

연습문제 풀이

문제 1

패키지 참조

데이터 가져오기

데이터 타입 확인

날짜 인덱스 설정

결측치 검사

데이터 검정

ARIMA 분석

분석 모델 만들기

학습 모델에 대한 예측치

학습한 내용을 토대로 1년간의 예상치 생성

시각화

연습문제 풀이

문제 1

패키지 참조

```
import sys
sys.path.append("../..")

from datetime import datetime as dt
from datetime import timedelta
from pandas import read_excel, to_datetime
from matplotlib import pyplot as plt
from matplotlib import dates as mdates
from statsmodels.tsa.arima.model import ARIMA
from pmdarima.arima import auto_arima
import seaborn as sb

from helper import set_datetime_index, exp_time_data
```

데이터 가져오기

```
origin = read_excel("https://data.hossam.kr/E06/temperatures_seoul.xlsx")
```

연습문제 풀이

문제 1

패키지 참조

데이터 가져오기

데이터 타입 확인

날짜 인덱스 설정

결측치 검사

데이터 검정

ARIMA 분석

분석 모델 만들기

학습 모델에 대한 예측치

학습한 내용을 토대로 1년간의 예상치 생성

시각화

```
origin.head()
```

	날짜	평균기온(°C)
0	1973-07-01	22.7
1	1973-07-02	23.8
2	1973-07-03	27.2
3	1973-07-04	25.2
4	1973-07-05	26.9

데이터 타입 확인

```
origin.dtypes
```

```
날짜          datetime64[ns]  
평균기온(°C)    float64  
dtype: object
```

날짜 인덱스 설정

```
df = set_datetime_index(origin, '날짜')  
df.head()
```

연습문제 풀이

문제 1

패키지 참조

데이터 가져오기

데이터 타입 확인

날짜 인덱스 설정

결측치 검사

데이터 검정

ARIMA 분석

분석 모델 만들기

학습 모델에 대한 예측치

학습한 내용을 토대로 1년간의 예상치 생성

시각화

	평균기온(°C)
1973-07-01	22.7
1973-07-02	23.8
1973-07-03	27.2
1973-07-04	25.2
1973-07-05	26.9

결측치 검사

```
df.isna().sum()
```

```
평균기온(°C)    0
dtype: int64
```

데이터 검정

데이터가 너무 많으므로 최신 1000건만으로 검정을 시도함

```
df2 = df.tail(365*5)
df2.head()
```

	평균기온(°C)
2018-08-08	31.0

연습문제 폴이

문제 1

패키지 참조

데이터 가져오기

데이터 타입 확인

날짜 인덱스 설정

결측치 검사

데이터 검정

ARIMA 분석

분석 모델 만들기

학습 모델에 대한 예측치

학습한 내용을 토대로 1년간의 예상치 생성

시각화

2018-08-09	평균기온(°C)
2018-08-10	30.7
2018-08-11	31.0
2018-08-12	30.6

```
exp_time_data(data=df2, yname="평균기온(°C)", sd_model="a", max_diff=10)
```

