



CreditKarma Suite

JUNE 2025 UPDATE – ML PIPELINE

Jennifer Poernomo

RECAP: CREDITKARMA SUITE

A unified ecosystem of tools that aggregates data from the organization's diverse systems to assess and analyze customer credit scores.

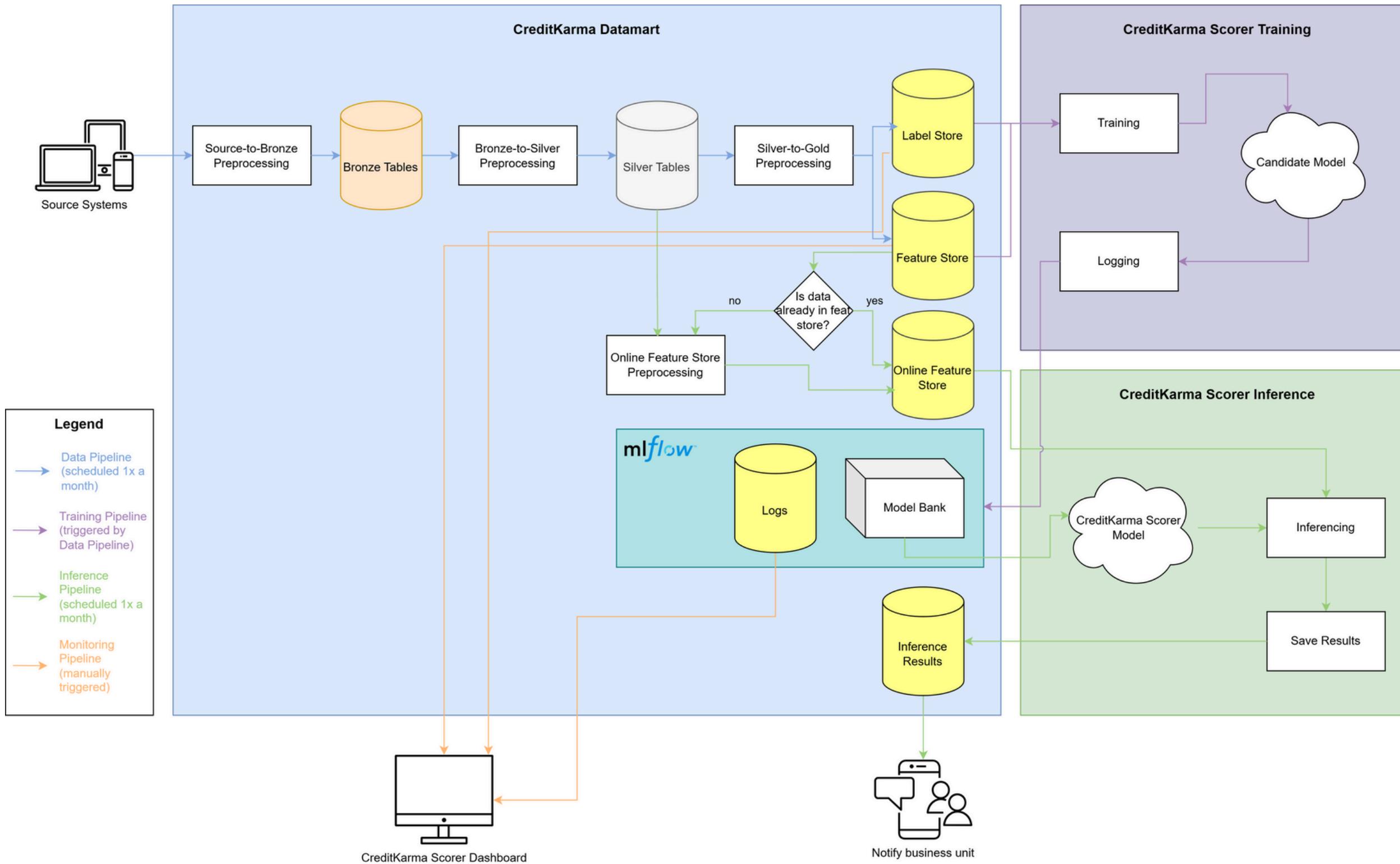
Currently comprises of two main components:

1. **CreditKarma Datamart**: Data repository aggregated from the organizations' systems that serves as a single source of truth applications can easily refer to.
2. **CreditKarma Scorer**: Binary classifier model that predicts the probability of customer default.

