

Machine Learning Regression Model Lab - Data Collection and Curation

Deploy a Regression model through AWS SageMaker studio using Autopilot

<https://dev.to/aws-builders/deploy-a-regression-model-through-aws-sagemaker-studio-using-autopilot-4ng4>

Using Amazon SageMaker Studio and Autopilot, create a Machine Learning model to predict house prices on AWS SageMaker.

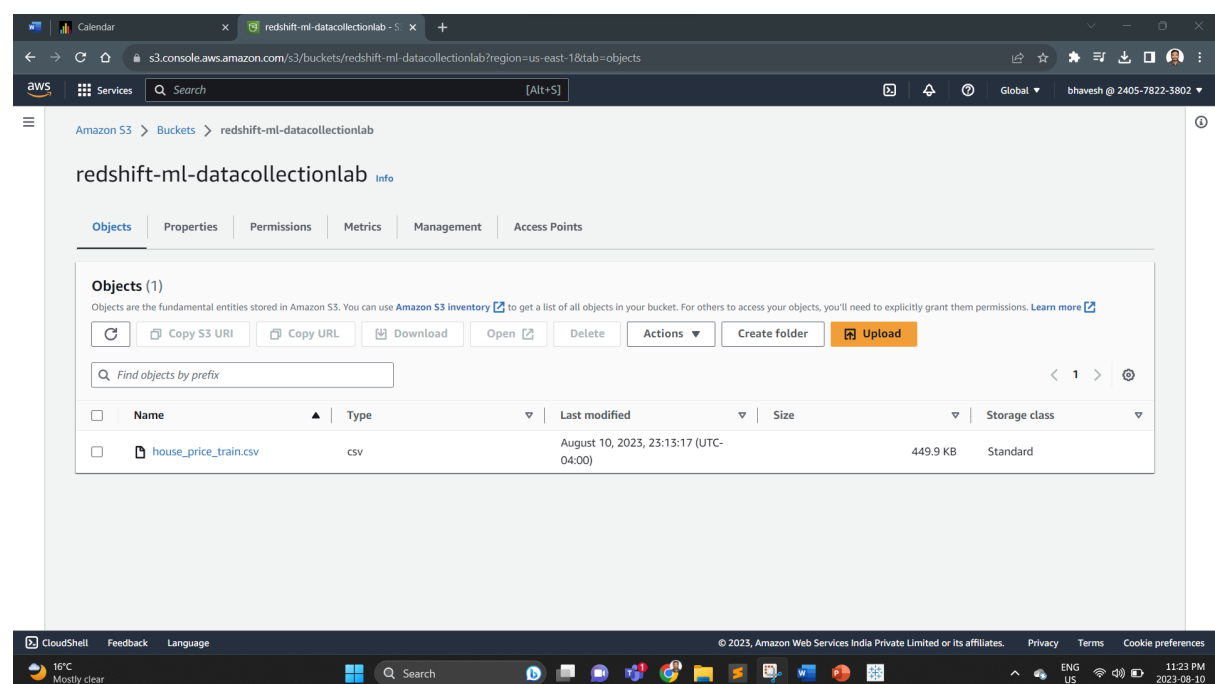
1. Upload dataset to Amazon S3

Signup for the competition to get dataset access and download the dataset from Kaggle:

<https://www.kaggle.com/competitions/house-prices-advanced-regression-techniques/rules>

Uploaded the CSV file for the training dataset into the S3 bucket "redshift-ml-datacollectionlab"

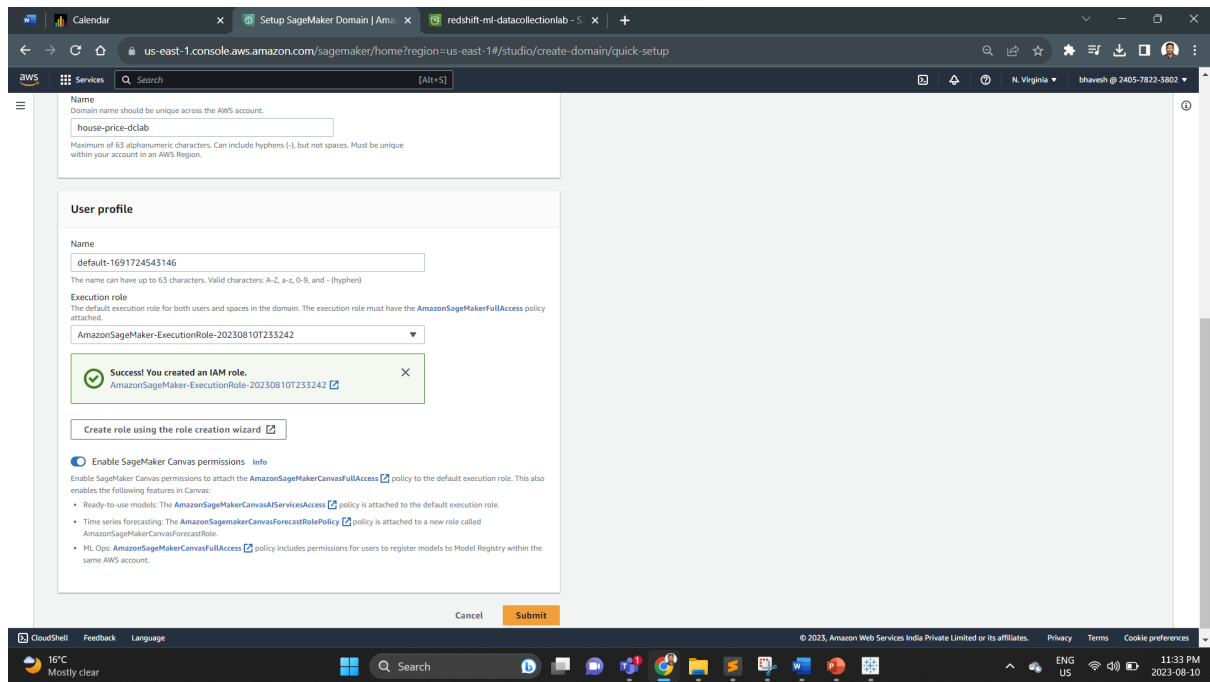
file: house_price_train.csv.



