Daily Demand Forecast

For Wheelchair-Accessible Taxies in Seoul

21900749 SEUNGRI CHOE

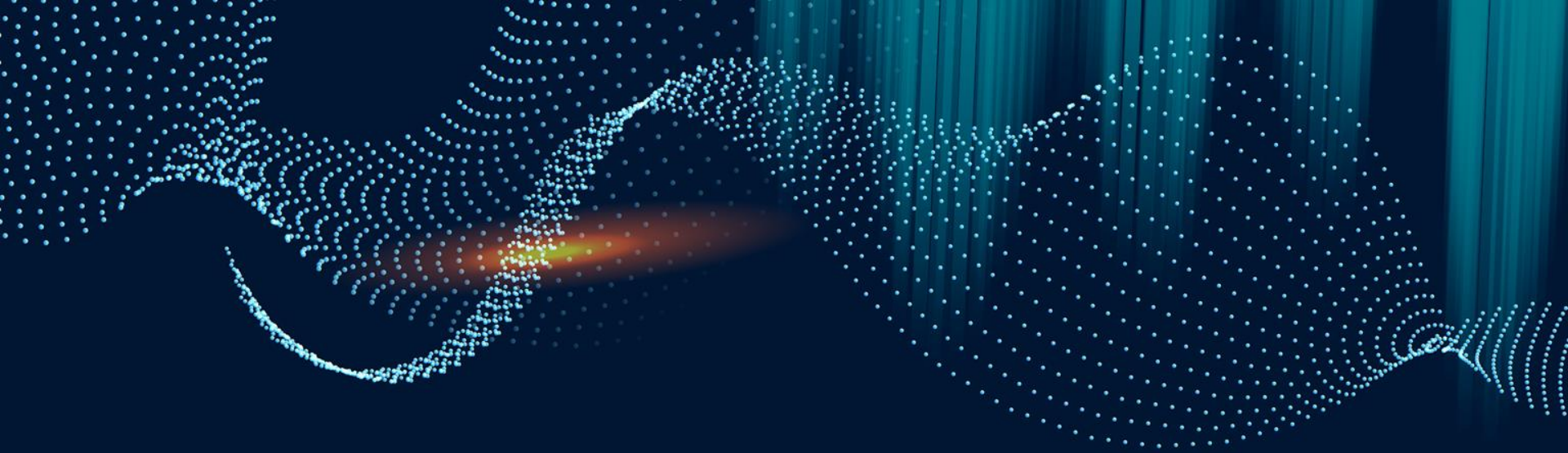
01 GOAL

02 DATA

03 MODELING

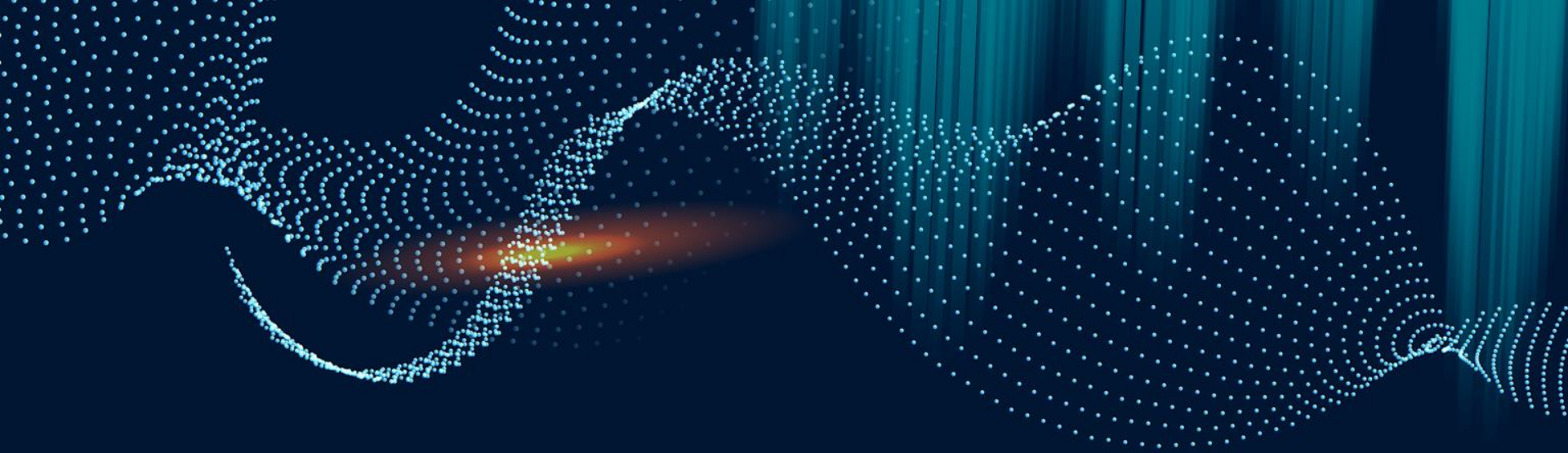
04 Forecast

05 CONCLUSION



01 | GOAL

The goal is to predict
the future daily demand
for wheelchair-accessible taxies in Seoul
to ensure the appropriate provision of
taxies can be done



02 | DATA

Data Acquisition

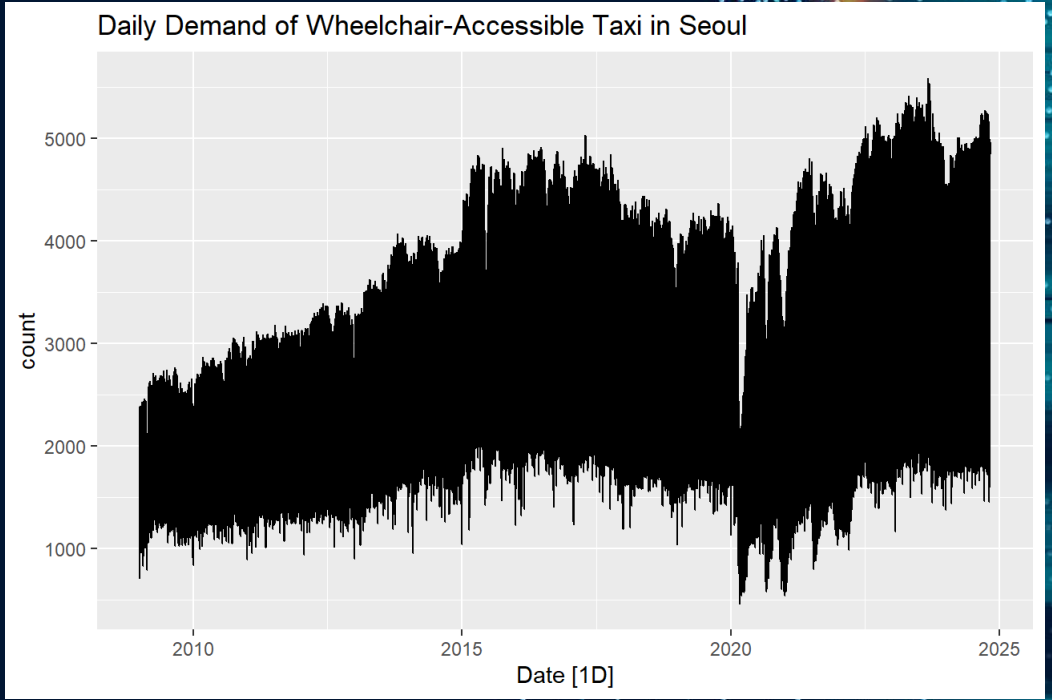
Wheelchair-accessible taxies infomation in Seoul
(서울 장애인콜택시 이용 정보)
2009/01/01 ~ 2024/10/31
서울 열린데이터 광장 (https://data.seoul.go.kr/)

