



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.

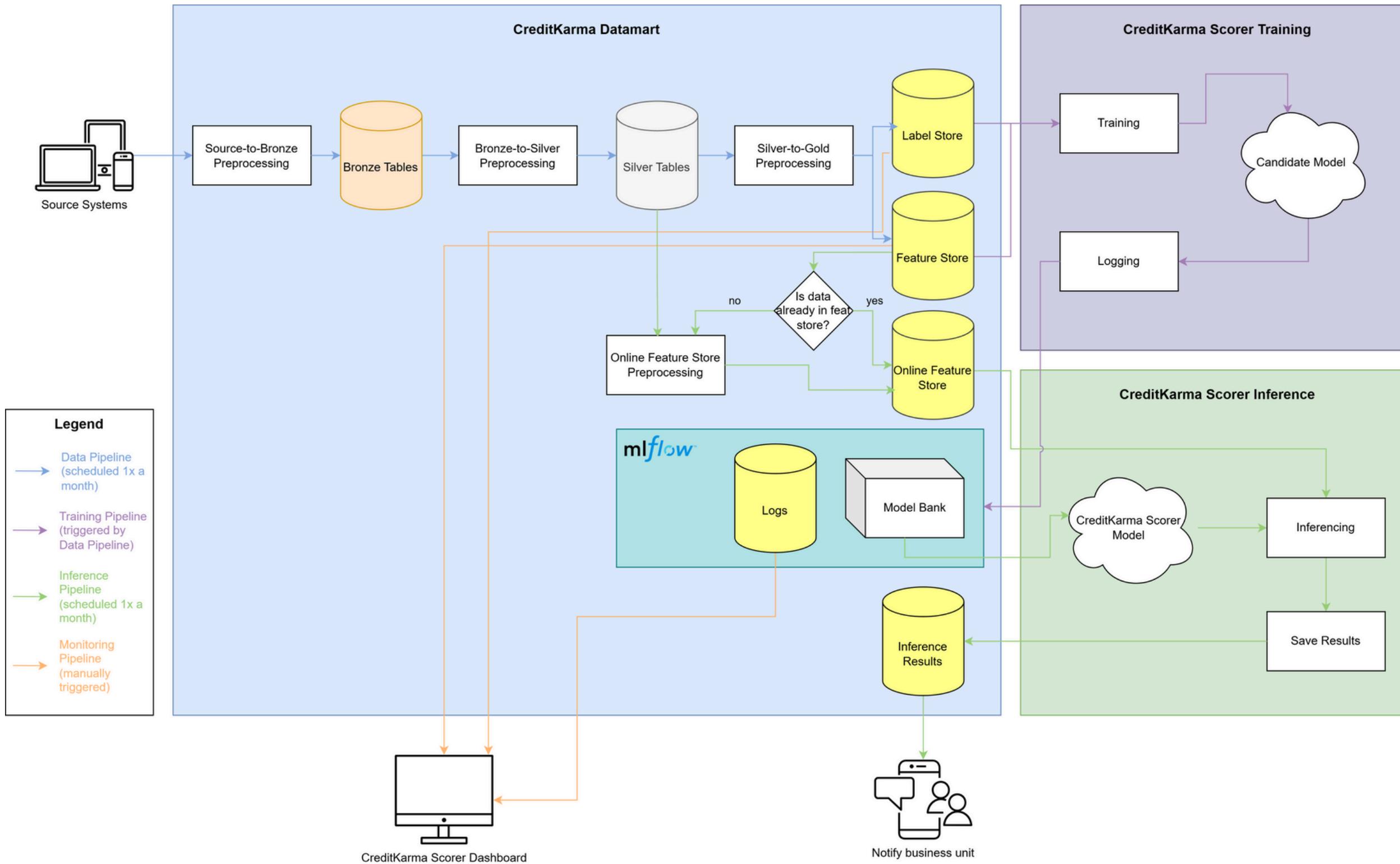
Currently comprises of two main components:

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

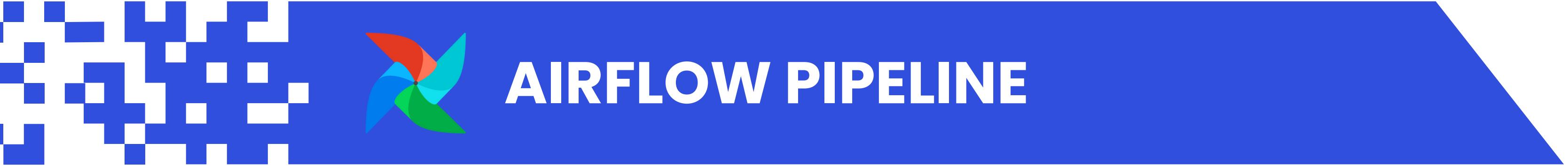
1. Deployed end-to-end machine learning pipeline using Airflow.
 - a. Monthly automated data processing pipeline.
 - b. Automated model training pipeline.
 - c. Model inferencing pipeline.
2. Created central model registry for CreditKarma Scorer versioning.
3. Created prototype model monitoring dashboard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

CREDITKARMA SUITE ARCHITECTURE

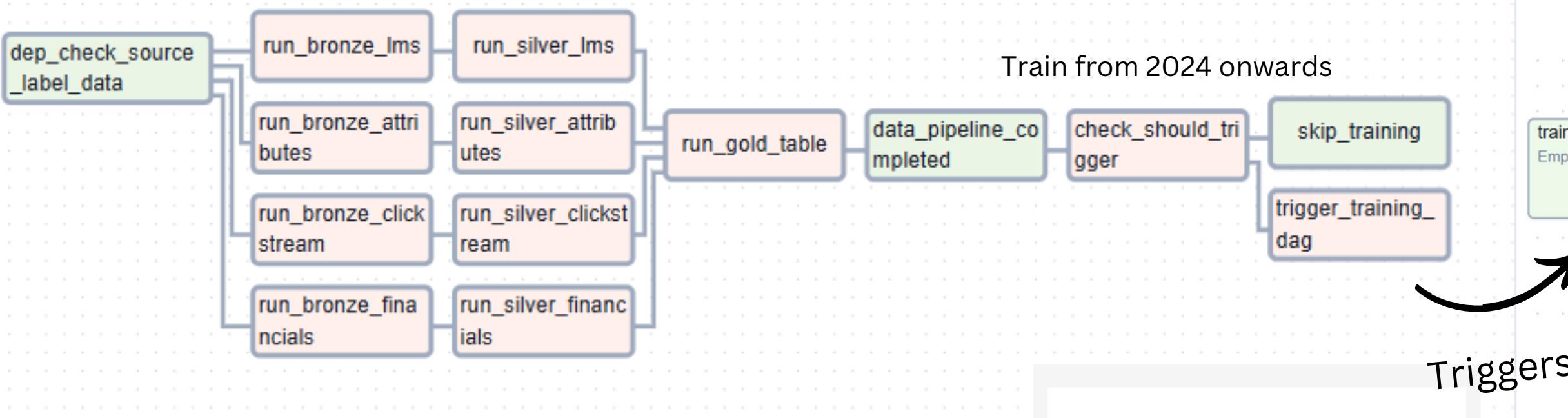


Pipeline is orchestrated using **Airflow**

- 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



Data Pipeline DAG

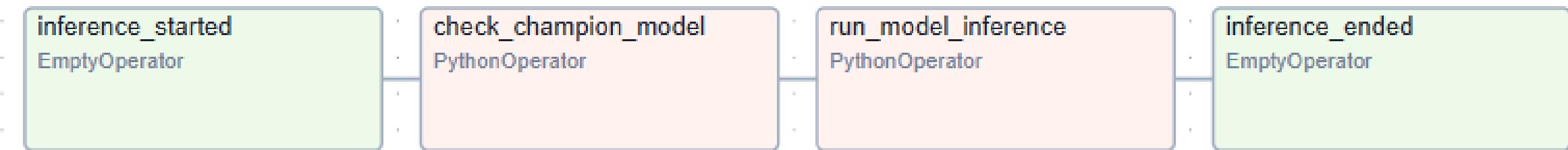


Training Pipeline DAG



Triggers

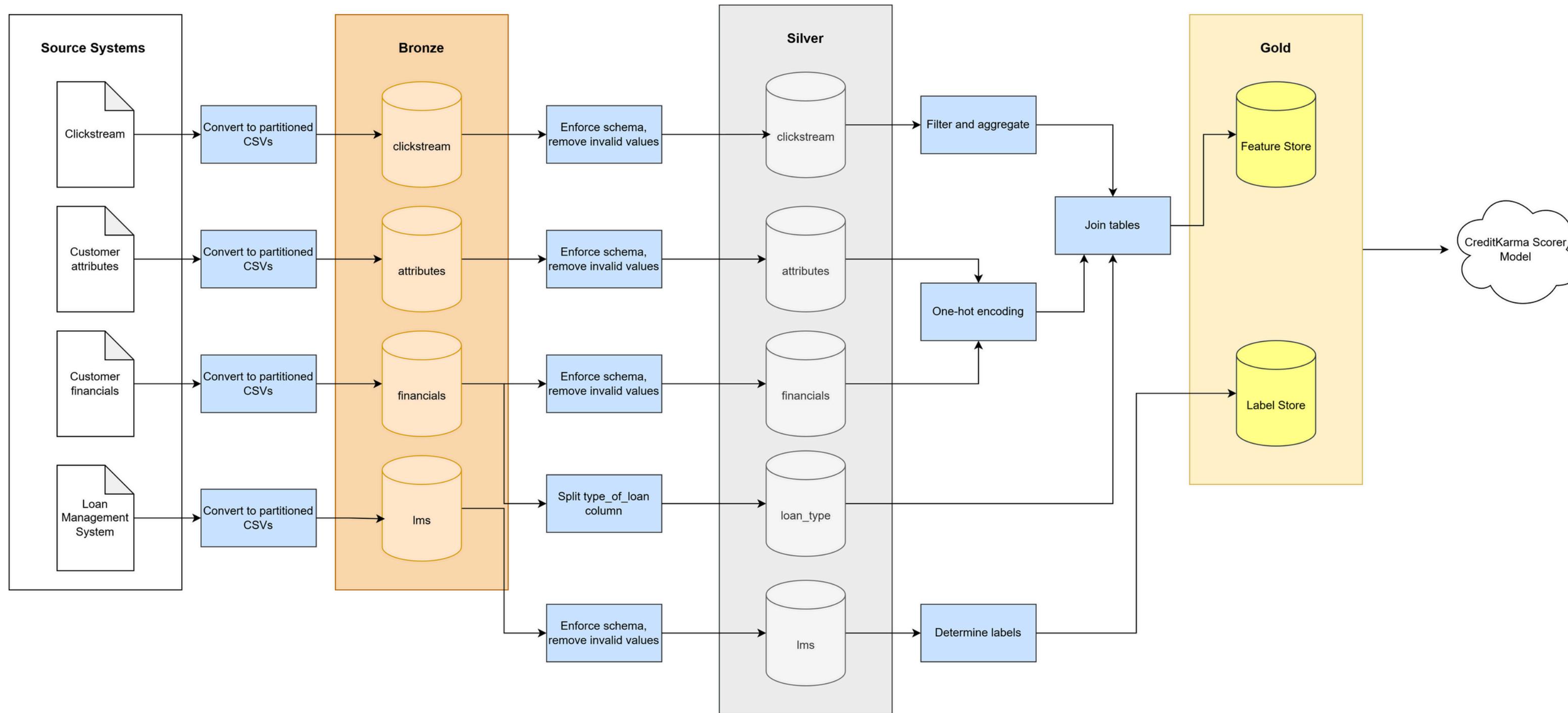
Inference Pipeline DAG



Manually triggered after new champion selected

CREDITKARMA DATAMART

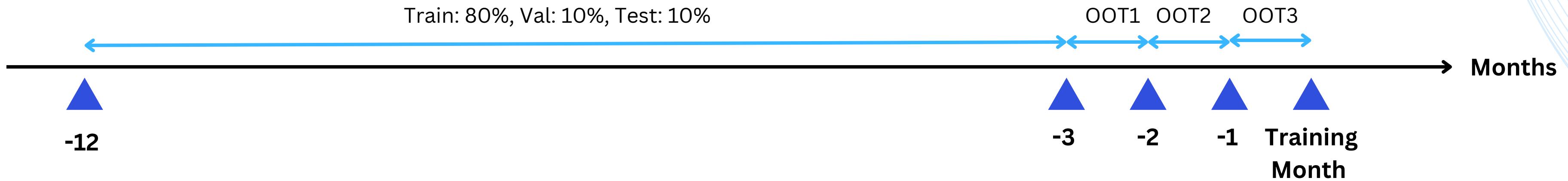
Follows the Medallion Architecture for consistency across applications and ease of maintainability



