Microsoft

Recommendation System

Contents

Goals and Requirements 3
Dataset for Recommendation-Restaurant 4
Dataset for Restaurant customer data 5
Clean Missing Data 6
Dataset for Restaurant Feature data 8
Dataset for Split Data 9
Train Matchbox Recommender Dataset 10
Score Matchbox Recommender Dataset 15
UNDERSTANDING THE MATCHBOX RECOMMENDATION RESULTS 20

Goals and Requirements

Estimated time to complete lab is 40-45 minutes

Goals

1. Implement and design a model for Recommending Restaurant based on Ratings and User Reviews.

2. Approach of using Match Box Recommendations

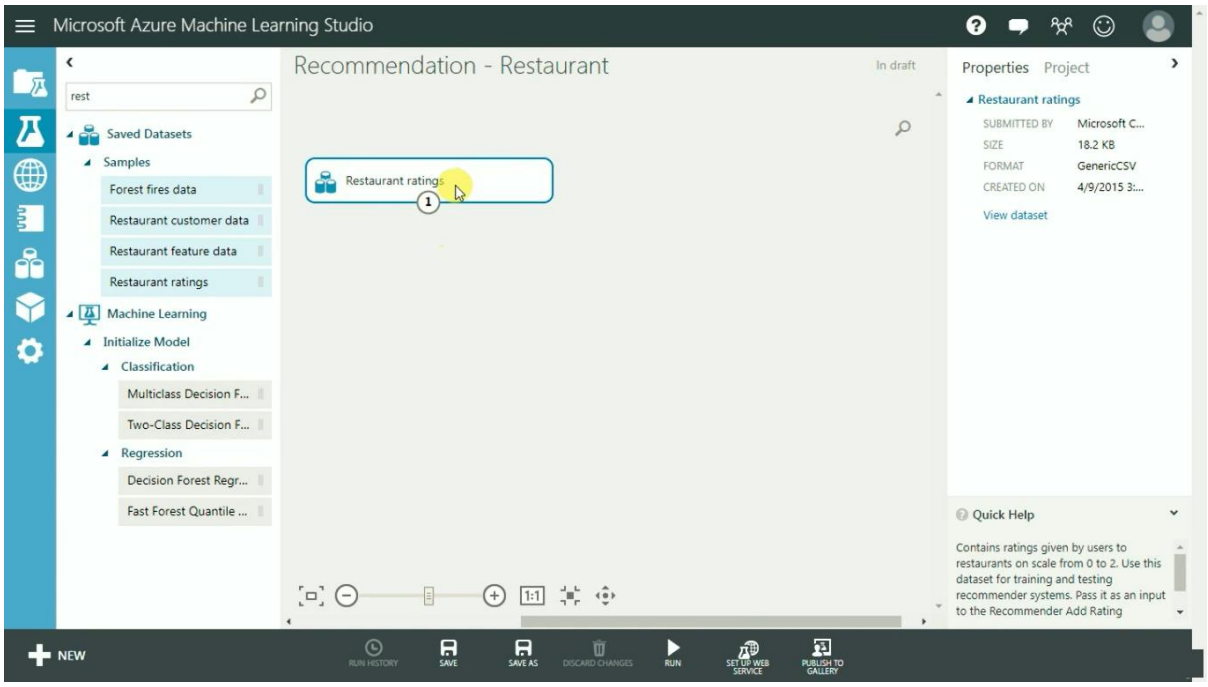
Requirements:

1. Access to an Azure Machine Learning Studio

Recommendation System

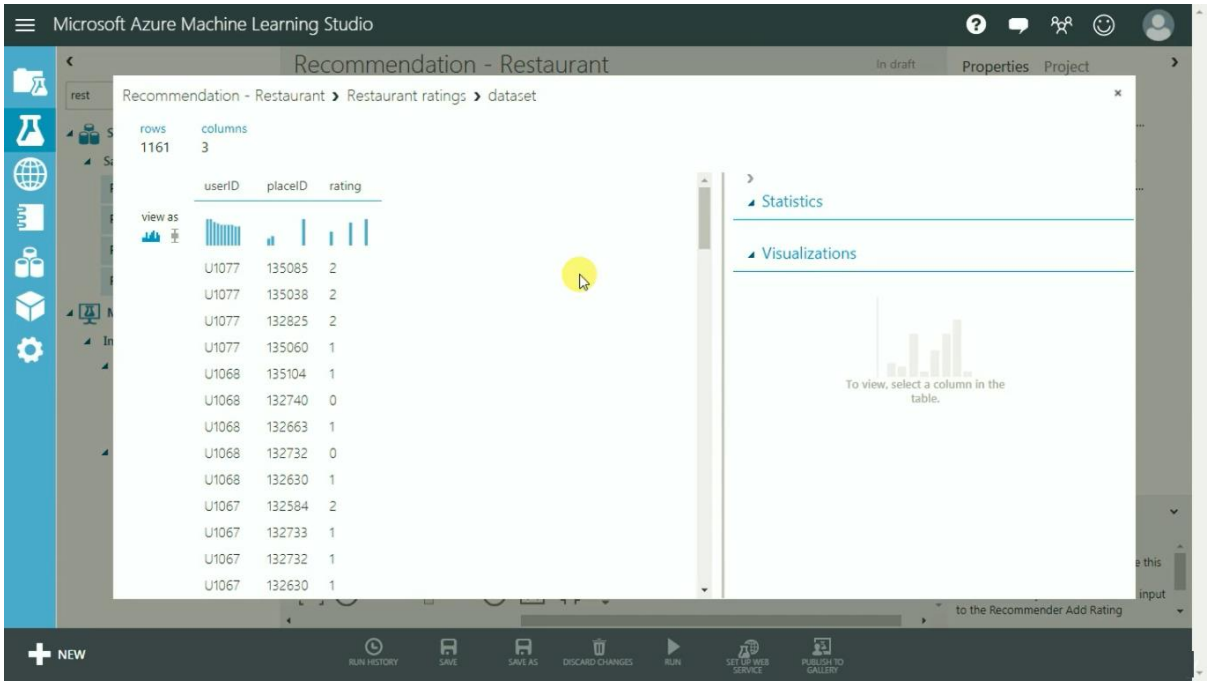
Dataset for Recommendation-Restaurant

Pick the sample data set provided by microsoft and drop in work space

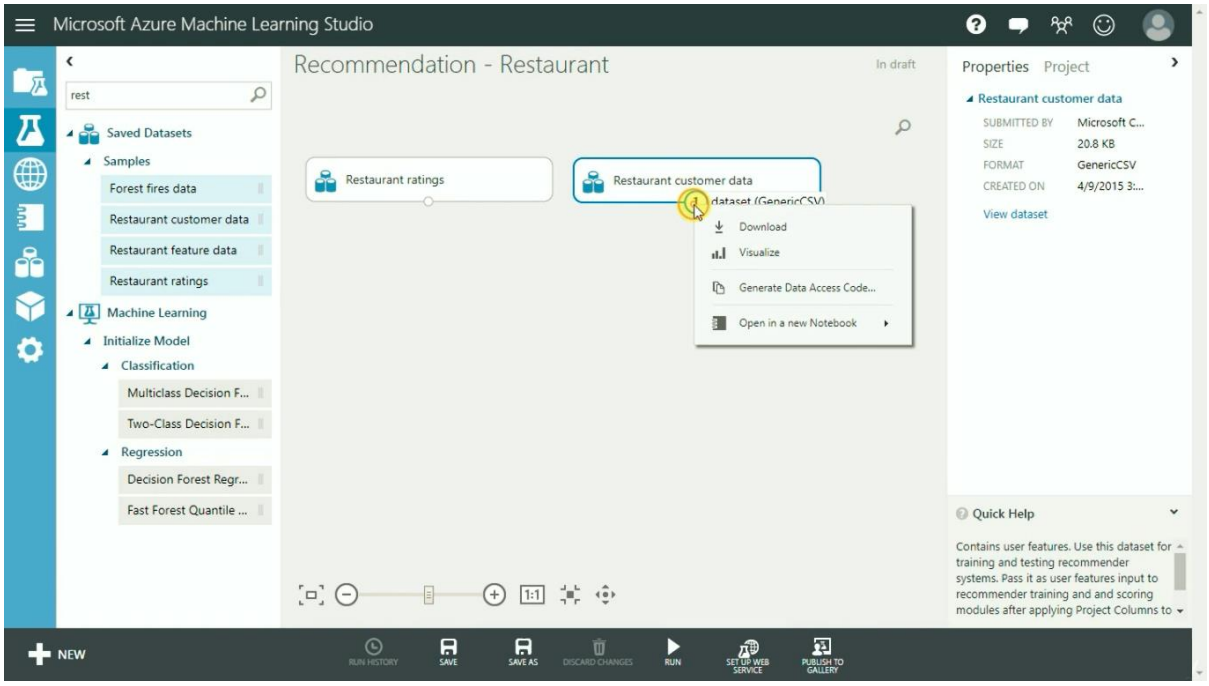


Dataset for Restaurant customer data

Visualize the dataset to know the features

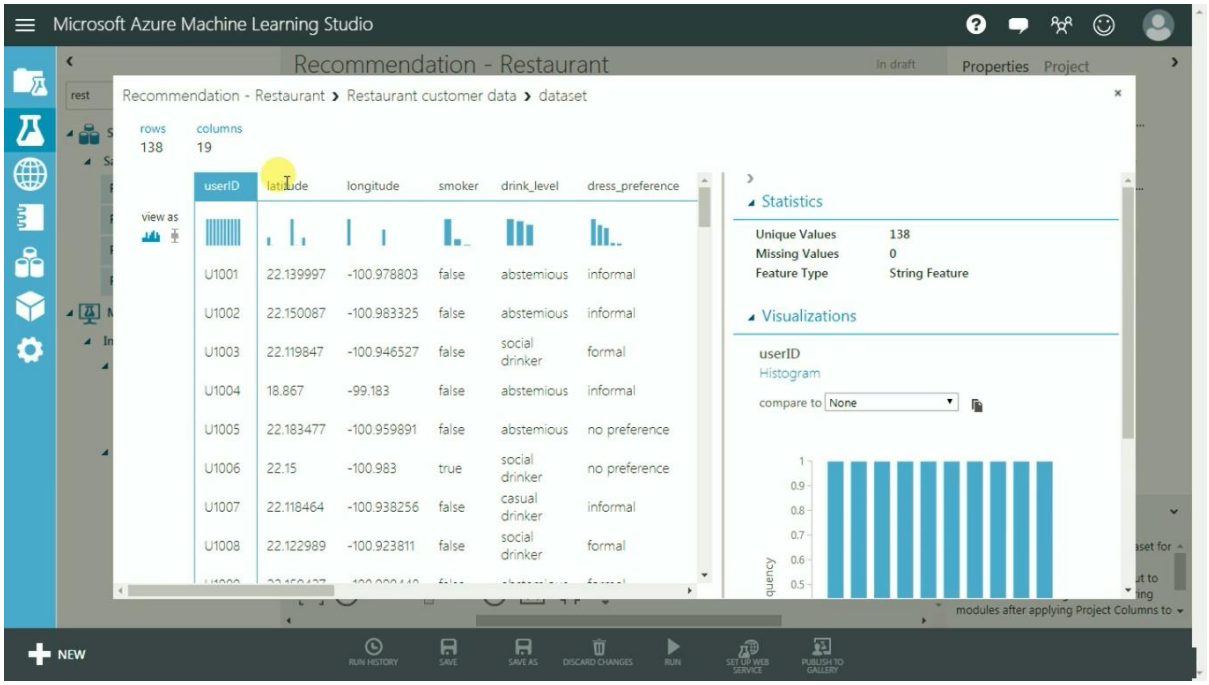


Add Restaurant customer data to the canvas and visualize

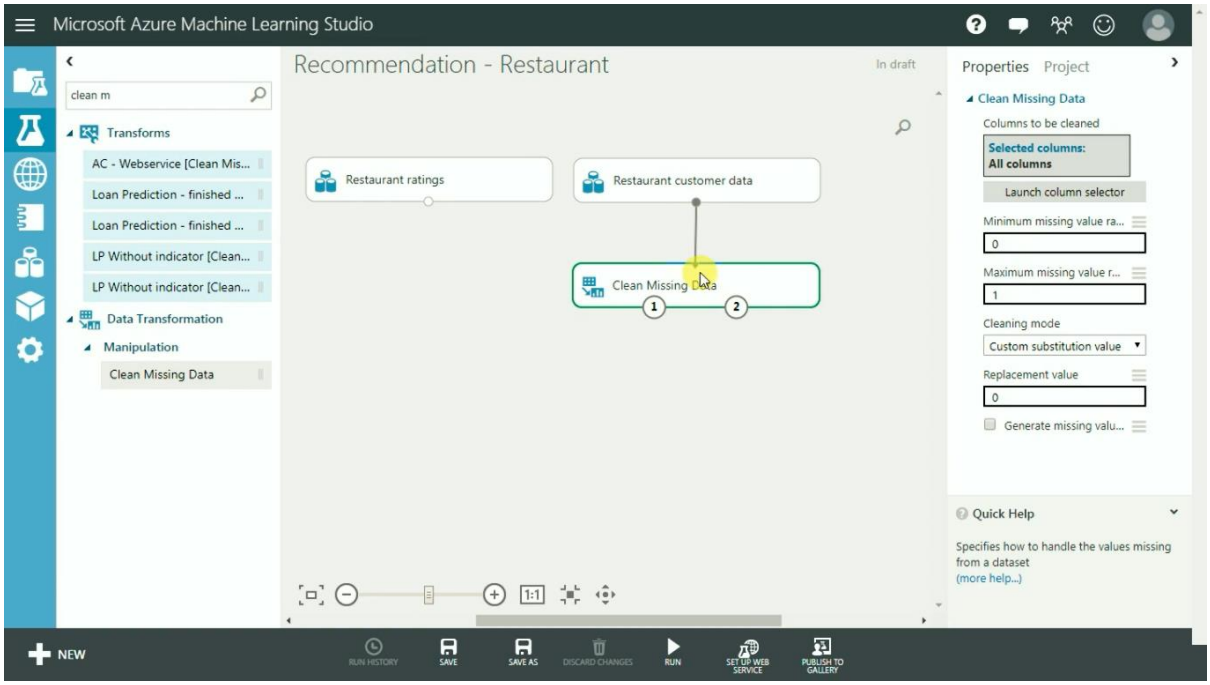


Clean Missing Data

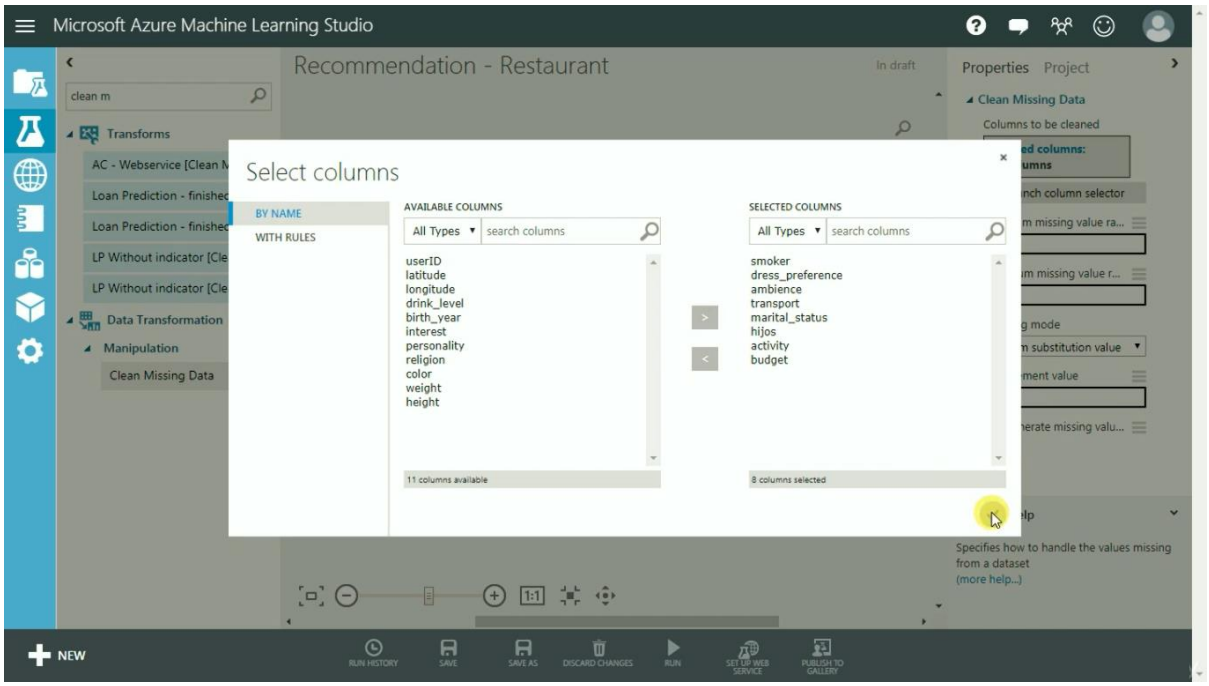
Check the result for missing values and string features



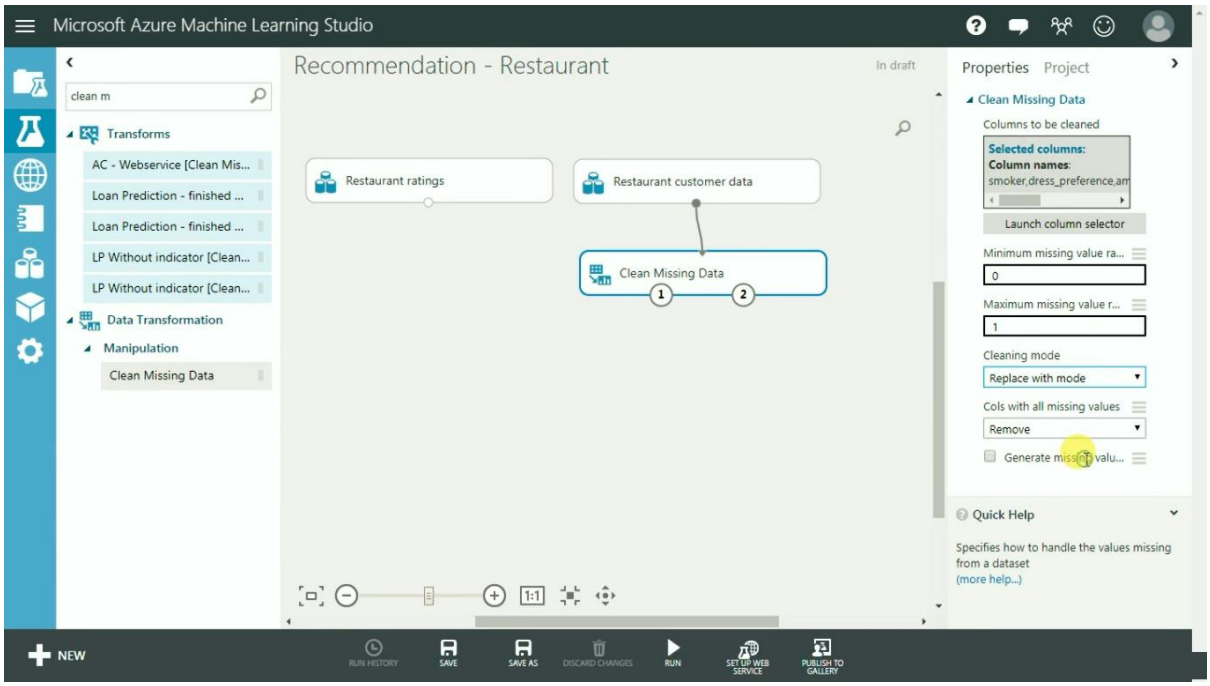
Add clean missing data and connect as required



