

Practice #2 / Part 1

Azure Machine Learning Studio

Auto Price Prediction



Azure Machine Learning

> #2 Azure Machine Learning Studio (Classic) / Part 1

INSTRUCTIONS:

Record a video with 5-10 minutes explaining how to use **Azure Machine Learning Studio** and your discoveries.

- Follow each of the steps detailed in the next slides and explain them.
- Explain that you are loading the file, select columns, cleaning missing data, splitting the file, etc.
- Explain why you are performing each step.
- Make your conclusion of the predicted values.
- Evaluate your model and explain the results.
- Analyze all the steps you performed and think how should be the behavior (processes) of an enterprise solution with the same purpose. What are the differences?

In your video explain how and why you performed each step and explain the final result.

EVALUATION:

Mark: 10 points (part 1 and 2)

- Ensure that you recorded yourself using the tool or using your own screenshots
- Ensure that you recorded all the performed steps
- Ensure that you analysed the results
- Explain what kind of ML you are using in this exercise and why

Will be considered: Your results, explanations, level of details, clarity to explain and presentation / video quality (preparation).

Due date: Week 6 class



> Azure Machine Learning Studio | Practice / Part 1

MAIN GOAL:

Your main goal is to predict the price of a car based on different variables such as make and technical specifications.

MAIN STEPS:

Create a model

- Get the data
- Prepare the data
- Define features

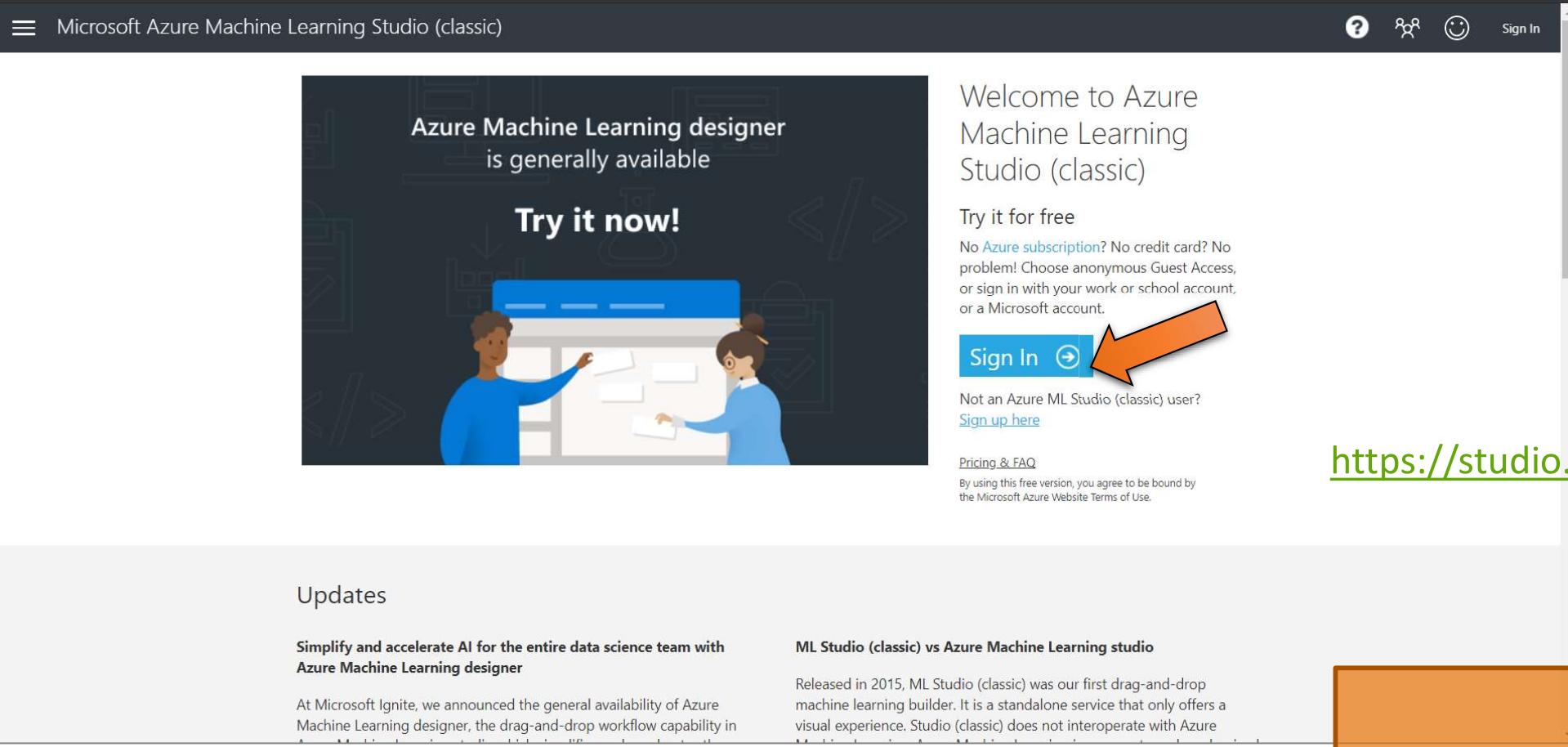
Train the model

- Choose and apply an algorithm

Score and test the model

- Predict new automobile prices

> Azure Machine Learning Studio | Practice



Welcome to Azure Machine Learning Studio (classic)

Try it for free

No Azure subscription? No credit card? No problem! Choose anonymous Guest Access, or sign in with your work or school account, or a Microsoft account.

[Sign In](#)

Not an Azure ML Studio (classic) user?
[Sign up here](#)

<https://studio.azureml.net/>

Updates

Simplify and accelerate AI for the entire data science team with Azure Machine Learning designer

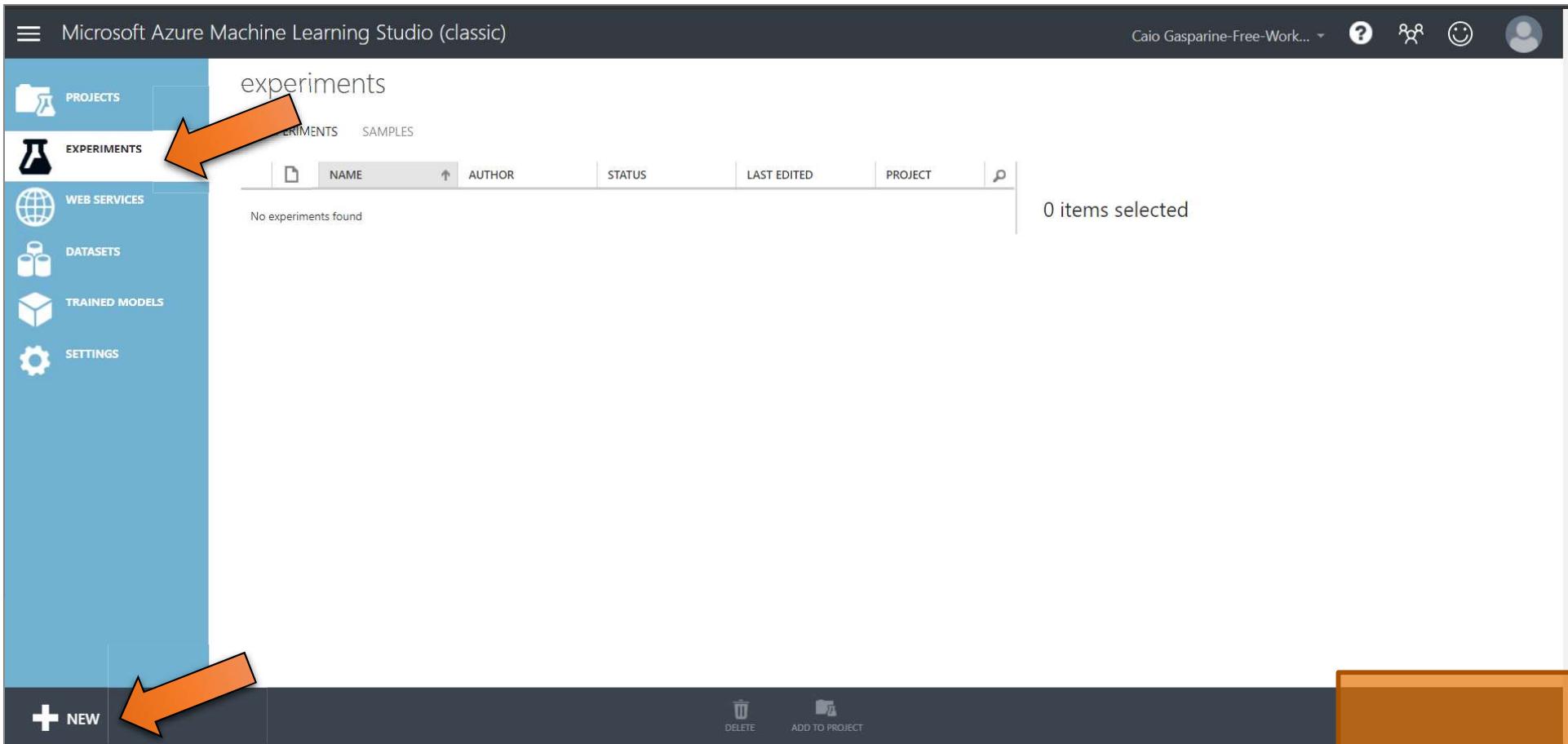
At Microsoft Ignite, we announced the general availability of Azure Machine Learning designer, the drag-and-drop workflow capability in

ML Studio (classic) vs Azure Machine Learning studio

Released in 2015, ML Studio (classic) was our first drag-and-drop machine learning builder. It is a standalone service that only offers a visual experience. Studio (classic) does not interoperate with Azure

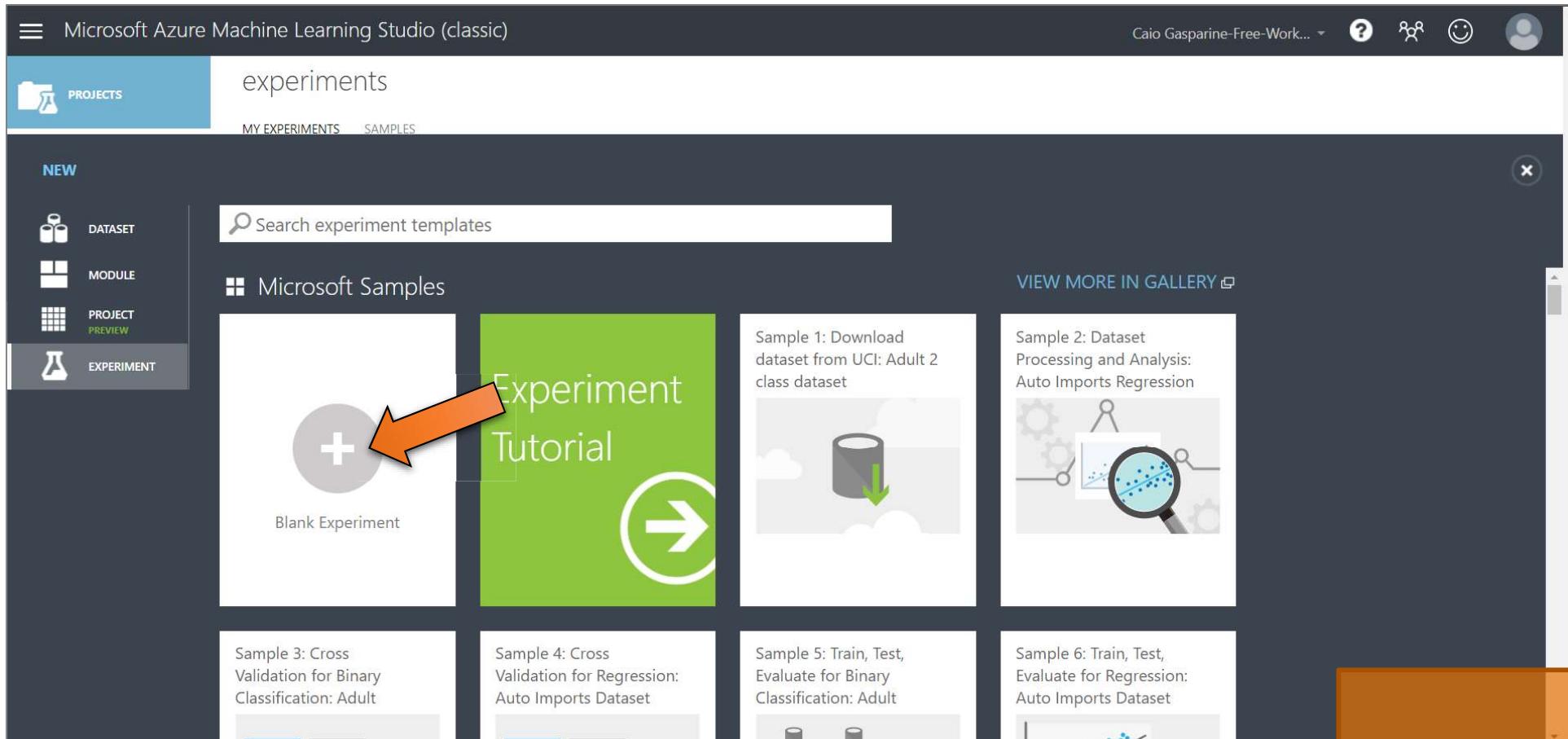
This is an example for academic purposes

