CS 461

Lab Assignment 10

Name: Gandhi Dhruv Vipulkumar

Institute ID: 202151053

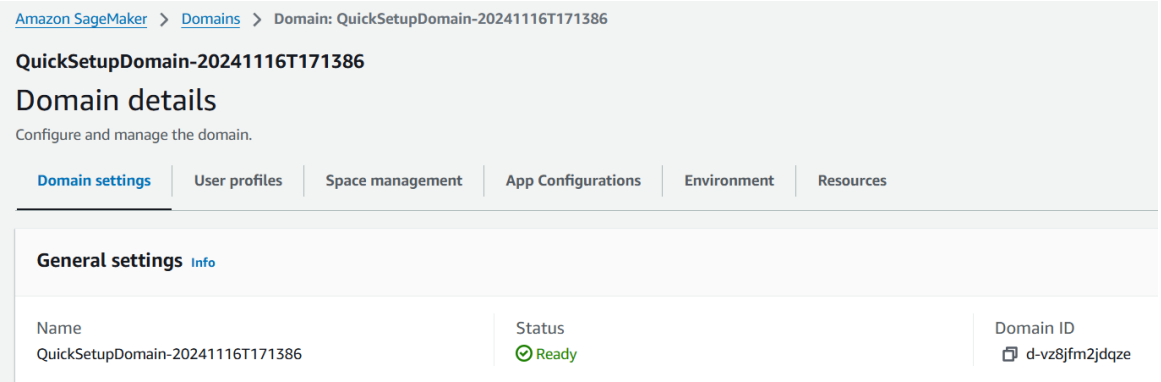
Date: 16-11-2024

Q. Working with AWS and Google Collab Architectures

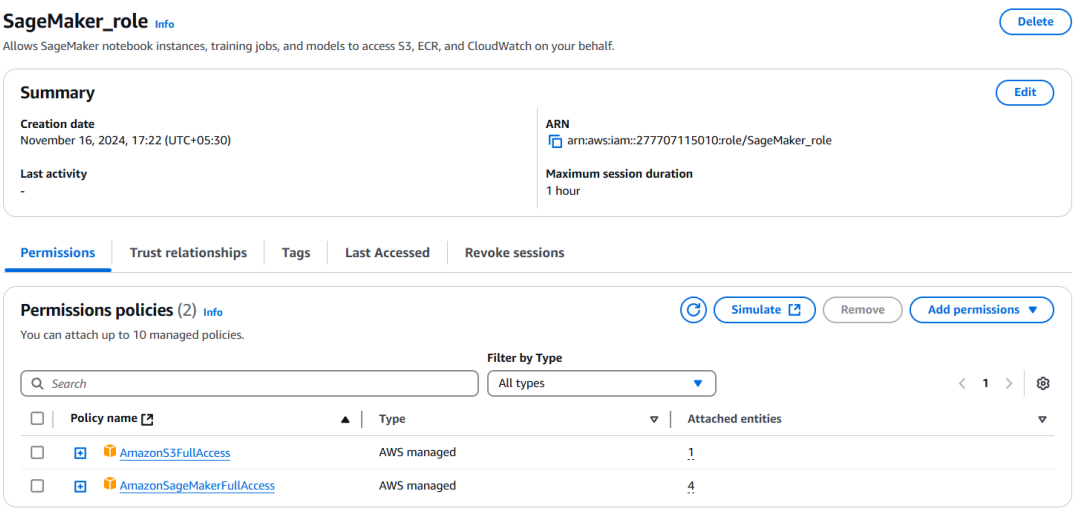
NOTE: No major changes done in provided ipynb file except updating the ARN role, the bucket name and installing s3fs thorough pip.

Steps Followed:

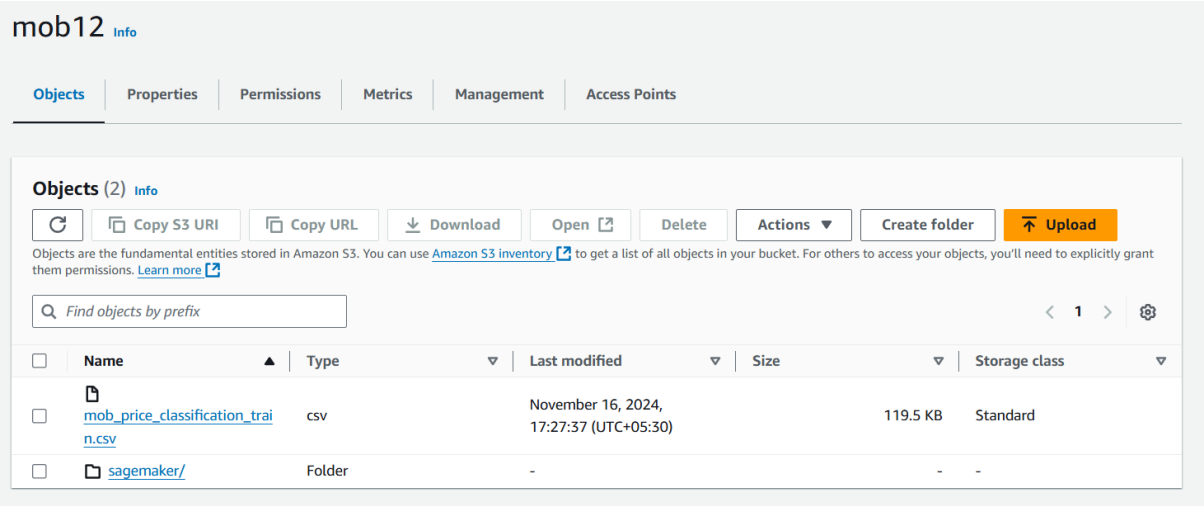
1. Create a Domain in SageMaker



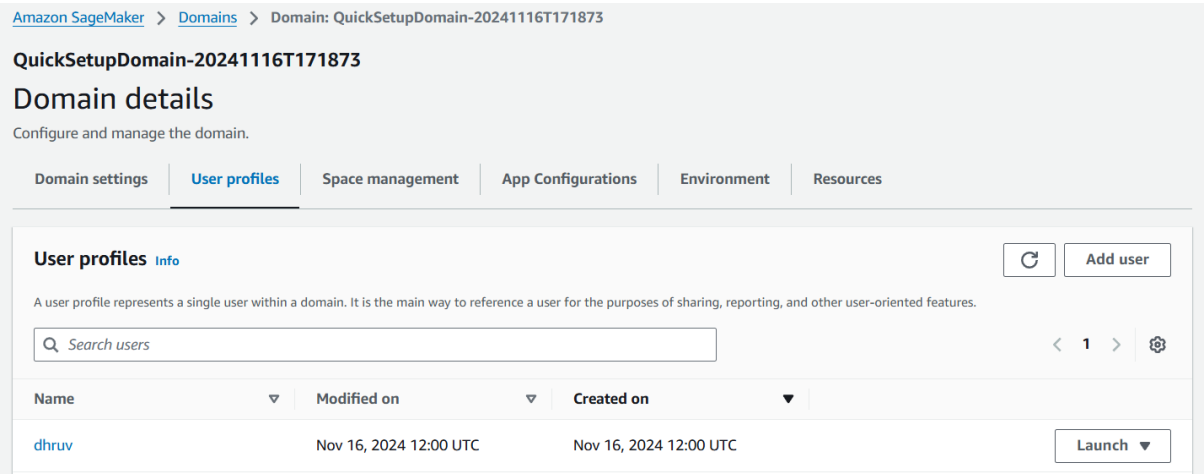
2. Create an IAM User Role and give AmazonSageMakerFullAccess permission and attach AmazonS3FullAccess policy for data access



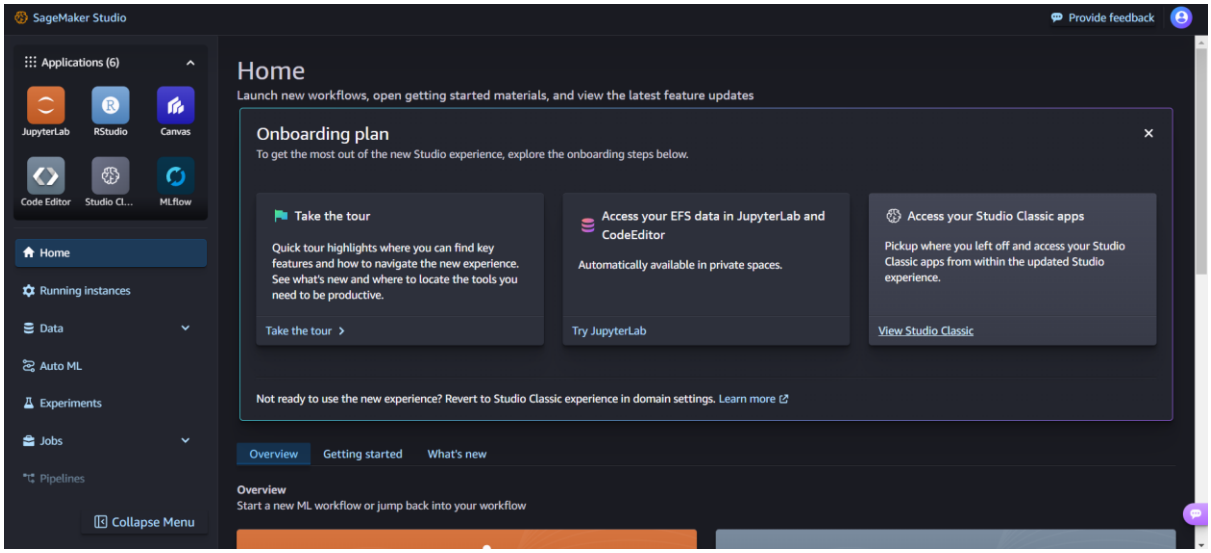
3. Create an S3 Bucket and Upload the Dataset i.e. mob_price_classification_train.csv



