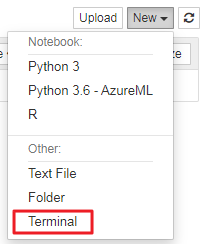
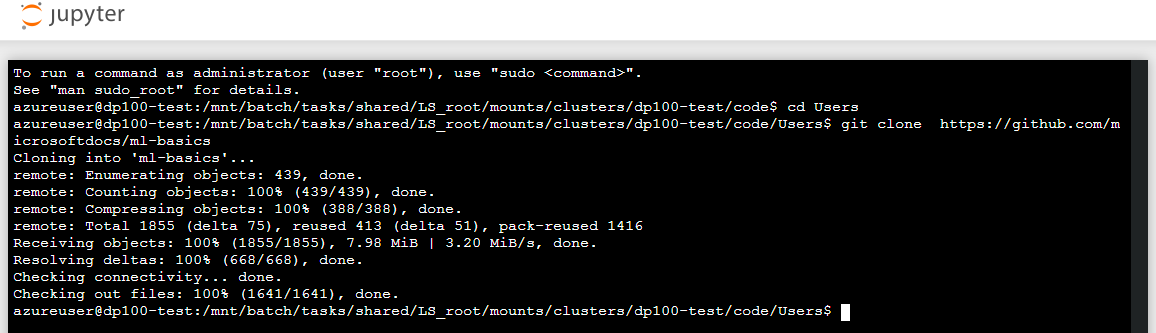
DP100-Exercise











210104

【python get series index name】: .name

【python multiple filters in dataframe】: &

【python type check】: isinstance()

【python type check numpy int64】: np.integer

【python group by ascending】: .nlargest(int)

【python dataframe correlation get nan】

【python get row with max value in column】: .argmax()

<https://stackoverflow.com/questions/10202570/find-row-where-values-for-column-is-maximal-in-a-pandas-dataframe>

210120(三)

【python standardscaler dataframe】

<https://stackoverflow.com/questions/35723472/how-to-use-sklearn-fit-transform-with-pandas-and-return-dataframe-instead-of-num>

【python dataframe delete column by name】

<https://stackoverflow.com/questions/13411544/delete-column-from-pandas-dataframe>

【python dataframe concat】

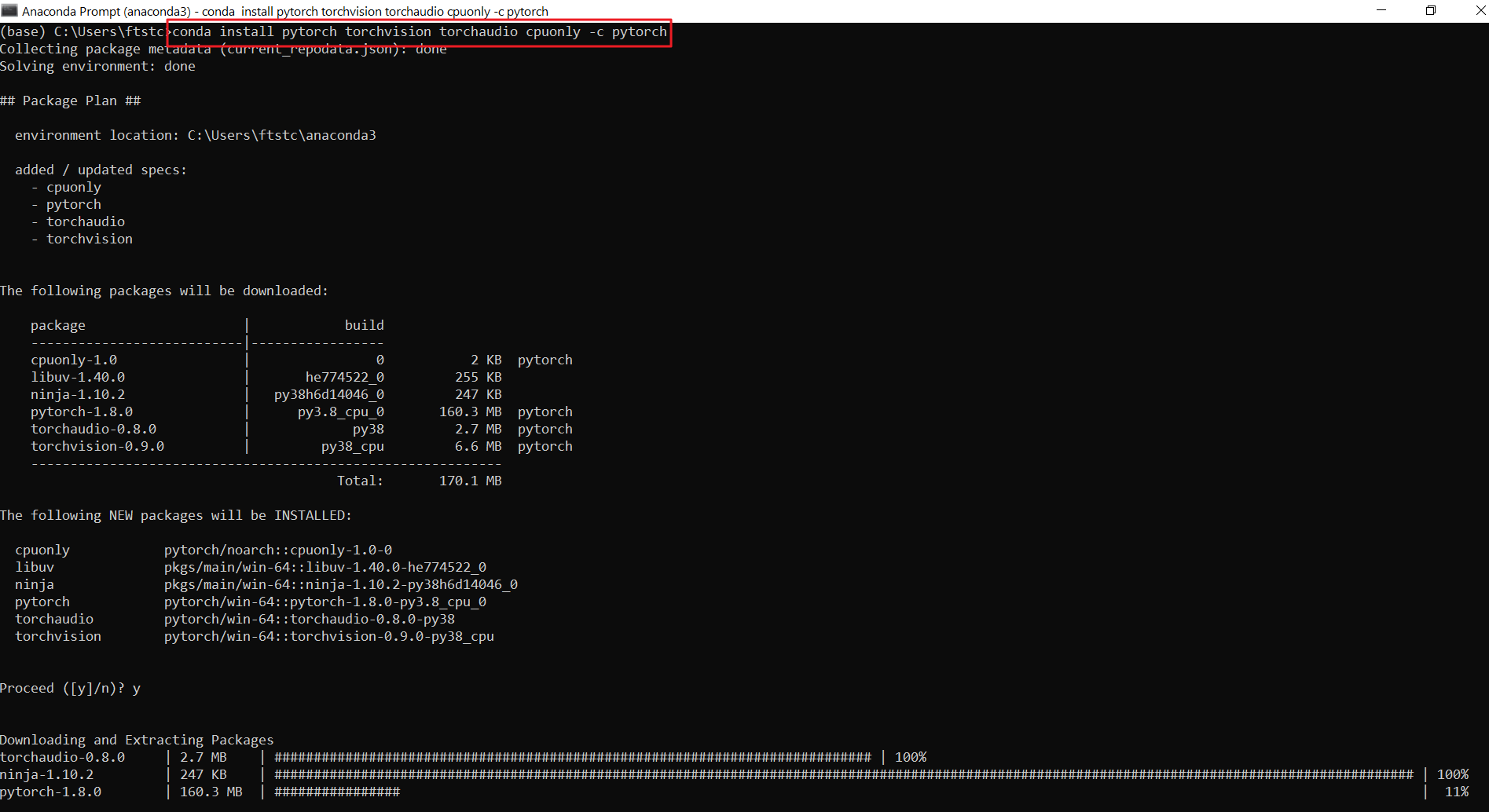
<https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.concat.html>

