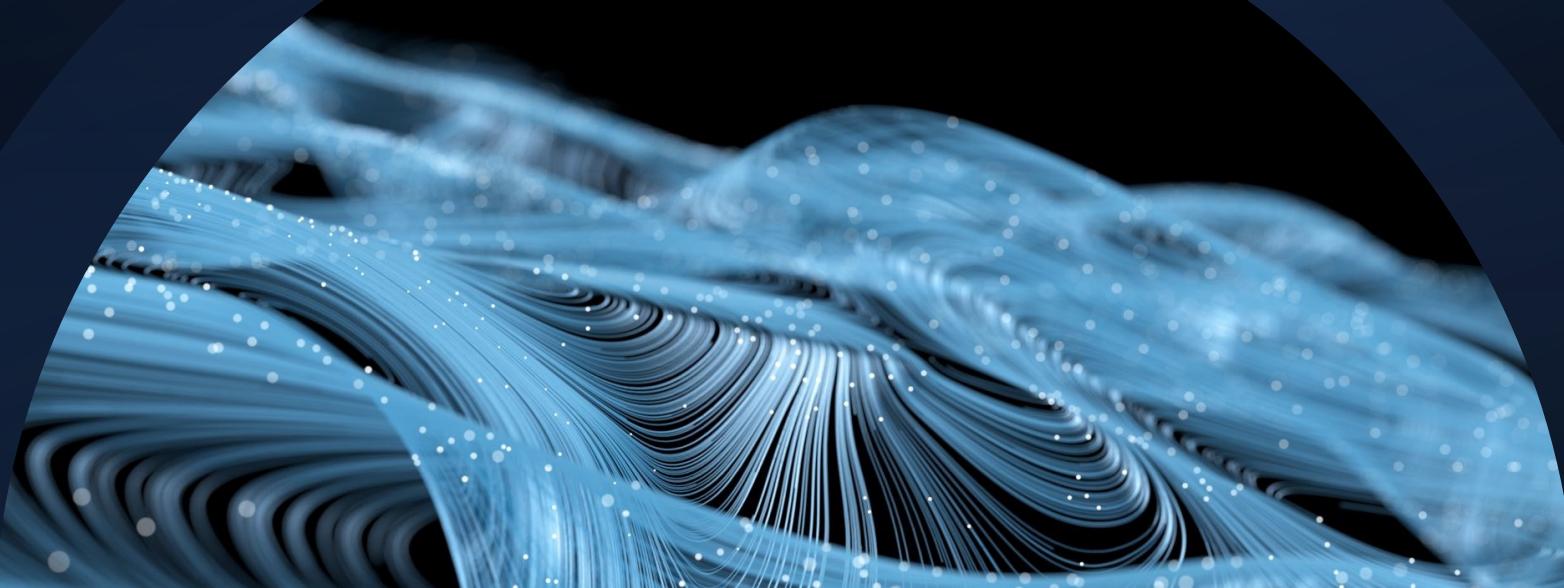


# Predicting Customer Churn for Telco

Mariya Graff



# Contents

**Methodology**

**Key Features**

**Feature Engineering**

**Baseline Model**

**Model Iterations & Final Model**

**Takeaways**

**Next steps**

# 1. Methodology

## Data

- Telco customer churn

Customer churn data for a fictional telecommunications company with 7043 observations and 33 quantitative and qualitative features from IBM.

<https://community.ibm.com/accelerators/catalog/content/Telco-customer-churn>

- Urban/Rural classifications

Six-level urban-rural classification scheme for U.S. counties and county-equivalent entities from National Center for Health Statistics (NCHS).

[https://www.cdc.gov/nchs/data\\_access/urban\\_rural.htm](https://www.cdc.gov/nchs/data_access/urban_rural.htm)

## Tools

- Data cleaning and analysis: Pandas, Numpy
- Modeling: scikit-learn, xgboost, imbalanced-learn
- Visualizations: Matplotlib, Seaborn, Tableau

