Web Traffic Analysis

Santa Monica Government Websites

Rahul Purushottam Gaonkar (rpg283)

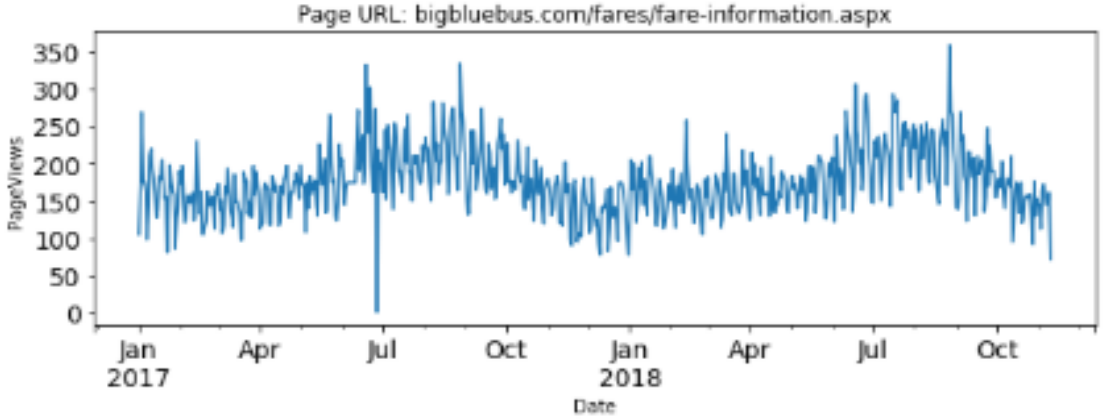
Bhushan Manohar Newalkar (bmn258)

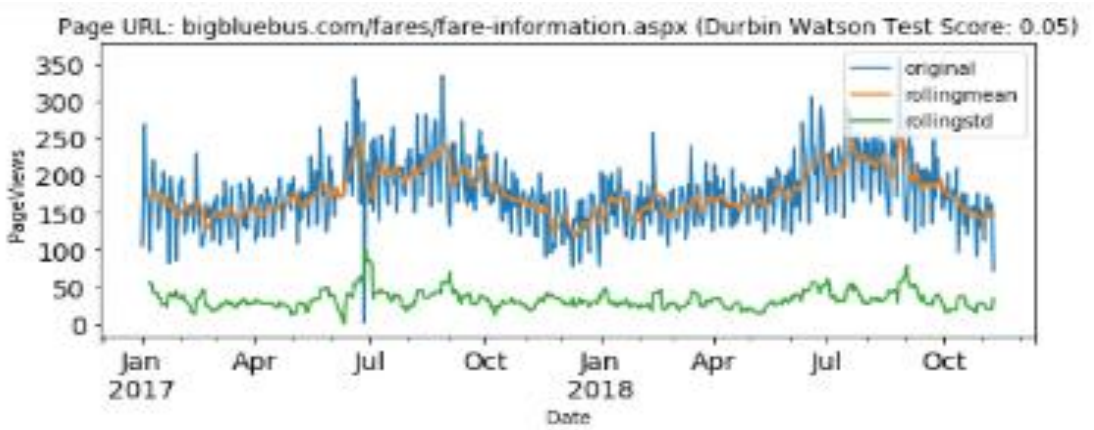
Background and Problem Statement

- Dataset is obtained from ‘Santa Monica Open Data’ which is collected using google analytics
- Goal is to forecast pageviews of top 20 Santa Monica government web pages using Time Series modeling techniques
- Forecast can help to take decisions regarding advertisements (to attract sponsors), server maintenance, etc

