Microsoft

Kosárelhagyók (Machine Learning)

Szávuly Krisztián

Data Science

MICROSOFT PROFESSIONAL PROGRAM

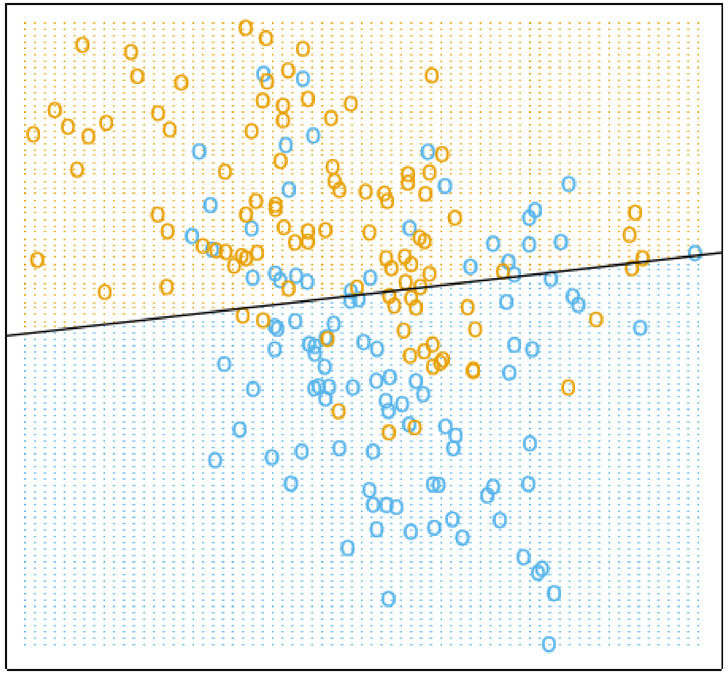
Konverzió

Machine Learning

Azure
Machine Learning
Studio

$$\{x_1, x_2, x_3, \ldots x_n\} \rightarrow y$$

$$f(\textstyle\sum x_n) \rightarrow \hat{y}$$



Cart Abandonment › E-Commerce Dataset › dataset

rows	columns
54251	13

view as

Column 0	X	SessionID	ItemID	Price	Quantity	Buy	Category	Year	Month	Day	Hour	Minute
9005	9005	1512504	214833755	523	1	1	0	2014	4	24	7	22
86949	86949	10124591	214853285	2093	1	1	1	2014	9	8	19	24
72456	72456	690696	214580387	6282	1	1	0	2014	4	12	13	31
28942	28942	940977	214748297	3350	1	1	0	2014	4	16	9	41
28837	28837	9780244	214712387	7643	1	1	2	2014	9	4	19	29
15547	15547	10718534	214853165	1151	1	1	7	2014	9	10	20	39
22926	22926	11100073	214853070	313	1	1	2	2014	9	20	18	22
13879	13879	820846	214601197	20839	1	1	0	2014	4	13	18	28
24755	24755	9964988	214856248	131	6	1	8	2014	9	6	11	25
35224	35224	1615967	214574777	7852	1	1	0	2014	4	23	22	3
97306	97306	10803463	214553837	292	1	1	3	2014	9	21	12	28
14540	14540	548052	214834865	523	1	1	0	2014	4	8	16	8
1580	1580	1546673	214829366	1360	1	1	0	2014	4	27	8	44
12994	12994	11376943	214859882	837	1	1	S	2014	9	27	13	35
31504	31504	9208924	214853709	837	1	1	S	2014	8	31	15	50
49921	49921	10086871	214853105	1046	1	1	S	2014	9	8	19	22
11433	11433	9125539	214853210	418	1	1	S	2014	8	25	12	43
70382	70382	9662634	214678376	1046	1	1	S	2014	8	31	8	59

