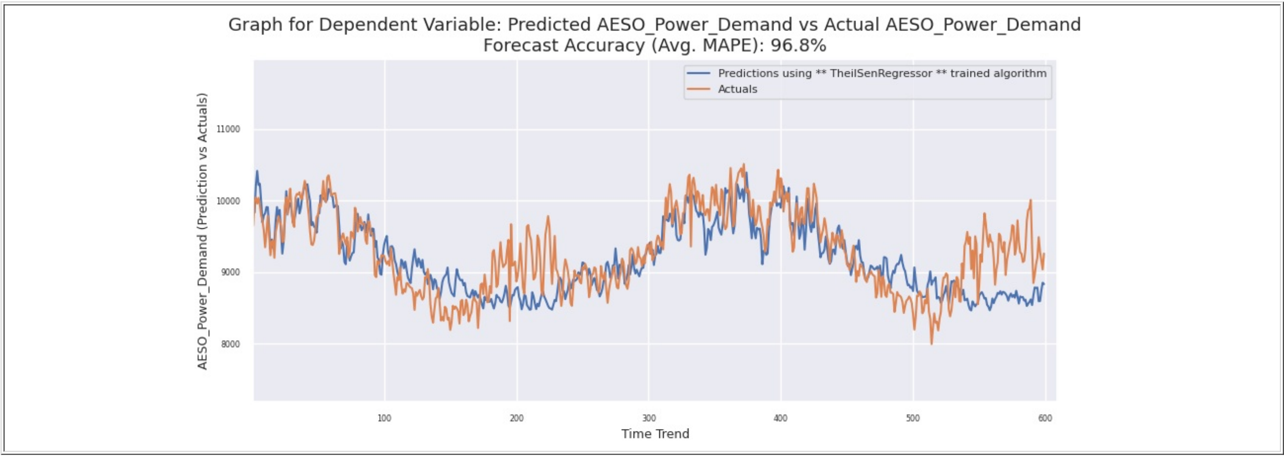
MAADSBML AutoML Report For ALBERTA-ELECTRIC-SYSTEM-OPERATOR_AESO

Generated On: 2024-04-17 02:20:35 (UTC)

Best Model(s) Report For admin_aesopowerdemand_csv



MODEL DESCRIPTION

Model Trained On: 2024/04/17
Training Start Time: 0217
Training End Time: 0220
Was Data Normalized: Yes
Was Data Shuffled: No
Deep Analysis: No
Total Training Data Set: 897
Training Data Percentage: 70%
Total Test Data Set: 383
Total # of Variables: 4
Adjusted for Seasonality: N
Total Algorithms Run: 3600
Removed Outliers: N
Best Distribution FOR ACTUAL Y: VONMISES
Dependent Variable: AESO_POWER_DEMAND
Independent Variables: ['Calgary_Weather', 'Edmonton_Weather', 'FtMac_Weather']

PREDICTION VARIABLE STATS

Mean: 9193.695
STD: 503.462
Kurtosis: -0.985
Skewness: 0.453
Coef. of Variation: 0.055
Shapiro Test for Normality: 0.936
Jarque-Bera Goodness of Fit: 44.819
Anderson: 12.205
KStat: 253896.817
KStatvar: 109187960.799
Wilcox: 0.000
Theil Slope: -1.021

ACTUAL VARIABLE STATS

Mean: 9281.250
STD: 545.222
Kurtosis: -0.891
Skewness: 0.098
Coef. of Variation: 0.059
Shapiro Test for Normality: 0.981
Jarque-Bera Goodness of Fit: 20.836
Anderson: 2.876
KStat: 297763.223
KStatvar: 164138770.163
Wilcox: 0.000
Theil Slope: -0.302

Statistics Showing Comparison Between Prediction and Actuals

Mood(actuals,predictions): 1.431
Pearson(actuals,predictions): 0.726
Kendall Tau(actuals,predictions): 0.473
Ansari(actuals,predictions): 178651.000
Jaccard_distance(actuals,predictions): 1.000
Minkowski_distance(actuals,predictions): 178919.938
Euclidean_distance(actuals,predictions): 9796.297

IMPORTANT FILE PATHS FOR RAW AND OUTPUT DATA

NOTE: These are DOCKER CONTAINER Paths. You can view these files inside the container by using the command: **docker exec -it {container id} bash If you have re-run the container, these files will be GONE but they exist on your HOST machine. The HOST MACHINE location is based on the volumes you mapped when you ran the Docker container.**
The Docker RUN Volume Mappings are :: (For example here is the docker run command (use multiple -v for multiple mappings):

DOCKER RUN COMMAND: docker run -d -p 5595:5595 -p 5495:5495 -p 10000:10000 -v {HOST MACHINE FOLDER}:{CONTAINER FOLDER}:z --env TRAININGPORT=5595 --env PREDICTIONPORT=5495 --env ABORTPORT=10000 --env COMPANYNAME=MYCOMPANY --env MAXRUNTIME=20 --env MAINHOST=127.0.0.1 maadsdocker/maads-batch-automl-otics

