# Traffic on I-94 in Minnesota

Madeline Chang

November 24, 2024

```
library(fpp3)
library(gt)
library(tidyverse)
library(ggfortify)
library(lubridate)
library(tibbletime)
library(mosaic)
library(fabletools)
library(here)
set.seed(506)
```

## Importing the Data

```
traffic<- read_csv(here::here('ADS 506/Metro_Interstate_Traffic_Volume.csv'))
```

```
## Rows: 48204 Columns: 9
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (3): holiday, weather_main, weather_description
## dbl  (5): temp, rain_1h, snow_1h, clouds_all, traffic_volume
## dttm (1): date_time
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
head(traffic)
```

```
## # A tibble: 6 x 9
##   holiday  temp rain_1h snow_1h clouds_all weather_main weather_description
##   <chr>   <dbl>   <dbl>   <dbl>      <dbl> <chr>        <chr>
## 1 None     288.       0       0         40 Clouds       scattered clouds
## 2 None     289.       0       0         75 Clouds       broken clouds
## 3 None     290.       0       0         90 Clouds       overcast clouds
## 4 None     290.       0       0         90 Clouds       overcast clouds
## 5 None     291.       0       0         75 Clouds       broken clouds
## 6 None     292.       0       0          1 Clear        sky is clear
## # i 2 more variables: date_time <dttm>, traffic_volume <dbl>
```

```
traffic_f<- tibble(traffic %>%
  mutate(temp = (temp * (9/5))-459.67, # changing unit of temperature to Fahrenheit
        time = hms(str_extract(date_time, "\\d+[[::punct::]]\\d+[[::punct::]]\\d+")),
        time = replace(time, is.na(time), period(hours = 0, minutes = 0, seconds = 0)),
        date = as.Date(date_time),
```

