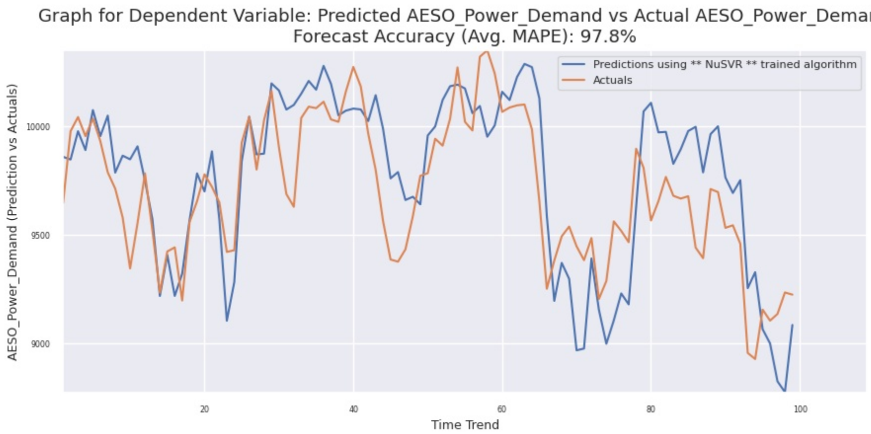
MAADSBML AutoML Report For OTICS ADVANCED ANALYTICS

Generated On: 2024-07-22 16:28:53 (UTC)

Best Model(s) Report For admin_aesopowerdemand_csv



MODEL DESCRIPTION

Model Trained On: 2024/07/22
Training Start Time: 1627
Training End Time: 1628
Was Data Normalized: Yes
Was Data Shuffled: Yes
Deep Analysis: No
Total Training Data Set: 961
Training Data Percentage: 75%
Total Test Data Set: 319
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: 9767.121
STD: 388.896
Kurtosis: -0.459
Skewness: -0.821
Coef. of Variation: 0.040
Shapiro Test for Normality: 0.900
Jarque-Bera Goodness of Fit: 12.116
Anderson: 3.654
KStat: 152767.880
KStatvar: 366028887.706
Wilcox: 0.000
Theil Slope: -2.536

ACTUAL VARIABLE STATS

Mean: 9702.870
STD: 324.119
Kurtosis: -0.698
Skewness: -0.100
Coef. of Variation: 0.033
Shapiro Test for Normality: 0.982
Jarque-Bera Goodness of Fit: 2.198
Anderson: 0.581
KStat: 106114.377
KStatvar: 148831236.050
Wilcox: 0.000
Theil Slope: -3.565

Statistics Showing Comparison Between Prediction and Actuals

Mood(actuals,predictions): -1.934
Pearson(actuals,predictions): 0.786
Kendall Tau(actuals,predictions): 0.610
Ansari(actuals,predictions): 5334.000
Jaccard_distance(actuals,predictions): 1.000
Minkowski_distance(actuals,predictions): 20727.703
Euclidean_distance(actuals,predictions): 2494.113

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/Otics Advanced Analytics_ADMIN_AESOPOWERDEMAND_CSVALLSEASON_AG1_4_NuSVR_normal_961_ensemble_.pkl
2. /maads/agentfilesdocker/networks/Otics Advanced Analytics_ADMIN_AESOPOWERDEMAND_CSVALLSEASON_AG1_4_NuSVR_normal_961_ensemble_scalerx_.pkl
3. /maads/agentfilesdocker/networks/Otics Advanced Analytics_ADMIN_AESOPOWERDEMAND_CSVALLSEASON_AG1_4_NuSVR_normal_961_ensemble_scalery_.pkl

