



Customer Churn

Bernard Opoku

INTRODUCTION

- What is it?
- Why does it matter?
- What are we going to do?



Procedure



GET OUR DATA



EXPLORE THE DATA



BUILD A MODEL

Data Acquisition & Storage



Data

Telco Dataset-Kaggle
21 Initial Features –
Mixed
~10K observations

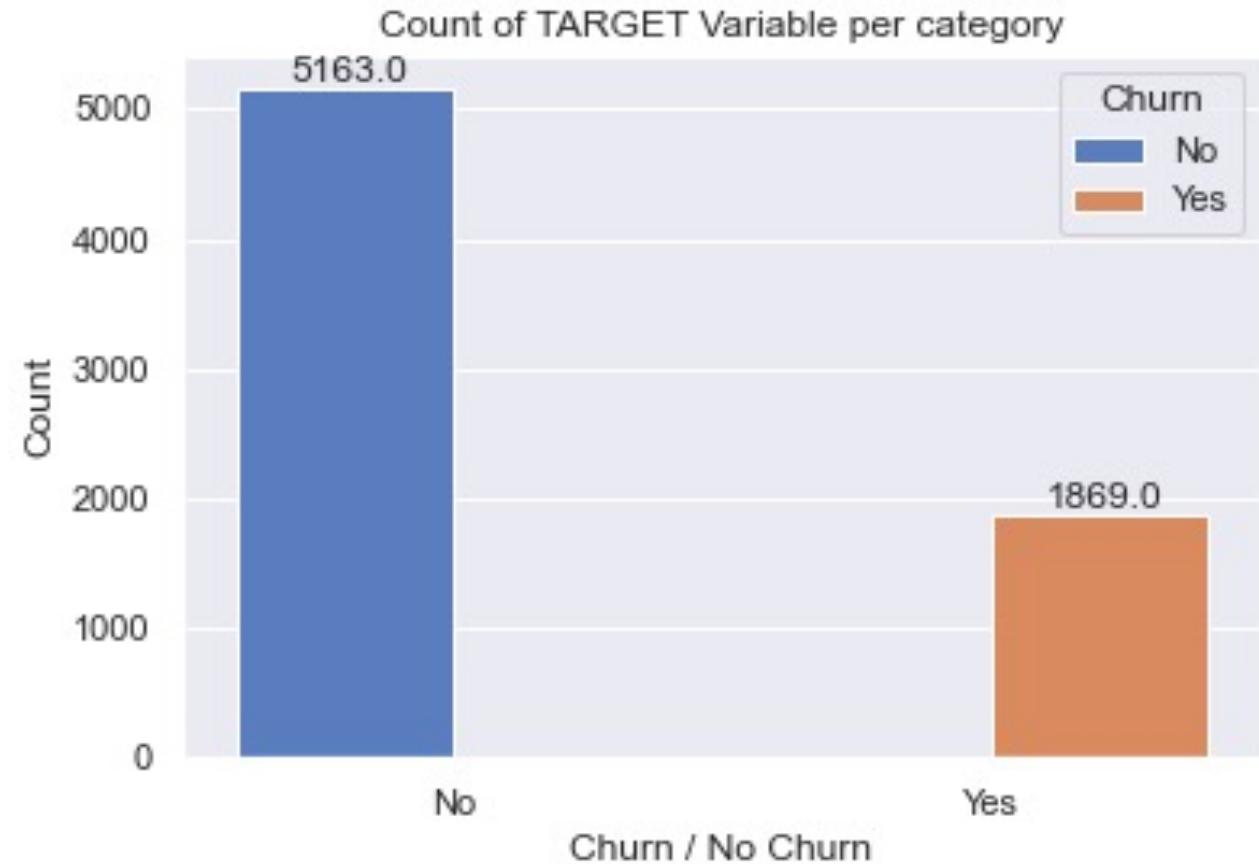


Storage

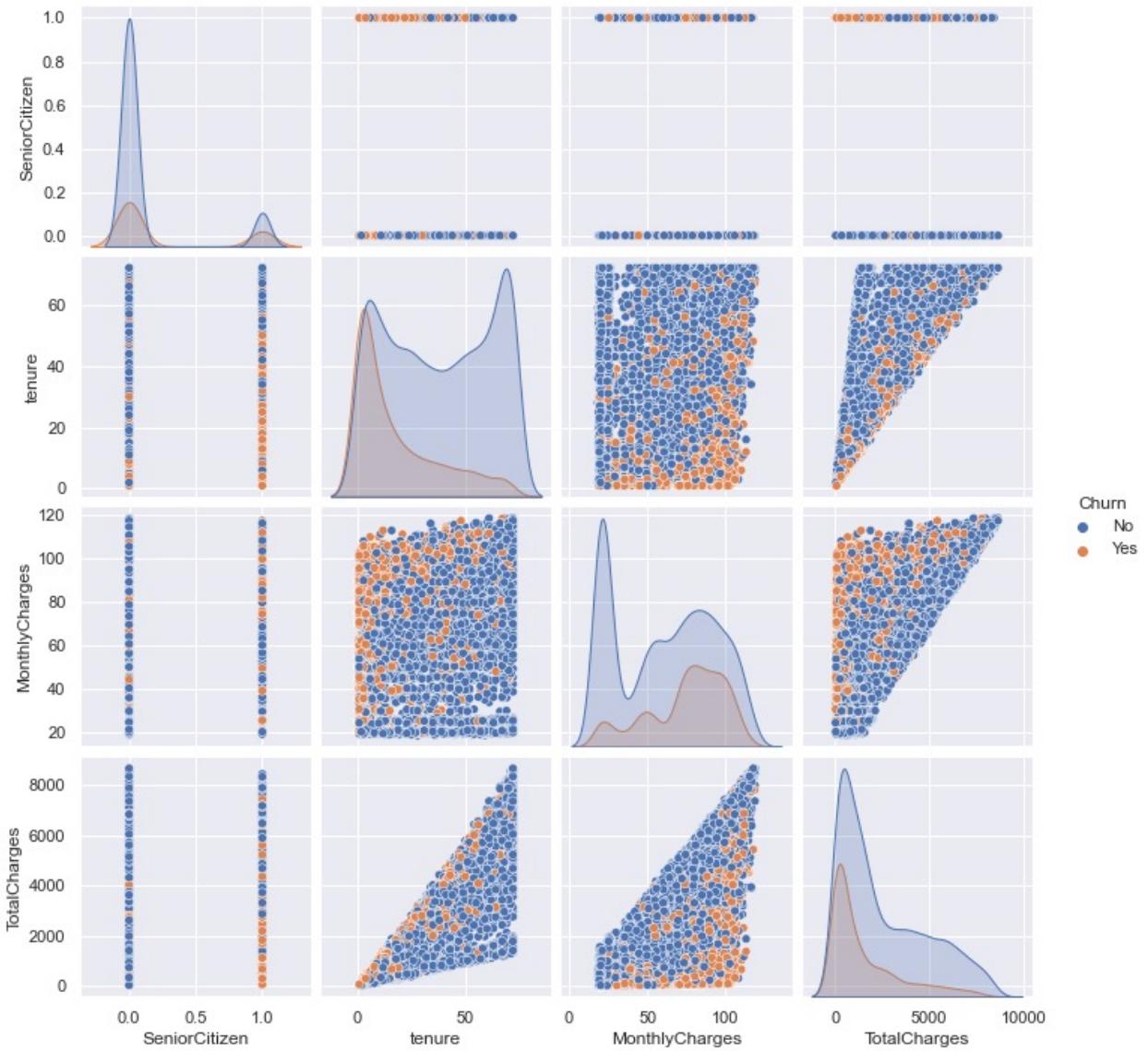
