



# Improving customer retention

# Outline

- 1.** Context & objectives of the meeting
- 2.** Presentation of the data analysis results
- 3.** Machine Learning methodology used
- 4.** Results and recommended marketing strategies



# **1. Context & objectives**

# Improving customer retention

**Objective:** Reduce customer churn by identifying at-risk customers and understanding the factors contributing to their decision to leave.

**Our analysis leverages data analytics and machine learning to:**

- Identify patterns and segments within the customer base.
- Predict customer churn using advanced predictive modeling.
- Provide actionable insights to reduce churn rates and enhance customer satisfaction.



# Key questions

**Which customers are most at risk of churning?**

**What common traits define customers who are likely to churn?**

**How can the company improve customer retention?**



# **2. Data Analysis**

## **results**

# Data

**Churn: target variable**

## **Services**

subscriptions of  
customers

**Customer account  
and billing details**

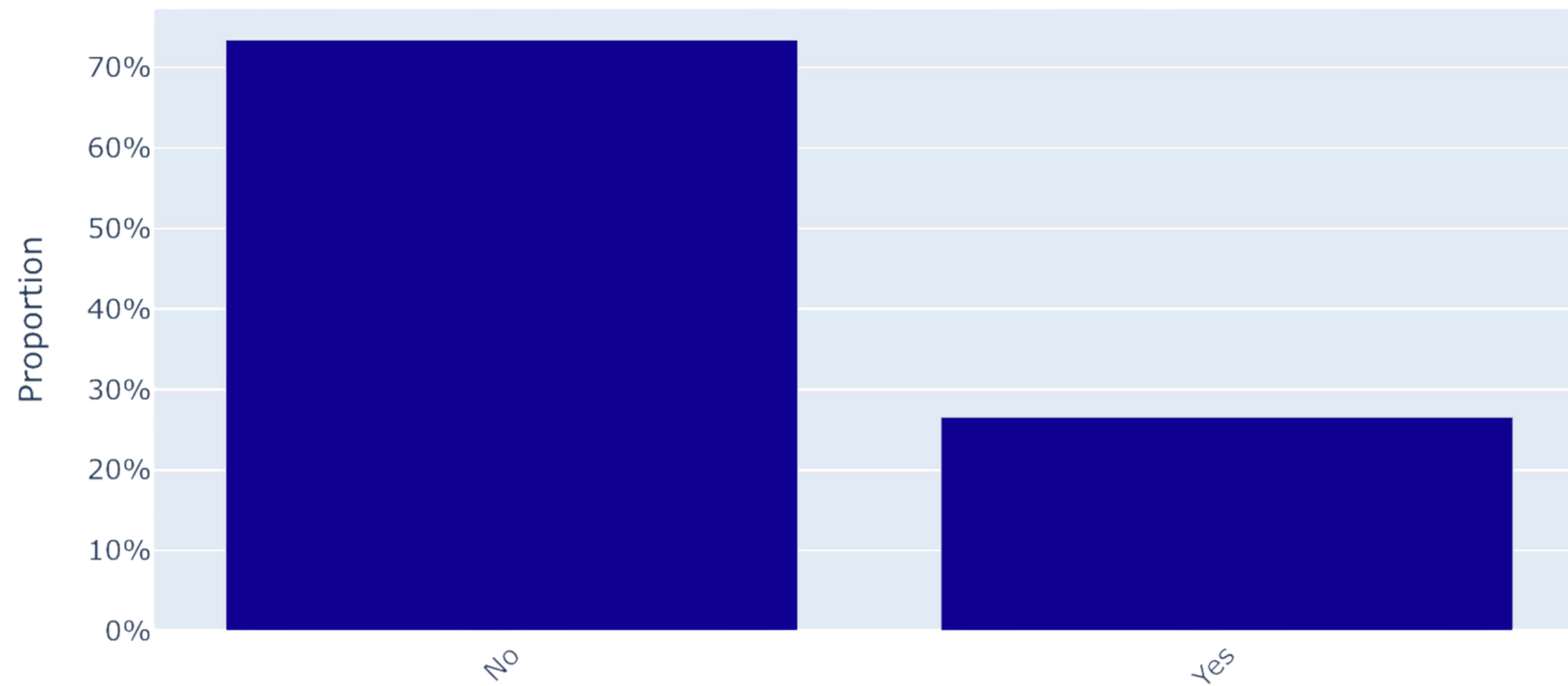
## **Demographic data**

Gender, age,  
partners, dependants

