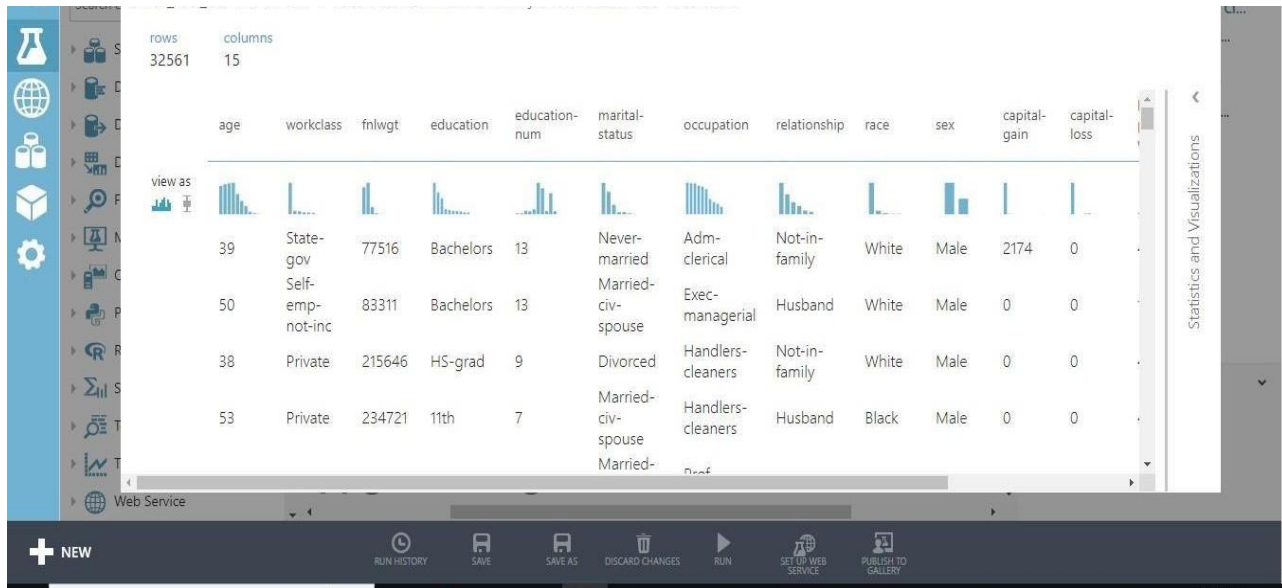
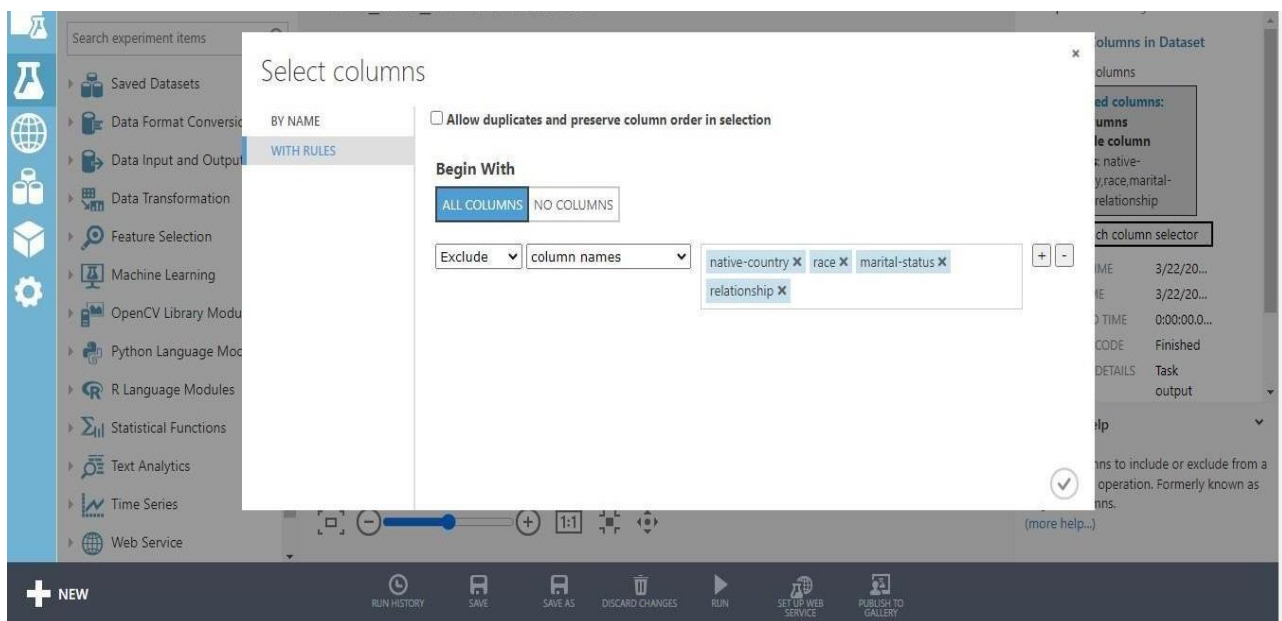


