

**MYEDUSOLVE**



**DATA SCIENCE B**

# Classification Churn Customer

---

**Mini Project**  
**By Group 1**

# Our Team

Agung Wahyu Prayogo



Information System  
UPN "Veteran" East Java

Informatics  
Bhayangkara Jakarta  
Raya University

Valent Aderiandra



Fauziya Alya Ramadhana



Mathematics  
Diponegoro University

Statistics  
Sepuluh Nopember  
Institute of Technology

Putri Windasari

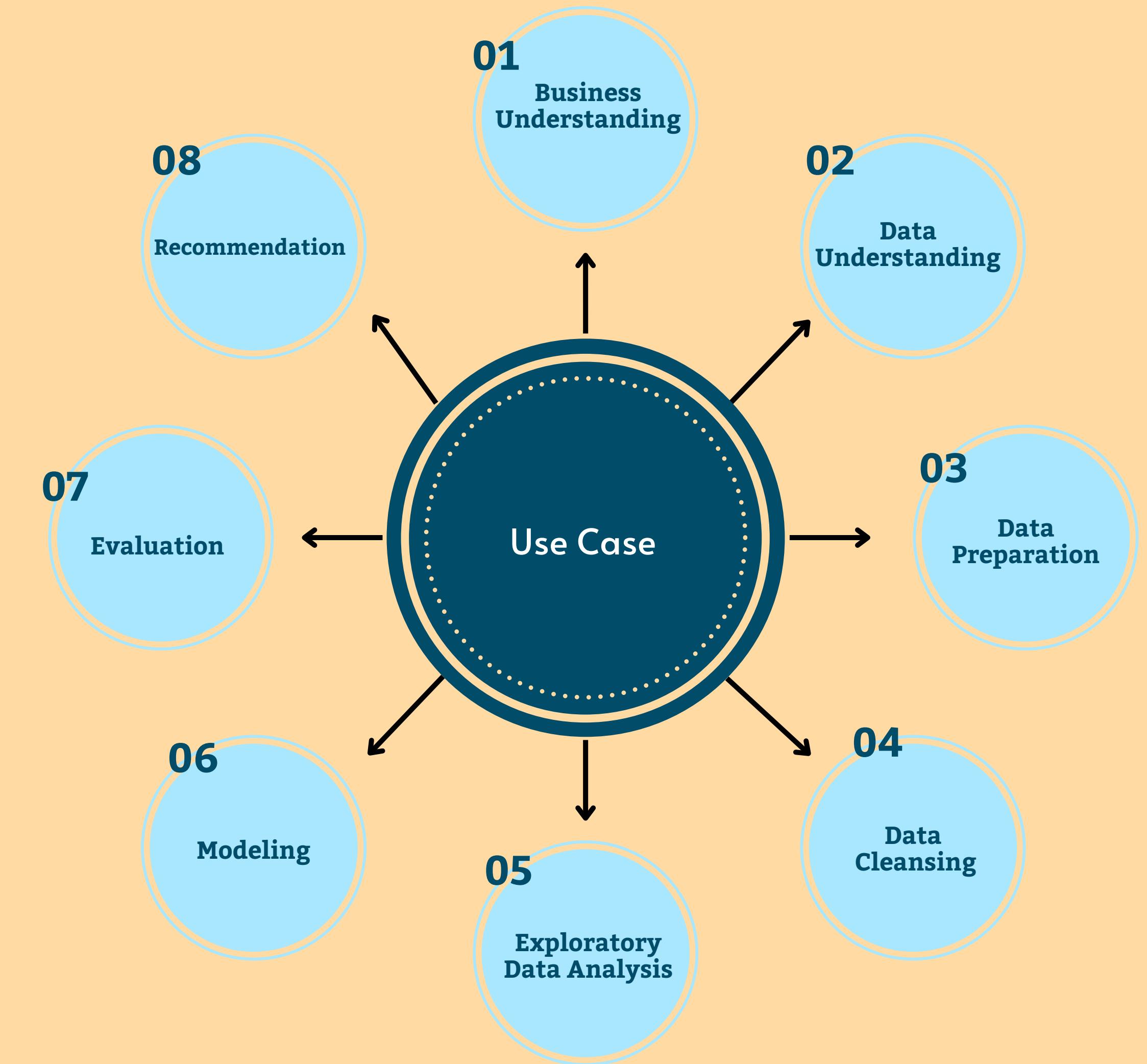
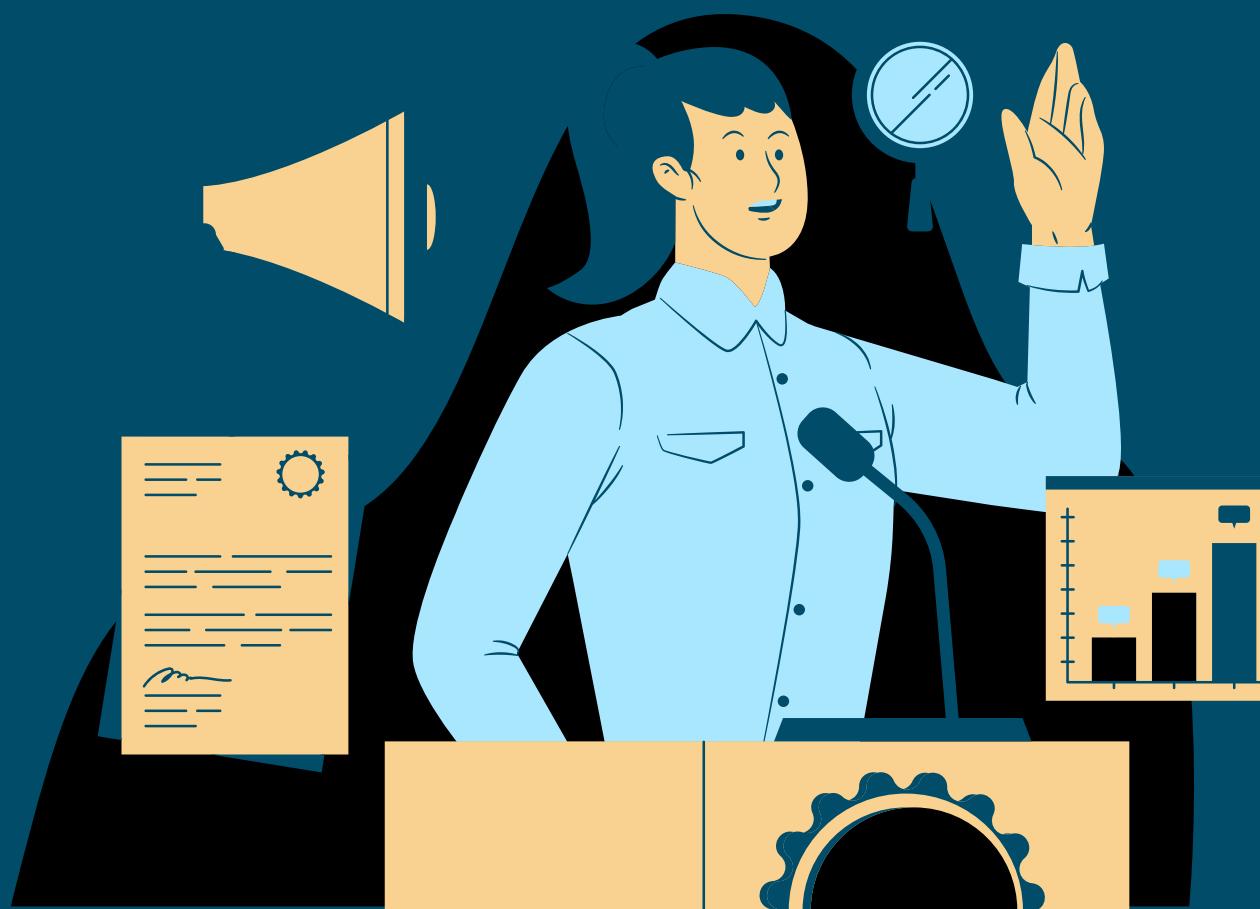


Reno Darin Khalifah P



Electrical  
Engineering  
Sepuluh Nopember  
Institute of  
Technology

# Work Flow Customer Churn Segmentation





**CUSTOMER CHURN  
SEGMENTATION**

# Use Case

# Use Case : Classification Churn Costumer

## Use Case Summary



### OBJECTIVE STATEMENT

- Get business insight about how many customer who churn or not churn
- Get business insight about how many female and male in churn or not churn customer
- Get business insight about how many customer who has a partner, has dependent or not, and old age or not
- Get business insight about how many customer used service facilities like : PhoneService, MultipleLines, InternetService, OnlineSecurity, OnlineBackup, DeviceProtection, TechSupport, StreamingTV, StreamingMovies, and Papperless Billing
- Get business insight about how long customer used service (Contract)
- Get business insight about how payment method from customer
- Get insight about 5 feature or facility that influenced customer to churn most. That is tenure, monthly charge, contract, gender male, fiber optic.
- Build models using machine learning to predict customer churn



# Use Case Summary



## >> CHALLANGE

- Large size of data, can not maintain by excel spreadsheet
- Fill in the missing value in the dataset
- Change the data type of column which doesn't match
- Hard to know the meaning of each column

## Challange | Metodeology

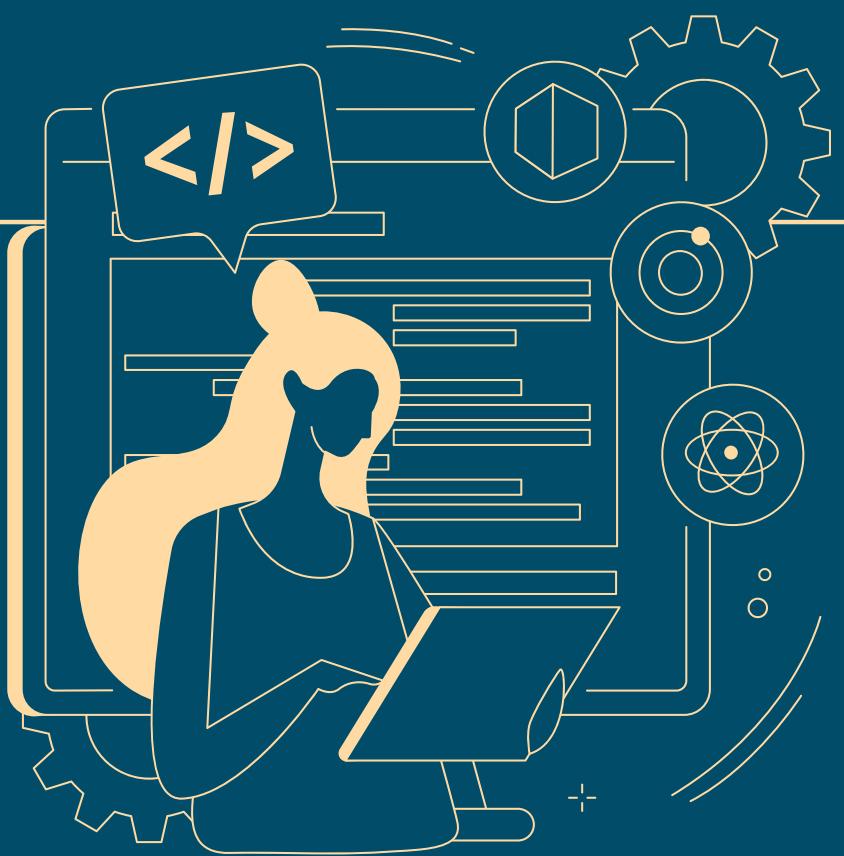


## >> METODEOLOGY

- Descriptive analysis
  - Describe the information such as, min/max value of each column, average, and the total count of data contained in each column. We also describe the distribution of the numerical data.
- Graph analysis
  - Elaborate the relation of each plot/graphic including the data percentage and assumption/hypothesis of its implication.
- Using Machine Learning Classification
  - Logistic regression

# Use Case Summary

Business Benefit  
Expected Outcome

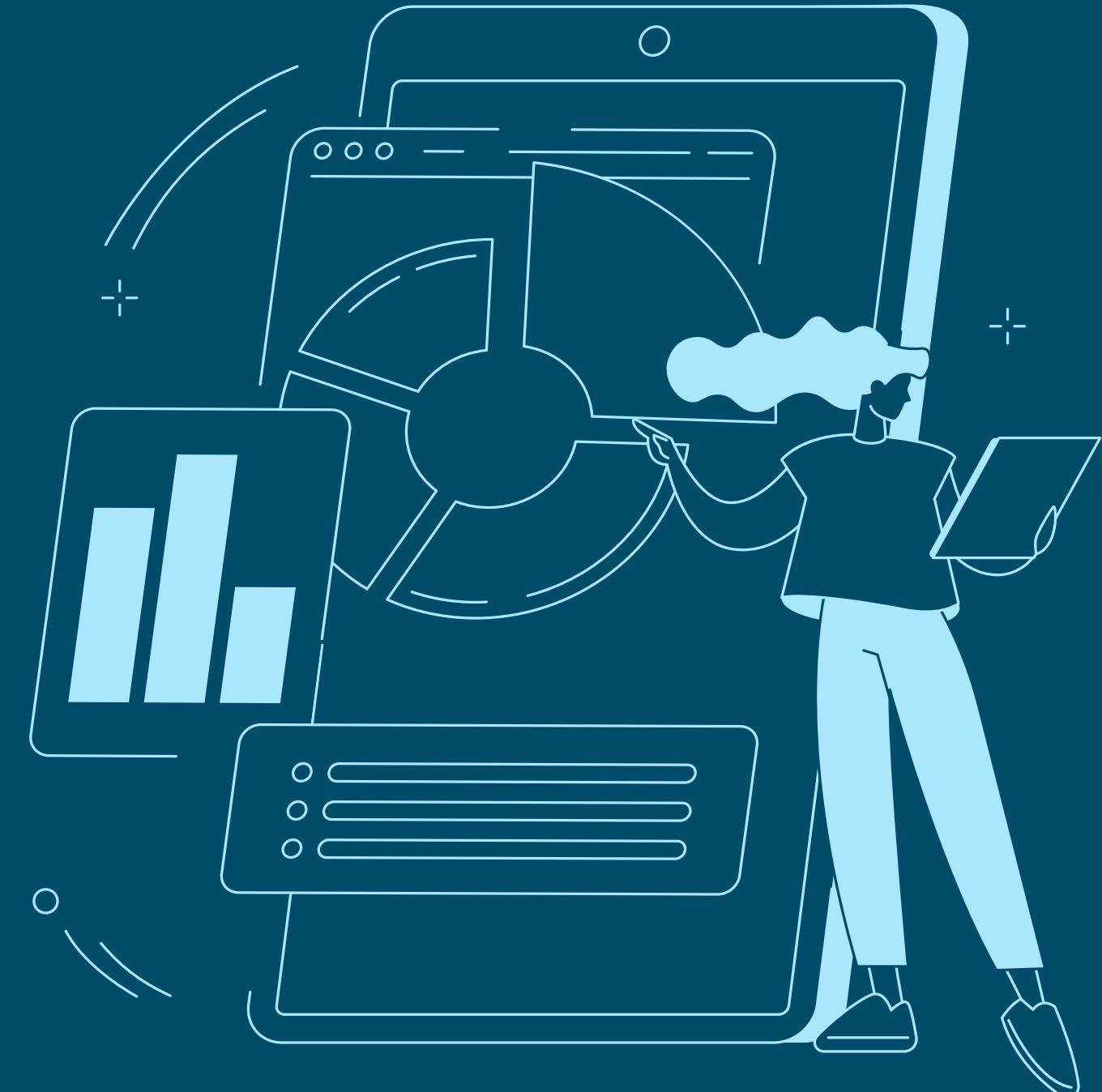


## >> BUSINESS BENEFIT

- Gain insight to keep customers from churn through modification of benefits/features provided to these customers
- Gain insight to improve the quality of company services so that customers remain loyal and gain more profit for the company
- Build models using machine learning to predict customer churn

## >> EXPECTED OUTCOME

- What are the trends these days that make your product or service possible, that may not know how many customer who churn or not churn
- Know how many female and male in churn or not churn customer
- Know how many customer who has a partner, the customer has dependent or not, and old age or not
- Know about frequently used facilities
- Know How about Among 'tenure', 'monthly Charge', 'Contract', 'Phone Service', 'gender', and others, which variable has the most influence on whether or not a customer churns churn
- Know about 4 feature variable that influenced customer to churn most.
- Know how to build models using machine learning to predict customer churn



**CUSTOMER CHURN  
SEGMENTATION**

# Business Understanding

- Data telco is a company engaged in telecommunication and internet services.
- This case has some business question using the data:
  - How many customer who churn or not churn?
  - How many female and male in churn or not churn customer?
  - How many customer who has a partner, the customer has dependent or not, and old age or not?
  - How about frequently used facilities?
  - How about Among 'tenure', 'monthly Charge', 'Contract', 'Phone Service', 'gender', and others, which variable has the most influence on whether or not a customer churns?
  - How about 4 feature or facility that influenced customer to churn most. That is, tenure, monthly charge, contract, gender male, fiber optic?
  - How to build models using machine learning to predict customer churn?





