

# ARIMA 모델 기반 전력 생산량 예측

시계열 자료 분석과 예측

---



|      |                  |    |                 |
|------|------------------|----|-----------------|
| 제출일  | 2022 년 12 월 22 일 | 소속 | 정보통신대학원         |
| 과목   | 시계열 자료 분석과 예측    | 팀명 | 6 팀             |
| 담당교수 | 김명석 교수님          | 이름 | 신요섭 서종환 서승덕 손소연 |

# 목차

|   |    |
|---|----|
| 초 록   | 3  |
| <b>I. 서론 (분석의 동기)</b>   | 4  |
| <b>II. 이론 및 선행문헌검토</b>  | 5  |
| 2.1 이론  | 5  |
| 2.2 선행문헌 검토 (문헌연구)  | 7  |
| <b>III. 연구설계(자료 및 방법론 설명)</b>   | 8  |
| 3.1 연구 질문   | 8  |
| 3.2 연구 방법   | 8  |
| 3.3 자료의 수집  | 9  |
| 3.4 데이터 선택, 전처리, 병합   | 9  |
| 3.5 데이터 분석(분석결과 정리보고)   | 10 |
| 1) 데이터 확인   | 10 |
| 2) Stationary 확인  | 11 |
| 3) Augmented Dickey-Fuller Test   | 12 |
| 4) Logarithmic Transformation with Box-Cox                                | 13 |
| 5) Removing Trend with Moving Average                                     | 14 |
| 6) Exponential Decay Transformation                                       | 14 |
| 7) Decomposition  | 16 |
| 8) ACF(AutoCorrelation Function) & PACF(Partial AutoCorrelation Function) | 17 |
| <b>IV. 분석내용</b>   | 18 |
| 4.1 예측 모델   | 18 |
| 1) Persistence Model  | 18 |
| 2) Autoregression Model(AR Model, 자동 회귀 모델)                               | 19 |
| 3) Moving Average Model(이동평균모형)   | 19 |
| 4) ARIMA(AutoRegressive Integrated Moving Average, 자기 회귀 누적 이동 평균 모델)     | 20 |
| 4.2 모델 비교   | 21 |
| <b>V. 결론</b>  | 22 |
| 5.1 분석결과  | 22 |
| 5.2 시사점 및 활용  | 22 |
| 5.3 한계 및 향후과제   | 22 |
| 5.4 느낀점   | 23 |

# ARIMA 모델 기반 전력 생산량 예측(제목)

## 초 록

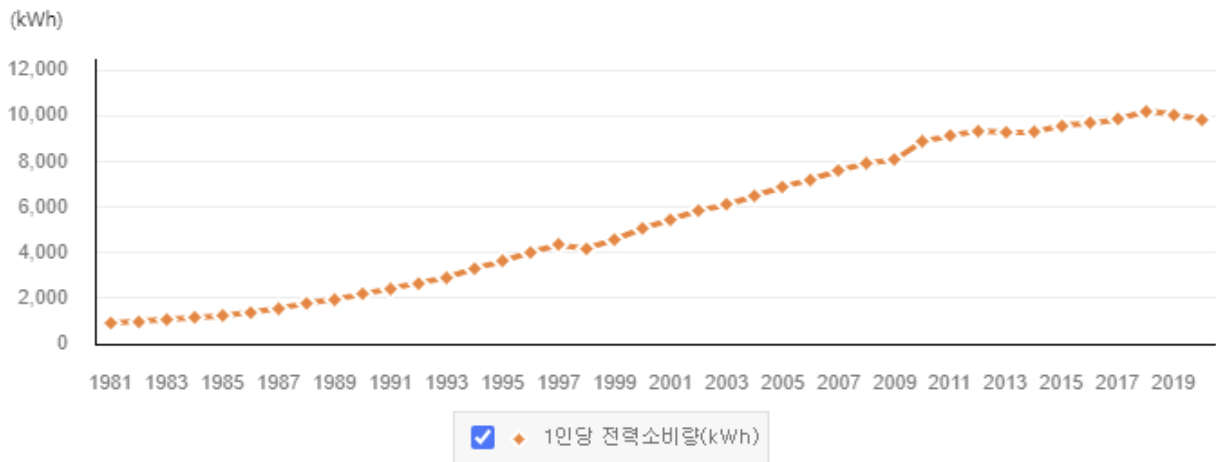
최근 안정적인 전력수급과 급증하는 전력수요를 예측하는 수요예측 기술에 대한 관심이 높아지고 있다. 또한 실시간 전력측정을 가능하게 하는 스마트 미터기의 보급의 증대로 인해 수요예측 기법에 대한 연구가 활발히 진행되고 있다. 본 연구에서는 실제 측정된 월별 전력 사용량 데이터를 학습하여 예측결과를 출력하는 시계열 예측모델 실험을 진행한다. 그리고 데이터 전처리 기법으로써 이동평균법을 도입하였다. 실제로 측정된 데이터를 학습한 모델의 예측량과 실제 전력 측정량을 비교한다. 이 예측량을 통해서 전력공급 예비율을 낮춰 사용되지 않고 낭비되는 예비전력을 줄일 수 있는 가능성을 제시한다.

