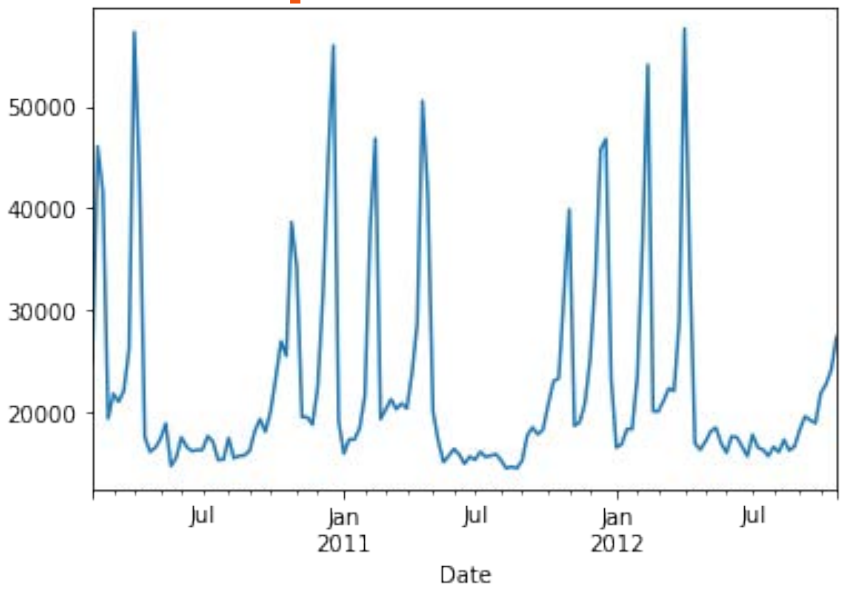
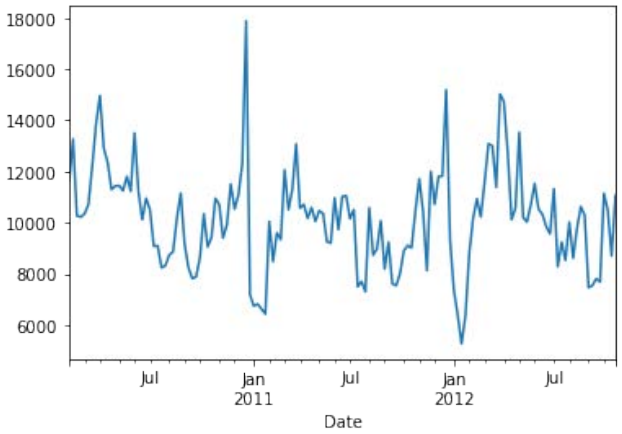
Forecasting Intro

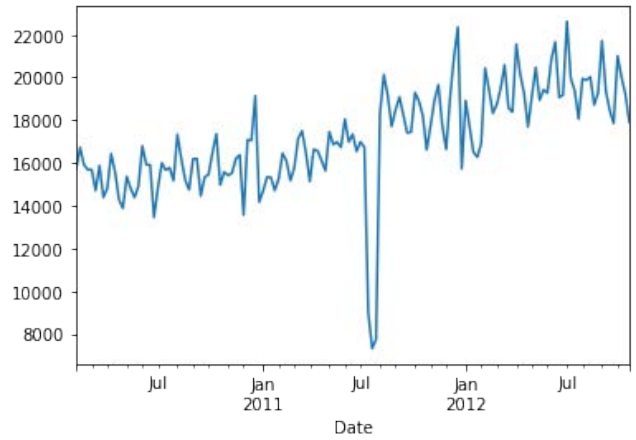
School of Information Studies
Syracuse University

Patterns in Temporal Data

- Obtain
- Scrub
- Explore
- Model
- iNterpret

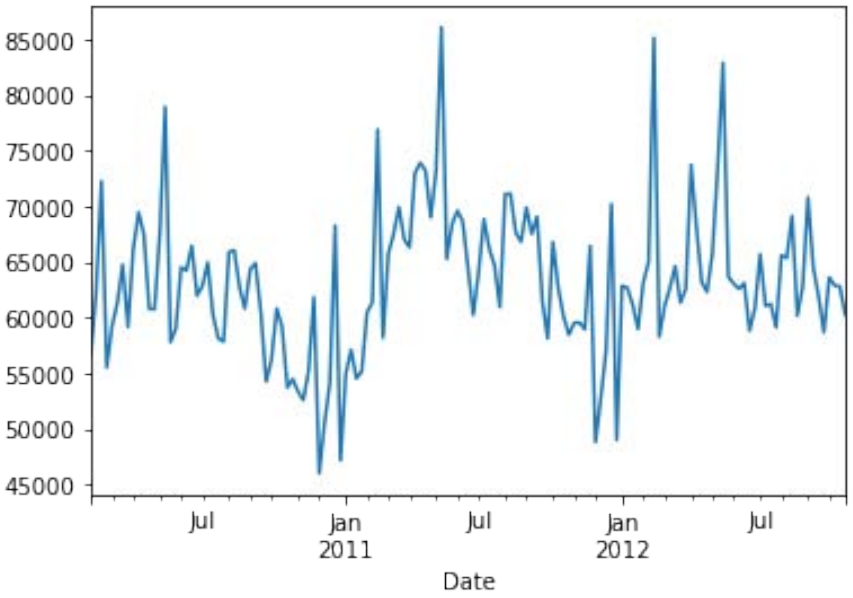




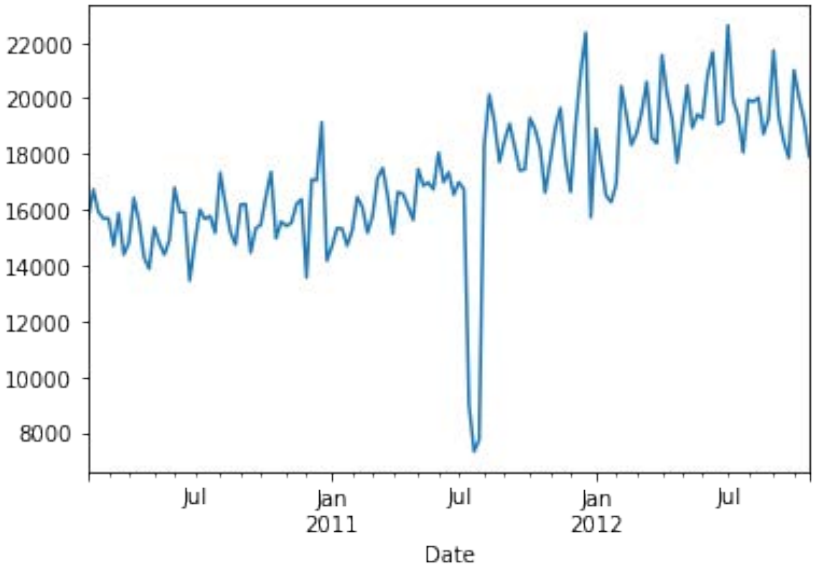


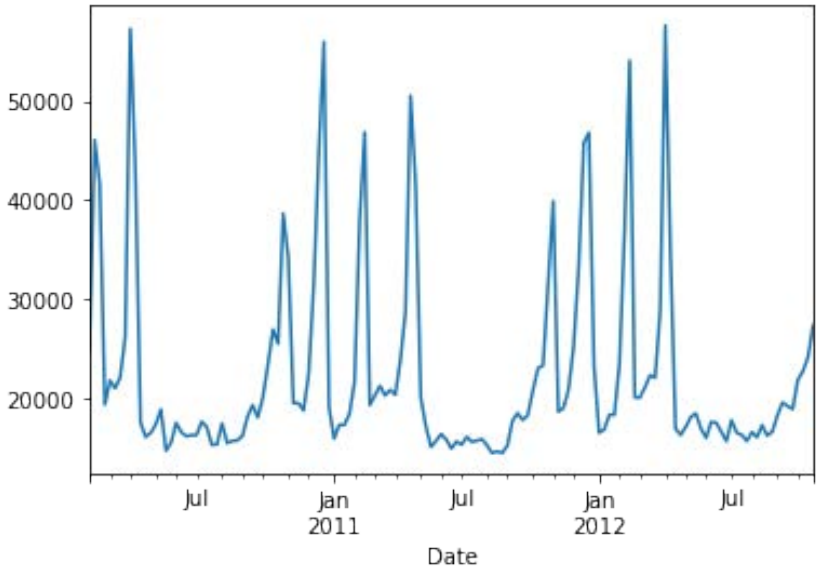
School of Information Studies
Syracuse University

Our Challenge This Week?

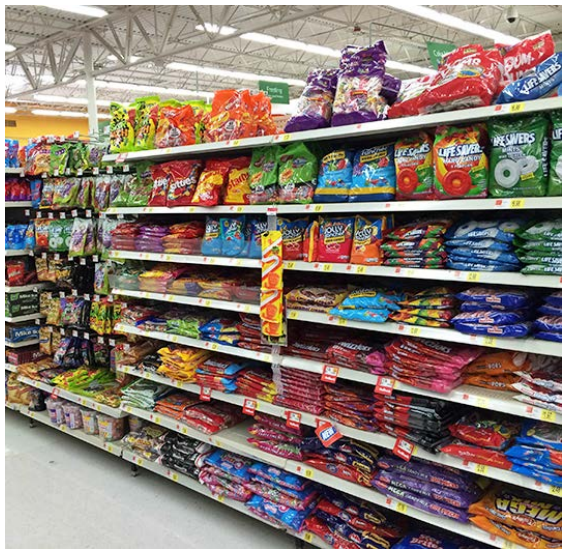


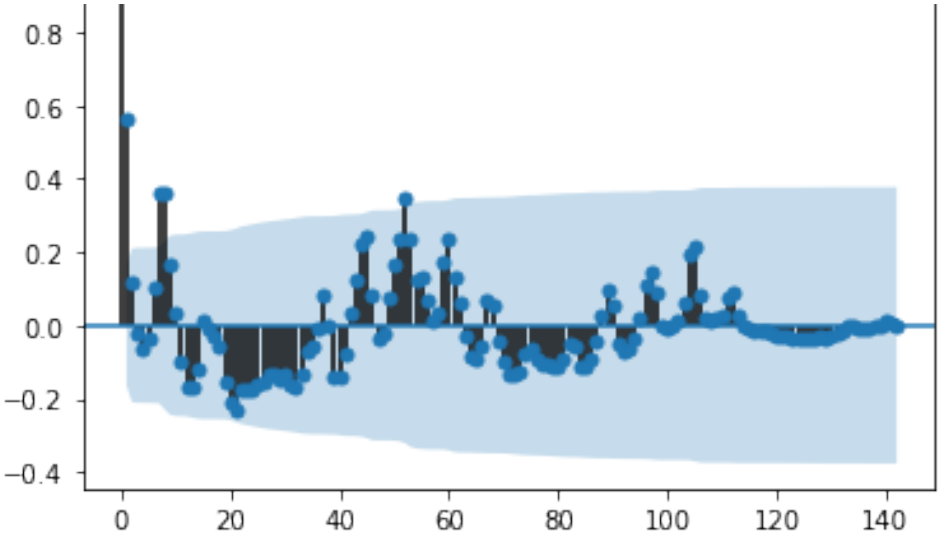
Will Sales Rise or Fall?