**CUSTOMER CHURN  
SEGMENTATION**

# Data Preparation

# Data Preparation



CODE USE :

Python 3.9

PACKAGES

Pandas, Numpy, Matplotlib,  
Seaborn, Sklearn, Warning  
and Feature Engine



**CUSTOMER CHURN  
SEGMENTATION**

# Data Understanding

## SOURCE DATA :

- The dataset used is data from <https://www.kaggle.com/datasets/yeanzc/telco-customer-churn-ibm-dataset>
- The raw data contains 7043 rows (customers) and 21 columns (features).



## DATA DICTIONARY :

- customerID : Customer ID
- gender : Whether the customer is a male or a female (Male, Female)
- SeniorCitizen : Whether the customer is a senior citizen or not (1, 0)
- Partner : Whether the customer has a partner or not (Yes, No)
- Dependents : Whether the customer has dependents or not (Yes, No)
- Tenure : Number of months the customer has stayed with the company
- PhoneService : Whether the customer has a phone service or not (Yes, No)
- MultipleLines : Whether the customer has multiple lines or not (Yes, No, No phone service)
- InternetService : Customer's internet service provider (DSL, Fiber optic, No)
- OnlineSecurity : Whether the customer has online security or not (Yes, No, No internet service)
- OnlineBackup : Whether the customer has online backup or not (Yes, No, No internet service)
- DeviceProtection : Whether the customer has device protection or not (Yes, No, No internet service)
- TechSupport : Whether the customer has tech support or not (Yes, No, No internet service)
- StreamingTV : Whether the customer has streaming TV or not (Yes, No, No internet service)
- Contract : Indicates the customer's current contract type: (Month-to-Month, One Year, Two Year)
- Paperless Billing: Indicates if the customer has chosen paperless billing: (Yes, No)
- Payment Method: Indicates how the customer pays their bill: (Bank Withdrawal, Credit Card, Mailed Check)
- Monthly Charge: Indicates the customer's current total monthly charge for all their services from the company.
- Total Charges: Indicates the customer's total charges, calculated to the end of the quarter specified above.
- Churn Label:
  - Yes = the customer left the company
  - No = the customer remained with the company

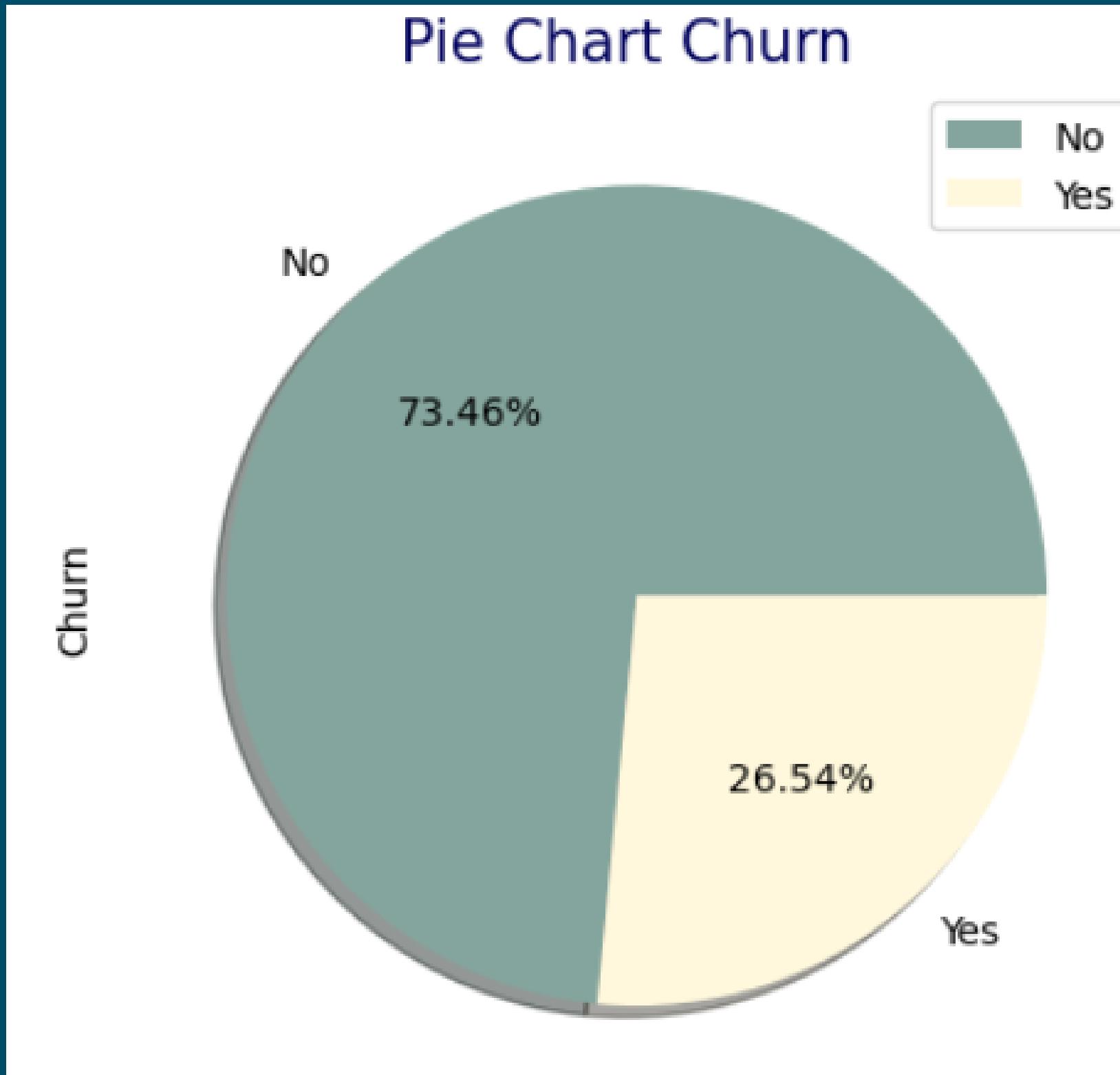


**CUSTOMER CHURN  
SEGMENTATION**

# Exploratory Data Analysis

# EDA

## How many customer who churn or not churn?

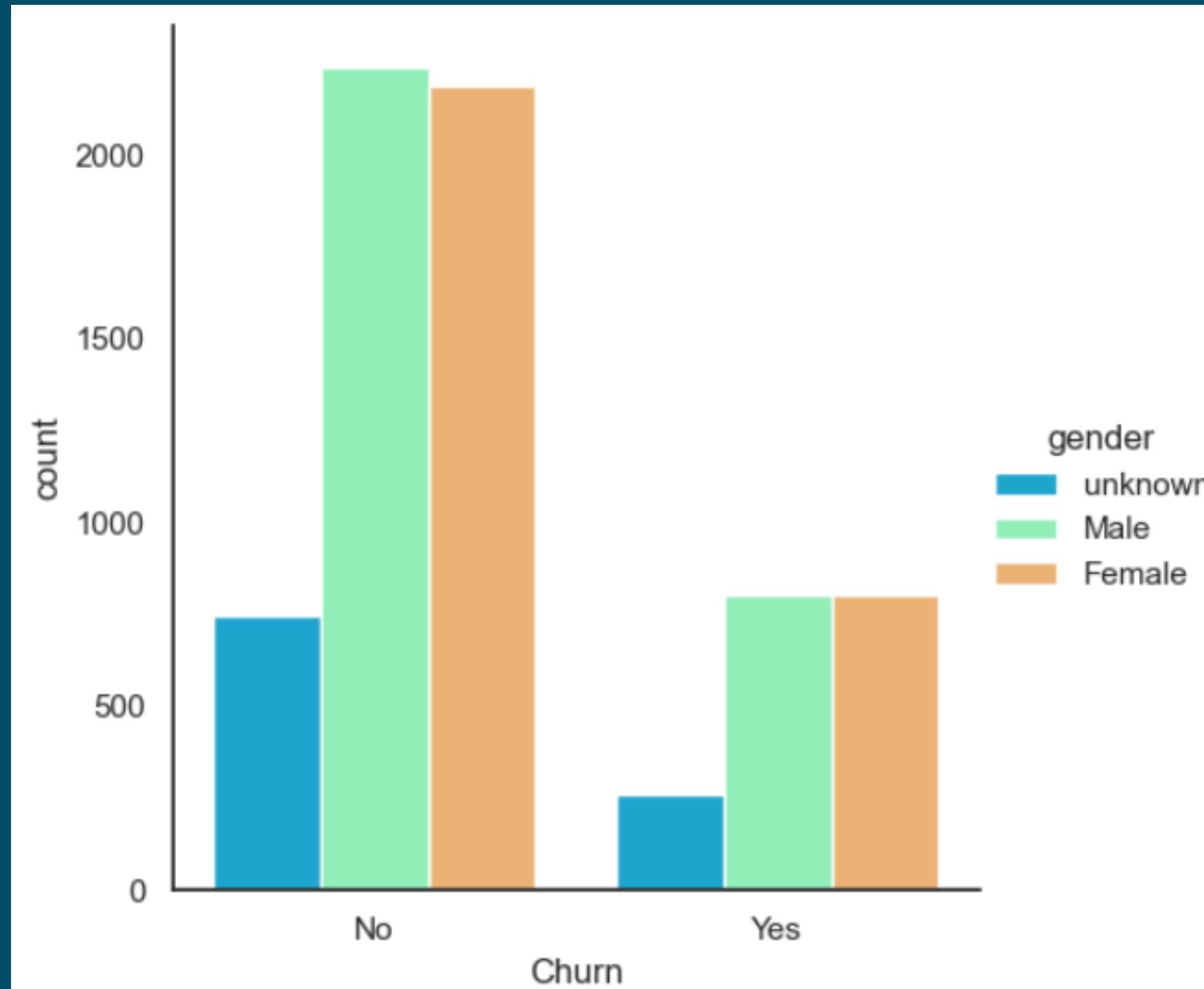


The bar chart above shows the number of customers who did not leave/stop telco services (No churn) more than customers who left/stopped telco services (Yes churn). Likewise, in the pie chart above, it can be seen that customers who do not Churn are 73.46% and customers who Churn are 26.54%.



# EDA

## How many female and male in churn or not churn customer?



Of the total 5174, those who did not churn were female as much as 42%, male as much as 43%, unknown 14%

Of the total 1869, the churn were 43% female, 42% male, Unknown 13%



# Exploratory Data Analysis



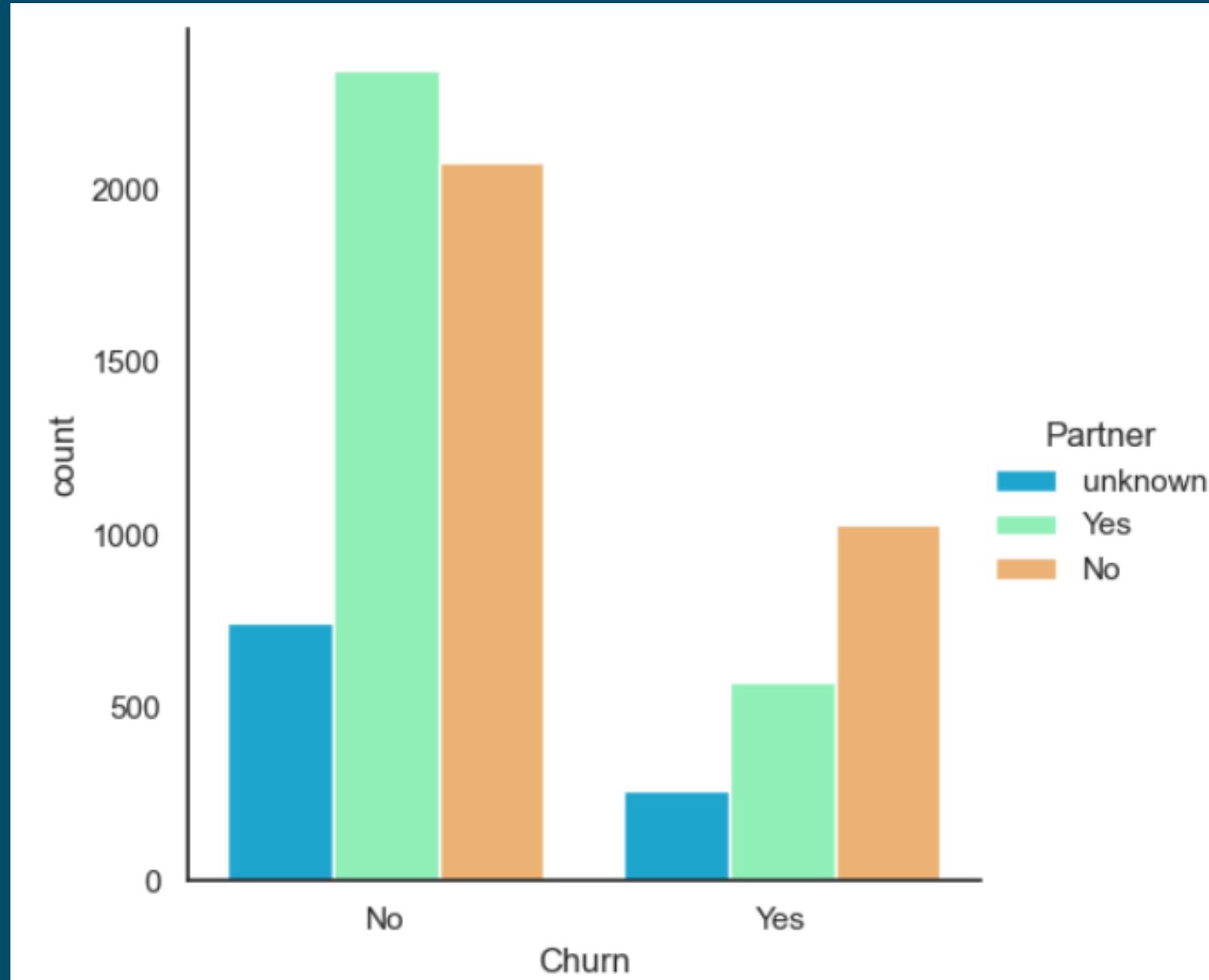
→ HOW MANY CUSTOMER HAS  
PARTNERS?

→ HOW MANY CUSTOMER HAS  
DEPENDENT?

→ HOW MANY CUSTOMERS WAS  
OLD?

# EDA

## How many customer has partners?



5174 customer who not churn, customer who no partner

40%, have a partner 45% and unknown 14%

1869 customer who churn, no partner 55%, have a partner

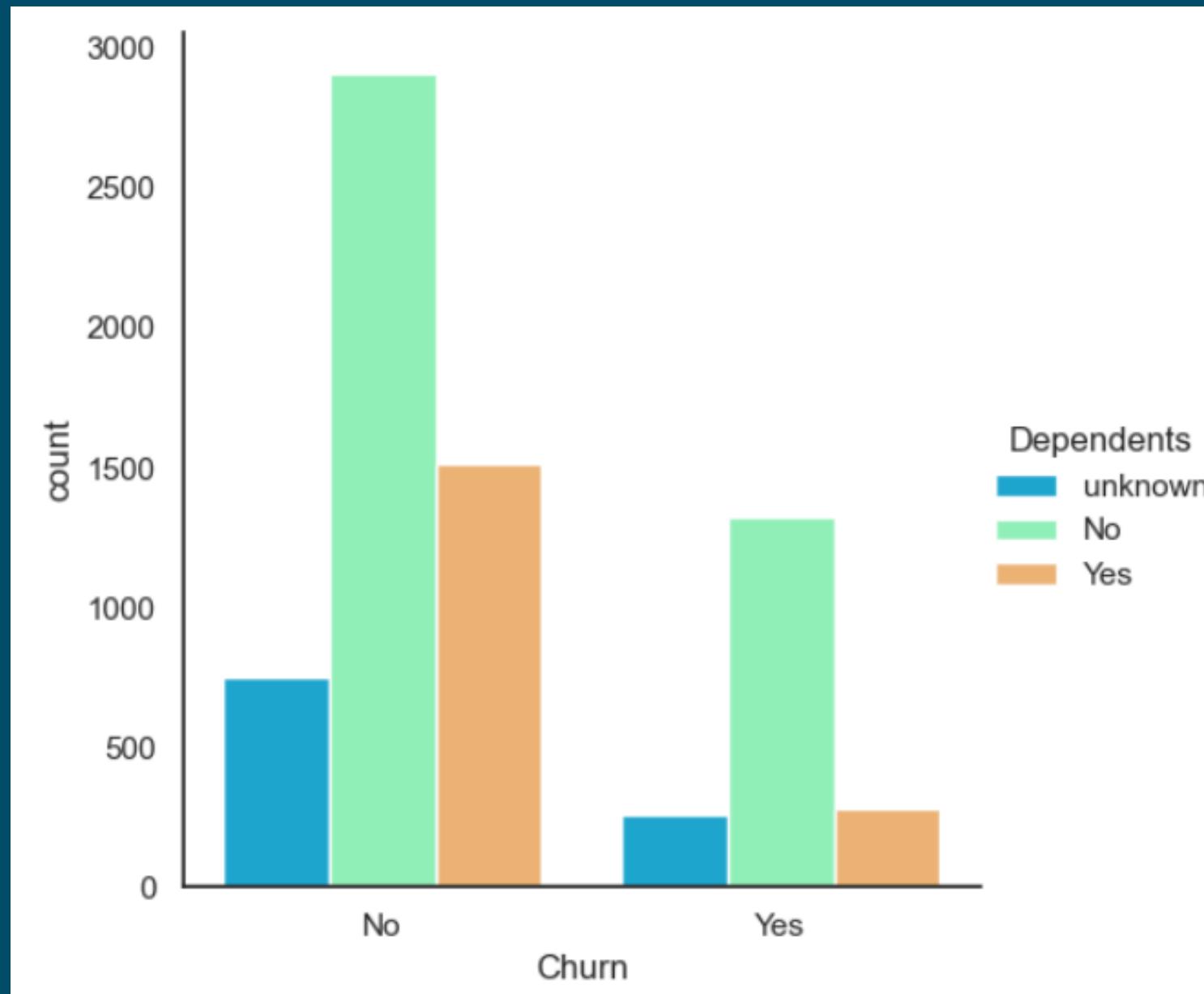
30%, unknown 13%

Why does not Churn have more partners, here we suspect that the customer is married and has a family so that the economy is quite stable.

