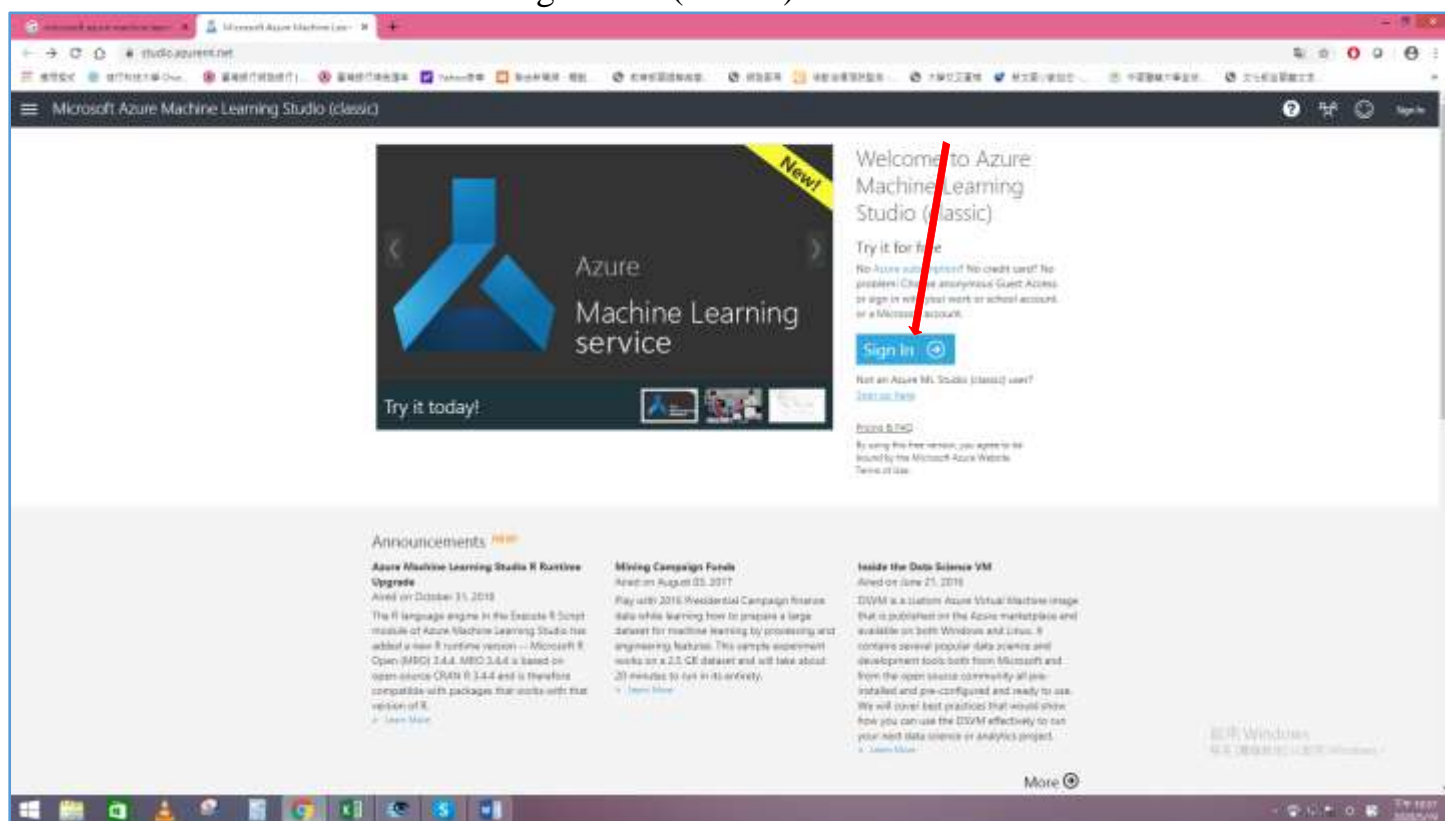


實作平台 Microsoft Azure Machine Learning Studio(classic)

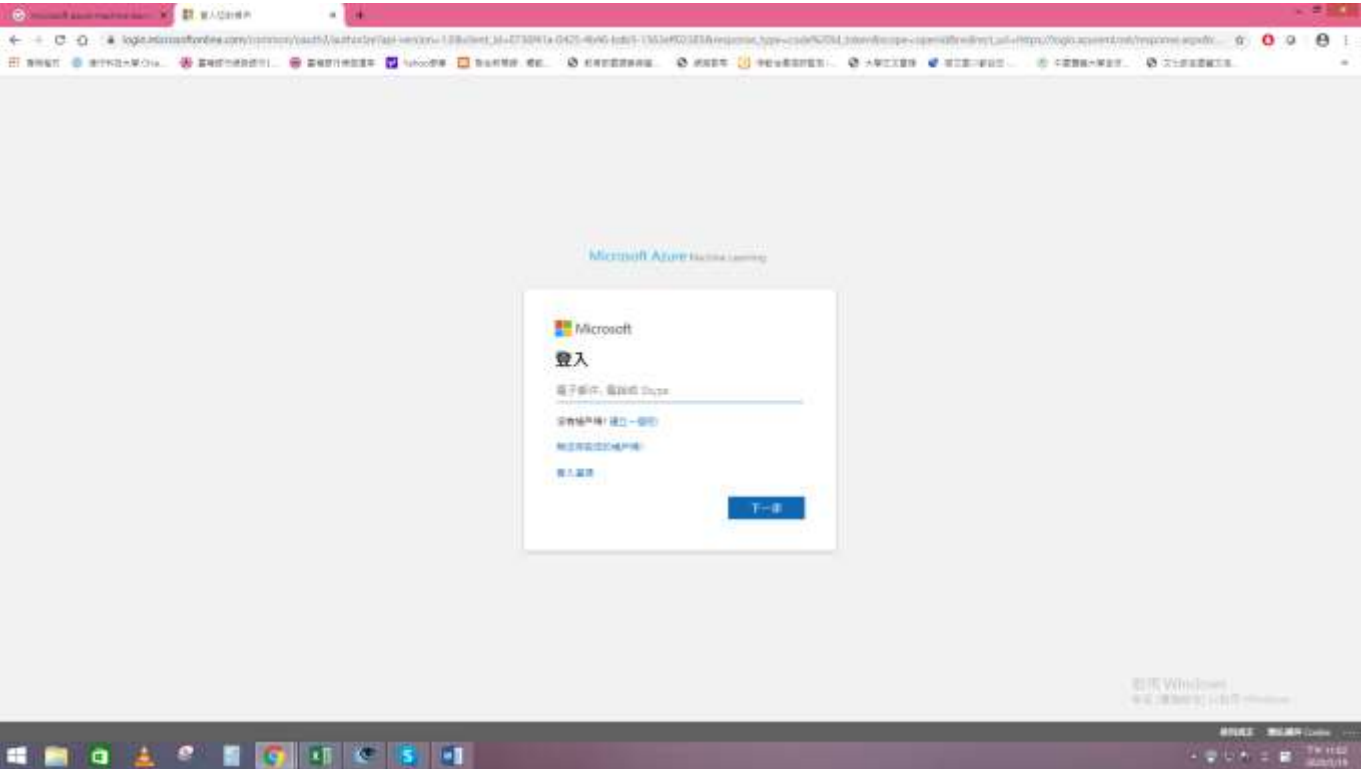
教學文件

<https://docs.microsoft.com/zh-tw/azure/machine-learning/studio/what-is-ml-studio>

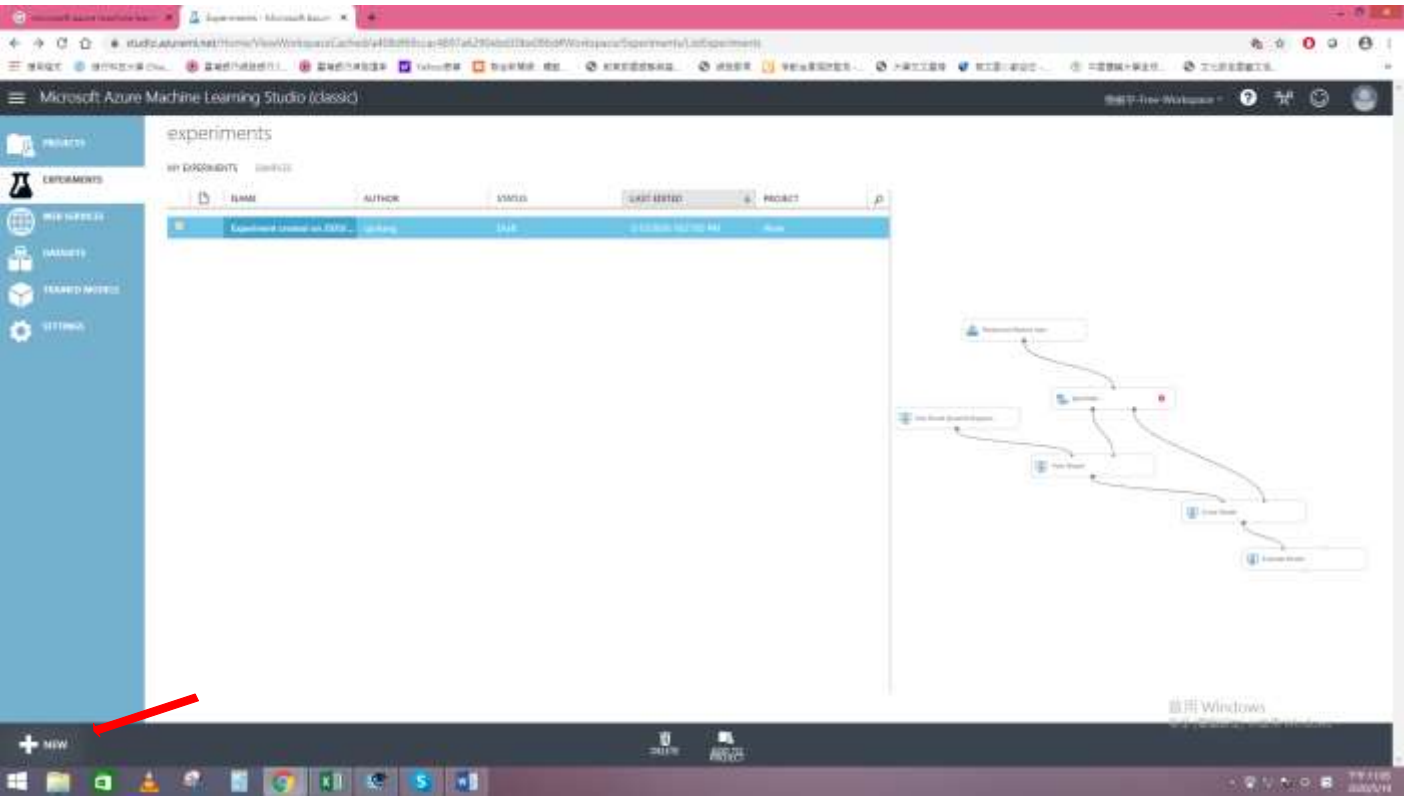
Microsoft Azure Machine Learning Studio (classic)



