

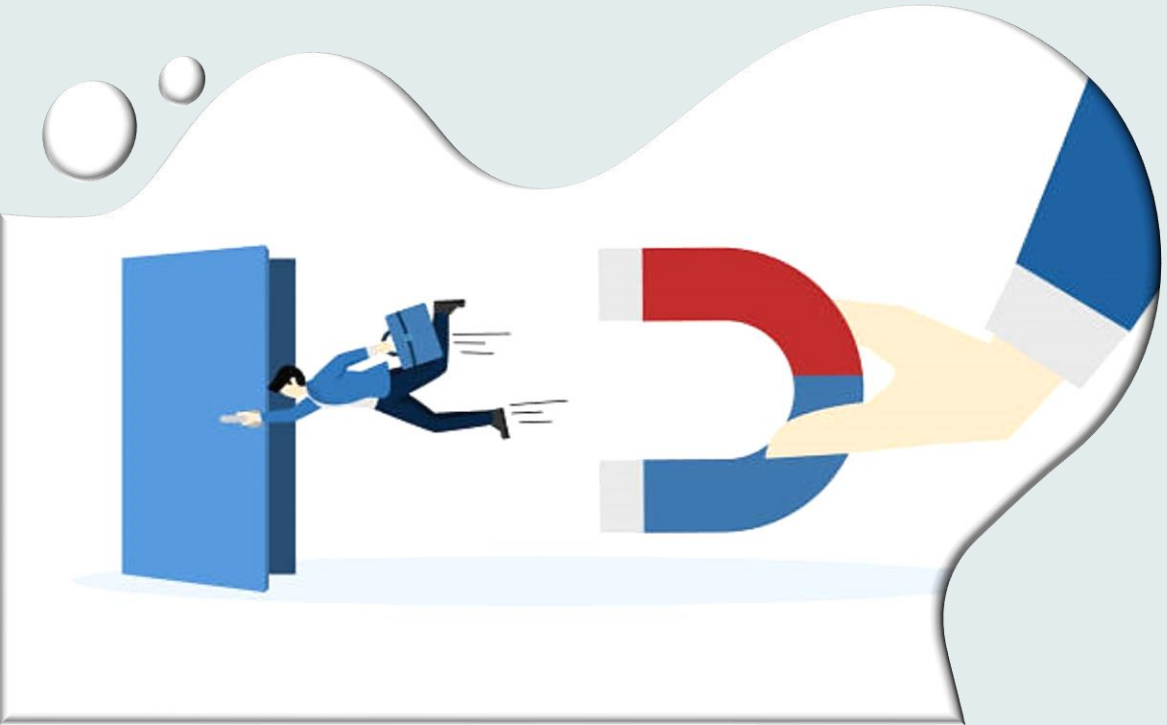
E-Commerce Customer Churn Analysis

Pamela Gatica

Marketing Analytics

September 2025

Introduction



Analyze the customer base of an e-commerce platform to understand churn.

Identify the key characteristics of customers who have stopped using the service (churn).

Use these insights to design strategies that target at-risk customers and reduce churn.



Dataset Overview

Dataset Size
- 5,630 rows (Customers)
- 20 columns (Features)

Target Variable
Churn (0 = No, 1= Yes)

