Gym Customer Churn Analysis

Predicting and Preventing Customer Attrition

Project Overview & Dataset

Project Goal

To identify customers at high risk of churning (leaving the gym) and develop targeted retention strategies based on predictive modeling.

The Data

Analysis is based on a clean dataset of 4,000 gym members. It includes demographic, behavioral, and contract data with no missing values.

EDA: Most Members Are New

The 'Lifetime' histogram is heavily skewed to the right, showing that a large portion of members are new.

- The median lifetime is only 1 month.

- This new member group is a key focus for retention efforts, as they churn most frequently.
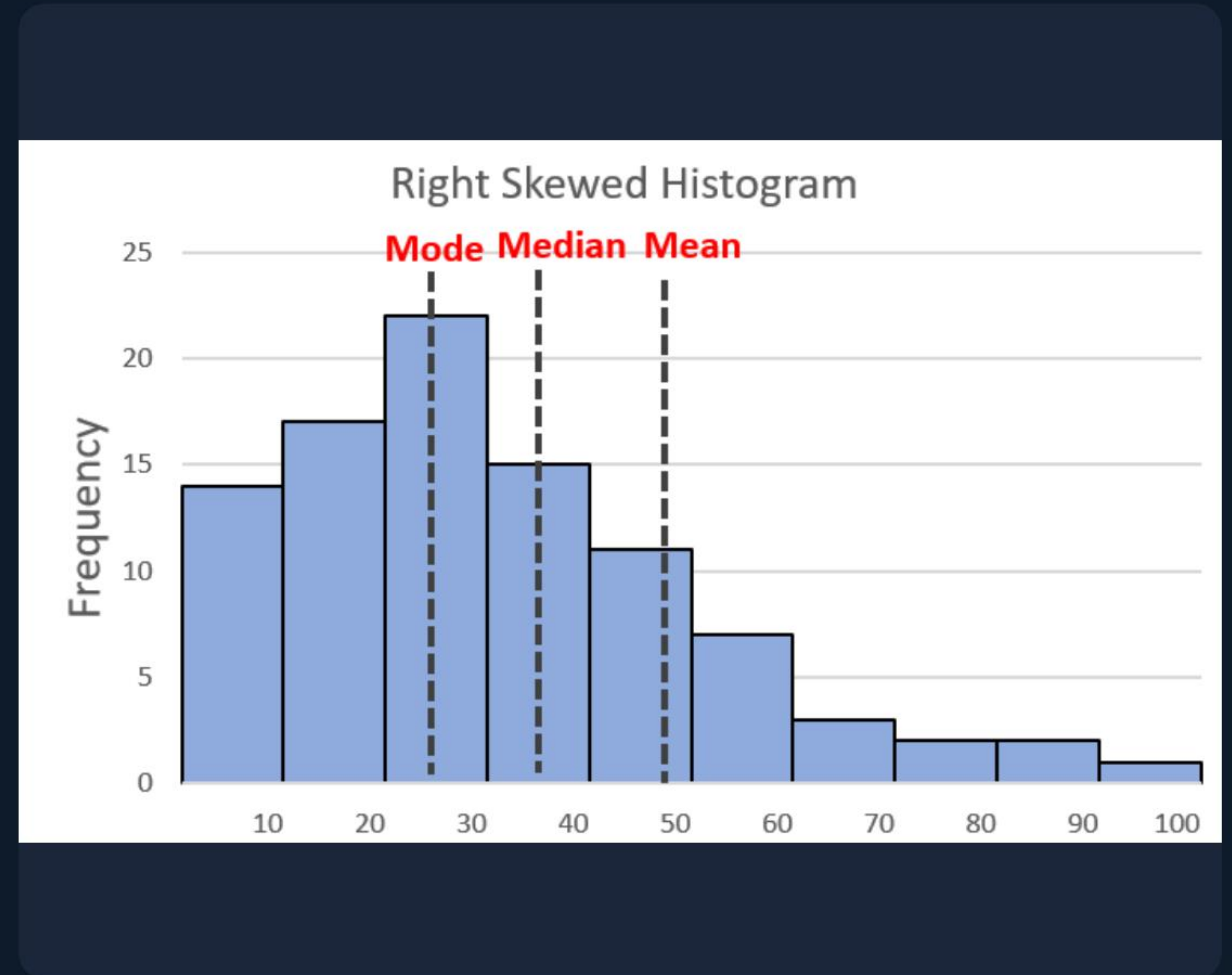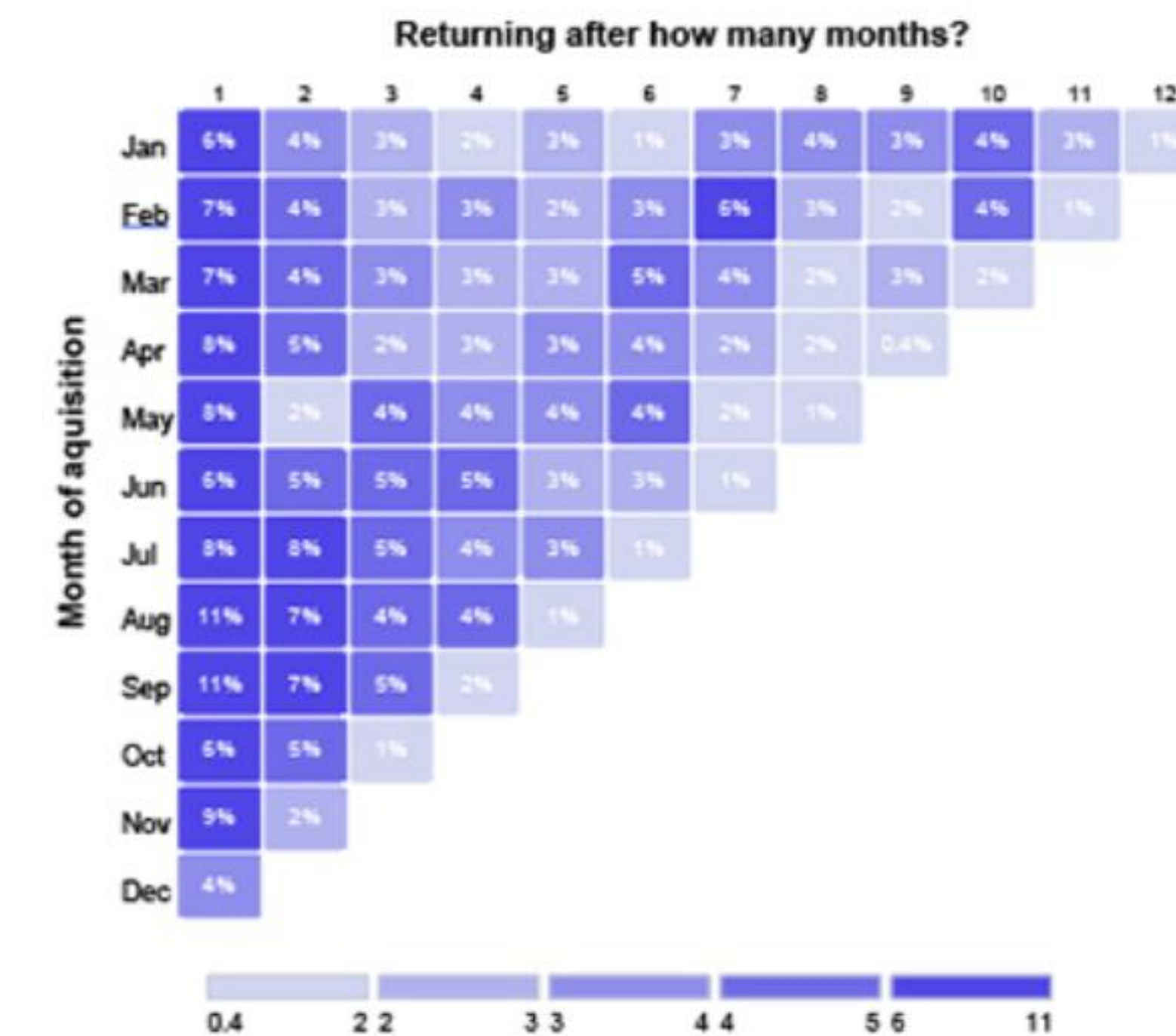
EDA: Churners Leave Early

Box plots show a dramatic difference in 'Lifetime' between customers who churn and those who stay:

- **Retained:** Median lifetime is ~4 months, with many members staying for 12+ months.