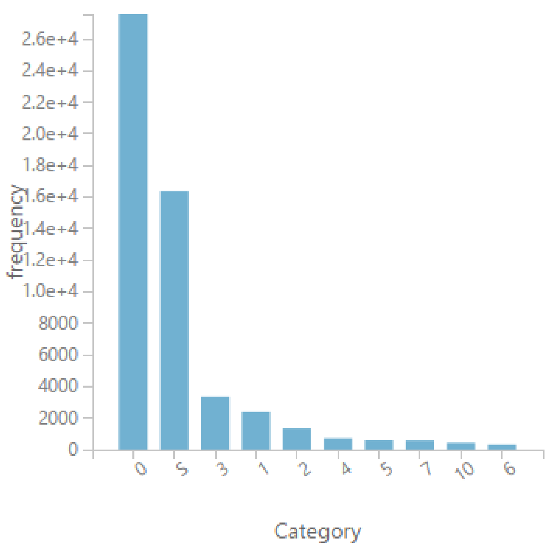
Statistics

Unique Values	49
Missing Values	0
Feature Type	String Feature

Visualizations

Category

Histogram



Cart Abandonment › E-Commerce Dataset › dataset

rows 54251
columns 13

view as

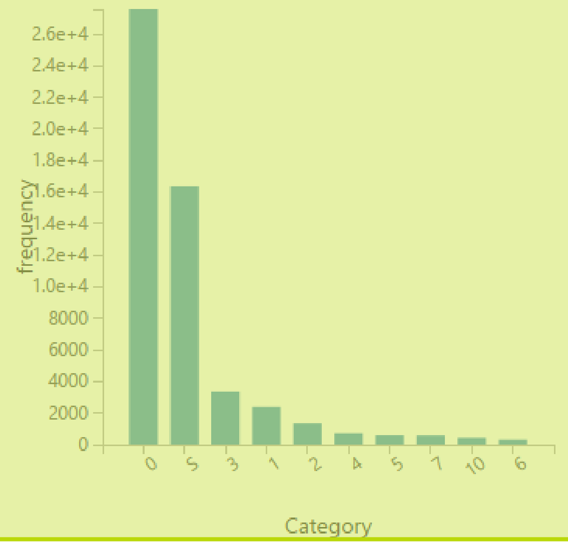
Column 0	X	SessionID	ItemID	Price	Quantity	Buy	Category	Year	Month	Day	Hour	Minute
9005	9005	1512504	214833755	523	1	1	0	2014	4	24	7	22
86949	86949	10124591	214853285	2093	1	1	1	2014	9	8	19	24
72456	72456	690696	214580387	6282	1	1	0	2014	4	12	13	31
28942	28942	940977	214748297	3350	1	1	0	2014	4	16	9	41
28837	28837	9780244	214712387	7643	1	1	2	2014	9	4	19	29
15547	15547	10718534	214853165	1151	1	1	7	2014	9	10	20	39
22926	22926	11100073	214853070	313	1	1	2	2014	9	20	18	22
13879	13879	820846	214601197	20839	1	1	0	2014	4	13	18	28
24755	24755	9964988	214856248	131	6	1	8	2014	9	6	11	25
35224	35224	1615967	214574777	7852	1	1	0	2014	4	23	22	3
97306	97306	10803463	214553837	292	1	1	3	2014	9	21	12	28
14540	14540	548052	214834865	523	1	1	0	2014	4	8	16	8
1580	1580	1546673	214829366	1360	1	1	0	2014	4	27	8	44
12994	12994	11376943	214859882	837	1	1	S	2014	9	27	13	35
31504	31504	9208924	214853709	837	1	1	S	2014	8	31	15	50
49921	49921	10086871	214853105	1046	1	1	S	2014	9	8	19	22
11433	11433	9125539	214853210	418	1	1	S	2014	8	25	12	43
70382	70382	9662634	214678376	1046	1	1	S	2014	8	31	8	59

