

# Database Management Systems

Azure Machine Learning\*

\* not among the exam topics

# Azure Machine Learning

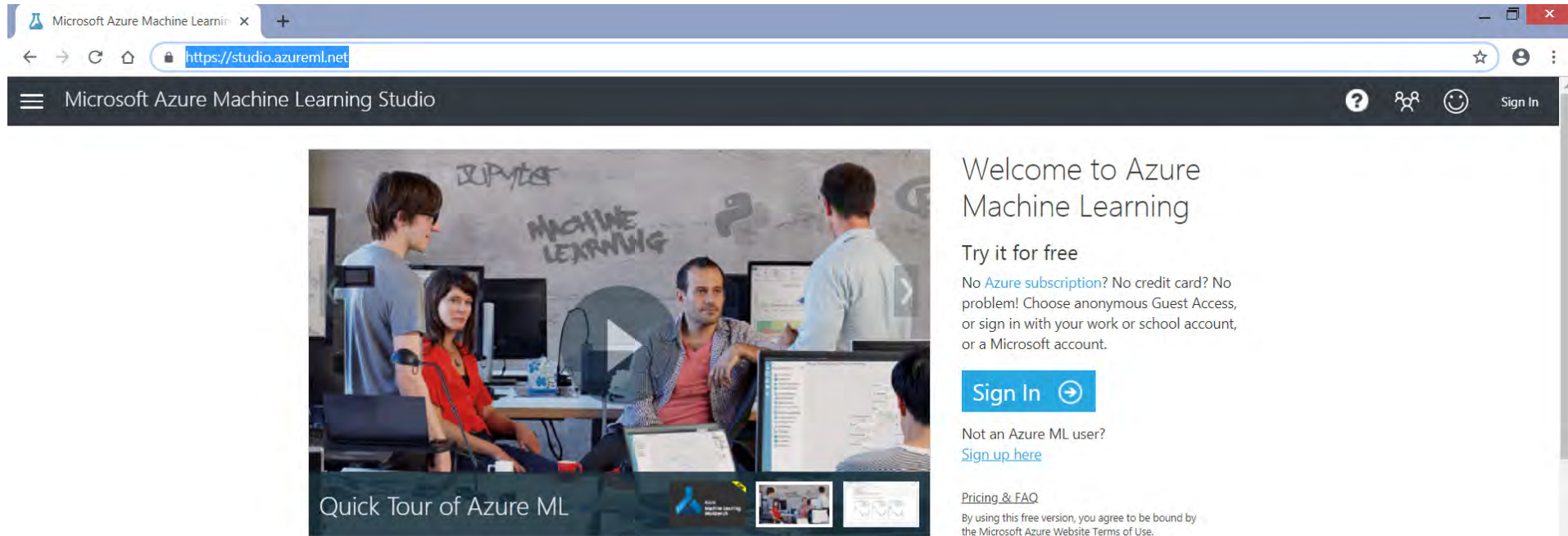
- build / train / deploy machine learning models
- Machine Learning Studio

## \* Machine Learning Studio

- create the model
  - obtain the data, prepare the data, define the features
- training
  - choose the algorithm, apply the algorithm
- testing the model
  - predict car prices for other vehicles
  - test the quality of the results

# Data science Experiment – Car Price Prediction

- <https://studio.azureml.net/> -> Sign in



Microsoft Azure Machine Learning Studio

Quick Tour of Azure ML

## Welcome to Azure Machine Learning

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 user?  
[Sign up here](#)

[Pricing & FAQ](#)

By using this free version, you agree to be bound by the Microsoft Azure Website Terms of Use.

## Announcements **NEW!**

### Azure Machine Learning Studio R Runtime Upgrade

Aired on October 31, 2018

The R language engine in the Execute R Script module of Azure Machine Learning Studio has added a new R runtime version -- Microsoft R Open (MRO) 3.4.4. MRO 3.4.4 is based on open-source CRAN R 3.4.4 and is therefore compatible with packages that works with that version of R.

[> Learn More](#)

### Mining Campaign Funds

Aired on August 03, 2017

Play with 2016 Presidential Campaign finance data while learning how to prepare a large dataset for machine learning by processing and engineering features. This sample experiment works on a 2.5 GB dataset and will take about 20 minutes to run in its entirety.

[> Learn More](#)