**주제어:** 전력 생산량 예측, ARIMA

# I. 서론 (분석의 동기)

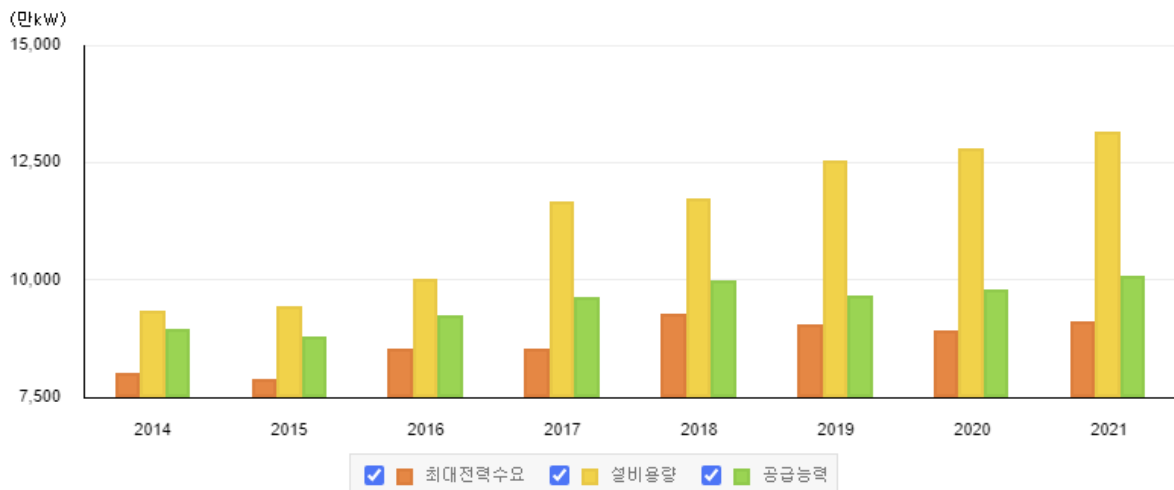
한국전력공사의 한국전력통계에 따르면, 인구 1 인당 전력소비량은 아래 그래프와 같이 매년 꾸준히 증가해 왔다. 전력은 1 차 에너지를 변형 또는 가공하여 우리 생활이나 산업 분야에서 다루기 쉽고 사용하기 편리한 에너지로 만든 것으로, 2 차 에너지에는 전기와 도시 가스, 석유 제품 등이 해당된다. 한국에서 전력은 경제성장의 원동력이자 보다 나은 삶을 가능하게 하는 수단이 되어 왔다. 1 인당 전력소비량은 1990 년 2,202kWh, 2000 년 5,067kWh 로 꾸준히 증가하다가 2020 년 기준 조금 감소하여 9,826kWh 다. 이는 시간의 흐름에 따라 시대가 바뀌면서 꾸준히 증가세를 보이던 수요도 상황에 따라 감소하기도 하는 등 수요 예측에 어려움이 있다는 것을 의미한다.

<표> 1 인당 전력소비량



아래 표는 한국전력거래소에서 발표한 2014 년부터 2021 년까지의 연도별 전력수급 실적 그래프이다. 그래프에서 최대전력수요는 일년 중 전력사용이 최대로 되는 순간의 전력수요이고, 발전설비용량은 전력수요를 충족시킬 수 있는 발전소 용량의 확보 정도를 의미한다. 전력은 저장이 불가능하고 매순간 수요와 공급이 일치되

<표> 연도별 전력수급 실적



어야 하므로 항상 최대 전력수요보다 어느 정도 여유를 가진 공급능력(발전설비용량)을 갖추어야 하며, 이를 통해 적절한 전력 생산량을 예측해 두어야 할 필요가 있다. 전력에 대한 수요는 5G 활성화, 전기차 수요, 이상기온, 4 차

산업혁명 등 여러 요인으로 인해 미래 예측이 매우 불확실하다. 그러나 발전설비는 단기간에 건설이 불가능한 설비이고 짧게는 5년, 길게는 20년 정도의 기간이 필요하므로 장기 전력 수요와 그에따른 생산량 예측을 통한 발전설비 건설계획이 매우 중요하다.

본 리포트는 미국의 1985년부터 2018년까지 미국의 전기산업의 월별 생산량 데이터를 시계열 분석해보고 그 결과를 통해 미래의 생산량을 예측하는 정량적 모델을 구축해 보고자 한다. 현재 우리나라의 산업통상자원부는 연도별 전력수급동향을 참고하여 장기 전력수급기본계획을 수립하고 있는데, 본 리포트와 같은 시계열적 접근이 많아진다면 우리나라의 향후 전력수급 안정에 이바지할 수 있을 것이다.

## II. 이론 및 선행문헌검토

### 2.1 이론

- Persistence Model

시계열 예측을 위한 좋은 Baseline 은 Persistence Model(지속성 모델)이다. 이것은 마지막 관측이 앞으로도 유지되는 예측 모델이다. 단순성때문에 종종 naive 예측이라 불린다.

- Autoregression Model(자기회귀모형)

자기회귀모형을 간단히 설명하면 마치 연도를 독립변인으로, 판매량을 종속변인인 것 처럼 간주했을 때 두 변인의 관계를 잘 설명할 수 있는 회귀식을 만든다. 이렇게 구한 회귀식을 통해서 미래의 어떤 연도에서 발생할 수 있는 값을 얻을 수 있다.

- Moving Average Model(이동평균모형)

차수  $m$ 의 이동 평균의 식은 아래와 같다.  $m = 2k+1$ 으로,  $k$ 기간 안의 시계열 값을 평균하여  $t$ 의 추세-주기를 측정하는 것이다.

$$\hat{T}_t = \frac{1}{m} \sum_{j=-k}^k y_{t+j},$$

측정 시기가 비슷하면 값도 비슷해진다. 따라서 평균이 데이터의 무작위성을 줄여주며 매끄러운 추세-주기 성분만 남긴다.