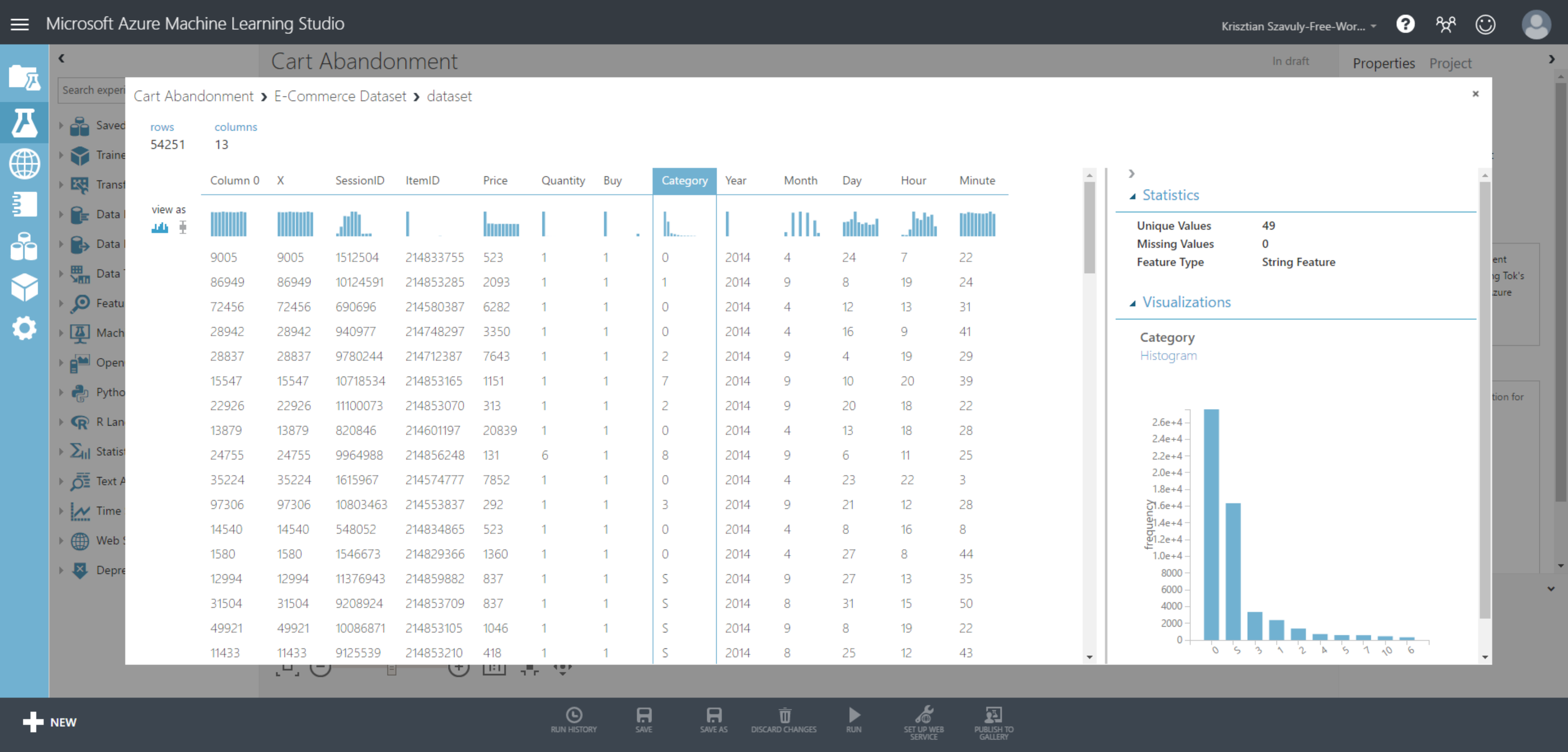
Statistics

Unique Values	49
Missing Values	0
Feature Type	String Feature

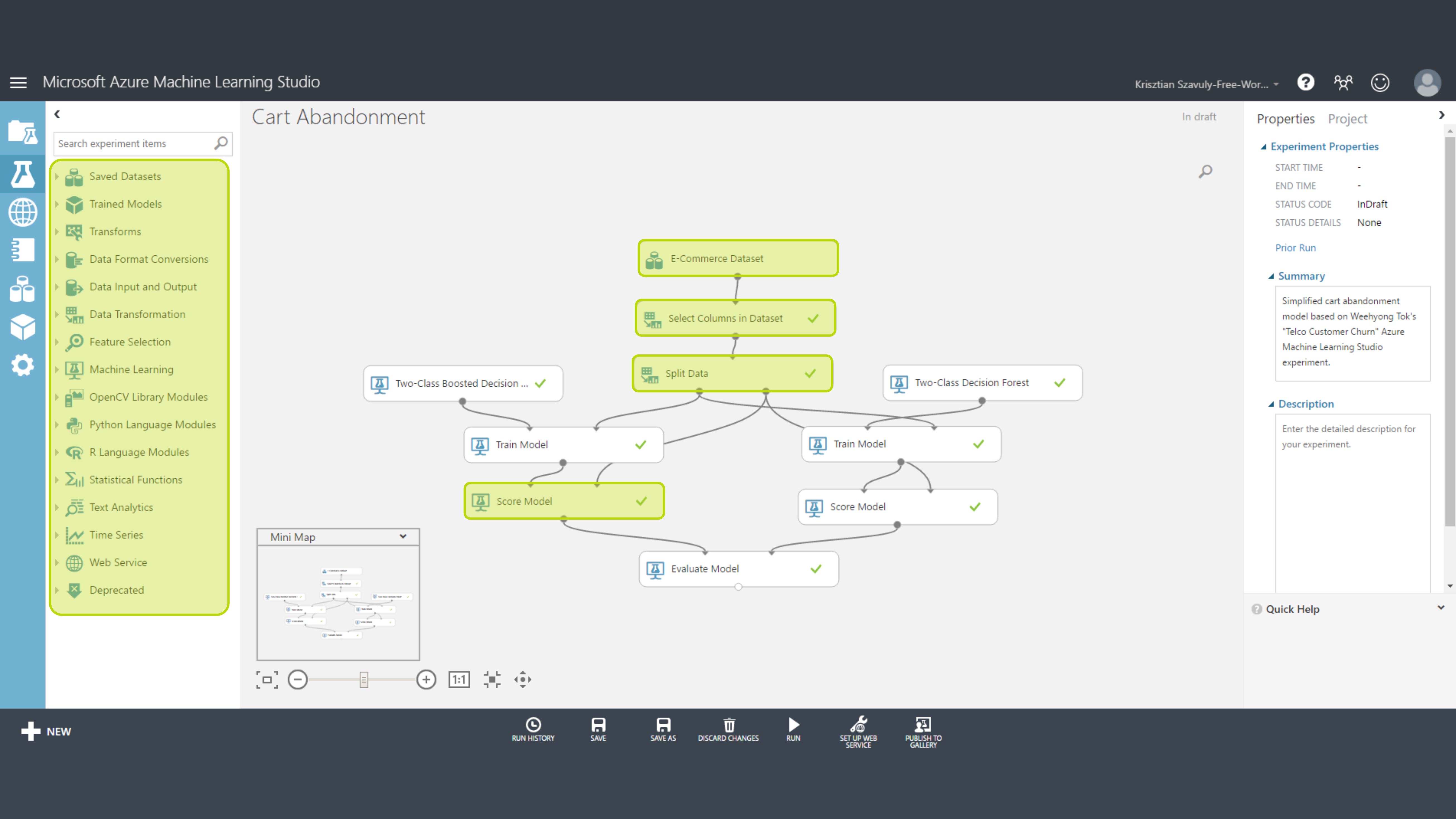
Visualizations

Category
Histogram

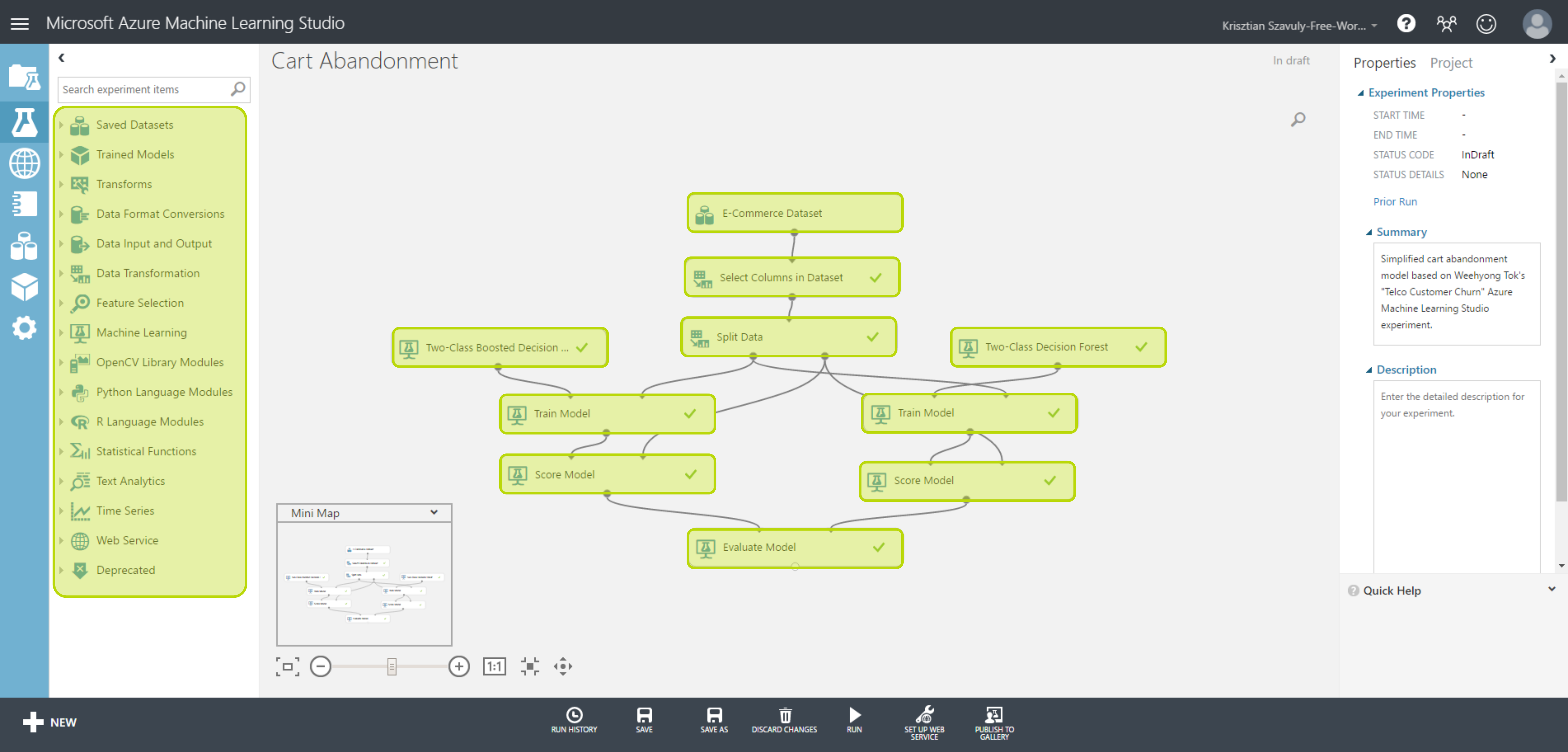


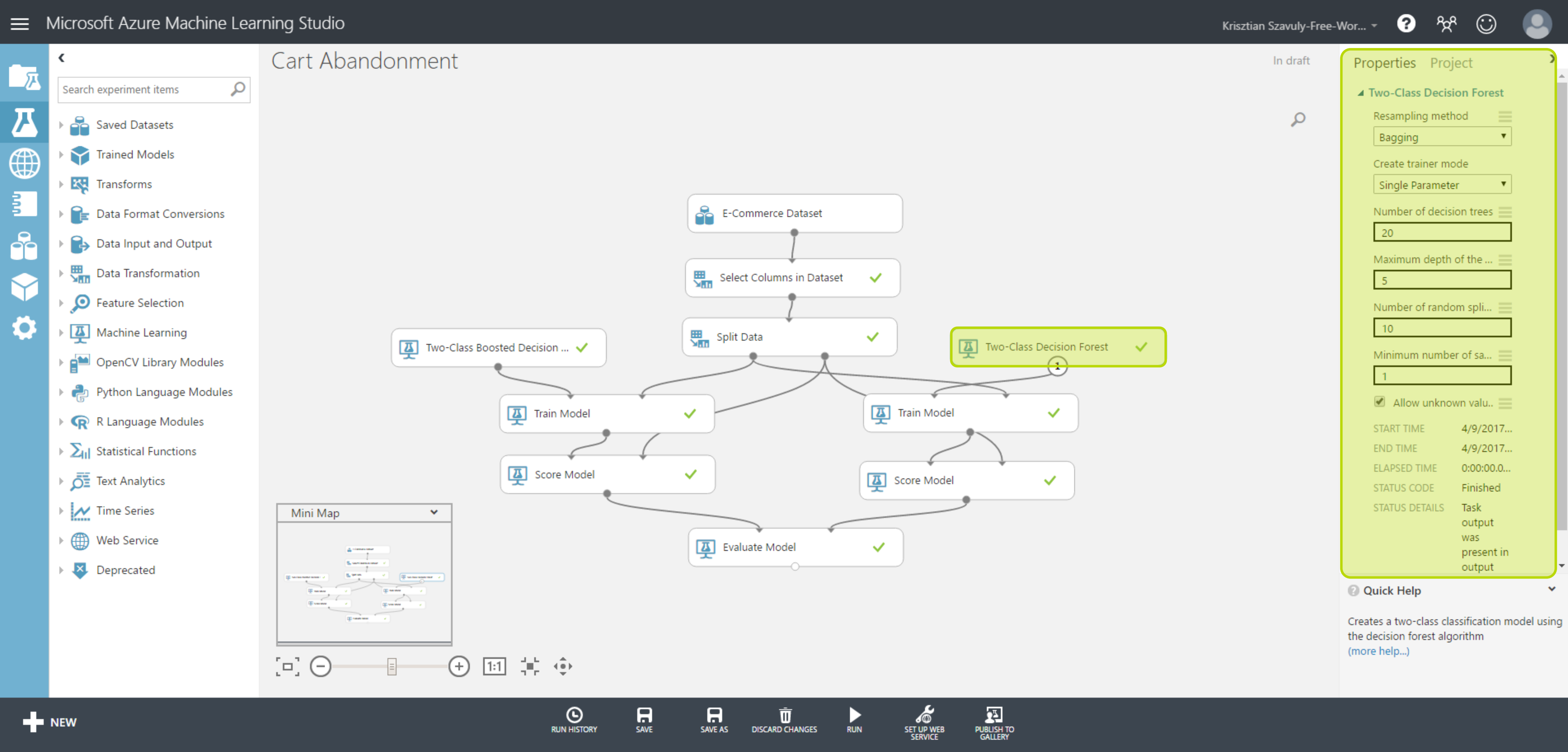


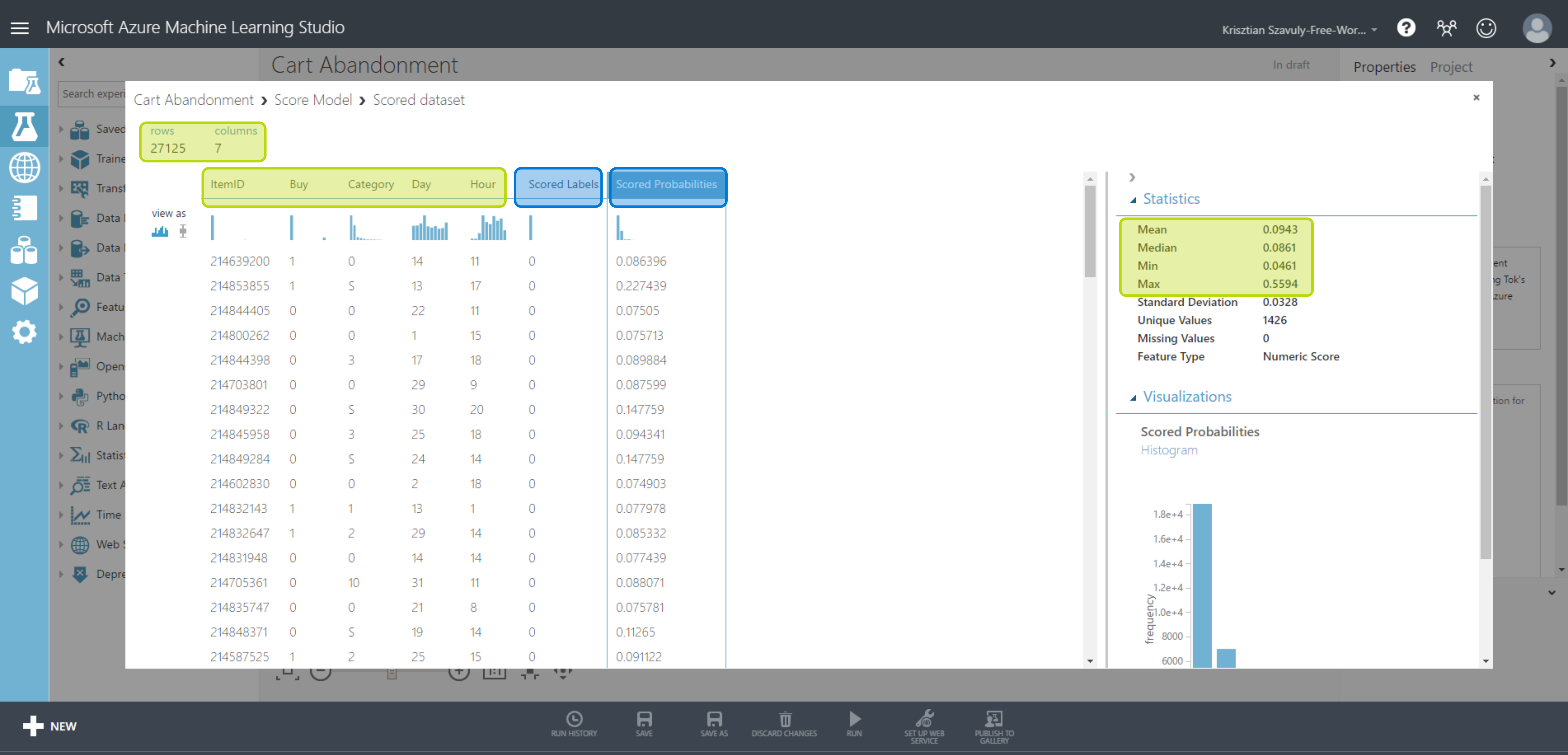
Column 0	X	SessionID	ItemID	Price	Quantity	Buy	Category	Year	Month	Day	Hour	Minute
9005	9005	1512504	214833755	523	1	1	0	2014	4	24	7	22