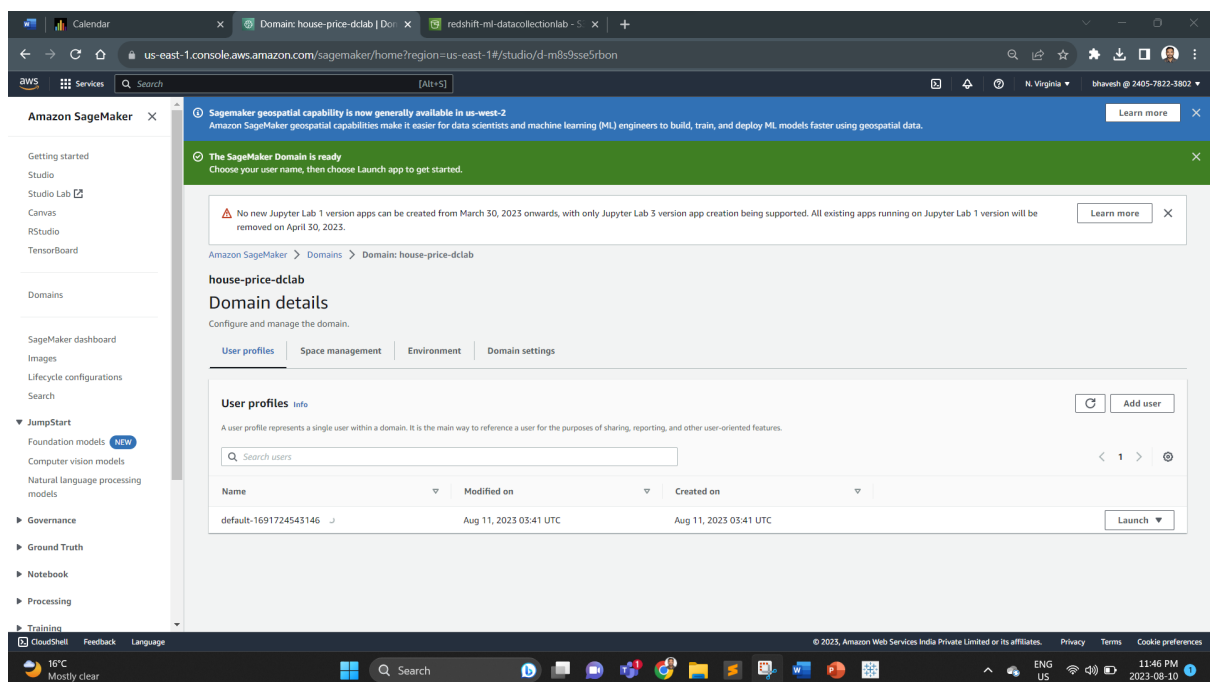
2. Setup SageMaker domain

Inside Amazon SageMaker created a new Domain with the following configurations

In order for SageMaker to access the bucket, an IAM role was created with the bucket name in it.



Domain Profile created successfully.



Once the domain and profile are created, I would be able to launch SageMaker Studio.

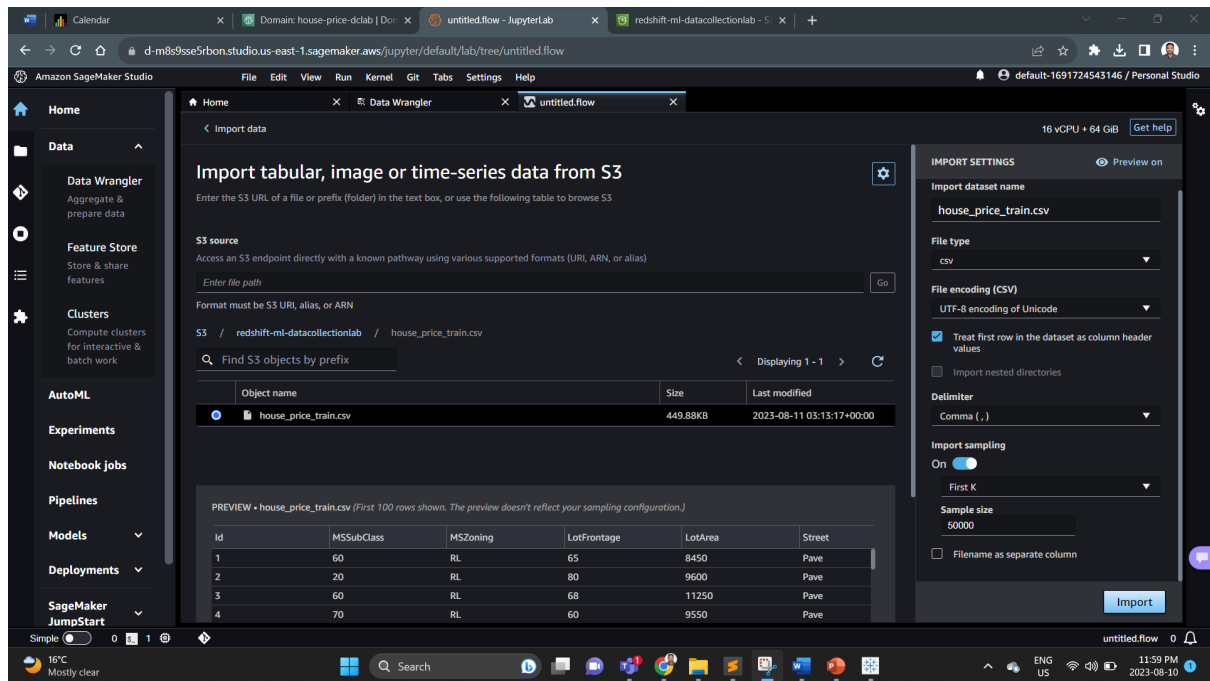
3. Import the dataset on SageMaker and add some preprocessing steps.

Inside the SageMaker Studio Loading the dataset from the S3 bucket is the first step.

Import the Dataset:

1. Click - *Data Wrangler*
2. Click - *Import Data*

3. Data Source: AWS S3
4. bucket selection
5. dataset selection - Click Import



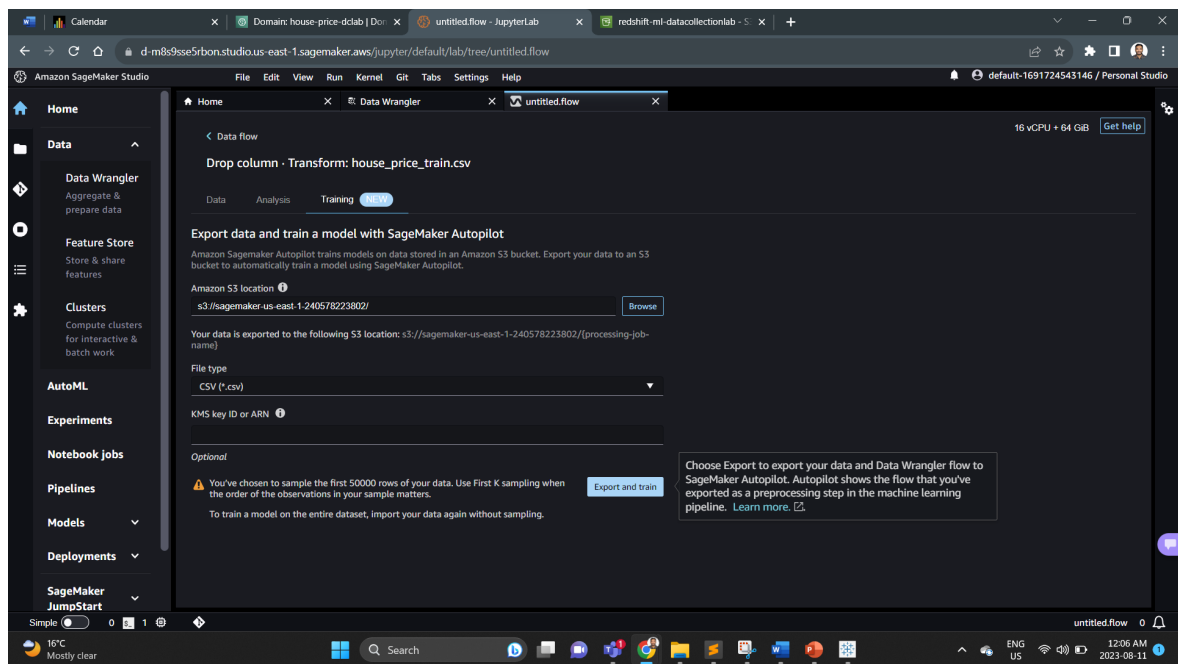
Data Transformation:

Remove column ID from the dataset so that it is not effected on the regression model.

