# Telco Customer Churn

By Group 3



## OUR TEAM



Agung Prayogi
Universitas Pembangunan
Nasional Veteran Yogyakarta

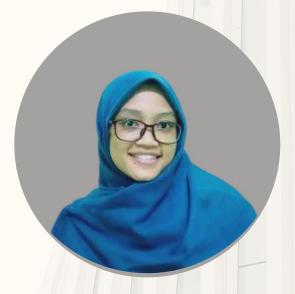


Universitas Pendidikan Indonesia

**Axel Eldrian H.** 



**Dian Maulida**Universitas Diponegoro

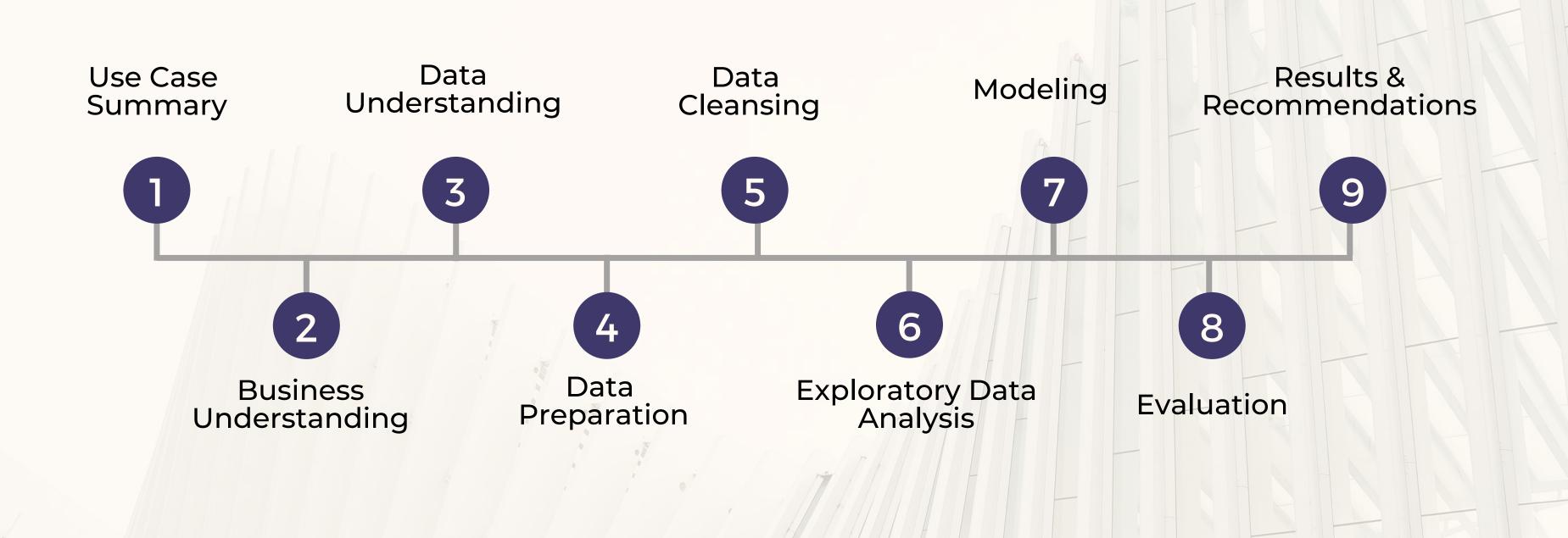


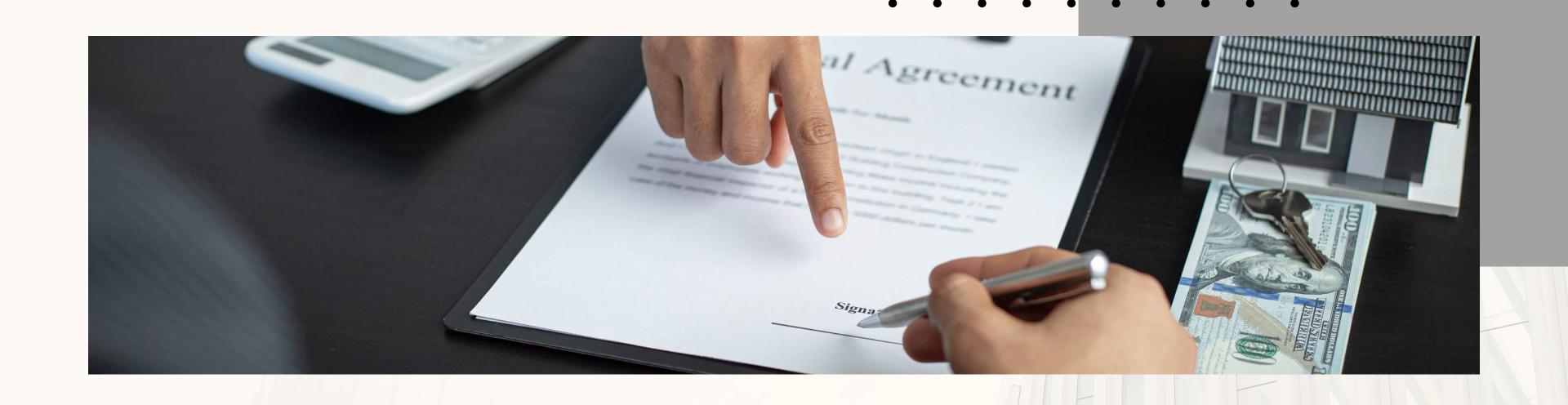
**Dwi Fitriawati F.**Universitas Pendidikan Indonesia



**Yesaya Arya D. K.**Universitas Merdeka Malang

## WORKFLOW





## USE CASE SUMMARY

#### **OBJECTIVE STATEMENTS**

- Get an insight into the influence of gender on customer churn.
- Describe customer satisfaction seen from the length of their contract that affects customer churn.
- Get to know how much influence the Senior Citizen has on customer churn.
- Get to know how much influence the Dependents have on customer churn.
- Get to know how much influence Paperless Billing has on customer churn.
- Get to know how important partners influence customer churn
- Get to know how important the type of internet service to influencing customer churn
- Get an insight into how much influence the Phone Services have on customer churn
- Get an insight into how much influence Online Security has on customer churn
- Get an insight into how much influence the Online Backup has on customer churn
- Get an insight into how much influence Device Protection has on customer churn
- Get an insight into how much influence Tech Support has on customer churn
- Get an insight into how much influence StreamingTV on customer churn
- Get an insight into how much influence Streaming Movie on customer churn
- Get an insight into how much influence the payment method has on Customer Churn
- Create machine learning models to predict customers who will churn

#### **CHALLENGES**

- The data have a lot missing value
- There is data that has an inappropriate data type
- The data scale is not all the same

#### METHODOLOGY / ANALYTIC TECHNIQUE

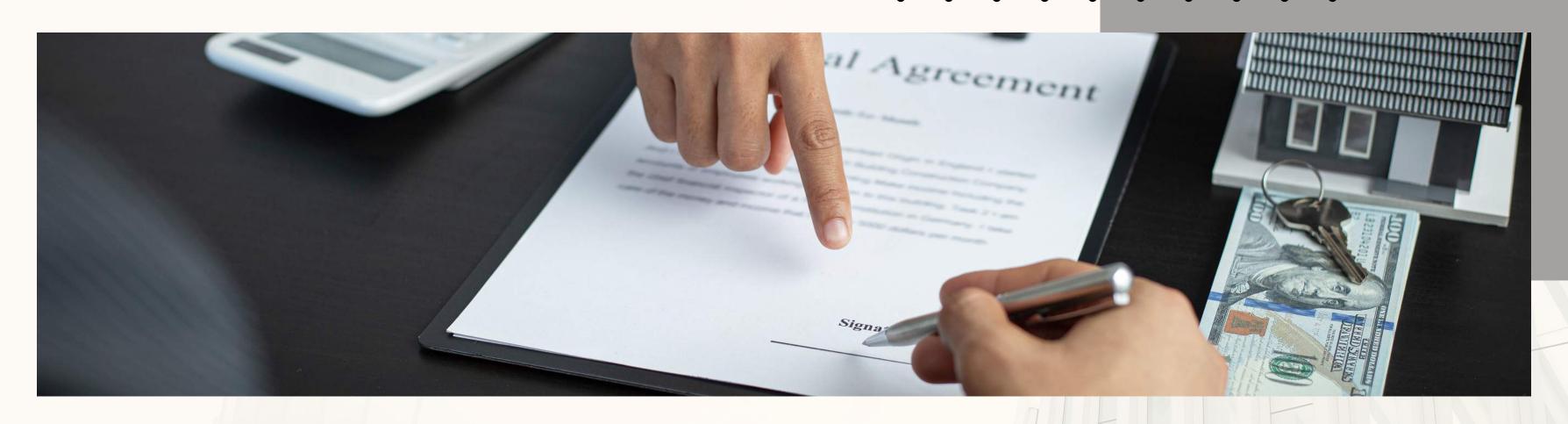
- Descriptive analysis
- Graph analysis
- Machine learning model (classification)

#### **BUSINESS BENEFIT**

- Help marketing strategies to improve products or services so that customer churn can be reduced.
- Help find out which aspects need to be developed to reduce customer churn.

