

Database Management Systems

Lecture 6

Azure Machine Learning*

Azure Stream Analytics*

* not among the exam topics

Azure Machine Learning

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

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

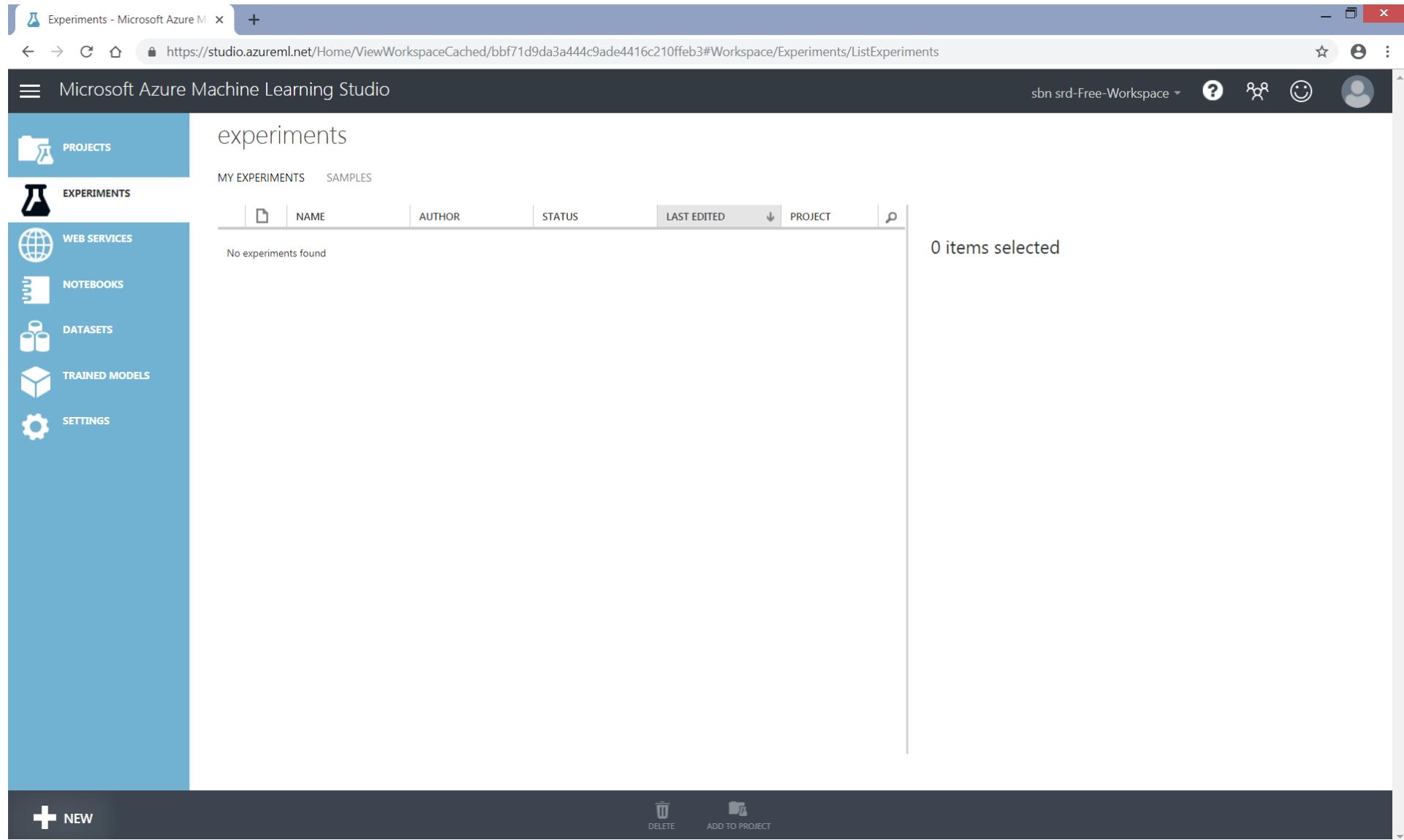
Aired on June 21, 2016

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](#)