Docker Volume Mappings:
1. {HOST MACHINE FOLDER}/csvuploads:/maads/agentfilesdocker/dist/maadsweb/csvuploads:z
2. {HOST MACHINE FOLDER}/pdfreports:/maads/agentfilesdocker/dist/maadsweb/pdfreports:z
3. {HOST MACHINE FOLDER}/autofeatures:/maads/agentfilesdocker/dist/maadsweb/autofeatures:z
4. {HOST MACHINE FOLDER}/outliers:/maads/agentfilesdocker/dist/maadsweb/outliers:z
5. {HOST MACHINE FOLDER}/sqlloads:/maads/agentfilesdocker/dist/maadsweb/sqlloads:z
6. {HOST MACHINE FOLDER}/networktemp:/maads/agentfilesdocker/dist/maadsweb/networktemp:z
7. {HOST MACHINE FOLDER}/networks:/maads/agentfilesdocker/networks:z
8. {HOST MACHINE FOLDER}/exception:/maads/agentfilesdocker/dist/maadsweb/exception:z
9. {HOST MACHINE FOLDER}/staging:/maads/agentfilesdocker/dist/staging:z

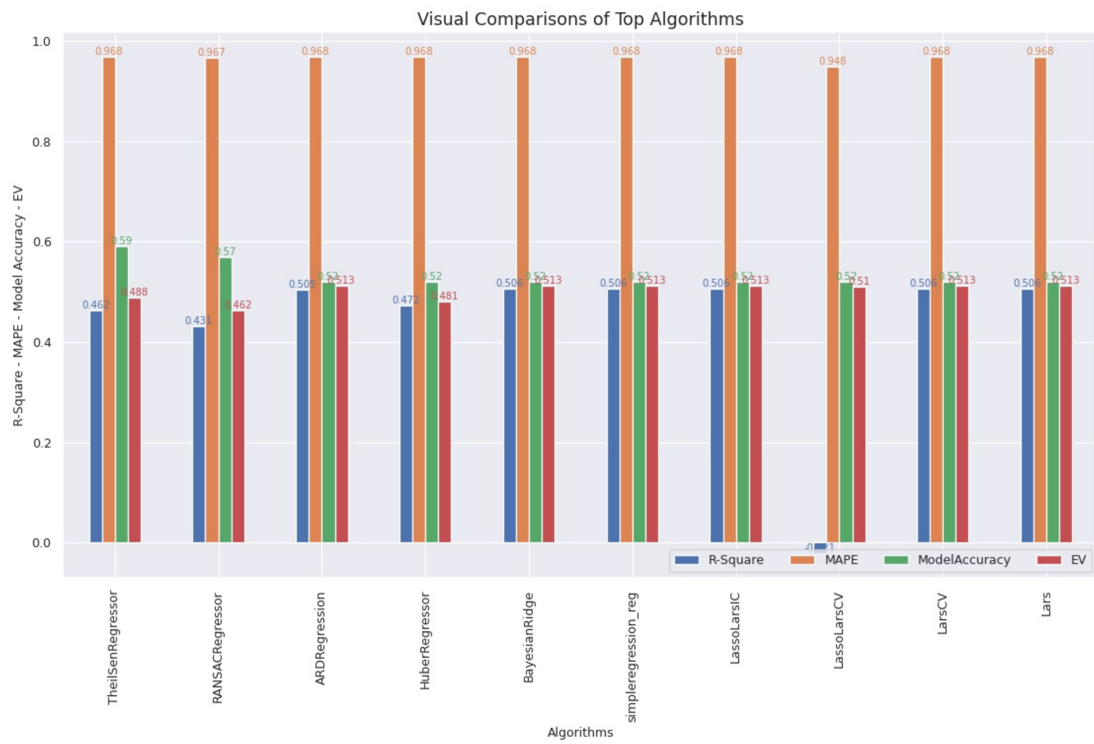
Path for Training Dataset File: /maads/agentfilesdocker/dist/maadsweb/csvuploads/aesopowerdemand.csv
Path for PDF Report (i.e. this file): /maads/agentfilesdocker/dist/maadsweb/pdfreports/admin_aesopowerdemand_csv_no_seasons.pdf
Path for AutoFeature File: /maads/agentfilesdocker/dist/maadsweb/autofeatures/admin_aesopowerdemand_csv_.csv
Path for Outliers File: /maads/agentfilesdocker/dist/maadsweb/outliers/admin_aesopowerdemand_csv.csv
Path for Algo JSON File: /maads/agentfilesdocker/dist/maadsweb/exception/admin_aesopowerdemand_csv_trained_algo_no_seasons.json
Folder Path for MySQL Scripts: /maads/agentfilesdocker/dist/maadsweb/sqlloads/
Path for Detailed Prediction File: /maads/agentfilesdocker/dist/maadsweb/csvuploads/admin_aesopowerdemand_csv_prediction_details.csv
Path for Algorithm Zip File (i.e pickle files): /maads/agentfilesdocker/dist/maadsweb/networktemp/admin_aesopowerdemand_csv.zip
Path for Algorithm Pickle Files:
1. /maads/agentfilesdocker/networks/Alberta-Electric-System-Operator_AESO)_ADMIN_AESOPOWERDEMAND_CSVALLSEASON_AG1_4_TheilSenRegressor_normal_897_ensemble_.pkl
2. /maads/agentfilesdocker/networks/Alberta-Electric-System-Operator_AESO)_ADMIN_AESOPOWERDEMAND_CSVALLSEASON_AG1_4_TheilSenRegressor_normal_897_ensemble_scalerx_.pkl
3. /maads/agentfilesdocker/networks/Alberta-Electric-System-Operator_AESO)_ADMIN_AESOPOWERDEMAND_CSVALLSEASON_AG1_4_TheilSenRegressor_normal_897_ensemble_scalery_.pkl

