	0.87	0.96	0.92	1527
1	0.74	0.43	0.55	376
accuracy			0.86	1903
macro avg	0.81	0.70	0.73	1903
weighted avg	0.85	0.86	0.84	1903

CONCLUSION

Therefore grad
use of any and all pictures or video
aterials including, but not limited to

THANK YOU