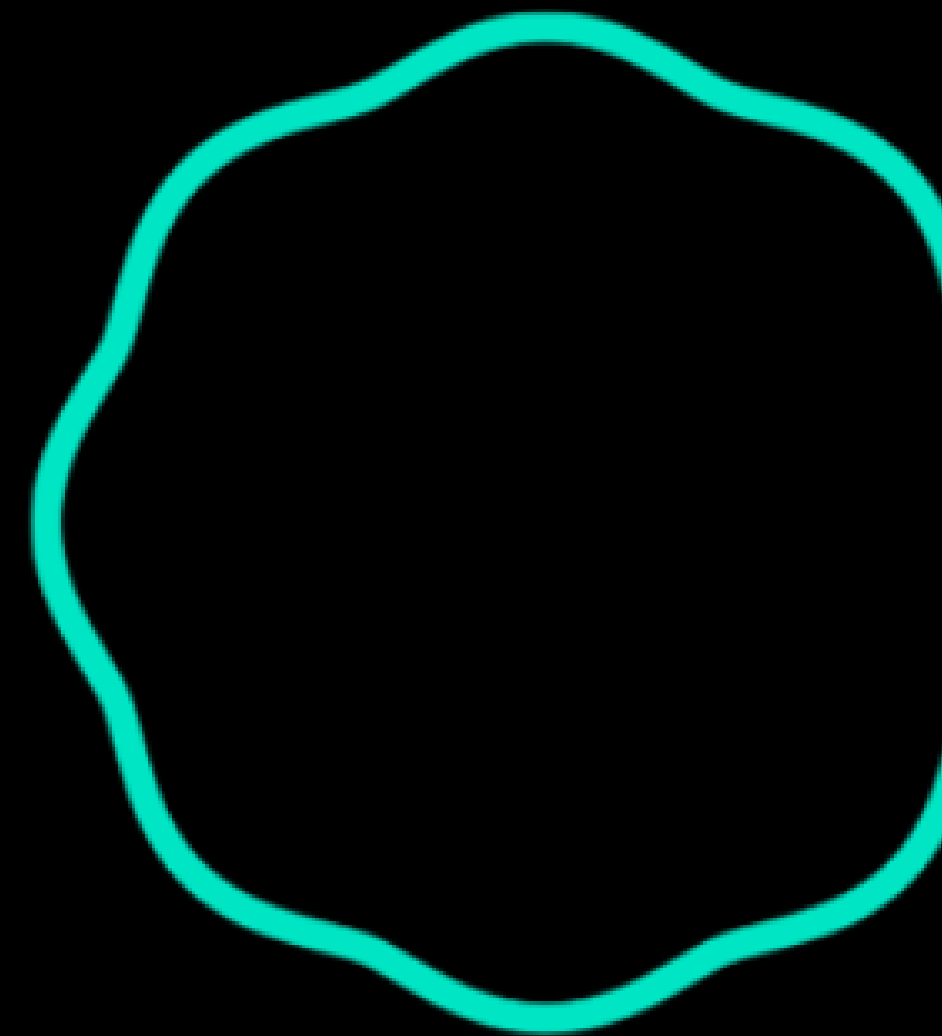MoneyLion®

Loan Risk Prediction Machine Learning Pipeline
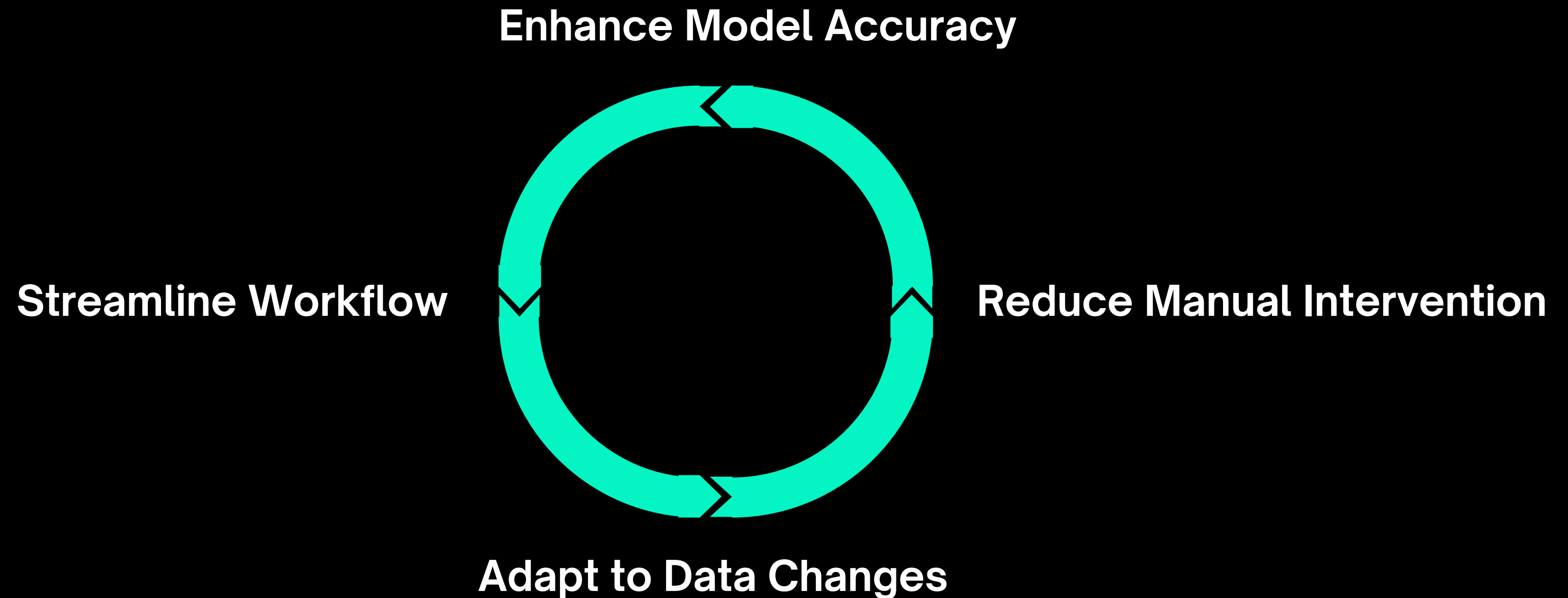
October 2, 2024

