

Database Management Systems

Azure Machine Learning*

* not among the exam topics

Car Price Prediction

* running the experiment

- *Run*

The screenshot displays the Microsoft Azure Machine Learning Studio interface. The main workspace shows a workflow titled "Predictii preturi masini" (Car Price Prediction). The workflow consists of three steps: "Automobile price data (Raw)", "Select Columns in Dataset" (marked with a green checkmark), and "Clean Missing Data" (also marked with a green checkmark). The status bar at the top right indicates "Finished running" with a green checkmark. On the right side, the "Properties" panel is open, showing "Experiment Properties" with fields for "START TIME", "END TIME", "STATUS CODE" (Finished), and "STATUS DETAILS" (None). Below this, there are sections for "Summary" and "Description", each with a text input area. 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".

Car Price Prediction

* displaying the data

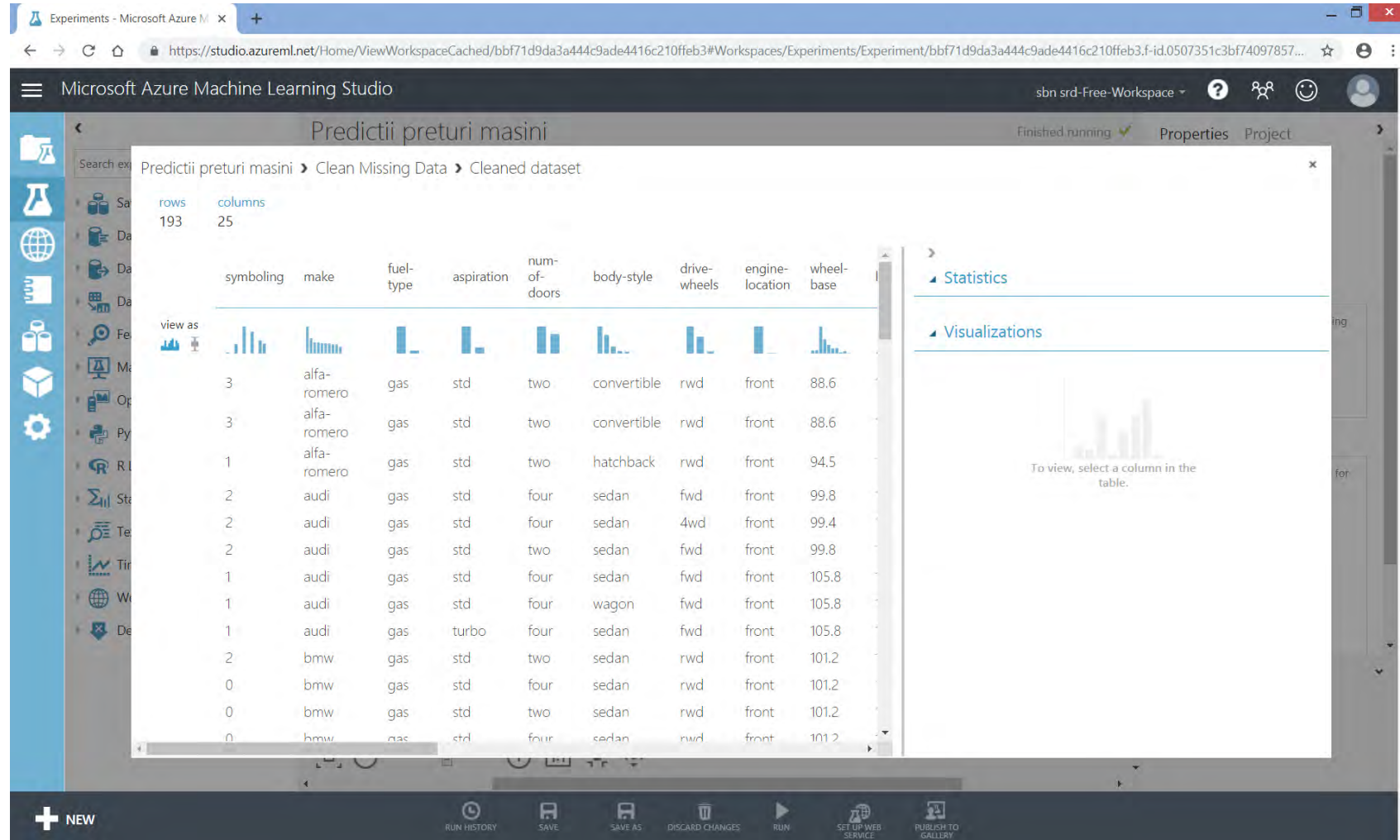
- *Clean Missing Data* module -> left output port -> *Visualize*