결측치 수: 0

연습문제 풀이

문제 1

패키지 참조

데이터 가져오기

데이터 타입 확인

날짜 인덱스 설정

결측치 검사

데이터 검정

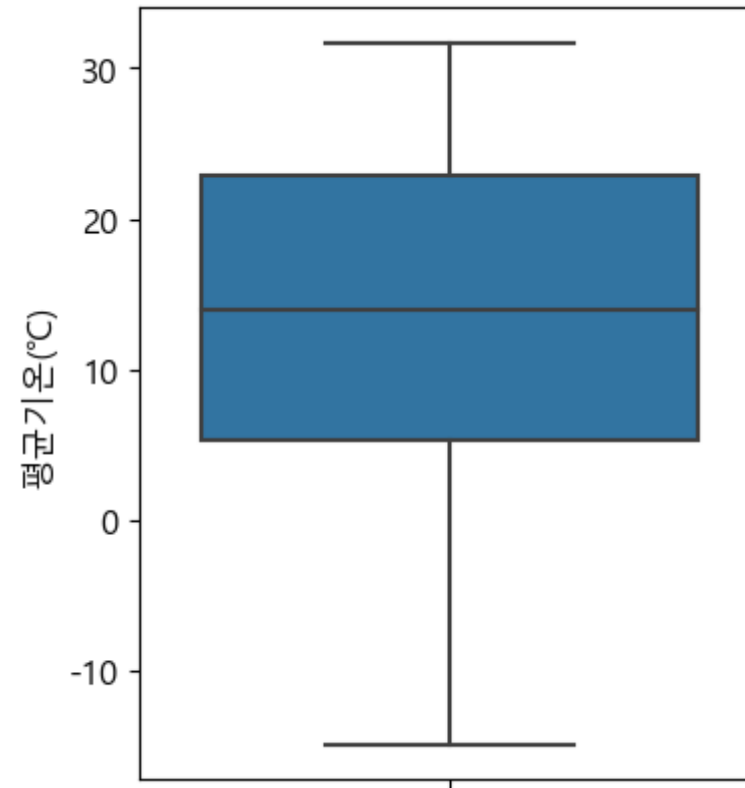
ARIMA 분석

분석 모델 만들기

학습 모델에 대한 예측치

학습한 내용을 토대로 1년간의 예상치 생성

시각화



연습문제 풀이

문제 1

패키지 참조