## 2. Key Features



**Customer**



**Demographic**



**Geographic**



**Service**



**Billing**

**Customer ID**

**Tenure in Months**

**Gender**

**Senior**

**Partner**

**Dependents**

**Country**

**State**

**City**

**Zipcode**

**Lat-Long**

**Phone Service**

**Multiple Lines**

**Internet Service**

**Online Security**

**Online Backup**

**Device Protection**

**Tech Support**

**Streaming TV**

**Streaming Movies**

**Contract**

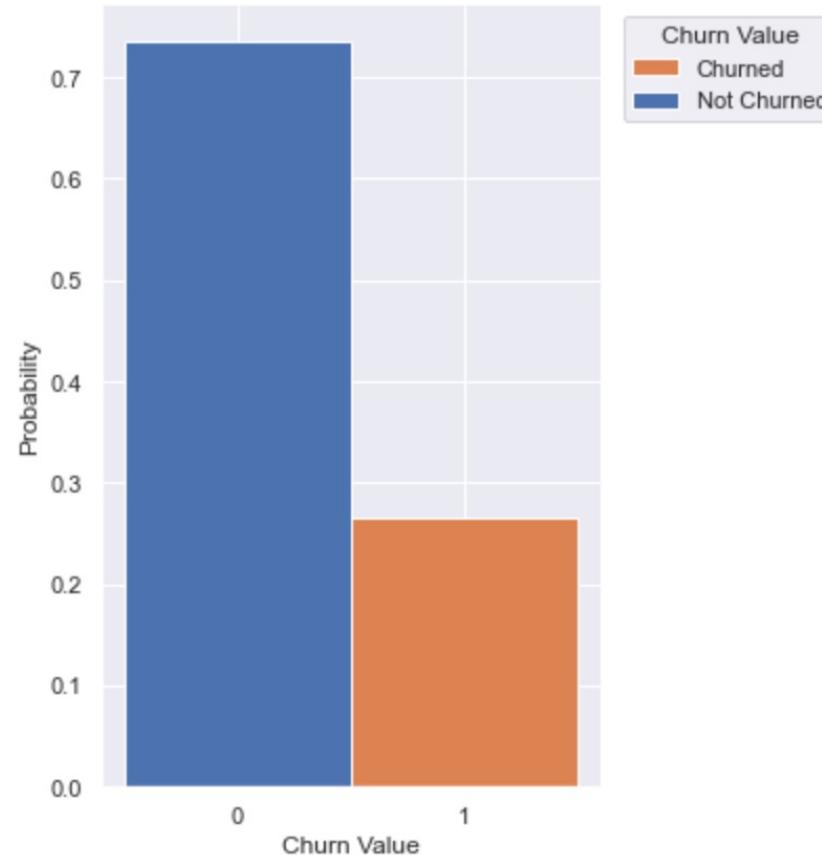
