Batch: HO – ML 1 **Experiment Number: 07**

Roll Number: 16010422234 **Name: Chandana Galgali**

Aim of the Experiment: Describe the following points with respect to the business under consideration.

Program/ Steps:

1. Problem faced by the business

The system is designed such that it helps to estimate the price of a Car based upon the different features present, so it helps the business to give the correct amount of pricing to their Car Models, so that the customers are attracted and also are satisfied with the price range assigned based upon its features. So, assigning Price range accordingly is challenging without the use of ML algorithms.

2. Approach/ Methodology followed by the business

The methodology which we followed was that we took the dataset of Automobile Price Data. Then we split the dataset into train and test. Then we used Linear Regression and trained the machine learning model with that dataset and in the end we tested that model.

3. Skill Sets, infrastructure and other impact on the business during implementation

Skillset: Cloud computing

Infrastructure: Microsoft Azure Cloud services

There were no as such impacts on the business, as any of the business service lines were not being used.

4. Similar approaches followed by other businesses

Step 1. Sign-in using Microsoft account on studio.azureml.net

Step 2. Creating a workspace for our Machine Learning project.

Step 3. Select New option on bottom right:

Step 4. Click on Blank experiment and write name and summary of experiment

Step 5. Select From Saved Datasets-> Samples-> dataset of your choice

Step 6. Now, search ‘Select columns in dataset’ from items and drag it

Step 7. Now, click on launch column selector-> with rules->exclude column normalized-losses as that column contains many rows/records with empty values.

Step 8. Search and select ‘Clean Missing Data’ from items list

Step 9. Now, select cleaning mode -> Remove entire row as it will remove the entire row wherever missing value is found

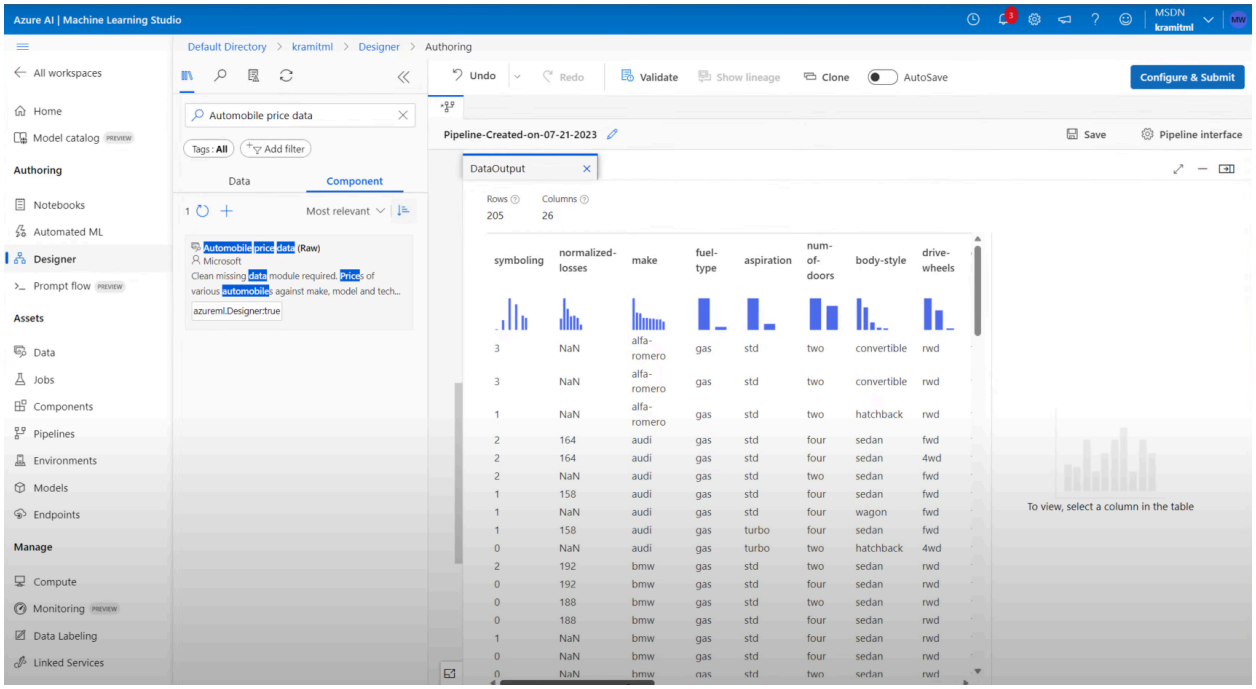
Step 10. Again choose ‘select columns in dataset’

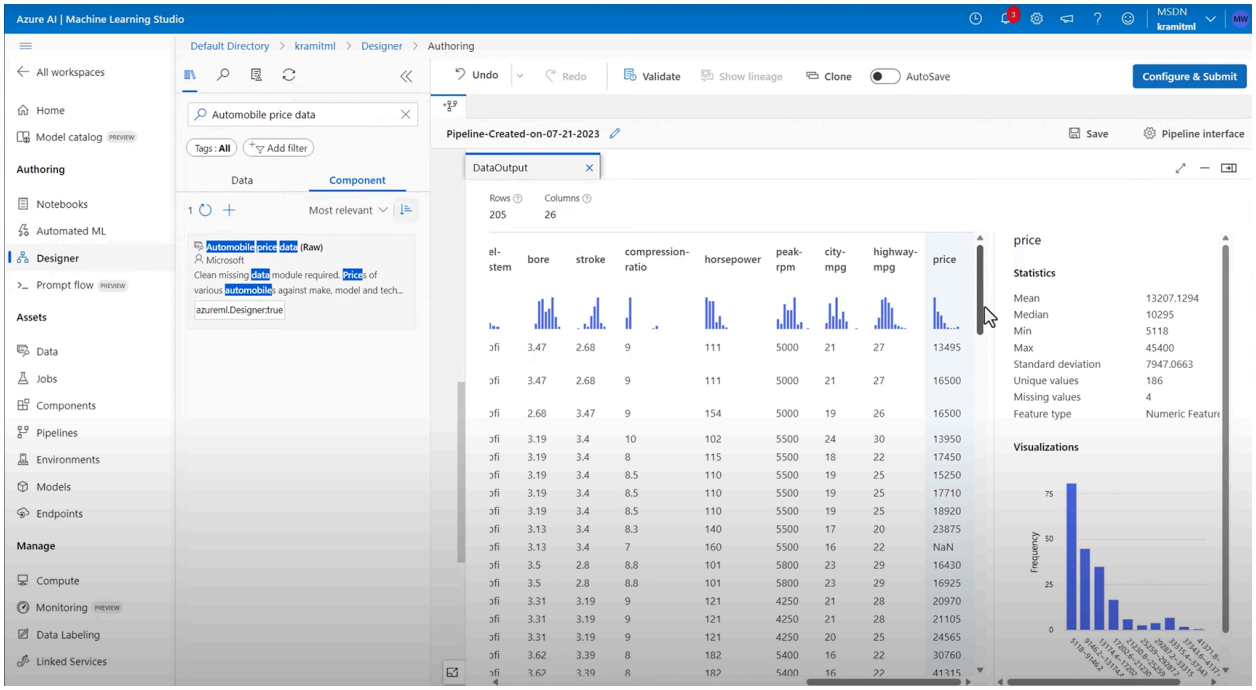
Step 11. Now, launch column selector and include all the columns based on which prediction is to be done: make, body-style, wheel-base, engine-size, horsepower, peak-rpm, highway-mpg, price

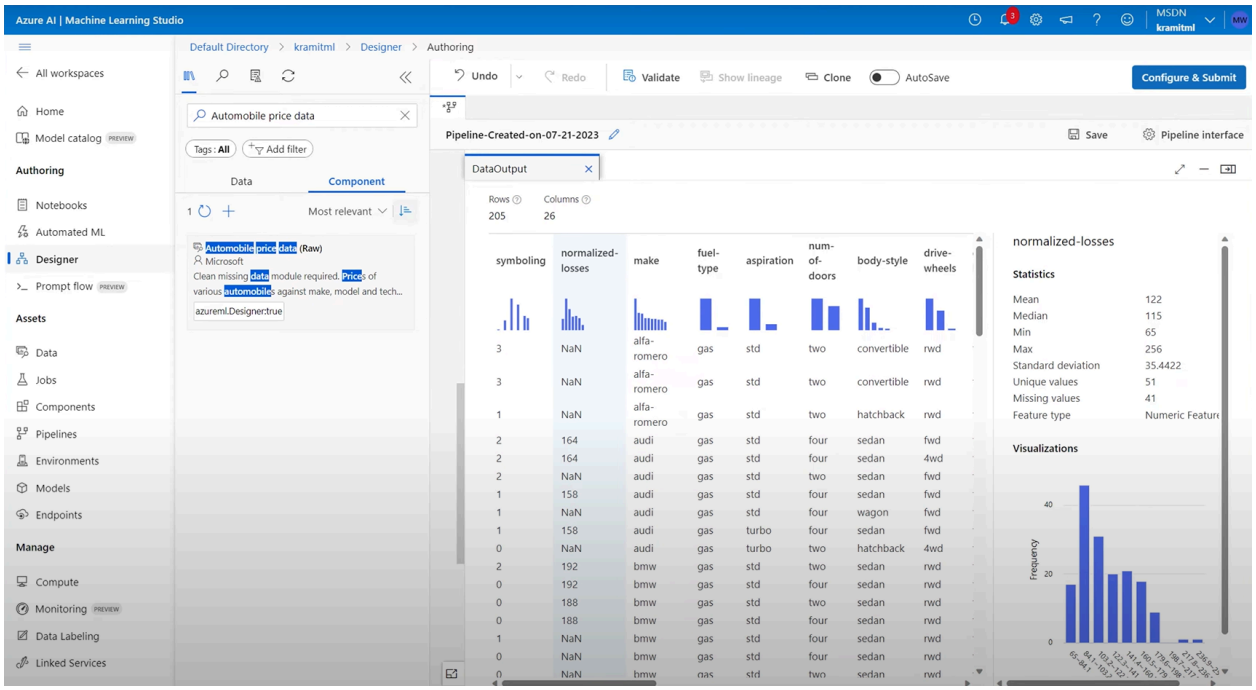
Step 12. Now, select ‘split data’ from list and drag it
Step 13. For Split data, enter the fraction of data which is needed for training while rest will be used for testing
Step 14. Now, Select ‘Linear Regression’ as the algorithm to be used and ‘Train Model’ from list
Step 15. For training model, click on launch column selector, include price column as Price is what is to be predicted
Step 16. Add Score Model from list drag it and make connections
Step 17. Now, Add Evaluate Model from list and make connections
Step 18. Now, Click on Run
Step 19. To check prediction results, right click on Score Model, select visualize
Step 20. To check Evaluation results, right click on Evaluation Model, select visualize