The screenshot displays the Microsoft Azure Machine Learning Studio interface. The main workspace shows a workflow titled "Predictii preturi masini" (Car Price Prediction). The workflow consists of three modules connected sequentially: "Automobile price data (Raw)", "Select Columns in Dataset", and "Clean Missing Data". The "Clean Missing Data" module has a green checkmark indicating it has completed successfully. A context menu is open over the "Clean Missing Data" module, showing options: "Download", "Save as Dataset", "Save as Trained Model", "Save as Transform", "Visualize" (highlighted), "Generate Data Access Code...", and "Open in a new Notebook". The left sidebar contains a search bar and a list of experiment items categories: "Saved Datasets", "Data Format Conversions", "Data Input and Output", "Data Transformation", "Feature Selection", "Machine Learning", "OpenCV Library Modules", "Python Language Modules", "R Language Modules", "Statistical Functions", "Text Analytics", "Time Series", "Web Service", and "Deprecated". The right sidebar shows the "Properties" and "Project" tabs. Under "Experiment Properties", it lists: "START TIME: 4/25/20...", "END TIME: 4/25/20...", "STATUS CODE: Finished", and "STATUS DETAILS: None". Under "Summary", it prompts to "Enter a few sentences describing your experiment (up to 140 characters)". Under "Description", it prompts to "Enter the detailed description for your experiment". 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

* displaying the data

- *Clean Missing Data* module -> left output port -> *Visualize*



Car Price Prediction

* defining the *features*

- used to create the predictive model
- *Select Columns in Dataset* module

The screenshot displays the Microsoft Azure Machine Learning Studio interface. The main workspace shows a workflow titled "Predictii preturi masini" (Car Price Predictions) in draft status. The workflow consists of four modules connected sequentially:

- Automobile price data (Raw)**: The starting data source.
- Select Columns in Dataset**: A module with a green checkmark, indicating it has been successfully configured.
- Clean Missing Data**: A module with a green checkmark, indicating it has been successfully configured.
- Select Columns in Dataset**: A module with a red exclamation mark and a "1" in a circle, indicating it is currently selected and needs configuration.

The left sidebar shows the "Data Transformation" category expanded, with the "Manipulation" sub-category selected. The "Select Columns in Dataset" module is highlighted in the sidebar.

The right sidebar shows the "Properties" pane for the selected "Select Columns in Dataset" module. It includes a "Launch column selector" button and a "Quick Help" section at the bottom.

At the bottom of the interface, there is a toolbar with icons for "NEW", "RUN HISTORY", "SAVE", "SAVE AS", "DISCARD CHANGES", "RUN", "SFT TIP WEB SERVICE", and "PUBLISH TO GALLERY".

