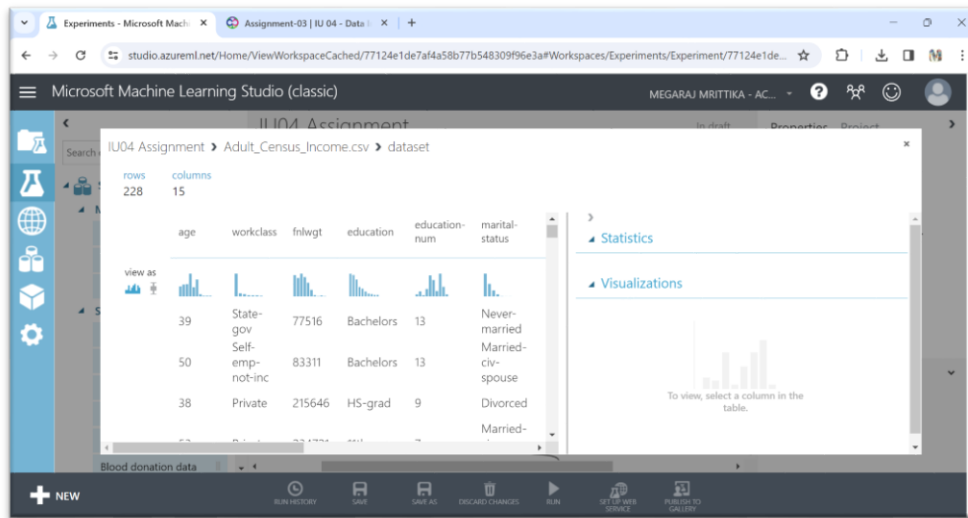
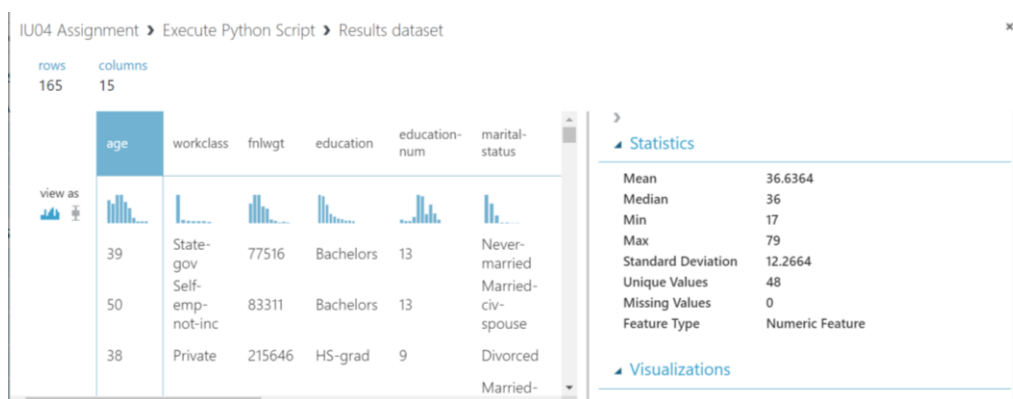
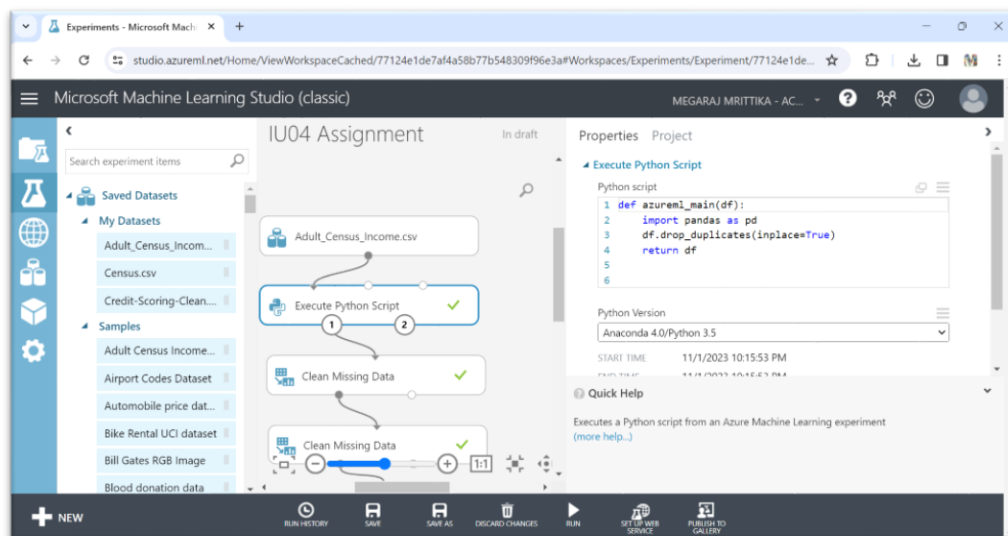