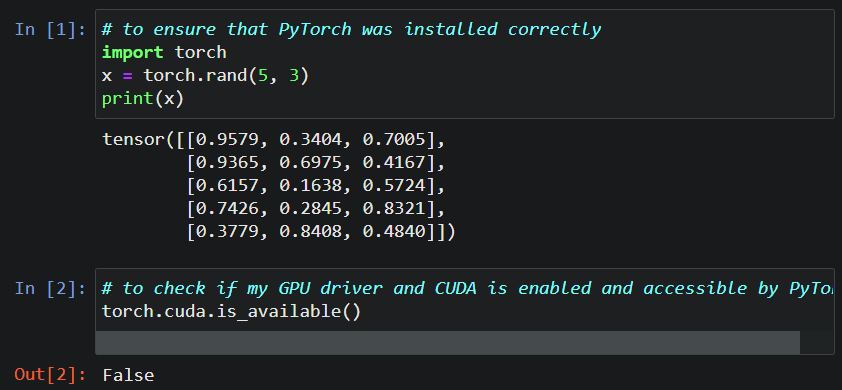
【python np.set\_printoptions】

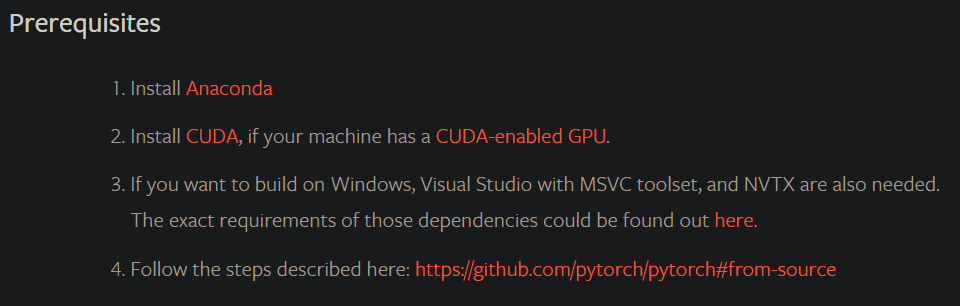
210318(四)

