

SET UP

1. Create Azure subscription.
2. <https://portal.azure.com/> you should be directed here. If you already have a subscription, head straight here.
3. Create a machine learning resource.
 - a. Search Machine Learning
 - b. Click create on the top left
4. Creation:
 - a. Ensure correct subscription is connected if you have multiple to choose from
 - b. For resource group – click create new. Name your resource group as you wish (maybe "MLworkshopWWC")
 - c. Name your workspace
 - d. Select "Australia East" as your region
 - e. If everything else has prepopulated for you, skip to i.
 - f. Click create new storage account and give it a name. Leave the replication as Locally Redundant Storage (LRS)
 - g. Key Vault and Application Insights will prefill themselves once you create a storage account
 - h. Leave container registry as None
 - i. Hit review and create!

This will begin to deploy the resources – should take approx 5mins. Click "go to resource" once deployed.

Machine learning ...

Create a machine learning workspace

Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription * ⓘ Visual Studio Enterprise Subscription

Resource group * ⓘ wwcML

[Create new](#)

Workspace details

Specify the name and region for the workspace.

Workspace name * ⓘ workspaceML ✓

Region * ⓘ Australia East

Storage account * ⓘ (new) workspaceml1477540488

[Create new](#)

Key vault * ⓘ (new) workspaceml3364164634

[Create new](#)

Application insights * ⓘ (new) workspaceml5270968088

[Create new](#)

Container registry * ⓘ None

[Create new](#)

[Review + create](#)

[< Previous](#)

[Next : Networking](#)

Microsoft Azure Search resources, services, and docs (G+/)

Home > wwcmlworkspace Machine learning

Search (Ctrl+/) « Download config.json Delete

Overview

- Activity log
- Access control (IAM)
- Tags
- Diagnose and solve problems
- Events

Settings

- Private endpoint connections
- Properties
- Locks

Monitoring

- Alerts
- Metrics

Essentials

- Resource group: wwcML
- Location: Australia East
- Subscription: Visual Studio Enterprise Subscription
- Subscription ID: [REDACTED]

JSON View

- Studio web URL: [https://ml.azure.com/\[REDACTED\]/workspace/738cws1...](https://ml.azure.com/[REDACTED]/workspace/738cws1...)
- Storage: wwcmlstorage
- Registry: ...
- Key Vault: wwcmlworkspace6301780899
- Application Insights: wwcmlworkspace7510817791

Manage your machine learning lifecycle

Use the Azure Machine Learning studio to build, train, evaluate, and deploy machine learning models. [Learn more](#)

[Launch studio](#)

HIT THE LAUNCH STUDIO! ML Studio will open in a new tab

ML STUDIO - COMPUTE RESOURCES

1. First thing will be to create compute resources. Lets create a virtual machine for us to use.
 - a. Find compute on the left panel
 - b. Click "Create compute instance"

Configure required settings

Compute name * ⓘ

akankshamalik961

Location ⓘ

australiaeast

Virtual machine type ⓘ

☒ CPU ☐ GPU

Virtual machine size ⓘ

☒ Select from recommended options ☐ Select from all options

Total available quota: 20 cores ⓘ

	Name	Category	Workload types	Av...	Cost
<input type="radio"/>	Standard_DS2_v2 2 cores, 7GB RAM, 14GB storage	General purpose	Development on Notebooks (or other IDE) and light weight testing	20 co...	\$0.17/hr
<input checked="" type="radio"/>	Standard_DS3_v2 4 cores, 14GB RAM, 28GB storage	General purpose	Classical ML model training, AutoML runs, pipeline runs (default compute)	20 co...	\$0.34/hr

Create

Back

Next

[Download a template for automation](#)

Cancel

2. Similarly create compute clusters. This will be a group of machines used for certain (bigger) tasks.
 - a. Head into the compute cluster tab and create new. The first screen should prefill. Click next.

☒ Virtual Machine
 ☐ Advanced Settings

Select virtual machine

Select the virtual machine size you would like to use for your compute cluster.

