



Data Import

Citizen Analytics – An Initiative by Data Science Team

START ►

© 2020 Petroliaam Nasional Berhad (PETRONAS)

All rights reserved. No part of this document may be reproduced in any form possible, stored in a retrieval system, transmitted and/or disseminated in any form or by any means (digital, mechanical, hard copy, recording or otherwise) without the permission of the copyright owner.

Learning Objectives

By the end of this module, you will be able to:



01

Import datasets from multiple sources

02

Create new datasets manually in ML Studio

03

Unpack zipped datasets within ML Studio

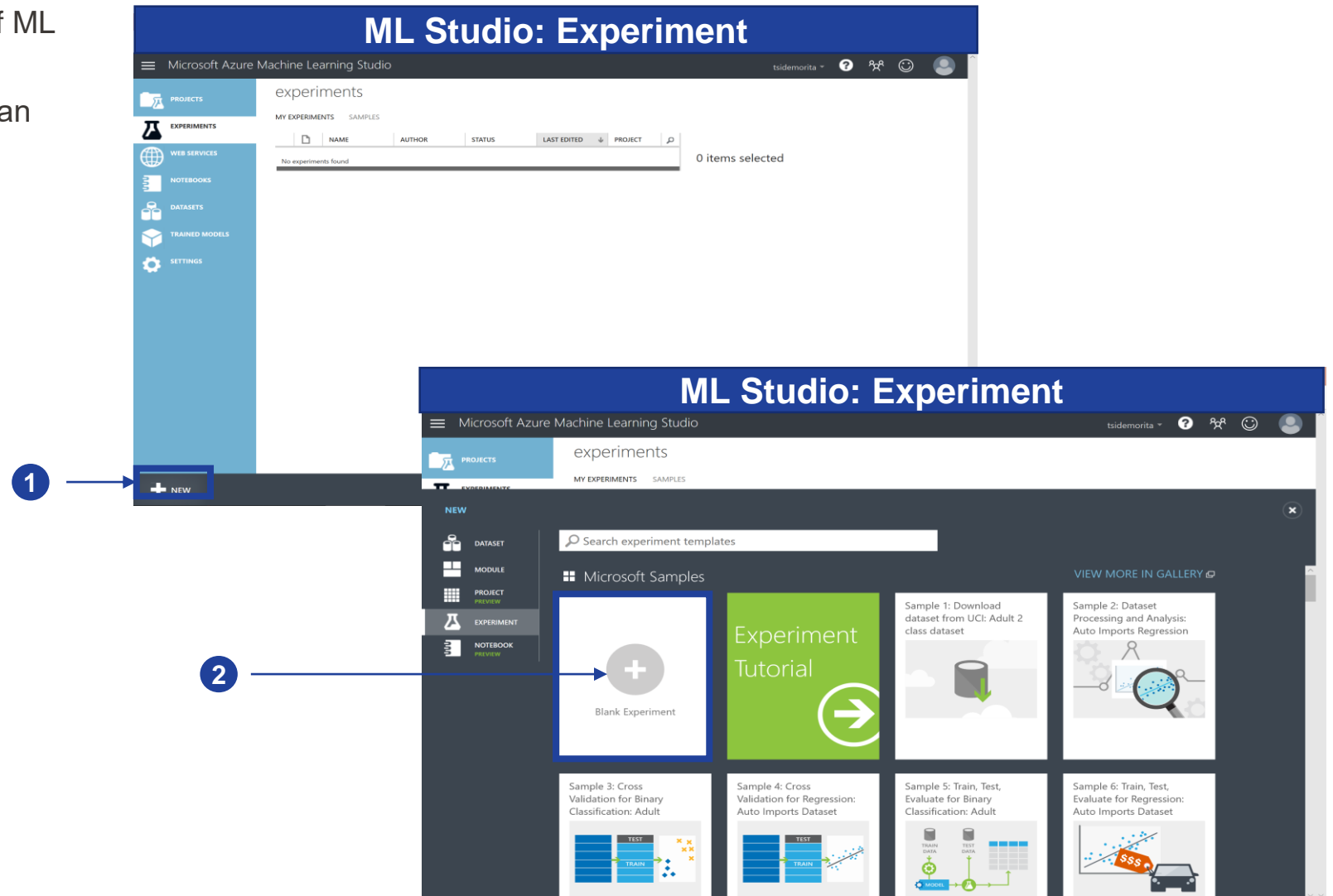
Content

01. Data Import	04
a. Create an experiment	
b. Import Data from a Blob Storage	
c. Import Data from a Web URL via http	
d. Import Data from Local Storage	
e. Create new dataset manually	
f. Unpack Zipped Dataset	
02. Summary	13