1. Add step
2. search 'remove col' - select Manage Columns
3. Transform: Drop column, Columns to drop: Id.

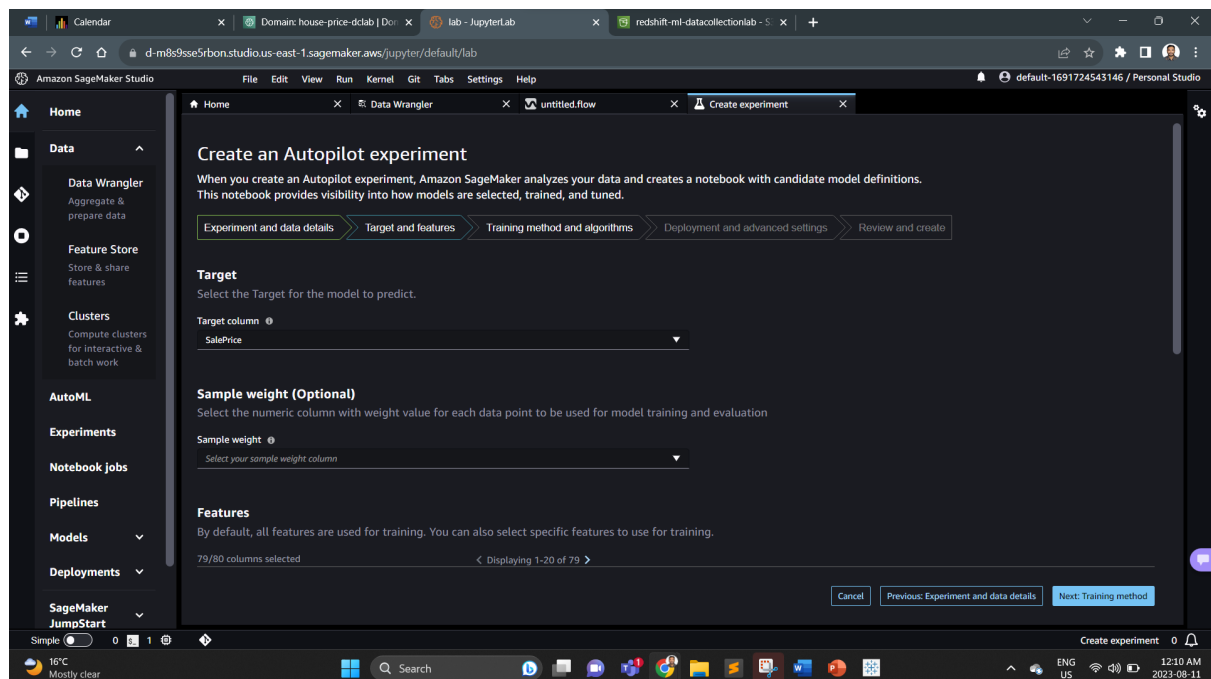
Export and Train:

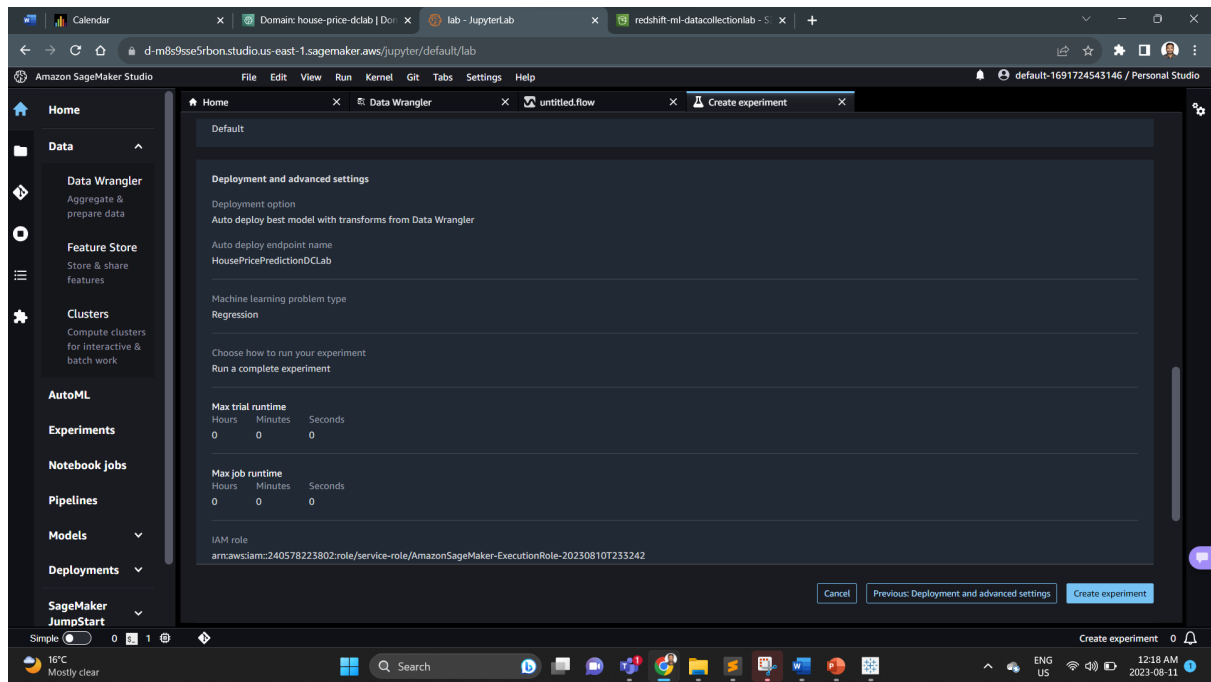
Export and Train the data set to feed to autopilot.



