

**Data Science Weekend 2023**  
**Student & Junior Pro**

# **Advanced Churn Analytics for Enhanced Retention Strategies in Telecommunications**

Lashani Team

# Table of Contents

---

Data Science Weekend 2023

- 1 Our Team
- 2 Problem
- 3 Proposed Solution
- 4 Methodology
- 5 Results
- 6 Recommendation
- 7 Conclusion



# Our Team

---

Data Science Weekend 2023



## **Mushab Tinumbang** Project Lead

Digital Business, Universitas Padjadjaran  
Data Scientist Intern at Pupuk Indonesia Holding Company



## **Gilang Setyawan Y. P.** Data Analyst

Computer Engineering, Universitas Indonesia  
Data Engineer Intern at Akasha Wira International



## **Miranty Anjani Putri** Data Scientist

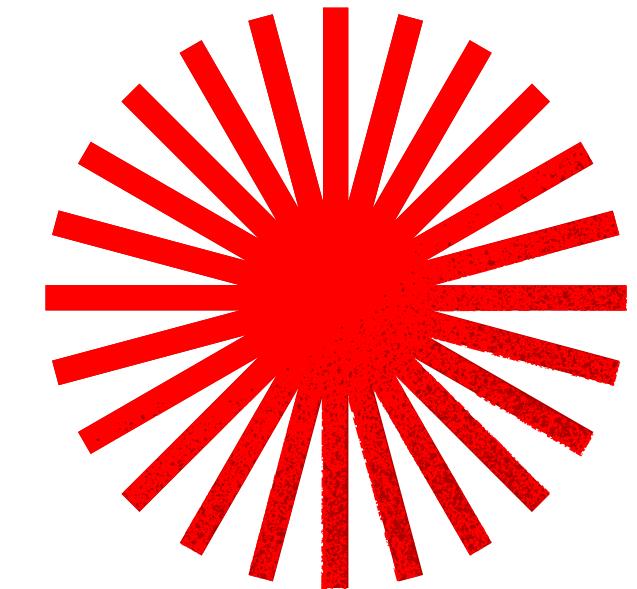
Computer Engineering, Universitas Indonesia  
Data Scientist Intern at Pupuk Indonesia Holding Company

# Problem

---

Data Science Weekend 2023

In the dynamic landscape of the telecommunications industry, **customer churn** is a persistent challenge affecting service providers. High customer turnover can lead to **revenue loss** and **diminished market share**. Understanding the factors influencing churn is crucial for devising effective retention strategies. Additionally, **comprehensive data-driven analytics** such as customer segmentation and churn prediction are **necessary** to allow the ability to tailor solutions to specific customer needs.



# Solution

---

Data Science Weekend 2023



This project aims to implement a robust **churn analytics framework** leveraging **machine learning** and data visualizations. By employing clustering techniques, we can identify distinct customer segments based on patterns and preferences. Simultaneously, the project will develop predictive models to forecast individual churn and estimate CLTV. A **customer dashboard** will also be created for real-time monitoring and decision-making. Insights gained from these analytics will guide personalized retention strategies. The goal is to **reduce churn, improve satisfaction, and optimize market position**.

# Methodology

---

Data Science Weekend 2023

## 1 Data Collection & Preprocessing

Gather **comprehensive customer data**, including customer patterns, demographic information, and historical interactions. **Standardize numerical features** and **encode categorical variables** for machine learning compatibility.

## 2 EDA & Feature Selection

Conduct **Exploratory Data Analysis** to gain insights into the distribution of features, identify correlations, and understand patterns that may influence churn. Utilize statistical tests and domain knowledge to select relevant features for clustering and predictive modeling.

## 3 Machine Learning Modeling

Apply clustering algorithms (e.g., **K-Means**) to group customers based on similar characteristics. Develop predictive models, such as **classification** and **regression** algorithms, to forecast individual churn label and estimate CLTV.

## 4 Insights Generation

Analyze results to extract impactful and actionable insights. **Utilize insights to formulate personalized retention strategies** for each identified customer segment. Develop a comprehensive customer dashboard with actionable visualizations.

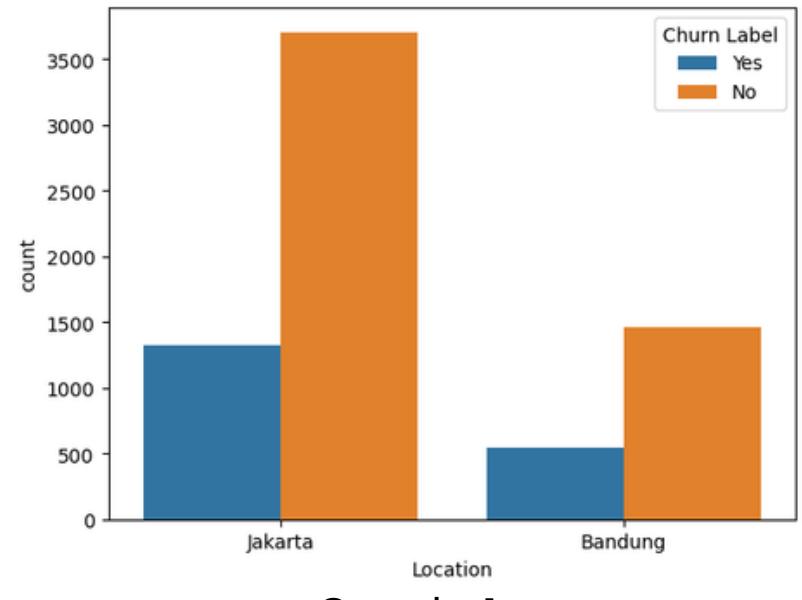
## 5 Deployment

Deploy visualization using dashboard, analytics, and machine learning models using **PowerBI** and **Streamlit**.