### Inside the Data Science VM

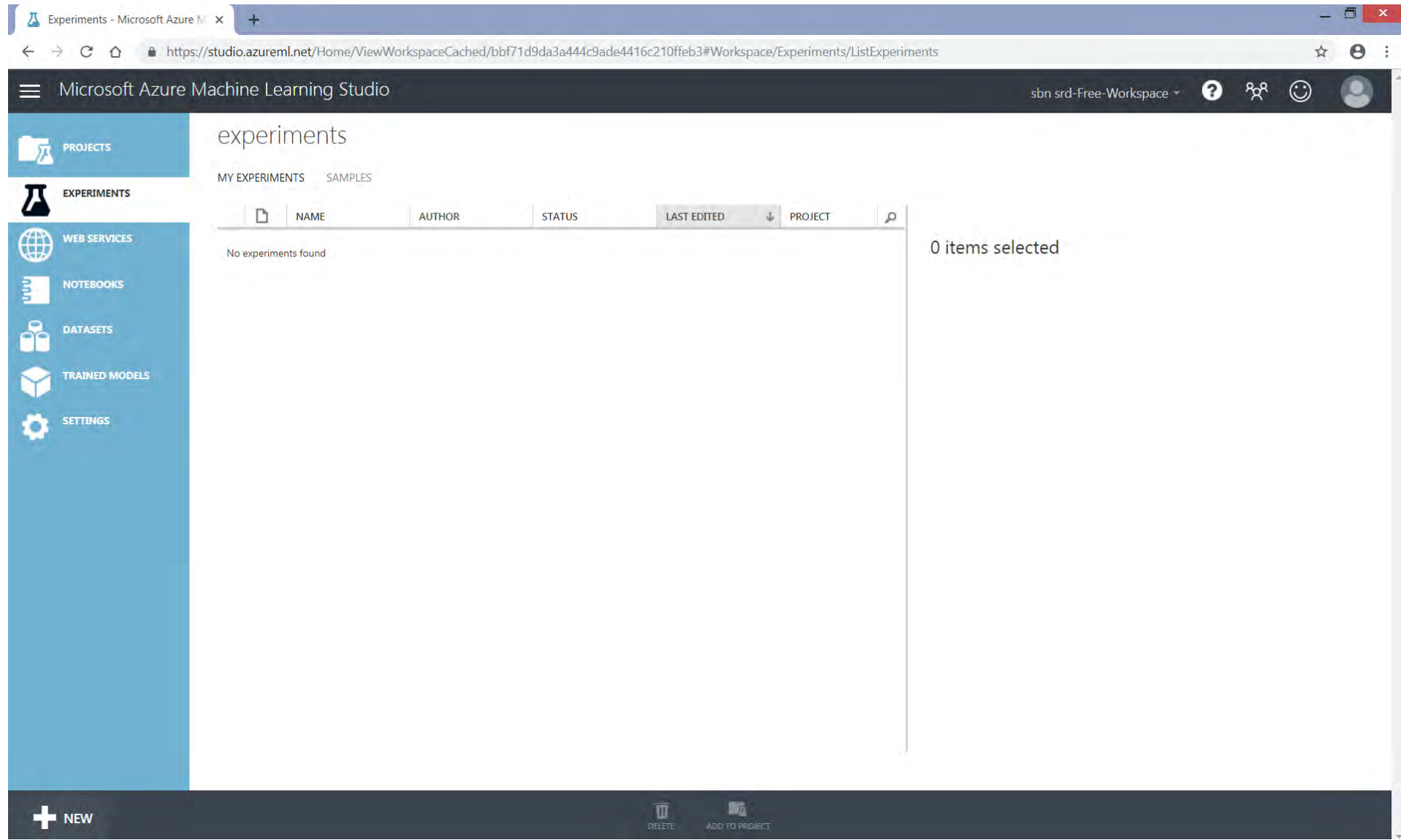
Aired on June 21, 2016

DSVM is a custom Azure Virtual Machine image that is published on the Azure marketplace and available on both Windows and Linux. It contains several popular data science and development tools both from Microsoft and from the open source community all pre-installed and pre-configured and ready to use. We will cover best practices that would show how you can use the DSVM effectively to run your next data science or analytics project.

[> Learn More](#)

