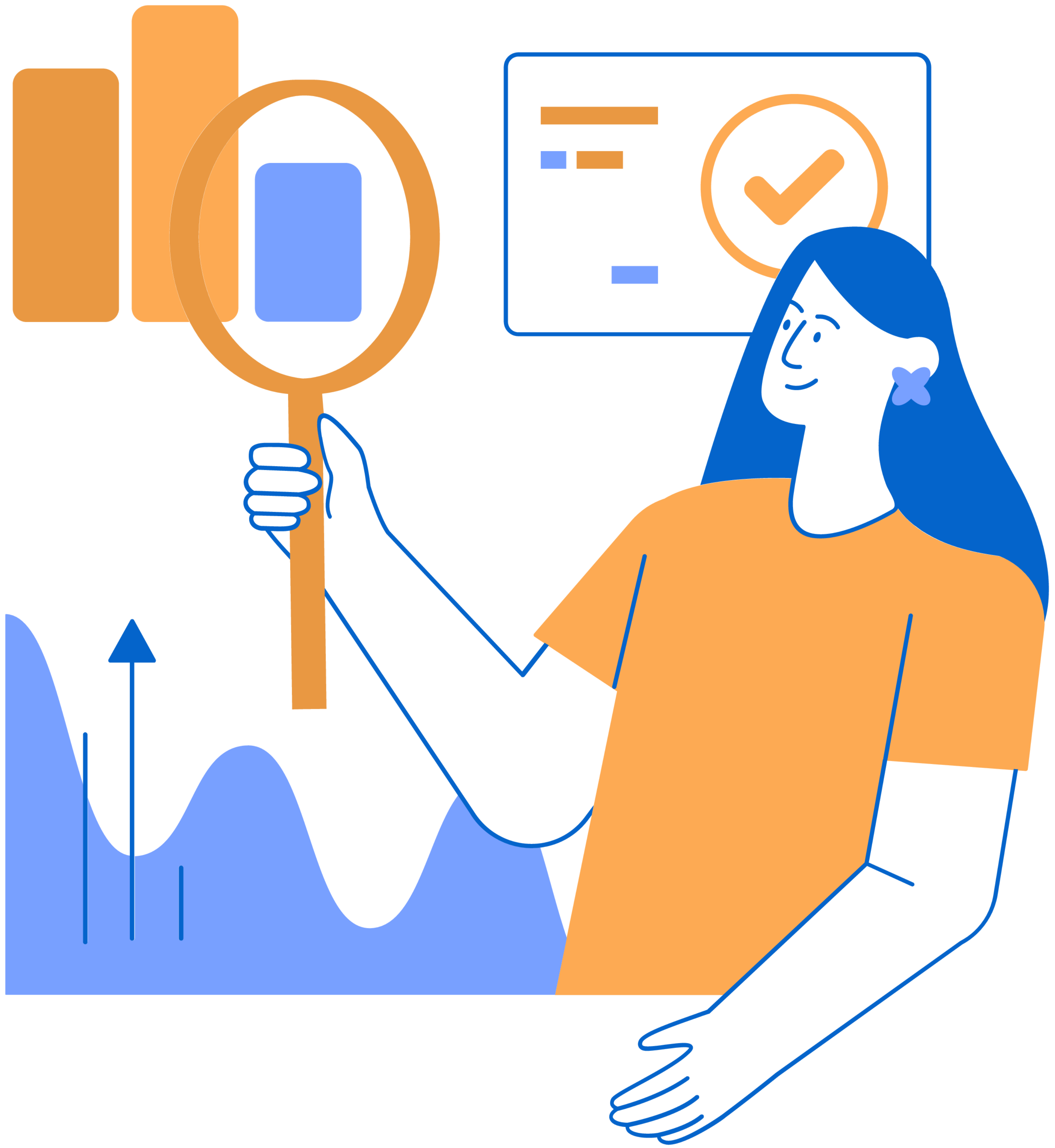
University of Regina

Customer Churn Analysis & Prediction



University of Regina

Customer Churn Analysis & Prediction

Presented By :

- Dev Gohel (200529608),
- Meet Patel (200513407),
- Niraj Asawale (200528919)

Guided By : Timothy D. Oleskiw

Date: April 8, 2025

Introduction & Motivation

Introduction & Motivation

What is churn?

- Churn means customers leaving the telecom service.

Introduction & Motivation

What is churn?

- Churn means customers leaving the telecom service.

Why it matters?

- Losing customers = losing revenue.
- Keeping customers is 5x cheaper than gaining new ones.
- Predicting churn helps keep valuable customers.

Source code: www.google.research.colab.com

Introduction & Motivation

What is churn?

- Churn means customers leaving the telecom service.

Why it matters?

- Losing customers = losing revenue.
- Keeping customers is 5x cheaper than gaining new ones.
- Predicting churn helps keep valuable customers.

Project overview:

- Predict churn using machine learning.
- Analyze customer behavior.
- Build a simple app for real-time predictions.

Source code: www.google.research.colab.com

Project Goals

Source code: www.google.research.colab.com

Project Goals

Primary goal:

- Build a predictive model to identify customers likely to churn.

Project Goals

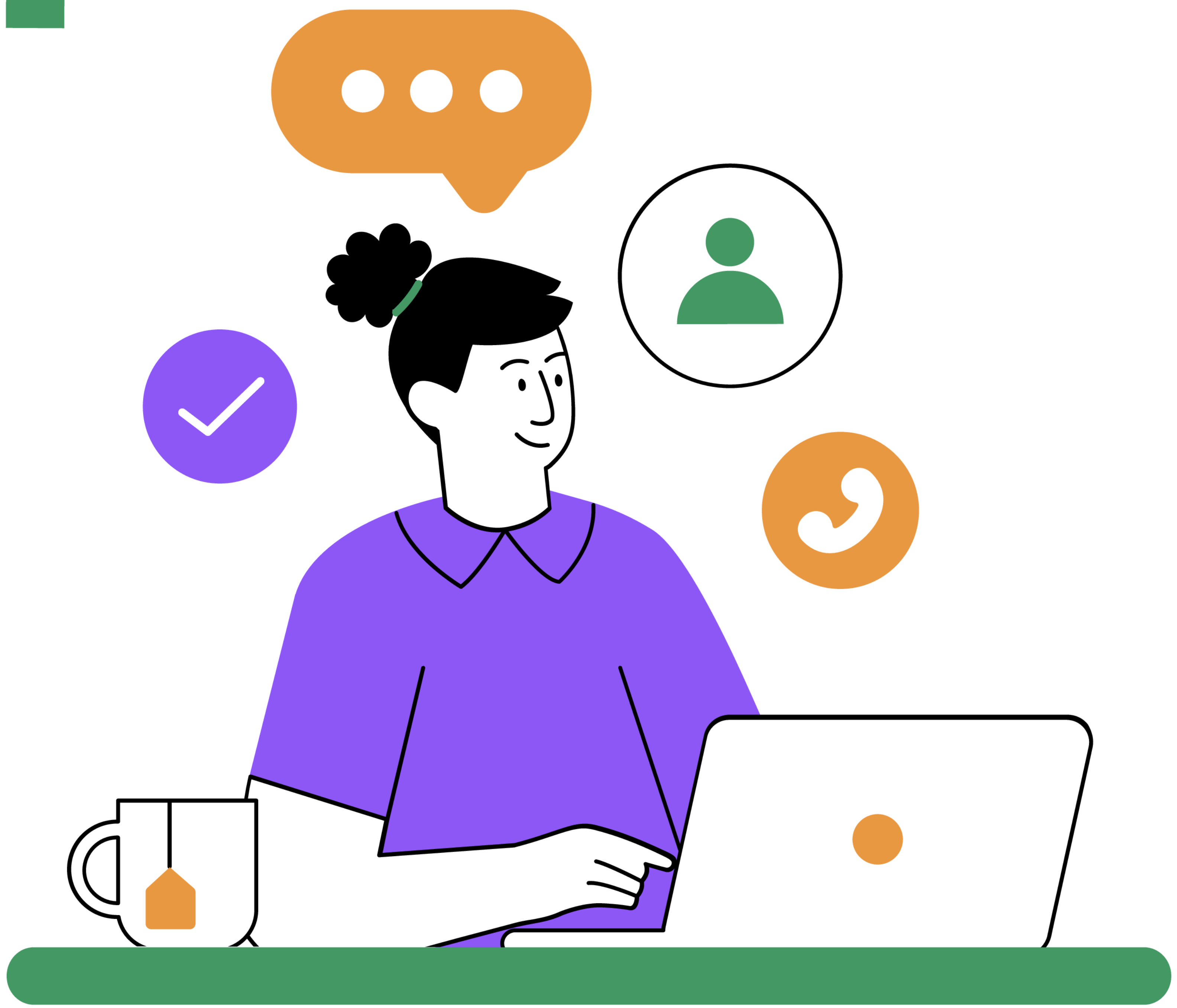
Primary goal:

- Build a predictive model to identify customers likely to churn.

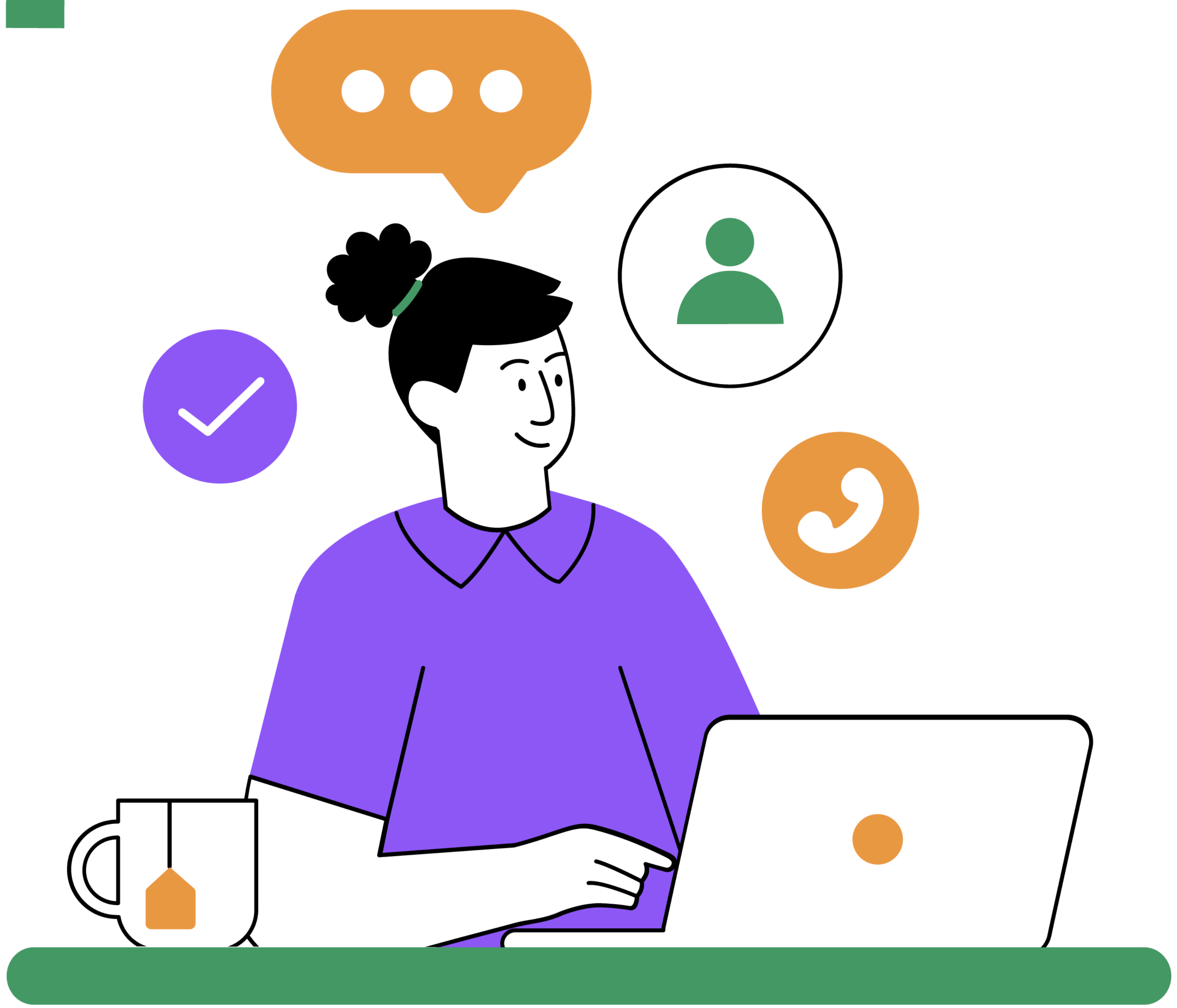
Secondary goal:

- Improve customer retention strategies.
- Provide insights into churn behavior.
- Develop a user-friendly web app for prediction.

Source code: www.google.research.colab.com



Executive Summary

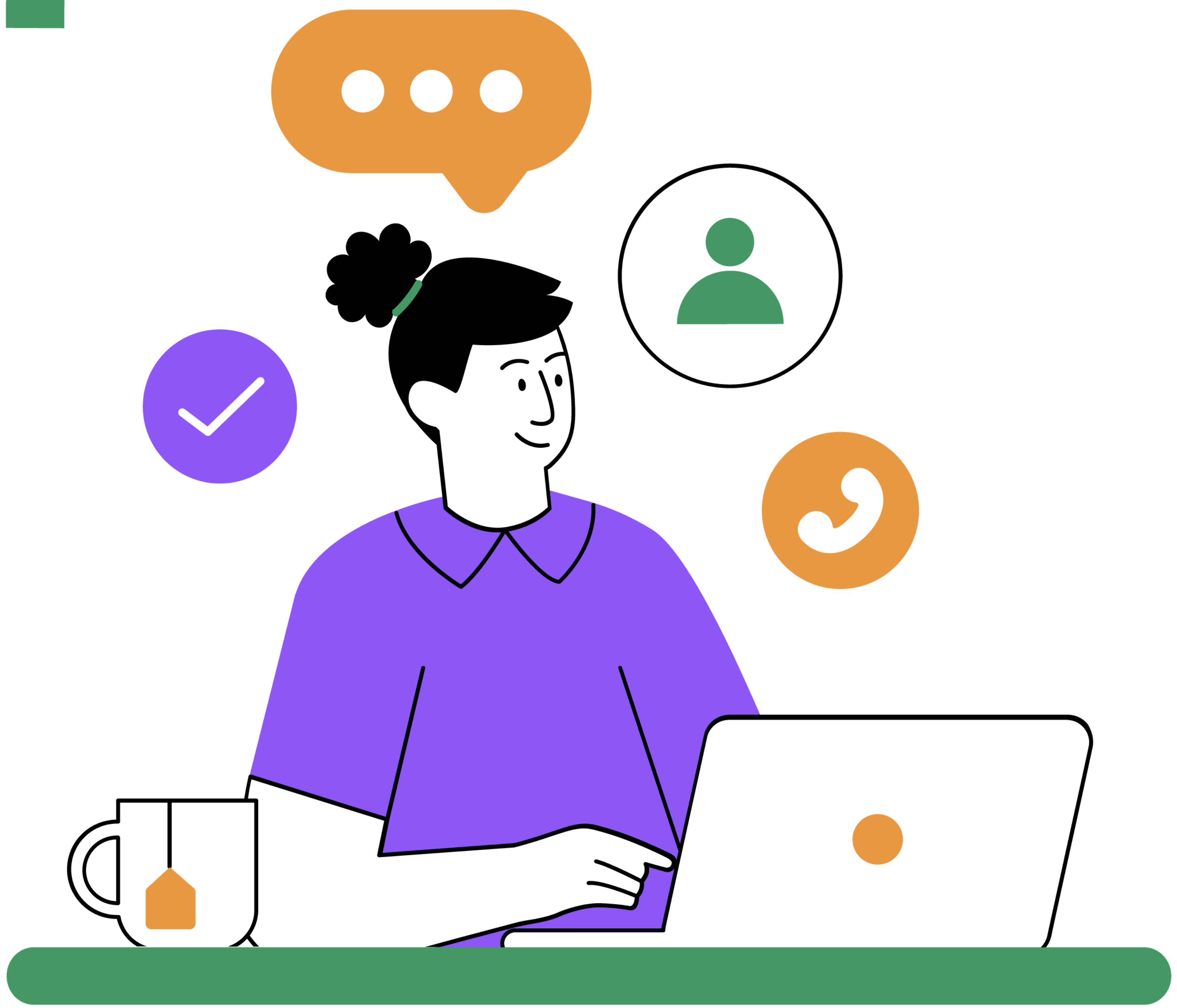


Executive Summary

Dataset used:

- Telco Customer Churn Dataset

Source code: www.google.research.colab.com



Executive Summary

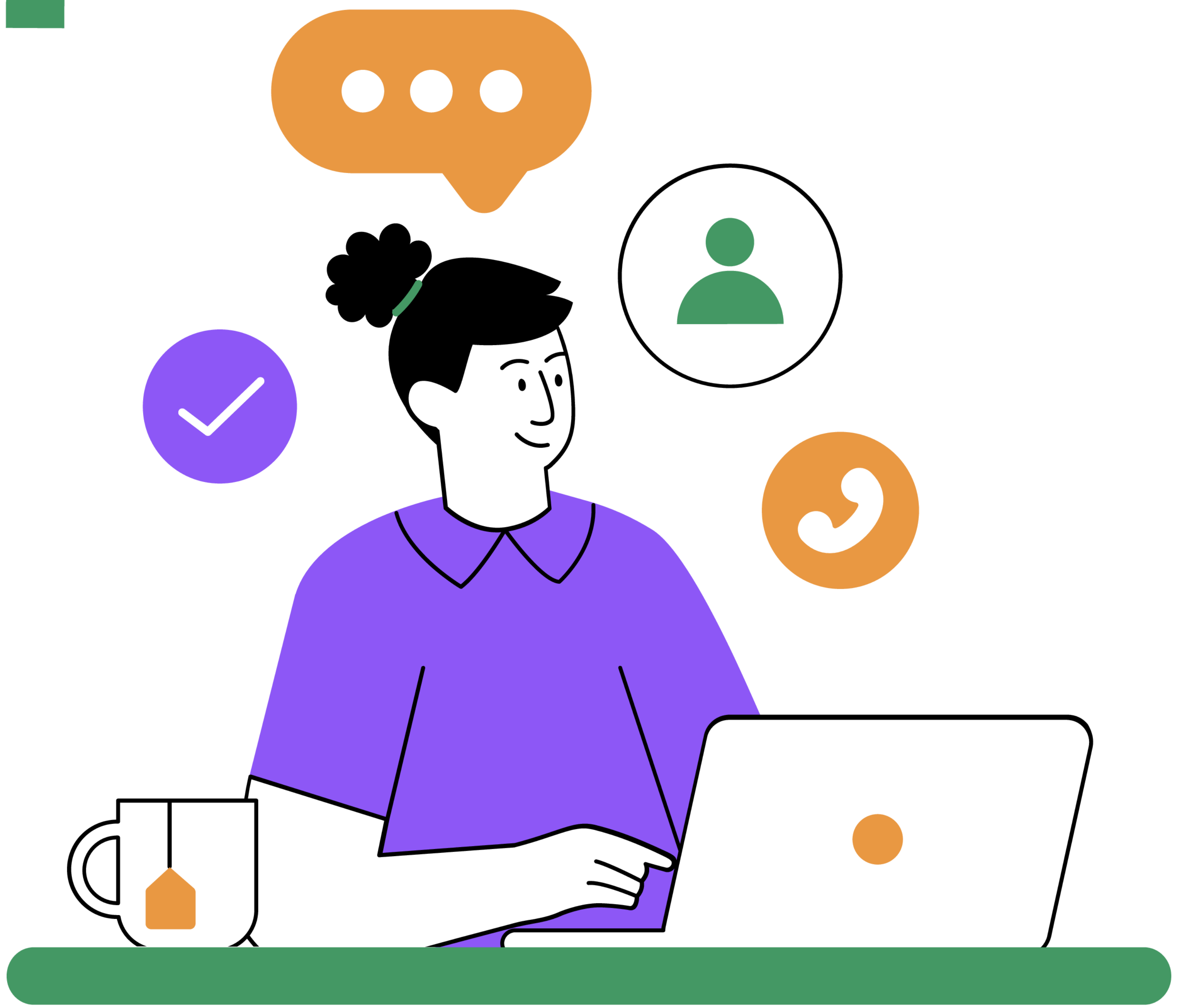
Dataset used:

- Telco Customer Churn Dataset

Tech stack:

- Python
- Pandas
- Scikit-learn
- Flask

Source code: www.google.research.colab.com



Executive Summary

Dataset used:

- Telco Customer Churn Dataset

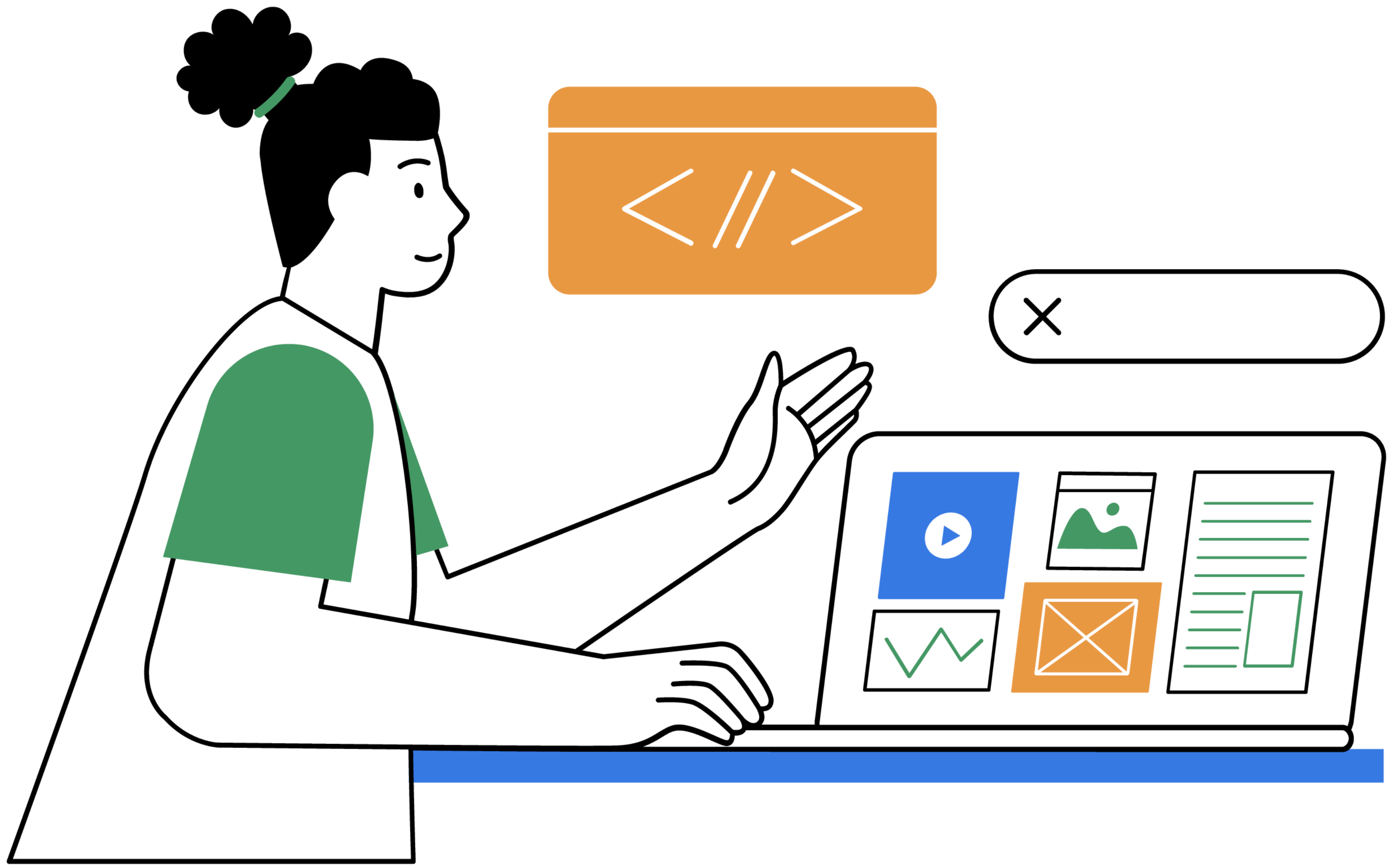
Tech stack:

- Python
- Pandas
- Scikit-learn
- Flask

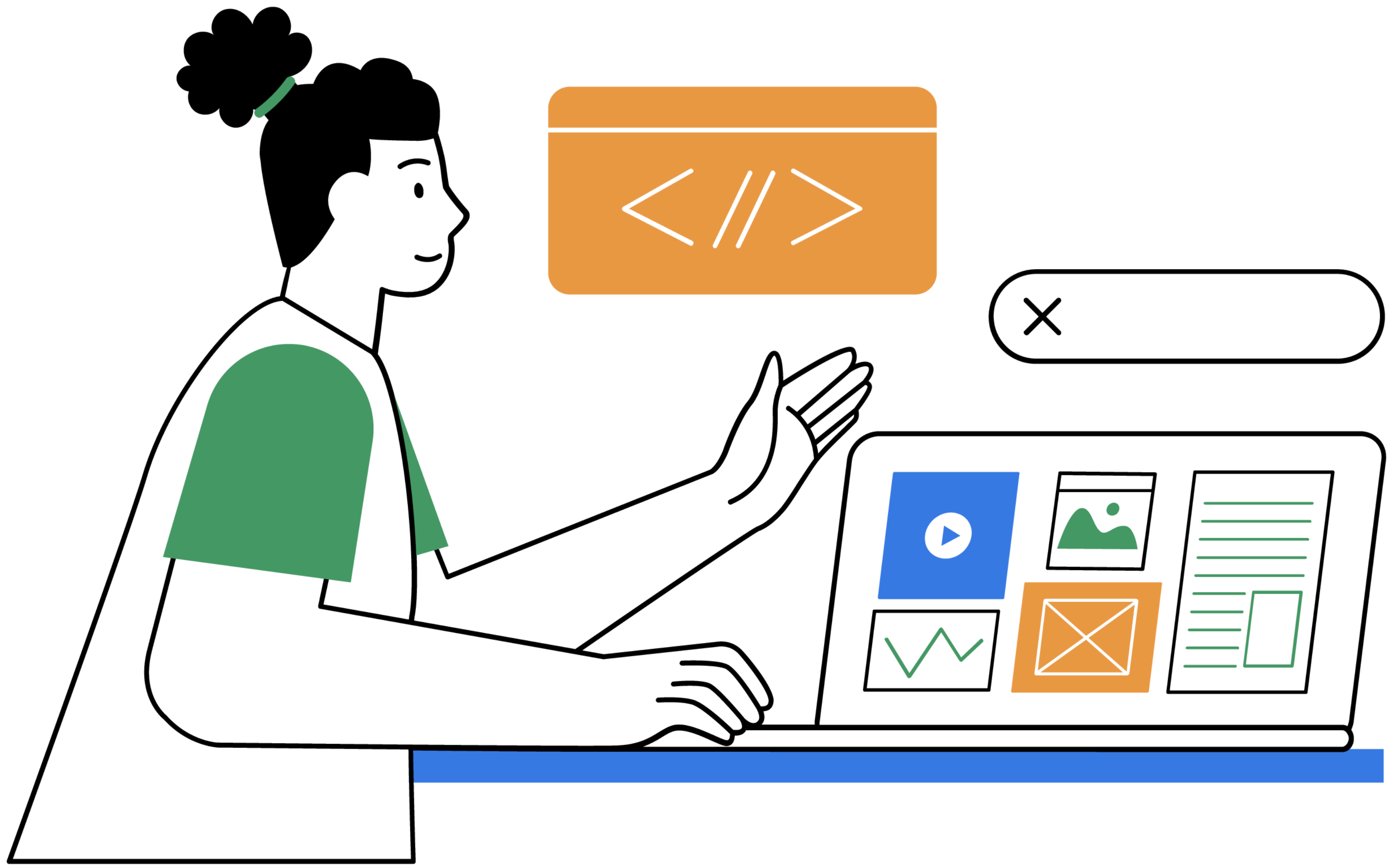
Models trained:

- **Gradient boosting:** High accuracy & robust performance.
- **Decision tree:** Simple but less accurate.
- **Random forest:** Powerful model, but less responsive.
- **LDA + SMOTEENN:** Balanced and accurate.
- **Random forest + PCA:** Lower performance.
- **Random forest + Grid Search:** Optimized hyperparameter tuning

Source code: www.google.research.colab.com



Approach

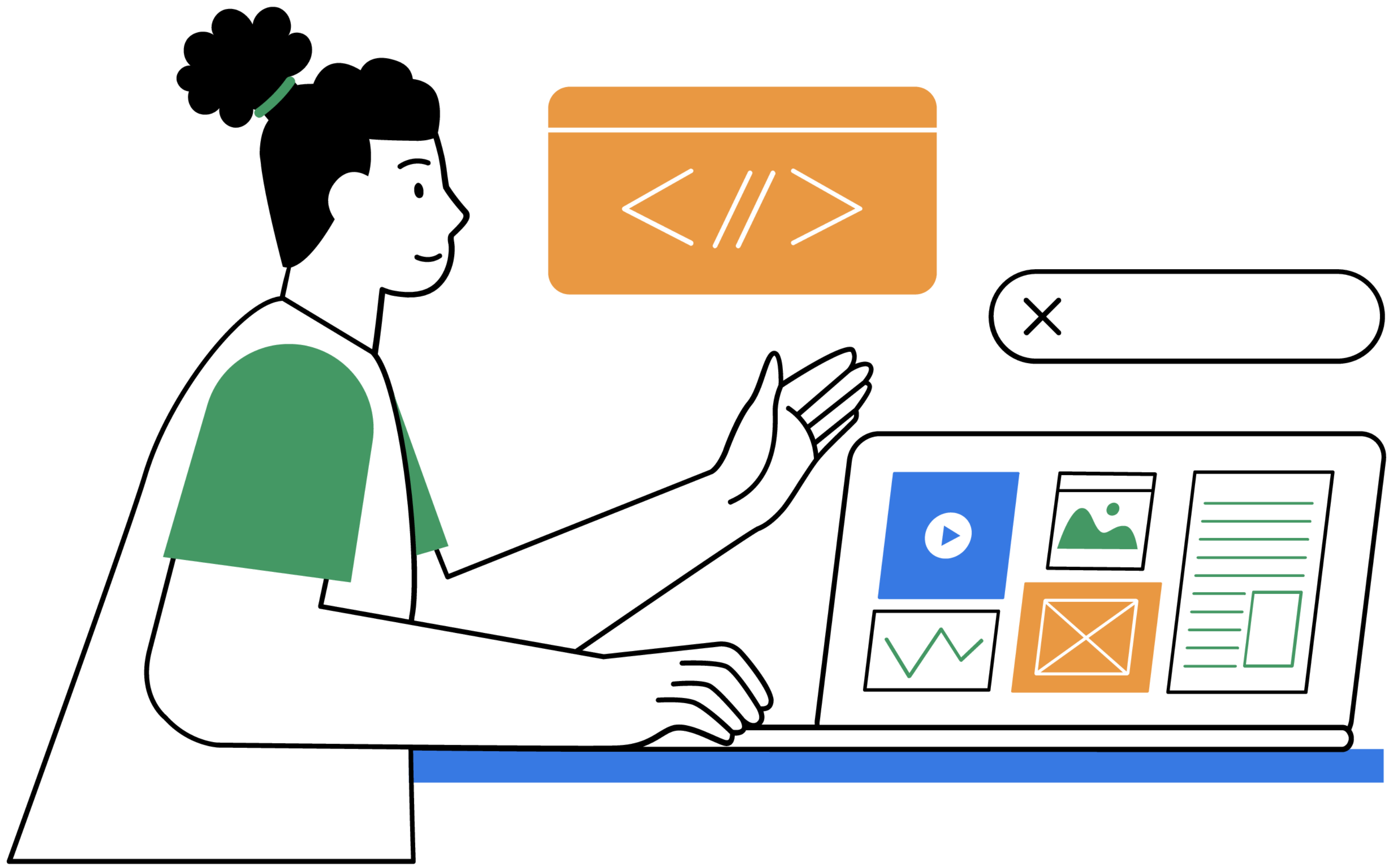


Approach

Data cleaning:

- Removed missing values, fixed types .

Source code: www.google.research.colab.com



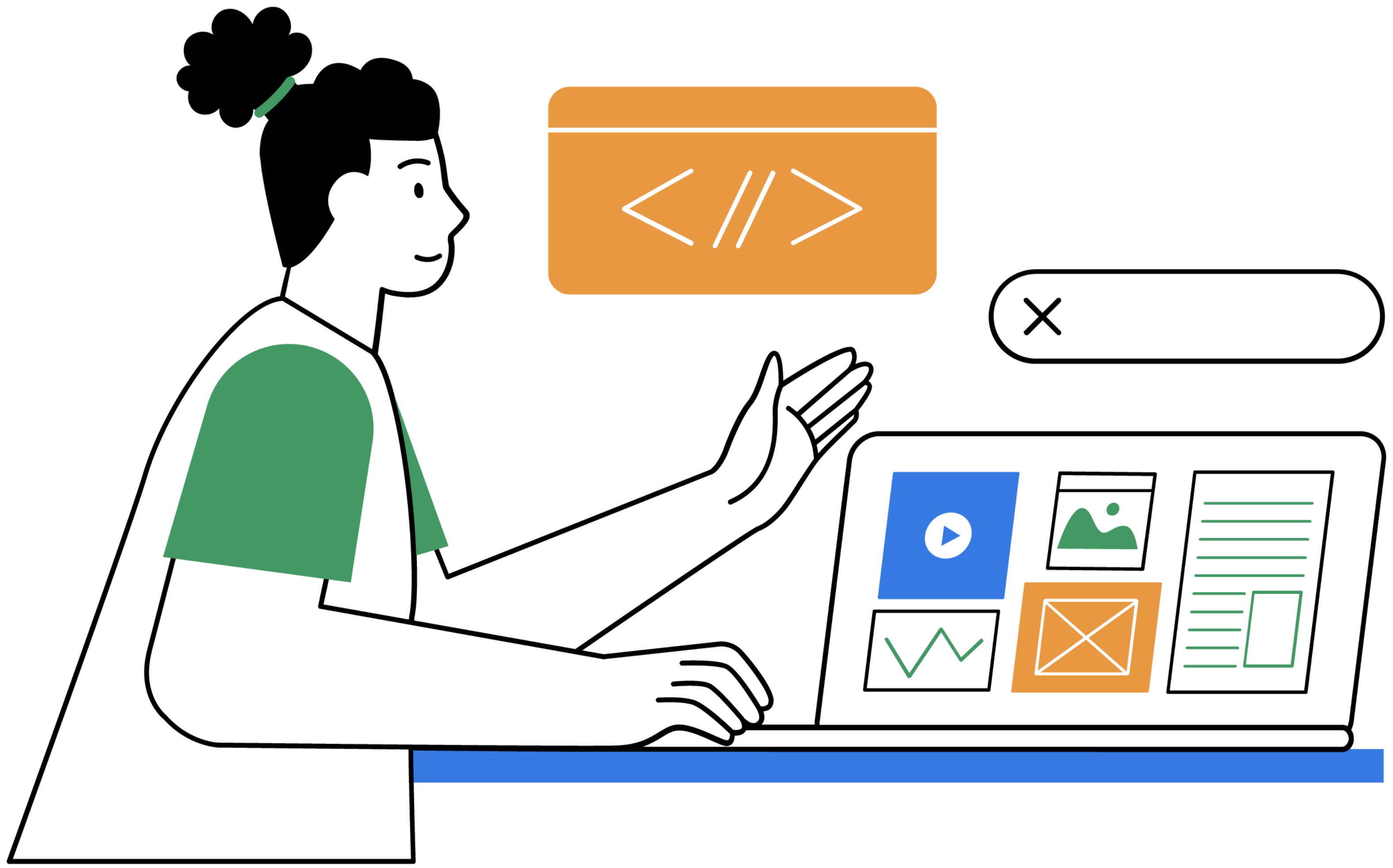
Approach

Data cleaning:

- Removed missing values, fixed types .

Feature engineering:

- Created features like “tenure_group”.



Approach

Data cleaning:

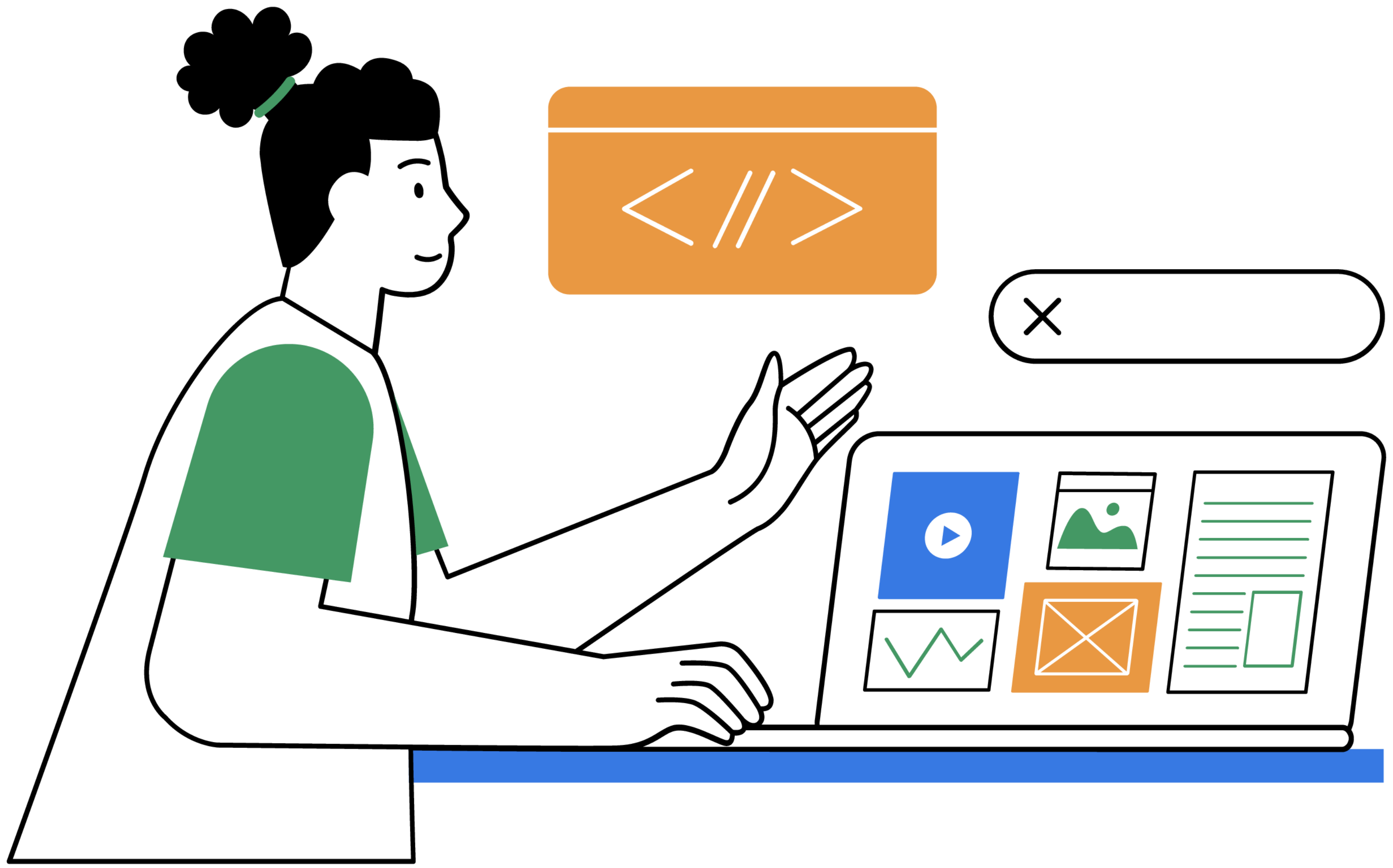
- Removed missing values, fixed types .

Feature engineering:

- Created features like “tenure_group”.

Exploratory data analysis:

- Visual trends in churn.



Approach

Data cleaning:

- Removed missing values, fixed types .

Feature engineering:

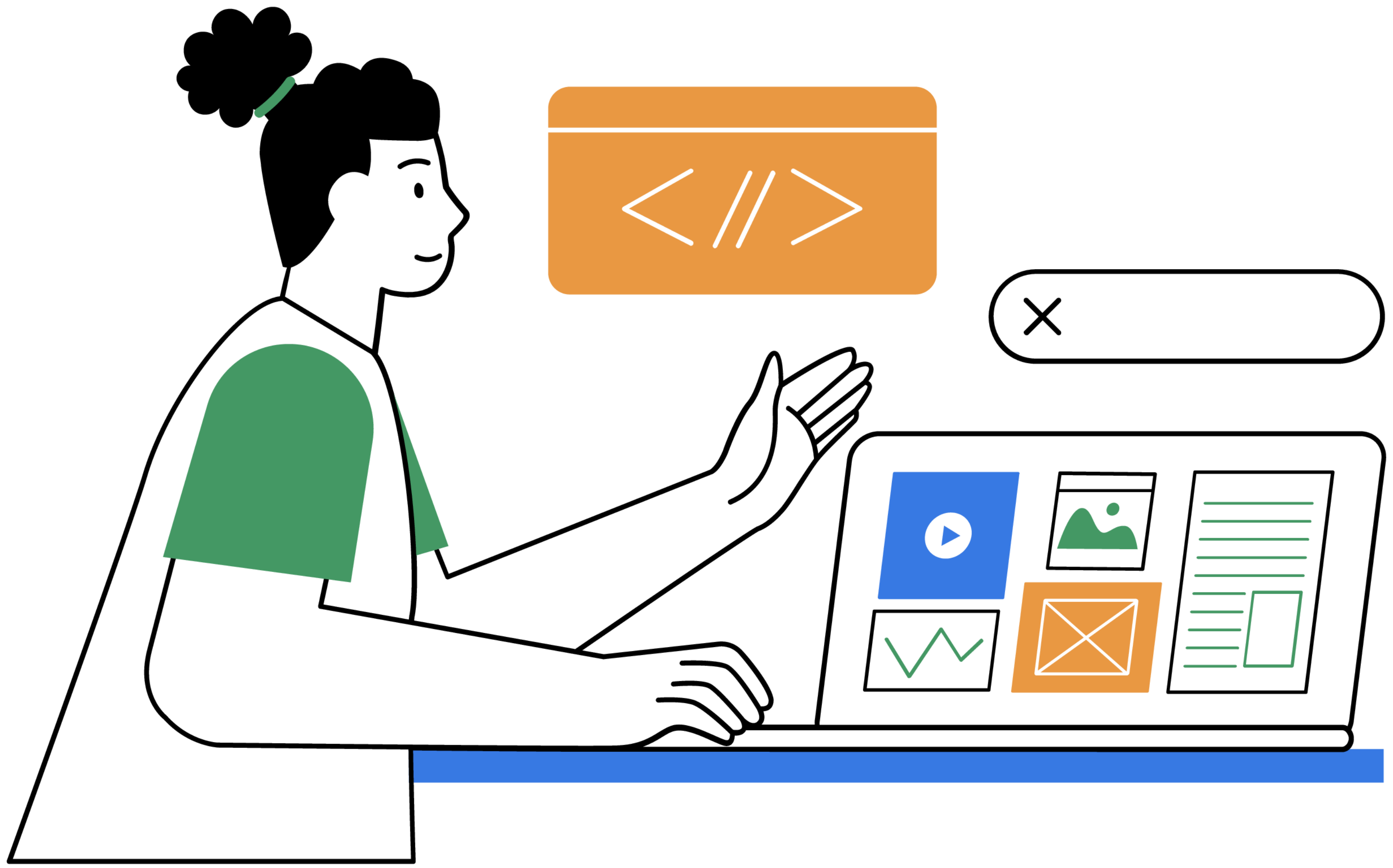
- Created features like “tenure_group”.

Exploratory data analysis:

- Visual trends in churn.

Model training & evaluation:

- Tested various ML models.



Approach

Data cleaning:

- Removed missing values, fixed types .

Feature engineering:

- Created features like “tenure_group”.

Exploratory data analysis:

- Visual trends in churn.

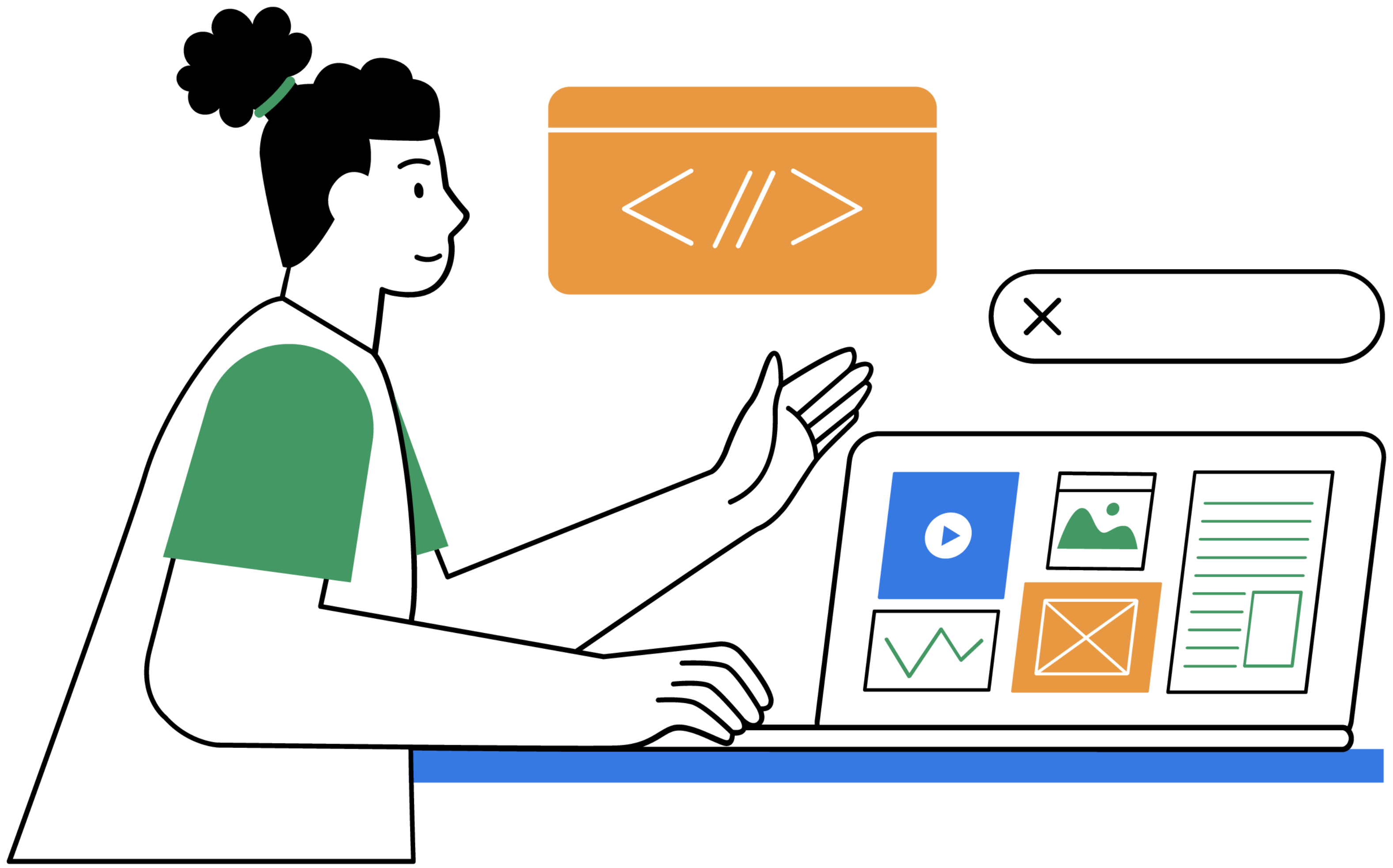
Model training & evaluation:

- Tested various ML models.

Hyperparameter tuning:

- Used “grid search” for enhanced performance.

Source code: www.google.research.colab.com



Approach

Data cleaning:

- Removed missing values, fixed types .

Feature engineering:

- Created features like “tenure_group”.

Exploratory data analysis:

- Visual trends in churn.

