Azure ML Classic Studio

Predicting Adult Income using Regression Model in AzureML Classic Studio.

This model (Pipeline) trains a linear regressor to predict a adult income prediction based on technical features.
Because you're trying to answer the question "How much?" this is called a regression problem.
However, you can apply the same fundamental steps in this example to tackle any type of machine learning problem whether it be regression, classification, clustering, and so on.

Gallery Link:

Adult Income prediction [19K45A0594]

Adult Income prediction using pre-available dataset and training the model using Linear Regression. Tags: Linear Regression, Azure ML, Microsoft

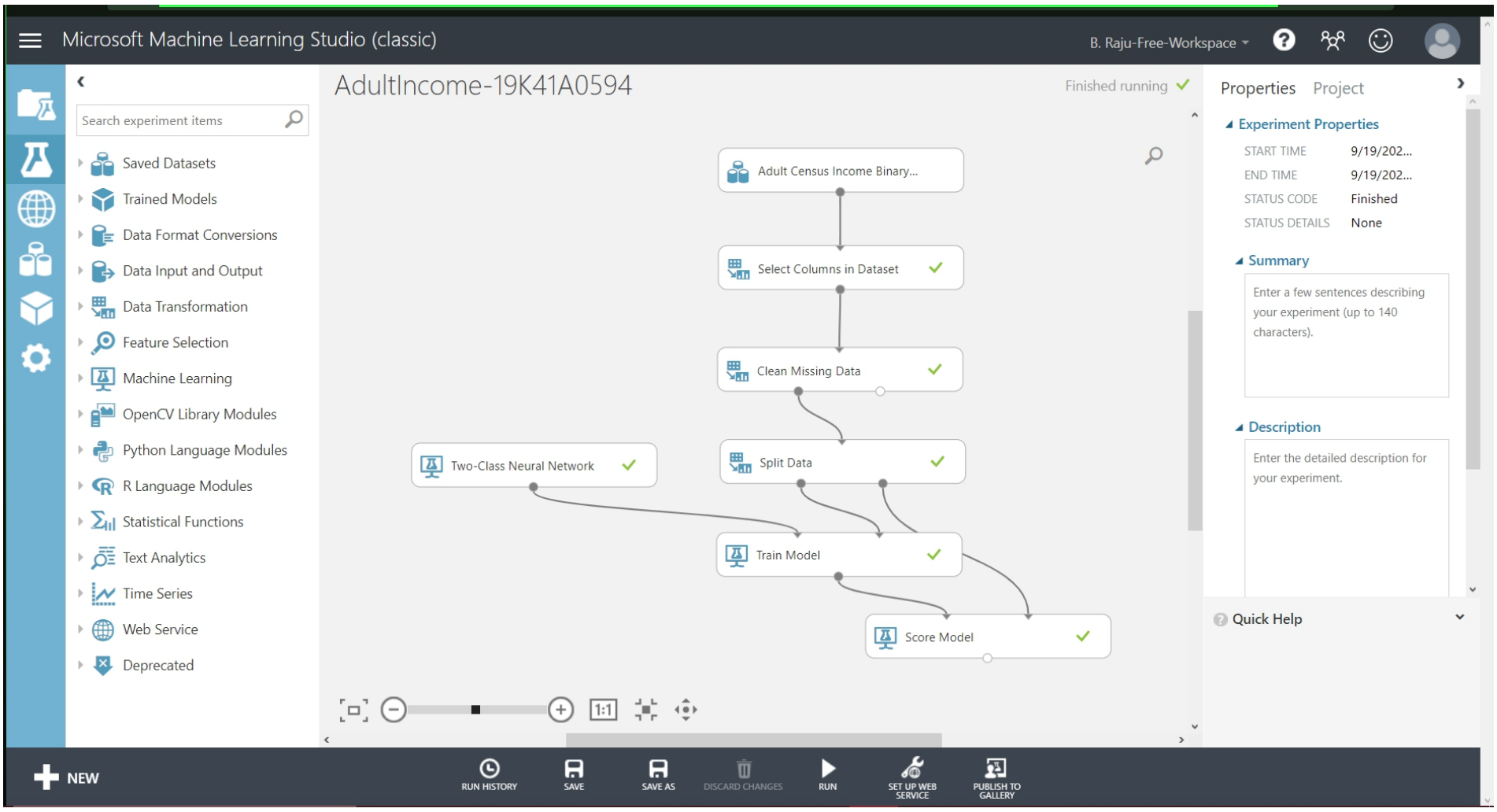
https://studio.azureml.net/Home/ViewWorkspaceCached/3375e3c3cf7340e89b23ca9dba1ba48a#Workspaces/Experiments/Experiment/3375e3c3cf7340e89b23ca9dba1ba48a.f-id.338270b762404f1aa394eaaf55274d26/ViewExperiment

navigate to the link to see the Workflow and you can download the project as well.

