



! DAG Import Errors (1)

The scheduler does not appear to be running. Last heartbeat was received 5 minutes ago.

The DAGs list may not update, and new tasks will not be scheduled.

Do not use **SQLite** as metadata DB in production – it should only be used for dev/testing. We recommend using Postgres or MySQL. [Click here](#) for more information.

Do not use the **SequentialExecutor** in production. [Click here](#) for more information.

DAGs

All 72	Active 0	Paused 72	Running 0	Failed 0	Filter DAGs by tag	Search DAGs		<input checked="" type="checkbox"/> Auto-refresh	C
---------------	-----------------	------------------	------------------	-----------------	--------------------	-------------	--	--	---

DAG	Owner	Runs	Schedule	Last Run		Next Run		Recent Tasks
Params Trigger UI <small>example params</small>	airflow		None					
Params UI tutorial <small>example params ui</small>	airflow		None					
Sample DAG with Display Name <small>example</small>	airflow		None					
conditional_dataset_and_time_based_timetable <small>dataset-time-based-timetable</small>	airflow		Dataset or 0 1 ** 3	2025-03-26, 01:00:00				
consume_1_and_2_with_dataset_expressions	airflow		Dataset		0 of 2 datasets updated			
consume_1_or_2_with_dataset_expressions	airflow		Dataset		0 of 2 datasets updated			
consume_1_or_both_2_and_3_with_dataset_expressions <small>Screenshot</small>	airflow		Dataset		0 of 3 datasets updated			



Do not use **SQLite** as metadata DB in production – it should only be used for dev/testing. We recommend using Postgres or MySQL. [Click here](#) for more information.

Do not use the **SequentialExecutor** in production. [Click here](#) for more information.

DAGs

All 72	Active 0	Paused 72	Running 0	Failed 0	Filter DAGs by tag	Search DAGs	Auto-refresh	C
DAG								
			Owner	Runs	Schedule	Last Run	Next Run	Recent Tasks
Params Trigger UI			airflow	None				
	example	params						
Params UI tutorial			airflow	None				
	example	params	ui					
Sample DAG with Display Name			airflow	None				
	example							
conditional_dataset_and_time_based_timetable			airflow	Dataset or 0 1 * * 3		2025-03-26, 01:00:00		
	dataset-time-based-timetable							
consume_1_and_2_with_dataset_expressions			airflow	Dataset		0 of 2 datasets updated		
consume_1_or_2_with_dataset_expressions			airflow	Dataset		0 of 2 datasets updated		
consume_1_or_both_2_and_3_with_dataset_expressions			airflow	Dataset		0 of 3 datasets updated		
dataset_alias_example_alias_consumer			airflow	Unresolved DatasetAlias				
	consumer	dataset-alias						
dataset_alias_example_alias_consumer_with_no_taskflow			airflow	Unresolved DatasetAlias				
	consumer	dataset-alias						
dataset_alias_example_alias_producer			airflow	None				
	dataset-alias	producer	Visual Studio Code					



Unpaused DAGs

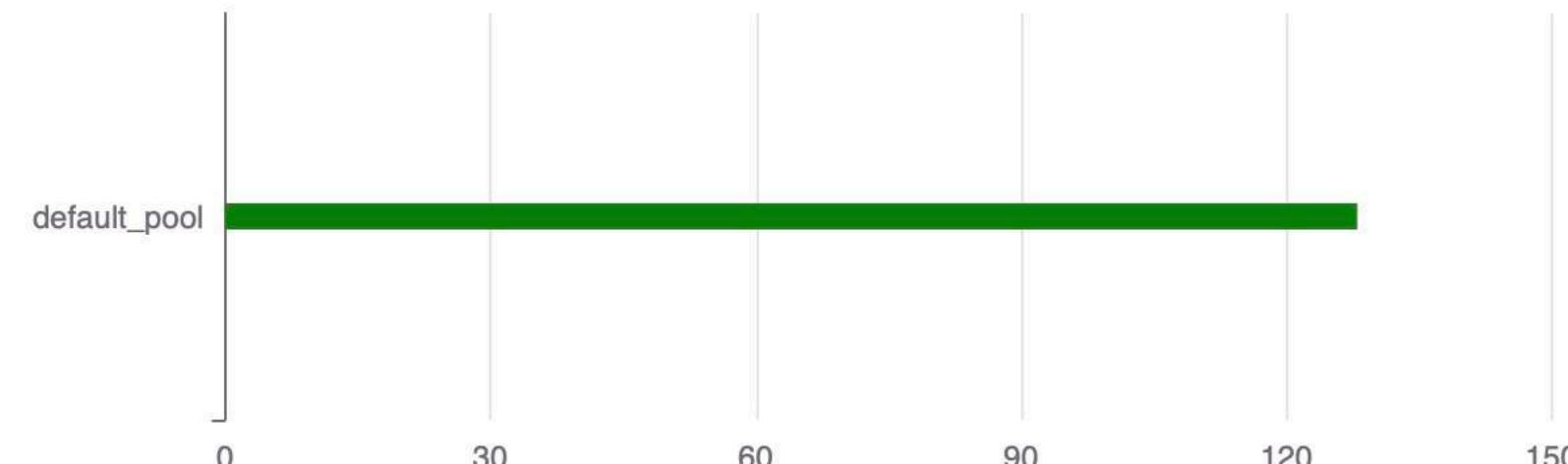
2

out of 72 total DAGs

Top 5 longest Dag Runs to finish

No dag running

out of 0 total running Dag Runs

open queued running scheduled deferred

Health

MetaDatabase

status: HEALTHY

Scheduler

status: UNHEALTHYlast heartbeat: 2025-03-30, 15:21:50 UTC

Triggerer

status: UNKNOWN

Historical metrics ⓘ

Start Date

03/29/2025 04:23:53 PM

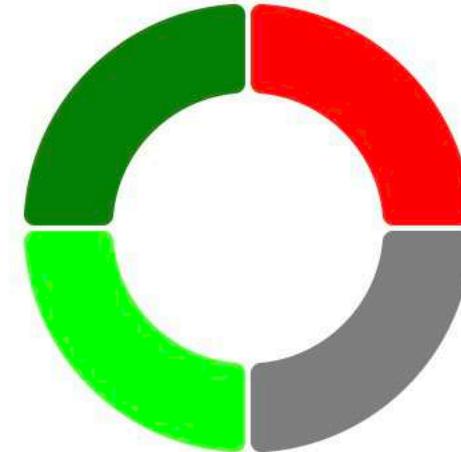
End Date

03/30/2025 04:23:53 PM

for a period of 1d00:00:00

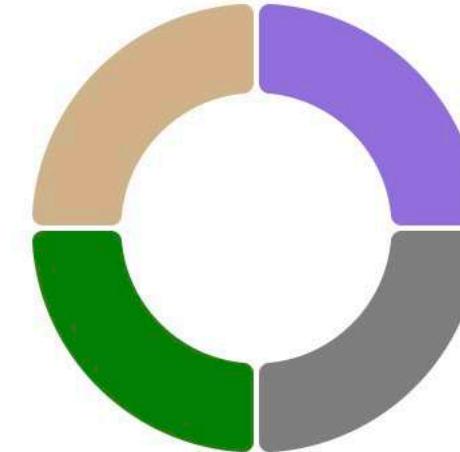
Clear Filters

Dag Run States

failed queued running success

on a total of 0

Dag Run Types

backfill datasetTriggered manual scheduled

on a total of 0

Task Instance States

deferred failed noStatus queued 1/5 ▶

on a total of 0

List Job

Search ▾



Record Count: 5

Id	Dag Id	State	Job Type	Start Date	End Date	Latest Heartbeat	Executor Class	Hostname	Unixname
5		success	SchedulerJob	2025-03-30, 15:21:38	2025-03-30, 15:21:53	2025-03-30, 15:21:50		anands-macbook-air.local	anandjha
4		success	SchedulerJob	2025-03-30, 14:42:14	2025-03-30, 14:45:23	2025-03-30, 14:45:19		anands-macbook-air.local	anandjha
3		success	SchedulerJob	2025-03-30, 14:35:36	2025-03-30, 14:39:45	2025-03-30, 14:39:40		anands-macbook-air.local	anandjha
2		success	SchedulerJob	2025-03-30, 13:55:55	2025-03-30, 14:23:43	2025-03-30, 14:23:37		anands-macbook-air.local	anandjha
1		success	SchedulerJob	2025-03-30, 13:47:48	2025-03-30, 13:52:28	2025-03-30, 13:52:25		anands-macbook-air.local	anandjha

The scheduler does not appear to be running. Last heartbeat was received 2 minutes ago.

The DAGs list may not update, and new tasks will not be scheduled.

