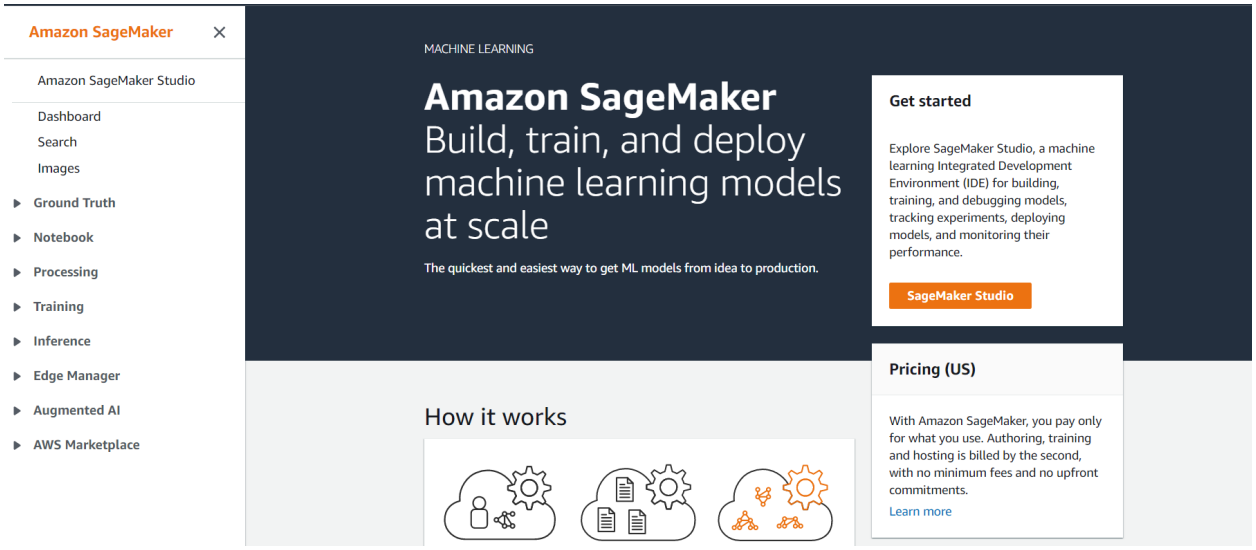
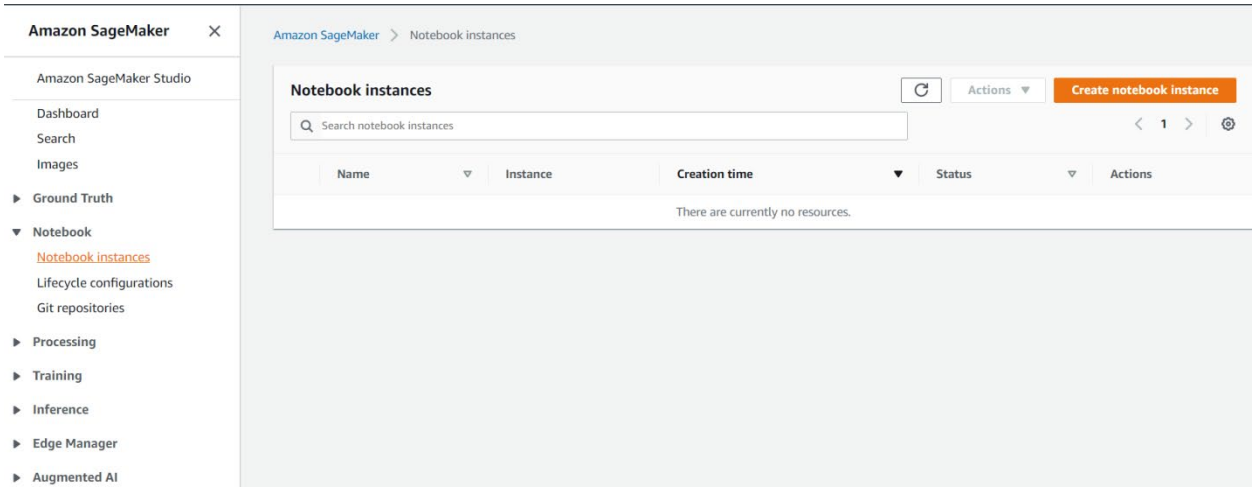
AWS SageMaker

