CHURN in Telecom

A case study on analysis churn in Telecom and building a machine learning model to classify the customer who is likely to churn

Nicole, Tang Sin Yu

Jul 2021

CONTENT

- *What is **CHURN***
- *Project Approach*
- *What does the **DATA** tell us?*
- *Our **PREDICTION** models*
- ***PERFORMANCE** evaluate*
- ***CONCLUSION** & **NEXT STEP***

Helps our client to predict whether their current customer will churn or not.

SCENARIO

Our client, a U.S. Telecom company who provided home and Internet services

Helps our client to predict whether their current customer will churn or not.

SCENARIO

- *IBM Sample Data Sets (Fictional)*
- *Hosted on Kaggle*
- *Includes 7,043 customers in California*

CHURN, what it is?

General revision on Churn, its business values and how can we deal with it.

*"The cost of **acquiring a new customer** can be higher than that of retaining a customer by as much as **700%**"*

Bain & Company's Research

By increasing customer retention rates ***5%*** *, we could* ***increase profits by 25% to 95%***

Bain & Company's Research

CHURN

- The rate at which customers stop doing business with an entity;
- Also known as the ***rate of attrition*** or ***customer churn;***
- Great evaluate tool for the industry with Subsricption/ yearly-renewal contract as the bases. ***(telecom)***

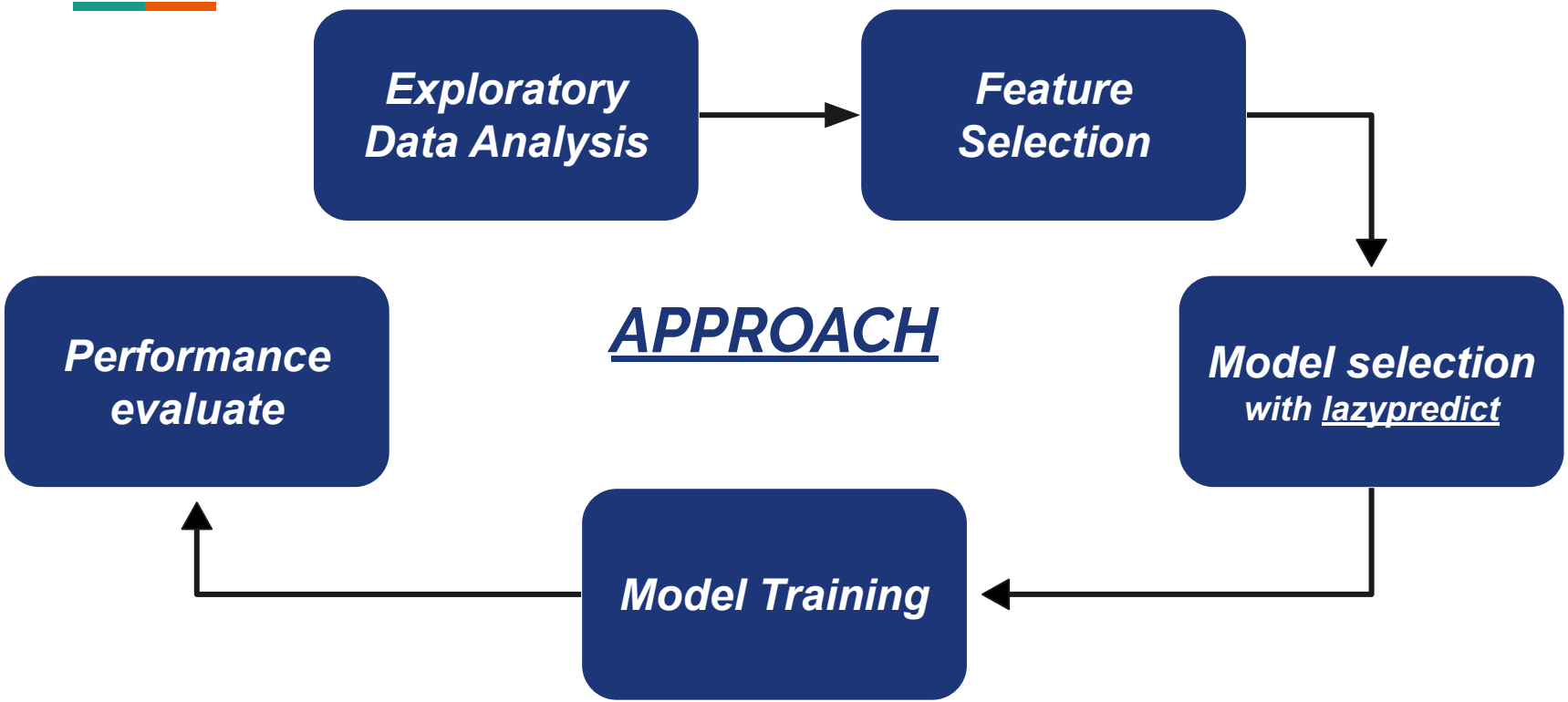
Why Dat Science

If we can predict whether a customer is likely to churn, we can take some actions and try to keep he/she with us!

@coletangsy

APPROACH

How data science helps to deal with churn, and how we are going to do with it.