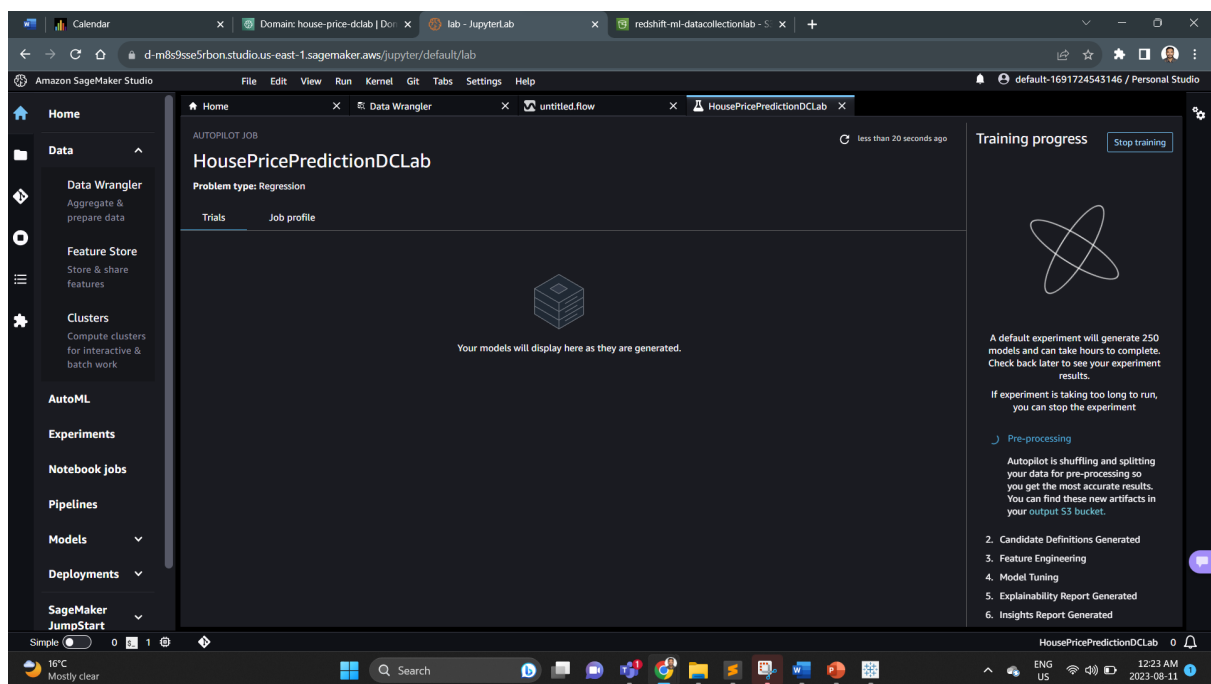
4. Create an Autopilot experiment

1. Name the Experiment: HousePricePredictionDCLab
2. Setting the Target Value as SalePrice and Featured to the remaining columns.
3. Select Training method and algorithms : Auto
4. Select the machine learning problem type: Regression, Objective Matrics: MSE





Training the Model



Experimentation results:

The experiment made 100 Trials and the Best Trail was on the top.

HousePricePredictionDCLab

Problem type: Regression

Trials Job profile

Trial name	Objective: Mse	RMSE	MAE	R2	Start time	Status
HousePricePredictionDCLab2TTPuk-054-11c7b22	635481472	14949.246	24678.584	0.902	24 minutes ago	Completed
HousePricePredictionDCLab2TTP...	47800127488	217674.922	206370.922	-6.873	26 minutes ago	Completed
HousePricePredictionDCLab2TTP...	42340241408	205767.438	181534.438	-6.852	18 minutes ago	Completed
HousePricePredictionDCLab2TTP...	41604427776	203971.641	182541.922	-5.394	26 minutes ago	Completed
HousePricePredictionDCLab2TTP...	41164783616	202891.063	177360.109	-4.976	25 minutes ago	Completed
HousePricePredictionDCLab2TTP...	37280239616	192978.5	175080.016	-5.155	26 minutes ago	Completed
HousePricePredictionDCLab2TTP...	33576947712	183198.703	167357.219	-4.369	20 minutes ago	Completed
HousePricePredictionDCLab2TTP...	31851958272	178394.328	162821.453	-4.185	19 minutes ago	Completed

100 results Results are cached Refresh Rows per page: 8 Go to page: 1 Page 1 of 13

Job Profile sharing the details of the experiment

HousePricePredictionDCLab

Problem type: Regression

Trials Job profile

Name: HousePricePredictionDCLab

Creation time: 54 minutes ago

Last updated: 7 minutes ago

End time: 7 minutes ago

ARN: arn:aws:sagemaker:us-east-1:240578223802:autoaml-job/HousePricePredictionDCLab

Role ARN: arn:aws:iam::240578223802:role/service-role/AmazonSageMaker-ExecutionRole-202308101723342

Problem type: Regression

Status: Completed

Secondary status: Completed

Failure reason: —

Job objective metric name: MSE

Generate candidate definitions only: —

Input data config

compressionType	targetAttributeName	sampleWeightAttributeName	s3DataType	s3URI
—	SalePrice	—	ManifestFile	s3://sagemaker-us-east-1-240578223802-jwpep-1691726814/manifest/data_wangler_output.manifest

Output data config

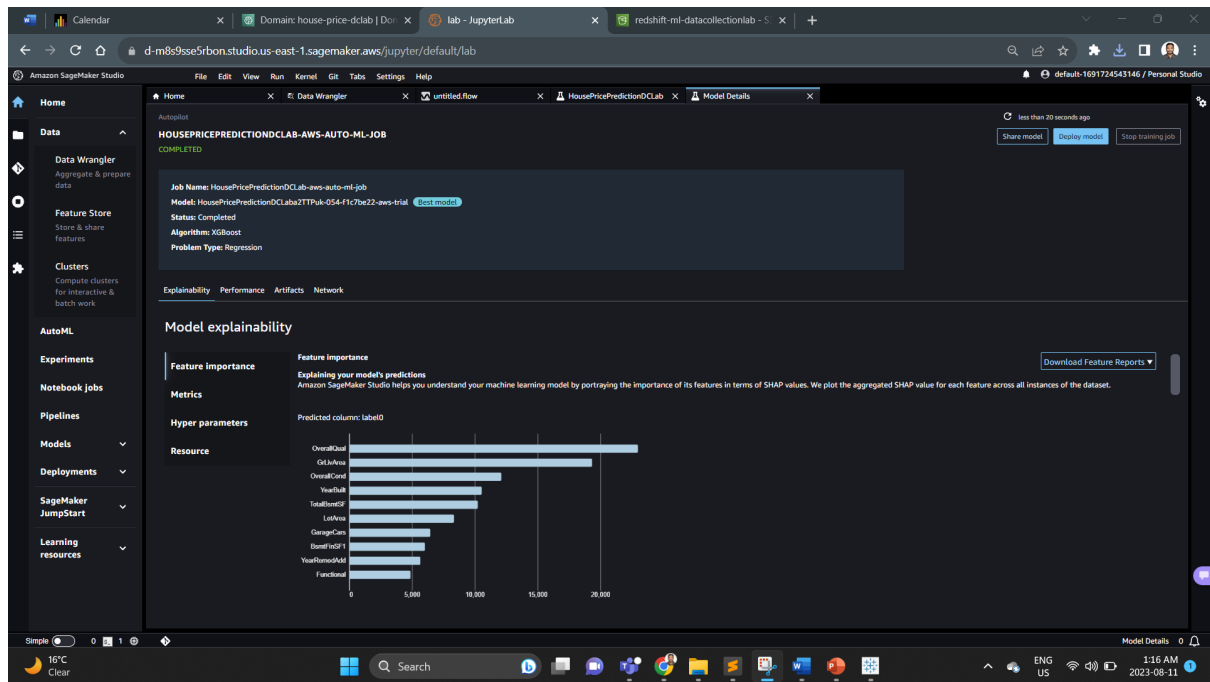
KMS key ID	S3 output path
—	s3://sagemaker-us-east-1-240578223802-output-1691726814/