# Distribution of churn

**No: 5163 - 73%**

**Yes: 1869 - 27%**

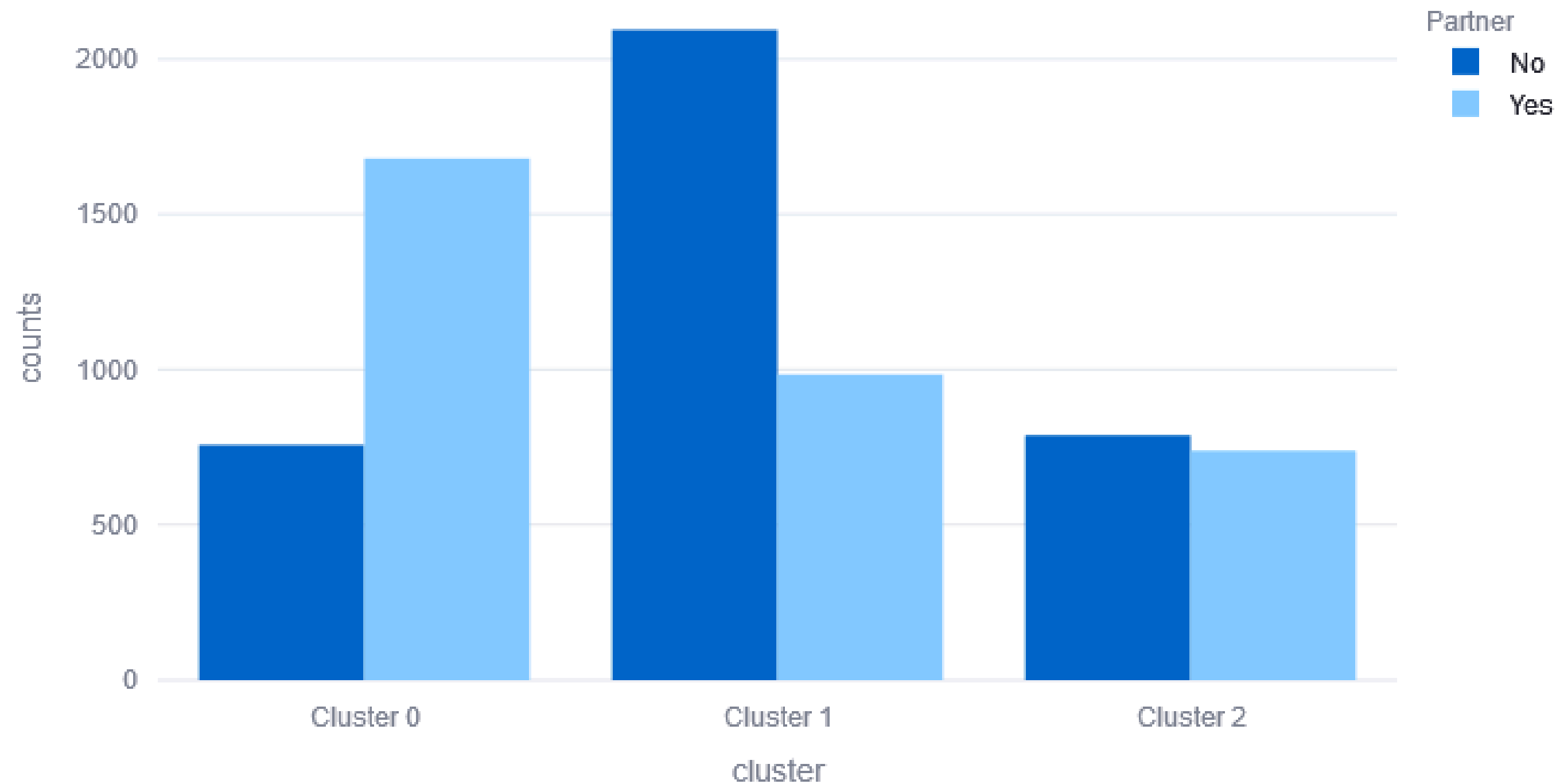




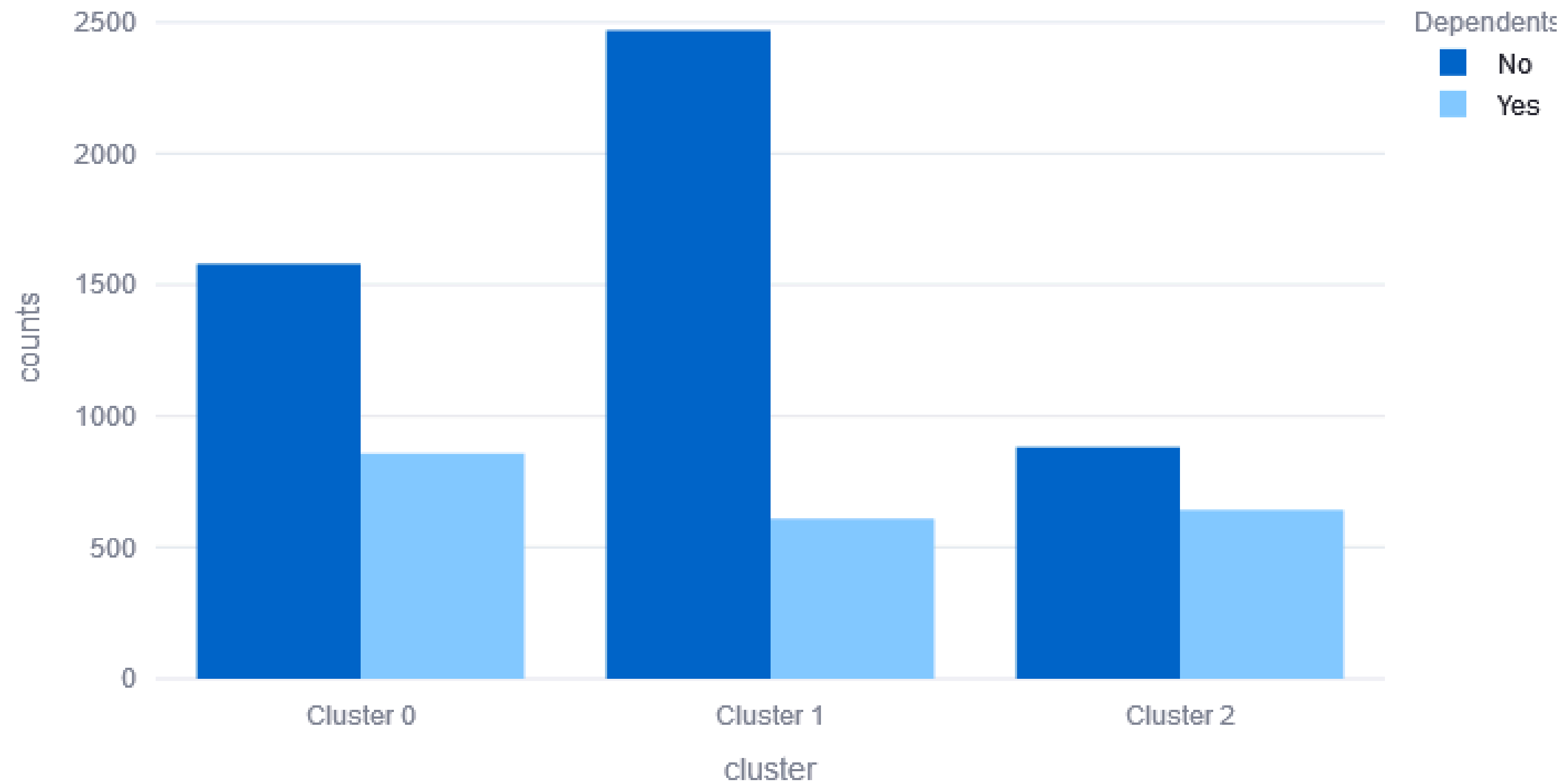
# Dashboard



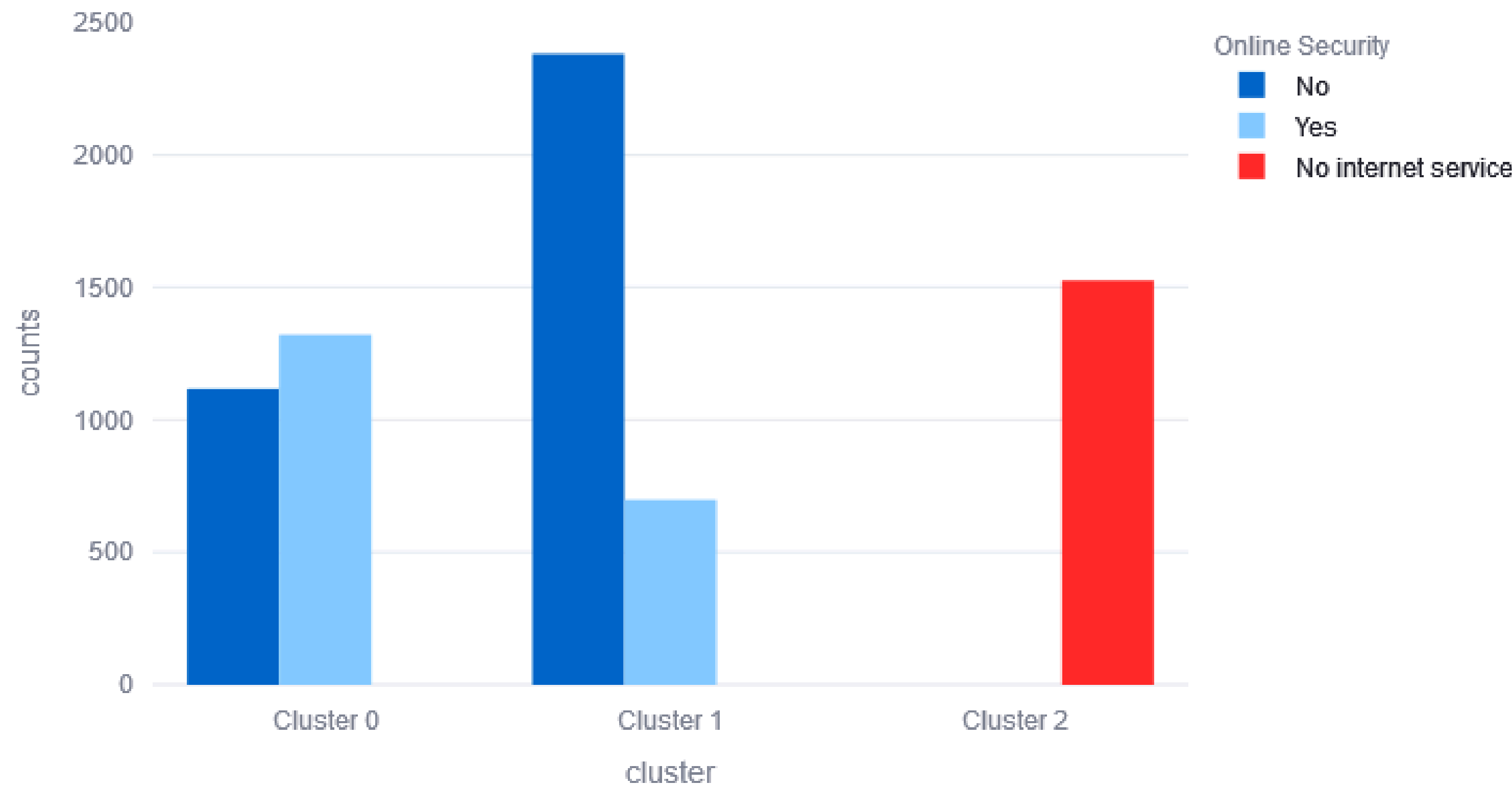
# Partner status by cluster



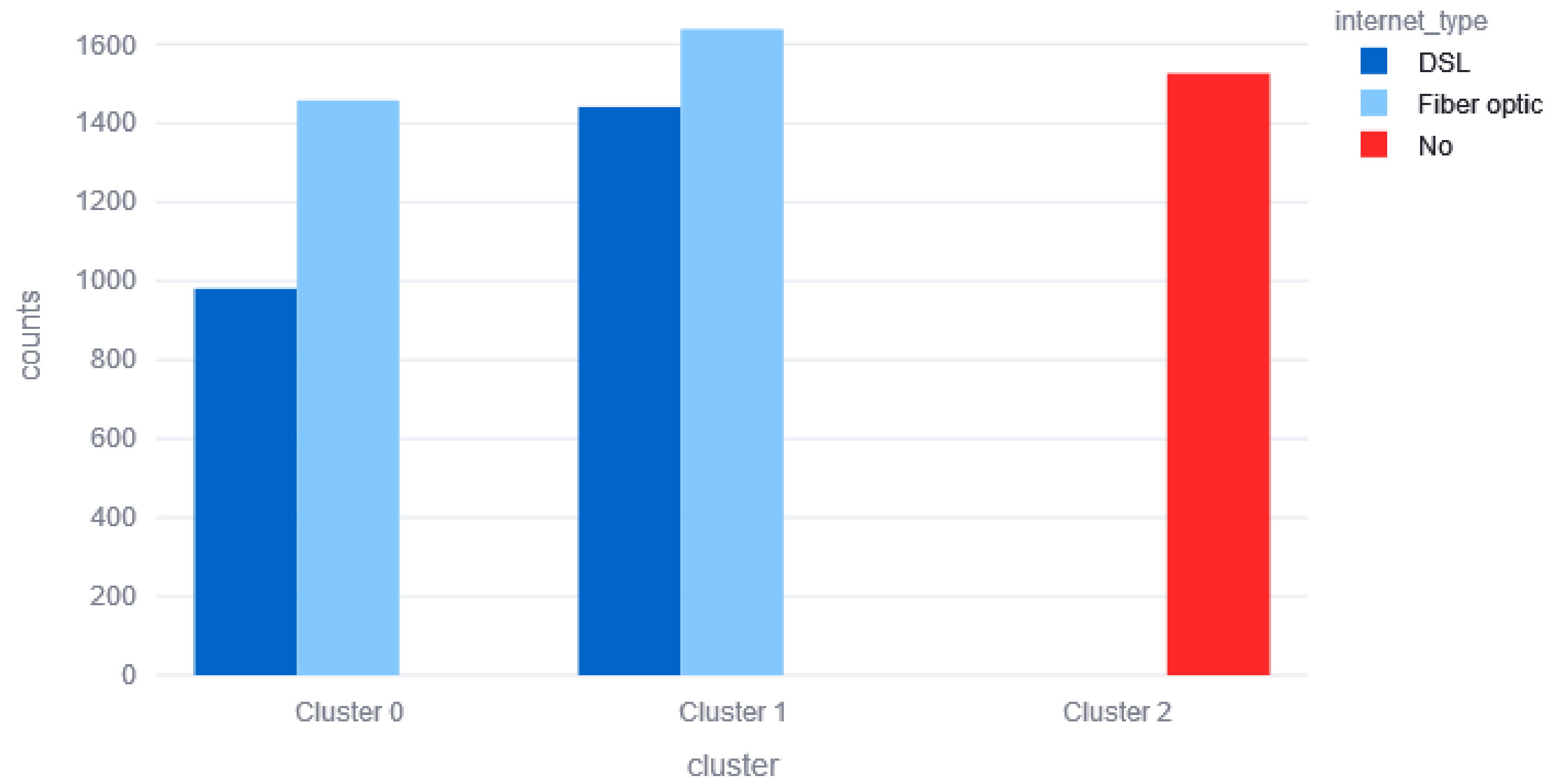
# Dependent by cluster



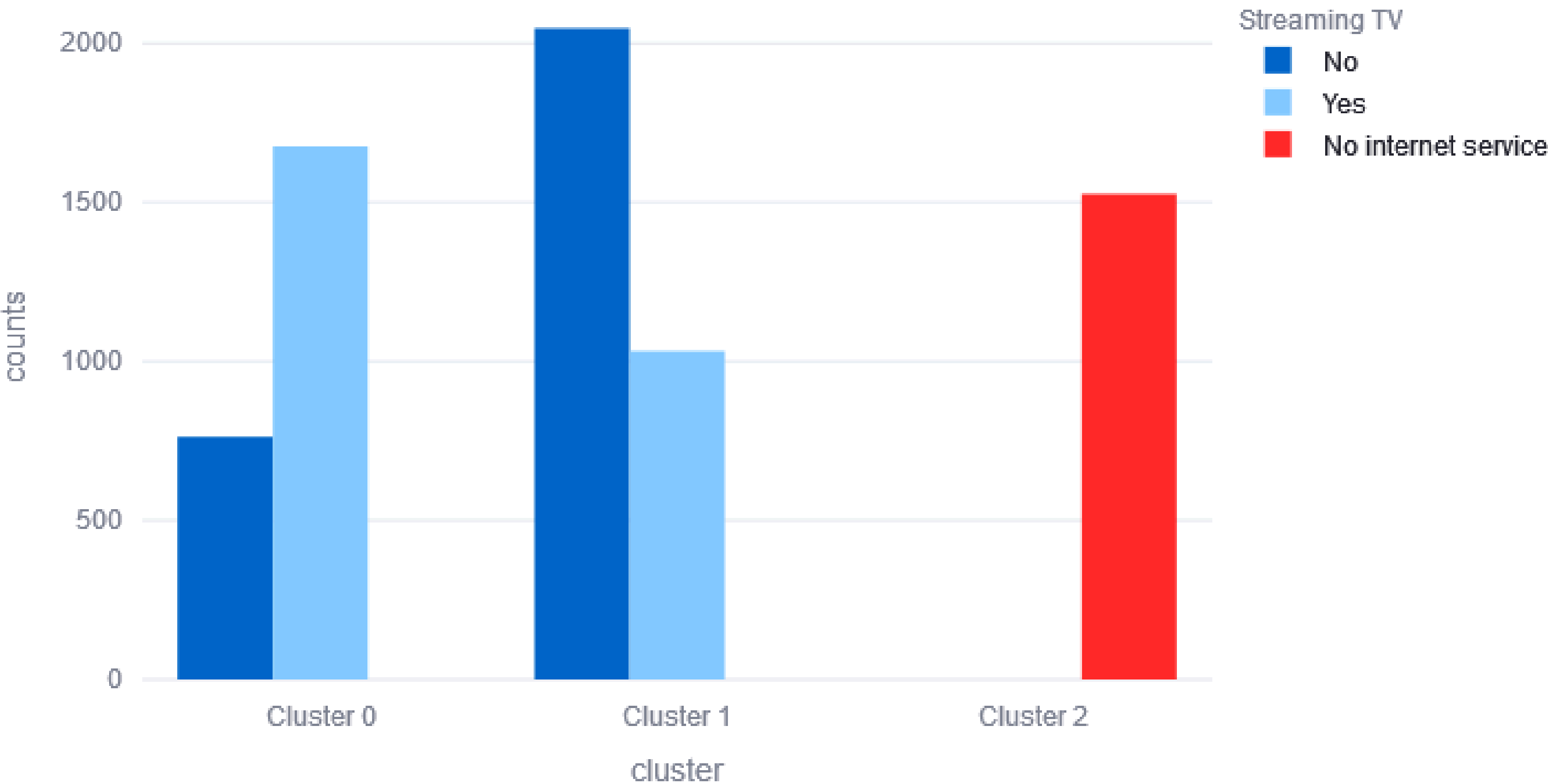
# Online Security by cluster



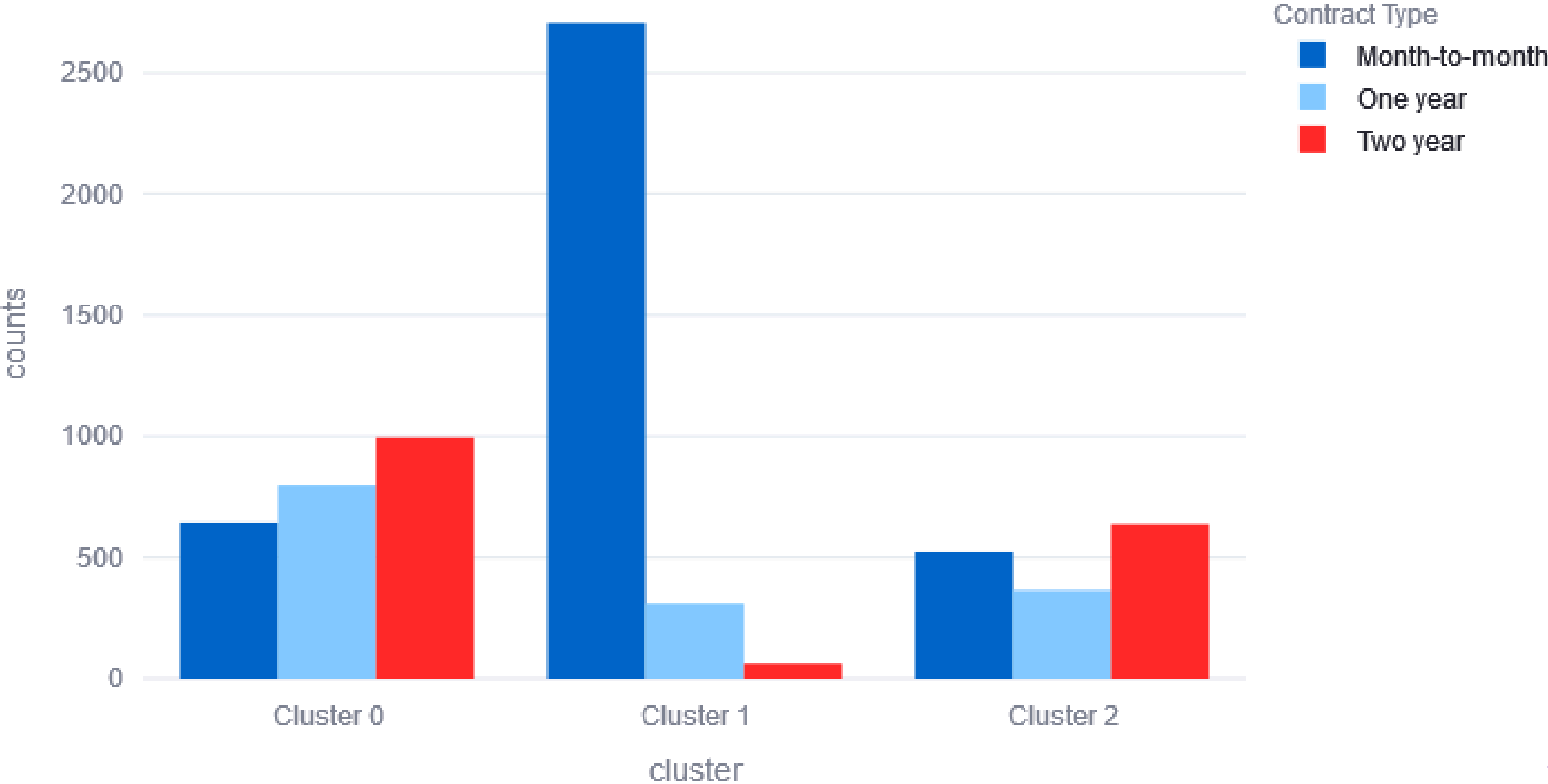