- ARIMA Model(AutoRegressive Integrated Moving Average(이동 평균을 누적한 자기회귀))

ARIMA 모델은 시계열을 예측할 때 가장 널리 사용하는 접근 방식이고, 주어진 문제를 상호 보완적으로 다루도록 하는 접근 방식이다. ARIMA 모델은 데이터에 나타나는 자기상관(autocorrelation)을 표현하는데 목적이 있다. 차분을 구하는 것을 자기회귀와 이동 평균 모델과 결합하면, 비-계절성(non-seasonal) ARIMA 모델을 얻을 수 있다.

ARIMA 는 AutoRegressive Integrated Moving Average(이동 평균을 누적한 자기회귀)의 약자인데 이 맥락에서 "누적(integration)"은 차분의 반대 의미를 갖는다.. 자기회귀(autoregression)와 이동 평균 모델에 사용되는 것과 같은 정상성(stationarity)과 가역성(Invertibility) 조건은 ARIMA 모델에도 적용된다. ARIMA 를 효과적으로 활용하기 위해서는 ARIMA 의 모수(parameter)를 잘 설정해야 한다. 이 파라미터가 올바른 예측식을 구하는데 핵심적인 요소가 된다.

p : 자기회귀 모형(AR)의 시차

q : 이동평균 모형(MA)의 시차

d : 차분누적(I) 횟수

p 와 q 는 일반적으로  $p + q < 2$ ,  $p * q = 0$  인 값을 사용한다. p 와 q 중 하나는 0 이라는 뜻인데, 시계열 데이터가 AR 이나 MA 중 하나의 경향만 가지기 때문이다.

- Mean Squared Errors(평균 제곱오차)

평균제곱오차(Mean Squared Error, MSE)는 이름에서 알 수 있듯이 오차(error)를 제곱한 값의 평균이다. 오차란 알고리즘이 예측한 값과 실제 정답과의 차이를 의미한다. 즉, 알고리즘이 정답을 잘 맞출수록 MSE 값은 작을 것이다. MSE 값은 작을수록 알고리즘의 성능이 좋다고 볼 수 있다.

## 2.2 선행문헌 검토 (문헌연구)

| 논문 제목  | 분석 방법                     | 내용  |
|--|---------------------------|---|
| ARIMA 모델을 이용한 주간 전력 수요 예측                          | ARIMA                     | 주간 전력 수요 특성을 고려하여 10 년의 평일 전력 수요를 ARIMA 모형으로 구성, 전력 수요에 대한 예측 및 결과 확인               |
| 계절 ARIMA 모델을 이용한 104 주 주간 최대 전력 수요 예측              | ARIMA, SARIMA             | 104 주간 최대 전력 수요 예측을 바탕으로 전력 수급 전망에 따른 발전기 예방정비 일정과 양수 발전기의 경제적인 운영계획 수립             |
| ARIMA 모델을 이용한 일일 건물 난방 전력 수요 예측                    | ARIMA                     | 86 일간의 데이터로 모형 추정 후 5 일간의 난방 전력 수요량 예측  |
| ARIMA 모델 기반 생활 기상 지수를 이용한 동·하계 최대 전력 수요 예측 알고리즘 개발 | ARIMA                     | 시계열 예측 모델과 생활 기상지수를 이용하여 동, 하계 최대 전력 수요를 예측하는 방법 제시                                 |
| ARIMA-Intervention 모델을 통한 최대전력 수요예측 연구             | ARIMA-Intervention, ARIMA | 코로나 19 에 대한 영향을 입력변수로 반영하여 전력수요예측 연구  |
| DNN 과 LSTM 활용한 일일 전력 수요 모델 개발 및 예측                 | DNN, LSTM                 | 공급중심전력정책과 수요관리정책이 전력수요에 미치는 영향, 경제성장 또는 하락이 전력수요에 미치는 영향을 DNN 과 LSTM 모델을 사용하여 결과 예측 |

## Ⅲ. 연구설계(자료 및 방법론 설명)

### 3.1 연구 질문

현대 사회에서 전기는 산업 및 일상에서 없어서는 안 될 중요한 필수 요소이다. 이런 전기의 원활한 공급을 위해 전력 생산 및 공급 시스템이 갖춰지고 전력 수요량을 예측하여 에너지 계획을 제대로 세워야 한다. 계획에 따라 정책을 수립하거나 생산 및 공급에 대한 결정을 해야 예산낭비를 하지 않을 수 있다. 따라서, 본 연구에서는 전력 수요가 생산과 비례하여 정확한 정책이 이루어졌고 생산량이 최적화되어 생산되었다는 가정 하에 전력 생산량 데이터 기반으로 전력 생산량에 대하여 분석하고 예측해보고자 한다.

### 3.2 연구 방법

데이터 확인을 통해 Original 데이터 및 평균과 표준편차를 확인하였다. Augmented Dicekey-Fuller Test 를 통해 시계열 데이터가 안정성이 있는지 확인하였으며, Decomposition 단계를 통해 trend 를 분리하였다. ACF & PACF 를 통해 관측 값과의 관계 강도를 확인하였다. 또한 Persistence Model, Autoregression Model, Moving Average Model, ARIMA Model 을 통해 10 일 간의 전력 생산량을 예측하고 결과를 MSE 로 판단하였다.