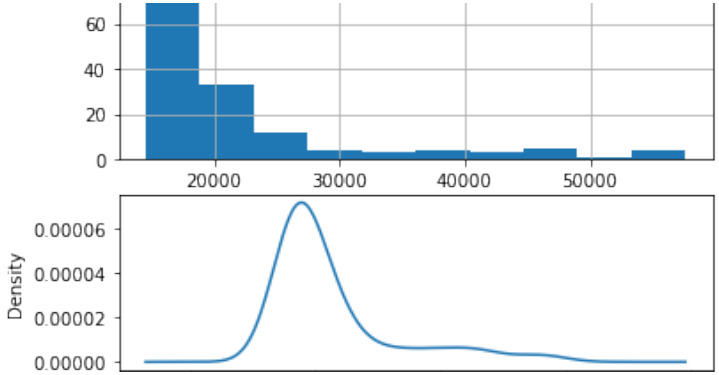




How Might We Know?





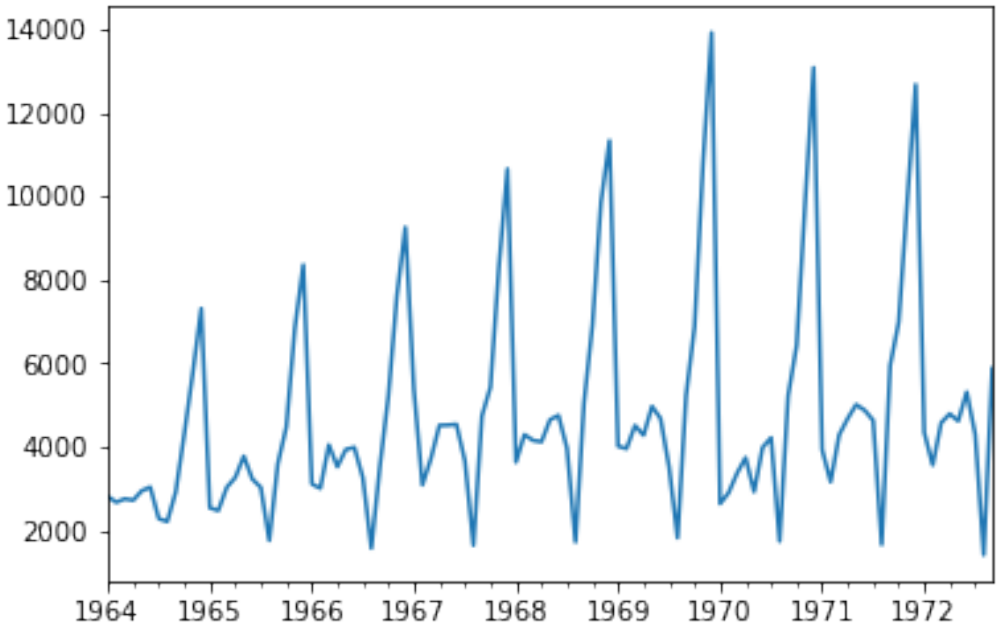


Date Review

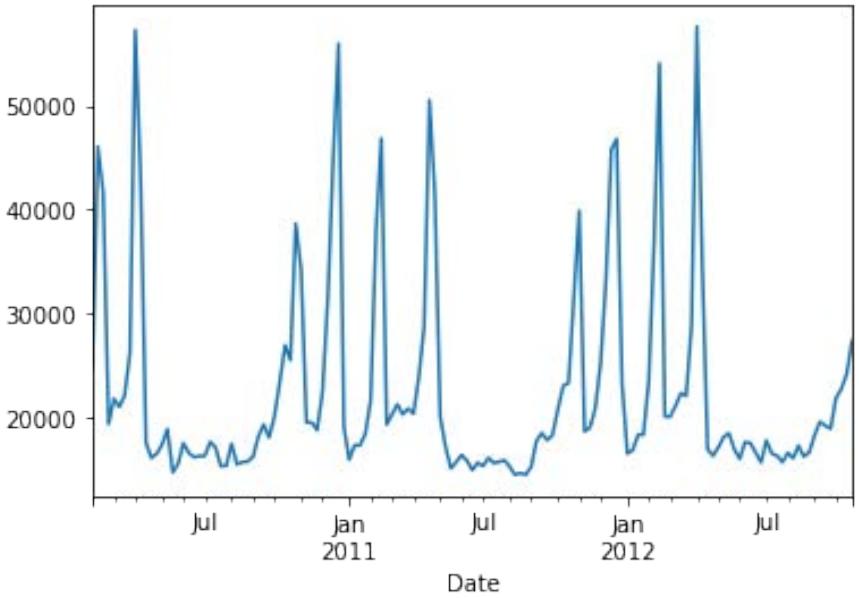
School of Information Studies
Syracuse University

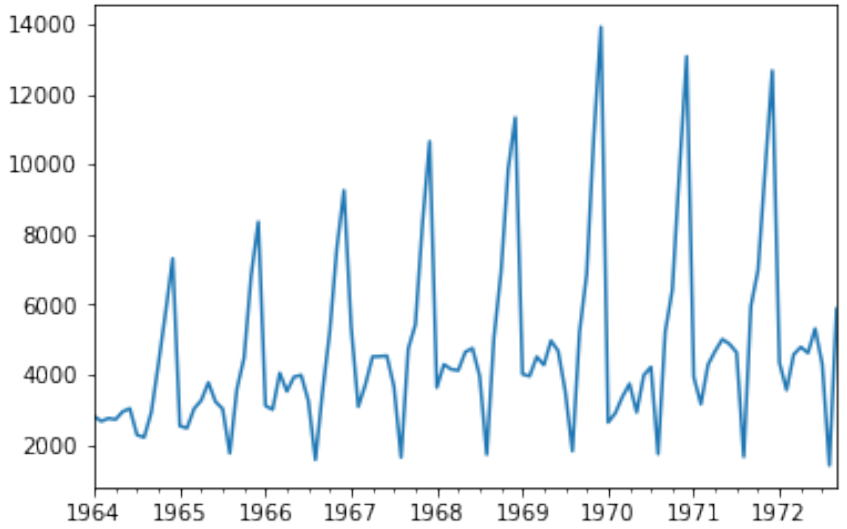
Temporal Data

1964-01	2815	1965-09	3595
1964-02	2672	1965-10	4474
1964-03	2755	1965-11	6838
1964-04	2721	1965-12	8357
1964-05	2946	1966-01	3113
1964-06	3036	1966-02	3006
1964-07	2282	1966-03	4047
1964-08	2212	1966-04	3523
1964-09	2922	1966-05	3937
1964-10	4301	1966-06	3986
1964-11	5764	1966-07	3260
1964-12	7312	1966-08	1573
1965-01	2541	1966-09	3528
1965-02	2475	1966-10	5211
1965-03	3031	1966-11	7614
1965-04	3266	1966-12	9254
1965-05	3776	1967-01	5375
1965-06	3230	1967-02	3088
1965-07	3028	1967-03	3718
1965-08	1759	1967-04	4514

