Project Deliverable 4

Group 6:

Rimjhim Jain
Sudhanshu Dalvi
Saumit Chinchkhandi
Naval Gaikwad
Aniket Ghumed

AWS Cost Analysis for Amazon Best Selling Prediction Project

AWS Services Utilized are:

- **AWS S3 (Simple Storage Service):**

  Purpose: Used for storing datasets.
  Cost Factors: Storage capacity, requests, and data transfer.

- **Amazon Athena:**

  Purpose: Querying transformed data.
  Cost Factors: Based on the amount of data scanned by the queries.

- **AWS Glue:**

  Purpose: ETL (Extract, Transform, Load) jobs for data transformation.
  Cost Factors: Number of Data Processing Units (DPUs) used and job run time.

- **Amazon QuickSight:**

  Purpose: Creating visualizations for data insights.
  Cost Factors: Monthly subscription and additional charges based on user count and SPICE (Super-fast, Parallel, In-memory Calculation Engine) capacity.

- **Amazon SageMaker:**

  Purpose: Model training, evaluation, and hyperparameter tuning.
  Cost Factors: Instance hours, data processing, and storage.

- **AWS Lambda (optional for scaling):**

  Purpose: Automating and scaling ETL jobs.
  Cost Factors: Number of requests and duration of code execution.

Cost and usage Info

Current month costs

$1.92

Forecasted month end costs

There isn't enough historical data to forecast your spend

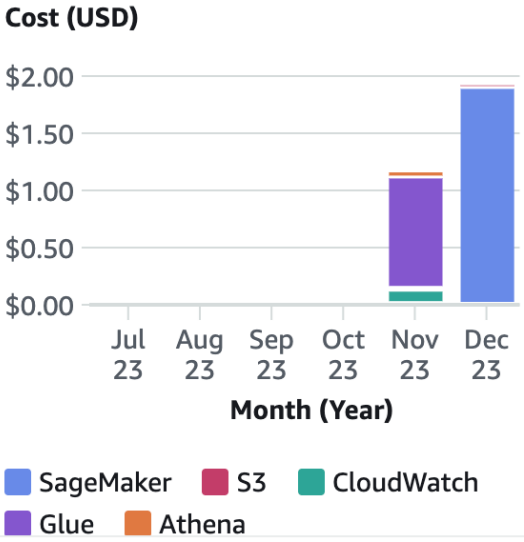
Last month costs

$1.16

Average month costs

$0.51

Total costs per month



Go to Billing and Cost Management

Cost summary Info

Month-to-date cost

$1.92

- compared to last month for same period

Last month's cost for same time period

$0.00

Nov 1 – 10

Total forecasted cost for current month

Access denied

Last month's total cost

$1.16

let's analyze the AWS costs for your Cloud-Based Machine Learning Project:

Month-to-Date Cost Analysis

Month-to-Date Cost: $1.92

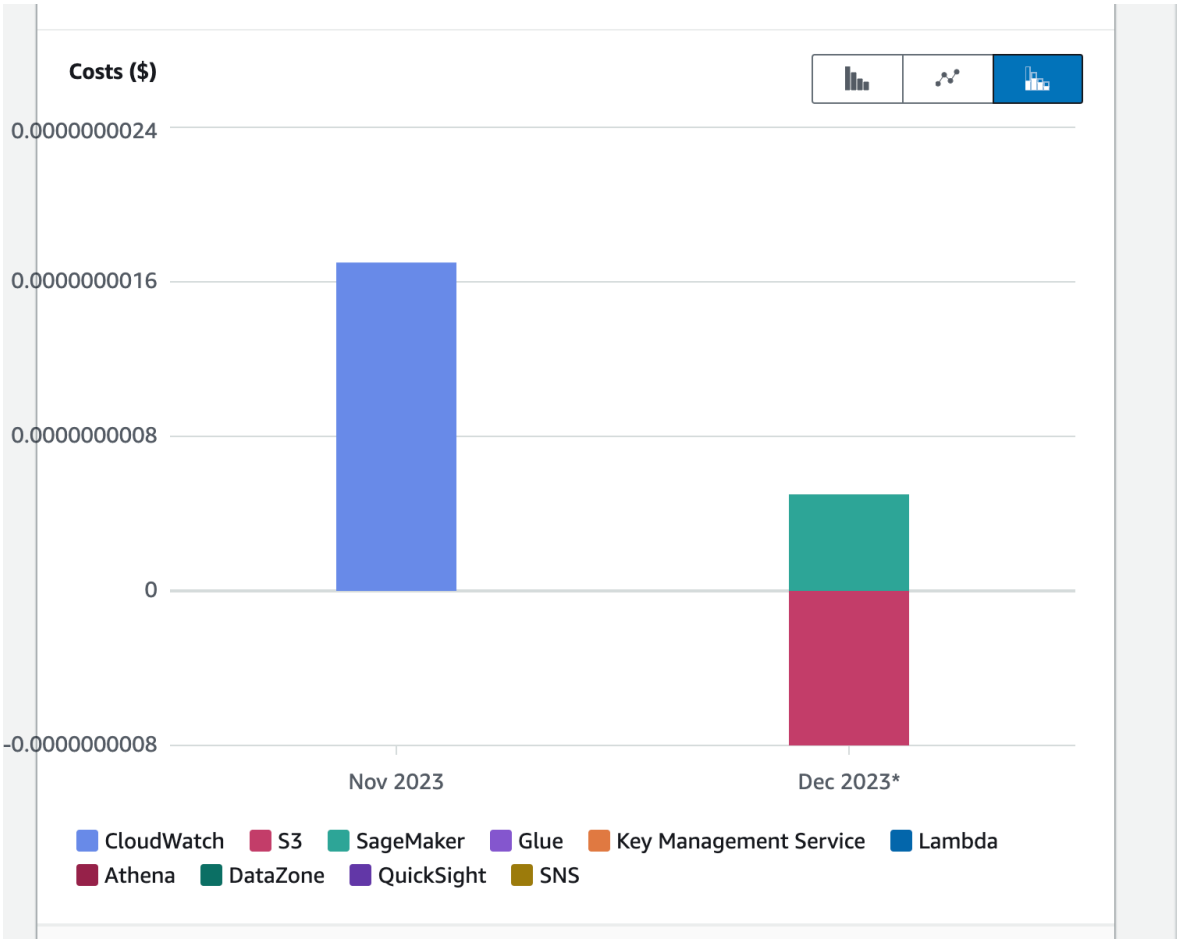
This is the total cost incurred from the start of the current month up to the present day.

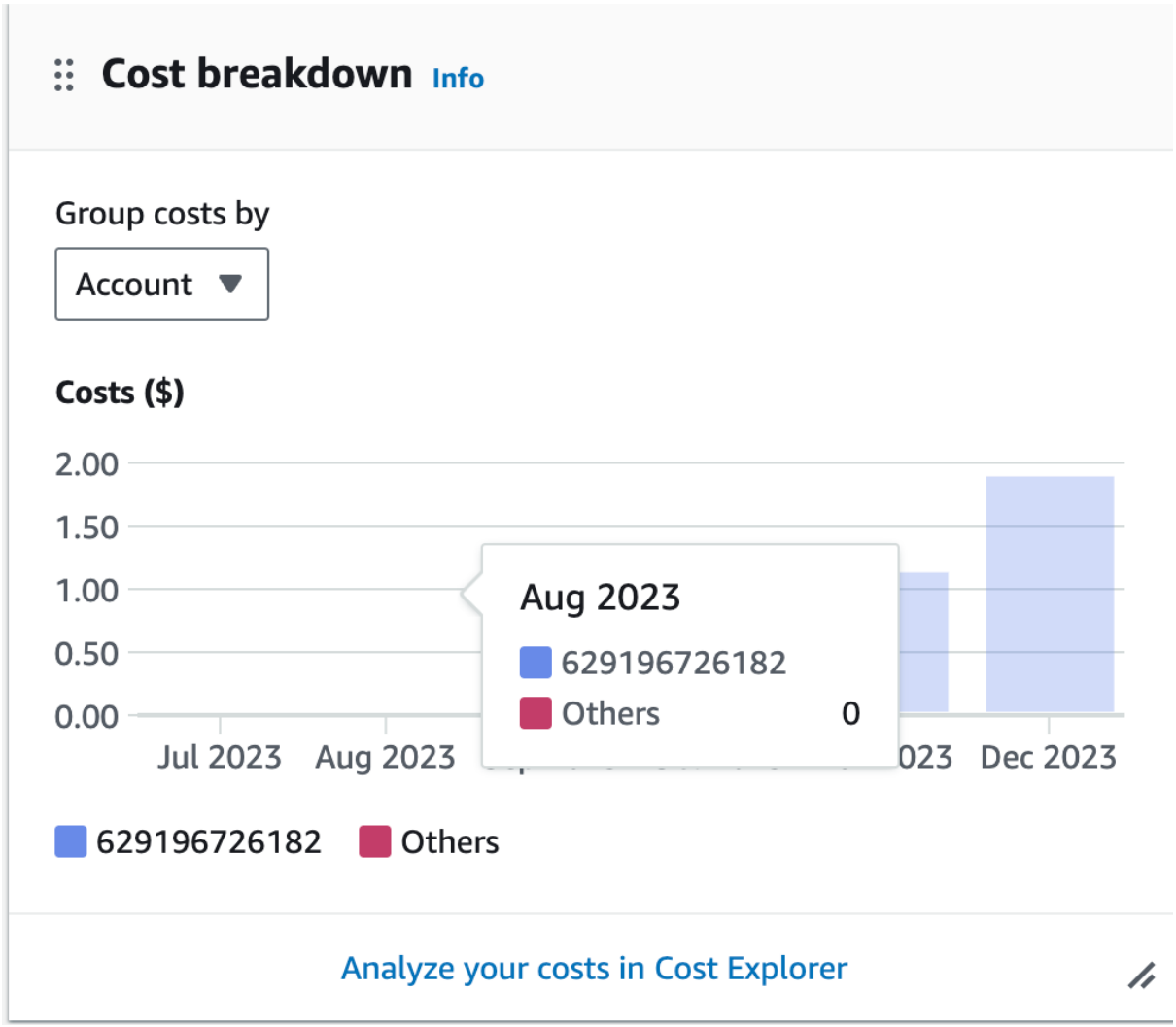
It shows an increase compared to the last month for the same period, where the cost was $0.00. This indicates new usage or a scaling up of services.