Build, Train, and Deploy Machine Learning models at scale

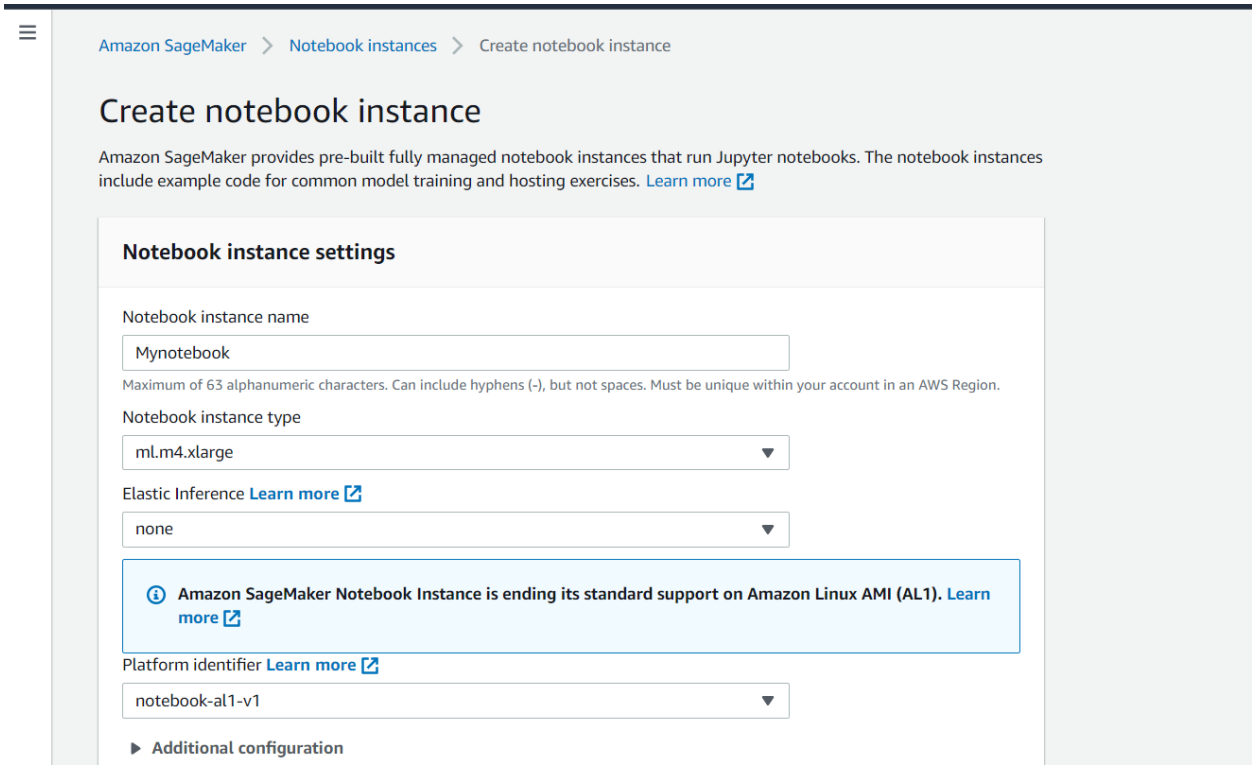
1. Creating and Importing Data



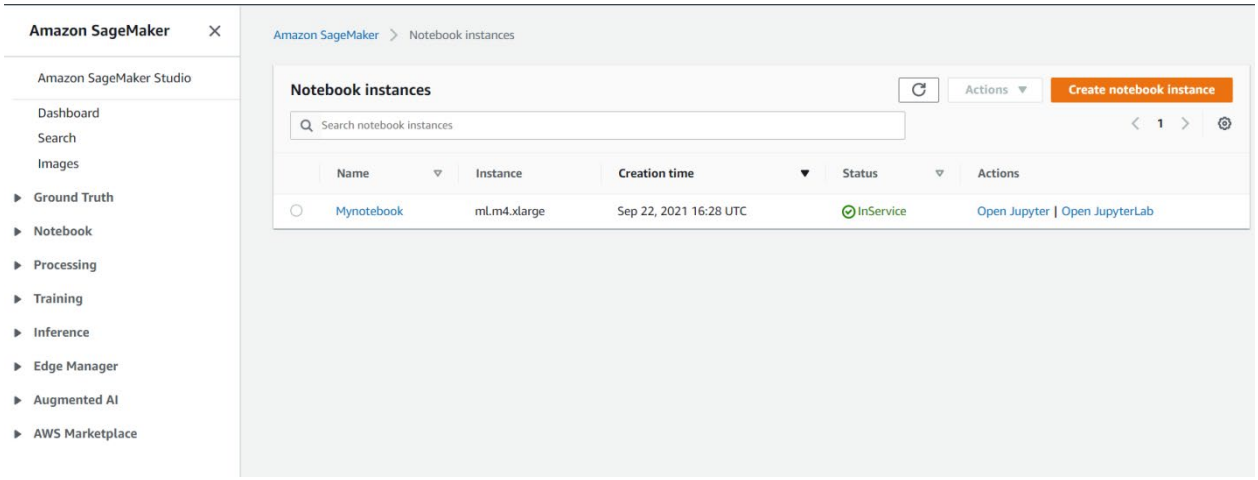
AWS SageMaker Dashboard



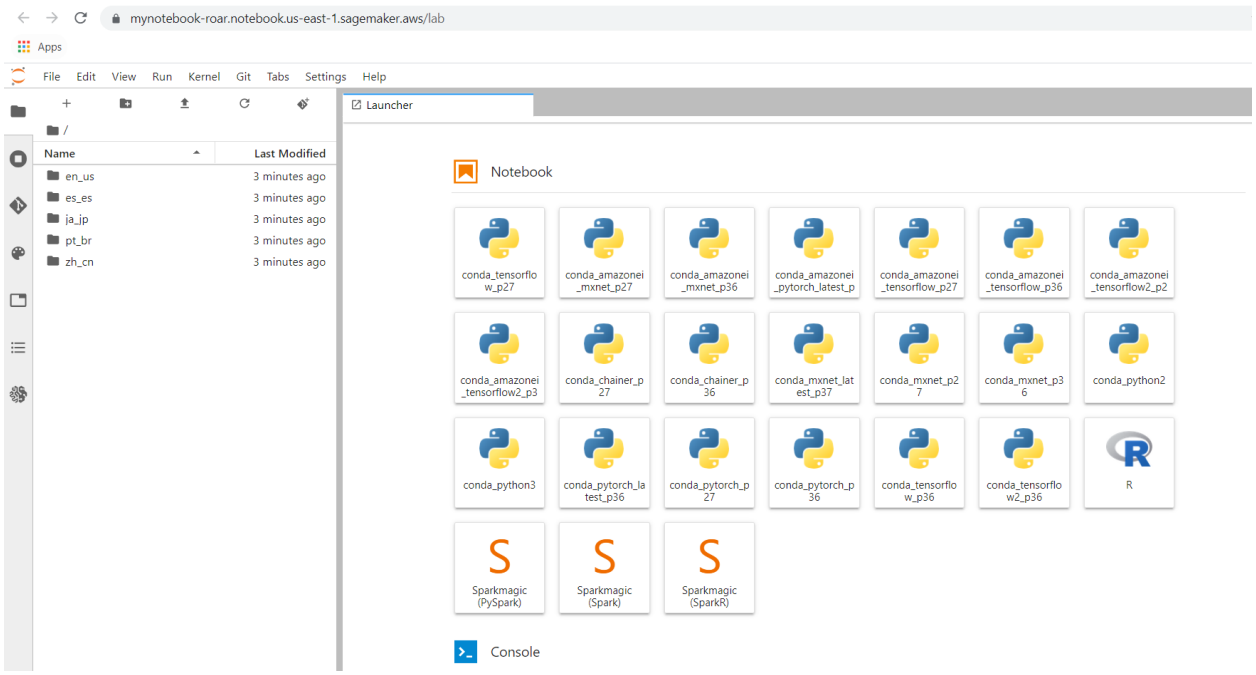
NoteBook Instances Page



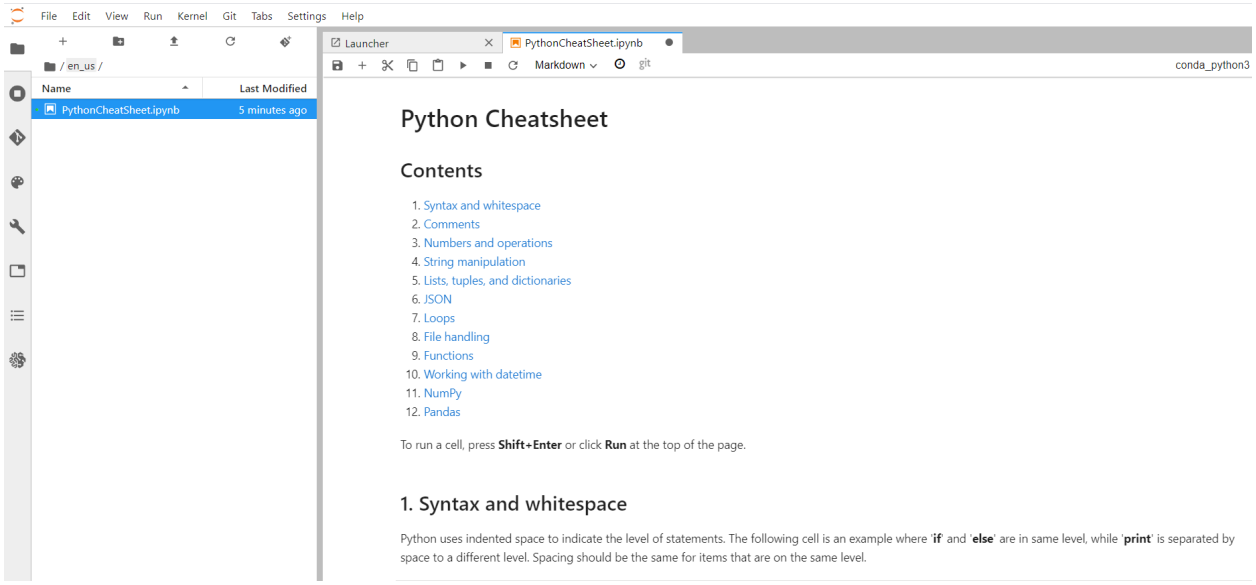
Creating Notebook Instance



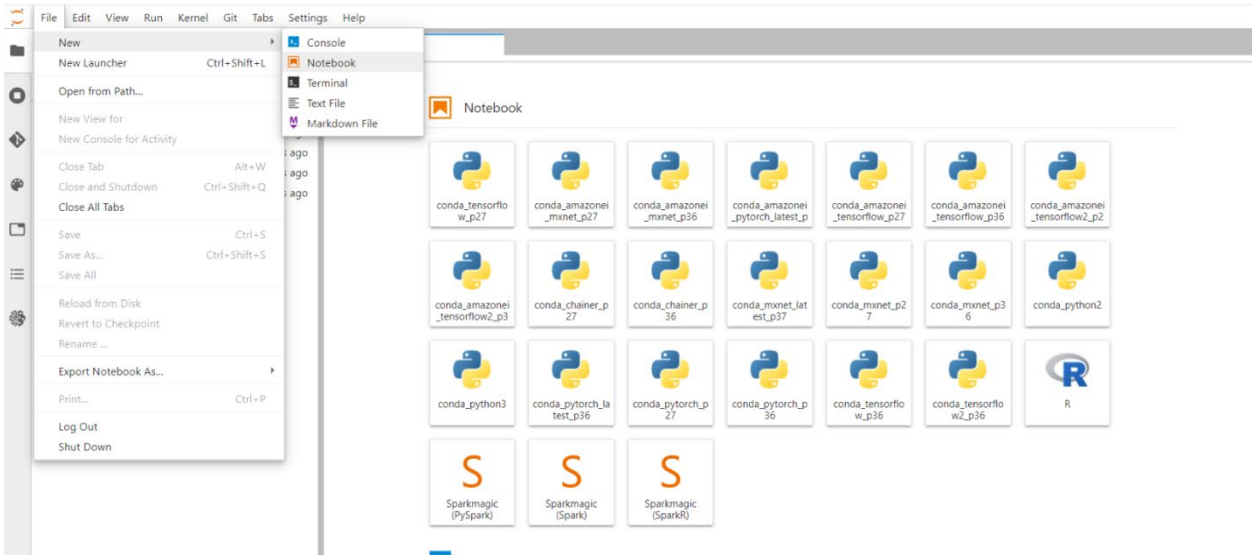