PROGRESS UPDATE

1. Deployed end-to-end machine learning pipeline using Airflow.
 - a. Monthly automated data processing pipeline.
 - b. Automated model training pipeline.
 - c. Model inferencing pipeline.
2. Created central model registry for CreditKarma Scorer versioning.
3. Created prototype model monitoring dashboard using Jupyter Notebook.

CREDITKARMA SUITE ARCHITECTURE

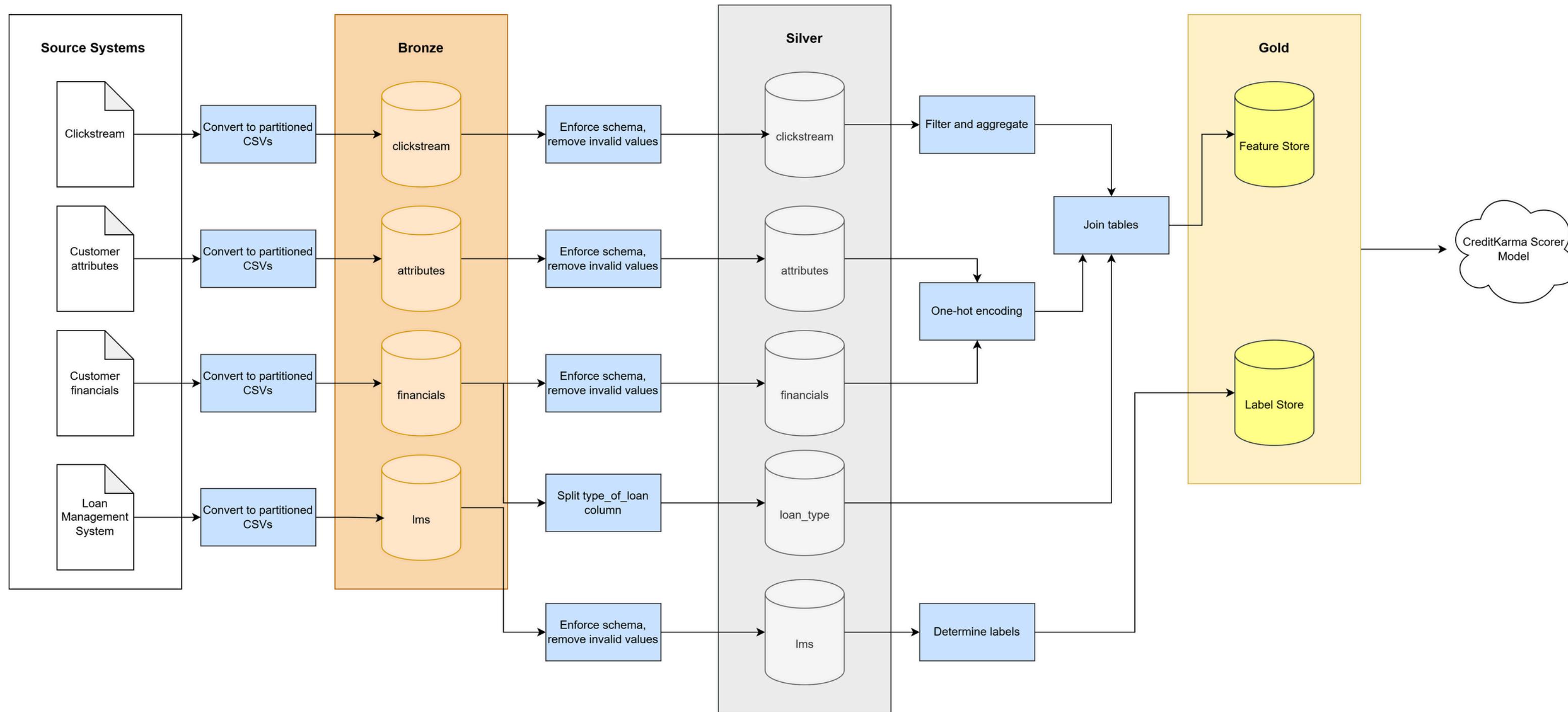


Pipeline is orchestrated using **Airflow**

- Data – Training pipeline is triggered on the 1st day of each month
- Inference pipeline is triggered on the 7th day of each month in case the team needs to analyze the metrics to choose the champion model

