Forecasting Average Weekly Traffic Volume

CSCI E-116 Final Project

Rebeca Ansar

2022-12-10

Purpose

To forecast the average weekly traffic volume for I-94 westbound traffic between Minneapolis and St Paul, MN crossing MN D0T ATR station 301, which is roughly midway between the two locations.

Several time series modeling methods were implemented and compared against each other in order to select the optimal forecasting model with minimal error.

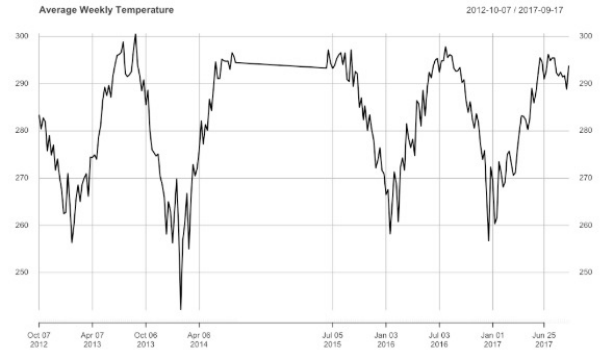
Data Processing

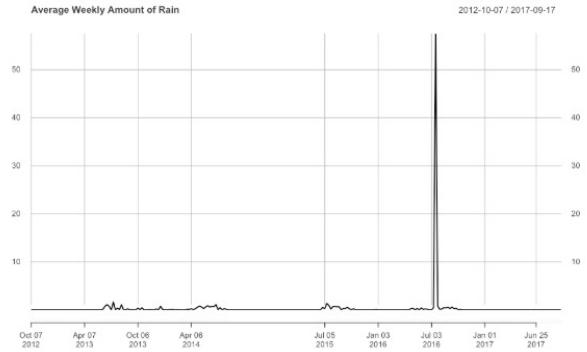
Variables used during model-building:

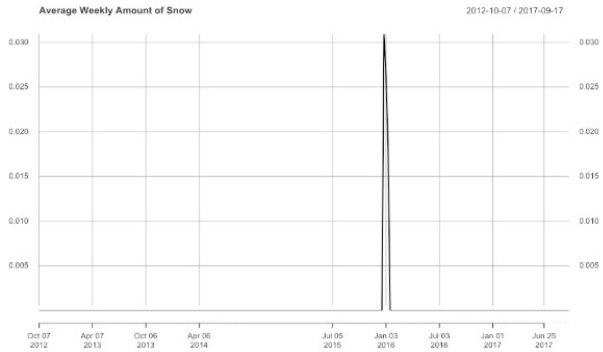
- avg_weekly_temp- average temp in kelvin
- avg_weekly_rain_1h- average weekly amount of hourly rain (mm)
- avg_weekly_snow_1h- average weekly amount of hourly snow (mm)
- avg_weekly_clouds_all- average weekly percentage of cloud cover
- avg_weekly_traffic_volume- average weekly traffic volume crossing MN D0T ATR station 301 measured hourly

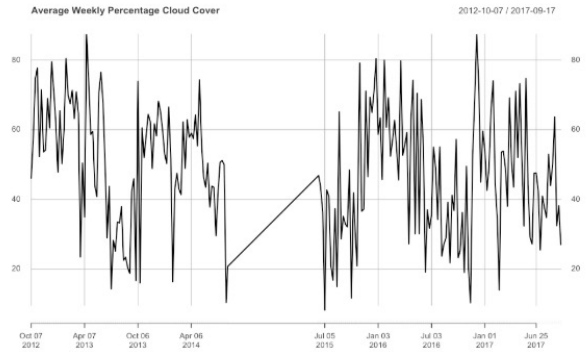
80% train / 20% test split:

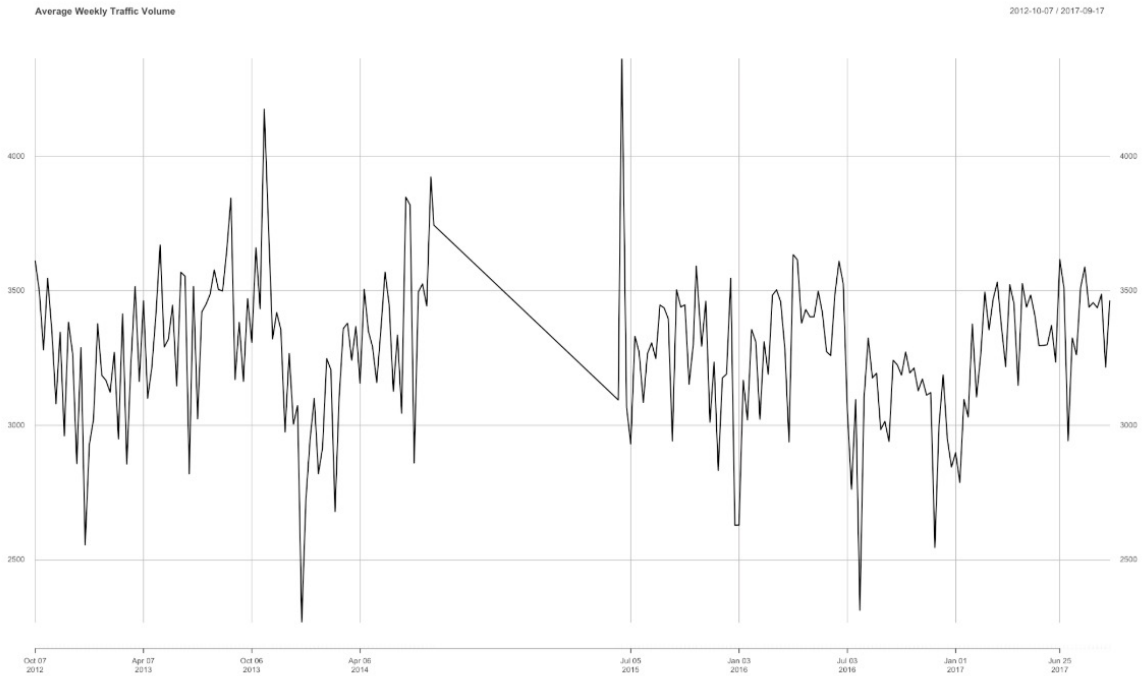
- 269 total rows
- 215 rows in training data
- 54 rows in test data











Evaluating for Stationarity

- None of the plots indicate that variable is non-stationarity
- Results of Augmented Dickey-Fuller unit root test were consistent with stationarity for each variable.

The benchmark: a naive time series model

- Provided a measure of baseline test accuracy

```
Call:
lm(formula = avg_weekly_traffic_volume ~ Lag.1, data = all_avg_weekly_predictors_train)

Residuals:
    Min      1Q  Median      3Q     Max
-947.75 -142.42   26.85  161.37 1141.87

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 2360.0843   216.0157  10.926  < 2e-16 ***
Lag.1          0.2783     0.0658   4.229 3.49e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 269.5 on 212 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.0778,    Adjusted R-squared:  0.07345
F-statistic: 17.89 on 1 and 212 DF,  p-value: 3.49e-05
```

