2024 Spring Semester
Professor Myungsuk, Kim
Team 1

Trend, Seasonality Analysis, and Enhancement Strategies for Japanese Tourists in Korea

: A Regression and Time Series Data Analysis

Contents

Introduction

Trend, Seasonality Analysis, and Enhancement Strategies for Japanese Tourists in Korea

: A Regression and Time Series Data Analysis

- Analyze the relationship between the number of times Japanese Users Searched for Keywords About Korea and the Number of Japanese Visitors to Korea with regression
- Supplementary forecasting analysis
- Find enhancement strategies for Japanese tourists in Korea using above analysis results

Problem Motivation

1) Increasing Rate of Domestic travel in Korea slows down after the COVID-19

*units 1000 times, %

Chart : Number of domestic travel and increasing rate

source: Ministry of Culture, Sports and Tourism, South Korea

Year	Annual	1Q	2Q	3Q	4Q
2019	344,750	84,061	80,915	90,658	89,116
2020	225,199	59,100	55,818	59,319	50,962
2021	245,127	54,277	65,810	64,588	60,452
2022	283,722	65,228	71,903	75,759	70,833
2023(B)	296,977	73,374	73,366	75,893	71,344
Growth rate (B-A)	4.7	17.1	2.0	0.2	0.7

Problem Motivation

2) Develop enhancement strategies of foreign tourists in Korea considering the increasement potential

- To make practical solution, we need to target the country that has meaningful impact to the total number of foreign travelers

- For travel purpose, more than 20% of the foreign tourists visited Korea are Japanese.

Chart : Annual number of travelers from Japan and total

source: Ministry of Culture, Sports and Tourism, South Korea
Unit: 명(person)

	2015Y	2016Y	2017Y	2018Y	2019Y	2020Y	2021Y	2022Y	2023Y	2024Y	Total
JAPAN	1,742,531	2,213,099	2,223,214	2,864,110	3,179,436	406,701	1,047	259,031	2,258,375	874,659	16,022,203
TOTAL	10,135,489	13,932,925	10,415,594	12,414,348	14,432,275	1,653,471	211,846	1,998,937	8,880,899	3,928,048	78,003,832
Percentage	17.19%	15.88%	21.35%	23.07%	22.03%	24.60%	0.49%	12.96%	25.43%	22.27%	20.54%

Problem Motivation

3) Find enhancement strategies for Japanese tourists in Korea to boost the domestic travel economy

- According to JNTO(Japan National Tourism Organization), the number of tourists in Japan who are Korean is as below chart
- Total population Japan in 24Y is 122,631,432 and that of Korea in 24Y is 51,751,065.
- Even considering the exchange rate problem, the number of Korean tourists visiting Japan is heavily outweighs than that of Japanese tourists visiting Korea, even though the Japan population is more than twice as large.

	Oct 23	Nov 23	Dec 23	Jan 24	Feb 24
Korea to Japan	631,120	649,880	782,730	857,000	818,500
Japan to Korea	250,235	272,686	191,892	137,713	177,633

* Each data from JNTO and Ministry of Culture, Sports and Tourism, South Korea

Literature Review

1. Molemeng. (2023). The Factors Influencing the Number of Visits and Behavioral Intentions of China, Japan and U.S. Tourists - Based on KTO 2019 International Visitor Survey. Master's thesis, Sejong University, Graduate School, Department of Hotel and Tourism Management.
2. Deukhee Park, Sanghoon Kang, and Gyehee Lee. "Demand Forecasting of Japanese Tourists to Korea for Sustainable Tourism Growth: An Application of Time-Series Econometric Models." Journal of Tourism Studies 34, no. 3 (2020): 47-60. doi: https://doi.org/10.21298/IJTHR.2020.3.34.3.47.
3. Soeon Park, Gunhee Lee, Inhye Lee. "Analysis of Factors of Visiting Tourists to Korea using Economic Variables and Media Data: A Case Study of Japan and Chinese Tourist." International Journal of Tourism and Hospitality Research 29, no. 7 (2015)

Literature Review

- Prior studies mainly focused on either forecasting or factor analysis (which factor affects the number of tourists) of the data regarding the tourism in Korea.
- While there are studies that forecast the number of visitors by adjusting for the impact of COVID-19 or analyze factors influencing visitation, there is a lack of research that comprehensively considers seasonality, trends, and influencing factors targeting a specific country to boost the number of tourists.

Statement of Research Objectives

1. Analyze What factors makes Japanese visit Korea using regression model between keyword data and tourist data

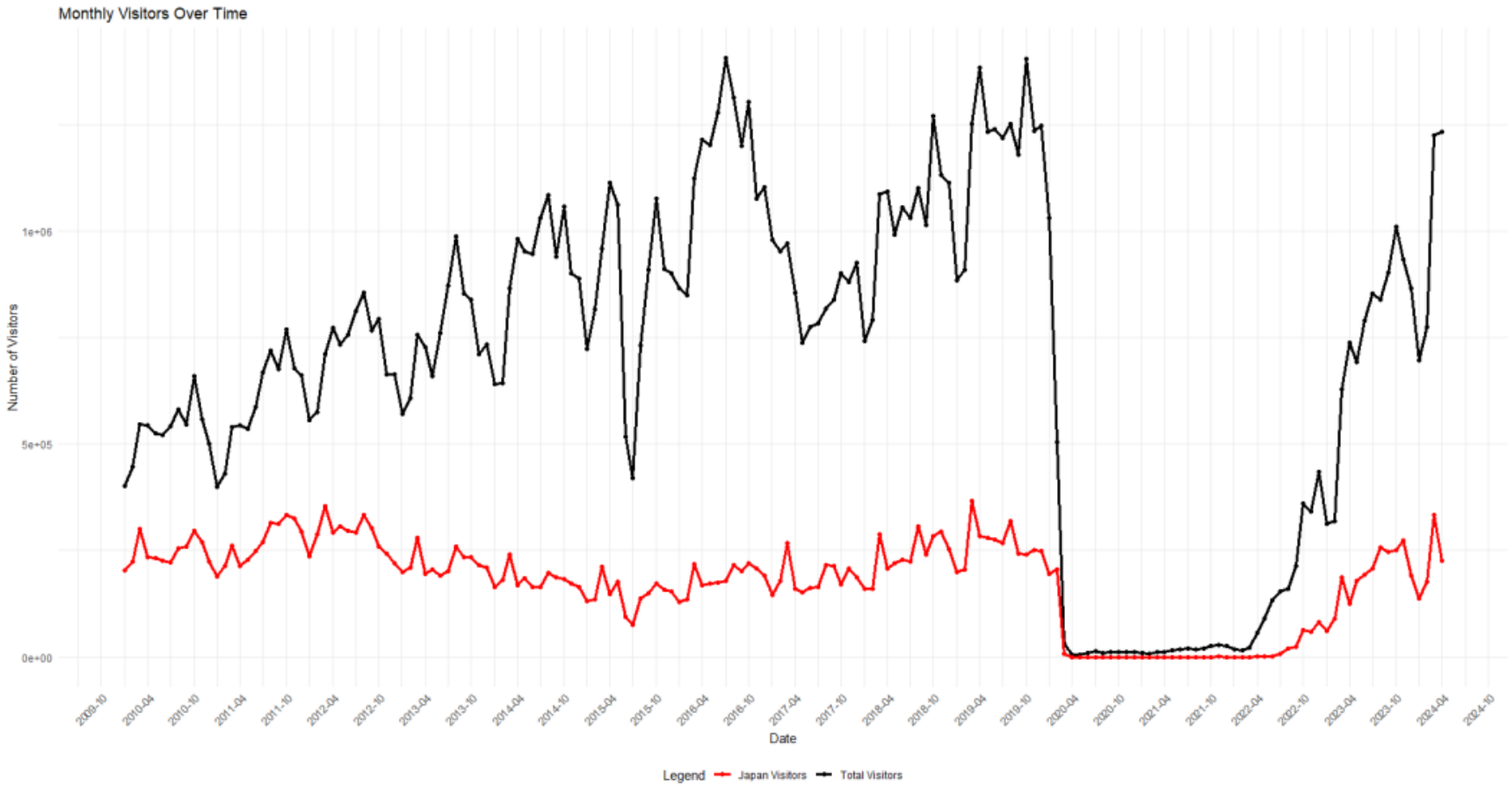
2. Interpret Trend and Seasonality of Japanese tourist data using decomposed model