Car Price Prediction

* defining the *features*

- *Select Columns in Dataset*

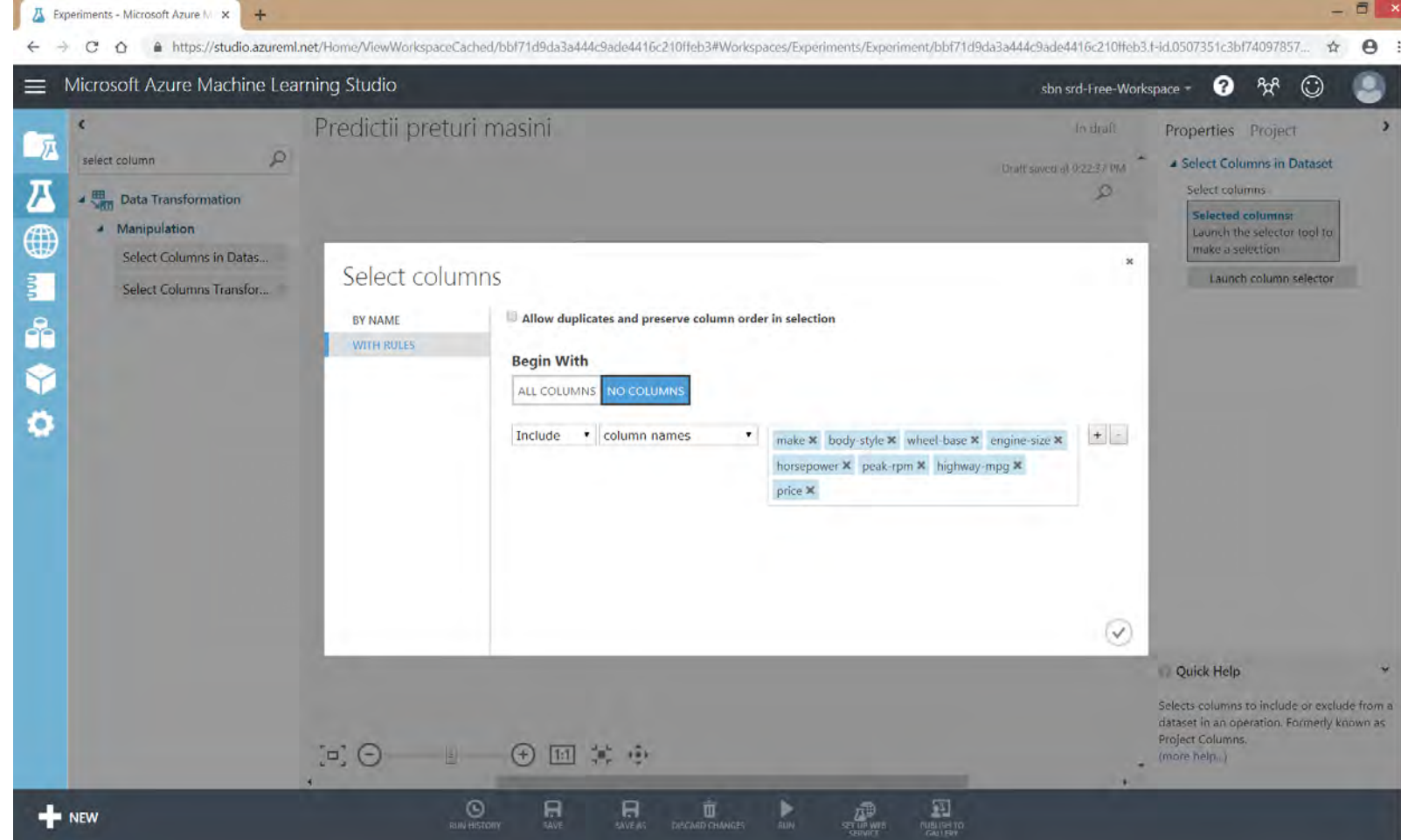
- *Launch column selector*

- *Begin With*

- *No columns*

- *Include*

- *make, body-style, wheel-base, engine-size, horsepower, peak-rpm, highway-mpg, price*



- goal: predict car price from selected features

Car Price Prediction

* choosing / applying the algorithm

- create the training dataset and the test dataset
- training dataset
 - dataset that includes the car price
 - the model is trained on this dataset
 - it searches for correlations between a car's features and its price
- test dataset
 - dataset that includes the car price
 - the model is tested on this dataset
 - the price estimated by the model for each car is compared with the real price

Car Price Prediction

* choosing / applying the algorithm

- create the training / test datasets - *Split Data* module

- *Split Data*

- *Fraction of rows in the first output dataset*
 - 0.75
 - i.e., training dataset - 75% of the data

- *Run experiment*

The screenshot displays the Microsoft Azure Machine Learning Studio interface. The main workspace shows a workflow titled "Predictii preturi masini" (Car Price Prediction). The workflow consists of the following modules in sequence:

- Automobile price data.(Raw)
- Select Columns in Dataset (with a green checkmark)
- Clean Missing Data (with a green checkmark)
- Select Columns in Dataset (with a green checkmark)
- Split Data (labeled with 1 and 2)

A "Mini Map" in the bottom left corner provides a visual overview of the entire workflow. The right-hand side of the interface shows the "Properties" pane for the "Split Data" module, which is currently in "Draft" mode. The properties include:

- Splitting mode: Split Rows
- Fraction of rows in...: 0.75
- Randomized split: ☒
- Random seed: 0
- Stratified split: False

At the bottom of the interface, there is a "Quick Help" section with the text: "Split the rows of a dataset into two distinct sets (more help...)". 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

* choosing / applying the algorithm

- *Machine Learning -> Initialize Model -> Regression -> Linear Regression*

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 the following steps:

