The Telco Customer Churn Analysis & Modeling

Study Case Exam

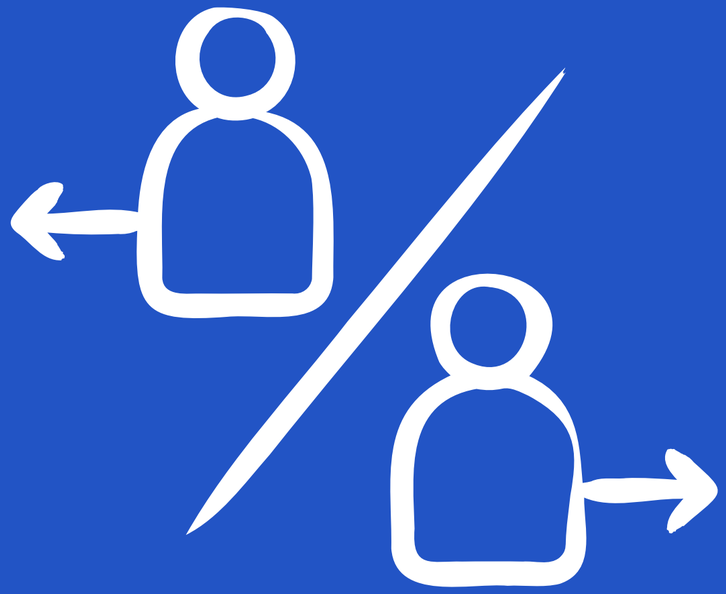
Almira Z.

A bachelor's degree in management, and was in an exchange program at the University of Applied Sciences Utrecht, contributing to **consulting projects for EDAG Group, Google Merchandise Store, and Ørsted.** Hands-on experience in market research, strategy development, and competitor analysis as a Junior Consultant Intern at EDAG Group. With strong analytical and communication abilities and proficiency in Microsoft Office, Google Suite, Canva, and Google Analytics.

About Dataset

The Telco Customer Churn dataset represents customer-level data from a fictional **telecommunications company that provided home phone and Internet services to customers in California during the third quarter (Q3).** The dataset **consists of 7,043 customer records,** each capturing **whether a customer churned (left the company), remained active during the observed period.**

This dataset is designed to support customer behavior analysis and retention strategy development.



Source: https://www.kaggle.com/datasets/hassanelfattmi/why-do-customers-leave-can-you-spot-the-churners/data

Dataset

Column Name	*Data Type*	*Descriptions*
satisfaction_score	int64	Customer satisfaction score; higher values indicate higher satisfaction
cltv	int64	Customer Lifetime Value – estimated total value of the customer to the company
churn_value	int64	Target variable: 1 = customer churned, 0 = customer stayed
tenure	int64	Length of time the customer has been with the company (in months)
multiple_lines	object	Indicates whether the customer has multiple phone lines
avg_monthly_gb_download	int64	Average monthly internet data usage (in gigabytes)
unlimited_data	object	Indicates whether the customer has an unlimited data plan
referred_a_friend	object	Indicates whether the customer has referred at least one friend
number_of_referrals	int64	Number of friends referred by the customer
online_security	object	Indicates whether the customer subscribes to online security services

Column Name	*Data Type*	*Descriptions*
online_backup	object	Indicates whether the customer subscribes to online backup services
device_protection	object	Indicates whether the customer subscribes to device protection services
premium_tech_support	object	Indicates whether the customer subscribes to premium technical support
streaming_tv	object	Indicates whether the customer uses streaming TV services
streaming_movies	object	Indicates whether the customer uses streaming movie services
streaming_music	object	Indicates whether the customer uses streaming music services
internet_type	object	Type of internet connection (e.g., DSL, Fiber, Cable)
contract	object	Customer contract type (Month-to-month, One year, Two year)
payment_method	object	Method used by the customer to pay their bills
monthly_ charges	float64	Monthly amount charged to the customer

Data Preparation

1. Duplicated & Missing Value

0 Duplicated Data and 1.526 Missing Values, *thus Missing Value need to be dropped*

2. Outliers

1.868 Outliers, and we drop it.
Data Before : 5.517
Data After. : 3.649

3. Normalization

Using ***Min-Max Scaler Method***

4. Featuring

Using ***Heatmap Correlation*** *to see the correlation between the numerical variables*

THE TELCO CUSTOMER CHURN ANALYSIS

ALMIRA Z.