Launch column selector and select the missing value columns and click ok

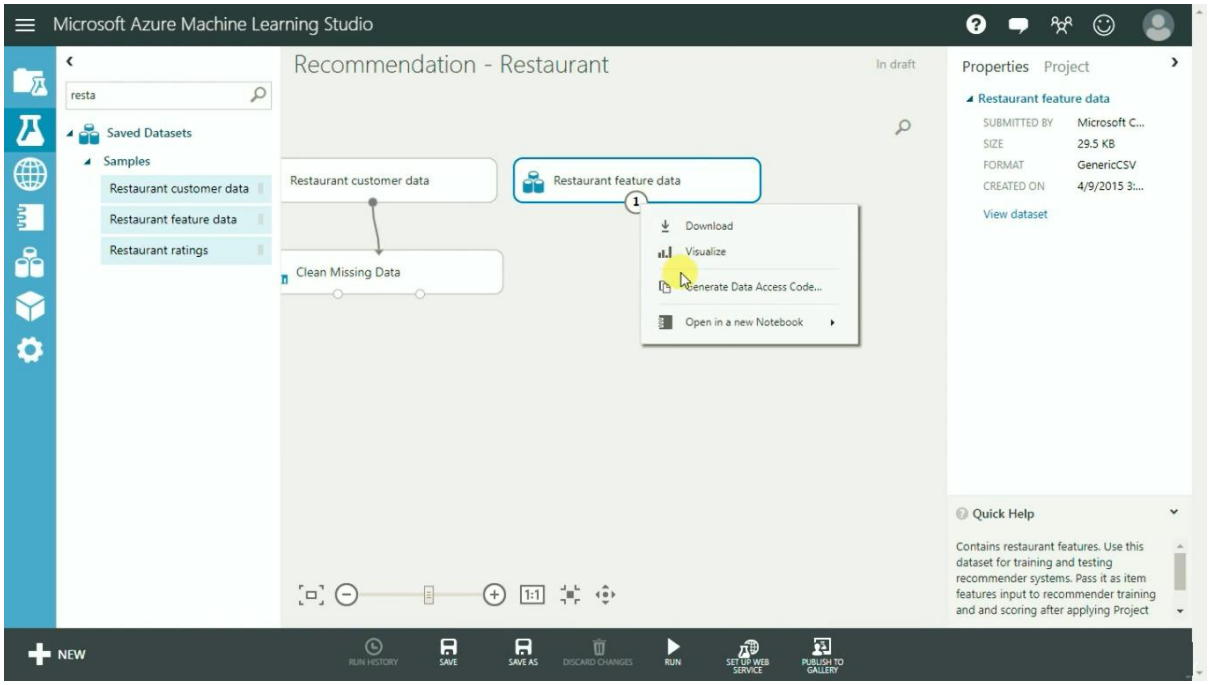


Parameters as follows

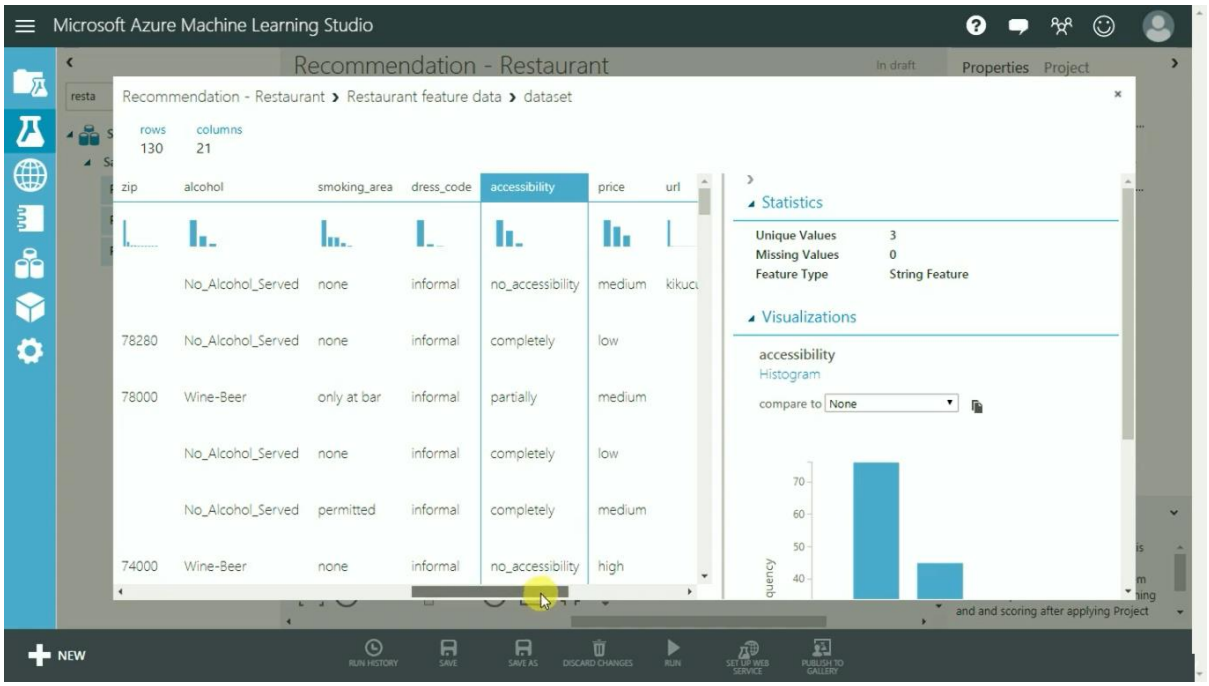


Dataset for Restaurant Feature data

Insert Restaurant feature data and visualize the same

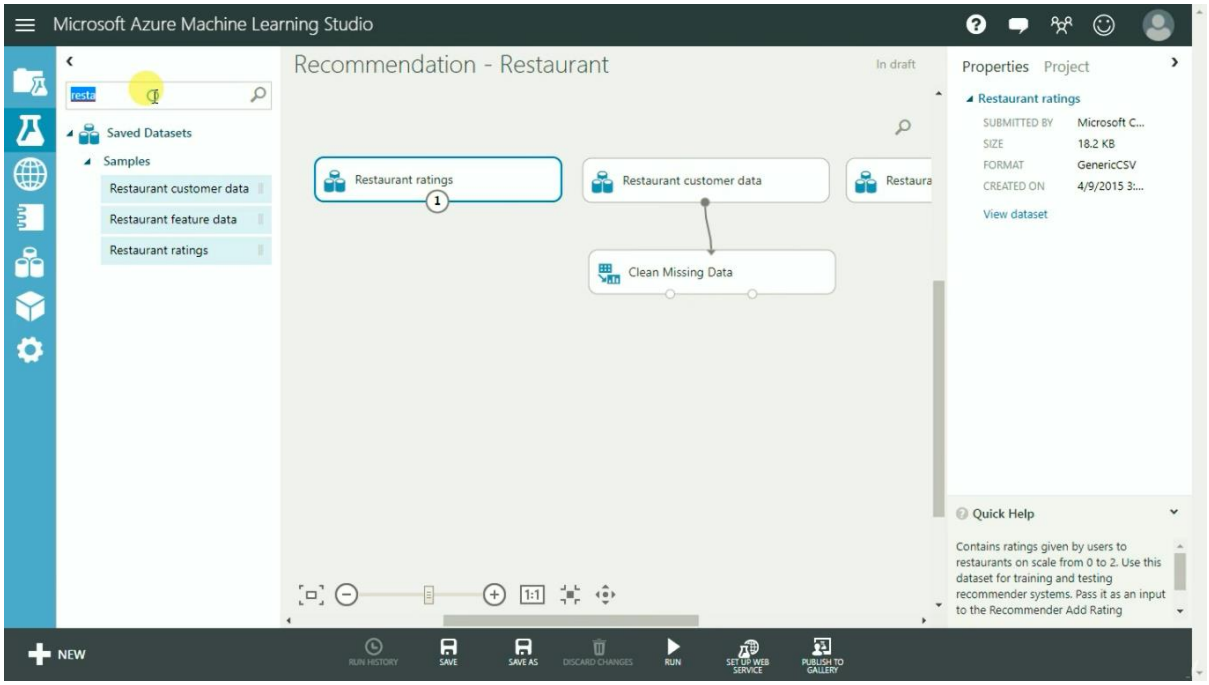


View the column with missing value that impacts on rating provided

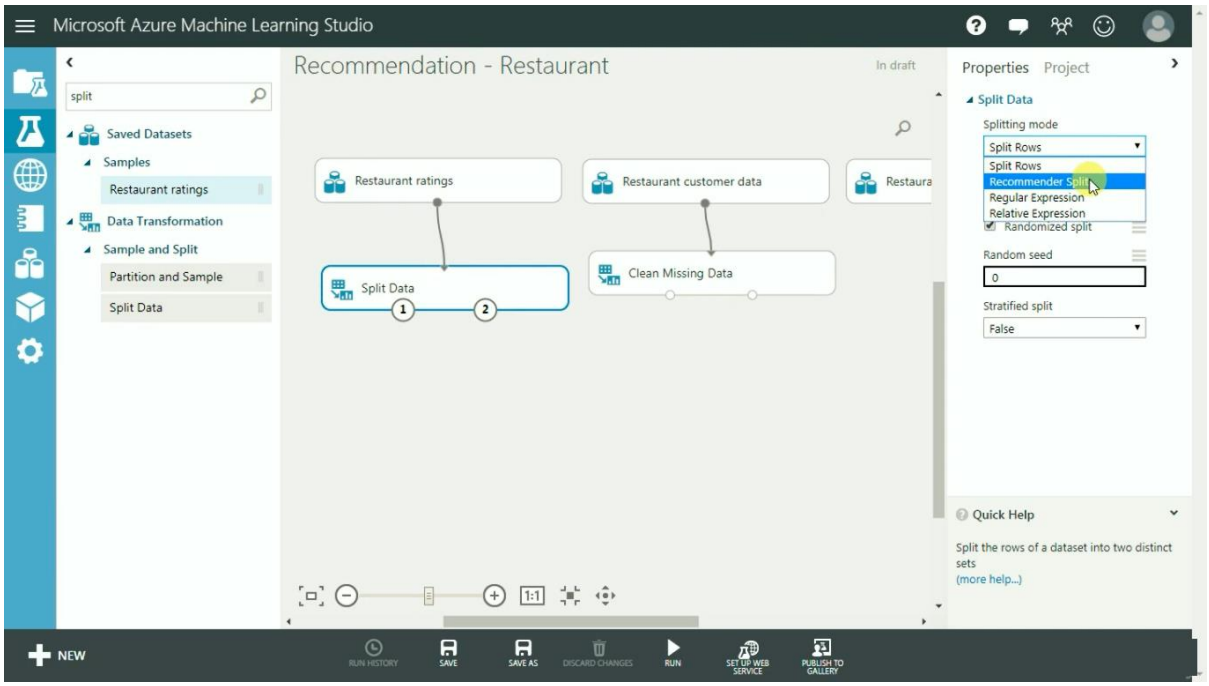


Dataset for Split Data

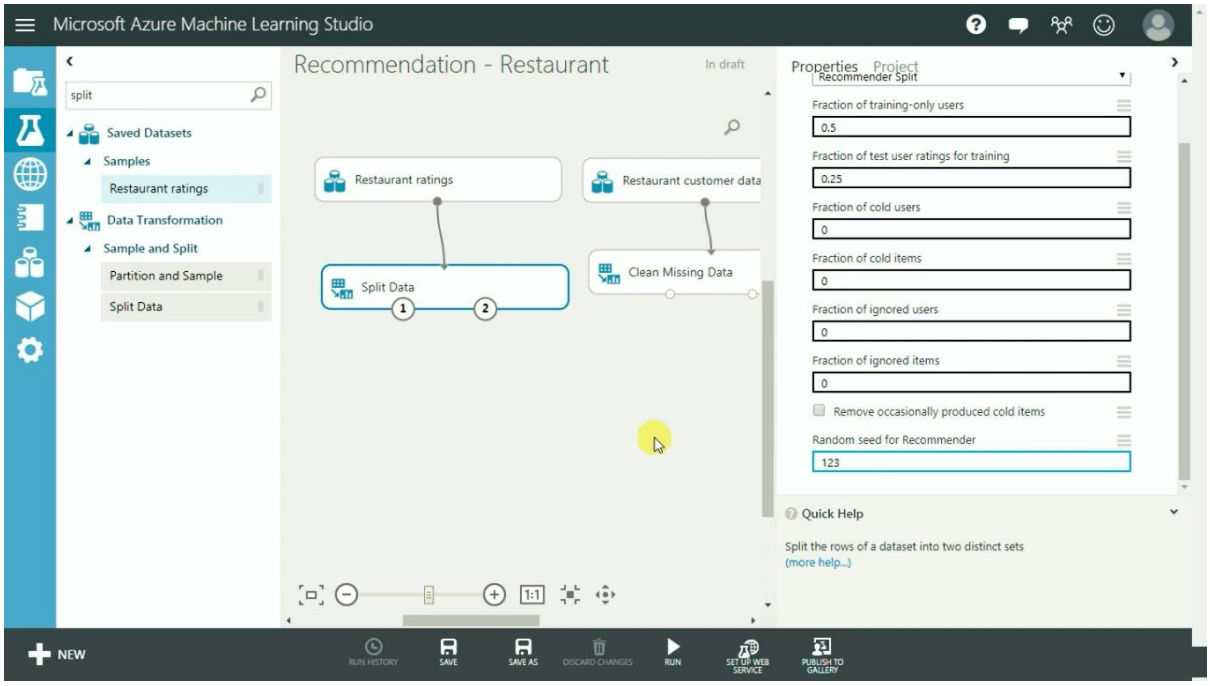
Now split the dataset restaurant ratings using split data



