



# Customer Churn Prediction Dataset

จัดทำโดย  
**นายศุภณัฐ แซ่เตีย**  
ID: 6505000270



# Agenda

- ที่มาของชุดข้อมูล
- ลักษณะของชุดข้อมูล
- คุณสมบัติของแต่ละ Feature
- Data Preprocess
- Supervised Learning
  - K Nearest Neighbors Classifier
- Unsupervised Learning
  - K Means Clustering

# ที่มาของชุดข้อมูล



# ที่มาของชุดข้อมูล

- Customer Churn Prediction เป็นชุดข้อมูลที่ถูก  
จำลองขึ้นจากลักษณะและรูปแบบข้อมูลของลูกค้า  
ทั่วไปที่ใช้บริการของบริษัทด้านโทรคมนาคม
- จัดทำขึ้นโดยคุณ Rashad Mammadov ในปี ค.ศ.  
2024
- เพื่อนำไปใช้ในการเทรนโมเดลในการประเมินว่า  
ลูกค้าคนใดมีแนวโน้มที่จะเลิกใช้บริการ เพื่อให้  
บริษัทสามารถส่งพนักงานเข้าไปดูแลเพื่อรักษา  
ไม่ยกเลิกการใช้บริการ



คุณ Rashad Mammadov

# តាកេណៈខំងុំដោយមូល



# ลักษณะของชุดข้อมูล

- Customer Churn Prediction Dataset ประกอบด้วยข้อมูลทั้งหมด 5,880 ตัวอย่าง
- โดยประกอบด้วย 21 ฟีเชอร์ ซึ่งแบ่งได้เป็น 4 กลุ่มได้ดังนี้
  - กลุ่ม Target Features ประกอบด้วย 1 ฟีเชอร์ ได้แก่ Churn
  - กลุ่ม Identifier Features ประกอบด้วย 1 ฟีเชอร์ ได้แก่ customerID
  - กลุ่ม Numerical Features ประกอบด้วย 3 ฟีเชอร์ ได้แก่
    - tenure
    - MonthlyCharges
    - TotalCharges

# ลักษณะของชุดข้อมูล

- กลุ่ม Categorical Features ประกอบด้วย 16 ฟีเชอร์ ได้แก่
  - gender
  - SeniorCitizen
  - Partner
  - Dependents
  - PhoneService
  - MultipleLines
  - InternetService
  - OnlineSecurity
  - OnlineBackup
  - DeviceProtection
  - TechSupport
  - StreamingTV
  - StreamingMovies
  - Contract
  - PaperlessBilling
  - PaymentMethod

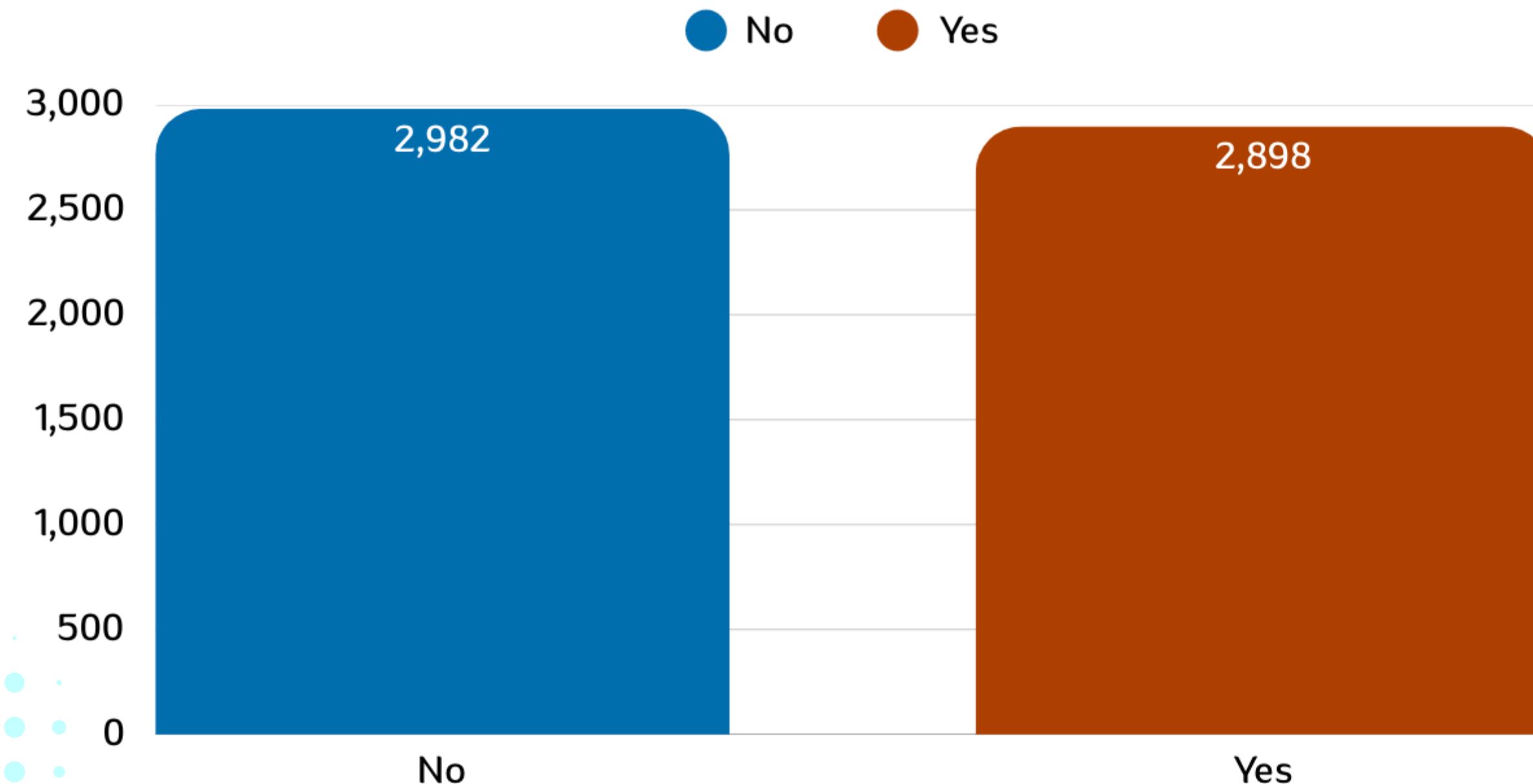
# คุณสมบัติของ Feature



# คุณสมบัติของ Feature ในกลุ่ม Target และ Identifier

# Churn

- เป็น Feature Target ประเภท Categorical ในกลุ่ม Binary
- กำหนดที่ระบุว่าลูกค้าตัดสินใจเลิกใช้บริการหรือไม่



# Customer ID

- เป็น Identifier Features ประเภท Categorical ชนิดมีลำดับ (Ordinal)
- กำหนดให้ชี้ว่าเป็นข้อมูลของลูกค้าคนใดในชุดข้อมูล
- ข้อมูลจะเป็นขึ้นต้นด้วยตัวอักษร “CUST” และตามด้วยตัวเลข 4 หลัก โดย customerID ในชุดข้อมูลจะเริ่มต้นที่ CUST0000 และไปจบที่ CUST5879
- โดย customerID จะเป็น Feature ที่ไม่มีผลต่อการคำนวณของโมเดล (Irrelevant Features)

# คุณสมบัติของ Feature ในกลุ่ม Numerical Features

# Tenure

- เป็นปีเชอร์ປະເກທ Numerical ຂັດ Ratio
- ກໍານຳທີ່ຮະບຸຈຳນວນເດືອນທີ່ລູກຄ້າໃຫ້ບົດກັບບົດບັນດາ

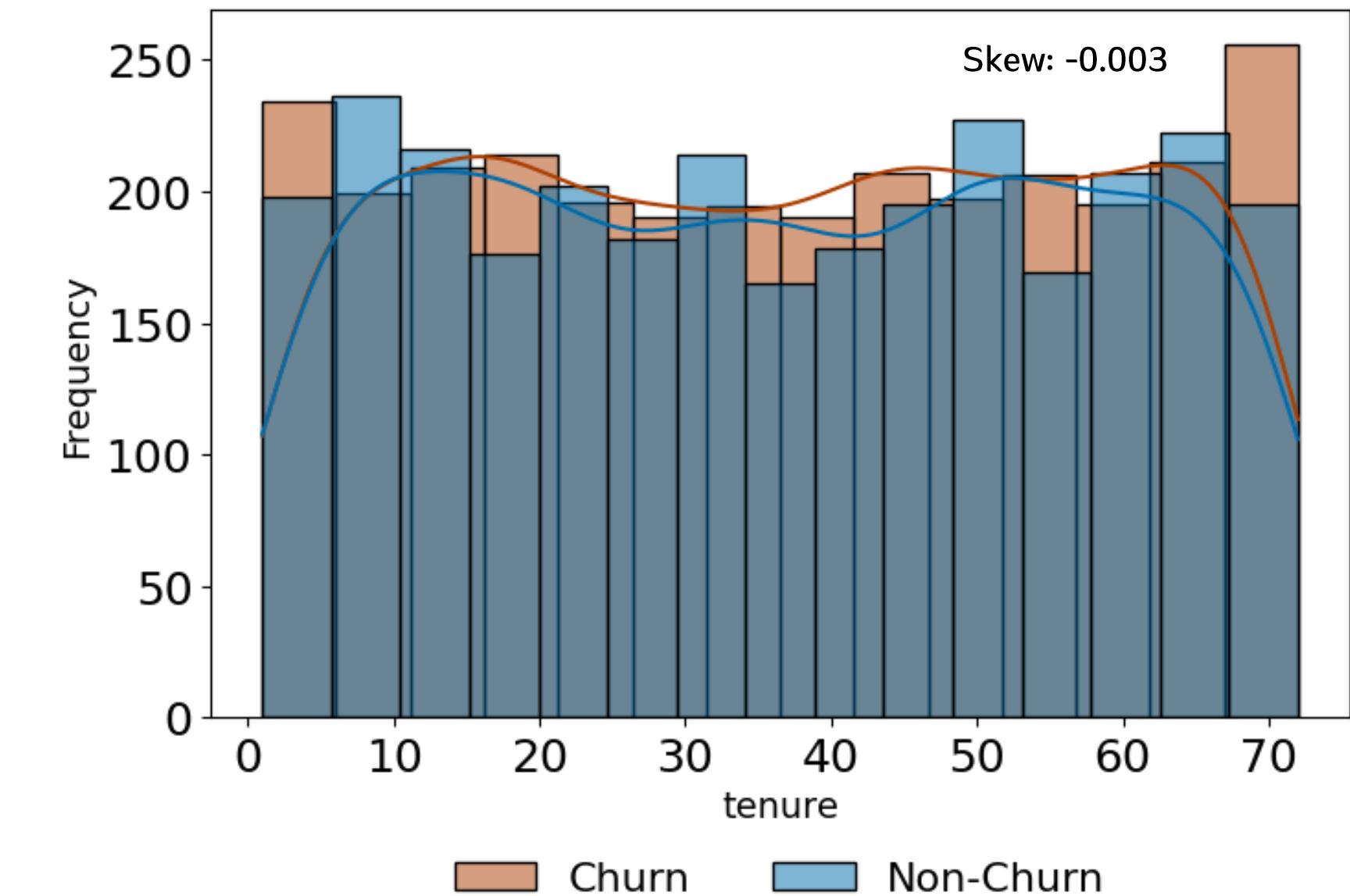
<b>Mean</b>	36.55
<b>Median</b>	37.00
<b>SD</b>	20.91
<b>Min</b>	1

**Q<sub>1</sub>** 18

**Q<sub>2</sub>** 37

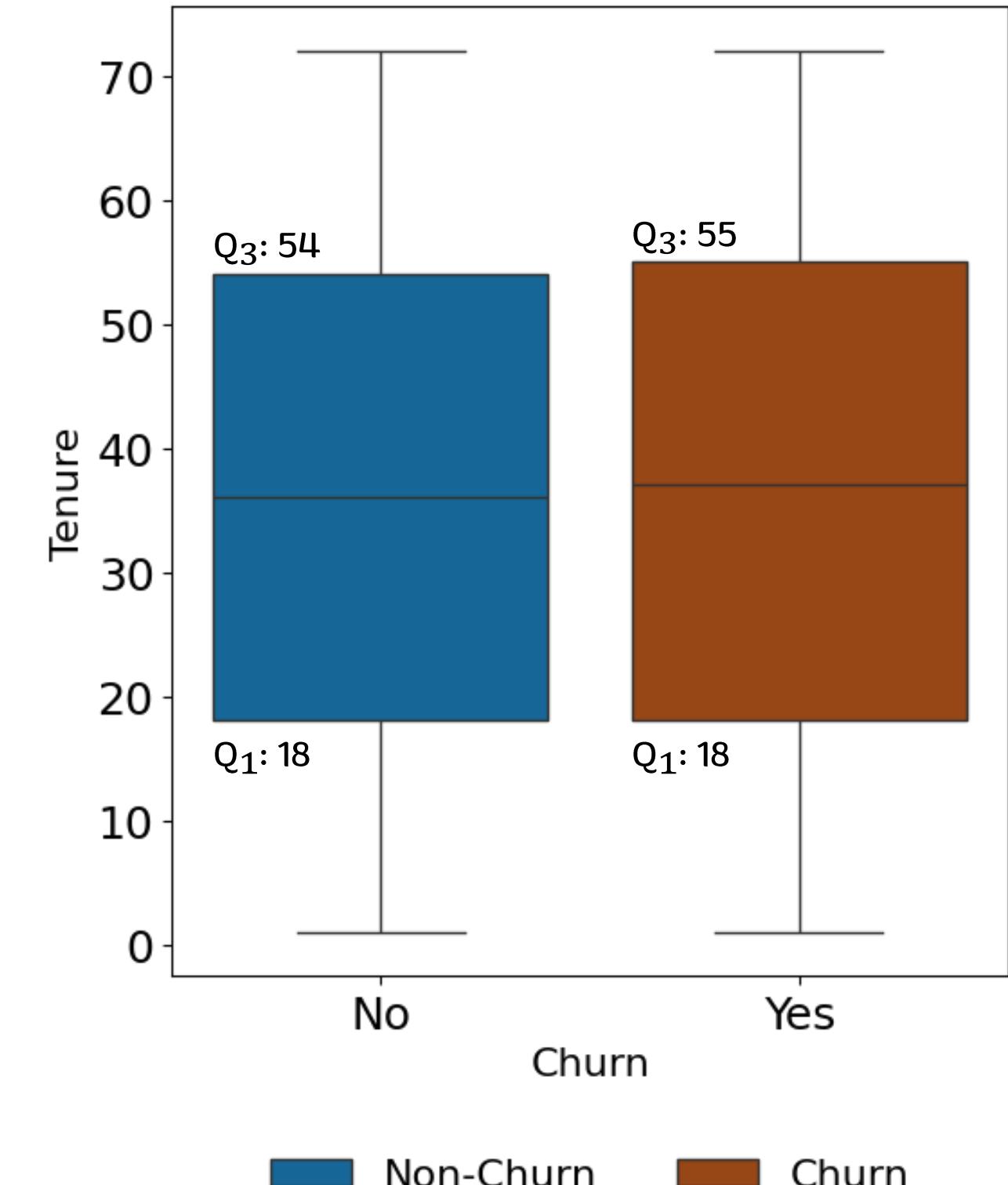
**Q<sub>3</sub>** 55

**Max** 72



# Tenure

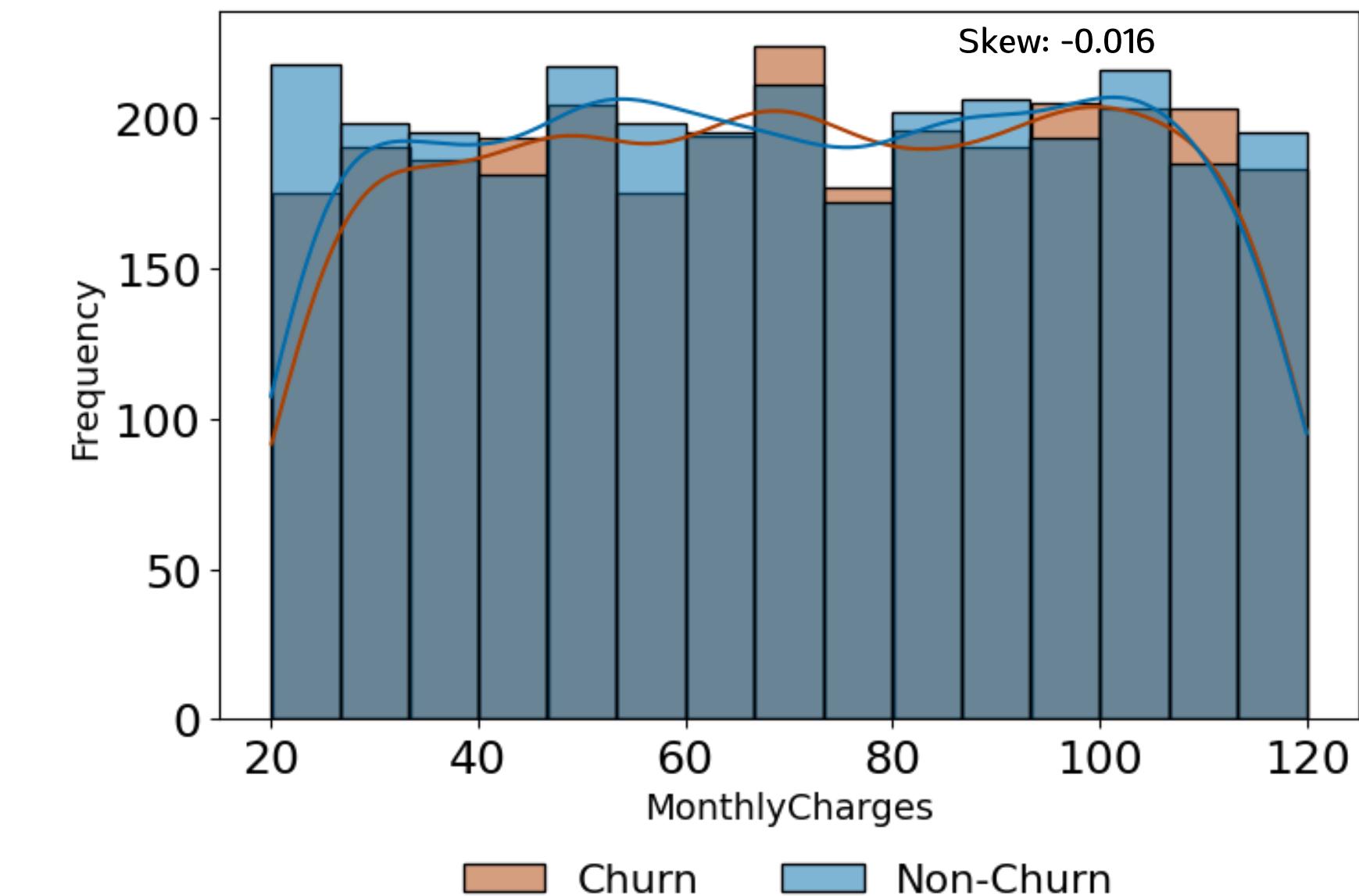
- จำนวนเดือนที่ลูกค้าใช้บริการกับบริษัทในช่วง
  - 18 - 54 เดือน จะเป็นช่วงที่ส่วนใหญ่ลูกค้าจะไม่ยกเลิกใช้บริการ
  - 18 - 55 เดือน จะเป็นช่วงที่ส่วนใหญ่ลูกค้าจะยกเลิกใช้บริการ