Introduction

Problem

Predicting loan application risk with a LightGBM model requiring monthly retraining and fine-tuning due to data drift and shifting feature importance.
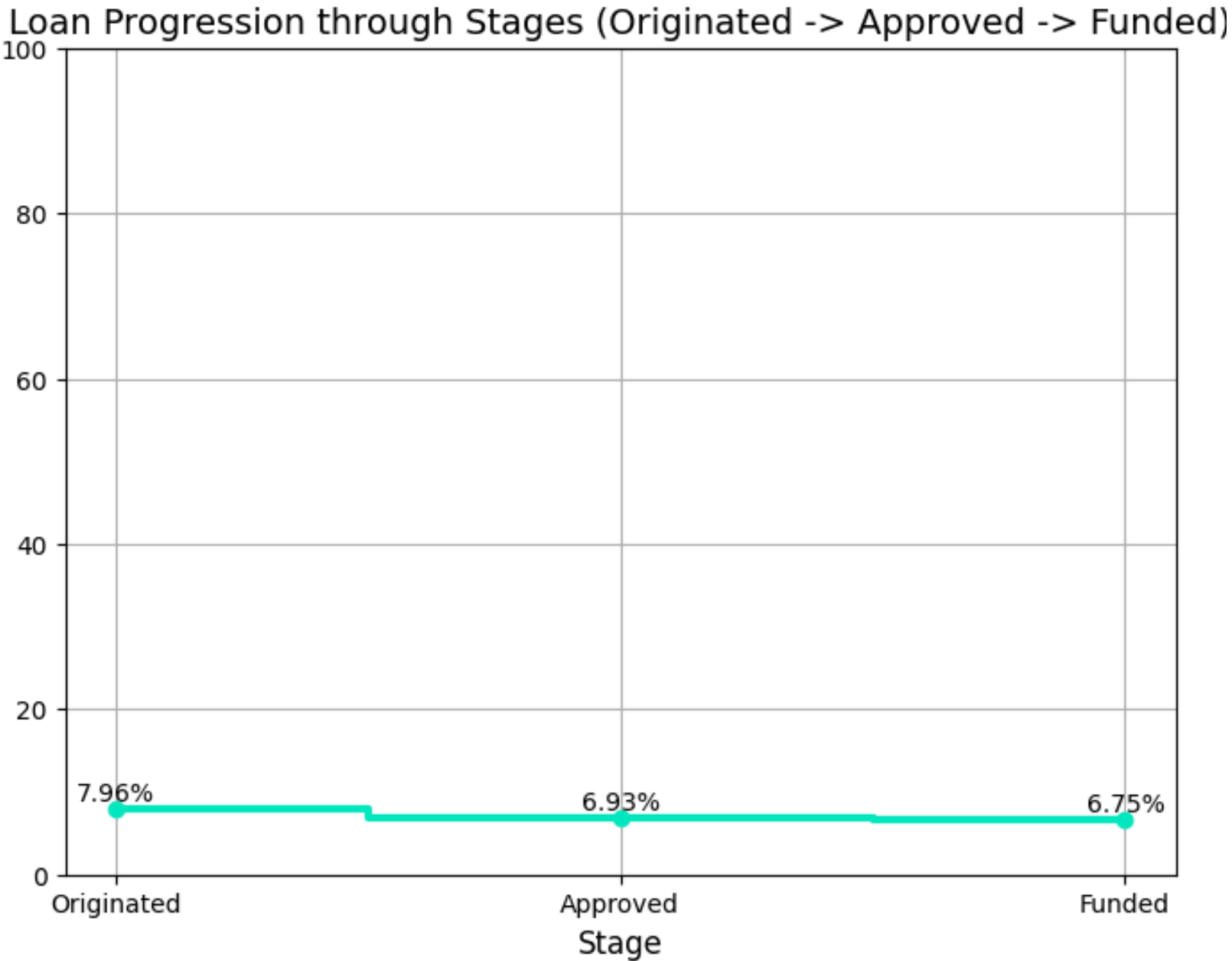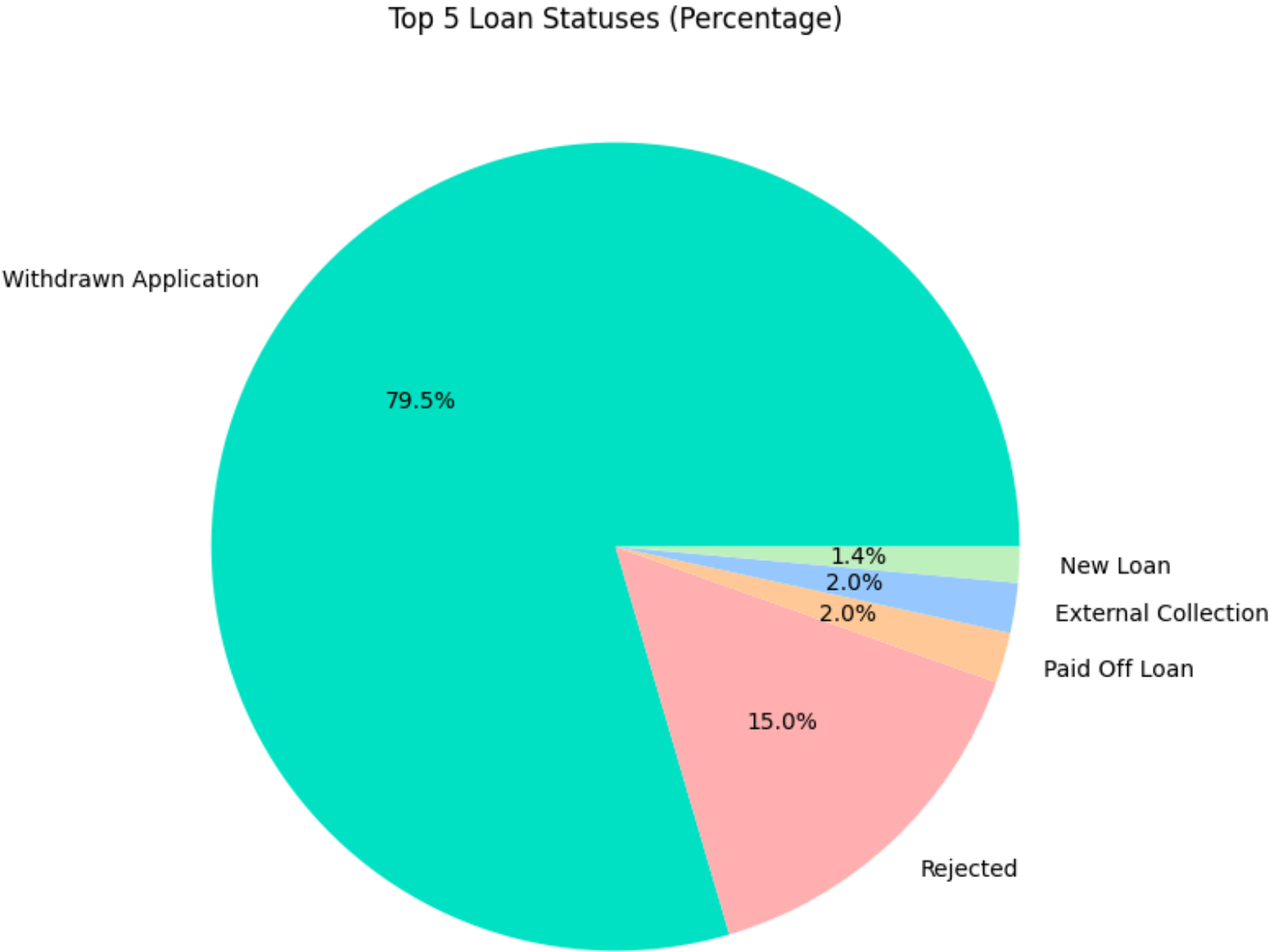
Solution