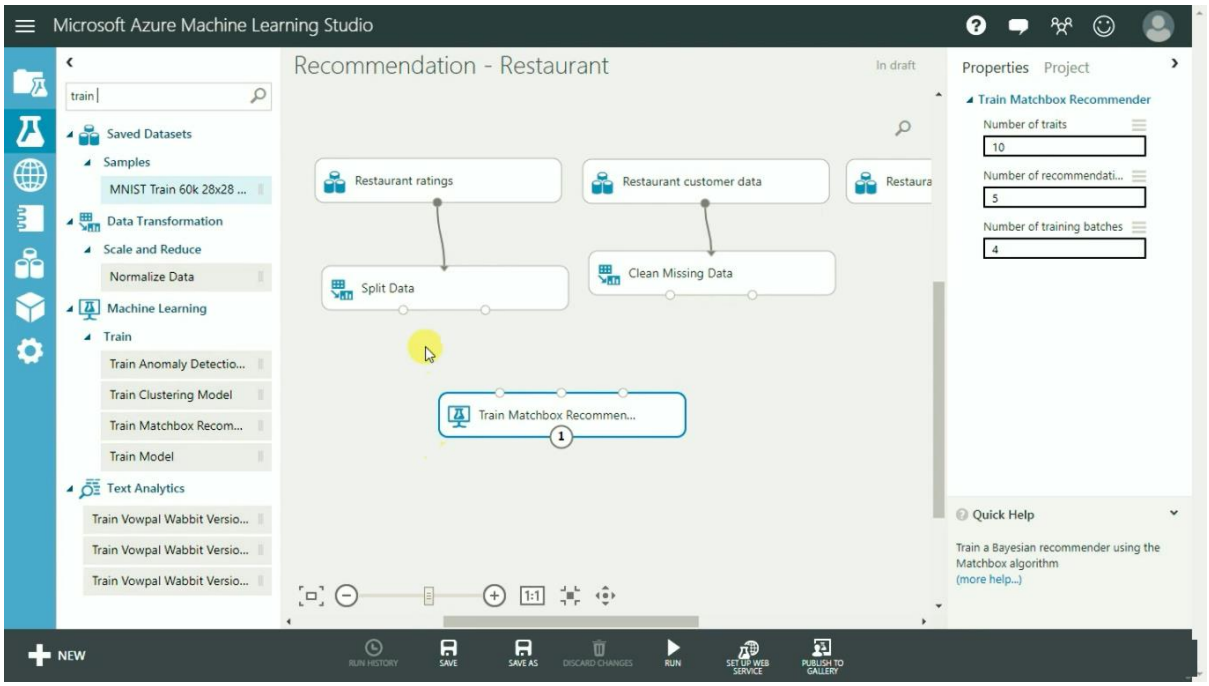
Add split data and connect with restaurant ratings and change parameter as required



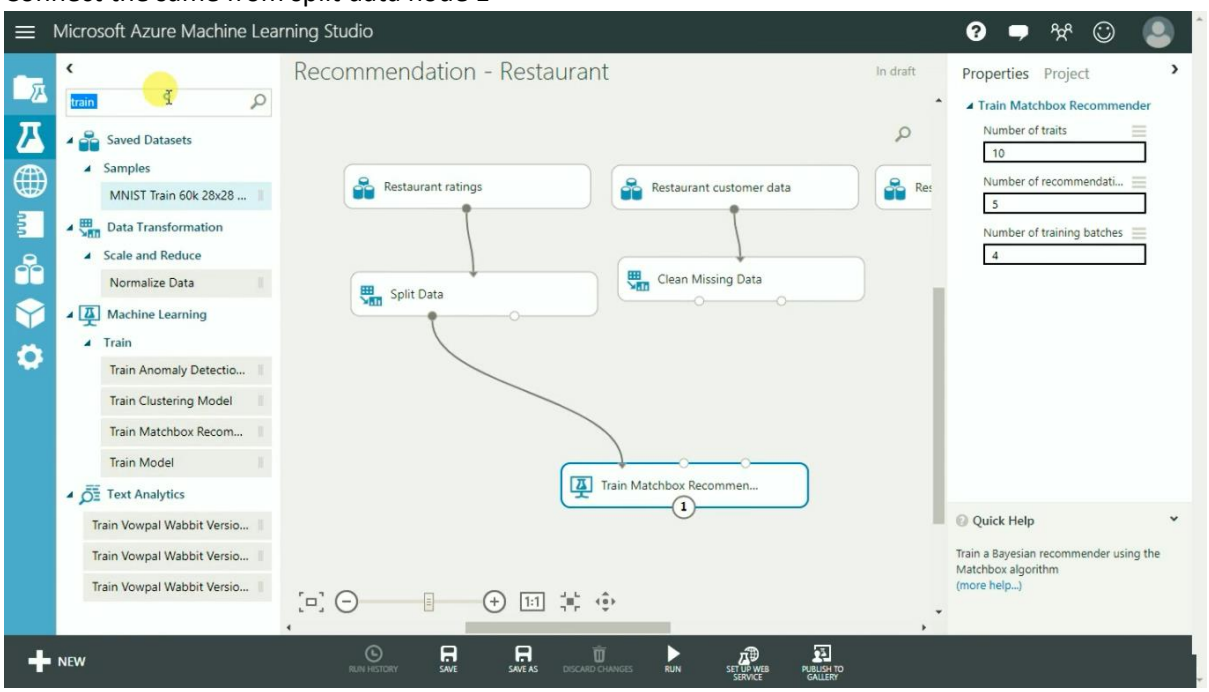
Train Matchbox Recommender Dataset



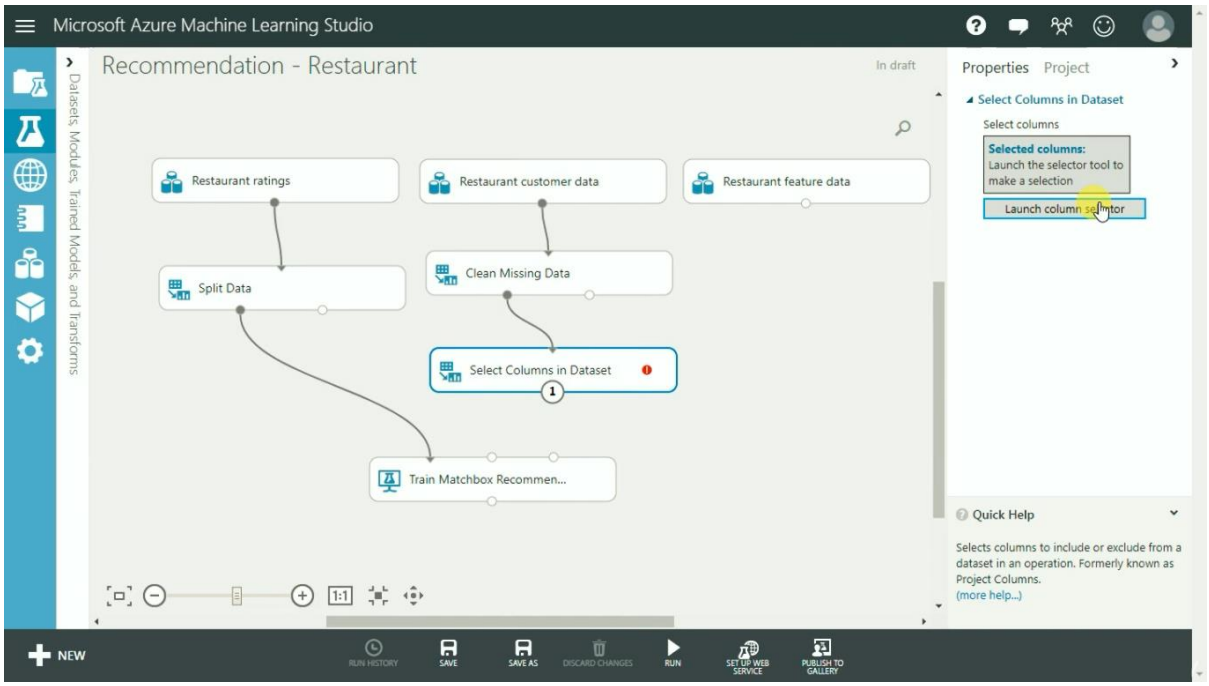
Add train matchbox recommender



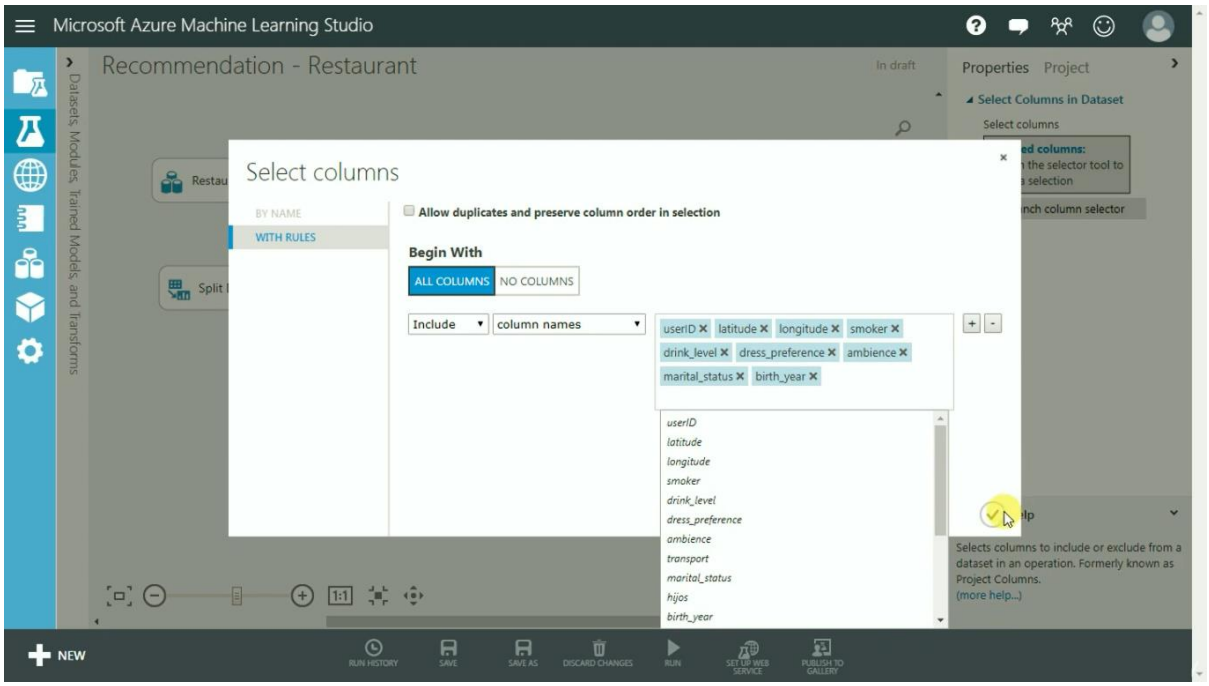
Connect the same from split data node 1



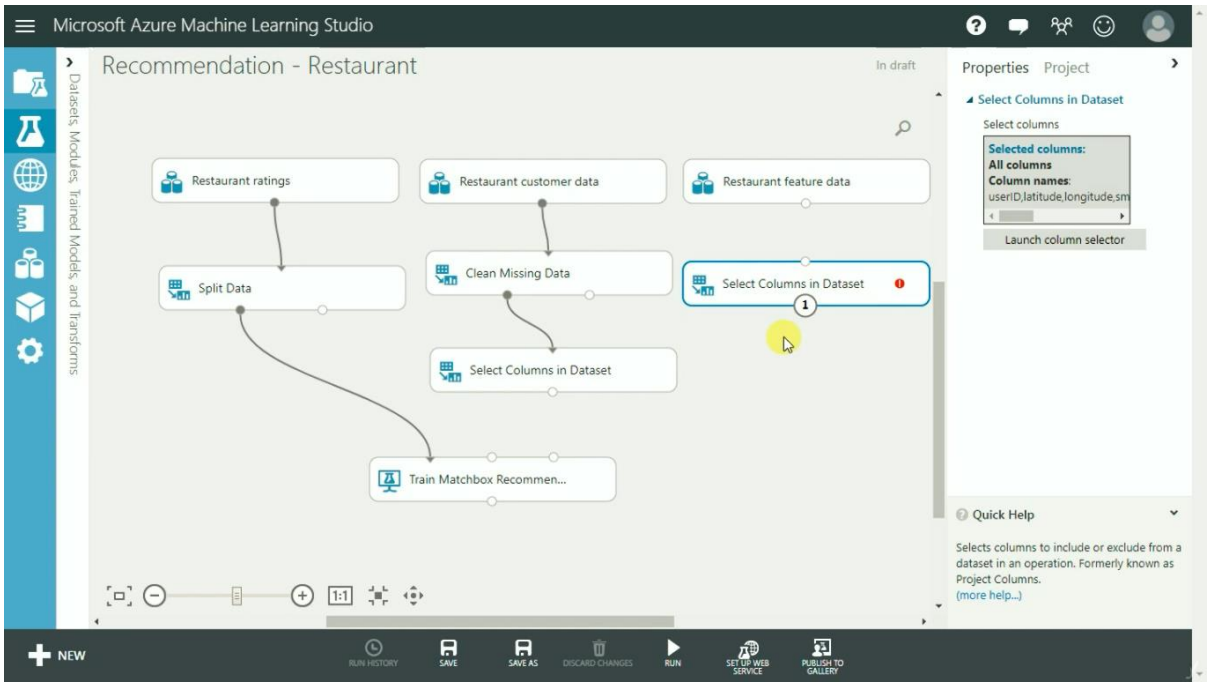
Insert select columns in dataset connected with clean missing data node and launch column selector