DESCRIPTIVE STATISTICS

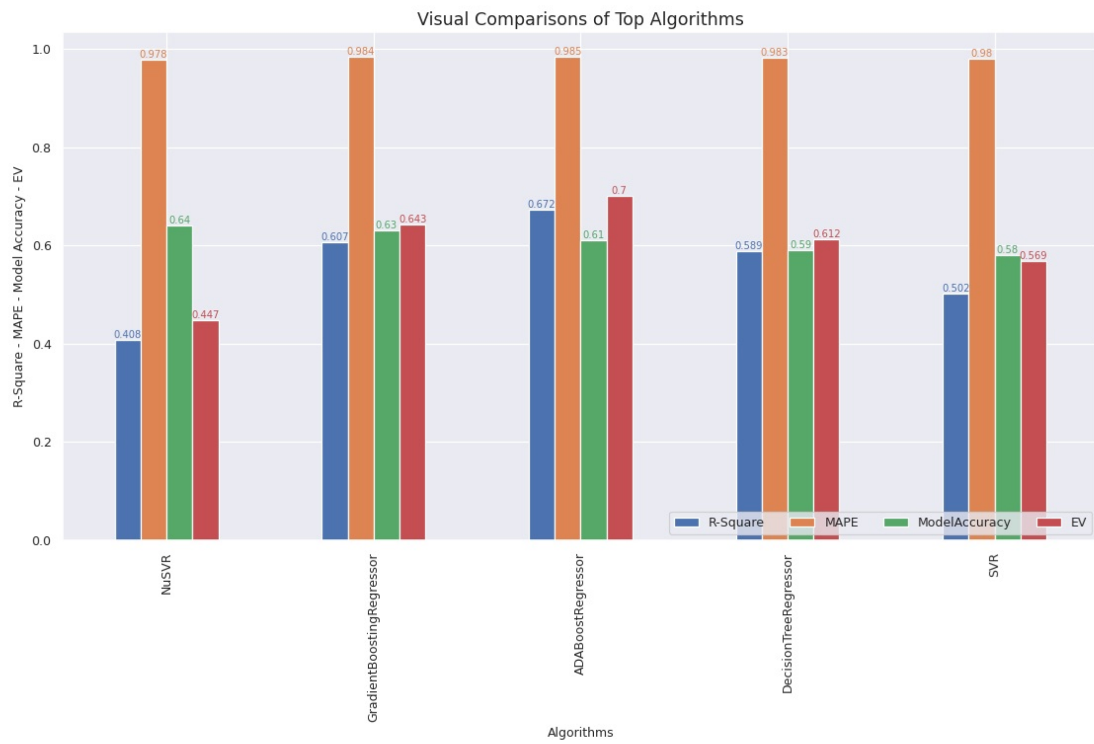
Variables	T-Statistic	Coefficients (alpha,beta)	Count	Mean	STD	MIN	25%	50%	75%	MAX
Calgary_Weather	-35.442	9492.661, -40.134	961.0	5.862	-27.75	10.113	-0.4	6.85	14.0	23.85
Edmonton_Weather	-38.8	9480.686, -37.077	961.0	5.916	-26.64	11.759	-2.25	7.16	16.1	25.75
FtMac_Weather	-40.088	9346.752, -32.447	961.0	2.367	-32.4	13.694	-7.65	4.56	14.8	23.85
AESO_POWER_DEMAND	NA	NA	961.0	9227.732	7611.0	577.371	8790.0	9225.0	9661.0	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
NuSVR	Nu Support Vector Regression.: Similar to NuSVC, for regression, uses a parameter nu to control the number of support vectors. However, unlike NuSVC, where nu replaces C, here nu replaces the parameter epsilon of epsilon-SVR.	NuSVR(C=7.774542443884447, degree=4, gamma='auto', nu=0.47052831690658303) **R-square:** 0.408 **Mean Squared Error (MSE):** 62205.972 **Skewness:** -0.576 **Kurtosis:** 2.329 **Mean Square Model (MSM):** 15536838.842 **F-Statistic (F):** 249.764 **Jarque-Bera (JB):** 7.403 **Explained Variance (EV):** 0.447 **Multicolinearity Test (Avg. VIF):** 19.321 **Heteroscedasticity Test (Avg P-Value):** 0.000 **(Based on White Test, there seems to be heteroscedasticity in the model)** **Autocorrelation (Durbin-Watson) Test:** 0.592 **(Based on DW Test - there seems to be autocorrelation in your model)**	0.978	1 - 12	allseason