# Monthly Charges

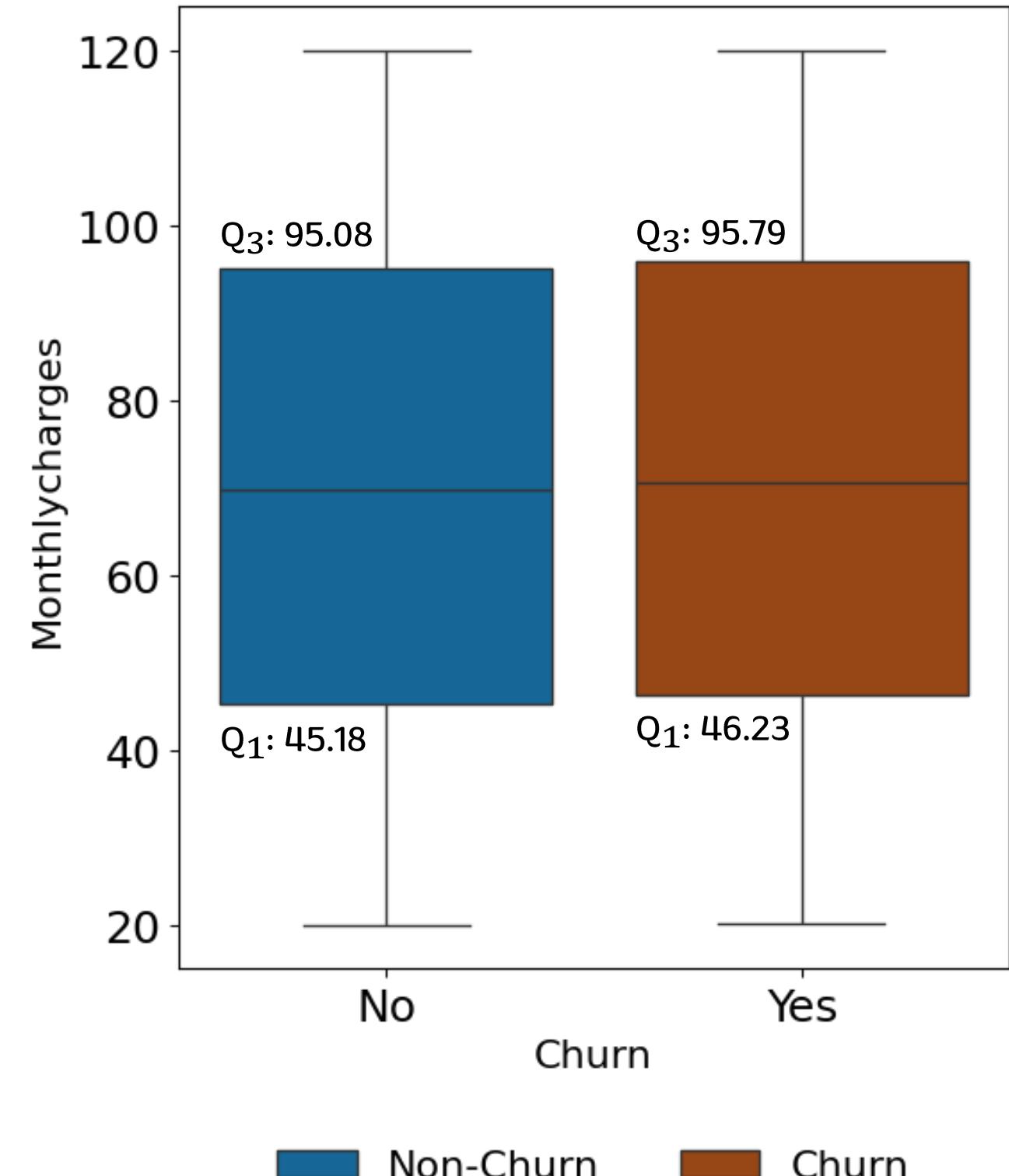
- เป็นฟีเชอร์ประเภท Numerical ชนิด Ratio
- กำหนดที่ระบุค่าบริการรายเดือนที่เรียกเก็บจากลูกค้า

<b>Mean</b>	70.16	<b>Q<sub>1</sub></b>	45.71
<b>Median</b>	70.16	<b>Q<sub>2</sub></b>	70.16
<b>SD</b>	28.81	<b>Q<sub>3</sub></b>	95.46
<b>Min</b>	20.00	<b>Max</b>	119.99



# Monthly Charges

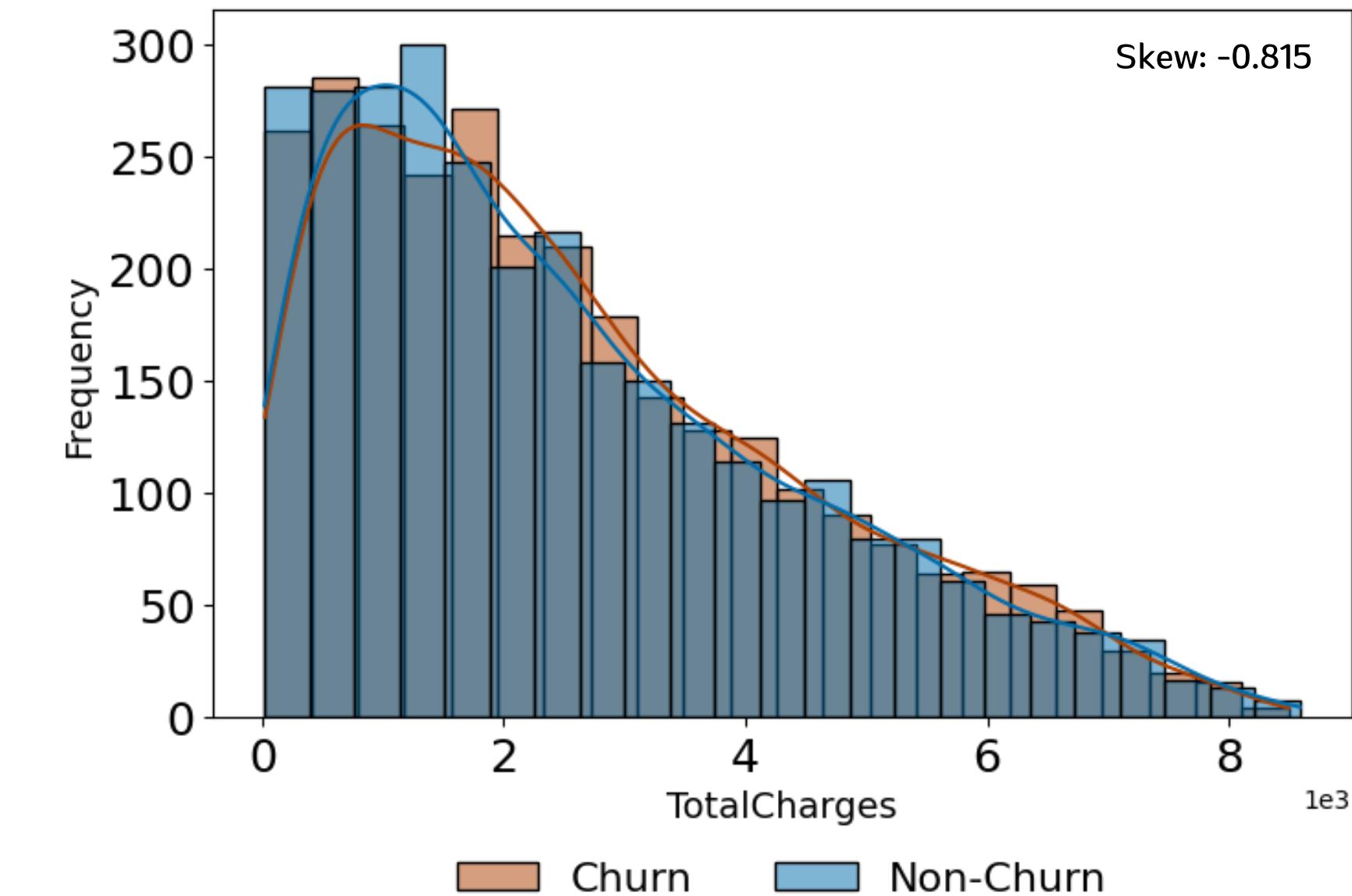
- ค่าบริการรายเดือนที่เรียกเก็บจากลูกค้าในช่วง
  - 45.18 - 95.08 จะเป็นช่วงที่ส่วนใหญ่ลูกค้าจะไม่ยกเลิกใช้บริการ
  - 46.23 - 95.79 จะเป็นช่วงที่ส่วนใหญ่ลูกค้าจะยกเลิกใช้บริการ



# Total Charges

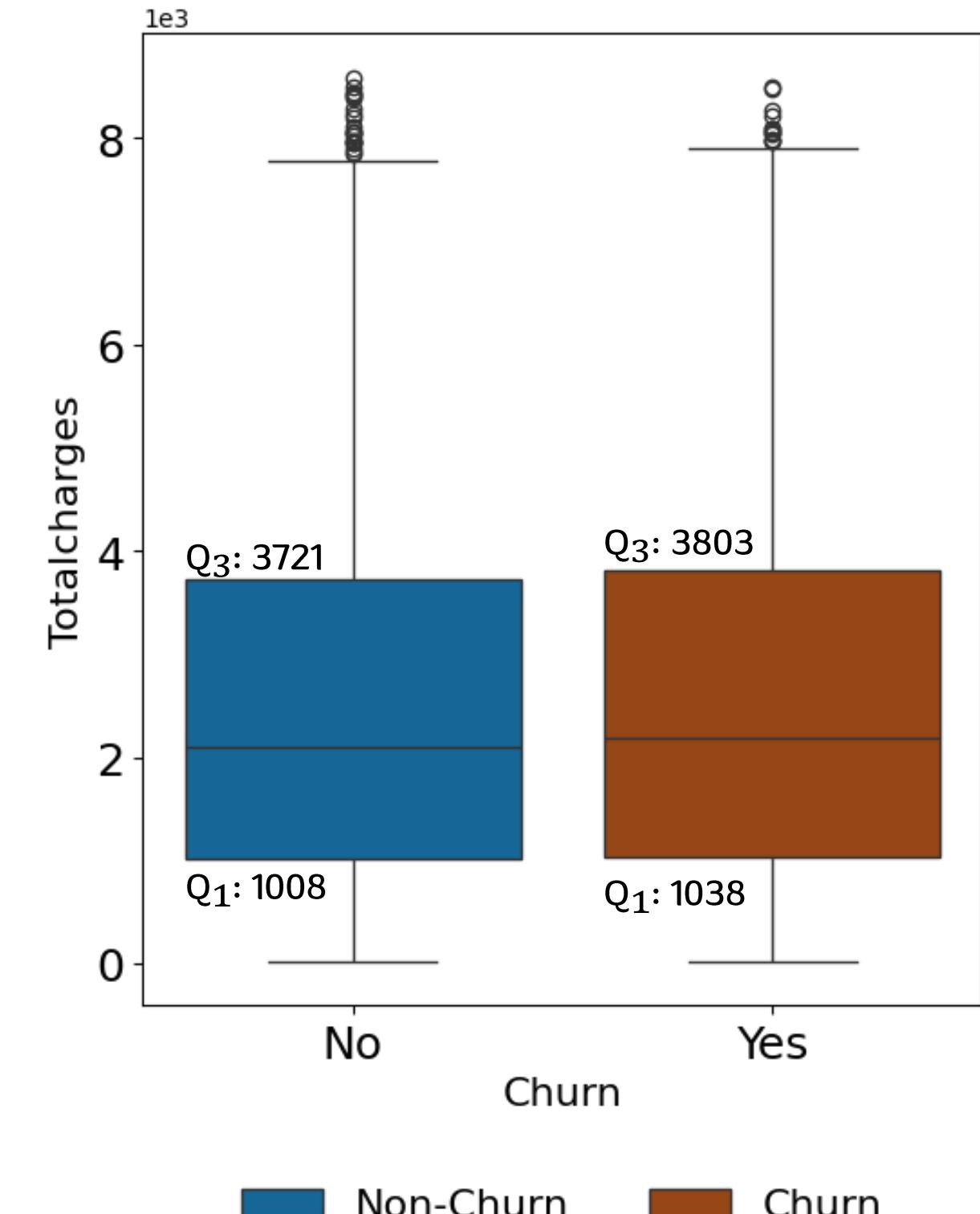
- เป็นฟีเชอร์ประเภท Numerical ชนิด Ratio
- กำหนดที่ระบุยอดรวมค่าบริการทั้งหมดที่ลูกค้าถูกเรียกเก็บ

<b>Mean</b>	2566.81	<b>Q<sub>1</sub></b>	1020.22
<b>Median</b>	2136.45	<b>Q<sub>2</sub></b>	2136.45
<b>SD</b>	1910.02	<b>Q<sub>3</sub></b>	3767.67
<b>Min</b>	20.03	<b>Max</b>	8589.60



# Total Charges

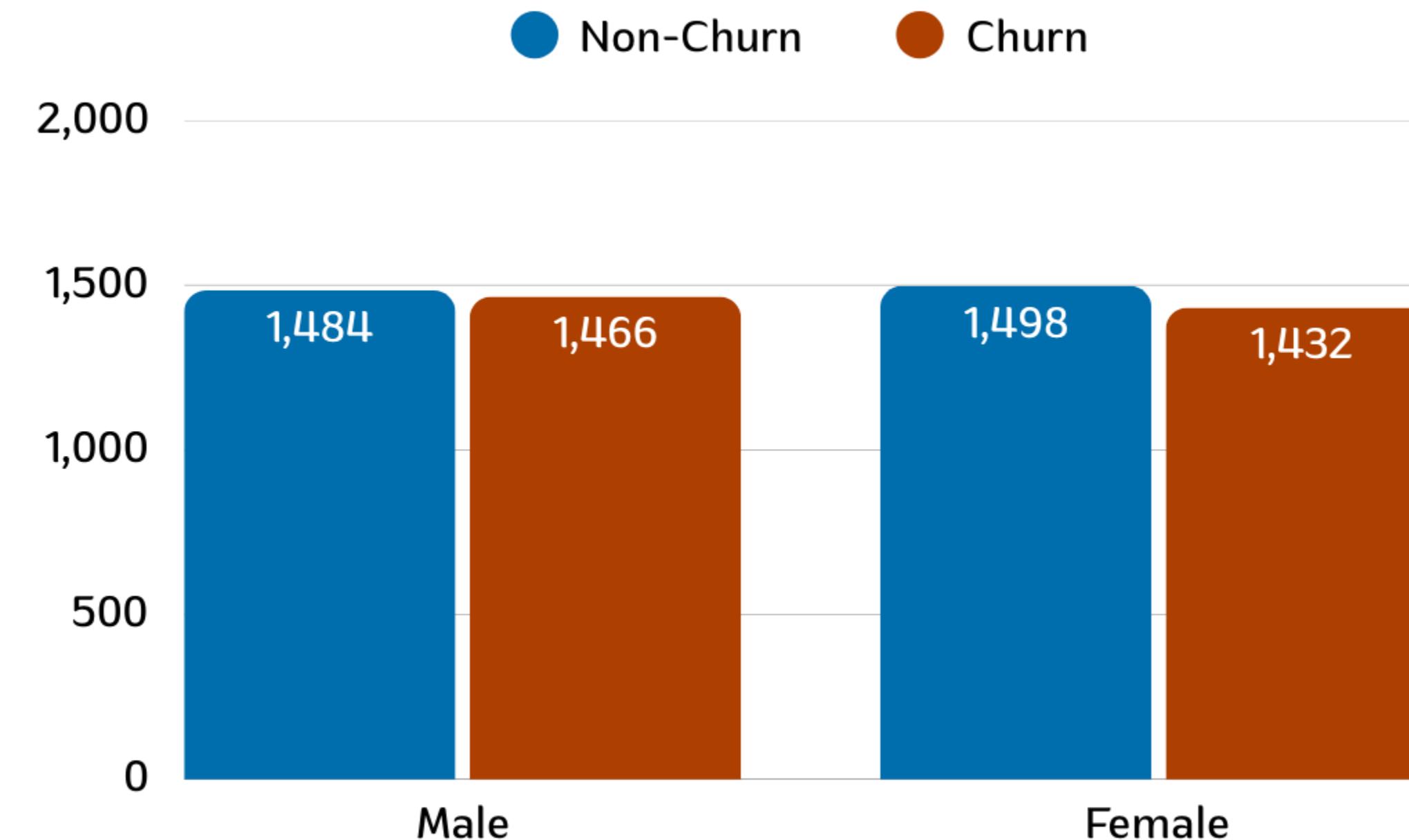
- ยอดรวมค่าบริการทั้งหมดของลูกค้าในช่วง
  - 1007.53 - 3721.34 จะเป็นช่วงที่ส่วนใหญ่ลูกค้าจะไม่ยกเลิกใช้บริการ
  - 1037.99 - 3803.15 จะเป็นช่วงที่ส่วนใหญ่ลูกค้าจะยกเลิกใช้บริการ



# คุณสมบัติของ Feature ในกลุ่ม Categorical Features

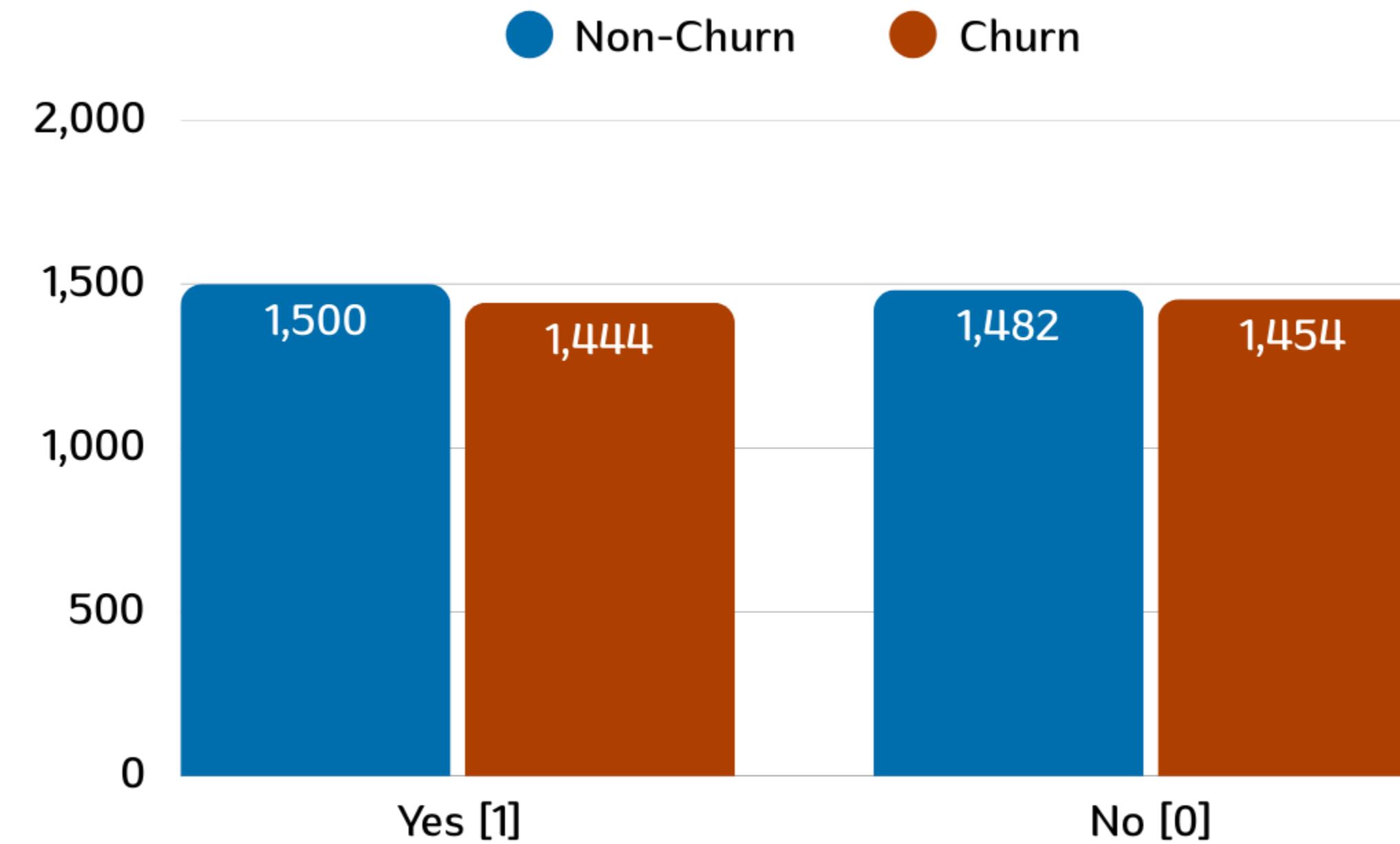
# Gender

- เป็นปีเซอร์ประเภท Categorical ชนิดไม่มีลำดับ (Nominal)
- กำหนดที่ระบุว่าเพศของลูกค้าเป็นเพศชายหรือหญิง