Resolved attributes

Objective	Problem type	Max candidates	Max runtime per training job in seconds	Max AutoML Job Runtime in seconds
MSE	Regression	250	86400	2550000

An error occurred rendering table

View Best Trail Model Details (report as a pdf attached)

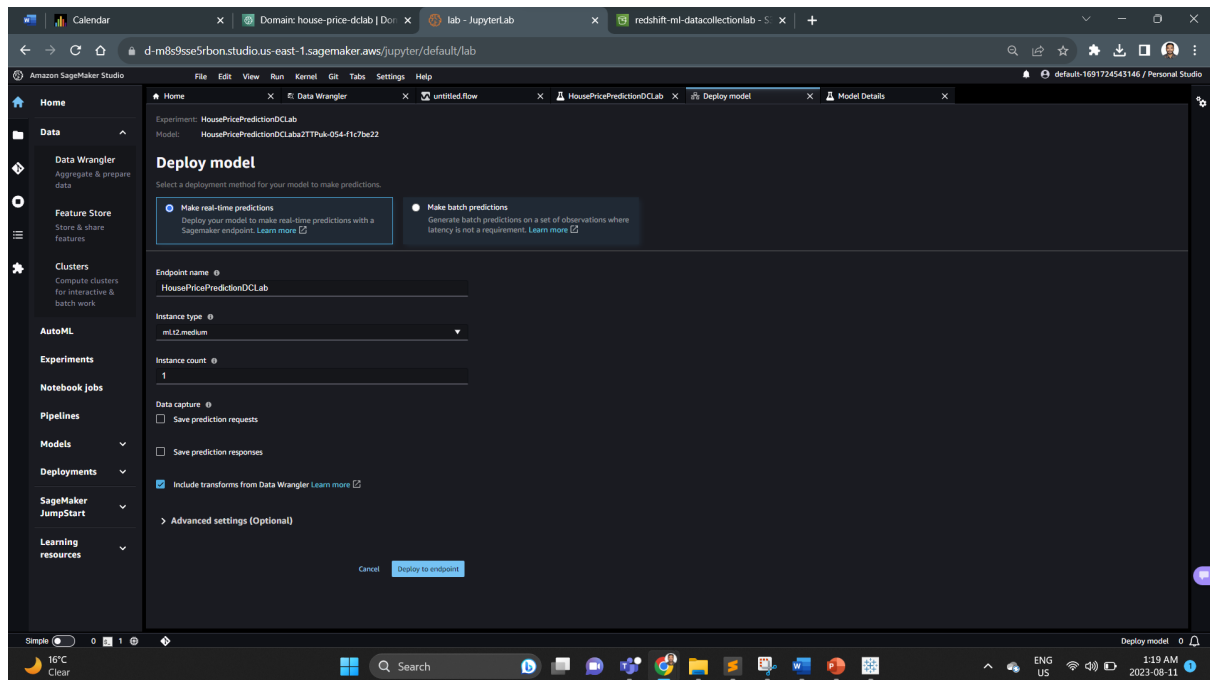


5. Deploy ML model

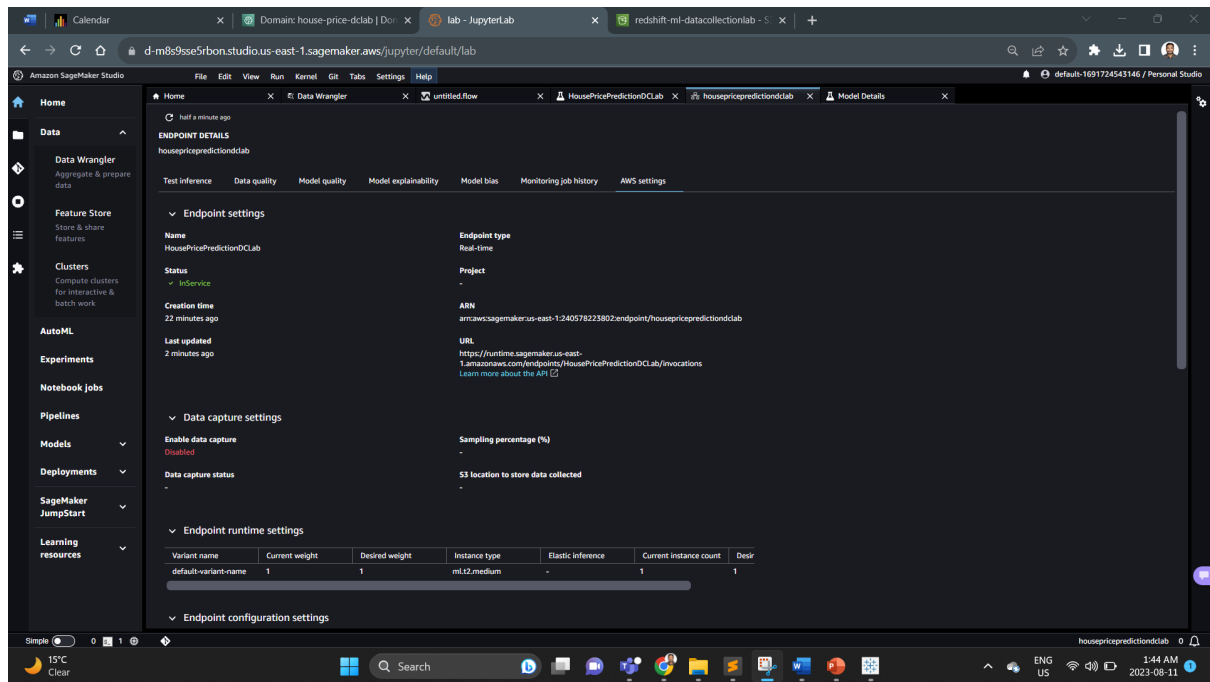
Deploy the best model that was created out of the experiment.

Deployment Configuration:

Endpoint Name, Instance type, Instance count, Data Capture.



I Waited until the endpoint is deployed. In my case, it took 23 minutes to finish the deployment.