#### **EXPECTED OUTCOME**

- Knowing the influence of gender on customer churn.
- Knowing customer satisfaction is seen from the length of their contract affects customer churn.
- Get to know how much influence the Senior Citizen on customer churn.
- Get to know how much influence the Dependents on customer churn.
- Get to know how much influence the Paperless Billing on customer churn.
- Knowing how important partners influence customer churn.
- Knowing how important the type of internet service to influence customer churn
- Get to know how much influence the Phone Services on customer churn.
- Get to know how much influence Online Security on customer churn.
- Get to know how much influence the Online Backup on customer churn.
- Get to know how much influence Device Protection on customer churn.
- Get to know how much influence Tech Support on customer churn.
- Get to know how much influence StreamingTV on customer churn.
- Get to know how much influence Streaming Movie on customer churn.
- Gain insight about how services affect the customer churn.
- Get to know how much influence the payment method on Customer Churn
- Making machine learning models to predict customers who will churn



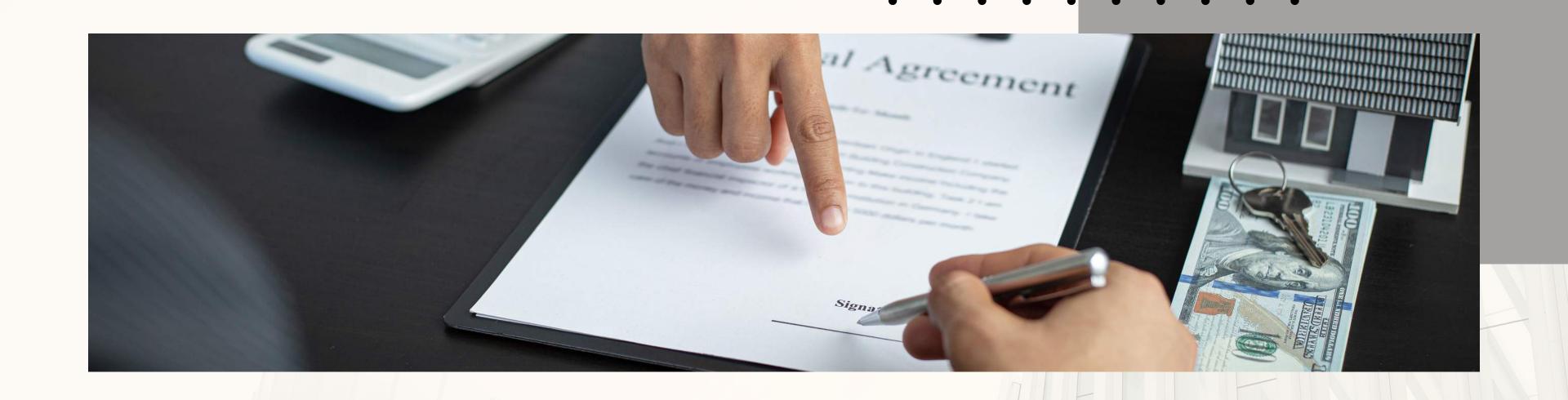
## BUSINESS UNDERSTANDING

Telco company is a kind of communications service provider, more precisely a telecommunications service provider, that provides telecommunications services such as telephony and data communications access.

# "Churn rate is a measure of the number of customers or employees who leave a company during a given period."

This case has some business questions using the data:

- How gender affects customer churn?
- How does customer satisfaction see from the length of their contract affects customer churn?
- How important services to affect customer churn?
- How important do partners influence customer churn?
- How important is the type of internet service to influence customer churn?
- How many percentage is predict to be churn?



## DATA UNDERSTANDING



#### **SOURCE DATA**

Data Telco

https://github.com/Yesaya03/Machin e-Learning-Regression-and-Timeseries/blob/main/data\_telco.csv

- 7043 rows
- 21 columns

#### **DATA DICTIONARY**

#### Services that each customer has signed

- OnlineSecurity: customer has online security or not (Yes, No, No internet service)
- OnlineBackup: customer has online backup or not (Yes, No, No internet service)
- PhoneService: customer has a phone service or not (Yes, No)
- MultipleLines: customer has multiple lines or not (Yes, No, No phone service)
- InternetService: customer's internet service provider (DSL, Fiber optic, No)
- DeviceProtection: customer has device protection or not (Yes, No, No internet service)
- TechSupport: customer has tech support or not (Yes, No, No internet service)
- StreamingTV: customer has streaming TV or not (Yes, No, No internet service)
- Streaming Movies: customer has streaming movies or not (Yes, No, No internet service)

#### Information Account for Customer

- tenure: number of months the customer has stayed with the company
- Contract: the contract term of the customer (Month-to-month, One year, Two year)

#### **DATA DICTIONARY**

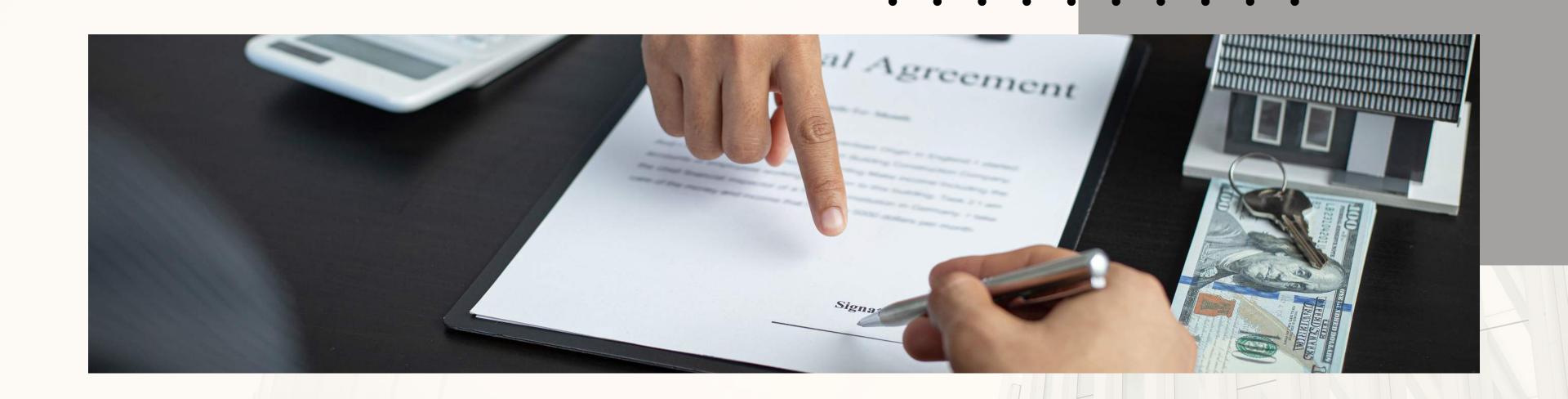
- PaymentMethod: the customer's payment method (Electronic check, Mailed check, Bank transfer (automatic), Credit card (automatic))
- PaperlessBilling: customer has paperless billing or not (Yes, No)
- MonthlyCharges: the amount charged to the customer monthly
- TotalCharges: the total amount charged to the customer

#### Demographic Info

- gender: male or a female (Male, Female)
- Partner: customer has a partner or not (Yes, No)
- Dependents: customer has dependents or not (Yes, No)
- customerID: customer ID
- SeniorCitizen: customer is a senior citizen or not (1, 0)

#### Label for Classification

• Churn: customer churned or not (Yes or No)



# DATA PREPARATION

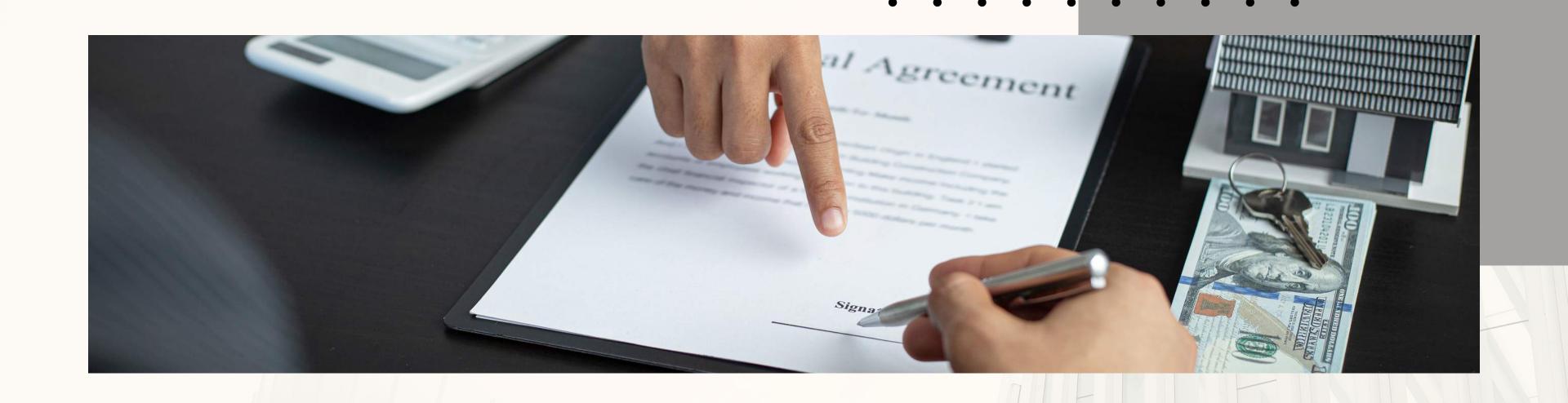
### **PYTHON VERSION**

3.9.12

#### **PACKAGES VERSION**

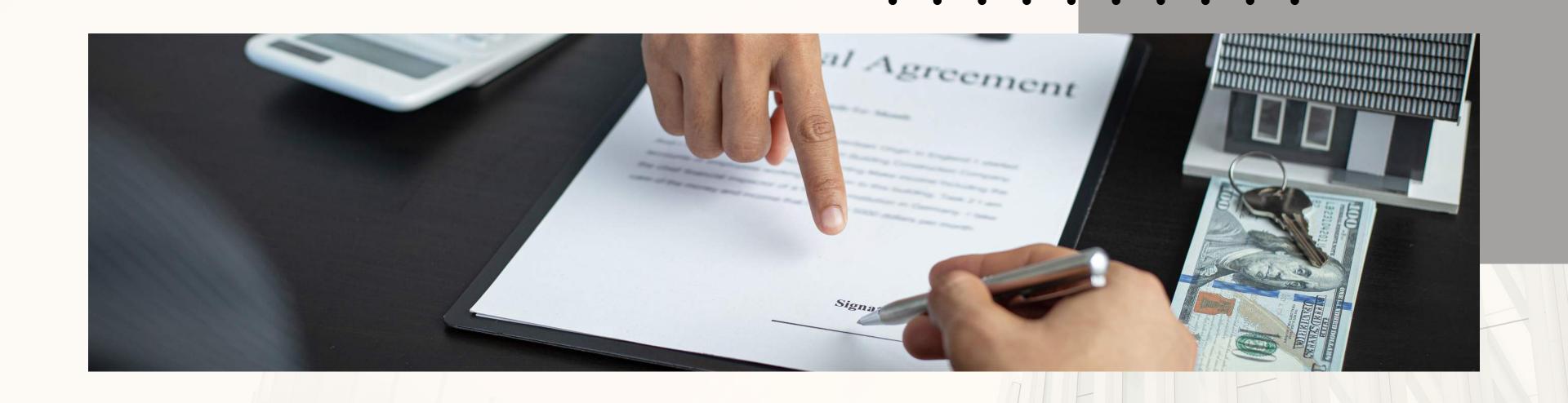
Pandas
Numpy
Matplotlib
Seaborn
Sklearn
Warnings