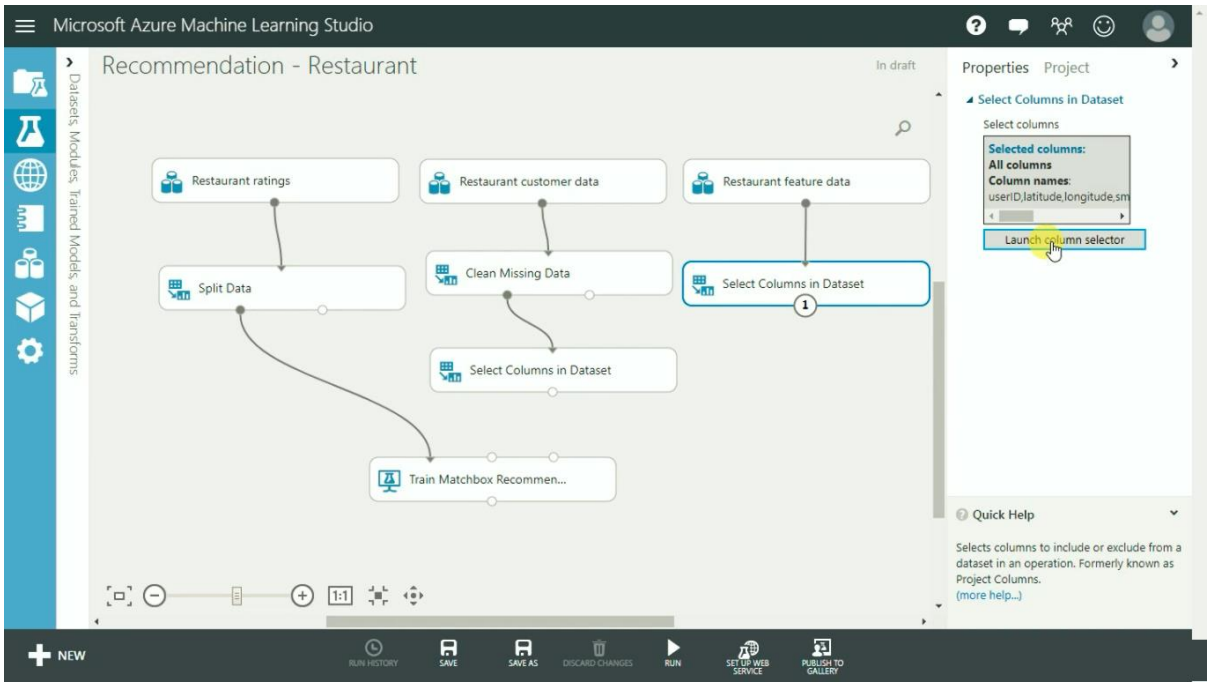
Add the columns as required and click ok



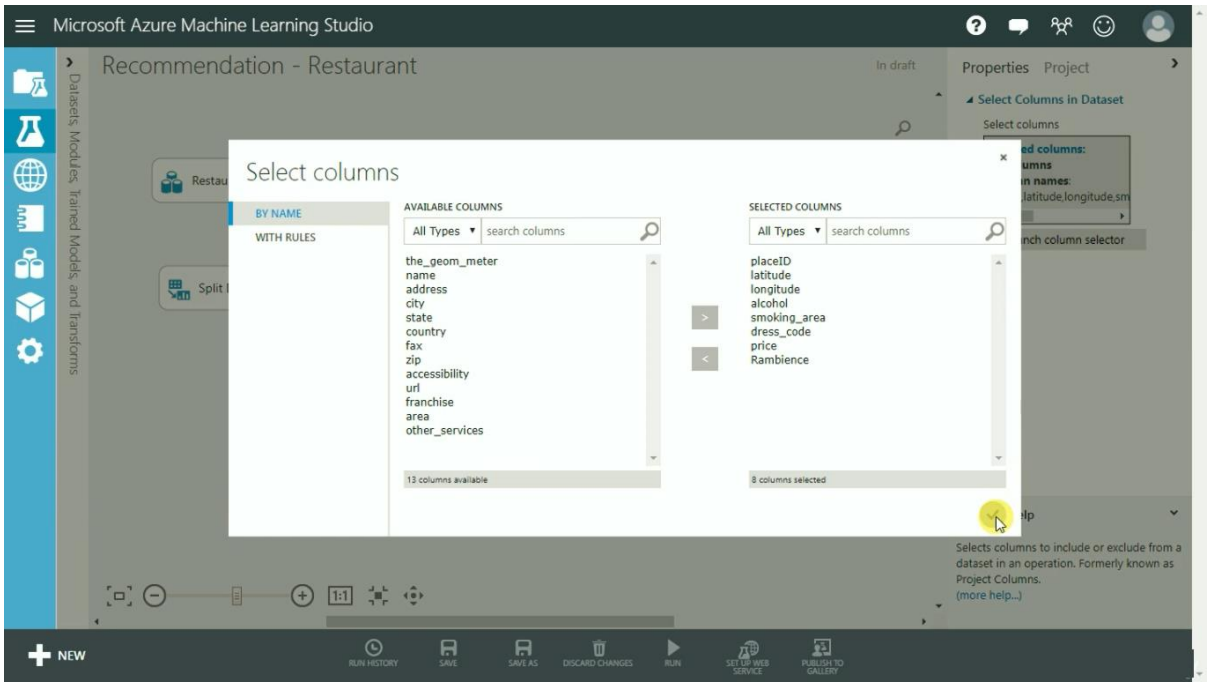
Repeat the step with Retaurant feature data



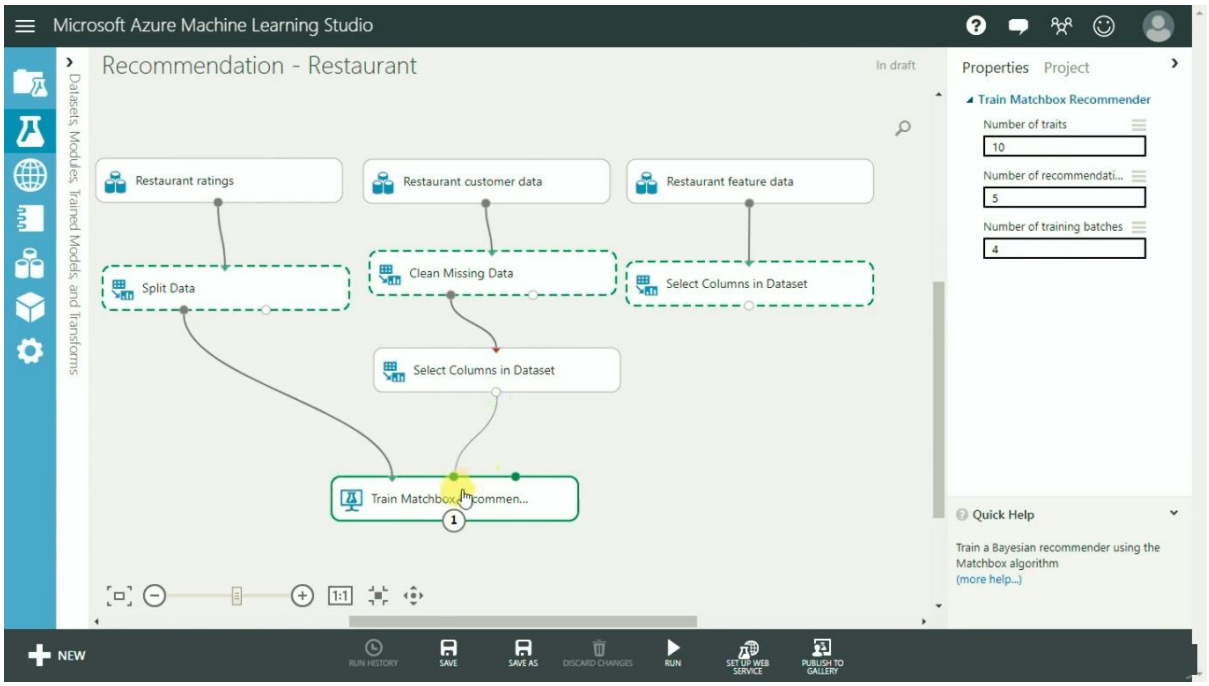
Connect with appropriate nodes and launch column selector