> Azure Machine Learning Studio | Practice



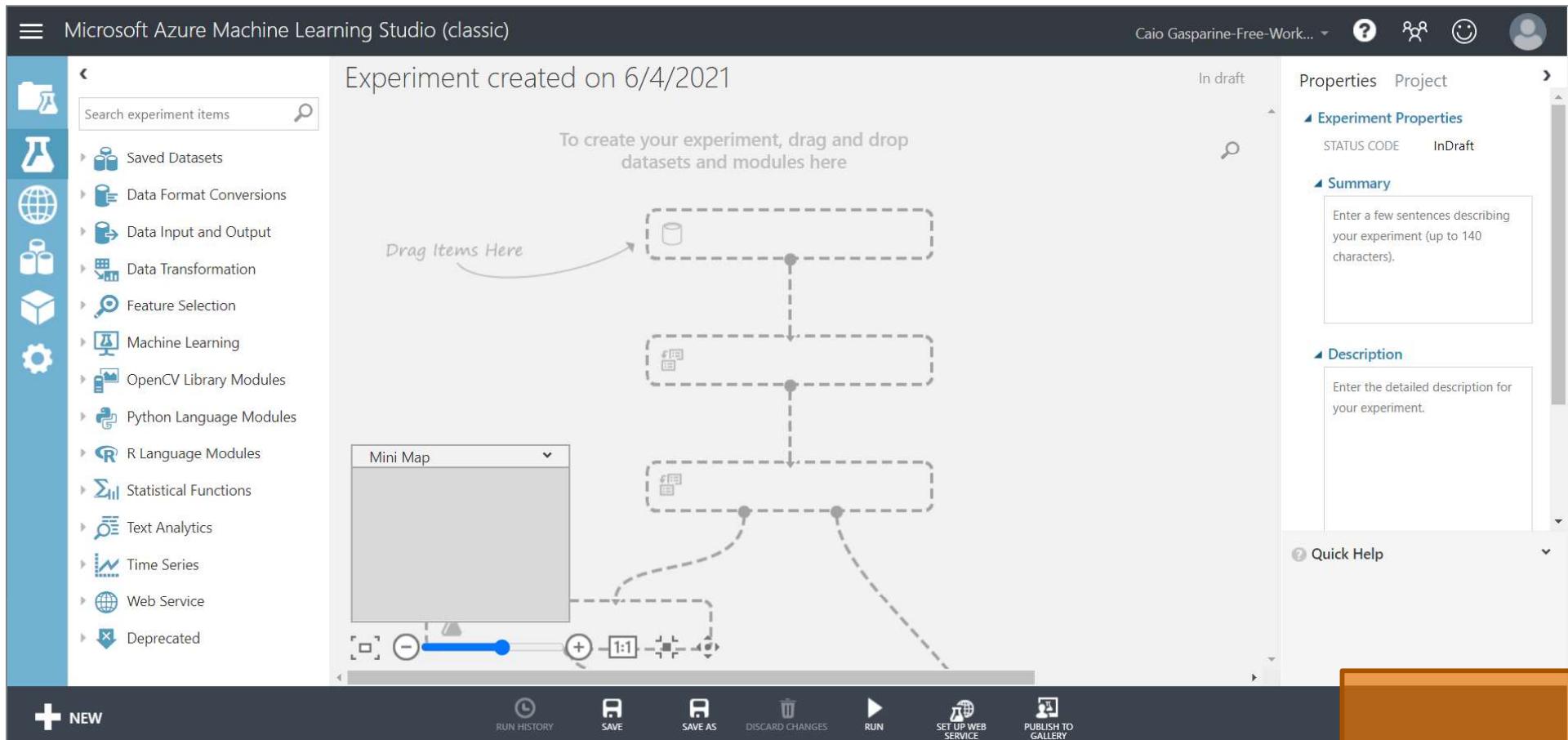
This is an example for
academic purposes

> Azure Machine Learning Studio | Practice



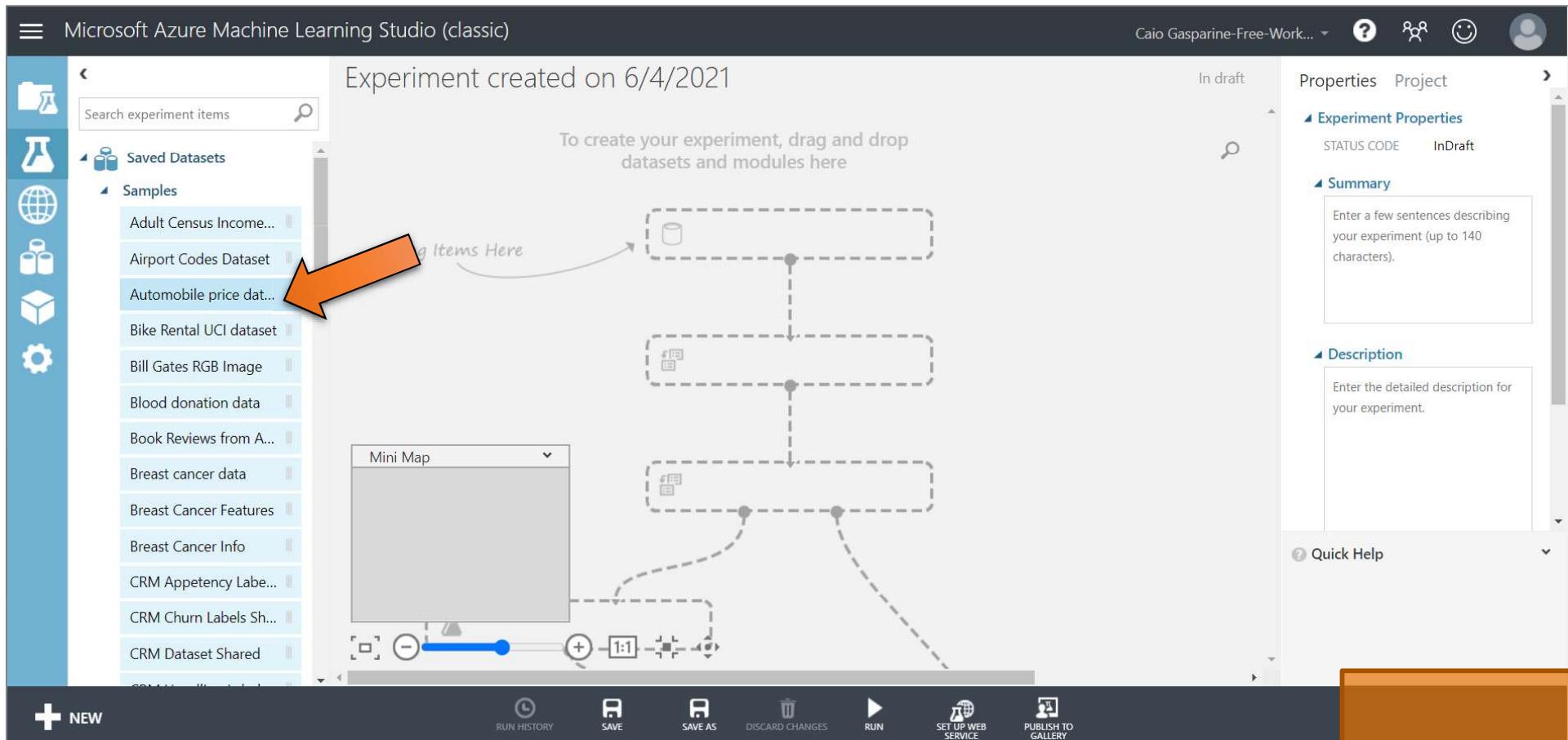
This is an example for
academic purposes

> Azure Machine Learning Studio | Practice



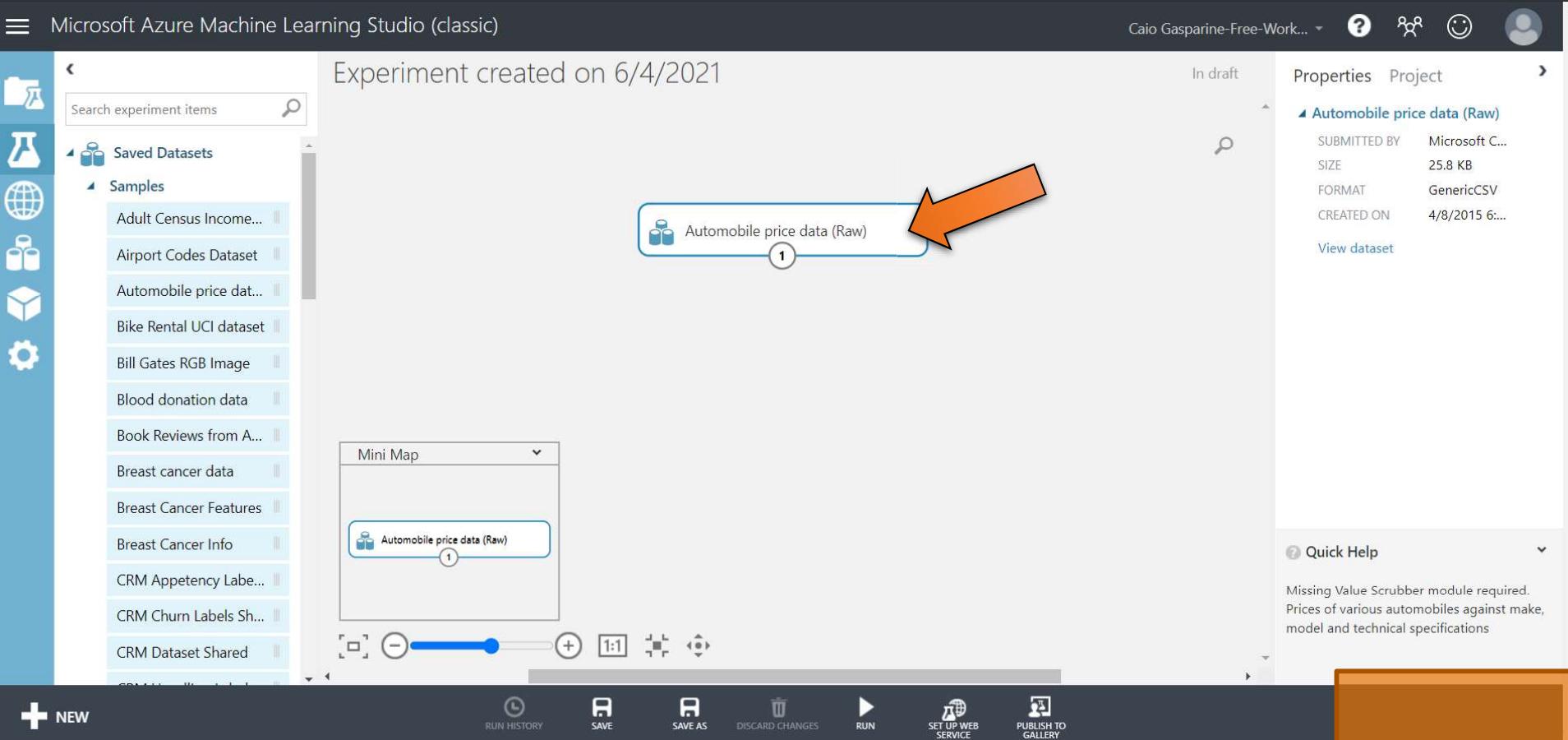
This is an example for
academic purposes

> Azure Machine Learning Studio | Practice



This is an example for
academic purposes

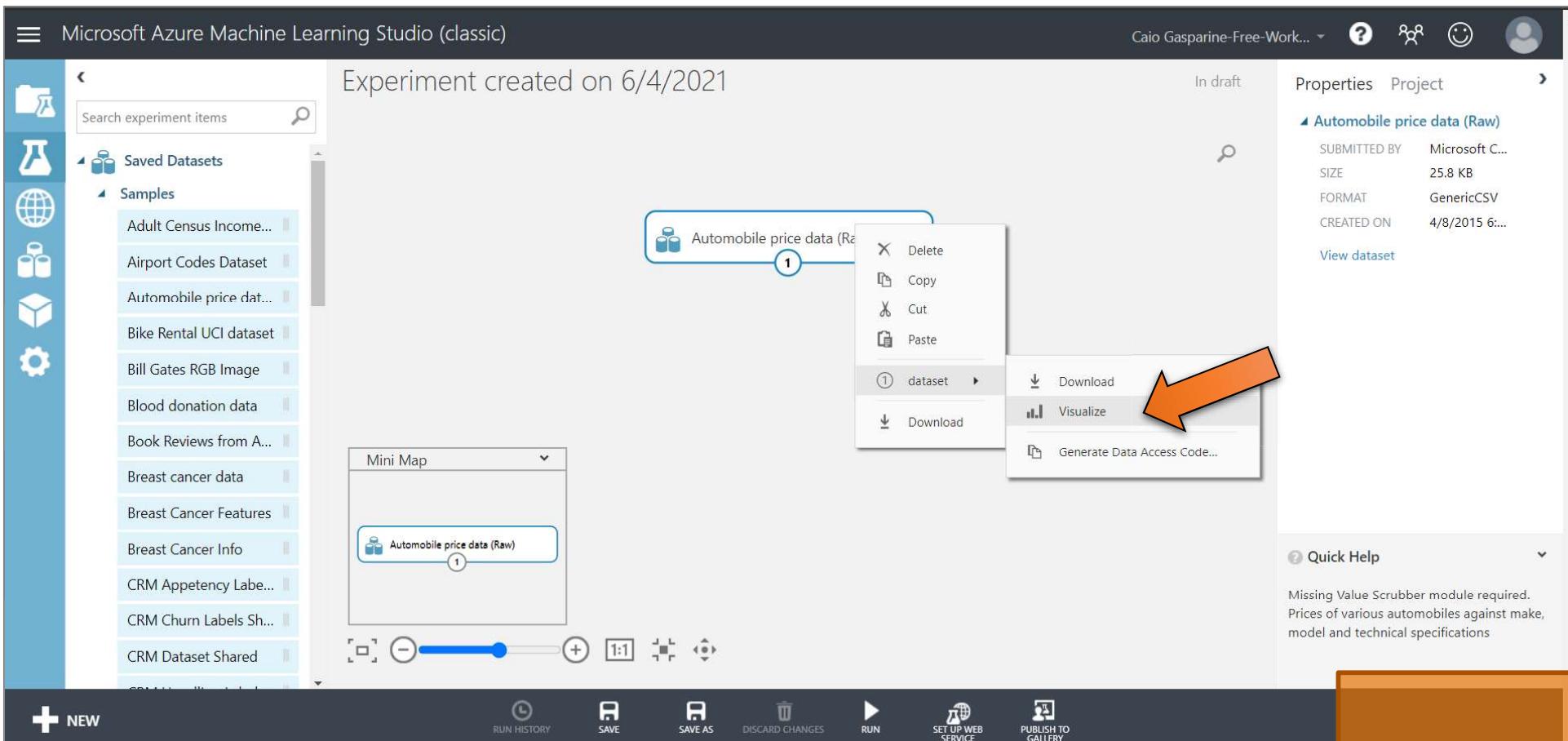
> Azure Machine Learning Studio | Practice



The screenshot shows the Microsoft Azure Machine Learning Studio (classic) interface. The main title bar reads "Microsoft Azure Machine Learning Studio (classic)" and "Experiment created on 6/4/2021". On the left, there's a sidebar with icons for Saved Datasets, Samples, and other data types. A search bar at the top says "Search experiment items". In the center, a "Mini Map" shows the flow of the experiment, with a node labeled "Automobile price data (Raw)". This node is highlighted with a blue rounded rectangle and an orange arrow points to it from the bottom right. To the right of the map, there's a "Properties" panel for the dataset, showing details like "SUBMITTED BY Microsoft C...", "SIZE 25.8 KB", "FORMAT GenericCSV", and "CREATED ON 4/8/2015 6...". Below the properties is a "Quick Help" section with the text: "Missing Value Scrubber module required. Prices of various automobiles against make, model and technical specifications". At the bottom of the studio window, there are buttons for "NEW", "RUN HISTORY", "SAVE", "SAVE AS", "DISCARD CHANGES", "RUN", "SET UP WEB SERVICE", and "PUBLISH TO GALLERY".