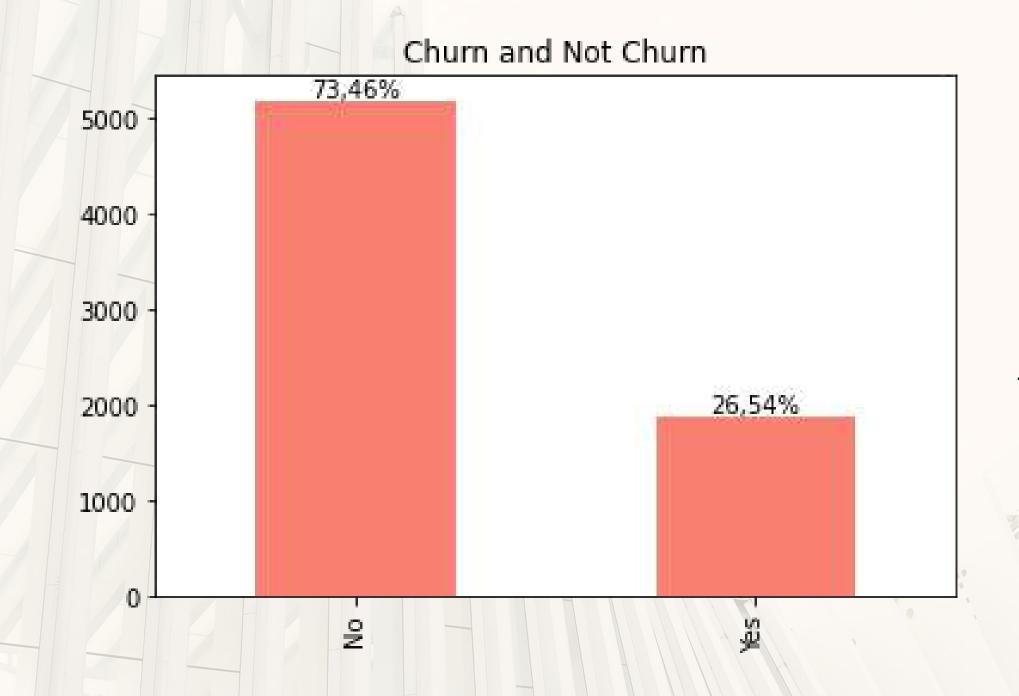
## DATA CLEANSING

There are 13 columns containing missing values, with 12 of them being categorical data so that it is imputed with "unknown" and the remaining 1 column is numeric data with positive-skew distribution so that it is imputed using the median. There is I column that has an inappropriate data type so it is necessary to adjust the data type. There are 16 categorical columns that need to be encoded, with 15 of them being nominal data so that they are encoded using One Hot Encoding and the remaining 1 column is ordinal data so that they are encoded using a map. The numeric columns in the dataset have different scales so it is necessary to scale the data using the StandardScaler. The identifier column (customerID) and the multicollinear column (TotalCharges) are removed.



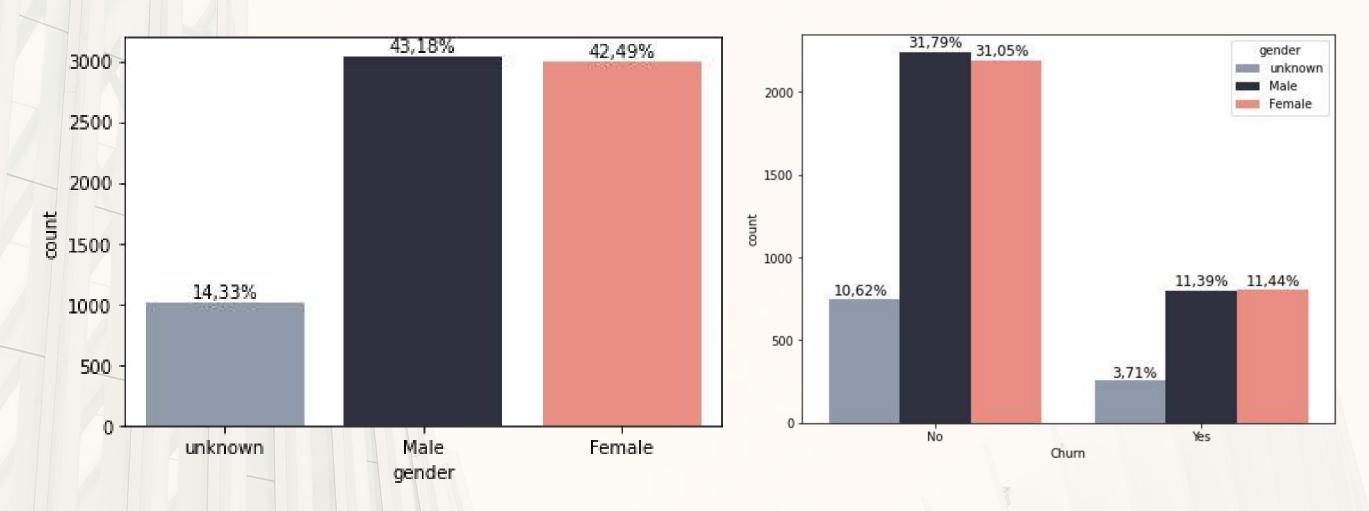
# EXPLORATORY DATA ANALYSIS

#### **CHURN RATE**



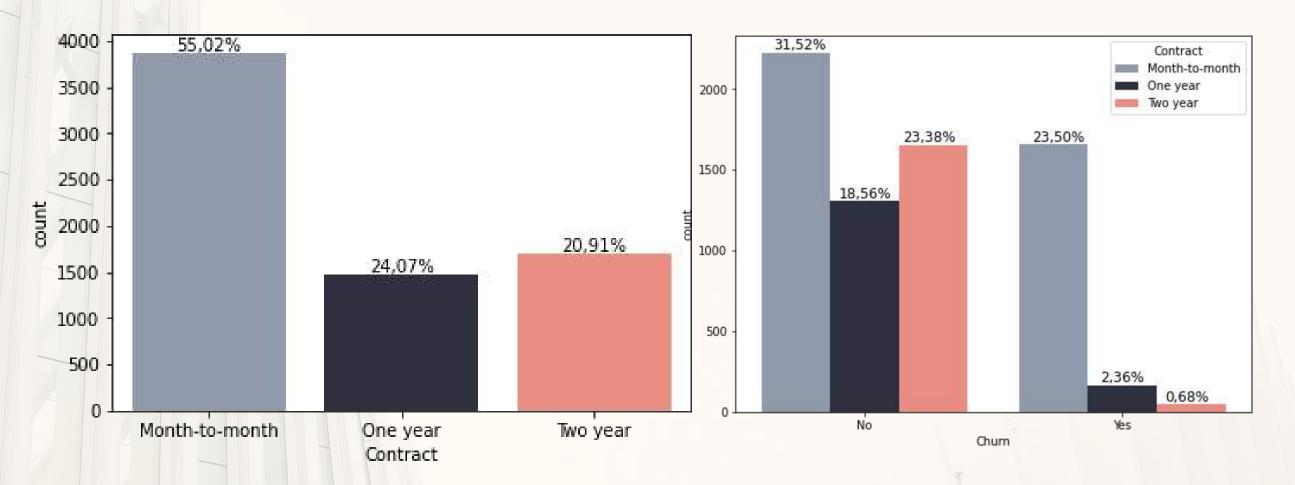
The graph above shows that most of the customers don't churn, the loyal customer is more than churn customer

#### **CUSTOMER CHURN BASED ON GENDER**



- Based on gender, telco customer distribution seems equal between males and female
- The churn rate is almost at an equal level, in other words, we can say that churn rates are not depending on the gender of the customer
- from 7043 customers there are 5147 customers who remain subscribed, or about 73.46% of telco data. and based on the results of EDA gender does not have much effect on customer churn, because the percentage results show the same thing, which is in the range of 31% of the total

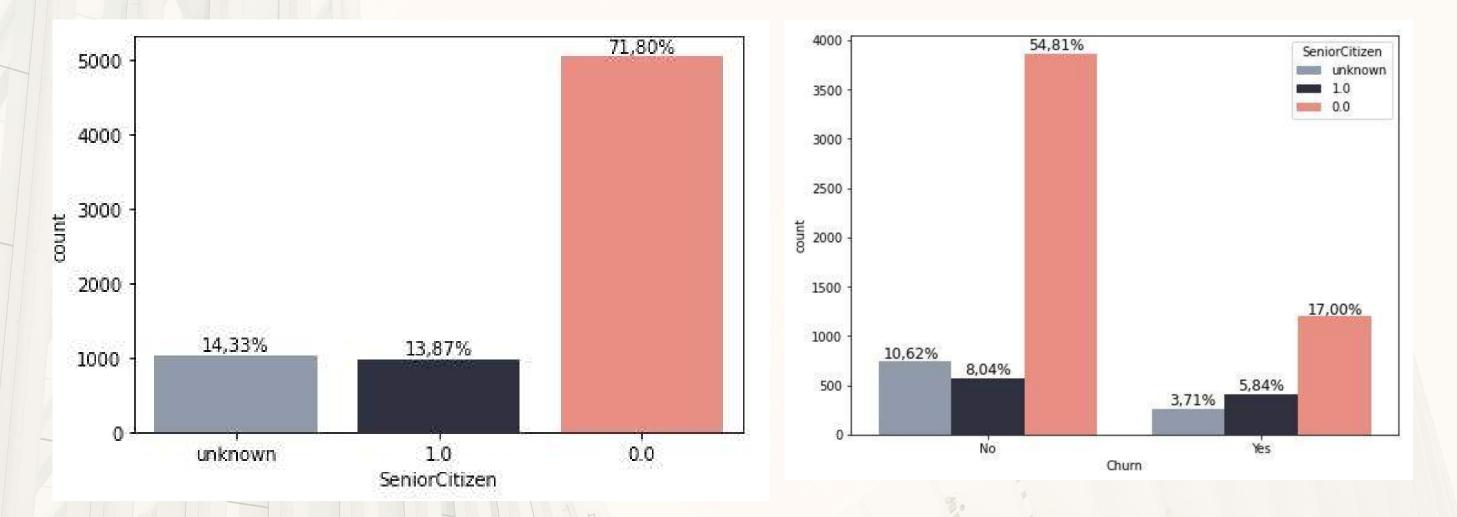
#### **CUSTOMER CHURN BASED ON CONTRACT**



- Only less than 4% of customers churn when they subscribe for 1 year and 2 years, and this explains that there is no fatal cause that makes customers churn.
- There is an increase in the percentage of about 4.82% between 1-year to 2-year subscriptions for customers who do not churn. that means some loyal customers are still satisfied with the services provided by the company.
- The high number of Month to Month customers who decide to churn is around 23.50%, which means that there are still many early customers who are not interested in the services provided by the company.