**Paperless Billing**

**Payment Method**

**Monthly Charges**

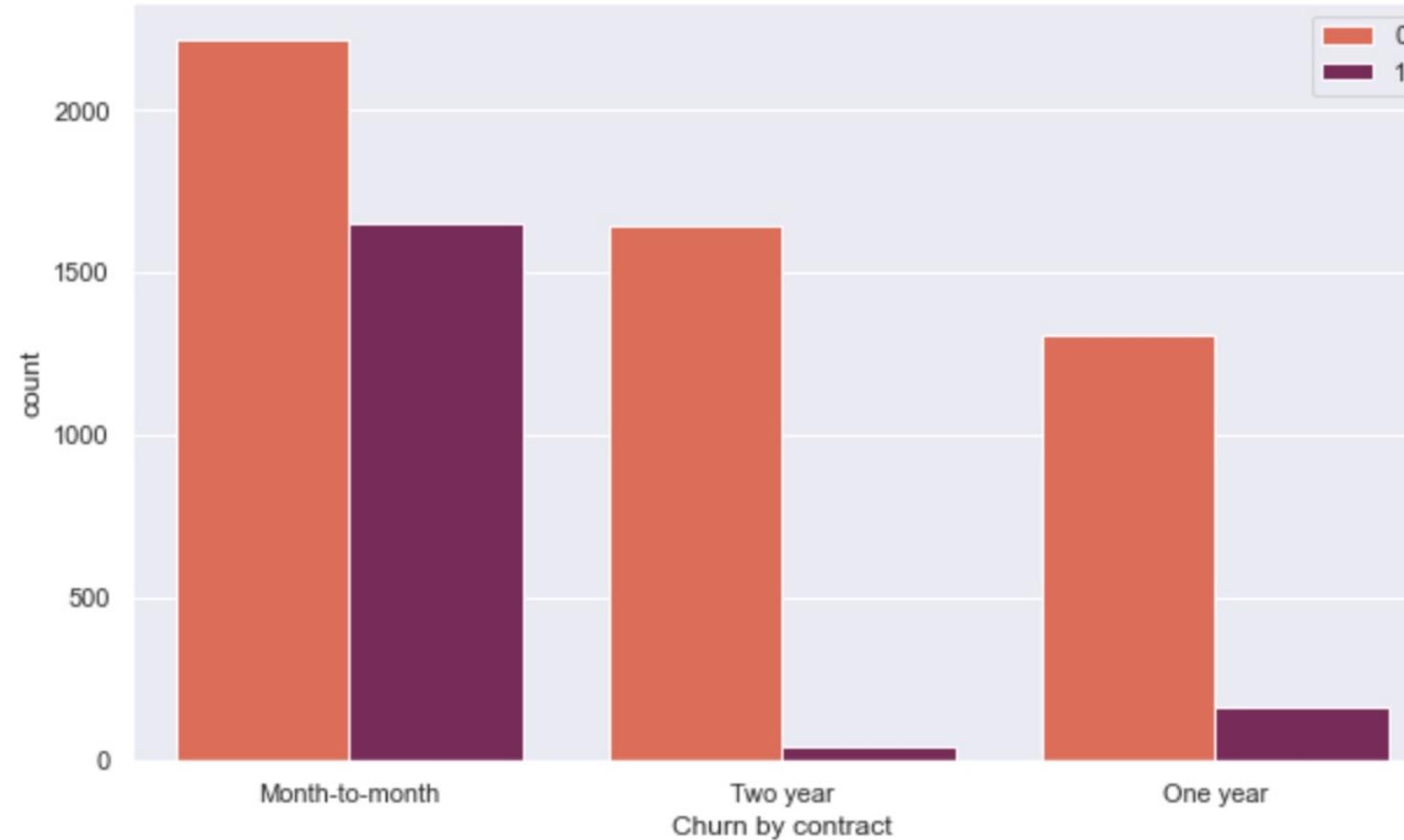
**Total Charges**

## 2. Key Features



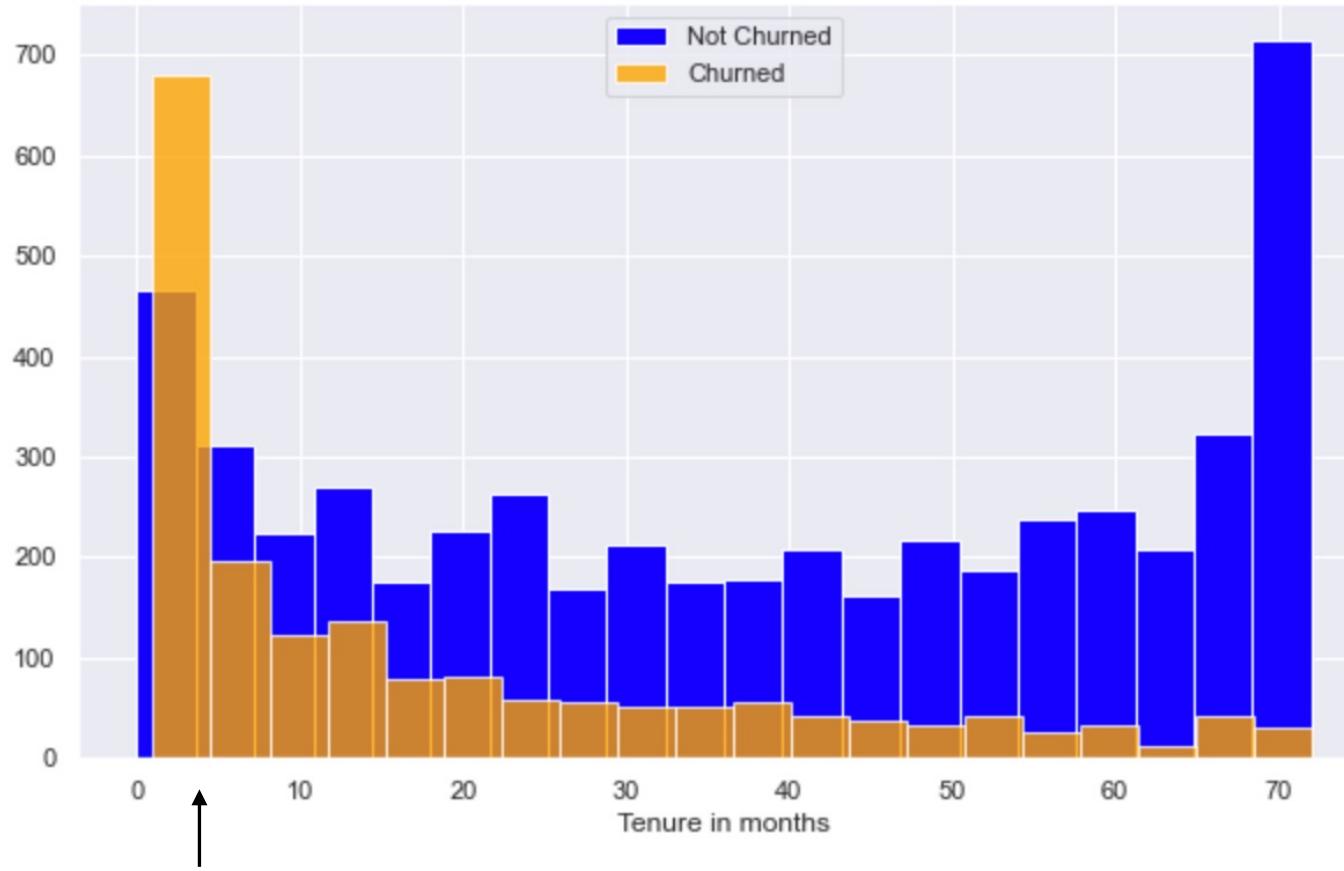
Churned is ~26% of target

## 2. Key Features



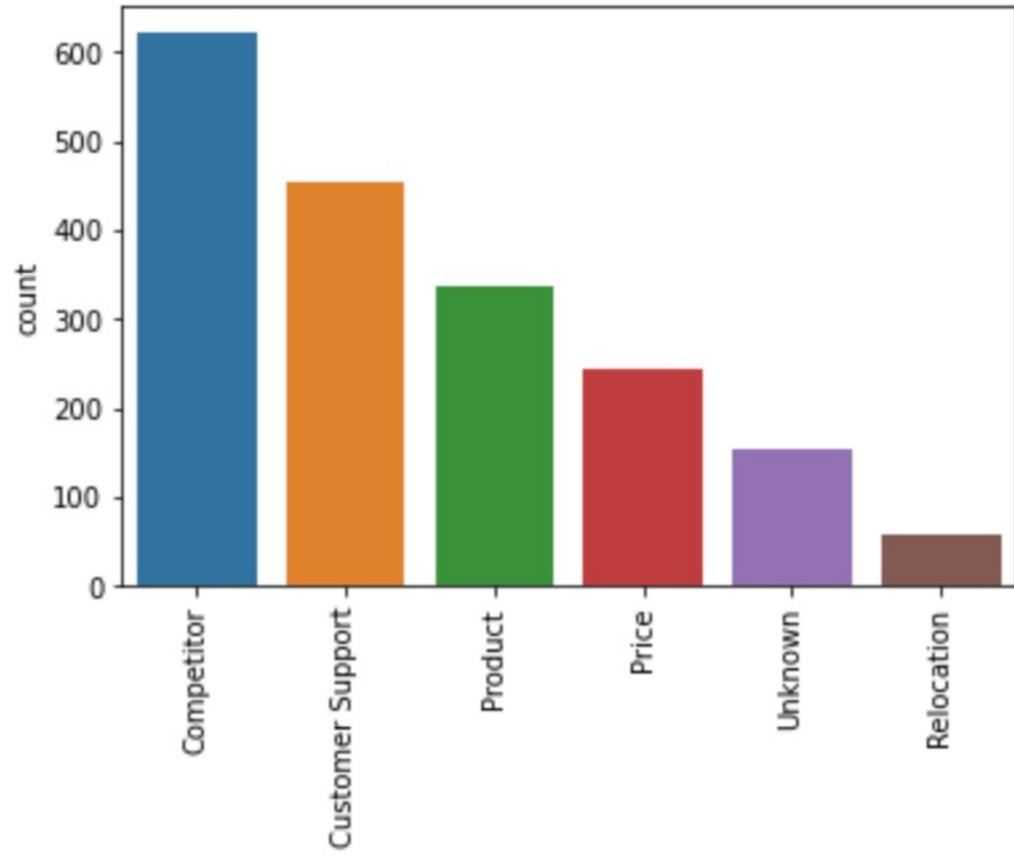
Monthly contracts have the highest number of churned customers

## 2. Key Features



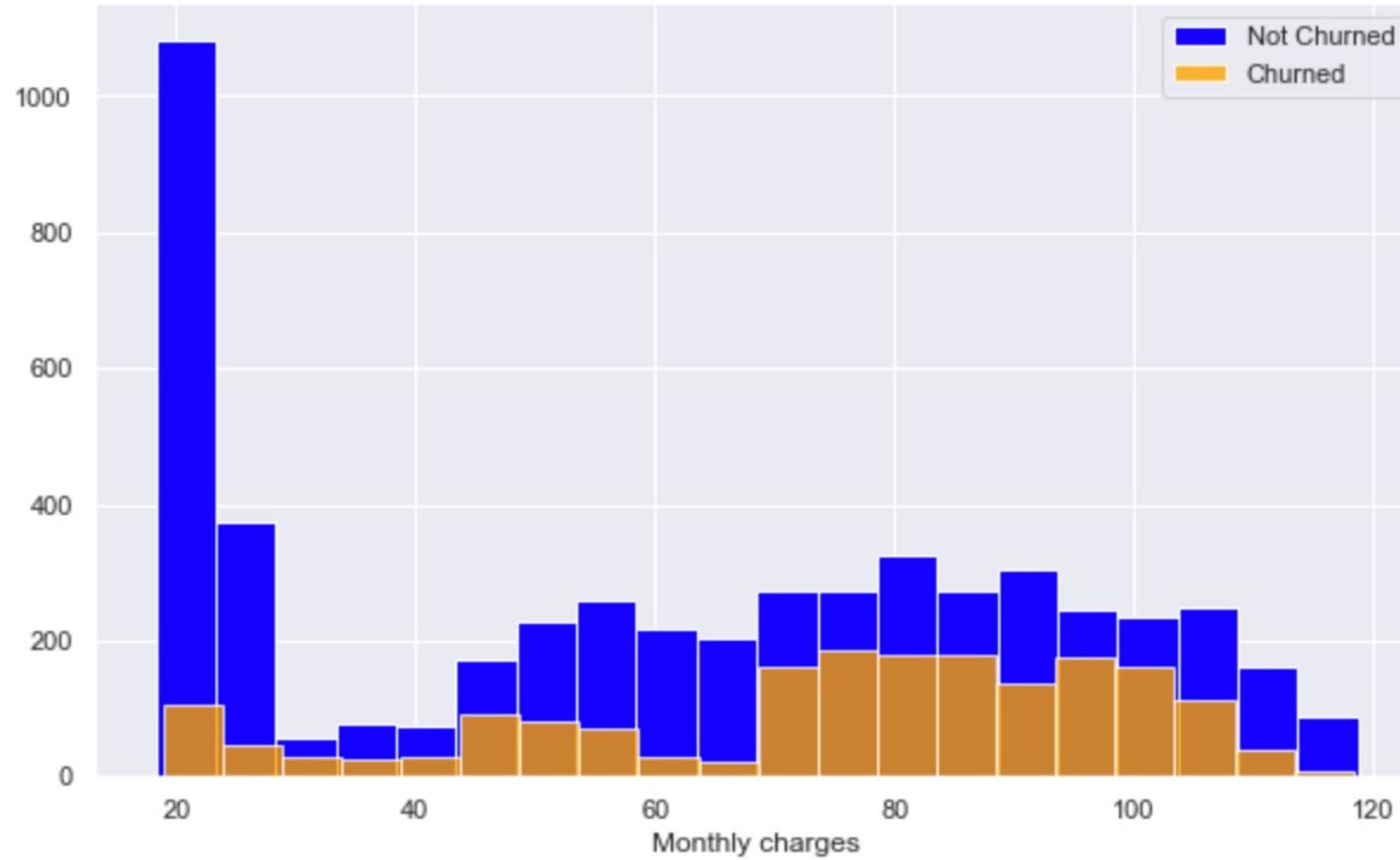
Most churn happens within the first 5 months