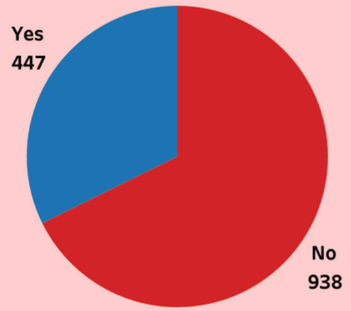
Churn	Score	Category	Satisfaction	CLV (Customer Lifetime Value)	Tenure
1,385	81.88	Dissatisfaction	1.66	$4,159	18.32

Churn Category
- Attitude
- Competitor
- Dissatisfaction

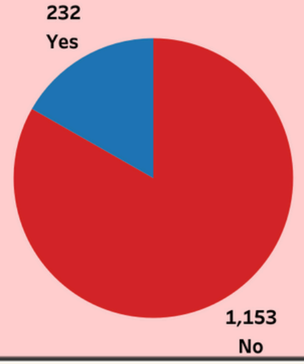
Churn Category
- ✔ Attitude
- ✔ Competitor
- ✔ Dissatisfaction
- ✔ Not Applicable
- ✔ Other

Limit
Top 3 by SUM([Churn Value])

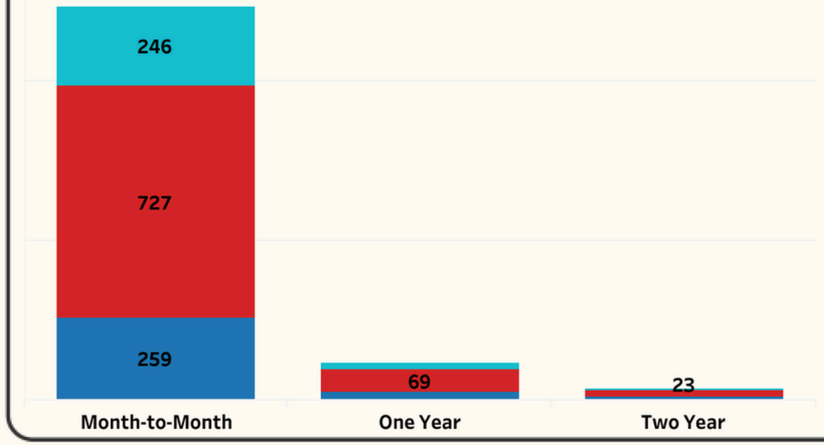
Device Protection



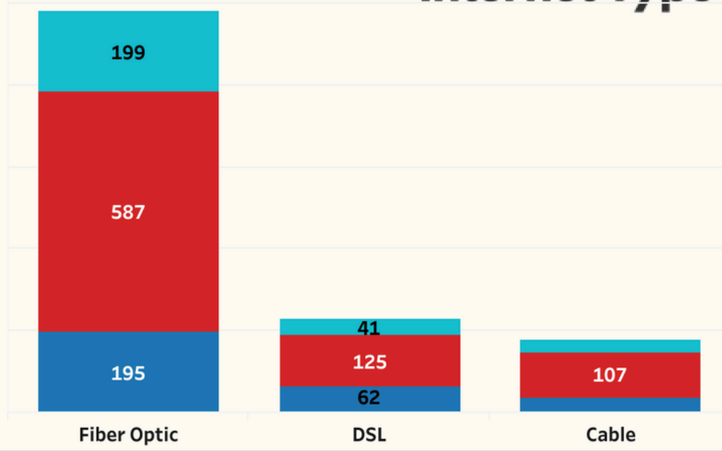
Premium Tech Support



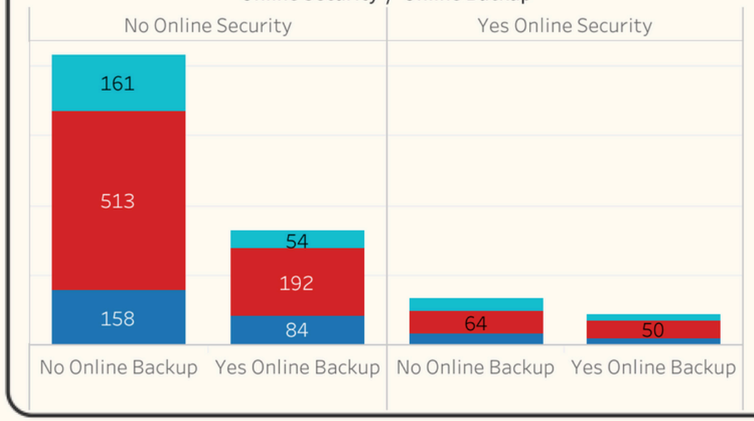
Contract



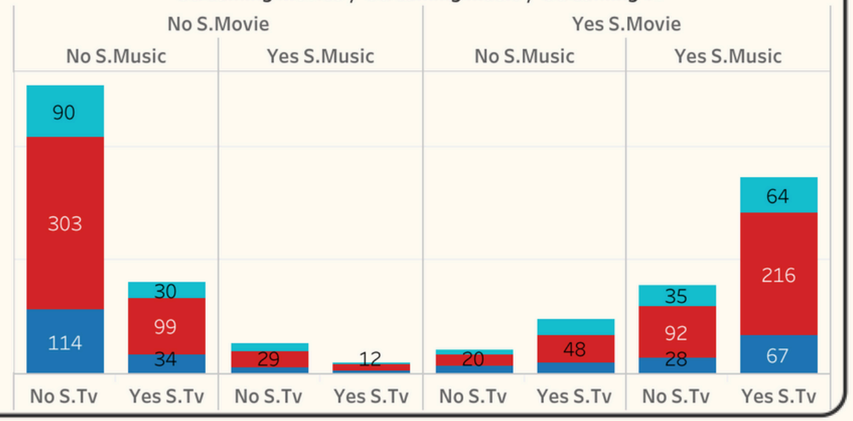