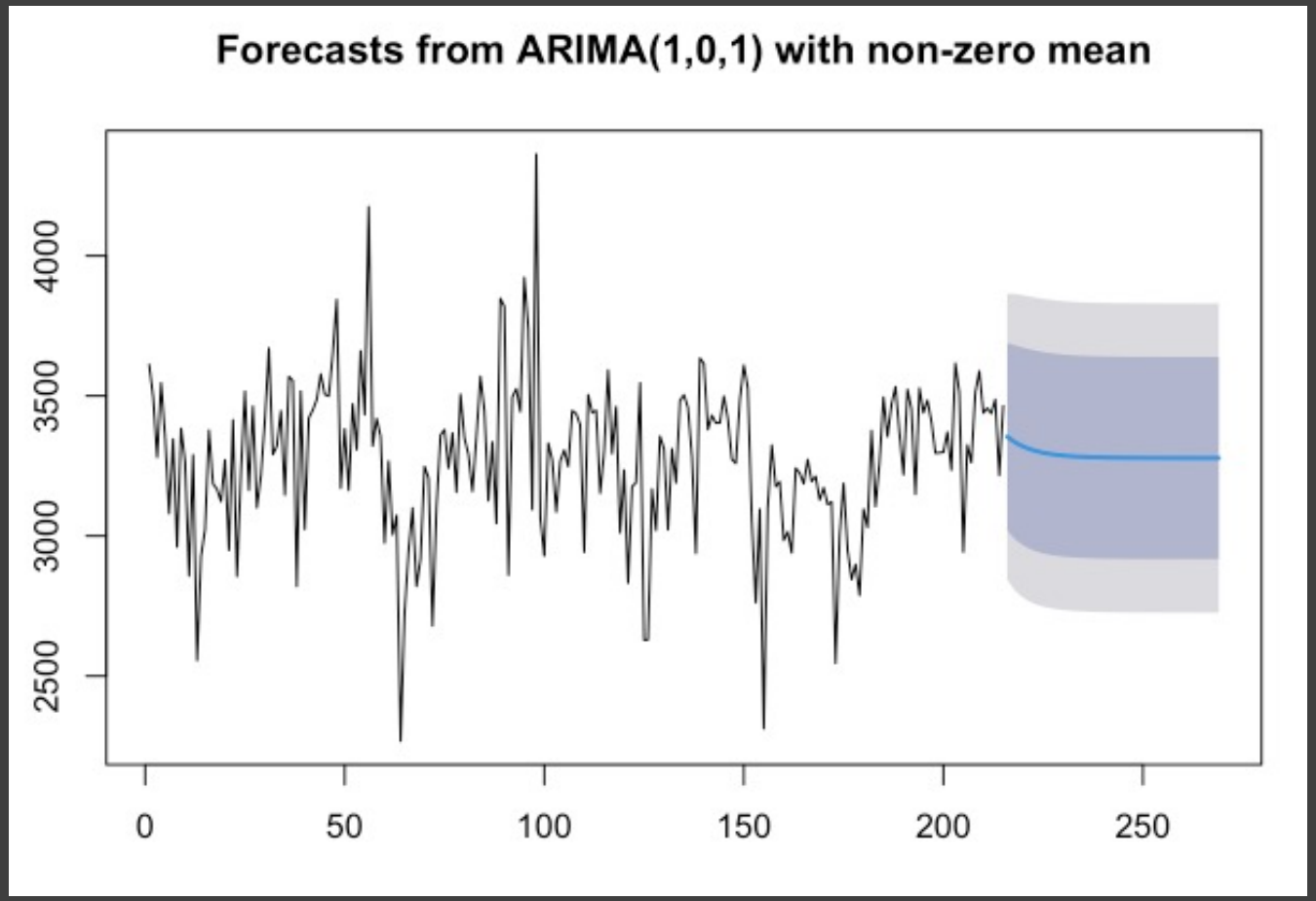
Using variable selection to build linear models

Best model by adjusted R-squared

```
Call:
lm(formula = avg_weekly_traffic_volume ~ avg_weekly_temp + avg_weekly_snow_1h +
    Lag.1 + Lag.2 + Lag.3, data = all_avg_weekly_predictors_train)

Residuals:
    Min      1Q  Median      3Q     Max
-919.61 -116.46   21.59  140.59  879.08

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)         7.312e+02  4.204e+02   1.739  0.08351 .
avg_weekly_temp     4.569e+00  1.686e+00   2.709  0.00731 **
avg_weekly_snow_1h -1.579e+04  5.762e+03  -2.740  0.00669 **
Lag.1               1.086e-01  6.836e-02   1.589  0.11352
Lag.2               1.348e-01  6.816e-02   1.978  0.04924 *
Lag.3               1.412e-01  6.779e-02   2.082  0.03855 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 252.1 on 206 degrees of freedom
  (3 observations deleted due to missingness)
Multiple R-squared:  0.2137,    Adjusted R-squared:  0.1946
F-statistic:  11.2 on 5 and 206 DF,  p-value: 1.465e-09
```