CREDITKARMA DATAMART

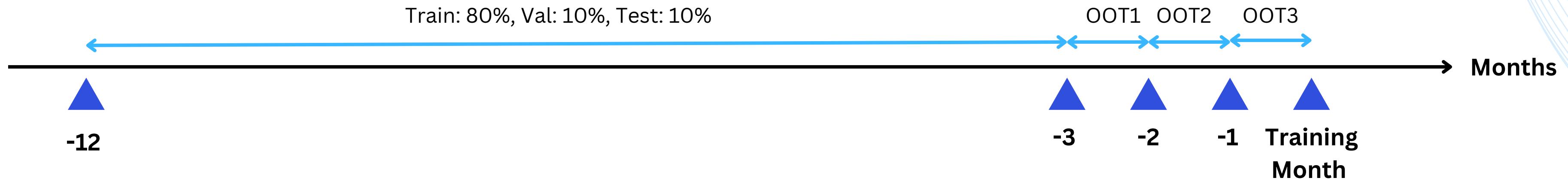
Follows the Medallion Architecture for consistency across applications and ease of maintainability



MODEL TRAINING

CreditKarma Scorer is trained on **the first day of each month** starting from 2024. The training pipeline performs automatic hyperparameter validation for **logistic regression** and **random forest classifier**. Model artifacts, alongside its data preprocessing pipeline, are logged into a model registry hosted using **mlflow**.

Dataset Split



Pipeline



MODEL REGISTRY

mlflow logs hyperparameter tuning results for each month of model training run and saves the best model into the **model_bank**.

GUI is served on localhost:5001 for the maintainers of CreditKarma Scorer to collaborate on.

The screenshot shows the mlflow UI interface. At the top, there's a navigation bar with tabs for Experiments, Models, and Prompts. The Experiments tab is selected. Below the navigation bar, the date "2024-01-01" is displayed, along with options to "Provide Feedback" and "Add Description".

The main area is divided into two sections: "Experiments" on the left and "Runs" on the right. The "Experiments" section lists monthly experiment runs from 2024-01-01 to 2024-11-01. The "Runs" section displays a detailed list of 100 matching runs for the specified date. The columns in the "Runs" table include Run Name, Created (sorted by time), Dataset, Duration, Source, and Models. Each row in the table represents a different run, with its status indicated by a colored circle (green checkmark, red X, or blue dot) and a brief description of the hyperparameters used.

Run Name	Created	Dataset	Duration	Source	Models
rf_best_model	1 day ago	-	8.5s	airflow	creditkarma-scorer v2
rf_n_estimators=200_max...	1 day ago	-	2.0s	airflow	-
rf_n_estimators=200_max...	1 day ago	-	2.3s	airflow	-
rf_n_estimators=200_max...	1 day ago	-	64ms	airflow	-
rf_n_estimators=200_max...	1 day ago	-	2.3s	airflow	-
rf_n_estimators=200_max...	1 day ago	-	2.3s	airflow	-
rf_n_estimators=200_max...	1 day ago	-	65ms	airflow	-
rf_n_estimators=200_max...	1 day ago	-	1.8s	airflow	-
rf_n_estimators=200_max...	1 day ago	-	2.1s	airflow	-
rf_n_estimators=200_max...	1 day ago	-	57ms	airflow	-
rf_n_estimators=200_max...	1 day ago	-	1.9s	airflow	-
rf_n_estimators=200_max...	1 day ago	-	2.1s	airflow	-
rf_n_estimators=200_max...	1 day ago	-	54ms	airflow	-
rf_n_estimators=200_max...	1 day ago	-	1.8s	airflow	-
rf_n_estimators=200_max...	1 day ago	-	2.0s	airflow	-
rf_n_estimators=200_max...	1 day ago	-	63ms	airflow	-
rf_n_estimators=200_max...	1 day ago	-	1.8s	airflow	-
rf_n_estimators=200_max...	1 day ago	-	1.9s	airflow	-
rf_n_estimators=200_max...	1 day ago	-	50ms	airflow	-

MODEL DEPLOYMENT

Model deployment is also done via **mlflow**. CreditKarma maintainers must assign the **“champion” alias** to the best model. This allows for easy **blue-green deployment**, which is safe enough for the CreditKarma Scorer as its primary users are the internal business unit who maintain active communication with the maintainers.