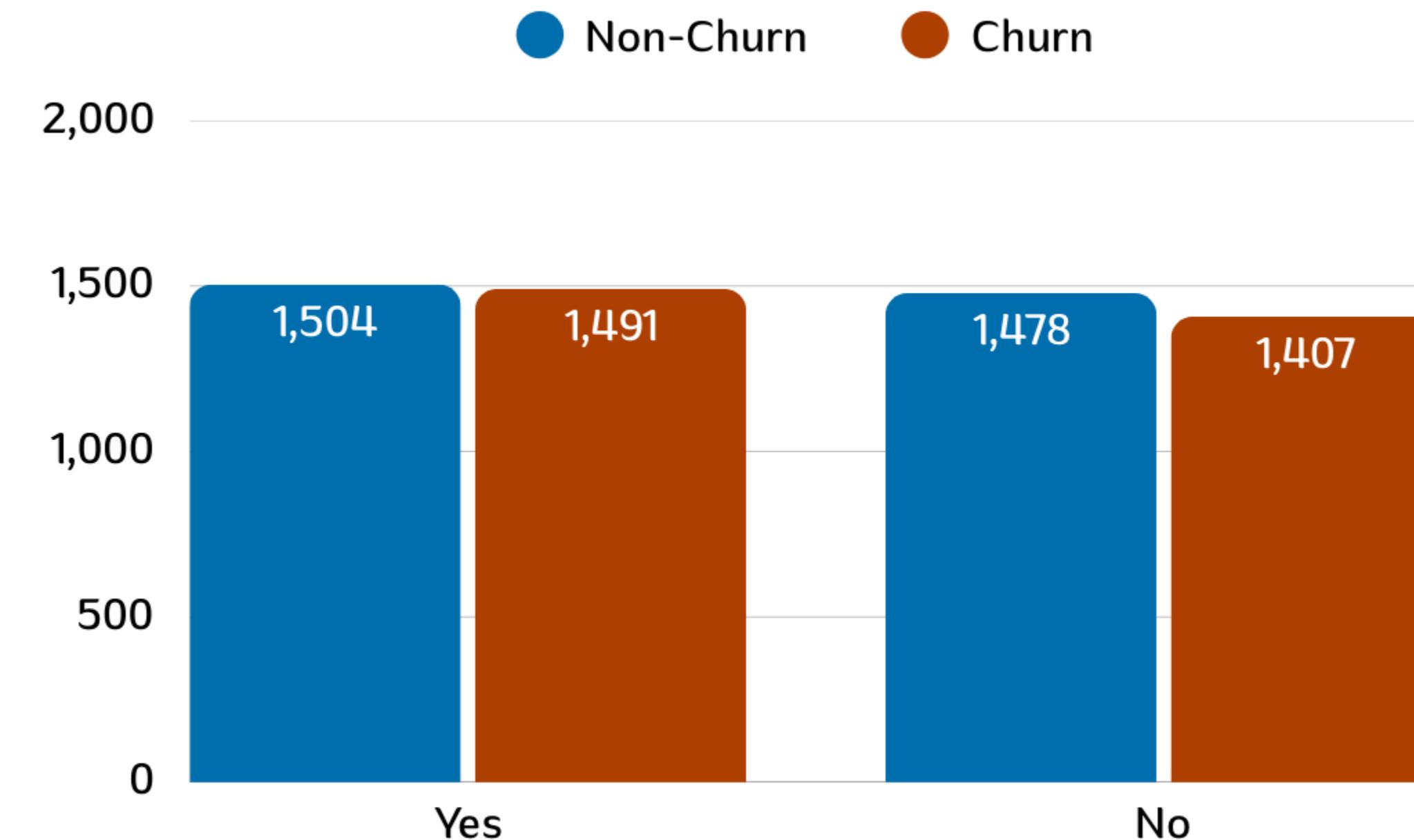
# Senior Citizen

- เป็นฟีเชอร์ประเภท Categorical ชนิดไม่มีลำดับ (Nominal)
- กำหนดที่ระบุว่าลูกค้าเป็นผู้สูงอายุหรือไม่ โดย 0 แทนไม่ใช่ และ 1 แทนใช่



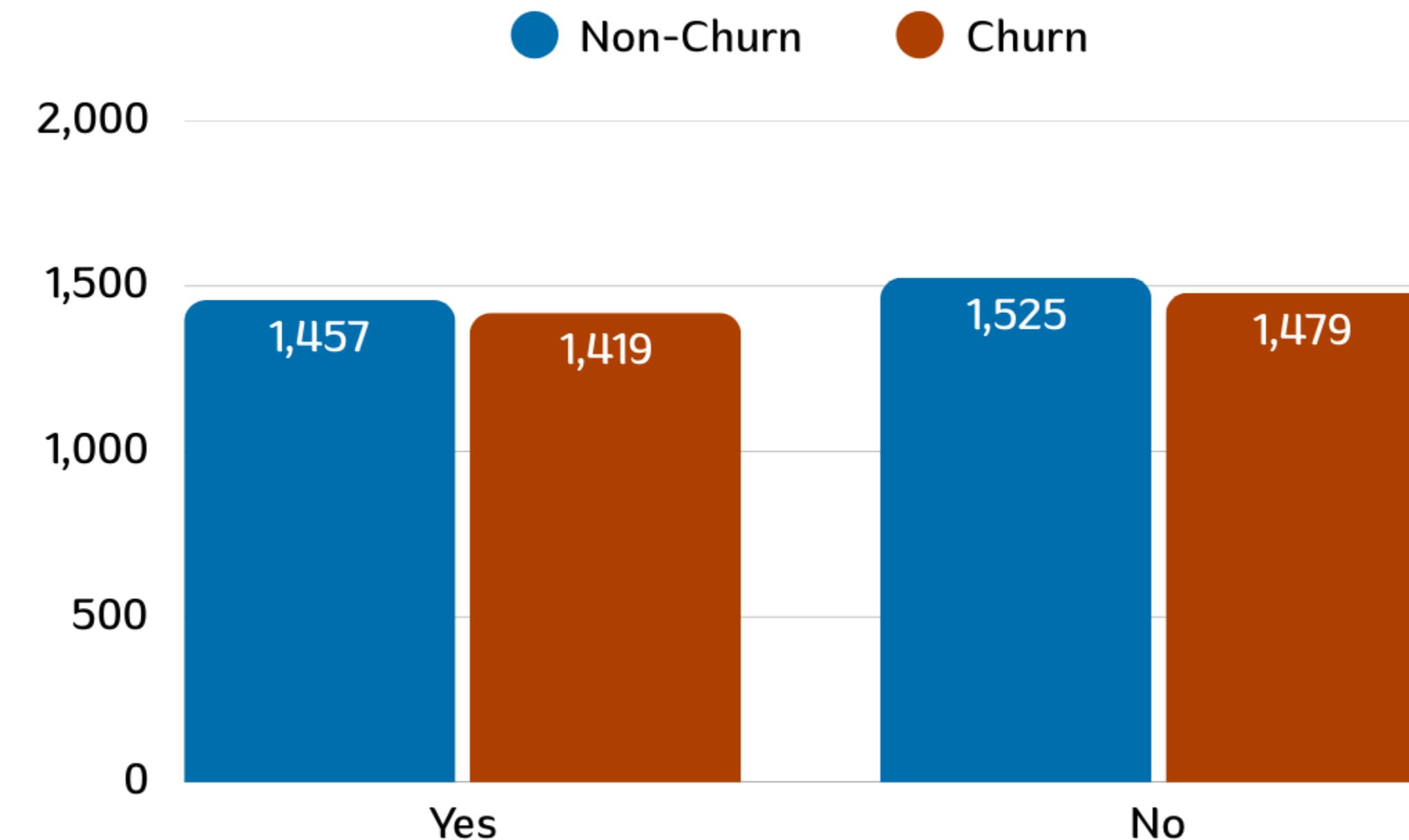
# Partner

- เป็นฟีเชอร์ประเภท Categorical ชนิดไม่มีลำดับ (Nominal)
- กำหนดที่ระบุว่าลูกค้ามีคู่สมรสใช่หรือไม่



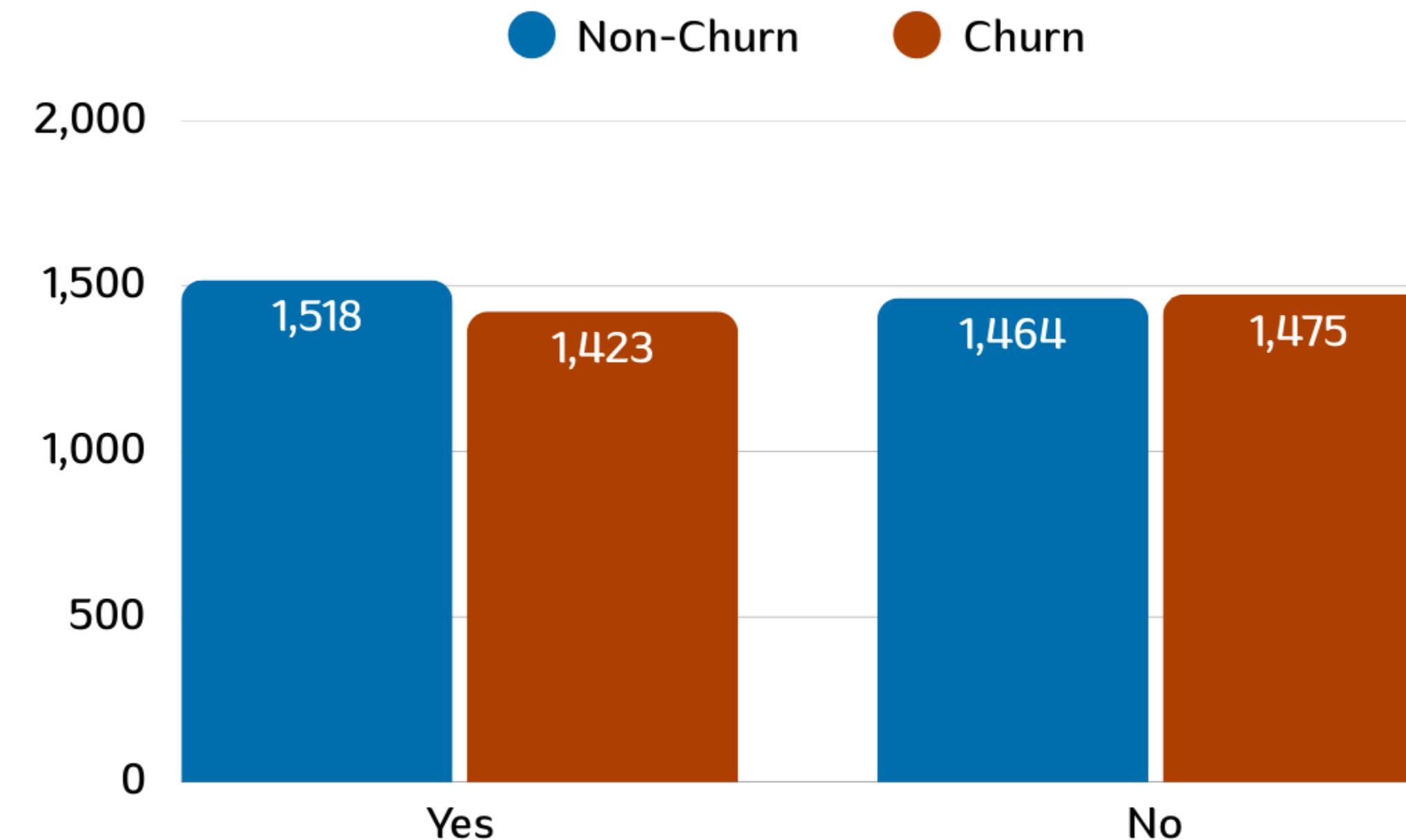
# Dependents

- เป็นฟีเชอร์ประเภท Categorical ชนิดไม่มีลำดับ (Nominal)
- กำหนดที่ระบุว่าลูกค้ามีผู้อยู่ในอุปการะชั่วคราวไม่



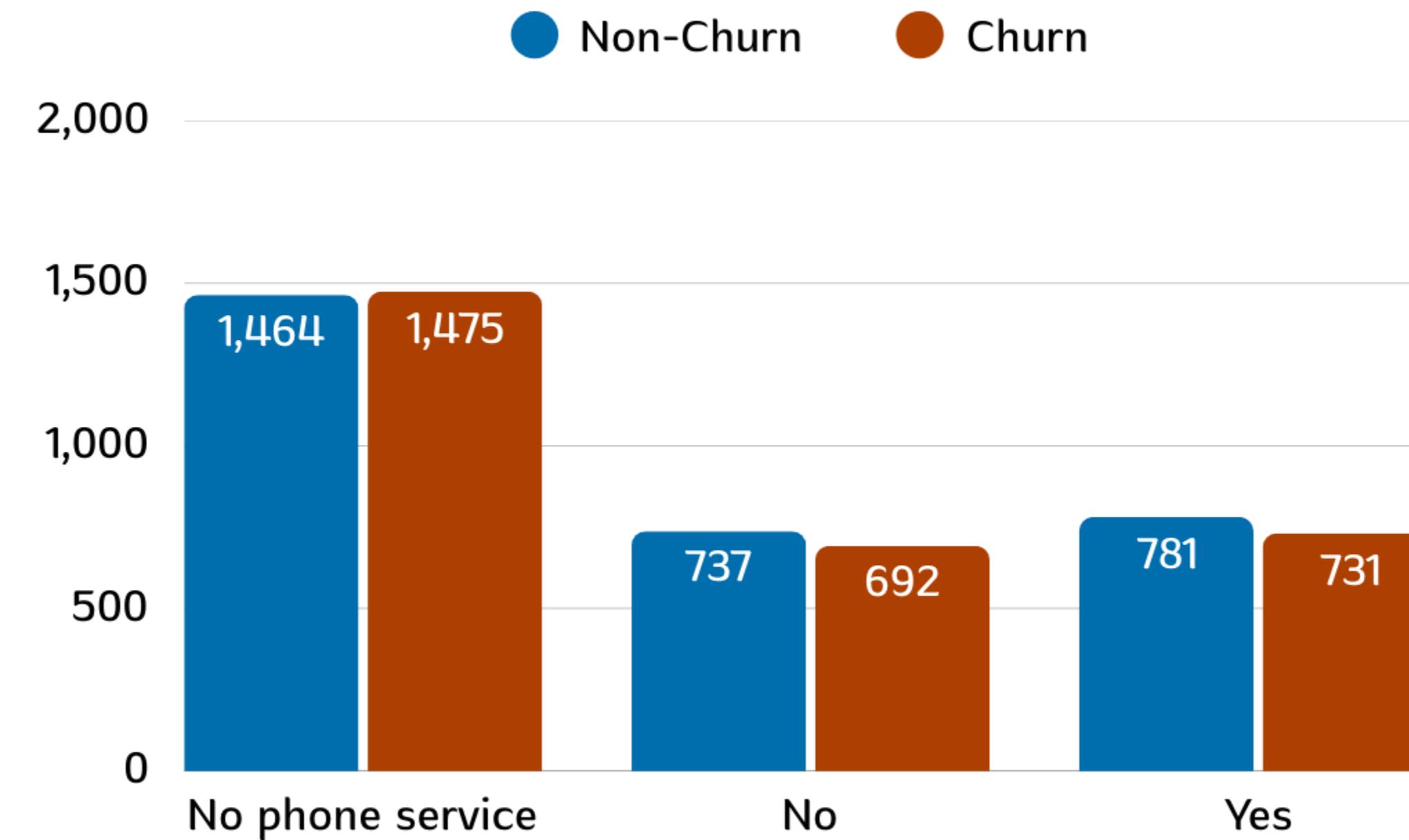
# Phone Service

- เป็นปีเซอร์ประเภท Categorical ชนิดไม่มีลำดับ (Nominal)
- กำหนดที่ระบุว่าลูกค้ามีการใช้บริการโทรศัพท์หรือไม่



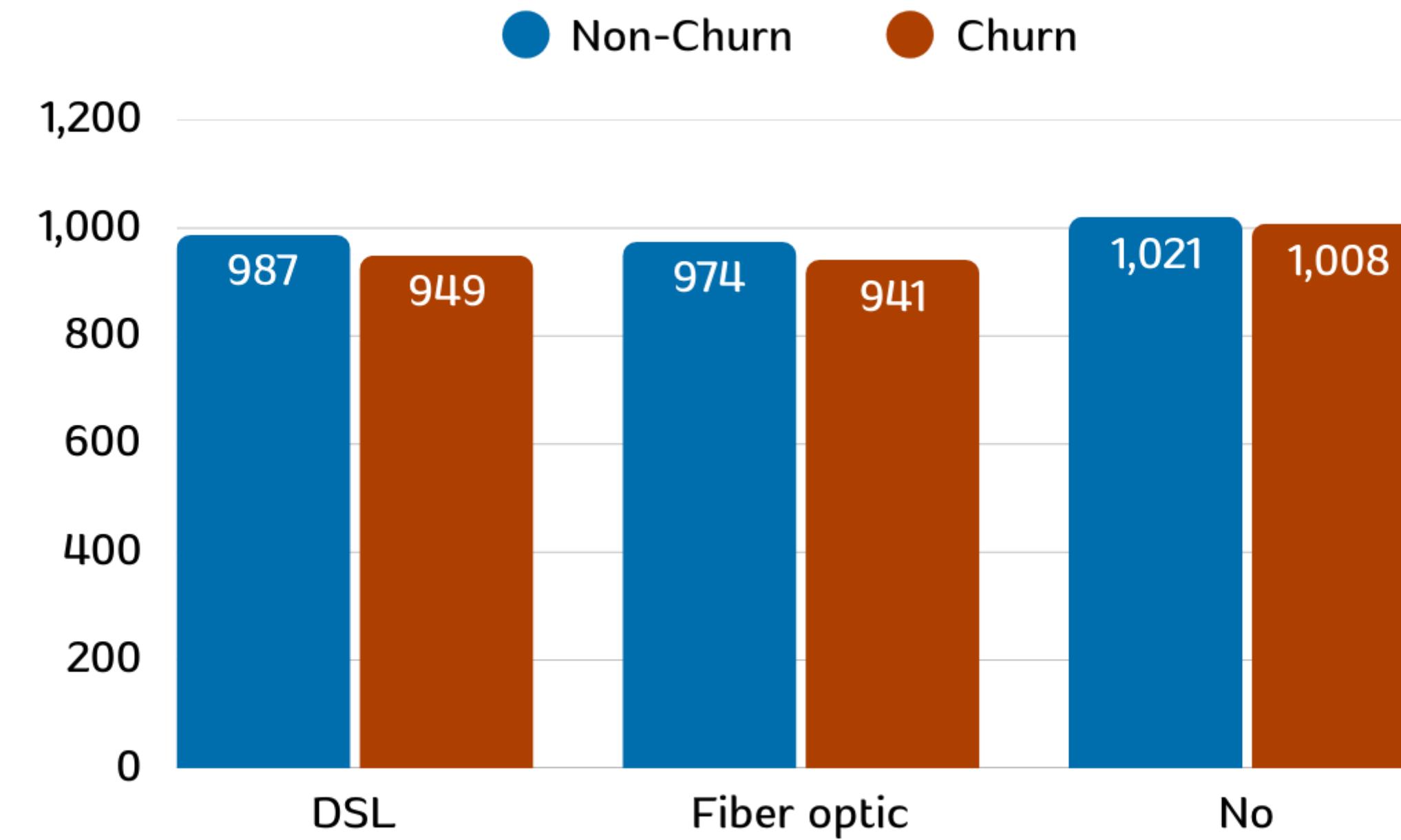
# Multiple Lines

- เป็นปีเซอร์ประเภท Categorical ชนิดไม่มีลำดับ (Nominal)
- กำหนดที่ระบุว่าลูกค้าใช้บริการเบอร์โทรศัพท์หลายเบอร์ใช่หรือไม่ หรือไม่ใช้บริการเลย



