A Restaurant's Future

MICHELLE LI

github.com/michellekli/visitor-forecasting

Photo by Helena Lopes from Pexels

2-6% profit margins (Jong, 2017)

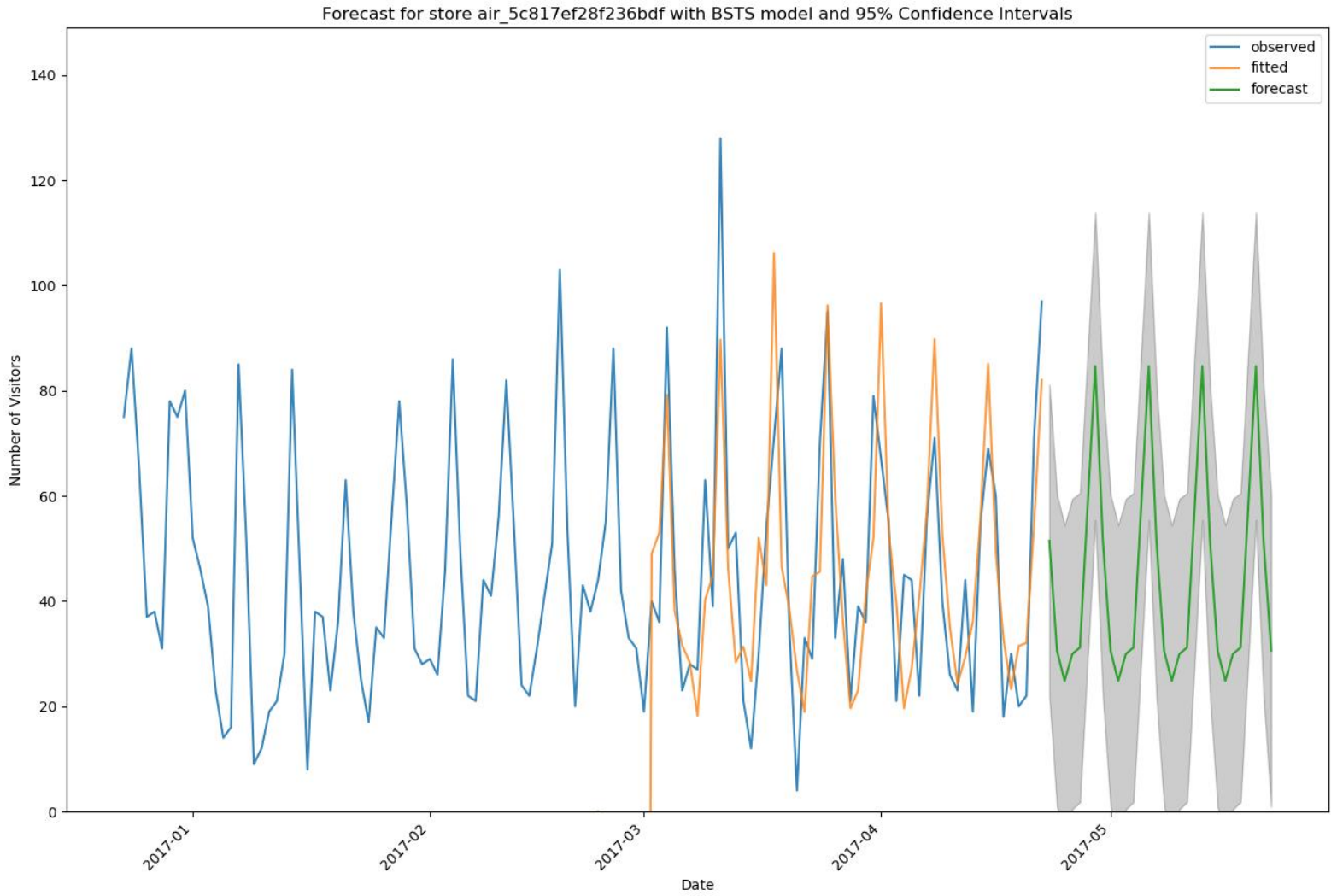
Photo by Daria Shevtsova from Pexels

$25 billion/year food waste (Amin, 2018)

Photo by Elle Hughes from Pexels

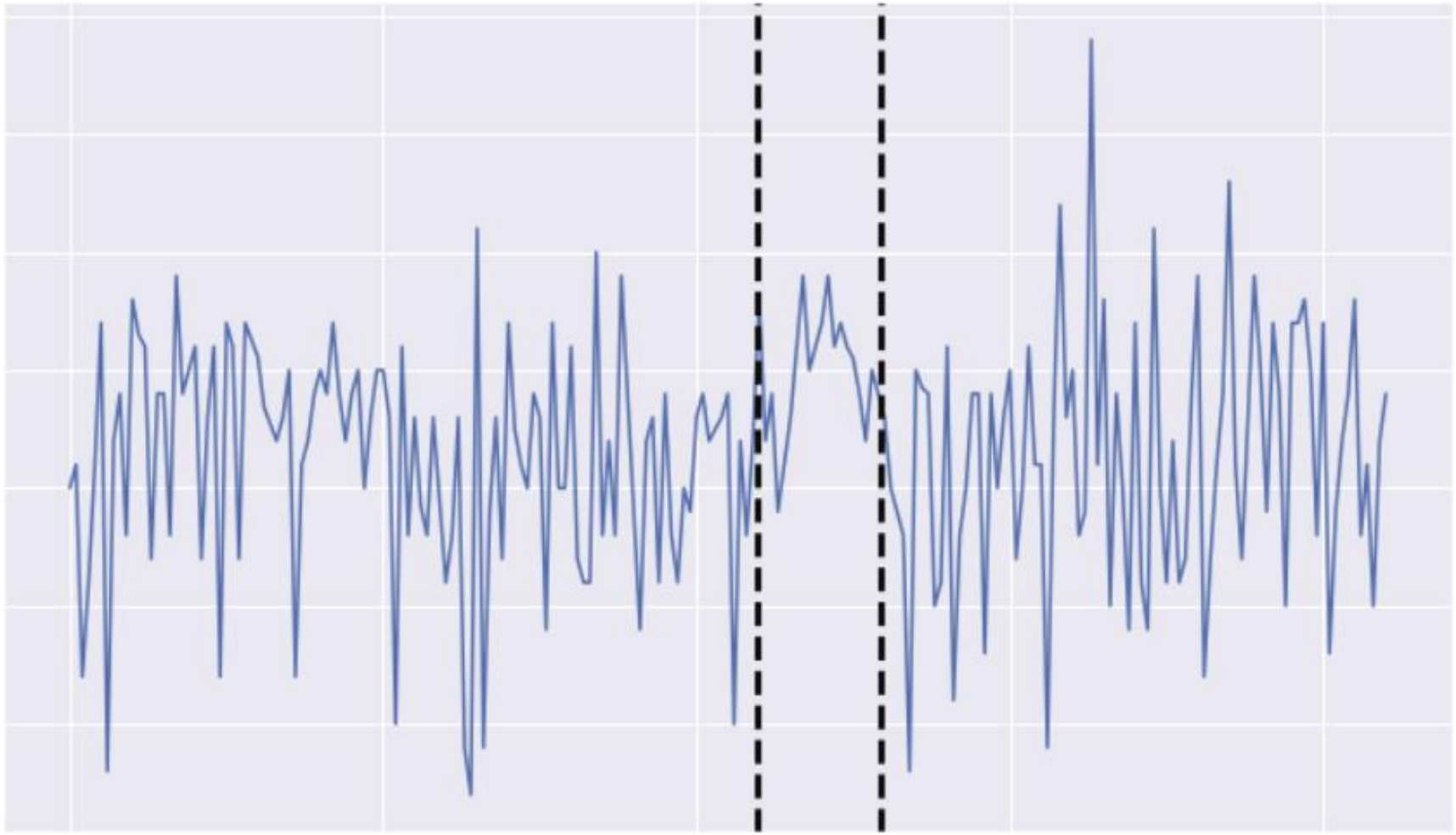
$150,000/year to staff turnover (Lachapelle, 2018)


SEE THE
FUTURE



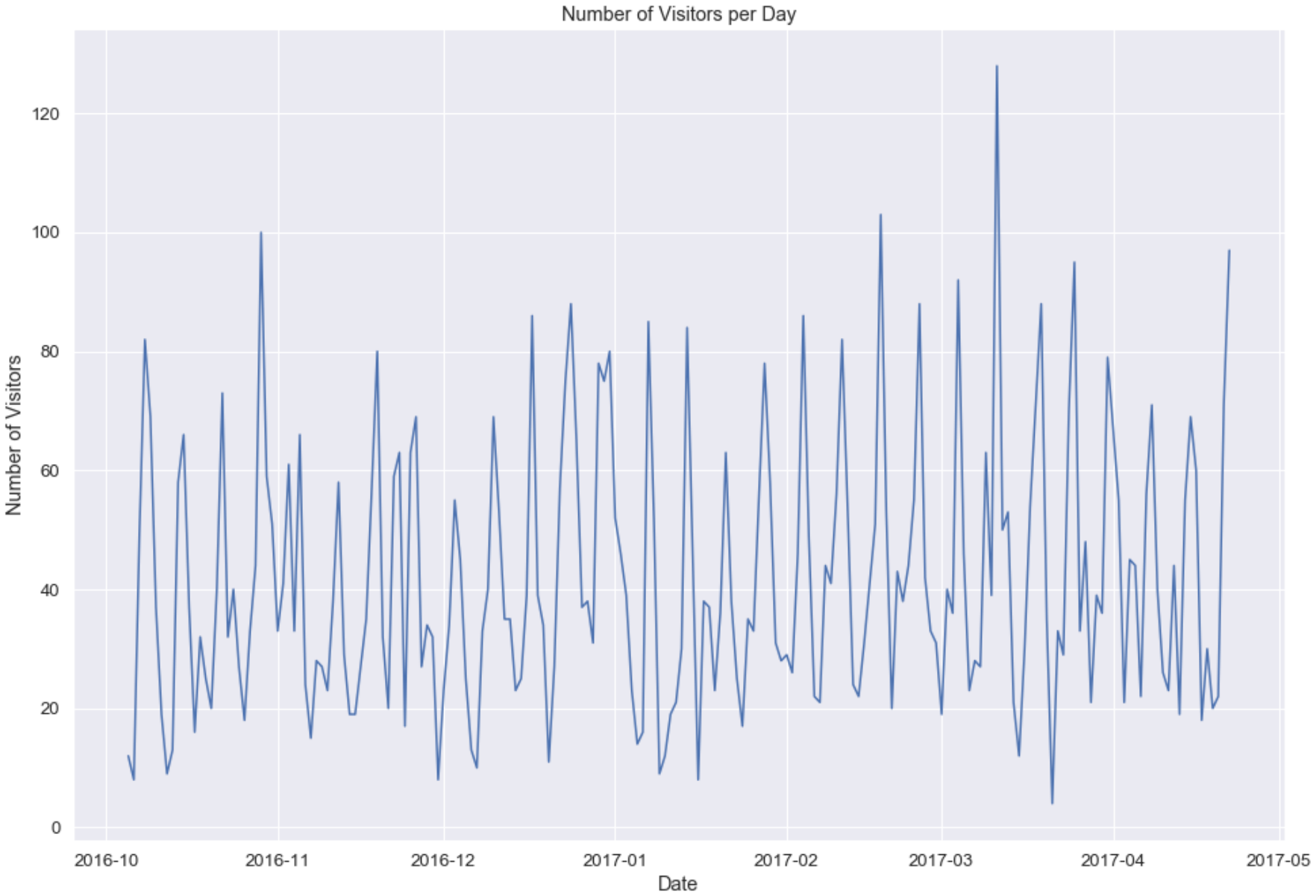
	visit_date	forecast	95_lower	95_upper
127	2017-04-29	84.681772	55.419824	113.943720
128	2017-04-30	51.479521	21.810771	81.148271
129	2017-05-01	30.561021	0.932295	60.189746
130	2017-05-02	24.827111	-4.756797	54.411018
131	2017-05-03	29.949439	0.441101	59.457776
132	2017-05-04	31.172617	1.798376	60.546858
133	2017-05-05	57.939979	28.630746	87.249211
134	2017-05-06	84.681772	55.419824	113.943720
135	2017-05-07	51.479521	21.810771	81.148271
136	2017-05-08	30.561021	0.932295	60.189746
137	2017-05-09	24.827111	-4.756797	54.411018
138	2017-05-10	29.949439	0.441101	59.457776
139	2017-05-11	31.172617	1.798376	60.546858
140	2017-05-12	57.939979	28.630746	87.249211
141	2017-05-13	84.681772	55.419824	113.943720
142	2017-05-14	51.479521	21.810771	81.148271
143	2017-05-15	30.561021	0.932295	60.189746
144	2017-05-16	24.827111	-4.756797	54.411018
145	2017-05-17	29.949439	0.441101	59.457776
146	2017-05-18	31.172617	1.798376	60.546858