Data Import

Create an experiment

1. Click new on the bottom left corner of ML Studio.
2. Click "+ Blank Experiment" to create an experiment



Import Data from a Blob Storage

1. Search for “**Import Data**” module and drag and drop to the canvas
2. On Data Source, select “**Azure Blob Storage**”
3. Fill up this space. DS will provide you with the details
4. Tick here if your data includes a header
5. Click “**Run**”

The screenshot displays the Microsoft Azure Machine Learning Studio interface for the 'Import Data from a Blob Storage' module. The interface is divided into several sections:

- Left Panel (Module List):** A search bar at the top contains the text 'impd'. Below it, the 'Data Input and Output' category is expanded, and the 'Import Data' module is highlighted. A blue arrow labeled '1' points from this module to the canvas.
- Canvas:** The 'Import Data' module is placed on the canvas, labeled with a '1'. A large blue arrow points from the module in the left panel to this instance on the canvas.
- Right Panel (Properties):** The 'Import Data' properties are shown. A blue arrow labeled '2' points to the 'Data source' dropdown, which is set to 'Azure Blob Storage'. A blue arrow labeled '3' points to the 'Account name' text box, which has a red exclamation mark icon next to it. A blue arrow labeled '4' points to the 'File has header row' checkbox, which is checked. A blue arrow labeled '5' points to the 'RUN' button at the bottom of the interface.
- Bottom Panel:** The 'RUN' button is highlighted with a blue box and a blue arrow labeled '5' pointing to it.

Import Data from a Web URL via HTTP

1. On Data Source, choose “**Web URL via HTTP**”
2. Fill up the space with your URL of choice
3. Tick here if your data includes header
4. Click “**Run**”

The screenshot displays the Microsoft Azure Machine Learning Studio interface. The title bar reads "ML Studio : Import Data from a Web URL". The main workspace shows an "Import Data" node in the center, with a "Mini Map" view below it. The left sidebar contains a "Search experiment items" bar and a list of "Samples" including "Adult Census Income...", "Airport Codes Dataset", "Automobile price dat...", "Bike Rental UCI dataset", "Bill Gates RGB Image", "Blood donation data", "Book Reviews from A...", "Breast cancer data", "Breast Cancer Features", "Breast Cancer Info", "CRM Appetency Labe...", "CRM Churn Labels Sh...", "CRM Dataset Shared", "CRM Upselling Labels...", "Energy Efficiency Reg...", "Flight Delays Data", and "Flight on-time perfor...". The right sidebar shows the "Properties" section for the "Import Data" node, with fields for "Data source" (set to "Web URL via HTTP"), "Data source URL" (set to "https://petronastraining.blob.core.windows.net/petrona/..."), "Data format" (set to "CSV"), and a checked checkbox for "CSV or TSV has header...". The bottom toolbar includes buttons for "NEW", "RUN HISTORY", "SAVE", "SAVE AS", "DISCARD CHANGES", "RUN", "SET UP WEB SERVICE", and "PUBLISH TO GALLERY".

1. On Data Source, choose “**Web URL via HTTP**”

2. Fill up the space with your URL of choice

3. Tick here if your data includes header

4. Click “**Run**”

Data Visualization

To perform data visualization, right click on **“Import Data”** and choose **“Visualize”**.