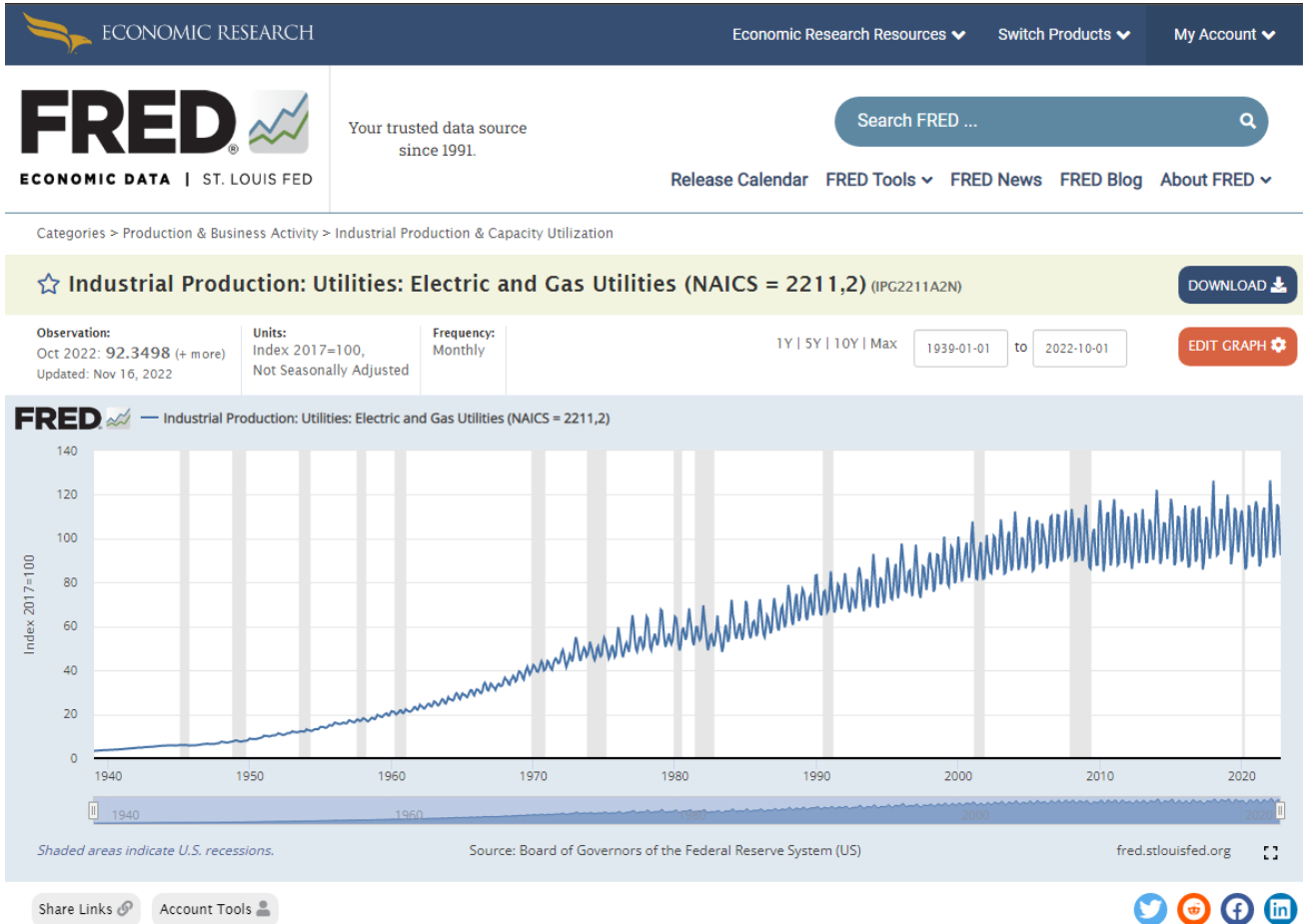


### 3.3 자료의 수집

- 데이터 출처 : FRED(연방 준비 은행 경제 데이터)