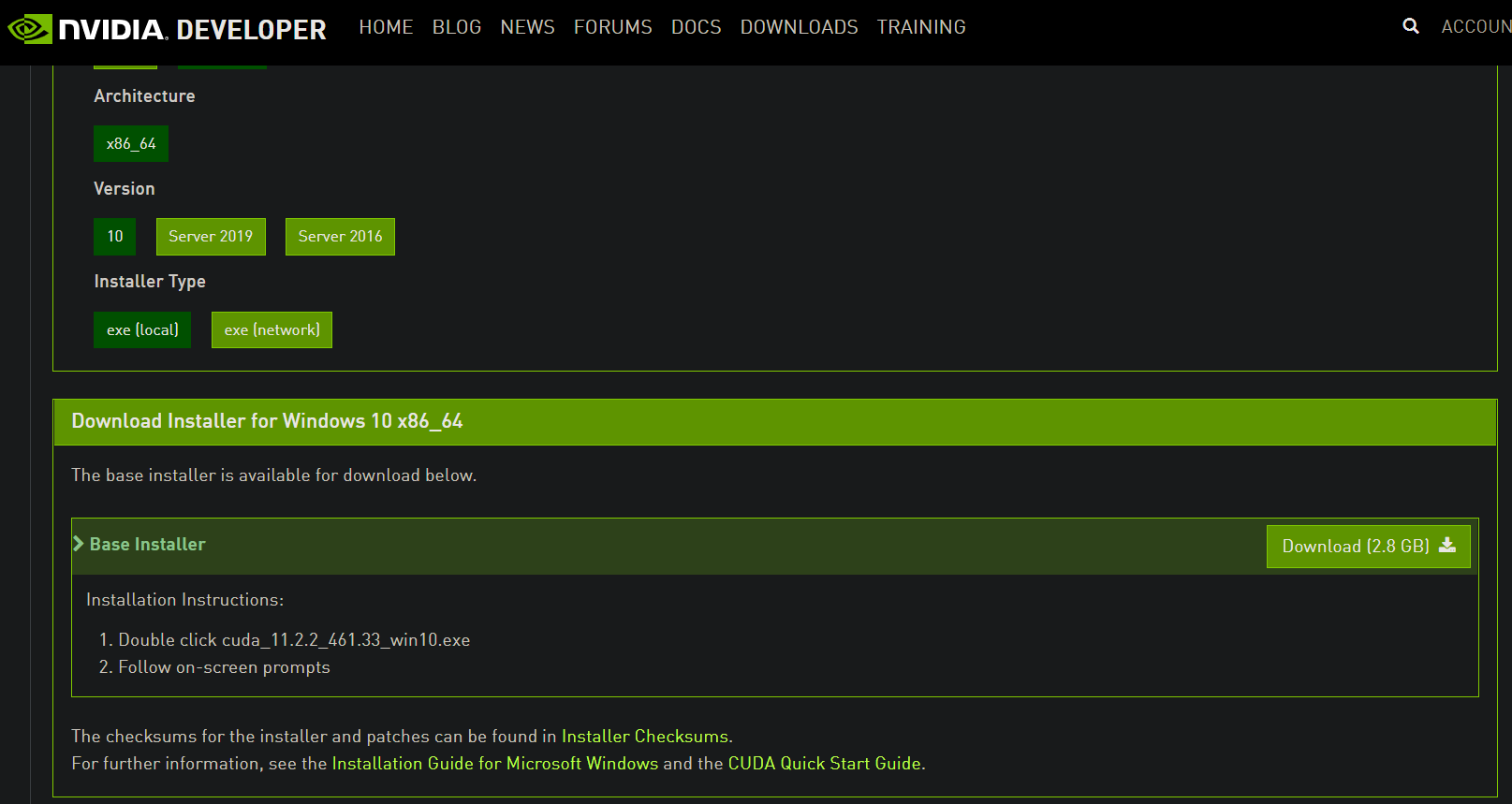
安裝pytorch

<https://pytorch.org/get-started/locally/#anaconda>

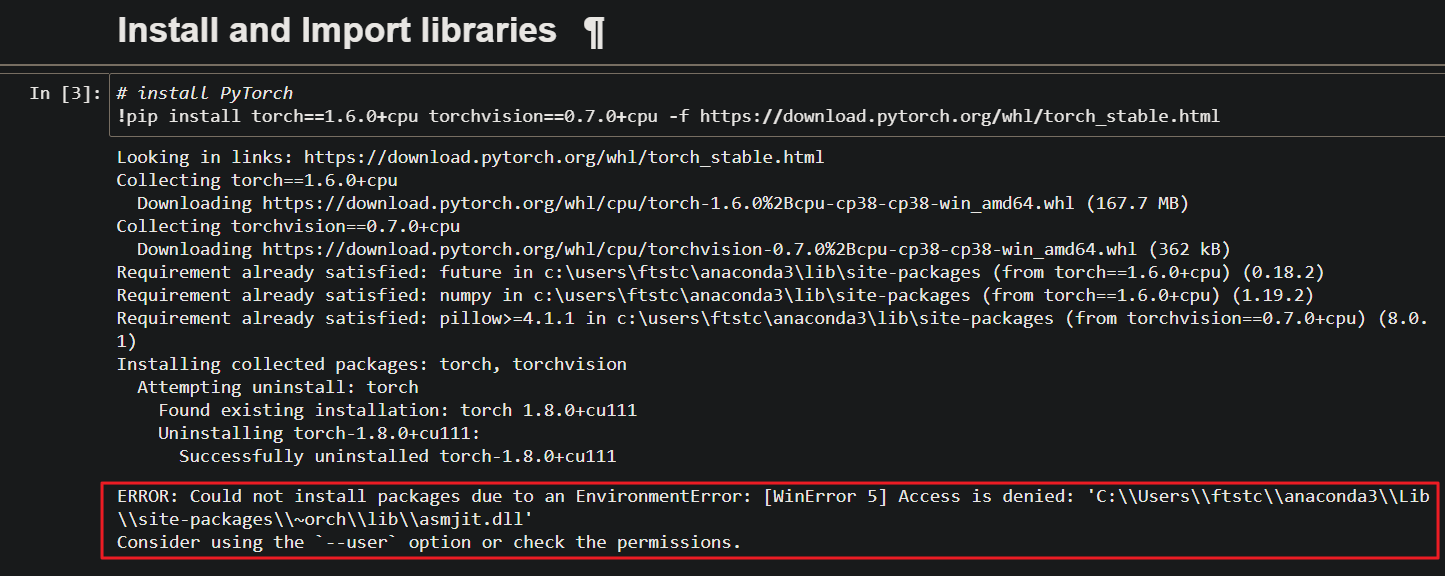




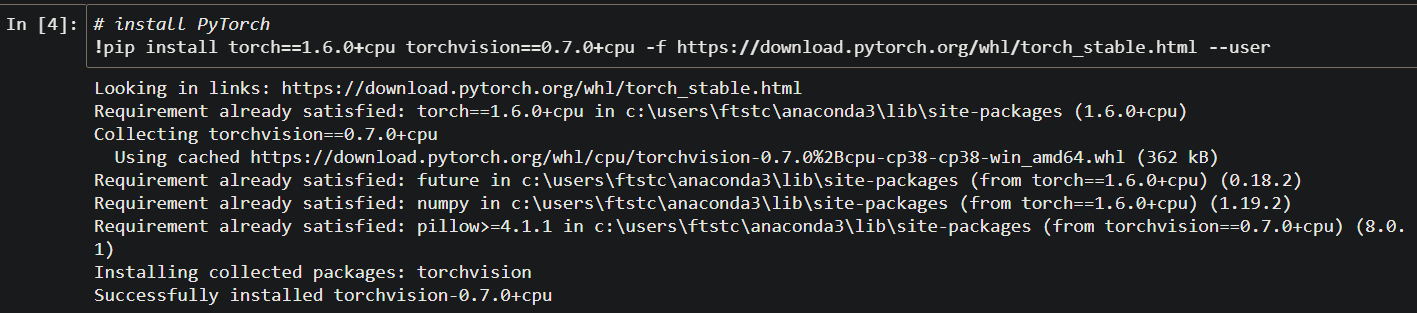




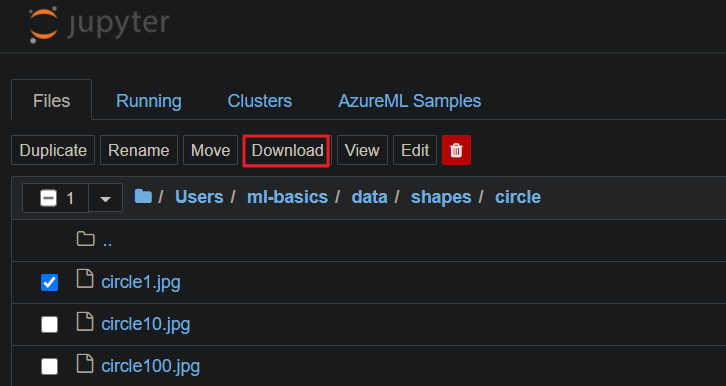
210415(Thur.)



<https://stackoverflow.com/questions/51912999/could-not-install-packages-due-to-an-environmenterror-winerror-5-access-is-de>



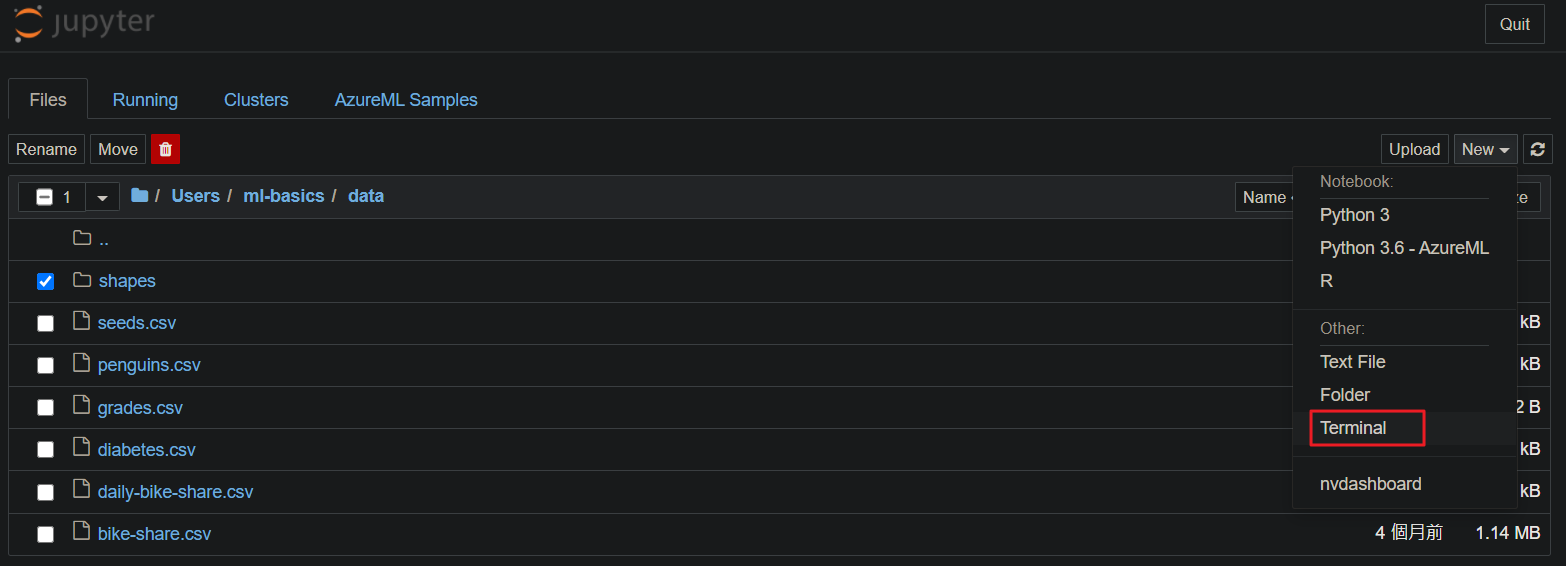
/// 壓縮jupyter notebook 上的file，以便於一起下載



(只能單一檔案下鮺)

