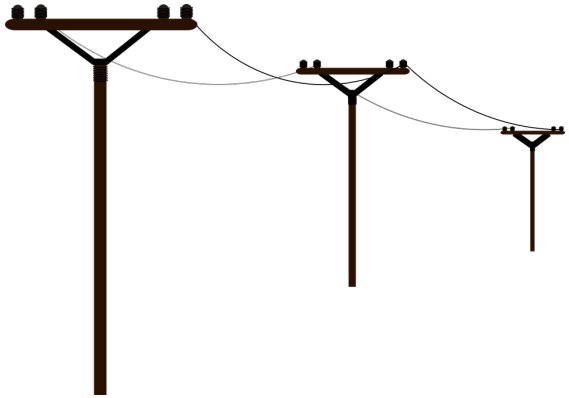
Telco Customer Churn



By Sam Ouimet

What is customer churn?

Definition

- Customer, user, or subscriber stops using the product or service of service of a particular company.

Why does it matter?

1. Cheaper to retain customers than to gain new ones
2. Money often flows to competitors
3. **IMPEDES GROWTH**

The Data

- **Services** that each customer has signed up for.
  - Phone, multiple lines, internet, streaming TV and movies, etc.
- **Customer** account information:
  - Tenure, contract, payment method, monthly charges and total charges, etc.
- **Demographic** info about customers – gender, age range, and if they have partners and dependents
- Customers who left within the last month:
  - **Churn (y/n)**

kaggle

The Process

Identify specific business problem

- Churn is hurting business.
- We need to identify which customers need more attention, deals, promotions, etc. to increase retention.

Explore data

- Gather data/clean
- EDA
- Feature Engineer
- Baseline
- Feature Selection
- Optimize

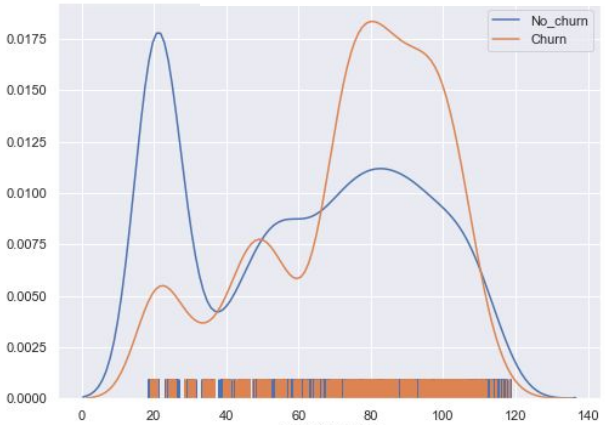
Test and report

- Pick final model with best results (**F1**)
- Build Tableau dashboard for internal company use

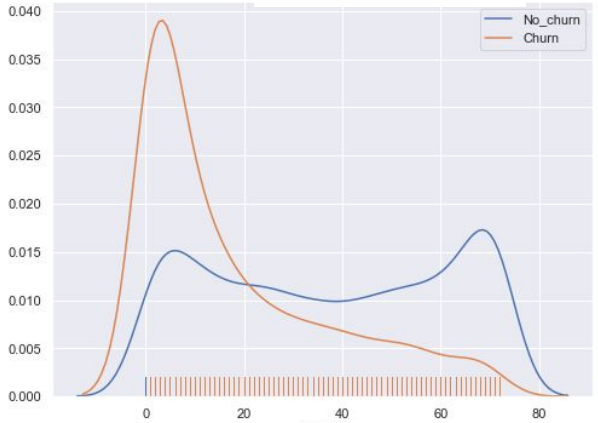
Early Findings

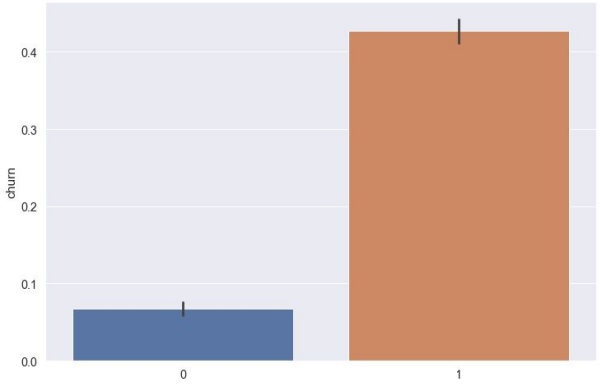
- Some linear separation present
- Certain features contribute a lot to churn rate:
  - Contract type
  - Monthly/charges
  - Tenure
  - Age group
  - Online Security
  - Tech Support