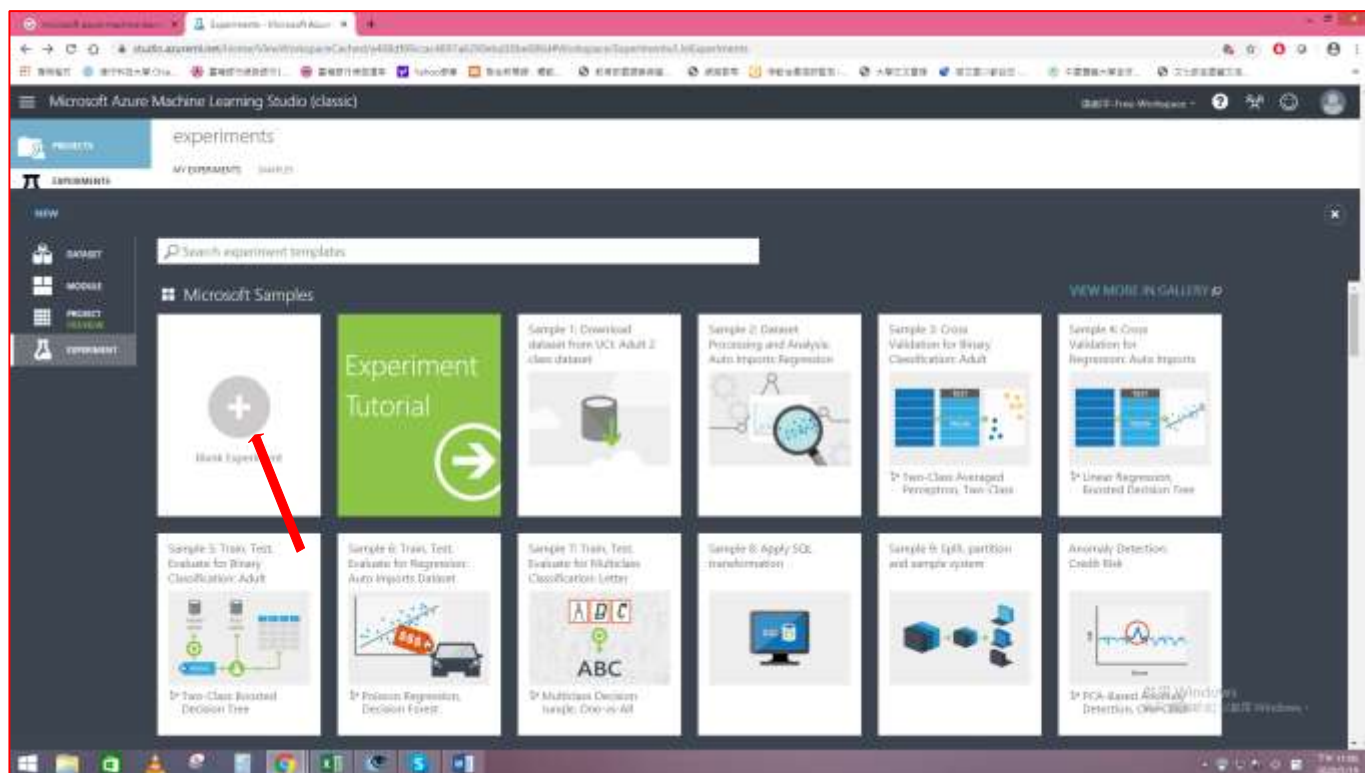
使用學校微軟 365 帳號登入



先 NEW 一個新的 Project



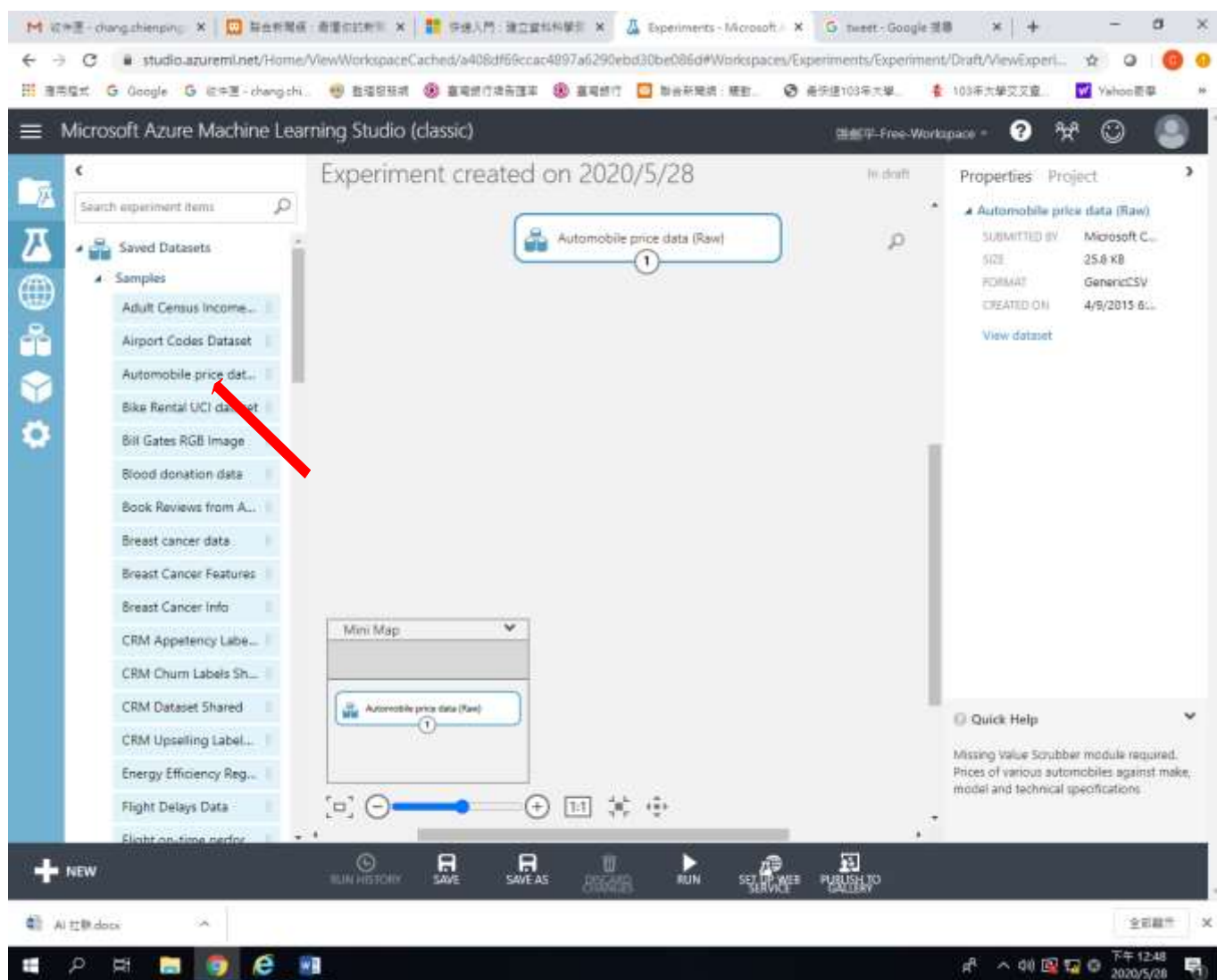
點擊 Blank Experiment



選 saved datasets 下拉 samples 汽車價格資料

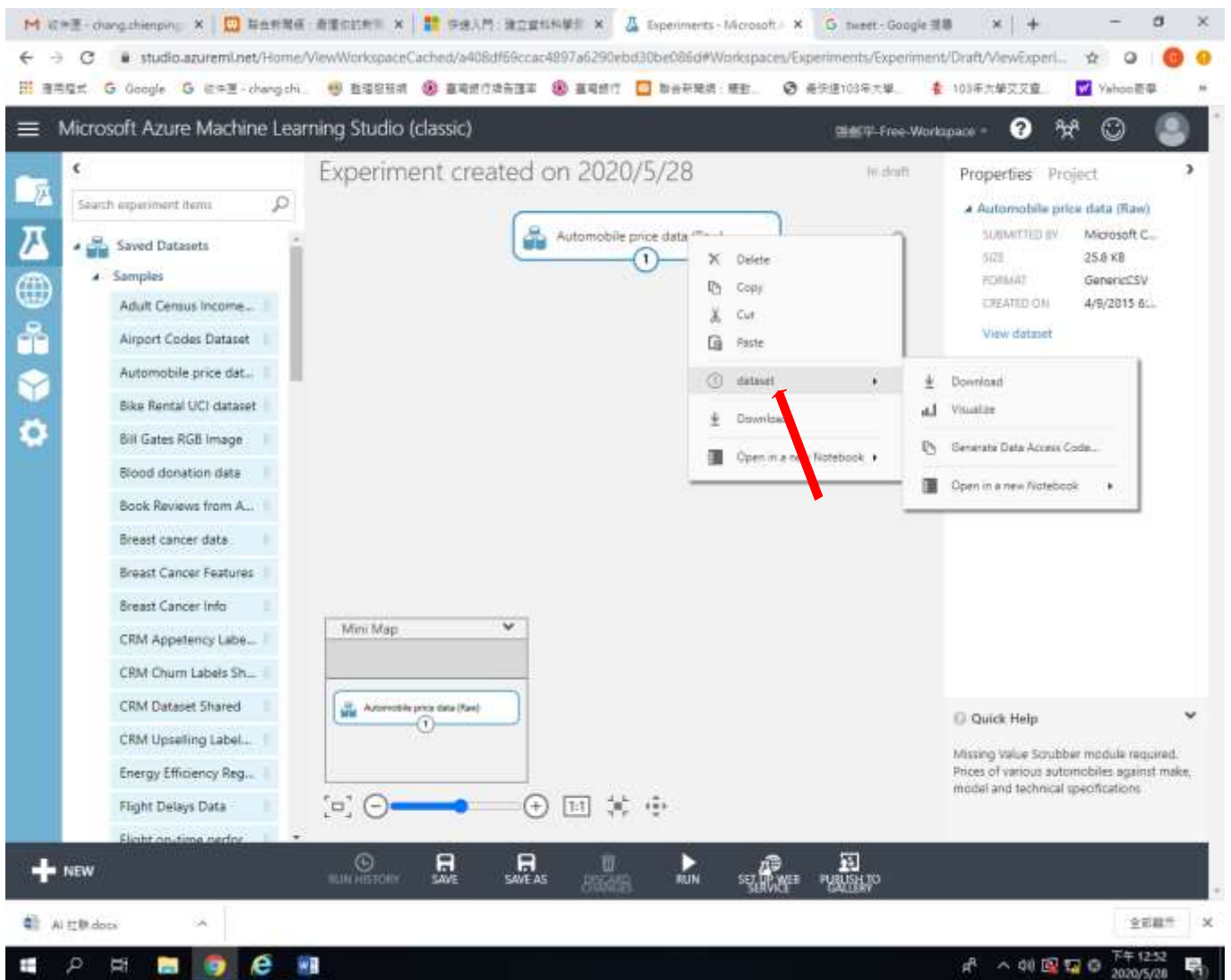
依構造和型號分類的汽車相關資訊，包括價格、性能（例如汽缸數和油耗），以及保險風險評分。風險分數一開始與自動價格相關聯。然後它會在精算師稱為符號化的程序中根據實際風險進行調整。若值為 +3，表示該汽車屬於高風險，若值為 -3，表示大致而言很安全。

1. 拖拉至桌面



資料集通常必須先經過某些前置處理，才能進行分析。您可能已經注意到在各種不同資料列的資料行中有遺漏的值。必須清除這些遺漏的值，讓模型才能正確地分析資料。我們將移除含有遺漏值的所有資料列。

2. 按滑鼠右鍵，可視化資料



顯示資料

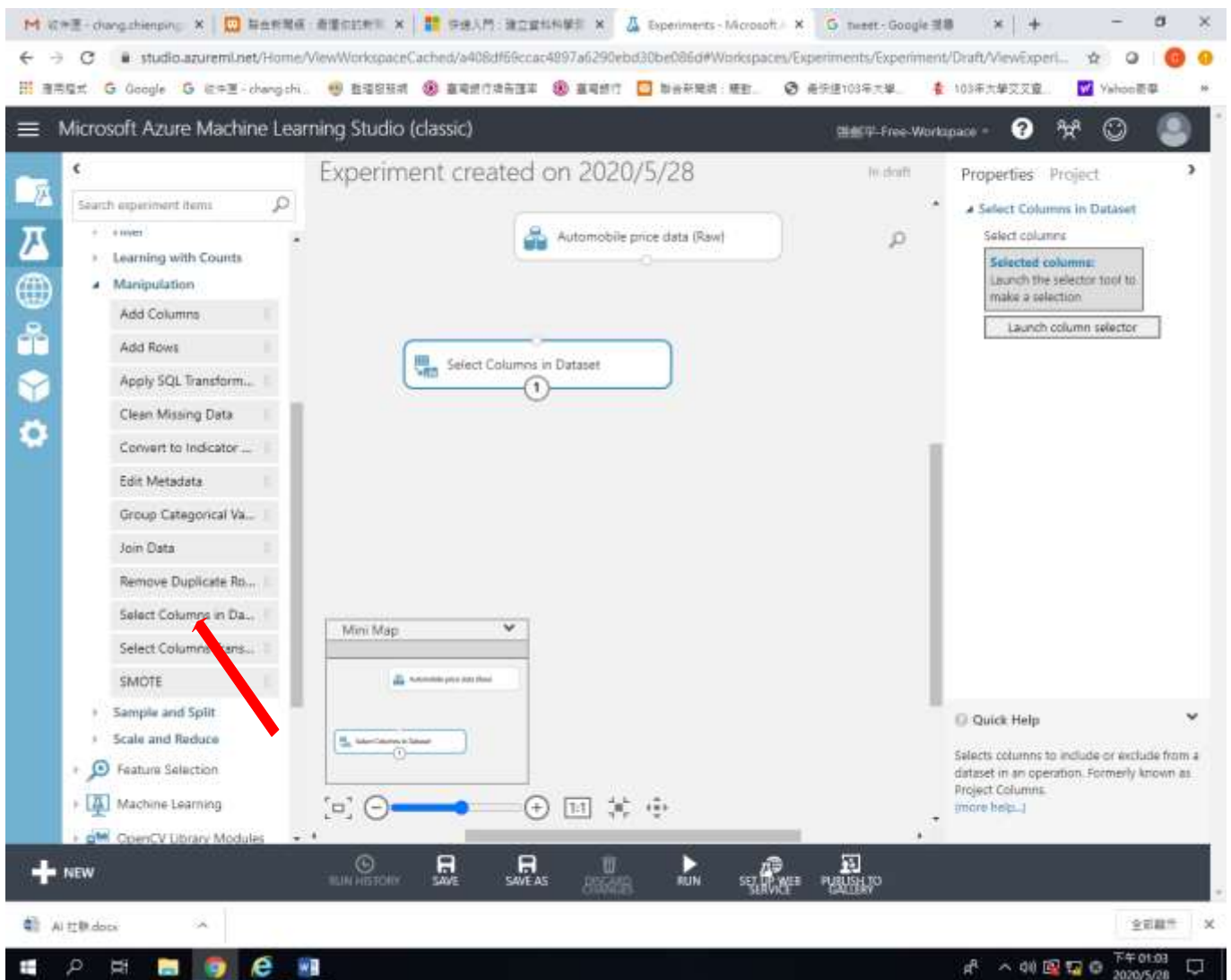
The screenshot displays the Microsoft Azure Machine Learning Studio (classic) interface. The main window shows a dataset titled "Automobile price data (Raw)" with 205 rows and 26 columns. The columns are: symboling, normalized-losses, make, fuel-type, aspiration, and num-of-doors. A red arrow points to the "normalized-losses" column. The dataset is displayed in a table view with the following data:

symboling	normalized-losses	make	fuel-type	aspiration	num-of-doors
3		alfa-romero	gas	std	two
3		alfa-romero	gas	std	two
1		alfa-romero	gas	std	two
2	164	audi	gas	std	four
2	164	audi	gas	std	four
2		audi	gas	std	two
1	158	audi	gas	std	four
1		audi	gas	std	four
1	158	audi	gas	turbo	four
0		audi	gas	turbo	two
2	192	bmw	gas	std	two
0	192	bmw	gas	std	four