Internet Type



Online Service



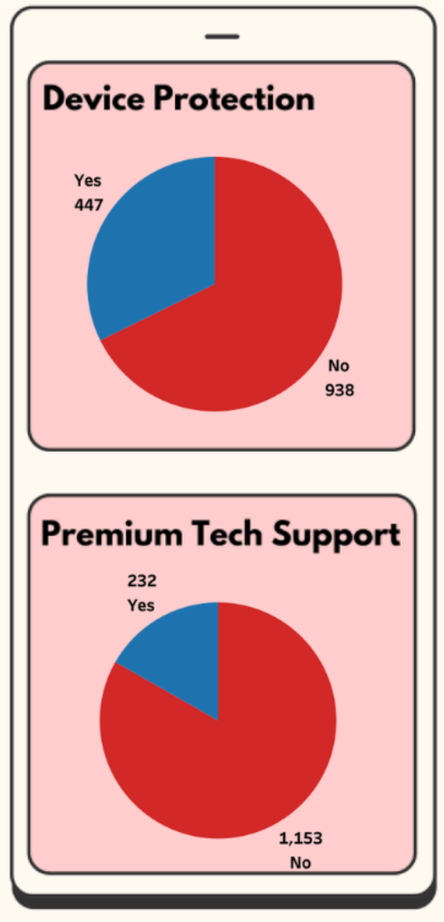
Streaming



The KPI

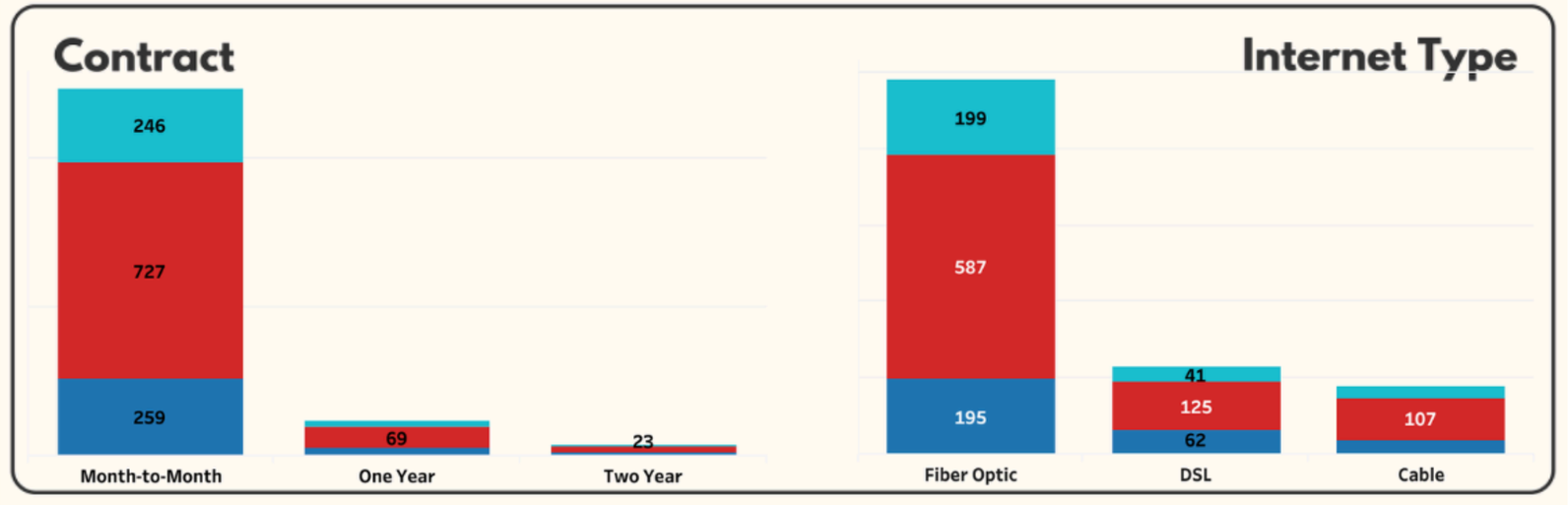
- **High churn volume (1,385 customers)** indicates a significant retention challenge.
- **Low satisfaction score (1.66)** is the primary driver of churn, with most cases classified as Dissatisfaction.
- **High churn risk score (81.88)** confirms that churners are accurately identified as high-risk customers.
- **High average CLV ($4,159)** shows that churned customers still represent substantial business value.
