

# DATA SCIENCE BOOTCAMP

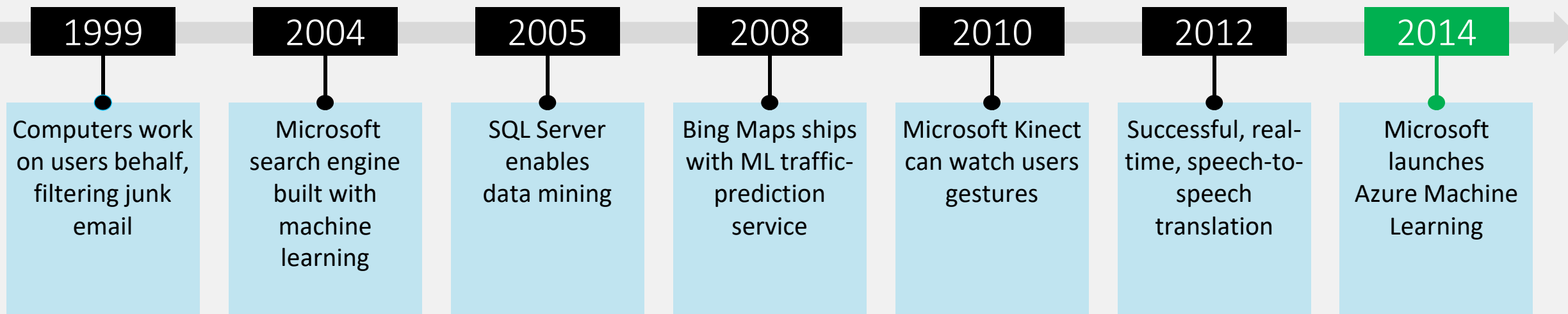
## MODULE 4

# Module Outline

- Overview of Azure ML Studio
- ML Process
- Creating ML Models in Azure ML Studio



## Microsoft & Machine Learning 15 years of realizing innovation



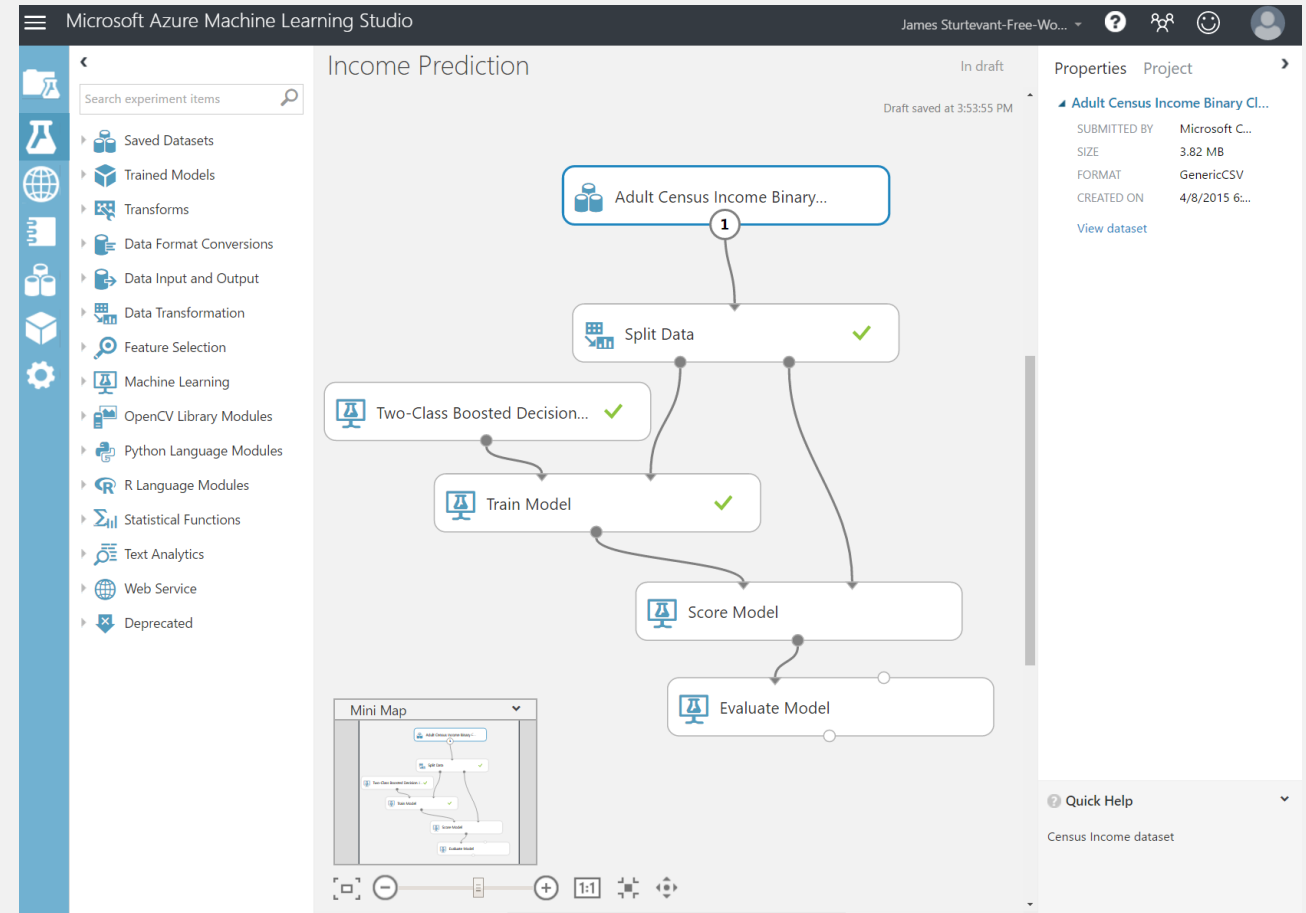
# Azure ML Studio (Cont'd)



Microsoft



- Cloud Based
- ML Studio for composing models
  - Supports numerous input formats
  - Supports R and Python



# Azure ML Studio (Cont'd)

## ML Algorithms

Machine Learning

- ▶ Evaluate
- ▶ Initialize Model
  - ▶ Anomaly Detection
    - One-Class Support Vector Machine
    - PCA-Based Anomaly Detection
  - ▶ Classification
  - ▶ Clustering
  - ▶ Regression

Machine Learning

- ▶ Evaluate
- ▶ Initialize Model
  - ▶ Anomaly Detection
  - ▶ Classification
    - Multiclass Decision Forest
    - Multiclass Decision Jungle
    - Multiclass Logistic Regression
    - Multiclass Neural Network
    - One-vs-All Multiclass
    - Two-Class Averaged Perceptron
    - Two-Class Bayes Point Machine
    - Two-Class Boosted Decision Tree
    - Two-Class Decision Forest
    - Two-Class Decision Jungle
    - Two-Class Locally-Deep Support...
    - Two-Class Logistic Regression
    - Two-Class Neural Network
    - Two-Class Support Vector Machine
  - ▶ Clustering
  - ▶ Regression

Machine Learning

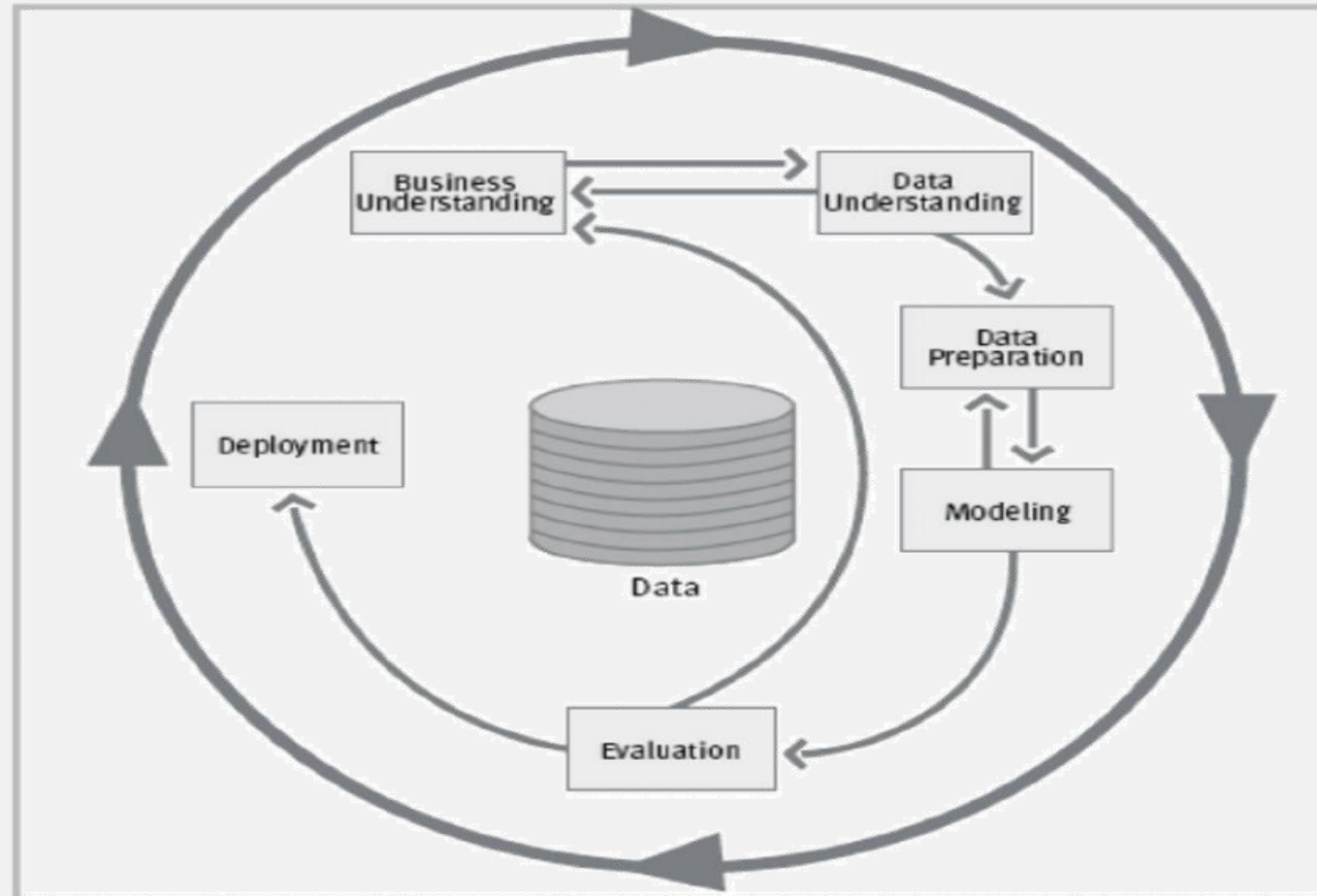
- ▶ Evaluate
- ▶ Initialize Model
  - ▶ Anomaly Detection
  - ▶ Classification
  - ▶ Clustering
    - K-Means Clustering
  - ▶ Regression

Machine Learning

- ▶ Evaluate
- ▶ Initialize Model
  - ▶ Anomaly Detection
  - ▶ Classification
  - ▶ Clustering
  - ▶ Regression
    - Bayesian Linear Regression
    - Boosted Decision Tree Regression
    - Decision Forest Regression
    - Fast Forest Quantile Regression
    - Linear Regression
    - Neural Network Regression
    - Ordinal Regression
    - Poisson Regression

# Machine Learning Process (Overview)

## CRISP DM Cycle



# ML Process – Business Understanding

- Understand the project objectives and requirements
- Can it be converted into a data mining problem
- Were any effort made in the past? If yes, what were the findings? Why are we doing it again? What has changed?
- Assess availability of time, technology and human resources. Do we have enough time and resources to execute the analytics project?
- Identify the success criteria, key risks and major stake holders.

# ML Process – Data Understanding

- Get familiar with the data. Is it enough to solve the stated business problem? If not, do we need to redesign the data collection process?
- What's needed vs. what's available
- Identify data quality problems
- Determine the structures and tools needed



# ML Process – Data Preparation

- Tasks include attribute selection as well as transformation and cleaning of data
- Understand what to keep and what to discard
- Extensive use of exploratory data analysis and visualization
- Process likely to be repeated multiple times

# ML Process – Modelling

- Selection of Machine Learning Algorithm according to the business problem you are trying to solve
- Train the model over training data
- Predict the class label for the test data

# ML Process – Evaluation

- Test robustness of the models under consideration by gauging their performances against hold-out data.
- Analyze if the models achieve the business objectives.

# ML Process – Deployment



Microsoft



- Important for the customer to understand up front the actions needed to actually make use of the created models.
- Define process to update and retrain the model, as needed

## (Explore the data dictionary)

- Go to the following github repository to download the LoanGrant\_Data\_Dictionary.docx
- <https://github.com/ahadmushir/MSDSBOOTCAMP>

# Creating ML Models In Azure ML Studio

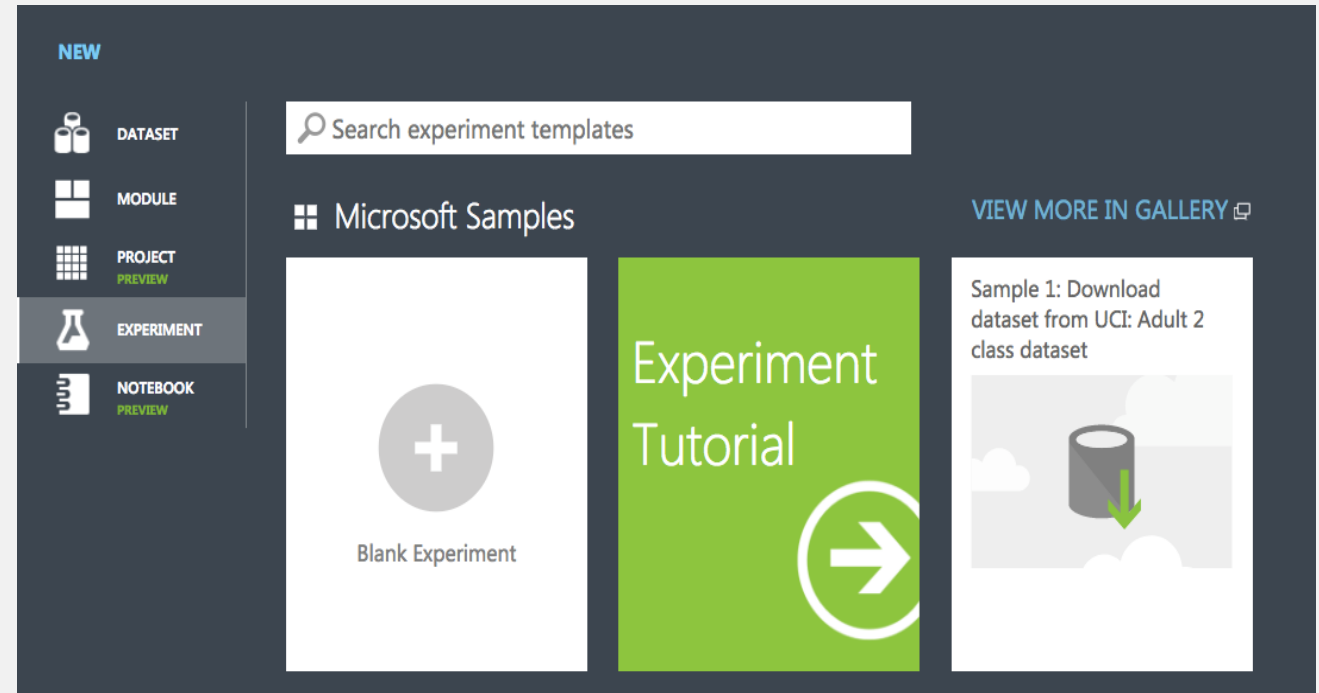


- Make sure you already have an Azure ML Studio account.
- Log on <https://studio.azureml.net/>

# Creating ML Models In Azure ML Studio

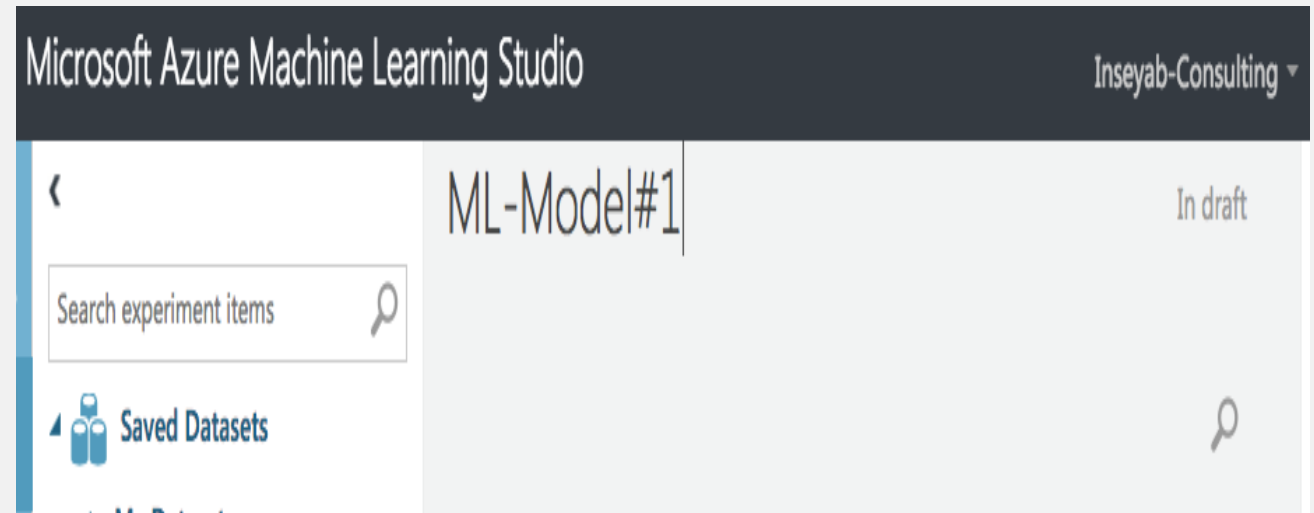


- Create a new blank experiment.