The Data

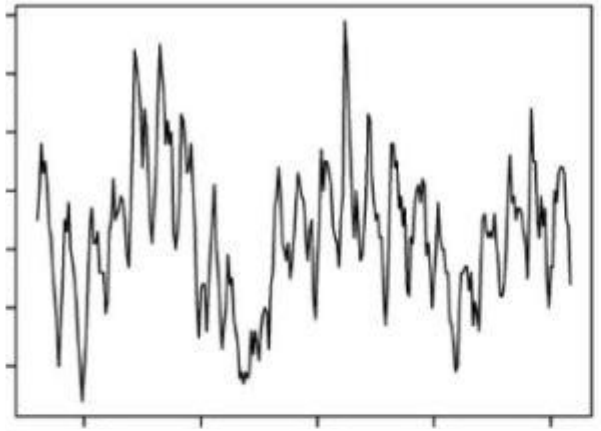


Structural Stability

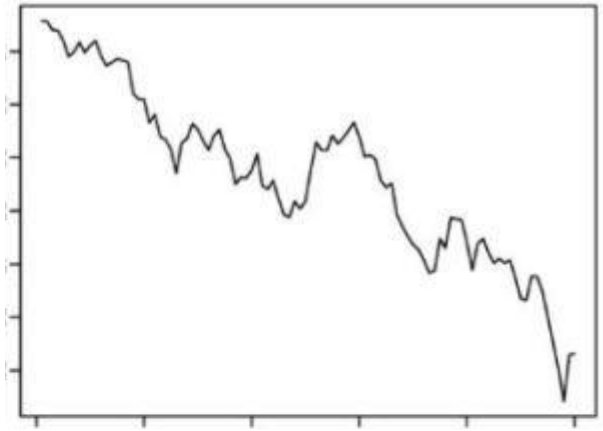
- Bottom-up segmentation
- Python library: ruptures


Number of Visitors per Day
Number of Visitors
0
20
40
60
80
100
120
2016-10
2016-11
2016-12
2017-01
2017-02
2017-03
2017-04
2017-05
Date

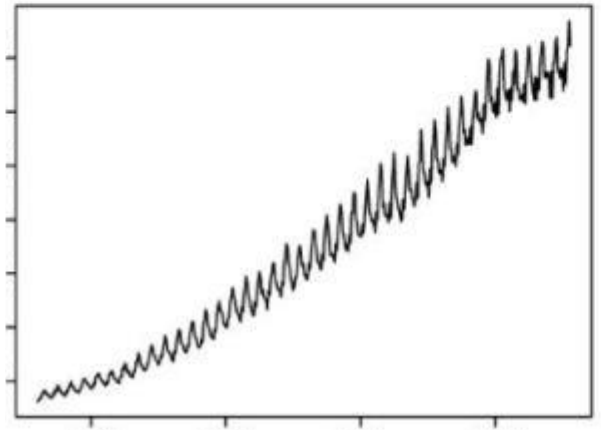
seasonality



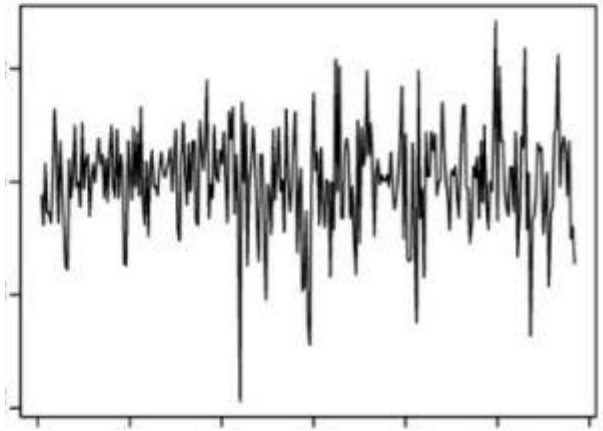
trend



seasonality and trend

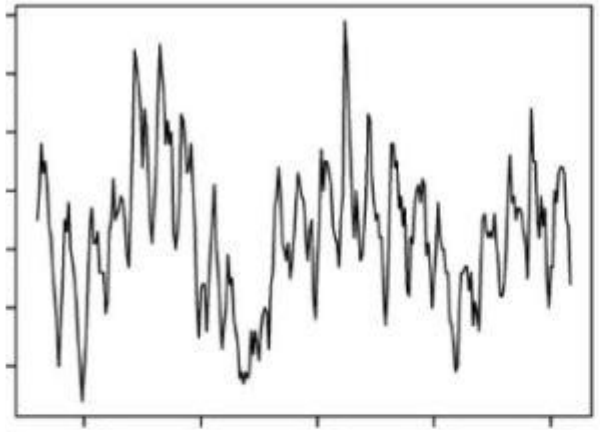


neither (stationary)

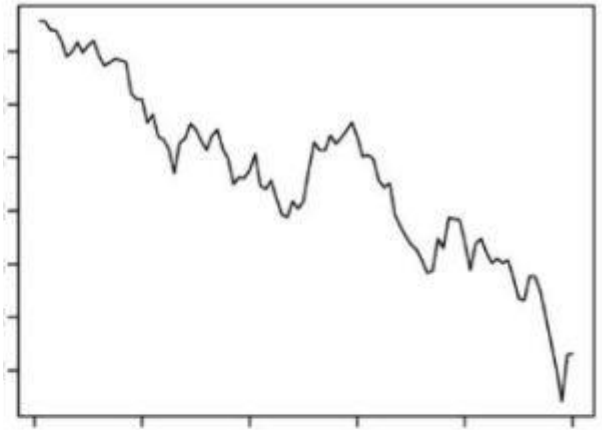


Source: (Seth, 2018)

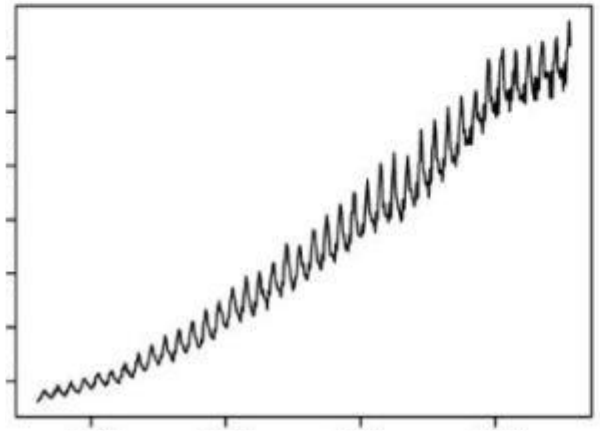
seasonality



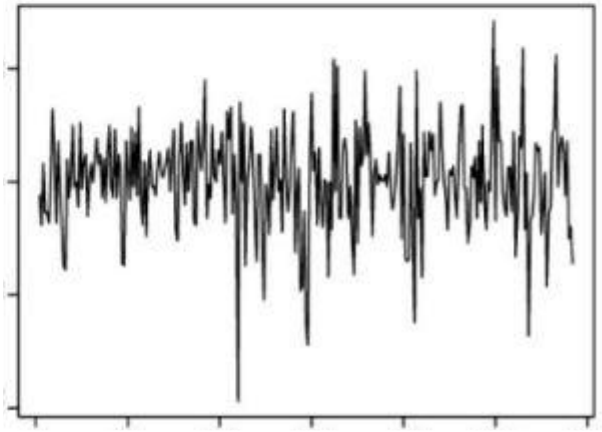
trend



seasonality and trend

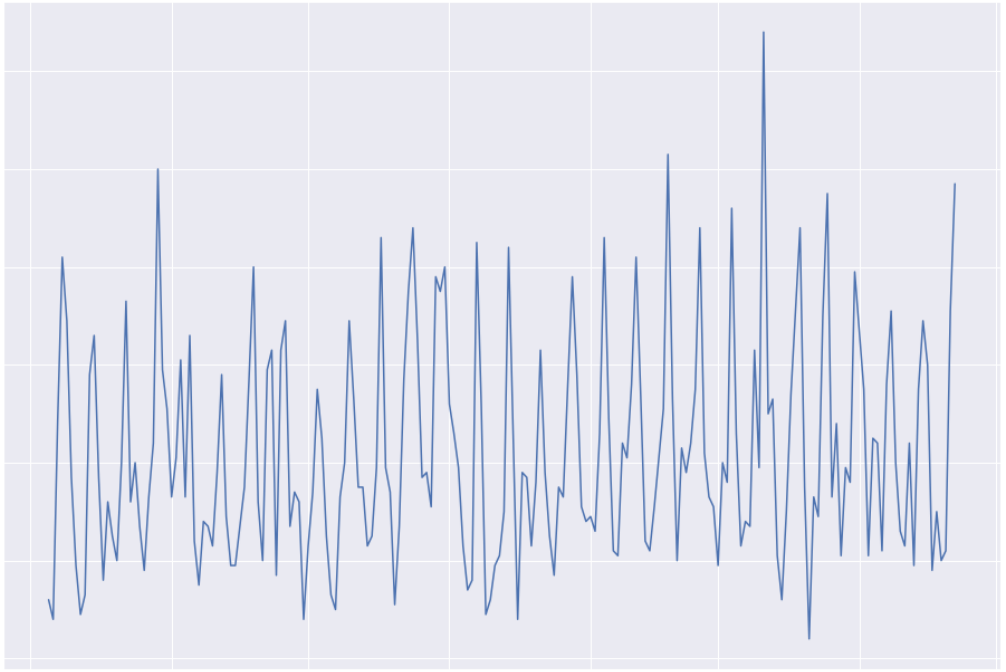


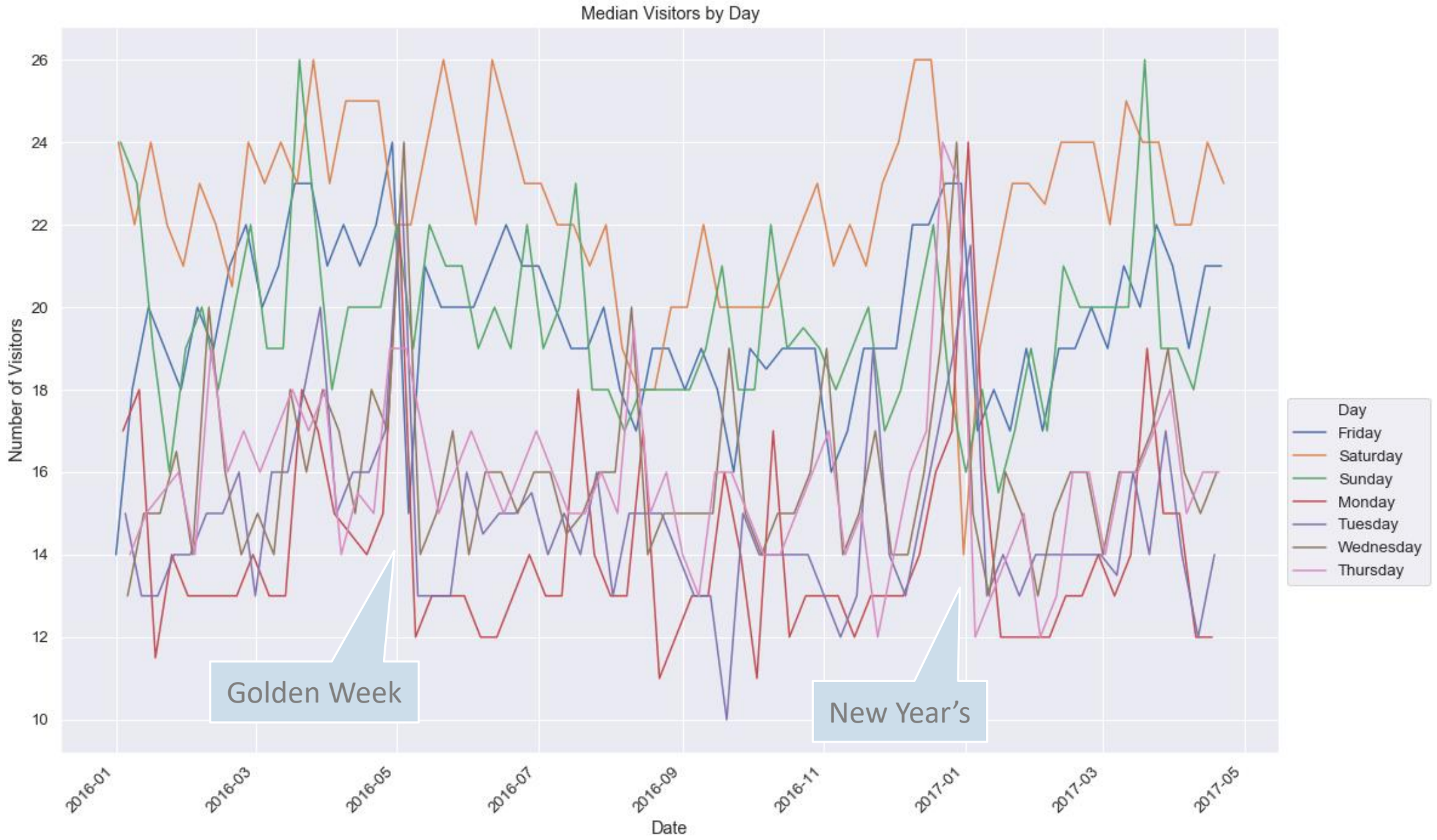
neither (stationary)