Location ⓘ
australiaeast

Virtual machine priority ⓘ
☒ Dedicated ☐ Low priority

Virtual machine type ⓘ
☒ CPU ☐ GPU

Virtual machine size ⓘ
☒ Select from recommended options ☐ Select from all options

Total available quota: 12 cores ⓘ

	Name ↑	Category	Workload types	Available ... ⓘ	Cost ⓘ
<input type="radio"/>	Standard_DS2_v2 2 cores, 7GB RAM, 14GB storage	General purpose	Development on Notebooks (or other IDE) and light weight testing	12 cores	\$0.17/hr
<input checked="" type="radio"/>	Standard_DS3_v2 4 cores, 14GB RAM, 28GB storage	General purpose	Classical ML model training, AutoML runs, pipeline runs (default compute)	12 cores	\$0.34/hr
<input type="radio"/>	Standard_DS12_v2	Memory optimized	Training on large datasets (>1GB) parallel run steps, batch	12 cores	\$0.40/hr

Back
Next
Cancel

- Name your cluster.
- Create as many nodes as you would like (more nodes = more money). Make sure the min is 0.
- Choose an appropriate time out setting for yourself (will scale down to 0 and not cost you when not being used)
- Click create

DATA LOAD

- Lets create some data! First we'll go through local files method.
 - In the repo you will find a dataset called "AB_NYC_2019.csv". Please download this file – it's a table of airBnB data from NY
 - In the left panel, find the datasets tab. In the tab, click "Create Dataset" and pick "From local files"
 - Give it a name (you will see this in the workspace - eg. AirBNB2019)
 - Select appropriate data type. As this is a csv, it is tabular.
 - Click "upload files" and select the file you downloaded.
 - When you click next, it will validate and upload.
 - Check the setting and preview is correct. Click next.
 - In the schema, you can check the data types are interpreted correctly and also remove any columns you do not want uploaded.
 - In the final step, check all the details and click "Profile this dataset". This will carry out some basic statistical analysis on your data.
 - Select one of the computes to carry out the profiling.
 - Click create!
- Lets try from Open Datasets
 - In the create a dataset option, chose "from an open dataset"
 - Search for NYC. You will see the NYC taxi dataset.
 - Here you will be asked to select a date range for the dataset. Pick the last month.

- d. Click create!
3. You should see all the datasets you have created now.

We now have machines and data ready to go. This is all we need to create ML models!