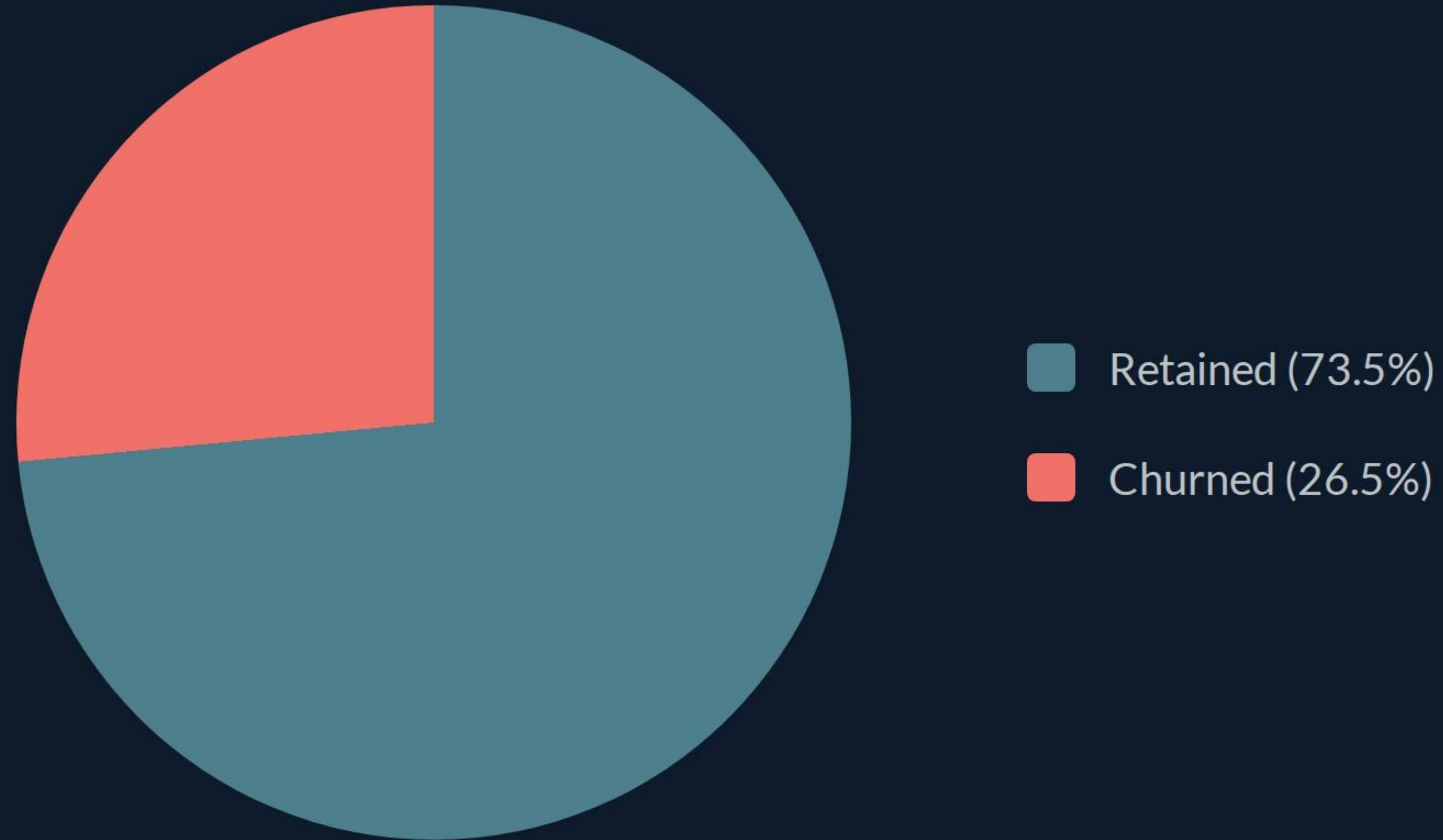- **Churned:** Median lifetime is only ~1 month.

This confirms that customers who leave do so very quickly.