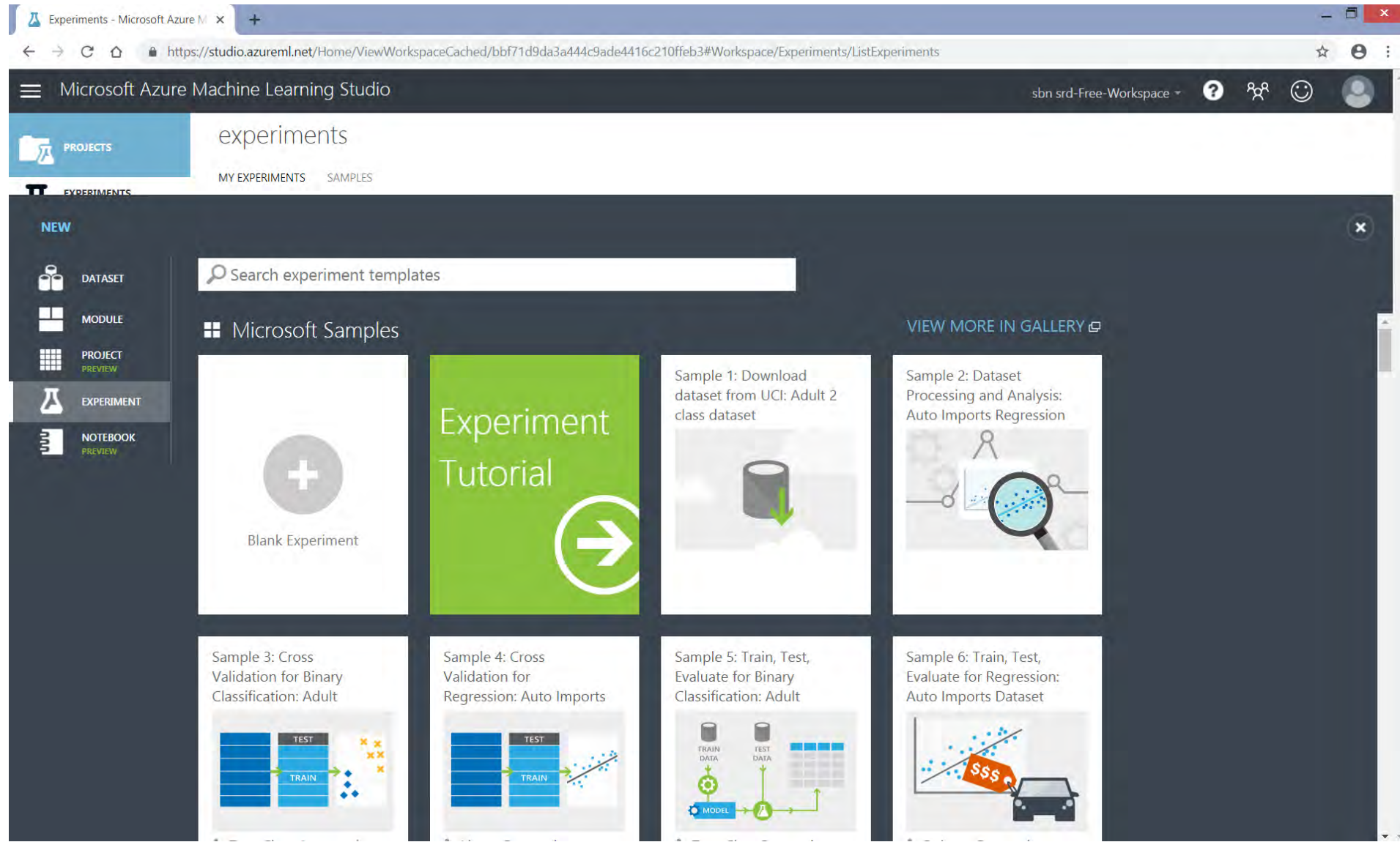
# Car Price Prediction

\* create an experiment: + New



# Car Price Prediction

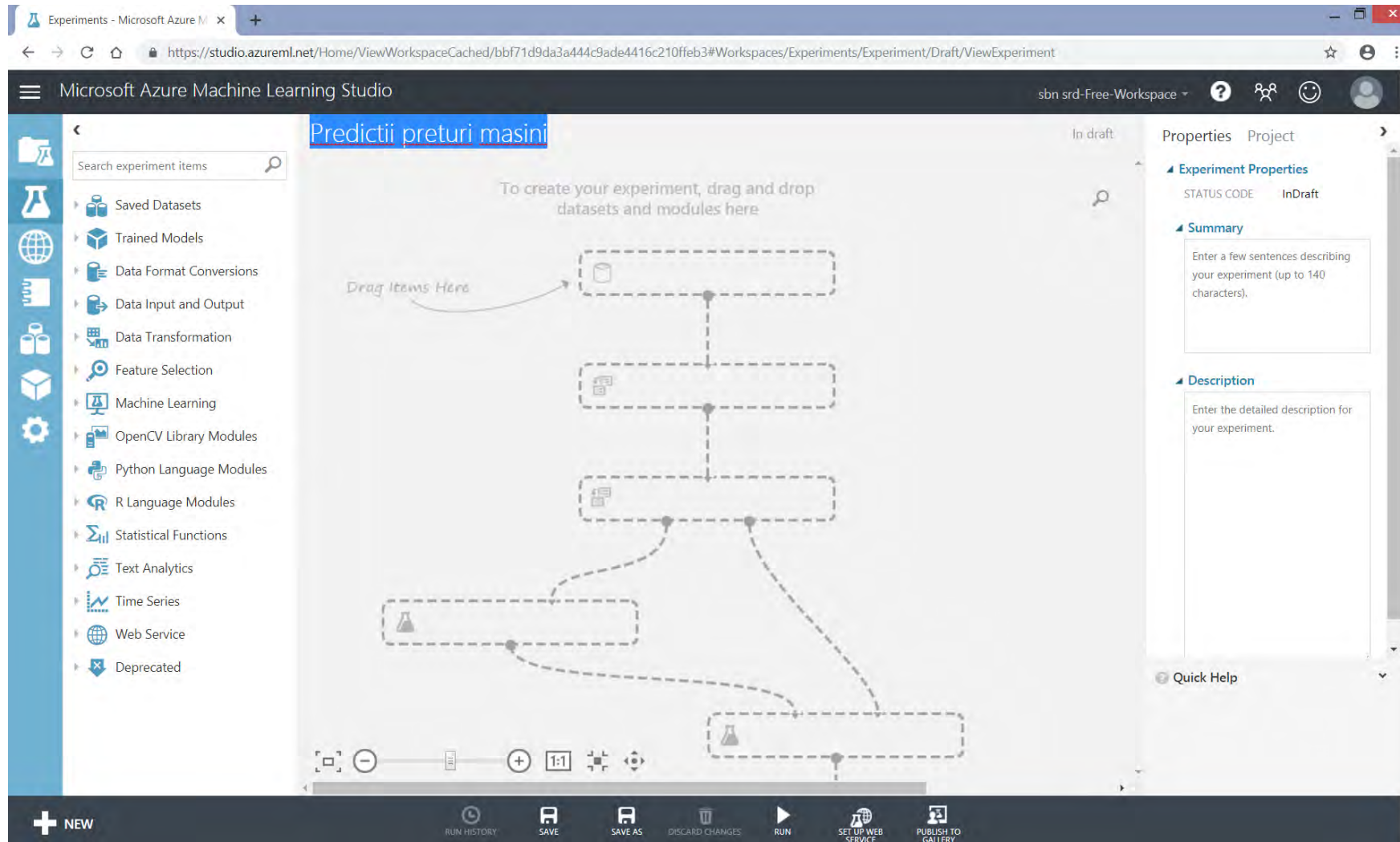
\* create an experiment: Blank Experiment





# Car Price Prediction

- \* create an experiment
  - experiment name



# Car Price Prediction

\* selecting the data source

- dataset *Automobile price data (raw)*

The screenshot displays the Microsoft Azure Machine Learning Studio interface. The browser address bar shows the URL: <https://studio.azureml.net/Home/ViewWorkspaceCached/bbf71d9da3a444c9ade4416c210ffeb3#Workspaces/Experiments/Experiment/Draft/ViewExperiment>. The page title is "Microsoft Azure Machine Learning Studio".