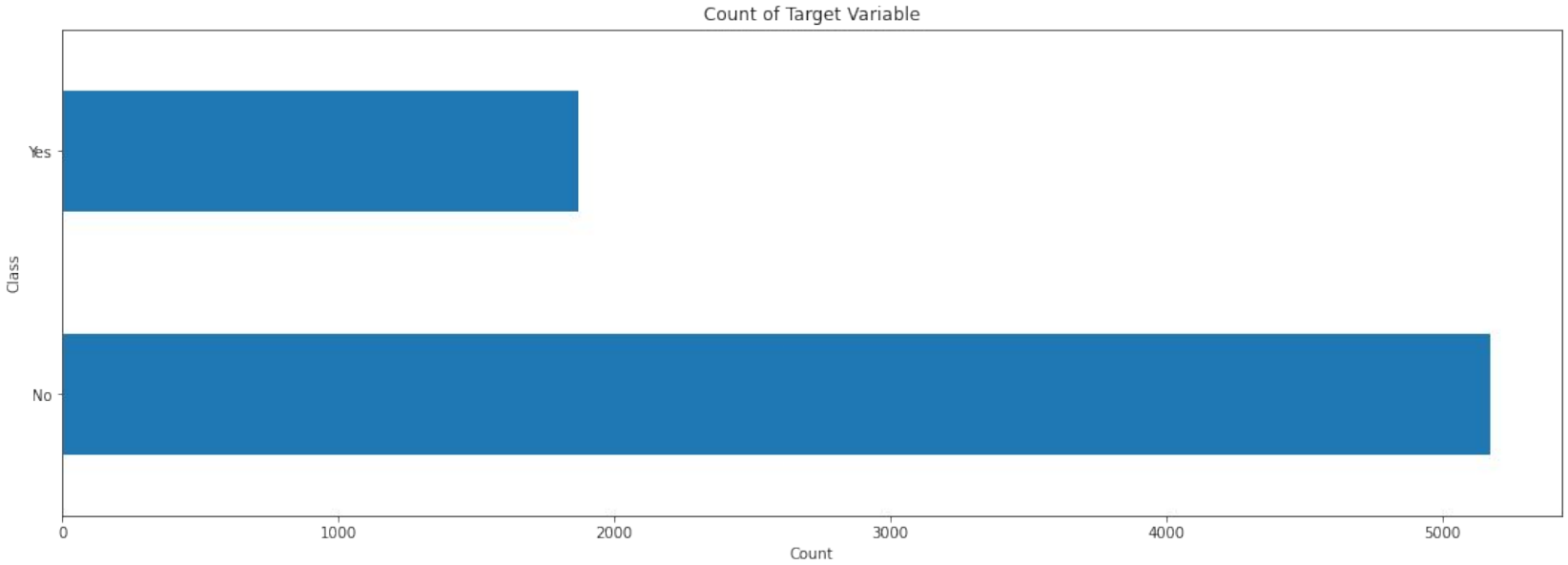

Exploratory Data Analysis
Feature Selection
Model selection with lazypredict
Model Training
Performance evaluate
APPROACH

@coletangsy

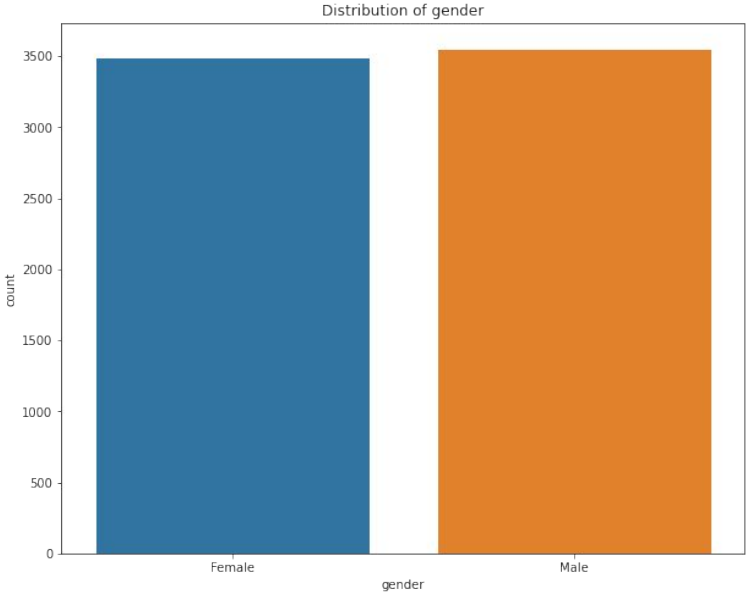
DATA ANAYLSIS

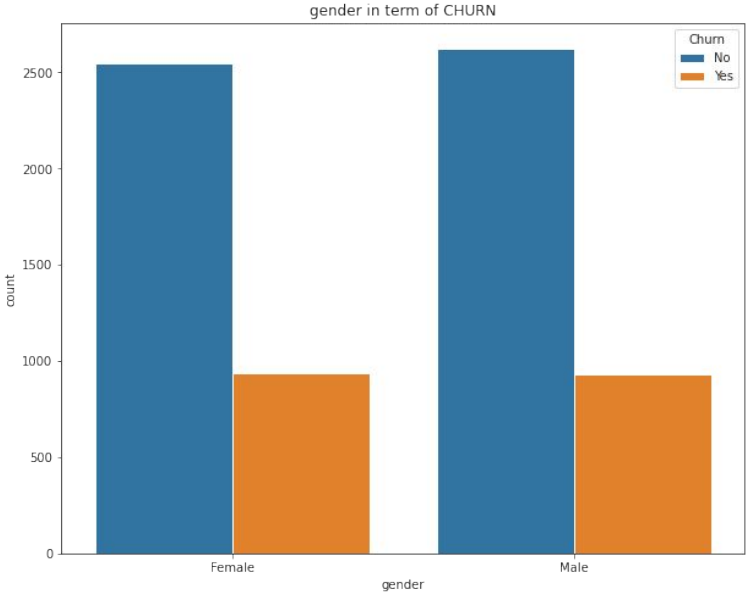
What the data told us? Let go for an EDA on the data set.