<https://fred.stlouisfed.org/series/IPG2211A2N>

- 데이터 설명 : 1985 년부터 2018 년까지 미국의 전기 산업의 월별 생산량

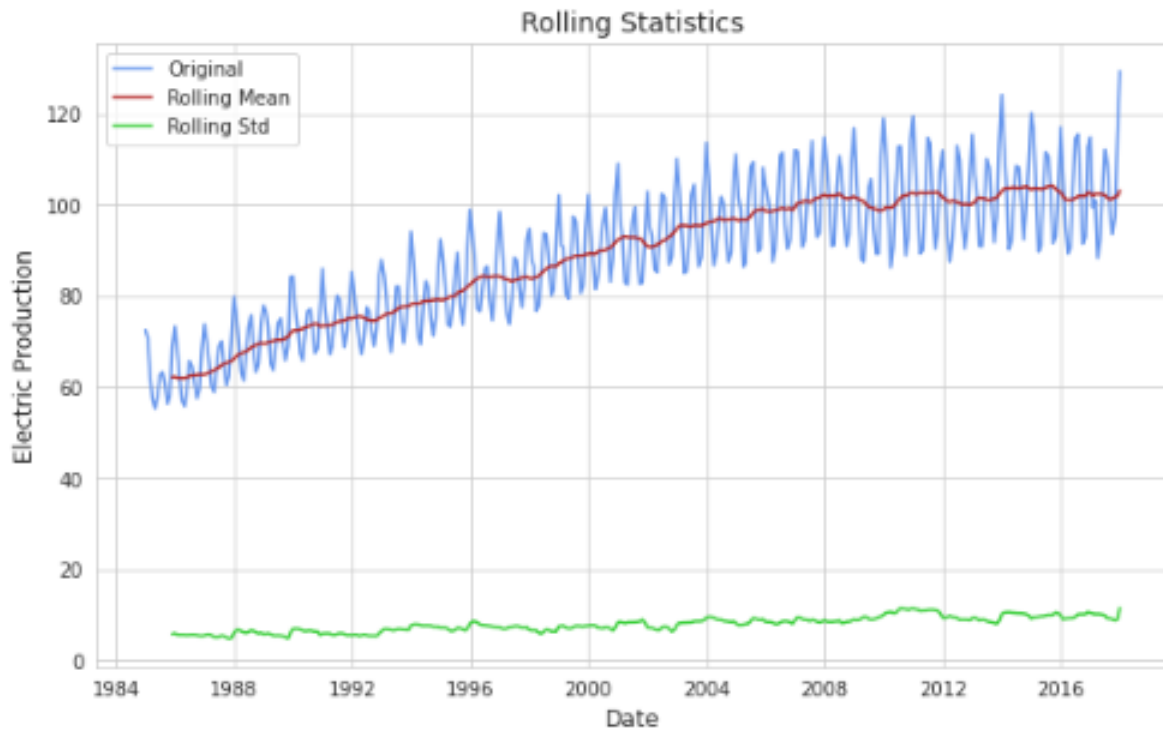


### 3.4 데이터 선택, 전처리, 병합

- train 데이터는 결측치가 없는 등 충분히 정제되어있는 데이터였다. 아무래도 시계열 데이터다보니 결측치를 보간하는 작업이 필요할 것이라고 예상했는데, 그렇지 않아서 전처리에 들이는 시간을 절약할 수 있었다.
- 분석을 용이하게 하기 위해 컬럼명을 영어로 변경해주는 작업을 진행하였다.
- date\_time 데이터의 type 을 datetime 으로 변경하고, 데이터의 날짜에 해당하는 월, 일, 요일, 주말 컬럼을 생성하였다.

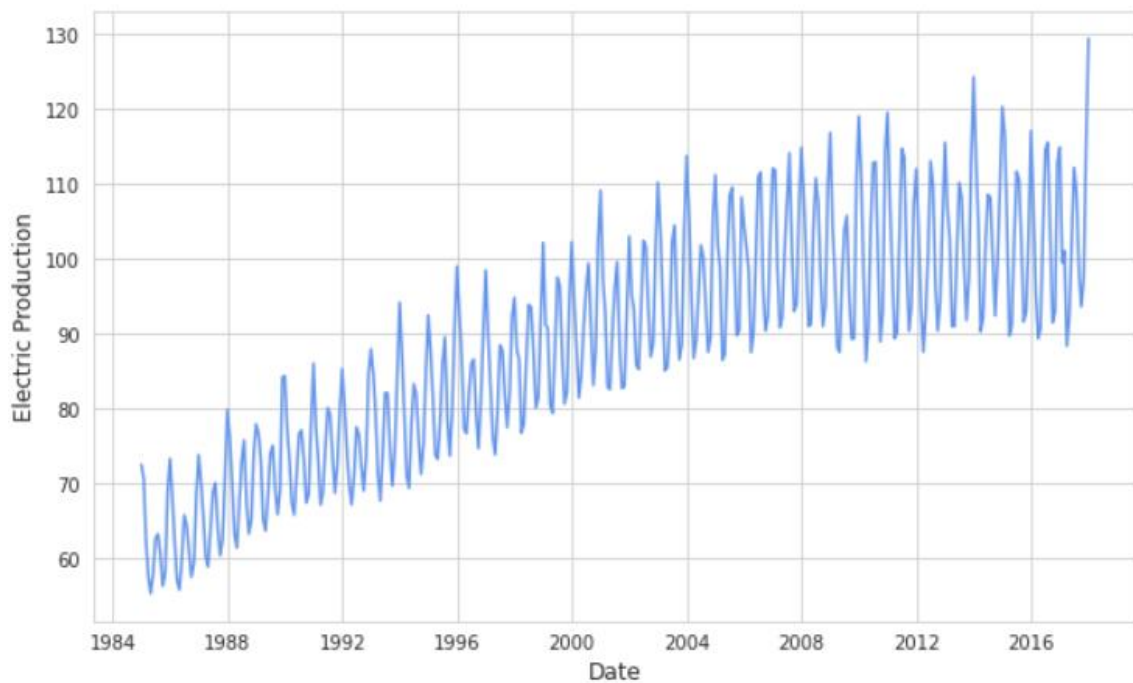
### 3.5 데이터 분석(분석결과 정리보고)