In the churn section, we think that we are still young people who are still in the stage of managing their finances and still leaving the house more often, so they prefer to use cellular internet services.

# EDA

## How many customer has dependent?



**5174 customers not churn**, customers has no dependents 56%, have dependents 29%, unknown 14%

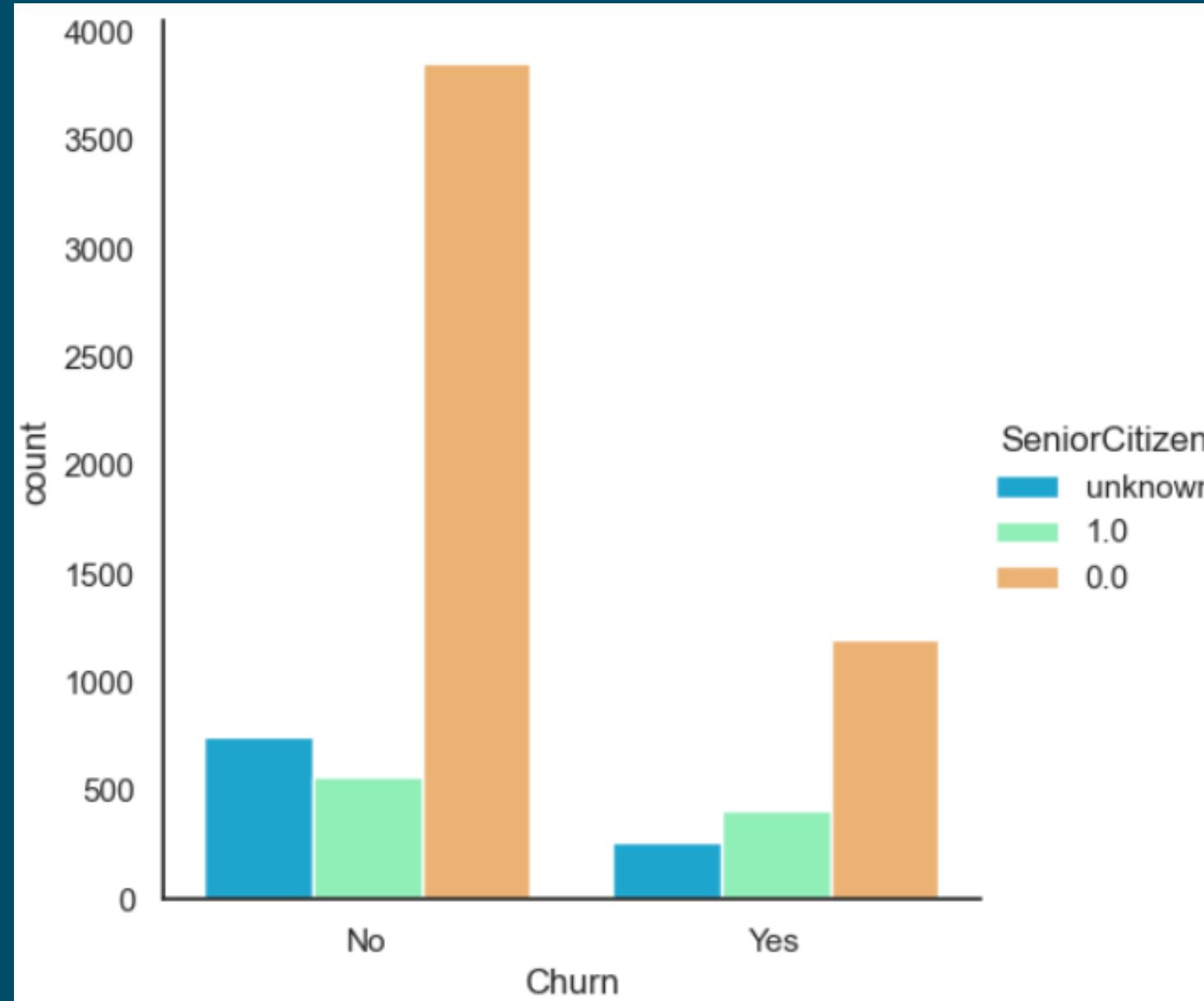
**1869 customer churn**, customer has no dependents 70%, have a dependent 15%, unknown 13%

Whether churn or not, have no dependents. because maybe many customers are not from the rich who have a lot of insurance



# EDA

## How many customers was old?



**From a total of 5174 not churn**, customers who not old age 74%, elderly 10%, and unknown 14%

**From a total of 1869 churn**, not old age 64%, elderly 21%, and Unknown 13%

From the data, both Churn and not, the elderly users are few. because also at this age we think many are 'technical' and can't keep up with the times, so if we give a discount for the elderly it will be in vain



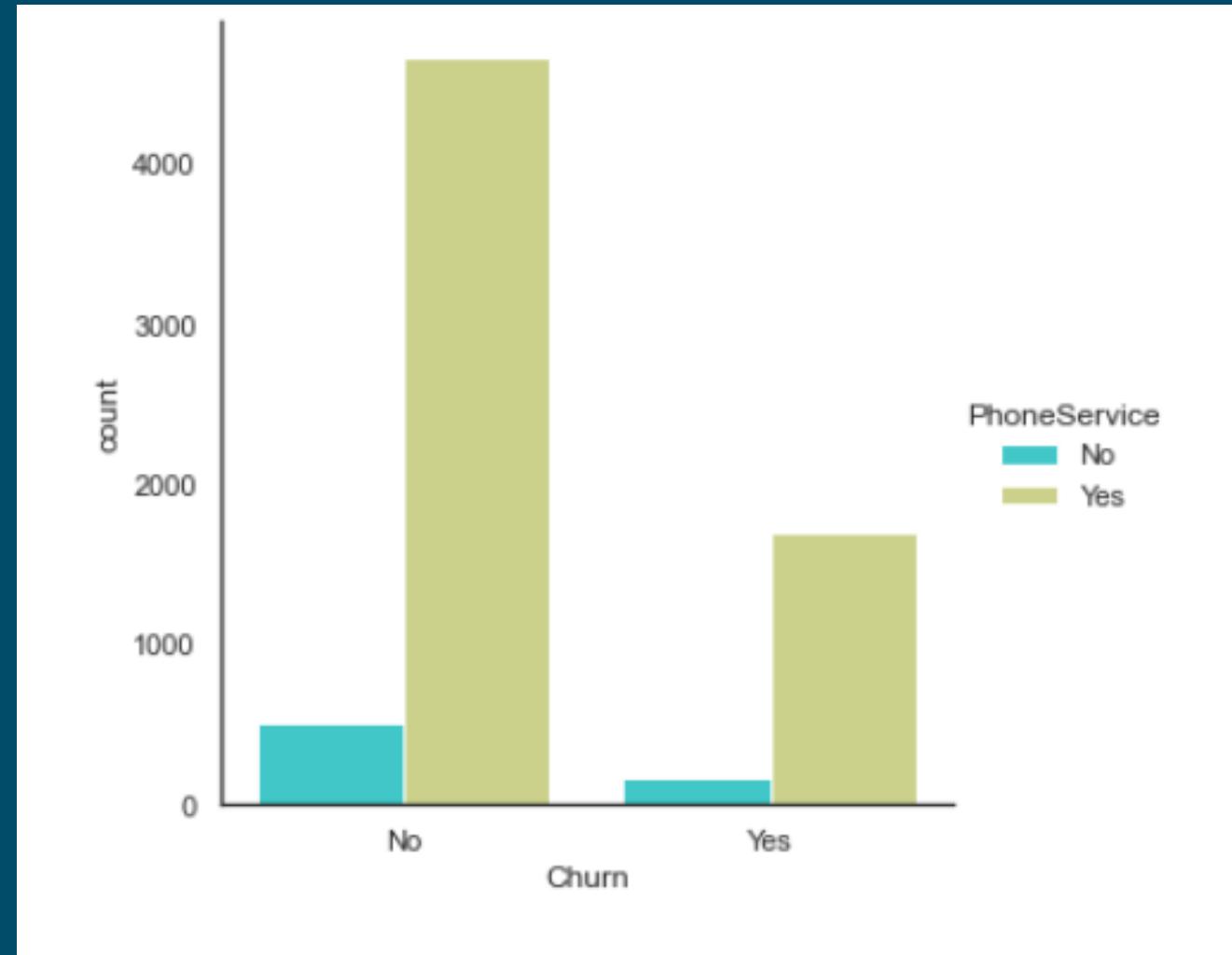
# EDA

## How many customers churn and don't churn from Phone Service

Of the total 1869, those who churn telco service and churn of Phone Service as 91 %,churn telco service and not churn of Phone Service as much as 9 %

Of the total 5174, those who did not churn telco service and churn of Phone Service as 90%, did not churn telco service and not churn of Phone Service as much as 10%

In our opinion, the Phone Service service is quite good, as evidenced by approximately 90% of customers who have Churn No or yes. In the future, in our opinion, it is better for customers who have churn yes and phone service yes so that discounts are given so that they can subscribe again (Churn == Yes)



Phone Service	No	Yes
Churn		
No	512	4662
Yes	170	1699

# EDA

## How many customers churn and don't churn from Multiple Lines



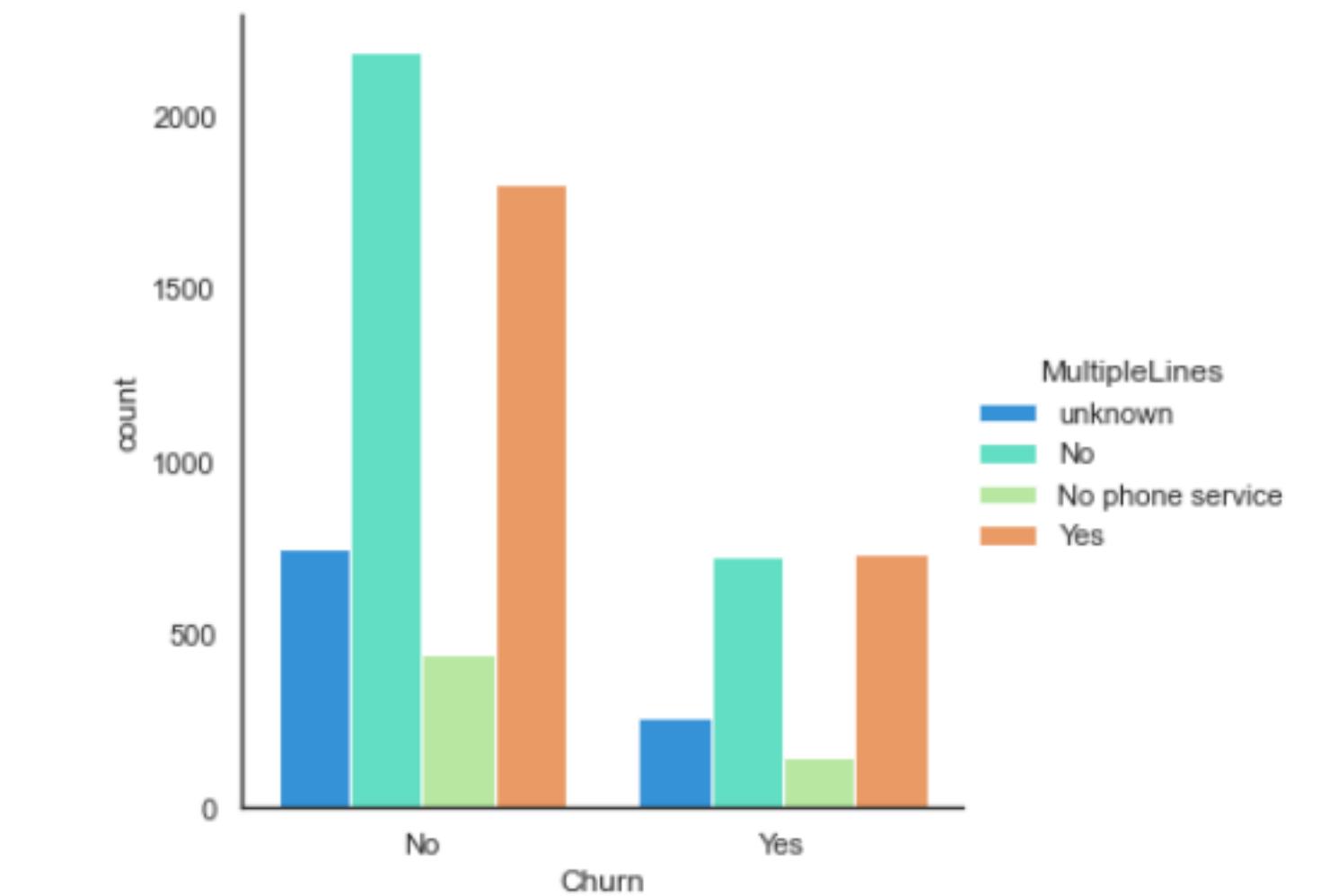
From Column Churn vs Multiple Lines

Churn no = from a total of 5174, subscribe to Multiple Lines 35% | not subscribed to Multiple Lines 42% | do not have phone service by 9% | customers whose data is unknown 14%

Churn yes = from a total of 1869, subscribed to Multiple Lines 39% | not subscribed to Multiple Lines 39% | do not have phone service by 8% | customers whose data is unknown 14%

In our opinion, multiple lines services are less affordable to customers because this is indicated by the multiple lines data for churn no (still subscribed) with a comparison of those who are still subscribed and unsubscribed +-8 : +-10. The comparison of churn yes (leaving the service) for multiple lines service with a ratio of those who are still subscribed and unsubscribed 1:1. So we can assume, the service is good (not bad) but in my opinion the service is expensive (only for certain circles)

MultipleLines	Churn				
	No	2184	442	1800	748
Yes	724	147	737	261	



# Exploratory Data Analysis



HOW MANY CHURN  
CUSTOMERS USED ADDITIONAL  
SERVICE?