데이터 가져오기

데이터 타입 확인

날짜 인덱스 설정

결측치 검사

데이터 검정

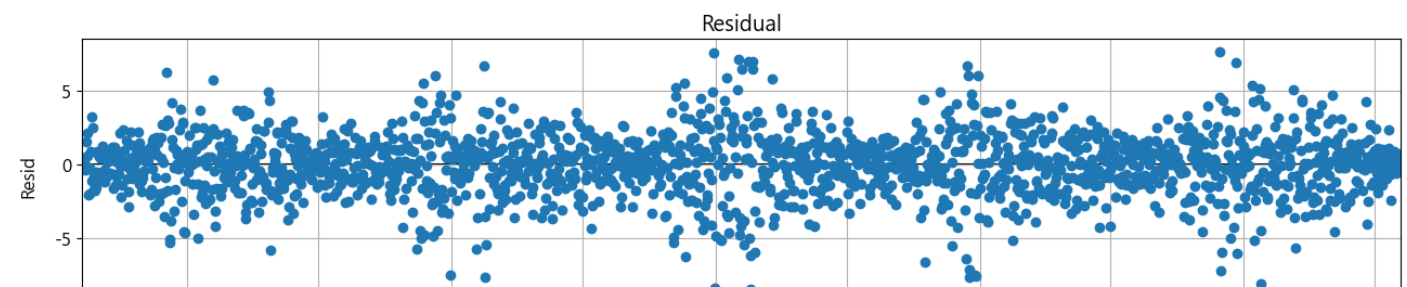
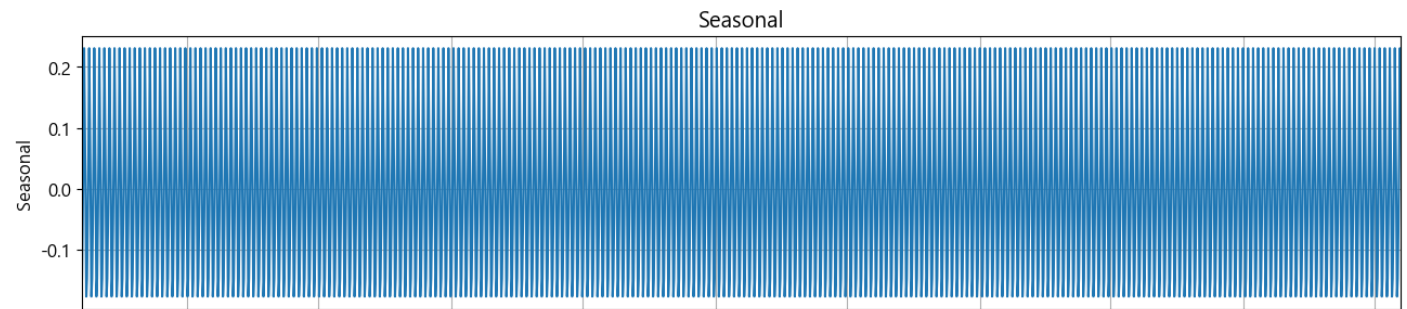
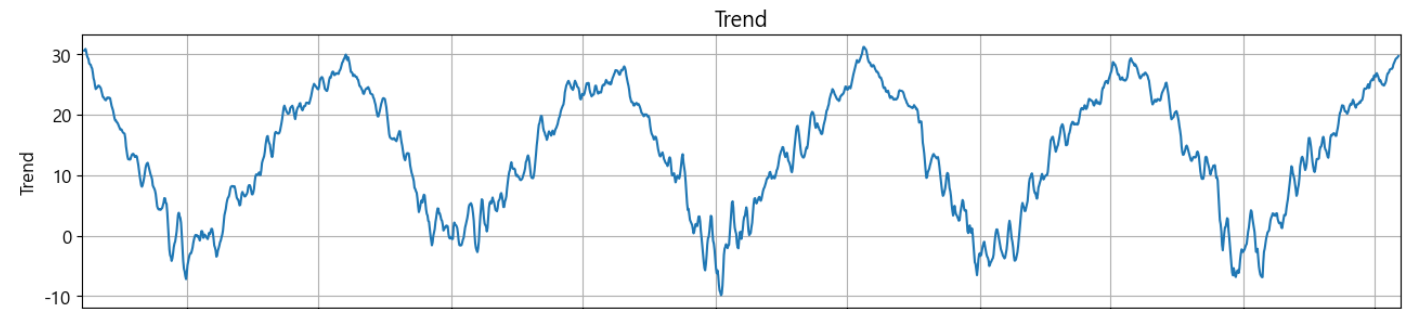
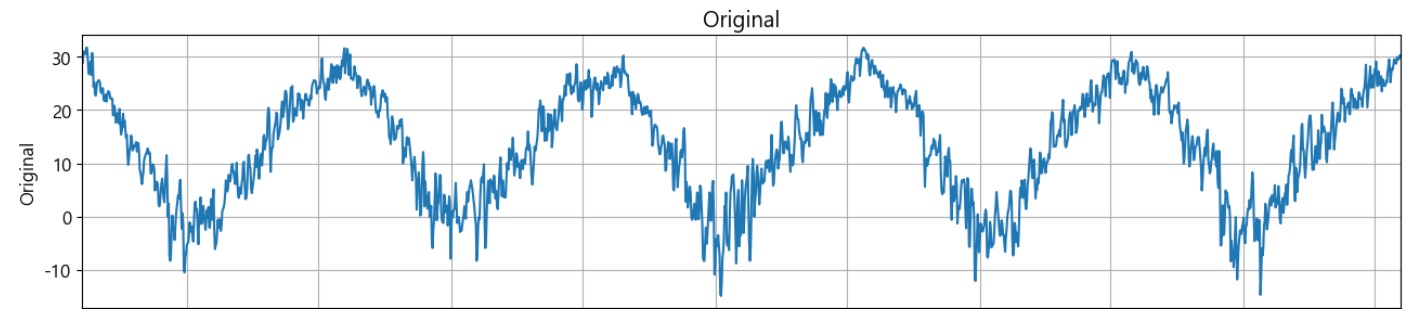
ARIMA 분석