Develop an **automated machine learning pipeline** to streamline updates, enhance scalability, and ensure compliance with best practices.
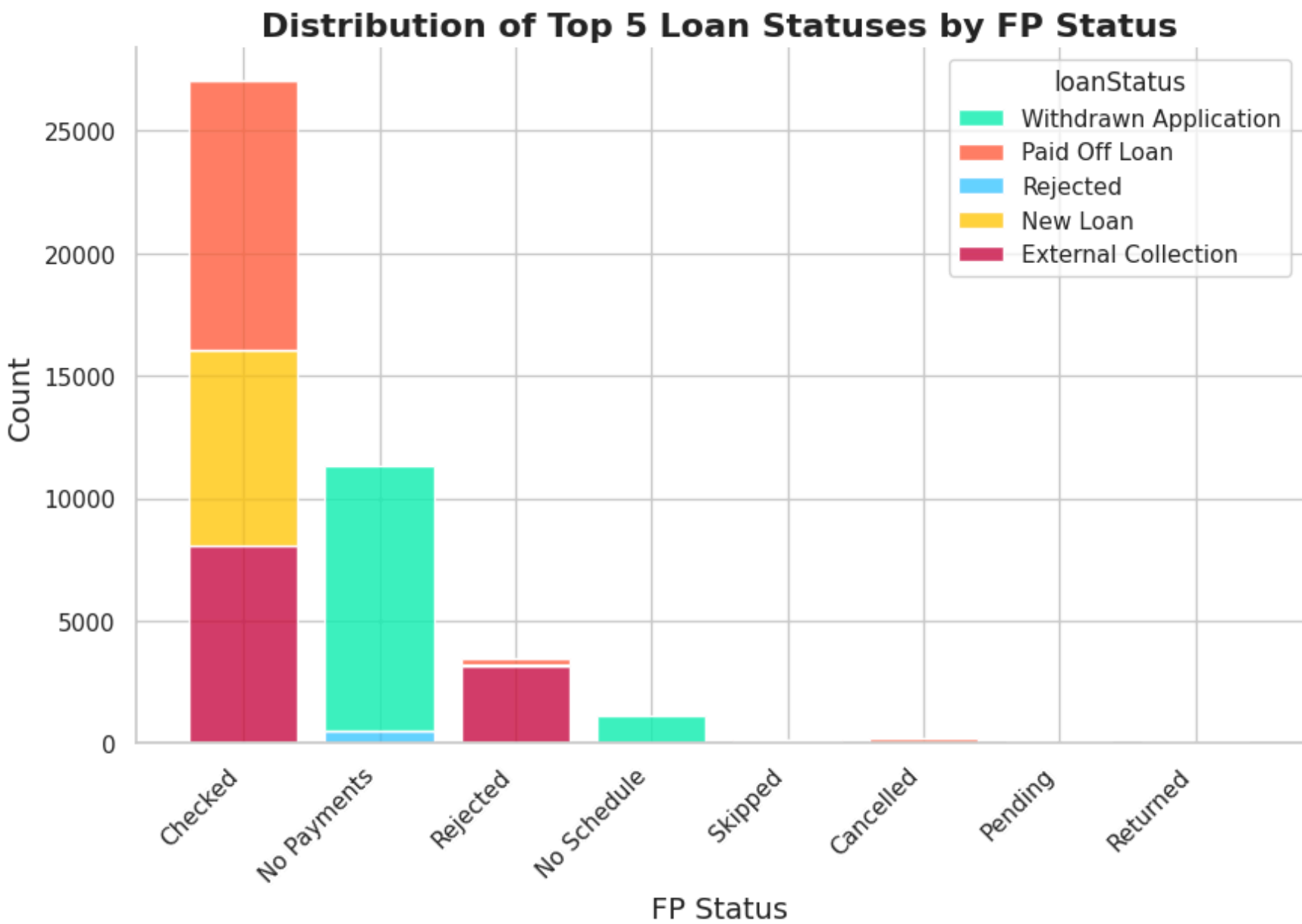
Business Objective