Model training & evaluation:

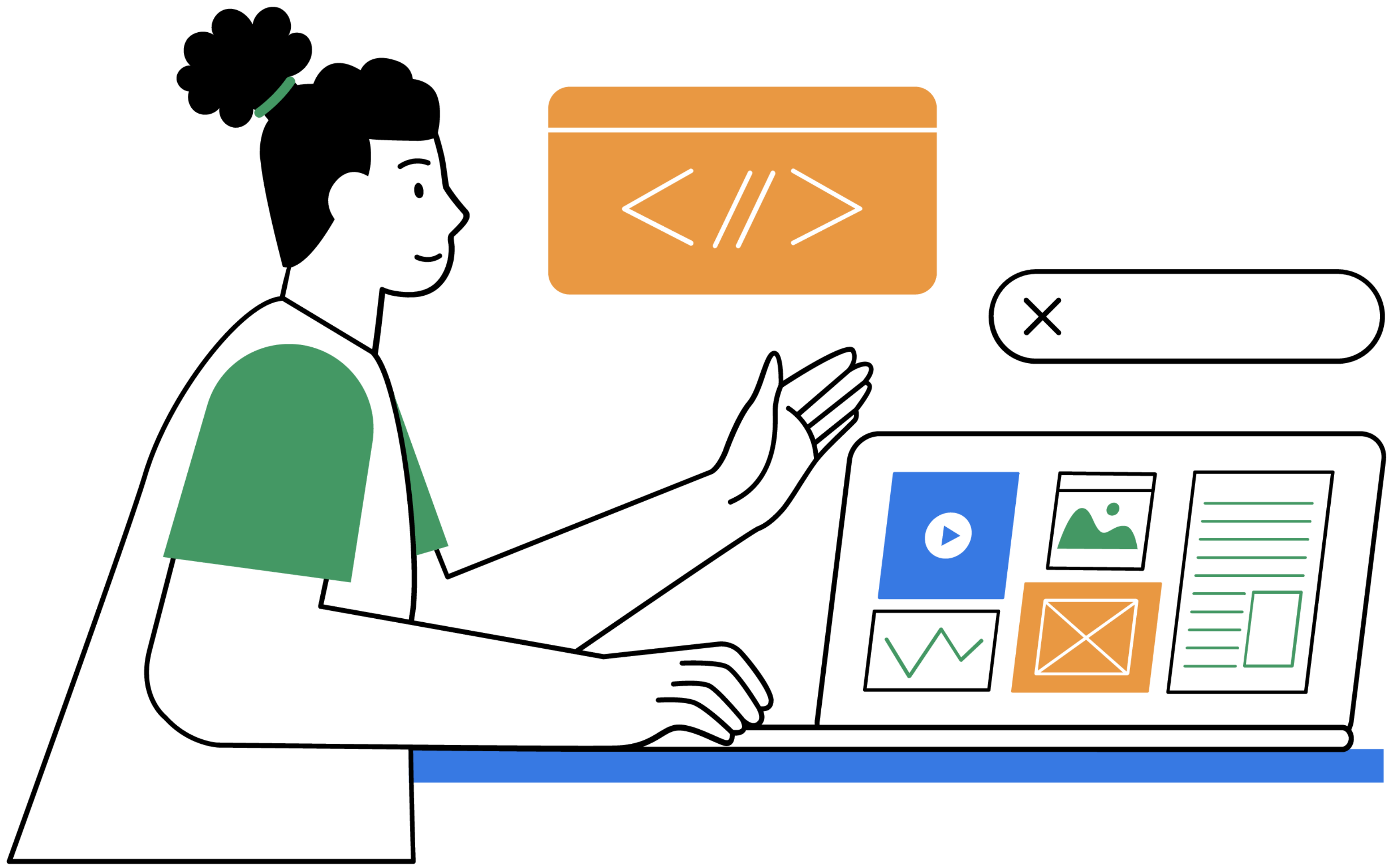
- Tested various ML models.

Hyperparameter tuning:

- Used “grid search” for enhanced performance.

Best model selection:

- Chose “gradient boosting”.



Approach

Data cleaning:
- Removed missing values, fixed types .

Feature engineering:
- Created features like “tenure_group”.

Exploratory data analysis:
- Visual trends in churn.

Model training & evaluation:
- Tested various ML models.

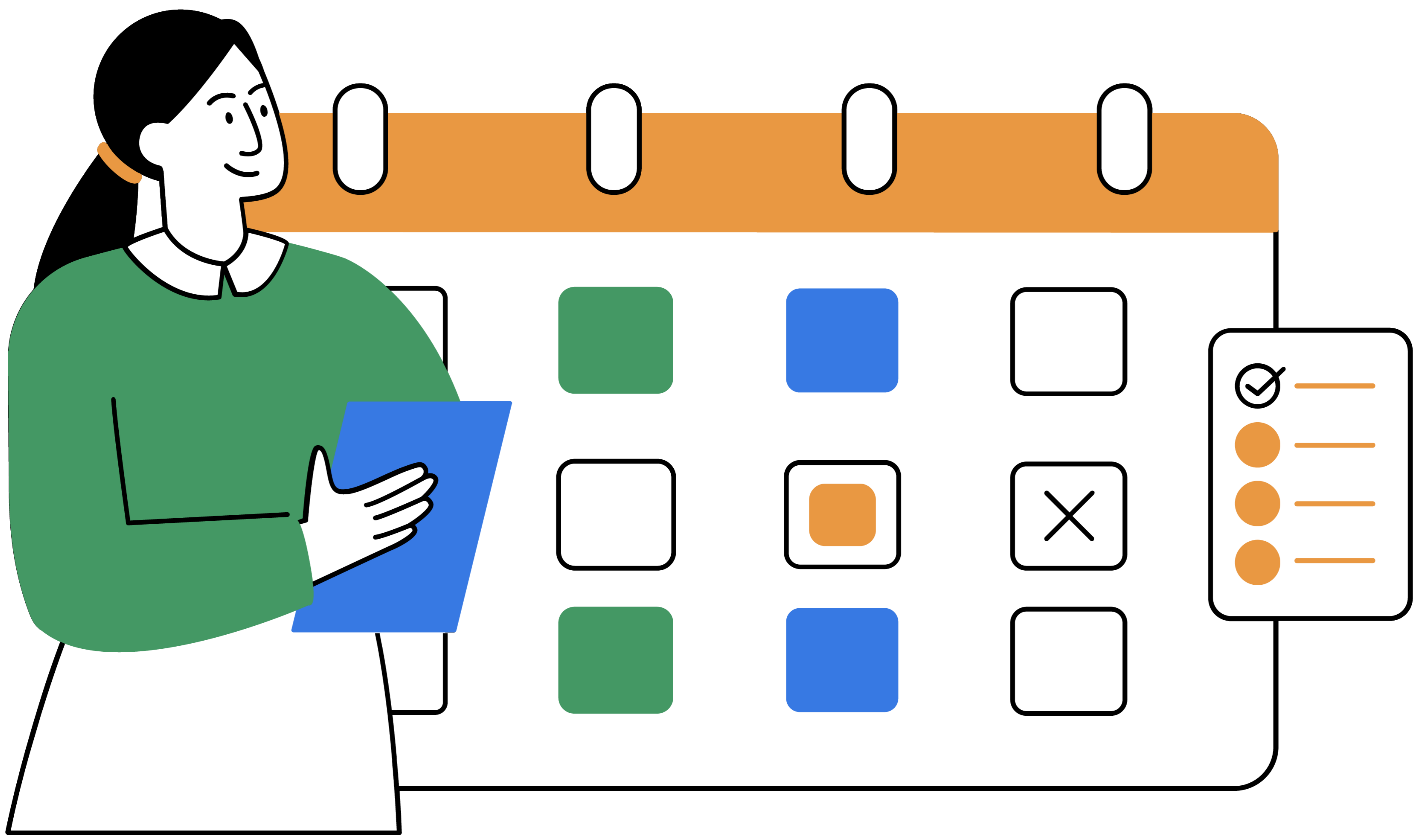
Hyperparameter tuning:
- Used “grid search” for enhanced performance.

Best model selection:
- Chose “gradient boosting”.

Deployment:
- Built a “flask” web app for making predictions.

Source code: www.google.research.colab.com

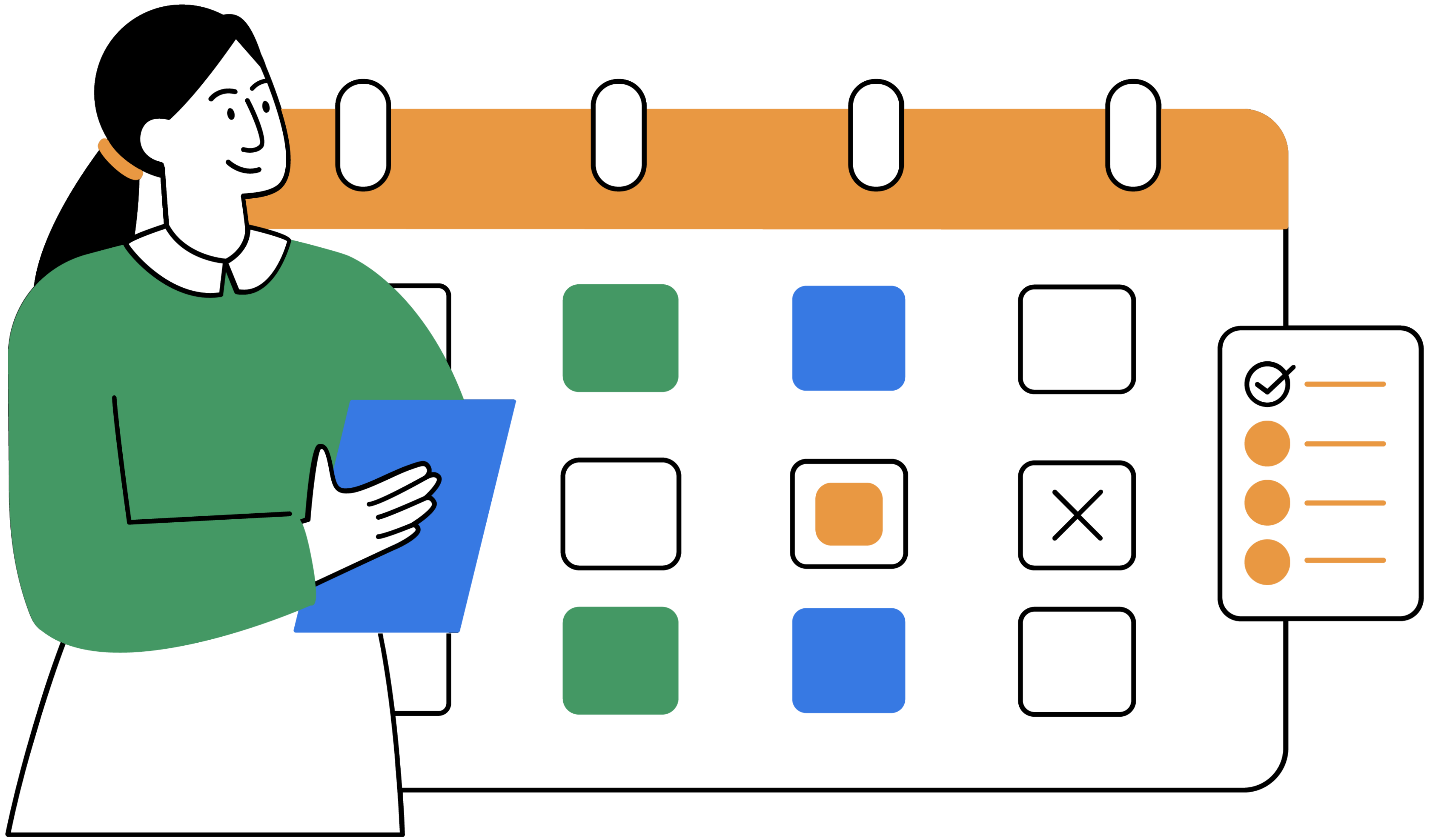
Data Analysis Process



Data Analysis Process

Why gradient boosting?

- Best overall performance.
- Handles class imbalance well.
- High precision, recall, and F1 - score.



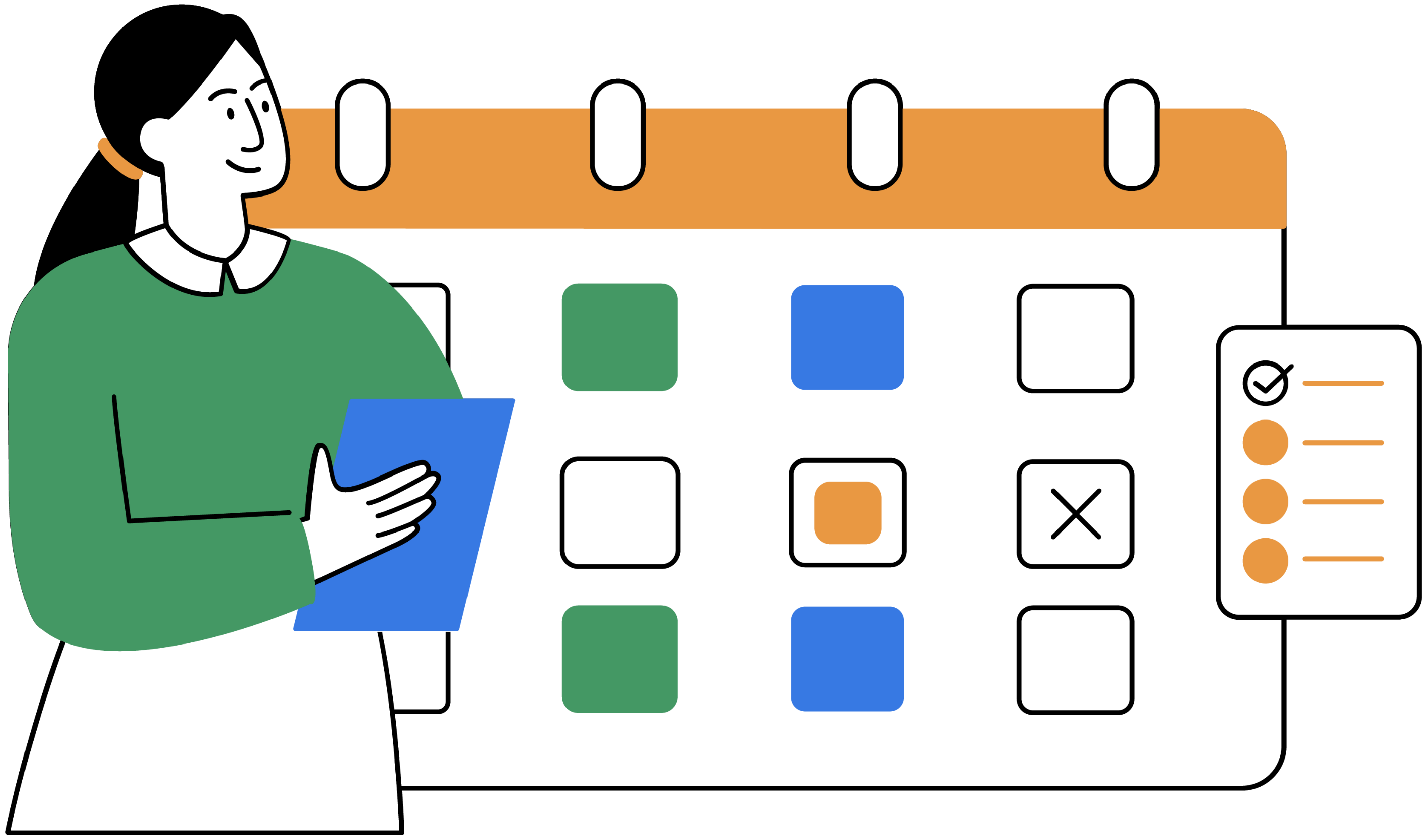
Data Analysis Process

Why gradient boosting?

- Best overall performance.
- Handles class imbalance well.
- High precision, recall, and F1 - score.

Input features used (19 total):

- Demographics.
- Contract type.
- Services used.
- Tenure & usage patterns.



Data Analysis Process

Why gradient boosting?

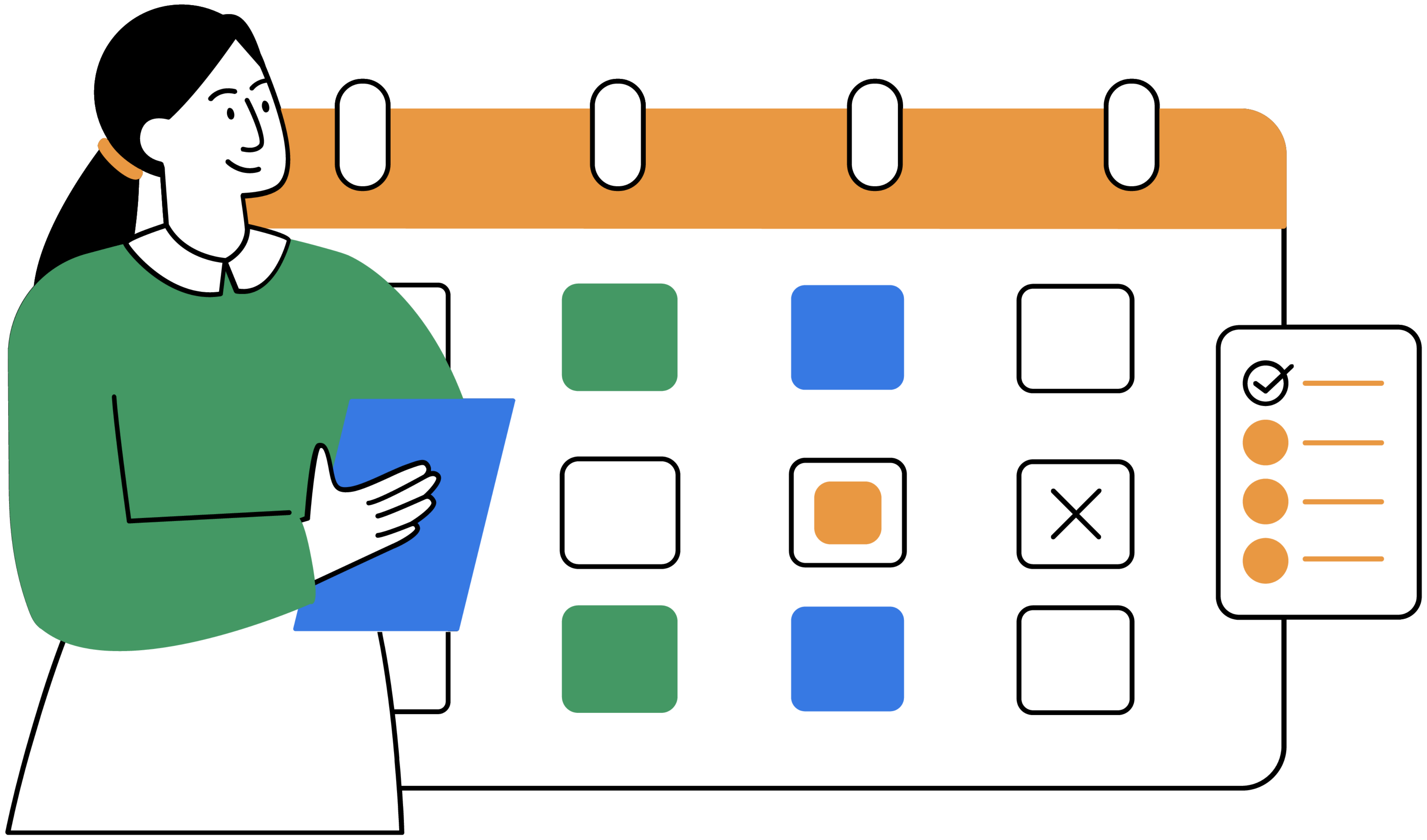
- Best overall performance.
- Handles class imbalance well.
- High precision, recall, and F1 - score.

Input features used (19 total):

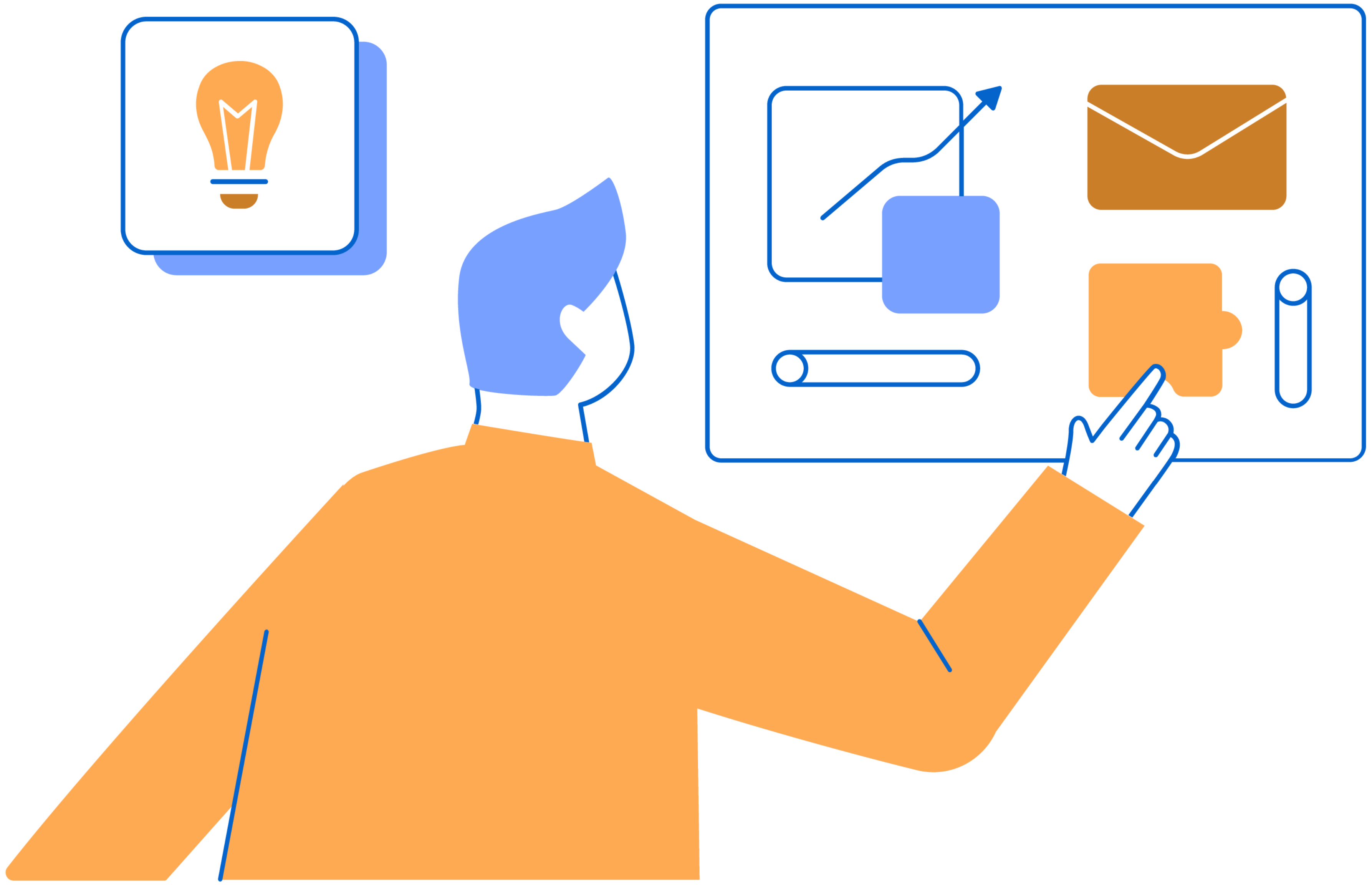
- Demographics.
- Contract type.
- Services used.
- Tenure & usage patterns.

Output:

- Churn prediction: Yes / No.