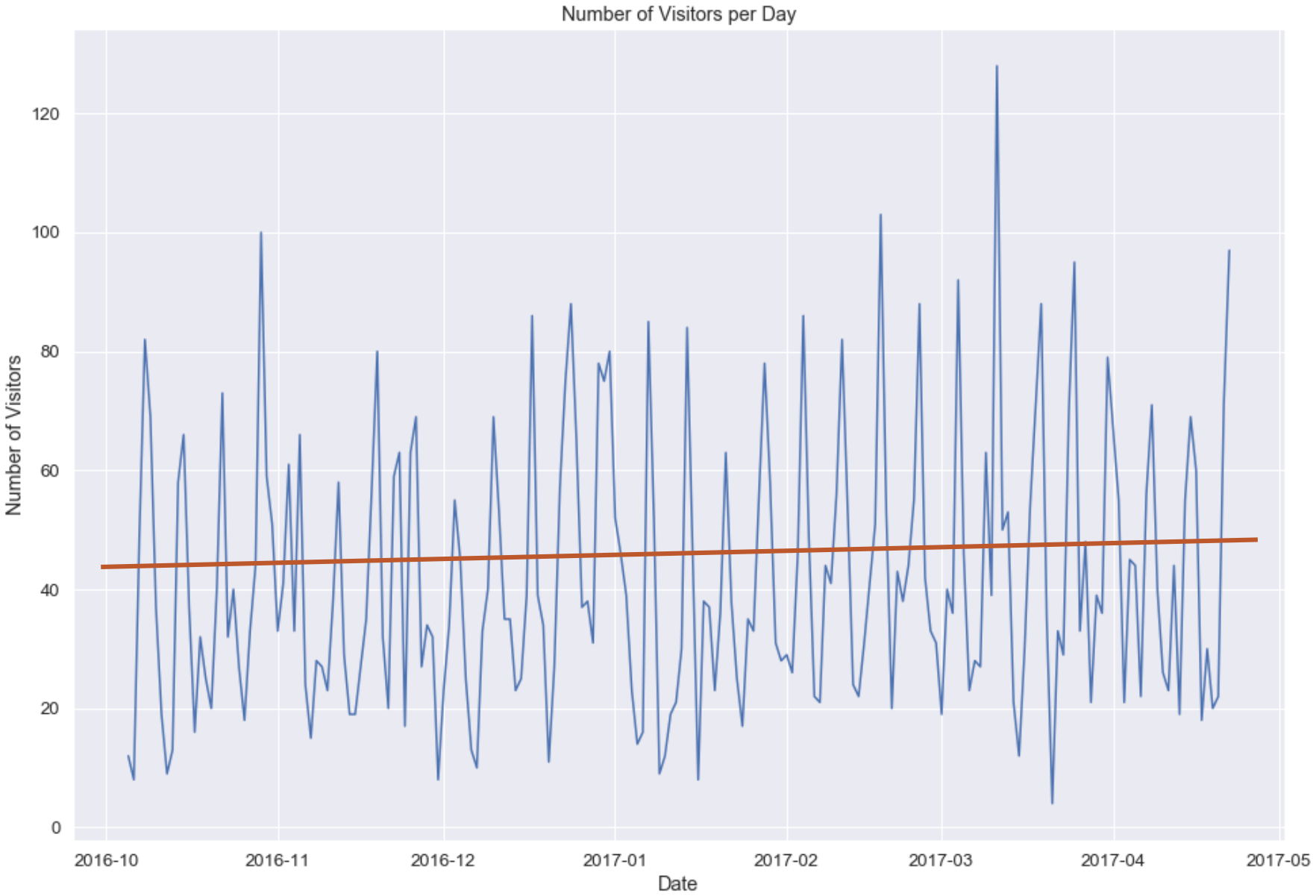
Source: (Seth, 2018)

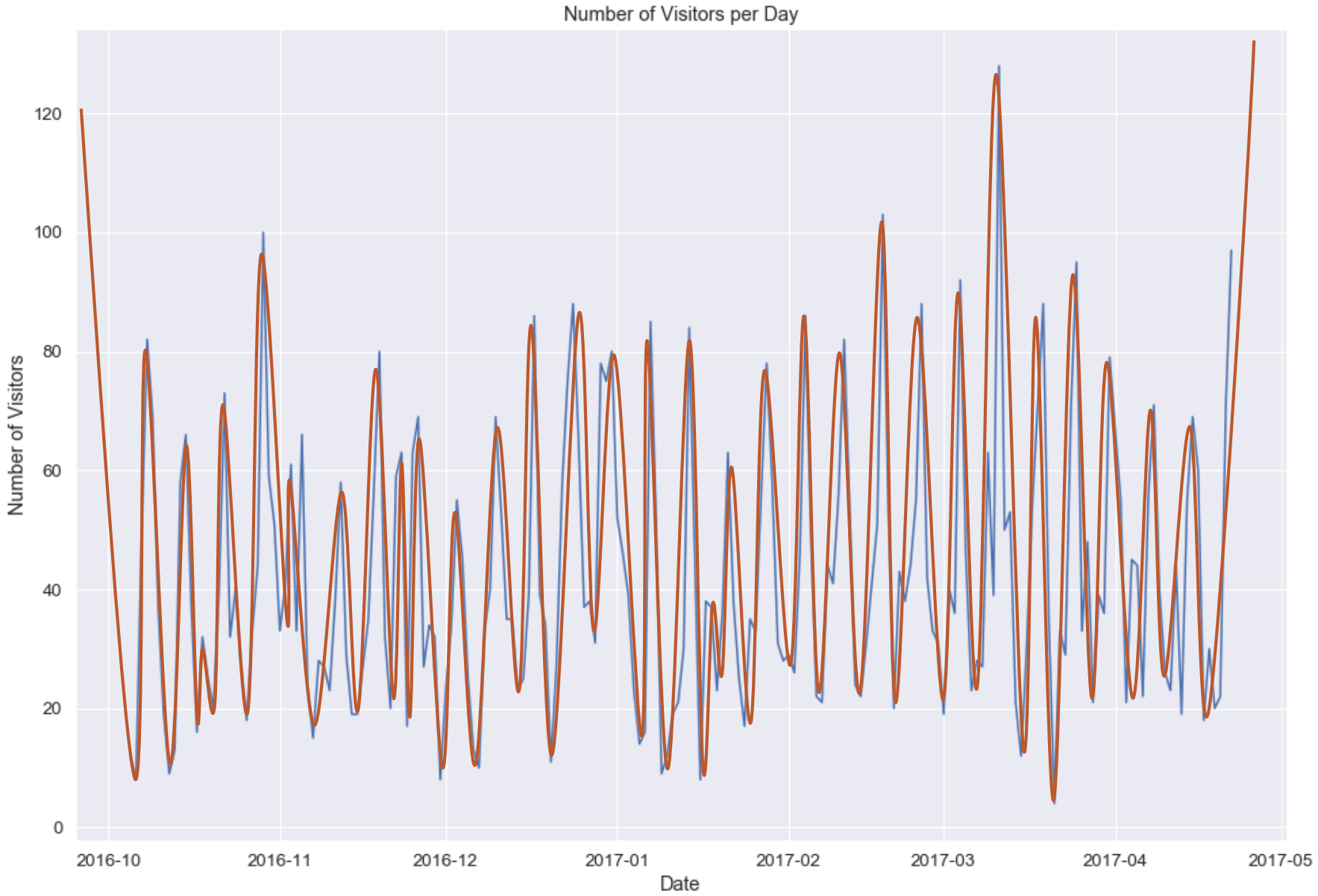
?




Median Visitors by Day
Number of Visitors
26
24
22
20
18
16
14
12
10
2016-01
2016-03
2016-05
2016-07
2016-09
2016-11
2017-01
2017-03
2017-05
Date
Day
Friday
Saturday
Sunday
Monday
Tuesday
Wednesday
Thursday
Golden Week
New Year's

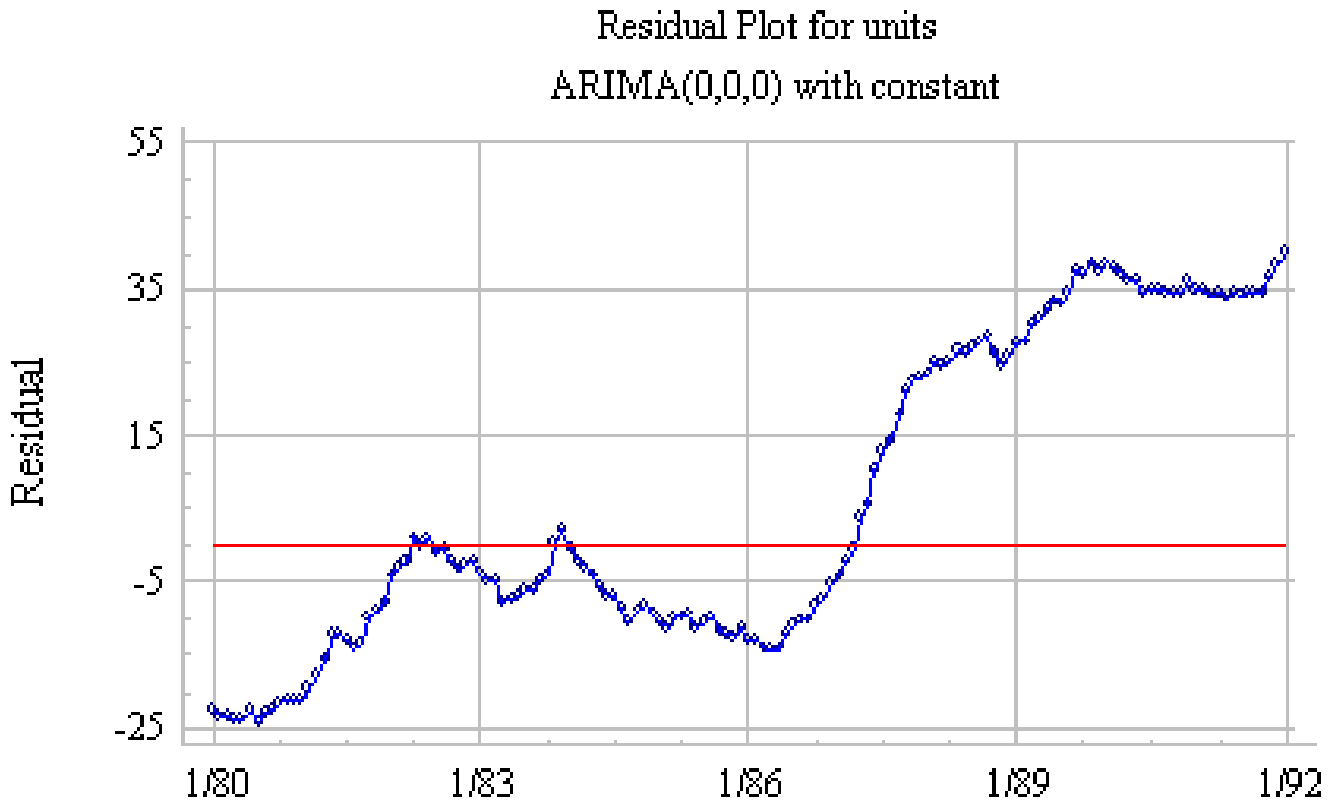
Model Selection


Number of Visitors per Day
Number of Visitors
0
20
40
60
80
100
120
2016-10
2016-11
2016-12
2017-01
2017-02
2017-03
2017-04
2017-05
Date

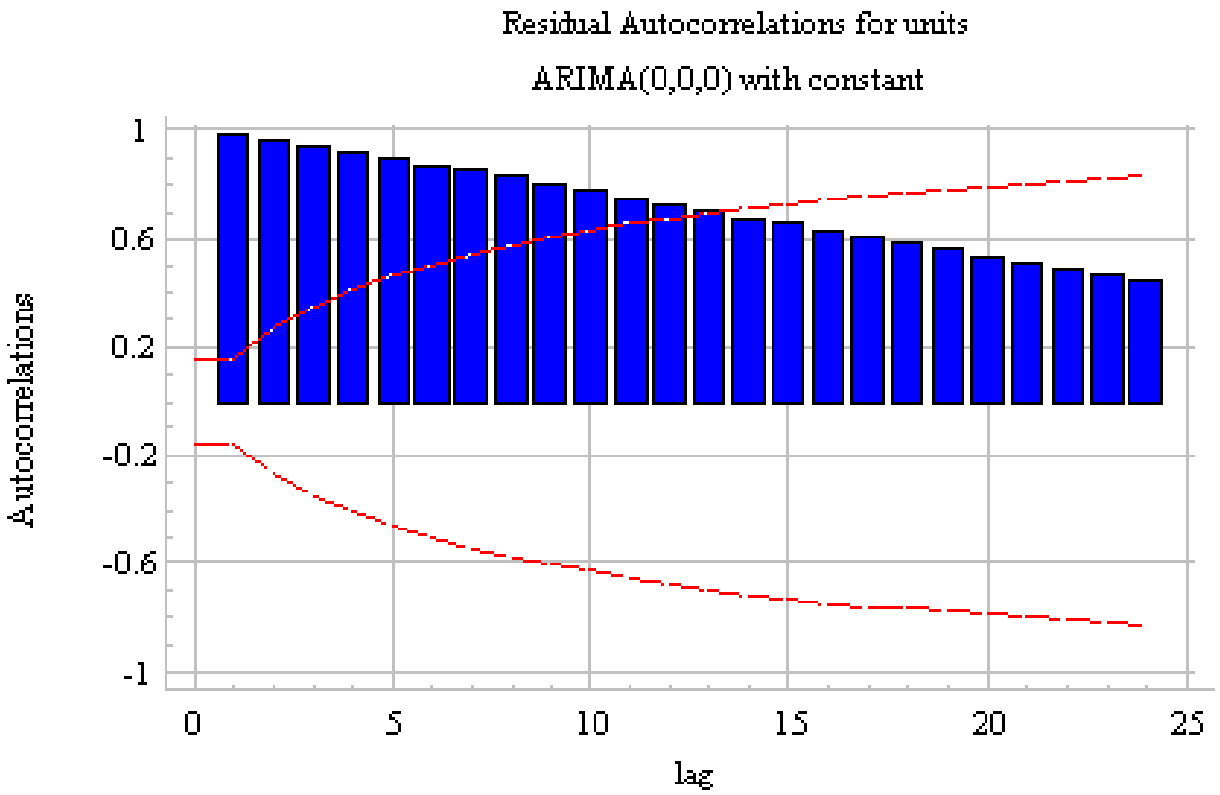

Number of Visitors per Day
Number of Visitors
0
20
40
60
80
100
120
2016-10
2016-11
2016-12
2017-01
2017-02
2017-03
2017-04
2017-05
Date