On the left sidebar, under "Saved Datasets", the "Automobile price data (...)" dataset is selected. The main workspace area is titled "Predictii preturi masini" and contains a diagram with several dashed boxes representing modules. A text overlay says "To create your experiment, drag and drop datasets and modules here" and "Drag Items Here" with an arrow pointing to the first module. The diagram shows a flow from a dataset module to a data preparation module, then to a model training module, and finally to a model evaluation module. There are also some additional modules at the bottom.

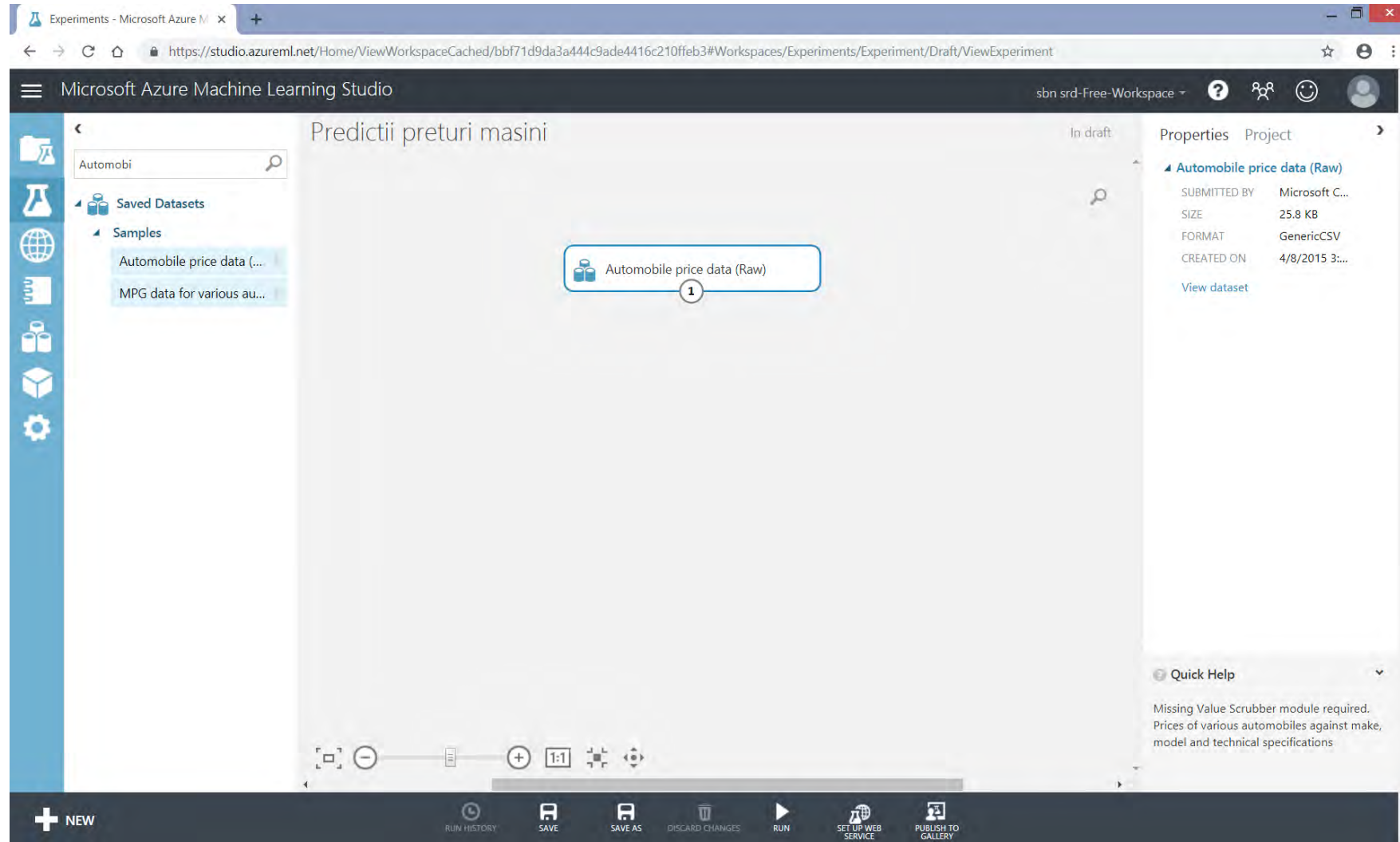
On the right sidebar, the "Properties" tab is active, showing "Experiment Properties" with a status of "InDraft". Below this is a "Summary" section with a text input field and a "Description" section with another text input field. A "Quick Help" link is at the bottom of the sidebar.

