I used Google Cloud Project.

The process was not clear until I found the quickstart guide and used that to run an ARIMA model on a subset of the stock data. There was one stock in the subset that seemed to throw off the model as it had a 197% error rate, but once removed, the remaining data had approximately a 14% error rate.

The quickstart guide is here (replace ‘logistic-reg’ with ‘arima-plus’):

<https://cloud.google.com/bigquery-ml/docs/create-machine-learning-model>

I had to use 2 scripts. One to create the model:

#standardSQL

CREATE MODEL `dailyPrices.sample\_model`

OPTIONS(MODEL\_TYPE = 'ARIMA\_PLUS',

TIME\_SERIES\_TIMESTAMP\_COL = 'dateString',

TIME\_SERIES\_DATA\_COL = 'adj\_close',

TIME\_SERIES\_ID\_COL = 'ticker\_symbol',

DATA\_FREQUENCY = 'DAILY') AS

SELECT

adj\_close,

cast(recordDate AS date) AS dateString,

ticker\_symbol

FROM

`subtle-bit-348822.dailyPrices.price`

WHERE

recordDate > CAST('1980-01-01' as date) AND recordDate < CAST('2021-01-01' as date)

ORDER BY ticker\_symbol, recordDate;

And one to evaluate it:

#standardSQL

SELECT

\*

FROM

ML.EVALUATE(MODEL `dailyPrices.sample\_model`, (

SELECT

adj\_close,

cast(recordDate AS date) AS dateString,

ticker\_symbol

FROM

`subtle-bit-348822.dailyPrices.price`

WHERE

recordDate > CAST('2021-01-01' as date) AND recordDate < CAST('2022-01-01' as date)

ORDER BY ticker\_symbol, recordDate));