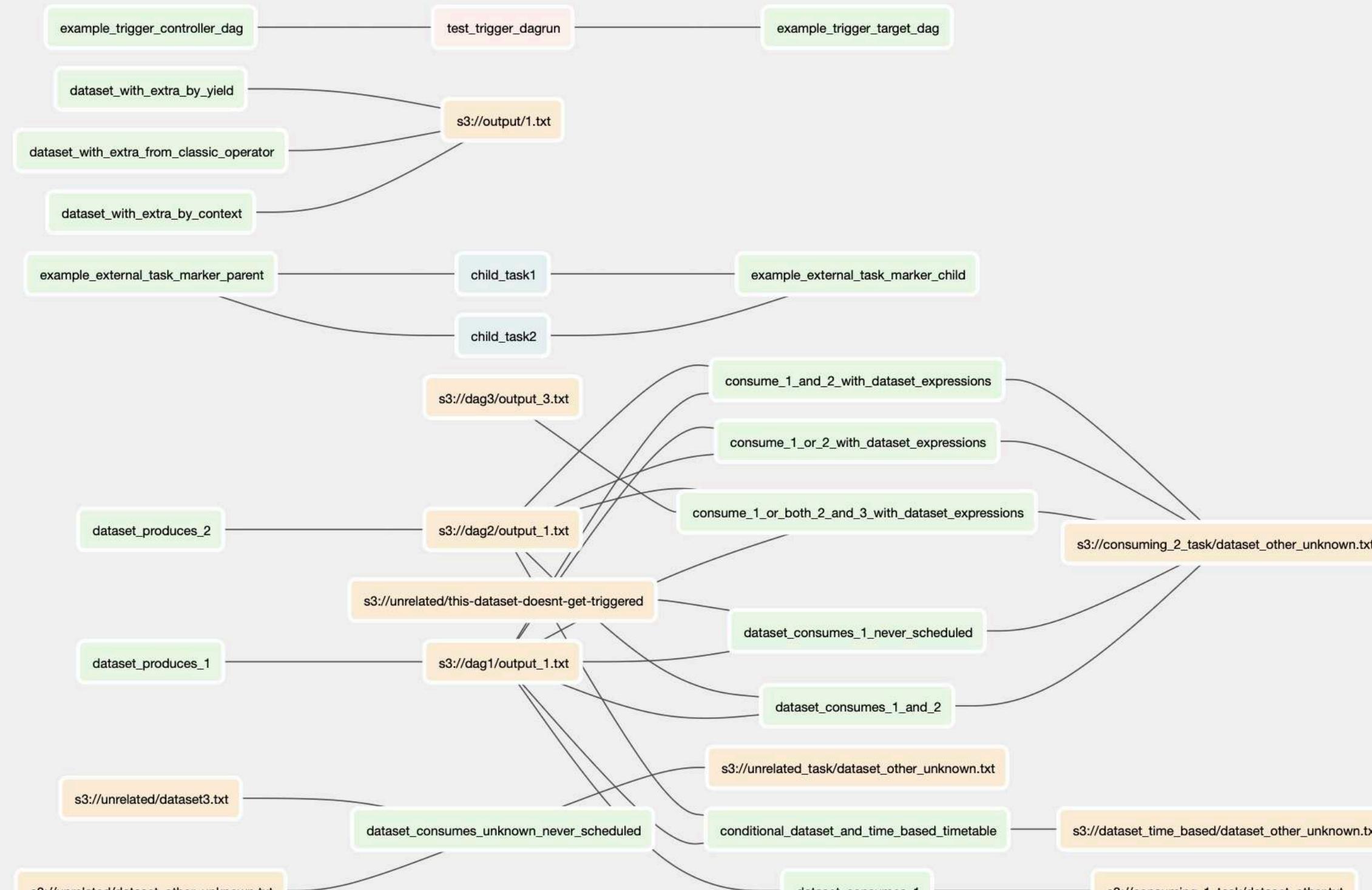
DAG Dependencies

Only show DAGs with dependencies

Search for...

`dag` `trigger` `sensor` `dataset` `dataset alias`

Last refresh: 2025-03-30, 15:24

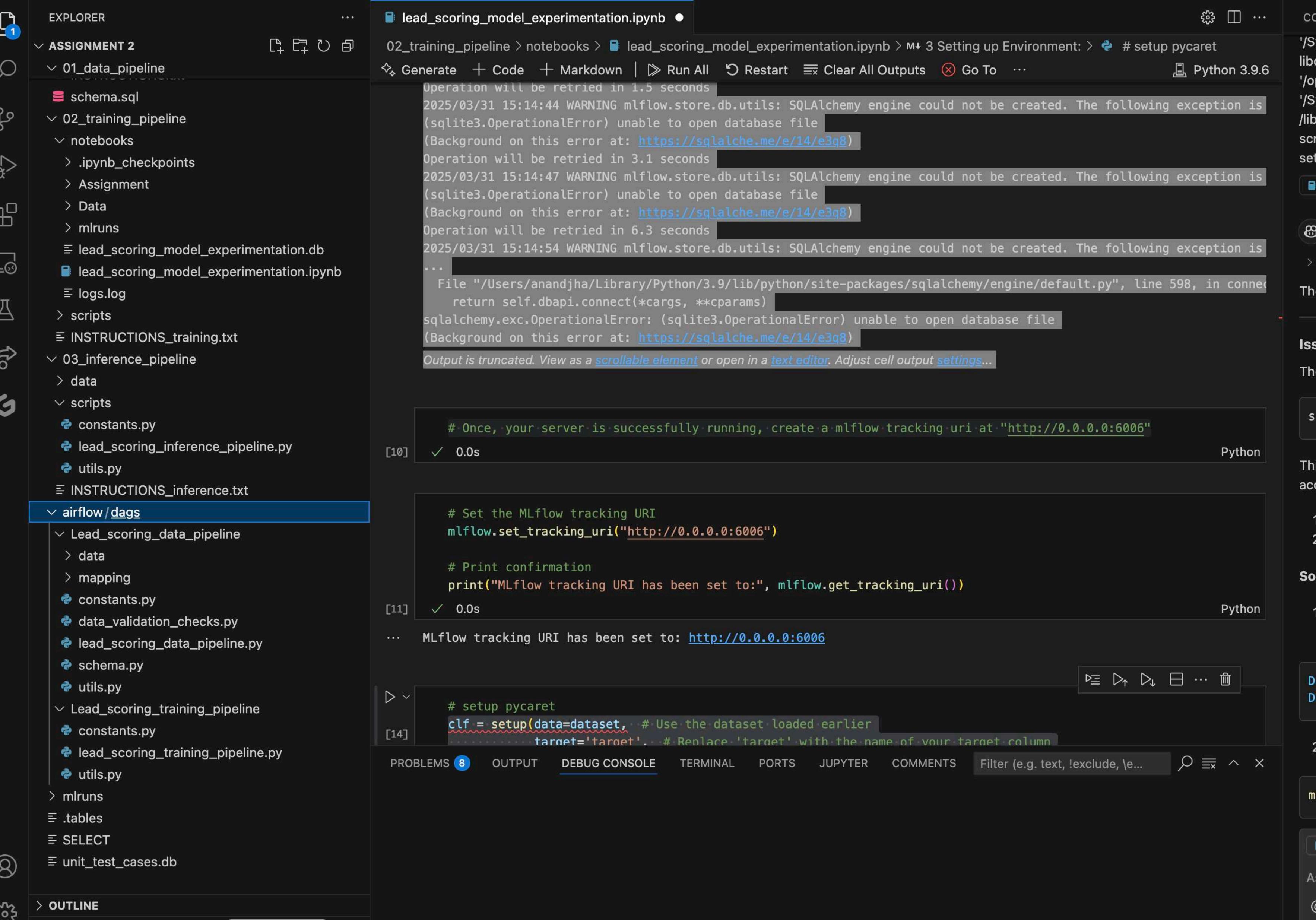


Screensho

dataset_s3_bucket_producer_with_no_taskflc

s3://bucket/my-task-with-no-task

dataset_s3_bucket_consumer_with_no_taskf



EXPLORER

ASSIGNMENT 2

- > .pytest_cache
- > .vscode
- < 01_data_pipeline
 - > notebooks
 - < scripts
 - > .ipynb_checkpoints
 - > .pytest_cache
 - < 01_data_pipeline/scripts
 - > data
 - < unit_test
 - > .ipynb_checkpoints
 - > 01_data_pipeline
 - > mapping
 - < constants.py
 - < data_validation_checks.py
 - < dummbby.ipynb
 - < lead_scoring_data_pipeline.py
 - < leadscoring_test.csv
 - < schema.py
 - < test_with_pytest.py
 - < unit_test_cases.db
 - < utils_output.db
 - < utils.py
 - < city_tier_mapping.py
 - < constants.py
 - < data_validation_checks.py
 - < dummbby.ipynb
 - < interaction_mapping.csv
 - < lead_scoring_data_pipeline.py
 - < leadscoring_test.csv
 - < leadscoring.csv
 - < schema.py
 - < significant_categorical_level.py
 - < test_with_pytest.py
 - < unit_test_cases.db
 - < utils_output.db
 - < utils.py
 - < INSTRUCTIONS.txt

lead_scoring_model_experimentation.ipynb

INSTRUCTIONS.txt

01_data_pipeline > INSTRUCTIONS.txt

```
77     |   └── data/leadscoreing.csv
78     |   └── mapping
79     |       └── mapping/city_tier_mapping.py
80     |       └── mapping/interaction_mapping.csv
81     |           └── mapping/significant_categorical_level.py
82     |   └── schema.py
83     |   └── data_validation_checks.py
84     |   └── lead_scoring_data_pipeline.py
85     |   └── constants.py
86     |   └── utils.py
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
```

Therefore, go back to the scripts folder and copy all the files and folders which are shown in the above ' ~/airflow/dags/lead_scoring_data_pipeline' folder.

