Database Management Systems

Lecture 12

Azure Machine Learning*

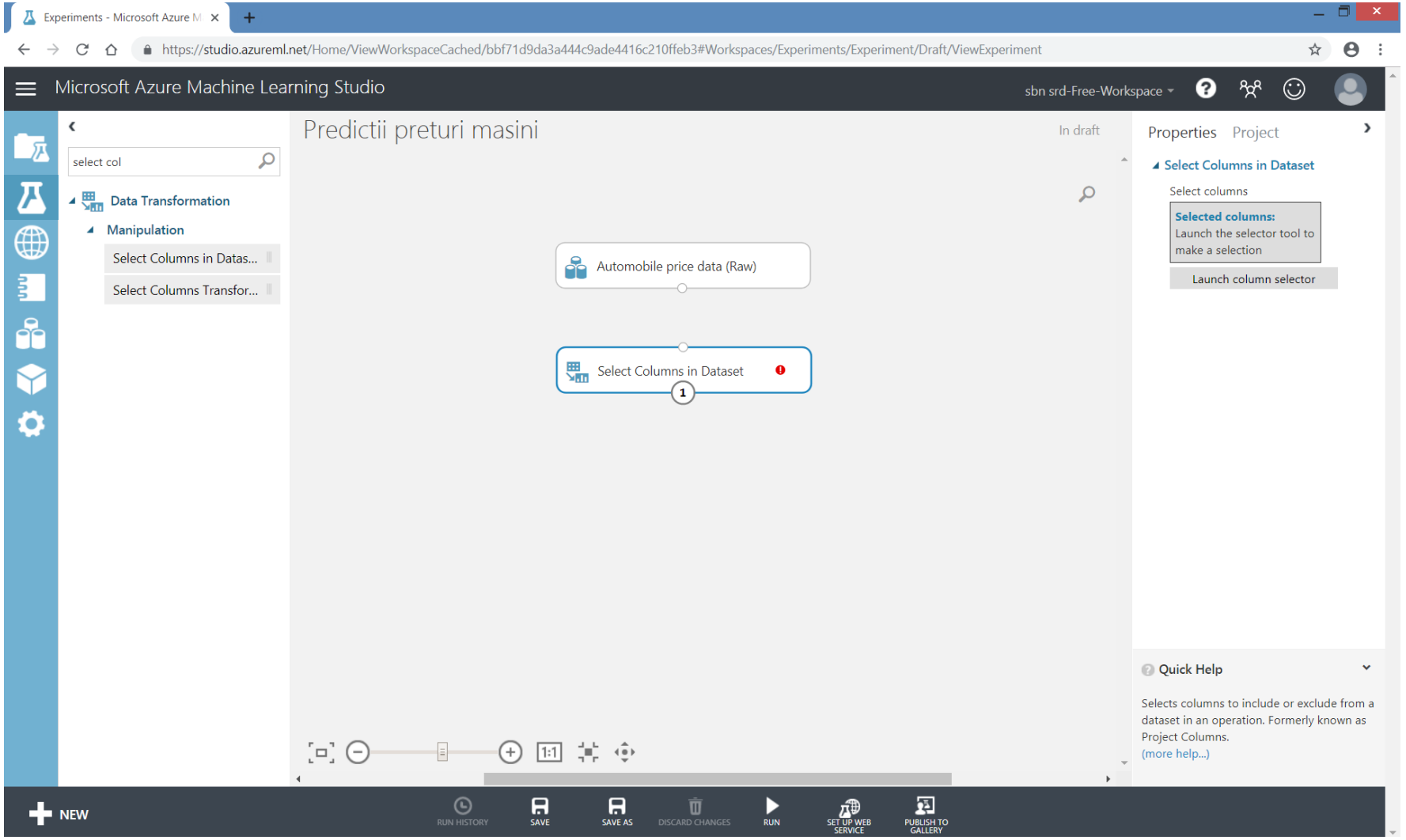
Azure Stream Analytics*

* not among the exam topics

Azure Machine Learning

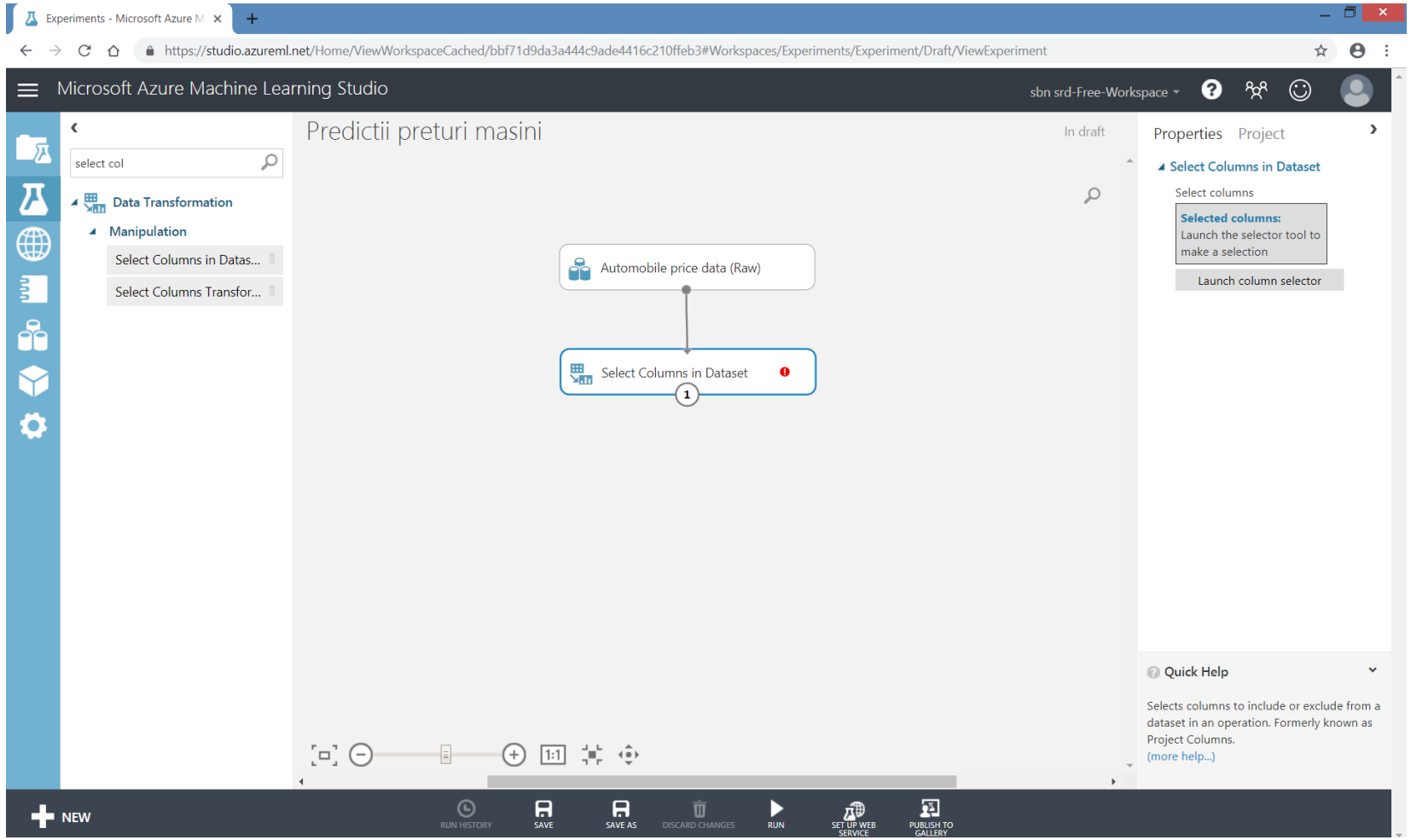
Car Price Prediction