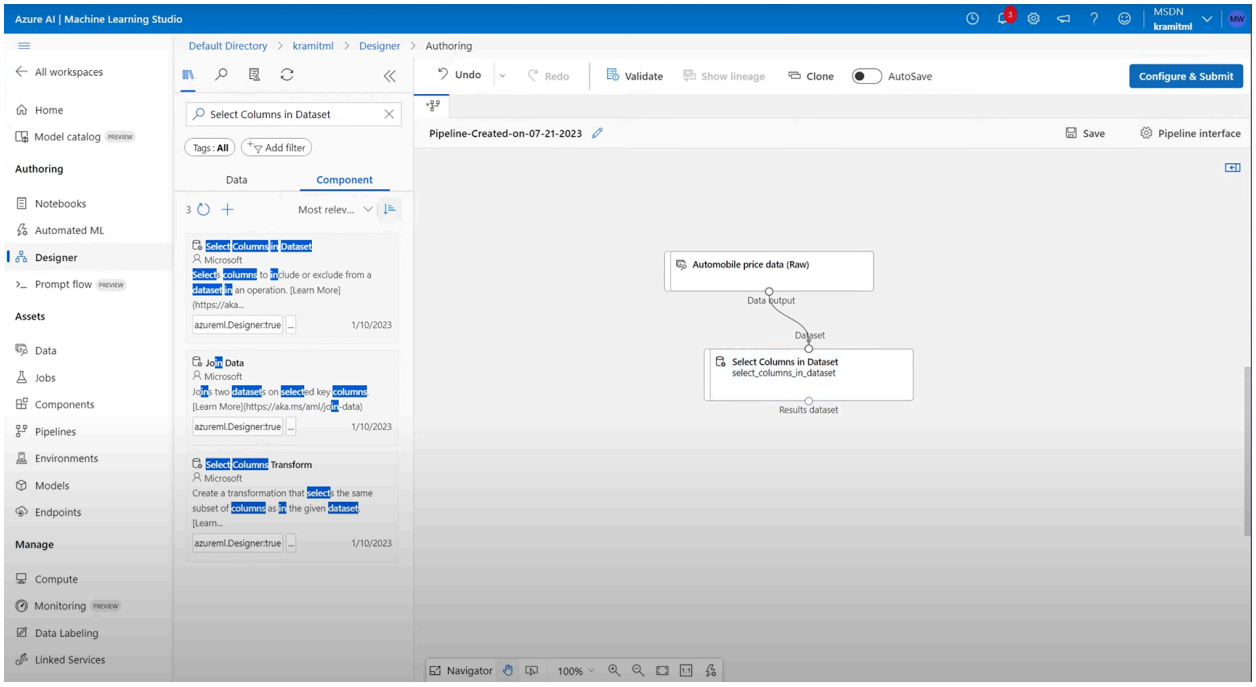
---

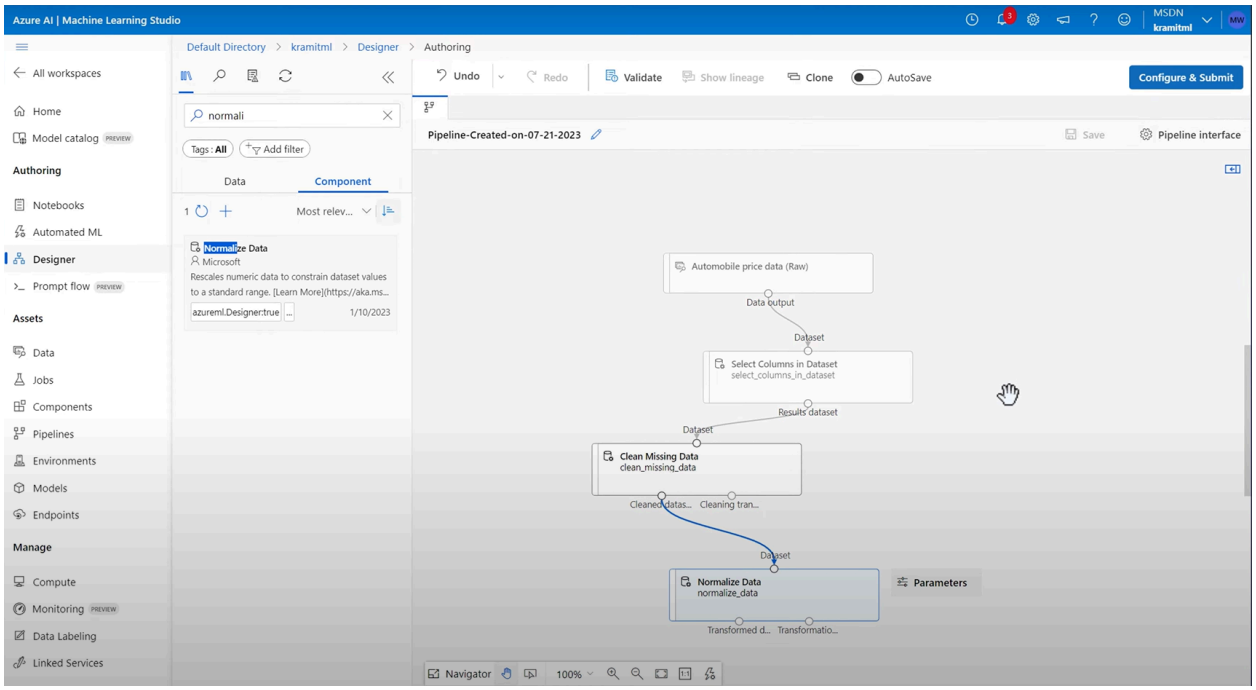
Output/Result:

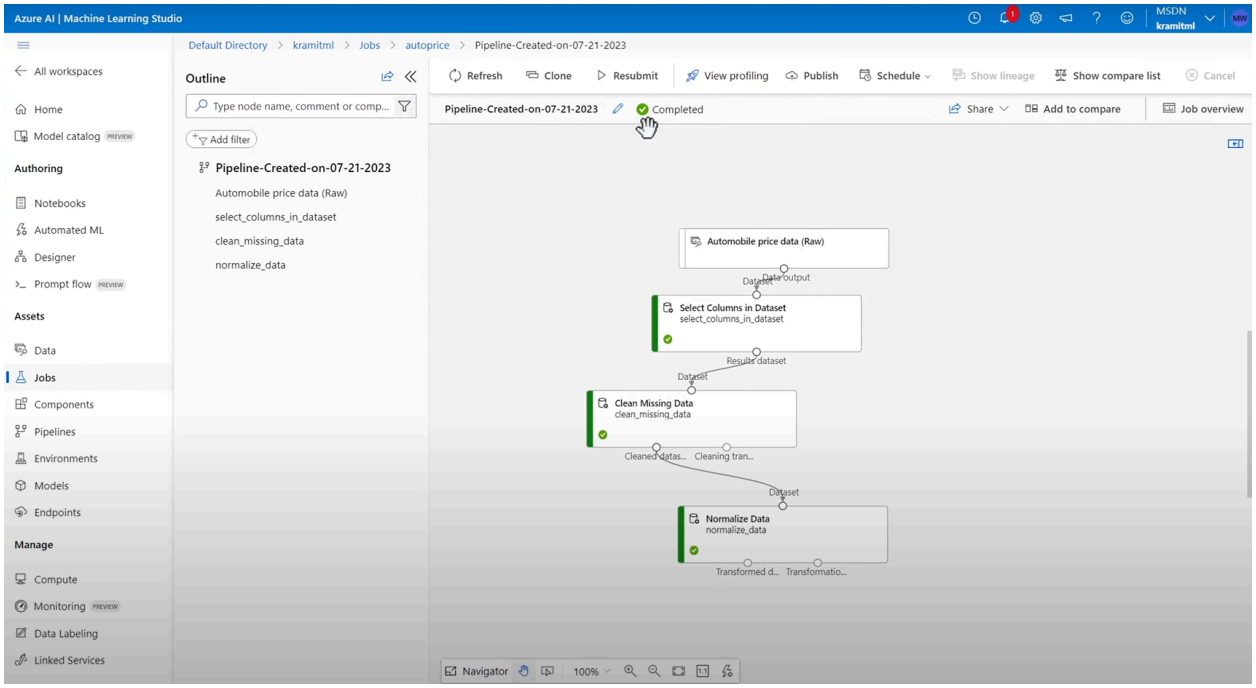


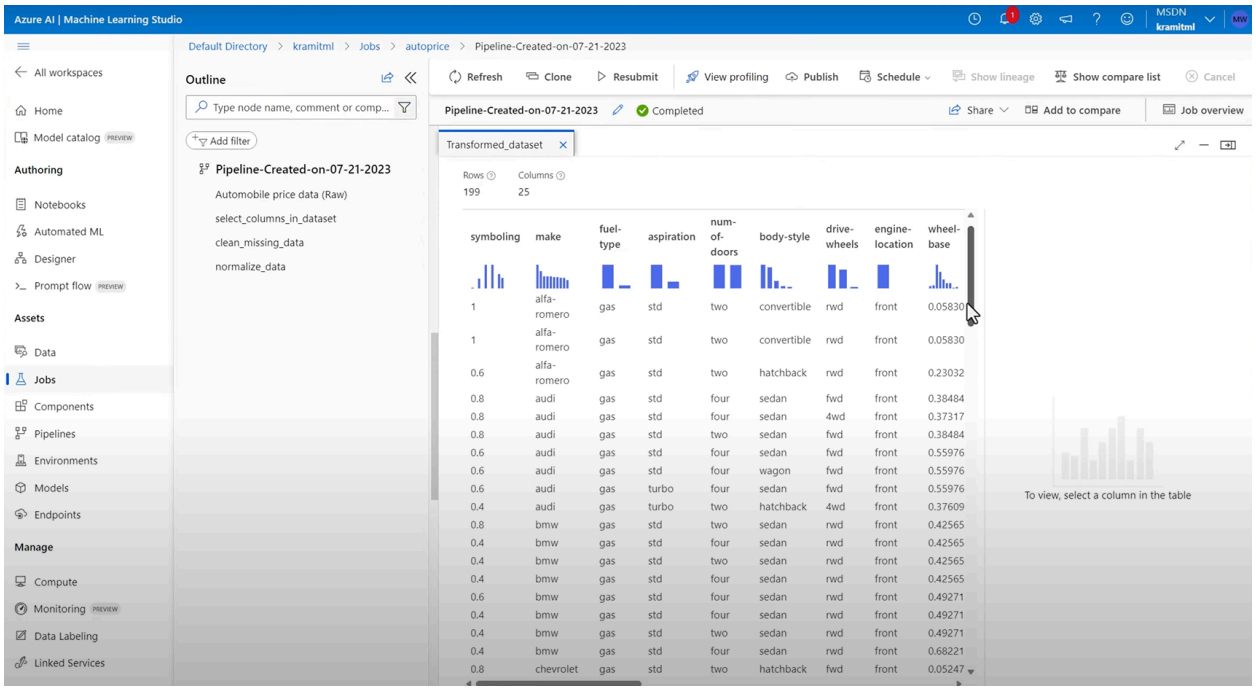

Azure AI | Machine Learning Studio
Default Directory > kramitml > Designer > Authoring
All workspaces
Home
Model catalog
Authoring
Notebooks
Automated ML
Designer
Prompt flow
Assets
Data
Jobs
Components
Pipelines
Environments
Models
Endpoints
Manage
Compute
Monitoring
Data Labeling
Linked Services
Automobile price data
Data
Component
Automobile price data (Raw)
Microsoft
Undo
Redo
Validate
Show lineage
Clone
AutoSave
Configure & Submit
Pipeline-Created-on-07-21-2023
Save
Pipeline interface
DataOutput
Rows 205
Columns 26
price
Statistics
Mean 13207.1294
Median 10295
Min 5118
Max 45400
Standard deviation 7947.0663
Unique values 186
Missing values 4
Feature type Numeric Feature
Visualizations
Frequency

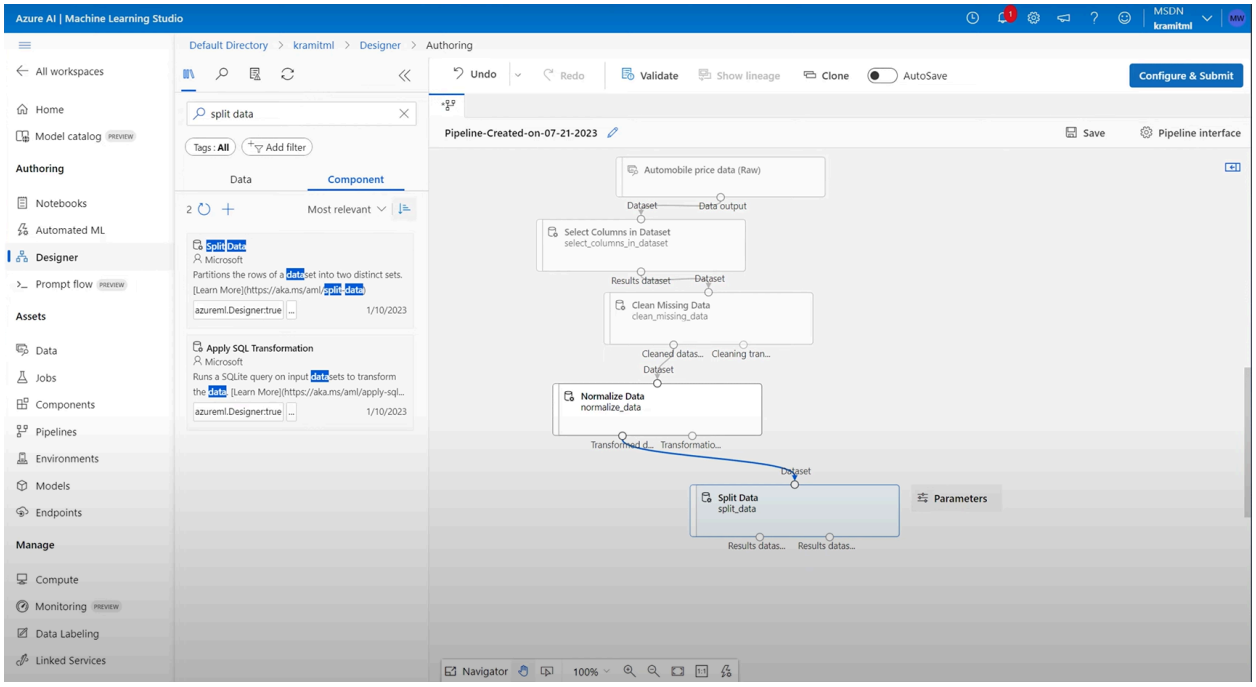

Azure AI | Machine Learning Studio
Default Directory > kramitml > Designer > Authoring
Automobile price data
Pipeline-Created-on-07-21-2023
DataOutput
Rows 205
Columns 26
symboling
normalized-losses
make
fuel-type
aspiration
num-of-doors
body-style
drive-wheels
normalized-losses
Statistics
Mean 122
Median 115
Min 65
Max 256
Standard deviation 35.4422
Unique values 51
Missing values 41
Feature type Numeric Feature
Visualizations
Frequency

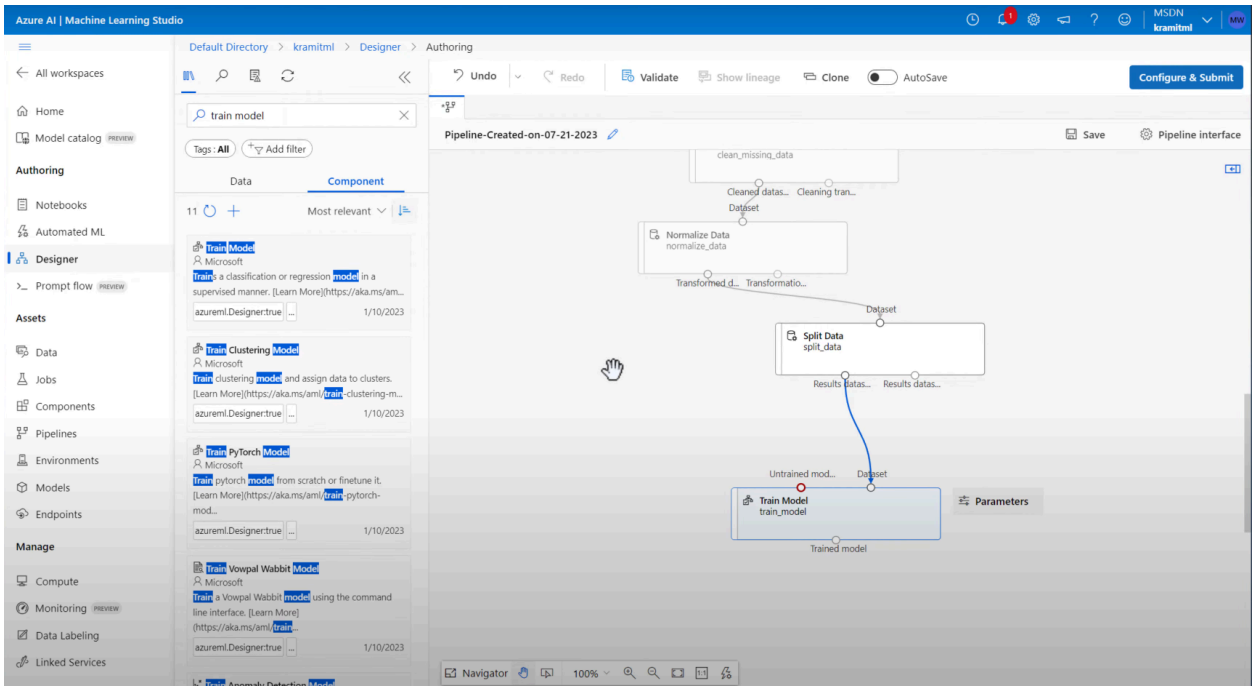

Azure AI | Machine Learning Studio
Default Directory > kramitml > Designer > Authoring
Undo
Redo
Validate
Show lineage
Clone
AutoSave
Configure & Submit
All workspaces
Home
Model catalog
Authoring
Notebooks
Automated ML
Designer
Prompt flow
Assets
Data
Jobs
Components
Pipelines
Environments
Models
Endpoints
Manage
Compute
Monitoring
Data Labeling
Linked Services
Select Columns in Dataset
Tags : All
Add filter
Data
Component
Most relev...
Select Columns in Dataset
Microsoft
Join Data
Microsoft
Select Columns Transform
Microsoft
Pipeline-Created-on-07-21-2023
Save
Pipeline interface
Automobile price data (Raw)
Data output
Dataset
Select Columns in Dataset
select_columns_in_dataset
Results dataset
Navigator
100%

