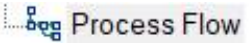
Tutorial 3

SAS EG doesn't have an easy way to do trailing moving average or centered moving average. So, here we are only going to deal with exponential smoothing, Holt's and Winter's models.

Before you start: Create a new process flow

This tutorial assumes that you have already created a project. You need to open it. For the next tutorial we will create a new Process Flow, which will let us separate tutorials from each other and keep our project more organized (described in tutorial 2).

Click on the icon that looks like a sheet of paper → Process Flow → Right click on the icon Process Flow once you see in the pane on the left  **→ Call it tutorial3.**

When you work with the materials for Tutorial3, make sure you left click on that Process Flow so that all the new datasets you open and create, get opened under that tutorial.

For this tutorial we are going to use QUERY_FOR_BIKESHARINGDAILY, that we created in tutorial 1.

Assign a permanent libname statement!

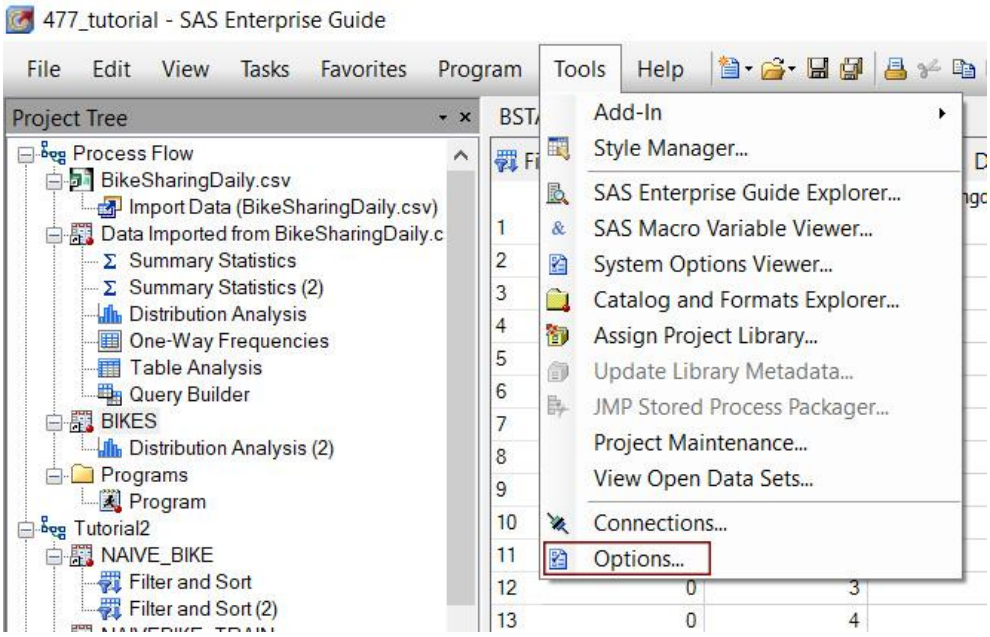
YOU ONLY NEED TO DO IT ONCE ON YOUR COMPUTER! IF YOU USE SCHOOL'S COMPUTERS, THEN YOU WILL NEED TO RUN OUR REGULAR LIBNAME STATEMENT.

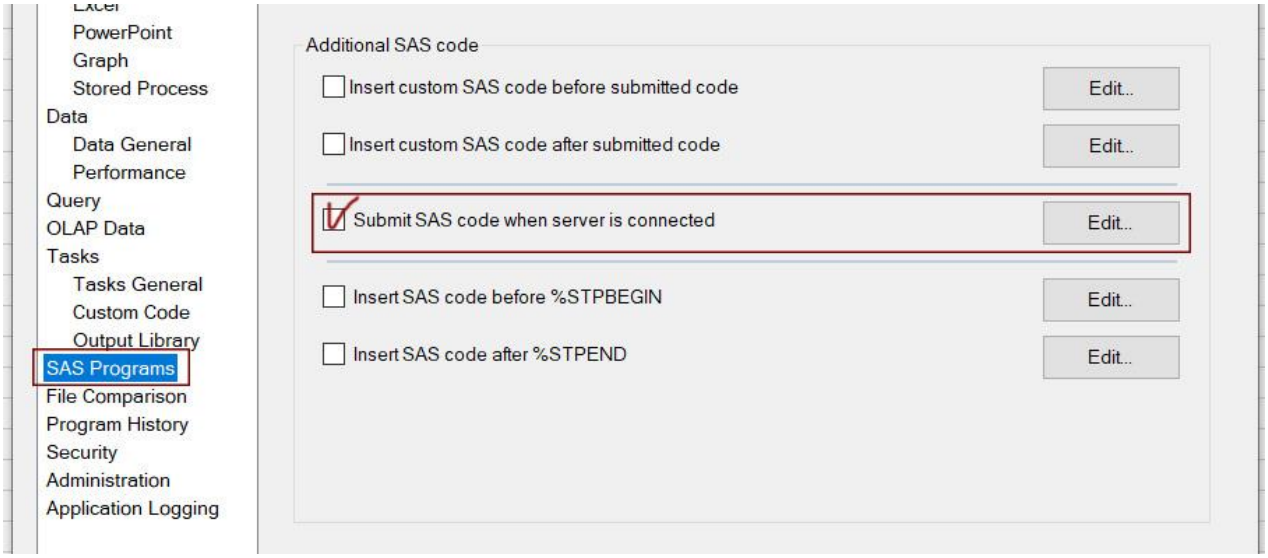
If you work from your own computer, you don't have to run a libname statement every time you open a new session. You can only do it once. Here is how:

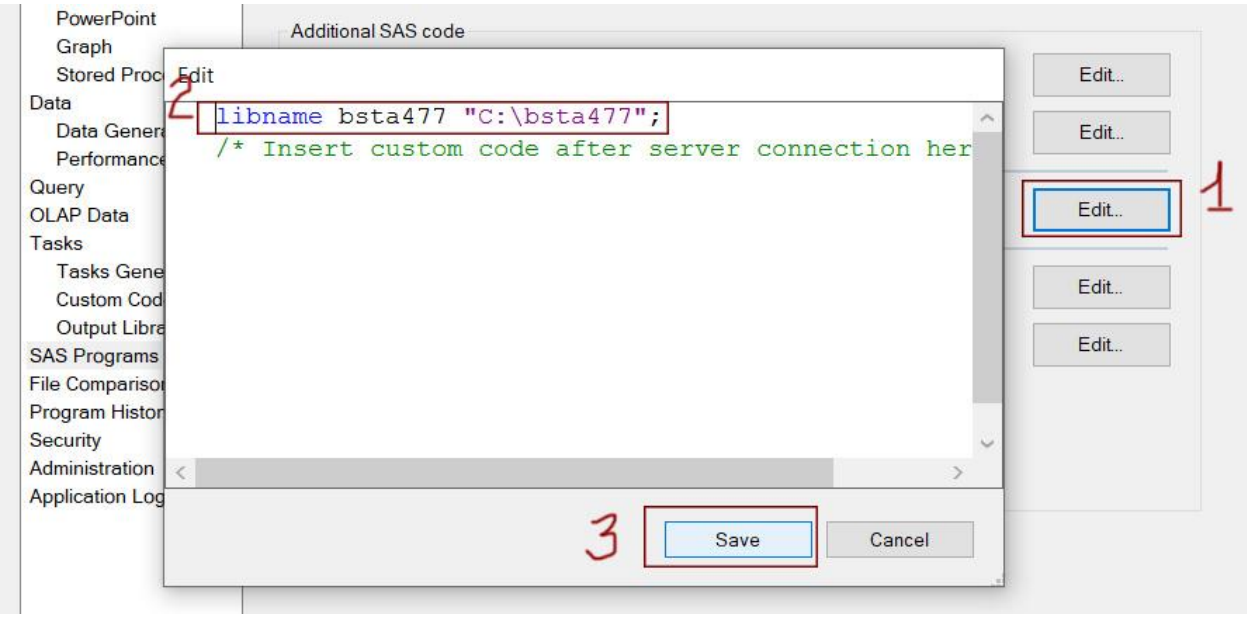
Open the existing project → Tools → Options → SAS Programs (in the tab on the left) → Check submit SAS code when server is connected → Edit → Insert the libname statement:

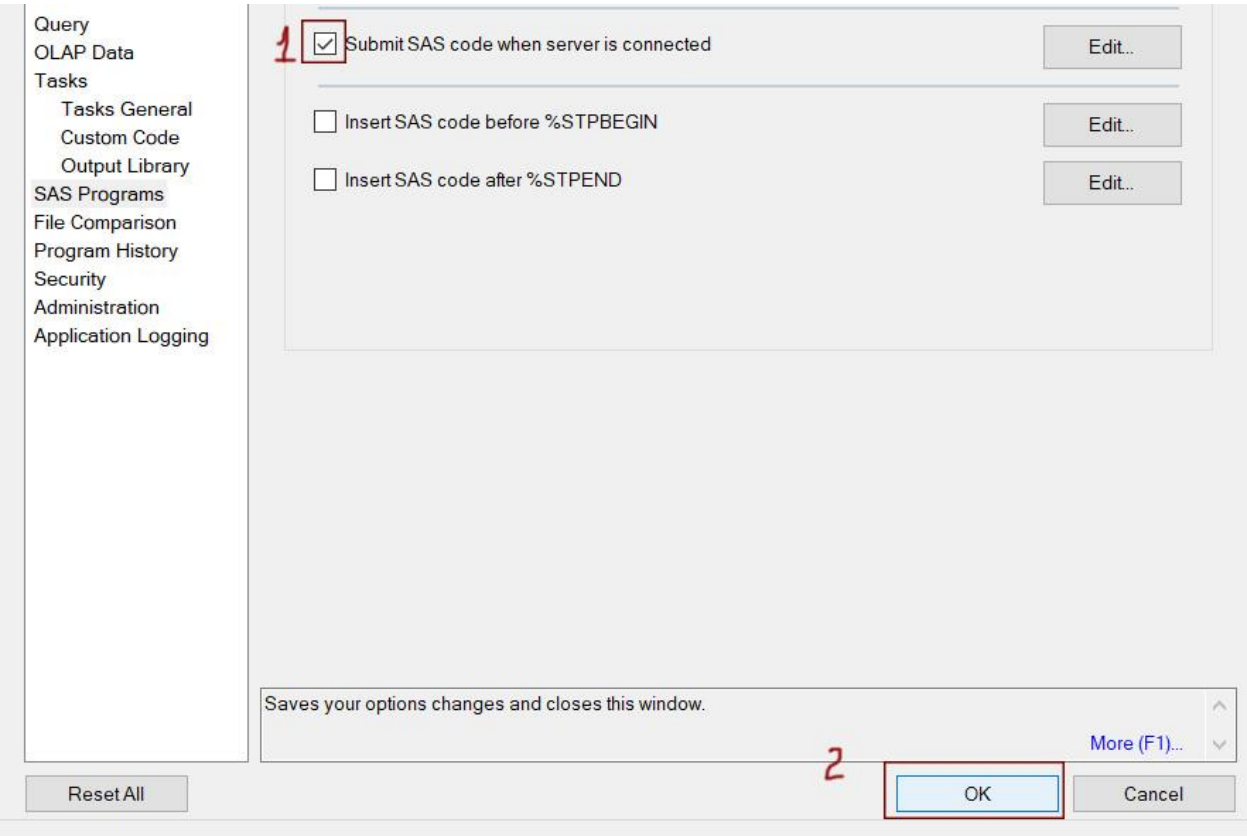
```
libname bsta477 "C:\bsta477";
```

→ Save → Ok







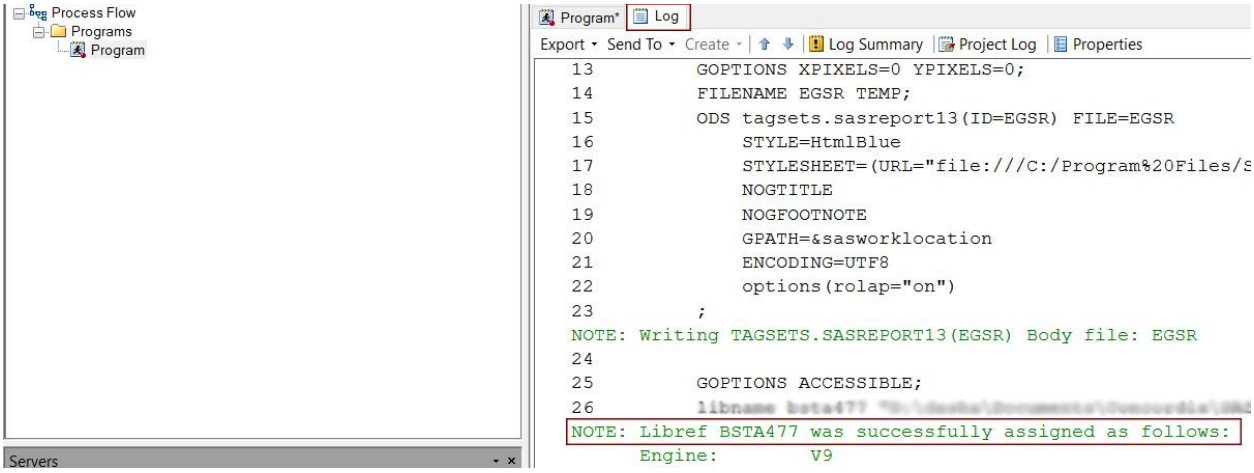


If you are working from the school's computer, then just open a new project and run a libname statement in a SAS program. Make sure that you are in the TUTORIAL3 Process Workflow.

Click on the icon that looks like a sheet of paper → Choose program → Type the libname statement (as shown below)

```
libname bsta477 "C:\bsta477";
```

Remember to always check the log. It will show whether the library has been successfully assigned or not!