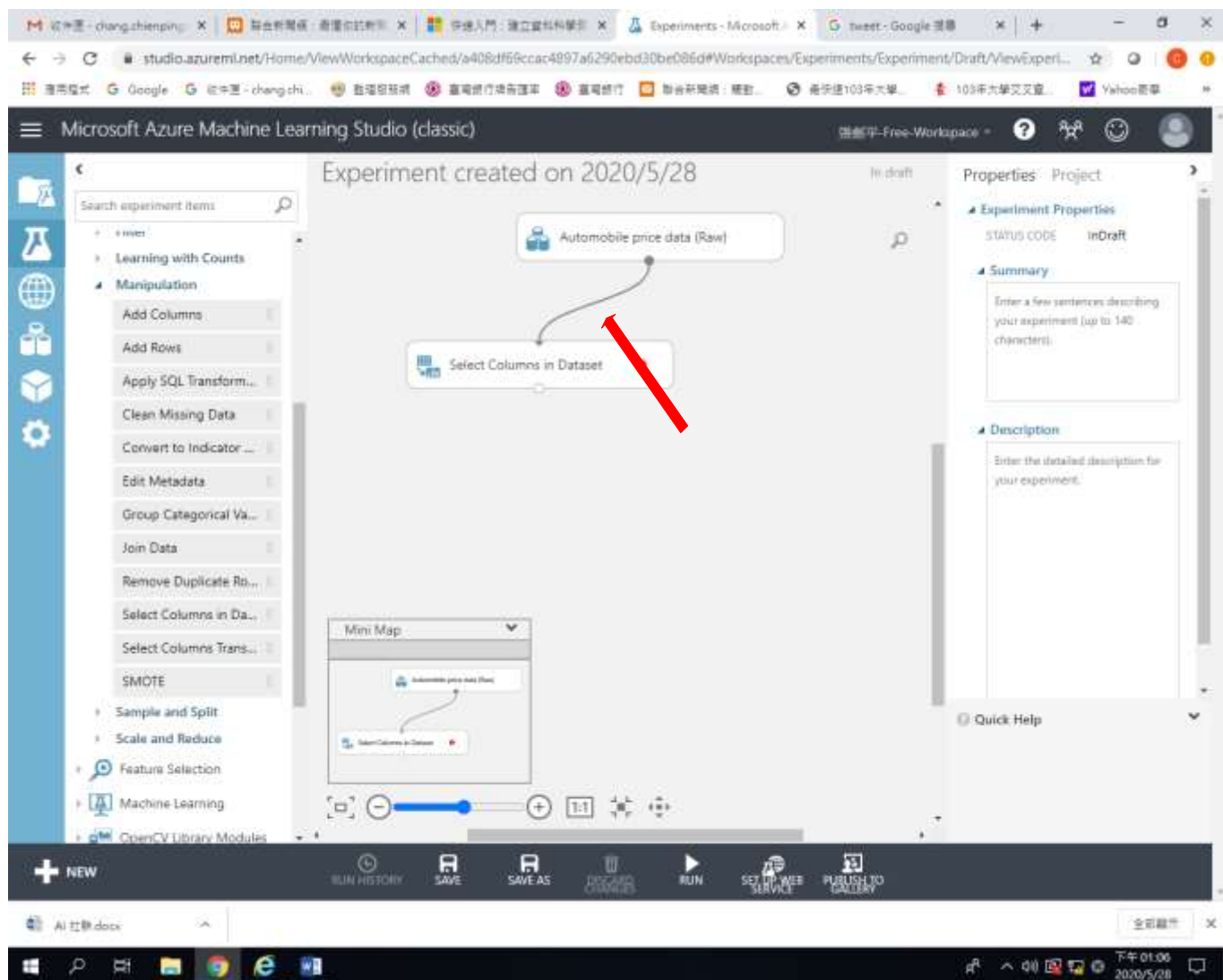
The interface also includes a sidebar with navigation icons, a top bar with the experiment name and date, and a bottom bar with various action buttons like "NEW", "RUN HISTORY", "SAVE", "SAVE AS", "DELETE", "RUN", "SET UP WEB SERVER", and "RUN WITH COLLECT".

首先，我們新增一個完全移除 [normalized losses] 資料行的模組。然後，再新增一個將任何含有遺漏資料的資料列移除的模組。

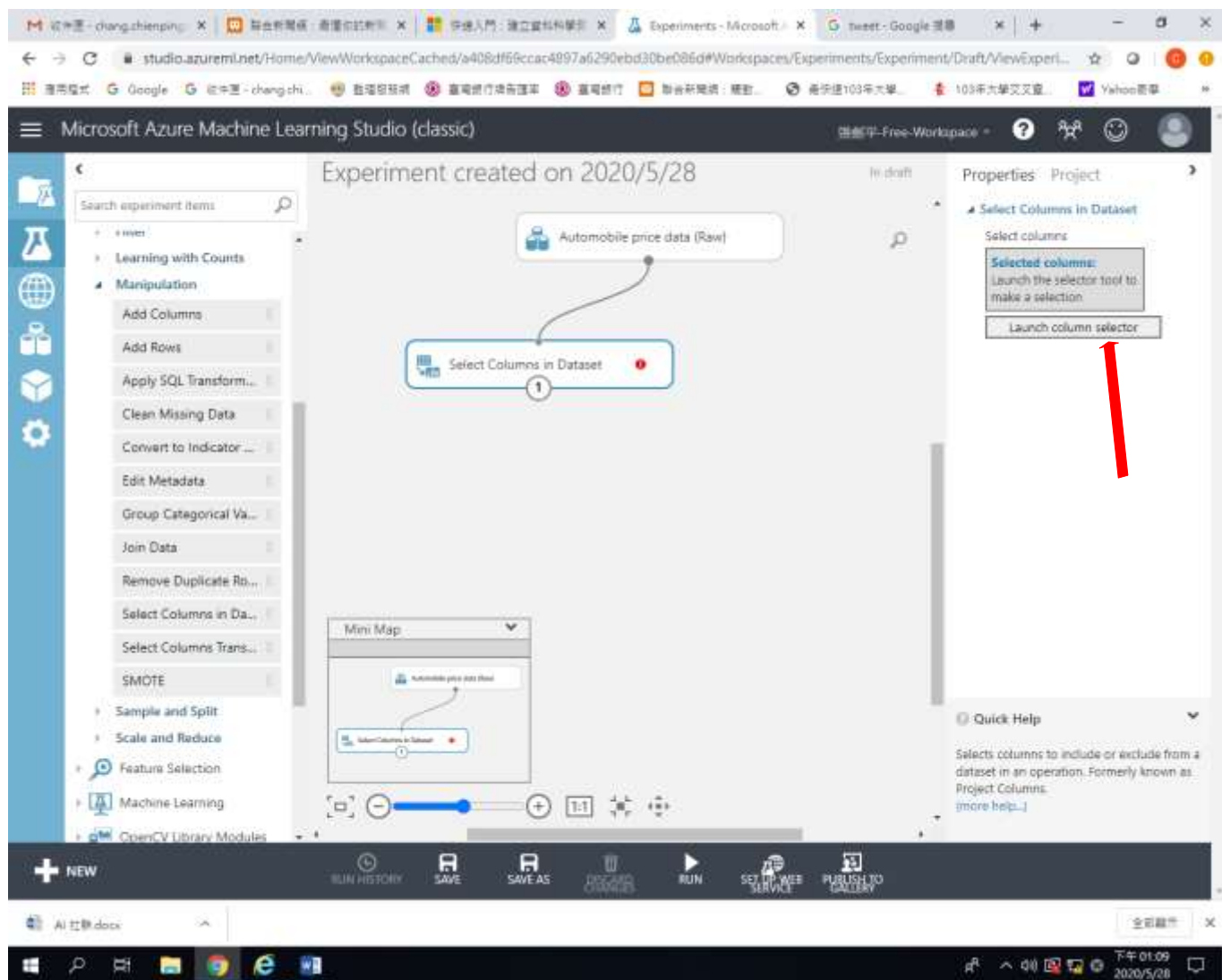
(1)在模組選擇區頂端的搜尋方塊中輸入選取資料行，以尋找選取資料集中的資料行模組。然後，將其拖曳到實驗畫布上。此模組可讓我們選取要將哪些資料行包含在模型中，或是從模型中排除。



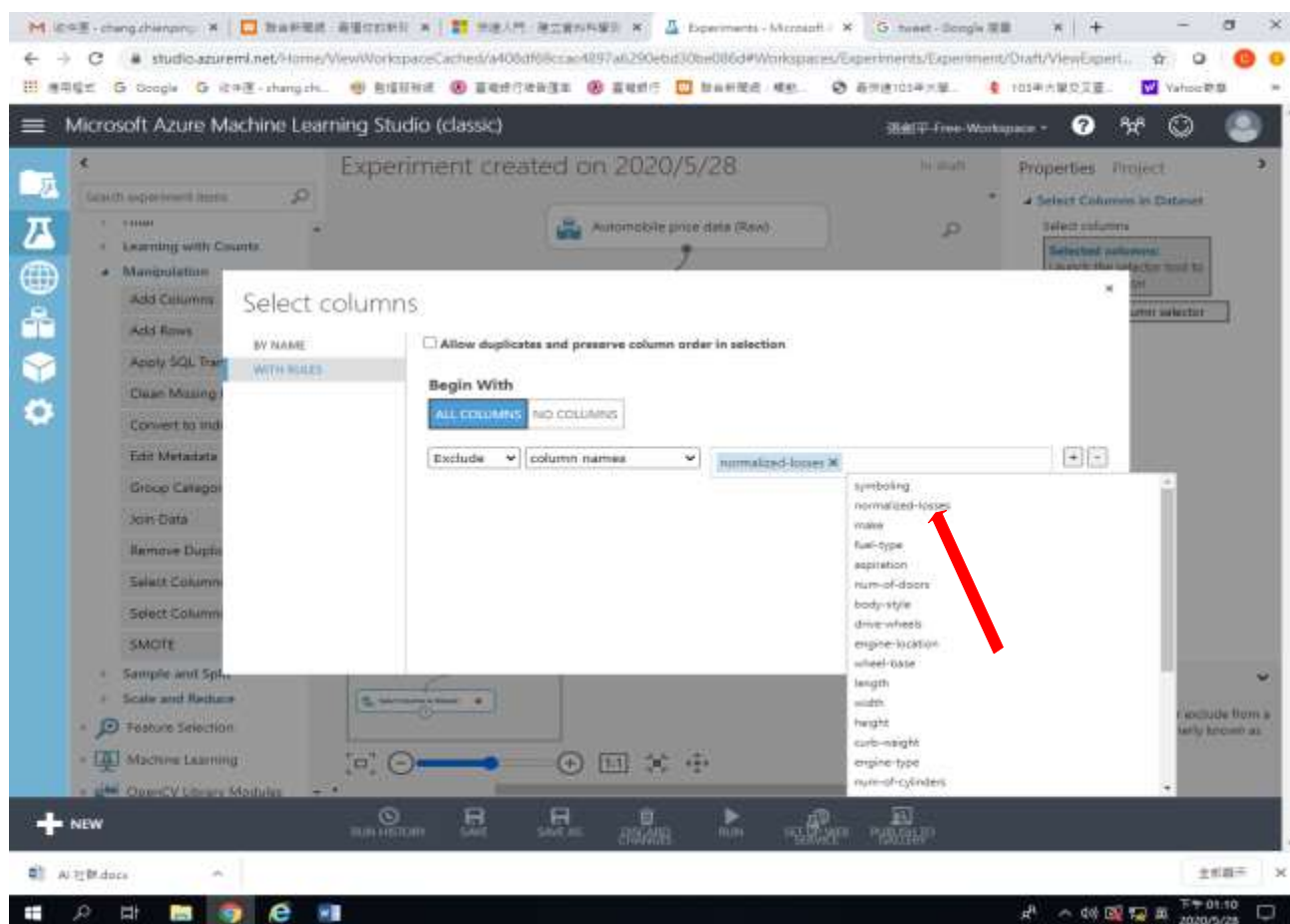
(2) 將汽車價格資料（原始）資料集的輸出連接埠連線至「選取資料集中的資料行」的輸入連接埠。