DESCRIPTIVE STATISTICS

Variables	T-Statistic	Count	Mean	STD	MIN	25%	50%	75%	MAX
Calgary_Weather	-35.442	897.0	5.129	-27.75	10.061	-0.9	6.05	12.8	23.85
Edmonton_Weather	-38.8	897.0	4.998	-26.64	11.629	-3.0	6.25	14.65	25.75
FtMac_Weather	-40.088	897.0	1.239	-32.4	13.472	-8.95	2.9	12.9	23.85
AESO_POWER_DEMAND	NA	897.0	9245.152	7611.0	586.467	8792.0	9256.0	9687.0	10510.0

BEST ALGORITHM FOUND FOR THIS DATASET (Note: This trained model will be used to predict AESO_POWER_DEMAND)					
Algorithm	**Description**	**Model Results**	**MAPE Accuracy**	**Forecast Months**	**Season**
TheilSenRegressor	Theil-Sen Estimator: Theil-Sen Estimator: robust multivariate regression model.	**COEFFICIENTS:** [0.38676439 -0.5151319 -0.49644178] **INTERCEPT:** 0.911761131528117 **R-square:** 0.462 **Mean Squared Error (MSE):** 159945.736 **Skewness:** 1.162 **Kurtosis:** 3.953 **Mean Square Model (MSM):** 156683729.715 **F-Statistic (F):** 979.606 **Jarque-Bera (JB):** 157.711 **Explained Variance (EV):** 0.488 **Multicolinearity Test (Avg. VIF):** 19.321 **Heteroscedasticity Test (Avg P-Value):** -1.000 **(Based on White Test, there seems to be heteroscedasticity in the model)** **Autocorrelation (Durbin-Watson) Test:** 0.370 **(Based on DW Test - there seems to be autocorrelation in your model)**	0.968	1 - 12	allseason

TOP 10 ALGORITHMS FOR ALLSEASON



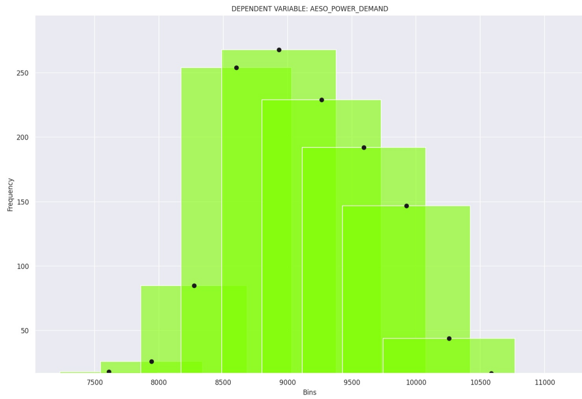
Num	Algorithm	Model Accuracy	Details	Season	Description
1	TheilSenRegressor	0.5895	**R-square:** 0.462 **MAPE:** 0.968 **Explained Variance (EV):** 0.488 **MSE:** 159945.736 **MSM:** 156683729.715 **Skewness:** 1.162 **Kurtosis:** 3.953 **F:** 979.606 **DW:** 0.37 **JB:** 157.711	allseason	THEIL-SEN ESTIMATOR: Theil-Sen Estimator: robust multivariate regression model.
2	RANSACRegressor	0.5698	**R-square:** 0.431 **MAPE:** 0.967 **Explained Variance (EV):** 0.462 **MSE:** 169115.428 **MSM:** 176298117.746 **Skewness:** 1.192 **Kurtosis:** 3.988 **F:** 1042.472 **DW:** 0.411 **JB:** 166.438	allseason	RANSAC (RANDOM SAMPLE CONSENSUS) ALGORITHM.: RANSAC is an iterative algorithm for the robust estimation of parameters from a subset of inliers from the complete data set. More information can be found in the general documentation of linear models.
3	ARDRegression	0.5244	**R-square:** 0.505 **MAPE:** 0.968 **Explained Variance (EV):** 0.513 **MSE:** 147107.923 **MSM:** 127358849.42 **Skewness:** 0.776 **Kurtosis:** 3.535 **F:** 865.751 **DW:** 0.381 **JB:** 67.32	allseason	BAYESIAN ARD: Fit the weights of a regression model, using an ARD prior. The weights of the regression model are assumed to be in Gaussian distributions. Also estimate the parameters lambda (precisions of the distributions of the weights) and alpha (precision of the distribution of the noise). The estimation is done by an iterative procedures (Evidence Maximization)
4	HuberRegressor	0.5235	**R-square:** 0.472 **MAPE:** 0.968 **Explained Variance (EV):** 0.481 **MSE:** 157089.942 **MSM:** 154347018.354 **Skewness:** 0.977 **Kurtosis:** 3.941 **F:** 982.539 **DW:** 0.379 **JB:** 117.577	allseason	HUBER REGRESSOR: Linear regression model that is robust to outliers.
5	BayesianRidge	0.5225	**R-square:** 0.506 **MAPE:** 0.968 **Explained Variance (EV):** 0.513 **MSE:** 146958.733 **MSM:** 128147002.277 **Skewness:** 0.773 **Kurtosis:** 3.531 **F:** 871.993 **DW:** 0.386 **JB:** 66.766	allseason	BAYESIAN RIDGE REGRESSION: Fit a Bayesian ridge model and optimize the regularization parameters lambda (precision of the weights) and alpha (precision of the noise).
			R-square: 0.506 **MAPE:** 0.968 **Explained Variance (EV):** 0.513 **MSE:** 146918.712		