# Creating ML Models In Azure ML Studio

- Rename the experiment to ***ML-Model#1***

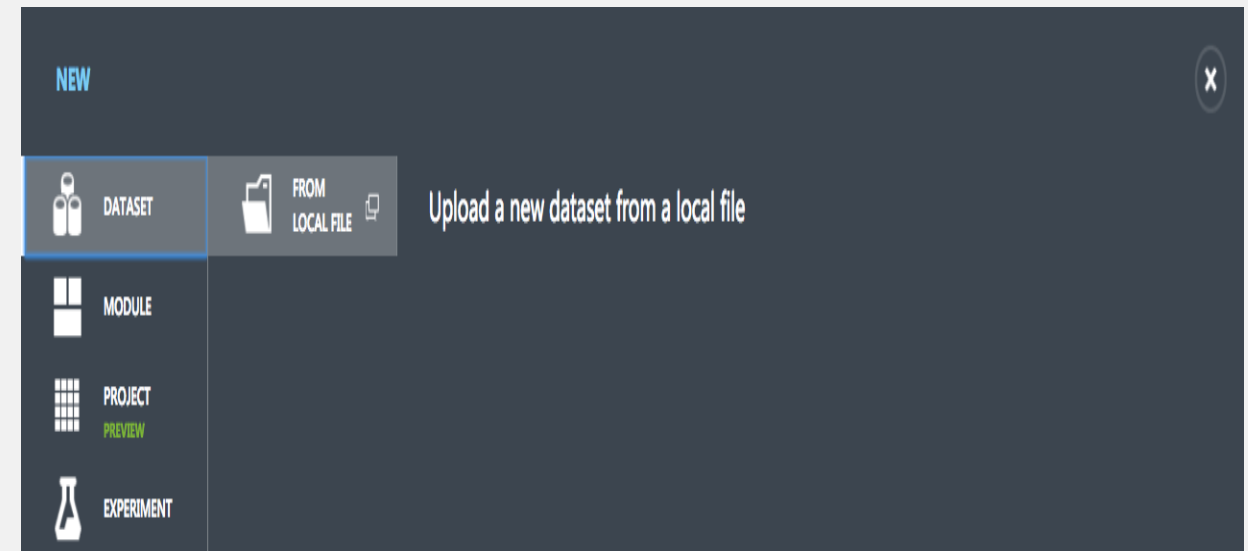




# Creating ML Models In Azure ML Studio

## (Uploading your data)

- Go to the following github repository to download the loangrant.csv
- <https://github.com/ahadmushir/MSDSBOOTCAMP>
- Upload the data in Azure ML Studio



# Creating ML Models In Azure ML Studio (Uploading your data)

×

Upload a new dataset

SELECT THE DATA TO UPLOAD:

Choose File

loangrant.csv

☐ This is the new version of an existing dataset

ENTER A NAME FOR THE NEW DATASET:

loangrant.csv

SELECT A TYPE FOR THE NEW DATASET:

Generic CSV File with a header (.csv) ▾

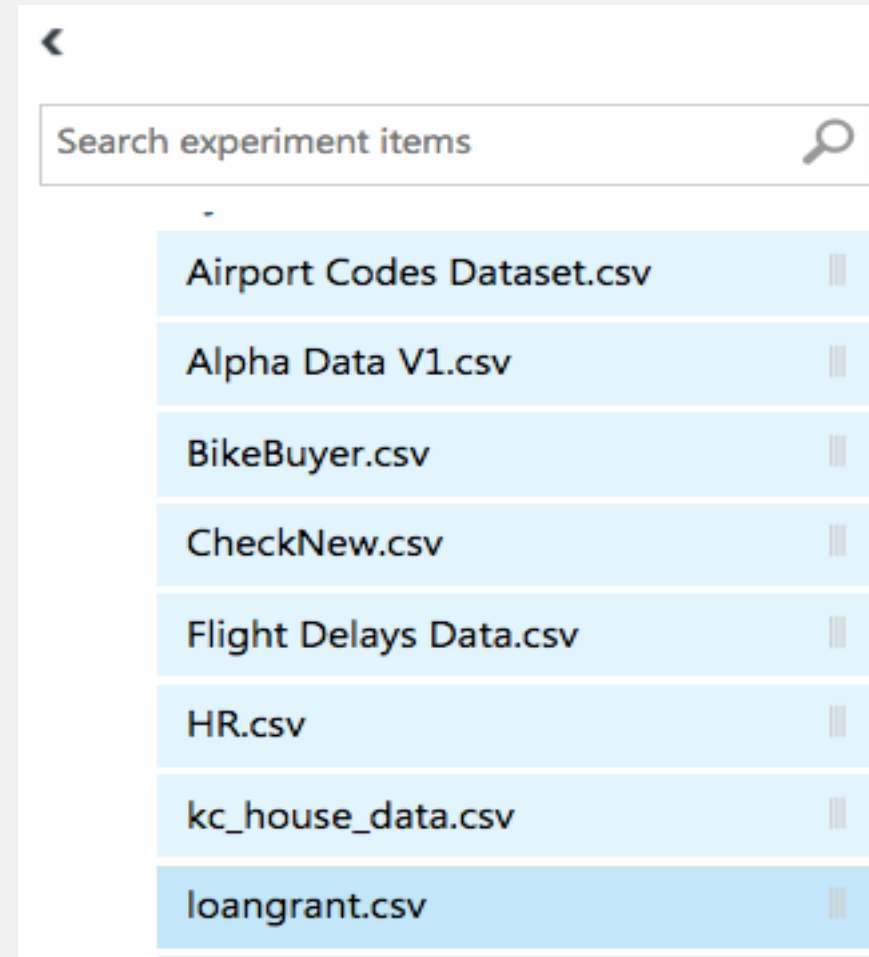
PROVIDE AN OPTIONAL DESCRIPTION:

✓

# Creating ML Models In Azure ML Studio

## (Uploading your data)

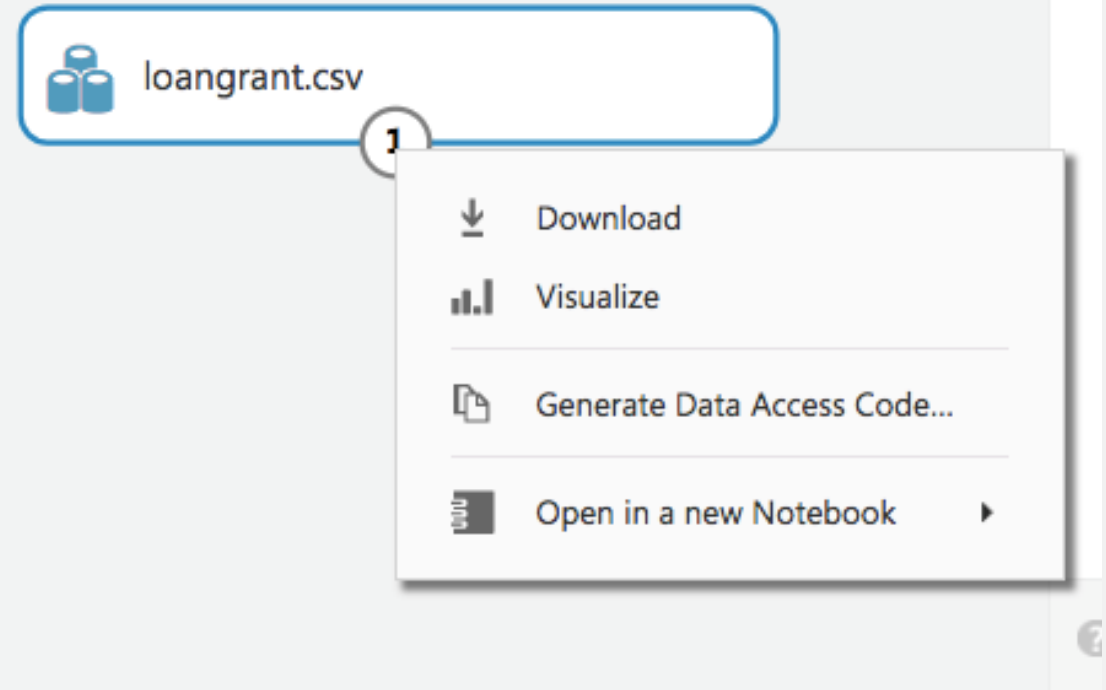
- Verify that your data is uploaded.



# Creating ML Models In Azure ML Studio

## (Explore your data)

- Right click on the dataset module.



# Creating ML Models In Azure ML Studio



Microsoft



## (Explore your data)

- Explore your dataset and its distributions.

Experiment created on 10/30/2017 > loangrant.csv > dataset

rows  
111107

columns  
19

Loan ID	Customer ID	Loan Status	Current Loan Amount	Term	Credit Score	Years in current job	Home Ownership	Annual Income	Purpose	
6cf51492-02a2-423e-b93d-676f05b9ad53	7c202b37-2add-44e8-9aea-d5b119aea935	Charged Off	12232	Short Term	7280	< 1 year	Rent	46643	Debt Consolidation	7
552e7ade-4292-4354-9ff9-c48031697d72	e7217b0a-07ac-47dd-b379-577b5a35b7c6	Charged Off	25014	Long Term	7330	10+ years	Home Mortgage	81099	Debt Consolidation	8
9b5e32b3-8d76-4801-afc8-472045e2e6b0	0a62fc41-16c8-40b5-92ff-0a4b762e7114	Charged Off	16117	Short Term	7240	9 years	Home Mortgage	60438	Home Improvements	1

# Creating ML Models In Azure ML Studio

## (Explore your data)



Microsoft



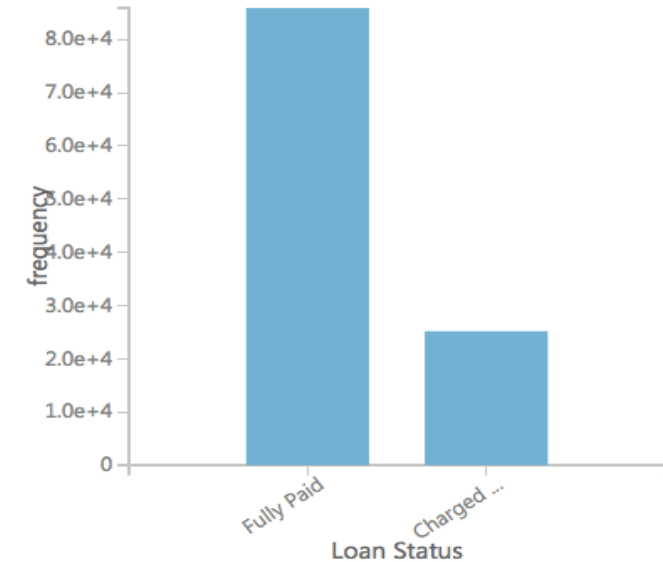
- Explore your class label and its distributions.

Experiment created on 10/30/2017 > loangrant.csv > dataset

rows  
111107

columns  
19

view as		Loan ID	Customer ID	Loan Status	Current Loan Amount	Terr
		6cf51492-02a2-423e-b93d-676f05b9ad53	7c202b37-2add-44e8-9aea-d5b119aea935	Charged Off	12232	Short Term
		552e7ade-4292-4354-9ff9-c48031697d72	e7217b0a-07ac-47dd-b379-577b5a35b7c6	Charged Off	25014	Long Term
		9b5e32b3-8d76-4801-afc8-172215206108	0a62fc41-16c8-40b5-92ff-014156205114	Charged Off	16117	Short Term



# Creating ML Models In Azure ML Studio

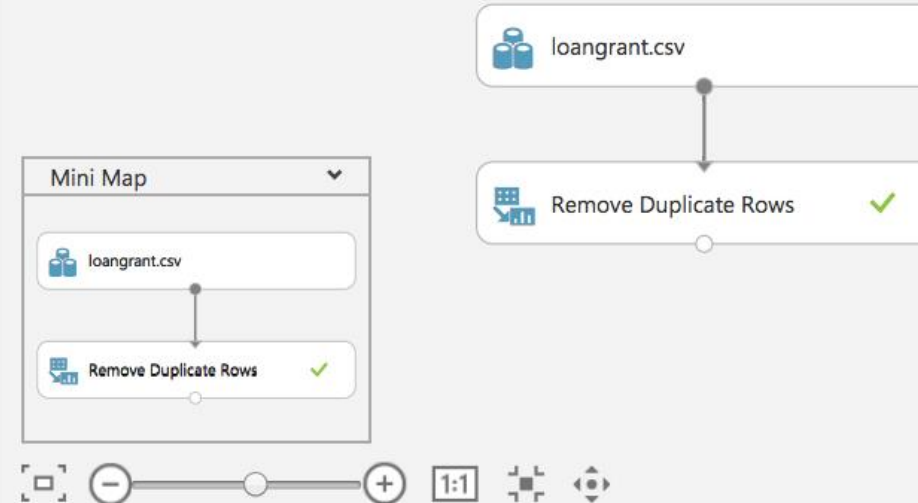
## (Removing Duplicates)

- Removing Duplicate Values on the basis of Loan ID.

ML-Model#1

In draft

Draft saved at 3:36:01 PM



# Creating ML Models In Azure ML Studio

## (Removing Duplicates)



Microsoft



- Explore the data.

ML-Model#1 > Remove Duplicate Rows > Results dataset

rows 88910 columns 19

	Loan ID	Customer ID	Loan Status	Current Loan Amount	Term	Credit Score	Years in current job	Home Ownership	Annual Income	Purpose
view as										
	6cf51492-02a2-423e-b93d-676f05b9ad53	7c202b37-2add-44e8-9aea-d5b119aea935	Charged Off	12232	Short Term	7280	< 1 year	Rent	46643	Debt Consolidat
	552e7ade-4292-4354-9ff9-c48031697d72	e7217b0a-07ac-47dd-b379-577b5a35b7c6	Charged Off	25014	Long Term	7330	10+ years	Home Mortgage	81099	Debt Consolidat
	9b5e32b3-8d76-4801-	0a62fc41-16c8-40b5-	Charged		Short			Home		Home

Statistics and Visualizations



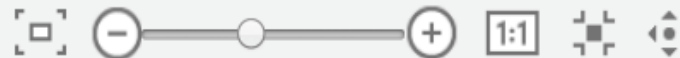
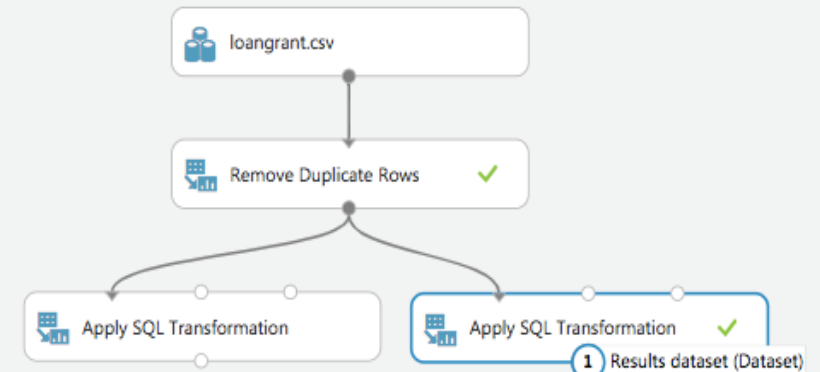
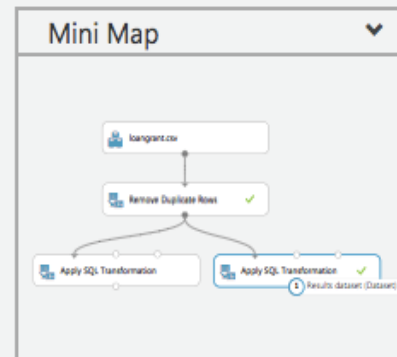
# Creating ML Models In Azure ML Studio

## (Handling “Credit Score”)

- Introducing “Apply SQL Transformation” module.
- Apply SQL syntax to your dataset.
- No need of any other IDE.