Monthly Charges

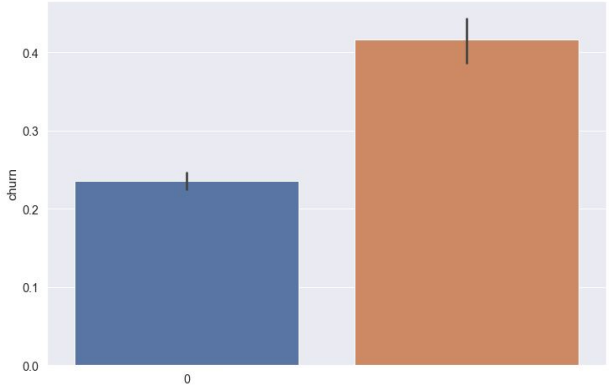


Tenure





Monthly Contract Type



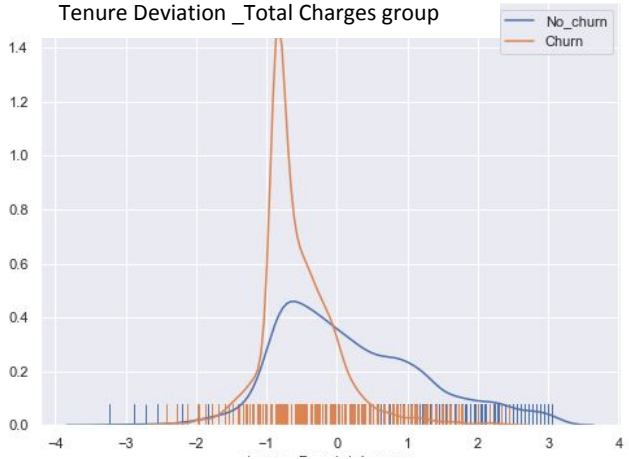
Senior Citizen

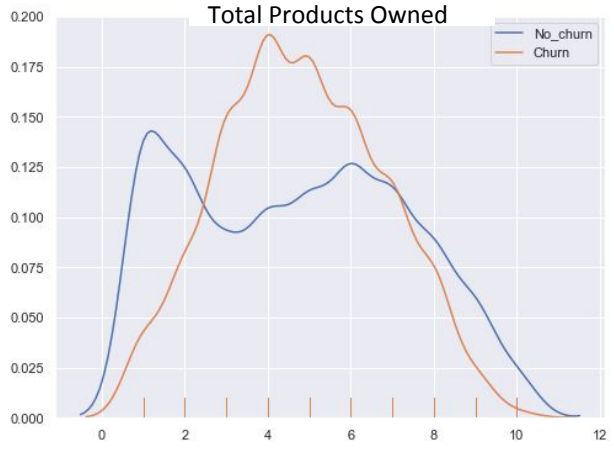
Feature Engineering

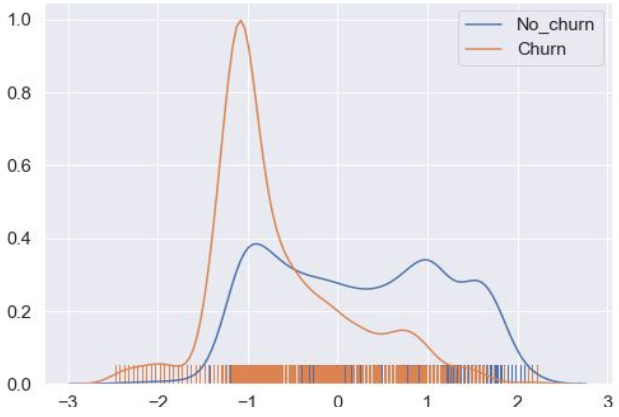
- Convert continuous variables into bins:
  - Monthly Charges
  - Total Charges
  - Tenure
  - # products owned
- Deviation features of newly created groups
- Total products owned

Top features:

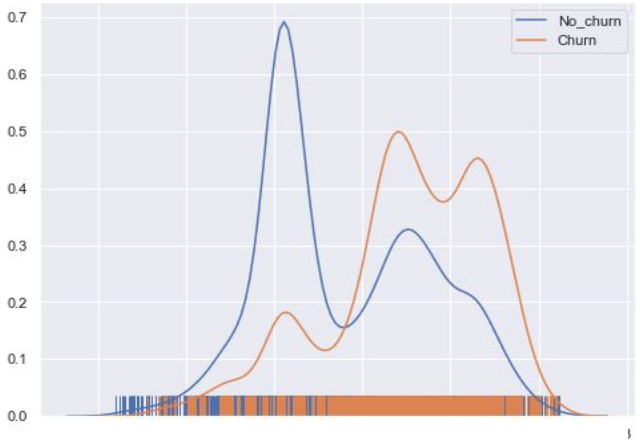
1. Tenure_dev_total_group
2. Tenure_dev_monthly_group
3. Monthly_charges_dev_prod_own_group