Notebook Successfully Created



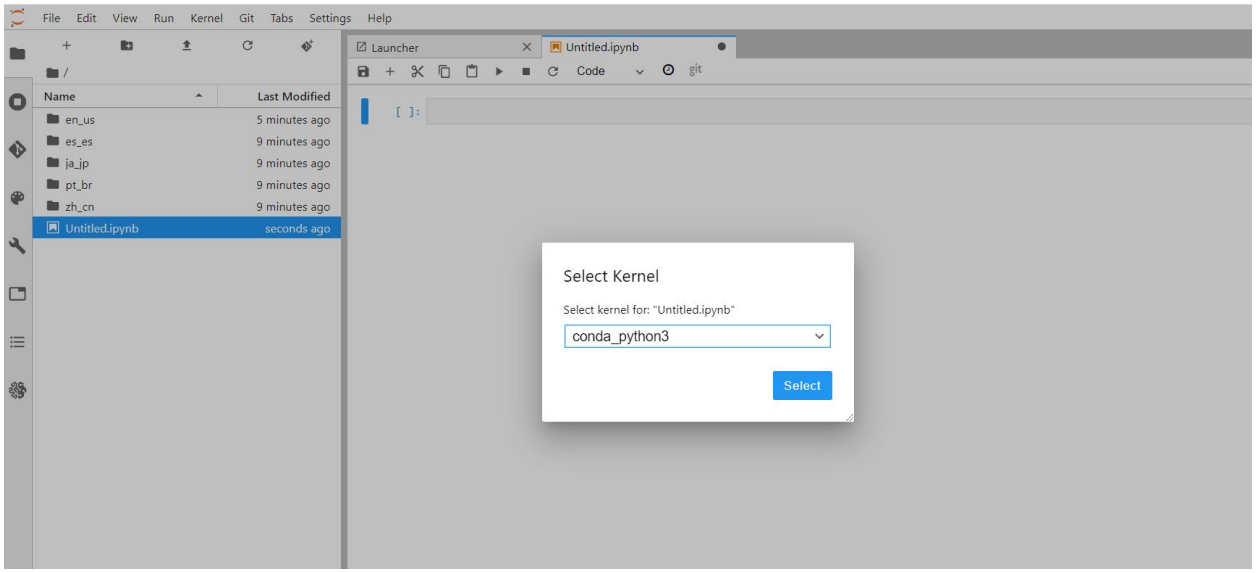
Jupyter Lab for the Notebook



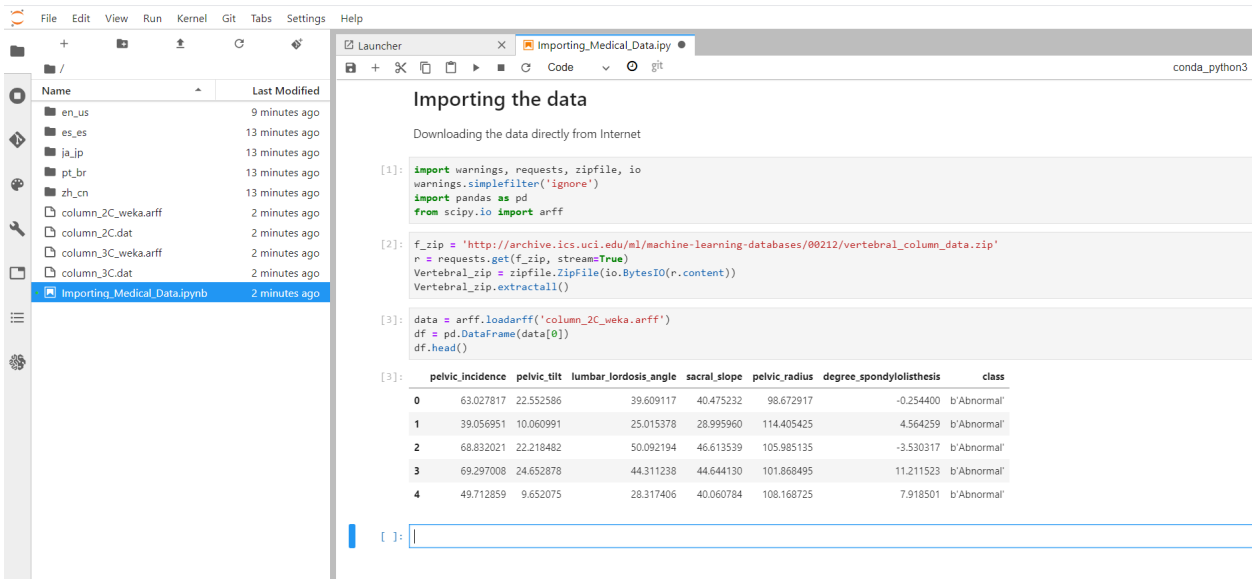
Ipynb files opens in a new window in Jupyter Lab



Creating New Notebook

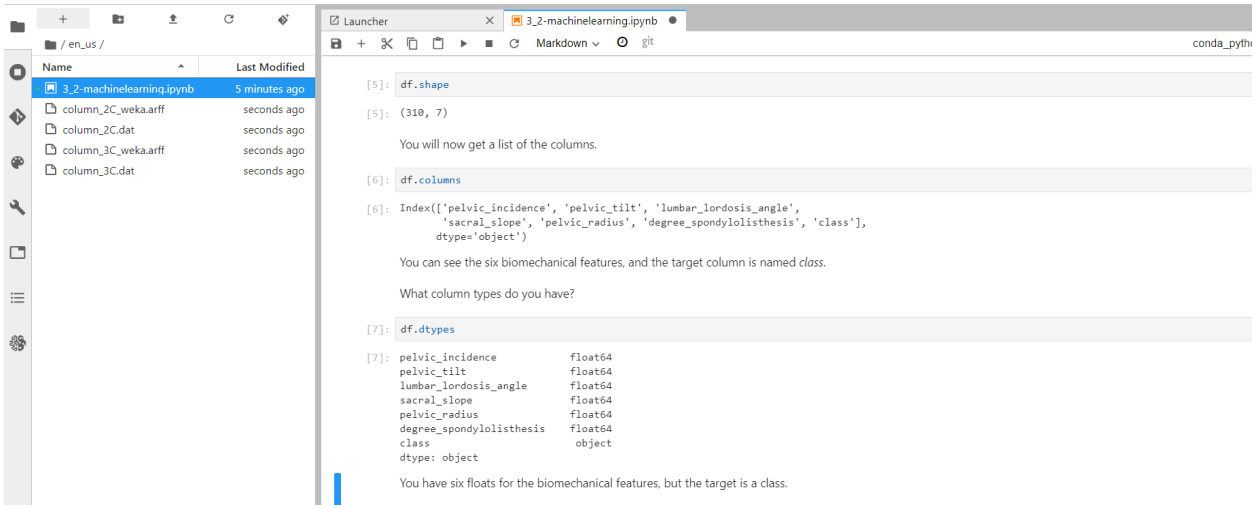


Selecting the Kernel for New Notebook



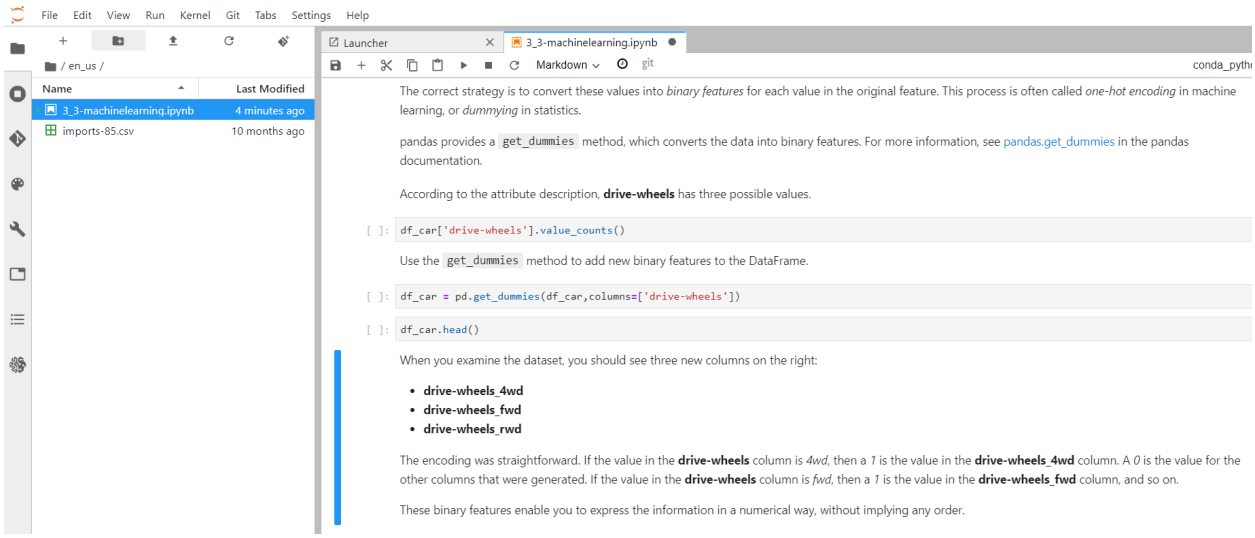
Working on New Notebook (Downloading and Importing Data)