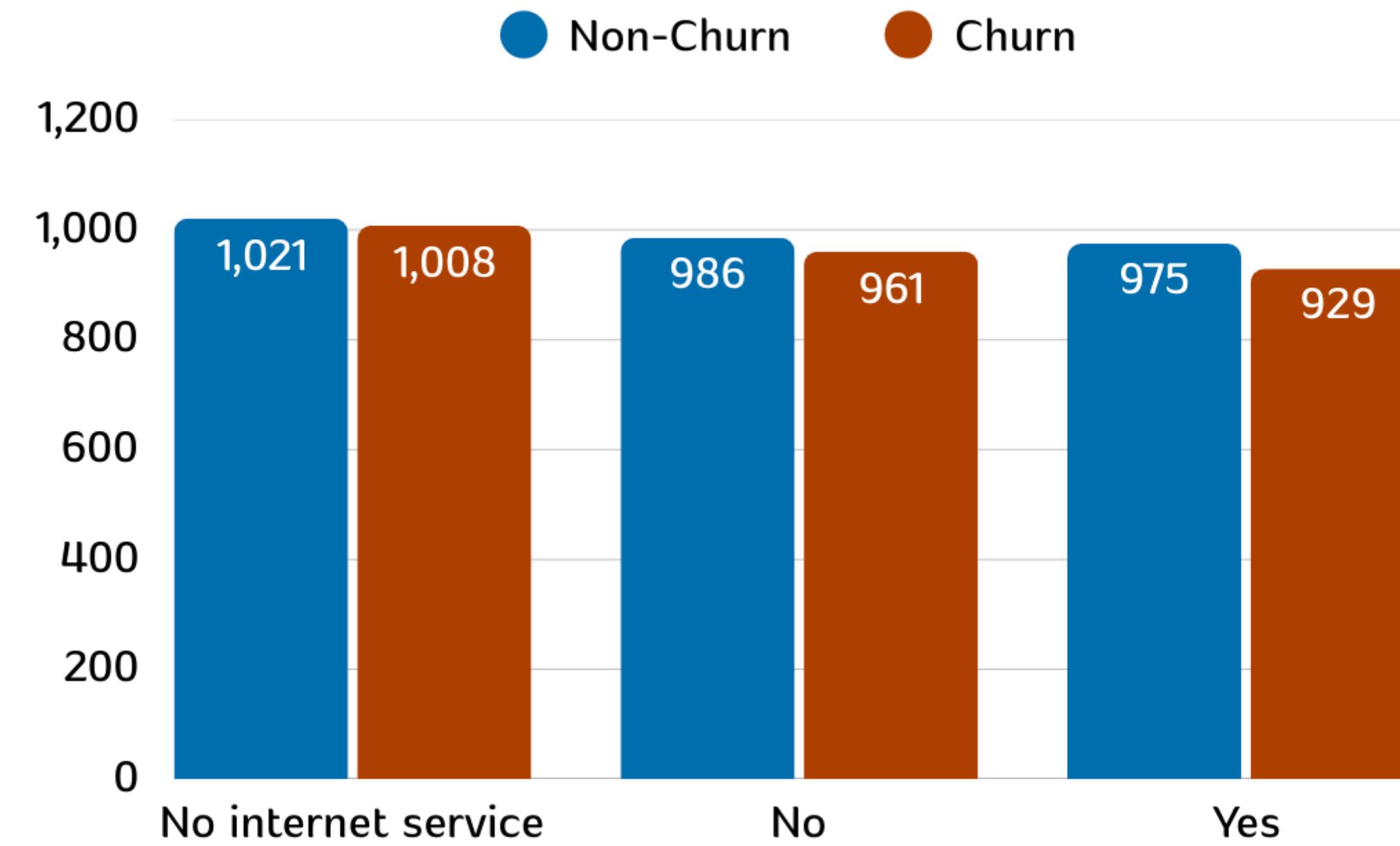
# Internet Service

- เป็นปีเซอร์ประเภท Categorical ชนิดไม่มีลำดับ (Nominal)
- กำหนดที่ระบุประเภทของบริการอินเทอร์เน็ตที่ลูกค้าใช้



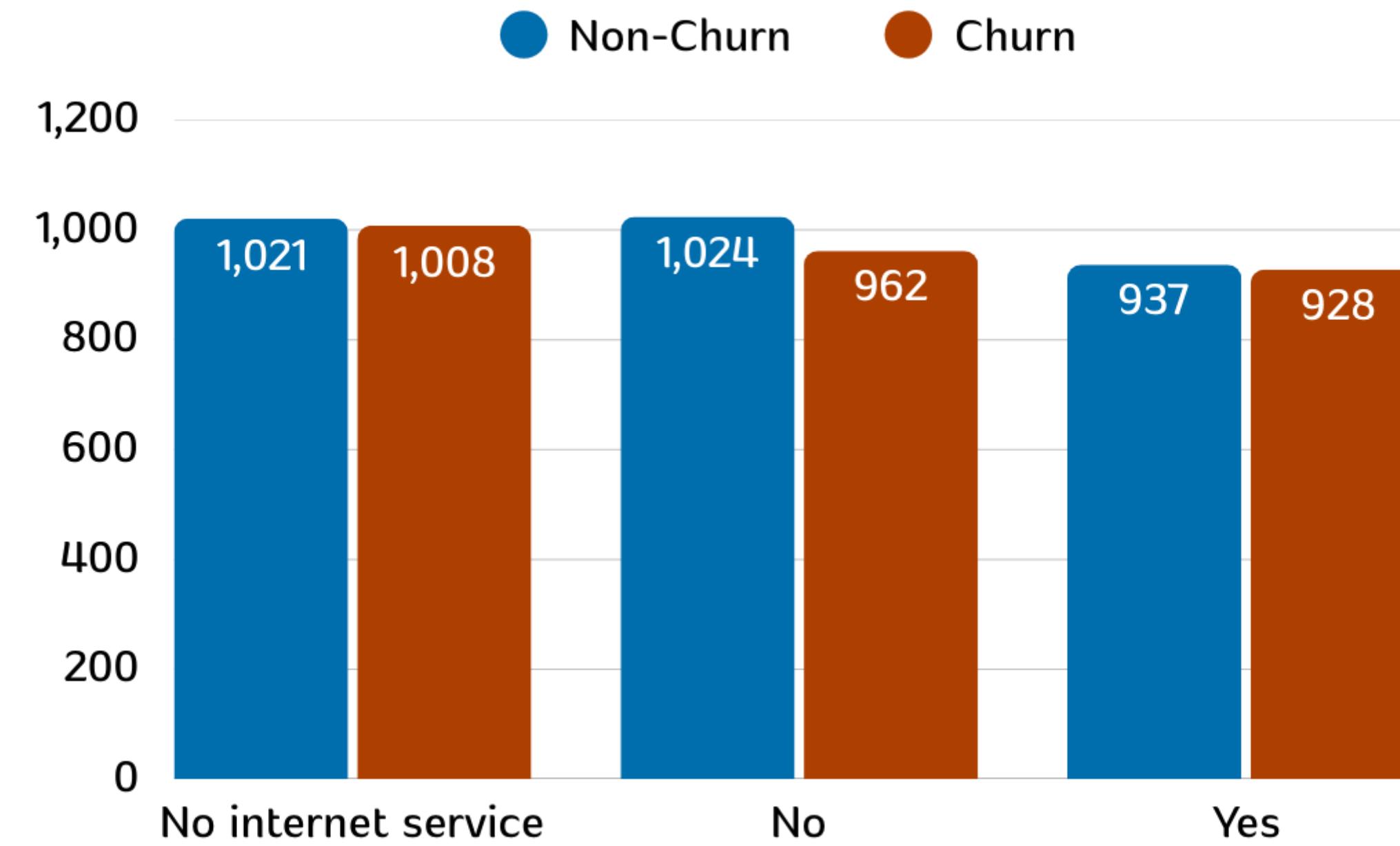
# Online Security

- เป็นปีเซอร์ประเภท Categorical ชนิดไม่มีลำดับ (Nominal)
- กำหนดที่ระบุสถานะการสมัครใช้บริการเสริมด้านความปลอดภัยออนไลน์ของลูกค้า



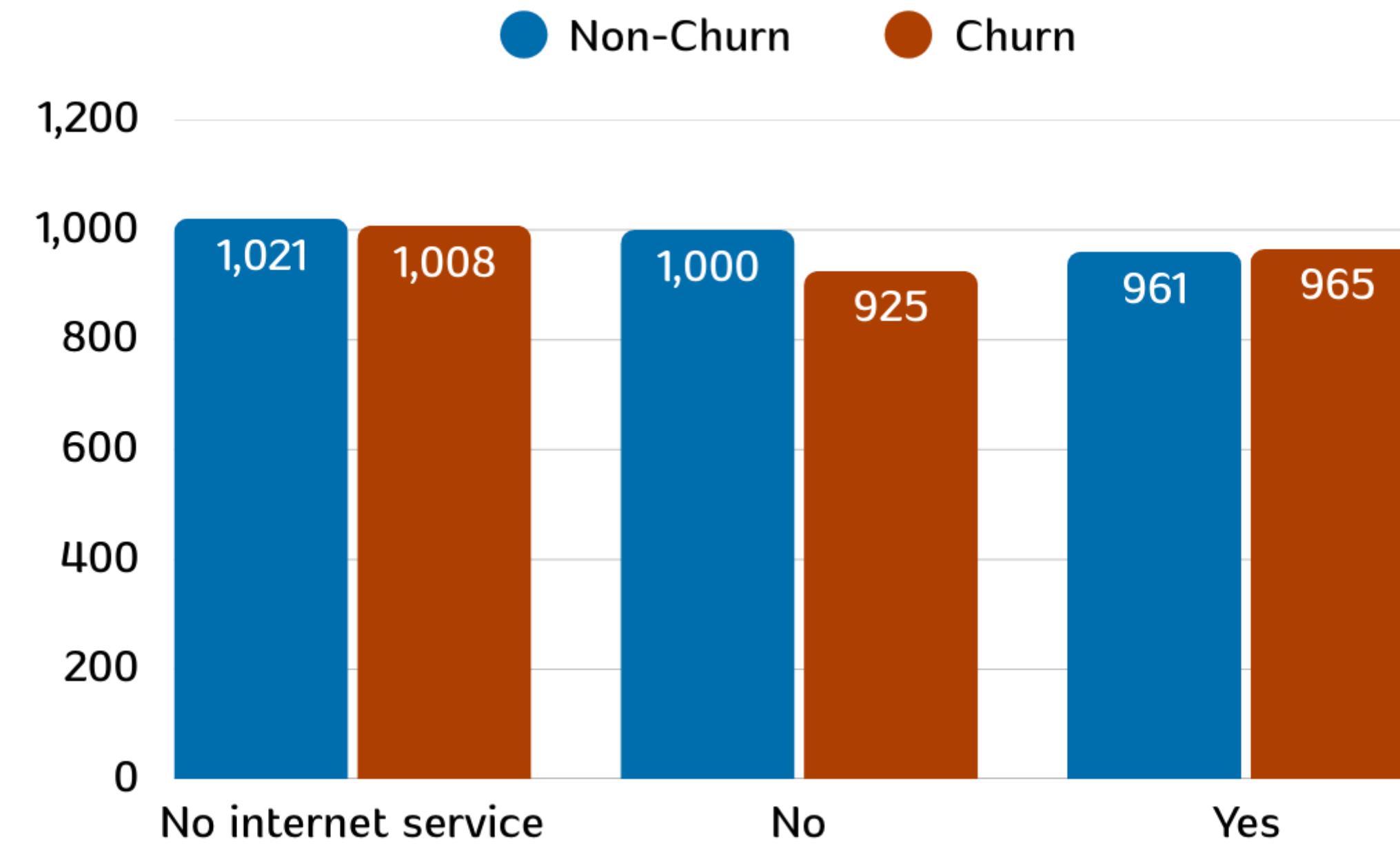
# Online Backup

- เป็นปีเซอร์ประเภท Categorical ชนิดไม่มีลำดับ (Nominal)
- กำหนดที่ระบุสถานะการสมัครใช้บริการสำรองข้อมูลออนไลน์ (Cloud) ของลูกค้า



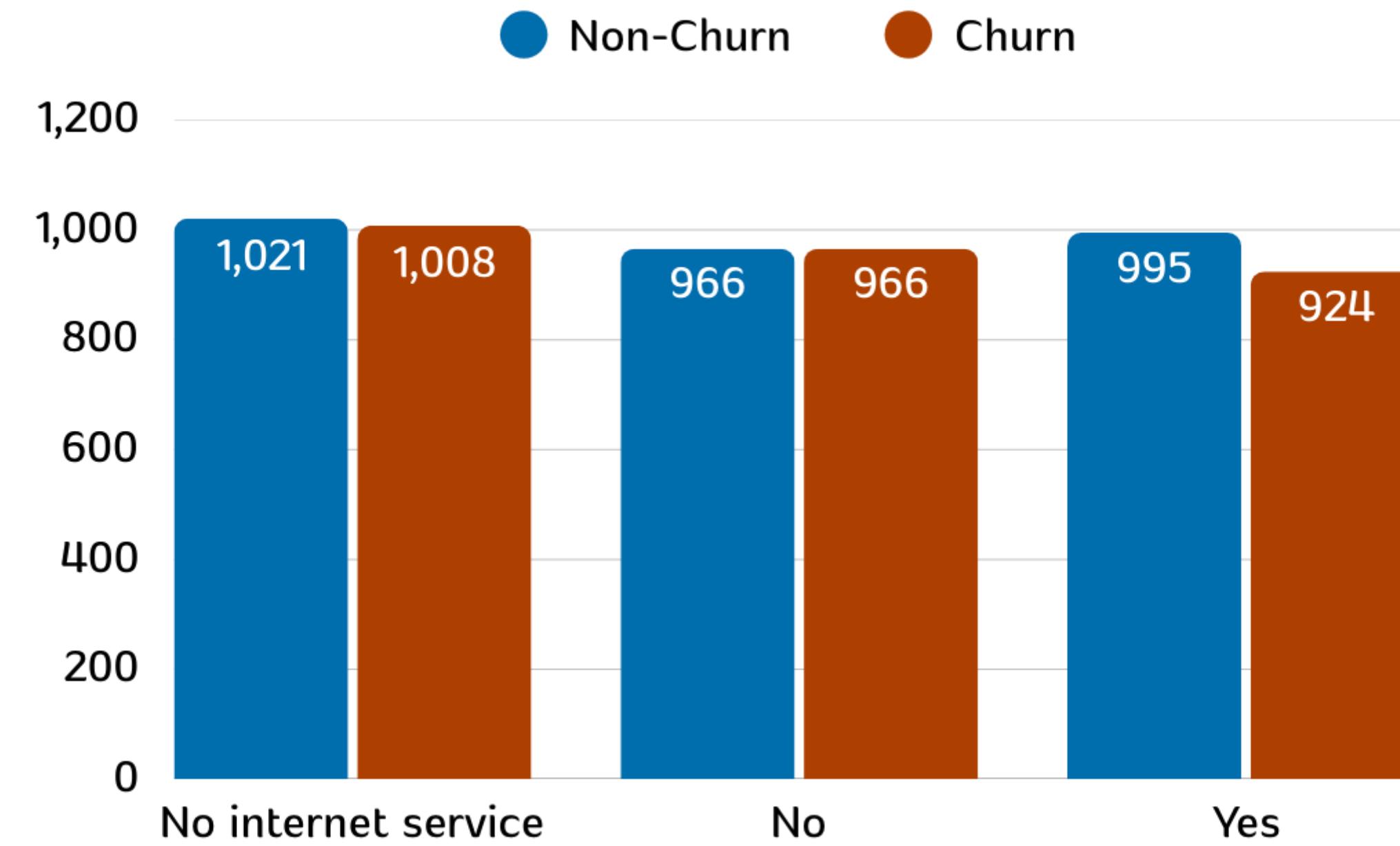
# Device Protection

- เป็นปีเซอร์ประเภท Categorical ชนิดไม่มีลำดับ (Nominal)
- กำหนดที่ระบุสถานะการสมัครใช้บริการป้องกันอุปกรณ์ของลูกค้า



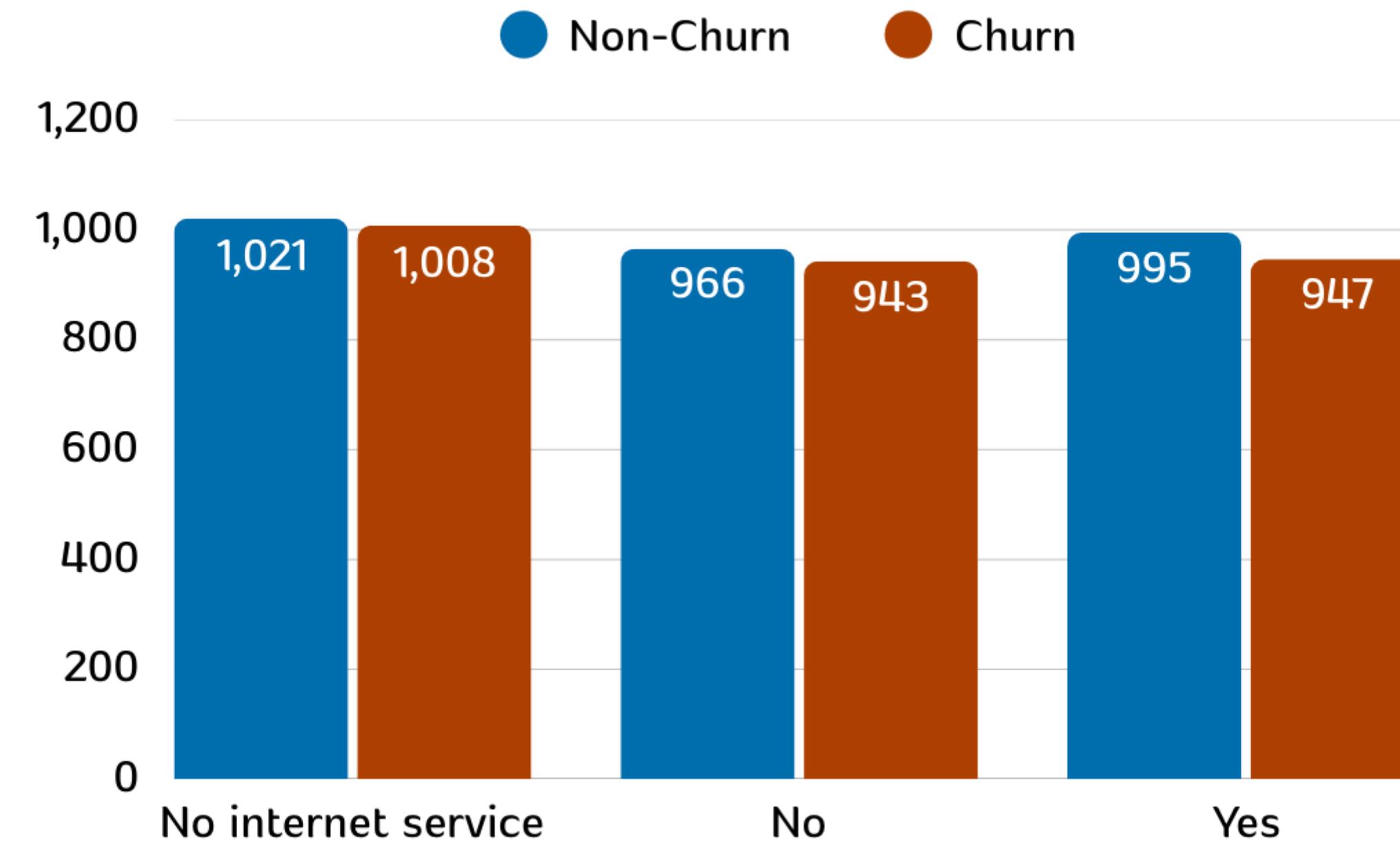
# Tech Support

- เป็นปีเซอร์ประเภท Categorical ชนิดไม่มีลำดับ (Nominal)
- กำหนดที่ระบุสถานะการสมัครใช้บริการสนับสนุนด้านเทคนิคของลูกค้า



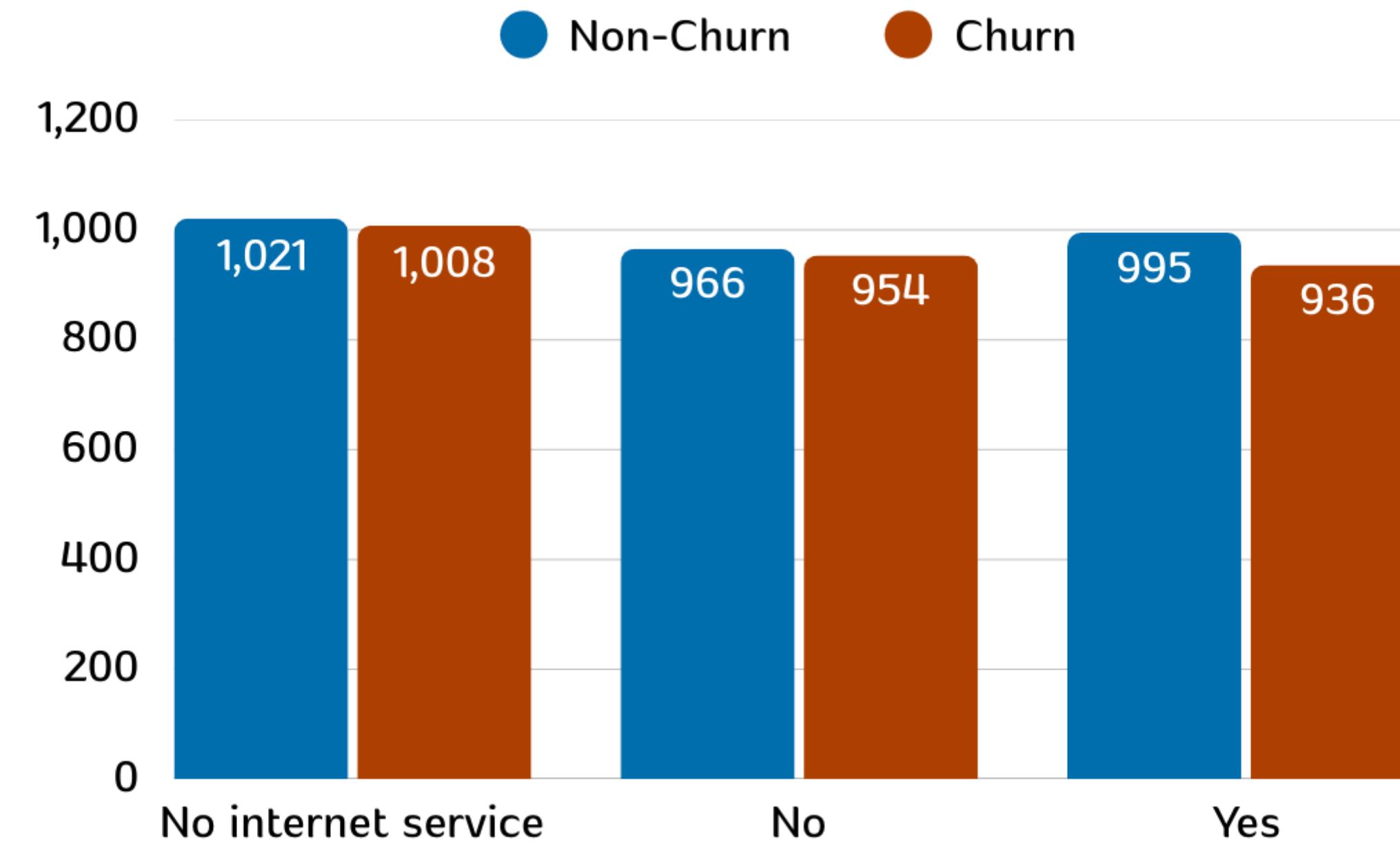
# Streaming TV

- เป็นฟีเชอร์ประเภท Categorical ชนิดไม่มีลำดับ (Nominal)
- กำหนดที่ระบุสถานะการสมัครใช้บริการสตรีมมิ่งผ่านทีวีของลูกค้า