SQL
Database

Data Exploration

Imbalance Data



Don't be tricked

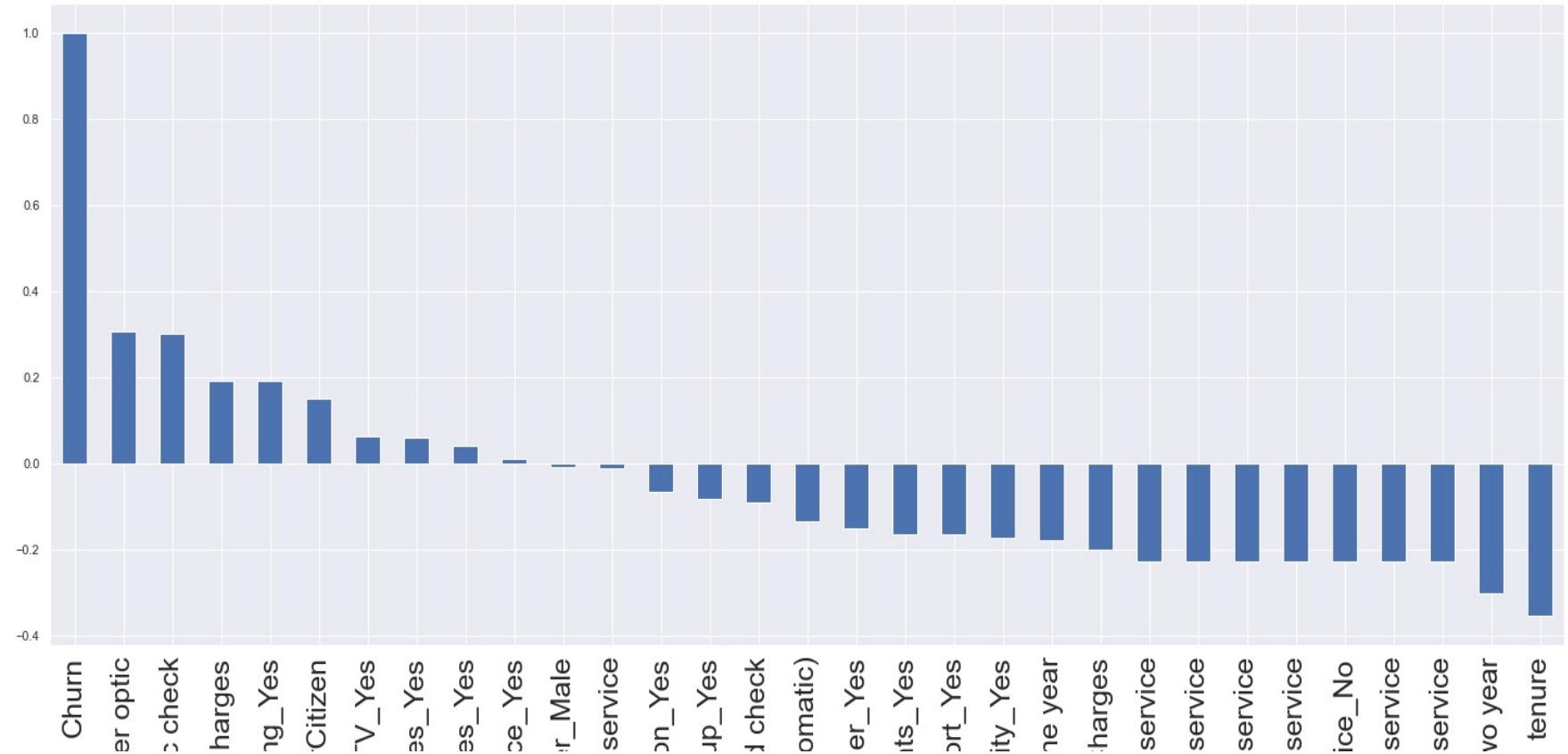


Tenure



Before Modeling

Any
relationship?