2. Exploring Data using Pandas, Matplotlib



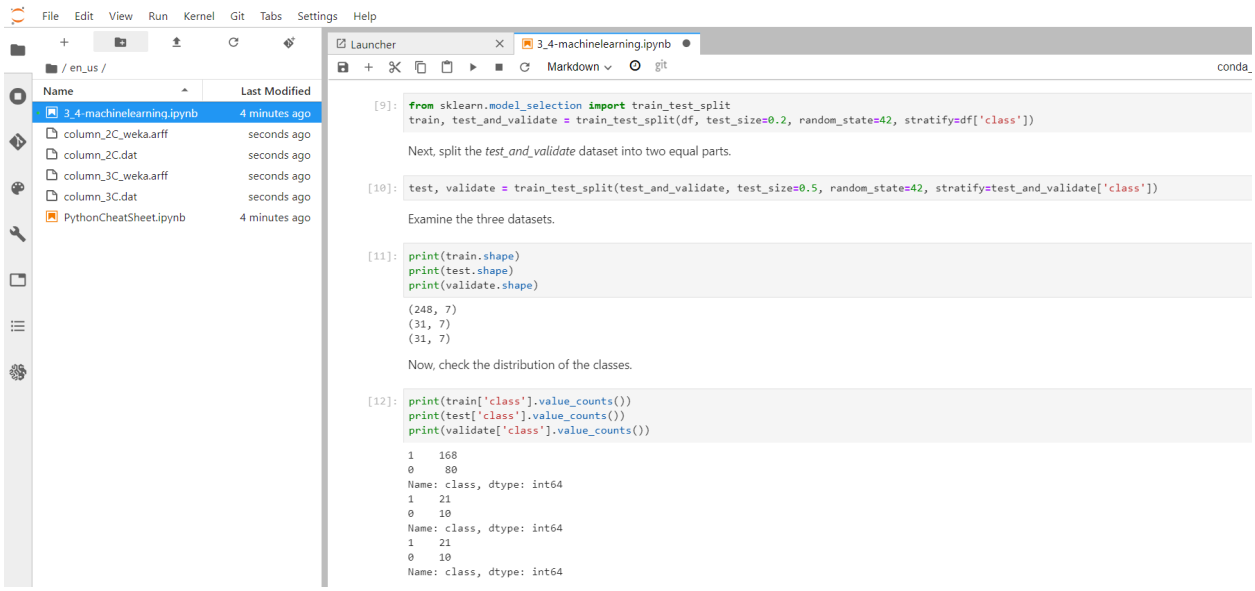
Exploratory Data Analysis

3. Feature Engineering
    - Cleaning Data
    - Dealing with Outliers and Selecting Features
    - Encoding Categorical Data

