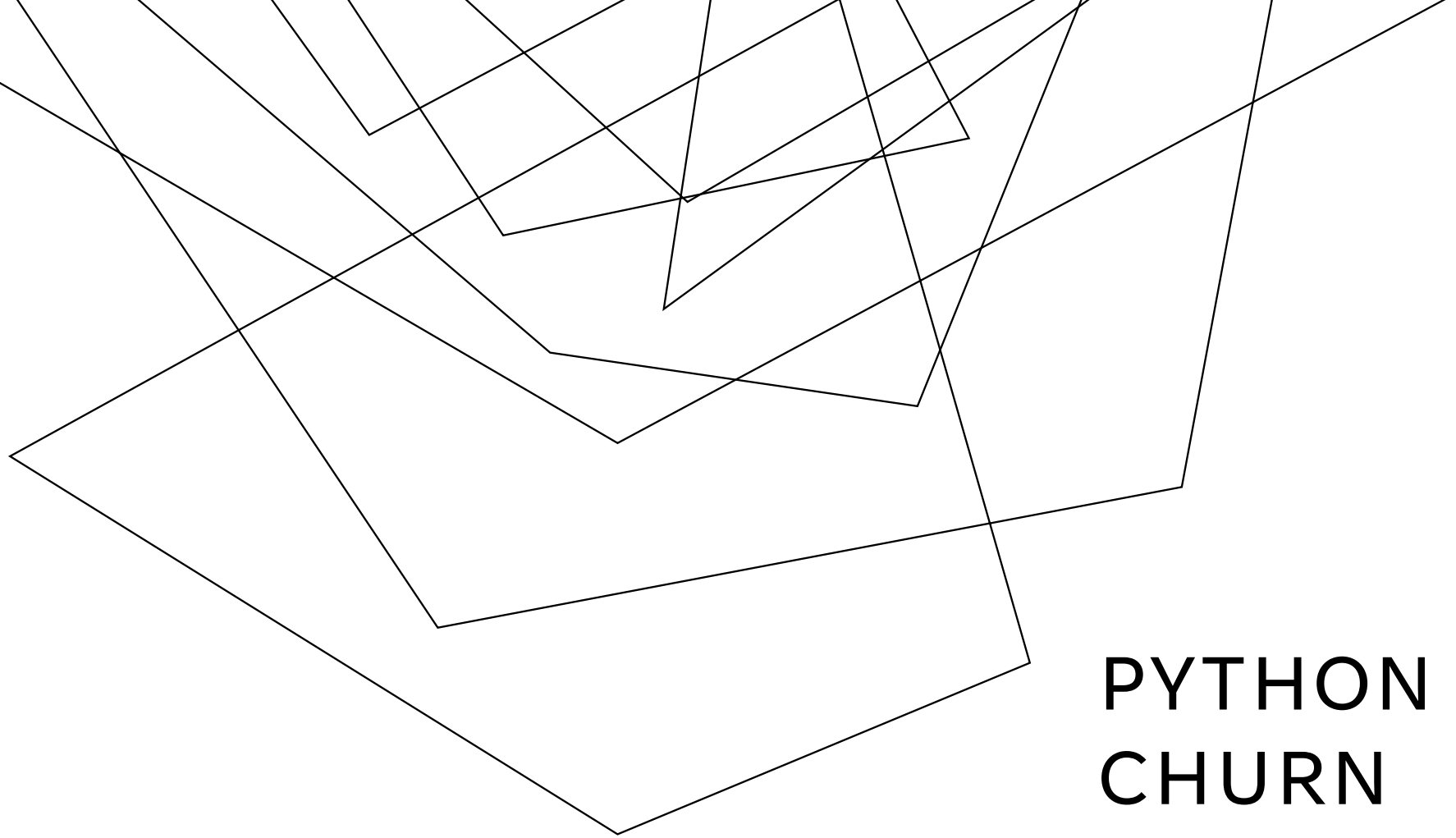


PYTHON PROJECT: CHURN

Presented by:

Margarita Dymshits

Natasha Shnaidman

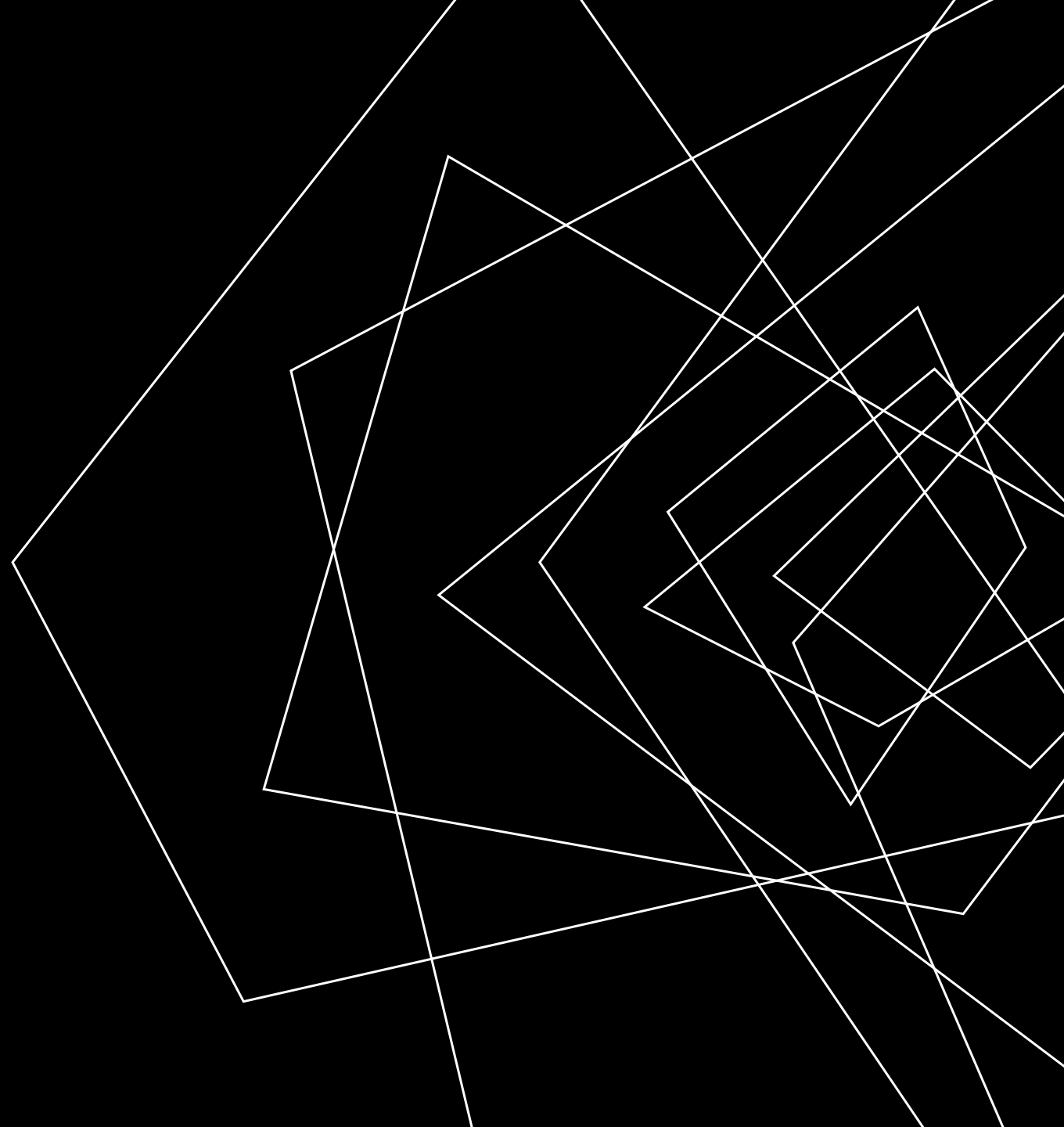
AGENDA

Introduction

Primary goals

Process

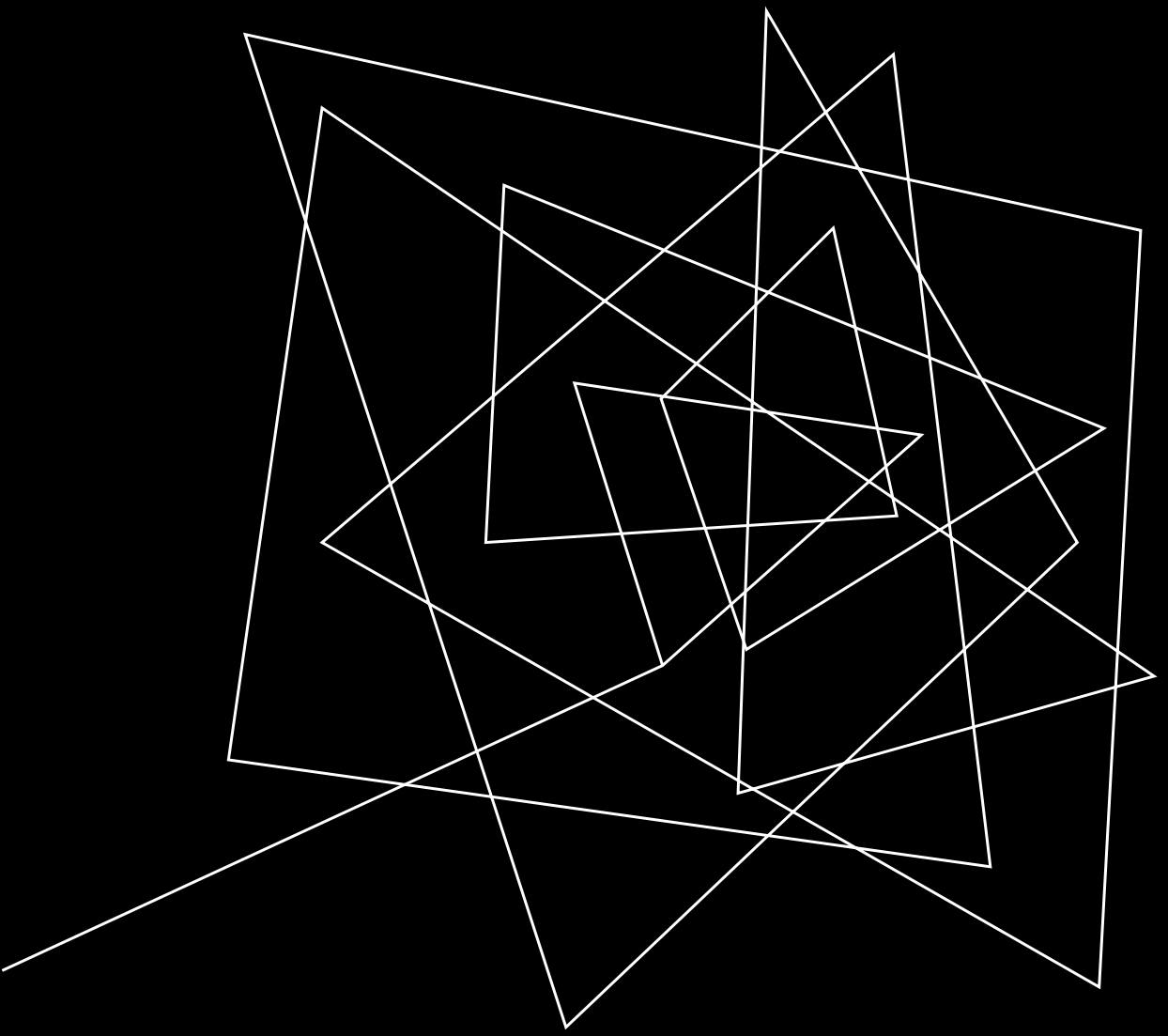
Summary



INTRODUCTION

"YOUR MOST UNHAPPY CUSTOMERS ARE YOUR GREATEST SOURCE OF LEARNING."