6	simpleregression_reg	0.5210	**MSM:** 128961343.199 **Skewness:** 0.77 **Kurtosis:** 3.526 **F:** 877.773 **DW:** 0.391 **JB:** 66.216	allseason	LINEAR REGRESSION: Simple multivariate linear regression
7	LassoLarsIC	0.5207	**R-square:** 0.506 **MAPE:** 0.968 **Explained Variance (EV):** 0.513 **MSE:** 146918.712 **MSM:** 128961343.199 **Skewness:** 0.77 **Kurtosis:** 3.526 **F:** 877.773 **DW:** 0.391 **JB:** 66.216	allseason	LASSO MODEL FIT WITH LARS USING BIC OR AIC FOR MODEL SELECTION: Lasso model fit with Lars using BIC or AIC for model selection
8	LassoLarsCV	0.5207	**R-square:** -0.021 **MAPE:** 0.948 **Explained Variance (EV):** 0.51 **MSE:** 303482.203 **MSM:** 225553876.692 **Skewness:** -1.268 **Kurtosis:** 2.128 **F:** 743.219 **DW:** 0.191 **JB:** 179.747	allseason	CROSS-VALIDATED LASSO, USING THE LARS ALGORITHM: The optimization objective for Lasso is:
9	LarsCV	0.5207	**R-square:** 0.506 **MAPE:** 0.968 **Explained Variance (EV):** 0.513 **MSE:** 146918.712 **MSM:** 128961343.199 **Skewness:** 0.77 **Kurtosis:** 3.526 **F:** 877.773 **DW:** 0.391 **JB:** 66.216	allseason	LEAST ANGLE REGRESSION WITH CROSS-VALIDATION: Cross-validated Least Angle Regression model
10	Lars	0.5207	**R-square:** 0.506 **MAPE:** 0.968 **Explained Variance (EV):** 0.513 **MSE:** 146918.712 **MSM:** 128961343.199 **Skewness:** 0.77 **Kurtosis:** 3.526 **F:** 877.773 **DW:** 0.391 **JB:** 66.216	allseason	LEAST ANGLE REGRESSION MODEL A.K.A. LAR: Least-angle regression (LARS) is a regression algorithm for high-dimensional data, developed by Bradley Efron, Trevor Hastie, Iain Johnstone and Robert Tibshirani. LARS is similar to forward stepwise regression. At each step, it finds the predictor most correlated with the response. When there are multiple predictors having equal correlation, instead of continuing along the same predictor, it proceeds in a direction equiangular between the predictors.

Detailed Histograms of Training and Test Data Sets

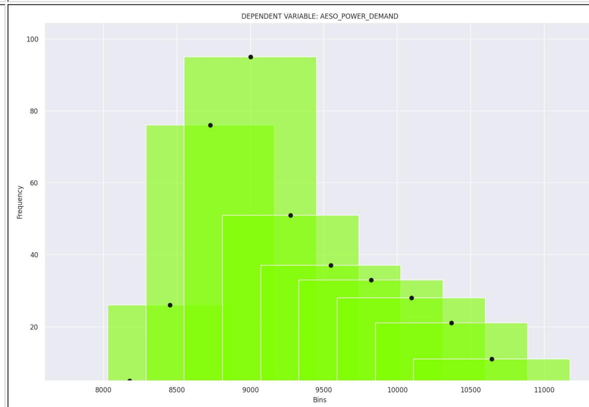
TRAINING VARIABLES | TEST VARIABLES

Training: AESO_POWER_DEMAND



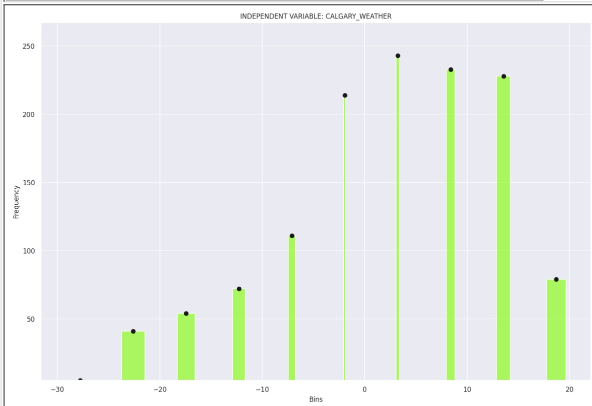
Bins	[7.61e+03, 7.94e+03]	[7.94e+03, 8.27e+03]	[8.27e+03, 8.60e+03]	[8.60e+03, 8.93e+03]	[8.93e+03, 9.26e+03]	[9.26e+03, 9.59e+03]
Count	18	26	85	254	268	229
Share	**1.0%**	**2.0%**	**7.0%**	**20.0%**	**21.0%**	**18.0%**
Total Rows	1280	1280	1280	1280	1280	1280
Min	7.61e+03	7.61e+03	7.61e+03	7.61e+03	7.61e+03	7.61e+03
Max	1.09e+04	1.09e+04	1.09e+04	1.09e+04	1.09e+04	1.09e+04
Number of Bins	6	6	6	6	6	6