- **Short average tenure (18.32 months)** suggests churn occurs early in the customer lifecycle.

Value-Added Services & Churn



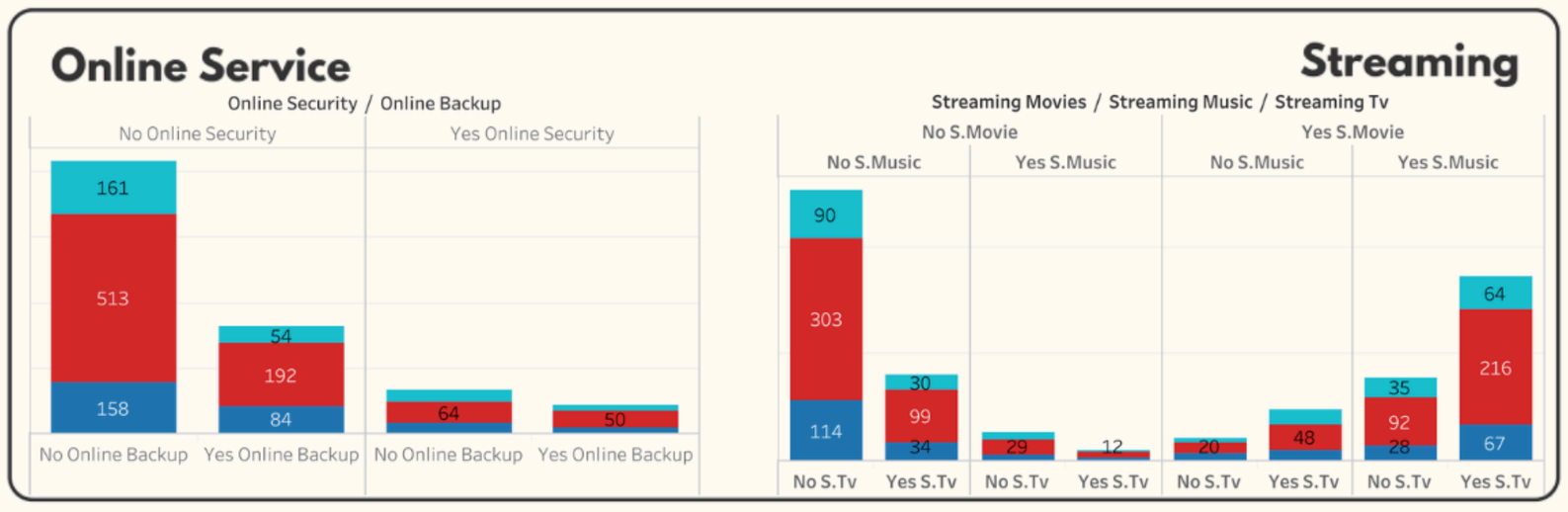
- **Most churned customers do not subscribe to Device Protection (938 vs 447),** indicating higher churn risk among customers without basic protection services.
- **Lack of Premium Tech Support is strongly associated with churn, with 1,153 churners not using the service, compared to only 232 who do.**
- **Customers who adopt value-added support services are significantly more likely to stay,** suggesting these services increase perceived reliability and trust.

