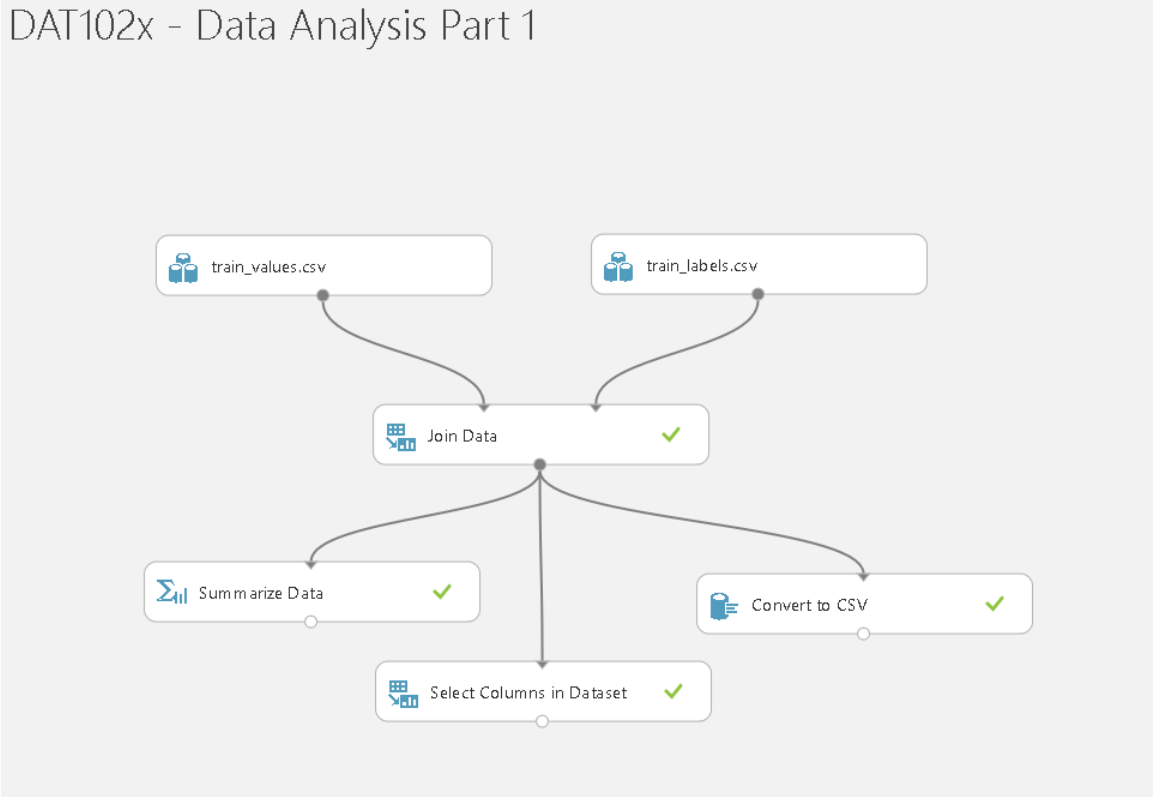
DAT102X Azure ML

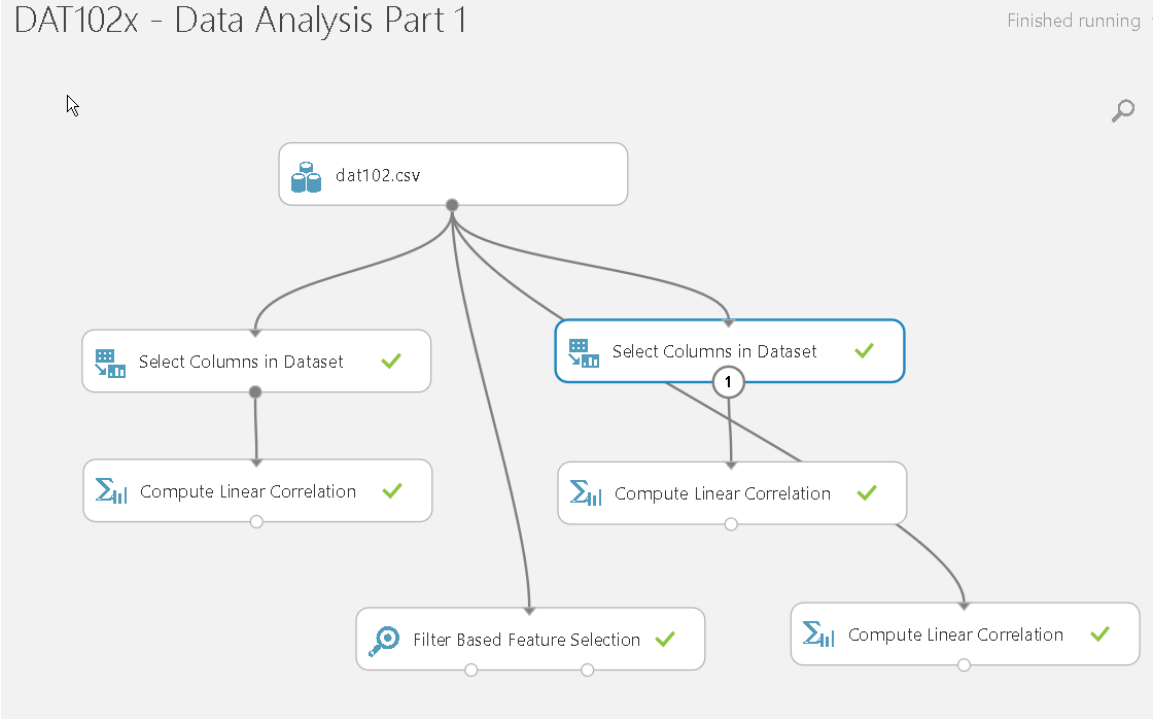
SUMMARY OF METHODS IMPLEMENTED

Dennis Lam | MPP Data Science Capstone Project | October 2018

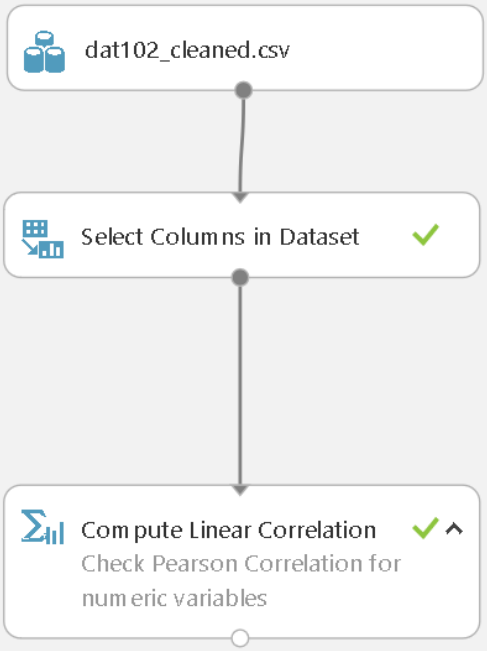
Data Analysis

Summary and Join 2 datasets for analysis.