Search Frequency for '코로나 (COVID-19)'
2021/01/01 ~ 2024/10/31
Google Trends

Holiday Data
2009/01/01 ~ 2024/10/31
공공데이터포털 (data.go.kr)

Trend

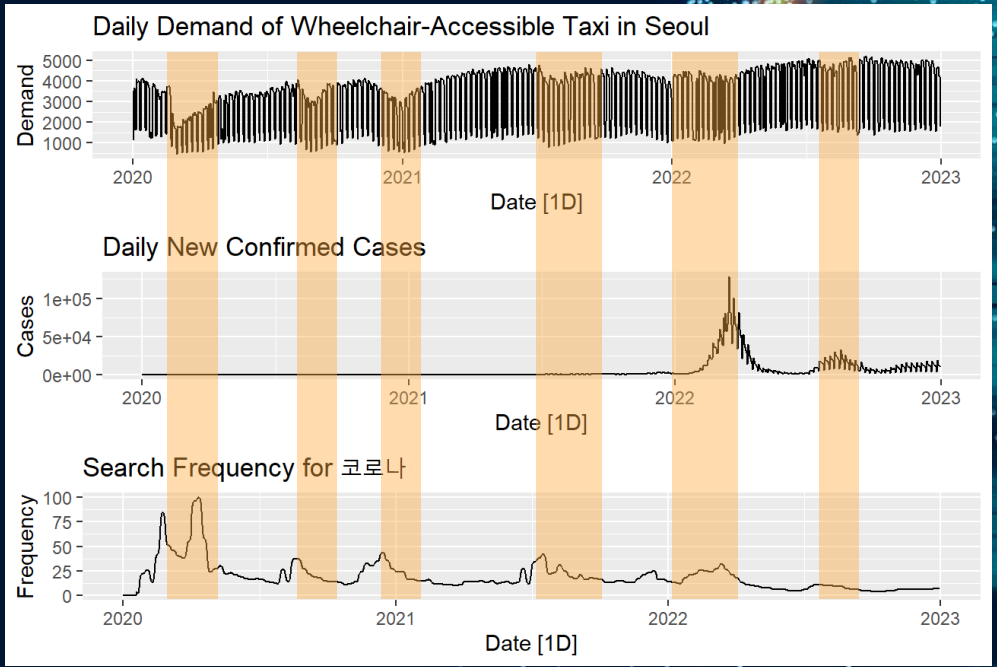
- The data exhibit an upward trend, with a drop occurring between 2017 and 2021, followed by a recovery
- There is a significant drop in 2020, coinciding with the onset of COVID-19
- The variation increases as the count becomes larger, indicating the need for a transformation



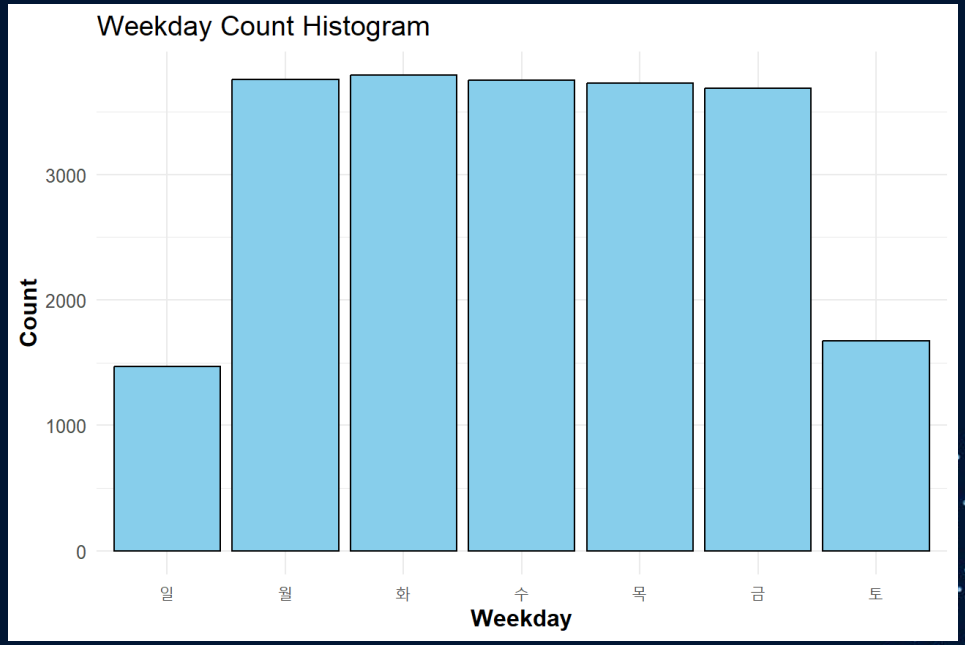
COVID-19

Search frequency is a more relevant variable

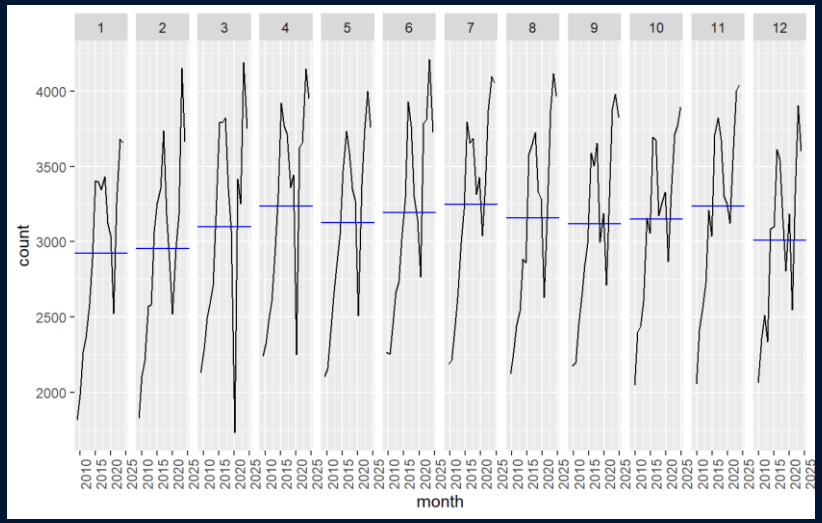
There were six breakouts:
2020/2 ~ 5
2020/8 ~ 10
2020/11 ~ 2021/2
2021/7 ~ 11
2022/1 ~ 4
2022/7 ~ 9