Machine Learning Project Workflow

1. **Import Data**
2. **Explore Data (Missing values, outliers)**
3. **Preprocess data (Missing value imputation, outlier treatment, normalization)**
4. **Model Selection**
5. **Model Training**
6. **Model Testing**
7. **Model Deployment**

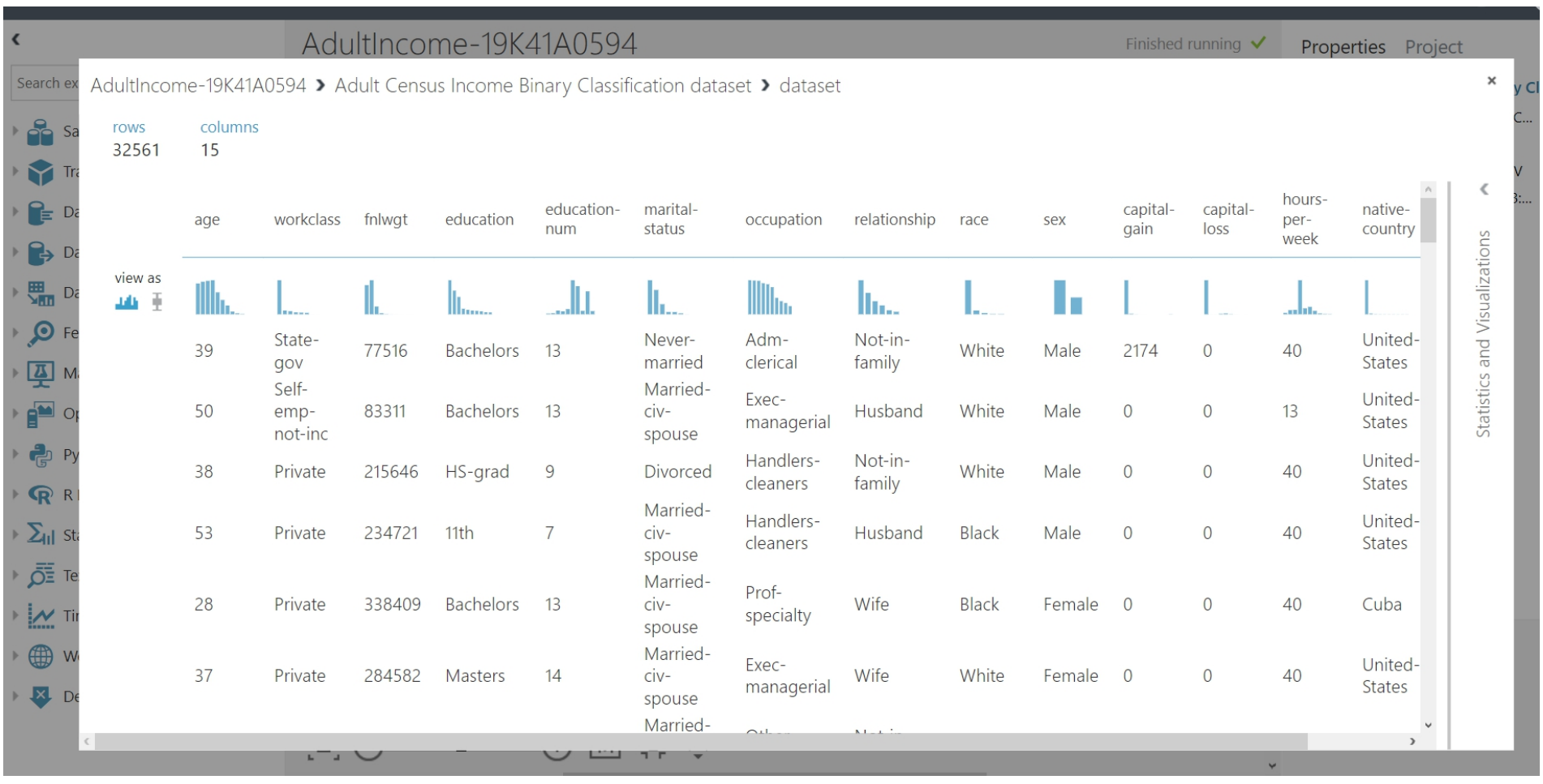
Workflow



Project Workflow

Import Data:

- importing the RAW dataset which is in CSV format.
- the dataset is pre-available in the Azure ML Classic Studio.