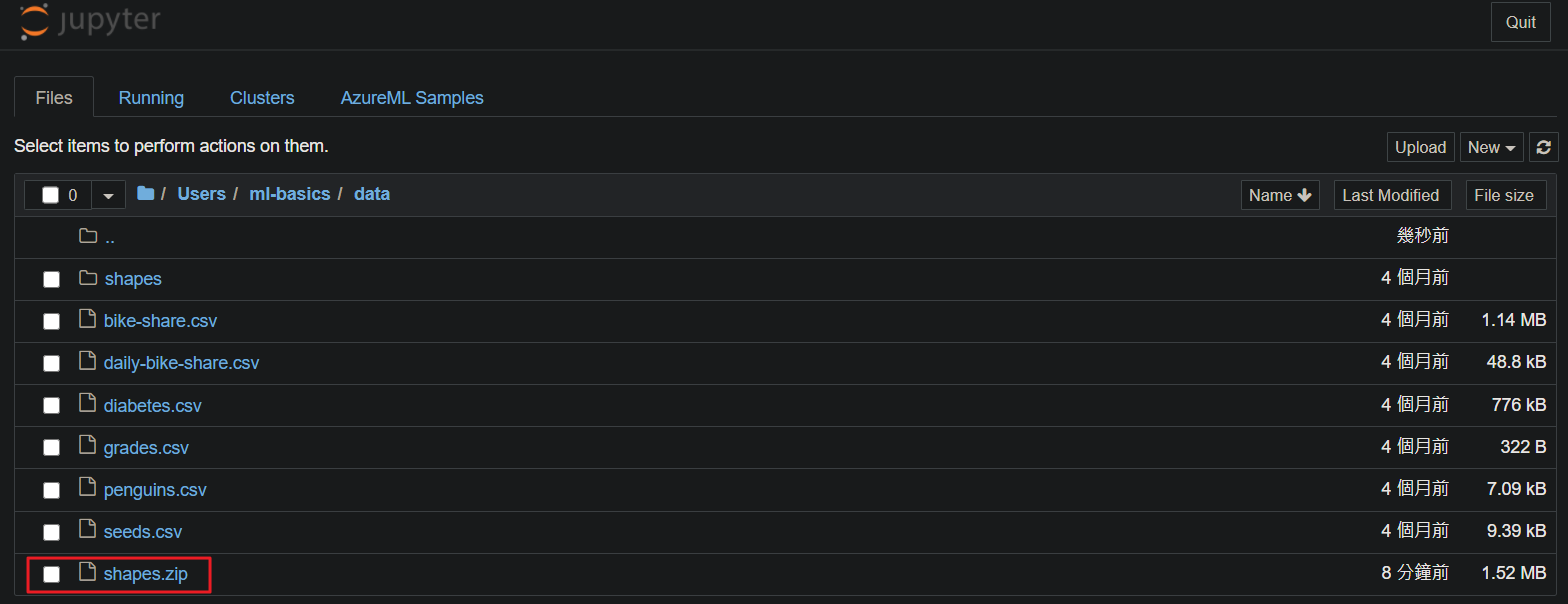
<https://github.com/jupyter/notebook/issues/3092>