OUR CUSTOMERS

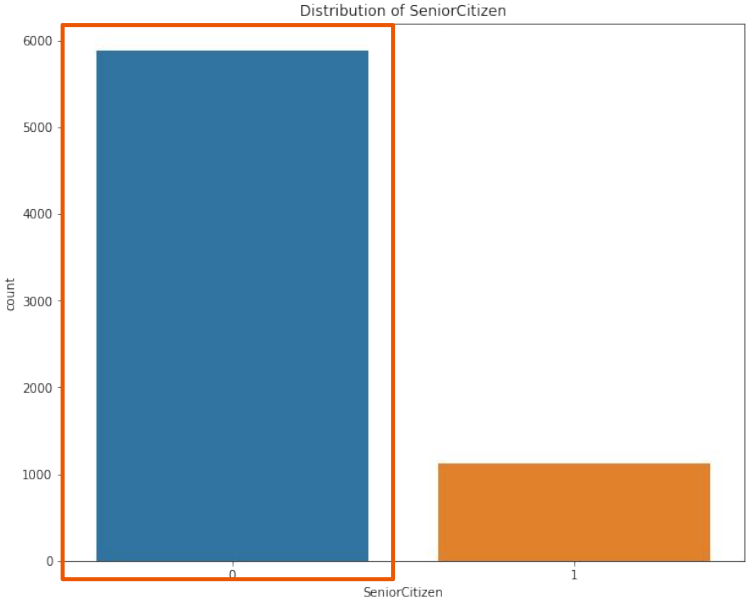


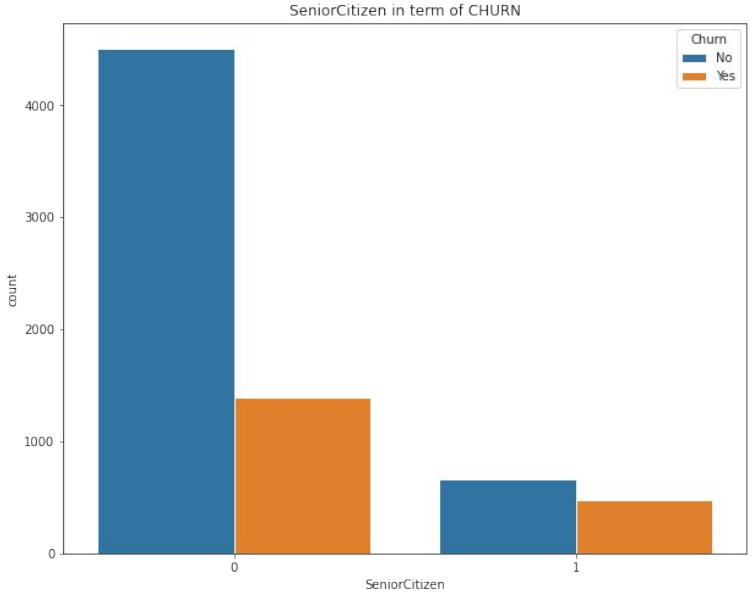
OUR CUSTOMERS



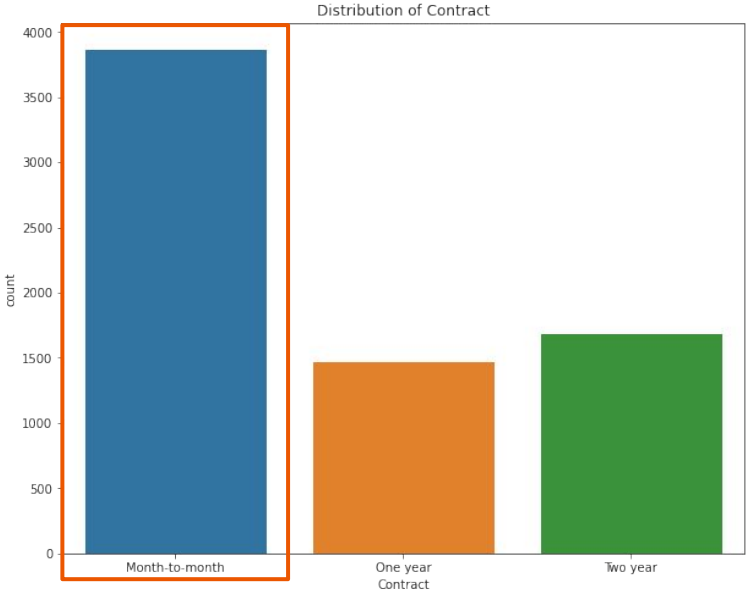


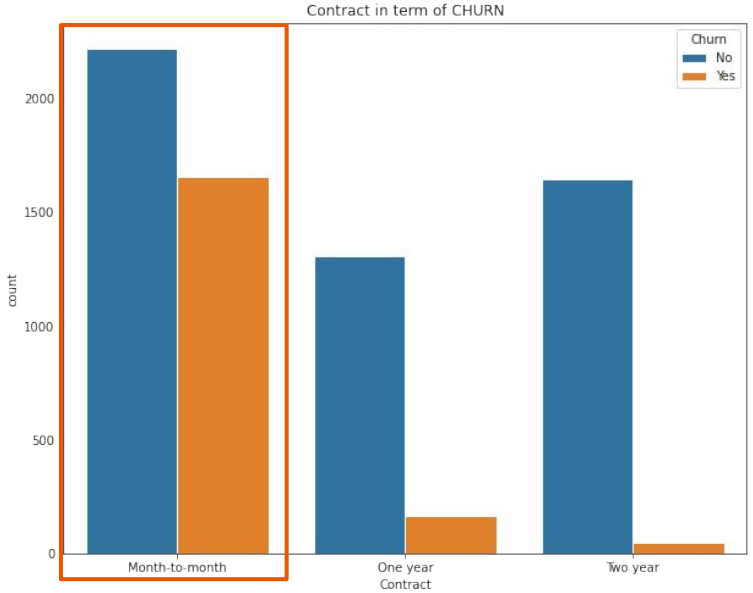
OUR CUSTOMERS



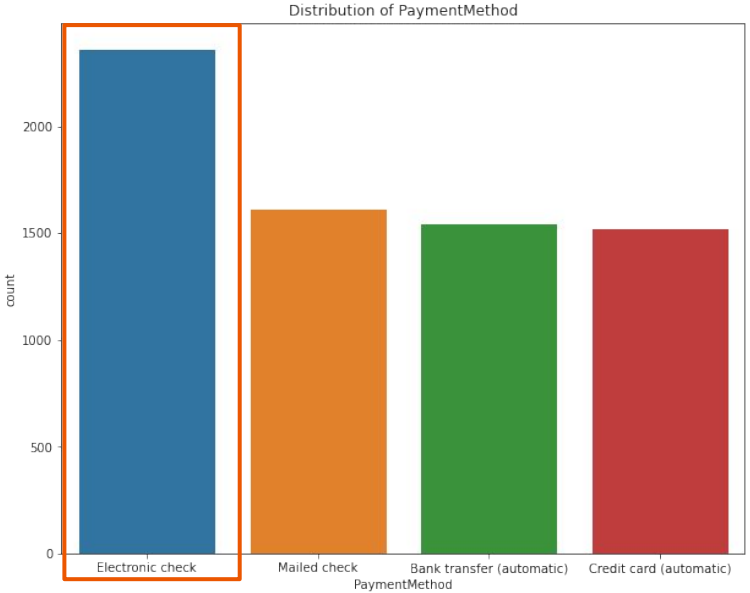