Bill Gates



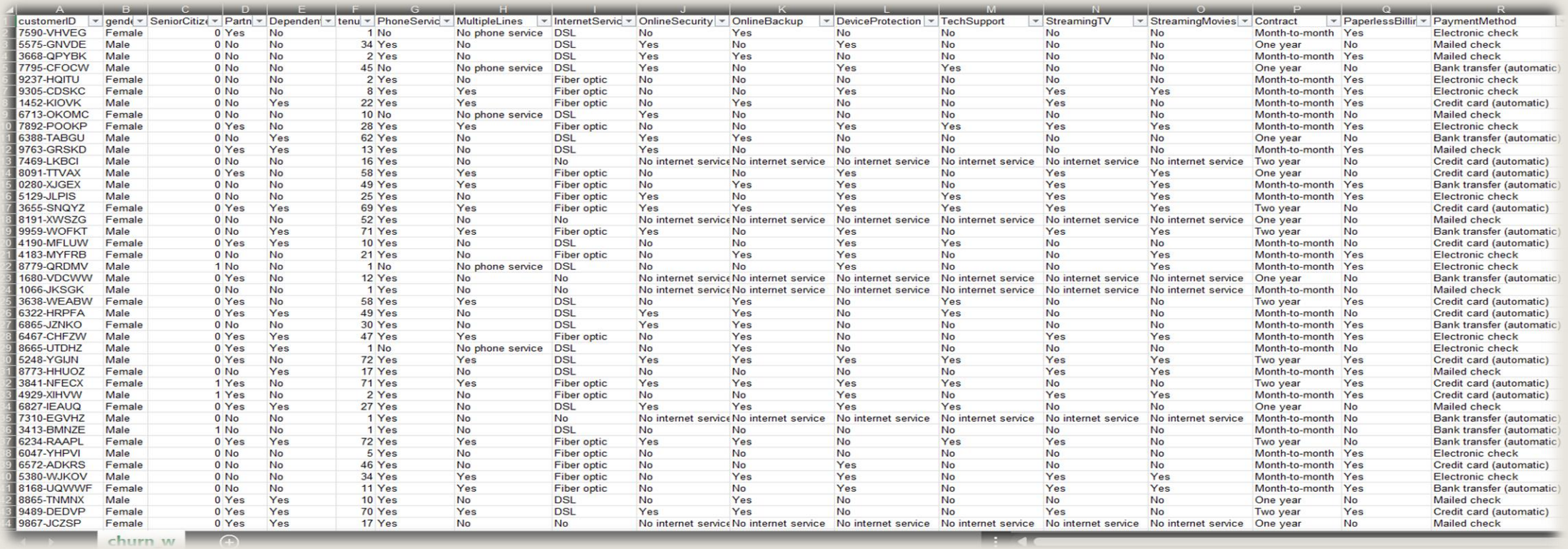
PRIMARY GOALS

Our main goal is to build a reliable prediction model, that will support company’s customer relationship department.

Our model will predict churn rate of existing clients. Churn rate is calculated from all possible information we have about our customers.

Importance of this model can hardly be overrated. Modern business world is built on principles of competitive market, where every company must fight for each customer. In this environment customer relations are crucial.

SO... THIS WAS OUR STARTING POINT



customerID	gende	SeniorCitize	Partn	Dependen	tenu	PhoneServic	MultipleLines	InternetServic	OnlineSecurity	OnlineBackup	DeviceProtection	TechSupport	StreamingTV	StreamingMovies	Contract	PaperlessBillir	PaymentMethod
7590-VHVEG	Female	0	Yes	No	1	No	No phone service	DSL	No	Yes	No	No	No	No	Month-to-month	Yes	Electronic check
5575-GNVDE	Male	0	No	No	34	Yes	No	DSL	Yes	No	Yes	No	No	No	One year	No	Mailed check
3668-QPYBK	Male	0	No	No	2	Yes	No	DSL	Yes	Yes	No	No	No	No	Month-to-month	Yes	Mailed check
7795-CFOCW	Male	0	No	No	45	No	No phone service	DSL	Yes	No	Yes	Yes	No	No	One year	No	Bank transfer (automatic)
9237-HQITU	Female	0	No	No	2	Yes	No	Fiber optic	No	No	No	No	No	No	Month-to-month	Yes	Electronic check
9305-CDSKC	Female	0	No	No	8	Yes	Yes	Fiber optic	No	No	Yes	No	Yes	Yes	Month-to-month	Yes	Electronic check
1452-KIOVK	Male	0	No	Yes	22	Yes	Yes	Fiber optic	No	Yes	No	No	Yes	No	Month-to-month	Yes	Credit card (automatic)
6713-OKOMC	Female	0	No	No	10	No	No phone service	DSL	Yes	No	No	No	No	No	Month-to-month	No	Mailed check
7892-POOKP	Female	0	Yes	No	28	Yes	Yes	Fiber optic	No	No	Yes	Yes	Yes	Yes	Month-to-month	Yes	Electronic check
6388-TABGU	Male	0	No	Yes	62	Yes	No	DSL	Yes	Yes	No	No	No	No	One year	No	Bank transfer (automatic)
9763-GRSKD	Male	0	Yes	Yes	13	Yes	No	DSL	Yes	No	No	No	No	No	Month-to-month	Yes	Mailed check
7469-LKBCI	Male	0	No	No	16	Yes	No	No	No internet service	No internet service	No internet service	No internet service	No internet service	No internet service	Two year	No	Credit card (automatic)
8091-TTVAX	Male	0	Yes	No	58	Yes	Yes	Fiber optic	No	No	Yes	No	Yes	Yes	One year	No	Credit card (automatic)
0280-XJGEX	Male	0	No	No	49	Yes	Yes	Fiber optic	No	Yes	Yes	No	Yes	Yes	Month-to-month	Yes	Bank transfer (automatic)
5129-JLPIS	Male	0	No	No	25	Yes	No	Fiber optic	Yes	No	Yes	Yes	Yes	Yes	Month-to-month	Yes	Electronic check
3655-SNQYZ	Female	0	Yes	Yes	69	Yes	Yes	Fiber optic	Yes	Yes	Yes	Yes	Yes	Yes	Two year	No	Credit card (automatic)
8191-XWSZG	Female	0	No	No	52	Yes	No	No	No internet service	No internet service	No internet service	No internet service	No internet service	No internet service	One year	No	Mailed check
9959-WOFKT	Male	0	No	Yes	71	Yes	Yes	Fiber optic	Yes	No	Yes	No	Yes	Yes	Two year	No	Bank transfer (automatic)
4190-MFLUW	Female	0	Yes	Yes	10	Yes	No	DSL	No	No	Yes	Yes	No	No	Month-to-month	No	Credit card (automatic)
4183-MYFRB	Female	0	No	No	21	Yes	No	Fiber optic	No	Yes	Yes	No	No	Yes	Month-to-month	Yes	Electronic check
8779-QRDMV	Male	1	No	No	1	No	No phone service	DSL	No	No	Yes	No	No	Yes	Month-to-month	Yes	Electronic check
1680-VDCWW	Male	0	Yes	No	12	Yes	No	No	No internet service	No internet service	No internet service	No internet service	No internet service	No internet service	One year	No	Bank transfer (automatic)
1066-JKSGK	Male	0	No	No	1	Yes	No	No	No internet service	No internet service	No internet service	No internet service	No internet service	No internet service	Month-to-month	No	Mailed check
3638-WEABW	Female	0	Yes	No	58	Yes	Yes	DSL	No	Yes	No	Yes	No	No	Two year	Yes	Credit card (automatic)
6322-HRPFA	Male	0	Yes	Yes	49	Yes	No	DSL	Yes	Yes	No	Yes	No	No	Month-to-month	No	Credit card (automatic)
6865-JZNKO	Female	0	No	No	30	Yes	No	DSL	Yes	Yes	No	No	No	No	Month-to-month	Yes	Bank transfer (automatic)
6467-CHFZW	Male	0	Yes	Yes	47	Yes	Yes	Fiber optic	No	Yes	No	No	Yes	Yes	Month-to-month	Yes	Electronic check
8665-UTDHZ	Male	0	Yes	Yes	1	No	No phone service	DSL	No	Yes	No	No	No	No	Month-to-month	No	Electronic check
5248-YGIJN	Male	0	Yes	No	72	Yes	Yes	DSL	Yes	Yes	Yes	Yes	Yes	Yes	Two year	Yes	Credit card (automatic)
8773-HHUOZ	Female	0	No	Yes	17	Yes	No	DSL	No	No	No	No	Yes	Yes	Month-to-month	Yes	Mailed check
3841-NFECX	Female	1	Yes	No	71	Yes	Yes	Fiber optic	Yes	Yes	Yes	Yes	No	No	Two year	Yes	Credit card (automatic)
4929-XIHVW	Male	1	Yes	No	2	Yes	No	Fiber optic	No	No	Yes	No	Yes	Yes	Month-to-month	Yes	Credit card (automatic)
6827-IEAUQ	Female	0	Yes	Yes	27	Yes	No	DSL	Yes	Yes	Yes	Yes	No	No	One year	No	Mailed check
7310-EGVHZ	Male	0	No	No	1	Yes	No	No	No internet service	No internet service	No internet service	No internet service	No internet service	No internet service	Month-to-month	No	Bank transfer (automatic)
3413-BMNZE	Male	1	No	No	1	Yes	No	DSL	No	No	No	No	No	No	Month-to-month	No	Bank transfer (automatic)
6234-RAAPL	Female	0	Yes	Yes	72	Yes	Yes	Fiber optic	Yes	Yes	No	Yes	Yes	No	Two year	No	Bank transfer (automatic)
6047-YHPVI	Male	0	No	No	5	Yes	No	Fiber optic	No	No	No	No	No	No	Month-to-month	Yes	Electronic check
6572-ADKRS	Female	0	No	No	46	Yes	No	Fiber optic	No	No	Yes	No	No	No	Month-to-month	Yes	Credit card (automatic)
5380-WJKOV	Male	0	No	No	34	Yes	Yes	Fiber optic	No	Yes	Yes	No	Yes	Yes	Month-to-month	Yes	Electronic check
8168-UQWWF	Female	0	No	No	11	Yes	Yes	Fiber optic	No	No	Yes	No	Yes	Yes	Month-to-month	Yes	Bank transfer (automatic)
8865-TNMNX	Male	0	Yes	Yes	10	Yes	No	DSL	No	Yes	No	No	No	No	One year	No	Mailed check
9489-DEDVP	Female	0	Yes	Yes	70	Yes	Yes	DSL	Yes	Yes	No	No	Yes	No	Two year	Yes	Credit card (automatic)
9867-JCZSP	Female	0	Yes	Yes	17	Yes	No	No	No internet service	No internet service	No internet service	No internet service	No internet service	No internet service	One year	No	Mailed check