## 2. Key Features



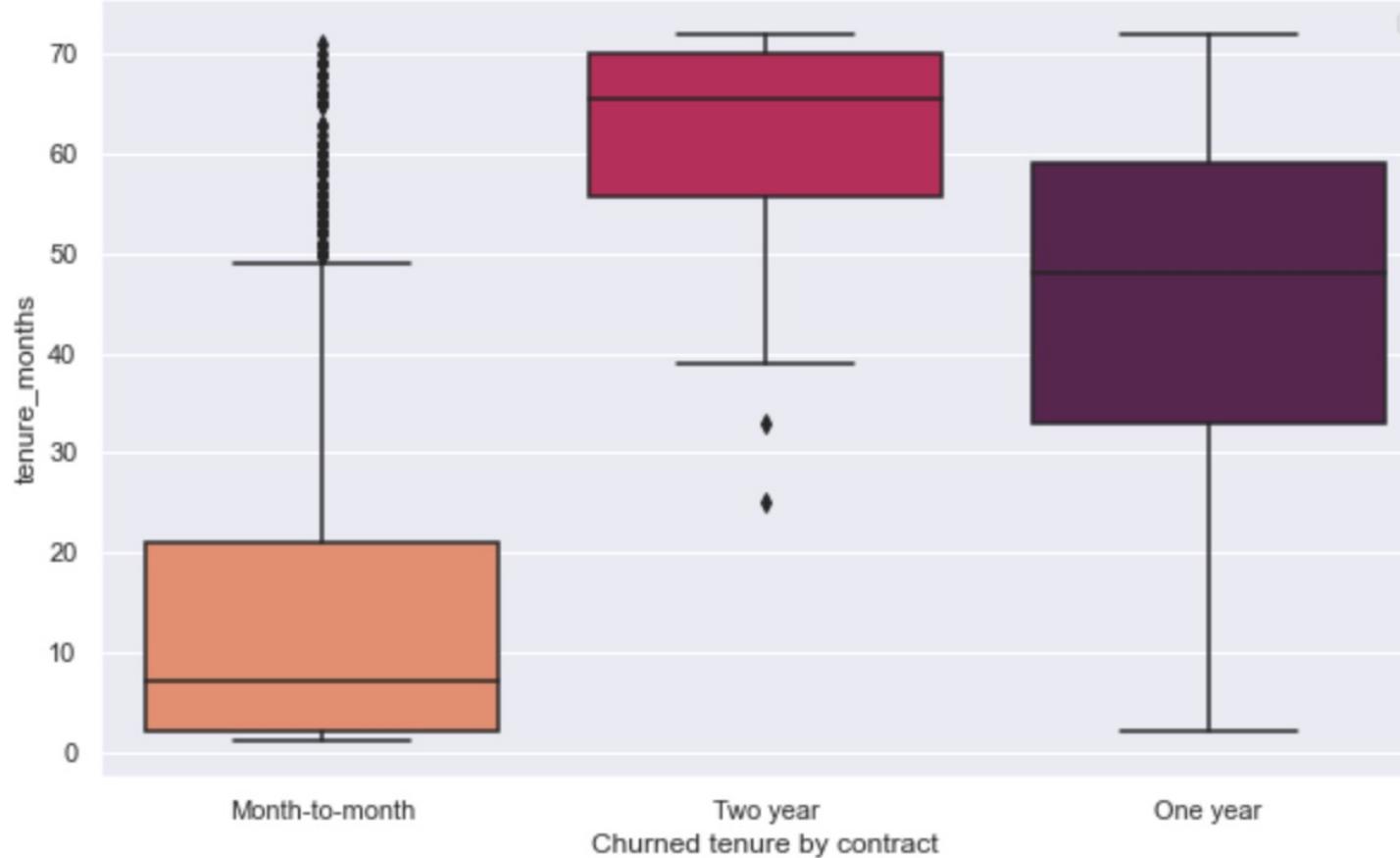
The highest reported reason for churn is competitor-related

## 2. Key Features



The higher the monthly chargers, the higher the churn

## 2. Key Features



Median tenure of churned customers by contract:

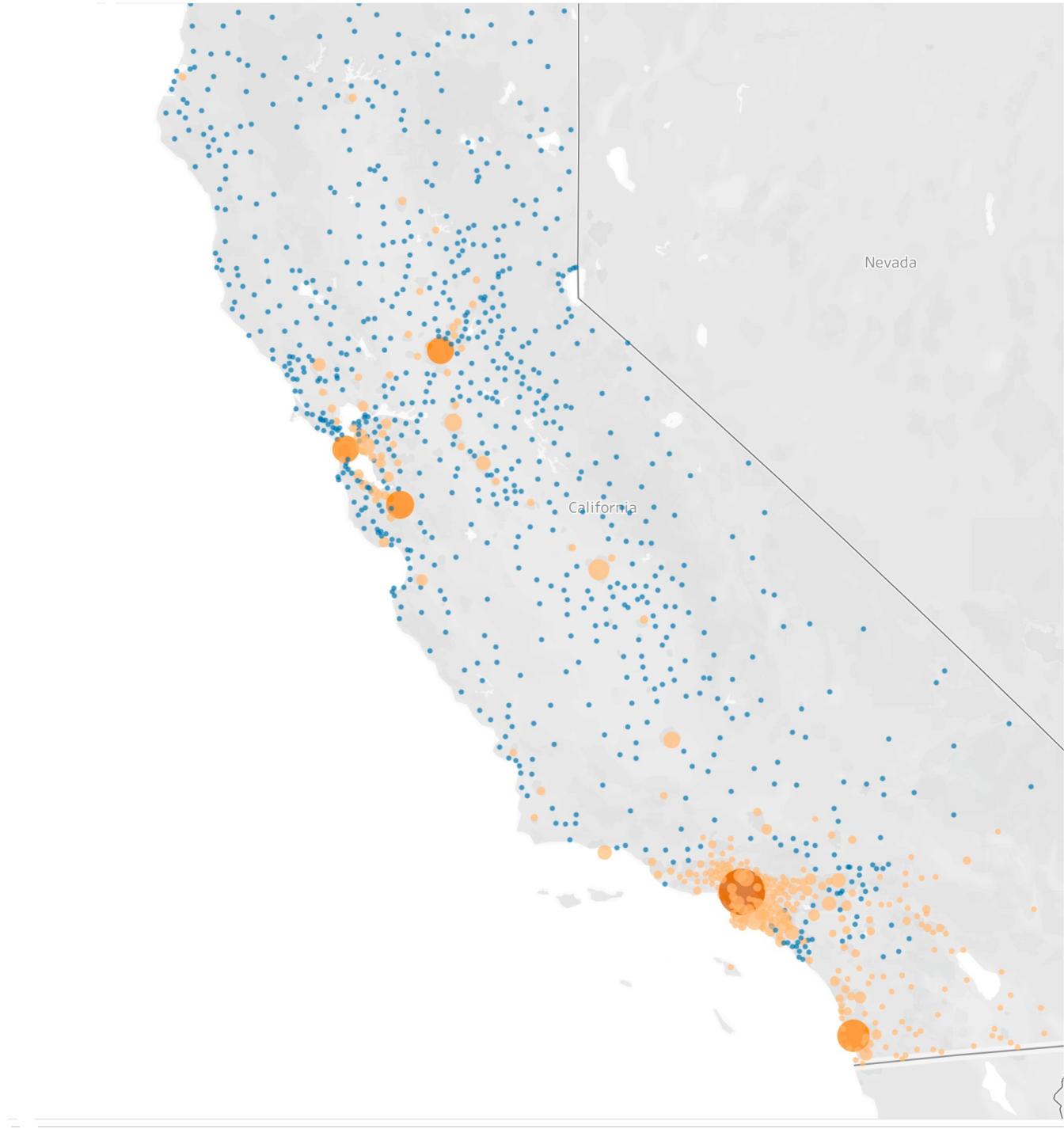
**Monthly:** 7 months

**One year:** 4 years

**Two year:** ~5.5 years

## 2. Key Features

Most customers (~52%) are located in Large Urban areas



### 3. Feature engineering

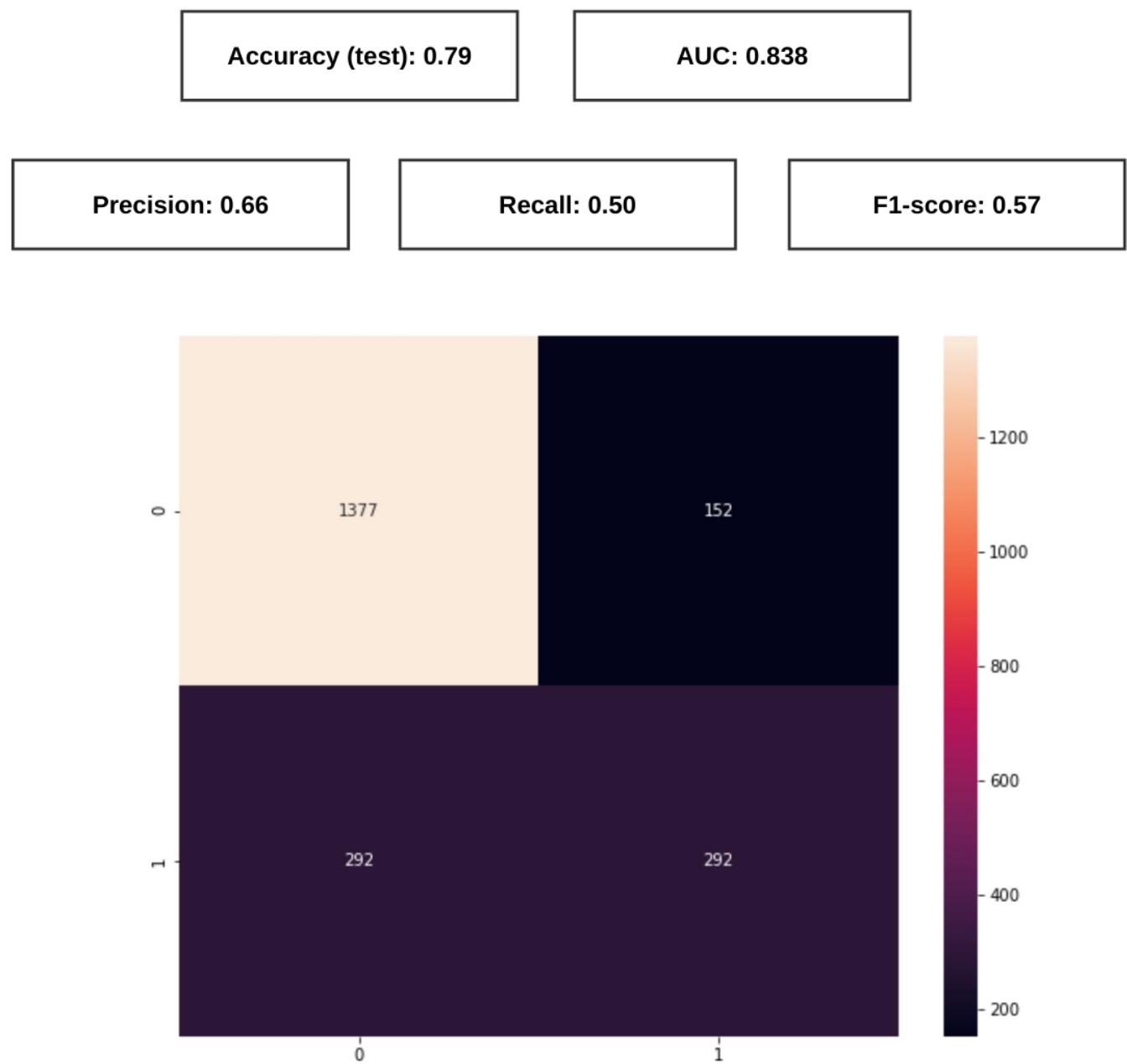
Locations:  
**Large Metro,**  
**Medium/Small Metro and**  
**Rural**