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



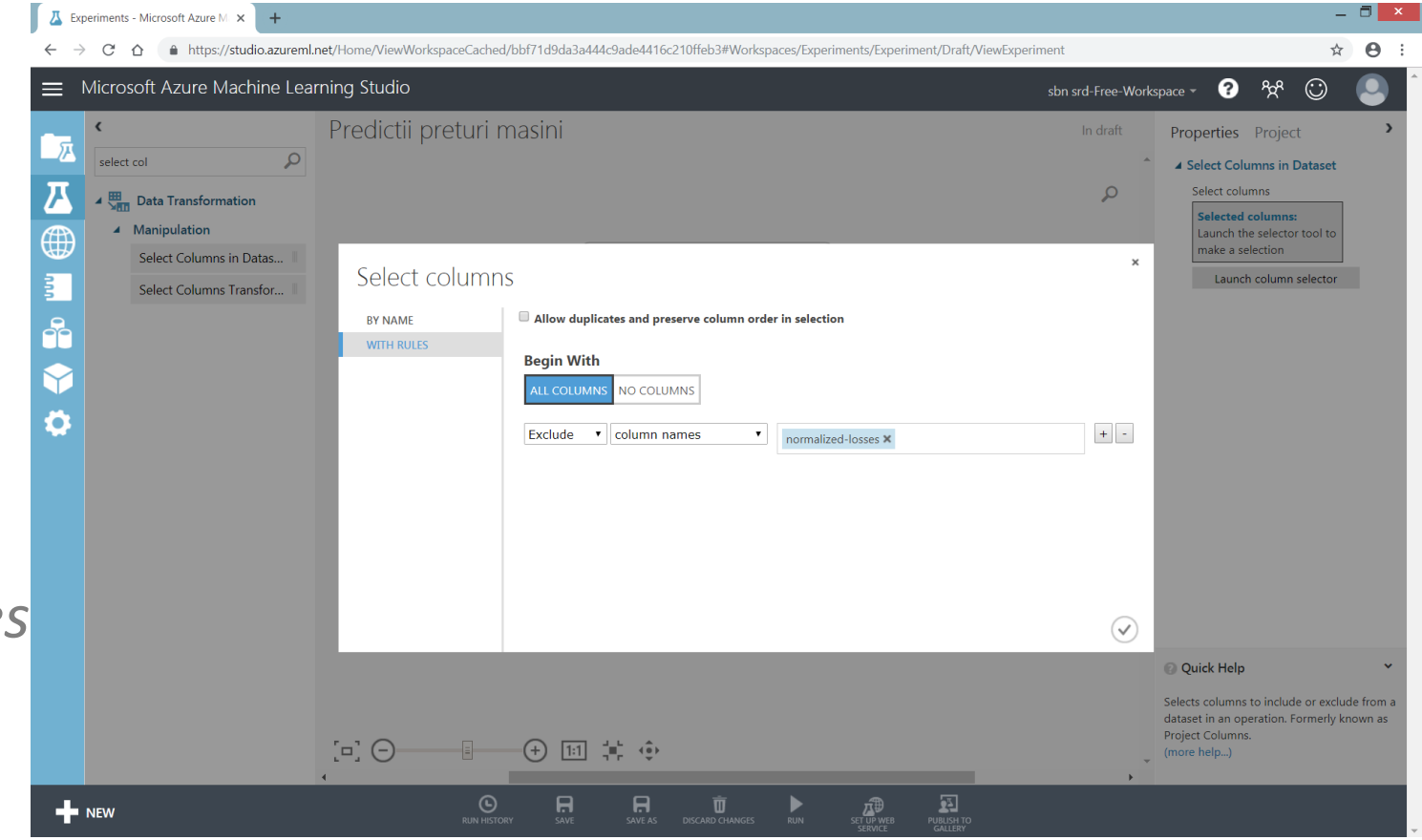
Car Price Prediction

* preparing the data
  - eliminate column with missing values



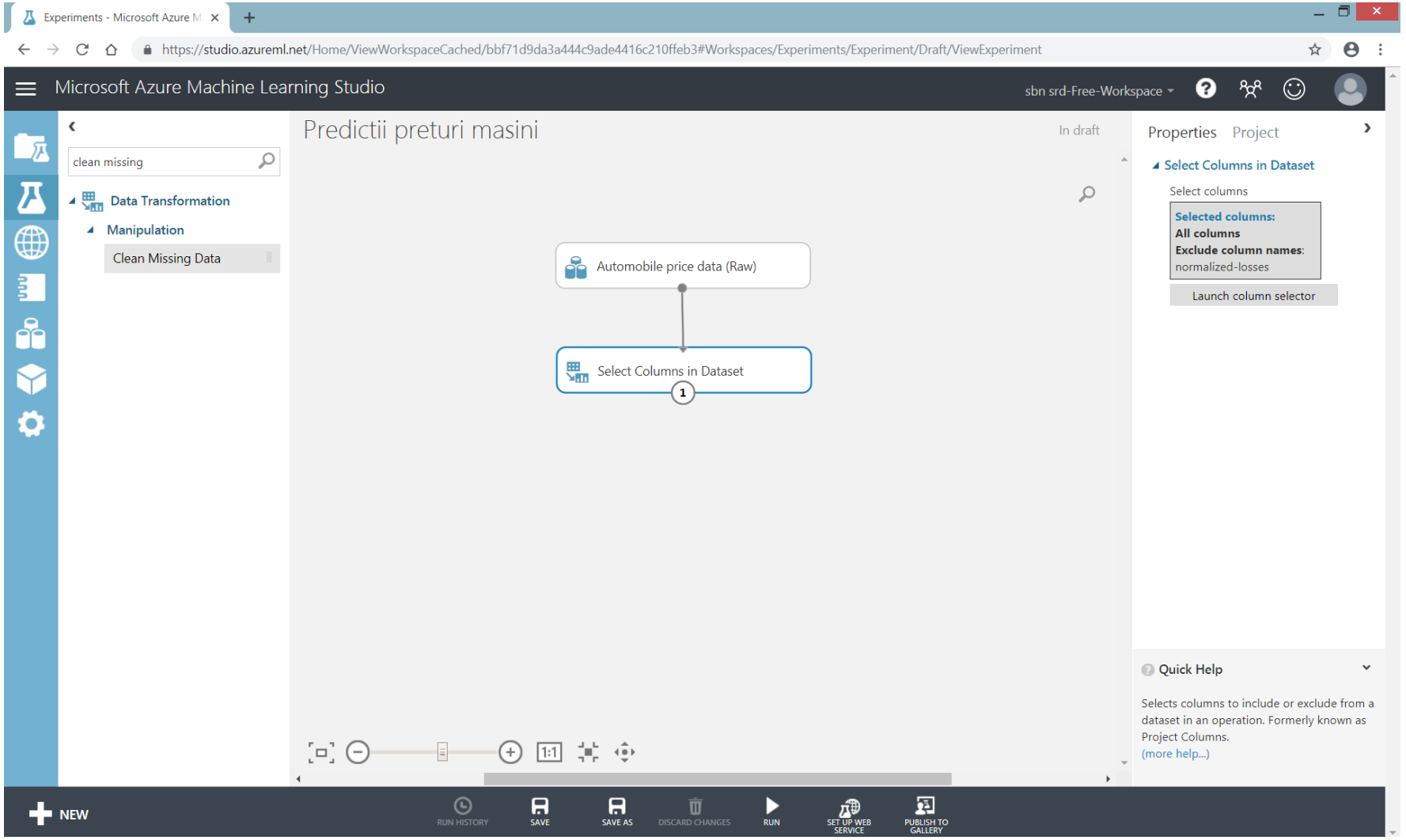
Car Price Prediction

* preparing the data
  - eliminate column with missing values
  - *Select Columns in Dataset*
    - *Launch column selector*
      - *With Rules*
        - *Begin With*
          - *All Columns*
        - *Exclude*
          - *normalized-losses*