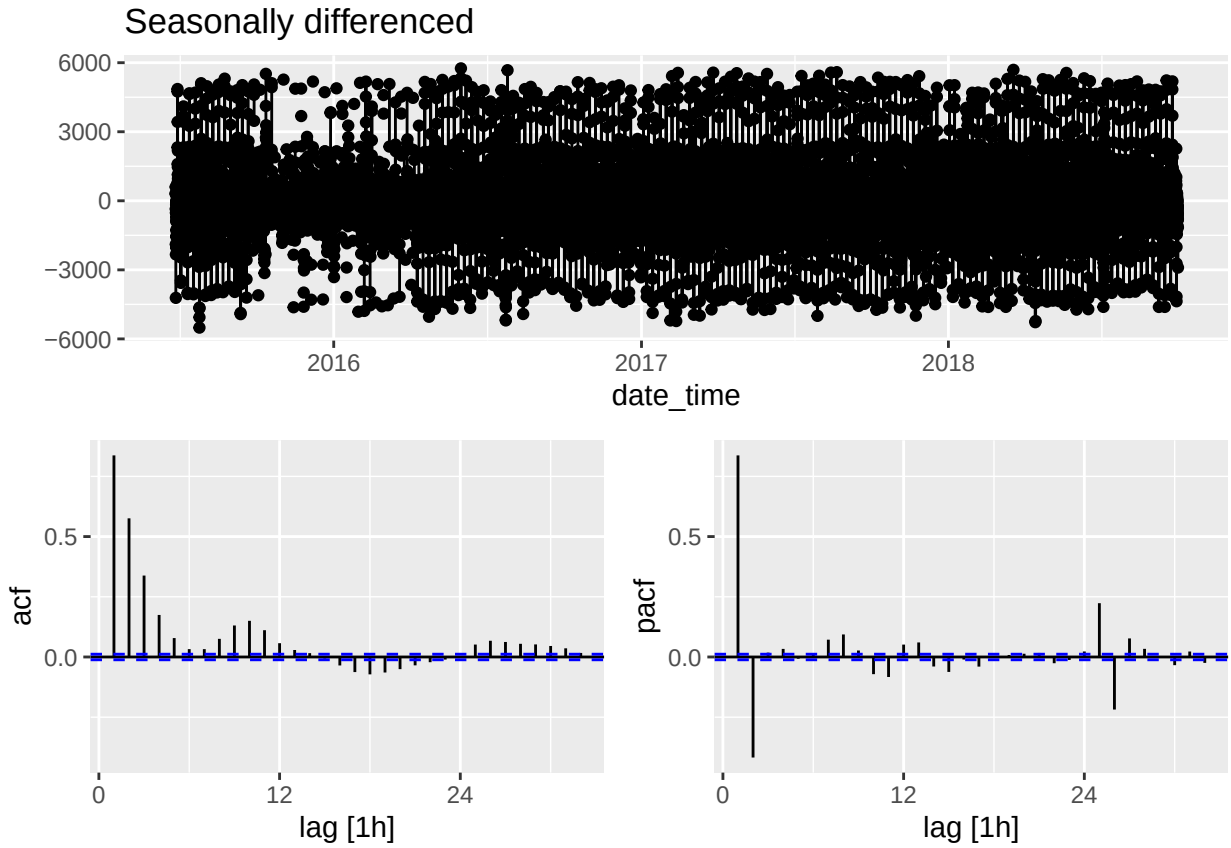
```
        is_holiday = as.factor(ifelse(holiday == 'None', 0, 1)),
        day = factor(weekdays(date), levels = c("Sunday", "Monday", "Tuesday", "Wednesday", "Thursday",
        temp = ifelse(temp < -100, 20.6, temp),
        rain_1h = ifelse(rain_1h > 2500, 21.4, rain_1h)
)
) %>%
  group_by(date_time) %>%
  distinct(traffic_volume, .keep_all = TRUE) %>%
  ungroup()


traffic_f <- traffic_f %>%
  group_by(date) %>%
  # Extend the holiday value to all hours of the day
  mutate(holiday = ifelse(holiday == "None", NA, holiday),  # Convert "None" to NA for proper propagati
         holiday = first(na.omit(holiday))) %>%
  # Replace NA values with "None" if there were no holidays for the day
  mutate(holiday = ifelse(is.na(holiday), "None", holiday)) %>%
  ungroup() %>%
  filter(date >= '2015-06-26') %>%
  as_tsibble(., index = date_time)
```

```
traffic_f |>
    tsibble::fill_gaps() %>%
  gg_tsdisplay(difference(traffic_volume, 24, differences = 1),
               plot_type='partial') +
  labs(title="Seasonally differenced", y="")
```

```
## Warning: Removed 24 rows containing missing values or values outside the scale range
## (`geom_line()`).
```

```
## Warning: Removed 3076 rows containing missing values or values outside the scale range
## (`geom_point()`).
```




```
traffic_train<- traffic_f %>%
  filter_index('2015-06-26'~'2018-08-31') %>%
  tsibble::fill_gaps()

traffic_val<- traffic_f %>%
  filter_index('2018-09-01'~.) %>%
  tsibble::fill_gaps()
```

```
traffic_fit<- traffic_train %>%
  fabletools::model(
    model_2 = ARIMA(traffic_volume ~ 0 +
                      pdq(0, 1, 1) + PDQ(0, 1, 0)),
    model_3 = ARIMA(traffic_volume~ 1+
                      date_time + holiday + pdq(1, 1, 0)+ PDQ(0, 1, 0)),
    model_4 = ARIMA(traffic_volume~ 1+
                      date_time + holiday + pdq(1, 1, 0)+ PDQ(0, 1, 0)),
    model_5 = ARIMA(traffic_volume~ 1+
                      date_time + weather_main + pdq(1, 1, 0)+ PDQ(0, 1, 0))
  )

traffic_naive<- traffic_train %>%
  fabletools::model(
    model_1 = ARIMA(traffic_volume)
  )
```

```
traffic_fit %>%
  accuracy()  # best RMSE is model2
```

```
## # A tibble: 4 x 10
##   .model  .type            ME    RMSE    MAE   MPE  MAPE    MASE   RMSSE    ACF1
##   <chr>   <chr>         <dbl>   <dbl>  <dbl> <dbl> <dbl>   <dbl>   <dbl>   <dbl>
## 1 model_2 Training      0.143  5.28e2   341.   NaN   Inf   0.614 5.20e-1 1.91e-5
## 2 model_3 Training      4.44   4.97e3   618.   Inf   Inf   1.11  4.90e+0 5.13e-1
## 3 model_4 Training      4.44   4.97e3   618.   Inf   Inf   1.11  4.90e+0 5.13e-1
## 4 model_5 Training -20186.     2.84e6 61556.  -Inf   Inf 111.    2.80e+3 5.07e-1
```

```
traffic_naive %>%
  accuracy()
```

```
## # A tibble: 1 x 10
##   .model  .type       ME  RMSE   MAE   MPE  MAPE  MASE RMSSE     ACF1
##   <chr>   <chr>    <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>    <dbl>
## 1 model_1 Training 0.305  473.  311.  -Inf   Inf 0.559 0.466 -0.00250
```

```
traffic_f2<- traffic_f %>%
  tsibble::fill_gaps() %>%
  fabletools::model(
    model_2 = ARIMA(traffic_volume ~ 0 +
                      pdq(0, 1, 1) + PDQ(0, 1, 0))
  )

traffic_n2<- traffic_f %>%
  tsibble::fill_gaps() %>%
  fabletools::model(
    model_1 = ARIMA(traffic_volume)
  )
```