TOP 5 ALGORITHMS FOR ALLSEASON

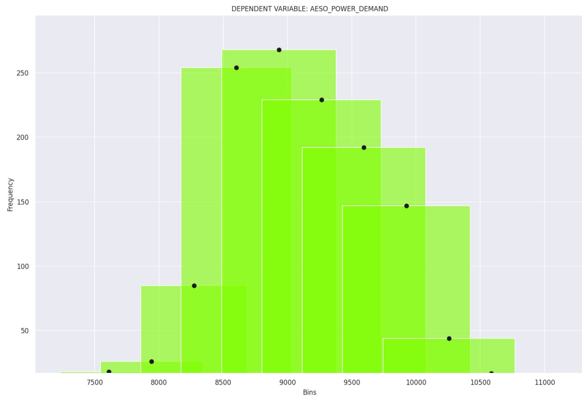


Num	Algorithm	Model Accuracy	Details	Season	Description
1	NuSVR	0.6376	**R-square:** 0.408 **MAPE:** 0.978 **Explained Variance (EV):** 0.447 **MSE:** 62205.972 **MSM:** 15536838.842 **Skewness:** -0.576 **Kurtosis:** 2.329 **F:** 249.764 **DW:** 0.592 **JB:** 7.403	allseason	NU SUPPORT VECTOR REGRESSION.: Similar to NuSVC, for regression, uses a parameter nu to control the number of support vectors. However, unlike NuSVC, where nu replaces C, here nu replaces the parameter epsilon of epsilon-SVR.
2	GradientBoostingRegressor	0.6330	**R-square:** 0.607 **MAPE:** 0.984 **Explained Variance (EV):** 0.643 **MSE:** 41311.165 **MSM:** 12482469.672 **Skewness:** -0.678 **Kurtosis:** 3.067 **F:** 302.157 **DW:** 0.757 **JB:** 7.677	allseason	GRADIENT BOOSTING FOR REGRESSION.: GB builds an additive model in a forward stage-wise fashion; it allows for the optimization of arbitrary differentiable loss functions. In each stage a regression tree is fit on the negative gradient of the given loss function.
3	ADABoostRegressor	0.6116	**R-square:** 0.672 **MAPE:** 0.985 **Explained Variance (EV):** 0.7 **MSE:** 34421.87 **MSM:** 12628011.395 **Skewness:** -0.28 **Kurtosis:** 3.402 **F:** 366.86 **DW:** 0.803 **JB:** 1.977	allseason	ADABOOST REGRESSOR: Ada boost
4	DecisionTreeRegressor	0.5903	**R-square:** 0.589 **MAPE:** 0.983 **Explained Variance (EV):** 0.612 **MSE:** 43149.155 **MSM:** 14106697.959 **Skewness:** -0.194 **Kurtosis:** 3.238 **F:** 326.929 **DW:** 0.816 **JB:** 0.861	allseason	DECISION TREE REGRESSOR: Decision Tree Regressor
5	SVR	0.5752	**R-square:** 0.502 **MAPE:** 0.980 **Explained Variance (EV):** 0.569 **MSE:** 52352.674 **MSM:** 16167378.75 **Skewness:** -0.808 **Kurtosis:** 2.51 **F:** 308.817 **DW:** 0.632 **JB:** 11.888	allseason	EPSILON-SUPPORT VECTOR REGRESSION.: The method of Support Vector Classification can be extended to solve regression problems. This method is called Support Vector Regression.