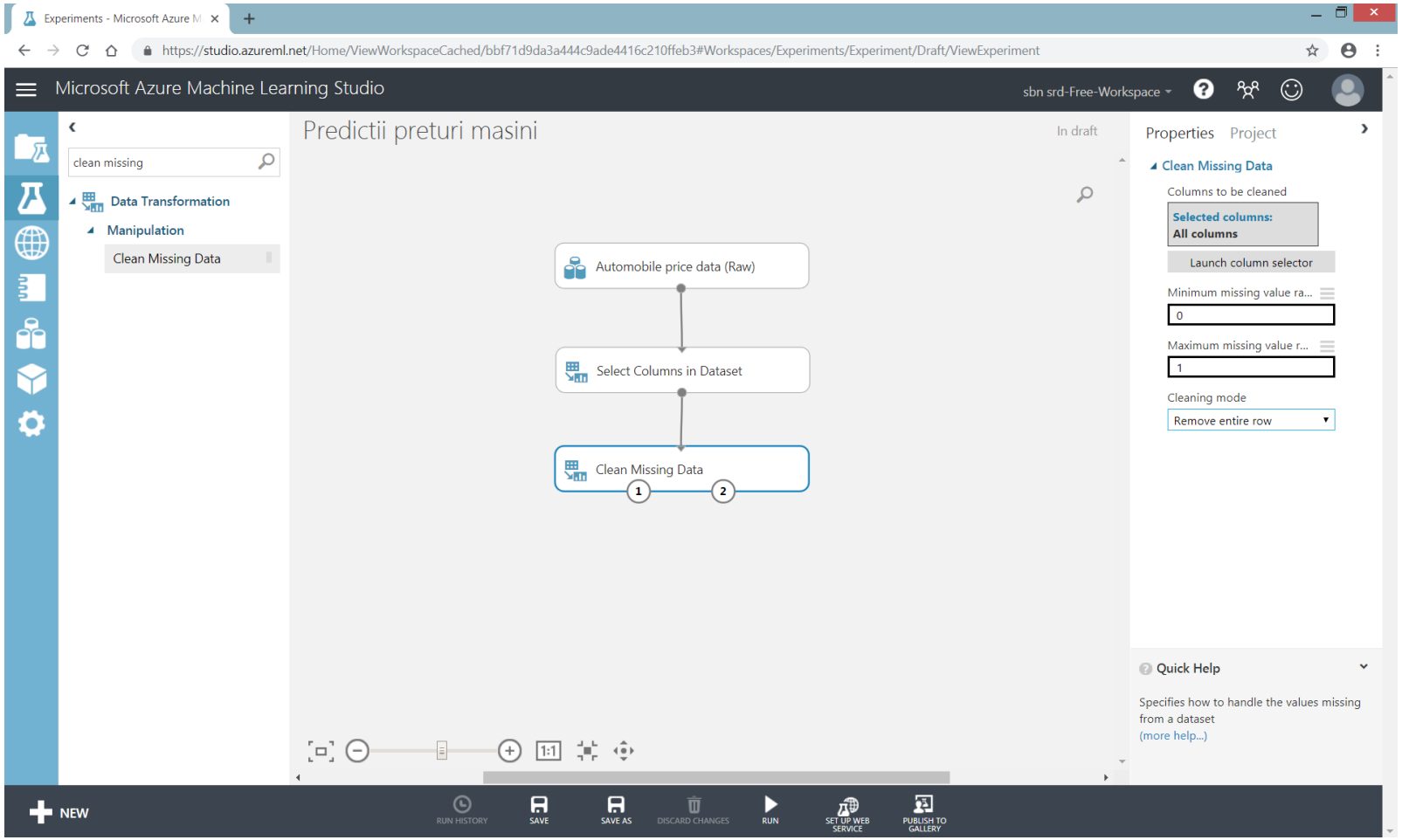
Car Price Prediction

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



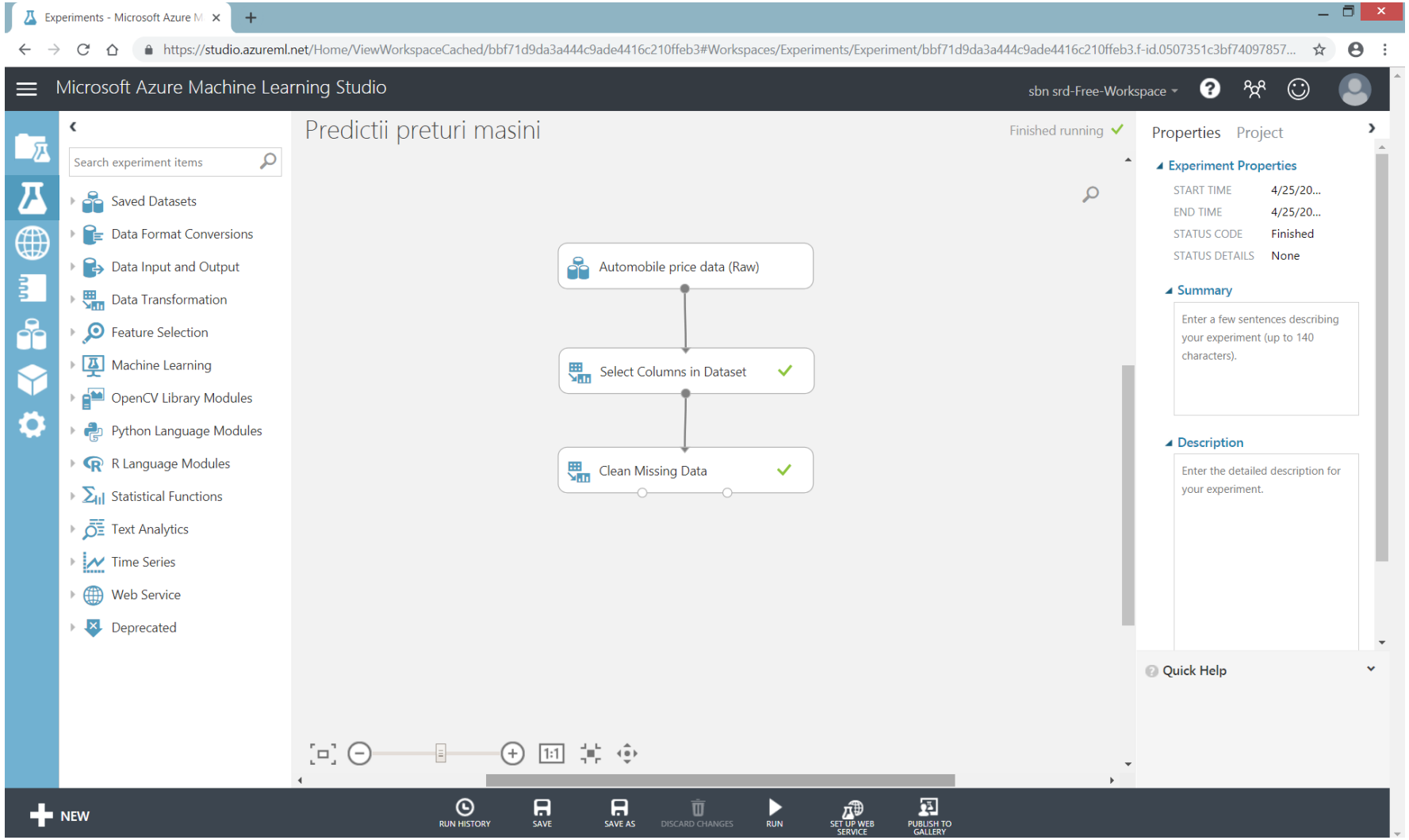
Car Price Prediction

* preparing the data
  - eliminate rows with missing values
  - *Clean Missing Data*
    - *Cleaning mode*
      - *Remove entire row*