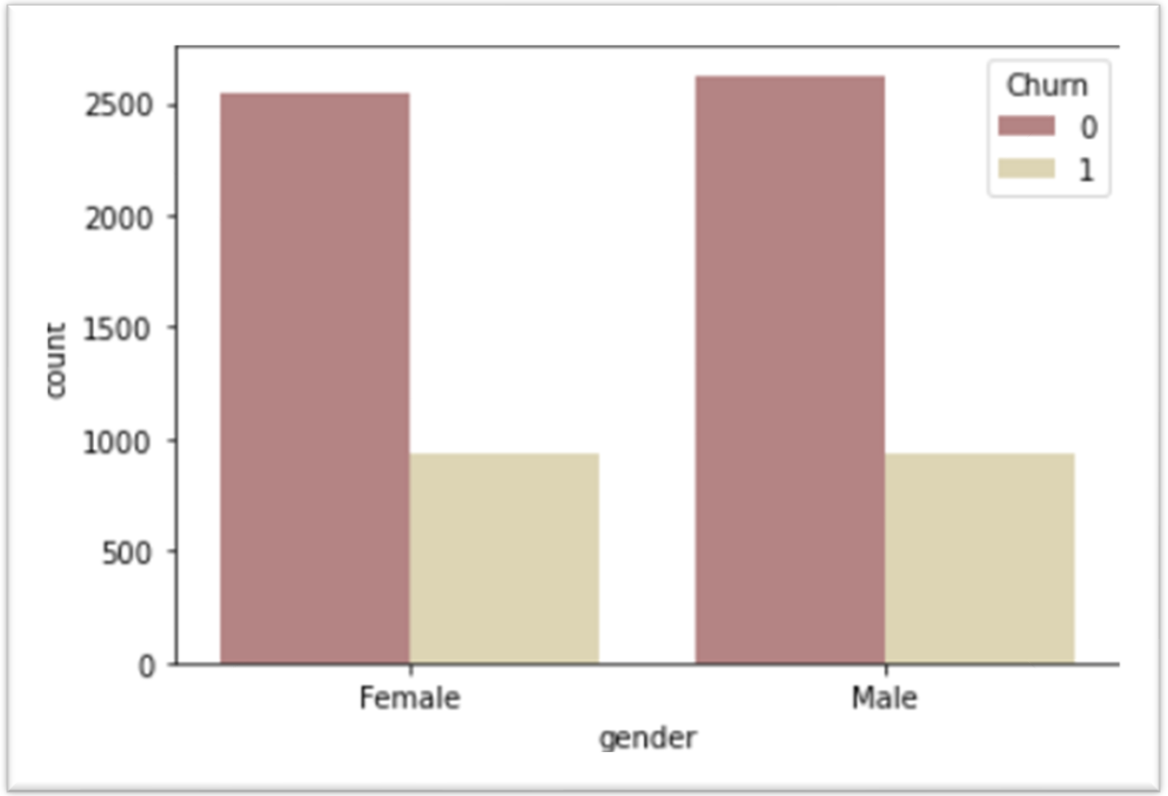
DATA DESCRIPTION AND FIRST CONCLUSIONS

General Data Characteristics	
Number of rows/ customers	7043
Number of features	21
Customers churned	1869
Types of values	object, int, float
Churn rate	26.5%
features	can be grouped by theme

EDA

Gender

Surprisingly, gender does not have a strong impact on churn rate. Churn rate is almost the same both in females and in males.



EDA

Gender & Seniority

GENDER	SENIOR CITIZEN	CHURN
FEMALE	YES	0.29
	NO	**0.42**
MALE	YES	0.23
	NO	**0.41**

Senior citizens churn less, both females and males.

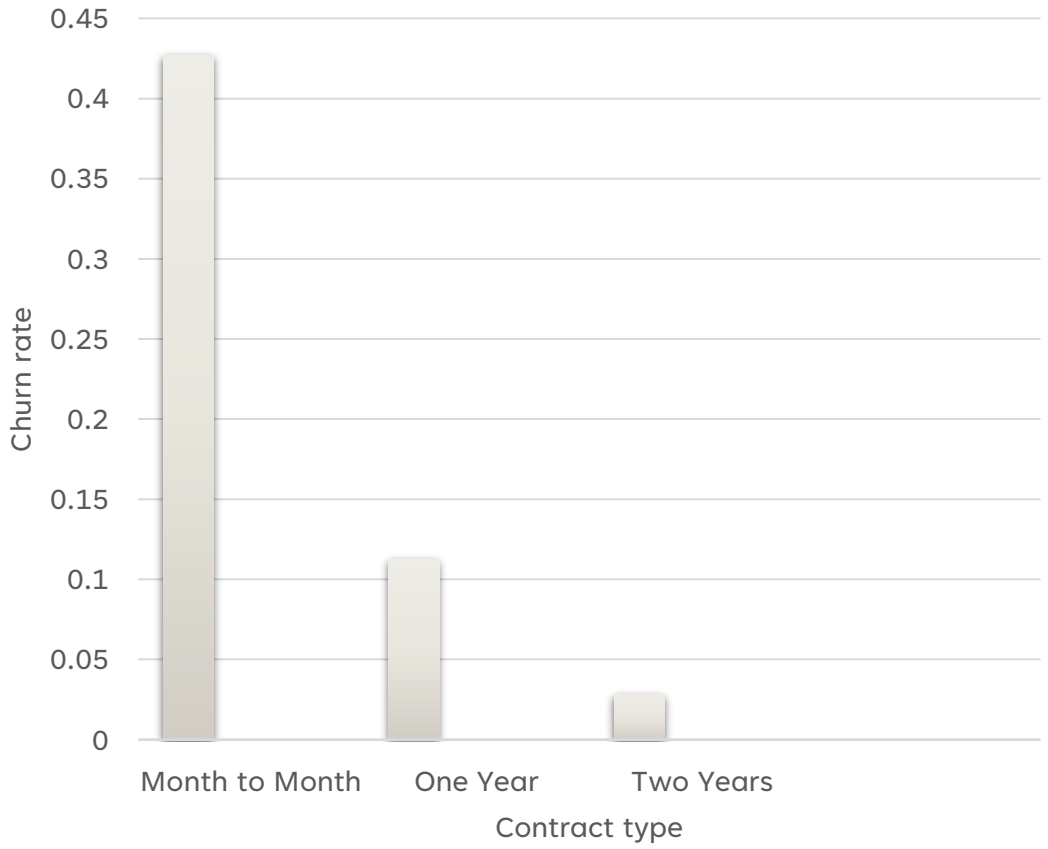
EDA

New feature: Seniority & Partnership (“Loneliness”)

	CHURN
Lonely Senior Citizen	**0.488**
Not Lonely Senior Citizen	0.345
Not Senior Citizen	0.236

Senior Citizens that have no partner churn more.

EDA

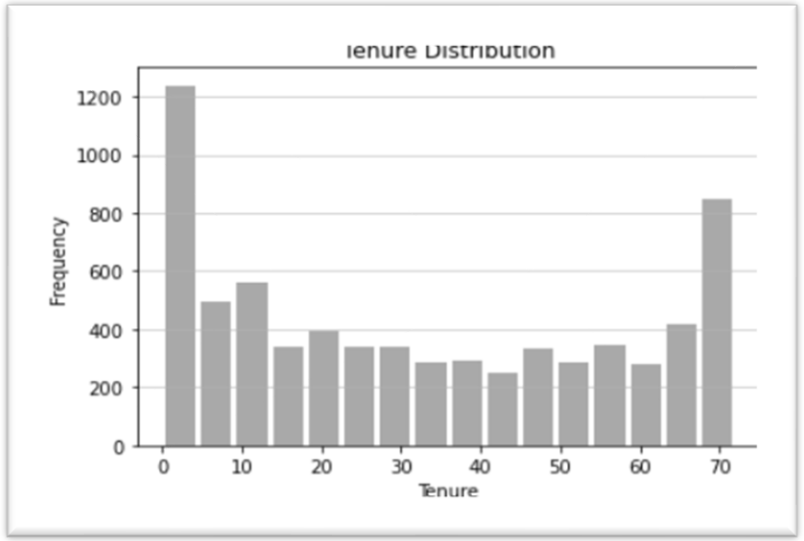


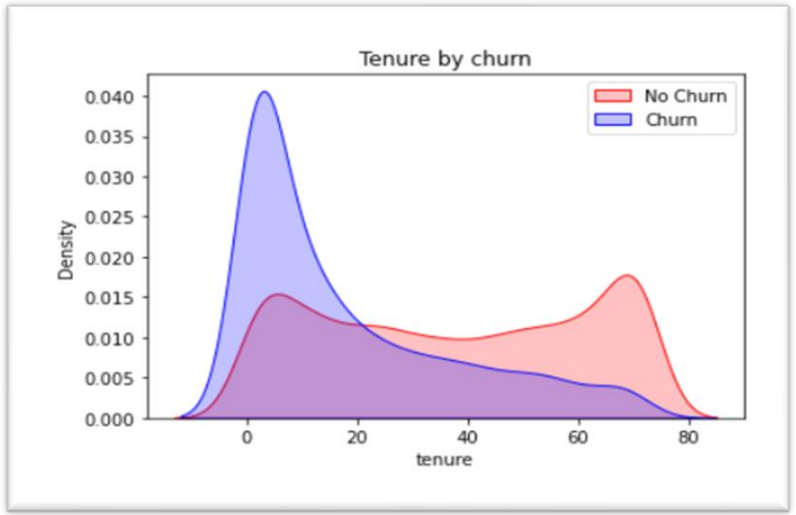
Contract type

Churn rate decreases as contract period increases.