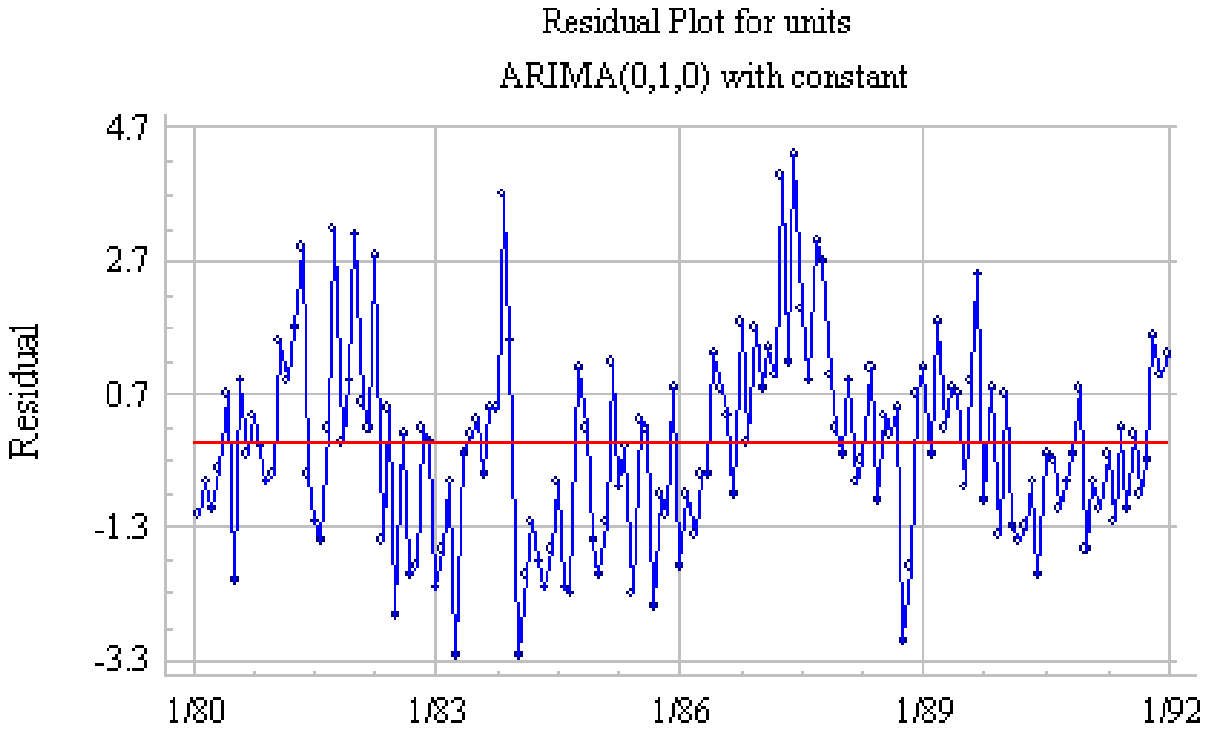
ARIMA(p,d,q)

- AutoRegressive (p)
- Integrated (d)
- Moving Average (q)



Source: (Nau, Identifying the order of differencing in an ARIMA model, n.d.)





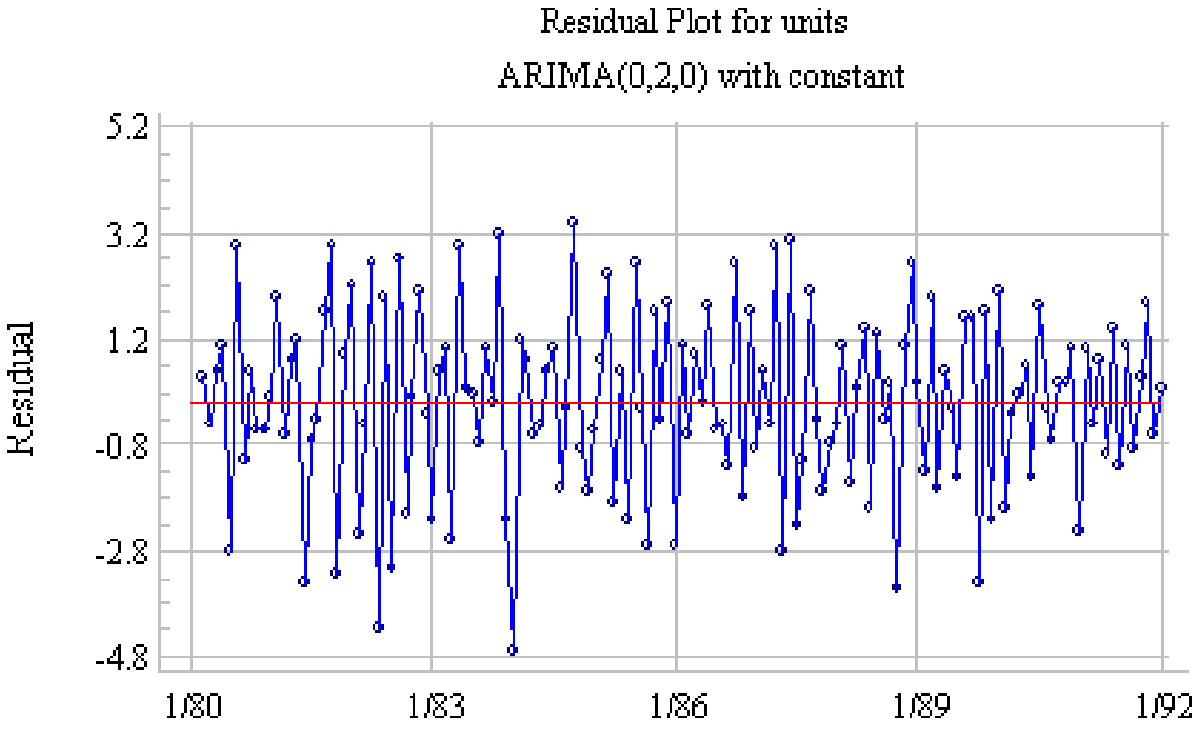
Source: (Nau, Identifying the order of differencing in an ARIMA model, n.d.)

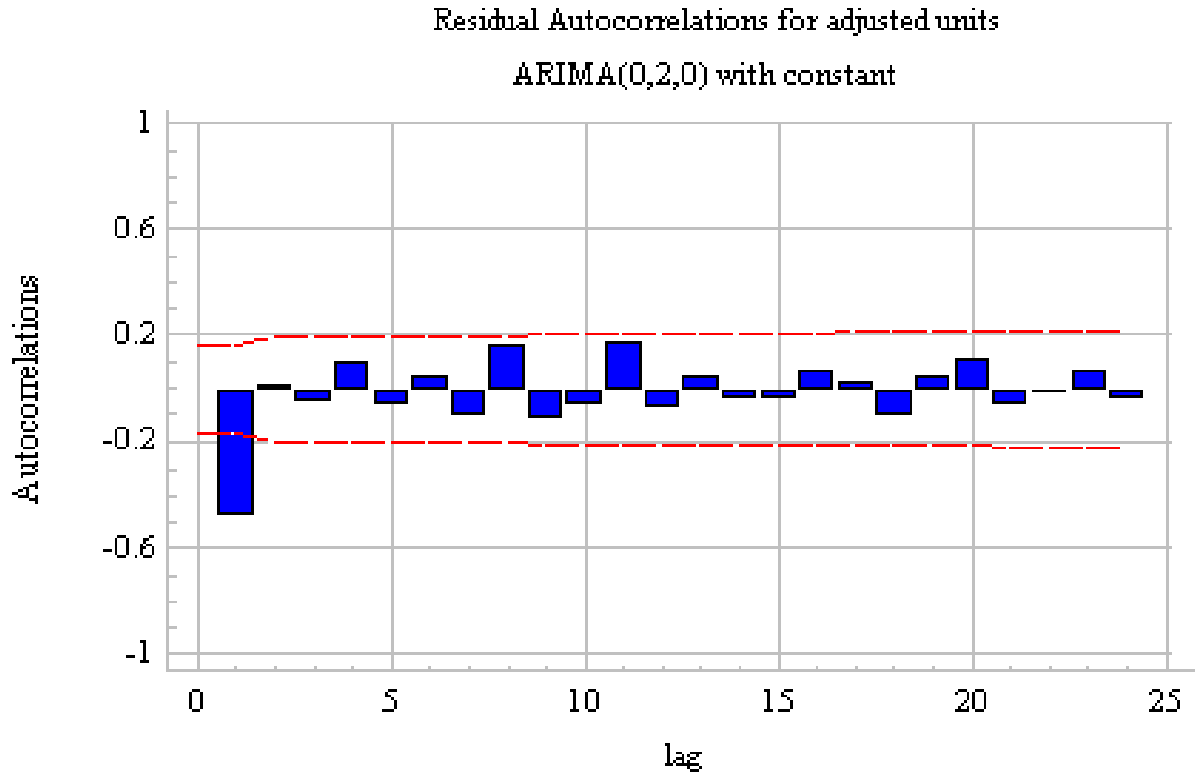
Integrated (d)

- Very slow linear decay in autocorrelation function (ACF) plot
- statsmodels.graphics.tsaplots: plot_acf, plot_pacf

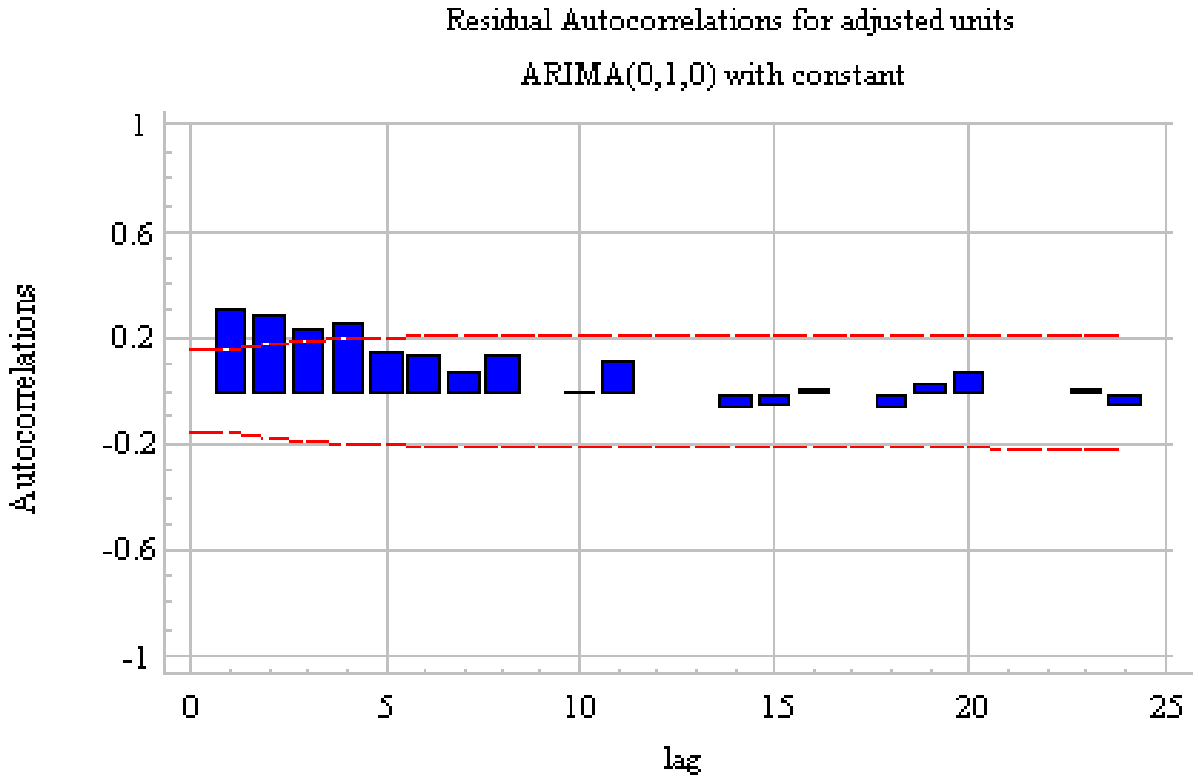
Over Differencing

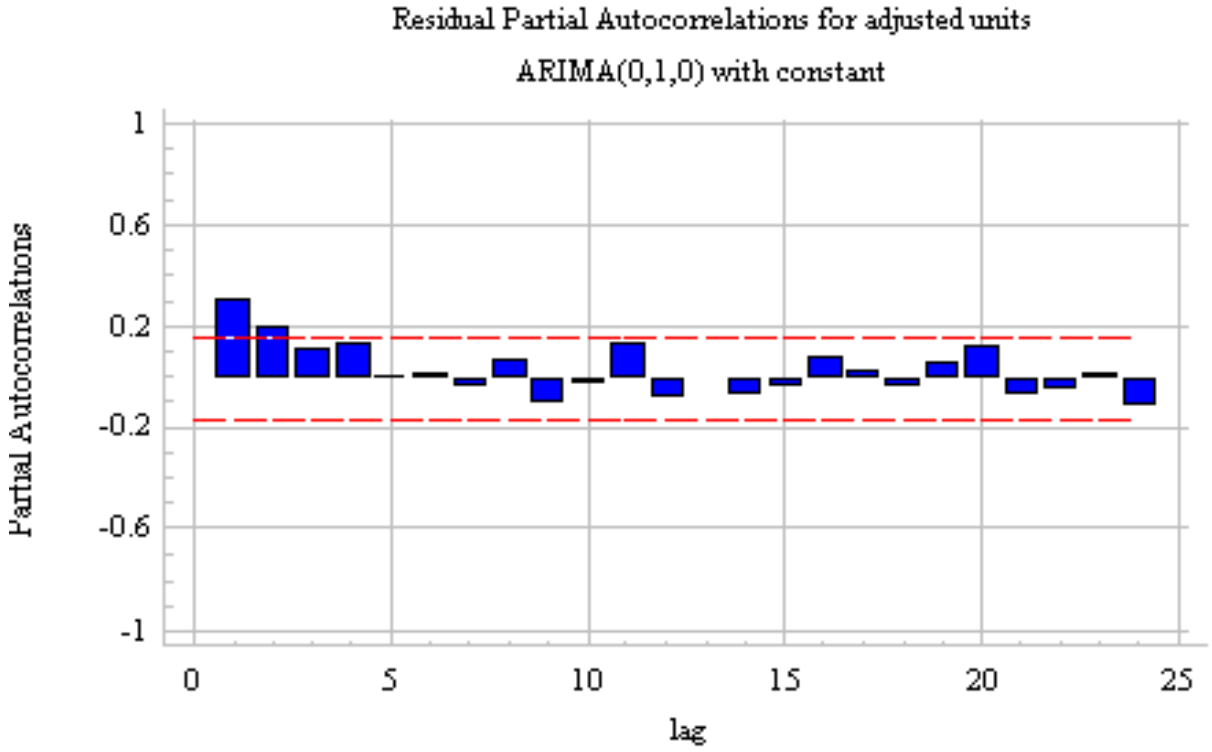
- Residuals change in sign from one observation to the next