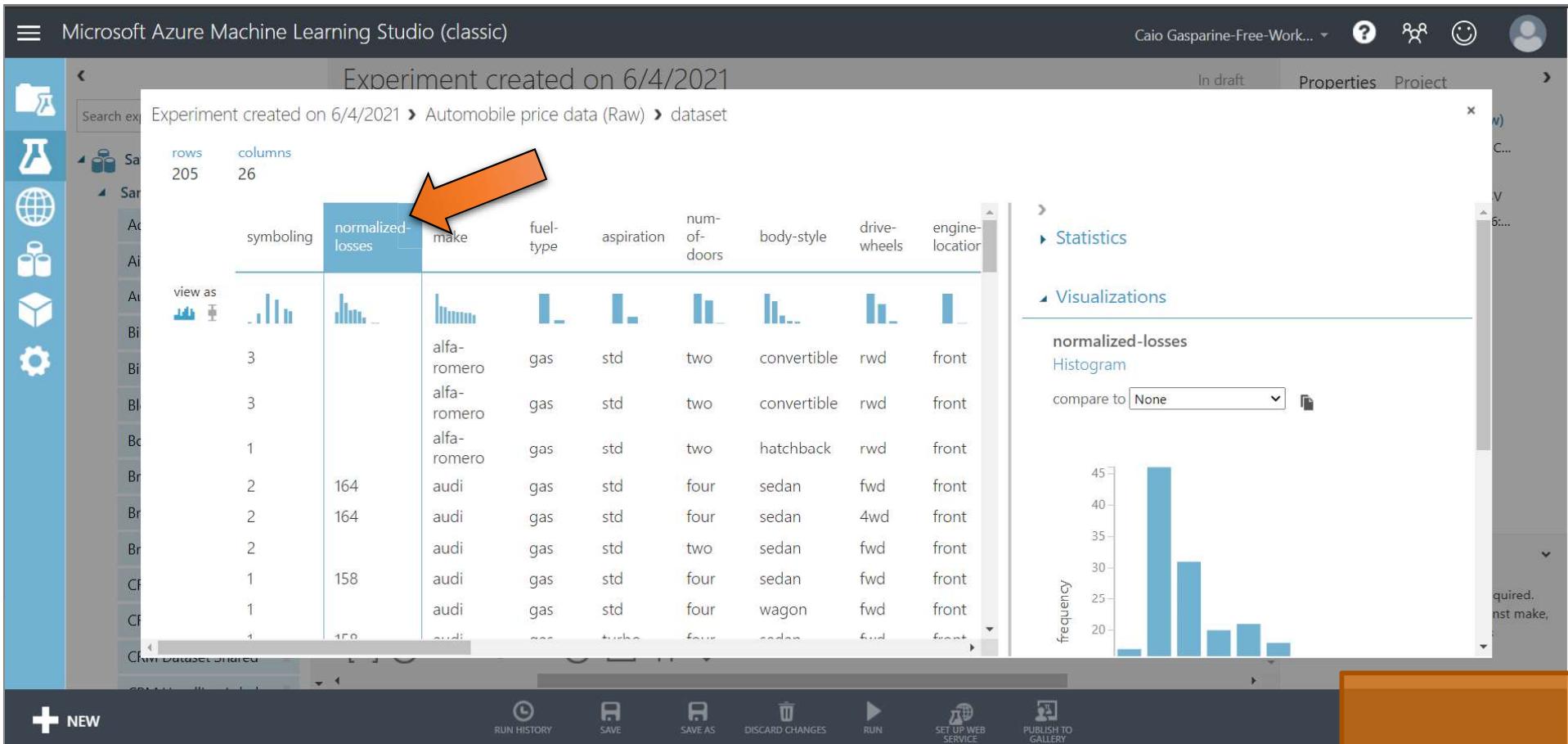
This is an example for
academic purposes

> Azure Machine Learning Studio | Practice



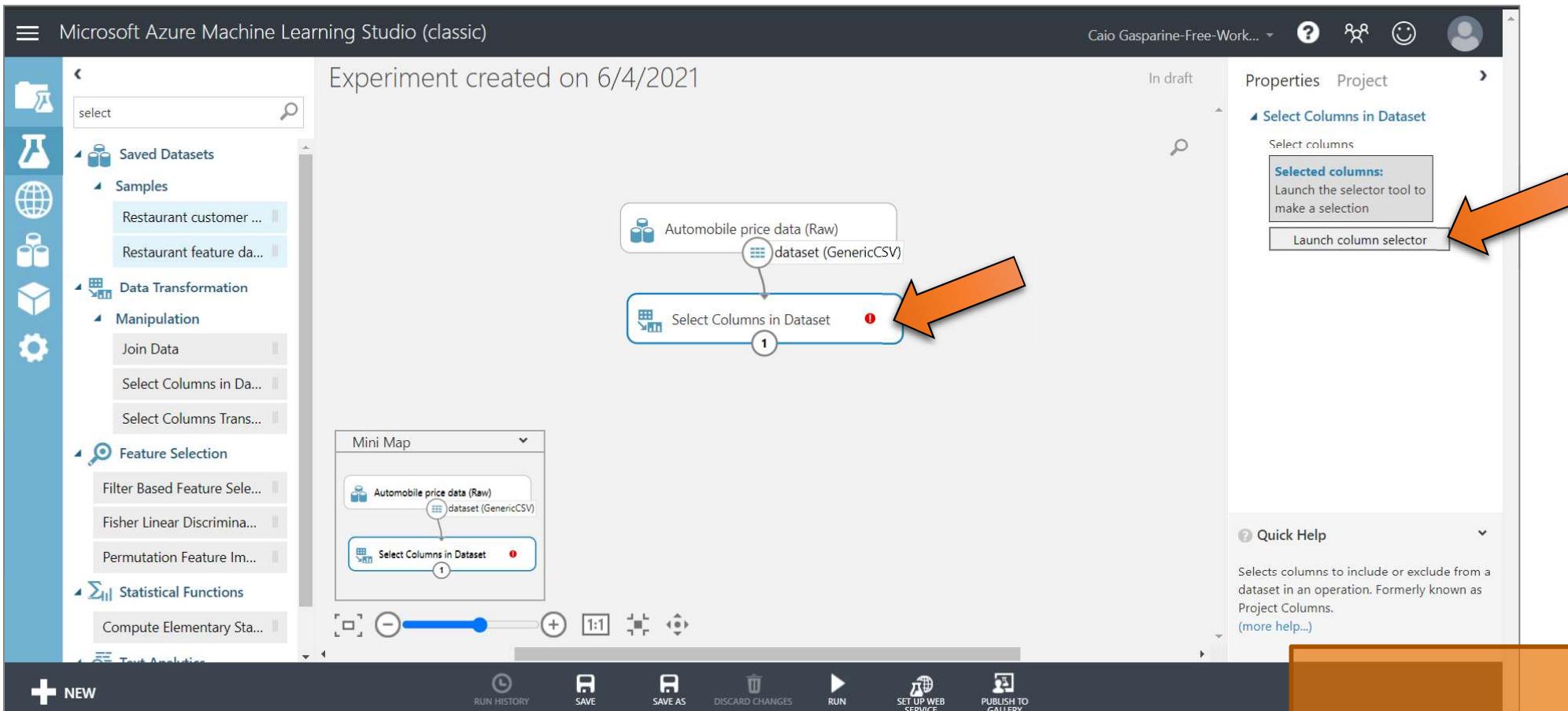
This is an example for
academic purposes

> Azure Machine Learning Studio | Practice



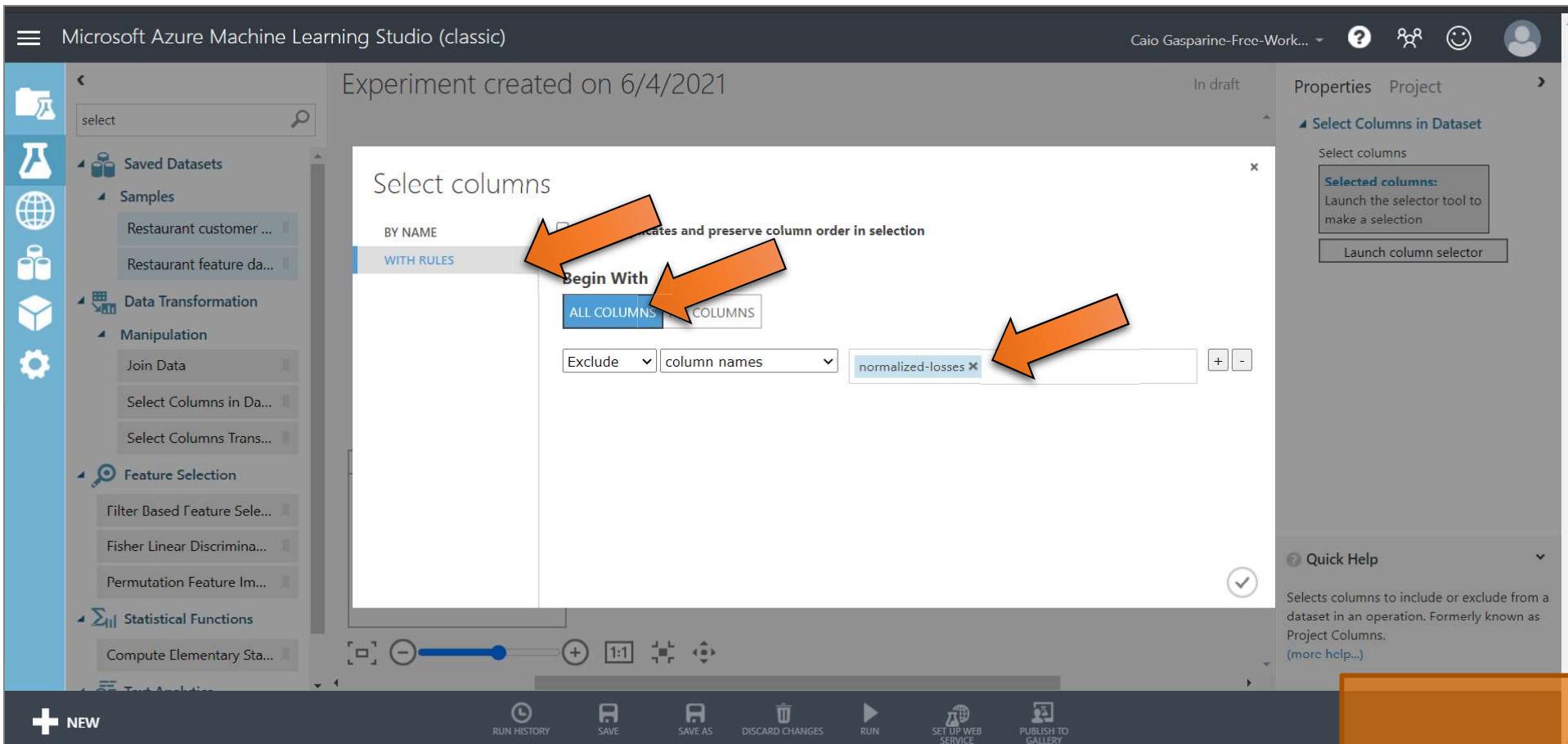
This is an example for
academic purposes