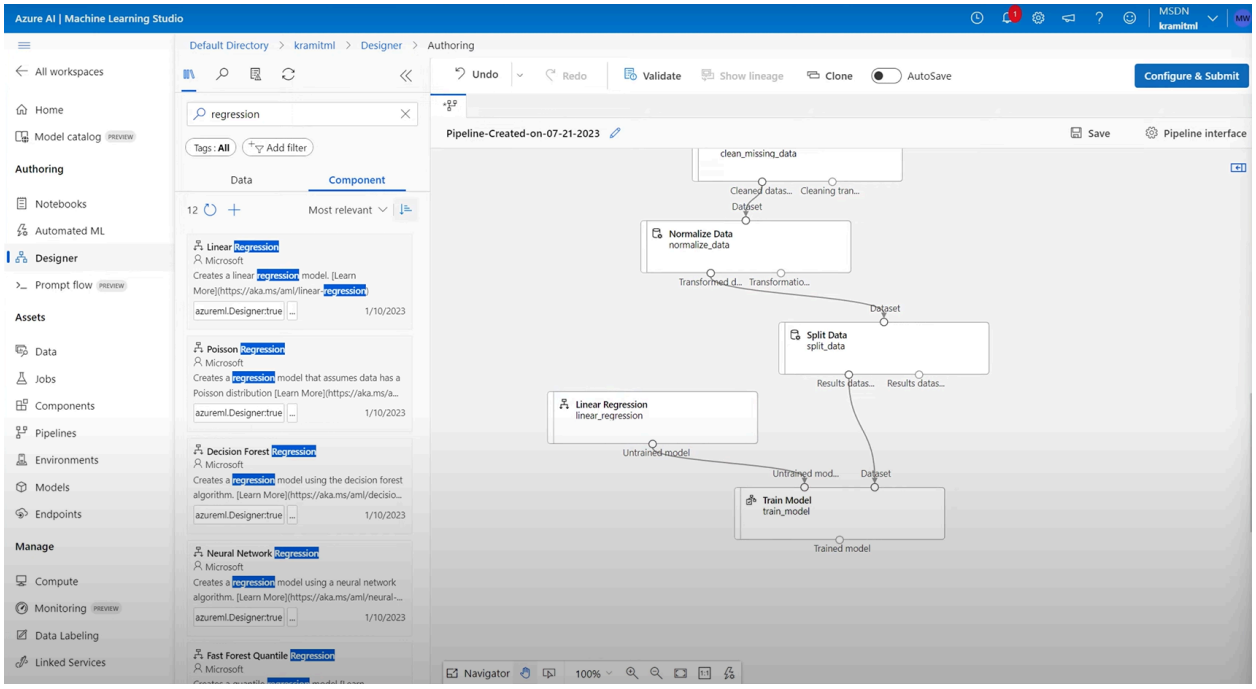

Azure AI | Machine Learning Studio
Default Directory > kramitml > Designer > Authoring
Undo
Redo
Validate
Show lineage
Clone
AutoSave
Configure & Submit
All workspaces
Home
Model catalog
Authoring
Notebooks
Automated ML
Designer
Prompt flow
Assets
Data
Jobs
Components
Pipelines
Environments
Models
Endpoints
Manage
Compute
Monitoring
Data Labeling
Linked Services
normali
Tags : All
Add filter
Data
Component
Most relev...
Normalize Data
Microsoft
Pipeline-Created-on-07-21-2023
Save
Pipeline interface
Automobile price data (Raw)
Data output
Dataset
Select Columns in Dataset
select_columns_in_dataset
Results dataset
Clean Missing Data
clean_missing_data
Cleaned datas...
Cleaning tran...
Normalize Data
normalize_data
Parameters
Transformed d...
Transformatio...
Navigator
100%

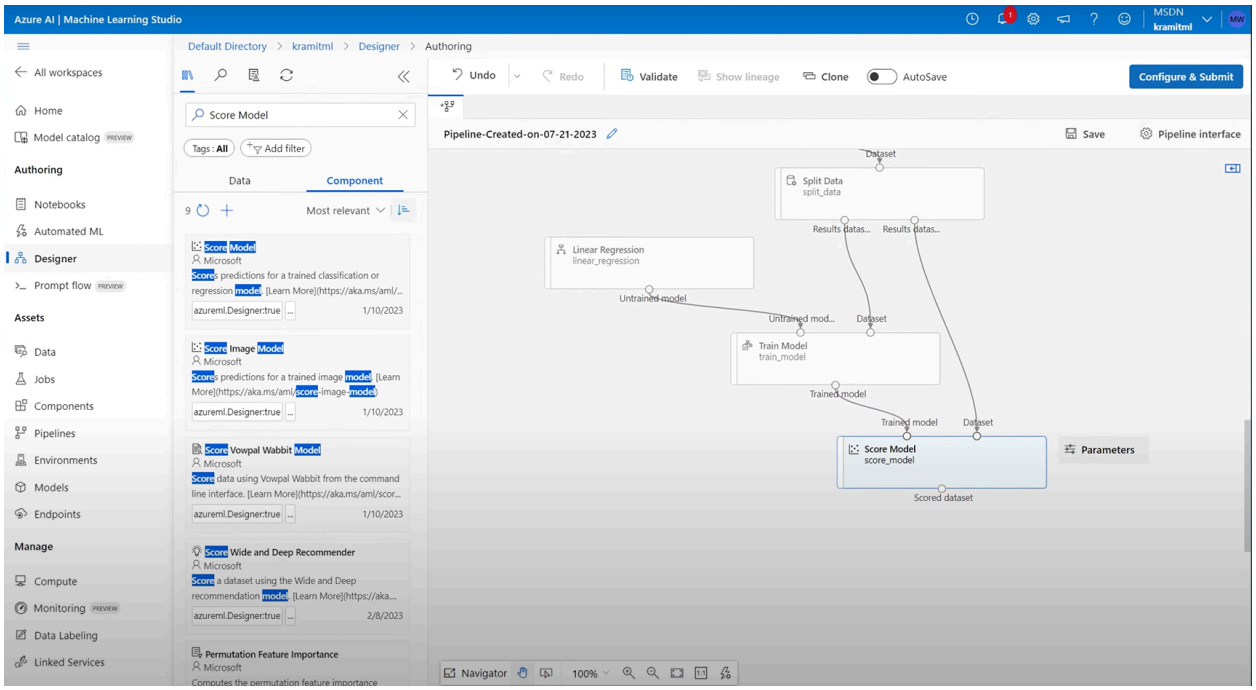

Azure AI | Machine Learning Studio
Default Directory > kramitml > Jobs > autoprice > Pipeline-Created-on-07-21-2023
All workspaces
Home
Model catalog
Authoring
Notebooks
Automated ML
Designer
Prompt flow
Assets
Data
Jobs
Components
Pipelines
Environments
Models
Endpoints
Manage
Compute
Monitoring
Data Labeling
Linked Services
Outline
Type node name, comment or comp...
Add filter
Pipeline-Created-on-07-21-2023
Automobile price data (Raw)
select_columns_in_dataset
clean_missing_data
normalize_data
Refresh
Clone
Resubmit
View profiling
Publish
Schedule
Show lineage
Show compare list
Cancel
Pipeline-Created-on-07-21-2023
Completed
Share
Add to compare
Job overview
Automobile price data (Raw)
Data output
Dataset
Select Columns in Dataset
select_columns_in_dataset
Results dataset
Dataset
Clean Missing Data
clean_missing_data
Cleaned datas...
Cleaning tran...
Dataset
Normalize Data
normalize_data
Transformed d...
Transformatio...
Navigator
100%