ML-Model#1

Finished running selected items ✓



# Creating ML Models In Azure ML Studio (Handling “Credit Score”)

## Apply SQL Transformation

SQL Query Script

```
1 select * from t1 where "Credit Score" <= 800 or "Credit Score" is null;
```

START TIME 4/18/2018 11:44:09 PM

END TIME 4/18/2018 11:44:10 PM

# Creating ML Models In Azure ML Studio (Handling “Credit Score”)

Finished running selected items ✓  
Draft saved at 3:51:06 PM



Properties Project

## Apply SQL Transformation

SQL Query Script

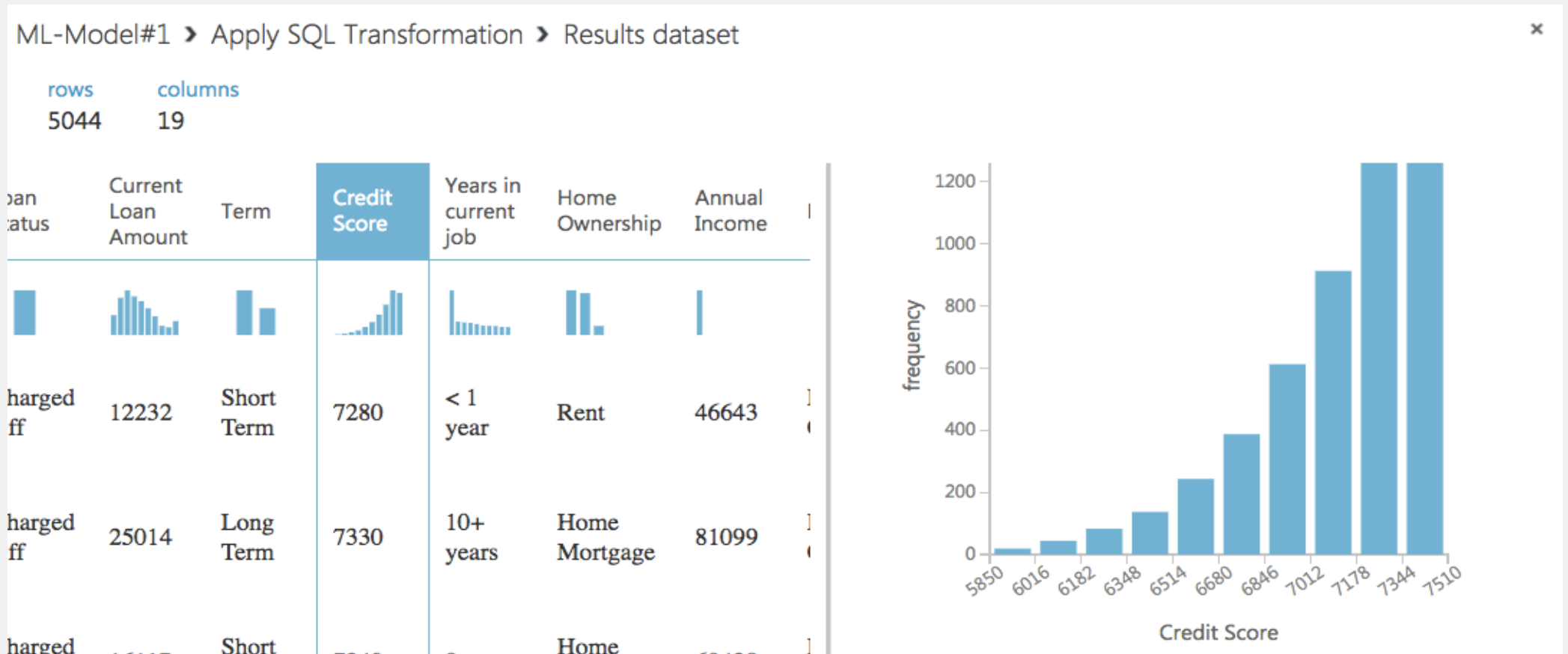
```
1 select * from t1 where "Credit Score" >= 1000;
```

## Quick Help

Runs a SQLite query on input datasets to transform the data  
([more help...](#))

# Creating ML Models In Azure ML Studio

## (Handling “Credit Score”)



# Creating ML Models In Azure ML Studio



Microsoft

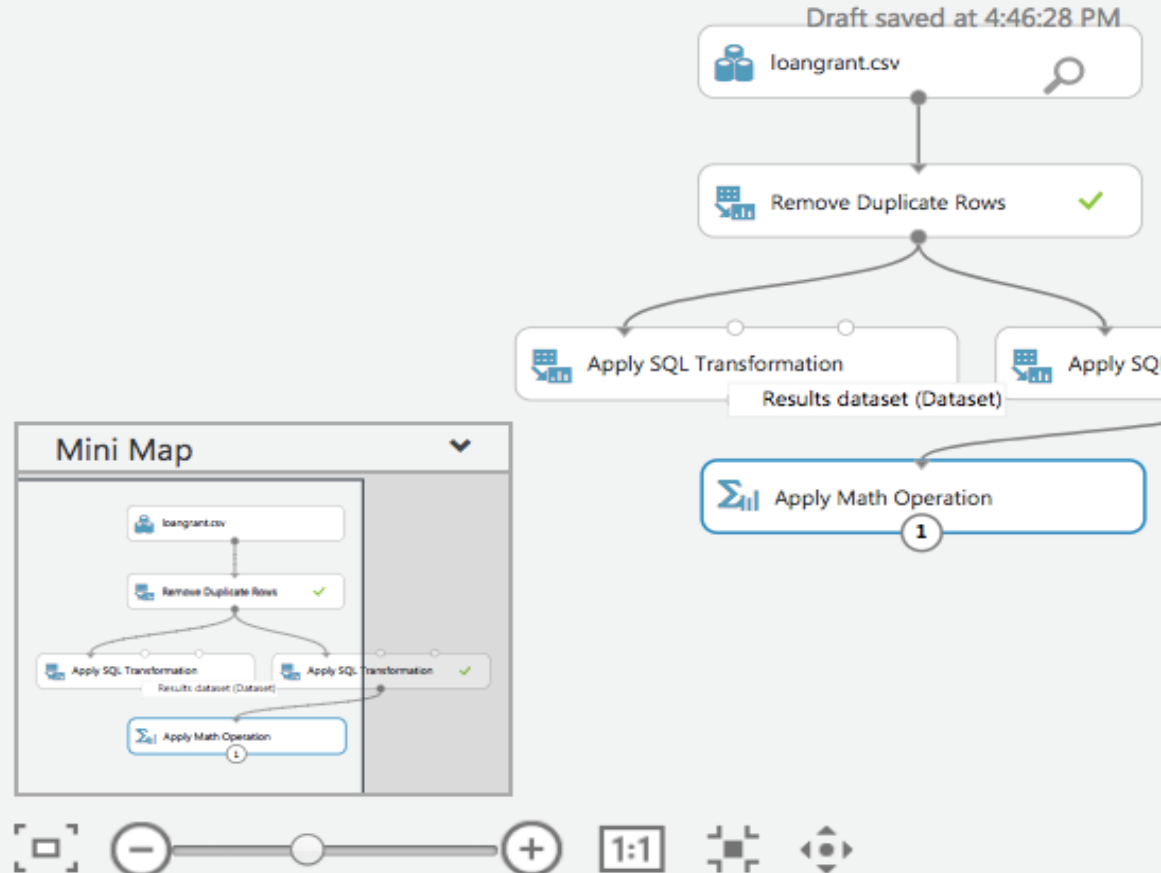


## (Handling “Credit Score”)

ML-Model#1

In draft

Draft saved at 4:46:28 PM



Properties Project

### Apply Math Operation

Category

Operations

Basic operation

Divide

Operation argument type

Constant

Constant operation argum...

10

Column set

### Quick Help

Applies a mathematical operation to column values

[\(more help...\)](#)

# Creating ML Models In Azure ML Studio



Microsoft



## (Handling “Credit Score”)

### Select columns

BY NAME

WITH RULES

AVAILABLE COLUMNS

All Types search columns

Loan ID  
Customer ID  
Loan Status  
Current Loan Amount  
Term  
Years in current job  
Home Ownership  
Annual Income  
Purpose  
Monthly Debt  
Years of Credit History  
Months since last delinquent  
Number of Open Accounts  
Number of Credit Problems  
Current Credit Balance  
Maximum Open Credit

18 columns available

SELECTED COLUMNS

All Types search columns

Credit Score

1 columns selected

>

<

✓

# Creating ML Models In Azure ML Studio

## (Handling “Credit Score”)

- Ensure that the output mode is “InPlace”.

Output mode

Inplace

? Quick Help

Applies a mathematical operation to column values

[\(more help...\)](#)

# Creating ML Models In Azure ML Studio

## (Handling “Credit Score”)



Microsoft

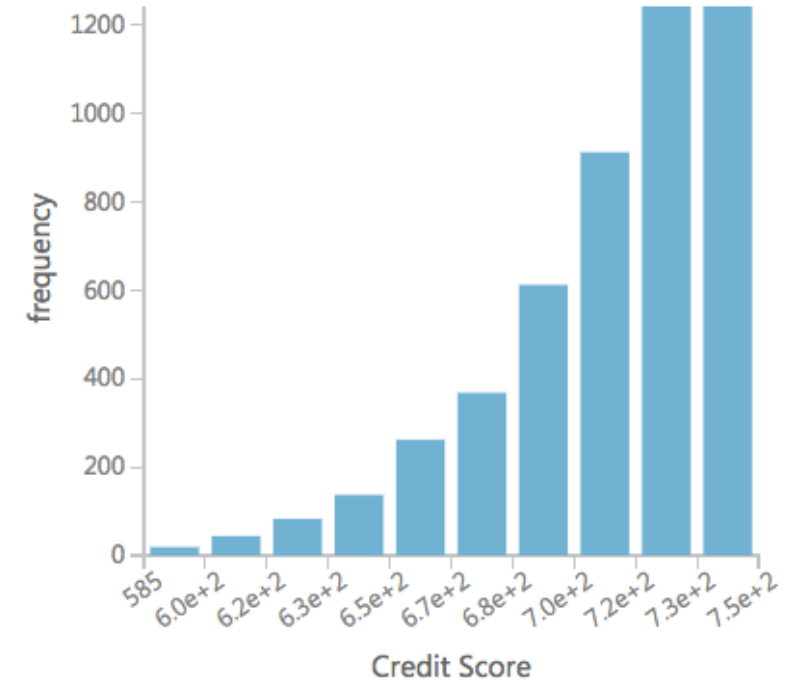


ML-Model#1 > Apply Math Operation > Results dataset

rows  
5044

columns  
19

Term	Credit Score	Years in current job	Home Ownership	Annual Income	Purpose	Mo Deb
Short Term	728	< 1 year	Rent	46643	Debt Consolidation	777
Long Term	733	10+ years	Home Mortgage	81099	Debt Consolidation	892
Short			Home		Home	





# Creating ML Models In Azure ML Studio



Microsoft

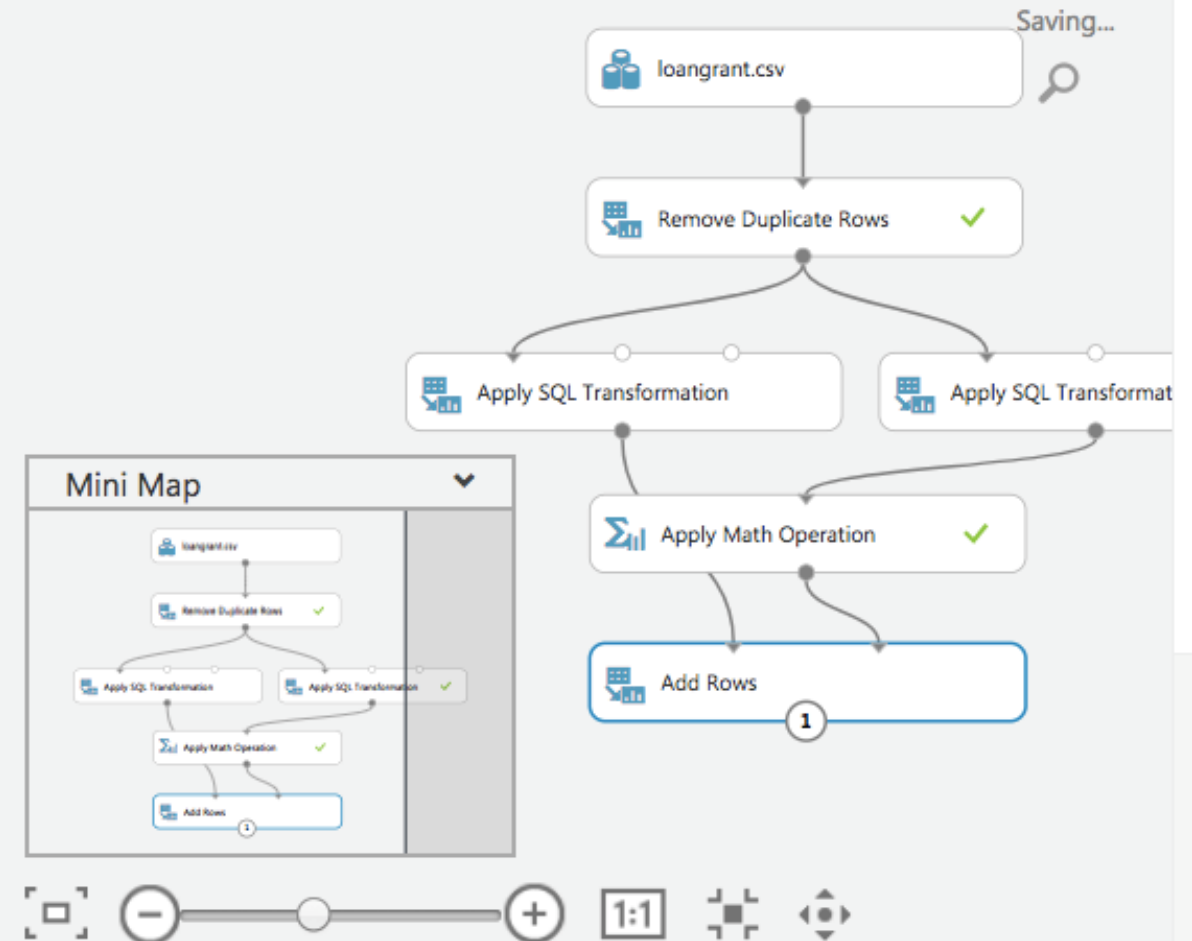


## (Handling “Credit Score”)

- Adding transformed data back to original dataset.