3. Predict the number of Japanese tourists in Korea using SARIMA model

4. Properly interpret the meaning of analysis results and find strategies accordingly

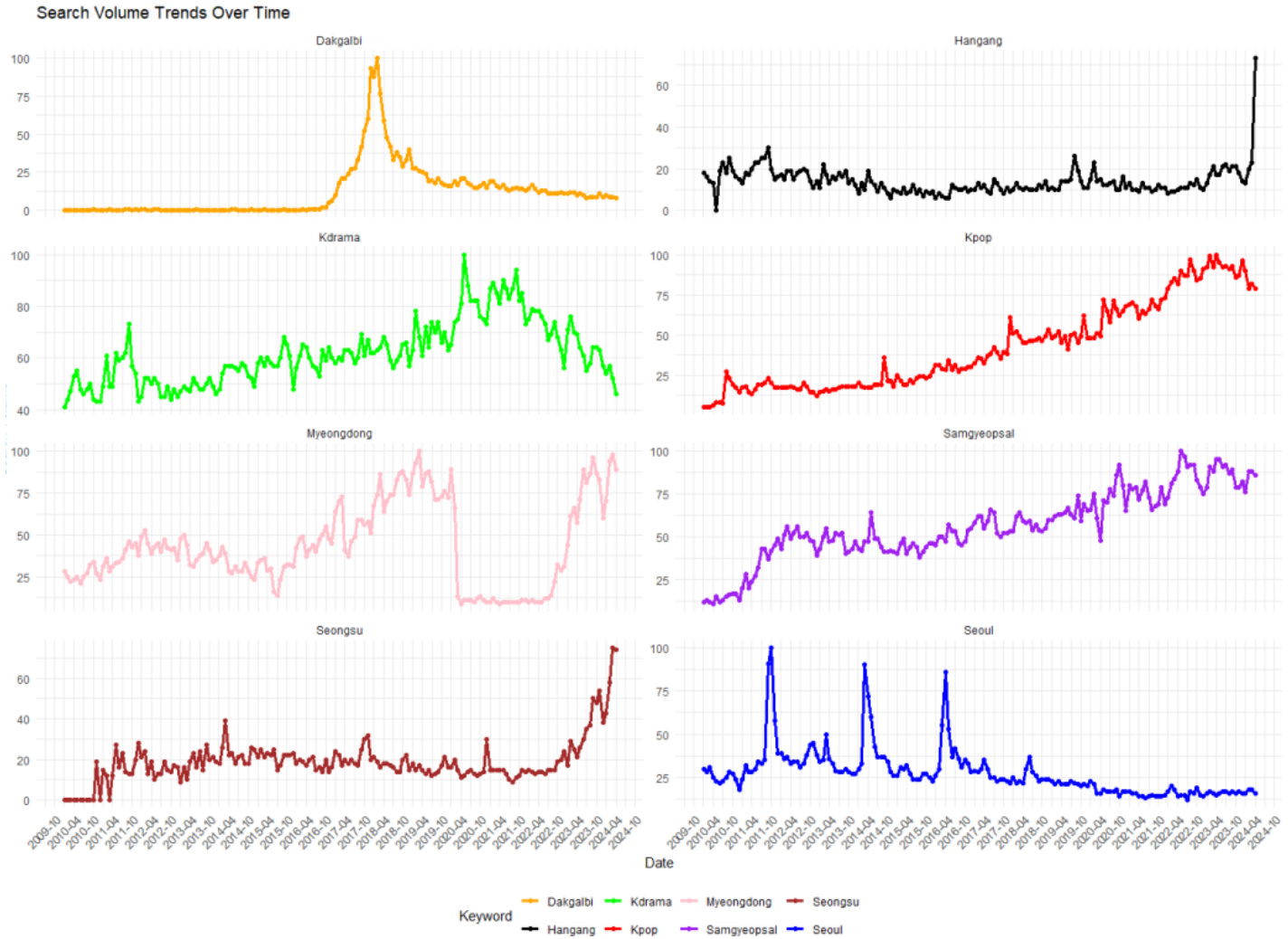
Methodology and Result

Dependent variable: Monthly visitors



Methodology and Result

Exogenous variables



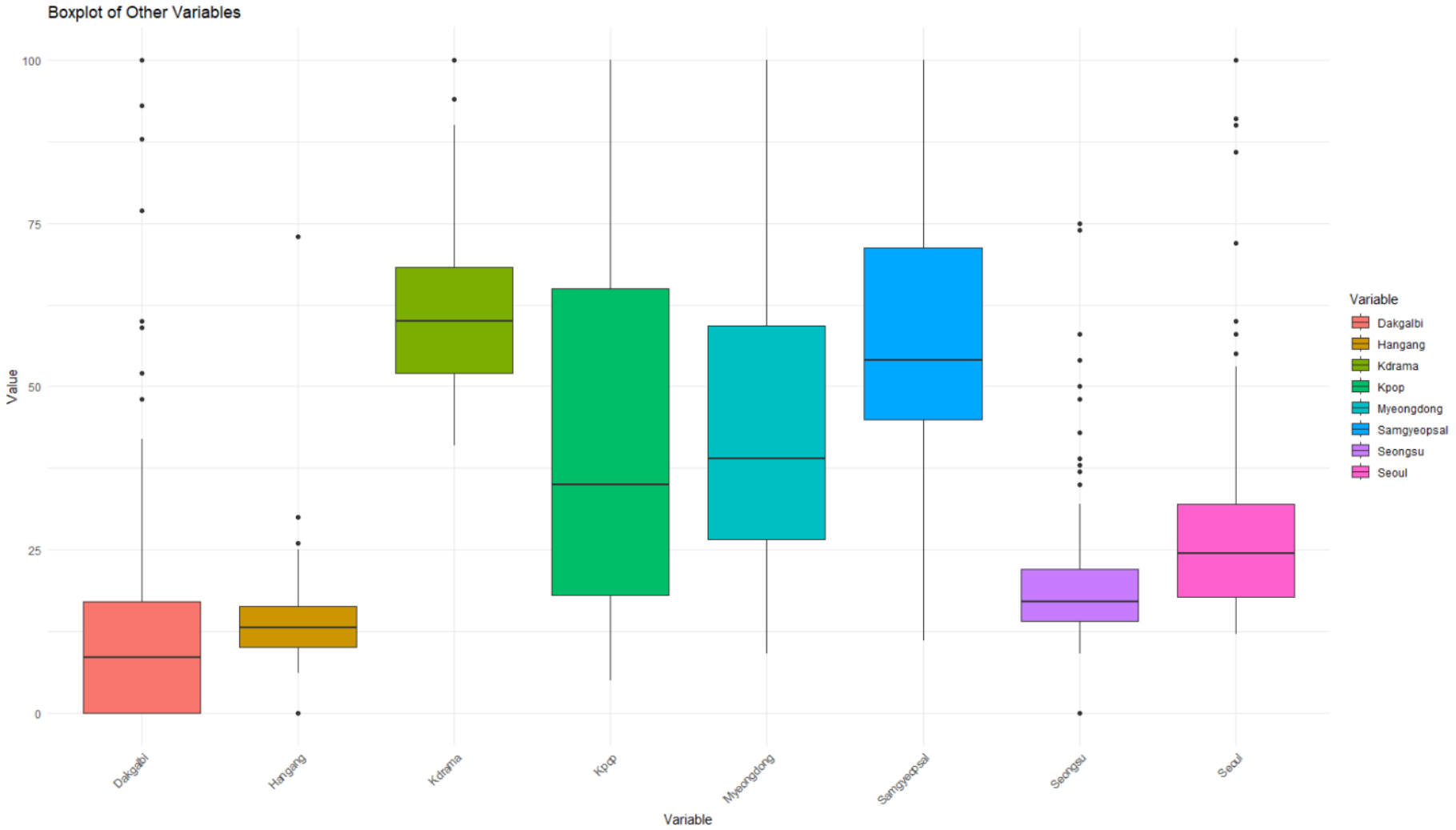
Methodology and Result

Descriptive Statistics for Variables

Metric	Japan_Visitors	Kpop	Seoul	Kdrama	Samgyeopsal	Dakgalbi	Seongsu	Myeongdong	Hangang
Mean	176772.40	42.28488	27.48837	61.66279	56.59884	11.88372	18.73837	43.00581	13.837209
Median	198194.00	35.00000	24.50000	60.00000	54.00000	8.50000	17.00000	39.00000	13.000000
Max	367157.00	100.00000	100.00000	100.00000	100.00000	100.00000	75.00000	100.00000	73.000000
Min	16.00	5.00000	12.00000	41.00000	11.00000	0.00000	0.00000	9.00000	0.000000
SD	99404.49	26.94667	13.96382	12.06232	20.53571	17.32417	10.95104	24.74371	6.572778