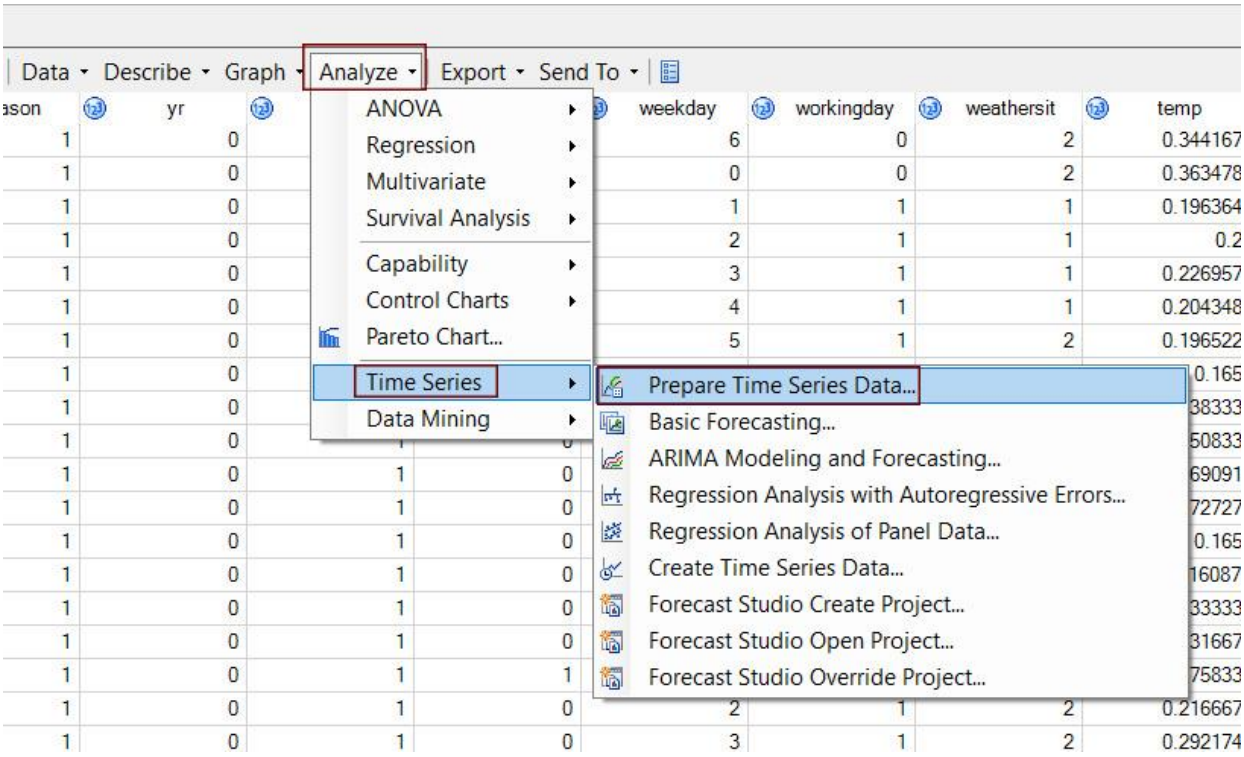


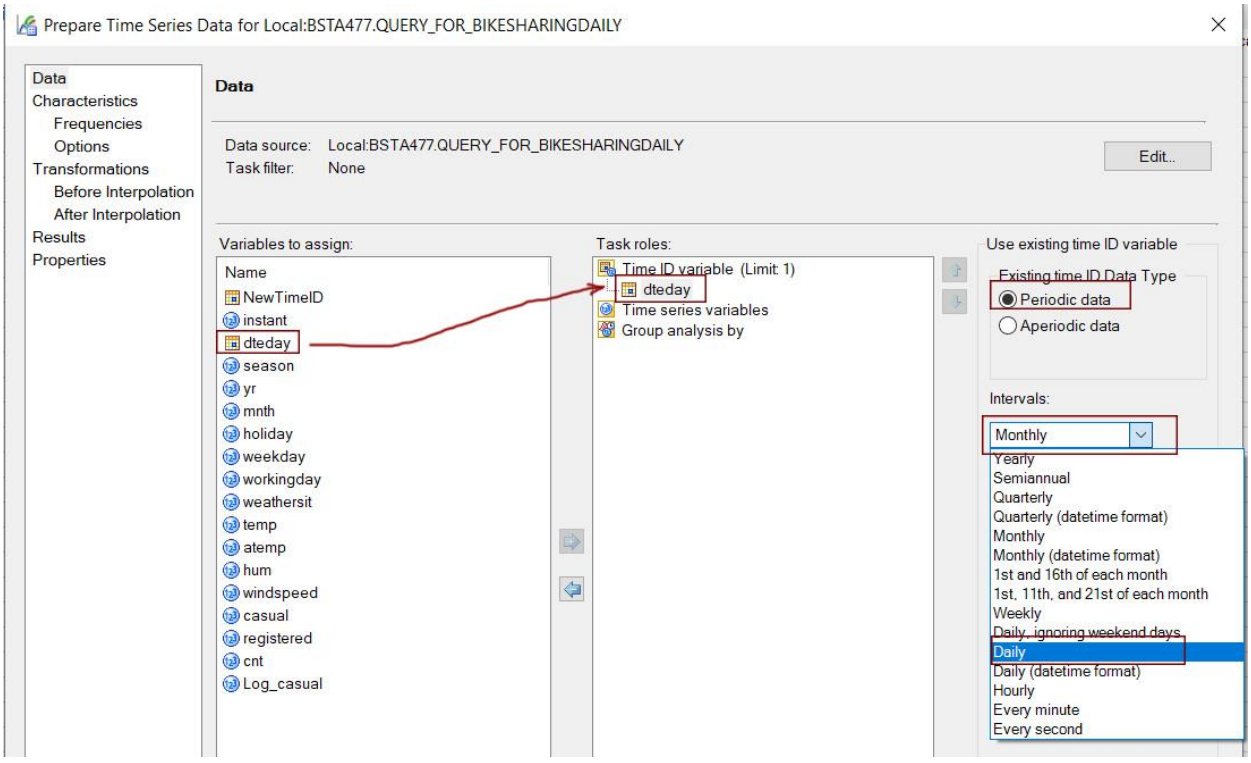
Preparing data for analysis

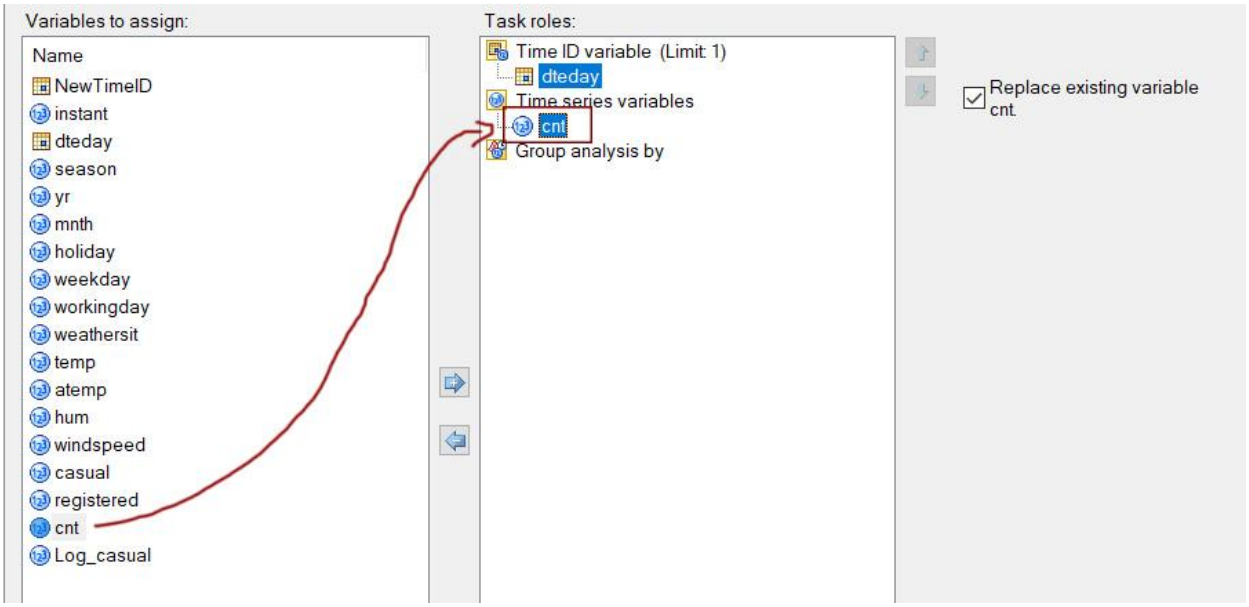
In our dataset we have rides accumulated on a daily basis, beginning the first of January 2011 and ending the 31st December 2012. We may think that predicting rides by days is not practical, and we may want to predict them on a monthly basis. To do that, first, we need to have rides per month, not per day. We are going to forecast a total number of rides that are represented by the **cnt** variable.

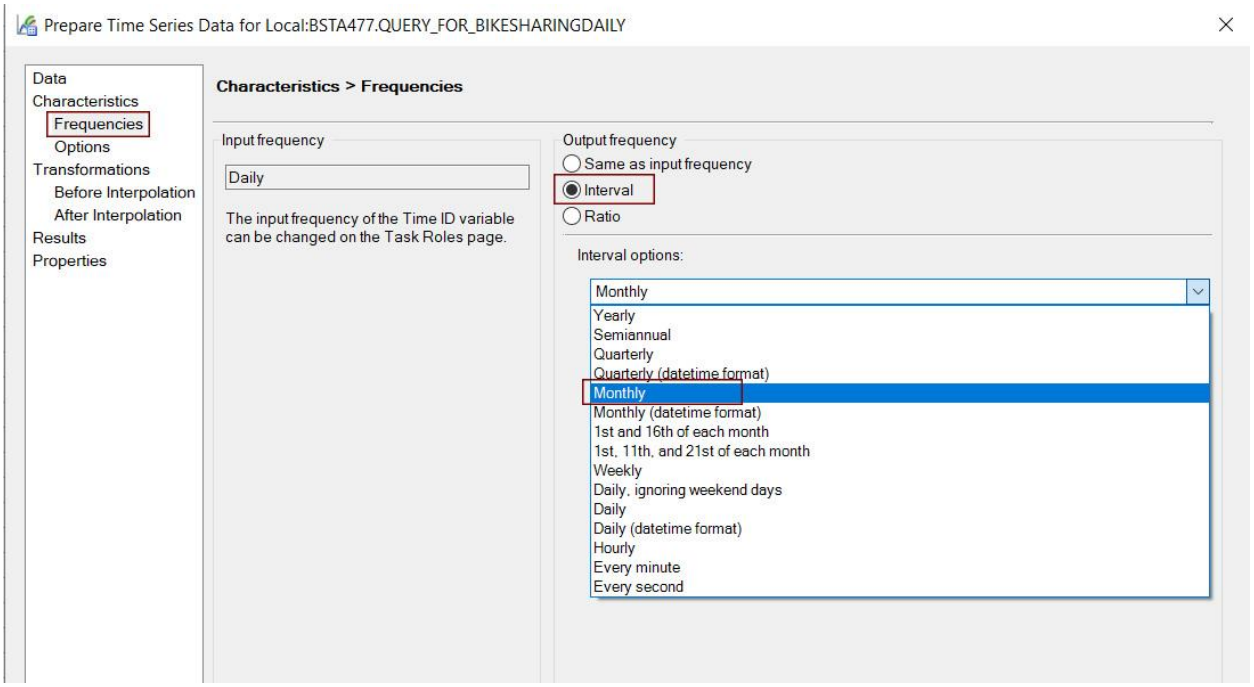
You can easily sum up daily data per month in SAS. It will add up rides per each day of the month and then will output a total number of rides per that month. Here is how to do it.

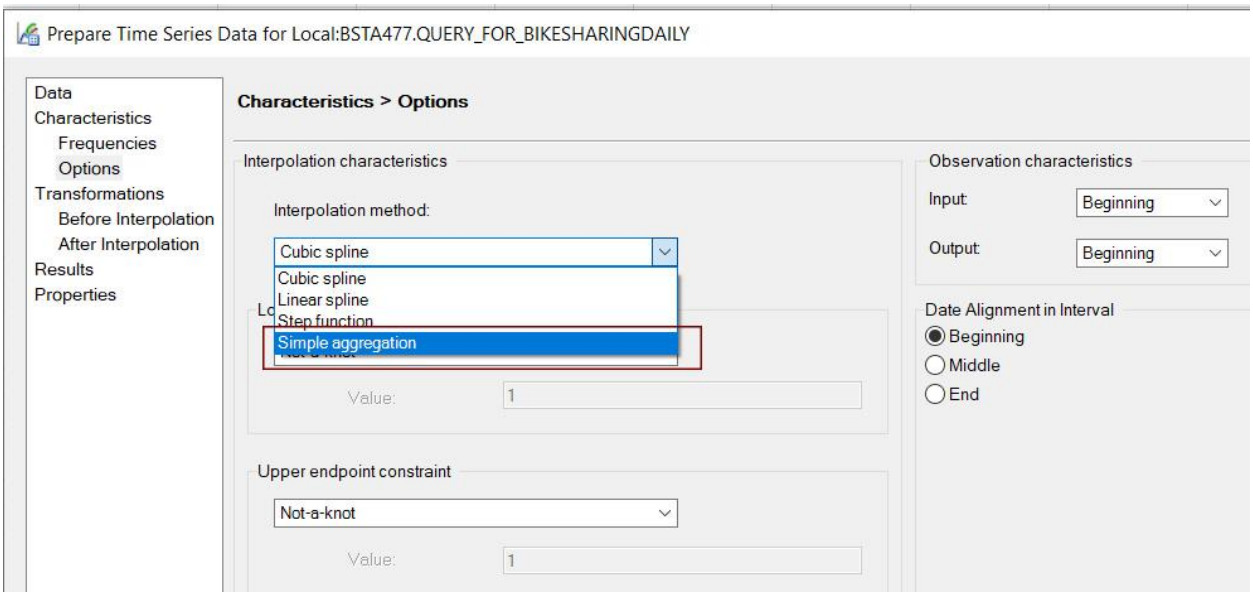
Open the file QUERY_FOR_BIKESHARING_DAILY → Analyze → Time Series → Prepare Time Series Data → Time ID Variable: dteday → Periodic data, Intervals: Daily → Drag the variable you are trying to predict (cnt – total count of member and non-member rides) → Frequencies → Interval: Monthly → Options: Simple Aggregation → Input: Total, Output: Total → Date Alignment in Interval: End → Results (on the left) → Create plots for input and output series → Run

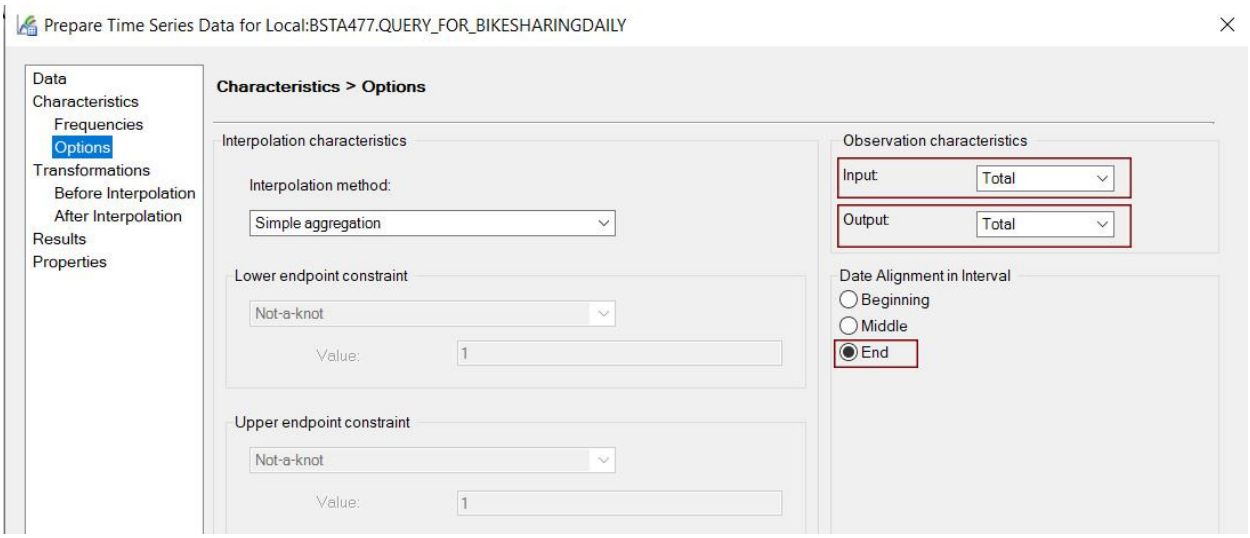

Data
Describe
Graph
Analyze
Export
Send To
ANOVA
Regression
Multivariate
Survival Analysis
Capability
Control Charts
Pareto Chart...
Time Series
Data Mining
Prepare Time Series Data...
Basic Forecasting...
ARIMA Modeling and Forecasting...
Regression Analysis with Autoregressive Errors...
Regression Analysis of Panel Data...
Create Time Series Data...
Forecast Studio Create Project...
Forecast Studio Open Project...
Forecast Studio Override Project...
yr
weekday
workingday
weathersit
temp