AUTOML

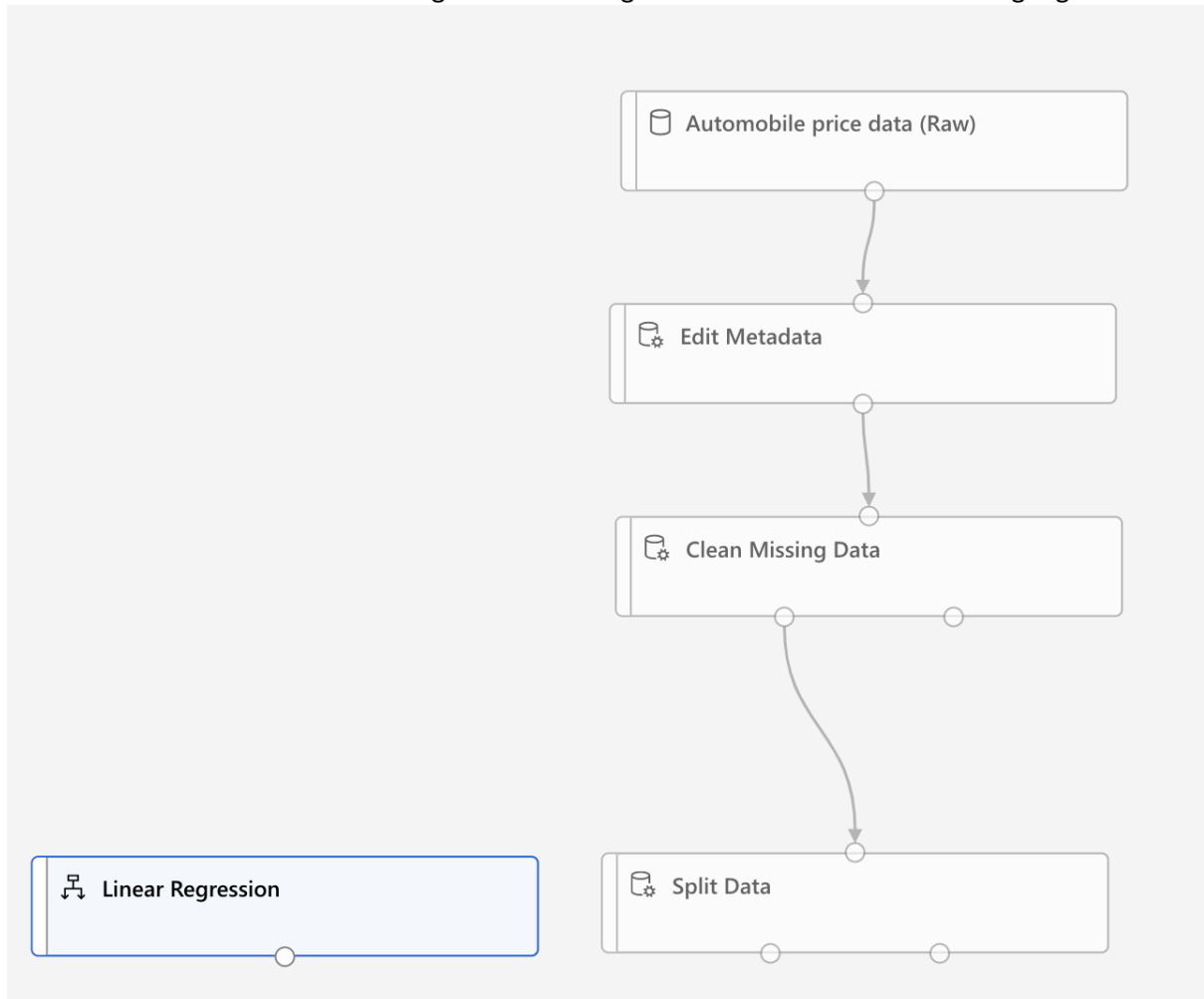
First we are going to set up an AutoML run. This can run in the background for us.

1. Pick AutoML from the panel on the left.
2. Select the AirBNB dataset that should pre-populate. Click Next.
3. Select "create new" experiment.
4. Give your experiment a name
5. Choose your target column. With the AirBNB data we are trying to predict the price of a rental based on some features - thus the target is the price column.
6. Select the compute cluster to carry out the work
7. Choose the type of ML you want to carry out - here it is regression as we are predicting the exact price based on features.
8. And click Finish!
9. This run can take some time so we will come back to see the completed run in the end (if it finishes in time!)

