

CUSTOMER CHURN PREDICTION

Contract

- ☐ Month-to-month
- ☐ One year
- ☐ Two year

Internet Service

- ☐ DSL
- ☐ Fiber optic
- ☐ No

64.76

AvgMonthlyCharges_All

74.44

AvgMonthlyCharges_Churned



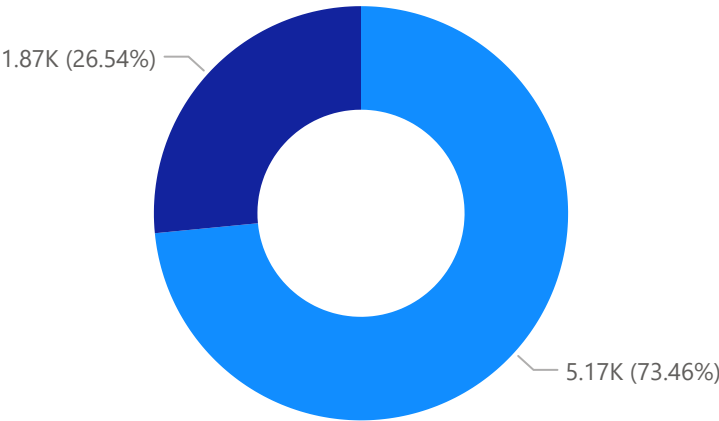
7043

Total Customers

Churn Distribution

(0-No, 1-Yes)