# EDA

## How many customers churn and don't churn from Internet service

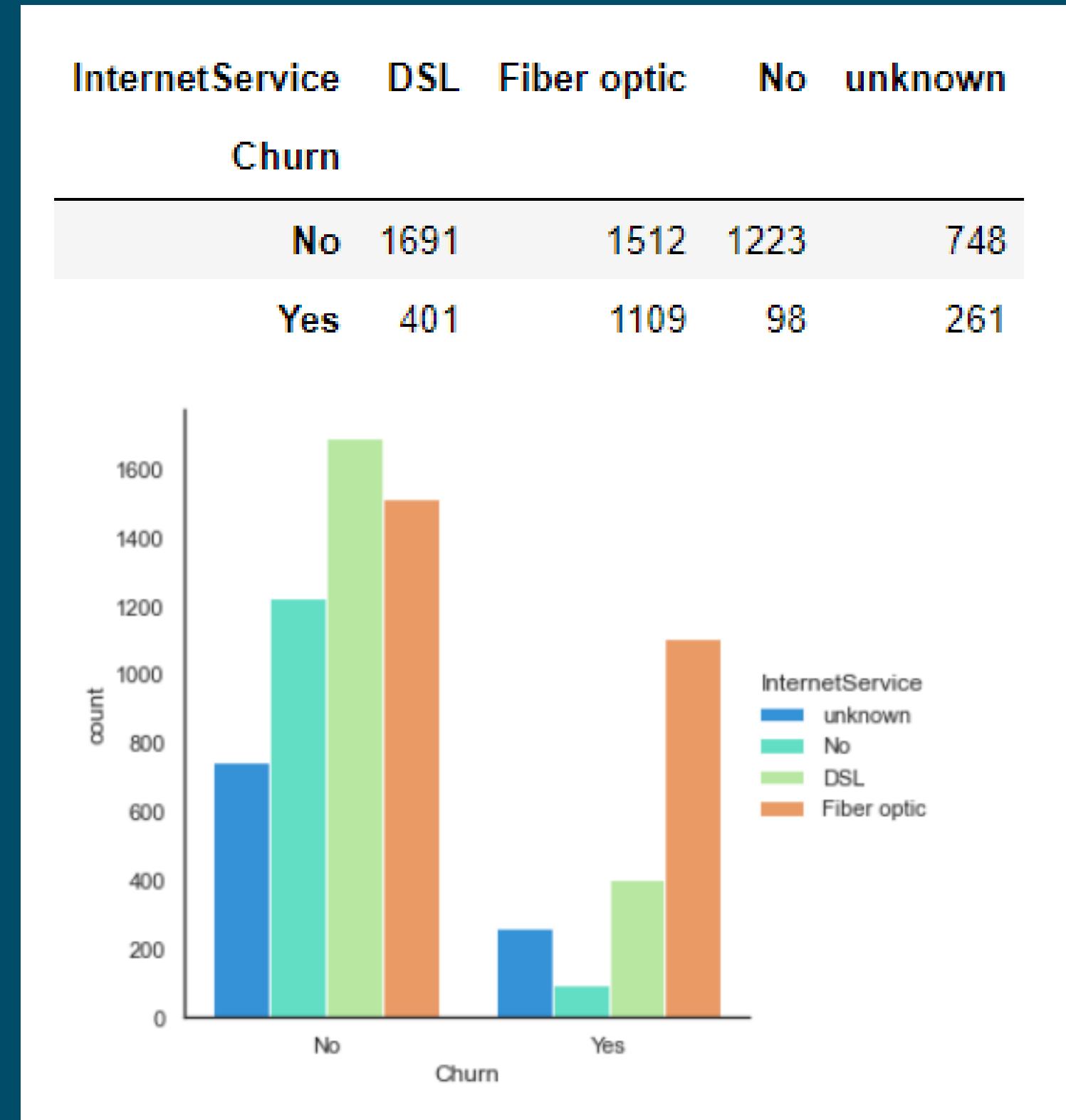


From Column Churn vs Internet Service

Churn no = out of a total of 5174, 33% DSL subscription | Fiber Optic subscription 29 % | do not subscribe to internet service by 24% | Unknown data 14%

Churn yes = out of 1869 total, 22% DSL subscription | 59% Fiber Optic subscription | don't subscribe to internet service by 5% | Unknown data 14%

In our opinion, DSL service is a superior service, as seen from the data churn No with the largest percentage, namely DSL at 33%. In our opinion, Fiber Optic service is a service that is not good, this can be seen from the churn value of yes (leaving our service) of 59%. So, every customer who leaves our service (Churn == Yes) has used Fiber Optic and Fiber Optic is worth the biggest



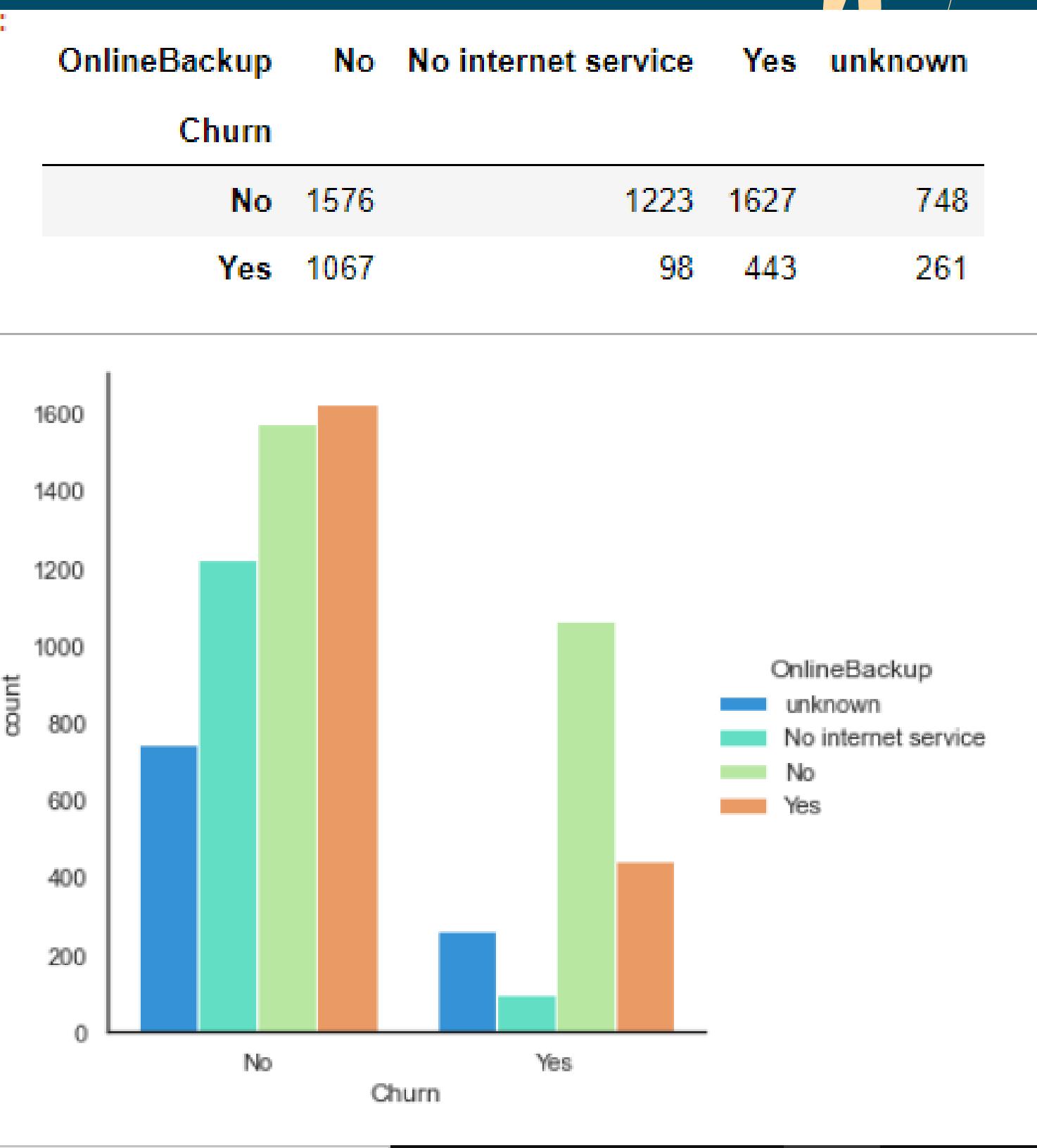
# EDA

## How many customers churn and don't churn from Online Backup

Of the 5174 customers whose Churn is No, as many as 30.45% of customers do not subscribe to Online Backup. Maybe they don't know about the Online Backup feature. as many as 23.6% do not have an internet connection. As many as 31.4% subscribed and 14.4% unknown status.

Of the 1869 customers whose Churn is Yes, as many as 57% of customers do not subscribe to Online Backup. Maybe they don't know the Online Backup feature. as much as 5.2% do not have an internet connection. A total of 23.7% subscribed and 13.4% unknown status.

Based on the data, more users use Online Backup for customers whose churn is Yes, maybe customers whose churn is Yes, most likely he has never tried the online backup feature. Because as many as 57% of 1869 customers of those whose Churn is Yes have not subscribed to Online Backup



# EDA

## How many customers churn and don't churn from Device Protection



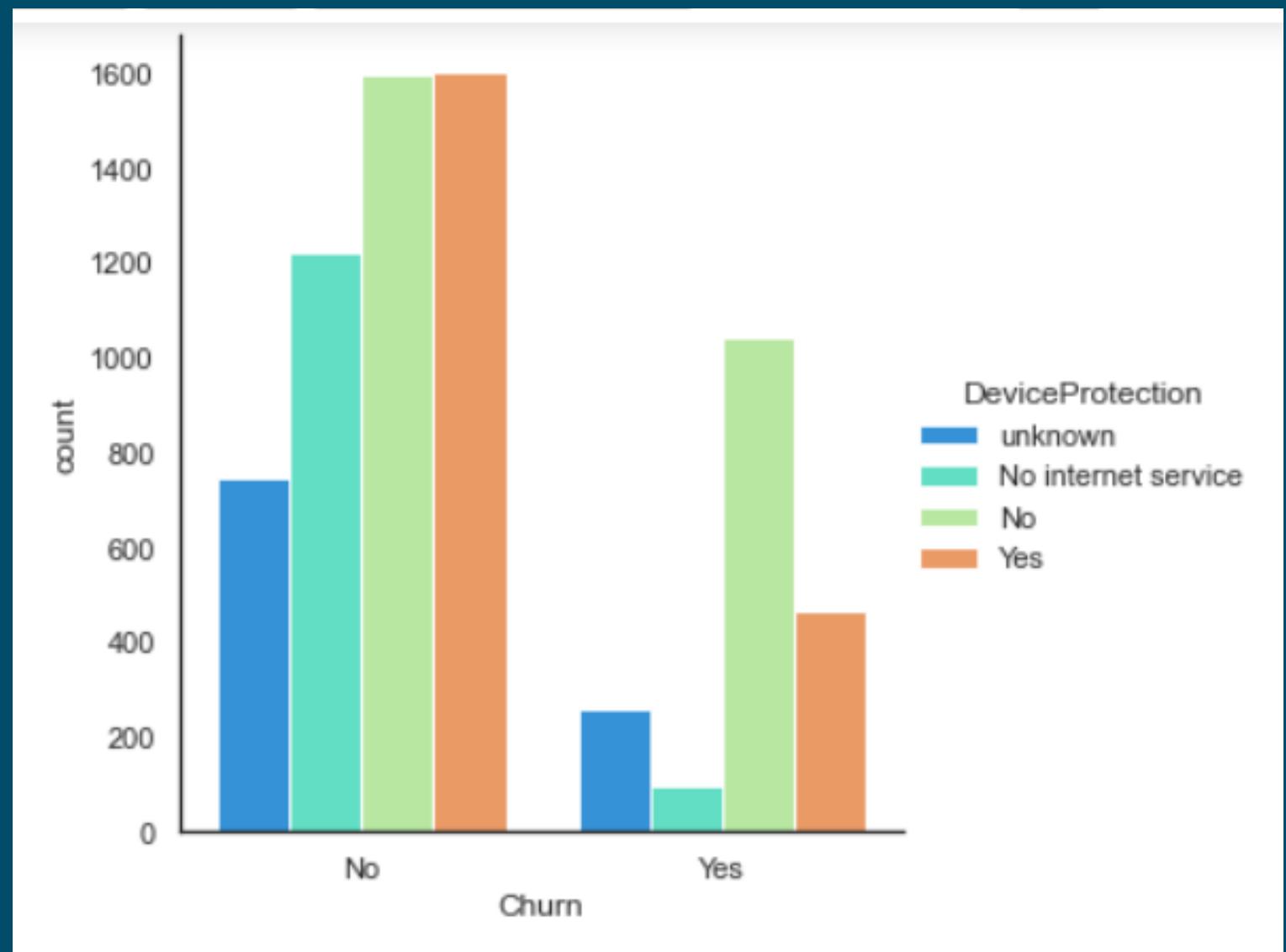
From Column Churn vs Device Protection

Churn no = from a total of 5174, subscribe to Device Protection 31% | not subscribed to Device Protection 31% | do not have internet service 24% | customers whose data is unknown 14%

Churn yes = from a total of 1869, subscribed to Device Protection 25% | not subscribed to Device Protection 56% | do not have internet service 5% | customers whose data is unknown 14%

In our opinion, device protection services are less attractive to customers. This is indicated by data on customers who churn and have never used device protection services as many as 1043 of 1869 customers (56%). customers who use and do not use device protection are 1:1. So, we assume that maybe the Device Protection service is less attractive to customers. In the future, this service can continue to be educated to customers so that customers know the advantages of this service.

DeviceProtection	No	No internet service	Yes	unknown
Churn				
No	1600	1223	1603	748
Yes	1043	98	467	261

