SAS® GLOBAL FORUM 2016

IMAGINE. CREATE. INNOVATE.

Case Study Using SAS® Forecast Studio: Daily Taxi Data from New York City - 4001

Simon Sheather
Professor & Academic Director
Texas A&M University

#SASGF

About the presenter

Simon Sheather received his BSc (Hons) degree from the University of Melbourne and a Ph.D. in Statistics from La Trobe University. Currently Simon is Professor and Academic Director of the Statistics Department at Texas A&M University. Previously he was a faculty member at the Australian Graduate School of Management at the University of New South Wales. His research interests are in the fields of flexible regression methods for big data, and smoothing methods in nonparametric statistics. Simon works with industry developing practical predictive models for big data.

SAS® Forecast Studio 13.2

- Automatically fits time series models which can at the same time
- Detect and allow for outliers and level shifts
- Determine which model(s) optimize the chosen fit criterion

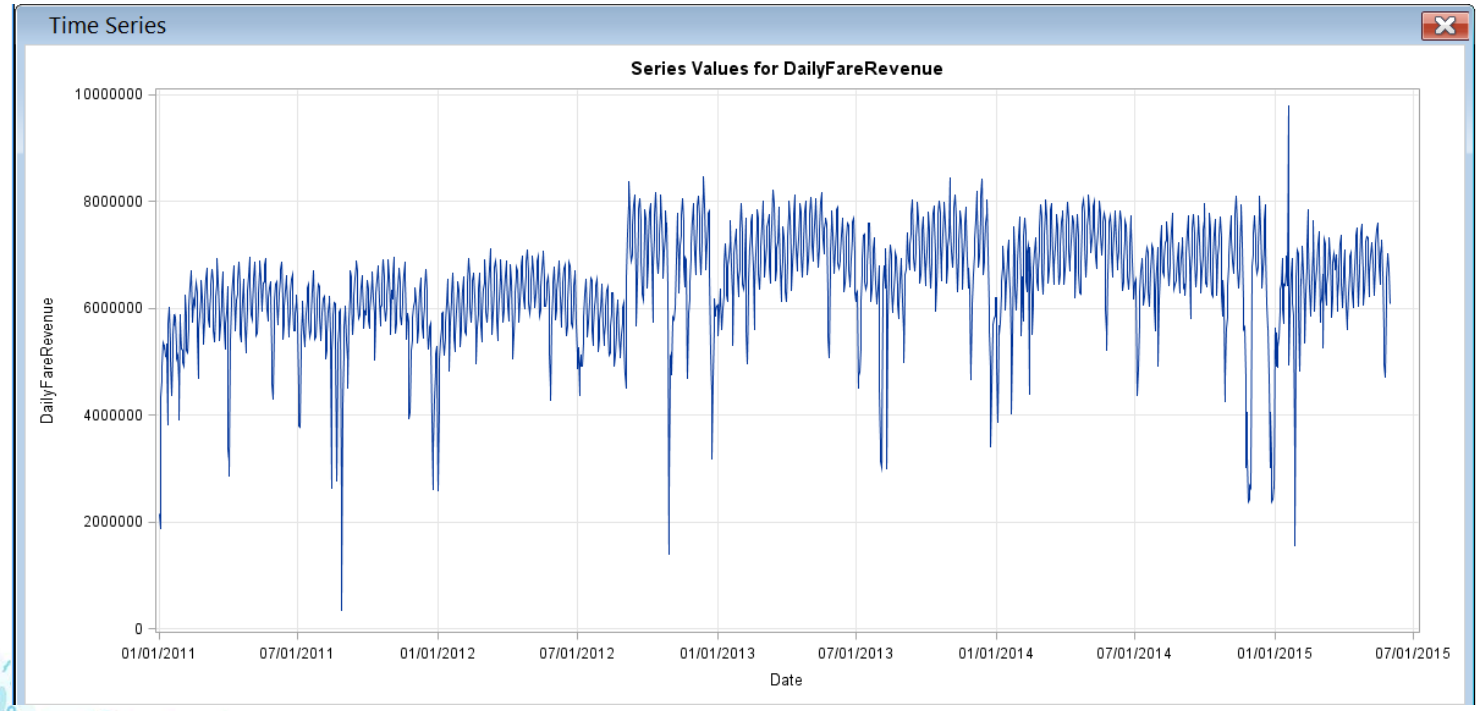
Outline

- Use SAS® Forecast Studio to develop times series models for New York City Yellow cabs
  1. Daily taxi fare revenue
  2. Daily taxi trips

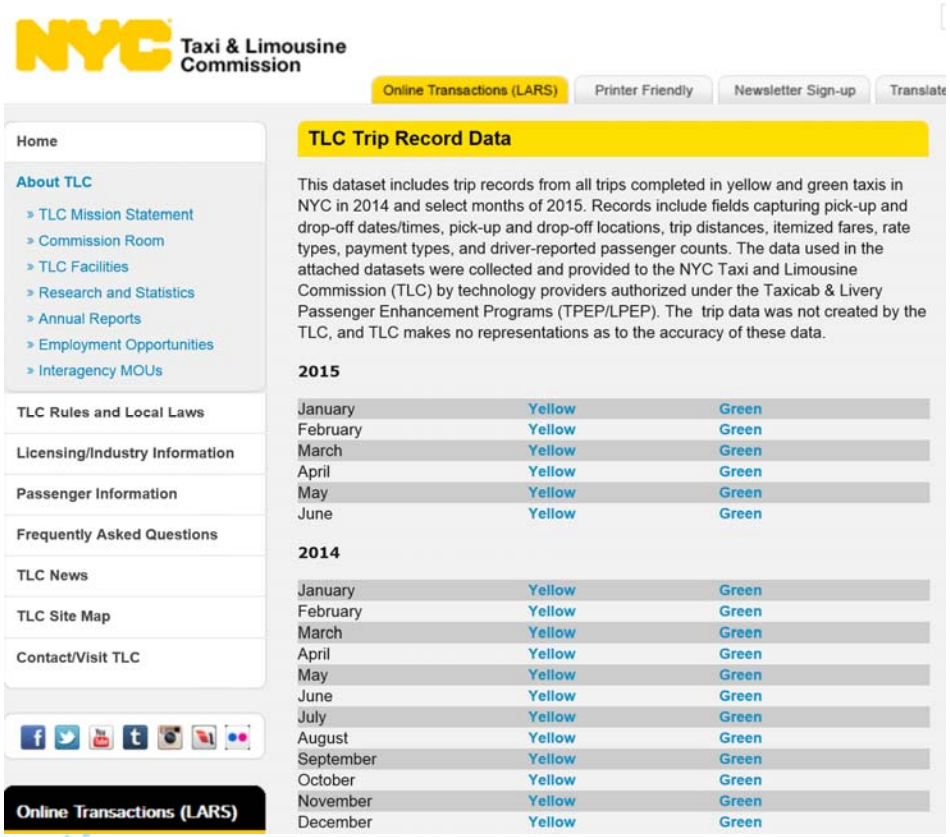
Part 1. Daily taxi fare revenue in New York City

- We wish to develop a time series model for the daily taxi fare revenue by Yellow Cabs in New York City using publicly available data from 1/1/2011 until 5/31/2015 in order to both
- Predict daily taxi fare revenue for June 2015
- Estimate the impact of events including
  - Price increase
  - Holidays
  - One of events such as hurricanes

Daily taxi fare revenue in New York City



Daily taxi fare revenue and trip data in New York City



NYC Taxi & Limousine Commission

Online Transactions (LARS) | Printer Friendly | Newsletter Sign-up | Translate

Home

About TLC
- » TLC Mission Statement
- » Commission Room
- » TLC Facilities
- » Research and Statistics
- » Annual Reports
- » Employment Opportunities
- » Interagency MOUs

TLC Rules and Local Laws

Licensing/Industry Information

Passenger Information

Frequently Asked Questions

TLC News

TLC Site Map

Contact/Visit TLC

Online Transactions (LARS)

TLC Trip Record Data

This dataset includes trip records from all trips completed in yellow and green taxis in NYC in 2014 and select months of 2015. Records include fields capturing pick-up and drop-off dates/times, pick-up and drop-off locations, trip distances, itemized fares, rate types, payment types, and driver-reported passenger counts. The data used in the attached datasets were collected and provided to the NYC Taxi and Limousine Commission (TLC) by technology providers authorized under the Taxicab & Livery Passenger Enhancement Programs (TPEP/LPEP). The trip data was not created by the TLC, and TLC makes no representations as to the accuracy of these data.

2015

January	Yellow	Green
February	Yellow	Green
March	Yellow	Green
April	Yellow	Green
May	Yellow	Green
June	Yellow	Green

2014

January	Yellow	Green
February	Yellow	Green
March	Yellow	Green
April	Yellow	Green
May	Yellow	Green
June	Yellow	Green
July	Yellow	Green
August	Yellow	Green
September	Yellow	Green
October	Yellow	Green
November	Yellow	Green
December	Yellow	Green

Source: http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml

New York City fare increase and weather events

The New York Times

As of Tuesday, Yellow Cabs Can Charge Higher Rates

By MATT FLEGENHEIMER SEPT. 3, 2012

Email
Share
Tweet
Save

Yellow taxis may begin charging more on Tuesday, ushering in an approved fare increase for riders, whose rates have remained virtually unchanged since 2006.

The city's Taxi and Limousine Commission announced late Monday that operators of yellow taxis would be allowed to put the new fares — which increase rates by about 17 percent — into effect as of 12:01 a.m. on Tuesday, once they have recalibrated their meters and updated external markings.

Storm Barrels Through Region, Le

By JAMES BARRON OCT. 29, 2012

Email
Share
Tweet
Save

Hurricane Sandy battered the mid-Atlantic region on Monday, its powerful gusts and storm surges causing once-in-a-generation flooding in coastal communities, knocking down trees and power lines and leaving more than five million people — including a large swath of Manhattan — in the rain-soaked dark. At least seven deaths in the New York region were tied to the storm.

theguardian

US world opinion sports soccer tech arts lifestyle fashion business travel environment science browse all sections

home › US

Hurricane Irene
News blog

Hurricane Irene hits New York - Sunday 28 August 2011

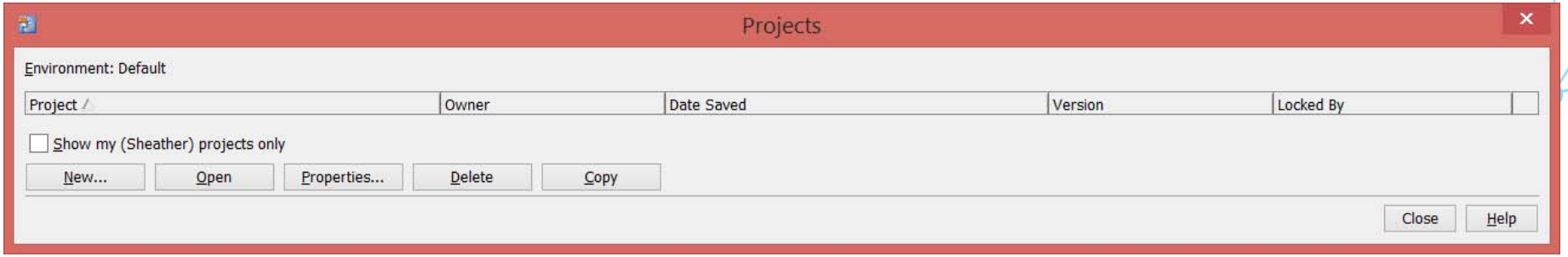
New York Today: Updates on the Winter Storm

By ANDY NEWMAN JANUARY 27, 2015 5:58 AM 25 Comments

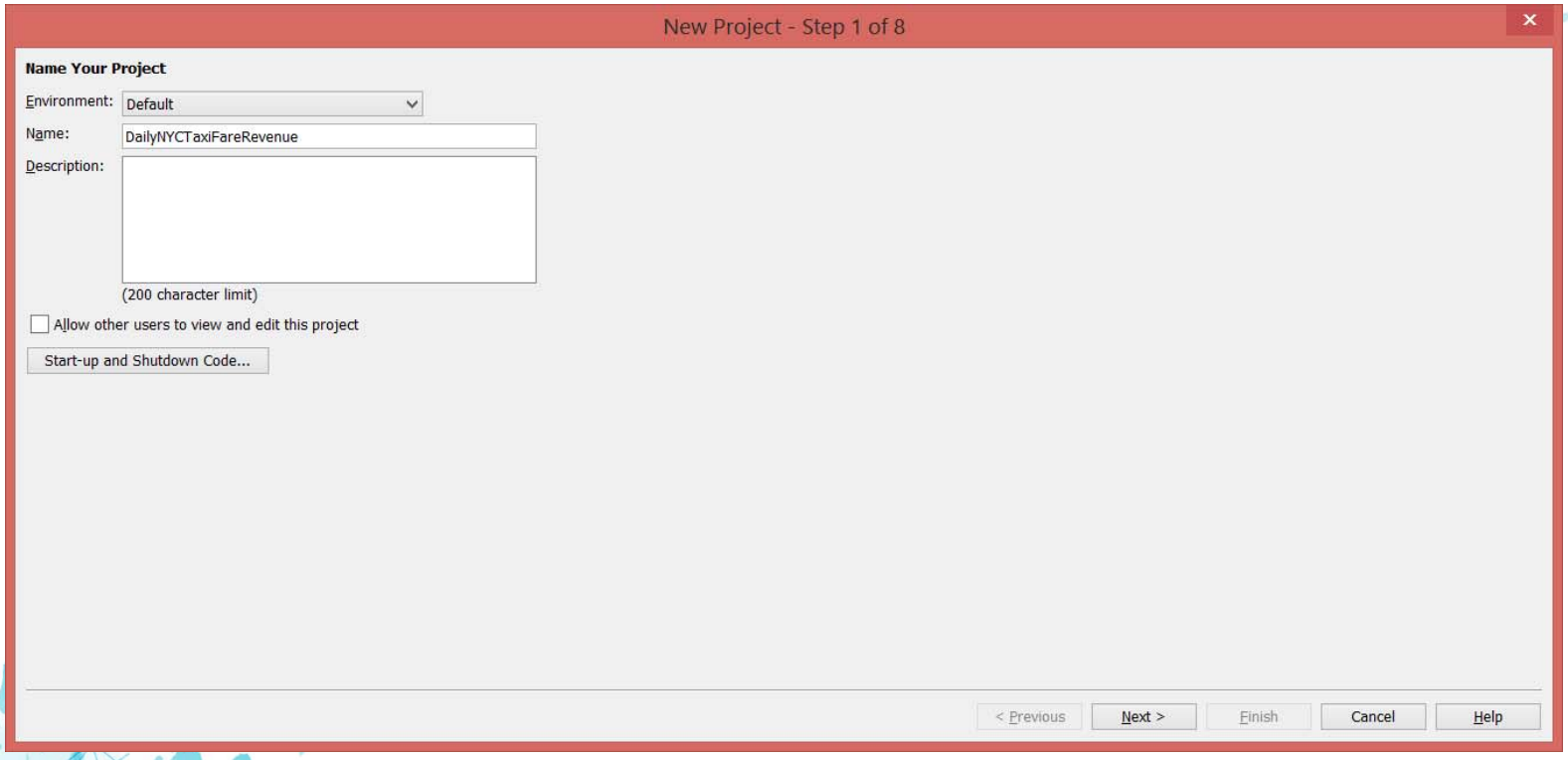
The great blizzard of 2015 did not turn out to be a blizzard in New York City. Only 7.8 inches had fallen in Central Park as of 7 a.m. and not much more after that.