# Internet Service by cluster



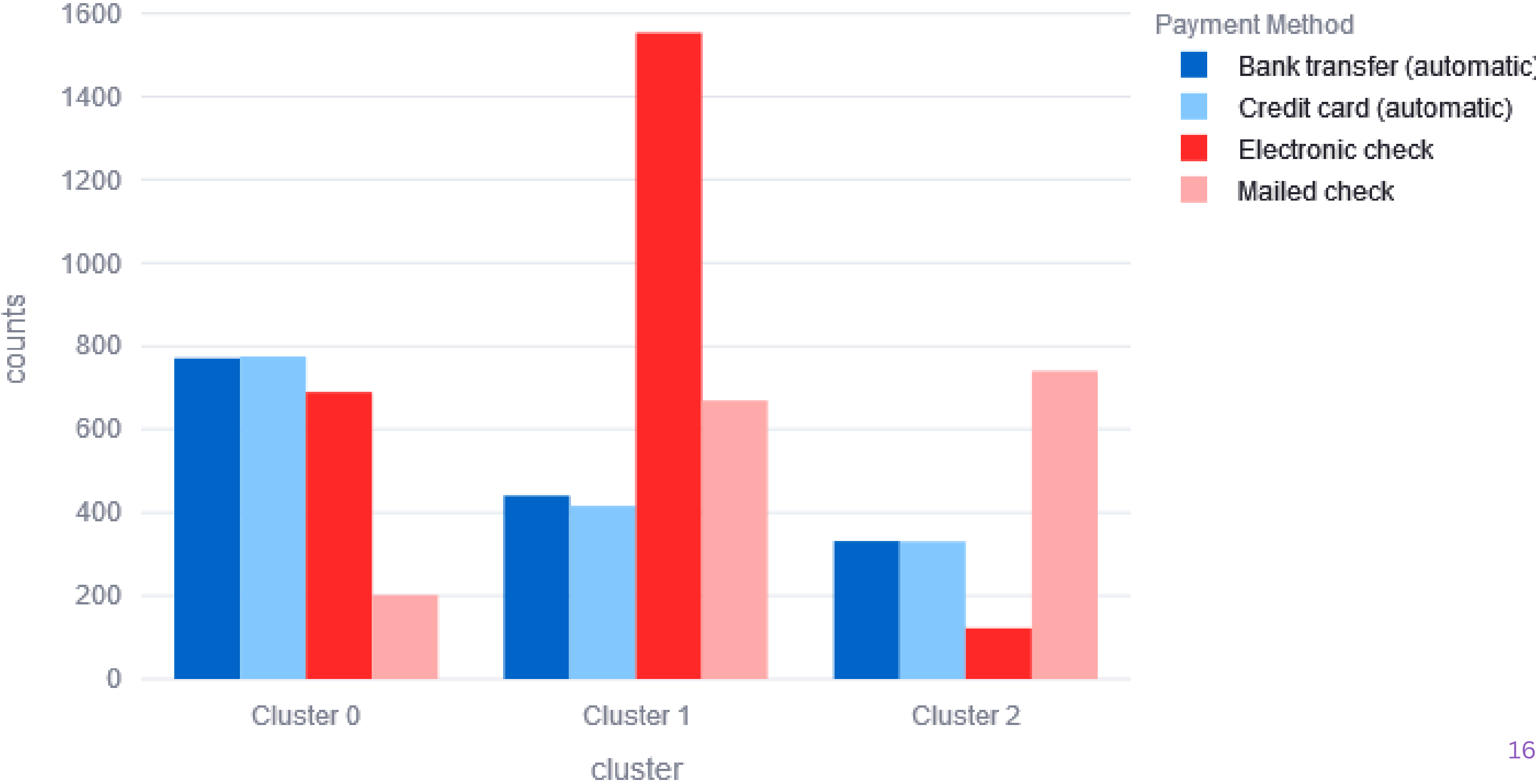
# Streaming TV by cluster



# Contract Type by cluster



# Payment method by cluster






# **3. Machine learning models**

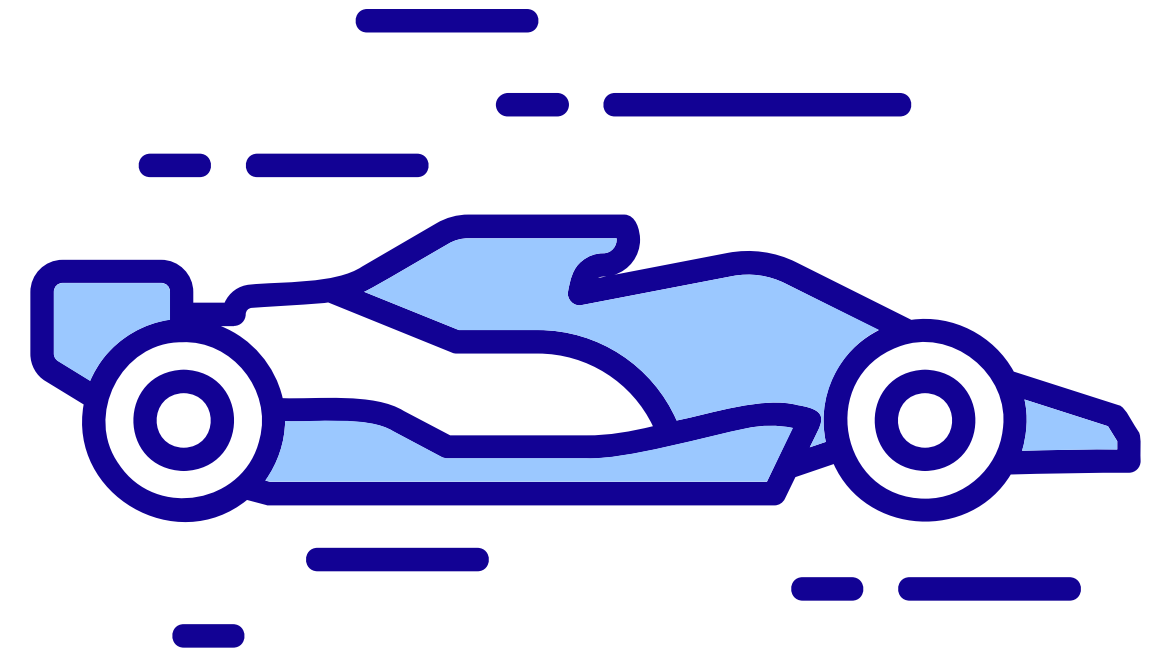
# Differents models

- Logistic Regression
- Random Forest
- Support Vector Machine
- XGBoost
- LightGBM



We use the grid  
search method  
to optimize them !