86949	86949	10124591	214853285	2093	1	1	1	2014	9	8	19	24
72456	72456	690696	214580387	6282	1	1	0	2014	4	12	13	31
28942	28942	940977	214748297	3350	1	1	0	2014	4	16	9	41
28837	28837	9780244	214712387	7643	1	1	2	2014	9	4	19	29
15547	15547	10718534	214853165	1151	1	1	7	2014	9	10	20	39
22926	22926	11100073	214853070	313	1	1	2	2014	9	20	18	22
13879	13879	820846	214601197	20839	1	1	0	2014	4	13	18	28
24755	24755	9964988	214856248	131	6	1	8	2014	9	6	11	25
35224	35224	1615967	214574777	7852	1	1	0	2014	4	23	22	3
97306	97306	10803463	214553837	292	1	1	3	2014	9	21	12	28
14540	14540	548052	214834865	523	1	1	0	2014	4	8	16	8
1580	1580	1546673	214829366	1360	1	1	0	2014	4	27	8	44
12994	12994	11376943	214859882	837	1	1	5	2014	9	27	13	35
31504	31504	9208924	214853709	837	1	1	5	2014	8	31	15	50
49921	49921	10086871	214853105	1046	1	1	5	2014	9	8	19	22
11433	11433	9125539	214853210	418	1	1	5	2014	8	25	12	43