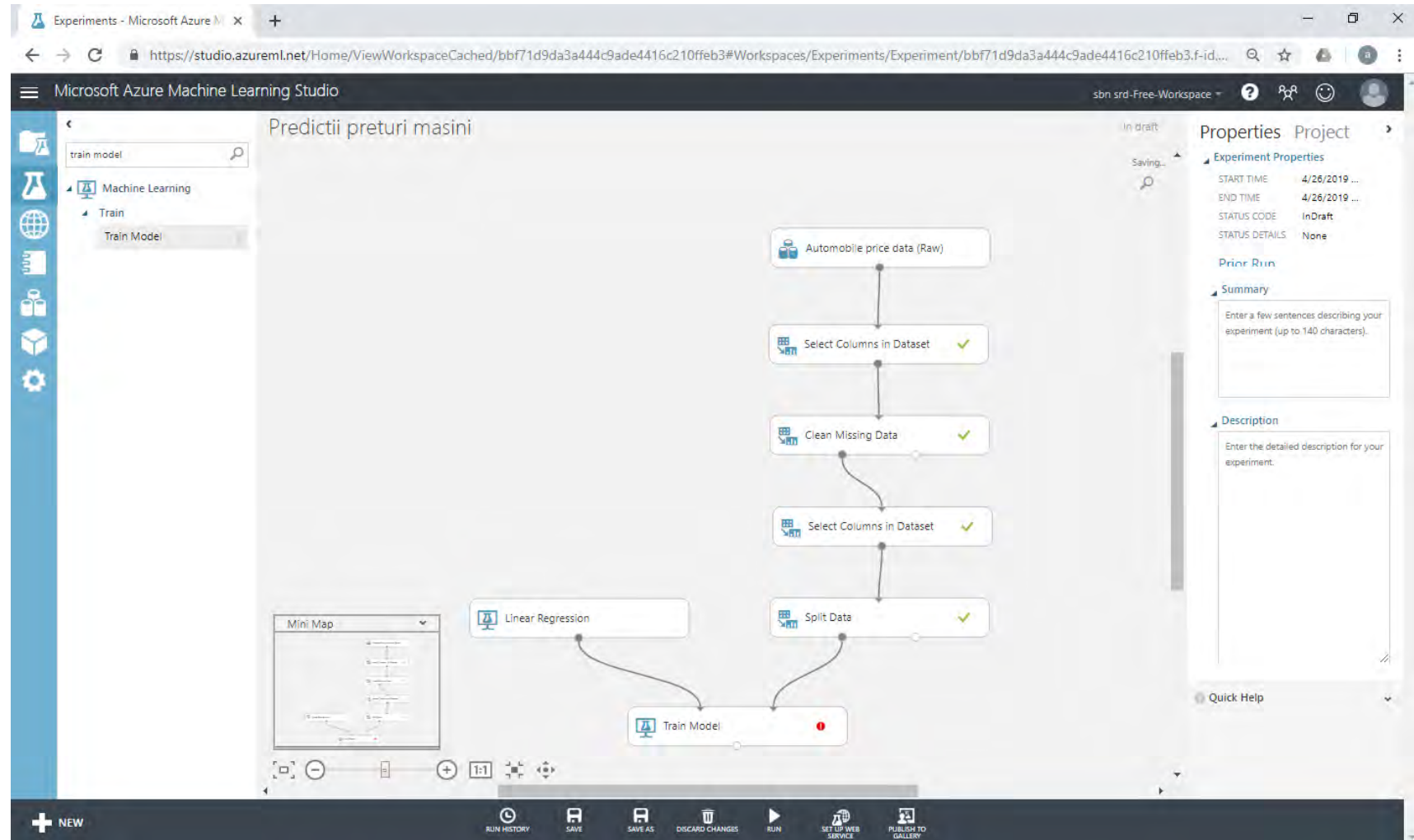
- Automobile price data (Raw)
- Select Columns in Dataset (with a green checkmark)
- Clean Missing Data (with a green checkmark)
- Select Columns in Dataset (with a green checkmark)
- Split Data (with a green checkmark)
- Linear Regression (highlighted with a blue box and a circled '1')

The left sidebar shows the "Machine Learning" category expanded, with "Initialize Model" selected. Under "Initialize Model", the "Regression" sub-category is expanded, and "Linear Regression" is selected. The right sidebar shows the "Properties" panel for the "Linear Regression" model, with the "Solution method" set to "Ordinary Least Squares". Other properties include "L2 regularization w..." set to "0.001", "Include interce..." checked, "Random number s..." set to a default value, and "Allow unknown..." checked. 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

