ConnectTel Telecom Company

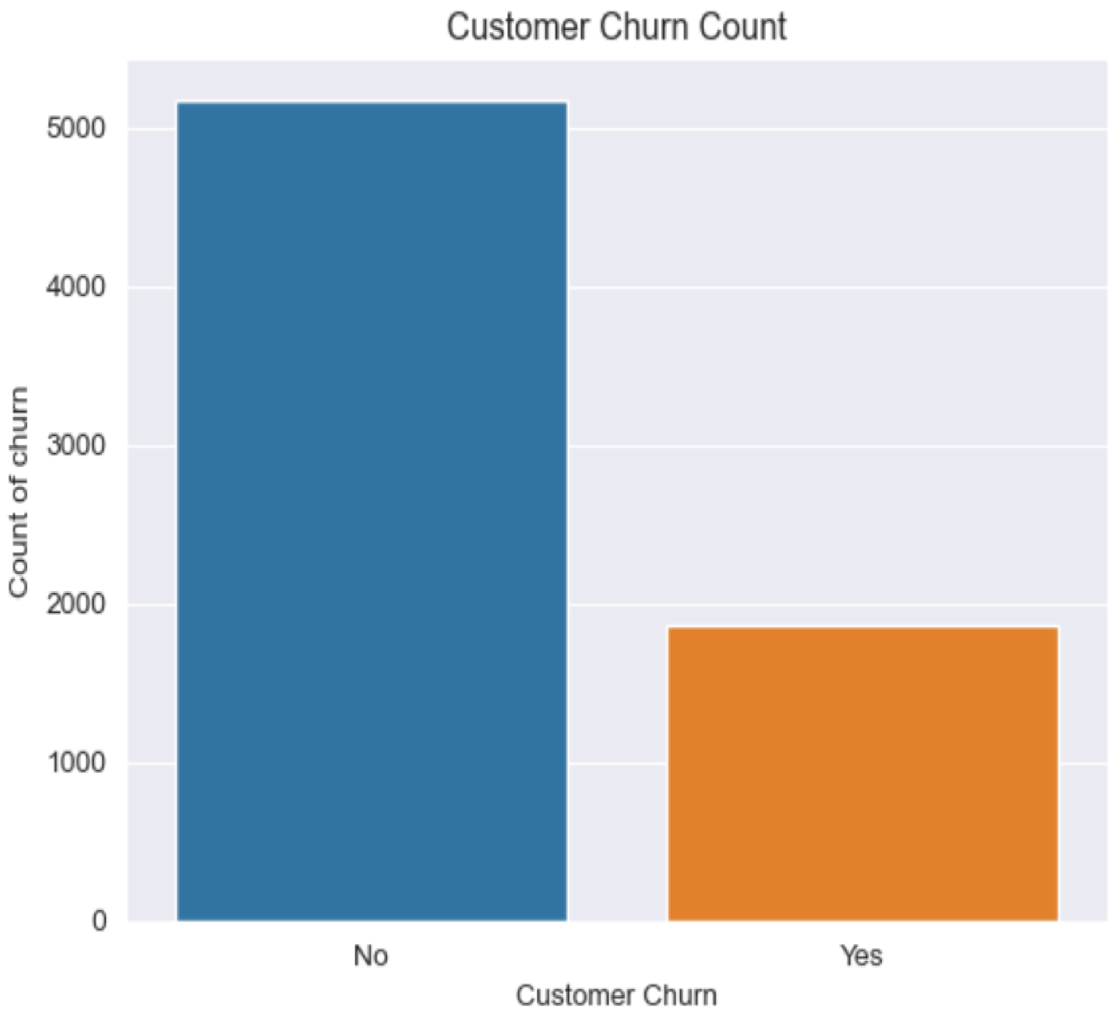
CUSTOMER ATTRITION ANALYSIS AND PREDICTION SYSTEM

AIMS AND OBJECTIVES

- Analysis of relevant service features.
- Analysis of the effect of service features on customer attrition(Churn).
- Create a machine learning model to accurately predict customer attrition(Churn).
- Reduction in customer attrition (Churn).
- Enhance customer loyalty.
- Maintain a competitive edge in the industry.