Test: AESO_POWER_DEMAND



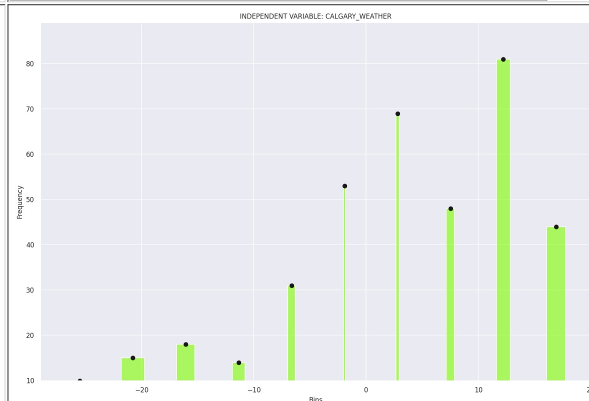
Bins	[8.18e+03, 8.45e+03]	[8.45e+03, 8.73e+03]	[8.73e+03, 9.00e+03]	[9.00e+03, 9.27e+03]	[9.27e+03, 9.55e+03]	[9.55e+03, 9.82e+03]
Count	5	26	76	95	51	37
Share	**1.0%**	**7.0%**	**20.0%**	**25.0%**	**13.0%**	**10.0%**
Total Rows	383	383	383	383	383	383
Min	8.18e+03	8.18e+03	8.18e+03	8.18e+03	8.18e+03	8.18e+03
Max	1.09e+04	1.09e+04	1.09e+04	1.09e+04	1.09e+04	1.09e+04
Number of Bins	6	6	6	6	6	6

Training: CALGARY_WEATHER



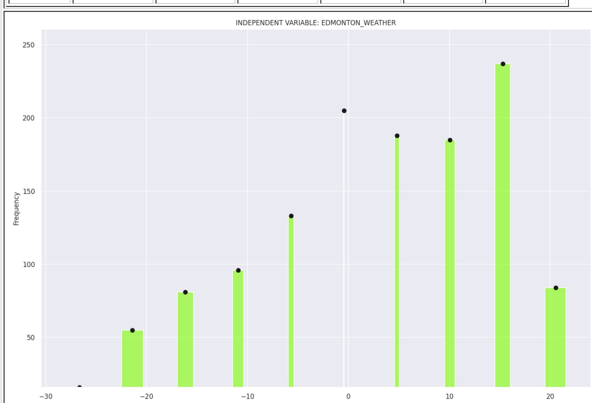
Bins	[-2.78e+01, -2.26e+01]	[-2.26e+01, -1.74e+01]	[-1.74e+01, -1.23e+01]	[-1.23e+01, -7.11e+00]	[-7.11e+00, -1.95e+00]	[-1.95e+00, 3.21e+00]
Count	5	41	54	72	111	214
Share	**0.0%**	**3.0%**	**4.0%**	**6.0%**	**9.0%**	**17.0%**
Total Rows	1280	1280	1280	1280	1280	1280
Min	-2.70e+01	-2.70e+01	-2.70e+01	-2.70e+01	-2.70e+01	-2.70e+01
Max	2.30e+01	2.30e+01	2.30e+01	2.30e+01	2.30e+01	2.30e+01
Number of Bins	6	6	6	6	6	6

Test: CALGARY_WEATHER

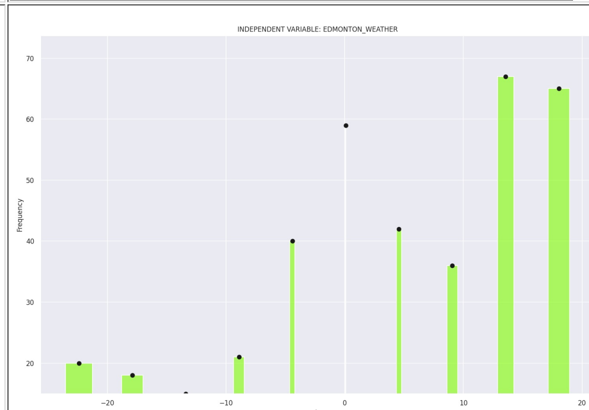


Bins	[-2.55e+01, -2.08e+01]	[-2.08e+01, -1.61e+01]	[-1.61e+01, -1.14e+01]	[-1.14e+01, -6.64e+00]	[-6.64e+00, -1.93e+00]	[-1.93e+00, 2.79e+00]
Count	10	15	18	14	31	53
Share	**3.0%**	**4.0%**	**5.0%**	**4.0%**	**8.0%**	**14.0%**
Total Rows	383	383	383	383	383	383
Min	-2.50e+01	-2.50e+01	-2.50e+01	-2.50e+01	-2.50e+01	-2.50e+01
Max	2.10e+01	2.10e+01	2.10e+01	2.10e+01	2.10e+01	2.10e+01
Number of Bins	6	6	6	6	6	6