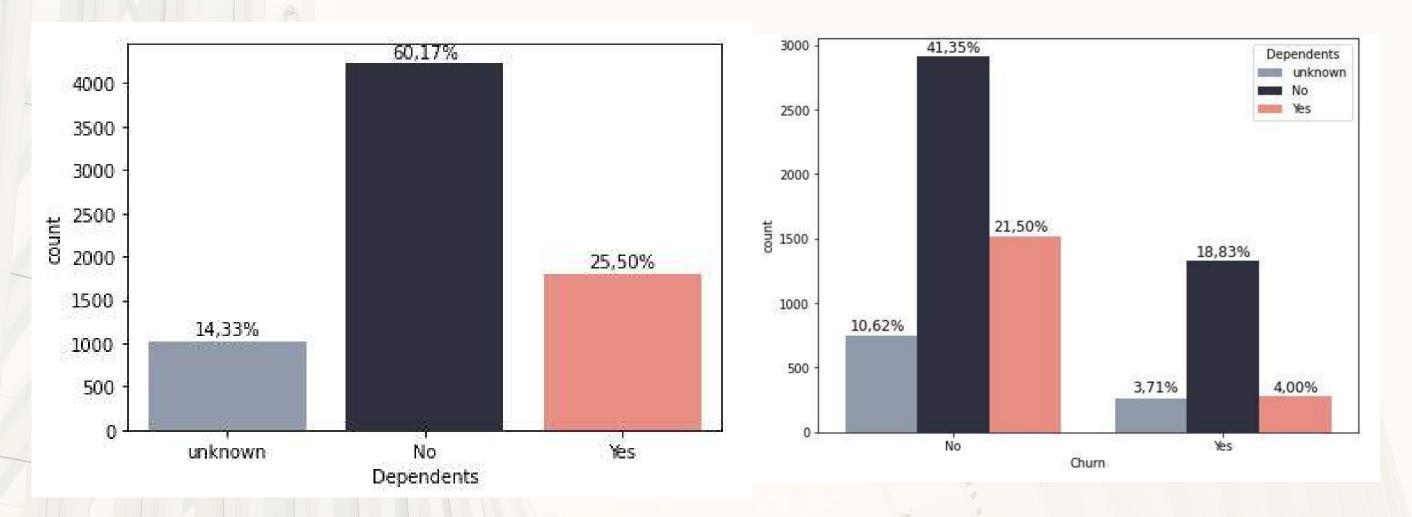






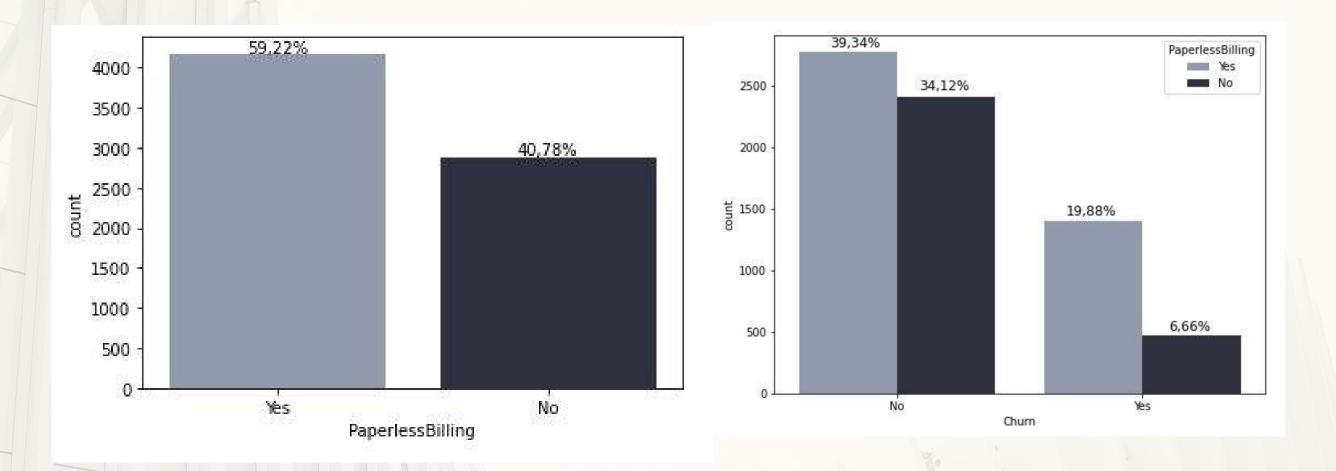
• Based on the senior citizen category, we can see that most telco customers are not senior citizen (younger citizen) in front of the percentage 71.80%, The level of customer churn of senior citizens is also low, the largest customer churn is from the non-senior citizen class with a percentage of 17%

## CUSTOMER CHURN BASED ON DEPENDENTS



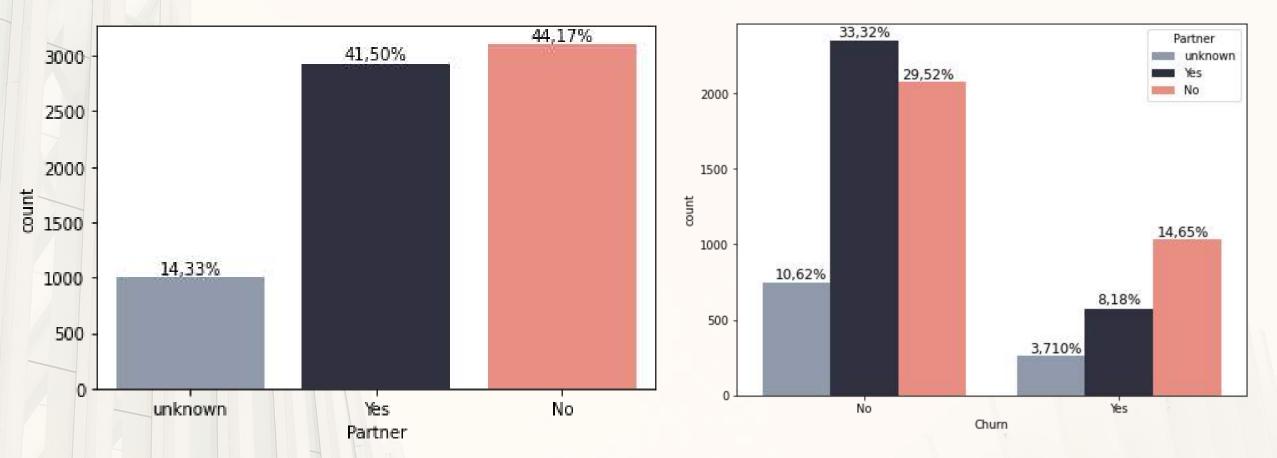
• Telco customers mostly do not have dependants with a percentage of 60.17% compared to customers who have a dependent that is equal to 25.50%. Customers who churn based on the most dependents are customers who do not have dependents with a percentage of 18.83%, even so, customers who still stick with telco services still quite a lot with a total of 73.47%

## CUSTOMER CHURN BASED ON PAPERLESS BILLING



- Telco customers based on paperless billing are the most customers who using paperless billing with a presentation of 59.22%, while customers who do not use paperless billing in amount 40.78%
- The customers who churn the most are customers who use paperless billing with a percentage of 19.88%

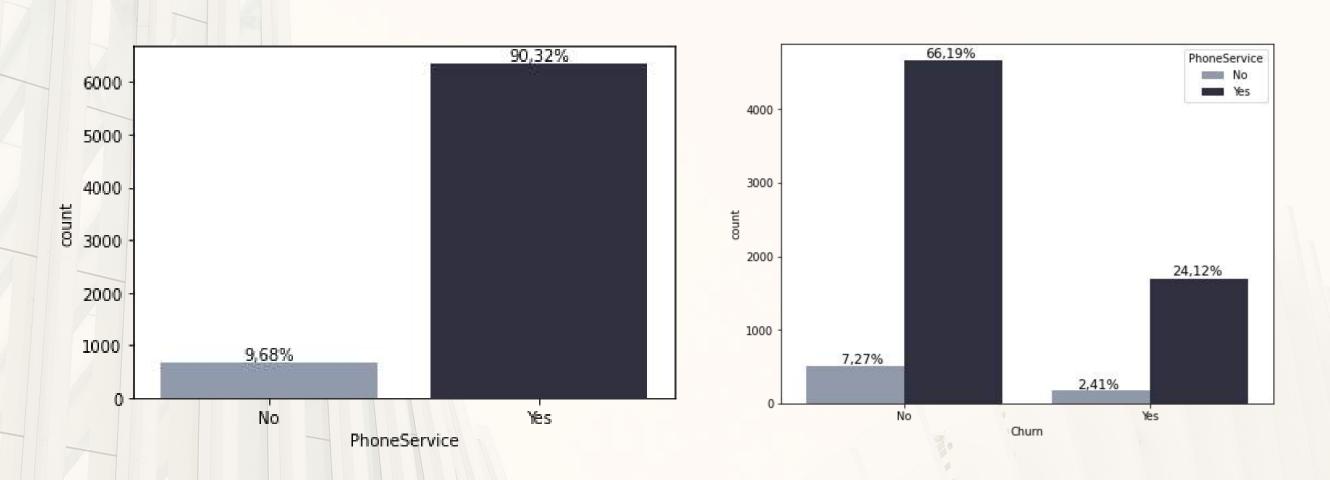
#### **CUSTOMER CHURN BASED ON PARTNER**



- Although the percentage of customers who do not churn and do not have partners is still high, namely 29.52%, partners are important things that can affect customer churn or not. It can be seen that there is a significant difference between churn customers who have partners and customers who do not churn without partners, namely 25.14%.
- From this we know, that most customers who enjoy the service are customers who do not have a partner. This is in line with the number of customers who churn, where more are customers who do not have partners. Companies can increase promotions on days associated with couples such as Valentine, to increase customers who want to use the company's services.