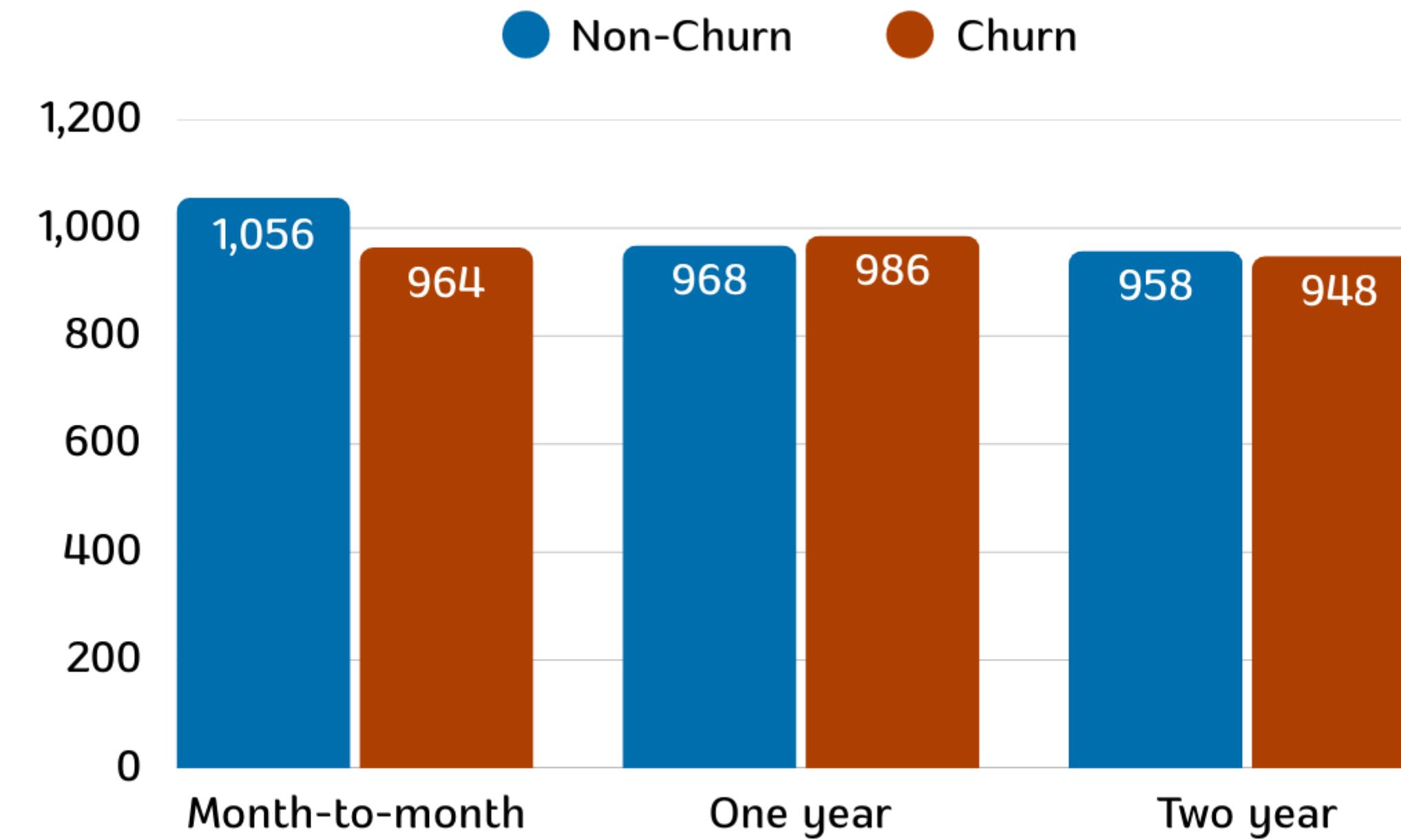
# Streaming Movies

- เป็นฟีเชอร์ประเภท Categorical ชนิดไม่มีลำดับ (Nominal)
- กำหนดที่ระบุสถานะการสมัครใช้บริการสตรีมมิ่งภาพยนตร์ของลูกค้า



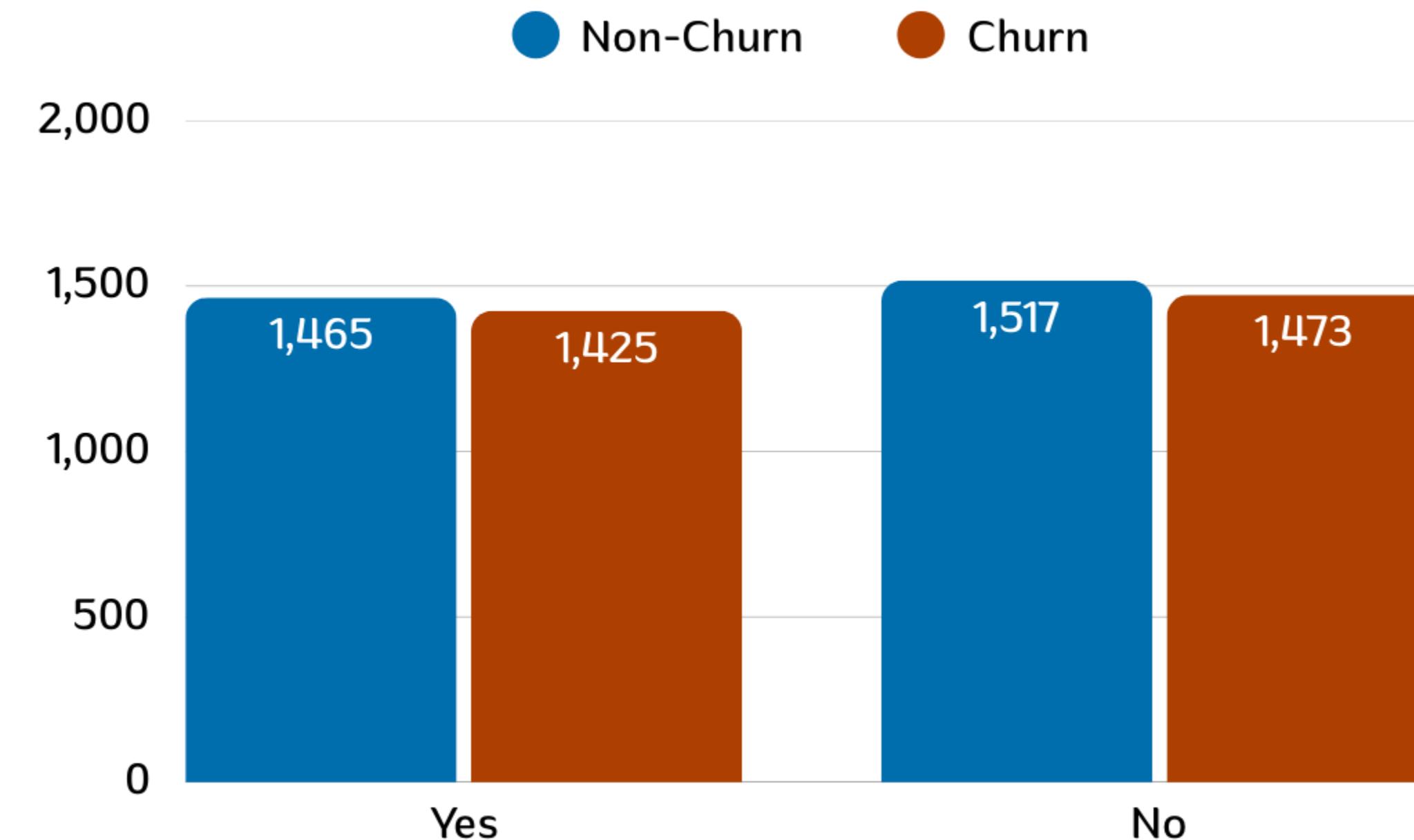
# Contract

- เป็นปีเซอร์ประเภท Categorical ชนิดไม่มีลำดับ (Nominal)
- กำหนดที่ระบุประเภทของลัญญาที่ลูกค้าใช้



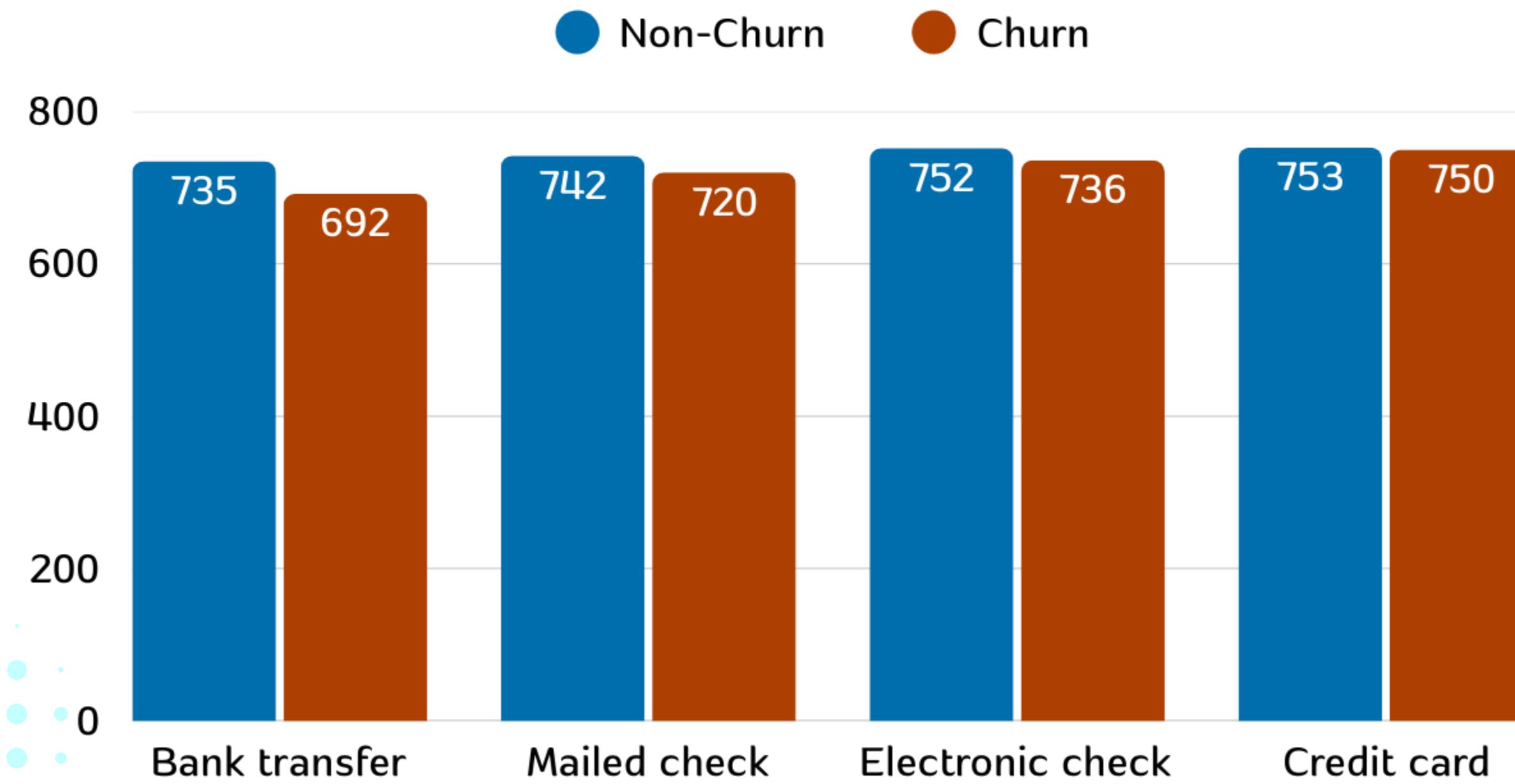
# Paperless Billing

- เป็นปีเซอร์ประเภท Categorical ชนิดไม่มีลำดับ (Nominal)
- กำหนดที่ระบุว่าลูกค้าใช้บริการรับบัลฟ์ผ่านระบบอิเล็กทรอนิกส์ใช่หรือไม่



# Payment Method

- เป็นปีเซอร์ประเภท Categorical ชนิดไม่มีลำดับ (Nominal)
- กำหนดที่ระบุวิธีการชำระเงินของลูกค้า

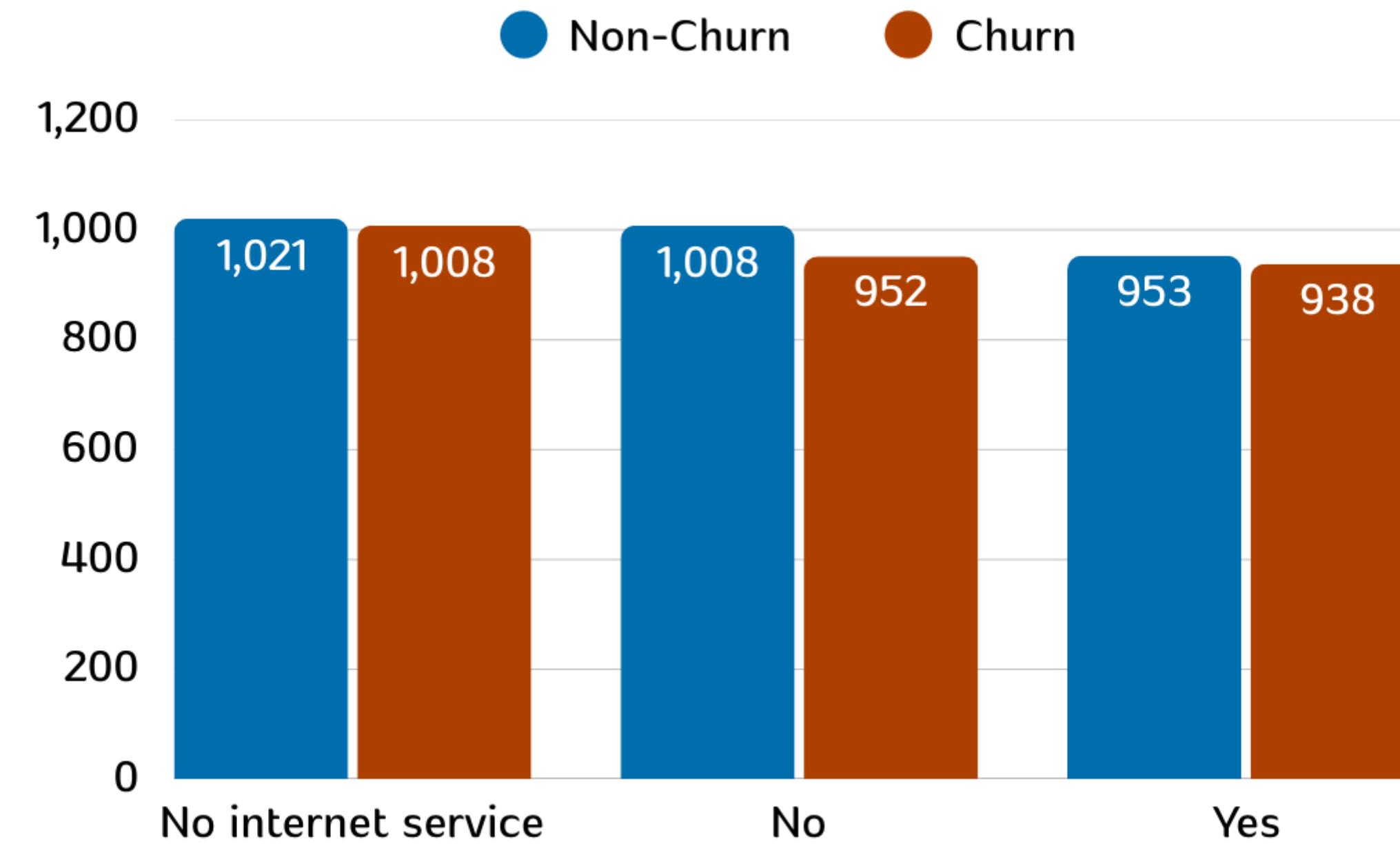


# Data Preprocess



# Combine Columns

- ทำการรวมฟีเจอร์ Online Security, Online Backup และ Device Protection เข้าด้วยกันเป็นฟีเจอร์ Security Service