6. Invoke deployed ML model

I found the deployed URL inside AWS Settings Tab which is:

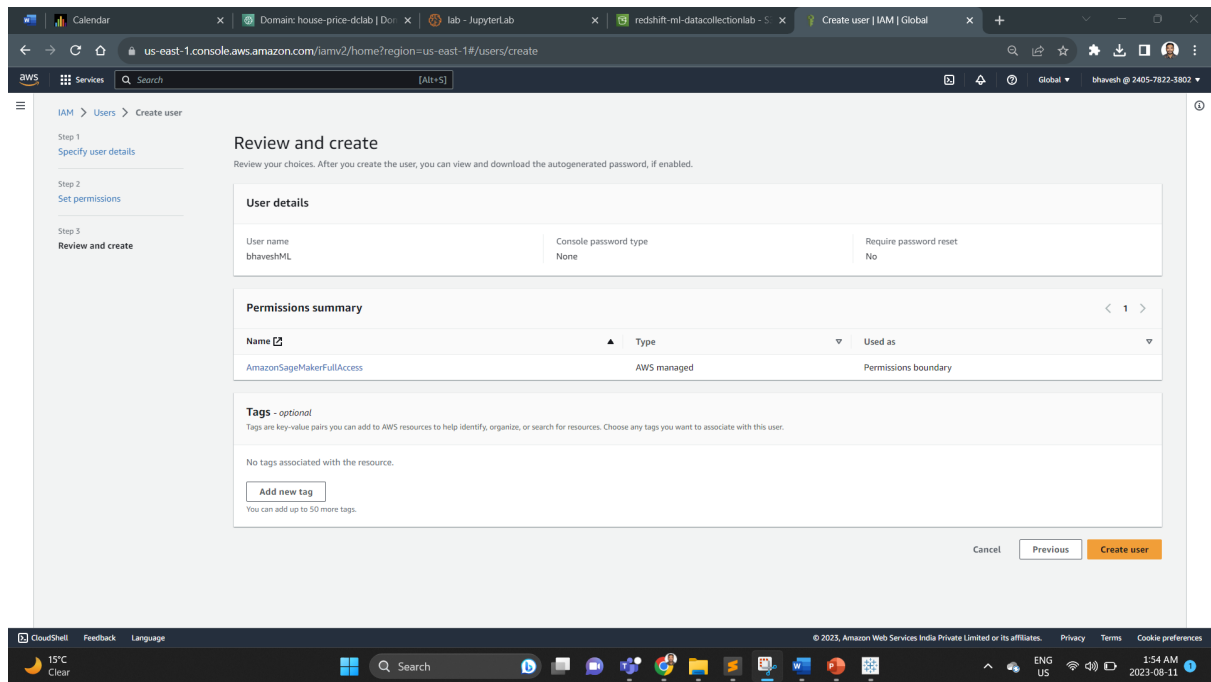
<https://runtime.sagemaker.us-east-1.amazonaws.com/endpoints/HousePricePredictionDCLab/invocations>

In order to invoke the URL, I need to have Access Key and a Secret Key.

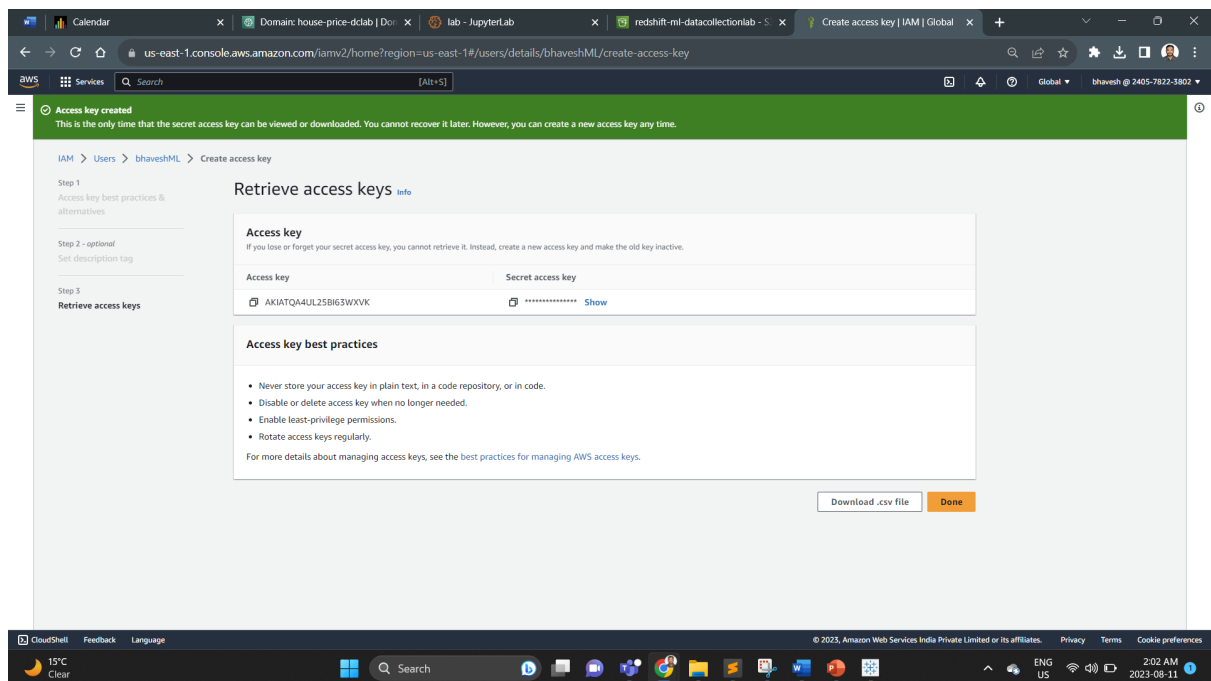
For this purpose, I create an IAM User and assign the AmazonSageMakerFullAccess permission policy.

To do that, follow these steps:

- I went to IAM Management Console
- Users
- Clicked on *Add Users*
- Typed the username: bhaveshML
- In the *Set permissions* step, I checked *Use a permissions boundary to control the maximum permissions* and search AmazonSageMakerFullAccess in the permissions policies list.
- Reviewed the details for the user and click on *Create user*.



Creating Access Key and Secret Access Key to access the Model Endpoint from Postman.