(3) 按一下選取資料集中的資料行模組，然後按一下 [屬性] 窗格中的 [啟動資料行選取器] 。

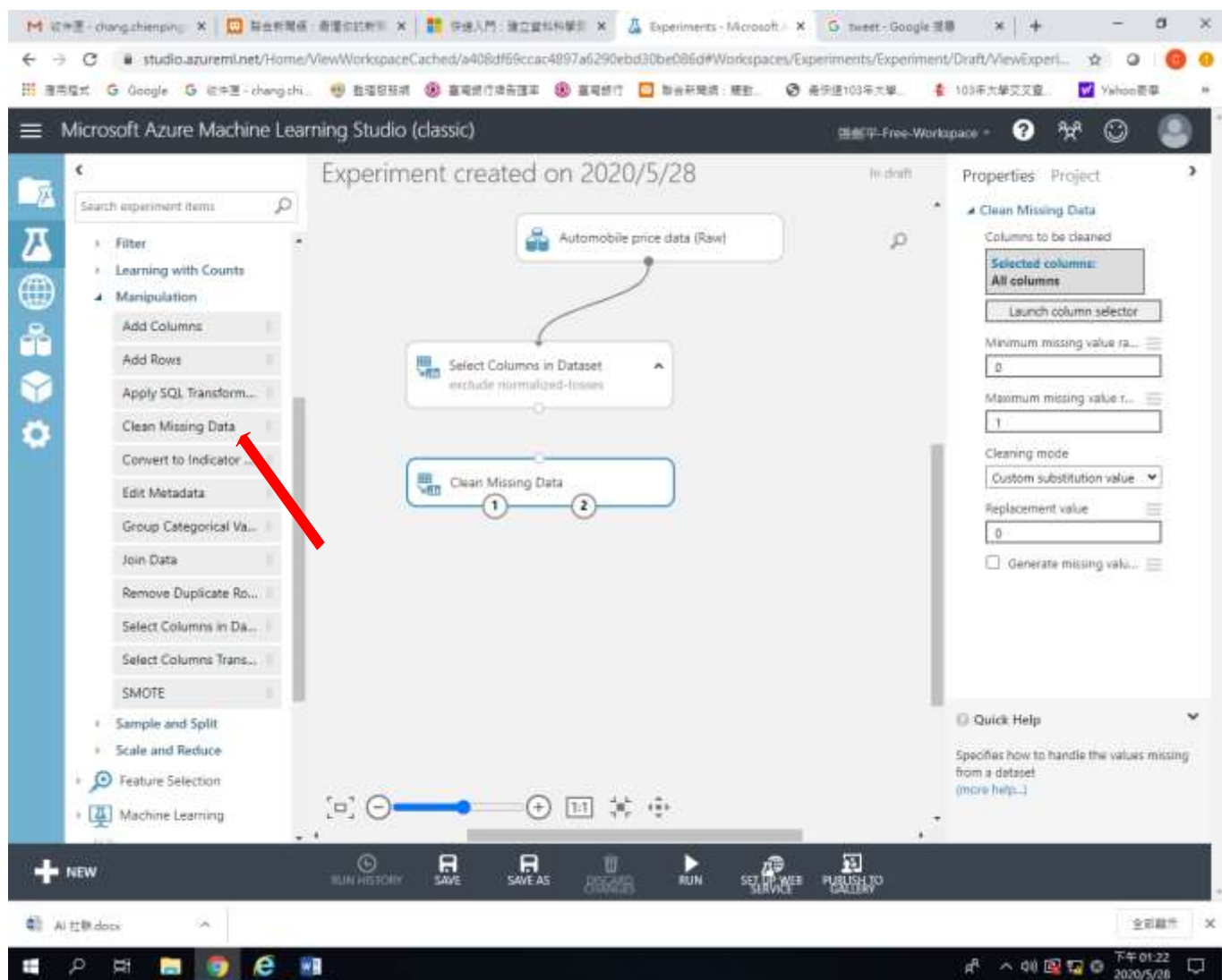


- 在左側按一下 [套用規則]
- 在 [開始於] 下，按一下 [所有資料行]。這些規則會指示[選取資料集中的資料行](#)傳遞所有資料行（但我們將排除的資料行除外）。
- 在下拉式清單中，選取 [排除] 和 [資料行名稱]，然後按一下文字方塊內部。資料行清單隨即顯示。選取 [normalized losses]，該資料行就會新增到文字方塊中。
- 按一下核取記號 ([確定]) 按鈕，以關閉資料行選取器（位於右下方）。



現在，[選取資料集中的資料行] 的屬性窗格指出它會傳遞資料集中的所有資料行，但 [normalized losses] 除外。

(4) 將清除遺漏的資料模組拖曳到實驗畫布，然後將其連線至選取資料集中的資料模組。在 [屬性] 窗格中，選取 [清除模式] 底下的 [移除整個資料列]。這些選項會指示清除遺漏的資料藉由移除含任何遺漏值的資料列來清除資料。按兩下模組，並輸入註解「移除遺漏值資料列」。



Microsoft Azure Machine Learning Studio (classic)

Experiment created on 2020/5/28

Filter

Learning with Counts

Manipulation

- Add Columns
- Add Rows
- Apply SQL Transform...
- Clean Missing Data
- Convert to Indicator...
- Edit Metadata
- Group Categorical Va...
- Join Data
- Remove Duplicate Ro...
- Select Columns in DA...
- Select Columns Trans...
- SMOTE

Sample and Split

Scale and Reduce

Feature Selection

Machine Learning

Automobile price data (Raw)

Select Columns in Dataset
exclude normalized-losses

Clean Missing Data

1 2

Properties: Project

Clean Missing Data

Columns to be cleaned

Selected columns:
All columns

Launch column selector

Minimum missing value ra...
0

Maximum missing value r...
1

Cleaning mode
Remove entire row

Quick Help

Specifies how to handle the values missing from a dataset
(more help...)

NEW

RUN HISTORY

SAVE

SAVE AS

DELETE

RUN

SET UP WEB SERVICE

PUBLISH TO GALLERY

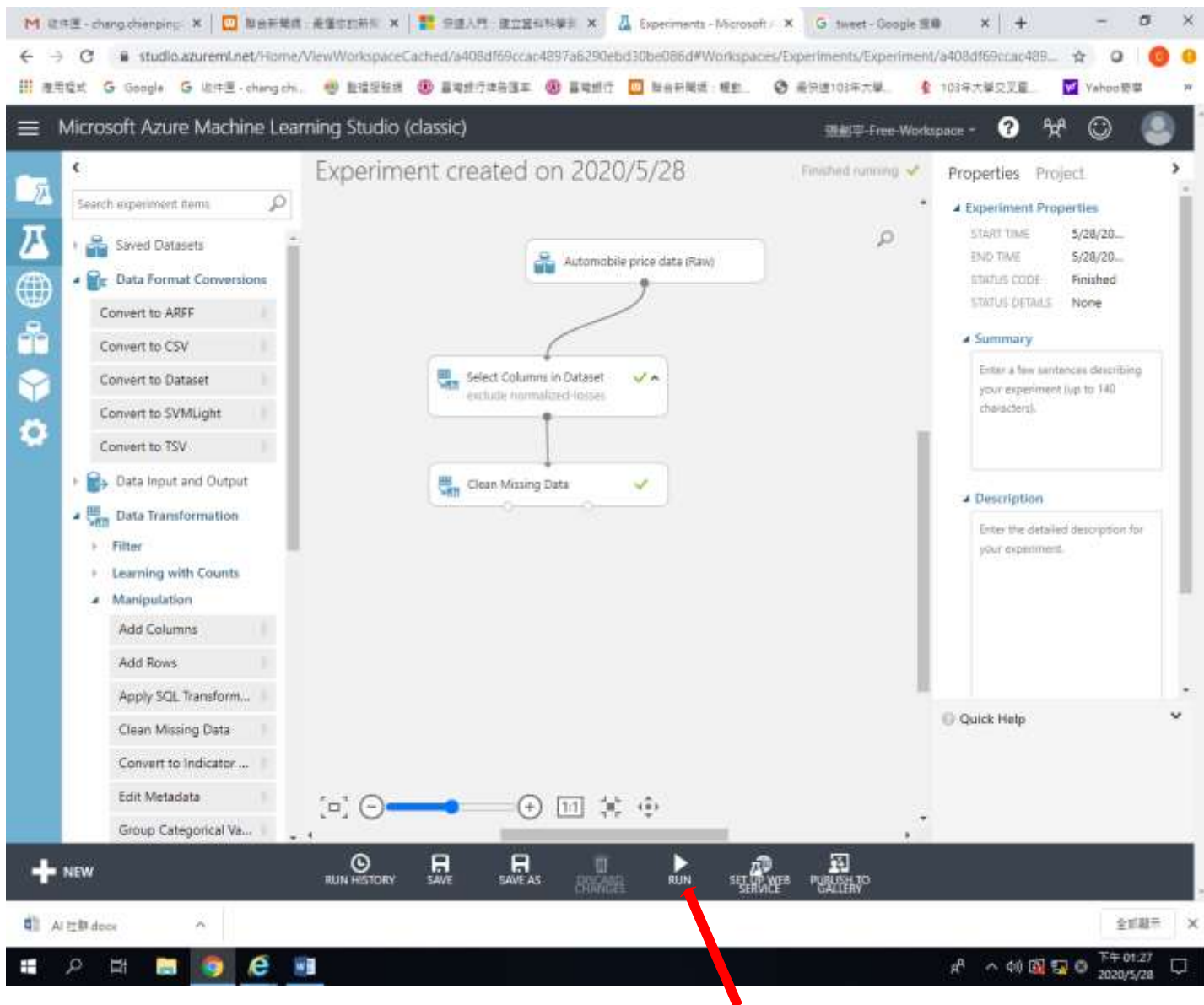
AI 社群.docx

全部顯示

下午 01:25 2020/5/28

(5) 按一下頁面底部的 [執行]，以執行實驗。

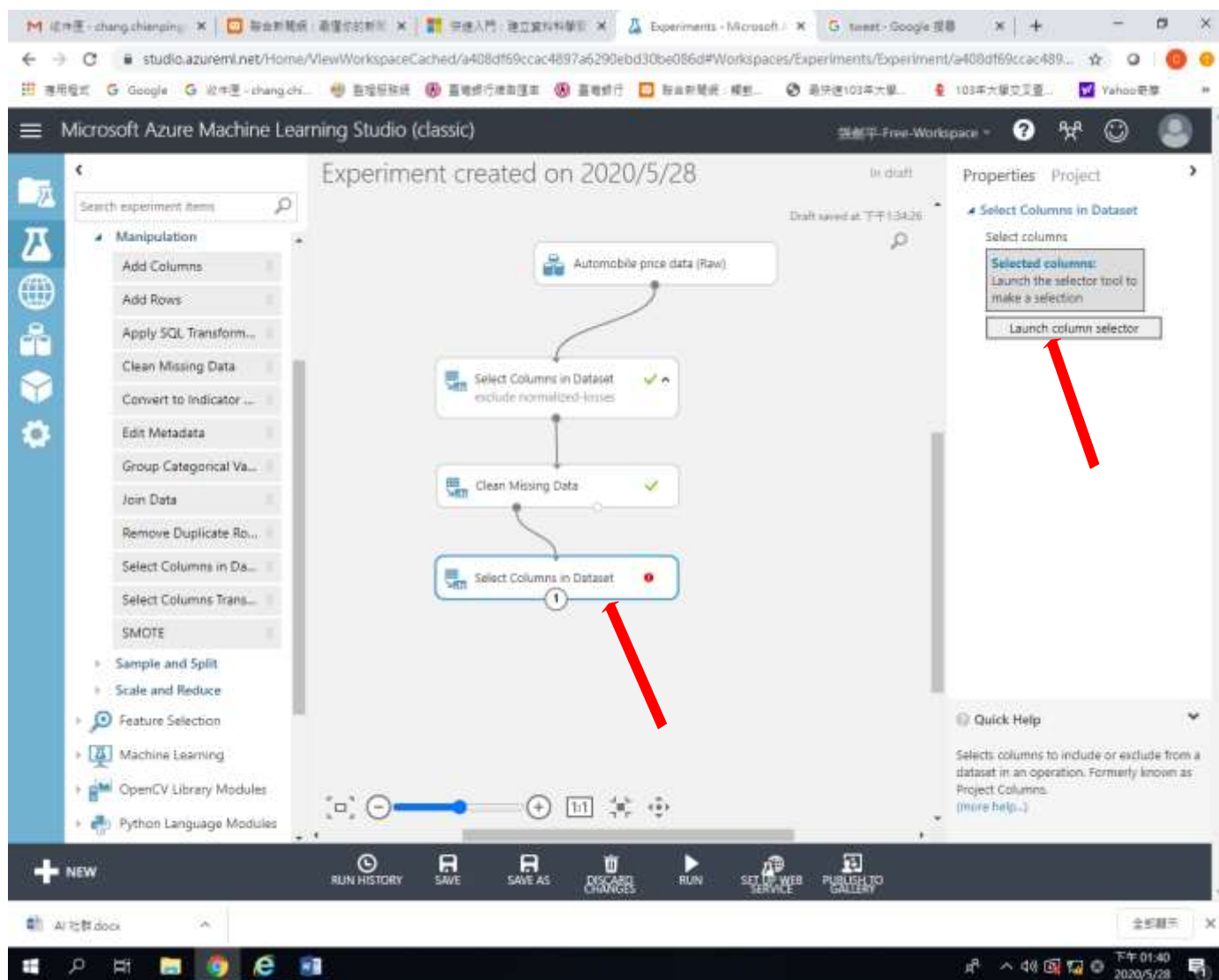
實驗執行完成時，所有模組都會呈現綠色核取標記，表示它們已順利完成。同時也請留意位於右上角的執行完成狀態。