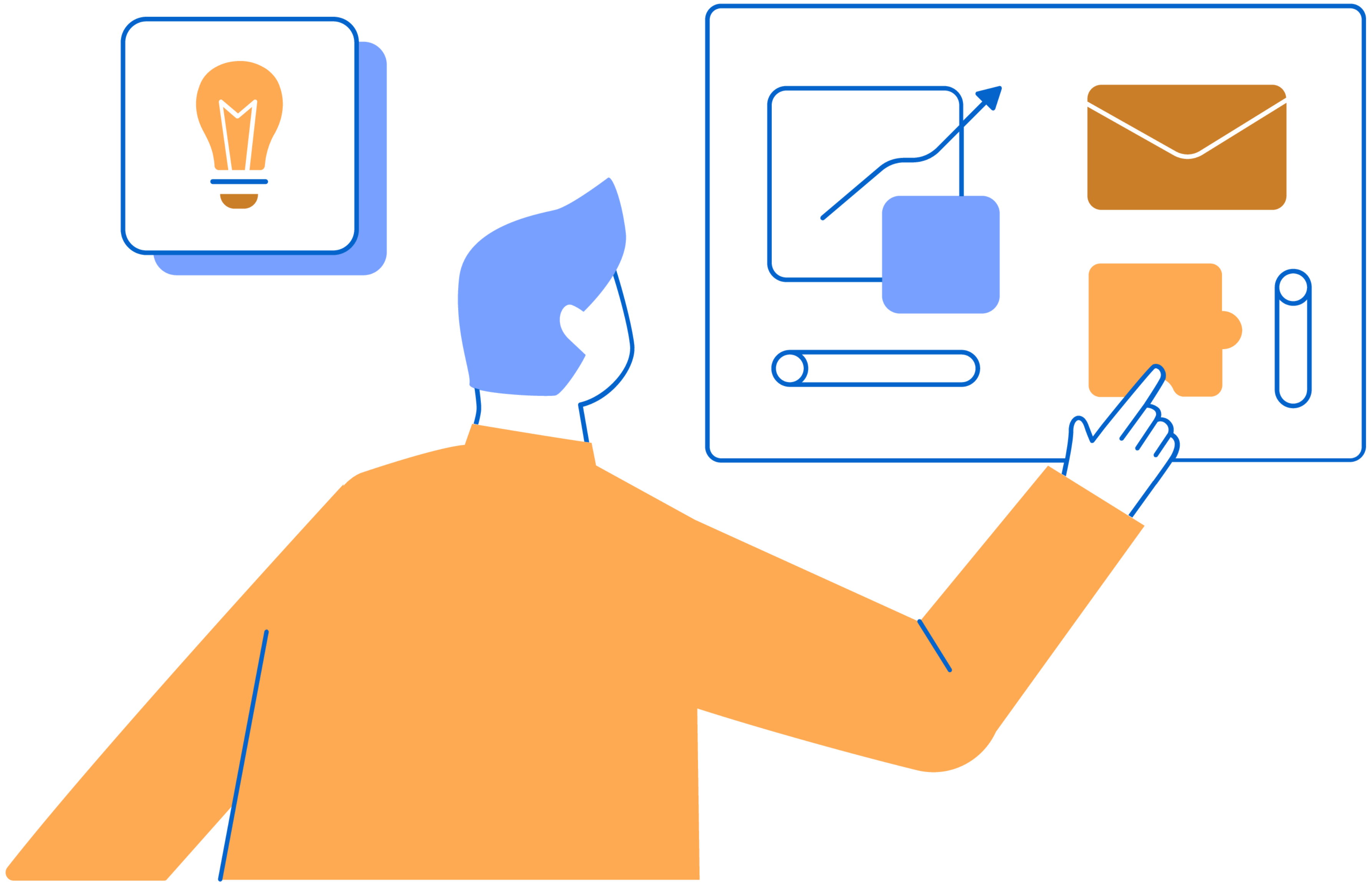
Key Points Supported with Data



Key Points Supported with Data

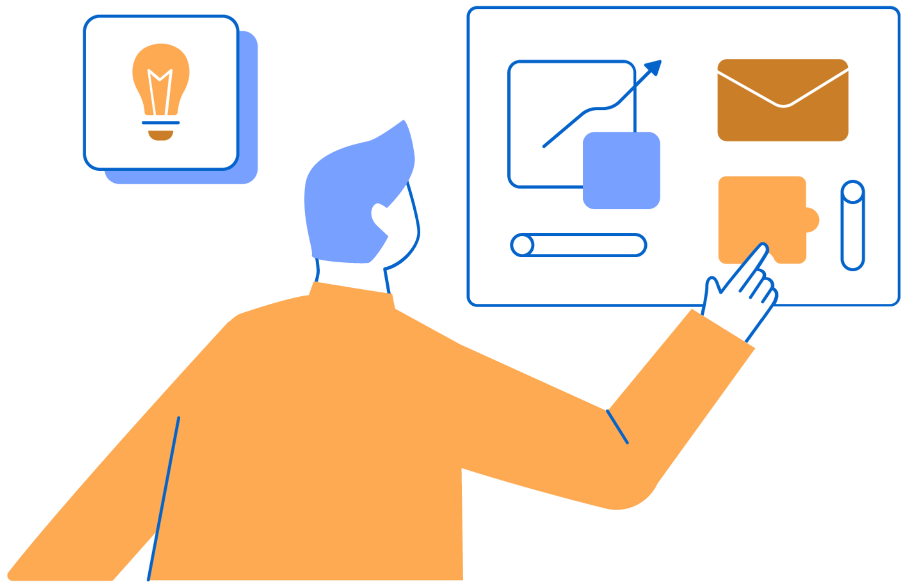
Customer behavior insights:

- Churning of customers.
- Distribution of internet services for churned customers.
- Distribution of online security plan for churned customers.
- Distribution of device protection plan for churned customers.
- Distribution of contract for churned customers.
- Distribution of payment methods for churned customers.
- Distribution of tenure groups for churned customers.
- Relationship between monthly charges and customer churn.

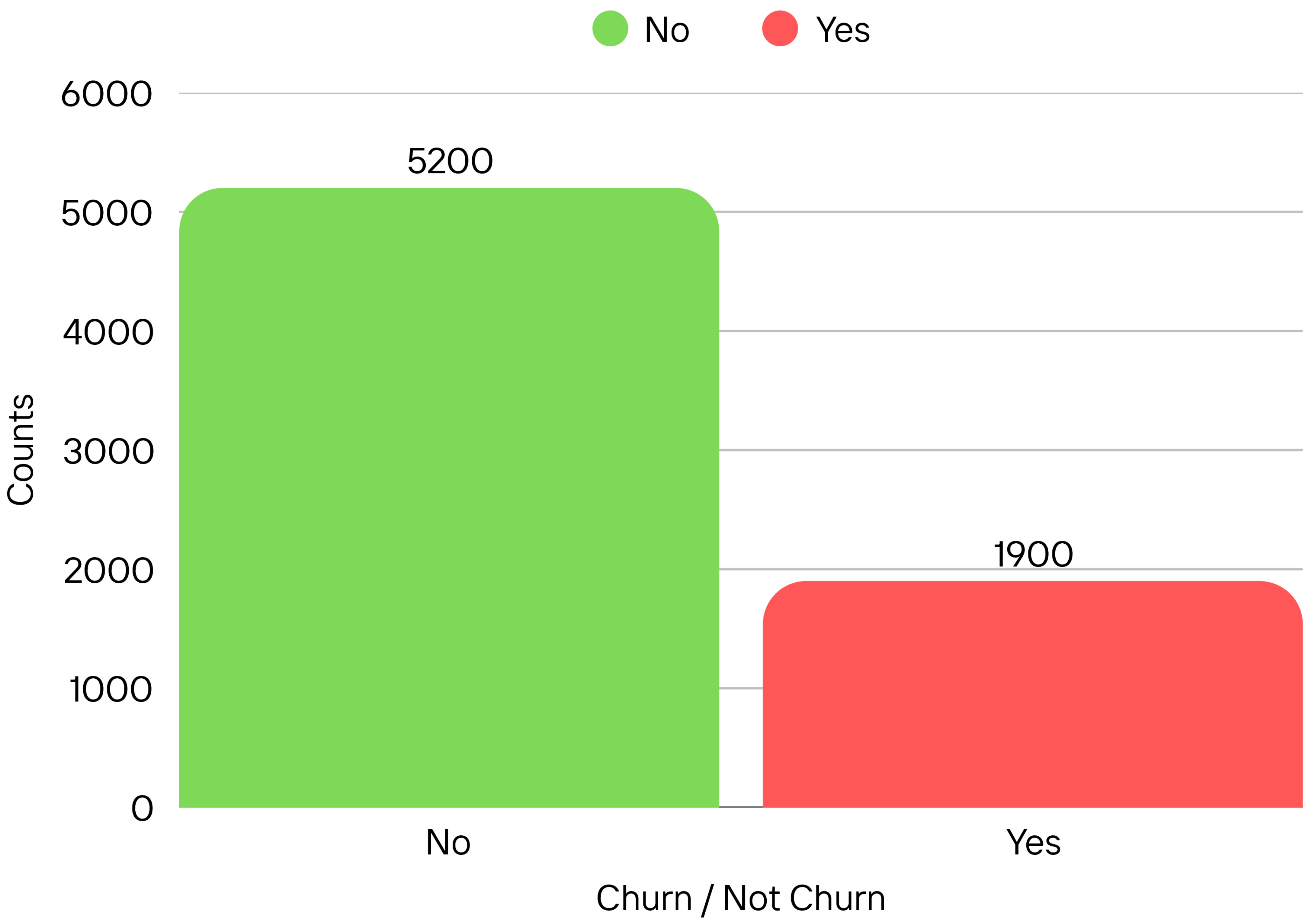


Source code: www.google.research.colab.com

Graph 1



Churning of customers

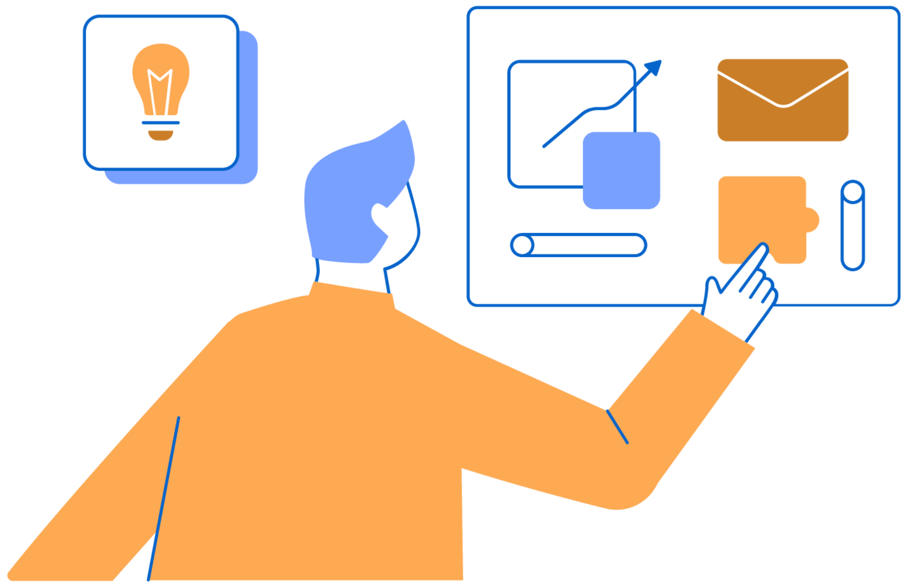


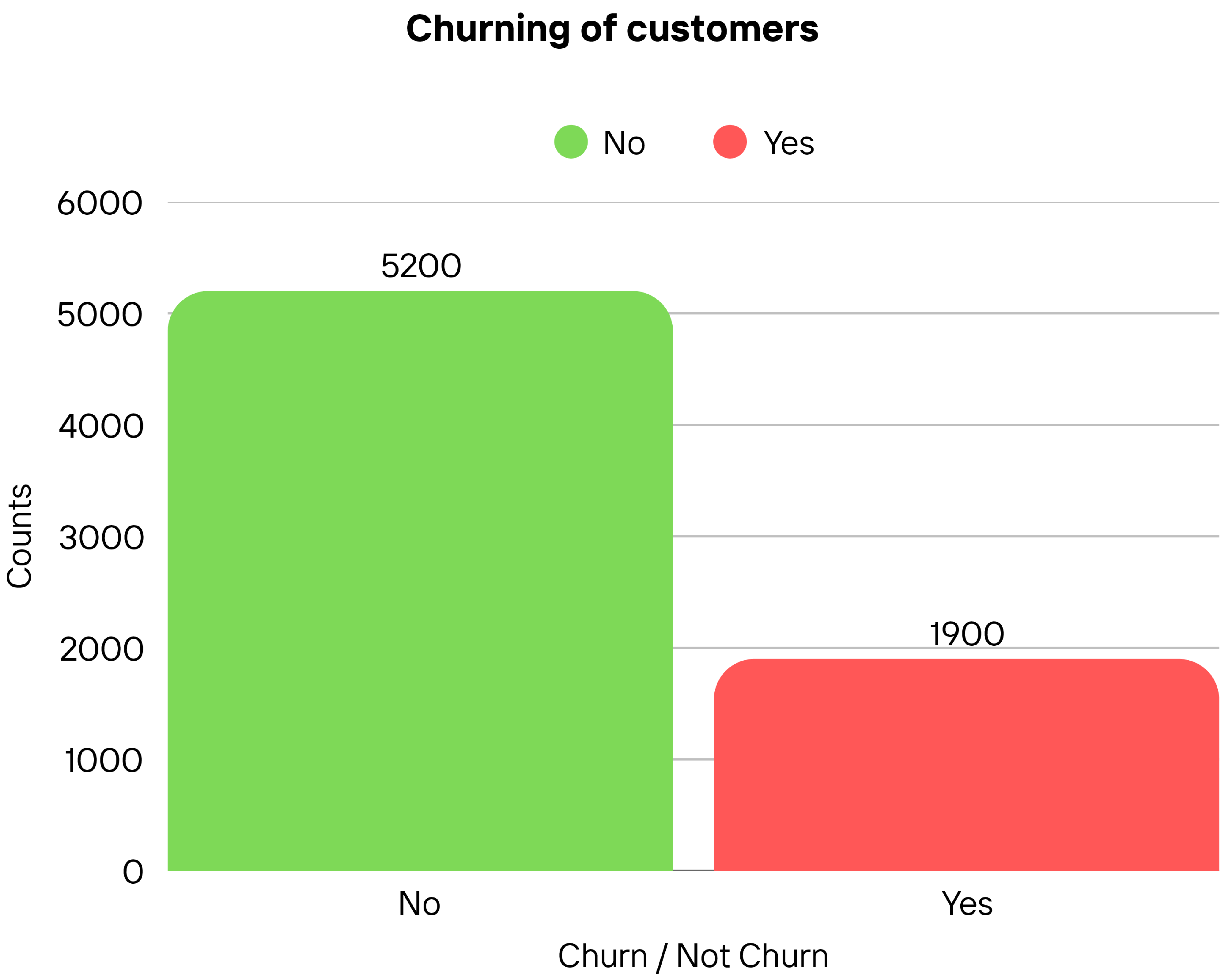
Source code: www.google.research.colab.com

Graph 1

Insights:

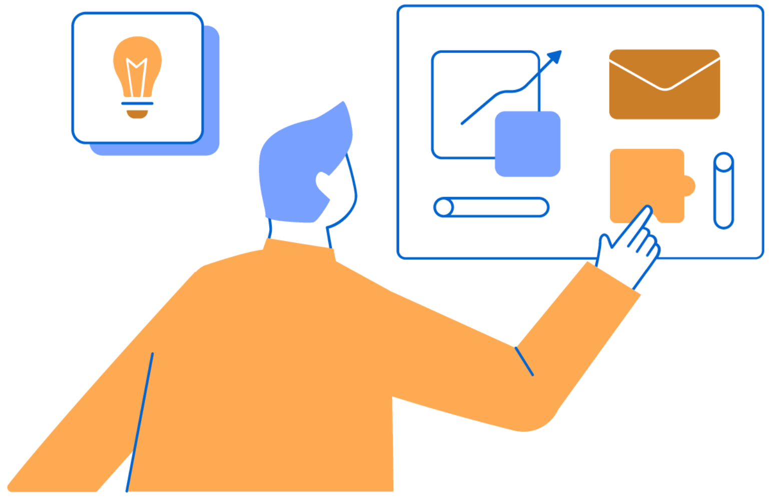
- Majority of customers “did not churn” i.e. 5,200 customers.
- Only “1,900” customers churned, indicating an approximate churn rate of “~27%”.

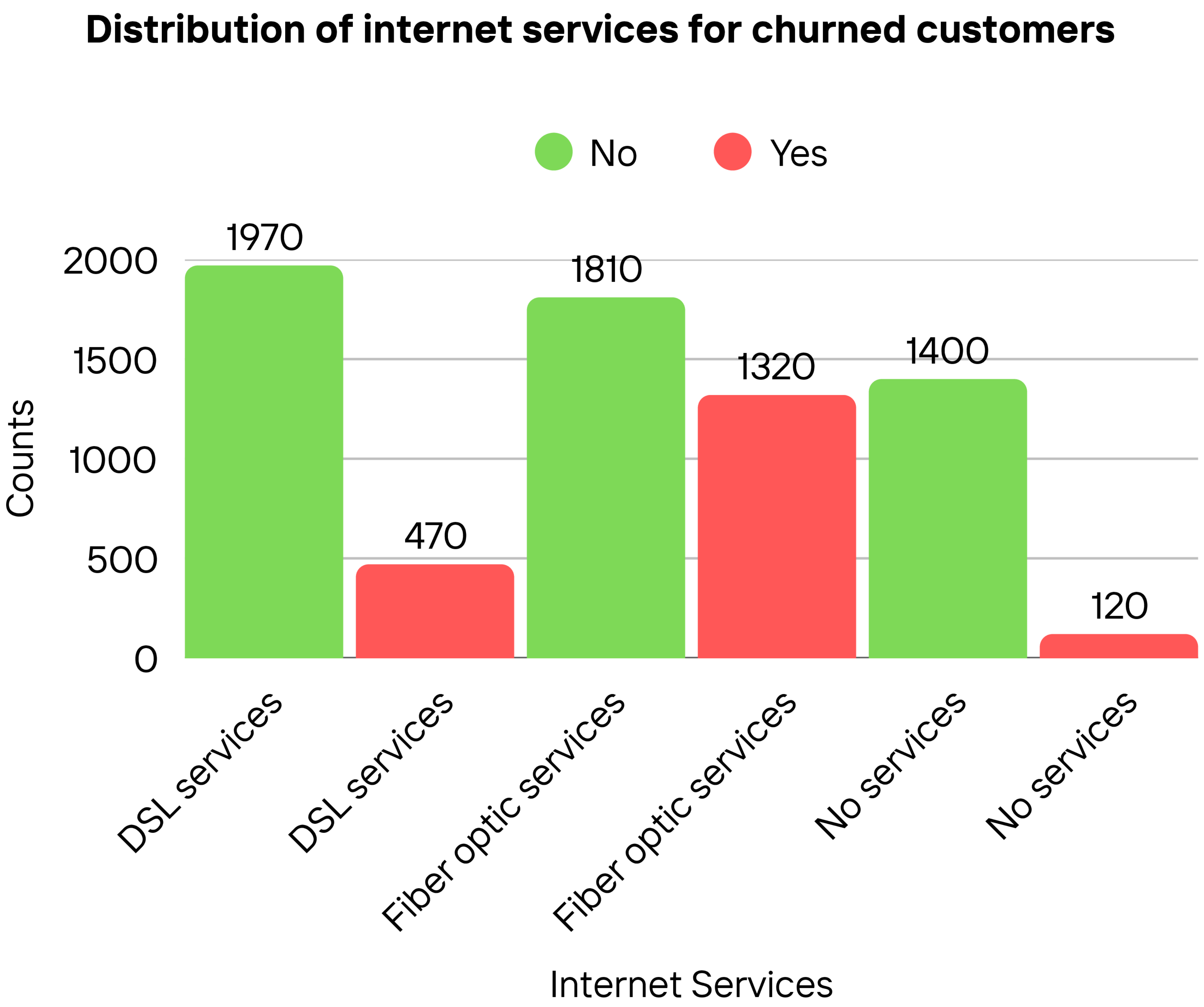




Source code: www.google.research.colab.com

Graph 2



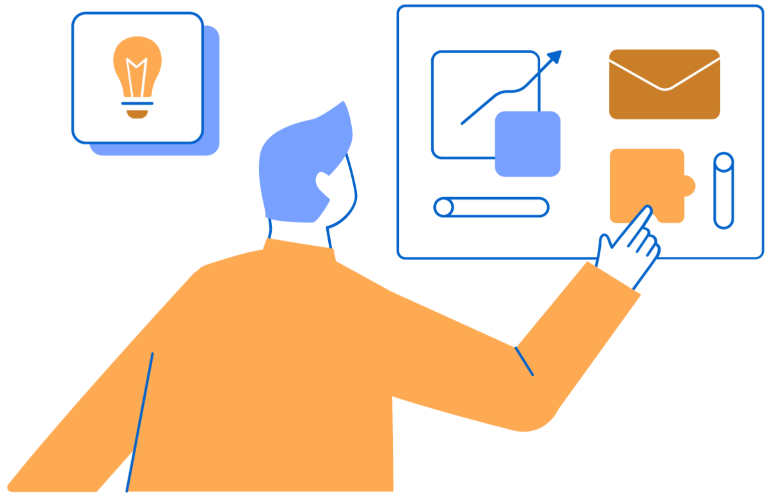


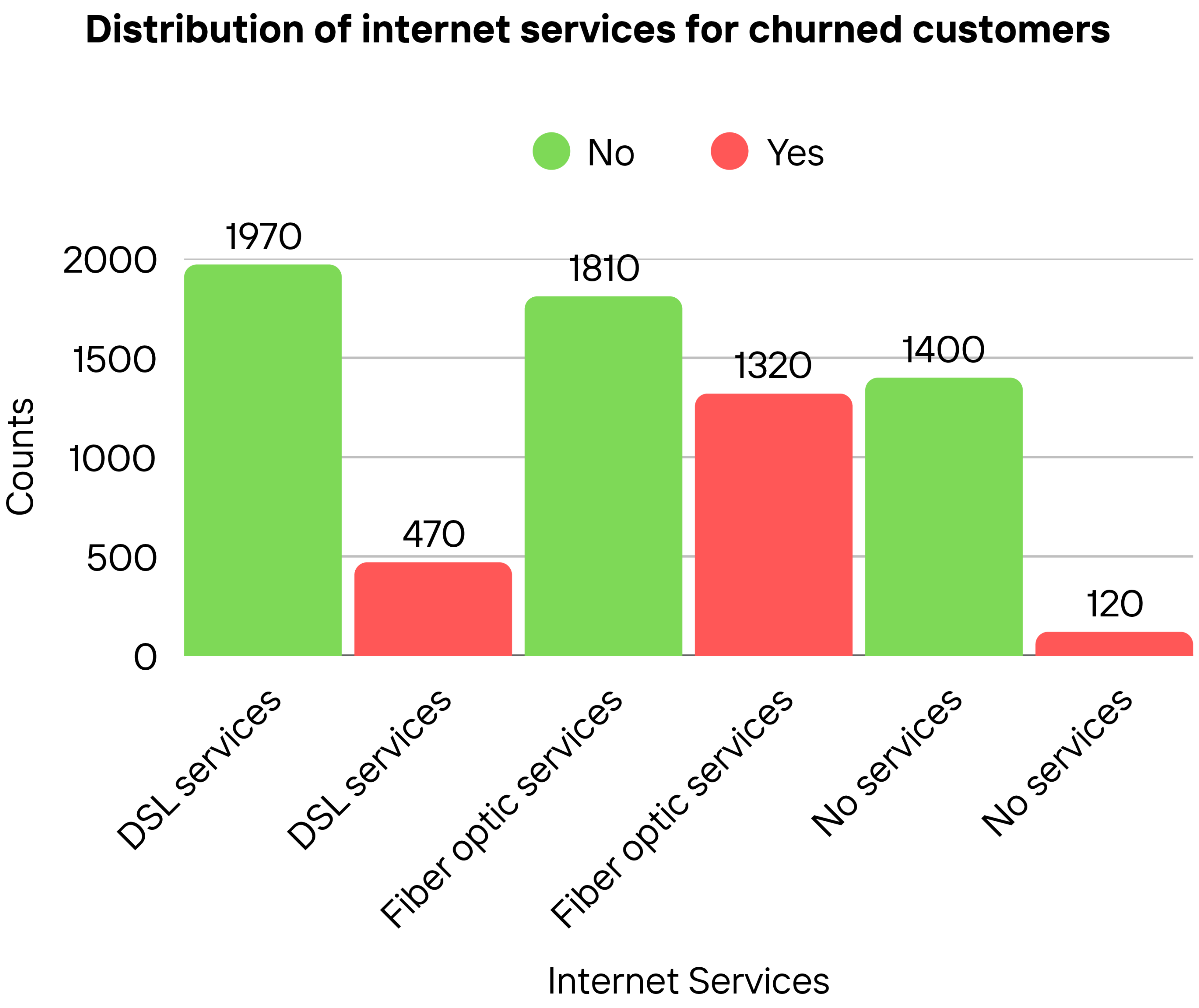
Source code: www.google.research.colab.com

Graph 2

Insights:

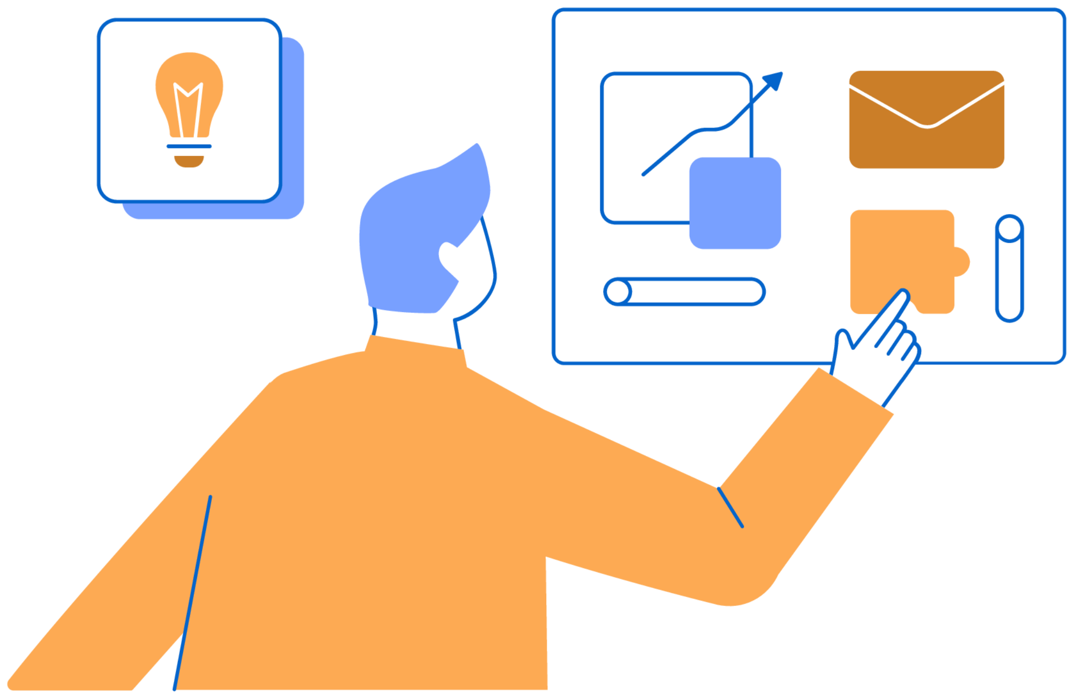
- Fiber optic users have the highest churn count i.e. “1,320".
- DSL users also churn but less than fiber users.
- Customers with no internet service churned the least “120” - likely because they are using minimal services.



