CHURN PREDICTION

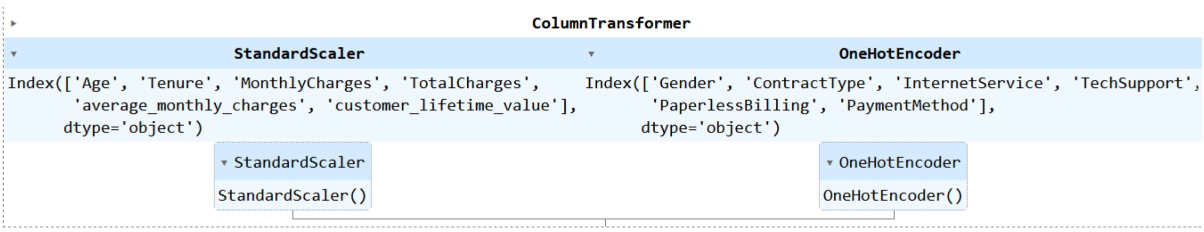
1. Exploratory Data Analysis (EDA)

1.1 Dataset Overview:

- **Features:**
  - **Age:** Integer
  - **gender:** Categorical (Female, Male)
  - **Contract:** Categorical (Month-to-month, One year, Two year)
  - **MonthlyCharges:** Float
  - **TotalCharges:** Float
  - **TechSupport:** Categorical (No, Yes)
  - **InternetService:** Categorical (DSL, Fiber optic, No)
  - **tenure:** Integer
  - **PaperlessBilling:** Categorical (No, Yes)
  - **PaymentMethod:** Categorical (Bank transfer, Credit card, Electronic check, Mail check)
  - **Churn:** Target Variable (Binary: 0 = No Churn, 1 = Churn)

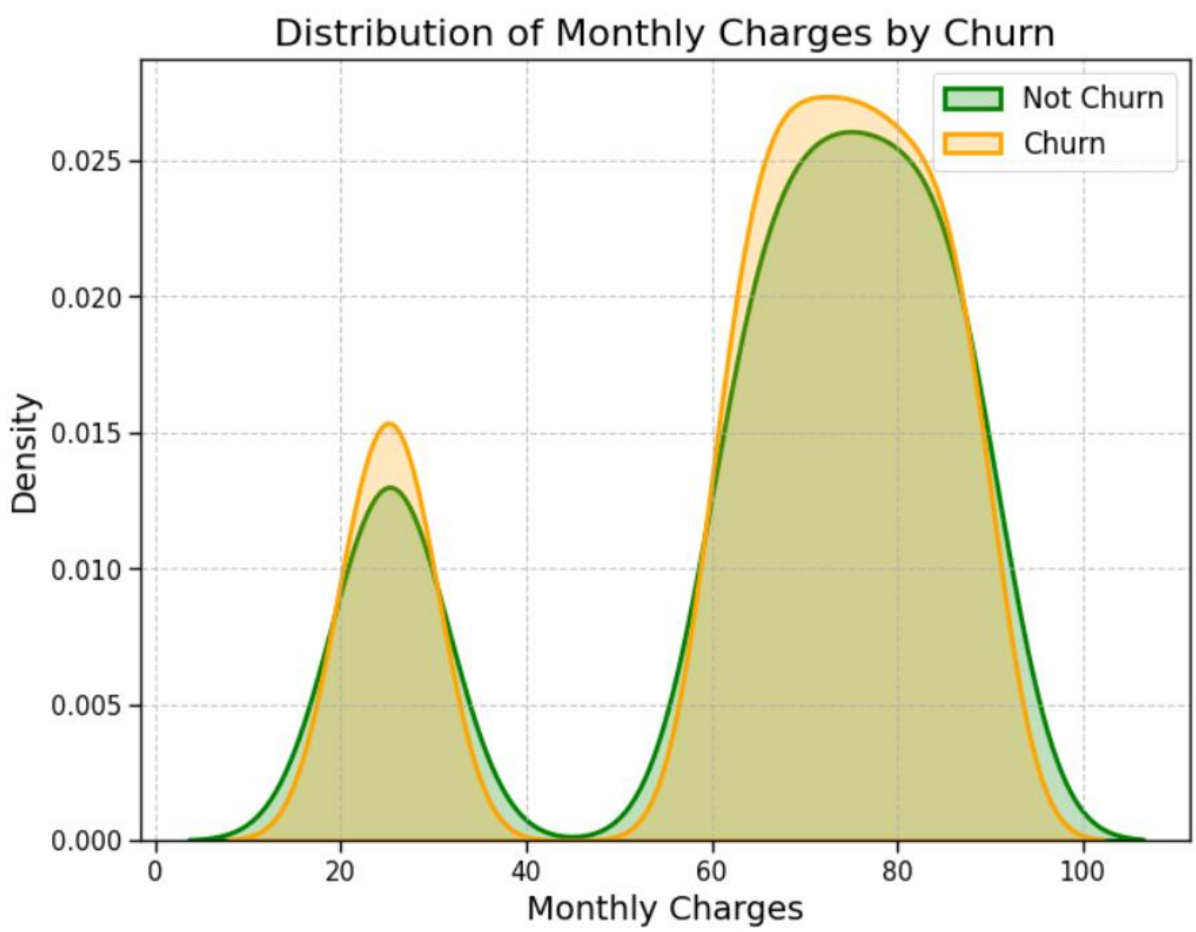
1.2 Data Cleaning and Preprocessing:

- Handling Missing Values: Replaced missing values in categorical columns with appropriate values or encoded as nan.
- Encoding Categorical Variables: Applied One hot Encoding to convert categorical variables into numerical values suitable for machine learning models.

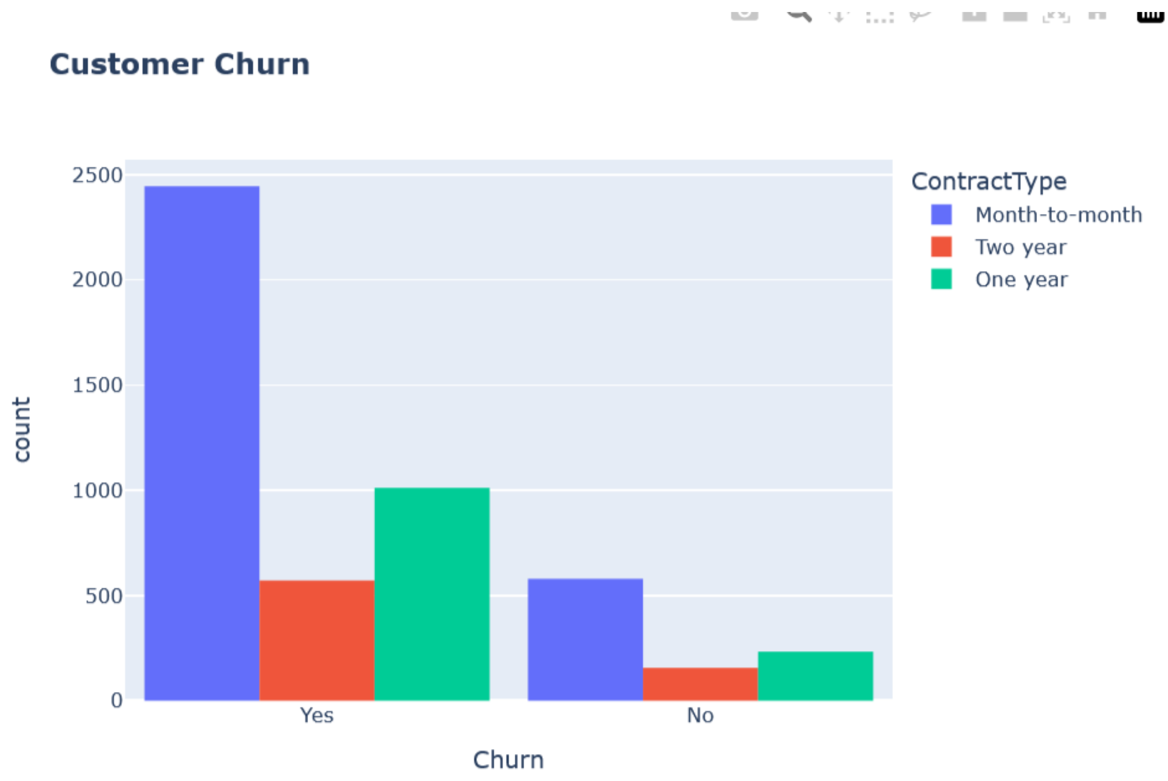


1.3 Visualization and Analysis:

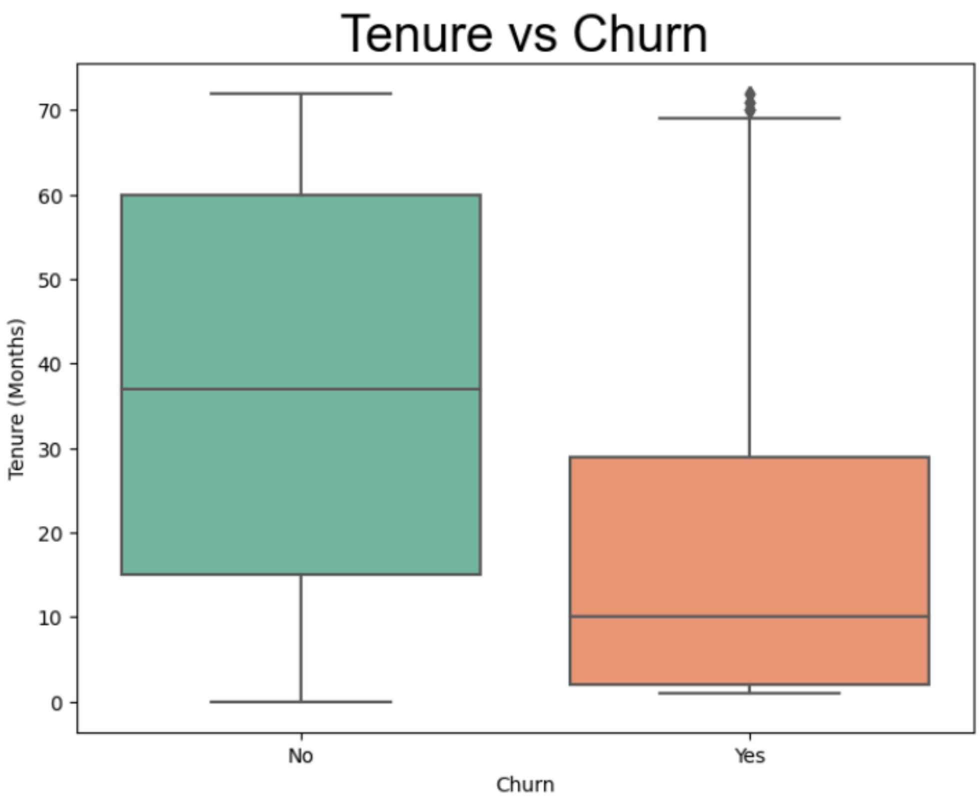
- Customer Monthly Charges Distribution w.r.t. Churn



- Churn Distribution w.r.t. Contract Type



- Tenure vs Churn



2. **Model Performance**

2.1 Models Evaluated:

- Logistic Regression
- Decision Tree Classifier
- Random Forest Classifier
- Support Vector Machine (SVM)
- K-Nearest Neighbors (KNN)
- Gradient Boosting Classifier
- CatBoostClassifier
- XGBClassifier
- ExtraTreesClassifier

2.2 Performance Metrics:

- Accuracy
- Precision
- Recall
- F1-Score
- ROC Curve
- AUC (Area Under the Curve)

2.5 Example Metrics for Logistic Regression:

- **Accuracy**: 79.98%
- **Precision**: 0.64
- **Recall**: 0.4906
- **F1-Score**: 0.55

3. Model Interpretation and Insights

3.1 Confusion Matrix: [[693 71]

[135 130]]

A confusion matrix was generated to visualize the performance of the classification models. It showed how many true positives, true negatives, false positives, and false negatives were identified.

3.2 Recommendations:

- **Customer Retention**: Focus on customers with high MonthlyCharges and low tenure as they are more likely to churn.
- **Service Improvement**: Enhance support and service for customers using Fiber optic internet service and those without TechSupport.

4. Conclusion

The analysis effectively identified key factors influencing customer churn and evaluated various machine learning models to predict churn with a focus on maximizing accuracy and interpretability. The best-performing models provide actionable insights for improving customer retention strategies and refining service offerings. Further analysis and model improvements can enhance predictive accuracy and customer insights.

5.Final OutPut