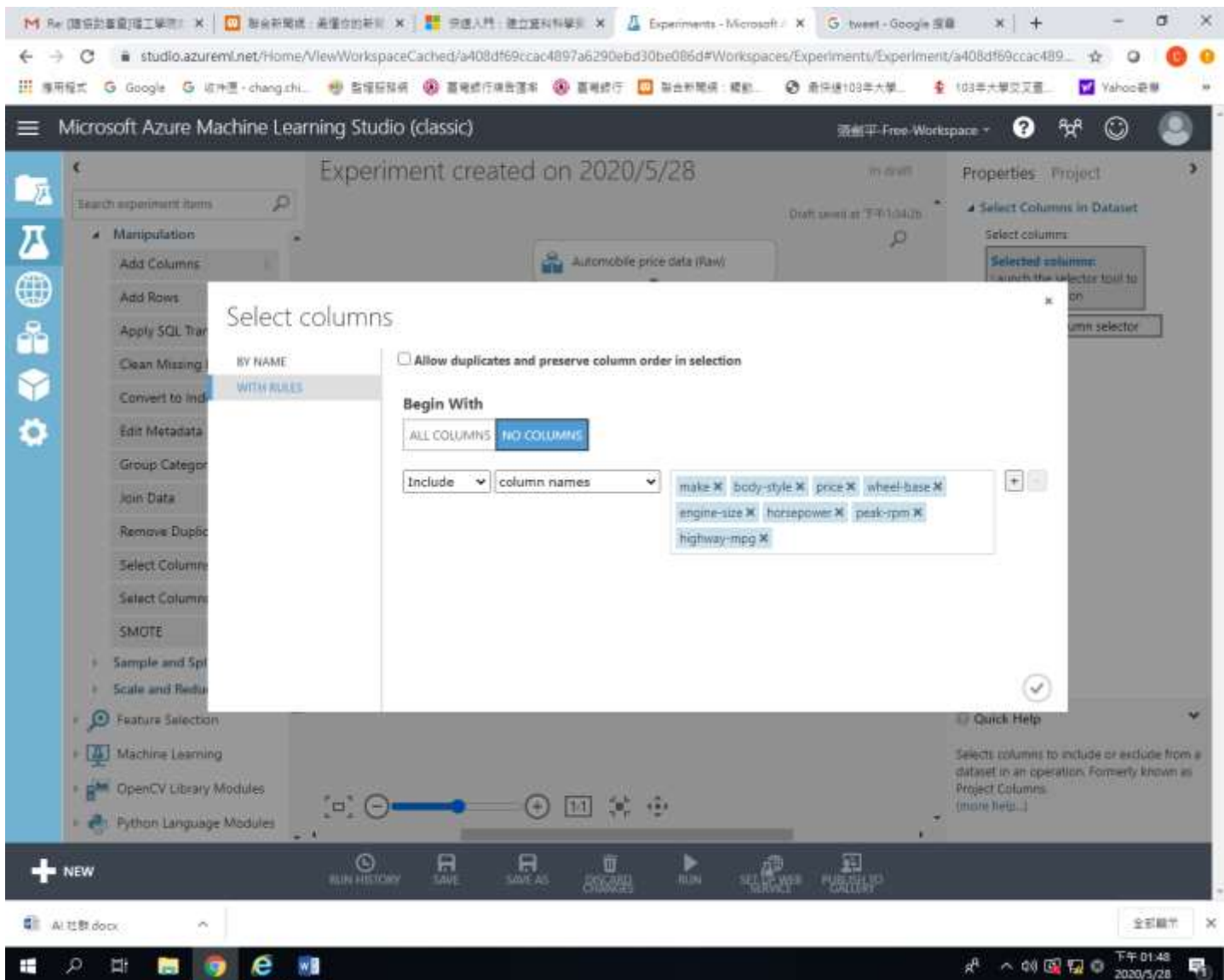
提示：為什麼要立即執行實驗？藉由執行實驗，就會從資料集傳遞我們資料的資料行定義（透過選取資料集中的資料行模組，及透過清除遺漏的資料模組）。這表示，我們連線到清除遺漏的資料的任何模組也會有相同的資訊。

開始分析步驟

1. 將另一個選取資料集中的資料行模組拖曳到實驗畫布。將清除遺漏資料模組的左側輸出連接埠連線到選取資料集中的資料行模組的輸入。



2. 按兩下模組，並輸入「選取要預測的功能」。
3. 按一下 [屬性] 窗格中的 [啟動資料行選取器]
4. 按一下 [套用規則]
5. 在 [開始於] 下，按一下 [無資料行]。在 [篩選] 資料列中，選取 [包含] 和 [資料行名稱]，然後在文字方塊中選取資料行名稱的清單。此篩選條件會指示模組不要傳遞我們未指定的任何資料行 (特性)
6. 按一下核取記號 ([確定]) 按鈕



選擇及套用演算法

現在，資料已備妥，訓練和測試是建構預測模型的要素。我們將使用前處理完的資料來訓練模型，然後測試模型，以檢驗它預測價格的精準度。

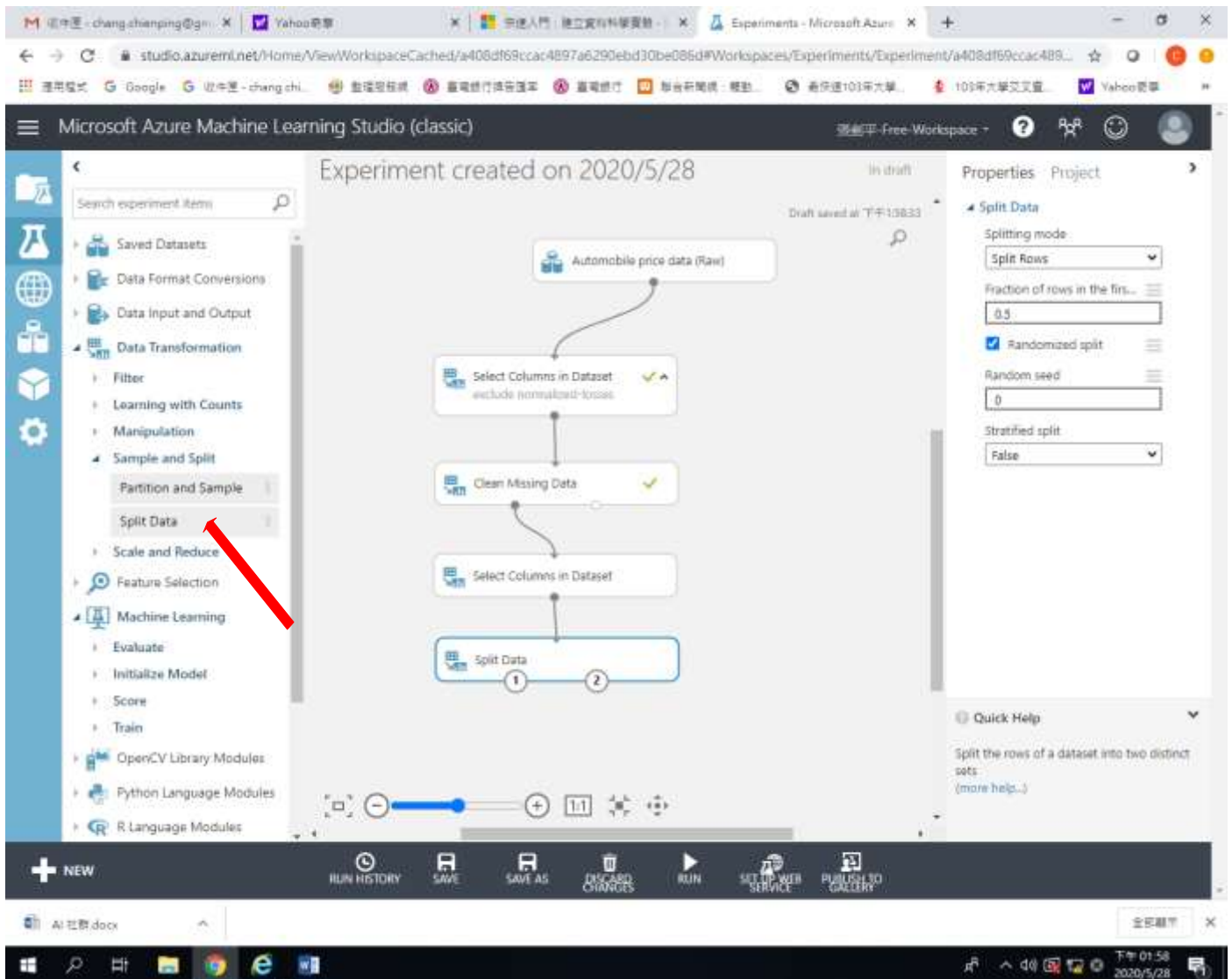
分類和**迴歸**是兩種受監督的機器學習演算法。「分類」可從一組已定義的類別預測答案，例如色彩（紅色、藍色或綠色）。「迴歸」可用來預測數字。

因為要預測價格，也就是一個數字，因此我們將使用迴歸演算法。在此範例中，我們將使用線性迴歸模型。

我們藉由提供一組包含價格的資料來訓練模型。模型會掃描的資料，然後尋找汽車功能與價格之間的關聯性。然後，我們將測試模型 - 提供它一組我們熟悉的汽車功能，接著檢驗模型預測已知價格的精準度。

我們將資料分割成個別的訓練和測試資料集，用來訓練和測試模型。

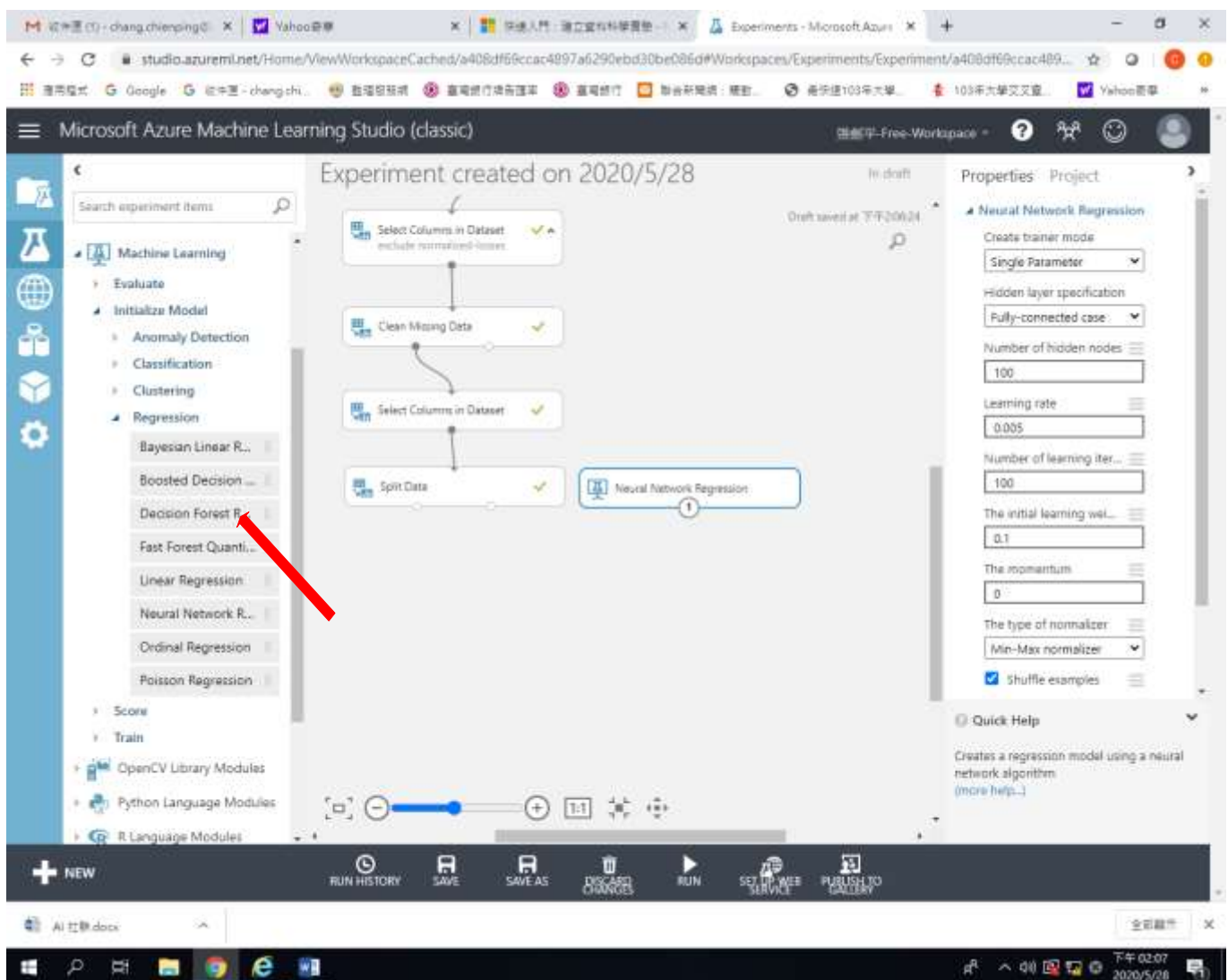
1. 選取 [分割資料](#) 模組並拖曳到實驗畫布，然後將其連線到最後一個[選取資料集](#)中的資料行模組。