開啟jupyter notebook的terminal



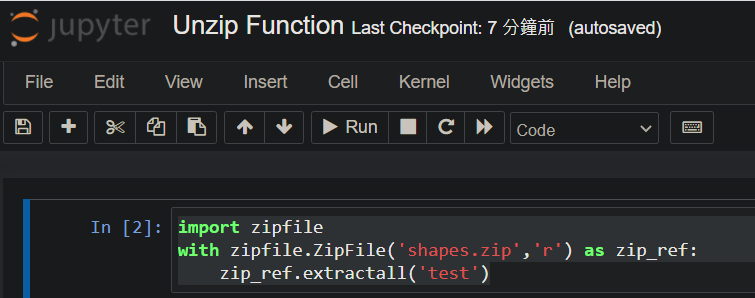
cd 到 shapes的檔案夾

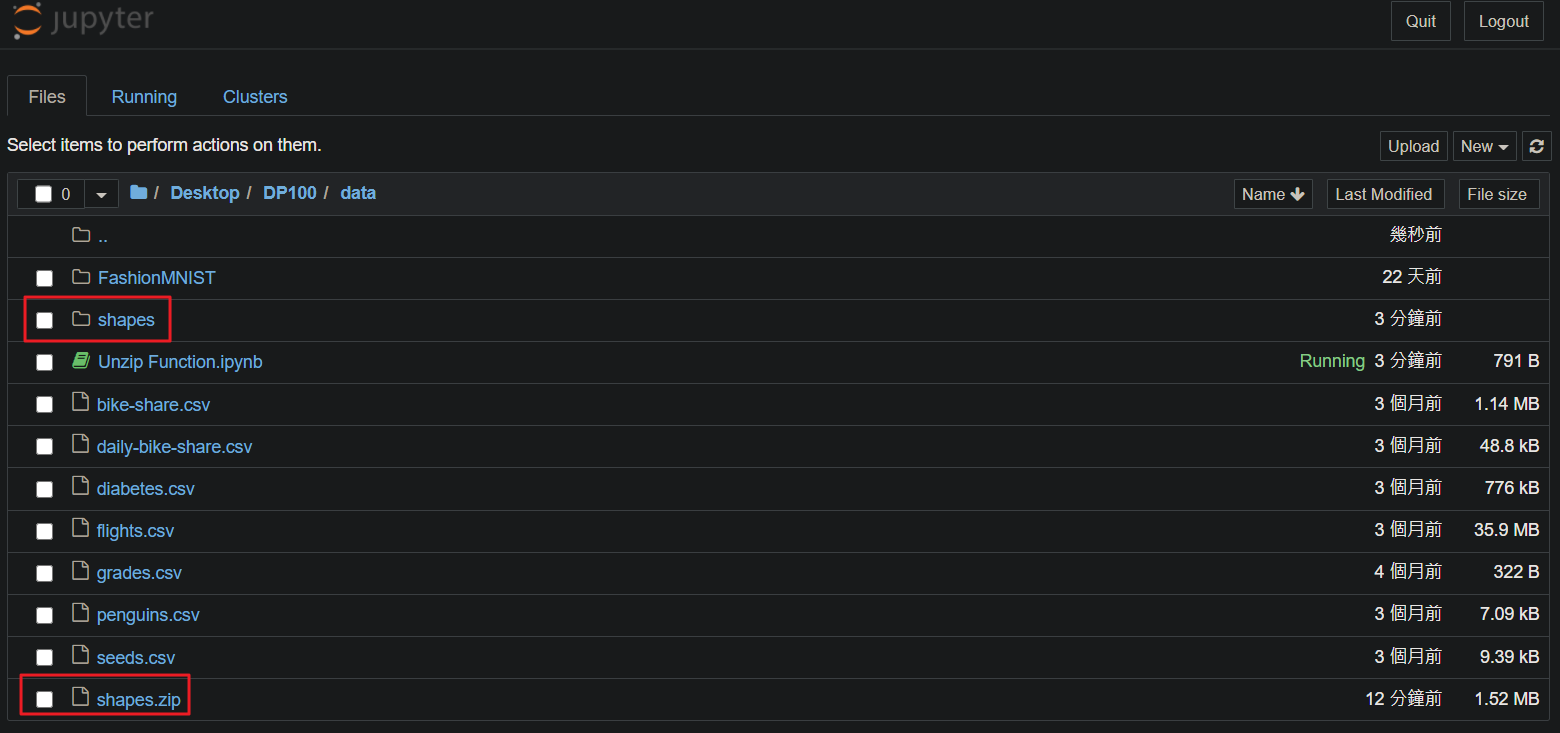
zip -r shapes.zip shapes



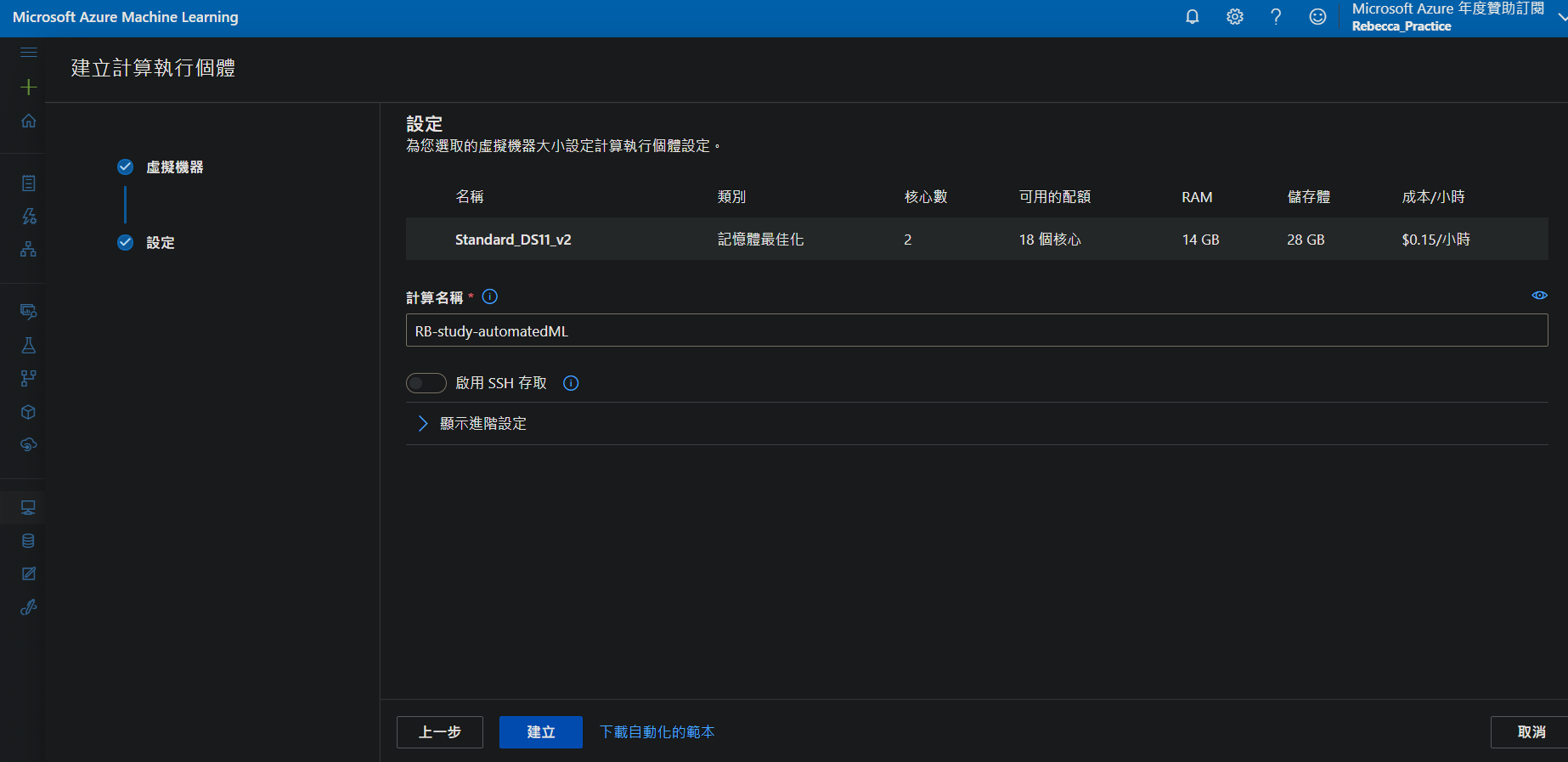
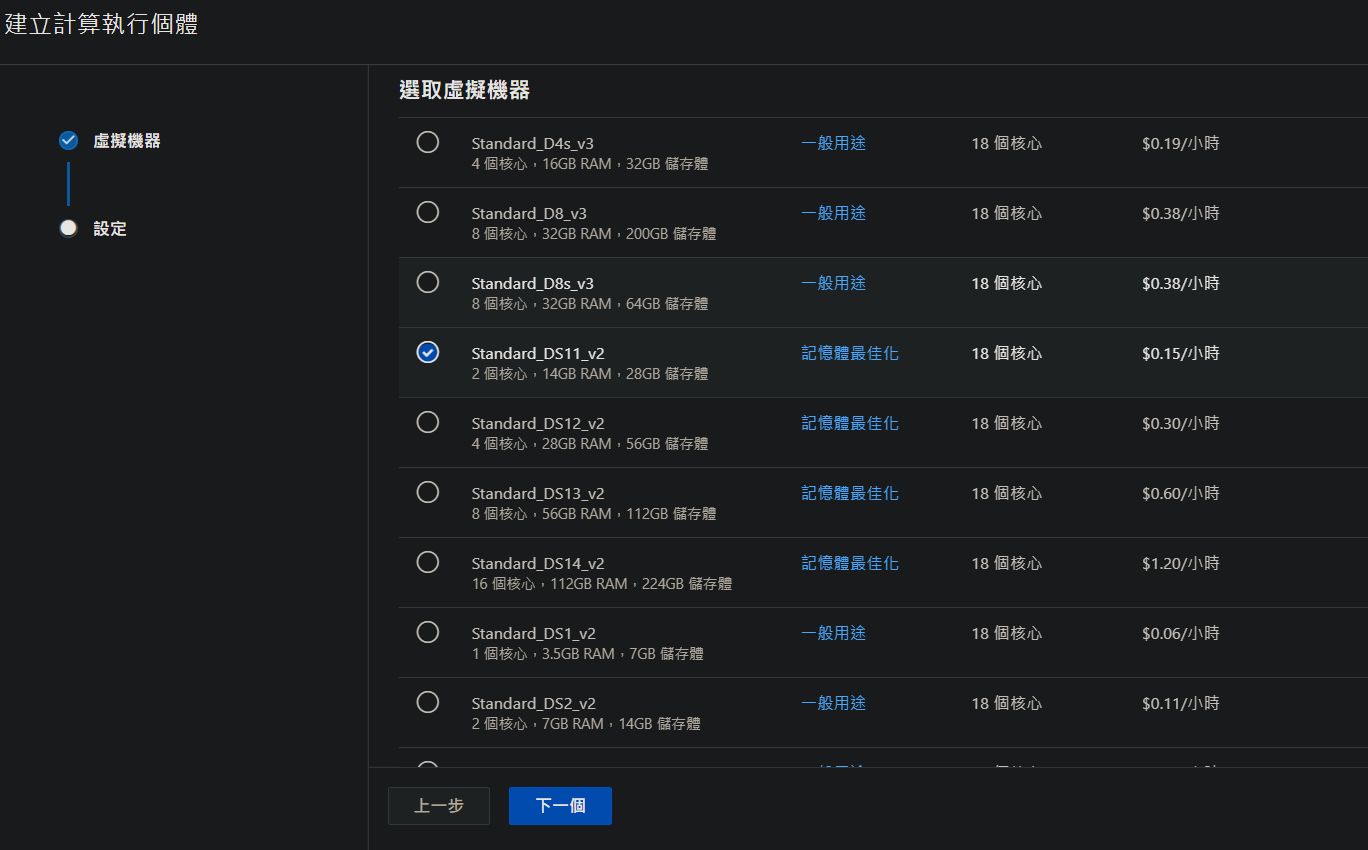
可以一次下載了