The bottom of the interface features a dark blue toolbar with icons for "NEW", "RUN HISTORY", "SAVE", "SAVE AS", "DISCARD CHANGES", "RUN", "SET UP WEB SERVICE", and "PUBLISH TO GALLERY".

# Car Price Prediction

\* selecting the data source

- drag & drop dataset *Automobile price data (raw)* onto the canvas

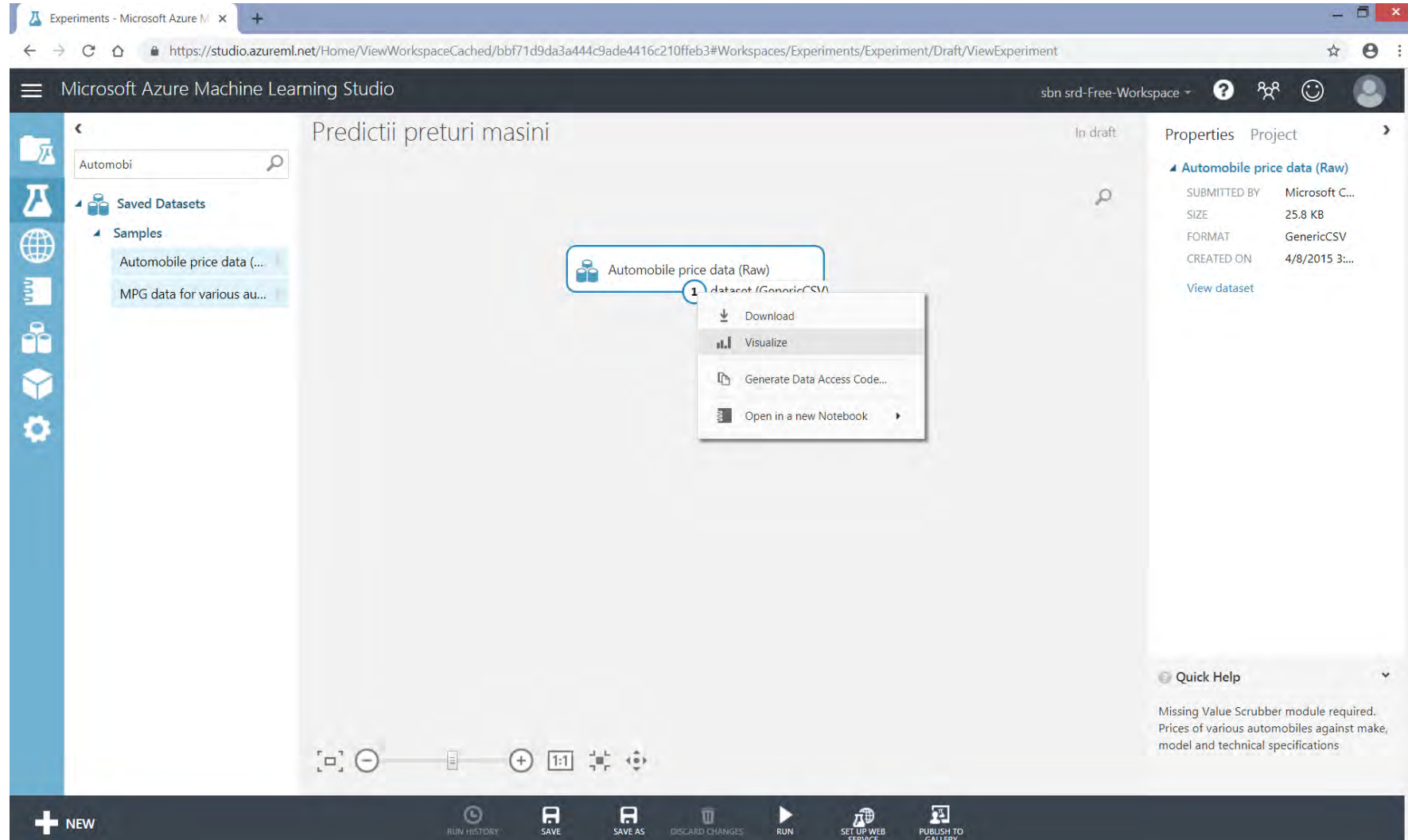




# Car Price Prediction

\* displaying the data

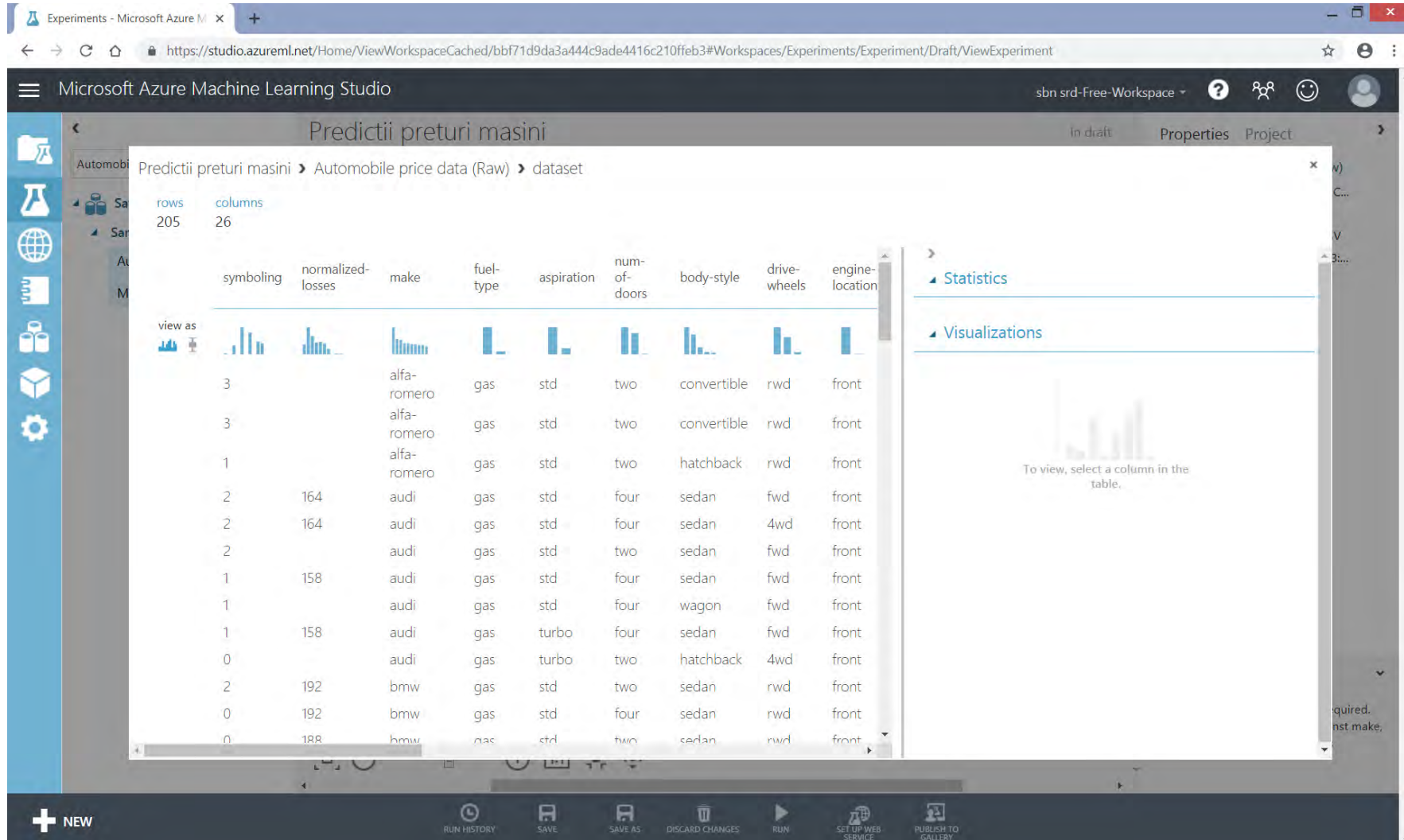
- dataset output port -> *Visualize*



# Car Price Prediction

\* displaying the data

- row – data about a car



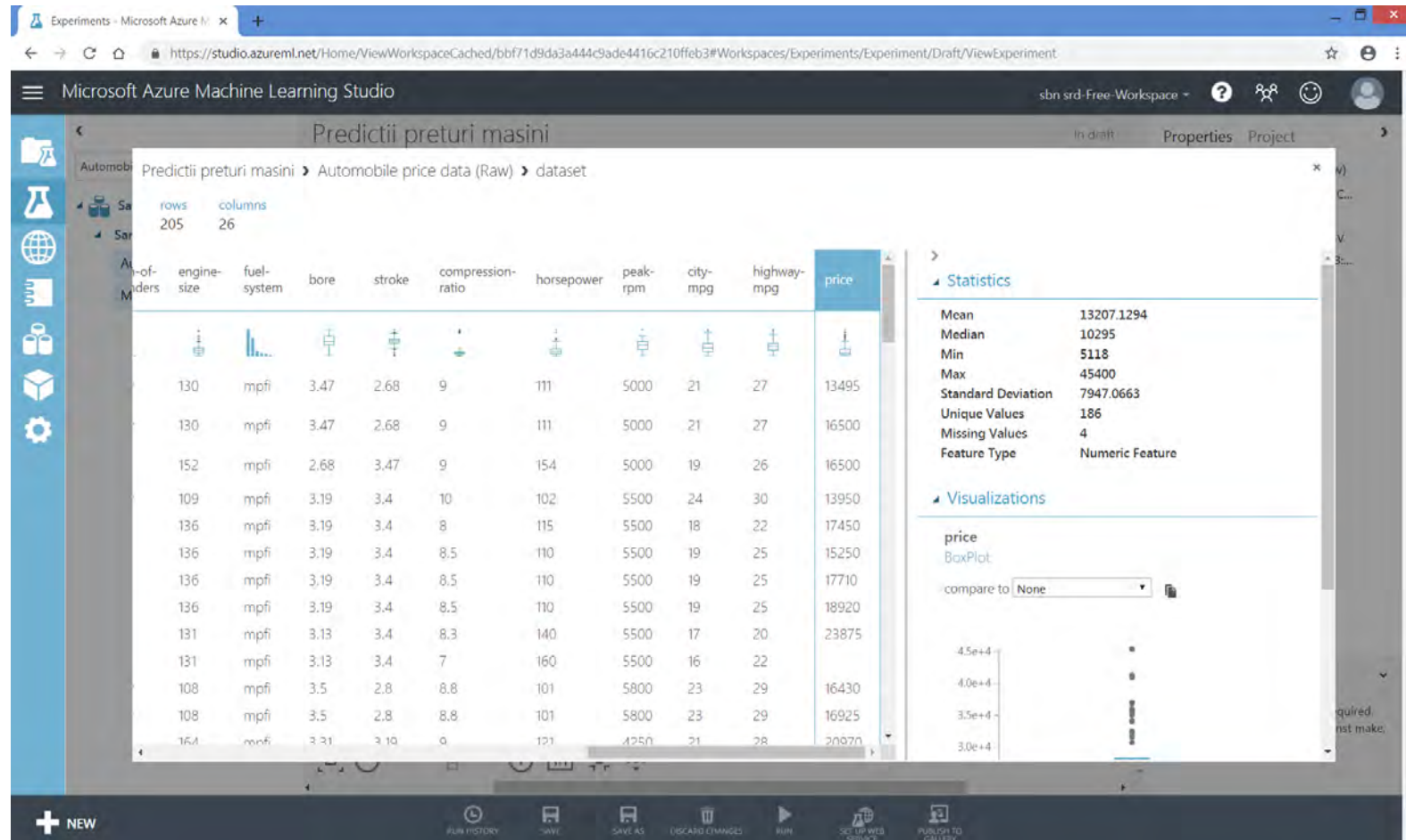
The screenshot displays the Microsoft Azure Machine Learning Studio interface. The main workspace shows a dataset titled "Predictii preturi masini" (Car Price Prediction) with 205 rows and 26 columns. The dataset is named "Automobile price data (Raw)". The columns are: symboling, normalized-losses, make, fuel-type, aspiration, num-of-doors, body-style, drive-wheels, and engine-location. The data is displayed in a table view, showing the first 10 rows. The interface includes a sidebar with navigation icons, a top navigation bar, and a bottom toolbar with buttons for "NEW", "RUN HISTORY", "SAVE", "SAVE AS", "DISCARD CHANGES", "RUN", "SET UP WEB SERVICE", and "PUBLISH TO GALLERY".