15. After copying all the necessary files and folder from 'scripts' folder to 'Lead_scoring_data_pipeline'

Notice that we did not copy paste "utils_output.db", this is because we plan to create a new db named

16. With this you should be able to run the data pipeline. Follow the instructions given below and ta

A. Go to airflow.cfg file and make the following changes:

1. Change the base_url to '<http://localhost:6007>'.
2. Change the web server port to 6007.

B. Now, run the following commands:

1. Run the airflow db init command to initialise the database.
2. Run the following command to create a user to access the database:

```
airflow users create --username <argument> --firstname <argument> --lastname <argument> --
```

Note: All the <argument> needs to replaced by your input. For eg: airflow users create --u

3. Start the Airflow server: airflow webserver

4. Go to a new terminal and run the command to start the scheduler: airflow scheduler

C. Now go to Jarvis dashboard and select the API Endpoint 1. Enter the credentials of the user th

PROBLEMS 8 OUTPUT DEBUG CONSOLE TERMINAL PORTS JUPYTER COMMENTS Filter (e.g. text, !exclude, \e...)

X

^

> OUTLINE

✓ 02_training_pipeline

✓ notebooks

✓ Data

cleaned_data.csv

mlruns

lead_scoring_model_experimentation.db

lead_scoring_model_experimentation.ipynb

logs.log

- ✓ airflow/dags
 - ✓ Lead_scoring_data_pipeline
 - ✓ data
 - leadscoring.csv
 - loaded_data_output.csv
 - ✓ mapping
 - city_tier_mapping.py
 - interaction_mapping.csv
 - significant_categorical_level.py
 - constants.py
 - data_validation_checks.py
 - lead_scoring_data_pipeline.py
 - schema.py
 - utils.py

✓ Lead_scoring_training_pipeline



constants.py



lead_scoring_training_pipeline.py



utils.py



__init__.py

- ✓ scripts
 - > .ipynb_checkpoints
 - > .pytest_cache
 - > 01_data_pipeline
 - > data
- ✓ unit_test [11]
 - > .ipynb_checkpoints
 - > 01_data_pipeline
 - > mapping
 - ⌚ constants.py
 - ⌚ data_validation_checks.py
 - ⌚ dummy.ipynb
 - ⌚ lead_scoring_data_pipeline.py
 - 📄 leadscoreing_test.csv
 - ⌚ schema.py
 - ⌚ test_with_pytest.py
 - ≡ unit_test_cases.db [16]
 - ≡ utils_output.db
 - ⌚ utils.py
 - ⌚ city_tier_mapping.py
 - ⌚ constants.py
 - ⌚ data_validation_checks.py
 - ⌚ dummy.ipynb

Assignment / 02_training_pipeline / notebooks

```

# Define the list of features (note: 'city_tier' is included only once)
significant_features = [
    'total_leads_dropped', 'city_tier', 'referred_lead', 'app_complete_flag',
    'first_platform_c', 'first_utm_medium_c', 'first_utm_source_c'
]

# Create a subset of the dataset with unique columns
dataset_subset = dataset[significant_features] # Remove duplicate 'city_tier'

# Setup PyCaret
clf = setup(data=dataset_subset, target='first_platform_c', normalize=False, transformation=False, session_id=123)

# Optionally, compare models
best_model = compare_models(exclude=['gbc', 'knn', 'qda', 'dummy', 'svm', 'ada'])
print("Best Model:", best_model)

```

[27] ✓ 6m 4.6s

	Description	Value
0	Session id	123
1	Target	first_platform_c
2	Target type	Multiclass
3	Target mapping	Level0: 0, Level1: 1, Level2: 2, Level3: 3, Level7: 4, Level8: 5, others: 6
4	Original data shape	(238964, 7)
5	Transformed data shape	(238964, 33)
6	Transformed train set shape	(167274, 33)
7	Transformed test set shape	(71690, 33)
8	Numeric features	4
9	Categorical features	2
10	Preprocess	This
11	Imputation type	simple
12	Numeric imputation	mean
13	Categorical imputation	mode
14	Maximum one-hot encoding	25

PROBLEMS 19 OUTPUT DEBUG CONSOLE TERMINAL PORTS JUPYTER COMMENTS

```

/Users/anandjha/Documents/GitHub/Assignment 2/venv/lib/python3.9/site-packages/urllib3/_init__.py:35: NotOpenSSLWarning: urllib3
ently the 'ssl' module is compiled with 'LibreSSL 2.8.3'. See: https://github.com/urllib3/urllib3/issues/3020
    warnings.warn(
/Users/anandjha/Documents/GitHub/Assignment 2/venv/lib/python3.9/site-packages/urllib3/_init__.py:35: NotOpenSSLWarning: urllib3
ently the 'ssl' module is compiled with 'LibreSSL 2.8.3'. See: https://github.com/urllib3/urllib3/issues/3020
    warnings.warn(
[2025-04-02 17:31:48 +0530] [21430] [INFO] Booting worker with pid: 21430
/Users/anandjha/Documents/GitHub/Assignment 2/venv/lib/python3.9/site-packages/urllib3/_init__.py:35: NotOpenSSLWarning: urllib3
ently the 'ssl' module is compiled with 'LibreSSL 2.8.3'. See: https://github.com/urllib3/urllib3/issues/3020

```

> OUTLINE

02_training_pipeline

- notebooks
 - mlflow.db
- scripts
 - constants.py
 - lead_scoring_training_pipeline.py
 - test_training_pipeline.ipynb
 - utils.py
- INSTRUCTIONS_training.txt

03_inference_pipeline

- data
 - leadscore_inference.csv
- scripts
- INSTRUCTIONS_inference.txt

airflow

mirluns

- .trash

Assignment/02_training_pipeline/notebooks

- .ipynb_checkpoints
- 01_data_pipeline/scripts/unit_test/.ipynb_ch...

Assignment/02_training_pipeline/notebooks

- lead_scoring_model_experimentation.db
- catboost_info
- Data
 - cleaned_data.csv
- mirluns
- .trash

309887832203812273

- f489b39ea35a4b01a15444c589755a98
- ! meta.yaml
- best_model.pkl
- lead_scoring_model_experimentation.ipynb
- logs.log
- test_training_pipeline.ipynb

mirluns

models

- mlflow.db

venv

tables

OUTLINE

Generate + Code + Markdown | ⚡ Interrupt ⚡ Restart ⚡ Clear All Outputs ⚡ Go To | 📊 Jupyter Variables ⚡ Outline ...

```
# Define the list of features (note: 'city_tier' is included only once)
significant_features = [
    'total_leads_dropped', 'city_tier', 'referred_lead', 'app_complete_flag',
    'first_platform_c', 'first_utm_medium_c', 'first_utm_source_c'
]

# Create a subset of the dataset with unique columns
dataset_subset = dataset[significant_features] # Remove duplicate 'city_tier'

# Setup PyCaret
clf = setup(data=dataset_subset, target='first_platform_c', normalize=False, transformation=False, session_id=123)

# Optionally, compare models
best_model = compare_models(exclude=['gbc', 'knn', 'qda', 'dummy', 'svm', 'ada'])
print("Best Model:", best_model)
```

[27] ✓ 6m 4.6s

...

	Description	Value
0	Session id	123
1	Target	first_platform_c
2	Target type	Multiclass
3	Target mapping	Level0: 0, Level1: 1, Level2: 2, Level3: 3, Level7: 4, Level8: 5, others: 6
4	Original data shape	(238964, 7)
5	Transformed data shape	(238964, 33)
6	Transformed train set shape	(167274, 33)
7	Transformed test set shape	(71690, 33)
8	Numeric features	4
9	Categorical features	2
10	Preprocess	This
11	Imputation type	simple
12	Numeric imputation	mean
13	Categorical imputation	mode
14	Maximum one-hot encoding	25

PROBLEMS 19 OUTPUT DEBUG CONSOLE TERMINAL PORTS JUPYTER COMMENTS

```
/Users/anandjha/Documents/GitHub/Assignment 2/venv/lib/python3.9/site-packages/urllib3/_init__.py:35: NotOpenSSLWarning: urllib3
ently the 'ssl' module is compiled with 'LibreSSL 2.8.3'. See: https://github.com/urllib3/urllib3/issues/3020
  warnings.warn(
/Users/anandjha/Documents/GitHub/Assignment 2/venv/lib/python3.9/site-packages/urllib3/_init__.py:35: NotOpenSSLWarning: urllib3
ently the 'ssl' module is compiled with 'LibreSSL 2.8.3'. See: https://github.com/urllib3/urllib3/issues/3020
  warnings.warn(
[2025-04-02 17:31:48 +0530] [21430] [INFO] Booting worker with pid: 21430
/Users/anandjha/Documents/GitHub/Assignment 2/venv/lib/python3.9/site-packages/urllib3/_init__.py:35: NotOpenSSLWarning: urllib3
ently the 'ssl' module is compiled with 'LibreSSL 2.8.3'. See: https://github.com/urllib3/urllib3/issues/3020
```

