

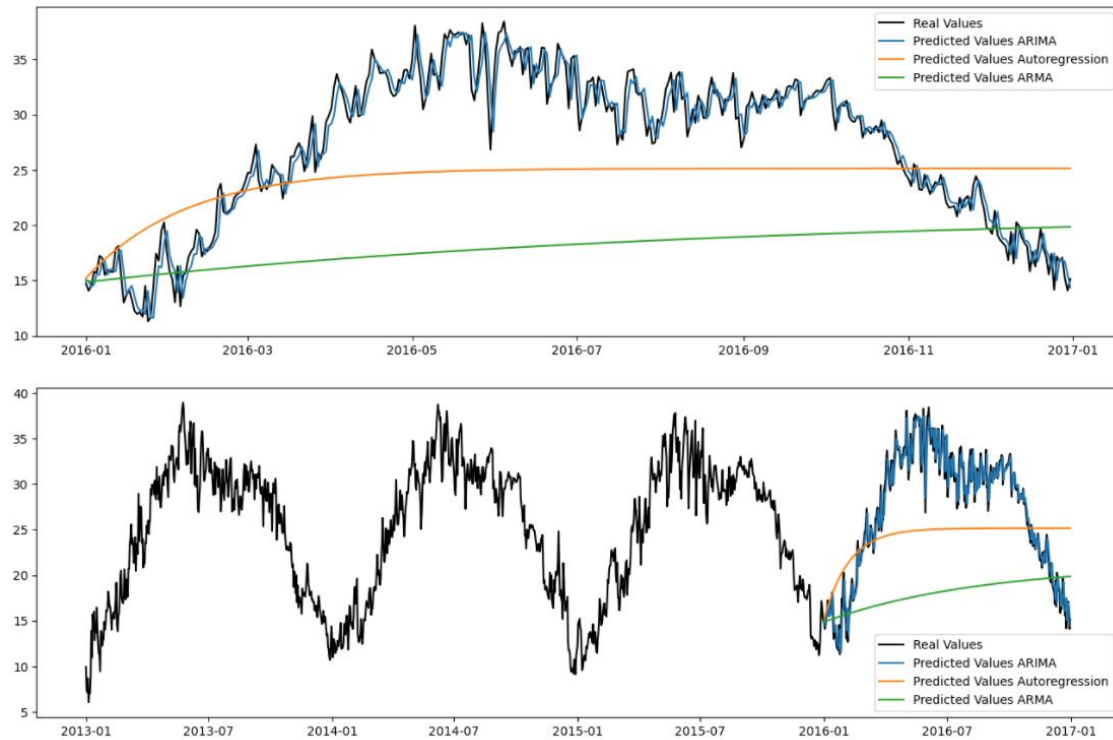
# M8 Practical Challenge: Time-Series Prediction— Python versus DataRobot

## Table of Contents

Python Models .....	2
DataRobot Model .....	3
Comparisons .....	6

## Python Models

For the reviews of the Python Models, they are explained and done in the jupyter Notebook. Here I'm only going to put the results:



Test ARIMA RMSE:	1.5885418984004303
Test Autoregression RMSE:	6.5555040314708455
Test ARMA RMSE:	11.261958580452083

We can see from the reviews that the bet model is the ARIMA model with a p, d, q of 5, 1, 0. It achieve the best RMSE, of only 1.58

# DataRobot Model

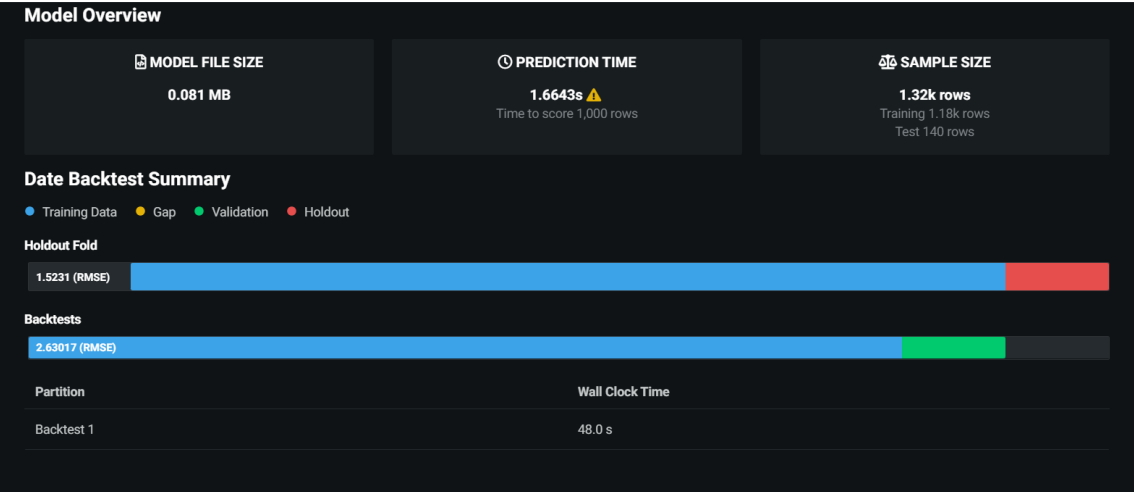
We get the next Information:

Menu Search Add new model Filter Models Deploy Automodel Export				Metric RMSE		
Model Name & Description		Feature List & Sample Size		Backtest 1	Holdout	
<div>Generalized Additive2 Model</div> <div>Ordinal encoding of categorical variables   Missing Values Imputed   Generalized Additive2 Model   Text fit on Residuals (L2 / Least-Squares Loss)</div> <div>M21 BP91 MONO</div> <div>RECOMMENDED FOR DEPLOYMENT</div> <div>PREPARED FOR DEPLOYMENT</div>		<div>Informative Features</div> <div>Start Date: 2013-10-05</div> <div>End Date: 2017-01-01</div>		N/A	N/A	
<div>Generalized Additive2 Model</div> <div>Ordinal encoding of categorical variables   Missing Values Imputed   Generalized Additive2 Model   Text fit on Residuals (L2 / Least-Squares Loss)</div> <div>M11 BP91 MONO</div>		<div>Informative Features</div> <div>3 years • 2 months • 26 days</div>		2.6302		
<div>AVG Blender</div> <div>Average Blender</div> <div>M23 M7+11</div>		<div>Informative Features</div> <div>3 years • 2 months • 26 days</div>		2.6930		
<div>eXtreme Gradient Boosted Trees Regressor</div> <div>Tree-based Algorithm Preprocessing v1</div> <div>M7 BP87 MONO</div>		<div>Informative Features</div> <div>3 years • 2 months • 26 days</div>		2.8205		
<div>Light Gradient Boosting on ElasticNet Predictions</div> <div>One Hot Encoding   Dummy Data Generation   Standardize   Ordinal encoding of categorical variables   Ridge Regressor</div>		<div>Informative Features</div>				

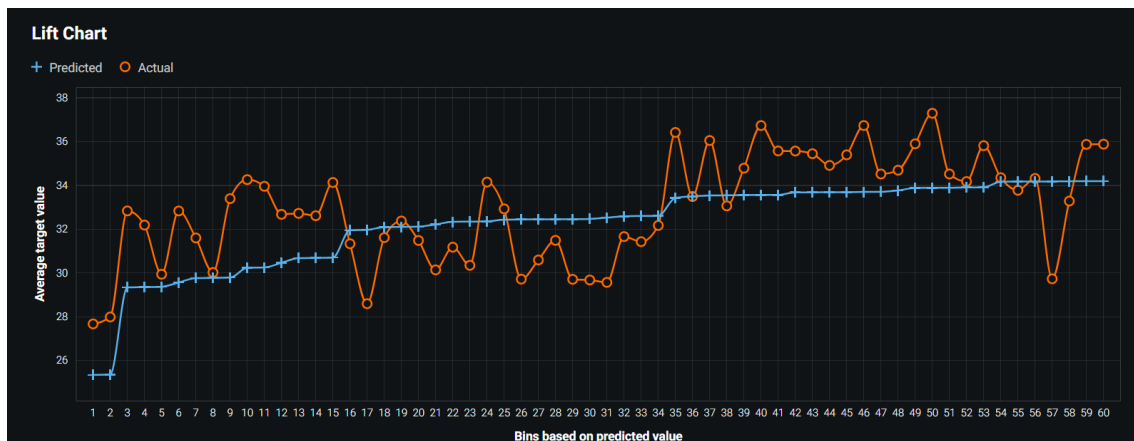
We can see that the best 3 Models are:

Generalized Additive2 Model (RMSE: 2.6302), AVG Blender (RMSE: 2.6930), eXtreme Gradient Boosted Trees Regressor (RMSE: 2.8205)

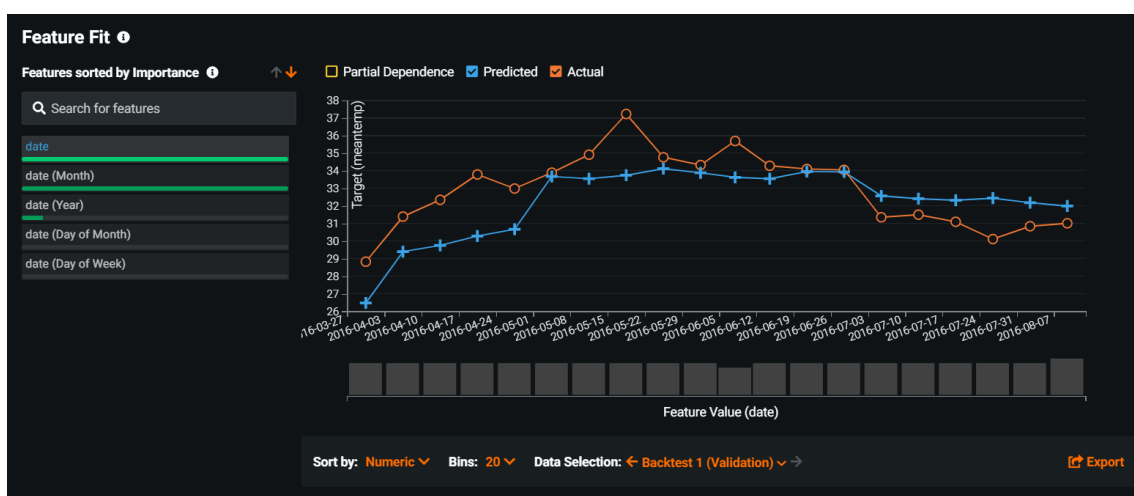
We can see from the RMSE, that the Generalized Additive2 Model is the best one, So lets see what it does:



We can see the Lift Chart



As well as the Feature Fit



Important to check the accuracy over time

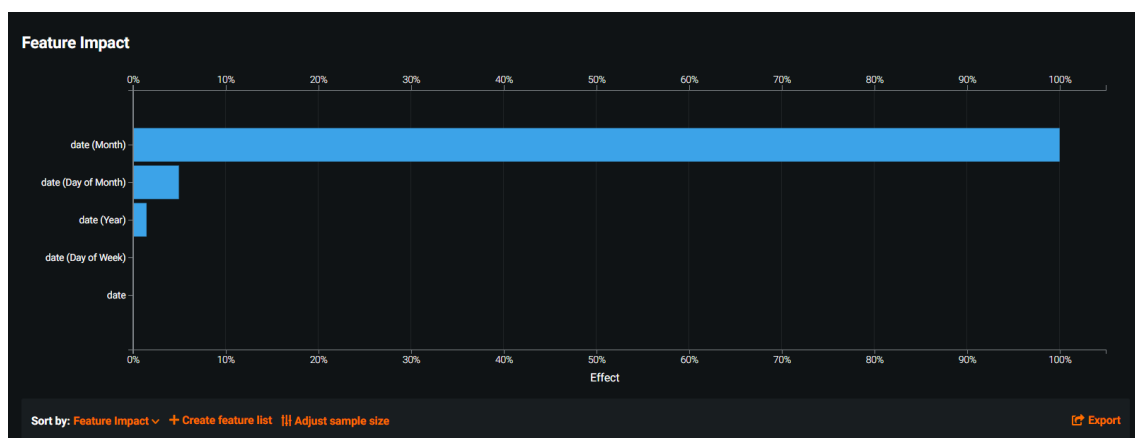


We can see that the predicted values are really close to the real values.

We can also see the Residuals, we can see that the highest error is 7, but this is an exception, as almost all the rest errors follow below 5.



We can also see that the highest impact is the Month by a huge margin. While the Day of the week have no effect. This is because we are measuring temperature, so its normal that the month have a huge impact.



## Comparisons

We can conclude that the ARIMA model on Python is a better model, as it gets a better RMSE, than the models that the Data Robot offers. The ARIMA RMSE is 1.58854, while the best model on Data Robot is Generalized Additive2 Model (RMSE: 2.6302),