symboling	normalized-losses	make	fuel-type	aspiration	num-of-doors	body-style	drive-wheels	engine-location
3		alfa-romero	gas	std	two	convertible	rwd	front
3		alfa-romero	gas	std	two	convertible	rwd	front
1		alfa-romero	gas	std	two	hatchback	rwd	front
2	164	audi	gas	std	four	sedan	fwd	front
2	164	audi	gas	std	four	sedan	4wd	front
2		audi	gas	std	two	sedan	fwd	front
1	158	audi	gas	std	four	sedan	fwd	front
1		audi	gas	std	four	wagon	fwd	front
1	158	audi	gas	turbo	four	sedan	fwd	front
0		audi	gas	turbo	two	hatchback	4wd	front

# Car Price Prediction

\* displaying the data

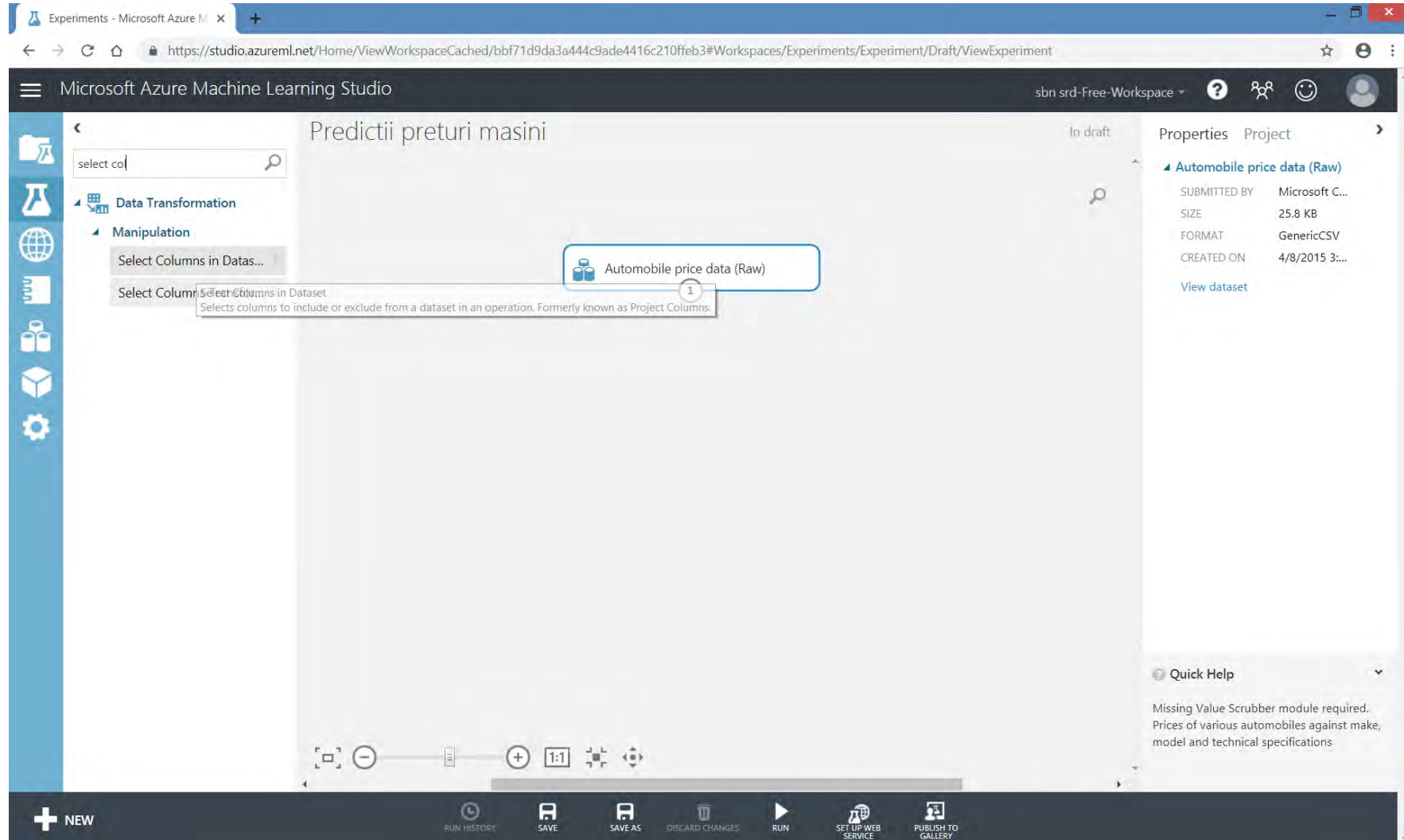
- columns – variables
- *target* column - *price*