# Adult census Income Prediction

## Reading Dataset Using Azure ML



## Remove duplicate values from dataset



## Clean missing values from dataset

Clean the missing values of fnlwgt, education-num by Remove entire row

The screenshot shows the Microsoft Machine Learning Studio interface. On the left, the 'My Datasets' list includes 'Adult\_Census\_Income.csv'. The main workspace displays a workflow titled 'IU04 Assignment' with the following steps: 'Adult\_Census\_Income.csv', 'Execute Python Script', 'Clean Missing Data' (step 1), 'Clean Missing Data' (step 2), and 'Clean Missing Data'. The 'Clean Missing Data' step 1 is selected, and its properties are shown on the right. The 'Columns to be cleaned' section lists 'Selected columns: fnlwgt, education-num'. The 'Cleaning mode' is set to 'Remove entire row'. The 'Minimum missing value ratio' is 0, and the 'Maximum missing value ratio' is 1.

The screenshot shows the 'Cleaned dataset' view of the 'IU04 Assignment' workflow. The dataset has 155 rows and 15 columns. The columns are: age, workclass, fnlwgt, education, education-num, marital-status, and native-country. The 'view as' section shows histograms for each column. The 'Statistics' and 'Visualizations' sections are also visible.

rows	columns
155	15

age	workclass	fnlwgt	education	education-num	marital-status
39	State-gov	77516	Bachelors	13	Never-married
50	Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse
38	Private	215646	HS-grad	9	Divorced
					Married-

Clean the missing values of workclass, occupation, native-country by Remove entire row

The screenshot shows the Microsoft Machine Learning Studio interface. On the left, the 'My Datasets' list includes 'Adult\_Census\_Income.csv'. The main workspace displays a workflow titled 'IU04 Assignment' with the following steps: 'Adult\_Census\_Income.csv', 'Execute Python Script', 'Clean Missing Data' (step 1), 'Clean Missing Data' (step 2), and 'Clean Missing Data'. The 'Clean Missing Data' step 1 is selected, and its properties are shown on the right. The 'Columns to be cleaned' section lists 'Selected columns: workclass, occupation, native-country'. The 'Cleaning mode' is set to 'Remove entire row'. The 'Minimum missing value ratio' is 0, and the 'Maximum missing value ratio' is 1.

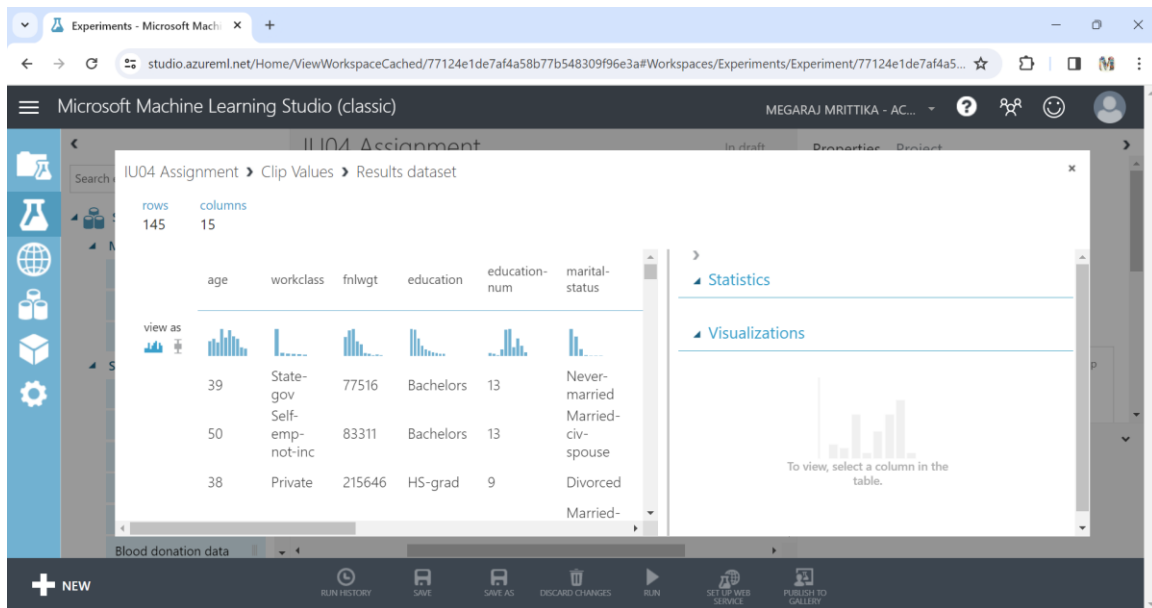
Clean the missing values of hours-per-week by Replace with Mean