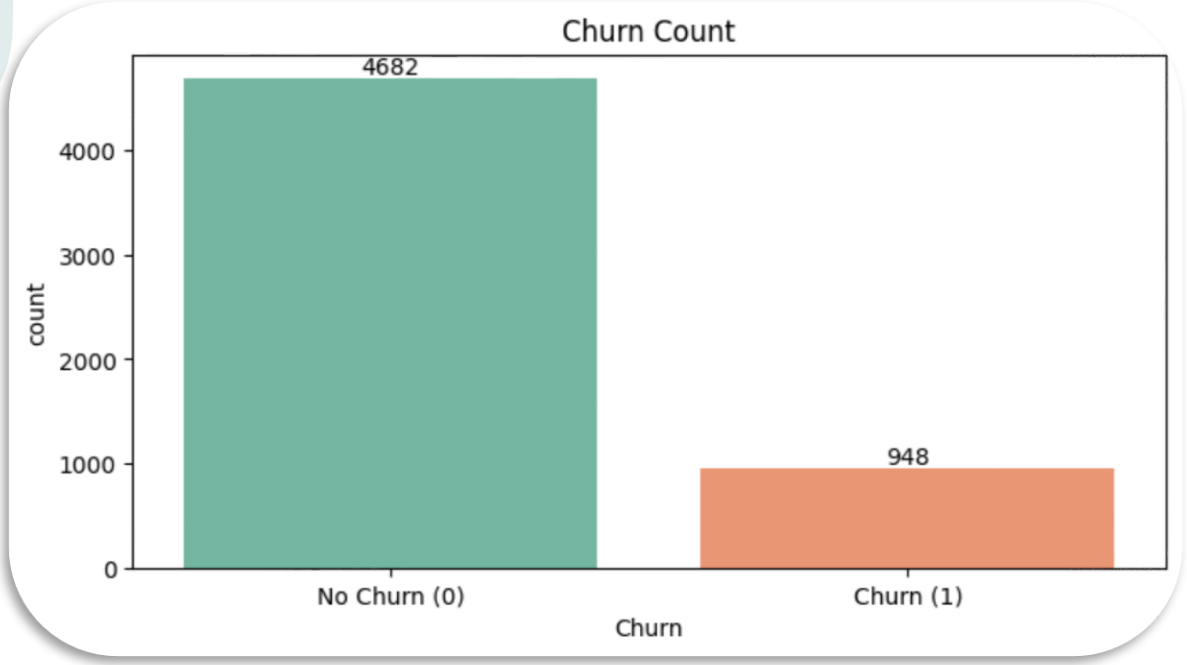
Features
- Numerical: 11
- Categorical: 7

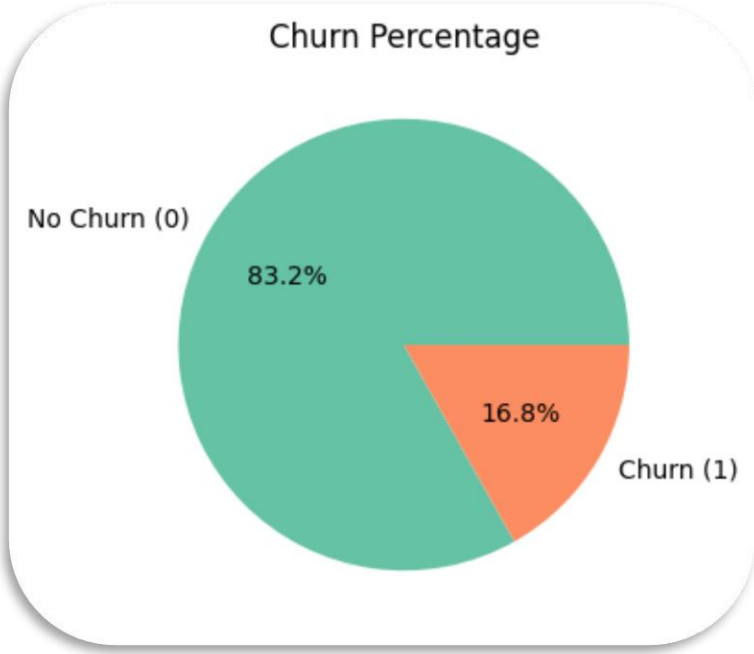
(Customer ID dropped from analysis)

* Missing values filled with 0 to preserve all customer records.