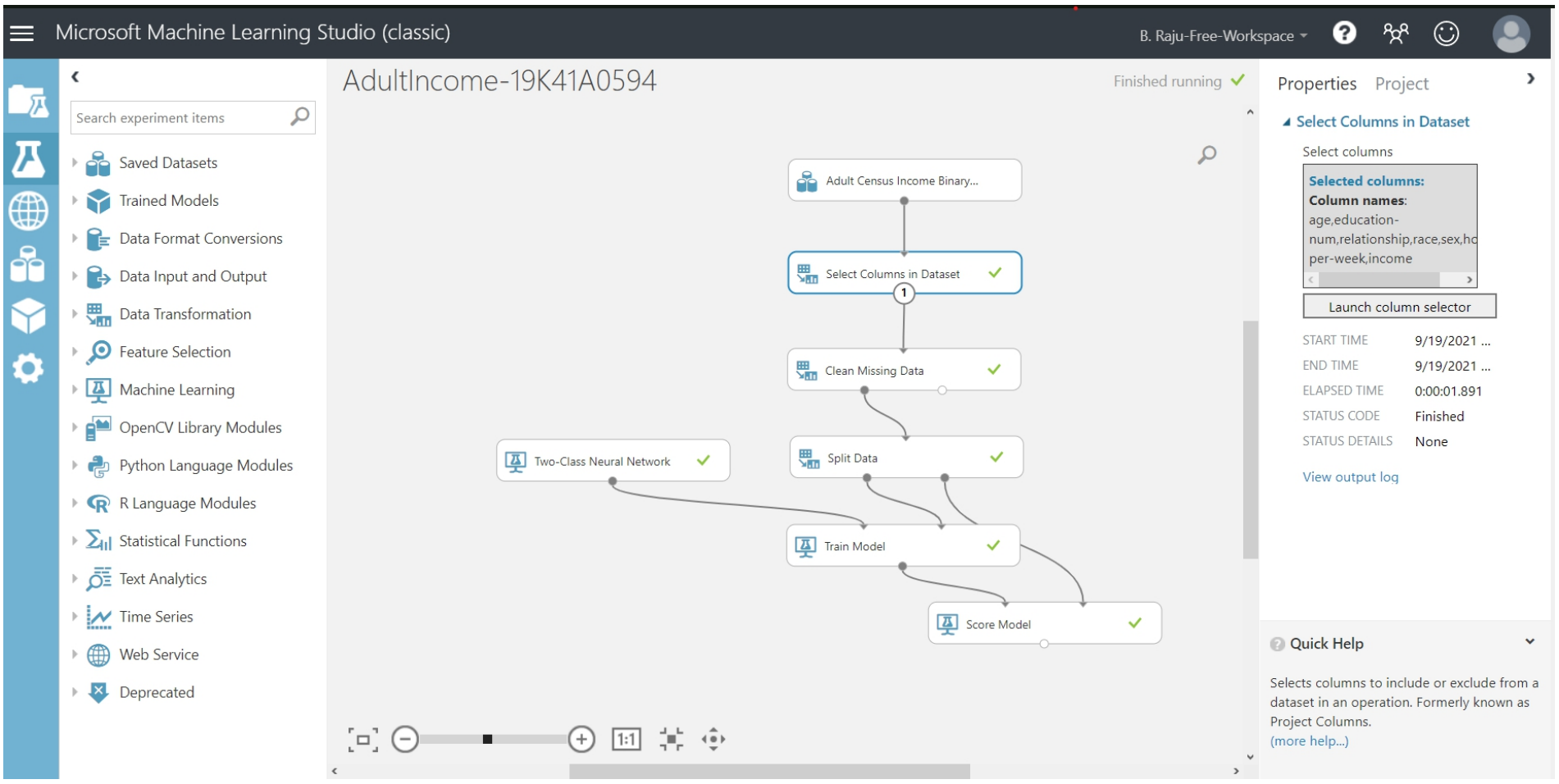


	age	workclass	fnlwgt	education	education-num	marital-status	occupation	relationship	race	sex	capital-gain	capital-loss	hours-per-week	native-country
view as														
	39	State-gov	77516	Bachelors	13	Never-married	Adm-clerical	Not-in-family	White	Male	2174	0	40	United-States
	50	Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male	0	0	13	United-States