DESIGNER

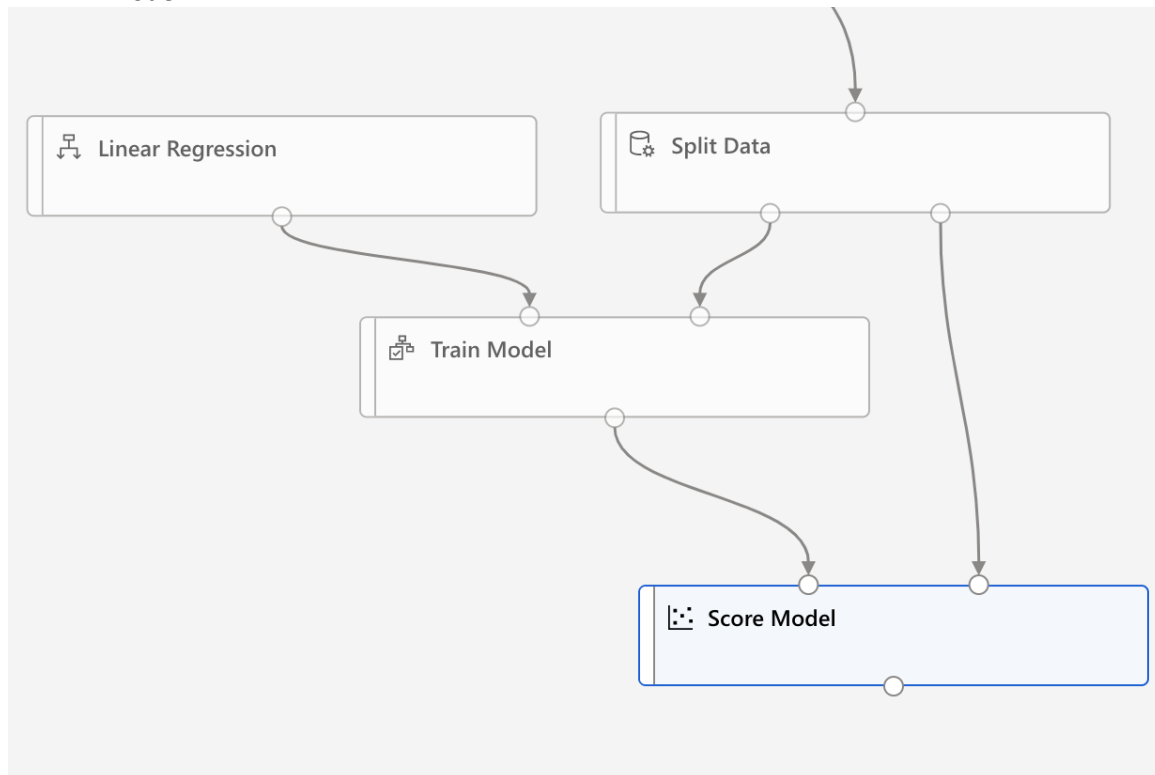
1. In the left panel, choose Designer
2. Replace "Pipeline-Created-on-06-24-2021" with a name for your pipeline
3. Click the settings symbol by the name, which will open a pop up on the right.
 - a. Select and set a compute target with which this pipeline will be evaluated.
4. Open "Sample Datasets" in modules
5. Drag "Automobile Price Data (Raw)" into the canvas
6. Click on the module in the canvas which will open a pop up on the right. In the pop up you can go into Outputs to have a look through the data available in this dataset.
7. Open the "Data Transformations" modules and drag out the "Edit Metadata" module onto the canvas.
8. Connect "Automobile Price Data (Raw)" to "Edit Metadata" (by clicking on the circle at the bottom of the module, this will create an arrow you can drag down to edit metadata)
9. Click "Edit Metadata" and in the pop up on the right choose "edit column"
 - a. Here you can remove any columns you want from the data.
 - b. So we are going to Include, "all columns" from the dropdown to start with.
 - c. Click the + button, now we can choose to exclude from the dropdown.
 - d. We are going to remove by "column name" and remove normalized-losses. Click save.
10. Next clean the data by dragging in "Clean Missing Data" module (also found in "Data Transformations"). Connect "Edit Metadata" to this module (by clicking on the circle at the bottom of the module, this will create an arrow you can drag down to edit metadata)
 - a. Click the module for the pop up on the right.

- b. We are going to click edit column and choose to clean all columns. (same as step 8b)
 - c. Open the Cleaning Mode dropdown. Here are the different ways of cleaning data. For today we are going to remove the entire row if there's any missing data.
- 11. Next drag out "Split Data" module. Connect this with the first circle in the "Clean Missing Data" module.
 - a. Edit the fraction of split to 0.7
 - b. This step has split our data into training and testing.
- 12. We now have our data ready to be trained. The other thing we need is the algorithm with which to train. Drag out "Linear Regression" from "Machine Learning Algorithms"