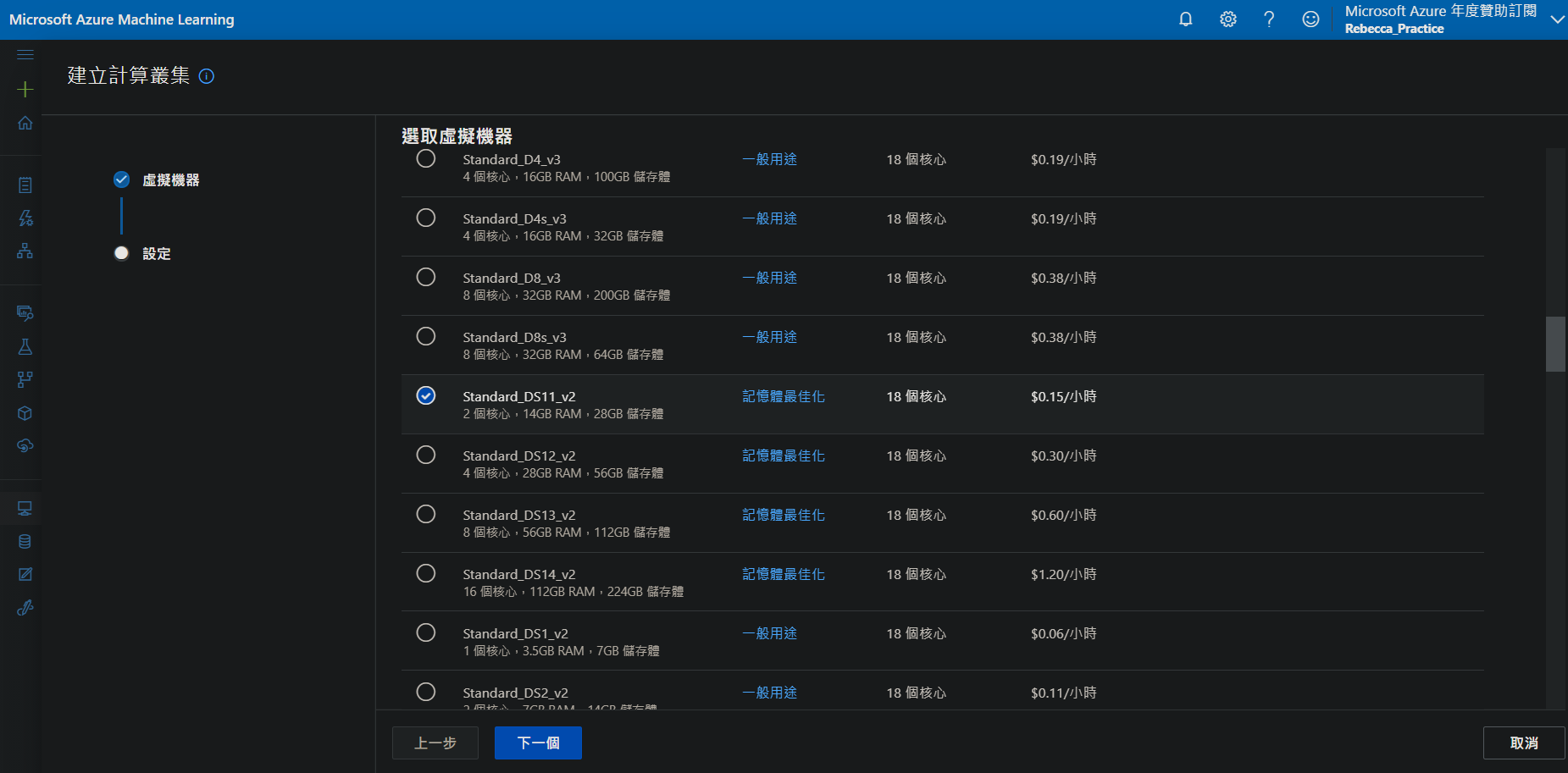
上傳自己的環境後，解壓縮



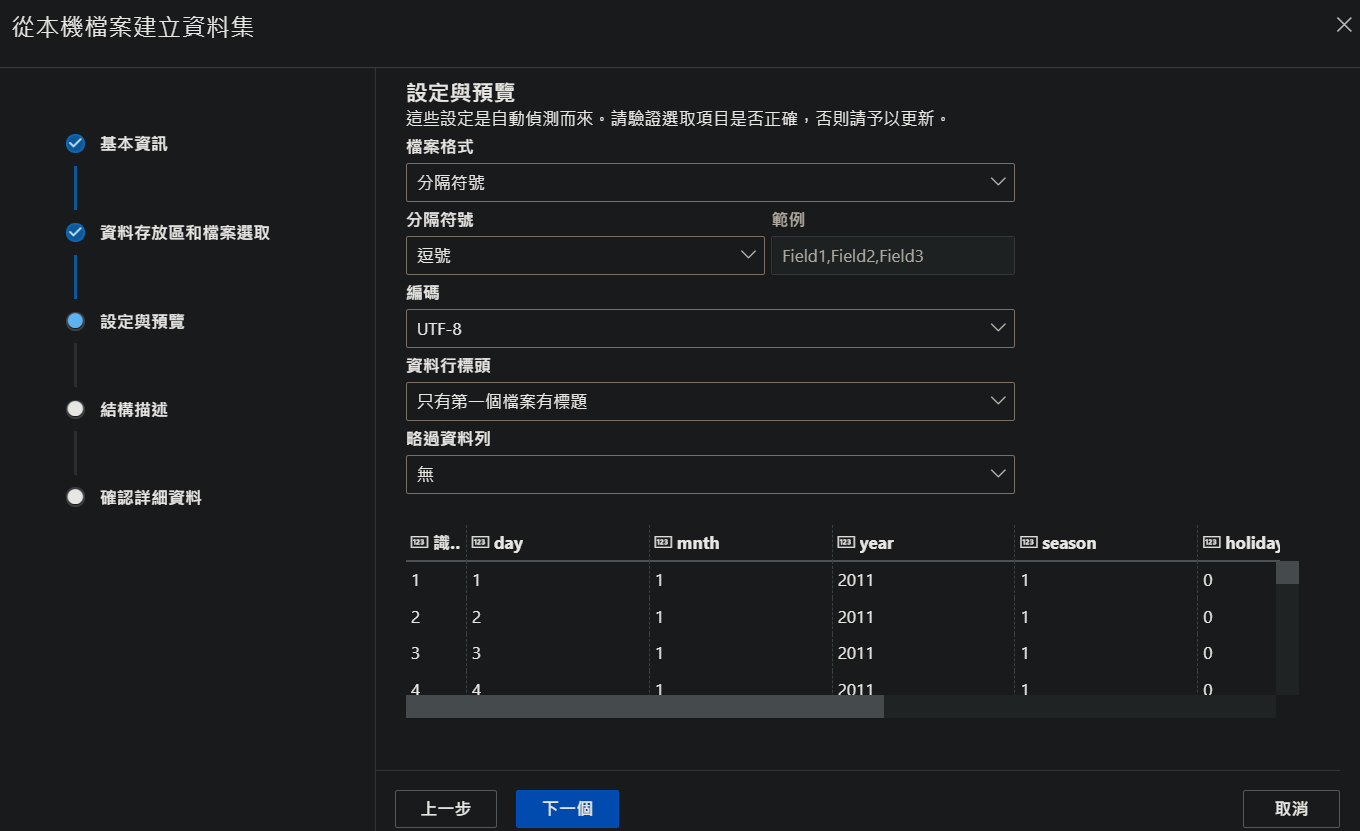


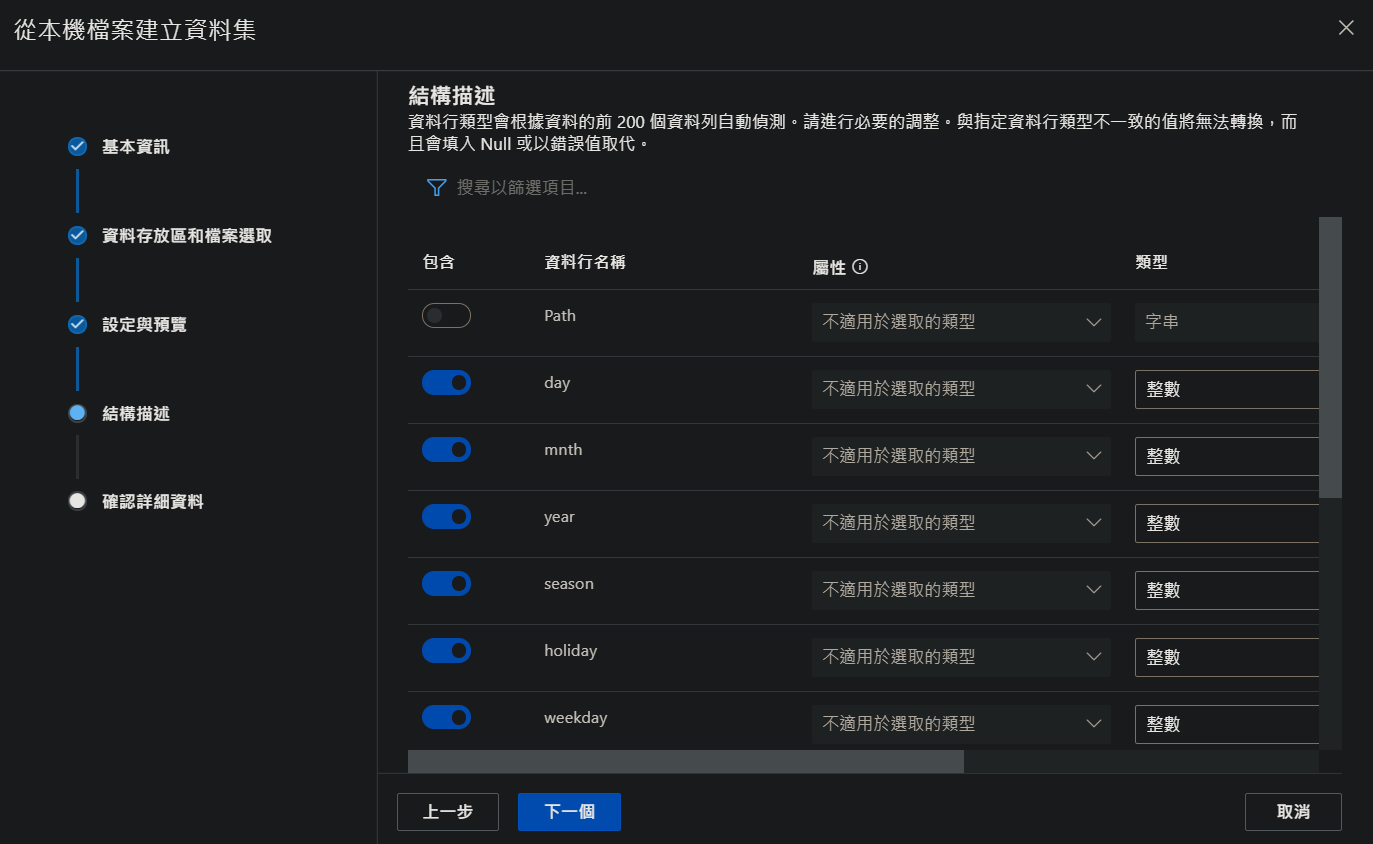