Last Month's Total Cost: $1.16

The total cost for the previous month was $1.16, which is less than the current month-to-date cost. This suggests either an expansion in the project's scope or increased usage of AWS services.

Cost Breakdown by Service
Amazon SageMaker, Amazon Simple Storage Service (S3), AWS Glue, etc.
The cost breakdown by service is useful for understanding which AWS services are contributing most to the overall cost.


Costs ($)
0.0000000024
0.0000000016
0.0000000008
0
-0.0000000008
Nov 2023
Dec 2023*
CloudWatch
S3
SageMaker
Glue
Key Management Service
Lambda
Athena
DataZone
QuickSight
SNS



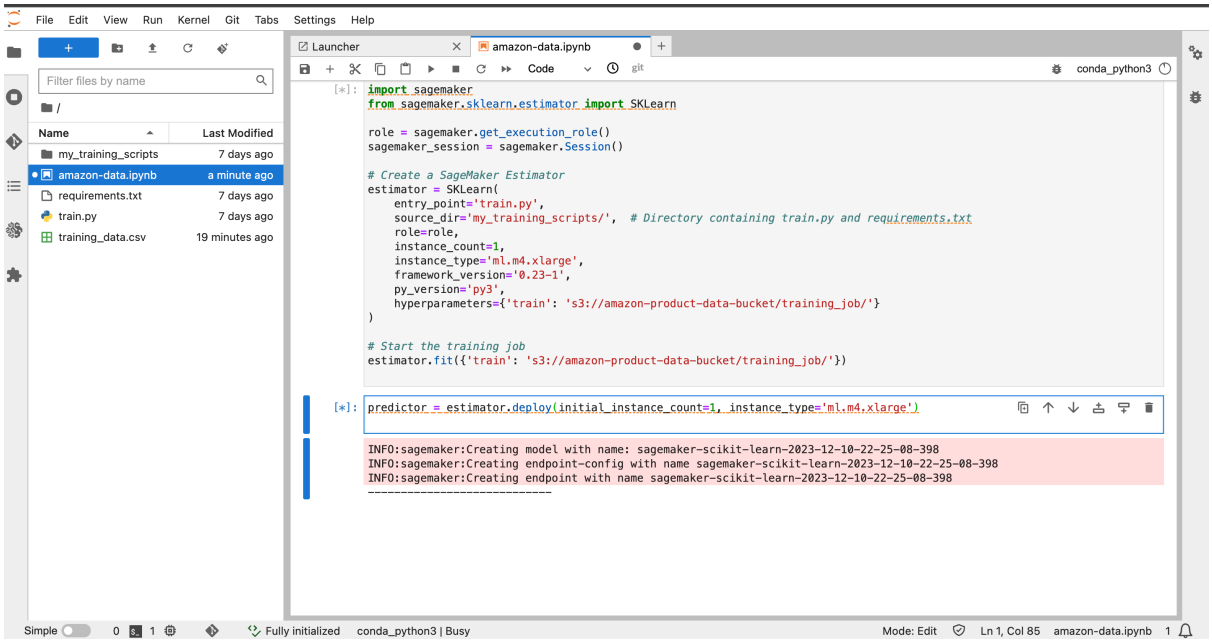
Cost Optimization Strategies:

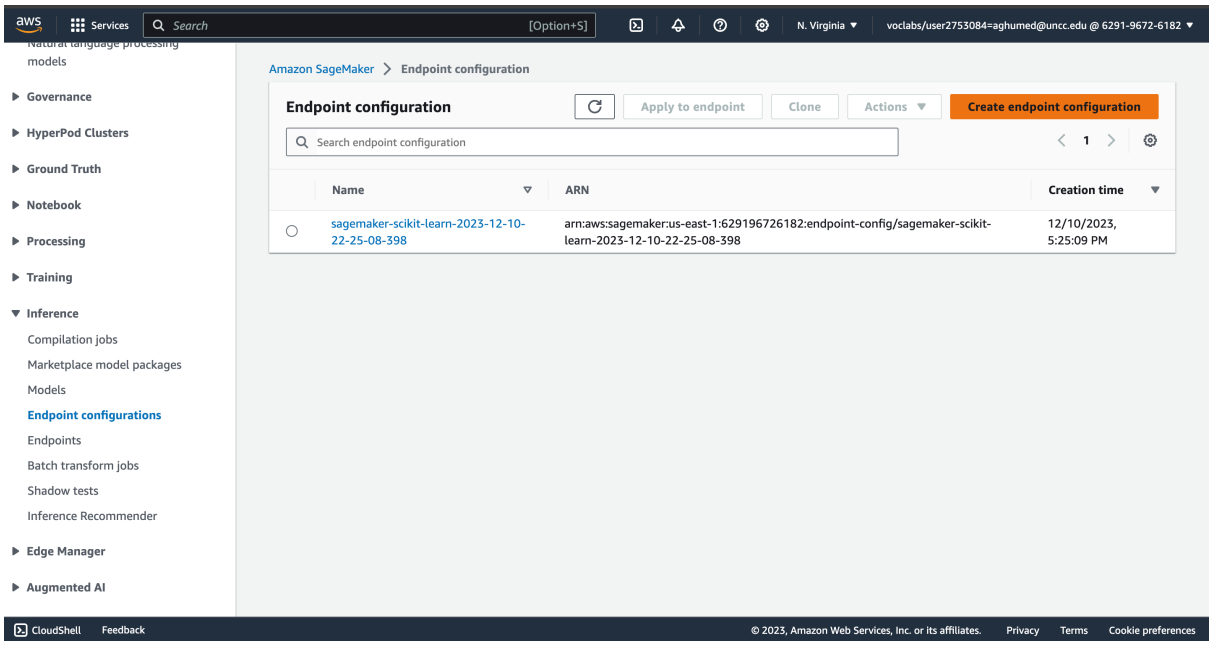
- Regularly monitor and review AWS usage and costs using AWS Cost Explorer.
- Optimize data storage by cleaning up unused S3 objects.
- Manage Athena query costs by optimizing query efficiency.
- Utilize SageMaker spot instances for cost-effective model training.
- Consider AWS reserved instances for predictable workloads.