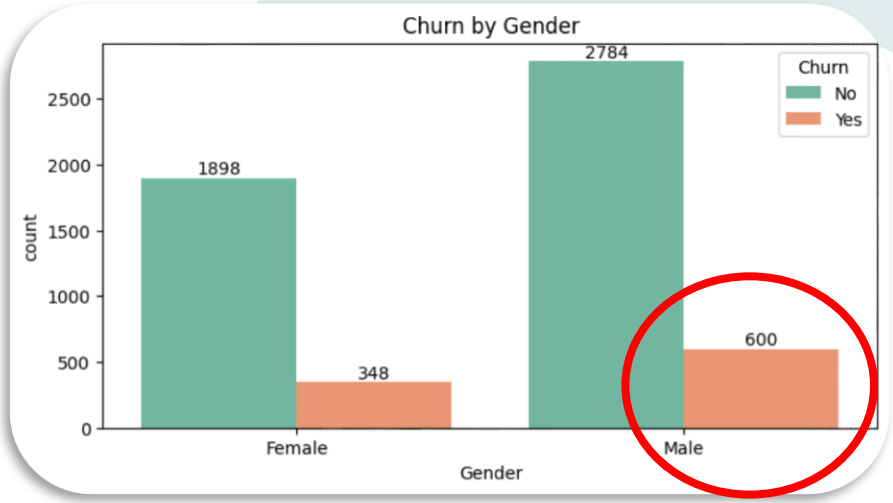
Churn Distribution

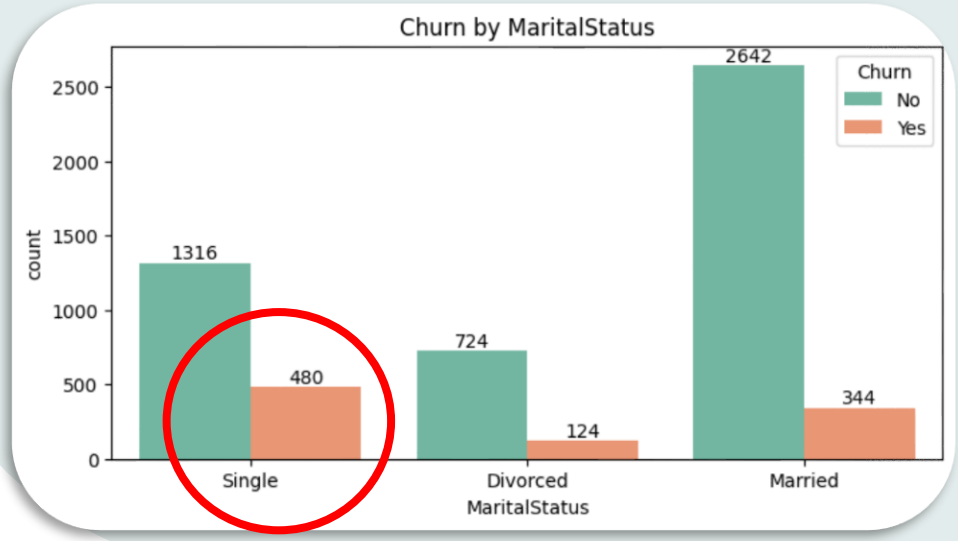
- Most customers remain with the company, but churn is still notable.
- Customer loss, even at this level, can affect revenue if not addressed.
- Highlights the need to explore numerical and categorical factors driving churn for better retention strategies.