ML Studio : Data Visualization

Microsoft Azure Machine Learning Studio (classic) | Experiment created on 5/2/2020 | Finished running

Import Data context menu options:

- Download
- Save as Dataset
- Save as Trained Model
- Save as Transform
- Visualize**
- Generate Data Access Code...
- Open in a new Notebook

Results dataset

Id	Observation	IsSafeAct	SubmittedDate	Status	ObservationDateTime
758dbbbc-f504-46a2-9a45-000133a07b47	Found Cotton Rags left at L-4010.	false	2020-01-15T01:42:12.673	Closed	2020-01-14T22:00:00
69d1f7df-7e3b-4901-8349-000259251bc5	Found safety harness inside FEX box	false	2020-01-20T15:26:51.797	Closed	2020-01-20T14:16:00
276da065-7f8e-47fa-805a-0003c2f398e7	Observed colleague do a proper shift handover with outgoing panelman and immediately perform SOSO after shift	true	2020-01-01T07:47:18.39	Closed	2020-01-01T06:20:00

Statistics

- Unique Values: 35274
- Missing Values: 357
- Feature Type: Object Feature

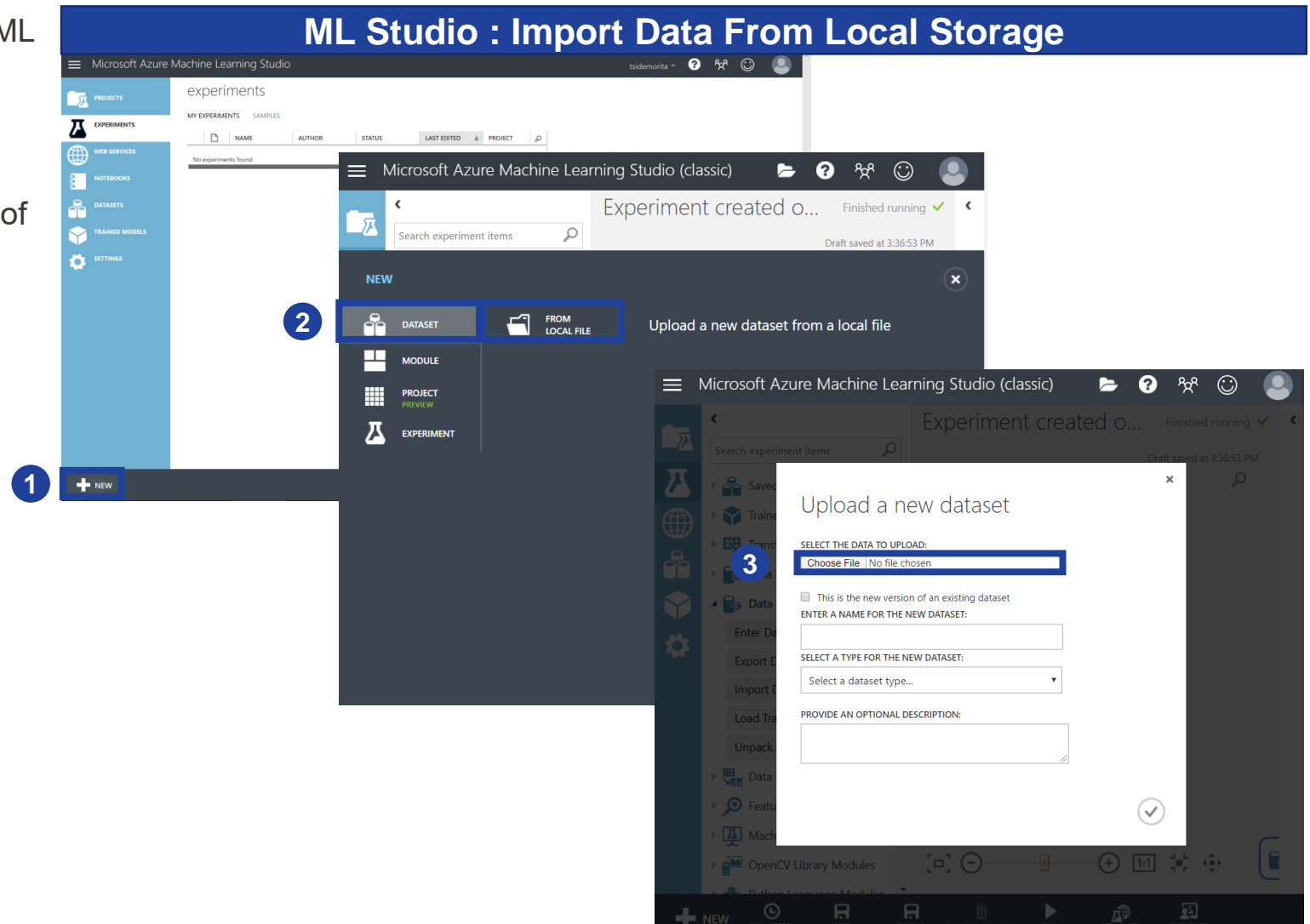
Visualizations

SubmittedDate Histogram

Frequency

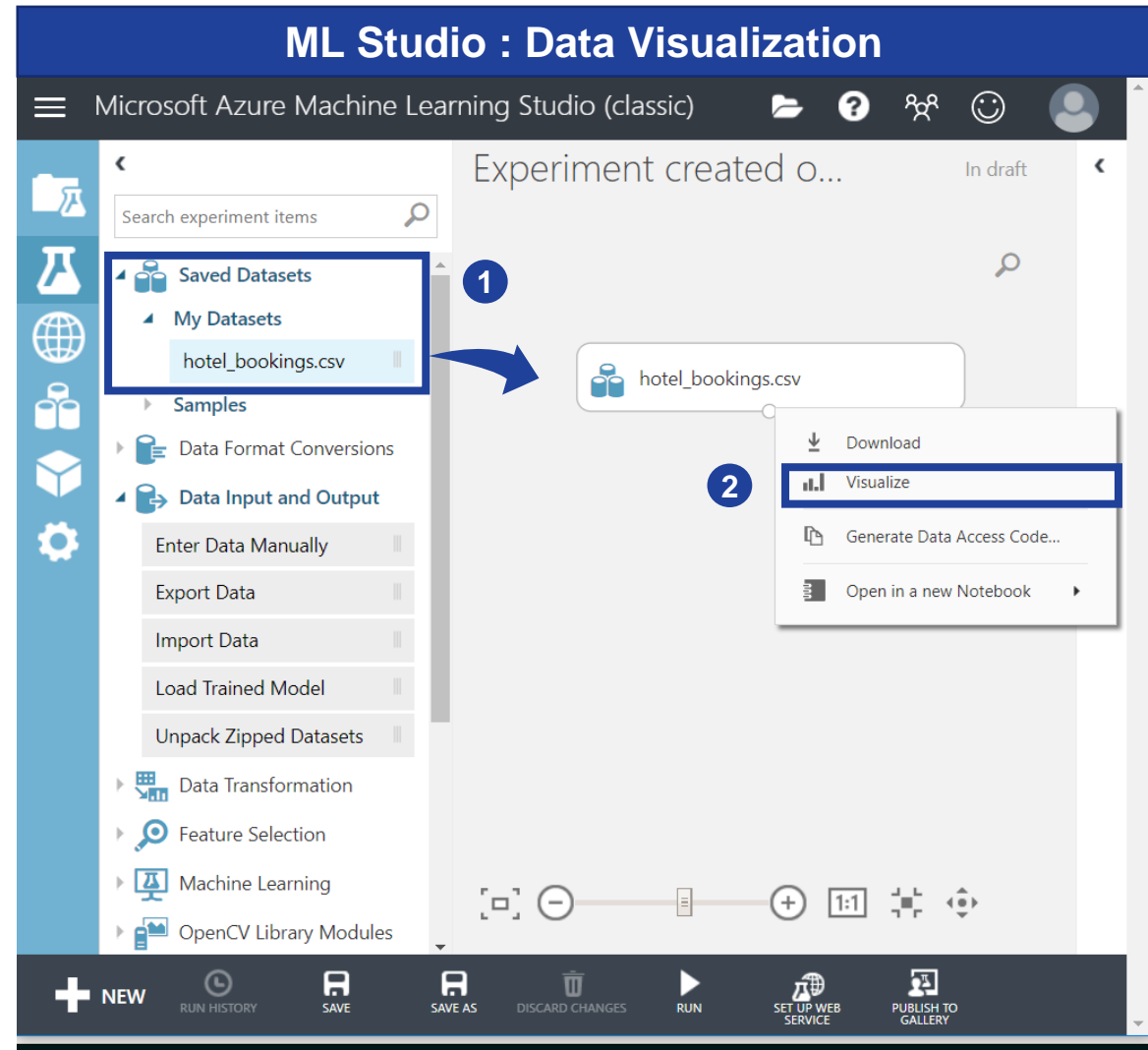
Import Data from Local Storage

1. Click “**New**” from the bottom left corner of ML Studio
2. Choose “**Dataset**” and click “**From Local File**”
3. Click on “**Choose File**” and select dataset of your choice



Visualize data from Saved Datasets

1. Click on “**Saved Datasets**” and drag dataset of your choice to the canvas
2. Right-click on the selected dataset and click “**Visualize**”



Create new dataset manually