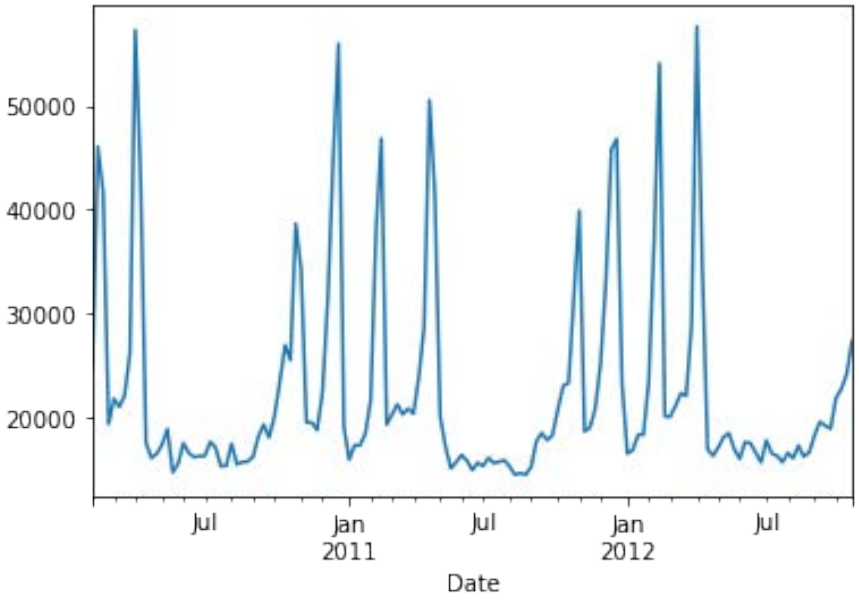


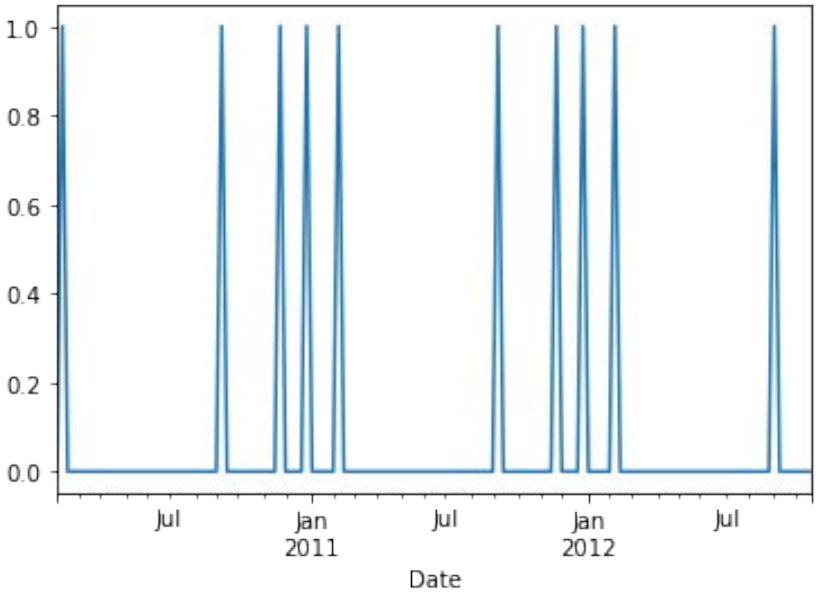
Seasonal Data



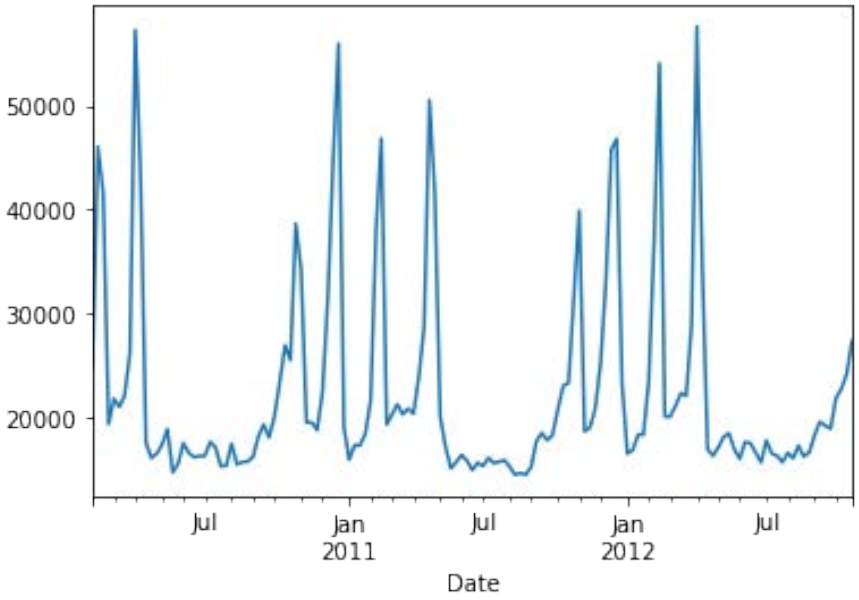


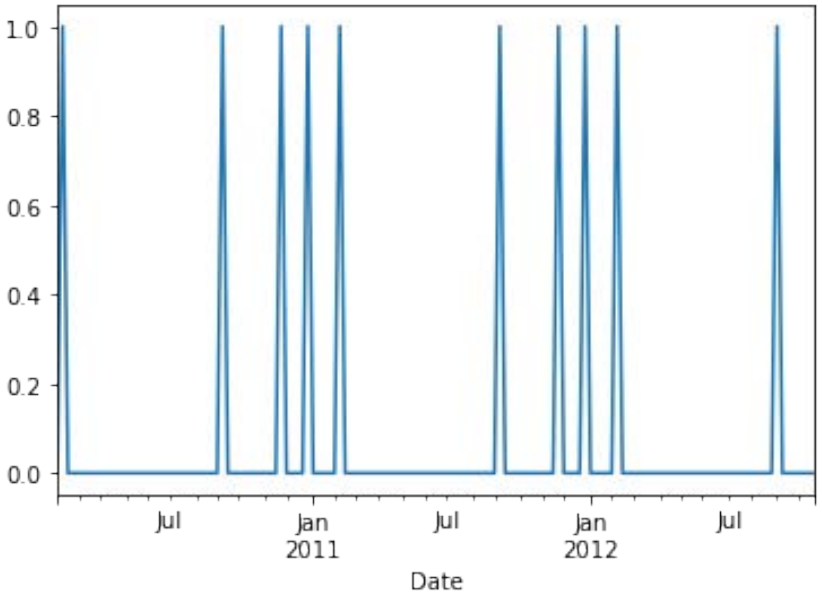
In Sample – Out of Sample



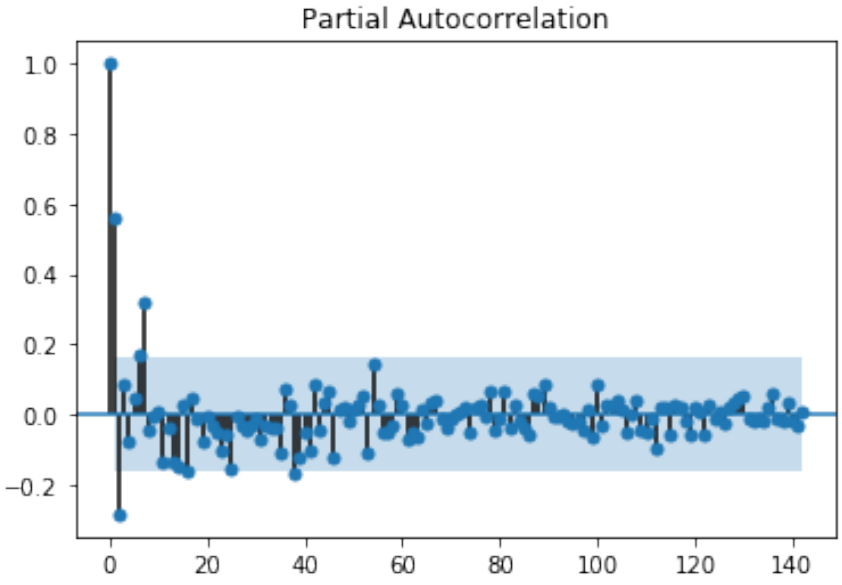


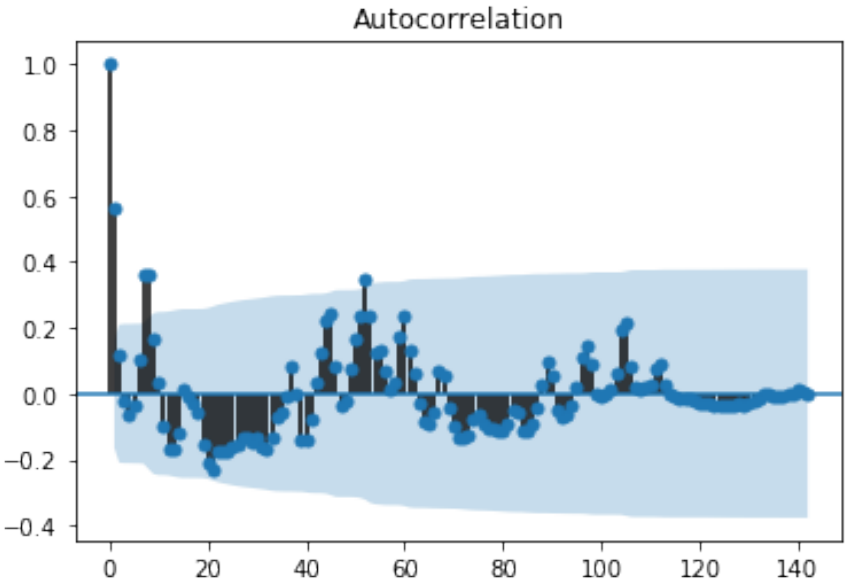
In Sample – Out of Sample (cont.)