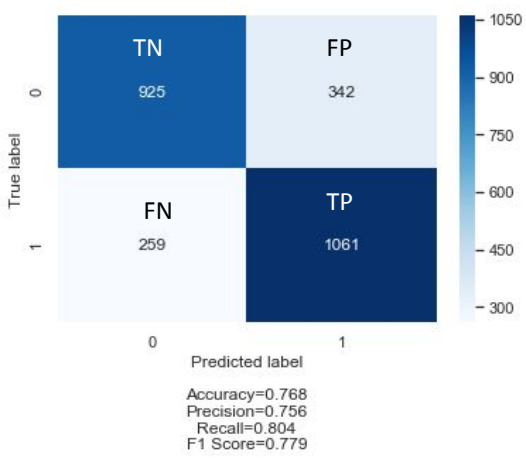


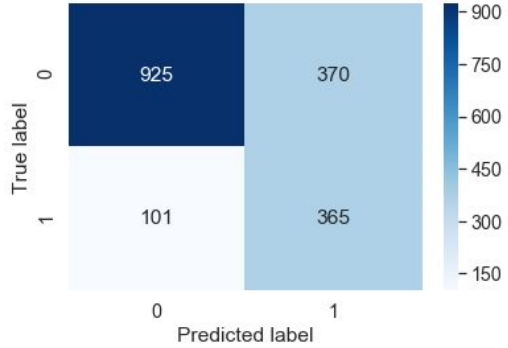
Tenure Deviation_ Monthly Charges Group

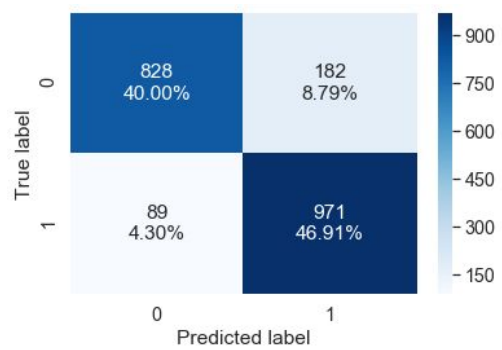


Monthly Charges Deviation _ Products Owned Group

Modeling







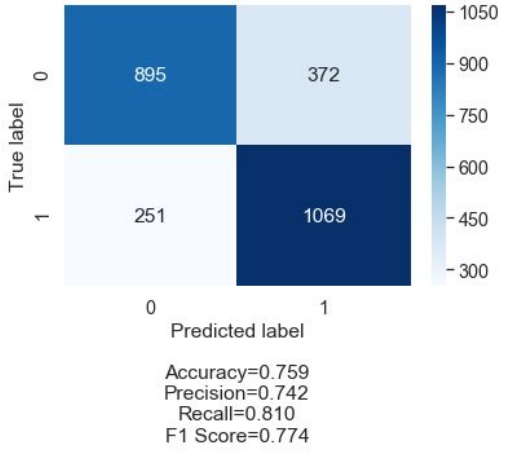
Logistic Regression

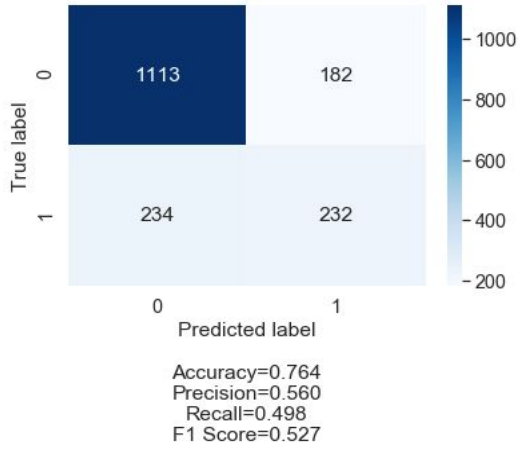
Naive Bayes (Gaussian)