And so the region gradually stirred back to life. The travel ban was lifted in the city and across New Jersey and southern New York at 7:30 a.m. New York City's subway system resumed operating at 9 a.m.

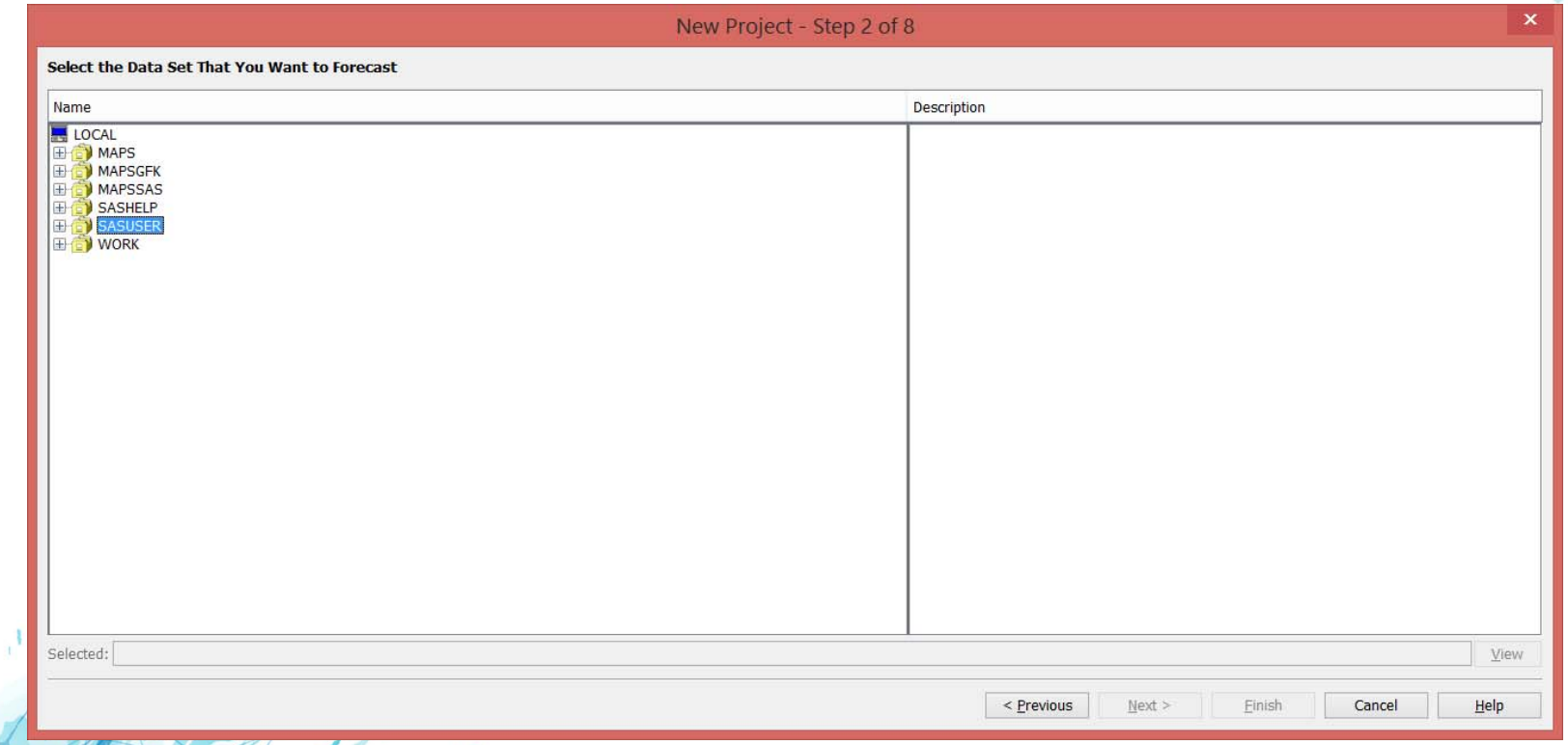
Open SAS Forecast Studio and click on New



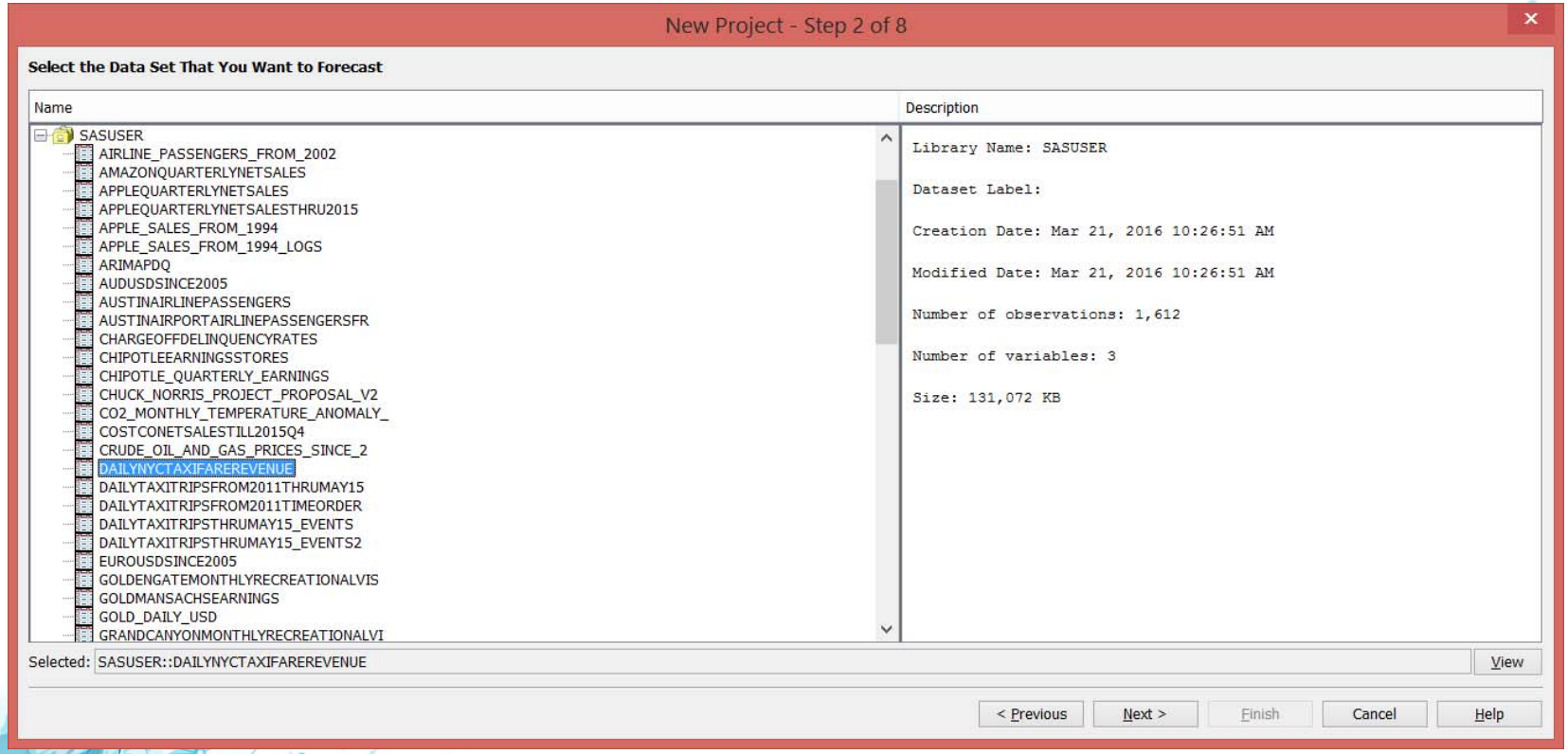
Enter a name for your project and click on Next



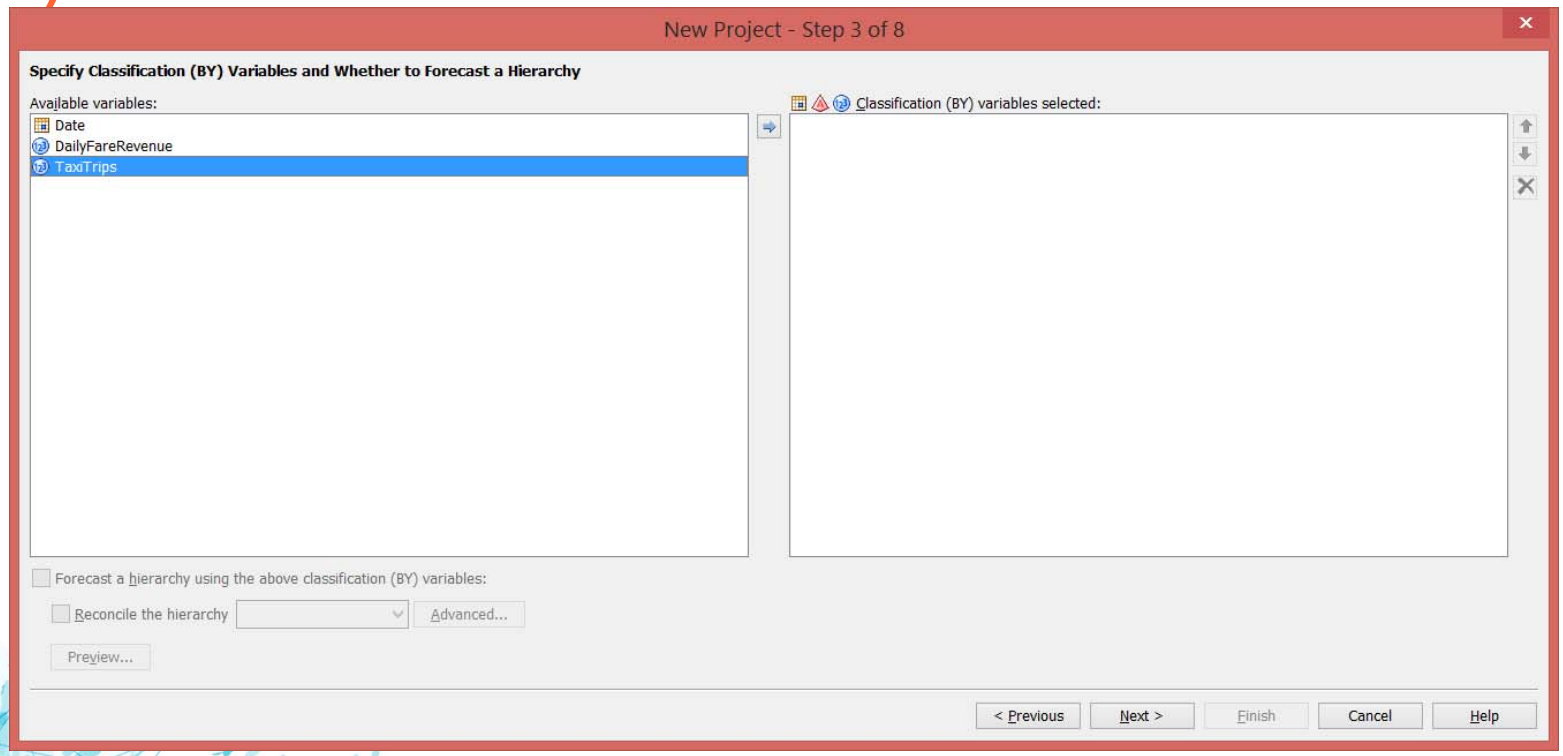
Double click on SASUSER



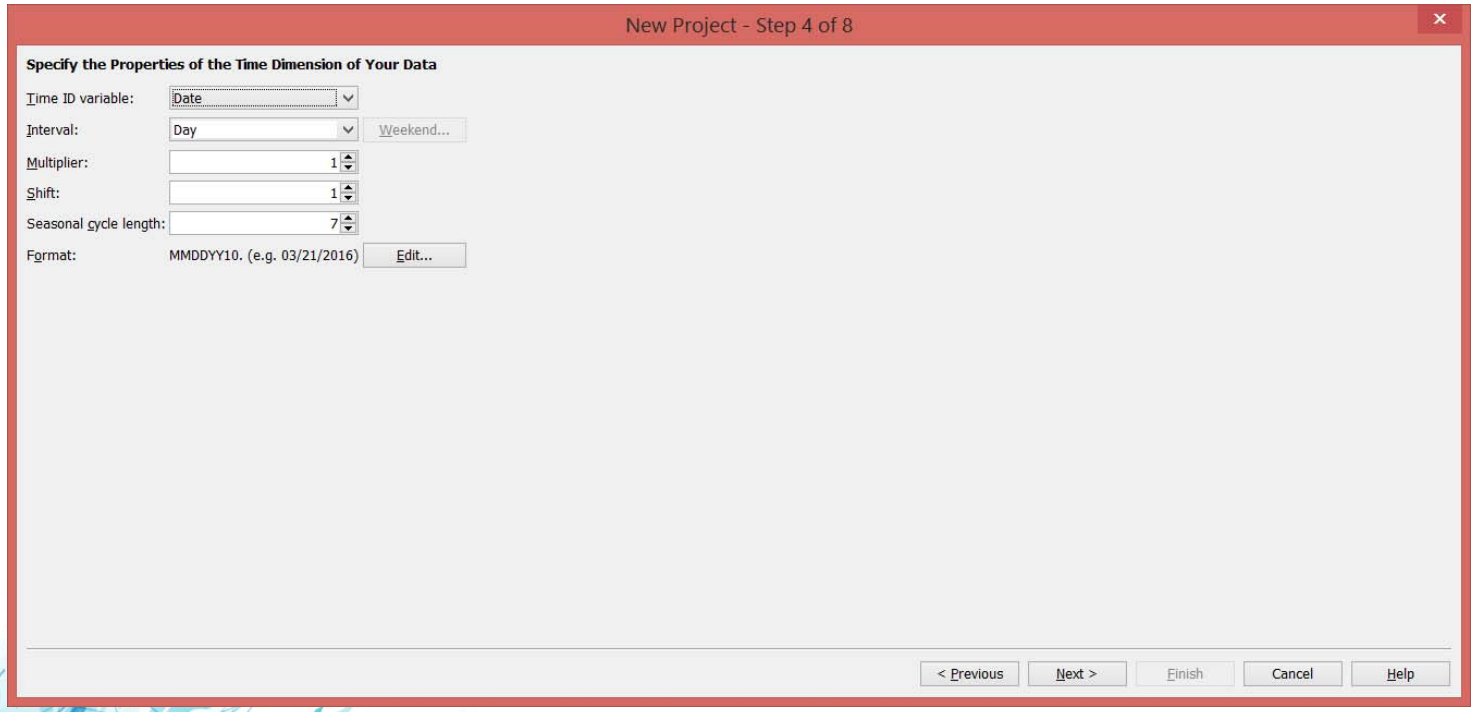
Select the SAS data set and click on Next