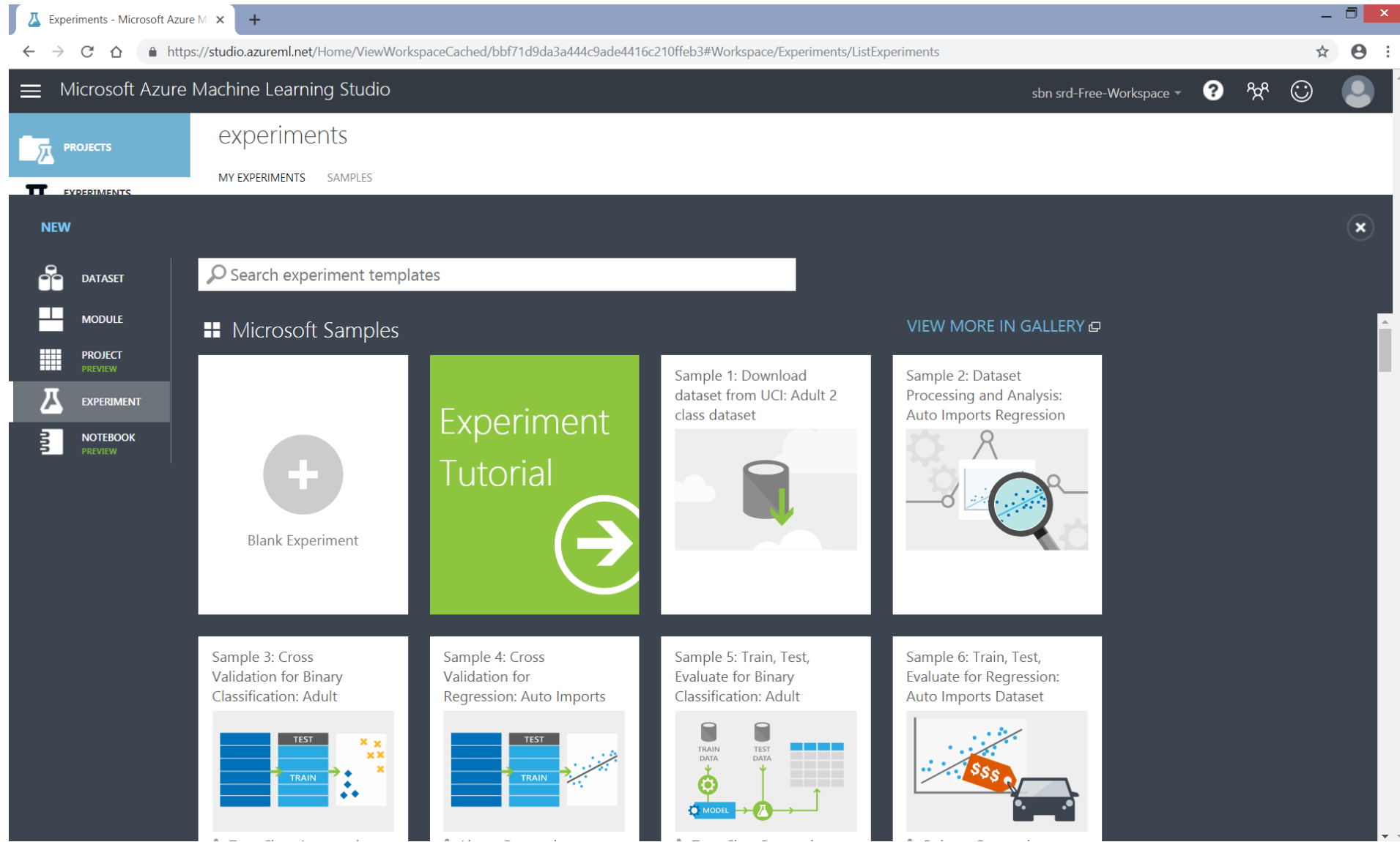
Car Price Prediction

* create an experiment: + New



Car Price Prediction

* create an experiment: Blank Experiment



Car Price Prediction

- * create an experiment
 - experiment name

The screenshot displays the Microsoft Azure Machine Learning Studio interface. The browser address bar shows the URL: <https://studio.azureml.net/Home/ViewWorkspaceCached/bbf71d9da3a444c9ade4416c210feb3#Workspaces/Experiments/Experiment/Draft/ViewExperiment>. The page title is "Microsoft Azure Machine Learning Studio". The left sidebar contains a search bar and a list of experiment items: Saved Datasets, Trained Models, 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 main workspace area is titled "Predictii preturi masini" and contains a diagram with dashed boxes and arrows, indicating a workflow. A text overlay says "To create your experiment, drag and drop datasets and modules here". The right sidebar shows the "Properties" tab with "Experiment Properties" and "Summary" sections. The "Summary" section has a text input field for a description. 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