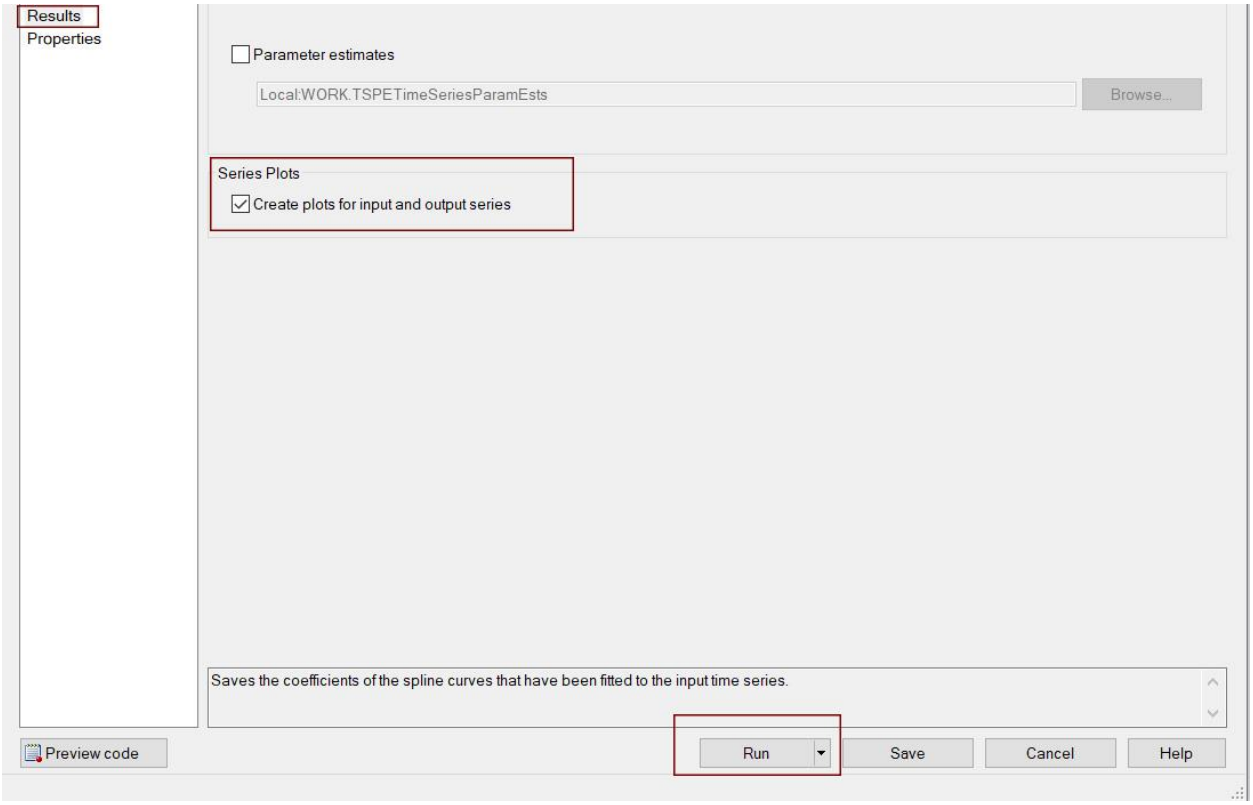

Prepare Time Series Data for Local:BSTA477.QUERY_FOR_BIKESHARINGDAILY
Data
Characteristics
Frequencies
Options
Transformations
Before Interpolation
After Interpolation
Results
Properties
Data source: Local:BSTA477.QUERY_FOR_BIKESHARINGDAILY
Task filter: None
Edit...
Variables to assign:
Name
NewTimeID
instant
dteday
season
yr
mnth
holiday
weekday
workingday
weathersit
temp
atemp
hum
windspeed
casual
registered
cnt
Log_casual
Task roles:
Time ID variable (Limit: 1)
dteday
Time series variables
Group analysis by
Use existing time ID variable
Existing time ID Data Type
Periodic data
Aperiodic data
Intervals:
Monthly
Yearly
Semiannual
Quarterly
Quarterly (datetime format)
Monthly
Monthly (datetime format)
1st and 16th of each month
1st, 11th, and 21st of each month
Weekly
Daily, ignoring weekend days
Daily
Daily (datetime format)
Hourly
Every minute
Every second


Variables to assign:
Name
NewTimeID
instant
dteday
season
yr
mnth
holiday
weekday
workingday
weathersit
temp
atemp
hum
windspeed
casual
registered
cnt
Log_casual
Task roles:
Time ID variable (Limit: 1)
dteday
Time series variables
cnt
Group analysis by
Replace existing variable
cnt


Prepare Time Series Data for Local:BSTA477.QUERY_FOR_BIKESHARINGDAILY
Data
Characteristics
Frequencies
Options
Transformations
Before Interpolation
After Interpolation
Results
Properties
Characteristics > Frequencies
Input frequency
Daily
The input frequency of the Time ID variable can be changed on the Task Roles page.
Output frequency
Same as input frequency
Interval
Ratio
Interval options:
Monthly
Yearly
Semiannual
Quarterly
Quarterly (datetime format)
Monthly
Monthly (datetime format)
1st and 16th of each month
1st, 11th, and 21st of each month
Weekly
Daily, ignoring weekend days
Daily
Daily (datetime format)
Hourly
Every minute
Every second


Prepare Time Series Data for Local:BSTA477.QUERY_FOR_BIKESHARINGDAILY
Data
Characteristics
Frequencies
Options
Transformations
Before Interpolation
After Interpolation
Results
Properties
Characteristics > Options
Interpolation characteristics
Interpolation method:
Cubic spline
Cubic spline
Linear spline
Step function
Simple aggregation
Value:
1
Upper endpoint constraint
Not-a-knot
Value:
1
Observation characteristics
Input:
Beginning
Output:
Beginning
Date Alignment in Interval
Beginning
Middle
End

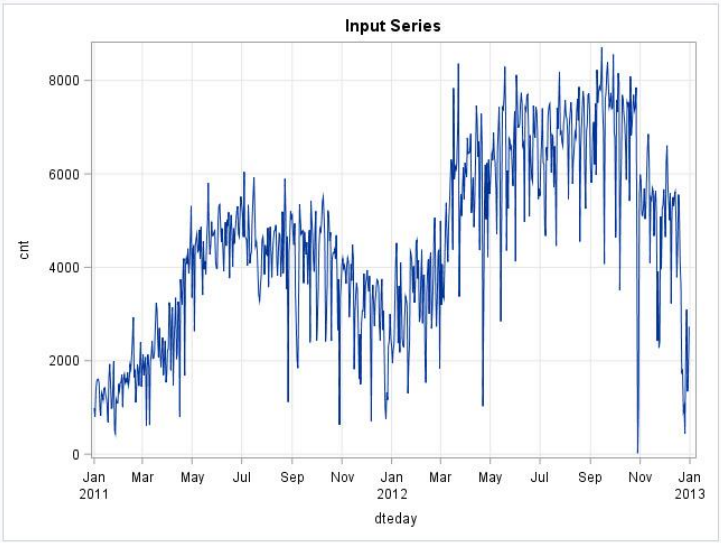

Prepare Time Series Data for Local:BSTA477.QUERY_FOR_BIKESHARINGDAILY
Data
Characteristics
Frequencies
Options
Transformations
Before Interpolation
After Interpolation
Results
Properties
Characteristics > Options
Interpolation characteristics
Interpolation method:
Simple aggregation
Lower endpoint constraint
Not-a-knot
Value:
1
Upper endpoint constraint
Not-a-knot
Value:
1
Observation characteristics
Input:
Total
Output:
Total
Date Alignment in Interval
Beginning
Middle
End

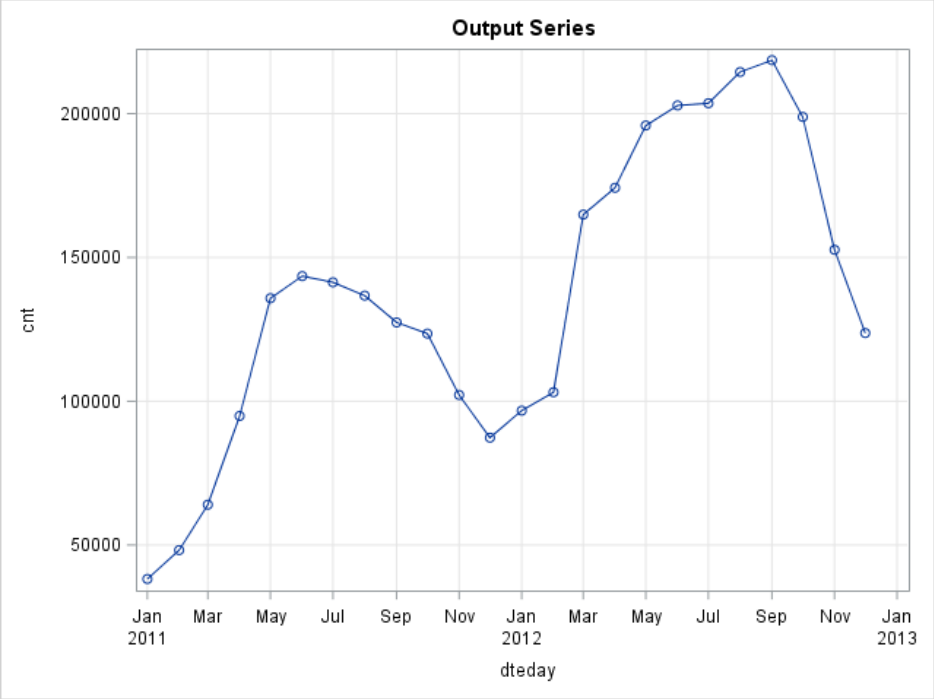


The EXPAND Procedure

Input Data Set	
Name	WORK.TMP0TEMPTABLEINPUT
Label	
Time ID Variable Name	dteday

Variable Information	
Input Variable Name	cnt
Number of Observations Read	731
Output Variable Name	cnt
Number of Observations Written	24





Prepare Time Series Data

Input Data | Code | Log | Output Data | Results

Modify Task | Filter and Sort | Query Builder | Where

	dteday	cnt
1	JAN2011	38E3
2	FEB2011	48E3
3	MAR2011	64E3
4	APR2011	95E3
5	MAY2011	14E4
6	JUN2011	14E4
7	JUL2011	14E4
8	AUG2011	14E4
9	SEP2011	13E4
10	OCT2011	12E4
11	NOV2011	1E5
12	DEC2011	87E3
13	JAN2012	97E3
14	FEB2012	1E5
15	MAR2012	16E4
16	APR2012	17E4
17	MAY2012	2E5
18	JUN2012	2E5
19	JUL2012	2E5
20	AUG2012	21E4
21	SEP2012	22E4
22	OCT2012	2E5
23	NOV2012	15E4
24	DEC2012	12E4

Choosing the best model

When we forecast, it is important to see whether we have a good model and also how well the model forecasts. For that we need to split the dataset we have into two part – the train and the validation datasets and then we need to estimate the model.

A. Do we have a good model?

To estimate that we need to compare the train and validation error terms, such as MAD, MAPE, MPE, MSE, RMSE. Their values should be as close as possible. We do this for each model we build. If the values are too far from each other, i.e. the difference is 10% or more, then we estimate that the model overfits the data and is not going to generalize well. These models will be excluded from the further analysis.

B. How well does our model forecast?

After estimating whether we have a good model, we then compare this model to others. Say, we will compare an exponential smoothing, Winter's and Holt's models to each other. How do we compare them? By comparing their VALIDATION error terms such as MAD, MAPE, MPE, MSE, RMSE. The model that has the lowest values is the best model and makes the most accurate forecasts.

