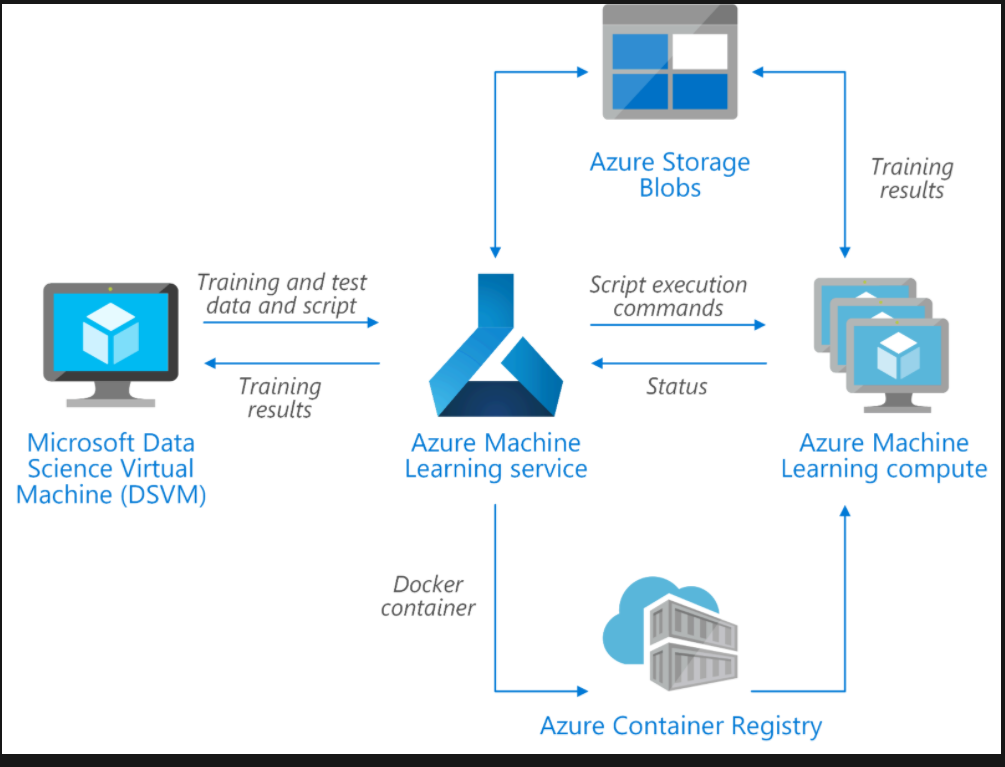
This document contains process flow of each project , Screenshots of their Implementation and Screencast Video

Architectural Diagram : [Reference](https://docs.microsoft.com/en-us/azure/architecture/reference-architectures/ai/training-python-models)



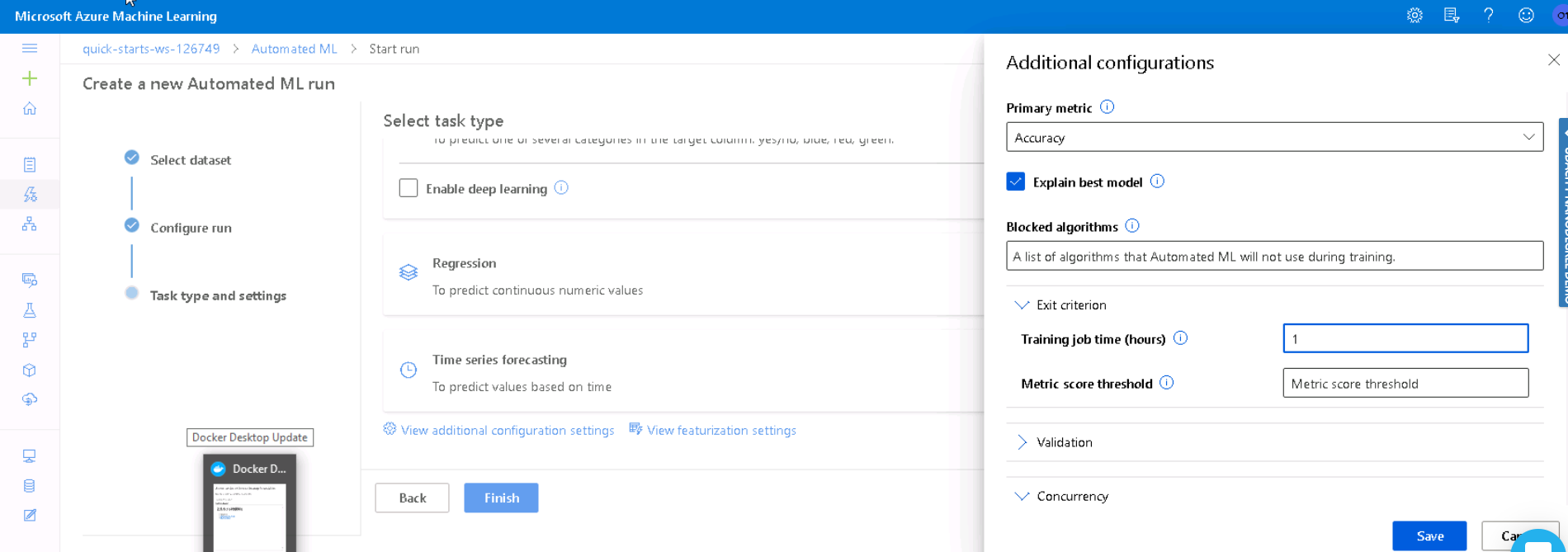
Process Flow Part 1

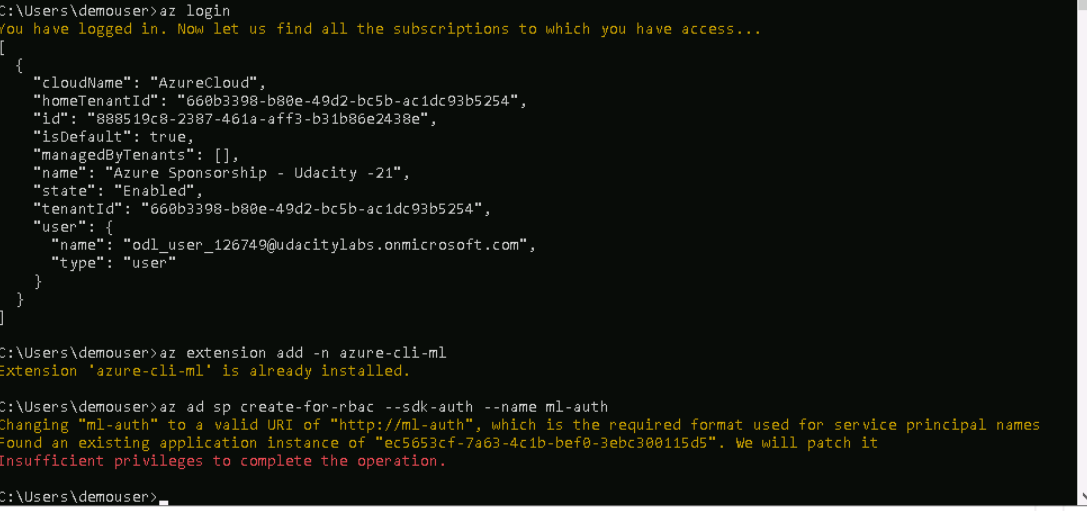
Process Flow Part 2

Steps in Deploying Auto ML Model

Authentication Step:

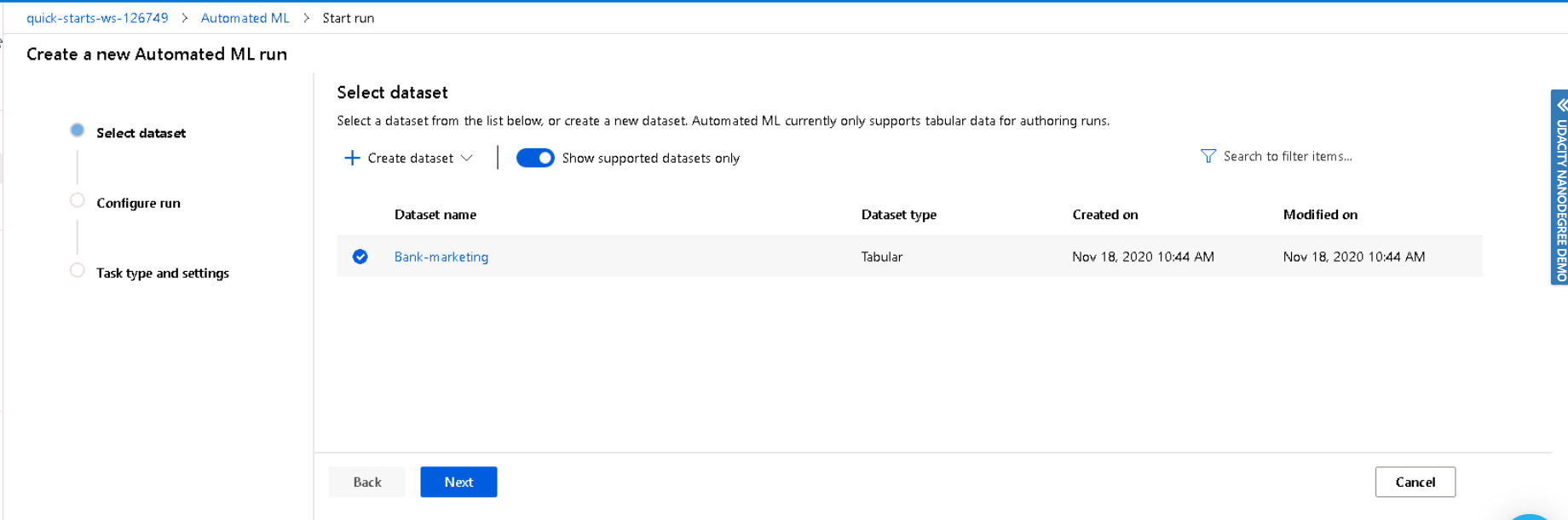
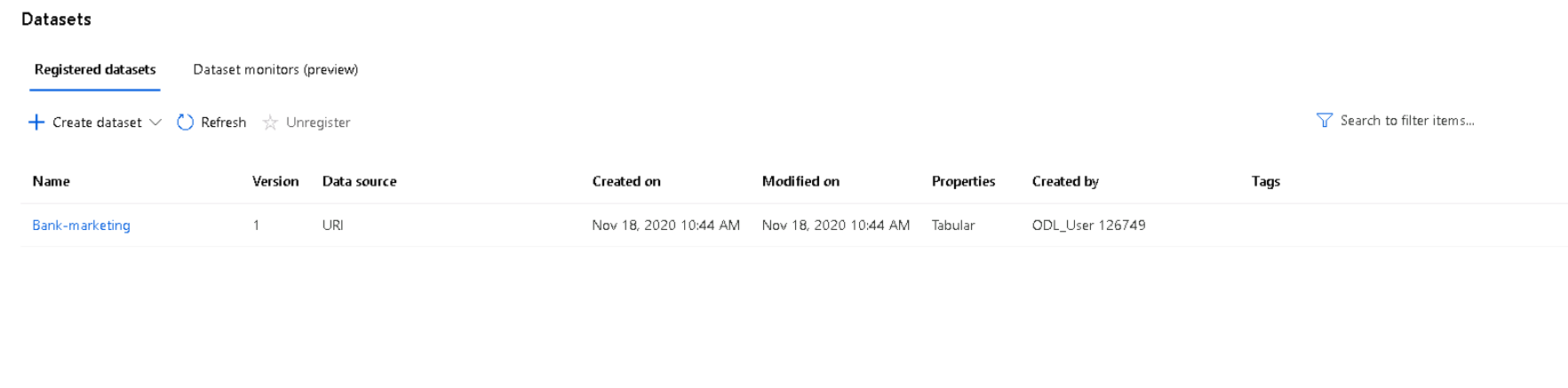
Creating Azure SP Authentication