* choosing / applying the algorithm

- *Train Model* module



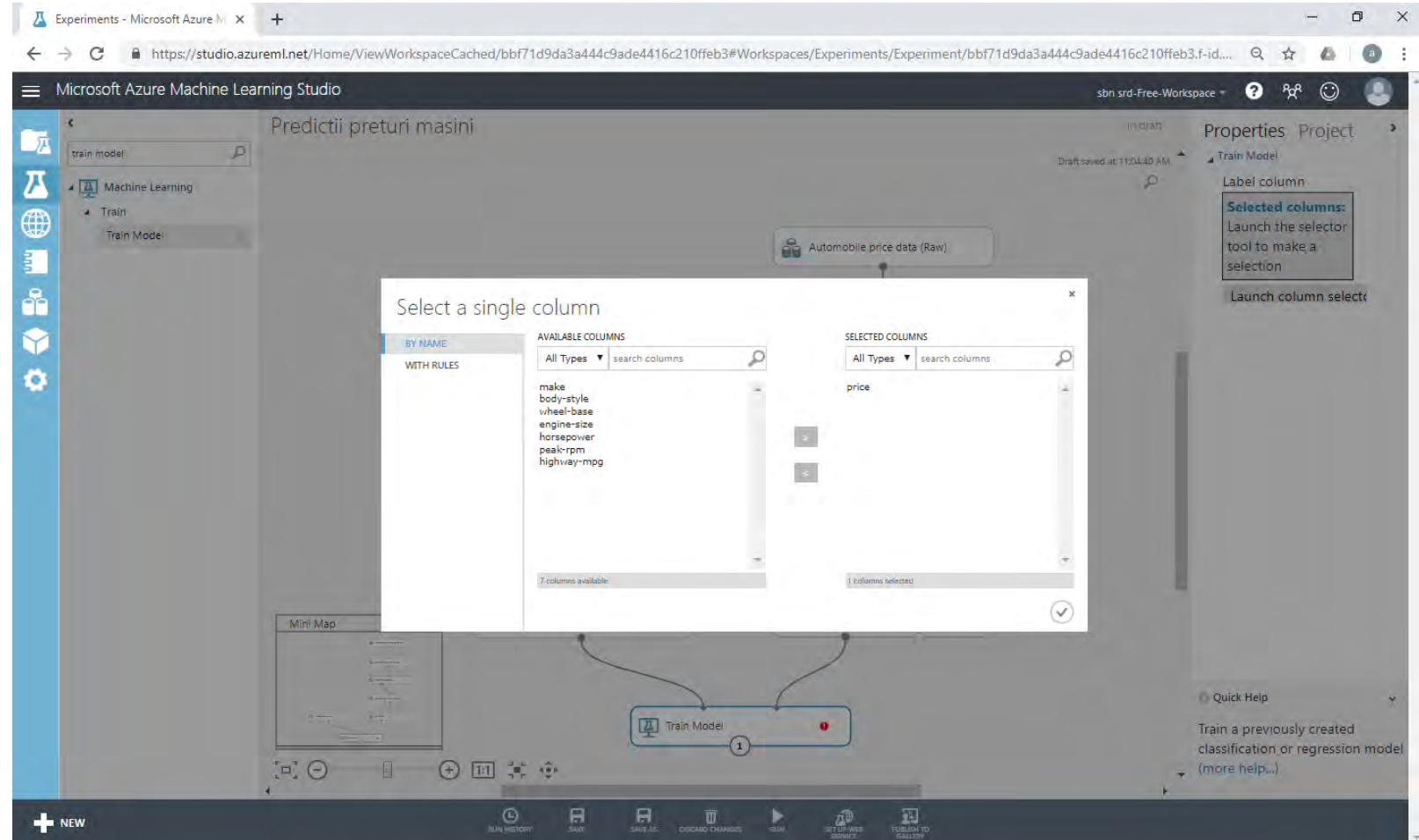
Car Price Prediction

* choosing / applying the algorithm

- *Train Model*

- *Launch column selector*
 - move column *price* from *Available columns* to *Selected columns*