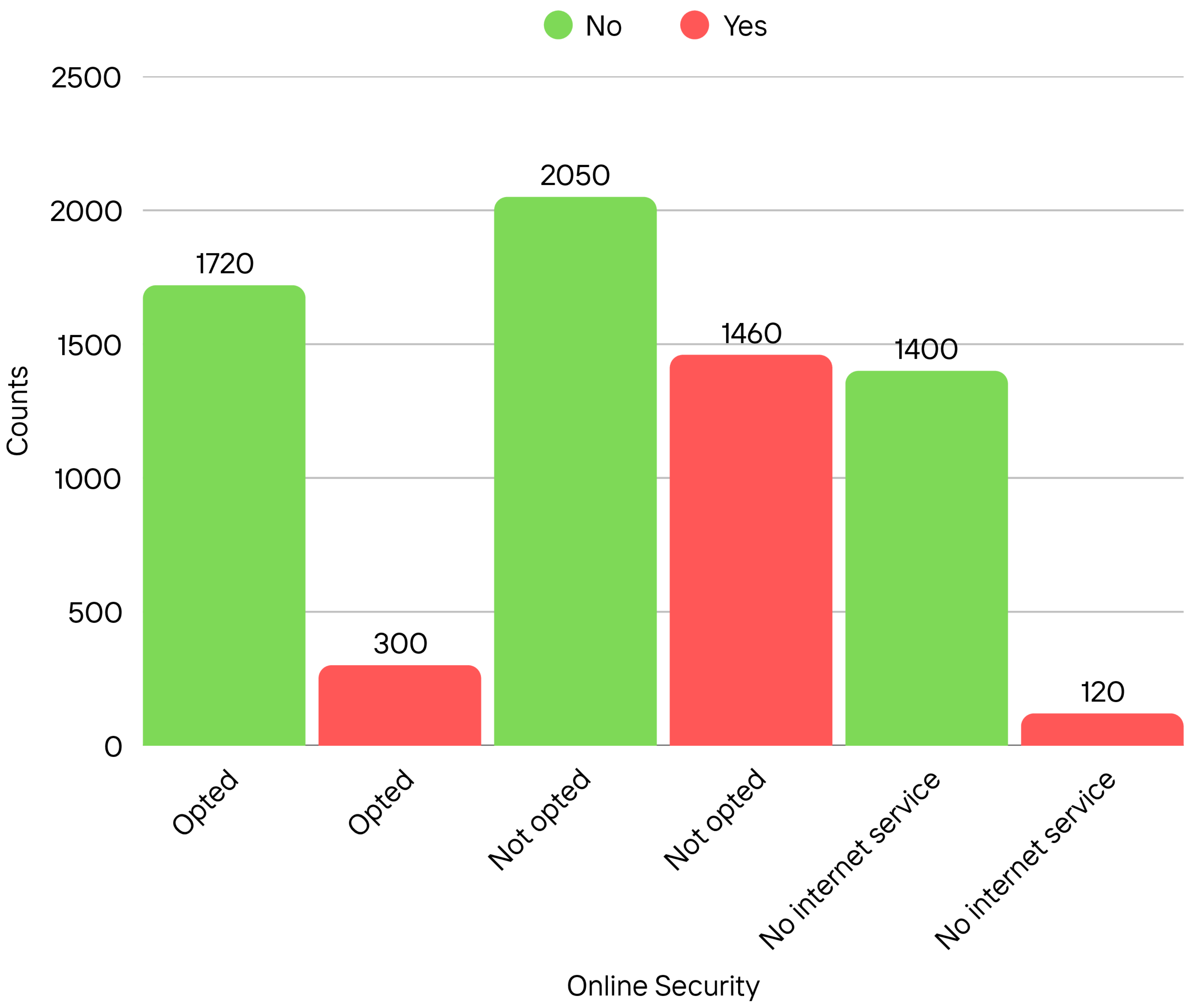


Source code: www.google.research.colab.com

Graph 3



Distribution of online security plan for churned customers

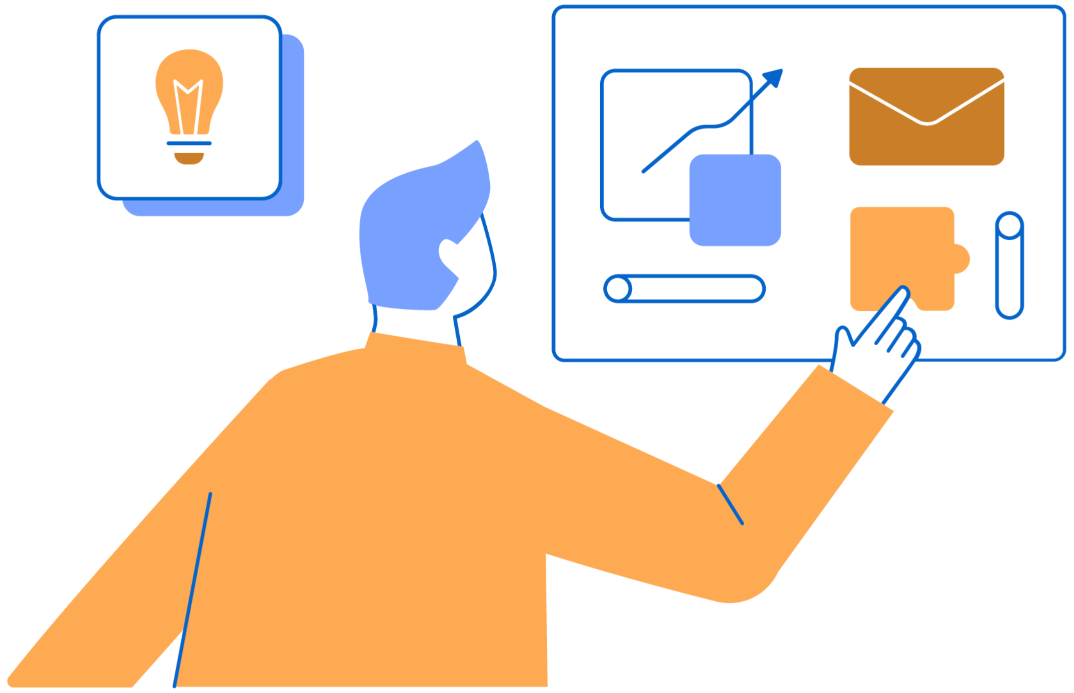


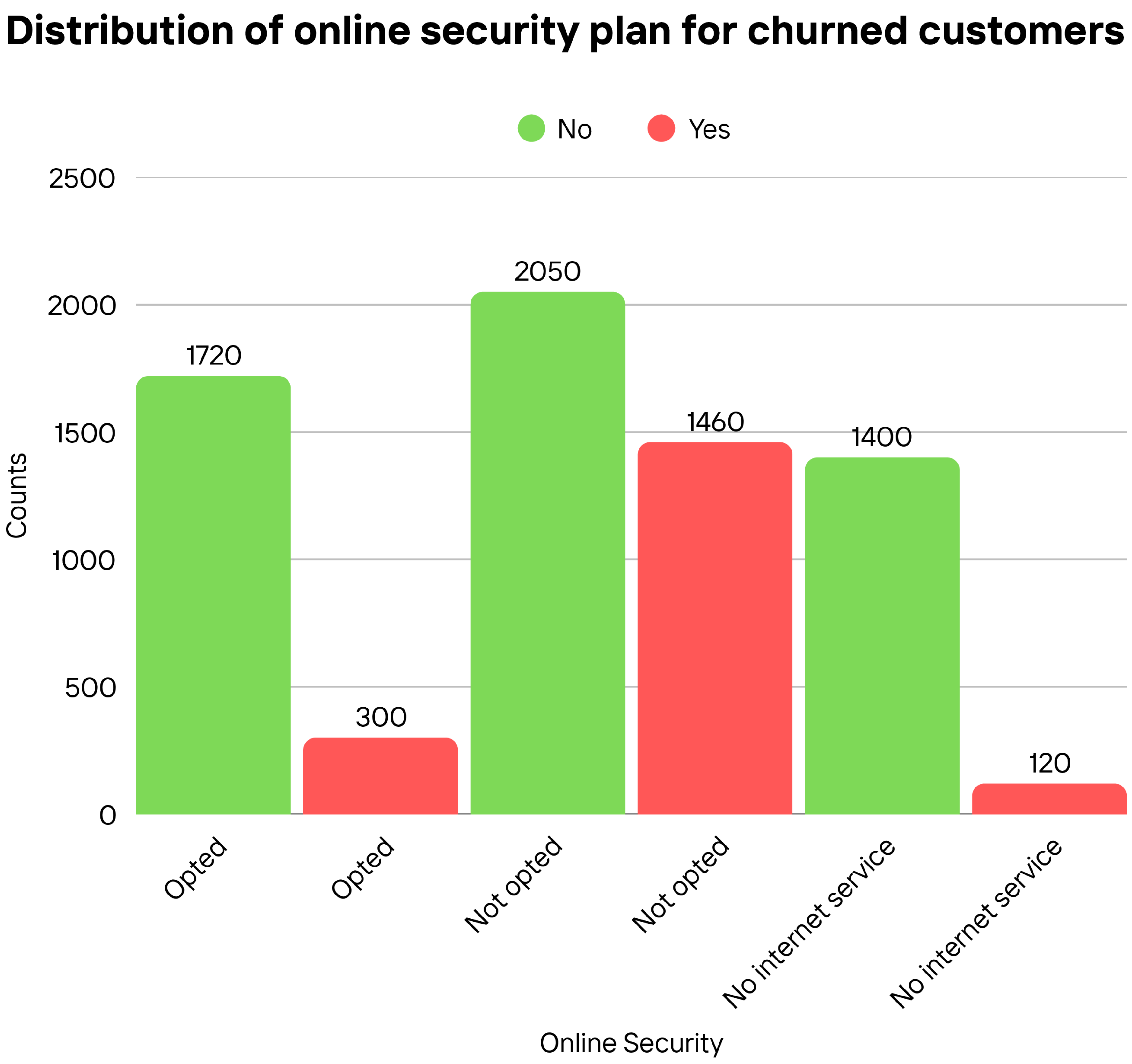
Source code: www.google.research.colab.com

Graph 3

Insights:

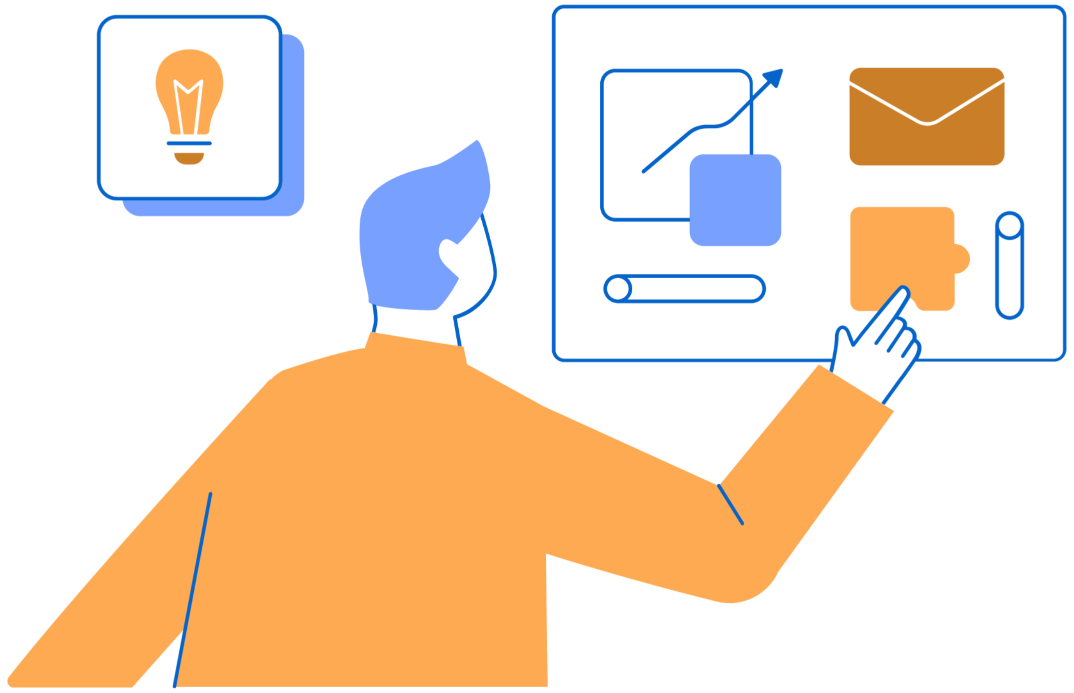
- Customers who "opted out" of online security show "higher churn" i.e. 1,460.
- Those who "opted in" had lower churn i.e. 300.
- Providing security services may help reduce churn.

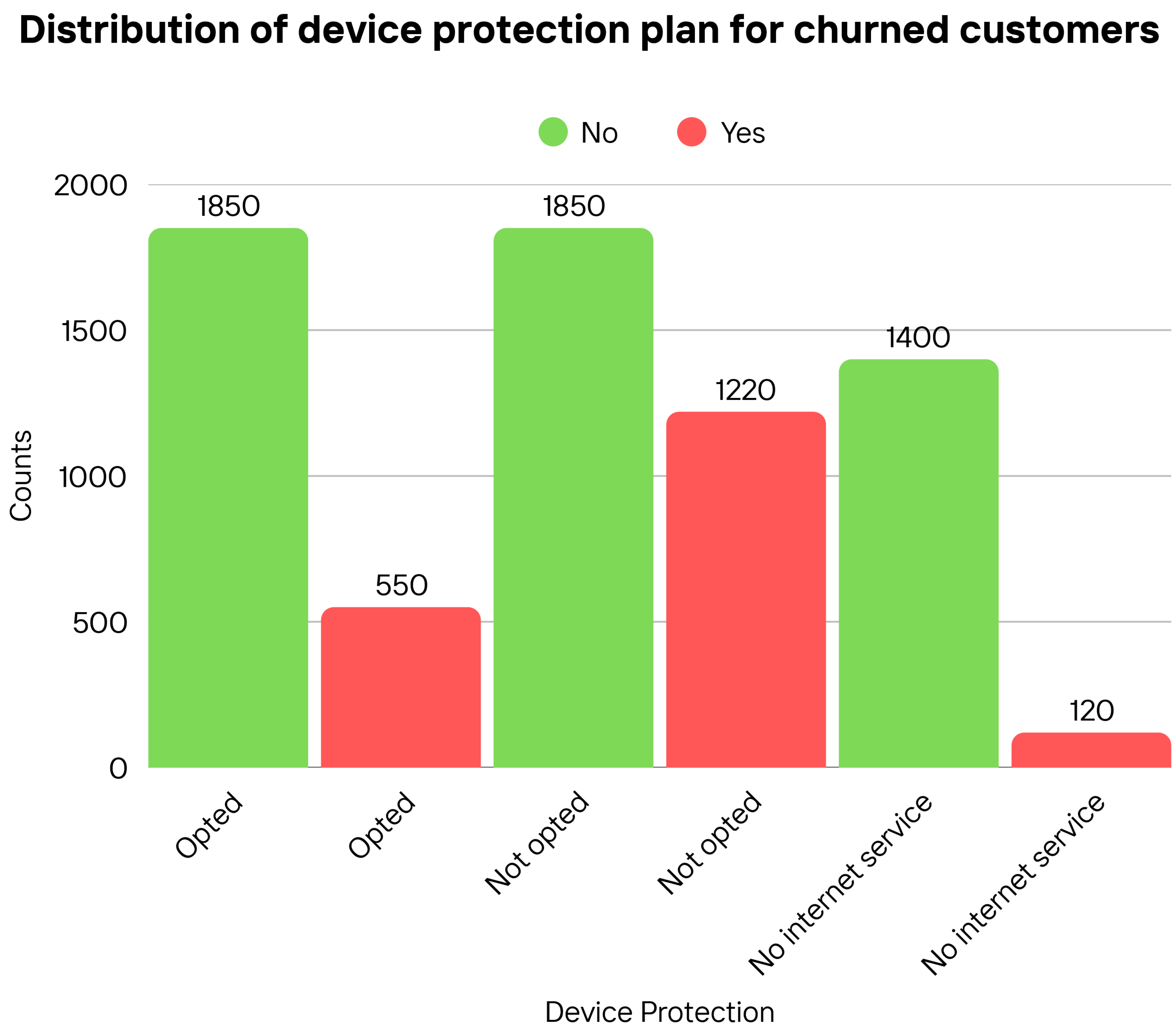




Source code: www.google.research.colab.com

Graph 4



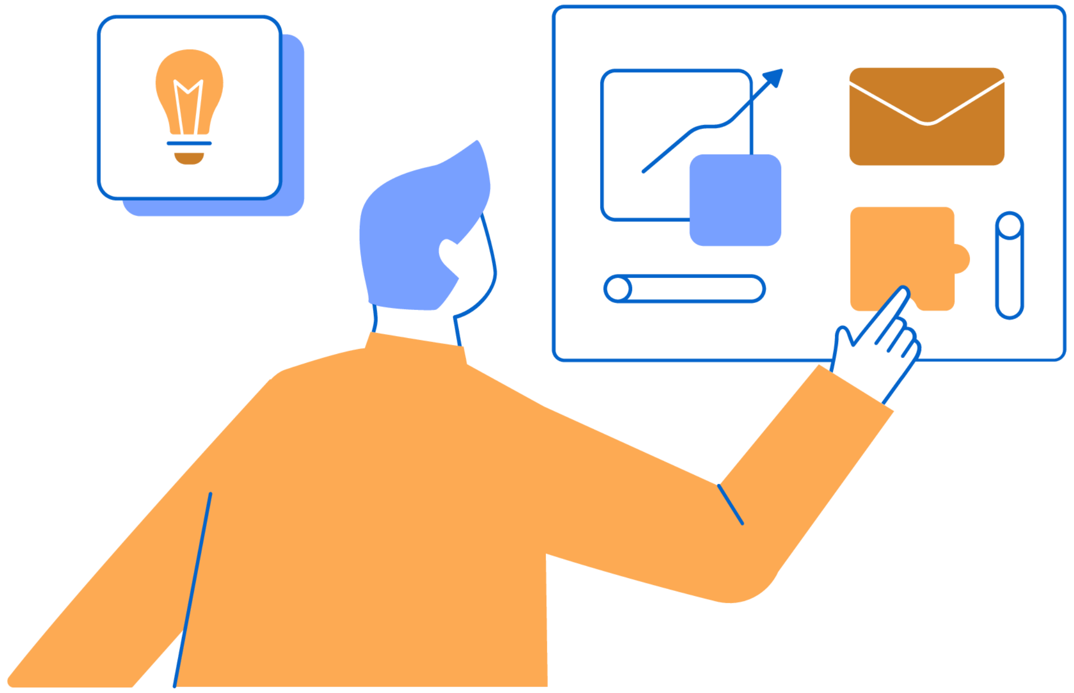


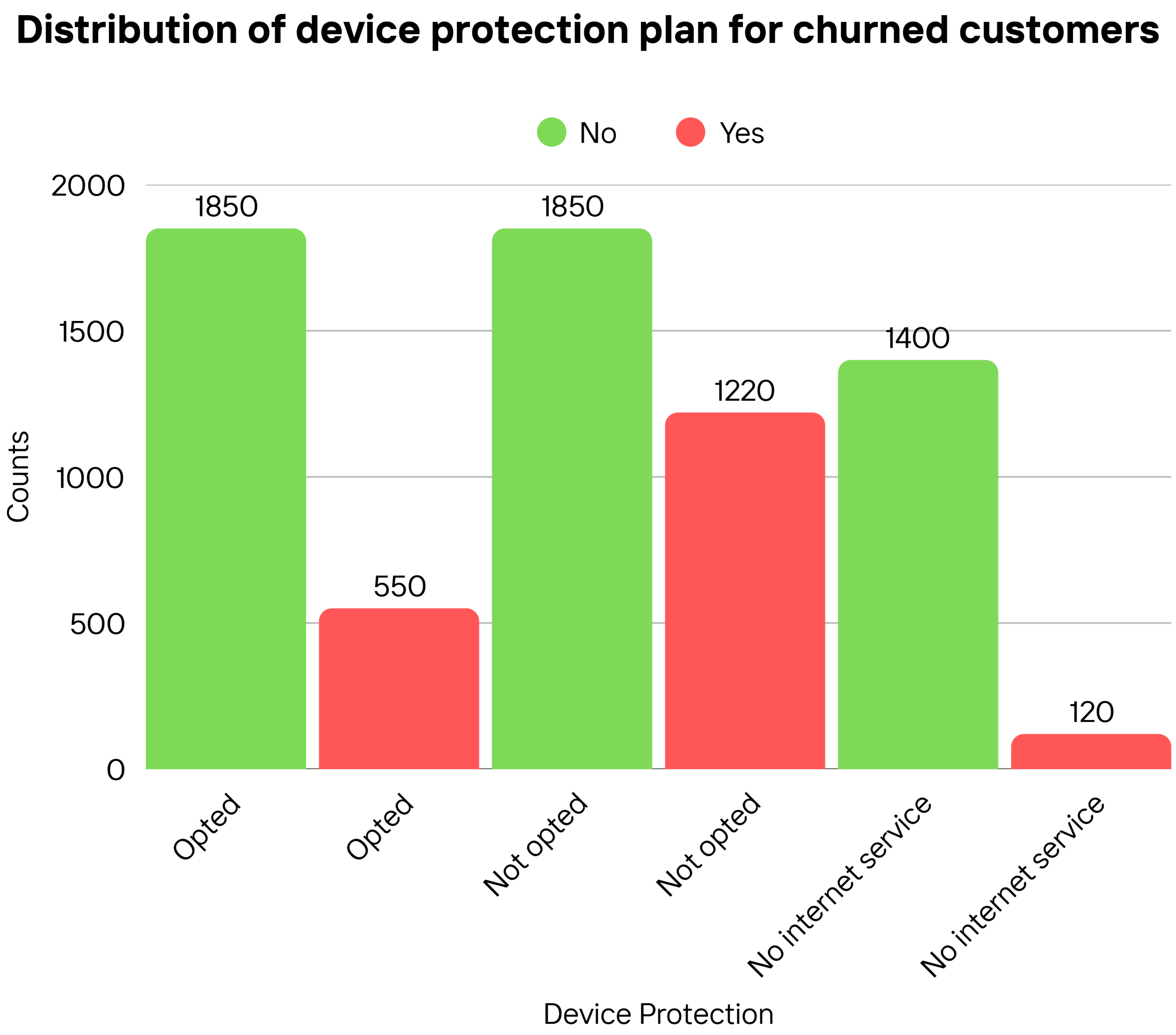
Source code: www.google.research.colab.com

Graph 4

Insights:

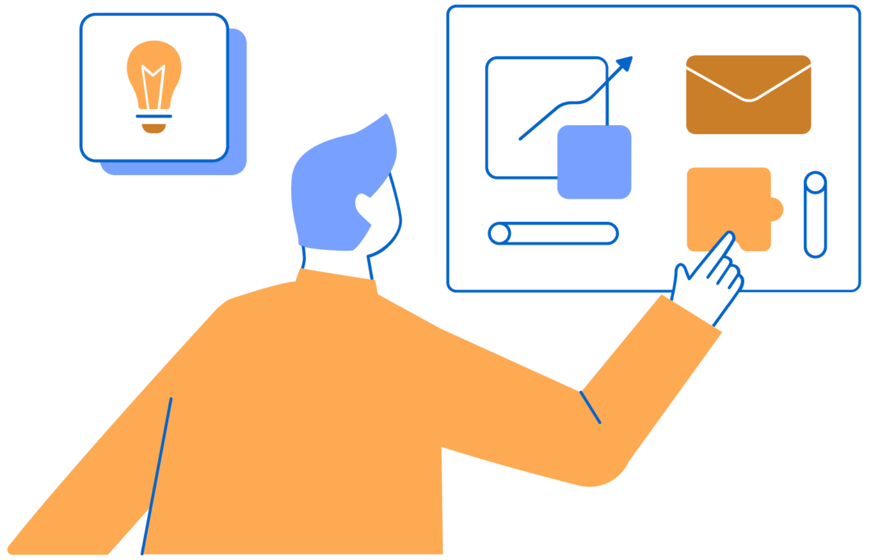
- Customers who "did not opt for device protection" churned more i.e. 1,220, compared to those who opted in i.e. 550.
- Customers with "no internet service" show very low churn here.