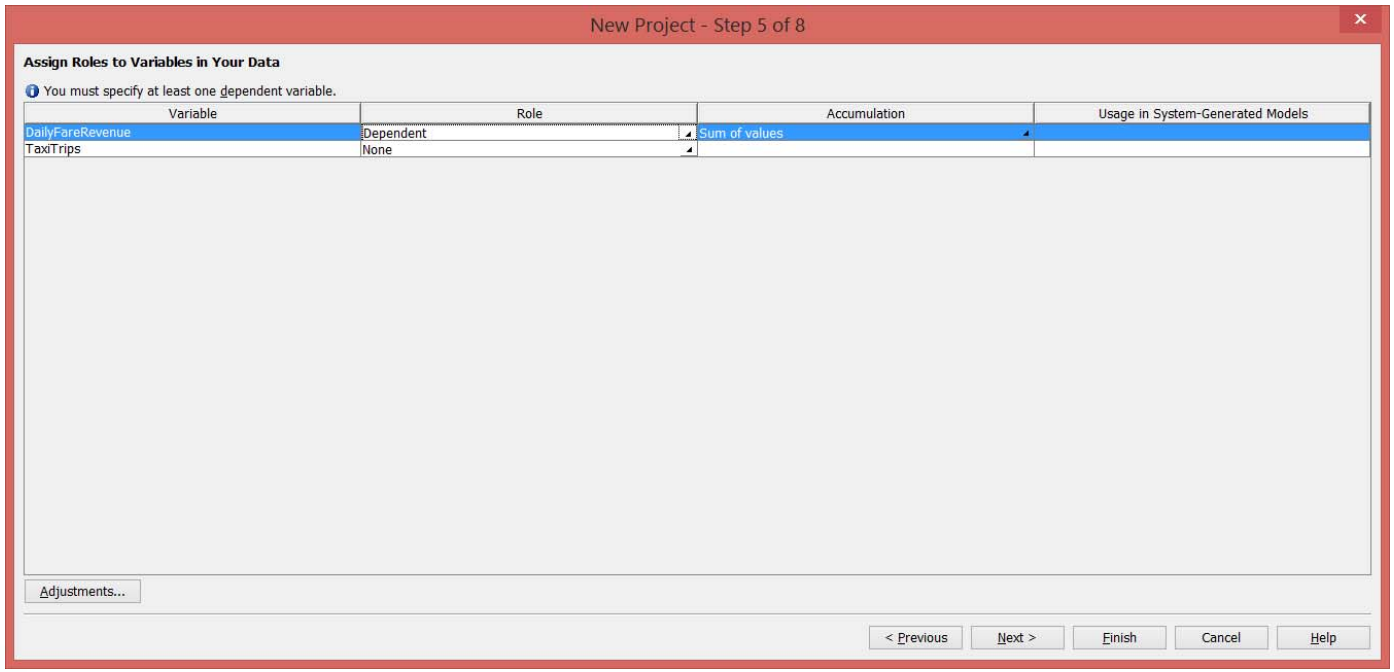
Click on Next since there are no Classification (BY) variables



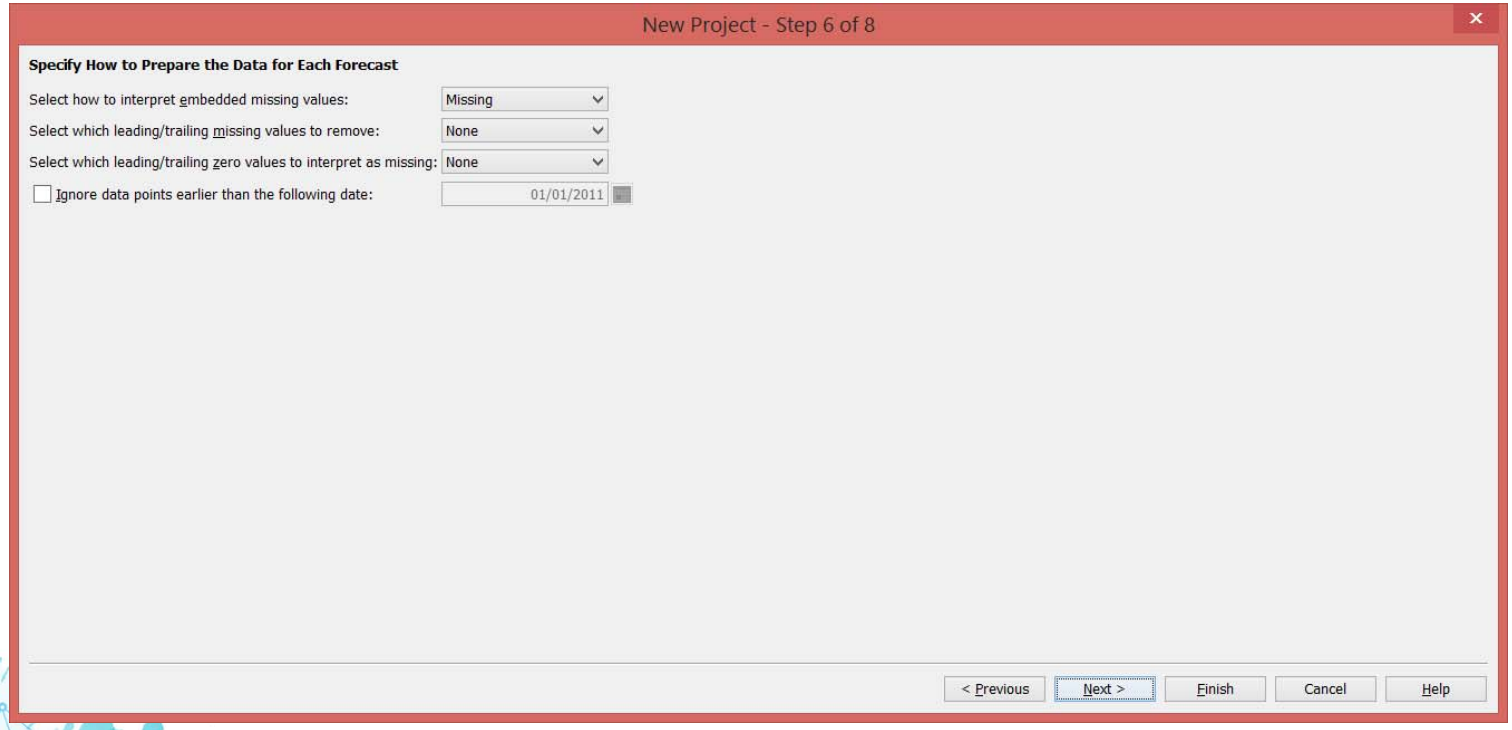
Select the Time ID variable (in this case Date) and click on Next



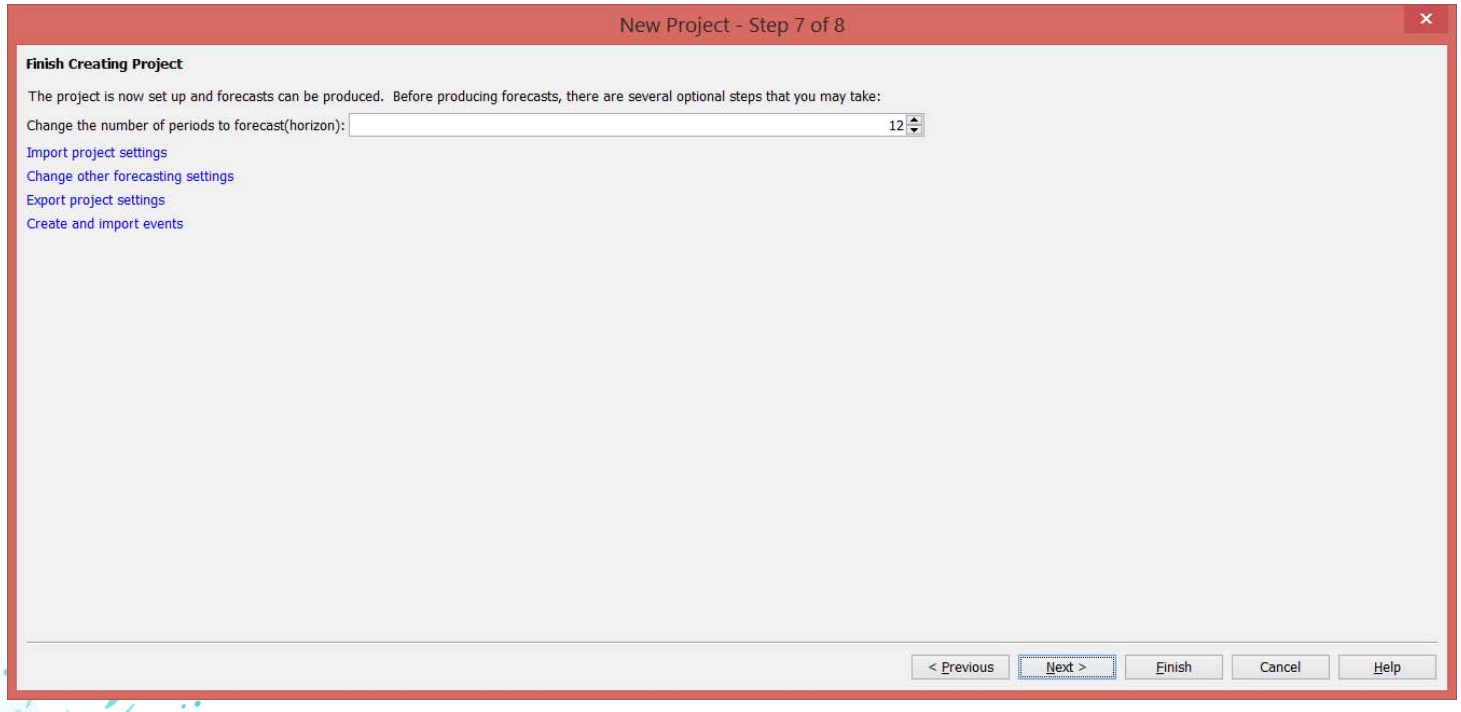
Select the Y variable (in this case DailyFareRevenue) and choose Dependent for its role then click on Next



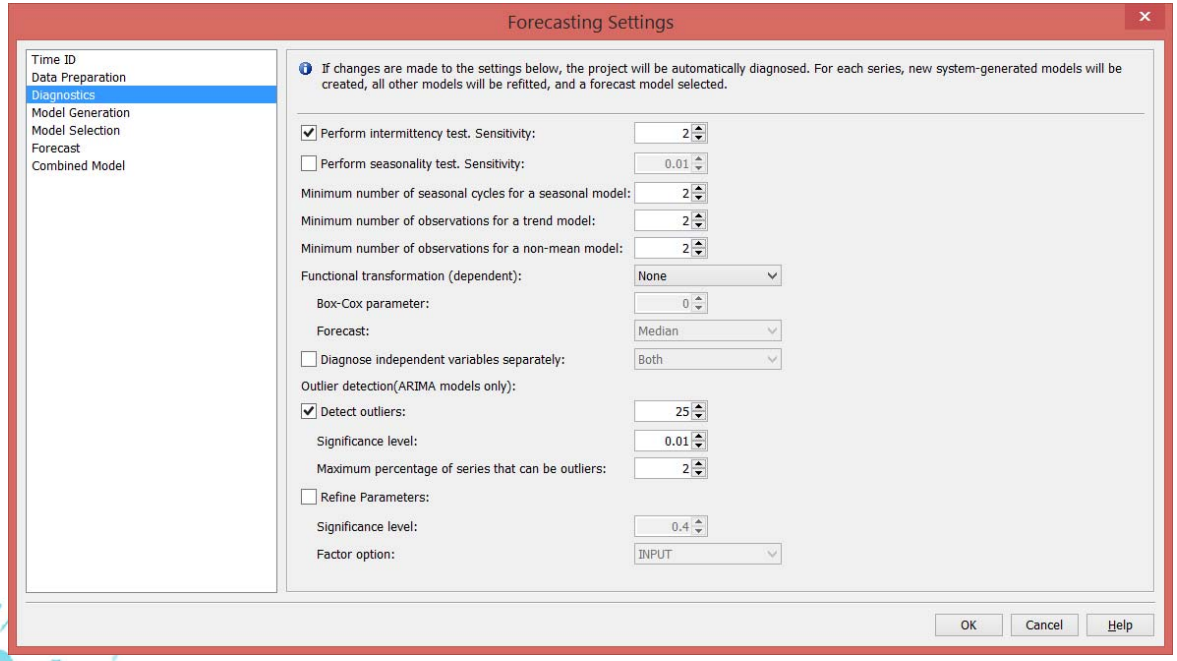
Click on Next (since there are no missing values)



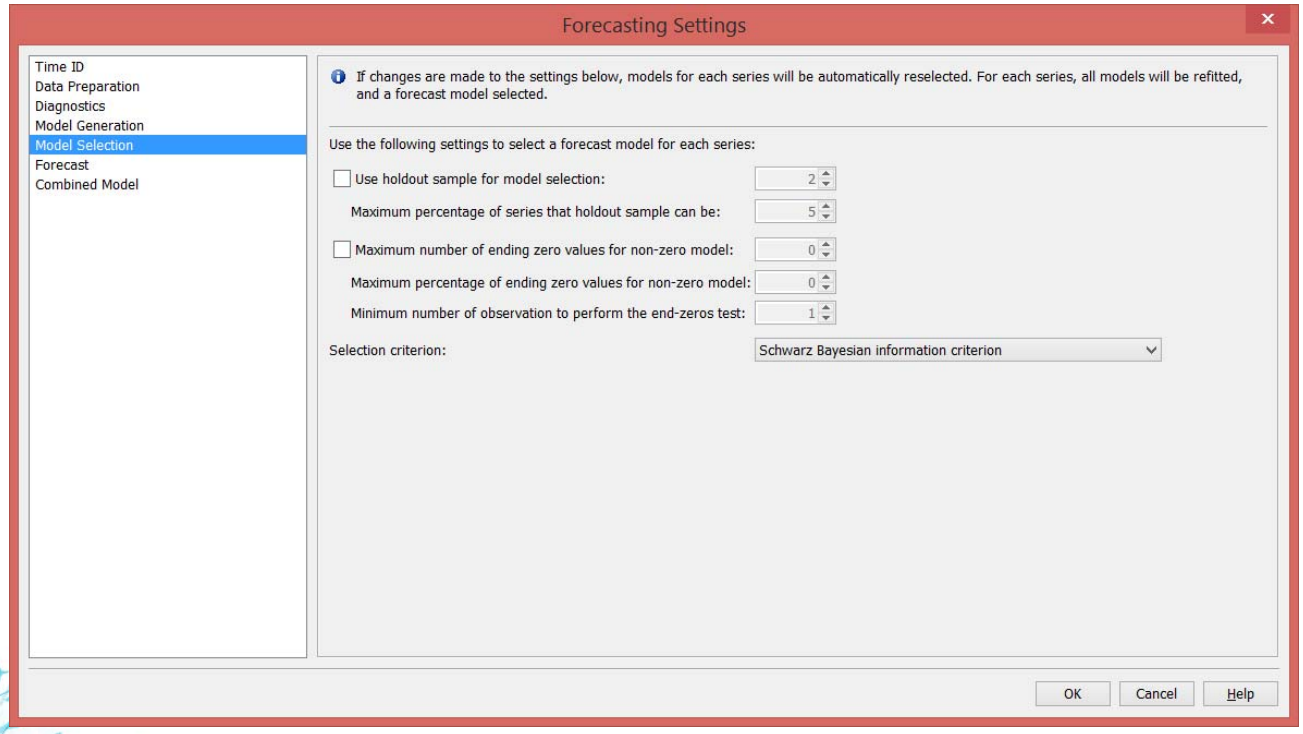
Click on Change other forecasting settings