Source: (Nau, Identifying the order of differencing in an ARIMA model, n.d.)

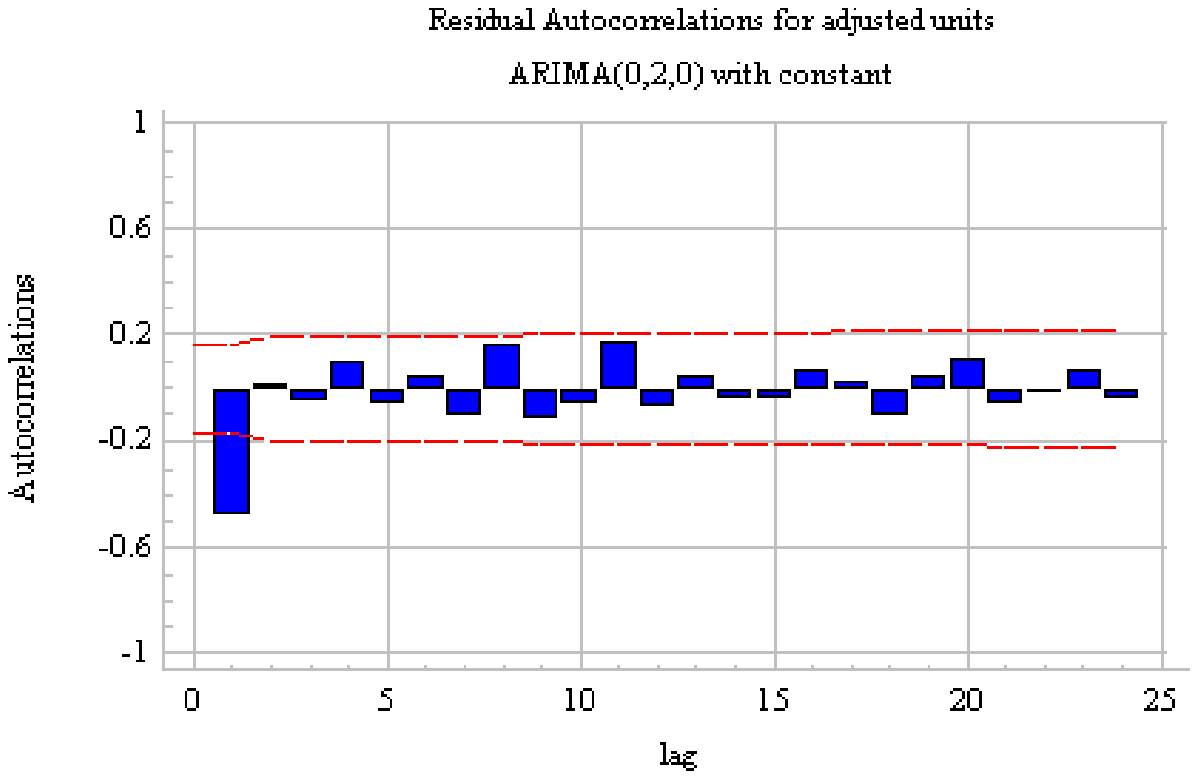


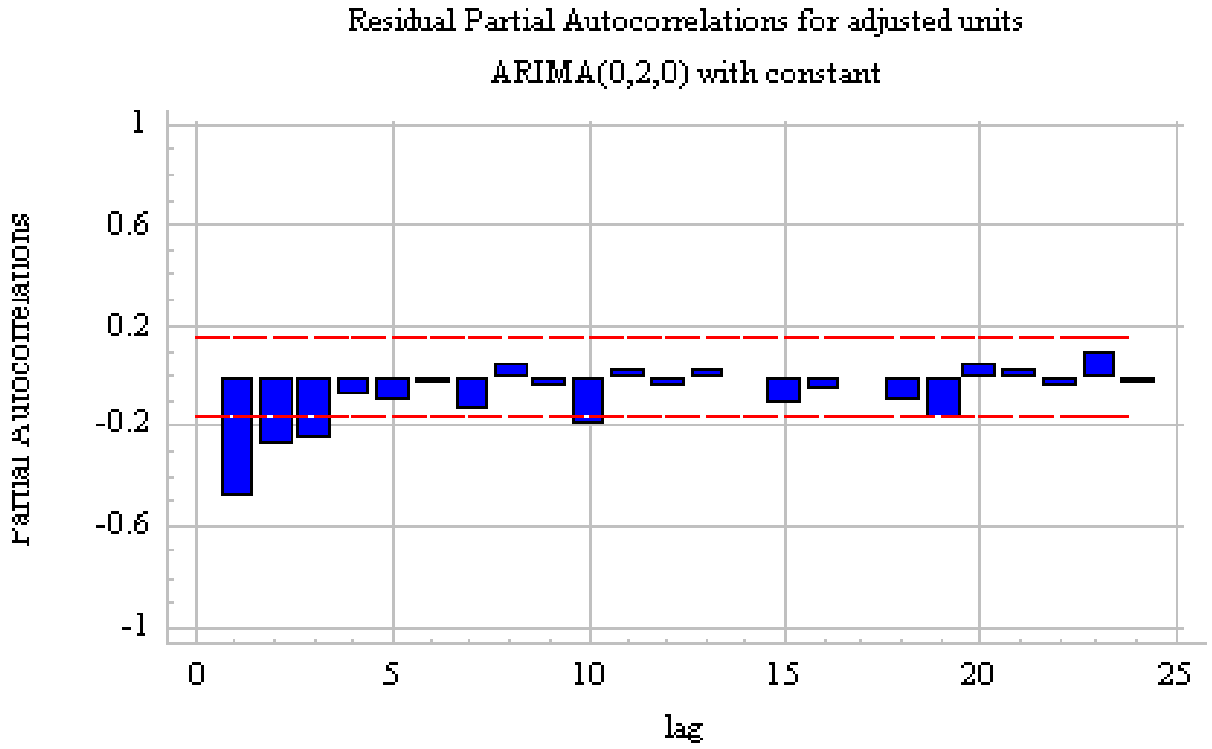


Source: (Nau, Identifying the numbers of AR or MA terms in an ARIMA model, n.d.)

AutoRegressive (p)

- Partial ACF (PACF) shows sharper cutoff than ACF
- Select AR term as the number of significant spikes in PACF

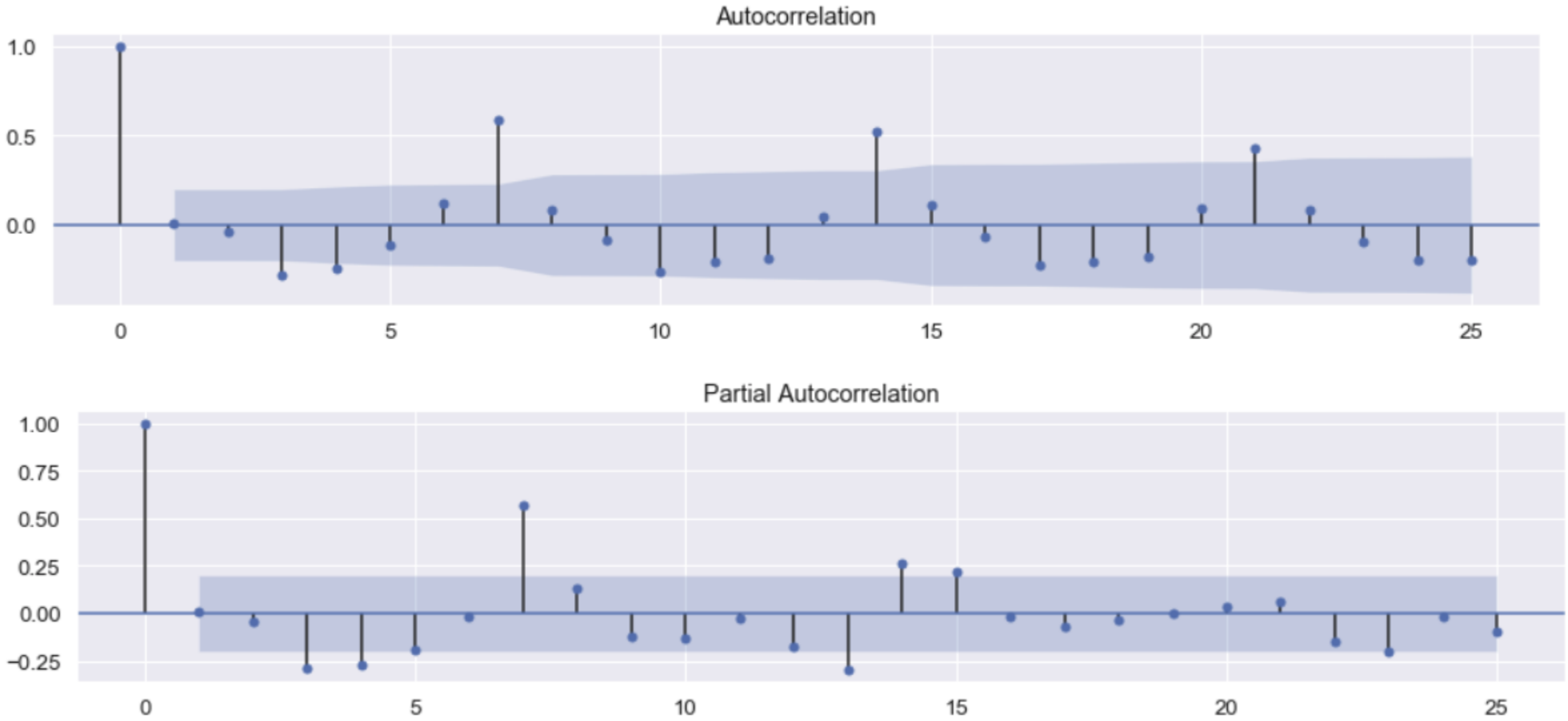




Source: (Nau, Identifying the numbers of AR or MA terms in an ARIMA model, n.d.)