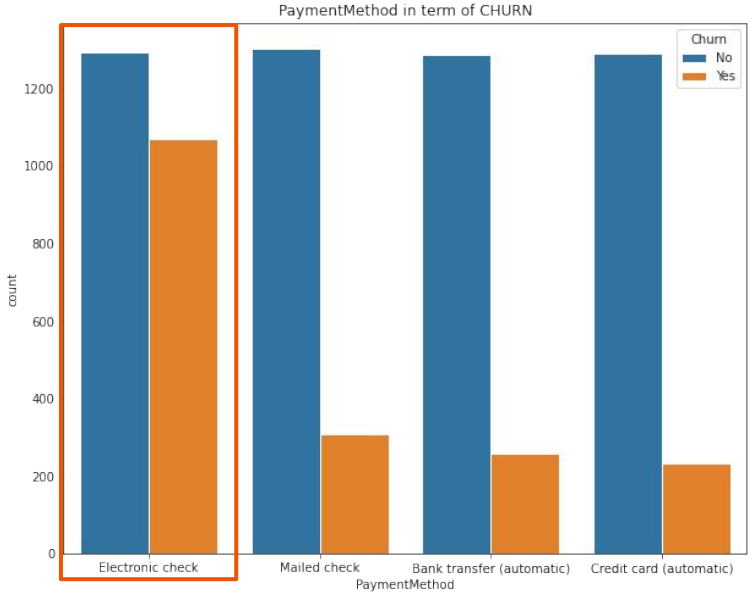
OUR CUSTOMERS



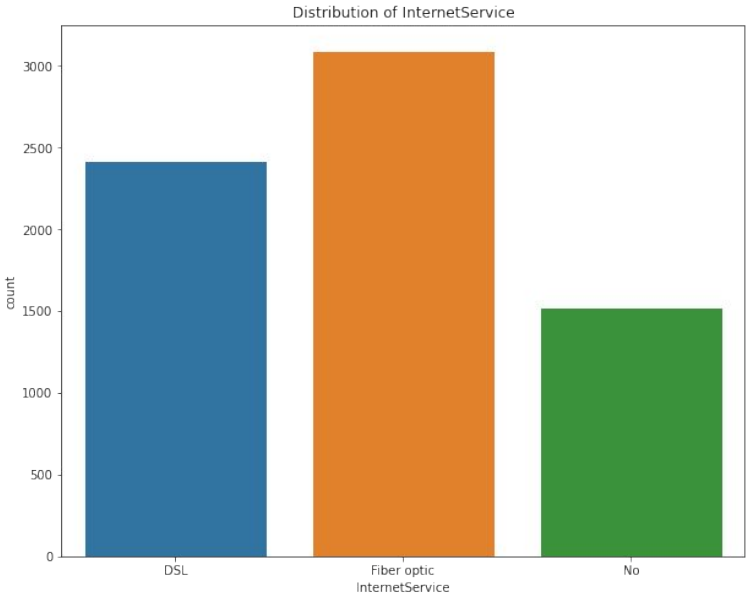


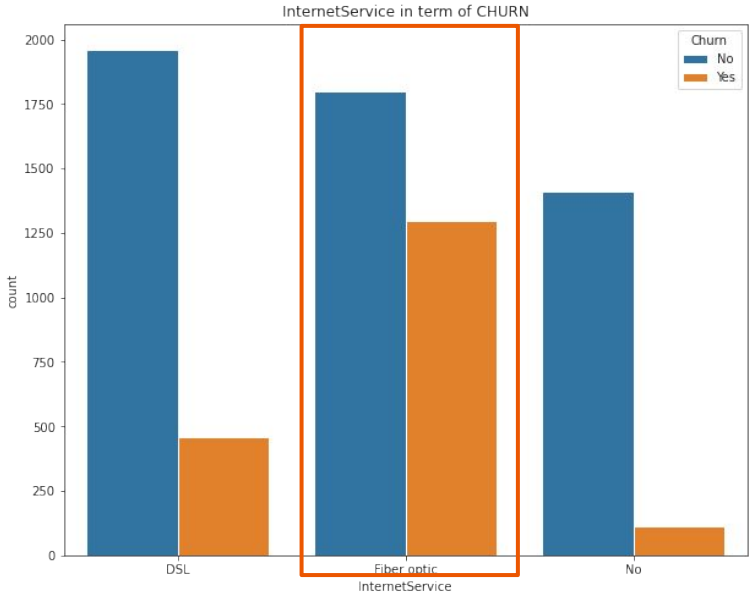
OUR CUSTOMERS



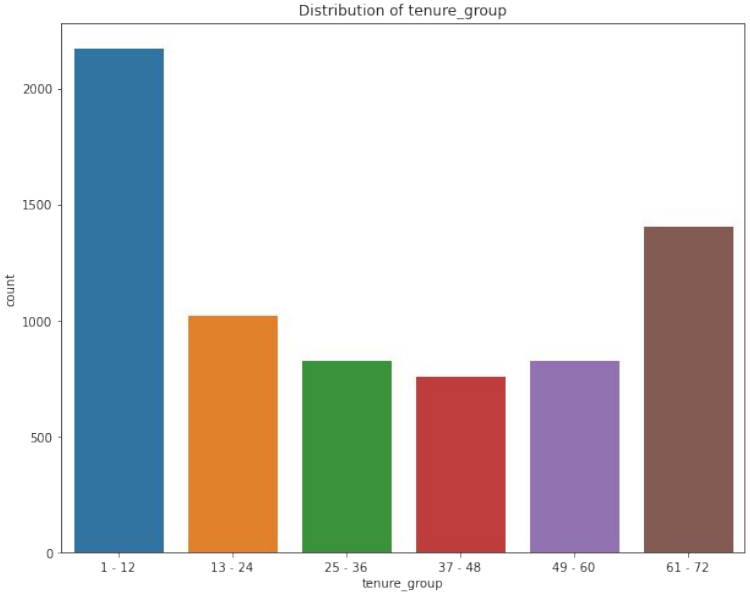


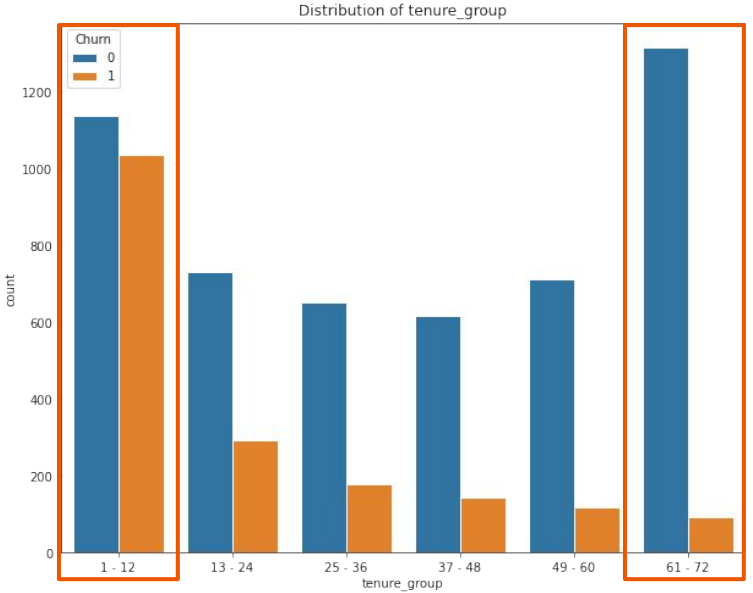