분석 모델 만들기

학습 모델에 대한 예측치

학습한 내용을 토대로 1년간의 예상치 생성

시각화



연습문제 풀이

문제 1

패키지 참조

데이터 가져오기

데이터 타입 확인

날짜 인덱스 설정

결측치 검사

데이터 검정

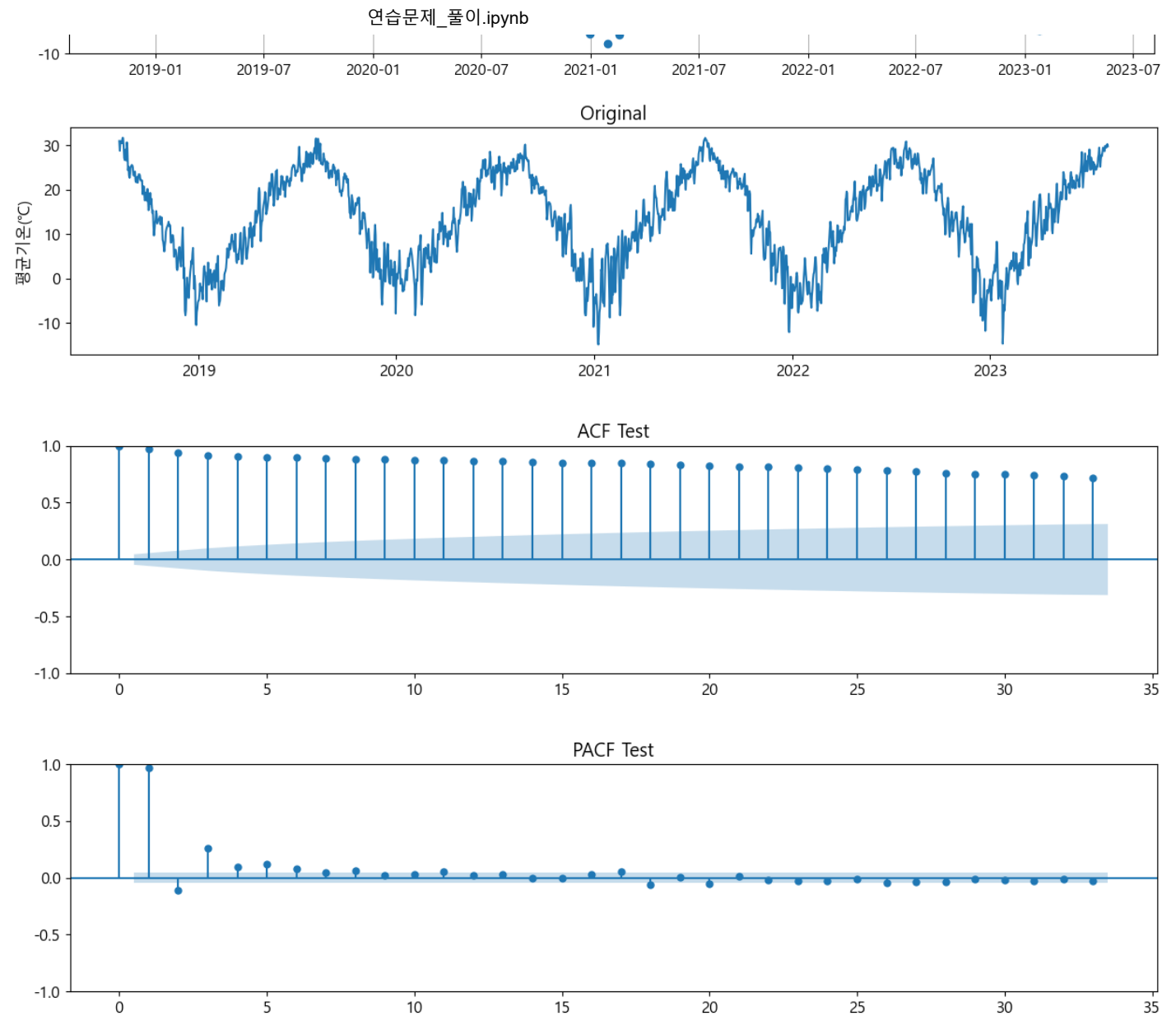
ARIMA 분석

분석 모델 만들기

학습 모델에 대한 예측치

학습한 내용을 토대로 1년간의 예상치 생성

시각화



===== 원본 데이터 =====

```

+-----+
| ADF Test |

```

연습문제 풀이

문제 1

패키지 참조

데이터 가져오기

데이터 타입 확인

날짜 인덱스 설정

결측치 검사

데이터 검토

ARIMA 분석

분석 모델 만들기

학습 모델에 대한 예측치

학습한 내용을 토대로 1년간의 예상치 생성

시각화

```

+-----+-----+
| 검정통계량(ADF Statistic) | -2.00483 |
| 유의수준(p-value) | 0.284458 |
| 최적차수(num of lags) | 19 |
| 관측치 개수(num of observations) | 1805 |
| 기각값(Critical Values) 1% | -3.43398 |
| 기각값(Critical Values) 5% | -2.86314 |
| 기각값(Critical Values) 10% | -2.56762 |
| 데이터 정상성 여부(0=False,1=True) | 0 |
+-----+-----+
===== 1차 차분 데이터 =====
+-----+-----+
| ADF Test | |
+-----+-----+
| 검정통계량(ADF Statistic) | -11.4068 |
| 유의수준(p-value) | 7.41633e-21 |
| 최적차수(num of lags) | 18 |
| 관측치 개수(num of observations) | 1805 |
| 기각값(Critical Values) 1% | -3.43398 |
| 기각값(Critical Values) 5% | -2.86314 |
| 기각값(Critical Values) 10% | -2.56762 |
| 데이터 정상성 여부(0=False,1=True) | 1 |
+-----+-----+

```

ARIMA 분석

분석 모델 만들기

날씨에 대한 데이터이므로 계절성은 1년마다 돌아온다고 보는 것이 맞지만 컴퓨터 성능상의 한계로 계절성을 1주일 단위(=7일)로 제한함

연습문제 풀이

문제 1

패키지 참조

데이터 가져오기

데이터 타입 확인

날짜 인덱스 설정

결측치 검사

데이터 검정

ARIMA 분석

분석 모델 만들기

학습 모델에 대한 예측치