Training: EDMONTON_WEATHER

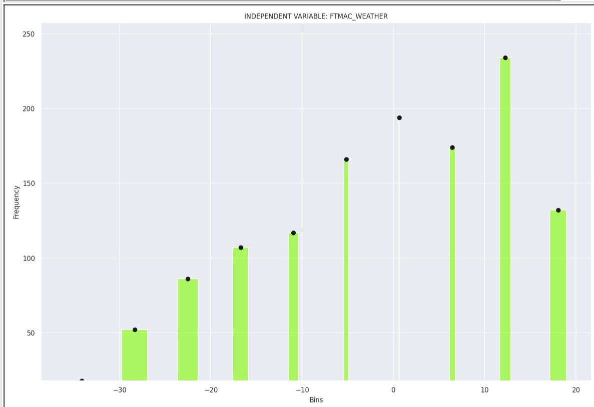


Test: EDMONTON_WEATHER

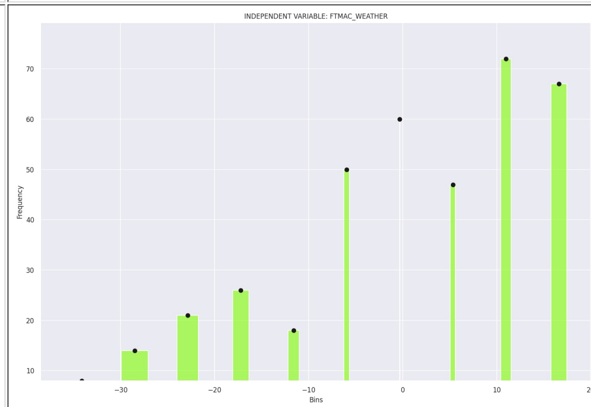


Bins	[-2.66e+01, -2.14e+01]	[-2.14e+01, -1.62e+01]	[-1.62e+01, -1.09e+01]	[-1.09e+01, -5.68e+00]	[-5.68e+00, -4.45e-01]	[-4.45e-01, 4.79e+00]
Count	16	55	81	96	133	205
Share	**1.0%**	**4.0%**	**6.0%**	**8.0%**	**10.0%**	**16.0%**
Total Rows	1280	1280	1280	1280	1280	1280
Min	-2.60e+01	-2.60e+01	-2.60e+01	-2.60e+01	-2.60e+01	-2.60e+01
Max	2.50e+01	2.50e+01	2.50e+01	2.50e+01	2.50e+01	2.50e+01
Number of Bins	6	6	6	6	6	6

Bins	[-2.24e+01, -1.79e+01]	[-1.79e+01, -1.34e+01]	[-1.34e+01, -8.91e+00]	[-8.91e+00, -4.42e+00]	[-4.42e+00, 7.50e-02]	[7.50e-02, 4.57e+00]
Count	20	18	15	21	40	59
Share	**5.0%**	**5.0%**	**4.0%**	**5.0%**	**10.0%**	**15.0%**
Total Rows	383	383	383	383	383	383
Min	-2.20e+01	-2.20e+01	-2.20e+01	-2.20e+01	-2.20e+01	-2.20e+01
Max	2.20e+01	2.20e+01	2.20e+01	2.20e+01	2.20e+01	2.20e+01
Number of Bins	6	6	6	6	6	6

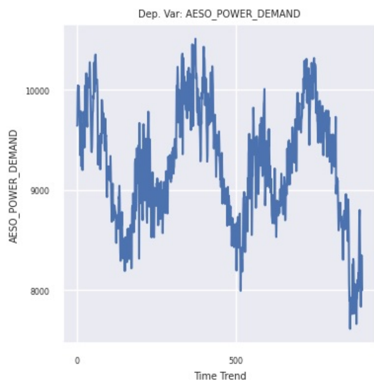


Bins	[-3.41e+01, -2.83e+01]	[-2.83e+01, -2.25e+01]	[-2.25e+01, -1.68e+01]	[-1.68e+01, -1.09e+01]	[-1.09e+01, -5.15e+00]	[-5.15e+00, 6.50e-01]
Count	18	52	86	107	117	166
Share	**1.0%**	**4.0%**	**7.0%**	**8.0%**	**9.0%**	**13.0%**
Total Rows	1280	1280	1280	1280	1280	1280
Min	-3.40e+01	-3.40e+01	-3.40e+01	-3.40e+01	-3.40e+01	-3.40e+01
Max	2.30e+01	2.30e+01	2.30e+01	2.30e+01	2.30e+01	2.30e+01
Number of Bins	6	6	6	6	6	6