OUR CUSTOMERS



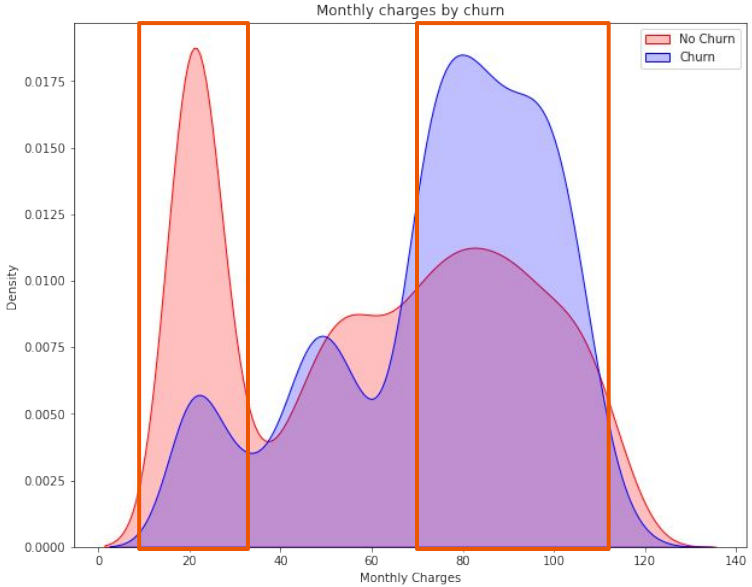


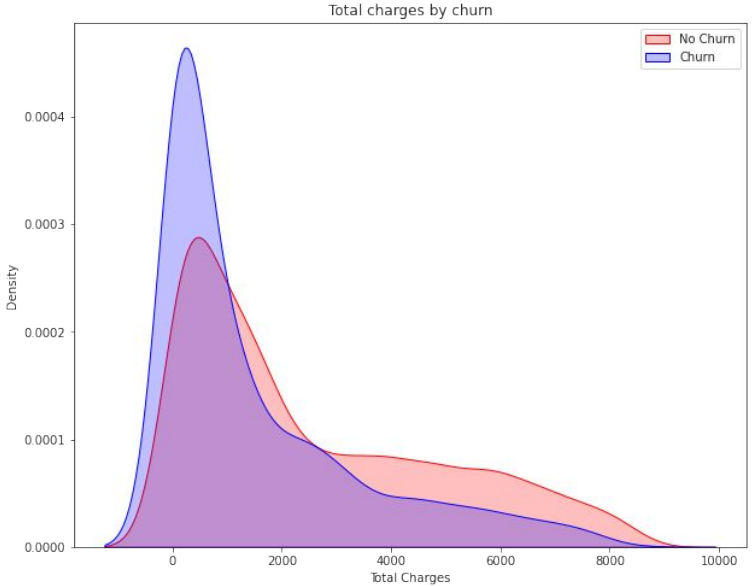
OUR CUSTOMERS





OUR CUSTOMERS



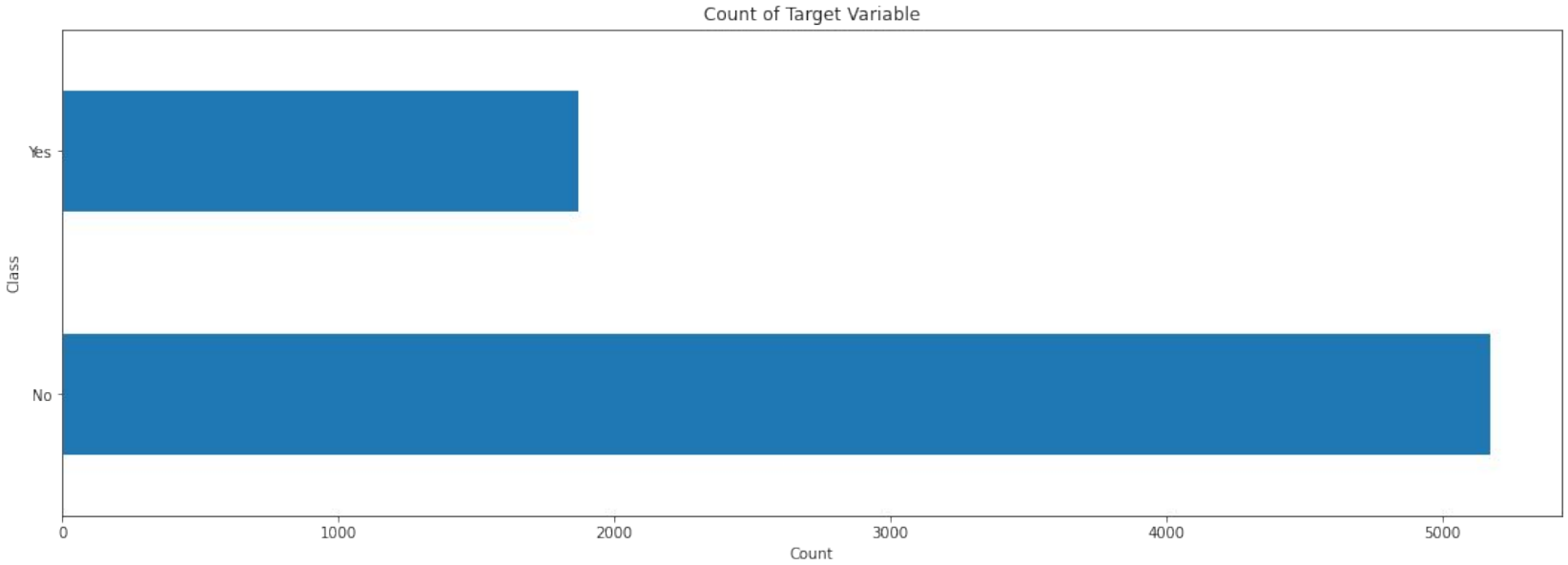


@coletangsy

PREDICTION MODEL

Going to build a machine learning model to predict it.