Correlation

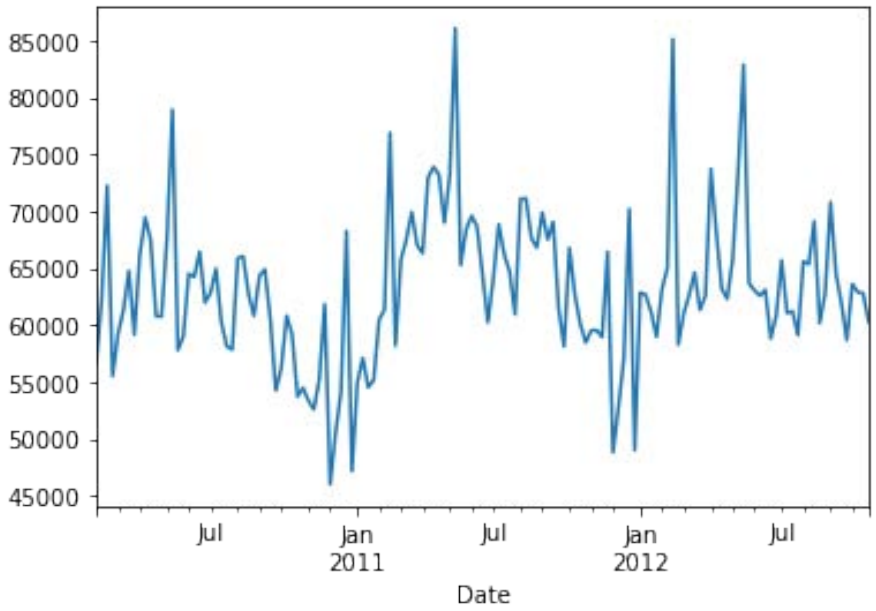




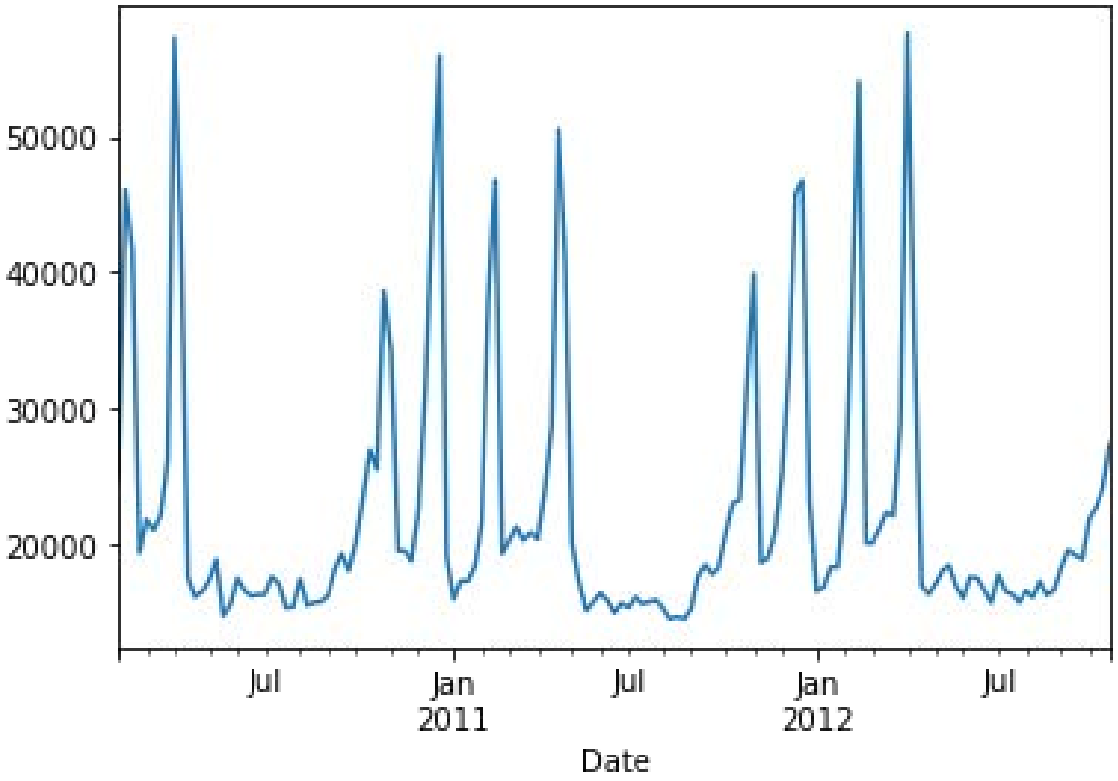
Recommendation

School of Information Studies
Syracuse University

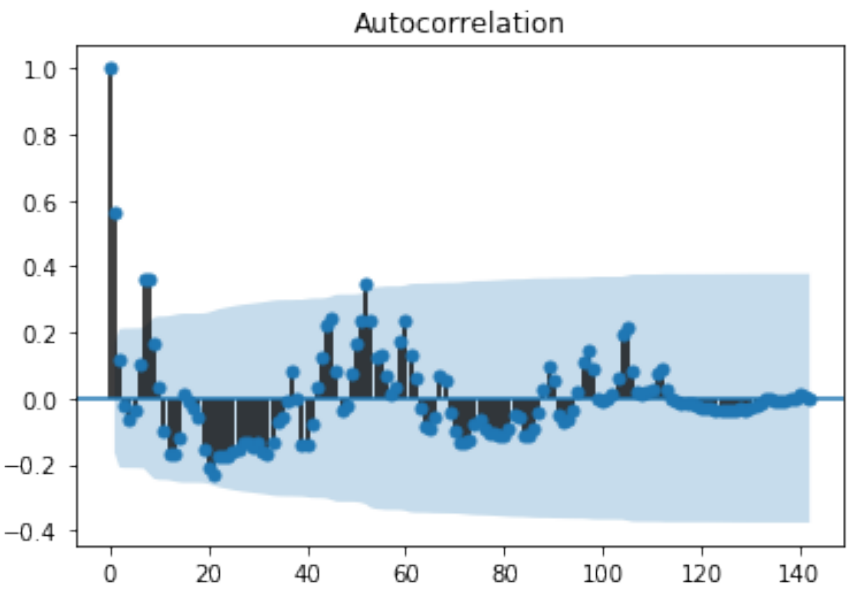
Our Challenge This Week?

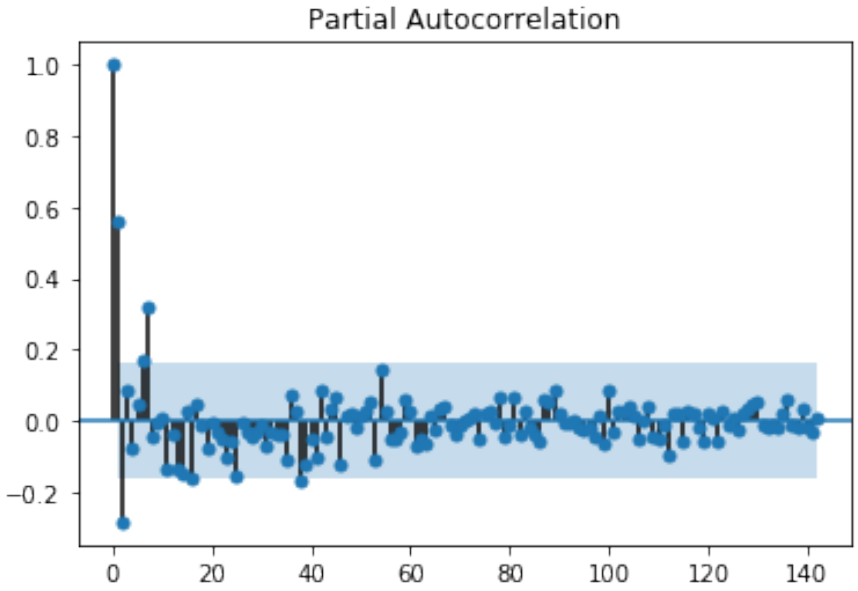


Will Sales Rise or Fall?



How Might We Know?





Stationary

ADF Statistic:

-8.892005

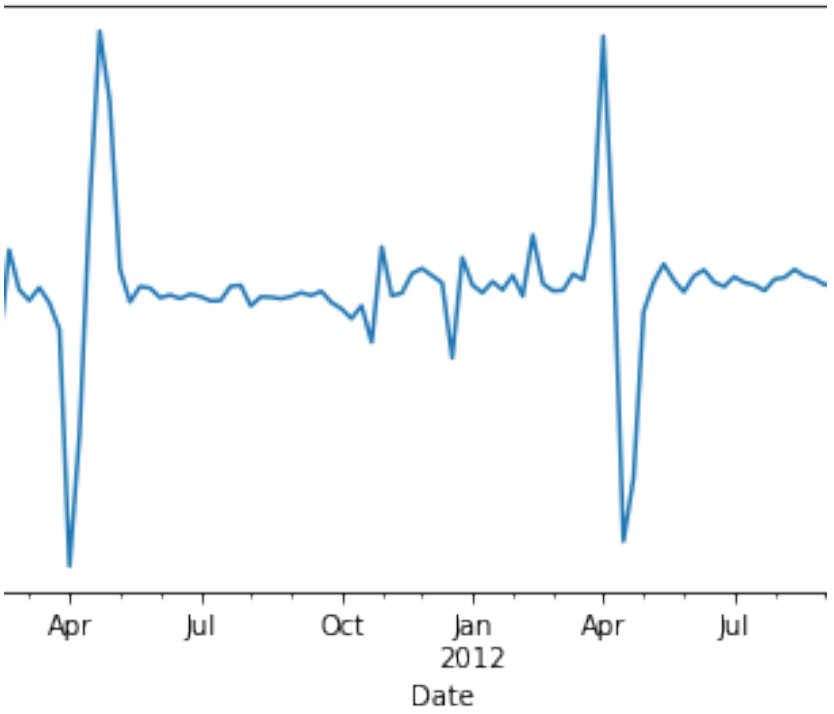
p-value: 0.000000

Critical Values:

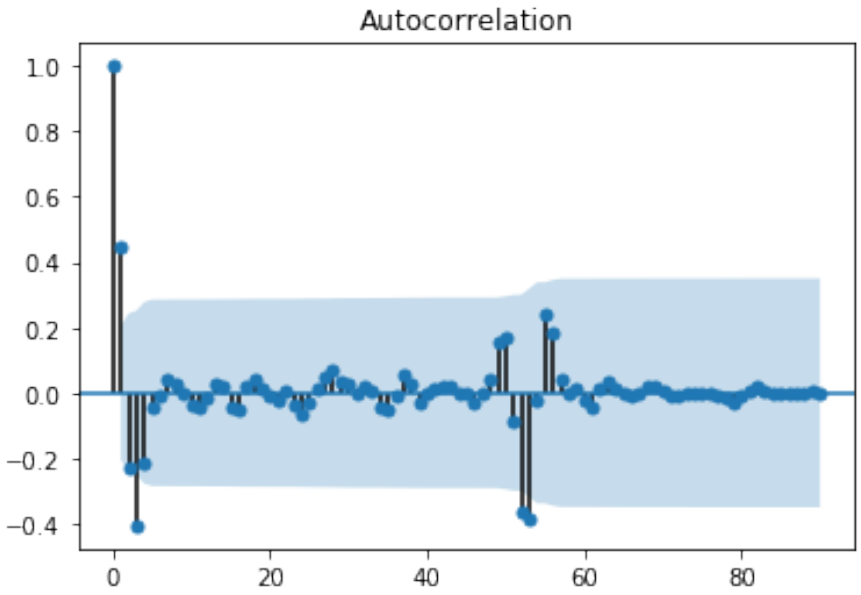
1%: -3.506

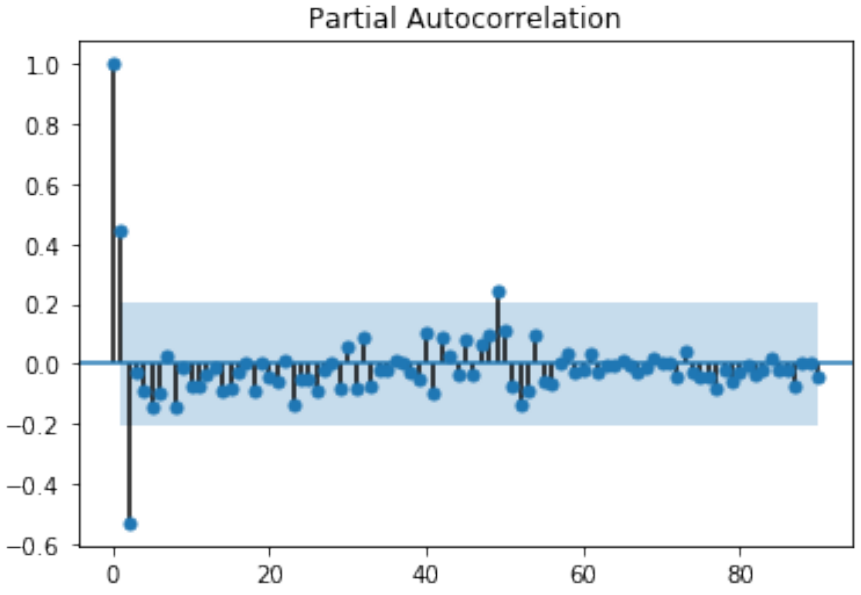
5%: -2.895

10%: -2.584



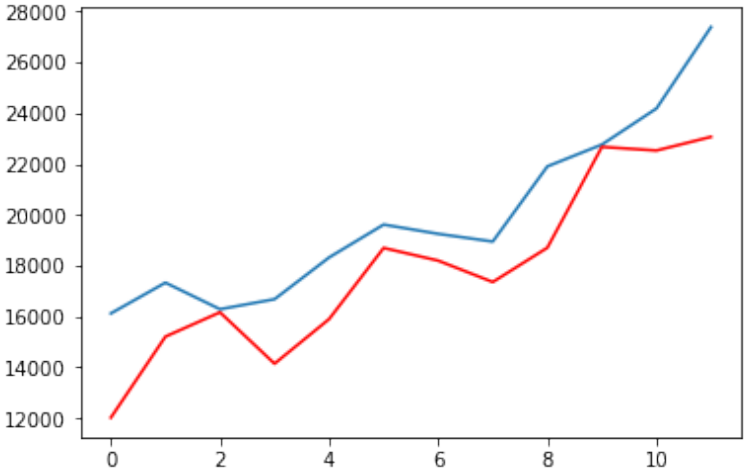
Modeling Input





Recommendation

- Increasing sales in next quarter
- Opportunity to close forecast gap
- Consider data transformations
- Consider methods to account for holiday



Forecasting

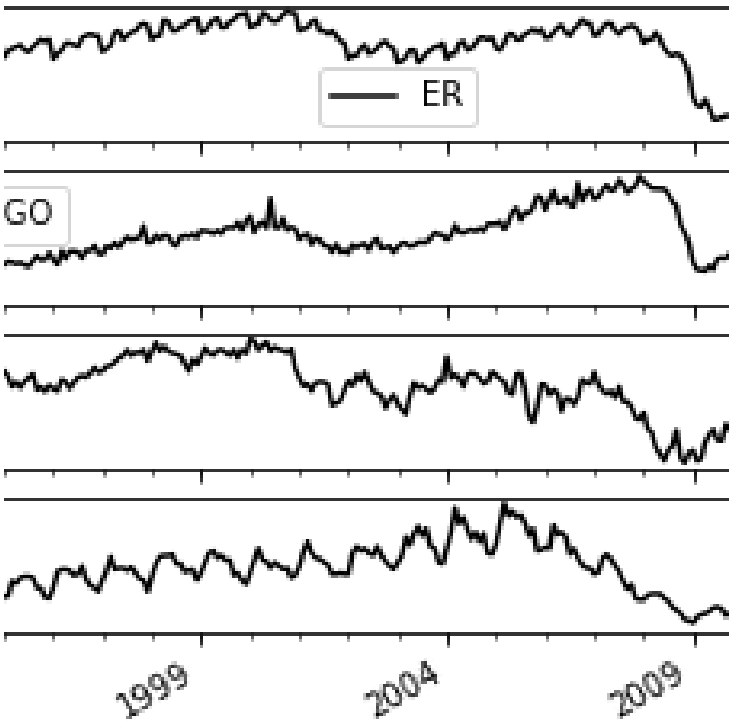
School of Information Studies
Syracuse University

Patterns in Temporal Data

	10 Feb 2012 12:21	Missed Call	077	Hel
	10 Feb 2012 12:29	2 Seconds	077	Hel
	10 Feb 2012 13:47	1 Minute	0779	Hel
	10 Feb 2012 15:05	2 Minutes	019	
	10 Feb 2012 16:48	1 Minute	077	Hel
	10 Feb 2012 16:53	7 Seconds	077	Hel
	10 Feb 2012 16:54	39 Seconds	077	Hel
	10 Feb 2012 17:13	3 Seconds	077	Hel
	10 Feb 2012 17:13	2 Seconds	0779	Hel
	10 Feb 2012 17:14	2 Seconds	0779	Hel
	10 Feb 2012 17:16	2 Seconds	0779	Hel
	10 Feb 2012 17:20	2 Seconds	0779	Hel
	10 Feb 2012 17:23	2 Seconds	0779	Hel
	10 Feb 2012 17:25	10 Seconds	0779	Hel
	10 Feb 2012 20:07	11 Seconds	077	Hel
	11 Feb 2012 09:41	1 Minute	077	Hel
	12 Feb 2012 10:54	1 Minute	077	Hel
	13 Feb 2012 12:50	2 Minutes	077	Hel
	13 Feb 2012 14:52	38 Seconds	077	Hel
	14 Feb 2012 09:30	2 Minutes	0779	Hel
	14 Feb 2012 18:19	49 Seconds	0142	Orie
	15 Feb 2012 17:14	1 Minute	0779	Hel
	15 Feb 2012 17:18	Cancelled	0779	Hel
	15 Feb 2012 17:18	1 Minute	0779	Hel
	15 Feb 2012 17:26	Missed Call	0779	Hel
	15 Feb 2012 17:27	5 Seconds	0779	Hel