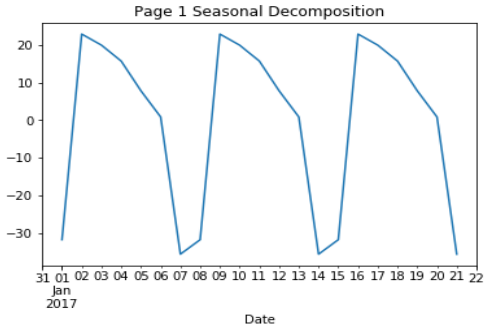
Work done so far

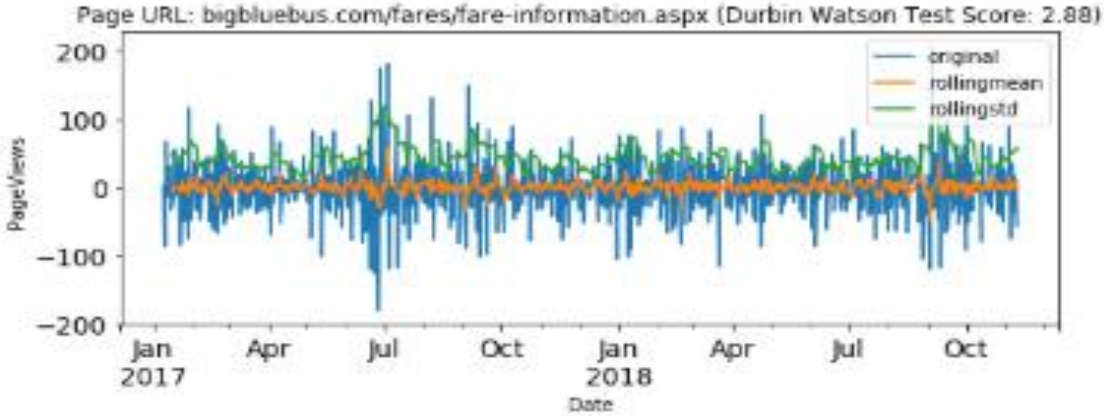
- Data pre-processing
- Identifying and removing trend and seasonality from the data
- ACF and PACF analysis to determine AR, MA, SAR, and SMA parameters for seasonal ARIMA model (SARIMAX)
- Seasonal ARIMA modeling with TimeSeriesSplit cross validator (10 folds)
- Plotting residuals and qq-plot
- Forecasting and Evaluation (MAE and MFE)

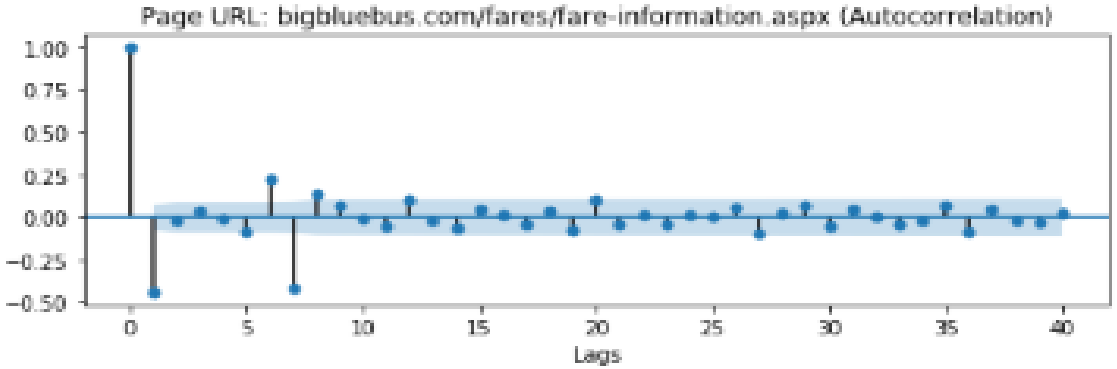


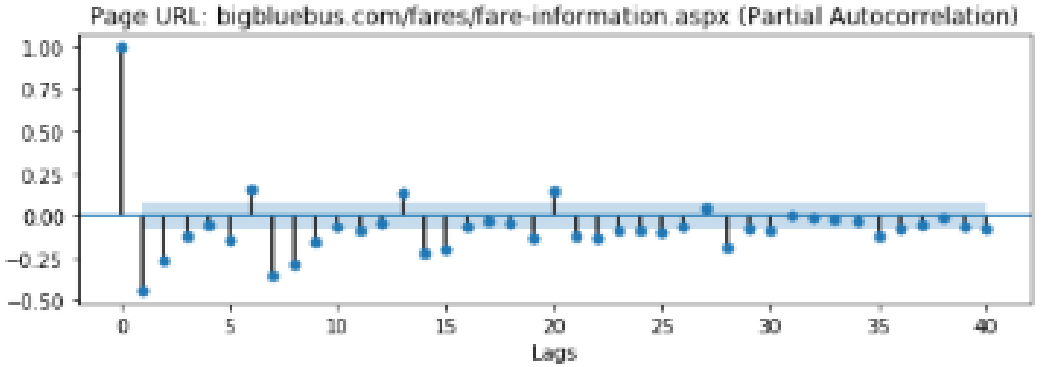


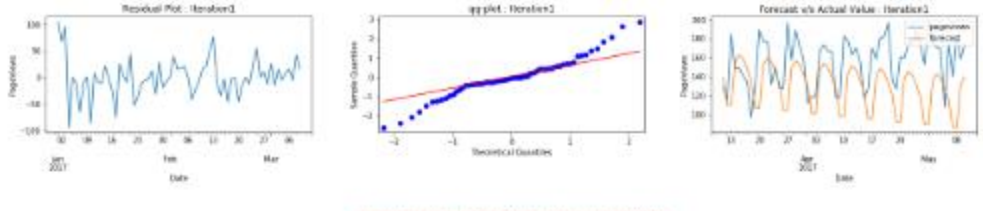

```
# Showing the seasonal component for Page 1 (Page URL: bigbluebus.com/fares/fare-information.aspx)
# We can see that the pattern is repeated after every 7 days i.e. a week
decomposition = sd.seasonal_decompose(page_data(0).pageviews)
seasonal = decomposition.seasonal.head(21).plot(title='Page 1 Seasonal Decomposition')
xlabel = seasonal.set_xlabel("Date")
```