# Scoring metric : F1 score



- Good when the data is unbalanced
- Combines precision and recall
- Precision : how many of the predicted positives were actually correct
- Recall : how many of the actual positives were correctly predicted

# Scoring metric : ROC and AUC



- **ROC Curve:** Plots how well a model separates two classes (True Positive Rate vs. False Positive Rate).
- **AUC:** A single score summarizing the ROC curve—higher means better at distinguishing classes
- **Perfect AUC:** 1.0 = always correct, 0.5 = random guessing.
- Compares models without needing a fixed threshold.

# Results :



**Logistic Regression**

**Random Forest**

**Support Vector Machine**

**XGBoost**

**LightGBM**

**F1 Score**

**0.64**

0.59

**0.64**

0.59

0.58

**AUC**

**0.86**

**0.86**

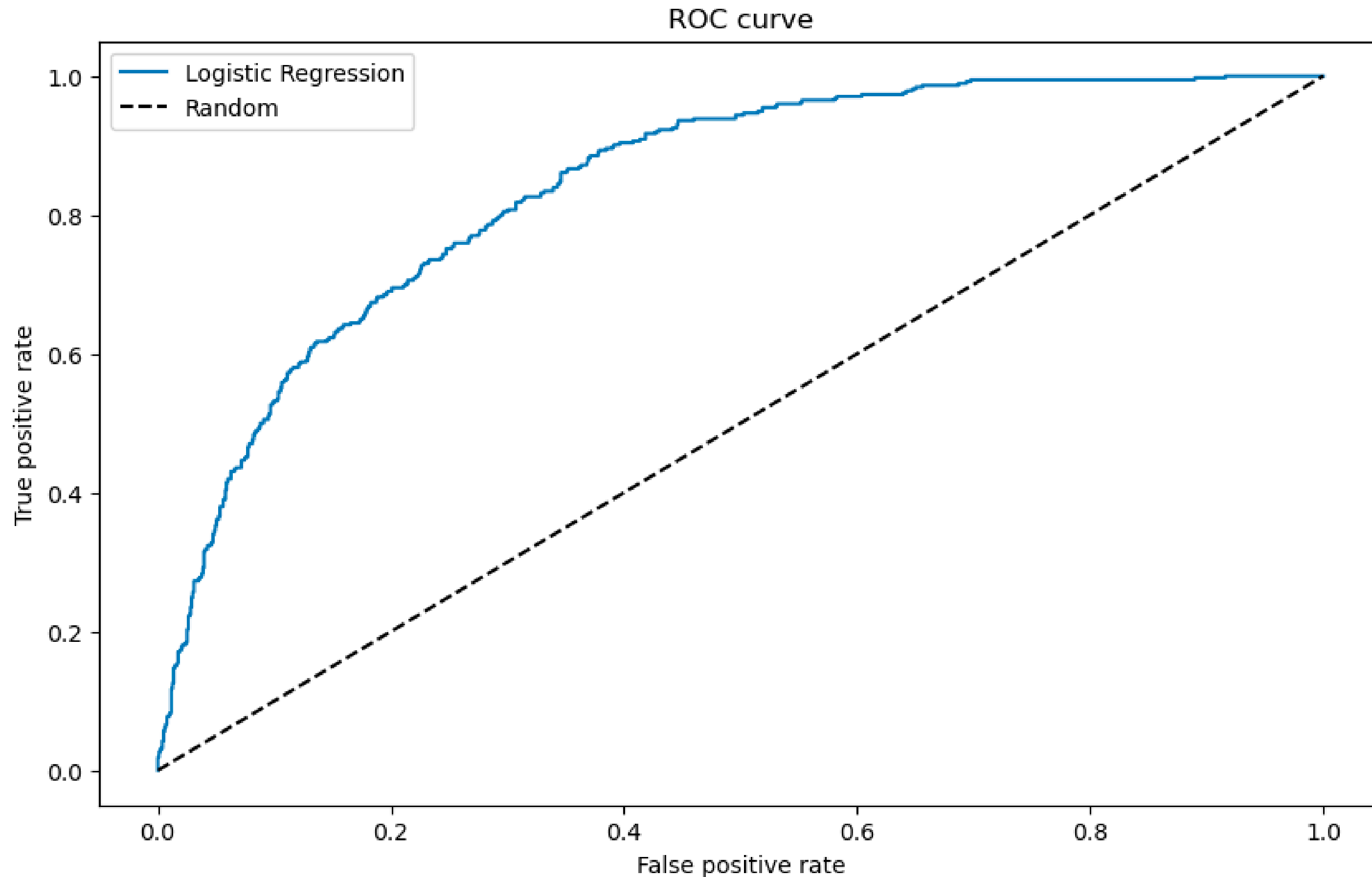
0.85

**0.86**

**0.86**

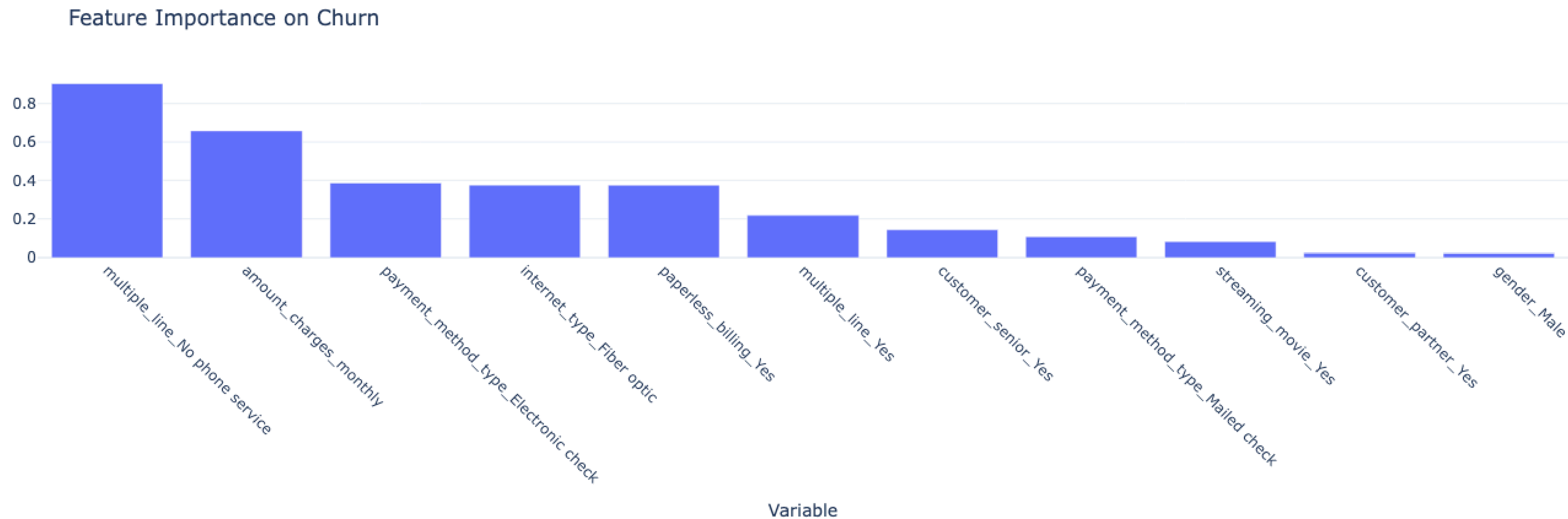
**Logistic  
Regression got  
the best results !**