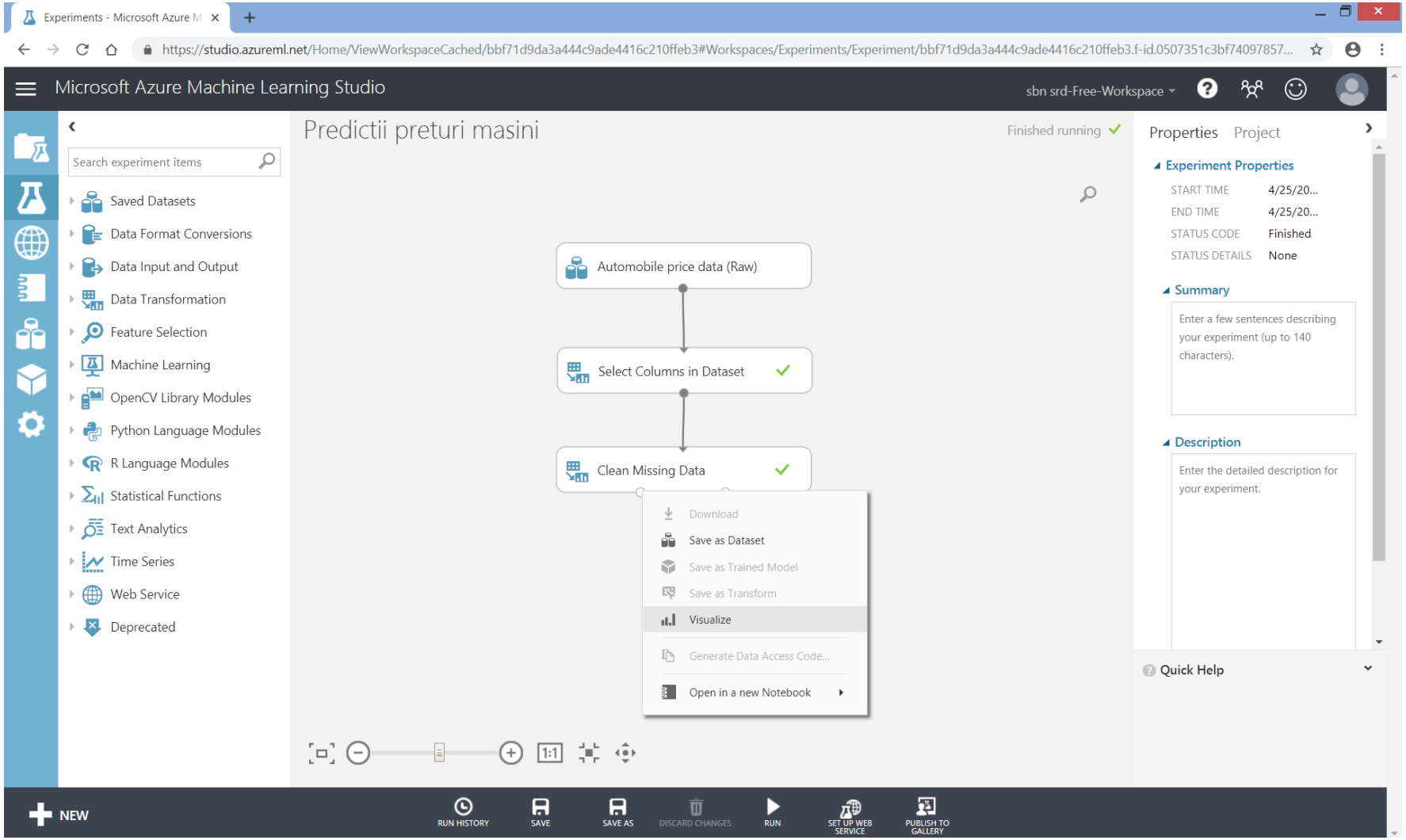
Car Price Prediction

* running the experiment
  - *Run*



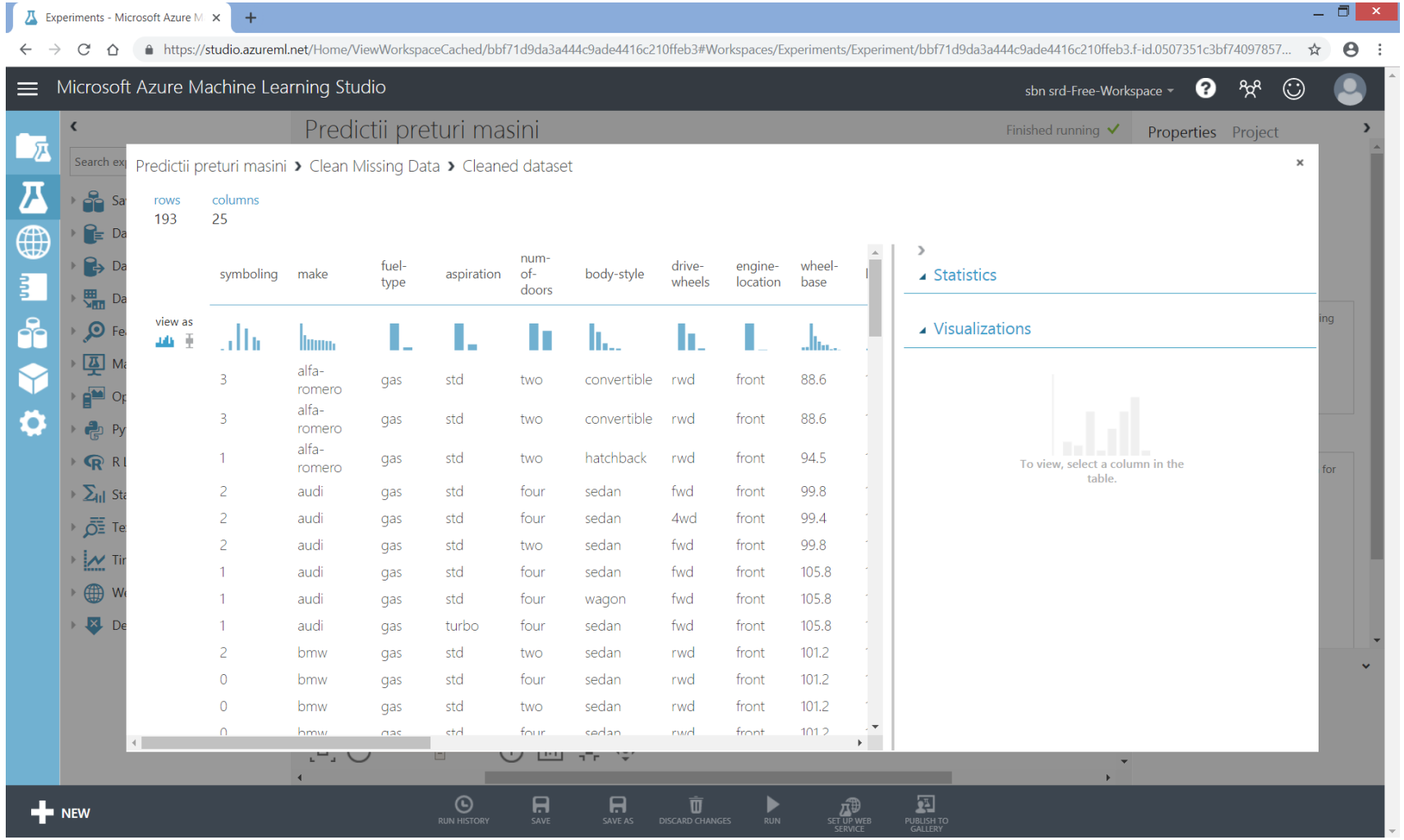
Car Price Prediction

* displaying the data
  - *Clean Missing Data* module -> left output port -> *Visualize*