02_training_pipeline

- notebooks
 - logs.log
 - mlflow.db
- scripts
 - constants.py
 - lead_scoring_training_pipeline.py
 - test_training_pipeline.ipynb
 - utils.py
- INSTRUCTIONS_training.txt

03_inference_pipeline

- data
 - leadscore_inference.csv
- scripts
- INSTRUCTIONS_inference.txt

airflow/dags

- Lead_scoring_data_pipeline
 - data
 - mapping
 - constants.py
 - data_validation_checks.py
 - lead_scoring_data_pipeline.py
 - schema.py
 - utils.py
- Lead_scoring_training_pipeline
 - constants.py
 - lead_scoring_training_pipeline.py
 - utils.py
- mlruns
 - .trash
- Assignment/02_training_pipeline/notebooks
 - .ipynb_checkpoints
 - 01_data_pipeline/scripts/unit_test/.ipynb_ch...
- Assignment/02_training_pipeline/notebooks
 - lead_scoring_model_experimentation.db
 - catboost_info
- Data
 - cleaned_data.csv
- mlruns

Generate + Code + Markdown | ⚡ Interrupt ⚡ Restart ⚡ Clear All Outputs ⚡ Go To ⚡ Jupyter Variables ⚡ Outline ...

```
# Define the list of features (note: 'city_tier' is included only once)
significant_features = [
    'total_leads_dropped', 'city_tier', 'referred_lead', 'app_complete_flag',
    'first_platform_c', 'first_utm_medium_c', 'first_utm_source_c'
]

# Create a subset of the dataset with unique columns
dataset_subset = dataset[significant_features] # Remove duplicate 'city_tier'

# Setup PyCaret
clf = setup(data=dataset_subset, target='first_platform_c', normalize=False, transformation=False, session_id=123)

# Optionally, compare models
best_model = compare_models(exclude=['gbc', 'knn', 'qda', 'dummy', 'svm', 'ada'])
print("Best Model:", best_model)
```

[27] ✓ 6m 4.6s

...	Description	Value
0	Session id	123
1	Target	first_platform_c
2	Target type	Multiclass
3	Target mapping	Level0: 0, Level1: 1, Level2: 2, Level3: 3, Level7: 4, Level8: 5, others: 6
4	Original data shape	(238964, 7)
5	Transformed data shape	(238964, 33)
6	Transformed train set shape	(167274, 33)
7	Transformed test set shape	(71690, 33)
8	Numeric features	4
9	Categorical features	2
10	Preprocess	This
11	Imputation type	simple
12	Numeric imputation	mean
13	Categorical imputation	mode
14	Maximum one-hot encoding	25

PROBLEMS 19 OUTPUT DEBUG CONSOLE TERMINAL PORTS JUPYTER COMMENTS

```
/Users/anandjha/Documents/GitHub/Assignment 2/venv/lib/python3.9/site-packages/urllib3/_init__.py:35: NotOpenSSLWarning: urllib3
ently the 'ssl' module is compiled with 'LibreSSL 2.8.3'. See: https://github.com/urllib3/urllib3/issues/3020
  warnings.warn(
/Users/anandjha/Documents/GitHub/Assignment 2/venv/lib/python3.9/site-packages/urllib3/_init__.py:35: NotOpenSSLWarning: urllib3
ently the 'ssl' module is compiled with 'LibreSSL 2.8.3'. See: https://github.com/urllib3/urllib3/issues/3020
  warnings.warn(
[2025-04-02 17:31:48 +0530] [21430] [INFO] Booting worker with pid: 21430
/Users/anandjha/Documents/GitHub/Assignment 2/venv/lib/python3.9/site-packages/urllib3/_init__.py:35: NotOpenSSLWarning: urllib3
ently the 'ssl' module is compiled with 'LibreSSL 2.8.3'. See: https://github.com/urllib3/urllib3/issues/3020
```

ASSIGNMENT 2

oring_model_experimentation.ipynb > 5 : Model Experimentation after dropping features > # You should get lightgbm as the best performing mod

> .pytest_cache
> .venv
> .vscode
✓ 01_data_pipeline
 ✓ notebooks
 > .ipynb_checkpoints
 ✓ Data
 > .ipynb_checkpoints
 cleaned_data.csv
 leadscoreing.csv
 ✓ Maps
 > __pycache__
 > .ipynb_checkpoints
 city_tier.py
 interaction_mapping.csv
✓ profile_report
 > .ipynb_checkpoints
 <> cleaned_data_report.html
 <> raw_data_report.html
 data_cleaning_template.ipynb
✓ scripts
 > .ipynb_checkpoints
 > .pytest_cache
 ✓ 01_data_pipeline / scripts
 unit_test_cases.db
 utils_output.db
 ✓ data
 leadscoreing.csv
 loaded_data_output.csv
 > unit_test
 city_tier_mapping.py
 constants.py
 data_validation_checks.py
 dummby.ipynb
 interaction_mapping.csv
 lead_scoring_data_pipeline.py
 leadscoreing_test.csv
 leadscoreing.csv

Generate + Code + Markdown | ⚡ Interrupt ⚡ Restart ⚡ Clear All Outputs ⚡ Go To ⚡ Jupyter Variables ⚡ Outline ...

```
# Define the list of features (note: 'city_tier' is included only once)
significant_features = [
    'total_leads_dropped', 'city_tier', 'referred_lead', 'app_complete_flag',
    'first_platform_c', 'first_utm_medium_c', 'first_utm_source_c'
]

# Create a subset of the dataset with unique columns
dataset_subset = dataset[significant_features] # Remove duplicate 'city_tier'

# Setup PyCaret
clf = setup(data=dataset_subset, target='first_platform_c', normalize=False, transformation=False, session_id=123)

# Optionally, compare models
best_model = compare_models(exclude=['gbc', 'knn', 'qda', 'dummy', 'svm', 'ada'])
print("Best Model:", best_model)
```

[27] ✓ 6m 4.6s

	Description	Value
0	Session id	123
1	Target	first_platform_c
2	Target type	Multiclass
3	Target mapping	Level0: 0, Level1: 1, Level2: 2, Level3: 3, Level7: 4, Level8: 5, others: 6
4	Original data shape	(238964, 7)
5	Transformed data shape	(238964, 33)
6	Transformed train set shape	(167274, 33)
7	Transformed test set shape	(71690, 33)
8	Numeric features	4
9	Categorical features	2
10	Preprocess	True
11	Imputation type	simple
12	Numeric imputation	mean
13	Categorical imputation	mode
14	Maximum one-hot encoding	25

PROBLEMS 19 OUTPUT DEBUG CONSOLE TERMINAL PORTS JUPYTER COMMENTS

```
/Users/anandjha/Documents/GitHub/Assignment 2/venv/lib/python3.9/site-packages/urllib3/_init__.py:35: NotOpenSSLWarning: urllib3
ently the 'ssl' module is compiled with 'LibreSSL 2.8.3'. See: https://github.com/urllib3/urllib3/issues/3020
    warnings.warn(
/Users/anandjha/Documents/GitHub/Assignment 2/venv/lib/python3.9/site-packages/urllib3/_init__.py:35: NotOpenSSLWarning: urllib3
ently the 'ssl' module is compiled with 'LibreSSL 2.8.3'. See: https://github.com/urllib3/urllib3/issues/3020
    warnings.warn(
[2025-04-02 17:31:48 +0530] [21430] [INFO] Booting worker with pid: 21430
/Users/anandjha/Documents/GitHub/Assignment 2/venv/lib/python3.9/site-packages/urllib3/_init__.py:35: NotOpenSSLWarning: urllib3
```

Experiments



lead_scoring

Provide Feedback

Add Description

Search experiments

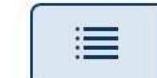
 lead_scoring my_experiment Assignment

Runs

Evaluation

Experimental

Traces



metrics.rmse < 1 and params.model = "tree"



Time created

State: Active



+ New



Sort: Created



Run Name

Created

Dataset

Duration

Source

Models



bald-grub-768

4 minutes ago

-

24ms

ipykern...