Random Forest

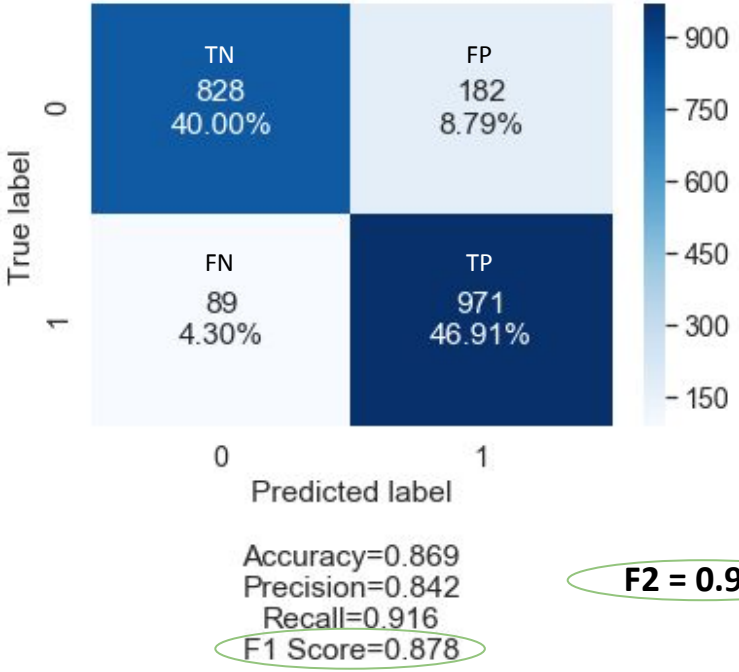
XGB Baseline

XGB Final





Final Model: XGBoost



F2 = 0.900

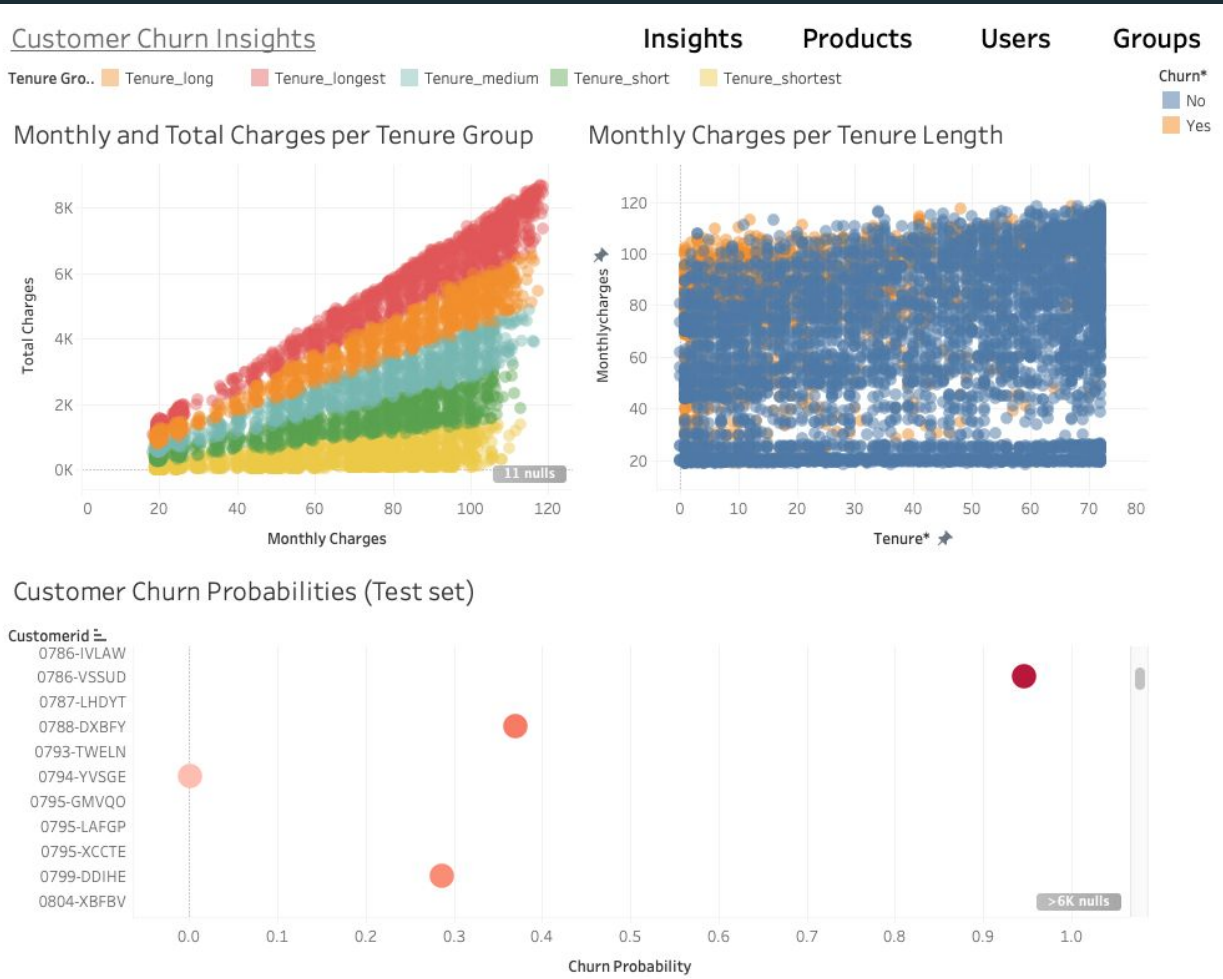
XGboost on balanced data (oversampled), with tuned parameters and L1 regularization.