Overall Churn Rate (26.5%)

Retained (73.5%)

Churned (26.5%)

The overall churn rate is 26.5%, highlighting a significant portion of the 4,000-member customer base at risk.
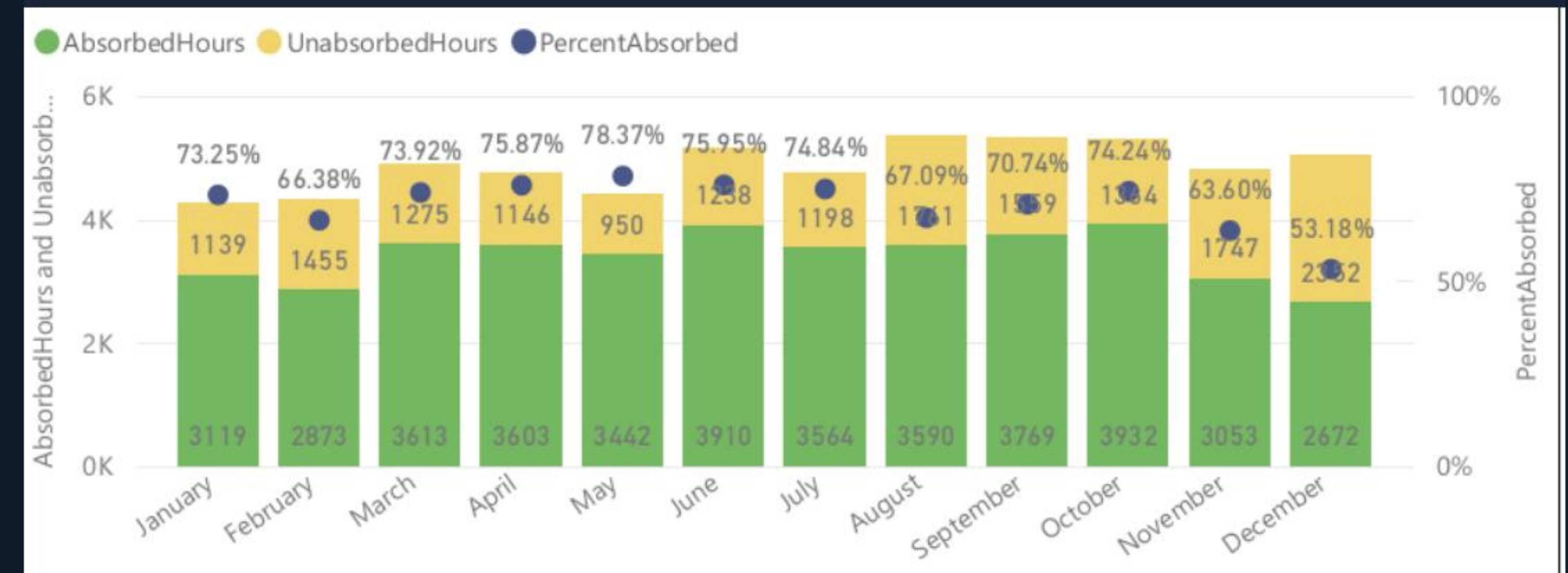
EDA: Contract Period is the Biggest Factor

The 1-month contract is the single biggest driver of churn.

- **1-Month Plan:** Has a churn rate of nearly 60%.

- **6-Month Plan:** Churn rate drops to ~15%.

- **12-Month Plan:** Churn rate is below 5%.

Longer contracts are highly effective at ensuring customer retention.

Correlation Analysis

A heatmap of numerical features reveals high multicollinearity, which is critical for modeling.

- **Key Finding:** `Avg_class_frequency_total` and `Avg_class_frequency_current_month` are highly correlated (0.95).

- This suggests that a member's recent behavior (current month) is a strong proxy for their total behavior.