Churn Reason:  
**Competitor, Price, Product,**  
**Customer Support &**  
**Relocation**

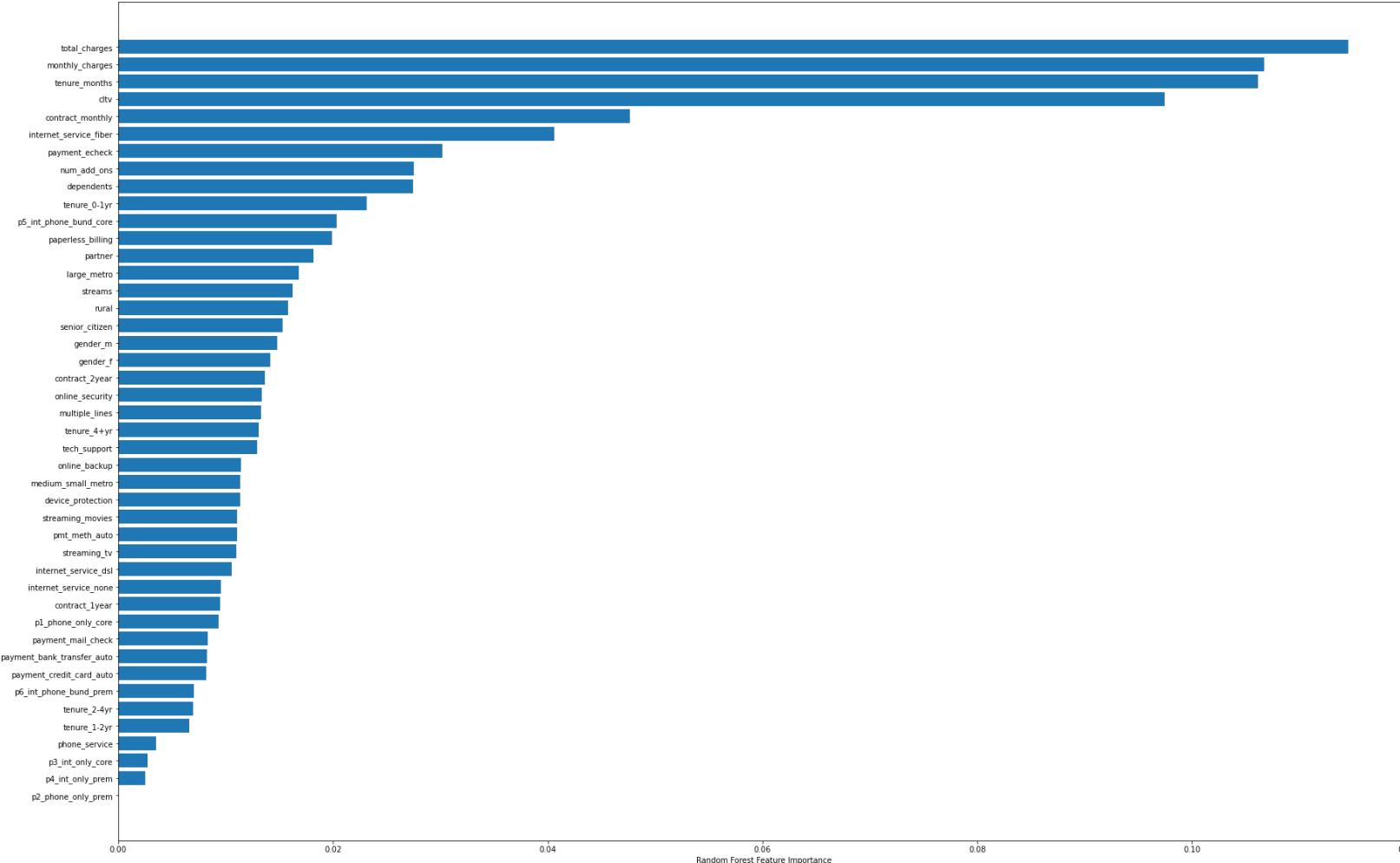
Product bundles:  
**Core & Premium**

Additional include:  
**Customer Tenure Groups,**  
**Number of Add-ons,**  
**Media Streaming &**  
**Auto-Payment Preference.**

### 3. Baseline Model – Random Forest



# 3. Baseline Model



Top important features:

- Total Charges
- Monthly Charges
- Tenure in months
- CLTV
- Contract Monthly

# 3. Baseline Model

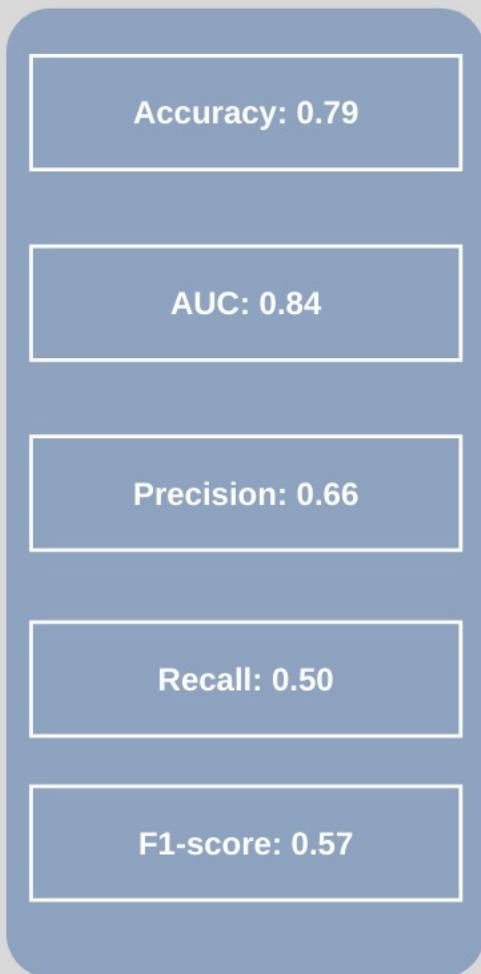
Let's remove low importance and high correlation (0.80+) features:

- Separation between Streaming Movies and TV is not significant
- Dropping all tenure buckets, except for Tenure 0 – 1 year
- Separation between phone-only and internet-only product categories is not contributing much