	38	Private	215646	HS-grad	9	Divorced	Handlers-cleaners	Not-in-family	White	Male	0	0	40	United-States
	53	Private	234721	11th	7	Married-civ-spouse	Handlers-cleaners	Husband	Black	Male	0	0	40	United-States
	28	Private	338409	Bachelors	13	Married-civ-spouse	Prof-specialty	Wife	Black	Female	0	0	40	Cuba
	37	Private	284582	Masters	14	Married-civ-spouse	Exec-managerial	Wife	White	Female	0	0	40	United-States

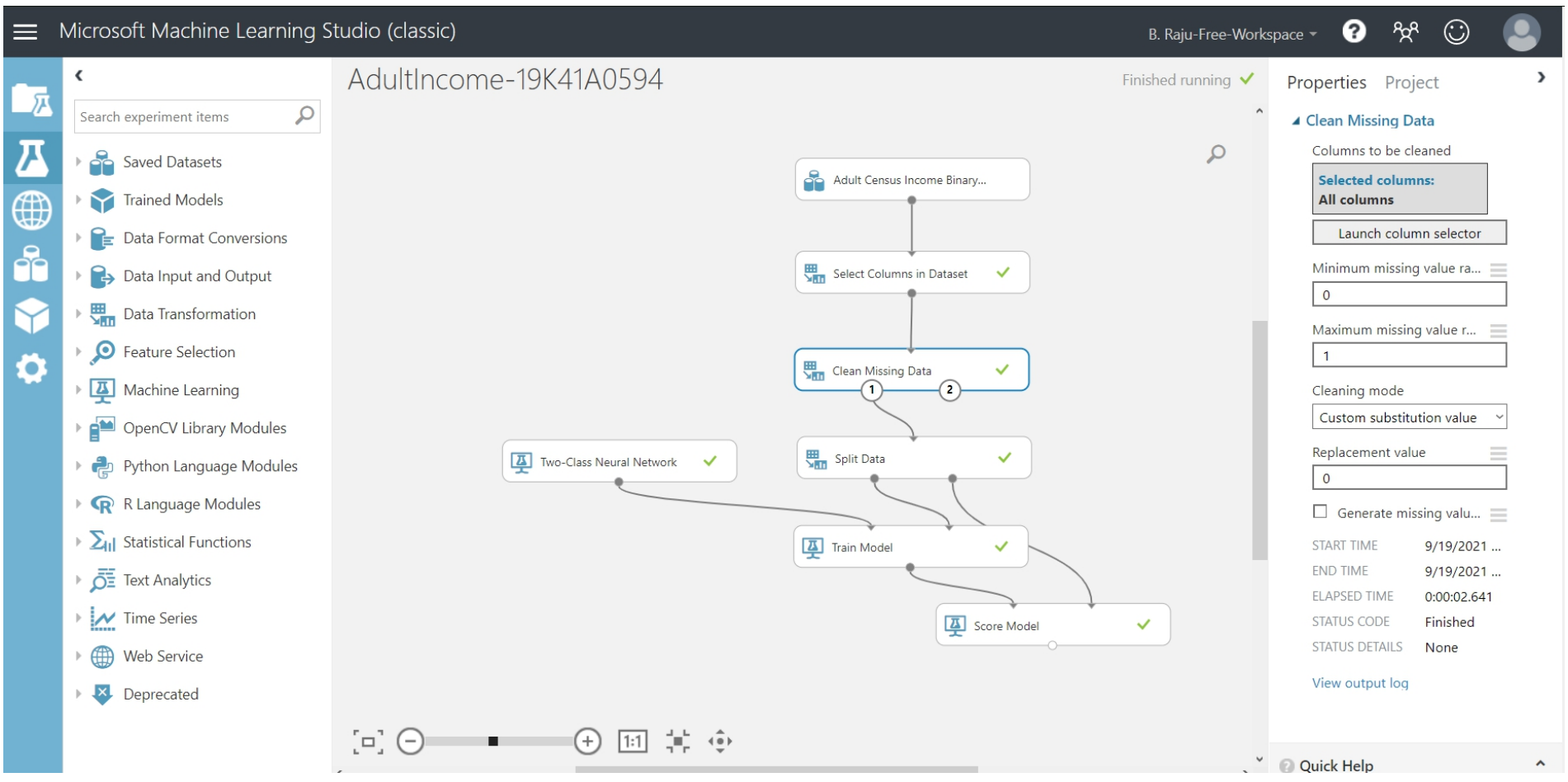
Adult Income Prediction RAW dataset (CSV format)