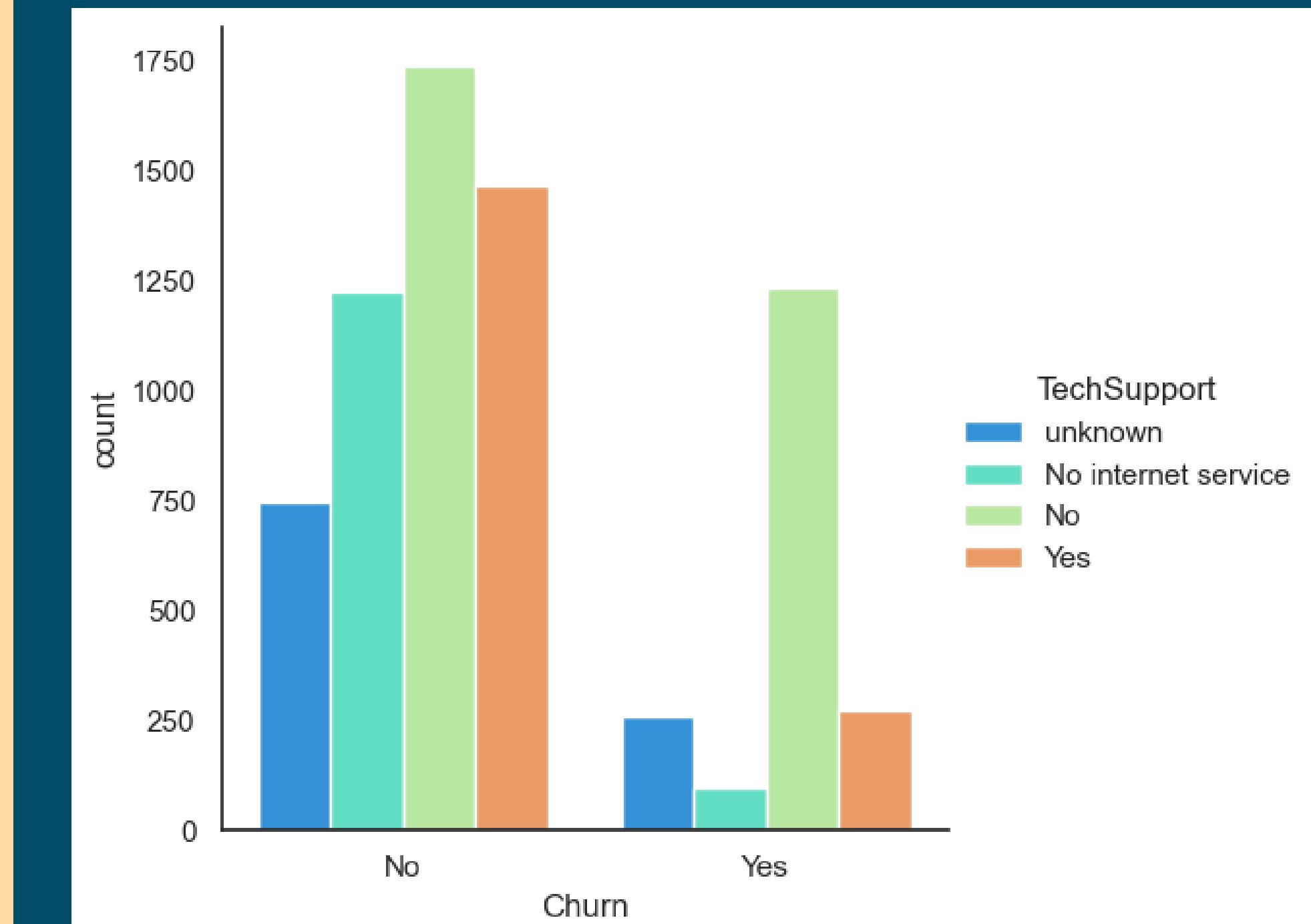


# Churn vs Tech Support

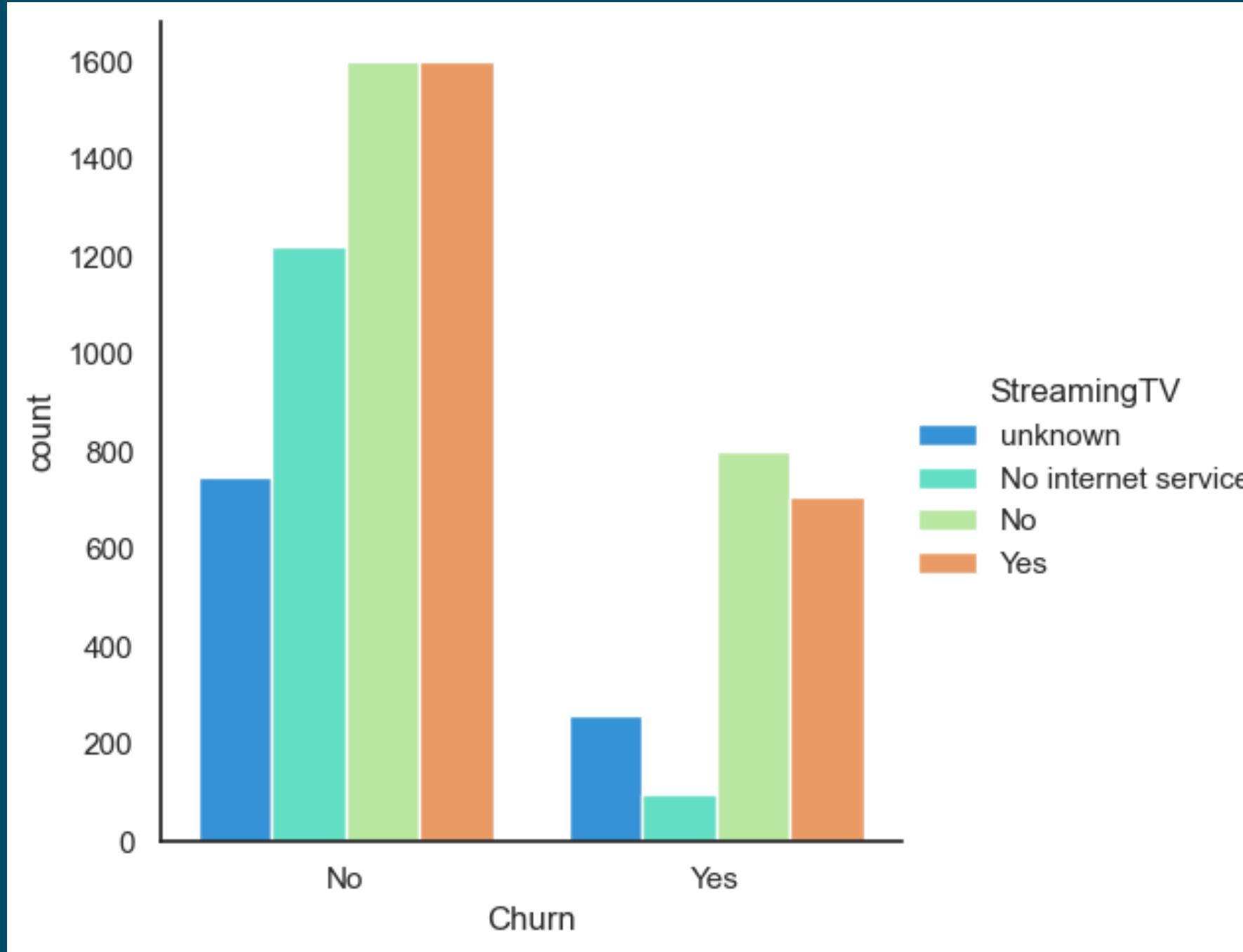
How many customer churn compared with their subscription in additional Tech Support?

→ **66% OF CHURN CUSTOMERS DID'NT SUBSCRIBE TECH SUPPORT.**

- Of the **1869 customers** whose Churn is Yes.
- Probably those customers found several technical issues and **was not able to resolve by them self.**
- Therefore, the marketing could **rebranding their strategy** for tech support plan to the customer



Tech Support	No	No internet service	Yes	unknown
Churn				
No	1738	1223	1465	748
Yes	1235	100	275	261

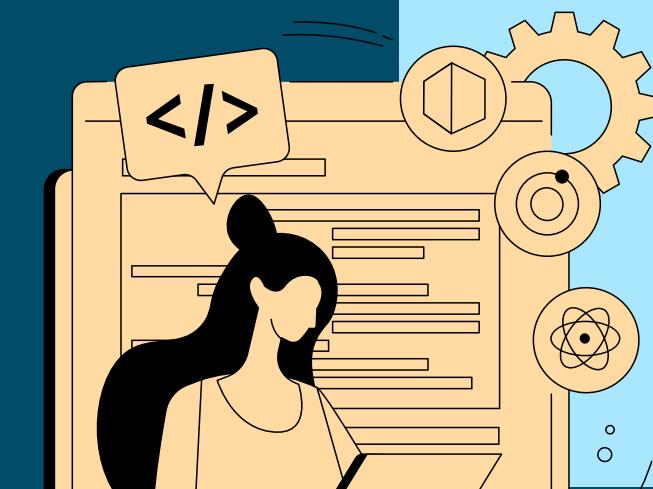


# Churn vs Streaming TV

How many customer churn compared with their usage of internet service for streaming tv programs?

- 43% OF CHURN CUSTOMERS DID'NT USE THEIR INTERNET SERVICE TO STREAM TV PROGRAM.**
- Of the **1869 customers** whose Churn is Yes.
- Probably those customers did **not interest** or did **not know yet** if they could watching tv program with internet service.
- Therefore, the creative team can **post creative contents** in the company ad or social media **to intrigue** the customers to enjoy their internet service by streaming tv.

StreamingTV	No	No internet service	Yes	unknown	
Churn	No	1601	1223	1602	748
Yes	802		98	708	261

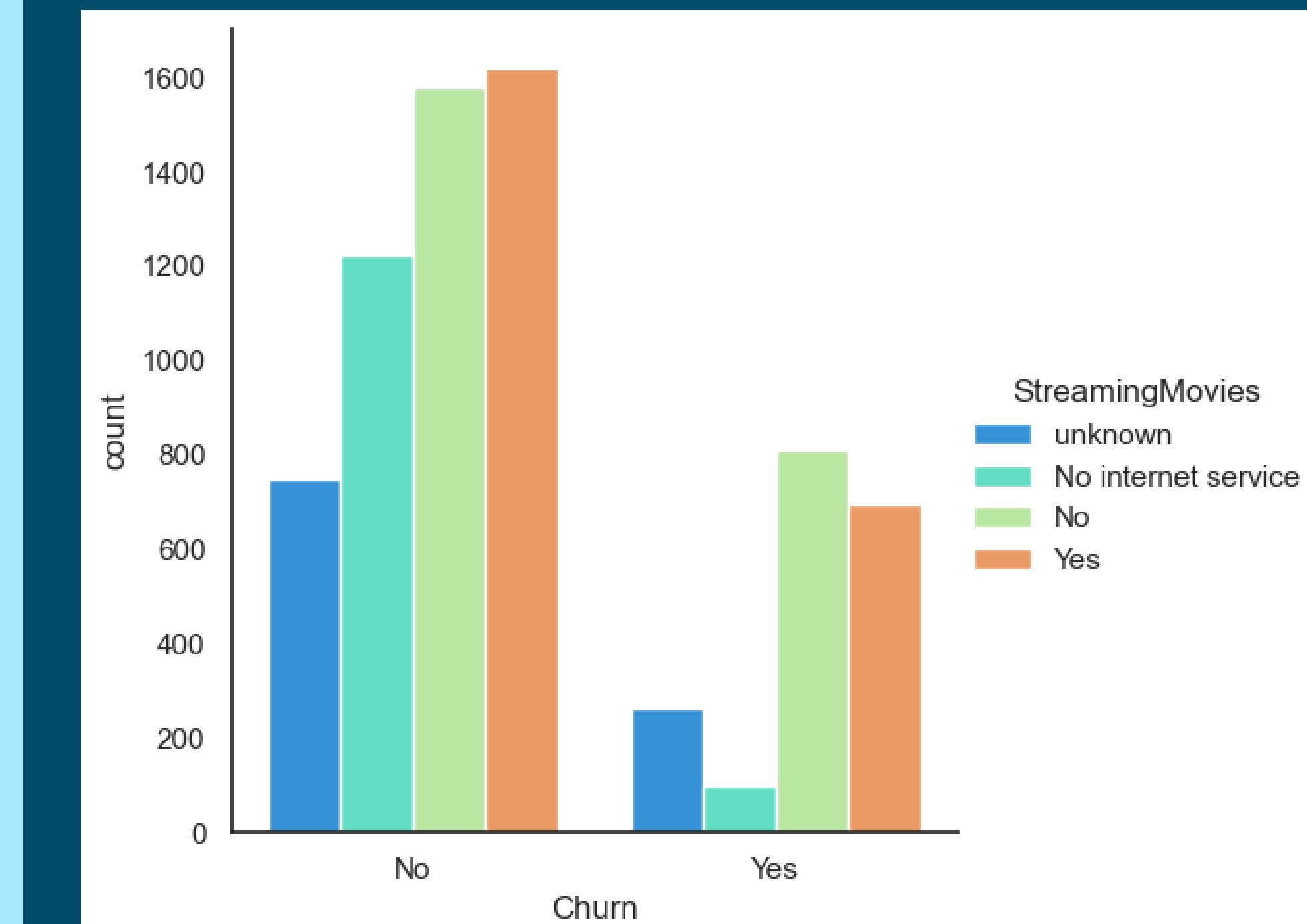


# Churn vs Streaming Movies

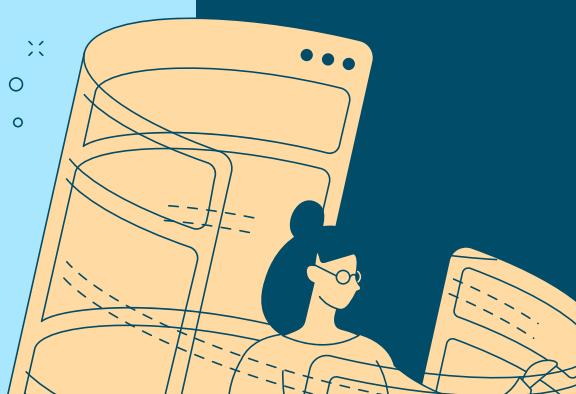
How many customer churn compared with their usage of internet service for streaming movies?

↓ **43.4% OF CHURN CUSTOMER DID'NT USE THEIR INTERNET SERVICE TO STREAM MOVIES.**

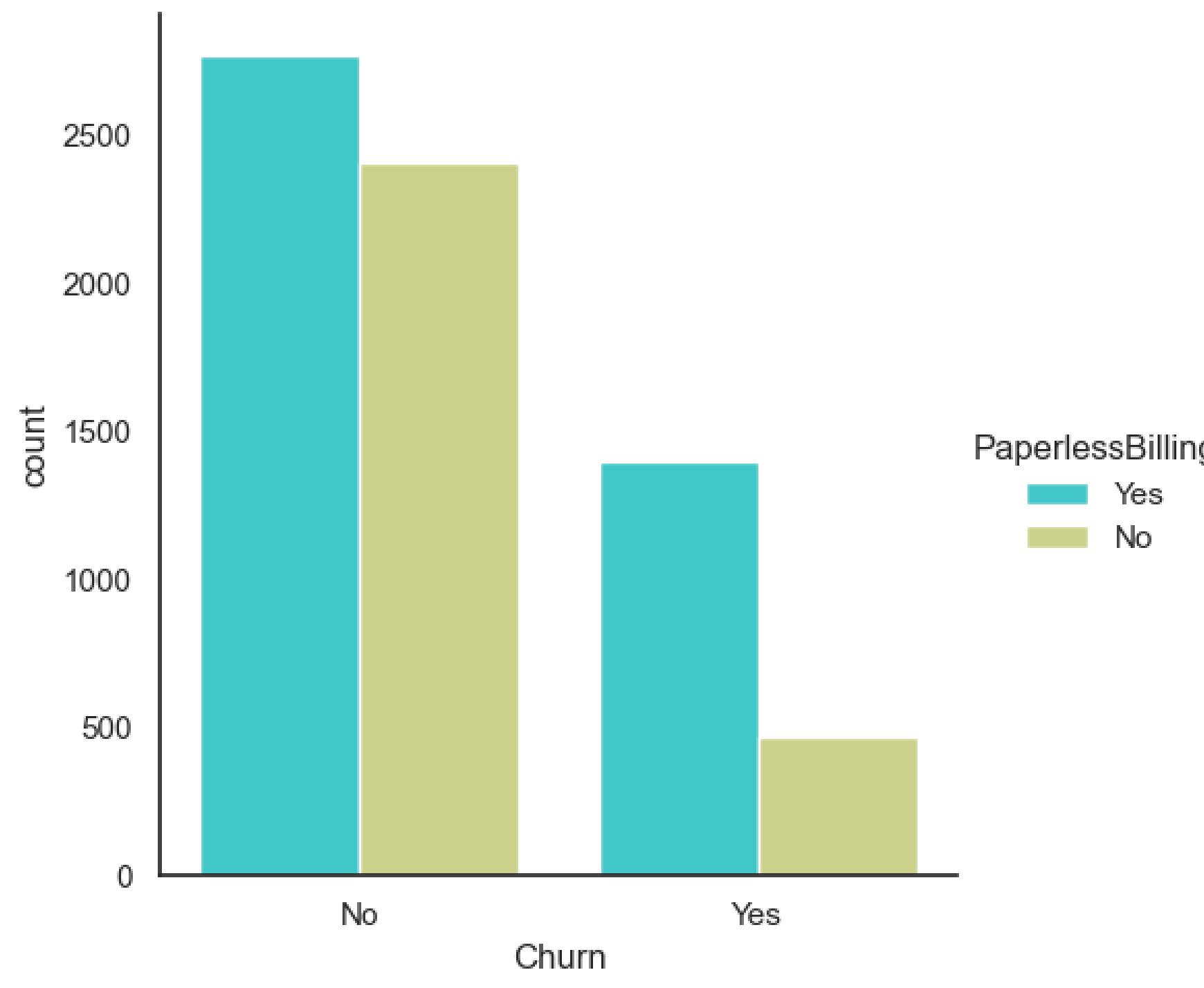
- Of the **1869 customers** whose Churn is Yes.
- Probably those customers did **not interest** or did **not know yet** if they could watching movies with internet service.
- Therefore, the creative team can **post creative contents** in the company ad or social media **to intrigue** the customers to enjoy their internet service by streaming movies.



StreamingMovies	No	No internet service	Yes	unknown
Churn				
No	1581	1223	1622	748
Yes	813	98	697	261



# Churn vs Paperless Billing



How many customer churn compared with use paperless billing and not use paperless billing?

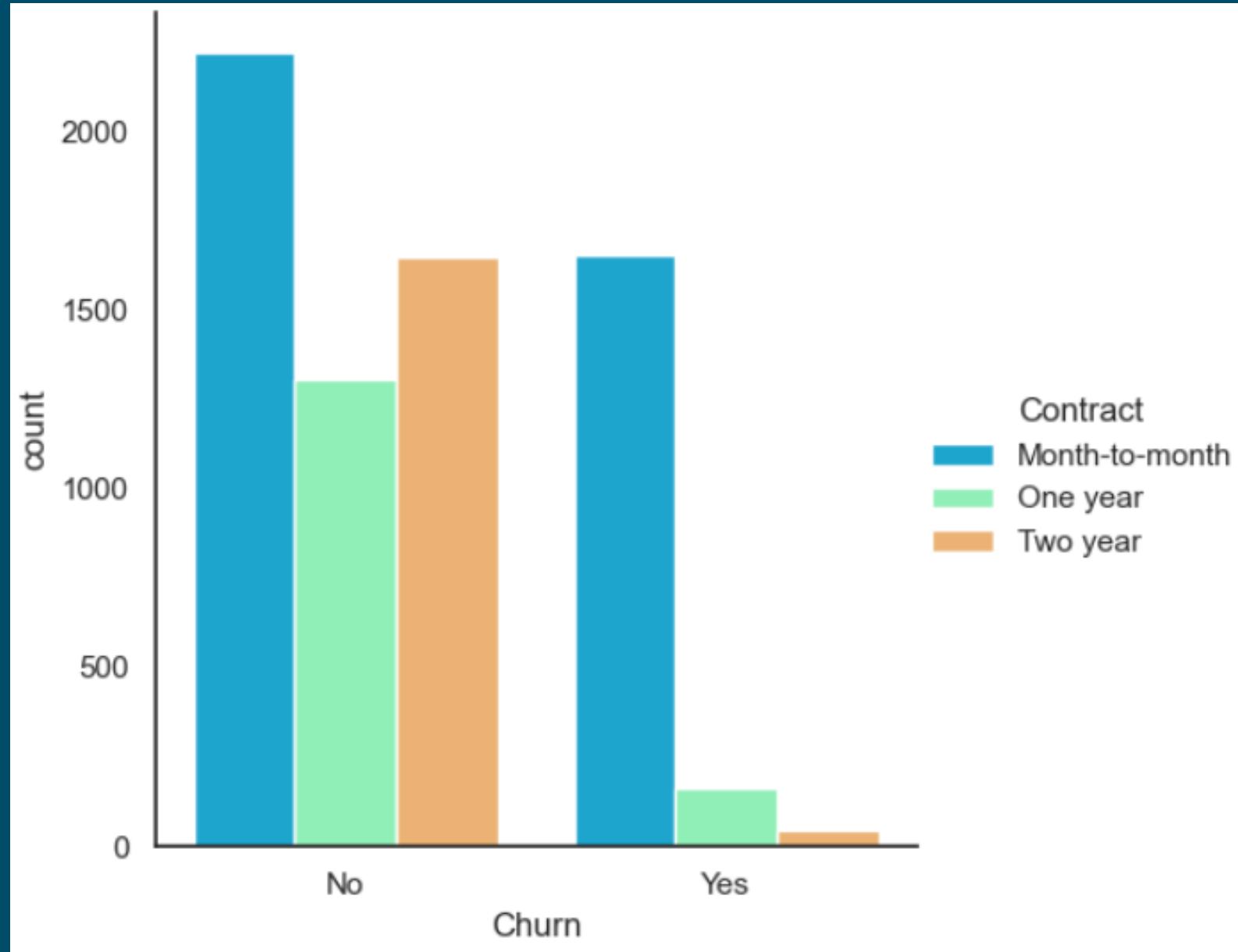
74.9% OF CHURN CUSTOMERS DID'NT CHOOSE PAPERLESS BILLING.