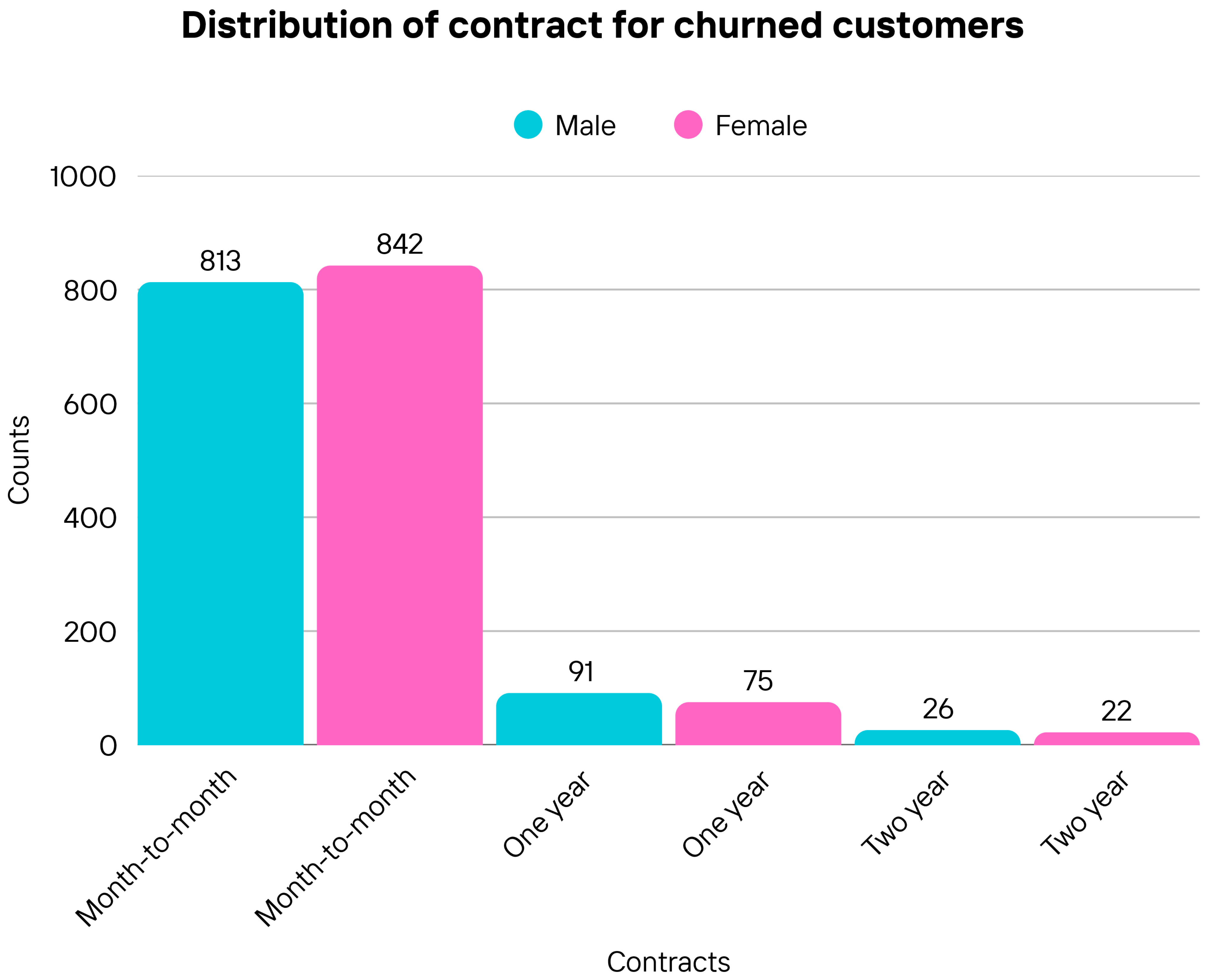




Source code: www.google.research.colab.com

Graph 5



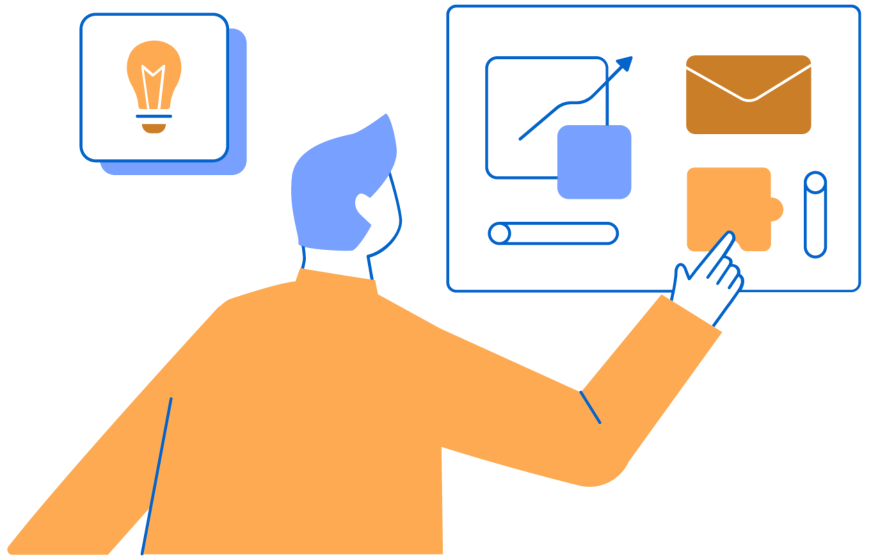


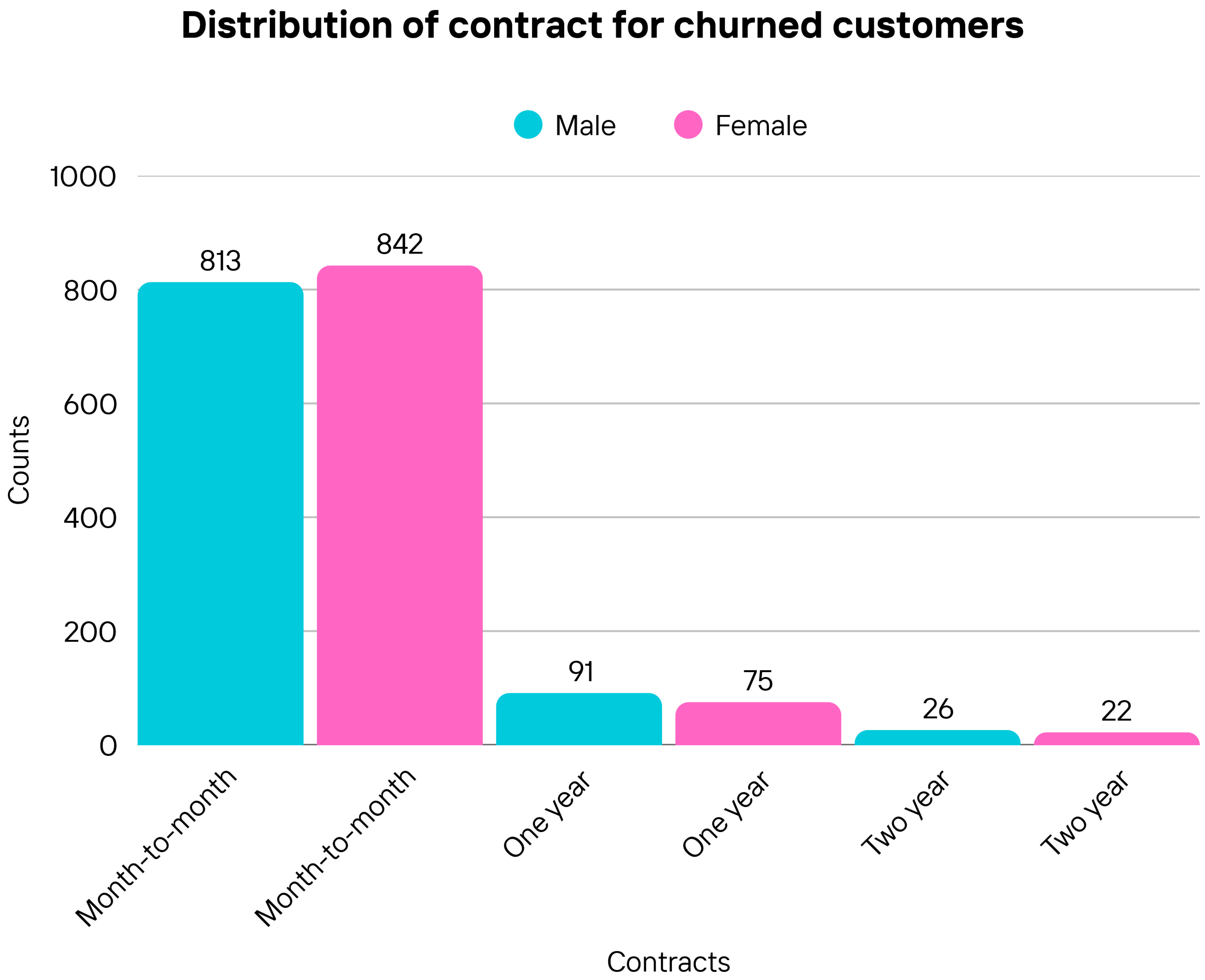
Source code: www.google.research.colab.com

Graph 5

Insights:

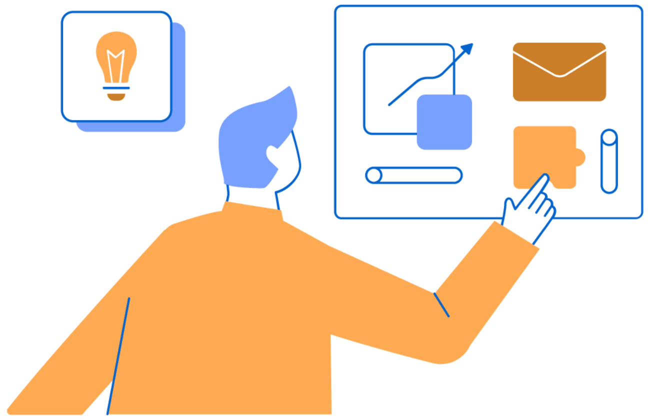
- Month-to-month - contracts account for the “majority of churn” for both males i.e. 813 and females i.e. 842.
- One-year and two-year contracts - show “significantly lower churn”, suggesting long-term contracts may improve retention.



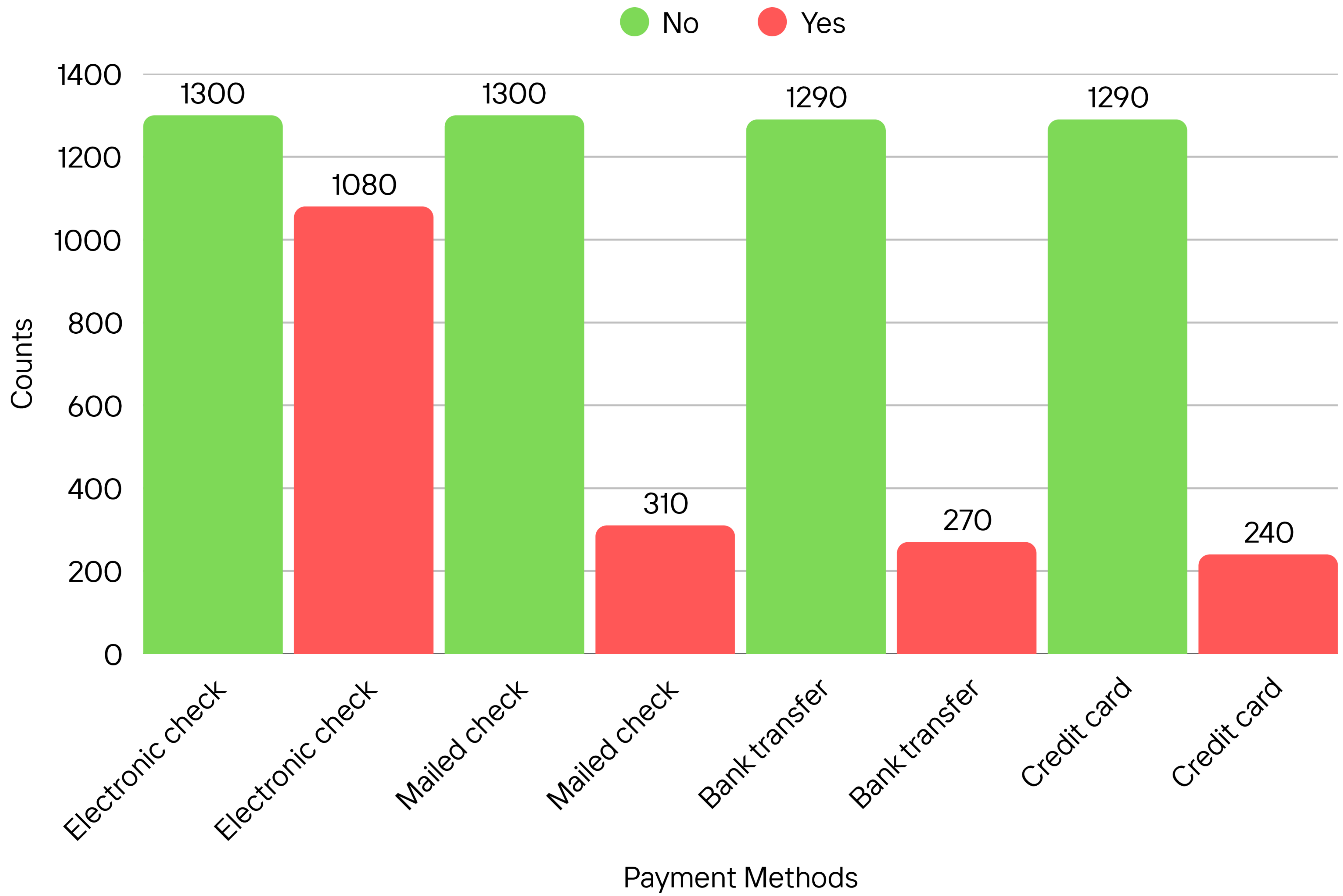


Source code: www.google.research.colab.com

Graph 6



Distribution of payment methods for churned customers

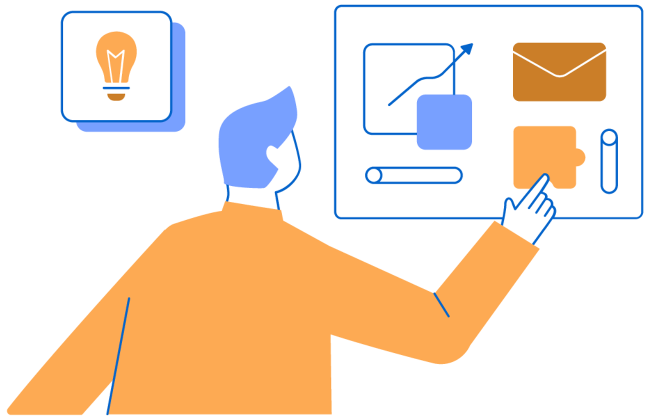


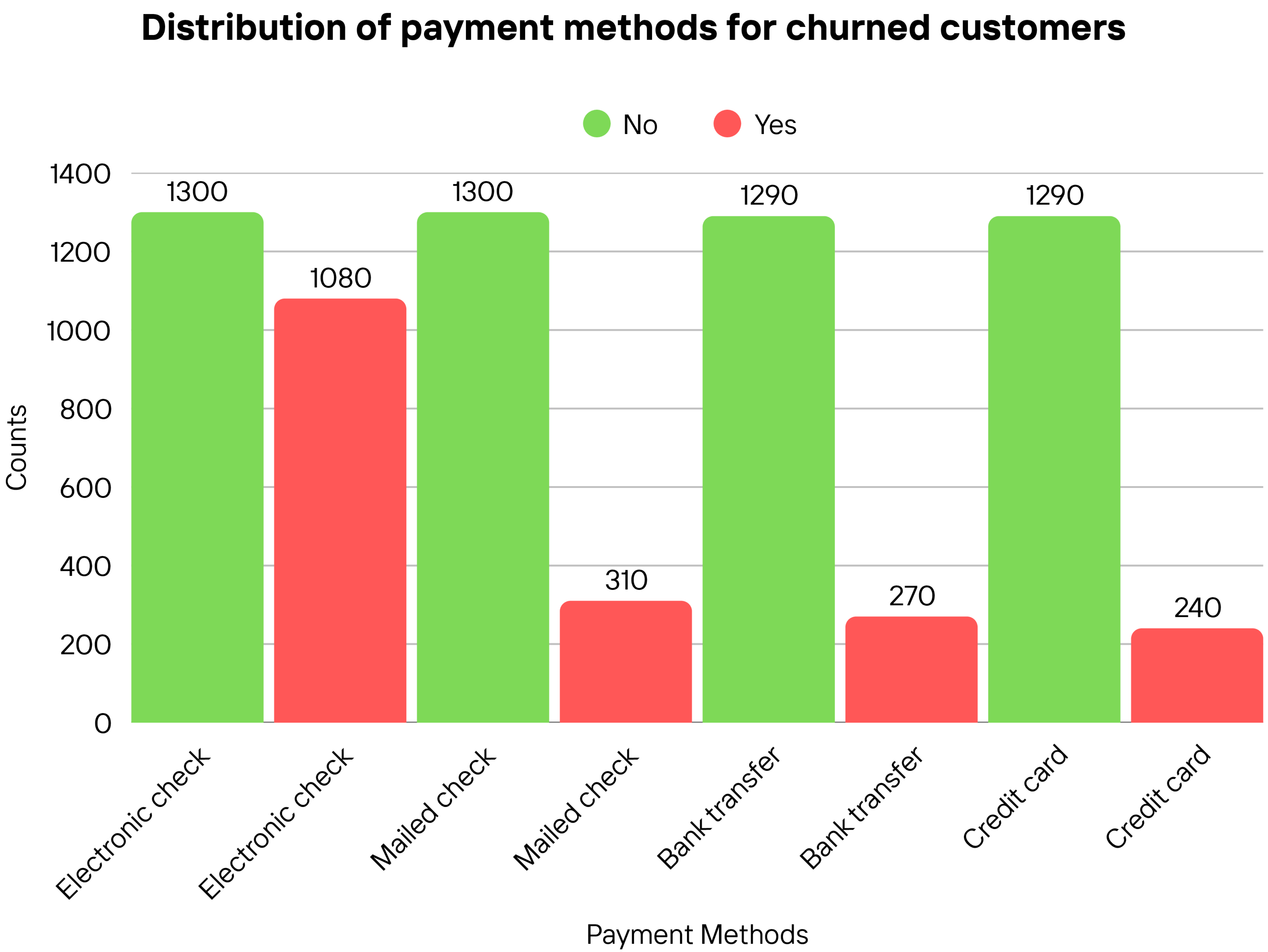
Source code: www.google.research.colab.com

Graph 6

Insights:

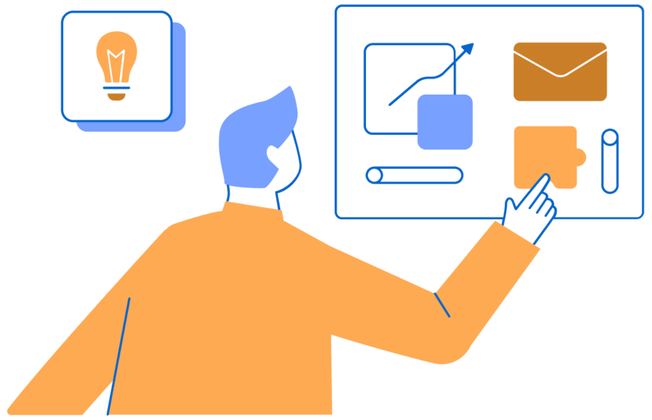
- Customers using “Electronic check” show “the highest churn i.e. 1,080).
- Other payment methods like “Mailed check, Bank transfer, and Credit card” show “much lower churn”, all below 350.
- Convenience and automation in payment might be linked to loyalty.