Explore Data

- this basically includes data visualization to search for any missing values in the Dataset.
- if any missing values are found, then they needs to be cleaned.
- selecting the required columns and clean the data using the Clean Missing Value module (Just Drag n' Drop)



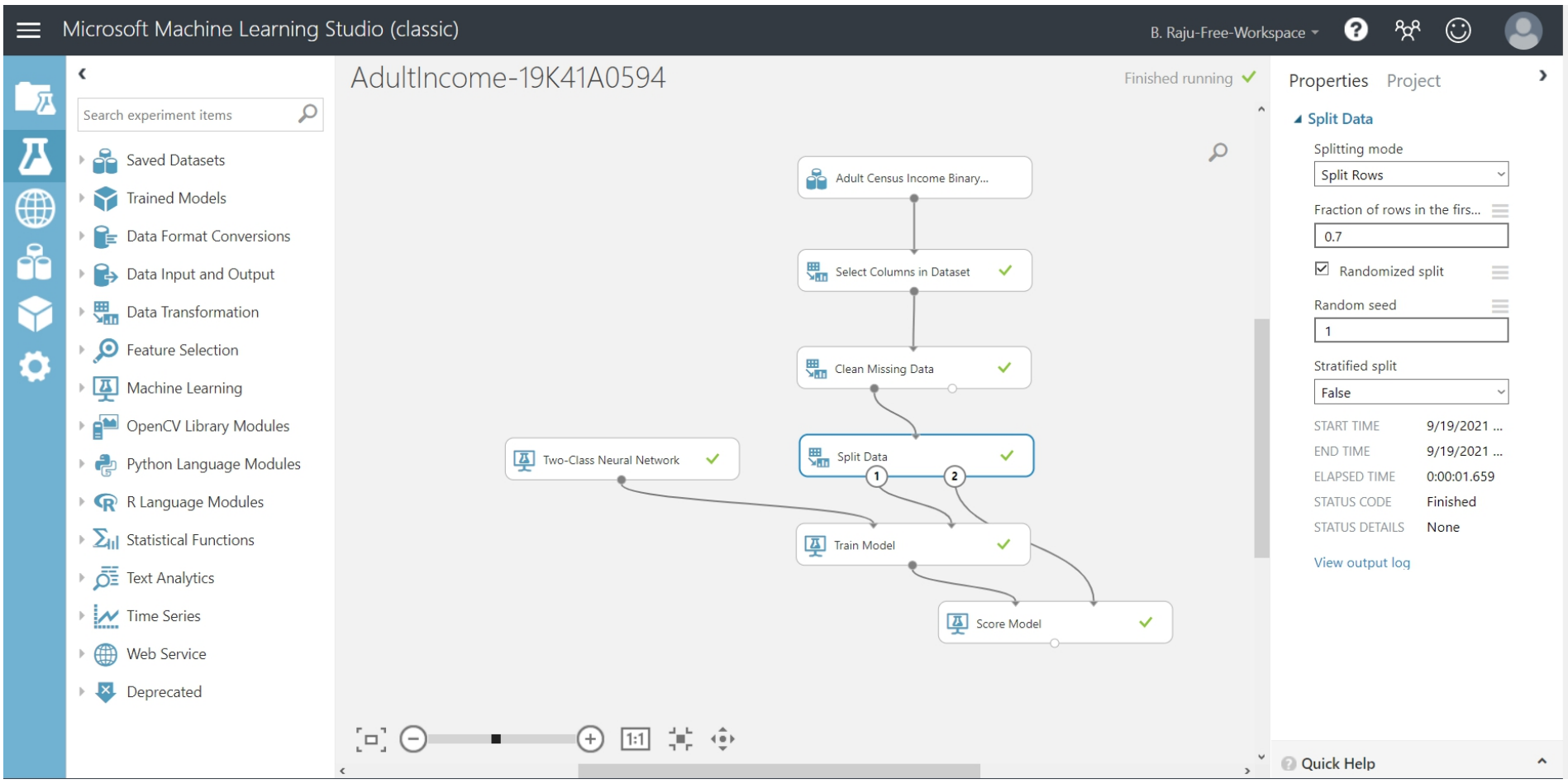
as the normalized loss has 41 missing values in the Dataset, those missing values are to be cleaned.