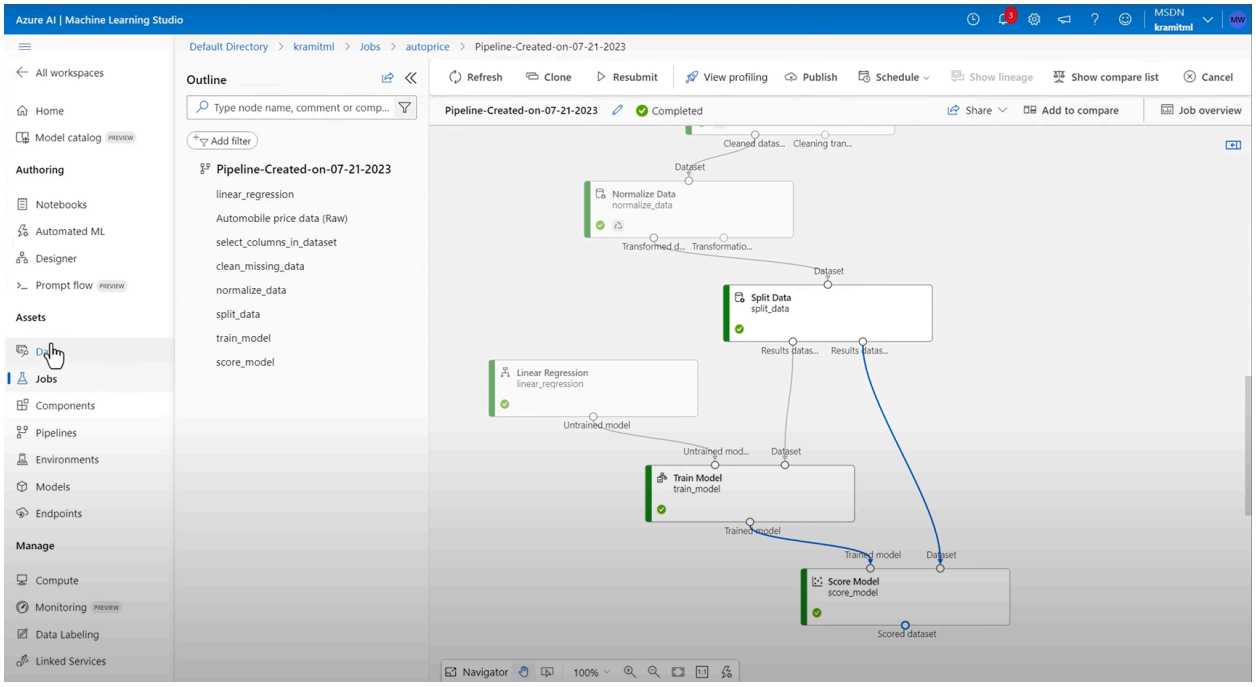

Azure AI | Machine Learning Studio
Default Directory > kramitml > Jobs > autoprice > Pipeline-Created-on-07-21-2023
All workspaces
Home
Model catalog
Authoring
Notebooks
Automated ML
Designer
Prompt flow
Assets
Data
Jobs
Components
Pipelines
Environments
Models
Endpoints
Manage
Compute
Monitoring
Data Labeling
Linked Services
Outline
Type node name, comment or comp...
Add filter
Pipeline-Created-on-07-21-2023
Automobile price data (Raw)
select_columns_in_dataset
clean_missing_data
normalize_data
Refresh
Clone
Resubmit
View profiling
Publish
Schedule
Show lineage
Show compare list
Cancel
Pipeline-Created-on-07-21-2023
Completed
Share
Add to compare
Job overview
Transformed_dataset
Rows
199
Columns
25
symboling make fuel-type aspiration num-of-doors body-style drive-wheels engine-location wheel-base
1 alfa-romero gas std two convertible rwd front 0.05830
1 alfa-romero gas std two convertible rwd front 0.05830
0.6 alfa-romero gas std two hatchback rwd front 0.23032
0.8 audi gas std four sedan fwd front 0.38484
0.8 audi gas std four sedan 4wd front 0.37317
0.8 audi gas std two sedan fwd front 0.38484
0.6 audi gas std four sedan fwd front 0.55976
0.6 audi gas std four wagon fwd front 0.55976
0.6 audi gas turbo four sedan fwd front 0.55976
0.4 audi gas turbo two hatchback 4wd front 0.37609
0.8 bmw gas std two sedan rwd front 0.42565
0.4 bmw gas std four sedan rwd front 0.42565
0.4 bmw gas std two sedan rwd front 0.42565
0.4 bmw gas std four sedan rwd front 0.42565
0.6 bmw gas std four sedan rwd front 0.49271
0.4 bmw gas std four sedan rwd front 0.49271
0.4 bmw gas std two sedan rwd front 0.49271
0.4 bmw gas std four sedan rwd front 0.68221
0.8 chevrolet gas std two hatchback fwd front 0.05247
To view, select a column in the table

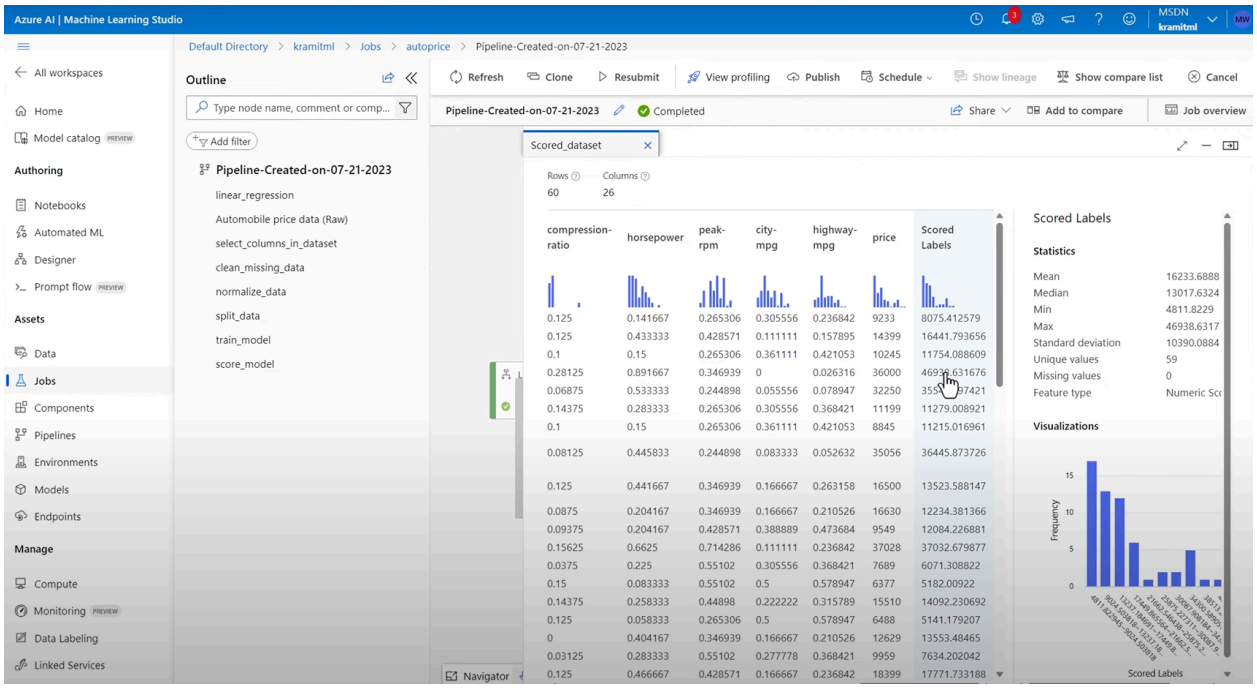

Azure AI | Machine Learning Studio
Default Directory > kramitml > Designer > Authoring
All workspaces
Home
Model catalog
Authoring
Notebooks
Automated ML
Designer
Prompt flow
Assets
Data
Jobs
Components
Pipelines
Environments
Models
Endpoints
Manage
Compute
Monitoring
Data Labeling
Linked Services
split data
Tags : All
Add filter
Data
Component
Most relevant
Split Data
Microsoft
Partitions the rows of a dataset into two distinct sets.
[Learn More](https://aka.ms/aml/split-data
azureml.Designer:true
1/10/2023
Apply SQL Transformation
Microsoft
Runs a SQLite query on input datasets to transform the data. [Learn More](https://aka.ms/aml/apply-sql...
azureml.Designer:true
1/10/2023
Undo
Redo
Validate
Show lineage
Clone
AutoSave
Configure & Submit
Pipeline-Created-on-07-21-2023
Save
Pipeline interface
Automobile price data (Raw)
Dataset
Data output
Select Columns in Dataset
select_columns_in_dataset
Results dataset
Clean Missing Data
clean_missing_data
Cleaned datas...
Cleaning tran...
Normalize Data
normalize_data
Transformed d...
Transformatio...
Split Data
split_data
Parameters
Results datas...
Navigator
100%

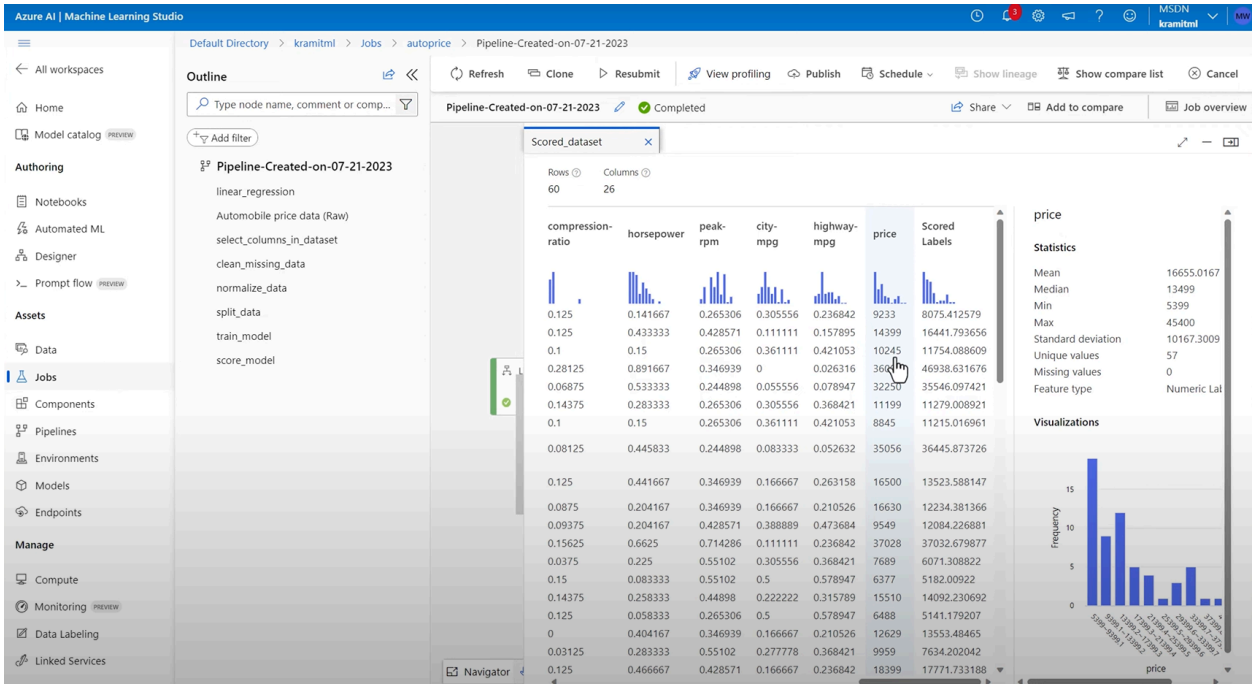

Azure AI | Machine Learning Studio
Default Directory > kramitml > Designer > Authoring
All workspaces
Home
Model catalog
Authoring
Notebooks
Automated ML
Designer
Prompt flow
Assets
Data
Jobs
Components
Pipelines
Environments
Models
Endpoints
Manage
Compute
Monitoring
Data Labeling
Linked Services
train model
Tags : All
Add filter
Data
Component
Most relevant
Train Model
Microsoft
Trains a classification or regression model in a supervised manner. [Learn More](https://aka.ms/am...
azureml.Designer:true
1/10/2023
Train Clustering Model
Microsoft
Train clustering models and assign data to clusters. [Learn More](https://aka.ms/aml/train-clustering-m...
azureml.Designer:true
1/10/2023
Train PyTorch Model
Microsoft
Train pytorch model from scratch or finetune it. [Learn More](https://aka.ms/aml/train-pytorch-mod...
azureml.Designer:true
1/10/2023
Train Vowpal Wabbit Model
Microsoft
Train a Vowpal Wabbit model using the command line interface. [Learn More](https://aka.ms/aml/train...
azureml.Designer:true
1/10/2023
Undo
Redo
Validate
Show lineage
Clone
AutoSave
Configure & Submit
Pipeline-Created-on-07-21-2023
Save
Pipeline interface
clean_missing_data
Cleaned datas...
Cleaning tran...
Dataset
Normalize Data
normalize_data
Transformed d...
Transformatio...
Split Data
split_data
Results datas...
Untrained mod...
Train Model
train_model
Parameters
Trained model
Navigator
100%