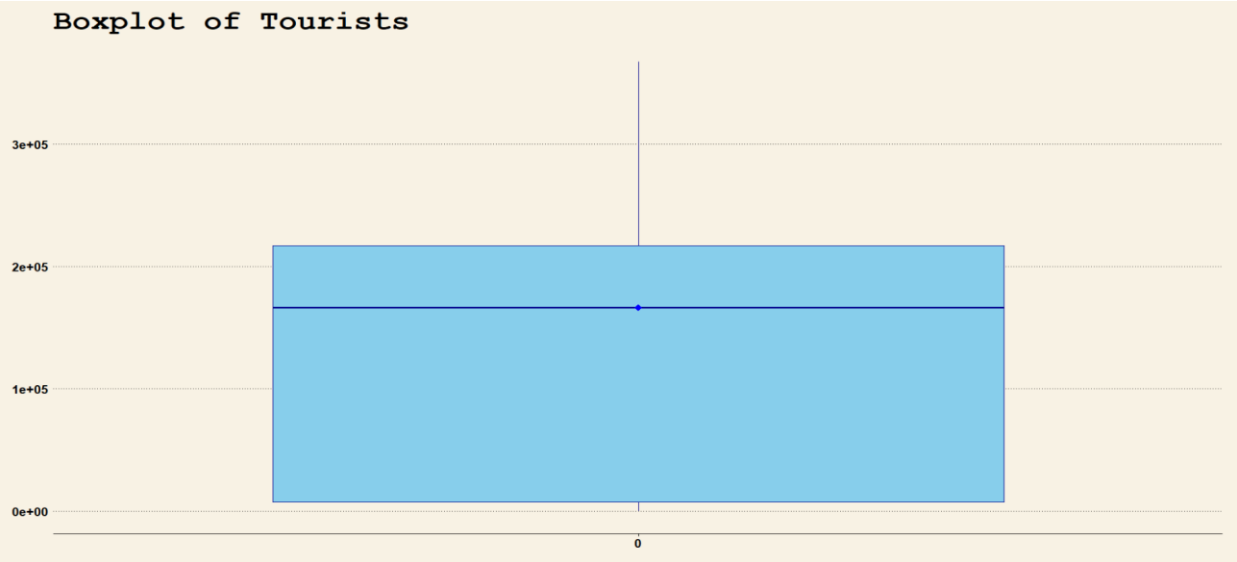
Methodology and Result

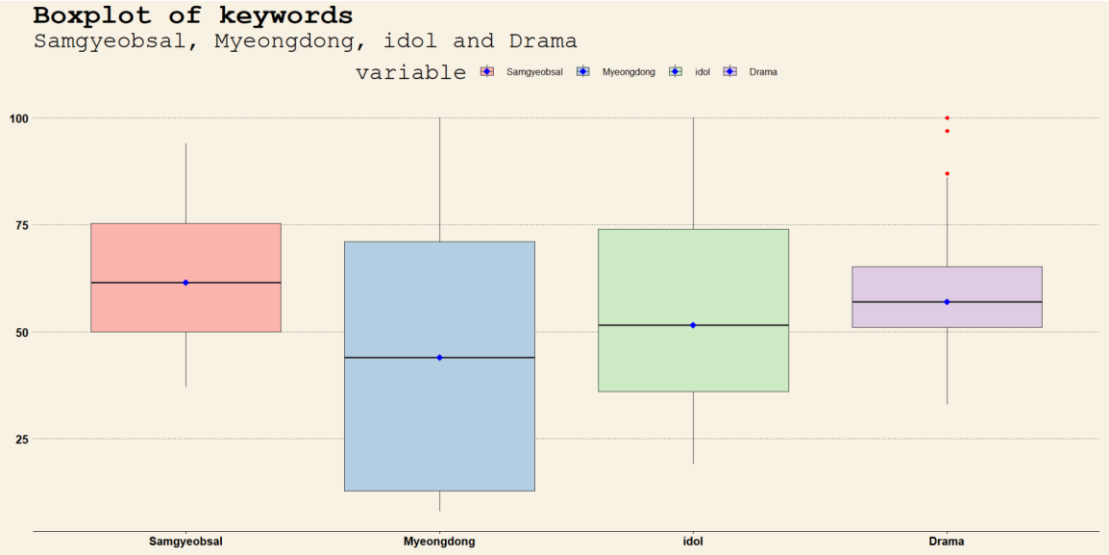
Boxplot for Variables



Methodology and Result

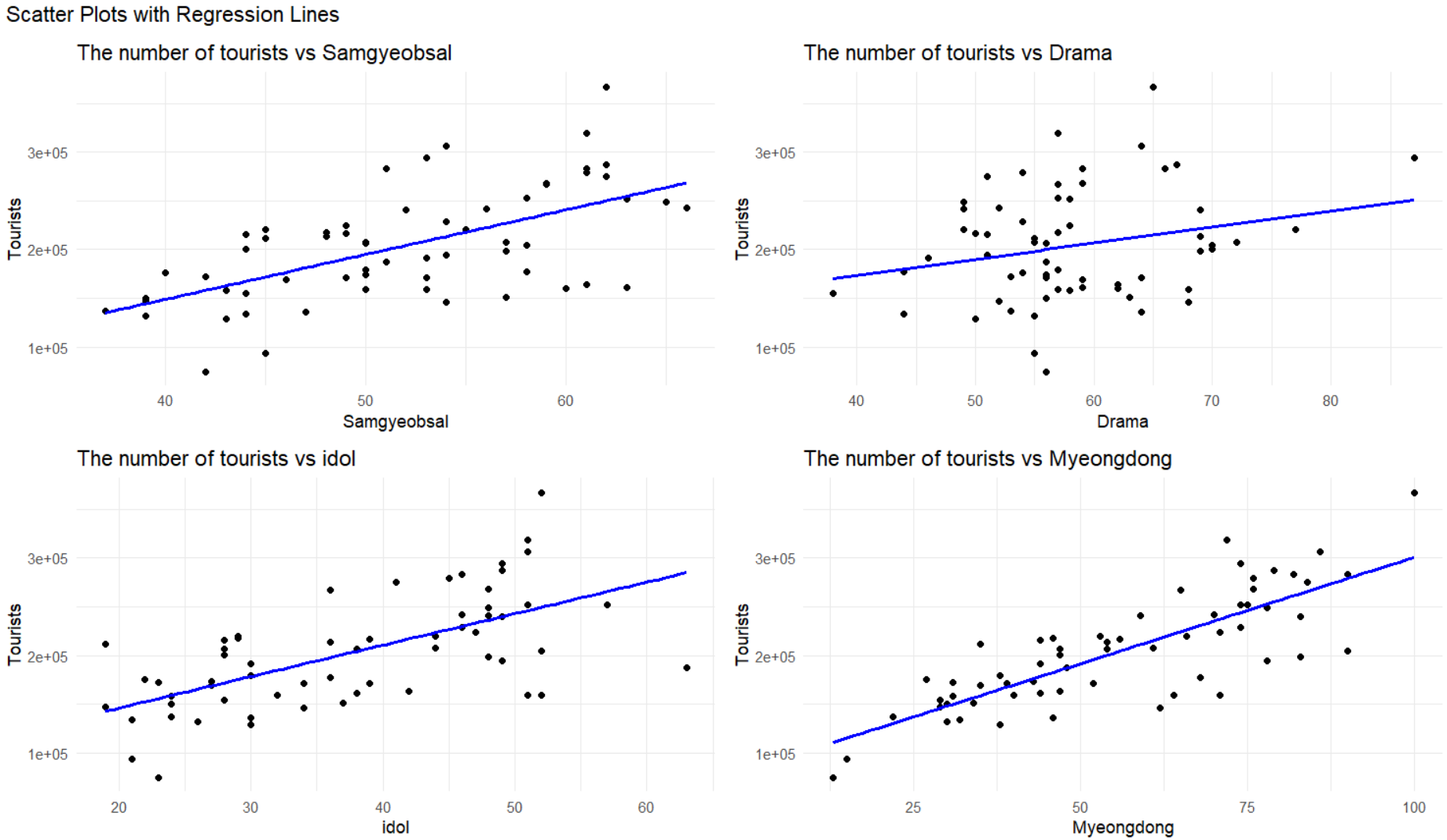
Boxplot





Methodology and Result

Scatter Plots with each keyword

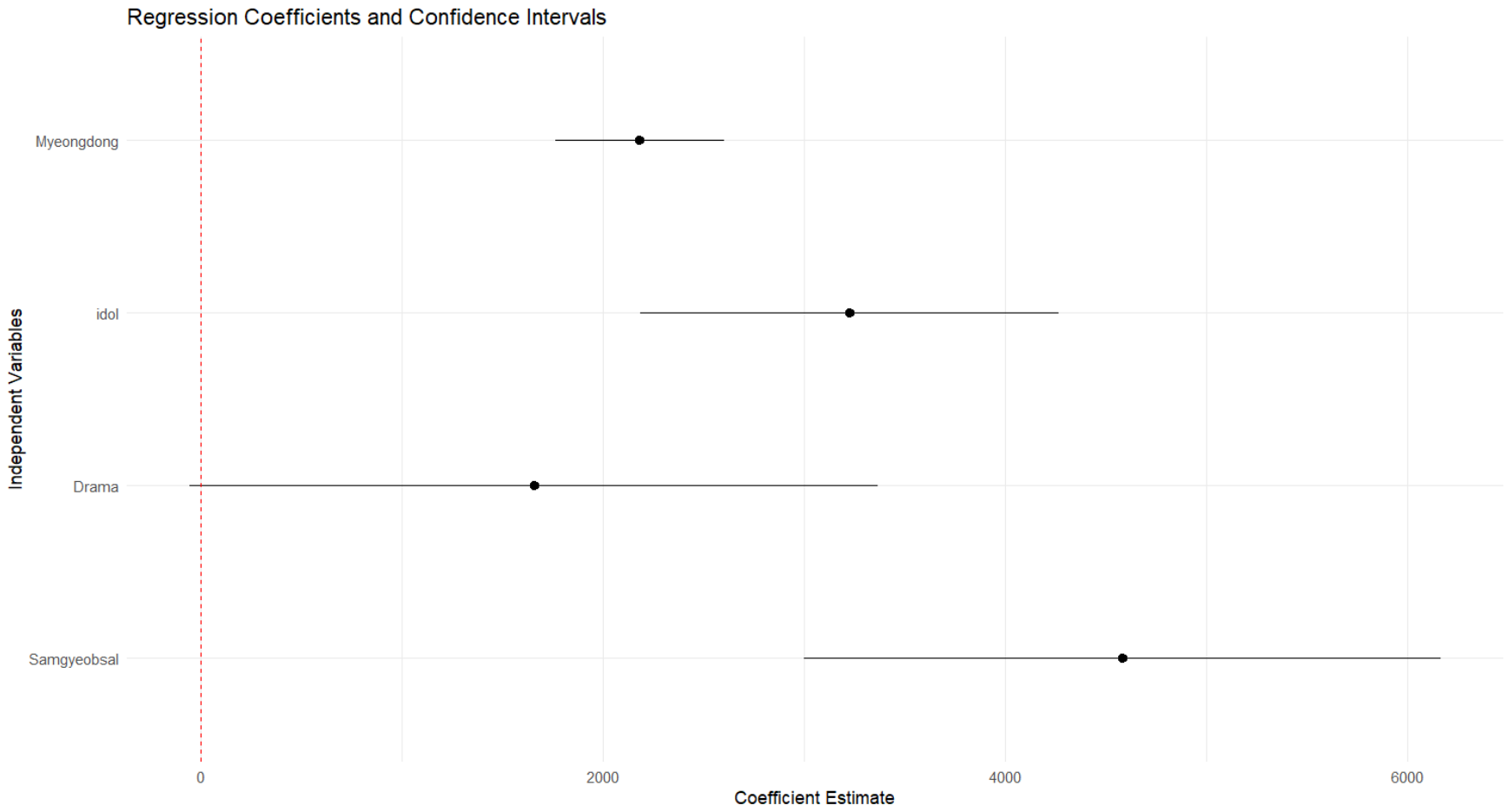


Results

- Positive slopes for every model

Methodology and Result

Coefficient Plot for each simple regression model



1. Visualisation

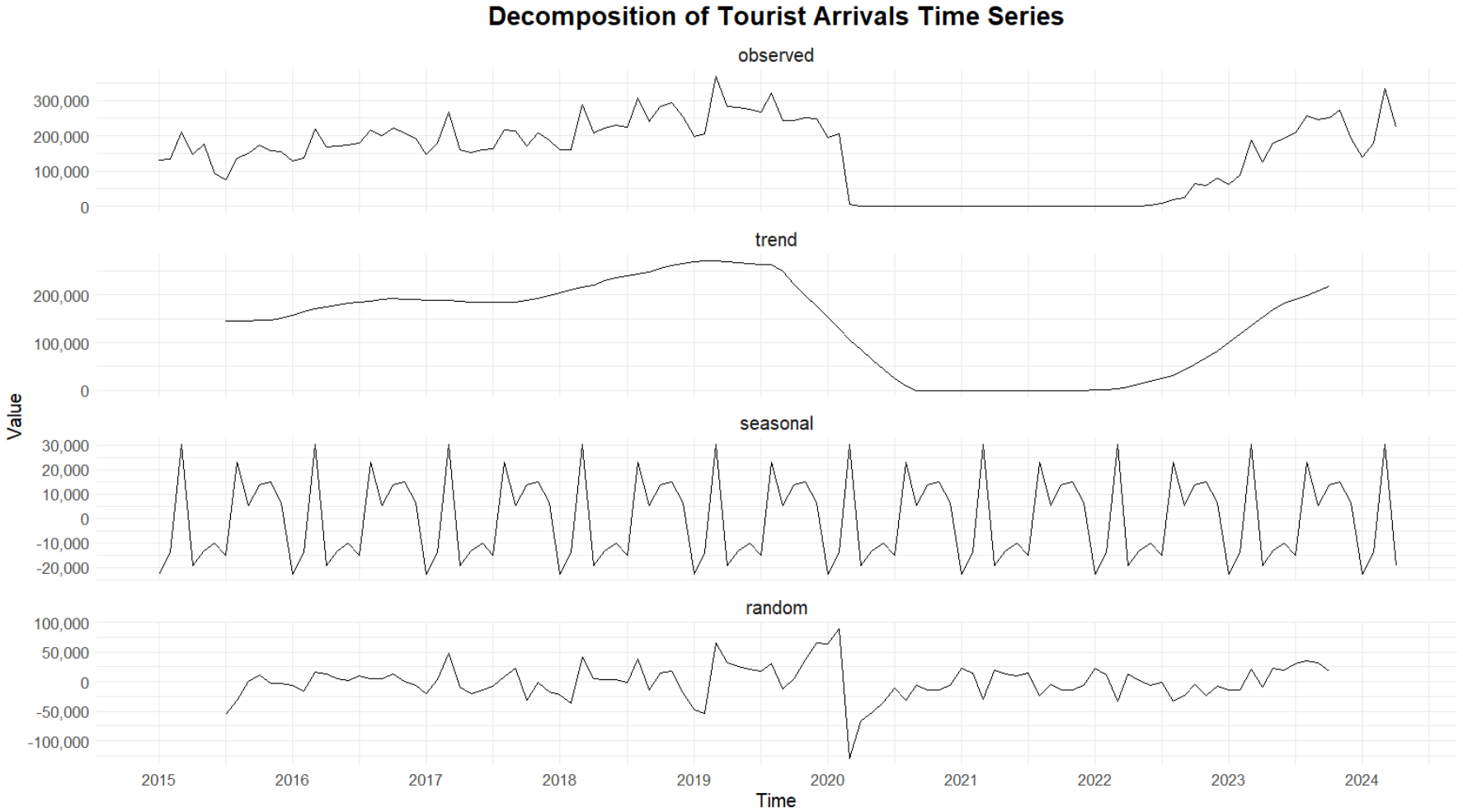
: Without Drama, Myeongdong, idol, and Samgyeobsal are significant at a significant level of 5%

2. Interpretation

: Japanese tourists came to Korea for
food, place, and idol.