210525(Tues.)









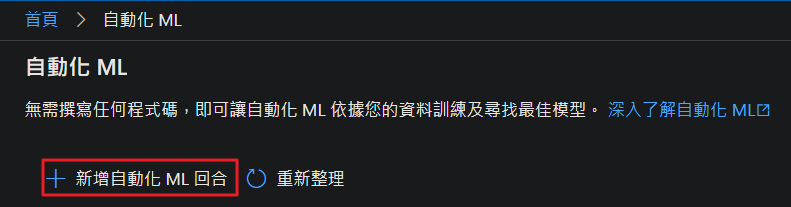


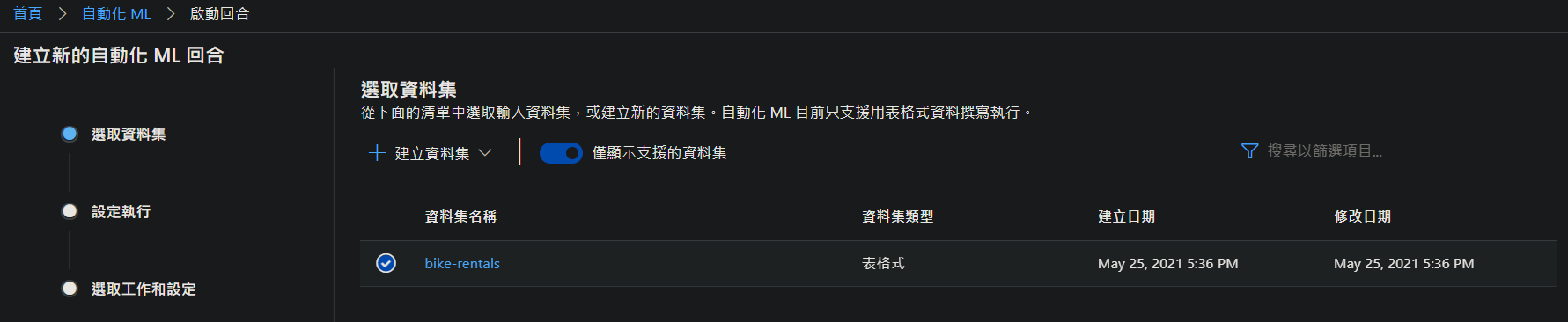


210526(Wens.)