Business Intelligence Dashboard

tableau

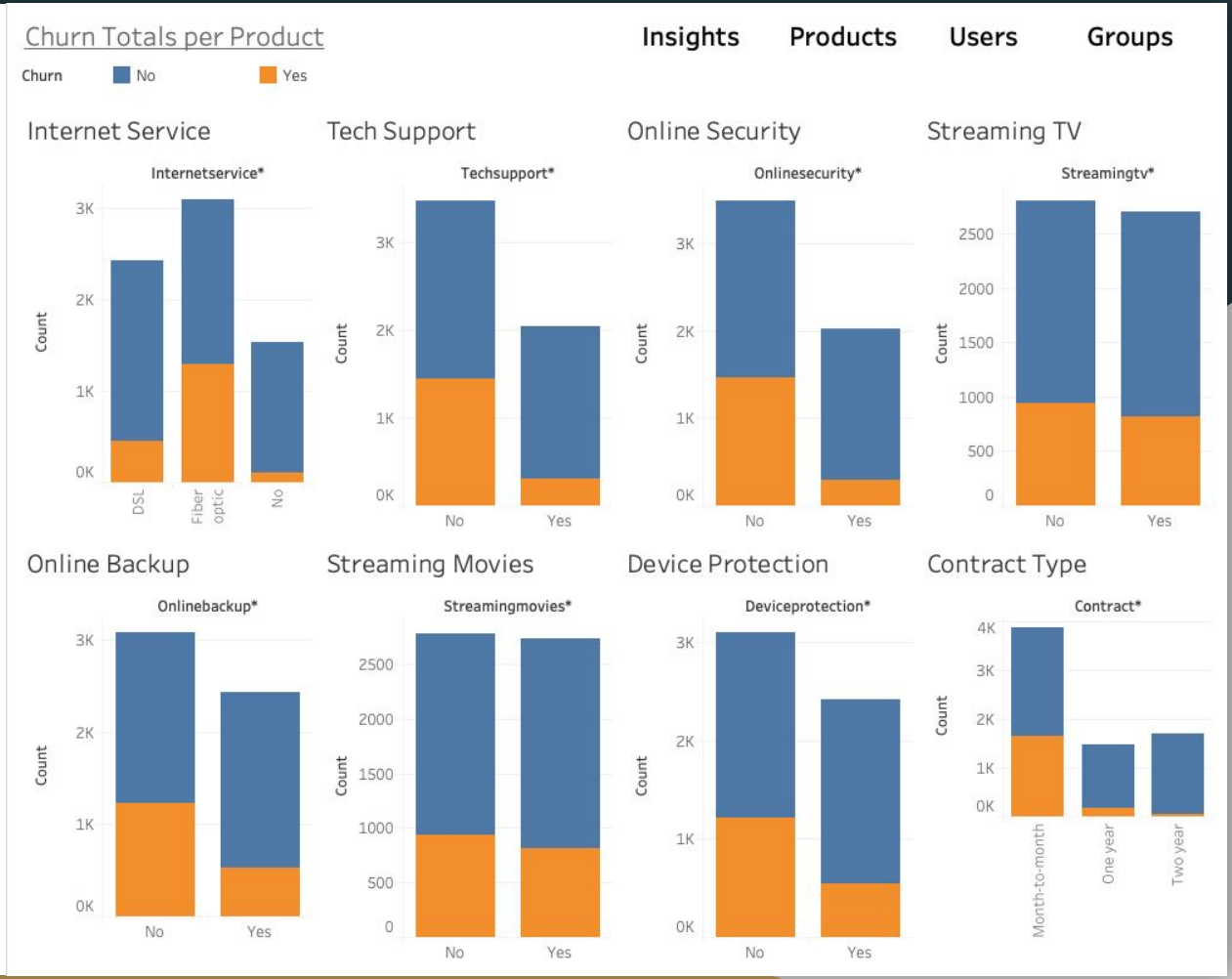
Insights

- Monthly and total charges per tenure group
  - Shows a **range of charges to expect** from each customer based on tenure group
- Monthly charges per tenure length
  - **New customer with high charges = likely to churn**
  - **Offer discounts, promos to new customers**
- Customer Churn Probabilities
  - Which customers are most likely to churn?