IMBALANCE TARGET CLASS



IMBALANCE TARGET CLASS

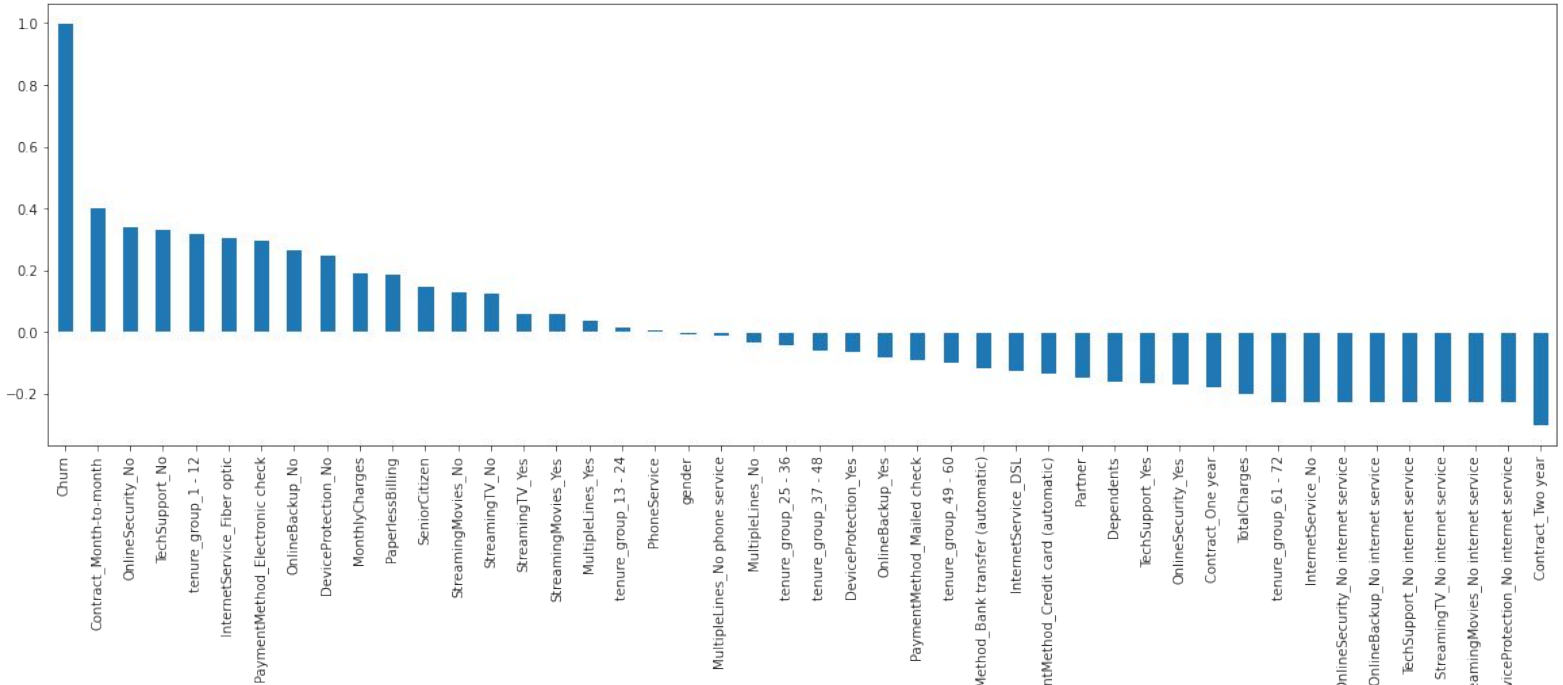
SMOTEENN()

- Combine over- and under-sampling using ***SMOTE*** (generate artificial data) and ***Edited Nearest Neighbours***.