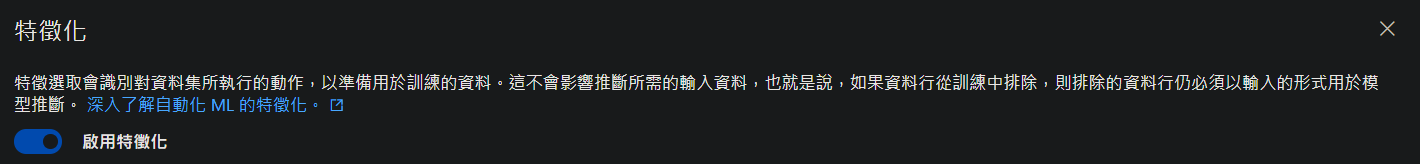
Run an automated machine learning experiment

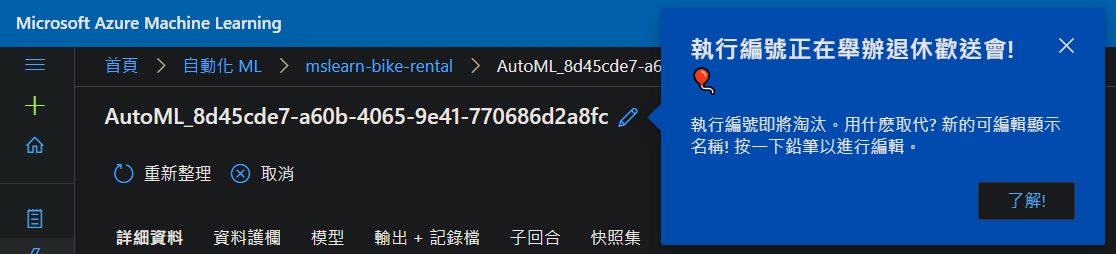












1. Select dataset

\* Dataset

1. Configure run

\* New experiment name

\* Target column (the label the model will be trained to predict)

\* Select compute cluster

1. Task type and settings:

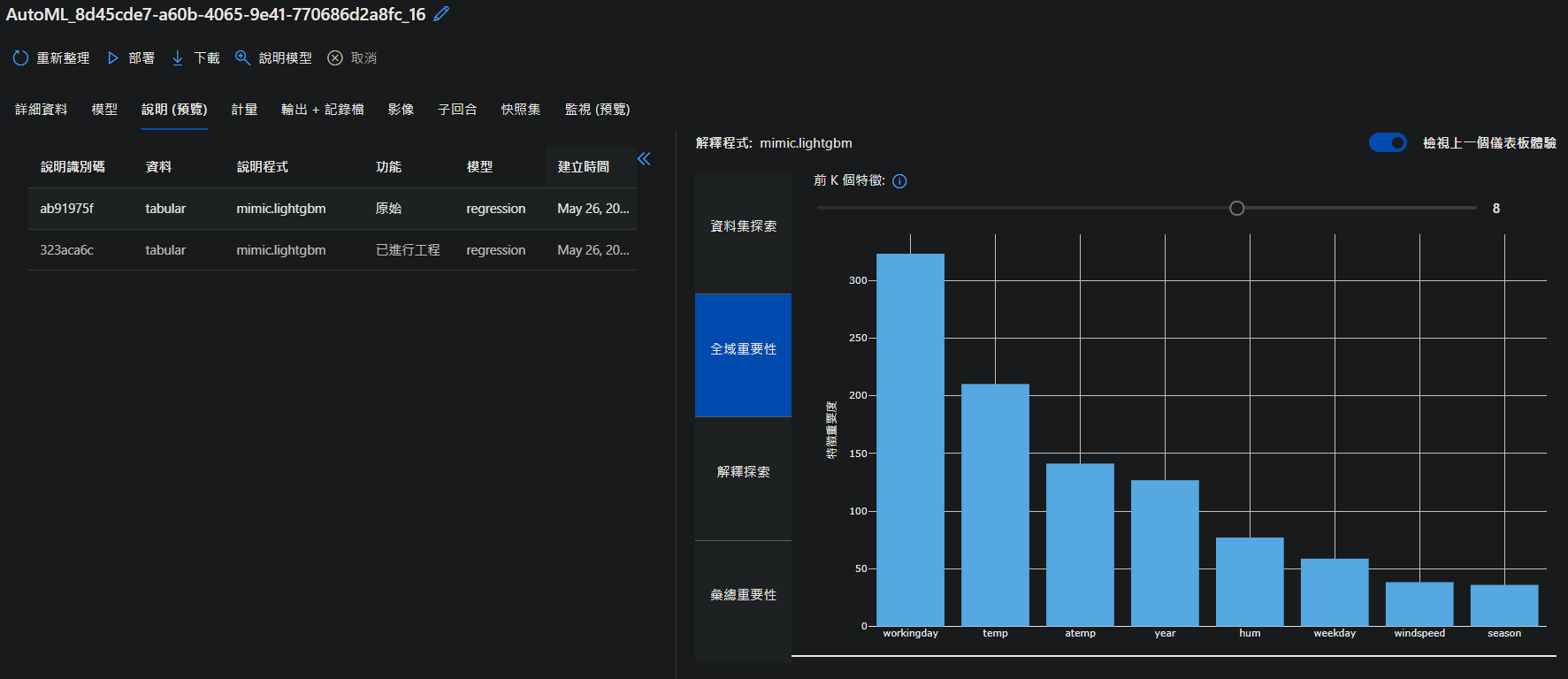
\* Task type: Regression (the model will predict a numeric value)

\* Additional configuration settings:

1. Primary metric: Select Normalized root mean squared
2. Explain best model: Selected - this option causes automated machine learning to calculate feature importance for the best model; making it possible to determine the influence of each feature on the predicted label.
3. Blocked algorithms: Block all other than RandomForest and LightGBM - normally you'd want to try as many as possible, but doing so can take a long time
4. Exit criterion:
5. Training job time (hours): causing the experiment to end after a maximum of X minutes.
6. Metric score threshold: causing the experiment to end if a model achieves a normalized root mean squared error metric score of X or less.
7. Featurization settings:

\* Enable featurization: Selected - this causes Azure Machine Learning to automatically preprocess the features before training.

The best model is identified based on the evaluation metric you specified (Normalized root mean squared error). To calculate this metric, the training process used some of the data to train the model, and applied a technique called cross-validation to iteratively test the trained model with data it wasn't trained with and compare the predicted value with the actual known value. The difference between the predicted and actual value (known as the residuals) indicates the amount of error in the model, and this particular performance metric is calculated by squaring the errors across all of the test cases, finding the mean of these squares, and then taking the square root. What all of this means is that smaller this value is, the more accurately the model is predicting.







210527(Thur.)