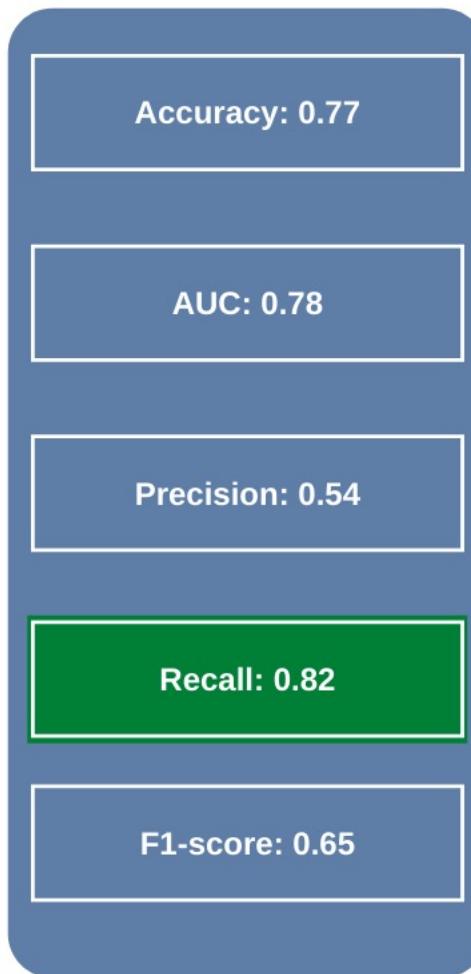
streaming_movies	streams	0.875754
streaming_tv	streams	0.875299
tenure_months	tenure_4+yr	0.858873
	total_charges	0.826178
phone_service	p4_int_only_prem	-0.889728

## 4. Model Iterations

# Base Model: Random Forest



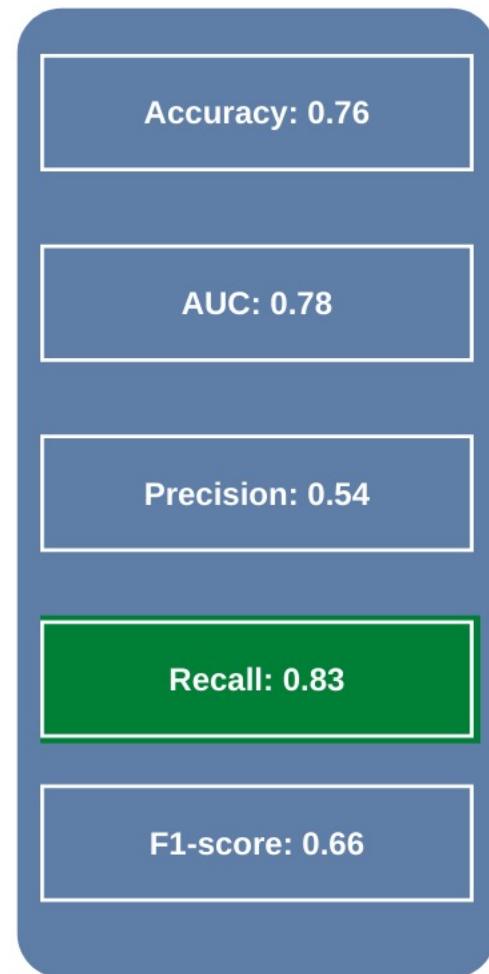
**Random Forest:**  
Class\_weight ='balanced',  
max\_depth=10,  
min\_samples\_split=0.1,  
n\_estimators=500  
decision\_threshold: 0.5



**XGBoost:**  
scale\_pos\_weight=1,  
subsample=0.6,  
learning\_rate=0.1,  
max\_depth=8,  
n\_estimators=500  
decision\_threshold: 0.3

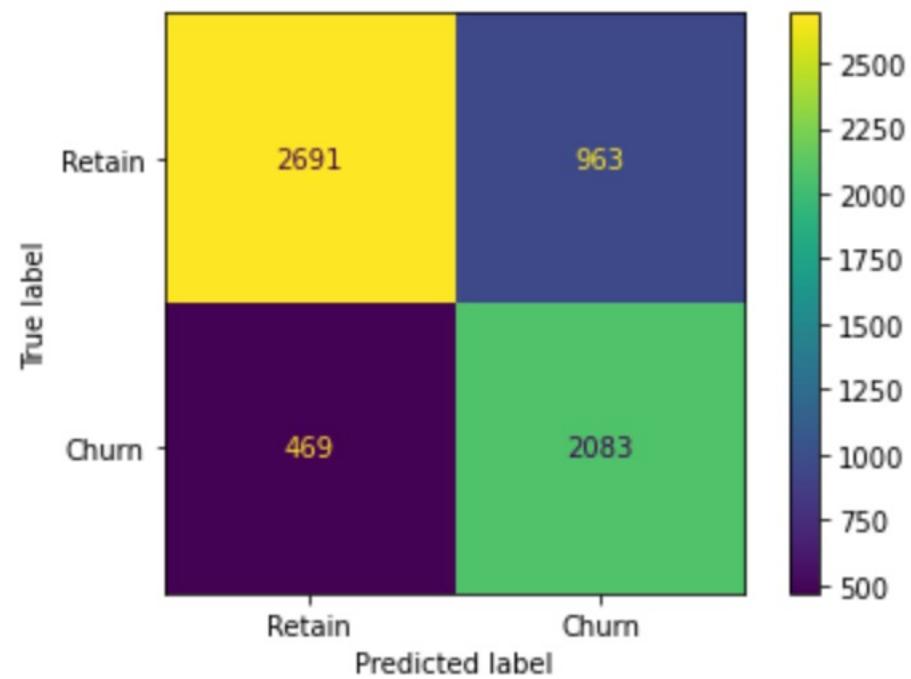


**Logistic Regression:**  
C=1,  
Class\_weight ='balanced', penalty=l1,  
min\_samples\_split=0.1,  
max\_iter=500  
decision\_threshold: 0.5