# Drop Feature

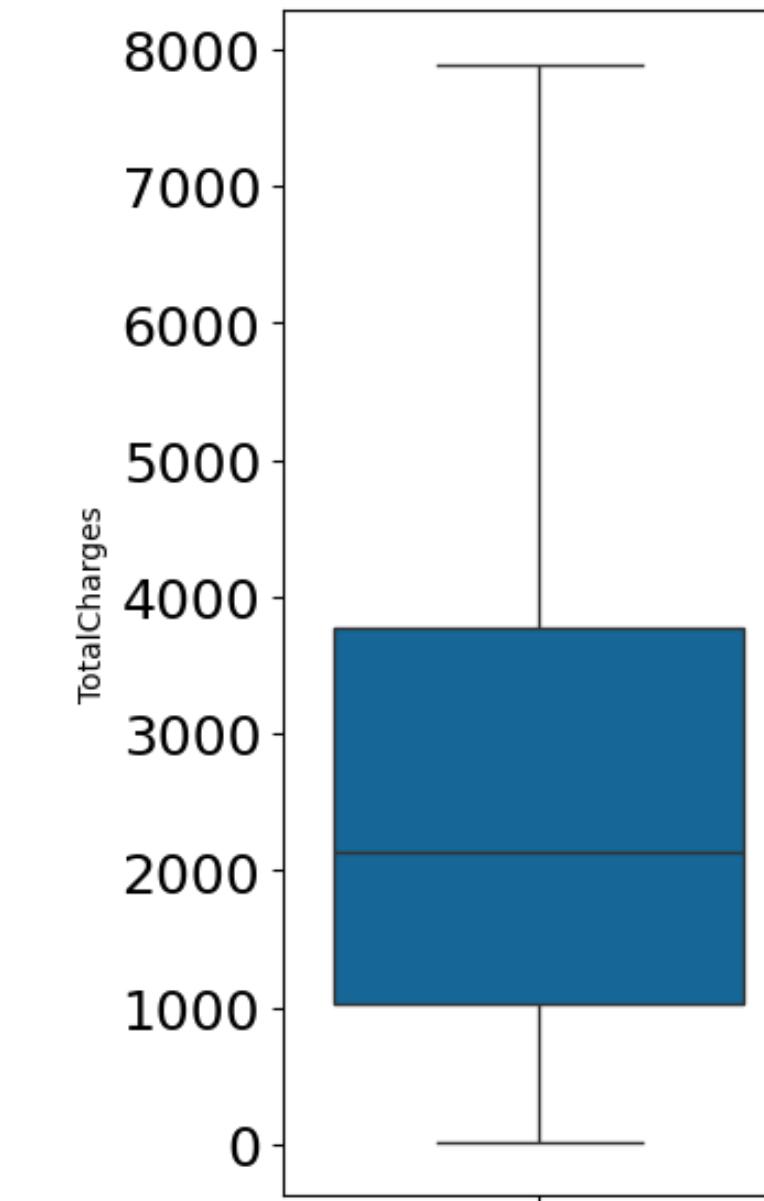
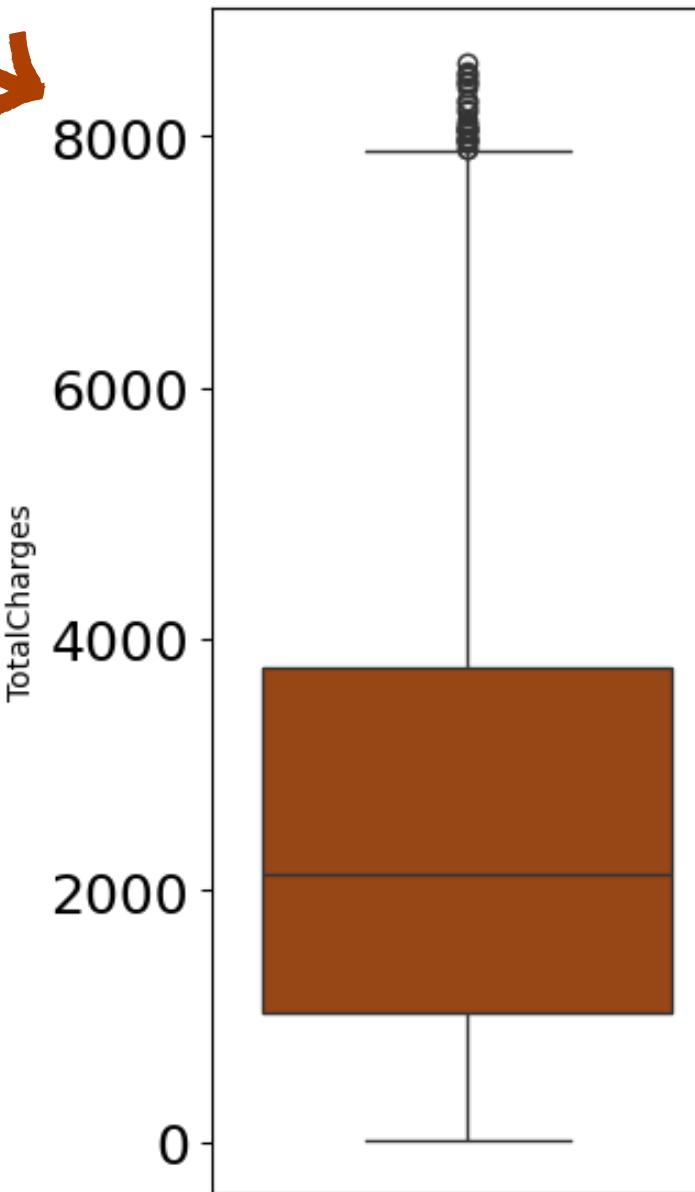
- ดรอปฟีเจอร์ที่ไม่มีผลต่อการคำนวณของโมเดล และฟีเจอร์ที่ไม่ได้ใช้งานแล้ว ซึ่งได้แก่ Customer ID, Online Security, Online Backup และ Device Protection

gender	object	MultipleLines	object	PaperlessBilling	object
SeniorCitizen	category	InternetService	object	PaymentMethod	object
Partner	object	TechSupport	object	MonthlyCharges	float64
Dependents	object	StreamingTV	object	TotalCharges	float64
tenure	int64	StreamingMovies	object	SecurityService	object
PhoneService	object	Contract	object	Churn	object

# Handling Outliers

- กำจัด Outliers ในฟีเชอร์ Total Charges

boxplot ก่อนการ  
กำจัด Outliers



boxplot หลังการ  
กำจัด Outliers

# Feature Encoding

- แปลง Categorical Features ให้เป็น Numerical Features โดยใช้ One-Hot Encoding

0	tenure	int64		
1	MonthlyCharges	float64		
2	TotalCharges	float64		
3	gender_Male	bool		
4	SeniorCitizen	bool		
5	Partner	bool		
6	Dependents	bool		
7	PhoneService	bool		
8	MultipleLines_No phone service	bool		
9	MultipleLines_Yes	bool		
10	InternetService_Fiber optic	bool		
11	InternetService_No	bool		
12	TechSupport_No internet service	bool		
13	TechSupport_Yes	bool		
14	StreamingTV_No internet service	bool		
15	StreamingTV_Yes	bool		
16	StreamingMovies_No internet service	bool		
17	StreamingMovies_Yes	bool		
18	Contract_One year	bool		
19	Contract_Two year	bool		
20	PaperlessBilling	bool		
21	PaymentMethod_Credit card	bool		
22	PaymentMethod_Electronic check	bool		
23	PaymentMethod_Mailed check	bool		
24	Churn	bool		
25	SecurityService_No internet service	bool		
26	SecurityService_Yes	bool		

ชุดข้อมูลหลังใช้ One-Hot Encoding

# Feature Scaling

- ทำให้ข้อมูลในฟีเชอร์ tenure, Monthly Charges และ Total Charges มีความใกล้เคียงกัน โดยใช้ Min-Max Scaling ทำให้ข้อมูลอยู่ในช่วง 0 - 1

tenure	MonthlyCharges	TotalCharges
23	49.85	1146.55
43	100.70	4330.10
51	97.33	4963.83
72	101.38	7299.36
25	52.22	1305.50

# ចុះមុនការការណ៍ Min-Max Scaling

tenure	MonthlyCharges	TotalCharges
0.309859	0.298530	0.143163
0.591549	0.807081	0.547741
0.704225	0.773377	0.628278
1.000000	0.813881	0.925087
0.338028	0.322232	0.163363

## ▪ ชุดข้อมูลหลังการทำ Min-Max Scaling

# Spilt Train and Test Data

- ใช้ train test split ของ sklearn ในการแบ่งข้อมูลสำหรับ Train และ Test
- โดยใช้สัดส่วนที่ Train: 70% และ Test 30%