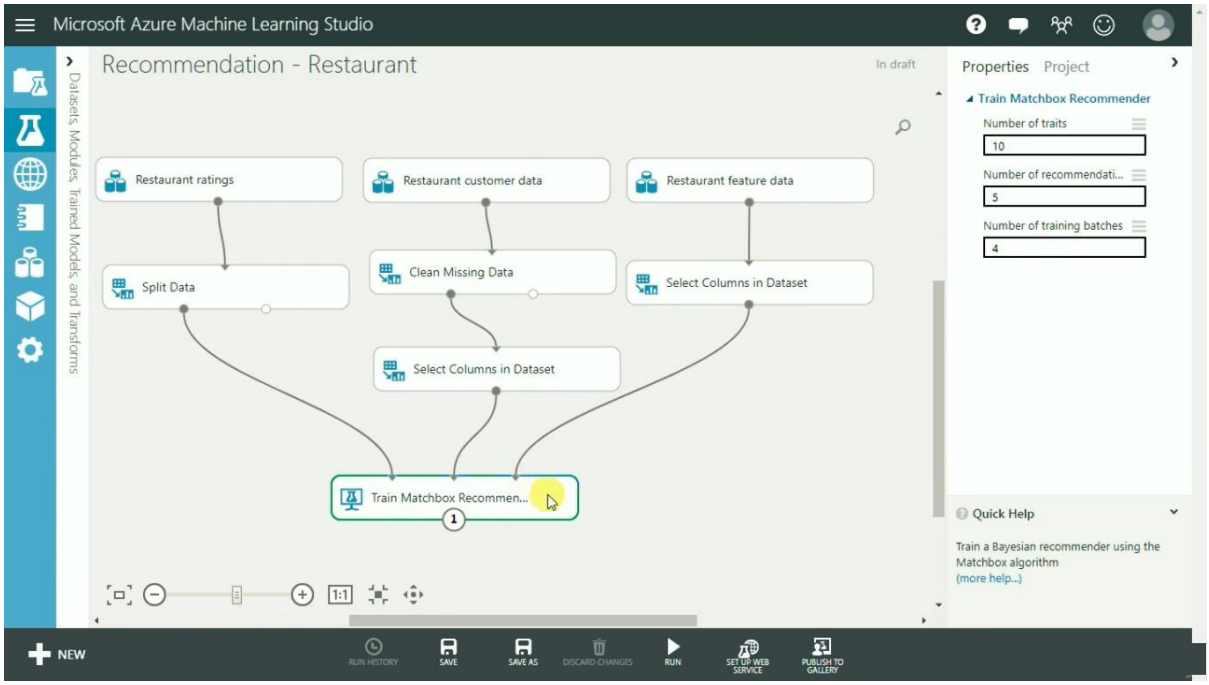


Select the required columns and click ok



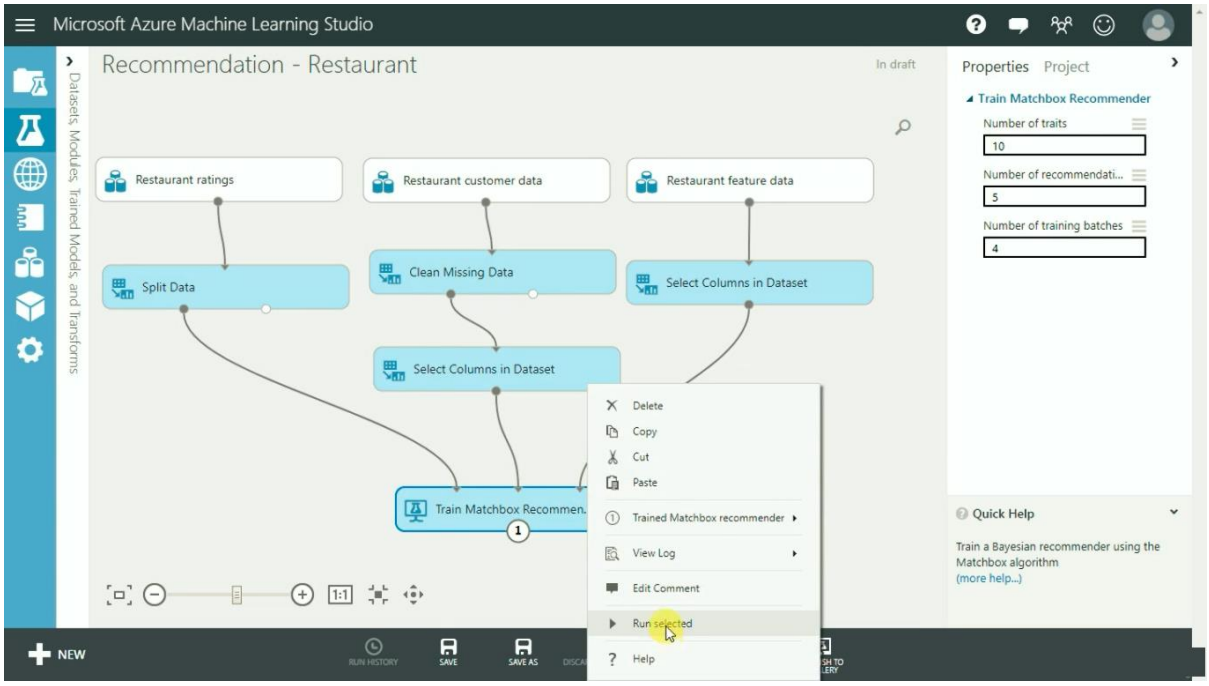
Connect select columns in data set to train matchbox recommender



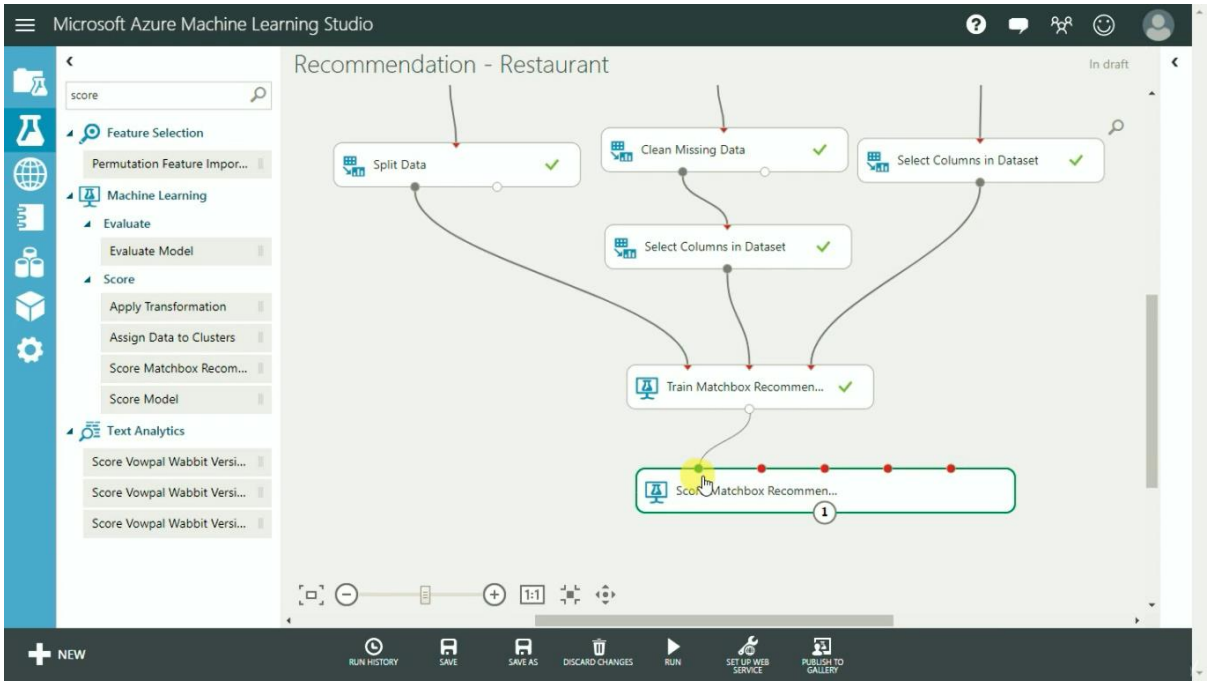


Score Matchbox Recommender Dataset

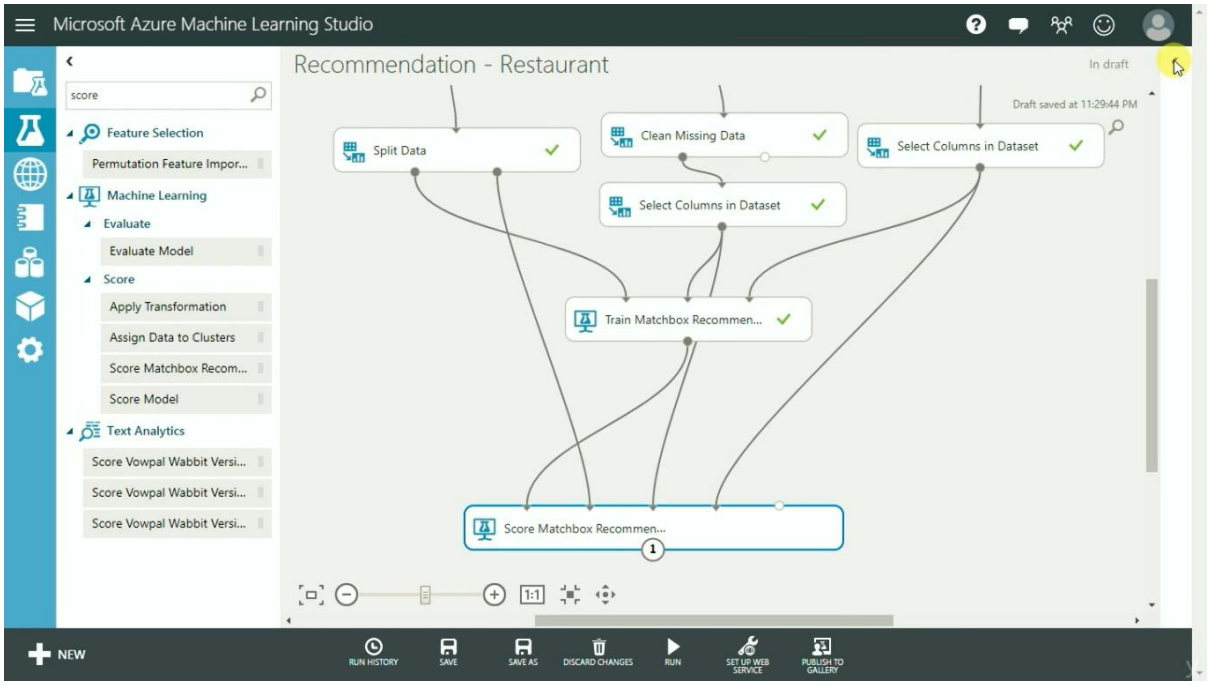
Run the module now



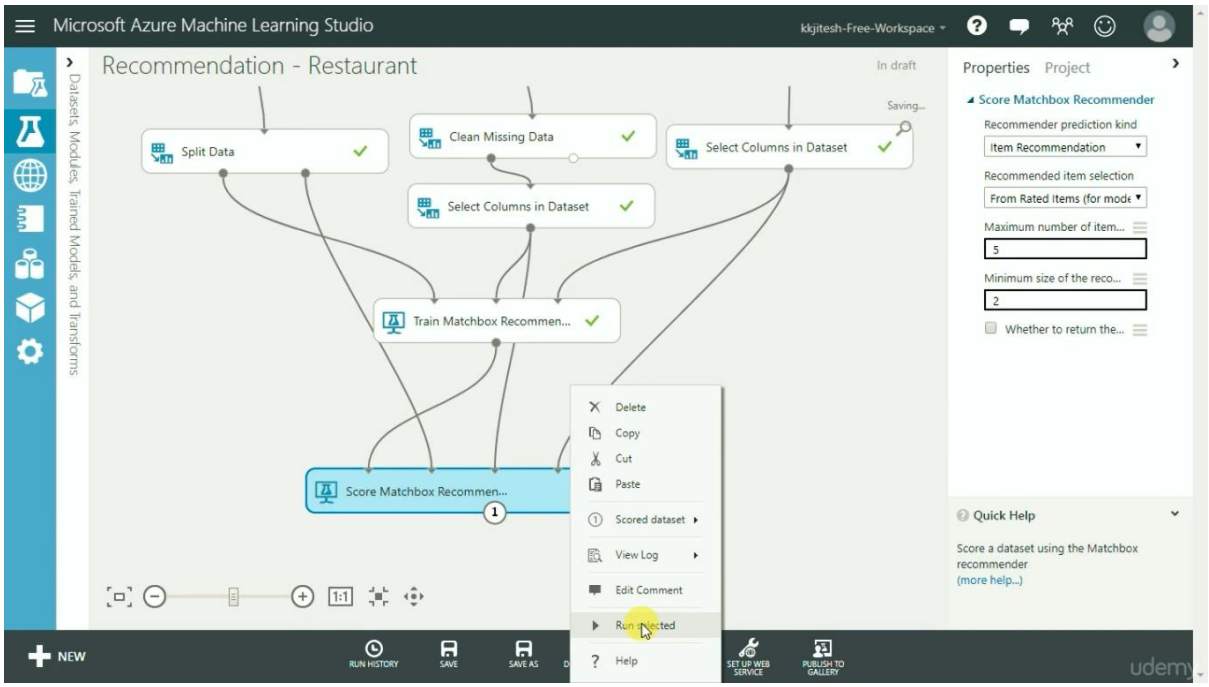
Insert and connect score matchbox recommender