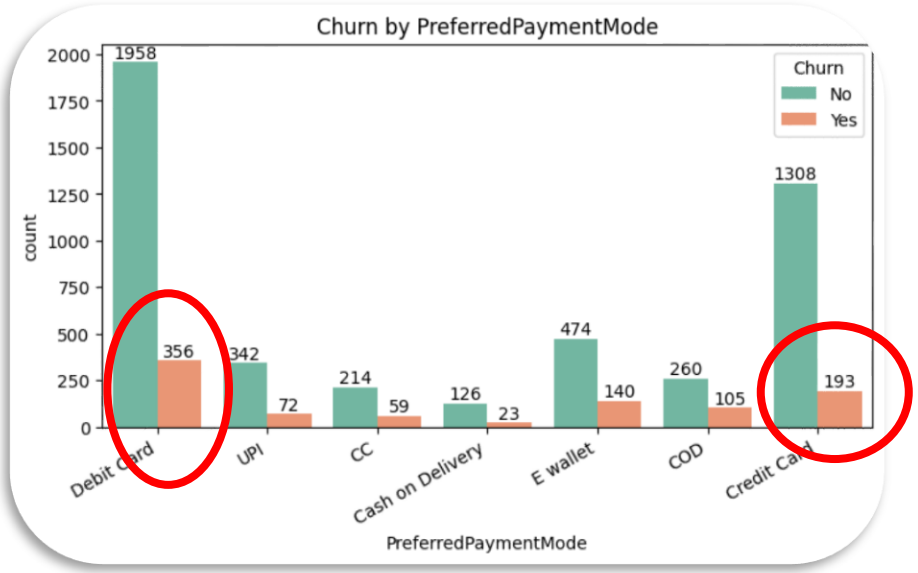


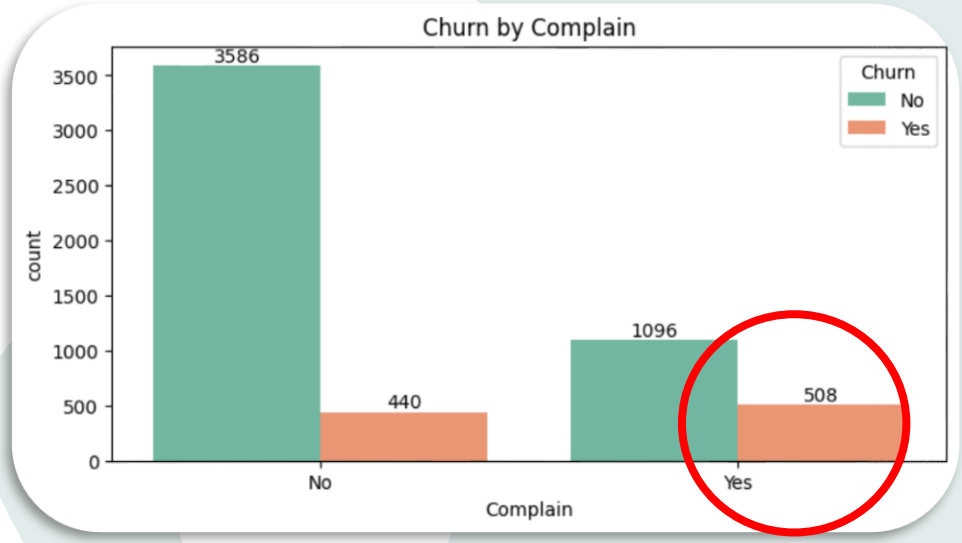


Categorical Variables vs. Churn



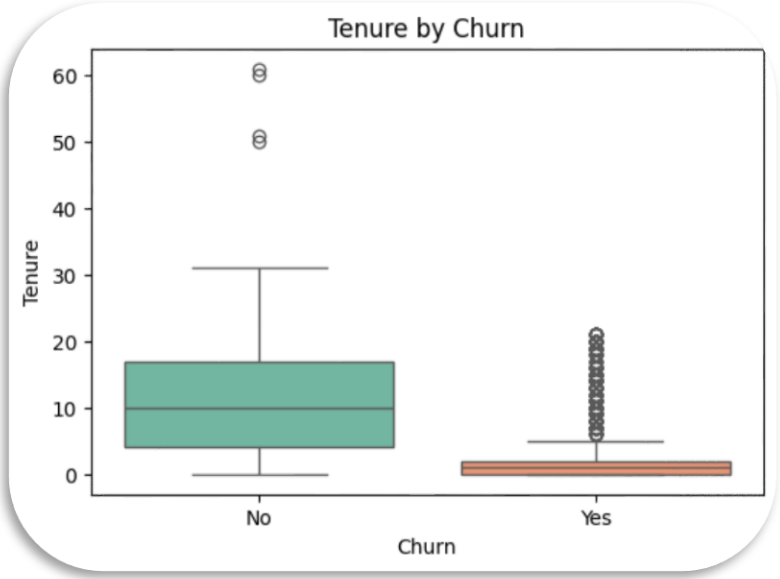




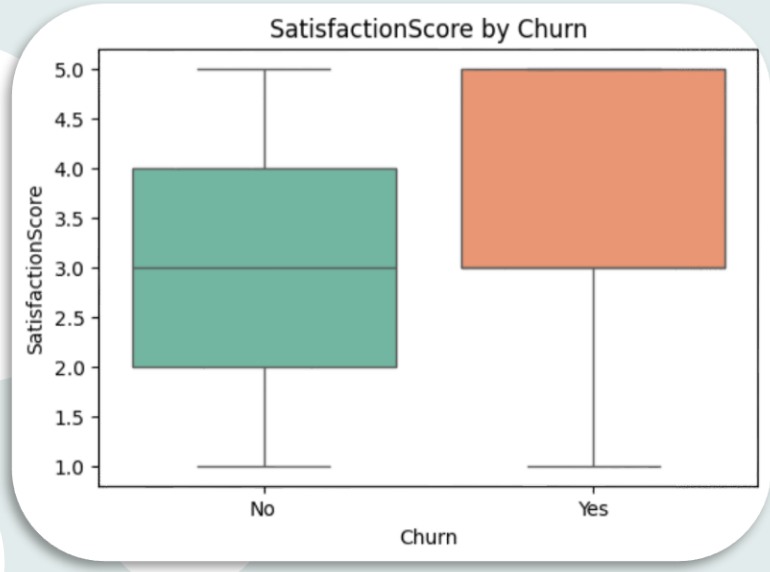


Numerical Variables vs. Churn

- Churners usually have **shorter tenure**, leaving earlier in the relationship.
- Churners often **live farther from the warehouse**.