mlflow 2.22.1 Experiments Models Prompts

Registered Models >
creditkarma-scorer

Created Time: 06/21/2025, 03:51:22 PM Last Modified: 06/21/2025, 05:26:23 PM

Description Edit

Tags

Versions Compare

Version	Registered at	Created by	Tags	Aliases	Description
Version 24	06/21/2025, 05:26:23 PM		train_date: 2024-11-01 model_type: rf	Add	
Version 23	06/21/2025, 05:20:58 PM		train_date: 2024-11-01 model_type: logreg	Add	
Version 22	06/21/2025, 05:19:31 PM		train_date: 2024-12-01 model_type: rf	Add	
Version 21	06/21/2025, 05:14:06 PM		train_date: 2024-12-01 model_type: logreg	@ champion	
Version 20	06/21/2025, 05:12:25 PM		train_date: 2024-10-01 model_type: rf	Add	
Version 19	06/21/2025, 05:06:50 PM		train_date: 2024-10-01 model_type: logreg	Add	
Version 18	06/21/2025, 05:05:15 PM		train_date: 2024-09-01 model_type: rf	Add	
Version 17	06/21/2025, 05:00:06 PM		train_date: 2024-09-01 model_type: logreg	Add	
Version 16	06/21/2025, 04:58:24 PM		train_date: 2024-08-01 model_type: rf	Add	
Version 15	06/21/2025, 04:53:32 PM		train_date: 2024-08-01 model_type: logreg	Add	

New model registry UI

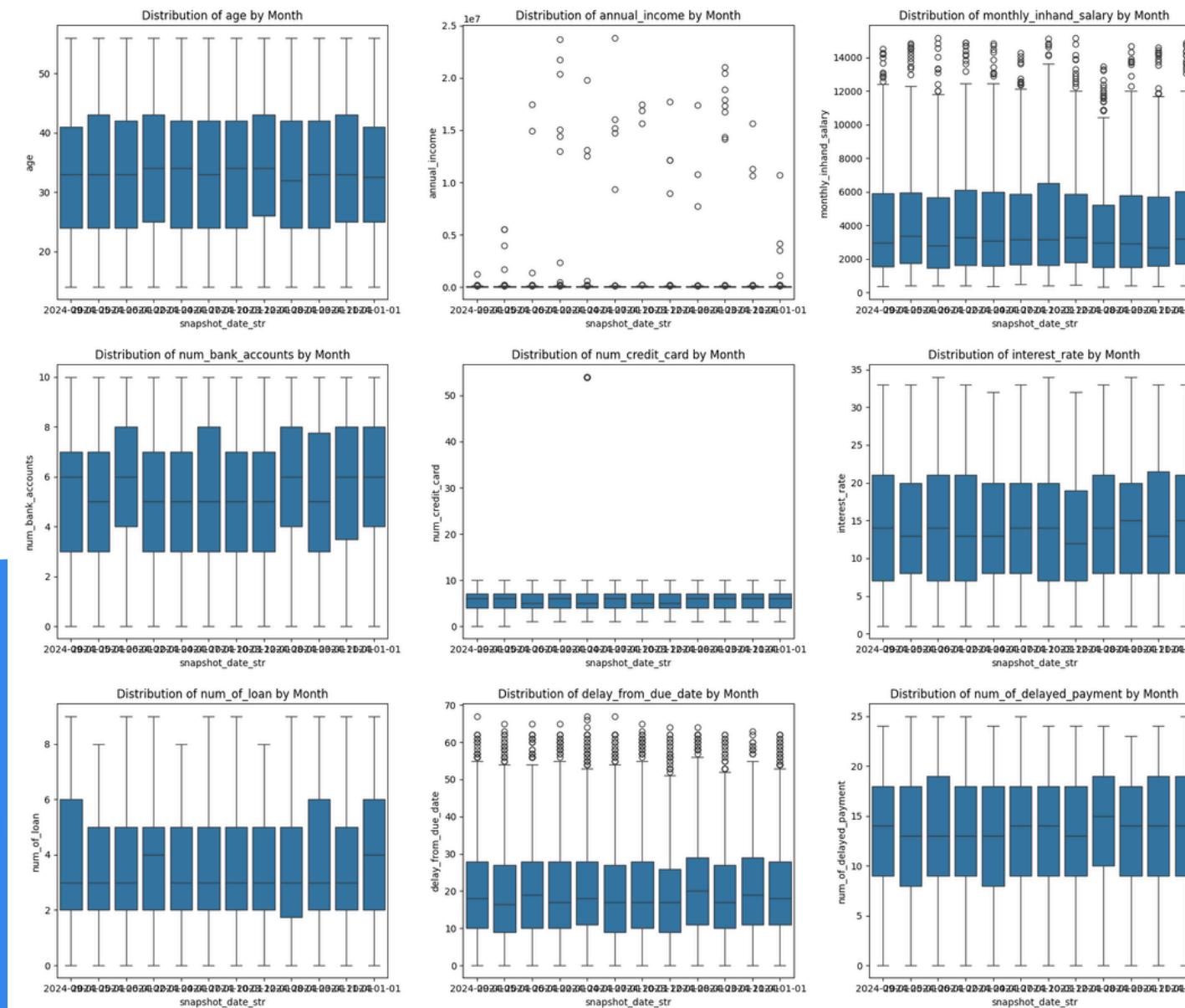
Updating between blue-green models is as easy as changing the alias