Detailed Histograms of Training and Test Data Sets

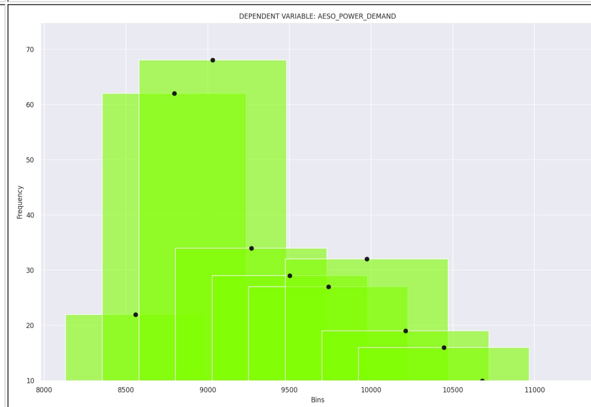
TRAINING VARIABLES	TEST VARIABLES

TRAINING VARIABLES

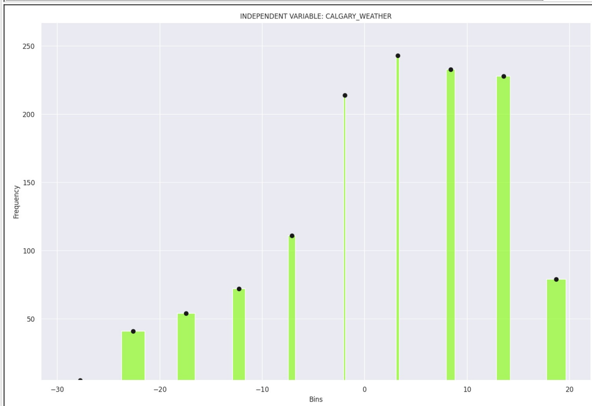


Bins	[7.61e+03, 7.94e+03]	[7.94e+03, 8.27e+03]	[8.27e+03, 8.60e+03]	[8.60e+03, 8.93e+03]	[8.93e+03, 9.26e+03]	[9.26e+03, 9.59e+03]
Count	18	26	85	254	268	229
Share	**1.0%**	**2.0%**	**7.0%**	**20.0%**	**21.0%**	**18.0%**
Total Rows	1280	1280	1280	1280	1280	1280
Min	7.61e+03	7.61e+03	7.61e+03	7.61e+03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

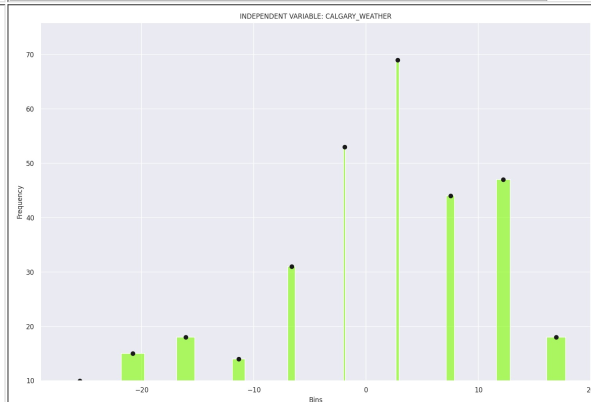
TEST VARIABLES



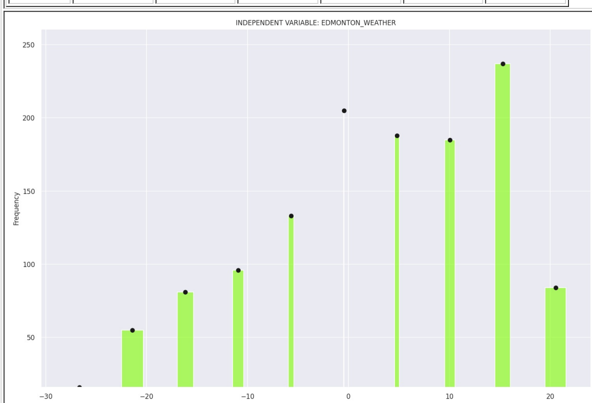
Bins	[8.56e+03, 8.79e+03]	[8.79e+03, 9.03e+03]	[9.03e+03, 9.27e+03]	[9.27e+03, 9.50e+03]	[9.50e+03, 9.74e+03]	[9.74e+03, 9.97e+03]
Count	22	62	68	34	29	27
Share	**7.0%**	**19.0%**	**21.0%**	**11.0%**	**9.0%**	**8.0%**
Total Rows	319	319	319	319	319	319
Min	8.56e+03	8.56e+03	8.56e+03	8.56e+03	8.56e+03	8.56e+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