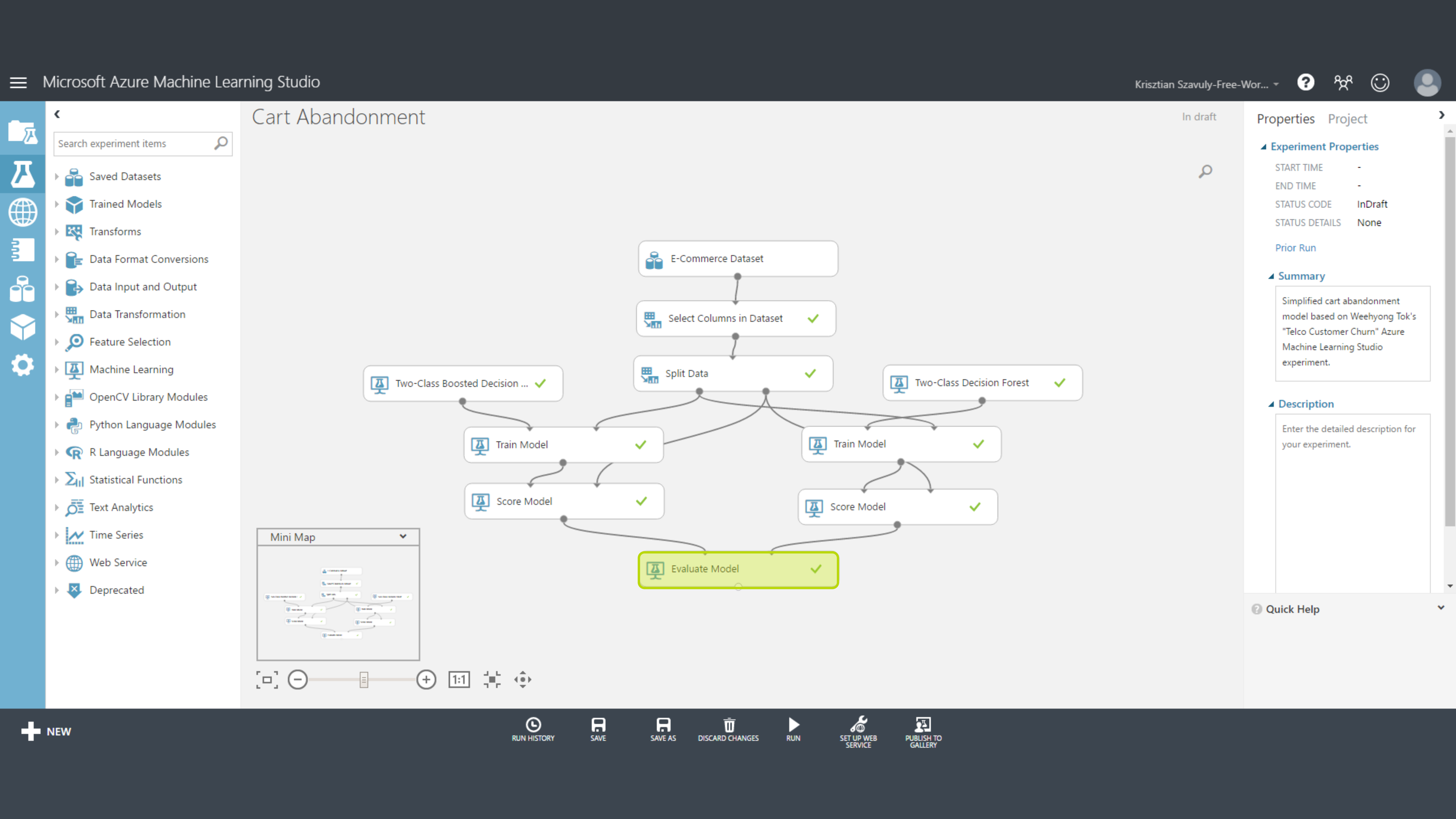

Microsoft Azure Machine Learning Studio
Krisztian Szavuly-Free-Wor...
Search experiment items
Saved Datasets
Trained Models
Transforms
Data Format Conversions
Data Input and Output
Data Transformation
Feature Selection
Machine Learning
OpenCV Library Modules
Python Language Modules
R Language Modules
Statistical Functions
Text Analytics
Time Series
Web Service
Deprecated
Cart Abandonment
In draft
E-Commerce Dataset
Select Columns in Dataset
Split Data
Two-Class Boosted Decision ...
Two-Class Decision Forest
Train Model
Train Model
Score Model
Score Model
Evaluate Model
Mini Map
1:1
Properties
Project
Experiment Properties
START TIME -
END TIME -
STATUS CODE InDraft
STATUS DETAILS None
Prior Run
Summary
Simplified cart abandonment model based on Weehyong Tok's "Telco Customer Churn" Azure Machine Learning Studio experiment.
Description
Enter the detailed description for your experiment.
Quick Help
NEW
RUN HISTORY
SAVE
SAVE AS
DISCARD CHANGES
RUN
SET UP WEB SERVICE
PUBLISH TO GALLERY

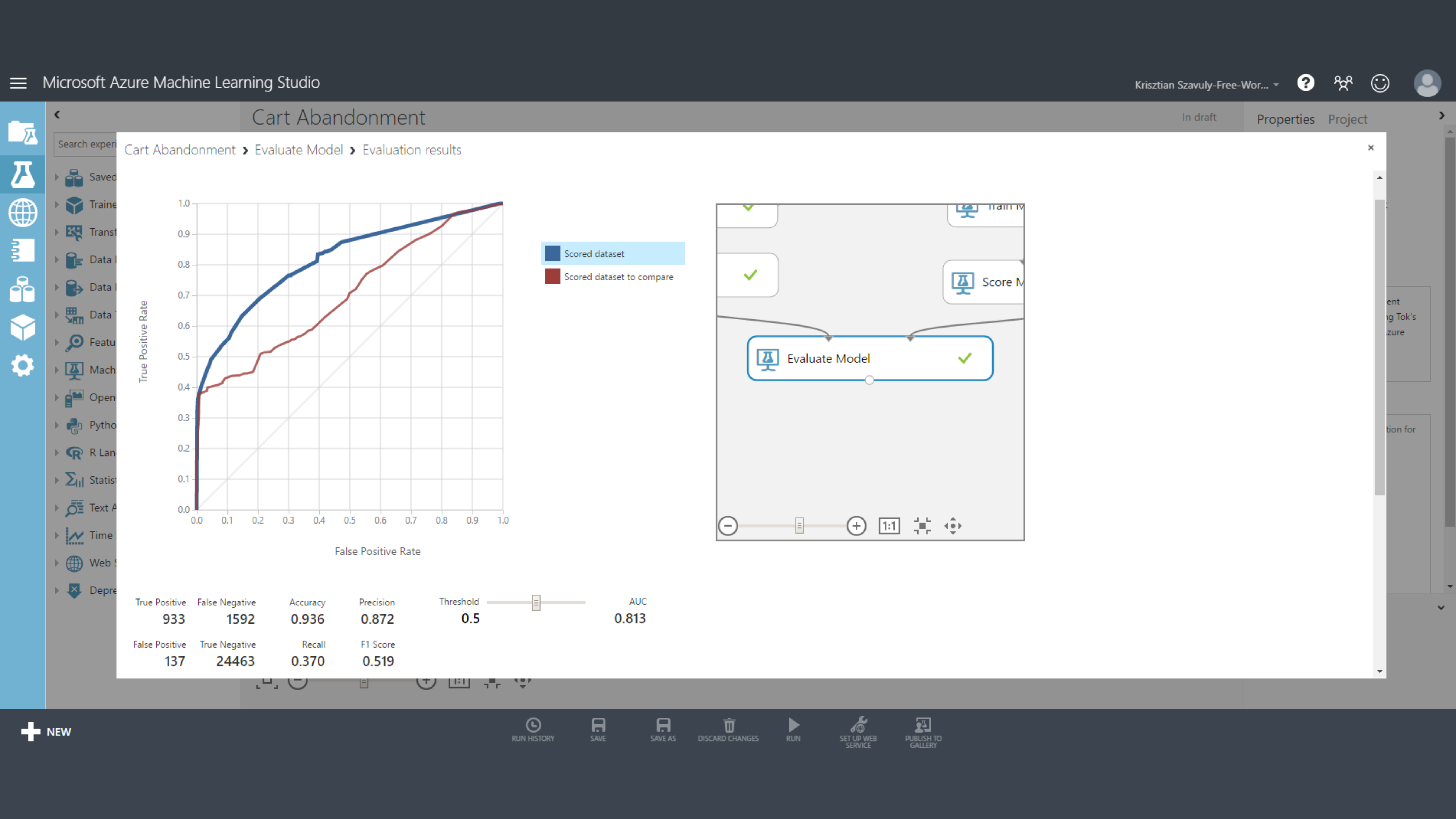
Tréning adat
Validációs adat
Jövőbeli adat