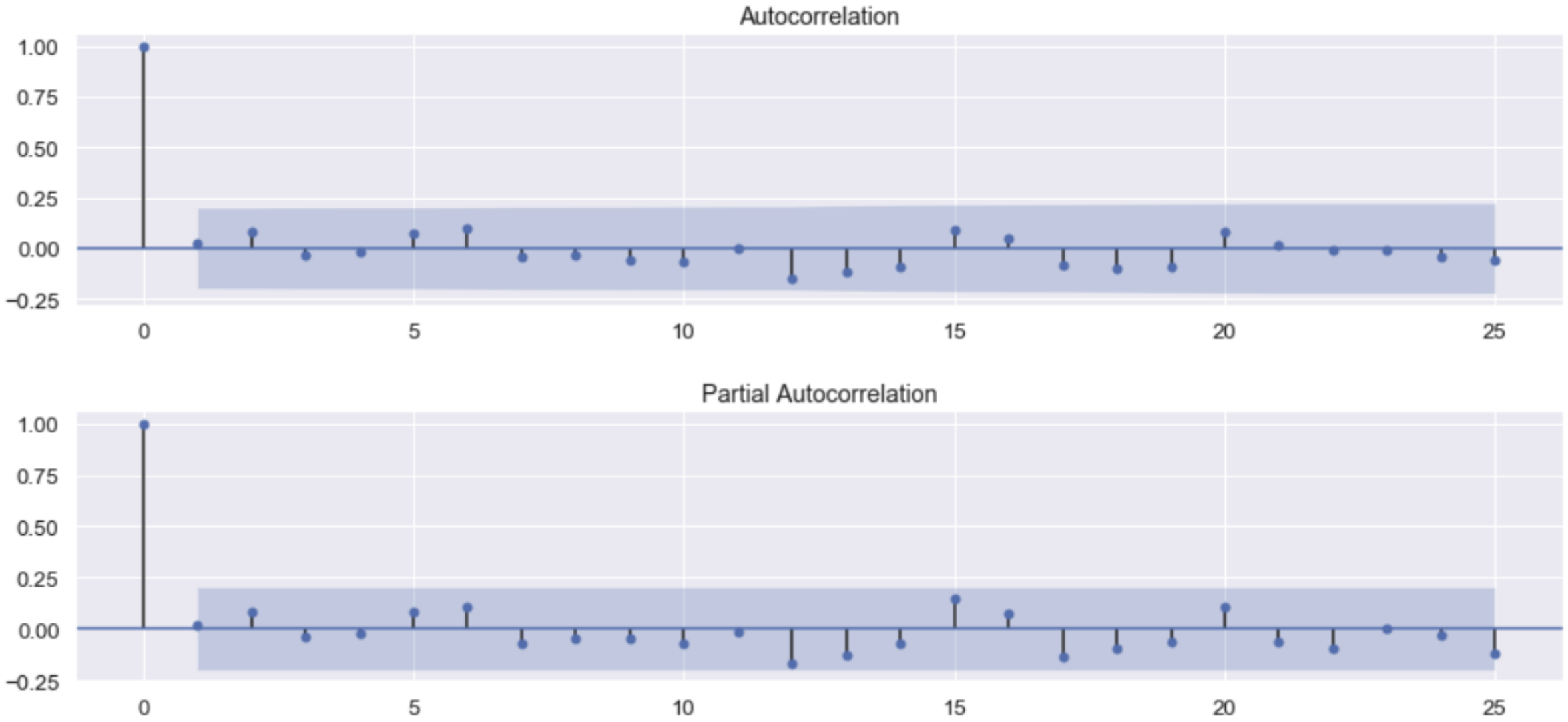
Moving Average (q)

- ACF shows sharper cutoff than PACF
- Select MA term as the number of significant spikes in ACF


Autocorrelation
1.0
0.5
0.0
0
5
10
15
20
25
Partial Autocorrelation
1.00
0.75
0.50
0.25
0.00
−0.25
0
5
10
15
20
25

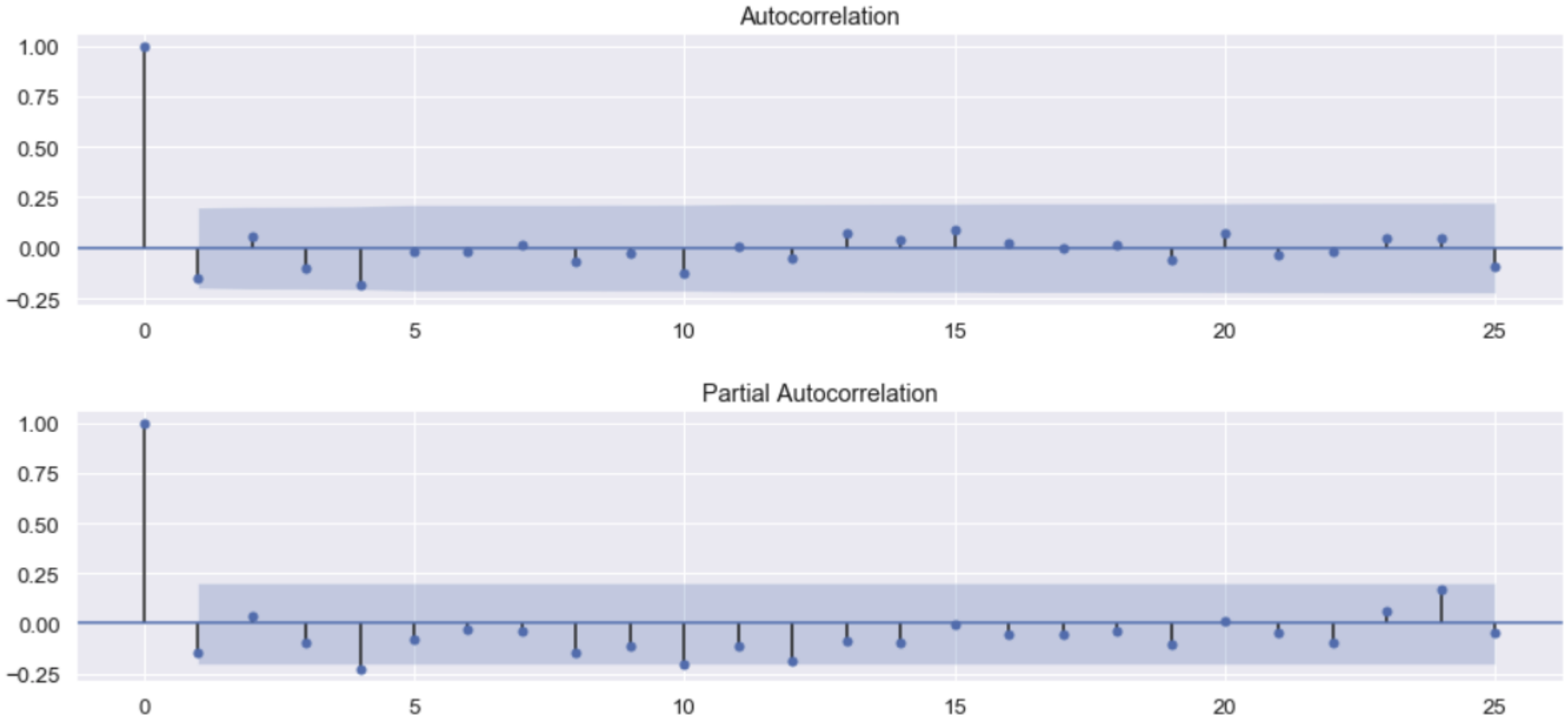
SARIMAX(p,d,q)(P,D,Q)m

- Seasonal AutoRegressive (P)
- Seasonal Integrated (D)
- Seasonal Moving Average (Q)
- Number of observations per seasonal cycle (m)
  - 7 – daily
  - 12 – monthly
  - 52 - weekly


Autocorrelation
1.00
0.75
0.50
0.25
0.00
−0.25
0
5
10
15
20
25
Partial Autocorrelation
1.00
0.75
0.50
0.25
0.00
−0.25
0
5
10
15
20
25

Bayesian Structural Time Series (BSTS)

- Trend
- Seasonality
- Auto-regression


Autocorrelation
1.00
0.75
0.50
0.25
0.00
−0.25
0
5
10
15
20
25
Partial Autocorrelation
1.00
0.75
0.50
0.25
0.00
−0.25
0
5
10
15
20
25

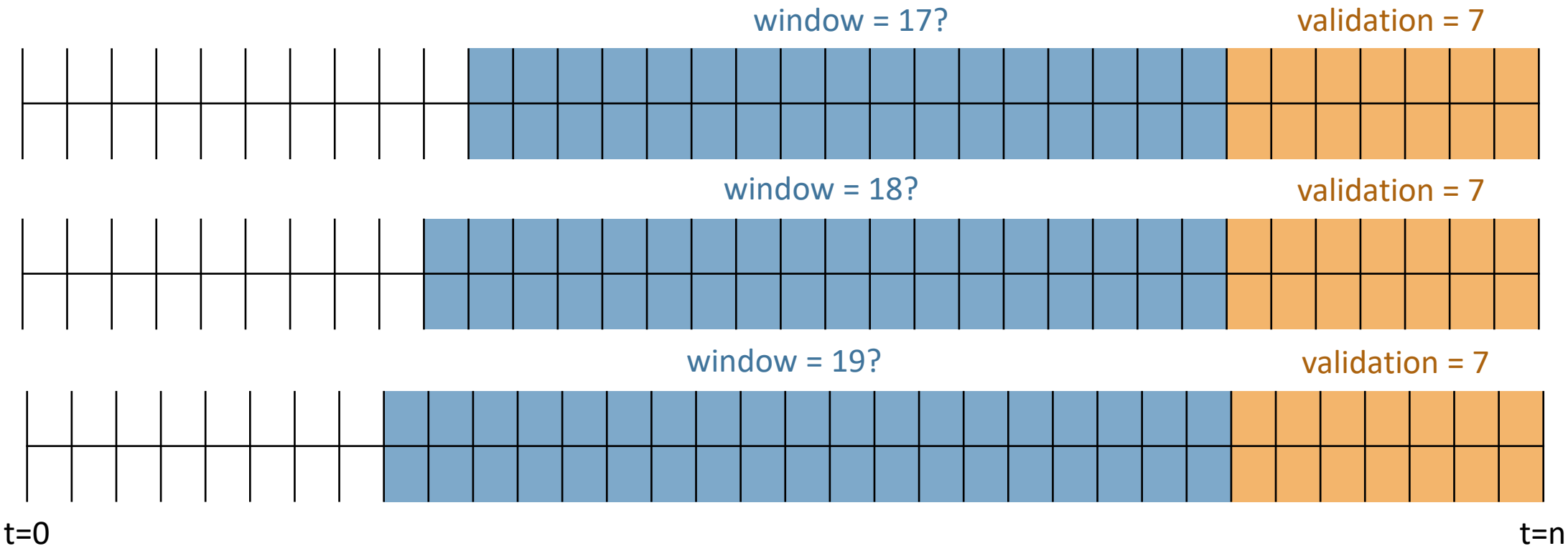
Forecasting 800+ Restaurants

- “pmdarima brings R’s beloved auto.arima to Python” (Smith, n.d.)
- Step-wise search for best model
  - Bias-corrected Akaike information criterion (AICc)