- Of the **1869 customers** whose Churn is Yes.
- Probably those customers did **not manage** papered billing quite well. So that, the record of transaction tend to be put aside. In addition, probably the customers **did not know** the **benefit** of paperless billing yet
- Therefore, The business team can make paperless billing **mandatory for each transaction** and optional for papered billing.



# EDA

## How long customer churn used service who has churn



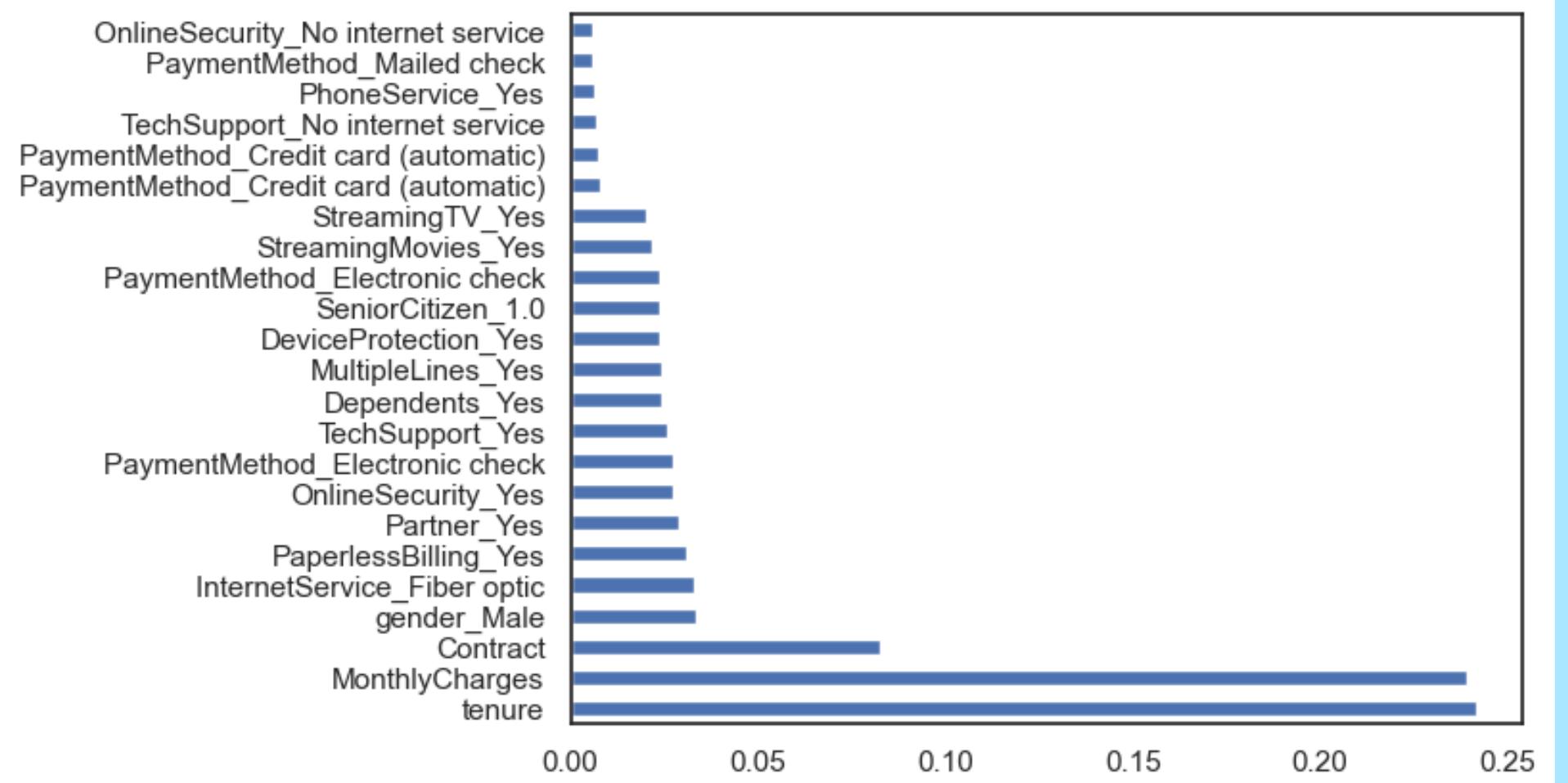
From a total of 1869 customer whose chure, month to month 88%, one year 8%, two years 2%

The barchart above shows the Churn amount by Contract category. Judging from the barchart, the number of churn customers is dominated by customers with month-to-month contracts. We think the customer is Churn because he just wants to try it, if the service is bad, the customer doesn't have to worry about extending it because it also costs less, not as much as one year and two years.

**Suggestions, improve service to new customers**

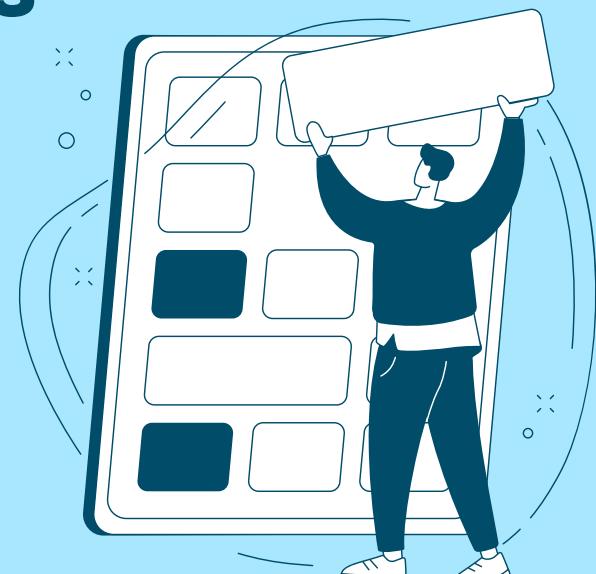
# 5

feature or variable that influenced customer to churn most



The barchart height on feature importance shows the score or how much the feature has an effect on Churn. The higher the score, the more significant it is. Based on the results besides, the 5 features that affect someone Churn or not are:

1. Tenure
2. Monthly Charges
3. Contract
4. Gender
5. Fiber Optic



# Modeling Data

## Logistic Regression



Logistic regression is an example of supervised learning. It is used to calculate or predict the probability of a binary (yes/no) event occurring. Logistic regression is used to solve classification problems.

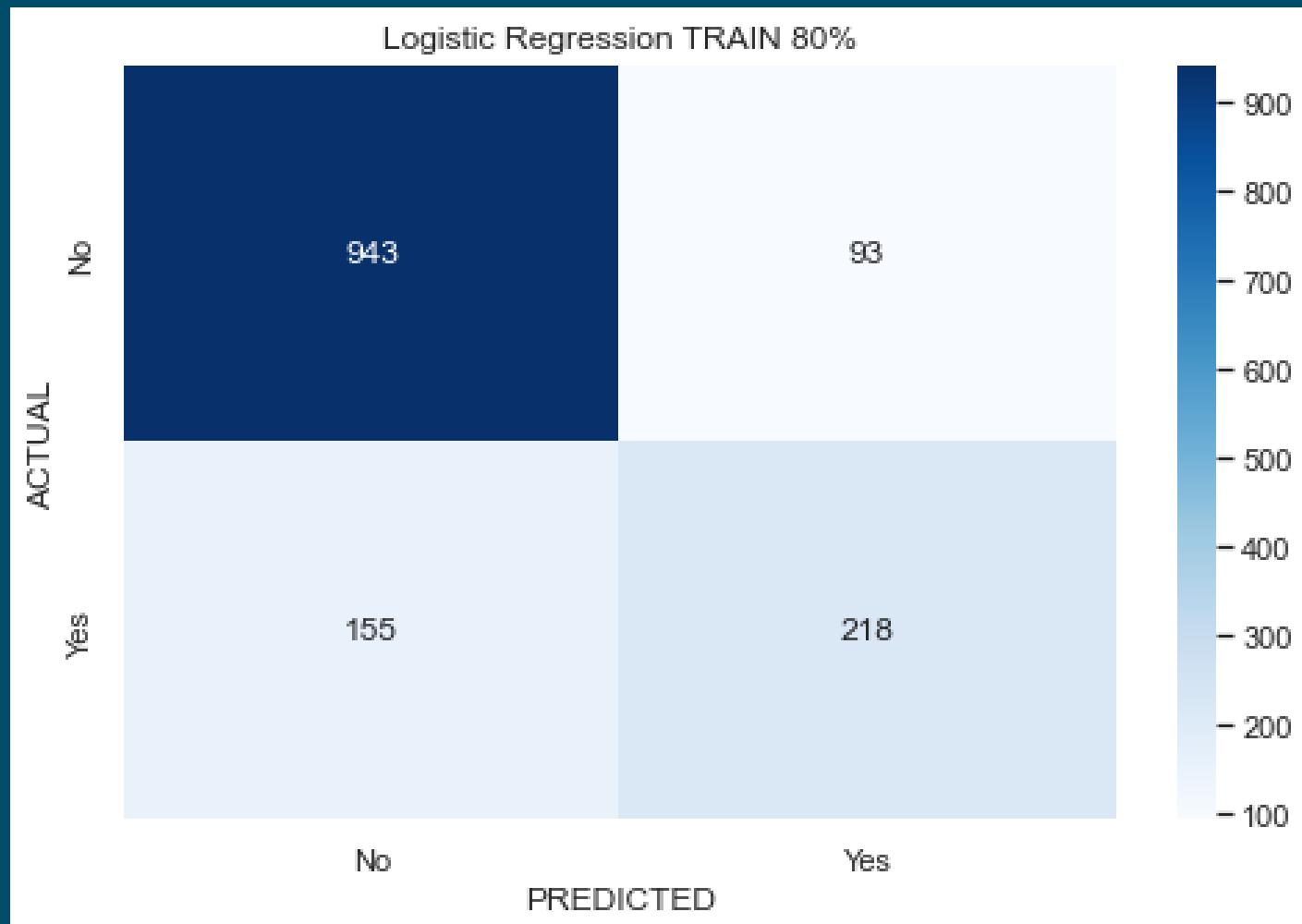
### ASSUMPTIONS :

- The factors, or the independent variables, that influence the outcome are independent of each other. In other words there is little or no multicollinearity among the independent variables.
- The independent variables can be linearly related to the log odds.
- Fairly large sample sizes.



# Logistic Regression Modeling Results

## EVALUATE MODEL



From the confusion matrix on the side, value:

- True Positive of 943 (meaning as many as 943 people who did not Churn, correctly predicted No Churn)
- True Negative of 218 (meaning as many as 218 people who Churn, correctly predicted Churn)
- False Positive of 155 (meaning as many as 155 people who Churn, incorrectly predicted No Churn)
- False Negative of 93 (meaning as many as 93 people who did not Churn, wrongly predicted Churn)

# Accuracy, Precision, Recall

From the confusion matrix, the results obtained are:

	precision	recall	f1-score	support
No	0.86	0.91	0.88	1036
Yes	0.70	0.58	0.64	373
accuracy			0.82	1409
macro avg	0.78	0.75	0.76	1409
weighted avg	0.82	0.82	0.82	1409

**Accuracy = 0,82**

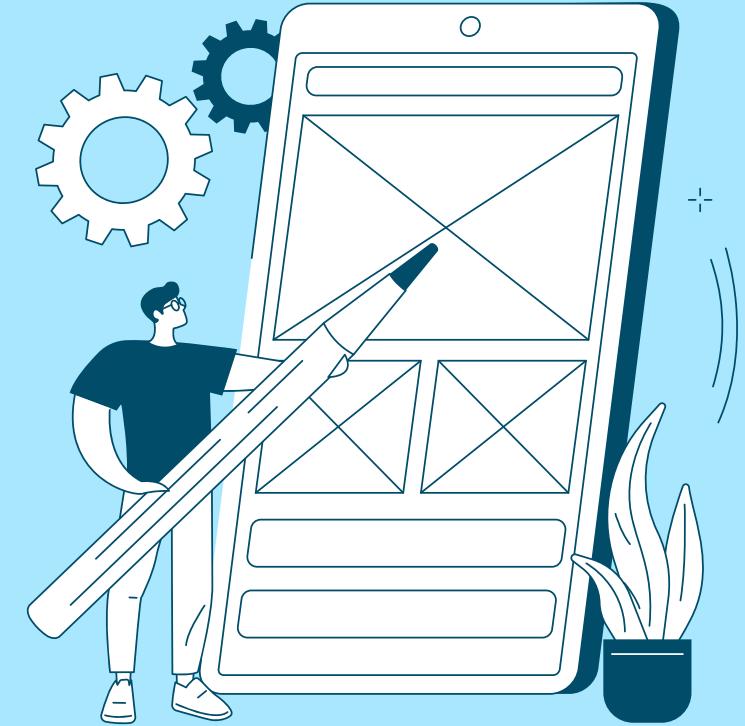
The correct customer predicted No Churn and Churn from the overall customer by 82%

**Precision = 0,86**

The correct customers did not Churn out of the total predicted customers No Churn by 86%

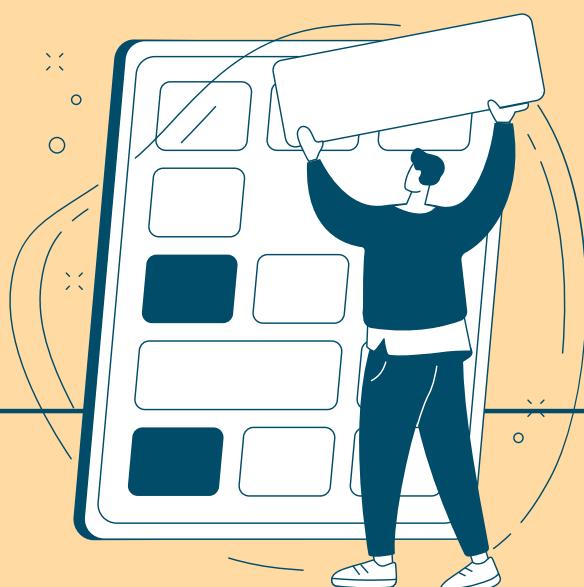
**Recall = 0,91**

Predicted customers No Churn compared to total customers who actually No Churn by 91%



## Is Model Overfit or Underfit?

Underfitting means the model has a low accuracy score on training data and testing data. Next is overfitting, overfitting has a high accuracy score on the training data, but a low score on the testing data. This means that the model made is not generalized.