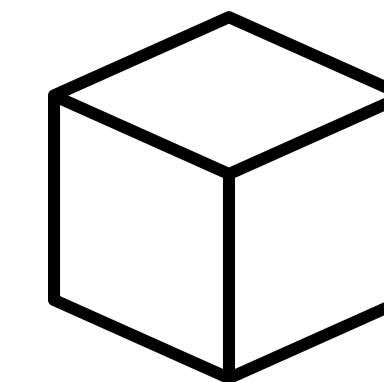
MODEL TRAINING

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 Bank

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 displays the mlflow 2.22.1 web interface. The top navigation bar includes links for mlflow 2.22.1, Experiments, Models, and Prompts, along with GitHub and Docs links. The main area shows the 'Experiments' tab selected, listing training runs for January 2024. One run, 'rf_best_model', is highlighted with a blue arrow pointing to its details in the 'Experimental' tab. This tab shows the run name, creation date ('1 day ago'), and various hyperparameters like 'rf_n_estimators'. A search bar at the top right filters results by 'metrics.rmse < 1 and params.model = "tree"'. To the right, a detailed view of the 'rf_best_model' run is shown, including an 'Overview' table with parameters like 'model_train_date_str' and 'train_test_period_months', and a 'Metrics' table with performance scores such as 'train_auc' and 'test_f1.5_score'.

Parameter	Value
model_train_date_str	2024-12-01
train_test_period_months	12
oot_period_months	3
model_train_date	2024-12-01
oot_end_date	2024-11-30
oot_start_date	2024-09-01
train_test_end_date	2024-08-31
train_test_start_date	2023-09-01
train_valtest_ratio	0.2
val_test_ratio	0.5
n_estimators	50
max_depth	20
min_samples_split	2
min_samples_leaf	1

Metric	Value
train_auc	1
val_auc	0.8651194596
test_auc	0.8451610644
oot1_auc	0.7697368421
oot2_auc	0.7319538017
oot3_auc	0.7811158798
train_f1.5_score	0.9898178526
val_f1.5_score	0.7326078507
test_f1.5_score	0.9780883174
oot1_f1.5_score	0.7697368421
oot2_f1.5_score	0.7319538017
oot3_f1.5_score	0.7811158798