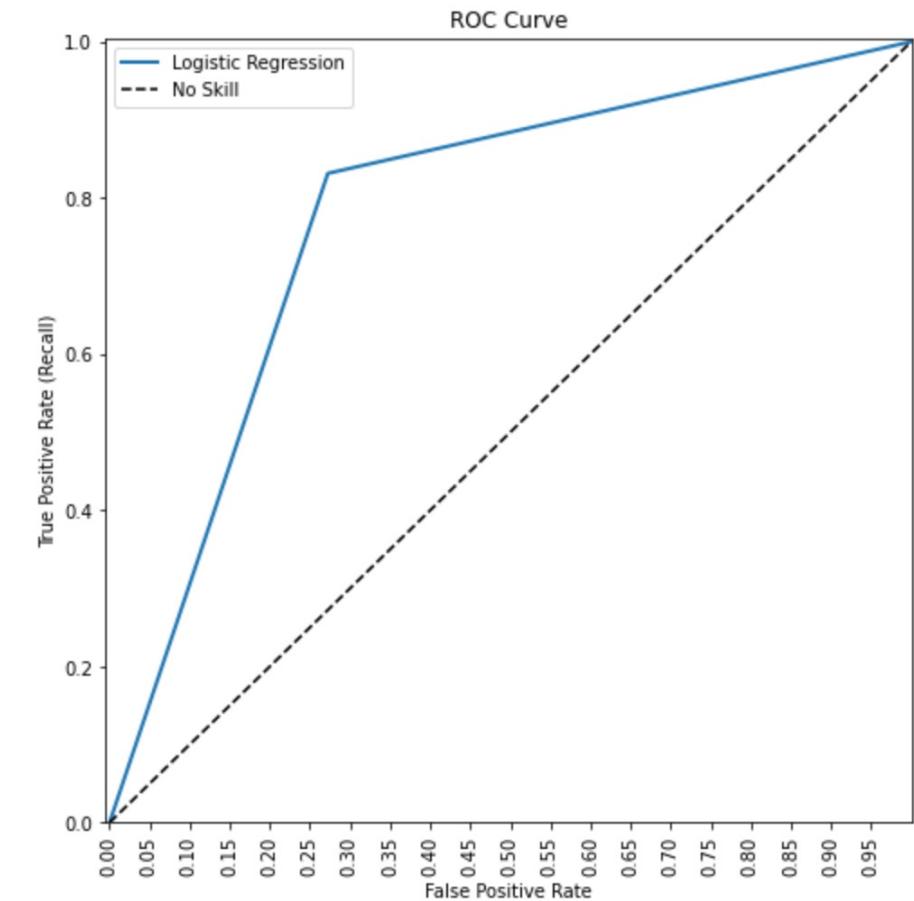


Oversampling by a factor of 2

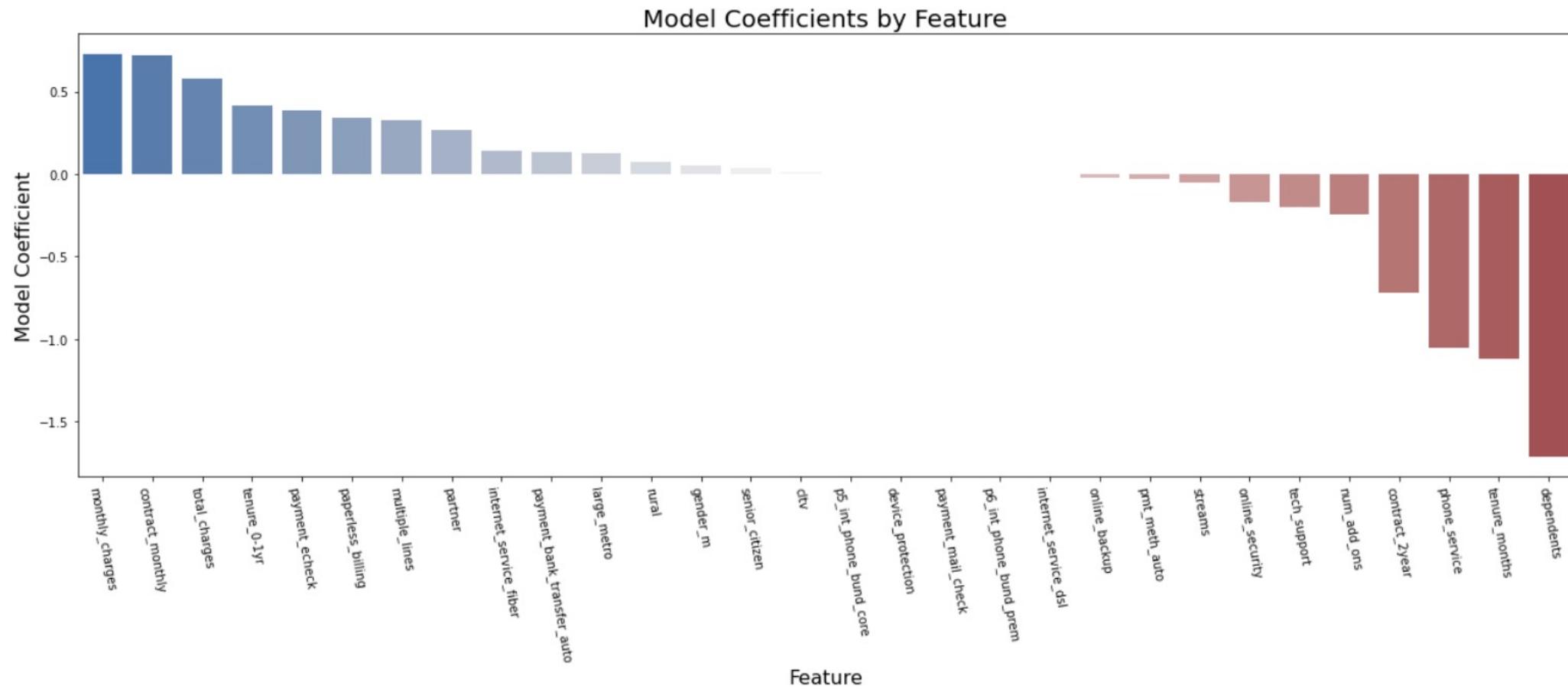
# 4. Final Model – Logistic Regression



Logistic Regression Model AUC 0.779



# 4. Final Model – Logistic Regression



## 5. Takeaways

### Main takeaways:

- **Monthly contract** are the most likely to churn (**within the first 5 months**)
- **Highly sensitive to changes in monthly charges:** increase of \$10 dollars, results in increase of odds of churning by ~36%
- Customers residing in **Large Metros** are 13% more likely to churn
- Customers with a phone line seem the most loyal
- Customers who subscribe to additional services such as Tech Support and Online Security are ~15% less likely to churn
- Overall, going up by 1 additional service reduces the odds of churning by ~22%
- Customers with dependents are significantly less likely to churn (~80%) , but customers with partners are ~30% more likely to churn

## 6. Next steps

- Roll out discounts for monthly customers to upgrade to annual plans
- Provide free trials of additional services
- Provide more personalized customer support
- Offer “buy one, get one free” deals for customers on single user plans

**THANK YOU!**