# Car Price Prediction

\* preparing the data

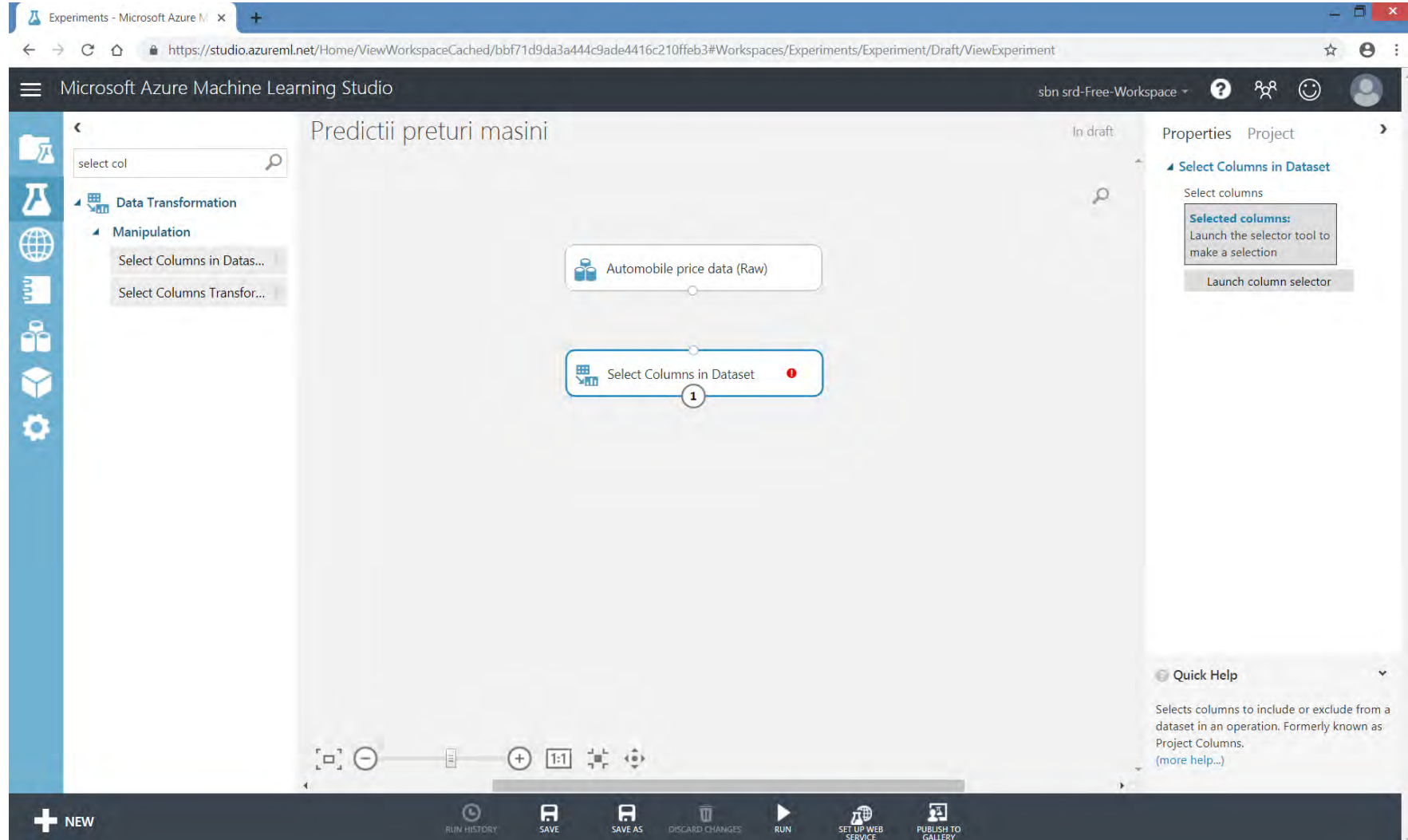
- eliminate column with missing values – *normalized-losses*