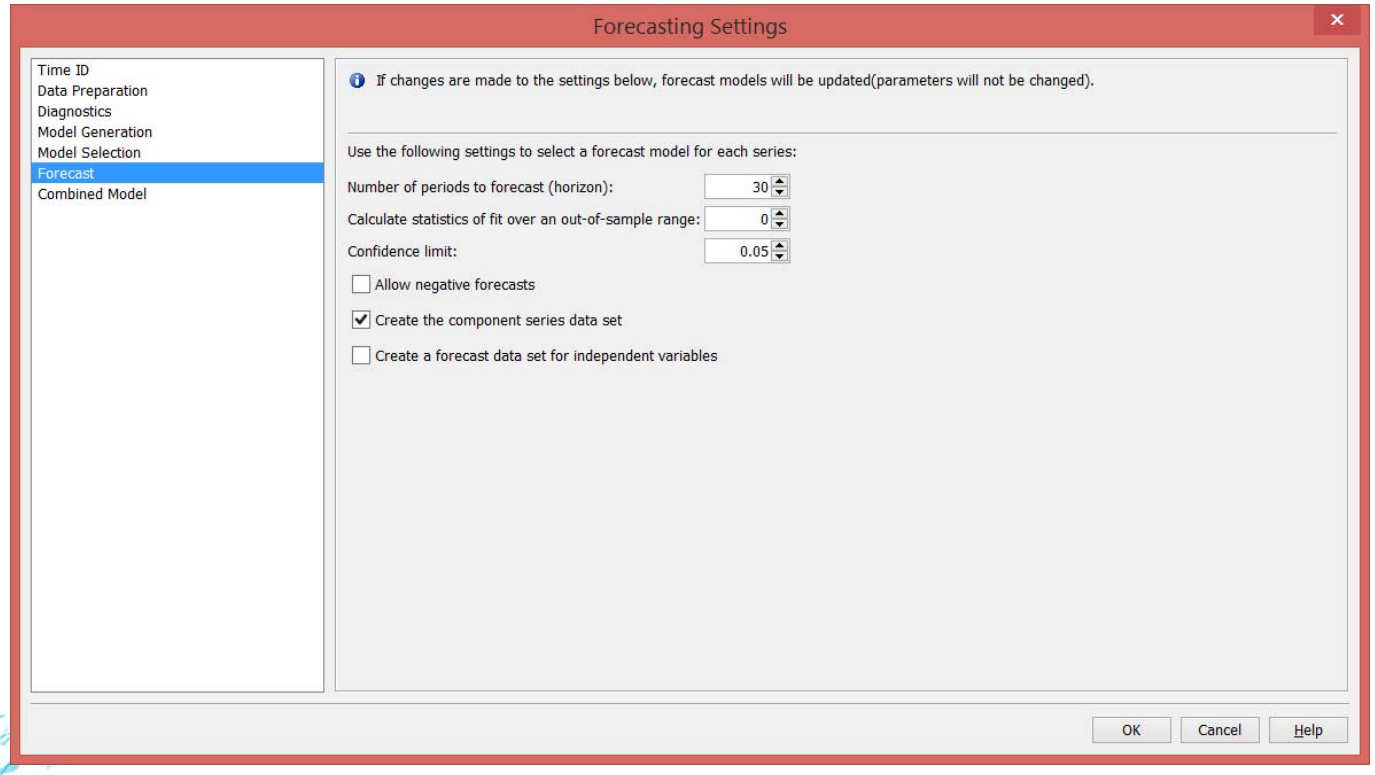
Click on Diagnostics, then click on the box next to Detect outliers change the settings to those below and then click on Model Generation



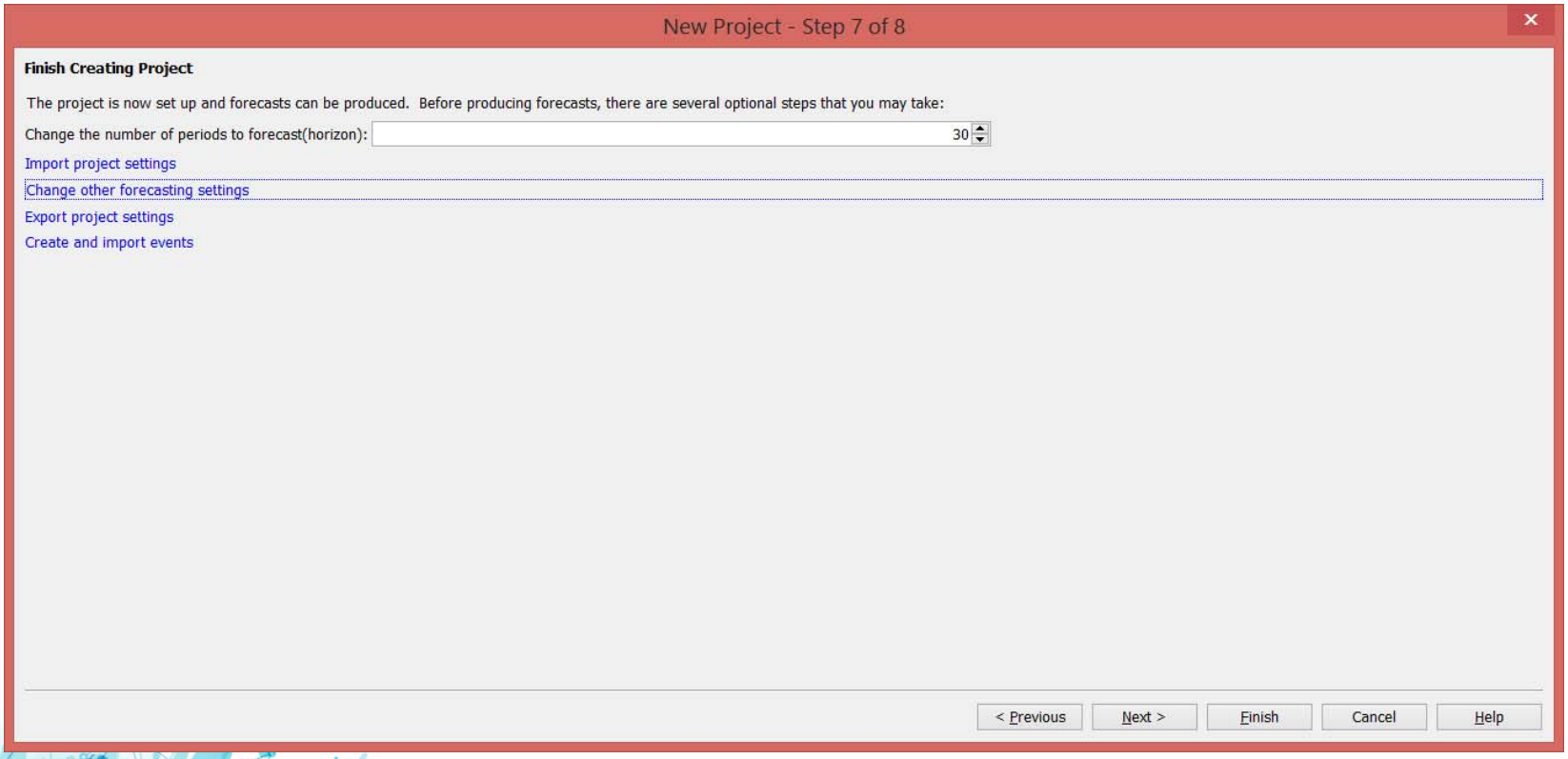
Click on Model Generation, then choose Schwarz Bayesian information criterion as the Selection criterion (since we are seeking a parsimonious model) and then click on Forecast



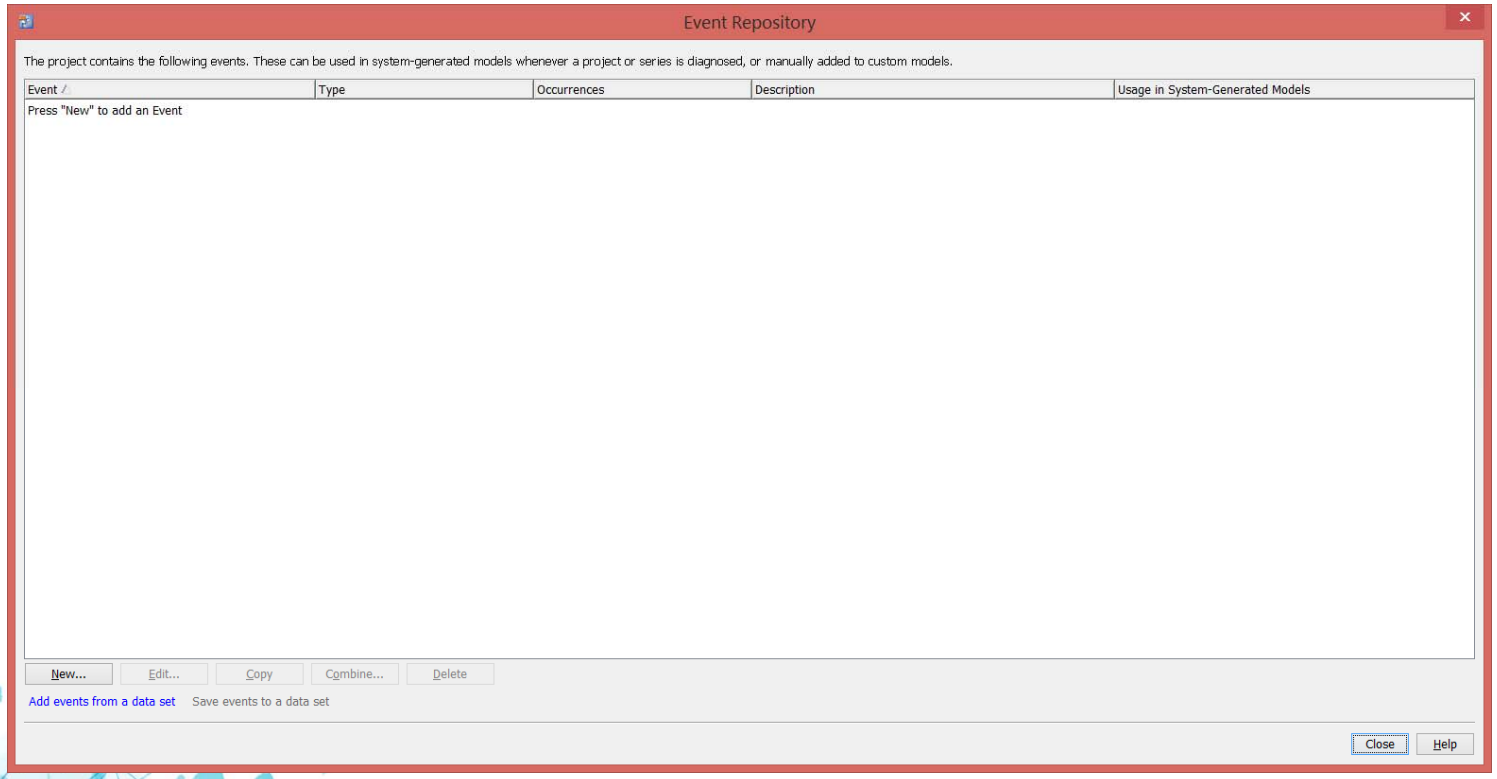
Click on Forecast, then change the Number of periods to forecast to 30 and then click on OK