What do we assess?	What do we compare	What to look for?
Goodness of fit	Training and validation error terms: MAD, MAPE, MPE, MSE, RMSE	A good model will have similar values of errors for both train and validation datasets
Forecasting power	A few models. Validation error terms: MAD, MAPE, MPE, MSE, RMSE	The best model will be the one with the lowest error terms

How to forecast in SAS

In SAS for exponential smoothing, Holt's and Winter's models, you will have to do the following:

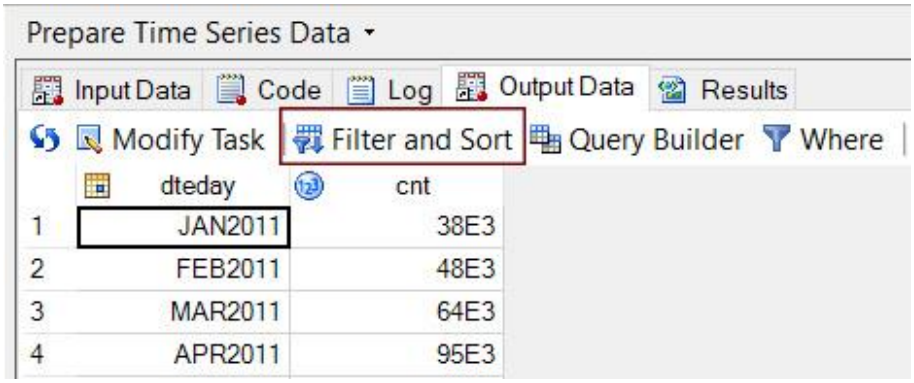
1. Create a train dataset

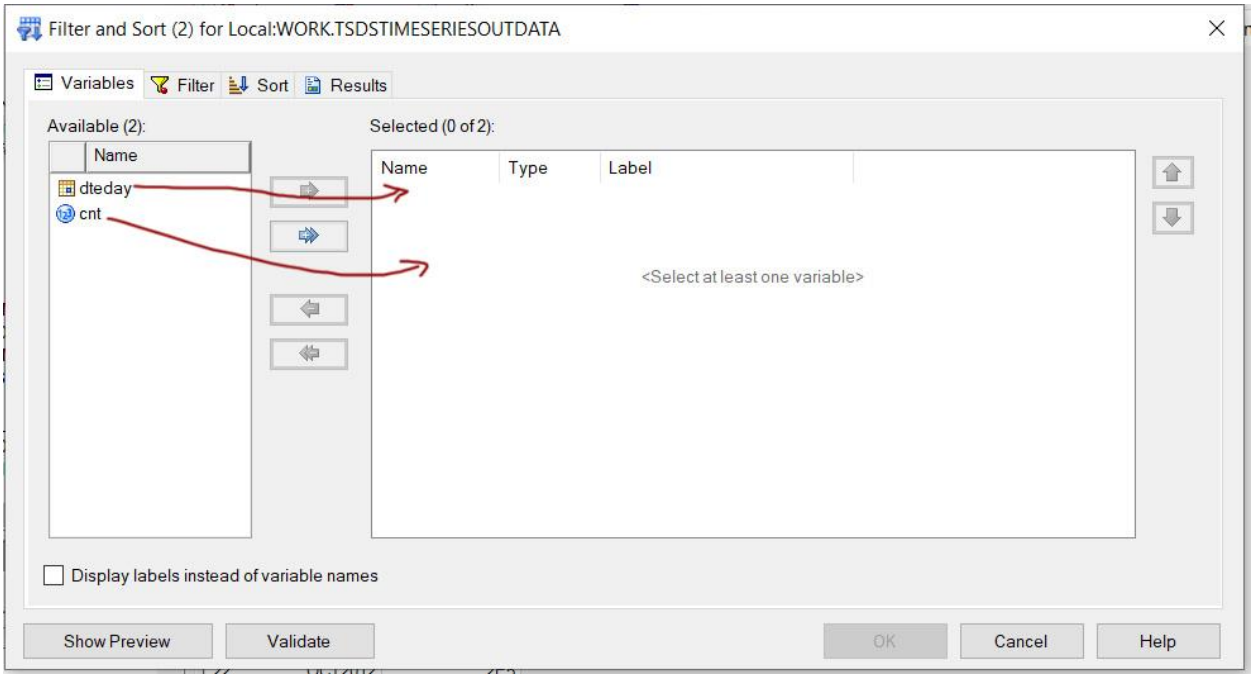
2. Run a forecasting model and get forecasts as well as the train errors

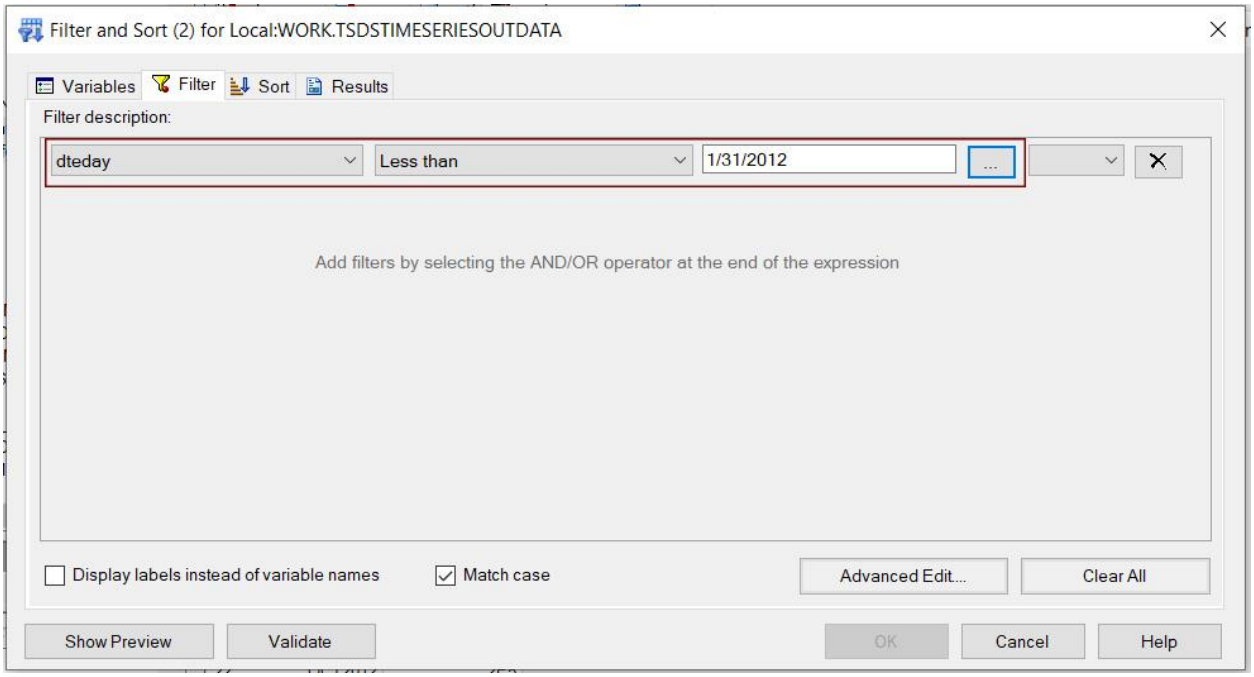
3. Export the created forecast into Excel and calculate errors for the validation part of the dataset.

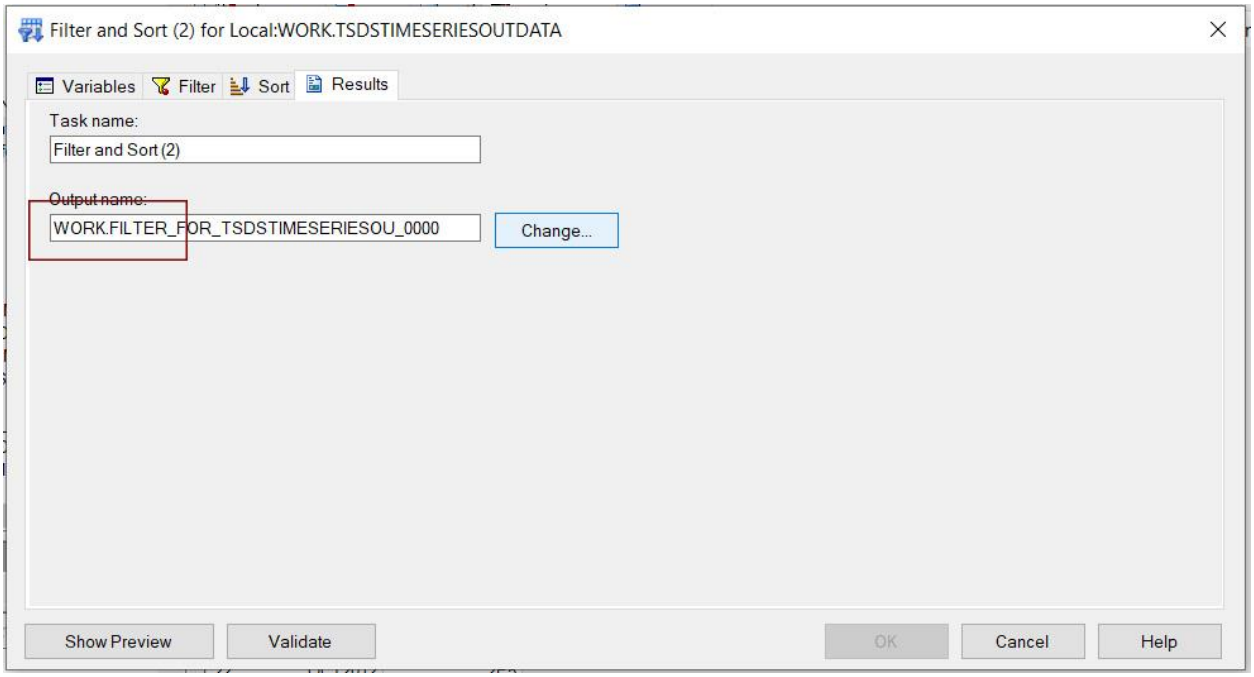
Creating a training dataset

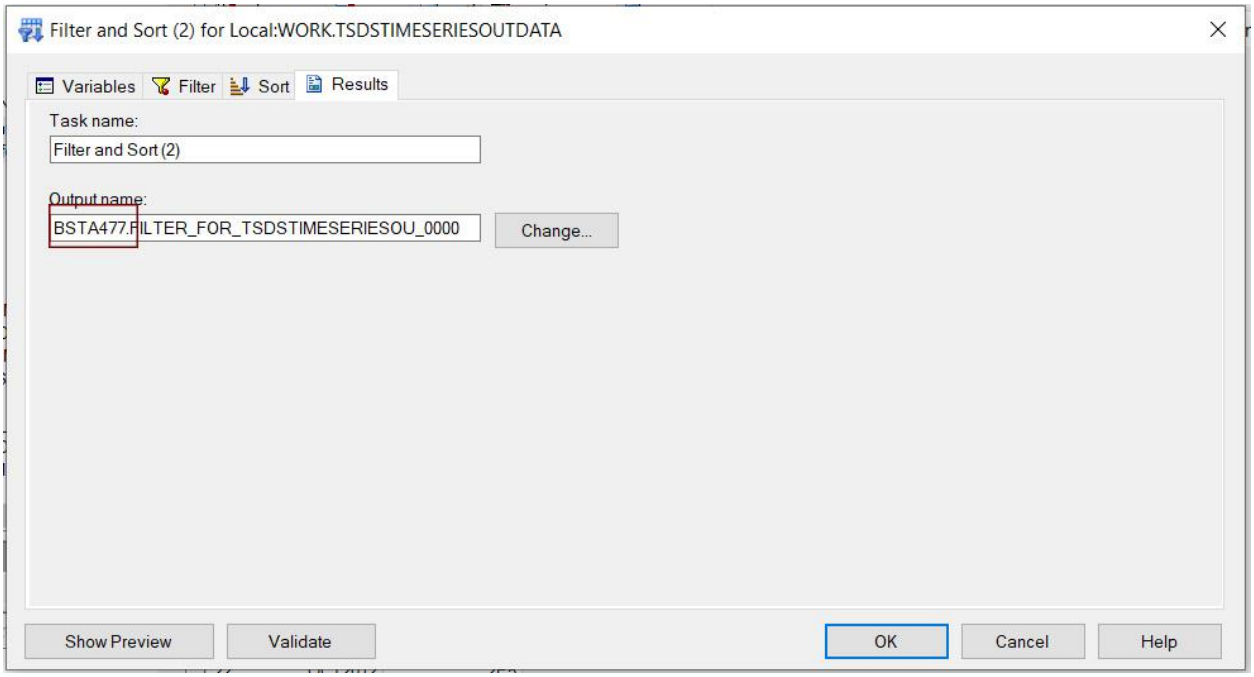
Open the dataset → Filter and Sort → Variables: Drag the variables you will need for the forecast to the right (date and the variable you want to forecast) → Filter: Set the filtering values (dteday; less then; December 31, 2012) → Results: Permanently save the dataset: Replace the WORK library with the one you ran at the beginning of the session – BSTA477 → OK






Filter and Sort (2) for Local:WORK.TSDSTIMESERIESOUTDATA
Variables
Filter
Sort
Results
Filter description:
dteday
Less than
1/31/2012
...
Add filters by selecting the AND/OR operator at the end of the expression
Display labels instead of variable names
Match case
Advanced Edit...
Clear All
Show Preview
Validate
OK
Cancel
Help

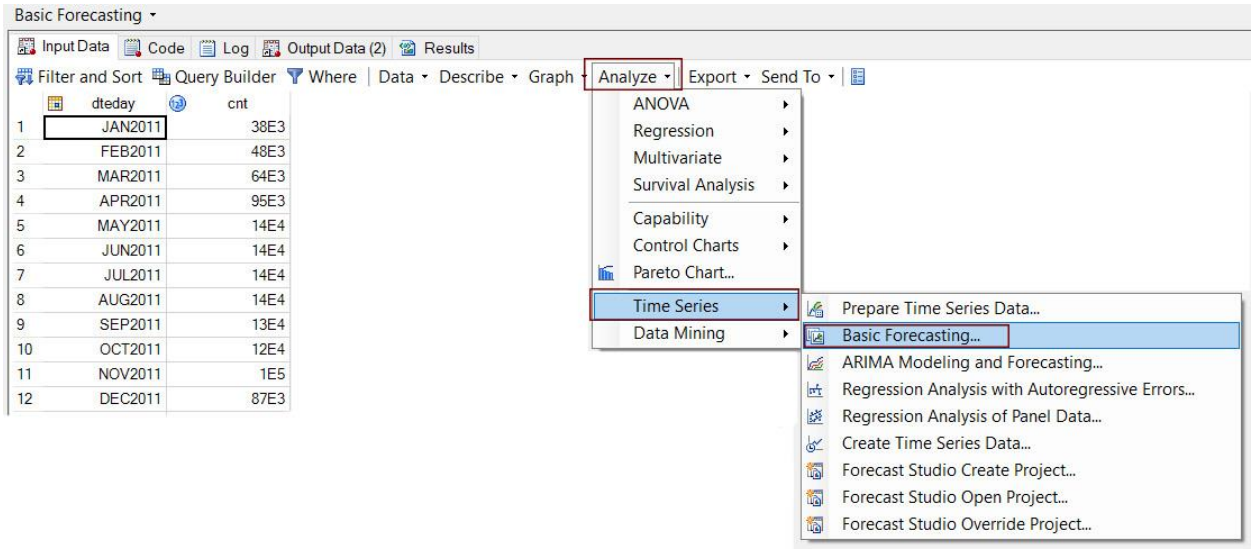

Filter and Sort (2) for Local:WORK.TSDSTIMESERIESOUTDATA
Variables
Filter
Sort
Results
Task name:
Filter and Sort (2)
Output name:
WORK.FILTER_FOR_TSDSTIMESERIESOU_0000
Change...
Show Preview
Validate
OK
Cancel
Help