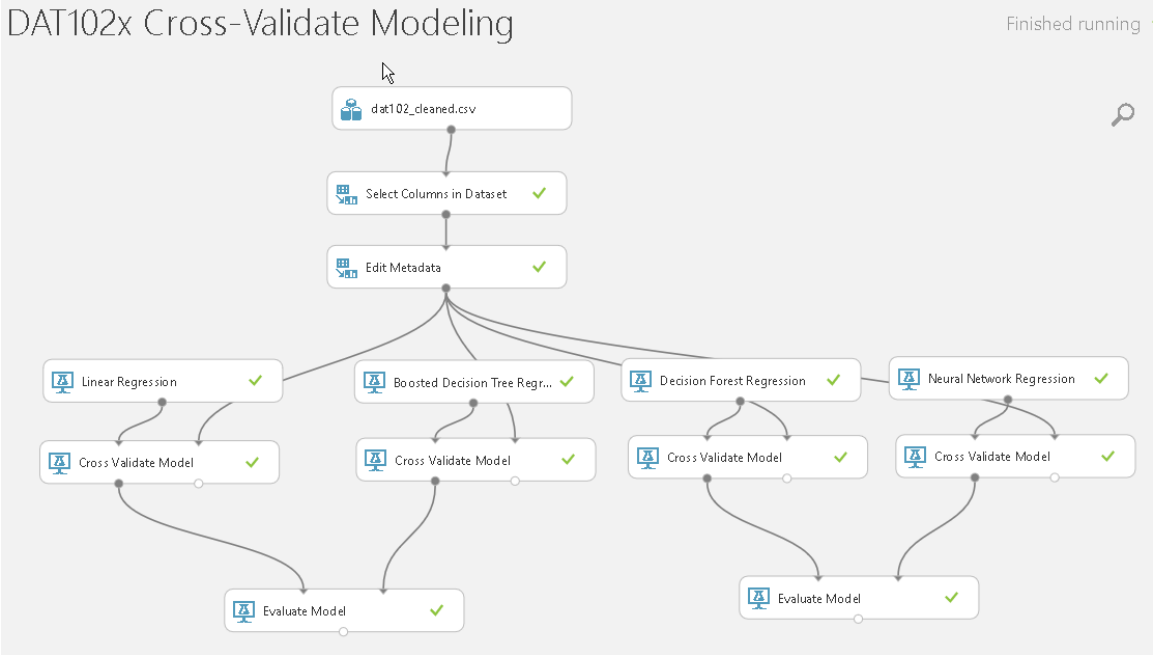


DAT102x - Data Analysis Part 2



Feature Selection

Try cross-validate and compare between Regression models



DAT102x Cross-Validate Modeling › Evaluate Model › Evaluation results

Metrics

Mean Absolute Error	5.813246
Root Mean Squared Error	7.880865
Relative Absolute Error	0.619212
Relative Squared Error	0.460893
Coefficient of Determination	0.539107

Metrics

Mean Absolute Error	1.915603
Root Mean Squared Error	2.876031
Relative Absolute Error	0.204045
Relative Squared Error	0.061382
Coefficient of Determination	0.938618