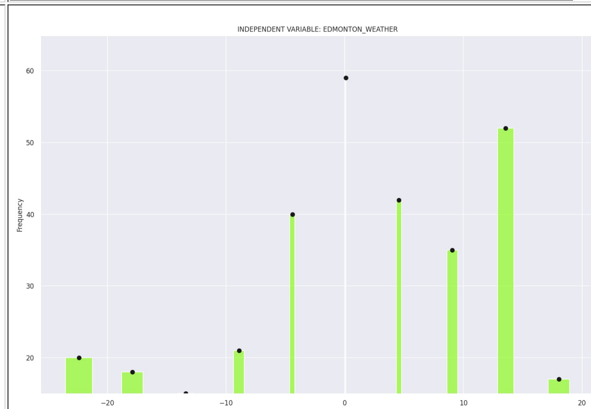


Bins	[-2.78e+01, -2.26e+01]	[-2.26e+01, -1.74e+01]	[-1.74e+01, -1.23e+01]	[-1.23e+01, -7.11e+00]	[-7.11e+00, -1.95e+00]	[-1.95e+00, 3.21e+00]
Count	5	41	54	72	111	214
Share	**0.0%**	**3.0%**	**4.0%**	**6.0%**	**9.0%**	**17.0%**
Total Rows	1280	1280	1280	1280	1280	1280
Min	-2.70e+01	-2.70e+01	-2.70e+01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

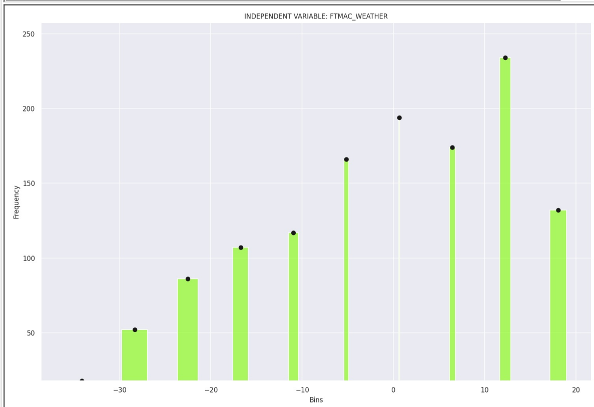


Bins	[-2.55e+01, -2.08e+01]	[-2.08e+01, -1.61e+01]	[-1.61e+01, -1.14e+01]	[-1.14e+01, -6.64e+00]	[-6.64e+00, -1.93e+00]	[-1.93e+00, 2.79e+00]
Count	10	15	18	14	31	53
Share	**3.0%**	**5.0%**	**6.0%**	**4.0%**	**10.0%**	**17.0%**
Total Rows	319	319	319	319	319	319
Min	-2.50e+01	-2.50e+01	-2.50e+01	-2.50e+01	-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



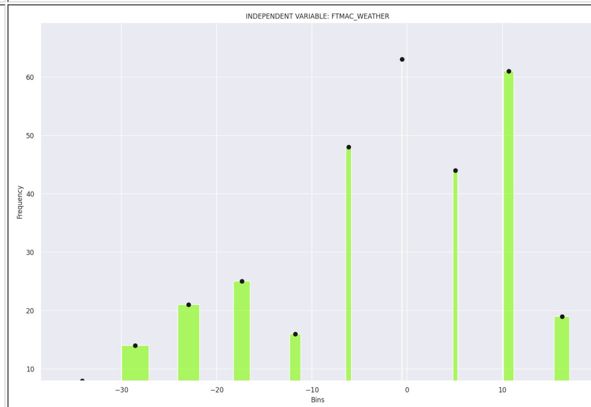


Bins	[-2.66e+01, -2.14e+01]	[-2.14e+01, -1.62e+01]	[-1.62e+01, -1.09e+01]	[-1.09e+01, -5.68e+00]	[-5.68e+00, -4.45e-01]	[-4.45e-01, 4.79e+00]
Count	16	55	81	96	133	205
Share	**1.0%**	**4.0%**	**6.0%**	**8.0%**	**10.0%**	**16.0%**
Total Rows	1280	1280	1280	1280	1280	1280
Min	-2.60e+01	-2.60e+01	-2.60e+01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