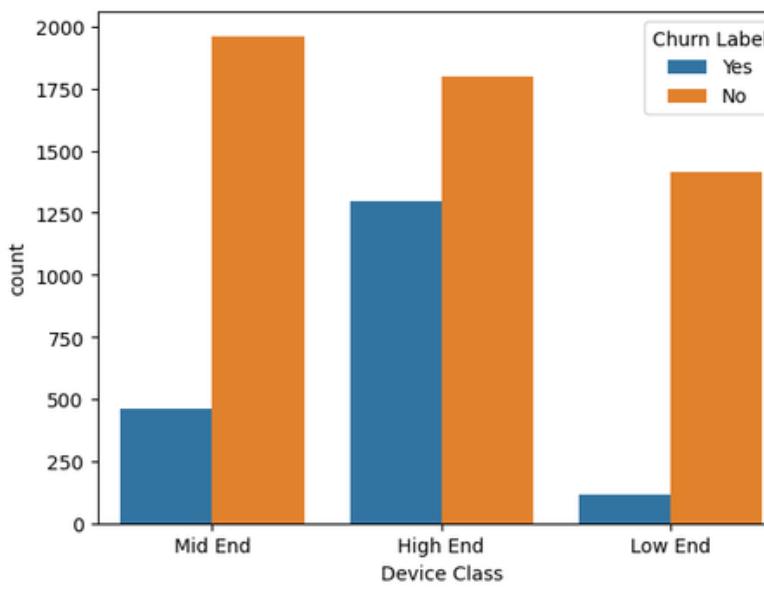
# **Results**

# Explanatory Data Analysis

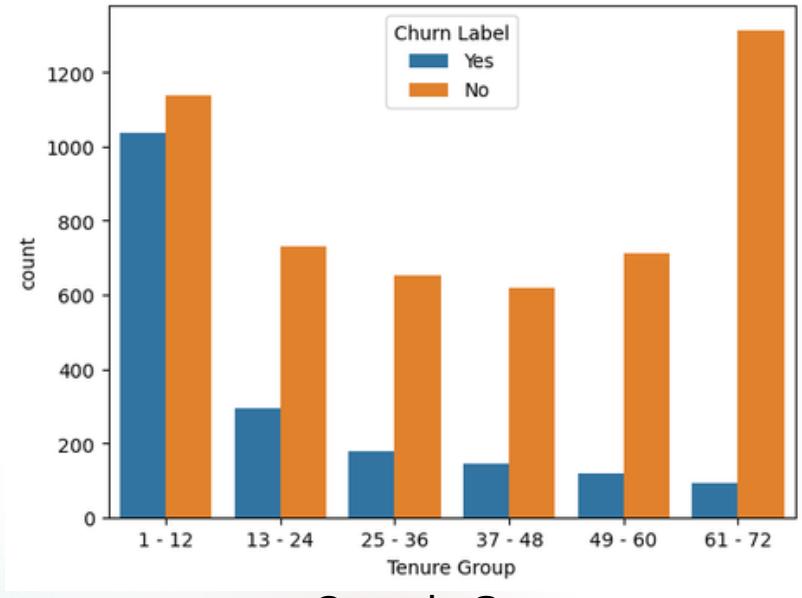
Data Science Weekend 2023



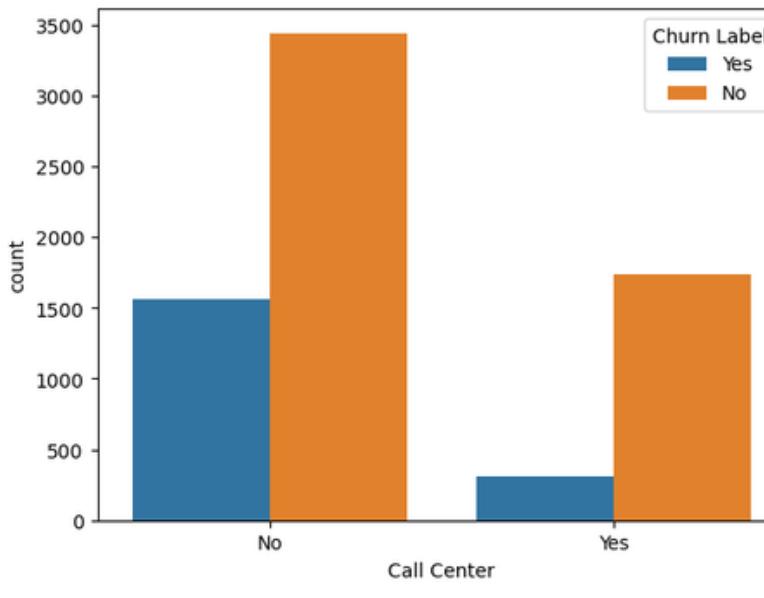
Graph 1



Graph 2



Graph 3



Graph 4

Through EDA, our team discovered several findings, such as:

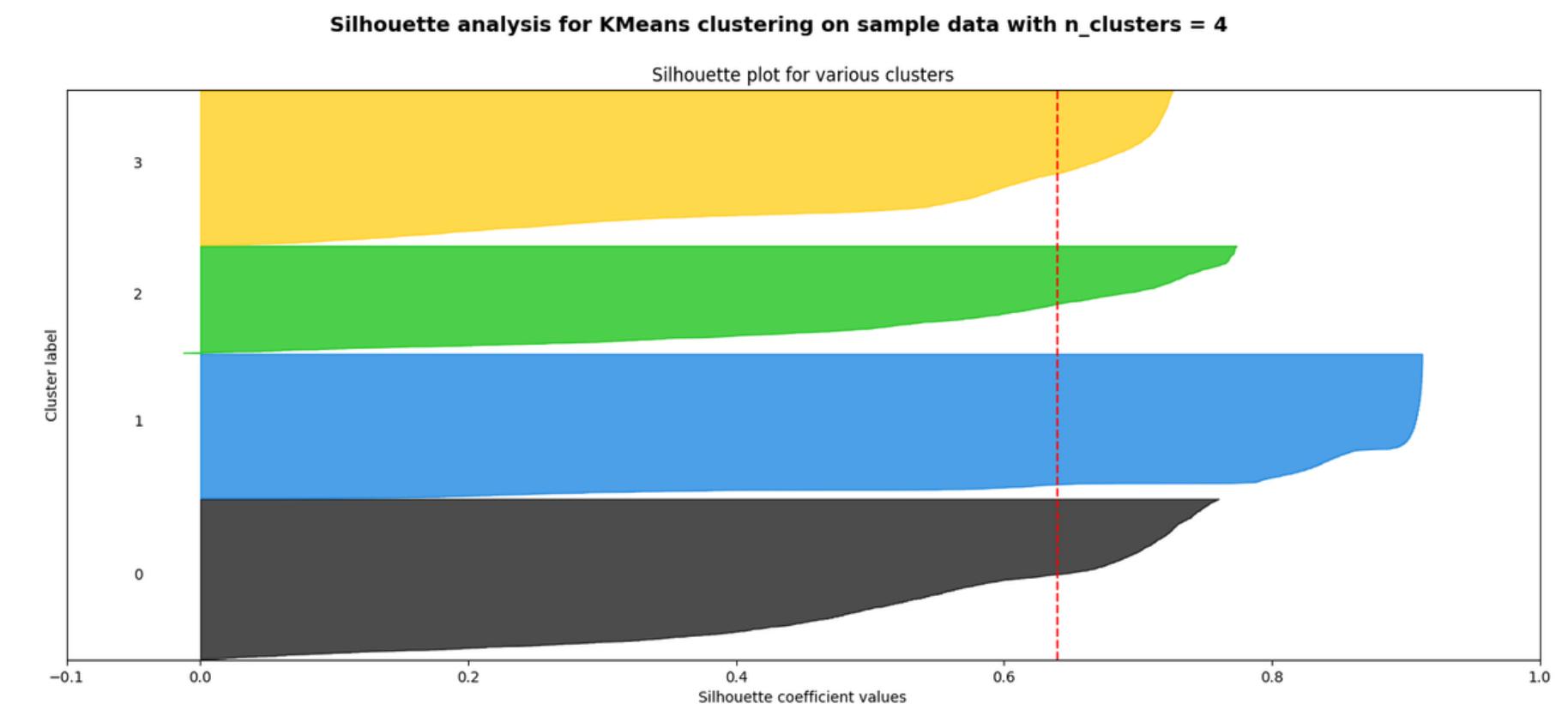
1. **Jakarta** has **more active customers** than Bandung.
2. **Customers with Low-End Devices** have the **lowest amount of Churn customers**. However, the **most active customers** come from those using **Mid-End Devices**.
3. Those who have been customers of the Telecommunication company for **61-72 months** have the **most active customers**, followed by the first 1-12 months. However, for **groups 1-12 months**, they also have the **highest Churn customers**.
4. The customers who **reached out to Call Centers** have **the least Churn customers**.

# Customer Segmentation

---

Data Science Weekend 2023

In this customer segmentation analysis, we utilized **K-Means** clustering to categorize customers. This customer segmentation analysis involved the preprocessing of features, standardizing numerical attributes, and determining the optimal number of clusters through the Elbow Method and Silhouette Score. We proceed to use **n\_clusters = 4** with a **silhouette score of 0.63** and **better distribution**.



# Customer Segmentation

---

Data Science Weekend 2023

The segmentation result reveals **four distinct customer clusters** (labeled 0 to 3), each characterized by specific behavior patterns; each of them is:

1. Cluster 0 (**High End, High Purchase, Moderate Product Usage**)
2. Cluster 1 (**Low End, Low Purchase, No Service**)
3. Cluster 2 (**Mid End, Moderate Purchase, Low Product Usage**)
4. Cluster 3 (**Mid End, Moderate Purchase, Moderate Product Usage**)

**Cluster 0 has the highest churn label count at 1297**, followed by Cluster 2 with 345 churns, Cluster 3 with 114 churns, and Cluster 1 with 113 churns.

Cluster	Games Product	Music Product	Education Product	Call Center	Video Product	Use MyApp	Monthly Purchase Level	Device	Total Churn Labels
0	Low Usage	Low Usage	Low Usage	Low Usage	Moderate Usage	Moderate Usage	High	High End	1297
1	No Service	No Service	No Service	No Tendency	No Service	No Service	Low	Low End	113
2	Low Usage	Low Usage	Low Usage	No Tendency	Low Usage	Low Usage	Moderate	Mid End	345
3	Moderate Usage	Moderate Usage	Moderate Usage	Using	Moderate Usage	Moderate Usage	Moderate	Mid End	114

# Customer Journey

Data Science Weekend 2023

For Customer Journey, our team proposes that the Company **track their customer's journey**, from their first interaction with them to the most recent one. It will track **each customer and the stages they go through monthly**. The stages are:

- Awareness,
- Research,
- Consideration,
- Purchase, and
- Support.



Due to the lack of data, the insight we can use is only a **Sankey Diagram** to display **the sum of Customers for each month**, from the first month to the current month (month 72). The last stage, or node, is divided into two stages (Month 72 Churn or Month 72 Active). As we can see, the **Telecommunication Company experiences growth in the number of customers**.