* selecting the data source

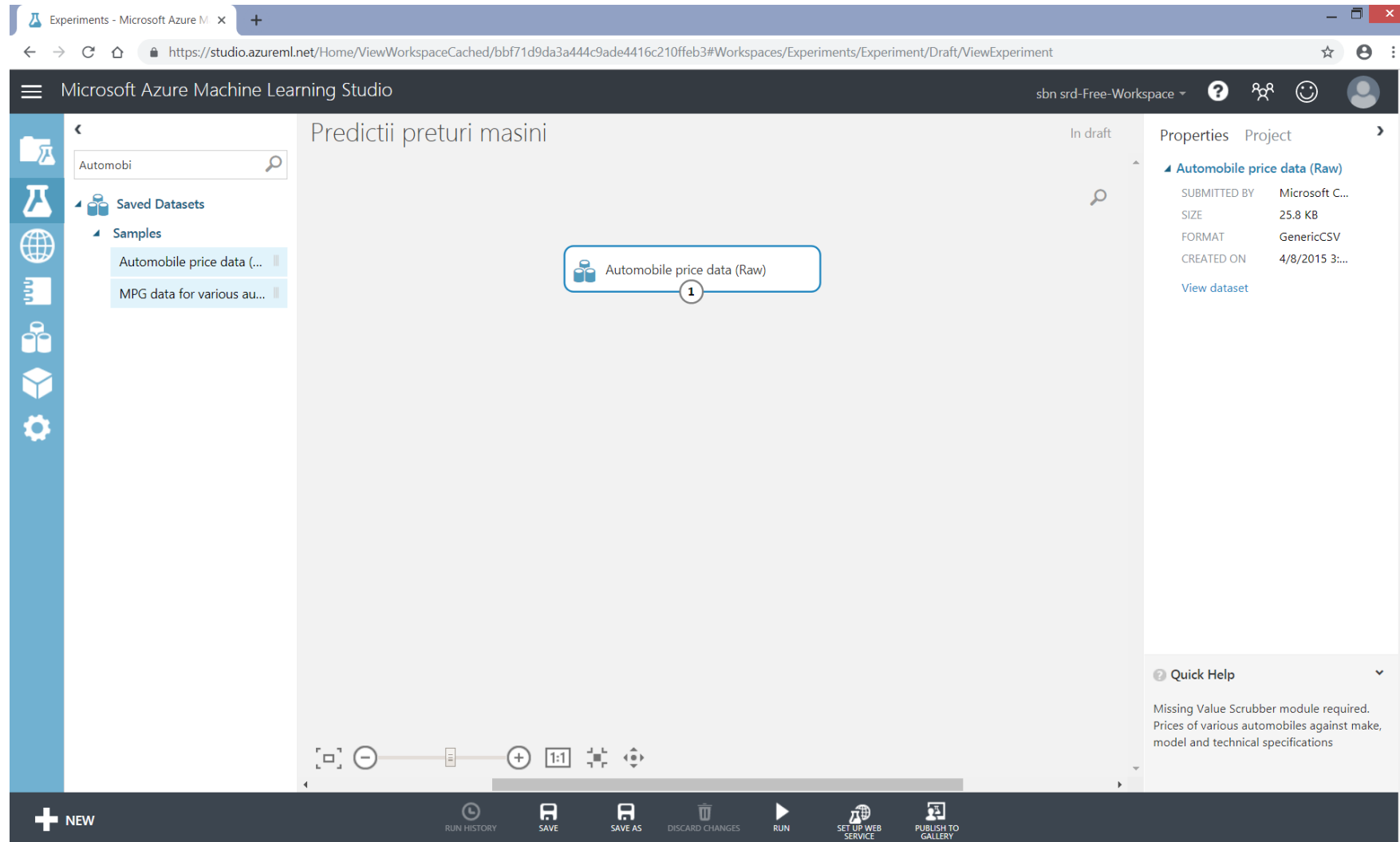
- dataset *Automobile price data (raw)*

The screenshot shows the Microsoft Azure Machine Learning Studio interface. The browser address bar displays the URL: <https://studio.azureml.net/Home/ViewWorkspaceCached/bbf71d9da3a444c9ade4416c210ffeb3#Workspaces/Experiments/Experiment/Draft/ViewExperiment>. The page title is "Microsoft Azure Machine Learning Studio". The left sidebar contains a search bar with "Automob" and a list of "Saved Datasets" under the "Samples" section, including "Automobile price data (...)" and "MPG data for various au...". The main workspace is titled "Predictii preturi masini" and contains a diagram with dashed boxes and arrows, indicating a workflow. A text overlay says "To create your experiment, drag and drop datasets and modules here" and "Drag Items Here" with an arrow pointing to a dashed box. The right sidebar shows the "Properties" tab with "Experiment Properties" (STATUS CODE: InDraft), "Summary" (Enter a few sentences describing your experiment (up to 140 characters)), and "Description" (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

* selecting the data source

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



Car Price Prediction

* displaying the data

- dataset output port -> *Visualize*

