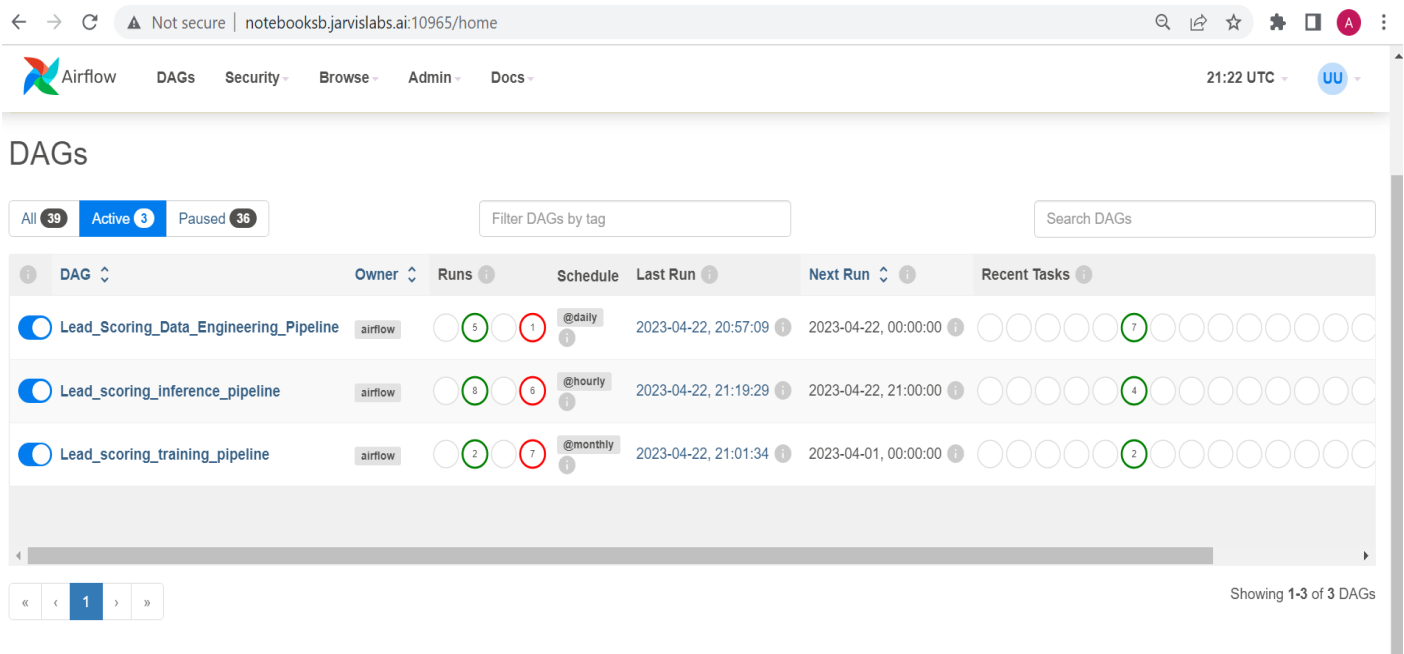# Airflow DAGs Dashboard