Best model by BIC

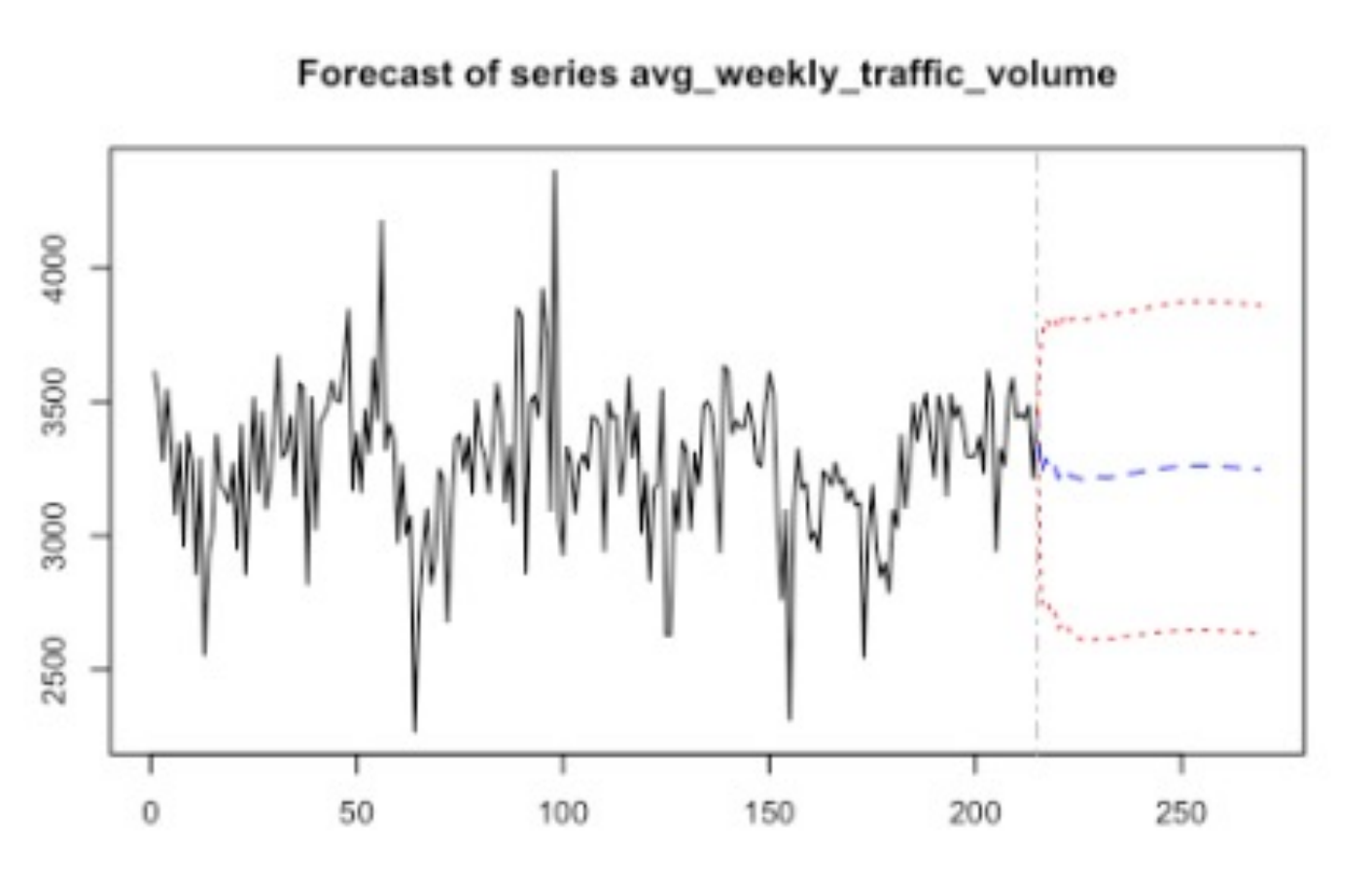
```
Call:
lm(formula = avg_weekly_traffic_volume ~ avg_weekly_temp + avg_weekly_snow_1h +
    Lag.3, data = all_avg_weekly_predictors_train)

Residuals:
     Min       1Q   Median       3Q      Max
-1026.29  -130.95    21.34   148.28   927.29

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)         9.032e+02  4.210e+02   2.146  0.03306 *
avg_weekly_temp     6.297e+00  1.583e+00   3.978 9.59e-05 ***
avg_weekly_snow_1h -1.652e+04  5.818e+03  -2.839  0.00497 **
Lag.3               1.833e-01  6.684e-02   2.743  0.00662 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 255.3 on 208 degrees of freedom
  (3 observations deleted due to missingness)
Multiple R-squared:  0.1853,    Adjusted R-squared:  0.1735
F-statistic: 15.77 on 3 and 208 DF,  p-value: 2.816e-09
```