Methodology and Result

Decomposition Model



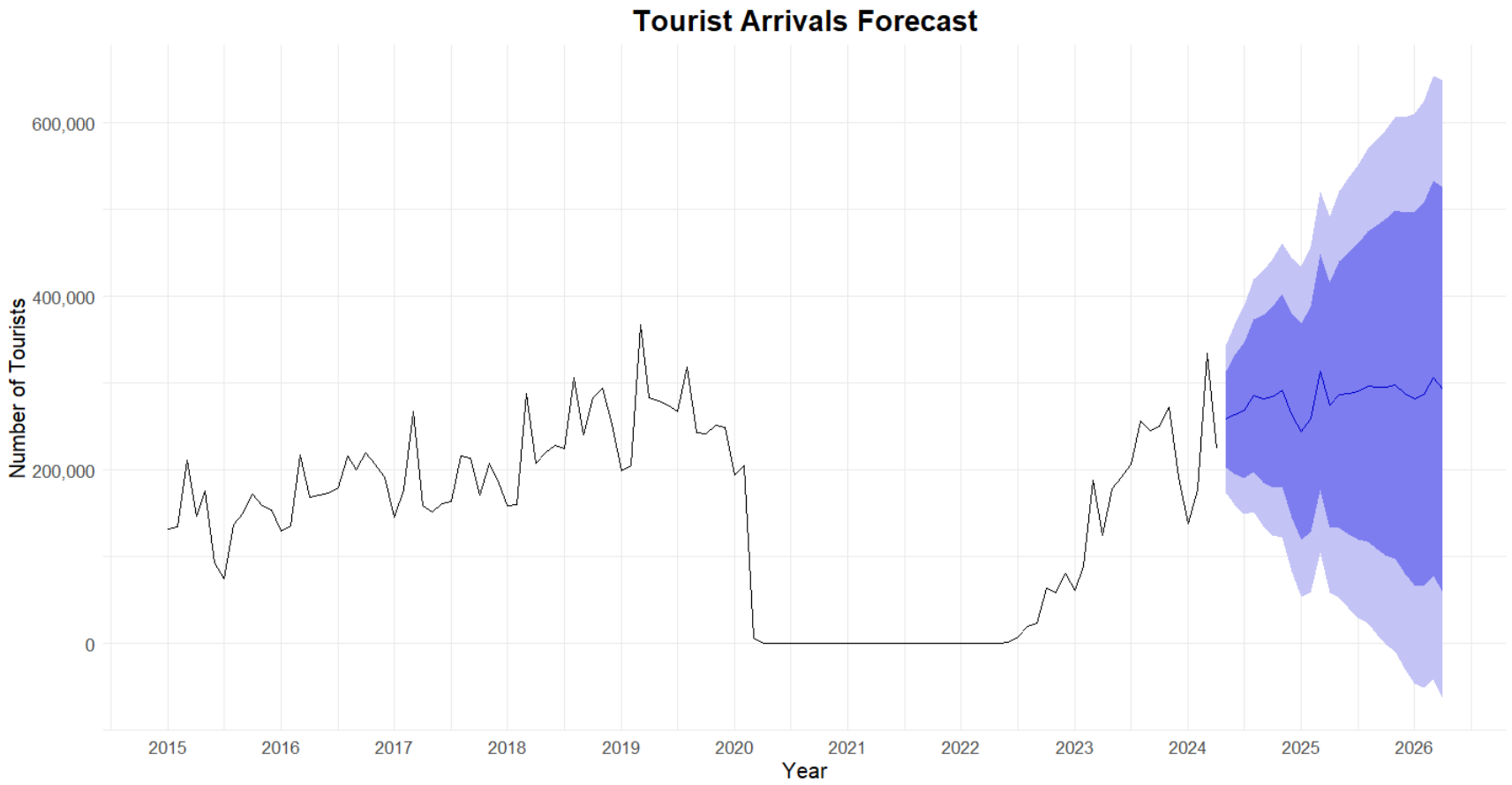
1. Result:

   Clear Seasonality

   Weird Trend due to COVID-19

Methodology and Result

Prediction using SARIMA



1. Result:

   ARIMA(0,1,1)(1,0,0)[12]

2. Point Forecast

Time	Tourists
May 2024	258816
Aug 2024	285926
Nov 2024	291745
Feb 2025	258432
May 2025	286884
Aug 2025	296384
Nov 2025	298424

Methodology and Result

Box-Ljung test

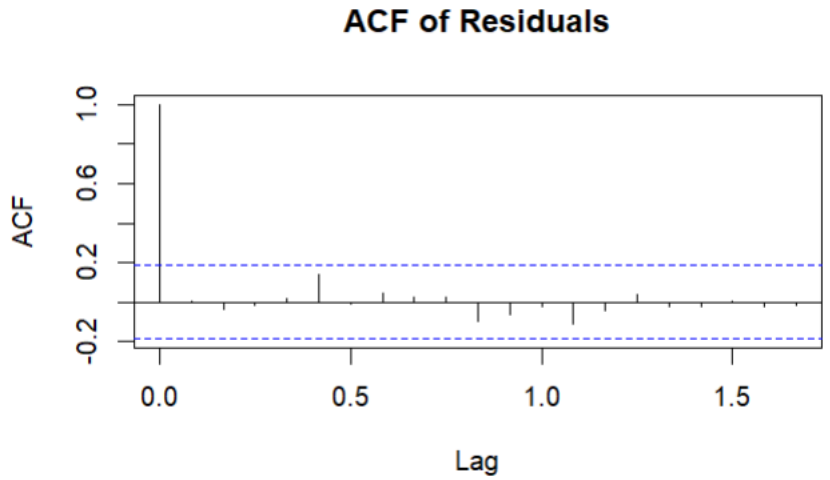
```r
Box.test(residuals, lag = 20, type = "Ljung-Box")
```

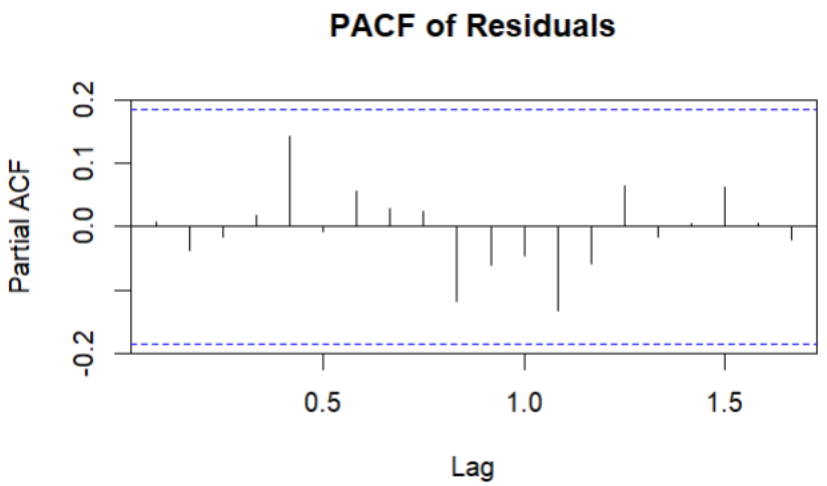
```
        Box-Ljung test

data:  residuals
X-squared = 7.0817, df = 20, p-value = 0.9964
```