1 2 3 >

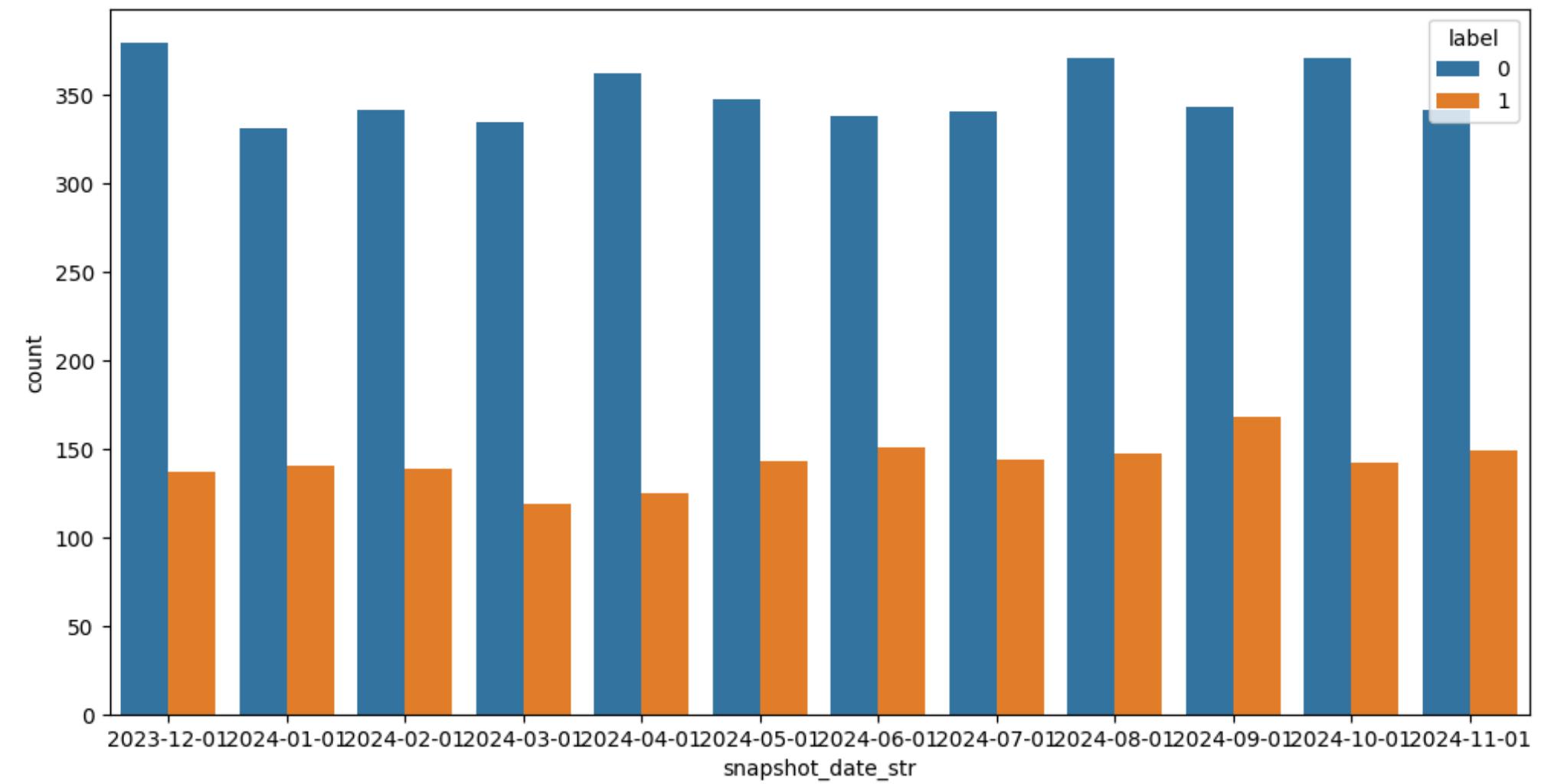
DATA MONITORING

Monitors the distribution of data over time (whether or not there is data / concept drift to be aware of).
Currently done using Jupyter Notebook, but planned to migrate to a Streamlit app in the future.