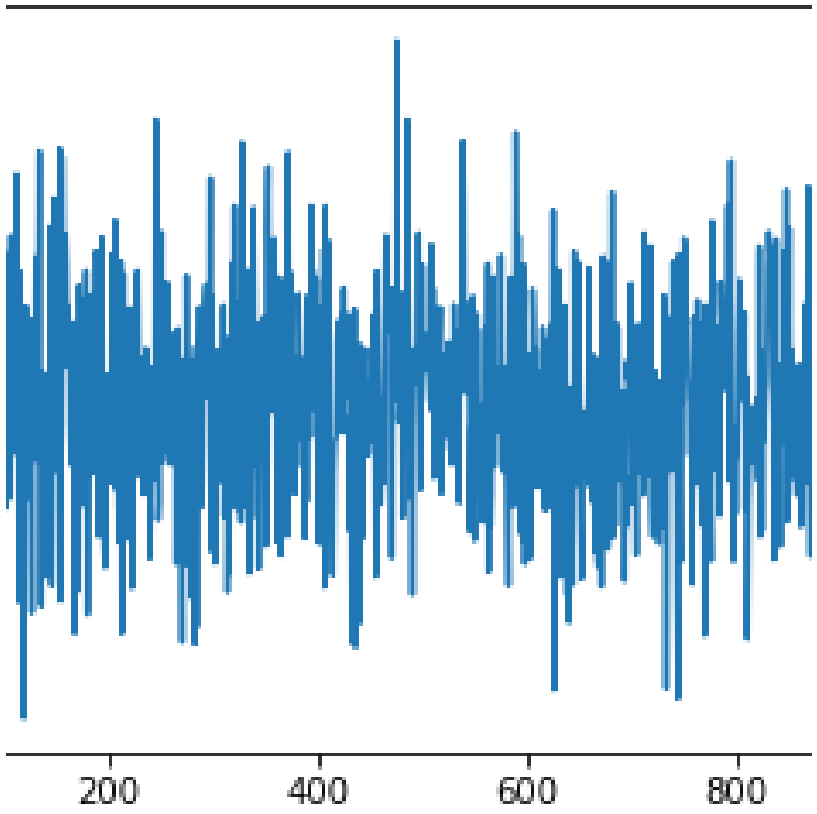
Forecasting

- White noise
- Covariance
- Seasonality
- Stationary

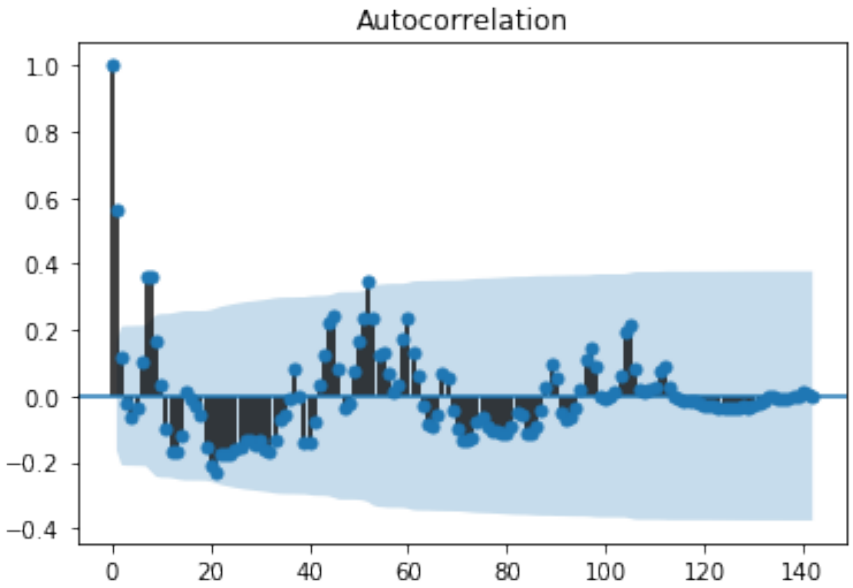


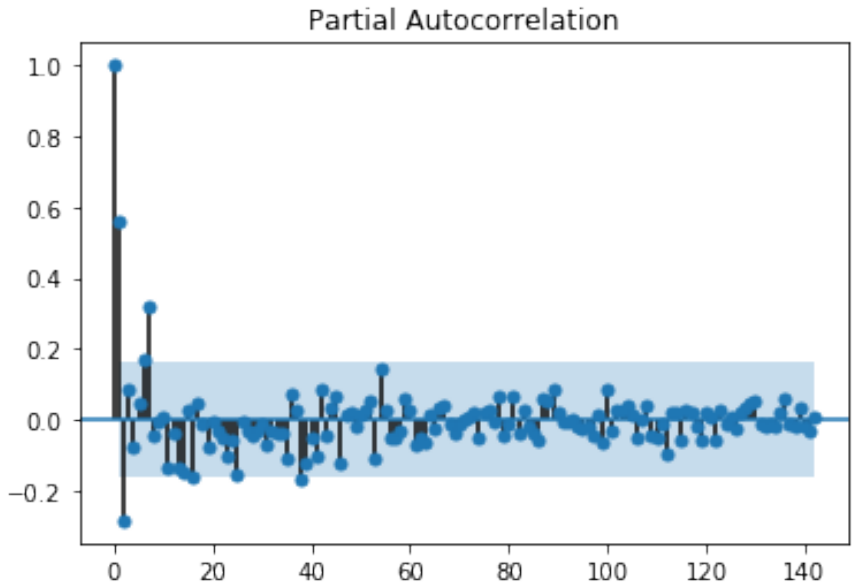
White Noise

- A random process
- Independent
- Identically distributed
- Mean = zero
- Variance = σ^2
- Correlation = zero



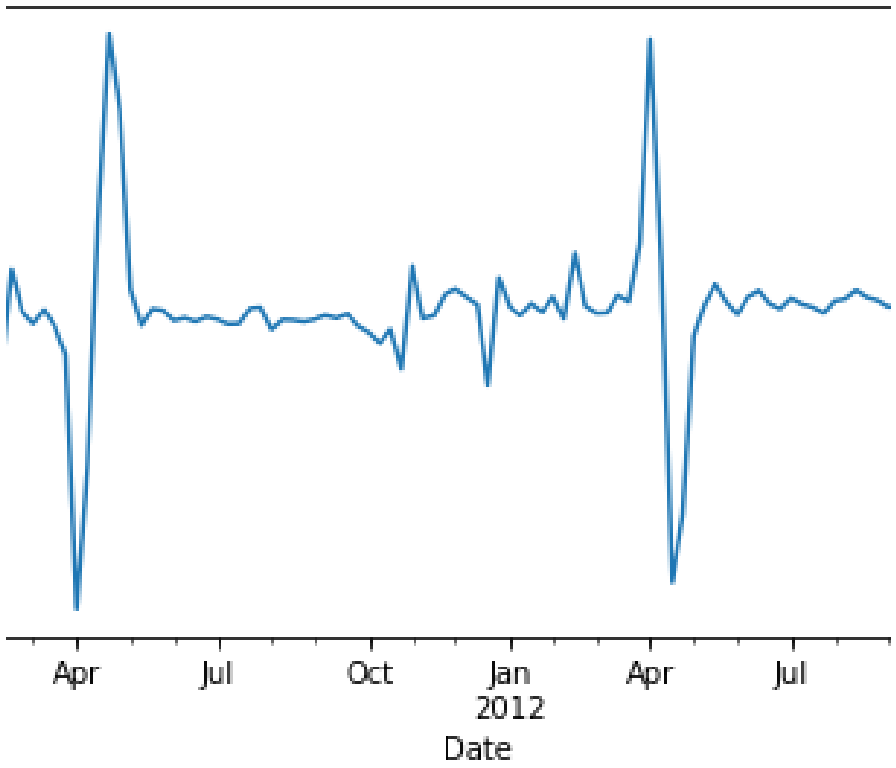
Covariance


Autocorrelation
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
0
20
40
60
80
100
120
140

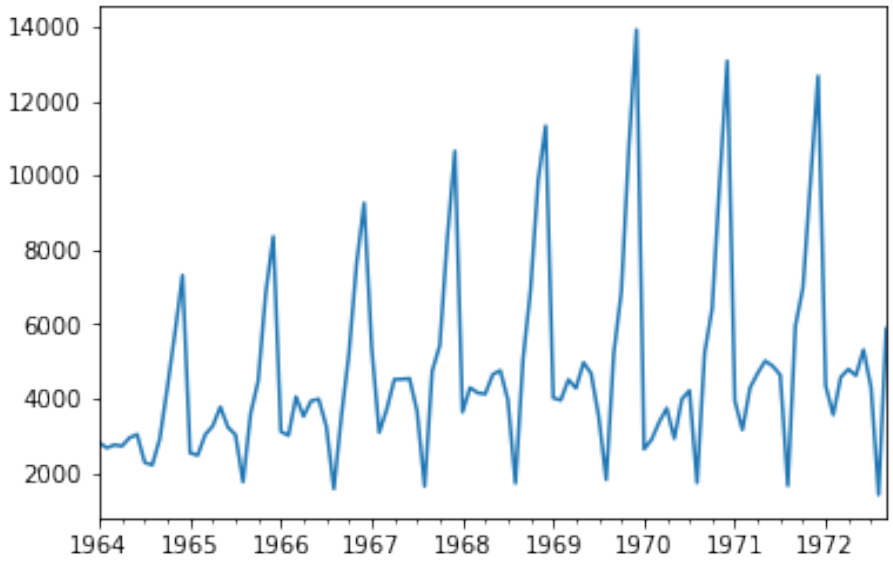

Partial Autocorrelation
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0
20
40
60
80
100
120
140

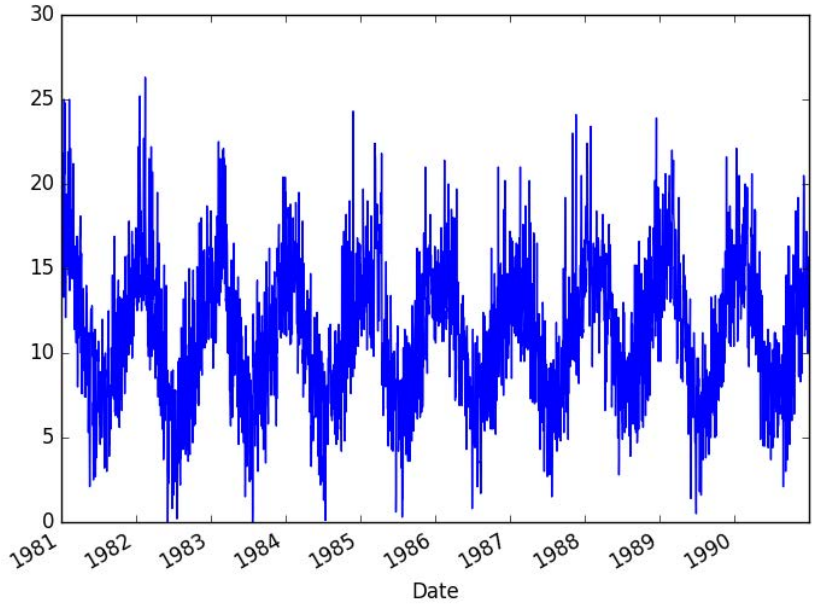
Stationary

- Process does not vary
- Mean stable over time
- Variance stable over time
- White noise?
  - Yes
- Seasonality?
  - No



Seasonality



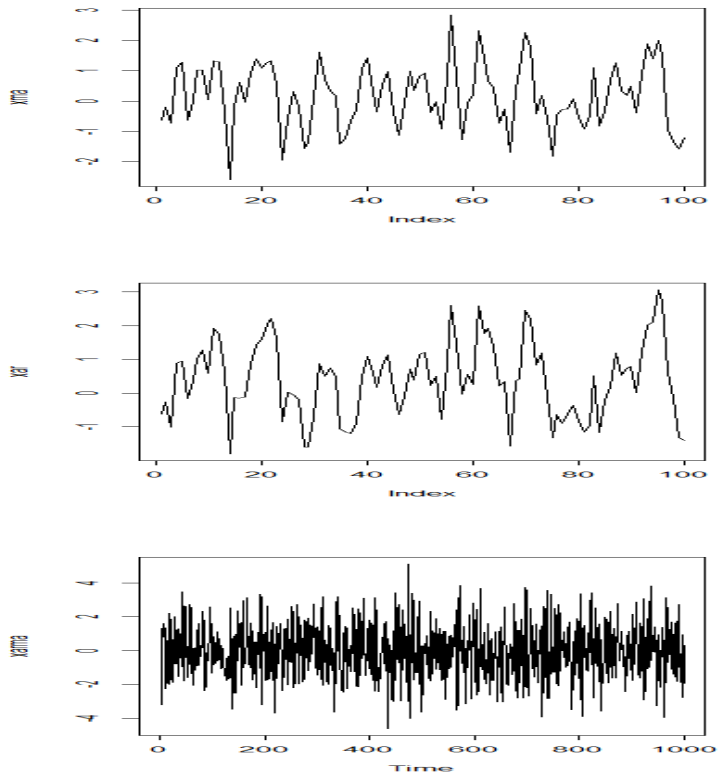


ARIMA

School of Information Studies
Syracuse University

Dynamics of a Time Series

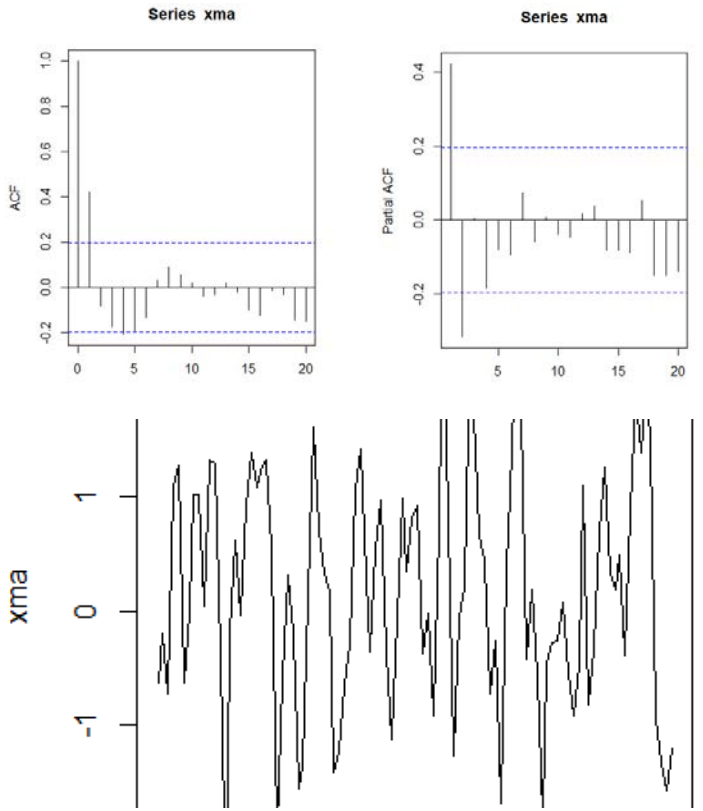
- Moving Average
- Autoregressive
- ARMA