	Train	Test
Churn	2,008	890
Non-Churn	2,108	874
Total	4,116	1,764

# Supervised Learning

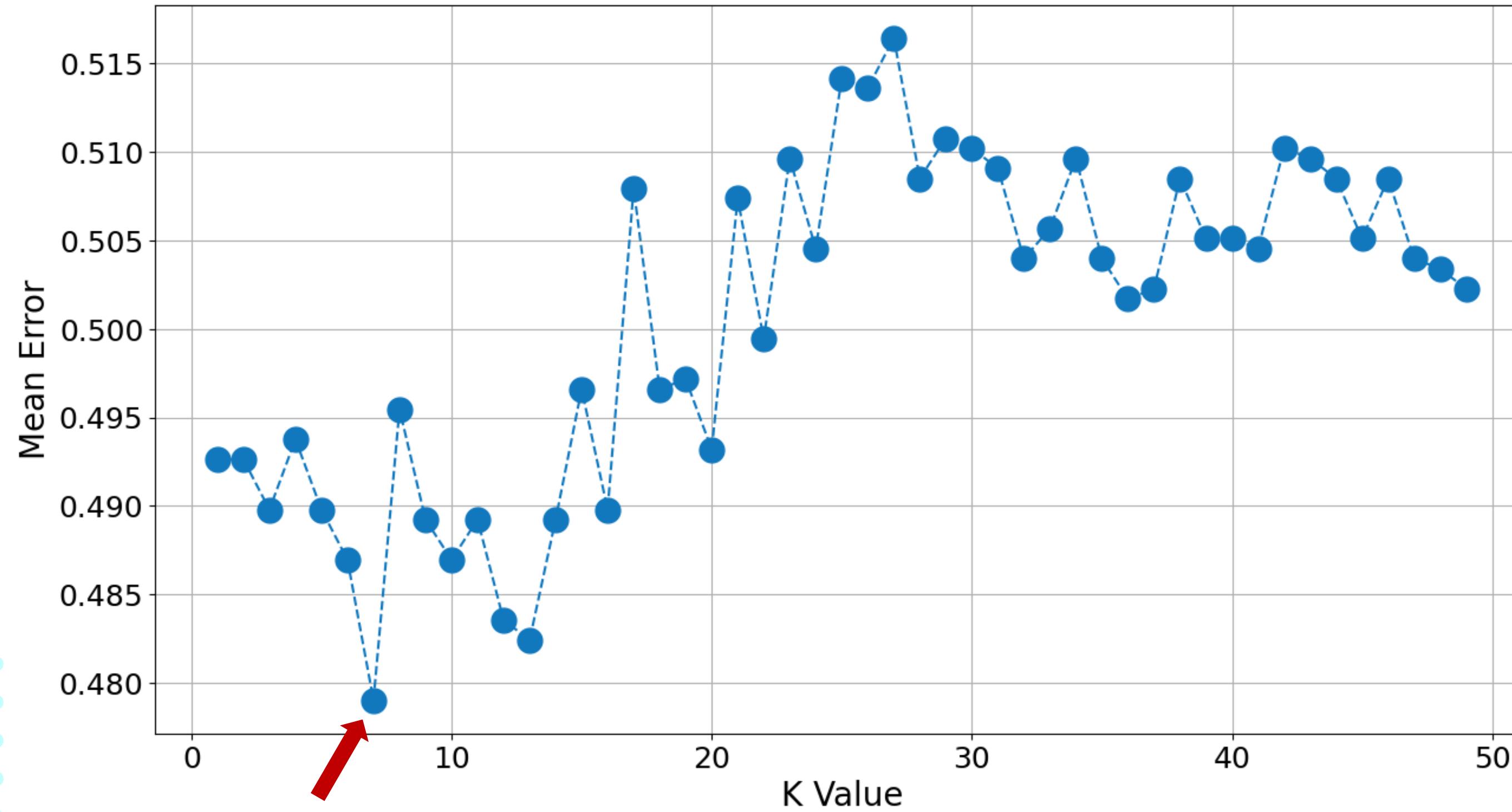
## K Nearest Neighbors Classifier

**without Principal Component Analysis**



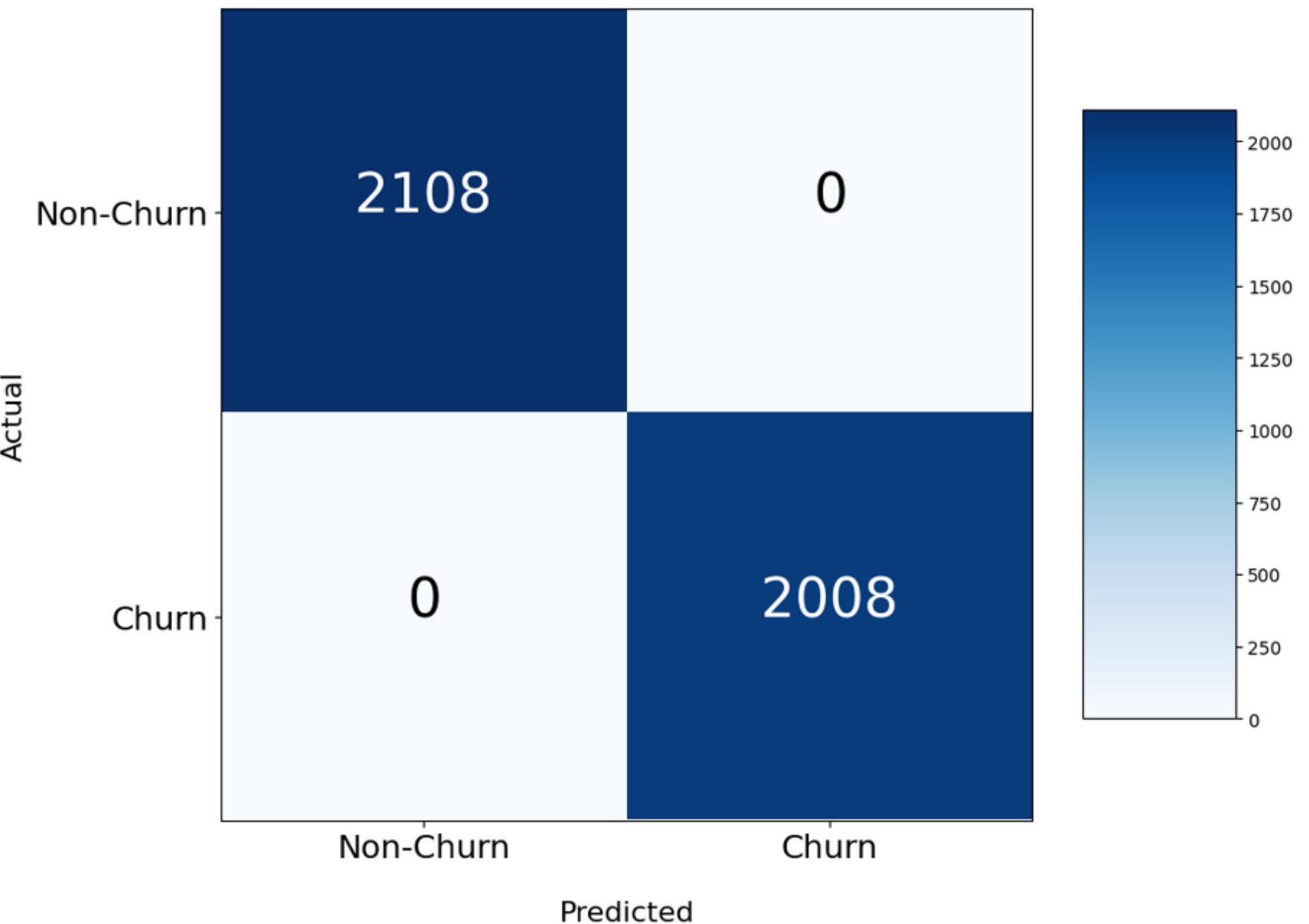
# KNN without PCA

- หากค่า K ที่จะนำมาใช้ Train model

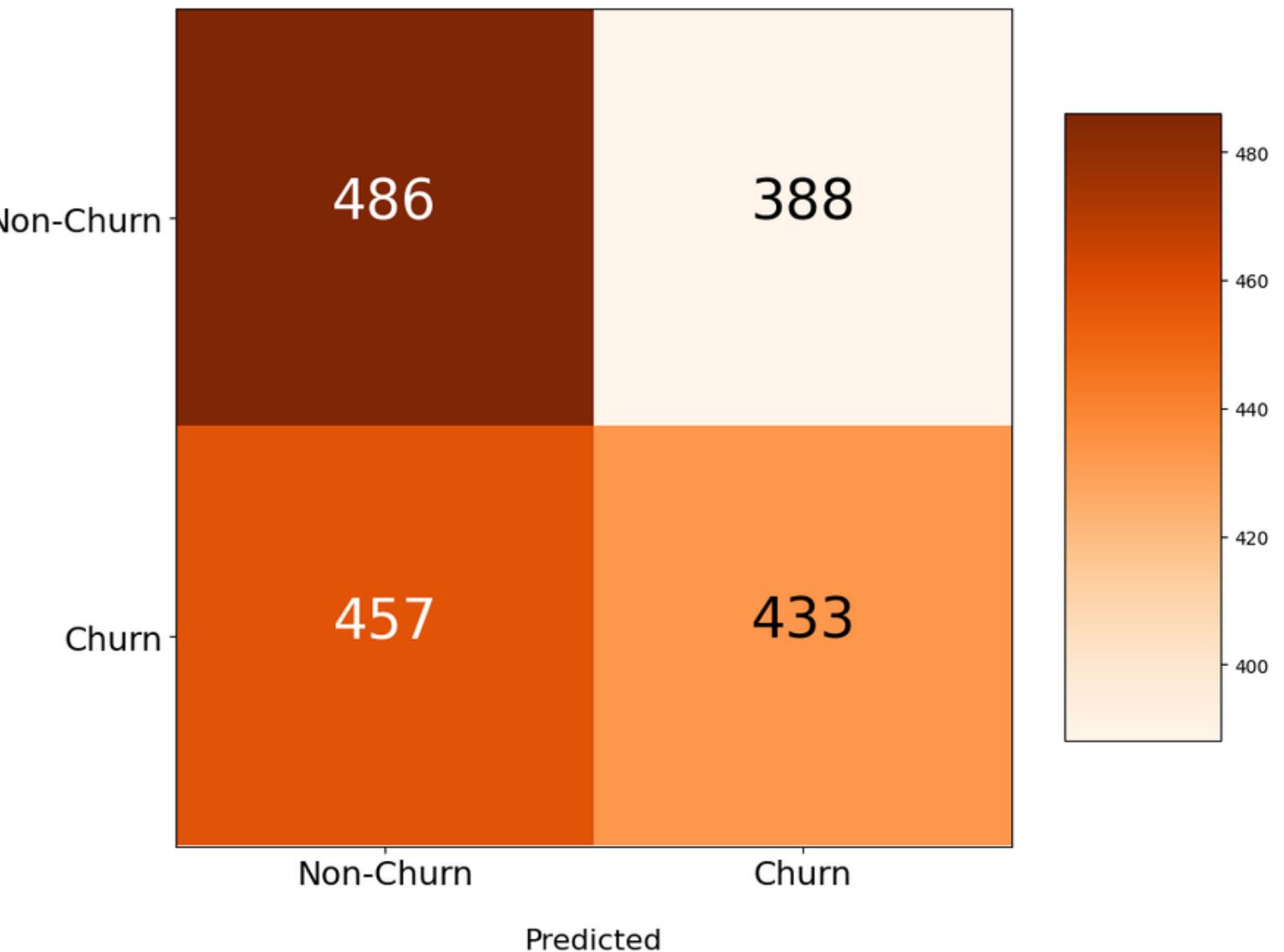


# KNN without PCA

**Train**



**Test**

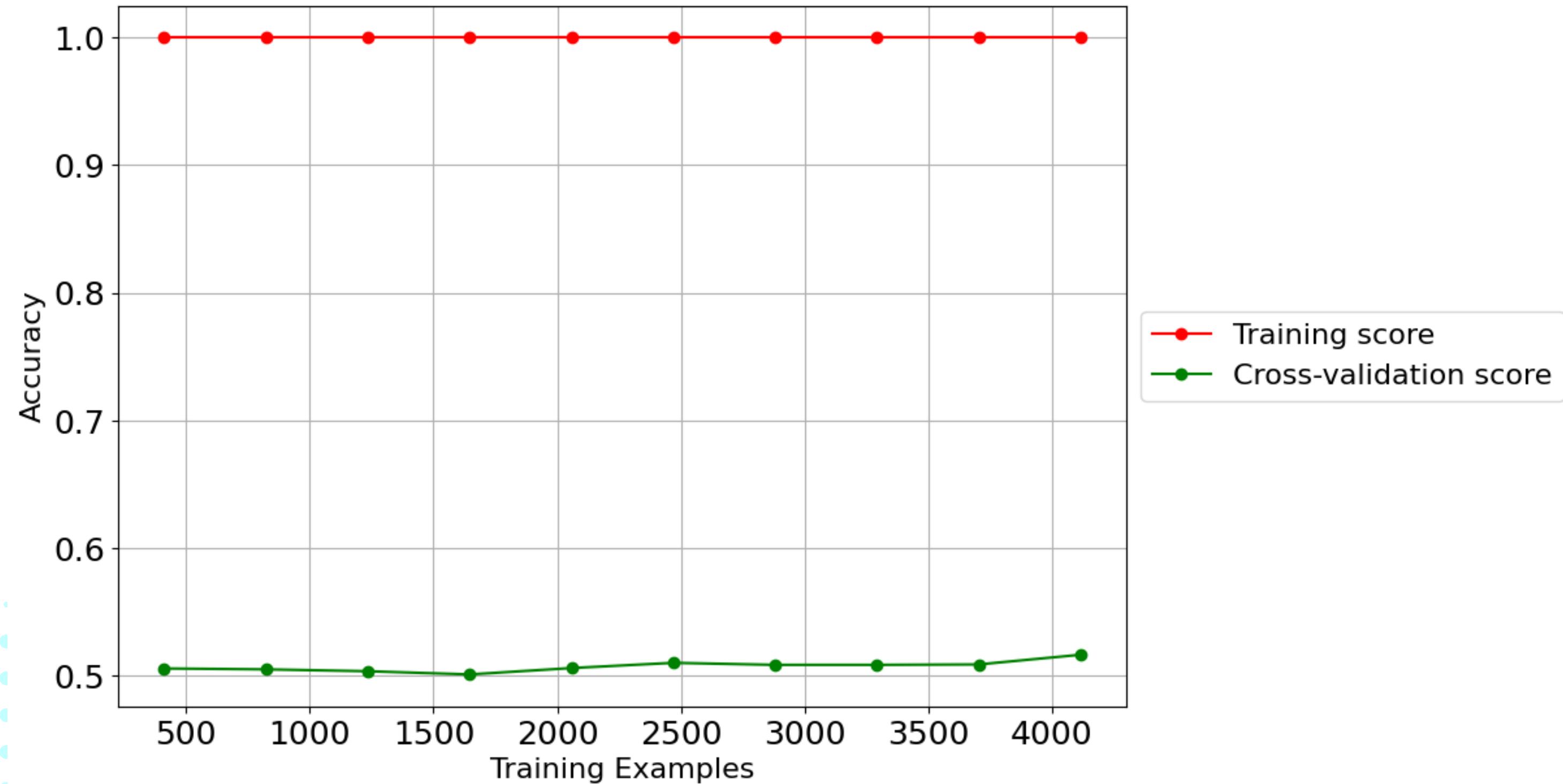


ความแม่นยำ: 100.00%

ความแม่นยำ: 52.10%

# KNN without PCA

- Learning curve



# Supervised Learning

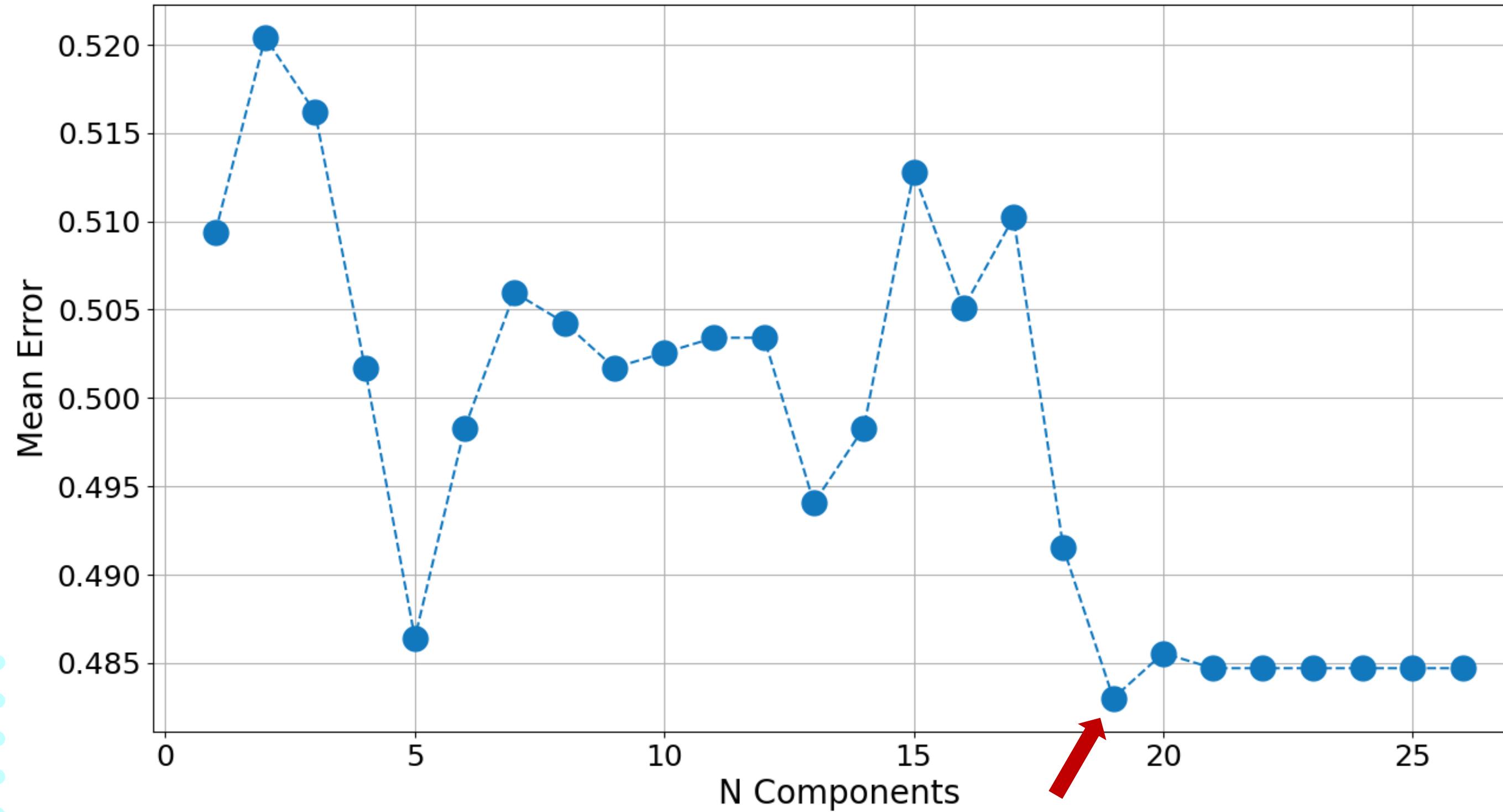
## K Nearest Neighbors Classifier

with Principal Component Analysis



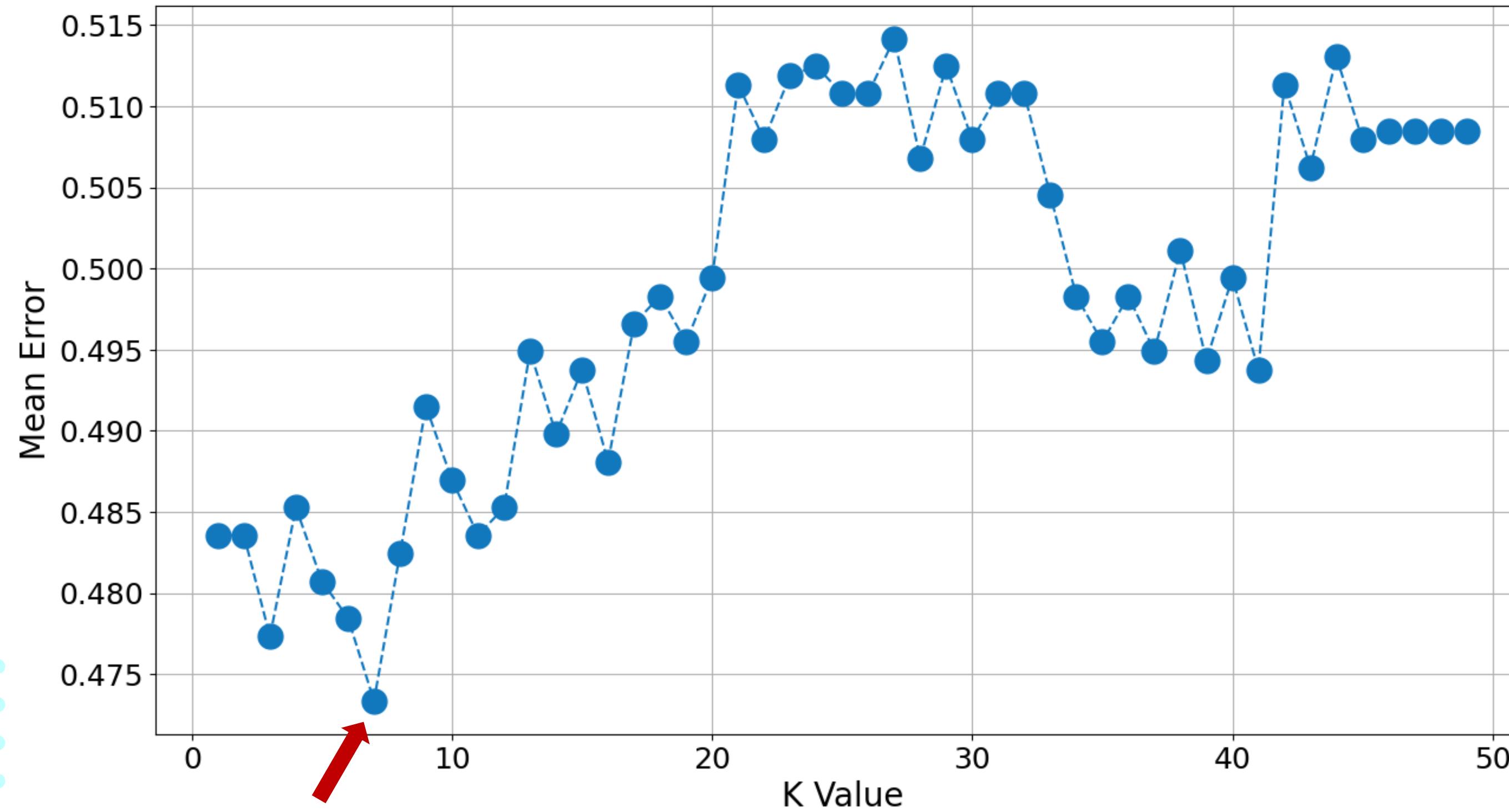
# KNN with PCA

- หากจำนวน component ที่จะนำมาใช้ในการทำ PCA



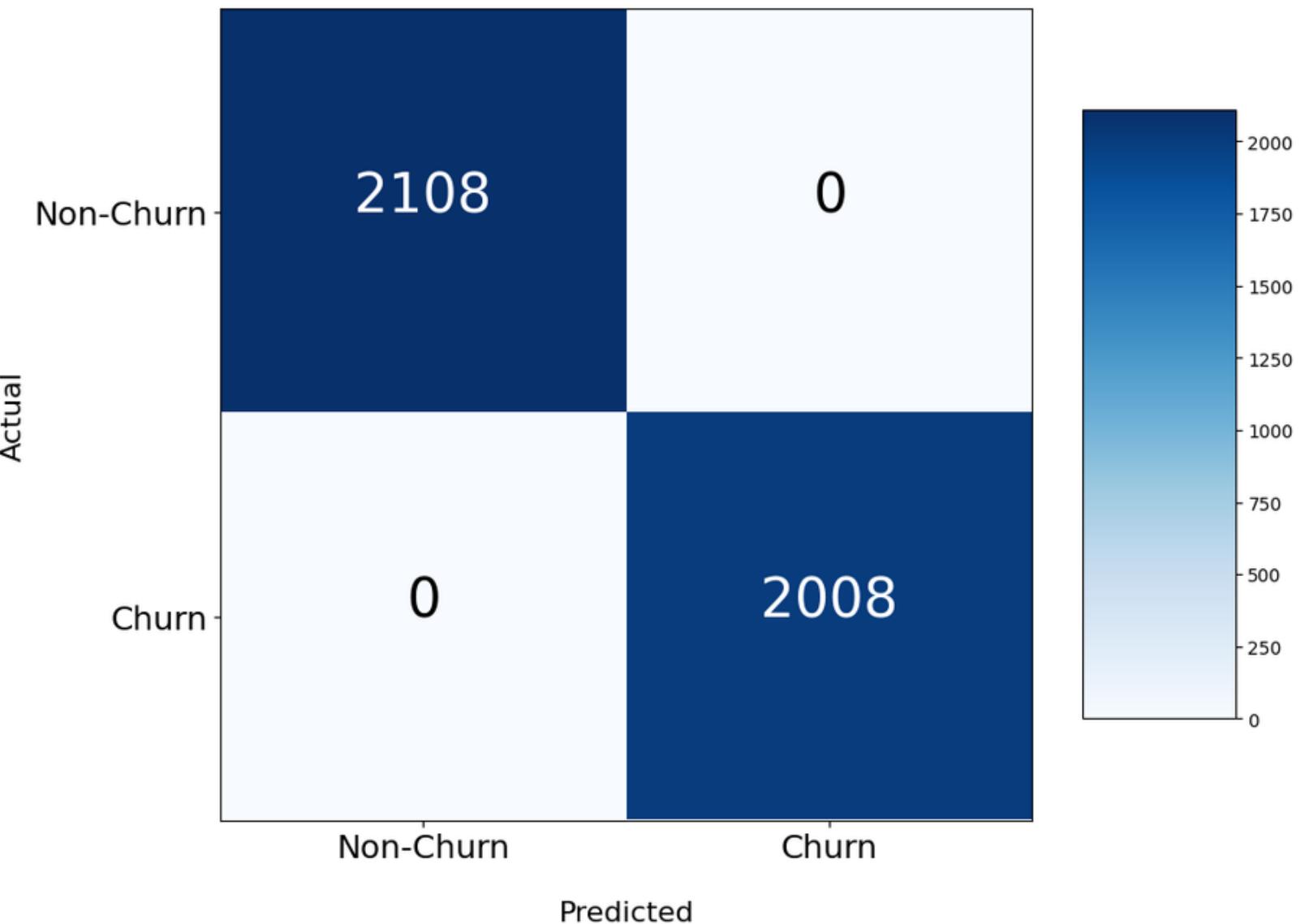
# KNN with PCA

- หากค่า K ที่จะนำมาใช้ Train model

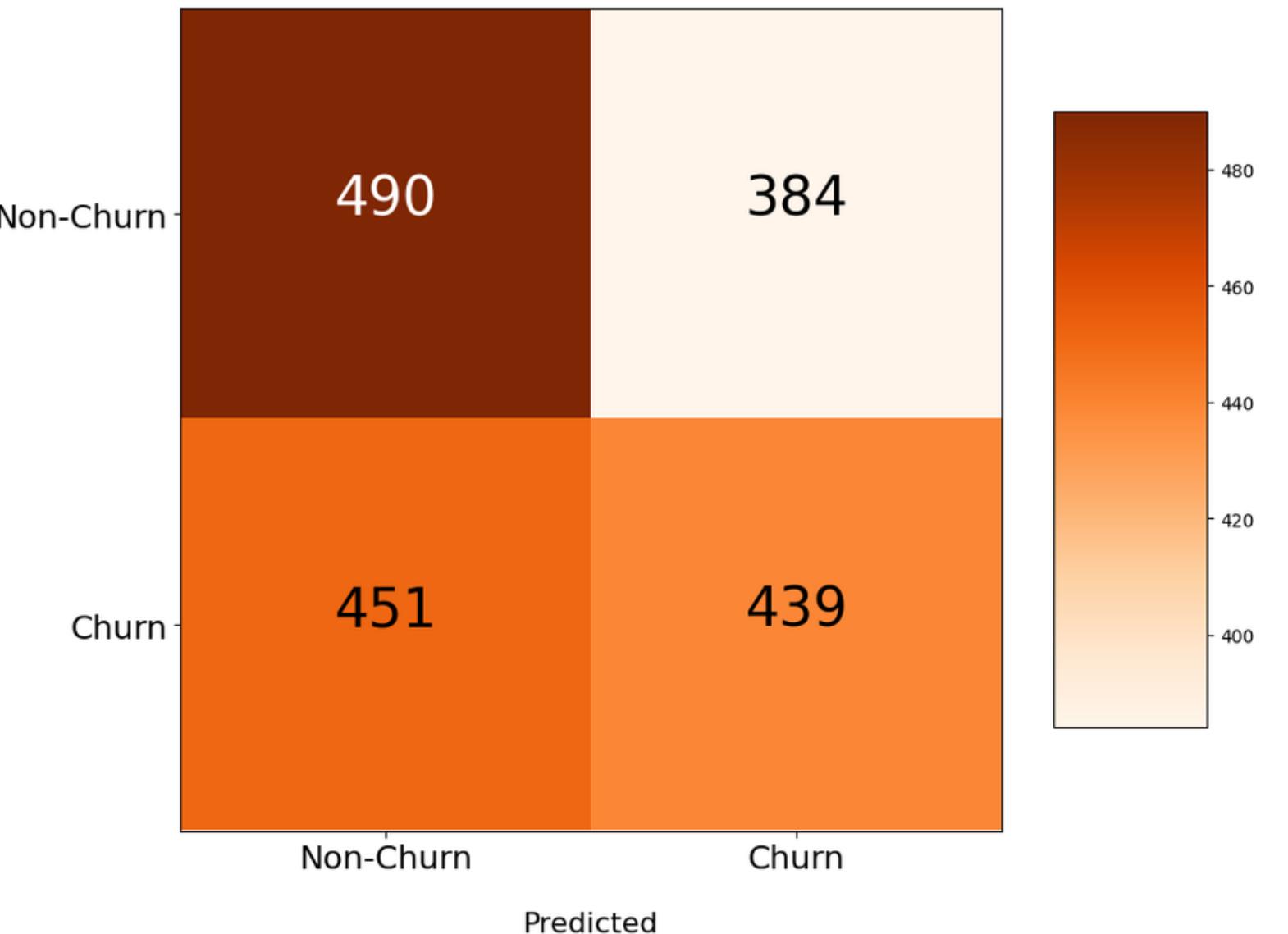


# KNN with PCA

**Train**

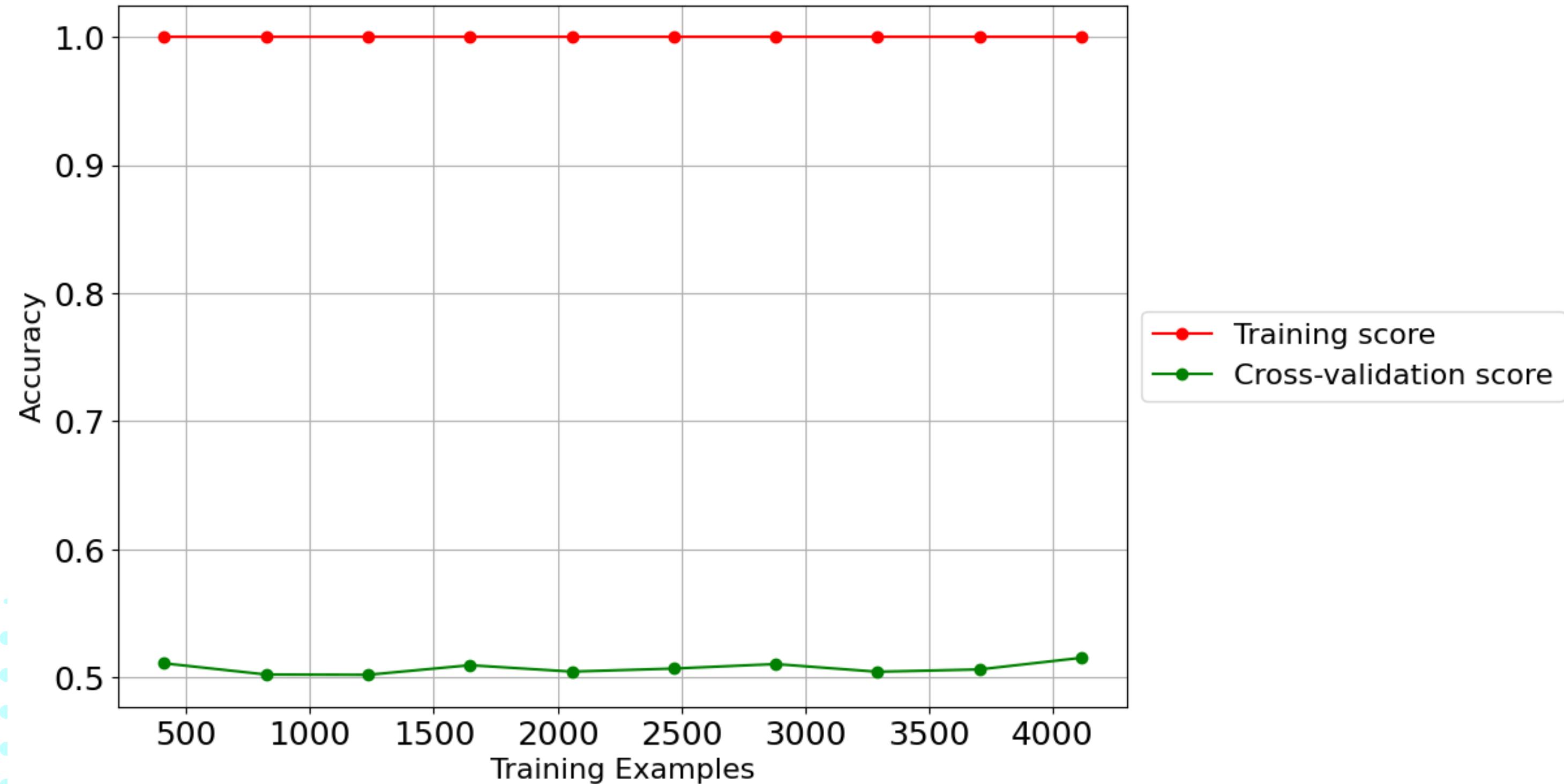


**Test**



# KNN without PCA

- Learning curve



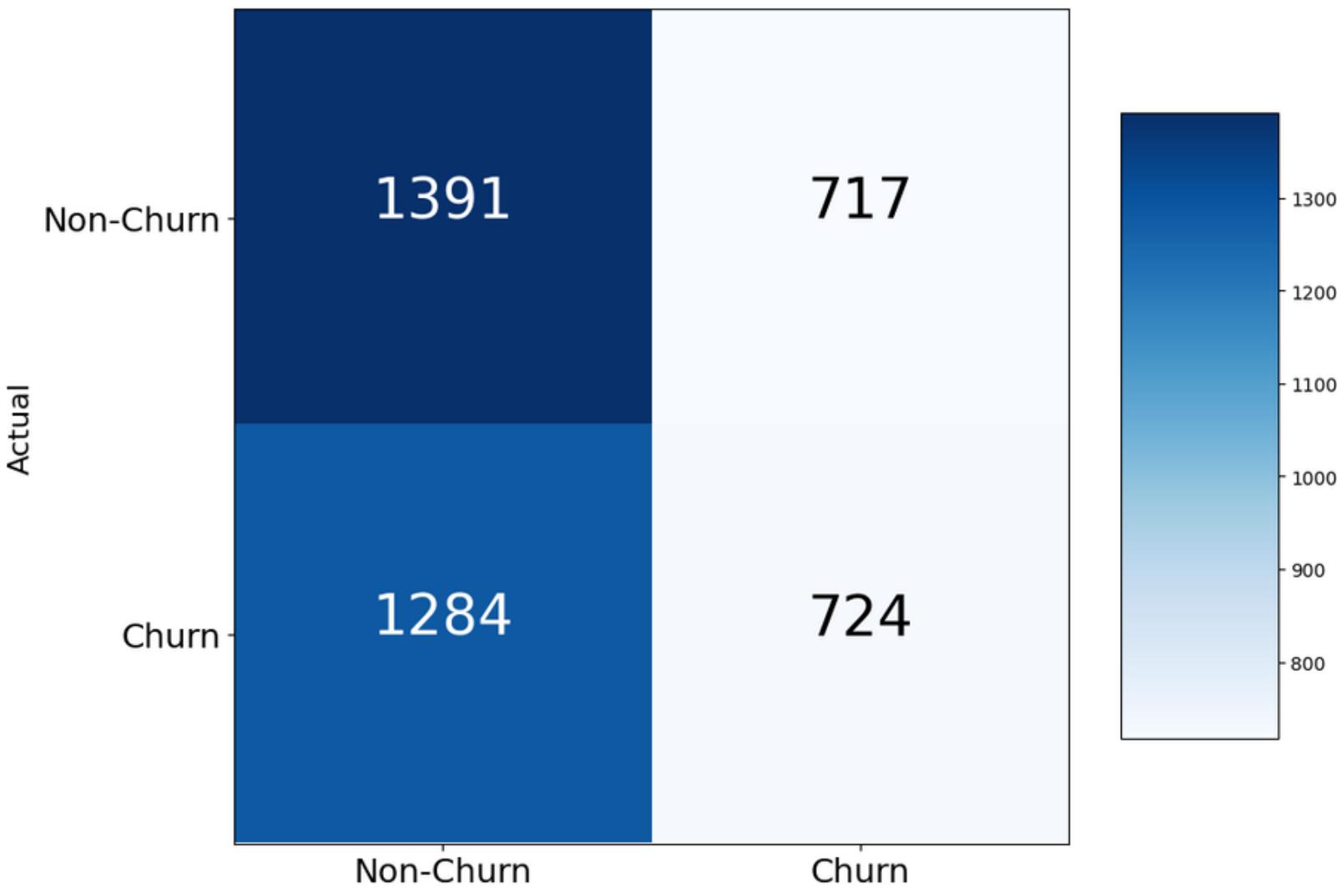
# Supervised Learning

## K Means Clustering



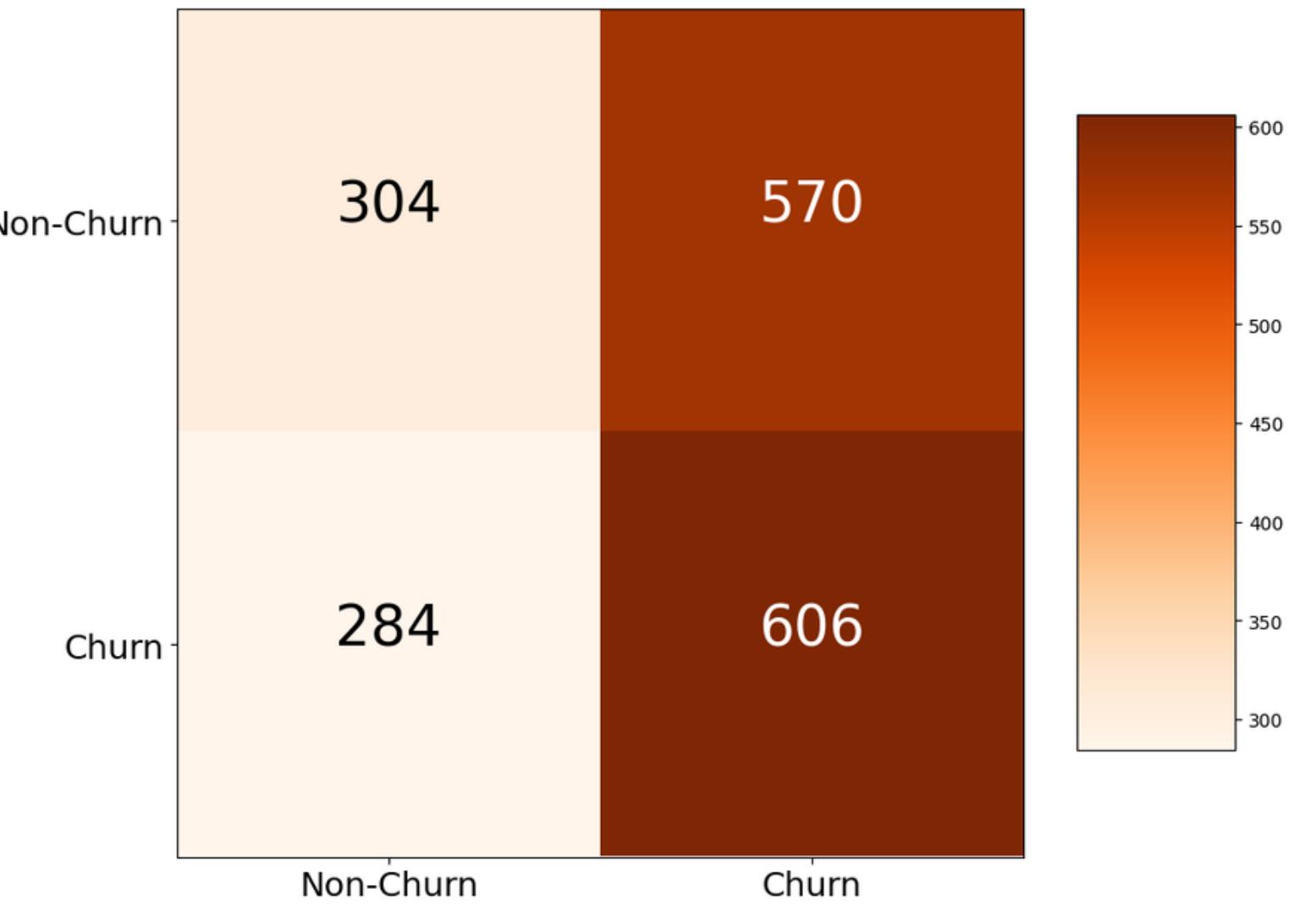
# K Means

**Train**



ความแม่นยำ: 51.38%

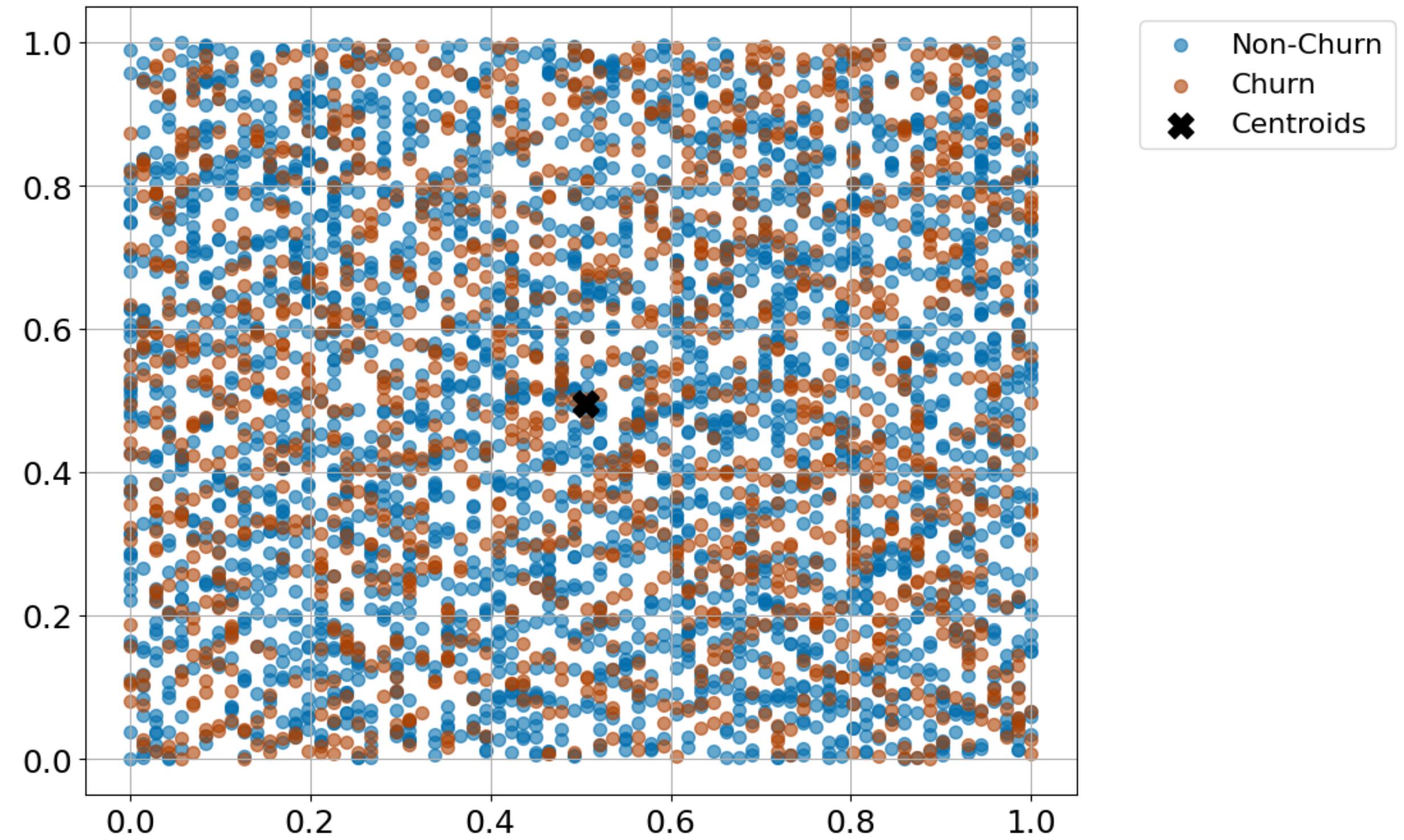
**Test**



ความแม่นยำ: 51.59%

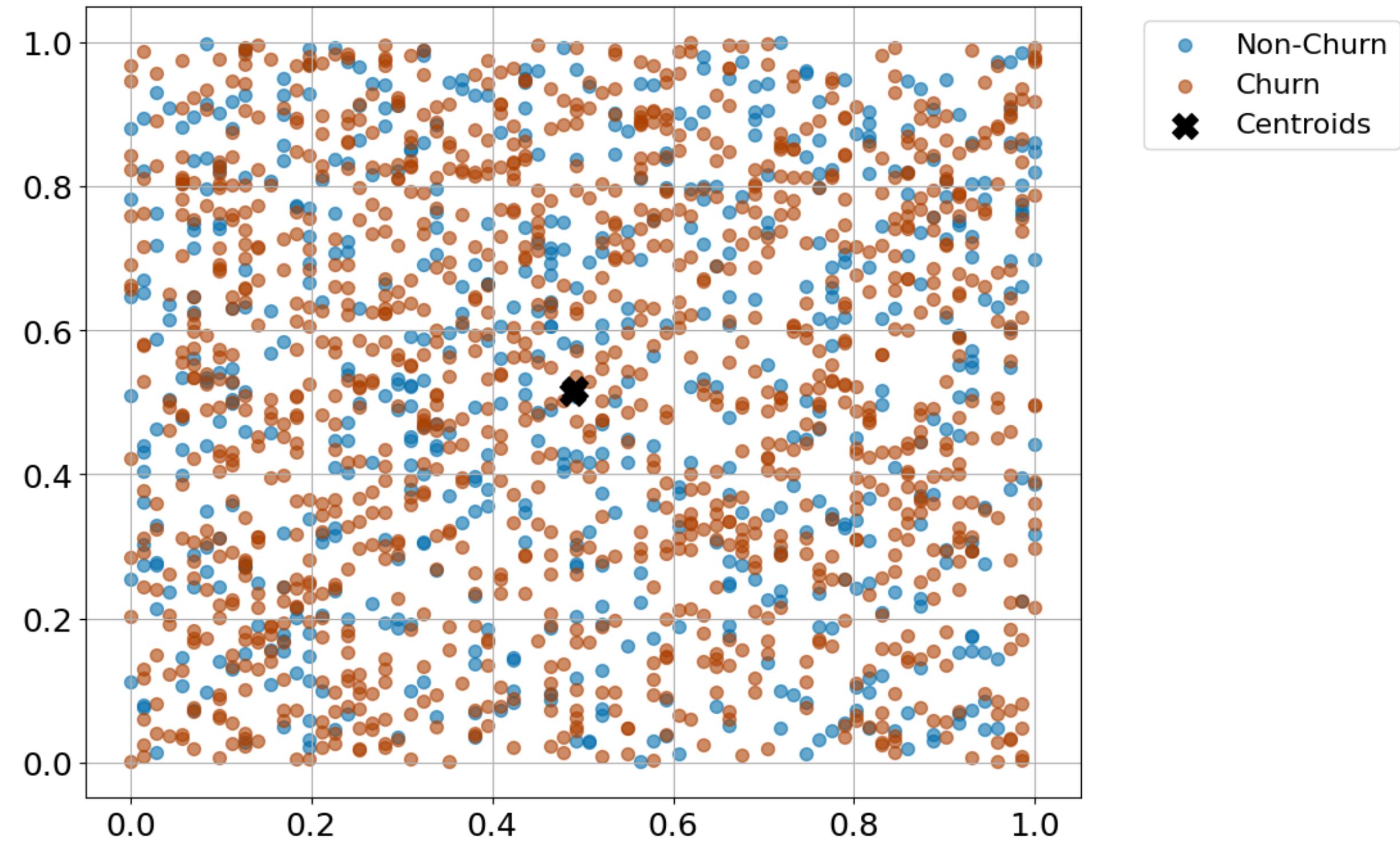
# K Means

- မာနရှိပံ့ clusters မျှ၏ Train



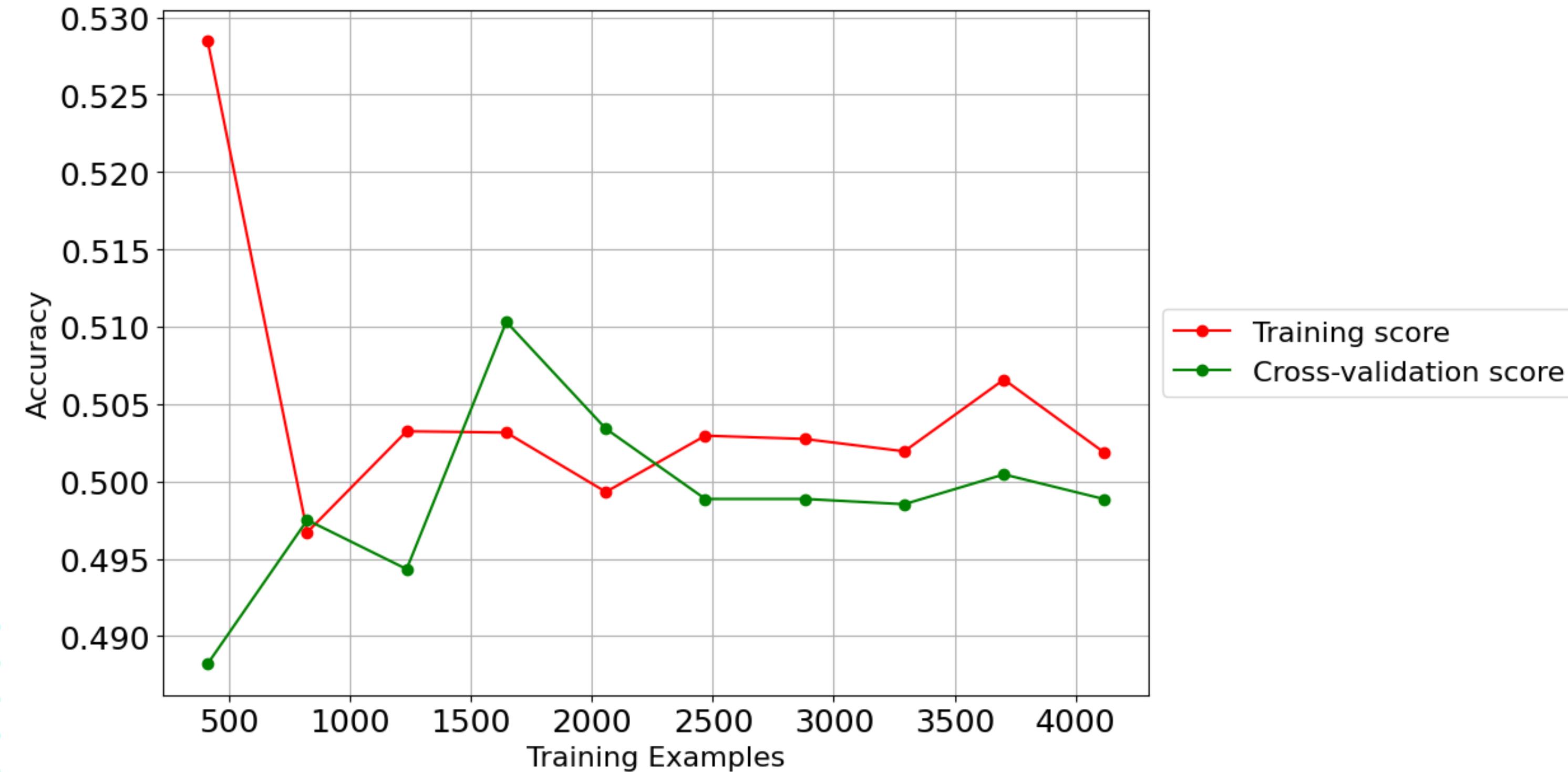
# K Means

- ພາມາຮັບປິ່ງ clusters ?unns Test



# K Means

- Learning curve