Contract Type & Internet Type



- **Month-to-month contracts account for the majority of churn,** far exceeding one-year and two-year contracts, indicating significantly higher churn risk among customers without long-term commitment.
- **Customers with longer contracts (one-year and two-year) show much lower churn,** highlighting contract length as a strong retention factor.
- **Fiber Optic users represent the largest share of churn,** despite being a premium service, suggesting potential issues with service quality, pricing expectations, or customer experience.
- **DSL and Cable customers churn less compared to Fiber Optic,** indicating that higher-speed offerings do not automatically translate into higher satisfaction.

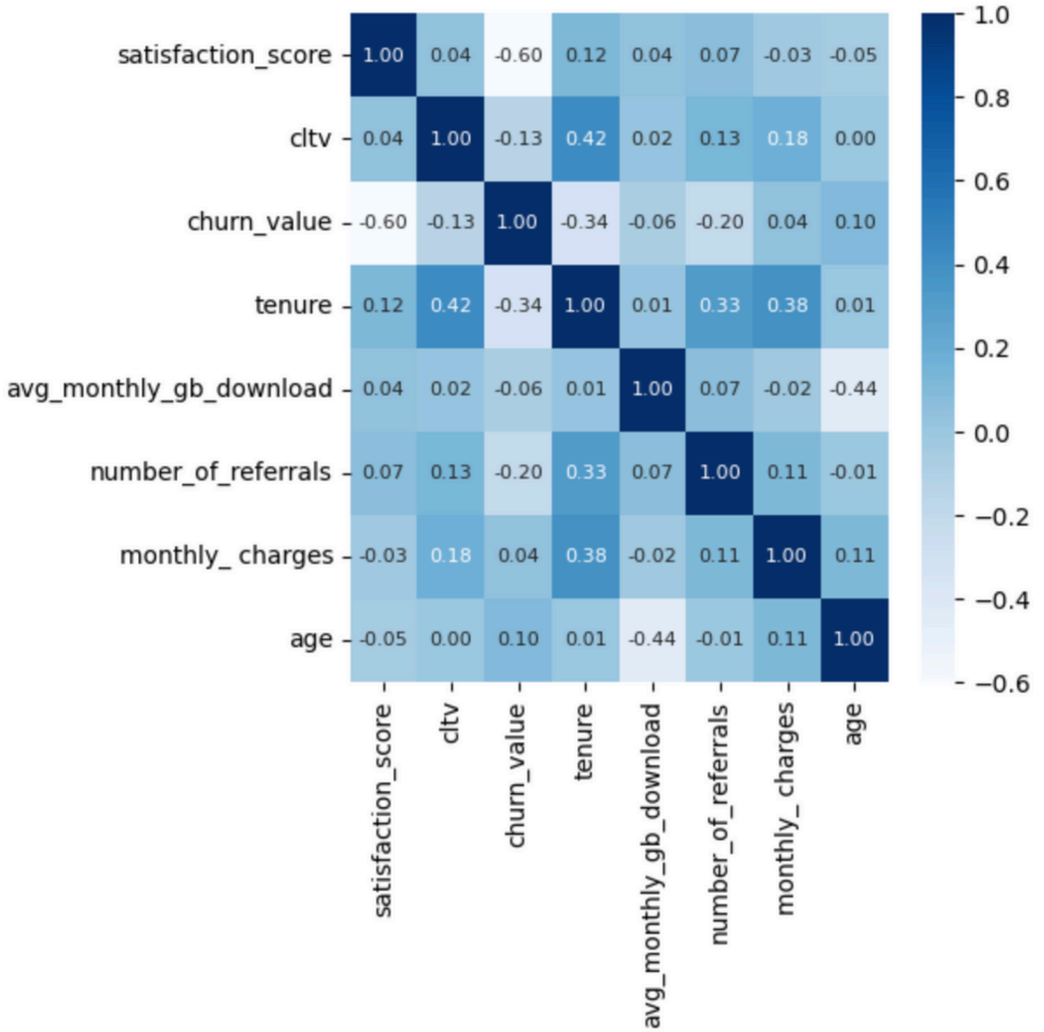
Online Services & Streaming Behavior



- **Customers without Online Security and Online Backup show significantly higher churn,** indicating that lack of basic digital protection increases churn risk.
- **Churn is substantially lower among customers who subscribe to both Online Security and Online Backup,** highlighting these services as strong retention drivers.