EXPLORATORY DATA ANALYSIS

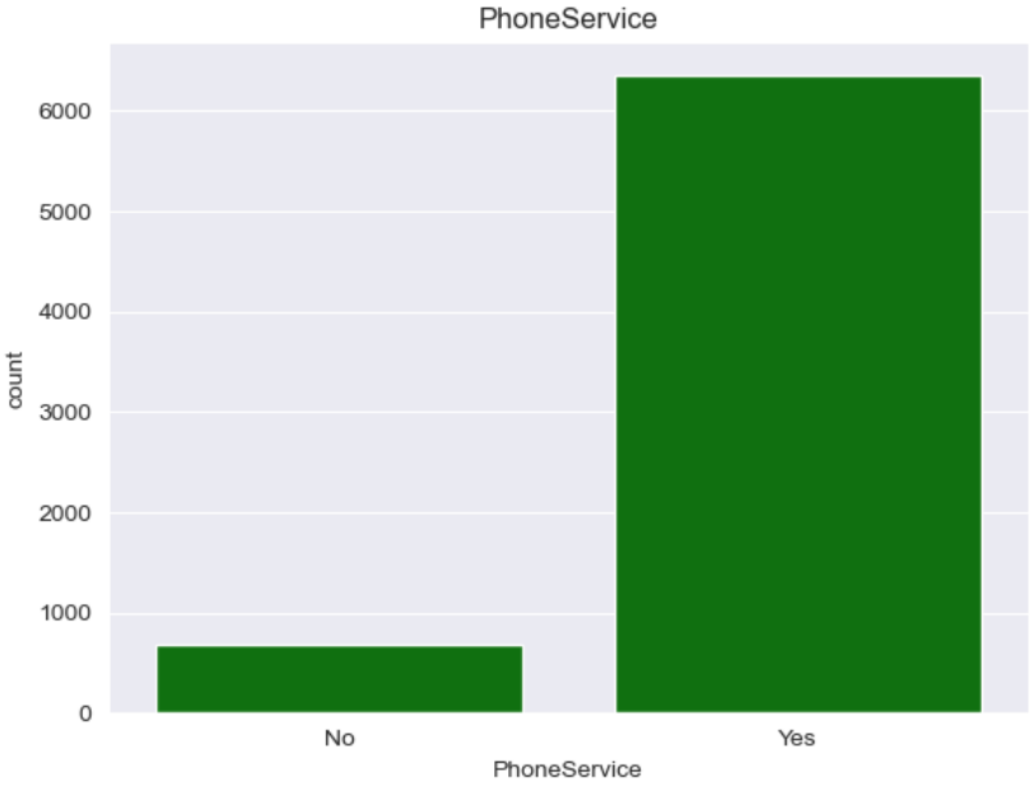
UNIVARIATE ANALYSIS

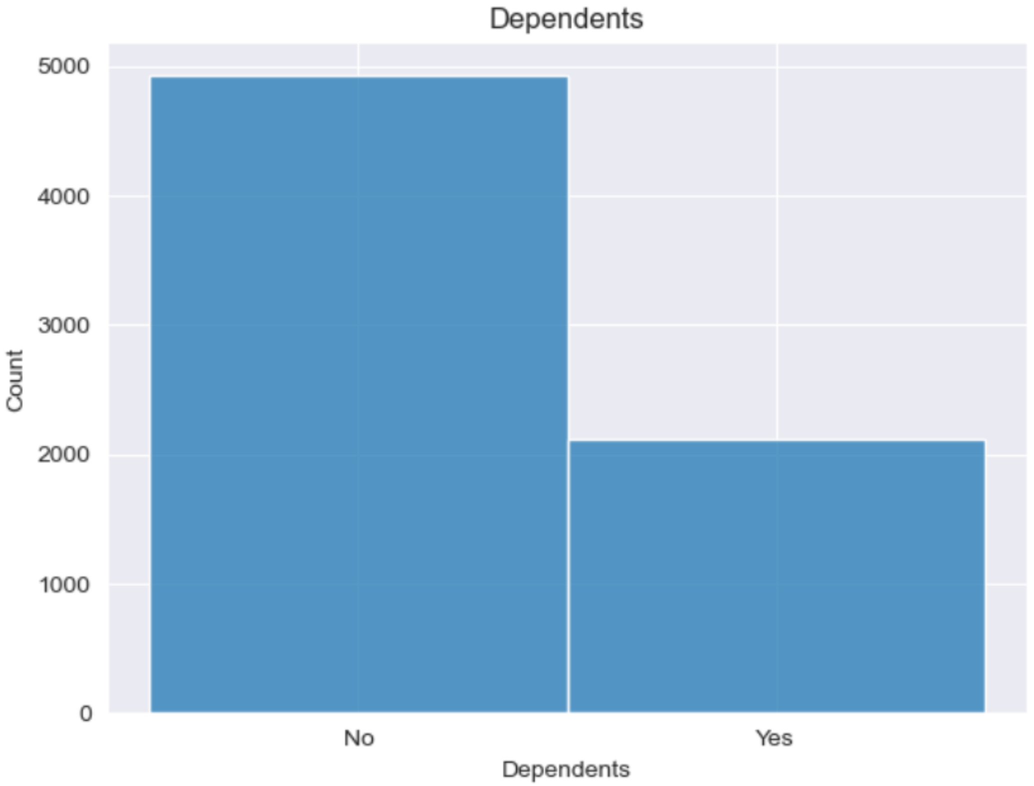


Customer churn count

- The visual shows the count of customer Churn.
- Yes count - 37,380
- No count - 103,480