Lab 05 - IT19184722

Data set



Select columns.



Clean the missing data.

The screenshot shows the Azure ML Studio interface with a workflow titled 'Adult Census Income Binary...'. The workflow consists of three steps: 'Select Columns in Dataset', 'Clean Missing Data', and 'Clean Missing Data'. The 'Clean Missing Data' step is highlighted with a green checkmark. The 'Clean Missing Data' step is configured with the following settings:

- Columns to be cleaned: **All columns**
- Launch column selector: **Launch column selector**
- Minimum missing value: **0**
- Maximum missing value: **1**
- Cleaning mode: **Remove entire row**
- START TIME: 3/22/20...
- END TIME: 3/22/20...

The 'Quick Help' section on the right states: 'Specifies how to handle the values missing from a dataset (more help...)'. The bottom toolbar includes buttons for NEW, RUN HISTORY, SAVE, SAVE AS, DISCARD CHANGES, RUN, SET UP WEB SERVICE, and PUBLISH TO GALLERY.

The screenshot shows the 'Statistics' and 'Visualizations' panels in Azure ML Studio. The 'Statistics' panel displays the following data:

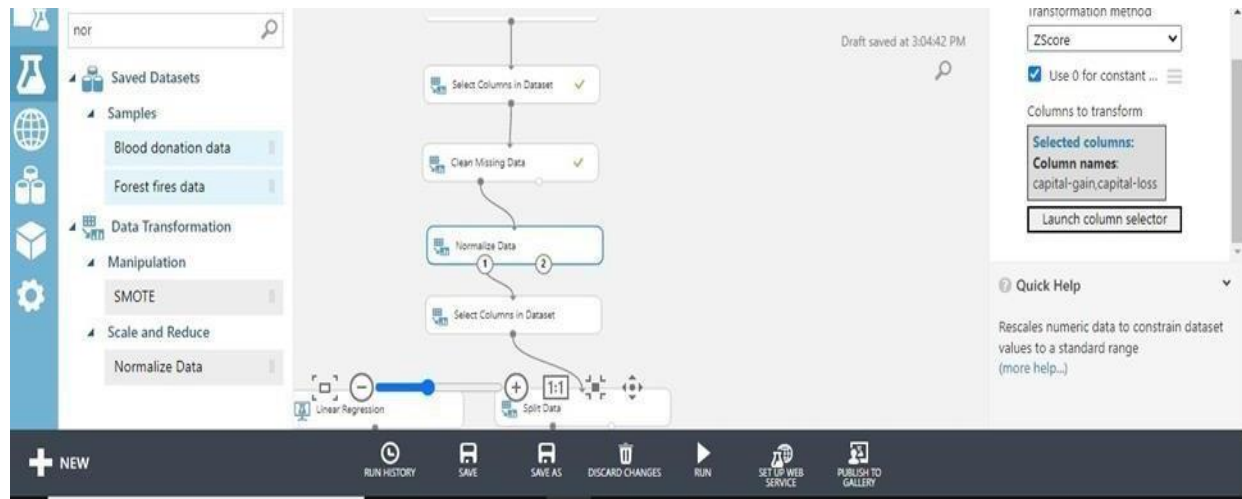
rows	columns
30718	11

The 'Visualizations' panel displays a bar chart with the following data:

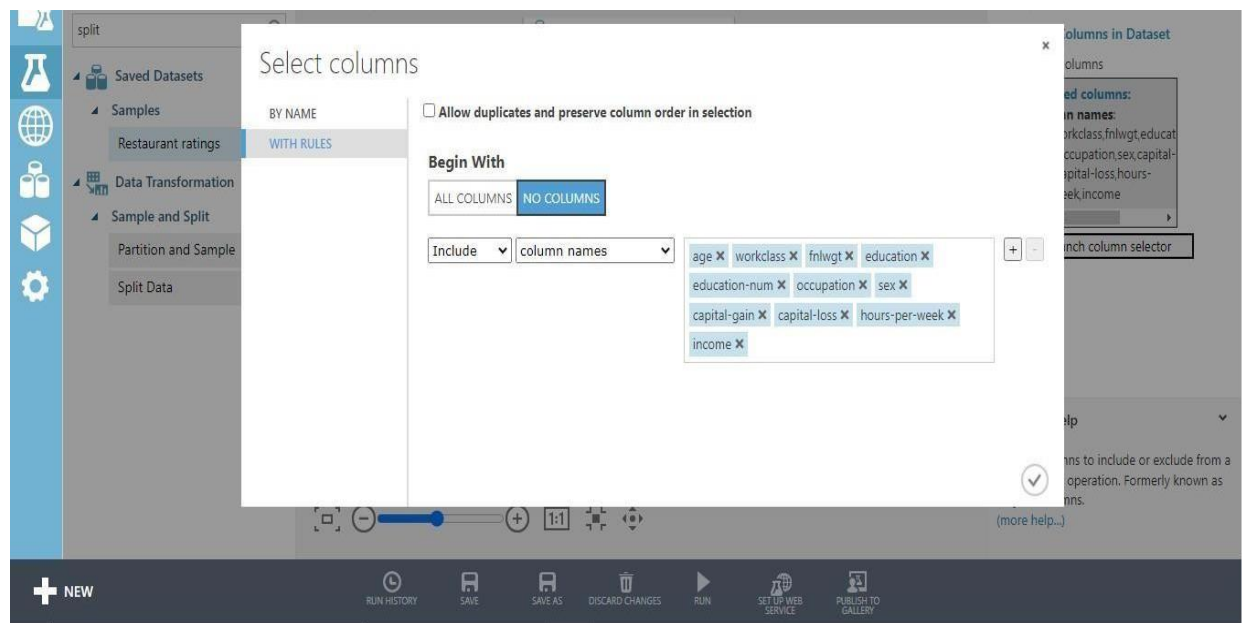
age	workclass	fnlwgt	education	education-num	occupation	sex
39	State-gov	77516	Bachelors	13	Adm-clerical	Male
50	Self-emp-not-inc	83311	Bachelors	13	Exec-managerial	Male
38	Private	215646	HS-grad	9	Handlers-cleaners	Male
53	Private	234721	11th	7	Handlers-cleaners	Male
28	Private	338409	Bachelors	13	Prof-specialty	Female

The bottom toolbar includes buttons for NEW, RUN HISTORY, SAVE, SAVE AS, DISCARD CHANGES, RUN, SET UP WEB SERVICE, and PUBLISH TO GALLERY.

Normalize the data.



Select Required Columns



Split data

split

Adult Census Income Binary...

Select Columns in Dataset

Clean Missing Data

Select Columns in Dataset

Split Data

Splitting mode: Split Rows

Fraction of rows in the first set: 0.75

Randomized split: ☒

Random seed: 0

Stratified split: False

Quick Help

Split the rows of a dataset into two distinct sets (more help...)

NEW RUN HISTORY SAVE SAVE AS DISCARD CHANGES RUN SET UP WEB SERVICE PUBLISH TO GALLERY

Train the model

nor

Normalize Data

Select Columns in Dataset

Linear Regression

Split Data

Train Model

Transformation method: ZScore

Use 0 for constant columns: ☒

Columns to transform: capital-gain, capital-loss

Quick Help

Rescales numeric data to constrain dataset values to a standard range (more help...)