Model Tuning

Window Size

- Validation size = forecast horizon



Model Evaluation

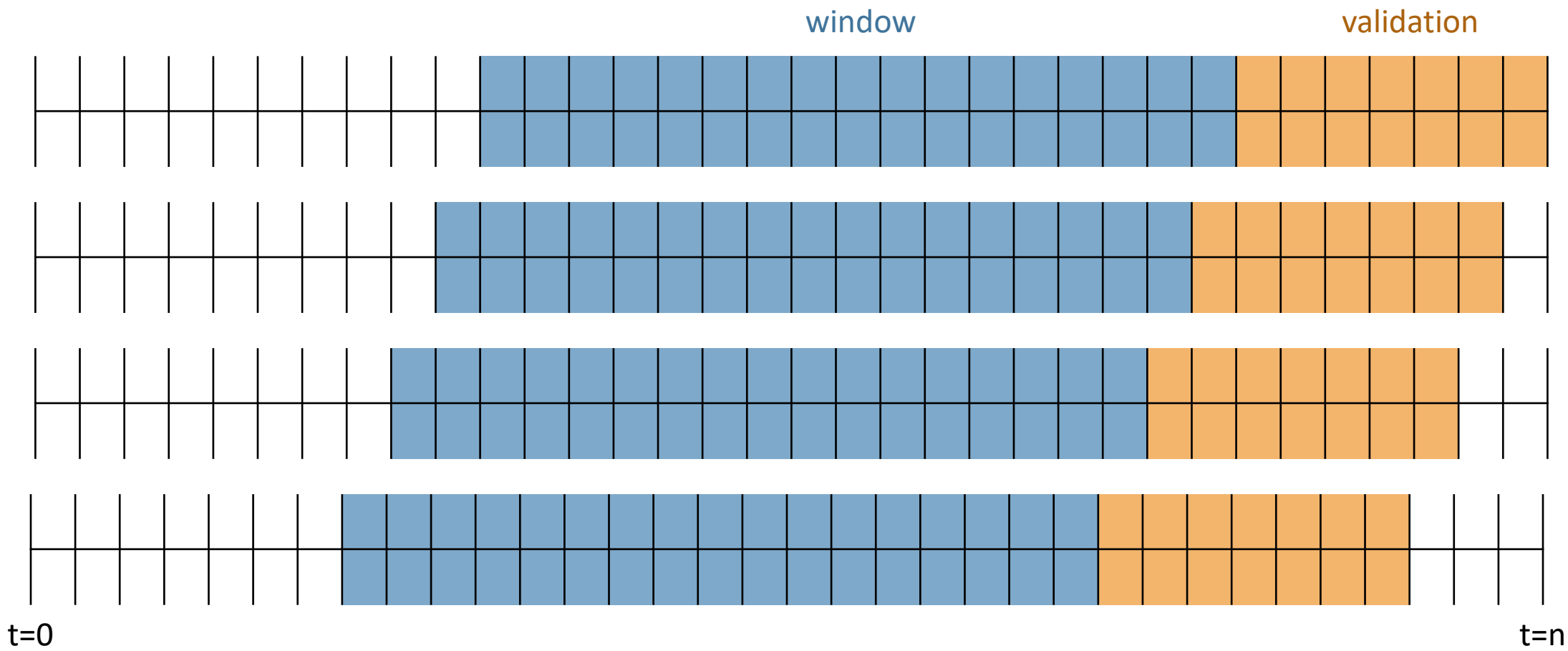
RMSLE over RMSE

- Robust to outliers
- Ignores scale of error
- Penalizes underestimation

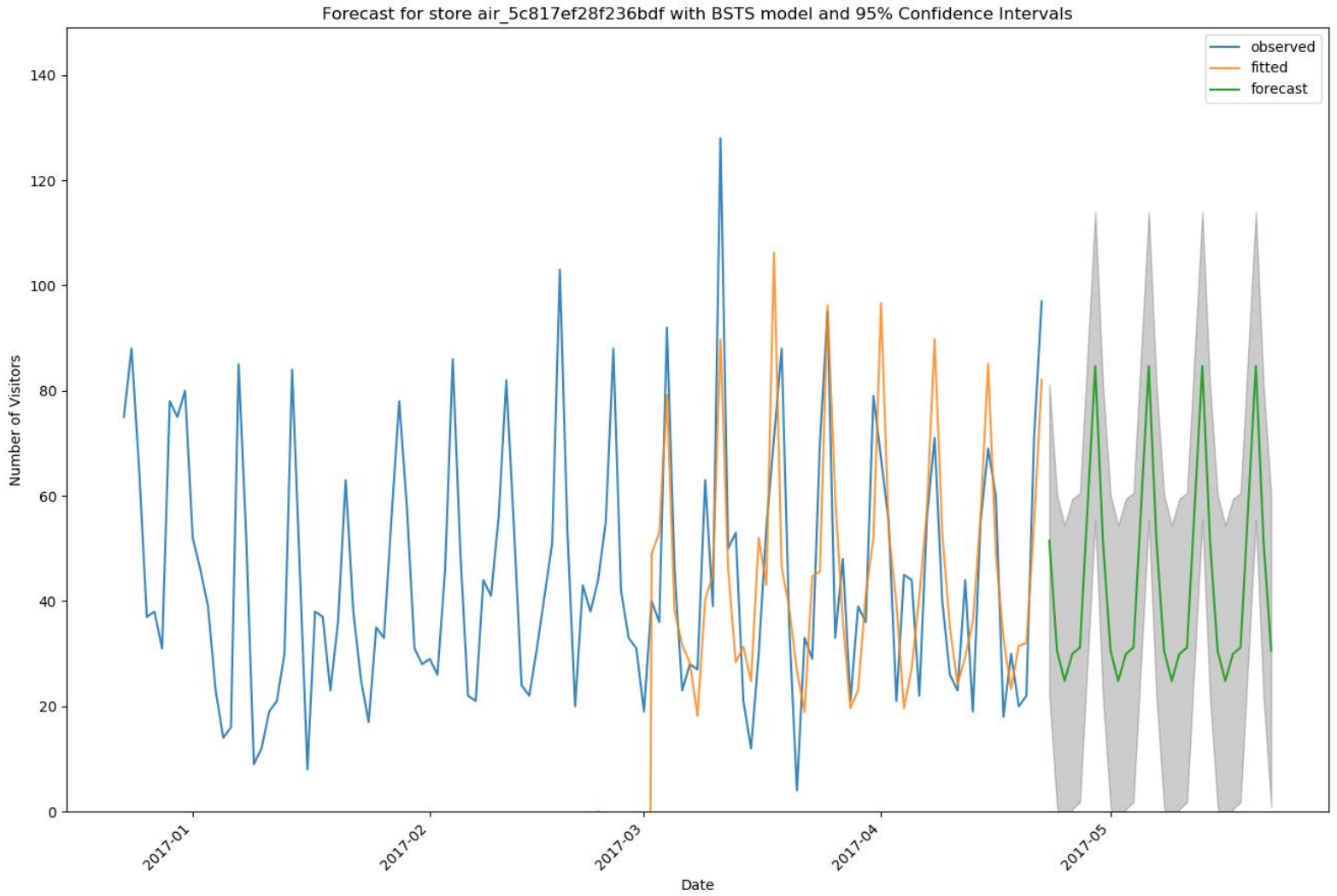
$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\ln(p_i+1)-\ln(a_i+1))^2}$$

- n = total number of observations
- p_i = prediction for number of visitors
- a_i = actual number of visitors