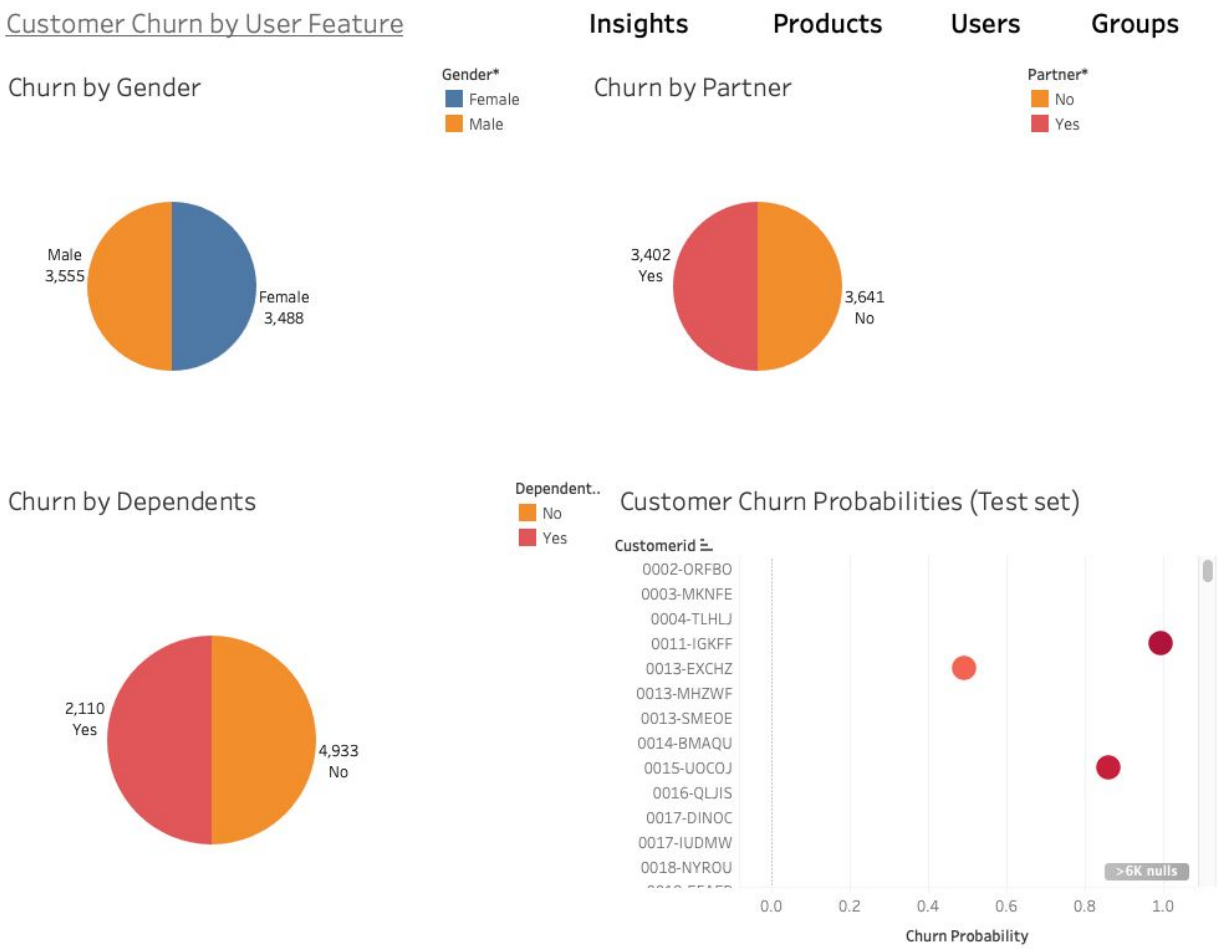
Product Info

- Tech support, online security and backup aid retention
  - Advertise these products to customers who don't have them
- Most customer who churn are on month-month contracts
  - **Upsell longer contract terms!**