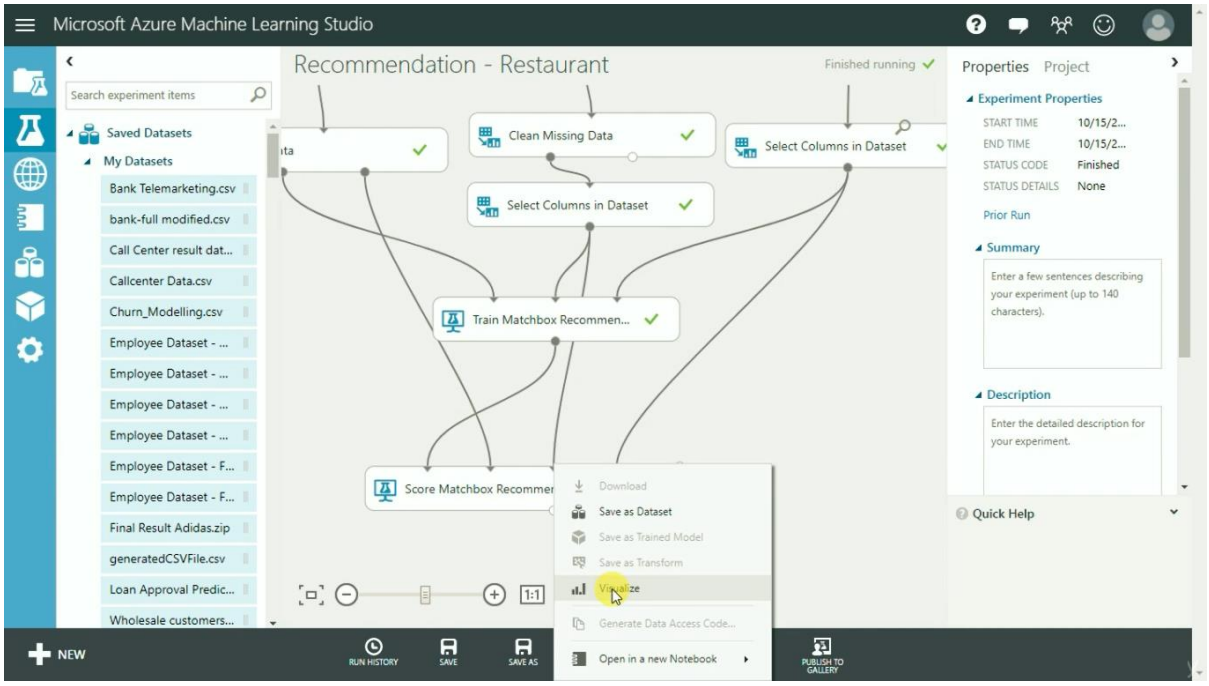
Connect the score input nodes as shown below



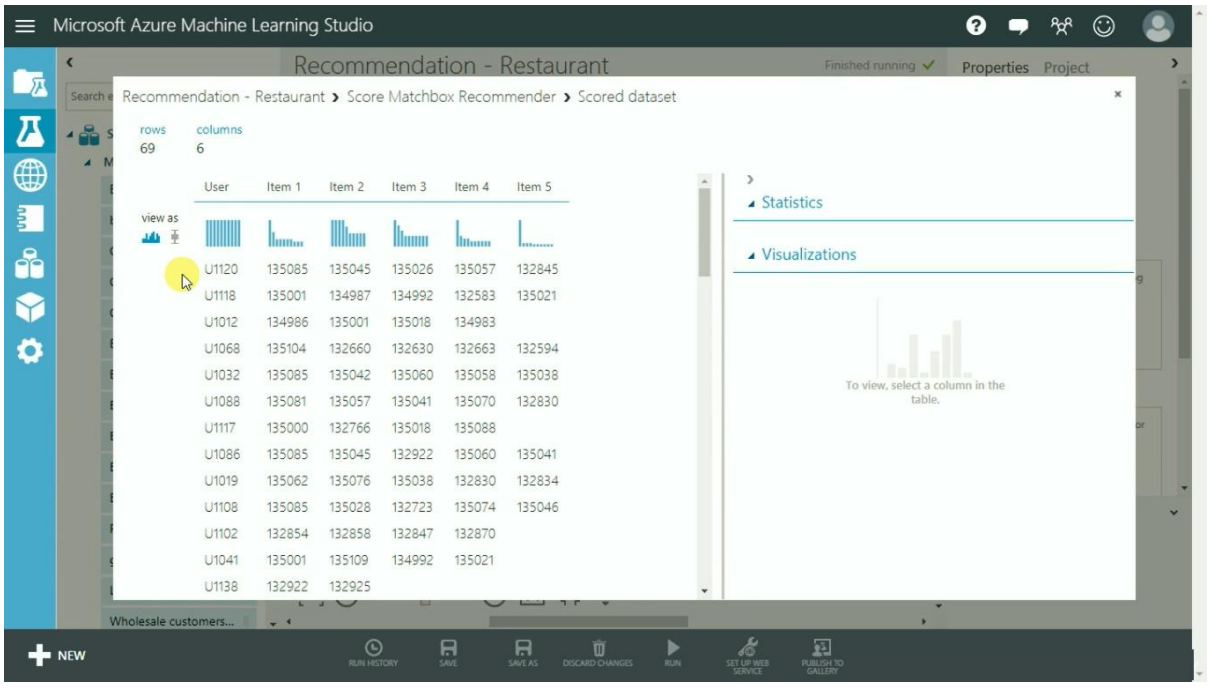
Check the required parameters and run the module now



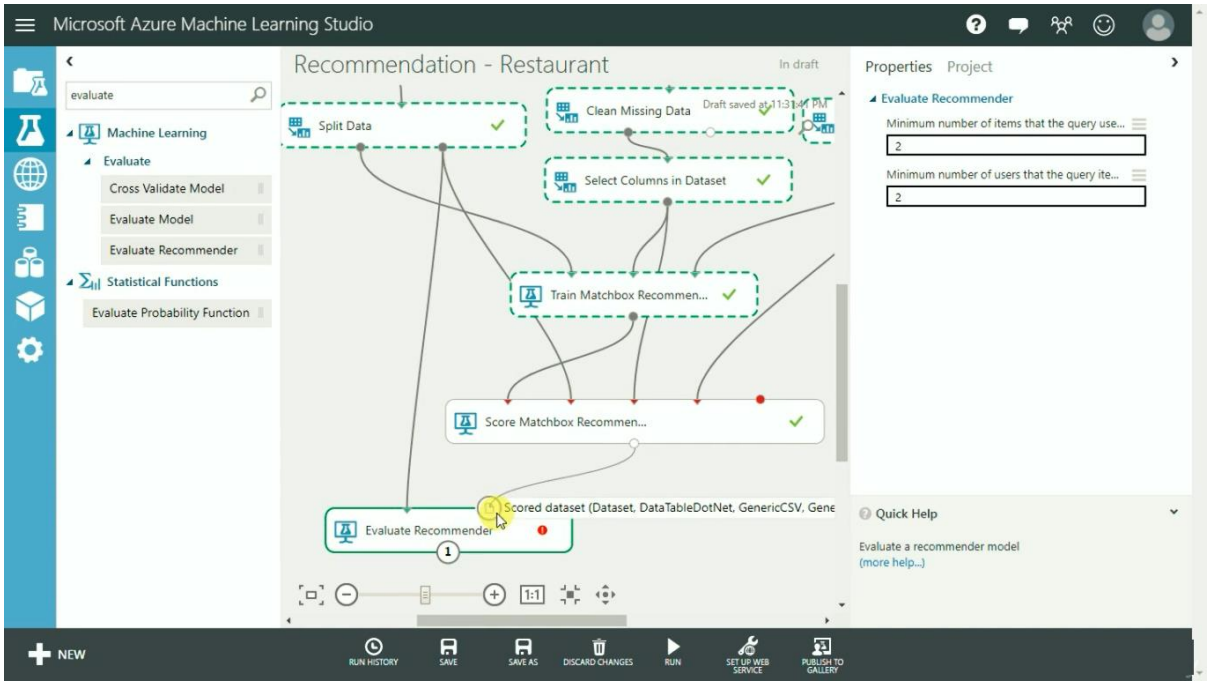
Visualize the result



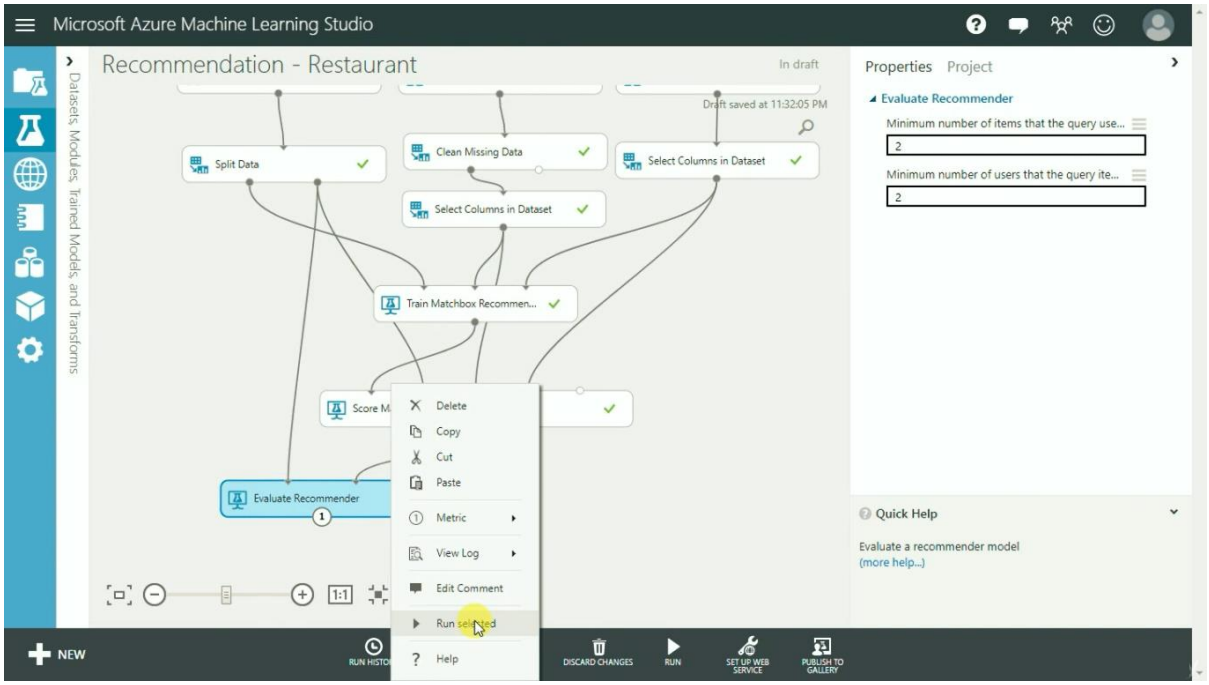
We have obtained the result . However, we have to evaluate the same



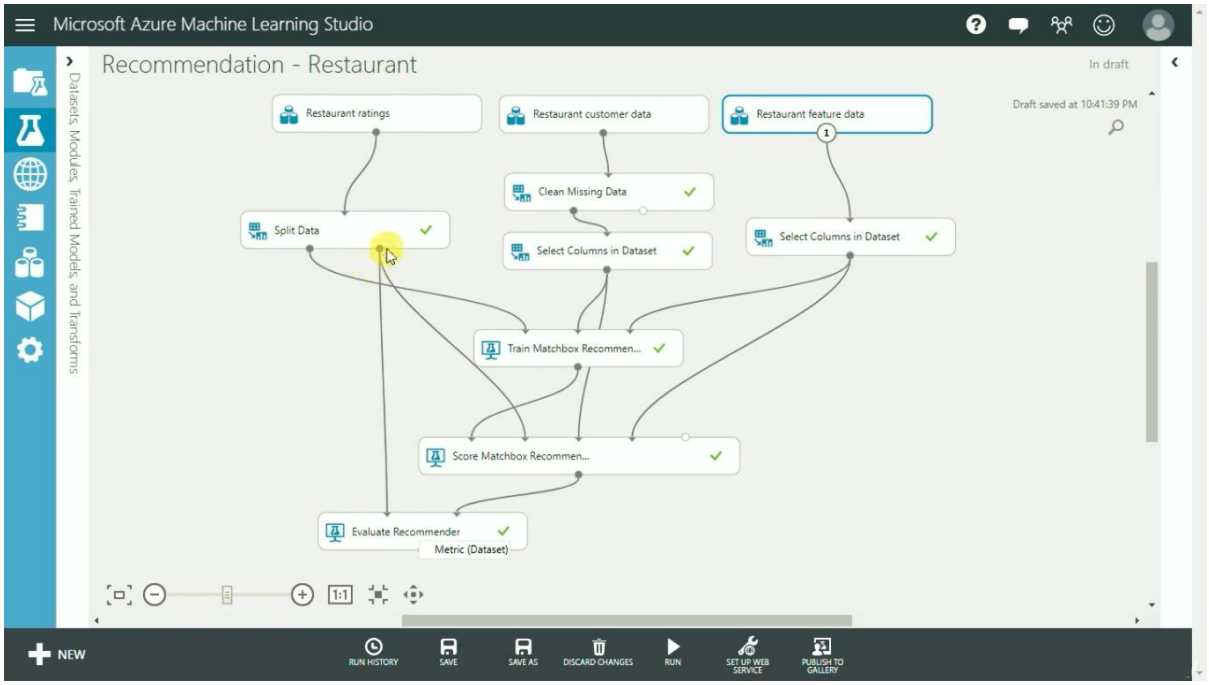
Insert Evaluate recommender and connect the same as below