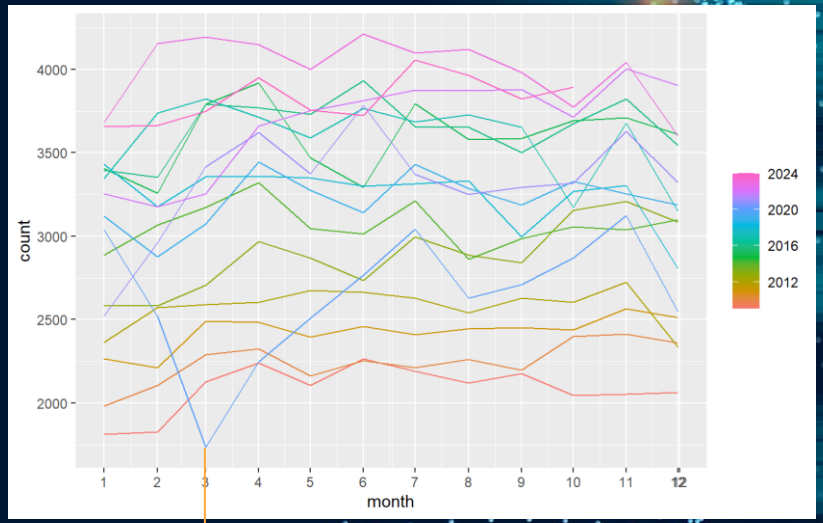


Weekly Seasonality



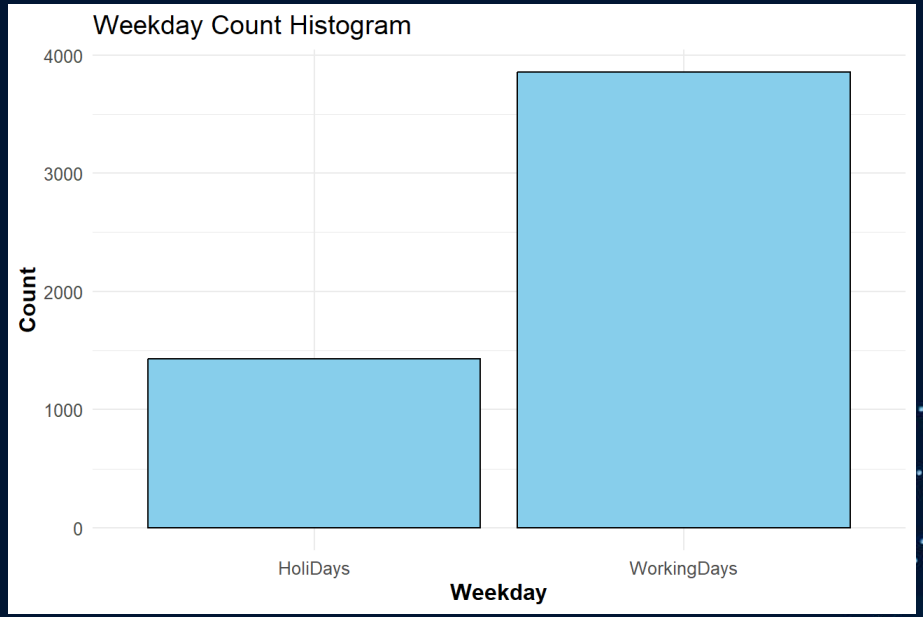
Monthly Seasonality

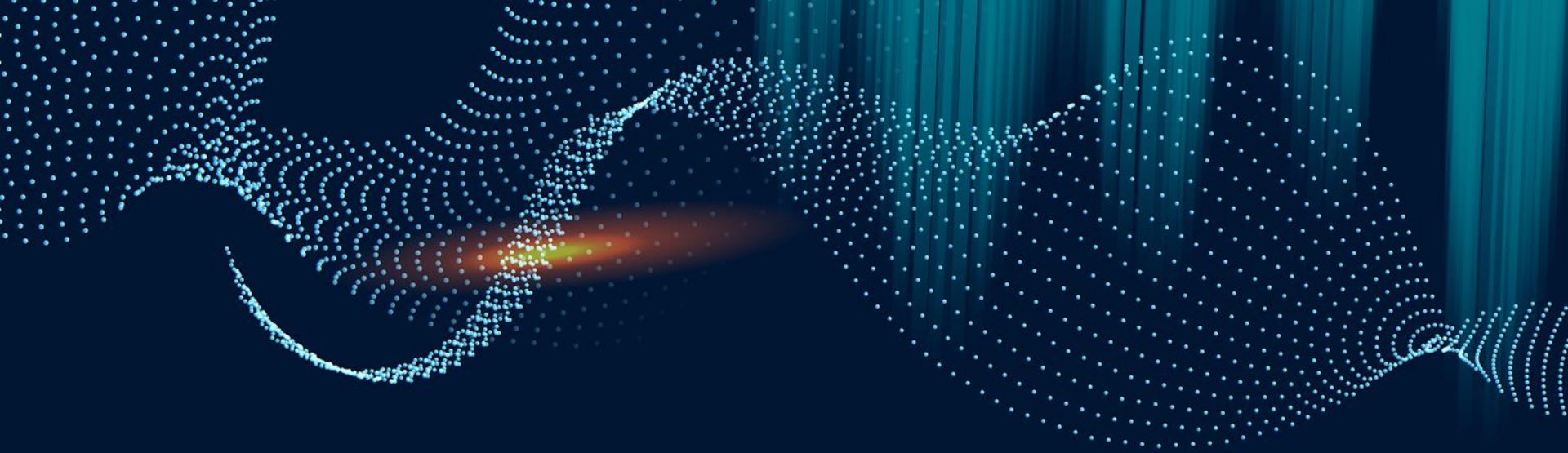




2020
COVID-19 effect

Holiday


Weekday Count Histogram
4000
3000
2000
1000
0
Count
HoliDays
WorkingDays
Weekday



03 | MODELLING

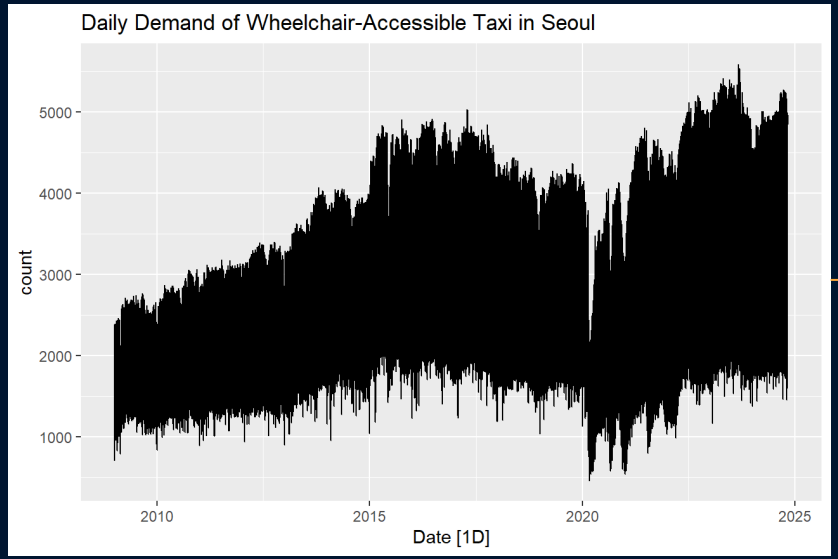
Model

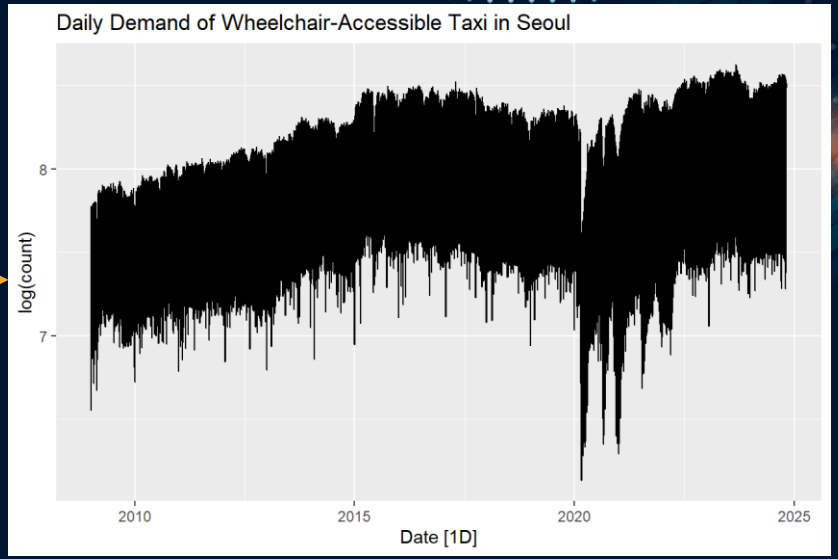
```
forecast_model <- df |>
  model(
    ARIMA = ARIMA(log(count) ~
                  trend_component +
                  PureUnusualHoliday +
                  fourier(period = 365.25, K=10)+
                  PDQ(D=1) + pdq(d=0) +
                  covid_trend))
```

→ ARIMA(2,0,1)(2,1,1)[7]

Dynamic Regression Model				
Trend	Holiday	Weekly Seasonality	Yearly Seasonality	COVID-19
SES	Dummy	SARIMA	Fourier term	Continous variable

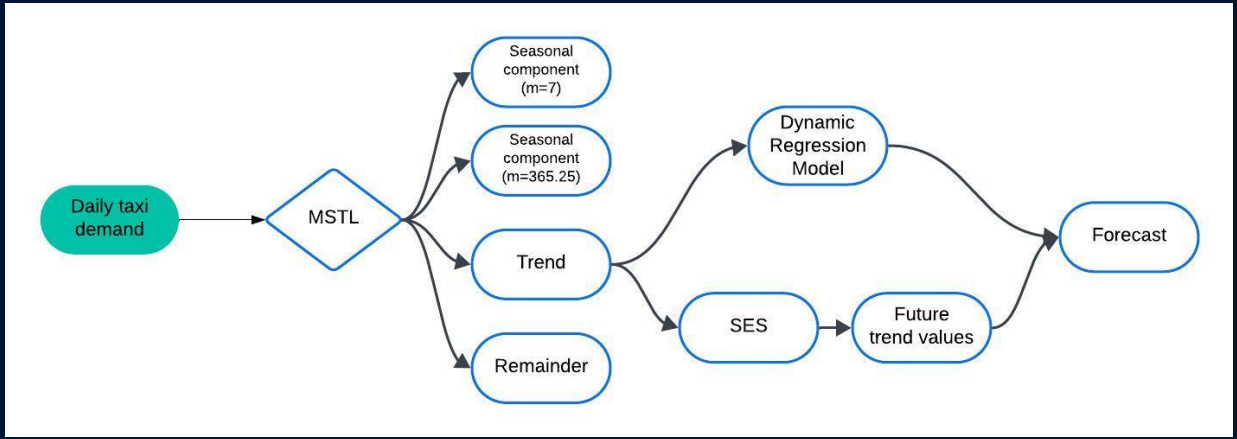
Log Transformation





Trend

- Seoul has been increasing the number of wheelchair-accessible taxies and is planning to increase the number of taxies[1].