Microsoft Azure Machine Learning Studio
Krisztian Szavuly-Free-Wor...
Search experiment items
Saved Datasets
Trained Models
Transforms
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
Deprecated
Cart Abandonment
In draft
E-Commerce Dataset
Select Columns in Dataset
Split Data
Two-Class Boosted Decision ...
Two-Class Decision Forest
Train Model
Train Model
Score Model
Score Model
Evaluate Model
Mini Map
1:1
Properties
Project
Experiment Properties
START TIME -
END TIME -
STATUS CODE InDraft
STATUS DETAILS None
Prior Run
Summary
Simplified cart abandonment model based on Weehyong Tok's "Telco Customer Churn" Azure Machine Learning Studio experiment.
Description
Enter the detailed description for your experiment.
Quick Help
NEW
RUN HISTORY
SAVE
SAVE AS
DISCARD CHANGES
RUN
SET UP WEB SERVICE
PUBLISH TO GALLERY

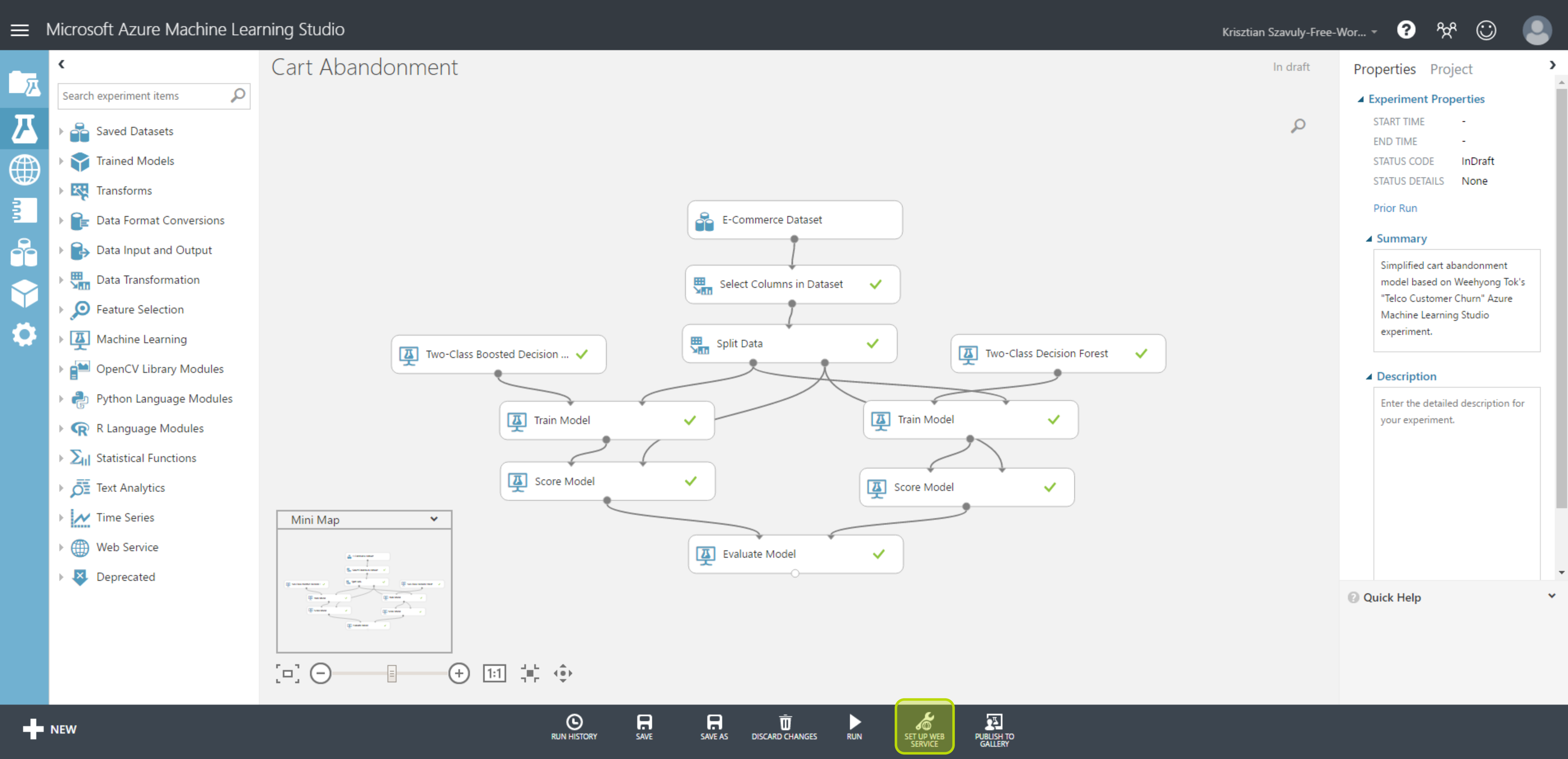

Microsoft Azure Machine Learning Studio
Krisztian Szavuly-Free-Wor...
Cart Abandonment
In draft
Search experiment items
Saved Datasets
Trained Models
Transforms
Data Format Conversions
Data Input and Output
Data Transformation
Feature Selection
Machine Learning
OpenCV Library Modules
Python Language Modules
R Language Modules
Statistical Functions
Text Analytics
Time Series
Web Service
Deprecated
E-Commerce Dataset
Select Columns in Dataset
Split Data
Two-Class Boosted Decision ...
Two-Class Decision Forest
Train Model
Train Model
Score Model
Score Model
Evaluate Model
Mini Map
Properties
Project
Two-Class Decision Forest
Resampling method
Bagging
Create trainer mode
Single Parameter
Number of decision trees
20
Maximum depth of the ...
5
Number of random spli...
10
Minimum number of sa...
1
Allow unknown valu...
START TIME 4/9/2017...
END TIME 4/9/2017...
ELAPSED TIME 0:00:00.0...
STATUS CODE Finished
STATUS DETAILS Task output was present in output
Quick Help
Creates a two-class classification model using the decision forest algorithm
(more help...)
NEW
RUN HISTORY
SAVE
SAVE AS
DISCARD CHANGES
RUN
SET UP WEB SERVICE
PUBLISH TO GALLERY


Microsoft Azure Machine Learning Studio
Krisztian Szavuly-Free-Wor...
Cart Abandonment
In draft
Properties
Project
Cart Abandonment › Score Model › Scored dataset
rows 27125
columns 7
view as
ItemID
Buy
Category
Day
Hour
Scored Labels
Scored Probabilities
214639200 1 0 14 11 0 0.086396
214853855 1 S 13 17 0 0.227439
214844405 0 0 22 11 0 0.07505
214800262 0 0 1 15 0 0.075713
214844398 0 3 17 18 0 0.089884
214703801 0 0 29 9 0 0.087599
214849322 0 S 30 20 0 0.147759
214845958 0 3 25 18 0 0.094341
214849284 0 S 24 14 0 0.147759
214602830 0 0 2 18 0 0.074903
214832143 1 1 13 1 0 0.077978
214832647 1 2 29 14 0 0.085332
214831948 0 0 14 14 0 0.077439
214705361 0 10 31 11 0 0.088071
214835747 0 0 21 8 0 0.075781
214848371 0 S 19 14 0 0.11265
214587525 1 2 25 15 0 0.091122
Statistics
Mean 0.0943
Median 0.0861
Min 0.0461
Max 0.5594
Standard Deviation 0.0328
Unique Values 1426
Missing Values 0
Feature Type Numeric Score
Visualizations
Scored Probabilities