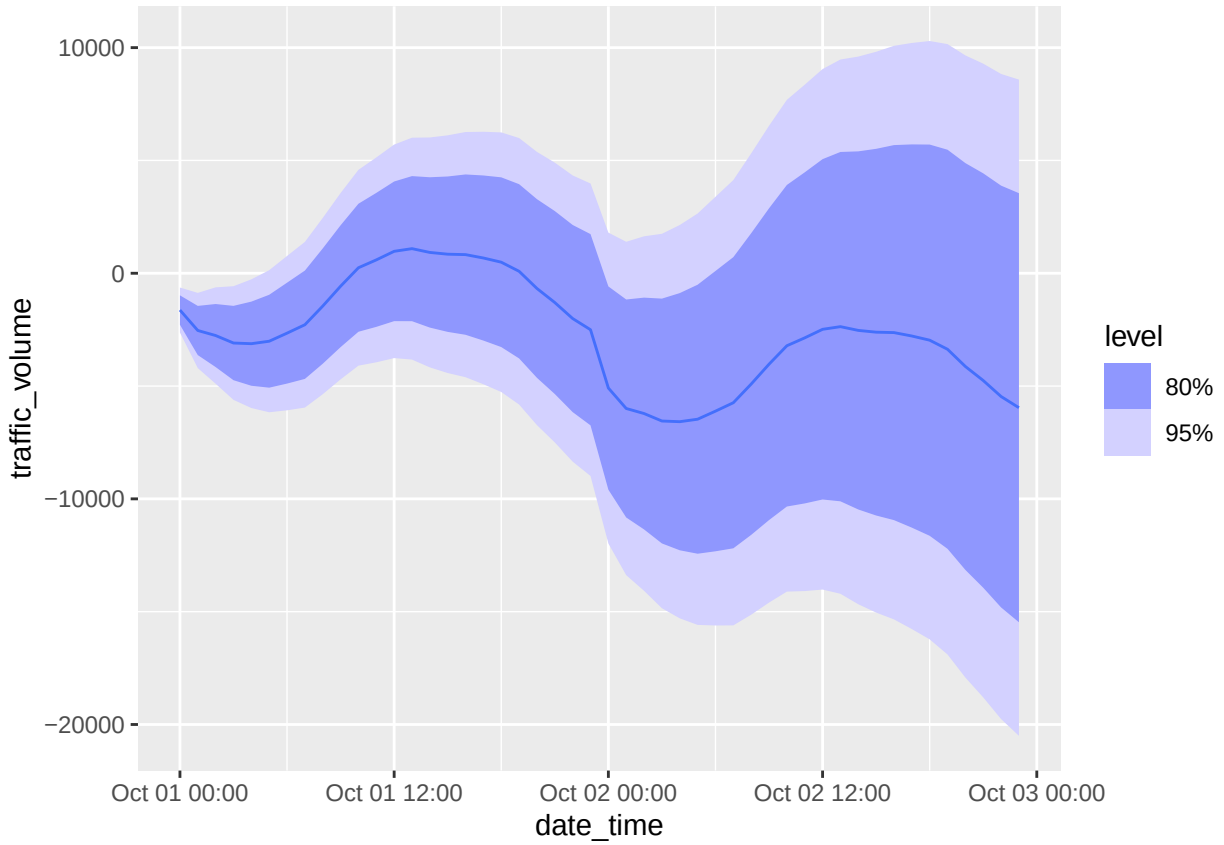
```
traffic_f2 %>%
  accuracy() %>%
  as_tibble() %>%
  select(.model, RMSE, ACF1)
```

```
## # A tibble: 1 x 3
##   .model   RMSE     ACF1
##   <chr>   <dbl>    <dbl>
## 1 model_2  527. 0.000914
```