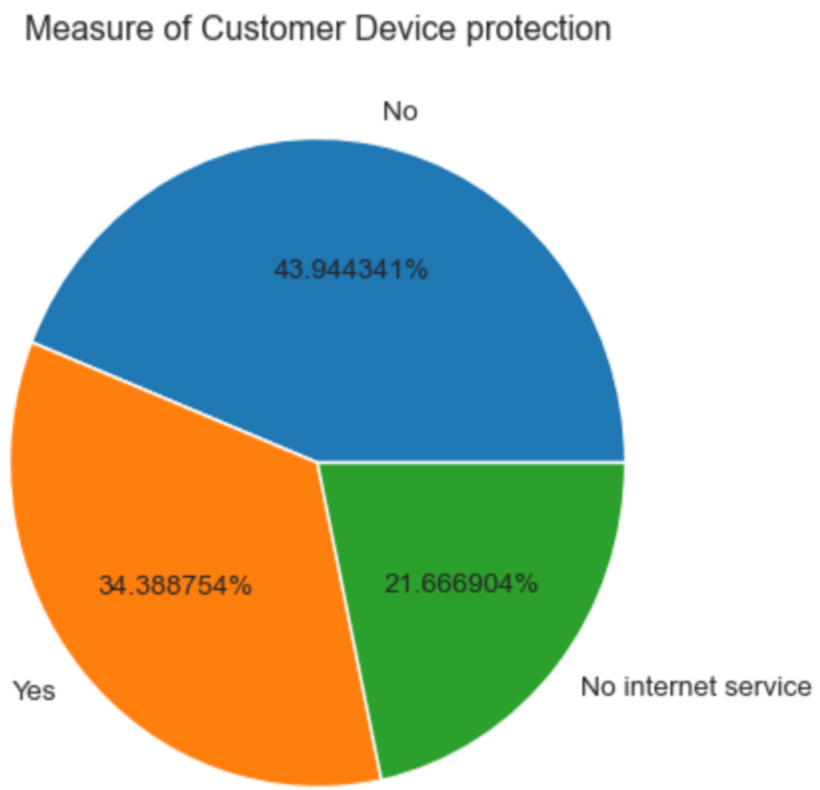
COUNT OF CUSTOMERS USING PHONE SERVICES AND CUSTOMERS WITH DEPENDENTS

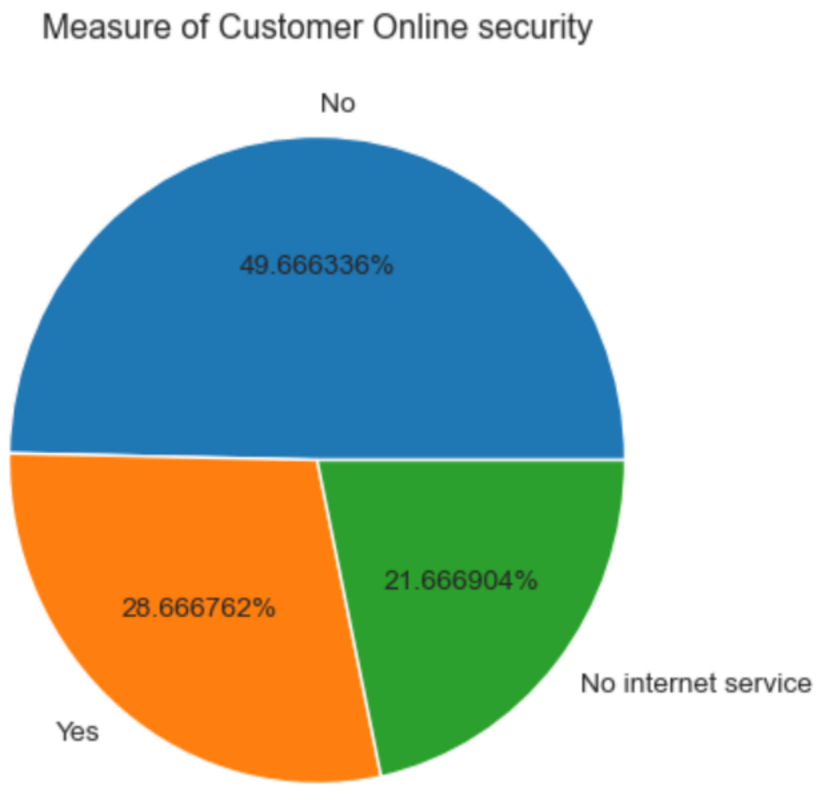




Exploratory data analysis

- Customer device protection
- Customer online security

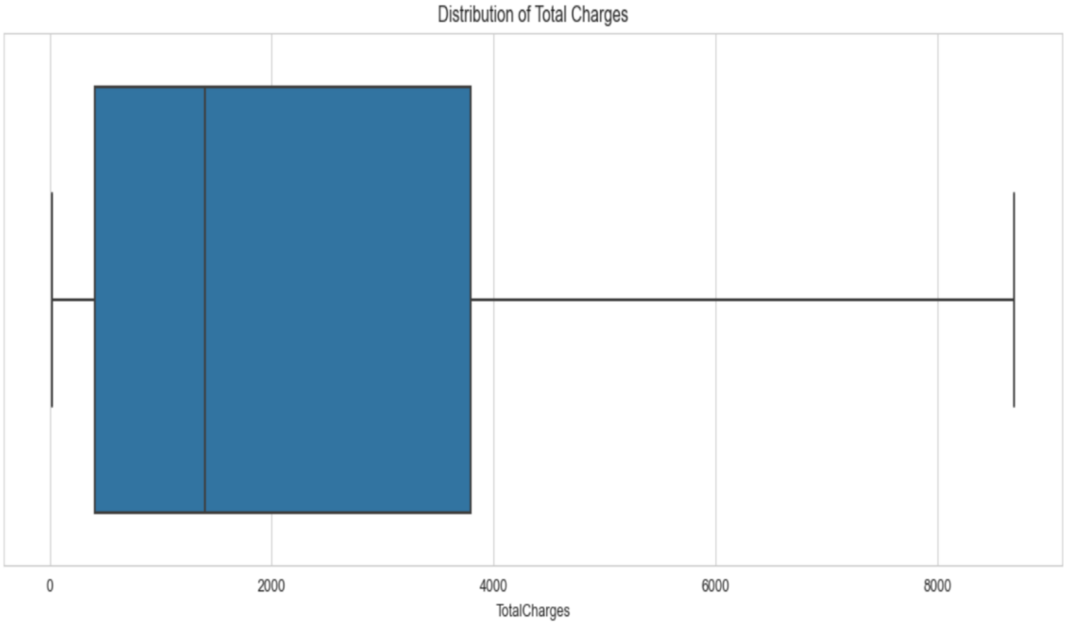




Exploratory Data Analysis of Numerical Variables

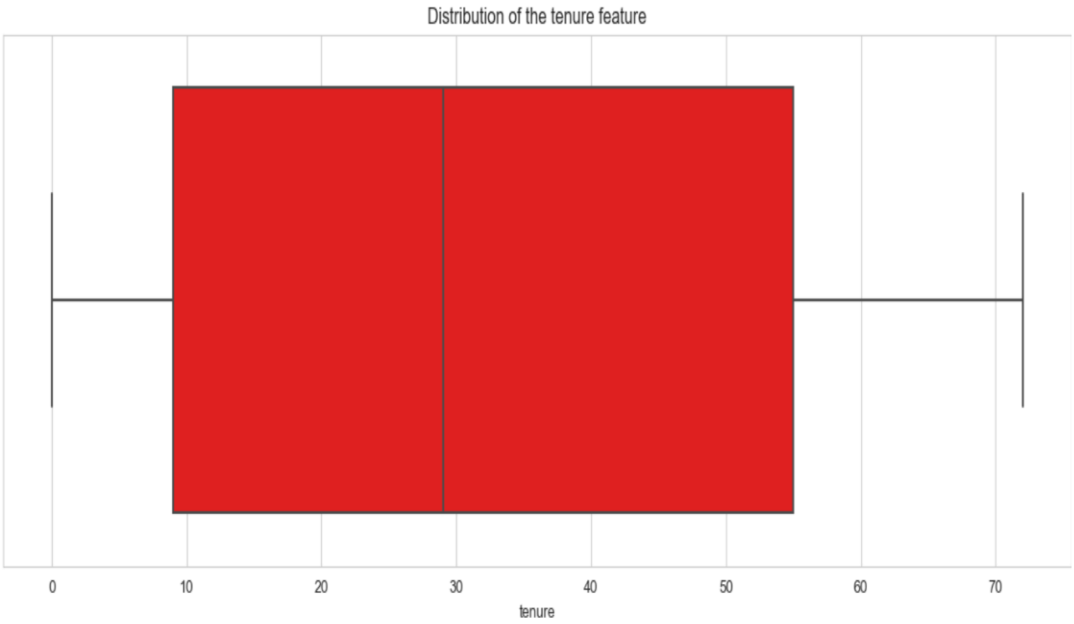
DISTRIBUTION OF TOTAL CHARGES

This distribution indicates no outliers in total charges