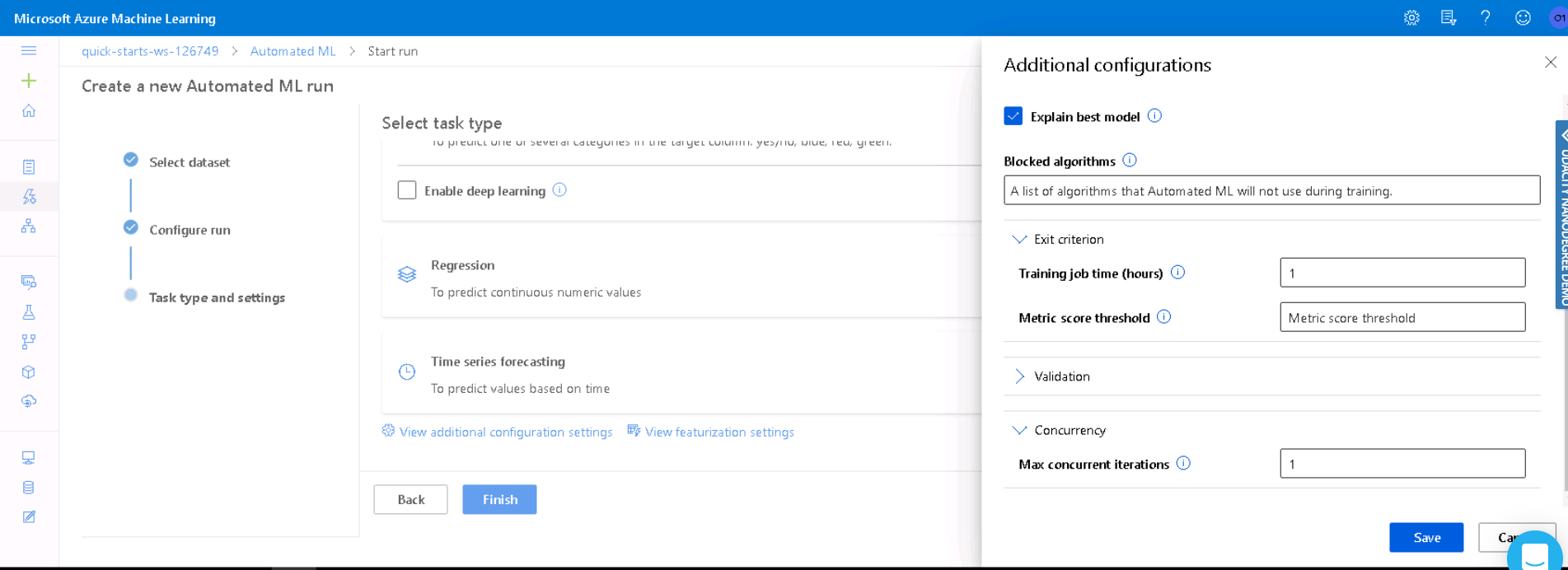




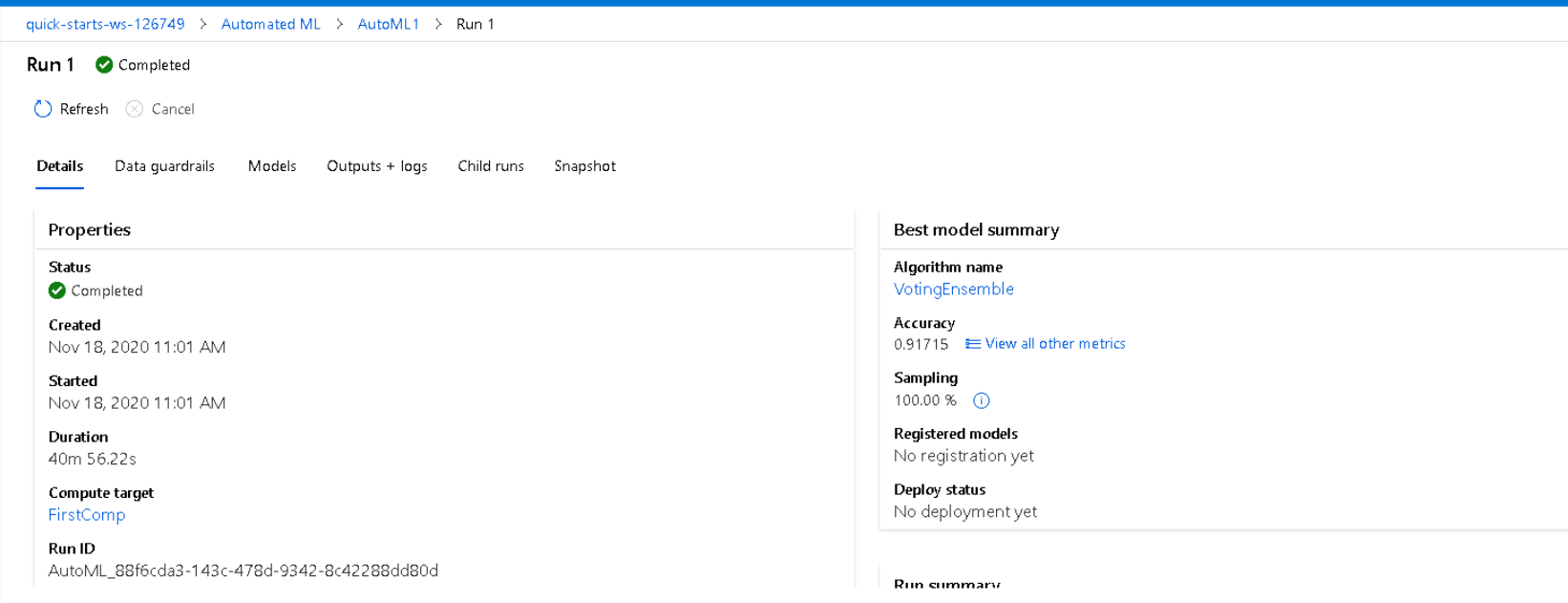
2. Bank Marketing data is considered as the raw dataset. It is registered and it is used in the Auto ML model

Registered Dataset



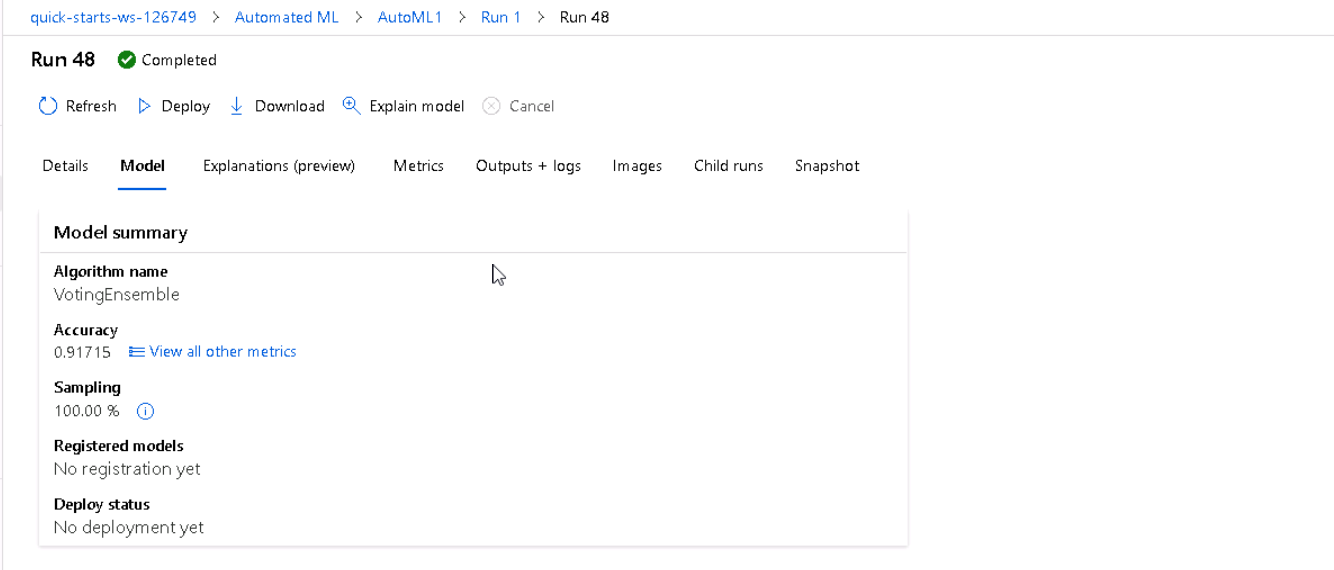


AutoML Completed Status: AutoML has provided an accuracy of 91.7 % and best algorithm is Voting Ensemble