To create new dataset manually, search “**Enter Data Manually**”, drag and drop it on the canvas and fill up the right form accordingly and proceed to “**Visualize**”

The screenshot displays the Microsoft Azure Machine Learning Studio (classic) interface. The title bar reads "ML Studio : Enter Data Manually". The left sidebar shows the "Data Input and Output" section with the "Enter Data Manually" module highlighted. A blue arrow points from this module in the sidebar to a "1" in a box on the canvas, which also contains the text "Enter Data Manually" with a green checkmark. The right sidebar shows the "Properties" pane for the "Enter Data Manually" module, with the "DataFormat" set to "CSV" and "HasHeader" checked. The "Data" section shows a table with 6 rows of data:

	Name	Age	Gender
1	Rina	32	Female
2	Ali	40	Male
3	Laila	25	Female
4	John	30	Male
5	Christina	25	Female
6			

The bottom status bar shows the "NEW" button and various action buttons: "RUN HISTORY", "SAVE", "SAVE AS", "DISCARD CHANGES", "RUN", "SET UP WEB SERVICE", and "PUBLISH TO GALLERY".

Unpack Zipped Dataset

Load zipped datasets in ML Studio using a similar manner as loading file from local storage:

1. Drag and drop your Zipped Dataset, then search “**Unpack Zipped Dataset**” module, drag it to the canvas
2. Click “**Run**”
3. Visualize by right click at (1)
4. Check this “**File has header row**” box if the dataset has a header