- However, increasing number of substitutes like Boucher taxies are mitigating the increase of demand for the wheelchair-accessible taxies[2].
- In this context, simple exponential smoothing(SES) method was selected to predict the trend to reflect the recent trend without increase or decrease.
- The trend is decomposed with MSTL method, used as a predictor, and the future value of it is predicted with SES.

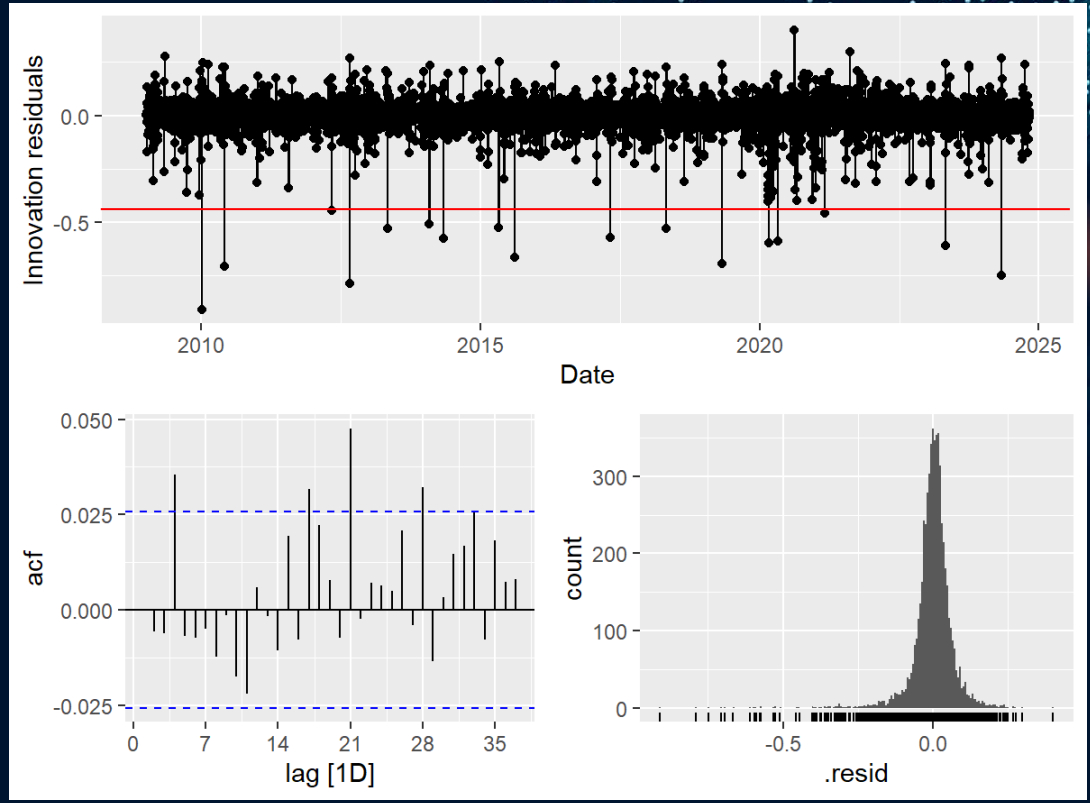


COVID-19

```
# A tibble: 6 × 8
  .model  sigma2 log_lik     AIC    AICc     BIC ar_roots    ma_roots
  <chr>    <dbl>   <dbl>   <dbl>   <dbl>   <dbl> <list>      <list>
1 lagno  0.00489   7177. -14295. -14295. -14102. <cpl [2]>  <cpl [16]>
2 lag0   0.00484   7211. -14362. -14362. -14162. <cpl [16]> <cpl [8]>
3 lag1   0.00545   6871. -13679. -13679. -13473. <cpl [18]> <cpl [0]>
4 lag2   0.00544   6870. -13677. -13676. -13463. <cpl [18]> <cpl [0]>
5 lag3   0.00544   6873. -13680. -13680. -13460. <cpl [18]> <cpl [0]>
6 lag4   0.00542   6881. -13694. -13694. -13467. <cpl [18]> <cpl [0]>
```