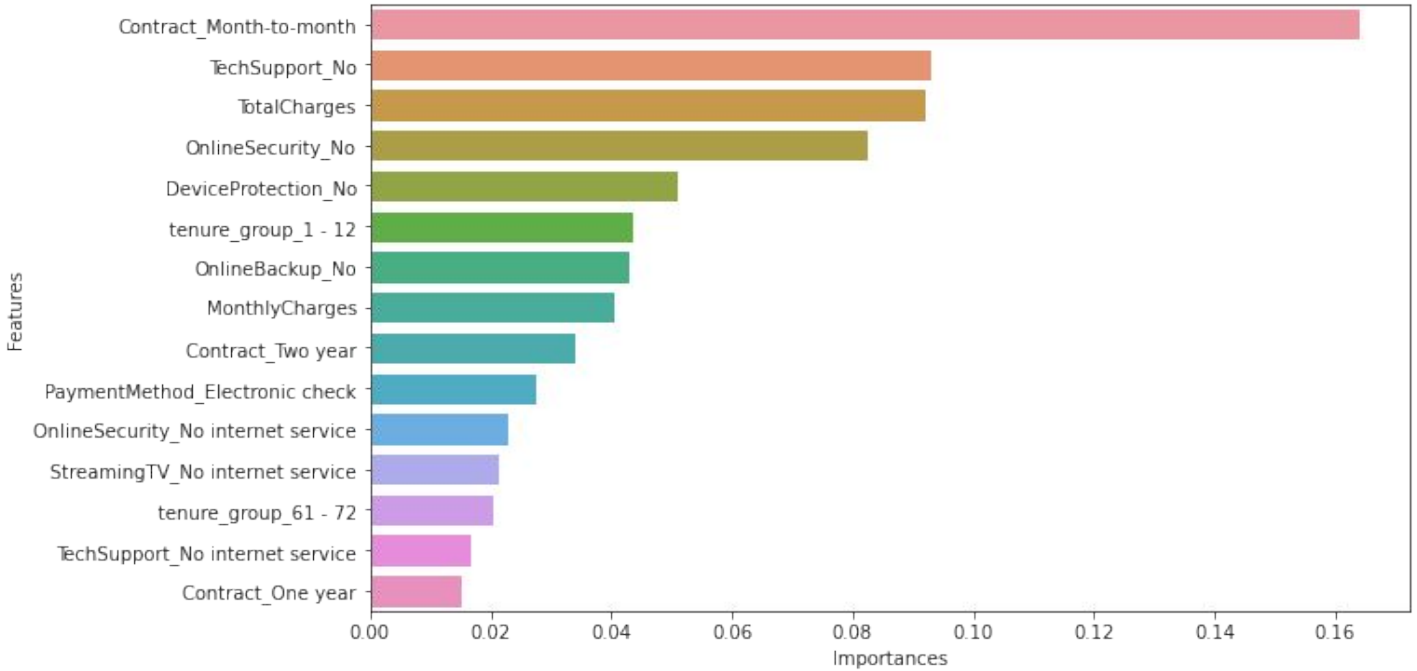
- Let the model learn ***study with the balanced training data*** and ***tested on the unbalance test data*** which is similar to the real situation.

```
from imblearn.combine import SMOTEENN
sm = SMOTEENN()
X_train_resampled, y_train_resampled = sm.fit_sample(X_train, y_train)
```

FEATURES SELECTION



FEATURES SELECTION



MODEL SELECTION (with lazypredict)

```
!pip install lazypredict

import lazypredict
from lazypredict.Supervised import LazyClassifier

clf = LazyClassifier(verbose=0,ignore_warnings=True, custom_metric=None)
models,predictions = clf.fit(X_train_resampled, X_test, y_train_resampled, y_test)


print(models)
```

MODEL SELECTION

Model	Accuracy	Balanced Accuracy	ROC AUC	F1 Score	Time Taken
AdaBoostClassifier	0.72	0.75	0.75	0.74	0.56
RandomForestClassifier	0.74	0.75	0.75	0.75	0.77
LogisticRegression	0.70	0.75	0.75	0.72	0.10
BaggingClassifier	0.75	0.75	0.75	0.76	0.35
CalibratedClassifierCV	0.70	0.74	0.74	0.72	2.00
ExtraTreesClassifier	0.74	0.74	0.74	0.75	0.59
SVC	0.73	0.74	0.74	0.74	0.73
LinearSVC	0.70	0.74	0.74	0.72	0.53
LGBMClassifier	0.73	0.74	0.74	0.75	0.86
LinearDiscriminantAnalysis	0.69	0.74	0.74	0.71	0.10
GaussianNB	0.70	0.74	0.74	0.72	0.04

Telecom-churn Model Building.ipynb

```
base_rf = RandomForestClassifier(n_estimators=200, random_state = 100)
base_rf.fit(X_train_resampled, y_train_resampled)

y_pred = base_rf.predict(X_test)

print(f"Model base score: {base_rf.score(X_test, y_test):.4}\n")
print(f"Model base ROC_AUC score: {roc_auc_score(y_test,y_pred):.4}\n")
print(classification_report(y_test, y_pred, labels=[0,1]))
```

MODEL BUILDING - RANDOM FOREST

	Accuracy	F1 score	**Recall**	Precision
base model	0.74	0.62	0.77	0.52
5 features	0.71	0.53	0.62	0.47
10 features	0.74	0.58	0.68	0.51
Final Tuned	0.74	0.57	0.82	0.44

PERFORMANCE EVALUATE

Going to build a machine learning model to predict it.