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

Current

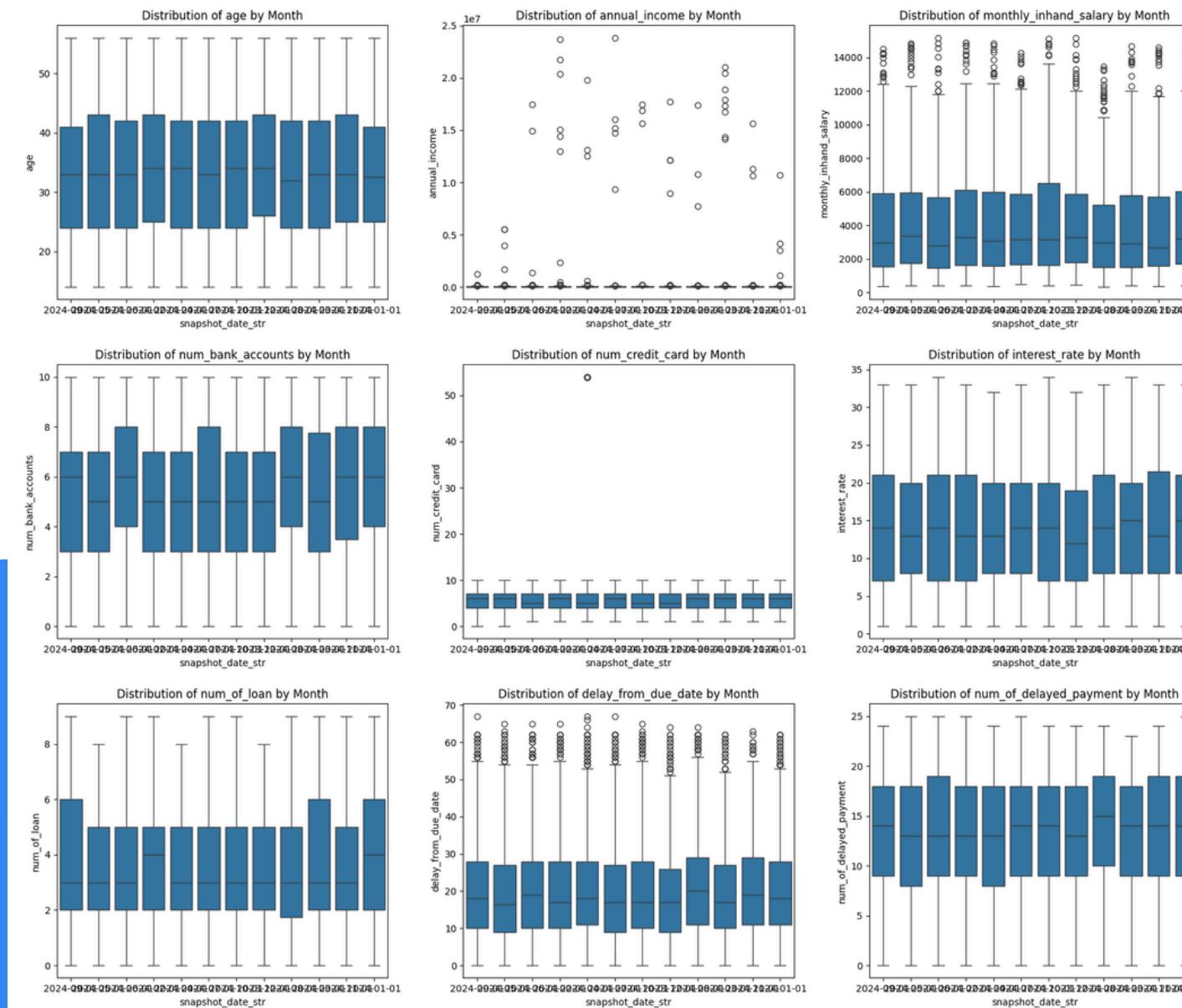
Old

1 2 3 >

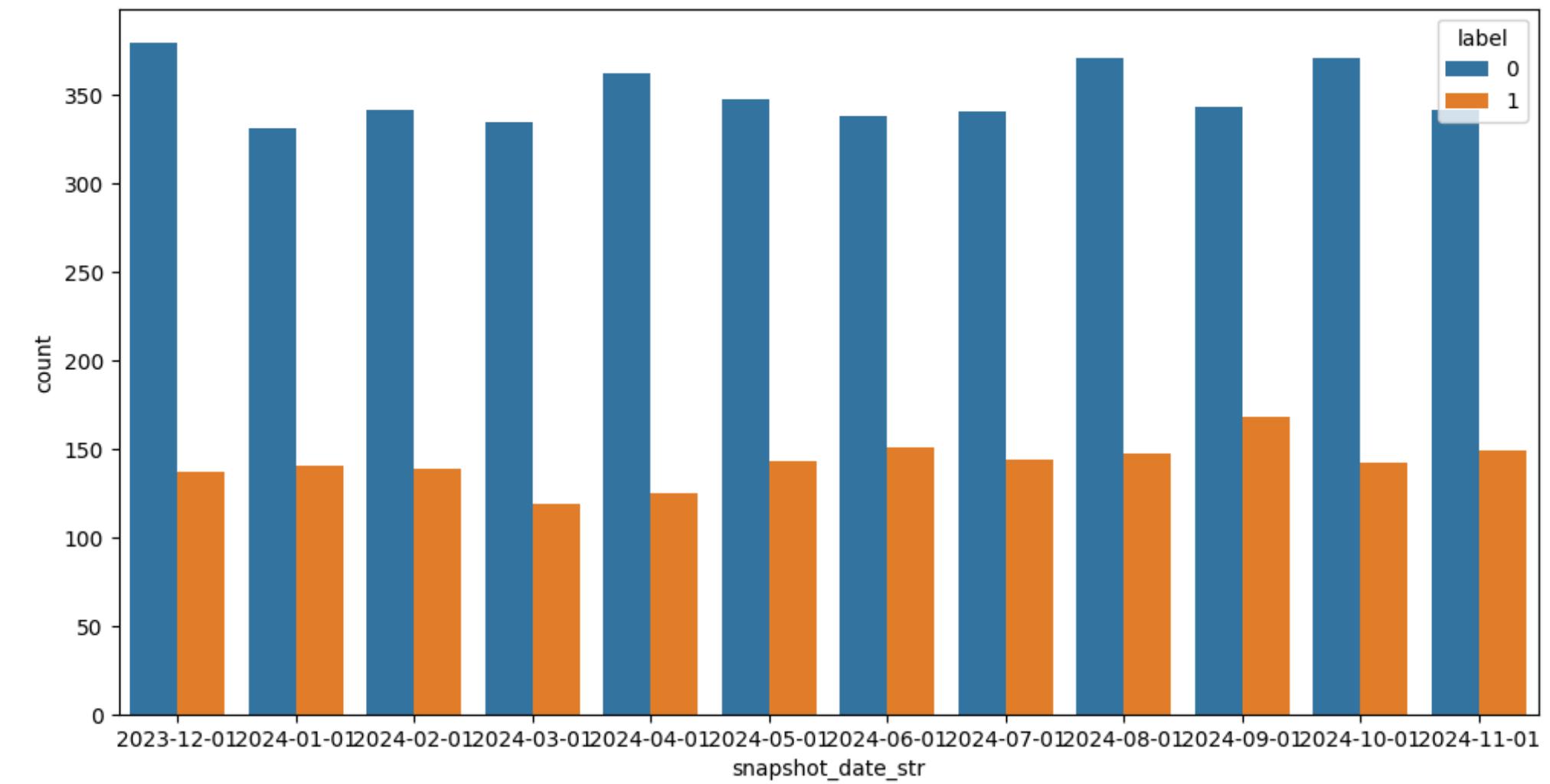
DATA MONITORING

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