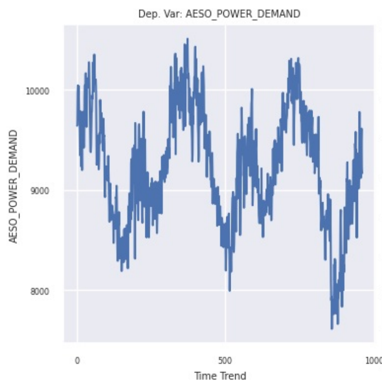
Bins	[-3.41e+01, -2.83e+01]	[-2.83e+01, -2.25e+01]	[-2.25e+01, -1.68e+01]	[-1.68e+01, -1.09e+01]	[-1.09e+01, -5.15e+00]	[-5.15e+00, 6.50e-01]
Count	18	52	86	107	117	166
Share	**1.0%**	**4.0%**	**7.0%**	**8.0%**	**9.0%**	**13.0%**
Total Rows	1280	1280	1280	1280	1280	1280
Min	-3.40e+01	-3.40e+01	-3.40e+01	-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	[-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	**6.0%**	**6.0%**	**5.0%**	**7.0%**	**13.0%**	**18.0%**
Total Rows	319	319	319	319	319	319
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

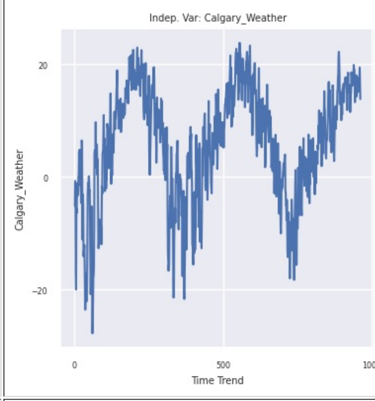


Bins	[-3.41e+01, -2.85e+01]	[-2.85e+01, -2.29e+01]	[-2.29e+01, -1.74e+01]	[-1.74e+01, -1.18e+01]	[-1.18e+01, -6.15e+00]	[-6.15e+00, -5.50e-01]
Count	8	14	21	25	16	48
Share	**3.0%**	**4.0%**	**7.0%**	**8.0%**	**5.0%**	**15.0%**
Total Rows	319	319	319	319	319	319
Min	-3.40e+01	-3.40e+01	-3.40e+01	-3.40e+01	-3.40e+01	-3.40e+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

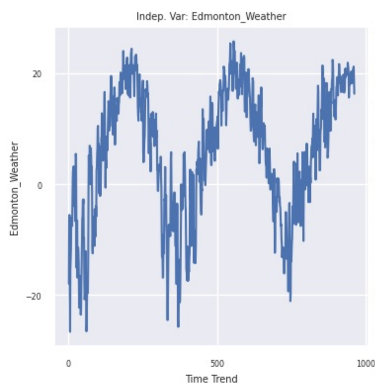
Detailed Graphs of Variables Against Time



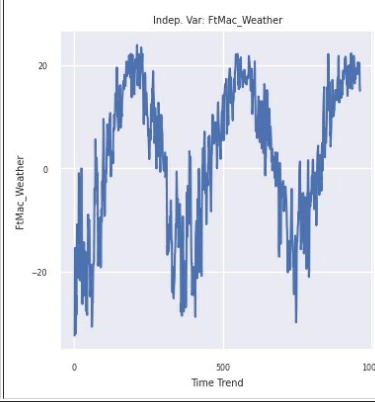
	Stats
Min	7611.0
Max	10510.0
Median	9225.0
Mean	9227.732
STD	577.071
Kurtosis	-0.538
Skewness	-0.122
Coeff. Variation	0.063
Shapiro Normality	0.989