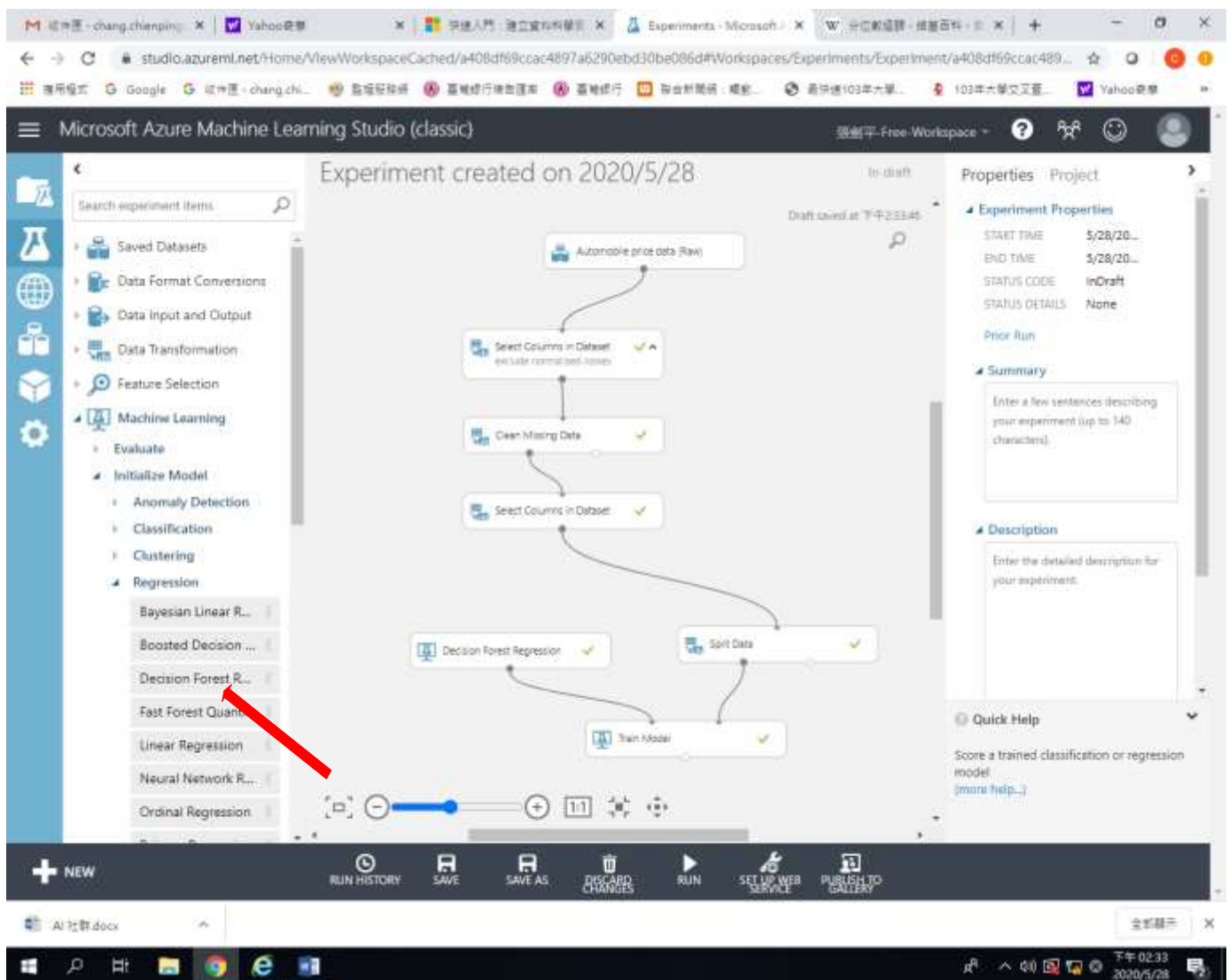
2. 按一下 [分割資料](#) 模組以選取它。尋找 [第一個輸出資料集中的資料列分數] (在畫布右邊的 [屬性] 窗格)，並將它設為 0.75。如此，我們將使用百分之 75 的資料來訓練模型，並保留百分之 25 做為測試之用。

The screenshot displays the Microsoft Azure Machine Learning Studio (classic) interface. The central workspace shows a workflow starting with 'Automobile price data (Raw)', followed by 'Select Columns in Dataset' (with 'exclude normalized losses' selected), 'Clean Missing Data', another 'Select Columns in Dataset', and finally the 'Split Data' module. The 'Split Data' module is highlighted with a red box and labeled with '1' and '2'. On the right, the 'Properties' pane for the 'Split Data' module is open. The 'Splitting mode' is set to 'Split Rows'. The 'Fraction of rows in the first...' is set to '0.75', which is highlighted with a red arrow. The 'Randomized split' checkbox is checked, and the 'Random seed' is set to '0'. The 'Stratified split' is set to 'False'. The bottom toolbar includes buttons for 'NEW', 'RUN HISTORY', 'SAVE', 'SAVE AS', 'DISCARD CHANGES', 'RUN', 'SET UP WEB SERVICE', and 'PUBLISH TO GALLERY'. The bottom status bar shows the date and time as '2020/5/28' and '下午 02:00'.

3. 執行實驗。執行實驗時，選取資料集中的資料行和分割資料模組會將資料行定義傳遞至我們接下來要新增的模組。
4. 若要選取學習演算法，請在畫布左側的模組調色盤中展開 [機器學習服務] 類別，然後展開 [初始化模型]。這會顯示數個可用來初始化機器學習演算法的模組類別。在此實驗中，請選取 [迴歸] 類別下的決策森林迴歸模組，然後將其拖曳到實驗畫布。(您也可以在調色盤搜尋方塊中輸入「決策森林迴歸」以尋找模組。)