- Churners surprisingly report **higher satisfaction score**, suggesting other churn drivers.
- **Hours on App is not a strong churn factor**, since usage is similar for both groups.







Correlation Analysis

Complain

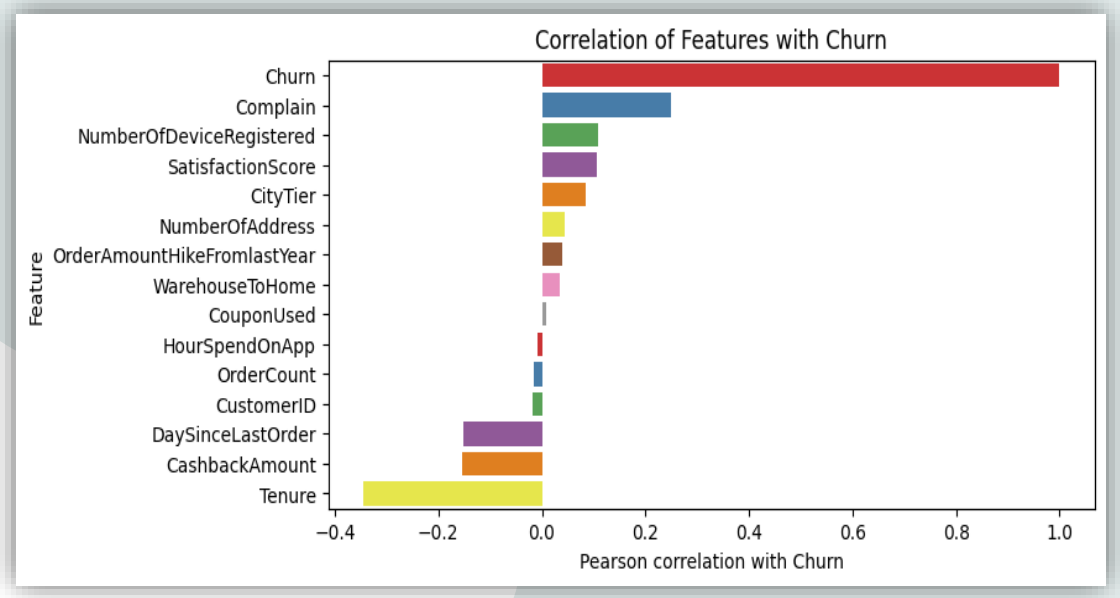
- Strongest positive correlation.
- Customers who complain are more likely to churn.

Tenure

- Clear negative correlation.
- Longer tenure are less likely to churn.