> Azure Machine Learning Studio | Practice



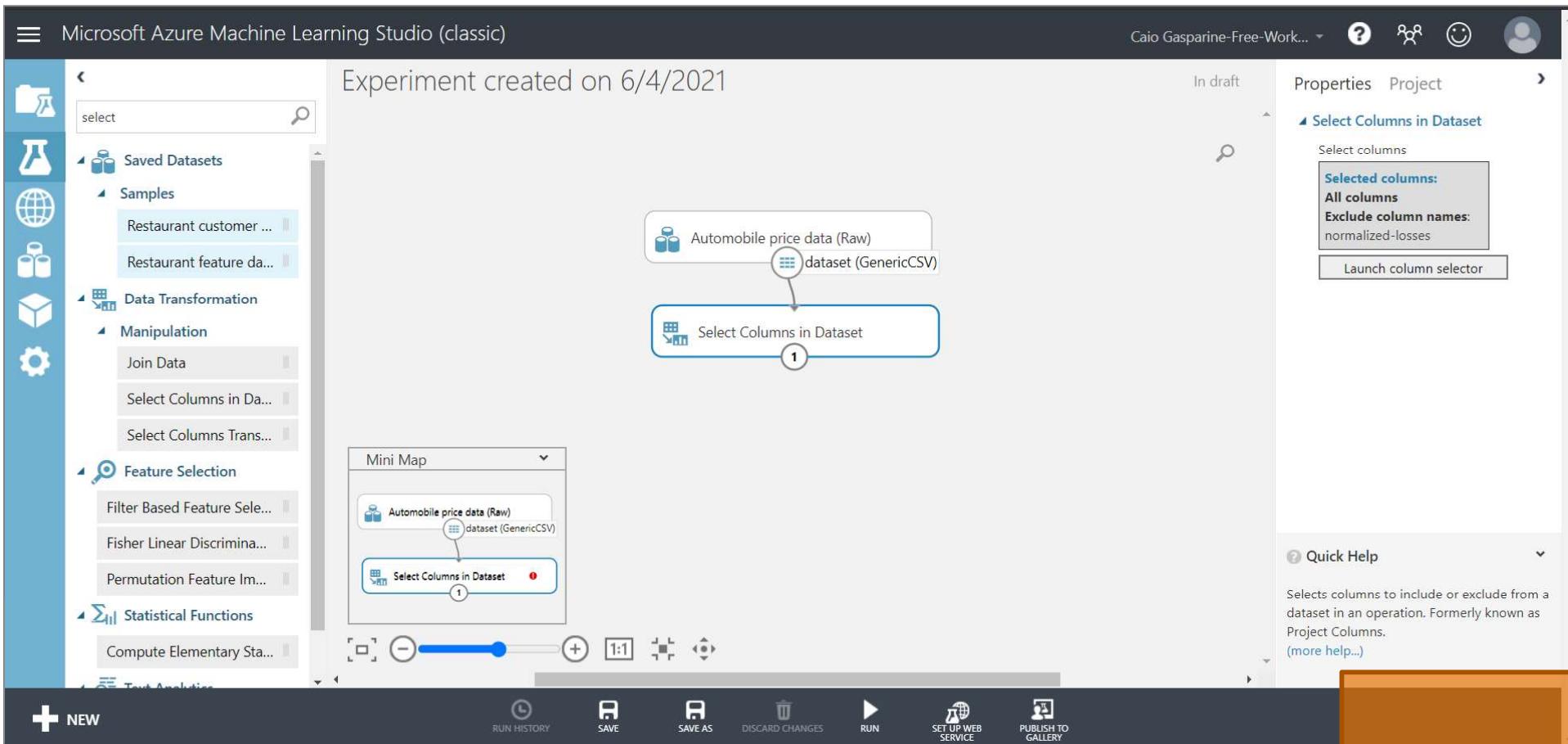
This is an example for
academic purposes

> Azure Machine Learning Studio | Practice



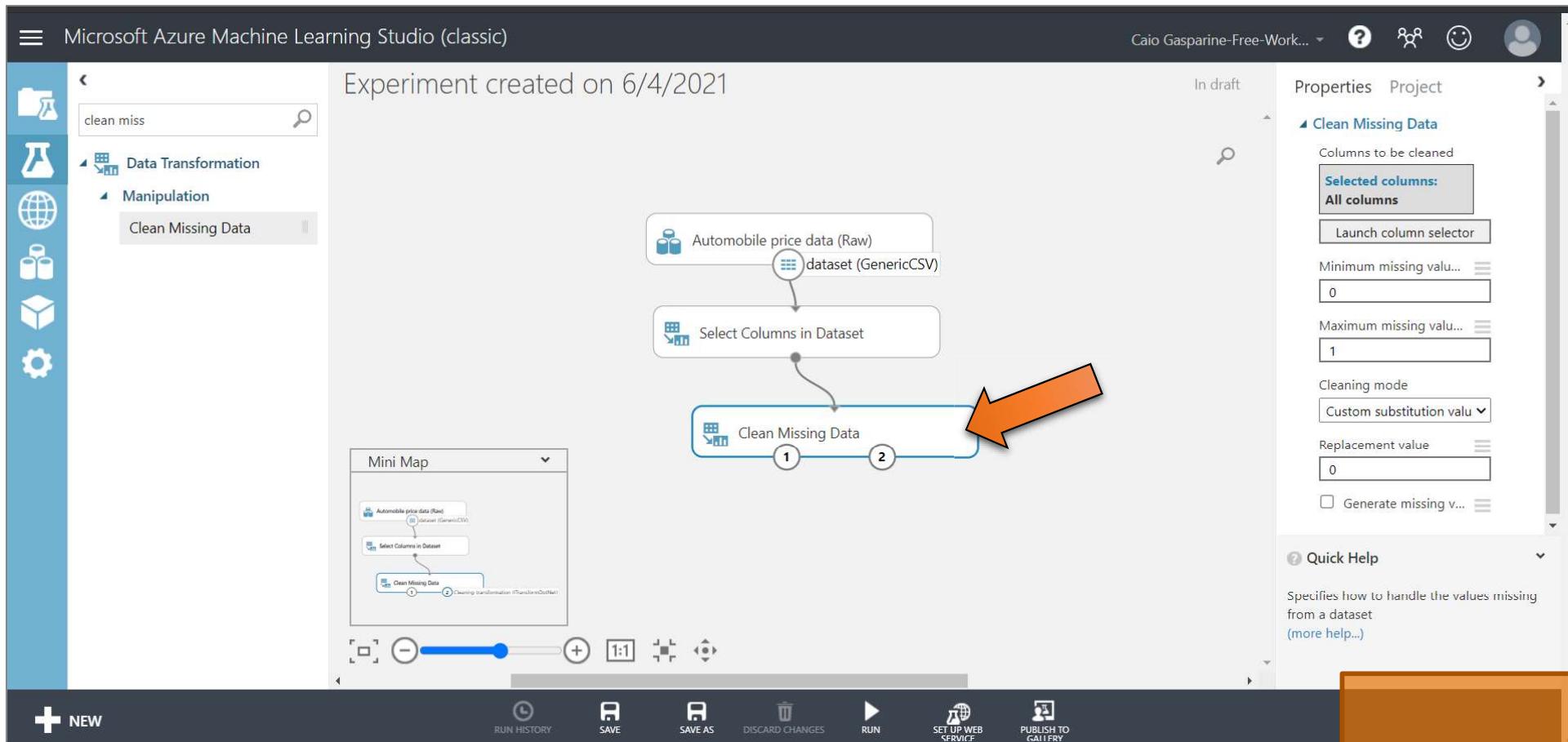
This is an example for
academic purposes

> Azure Machine Learning Studio | Practice



This is an example for
academic purposes

> Azure Machine Learning Studio | Practice



This is an example for
academic purposes

> Azure Machine Learning Studio | Practice

The screenshot shows the Microsoft Azure Machine Learning Studio (classic) interface. On the left, the sidebar includes icons for Data Transformation, Manipulation, and a search bar containing "clean miss". The main workspace displays an experiment titled "Experiment created on 6/4/2021" in draft mode. The experiment flow starts with "Automobile price data (Raw)" connected to "Select Columns in Dataset", which then connects to "Clean Missing Data". A Mini Map on the left shows the same sequence of steps. The "Clean Missing Data" component has two numbered outputs: 1 and 2. An orange arrow points from the output of step 2 to the "Properties" panel on the right. The "Properties" panel is expanded to show the "Clean Missing Data" configuration, specifically the "Cleaning mode" set to "Remove entire row". A second orange arrow points from the "Remove entire row" dropdown in the Properties panel to the "Quick Help" section below it. The "Quick Help" section defines the mode as "Specifies how to handle the values missing from a dataset".

This is an example for academic purposes

> Azure Machine Learning Studio | Practice

Microsoft Azure Machine Learning Studio (classic)

Experiment created on 6/4/2021

In draft

Properties Project

Clean Missing Data

Selected columns: All columns

Launch column selector

Minimum missing value ratio: 0

Maximum missing value ratio: 1

Cleaning mode: Remove entire row

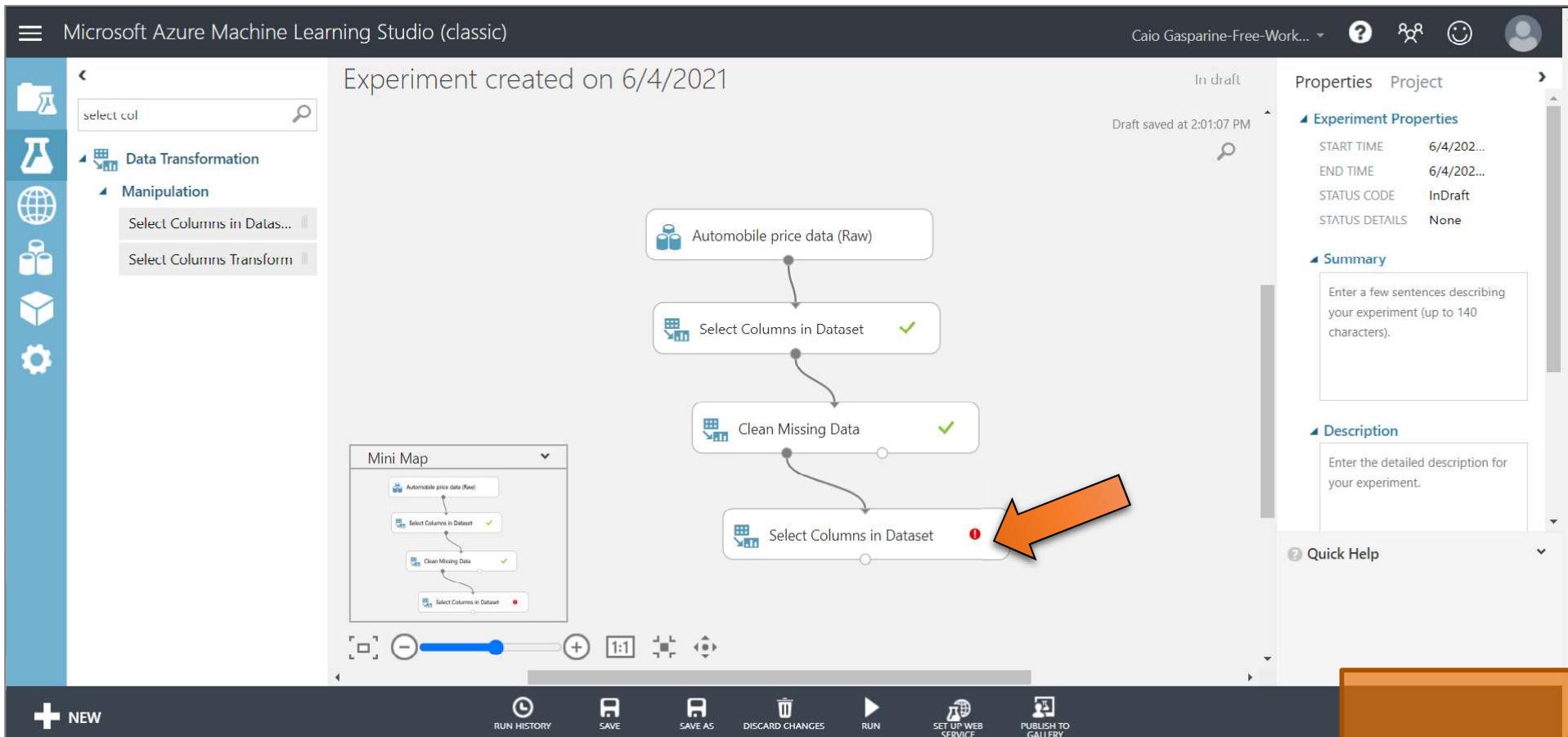
Quick Help

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

This is an example for academic purposes

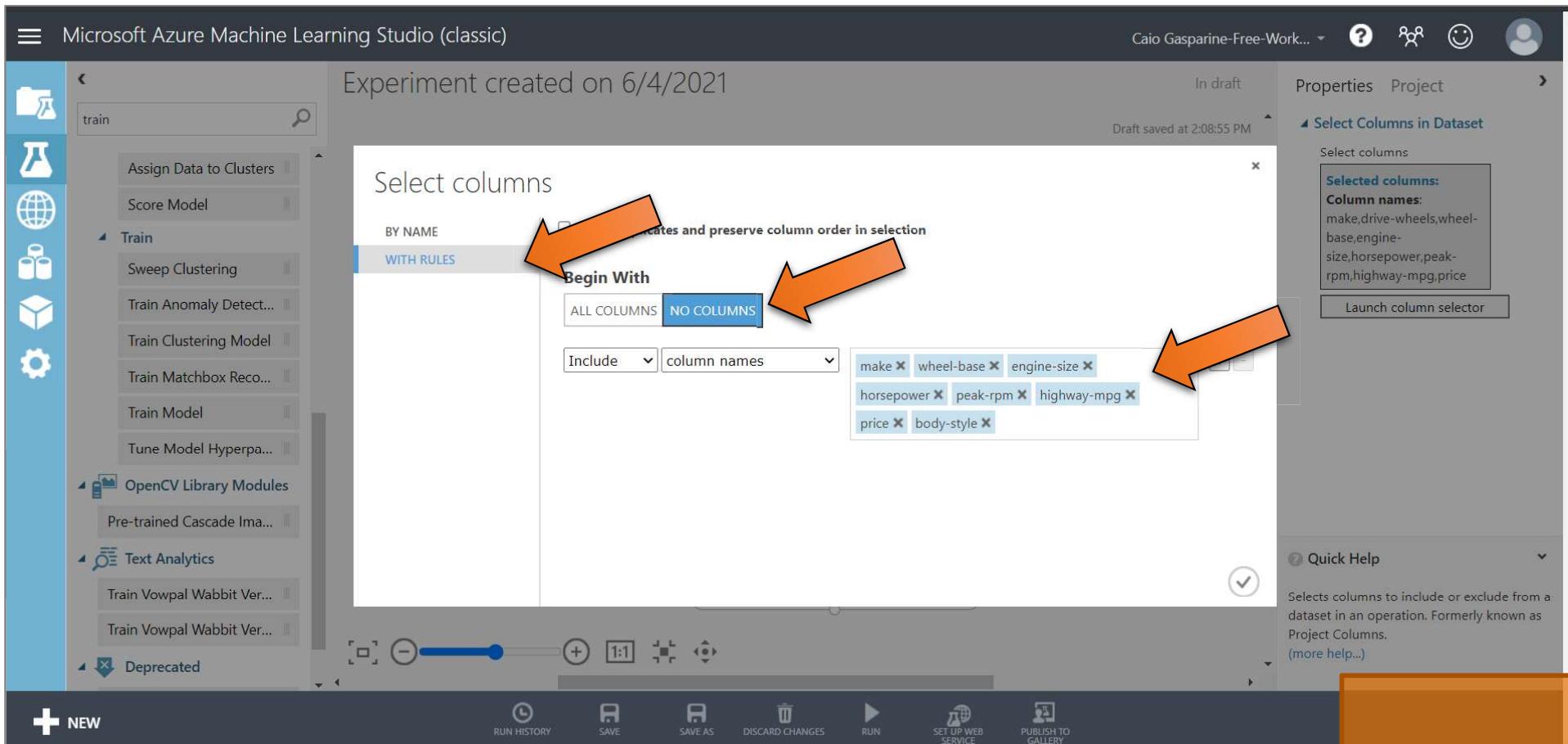
```
graph TD; A[Automobile price data (Raw)] --> B[Select Columns in Dataset]; B --> C[Clean Missing Data]
```