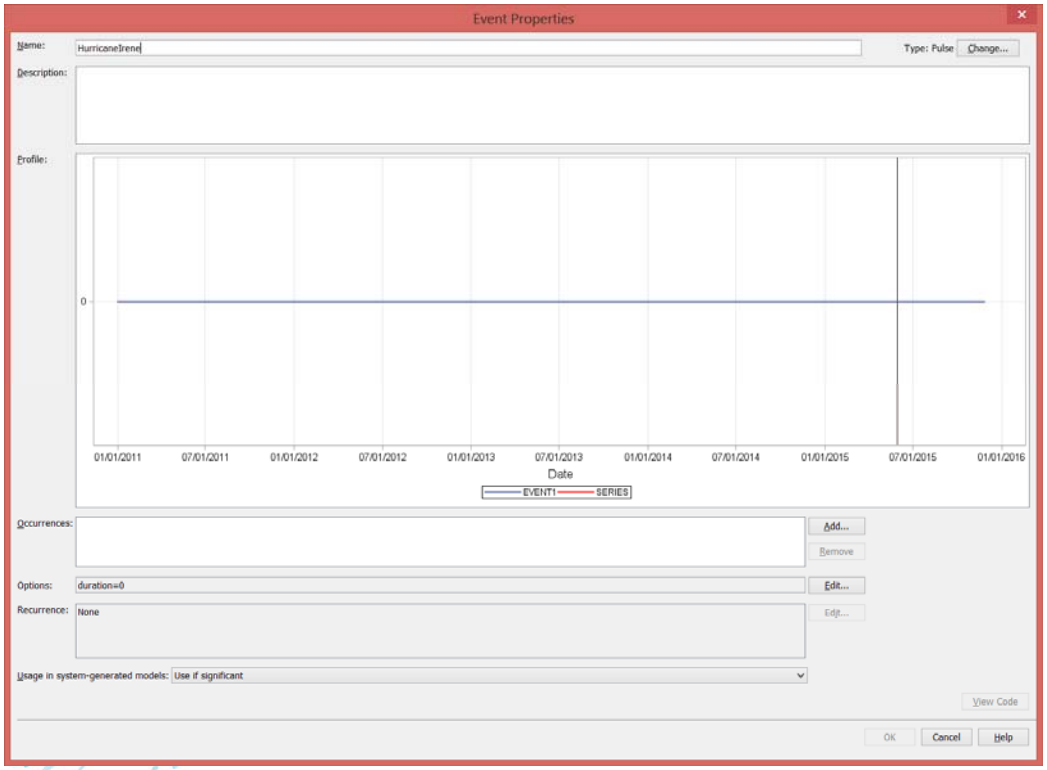
Click on Create and import events



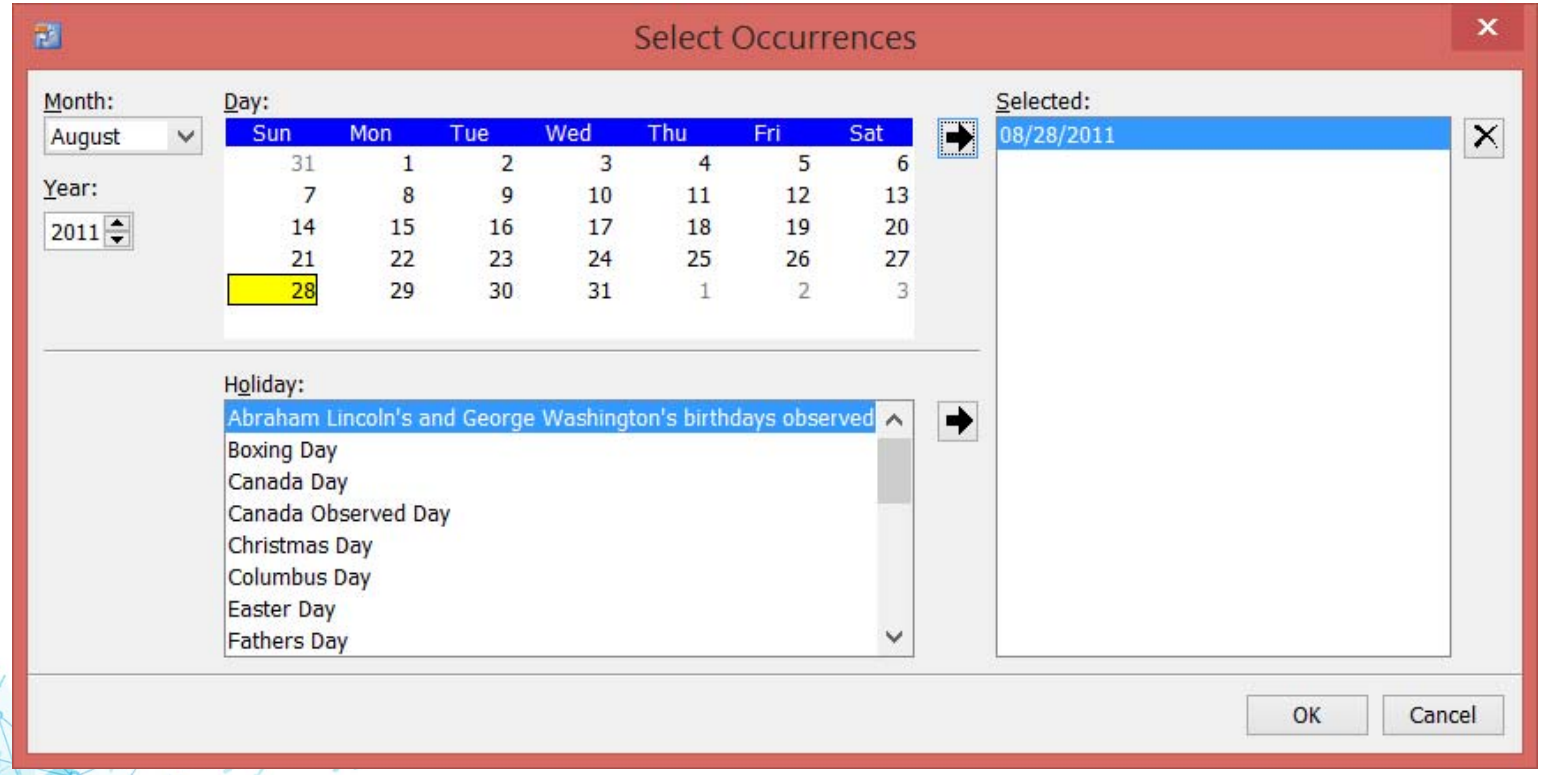
Click on New


Event Repository
The project contains the following events. These can be used in system-generated models whenever a project or series is diagnosed, or manually added to custom models.
Event
Type
Occurrences
Description
Usage in System-Generated Models
Press "New" to add an Event
New...
Edit...
Copy
Combine...
Delete
Add events from a data set
Save events to a data set
Close
Help

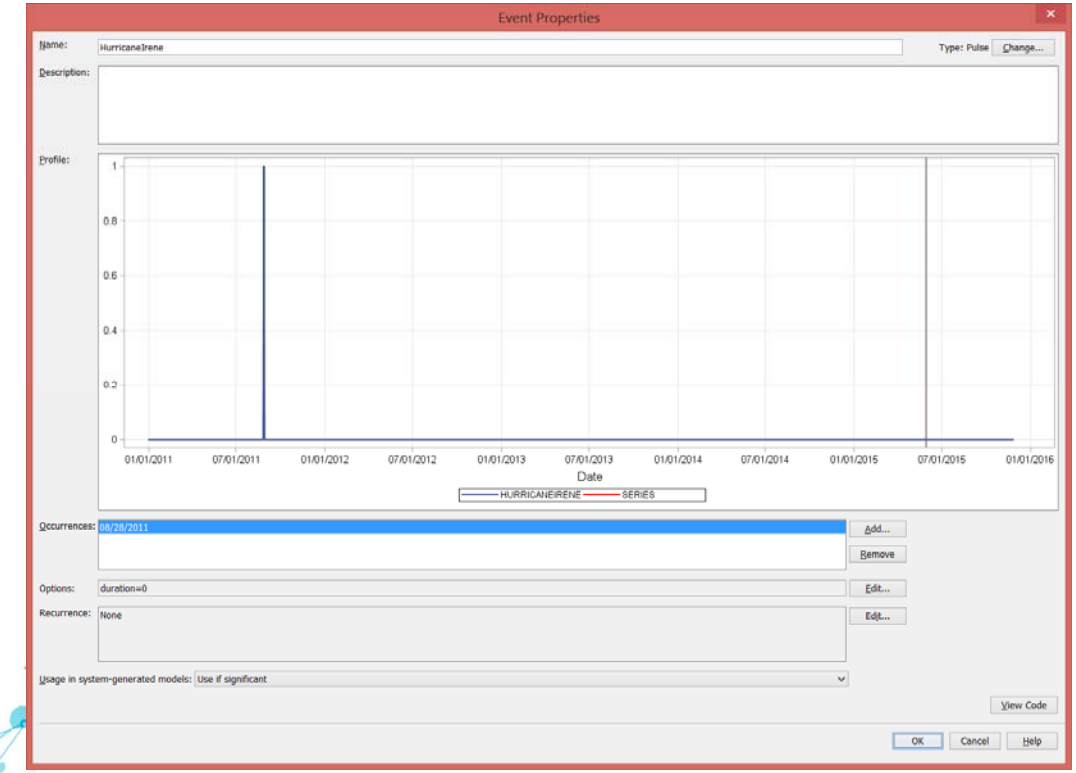
Enter the Name of the event (HurricaneIrene) and click on Add



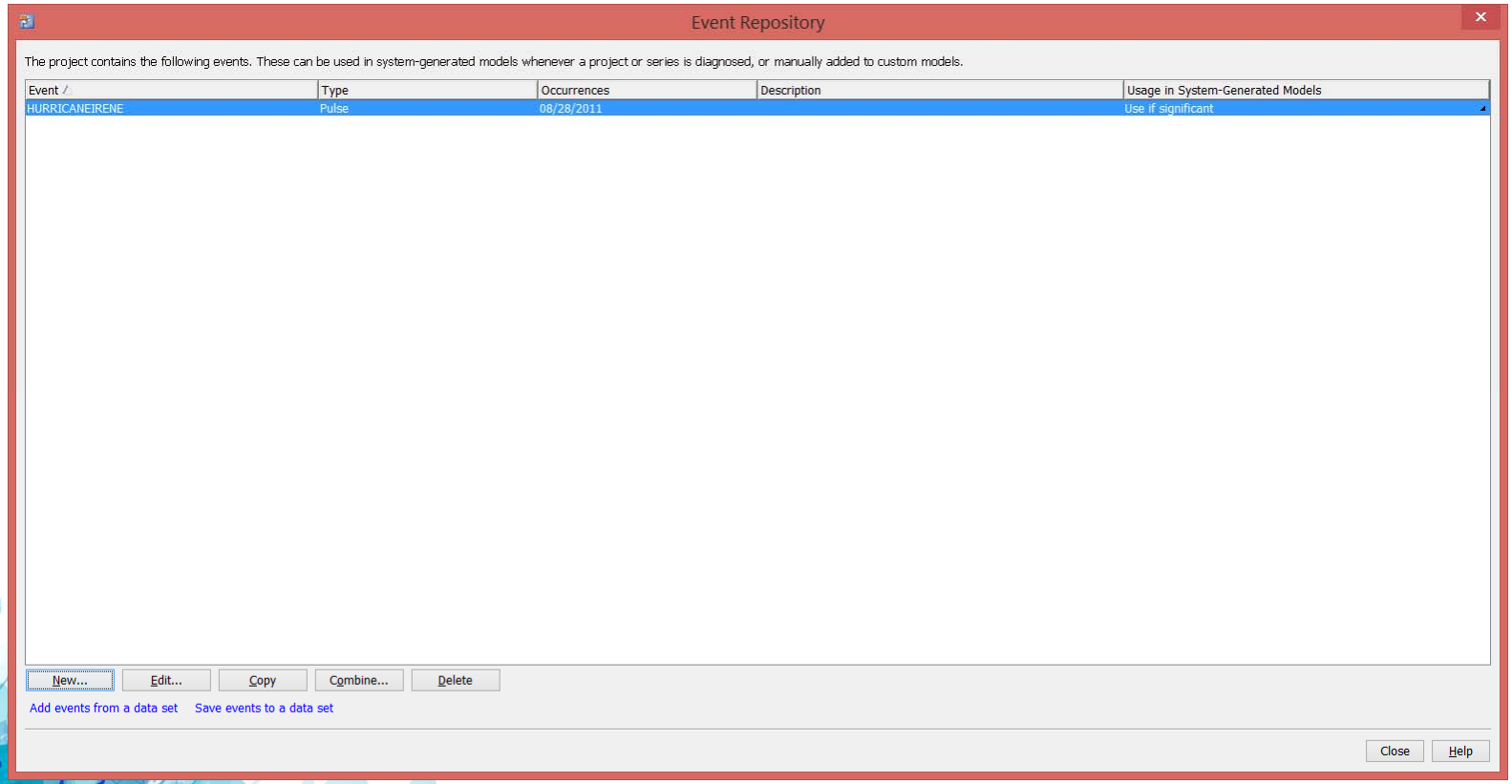
Select the Date of the Event, click on the top arrow and then click on OK



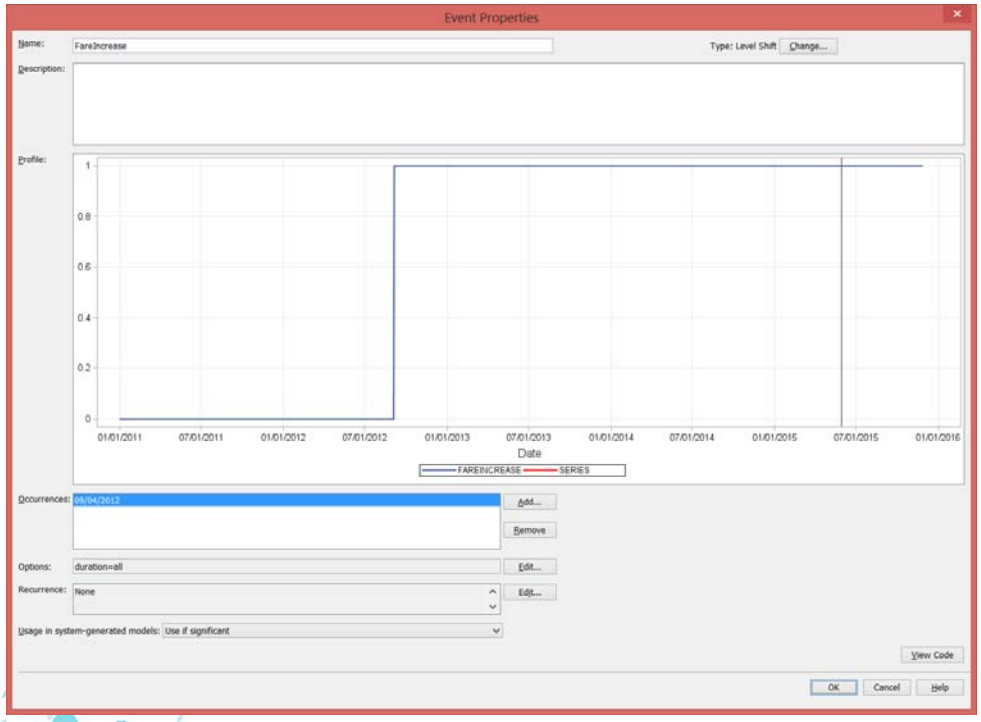
Click on OK

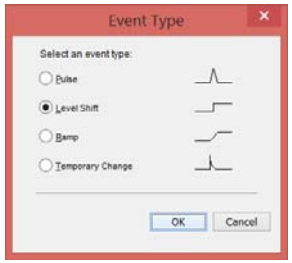

Event Properties
Name:
HurricaneIrene
Type: Pulse
Change...
Description:
Profile:
1
0.8
0.6
0.4
0.2
0
01/01/2011
07/01/2011
01/01/2012
07/01/2012
01/01/2013
07/01/2013
01/01/2014
07/01/2014
01/01/2015
07/01/2015
01/01/2016
Date
HURRICANEIRENE
SERIES
Occurrences:
08/28/2011
Add...
Remove
Options:
duration=0
Edit...
Recurrence:
None
Edit...
Usage in system-generated models:
Use if significant
View Code
OK
Cancel
Help