Data Cleaning

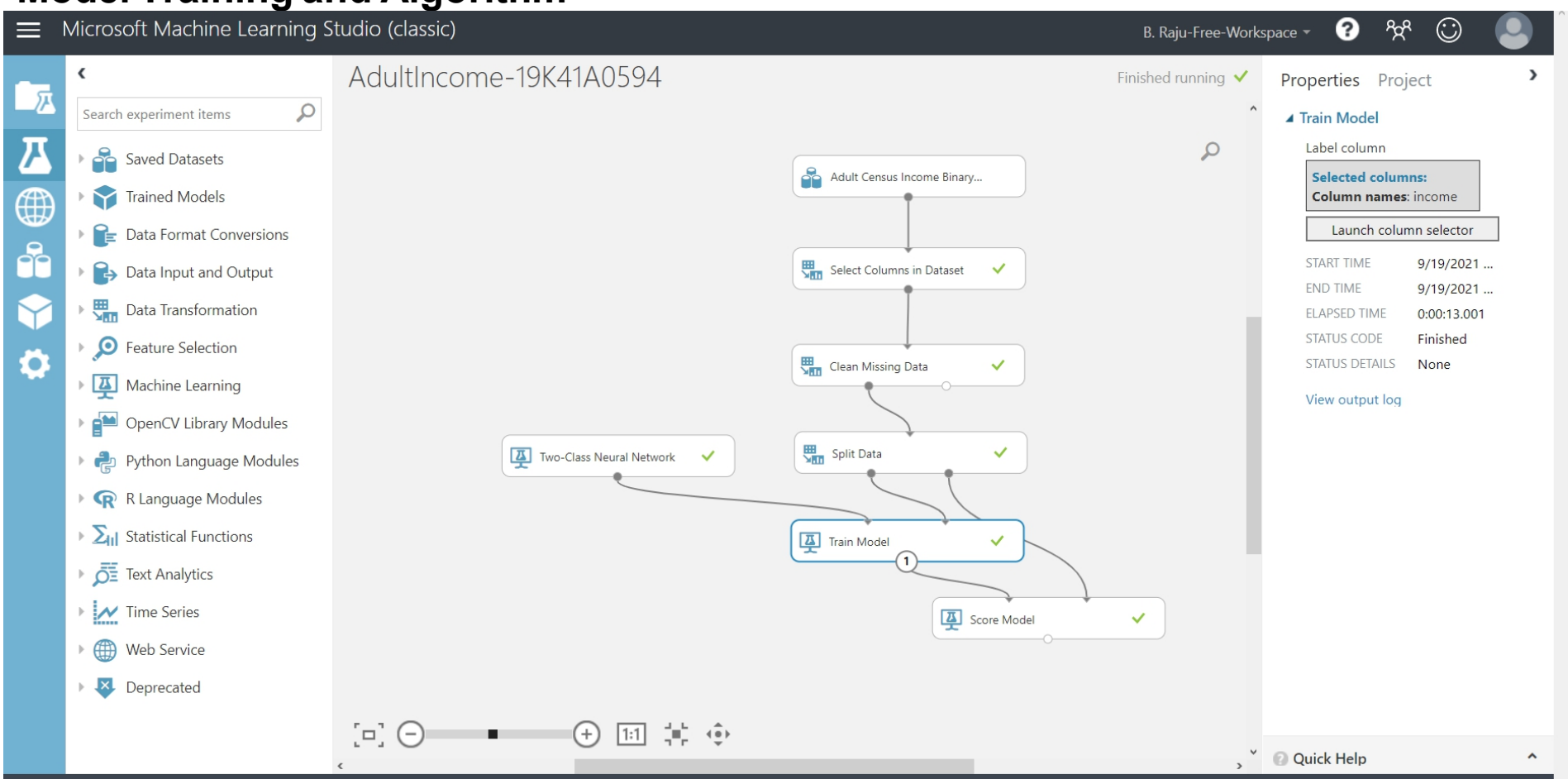
Split Data

- Use the Split Data module to randomly divide the input data so that the training dataset contains 70% of the original data and the testing dataset contains 30% of the original data.