Car Price Prediction

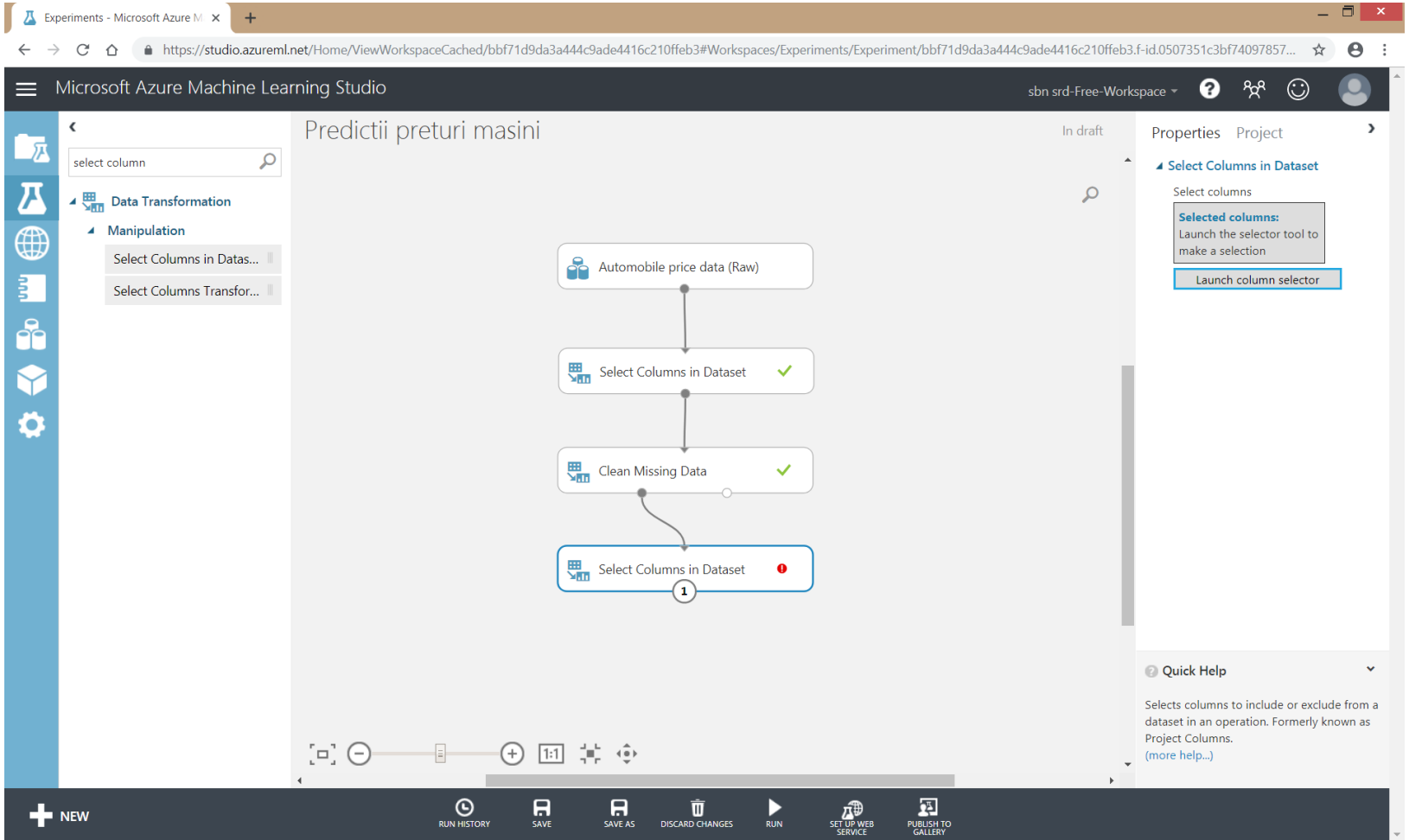
* displaying the data
  - *Clean Missing Data* module -> left output port -> *Visualize*



symboling	make	fuel-type	aspiration	num-of-doors	body-style	drive-wheels	engine-location	wheel-base
3	alfa-romero	gas	std	two	convertible	rwd	front	88.6
3	alfa-romero	gas	std	two	convertible	rwd	front	88.6
1	alfa-romero	gas	std	two	hatchback	rwd	front	94.5
2	audi	gas	std	four	sedan	fwd	front	99.8
2	audi	gas	std	four	sedan	4wd	front	99.4
2	audi	gas	std	two	sedan	fwd	front	99.8
1	audi	gas	std	four	sedan	fwd	front	105.8
1	audi	gas	std	four	wagon	fwd	front	105.8
1	audi	gas	turbo	four	sedan	fwd	front	105.8
2	bmw	gas	std	two	sedan	rwd	front	101.2
0	bmw	gas	std	four	sedan	rwd	front	101.2
0	bmw	gas	std	two	sedan	rwd	front	101.2

Car Price Prediction

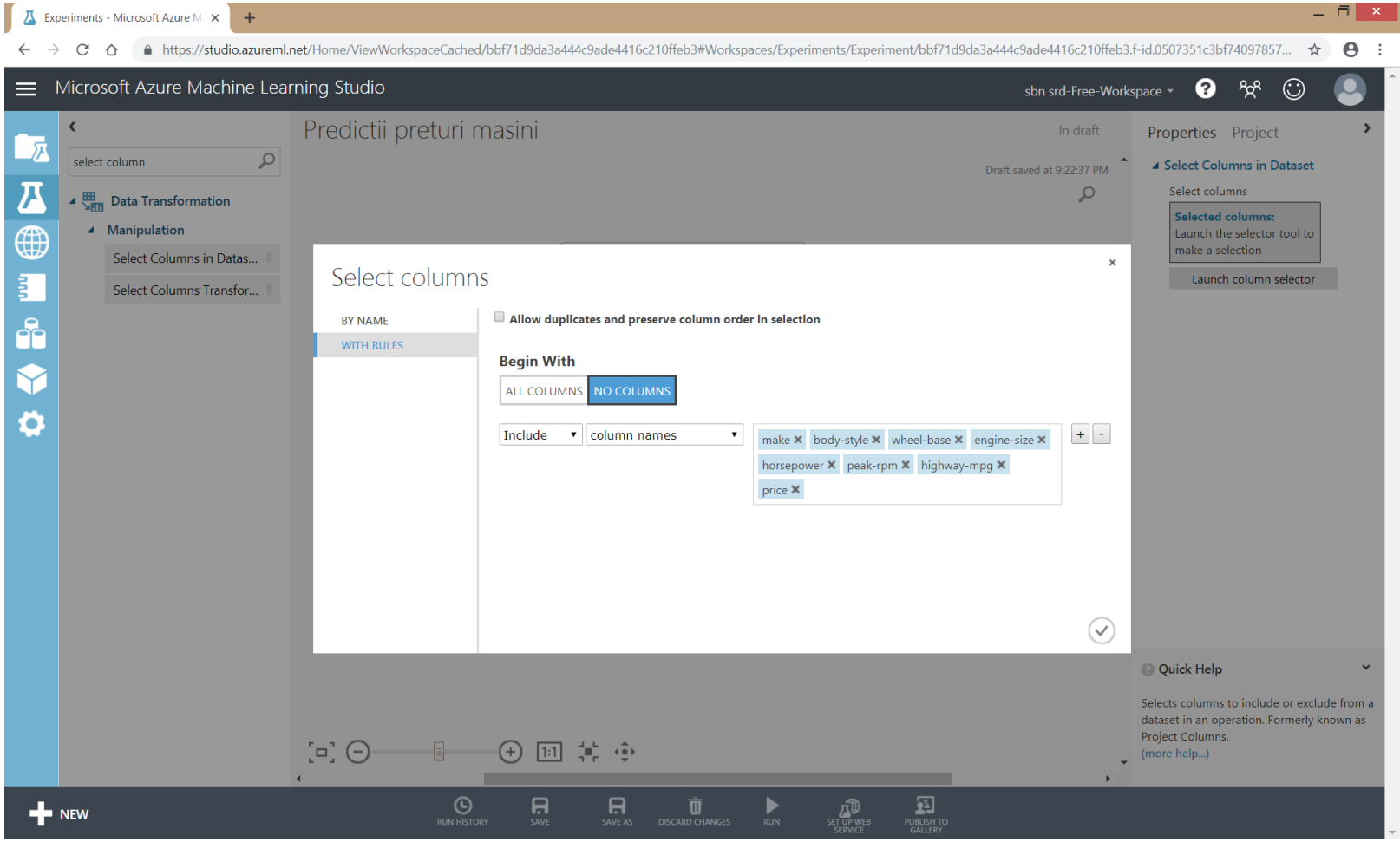
* defining the *features*
  - used to create the predictive model
  - *Select Columns in Dataset* module