Data	Accuracy
Training	79,9787%
Testing	82,3989%

In the results of the logistic regression model from telco data, the accuracy of testing data with training data is not too far away. This shows that **there is no overfitting or underfitting of the model.**

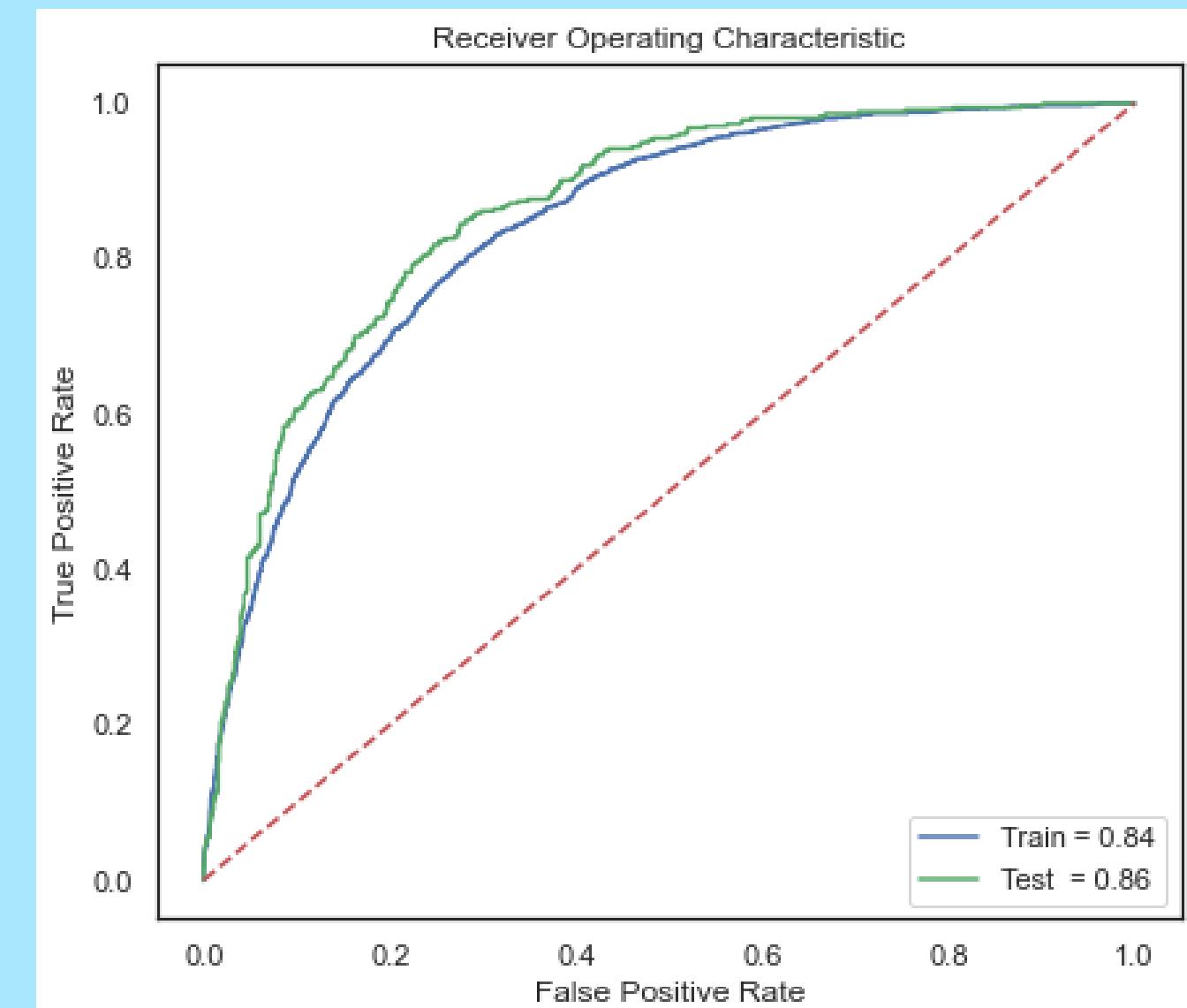
# ROC AUC

AUC is the area under the ROC curve. AUC makes it easy to compare models.

```
Train Confusion Matrix  
[[3728  410]  
 [ 718  778]]  
Test Confusion Matrix  
[[943   93]  
 [155  218]]  
  
Area Under Curve  
AUC train & test      : 83.86% & 86.06%  
  
Confusion Matrix Evaluation  
Accuracy train & test    : 79.98% & 82.40%  
Recall train & test       : 52.01% & 58.45%  
Specificity train & test: 90.09% & 91.02%  
Precision train & test   : 65.49% & 70.10%  
F1 Score train & test    : 57.97% & 63.74%  
Log Loss train & test     : 6.9152 & 6.0793
```

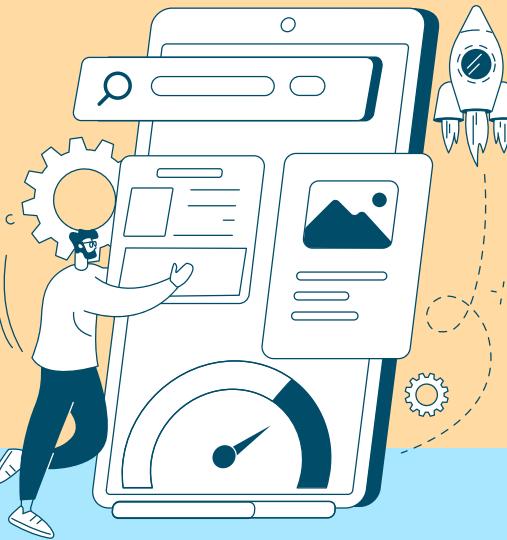


## Plot of ROC AUC Regression Logistic



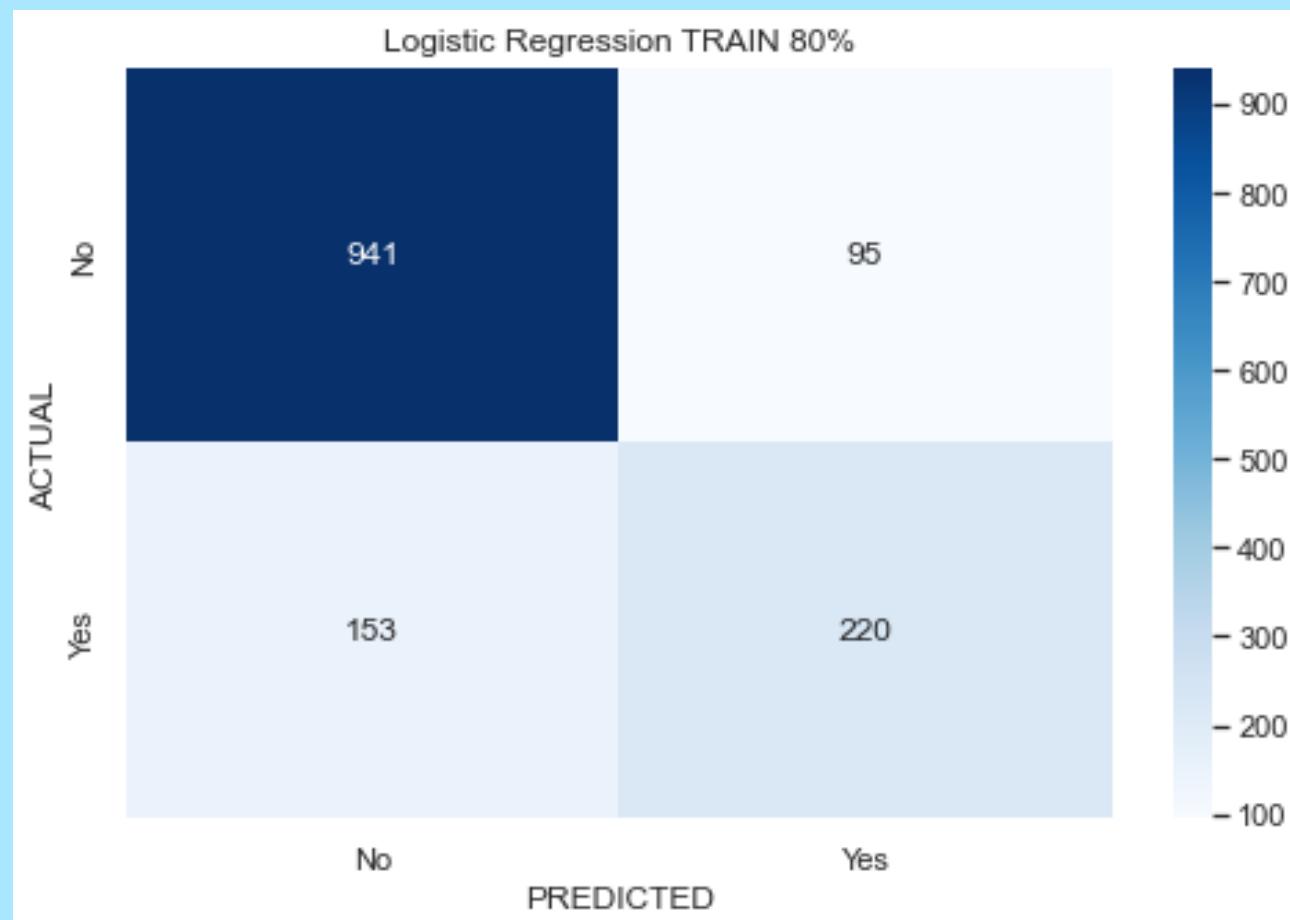
The AUC value of the **training data** is 83.86% and the AUC value of the **testing data** is 86.06%.

# Regression Logistic Model With Cross Validation and Hyperparameter Tuning Results



In logistic regression modeling, cross validation can be used in the selection of training data and testing data. Then also can use the optimum parameters by using hyperparameter tuning.

## EVALUATE MODEL



**From the confusion matrix on the side, value:**

- True Positive of 941 (meaning as many as 941 people who did not Churn, correctly predicted No Churn)
- True Negative of 220 (meaning as many as 220 people who Churn, correctly predicted Churn)
- False Positive of 153 (meaning as many as 153 people who Churn, incorrectly predicted No Churn)
- False Negative of 95 (meaning as many as 95 people who did not Churn, wrongly predicted Churn)

# Accuracy, Precision, Recall

	precision	recall	f1-score	support
No	0.86	0.91	0.88	1036
Yes	0.70	0.59	0.64	373
accuracy			0.82	1409
macro avg	0.78	0.75	0.76	1409
weighted avg	0.82	0.82	0.82	1409

↓ **ACCURACY = 0,82**

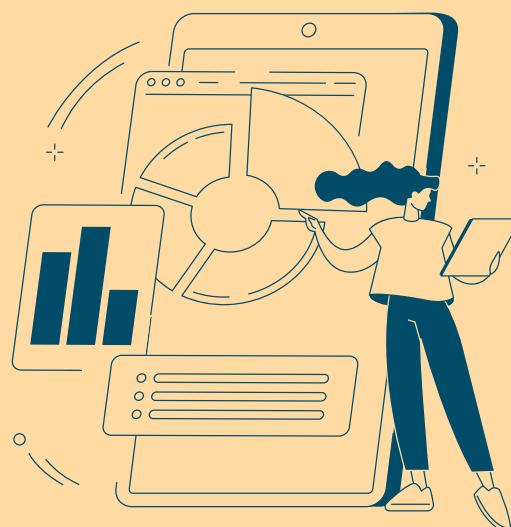
The correct customer predicted No Churn and Churn from the overall customer by 82%

↓ **PRECISION = 0,86**

The correct customers did not Churn out of the total predicted customers No Churn by 86%

↓ **RECALL = 0,91**

Predicted customers No Churn compared to total customers who actually No Churn by 91%





## Is Model Overfit or Underfit?

Accuracy value of training data and testing data from logistic regression modeling with cross validation and hyperparameter tuning :

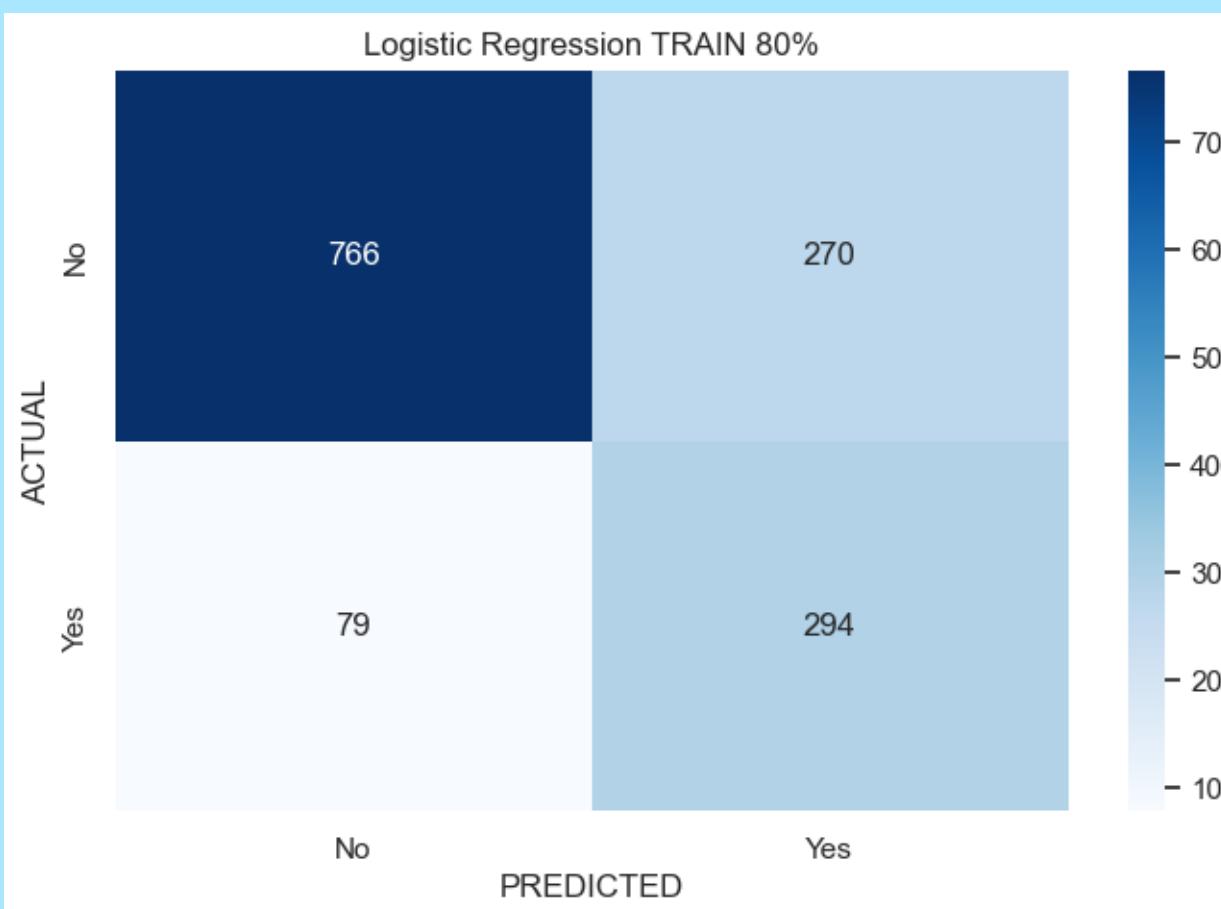
Data	Accuracy
Training	79,9978%
Testing	82,3989%

In the results of the logistic regression model from telco data, the accuracy of testing data with training data is not too far away. This shows that **there is no overfitting or underfitting** of the model.

# Oversampling SMOTE for Logistic Regression Model Result

In Churn data, the value between Churn and non Churn Customers is imbalanced. That way oversampling is done using SMOTE (Synthetic Minority Oversampling Technique) to overcome this.

## EVALUATE MODEL



From the confusion matrix on the side, value:

- True Positive of 766 (meaning as many as 766 people who did not Churn, correctly predicted No Churn)
- True Negative of 294 (meaning as many as 294 people who Churn, correctly predicted Churn)
- False Positive of 79 (meaning as many as 153 people who Churn, incorrectly predicted No Churn)
- False Negative of 270 (meaning as many as 270 people who did not Churn, wrongly predicted Churn)



# Accuracy, Precision, Recall

	precision	recall	f1-score	support
No	0.91	0.74	0.81	1036
Yes	0.52	0.79	0.63	373
accuracy			0.75	1409
macro avg	0.71	0.76	0.72	1409
weighted avg	0.80	0.75	0.76	1409

**Accuracy = 0,75**

The correct customer predicted No Churn and Churn from the overall customer by 75%

**Precision = 0,91**

The correct customers did not Churn out of the total predicted customers No Churn by 91%

**Recall = 0,74**

Predicted customers No Churn compared to total customers who actually No Churn by 74%



# Is Model Overfit or Underfit?

Accuracy value of training data and testing data from logistic regression modeling with oversampling SMOTE :

Data	Accuracy
Training	78,6129%
Testing	75,2397%

In the results of the logistic regression model from telco data, the accuracy of testing data with training data is not too far away. This shows that **there is no overfitting or underfitting** of the model.

## Comparing the data accuracy of the three models (without CV and hyperparameter tuning, with CV and hyperparameter tuning, and using SMOTE oversampling)



Data	Accuracy Without CV and Hyperparameter Tuning	Accuracy With CV and Hyperparameter Tuning	Accuracy With Oversampling SMOTE
Training	79,9787%	79,9978%	78,6129%
Testing	82,3989%	82,3989%	75,2397%

- The accuracy value after using CV and hyperparameter tuning is not much different from the model without CV and hyperparameter tuning. However, the accuracy value of the oversampling model in the testing data is far from other models. So SMOTE over sampling is optional, it turns out that when applied to telco data, the accuracy decreases.
- The greatest value of training and testing data accuracy is to use Cross Validation and Hyperparameter Tuning.