Enhance Model Accuracy

Streamline Workflow

Reduce Manual Intervention

Adapt to Data Changes

Exploratory Data Analysis

Top 5 Loan Statuses (Percentage)
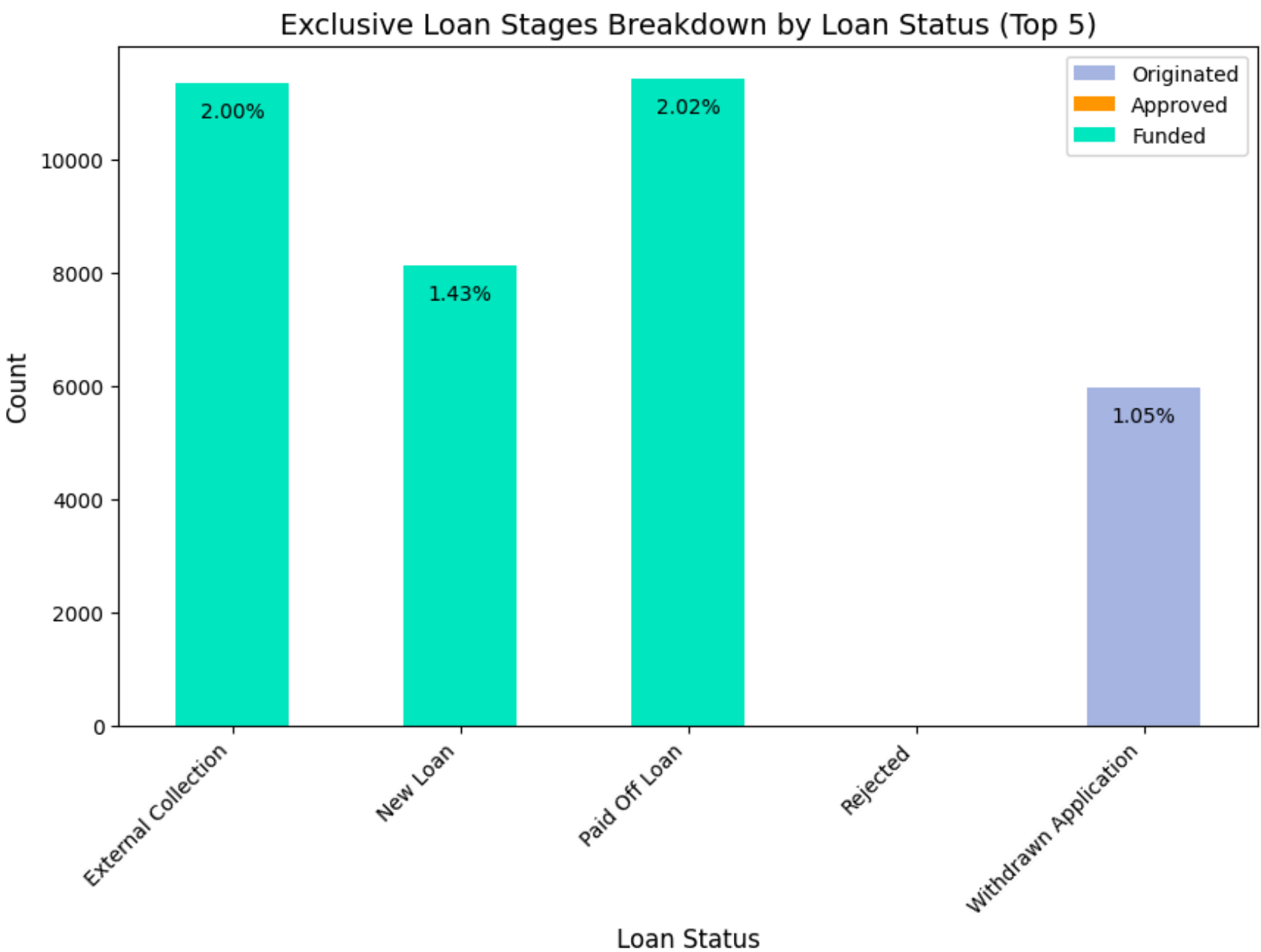
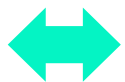Loan Progression through Stages (Originated -> Approved -> Funded)

Only 2% loan paid off, with 78% withdrawn and 14.7% rejected.