Moving Average Process

$$Y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1}$$

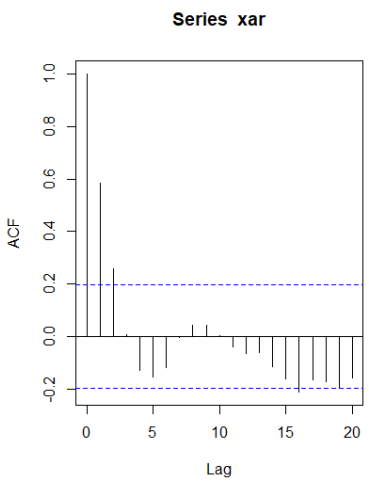
- Weighted sum of most recent
- ACF – contains spikes with rest zero
- Covariance stationary
  - Mean and autocovariance are not functions of time

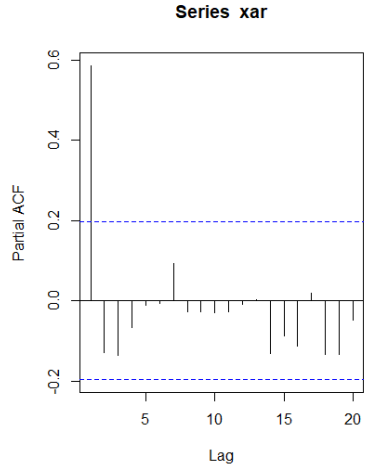


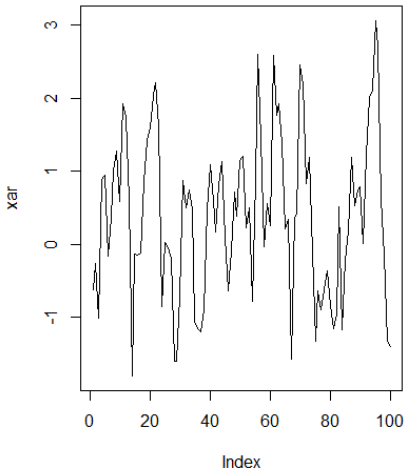
Autoregressive Process

$$Y_t = c + \varphi Y_{t-1} + \varepsilon_t$$

- Dependent on previous values
- ACF – decay to zero
- PACF – identify order at decay
- Stationary?



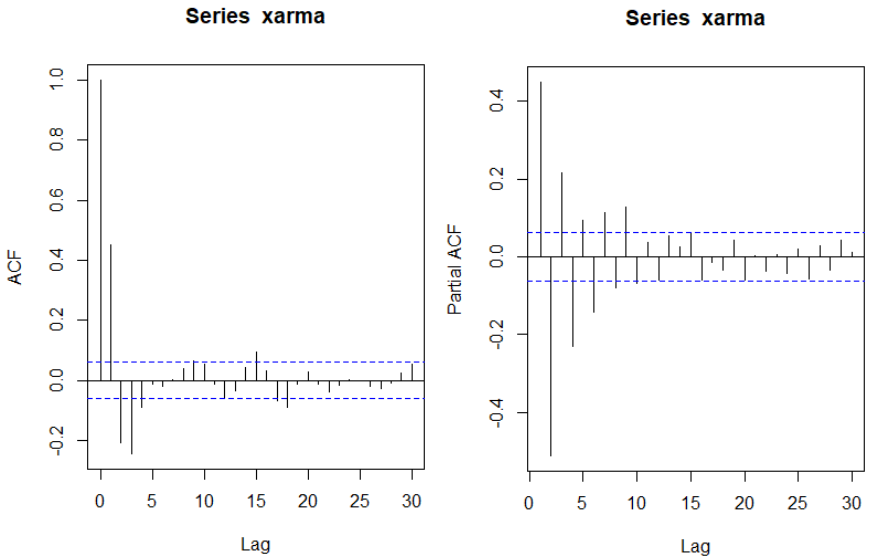


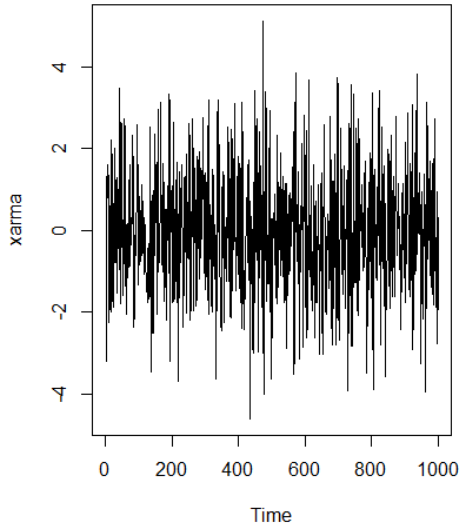


Arma/Arima

$$Y_t = c + \varphi Y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}$$

- Mixture of AR and MA processes
- ACF – decay after lags
- PACF – identify order at decay

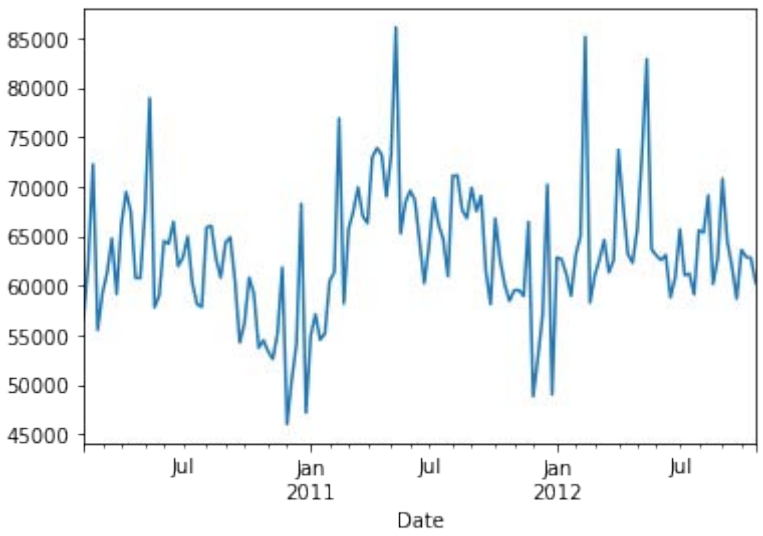




Forecasting

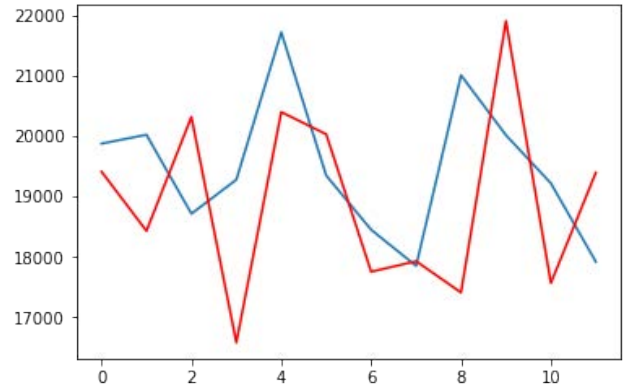
School of Information Studies
Syracuse University

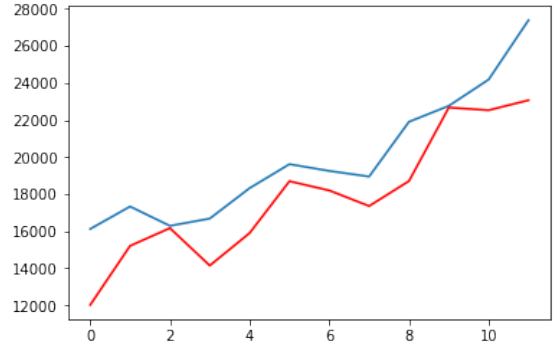
Our Challenge This Week?