EDA

New Feature: Tenure Group



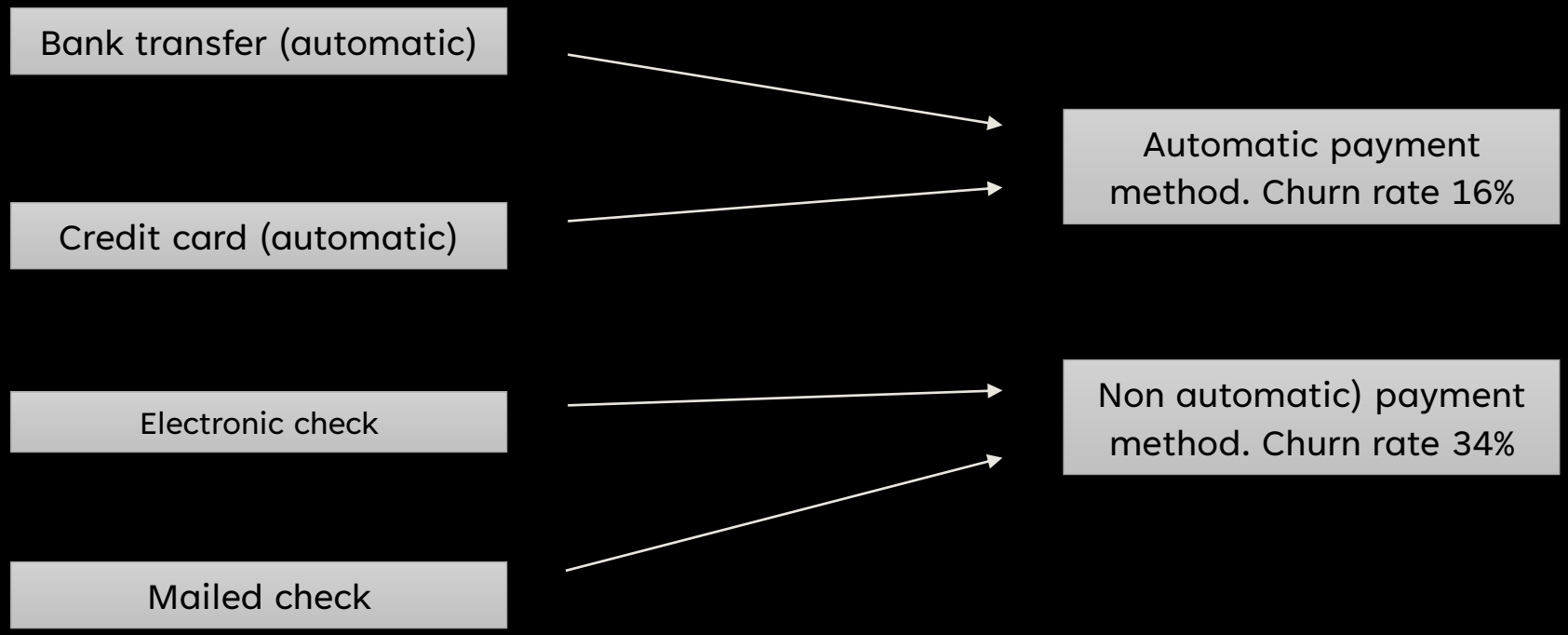


Tenure Group allows to see that highest risk of churn is in first 6 month: the longer the customer staying, the lower churn risk.

- Tenure 0-6 months - Churn rate 0.52
- Tenure 6-12 month - Churn rate 0.35
- Tenure up to 2 years - Churn rate 0.28
- Tenure up to 4 years - Churn rate 0.2
- Tenure over 4 years - Churn rate 0.09

EDA

New feature: Payment method



Automatic way of payments decreases customer’s churn.

EDA

Automatic Payments & Paperless Billing

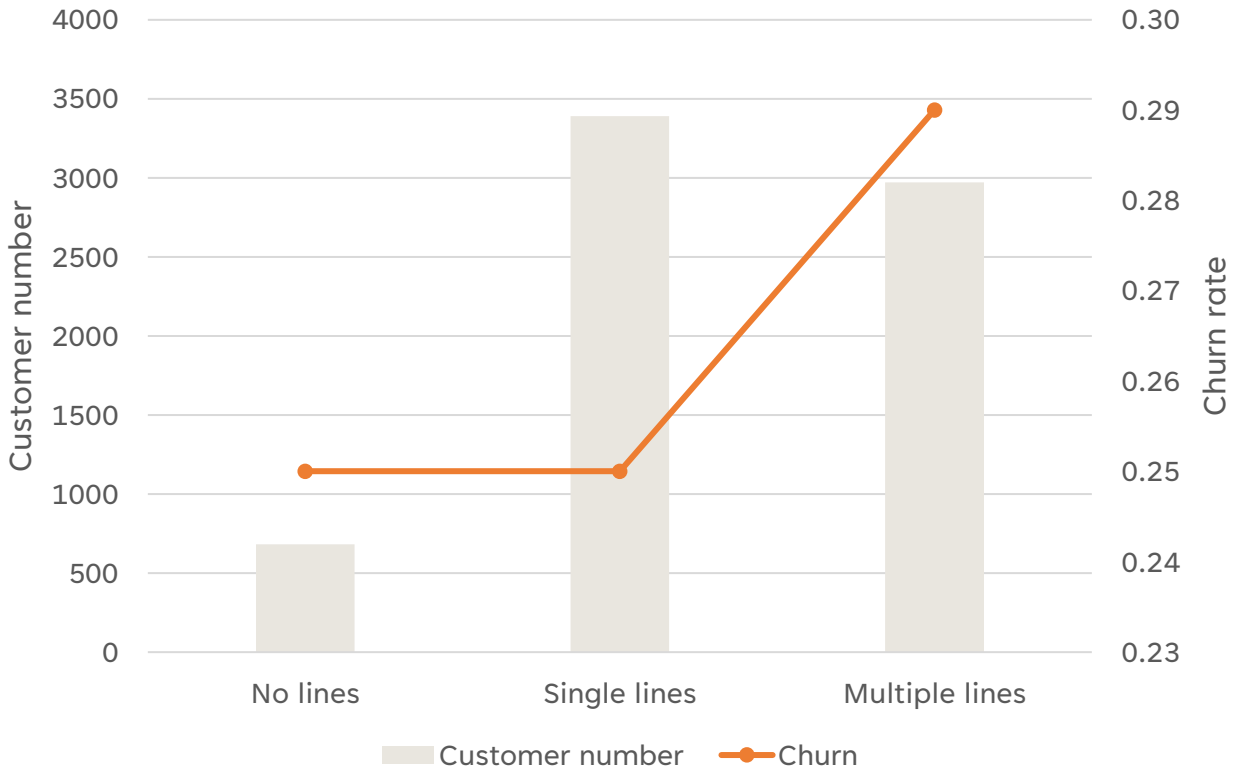
PAYMENT METHOD	PAPER BILLING	CHURN
AUTOMATIC	YES	0.21
	NO	**0.43**
NON-AUTOMATIC	YES	0.1
	NO	0.2

Automatic payments method and paper billing reduce churn rate.

EDA

New Feature: Phone Service New

Distribution of phone services by line number. Customer with multiple lines tend to leave slightly more then those with single line.



EDA

NEW FEATURE: SERVICE CATEGORY

PHONE SERVICES

Phone services

Churn Rate 0.07

INTERNET SERVICES

Internet services include:

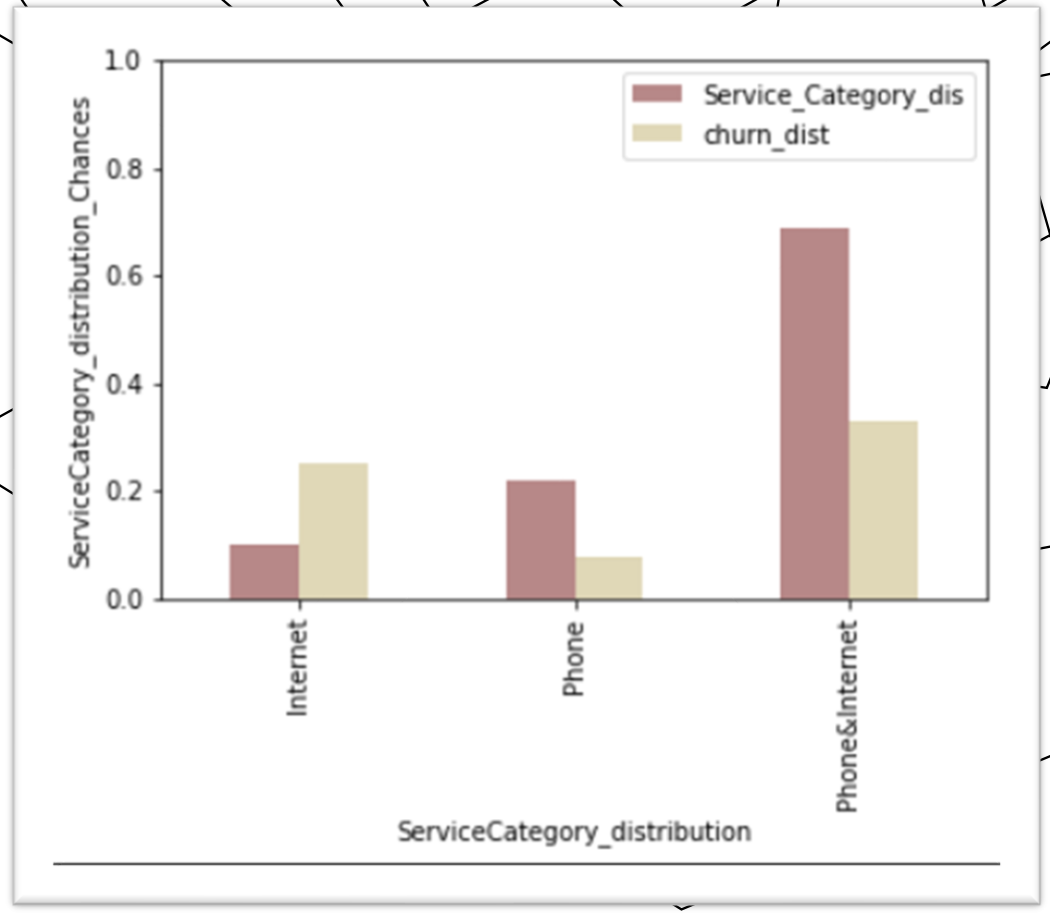
1. DSL
2. Fiber optic

Total Churn rate for internet services 0.38

PHONE & INTERNET

Bundling. Phone and internet at the same time

Churn Rate 0.16



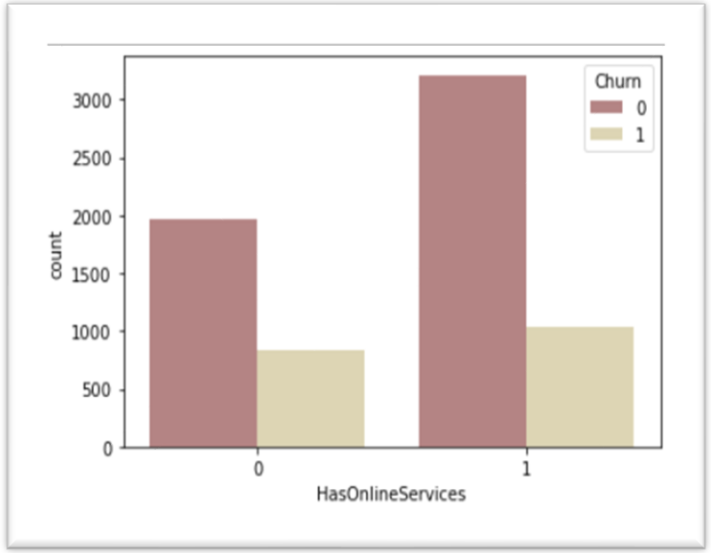
CHURN BY SERVICE CATEGORY

Customers tend to stay longer with phone services. Bundling phone and internet services will cause a significant raise in income and minor raise in churn rate.

INTERNET SERVICES DIVE IN
NEW FEATURE: HAS ONLINE SERVICES
NEW FEATURE: HAS STREAMING SERVICES

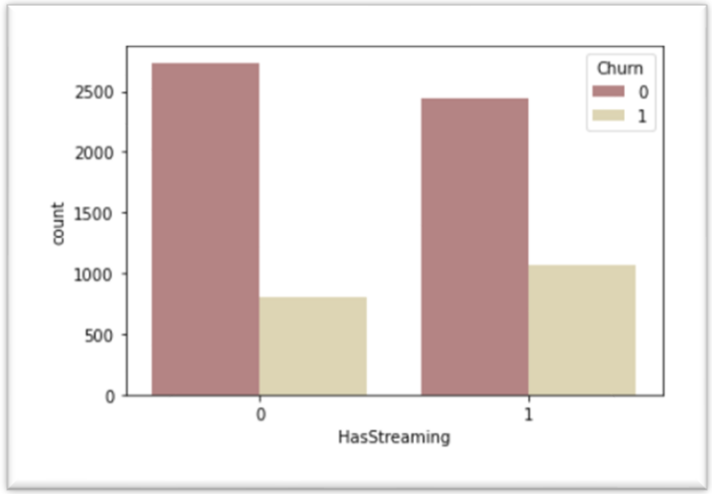
"Has online services" – customers that purchased at list one of internet services:

- online backup
- online security
- device protection
- tech support



"Has streaming services" – customers that purchased at least one of streaming services:

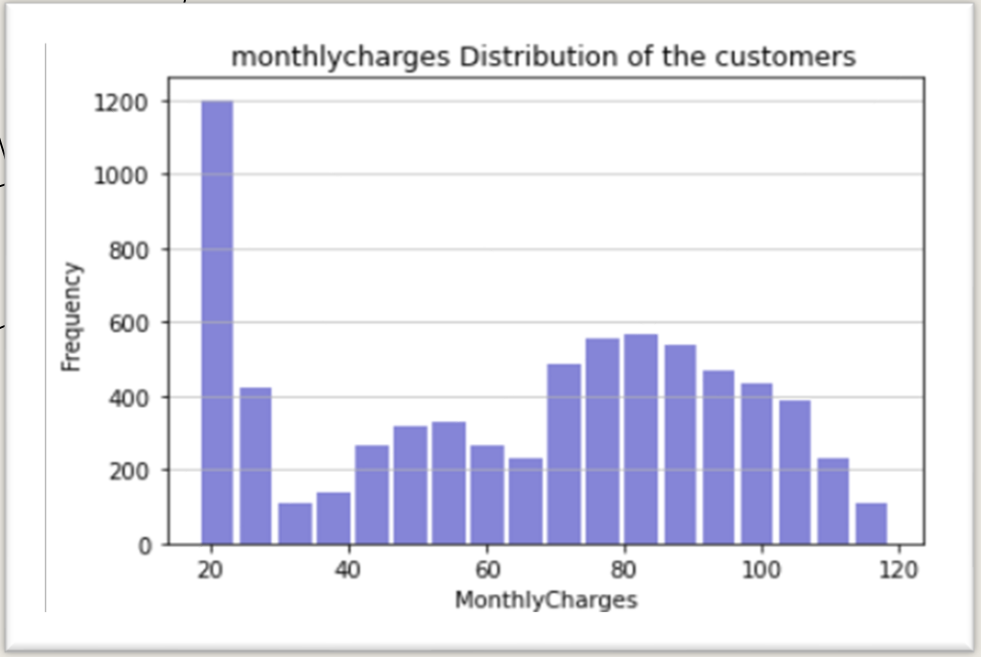
- streaming TV
- streaming movies

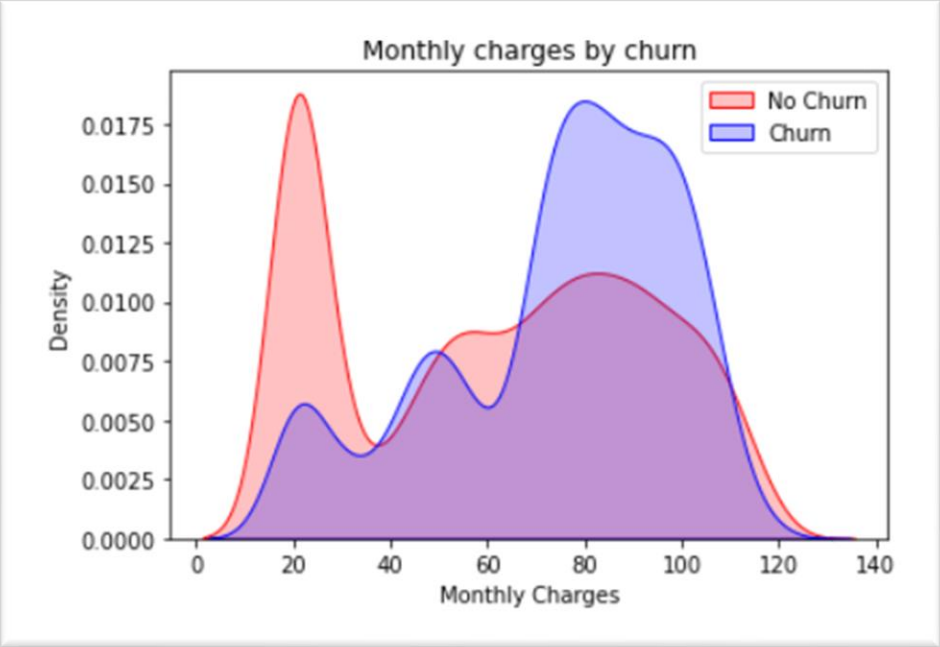


Online services and streaming services do not impact churn rate dramatically

EDA

Total Charge & Monthly Charges





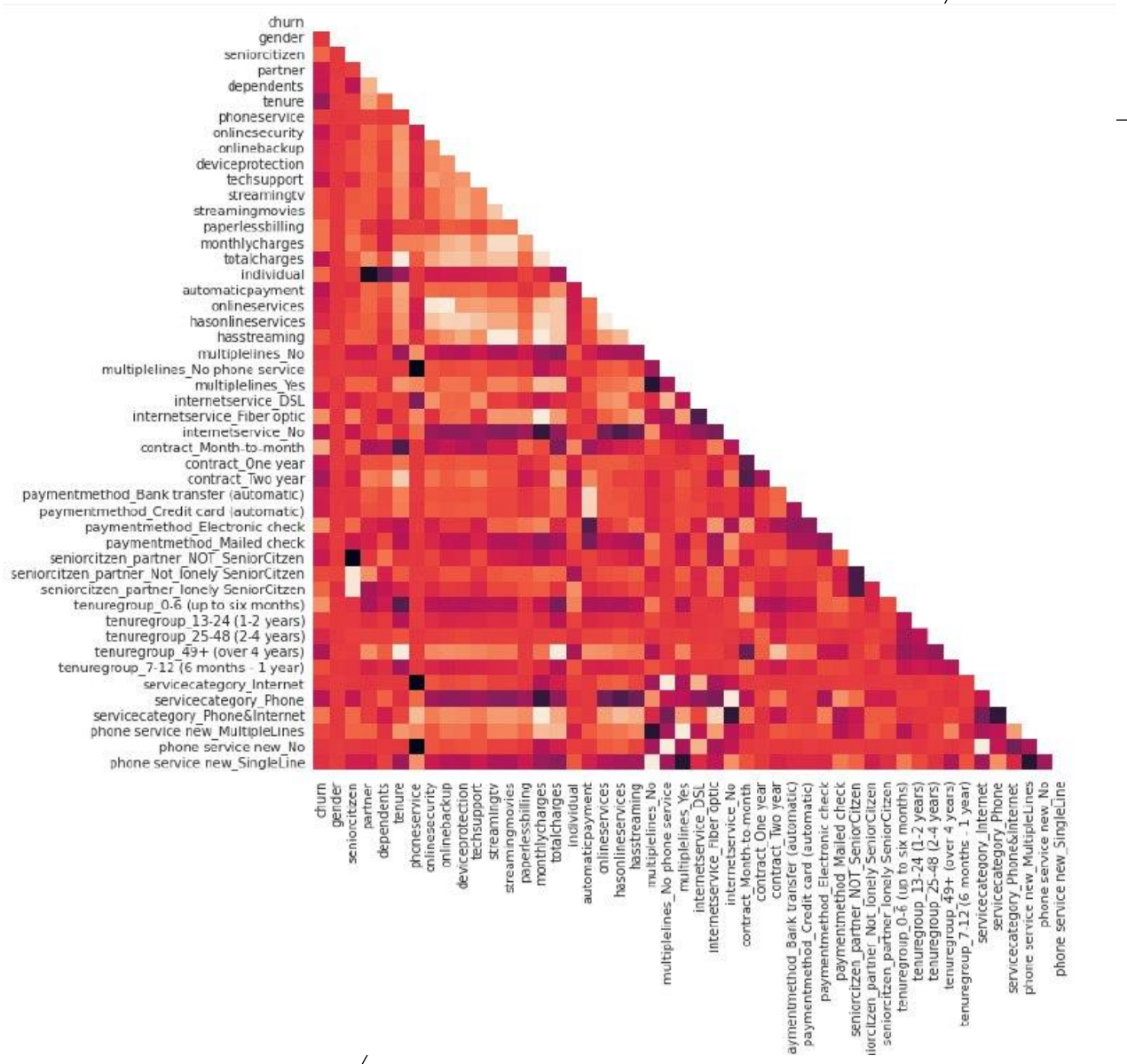
Had to fix data in Total charges that had hidden nulls.