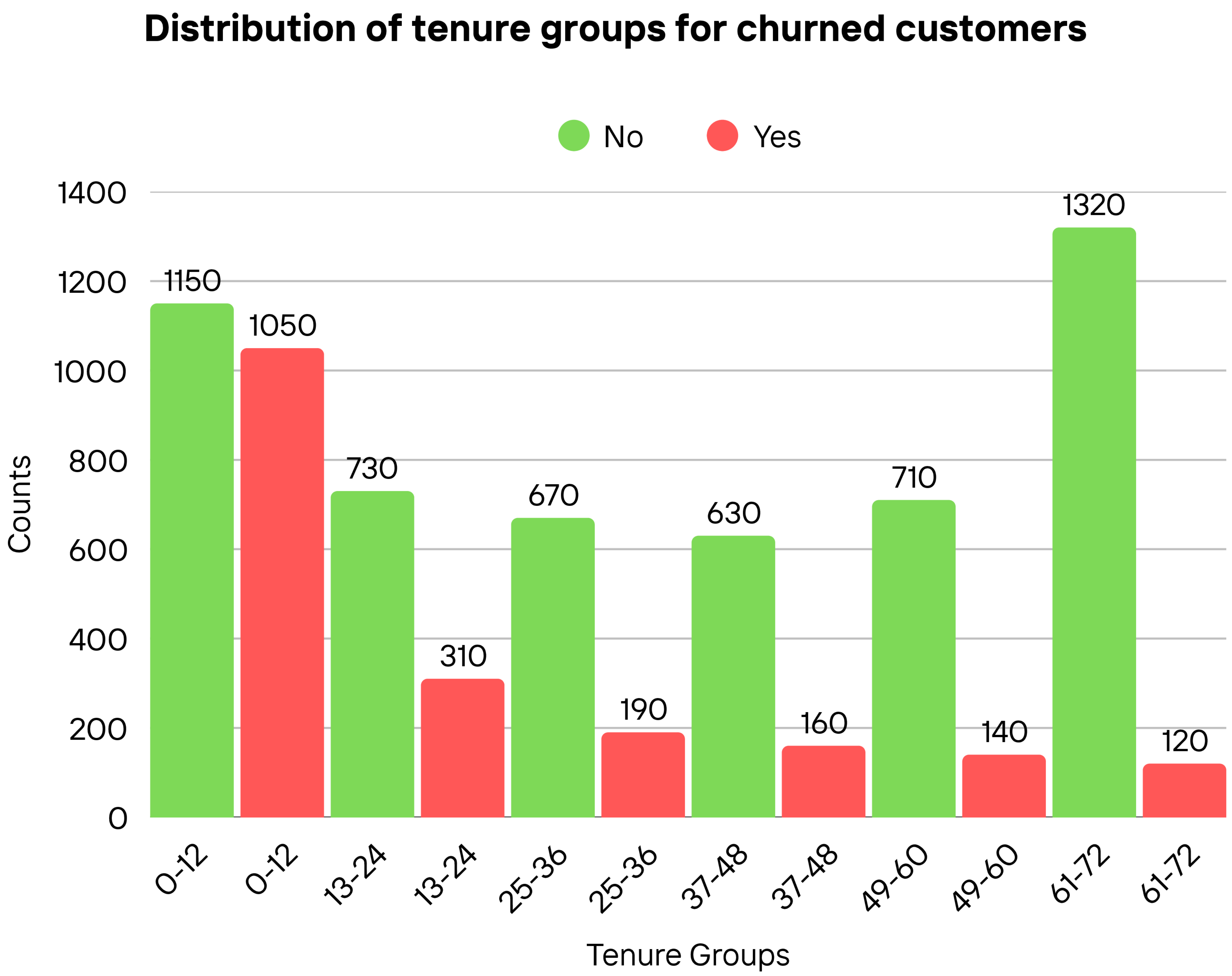




Source code: www.google.research.colab.com

Graph 7



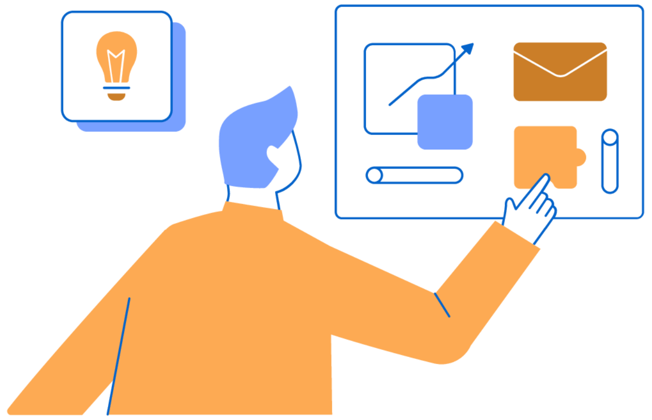


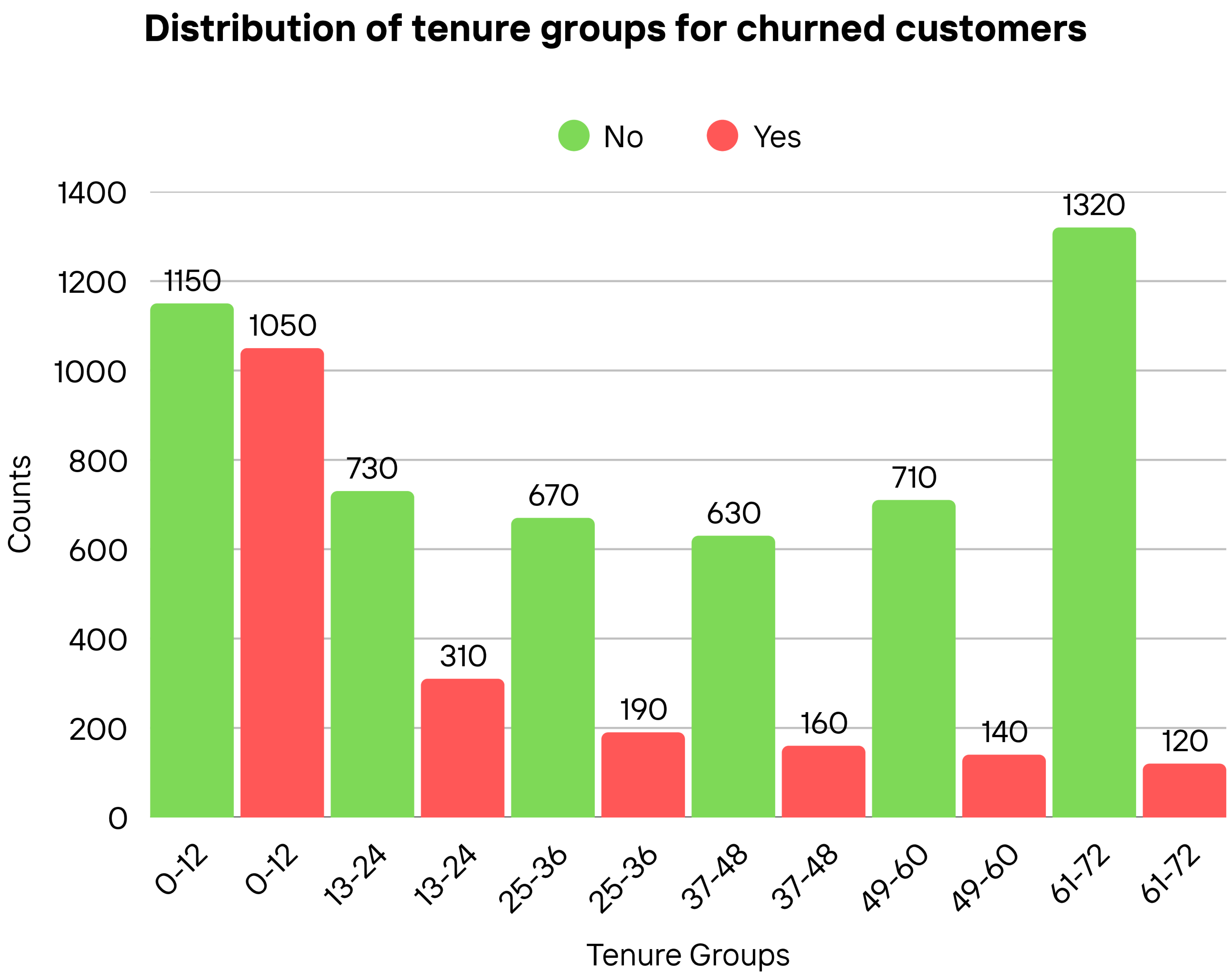
Source code: www.google.research.colab.com

Graph 7

Insights:

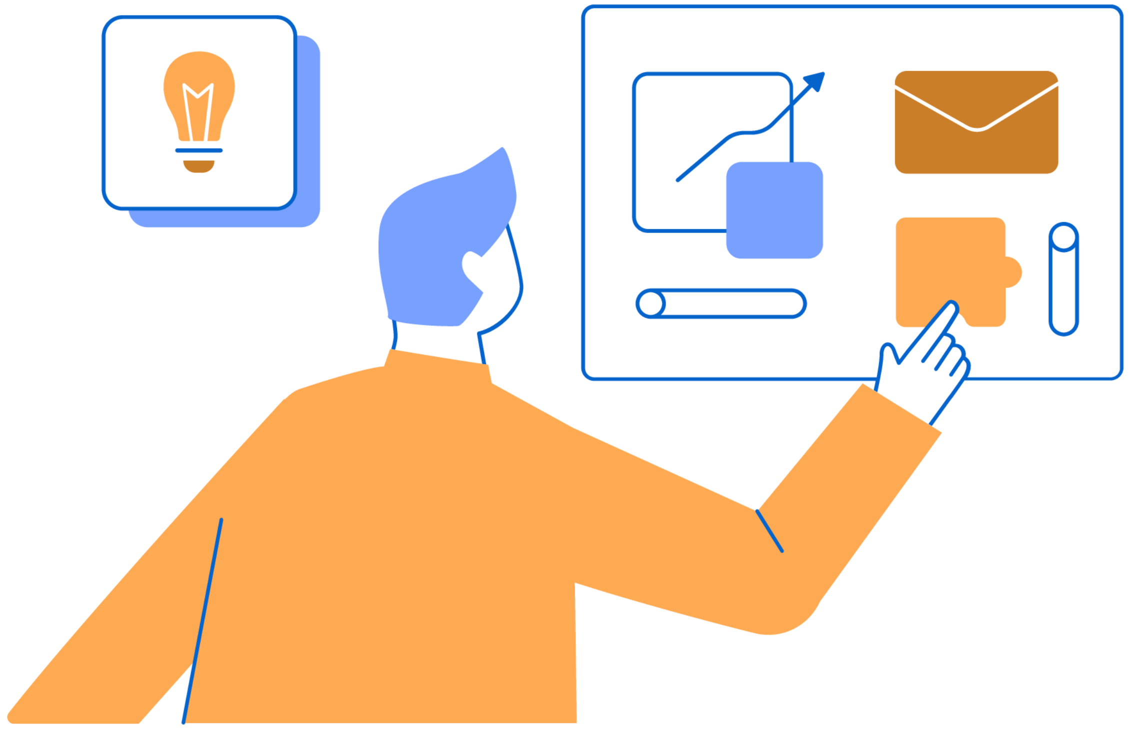
- Churn is “highest in the early months from 0–12 months i.e. 1,050 churned customers.
- Churn “decreases as tenure increases”, suggesting that the longer customers stay, the less likely they are to leave.
- Only 120 customers churned from 61–72 months of tenure group.



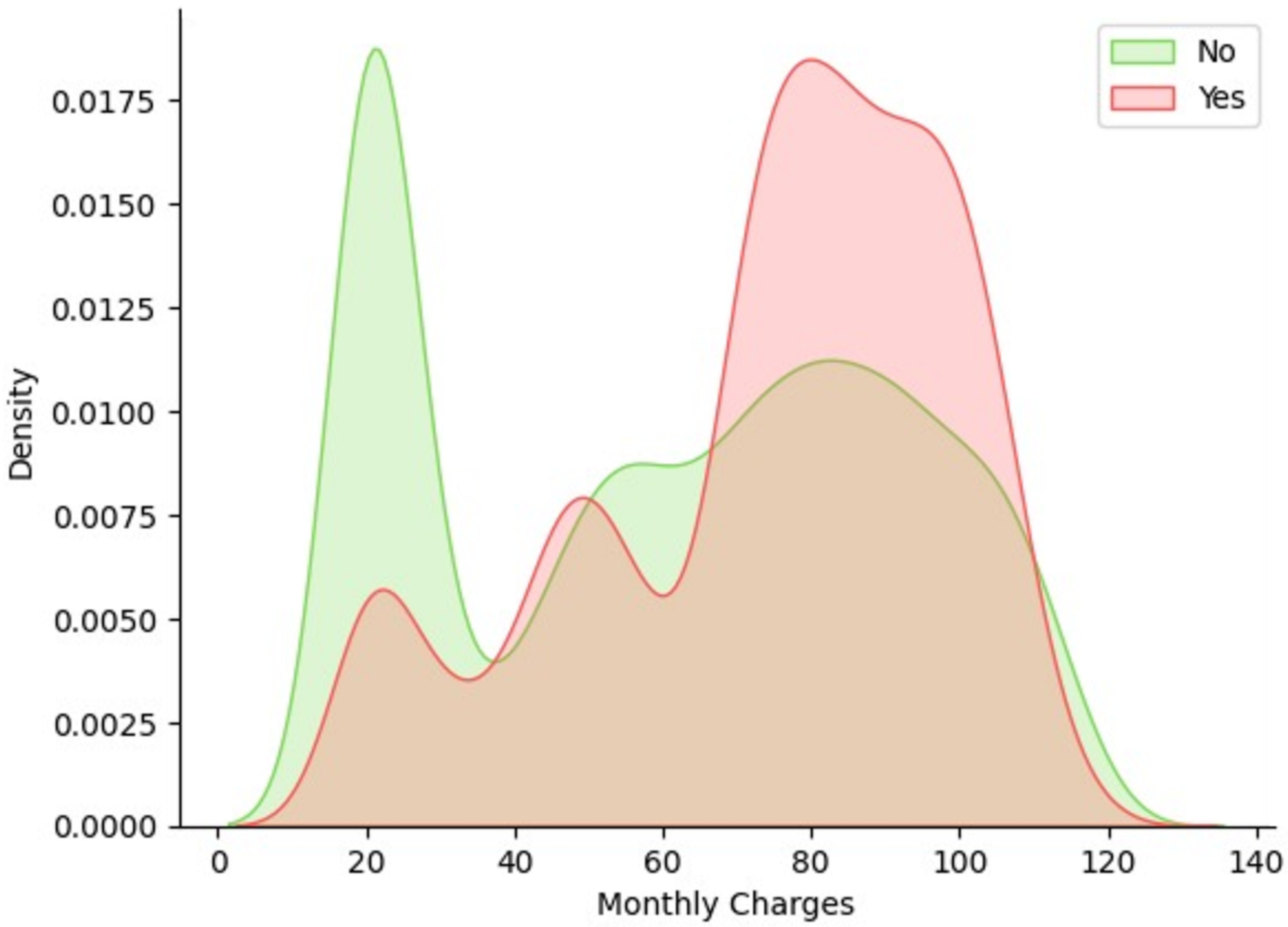


Source code: www.google.research.colab.com

Graph 8



Relationship between monthly charges and customer churn

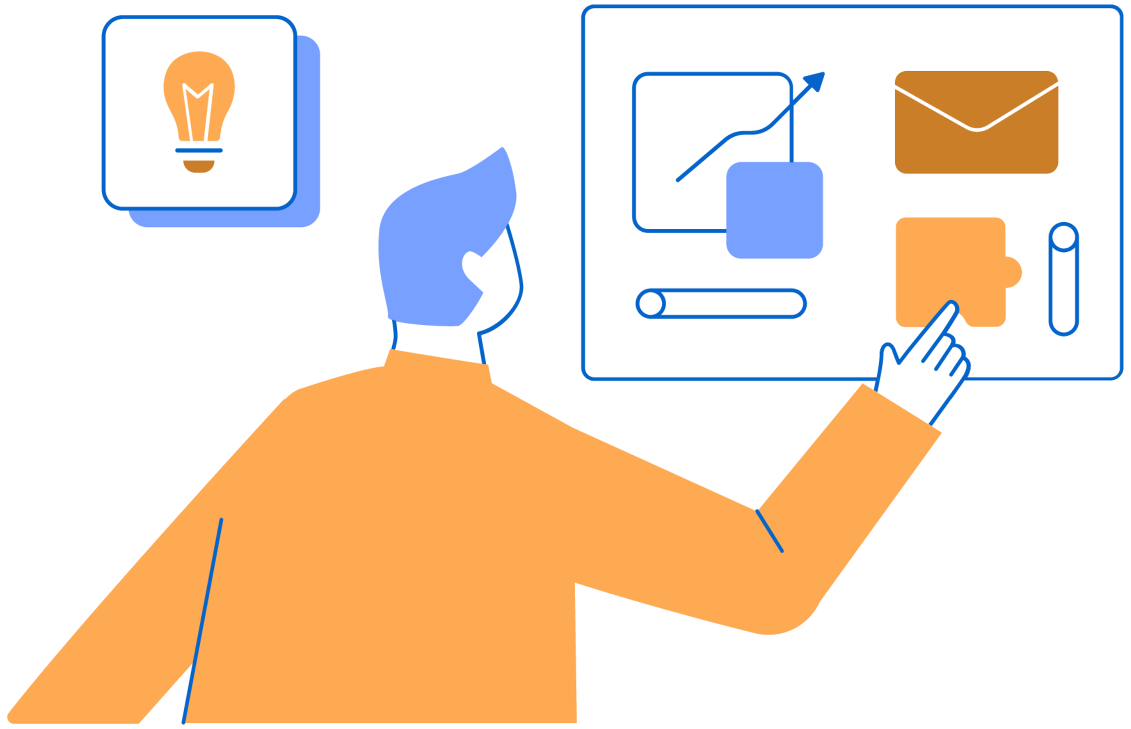


Source code: www.google.research.colab.com

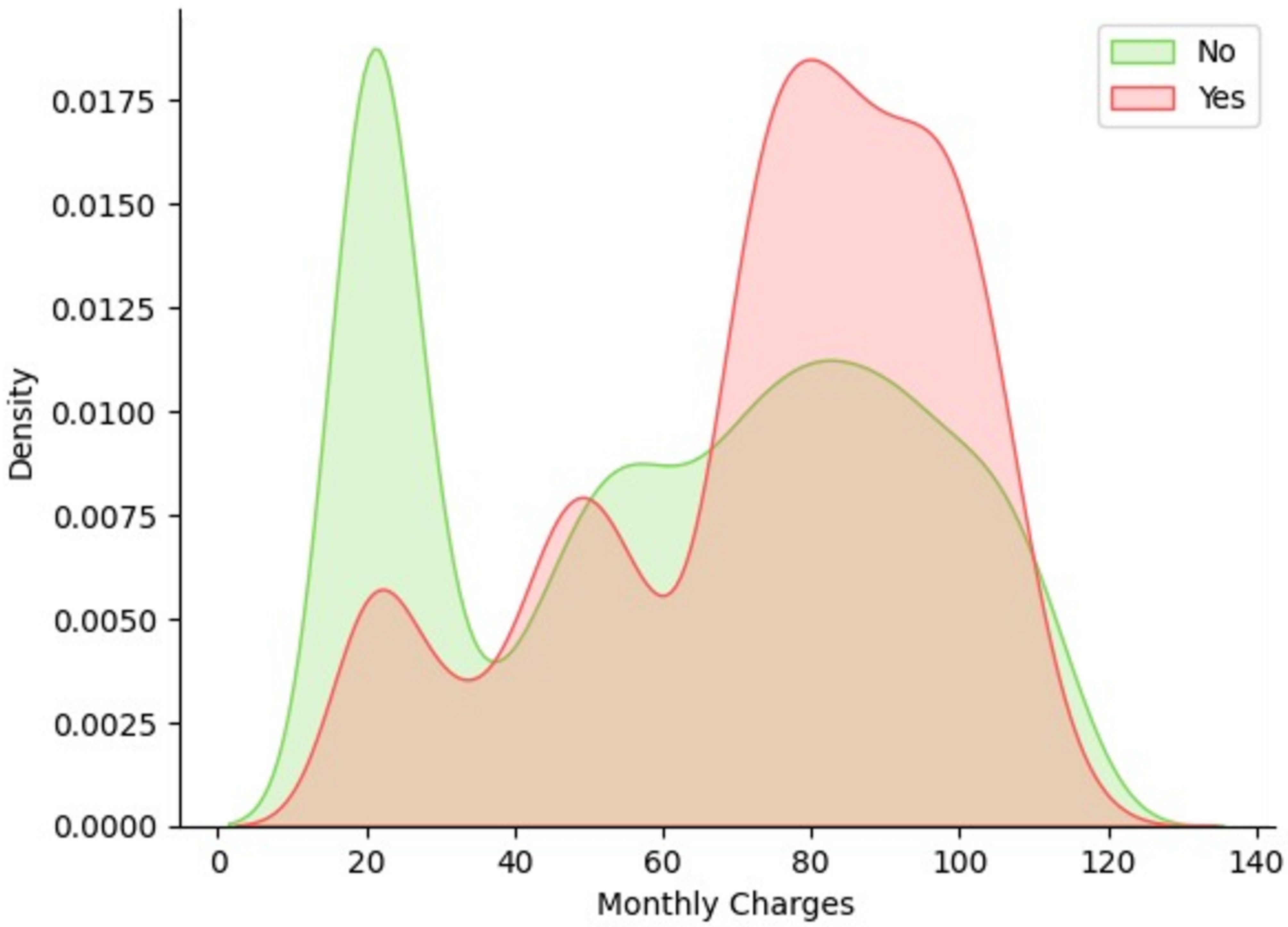
Graph 8

Insights:

- Customers with higher monthly charges i.e. around $75–$90 are more likely to churn, while those paying less than $30 tend to stay longer.

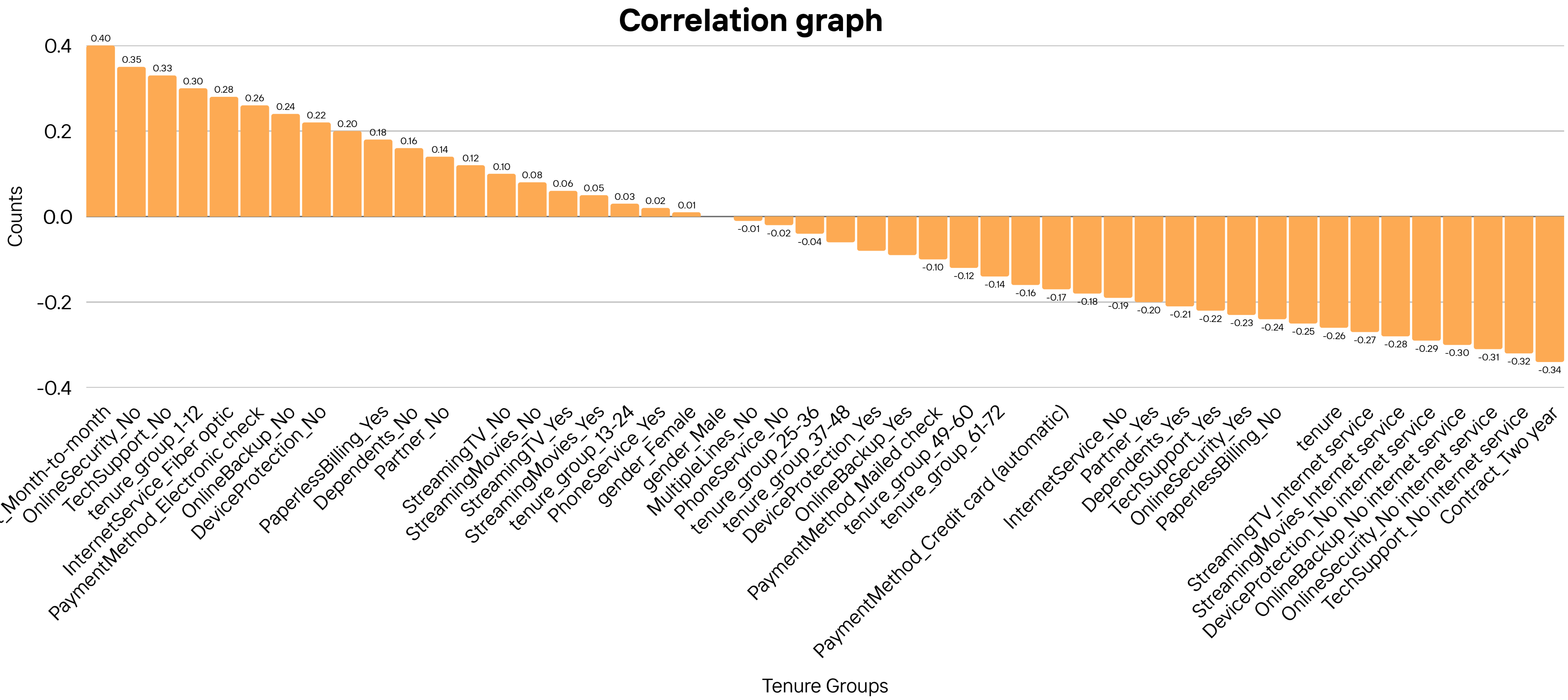


Relationship between monthly charges and customer churn



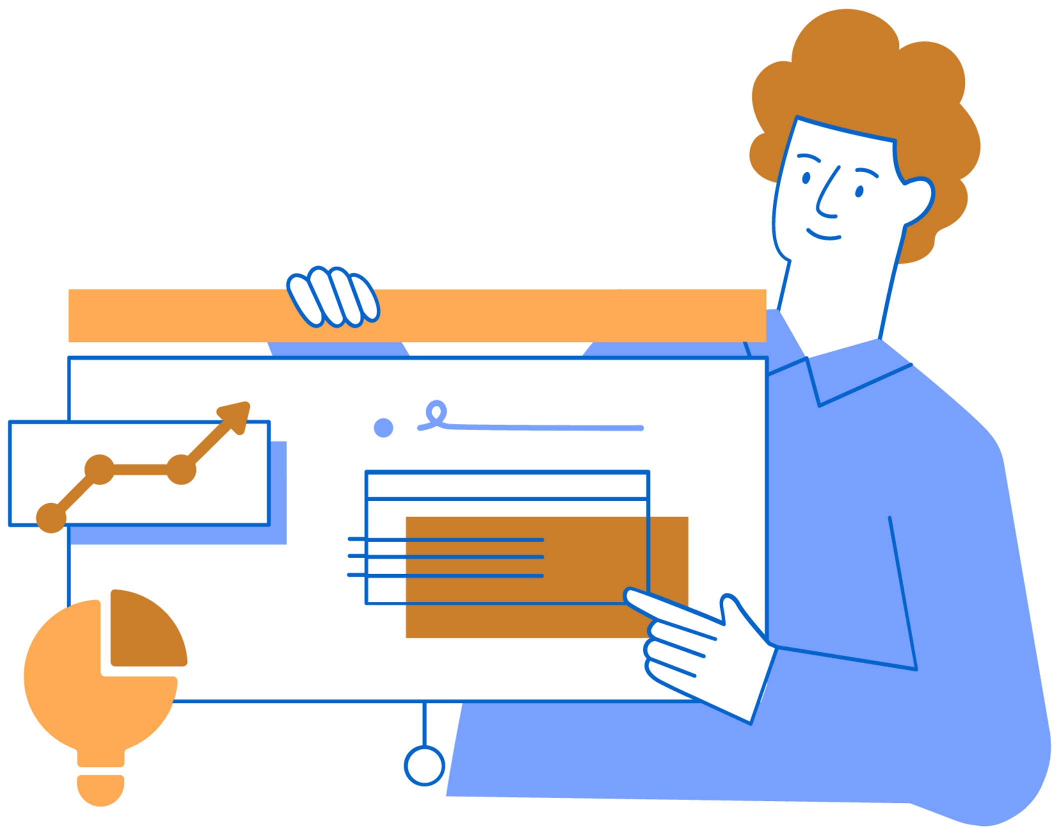
Source code: www.google.research.colab.com