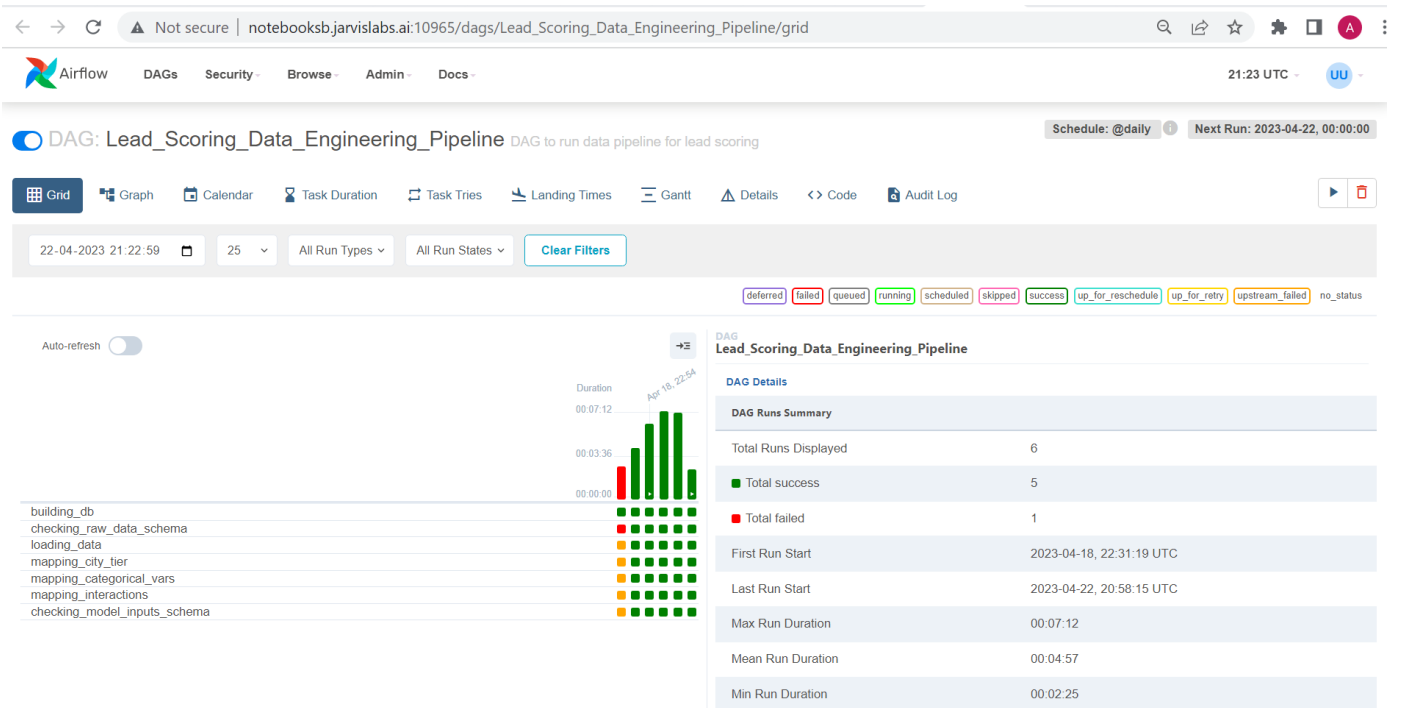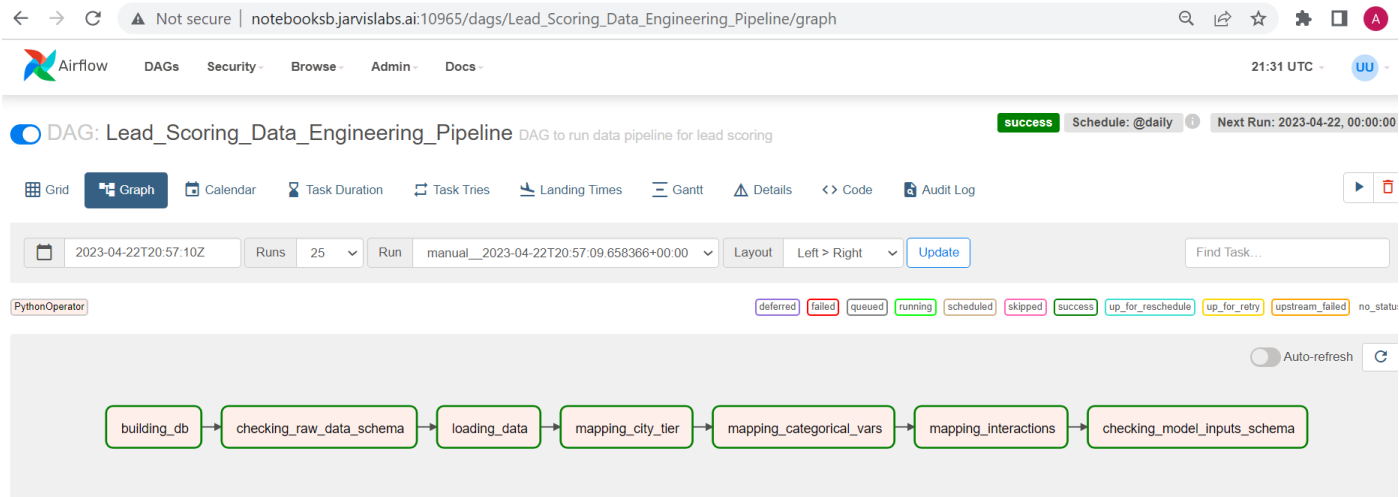