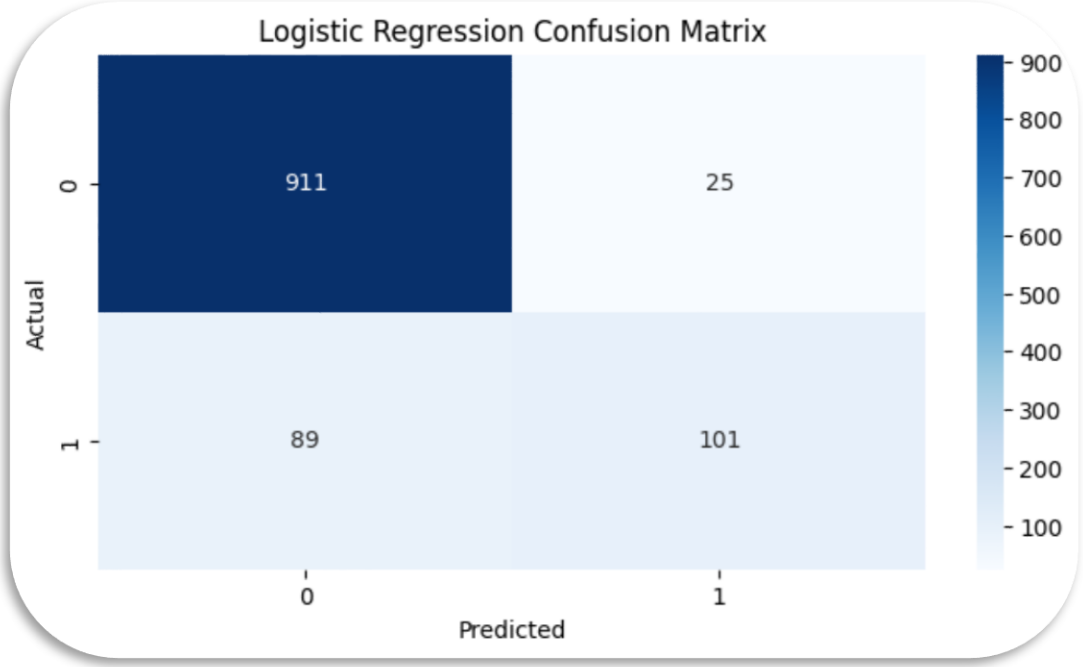
Other Features

- Mostly show weak positive correlations.
- Most have correlations close to zero.



Modeling Results – Logistic Regression

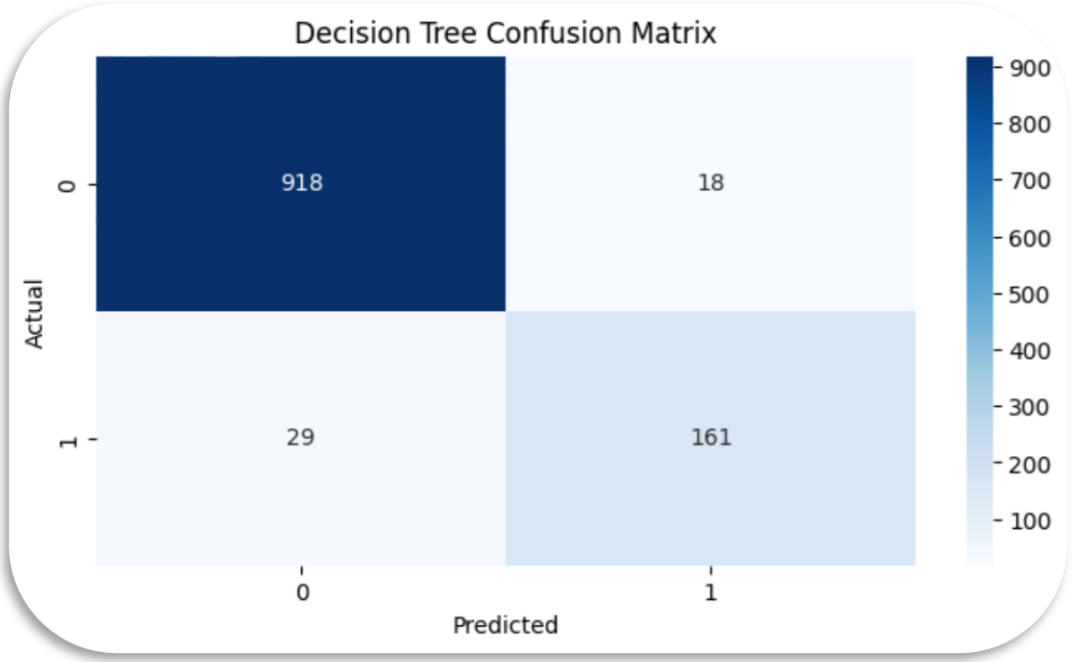
Accuracy Score: 0.8987566607460036
Precision Score: 0.8015873015873016
Recall Score: 0.531578947368421
F1 Score: 0.6392405063291139



- Model shows **high accuracy (~90%)**, but mainly due to large number of non-churners.
- While correctly identifies most loyal customers, **recall is low (~53%)**, meaning many churners are missed by the model.
- Model appears accurate, but **Logistic Regression is not the most effective** for reducing customer loss.