Feature box plots over months

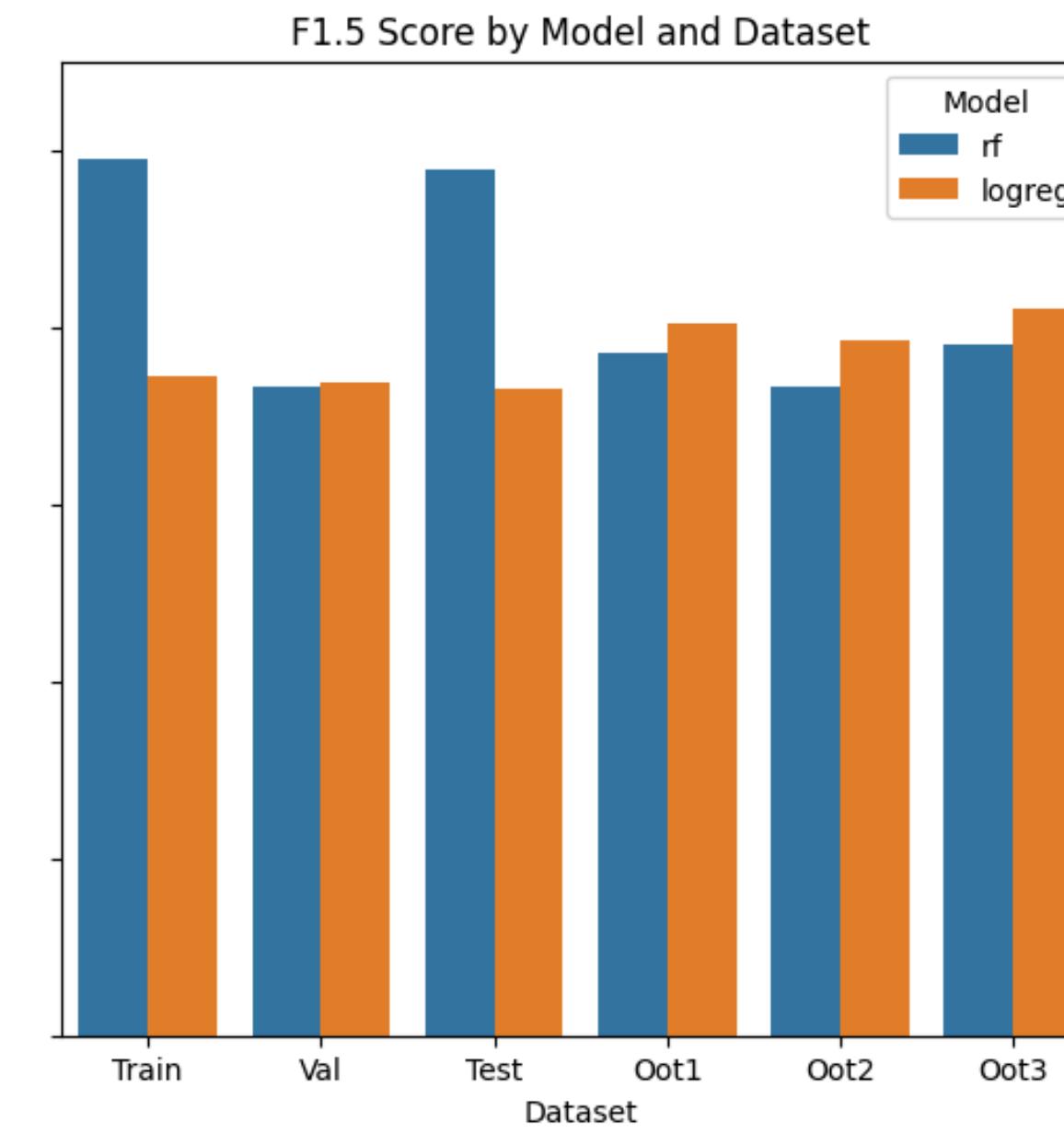