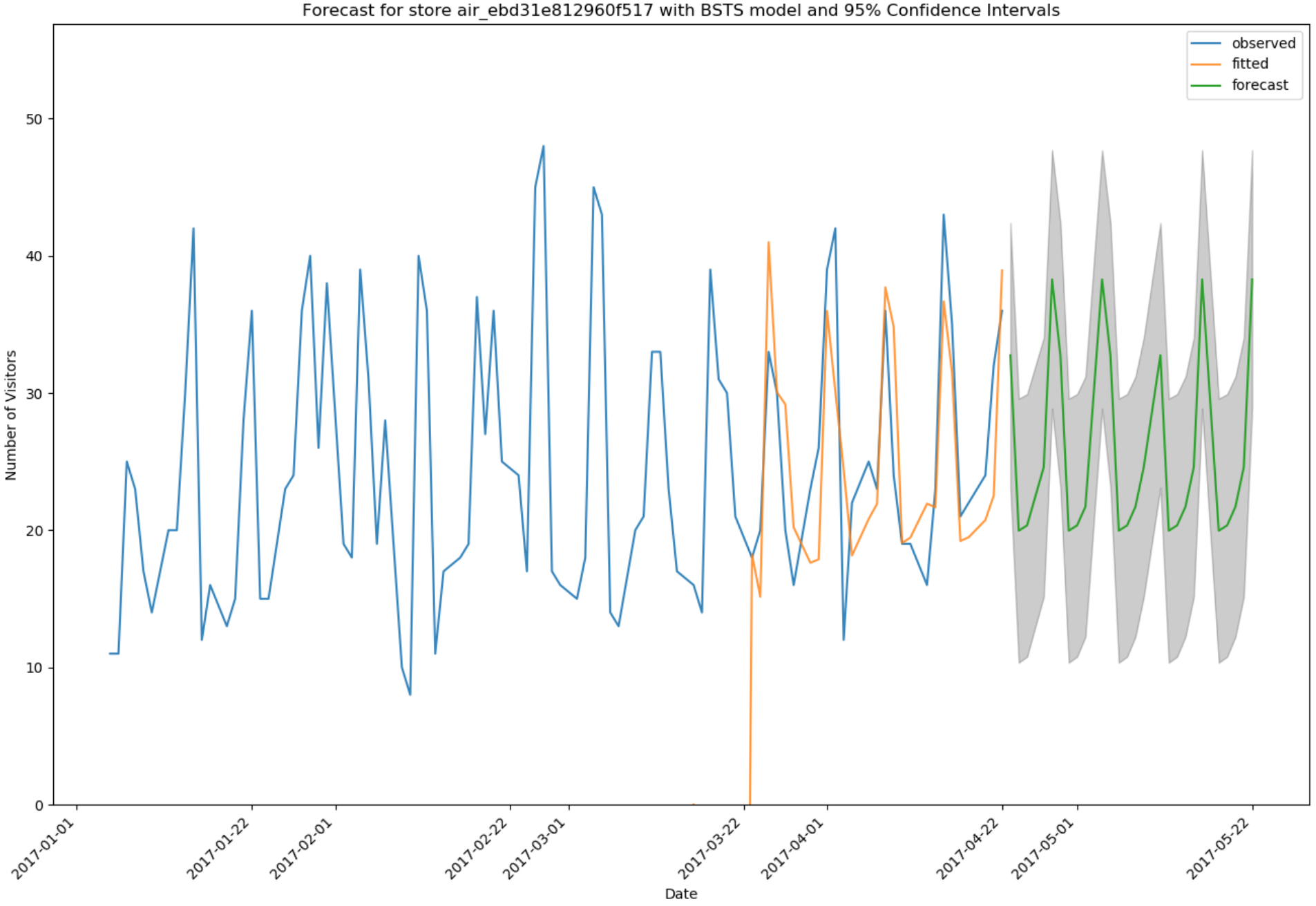
Rolling Window Analysis

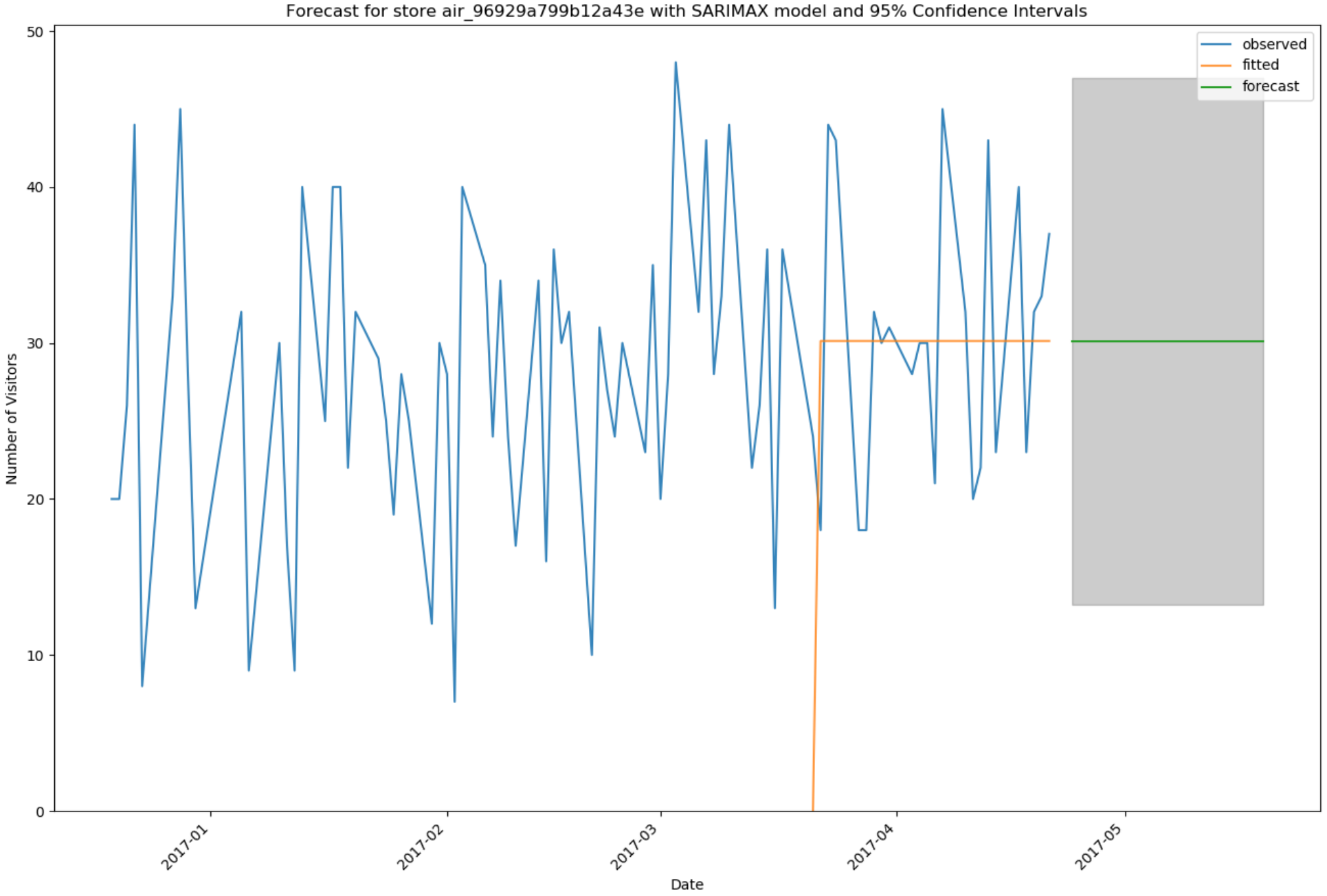


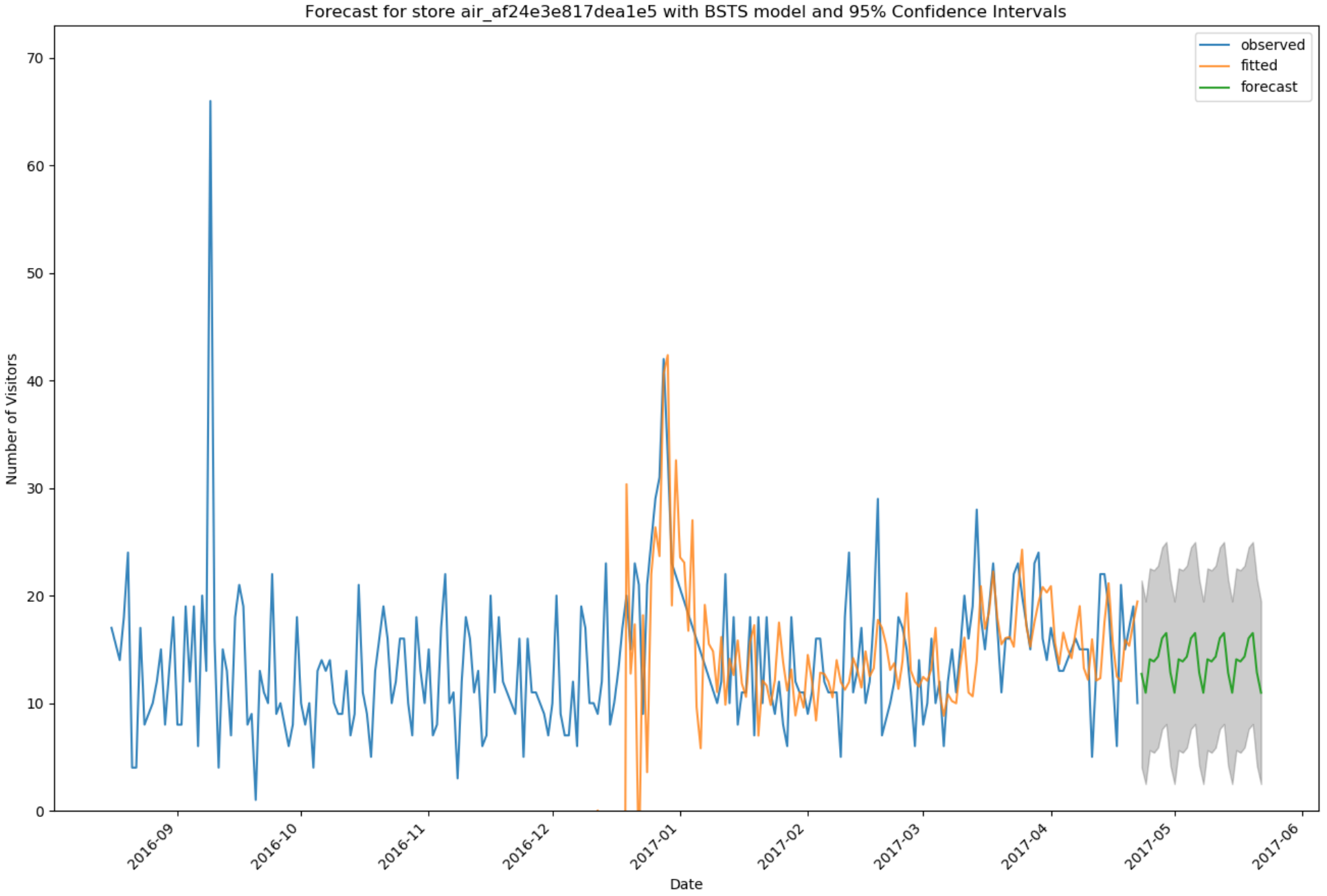
Forecasts

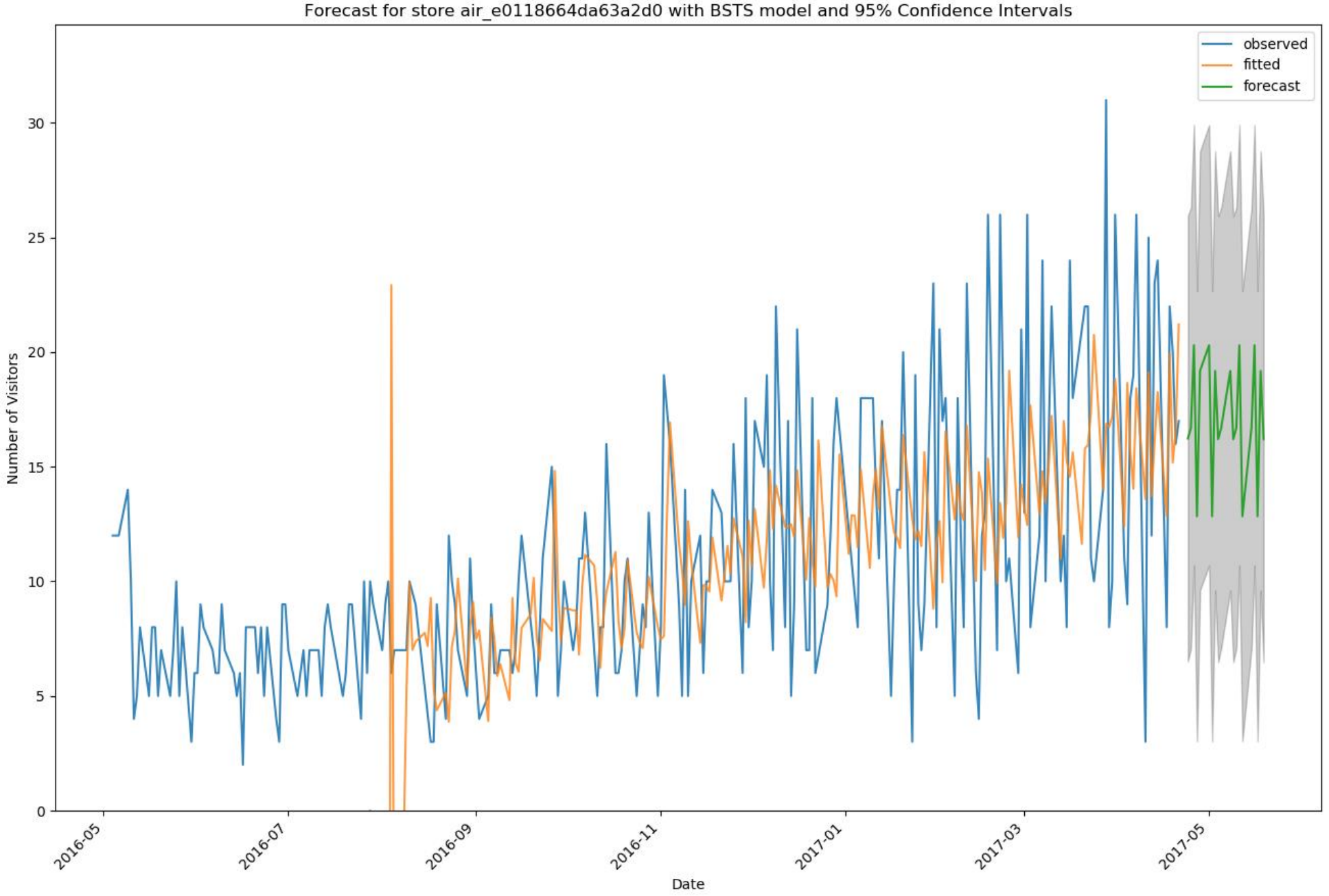


	visit_date	forecast	95_lower	95_upper
127	2017-04-29	84.681772	55.419824	113.943720
128	2017-04-30	51.479521	21.810771	81.148271
129	2017-05-01	30.561021	0.932295	60.189746
130	2017-05-02	24.827111	-4.756797	54.411018
131	2017-05-03	29.949439	0.441101	59.457776
132	2017-05-04	31.172617	1.798376	60.546858
133	2017-05-05	57.939979	28.630746	87.249211
134	2017-05-06	84.681772	55.419824	113.943720
135	2017-05-07	51.479521	21.810771	81.148271
136	2017-05-08	30.561021	0.932295	60.189746
137	2017-05-09	24.827111	-4.756797	54.411018
138	2017-05-10	29.949439	0.441101	59.457776
139	2017-05-11	31.172617	1.798376	60.546858
140	2017-05-12	57.939979	28.630746	87.249211
141	2017-05-13	84.681772	55.419824	113.943720
142	2017-05-14	51.479521	21.810771	81.148271
143	2017-05-15	30.561021	0.932295	60.189746
144	2017-05-16	24.827111	-4.756797	54.411018
145	2017-05-17	29.949439	0.441101	59.457776
146	2017-05-18	31.172617	1.798376	60.546858


Forecast for store air_ebd31e812960f517 with BSTS model and 95% Confidence Intervals
observed
fitted
forecast
Number of Visitors
0
10
20
30
40
50
2017-01-01
2017-01-22
2017-02-01
2017-02-22
2017-03-01
2017-03-22
2017-04-01
2017-04-22
2017-05-01
2017-05-22
Date


Forecast for store air_96929a799b12a43e with SARIMAX model and 95% Confidence Intervals
observed
fitted
forecast
Number of Visitors
Date
0
10
20
30
40
50
2017-01
2017-02
2017-03
2017-04
2017-05


Forecast for store air_af24e3e817dea1e5 with BSTS model and 95% Confidence Intervals
observed
fitted
forecast
Number of Visitors
Date
0
10
20
30
40
50
60
70
2016-09
2016-10
2016-11
2016-12
2017-01
2017-02
2017-03
2017-04
2017-05
2017-06


Forecast for store air_e0118664da63a2d0 with BSTS model and 95% Confidence Intervals
observed
fitted
forecast
Number of Visitors
0
5
10
15
20
25
30
2016-05
2016-07
2016-09
2016-11
2017-01
2017-03
2017-05
Date

Future Work

Possible Enhancements

- Add user interface
- Read data from database
- Parallelize with pySpark
- Add multivariate models

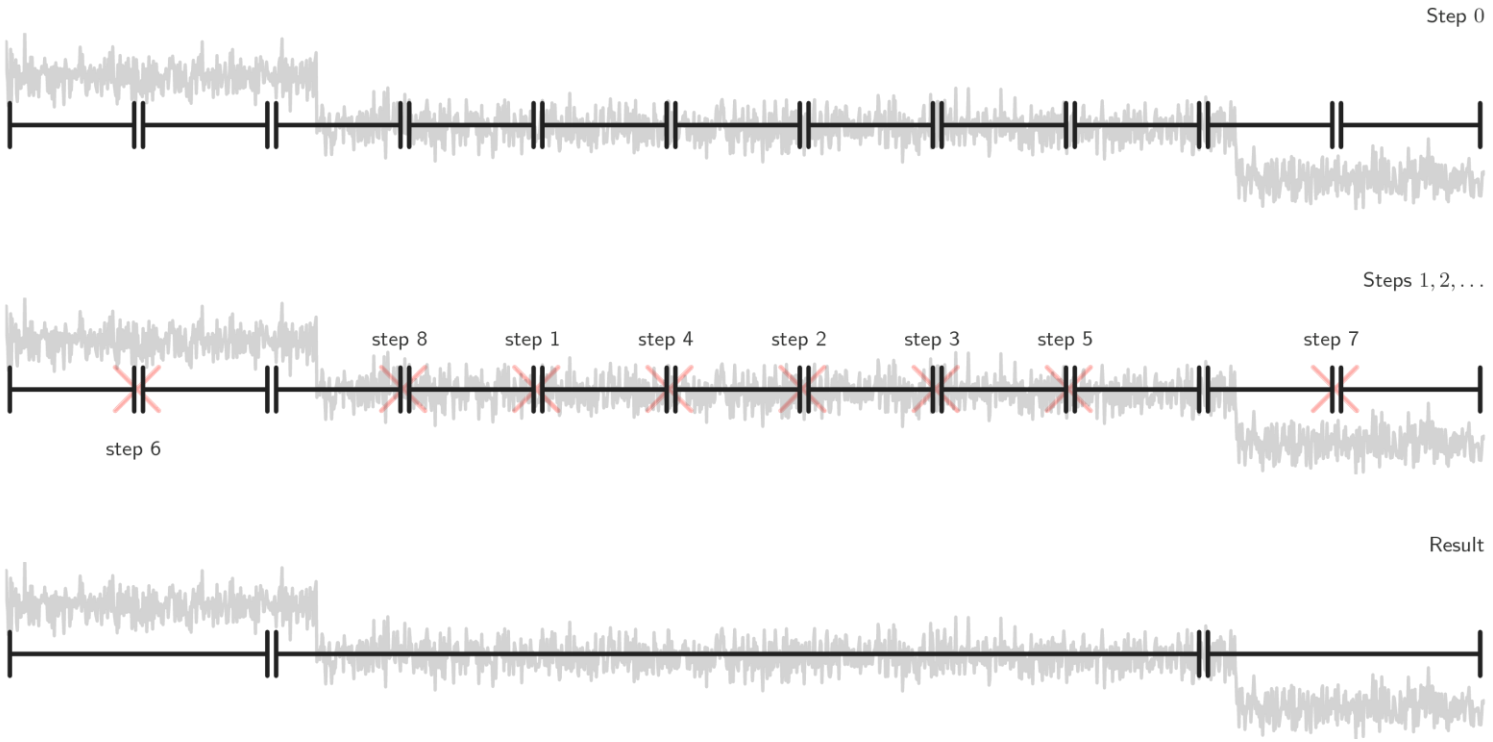
References

- Amin, H. (2018, October 2). Restaurant Food Waste is Killing Your Profit Margin—Here's How to Fix It. Retrieved from ChefHero | The Online Marketplace for Restaurant Supplies: https://www.chefhero.com/blog/restaurant-food-waste-profit-margin
- Hall, M. (2018, September 5). How to Efficiently Manage Labor Costs for Your Restaurant. Retrieved from Restaurant Insider: https://upserve.com/restaurant-insider/labor-cost-guidelines-restaurant/
- Jong, C. d. (2017, November 30). 3 Reasons Your Restaurant Employee Schedule is Costing You Money. Retrieved from Toast POS: https://pos.toasttab.com/blog/restaurant-employee-schedule-cost
- Lachapelle, J. (2018, October 4). The Restaurant Turnover Rate is Astronomical. Here's Why (And How to Fix It). Retrieved from ChefHero | The Online Marketplace for Restaurant Supplies: https://www.chefhero.com/blog/restaurant-turnover-rate
- Nau, R. (n.d.). Identifying the numbers of AR or MA terms in an ARIMA model. Retrieved from Statistical forecasting: notes on regression and time series analysis: https://people.duke.edu/~rnau/411arim3.htm
- Nau, R. (n.d.). Identifying the order of differencing in an ARIMA model. Retrieved from Statistical forecasting: notes on regression and time series analysis: https://people.duke.edu/~rnau/411arim2.htm
- Seth, Y. (2018, January 19). Understanding Time Series Modelling and Forecasting – Part 1. Retrieved from Let the Machines Learn: https://yashuseth.blog/2018/01/19/time-series-analysis-forecasting-modelling-arima/
- Smith, T. G. (n.d.). pmdarima: ARIMA estimators for Python. Retrieved from pmdarima 1.2.1 documentation: https://www.alkaline-ml.com/pmdarima/index.html

Appendix

Bottom-up Segmentation

- Complexity: $O(n \log n)$



Source: http://ctruong.perso.math.cnrs.fr/ruptures-docs/build/html/detection/bottomup.html

Bias-corrected Akaike information criterion (AICc)

- Estimates information lost
- Lower is better
- Balances complexity and goodness of fit
- Corrects for preference of AIC to select model with *n* parameters

Additional resources:

- http://pages.stern.nyu.edu/~churvich/Forecasting/Handouts/AICC.pdf