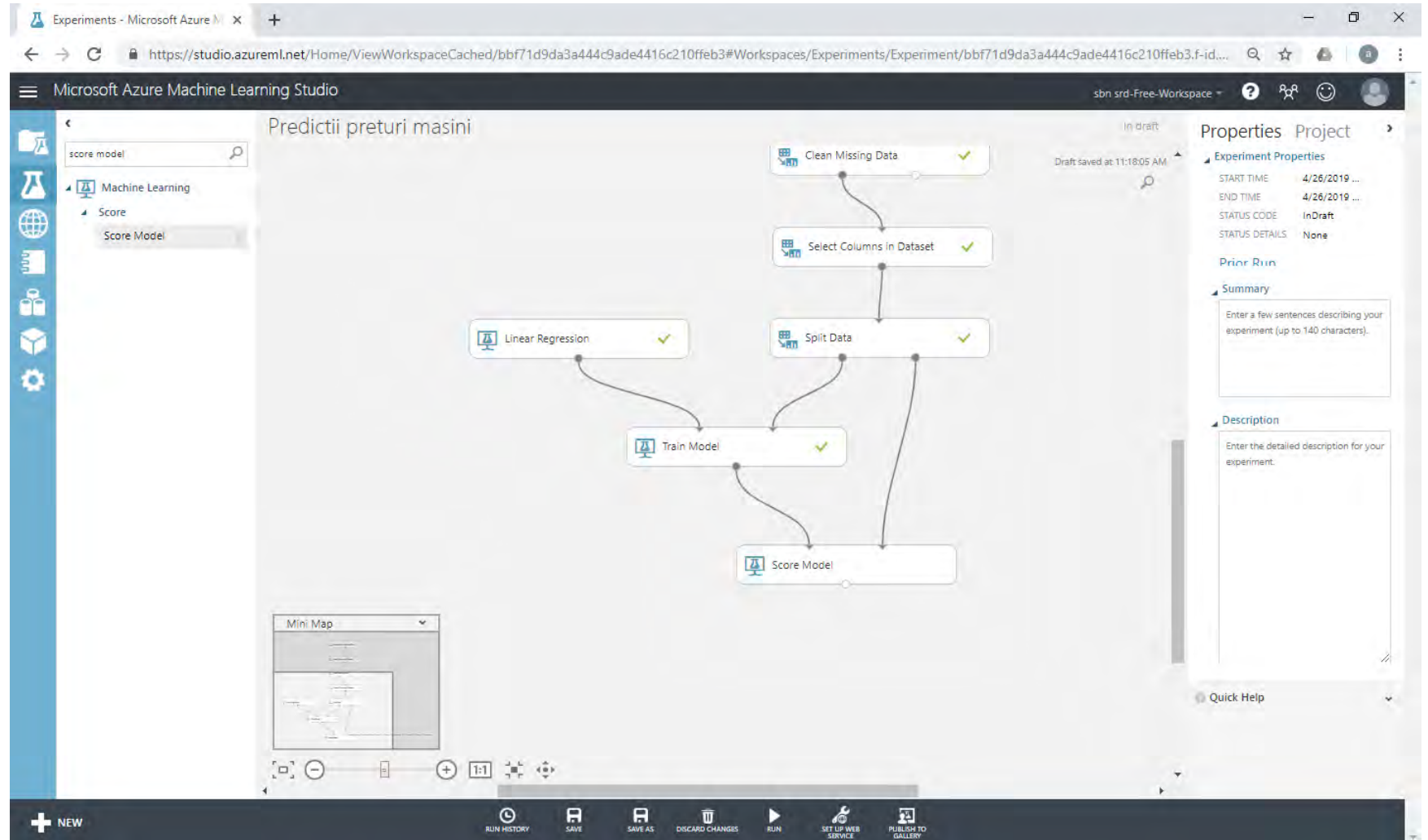
- *Run experiment*



Car Price Prediction

* testing the model - *Score Model* module

• *Run experiment*

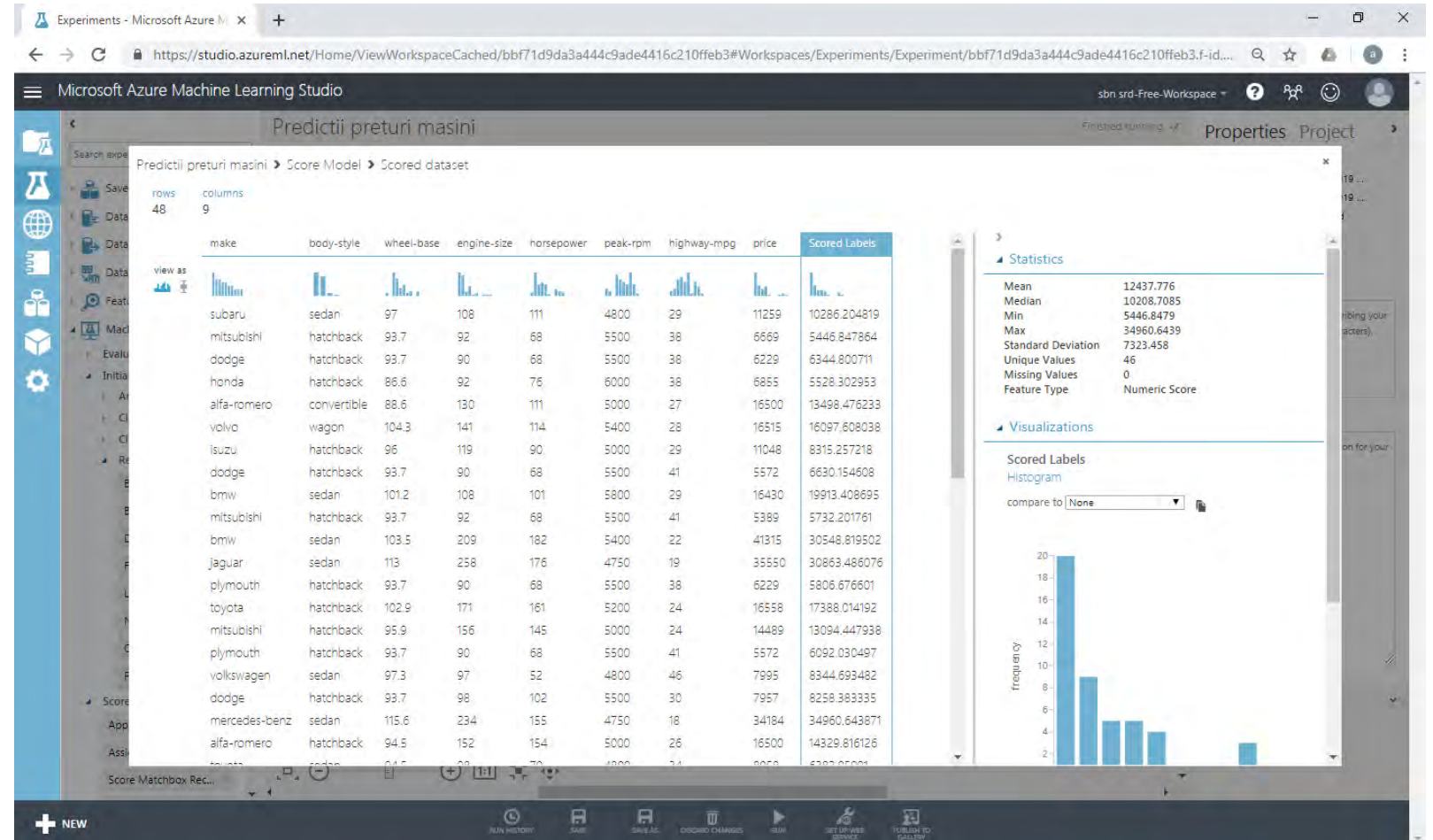