Showing DAGs for the assignment
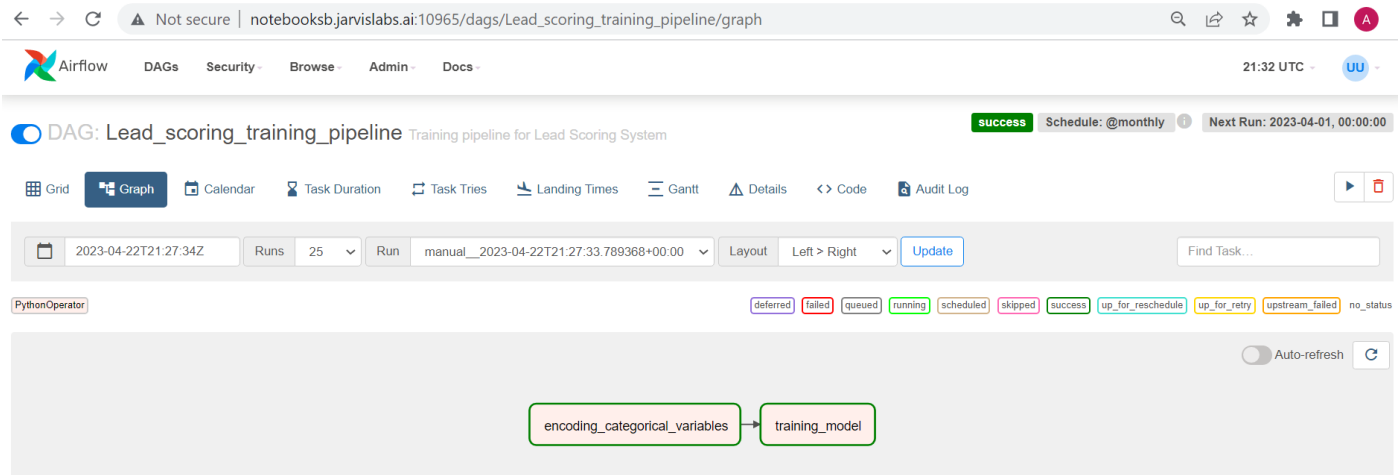