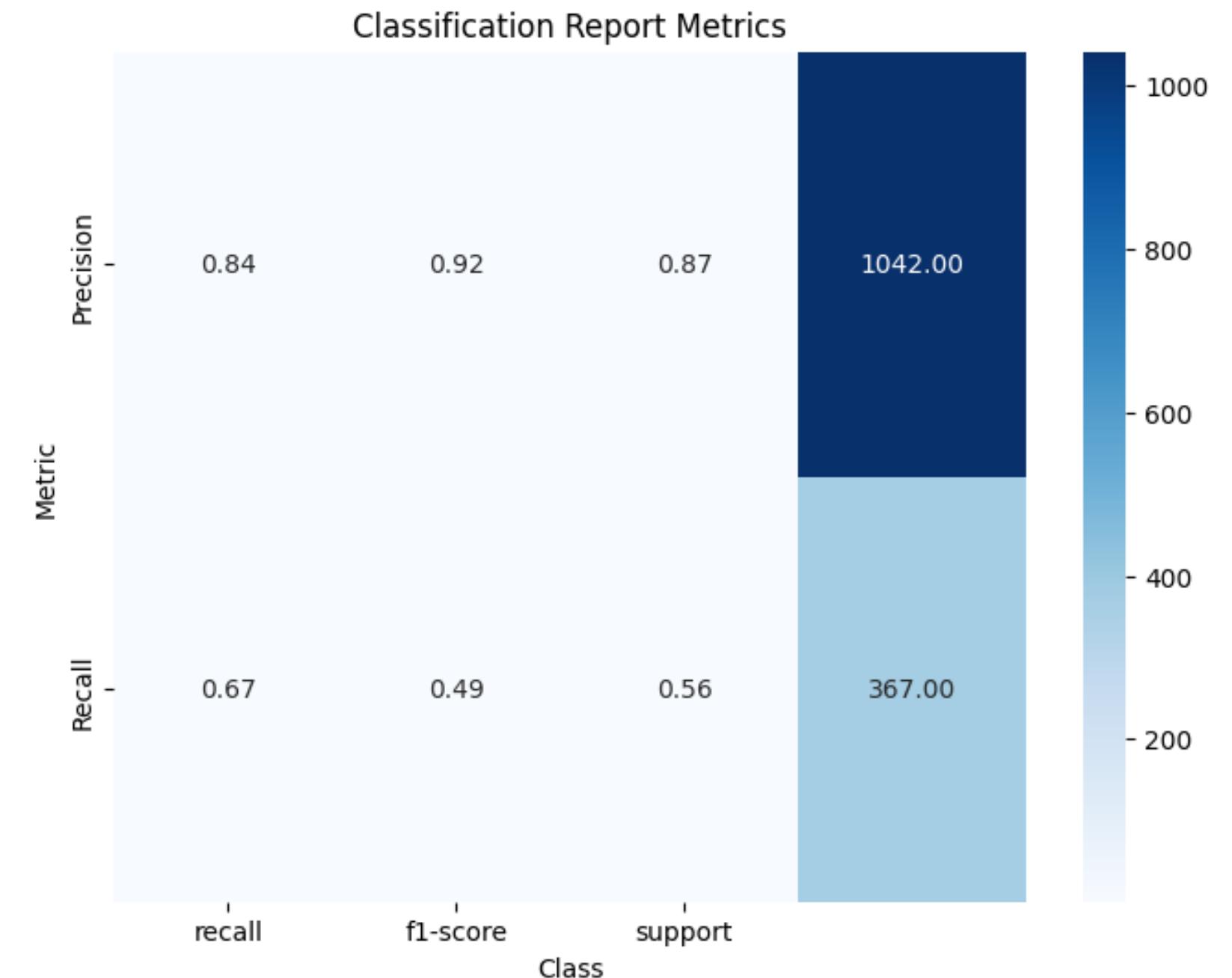
Reference: [Forbes](#)