Car Price Prediction

* testing the model

- *Score Model* output port-> *Visualize*

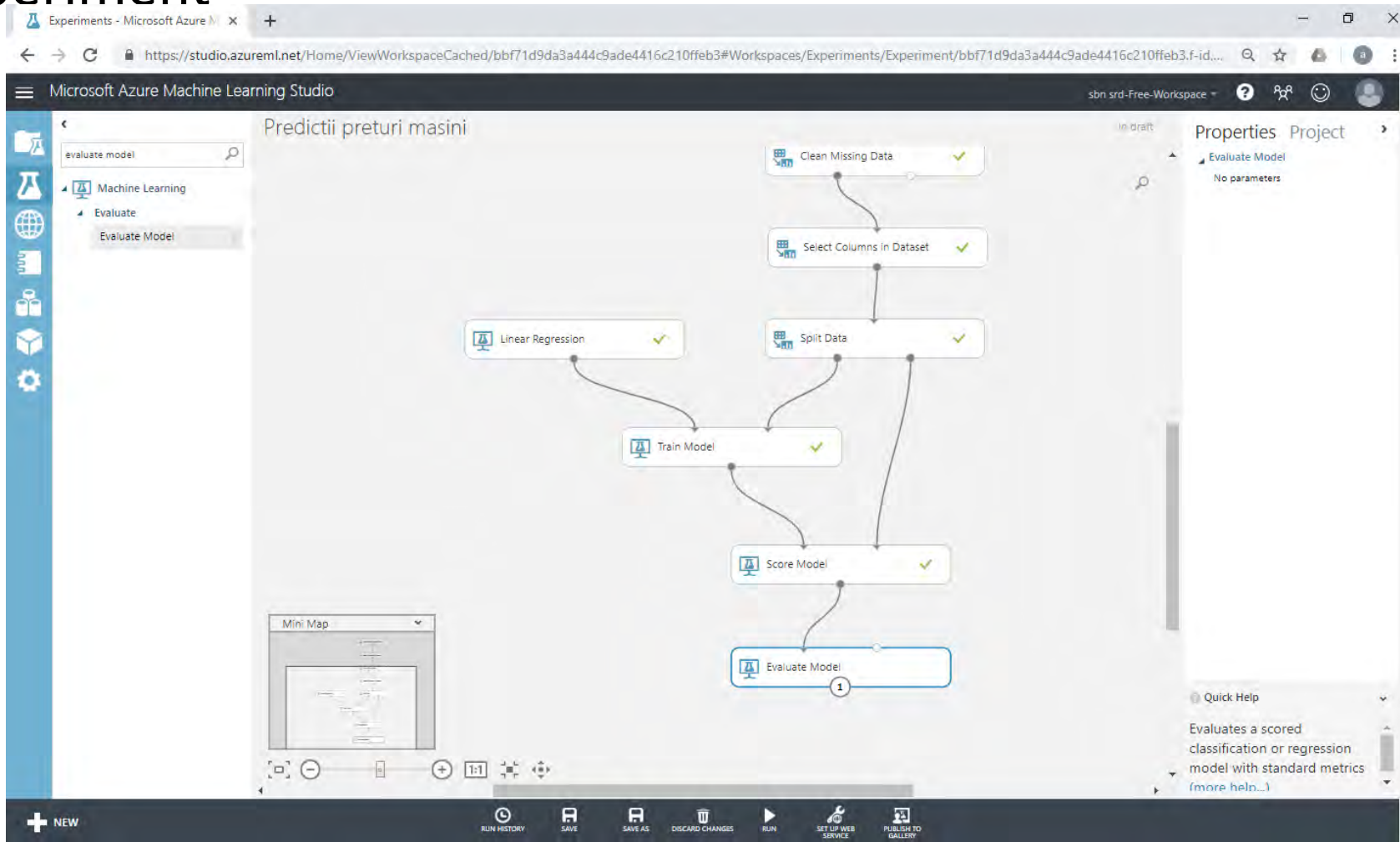
- estimated / actual values for the *price* column