Methodology and Result

ACF, PACF of Residuals

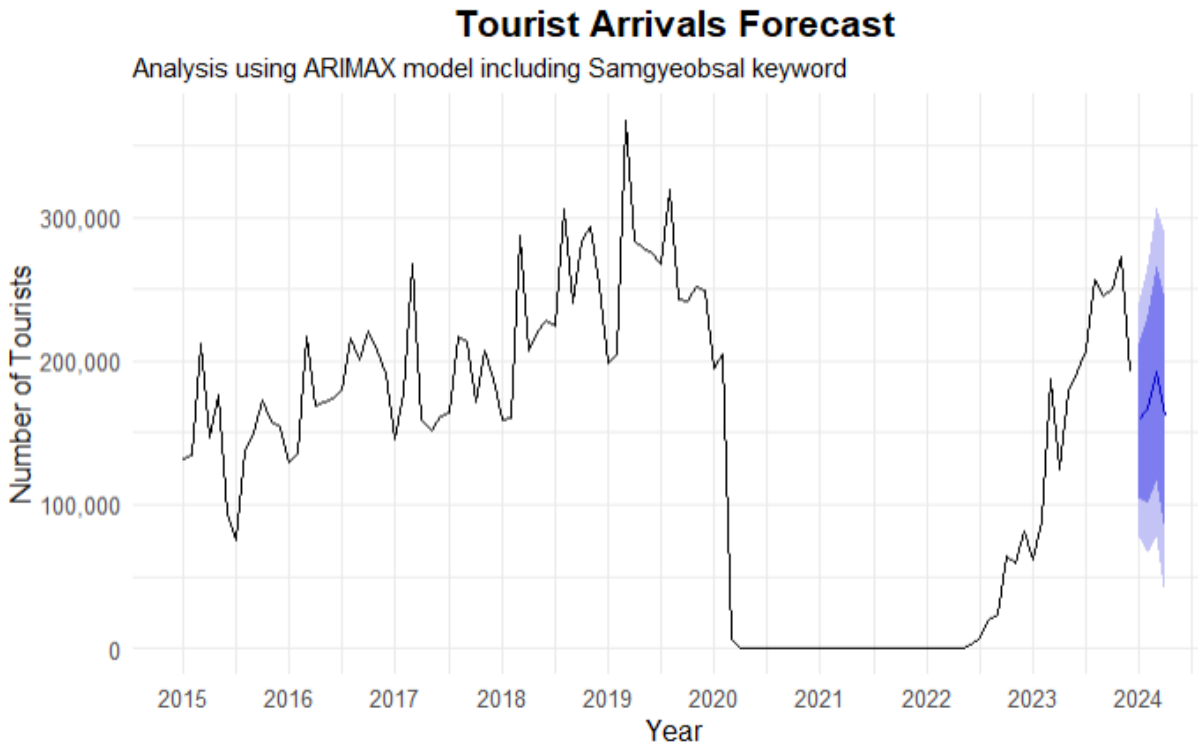




Methodology and Result

SARIMAX model including Samgyeosal keyword

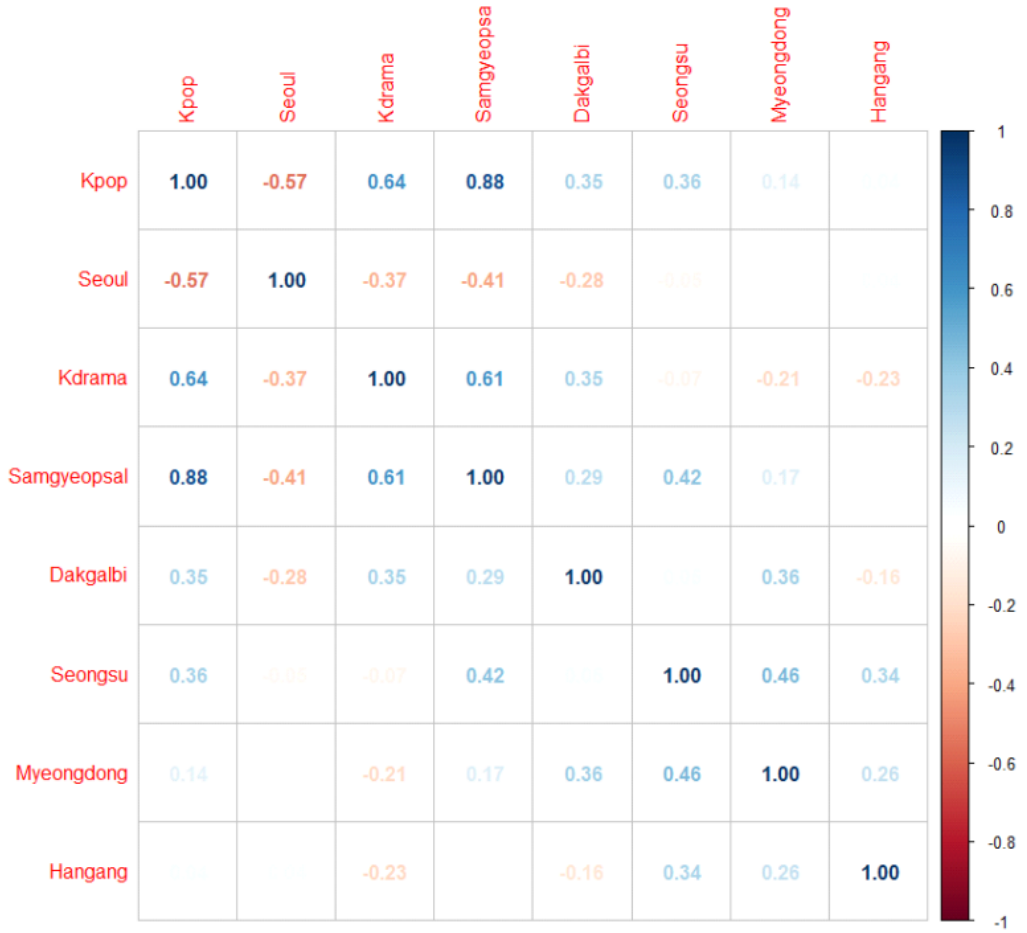
	SARIMA model	SARIMAX (Samgyeobsal)	SARIMAX (Myeongdong)	SARIMAX (idol)
RMSE	56978	24197	82758	116609



Using the SARIMAX model improves the accuracy of the tourist arrival forecasts compared to the basic ARIMA model, as evidenced by the lower RMSE values

Methodology and Result

SARIMAX model including multiple keywords



1. Classification of keywords

: Keywords can be classified into place, food, and so on

: Rather than using one exogenous variable, the model could include more than two exogenous variables

2. Hit-map representation

: Hit-map shows that Samgyeopsal and K-pop have a high correlation(0.88)

: Samgyeopsal has a more powerful impact on the dependent variable than that of K-pop

: To avoid multicollinearty, K-pop is removed.

Methodology and Result

SARIMAX model including multiple keywords

```
[1] "Lags - Seoul: 0 Kdrama: 0 Samgyeopsal: 0 Seongsu: 0 Myeongdong: 0 Hangang: 0 RMSE: 79076.925640823"
[1] "Lags - Seoul: 0 Kdrama: 0 Samgyeopsal: 0 Seongsu: 0 Myeongdong: 0 Hangang: 1 RMSE: 60693.9451000964"
[1] "Lags - Seoul: 0 Kdrama: 0 Samgyeopsal: 0 Seongsu: 0 Myeongdong: 1 Hangang: 1 RMSE: 47913.7061307375"
[1] "Lags - Seoul: 0 Kdrama: 0 Samgyeopsal: 0 Seongsu: 1 Myeongdong: 1 Hangang: 1 RMSE: 47846.0603930127"
[1] "Lags - Seoul: 0 Kdrama: 0 Samgyeopsal: 0 Seongsu: 2 Myeongdong: 1 Hangang: 1 RMSE: 44727.6744806252"
[1] "Lags - Seoul: 0 Kdrama: 1 Samgyeopsal: 0 Seongsu: 1 Myeongdong: 0 Hangang: 3 RMSE: 42496.621128885"
[1] "Lags - Seoul: 0 Kdrama: 1 Samgyeopsal: 3 Seongsu: 3 Myeongdong: 0 Hangang: 4 RMSE: 40561.4724101601"
[1] "Lags - Seoul: 0 Kdrama: 1 Samgyeopsal: 3 Seongsu: 4 Myeongdong: 0 Hangang: 4 RMSE: 40484.9237790434"
[1] "Lags - Seoul: 0 Kdrama: 2 Samgyeopsal: 0 Seongsu: 2 Myeongdong: 0 Hangang: 4 RMSE: 38473.492382454"
[1] "5000!"
[1] "Lags - Seoul: 2 Kdrama: 1 Samgyeopsal: 3 Seongsu: 0 Myeongdong: 0 Hangang: 1 RMSE: 36634.8516281763"
[1] "Lags - Seoul: 2 Kdrama: 1 Samgyeopsal: 3 Seongsu: 0 Myeongdong: 0 Hangang: 4 RMSE: 36036.9926211054"
[1] "Lags - Seoul: 2 Kdrama: 1 Samgyeopsal: 3 Seongsu: 1 Myeongdong: 0 Hangang: 1 RMSE: 34595.0947397314"
[1] "Lags - Seoul: 2 Kdrama: 1 Samgyeopsal: 3 Seongsu: 1 Myeongdong: 0 Hangang: 4 RMSE: 34217.9238013849"
[1] "Lags - Seoul: 2 Kdrama: 4 Samgyeopsal: 3 Seongsu: 4 Myeongdong: 0 Hangang: 4 RMSE: 34038.6386693189"
[1] "5000!"
[1] "5000!"
> print(best_lags)
```

```
> print(best_lags)
$Seoul
[1] 2

$Kdrama
[1] 4

$Samgyeopsal
[1] 3

$Seongsu
[1] 4

$Myeongdong
[1] 0

$Hangang
[1] 4

> print(paste("Best RMSE:", best_rmse))
[1] "Best RMSE: 34038.6386693189"
```

```
> forecast(fit, xreg = as.matrix(exog_test), h = nrow(test_y))
     Point Forecast    Lo 80    Hi 80     Lo 95    Hi 95
169       153366.1 117022.4 189709.7  97783.28 208948.8
170       209607.8 170005.7 249209.9 149041.59 270173.9
171       278931.1 238290.6 319571.6 216776.86 341085.4
172       243072.9 202285.8 283860.0 180694.40 305451.4
```

1. Methodology

: If a Japanese searched any keywords in Jan, 2024, it is more likely to come Korea after one to four months.