Error Histogram

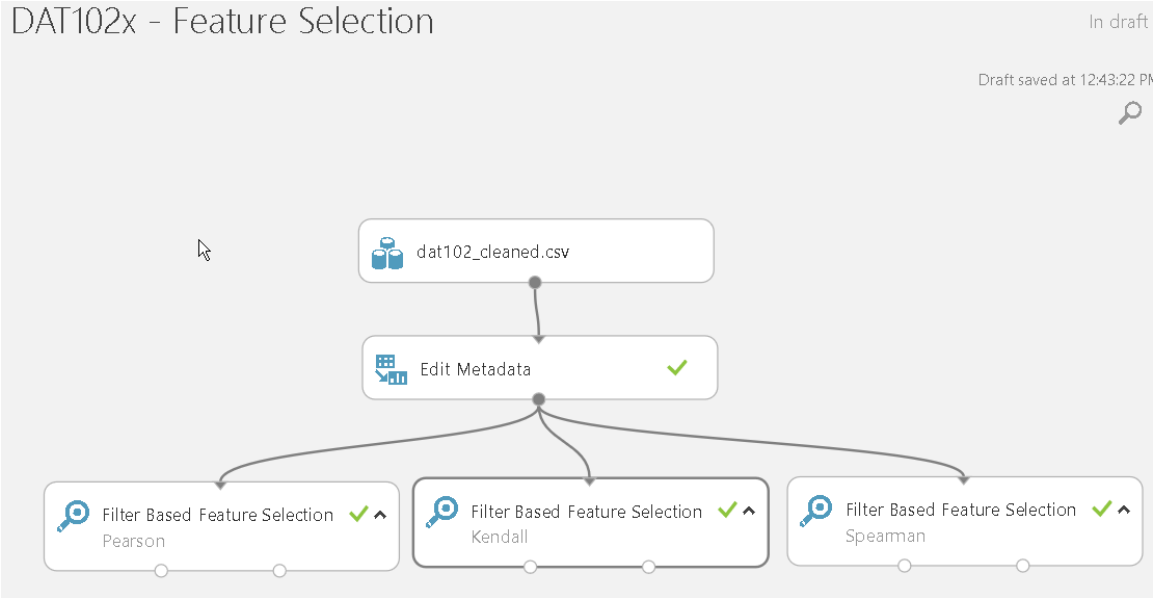
Error Histogram

DAT102x Cross-Validate Modeling › Evaluate Model › Evaluation results

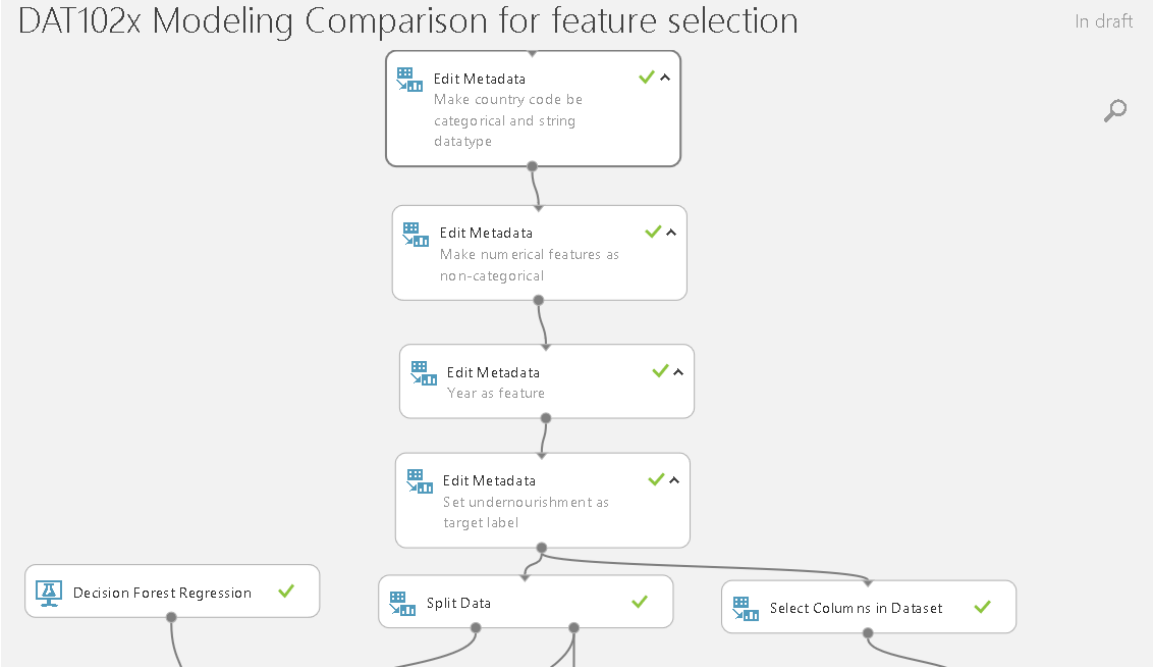
rows 2 columns 6

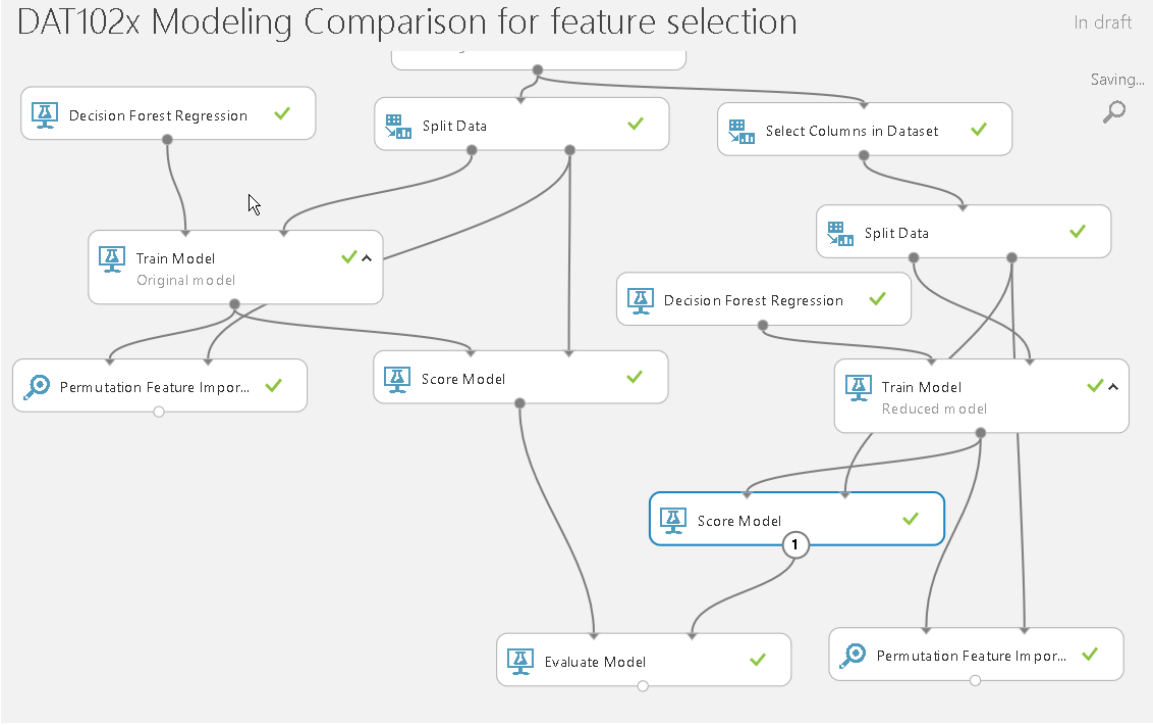
	Negative Log Likelihood	Mean Absolute Error	Root Mean Squared Error	Relative Absolute Error	Relative Squared Error	Coefficient of Determination
view as						
	2899.838857	2.212396	3.442841	0.235659	0.08796	0.91204
	Infinity	4.080628	5.67607	0.434658	0.239082	0.760918