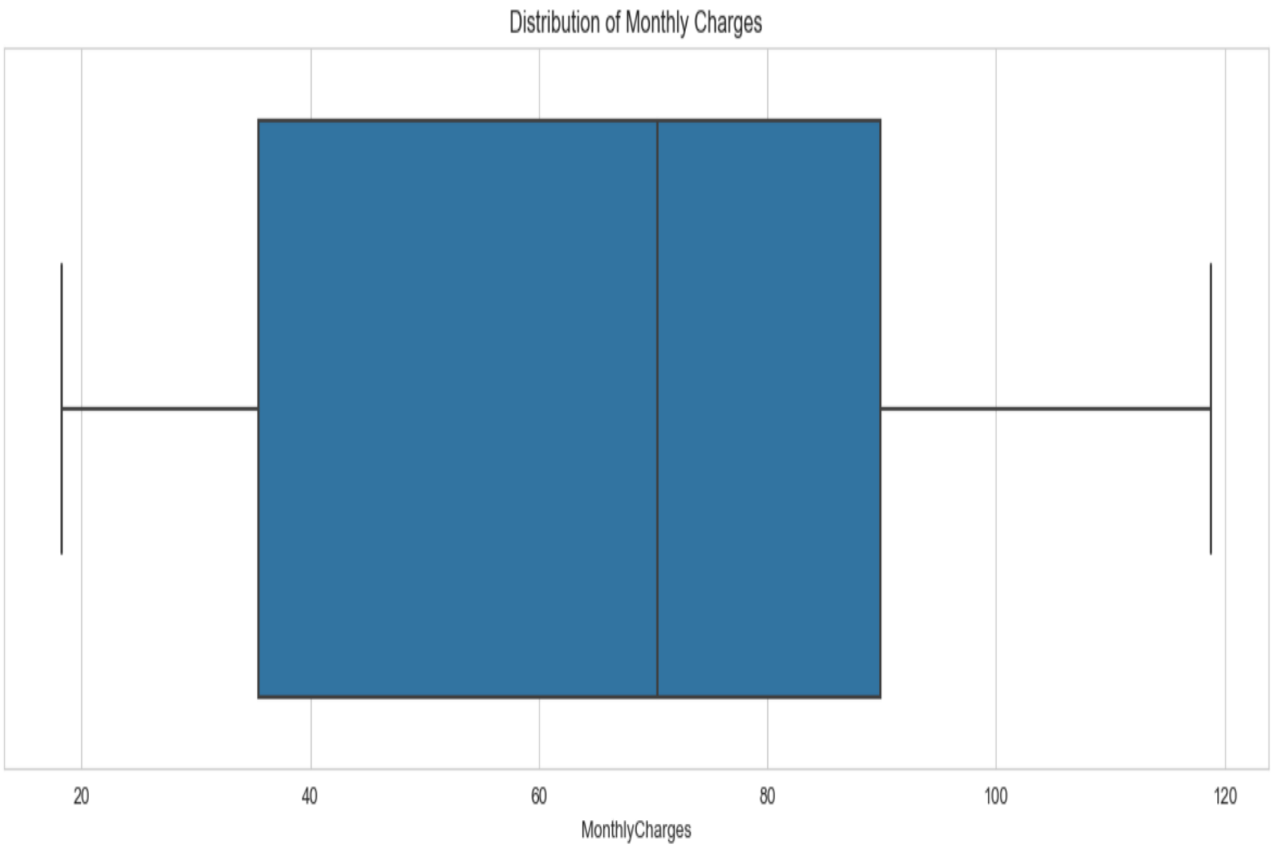
DISTRIBUTION OF TENURE FEATURE

This distribution indicates no outliers in tenure



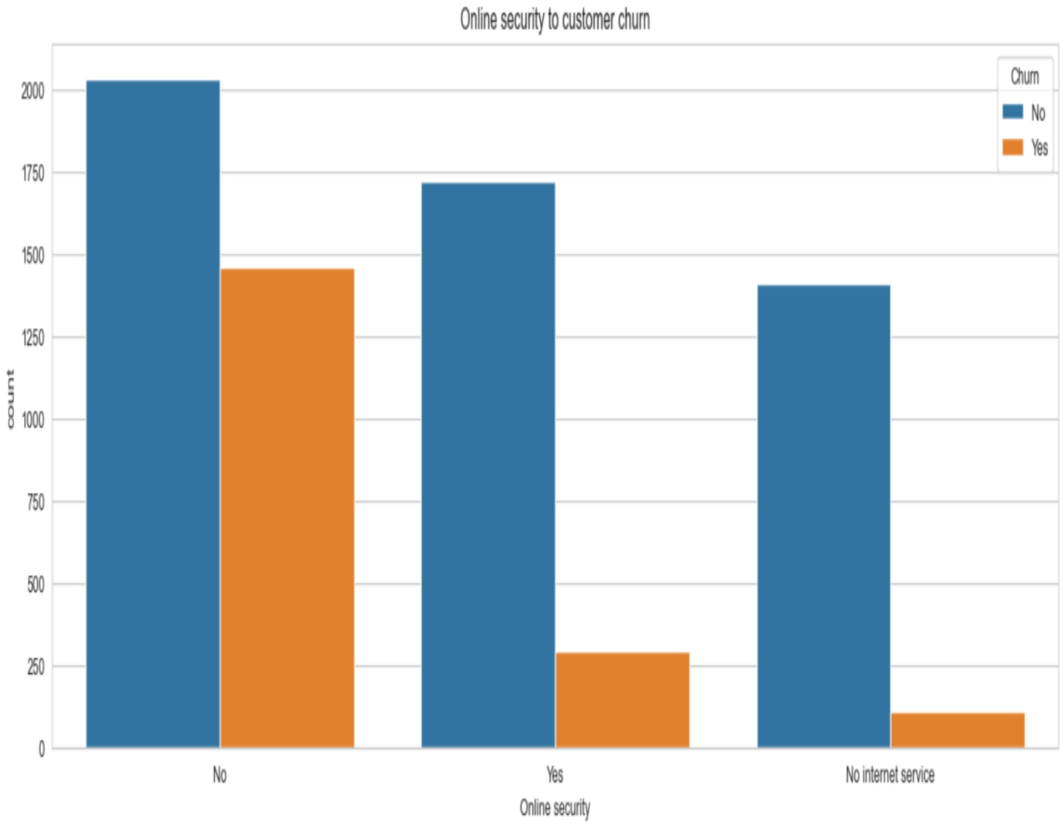
Distribution of Monthly Charges

- This distribution indicates no outliers in the data for monthly charges



BIVARIATE ANALYSIS:

The target variable, customer churn will be plotted against relevant features deduced from univariate analysis.