Churn and monthly charges are correlated. When monthly charge increases, chur increases as well.

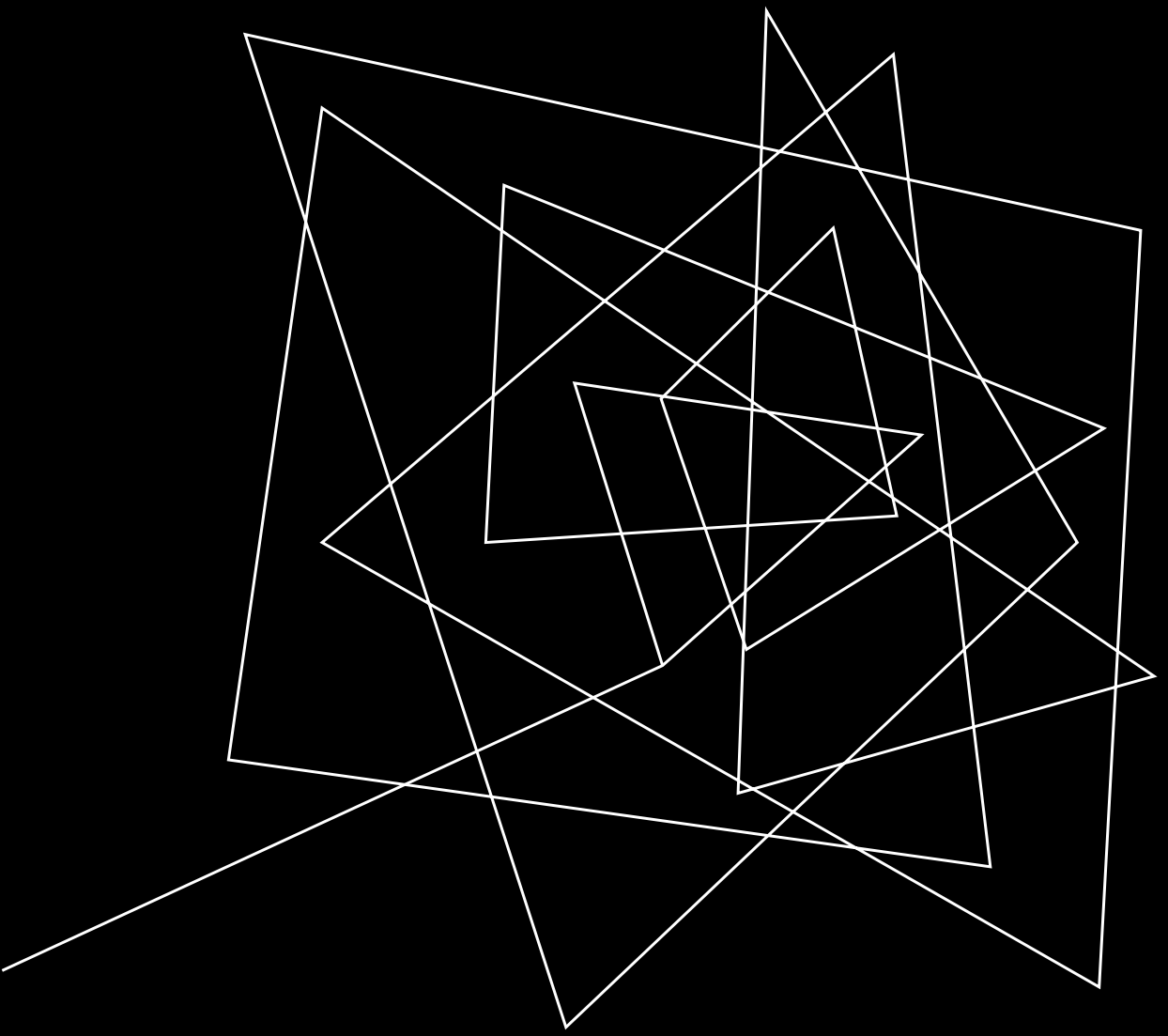
DATA CLEANING

1. Removed Customer id, although we did see that there are 2 parts in each customer ID: letters part and numbers part. We noticed that while letters parts are unique part, some number parts are repeating. If we had some more time, we could make some new feature out of it.

2. Calculated the correlations. Tenure, Contract are more correlated with churn.



MACHINE LEARNING



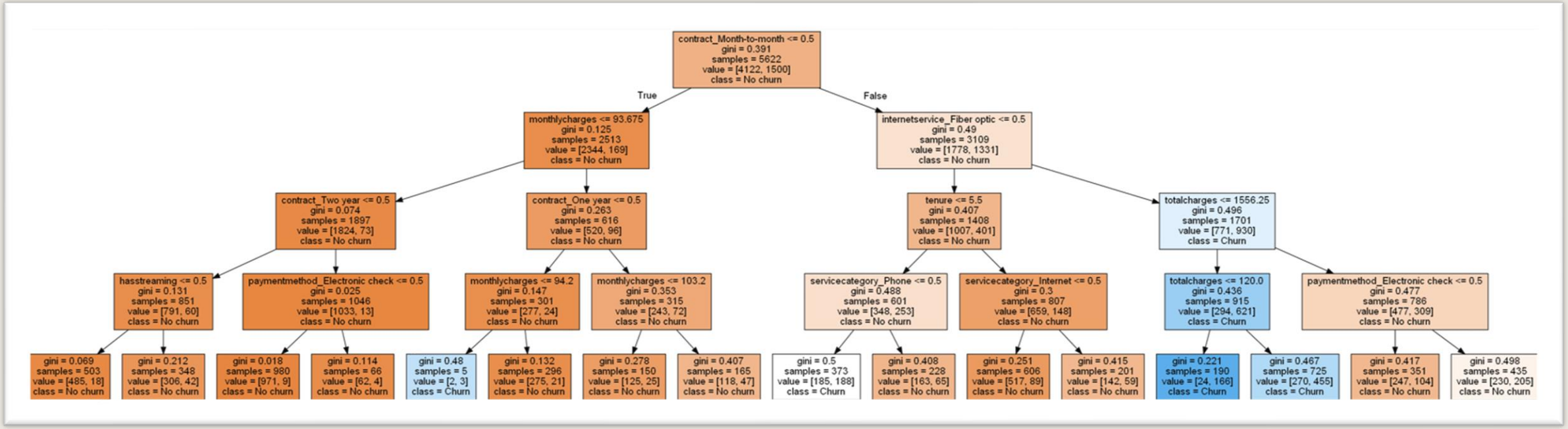
SPLIT THE DATA TO TRAIN AND TEST SETS

TRAIN (80%) – 5622 ROWS
TEST (20%) – 1410 ROWS

NOW WE CAN BUILD OUR PREDICTIONS MODELS:

1. DECISION TREE
2. RANDOM FOREST
3. KNN

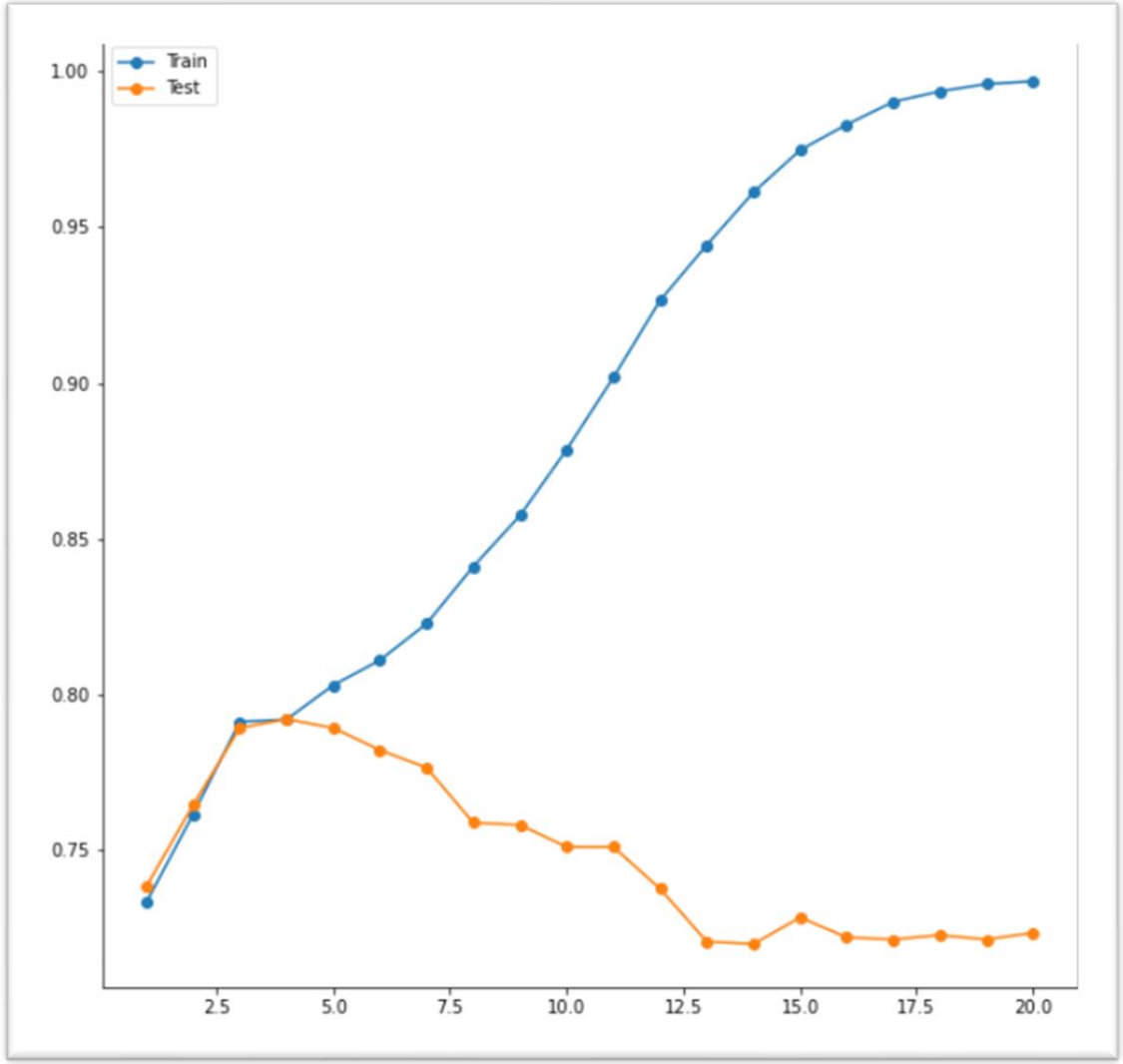
DECISION TREE



MAX DEPTH – 4 (CALCULATED)