ML-Model#1

In draft

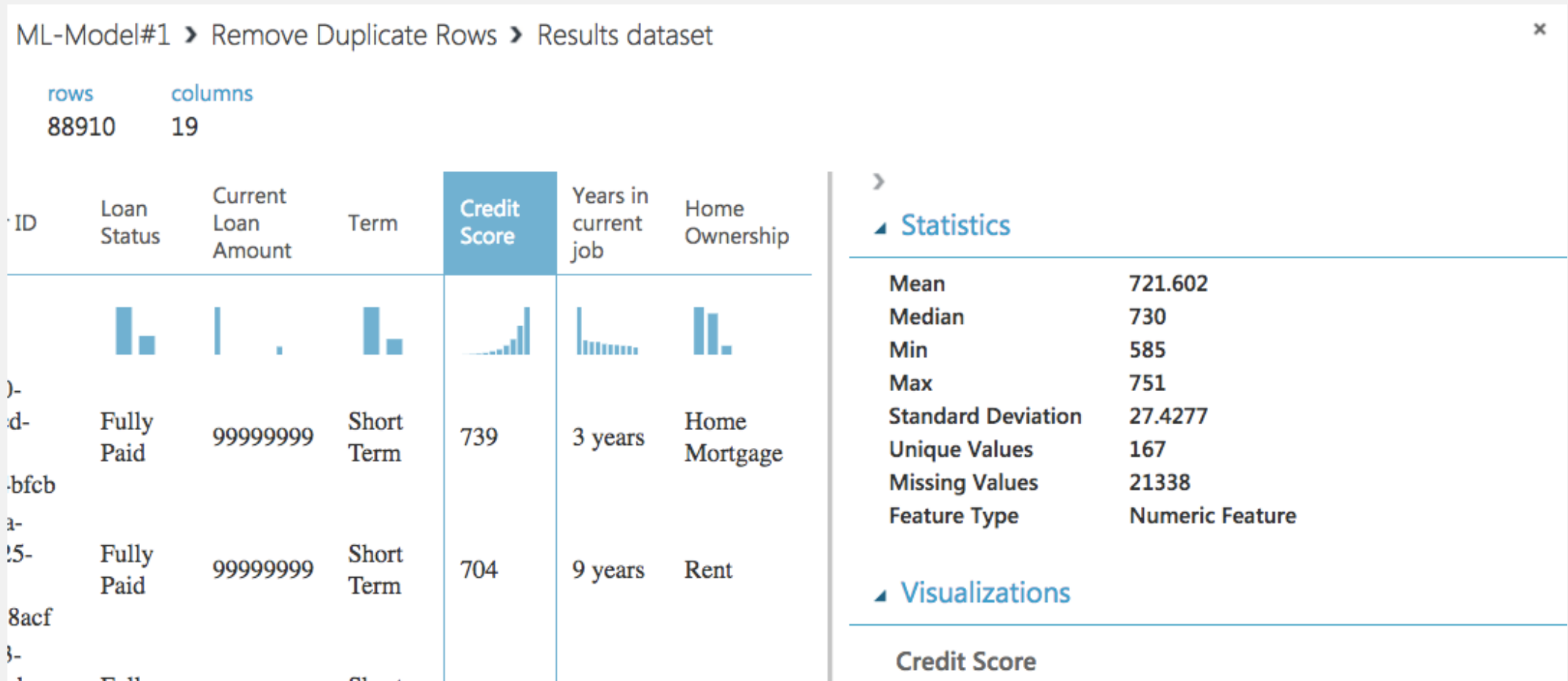


# Creating ML Models In Azure ML Studio

## (Handling “Credit Score”)



Microsoft



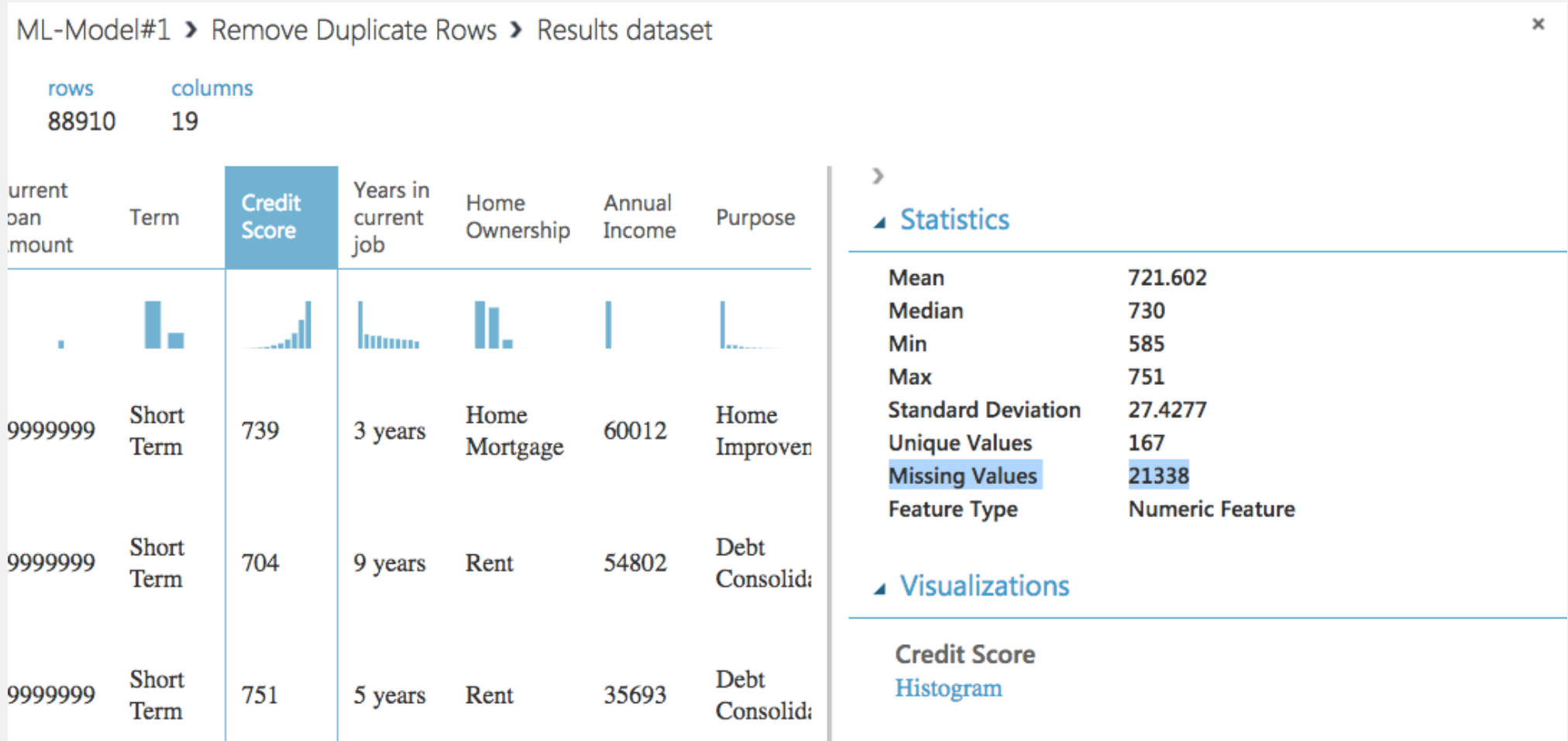
# Creating ML Models In Azure ML Studio



Microsoft



## (Handling “Credit Score”)



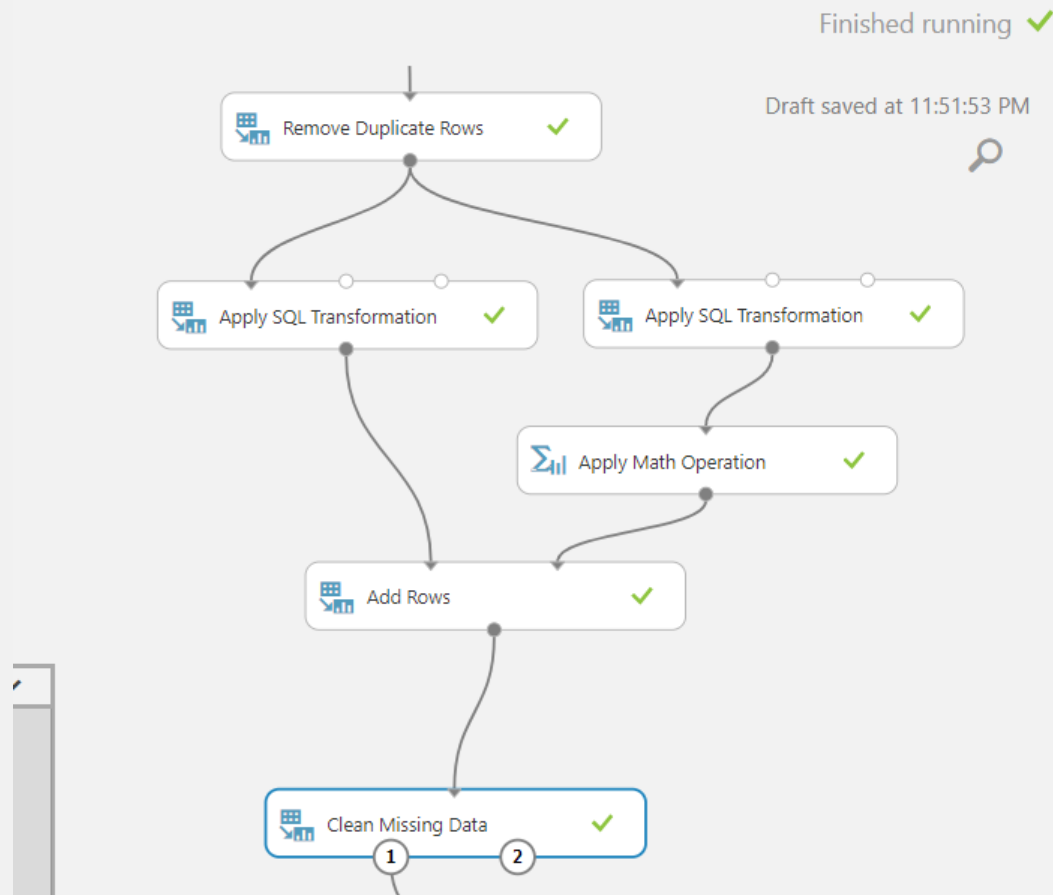
# Creating ML Models In Azure ML Studio



Microsoft



## (Handling “Credit Score”)



### Clean Missing Data

Columns to be cleaned

**Selected columns:**

**Column names:** Credit Score

Launch column selector

Minimum missing value ratio

0

Maximum missing value ratio

1

Cleaning mode

Replace with median

Cols with all missing values

Remove

☐ Generate missing value indicator column

START TIME 4/18/2018 11:44:33 PM

END TIME 4/18/2018 11:44:38 PM

Quick Help

# Creating ML Models In Azure ML Studio

## (Handling “Credit Score”)



Microsoft



### Select columns

BY NAME

WITH RULES

AVAILABLE COLUMNS

All Types search columns

Loan ID  
Customer ID  
Loan Status  
Current Loan Amount  
Term  
Years in current job  
Home Ownership  
Annual Income  
Purpose  
Monthly Debt  
Years of Credit History  
Months since last delinquent  
Number of Open Accounts  
Number of Credit Problems  
Current Credit Balance  
Maximum Open Credit

18 columns available

SELECTED COLUMNS

All Types search columns

Credit Score

1 columns selected

>


<

✓


# Creating ML Models In Azure ML Studio

## (Handling “Credit Score”)

- Ensure to replace with Median.


Minimum missing value ra... 


0


Maximum missing value r... 


1

Cleaning mode

Replace with median 

Cols with all missing values 

Remove 

☐ Generate missing valu... 

# Creating ML Models In Azure ML Studio



Microsoft



## (Handling “Credit Score”)

ML-Model#1 > Clean Missing Data > Cleaned dataset

rows 88910  
columns 19

Loan Status	Current Loan Amount	Term	Credit Score	Years in current job	Home Ownership	Annual Income
Fully Paid	99999999	Short Term	739	3 years	Home Mortgage	60012
Fully Paid	99999999	Short Term	704	9 years	Rent	54802
Fully Paid	99999999	Short Term	751	5 years	Rent	35693

### Statistics

Mean	723.6174
Median	730
Min	585
Max	751
Standard Deviation	24.1785
Unique Values	167
Missing Values	0
Feature Type	Numeric Feature

### Visualizations

Credit Score  
[Histogram](#)

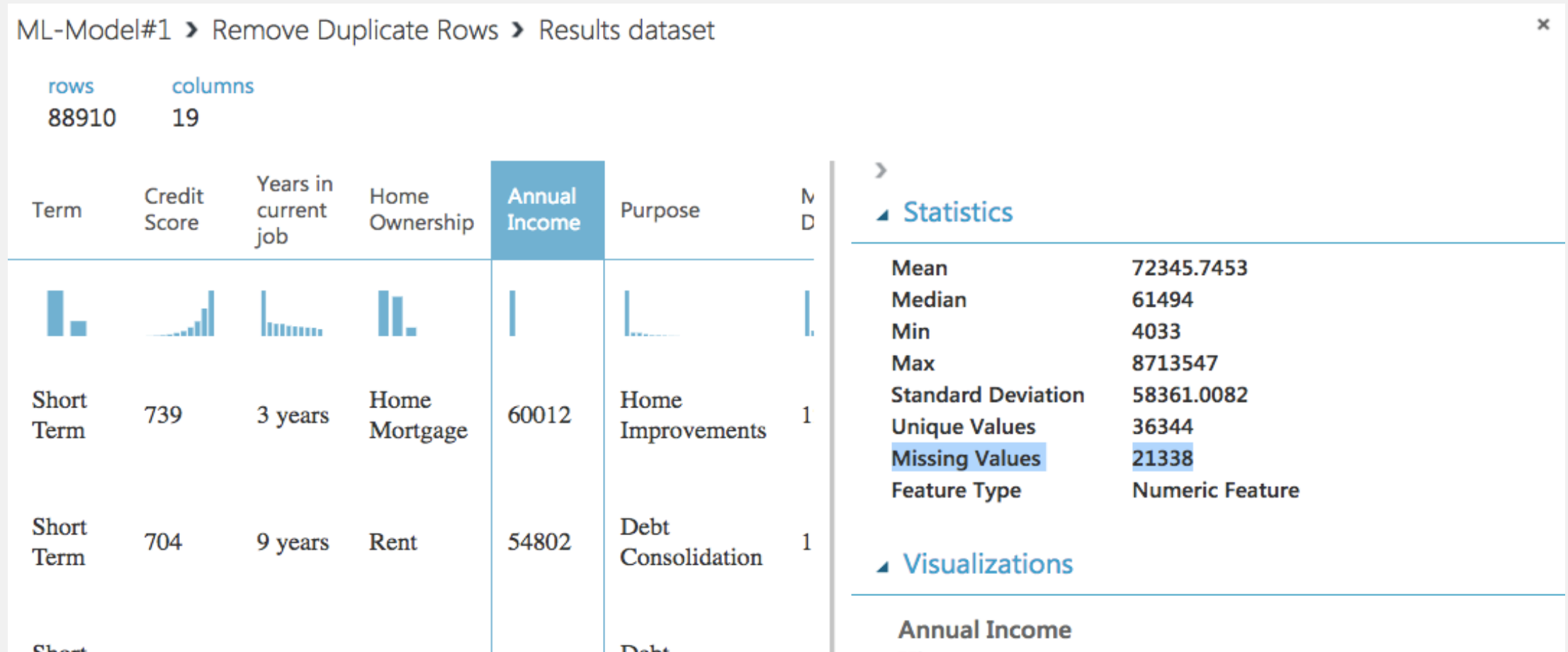
# Creating ML Models In Azure ML Studio



Microsoft



## (Handling “Annual Income”)





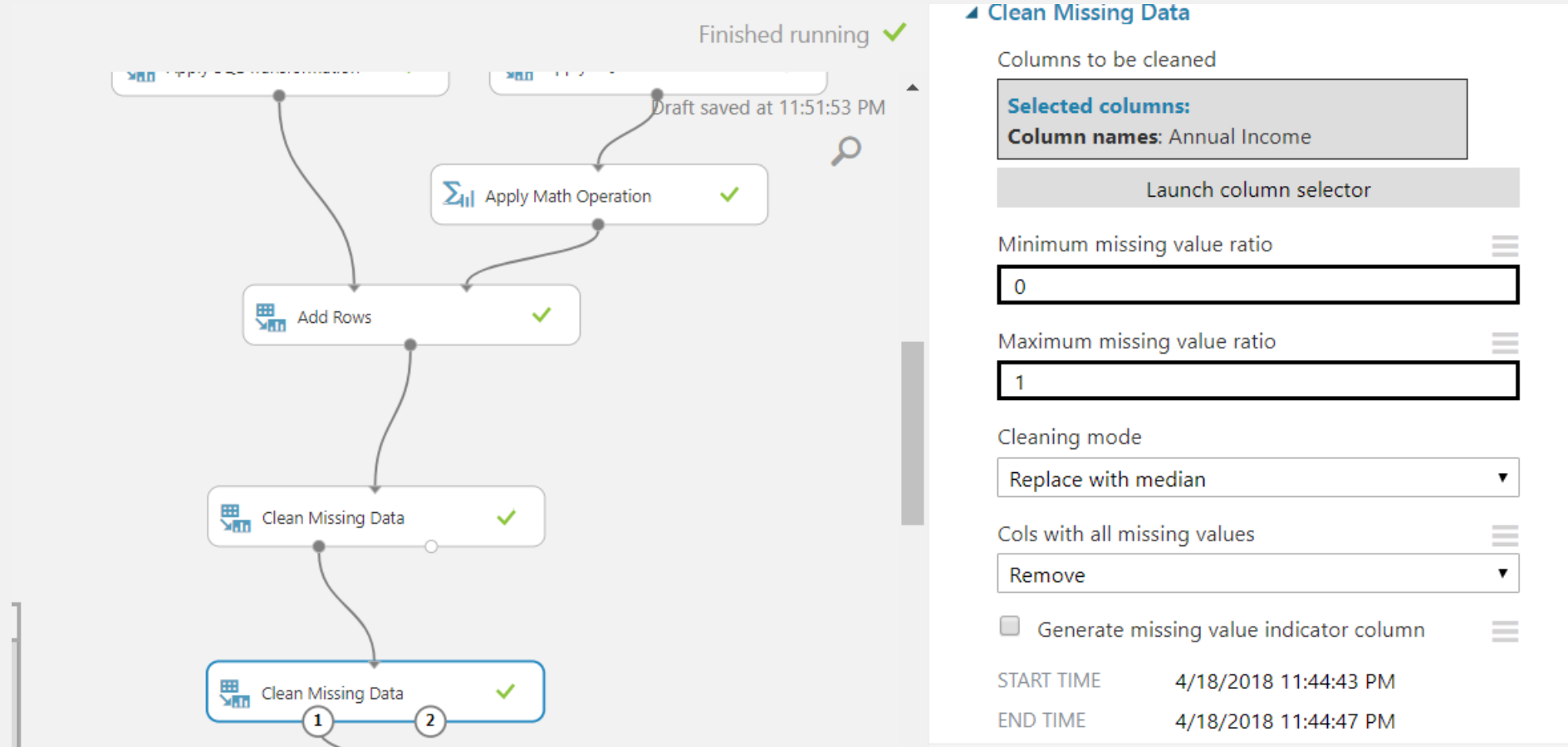
# Creating ML Models In Azure ML Studio



Microsoft



## (Handling “Annual Income”)



# Creating ML Models In Azure ML Studio

## (Handling “Annual Income”)

### Select columns

BY NAME

WITH RULES

AVAILABLE COLUMNS

All Types

search columns

Loan ID  
Customer ID  
Loan Status  
Current Loan Amount  
Term  
Credit Score  
Years in current job  
Home Ownership  
Purpose  
Monthly Debt  
Years of Credit History  
Months since last delinquent  
Number of Open Accounts  
Number of Credit Problems  
Current Credit Balance  
Maximum Open Credit

18 columns available

SELECTED COLUMNS

All Types

search columns

Annual Income

1 columns selected

>


<

✓


# Creating ML Models In Azure ML Studio

## (Handling “Annual Income”)

- Ensure to replace with Median.