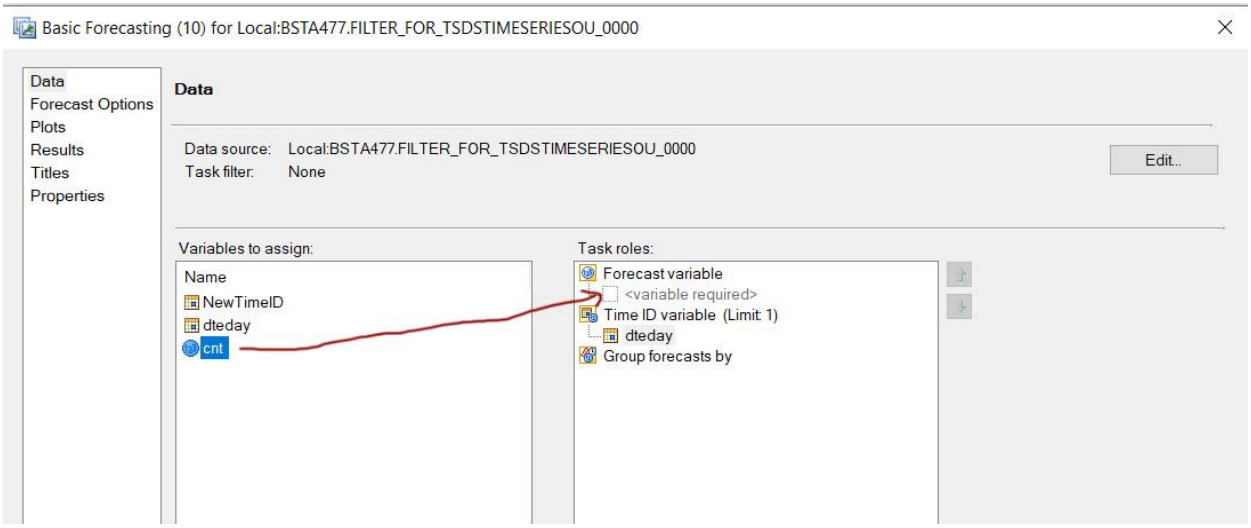


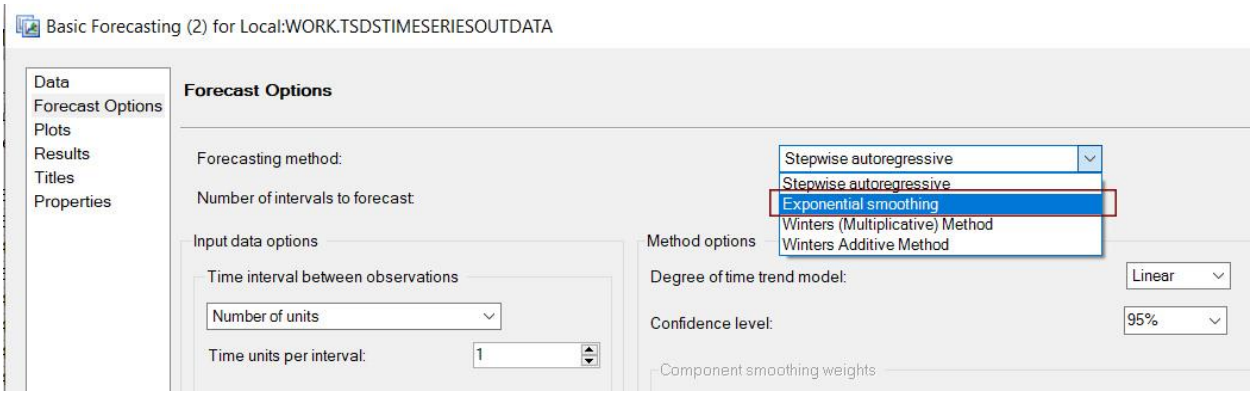
Exponential Smoothing

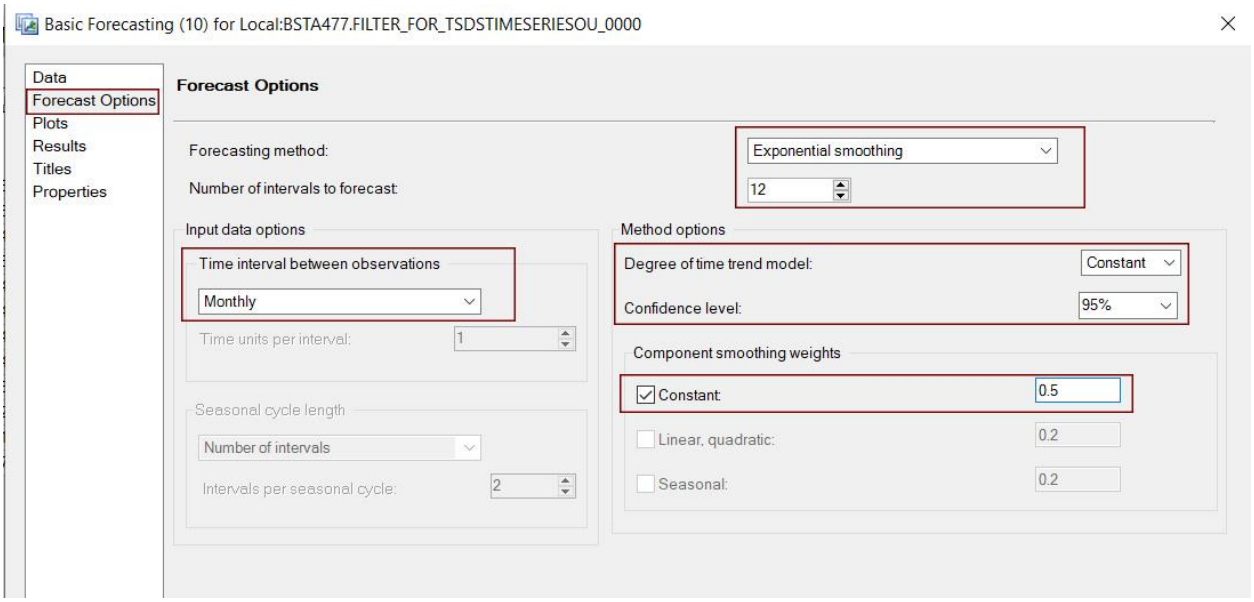
We are going to use the newly created **TRAIN dataset** with accumulated values of total member and non-member rides on a monthly basis. As we can see from the table above, we now only have 12 points of data. Let's see how we can use exponential smoothing to create a forecast. We are going to do one forecast by using the exponential smoothing constant (alpha) = 0.5. We can also change the alpha values to 0.2 or 0.3.

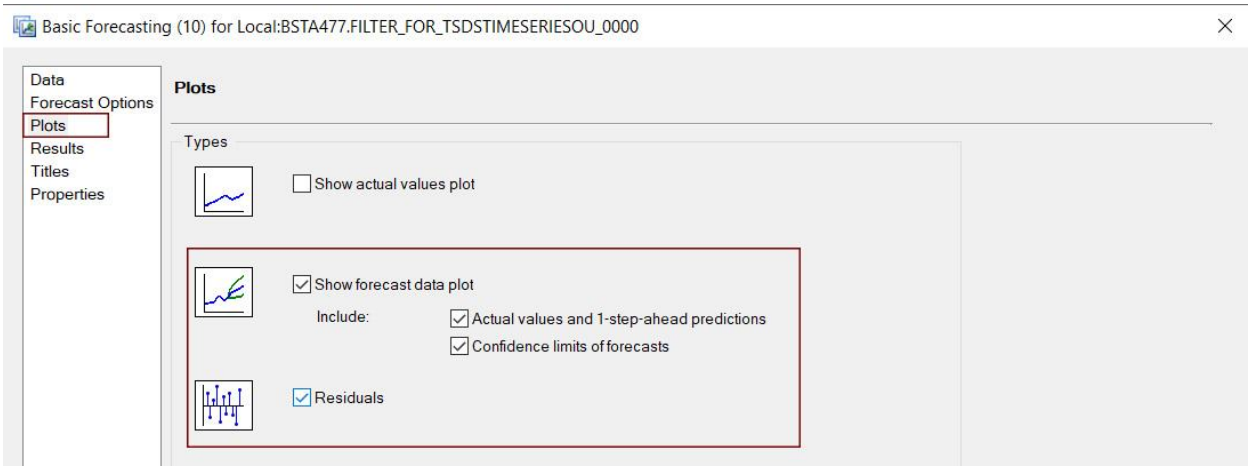
Open the file you want to use for forecasting → Analyze → Time Series → Basic Forecasting → Data: Drag the variable for forecast → Forecast options: Exponential Smoothing (set the number of points to forecast (right underneath the model to the number of periods for forecast; we are going to forecast a year of rides, so we will be forecasting 12 months) → Method options: Degree of time trend model: Constant → Confidence level: 95% → Time interval between observations: Monthly → Constant: 0.5 → Plots: Show forecast data plot, Residuals → Results: Parameter estimates, R-squared-type forecast accuracy → Run → Output data → Open the needed dataset → Export → Export (the file name) → Place the file into a folder on your computer, give it a name, select the needed extension (xlsx) → Repeat for the second file

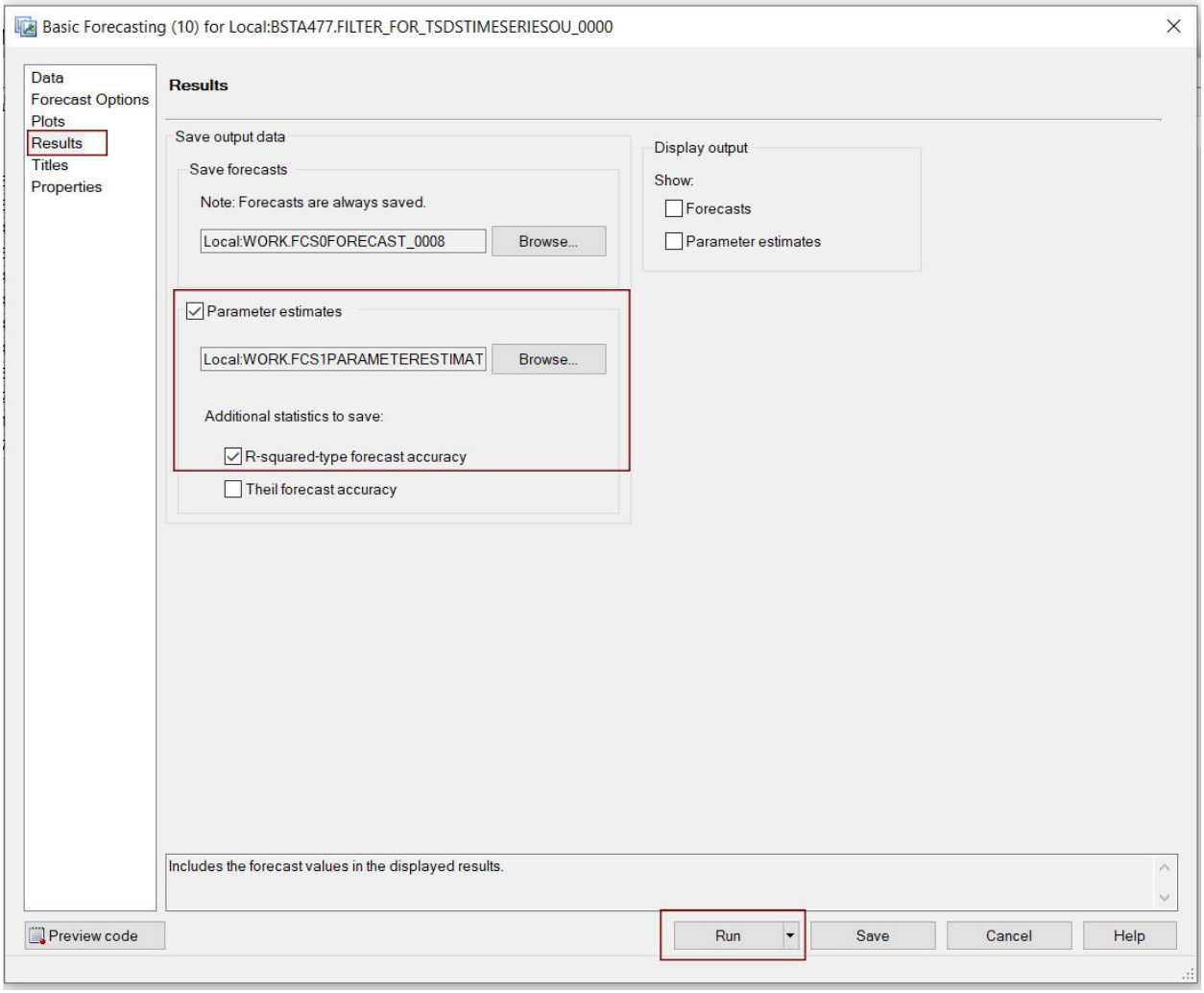

Basic Forecasting
Input Data
Code
Log
Output Data (2)
Results
Filter and Sort
Query Builder
Where
Data
Describe
Graph
Analyze
Export
Send To
dteday
cnt
1 JAN2011 38E3
2 FEB2011 48E3
3 MAR2011 64E3
4 APR2011 95E3
5 MAY2011 14E4
6 JUN2011 14E4
7 JUL2011 14E4
8 AUG2011 14E4
9 SEP2011 13E4
10 OCT2011 12E4
11 NOV2011 1E5
12 DEC2011 87E3
ANOVA
Regression
Multivariate
Survival Analysis
Capability
Control Charts
Pareto Chart...
Time Series
Data Mining
Prepare Time Series Data...
Basic Forecasting...
ARIMA Modeling and Forecasting...
Regression Analysis with Autoregressive Errors...
Regression Analysis of Panel Data...
Create Time Series Data...
Forecast Studio Create Project...
Forecast Studio Open Project...
Forecast Studio Override Project...