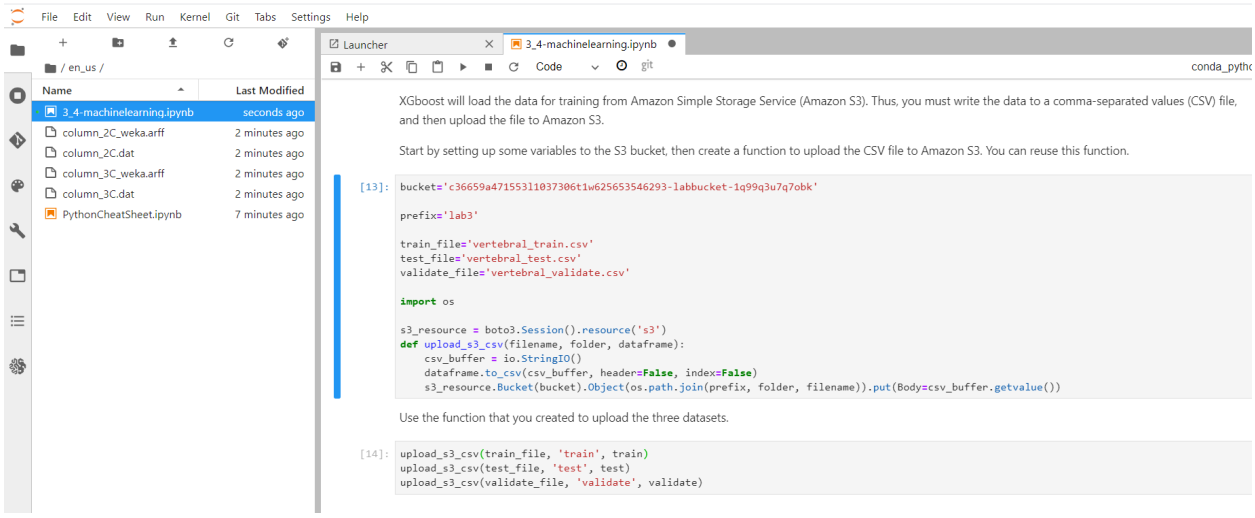


Encoding Categorical Data using Pandas (get_dummies) – AWS SageMaker Notebook

4. Training
    - Train / Test / Validation Split

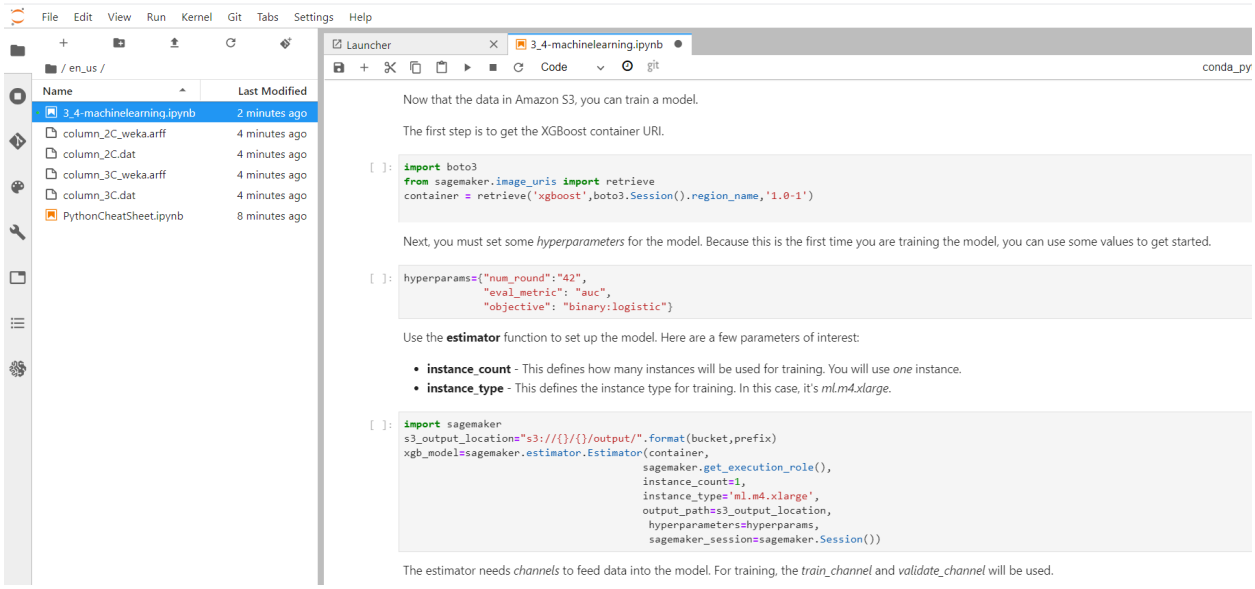


- Uploading Data to S3 Bucket for Model Building



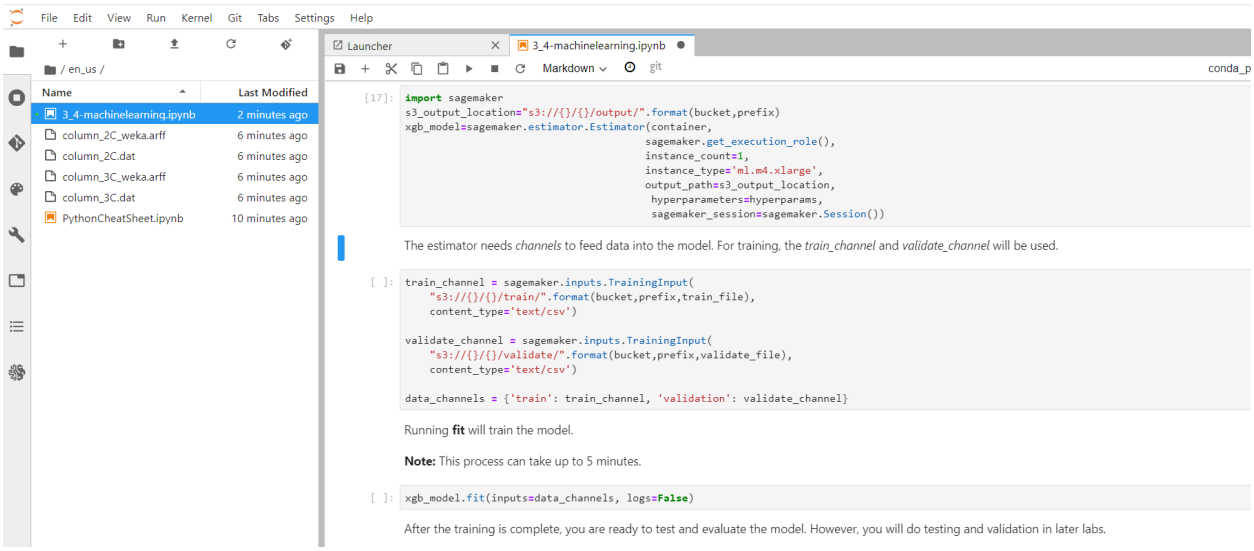
XGBoost Model loads data from S3 bucket (uploading data to S3 bucket)

- Training the Model



XGBoost Model Retrieved from Container and setting up

- Fitting the Model



Creating Channels and Fitting the Model

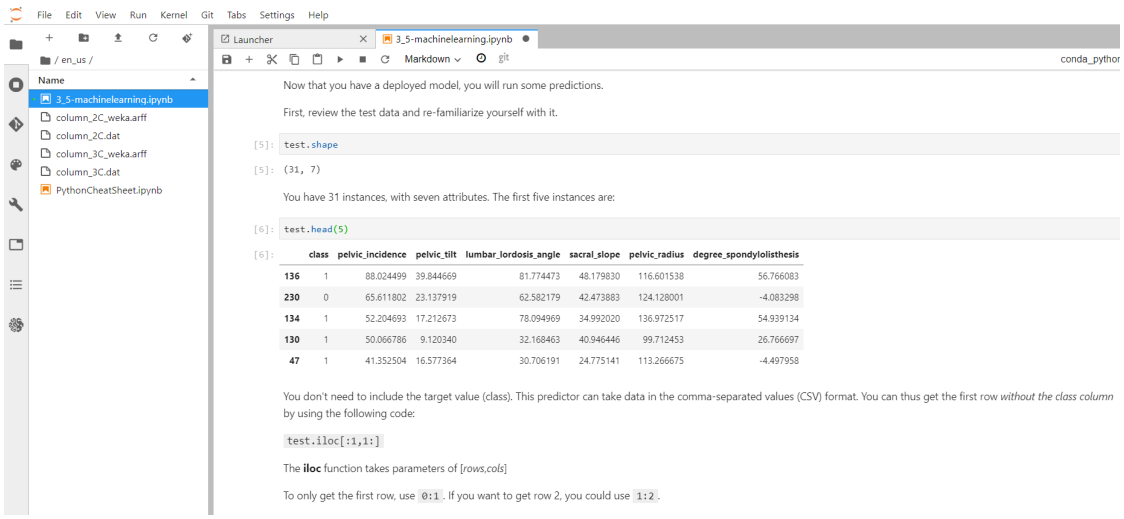
5. Hosting & Using the Model

Now that you have a trained model, you can host it by using Amazon SageMaker hosting services.

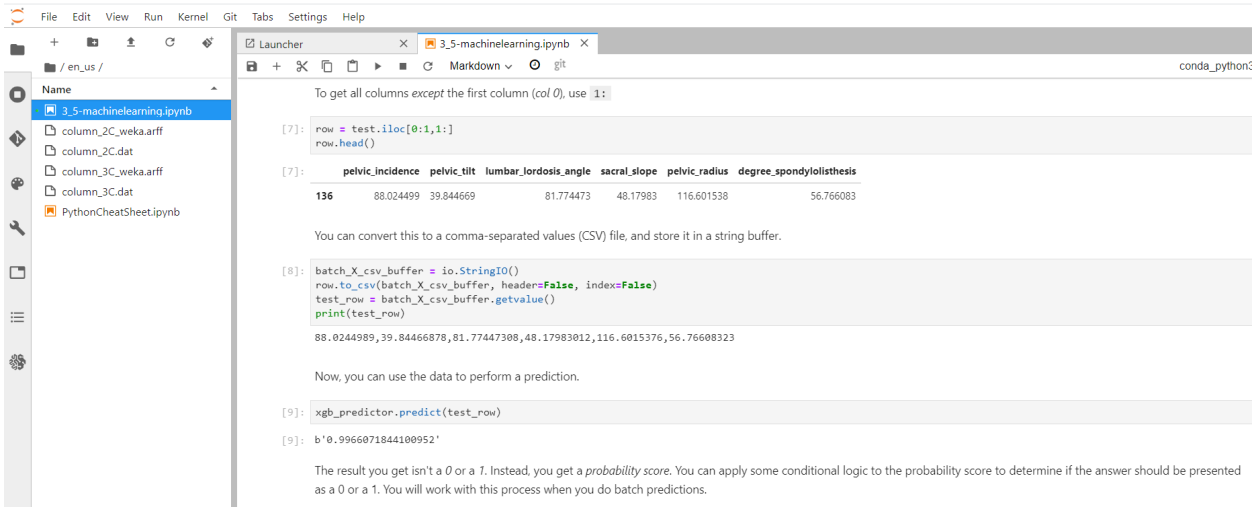
The first step is to deploy the model. Because you have a model object, *xgb_model*, you can use the **deploy** method. For this lab, you will use a single ml.m4.xlarge instance.

```
[ ]: xgb_predictor = xgb_model.deploy(initial_instance_count=1,
                    serializer = sagemaker.serializers.CSVSerializer(),
                    instance_type='ml.m4.xlarge')
```

Deploying the Model



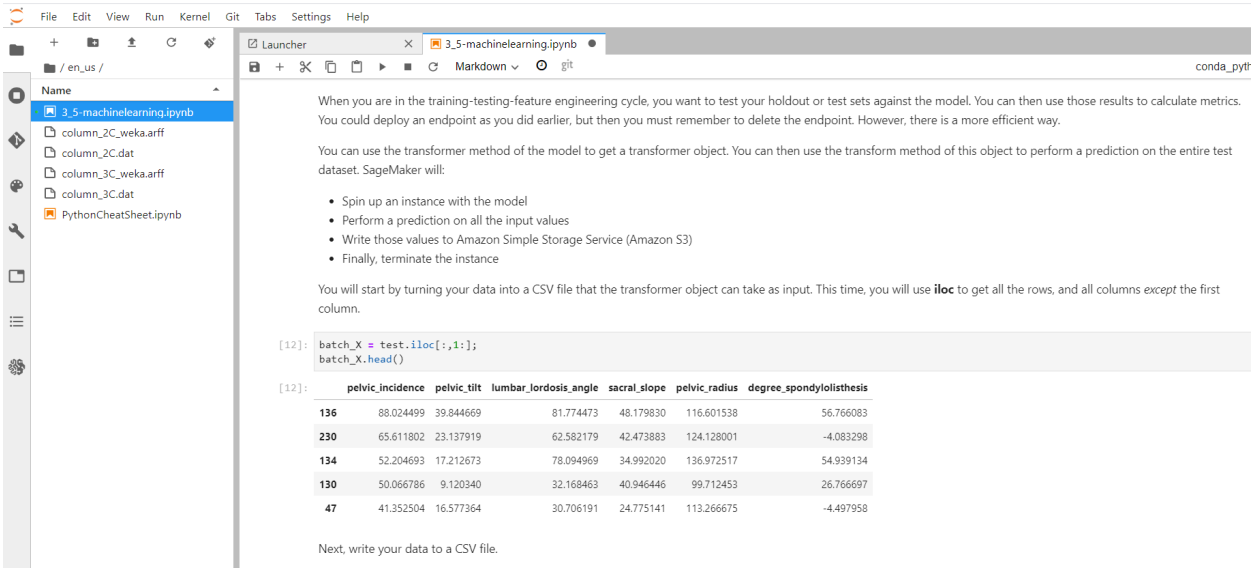
Test data (for Predictions)