# Churn Prediction

---

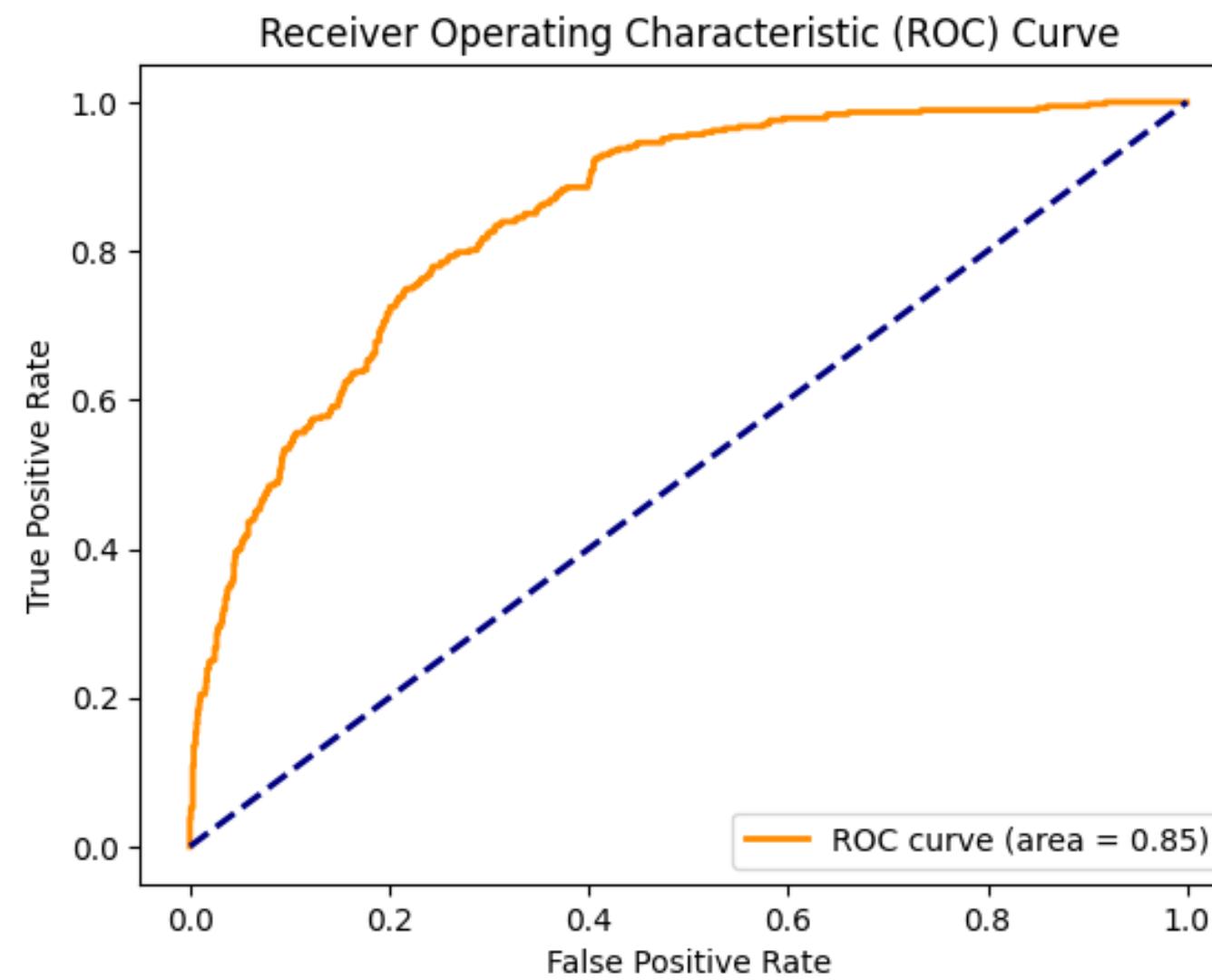
Data Science Weekend 2023

To create a **churn prediction model**, we initiated the process by preprocessing the dataset, excluding irrelevant features, and subsequently splitting it into training and testing sets with an 80-20 ratio. Our logistic regression model achieved **80% accuracy** in predicting churn on a dataset of 1409 instances. The classification report highlights good precision for no churn (class 0) at 84%, with somewhat lower recall for churn (class 1) at 49%. The model demonstrates balanced performance, with a satisfactory weighted **average F1-score of 79%**.



# Churn Prediction

Data Science Weekend 2023



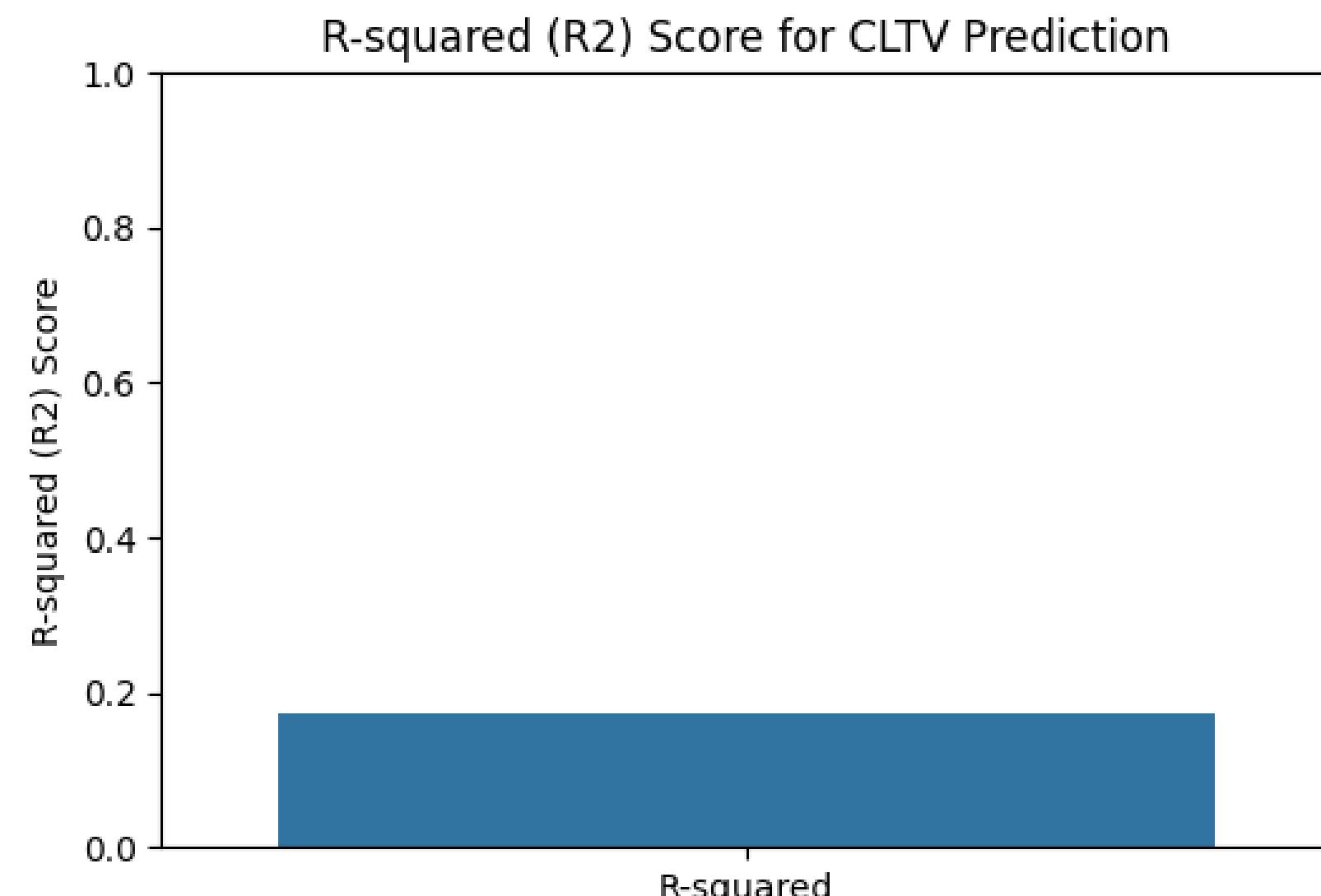
The confusion matrix reveals that our model correctly identified 954 instances of no-churn, but misclassified 88 instances as churn and 188 instances as no churn. On the positive side, it accurately predicted 179 instances of churn. The **ROC AUC score of 0.85** indicates **strong discriminatory power**, affirming the model's proficiency in distinguishing between churn and no-churn instances, making it a **promising tool for predicting customer churn**.

