

A decorative graphic on the left side of the slide, featuring a dark blue background with several bright blue diagonal stripes and a white hexagonal shape containing an orange hexagon.

Initial  
Findings

# Logistic Regression

**No Oversample vs Oversample**

# Class Imbalance

**27%** Average Rate of Churn

**73%** Negative class (Typically problematic at 80%)

- Addressed class imbalance with oversampling
- Optimizing for recall metric

# Scores

## Model: No Oversampling (LogReg)

- ROC AUC .83
  - Excellent discrimination
- Recall .51
  - Establishes a benchmark

## Model: Oversampling (LogReg)

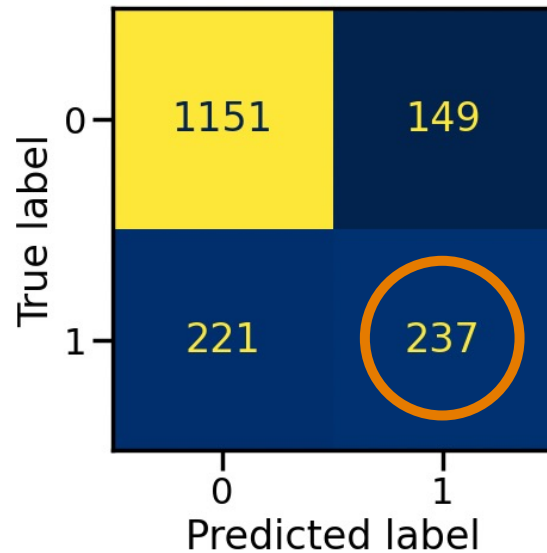
- ROC AUC .75
  - Acceptable discrimination
- Recall .80
  - Improvement over no oversampling

# Confusion Matrix

Compare models based on default threshold

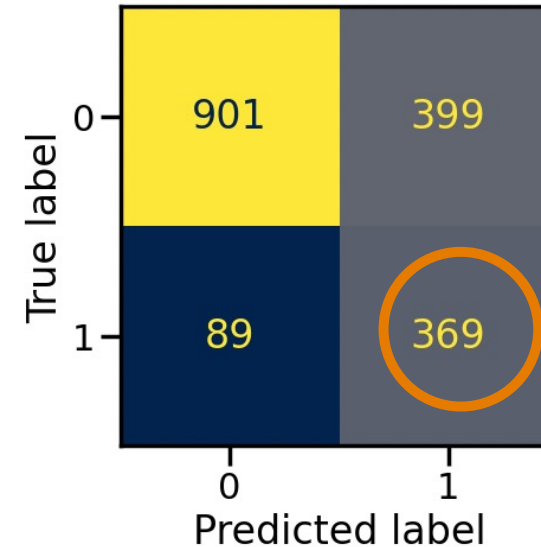
Oversampling improves the model's ability to identify people who cancelled their subscription

**Model: No Oversampling (LogReg)**



- Correctly classified 1151 customers who didn't cancel their subscription.
- Correctly classified 237 customers as people who did cancel their subscription.
- Incorrectly classified 149 customers as cancelled their subscription when they kept their subscription.
- Incorrectly classified 221 customers as people who kept their subscription but actually cancelled it.

**Model: Oversampling (LogReg)**



- Correctly classified 901 customers who didn't cancel their subscription.
- Correctly classified 369 customers as people who did cancel their subscription.
- Incorrectly classified 399 customers as cancelled their subscription when they kept their subscription.
- Incorrectly classified 89 customers as customers who kept their subscription but actually cancelled it.