-



File View Settings Help

<input type="checkbox"/> Files	<input checked="" type="checkbox"/> Running	
<input type="button" value="Open"/> <input type="button" value="Download"/> <input type="button" value="Rename"/> <input type="button" value="Duplicate"/> <input type="button" value="Delete"/>		<input type="button" value="New"/> <input type="button" value="Upload"/> <input type="button" value="C"/>
<input type="checkbox"/> /		
Name	Last Modified	File Size
<input type="checkbox"/> Data	6 days ago	
<input type="checkbox"/> mlrns	2 days ago	
<input checked="" type="checkbox"/> lead_scoring_model_experimentation.ipynb	yesterday	78.2 KB
<input type="checkbox"/> lead_scoring_model_experimentation.db	5 days ago	220 KB
<input type="checkbox"/> logs.log	2 days ago	6.3 KB



Experiments Models Prompts

GitHub D

Experiments



Default ⓘ Provide Feedback ⚡ Add Description



Search experiments

 Default

Runs Evaluation Experimental Traces



Q metrics.rmse < 1 and params.model = "tree" ⓘ



Time created ▾

State: Active ▾



+ New

Datasets ▾

Sort: Created ▾

Columns ▾

Group by ▾



Run Name

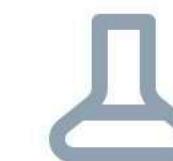
Created ⏪

Dataset

Duration

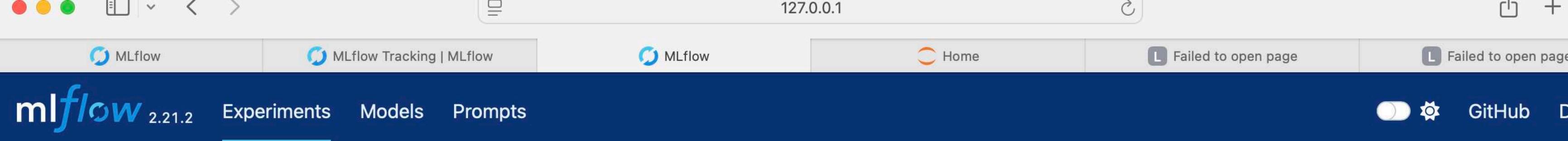
Source

Models



No runs logged

No runs have been logged yet. [Learn more](#) about how to create ML model training runs in this experiment.



Experiments



lead_scoring

Provide Feedback

Add Description

Runs

Evaluation

Experimental

Traces



metrics.rmse < 1 and params.model = "tree"



Time created

State: Active



+ New

Datasets

Sort: Created

Columns

Group by



Run Name

Created



Dataset

Duration

Source

Models



bald-grub-768



23 minutes ago

-

24ms

ipykern...

-

lead_scoring >

bald-grub-768

[Overview](#) [Model metrics](#) [System metrics](#) [Traces](#) [Artifacts](#)

Description

No description

Details

Created at	04/02/2025, 08:06:10 PM
Created by	anandjha
Experiment ID	706737907547362779 
Status	 Finished
Run ID	eb548ec10168428caa788d46bd5cca0b 
Duration	24ms
Datasets used	—
Tags	model: test_model 
Source	 ipykernel_launcher.py
Logged models	—
Registered models	—
Registered prompts	—

Parameters (1)

 Search parameters

Parameter	Value
Screenshot	

Metrics (1)

 Search metrics

Metric	Value
Screenshot	

lead_scoring >

bald-grub-768

[Overview](#) [Model metrics](#) [System metrics](#) [Traces](#) [Artifacts](#)

Description

No description

Details

Created at	04/02/2025, 08:06:10 PM
Created by	anandjha
Experiment ID	706737907547362779 
Status	 Finished
Run ID	eb548ec10168428caa788d46bd5cca0b 
Duration	24ms
Datasets used	—
Tags	model: test_model 
Source	 ipykernel_launcher.py
Logged models	—
Registered models	—
Registered prompts	—

Parameters (1)

 Search parameters

Parameter	Value
learning_rate	0.01

Screenshot

Metrics (1)

 Search metrics

Metric	Value
accuracy	0.95

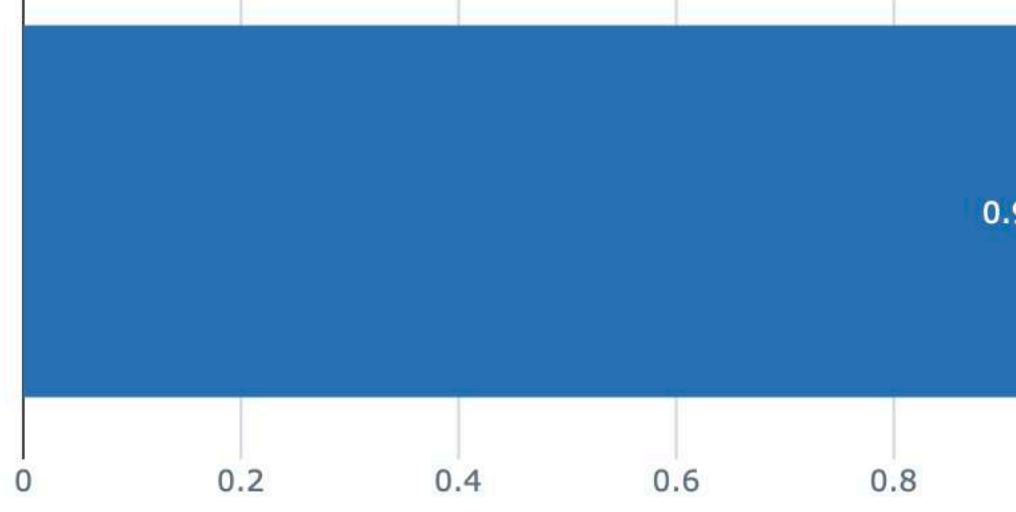
lead_scoring >

bald-grub-768

[Overview](#) [Model metrics](#) [System metrics](#) [Traces](#) [Artifacts](#) Search metric charts Auto-refresh

Model metrics (1)

accuracy

[bald-grub-768](#)[+ Add section](#)

lead_scoring_model_experimentation.ipynb

! meta.yaml



mlruns > Assignment > 02_training_pipeline > notebooks > lead_scoring_model_experimentation.ipynb > m+ 3 Setting up Environment: > # Set the MLflow tracking URI

Generate + Code + Markdown | Run All Restart Clear All Outputs Jupyter Variables Outline ...

Python

9	Rows with missing values	100.0%
10	Preprocess	Fit
11	Imputation type	simple
12	Numeric imputation	mean
13	Categorical imputation	mode
14	Maximum one-hot encoding	25
15	Encoding method	None
16	Fold Generator	StratifiedKFold
17	Fold Number	10
18	CPU Jobs	-1
19	Use GPU	False
20	Log Experiment	False
21	Experiment Name	clf-default-name
22	USI	59c0

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
dt	Decision Tree Classifier	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.6210
rf	Random Forest Classifier	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.2630
catboost	CatBoost Classifier	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	7.3730
lr	Logistic Regression	0.9994	1.0000	0.9995	0.9993	0.9994	0.9987	0.9987	2.9790
nb	Naive Bayes	0.9950	0.9994	0.9986	0.9915	0.9950	0.9900	0.9900	0.6090
xgboost	Extreme Gradient Boosting	0.9940	1.0000	0.9910	0.9969	0.9939	0.9879	0.9880	0.8580
lightgbm	Light Gradient Boosting Machine	0.9933	0.9988	0.9996	0.9872	0.9933	0.9866	0.9867	0.6430
et	Extra Trees Classifier	0.9856	0.9989	0.9882	0.9831	0.9856	0.9712	0.9712	1.2270
lda	Linear Discriminant Analysis	0.9760	0.9752	0.9844	0.9683	0.9762	0.9521	0.9522	0.6550
ridge	Ridge Classifier	0.9759	0.9752	0.9841	0.9682	0.9761	0.9518	0.9519	0.4550

... Best Model: DecisionTreeClassifier(ccp_alpha=0.0, class_weight=None, criterion='gini',
max_depth=None, max_features=None, max_leaf_nodes=None,
min_impurity_decrease=0.0, min_samples_leaf=1,
min_samples_split=2, min_weight_fraction_leaf=0.0,
monotonic_cst=None, random_state=123, splitter='best')

PROBLEMS 22 OUTPUT DEBUG CONSOLE TERMINAL PORTS JUPYTER COMMENTS

```
[2025-04-02 19:24:08 +0530] [24034] [ERROR] Connection in use: ('127.0.0.1', 5000)
[2025-04-02 19:24:08 +0530] [24034] [ERROR] connection to ('127.0.0.1', 5000) failed: [Errno 48] Address already in use
[2025-04-02 19:24:09 +0530] [24034] [ERROR] Connection in use: ('127.0.0.1', 5000)
[2025-04-02 19:24:09 +0530] [24034] [ERROR] connection to ('127.0.0.1', 5000) failed: [Errno 48] Address already in use
[2025-04-02 19:24:10 +0530] [24034] [ERROR] Connection in use: ('127.0.0.1', 5000)
[2025-04-02 19:24:10 +0530] [24034] [ERROR] connection to ('127.0.0.1', 5000) failed: [Errno 48] Address already in use
[2025-04-02 19:24:11 +0530] [24034] [ERROR] Connection in use: ('127.0.0.1', 5000)
[2025-04-02 19:24:11 +0530] [24034] [ERROR] connection to ('127.0.0.1', 5000) failed: [Errno 48] Address already in use
[2025-04-02 19:24:12 +0530] [24034] [ERROR] Can't connect to ('127.0.0.1', 5000)
Running the mlflow server failed. Please see the logs above for details.
```

Pyt

Pyt

zsh

lead_scoring_model_experimentation.ipynb

! meta.yaml



mlruns > Assignment > 02_training_pipeline > notebooks > lead_scoring_model_experimentation.ipynb > m+ 3 Setting up Environment: > # Set the MLflow tracking URI

Generate + Code + Markdown | Run All Restart Clear All Outputs Jupyter Variables Outline ...