Unusual Days

- There are unusual days that were not taken into account
- The days with innovation residuals that are less than -0.4 were investigated



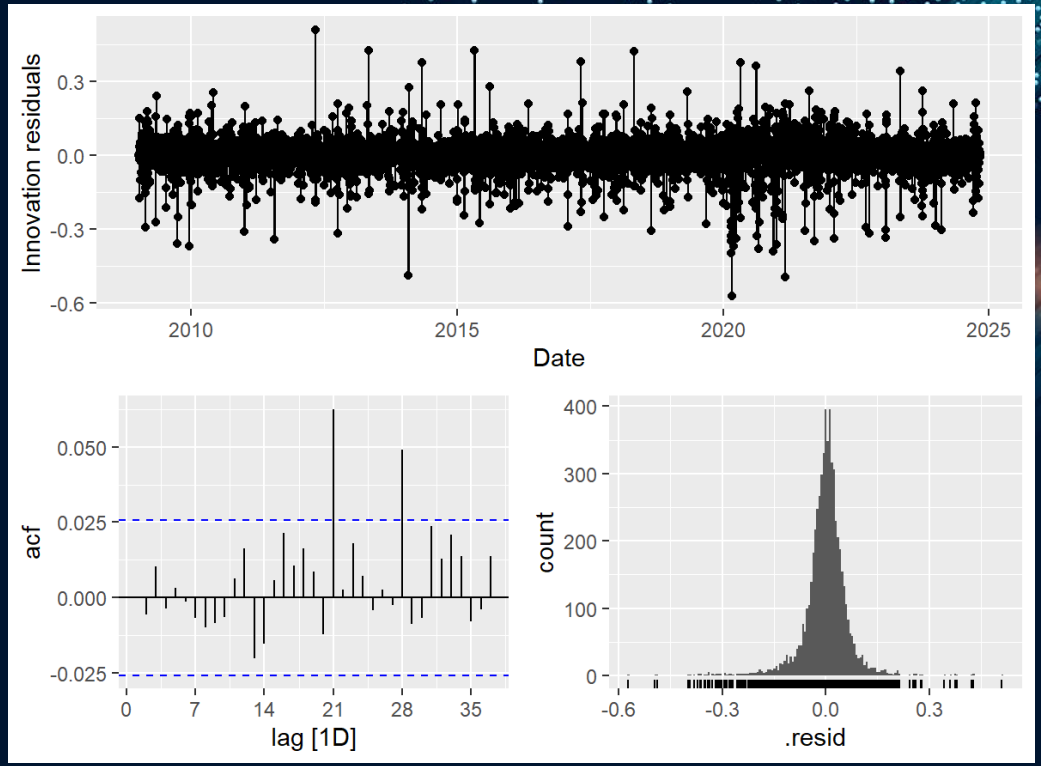
Unusual Days

- January 4, 2010: Central Region Heavy Snowfall Incident in the Korean Peninsula
- June 2, 2010: 2010 Local Elections
- August 28, 2012: Typhoon Bolaven
- May 1: Labor Day
- ...

```
> augment(rest_model) |> filter(.innov < -0.4)
# A tsibble: 18 x 6 [1D]
# Key:       .model [1]
   .model Date       rest_component  .fitted .resid .innov
   <chr>  <date>              <dbl>    <dbl>  <dbl>  <dbl>
 1 ARIMA  2010-01-04          -0.753  0.157   -0.910 -0.910
 2 ARIMA  2010-06-02          -0.427  0.281   -0.708 -0.708
 3 ARIMA  2012-05-01          -0.193  0.253   -0.446 -0.446
 4 ARIMA  2012-08-28          -0.547  0.243   -0.790 -0.790
 5 ARIMA  2013-05-01          -0.236  0.293   -0.529 -0.529
 6 ARIMA  2014-02-01          -1.08  -0.573   -0.510 -0.510
 7 ARIMA  2014-05-01          -0.291  0.285   -0.577 -0.577
 8 ARIMA  2015-05-01          -0.204  0.324   -0.527 -0.527
 9 ARIMA  2015-08-14          -0.422  0.245   -0.667 -0.667
10 ARIMA  2017-05-01          -0.233  0.340   -0.573 -0.573
11 ARIMA  2018-05-01          -0.245  0.287   -0.531 -0.531
12 ARIMA  2019-05-01          -0.372  0.322   -0.693 -0.693
13 ARIMA  2020-02-23          -1.30  -0.903   -0.402 -0.402
14 ARIMA  2020-03-01          -1.89  -1.29    -0.596 -0.596
15 ARIMA  2020-05-01          -0.583  0.00691 -0.590 -0.590
16 ARIMA  2021-03-01          -1.12  -0.659   -0.456 -0.456
17 ARIMA  2023-05-01          -0.209  0.402   -0.611 -0.611
18 ARIMA  2024-05-01          -0.445  0.304   -0.748 -0.748
```

Residuals

- The unusual days were treated just as other holidays
- The graphs on the right are after the treatment
- ACF implies that the residuals are still autocorrelated