	Stats
Min	-27.75
Max	23.85
Median	6.85
Mean	5.862
STD	10.108
Kurtosis	-0.137
Skewness	-0.618
Coeff. Variation	1.724
Shapiro Normality	0.964

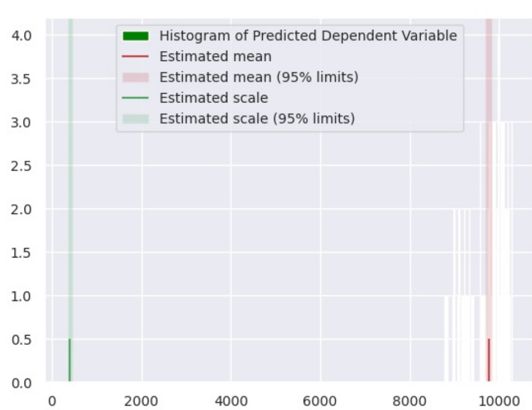


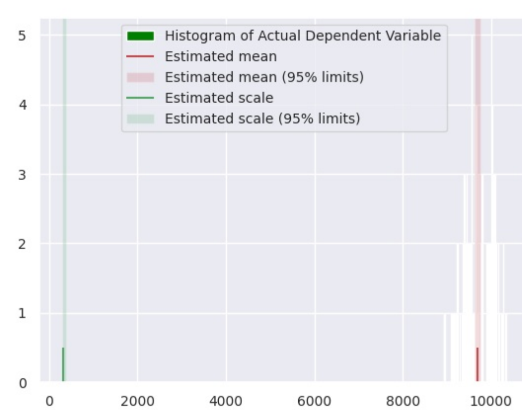
	Stats
Min	-26.64
Max	25.75
Median	7.16
Mean	5.916
STD	11.753
Kurtosis	-0.6
Skewness	-0.518
Coeff. Variation	1.987
Shapiro Normality	0.956

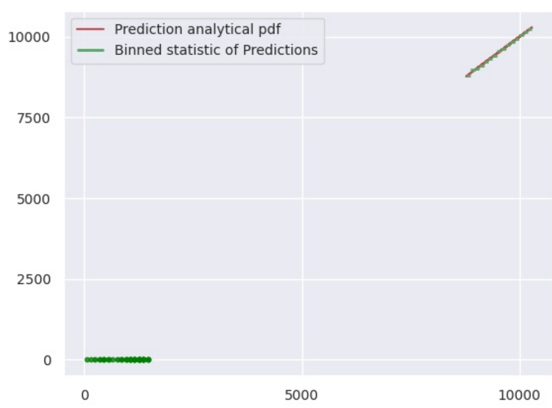
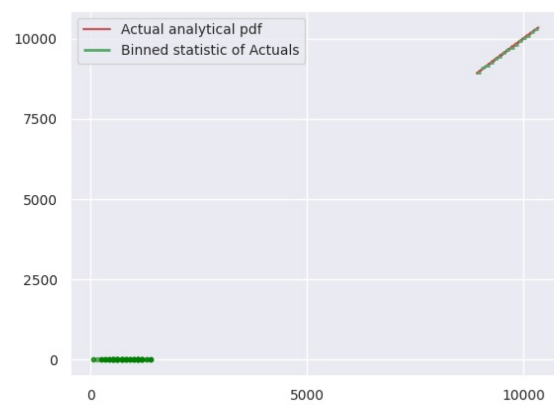