ACCURACY – 0.792

DECISION TREE – CALCULATION OF HIPER PARAMETERS

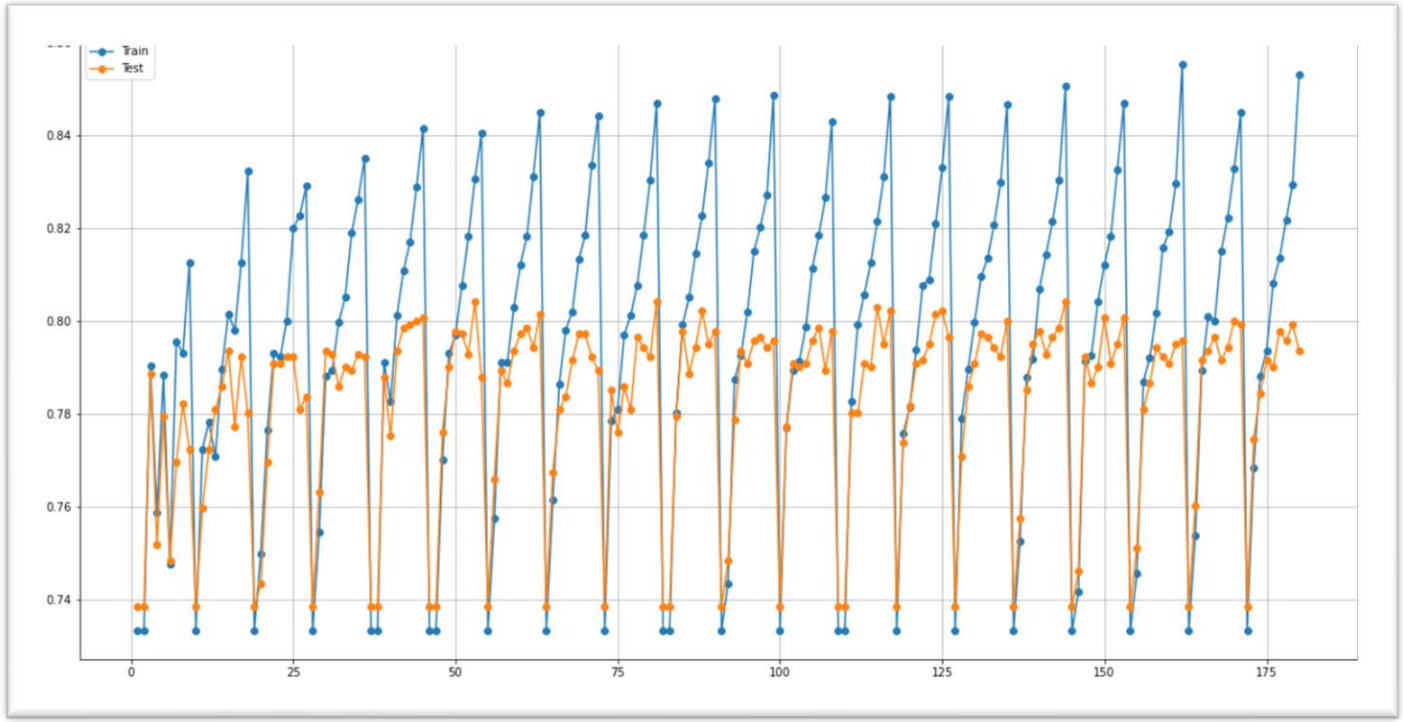


IN ORDER TO FIND BEST MAX DEPTH OF THE TREE WE RAN A LOOP THAT CALCULATES ACCURACY FOR ALL DEPTH LEVELS FROM 1 TO 20 ON BOTH TRAIN AND TEST SETS. THEN WE FOUND THE OPTIMUM.

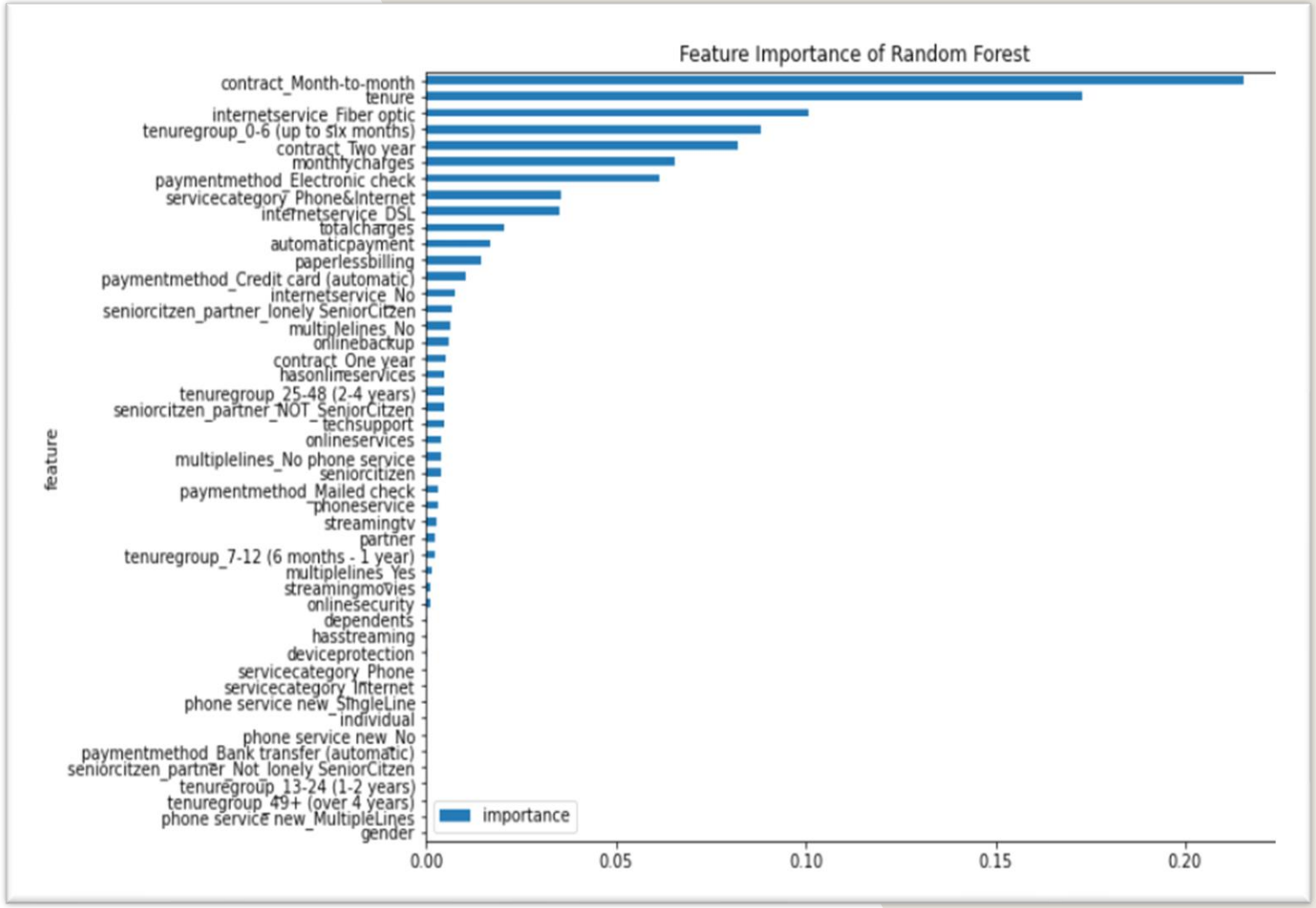
RANDOM FOREST

TREE NUMBER – 10, MAX DEPTH – 4
CALCULATED SAME WAY AS DECISION TREE, THROUGH THE LOOP. THEN TRIED TO FIND BEST OVERLAPPING POINTS OF BOTH SETS.