Car Price Prediction

* testing the model

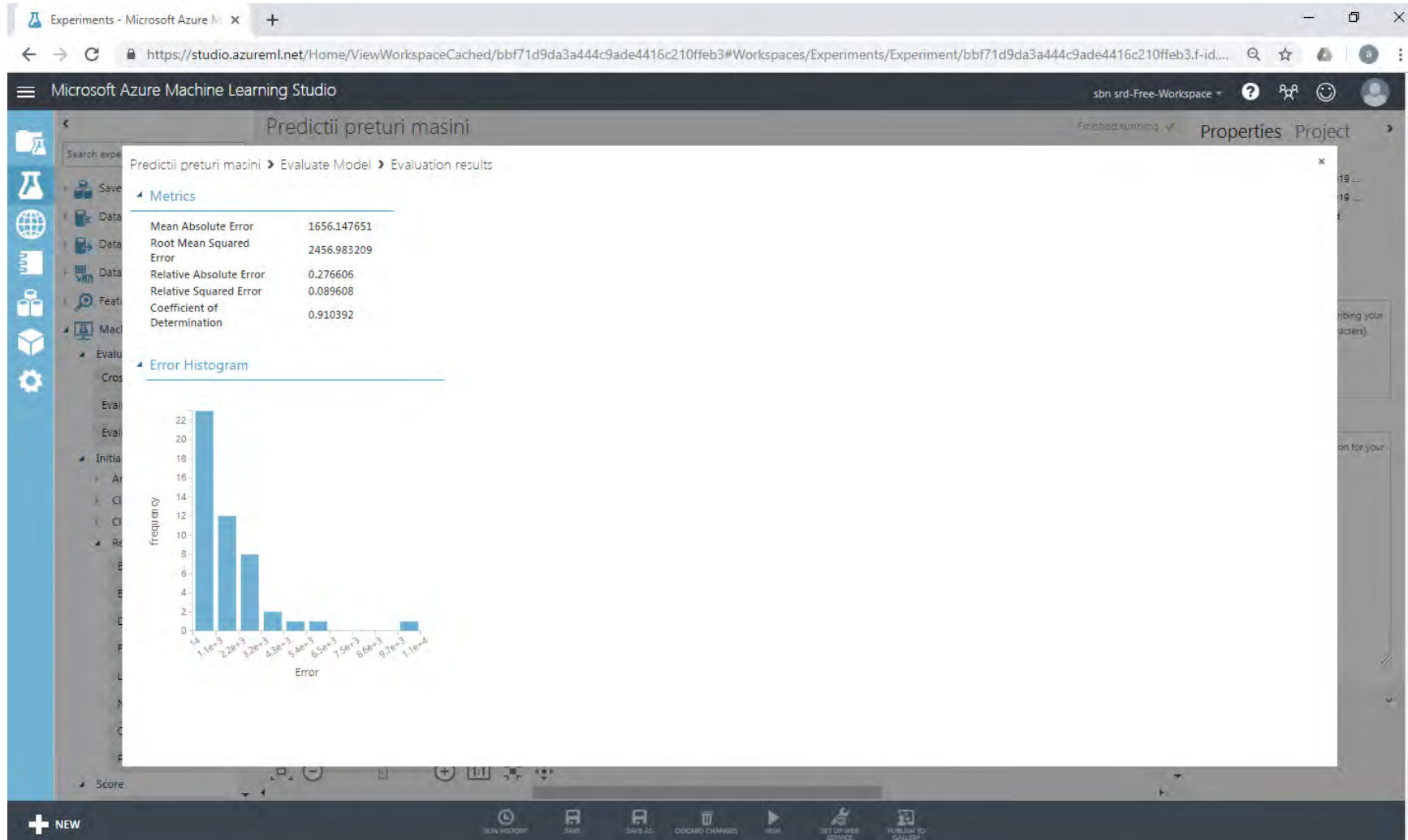
- *Evaluate Model* module
- *Run experiment*



Car Price Prediction

* testing the model

- *Evaluate Model* output port -> *Visualize*



Car Price Prediction

* eliminate resources

The screenshot displays the Microsoft Azure Machine Learning Studio interface. On the left, a sidebar contains navigation links: PROJECTS, EXPERIMENTS, WEB SERVICES, NOTEBOOKS, DATASETS, TRAINED MODELS, and SETTINGS. The main area is titled 'experiments' and shows a table of experiments under the 'MY EXPERIMENTS' tab.

	NAME	AUTHOR	STATUS	LAST EDITED	PROJECT
<input type="checkbox"/>	Predictive Experiment - Mini ...	surdusabina	Finished	4/26/2019 3:13:51 PM	None
<input checked="" type="checkbox"/>	Predictii preturi masini	surdusabina	Finished	4/26/2019 11:28:46 AM	None

On the right side of the interface, a workflow diagram is visible, illustrating the process of building a car price prediction model. The workflow consists of the following steps:

- Automobile price data (Raw)
- Select Columns in Dataset
- Clean Missing Data
- Select Columns in Dataset
- Split Data
- Linear Regression
- Train Model
- Score Model
- Evaluate Model

The workflow is a vertical sequence of steps, with 'Split Data' branching into 'Linear Regression' and 'Train Model'. 'Train Model' then leads to 'Score Model', which finally leads to 'Evaluate Model'.

Car Price Prediction

* eliminate resources

- delete workspace: *Settings -> Delete*