The screenshot displays the Microsoft Azure Machine Learning Studio interface. The main workspace is titled "Predictii preturi masini" and shows a draft experiment. A dataset named "Automobile price data (Raw)" is highlighted, and a context menu is open over it, with the "Visualize" option selected. The left sidebar shows the "Automobile price data (Raw)" dataset under the "Samples" section. The right sidebar shows the "Properties" tab for the dataset, listing details such as "SUBMITTED BY: Microsoft C...", "SIZE: 25.8 KB", "FORMAT: GenericCSV", and "CREATED ON: 4/8/2015 3:...". The bottom status bar includes icons for "NEW", "RUN HISTORY", "SAVE", "SAVE AS", "DISCARD CHANGES", "RUN", "SET UP WEB SERVICE", and "PUBLISH TO GALLERY".

Microsoft Azure Machine Learning Studio

Predictii preturi masini

In draft

Automobile price data (Raw)

Visualize

Download

Generate Data Access Code...

Open in a new Notebook

Properties Project

Automobile price data (Raw)

SUBMITTED BY: Microsoft C...

SIZE: 25.8 KB

FORMAT: GenericCSV

CREATED ON: 4/8/2015 3:...

View dataset

Quick Help

Missing Value Scrubber module required.
Prices of various automobiles against make,
model and technical specifications

Car Price Prediction

* displaying the data

- row – data about a car

The screenshot displays the Microsoft Azure Machine Learning Studio interface. The main window shows a dataset named "Predictii preturi masini" (Automobile price data (Raw)) with 205 rows and 26 columns. The dataset is displayed in a table view. The table has columns: symboling, normalized-losses, make, fuel-type, aspiration, num-of-doors, body-style, drive-wheels, and engine-location. The data is sorted by symboling in descending order. The first few rows of data are:

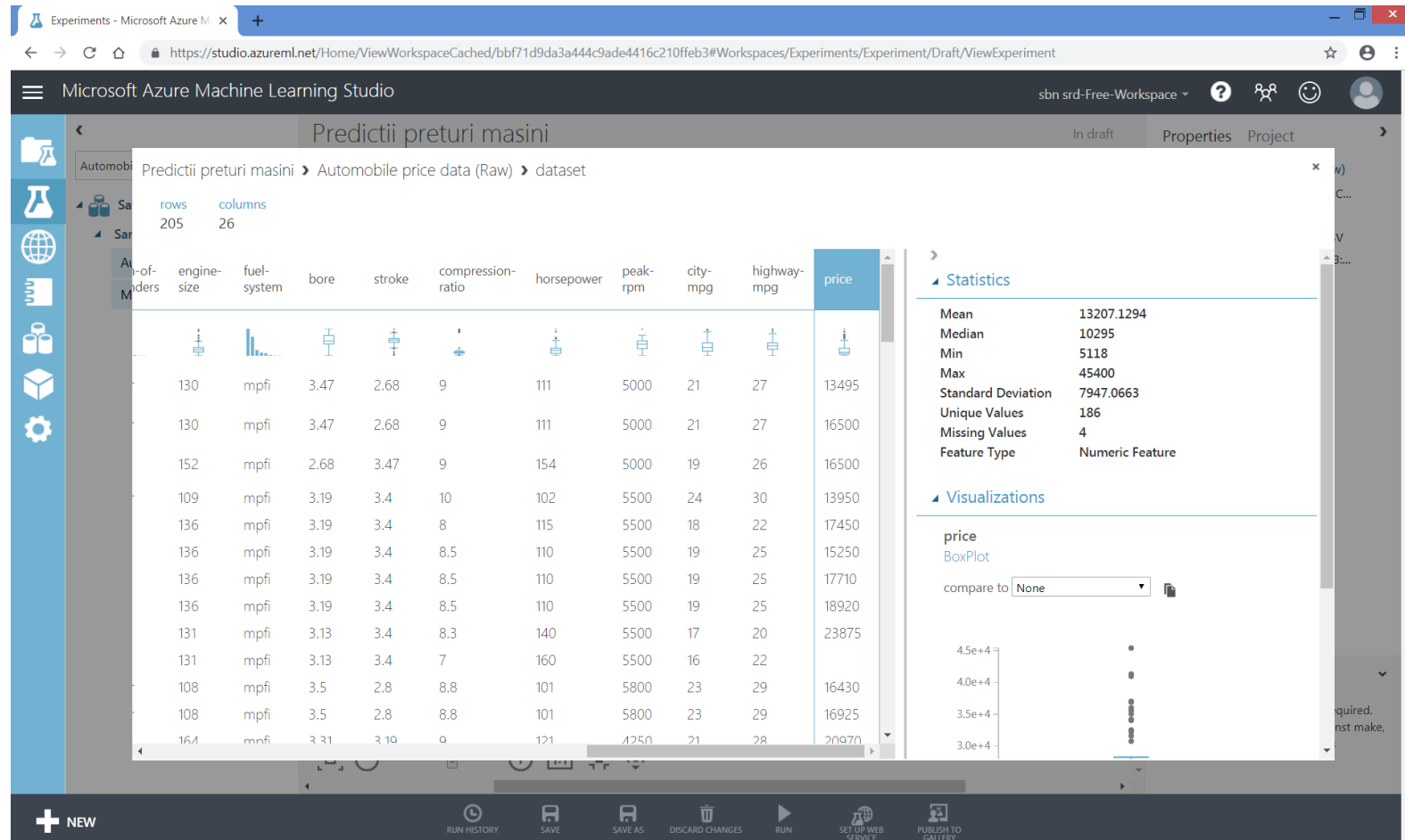
symboling	normalized-losses	make	fuel-type	aspiration	num-of-doors	body-style	drive-wheels	engine-location
3		alfa-romero	gas	std	two	convertible	rwd	front
3		alfa-romero	gas	std	two	convertible	rwd	front
1		alfa-romero	gas	std	two	hatchback	rwd	front
2	164	audi	gas	std	four	sedan	fwd	front
2	164	audi	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
2	192	bmw	gas	std	two	sedan	rwd	front
0	192	bmw	gas	std	four	sedan	rwd	front
0	188	bmw	gas	std	two	sedan	rwd	front

The interface also includes a sidebar with navigation icons, a top bar with the workspace name "sbn srd-Free-Workspace", and a bottom bar with action buttons like "NEW", "RUN HISTORY", "SAVE", "SAVE AS", "DISCARD CHANGES", "RUN", "SET UP WEB SERVICE", and "PUBLISH TO GALLERY".

Car Price Prediction

* displaying the data

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



Car Price Prediction

* preparing the data

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