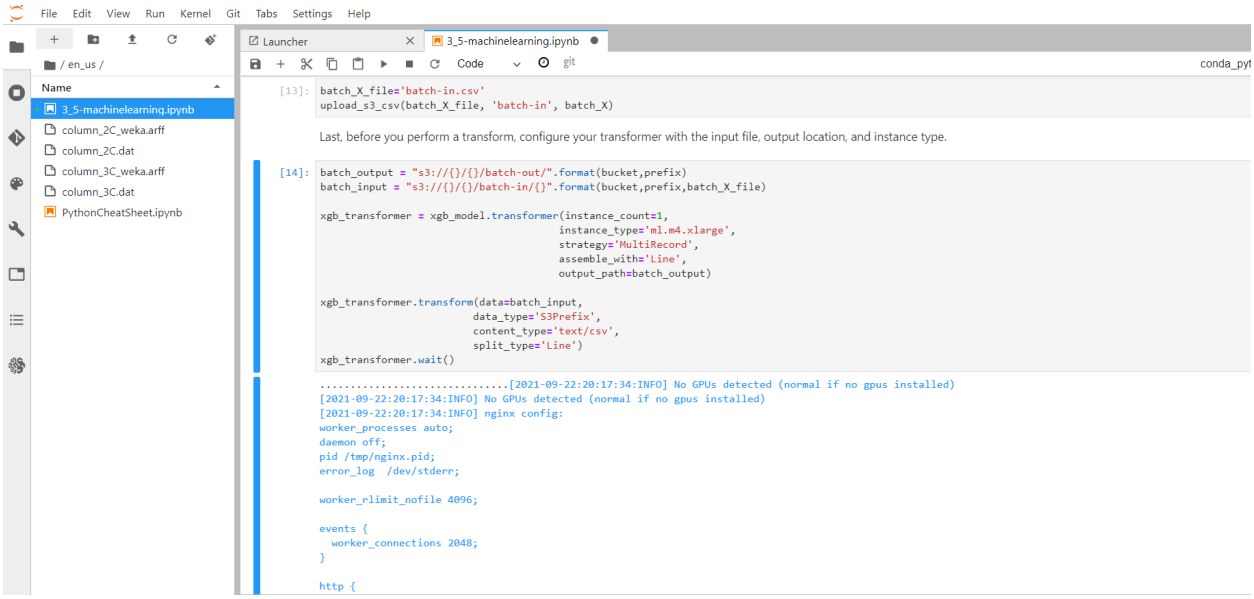
Predictions on Test Data

After Prediction, Delete the endpoint Manually (No an efficient method)