Modeling

- What do we want to achieve?
- What is our use case?
- What will be our solution to churn?

Predictive Modeling

- Logistic Regression as baseline
- Recall: 0.4328
- F2 Score : 0.4636
- ...Ooopsss!!
- SO WHERE FROM HERE?
- Test Different Models
- Every Model:
 - Bring in all our our feature
 - Make them usable by the model (dummy)
 - Cross- Validation & Hyperparameter Tunning
 - Class Imbalance Techniques

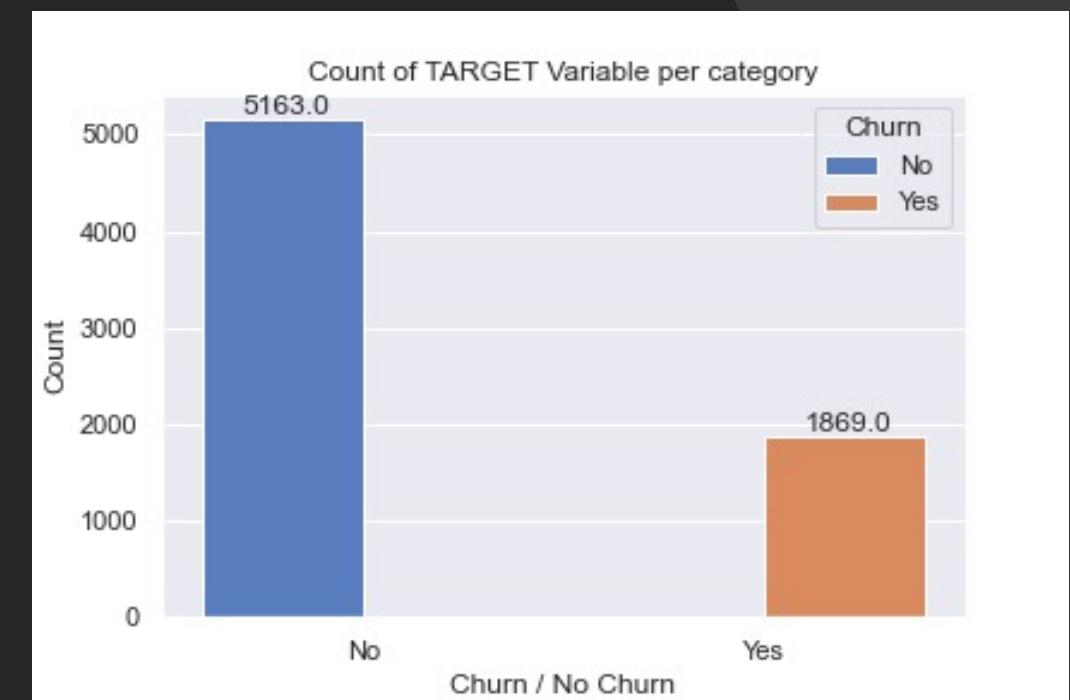
Class Imbalance Techniques

Threshold Adjustment

Oversampling

- SMOTE
- ADASYN

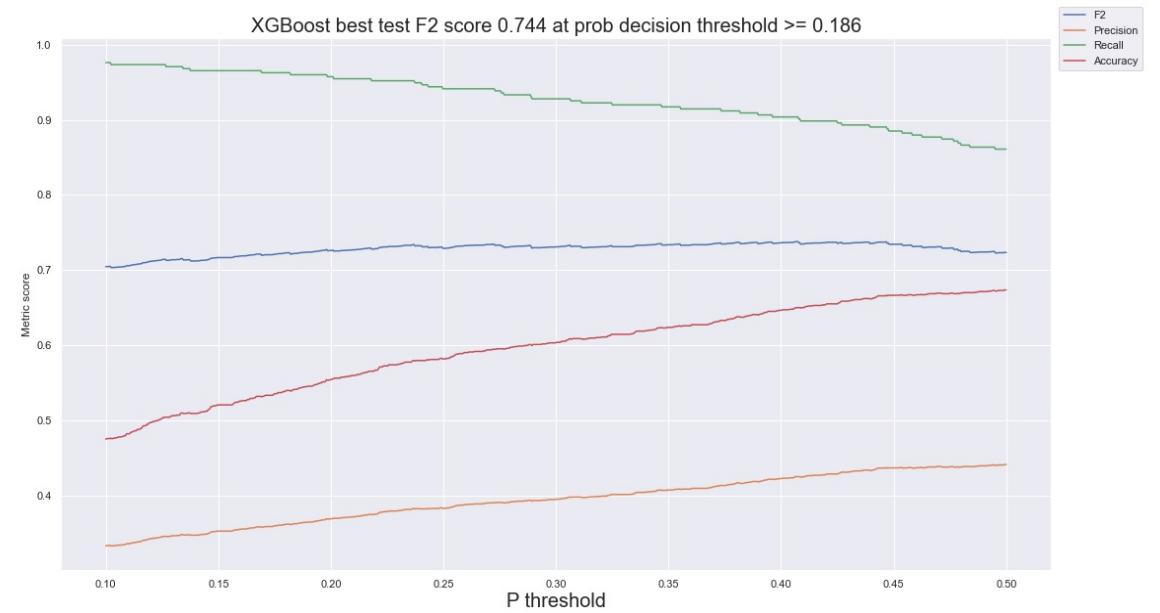