# THANK YOU

นายศุภณัฐ แซ่เตีย

ID: 6505000270



## อ้างอิง

- ผศ. ดร.ชญาพร เกตุณณี. (2568). เอกสารประกอบการเรียนวิชา COS3302 การบูรณาการศาสตร์ทางข้อมูล. สืบคันเมื่อ 17 เมษายน 2568
- BHARTI PRASAD. (2564). CUSTOMER CHURN PREDICTION. สืบคันเมื่อ 17 เมษายน 2568, จาก <https://www.kaggle.com/code/bhartiprasad17/customer-churn-prediction>
- OpenAI. (2567). ChatGPT (Version 4) [Large language model]. สืบคันเมื่อ 17 เมษายน 2568, จาก <https://chatgpt.com/>
- Rashad Mammadov . (2567). Customer Churn Prediction. สืบคันเมื่อ 17 เมษายน 2568, จาก <https://www.kaggle.com/datasets/rashadrmammadov/customer-churn-dataset/data>

## อ้างอิง

- Rashad Mammadov . (2567). Rashad Mammadov Profile. สืบค้นเมื่อ 17 เมษายน 2568, จาก <https://www.kaggle.com/rashadrmammadov>
- scikit-learn developer. (2568). KNeighborsClassifier. สืบค้นเมื่อ 23 กุมภาพันธ์ 2568, จาก <https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html>
- scikit-learn developer.. (2568). PCA. สืบค้นเมื่อ 17 เมษายน 2568, จาก <https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html>

## อ้างอิง

- scikit-learn developers. (2568). KMeans. สืบคันเมื่อ 17 เมษายน 2568, จาก <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html>
- scikit-learn developers. (2568). train\_test\_split. สืบคันเมื่อ 23 กุมภาพันธ์ 2568, จาก [https://scikit-learn.org/stable/modules/generated/sklearn.model\\_selection.train\\_test\\_split.html](https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html)