: Including lag from 0 to 4 and 5 number of keywords, the loop generated by data from January 2010 to December 2023, tests 5^6 models by RMSE criteria which is calculated by out-of-sample data, from January 2024 to April 2024.

: Chicken Rib is removed as the data of earlier period is almost 0 and the number of data is different from that of DV.

2. Result

: The lags of variables are 2,4,3,4,0,4 for Seoul, K-Drama, Samgeopsal, Seongsu, Myeongdong, and Hangang.

: RMSE is lower than that of SARIMAX with one keyword.

Methodology and Result

SARIMAX model including multiple keywords

```
> summary(fit)
Series: as.numeric(train_y)
Regression with ARIMA(3,1,3) errors

Coefficients:
          ar1      ar2      ar3      ma1     ma2      ma3       drift  Seoul_lag2  Kdrama_lag4  Samgyeopsal_lag3
      -0.1677  -0.8811  -0.0571  -0.3993  0.605  -0.8200  -2255.4608    113.2918     110.6194         1779.4424
s.e.   0.1050   0.0471   0.0969   0.0779  0.042   0.0917    436.4001    191.7720     399.0480          320.2614
      Seongsu_lag4  Myeongdong_lag0  Hangang_lag4
          650.0689          3279.8770    -1294.0085
s.e.      406.1910           185.5412      495.8653

sigma^2 = 803663858:  log likelihood = -1945.15
AIC=3918.31   AICc=3921.07   BIC=3961.96

Training set error measures:
                    ME     RMSE      MAE       MPE     MAPE      MASE         ACF1
Training set 163.2184 27142.07 21182.66 -461.2268 4297.359 0.7659747 -0.0009930084
> fit
Series: as.numeric(train_y)
Regression with ARIMA(3,1,3) errors

Coefficients:
          ar1      ar2      ar3      ma1     ma2      ma3       drift  Seoul_lag2  Kdrama_lag4  Samgyeopsal_lag3
      -0.1677  -0.8811  -0.0571  -0.3993  0.605  -0.8200  -2255.4608    113.2918     110.6194         1779.4424
s.e.   0.1050   0.0471   0.0969   0.0779  0.042   0.0917    436.4001    191.7720     399.0480          320.2614
      Seongsu_lag4  Myeongdong_lag0  Hangang_lag4
          650.0689          3279.8770    -1294.0085
s.e.      406.1910           185.5412      495.8653

sigma^2 = 803663858:  log likelihood = -1945.15
AIC=3918.31   AICc=3921.07   BIC=3961.96
```

1. Full Model

: The impact of each variable can be interpreted by coefficients.

: The values of coefficients are on the left.

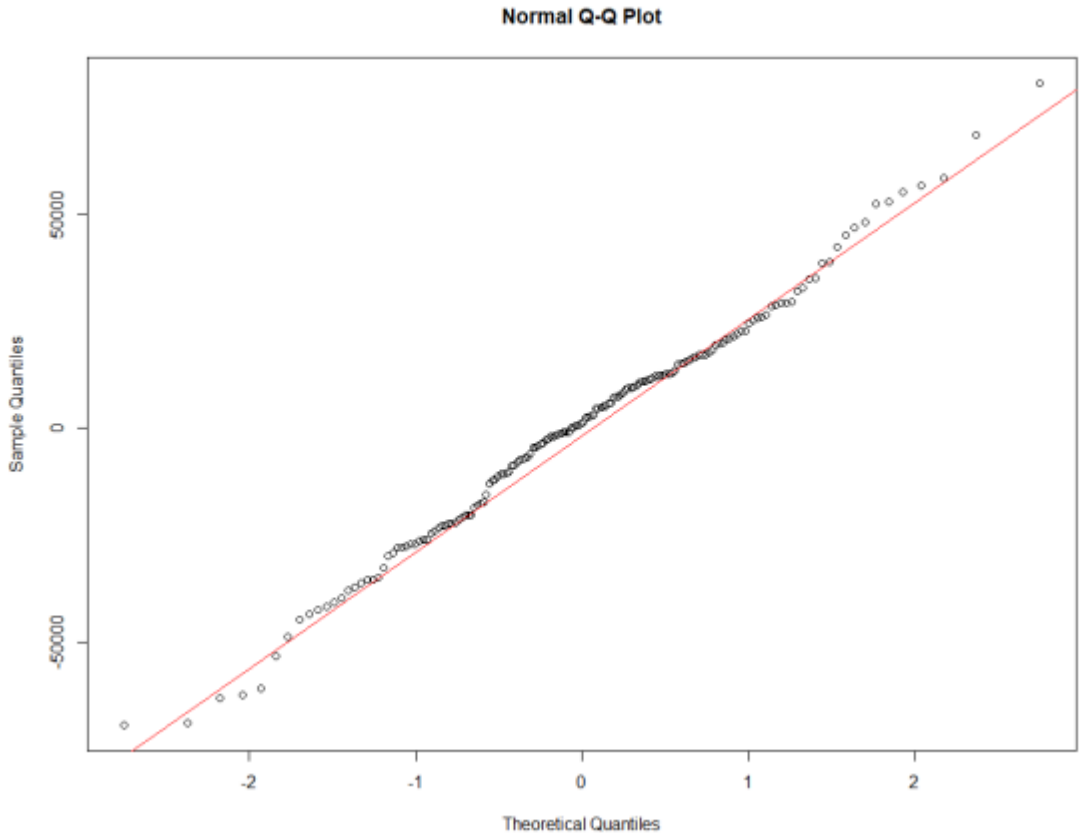
2. Result

: Searches for samgyeopsal before a trip significantly influence the number of tourists in the subsequent period

: Searches for Myeongdong during the current month influence the number of tourists in that same month

Methodology and Result

Q-Q plot



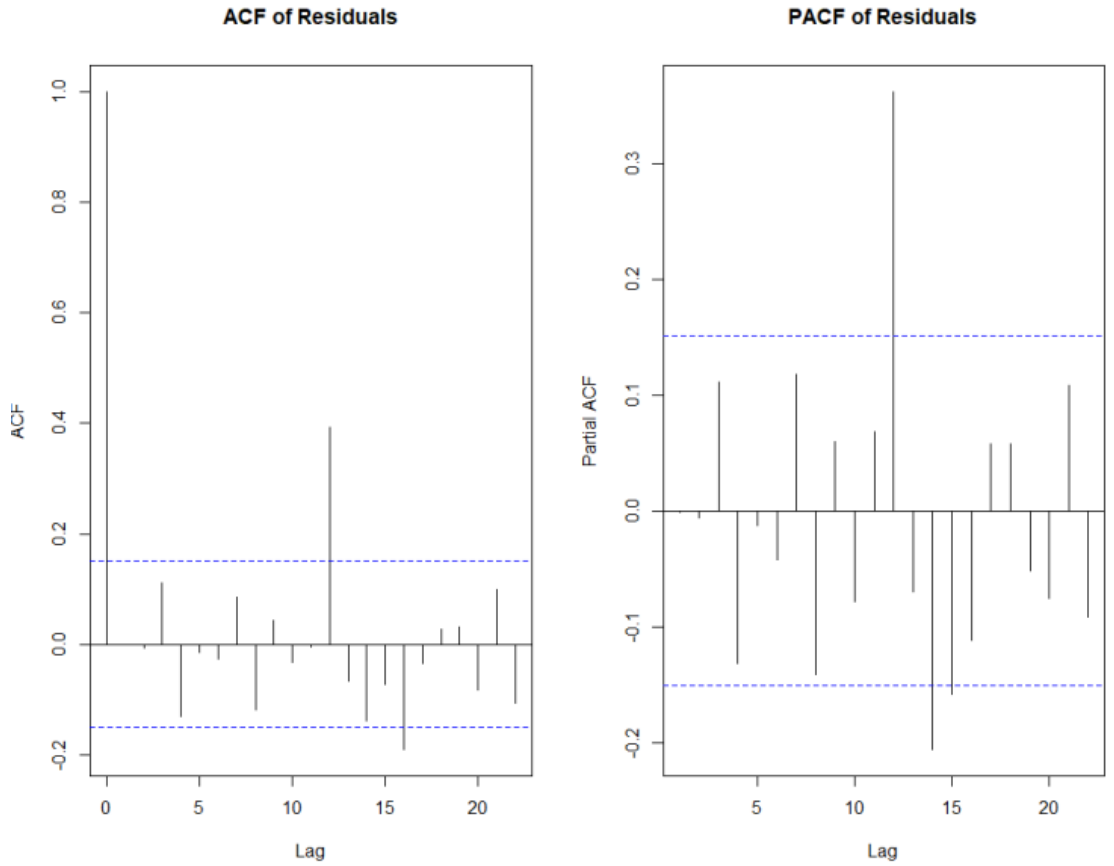
Methodology and Result

Box-Ljung test and ACF and PACF of residuals

```
> print(ljung_box_test)

        Box-Ljung test

data:  residuals
X-squared = 51.79, df = 20, p-value = 0.0001222
```