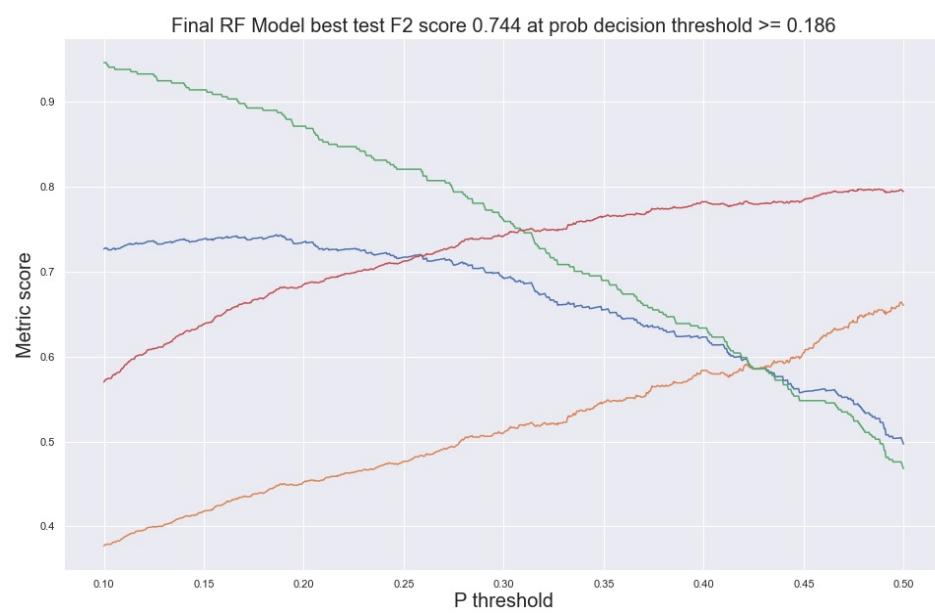
Both



Compare

Model	Base	Class Imbalance Technique , CV, GS - Val	Test
Logistic Regression (CV)	F2_Score:0.555 Recall:.534	F2: 0.770 Threshold = 0.196	
KNN	F2 Score: 4.84		
Decision Tree	F2 score: 0.539 Recall:0.5320	F2 score 0.757 Threshold >= 0.200	F2 Score: 0.710 Threshold >= 0.100
Random Forest	F2 score: 0.535 Recall: 0.506	F2 Score: 0.770 Threshold >= 0.176	F2_Score: 0.745 Threshold >= 0.179
XGBoost	F2_Score: 0.697 Recall: 0.813	Recall score: F2 score: 0.756	F2 score: 0.734 Threshold >= 0.439