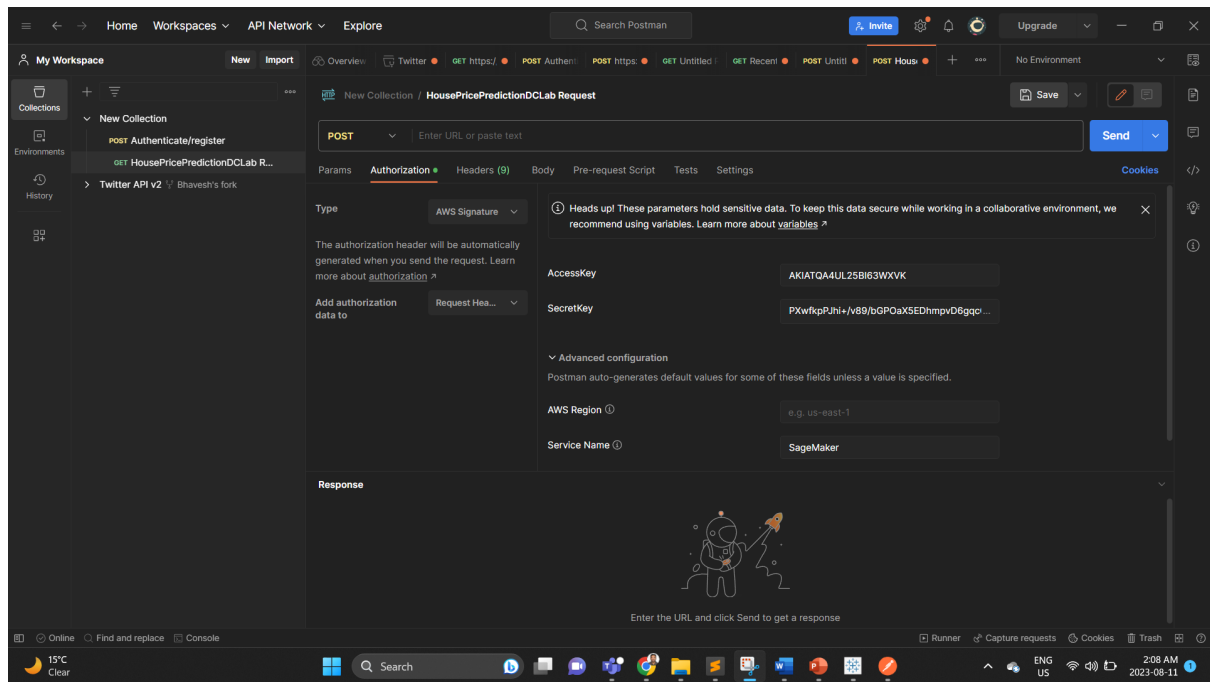


Invoke the URL from Postman.

I Select POST as the HTTP method.

In the *Authorization* tab, selected *AWS Signature* as the type.

Paste the access key and secret key. Also added the *Service Name: SageMaker*.



In the *Headers* tab, changed the *Content-Type* to *text/csv*.

In the *Body* tab, select raw and Text.

Paste 1 record from the test dataset. Removed the 'id' and 'SalePrice' from the record.

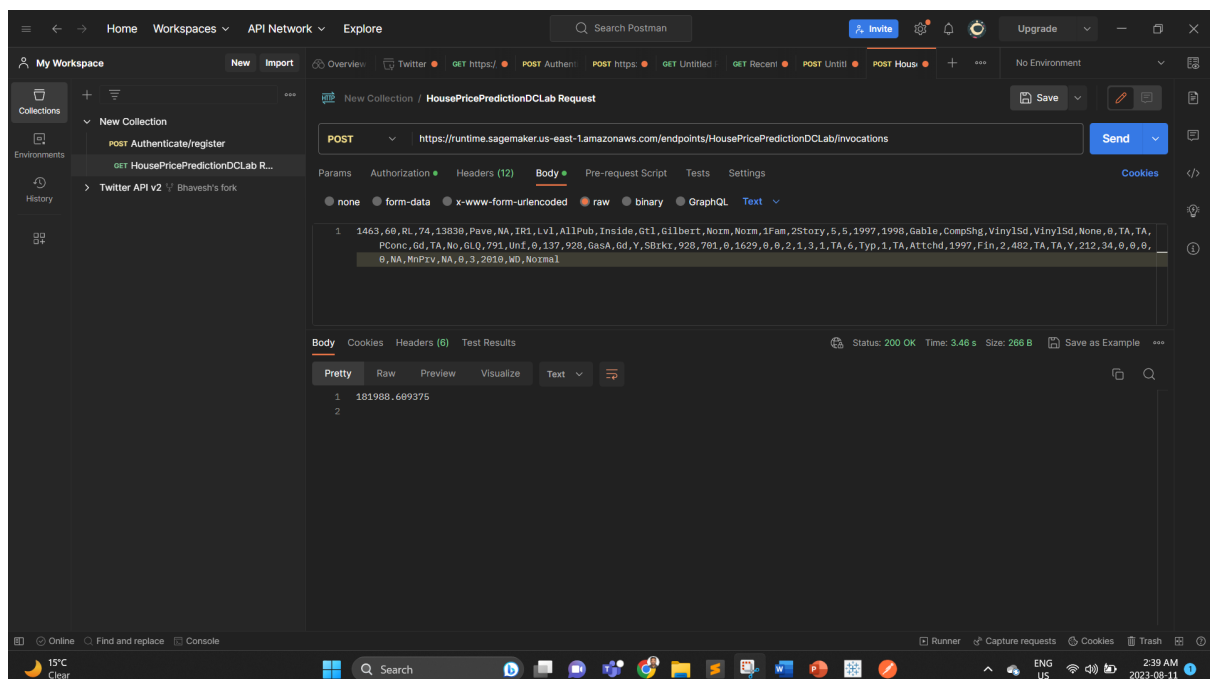
Send the request to the SageMaker Deployed Regression Model Endpoint to fetch the sale price.

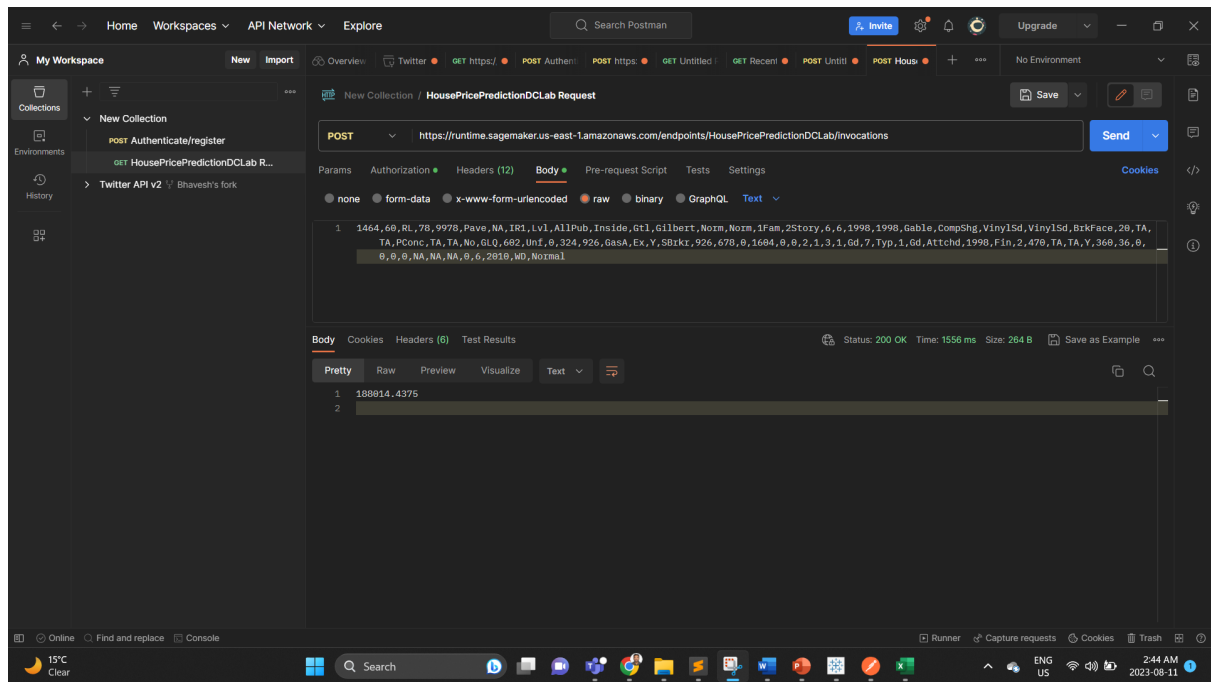
Predicted House Price Based on Test Dataset with 2 example.