Minimum missing value ra... 


0


Maximum missing value r... 


1

Cleaning mode

Replace with median 

Cols with all missing values 

Remove 

☐ Generate missing valu... 

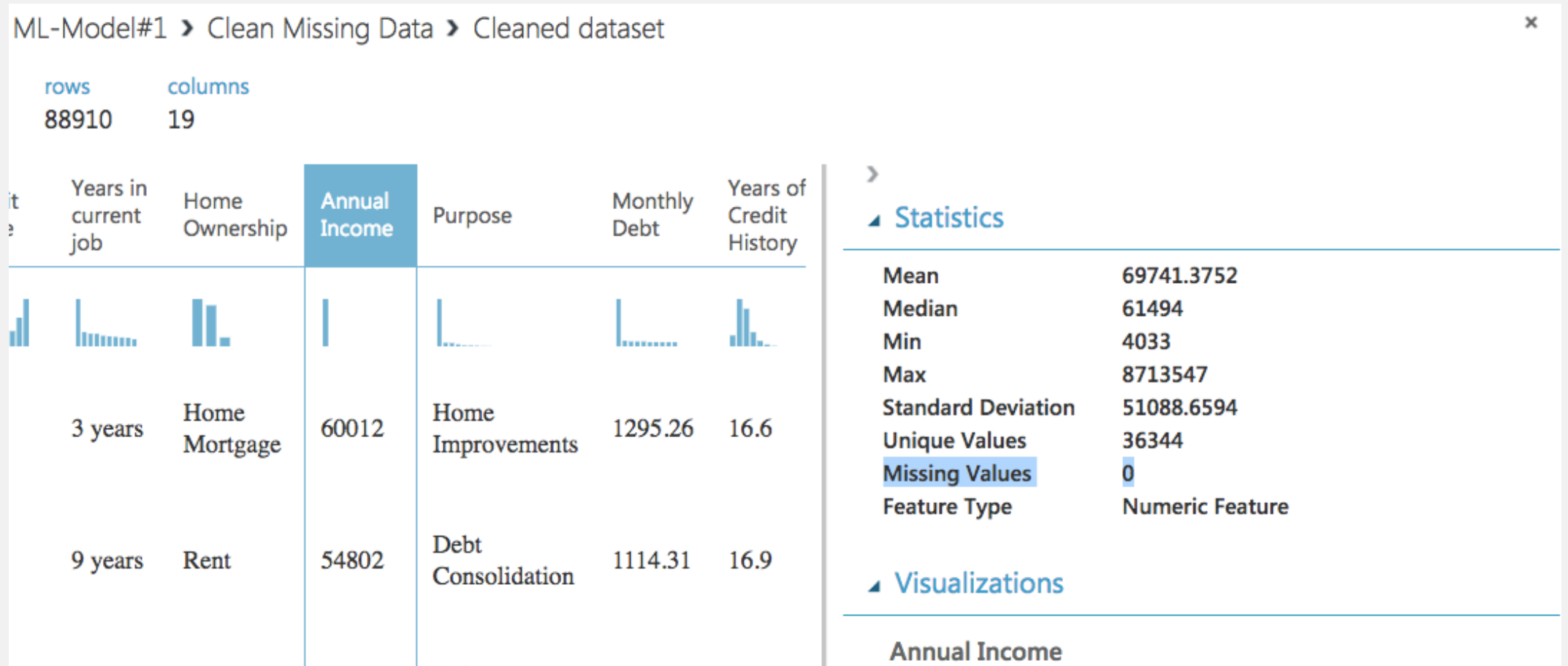
# Creating ML Models In Azure ML Studio



Microsoft



## (Handling “Annual Income”)



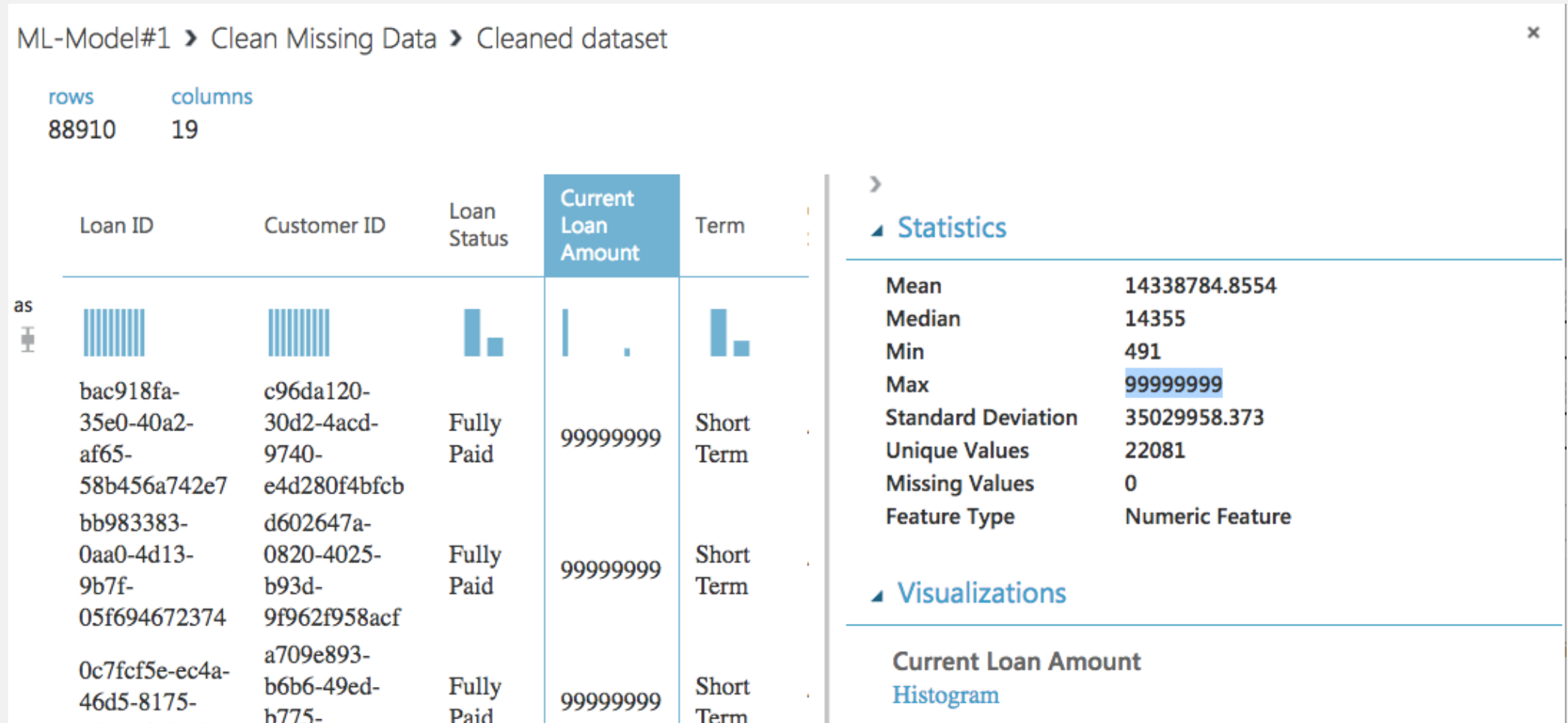
# Creating ML Models In Azure ML Studio



Microsoft



## (Handling “Current Loan Amount”)



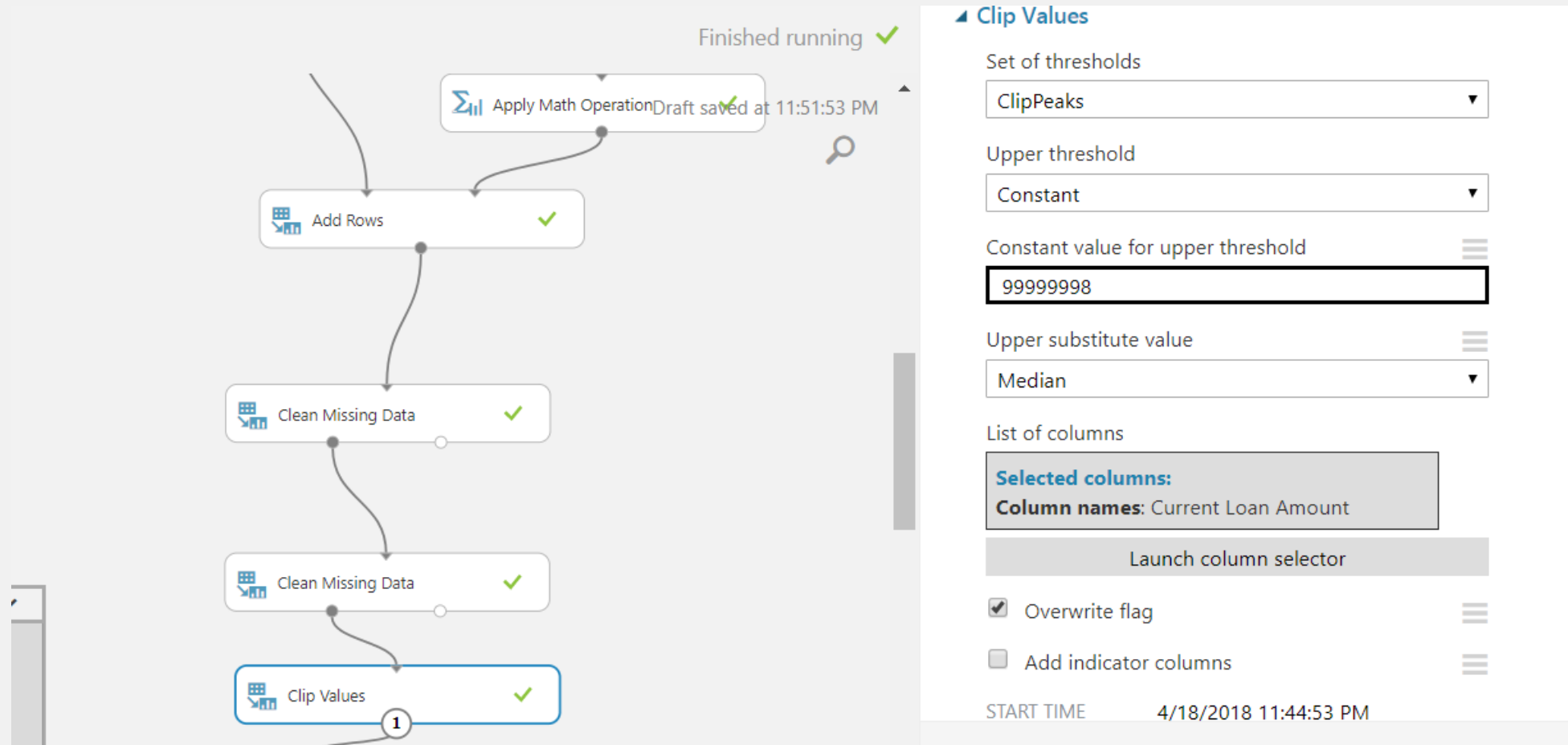
# Creating ML Models In Azure ML Studio



Microsoft



## (Handling “Current Loan Amount”)



# Creating ML Models In Azure ML Studio



Microsoft



## (Handling “Current Loan Amount”)

Select columns

BY NAME

WITH RULES

AVAILABLE COLUMNS

All Types

search columns

Loan ID

Customer ID

Loan Status

Term

Credit Score

Years in current job

Home Ownership

Annual Income

Purpose

Monthly Debt

Years of Credit History

Months since last delinquent

Number of Open Accounts

Number of Credit Problems

Current Credit Balance

Maximum Open Credit

18 columns available

SELECTED COLUMNS

All Types

search columns

Current Loan Amount

1 columns selected

>

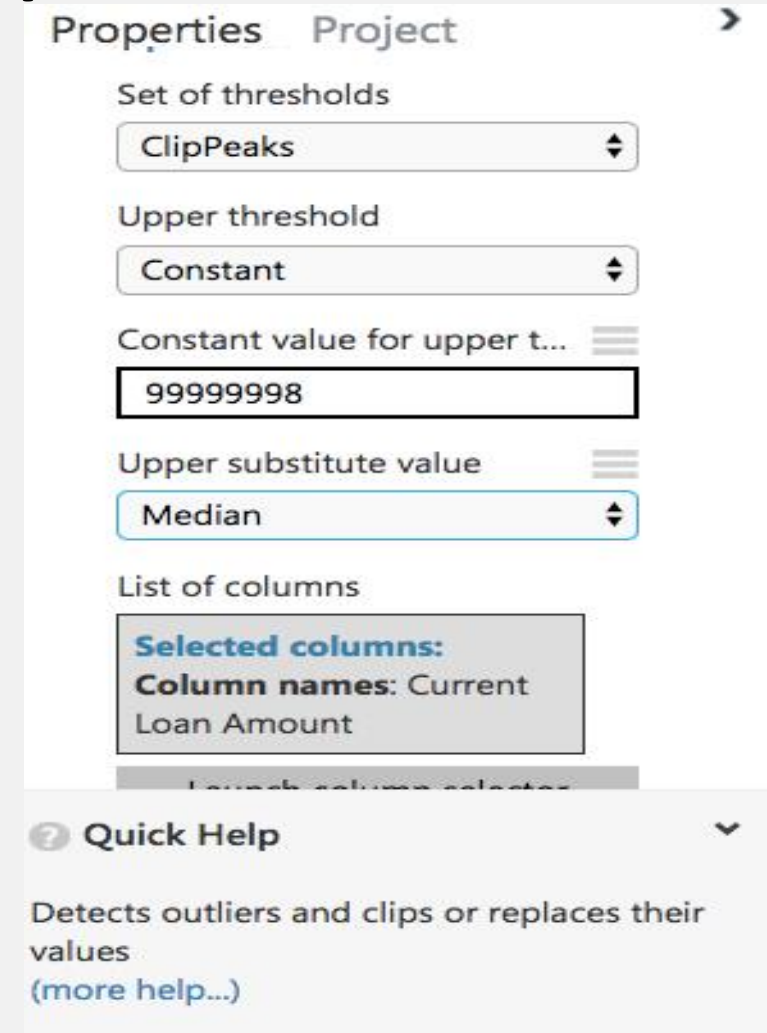
<

✓

# Creating ML Models In Azure ML Studio

## (Handling “Current Loan Amount”)

- Ensure the following properties.



The screenshot shows the 'Properties' tab of the 'ClipPeaks' module in Azure ML Studio. The settings are as follows:

- Set of thresholds:** ClipPeaks
- Upper threshold:** Constant
- Constant value for upper t...:** 99999998
- Upper substitute value:** Median
- List of columns:**
  - Selected columns:**
  - Column names:** Current Loan Amount

Below the properties panel, a 'Quick Help' section is visible, stating: 'Detects outliers and clips or replaces their values' with a link to '(more help...)'. A 'Launch column selector' button is partially visible at the bottom of the properties panel.