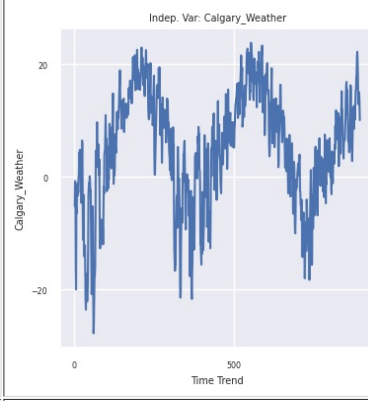


Bins	[-3.41e+01, -2.85e+01]	[-2.85e+01, -2.29e+01]	[-2.29e+01, -1.72e+01]	[-1.72e+01, -1.16e+01]	[-1.16e+01, -5.95e+00]	[-5.95e+00, -3.10e-01]
Count	8	14	21	26	18	50
Share	**2.0%**	**4.0%**	**5.0%**	**7.0%**	**5.0%**	**13.0%**
Total Rows	383	383	383	383	383	383
Min	-3.40e+01	-3.40e+01	-3.40e+01	-3.40e+01	-3.40e+01	-3.40e+01
Max	2.20e+01	2.20e+01	2.20e+01	2.20e+01	2.20e+01	2.20e+01
Number of Bins	6	6	6	6	6	6

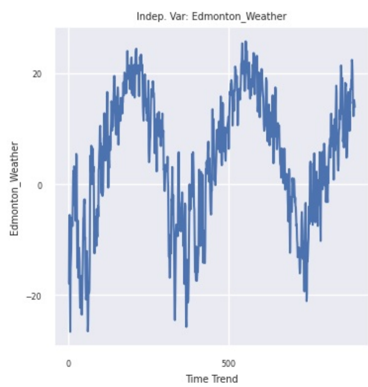
Detailed Graphs of Variables Against Time



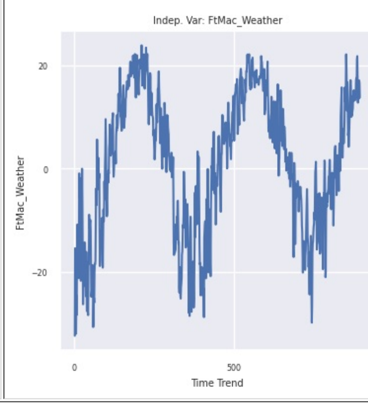
	Stats
Min	7611.0
Max	10510.0
Median	9256.0
Mean	9245.152
STD	586.14
Kurtosis	-0.565
Skewness	-0.173
Coeff. Variation	0.063
Shapiro Normality	0.987



	Stats
Min	-27.75
Max	23.85
Median	6.05
Mean	5.129
STD	10.056
Kurtosis	-0.152
Skewness	-0.538
Coeff. Variation	1.96
Shapiro Normality	0.974

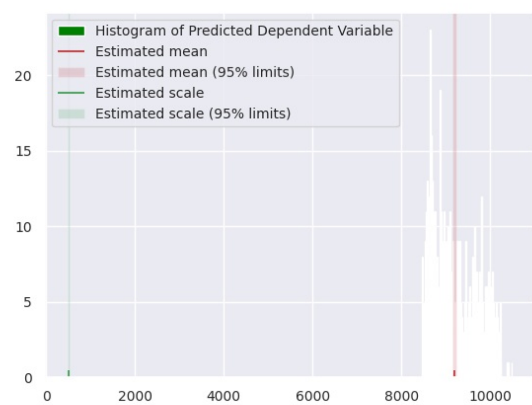


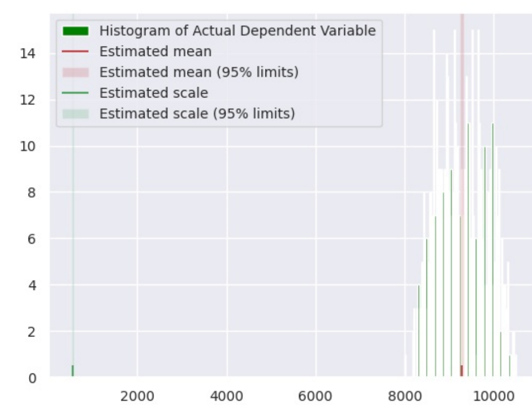
	Stats
Min	-26.64
Max	25.75
Median	6.25
Mean	4.998
STD	11.622
Kurtosis	-0.607
Skewness	-0.44
Coeff. Variation	2.325
Shapiro Normality	0.968



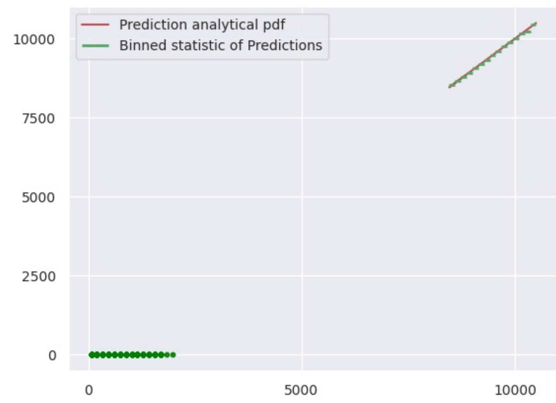
	Stats
Min	-32.4
Max	23.85
Median	2.9
Mean	1.239
STD	13.464
Kurtosis	-0.859
Skewness	-0.388
Coeff. Variation	10.871
Shapiro Normality	0.959