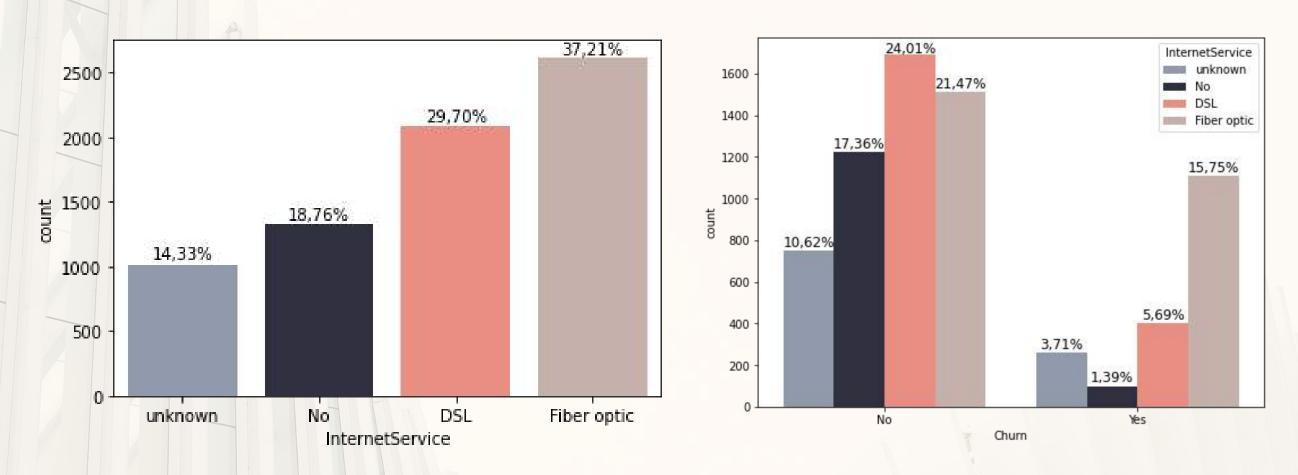


## CUSTOMER CHURN BASED ON PHONE SERVICE



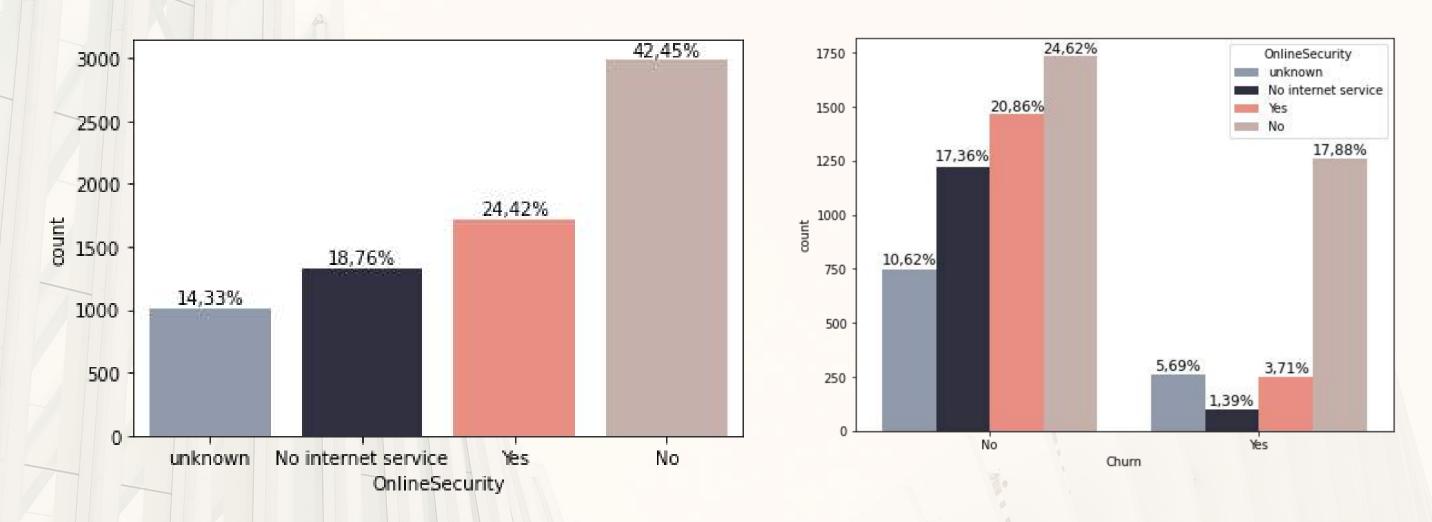
• The churn rate for phone service is low, while customers who have a lot of churn are customers who use phone service, this is in line with the number of phone service subscribers who are more (6361) than those who do not subscribe (682)

## CUSTOMER CHURN BASED ON INTERNET SERVICE



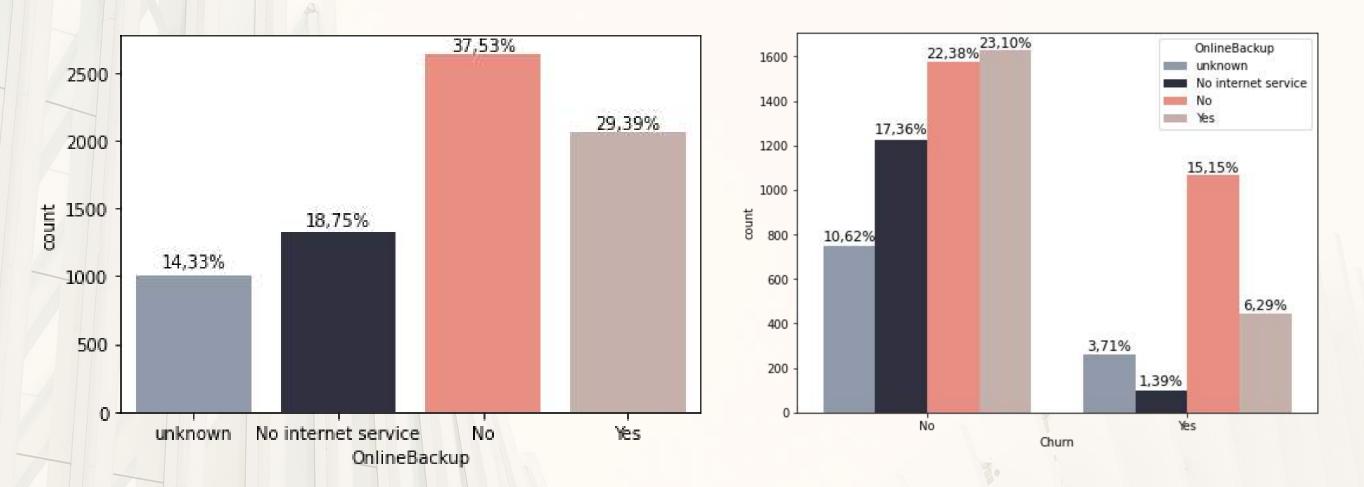
- Not only being the highest at 24.01%, but DSL also has a much lower percentage in influencing customer churn than Fiber Optic, which is a difference of 10%. This shows that DSL InternetService is still superior for customer use.
- From this we know that most customers who enjoy the service use Fiber Optic internet services, and the second is DSL. However, the number of customers who churn more is using Fiber Optic internet services. Companies can improve or maintain Fiber Optic-based services, so that customers can continue to survive or add new customers

## CUSTOMER CHURN BASED ON ONLINE SECURITY



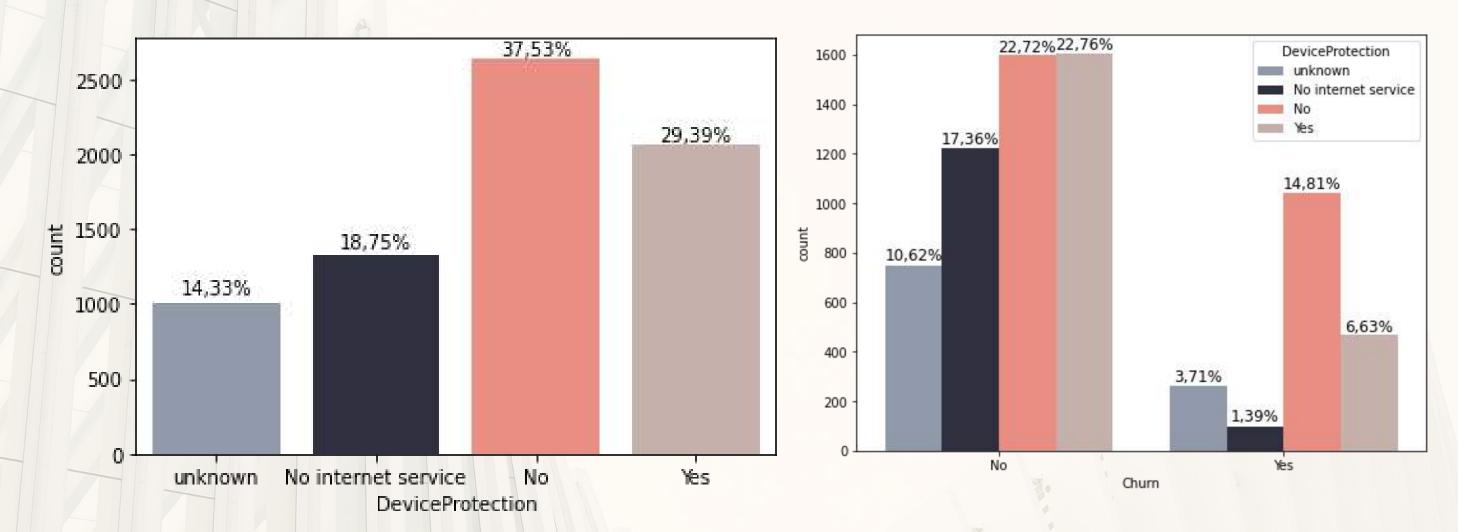
• From the data above, it is known that most customers do not subscribe to Online Security, followed by 1720 customers who subscribe to Online Security, 1321 customers who do not have internet service, and 1009 unknown customers. The churn rate in Online Security is dominated by customers who do not churn, while customers who churn are dominated by customers who do not use Online Security services.

## CUSTOMER CHURN BASED ON ONLINE BACKUP



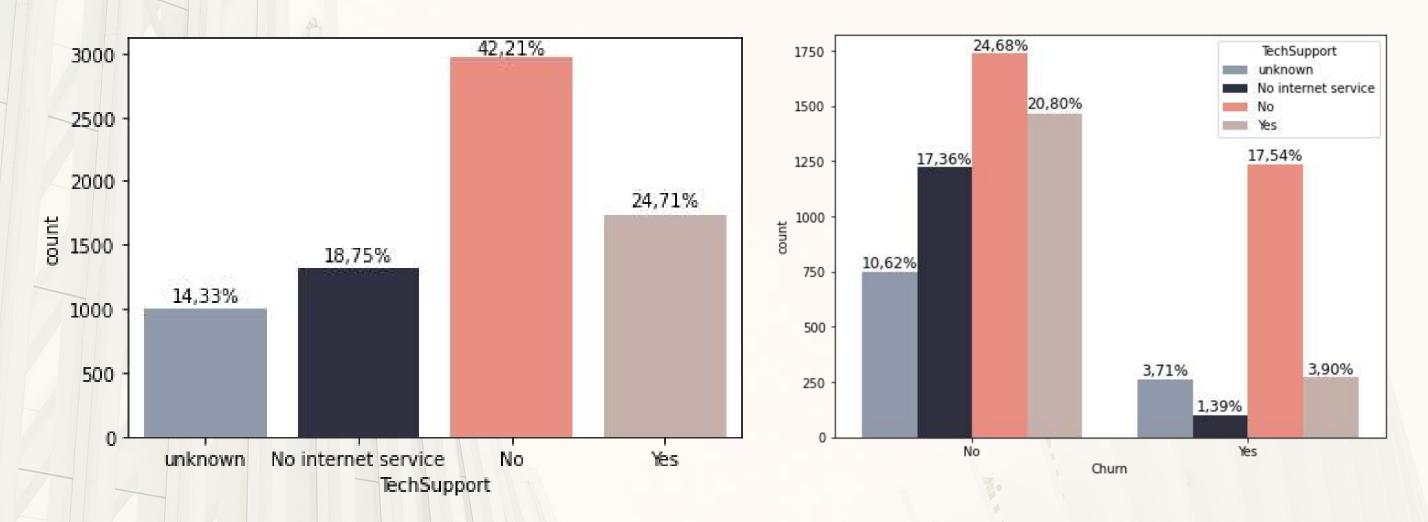
• From the data above, it is known that most customers do not subscribe to Online Backup, followed by 2070 customers who do not subscribe to Online Backup, 1321 customers who do not have internet service, and 1009 unknown customers. The churn rate in Online Backup is dominated by customers who do not have internet service. do not churn, while customers who churn are dominated by customers who do not use Online Backup services.