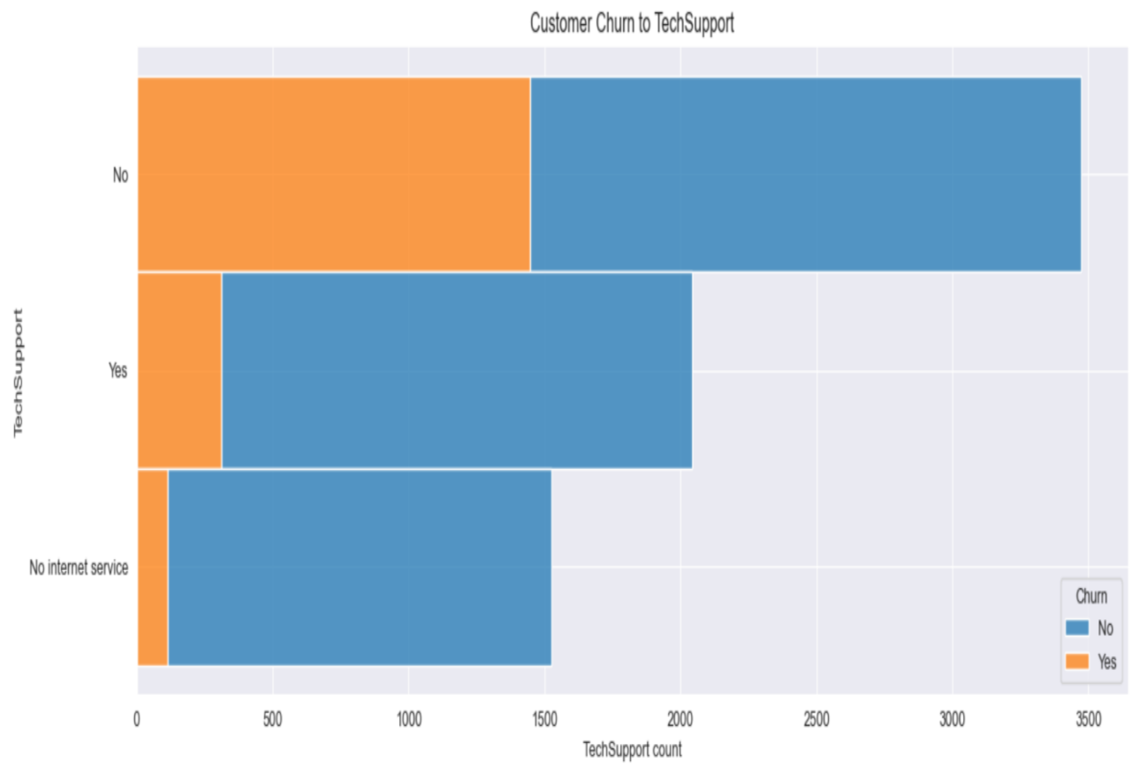


Online security to customer churn

- The visual shows the high count of customers without online security with a the highest count of customer churn, indicating the online security feature is relevant to customer churn.

BIVARIATE ANALYSIS:

The target variable, customer churn will be plotted against relevant features deduced from univariate analysis.



Tech support to customer churn

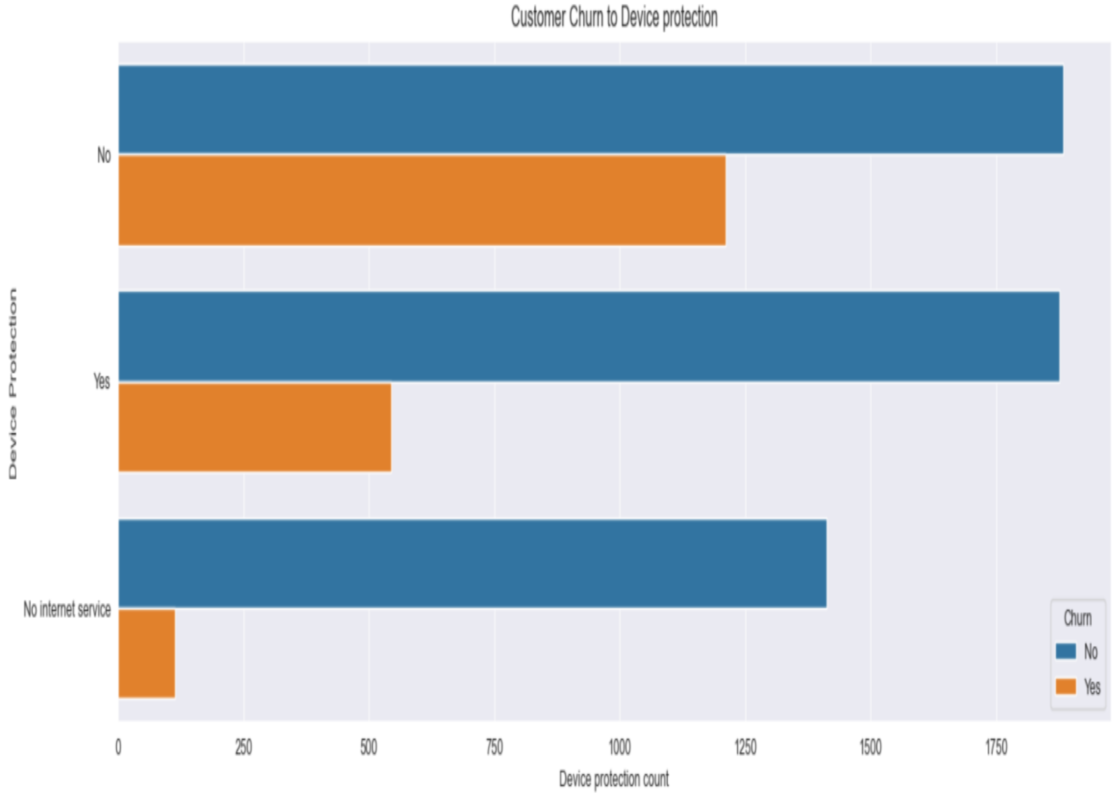
- The visual shows the high count of customers without tech support with a the highest count of customer churn, indicating technical support feature is relevant to customer churn.