Histogram
frequency
1.8e+4
1.6e+4
1.4e+4
1.2e+4
1.0e+4
8000
6000
NEW
RUN HISTORY
SAVE
SAVE AS
DISCARD CHANGES
RUN
SET UP WEB SERVICE
PUBLISH TO GALLERY


Microsoft Azure Machine Learning Studio
Krisztian Szavuly-Free-Wor...
Cart Abandonment
In draft
Search experiment items
Saved Datasets
Trained Models
Transforms
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
Deprecated
E-Commerce Dataset
Select Columns in Dataset
Split Data
Two-Class Boosted Decision ...
Two-Class Decision Forest
Train Model
Train Model
Score Model
Score Model
Evaluate Model
Mini Map
1:1
Properties
Project
Experiment Properties
START TIME -
END TIME -
STATUS CODE InDraft
STATUS DETAILS None
Prior Run
Summary
Simplified cart abandonment model based on Weehyong Tok's "Telco Customer Churn" Azure Machine Learning Studio experiment.
Description
Enter the detailed description for your experiment.
Quick Help
NEW
RUN HISTORY
SAVE
SAVE AS
DISCARD CHANGES
RUN
SET UP WEB SERVICE
PUBLISH TO GALLERY


Microsoft Azure Machine Learning Studio
Krisztian Szavuly-Free-Wor...
Cart Abandonment
In draft
Properties
Project
Cart Abandonment › Evaluate Model › Evaluation results
True Positive Rate
False Positive Rate
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Scored dataset
Scored dataset to compare
Score M
Evaluate Model
1:1
True Positive
933
False Negative
1592
Accuracy
0.936
Precision
0.872
Threshold
0.5
AUC
0.813
False Positive
137
True Negative
24463
Recall
0.370
F1 Score
0.519
NEW
RUN HISTORY
SAVE
SAVE AS
DISCARD CHANGES
RUN
SET UP WEB SERVICE
PUBLISH TO GALLERY

Microsoft Azure Machine Learning Studio
Krisztian Szavuly-Free-Wor...
Cart Abandonment
In draft
Properties
Project
Cart Abandonment › Evaluate Model › Evaluation results
True Positive Rate
False Positive Rate
Scored dataset
Scored dataset to compare
Evaluate Model
True Positive
933
False Negative
1592
False Positive
137
True Negative
24463
Accuracy
0.936
Precision
0.872
Threshold
0.5
AUC
0.813
Recall
0.370
F1 Score
0.519
NEW
RUN HISTORY
SAVE
SAVE AS
DISCARD CHANGES
RUN
SET UP WEB SERVICE
PUBLISH TO GALLERY


Microsoft Azure Machine Learning Studio
Krisztian Szavuly-Free-Wor...
Search experiment items
Saved Datasets
Trained Models
Transforms
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
Deprecated
Cart Abandonment
In draft
E-Commerce Dataset
Select Columns in Dataset
Split Data
Two-Class Boosted Decision ...
Two-Class Decision Forest
Train Model
Train Model
Score Model
Score Model
Evaluate Model
Mini Map
1:1
Properties
Project
Experiment Properties
START TIME -
END TIME -
STATUS CODE InDraft
STATUS DETAILS None
Prior Run
Summary
Simplified cart abandonment model based on Weehyong Tok's "Telco Customer Churn" Azure Machine Learning Studio experiment.
Description
Enter the detailed description for your experiment.
Quick Help
NEW
RUN HISTORY
SAVE
SAVE AS
DISCARD CHANGES
RUN
SET UP WEB SERVICE
PUBLISH TO GALLERY

Szávuly Krisztián
Microsoft

Data Science

MICROSOFT PROFESSIONAL PROGRAM

@szavuly