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

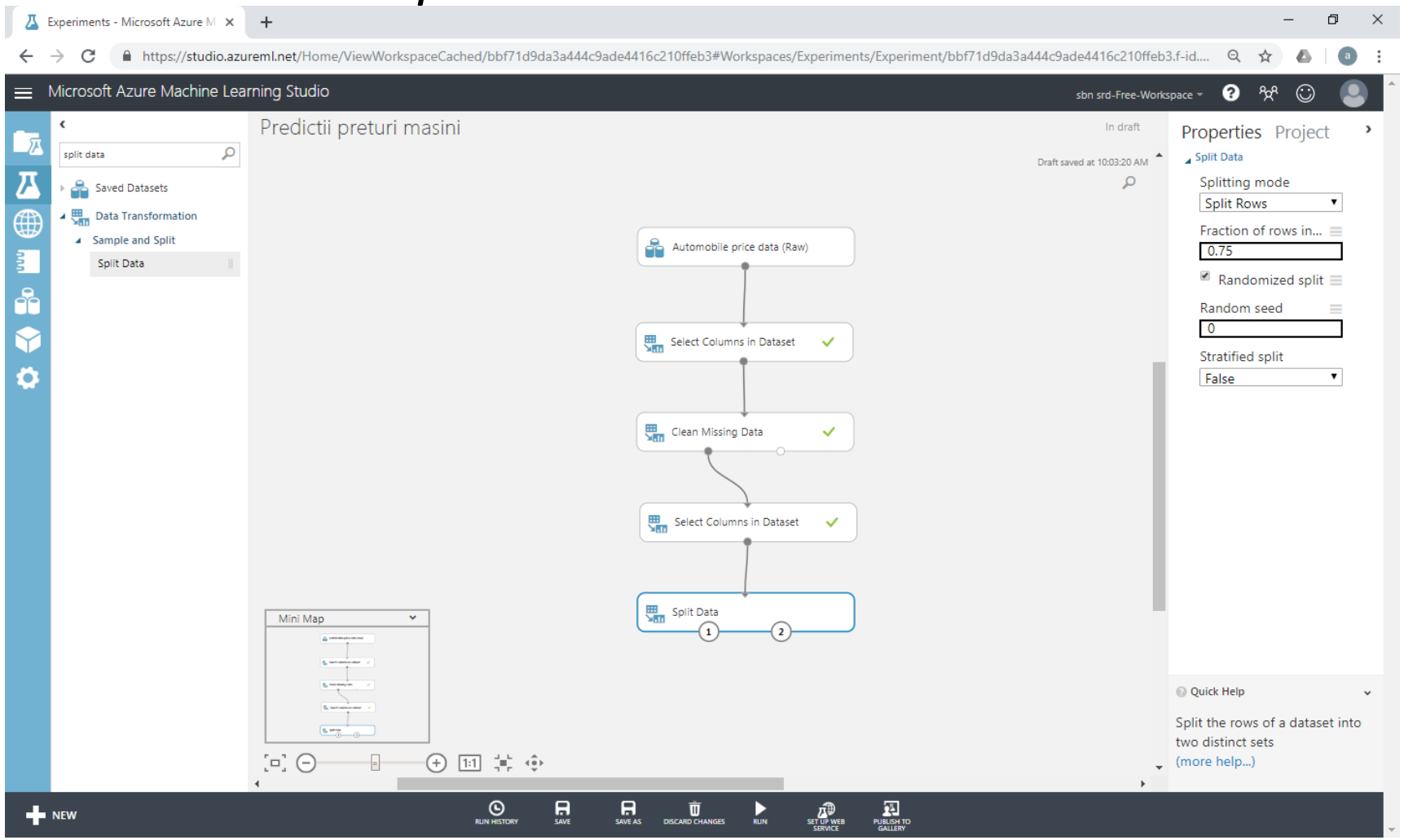
* choosing / applying the algorithm
  - create the training dataset and the test dataset
  - training dataset
    - dataset that includes the car price
    - the model is trained on this dataset
      - it searches for correlations between a car's features and its price
  - test dataset
    - dataset 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