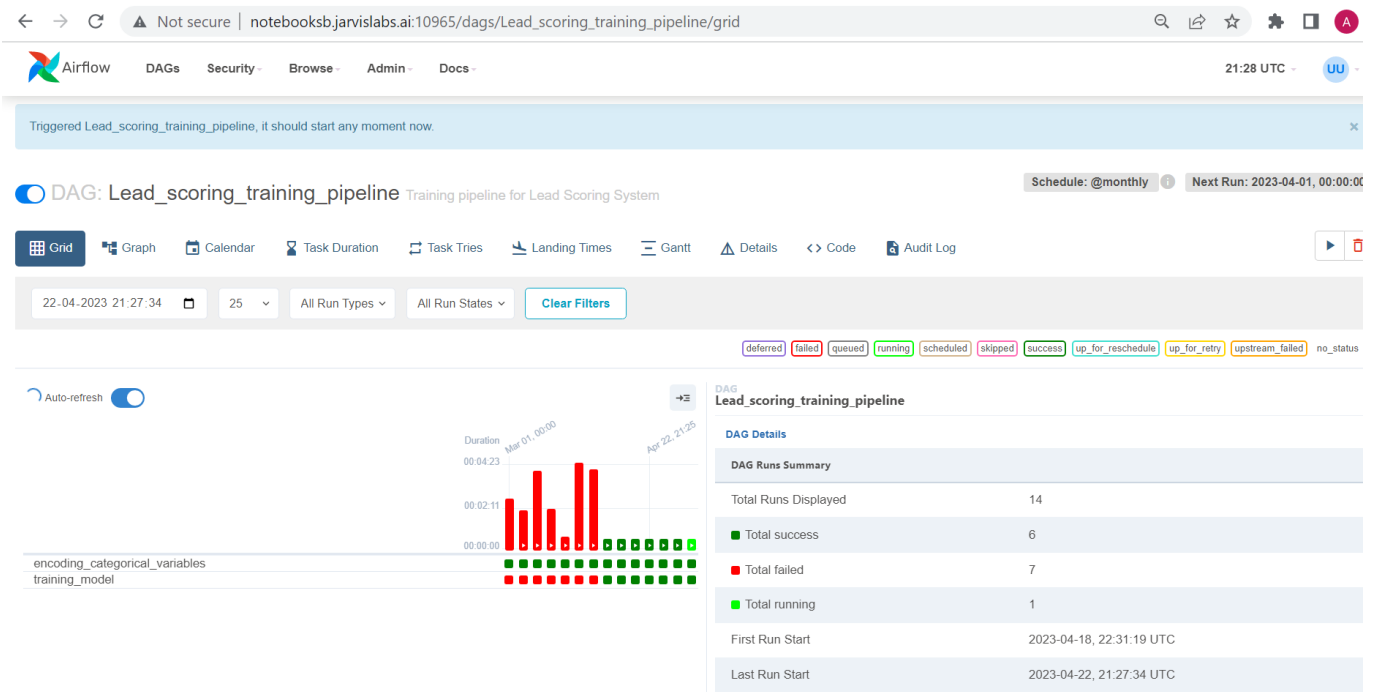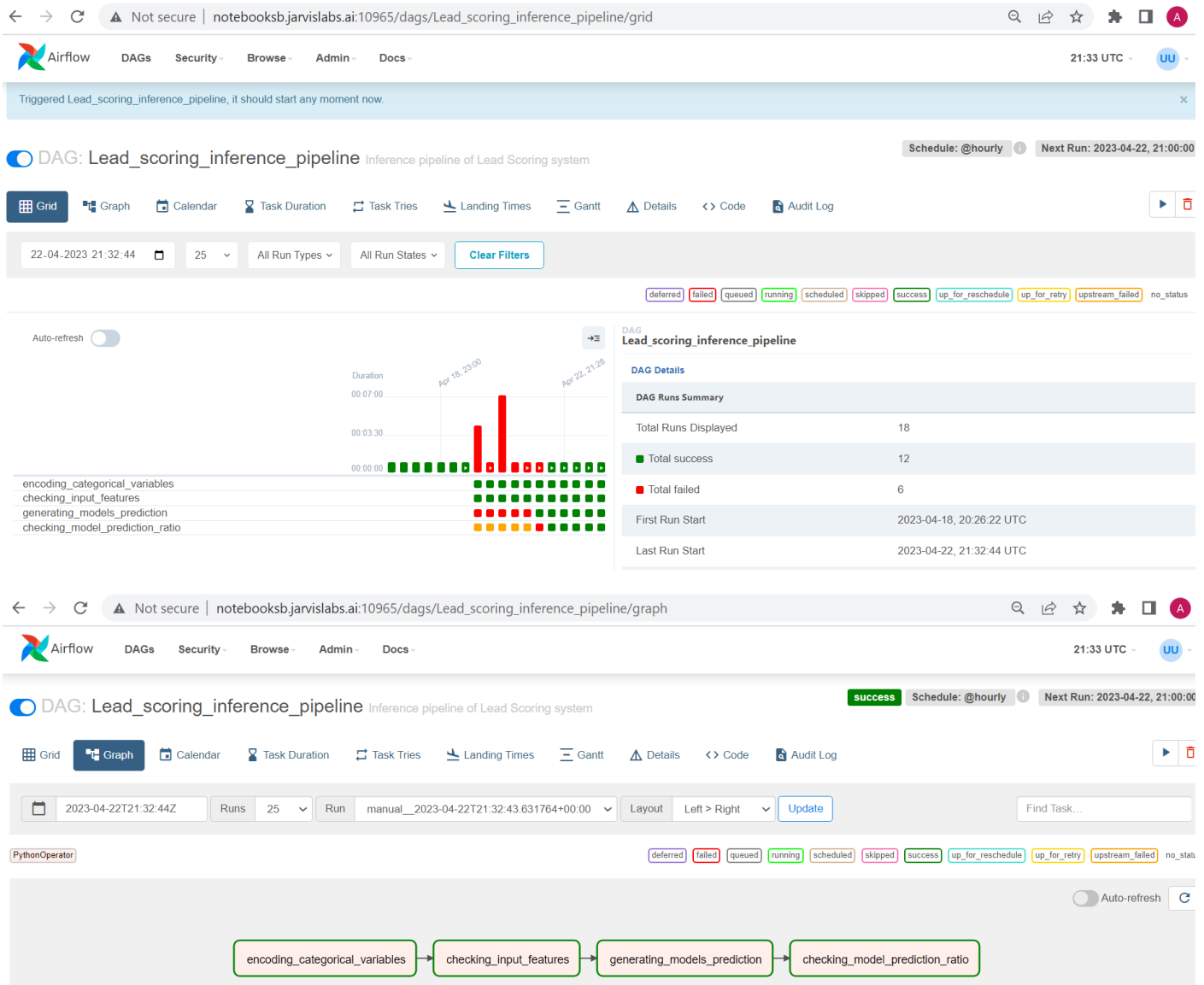