BIVARIATE ANALYSIS:

The target variable, customer churn will be plotted against relevant features deduced from univariate analysis.

Device protection to customer churn

The visual indicates the high count of customers without device protection. It also indicates the high count of customer churn from the category of customers without device protection.

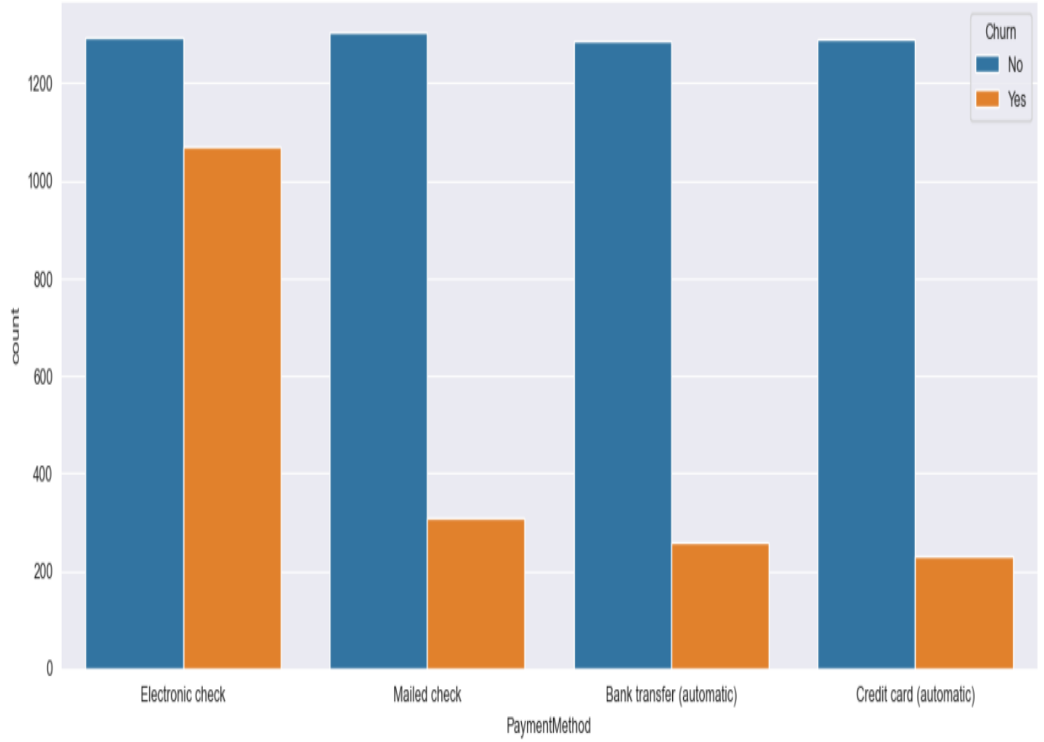


BIVARIATE ANALYSIS:

The target variable, customer churn will be plotted against relevant features deduced from univariate analysis.

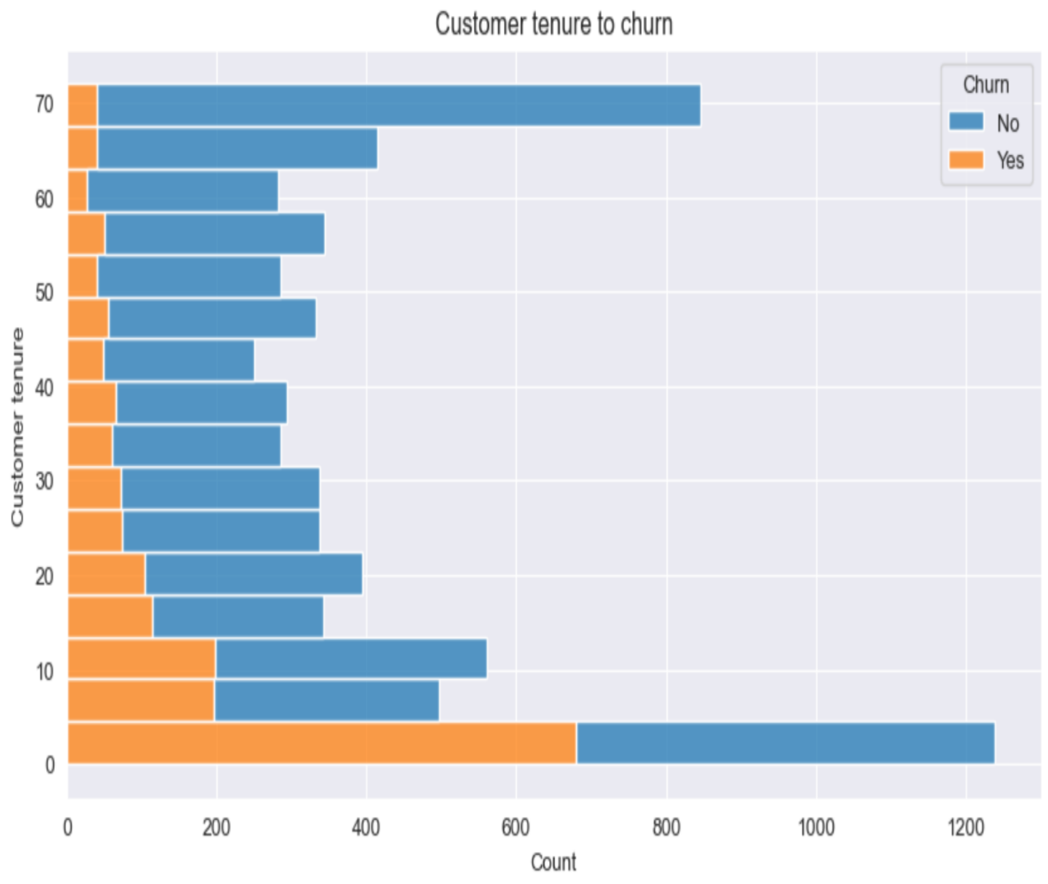
Payment Methods to customer churn

The visual indicates the high count of customers that use electronic check have the highest customer churn count. In as much as the other payment methods have equal no counts.