The screenshot displays the Microsoft Azure Machine Learning Studio (classic) interface. The title bar reads "ML Studio : Unpack Zipped Dataset". The main workspace shows an experiment created on 5/2/2020. On the left sidebar, under "Saved Datasets", the "hotel_booking_demand.zip" file is highlighted. In the center canvas, the "Unpack Zipped Datasets" module is added, with the file name "hotel_booking_demand.zip" connected to its input. The module is numbered with a blue circle '1'. At the bottom of the canvas, the "RUN" button is highlighted with a blue circle '2'. On the right sidebar, the "Properties" pane for the "Unpack Zipped Datasets" module is shown. The "Dataset to Unpack" field is empty. The "Dataset file format" is set to "CSV". The "File has header row" checkbox is checked and highlighted with a blue circle '4'. The "Compression file format" is set to "Zip". Below these settings, the execution status is shown: "START TIME: 5/2/2020 4:17:01 PM", "END TIME: 5/2/2020 4:17:15 PM", "ELAPSED TIME: 0:00:13.265", "STATUS CODE: Finished", and "STATUS DETAILS: None". A "Quick Help" section at the bottom right explains that the module unpacks datasets from a zip package in user storage.

Summary

Summary

In this module, you learnt to import data into Azure ML Studio:

1. Import dataset from multiple sources:
 - Import data from a Blob Storage
 - Import Data from a Web URL via HTTP
 - Import Data from Local Storage
2. Create new datasets manually in ML Studio
3. Unpack zipped datasets within ML Studio

Thank you for your passion!