## CUSTOMER CHURN BASED ON DEVICE PROTECTION



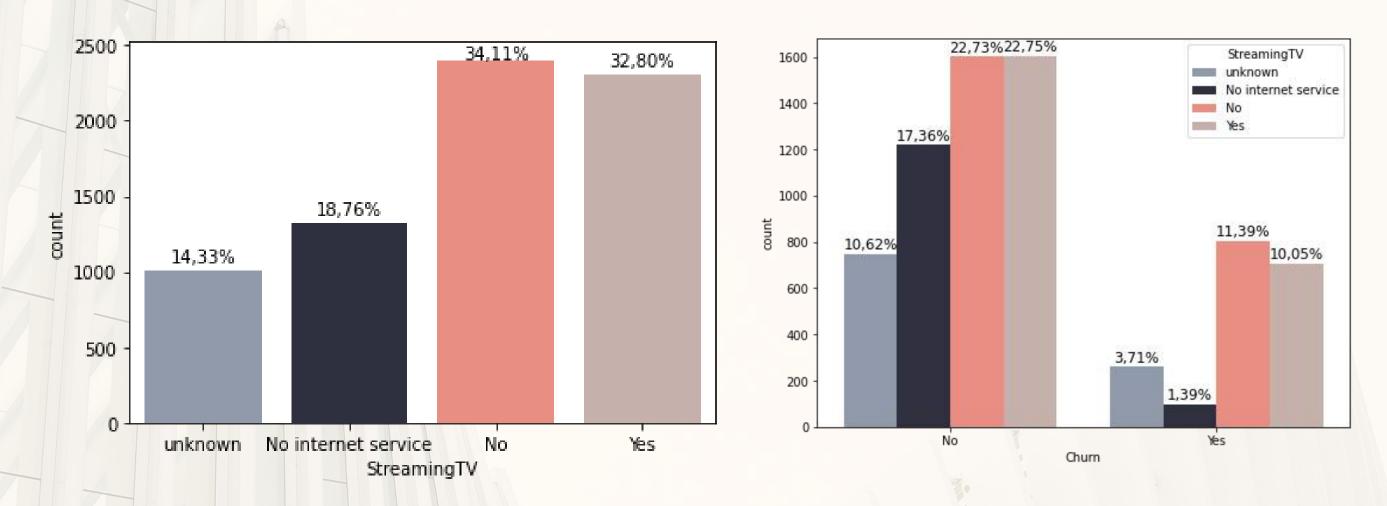
• From the data above, it is known that most customers do not subscribe to Device Protection, followed by 2070 customers who do not subscribe to Device Protection, 1321 customers who do not have internet service, and 1009 unknown customers. The churn rate in Device Protection is dominated by customers who do not have internet service. do not churn, while customers who churn are dominated by customers who do not use Device Protection services.

## CUSTOMER CHURN BASED ON TECH SUPPORT



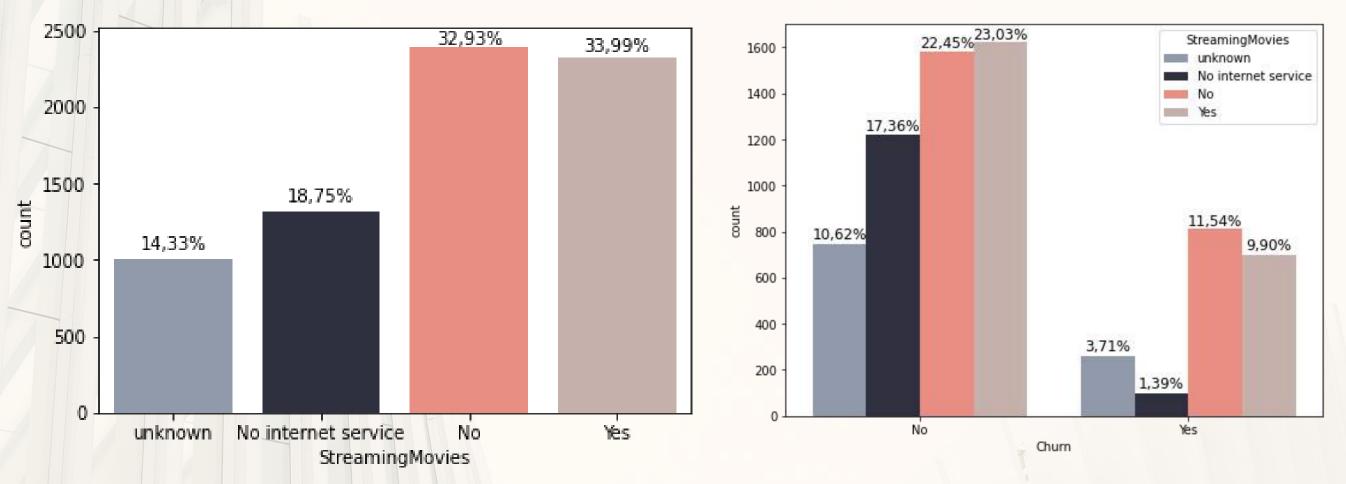
• From the data above, it is known that most customers do not subscribe to Tech Support, followed by 1740 customers who subscribe to Tech Support, 1321 customers who do not have internet service, and 1009 unknown customers. The churn rate in Tech Support is dominated by customers who do not churn, while customers who churn are dominated by customers who do not use Tech Support services.

## CUSTOMER CHURN BASED ON STREAMING TV



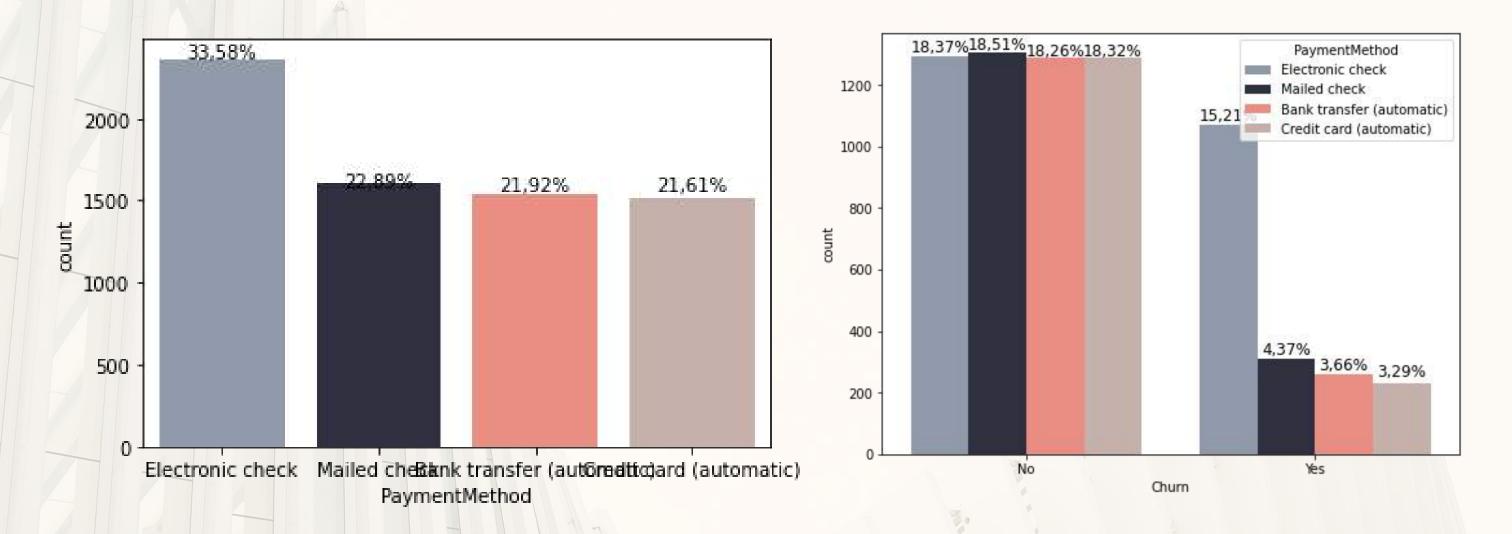
• From the data above, it is known that most customers do not subscribe to Streaming TV, followed by customers who subscribe to 2310 customers, 1321 customers who do not have internet service, and 1009 unknown customers. The churn rate on Streaming TV is dominated by customers who do not churn, while subscribers who churn are not much different between those who subscribe and do not subscribe to Streaming TV

## CUSTOMER CHURN BASED ON STREAMING MOVIES



- From the data above, it is known that most customers do not subscribe to Streaming Movies, followed by 2319 subscribers, 1321 customers who do not have internet service, and 1009 unknown customers. The churn rate for Streaming Movies is dominated by customers who do not churn, while subscribers who churn are not much different from those who subscribe and do not subscribe to Streaming Movies
- From all the data services above, it shows that the services provided are still not superior that can be utilized by the company because there are still many customers who continue to subscribe even though they do not use the services provided by the company.