# CLTV Model

---

Data Science Weekend 2023

In this analysis, we used a **Linear Regression** model to predict **Customer Lifetime Value (CLTV)**. The dataset was split into training and test sets, and the model was trained on the former to predict CLTV values. Key performance metrics, including Mean Absolute Error (MAE) and R-squared (R<sup>2</sup>) Score, were then calculated to assess the model's accuracy and explanatory power. The dataset's **standard deviation of 1537.87** suggests significant variability in CLTV. The **Mean Absolute Error of 1195.41** and the **R-squared (R<sup>2</sup>) score is 0.17**.



# Dashboard

---

Data Science Weekend 2023

Our first dashboard is centered around the **distribution of customer locations** and the **types of devices** they use. On the map, you can see how our customers are spread across different cities. Additionally, we provide information on the usage of specific device classes by customers. This graph helps us understand **customer preferences** for the types of devices they use. We also display the number of customers in each region using specific device classes.

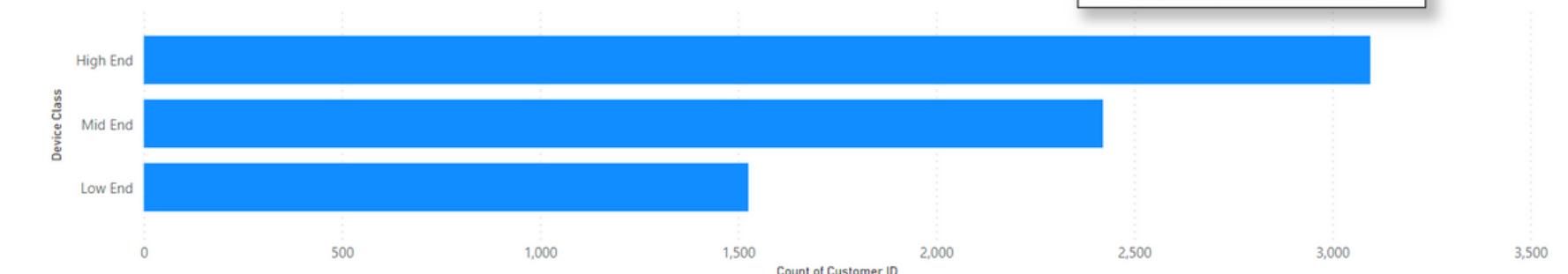
Dashboard Distribution Of Customer Location and Device Class

Distribution Of Customer Location

Location ● Bandung ● Jakarta



Device Class of customer



7043  
Count of Customer

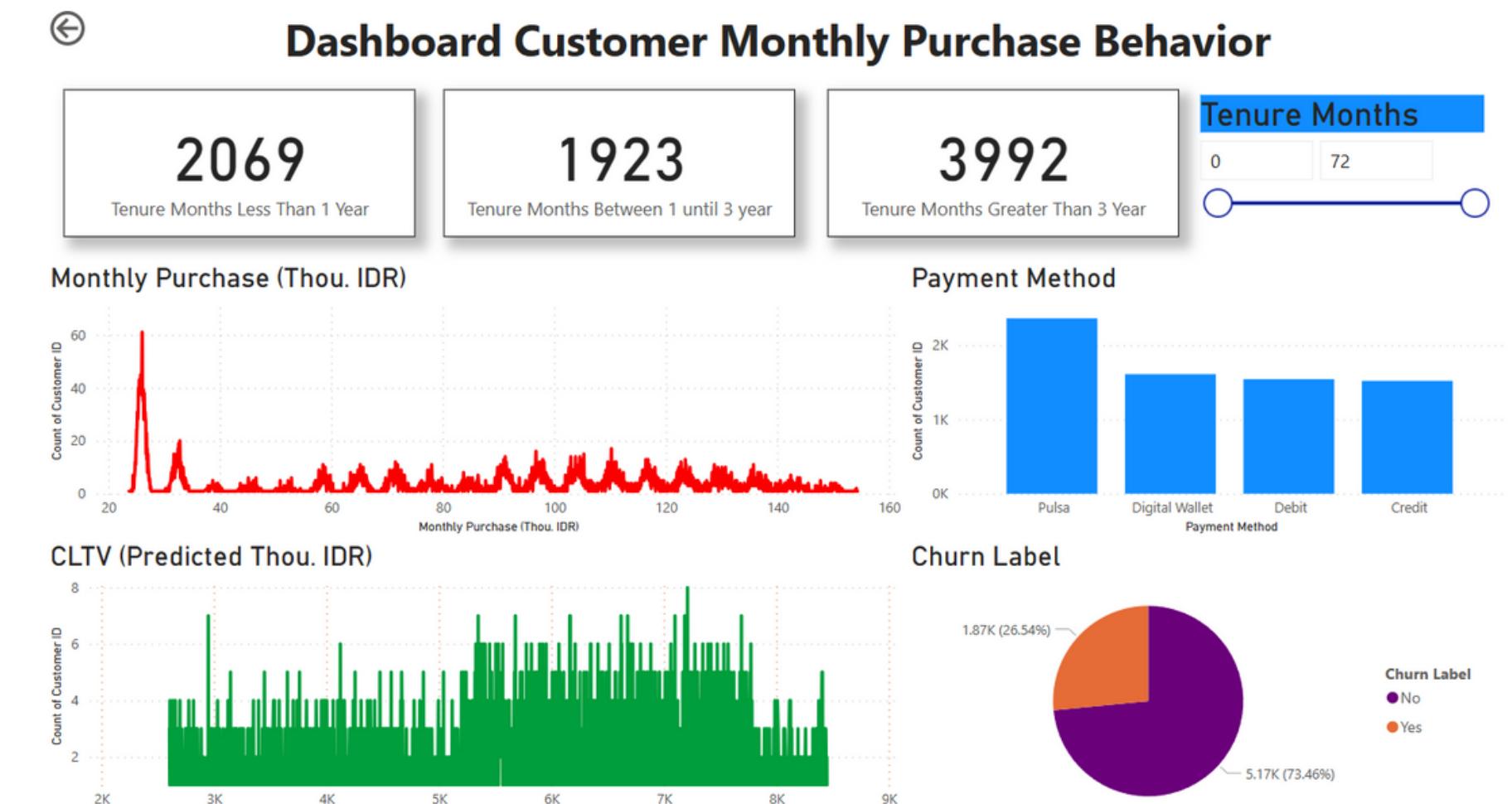
2012  
Count of Customer in Bandung

5031  
Count of Customer in Jakarta

# Dashboard

Data Science Weekend 2023

The second dashboard is **centered around customer monthly purchases**. We display a graph of monthly purchases, the Customer Lifetime Value (CLTV), the most commonly used payment methods, churn labels (subscription termination), and the number of customers who have subscribed for a specific time range. This helps us observe **customer purchase patterns**, assess the **value of customers** to the company, and understand **why** some customers unsubscribe.