Graph 9



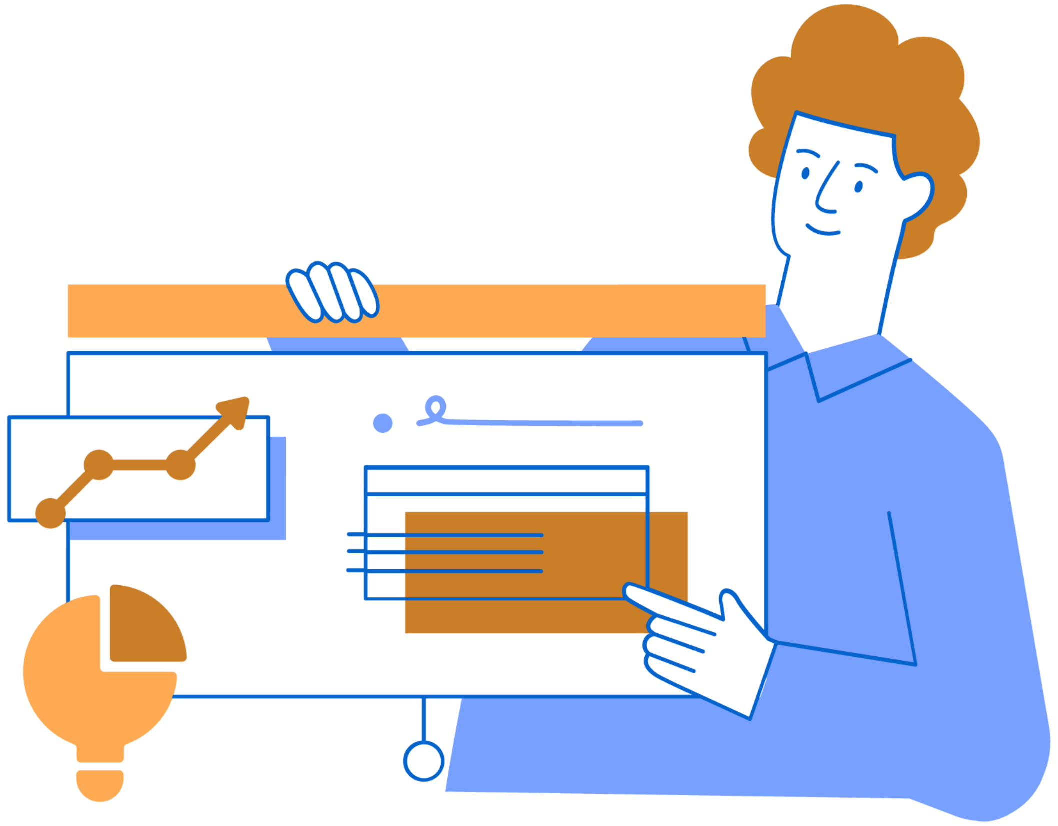
Correlation graph
Counts
Tenure Groups
0.4
0.2
0.0
-0.2
-0.4
Contract_Month-to-month
OnlineSecurity_No
TechSupport_No
tenure_group_1-12
InternetService_Fiber optic
PaymentMethod_Electronic check
OnlineBackup_No
DeviceProtection_No
PaperlessBilling_Yes
Dependents_No
Partner_No
StreamingTV_No
StreamingMovies_No
StreamingTV_Yes
StreamingMovies_Yes
tenure_group_13-24
PhoneService_Yes
gender_Female
gender_Male
MultipleLines_No
PhoneService_No
tenure_group_25-36
tenure_group_37-48
DeviceProtection_Yes
OnlineBackup_Yes
PaymentMethod_Mailed check
tenure_group_49-60
tenure_group_61-72
PaymentMethod_Credit card (automatic)
InternetService_No
Partner_Yes
Dependents_Yes
TechSupport_Yes
OnlineSecurity_Yes
PaperlessBilling_No
tenure
StreamingTV_No internet service
StreamingMovies_No internet service
DeviceProtection_No internet service
OnlineBackup_No internet service
OnlineSecurity_No internet service
TechSupport_No internet service
Contract_Two year

Model Details



Model Details

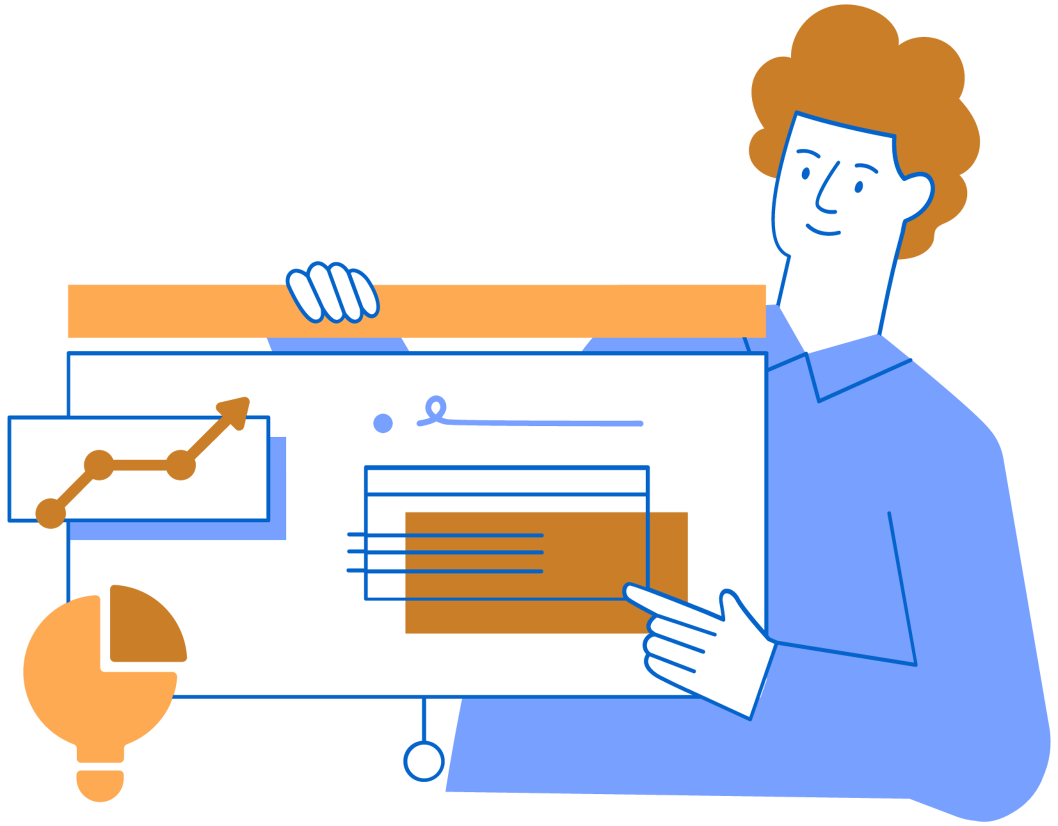
Model	Accuracy	Precision	Recall	F1 Score
Decision tree				
Decision tree (SMOOTEENN)				
Random forest				
Random forest (SMOOTEENN)				
Random forest (PCA)				
Grid search				
LDA (SMOOTEENN)				
Gradient boost				



Source code: www.google.research.colab.com

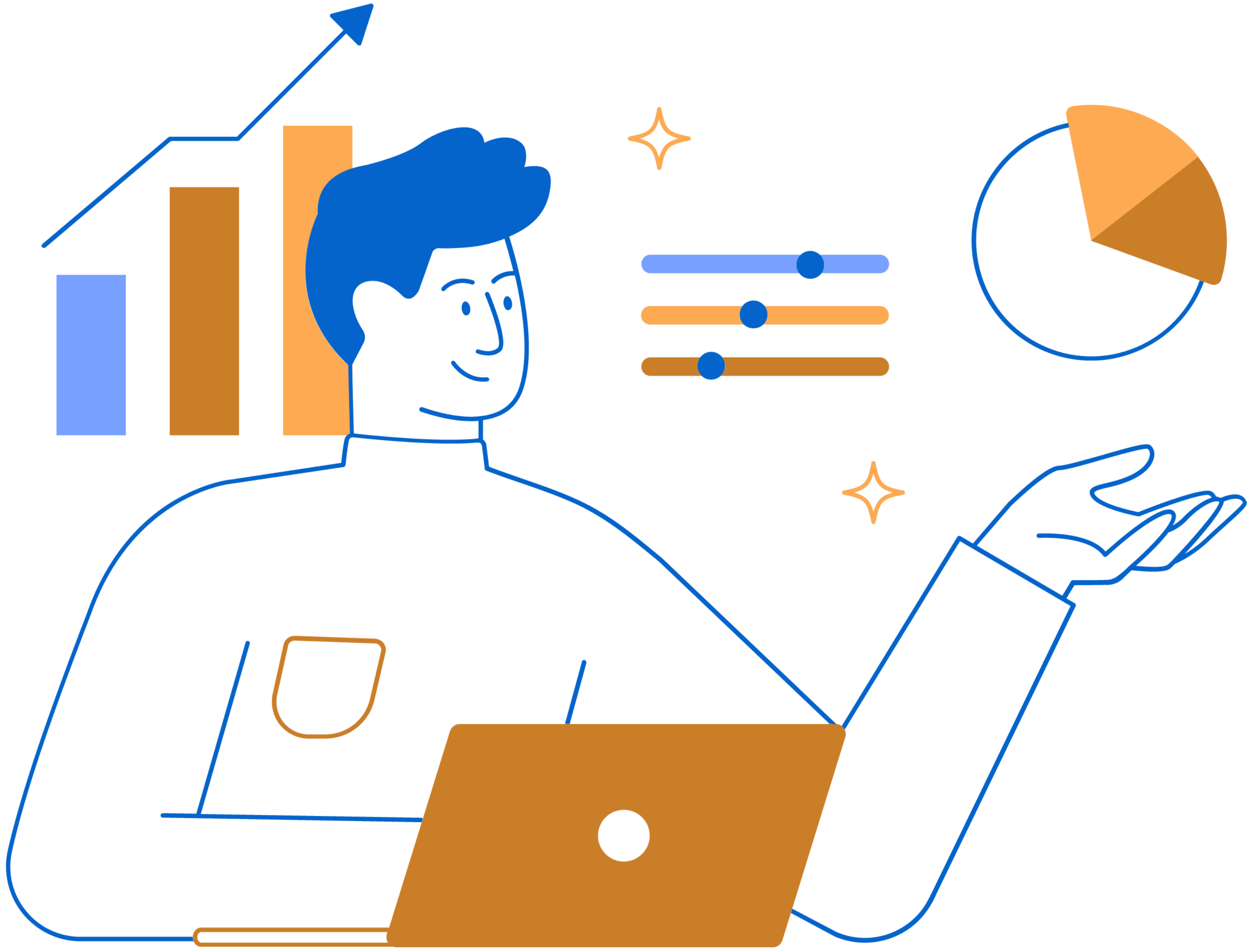
Model Details

Model	Accuracy	Precision	Recall	F1 Score
Decision tree	79%	63%	45%	53%
Decision tree (SMOOTEENN)	93%	94%	95%	94%
Random forest	81%	68%	49%	57%
Random forest (SMOOTEENN)	94%	93%	97%	95%
Random forest (PCA)	72%	73%	78%	75%
Random Forest + Grid search	95%	95%	96%	96%
LDA (SMOOTEENN)	94%	96%	94%	95%
Gradient boost	96%	96%	96%	96%

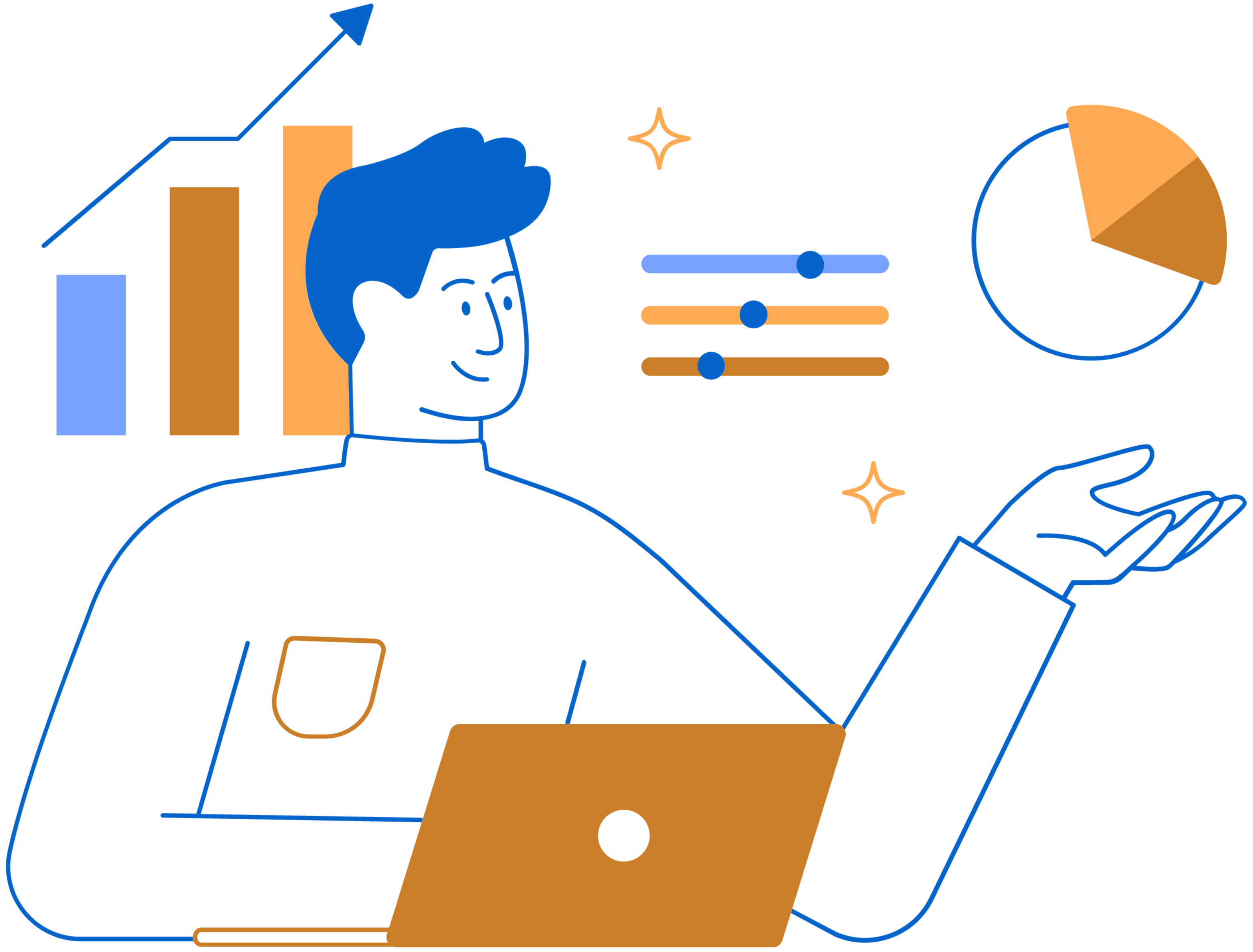


Source code: www.google.research.colab.com

Recommendations



Source code: www.google.research.colab.com



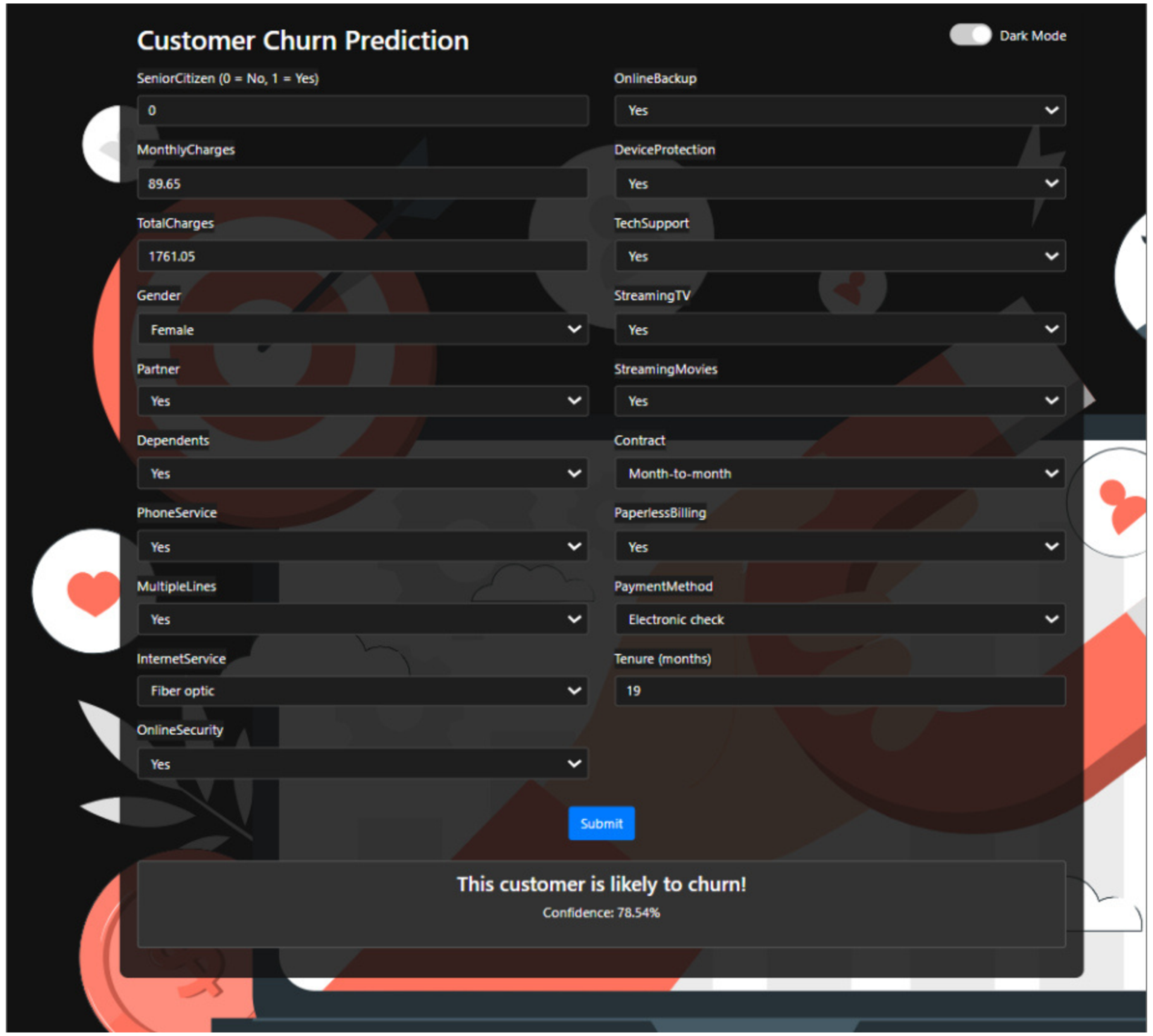
Recommendations

Actions to reduce churn:

- Incentivize long-term contracts.
- Target users lacking tech support or online security.
- Use churn prediction tools regularly in customer service.

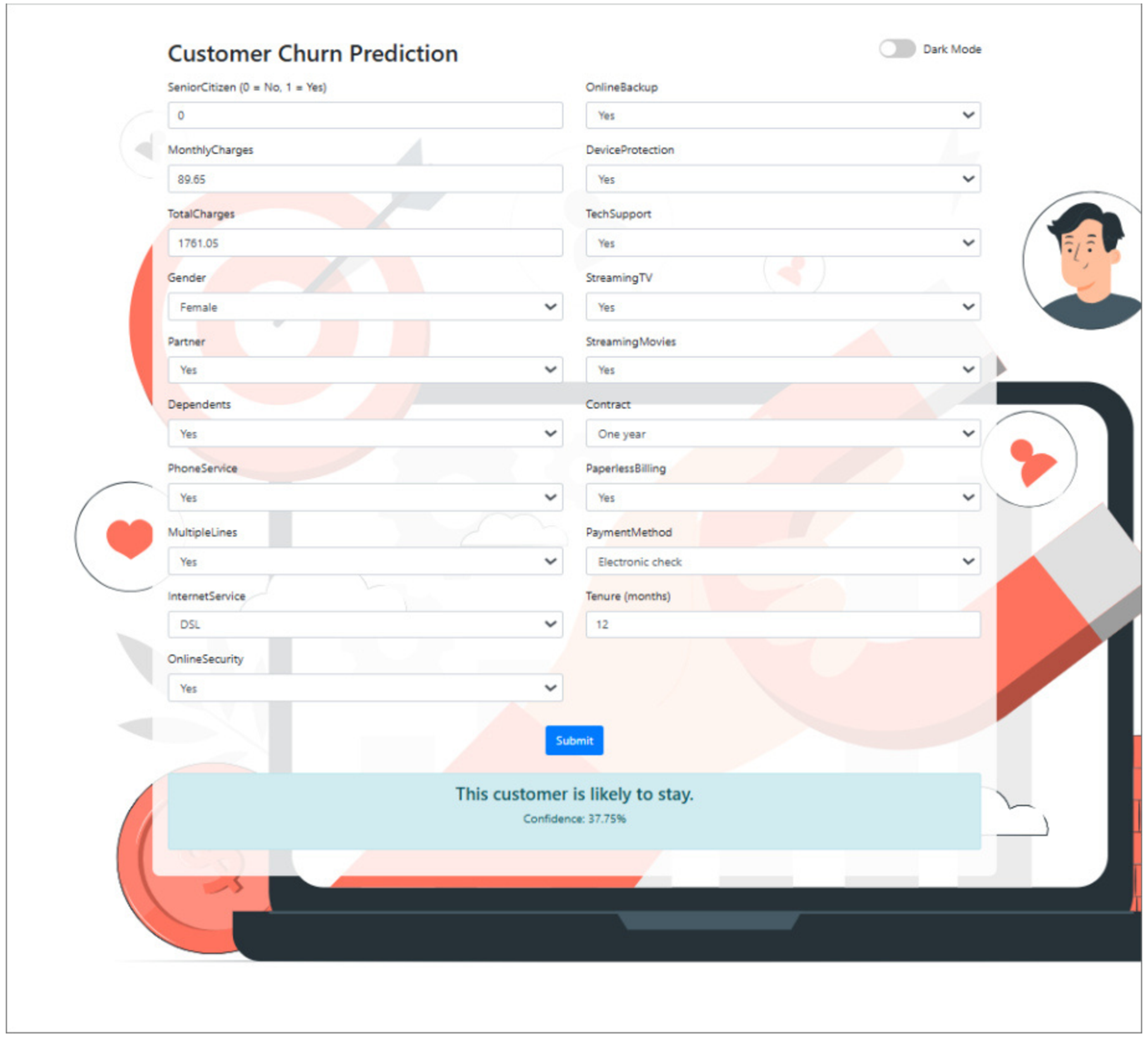
Source code: www.google.research.colab.com

Web App Implementation



Source code: www.google.research.colab.com

Web App Implementation



Source code: www.google.research.colab.com

Thank You

Thank You

Questions ?