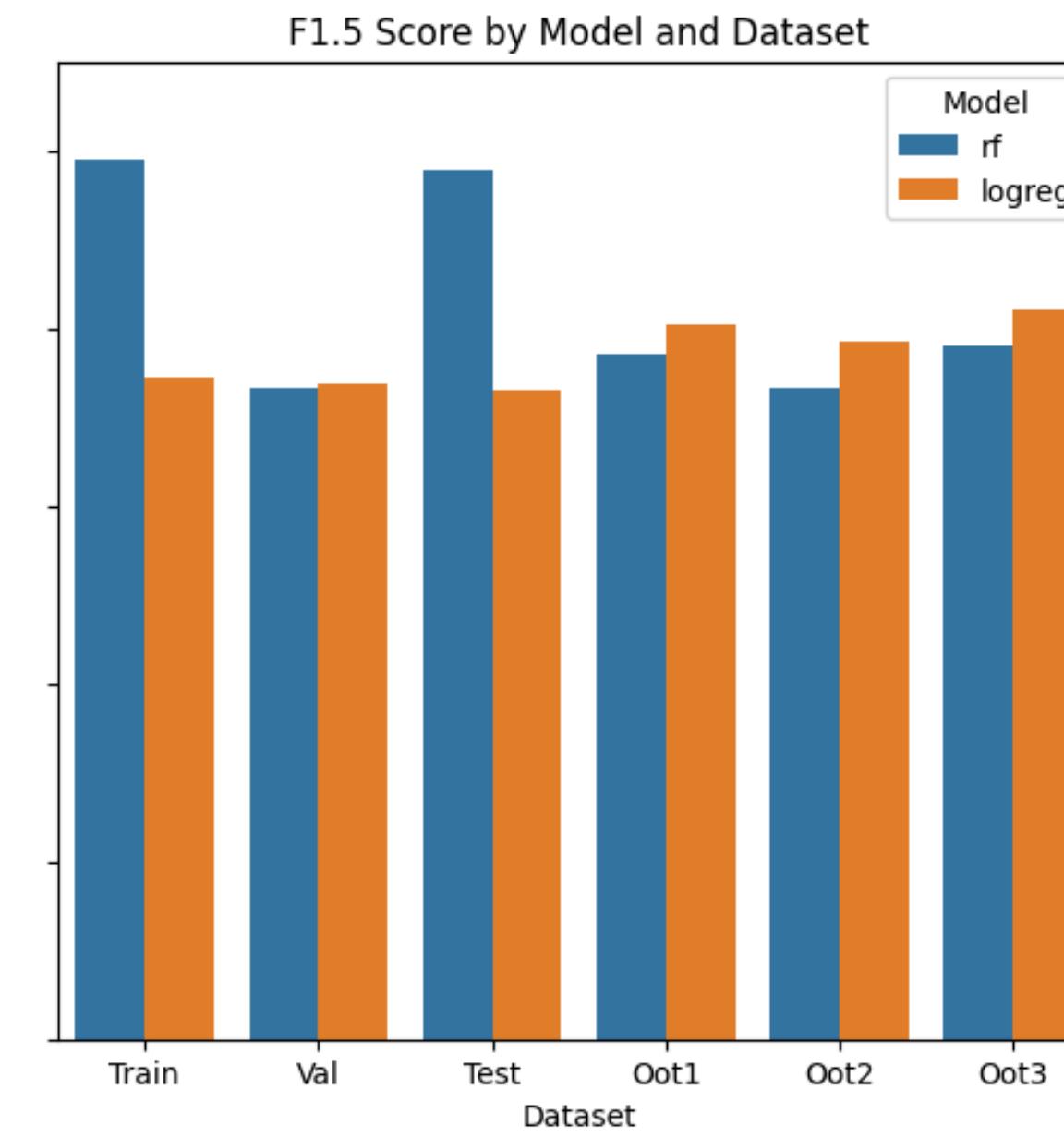
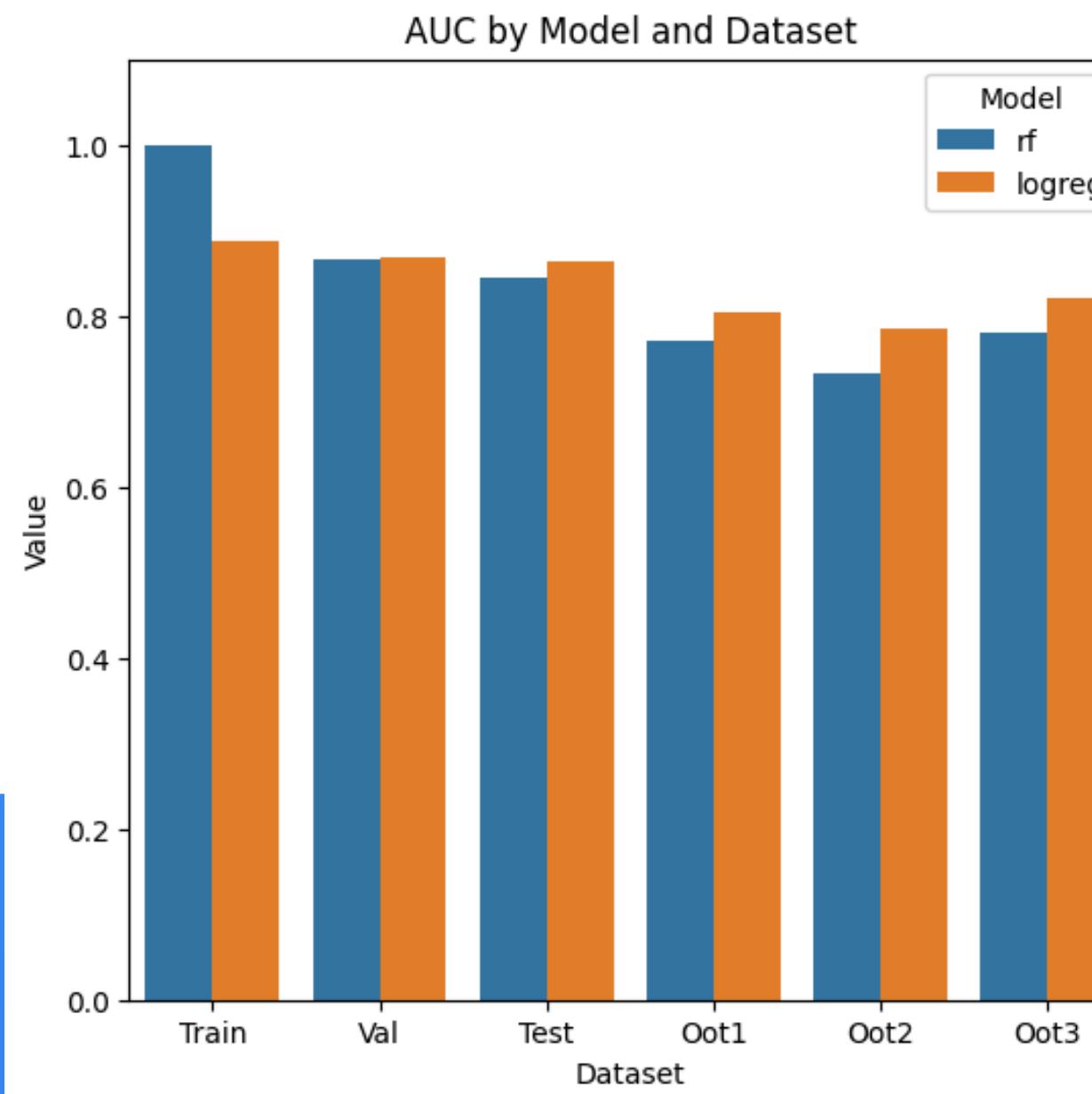


True label distribution over months



MODEL PERFORMANCE MONITORING

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

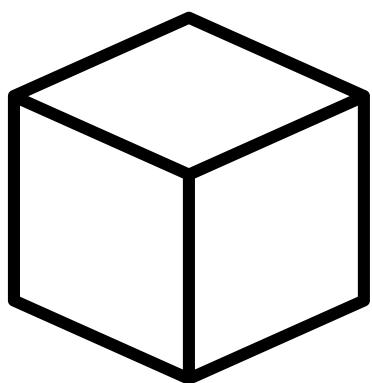


- Random Forest
Classifier seems to overfit on the train data, with a steep drop in performance in every set other than Test
- Therefore, **Logistic Regression** is chosen as the model of choice for CreditKarma Scorer