Data Splitting

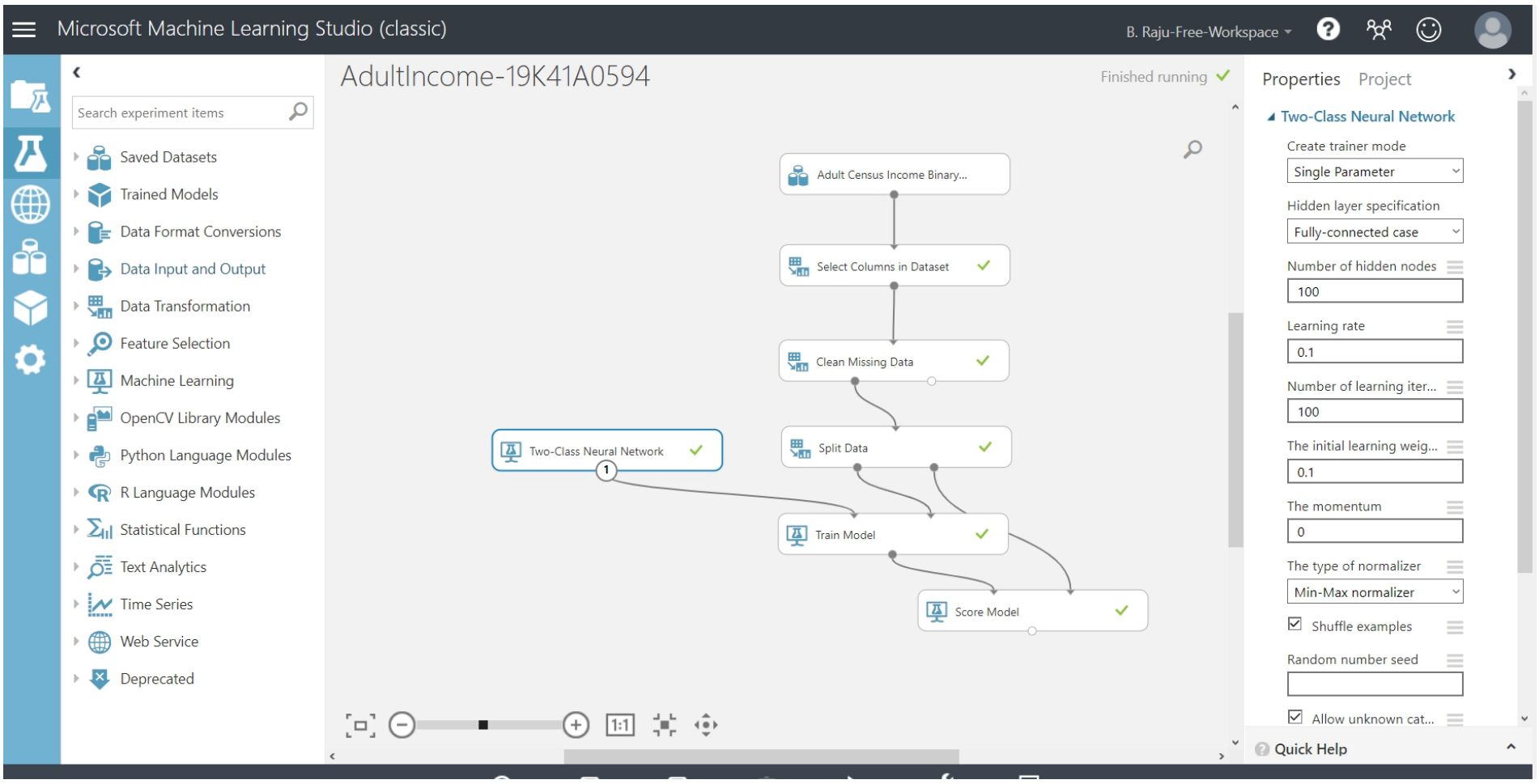
Model Training and Algorithm



Model Training

Using two-class neural network to train the model

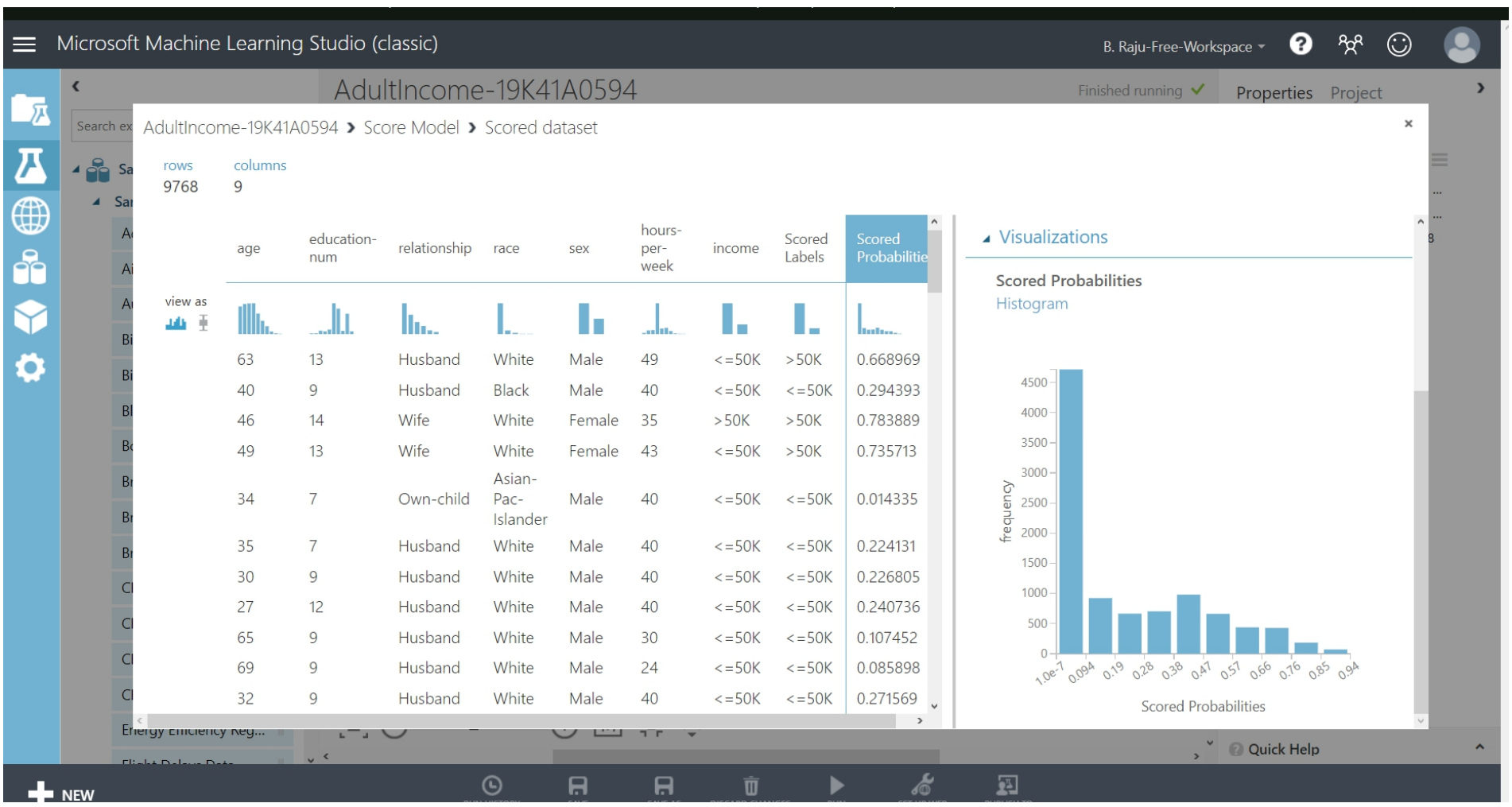
- Since the goal of this sample is to predict adult income prediction, and because the label column (income) is continuous data,We use two-class neural network for this pipeline.



Linear Regression

Score Model and Evaluate Model

- After the model is trained, we can use the Score Model module to generate predicted results.



Score Labels