Random Forest vs XGBoost



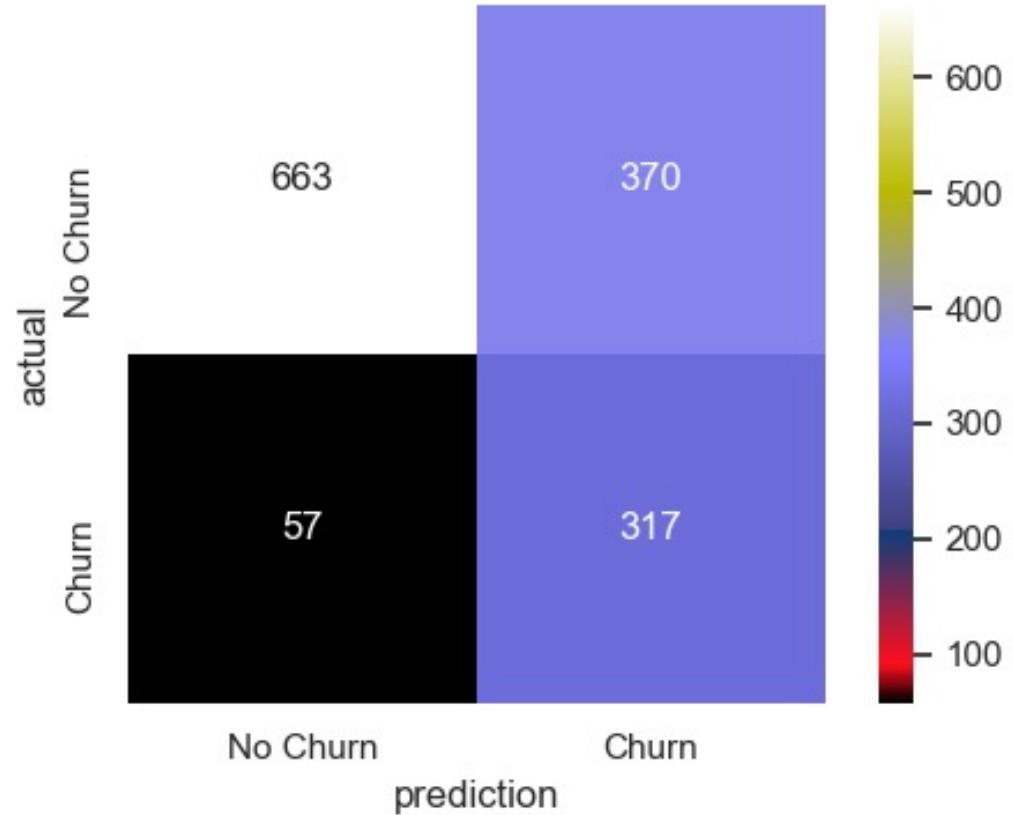
Winner!

Model	Base	Class Imbalance Technique , CV, GS - Val	Test
Logistic Regression (CV)	F2_Score:0.555 Recall:.534	F2: 0.770 Threshold = 0.196	
KNN	F2 Score: 4.84		
Decision Tree	F2 score: 0.539 Recall:0.5320	F2 score 0.757 Threshold >= 0.200	F2 Score: 0.710 Threshold >= 0.100
Random Forest	F2 score: 0.535 Recall: 0.506	F2 Score: 0.770 Threshold >= 0.176	F2_Score: 0.745 Threshold >= 0.179
XGBoost	F2_Score: 0.697 Recall: 0.813	Recall score: F2 score: 0.756	F2 score: 0.734 Threshold >= 0.439

Random Forest

- Combine Val and test
- Fit
- Results:

Threshold @ .21



Questions??

Thank you