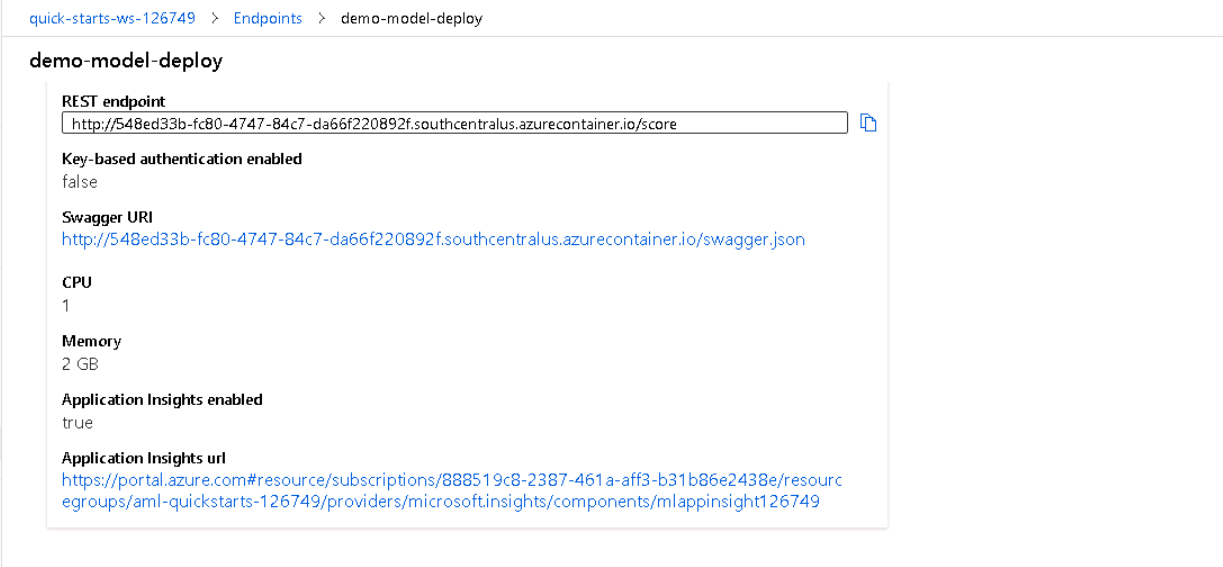


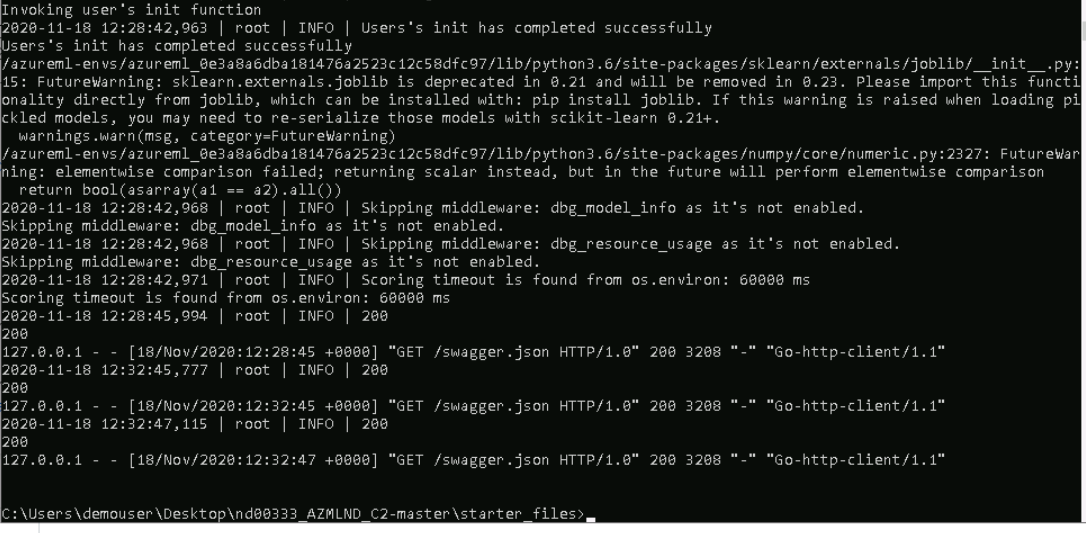
Best Model:

The best model is selected and deployed



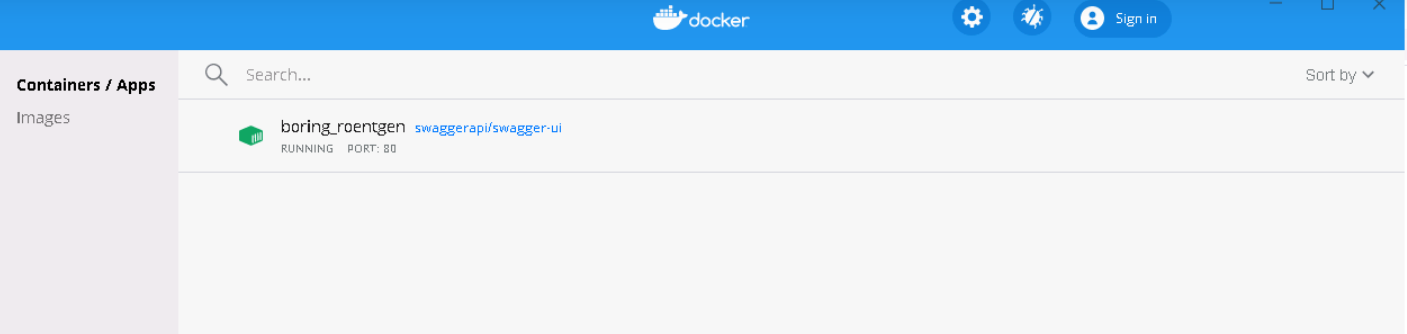
The application Insights has been enabled using the command Service.update() in logs.py file