Basic Forecasting (10) for Local:BSTA477.FILTER_FOR_TSDSTIMESERIESOU_0000
Data
Forecast Options
Plots
Results
Titles
Properties
Data
Data source: Local:BSTA477.FILTER_FOR_TSDSTIMESERIESOU_0000
Task filter: None
Edit...
Variables to assign:
Name
NewTimeID
dteday
cnt
Task roles:
Forecast variable
<variable required>
Time ID variable (Limit: 1)
dteday
Group forecasts by

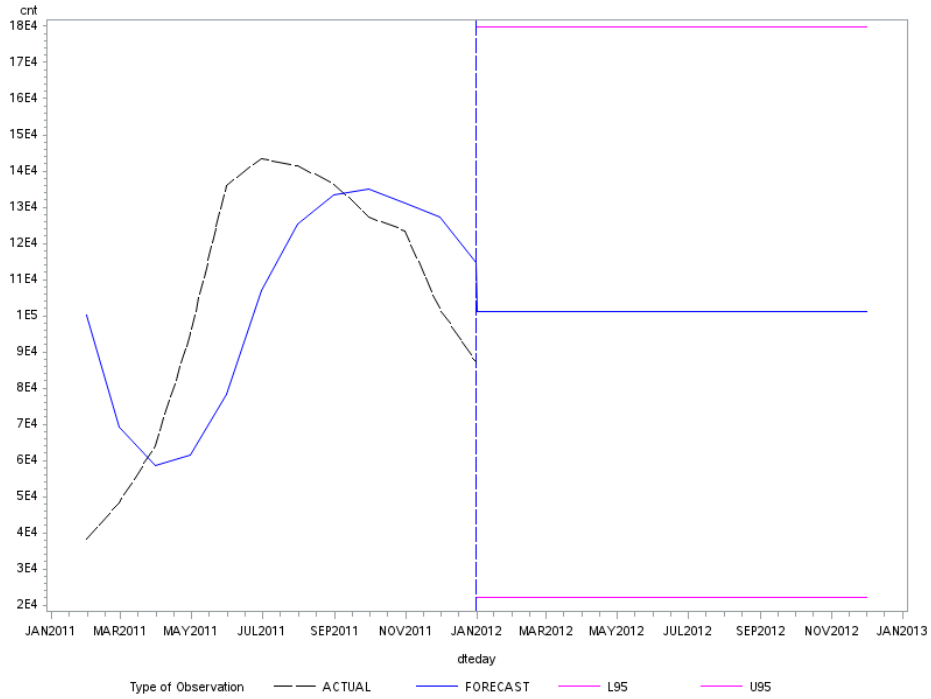

Basic Forecasting (2) for Local:WORK.TSDSTIMESERIESOUTDATA
Data
Forecast Options
Plots
Results
Titles
Properties
Forecast Options
Forecasting method:
Stepwise autoregressive
Stepwise autoregressive
Exponential smoothing
Winters (Multiplicative) Method
Winters Additive Method
Number of intervals to forecast:
Input data options
Time interval between observations
Number of units
Time units per interval:
1
Method options
Degree of time trend model:
Linear
Confidence level:
95%
Component smoothing weights


Basic Forecasting (10) for Local:BSTA477.FILTER_FOR_TSDSTIMESERIESOU_0000
Data
Forecast Options
Plots
Results
Titles
Properties
Forecast Options
Forecasting method:
Exponential smoothing
Number of intervals to forecast:
12
Input data options
Time interval between observations
Monthly
Time units per interval:
1
Seasonal cycle length
Number of intervals
Intervals per seasonal cycle:
2
Method options
Degree of time trend model:
Constant
Confidence level:
95%
Component smoothing weights
Constant:
0.5
Linear, quadratic:
0.2
Seasonal:
0.2

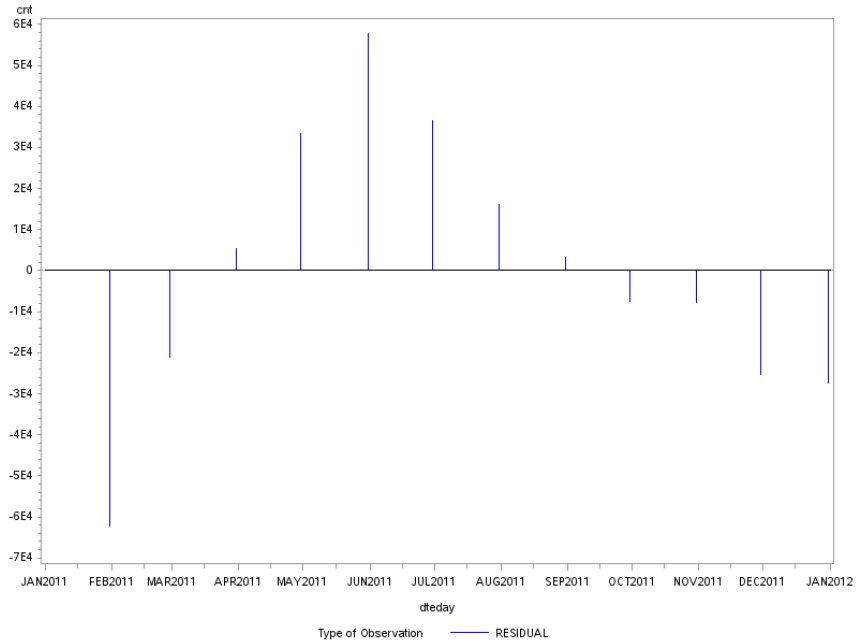

Basic Forecasting (10) for Local:BSTA477.FILTER_FOR_TSDSTIMESERIESOU_0000
Data
Forecast Options
Plots
Results
Titles
Properties
Plots
Types
Show actual values plot
Show forecast data plot
Include:
Actual values and 1-step-ahead predictions
Confidence limits of forecasts
Residuals

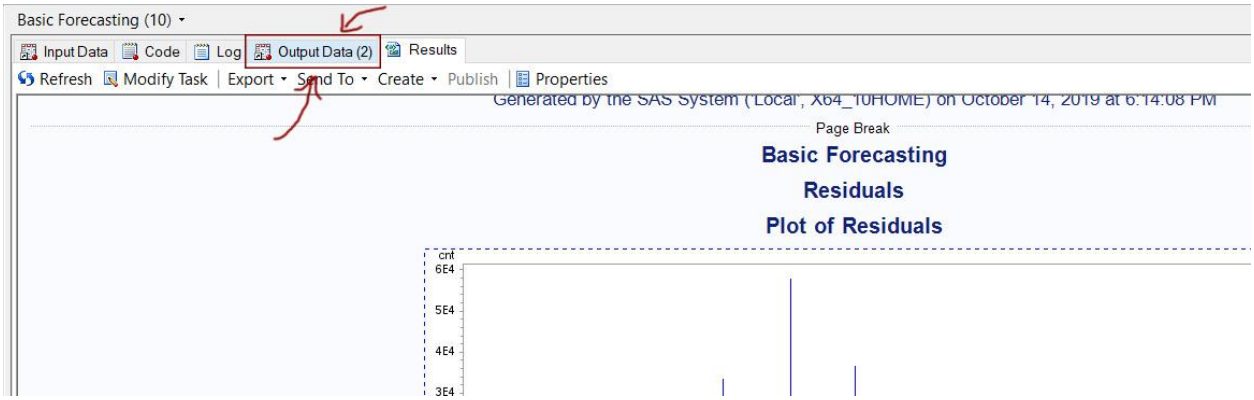

Basic Forecasting (10) for Local:BSTA477.FILTER_FOR_TSDSTIMESERIESOU_0000
Data
Forecast Options
Plots
Results
Titles
Properties
Results
Save output data
Save forecasts
Note: Forecasts are always saved.
Local:WORK.FCS0FORECAST_0008
Browse...
Parameter estimates
Local:WORK.FCS1PARAMETERESTIMAT
Browse...
Additional statistics to save:
R-squared-type forecast accuracy
Theil forecast accuracy
Display output
Show:
Forecasts
Parameter estimates
Includes the forecast values in the displayed results.
Preview code
Run
Save
Cancel
Help

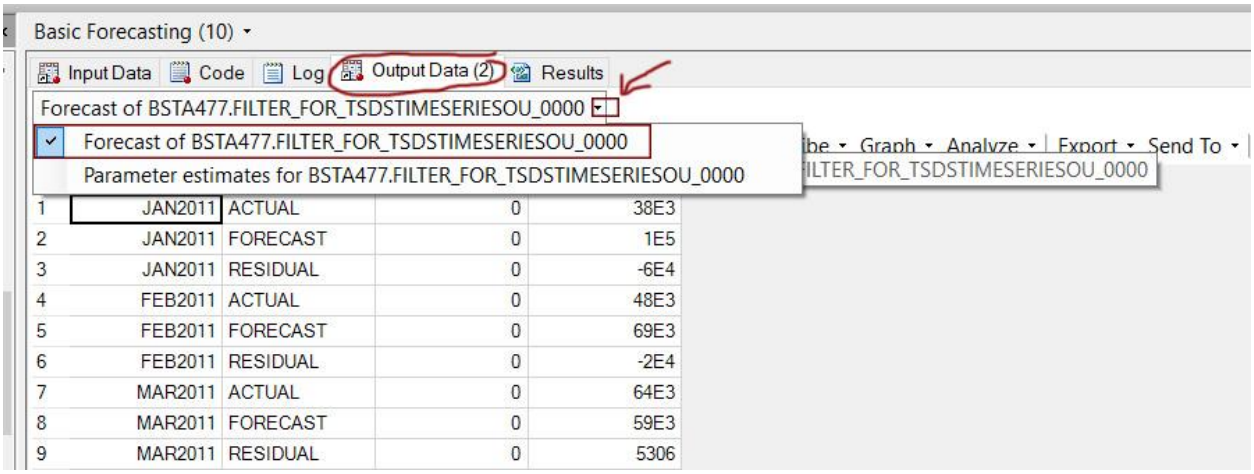
The plot

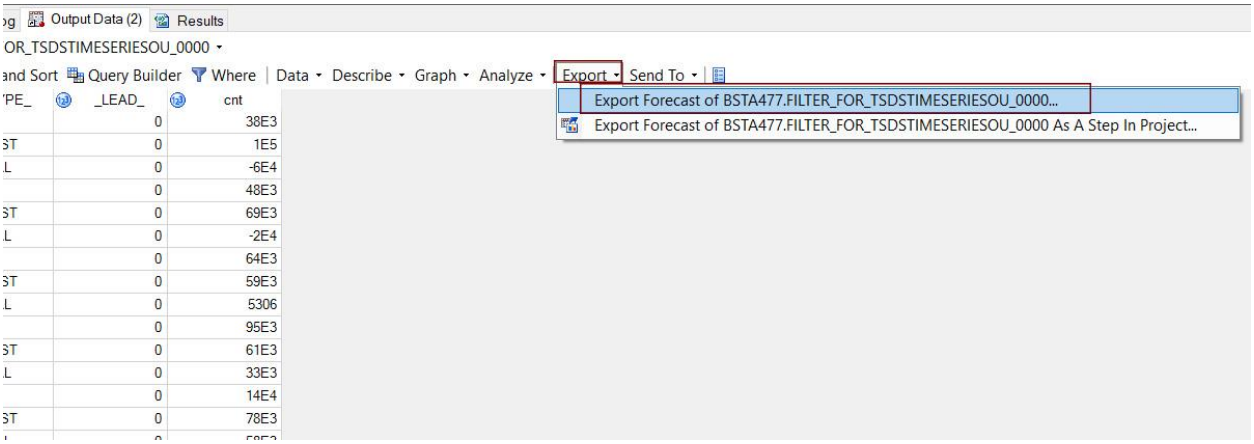


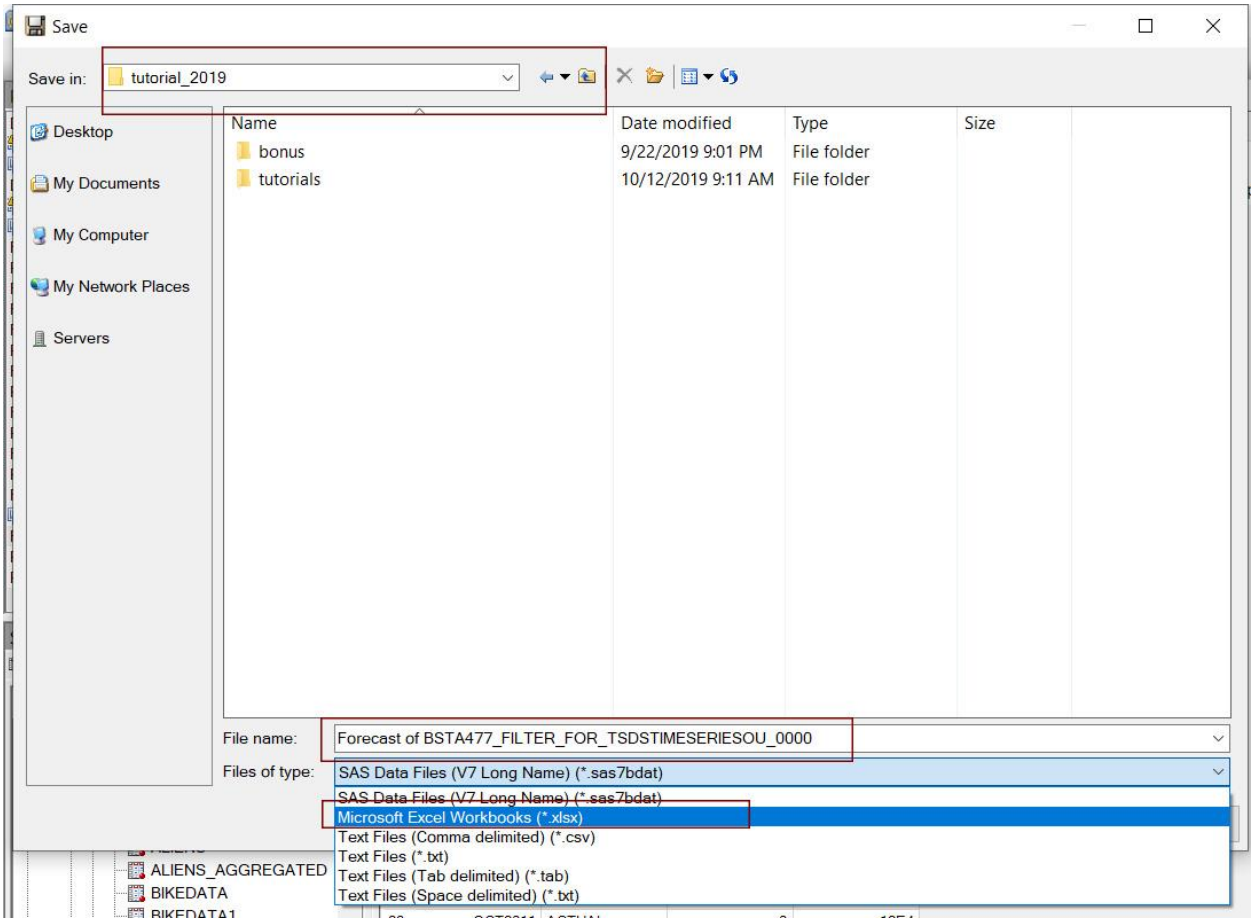
Residuals




Basic Forecasting (10)
Input Data
Code
Log
Output Data (2)
Results
Refresh
Modify Task
Export
Send To
Create
Publish
Properties
Page Break
Basic Forecasting
Residuals
Plot of Residuals
cnt
6E4
5E4
4E4
3E4


Basic Forecasting (10)
Input Data
Code
Log
Output Data (2)
Results
Forecast of BSTA477.FILTER_FOR_TSDSTIMESERIESOU_0000
Forecast of BSTA477.FILTER_FOR_TSDSTIMESERIESOU_0000
Parameter estimates for BSTA477.FILTER_FOR_TSDSTIMESERIESOU_0000
Graph
Analyze
Export
Send To
1 JAN2011 ACTUAL 0 38E3
2 JAN2011 FORECAST 0 1E5
3 JAN2011 RESIDUAL 0 -6E4
4 FEB2011 ACTUAL 0 48E3
5 FEB2011 FORECAST 0 69E3
6 FEB2011 RESIDUAL 0 -2E4
7 MAR2011 ACTUAL 0 64E3
8 MAR2011 FORECAST 0 59E3
9 MAR2011 RESIDUAL 0 5306


Output Data (2)
Results
Query Builder
Where
Data
Describe
Graph
Analyze
Export
Send To
LEAD
cnt
Export Forecast of BSTA477.FILTER_FOR_TSDSTIMESERIESOU_0000...
Export Forecast of BSTA477.FILTER_FOR_TSDSTIMESERIESOU_0000 As A Step In Project...



Now, as you have forecasts for the 2012, you can use the data from the full file (the one that has both years) and you can use the actual data and the forecasts to calculate the error terms.