Data Pipeline

Training pipeline

Inference pipeline

Changes done in Data pipeline for inference file

(*As per Automation Workflow of Inference Pipeline, 11th bullet point*)
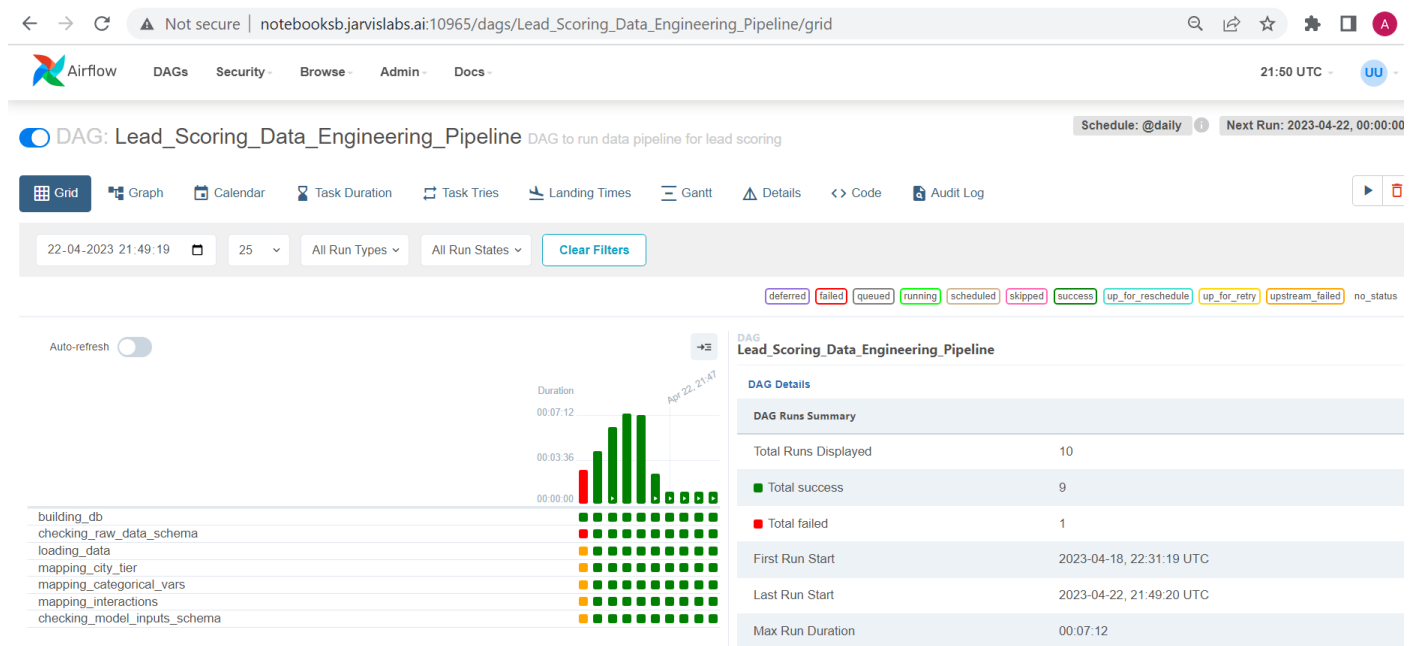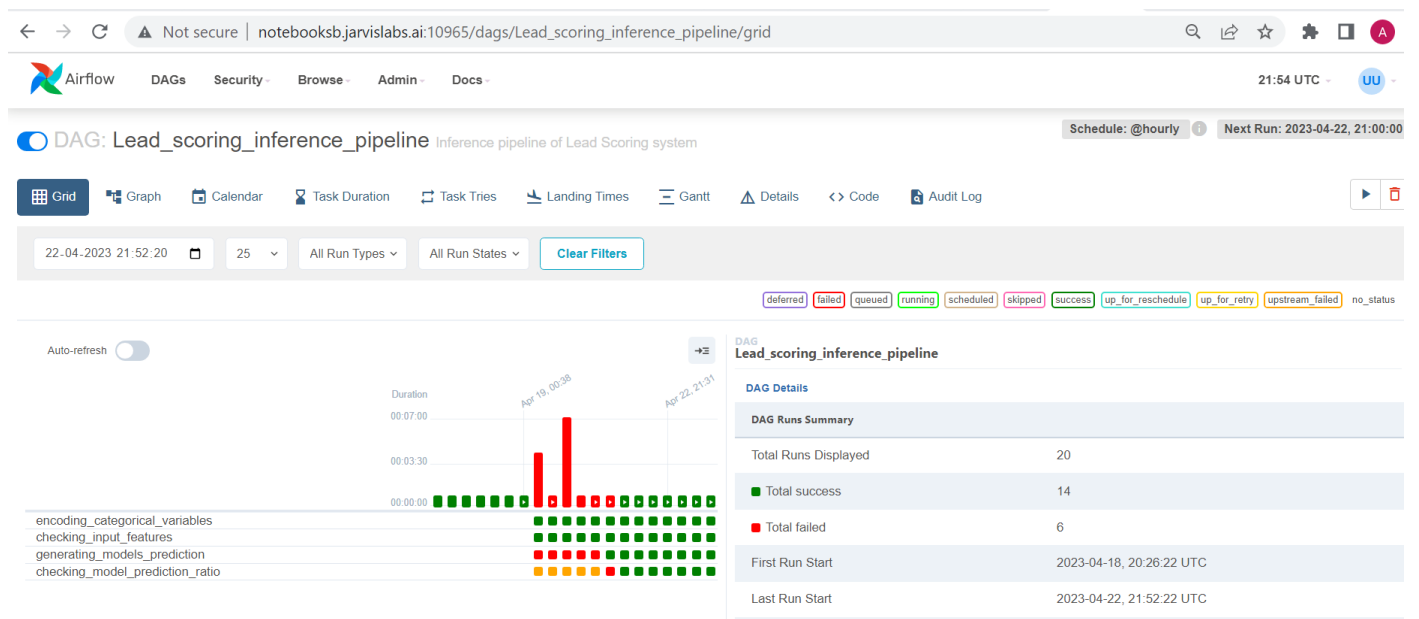
Lead Scoring Data Pipeline -> Constants file