Python

3	Original data shape	(240000, 46)
4	Transformed data shape	(240000, 38)
5	Transformed train set shape	(216000, 38)
6	Transformed test set shape	(24000, 38)
7	Numeric features	40
8	Categorical features	5
9	Rows with missing values	100.0%
10	Preprocess	True
11	Imputation type	simple
12	Numeric imputation	mean
13	Categorical imputation	mode
14	Maximum one-hot encoding	25
15	Encoding method	None
16	Fold Generator	StratifiedKFold
17	Fold Number	10
18	CPU Jobs	-1
19	Use GPU	False
20	Log Experiment	False
21	Experiment Name	clf-default-name
22	USI	3e02

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
knn	K Neighbors Classifier	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	7.8430
dt	Decision Tree Classifier	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.8140
rf	Random Forest Classifier	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.3260
ada	Ada Boost Classifier	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.4880
gbc	Gradient Boosting Classifier	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	2.4600
catboost	CatBoost Classifier	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	7.6610
lr	Logistic Regression	0.9994	1.0000	0.9995	0.9993	0.9994	0.9987	0.9987	3.2400
nb	Naive Bayes	0.9950	0.9994	0.9986	0.9915	0.9950	0.9900	0.9900	0.4660
xgboost	Extreme Gradient Boosting	0.9940	1.0000	0.9910	0.9969	0.9939	0.9879	0.9880	0.6770

PROBLEMS 22 OUTPUT DEBUG CONSOLE TERMINAL PORTS JUPYTER COMMENTS

```
[2025-04-02 19:24:08 +0530] [24034] [ERROR] Connection in use: ('127.0.0.1', 5000)
[2025-04-02 19:24:08 +0530] [24034] [ERROR] connection to ('127.0.0.1', 5000) failed: [Errno 48] Address already in use
[2025-04-02 19:24:09 +0530] [24034] [ERROR] Connection in use: ('127.0.0.1', 5000)
[2025-04-02 19:24:09 +0530] [24034] [ERROR] connection to ('127.0.0.1', 5000) failed: [Errno 48] Address already in use
[2025-04-02 19:24:10 +0530] [24034] [ERROR] Connection in use: ('127.0.0.1', 5000)
[2025-04-02 19:24:10 +0530] [24034] [ERROR] connection to ('127.0.0.1', 5000) failed: [Errno 48] Address already in use
[2025-04-02 19:24:11 +0530] [24034] [ERROR] Connection in use: ('127.0.0.1', 5000)
[2025-04-02 19:24:11 +0530] [24034] [ERROR] connection to ('127.0.0.1', 5000) failed: [Errno 48] Address already in use
[2025-04-02 19:24:12 +0530] [24034] [ERROR] Can't connect to ('127.0.0.1', 5000)
Running the mlflow server failed. Please see the logs above for details.
```

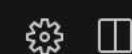
Pyt

Pyt

zsh

lead_scoring_model_experimentation.ipynb

! meta.yaml

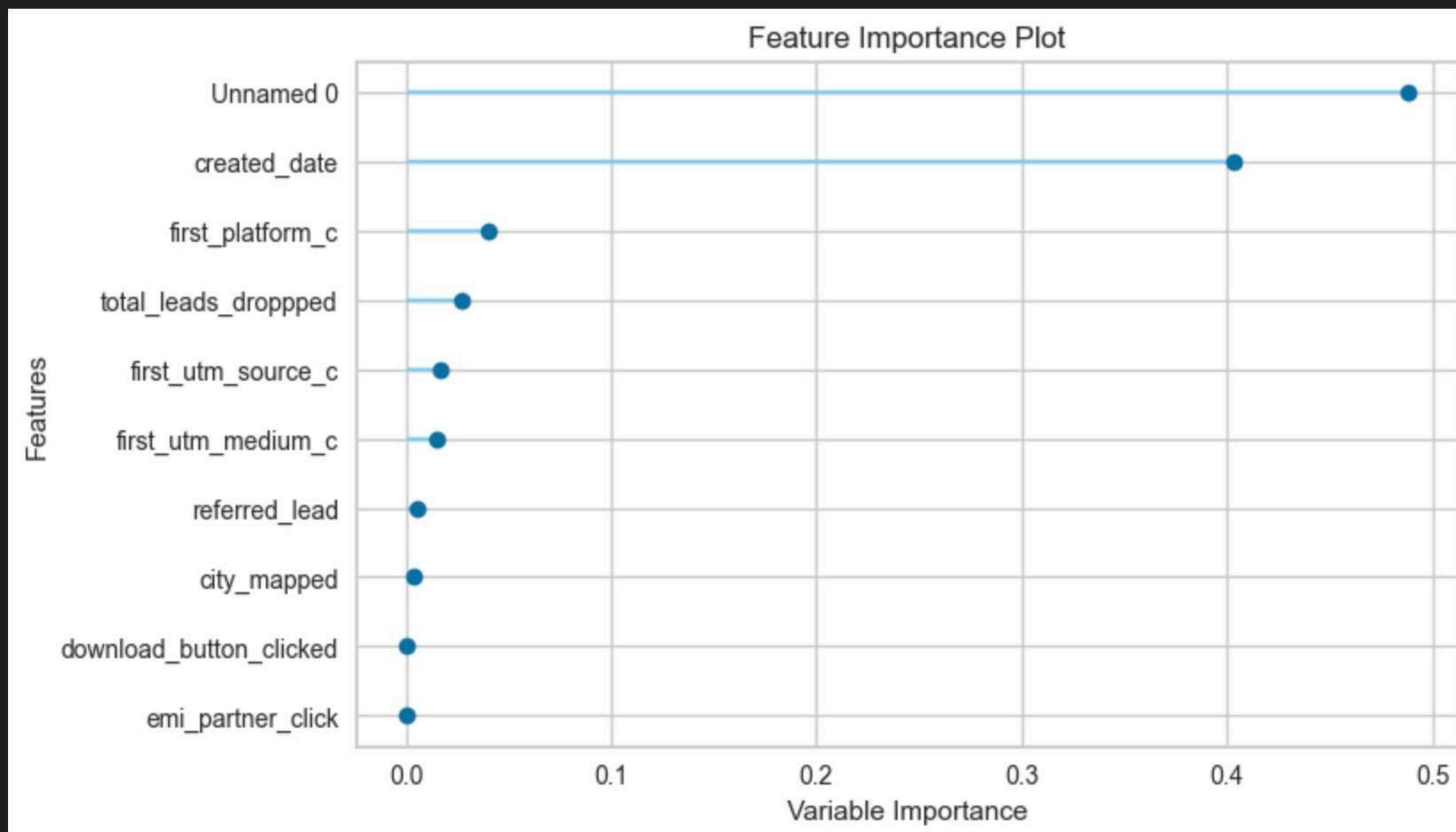


mlruns > Assignment > 02_training_pipeline > notebooks > lead_scoring_model_experimentation.ipynb > 3 Setting up Environment: # Set the MLflow tracking URI

Generate + Code + Markdown | Run All Restart Clear All Outputs Jupyter Variables Outline ...

Python 3

4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
7	1.0000	1.0000	0.9999	1.0000	1.0000	0.9999	0.9999	1.0000
8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Mean	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Std	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000