Correlation Heatmap for eCommerce Store Metrics
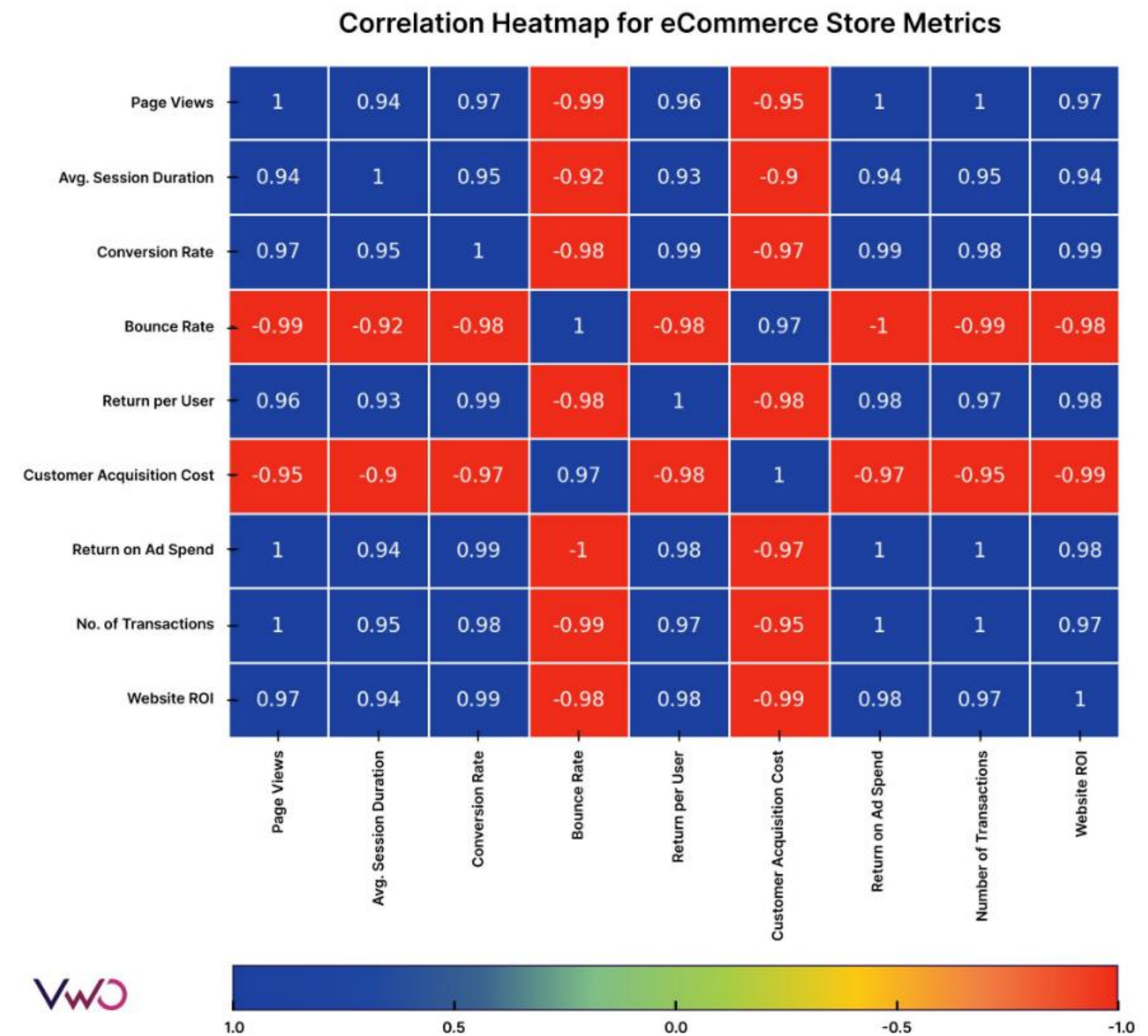
Predictive Modeling Methodology

1. Data Prep

The dataset was split into a 70% Training set (2800) and a 30% Testing set (1200). All numerical features were scaled.

2. Models Tested

Three models were trained to predict churn: Logistic Regression, Random Forest, and Gradient Boosting.