- **Customers with no streaming services (TV, Movies, Music) churn more frequently,** suggesting lower engagement among customers with fewer bundled services.
- **Streaming service adoption is associated with lower churn,** particularly when customers subscribe to multiple streaming options.

Correlation Heatmap

- **Customer satisfaction is the strongest driver of churn (r = -0.60):** customers with low satisfaction are significantly more likely to churn.
- **Shorter-tenure customers have a higher churn risk (r = -0.34),** highlighting the importance of early-stage retention.
- **Customer engagement reduces churn:** customers who refer others are less likely to leave (r = -0.20).
- **No multicollinearity issue detected:** no feature pairs exceed the correlation threshold of |0.8|, so all numerical variables are retained for modeling.



Modeling

Train & Test Split

Data Train : 2.847
Data Test : 802

Tuning Hyperparameter

Only for the KNN Model
with ***K - Value : 3***

Class Imbalance

Label 0 : 2.919
Label 1 : 730

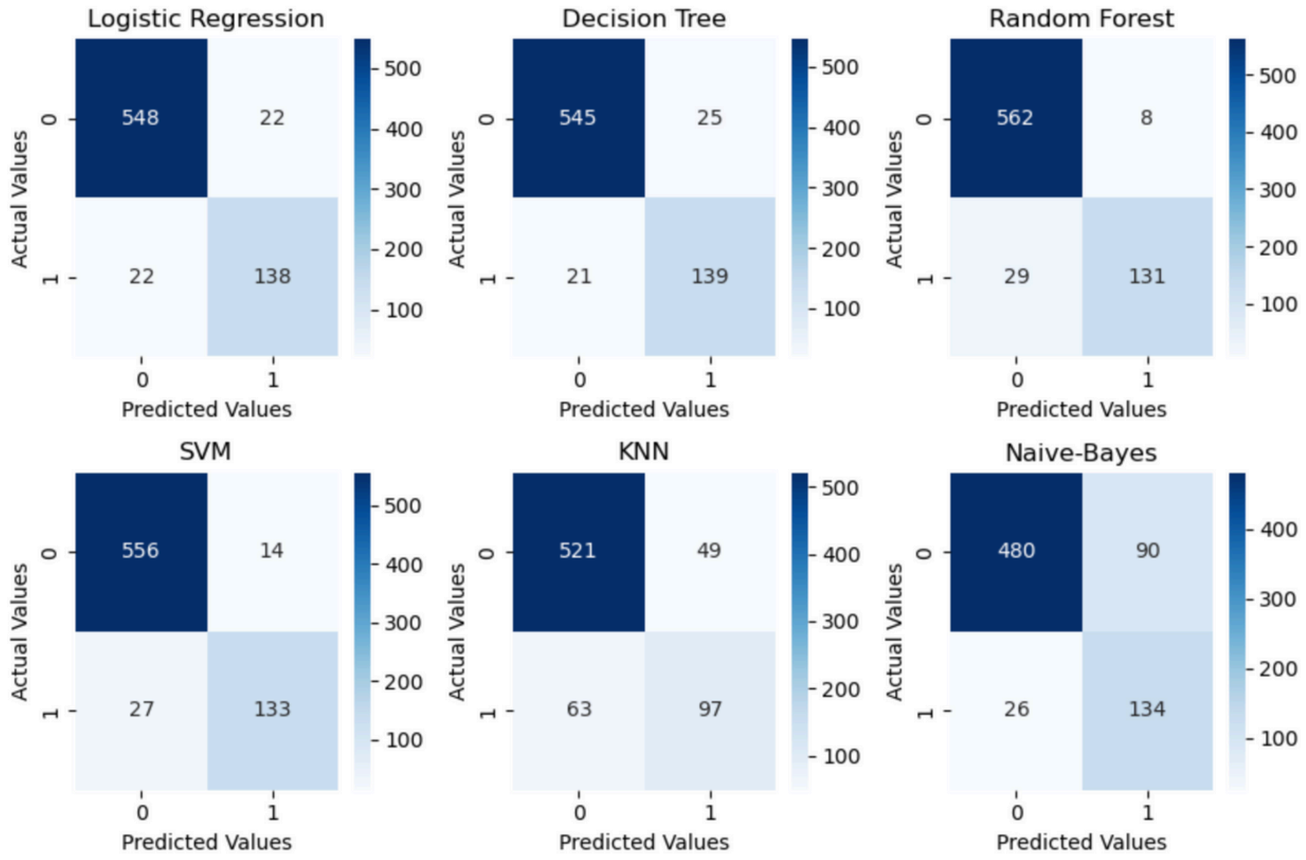
Needs Balancing by
Handle Class Imbalance:
2.277