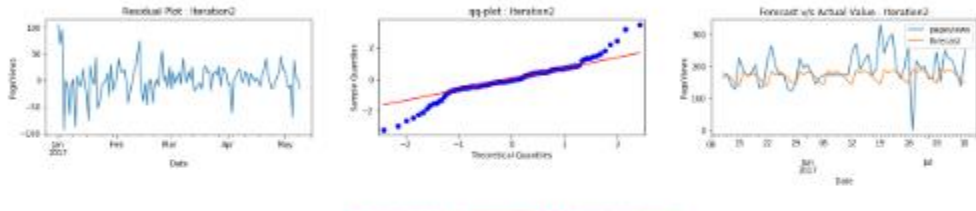


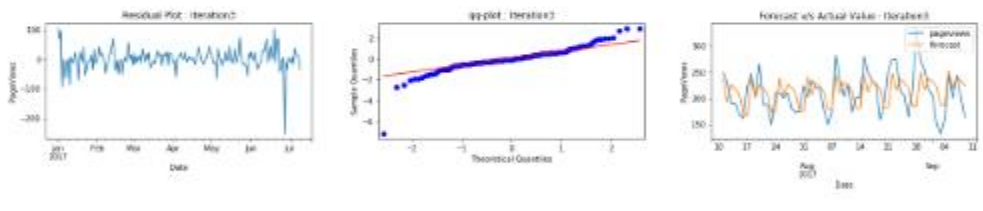


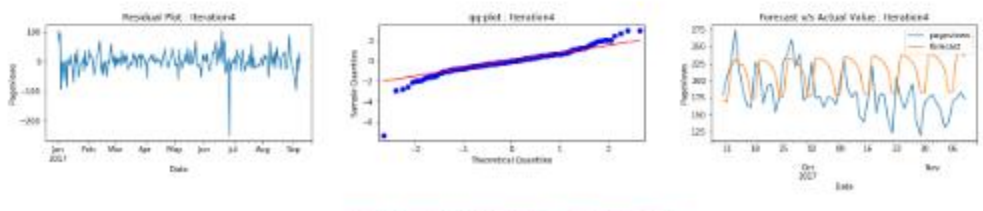


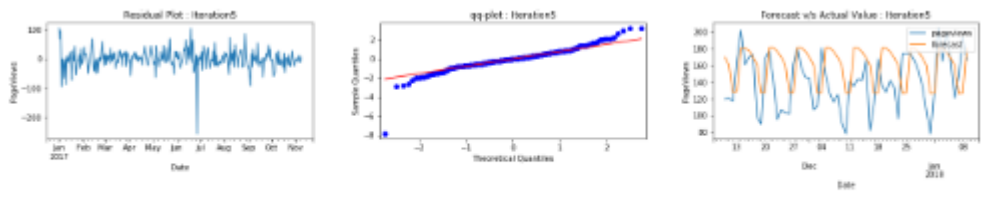


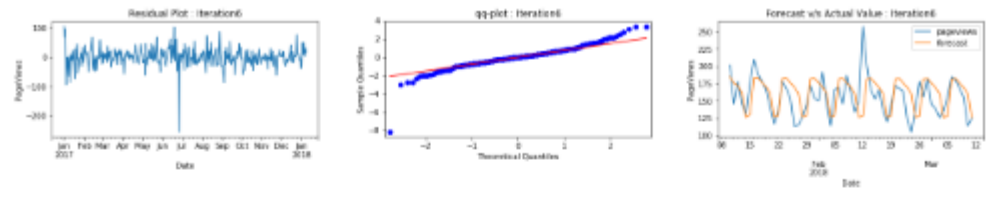

Residual Plot : Iteration1
qq-plot : Iteration1
Forecast v/s Actual Value : Iteration1
pageviews
Forecast
Pageviews
Sample Quantiles
Theoretical Quantiles
Date