Coefficients

- These are coefficients of the model after training with all available data

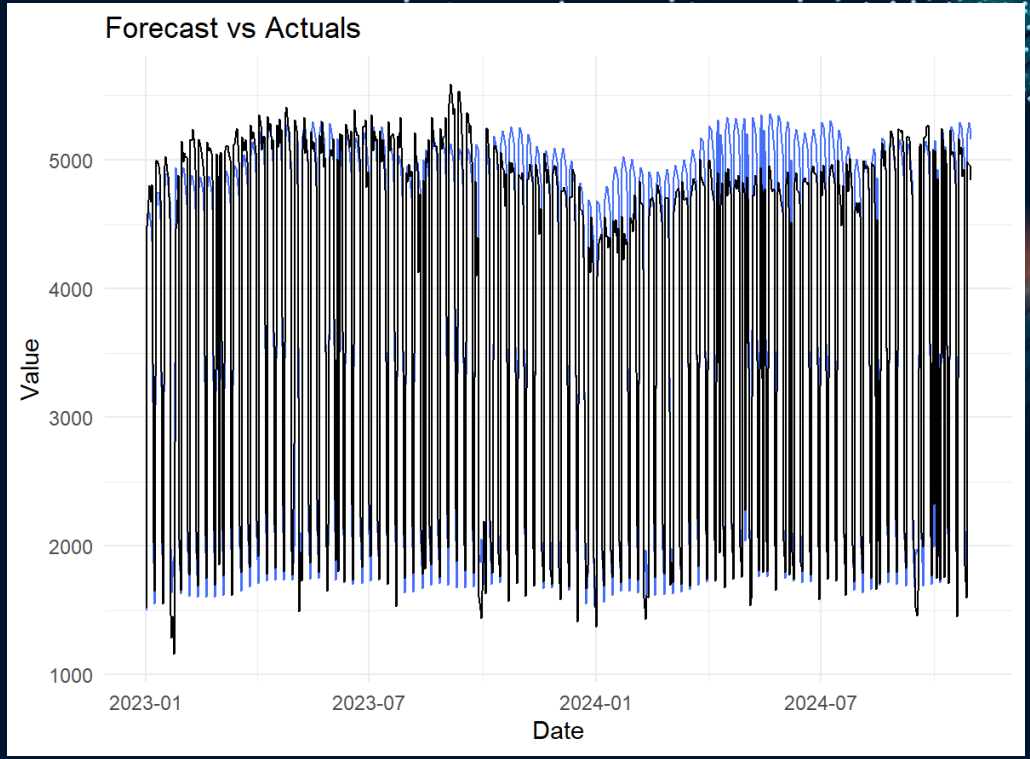
```
Model: LM w/ ARIMA(2,0,1)(2,1,1)[7] errors
Transformation: log(count)

Coefficients:
         ar1      ar2      ma1     sar1    sar2     sma1  trend_component
      0.9113  -0.0556  -0.5118  0.1461  0.1042  -0.9092           0.9371
s.e.  0.0479   0.0307   0.0457  0.0165  0.0152   0.0078           0.2701
      PureUnusualHolidayTRUE  fourier(period = 365.25, K = 10)C1_365
                     -0.9605                             -0.0318
s.e.                  0.0044                              0.0064
      fourier(period = 365.25, K = 10)S1_365  fourier(period = 365.25, K = 10)C2_365
                                      0.0006                             -0.0187
s.e.                                  0.0064                              0.0052
      fourier(period = 365.25, K = 10)S2_365  fourier(period = 365.25, K = 10)C3_365
                                     -0.0181                             -0.0044
s.e.                                  0.0052                              0.0047
      fourier(period = 365.25, K = 10)S3_365  fourier(period = 365.25, K = 10)C4_365
                                      0.0009                             -0.0037
s.e.                                  0.0048                              0.0043
      fourier(period = 365.25, K = 10)S4_365  fourier(period = 365.25, K = 10)C5_365
                                      0.0053                             -0.0081
s.e.                                  0.0043                              0.0040
      fourier(period = 365.25, K = 10)S5_365  fourier(period = 365.25, K = 10)C6_365
                                      0.0071                             -0.0026
s.e.                                  0.0040                              0.0037
      fourier(period = 365.25, K = 10)S6_365  fourier(period = 365.25, K = 10)C7_365
                                      0.0004                             -0.0101
s.e.                                  0.0037                              0.0033
      fourier(period = 365.25, K = 10)S7_365  fourier(period = 365.25, K = 10)C8_365
                                     -0.0033                             -0.0018
s.e.                                  0.0033                              0.0031
      fourier(period = 365.25, K = 10)S8_365  fourier(period = 365.25, K = 10)C9_365
                                      0.0066                             -0.0043
s.e.                                  0.0031                              0.0028
      fourier(period = 365.25, K = 10)S9_365  fourier(period = 365.25, K = 10)C10_365
                                     -0.0005                             -0.0024
s.e.                                  0.0028                              0.0026
      fourier(period = 365.25, K = 10)S10_365  covid_trend
                                       0.0029      -0.0035
s.e.                                   0.0026       0.0005

sigma^2 estimated as 0.003996:  log likelihood=7763.15
AIC=-15466.31   AICc=-15465.98   BIC=-15266.46
```

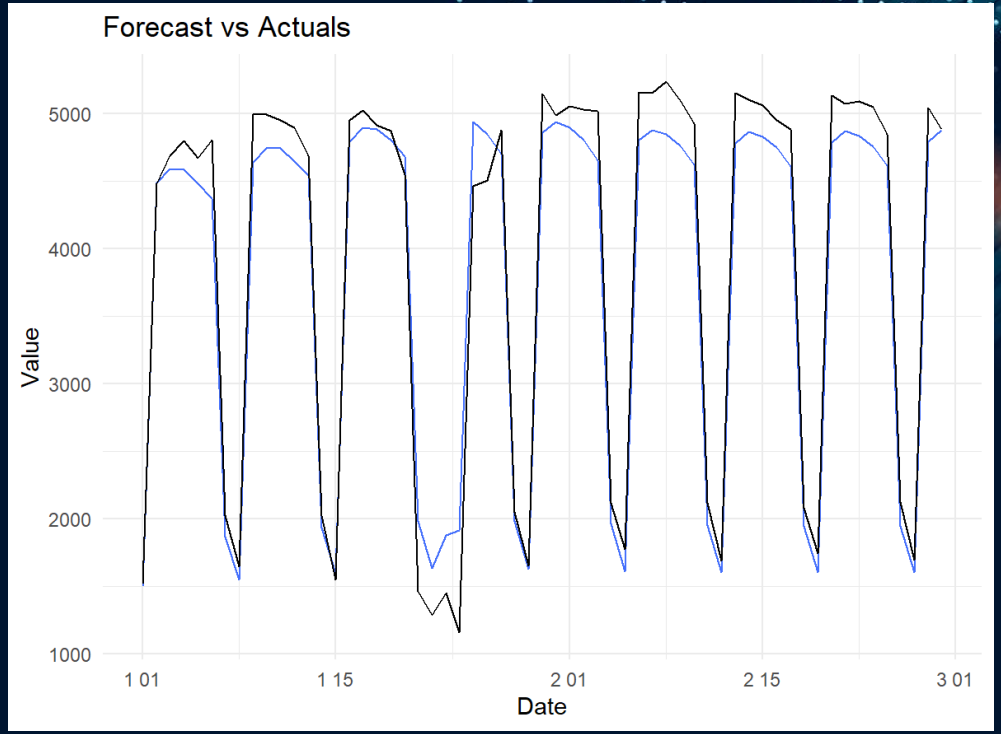
Evaluation

- Data was splitted into a training set(~2022) and a test set(2023~) for evaluation