The screenshot shows the Microsoft Machine Learning Studio (classic) interface. The main workspace displays a workflow titled "IU04 Assignment" in draft mode. The workflow starts with a dataset "Adult\_Census\_Income.csv", followed by an "Execute Python Script" step, and then three "Clean Missing Data" steps. The first "Clean Missing Data" step is selected, and its properties are shown on the right. The "Columns to be cleaned" section lists "Selected columns: hours-per-week". The "Cleaning mode" is set to "Replace with mean". The "Quick Help" section explains that this specifies how to handle missing values from a dataset.

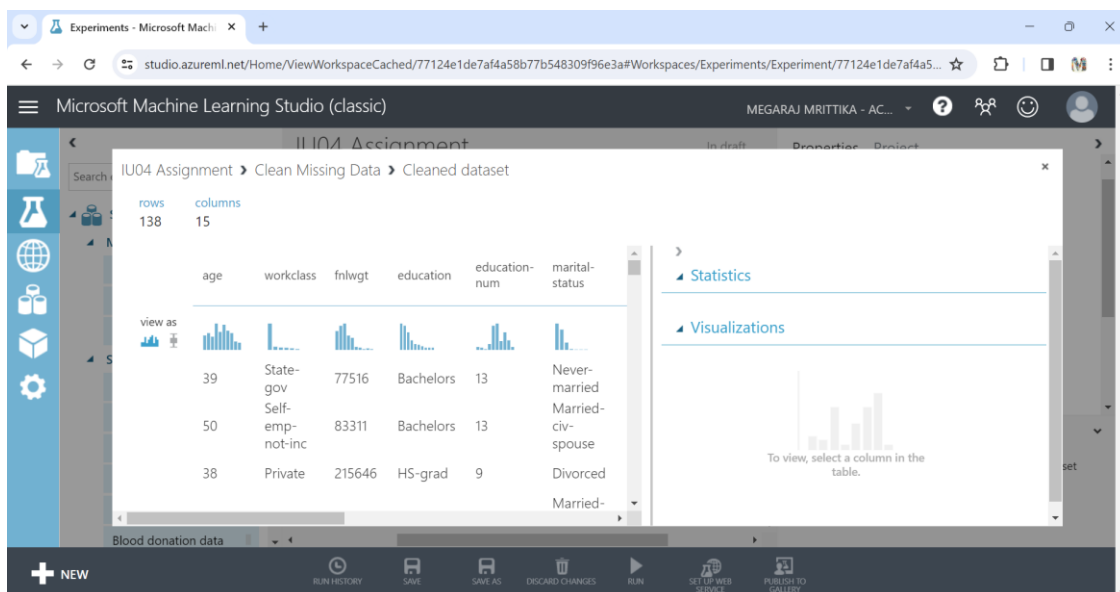
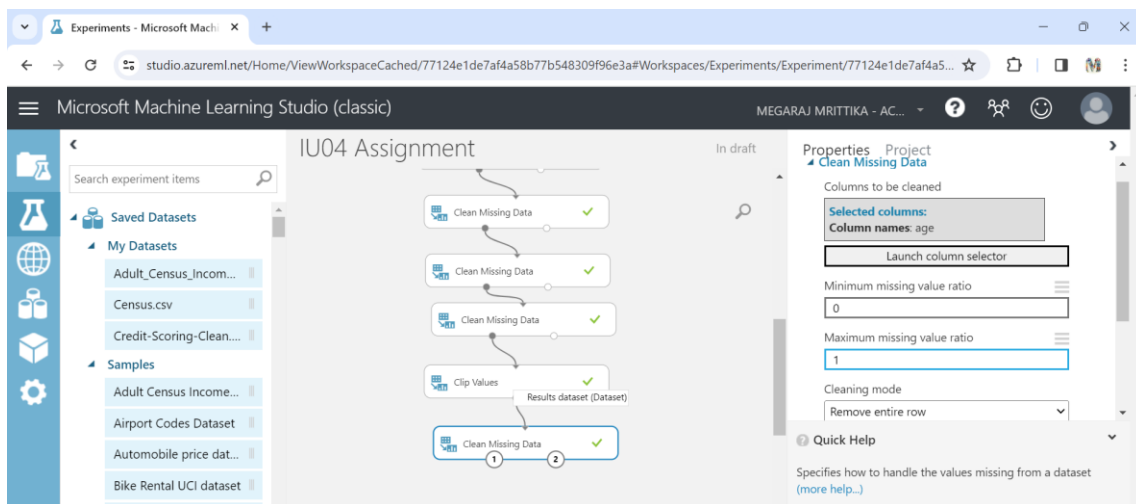
Remove outlier from data give python example