Click on New add a level shift associated with the fare increase

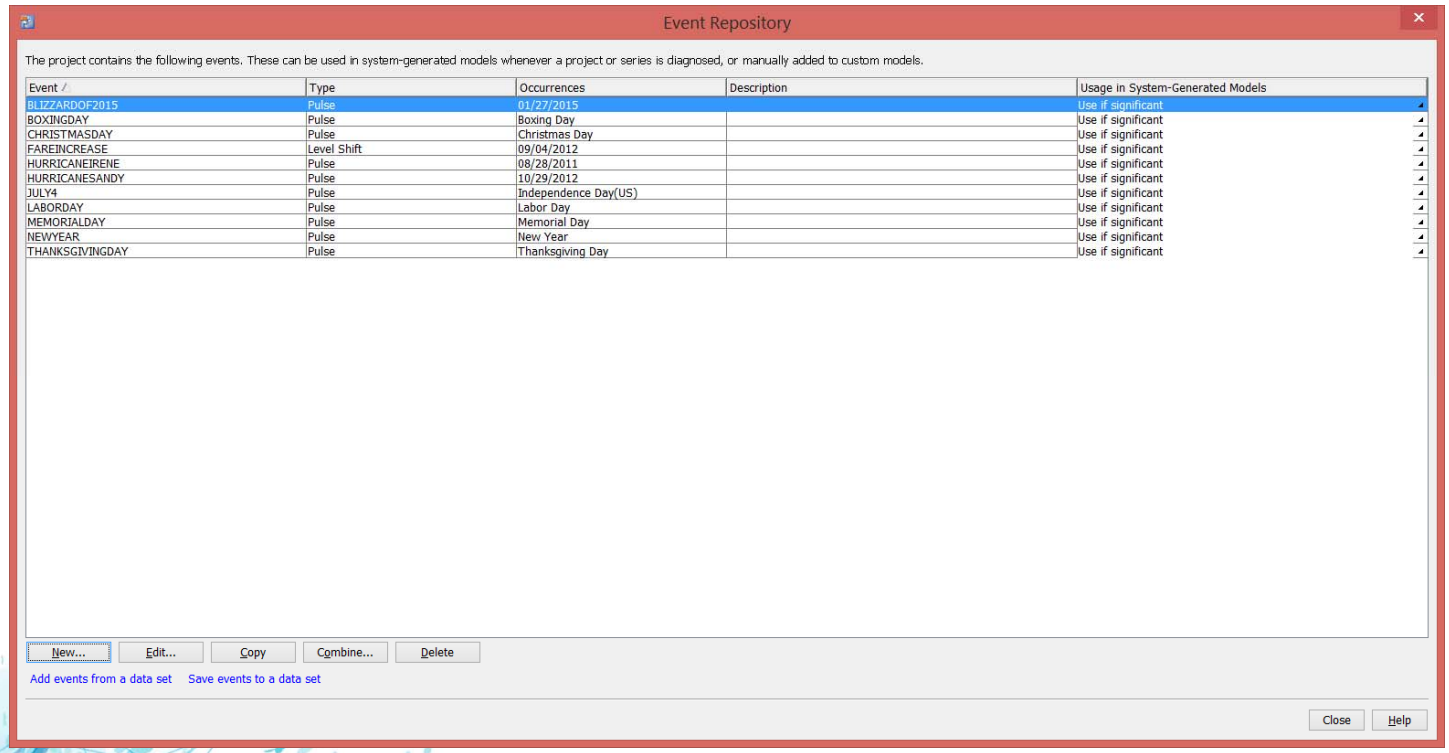


Click on New add a level shift associated with the fare increase on 9/04/2012

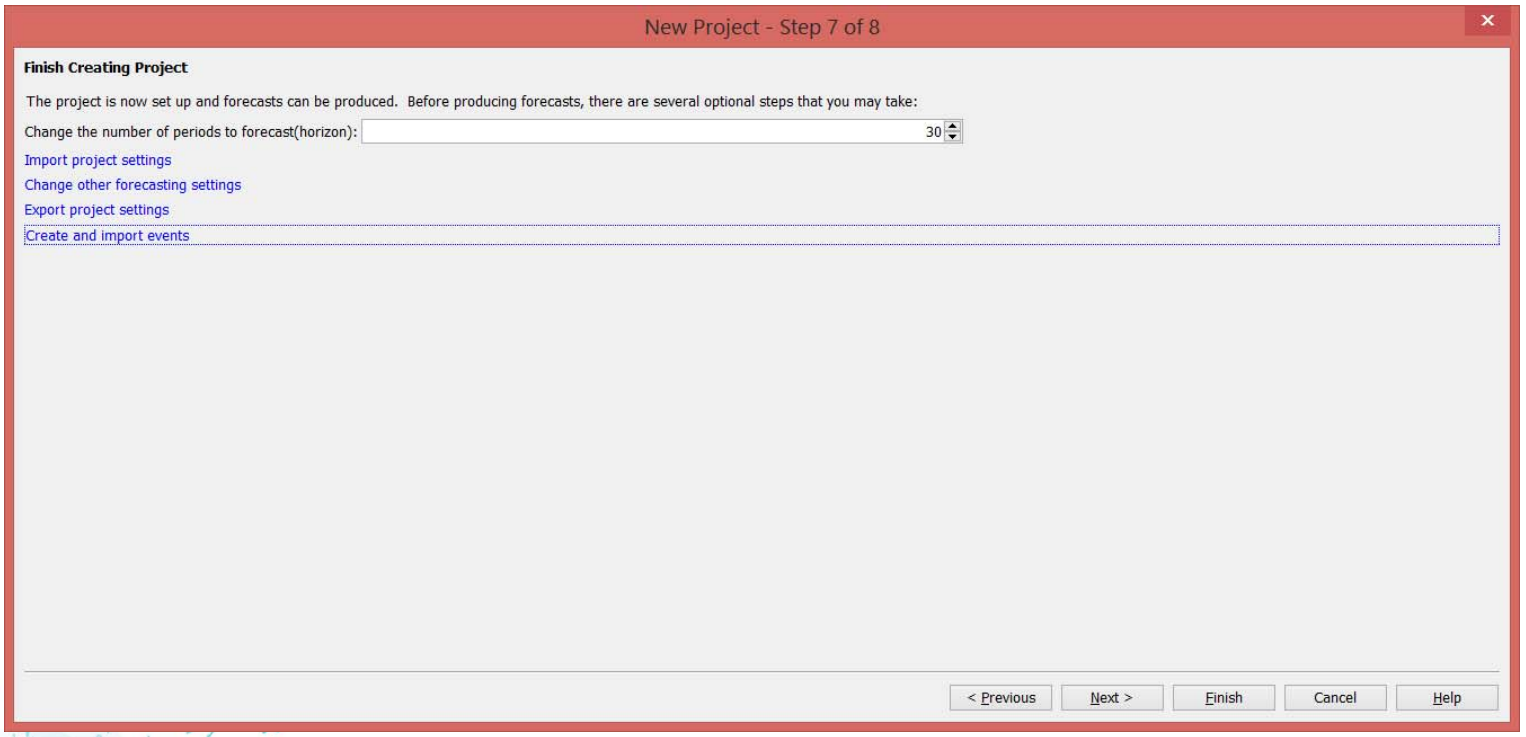




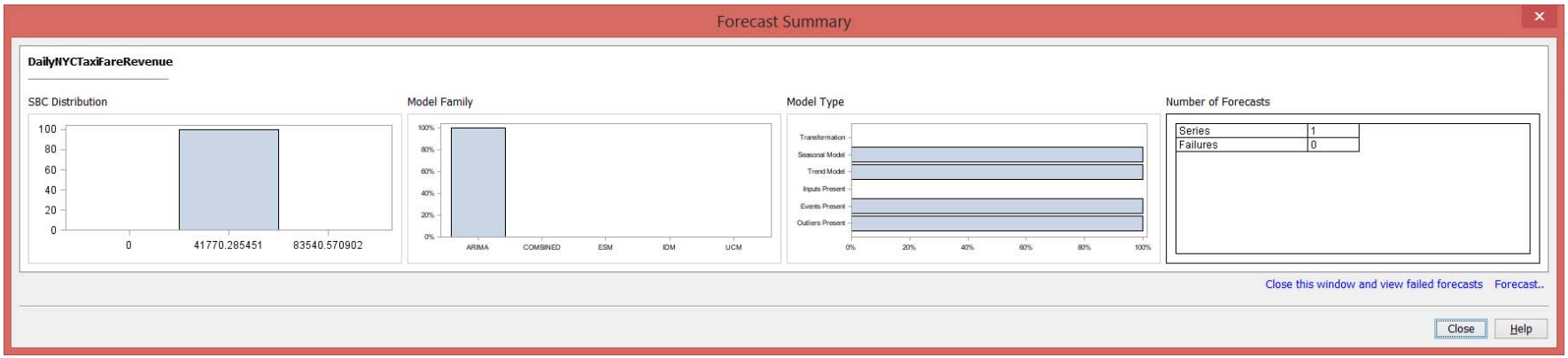
Adding the other events and holidays and then click on Close



Click on Finish

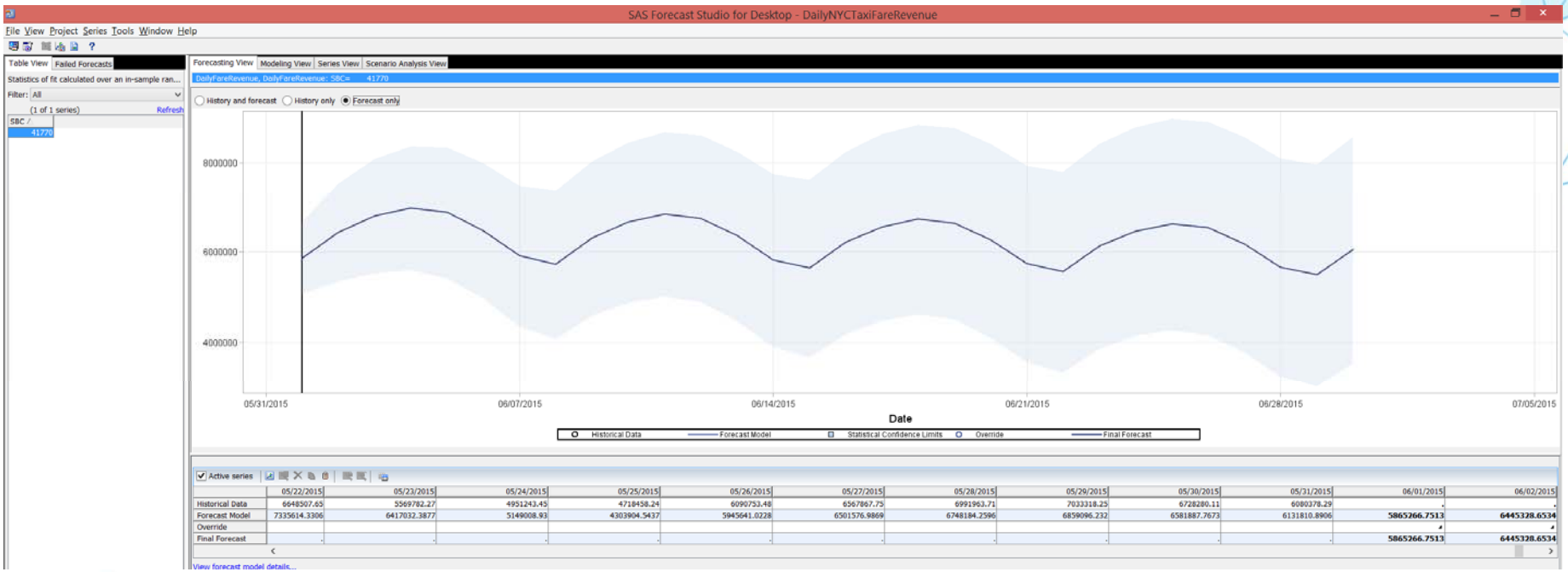


Daily taxi fare revenue in New York City – Forecast Summary

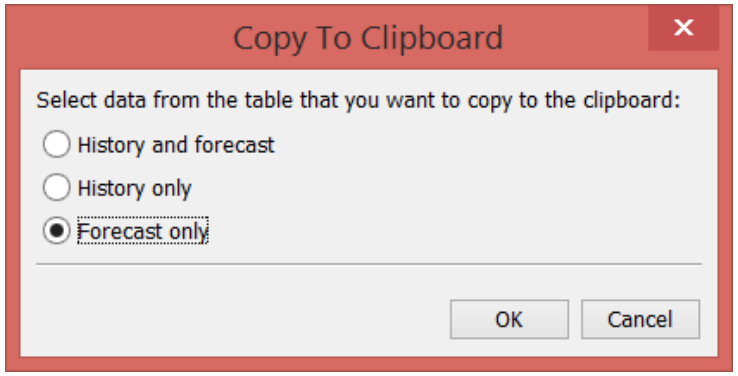


Daily taxi fare revenue in New York City – Forecasting View

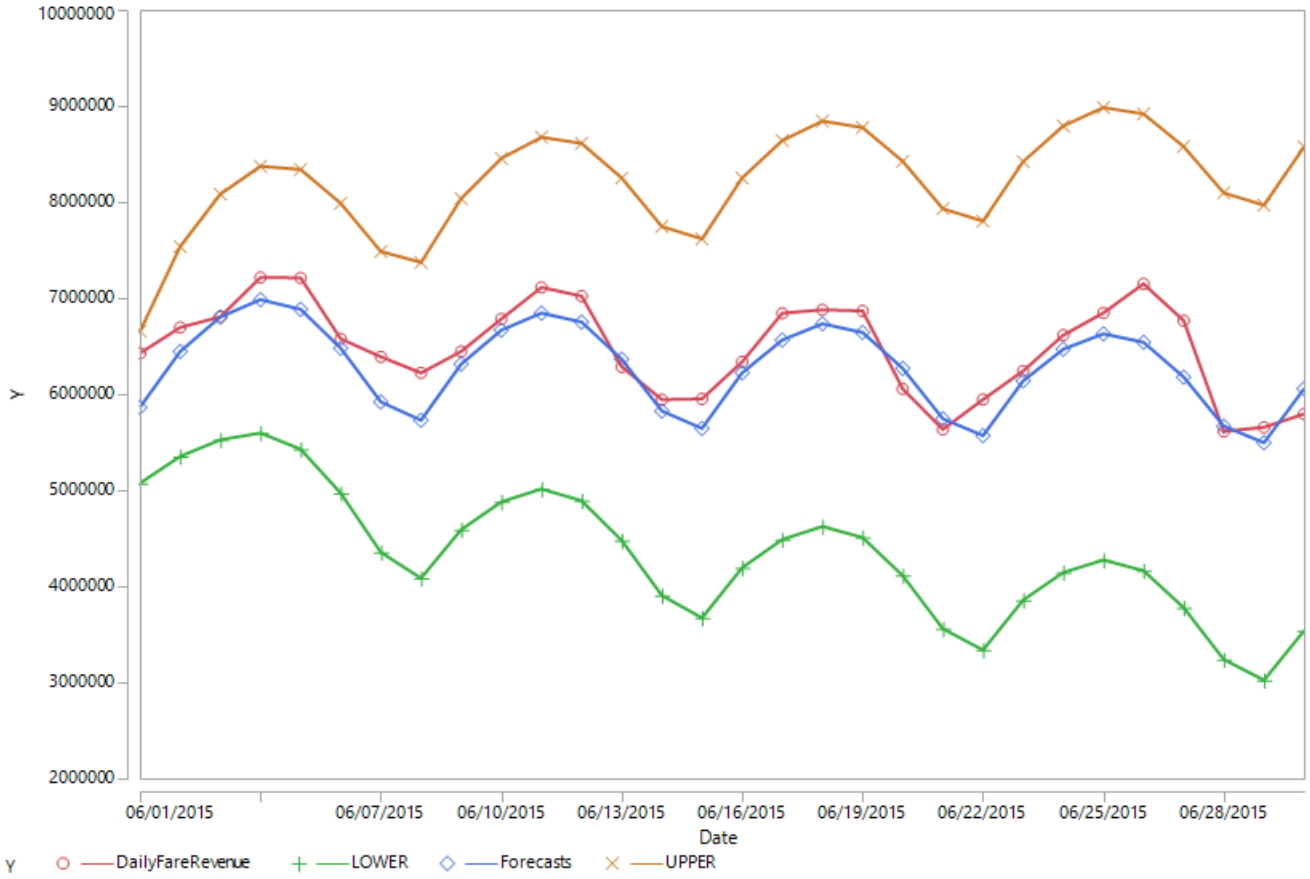
Daily taxi fare revenue in New York City – Forecasting View