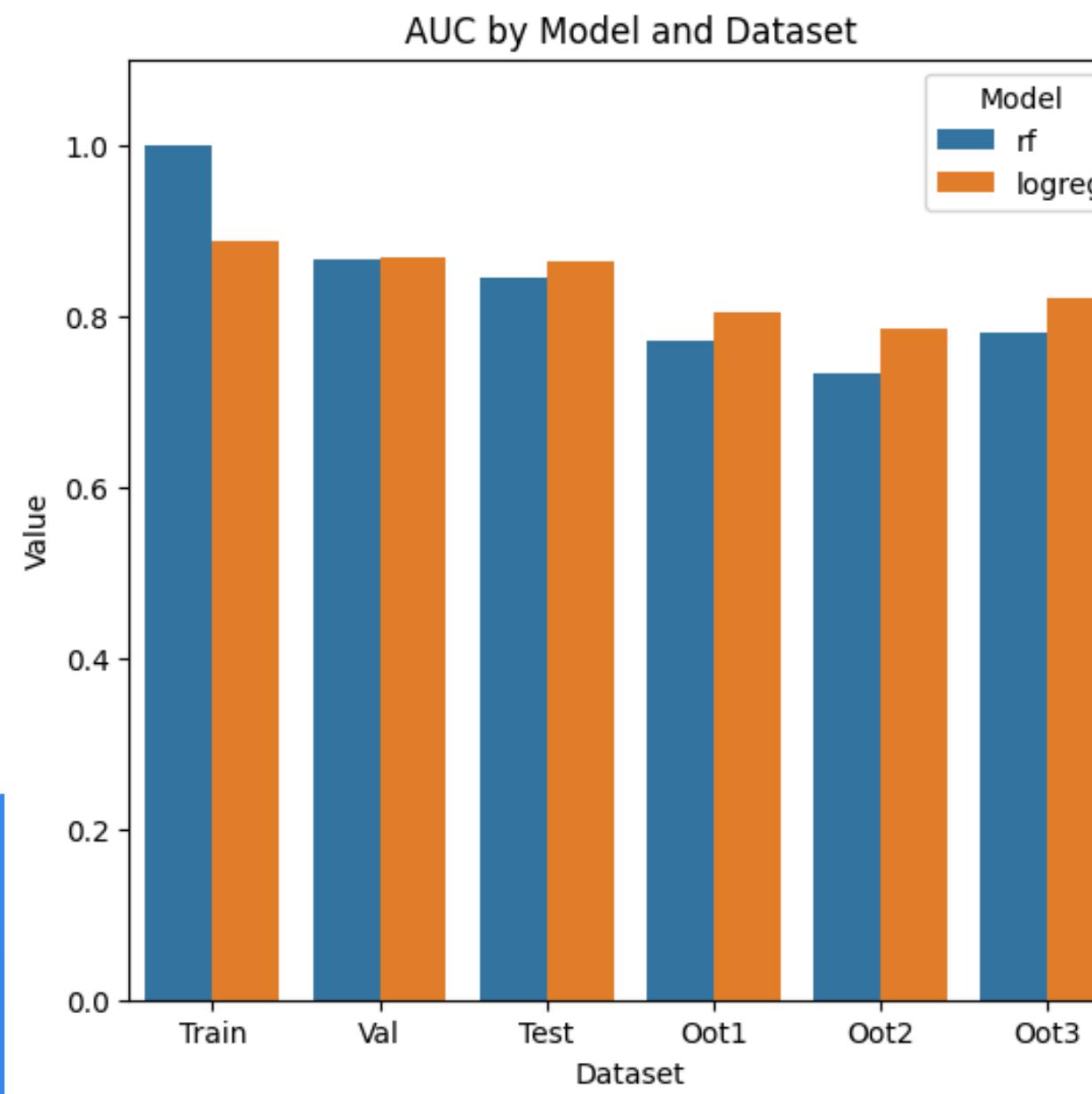