Forecasting

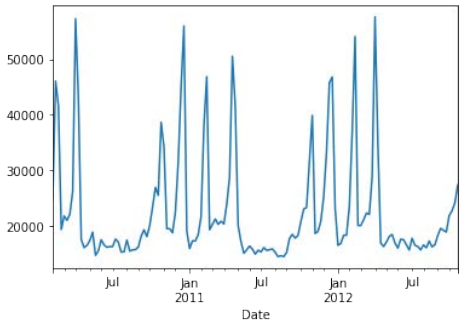
- Moving average smoothing
- Exponential smoothing
- Box Jenkins models

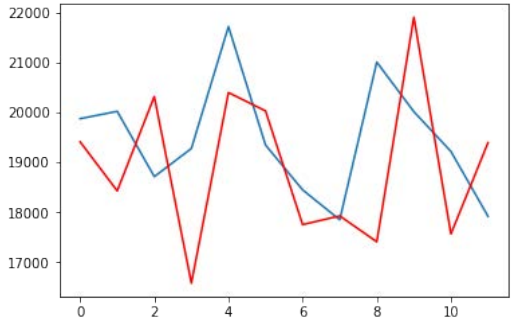




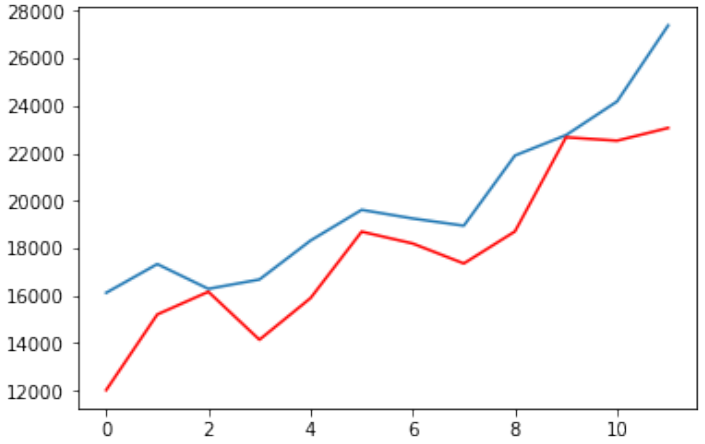
Box Jenkins or ARIMA Models

$$Y_t = c + \varphi Y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}$$



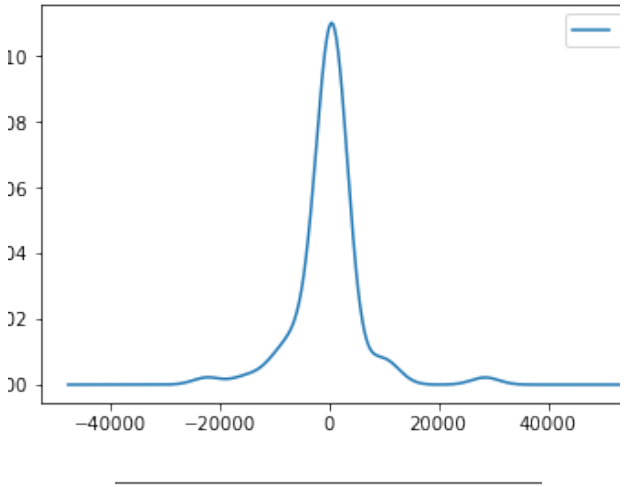


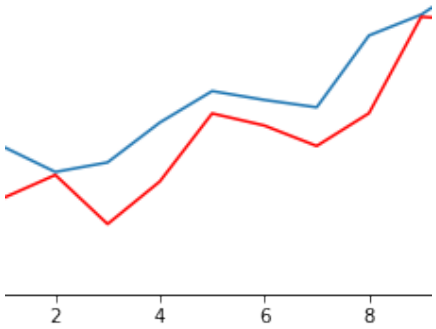
- Assumes stationary
- Transformation can help
- Model identification
- Model estimation
- Model validation



Model Validation

- Residual plot
- ACF – should decay to zero
- Root mean square error
  - mse = mean_squared_error(y, predictions)
  - rmse = sqrt(mse)



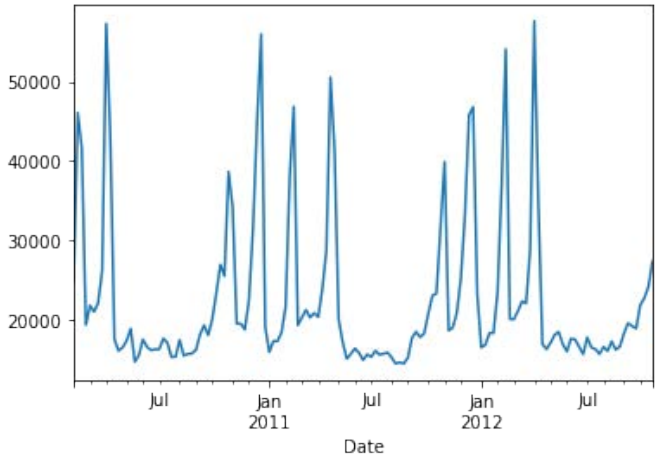


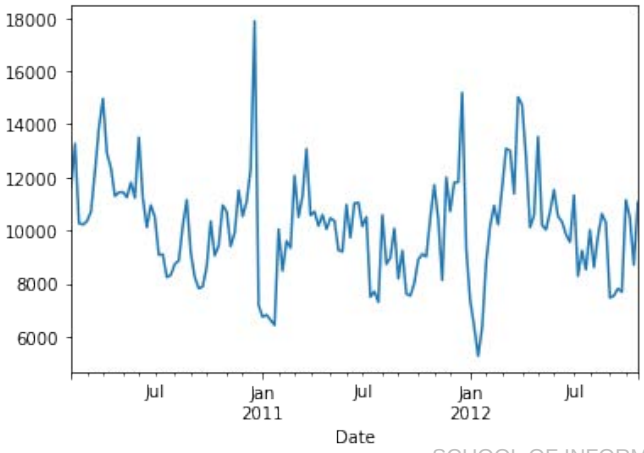
Data Structure

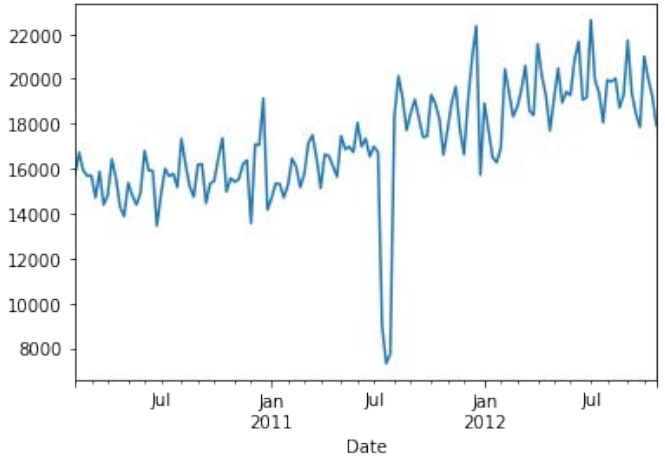
School of Information Studies
Syracuse University

Temporal Data Structure

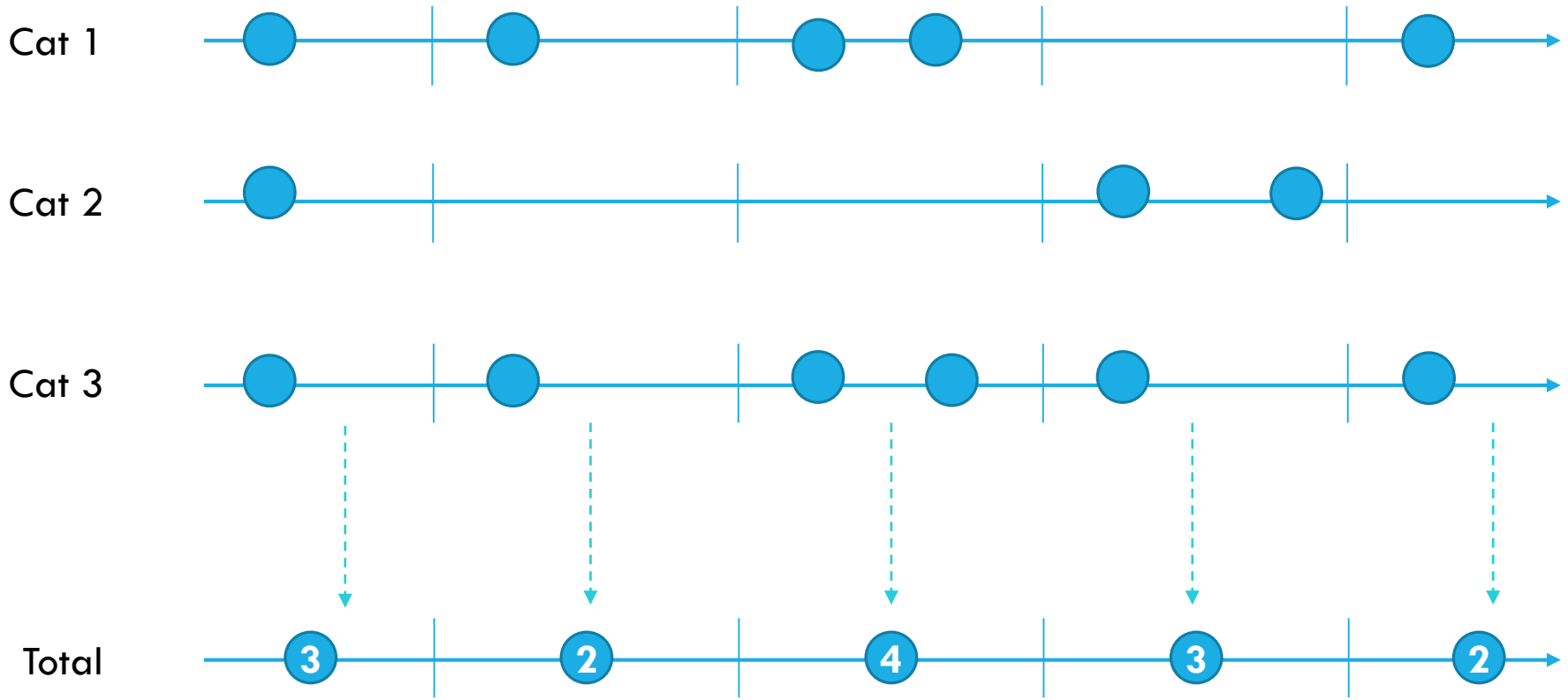
- Exploratory data analysis
- Temporal aggregation
- New features
- Transformation







Temporal Aggregation



New Features

- Function of the data
- Cumulative sums
- Frequencies
- Counts

- NOTE: Always perform causal modeling – never use future information to predict now

Transformations

- Normalize
  - Subtracting mean
  - Dividing maximum value
- Take the log
- Make categorical – bucketize
- Define threshold and make binary
  - >= threshold? 1
  - < threshold? 0