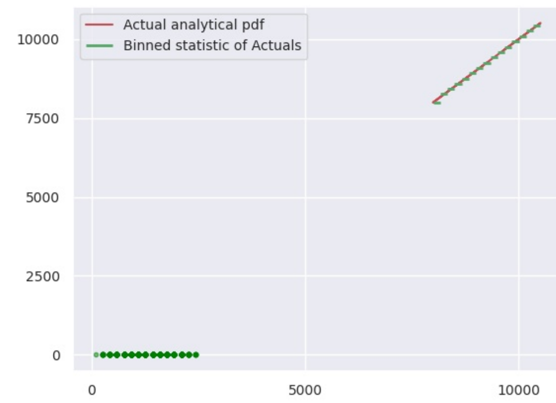
Bayesian Distribution

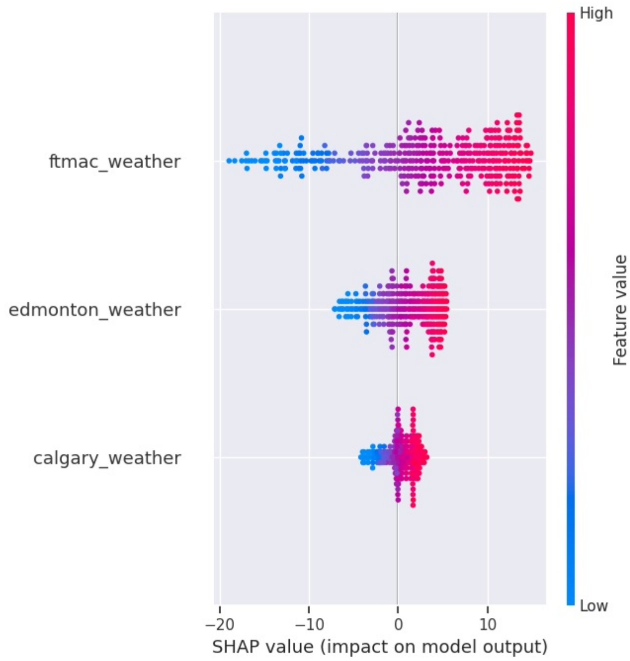




Binned Statistic





MODEL EXPLANATION


- The x-axis represents the model's output values of **AESO_POWER_DEMAND**
- The plot is centered on the x-axis at explainer.expected_value.
- All values are relative to the model's expected value like a linear model's effects are relative to the intercept.
- The y-axis lists the model's features. By default, the features are ordered by descending importance.
- The importance is calculated over the observations plotted. _This is usually different than the importance ordering for the entire dataset.
- In addition to feature importance ordering, the decision plot also supports hierarchical cluster feature ordering and user-defined feature ordering.
- Each observation's prediction is represented by a colored line.
- At the top of the plot, each line strikes the x-axis at its corresponding observation's predicted value. This value determines the color of the line on a spectrum.
- Moving from the bottom of the plot to the top, SHAP values for each feature are added to the model's base value.
- This shows how each feature contributes to the overall prediction.
- At the bottom of the plot, the observations converge at explainer.expected_value.
- The points in the graph are the values of the feature in the training dataset.

FEATURE SELECTION	
RFE Variable (Most important to Least Important)	**Value**
AESO_Power_Demand	0.304
Calgary_Weather	0.235
Edmonton_Weather	0.232
FtMac_Weather	0.229
Best Variable(s) From Genetic Algorithm	
AESO_Power_Demand	
FtMac_Weather	
FtMac_Weather	
Excluded Variable(s)	
Calgary_Weather	
Edmonton_Weather	
PCA for Best Variable(s)	Value
AESO_Power_Demand_pca_1	0.707
AESO_Power_Demand_pca_2	0.707
AESO_Power_Demand_pca_3	-0.081
Calgary_Weather_pca_1	-0.590
Calgary_Weather_pca_2	0.448
Calgary_Weather_pca_3	0.672
FtMac_Weather_pca_1	-0.707
FtMac_Weather_pca_2	0.707
FtMac_Weather_pca_3	-0.736
PCA Explained Variance	Value
PCA1	0.873
PCA2	0.127
PCA3	0.023

- Feature selection shows which variables were more influential than other variables
- It uses two core algorithms: Recursive Feature Elimination (RFE) and Genetic Algorithm to determine influence
- It also performs PCA (principal component analysis) analysis to determine the influence of the best variables in the model
- These results should be used in conjunction with other information as well as theory to establish relevance and confidence in the chosen model formulation

CORRELATION MATRIX

CORRELATED FEATURES

	Feature(s)	Feature(s)	Correlation >= 0.100
0	Calgary_Weather	FtMac_Weather	0.929
1	Edmonton_Weather	FtMac_Weather	0.964
2	Calgary_Weather	Edmonton_Weather	0.970
3	Calgary_Weather	Calgary_Weather	NaN

SUGGESTED CORRELATED FEATURES TO DELETE

	2 Feature(s) to Delete	Correlation
0	Calgary_Weather	0.929
1	Edmonton_Weather	0.964

END OF REPORT

MAADSBML Python Library:https://pypi.org/project/maadsbml/
MAADSBML Docker Container For Windows: https://hub.docker.com/r/maadsdocker/maads-batch-automl-otics
MAADSBML Docker Container For MAC: https://hub.docker.com/r/maadsdocker/maads-batch-automl-otics-arm64
MAADSBML Sample Code and Setup: https://github.com/smaurice101/raspberrypi/tree/main/maadsbml

MAADSBML
Developed and Maintained by: Otics Advanced Analytics, Inc.
Toronto, Ontario, Canada
https://www.otics.ca
Email: support@otics.ca