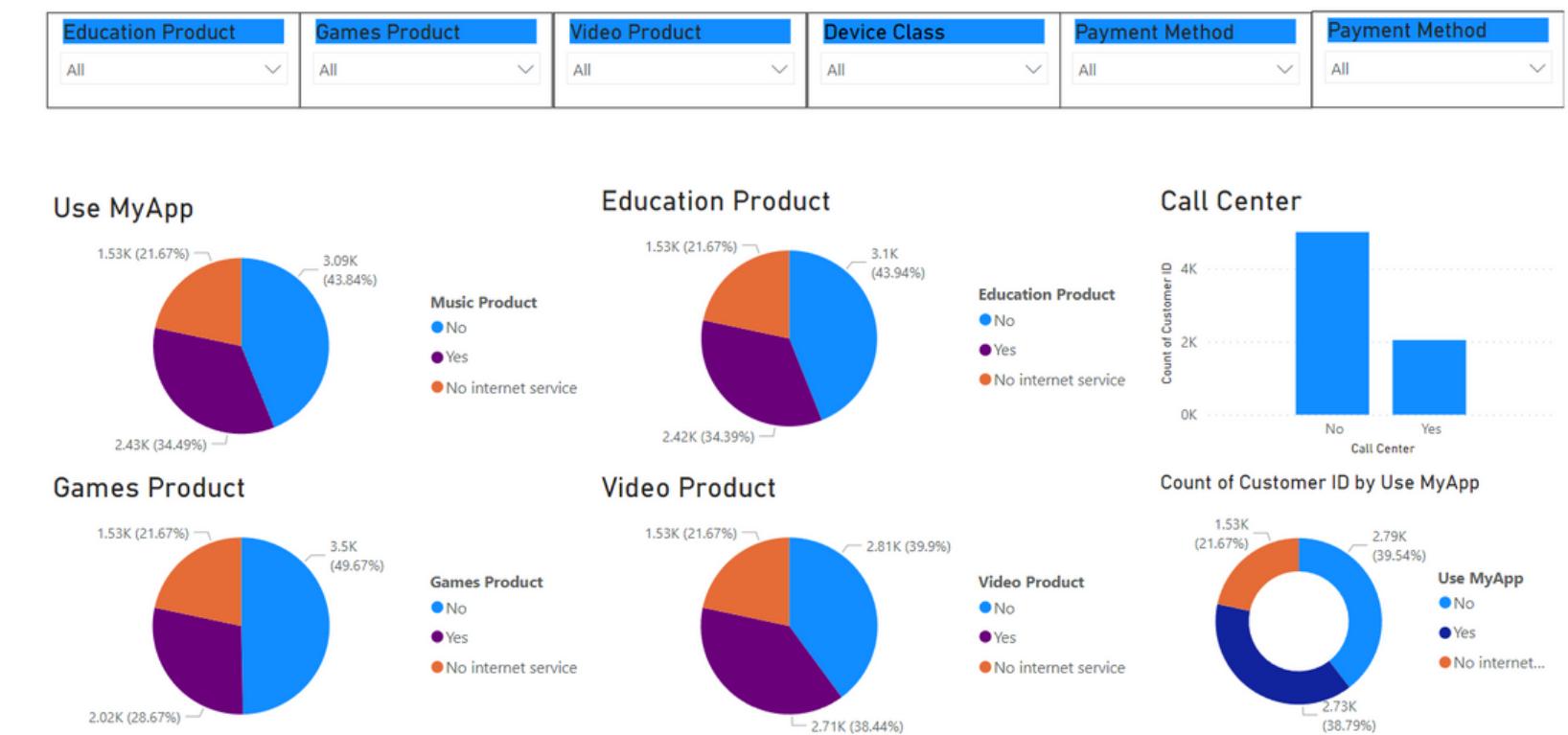


# Dashboard

Data Science Weekend 2023

The third dashboard is about **product orders**. We present a pie chart showing the **quantity of products ordered** by customers. Additionally, we show how many customers placed orders through MyApp and provide a graph for orders through the call center. You can also use filters to **adjust the quantity displayed** in this dashboard.