92% of loans fail to originate, but most originated loans are funded.

Exploratory Data Analysis
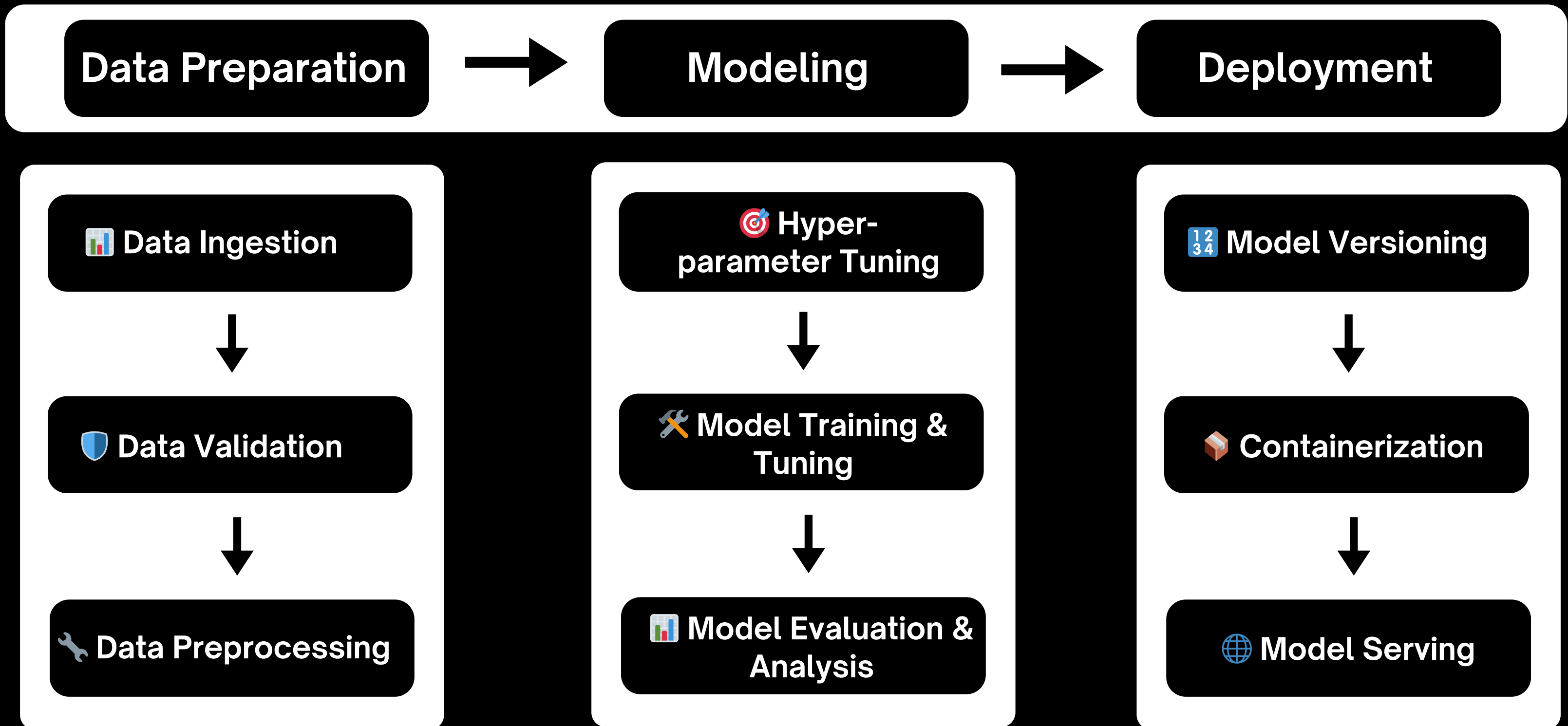

Exclusive Loan Stages Breakdown by Loan Status (Top 5)

Distribution of Top 5 Loan Statuses by FP Status

All External Collection, New Loan, and Paid Off Loan are funded ⬌ Paid Off, New Loan, and External Collection are 100% checked.