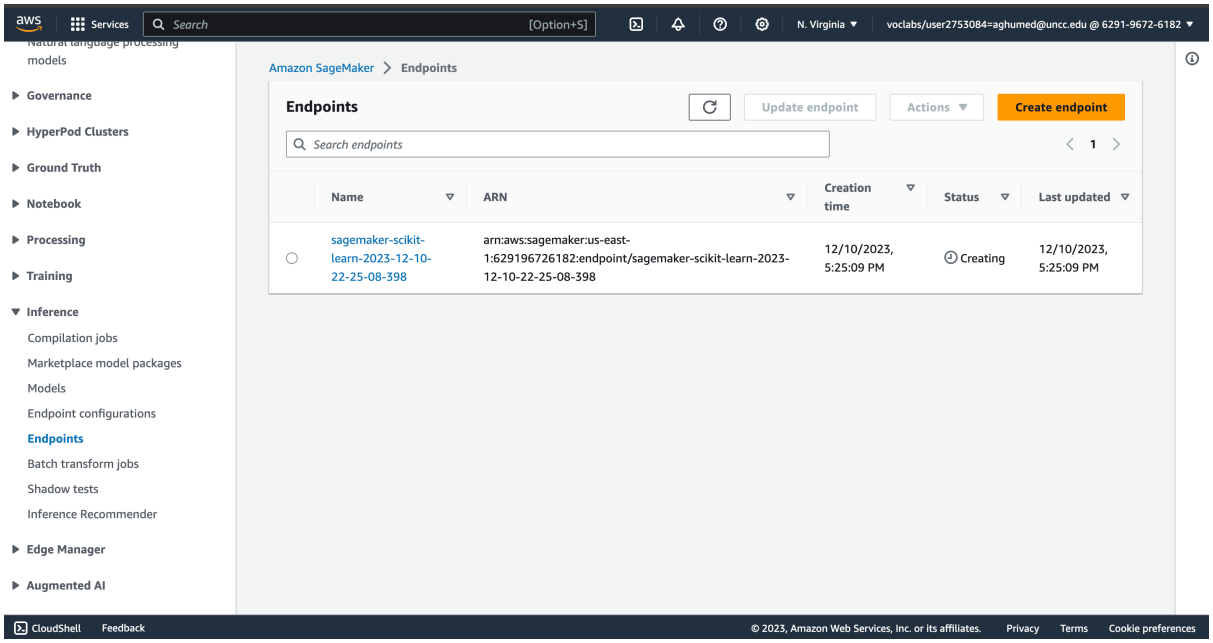
Conclusion after cost analysis:

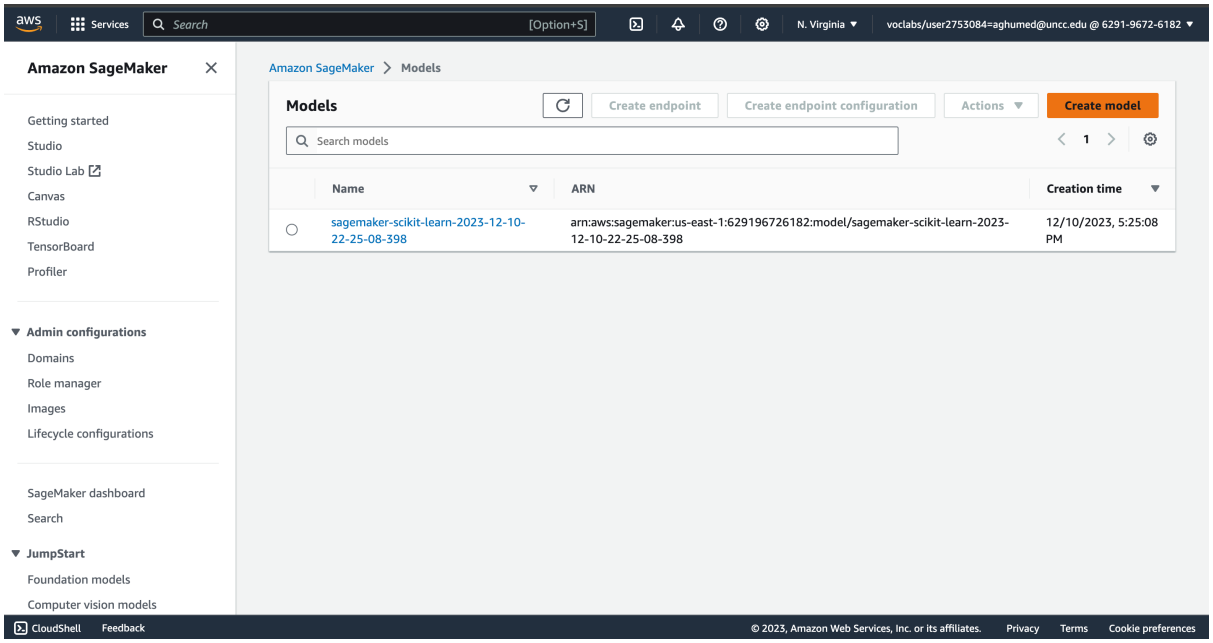
Effective cost management is crucial for the sustainability of the project. Regular monitoring and optimization strategies should be implemented to ensure that the project remains within budget while achieving its objectives.