Holt's (with a trend) and Winters (with a trend and seasonality)

Holt's and Winter's are very similar. Holt's includes a trend while Winter's, in addition to the trend, also includes seasonality. Depending on the trend, you need to choose either the multiplicative or the additive model. The model can be either linear, quadratic or constant. In SAS you can set the values of:

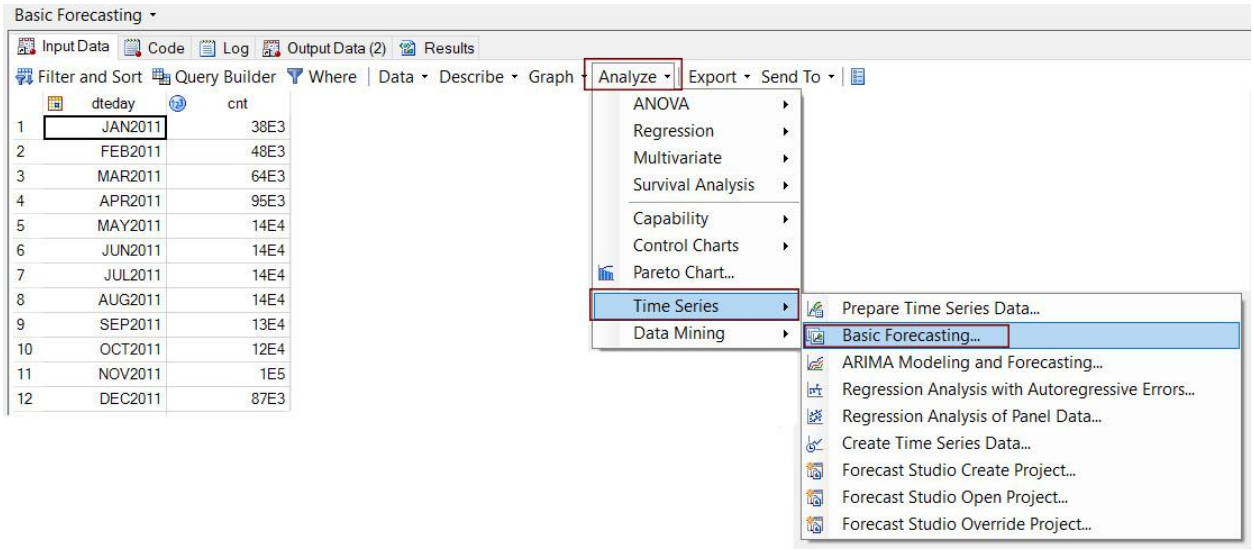
- **Alpha** – the coefficient for the level smoothing;
- **Beta** – the coefficient for the trend smoothing;
- **Gamma** – the coefficient for the seasonal smoothing.

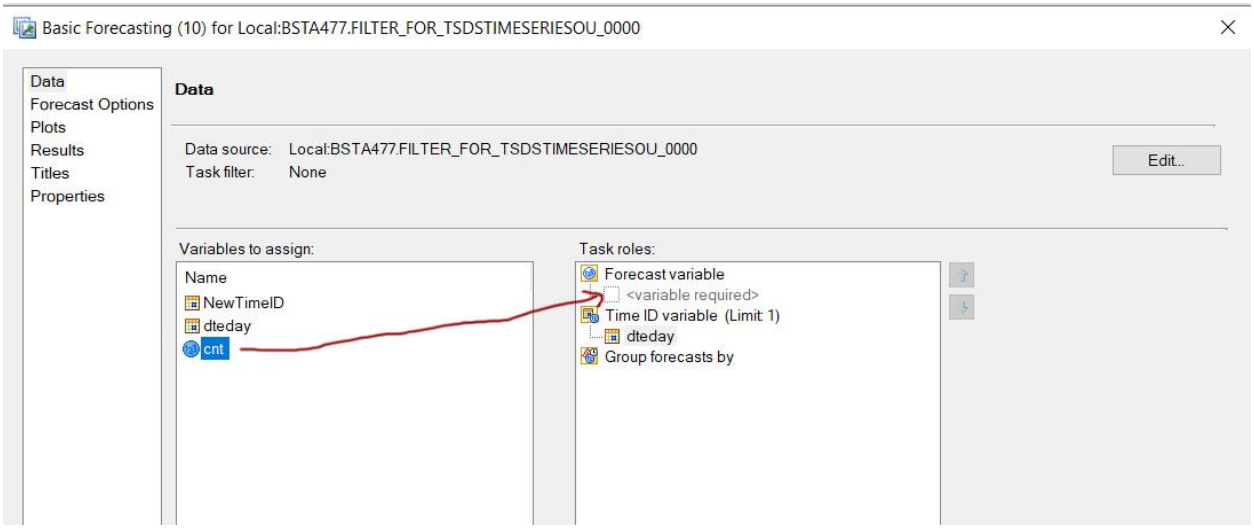
For Holt's you will only set alpha and beta coefficients while Winter's – all three. In SAS you have to make a decision on which values to set. If you want to get the best values, you can run the Excel macro and then use those values in SAS.

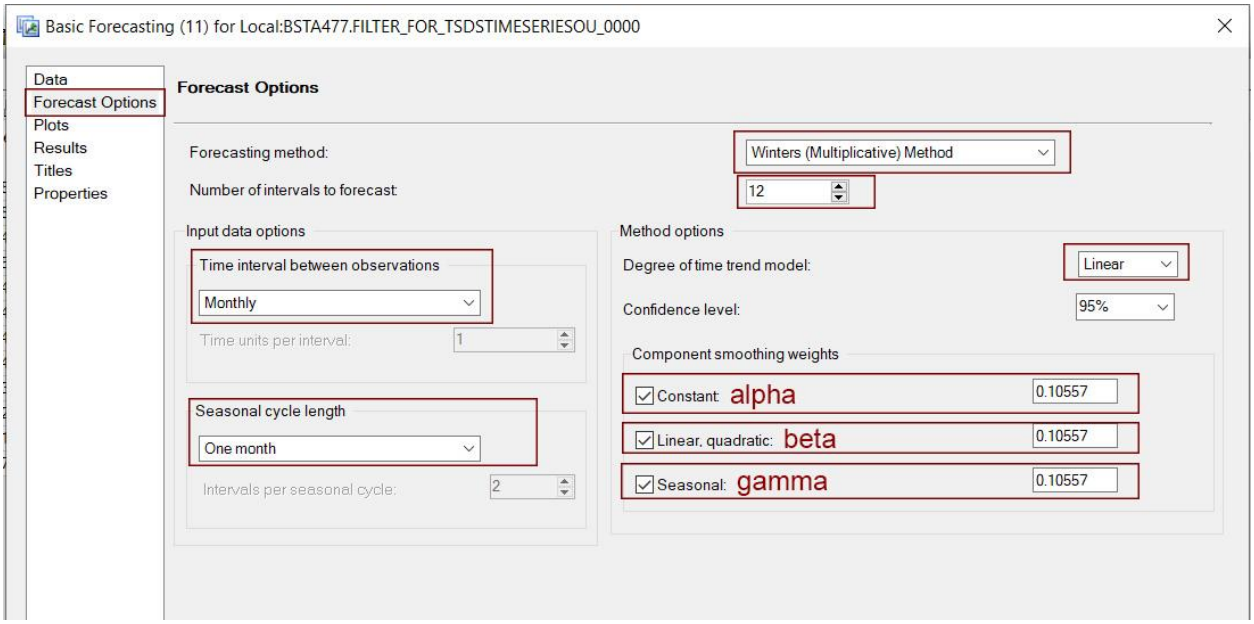
A. Winter's (Multiplicative)

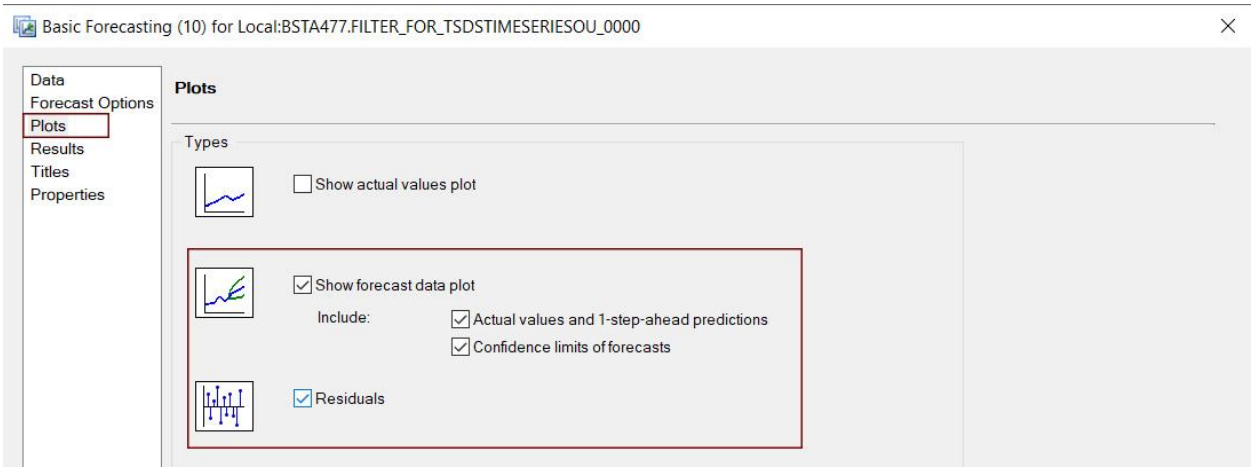
Open the file you want to forecast (Train dataset) → Analyze → Time Series → Basic forecasting → Data: Drag the variable you want to analyze to the right → Forecast Options: Winter's (Multiplicative) Method, 12 right below → Degree of time trend model (linear trend) → Time interval between observations: Monthly → Seasonal cycle length: Monthly (in this case) → Component smoothing weights: set the alpha, beta and gamma values (either keep at default, run a macro or just play around with several values) → Plots: Show forecast data plot, Residuals → Results: Parameter estimates, R-squared-type forecast accuracy → Run → Output data → Open the needed dataset → Export → Export (the file name) → Place the file into a folder on your computer, give it a name, select the needed extension (xlsx) → Repeat for the second file

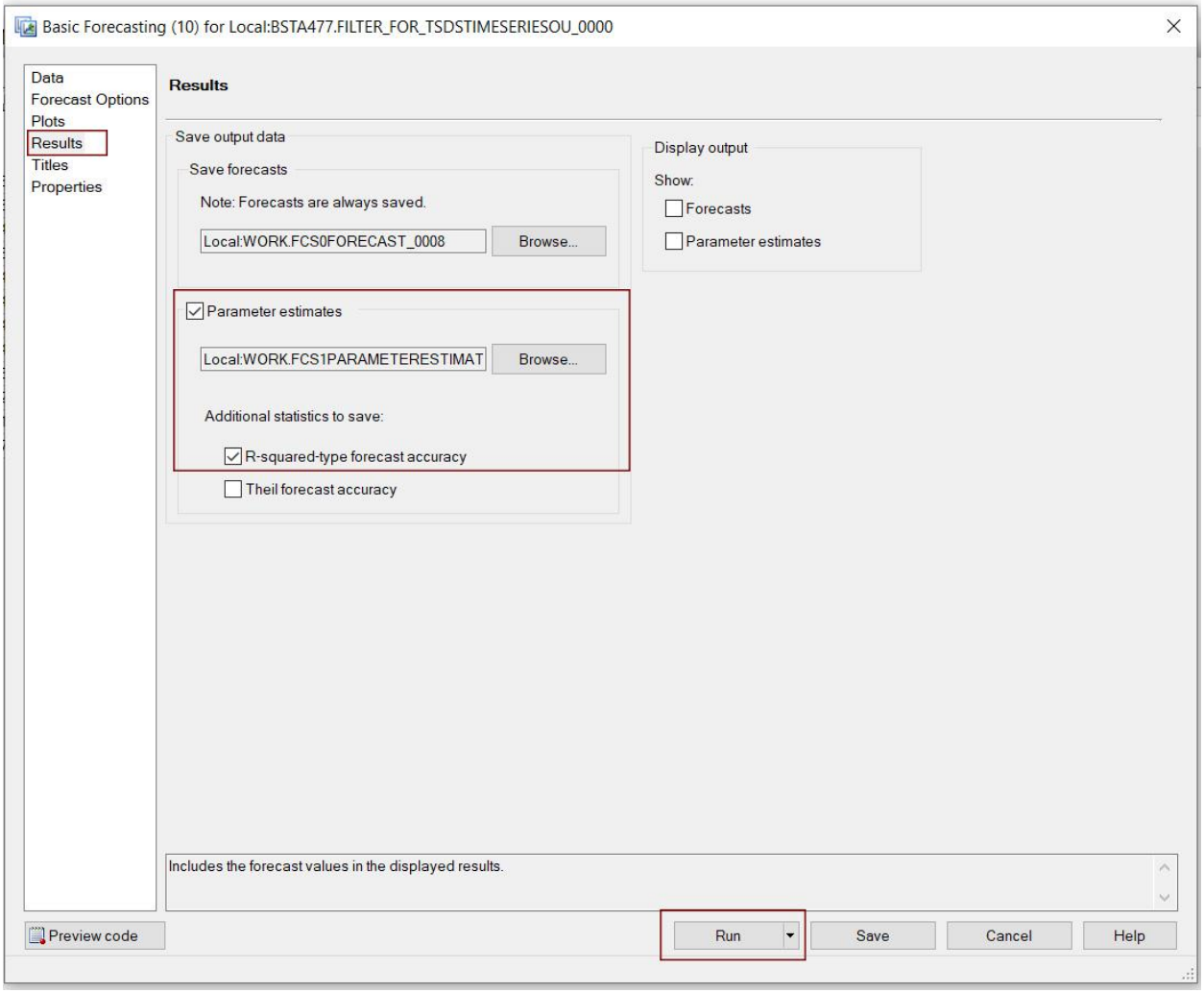
To get Holt's, just don't check the gamma value (seasonal).