Clip values in column selector using age

The screenshot shows the Microsoft Machine Learning Studio (classic) interface. The main workspace displays a workflow titled "IU04 Assignment" in draft mode. The workflow starts with a dataset "Adult\_Census\_Income.csv", followed by three "Clean Missing Data" steps, and then a "Clip Values" step. The "Clip Values" step is selected, and its properties are shown on the right. The "Set of thresholds" is set to "ClipPeaks". The "Upper threshold" is set to "Percentile", and the "Percentile number for upper threshold" is set to "95". The "Upper substitute value" is set to "Missing". The "List of columns" section lists "Selected columns: age". The "Quick Help" section explains that this detects outliers and clips or replaces their values.



Clean missing data in age



## Split Data

Microsoft Machine Learning Studio (classic) - MEGARAJ MRITTIKA - AC...

Training experiment | Predictive experiment

IU04 Assignment

Finished running ✓

Properties | Project

**Split Data**

Splitting mode: Split Rows

Fraction of rows in the f...: 0.5

☒ Randomized split

Random seed: 0

Stratified split: False

**Quick Help**

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

## Train Data

Microsoft Machine Learning Studio (classic) - MEGARAJ MRITTIKA - AC...

IU04 Assignment > Split Data > Results dataset1

rows: 69, columns: 15

age	workclass	fnlwgt	education	education-num	marital-status
42	Private	124692	HS-grad	9	Married-civ-spouse
27	Private	213921	HS-grad	9	Never-married
45	Private	109434	Bachelors	13	Married-civ-spouse

view as: [Histogram]

**Statistics**

**Visualizations**

To view, select a column in the table.

## Test Data

Microsoft Machine Learning Studio (classic) - MEGARAJ MRITTIKA - AC...

IU04 Assignment > Split Data > Results dataset2

rows: 69, columns: 15

age	workclass	fnlwgt	education	education-num	marital-status
44	Self-emp-inc	78374	Masters	14	Divorced
20	Private	188300	Some-college	10	Never-married
37	Private	284582	Masters	14	Married-civ-spouse

view as: [Histogram]

**Statistics**

**Visualizations**

To view, select a column in the table.

## Train Model

The screenshot shows the Microsoft Machine Learning Studio (classic) interface. The main workspace displays a workflow titled "IU04 Assignment" with the following steps: Clean Missing Data, Clip Values, Clean Missing Data, Split Data, Train Model, Score Model, and Evaluate Model. The "Train Model" step is selected, and its properties are shown on the right. The "Label column" is set to "income", and the "Column names" are set to "income". The "Launch column selector" button is visible. The "Properties" pane on the right also shows the "Train Model" step's details, including "START TIME", "END TIME", "ELAPSED TIME", "STATUS CODE", and "STATUS DETAILS".

The screenshot shows the Microsoft Machine Learning Studio (classic) interface. The main workspace displays a workflow titled "IU04 Assignment" with the following steps: Clean Missing Data, Clip Values, Clean Missing Data, Split Data, Train Model, Score Model, and Evaluate Model. The "Trained model" step is selected, and its properties are shown on the right. The "Label column" is set to "income", and the "Column names" are set to "income". The "Launch column selector" button is visible. The "Properties" pane on the right also shows the "Trained model" step's details, including "START TIME", "END TIME", "ELAPSED TIME", "STATUS CODE", and "STATUS DETAILS".