Testing Input 1:

"1463,60,RL,74,13830,Pave,NA,IR1,Lvl,AllPub,Inside,Gtl,Gilbert,Norm,Norm,1Fam,2Story,5,5,1997,1998,Gable,CompShg,VinylSd,VinylSd,None,0,

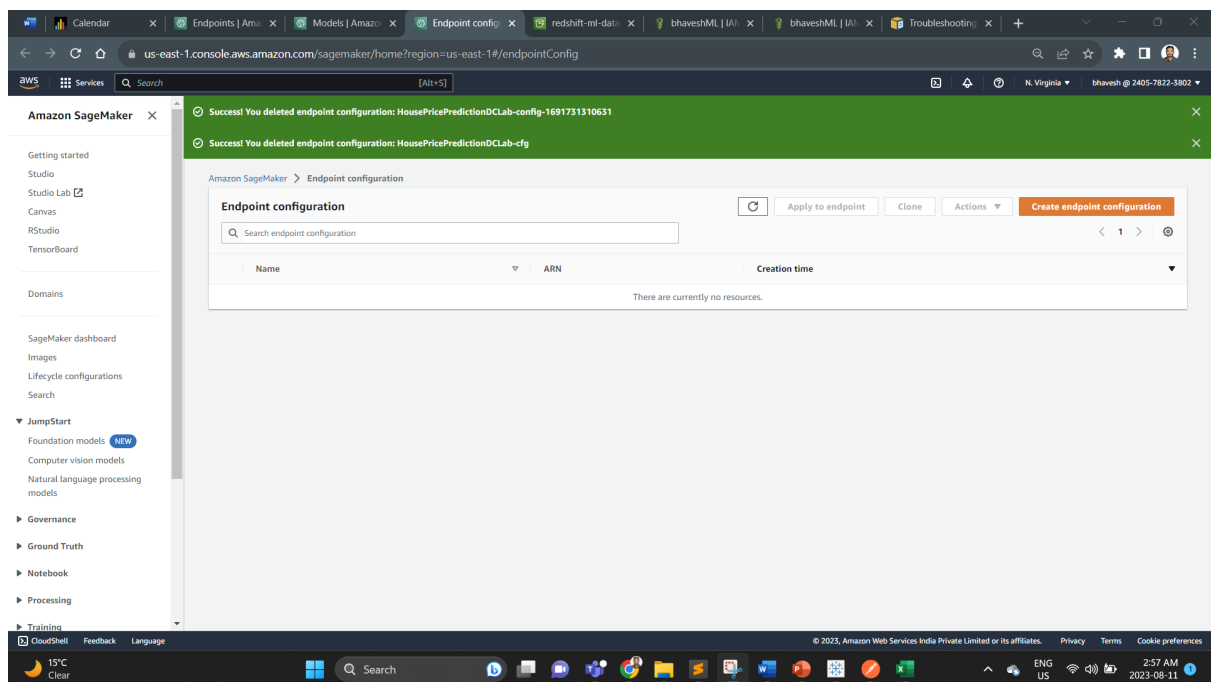
Predicted SalePrice: 181988.609375



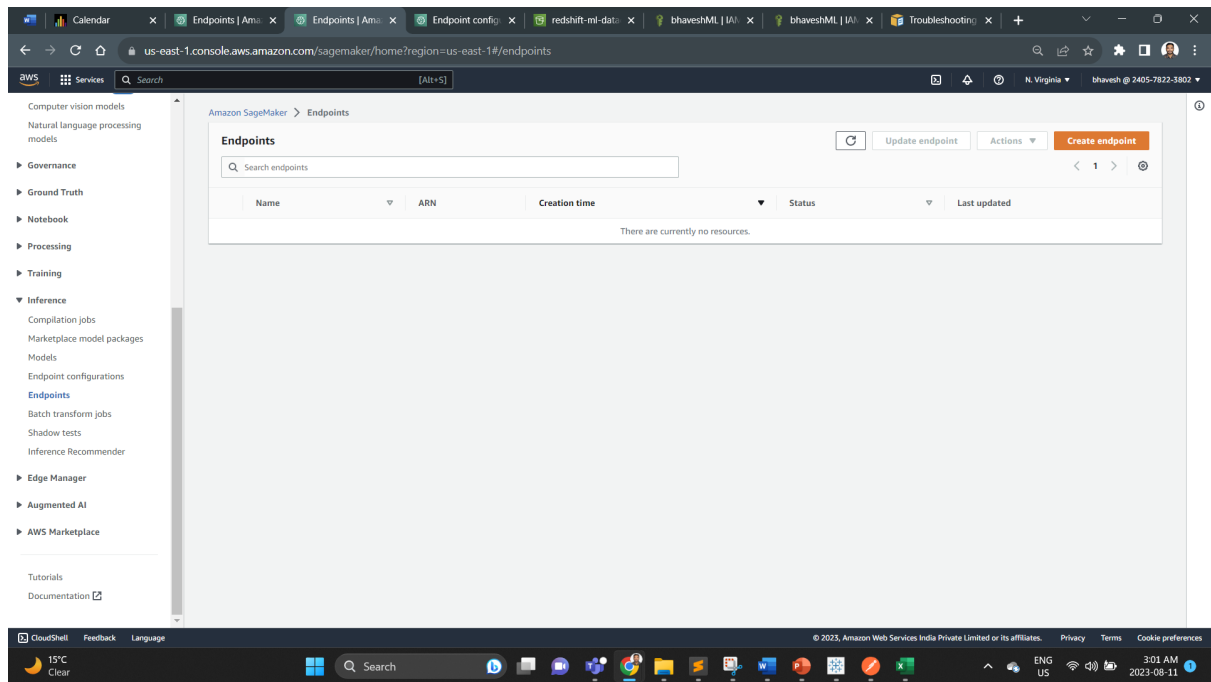


7. Clean up the Resources utilized

Delete Endpoint-configuration



Delete Endpoint



Delete Predictive Models Created by Auto Pilot