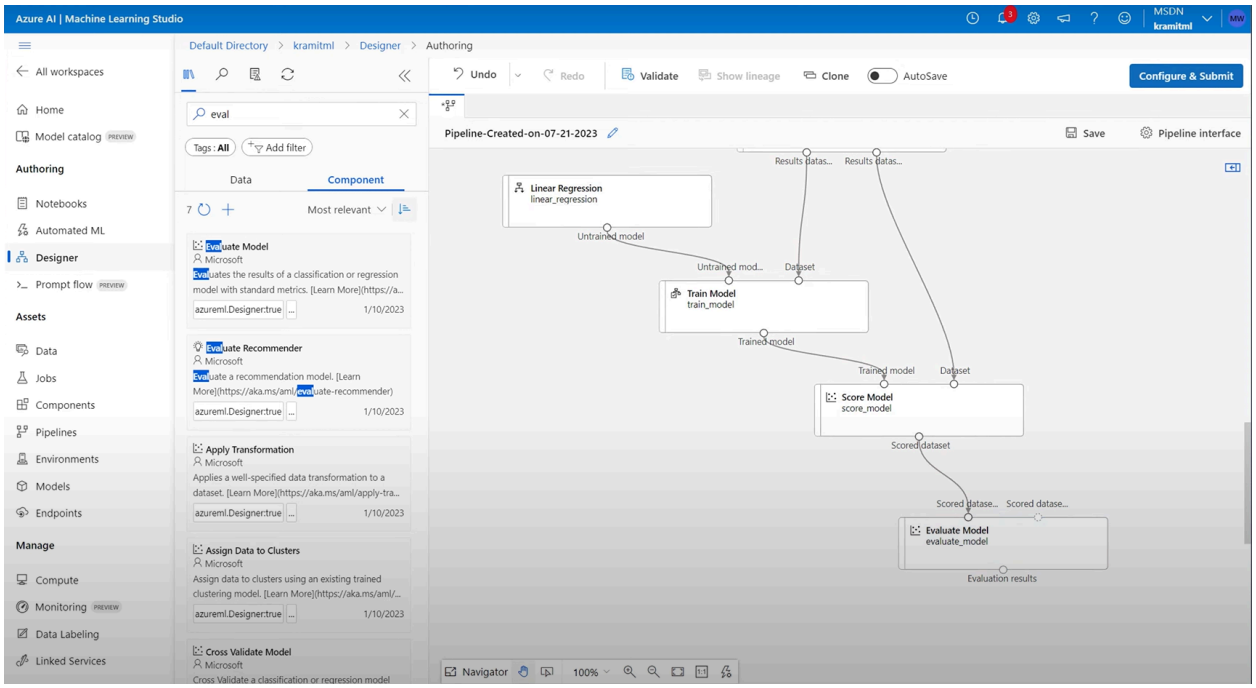

Azure AI | Machine Learning Studio
Default Directory > kramitml > Designer > Authoring
Undo
Redo
Validate
Show lineage
Clone
AutoSave
Configure & Submit
All workspaces
Home
Model catalog
Authoring
Notebooks
Automated ML
Designer
Prompt flow
Assets
Data
Jobs
Components
Pipelines
Environments
Models
Endpoints
Manage
Compute
Monitoring
Data Labeling
Linked Services
regression
Tags : All
Add filter
Data
Component
Most relevant
Linear Regression
Microsoft
Creates a linear regression model. [Learn More](https://aka.ms/aml/linear-regression)
azureml.Designer:true
1/10/2023
Poisson Regression
Creates a regression model that assumes data has a Poisson distribution [Learn More](https://aka.ms/a...
Decision Forest Regression
Creates a regression model using the decision forest algorithm. [Learn More](https://aka.ms/aml/decisio...
Neural Network Regression
Creates a regression model using a neural network algorithm. [Learn More](https://aka.ms/aml/neural-...
Fast Forest Quantile Regression
Pipeline-Created-on-07-21-2023
Save
Pipeline interface
clean_missing_data
Cleaned datas...
Cleaning tran...
Dataset
Normalize Data
normalize_data
Transformed d...
Transformatio...
Split Data
split_data
Results datas...
Linear Regression
linear_regression
Untrained model
Train Model
train_model
Trained model
Navigator
100%

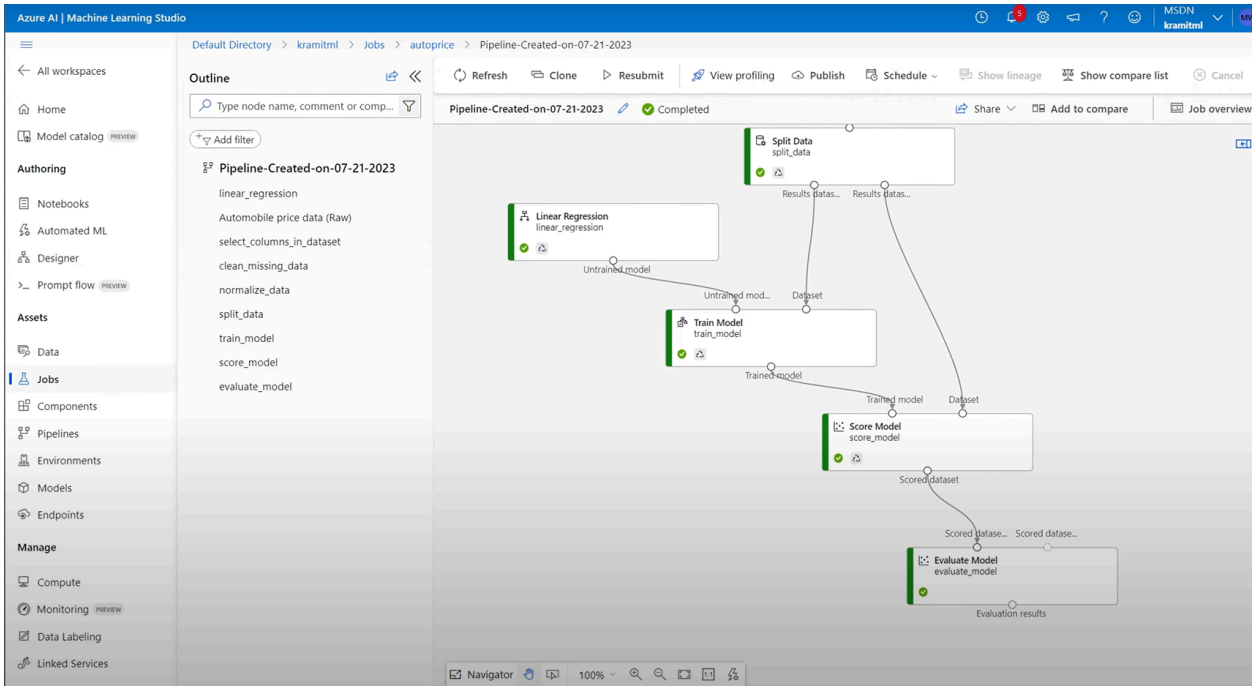

Azure AI | Machine Learning Studio
Default Directory > kramitml > Designer > Authoring
Undo
Redo
Validate
Show lineage
Clone
AutoSave
Configure & Submit
All workspaces
Home
Model catalog
Authoring
Notebooks
Automated ML
Designer
Prompt flow
Assets
Data
Jobs
Components
Pipelines
Environments
Models
Endpoints
Manage
Compute
Monitoring
Data Labeling
Linked Services
Score Model
Tags : All
Add filter
Data
Component
Most relevant
Score Model
Microsoft
Scores predictions for a trained classification or regression model. [Learn More](https://aka.ms/aml/...
azureml.Designer:true
1/10/2023
Score Image Model
Microsoft
Scores predictions for a trained image model. [Learn More](https://aka.ms/aml/score-image-model)
Score Vowpal Wabbit Model
Microsoft
Score data using Vowpal Wabbit from the command line interface. [Learn More](https://aka.ms/aml/scor...
2/8/2023
Score Wide and Deep Recommender
Microsoft
Score a dataset using the Wide and Deep recommendation model. [Learn More](https://aka....
Permutation Feature Importance
Microsoft
Computes the permutation feature importance
Pipeline-Created-on-07-21-2023
Save
Pipeline interface
Dataset
Split Data
split_data
Results datas...
Linear Regression
linear_regression
Untrained model
Train Model
train_model
Trained model
Score Model
score_model
Scored dataset
Parameters
Navigator
100%