PROBLEMS 22

OUTPUT

DEBUG CONSOLE

TERMINAL

PORTS

JUPYTER

COMMENTS

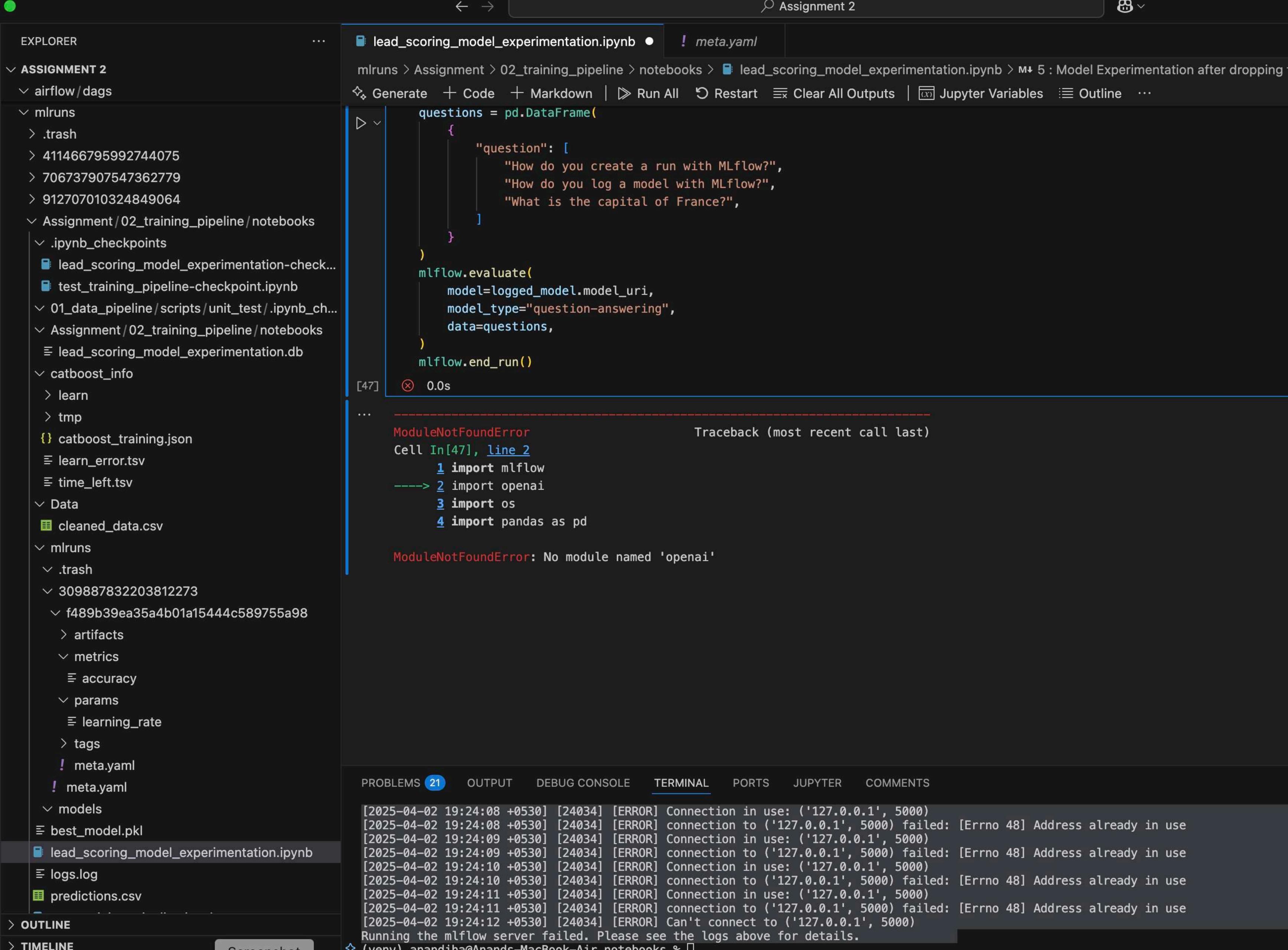
```
[2025-04-02 19:24:08 +0530] [24034] [ERROR] Connection in use: ('127.0.0.1', 5000)
[2025-04-02 19:24:08 +0530] [24034] [ERROR] connection to ('127.0.0.1', 5000) failed: [Errno 48] Address already in use
[2025-04-02 19:24:09 +0530] [24034] [ERROR] Connection in use: ('127.0.0.1', 5000)
[2025-04-02 19:24:09 +0530] [24034] [ERROR] connection to ('127.0.0.1', 5000) failed: [Errno 48] Address already in use
[2025-04-02 19:24:10 +0530] [24034] [ERROR] Connection in use: ('127.0.0.1', 5000)
[2025-04-02 19:24:10 +0530] [24034] [ERROR] connection to ('127.0.0.1', 5000) failed: [Errno 48] Address already in use
[2025-04-02 19:24:11 +0530] [24034] [ERROR] Connection in use: ('127.0.0.1', 5000)
[2025-04-02 19:24:11 +0530] [24034] [ERROR] connection to ('127.0.0.1', 5000) failed: [Errno 48] Address already in use
[2025-04-02 19:24:12 +0530] [24034] [ERROR] Can't connect to ('127.0.0.1', 5000)
Running the mlflow server failed. Please see the logs above for details.
```

+ ↻ ⌂ ⌄ ⌁

Python

Python

zsh not





Experiments Models Prompts

GitHub

Experiments



my_experiment

Provide Feedback

Add Description

Runs

Evaluation

Experimental

Traces



metrics.rmse < 1 and params.model = "tree"



Time created

State: Active



Datasets

Sort: Created

Columns

Group by

Run Name	Created	Dataset	Duration	Source	Models
----------	---------	---------	----------	--------	--------



No runs logged

No runs have been logged yet. [Learn more about how to create ML model training runs in this experiment.](#)

Do not use **SQLite** as metadata DB in production – it should only be used for dev/testing. We recommend using Postgres or MySQL. [Click here](#) for more information.

Do not use the **SequentialExecutor** in production. [Click here](#) for more information.

DAGs

All 72	Active 6	Paused 66	Running 0	Failed 0	Filter DAGs by tag	Search DAGs	Auto-refresh	C
DAG ▾		Owner	Runs	Schedule	Last Run	Next Run	Recent Tasks	
<input type="checkbox"/> Params Trigger UI		airflow	○○○○	None ⓘ			○○○○	
example params								
<input type="checkbox"/> Params UI tutorial		airflow	○○○○	None ⓘ			○○○○	
example params ui								
<input type="checkbox"/> Sample DAG with Display Name		airflow	○○○○	None ⓘ			○○○○	
example								
<input type="checkbox"/> conditional_dataset_and_time_based_timetable		airflow	○○○○	Dataset or 0 1 * * 3 ⓘ	2025-03-26, 01:00:00 ⓘ		○○○○	
dataset-time-based-timetable								
<input checked="" type="checkbox"/> consume_1_and_2_with_dataset_expressions		airflow	○○○○	Dataset ⓘ	0 of 2 datasets updated		○○○○	
<input checked="" type="checkbox"/> consume_1_or_2_with_dataset_expressions		airflow	○○○○	Dataset ⓘ	0 of 2 datasets updated		○○○○	
<input checked="" type="checkbox"/> consume_1_or_both_2_and_3_with_dataset_expressions		airflow	○○○○	Dataset ⓘ	0 of 3 datasets updated		○○○○	
<input checked="" type="checkbox"/> dataset_alias_example_alias_consumer		airflow	○○○○	Unresolved DatasetAlias ⓘ			○○○○	
consumer dataset-alias								
<input checked="" type="checkbox"/> dataset_alias_example_alias_consumer_with_no_taskflow		airflow	○○○○	Unresolved DatasetAlias ⓘ			○○○○	
consumer dataset-alias								
<input checked="" type="checkbox"/> dataset alias example alias consumer with no taskflow								

Experiments



Search experiments

 Default lead_scoring_model_expe... 

lead_scoring_model_experimentation

Provide Feedback

Add Description

Runs

Evaluation

Experimental

Traces

metrics.rmse < 1 and params.model = "tree"

Time created

State: Active

Datasets

+ New

Sort: Created



Run Name

Table



Filter by

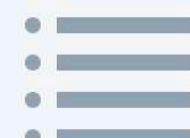
Group by

Compare:



No runs logged

No runs have been logged yet. [Learn more](#) about how to create ML model training runs in this experiment.



No runs selected

Make sure that at least one experiment run is visible and available to compare

Experiments



Search experiments

 Default

Default ⓘ Provide Feedback Add Description

Runs Evaluation Experimental Traces



Q metrics.rmse < 1 and params.model = "tree" ⓘ



Time created ▾

State: Active ▾



+ New

Datasets ▾

Sort: Created ▾

Columns ▾

Group by ▾



Run Name

Created

Dataset

Duration

Source

Models



No runs logged

No runs have been logged yet. [Learn more](#) about how to create ML model training runs in this experiment.

Unpaused DAG Lead_scoring_training_pipeline.

Triggered Lead_scoring_training_pipeline with new Run ID manual_2025-04-03T10:21:53.533097+00:00, it should start any moment now.

DAG: **Lead_scoring_training_pipeline** Training pipeline for Lead Scoring System

Schedule: @monthly i Next Run ID: 2025-03-01, 00:00:00 UTC ▶ 🔗

04/03/2025 10:21:53 AM All Run Types ▼ All Run States ▼ Clear Filters

Press shift + / for Shortcuts deferred failed queued removed restarting running scheduled shutdown skipped success up_for_reschedule up_for_retry upstream_failed no_status

DAG **Lead_scoring_training_pipeline** / Run **2025-04-03, 10:21:53 UTC** Clear Mark state as...

Duration 00:00:08

encode_features 
train_model 

△ Details Graph Gantt Code Event Log Add Note

DAG Run Notes

Dag Run Details

Status	running
Run ID	manual_2025-04-03T10:21:53.533097+00:00 🔗
Run type	► manual
Run duration	00:00:08
Last scheduling decision	2025-04-03, 10:22:01 UTC
Queued at	2025-04-03, 10:21:53 UTC
Started	2025-04-03, 10:21:54 UTC
Data interval start	2025-03-01, 00:00:00 UTC

Screenshot

127.0.0.1

MLflow MLflow Tracking | MLflow MLflow Home MLflow Lead_scoring_training_pipeline

Airflow DAGs Cluster Activity Datasets Security Browse Admin Docs 16:37 UTC AJ

DAG: Lead_scoring_training_pipeline Training pipeline for Lead Scoring System Schedule: @monthly i Next Run ID: 2025-04-01, 00:00:00 UTC ▶ ⌂

04/03/2025 04:37:04 PM All Run Types All Run States Clear Filters Auto-refresh 25 Press shift + / for Shortcuts deferred failed queued removed restarting running scheduled shutdown skipped success up_for_reschedule up_for_retry upstream_failed no_status

DAG Lead_scoring_training_pipeline / Run 2025-04-03, 10:21:53 UTC Clear Mark state as...