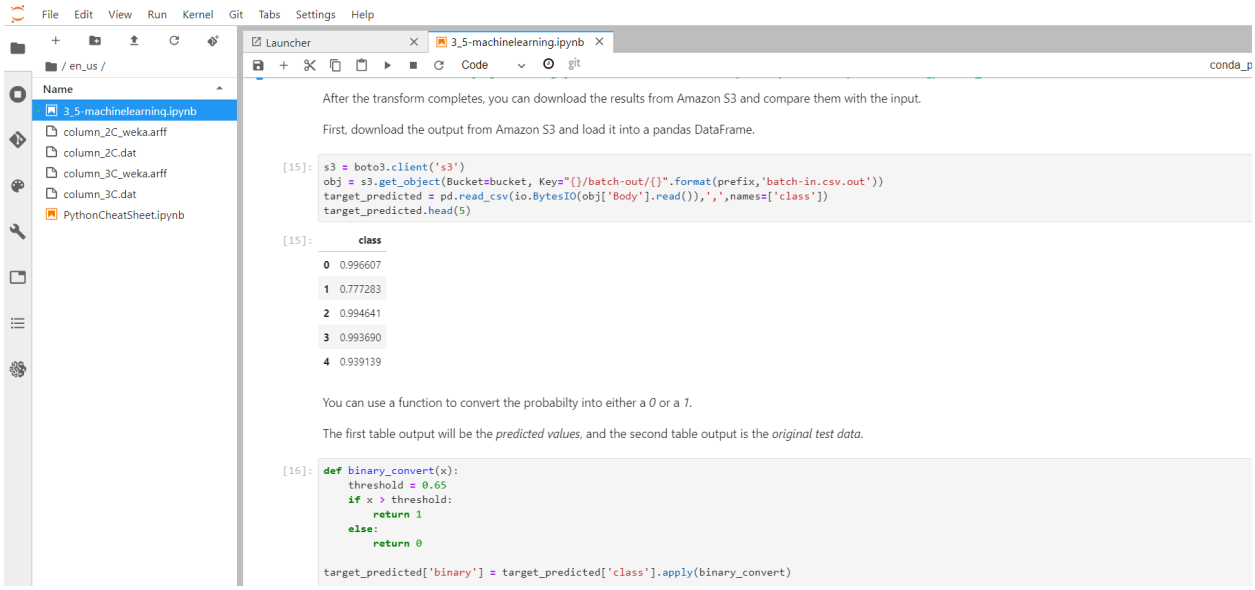
- Using Batch Transformer



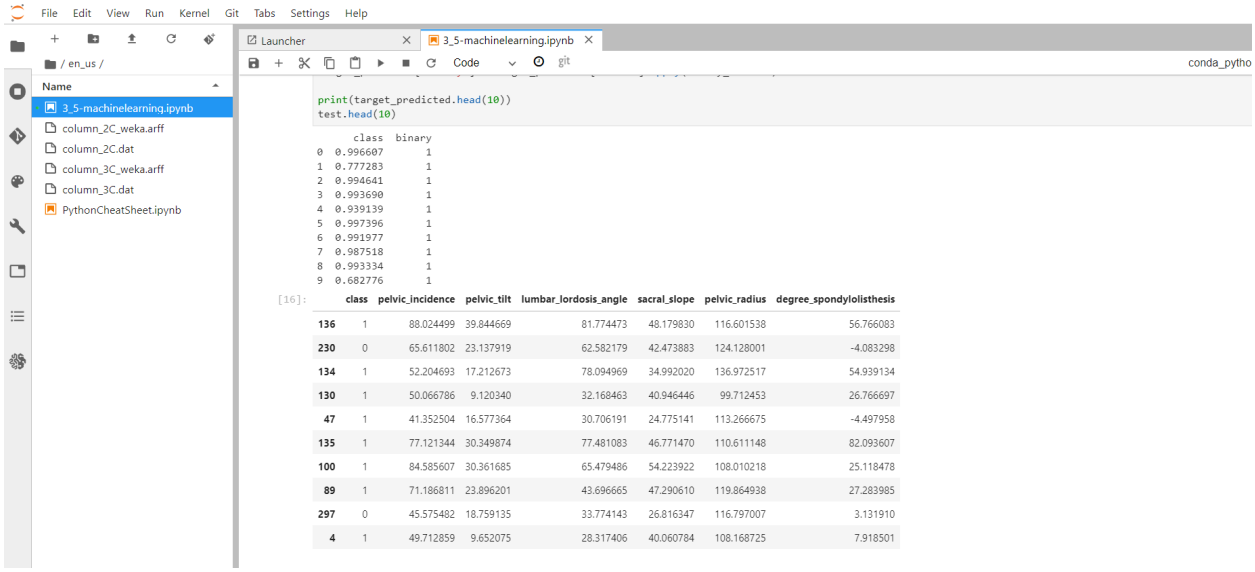
Passing Data to Batch Transformer



Batch Transformer Running



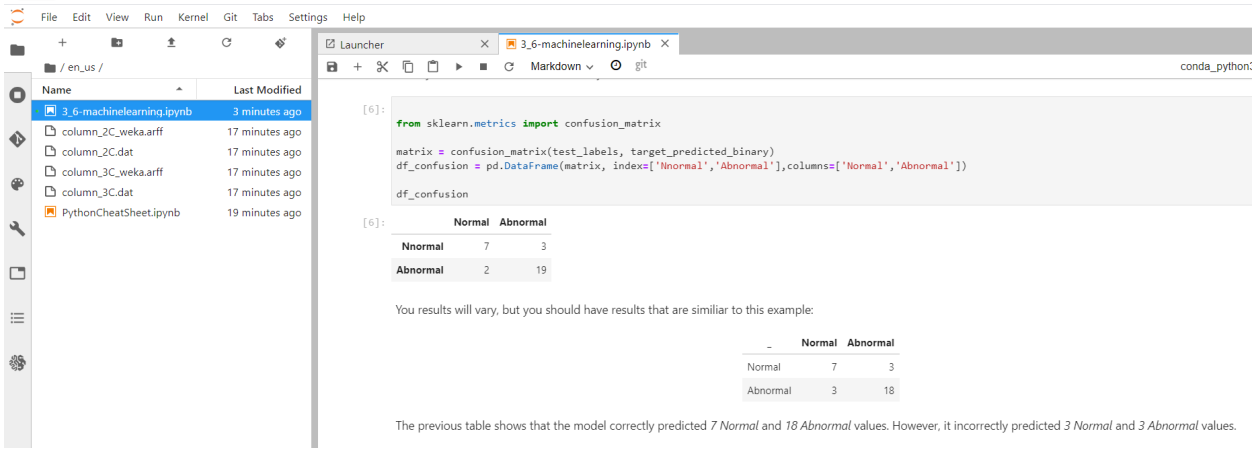
Getting results from S3 bucket and storing it in Dataframe



	class	pelvic_incidence	pelvic_tilt	lumbar_lordosis_angle	sacral_slope	pelvic_radius	degree_spondylolisthesis
136	1	88.024499	39.844669	81.774473	48.179830	116.601538	56.766083
230	0	65.611802	23.137919	62.582179	42.473883	124.128001	-4.083298
134	1	52.204693	17.212673	78.094969	34.992020	136.972517	54.939134
130	1	50.066786	9.120340	32.168463	40.946446	99.712453	26.766697
47	1	41.352504	16.577364	30.706191	24.775141	113.266675	-4.497958
135	1	77.121344	30.349874	77.481083	46.771470	110.611148	82.093607
100	1	84.585607	30.361685	65.479486	54.223922	108.010218	25.118478
89	1	71.186811	23.896201	43.696665	47.290610	119.864938	27.283985
297	0	45.575482	18.759135	33.774143	26.816347	116.797007	3.131910
4	1	49.712859	9.652075	28.317406	40.060784	108.168725	7.918501

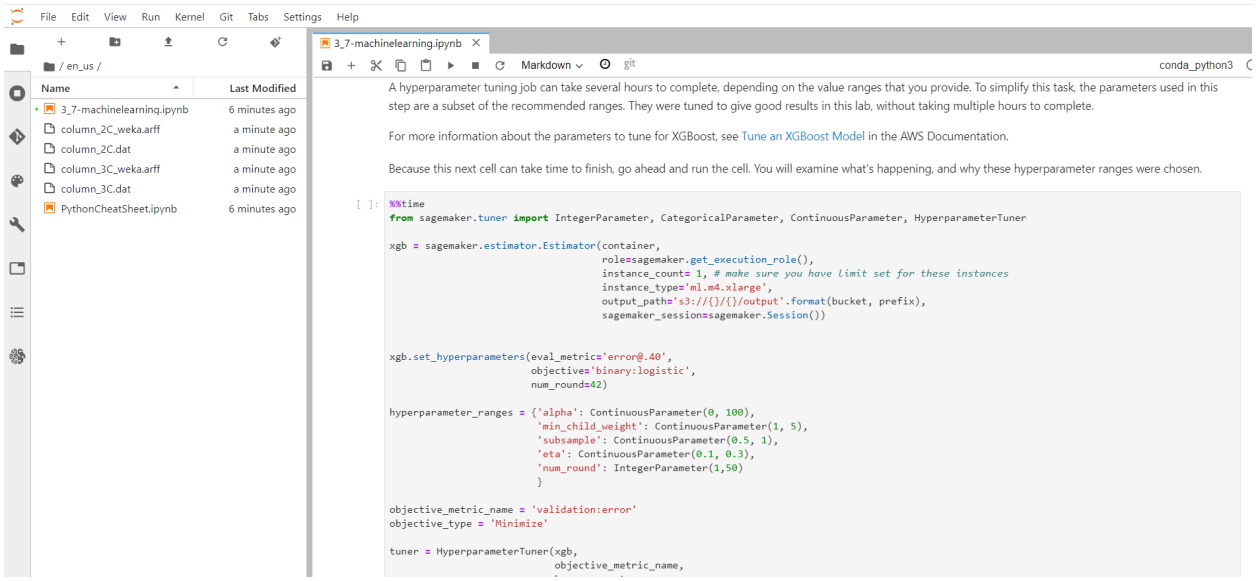
Predictions made and instance terminated automatically