> Azure Machine Learning Studio | Practice

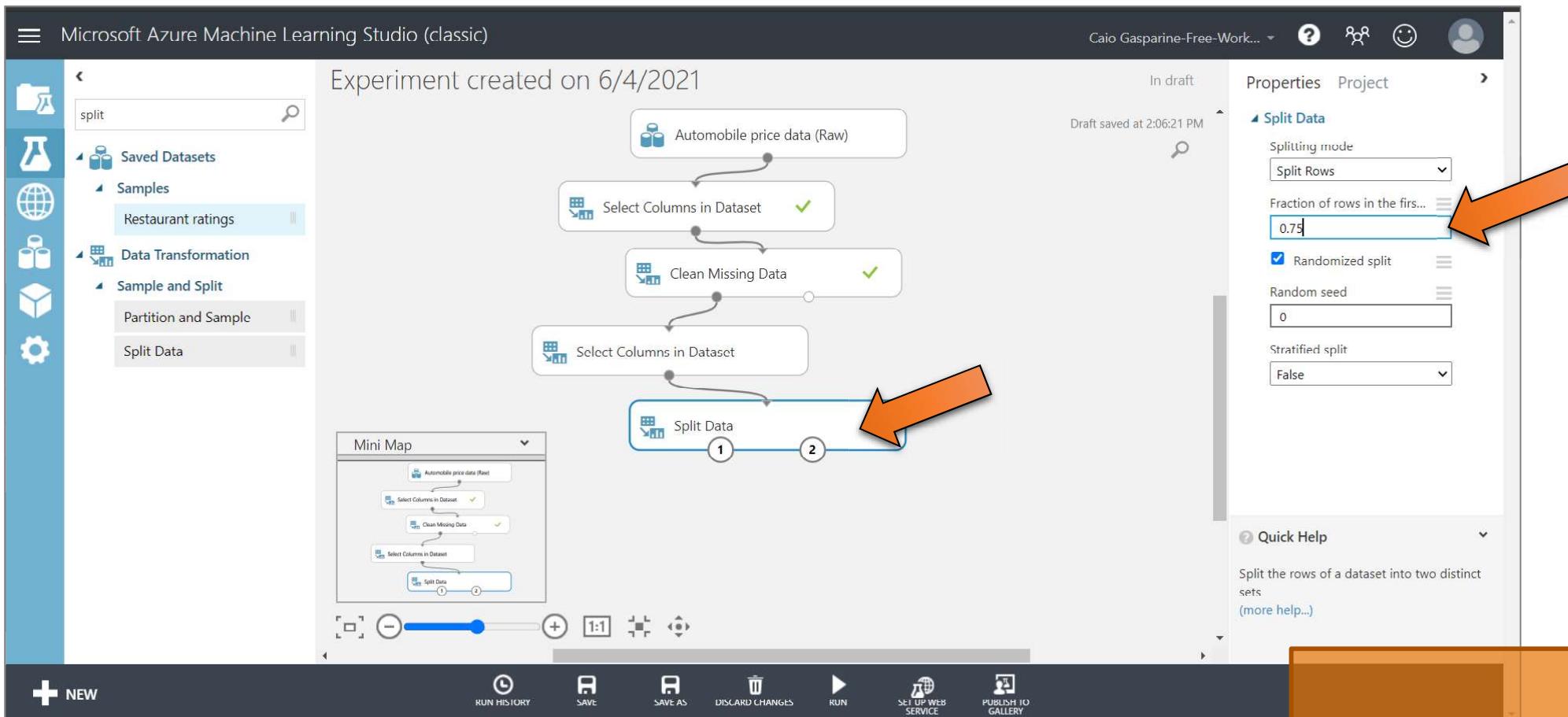


This is an example for
academic purposes

> Azure Machine Learning Studio | Practice

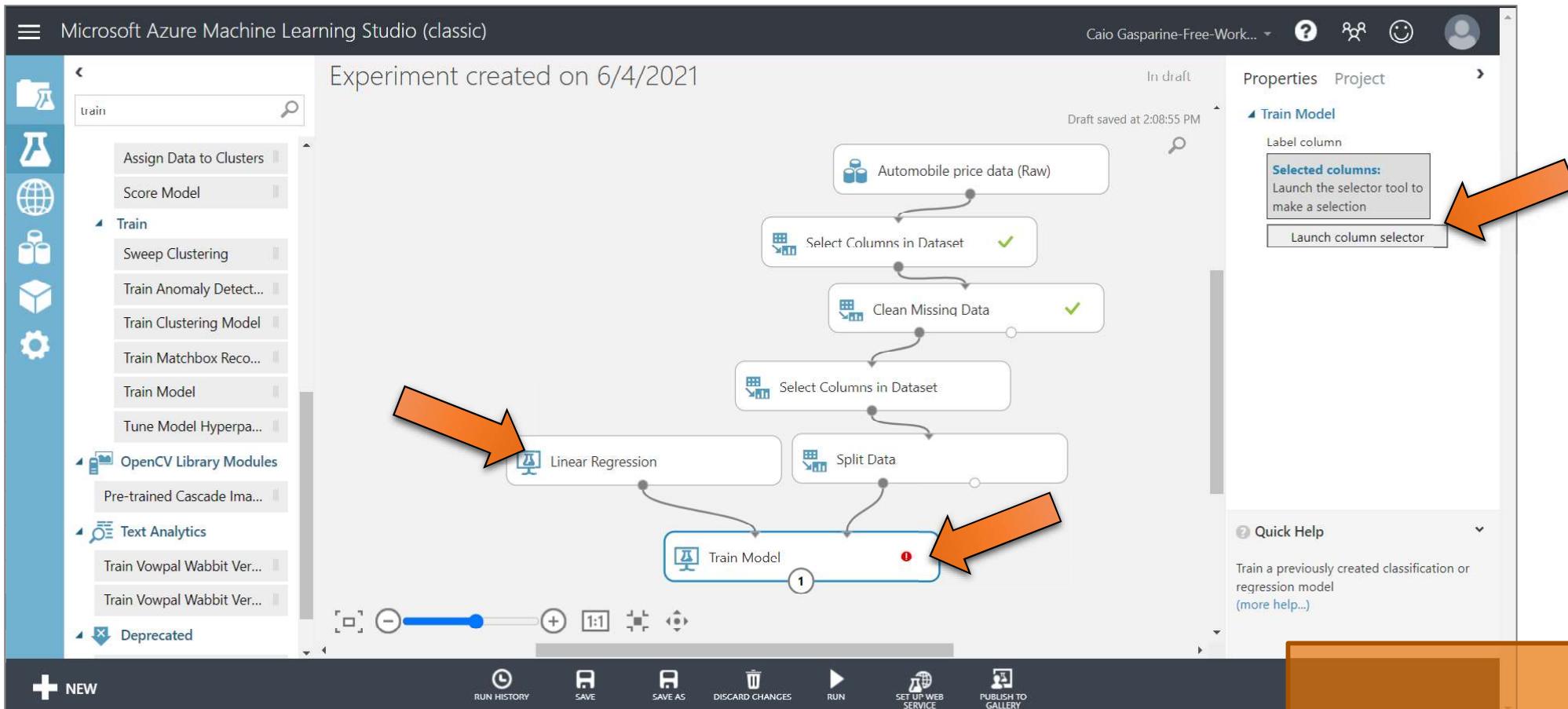


> Azure Machine Learning Studio | Practice



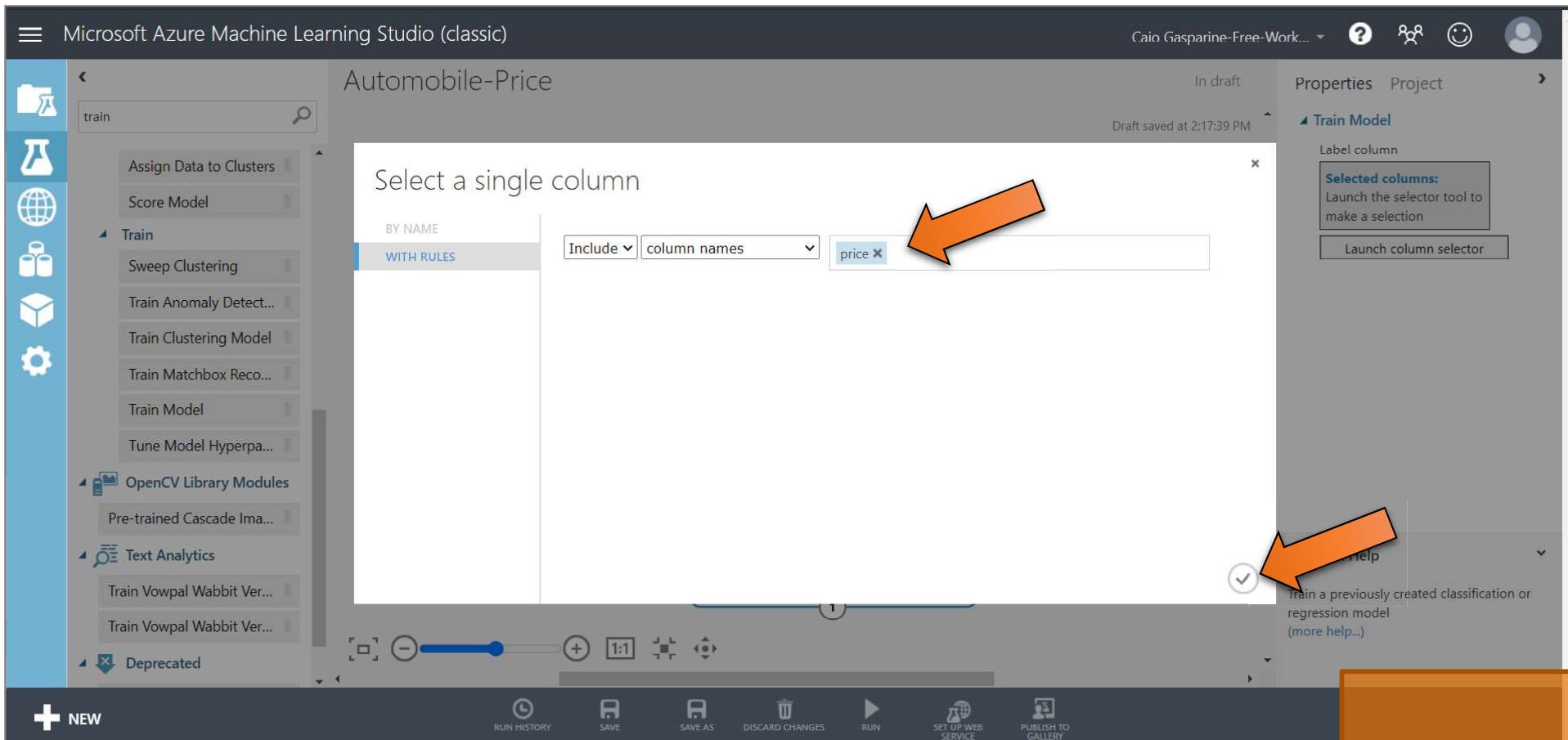
This is an example for
academic purposes

> Azure Machine Learning Studio | Practice



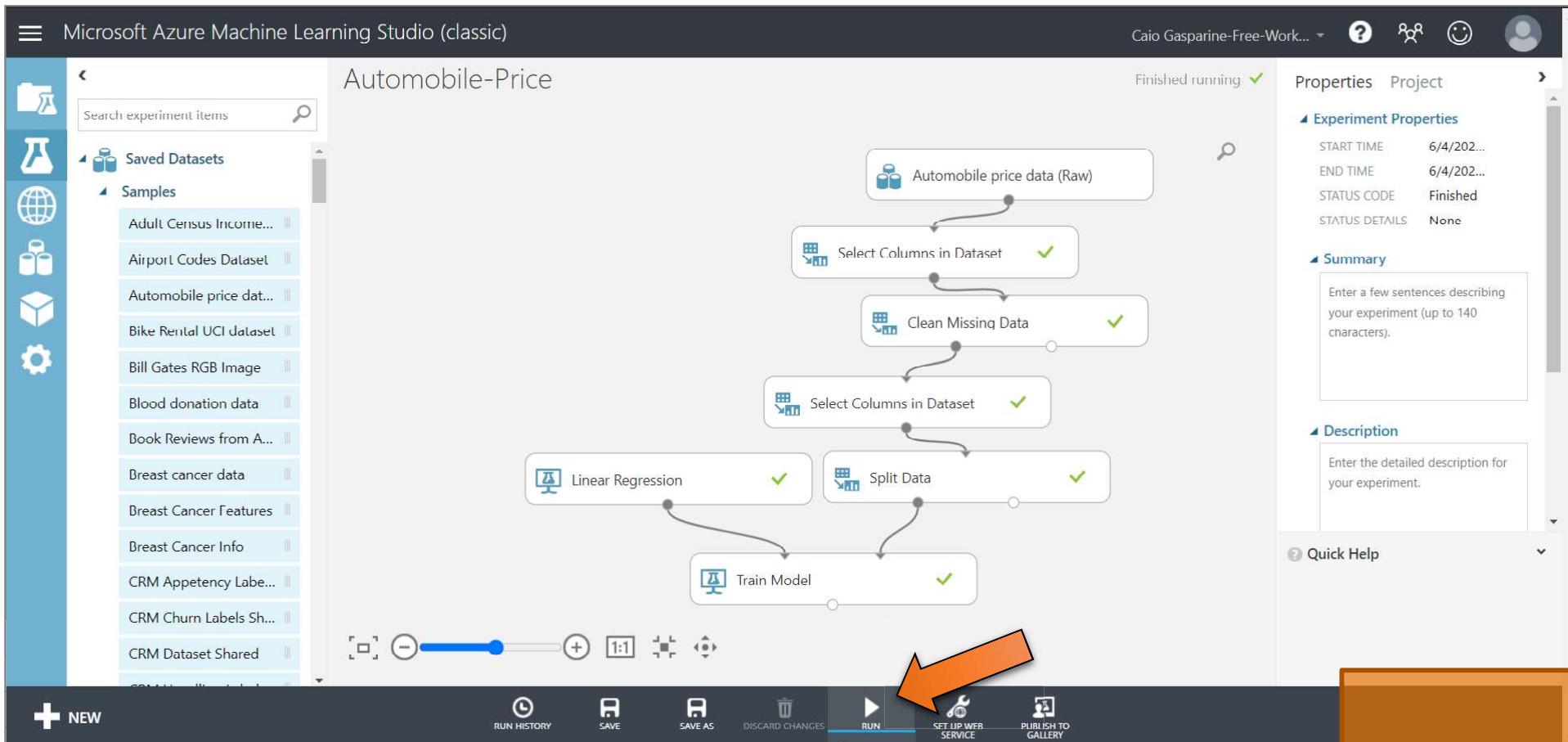
This is an example for
academic purposes