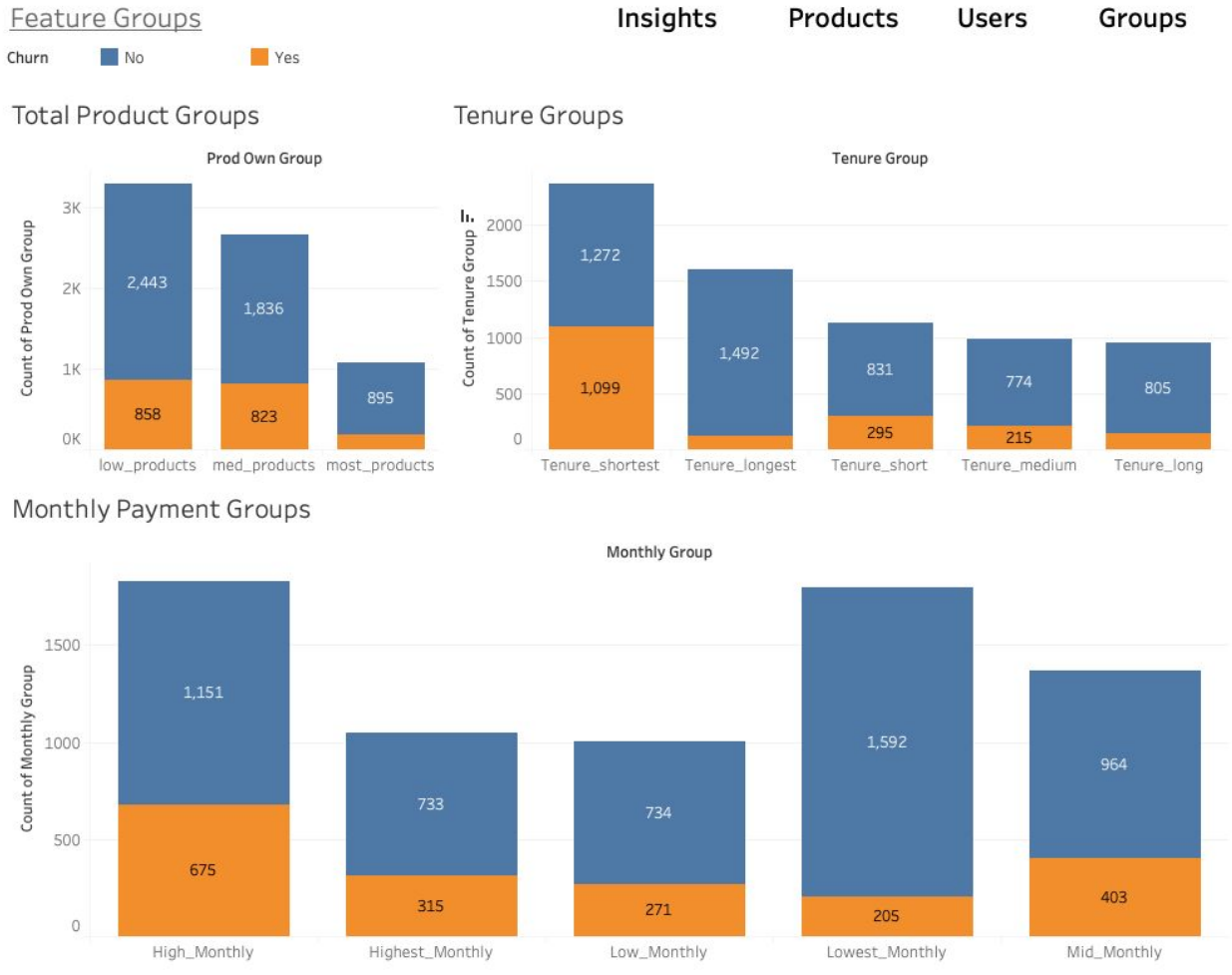
User Info

- A look at how customer gender, partners, and dependants impact churn
  - They don't really, but may be useful to **monitor for potential trend changes**



Engineered Groups

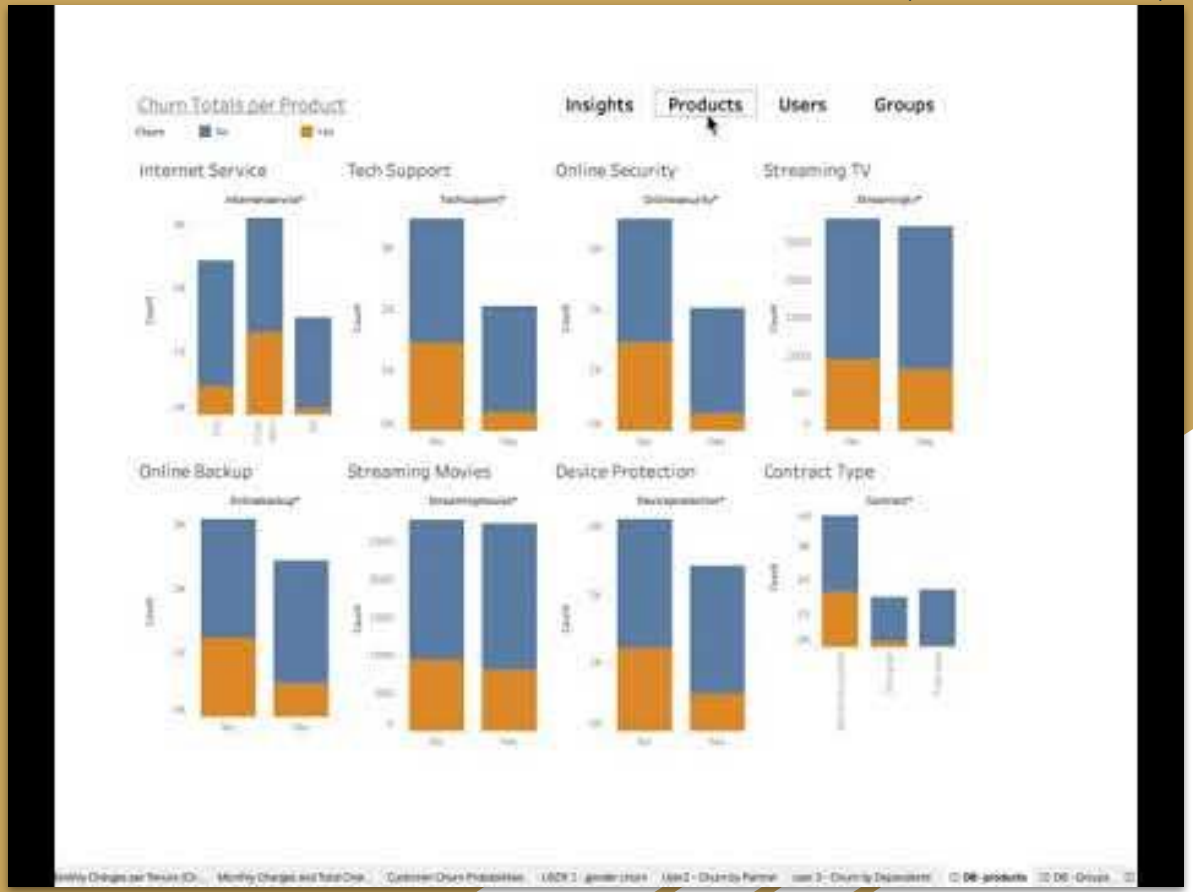
- Total owned product group:
  - Most customers fall in the “low” to “medium” products groups
  - (1-6 products)
- Tenure group:
  - The longer the customer stays, the less likely the churn
  - **Offer incentive for customers to stay past X cycles**
- Monthly payment group:
  - Lower the customer is paying, the less likely the churn



(Dashboard demo)

Thank you!