Daily taxi fare revenue in New York City – Forecasting View
Click on the right most icon in line with “Active series”

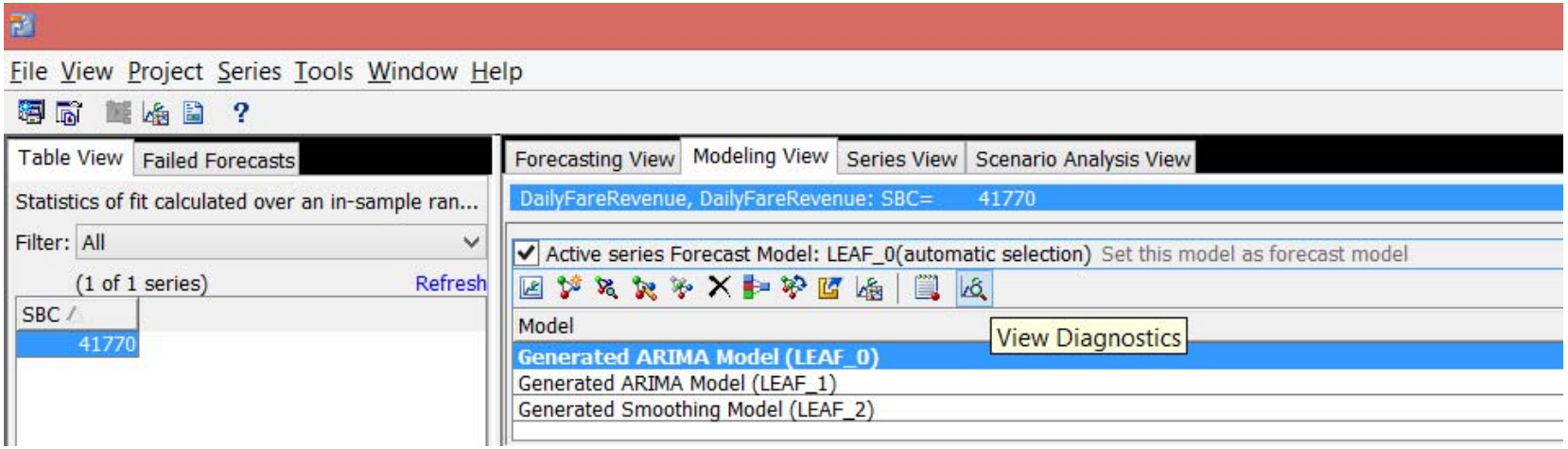


Daily taxi fare revenue in New York City – Actuals and Forecasts

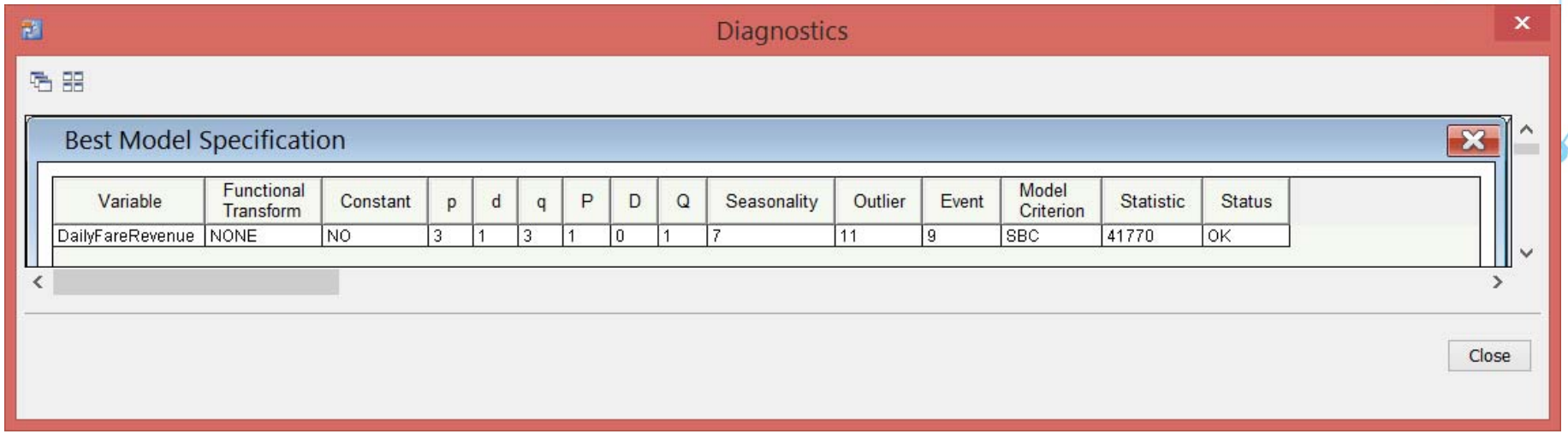


Daily taxi fare revenue in New York City – Modeling View

Under the Modeling View, click on the right most icon to “View Diagnostics”

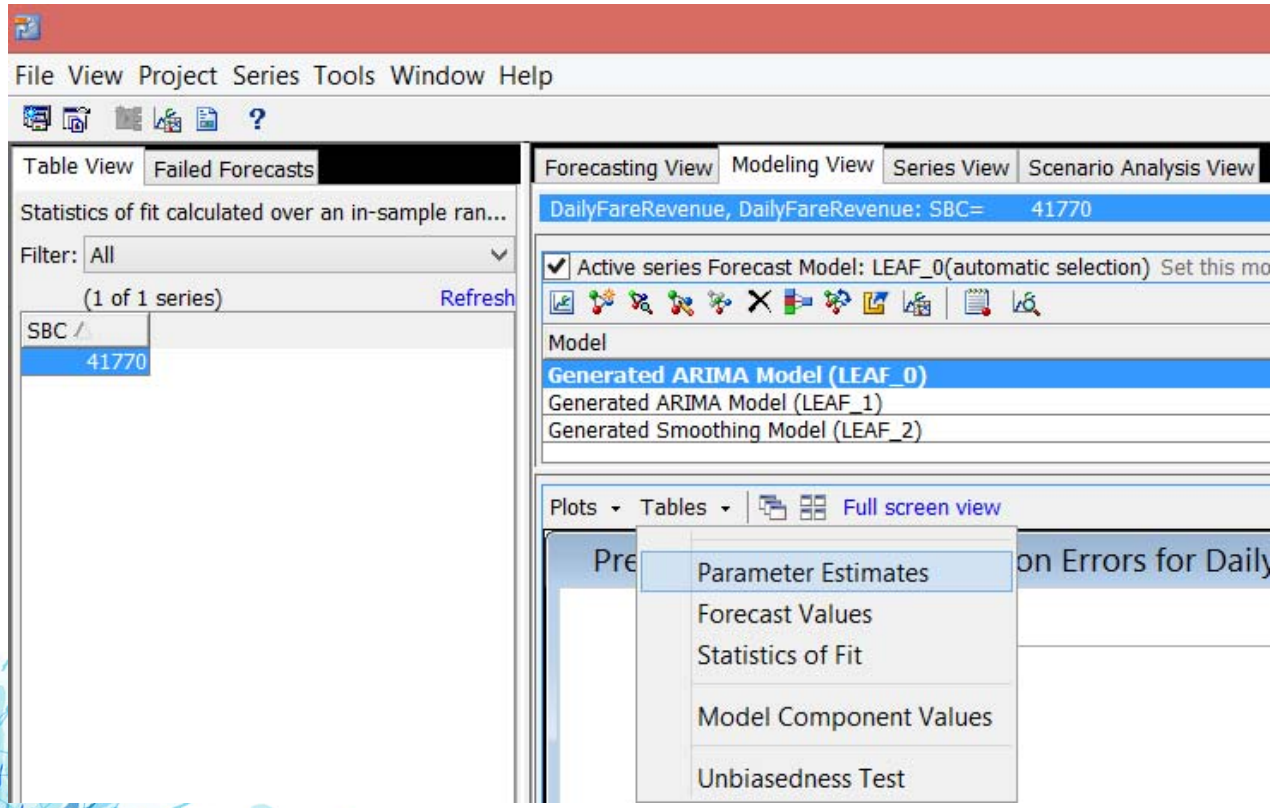


Daily taxi fare revenue in New York City – Diagnostics (partial output)



Variable	Functional Transform	Constant	p	d	q	P	D	Q	Seasonality	Outlier	Event	Model Criterion	Statistic	Status
DailyFareRevenue	NONE	NO	3	1	3	1	0	1	7	11	9	SBC	41770	OK

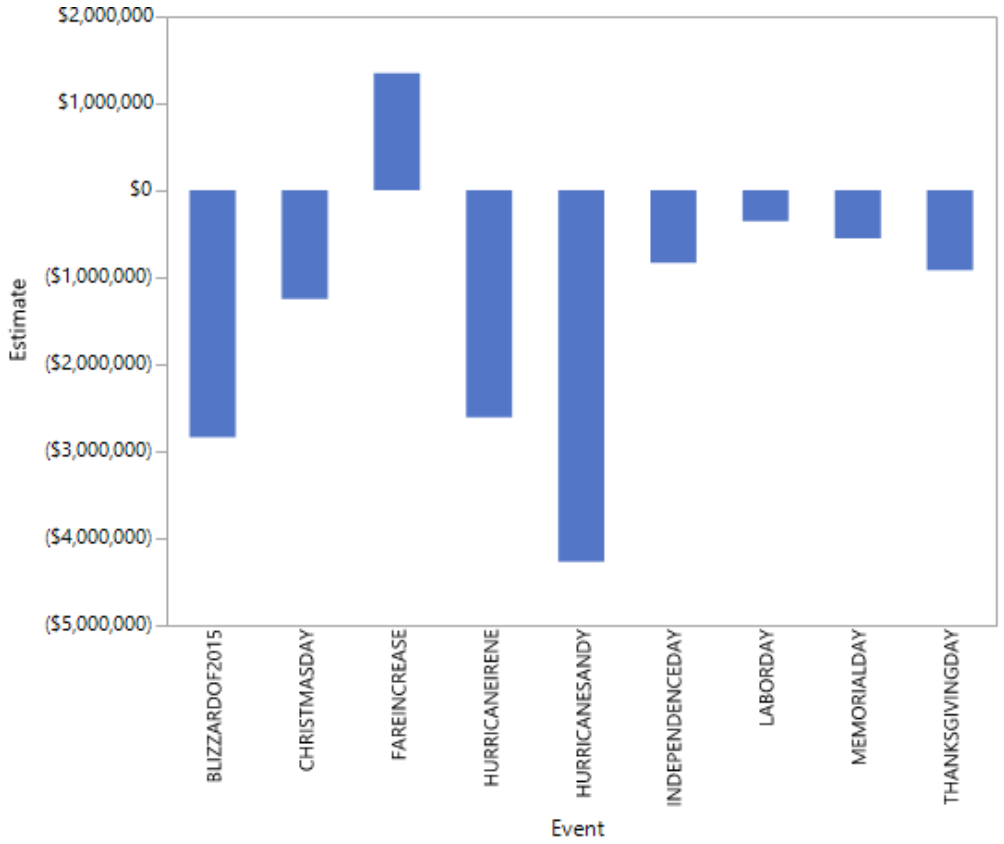
Under the Modeling View, click on Tables and then click on Parameter Estimates



Daily taxi fare revenue in New York City – Parameter Estimates

Component	Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|
DailyFareRevenue	MA1_1	2.06587	0.03711	55.67	<.0001
DailyFareRevenue	MA1_2	-1.93025	0.04888	-39.49	<.0001
DailyFareRevenue	MA1_3	0.79967	0.03001	26.65	<.0001
DailyFareRevenue	MA2_7	0.95528	0.01129	84.64	<.0001
DailyFareRevenue	AR1_1	2.01692	0.04306	46.84	<.0001
DailyFareRevenue	AR1_2	-1.94354	0.05280	-36.81	<.0001
DailyFareRevenue	AR1_3	0.77606	0.04019	19.31	<.0001
DailyFareRevenue	AR2_7	0.99561	0.0040845	243.76	<.0001
HURRICANEIRENE	SCALE	-2607042.8	290485.7	-8.97	<.0001
FAREINCREASE	SCALE	1350446.9	401937.7	3.36	0.0008
HURRICANESANDY	SCALE	-4268159.6	335583.5	-12.72	<.0001
BLIZZARDOF2015	SCALE	-2838596.2	288736.6	-9.83	<.0001
CHRISTMASDAY	SCALE	-1245637.5	143927.1	-8.65	<.0001
JULY4	SCALE	-833331.5	143795.7	-5.80	<.0001
LABORDAY	SCALE	-351219.0	153208.0	-2.29	0.0220
MEMORIALDAY	SCALE	-552486.0	129617.8	-4.26	<.0001
THANKSGIVINGDAY	SCALE	-915862.3	144265.6	-6.35	<.0001
LS01JAN2015D	SCALE	2498994.8	390279.1	6.40	<.0001
AO13FEB2014D	SCALE	-2890060.9	287751.8	-10.04	<.0001
LS05AUG2013D	SCALE	2537055.4	384968.4	6.59	<.0001
AO11AUG2013D	SCALE	-2921294.4	287877.6	-10.15	<.0001
LS27AUG2011D	SCALE	-2379733.4	380531.1	-6.25	<.0001
LS04APR2011D	SCALE	2673861.4	385258.8	6.94	<.0001
LS01APR2011D	SCALE	-3183052.7	385269.2	-8.26	<.0001
AO18JAN2015D	SCALE	4211287.5	288236.3	14.61	<.0001
LS01DEC2014D	SCALE	3668954.9	386812.7	9.49	<.0001
AO30OCT2012D	SCALE	-3015633.0	335561.0	-8.99	<.0001
AO01JAN2012D	SCALE	-1774041.6	287781.6	-6.16	<.0001