#### 1) 데이터 확인

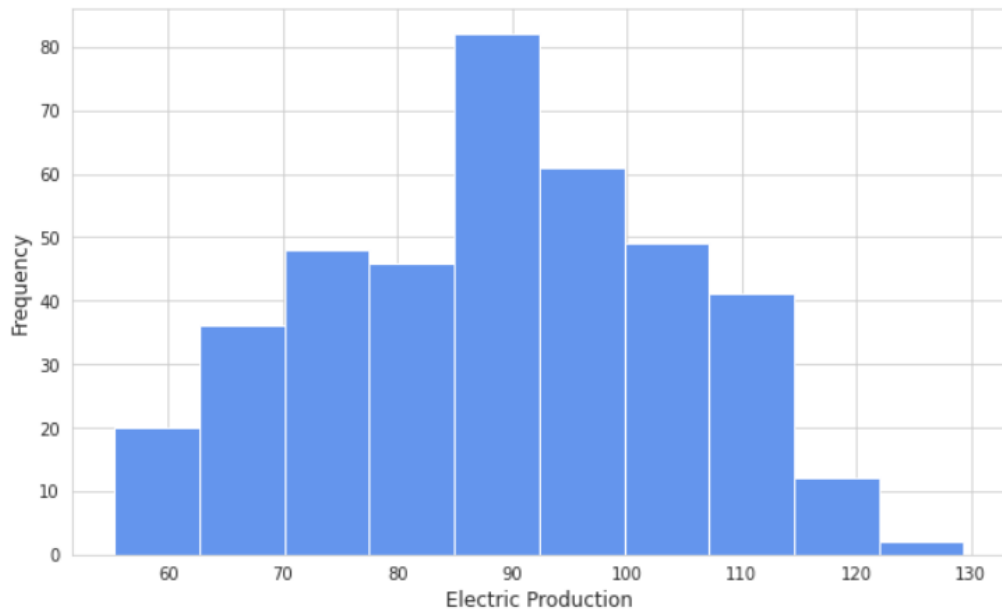


분석에 앞서 데이터를 시각해 보았다. 해당 데이터는 1985년 1월부터 2018년 1월까지의 전력생산량이다. Original(파란 선)은 모든 데이터를 표시한 것이며 Rolling Mean(빨간 선)은 평균 값, Rolling Std (초록색 선)은 표준편차 값이다. 데이터와 평균을 확인해 보았을 때 시간의 흐름에 따라 전기 생산량은 우상향하고 있는 것이 확인된다.

## 2) Stationary 확인

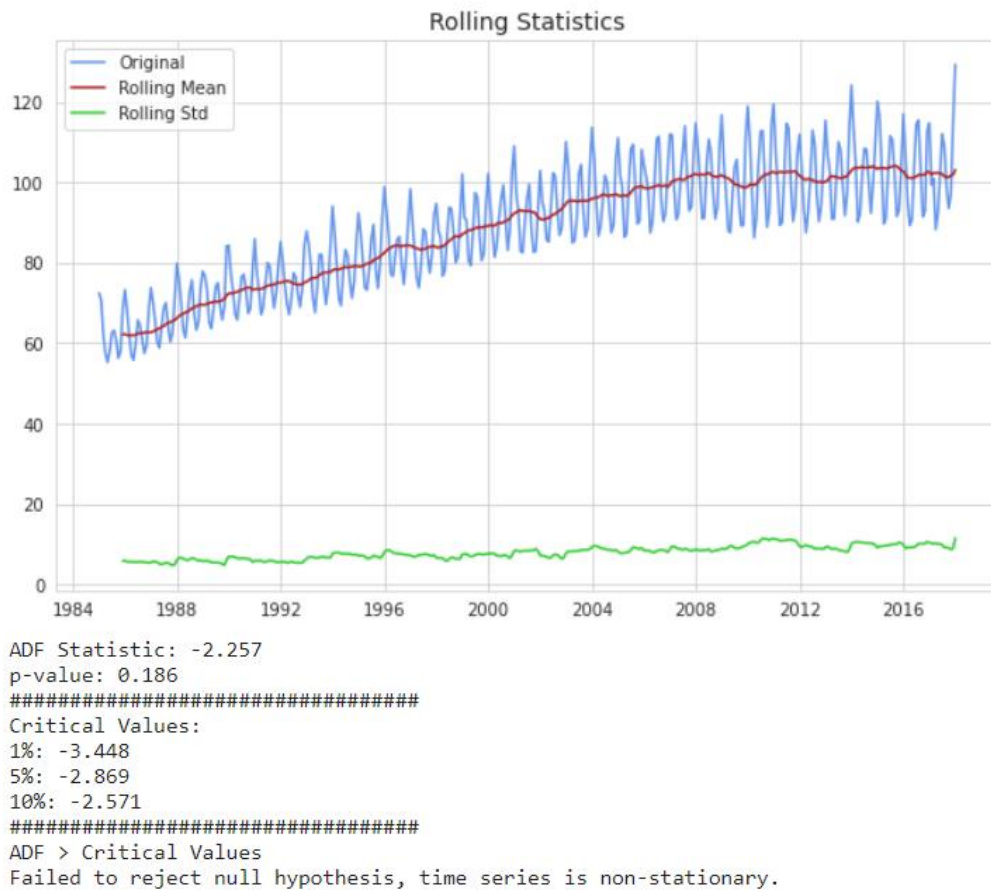


데이터들만을 가지고 plot 을 확인해봤을 때 전력 생산량은 계속해서 상승 추세가 보이는 것으로 나타난다.