Modeling Results – Decision Tree

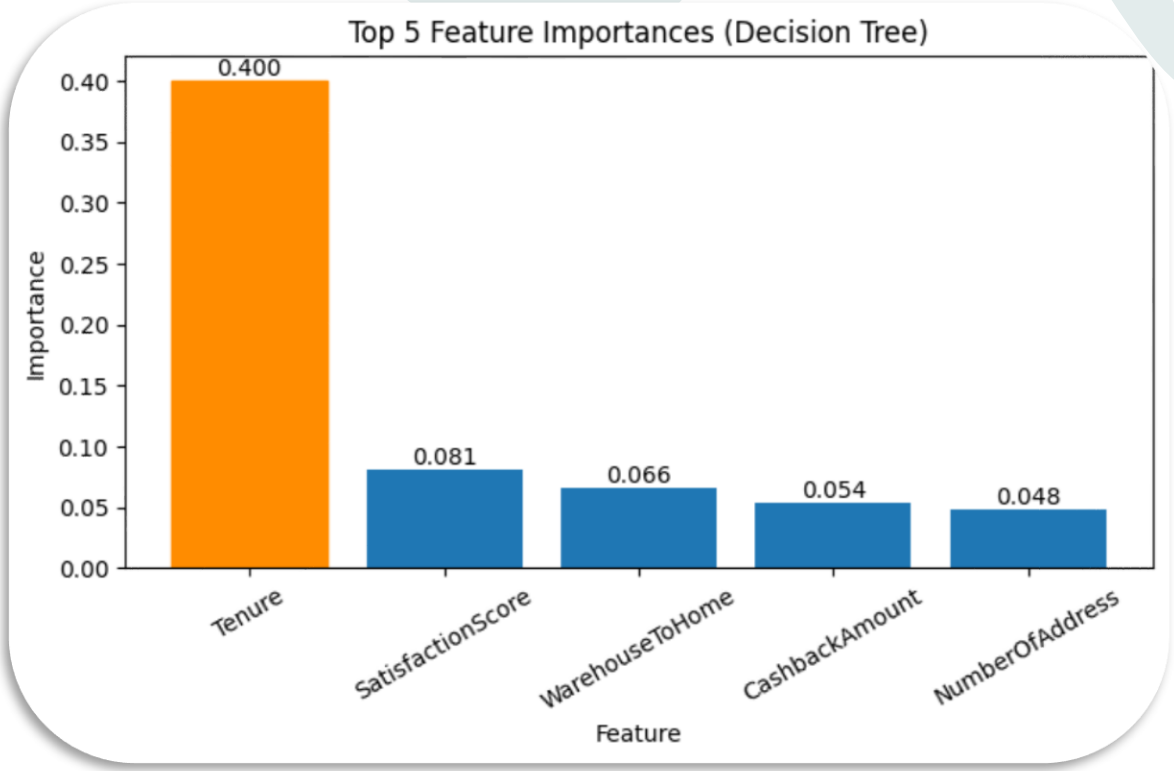
Accuracy Score: 0.9582593250444049
Precision Score: 0.8994413407821229
Recall Score: 0.8473684210526315
F1 Score: 0.8726287262872628



- Very **high accuracy (~96%)**, with strong performance in both **precision (~90%)** and **recall (~85%)**.
- Meaning the model **effectively identifies churners** while still **performing well for non-churners**, unlike Logistic Regression.
- **More balanced model**, making it much better suited for reducing customer loss.

Key insights

- **Tenure is the strongest predictor** of churn (customers with shorter tenure are much likely to leave)
- Others factors also contribute, but with less weight.
- These insights help the company to focus on high-risk groups, especially new customers, to design better retention strategies.





Conclusions

- **Churn rate is 16,8%** (about 1 in 6 customers).
- **Tenure and complaints** are the strongest drivers of churn.
- The **Decision Tree model** proved most effective, capturing churners with high accuracy and recall.

Recommendations

- Focus on **early retention programs** for new customers.
- Prioritize **fast resolution of complaints** to prevent exits.
- Pay special attention to **male and single customers**, who show higher churn rates.
- Review offers and experiences for **debit/credit cards users**, as they are more likely to churn.


THANK YOU