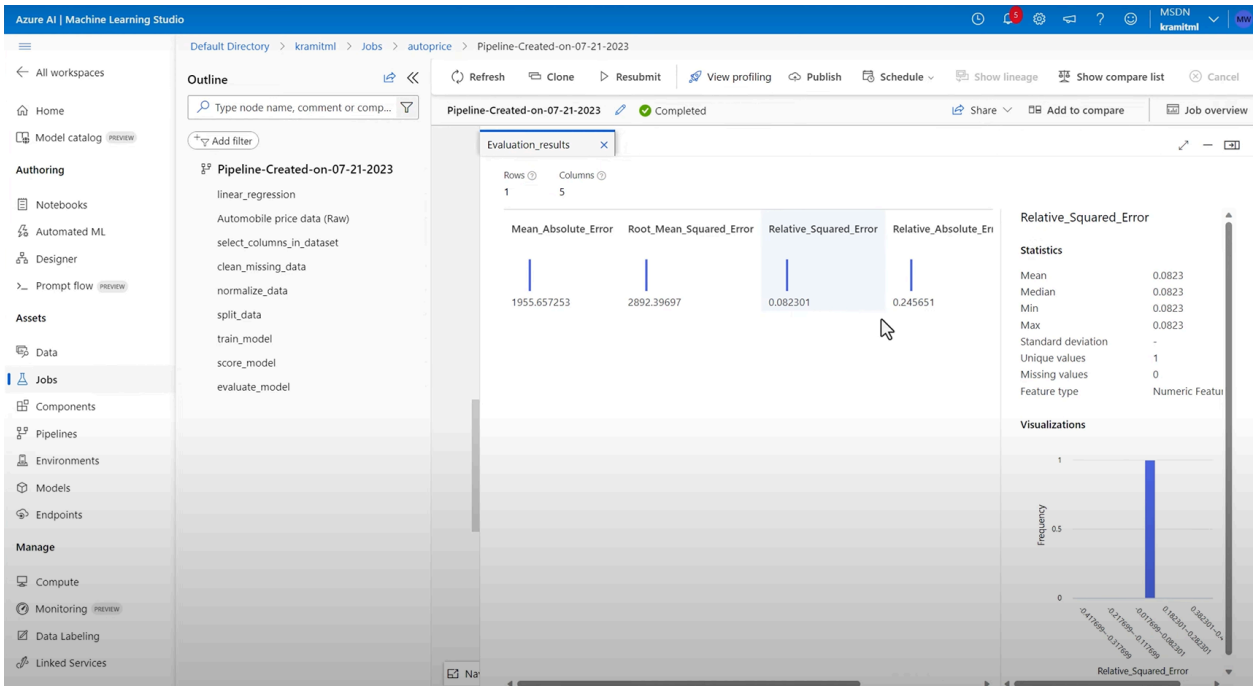

Azure AI | Machine Learning Studio
Default Directory > kramitml > Jobs > autoprice > Pipeline-Created-on-07-21-2023
Refresh
Clone
Resubmit
View profiling
Publish
Schedule
Show lineage
Show compare list
Cancel
Completed
Share
Add to compare
Job overview
Outline
Pipeline-Created-on-07-21-2023
linear_regression
Automobile price data (Raw)
select_columns_in_dataset
clean_missing_data
normalize_data
split_data
train_model
score_model
Normalize Data
Split Data
Linear Regression
Train Model
Score Model
Scored dataset
Navigator

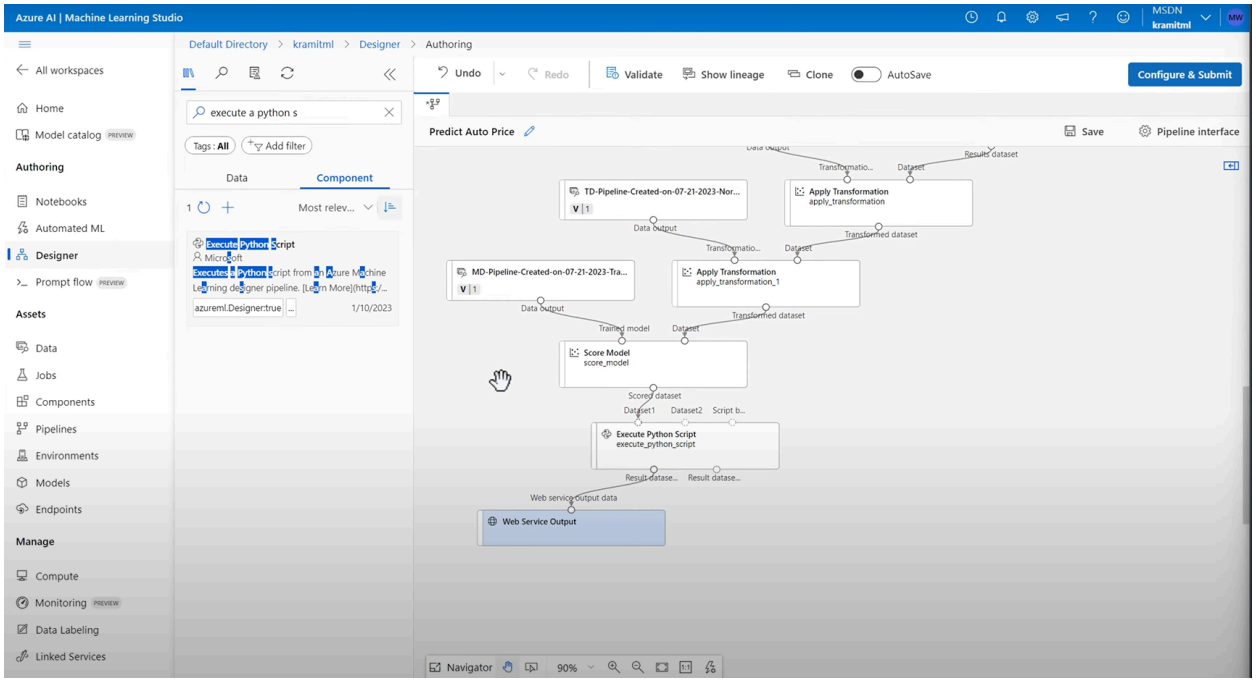

Azure AI | Machine Learning Studio
Scored_dataset
Rows 60
Columns 26
compression-ratio
horsepower
peak-rpm
city-mpg
highway-mpg
price
Scored Labels
Statistics
Mean 16233.6888
Median 13017.6324
Min 4811.8229
Max 46938.6317
Standard deviation 10390.0884
Unique values 59
Missing values 0
Feature type Numeric Sc
Visualizations
Frequency
Scored Labels

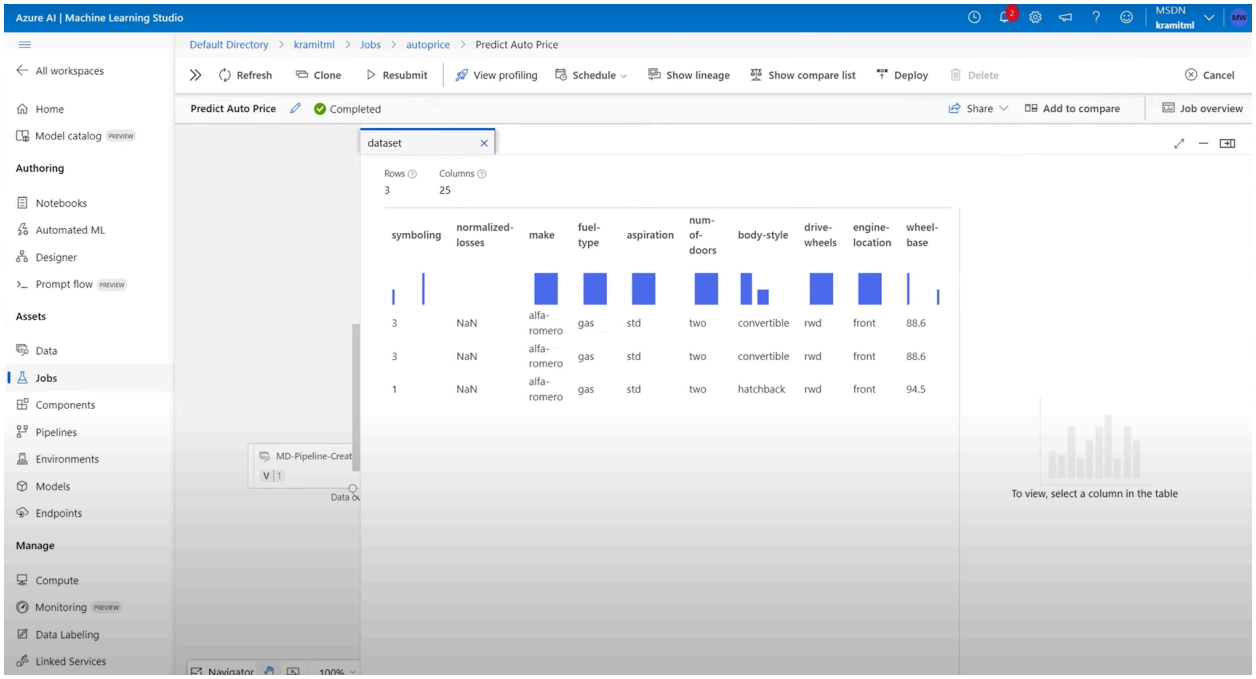

Azure AI | Machine Learning Studio
Default Directory > kramitml > Jobs > autoprice > Pipeline-Created-on-07-21-2023
Refresh
Clone
Resubmit
View profiling
Publish
Schedule
Show lineage
Show compare list
Cancel
Pipeline-Created-on-07-21-2023
Completed
Share
Add to compare
Job overview
Outline
Type node name, comment or comp...
Add filter
Pipeline-Created-on-07-21-2023
linear_regression
Automobile price data (Raw)
select_columns_in_dataset
clean_missing_data
normalize_data
split_data
train_model
score_model
Scored_dataset
Rows
60
Columns
26
compression-ratio
horsepower
peak-rpm
city-mpg
highway-mpg
price
Scored Labels
price
Statistics
Mean 16655.0167
Median 13499
Min 5399
Max 45400
Standard deviation 10167.3009
Unique values 57
Missing values 0
Feature type Numeric Lal
Visualizations
Frequency
Navigator