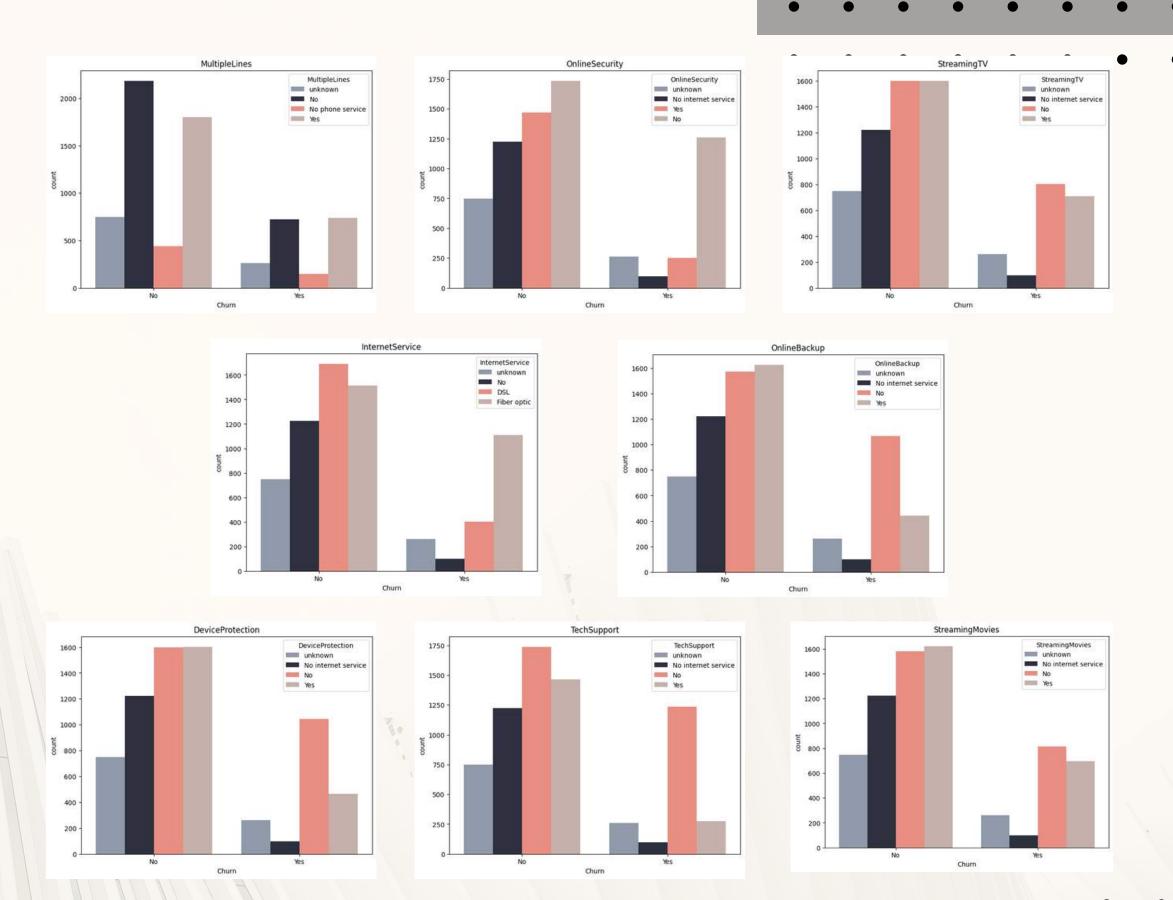
## CUSTOMER CHURN BASED ON PAYMENT METHOD

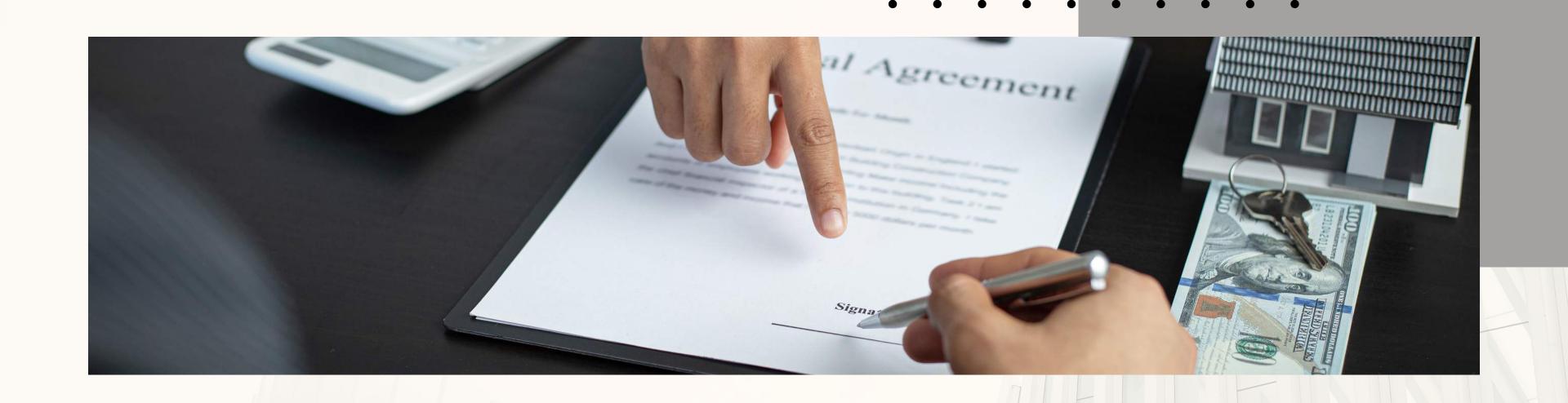


• From the data above, it is known that most customers use Electronic Check for the Payment Method, the churn rate in the Payment Method is dominated by customers who do not churn, while customers who churn are dominated by customers who use Electronic check for the payment method.

## HOW SERVICES AFFECT THE CUSTOMER CHURN

- From the overall data, there are still many customers who continue in telco even though they do not use the services provided. can be seen if the average is almost more than 22% of the total data, customers do not use the services provided by the company.
- From the overall data, there are still many customers who still in telco but they do not have internet services in it, about 17% of the total data.

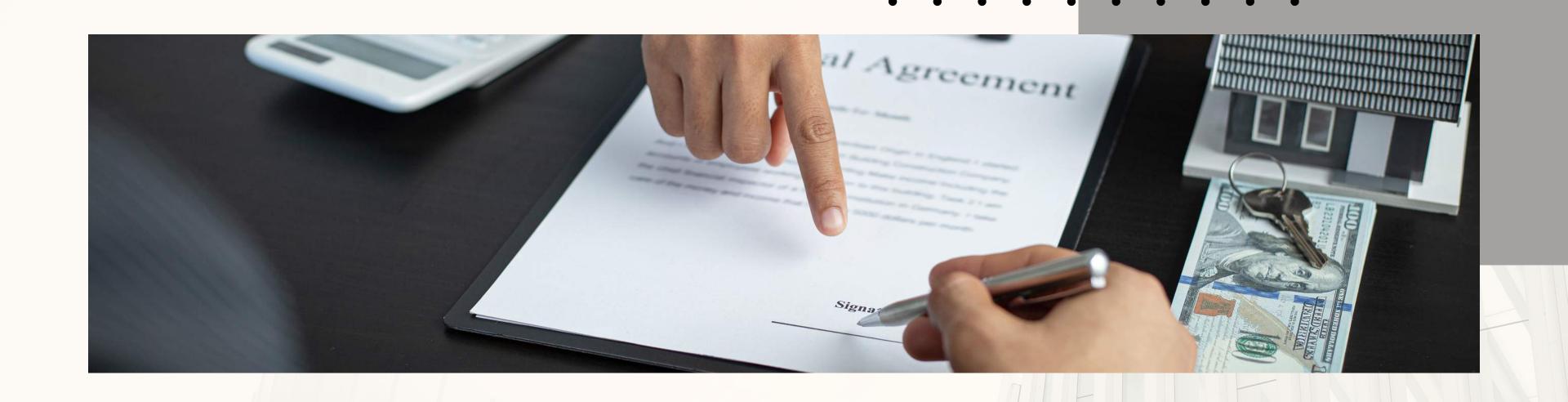




## MODELING

## MODELING: LOGISTIC REGRESSION

- Logistic Regression algorithm is an supervised machine learning algorithm for predicting the categorical dependant variable using a given set of independent variables
- Logistic gives best result under the following conditions:
  - The dependant variable in binary logistic regression must be binary
  - only the variables that are relevant should be included
  - the independent variable must be unrelated to one another (Have minimal or no multicollinearity in the model)
  - The log chance are proportional to the independent variables
  - Required large sampe size

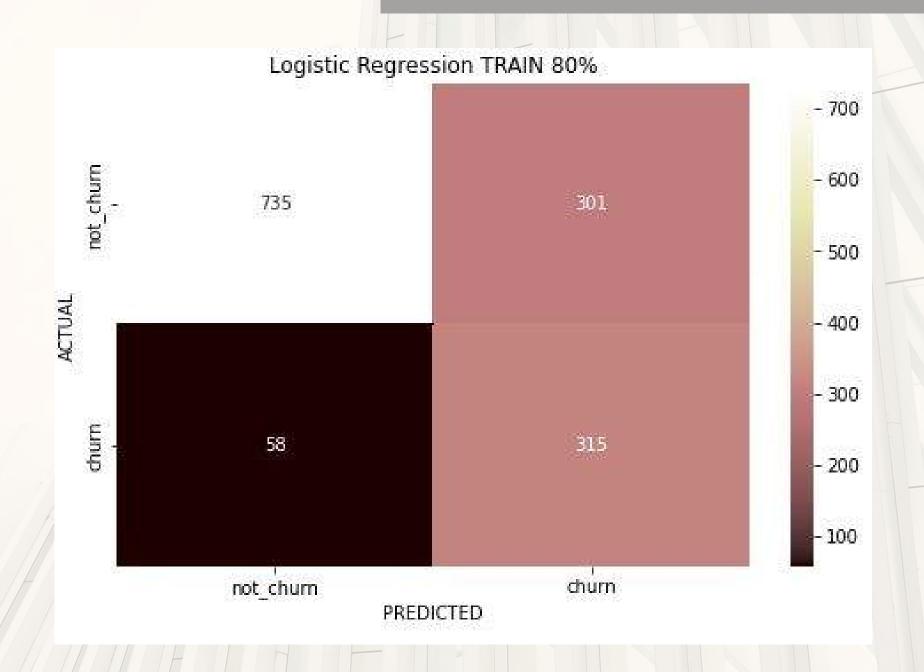


# EVALUATION

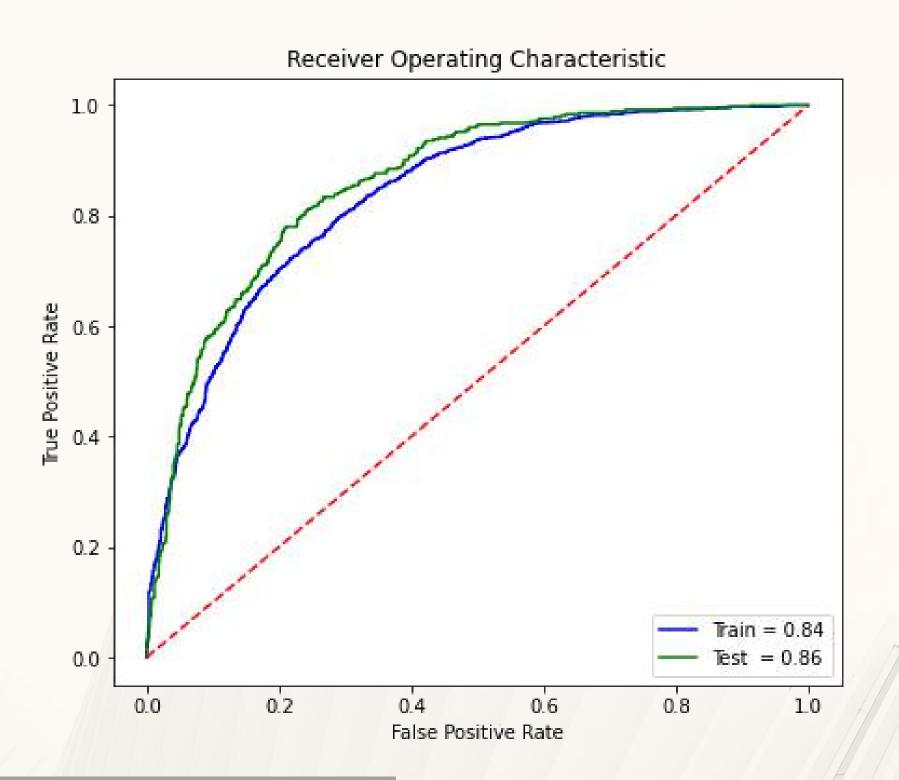
In the logistic regression model, for the evaluation model, we use the Confusion Matrix. The Confusion Matrix contains precision, recall, fl-score, accuracy, and support.