Trying out and measure each correlation for all features using each method:

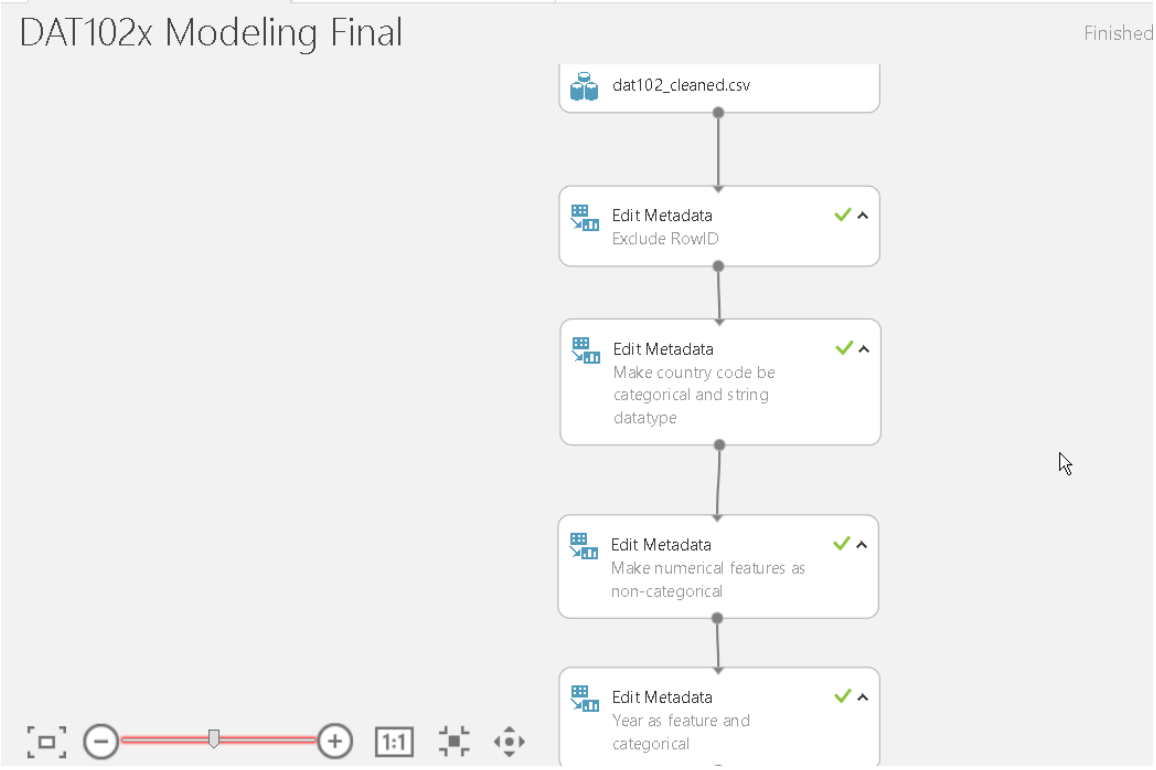


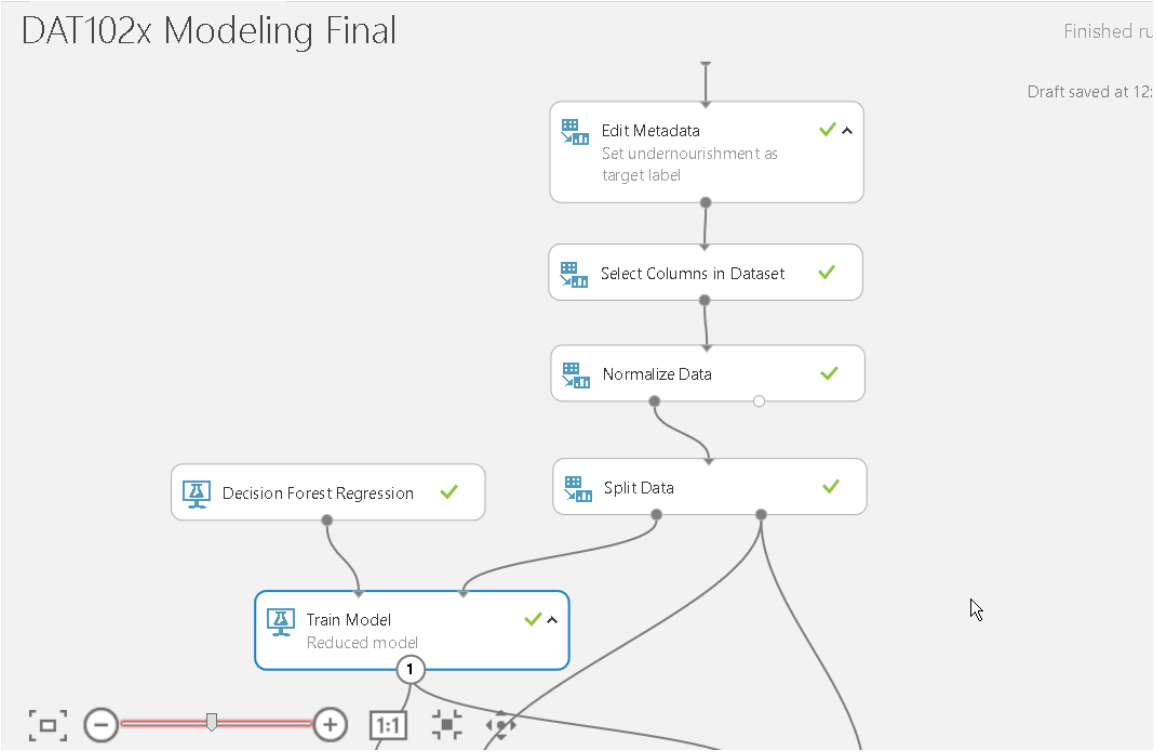
Compare models with or without feature selection