학습한 내용을 토대로 1년간의 예상치 생성

시각화

Warnings:

```
[1] Covariance matrix calculated using the outer product of gradients (c
```

학습 모델에 대한 예측치

```
fv = fit.fittedvalues
fv.head()
```

```
2018-08-08    0.000000
2018-08-09    31.000016
2018-08-10    28.799999
2018-08-11    30.700002
2018-08-12    31.000000
Freq: D, dtype: float64
```

학습한 내용을 토대로 1년간의 예상치 생성

```
fc = fit.forecast(365)
fc.head()
```

```
2023-08-07    28.685755
2023-08-08    29.648218
2023-08-09    30.412096
2023-08-10    28.447706
```

연습문제 풀이

문제 1

패키지 참조

데이터 가져오기

데이터 타입 확인

날짜 인덱스 설정

결측치 검사

데이터 검정

ARIMA 분석

분석 모델 만들기

학습 모델에 대한 예측치

학습한 내용을 토대로 1년간의 예상치 생성

시각화

2023-08-11 29.666272
 Freq: D, Name: predicted_mean, dtype: float64

시각화

```
last = df.index.max()
xmin = last-timedelta(days=365)
xmax = last+timedelta(days=365+10)
ymax = df['평균기온(℃)'][xmin:xmax].max()
ymin = df['평균기온(℃)'][xmin:xmax].min()
xmin, xmax, ymax, ymin
```

```
(Timestamp('2022-08-06 00:00:00'),
 Timestamp('2024-08-15 00:00:00'),
 30.3,
 -14.7)
```

```
plt.figure(figsize=(20,8))
```

원본 데이터

```
sb.lineplot(data=df, x=df.index, y='평균기온(℃)', label='Original')
```

원본에 대한 학습결과

```
sb.lineplot(x=fv.index, y=fv.values, label='FittedValues', linestyle='--')
```

향후 1년간의 예측값

```
sb.lineplot(x=fc.index, y=fc.values, label='Predict', linestyle='--', co
```