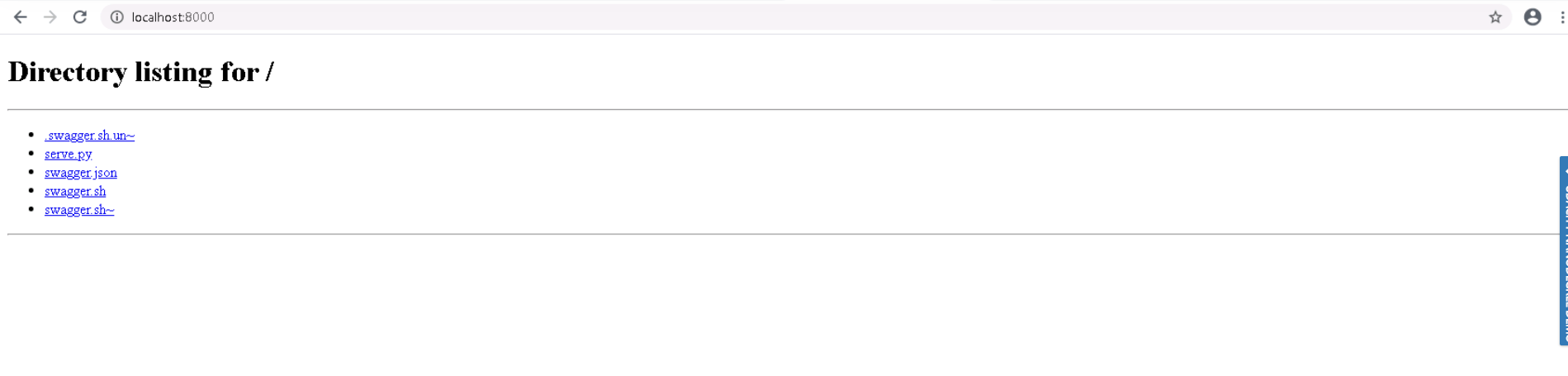


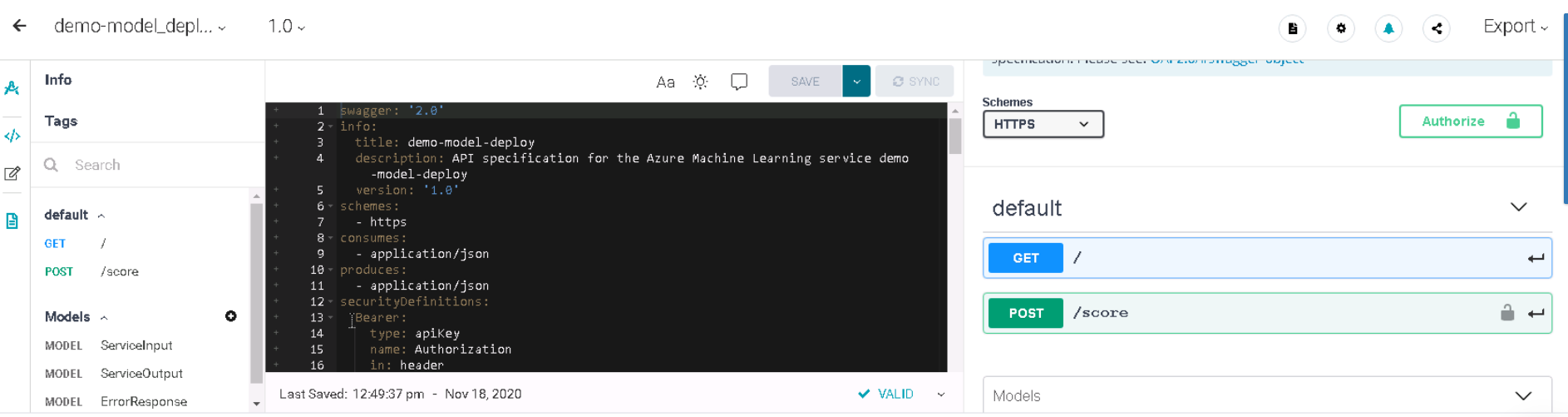


Swagger:

Docker instance is running on port 80 and the files available are given below

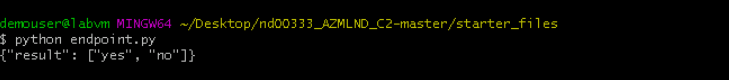




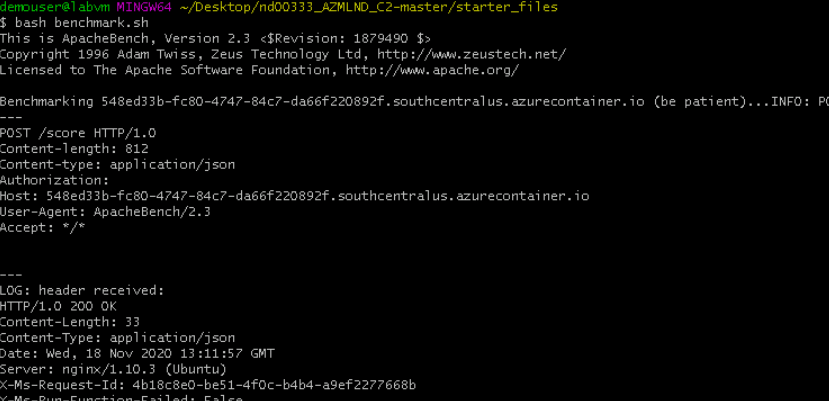


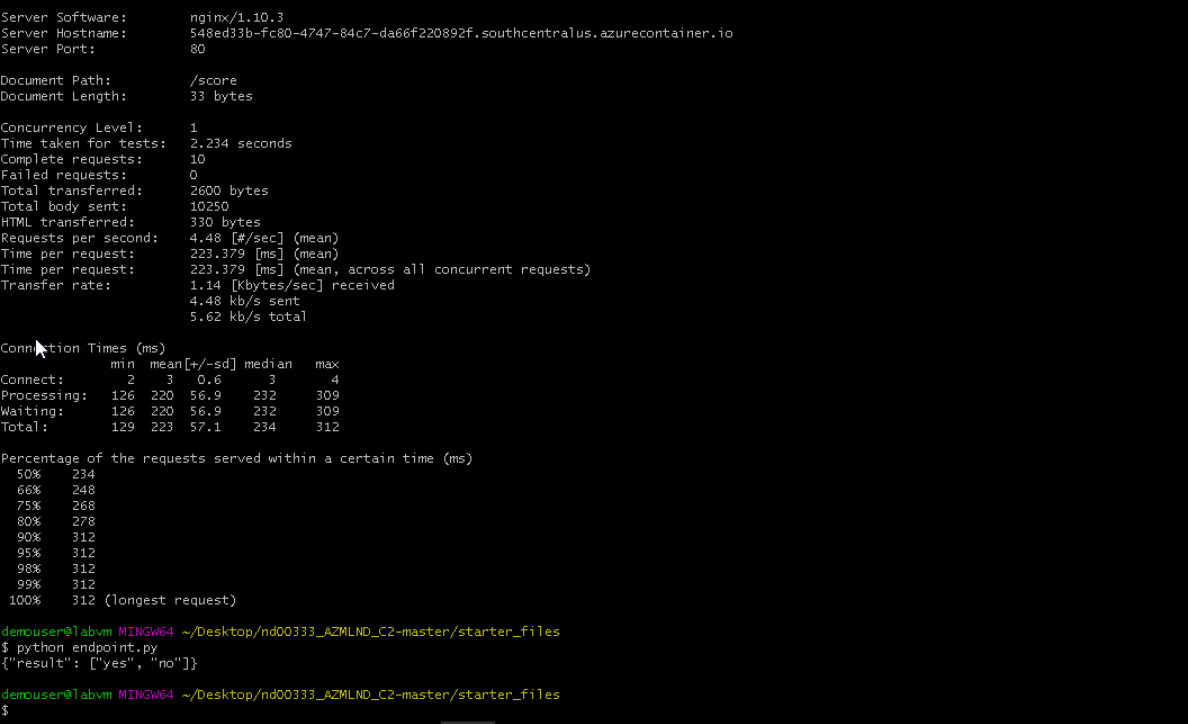
Endpoint Result:

After documenting endpoint using swagger used endpoint.py to post data.json input file and response obtained as response.json file



Benchmark.sh screenshot



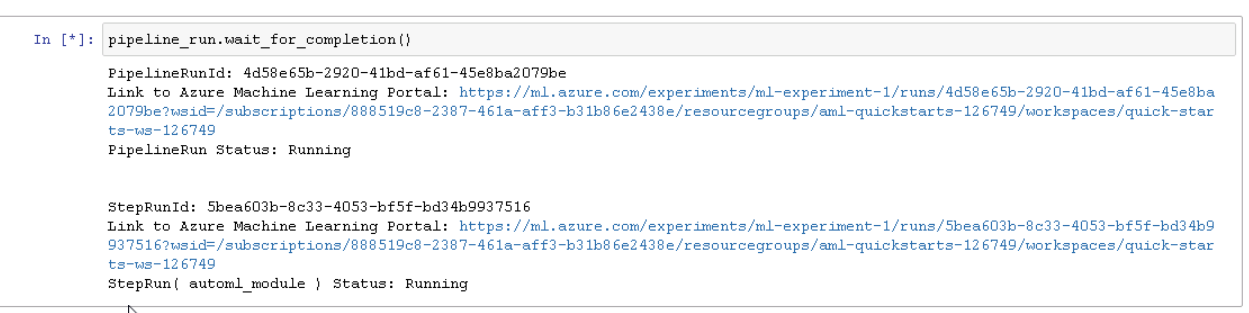


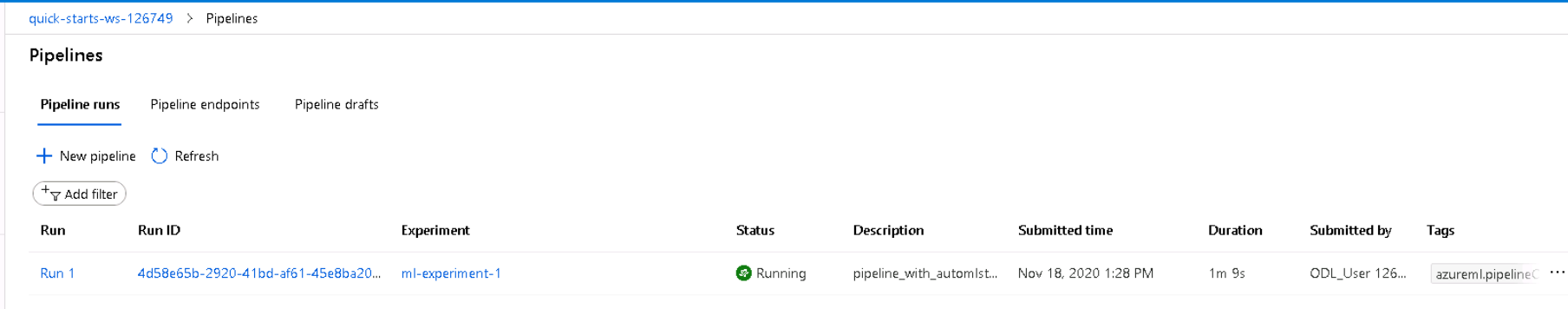
Python SDK Implementation:

After creating new notebook and connecting with the compute instance . Pipelines are built to implement a AutoML model on the same dataset. Pipeline has been configured for AutoML and scoring

After submitting the pipeline the following screen shot is captured displaying the status of pipeline in SDK and designer studio

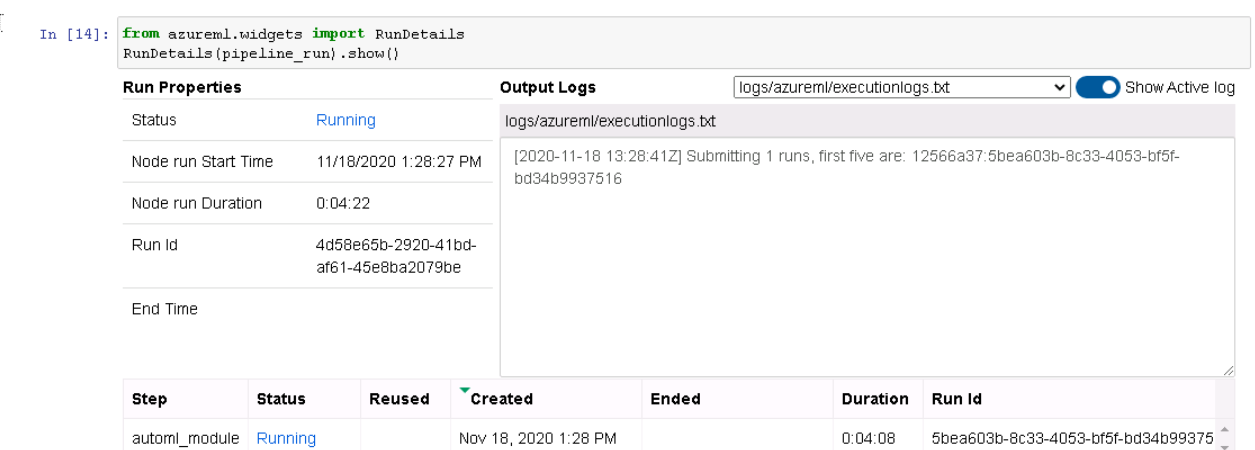
Pipeline Section in AzureML Studio:

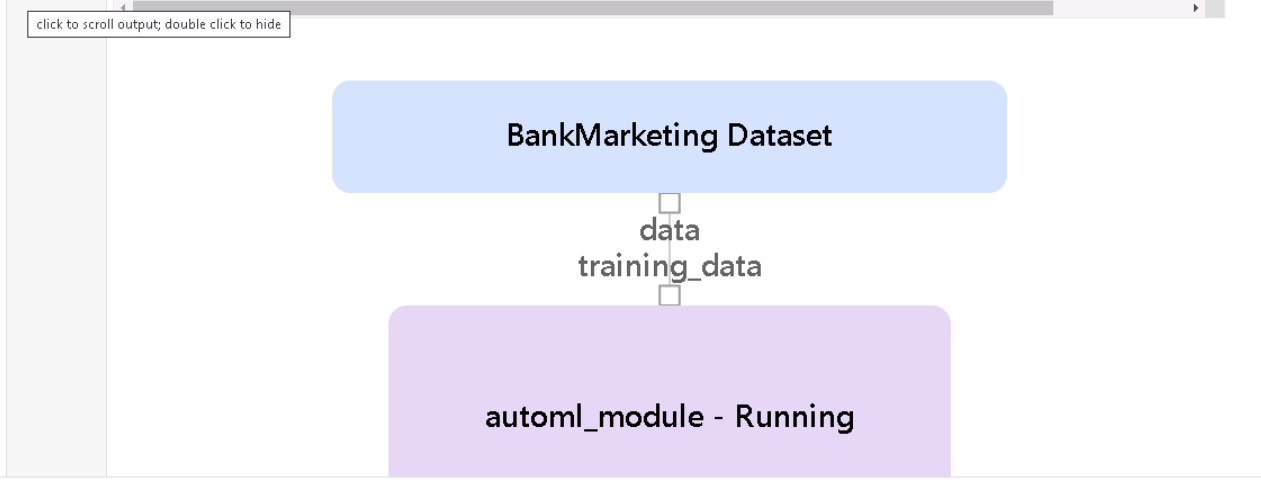




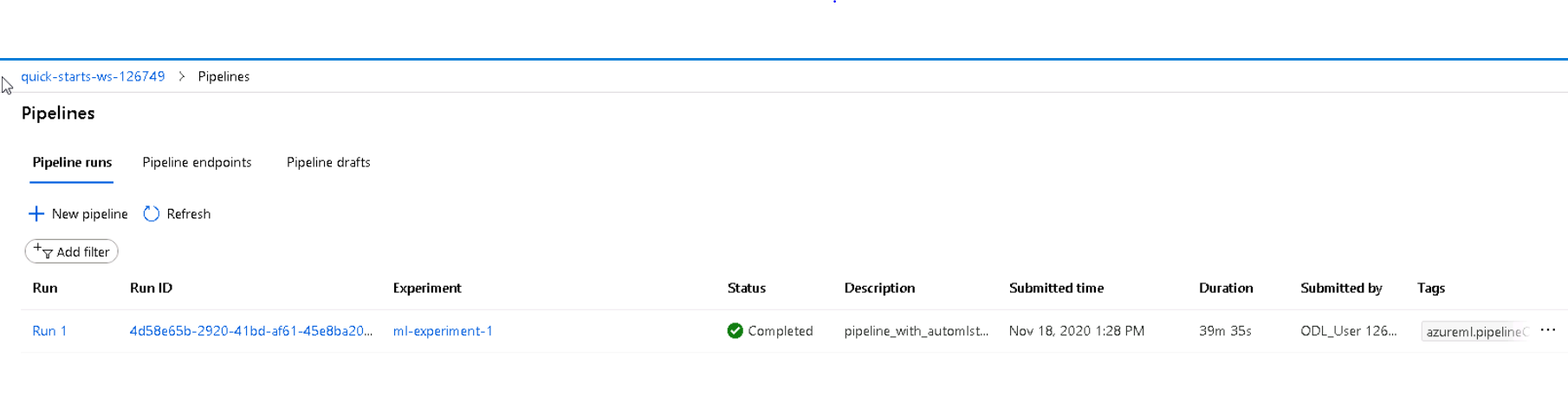
Run Details widget to show step runs :

This widget displays the Runs of the pipeline .