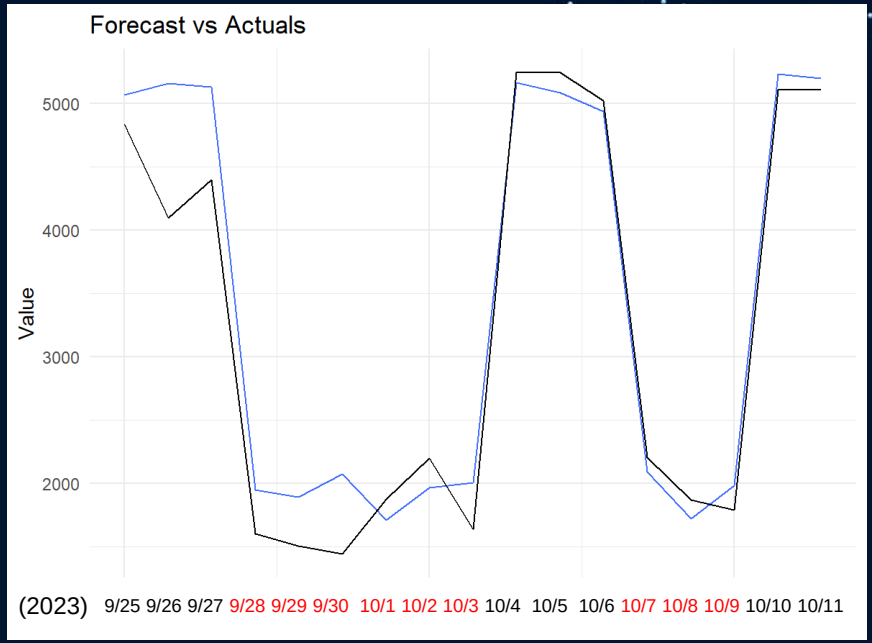
Evaluation

Weekly seasonality was captured properly



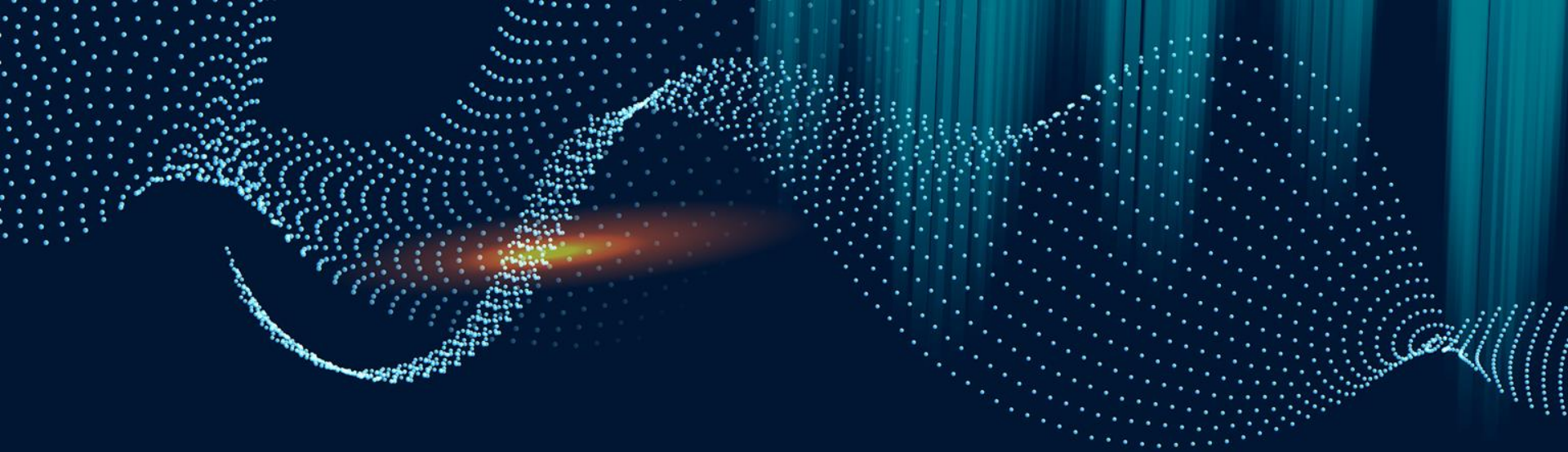
Evaluation

Holidays are well reflected



Evaluation

.model	.type	ME	RMSE	MAE	MPE	MAPE	MASE	RMSSE	ACF1
<chr>	*<chr>*	*<dbl>*	*<dbl>*	*<dbl>*	*<dbl>*	*<dbl>*	*<dbl>*	*<dbl>*	*<dbl>*
ARIMA	Test	-59.8	249.	197.	-1.61	5.69	NaN	NaN	0.651



04 | Forecast

Predictors

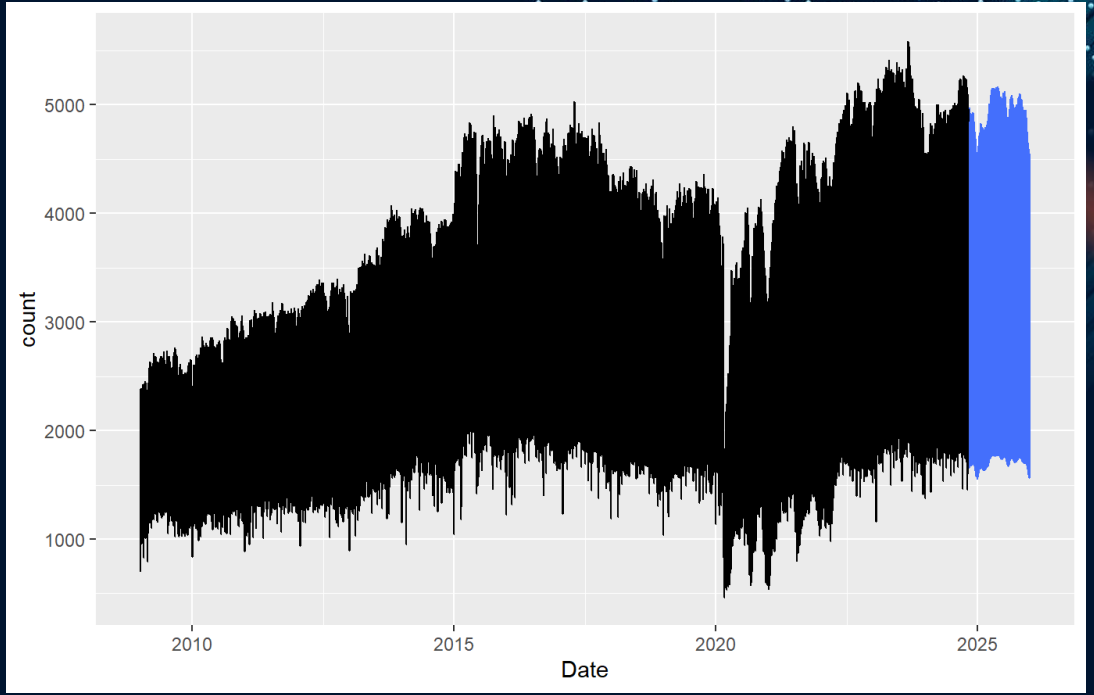
- **Trend component**
- SES is used to for future trend values

- **Search frequency for COVID-19**
- Value 1 is used because the majority of the values in 2024 is 1 and it is likely to be continued.
- No significant breakout is assumed

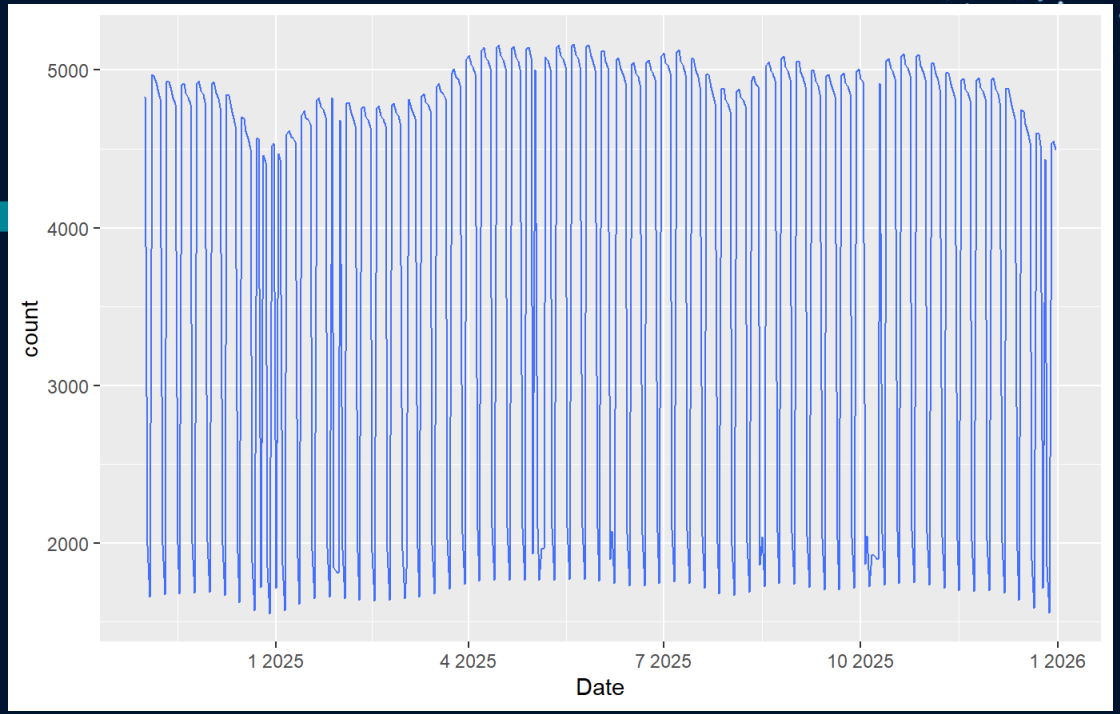
- **Holidays**
- Holidays of the future are known