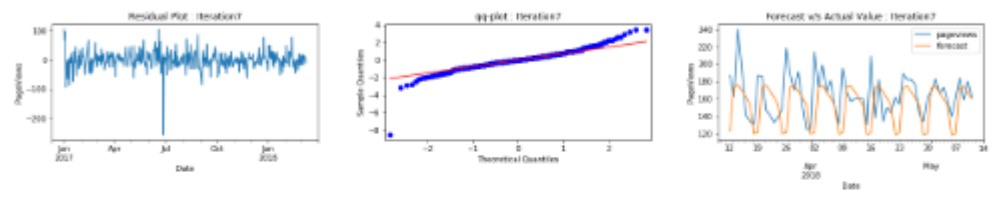

Residual Plot : Iteration2
qq-plot : Iteration2
Forecast v/s Actual Value : Iteration2
pageviews
Forecast
Pageviews
Sample Quantiles
Theoretical Quantiles
Date

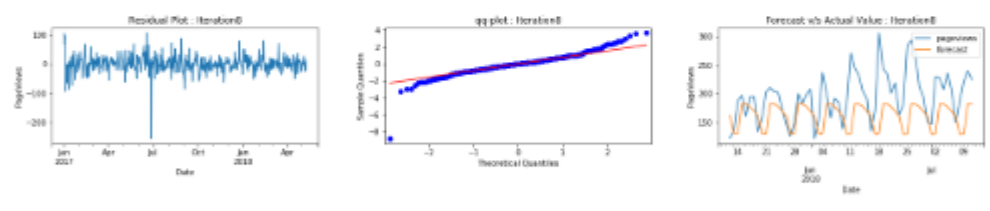

Residual Plot : Iteration3
qq-plot : Iteration3
Forecast v/s Actual Value : Iteration3
pageviews
Forecast
Pageviews
Sample Quantiles
Theoretical Quantiles
Date

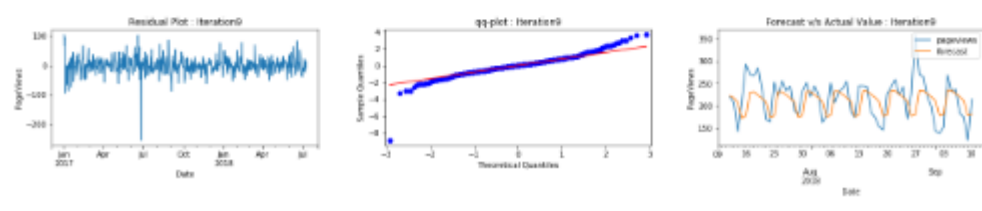

Residual Plot : Iteration4
qq-plot : Iteration4
Forecast v/s Actual Value : Iteration4
pageviews
Forecast
Pageviews
Sample Quantiles
Theoretical Quantiles
Date

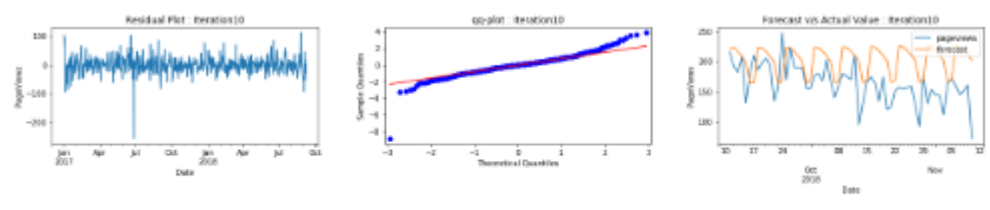

Residual Plot : Iteration5
qq-plot : Iteration5
Forecast v/s Actual Value : Iteration5
pageviews
Forecast
Pageviews
Sample Quantiles
Theoretical Quantiles
Date


Residual Plot : Iteration6
qq-plot : Iteration6
Forecast v/s Actual Value : Iteration6
pageviews
Forecast
Pageviews
Sample Quantiles
Theoretical Quantiles
Date


Residual Plot : Iteration7
qq-plot : Iteration7
Forecast v/s Actual Value : Iteration7
pageviews
Forecast
Pageviews
Sample Quantiles
Theoretical Quantiles
Date


Residual Plot : Iteration8
qq-plot : Iteration8
Forecast v/s Actual Value : Iteration8
pageviews
Forecast
Pageviews
Sample Quantiles
Theoretical Quantiles
Date


Residual Plot : Iteration9
qq-plot : Iteration9
Forecast v/s Actual Value : Iteration9
pageviews
Forecast
Pageviews
Sample Quantiles
Theoretical Quantiles
Date


Residual Plot : Iteration10
qq-plot : Iteration10
Forecast v/s Actual Value : Iteration10
pageviews
Forecast
Pageviews
Sample Quantiles
Theoretical Quantiles
Date

Future Scope

- Identifying the importance of the confounding variables with respect to pageviews
- We have identified following confounding variables:
  - avg_page_load_time
  - avg_time_on_page
  - bounces

Thank You