# ROC curve for Logistic Regression :



# **4. Results and recommendations**

# Feature importance on Churn





# Most important features

**No phone service**

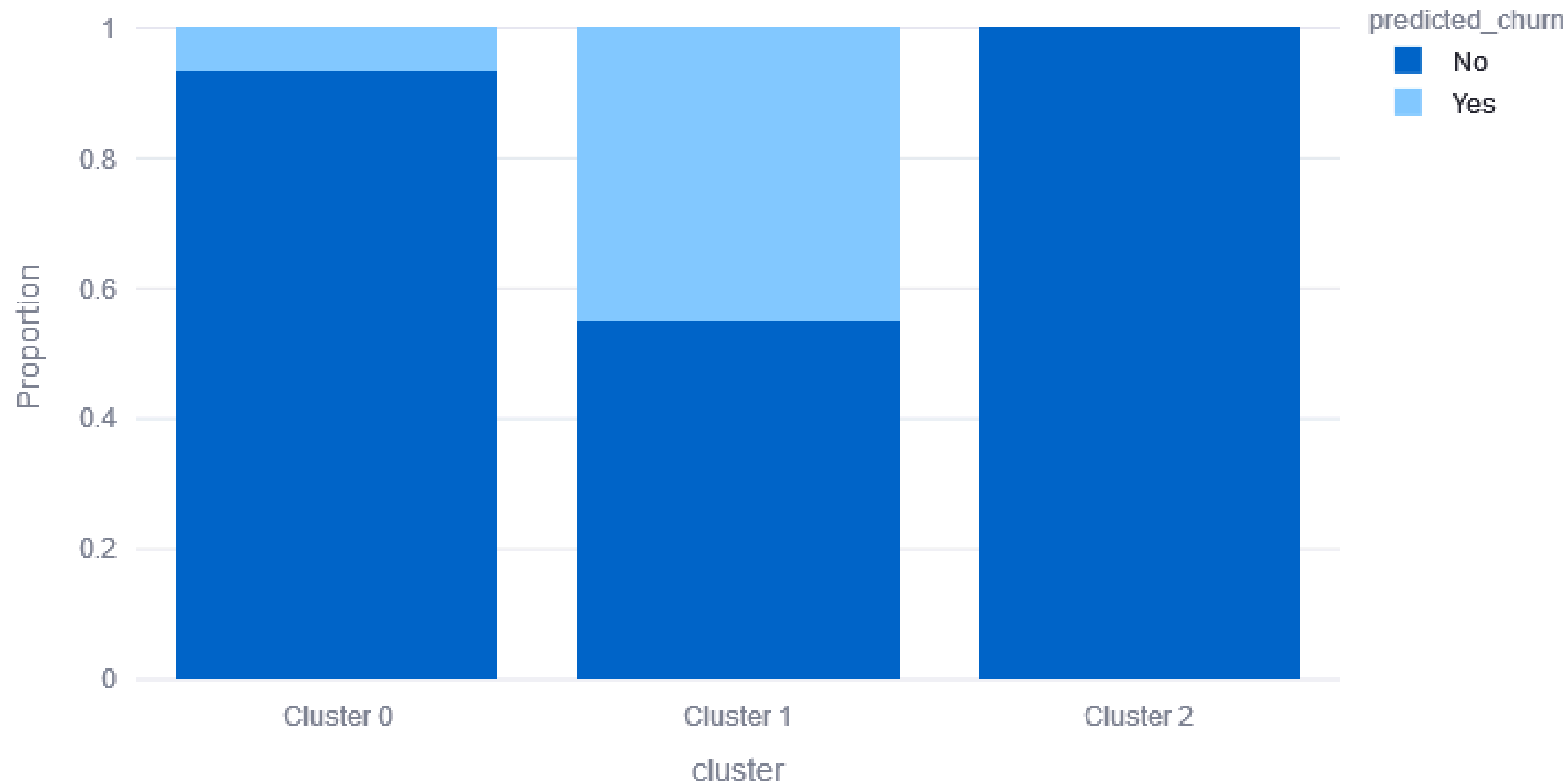
**Amount charges  
monthly**

**Electric check**

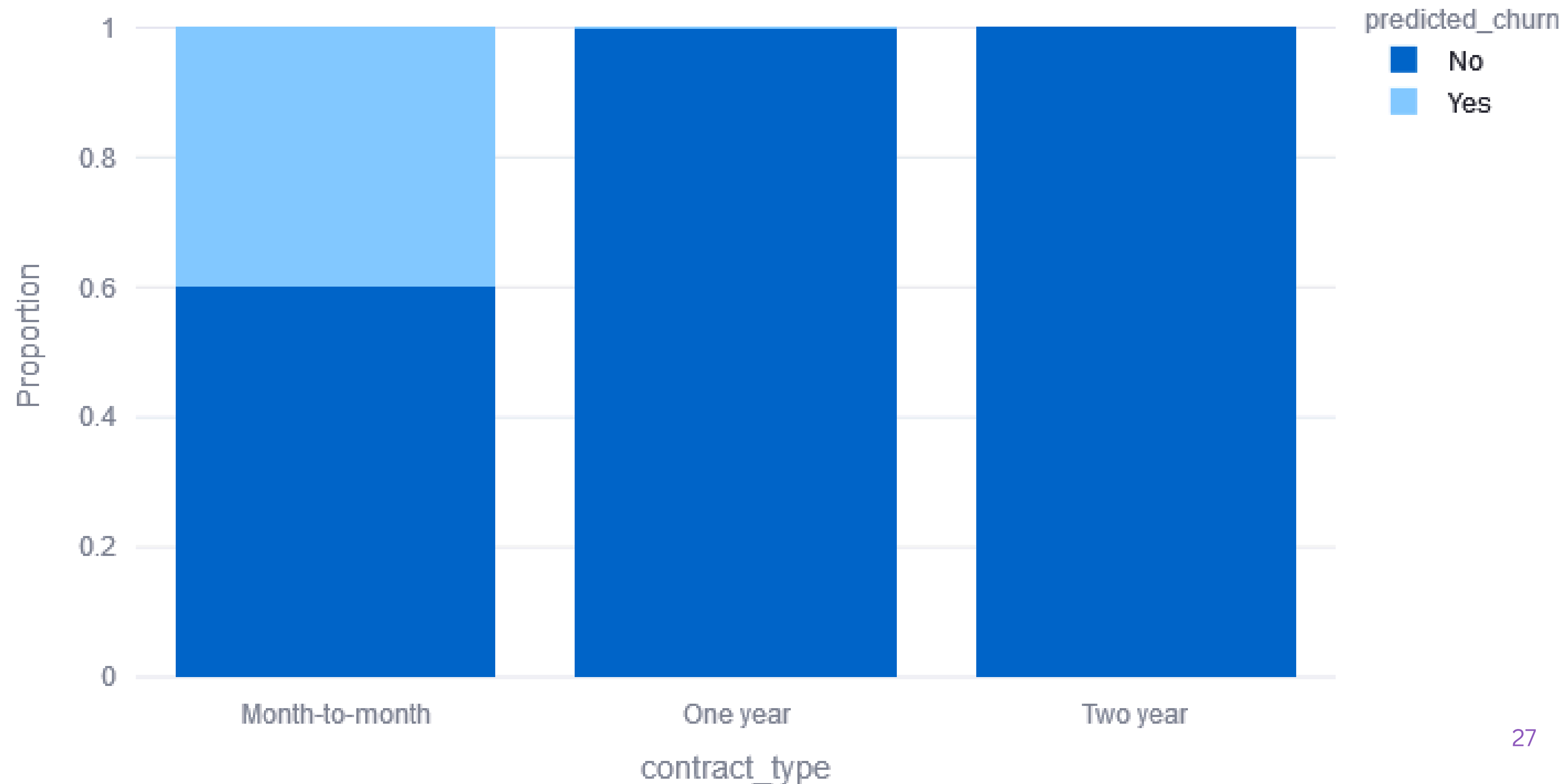
**Fiber optic**

**Paperless billing**

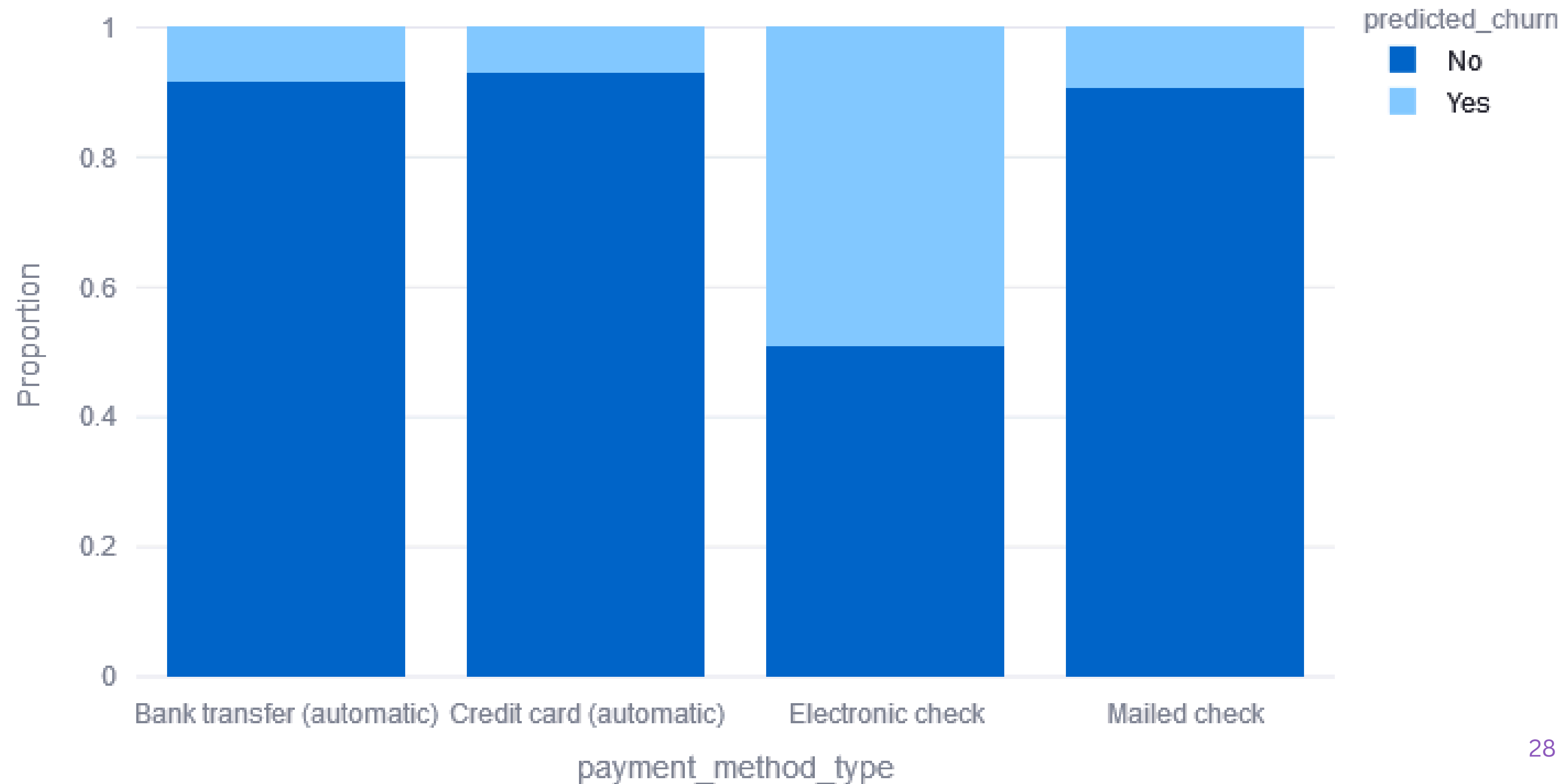
# Predicted Churn rate by cluster



# Predicted Churn rate by contract type



# Predicted Churn rate by Payment method



## Cluster 0: Loyal, High-Value Customers

**Enhance Loyalty Programs:** Offer exclusive deals and recognition

**Cross-Selling Opportunities:**  
Introduce new premium services

**Maintain Service Excellence:** Ensure high-quality customer service

**Solicit Feedback:** Engage in feedback programs



## Cluster 2: Phone Service Only Customers

**Upselling Internet Services:** Introduce attractive bundled packages

**Loyalty Programs:** Implement rewards for continued patronage.

**Promote Convenience Features:**  
Highlight benefits of paperless billing.



## Cluster 1: Newer, Price-sensitive customers

**Retention Efforts:** Implement retention campaigns focusing on satisfaction

**Promote Long-Term Contracts:** Offer incentives for longer-term contracts.

**Upsell Add-on Services:** Educate on the value of additional services.

**Personalized Communication:**  
Use targeted messaging