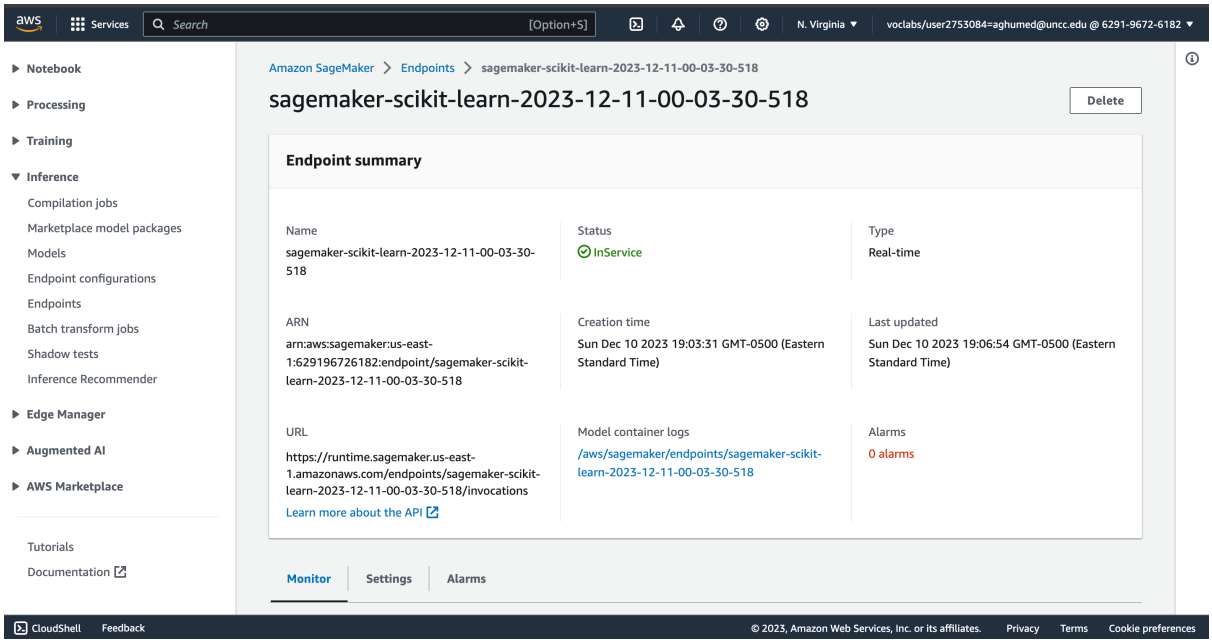
Deployment:

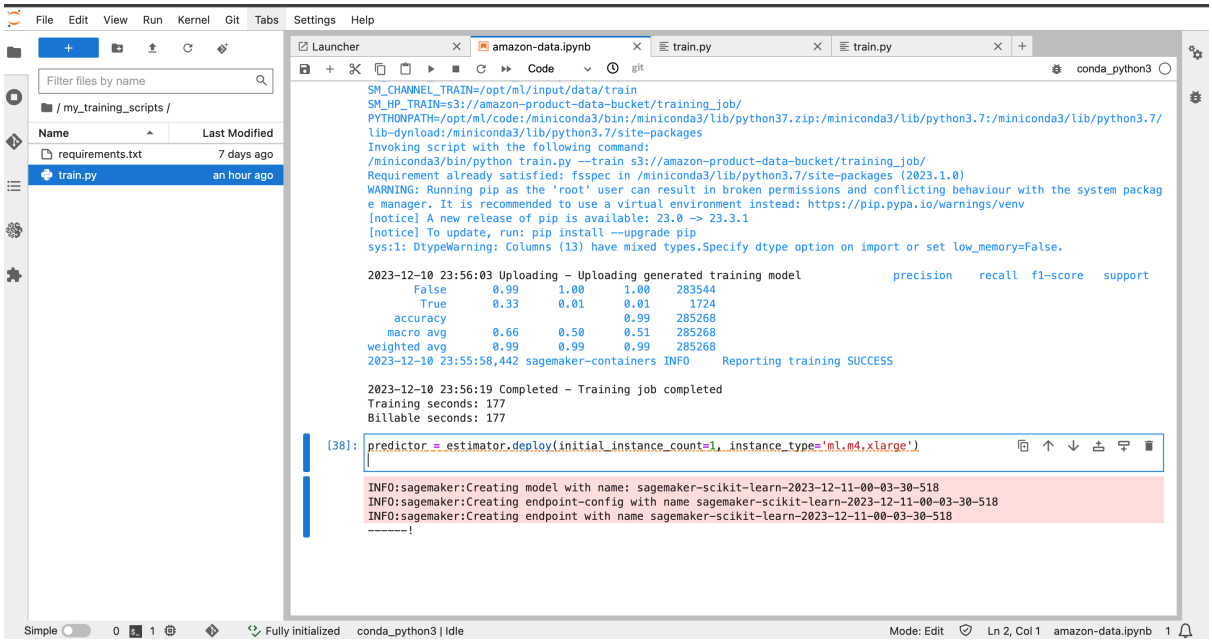

File Edit View Run Kernel Git Tabs Settings Help
Launcher
amazon-data.ipynb
Code
conda_python3
Filter files by name
Name
Last Modified
my_training_scripts 7 days ago
amazon-data.ipynb a minute ago
requirements.txt 7 days ago
train.py 7 days ago
training_data.csv 19 minutes ago
import sagemaker
from sagemaker.sklearn.estimator import SKLearn
role = sagemaker.get_execution_role()
sagemaker_session = sagemaker.Session()
# Create a SageMaker Estimator
estimator = SKLearn(
entry_point='train.py',
source_dir='my_training_scripts/', # Directory containing train.py and requirements.txt
role=role,
instance_count=1,
instance_type='ml.m4.xlarge',
framework_version='0.23-1',
py_version='py3',
hyperparameters={'train': 's3://amazon-product-data-bucket/training_job/'}
)
# Start the training job
estimator.fit({'train': 's3://amazon-product-data-bucket/training_job/'})
predictor = estimator.deploy(initial_instance_count=1, instance_type='ml.m4.xlarge')
INFO:sagemaker:Creating model with name: sagemaker-scikit-learn-2023-12-10-22-25-08-398
INFO:sagemaker:Creating endpoint-config with name sagemaker-scikit-learn-2023-12-10-22-25-08-398
INFO:sagemaker:Creating endpoint with name sagemaker-scikit-learn-2023-12-10-22-25-08-398
Simple
Fully initialized
conda_python3 | Busy
Mode: Edit
Ln 1, Col 85
amazon-data.ipynb


aws
Services
Search
[Option+S]
N. Virginia
voclabs/user2753084=aghumed@uncc.edu @ 6291-9672-6182
models
Governance
HyperPod Clusters
Ground Truth
Notebook
Processing
Training
Inference
Compilation jobs
Marketplace model packages
Models
Endpoint configurations
Endpoints
Batch transform jobs
Shadow tests
Inference Recommender
Edge Manager
Augmented AI
Amazon SageMaker > Endpoint configuration
Endpoint configuration
Apply to endpoint
Clone
Actions
Create endpoint configuration
Search endpoint configuration
Name
ARN
Creation time
sagemaker-scikit-learn-2023-12-10-22-25-08-398
arn:aws:sagemaker:us-east-1:629196726182:endpoint-config/sagemaker-scikit-learn-2023-12-10-22-25-08-398
12/10/2023, 5:25:09 PM
CloudShell
Feedback
© 2023, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences


Services
Search
[Option+S]
N. Virginia
voclabs/user2753084=aghumed@uncc.edu @ 6291-9672-6182
models
Governance
HyperPod Clusters
Ground Truth
Notebook
Processing
Training
Inference
Compilation jobs
Marketplace model packages
Models
Endpoint configurations
Endpoints
Batch transform jobs
Shadow tests
Inference Recommender
Edge Manager
Augmented AI
Amazon SageMaker > Endpoints
Endpoints
Update endpoint
Actions
Create endpoint
Search endpoints
1
Name
ARN
Creation time
Status
Last updated
sagemaker-scikit-learn-2023-12-10-22-25-08-398
arn:aws:sagemaker:us-east-1:629196726182:endpoint/sagemaker-scikit-learn-2023-12-10-22-25-08-398
12/10/2023, 5:25:09 PM
Creating
12/10/2023, 5:25:09 PM
CloudShell
Feedback
© 2023, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences


Services
Search
[Option+S]
N. Virginia
voclabs/user2753084=aghumed@uncc.edu @ 6291-9672-6182
Amazon SageMaker
Getting started
Studio
Studio Lab
Canvas
RStudio
TensorBoard
Profiler
Admin configurations
Domains
Role manager
Images
Lifecycle configurations
SageMaker dashboard
Search
JumpStart
Foundation models
Computer vision models
Amazon SageMaker > Models
Models
Create endpoint
Create endpoint configuration
Actions
Create model
Search models
1
Name
ARN
Creation time
sagemaker-scikit-learn-2023-12-10-22-25-08-398
arn:aws:sagemaker:us-east-1:629196726182:model/sagemaker-scikit-learn-2023-12-10-22-25-08-398
12/10/2023, 5:25:08 PM
CloudShell
Feedback
© 2023, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences


Services
Search
[Option+S]
N. Virginia
voclabs/user2753084=aghumed@uncc.edu @ 6291-9672-6182
Notebook
Processing
Training
Inference
Compilation jobs
Marketplace model packages
Models
Endpoint configurations
Endpoints
Batch transform jobs
Shadow tests
Inference Recommender
Edge Manager
Augmented AI
AWS Marketplace
Tutorials
Documentation
Amazon SageMaker > Endpoints > sagemaker-scikit-learn-2023-12-11-00-03-30-518
sagemaker-scikit-learn-2023-12-11-00-03-30-518
Delete
Endpoint summary
Name
sagemaker-scikit-learn-2023-12-11-00-03-30-518
Status
InService
Type
Real-time
ARN
arn:aws:sagemaker:us-east-1:629196726182:endpoint/sagemaker-scikit-learn-2023-12-11-00-03-30-518
Creation time
Sun Dec 10 2023 19:03:31 GMT-0500 (Eastern Standard Time)
Last updated
Sun Dec 10 2023 19:06:54 GMT-0500 (Eastern Standard Time)
URL
https://runtime.sagemaker.us-east-1.amazonaws.com/endpoints/sagemaker-scikit-learn-2023-12-11-00-03-30-518/invocations
Learn more about the API
Model container logs
/aws/sagemaker/endpoints/sagemaker-scikit-learn-2023-12-11-00-03-30-518
Alarms
0 alarms
Monitor
Settings
Alarms
CloudShell
Feedback
© 2023, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences


File Edit View Run Kernel Git Tabs Settings Help
Launcher
amazon-data.ipynb
train.py
train.py
Code
conda_python3
Filter files by name
/ my_training_scripts /
Name
Last Modified
requirements.txt
7 days ago
train.py
an hour ago
SM_CHANNEL_TRAIN=/opt/ml/input/data/train
SM_HP_TRAIN=s3://amazon-product-data-bucket/training_job/
PYTHONPATH=/opt/ml/code:/miniconda3/bin:/miniconda3/lib/python37.zip:/miniconda3/lib/python3.7:/miniconda3/lib/python3.7/lib-dynload:/miniconda3/lib/python3.7/site-packages
Invoking script with the following command:
/miniconda3/bin/python train.py --train s3://amazon-product-data-bucket/training_job/
Requirement already satisfied: fsspec in /miniconda3/lib/python3.7/site-packages (2023.1.0)
WARNING: Running pip as the 'root' user can result in broken permissions and conflicting behaviour with the system package manager. It is recommended to use a virtual environment instead: https://pip.pypa.io/warnings/venv
[notice] A new release of pip is available: 23.0 -> 23.3.1
[notice] To update, run: pip install --upgrade pip
sys:1: DtypeWarning: Columns (13) have mixed types.Specify dtype option on import or set low_memory=False.
2023-12-10 23:56:03 Uploading - Uploading generated training model
precision recall f1-score support
False 0.99 1.00 1.00 283544
True 0.33 0.01 0.01 1724
accuracy 0.99 285268
macro avg 0.66 0.50 0.51 285268
weighted avg 0.99 0.99 0.99 285268
2023-12-10 23:55:58,442 sagemaker-containers INFO Reporting training SUCCESS
2023-12-10 23:56:19 Completed - Training job completed
Training seconds: 177
Billable seconds: 177
[38]:
predictor = estimator.deploy(initial_instance_count=1, instance_type='ml.m4.xlarge')
INFO:sagemaker:Creating model with name: sagemaker-scikit-learn-2023-12-11-00-03-30-518
INFO:sagemaker:Creating endpoint-config with name sagemaker-scikit-learn-2023-12-11-00-03-30-518
INFO:sagemaker:Creating endpoint with name sagemaker-scikit-learn-2023-12-11-00-03-30-518
------!
Simple
Fully initialized
conda_python3 | Idle
Mode: Edit
Ln 2, Col 1
amazon-data.ipynb