BIVARIATE ANALYSIS:

The target variable, customer churn will be plotted against relevant features deduced from univariate analysis.

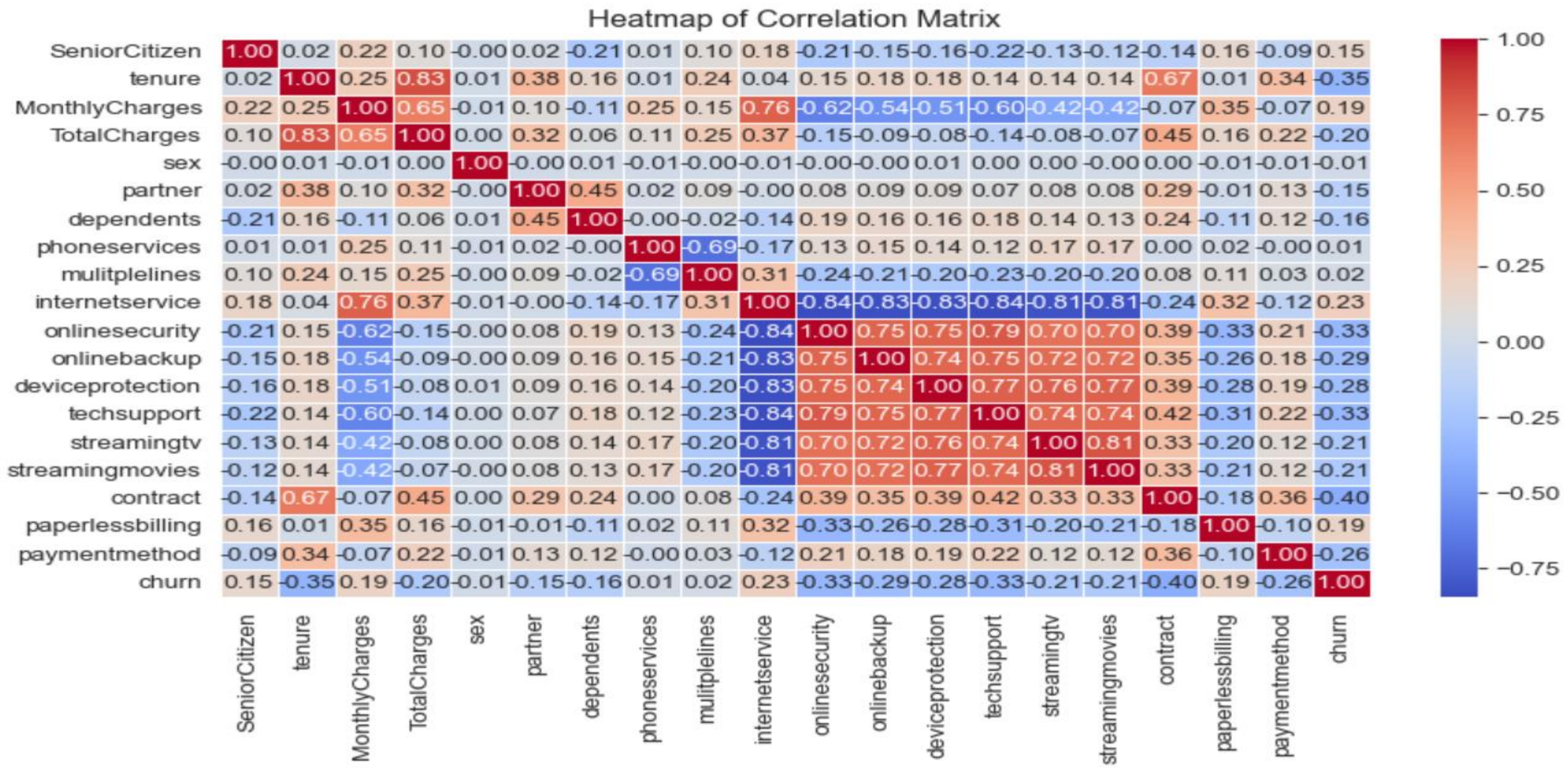


Tenure to customer churn count

- The visual indicates the highest counts of customer churn with regards to tenure is from the category of new customers

MULTIVARIATE EXPLORATORY ANALYSIS

Heat map of all data features in relation to the target variable customer churn



HEAT MAP OBSERVATIONS

There is positive correlation between the target variable ‘churn' and the following features:

- paperless bill
- internet service
- multiple lines
- phone services
- monthly service
- senior citizen

This positive correlation indicates the increase in the target variable(churn) a these features increase as well, highlighting a direct relationship between the target variable(church) and the features with positive relation stated above.