> Azure Machine Learning Studio | Practice



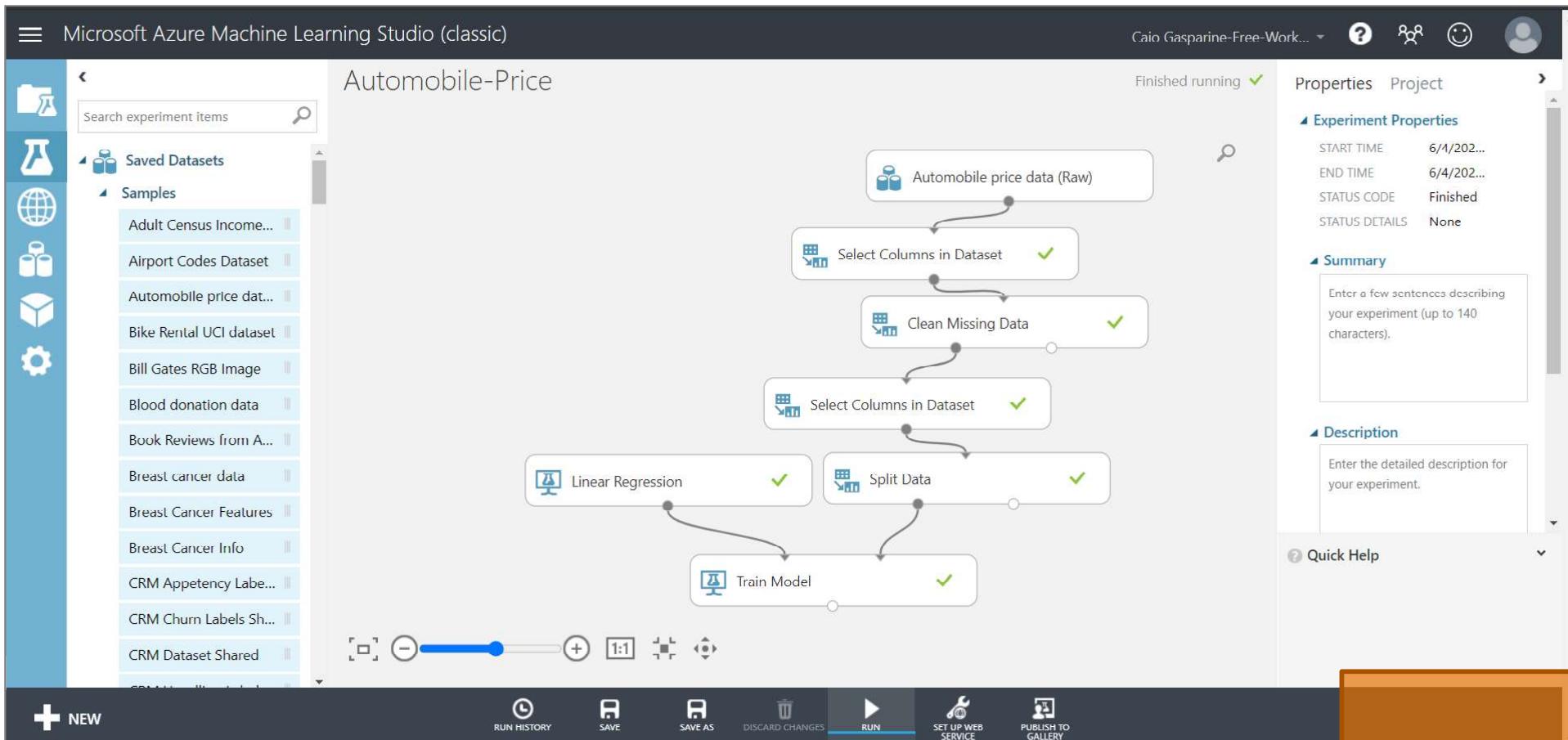
This is an example for
academic purposes

> Azure Machine Learning Studio | Practice



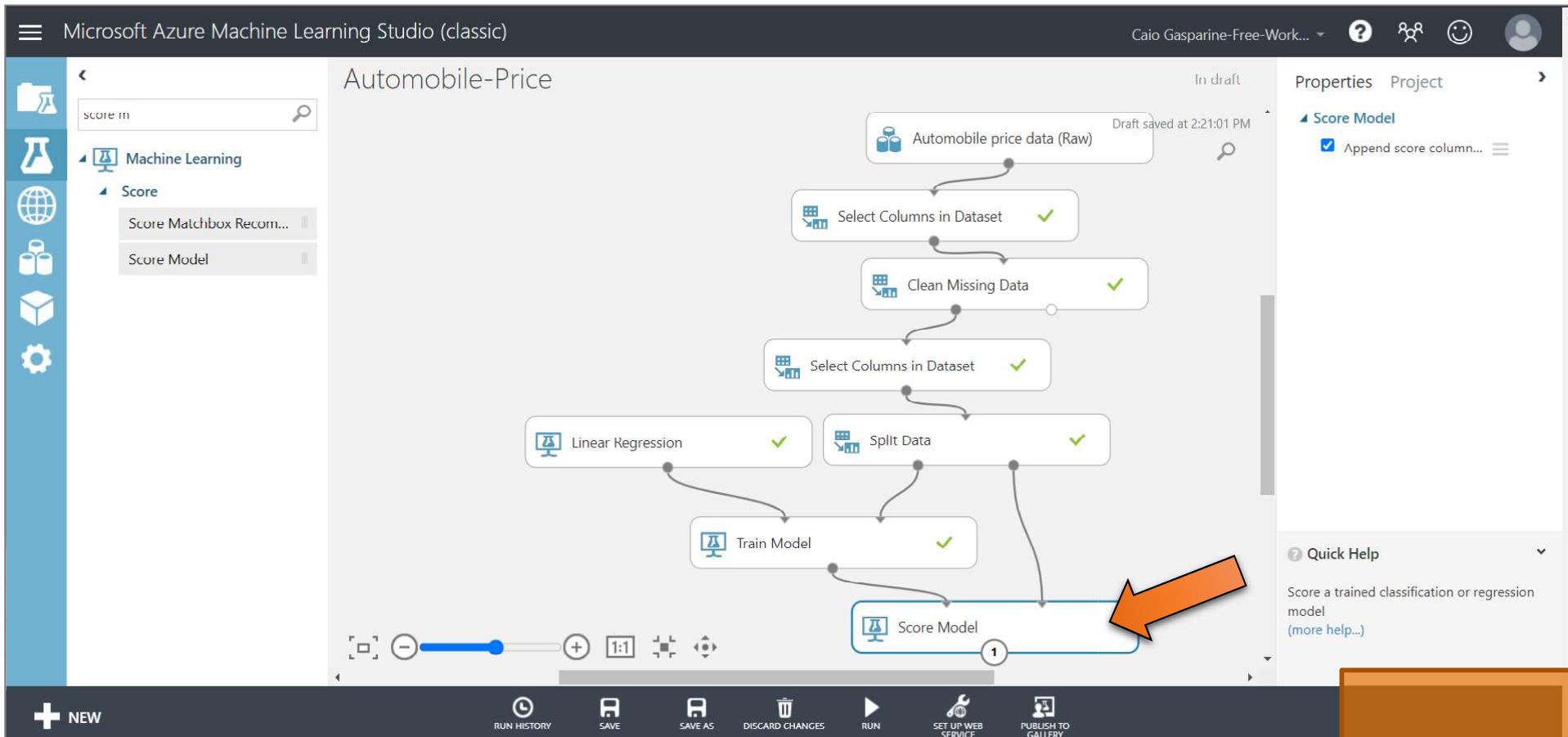
This is an example for
academic purposes

> Azure Machine Learning Studio | Practice



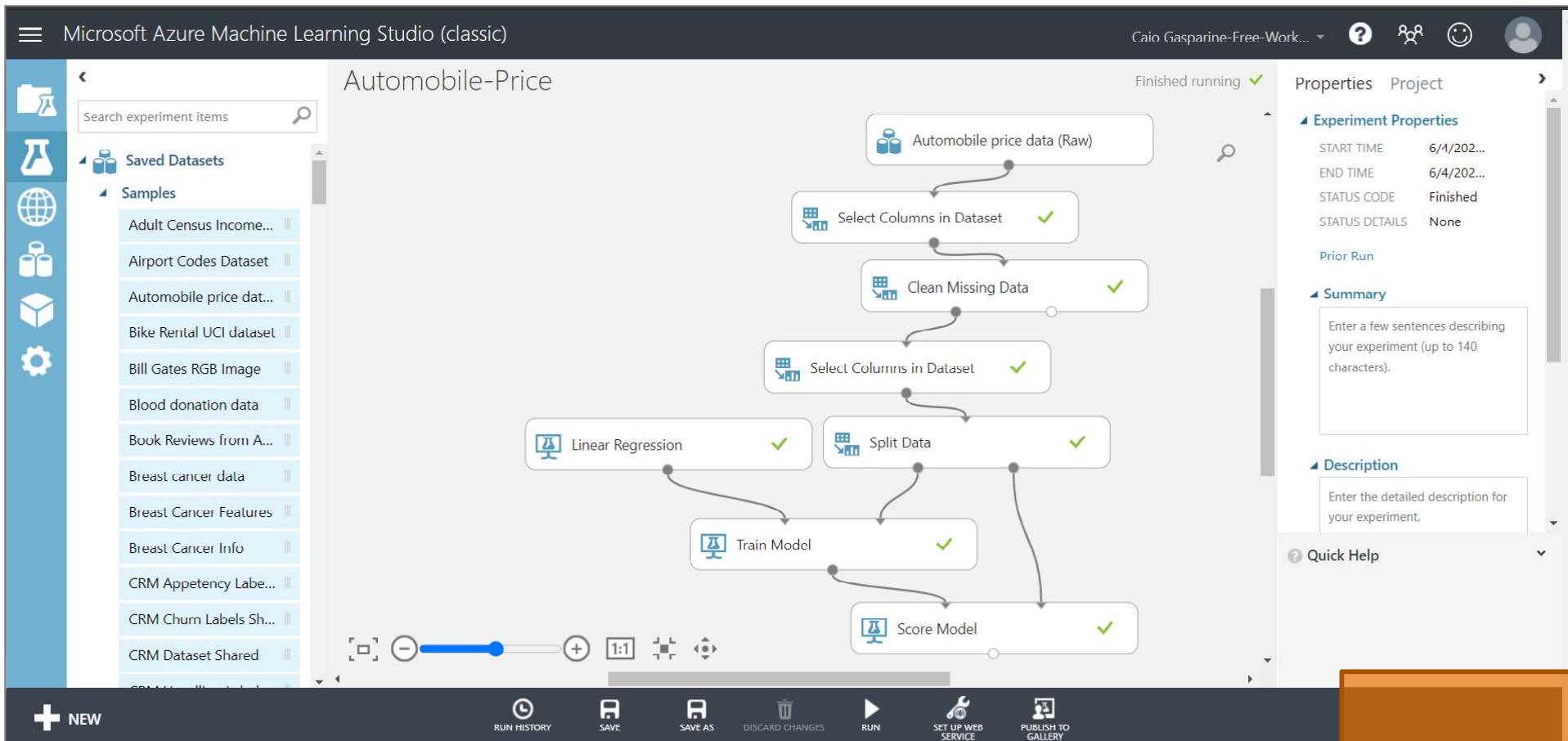
This is an example for
academic purposes