Data Pipeline

Once this is done first trigger the Data pipeline manually and then trigger the inference pipeline once the data pipeline's run is complete.

Inference pipeline

ML Flow server

Models

Metrics

mlflow 1.26.1   Experiments   Models

Default > Lead_scoring_mlflow_production

Lead_scoring_mlflow_production

Date: 2023-04-23 02:57:56   Source: ▢ airflow   User: root
Duration: 4.3s   Status: FINISHED   Lifecycle Stage: active

▸ Description   Edit

▸ Parameters (20)

▾ Metrics (12)

Name	Value
False Negative	3956
Precision	0.716
Precision_0	0.831
Precision_1	0.661
Recall	0.746
Recall_0	0.542
Recall_1	0.89
True Negative	19502
f1_0	0.656
f1_1	0.759
roc_auc	0.716
test_accuracy	0.716

▸ Tags

▾ Artifacts

▾ ▇ models
   ▪ MLmodel
   ▪ conda.yaml
   ▪ model.pkl

Full Path:/home/mlruns/0/0b12c9f66942403bb2b59c890b7ecab9/artifacts/models ▢

MLflow Model

Parameters

▸ Description   Edit

▾ Parameters (20)

Name	Value
boosting_type	gbdt
class_weight	None
colsample_bytree	1.0
importance_type	split
learning_rate	0.1
max_depth	-1
min_child_samples	20
min_child_weight	0.001
min_split_gain	0.0
n_estimators	100
n_jobs	-1
num_leaves	31
objective	None
random_state	42
reg_alpha	0.0
reg_lambda	0.0
silent	warn
subsample	1.0
subsample_for_bin	200000
subsample_freq	0

▾ Metrics (12)

File details:

- Complete assignment code is kept at git repo shared in AssignmentCodeAtGitRepo.txt (https://github.com/akashpd/MLOps_Assignment_PipeLine_MLFlow_and_AirFlow)
- All DAGs which were run are kept in DAGs_4_AP.zip
- inference_test.ipynb file is also kept in unit_test folder which was used to test inference pipeline.

My Jarvis lab folder structure highlighted

My DAGs folder location

Airflow Web server port configuration
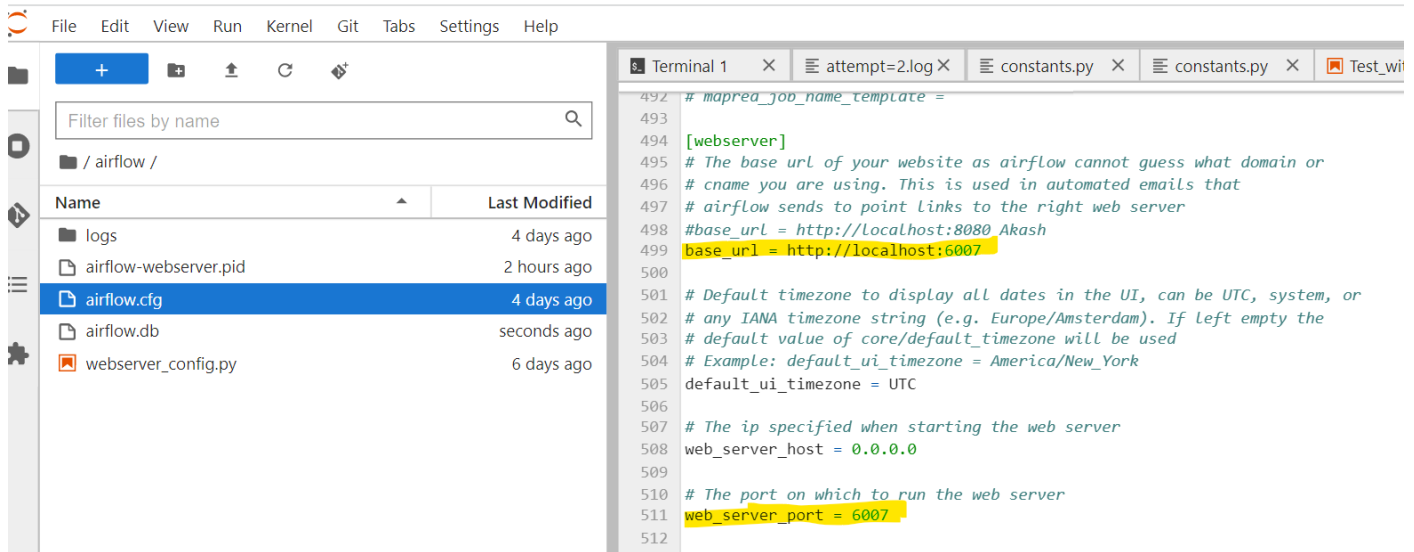
After all code is done and working correctly then run below steps in summary to run airflow and DAGs (in order)

➢ **on Terminal 1**

airflow db init
airflow webserver

➢ **on Terminal 2**

airflow scheduler

➢ **Run DAG Lead_scoring_data_pipeline from Airflow DAGs dashboard**

➢ **On Terminal 3**

cd dags

mlflow server --backend-store-
uri='sqlite:///./Lead_scoring_training_pipeline/Lead_scoring_mlflow_production.db' --default-artifact-
root="/home/mlruns" --port=6006 --host=0.0.0.0

➢ **Run DAG Lead_scoring_training_pipeline from Airflow DAGs dashboard**

➢ **Register new version Model from mlflow/models**

➢ **Run DAG Lead_scoring_inference_pipeline from Airflow DAGs dashboard**