Machine Learning Pipeline
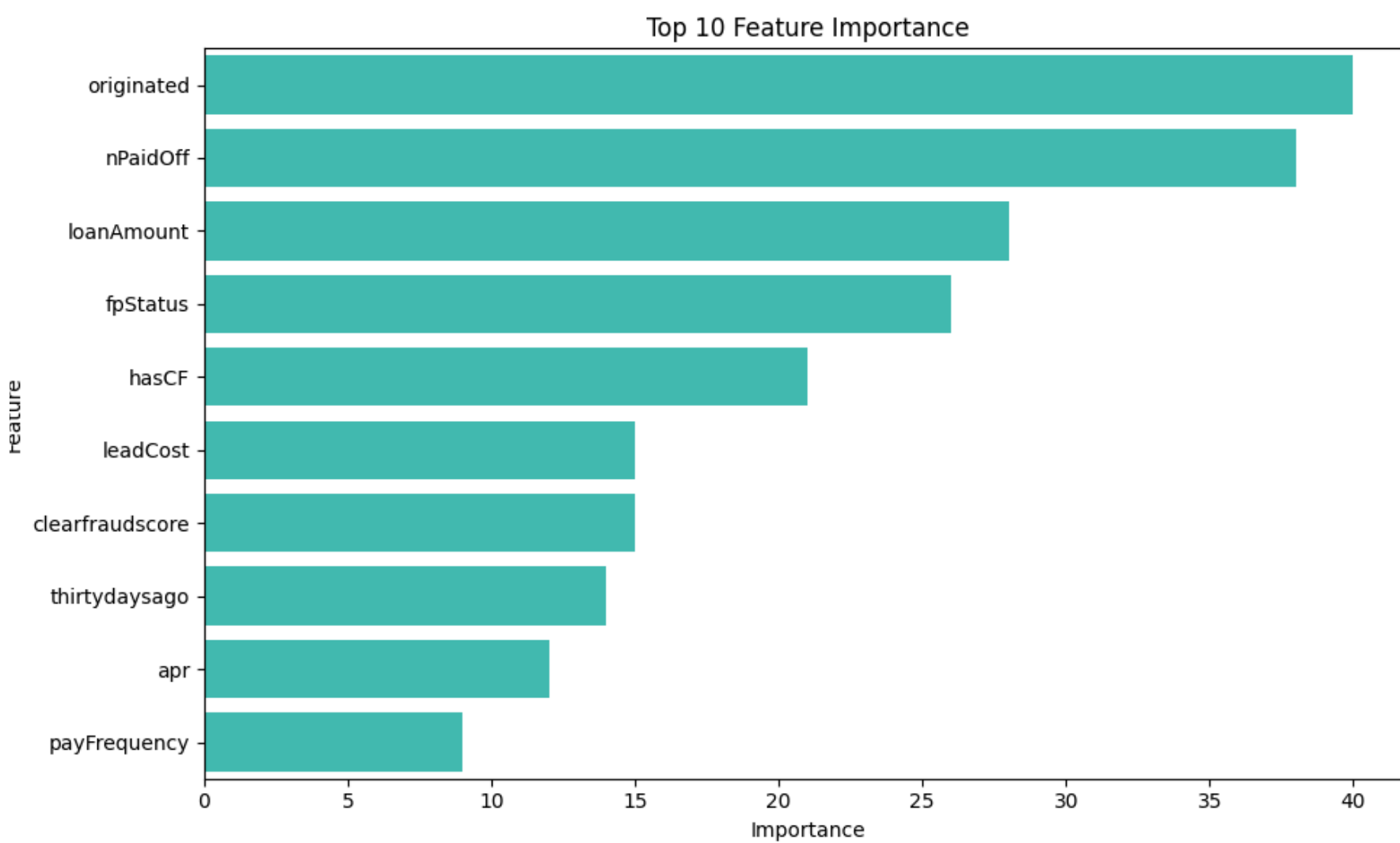
Data Preparation ➡️ **Modeling** ➡️ **Deployment**

Data Preparation	Modeling	Deployment
📊 Data Ingestion	🎯 Hyper-parameter Tuning	🔢 Model Versioning
⬇️	⬇️	⬇️
🛡️ Data Validation	🛠️ Model Training & Tuning	📦 Containerization
⬇️	⬇️	⬇️
🔧 Data Preprocessing	📊 Model Evaluation & Analysis	🌐 Model Serving