Forecast

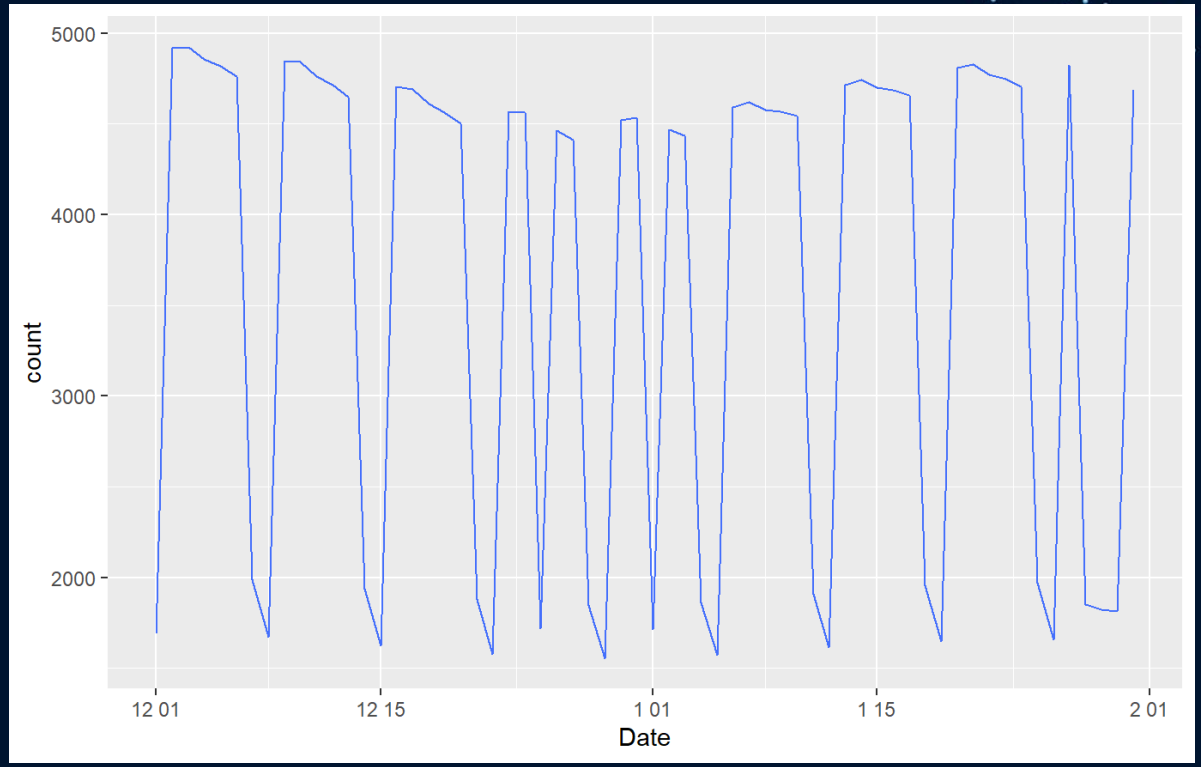
The period for forecasting is
2024/11/01 ~ 2025/12/31

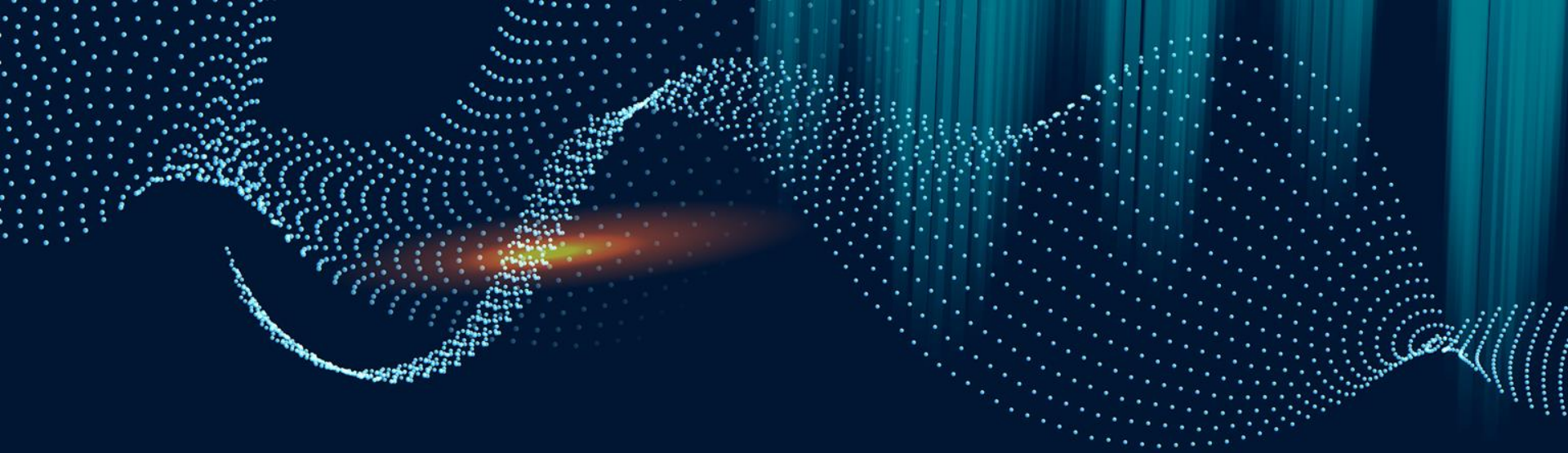


Forecast


5000
4000
3000
2000
count
1 2025
4 2025
7 2025
10 2025
1 2026
Date

Forecast





05 CONCLUSION

Conclusion

- **Limitations**

- The remainder is still autocorrelated; probably there is more pattern to be captured

- **Implications**

- The model captured seasonality and holidays even though there’s more to be captured
- Incorporating the COVID-19 effect using search frequency data improved the model's performance

References

[1] Choi, J. (2023, July 25). Seoul City significantly expands disabled call taxis... 'Jeon Jang-yeon, are you watching?'. Financial News. https://www.fnnews.com/news/202307250849429623
[2] Seoul Metropolitan Government Press Release:Seoul Metropolitan Government. (2023, September 21). Seoul City significantly improves convenience of voucher taxis for the disabled... 'Reducing waiting times' [Press release]. https://opengov.seoul.go.kr/press/29320056


THANK YOU