Dashboard Product Order By Customer

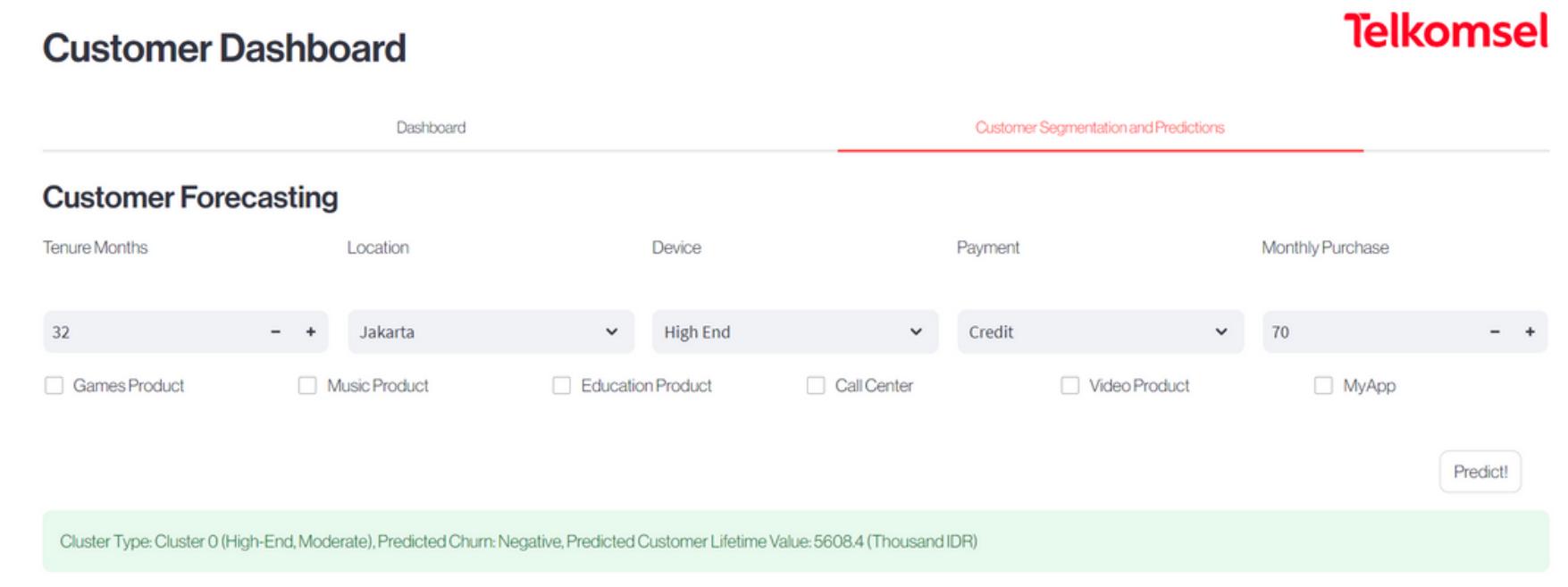


# Deployment

---

Data Science Weekend 2023

We are using **Streamlit** to deploy our project. Users can seamlessly interact with an actionable **PowerBI dashboard**, gaining real-time access to dynamic visualizations. The application empowers users to **predict customer segmentation, churn labels, and Customer Lifetime Value (CLTV)** by inputting relevant customer data. Furthermore, the app provides insightful results regarding customer segmentations, showcasing the outcomes of the analytics. Lastly, we deployed the Streamlit app on **GitHub**.



# Deployment

---

Data Science Weekend 2023

## Project Links:

Streamlit: <https://dsw-2023-lashani.streamlit.app/>

PowerBI: <https://bit.ly/PowerBILashani>

Github: <https://github.com/LASHANI-DSW23/LASHANI-Data-Challenge>

# **Solution / Recommendation**

# Retention Strategy

---

Data Science Weekend 2023

## **Cluster 0** (High End, High Purchase, Moderate Product Usage):

- Engagement Campaigns:
  - Develop targeted re-engagement campaigns, especially through the **Video** product and **MyApp**.
  - Offer **exclusive content, features, or discounts** to encourage increased usage.
- Personalization:
  - Leverage data to **personalize recommendations** and content within the games, music, and video products.
  - Tailor **promotions** based on their historical preferences and usage patterns.
- Customer Feedback and Improvement:
  - Proactively seek feedback from Cluster 0 customers to **understand their dissatisfaction**.
  - Address and communicate improvements in response to their concerns.



# Retention Strategy

---

Data Science Weekend 2023

## **Cluster 1** (Low End, Low Purchase, No Service):

- Service Restoration:
  - Investigate and address the **issues causing "No Service"** across all products promptly.
  - Clearly communicate the steps taken to improve service reliability and quality.
- Re-Introduction Campaigns:
  - Develop **targeted campaigns** to re-introduce the value of each product.

## **Cluster 2 & 3** (Mid End, Moderate Purchase, Low & Moderate Product Usage):

- Increase Product Engagement:
  - Launch **targeted marketing campaigns** to promote key product features.
  - Provide incentives or rewards for increased product usage.
- Personalized Loyalty Programs:
  - Develop **loyalty programs** to reward consistent engagement.
  - Provide **personalized recommendations** to enhance their experience.
  - Implement **proactive customer support** to address any concerns promptly.



# Retention Strategy

---

Data Science Weekend 2023

## General Strategies (Across Clusters):

- Enhance Customer Satisfaction:
  - Maintain **consistent high-quality** products or services
- Cross-Product Bundling:
  - **Offer bundled packages or discounts** for using multiple products, encouraging users to explore different offerings.
- Enhance Customer Support:
  - **Strengthen customer support** for all clusters, with a focus on quicker issue resolution.
  - Provide proactive assistance for users facing any challenges or technical issues.
- Regular Communication:
  - **Maintain regular communication** with all customers
  - **Highlight** new features, content, or improvements to maintain user interest



# Analytics

---

Data Science Weekend 2023

- **Informed Decision-Making:**
  - Our **customer dashboard provides real-time insights**, allowing stakeholders to make data-driven decisions.
- **Targeted Marketing:**
  - Our **customer segmentation model enables targeted marketing campaigns**. By understanding different customer segments, marketing efforts can be tailored to meet the specific needs and preferences of each group, improving the effectiveness of campaigns.
- **Churn Prevention:**
  - Our Churn label prediction **identifies customers at risk of leaving**, allowing the company to implement proactive measures to retain them. Early intervention strategies, such as personalized offers or engagement initiatives, can reduce churn rates.



# Conclusion

---

Data Science Weekend 2023

In conclusion, our data-driven analytics approach, featuring a customer segmentation model, customer journey, visualization dashboards, churn prediction, and CLTV forecasting, equips us with the tools needed to make informed decisions, enhance customer experiences, and proactively address challenges. Through personalized marketing strategies, targeted retention efforts, and general strategies, the company could reduce churn, improve customer satisfaction, and optimize its market position, leading to increased revenue and overall company success.

“



**Data Science Weekend 2023**  
**Student & Junior Pro**

# Thank you

---

LASHANI

---