# Creating ML Models In Azure ML Studio



Microsoft



## (Handling “Current Loan Amount”)

ML-Model#1 > Clip Values > Results dataset

rows  
88910

columns  
19

ID	Customer ID	Loan Status	Current Loan Amount	Term	Credit Score	Year current job
8fa-40a2-56a742e73383-4d13-4672374f5e-ec4a-8175-16df7c74	c96da120-30d2-4acd-9740-e4d280f4bfcbb602647a-0820-4025-b93d-9f962f958acfa709e893-b6b6-49ed-b775-bdef1155c78d	Fully Paid	14355	Short Term	739	3 ye
		Fully Paid	14355	Short Term	704	9 ye
		Fully Paid	14355	Short Term	751	5 ye

### Statistics

Mean	13994.2439
Median	14355
Min	491
Max	35875
Standard Deviation	7624.5725
Unique Values	22080
Missing Values	0
Feature Type	Numeric Feature

### Visualizations

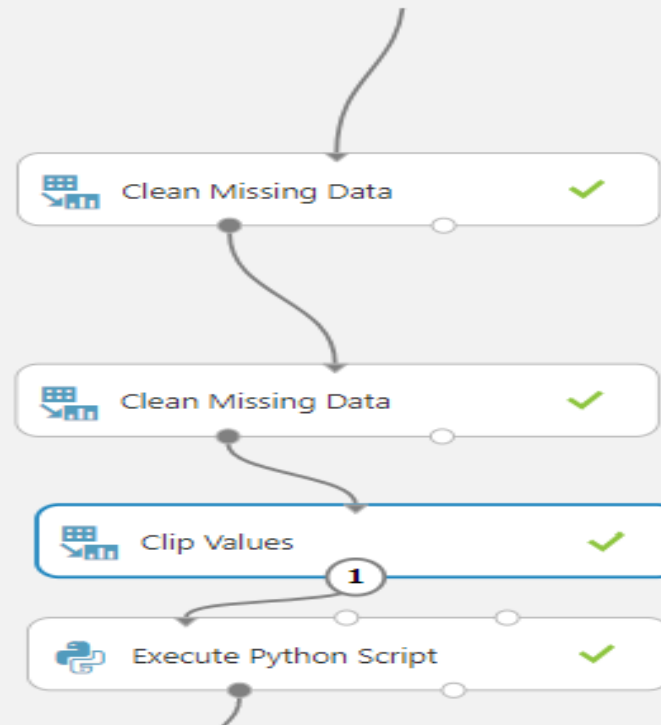
Current Loan Amount  
Histogram

# Creating ML Models In Azure ML Studio

## (Handling Features with Python Module)

Finished running

Draft saved at 11:51:53



# Creating ML Models In Azure ML Studio



Microsoft



## (Handling Features with Python Module)

Properties Project

### Execute Python Script

Python script

```
1 : The script MUST contain a function named azureml_main
2 : which is the entry point for this module.
3
4 : imports up here can be used to
5 import pandas as pd
6
7 : The entry point function can contain up to two input arguments:
8 :   Param<dataframe1>: a pandas.DataFrame
9 :   Param<dataframe2>: a pandas.DataFrame
10 def azureml_main(dataframe1 = None, dataframe2 = None):
11
12     # Execution logic goes here
13     print('Input pandas.DataFrame #1:\n\n{0}'.format(dataframe1))
14
15     # If a zip file is connected to the third input port is connected
16     # it is unzipped under ".\Script Bundle". This directory is added
```

### ? Quick Help

Executes a Python script from an Azure Machine Learning experiment  
([more help...](#))

# Creating ML Models In Azure ML Studio



Microsoft



## (Handling Features with Python Module)

```
def azureml_main(dataframe1 = None, dataframe2 = None):  
  
    dataframe1.loc[dataframe1['Months since last delinquent'] == 'NA', 'Months since last delinquent'] = "0"  
    dataframe1.loc[dataframe1['Bankruptcies'] == 'NA', 'Bankruptcies'] = "0"  
    dataframe1.loc[dataframe1['Tax Liens'] == 'NA', 'Tax Liens'] = "0"  
    dataframe1.loc[dataframe1['Home Ownership'] == 'HaveMortgage', 'Home Ownership'] = "Home Mortgage"  
    return dataframe1,
```

# Creating ML Models In Azure ML Studio



Microsoft



## (Handling Features with Python Module)

Properties Project

### Execute Python Script

Python script

```
3
4 # imports up here can be used to
5 import pandas as pd
6
7 # The entry point function can contain up to two input arguments
8 #   Param<dataframe1>: a pandas.DataFrame
9 #   Param<dataframe2>: a pandas.DataFrame
10 def azureml_main(dataframe1 = None, dataframe2 = None):
11
12     dataframe1.loc[dataframe1['Months since last delinquent'] > 12,
13                    dataframe1['Bankruptcies'] == 'NA', 'Bankruptcies'] = 'HaveMoreThan12Months'
14     dataframe1.loc[dataframe1['Tax Liens'] == 'NA', 'Tax Liens'] = 'HaveMoreThan12Months'
15     dataframe1.loc[dataframe1['Home Ownership'] == 'HaveMoreThan12Months', 'Home Ownership'] = 'HaveMoreThan12Months'
16     return dataframe1,
17
```








### ? Quick Help

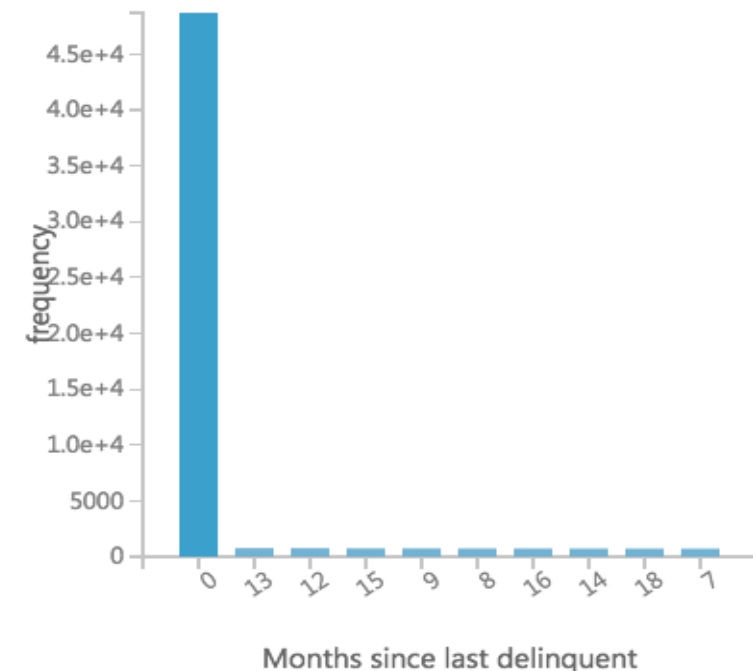
Executes a Python script from an Azure Machine Learning experiment  
([more help...](#))

# Creating ML Models In Azure ML Studio

## (Handling Features with Python Module)

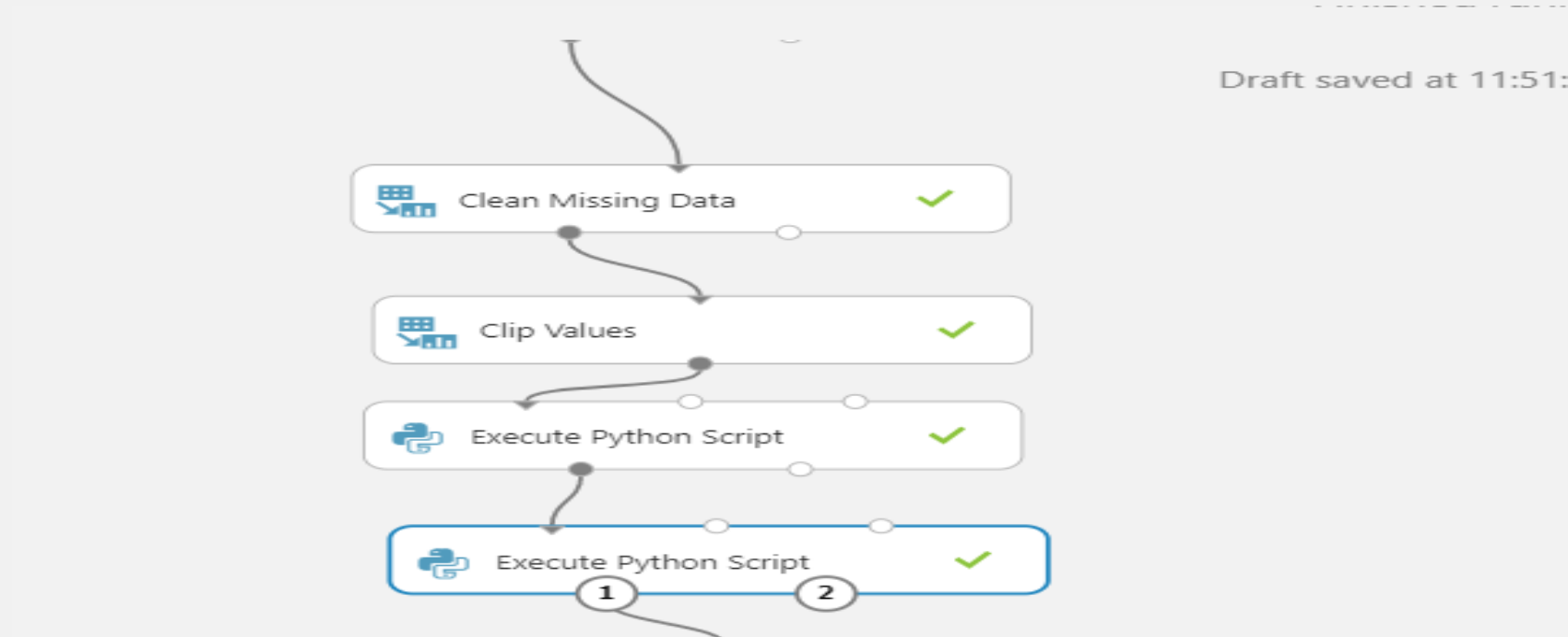
ML-Model#1 > Execute Python Script > Results dataset

rows		columns				
88910		19				
Monthly Debt	Years of Credit History	Months since last delinquent	Number of Open Accounts	Number of Credit Problems	Current Credit Balance	Maxim Open Credit
						
1295.26	16.6	33	16	0	427	8542
1114.31	16.9	30	12	0	4716	5822
626.12	11	0	7	0	742	9166



# Creating ML Models In Azure ML Studio

## (Handling Features with Python Module: Converting Data Type)



## (Handling Features with Python Module: Converting Data Type)

```
In [ ]: def azureml_main(dataframe1 = None, dataframe2 = None):  
    # dataframe1['Monthly Debt'] = dataframe1['Monthly Debt'].astype('float')  
    dataframe1['Monthly Debt'] = dataframe1['Monthly Debt'].convert_objects(convert_numeric=True)  
    dataframe1['Maximum Open Credit'] = dataframe1['Maximum Open Credit'].convert_objects(convert_numeric=True)  
    dataframe1['Months since last delinquent'] = dataframe1['Months since last delinquent'].convert_objects(convert_num  
  
    return dataframe1
```



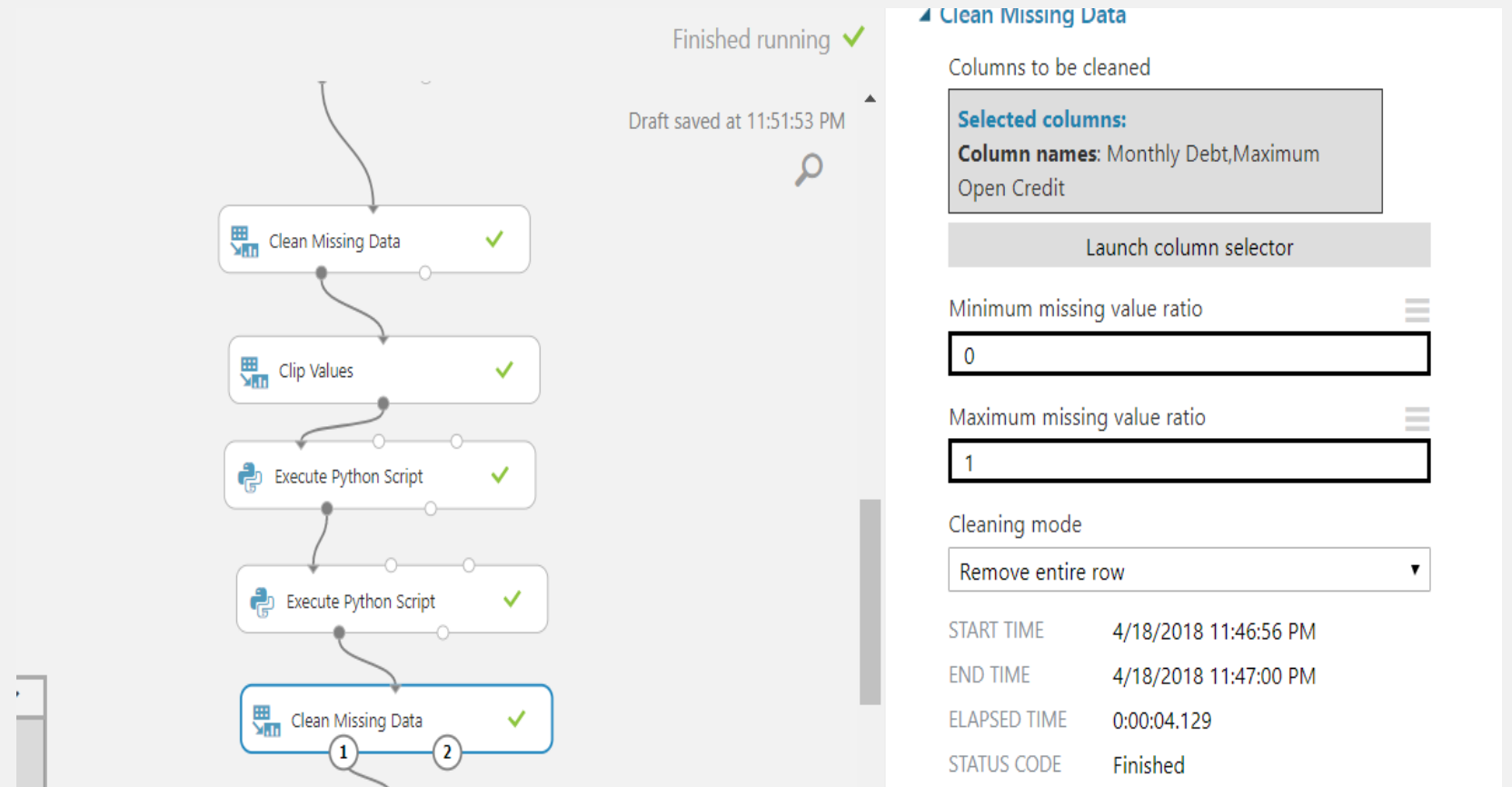
# Creating ML Models In Azure ML Studio

## (Handling Features with Python Module: Converting Data Type)

- Clean missing values created when converting to numeric values.

Finished running ✓

Draft saved at 11:51:53 PM



The screenshot displays the Azure ML Studio interface. On the left, a vertical workflow is shown with five steps: 'Clean Missing Data' (green checkmark), 'Clip Values' (green checkmark), 'Execute Python Script' (green checkmark), 'Execute Python Script' (green checkmark), and 'Clean Missing Data' (green checkmark). The first and last steps are highlighted with a blue border and numbered 1 and 2 respectively. On the right, a configuration panel for the 'Clean Missing Data' step is visible. It includes a 'Columns to be cleaned' section with 'Selected columns: Monthly Debt, Maximum Open Credit'. Below this is a 'Launch column selector' button. The 'Minimum missing value ratio' is set to 0, and the 'Maximum missing value ratio' is set to 1. The 'Cleaning mode' is set to 'Remove entire row'. At the bottom, a table shows the execution details: START TIME (4/18/2018 11:46:56 PM), END TIME (4/18/2018 11:47:00 PM), ELAPSED TIME (0:00:04.129), and STATUS CODE (Finished).