The residuals of the model includes autocorrelation

Methodology and Result

Manually fitted SARIMAX model including multiple keywords

```
> fit_seasonal <- Arima(train_y, order = c(3, 1, 3), seasonal = list(order = c(1, 1, 3), period = 12), xreg =
as.matrix(exog_train_ts))
> summary(fit_seasonal)
Series: train_y
Regression with ARIMA(3,1,3)(1,1,3)[12] errors

Coefficients:
          ar1      ar2      ar3      ma1     ma2      ma3     sar1     sma1     sma2     sma3     Seoul
Kdrama  Samgyeopsal  Seongsu
      -0.3776  -0.9226  -0.1476  -0.1705  0.6588  -0.4847  -0.1698  -0.5424  -0.2158  -0.1032   53.1783
-705.3013     952.8058  249.3340
s.e.   0.1717   0.0696   0.1587   0.1627  0.0857   0.1538   0.5717   0.5929   0.4123   0.1397  163.6108  338.6328
   305.3494  329.6895
      Myeongdong     Hangang
       3233.2863  -402.9109
s.e.    183.8627   487.3961

sigma^2 = 458717199:  log likelihood = -1765.09
AIC=3564.18   AICc=3568.65   BIC=3615.92

Training set error measures:
                    ME     RMSE      MAE       MPE     MAPE      MASE         ACF1
Training set 1152.745 19481.63 14320.72 2046.522 6410.242 0.5178436 -0.0007753541
> forecast_values <- forecast(fit_seasonal, xreg = as.matrix(exog_test_ts), h = length(test_y_ts))
> predicted_values <- as.numeric(forecast_values$mean)
> rmse_value_seasonal <- rmse(as.numeric(test_y_ts), predicted_values)
> print(paste("Test Data RMSE:", rmse_value_seasonal))
[1] "Test Data RMSE: 21798.8658018887"
```

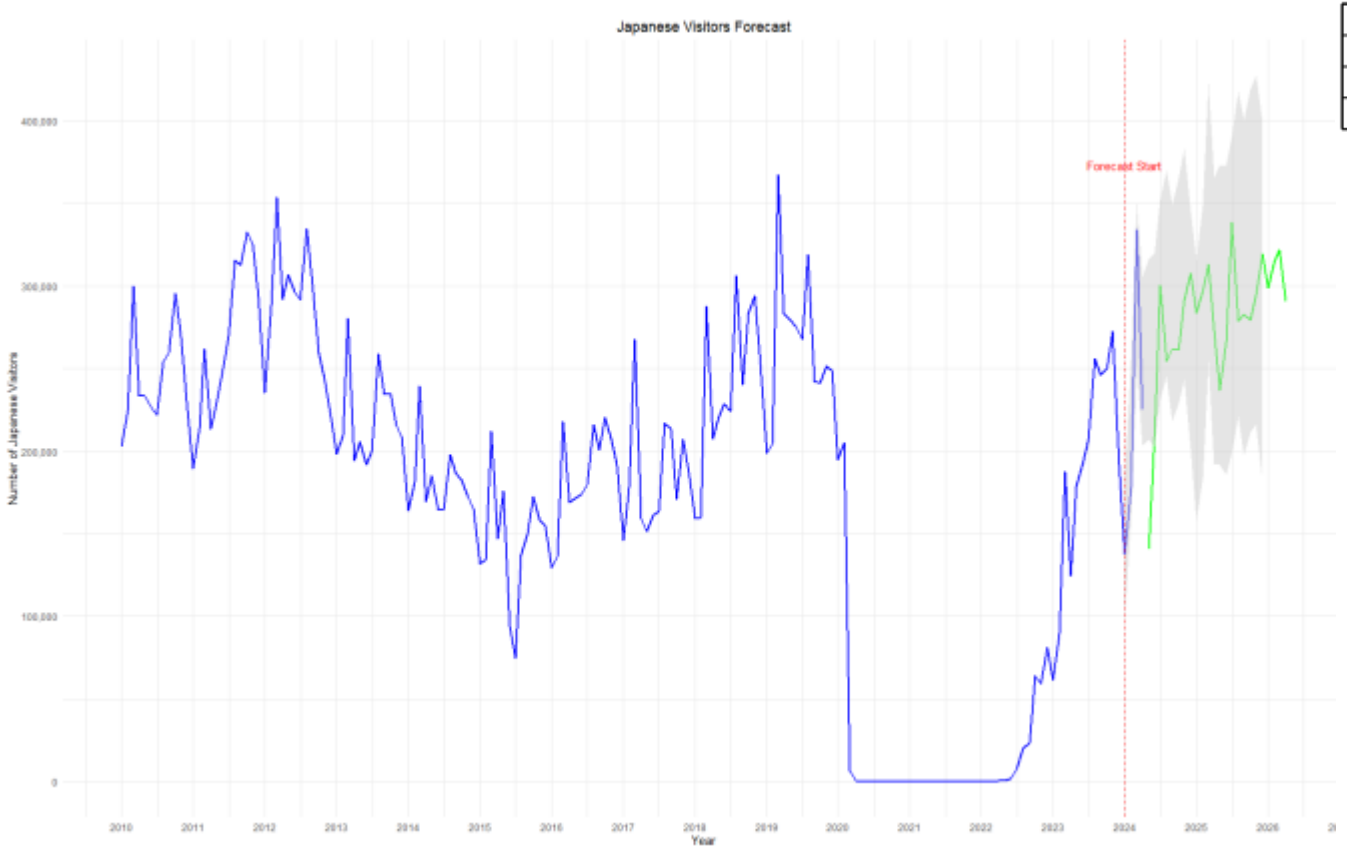
```
> forecast_values$mean
Time Series:
Start = 169
End = 172
Frequency = 1
[1] 140679.8 205918.3 306263.6 242939.0
```

Result

: The new model has much more improved prediction RMSE results than the both prior full model and the each model using just one key-word data

Methodology and Result

Manually fitted SARIMAX model including multiple keywords



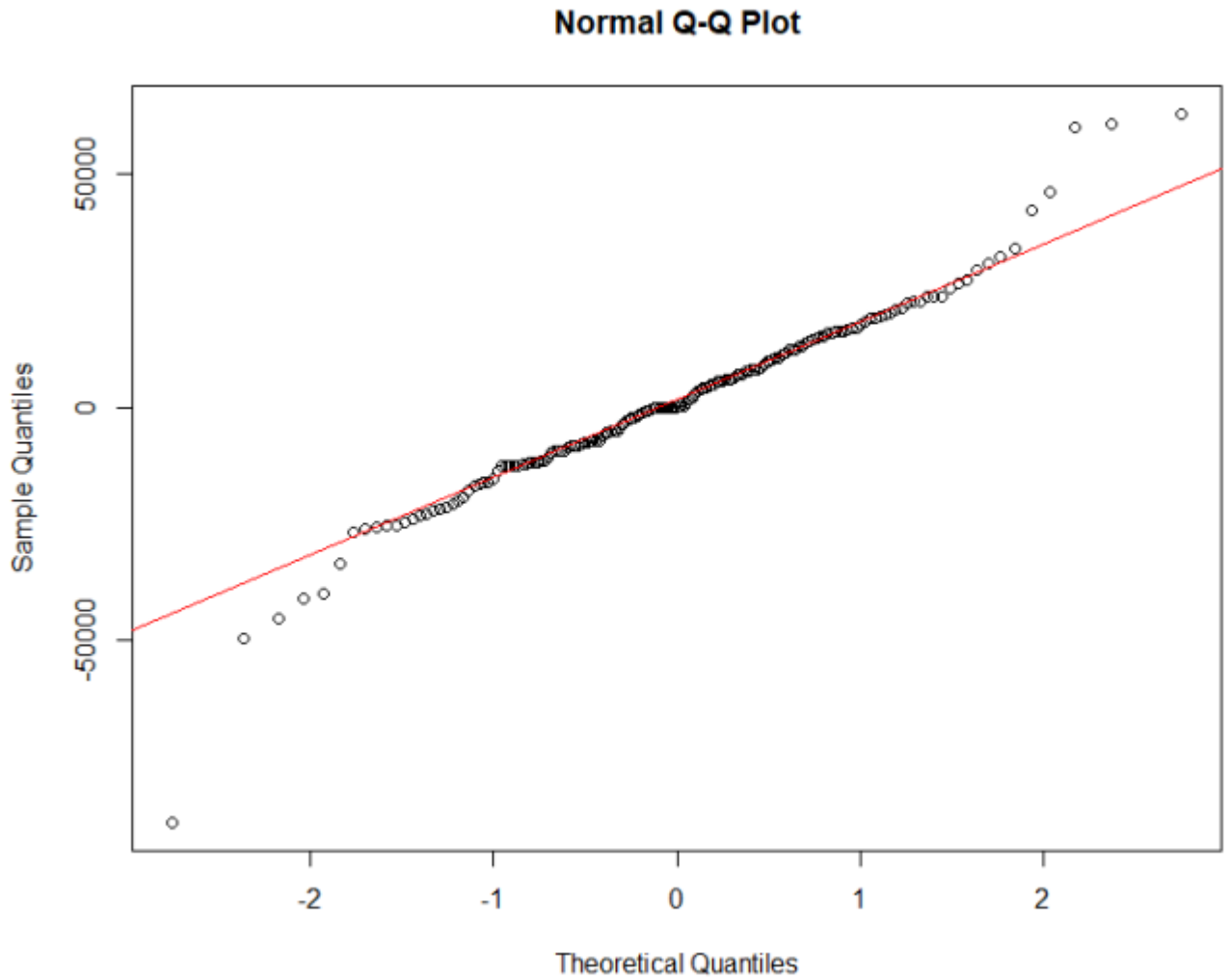
Y-M	24-05	24-06	24-07	24-08	24-09	24-10	24-11	24-12	25-01	25-02
unit	261743	261166	291340	307600	283644	296923	313303	272992	236685	266208
Y-M	25-03	25-04	25-05	25-06	25-07	25-08	25-09	25-10	25-11	25-12
unit	338451	278730	282611	279437	294538	319504	298741	314916	321939	290615

Result

: The prediction result of the Japanese visitors during May 2024 to Dec 2025 are shown above chart and the left graph is the forecasting plot by using the final model