6. Evaluation Metrics

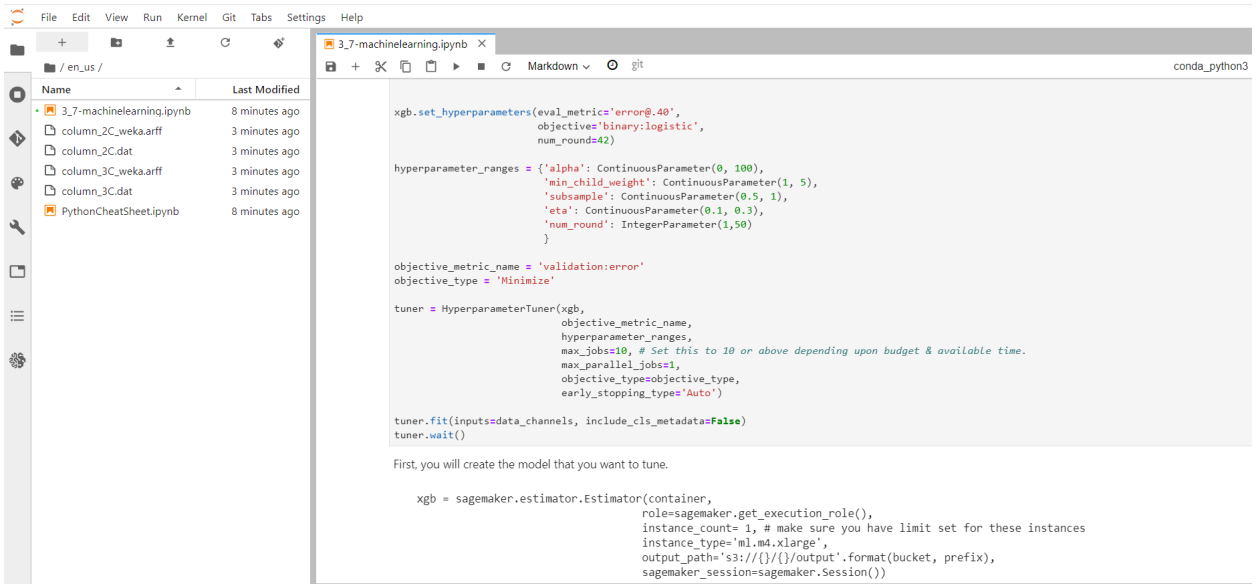


Confusion Matrix

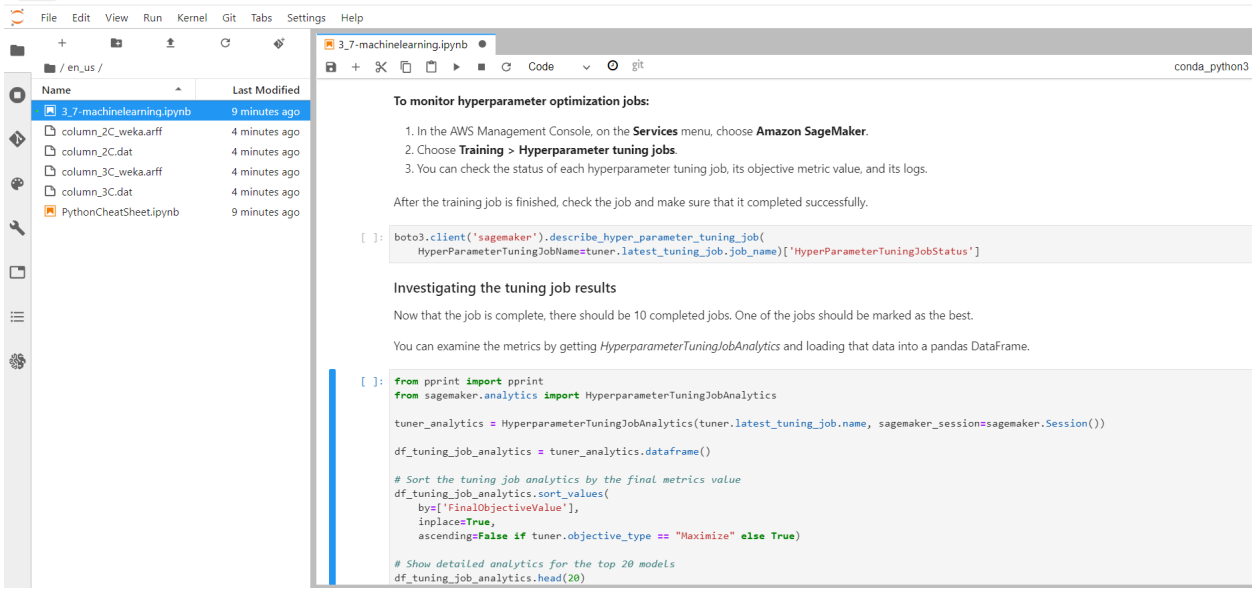
7. HyperParameter Tuning https://docs.aws.amazon.com/sagemaker/latest/dg/xgboost-tuning.html



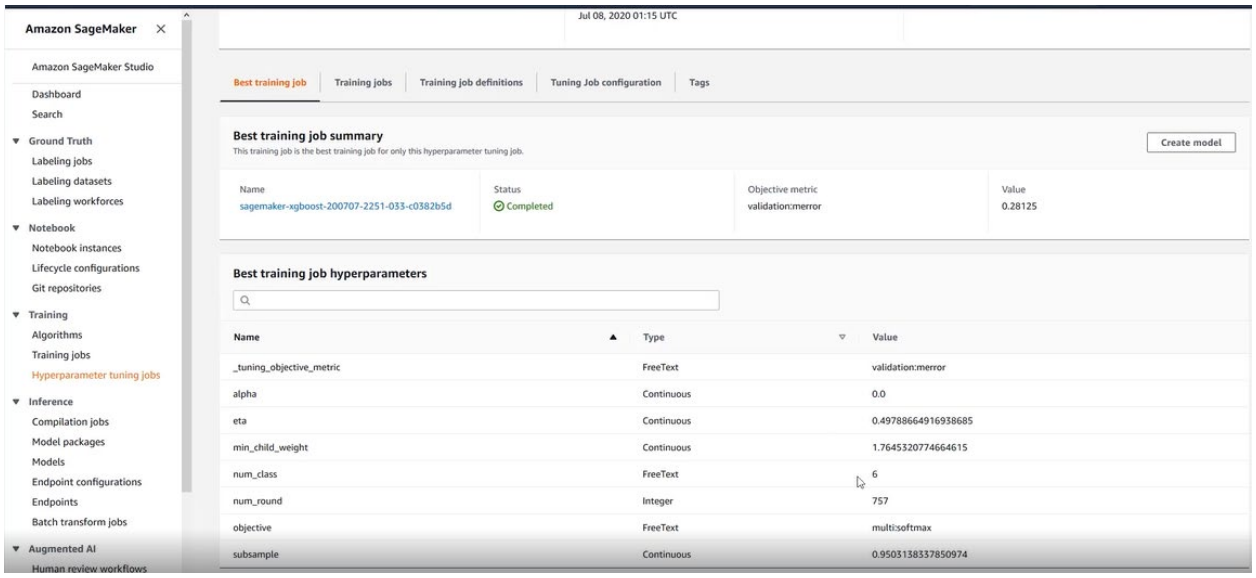
HyperParameter Tuning Job



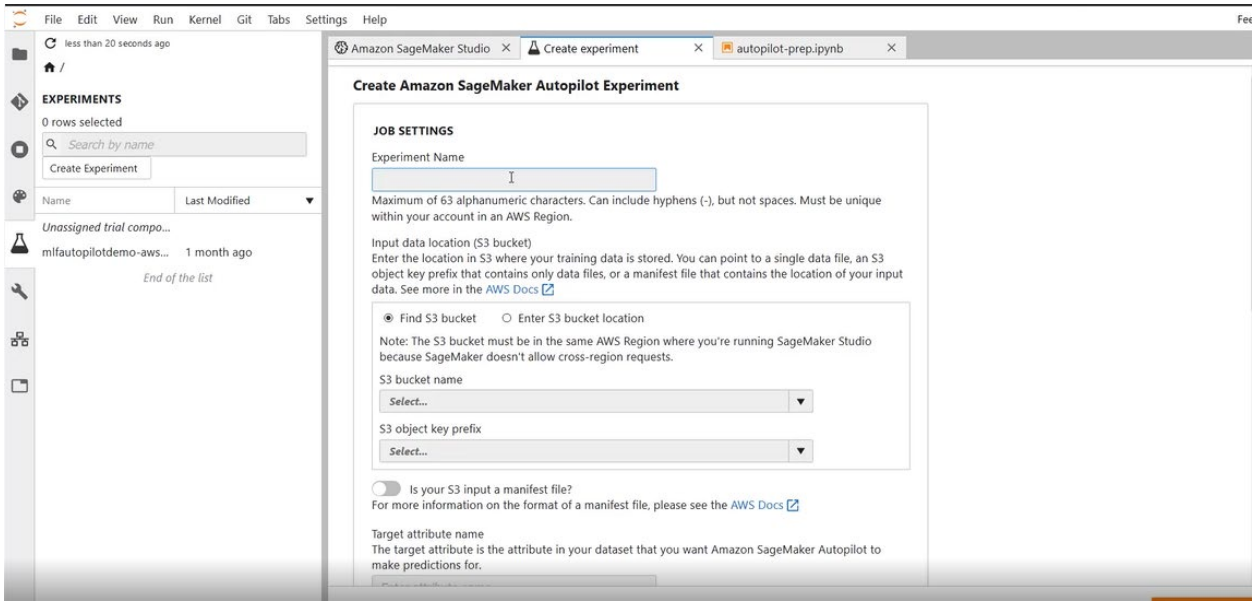
HyperParameter Tuning Job (continued)



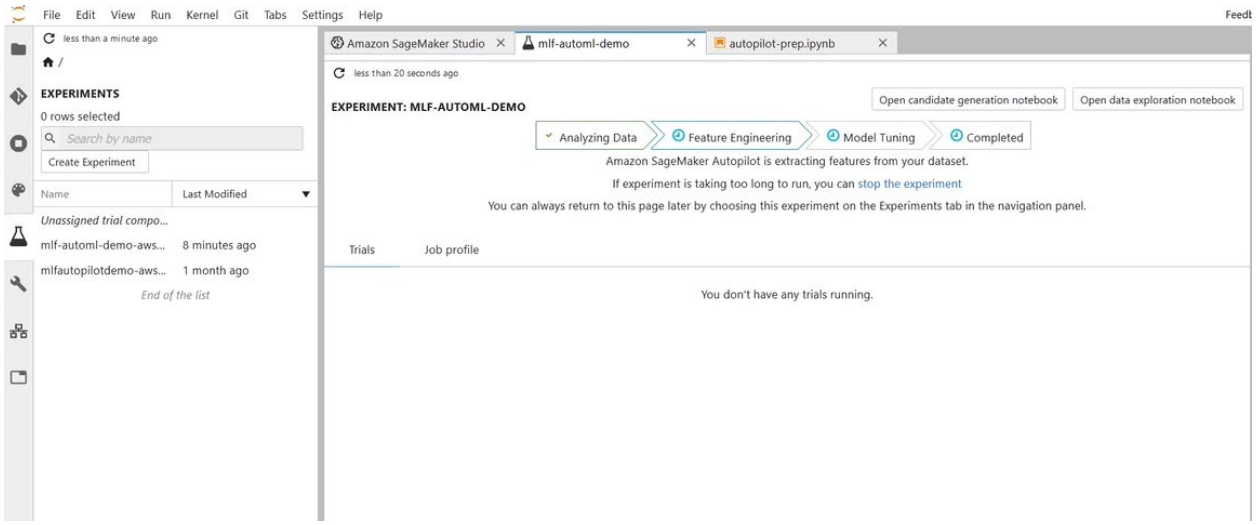
HyperParameter Tuning Optimization Jobs (can be seen in Services Option)



Best Training Job Parameters



AWS SageMaker AutoPilot



Experiments created in SageMaker AutoPilot