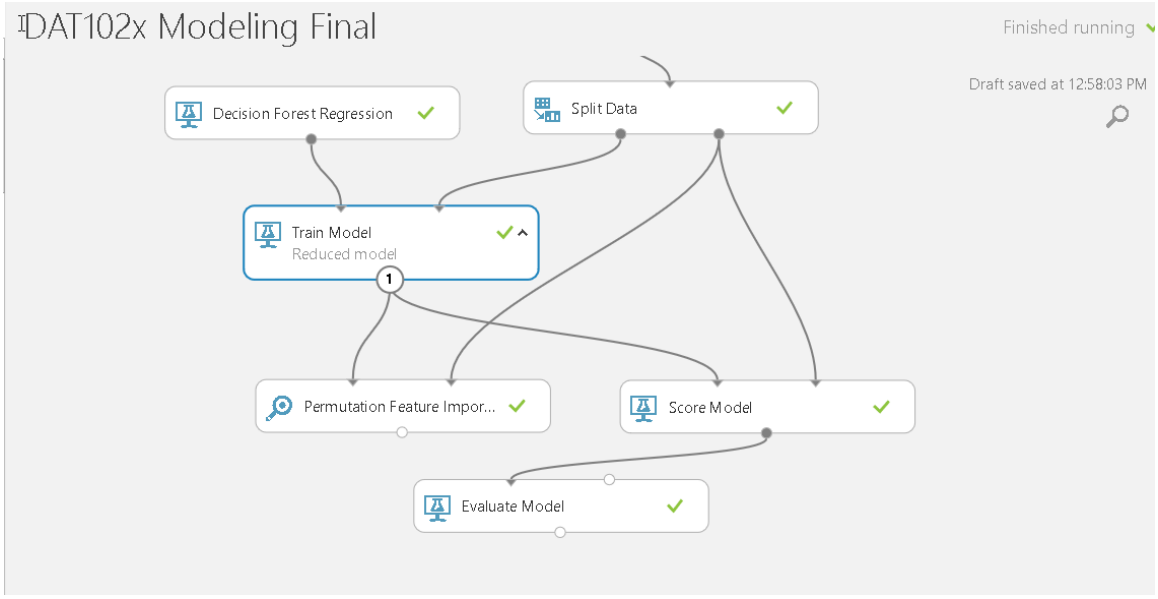




Model Training and Web Publishing



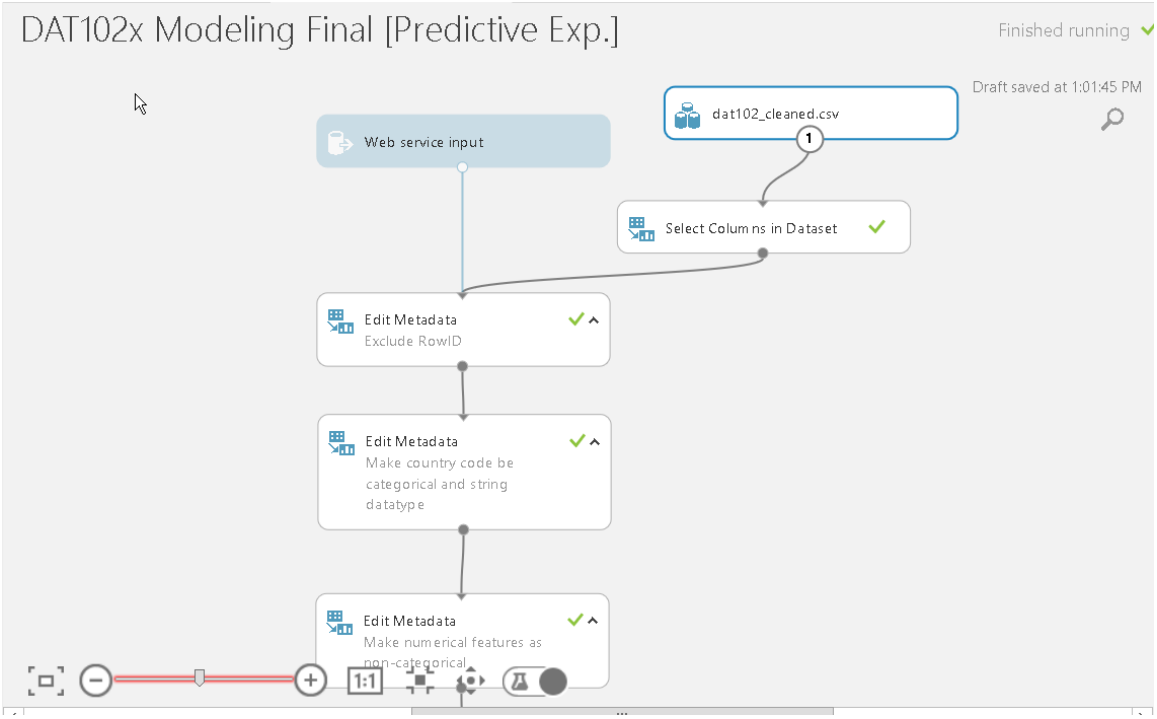


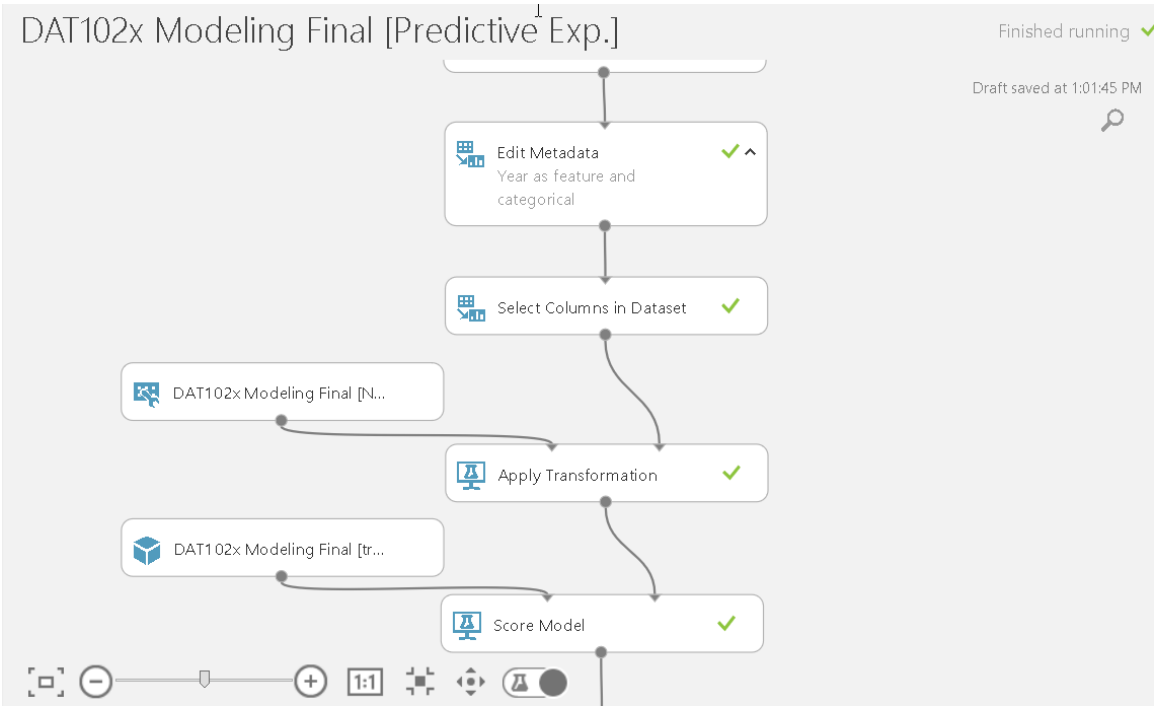


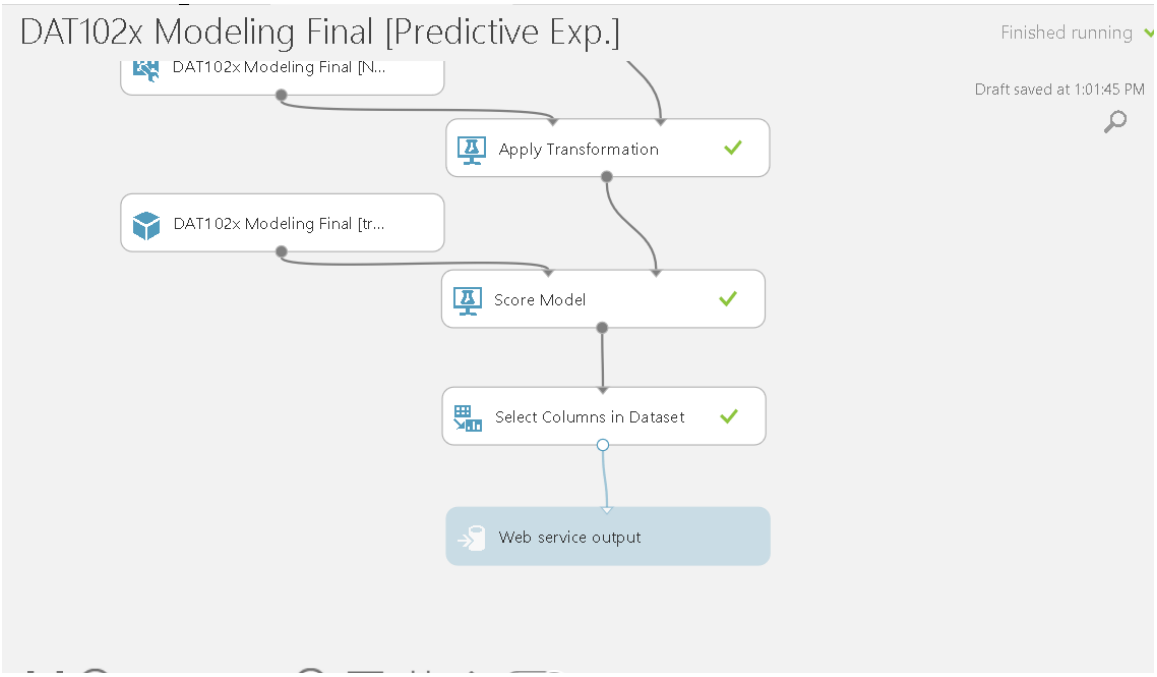
DAT102x Modeling Final › Evaluate Model › Evaluation results

rows 1 columns 6

	Negative Log Likelihood	Mean Absolute Error	Root Mean Squared Error	Relative Absolute Error	Relative Squared Error	Coefficient of Determination
view as						
	571.90844	1.7039	2.616738	0.18442	0.052171	0.947829


DAT102x Modeling Final [Predictive Exp.]
Finished running
Draft saved at 1:01:45 PM
dat102_cleaned.csv
Web service input
Select Columns in Dataset
Edit Metadata
Exclude RowID
Edit Metadata
Make country code be categorical and string datatype
Edit Metadata
Make numerical features as non-categorical


DAT102x Modeling Final [Predictive Exp.]
Finished running
Draft saved at 1:01:45 PM
Edit Metadata
Year as feature and categorical
Select Columns in Dataset
DAT102x Modeling Final [N...
Apply Transformation
DAT102x Modeling Final [tr...
Score Model


DAT102x Modeling Final [Predictive Exp.]
Finished running
Draft saved at 1:01:45 PM
DAT102x Modeling Final [N...
Apply Transformation
DAT102x Modeling Final [tr...
Score Model
Select Columns in Dataset
Web service output