CONFUSION MATRIX

Actual Class		Predicted Class: STAY (0)	Predicted Class: CHURN (1)
	STAY (0)	TN	FP
	CHURN (1)	FN	TP

CONFUSION MATRIX

Actual Class	Predicted Class: STAY (0)	Predicted Class: CHURN (1)
STAY (0)	TN	FP
CHURN (1)	FN	TP

FP: I predicted he will leave, but he stayed

VS

FN: I predicted he will stay, but he leaved

CONFUSION MATRIX

Actual Class	Predicted Class: STAY (0)	Predicted Class: CHURN (1)
STAY (0)	TN	FP
CHURN (1)	FN	TP

Lucky...

FP: I predicted he will leave, but he stayed

VS

FN: I predicted he will stay, but he leaved

CONFUSION MATRIX

Actual Class	Predicted Class: STAY (0)	Predicted Class: CHURN (1)
STAY (0)	TN	FP
CHURN (1)	FN	TP

Accuracy

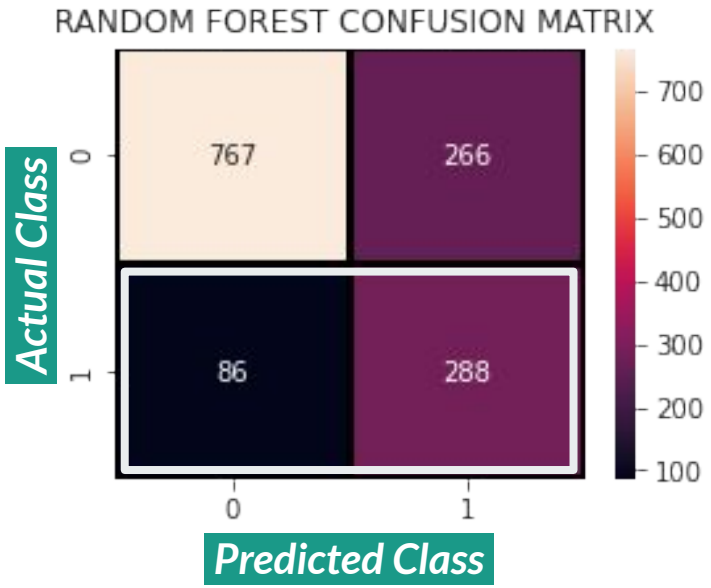
the testing data is imbalanced that can mislead the accuracy score.

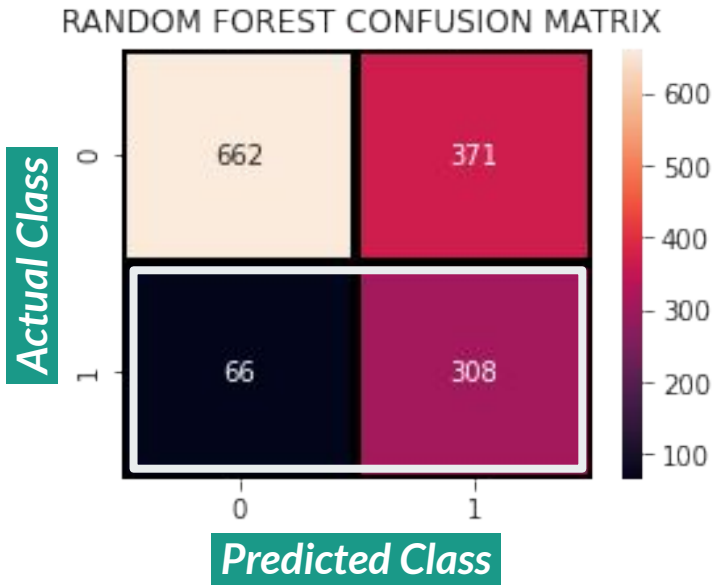
VS

Recall

How many churn case is correctly predicted

MODELS PERFORMANCES





-- Parameter Tuning -->

MODELS PERFORMANCES

	Accuracy	F1 score	**Recall**	Precision
base model	0.74	0.62	0.77	0.52
5 features	0.71	0.53	0.62	0.47
10 features	0.74	0.58	0.68	0.51
10 features (fine-tuned)	0.74	0.57	***0.82***	0.44

CONCLUSION

CUSTOMER CHARACTERISTICS

Customers who have the following characteristics,

- ***Having month-to-month contract;***
- ***Tenure within 1- 12 months;***
- ***Using electronic check to pay;***
- ***Do not use Online Security, Technical Support, Online Backup, Device Protection;***
- ***Using Fiber optic;***

are likely to churn.

BUSINESS VALUES

*"The cost of **acquiring a new customer** can be higher than that of retaining a customer by as much as **700%"***

*By increasing customer retention rates **5%** , we could **increase profits by 25% to 95%***

How to turn these into numbers, profit for boss?

if we can have 80% rate to predict the one who are gonna churn, we can target them

we can target them and offer more to retent them and which increasing our customer retention rates, and lower the cost of acquiring a new customer.

WHAT CAN WE DO

General

1. enhance ***Fiber optic*** services' quality;
2. Transfer ***Month-to-month*** customer into a longer contract term;
3. Review the current ***internet support services*** and develop new strategies

Target customer who are likely to churn

1. Develop ***specific offers*** for them to keep them with us;
2. Identify ***which kind of churn*** and the ***reason behind***

LIMITATION & NEXT STEP

HOW TO IMPROVE

1. ***Only apply Random Forest classifier***
    a. Apply ***more models*** and compare the tuned performance.

2. ***Cross check hyperparameter tuning process***
    a. Hyperparameter tuning with mircosoft auto-ml library ***FLAML***, to check is there any better way to fine-tuned the models

Q & A

END.

thank you for your time.