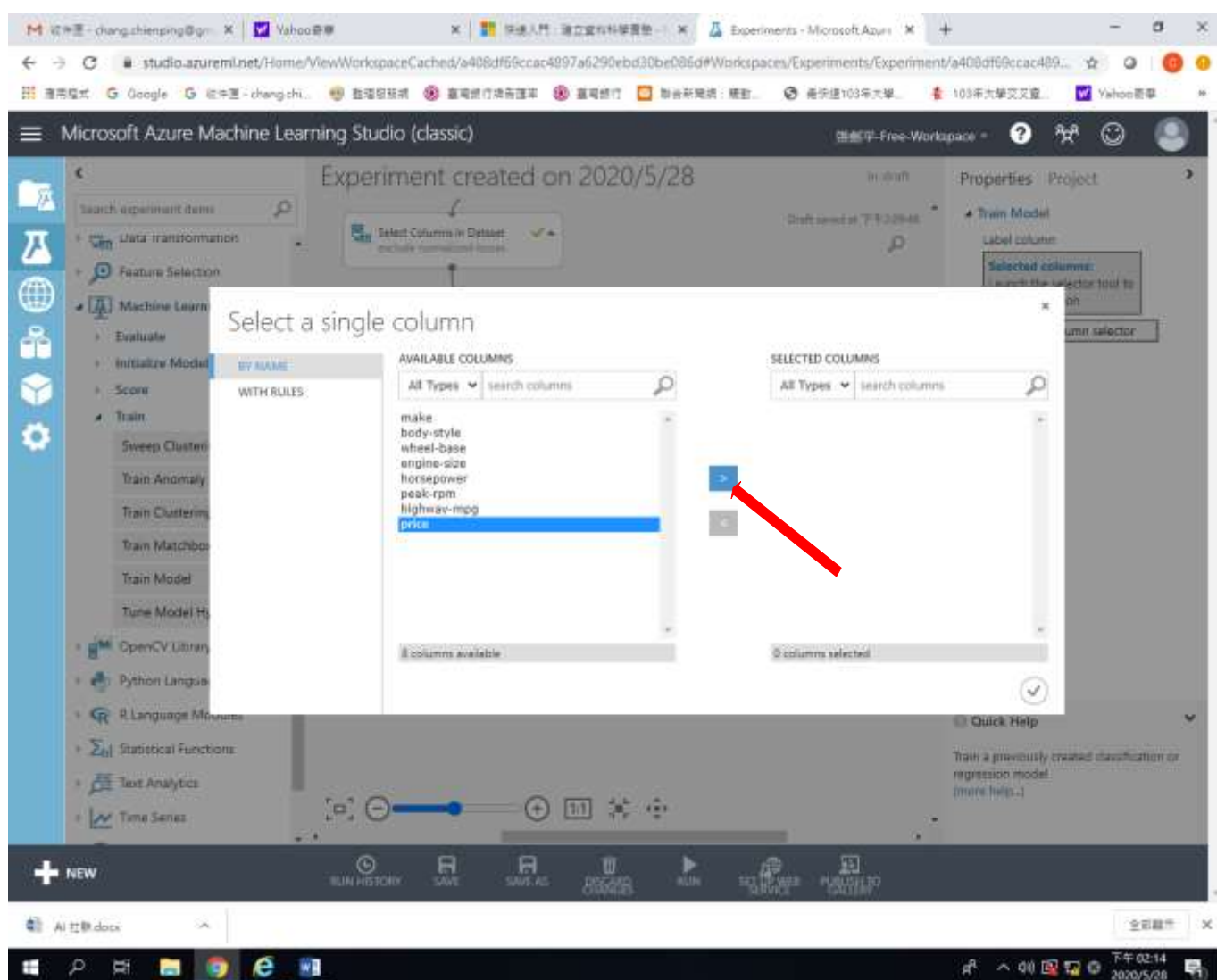


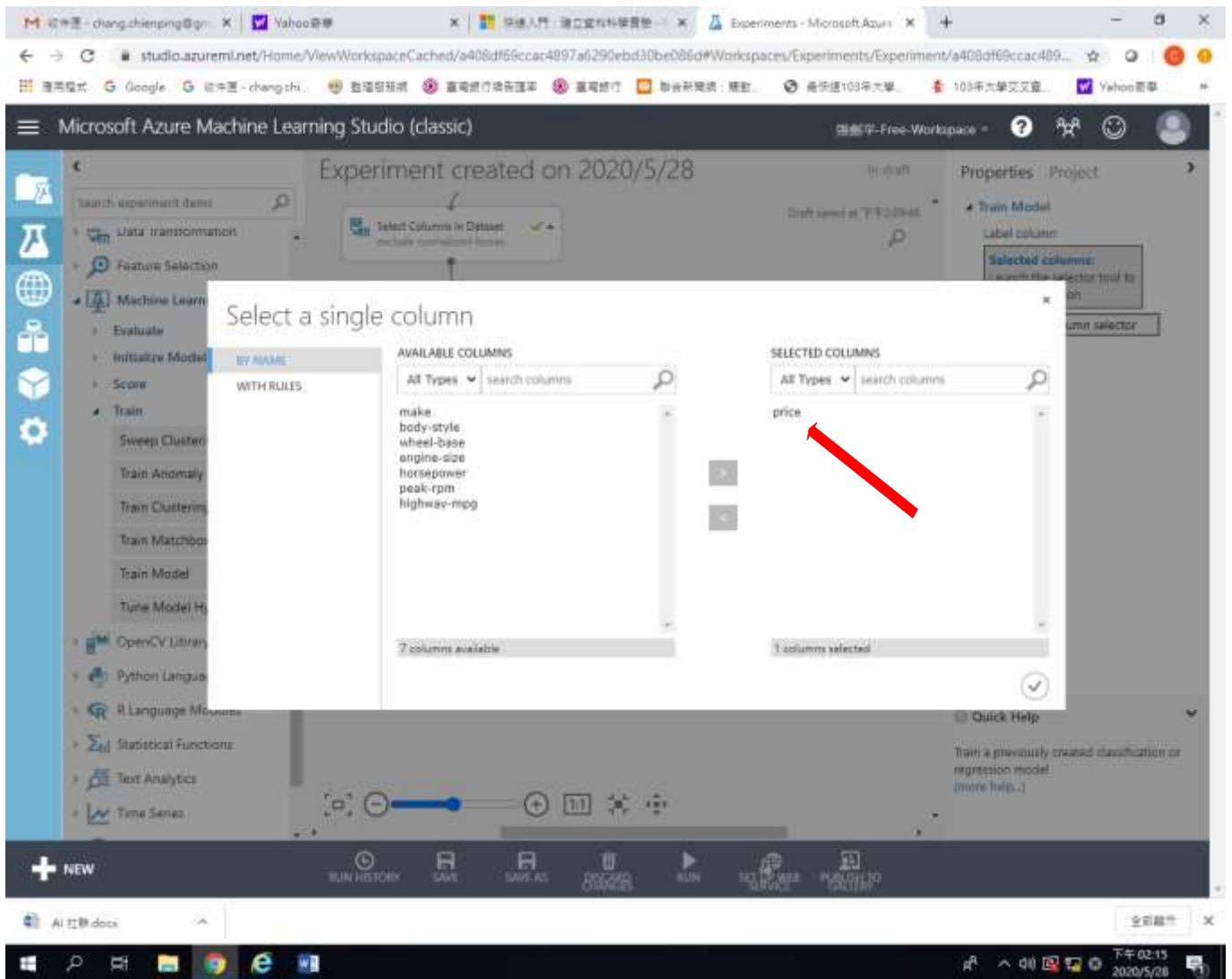
5. 找出訓練模型模組，並將其拖曳到實驗畫布。將決策森林迴歸模組的輸出連線至訓練模型模組的左側輸入，並將分割資料模組的訓練資料輸出（左側連接埠）連線至訓練模型模組的右側輸入。



6. 按一下 訓練模型 模組，按一下 [屬性] 窗格中的 [啟動資料行選取器]，然後選取 [價格] 資料行。**價格** 是我們的模型所將預測的值。

在資料行選取器中選取 [價格] 資料行（將它從 [可用的資料行] 清單 移到 [選取的資料行] 清單）。





7. 執行實驗。

我們現在有經過**訓練的迴歸模型**可用來為新的汽車資料評分，以藉此預測價格。

The screenshot displays the Microsoft Azure Machine Learning Studio (classic) interface. The main workspace shows a workflow diagram for an experiment created on 2020/5/28. The workflow consists of the following steps:

- Automobile price data (Raw)
- Select Columns in Dataset (exclude normalized-income)
- Clean Missing Data
- Select Columns in Dataset
- Decision Forest Regression
- Split Data
- Train Model

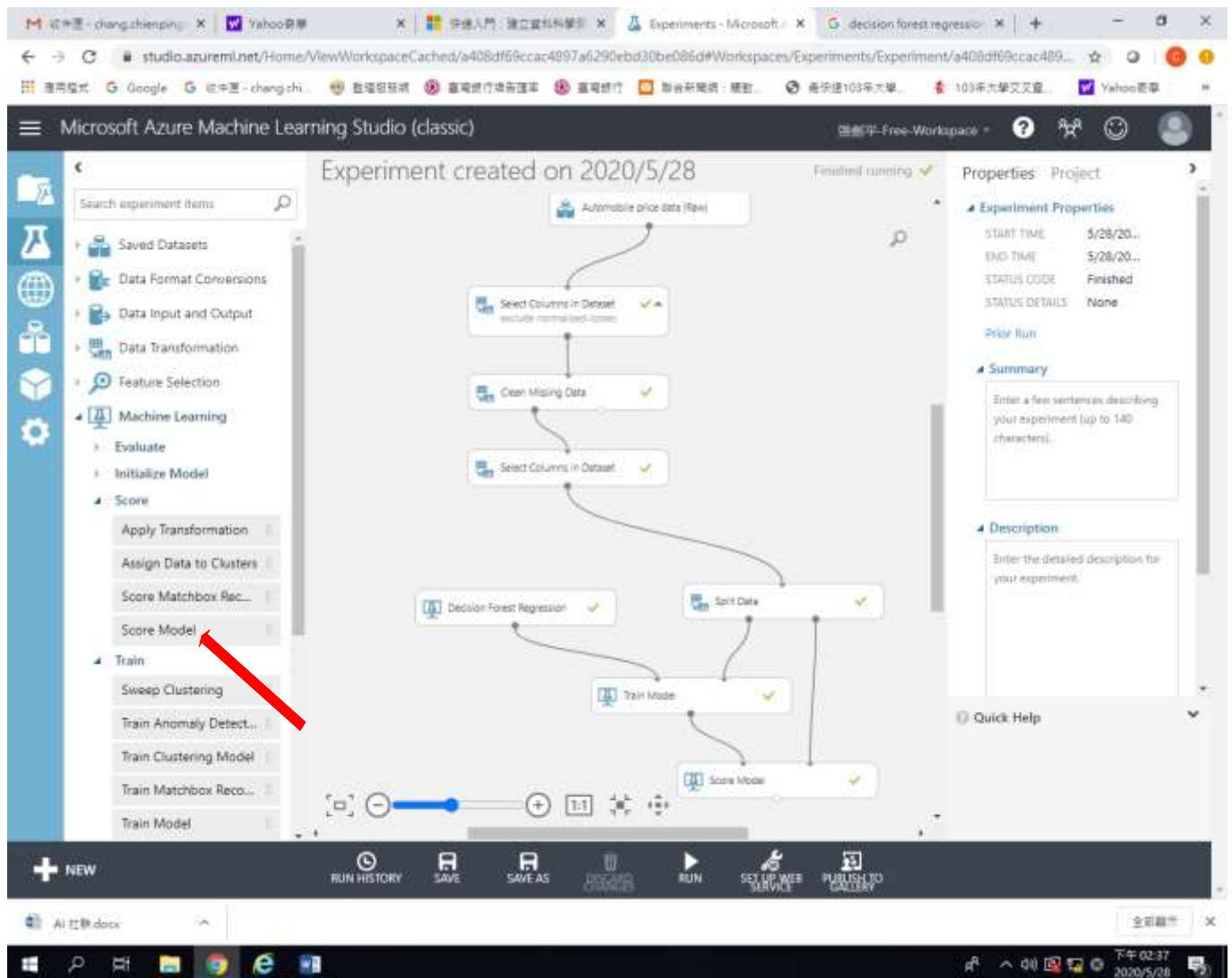
The left sidebar contains a search bar and a list of experiment items, including Saved Datasets, Data Format Conversions, Data Input and Output, Data Transformation, Feature Selection, Machine Learning, Evaluate, Initialize Model, Anomaly Detection, Classification, Clustering, and Regression. The right sidebar shows the Properties and Project tabs, with the Experiment Properties tab selected. It displays the experiment's start time (5/28/2020), end time (5/28/2020), status (inDraft), and status details (None). There are also sections for Summary and Description.

At the bottom of the interface, there is a toolbar with buttons for NEW, RUN HISTORY, SAVE, SAVE AS, DISCARD CHANGES, RUN, SETUP WEB SERVICE, and PUBLISH TO GALLERY. A red arrow points to the RUN button, indicating the next step in the process.

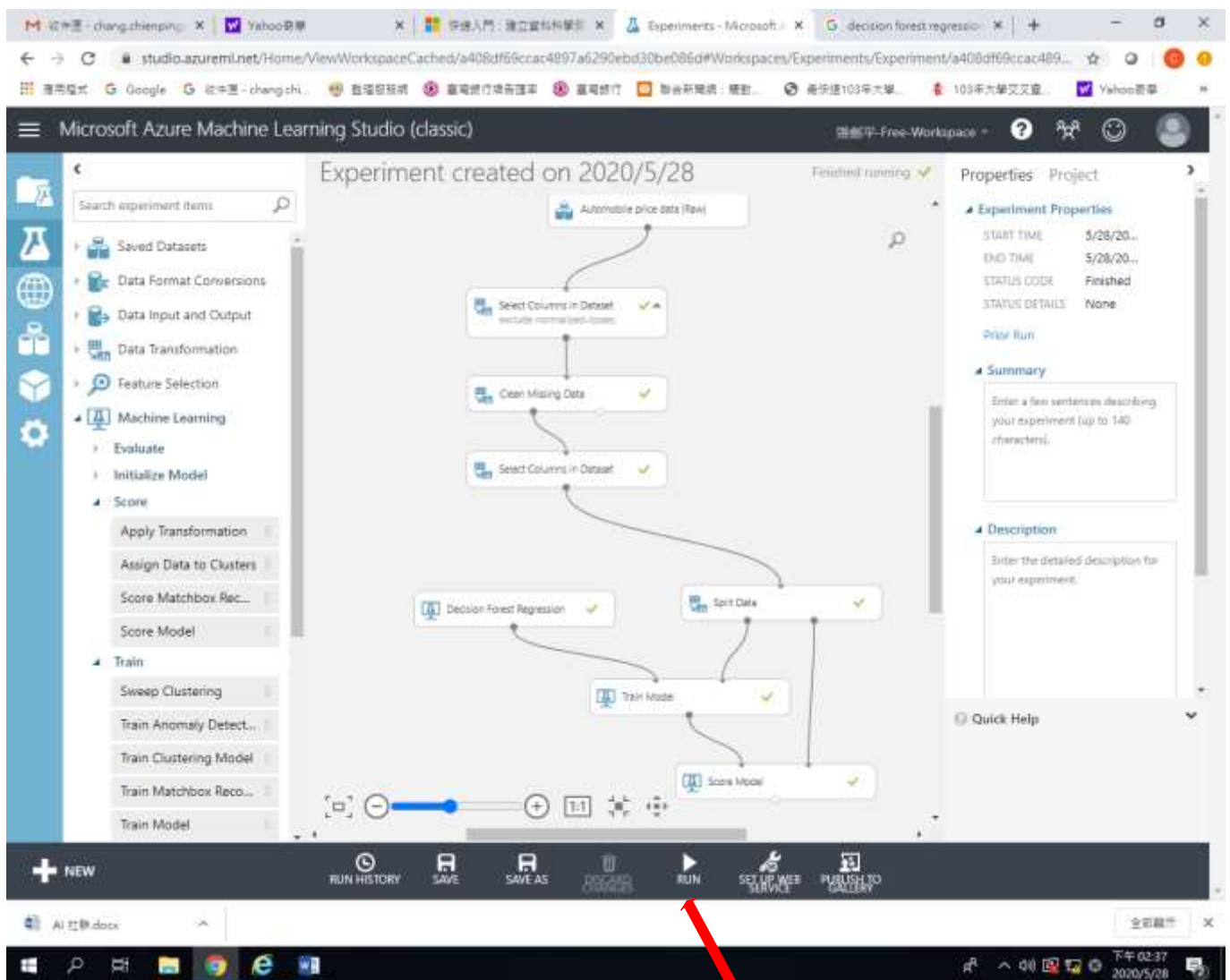
預測新的汽車價格

現在已完成使用百分之 75 資料模型的訓練，**我們可以用它來為其他百分之 25 的資料評分，以了解模型的運作是否理想。**

1. 找出評分模型模組，並將其拖曳到實驗畫布。將訓練模型模組的輸出連線至評分模型的左側輸入連接埠。將分割資料模組的測試資料輸出（右側連接埠）連線至評分模型的右側輸入連接埠。



2. 執行實驗，然後按一下評分模型的輸出連接埠並選取 [視覺化]，以檢視評分模型模組的輸出。



Microsoft Azure Machine Learning Studio (classic)

Experiment created on 2020/5/28

Automobile price data (Raw)

Select Columns in Dataset

Clean Missing Data

Select Columns in Dataset

Decision Forest Regression

Split Data

Train Model

Score Model

Properties: Project

Score Model

Append score column...

START TIME 5/28/2020 ...

END TIME 5/28/2020 ...

ELAPSED TIME 0:00:00.000

STATUS CODE Finished

STATUS DETAILS Task output was present in output cache

Download

Save as Dataset

Save as Trained Model

Save as Transform

Visualize

Generate Data Access Code...