Duration 00:00:13

encode_features train_model

DAG Run Notes Add Note

Dag Run Details

Status	success
Run ID	manual_2025-04-03T10:21:53.533097+00:00
Run type	▶ manual
Run duration	00:00:13
Last scheduling decision	2025-04-03, 10:22:08 UTC
Queued at	2025-04-03, 10:21:53 UTC
Started	2025-04-03, 10:21:54 UTC
Ended	2025-04-03, 10:22:08 UTC
Data interval start	2025-03-01, 00:00:00 UTC
Data interval end	2025-04-01, 00:00:00 UTC

Screenshot

127.0.0.1

MLflow MLflow Tracking | MLflow MLflow Home MLflow Lead_scoring_training_pipeline

Airflow DAGs Cluster Activity Datasets Security Browse Admin Docs 16:37 UTC AJ

04/03/2025 04:37:04 PM All Run Types All Run States Clear Filters Auto-refresh 25 Press shift + / for Shortcuts deferred failed queued removed restarting running scheduled shutdown skipped success up_for_reschedule up_for_retry upstream_failed no_status

DAG Lead_scoring_training_pipeline / encode_features Task Details Graph Gantt Code Event Log Filter DAG by task

Duration 00:00:13 00:00:06 00:00:00

Task Instance Duration Median total duration Median queued duration

Duration (seconds) 6 5 4 3 2 1 0

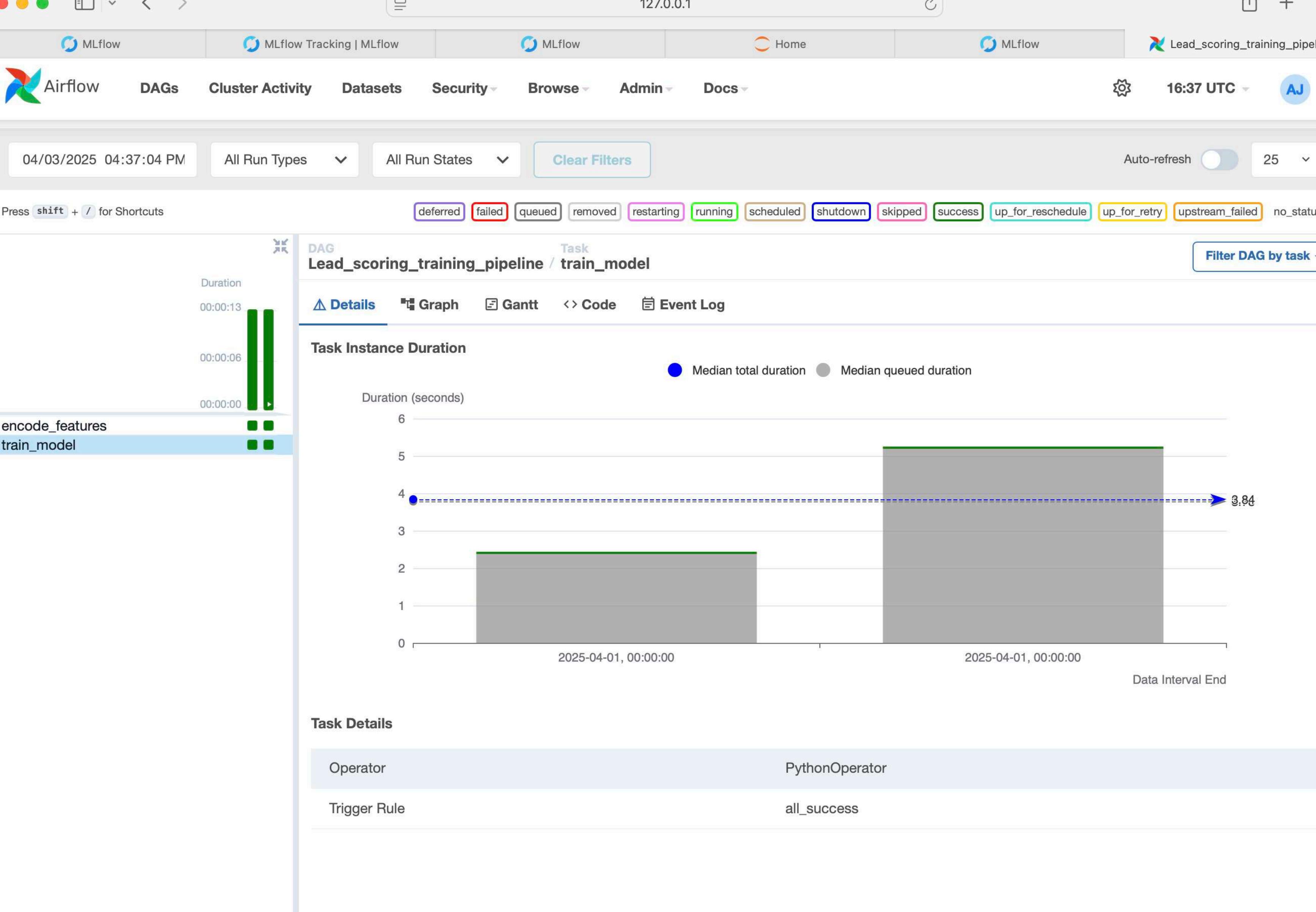
2025-04-01, 00:00:00 2025-04-01, 00:00:00 Data Interval End

Task Details

Operator	PythonOperator
Trigger Rule	all_success

encode_features train_model

Version: v2.10.5 Screenshot Git Version: .release:b93c3 0bd15ac7d05bc58ff2cccbf



antId= ← → C 127.0.0.1:5001/home ⋮

Airflow DAGs Cluster Activity Datasets Security Browse Admin Docs ⋮ AJ

Do not use **SQLite** as metadata DB in production – it should only be used for dev/testing. We recommend using Postgres or MySQL. [Click here](#) for more information.

Do not use the **SequentialExecutor** in production. [Click here](#) for more information.

DAGs

All 76 Active 2 Paused 74 Running 2 Failed 0 Filter DAGs by tag Search DAGs

Auto-refresh

i DAG	Owner	Runs	Schedule	Last Run	Next Run
<input checked="" type="checkbox"/> Lead_Scoring_Data_Engineering_Pipeline	airflow	2	@daily	2025-04-04, 09:34:53	2025-04-04, 00:00
<input checked="" type="checkbox"/> Lead_scoring_training_pipeline	airflow	1	@monthly	2025-04-04, 09:35:08	2025-04-01, 00:00
<input checked="" type="checkbox"/> Params Trigger UI example params	airflow	None	None		
<input checked="" type="checkbox"/> Params UI tutorial example params ui	airflow	None	None		
<input checked="" type="checkbox"/> Sample DAG with Display Name example	airflow	None	None		
<input checked="" type="checkbox"/> conditional_dataset_and_time_based_timetable dataset-time-based-timetable	airflow	Dataset or 0 1 ** 3	Dataset or 0 1 ** 3		2025-04-02, 01:00
<input checked="" type="checkbox"/> consume_1_and_2_with_dataset_expressions	airflow	Dataset	Dataset		0 of 2 datasets updated
<input checked="" type="checkbox"/> consume_1_or_2_with_dataset_expressions	airflow	Dataset	Dataset		0 of 2 datasets updated
<input checked="" type="checkbox"/> consume_1_or_both_2_and_3_with_dataset_expressions	airflow	Dataset	Dataset		0 of 3 datasets updated
<input checked="" type="checkbox"/> dataset_usage_example_dags_consumer	airflow	Unresolved DatasetAlias	Unresolved DatasetAlias		

Screenshot

The scheduler does not appear to be running. Last heartbeat was received 17 minutes ago.
The DAGs list may not update, and new tasks will not be scheduled.

DAG: Lead_scoring_training_pipeline

Training pipeline for Lead Scoring System

Schedule: @monthly | Next Run ID: 2025-04-01, 00:00:00 UTC

04/04/2025 10:06:21 AM All Run Types All Run States Clear Filters Auto-refresh 25

Press shift + / for Shortcuts deferred failed queued removed restarting running scheduled shutdown skipped success up_for_reschedule up_for_retry upstream_failed no_status

Duration: 00:00:13

encode_features

train_model

DAG: Lead_scoring_training_pipeline

Details Graph Gantt Code Event Log Run Duration Task Duration Calendar

DAG Runs Summary

Total Runs Displayed	3
Total success	3

First Run Start: 2025-04-03, 10:21:54 UTC

Last Run Start: 2025-04-04, 09:35:09 UTC

Max Run Duration: 00:00:13

Mean Run Duration: 00:00:12

Min Run Duration: 00:00:08

DAG Summary

Total Tasks	2
PythonOperators	2

Screenshot