3. Evaluation

Models were compared using Accuracy, Precision, Recall, F1-Score, and AUC-ROC (Area Under Curve) to find the best predictor.
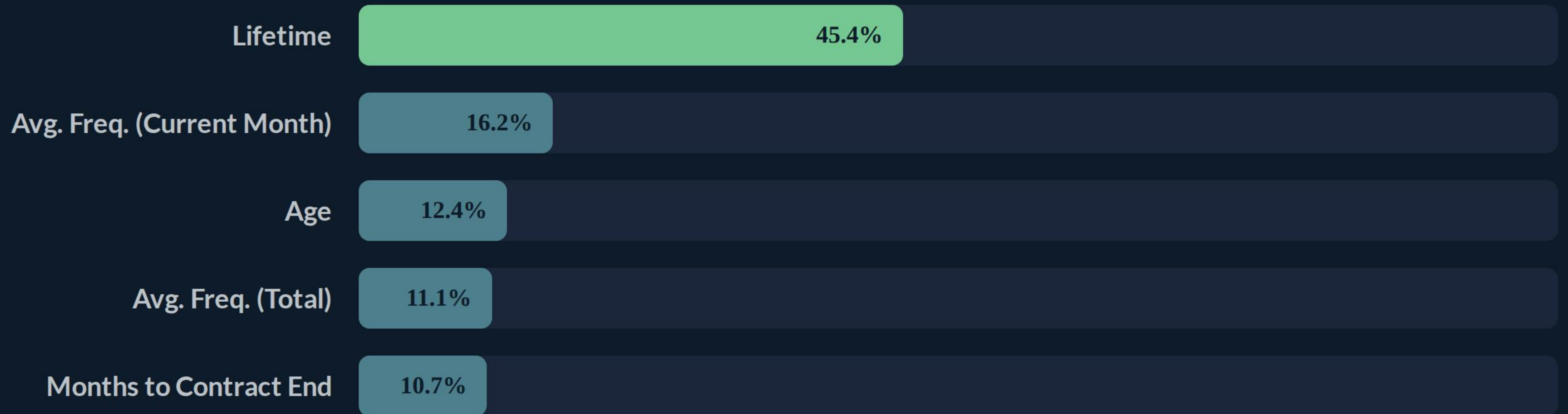
Model Performance Comparison

Model	Accuracy	Precision	Recall	F1-Score	ROC AUC
Logistic Regression	0.927	0.894	0.821	0.856	0.975
Gradient Boosting	**0.935**	**0.917**	**0.830**	**0.871**	**0.977**
Random Forest	0.914	0.872	0.792	0.830	0.968

Conclusion: *Gradient Boosting and Logistic Regression show top-tier performance. Gradient Boosting was selected for its highest Accuracy and AUC-ROC score.*

Key Churn Predictors (from Gradient Boosting Model)

Predictor	Value
Lifetime	45.4%
Avg. Freq. (Current Month)	16.2%
Age	12.4%
Avg. Freq. (Total)	11.1%
Months to Contract End	10.7%

*The model confirms the EDA: **'Lifetime'** is the single most dominant predictor of churn, followed by recent class frequency.*

Conclusions & Recommendations

📅 **Promote Longer Contracts:** Aggressively promote 6 and 12-month contracts. The 1-month plan has a ~60% churn rate, while 12-month plans have <5% churn.

👥 **Leverage Social Programs:** Members who join via 'Partner' or 'Promo_friends' programs are 2-3 times more loyal. These programs should be expanded.

🔔 **Create Engagement Alerts:** 'Lifetime' (< 3 months) and 'Current Month Frequency' (drops to 0) are the top predictors. Flag these users for proactive outreach and support.

Questions?

Thank you for your time.

Image Sources

https://blogger.googleusercontent.com/img/b/R29vZ2xl/AVvXsEj5KFeJi9ZKEAKre1LOQ3i2f8xS2pqColu2XM3nuhniiTg1qqZbKTa5RJ5eCwZPXwhNVUYCn1Y3uj8XWXL3GU1efTIfYZ5I7Df-095PNv2UAW3TvDL6iU6LcoPC0zDgeNL2XX7hInzNe_Qjc1tiZhjC_M_O4KJi_DjpeIG6WdPNq11plxPHuZvb2CRto0oq/s1600-rw/right_skewed_histogram2.png
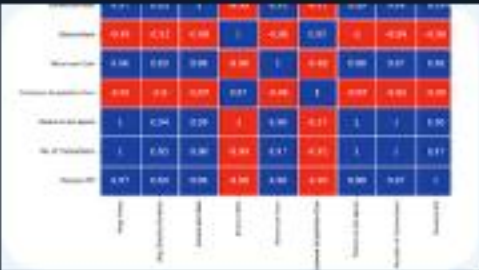
Source: www.listendata.com

https://chartexpo.com/blog/wp-content/uploads/2024/10/churn-rate.jpg

Source: chartexpo.com

https://community.fabric.microsoft.com/t5/image/serverpage/image-id/668166i96413B758C398012?v=v2

Source: community.powerbi.com

https://static.wingify.com/gcp/uploads/2024/08/11.png

Source: vwo.com