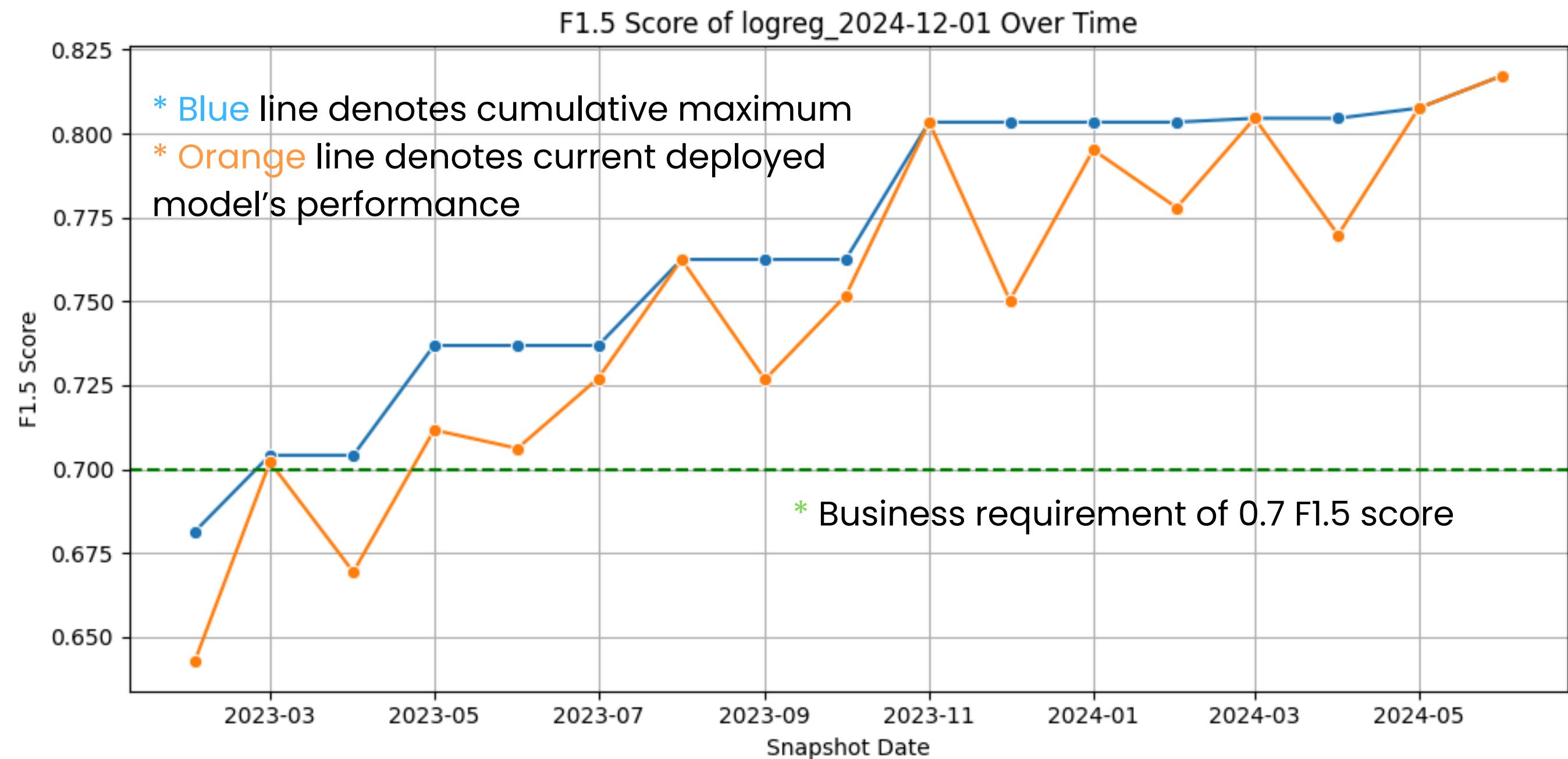
* **F1.5 Score** is chosen as the primary validation metric because we want to slightly prioritize recall over precision. Giving out loans to customers who default will result in greater loss, but we don't want to be too conservative.

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