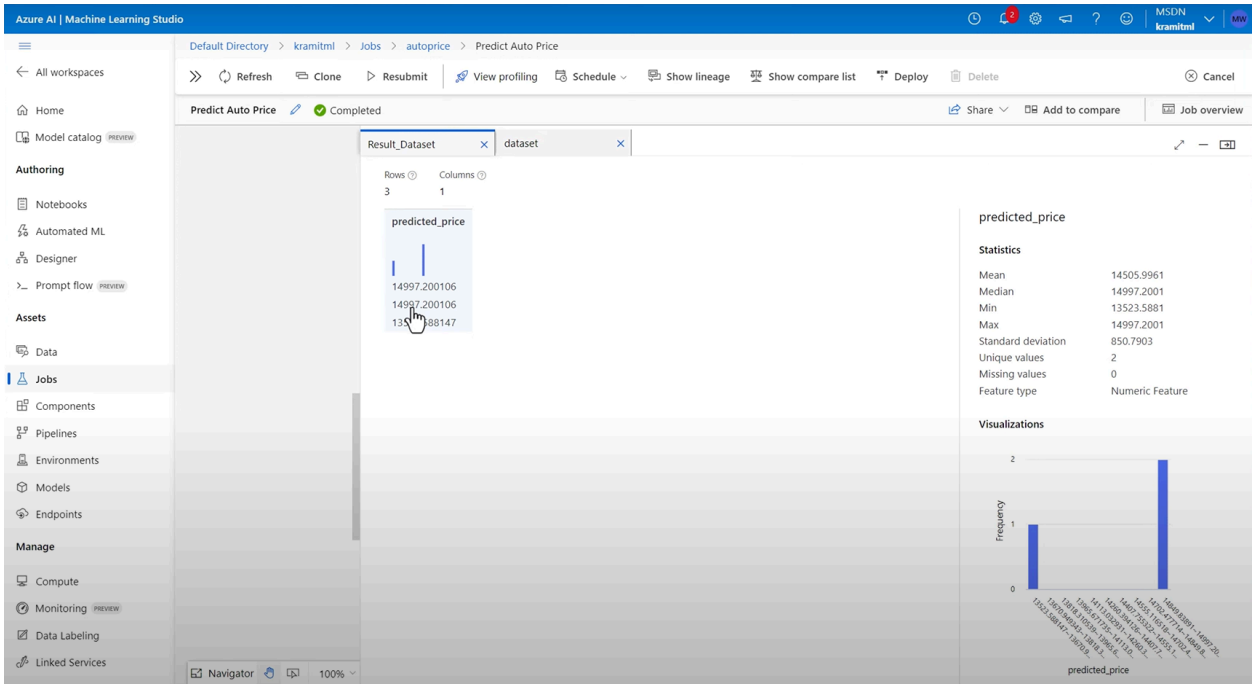

Azure AI | Machine Learning Studio
Default Directory > kramitml > Designer > Authoring
Undo
Redo
Validate
Show lineage
Clone
AutoSave
Configure & Submit
Pipeline-Created-on-07-21-2023
Save
Pipeline interface
eval
Tags : All
Add filter
Data
Component
Most relevant
Evaluate Model
Microsoft
Evaluates the results of a classification or regression model with standard metrics. [Learn More](https://a...
azureml.Designer:true
1/10/2023
Evaluate Recommender
Microsoft
Evaluate a recommendation model. [Learn More](https://aka.ms/aml/evaluate-recommender)
azureml.Designer:true
1/10/2023
Apply Transformation
Microsoft
Applies a well-specified data transformation to a dataset. [Learn More](https://aka.ms/aml/apply-tra...
azureml.Designer:true
1/10/2023
Assign Data to Clusters
Microsoft
Assign data to clusters using an existing trained clustering model. [Learn More](https://aka.ms/aml/...
azureml.Designer:true
1/10/2023
Cross Validate Model
Microsoft
Linear Regression
linear_regression
Untrained model
Train Model
train_model
Trained model
Dataset
Score Model
score_model
Scored dataset
Evaluate Model
evaluate_model
Evaluation results
Navigator
100%

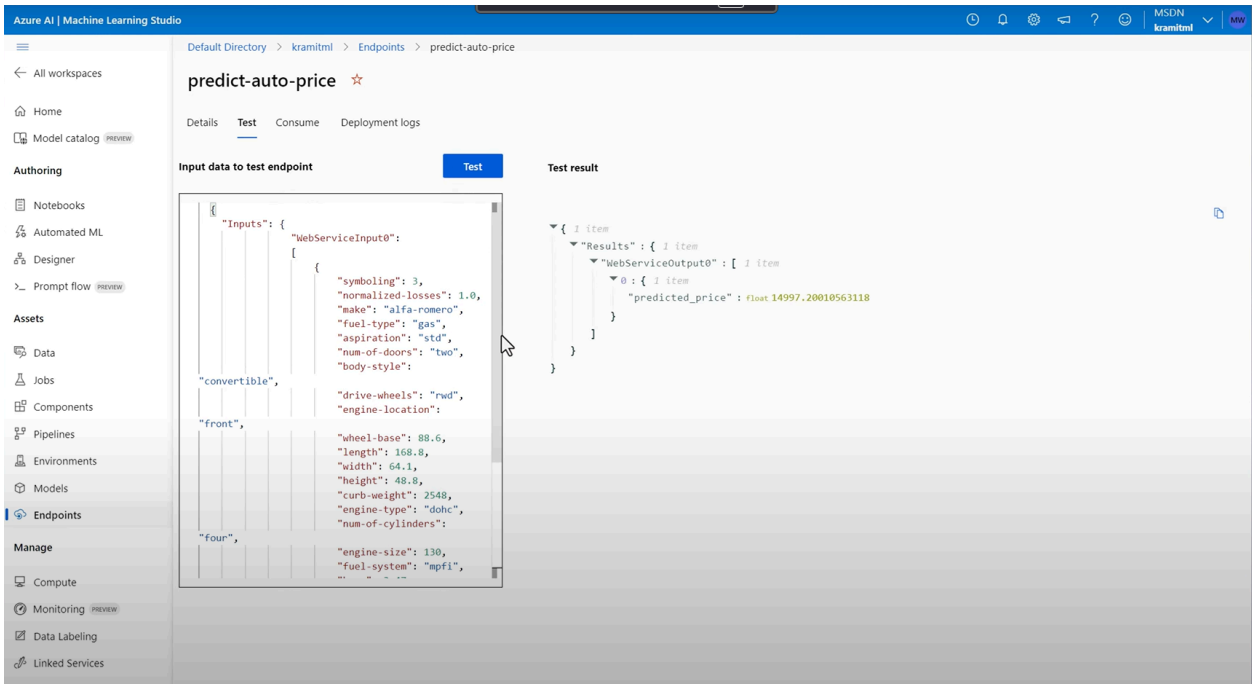

Azure AI | Machine Learning Studio
Default Directory > kramitml > Jobs > autoprice > Pipeline-Created-on-07-21-2023
Refresh
Clone
Resubmit
View profiling
Publish
Schedule
Show lineage
Show compare list
Cancel
Pipeline-Created-on-07-21-2023
Completed
Share
Add to compare
Job overview
Outline
Pipeline-Created-on-07-21-2023
linear_regression
Automobile price data (Raw)
select_columns_in_dataset
clean_missing_data
normalize_data
split_data
train_model
score_model
evaluate_model
Split Data
split_data
Linear Regression
linear_regression
Train Model
train_model
Score Model
score_model
Evaluate Model
evaluate_model
Evaluation results
Navigator
100%


Azure AI | Machine Learning Studio
Default Directory > kramitml > Jobs > autoprice > Pipeline-Created-on-07-21-2023
Refresh
Clone
Resubmit
View profiling
Publish
Schedule
Show lineage
Show compare list
Cancel
Pipeline-Created-on-07-21-2023
Completed
Share
Add to compare
Job overview
Evaluation_results
Rows
Columns
1
5
Mean_Absolute_Error
Root_Mean_Squared_Error
Relative_Squared_Error
Relative_Absolute_Err
1955.657253
2892.39697
0.082301
0.245651
Relative_Squared_Error
Statistics
Mean 0.0823
Median 0.0823
Min 0.0823
Max 0.0823
Standard deviation -
Unique values 1
Missing values 0
Feature type Numeric Featu
Visualizations
Frequency
Relative_Squared_Error


Azure AI | Machine Learning Studio
Default Directory > kramitml > Designer > Authoring
Undo
Redo
Validate
Show lineage
Clone
AutoSave
Configure & Submit
All workspaces
Home
Model catalog
Authoring
Notebooks
Automated ML
Designer
Prompt flow
Assets
Data
Jobs
Components
Pipelines
Environments
Models
Endpoints
Manage
Compute
Monitoring
Data Labeling
Linked Services
execute a python s
Tags : All
Add filter
Data
Component
Execute Python Script
Microsoft
Executes a Python script from an Azure Machine Learning designer pipeline. [Learn More](https://...
azureml.Designer:true
1/10/2023
Predict Auto Price
Save
Pipeline interface
TD-Pipeline-Created-on-07-21-2023-Nor...
Apply Transformation
apply_transformation
MD-Pipeline-Created-on-07-21-2023-Tra...
apply_transformation_1
Score Model
score_model
Execute Python Script
execute_python_script
Web Service Output
Navigator
90%


Azure AI | Machine Learning Studio
Default Directory > kramitml > Jobs > autoprice > Predict Auto Price
Refresh
Clone
Resubmit
View profiling
Schedule
Show lineage
Show compare list
Deploy
Delete
Cancel
Predict Auto Price
Completed
Share
Add to compare
Job overview
dataset
Rows
3
Columns
25
symboling
normalized-losses
make
fuel-type
aspiration
num-of-doors
body-style
drive-wheels
engine-location
wheel-base
3 NaN alfa-romero gas std two convertible rwd front 88.6
3 NaN alfa-romero gas std two convertible rwd front 88.6
1 NaN alfa-romero gas std two hatchback rwd front 94.5
To view, select a column in the table
Navigator





Post Lab Question-Answers:

1. Differentiate between linear and nonlinear regression.

Linear Regression: Assumes a linear relationship between the independent variables and the dependent variable. The model can be expressed in a straight-line formula (y = mx + b).

Nonlinear Regression: Involves modeling relationships that do not follow a straight line. It uses polynomial, exponential, or logarithmic functions to fit data that exhibits a more complex relationship.

2. Write a note on converting non-linear models into linear models.

Converting Non-linear Model into Linear Model Non-linear models can often be transformed into linear models through mathematical manipulation:

1) Transformation: Applying mathematical transformations to the variables (e.g., logarithmic or polynomial transformations) to achieve linearity.

2) Linearization: Using techniques such as Taylor series expansion to approximate non-linear relationships as linear over a limited range.

Outcomes: Comprehend radial-basis-function (RBF) networks and Kernel learning method

---

Conclusion (based on the Results and outcomes achieved):
The experiment effectively illustrated the use of machine learning to address a specific business challenge. By implementing a systematic approach that included data collection, model training, and evaluation, we demonstrated how predictive analytics can lead to more informed decision-making. This process not only enhances operational efficiency but also improves customer satisfaction, highlighting the importance of data-driven strategies in today's competitive landscape.

References:

Books/ Journals/ Websites:

1. Han, Kamber, "Data Mining Concepts and Techniques", Morgan Kaufmann 3nd Edition

---

https://archive.ics.uci.edu/dataset/10/automobile