Questions?

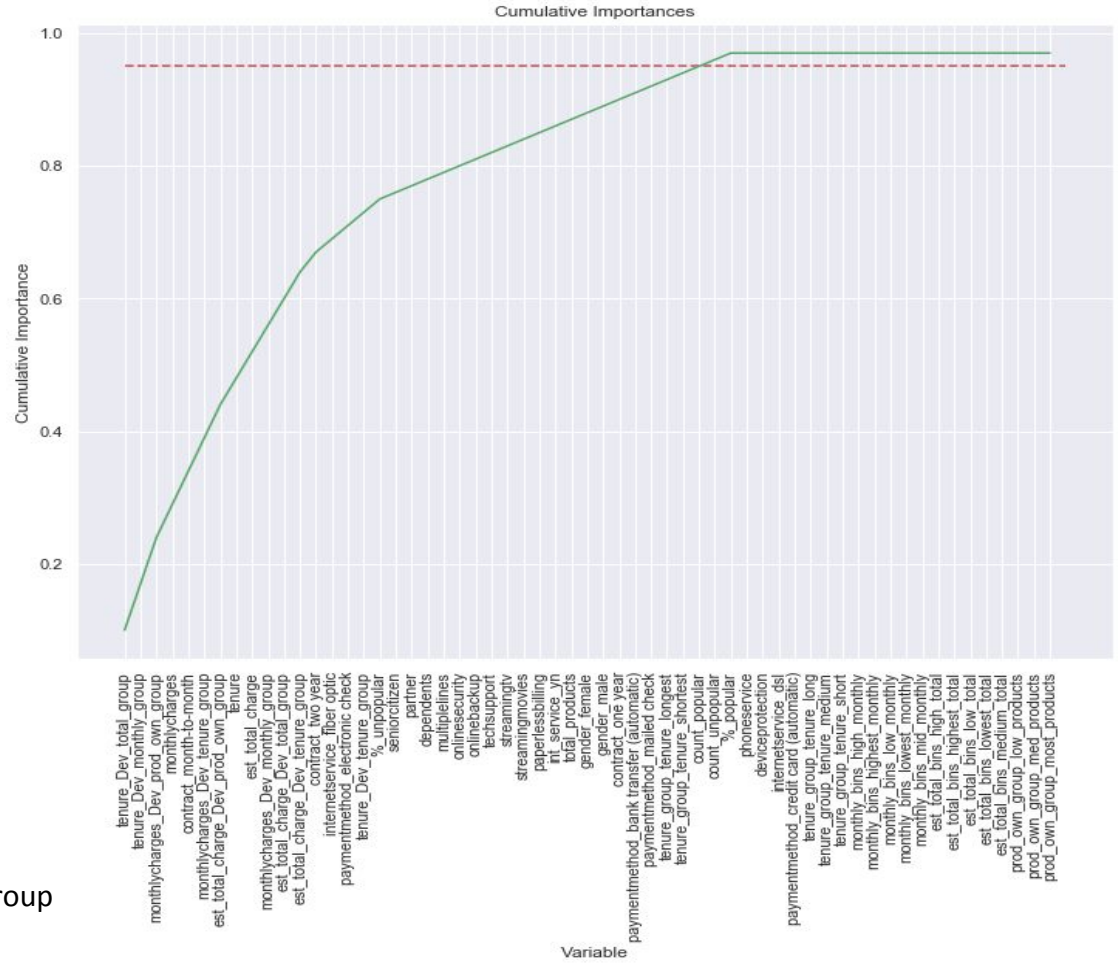


Appendix

	Log Baseline (unbalanecd, all features)	Log (balance, tuned, L2)	NB (Balanced, reduced feature set)	RF, (balancd)	XGboost Baseline (unb, all features)	**XGboost Final (bal, tuned, L1)**	
Accuracy	0.793	0.768	0.752	0.735	0.798	**0.869**	
Precision	0.630	0.756	0.735	0.500	0.5171	**0.916**	
Recall	0.527	0.804	0.806	0.793	0.663	**0.842**	
F1	0.574	0.779	0.769	0.614	0.581	**0.875**	
F2	0.545	0.796	0.790	0.710	0.54	**0.900**	

Feature Selection

- Feature importance results from random forest model
- Ranked importance from most to least (left to right)
- Dropped remaining features after 95% importance was reached
- Features left: 38



Top features:

1. Tenure_dev_total_group
2. Tenure_dev_monthly_group
3. Monthly_charges_dev_prod_own_group

Resources:

https://www.kaggle.com/blastchar/telco-customer-churn

https://www.analyticsvidhya.com/blog/2016/03/complete-guide-parameter-tuning-xgboost-with-codes-python/