• From the modeling, we get 75% churn accuracy, 51% Precision churn, 84% Recall churn, and 64% F1 Score churn. In this case modeling, it is better to see someone who does not churn into churn so it is better to use recall as an indication of model performance

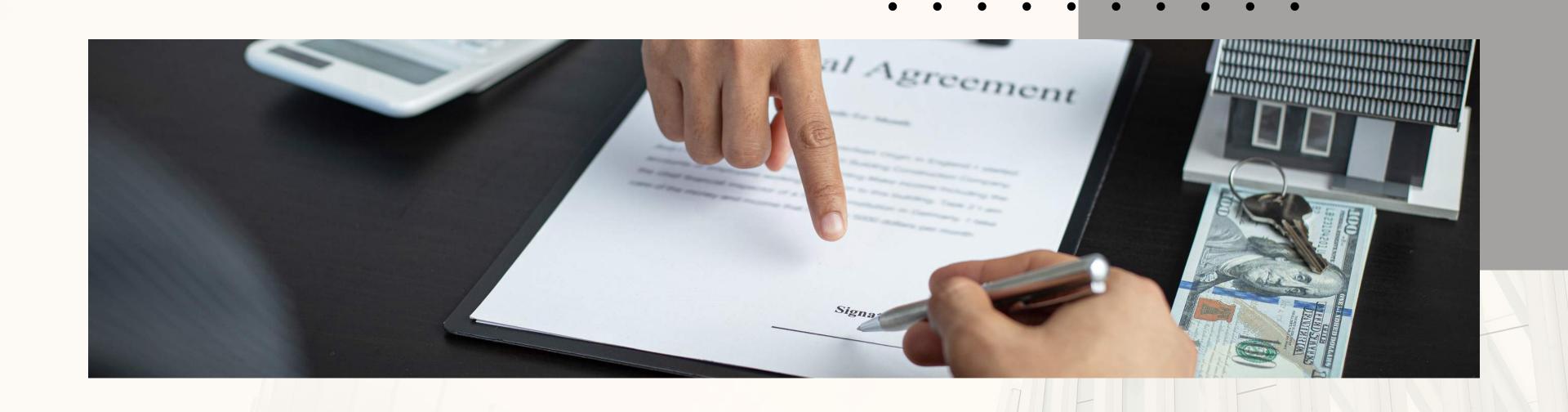
precision	recall	f1-score	support
0.93	0.71	0.80	1036
0.51	0.84	0.64	373
		0.75	1409
0.72	0.78	0.72	1409
0.82	0.75	0.76	1409
	0.93 0.51 0.72	<ul> <li>0.93</li> <li>0.71</li> <li>0.51</li> <li>0.84</li> <li>0.72</li> <li>0.78</li> </ul>	0.93 0.71 0.80 0.51 0.84 0.64 0.75 0.72 0.78 0.72



## **AUC-ROC**



From the AUC-ROC graph and confusion matrix evaluation, we can know that the AUC-ROC value is 86% and also the model is good fit.



# RESULT

#### • [Gender on churn]

- 1. Based on gender, telco customer distribution seems equal between males and female
- 2.The churn rate is almost at an equal level, in other words, we can say that churn rates are not depending on the gender of the customer
- 3.from 7043 customers there are 5147 customers who remain subscribed, or about 73.46% of telco data. and based on the results of EDA gender does not have much effect on customer churn, because the percentage results show the same thing, which is in the range of 31% of the total

#### • [Contract on churn]

- 1.Only less than 4% of customers churn when they subscribe for 1 year and 2 years, and this explains that there is no fatal cause that makes customers churn.
- 2. There is an increase in the percentage of about 4.82% between 1-year to 2-year subscriptions for customers who do not churn. that means some loyal customers are still satisfied with the services provided by the company.
- 3.The high number of Month to Month customers who decide to churn is around 23.50%, which means that there are still many early customers who are not interested in the services provided by the company.

#### • [Services on churn]

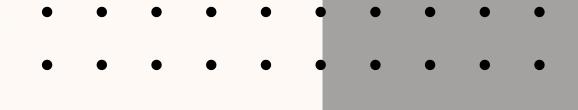
- 1. From the overall data, there are still many customers who continue in telco even though they do not use the services provided. can be seen if the average is almost more than 22% of the total data, customers do not use the services provided by the company.
- 2. From the overall data, there are still many customers who still in telco but they do not have internet services in it, about 17% of the total data.

#### • [Partner on churn]

- Although the percentage of customers who do not churn and do not have partners is still high, namely 29.52%,
  partners are important things that can affect customer churn or not. It can be seen that there is a significant
  difference between churn customers who have partners and customers who do not churn without partners,
  namely 25.14%.
- From this we know, that most customers who enjoy the service are customers who do not have a partner. This is in line with the number of customers who churn, where more are customers who do not have partners. Companies can increase promotions on days associated with couples such as Valentine, to increase customers who want to use the company's services.

#### • [Internet Service on churn]

- 1. Not only being the highest at 24.01%, but DSL also has a much lower percentage in influencing customer churn than Fiber Optic, which is a difference of 10%. This shows that DSL InternetService is still superior for customer use.
- 2. From this we know that most customers who enjoy the service use Fiber Optic internet services, and the second is DSL. However, the number of customers who churn more is using Fiber Optic internet services. Companies can improve or maintain Fiber Optic-based services, so that customers can continue to survive or add new customers



#### • [Senior Citizen on churn]

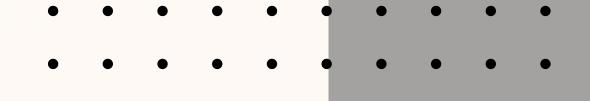
Based on the senior citizen category, we can see that most telco customers are not senior citizen (younger citizen) in front of the percentage 71.80%, The level of customer churn of senior citizens is also low, the largest customer churn is from the non-senior citizen class with a percentage of 17%

#### • [Dependents on churn]

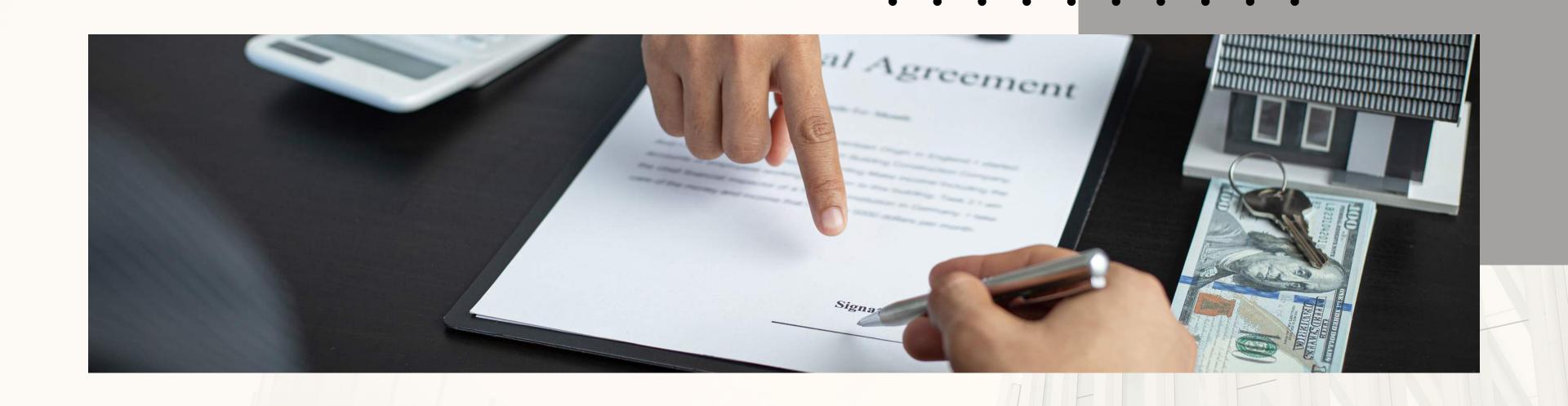
Telco customers mostly do not have dependents with a percentage of 60.17% compared to customers who have a dependent that is equal to 25.50%. Customers who churn based on the most dependents are customers who do not have dependents with a percentage of 18.83%, even so, customers who still stick with telco services still quite a lot with a total of 73.47%

#### • [Paperless Billing on churn]

- 1.Telco customers based on paperless billing are the most customers who using paperless billing with a presentation of 59.22%, while customers who do not use paperless billing in amount 40.78%.
- 2. The customers who churn the most are customers who use paperless billing with a percentage of 19.88%.



- From the modeling, we get 75% churn accuracy, 51% Precision churn, 84% Recall churn, and 64% F1 Score churn. In this case modeling, it is better to see someone who does not churn into churn so it is better to use recall as an indication of model performance.
- To see further performance of the model, a probability check was carried out with ROC-AUC, it was found that the difference in accuracy between the train data and the test data was less than 0.05, so it can be concluded that the model made is good.
- When parameter adjustment is made with hyperparameter tuning, there is not much change in accuracy.



## RECOMMENDATION



### RECOMMENDATION

#### Recommendation for "Churn" Customer:

• Focus on promotion, campaign, or give advantage for return old customers such discount reactivation or free services. Make a point reactivation strategy, give the big advantage to old customers who already churns

#### Recommendation for "Not Churn" Customer

- Focus on a strategy to raise income such as down-selling such as bundles for a few products
- Indeed, services have not been very influential in terms of customer churn, but internet services need to be improved because there are still many customers who do not have internet services, which could be one of the big influences on why customers have not used the service.

. . . . . .