**Clean Missing Data**

Columns to be cleaned

**Selected columns:**  
Column names: Monthly Debt, Maximum Open Credit

Launch column selector

Minimum missing value ratio  
0

Maximum missing value ratio  
1

Cleaning mode  
Remove entire row

START TIME 4/18/2018 11:46:56 PM  
END TIME 4/18/2018 11:47:00 PM  
ELAPSED TIME 0:00:04.129  
STATUS CODE Finished

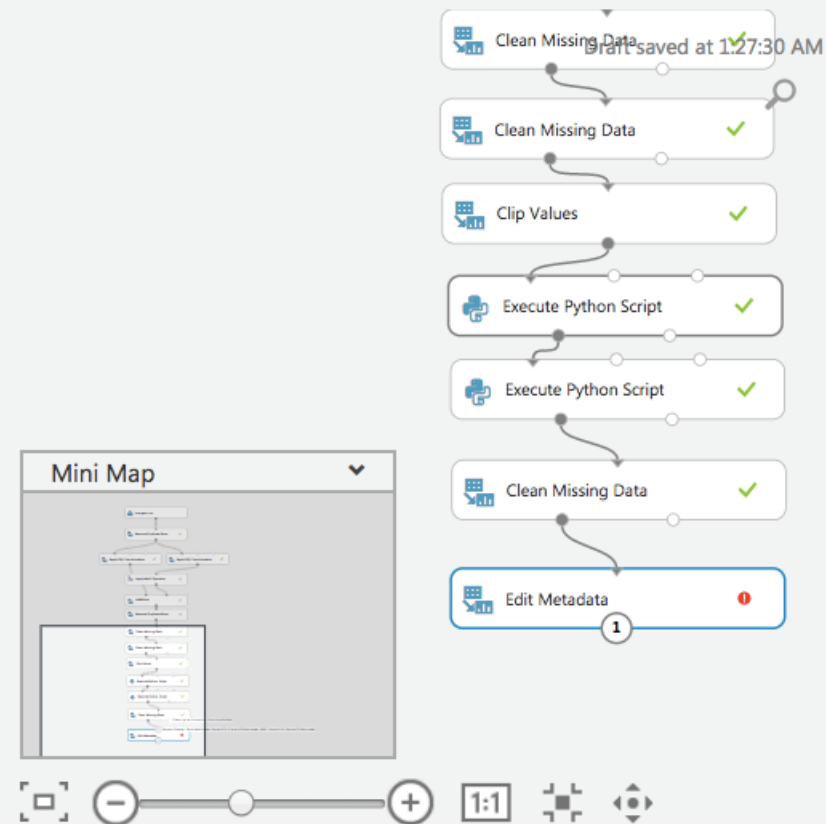
# Creating ML Models In Azure ML Studio



## (Handling Features with “Edit Metadata”)

ML-Model#1

In draft



Properties Project

### Edit Metadata

Column

#### Selected columns:

Launch the selector tool to make a selection

Launch column selector

Data type

Unchanged

Categorical

Unchanged

Fields

Unchanged

### Quick Help

Edits metadata associated with columns in a dataset. Formerly known as Metadata Editor.  
[\(more help...\)](#)

# Creating ML Models In Azure ML Studio

## (Handling Features with “Edit Metadata”)

Select columns

BY NAME

WITH RULES

AVAILABLE COLUMNS

All Types search columns

Loan ID  
Customer ID  
Current Loan Amount  
Credit Score  
Annual Income  
Monthly Debt  
Years of Credit History  
Months since last delinquent  
Number of Open Accounts  
Number of Credit Problems  
Current Credit Balance  
Maximum Open Credit

12 columns available

SELECTED COLUMNS

All Types search columns

Loan Status  
Term  
Years in current job  
Home Ownership  
Bankruptcies  
Tax Liens  
Purpose

7 columns selected

>

<

✓

# Creating ML Models In Azure ML Studio

## (Handling Features with “Edit Metadata”)

- Ensure the following properties.

Properties Project >

Status,Term,Years in current job,Home Ownership,Bankruptcies,Tax Liens,Purpose

Launch column selector

Data type

Unchanged

Categorical

Make categorical

Fields

Unchanged

New column names

# Creating ML Models In Azure ML Studio

## (Handling “Years In Current Job”)

- Using “Grouping Categorical Values” component.

ML-Model#1

In draft

Draft saved at 12:58:44 PM

Clean Missing Data ✓

Clean Missing Data ✓

Clip Values ✓

Execute Python Script ✓

Execute Python Script ✓

Clean Missing Data ✓

Edit Metadata ✓

Results dataset (Dataset)

Group Categorical Values 1

Mini Map

Properties Project

Group Categorical Values

Selected columns

Selected columns: Column type: Categorical, All

Launch column selector

Output mode

ResultOnly

Default level name

New number of levels

1

Quick Help

Groups data from multiple categories into a new category (more help...)

# Creating ML Models In Azure ML Studio

## (Handling “Years In Current Job”)

### Select columns

BY NAME

WITH RULES

AVAILABLE COLUMNS

All Types search columns

Loan ID  
Customer ID  
Loan Status  
Current Loan Amount  
Term  
Credit Score  
Home Ownership  
Annual Income  
Purpose  
Monthly Debt  
Years of Credit History  
Months since last delinquent  
Number of Open Accounts  
Number of Credit Problems  
Current Credit Balance  
Maximum Open Credit

18 columns available

SELECTED COLUMNS

All Types search columns

Years in current job

1 columns selected

>

<

✓

# Creating ML Models In Azure ML Studio

## (Handling “Years In Current Job”)

Properties Project >

▲ Group Categorical Values

Selected columns

**Selected columns:**  
Column names: Years in current job

Launch column selector

Output mode

Default level name

New number of levels

Name of new level 1

Comma-separated list of old l...

Name of new level 2

Comma-separated list of old l...

Name of new level 3

Comma-separated list of old l...

Name of new level 4

Comma-separated list of old l...

Name of new level 4

Comma-separated list of old l...

Name of new level 5

Comma-separated list of old l...

# Creating ML Models In Azure ML Studio



Microsoft



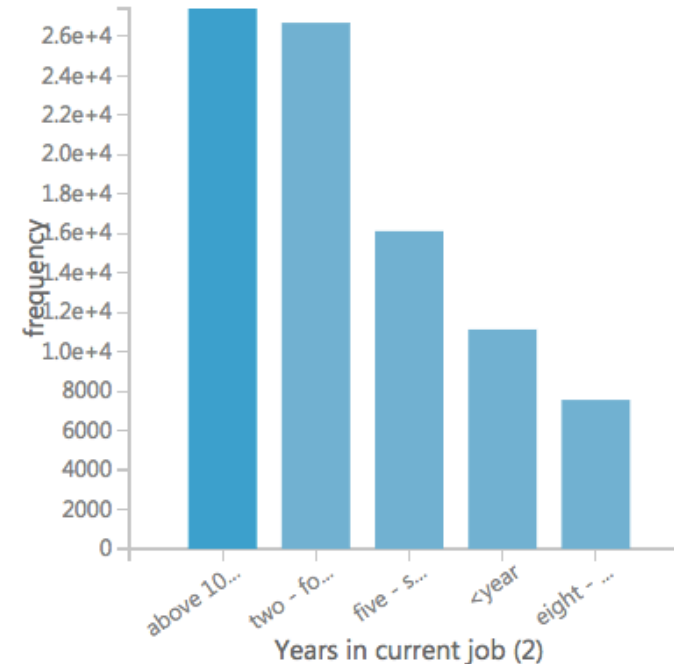
## (Handling “Years In Current Job”)

ML-Model#1 > Group Categorical Values > Results dataset

rows  
88902

columns  
20

Number of Open Accounts	Number of Credit Problems	Current Credit Balance	Maximum Open Credit	Bankruptcies	Tax Liens	Years in current job (2)
0	0	427	8542	0	0	two - four years
0	0	4716	5822	0	0	eight - nine years
0	0	742	9166	0	0	five - seven years

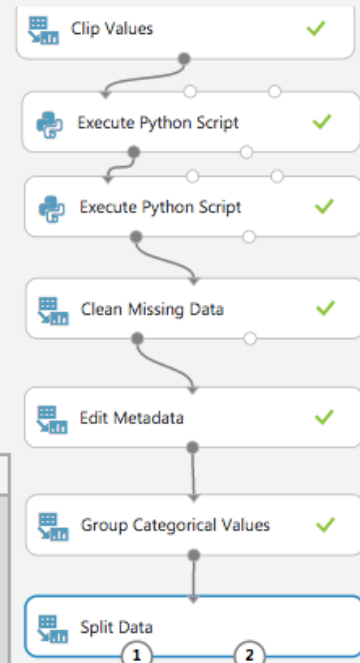




# Creating ML Models In Azure ML Studio (Splitting Data)

- Splitting data into Test and Training dataset.
- Split on the basis of 70% training to 30% testing dataset.

ML-Model#1 In draft

  
Clip Values ✓  
Execute Python Script ✓  
Execute Python Script ✓  
Clean Missing Data ✓  
Edit Metadata ✓  
Group Categorical Values ✓  
Split Data 1 2

Mini Map

Properties Project

**Split Data**

Splitting mode  
Split Rows

Fraction of rows in the first o...  
0.7

☒ Randomized split

Random seed  
7688

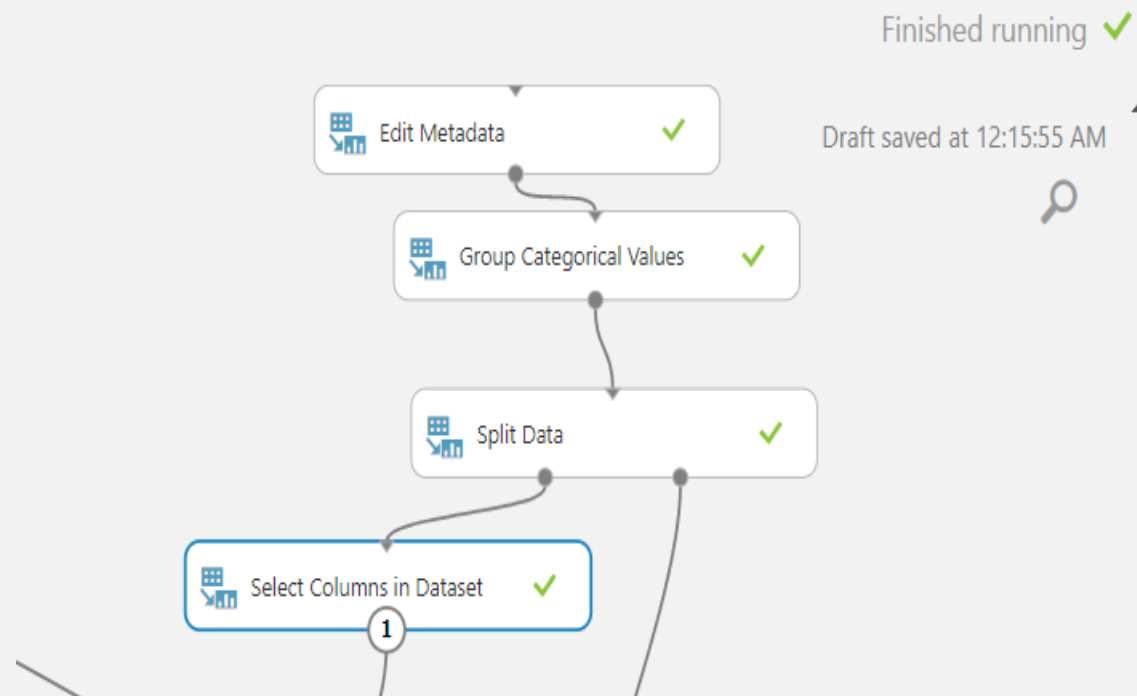
Stratified split  
False

**Quick Help**

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

# Creating ML Models In Azure ML Studio

## (Selecting Columns in a Dataset)



### Select Columns in Dataset

Select columns

#### Selected columns:

**Column names:** Loan Status,Current Loan Amount,Term,Credit Score,Home Ownership,Annual Income,Purpose,Monthly Debt,Years of Credit History,Months since last delinquent,Number of Open Accounts,Number of Credit Problems,Current Credit Balance,Maximum Open Credit,Bankruptcies,Tax Liens,Years in current job (2)

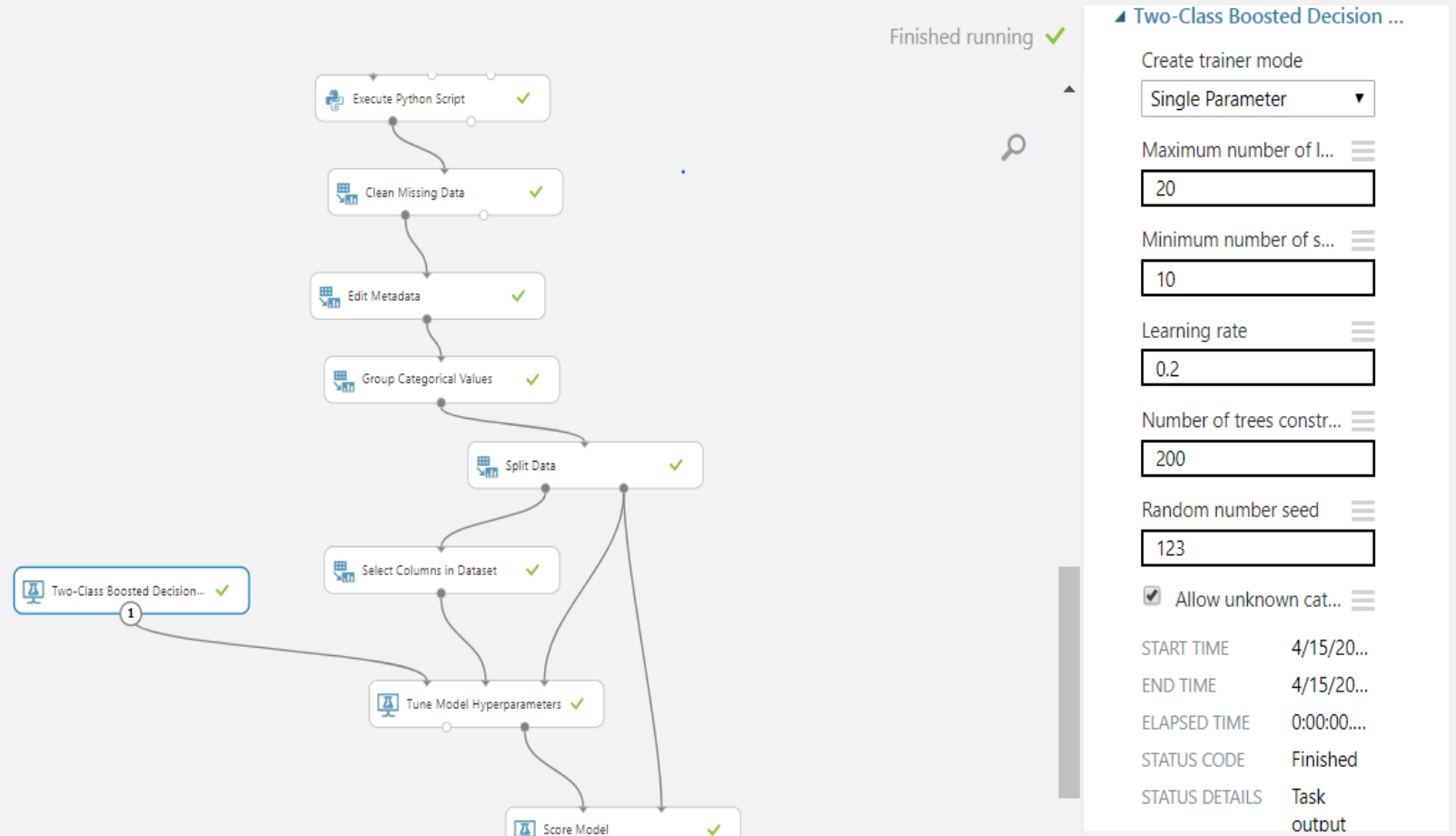
Launch column selector

START TIME 4/19/2018 12:12:02 AM

# Creating ML Models In Azure ML Studio

## (Selecting ML Algorithm)

- Using ML model of “Two Class Boosted Decision Tree”.