True label distribution over months



MODEL PERFORMANCE MONITORING

Monitors the performance of newly-trained models to select the best one.

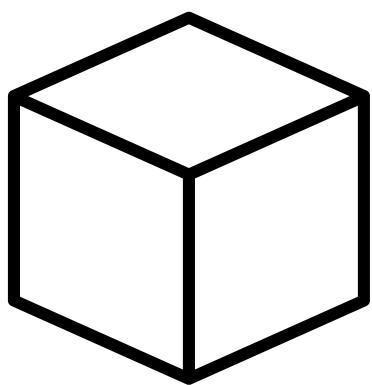


- Random Forest
Classifier seems to overfit on the train data, with a steep drop in performance in every set other than Test
- Therefore, **Logistic Regression** is chosen as the model of choice for CreditKarma Scorer

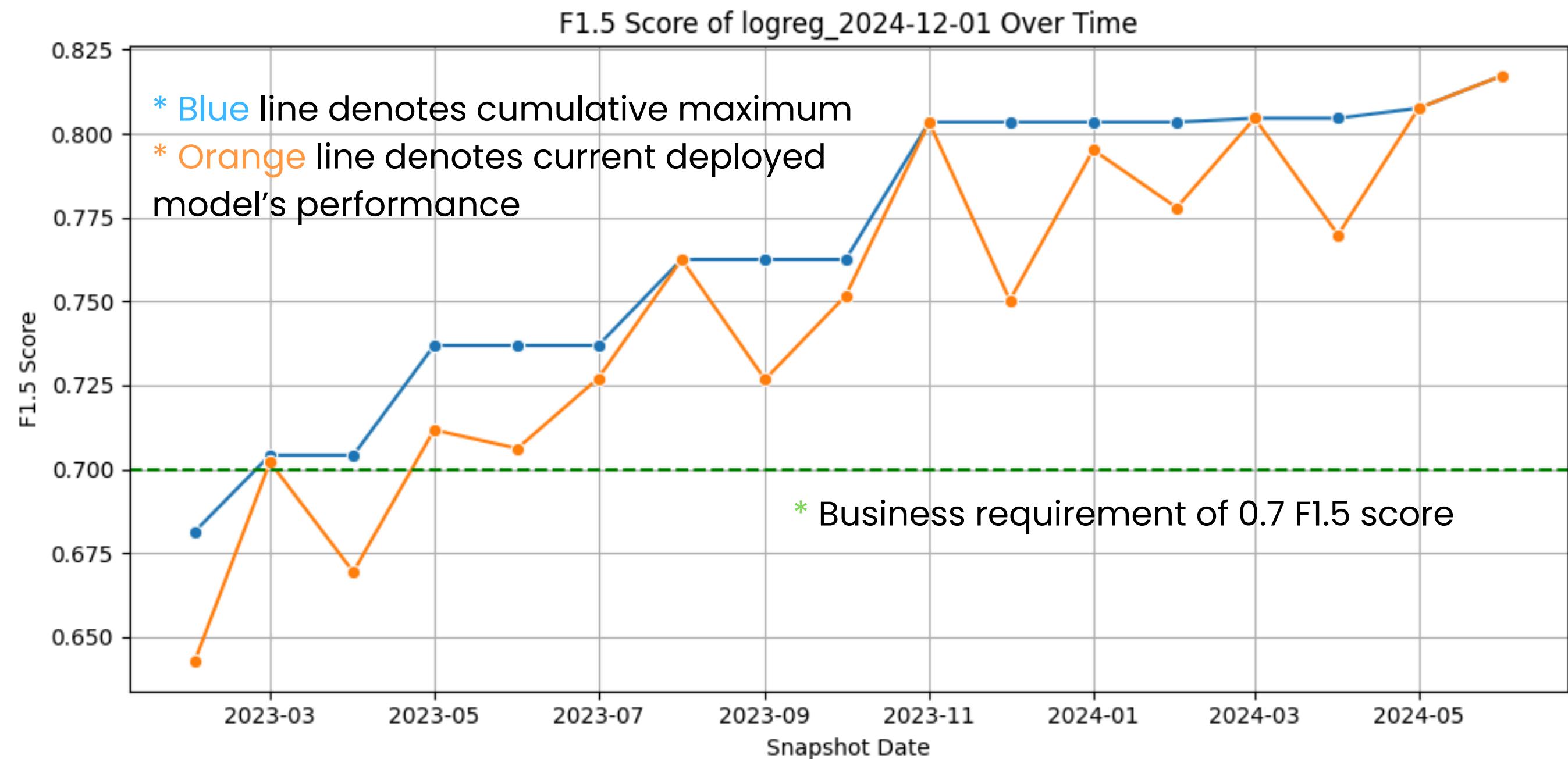
* **F1.5 Score** is chosen as the primary validation metric because we want to slightly prioritize recall over precision. Giving out loans to customers who default will result in greater loss, but we don't want to be too conservative.

MODEL STABILITY MONITORING

Monitors the stability of the deployed model version over time to alert for potentially having to deploy a new version.



Pull
“champion”
model



* Inference results are backfilled by Airflow when a new champion model is deployed