연습문제 풀이

문제 1

패키지 참조

데이터 가져오기

데이터 타입 확인

날짜 인덱스 설정

결측치 검사

데이터 검정

ARIMA 분석

분석 모델 만들기

학습 모델에 대한 예측치

학습한 내용을 토대로 1년간의 예상치 생성

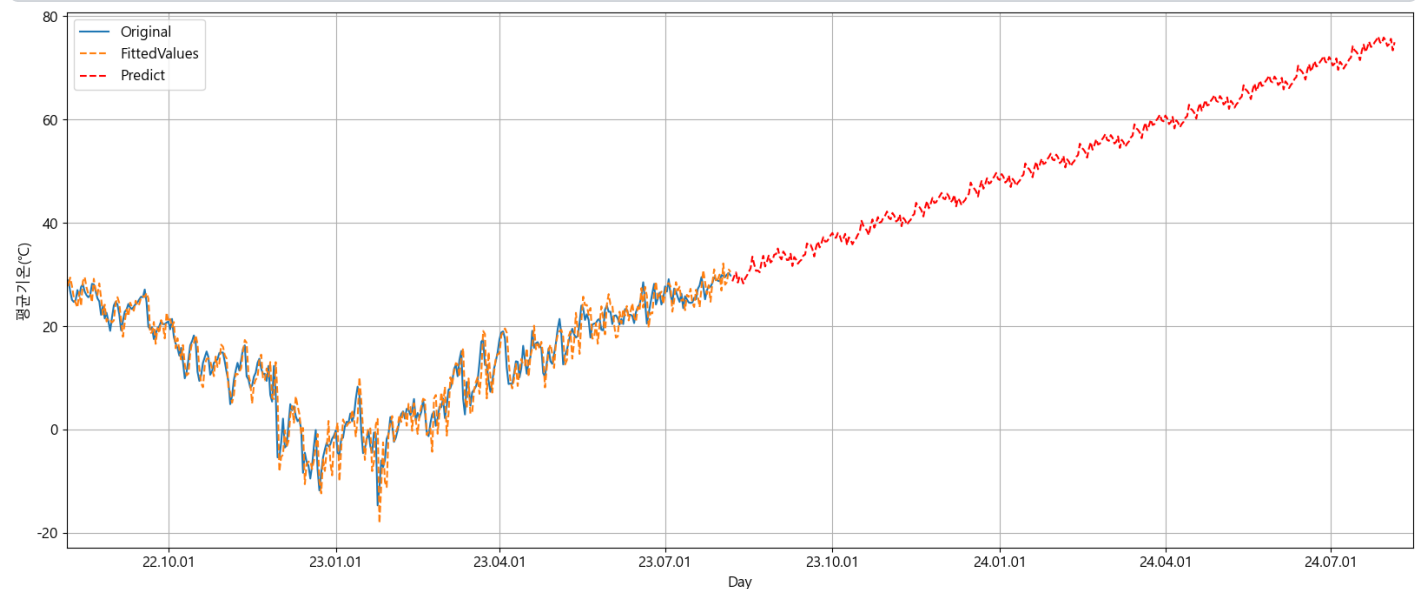
시각화

```
plt.xlabel('Day')
plt.ylabel('평균기온(℃)')
plt.legend()

plt.xlim([xmin, xmax])
#plt.ylim([ymin * 0.8, ymax*1.2])

# 그래프의 x축이 날짜로 구성되어 있을 경우 형식 지정
monthyearFmt = mdates.DateFormatter('%y.%m.%d')
plt.gca().xaxis.set_major_formatter(monthyearFmt)

plt.grid()
plt.show()
plt.close()
```



연습문제 풀이

문제 1

패키지 참조

데이터 가져오기

데이터 타입 확인

날짜 인덱스 설정

결측치 검사

데이터 검정

ARIMA 분석

분석 모델 만들기

학습 모델에 대한 예측치

학습한 내용을 토대로 1년간의 예상치 생성

시각화