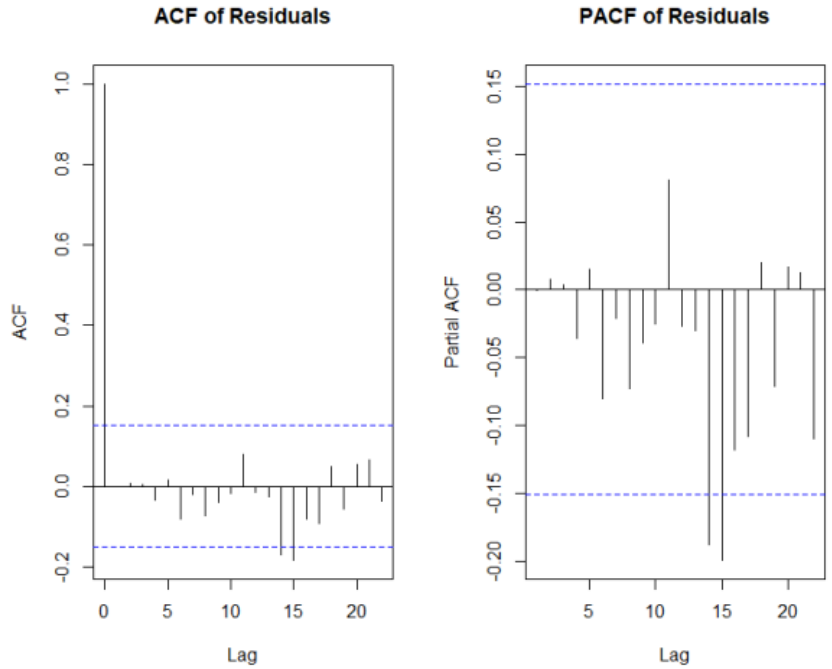
Methodology and Result

Q-Q plot



Methodology and Result

Revised Full-model test results




```
> print(ljung_box_test)

        Box-Ljung test

data:  residuals
X-squared = 18.509, df = 17, p-value = 0.3574
```

Result

: No autocorrelation is found in the model

Expected Original Contribution

Significance

1. The number of Japanese tourists visiting Korea declined in the late 2010s and around the time of COVID-19, but is recovering after COVID-19 and has reached close to its previous peak.

2. Examining the forecasting models used to predict future demand for Japanese tourists visiting Korea to provide an opportunity for practitioners to supplement and develop the results of this study

3. We checked what factors influence Japanese tourists through Google Trends and considered various related policies.

4. The Most interesting part is the search volume of Korean dramas in Japan has a weaker relationship with Japanese people's visits to Korea than other factors

Expected Original Contribution

What we thoughts on Tourism Policy by Time-Series

Korean wave elements such as K-drama didn't have an effect on the increase in the number of Japanese tourists, making it difficult to reflect them in policy.

Expected Original Contribution

What we thoughts on Tourism Policy by Time-Series

Like the German Beer Festival, develop a cultural festival where foreigners can sample and experience the taste of one Korean food element, such as Korean style Pork belly and pancakes, for free event.

Detail Policy Recommendations - Place

For Myeongdong, it was found that tourists tend to search for this specific area in advance of their visit, with a very significant impact observed within a one-month lag. In contrast, the effect for samgyeopsal was observed with a three-month lag.

Based on these results, for Myeongdong, it is clear that tourists do not plan their visits far in advance. Therefore, continuously promoting this area through events could be effective. For instance, marketing campaigns targeting Myeongdong should not be executed three to four months in advance, as such long lead times do not significantly influence travel decisions. Given the short lag, we suggest small-scale, continuous promotions.

Detail Policy Recommendations - Food

For samgyeopsal, the analysis shows a significant impact with a relatively long lag. Typically, tourists make specific plans for their trips closer to the departure date, but they consider what to eat much earlier. This is further supported by the high correlation between samgyeopsal and k-idol searches. The longer lag can be easily understood when considering that planning to attend an idol concert or fan meeting might be reflected in the search data well in advance of the trip for about 3 months.

Therefore, we propose conducting promotional events centered around well-established franchise restaurants or reputable individual eateries that properly represent Korean samgyeopsal culture, focusing on tourist hotspots in Seoul popular among Japanese visitors. Additionally, such promotions can be integrated into group tour packages, as suggested by previous research.

Detail Policy Recommendations - Influencer

Next, idols may be able to supplement the low season 'Summer' as well as make stronger at spring and fall.

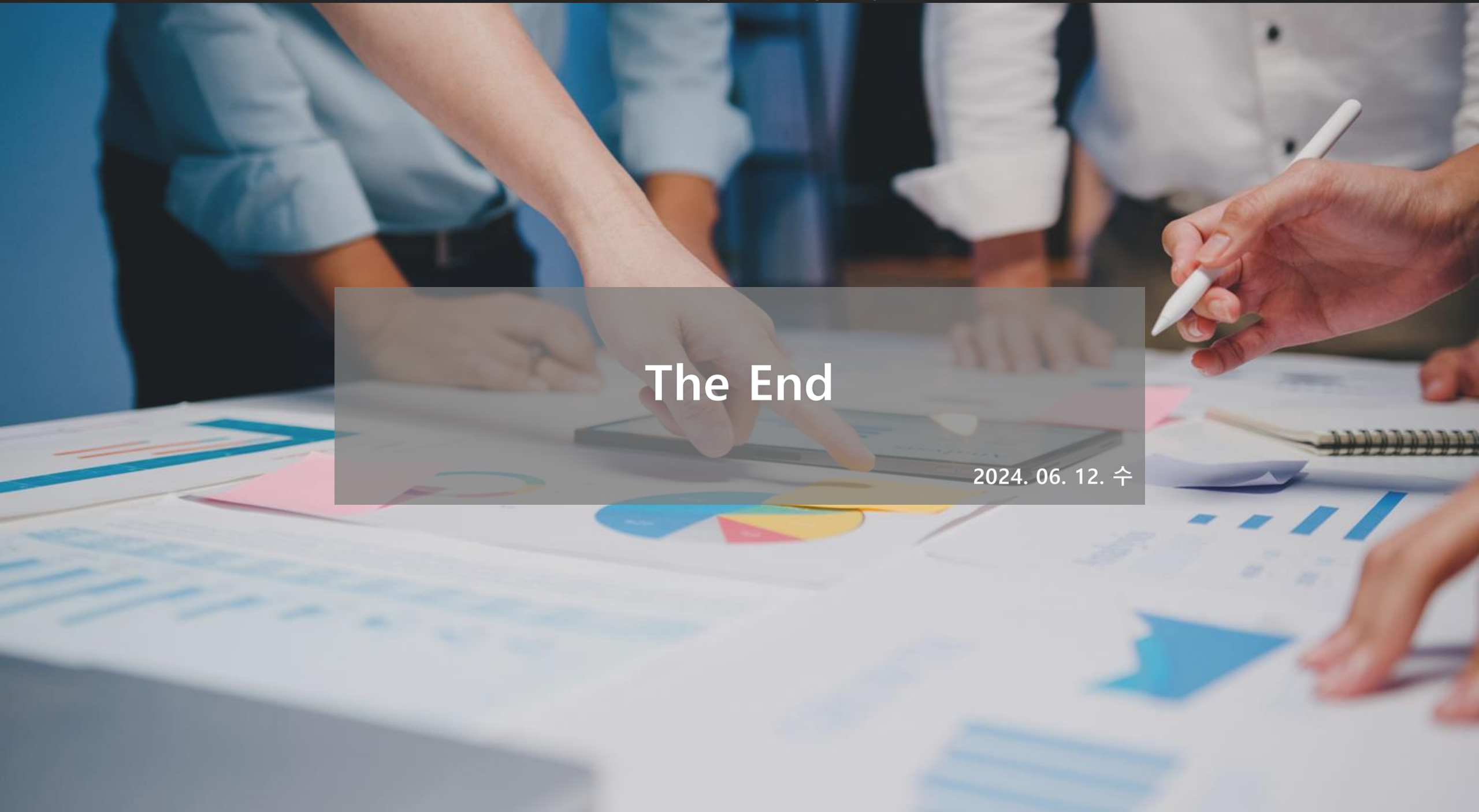
Through additional research on Idol groups that Japanese prefer, it is possible to make a waterbomb concert that is usually held in summer and implement strategies to induce them to visit Korea more easily by linking tour packages including these festivals.

Through this approach, we will be able to find ways to increase the number of Japanese visitors to Korea in summer.

Meanwhile, the importance (search volume, etc.) between keywords not covered in the PPT will be covered in detail in the report.

References

1. Ministry of Culture, Sports, and Tourism, South Korea. "Number of Domestic Travel and Increasing Rate." Accessed [JUNE 20, 2024].
2. 김보경. "한류 알려온 '케이콘' 12년 역사 처음으로 9월 독일서 개최." Naver News. June 21, 2024.
3. Wells, Pete. "The Top Dishes of 2023 in NYC." The New York Times. December 13, 2023, https://www.nytimes.com/2023/12/13/dining/top-dishes-nyc-2023.html
4. Ministry of Culture, Sports, and Tourism, South Korea. "Annual Number of Travelers from Japan and Total." Accessed [June 20, 2024].
5. Japan National Tourism Organization (JNTO) and Ministry of Culture, Sports, and Tourism, South Korea. "Monthly Number of Korean Tourists Visiting Japan and Japanese Tourists Visiting Korea." Accessed June 23, 2024.
6. Molemeng. (2023). The Factors Influencing the Number of Visits and Behavioral Intentions
7. Deukhee Park, Sanghoon Kang, and Gyehee Lee. "Demand Forecasting of Japanese Tourists to Korea for Sustainable Tourism Growth: An Application of Time-Series Econometric Models." Journal of Tourism Studies 34, no. 3 (2020): 47-60.
8. Soeon Park, Gunhee Lee, and Inhye Lee. "Analysis of Factors of Visiting Tourists to Korea using Economic Variables and Media Data: A Case Study of Japan and Chinese Tourist." International Journal of Tourism and Hospitality Research 29, no. 7 (2015): 57-72



The End

2024. 06. 12. 수