NEW RUN HISTORY SAVE SAVE AS DISCARD CHANGES RUN SET UP WEB SERVICE PUBLISH TO GALLERY

Batch Linear Regressor

Settings

Setting	Value
Bias	True
Regularization	0.001
Allow Unknown Levels	True
Random Number Seed	

Feature Weights

Feature	Weight
Bias	-0.223065
sex_Female_0	-0.188294
occupation_Armed-Forces_1	-0.141213

Score Model

score

Machine Learning

- Score
- Score Matchbox Recom...
- Score Model

Text Analytics

- Score Vowpal Wabbit Versi...
- Score Vowpal Wabbit Versi...

Deprecated

- Score Vowpal Wabbit Versi...

Workflow:

```
graph TD; A[Select Columns in Dataset] --> B[Split Data]; A --> C[Linear Regression]; B --> D[Train Model]; B --> E[Score Model]; C --> D; D --> E;
```

Experiment Properties

START TIME 3/22/20...
END TIME 3/22/20...
STATUS CODE InDraft
STATUS DETAILS None

Prior Run

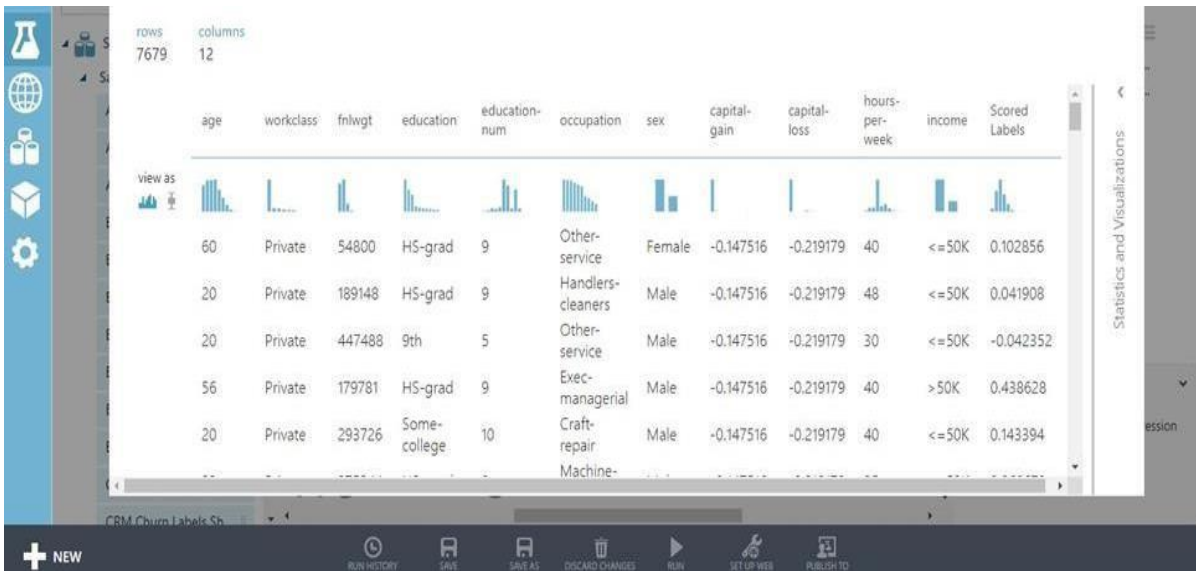
Summary

Enter a few sentences describing your experiment (up to 140 characters).

Quick Help

NEW RUN HISTORY SAVE SAVE AS DISCARD CHANGES RUN SET UP WEB SERVICE PUBLISH TO GALLERY

Scored Labels



Based on the Analysis and the statistical data,

- i. What features do you think will be the most appropriate for the Model Training?
 - 1. Age
 - 2. Work class
 - 3. education-num
 - 4. capital-gain.
 - 5. capital-loss.
 - 6. hours-per-week
 - 7. Education
 - 8. Income
 - 9. sex

- ii. What features should be cleaned?
 - 1. Relationship
 - 2. marital status
 - 3. Native Country
 - 4. Race

- iii. No

- iv. What features should be scaled?
 - 1. Capital-gain.
 - 2. Capital-loss.