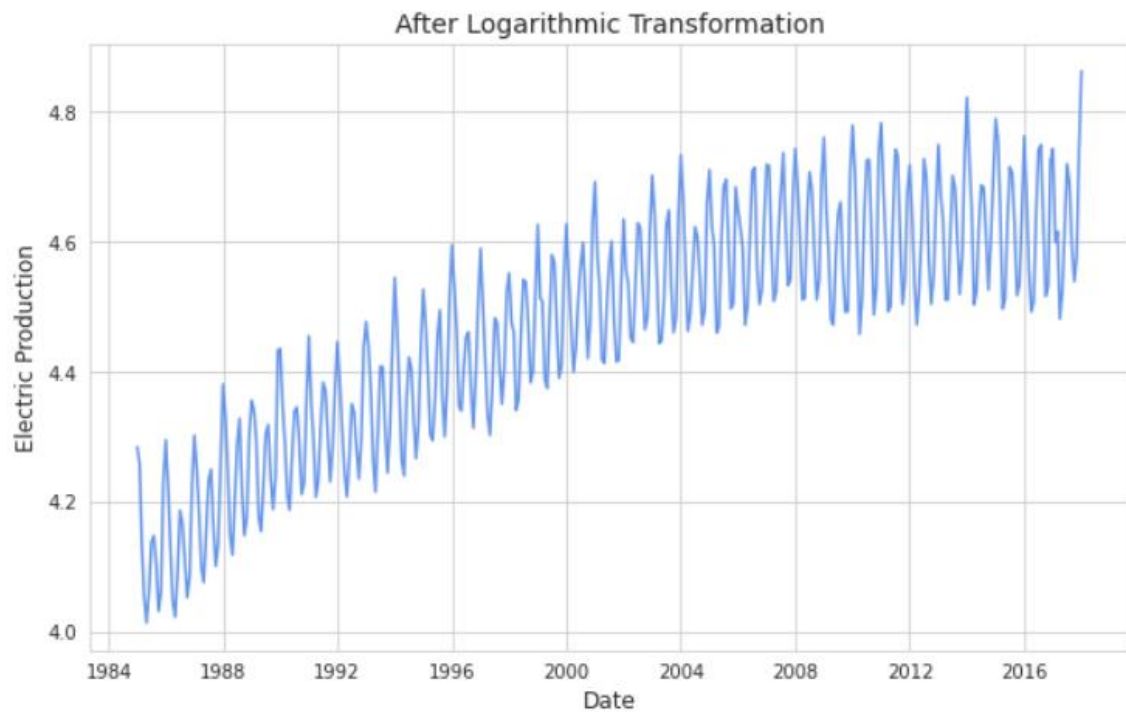
전체 데이터의 분포도는 위와 같이 나왔으며, 정규성을 따르는 것으로 보인다.

### 3) Augmented Dickey-Fuller Test



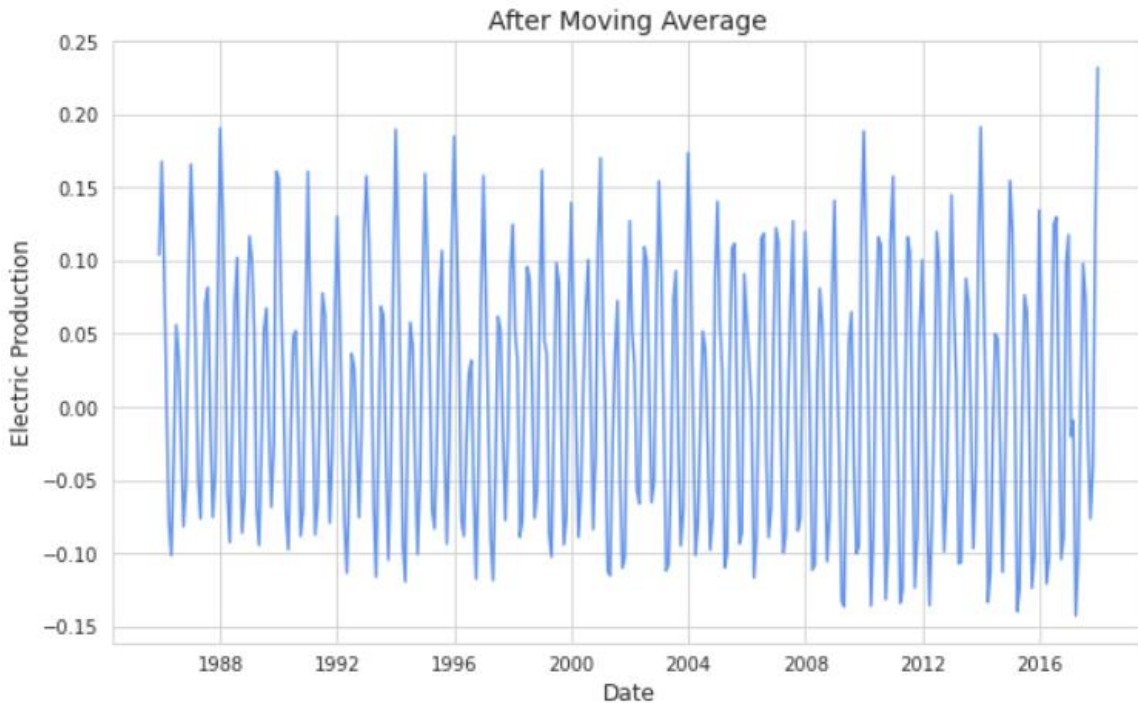
이 테스트에서 귀무가설은 시계열이 비정상적이고, 시간에 따른 추세를 제공한다는 것이고, 대립가설은 시계열은 고정되어 있다는 것이다. 만약 ADF 값이 t 통계량보다 작다면 귀무가설을 기각할 수 있고, 시계열이 고정되어 있다는 것을 의미한다. 테스트 결과, p-value 는 유의수준 0.05 보다 크고 ADF 는 critical values(임계값)보다 크다. 따라서, 시계열은 비정상적이고 시간에 따른 추세를 제공한다는 귀무가설을 기각하지 못하며, 시계열은 non-stationary 하다. (\* Augmented Dickey-Fuller Test 는 ADF Test 라고도 하며 시계열 데이터의 안정성을 테스트하는 통계적 방법이다.)