Churn 0 1



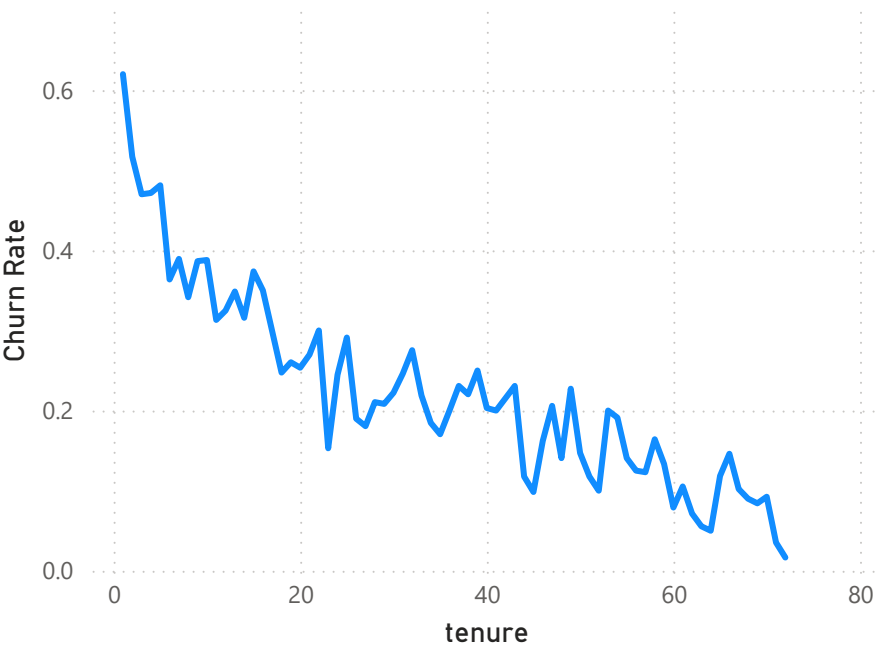
1869

Churned Customers

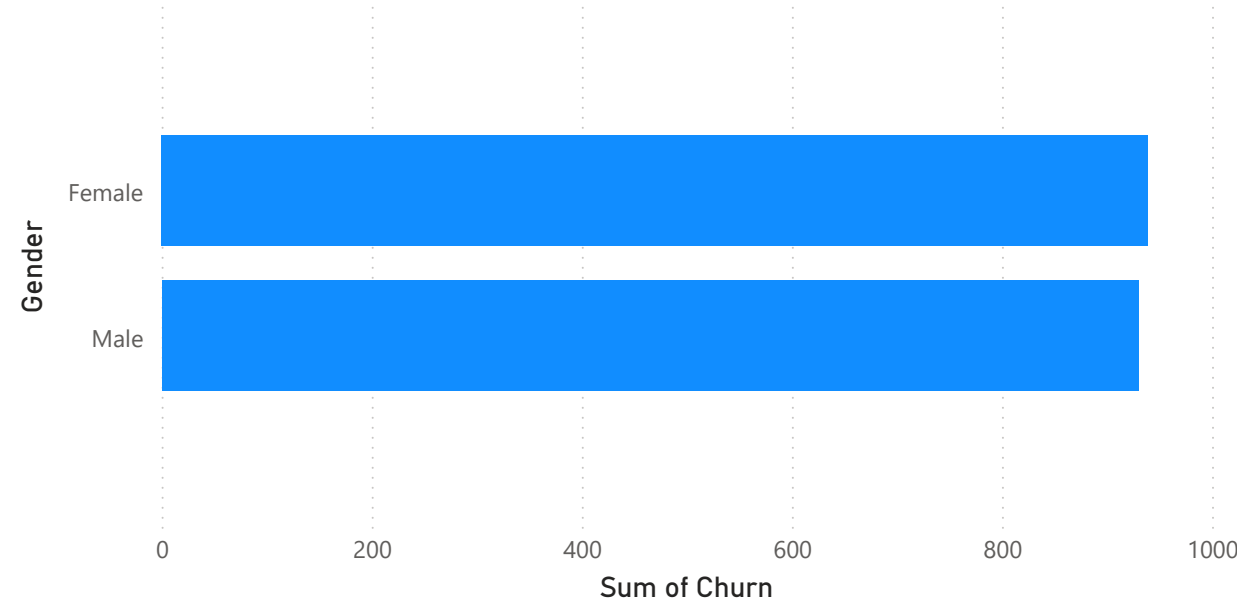
Total Churn by Contract Type



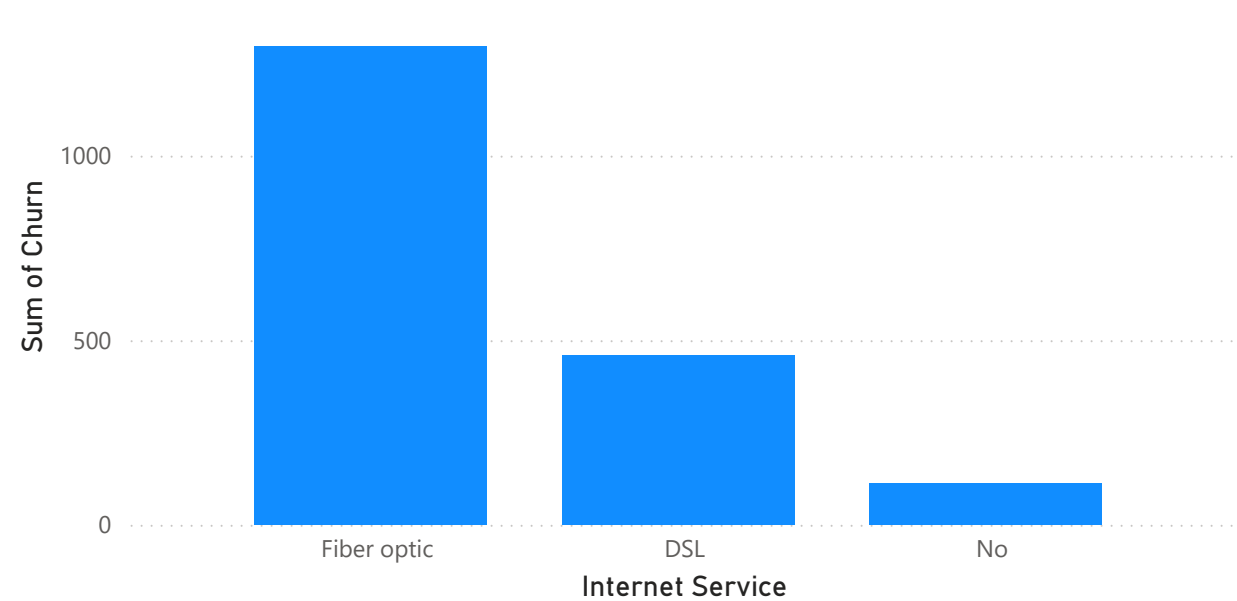
Churn Rate by Tenure



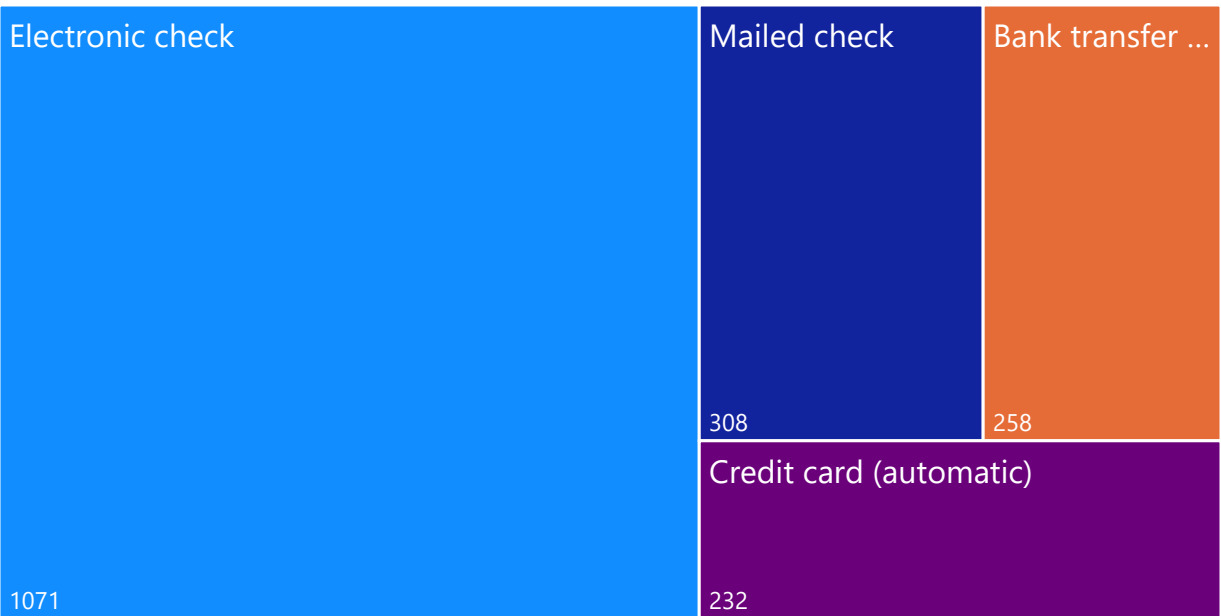
Churn Distribution by Gender



Churn Distribution by Internet Service



Churn Tree Map by Payment Method



7043

Total Customers

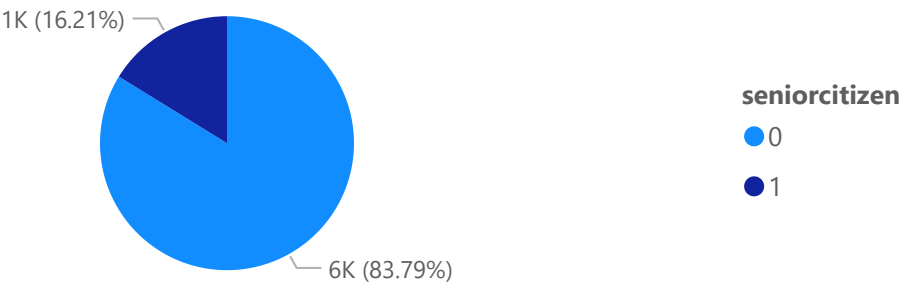
1869

Churned Customers

0.27

Churn Rate

Senior Citizen Distribution



**Average Churn Probability
(Random Forest)**

25.96%

**Average Churn Probability
(Logistic Regression)**

26.70%

**Average Churn Probability
(XGBoost)**

26.03%

Actual Churn Rate

26.54%

**High Risk Count Prediction
(Random Forest)**

172

**High Risk Count Prediction
(Logistic Regression)**

212

**High Risk Count Prediction
(XGBoost)**

260

Actual Churn Count

374

**80% data was used to train the 3 models
and 20% data (1409) customer data was
used to test these models and they yielded
the above results**

**Total Customers used for testing
models based on which we have
received above results**

1409

**also do checkout the
kaggle notebook for more
analytics on the ML
models**

FINAL CONCLUSION

We trained three machine learning models to predict customer churn: **Logistic Regression**, **Random Forest**, and **XGBoost**.

◆ **Logistic Regression**

- Accuracy: **80%**
- Strengths: Simple, interpretable, fast to train.
- Weakness: Recall for churn is lower, meaning it misses some actual churners.

◆ **Random Forest**

- Accuracy: **80.4%**
- Strengths: Handles non-linear relationships well, very strong at predicting "No Churn".
- Weakness: Recall for churn (49%) is weaker, indicating it struggles to capture all churn cases.

◆ **XGBoost**

- Accuracy: **78%**
- Strengths: Good precision, flexible model with high potential after tuning.
- Weakness: Slightly worse accuracy out-of-the-box compared to Logistic Regression and Random Forest.

✅ **Conclusion**

- Both **Logistic Regression** and **Random Forest** provide competitive results around 80% accuracy.
- For **business use cases**, recall for churn is more important since companies want to **catch as many potential churners as possible**.
- Logistic Regression is a great **baseline model** due to its balance and interpretability.
- Random Forest can be further tuned to improve recall, while XGBoost could be optimized with hyperparameter tuning.