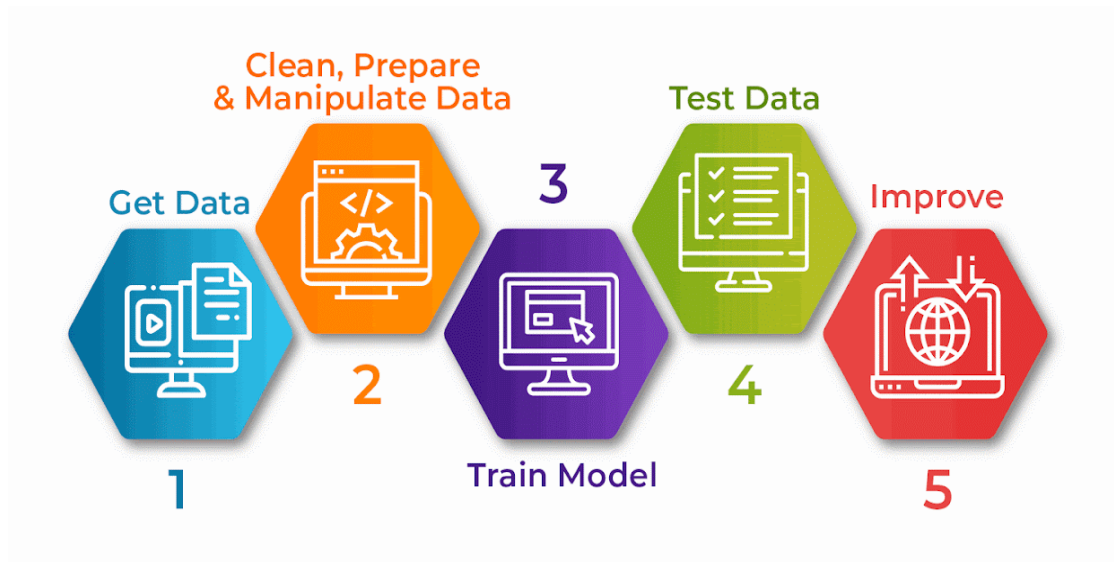




## AZURE ML STUDIO

### REGRESSION MODEL

#### Machine Learning Workflow



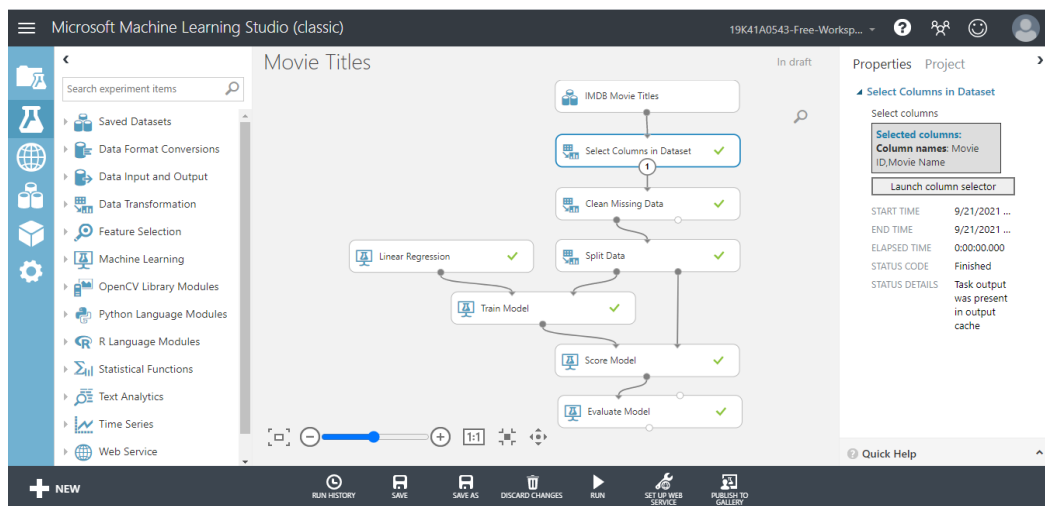
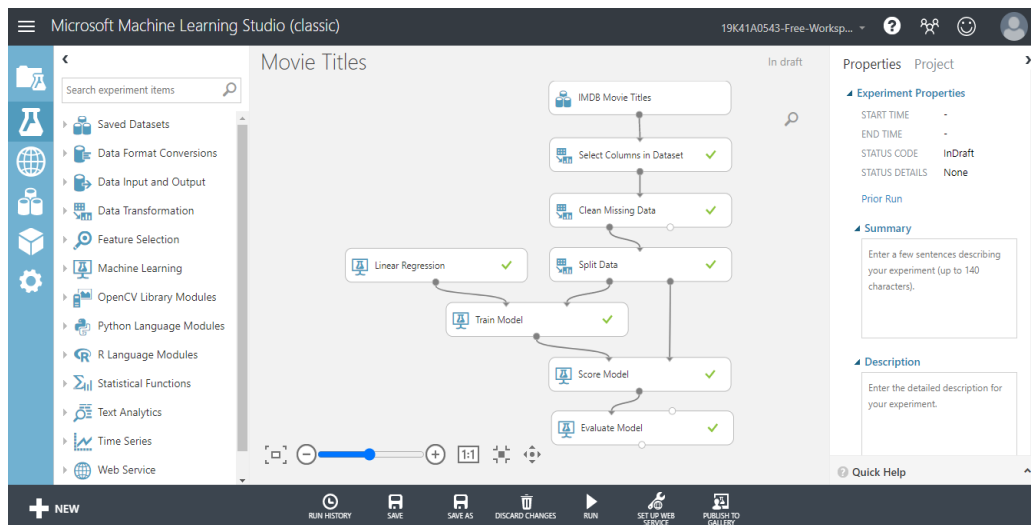
#### IMDB Movie Titles Using Regression Model