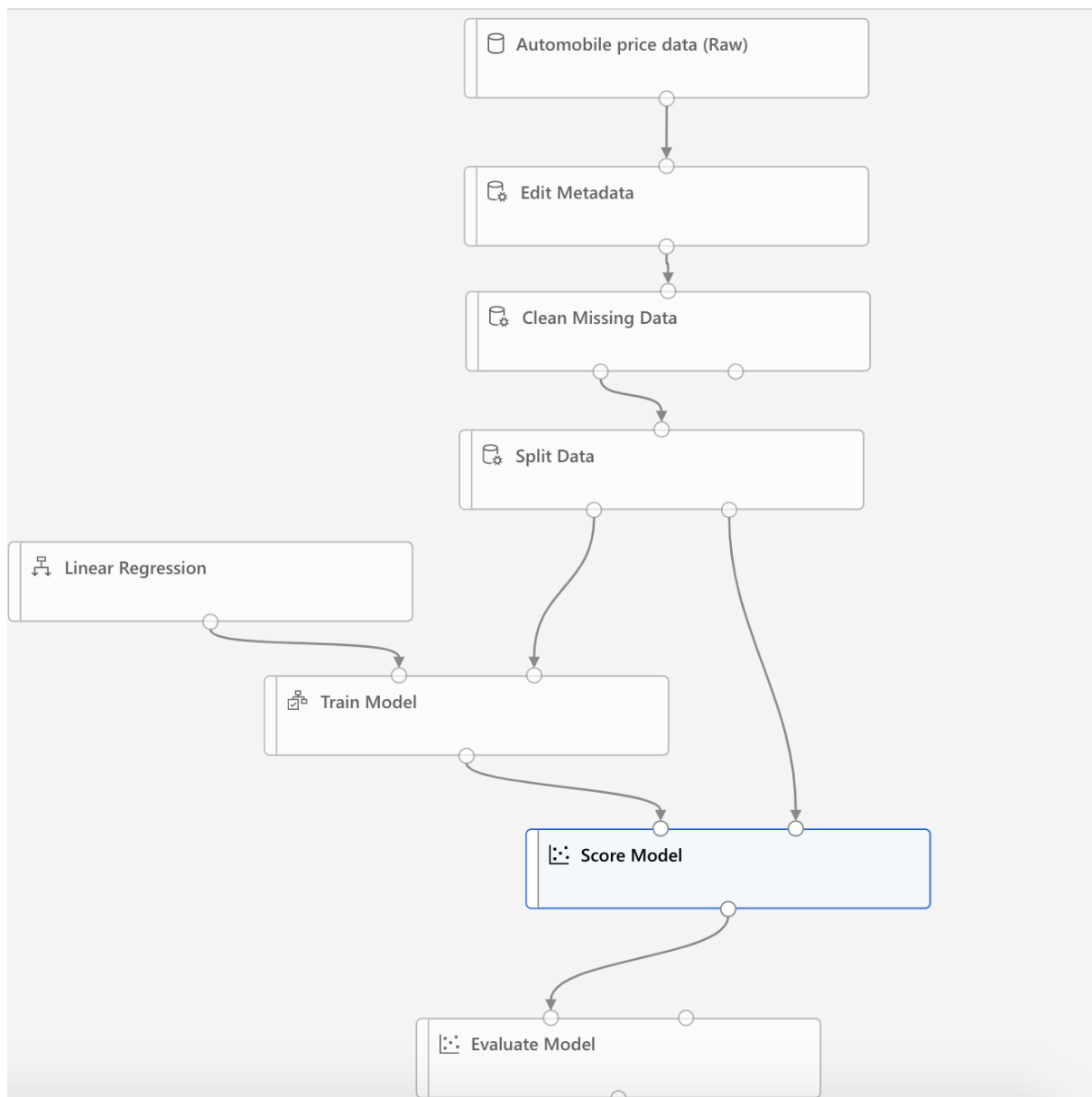


- 13. Now that we have an algorithm chosen and the data ready, it's time to train. Drag out "Train model" module.
 - a. If you hover over the circles at the top of the "Train Model" it will briefly tell you what each of them expects. One expects an algorithm and one a dataframe.
 - b. As such connect "Linear Regression" to the first circle.
 - c. Connect the left/first circle of "Split Data" to the second circle in "Train Model"
 - d. Now click into "Train Model" and head over to the pop up on the right.
 - e. As we did in AutoML, we need to tell it what column to target. Click "Edit column"

- f. In the dropdown options, chose "price" as we will aim to predict the price of a car based on its features.
- 14. Once the model is trained on the training set of 70% of the data we gave it, we want to see how well it did. So we are going to score it against test data we set aside in the "Split data" (the leftover 30%)
 - a. Drag out the module "Score Model"
 - b. Connect the outcome circle of "Train Model" to the first circle in the "Score Model"
 - c. Connect the second/right circle of "Split Data" to the second circle in "Score Model"



- 15. As a final step, we would like to evaluate the overall model. Drage out "Evaluate Model" and connect it with "Score Model"



16. Hit Submit!