MoneyLion®

Why Light GBM

Speed and Efficiency

Regularization

Scalability

Feature Importance Insights

Native Categorical Support

High Accuracy
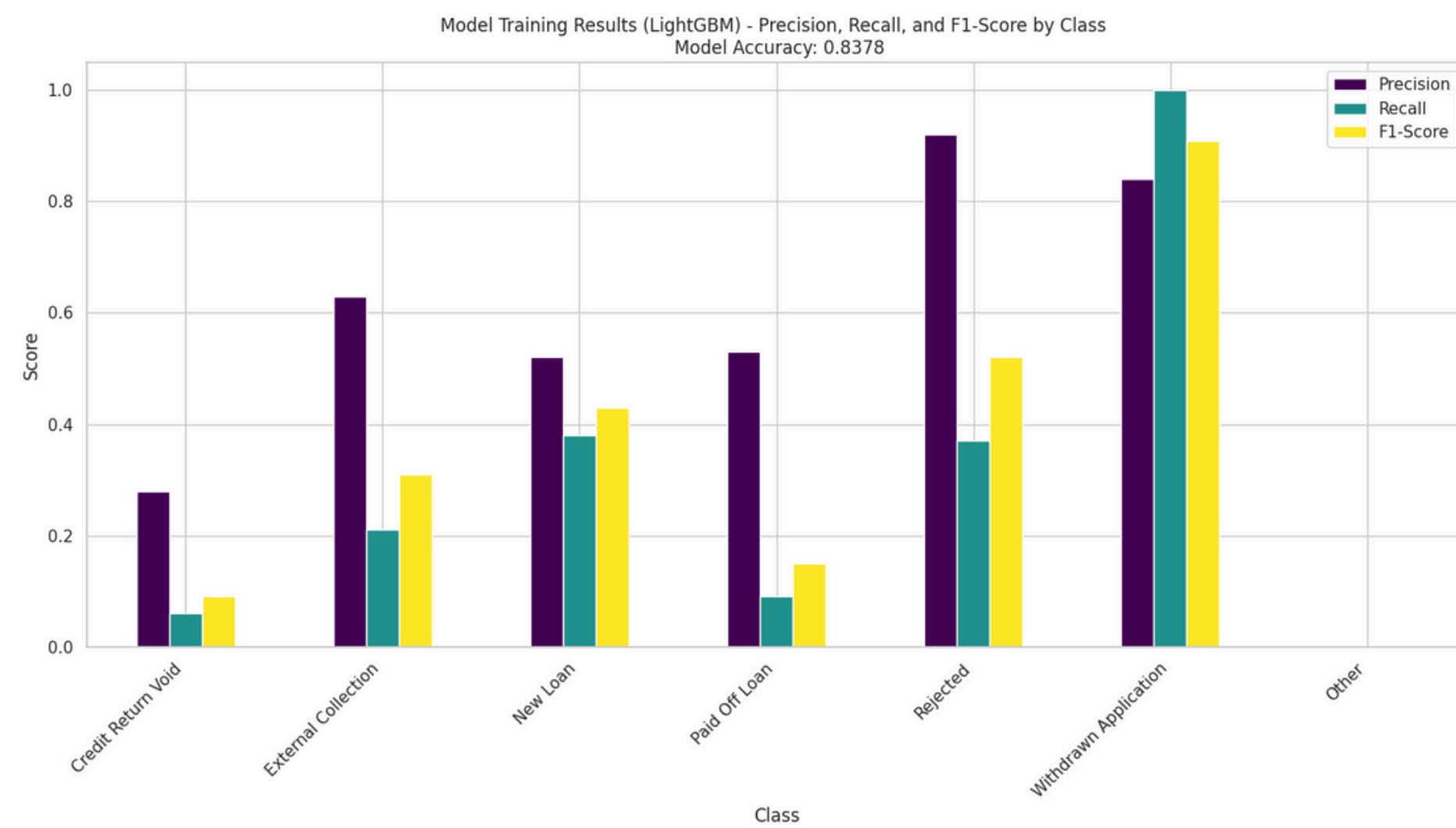
MoneyLion®

Model Evaluation & Analysis

The model achieved 83.78% accuracy. Withdrawn Application performed best with 0.84 Precision and 1.00 Recall

originated, nPaidOff, and loanAmount as top factors, while anon_ssn and loan_age have minimal impact.

Thank You