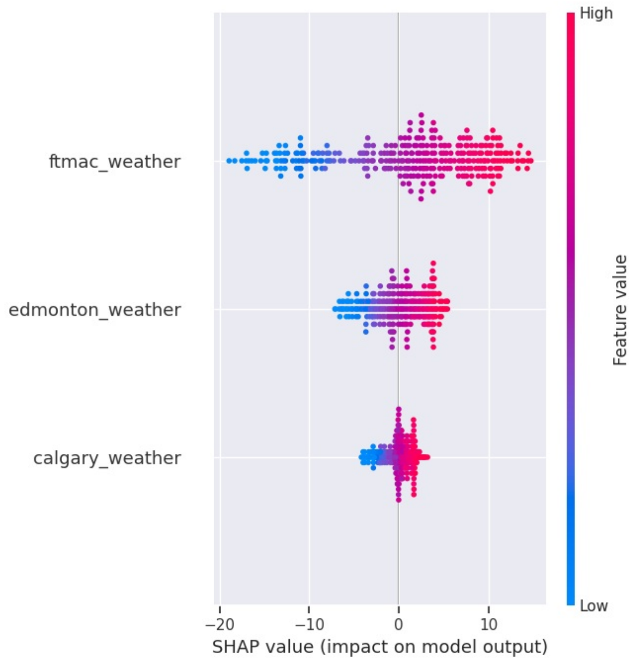
	Stats
Min	-32.4
Max	23.85
Median	4.56
Mean	2.367
STD	13.687
Kurtosis	-0.848
Skewness	-0.46
Coeff. Variation	5.782
Shapiro Normality	0.947

Bayesian Distribution





Binned Statistic	
	

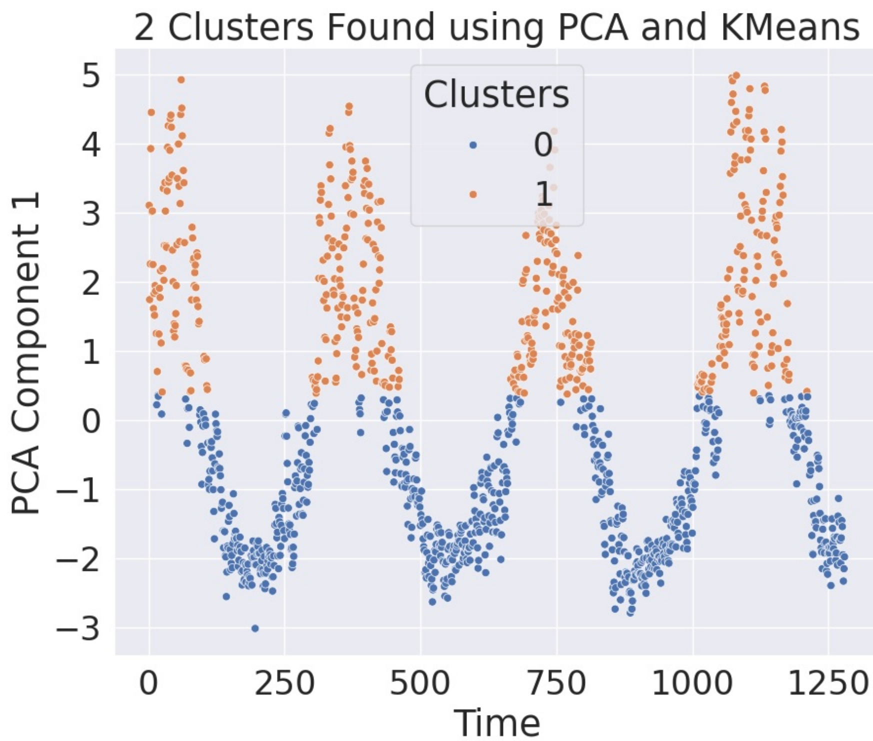
MODEL EXPLANATION


- 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**
Calgary_Weather	0.234
Edmonton_Weather	0.232
FtMac_Weather	0.227
Best Variable(s) From Genetic Algorithm	
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
FtMac_Weather_pca_1	-0.707
FtMac_Weather_pca_2	0.707
PCA Explained Variance	Value
PCA1	0.873
PCA2	0.127

- 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

CLUSTER ANALYSIS: PCA and KMeans



Over 80% of the variance in your data is explaned by 1 principal components.

Optimal clusters are **2** with a silhouette score of **0.534** using euclidean distance.
NOTE: Only two principal components are shown in the 2D graph because your variables have been reduced by dimensionality reduction using PCA. However, the entire data set can be found here:
/maads/agentfilesdocker/dist/maadsweb/csvuploads/admin_aesopowerdemand_csv_clusters_pcakmeans.csv for further analysis.

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