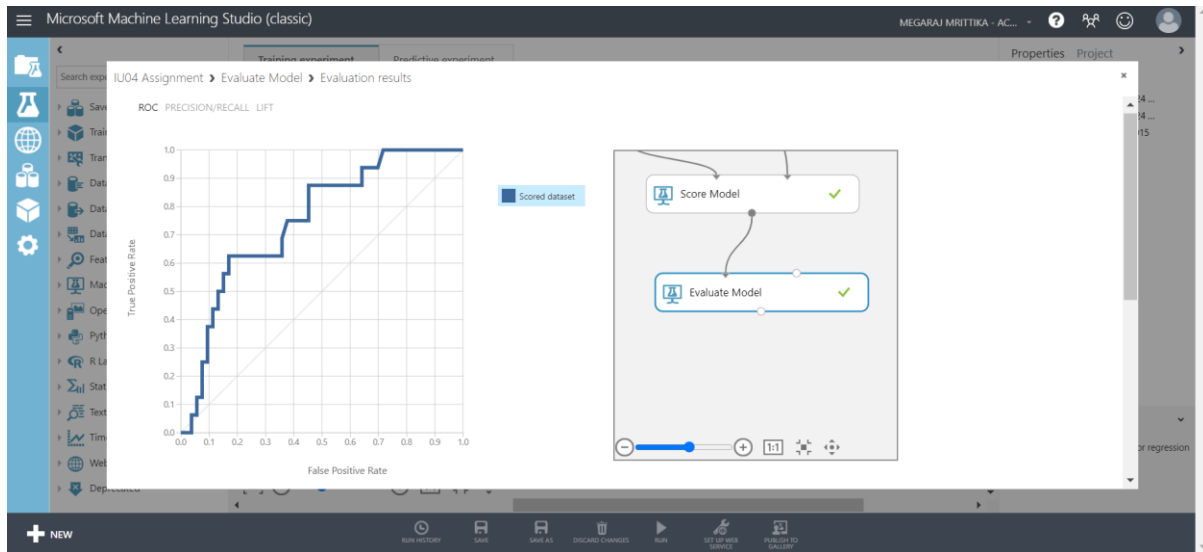
## Score Model

The screenshot shows the Microsoft Machine Learning Studio (classic) interface. The main workspace displays a workflow titled "IU04 Assignment" with the following steps: Clean Missing Data, Clip Values, Clean Missing Data, Split Data, Train Model, Score Model, and Evaluate Model. The "Scored dataset" step is selected, and its properties are shown on the right. The "Label column" is set to "income", and the "Column names" are set to "income". The "Launch column selector" button is visible. The "Properties" pane on the right also shows the "Scored dataset" step's details, including "START TIME", "END TIME", "ELAPSED TIME", "STATUS CODE", and "STATUS DETAILS".

rows	columns
69	17

age	workclass	fnlwgt	education	education-num	marital-status
44	Self-emp-inc	78374	Masters	14	Divorced
20	Private	188300	Some-college	10	Never-married
37	Private	284582	Masters	14	Married-civ-spouse

## Evaluate Model



Search explorer      IU04 Assignment > Evaluate Model > Evaluation results

False Positive Rate

True Positive      False Negative      Accuracy      Precision      Threshold      AUC

8      8      0.768      0.500      0.5      0.752

False Positive      True Negative      Recall      F1 Score

8      45      0.500      0.500

Positive Label      Negative Label

> 50K      <= 50K

Score Bin	Positive Examples	Negative Examples	Fraction Above Threshold	Accuracy	F1 Score	Precision	Recall	Negative Precision	Negative Recall	Cumulative AUC
(0.900,1.000]	2	3	0.072	0.754	0.190	0.400	0.125	0.781	0.943	0.001
(0.800,0.900]	2	2	0.130	0.754	0.320	0.444	0.250	0.800	0.906	0.008
(0.700,0.800]	3	2	0.203	0.768	0.467	0.500	0.438	0.836	0.868	0.024
(0.600,0.700]	1	0	0.217	0.783	0.516	0.533	0.500	0.852	0.868	0.024
(0.500,0.600]	0	1	0.232	0.768	0.500	0.500	0.500	0.849	0.849	0.033
(0.400,0.500]	1	1	0.261	0.768	0.529	0.500	0.563	0.863	0.830	0.044
(0.300,0.400]	1	3	0.319	0.739	0.526	0.455	0.625	0.872	0.774	0.079
(0.200,0.300]	0	4	0.377	0.681	0.476	0.385	0.625	0.860	0.698	0.126
(0.100,0.200]	2	5	0.478	0.638	0.490	0.364	0.750	0.889	0.604	0.190
(0.000,0.100]	4	32	1.000	0.232	0.376	0.232	1.000	1.000	0.000	0.752

NEW

Run history      Save      Save As      Discard changes      Run      Set up web service      Publish to gallery

## Set Up the web Services