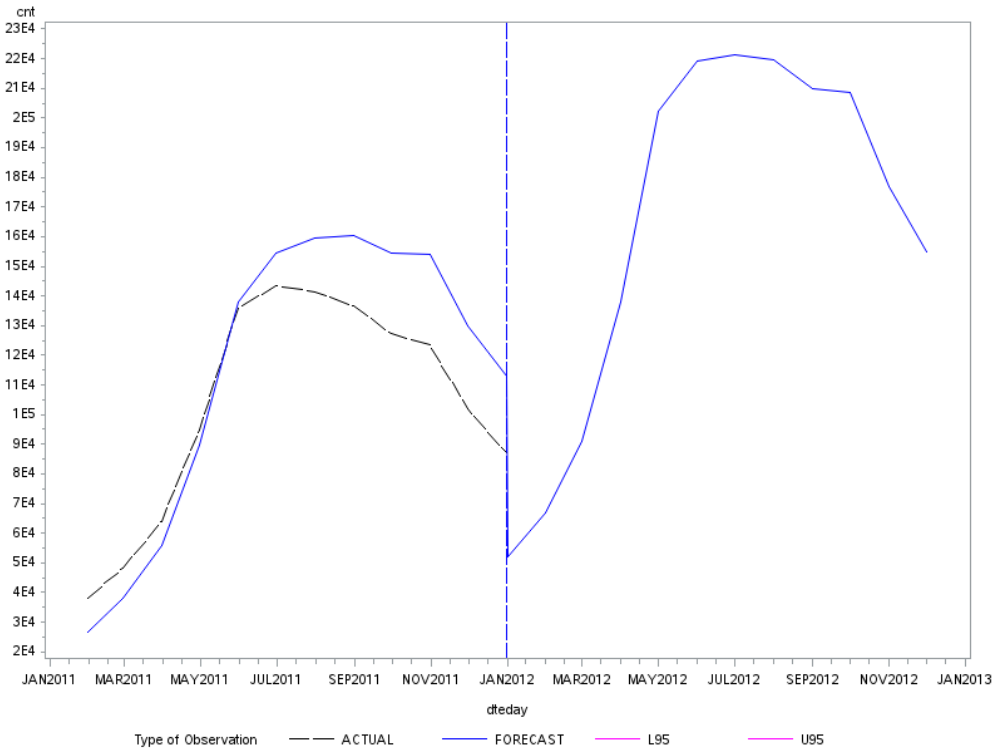

Basic Forecasting
Input Data
Code
Log
Output Data (2)
Results
Filter and Sort
Query Builder
Where
Data
Describe
Graph
Analyze
Export
Send To
dteday
cnt
1 JAN2011 38E3
2 FEB2011 48E3
3 MAR2011 64E3
4 APR2011 95E3
5 MAY2011 14E4
6 JUN2011 14E4
7 JUL2011 14E4
8 AUG2011 14E4
9 SEP2011 13E4
10 OCT2011 12E4
11 NOV2011 1E5
12 DEC2011 87E3
ANOVA
Regression
Multivariate
Survival Analysis
Capability
Control Charts
Pareto Chart...
Time Series
Data Mining
Prepare Time Series Data...
Basic Forecasting...
ARIMA Modeling and Forecasting...
Regression Analysis with Autoregressive Errors...
Regression Analysis of Panel Data...
Create Time Series Data...
Forecast Studio Create Project...
Forecast Studio Open Project...
Forecast Studio Override Project...

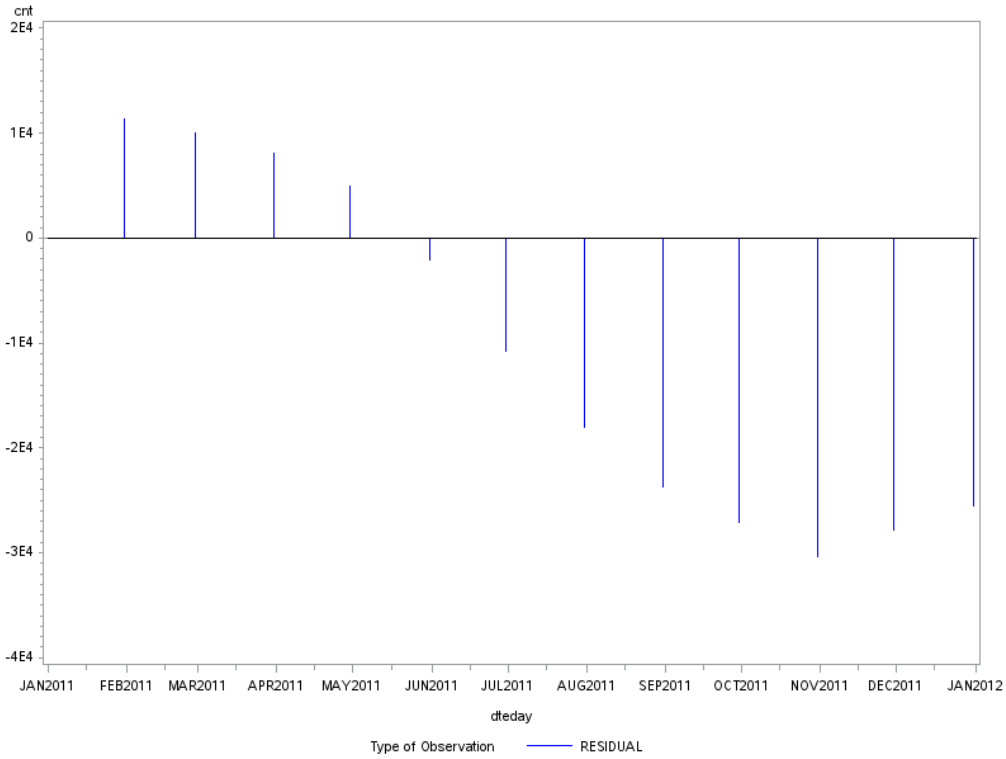

Basic Forecasting (10) for Local:BSTA477.FILTER_FOR_TSDSTIMESERIESOU_0000
Data
Forecast Options
Plots
Results
Titles
Properties
Data
Data source: Local:BSTA477.FILTER_FOR_TSDSTIMESERIESOU_0000
Task filter: None
Edit...
Variables to assign:
Name
NewTimeID
dteday
cnt
Task roles:
Forecast variable
<variable required>
Time ID variable (Limit: 1)
dteday
Group forecasts by

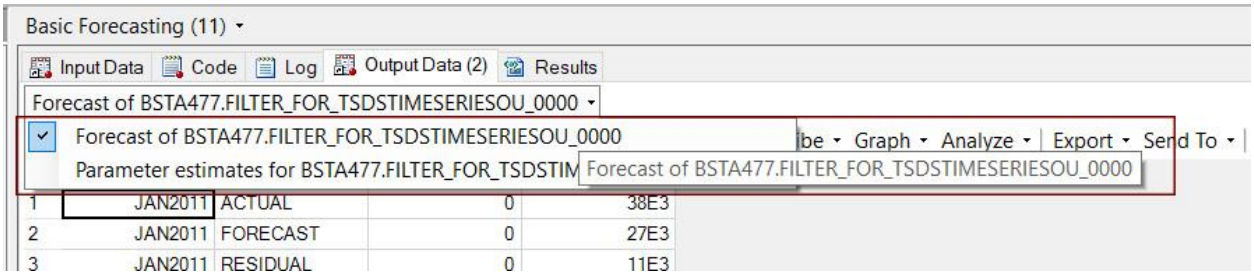

Basic Forecasting (11) for Local:BSTA477.FILTER_FOR_TSDSTIMESERIESOU_0000
Data
Forecast Options
Plots
Results
Titles
Properties
Forecast Options
Forecasting method:
Winters (Multiplicative) Method
Number of intervals to forecast:
12
Input data options
Time interval between observations
Monthly
Time units per interval:
1
Seasonal cycle length
One month
Intervals per seasonal cycle:
2
Method options
Degree of time trend model:
Linear
Confidence level:
95%
Component smoothing weights
Constant: alpha
0.10557
Linear, quadratic: beta
0.10557
Seasonal: gamma
0.10557

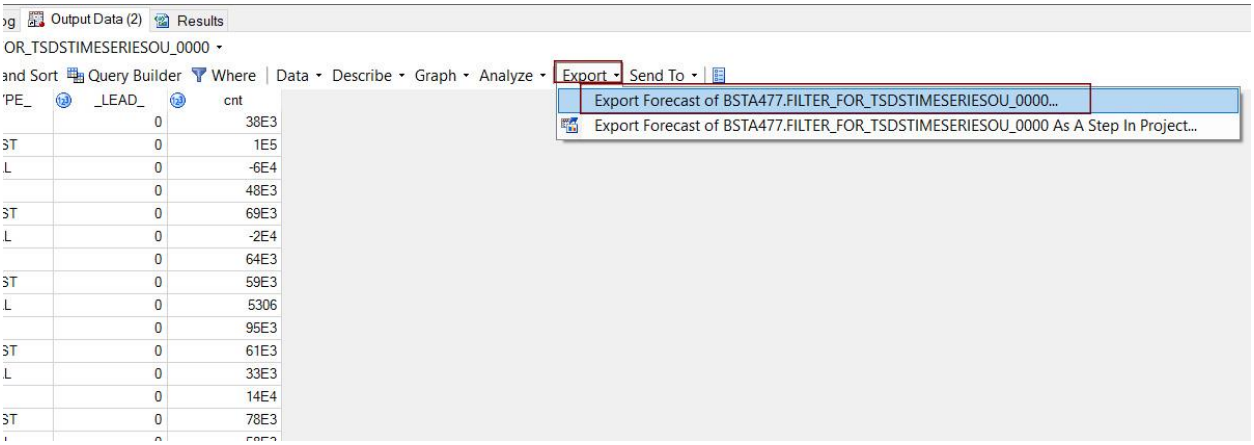

Basic Forecasting (10) for Local:BSTA477.FILTER_FOR_TSDSTIMESERIESOU_0000
Data
Forecast Options
Plots
Results
Titles
Properties
Plots
Types
Show actual values plot
Show forecast data plot
Include:
Actual values and 1-step-ahead predictions
Confidence limits of forecasts
Residuals

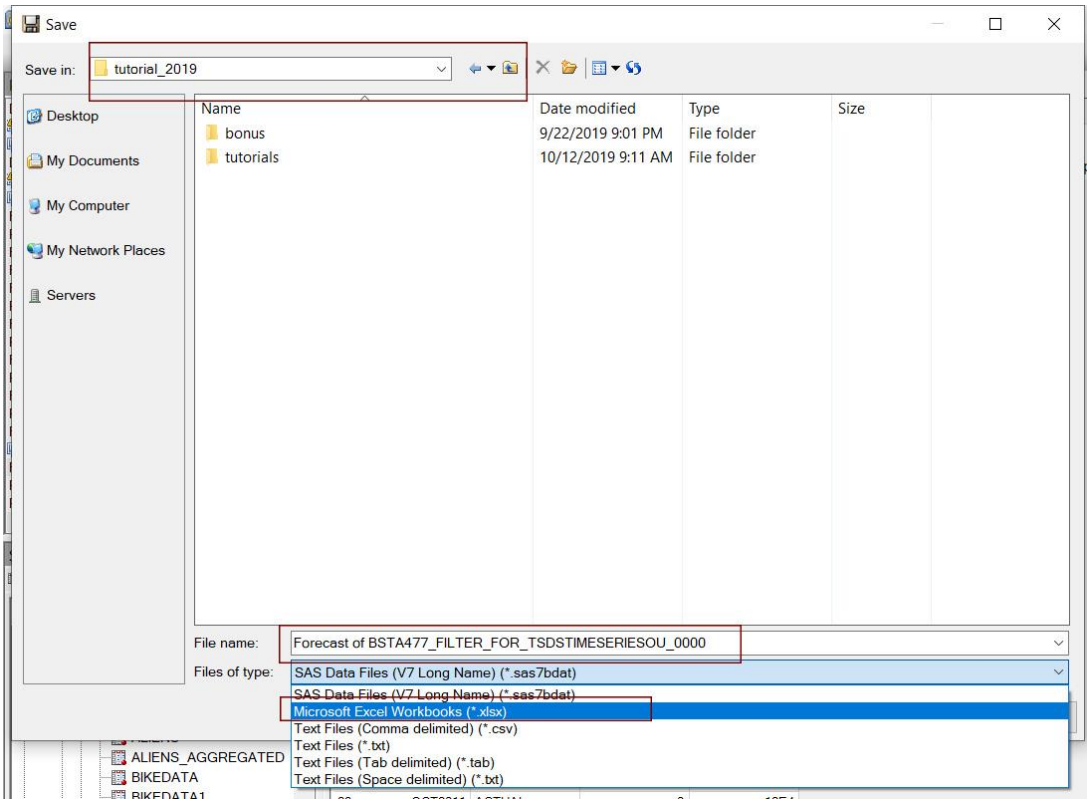

Basic Forecasting (10) for Local:BSTA477.FILTER_FOR_TSDSTIMESERIESOU_0000
Data
Forecast Options
Plots
Results
Titles
Properties
Results
Save output data
Save forecasts
Note: Forecasts are always saved.
Local:WORK.FCS0FORECAST_0008
Browse...
Parameter estimates
Local:WORK.FCS1PARAMETERESTIMAT
Browse...
Additional statistics to save:
R-squared-type forecast accuracy
Theil forecast accuracy
Display output
Show:
Forecasts
Parameter estimates
Includes the forecast values in the displayed results.
Preview code
Run
Save
Cancel
Help


cnt
23E4
22E4
21E4
2E5
19E4
18E4
17E4
16E4
15E4
14E4
13E4
12E4
11E4
1E5
9E4
8E4
7E4
6E4
5E4
4E4
3E4
2E4
JAN2011
MAR2011
MAY2011
JUL2011
SEP2011
NOV2011
JAN2012
MAR2012
MAY2012
JUL2012
SEP2012
NOV2012
JAN2013
dteday
Type of Observation
ACTUAL
FORECAST
L95
U95


cnt
2E4
1E4
0
-1E4
-2E4
-3E4
-4E4
JAN2011
FEB2011
MAR2011
APR2011
MAY2011
JUN2011
JUL2011
AUG2011
SEP2011
OCT2011
NOV2011
DEC2011
JAN2012
dteday
Type of Observation
RESIDUAL


Basic Forecasting (11)
Input Data
Code
Log
Output Data (2)
Results
Forecast of BSTA477.FILTER_FOR_TSDSTIMESERIESOU_0000
Forecast of BSTA477.FILTER_FOR_TSDSTIMESERIESOU_0000
Parameter estimates for BSTA477.FILTER_FOR_TSDSTIM
Graph
Analyze
Export
Send To
Forecast of BSTA477.FILTER_FOR_TSDSTIMESERIESOU_0000
1 JAN2011 ACTUAL 0 38E3
2 JAN2011 FORECAST 0 27E3
3 JAN2011 RESIDUAL 0 11E3


Output Data (2)
Results
Query Builder
Where
Data
Describe
Graph
Analyze
Export
Send To
Export Forecast of BSTA477.FILTER_FOR_TSDSTIMESERIESOU_0000...
Export Forecast of BSTA477.FILTER_FOR_TSDSTIMESERIESOU_0000 As A Step In Project...
LEAD
cnt
38E3
1E5
-6E4
48E3
69E3
-2E4
64E3
59E3
5306
95E3
61E3
33E3
14E4
78E3


Save
Save in:
tutorial_2019
Desktop
My Documents
My Computer
My Network Places
Servers
Name
Date modified
Type
Size
bonus
9/22/2019 9:01 PM
File folder
tutorials
10/12/2019 9:11 AM
File folder
File name:
Forecast of BSTA477_FILTER_FOR_TSDSTIMESERIESOU_0000
Files of type:
SAS Data Files (V7 Long Name) (*.sas7bdat)
SAS Data Files (V7 Long Name) (*.sas7bdat)
Microsoft Excel Workbooks (*.xlsx)
Text Files (Comma delimited) (*.csv)
Text Files (*.txt)
Text Files (Tab delimited) (*.tab)
Text Files (Space delimited) (*.txt)
ALIENS_AGGREGATED
BIKEDATA