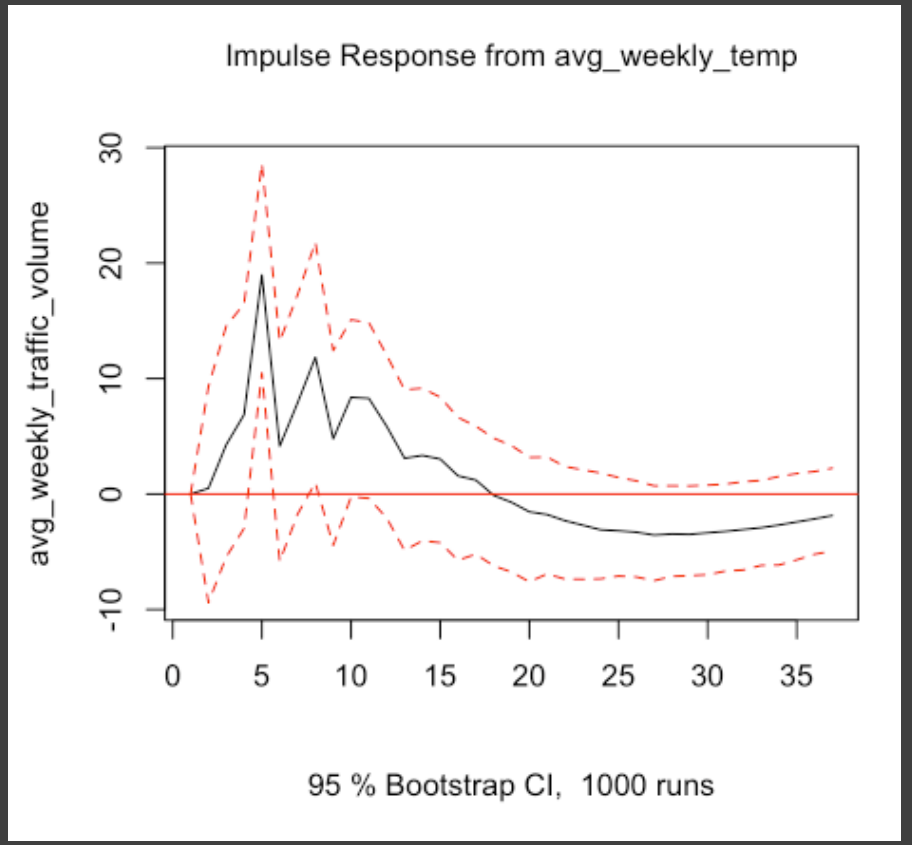



```
Series: arima_train_data
ARIMA(1,0,1) with non-zero mean

Coefficients:
          ar1      ma1       mean
      0.8573  -0.6433  3278.1075
s.e.  0.0600   0.0830    43.2888

sigma^2 = 67574:  log likelihood = -1499.2
AIC=3006.39   AICc=3006.58   BIC=3019.87
```

ARIMA model

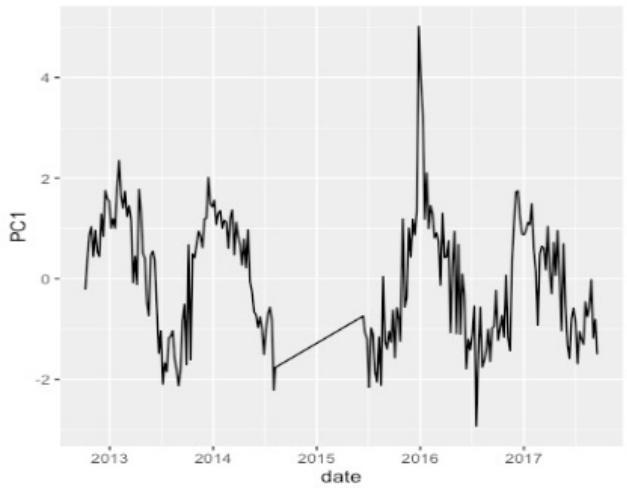
Built using the historical values of the outcome: avg_weekly_traffic_volume