> Azure Machine Learning Studio | Practice



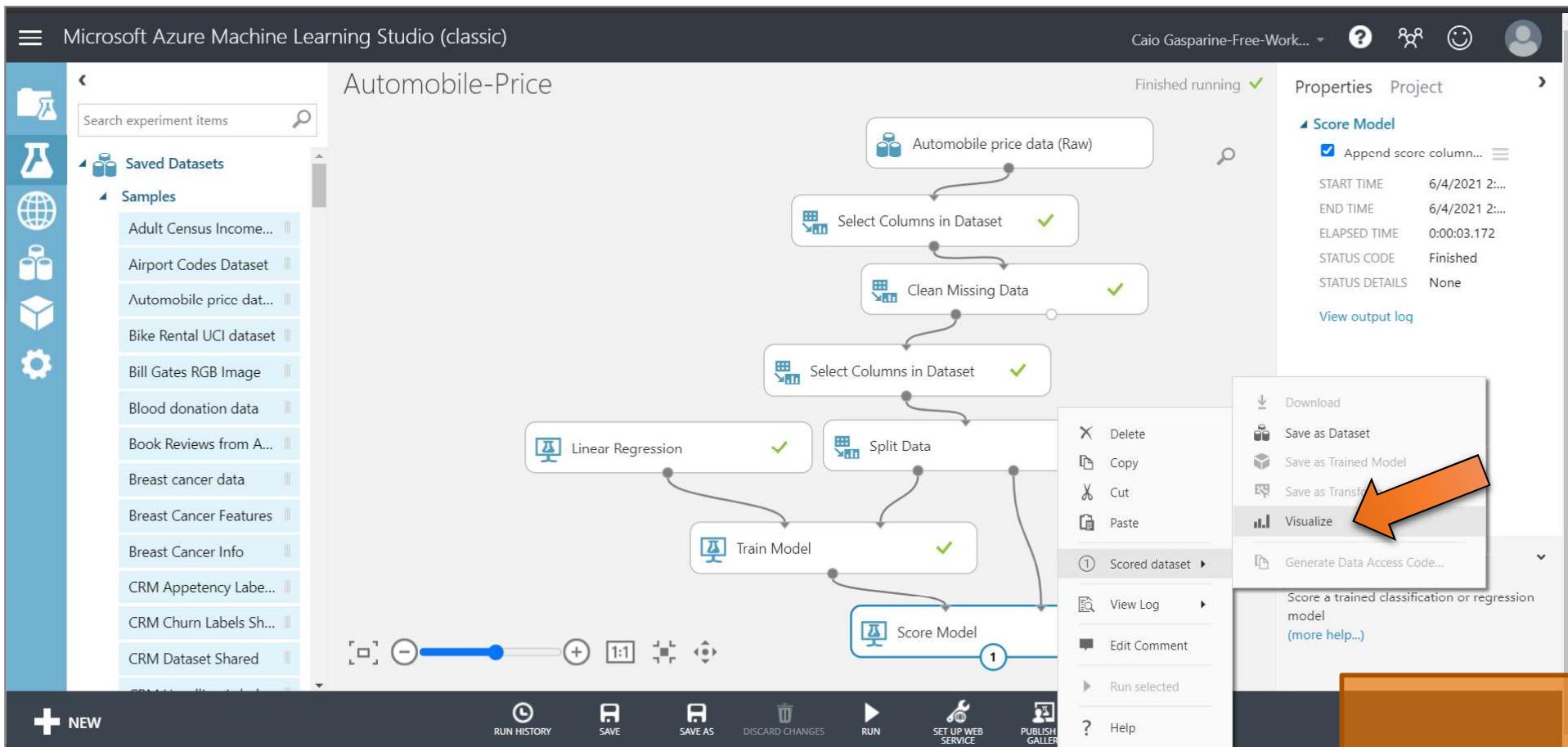
This is an example for
academic purposes

> Azure Machine Learning Studio | Practice



This is an example for
academic purposes

> Azure Machine Learning Studio | Practice



This is an example for
academic purposes

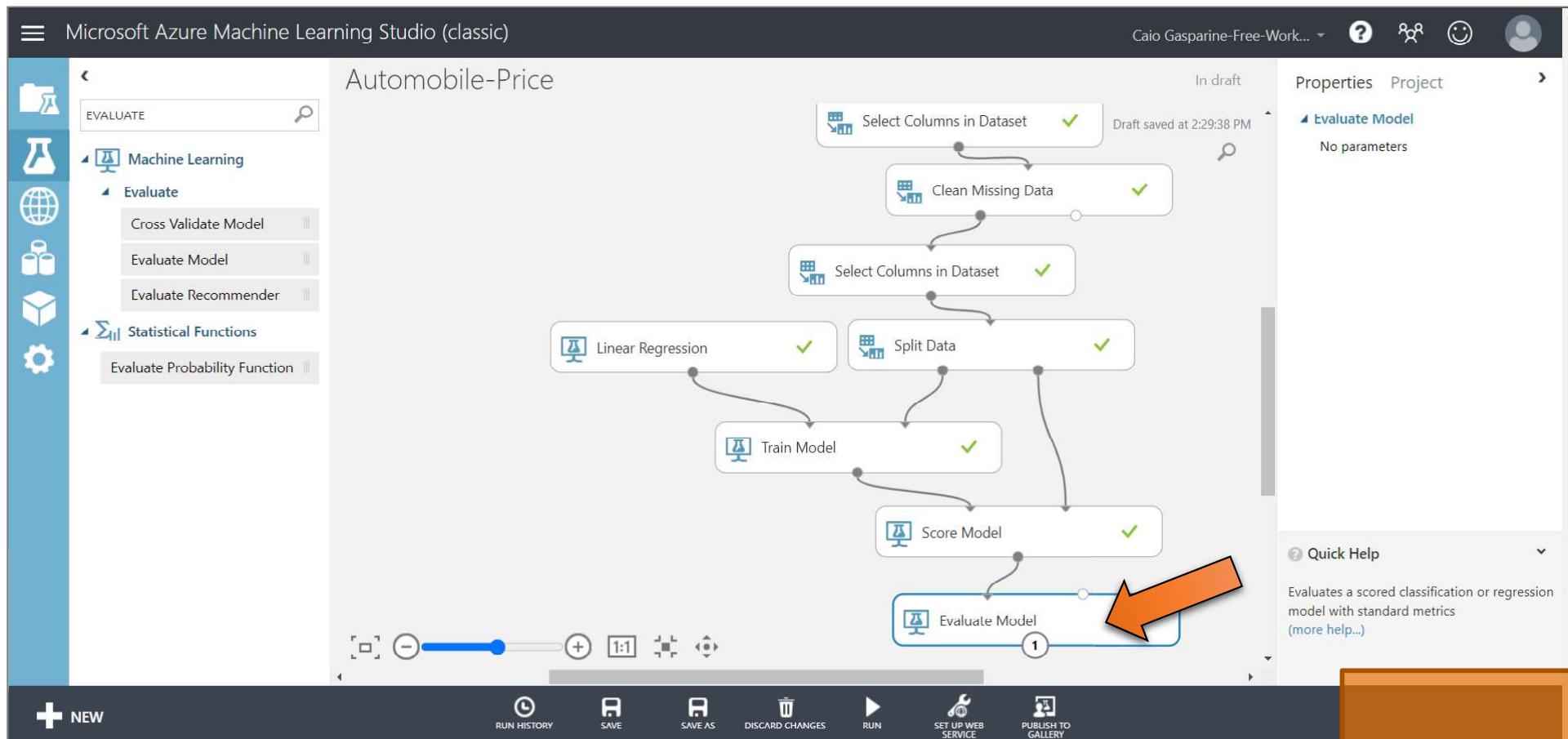
> Azure Machine Learning Studio | Practice

The screenshot shows the Microsoft Azure Machine Learning Studio (classic) interface. The main area displays a dataset titled "Automobile-Price" with 48 rows and 9 columns. The columns are: make, body-style, wheel-base, engine-size, horsepower, peak-rpm, highway-mpg, price, and Scored Labels. Below the table, there are two large blue buttons: "KNOWN VALUES" and "PREDICTED VALUES". To the right of the table, a sidebar provides statistical information and visualizations for the "Scored Labels" column. An orange arrow points from the sidebar towards the "Scored Labels" column header. The sidebar includes sections for "Statistics" and "Visualizations". The "Statistics" section lists various metrics: Mean (12437.776), Median (10208.7085), Min (5446.8479), Max (34960.6439), Standard Deviation (7323.458), Unique Values (46), Missing Values (0), and Feature Type (Numeric Score). The "Visualizations" section shows a histogram for the "Scored Labels" column, comparing it to "None".

make	body-style	wheel-base	engine-size	horsepower	peak-rpm	highway-mpg	price	Scored Labels
subaru	sedan	97	108	111	4800	29	11259	10286.204819
mitsubishi	hatchback	93.7	92	68	5500	38	6669	5446.847864
dodge	hatchback	93.7	90	68	5500	38	6229	6344.800711
honda	hatchback	86.6	92	76	6000	38	6855	5528.302953
alfa-romero	convertible	88.6	130	111	5000	27	16500	13498.476233
volvo	wagon	104.3	141	114	5400	28	16515	16097.608038
isuzu	hatchback	96	119	90	5000	29	11048	8315.257218

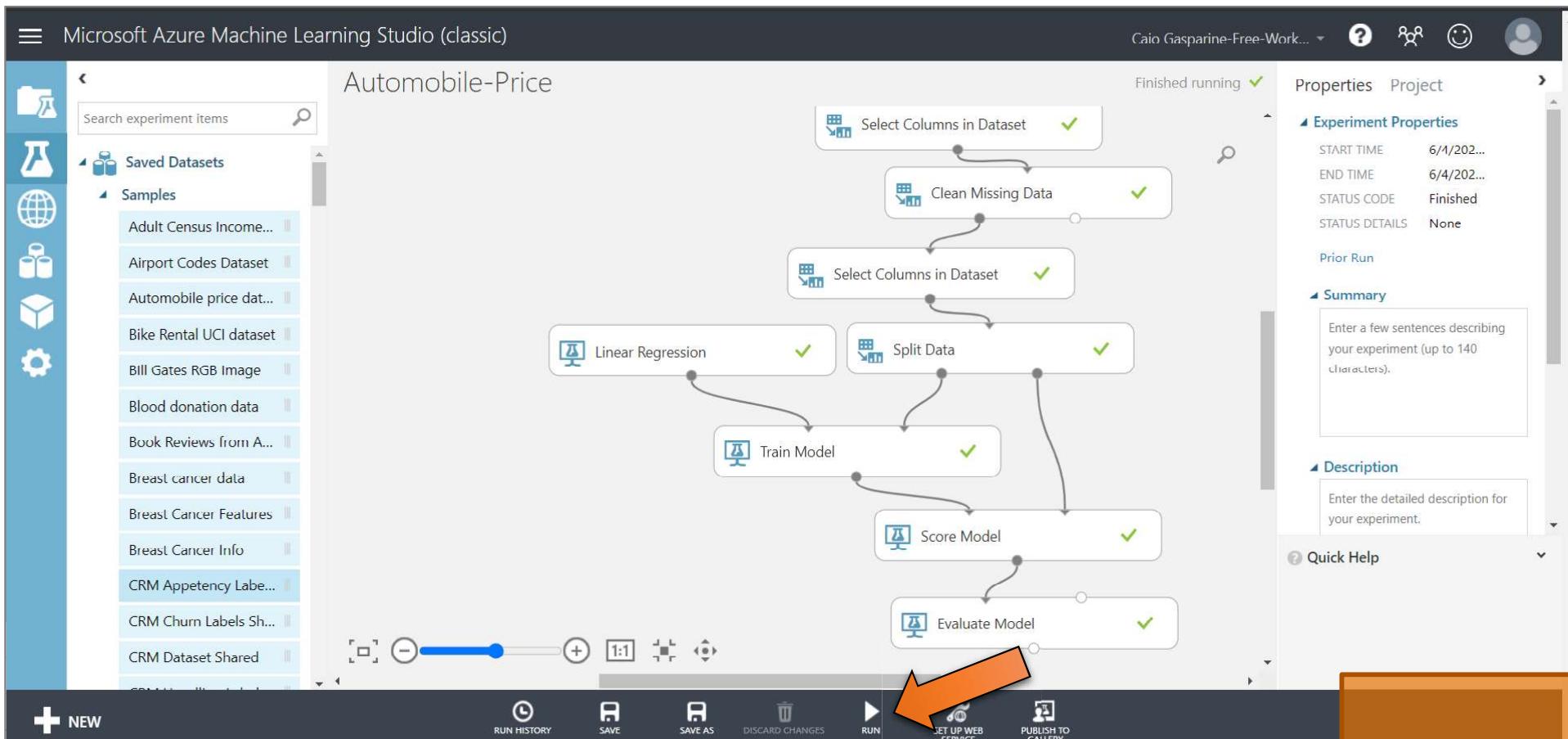
This is an example for
academic purposes

> Azure Machine Learning Studio | Practice



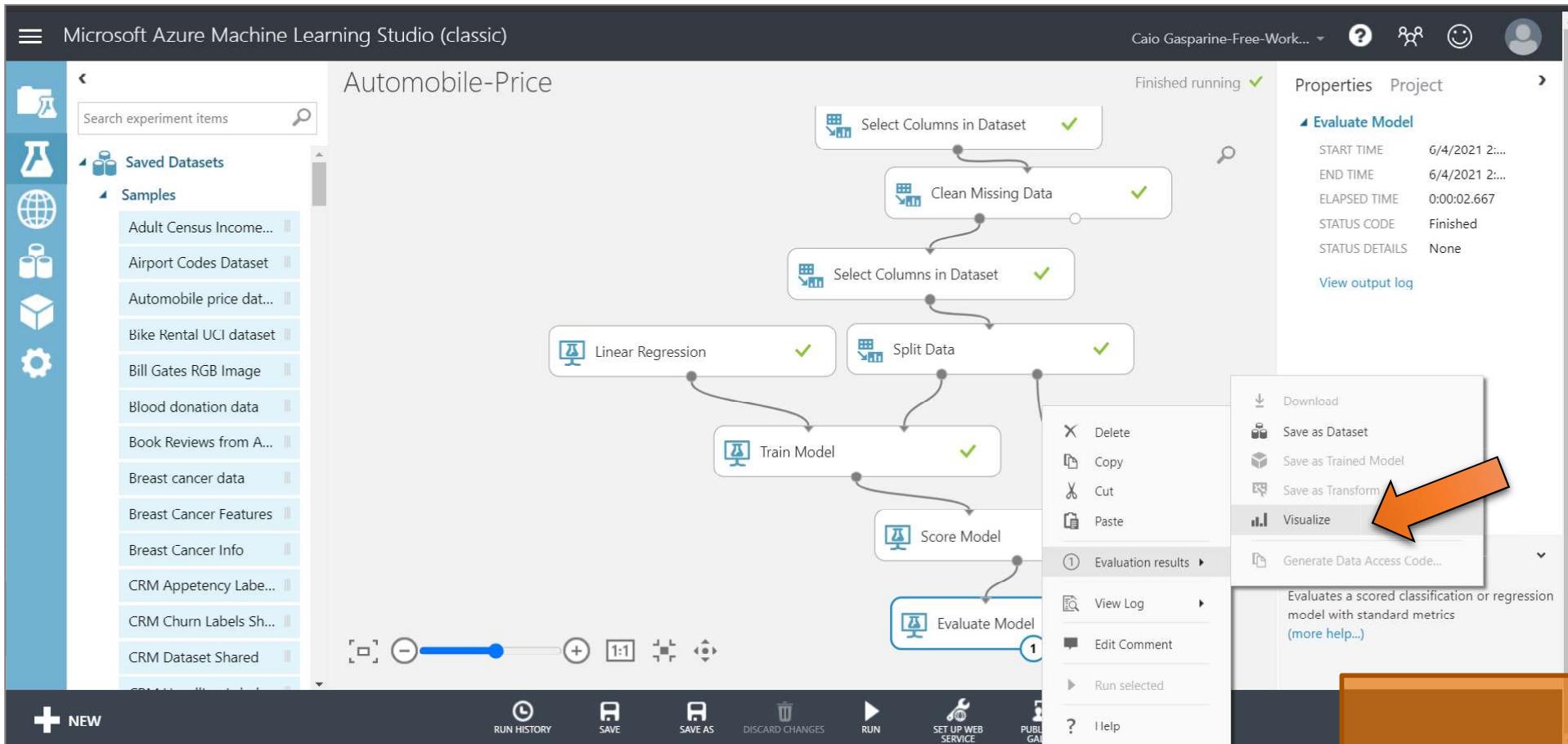
This is an example for
academic purposes