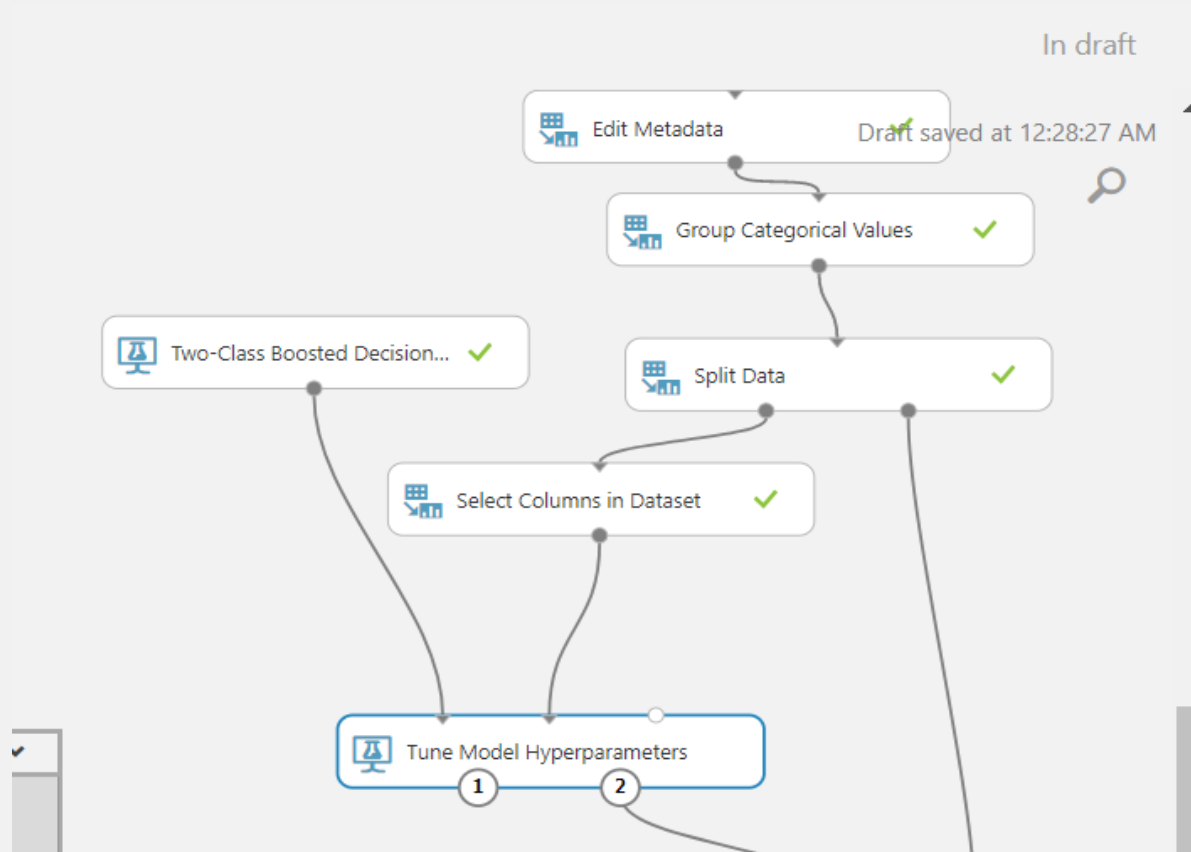


# Creating ML Models In Azure ML Studio

## (Add Hypertuning Parameters)

In draft

Draft saved at 12:28:27 AM



```
graph TD; EditMetadata[Edit Metadata] --> GroupCategorical[Group Categorical Values]; GroupCategorical --> SplitData[Split Data]; SplitData --> SelectColumns[Select Columns in Dataset]; SplitData --> TuneModel[Tune Model Hyperparameters]; SelectColumns --> TuneModel; TwoClassBoosted[Two-Class Boosted Decision...] --> TuneModel; TuneModel --> End[ ];
```

**Tune Model Hyperparameters**

Specify parameter sweeping mode

Random sweep

Maximum number of runs on random sweep

25

Random seed

0

Label column

**Selected columns:**

**Column names:** Loan Status

Launch column selector

Metric for measuring performance for classification

Accuracy

Metric for measuring performance for regression

Mean absolute error

# Creating ML Models In Azure ML Studio (Evaluation Metrics)

- Confusion Matrix

		Predicted class	
		Class 1	Class 0
Actual class	Class 1	10 true positives (TP)	2 false negatives (FN)
	Class 0	3 false positives (FP)	35 true negatives (TN)

- In this example of a confusion matrix, among the 50 data points that are classified, 45 are correctly classified and the 5 are misclassified.

# Creating ML Models In Azure ML Studio



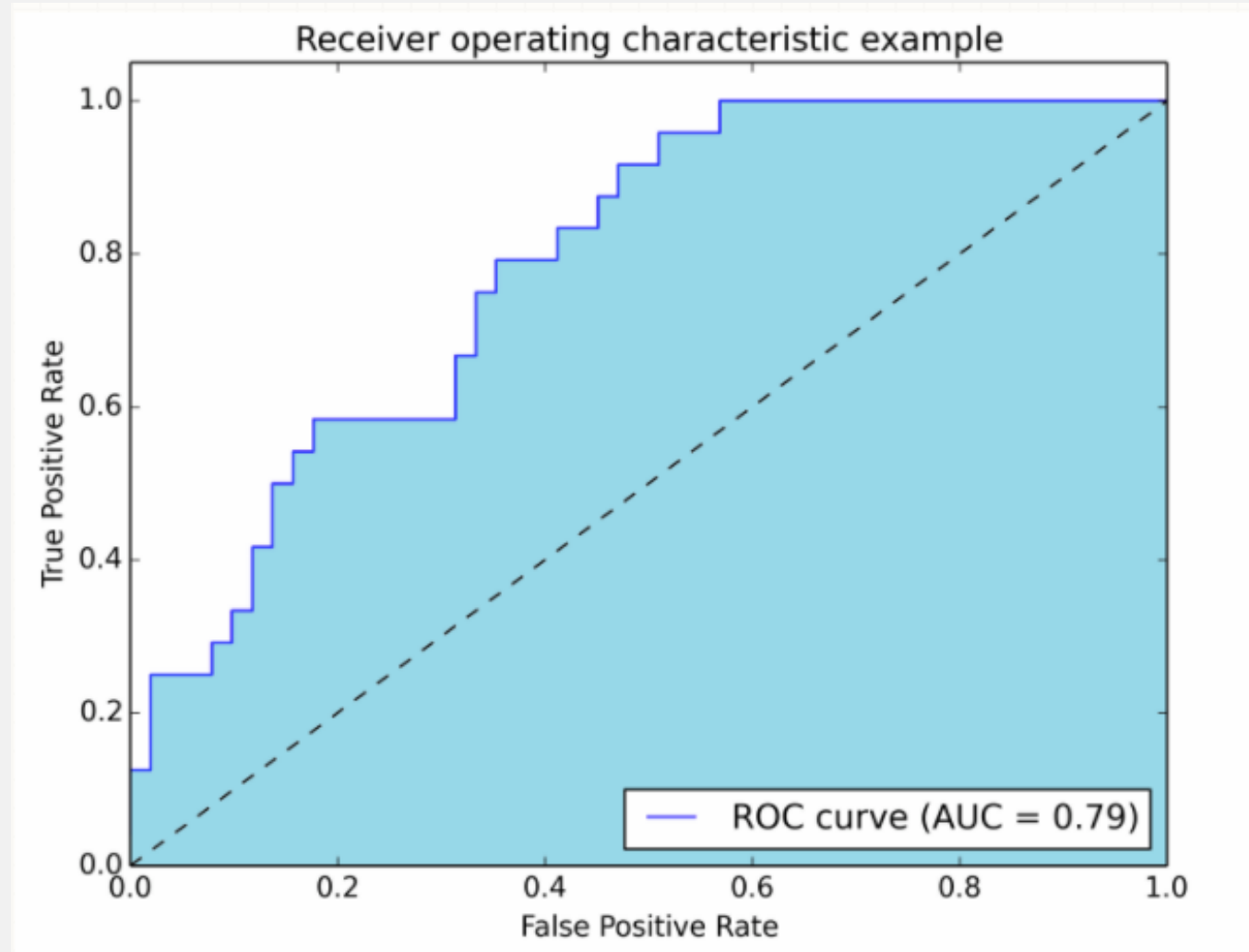
## (Evaluation Metrics)

- True Positive Rate (**TPR**), known as sensitivity, which is defined as  $TP / (TP + FN)$ .
- False Positive Rate (**FPR**), which is defined as  $FP / (FP + TN)$ .

# Creating ML Models In Azure ML Studio

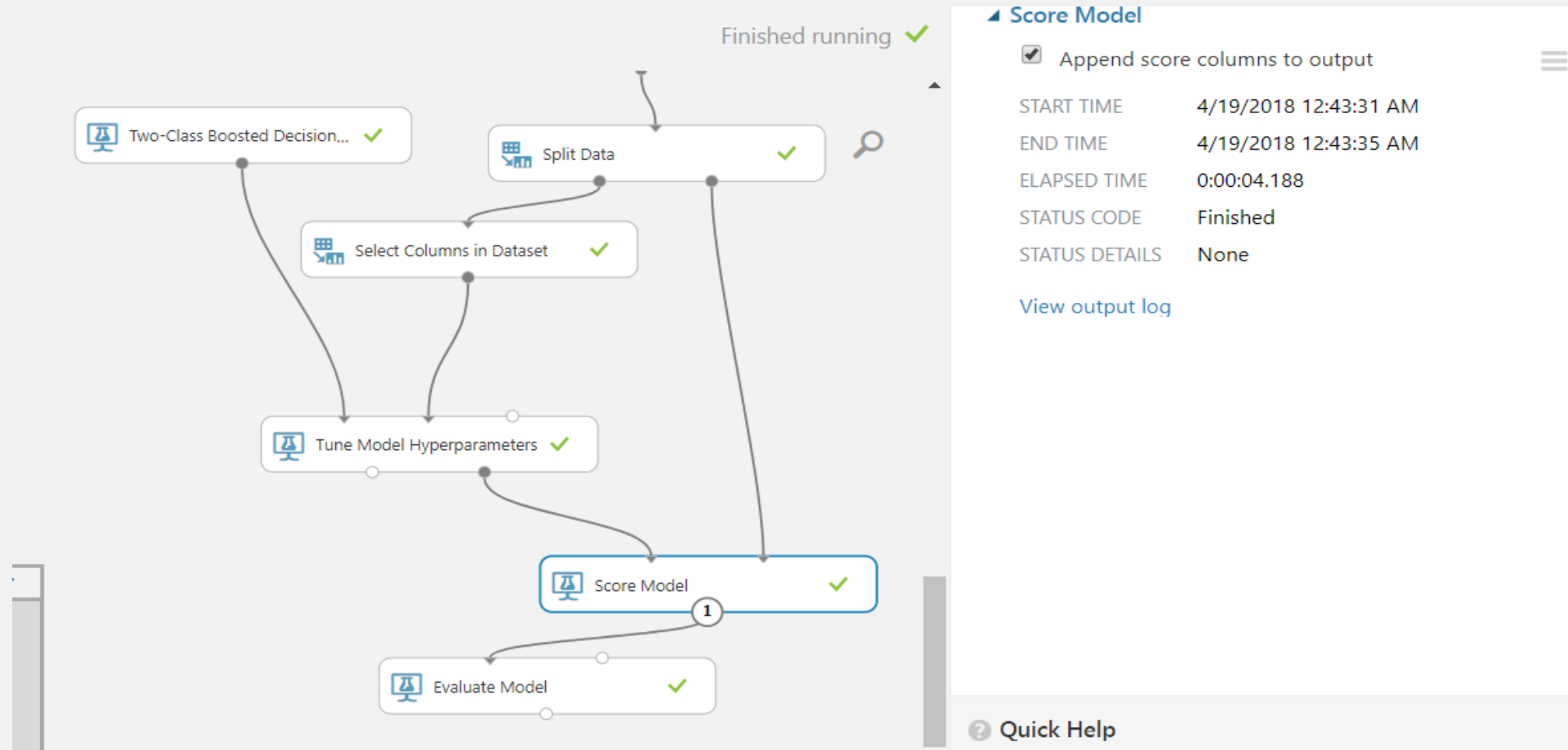


## (Evaluation Metrics)



# Creating ML Models In Azure ML Studio

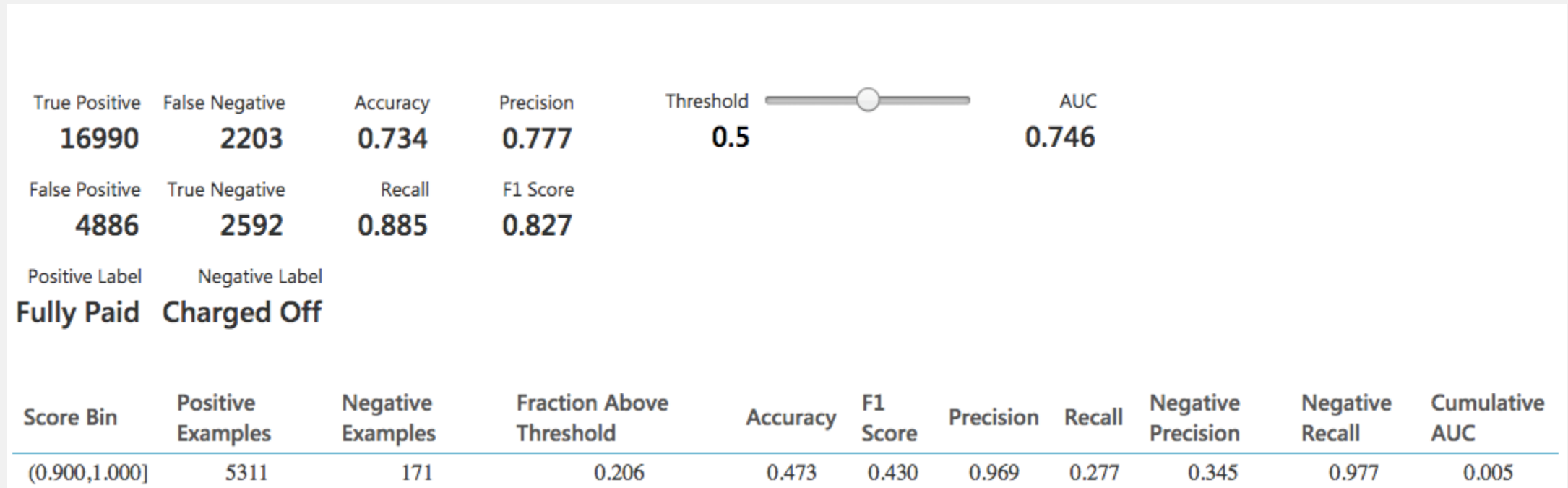
## (Score & Evaluate Model)





# Creating ML Models In Azure ML Studio

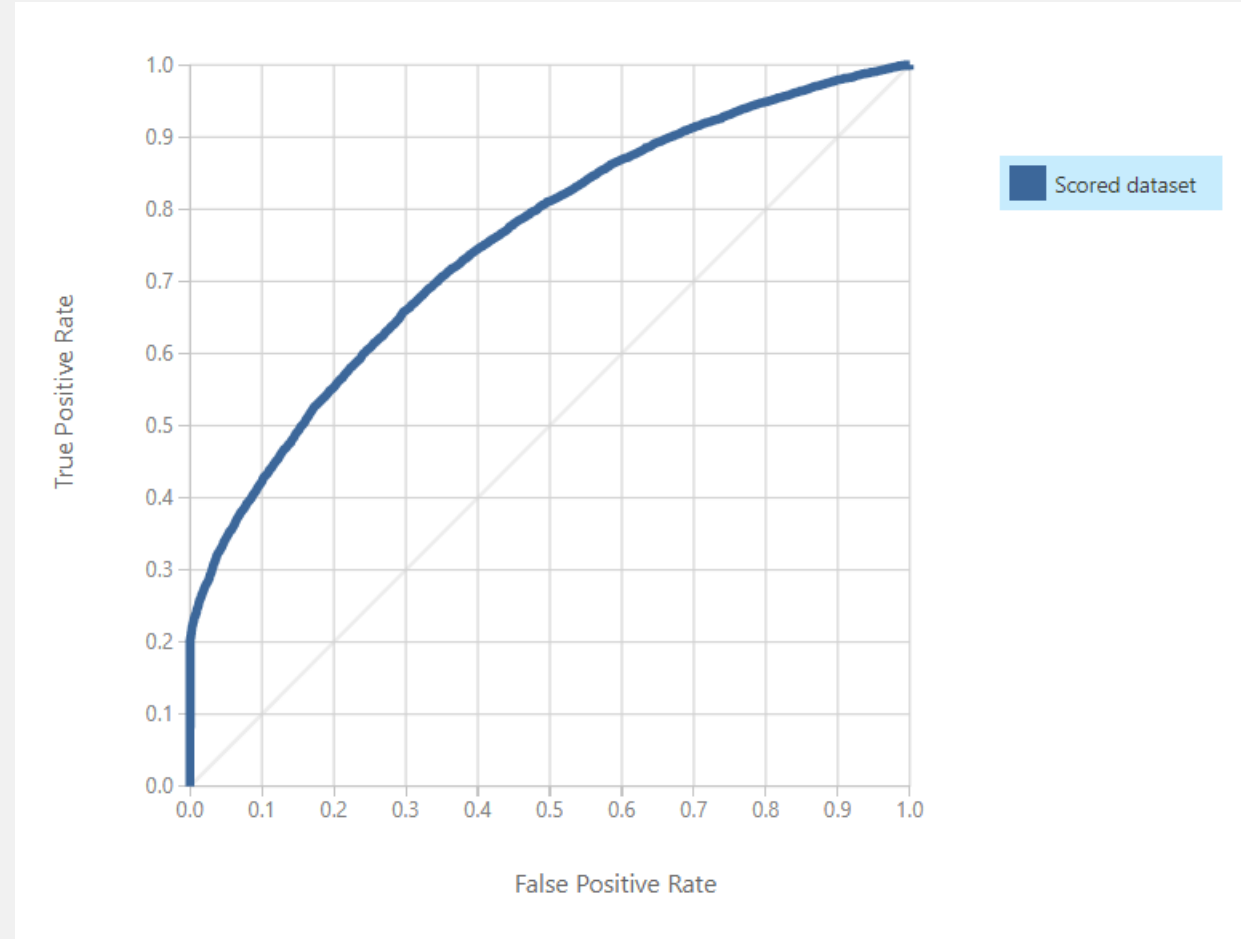
## (Score & Evaluate Model)



# Creating ML Models In Azure ML Studio



- **ROC Curve**



# END OF MODULE 4