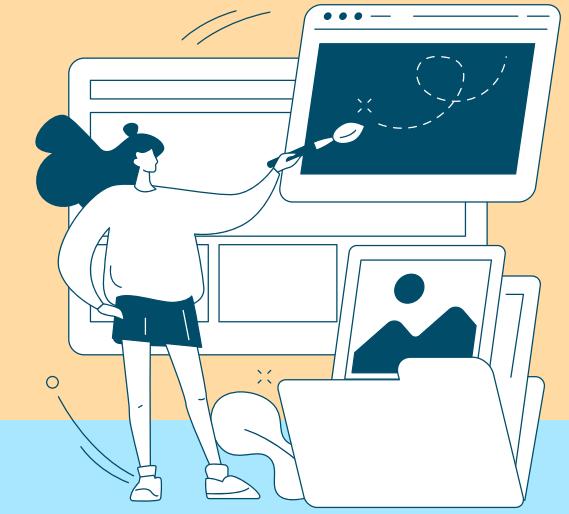
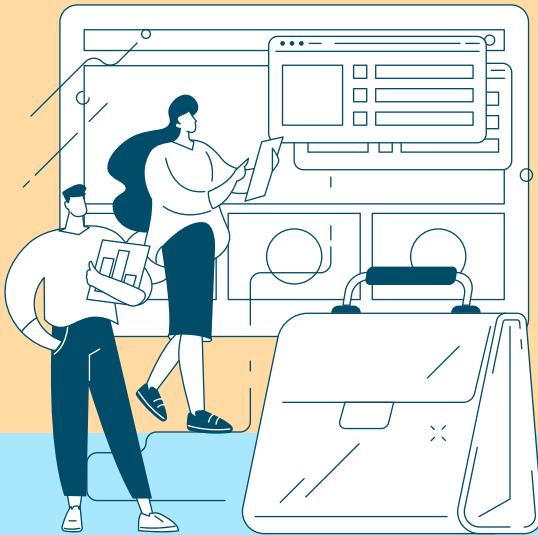
**From the results of good accuracy, the model chosen is logistic regression using Cross Validation and Hyperparameter tuning.**



**CUSTOMER CHURN  
SEGMENTATION**

# Result

# Result



## HOW MANY CUSTOMER CHURN OR NOT CHURN?

From a total of 7043 customers :

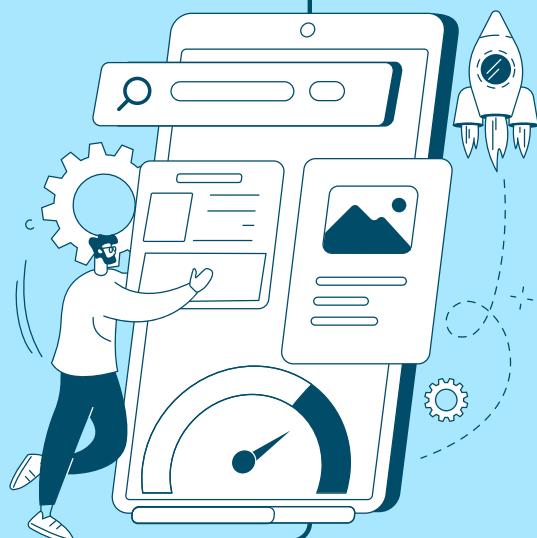
- 73.46% No churn
- 26.54% Churn



## HOW MANY CUSTOMER FEMALE AND MALE?

Of the total 5174, those who did not churn were female as much as 42%, male as much as 43%, unknown 14%

Of the total 1869, the churn were 43% female, 42% male, Unknown 13%



# Result



## HOW MANY CUSTOMER CHURN, WHO HAS A PARTNER, HAS DEPENDENT OR NOT, AND OLD AGE OR NOT?

### PARTNER

Out of a total of 1869 customer churn :

Customer has no partner as much as 55%, customer who has a partner 30%, unknown 13%

### DEPENDENT

Out of a total of 1869 customers churn :

Customer has no dependents as much as 70%, customer have a dependent 15%, unknown 13%

### SENIOR CITIZENS

Out of the total 1869 customer churn :

non-elderly customers as much as 64%, elderly customers as much as 21%, Unknown 13%



# Result



## HOW MANY CHURN CUSTOMERS USED ADDITIONAL SERVICE?

### TELEPHONE SERVICE

Out of a total of 1869 customer churn :

91% of customers used Telephone Service, 9% not used Telephone Service

### MULTIPLE LINES

of a total of 1869 customer churn:

Customer use Multiple Lines 39%, not use Multiple Lines 39%, do not have telephone service by 8%, customers whose data is unknown 14%

### INTERNET SERVICES

of 1869 total customer churn :

22% use DSL, 59% use Fiber Optic, 5% not using internet service, unknown data 14%



# Result



## HOW MANY CHURN CUSTOMERS USED ADDITIONAL SERVICE?

### ONLINE BACKUP

Of the 1869 customers whose Churn:

57% not use Online Backup. as much as 5.2% do not have an internet connection. A total of 23.7% use Online Backup and 13.4% unknown status

### DEVICE PROTECTION

Of the 1869 customers whose Churn :

55% of customers not use Device Protection. as much as 5.2% do not have an internet connection. A total of 24.9% subscribed and 13% unknown status

### TECH SUPPORT

Of the 1869 customers whose Churn :

as many as 66% of customers have not done Tech Support. Probably those customers found several technical issues and was not able to resolve by them self.



# Result



## HOW LONG CUSTOMER CHURN USED SERVICE WHO HAS CHURN?

### STREAMING TV

Of the 1869 customers whose Churn :

not using their internet service for streaming tv 43%, do not have internet service by 5.2%, subscribe to tech support 37.2, unknowns of 13.9%

### STREAMING MOVIES

Of the 1869 customers whose Churn :  
not using their internet service for streaming movies 43.4%, do not have internet service by 5.2%, subscribe to tech support 37.2%, unknowns of 13.96%

### PAPERLESS BILLING

Of a total of 1869 whose churn:

billing without receipts 25%, with 74% receipt

# Result



## HOW MANY CHURN CUSTOMERS USED ADDITIONAL SERVICE?

### CONTRACT

From a total of 1869 costumer who churn :

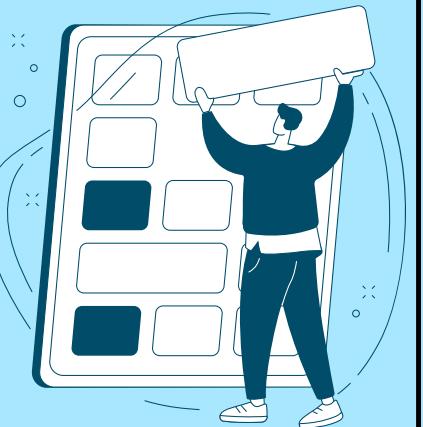
customers who use month to month 88%, one year 8%, and two years 2%

# Result



Based on the result of the feature importance, there are 5 features that likely affect the customers to Churn or not are:

- 1. Tenure Months**
- 2. Monthly Charges**
- 3. Contract**
- 4. Gender**
- 5. Fiber Optic**

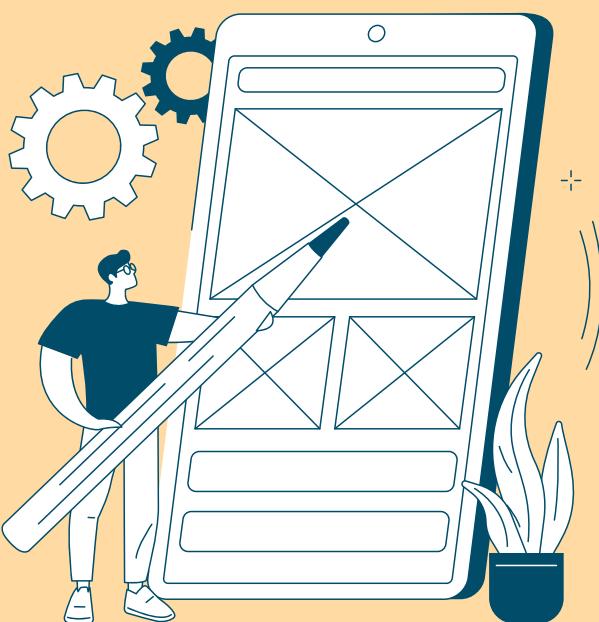


# Result

## Machine Learning

From the third model that has been tried/done, the best model with the highest accuracy is the With CV and Hyperparameter Tuning model with training data accuracy of 79.9%, and testing data of 82.3%

Descriptive analysis



Graph analysis



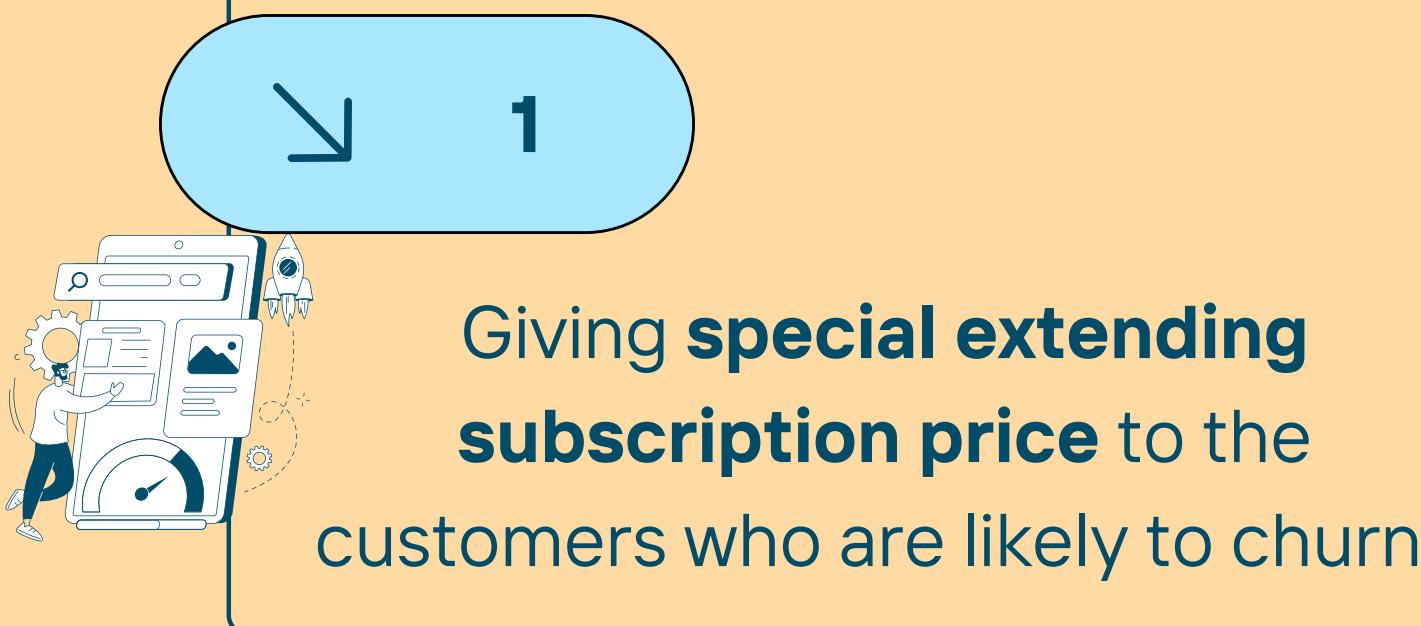
Logistic regression



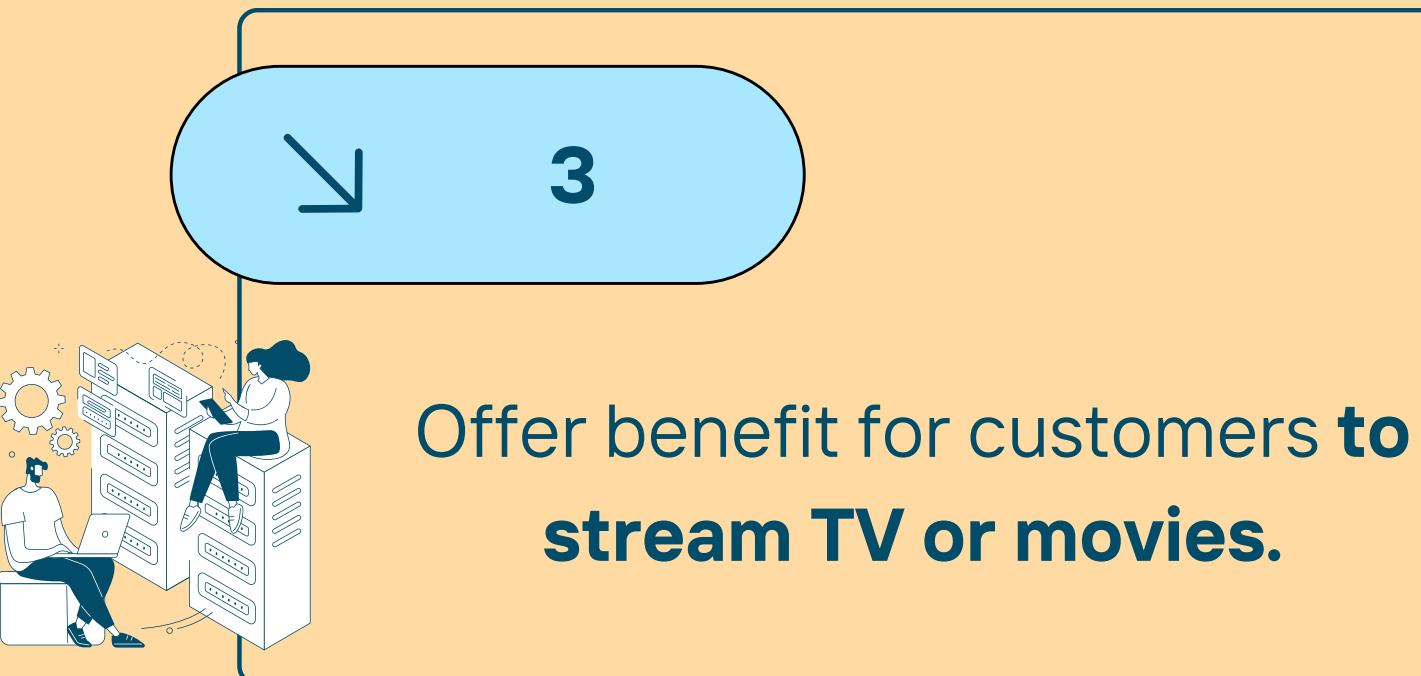
**CUSTOMER CHURN  
SEGMENTATION**

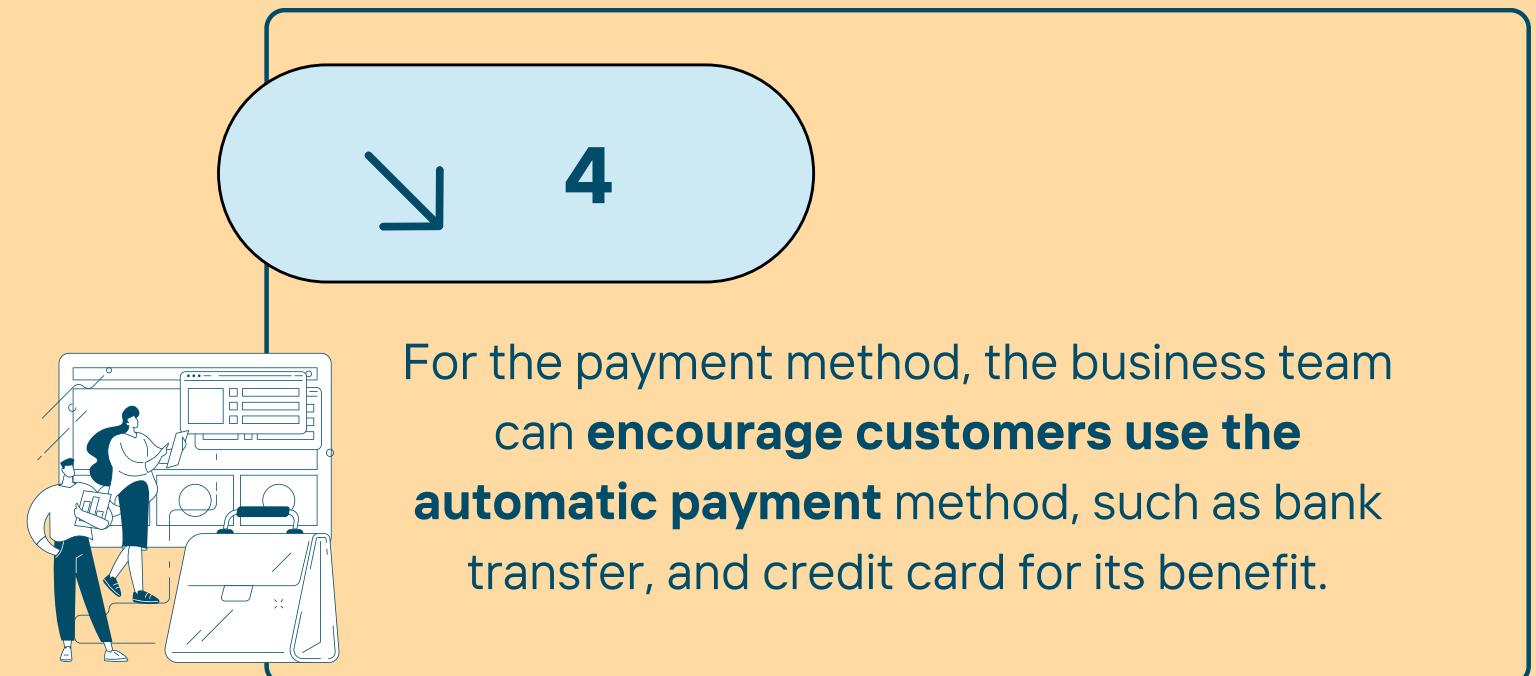
# Recommendation

# Recommendation

- 

1 Giving **special extending subscription price** to the customers who are likely to churn.
- 

2 Offer **free trial additional support** for the new customers who have short contract or tenure months
- 

3 Offer benefit for customers **to stream TV or movies.**
- 

4 For the payment method, the business team can **encourage customers use the automatic payment** method, such as bank transfer, and credit card for its benefit.

# THANK YOU!