> Azure Machine Learning Studio | Practice



This is an example for
academic purposes

> Azure Machine Learning Studio | Practice



This is an example for
academic purposes

> Azure Machine Learning Studio | Practice

Microsoft Azure Machine Learning Studio (classic)

Automobile-Price

Automobile-Price > Evaluate Model > Evaluation results

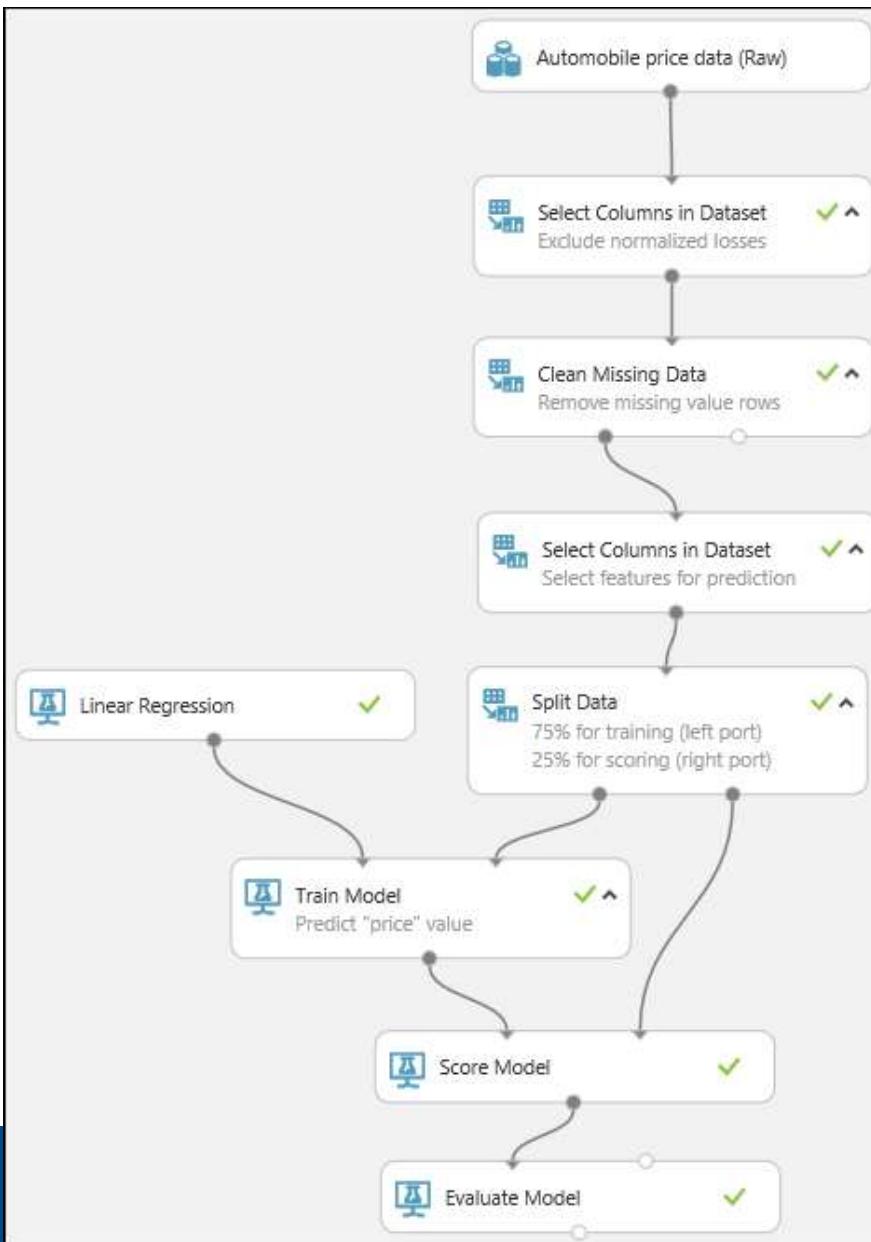
Metrics

Mean Absolute Error	1656.147651
Root Mean Squared Error	2456.983209
Relative Absolute Error	0.276606
Relative Squared Error	0.089608
Coefficient of Determination	0.910392

Error Histogram

This is an example for academic purposes

> Azure Machine Learning Studio | Practice

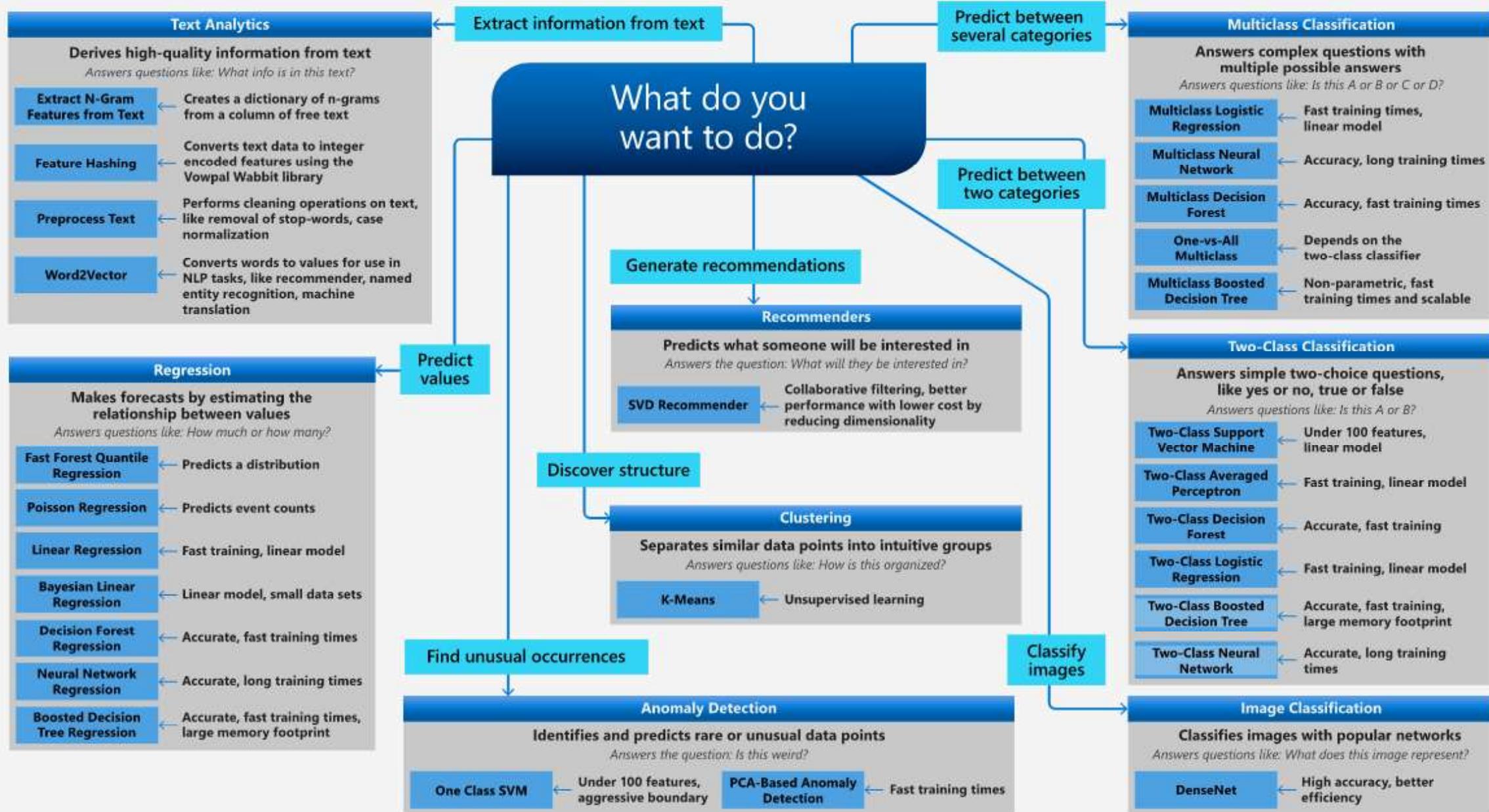


This is an example for
academic purposes



Microsoft Azure Machine Learning Algorithm Cheat Sheet

This cheat sheet helps you choose the best machine learning algorithm for your predictive analytics solution. Your decision is driven by both the nature of your data and the goal you want to achieve with your data.



Practice #2 / Part 2

Azure Machine Learning Studio

German Credit Data



Azure Machine Learning

> #2 Azure Machine Learning Studio (Classic) / Part 2

INSTRUCTIONS:

Record a video with 5-10 minutes explaining how to use **Azure Machine Learning Studio** and your discoveries.

- Follow each of the steps detailed in the Microsoft Tutorial and explain them.
- Explain that you are loading the file, select columns, cleaning missing data, splitting the file, etc.
- Explain why you are performing each step.
- Make your conclusion of the predicted values.
- Evaluate your model and explain the results.
- Analyze all the steps you performed and think how should be the behavior (processes) of an enterprise solution with the same purpose. What are the differences?

In your video explain how and why you performed each step and explain the final result.

EVALUATION:

Mark: 10 points (part 1 and 2)

- Ensure that you recorded yourself using the tool or using your own screenshots
- Ensure that you recorded all the performed steps
- Ensure that you analysed the results
- Explain what kind of ML you are using in this exercise and why

Will be considered: Your results, explanations, level of details, clarity to explain and presentation / video quality (preparation).

Due date: Week 6 class



> Azure Machine Learning Studio | Practice / Part 2

MAIN GOAL:

**Develop a predictive analytics
solution and deploy the model as an
Azure Machine Learning web service.**

MAIN STEPS:

Create a predictive experiment

<https://docs.microsoft.com/en-us/azure/machine-learning/classic/tutorial-part1-credit-risk>

Train and evaluate

<https://docs.microsoft.com/en-us/azure/machine-learning/classic/tutorial-part2-credit-risk-train>

Deploy web service

<https://docs.microsoft.com/en-us/azure/machine-learning/classic/tutorial-part3-credit-risk-deploy>

References



> References (1)

- Big Data Analytics Program, 2019/2020 – Georgian College, Barrie, Ontario
- Git Hub, ss - <https://mlops.githubapp.com/>
- Informatica – Data Catalog – <https://www.informatica.com/ca/products/data-catalog.html>
- Google Research – Publication Data Management Challenges in Production Machine Learning, Alkis Polyzotis, Martin A. Zinkevich, Steven Whang, Sudip Roy - <https://research.google/pubs/pub46178/>
- <https://research.google/pubs/pub46178/>
- https://databricks.com/session_na20/an-approach-to-data-quality-for-netflix-personalization-systems
- Microsoft Azure Machine Learning Studio – Official Documentation, December 2020 - <https://docs.microsoft.com/en-us/azure/machine-learning>
- Microsoft Azure Machine Learning Cheat Sheet. Machine Learning Algorithm Cheat Sheet for Azure Machine Learning designer, December 2020. <https://docs.microsoft.com/en-us/azure/machine-learning/algorithm-cheat-sheet>
- DataCookbook, Understanding Data Lineage, <https://youtu.be/lUxgWb6WpF0>
- Microsoft, What is Azure Data Catalog?, <https://docs.microsoft.com/en-us/azure/data-catalog/overview>
- Microsoft, Azure Data Platform End-to-End, Implement a Modern Data Platform Architecture, Official Material
- H2O.ai, RMSLE, website, <https://www.h2o.ai/community/glossary/rmsle-root-mean-squared-logarithmic-error>
- Kaggle, Understanding the metric: RMSLE, website, <https://www.kaggle.com/carlolepelaars/understanding-the-metric-rmsle>

> References (2)

- Big Data Analytics Program, 2019/2020 – Georgian College, Barrie, Ontario
- Kaggle, dataset: Forest Fires in Brazil, Luis Gustavo Modeli, <https://www.kaggle.com/gustavomodelli/forest-fires-in-brazil>
- Data Iku, ebook, 2021 Trends: Where Enterprise AI is headed next?
- Microsoft, Success by Design Implementation Guide, First Edition, 2021
- Microsoft, Quickstart: Create your first data science experiment in Machine Learning Studio (classic), <https://docs.microsoft.com/en-us/azure/machine-learning/classic/create-experiment>
- Microsoft, Automated ML, <https://docs.microsoft.com/en-us/azure/machine-learning/concept-automated-ml>
- KD Nuggets, website, May 2021, <https://www.kdnuggets.com/2020/02/data-scientists-automl-replace.html>
- Microsoft, Tutorial 1: Predict credit risk, <https://docs.microsoft.com/en-us/azure/machine-learning/classic/tutorial-part1-credit-risk>
- Microsoft, Tutorial 2: Train credit risk models, <https://docs.microsoft.com/en-us/azure/machine-learning/classic/tutorial-part2-credit-risk-train>
- Microsoft, Tutorial 3: Deploy credit risk model, <https://docs.microsoft.com/en-us/azure/machine-learning/classic/tutorial-part3-credit-risk-deploy>
- Towards Data, Understanding AUC - ROC Curve, website, <https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5>



Georgian

END OF DAY 5