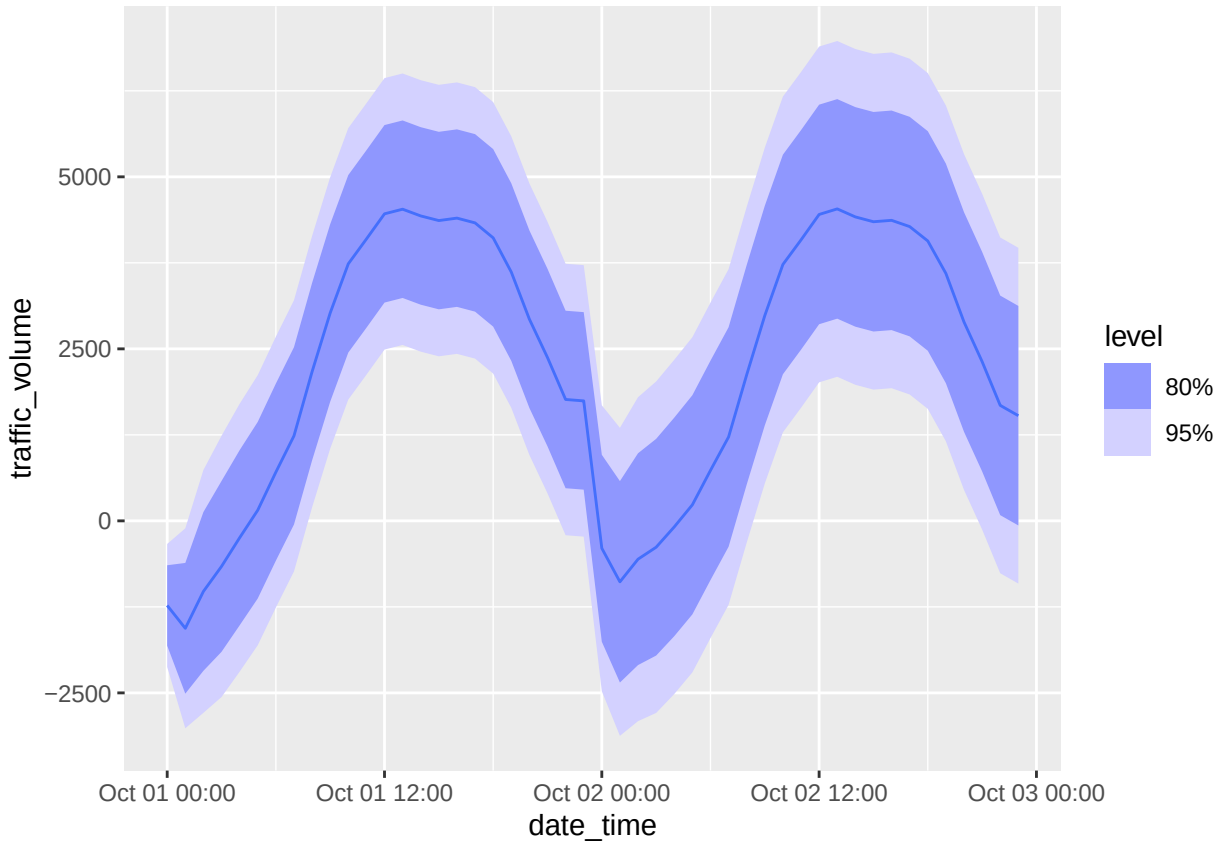
```
traffic_n2 %>%
  accuracy() %>%
  as_tibble() %>%
  select(.model, RMSE, ACF1)
```

```
## # A tibble: 1 x 3
##   .model   RMSE     ACF1
##   <chr>   <dbl>    <dbl>
## 1 model_1  472. -0.00241
```

```
traffic_f2 %>%
  forecast() %>%
  autoplot()
```




```
traffic_n2 %>%
  forecast() %>%
  autoplot()
```


traffic_volume
5000
2500
0
−2500
Oct 01 00:00
Oct 01 12:00
Oct 02 00:00
Oct 02 12:00
Oct 03 00:00
date_time
level
80%
95%