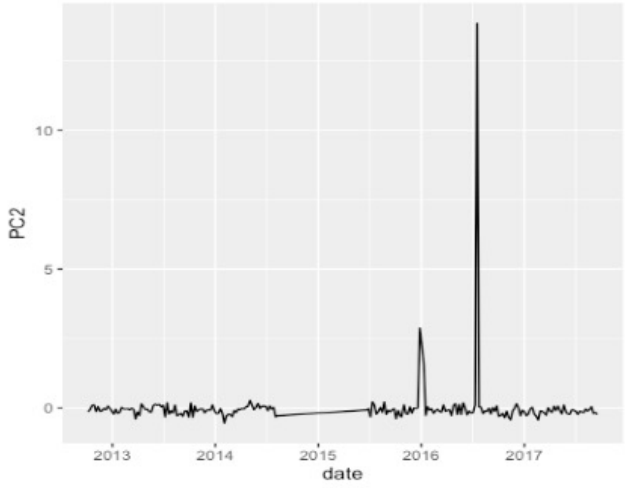


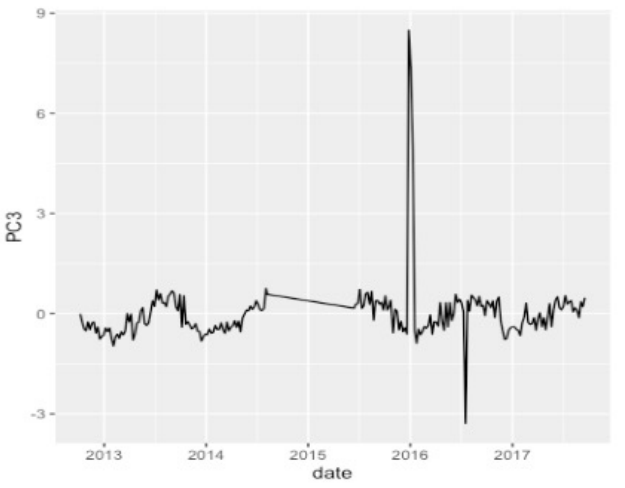


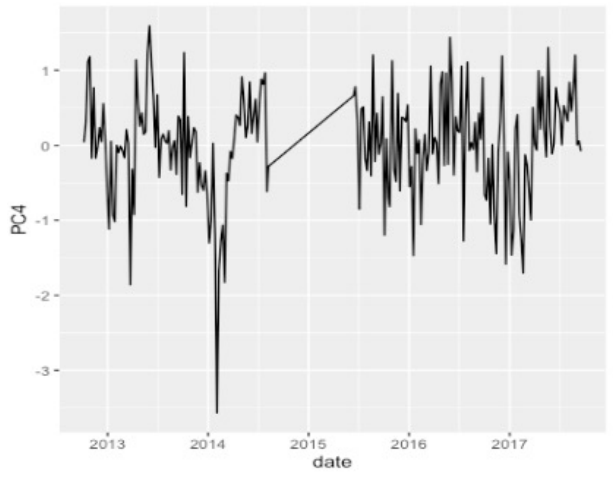
Vector autoregressive model (VAR)

- Built using all 5 numeric variables
- Optimal lag = 8










```
Importance of components:
                          PC1    PC2    PC3    PC4
Standard deviation     1.2406 1.0009 0.9655 0.7258
Proportion of Variance 0.3848 0.2505 0.2331 0.1317
Cumulative Proportion  0.3848 0.6352 0.8683 1.0000
```

Dynamic factor model (DFM)

- Built using 4 principal components (PC's)

1. Linear model that predicts avg_weekly_traffic_volume from all 4 principal components
2. VAR model utilized to forecast PC values for 54 weeks
3. Initial linear model used to forecast avg_weekly_traffic_volume

Results

	test_rmse	models	regressors
3	223.85	best_bic_3var_model	avg_weekly_snow_1h + avg_weekly_temp + lag(avg_weekly_traffic_volume, 3)
2	225.19	best_adjr2_5var_model	avg_weekly_snow_1h + avg_weekly_temp + + lag(avg_weekly_traffic_volume, 1)+ lag(avg_weekly_traffic_volume, 2) + lag(avg_weekly_traffic_volume, 3)
4	228.30	arima_model	auto-ARIMA with avg_weekly_traffic_volume
1	229.11	naive_lag_1_linear_model	lag(avg_weekly_traffic_volume, 1)
5	233.96	var_level_model	VAR with avg_weekly_traffic_volume, avg_weekly_temp, avg_weekly_rain_1h, avg_weekly_snow_1h, avg_weekly_clouds_all
6	235.67	dfm_lm1	DFM with 4 principle components