Regression model is used to predict the required solution for a problem statement based on the available data.

#### Workflow

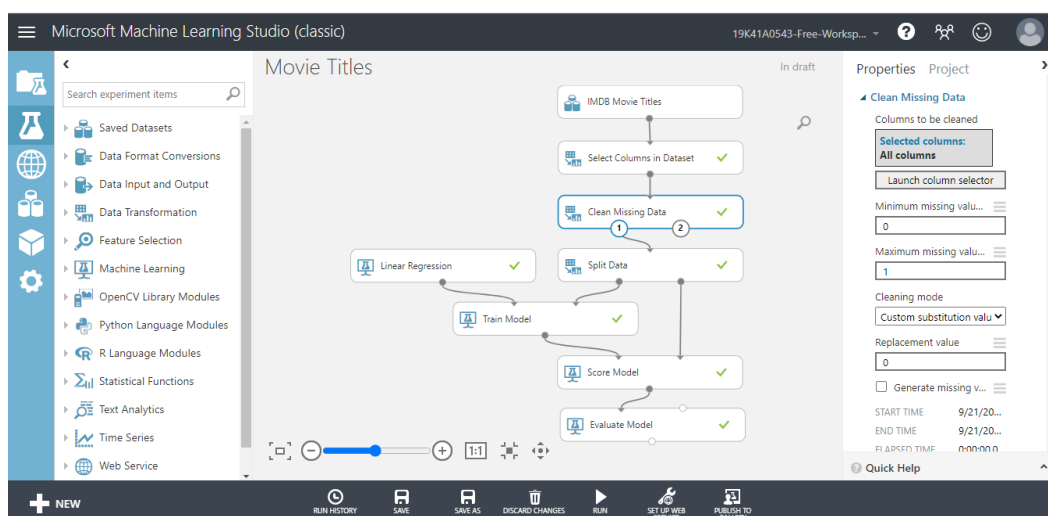
- Data collection and prepare data
- Develop the model
- Train an ML model on data
- Train model
- Evaluate model accuracy
- Deploy trained model
- Send prediction requests to your model
- Manage your models
- Upgrade or improve data

## Collect Data & Prepare Data



## Clean & Manipulate Data

Missing data in the dataset can be ignored and removed or filled manually, basically the mean, median or highest frequency value.



## Split Data

The screenshot shows the Microsoft Machine Learning Studio interface. The workflow is titled "Movie Titles" and is in draft status. The workflow steps are: IMDb Movie Titles, Select Columns in Dataset, Clean Missing Data, Split Data, Train Model, Score Model, and Evaluate Model. The Split Data module is highlighted, and its properties are shown on the right. The properties include: Splitting mode (Split Rows), Fraction of rows in the ... (0.5), Randomized split (checked), Random seed (0), Stratified split (False), START TIME (9/21/20...), END TIME (9/21/20...), ELAPSED TIME (0:00:00.0...), STATUS CODE (Finished), and STATUS DETAILS (Task output was present in ...).

## Train Model

Train is a method to measure the accuracy of your model.

The screenshot shows the Microsoft Machine Learning Studio interface. The workflow is titled "Movie Titles" and is in draft status. The workflow steps are: IMDb Movie Titles, Select Columns in Dataset, Clean Missing Data, Split Data, Train Model, Score Model, and Evaluate Model. The Train Model module is highlighted, and its properties are shown on the right. The properties include: Label column (Selected columns: Column names: Movie Name), START TIME (9/21/2021 ...), END TIME (9/21/2021 ...), ELAPSED TIME (0:00:00.000), STATUS CODE (Finished), and STATUS DETAILS (Task output was present in output cache).