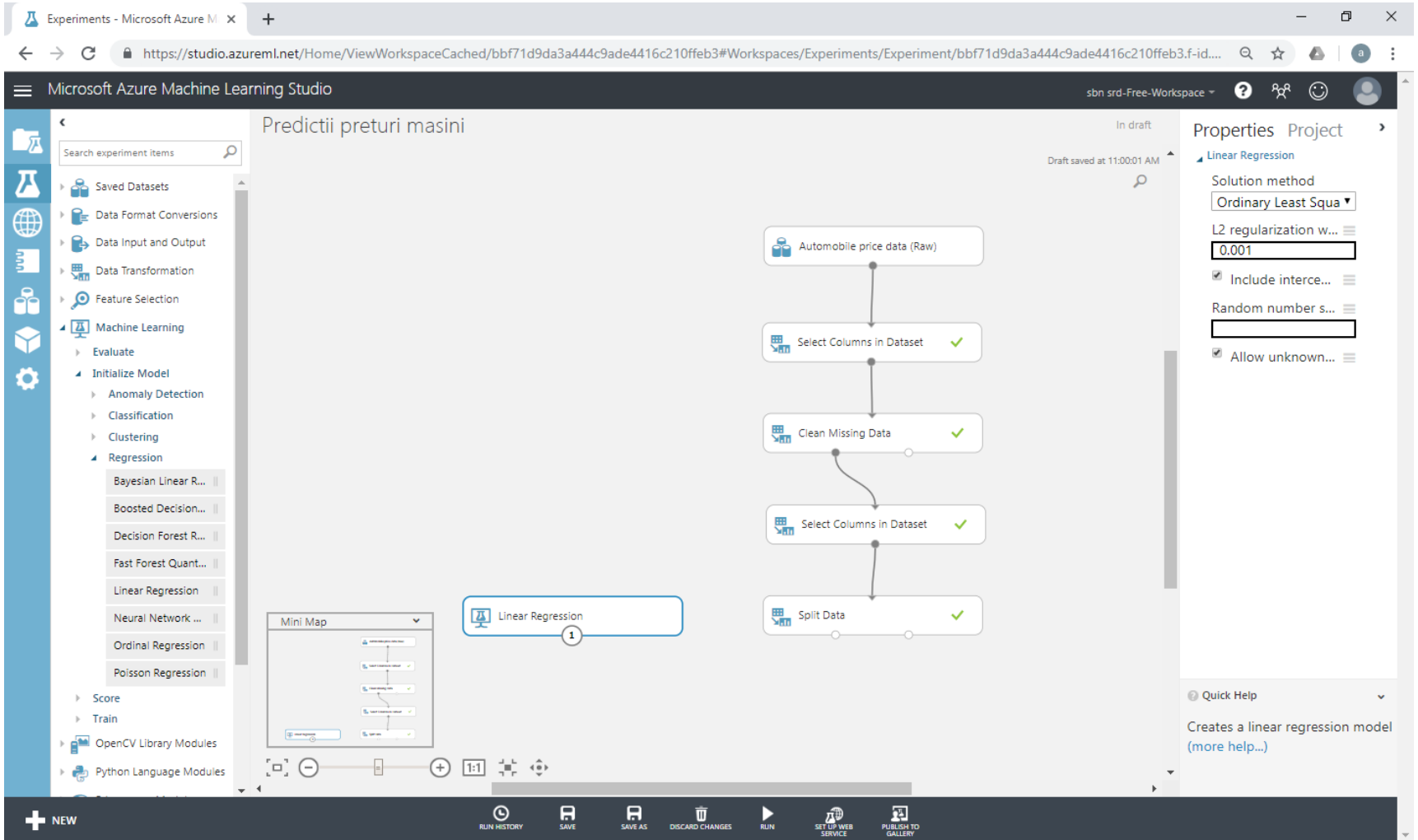
- *Split Data*
  - *Fraction of rows in the first output dataset*
    - *0.75*
    - i.e., training dataset - 75% of the data

- *Run* experiment



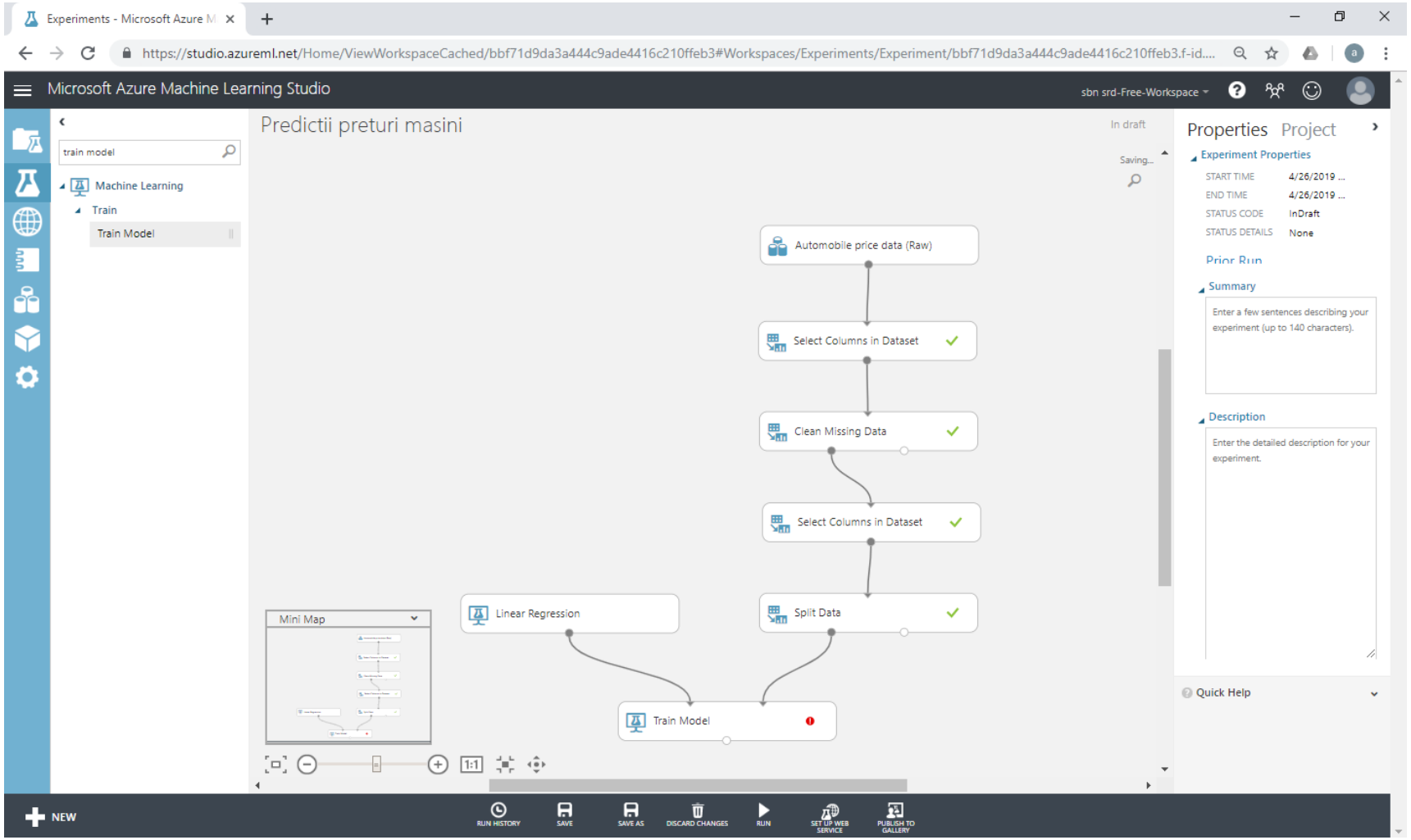
Car Price Prediction

* choosing / applying the algorithm
    - *Machine Learning -> Initialize Model -> Regression -> Linear Regression*



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