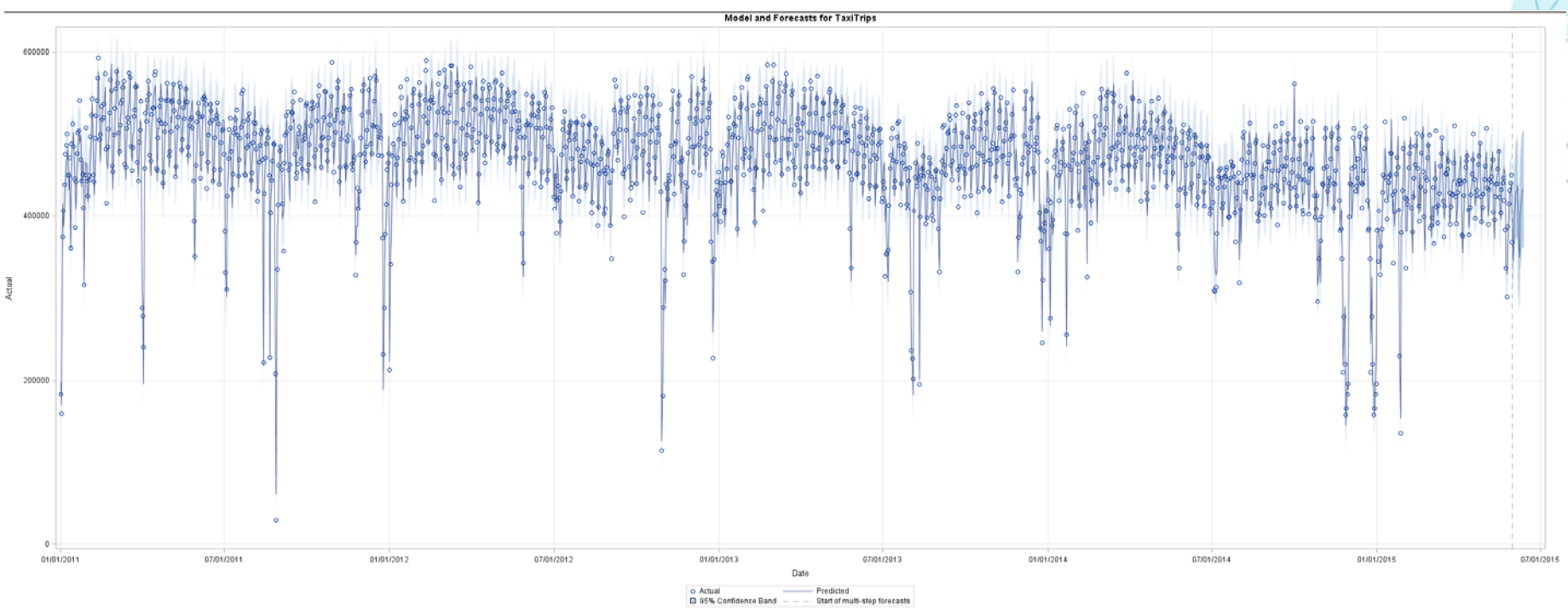
Daily taxi fare revenue in New York City – Parameter Estimates



Part 2. Daily taxi trips in New York City

- We wish to develop a time series model for the daily number of taxi trips by Yellow Cabs in New York City using publicly available data from 1/1/2011 until 5/31/2015 in order to both
- Predict daily taxi trips for June 2015
- Estimate the impact of events including
  - Price increase
  - Holidays
  - One of events such as hurricanes

Daily taxi trips in New York City – Forecasting View



Daily taxi trips in New York City – Modeling View

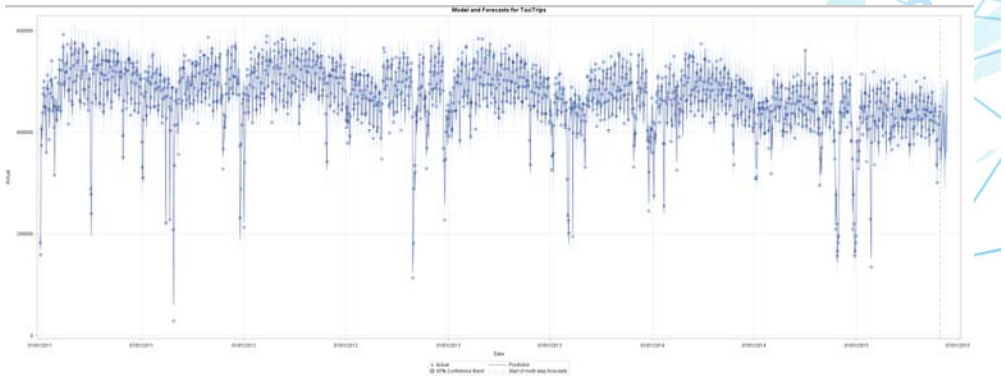
Component	Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|
TaxiTrips	CONSTANT	193950.5	20145.4	9.63	<.0001
TaxiTrips	MA1_7	0.66573	0.02049	32.50	<.0001
TaxiTrips	AR1_1	0.83603	0.01454	57.49	<.0001
TaxiTrips	AR2_7	0.98939	0.0046384	213.31	<.0001
HURRICANEIRENE	SCALE	-334459.4	17734.8	-18.86	<.0001
HURRICANESANDY	SCALE	-239077.6	17394.9	-13.74	<.0001
BLIZZARDOF2015	SCALE	-184863.2	15136.8	-12.21	<.0001
BOXINGDAY	SCALE	-44047.0	9367.8	-4.70	<.0001
CHRISTMASDAY	SCALE	-125750.3	8953.2	-14.05	<.0001
INDEPENDENCEDAY	SCALE	-61274.3	7569.2	-8.10	<.0001
LABORDAY	SCALE	-40070.9	7582.9	-5.28	<.0001
MEMORIALDAY	SCALE	-57867.3	6879.1	-8.41	<.0001
NEWYEAR	SCALE	1421.8	8514.0	0.17	0.8674
THANKSGIVINGDAY	SCALE	-52018.4	8391.1	-6.20	<.0001

Best Model Specification

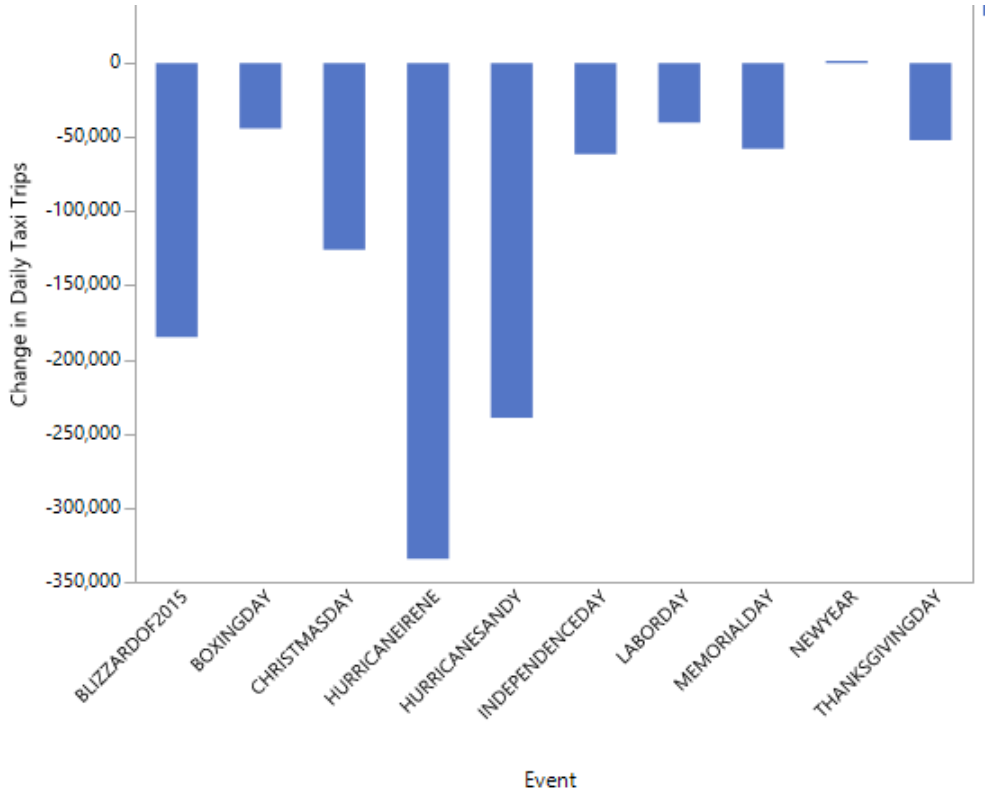
Variable	Functional Transform	Constant	p	d	q	P	D	Q	Seasonality	Outlier	Event	Model Criterion	Statistic	Status
TaxiTrips	NONE	YES	1	0	0	1	0	1	7	32	10	SBC	32454	OK

Daily taxi trips in New York City – Modeling View

Component	Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|
AO17FEB2013D	SCALE	67374.5	15201.2	4.43	<.0001
LS24NOV2011D	SCALE	-81926.9	19574.7	-4.19	<.0001
AO27JAN2011D	SCALE	-118600.7	15150.4	-7.83	<.0001
LS01APR2011D	SCALE	-261725.4	18989.1	-13.78	<.0001
LS04NOV2012D	SCALE	75519.6	18724.9	4.03	<.0001
LS27DEC2014D	SCALE	-138905.3	20623.9	-6.74	<.0001
LS05AUG2013D	SCALE	209675.5	19432.7	10.79	<.0001
AO05MAR2015D	SCALE	-77884.7	15127.3	-5.15	<.0001
LS24DEC2011D	SCALE	-79525.7	18588.7	-4.28	<.0001
AO01JAN2012D	SCALE	-185341.6	17332.3	-10.69	<.0001
LS01AUG2013D	SCALE	-150976.6	19421.5	-7.77	<.0001
AO11AUG2013D	SCALE	-204879.4	15229.3	-13.45	<.0001
AO08FEB2013D	SCALE	-119024.0	17396.2	-6.84	<.0001
LS01JAN2015D	SCALE	170144.1	20114.2	8.46	<.0001
AO09FEB2013D	SCALE	-160075.2	17454.5	-9.17	<.0001
AO14AUG2011D	SCALE	-196215.0	15508.8	-12.65	<.0001
AO13FEB2014D	SCALE	-152227.5	15134.5	-10.06	<.0001
LS03JAN2011D	SCALE	248948.7	19097.1	13.04	<.0001
LS02AUG2013D	SCALE	-71066.4	19475.4	-3.65	0.0003
LS04APR2011D	SCALE	245610.4	18997.8	12.93	<.0001
AO21JAN2013D	SCALE	-73075.4	15137.7	-4.83	<.0001
LS22NOV2012D	SCALE	-83898.7	19697.0	-4.26	<.0001
LS27OCT2014D	SCALE	-112633.7	18593.3	-6.06	<.0001
AO30OCT2012D	SCALE	-156026.7	17391.5	-8.97	<.0001
AO21JAN2014D	SCALE	-145163.4	15131.8	-9.59	<.0001
AO03JAN2014D	SCALE	-123591.6	15149.8	-8.16	<.0001
AO21AUG2011D	SCALE	-169790.8	15637.6	-10.86	<.0001
LS24NOV2014D	SCALE	-149887.9	18693.5	-8.02	<.0001
AO27AUG2011D	SCALE	-237780.3	17394.2	-13.67	<.0001
AO12JAN2011D	SCALE	-113624.4	15149.7	-7.50	<.0001
LS01DEC2014D	SCALE	205424.4	18672.3	11.00	<.0001
AO01AUG2014D	SCALE	-104115.9	15137.4	-6.88	<.0001



Daily taxi trips in New York City – Modeling View

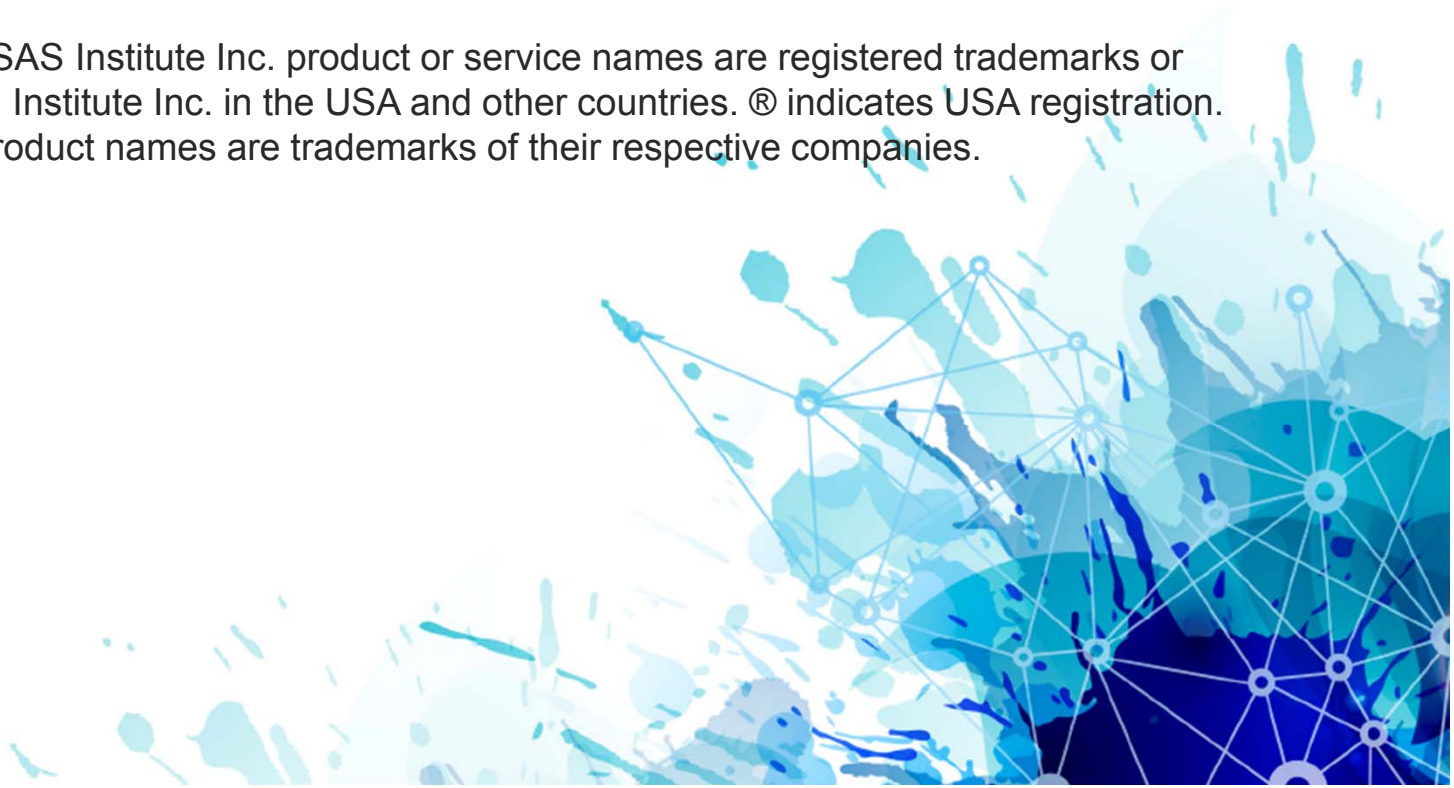


SAS® Forecast Studio 13.2

- Automatically fits time series models which can at the same time
- Detect and allow for outliers and level shifts
- Determine which model(s) optimize the chosen fit criterion

SAS® GLOBAL FORUM 2016

IMAGINE. CREATE. INNOVATE.



#SASGF