The screenshot shows the Microsoft Machine Learning Studio interface. The workflow is titled "Movie Titles" and is in draft status. The workflow steps are: IMDb Movie Titles, Select Columns in Dataset, Clean Missing Data, Split Data, Train Model, Score Model, and Evaluate Model. The Train Model module is highlighted, and its properties are shown on the right. The properties include: Label column (Selected columns: Column names: Movie Name), START TIME (9/21/2021 ...), END TIME (9/21/2021 ...), ELAPSED TIME (0:00:00.000), STATUS CODE (Finished), and STATUS DETAILS (Task output was present in output cache).

Batch Linear Regressor

Settings

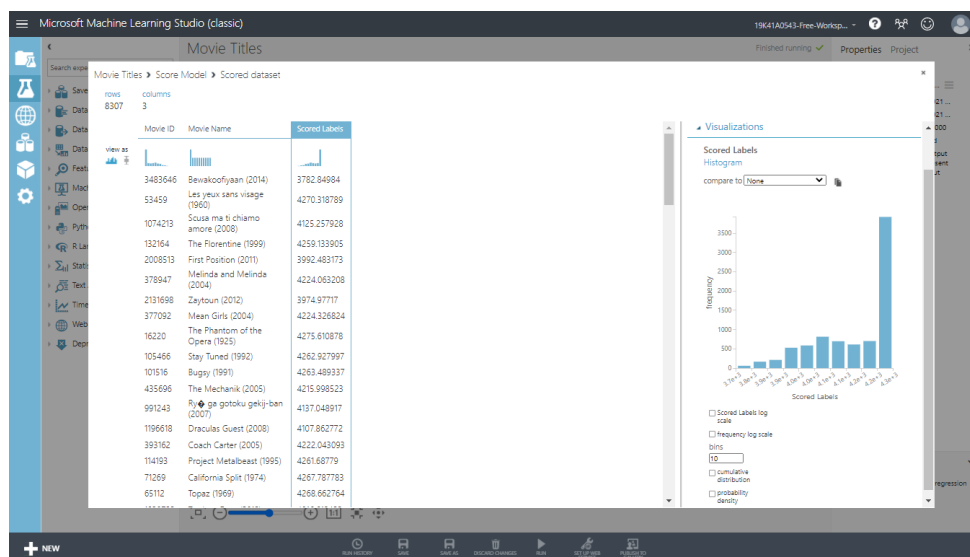
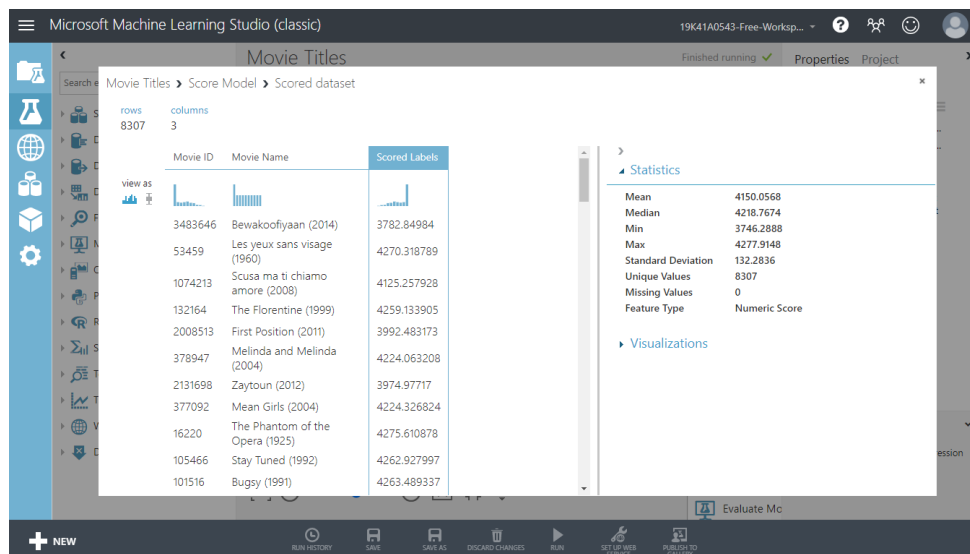
Setting	Value
Bias	True
Regularization	0.001
Allow Unknown Levels	True
Random Number Seed	

Feature Weights

Feature	Weight
Bias	4277.92
Movie ID	-0.000142111

## Score Model

Scoring is the process of generating values based on a trained machine learning model, given some new input data.



# Testing & Evaluation