# Car Price Prediction

\* preparing the data

- eliminate column with missing values - *Select Columns in Dataset* module

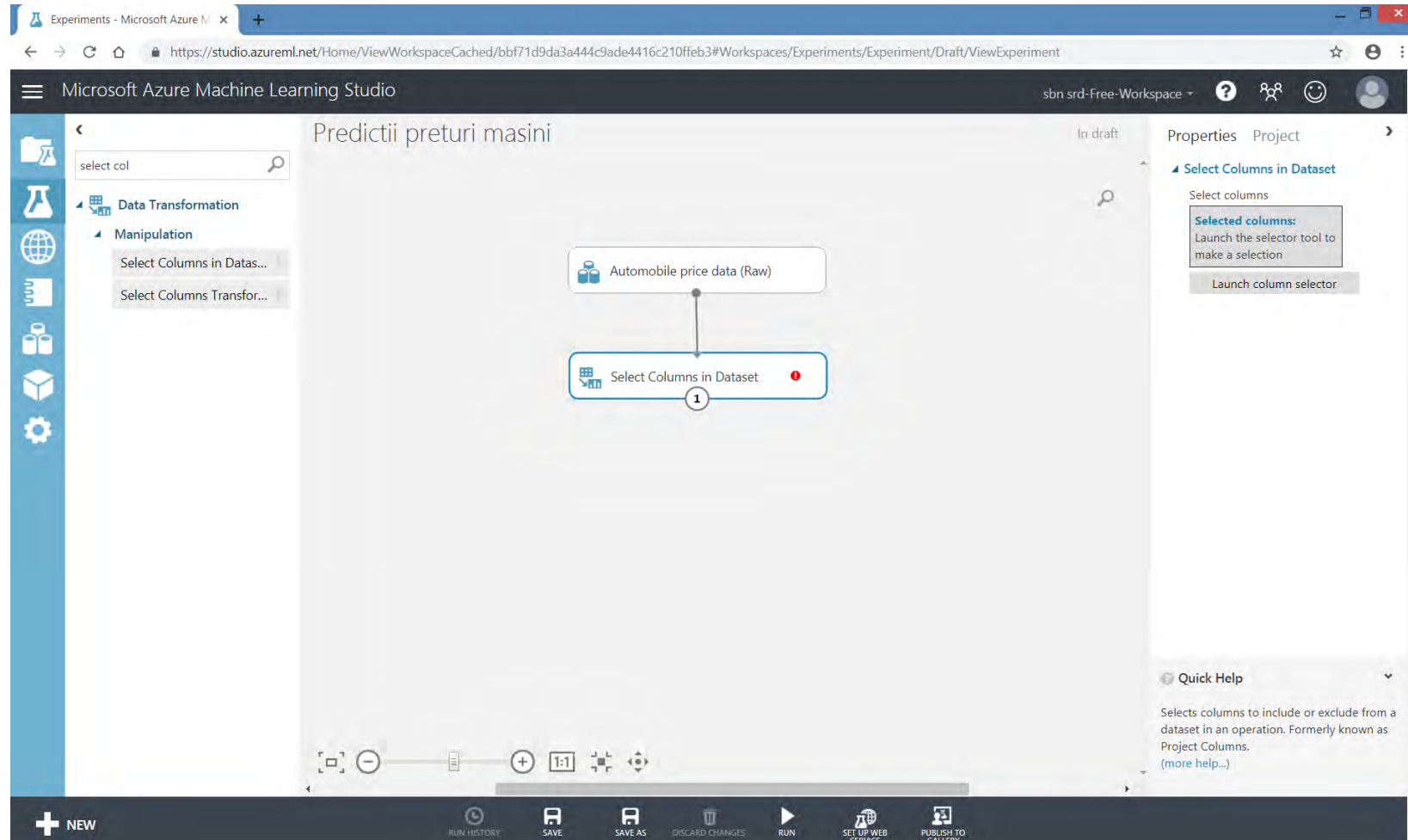




# Car Price Prediction

\* preparing the data

- eliminate column with missing values



# Car Price Prediction

\* preparing the data

- eliminate column with missing values

- *Select Columns in Dataset*

- *Launch column selector*

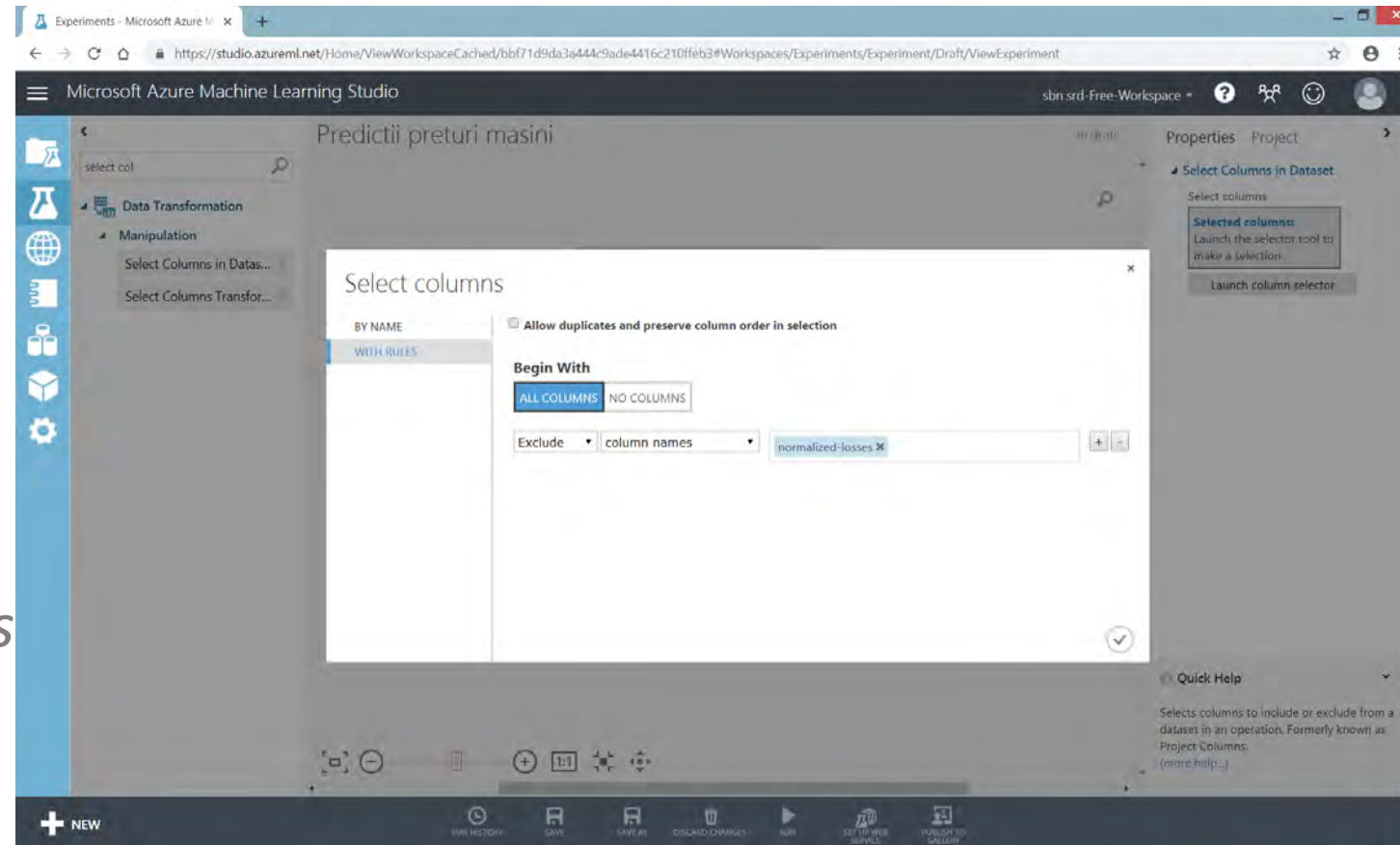
- *With Rules*

- *Begin With*

- *All Columns*

- *Exclude*

- *normalized-losses*



# Car Price Prediction

\* preparing the data

- eliminate rows with missing values – *Clean Missing Data* module

The screenshot displays the Microsoft Azure Machine Learning Studio interface. The top navigation bar shows the workspace name 'sbn srd-Free-Workspace'. The main workspace area is titled 'Predictii preturi masini' and contains a data pipeline with two modules: 'Automobile price data (Raw)' and 'Select Columns in Dataset'. The 'Select Columns in Dataset' module is highlighted with a blue border and a circled '1' below it. On the left sidebar, the 'Data Transformation' section is expanded, showing the 'Clean Missing Data' module under the 'Manipulation' category. The right sidebar shows the 'Properties' panel for the 'Select Columns in Dataset' module, with the 'Select Columns in Dataset' tab selected. The 'Selected columns' section shows 'All columns' and 'Exclude column names: normalized-losses'. The bottom status bar includes icons for 'NEW', 'RUN HISTORY', 'SAVE', 'SAVE AS', 'DISCARD CHANGES', 'RUN', 'SET UP WEB SERVICE', and 'PUBLISH TO GALLERY'.

# Car Price Prediction

## \* preparing the data

- eliminate rows with missing values

- *Clean Missing Data*

- *Cleaning mode*

- *Remove entire row*

The screenshot displays the Microsoft Azure Machine Learning Studio interface. The main workspace shows a workflow titled "Predictii preturi masini" (Car Price Predictions). The workflow consists of three steps: "Automobile price data (Raw)", "Select Columns in Dataset", and "Clean Missing Data". The "Clean Missing Data" step is highlighted with a blue border and numbered 1 and 2. On the left sidebar, the "Data Transformation" section is expanded, showing the "Clean Missing Data" option under "Manipulation". On the right sidebar, the "Properties" panel for the "Clean Missing Data" step is visible. It shows the "Columns to be cleaned" as "All columns", the "Minimum missing value ratio" as "0", the "Maximum missing value ratio" as "1", and the "Cleaning mode" set to "Remove entire row". The bottom of the interface features a toolbar with icons for "NEW", "RUN HISTORY", "SAVE", "SAVE AS", "DISCARD CHANGES", "RUN", "SET UP WEB SERVICE", and "PUBLISH TO GALLERY".