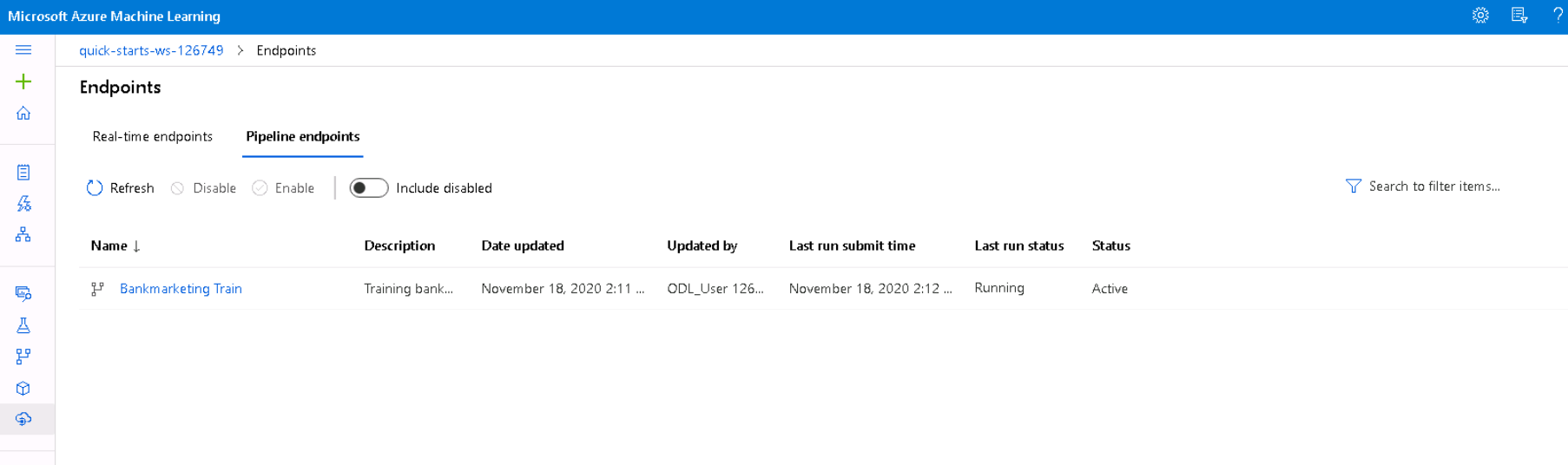


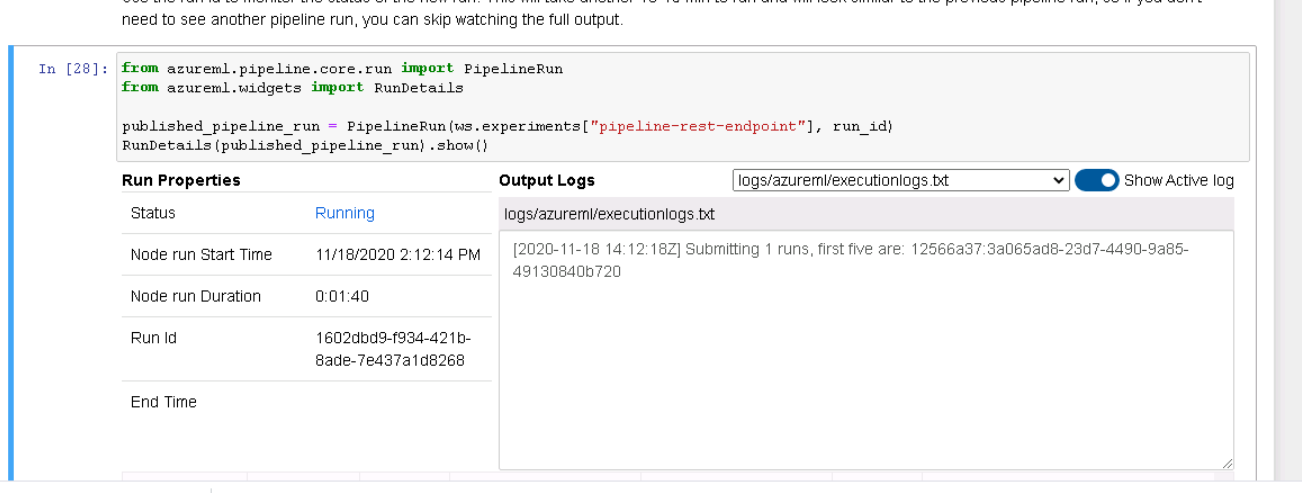


Pipeline Runs in Studio

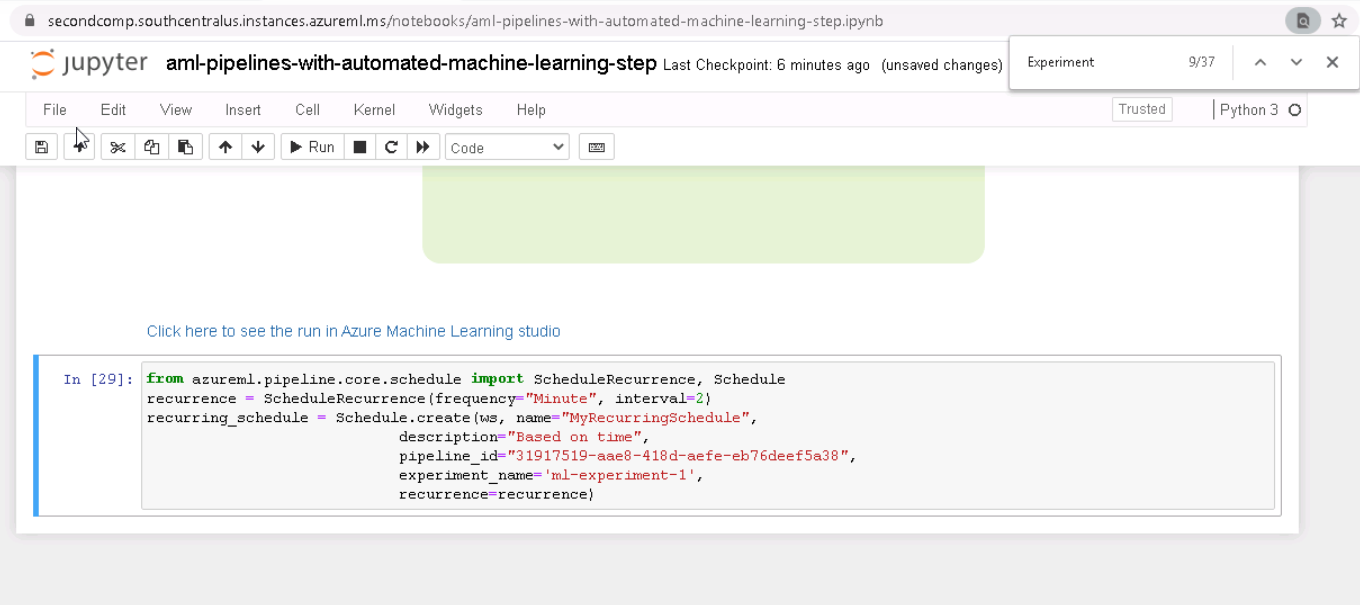


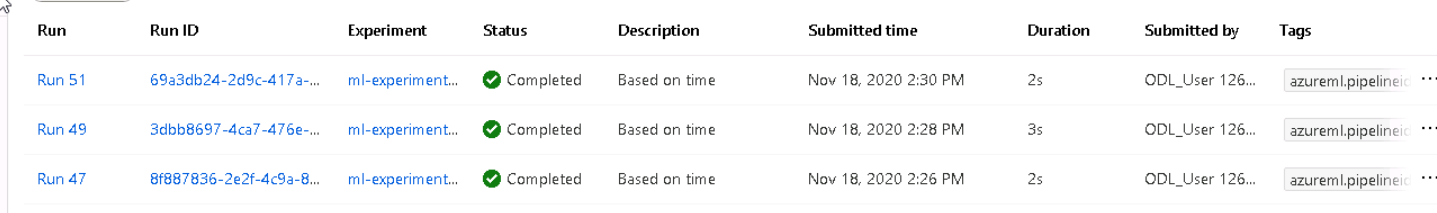
Pipeline endpoints in Studio:





The pipeline runs were scheduled at a frequency of 2 minutes as per the below script and attaching a screen shot on ml studio here

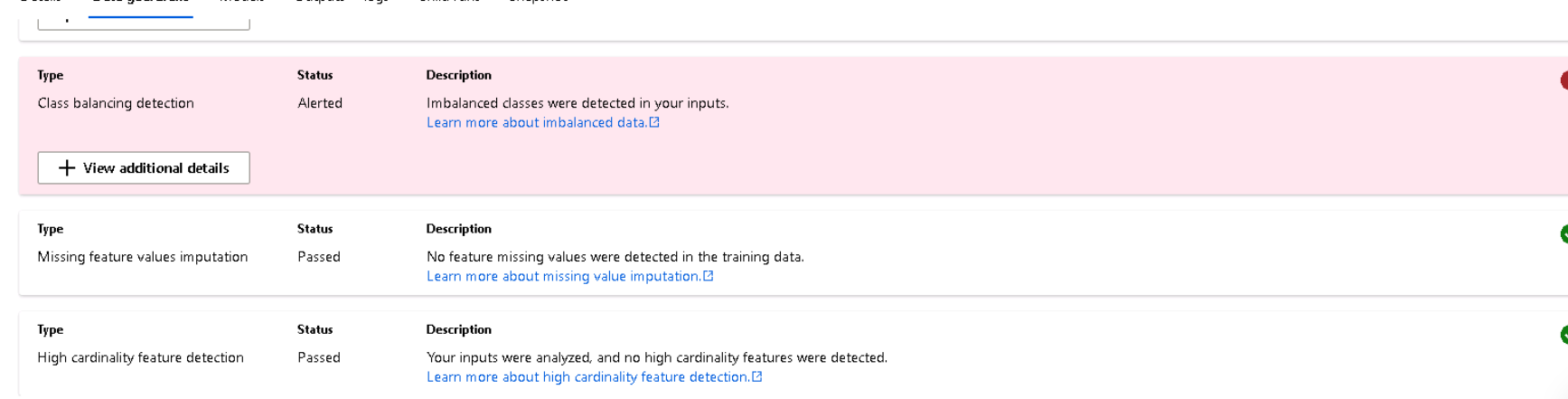




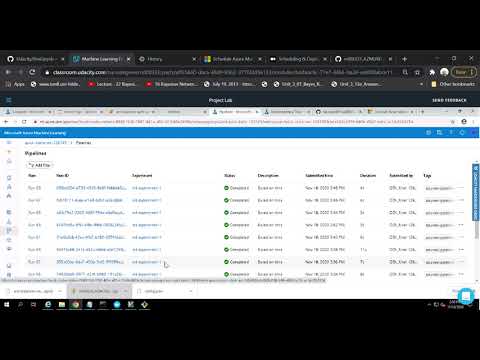
Areas of Improvement :

1.Data Imbalance has been detected as displayed in screenshot . Undersampling and Oversampling can be used to improve the distribution of data

2. During sampling stratified sampling can be used



Screencast Video

[](https://www.youtube.com/embed/-yH1SSQmINk?feature=oembed)