Open in New Notebook

View Log

Edit Comment

Run selected

Help

NEW

RUN HISTORY

SAVE

SAVE AS

DISCARD CHANGES

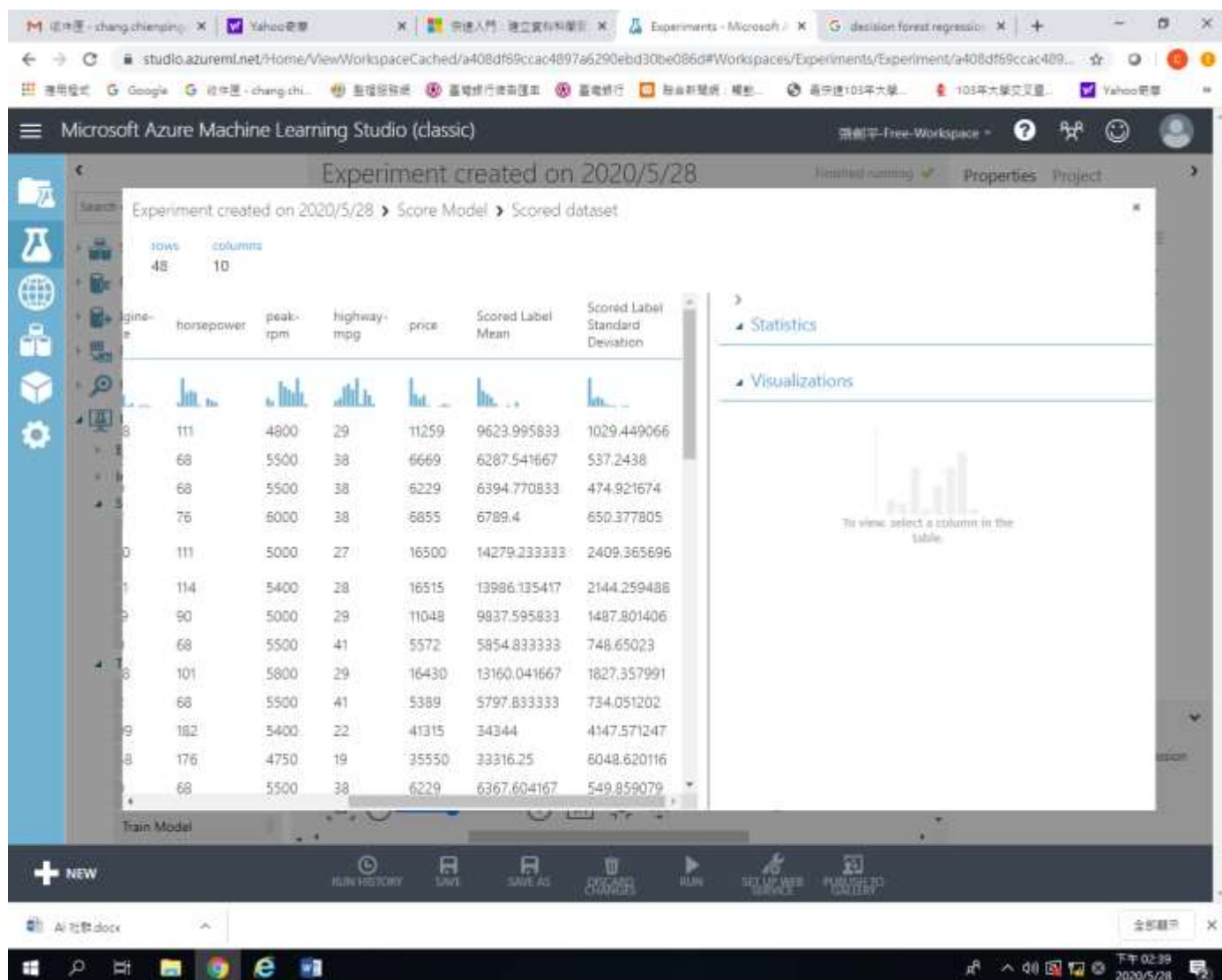
RUN

SET UP WEB SERVICE

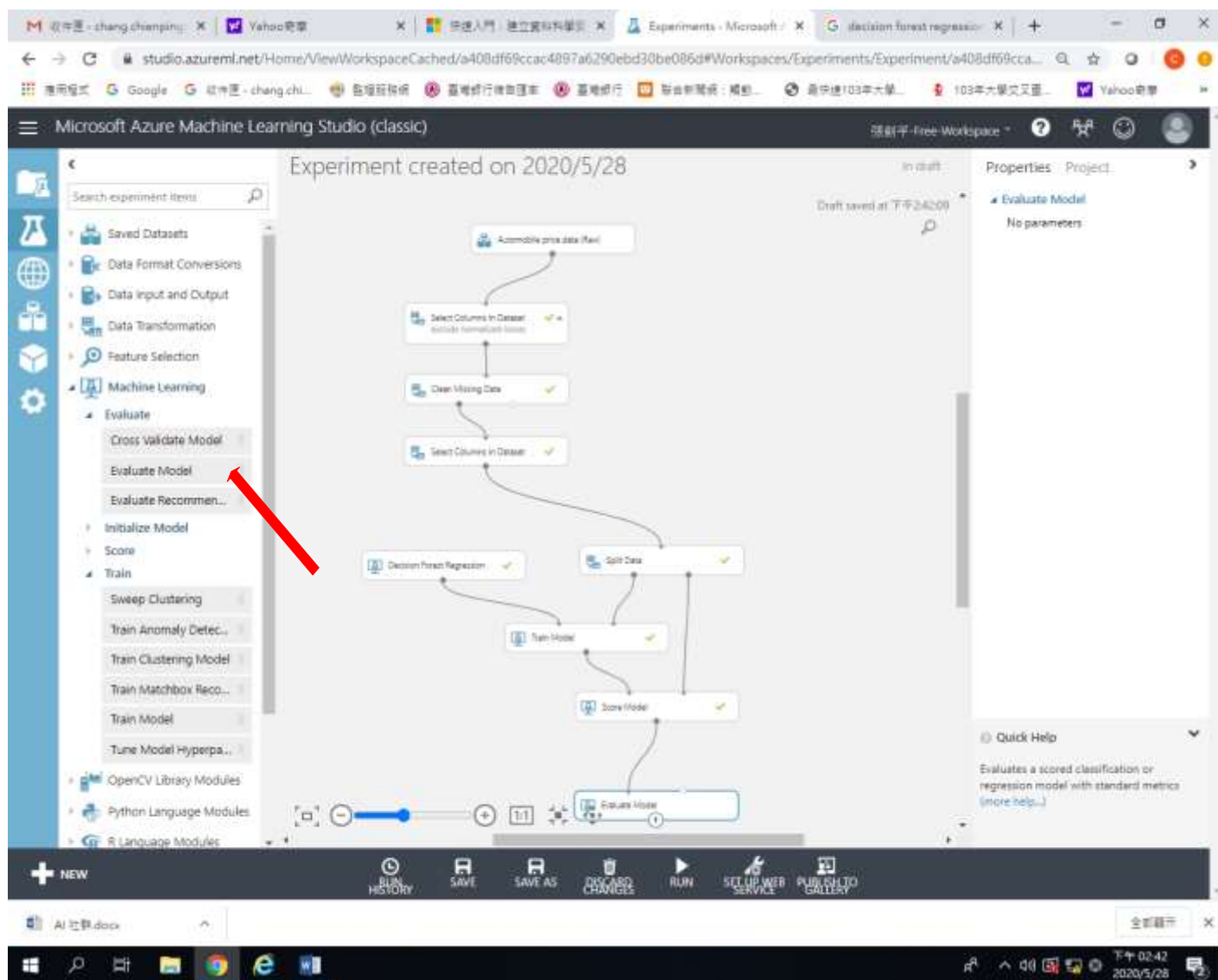
AI 社群.docx

下午 02:38 2020/5/28

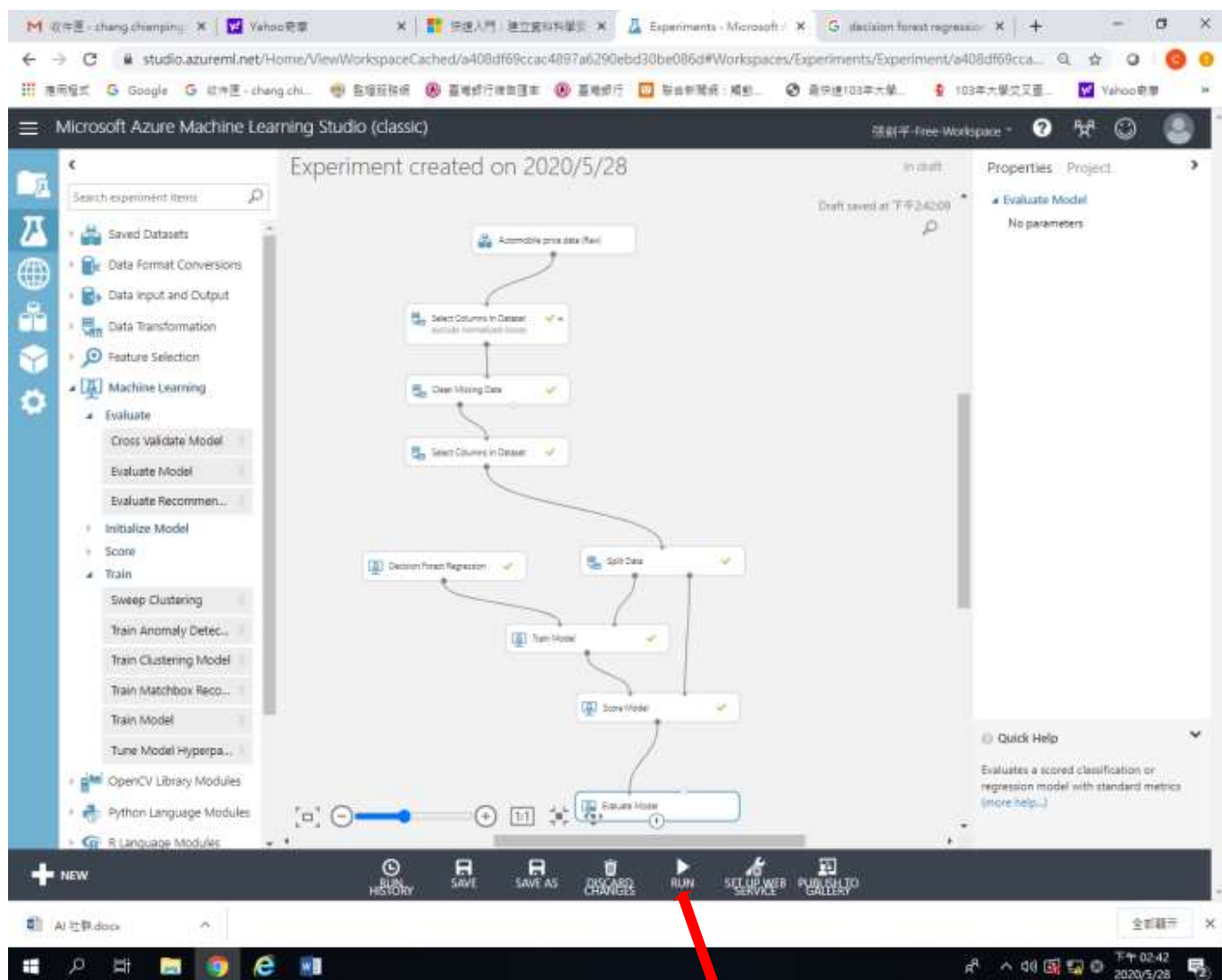
輸出會顯示價格的預測值，以及來自測試資料的已知值。



3. 最後，我們要測試結果的品質。選取評估模型模組，並將其拖曳至實驗畫布，然後將評分模型模組的輸出連線到評估模型的左側輸入。實驗最終應呈現如下：



執行實驗。



若要檢視評估模型模組的輸出，請按一下輸出連接埠，然後選取 [視覺化] 。

The screenshot displays the Microsoft Azure Machine Learning Studio (classic) interface. The main workspace shows an experiment workflow titled "Experiment created on 2020/5/28". The workflow includes the following modules: "Select Columns in Dataset", "Split Data", "Decision Forest Regression", "Train Model", "Score Model", and "Evaluate Model". The "Evaluate Model" module is selected, and its output port is highlighted with a red circle. A context menu is open over the "Evaluate Model" module, showing options such as "Delete", "Copy", "Cut", "Paste", "Evaluation results", "View Log", "Edit Comment", "Run selected", and "Help". The "Evaluation results" option is highlighted with a red arrow. The right sidebar shows the "Properties" tab for the "Evaluate Model" module, displaying details like "START TIME", "END TIME", "ELAPSED TIME", "STATUS CODE", and "STATUS DETAILS". The bottom status bar shows the "RUN" button and other controls.

會顯示下列統計資料：

- **平均絕對誤差 (MAE)：**絕對誤差的平均值（「誤差」是指預測值與實際值之間的差異）。
- **均方根誤差 (RMSE)：**對測試資料集所做之預測的平方誤差的評分根平均值。
- **相對絕對誤差：**相對於實際值與所有實際值之平均值之間的絕對差異的絕對誤差平均值。
- **相對平方誤差：**相對於實際值與所有實際值之平均值之間的平方差異的平方誤差平均值。
- **決定係數：**也稱為 **R 平方值**，這是一個統計度量，可指出模型對於資料的適用程度。

針對每個誤差統計資料，越小越好。值越小，表示預測越接近實際值。就 [決定係數] 而言，其值愈接近一 (1.0)，預測就愈精準。