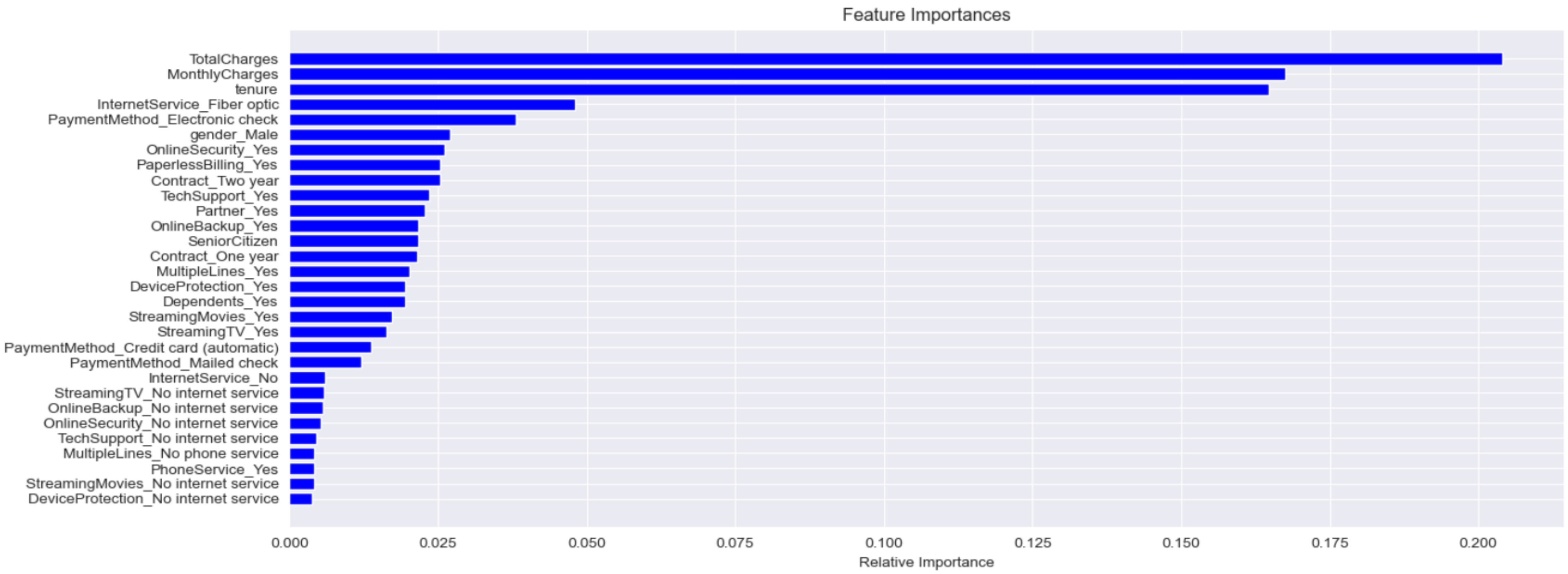
HEAT MAP OBSERVATIONS CONT.

There is negative Correlation between the target variable ‘churn' and the following features:

- payment method
- contract
- streaming movie
- streaming tv
- tech support
- device protection
- online backup
- online security
- dependent
- partner
- sex
- total charges
- tenure

This negative correlation indicates when there is increase in these variables the target variable(churn)decrease, highlighting an inverse relationship between the target variable(church) and the features with negative relation.

FEATURE IMPORTANCE CHART



BUILDING MACHINE LEARNING MODELS

The following are machine learning models to explore:

- Logistic Regression model
- Decision Trees Classifier model
- Random Forest Classifier model
- Support Vector Machines model

Logistic Regression model

ACCURCY = 0.81

	precision	recall	f1-score	support
0	0.86	0.89	0.88	3649
1	0.65	0.57	0.61	1258
accuracy			0.81	4907
macro avg	0.75	0.73	0.74	4907
weighted avg	0.80	0.81	0.81	4907

Confusion Matrix insights

- The precision of 86% and a recall of 89% the f1 score is 88% achieving a good balance between recall and precision for class o
- The precision of 65% and a recall of 57% the f1 score is 61% achieving a average performance balance between recall and precision for class 1
- Overall, the weighted average metrics (precision, recall, and F1-score) provide a balanced view considering the class imbalances.

Decision Trees Classifier model

ACCURCY = 0.721

	precision	recall	f1-score	support
0	0.83	0.79	0.81	3649
1	0.46	0.53	0.49	1258
accuracy			0.72	4907
macro avg	0.65	0.66	0.65	4907
weighted avg	0.73	0.72	0.73	4907

Confusion Matrix insights

- The precision of 83% and a recall of 79% the f1 score is 81% achieving a good balance between recall and precision for class o
- The precision of 46% and a recall of 53% the f1 score is 49% achieving a average performance balance between recall and precision for class 1
- Overall, the weighted average metrics (precision, recall, and F1-score) provide a balanced view considering the class imbalances.

Random Forest Classifier Model

ACCURCY = 0.791

	precision	recall	f1-score	support
0	0.84	0.89	0.86	3649
1	0.61	0.52	0.56	1258
accuracy			0.79	4907
macro avg	0.73	0.70	0.71	4907
weighted avg	0.78	0.79	0.79	4907

Confusion Matrix insights

- The precision of 83% and a recall of 79% the f1 score is 81% achieving a good balance between recall and precision for class o
- The precision of 46% and a recall of 53% the f1 score is 49% achieving a average performance balance between recall and precision for class 1
- Overall, the weighted average metrics (precision, recall, and F1-score) provide a balanced view considering the class imbalances.

Support Vector Machine Model

ACCURCY = 0.801

	precision	recall	f1-score	support
0	0.84	0.91	0.87	3649
1	0.65	0.50	0.56	1258
accuracy			0.80	4907
macro avg	0.74	0.70	0.72	4907
weighted avg	0.79	0.80	0.79	4907

Confusion Matrix insights

- The precision of 84% and a recall of 91% the f1 score is 87% achieving a good balance between recall and precision for class o
- The precision of 65% and a recall of 50% the f1 score is 56% achieving a average performance balance between recall and precision for class 1
- Overall, the weighted average metrics (precision, recall, and F1-score) provide a balanced view considering the class imbalances.

MODELS ACCURACY TRADE-OFFS

Accuracy trade off between true positives and true negatives

- Logistic Regression model - 3261 (true positive) and 717 (true negative)
- Decision Trees Classifier model - 2875 (true positive) and 664 (true negative)
- Random Forest Classifier model - 3234 (true positive) and 651 (true negative)
- Support Vector Machines model - 3304 (true positive) and 627 (true negative)

Accuracy trade off between false positives and false negatives

- Logistic Regression model - 541 (false positive) and 388 (false negative)
- Decision Trees Classifier model - 594 (false positive) and 774 (false negative)
- Random Forest Classifier model - 607 (false positive) and 415 (false negative)
- Support Vector Machines model - 631 (false positive) and 342 (false negative)

Model insight

- The logistic regression model is the most efficient with the highest accuracy

- The matrixes that are most important for the problem are the false positives and false negative matrixes as these reduce the accuracy of the model.

- Conduct hyperparameter tuning, adjust the hyperparameters of the logistic regression model to find better configuration that improves performance.

Recommendations

- The company should focus on customers without internet
- Improve their online security
- Improve their tech support
- Device protection
- Electronic check payment method should be investigated