Training a Modeling

KNN : 92,2%
Logistic : 93,2%
Decision Tree: 100%
Random Forest : 100%
Naive-Bayes: 84,9%
SVM: 94,7%

Modeling

Model	Akurasi	Precision	Recall	F1 Score
Logistic Regression	0.940	0.862	0.862	0.862
Decision Tree	0.937	0.848	0.869	0.858
Random Forest	0.949	0.942	0.819	0.876
Support Vector Machine	0.944	0.905	0.831	0.866
K-Nearest Neighbor	0.847	0.664	0.606	0.634
Naive-Bayes	0.841	0.598	0.838	0.698



EVALUATION

Best Overall Model

Random Forest

Accuracy 0.949, F1-score 0.876. Delivers the strongest overall predictive performance and the fewest false positives.
Confusion Matrix: TN = 562, FP = 8, FN = 29, TP = 131
Best suited when overall accuracy and prediction reliability are the main priorities, though it misses more churners than some alternatives.

Stable & Interpretable Model

Logistic Regression

Accuracy 0.940, with balanced precision and recall (0.862). Reliable, stable, and easy to interpret.
Confusion Matrix: TN = 548, FP = 22, FN = 22, TP = 138
Provides a strong balance between churn detection and interpretability, making it suitable for business-driven decision-making.

Best Model for Churn Detection (High Recall)

Decision Tree

Accuracy 0.937, with the highest recall (0.869) among all models.
Confusion Matrix: TN = 545, FP = 25, FN = 21, TP = 139
Most effective at identifying churners and minimizing false negatives, making it the best choice when the priority is to avoid missing churn cases.

High-Precision Model

Support Vector Machine

AAccuracy 0.944, precision 0.905, indicating strong confidence in positive predictions.
Confusion Matrix: TN = 556, FP = 14, FN = 27, TP = 133
Performs well but captures fewer churners than Logistic Regression and Decision Tree, with limited interpretability.

Less Suitable Model

K-Nearest Neighbors

Accuracy 0.847, with weak balance between precision and recall.
Confusion Matrix: TN = 521, FP = 49, FN = 63, TP = 97
Misses a large number of churners, making it unsuitable when churn detection is critical.

Not Suitable Model

Naive Bayes

Accuracy 0.841, with unstable and inconsistent predictions.
Confusion Matrix: TN = 480, FP = 90, FN = 26, TP = 134
Produces a high number of incorrect classifications and lacks reliability for churn prediction.

Recommendation

Since **the primary business objective is to maximize the identification of churners and minimize false negatives, recall is selected as the main evaluation metric.** Among all evaluated models, **the Decision Tree model achieves the highest recall (0.869) and the lowest number of false negatives (21),** making it the most effective model for detecting customers who are likely to churn.

Although the Random Forest model delivers the best overall accuracy and F1-score, it misses more churners compared to Decision Tree and Logistic Regression. Therefore, it is less suitable when the cost of missing churners is high. Logistic Regression provides a strong alternative with slightly lower recall but greater stability and interpretability, which is valuable for understanding churn drivers and supporting business decisions.

Thank You

Churn is not the end of a relationship; it is feedback we failed to act on in time.

We are ready to assist you

Canva Link

Dashboard Link

Github Link