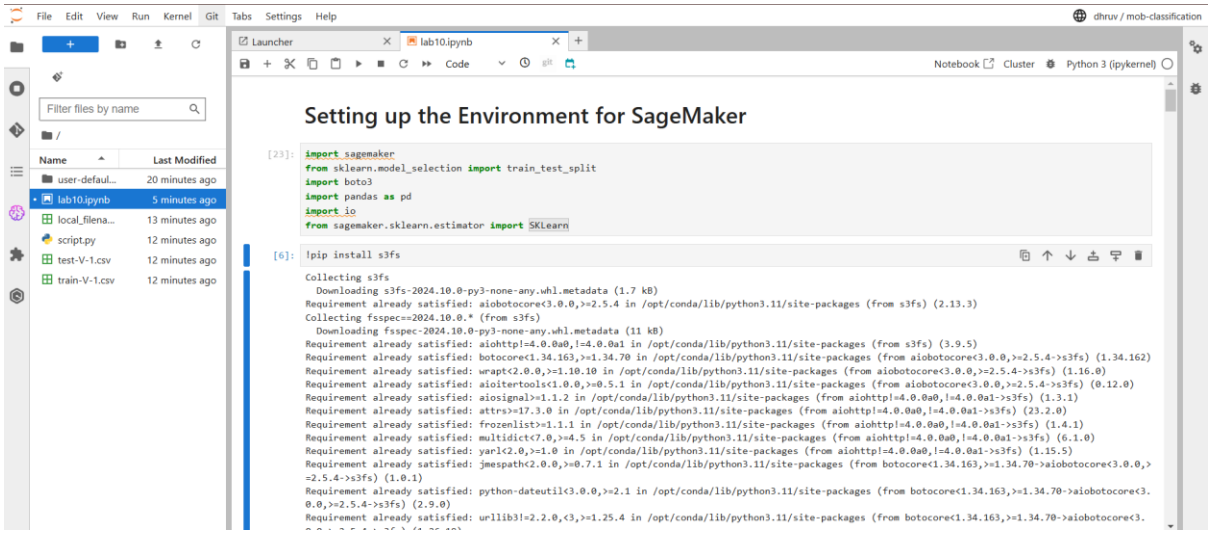
4. Create a new user profile in the created domain of SageMaker console:



5. Launch Studio from the SageMaker console.



6. In Studio, select Jupyter Lab and upload the .ipynb file.



7. Update the bucket name, ARN role or any other dependencies and run it.

Final Output:

```
building tree 92 of 100
building tree 93 of 100
building tree 94 of 100
building tree 95 of 100building tree 96 of 100
building tree 97 of 100
building tree 98 of 100
building tree 99 of 100
building tree 100 of 100
[Parallel(n_jobs=-1)]: Done 100 out of 100 | elapsed:    0.3s finished
Model persisted at /opt/ml/model/model.joblib
[Parallel(n_jobs=2)]: Using backend ThreadingBackend with 2 concurrent workers.
[Parallel(n_jobs=2)]: Done  28 tasks      | elapsed:    0.0s
[Parallel(n_jobs=2)]: Done 100 out of 100 | elapsed:    0.0s finished
---- METRICS RESULTS FOR TESTING DATA ----
Total Rows are:  300
[TESTING] Model Accuracy is:  0.8833333333333333
[TESTING] Testing Report:
              precision    recall  f1-score   support

           0       0.95      1.00      0.97        69
           1       0.85      0.80      0.83        66
           2       0.80      0.77      0.79        74
           3       0.91      0.95      0.93        91

    accuracy                           0.88       300
   macro avg       0.88      0.88      0.88       300
weighted avg       0.88      0.88      0.88       300
2024-11-16 12:19:26,817 sagemaker-containers INFO     Reporting training SUCCESS

2024-11-16 12:19:31 Training - Training image download completed. Training in progress.
2024-11-16 12:19:31 Uploading - Uploading generated training model
2024-11-16 12:19:44 Completed - Training job completed
Training seconds: 84
Billable seconds: 30
Managed Spot Training savings: 64.3%
```

Successfully trained and attained accuracy of 88%.

Conclusion: Successfully able to utilize Amazon SageMaker to run and train machine learning model and S3 for storage.