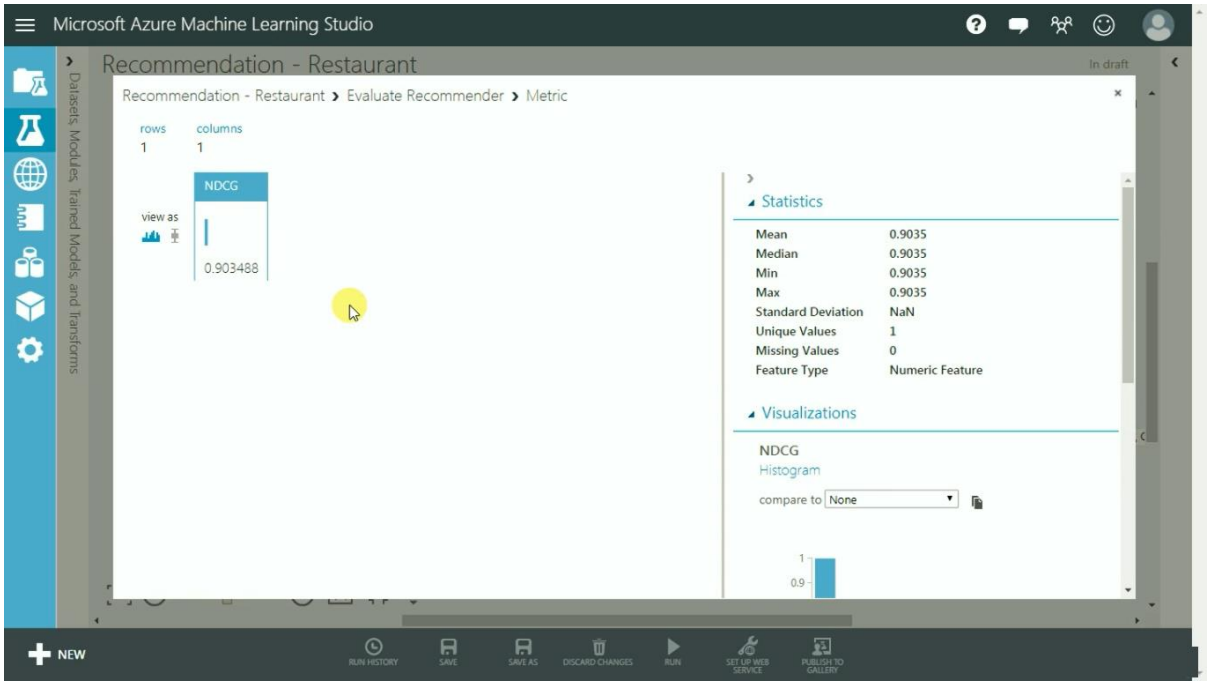
Run the result



UNDERSTANDING THE MATCHBOX RECOMMENDATION RESULTS



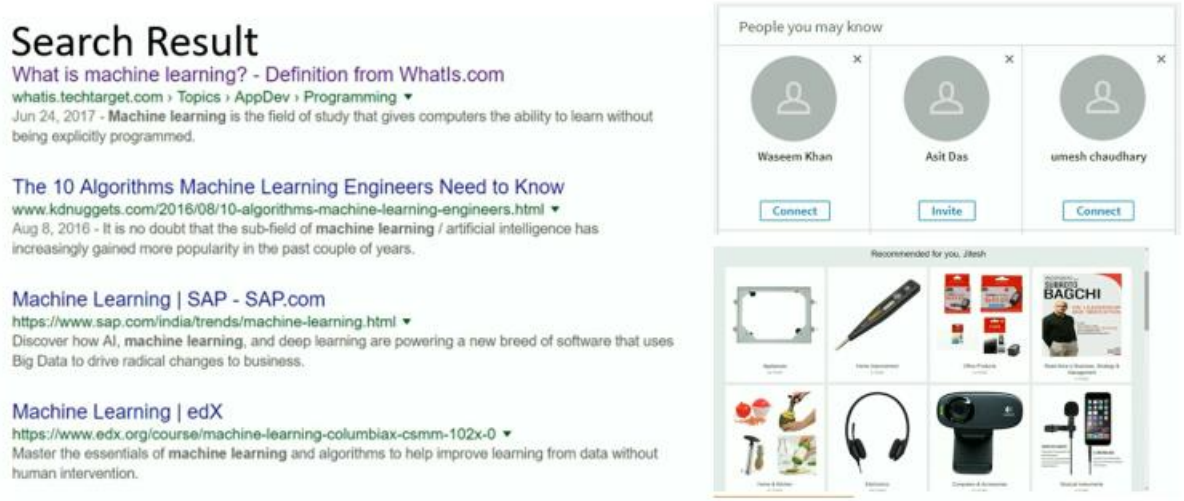
Results



Understanding the Recommender Result

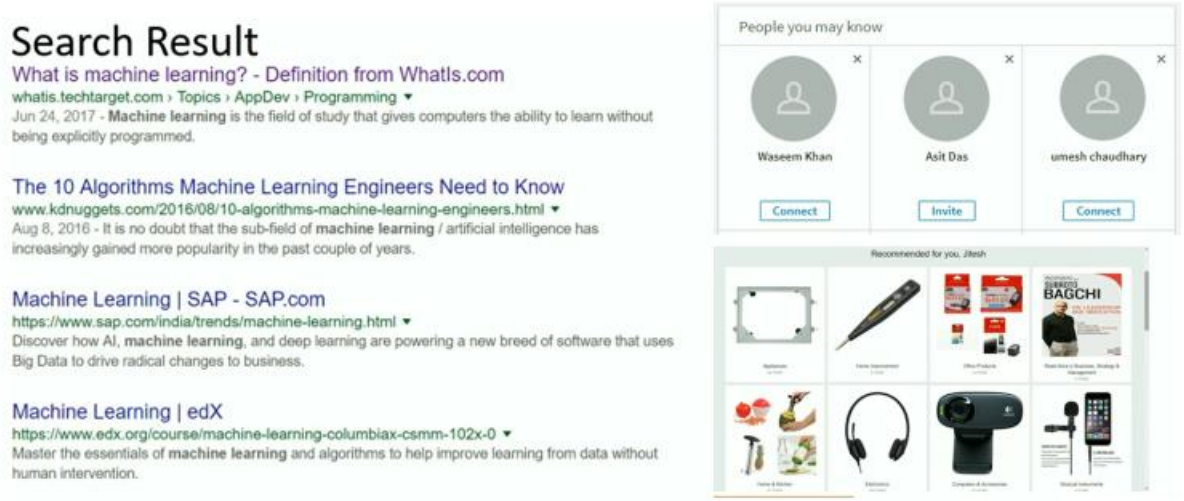
- For Item Recommendations
- Normalized Discounted Cumulative Gain (NDCG)

Ranking Quality



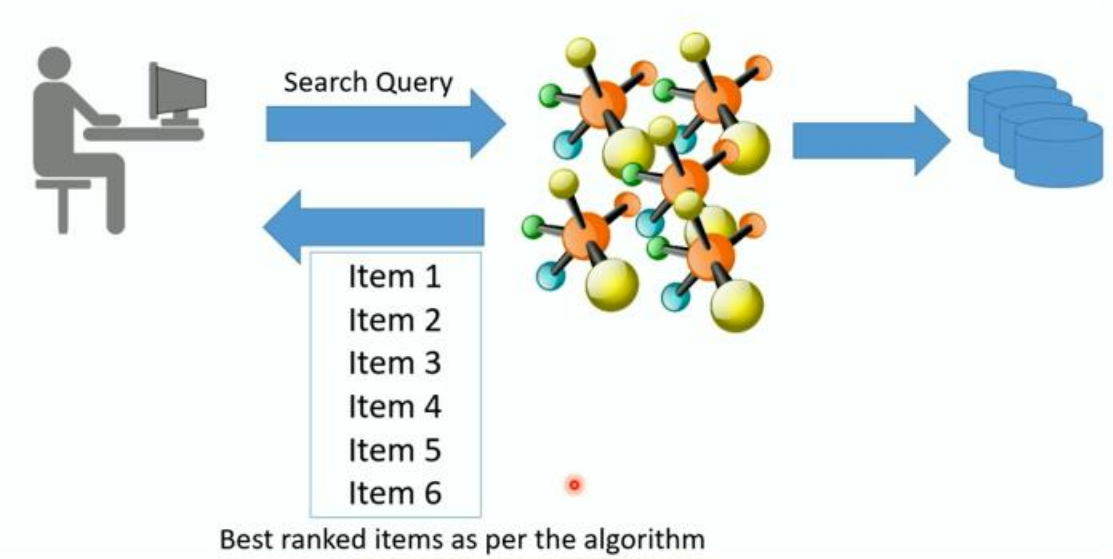
The highest ranked item should result in highest gain.

Ranking Quality



The highest ranked item should result in highest gain.

Ranking Quality



Best ranked items as per the algorithm

Discounted Cumulative Gain

Ranked items by the algorithms → I1, I2, I3, I4, I5, I6

Gain perceived by the user → 4, 3, 4, 0, 1, 2

$CG = 4 + 3 + 4 + 0 + 1 + 2 = 14$

i	rel_i	$\log_2 (i+1)$	$\frac{rel_i}{\log_2 (i+1)}$
1	4	1	4.00
2	3	1.585	1.89
3	4	2	2
4	0	2.322	0
5	1	2.585	0.39
6	2	2.81	0.71
		DCG	**8.99**

Ideal Discounted Cumulative Gain

Ideal Ranking by algorithm → I1, I3, I2, I6, I5, I4

Ideal Gain perceived by the user → 4, 4, 3, 2, 1, 0

$CG = 4 + 4 + 3 + 2 + 1 + 0 = 14$

i	rel_i	$\log_2 (i+1)$	$\frac{rel_i}{\log_2 (i+1)}$
1	4	1	4.00
2	4	1.585	2.52
3	3	2	1.5
4	2	2.322	0.86
5	1	2.585	0.38
6	0	2.81	0
		IDCG	**9.27**

Normalised DCG

DCG = 8.99 IDCG = 9.27

$$NDCG = \frac{DCG}{IDCG} = \frac{8.99}{9.27} = 0.9697$$

Highly relevant documents are more useful than marginally relevant documents, which are in turn more useful than non-relevant documents

Understanding the Result

- Normalized Discounted Cumulative Gain (NDCG)