ACCURACY – 0.790



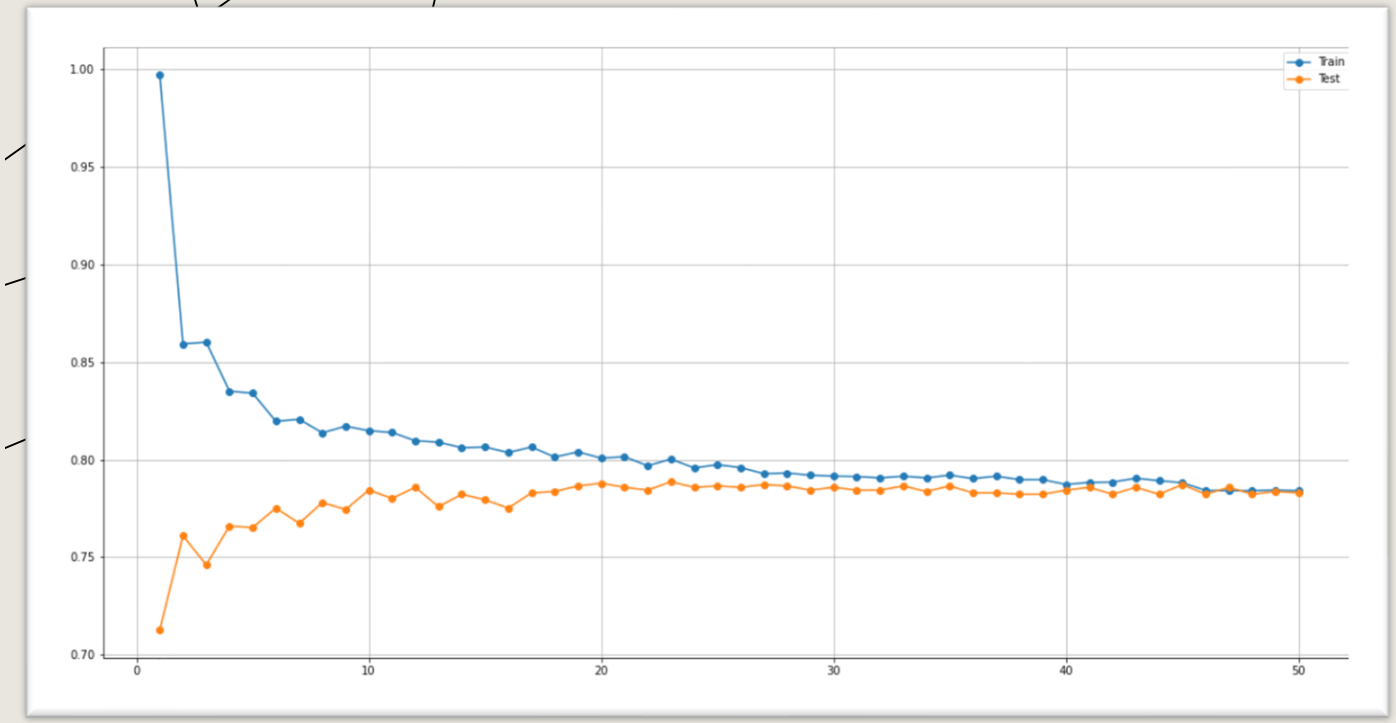
RANDOM FOREST – FEATURE IMPORTANCE



MOST IMPORTANT FEATURES ARE:

- TENURE
- CONTRACT
- FIBER OPTIC SERVICE
- MONTHLY CHARGES
- PAYMENT METHOD

KNN



NEIGHBORS NUMBER – 45 (CALCULATED)
ACCURACY – 0.787

Neighbors number was calculated by a loop that finds accuracy for every neighbor number from 1 to 50.

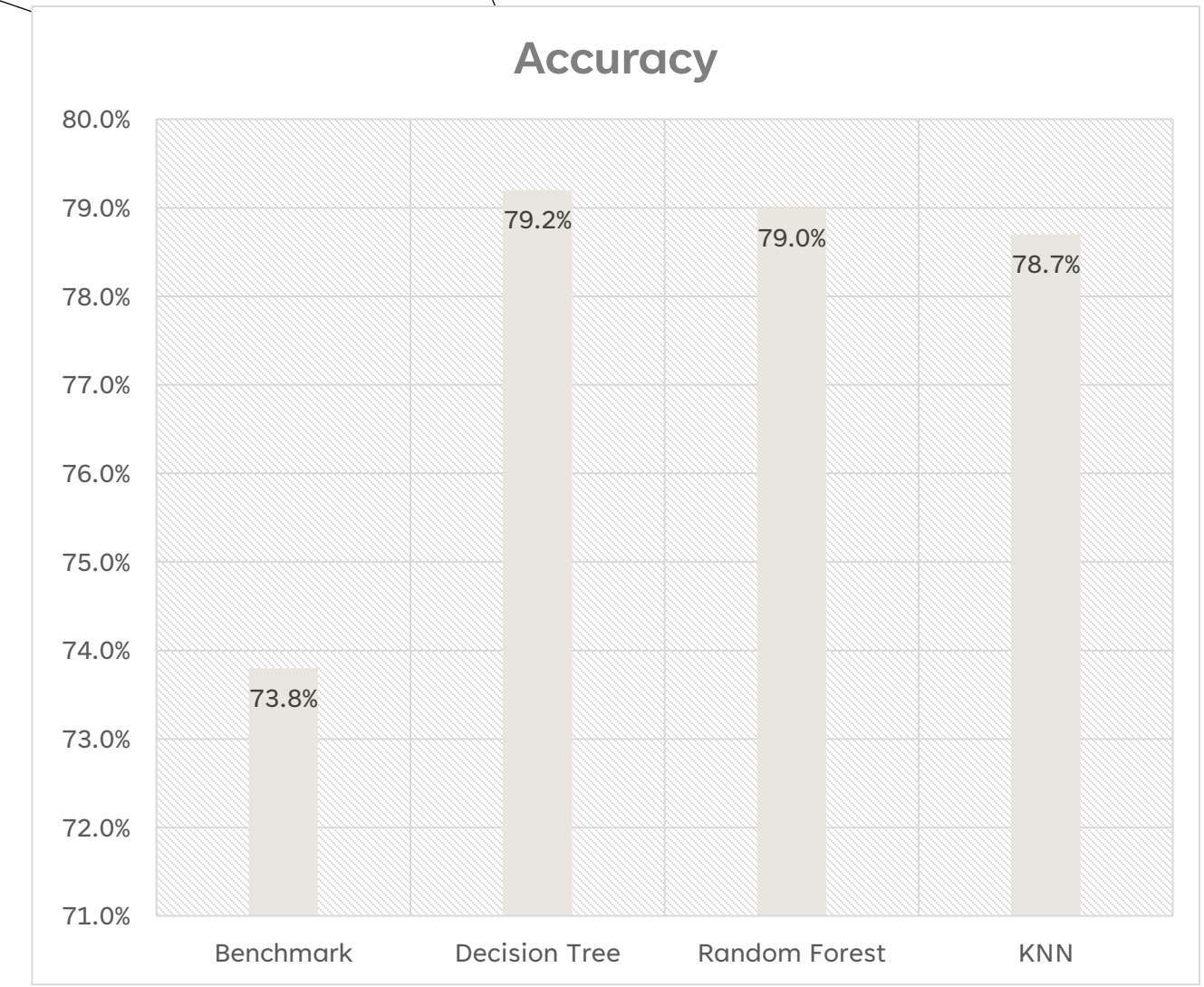
We wanted to be sure that KNN model is not overfitted, so we made a check up by Standard Scaler.
KNN Scaled Accuracy 0.767, so we concluded that original model was not overfitted.

BENCHMARK

For better understanding the quality of our models we want to set a simple benchmark and compare our results to it.

Let’s say non of the customers in train data set churned.
In this case, accuracy was 0.738

BENCHMARK



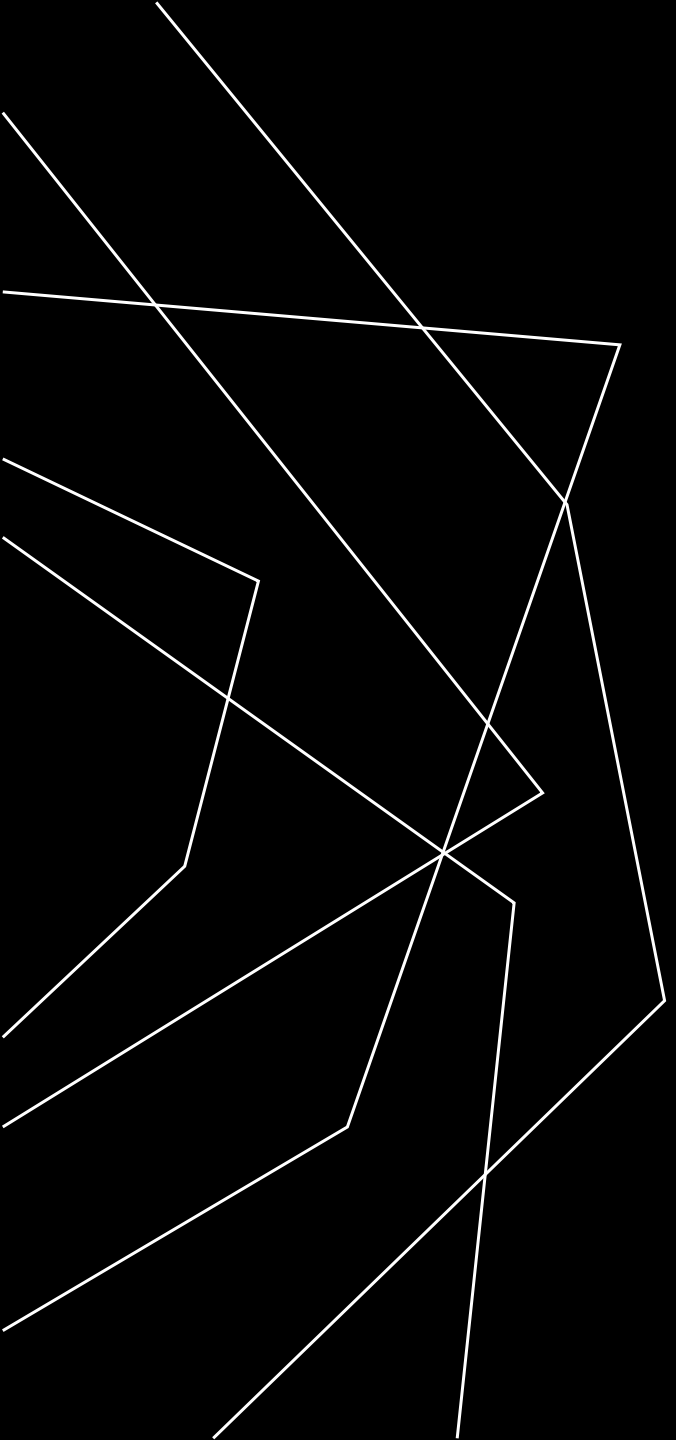
USING EVERY ONE OF ALGORITHMS IMPROVES ACCURACY OF PREDICTION SIGNIFICANTLY. BEST ACCURACY IS SHOWN BY DECISION TREE.

SUMMARY

IN THIS PROJECT WE TRIED TO CREATE A USEFULL MODEL THAT BUILDS CHURN PREDICTIONS.
WE USED 3 DIFFERENT ALGORITHMS AND COMPARED THE RESULTS TO A BENCHMARK.

WE SEE CLEARLY THAT USING THESE MODELS CAN IMPROVE THE QUALITY OF CHURN PREDICTION AND CREATE REAL VALUE FOR THE COMPANY.

IN OUR CASE, A MODEL WITH BEST PREDICTION RESULTS WAS DECISION TREE THAT PROVIDED ACURACY OF 79.2% COMPARED TO BENCHMARK ACURACY OF 73.8%



THANK YOU