#### 4) Logarithmic Transformation with Box-Cox



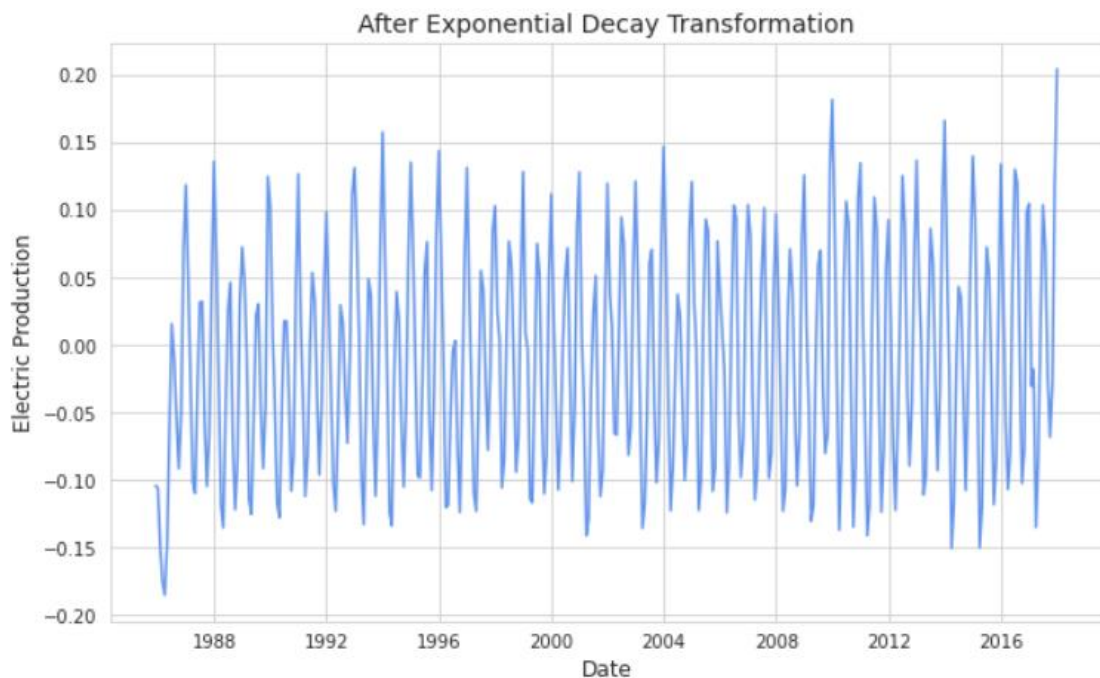
Box-Cox 변환은 데이터를 정규분포에 가깝게 만들거나 데이터의 분산을 안정화하는 것이다. 표본평균과 종속변수값 중 독립변수에 의해 설명된 부분과의 차이를 제공하여 합한 값인 SSE 오차를 최소화하기 위하여 Box-cox 변환을 한 결과이다.

## 5) Removing Trend with Moving Average

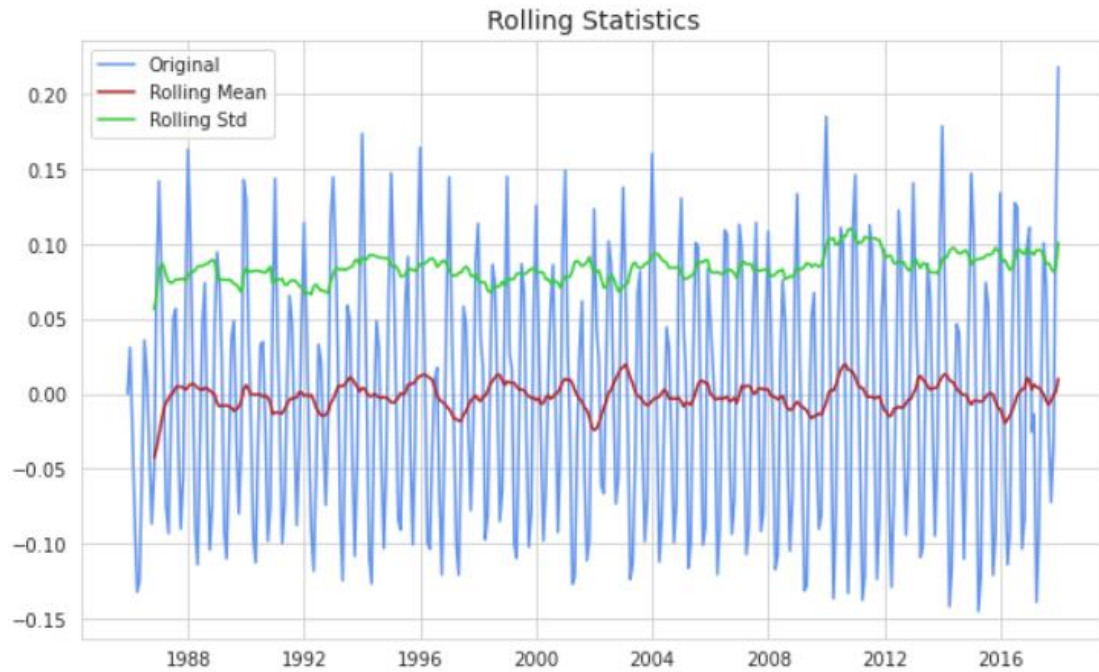


앞서 augmented dickey-fuller test 를 통해 trend 가 존재하는 non-stationary 한 시계열임을 확인하였다. trend 가 있는 시계열을 비정상 시계열이라고 한다. 시계열 모형을 만들기 위해 데이터의 확률분포를 수치화해야하는데 데이터의 추세(trend)가 있으면 시간에 따라 값이나 확률분포의 의미가 달라진다. 즉 어느 구간에서도 평균과 분산이 거의 같아야한다. 따라서 이동평균을 통해 trend 를 제거하였다. 제거 후 data set 은 stationary 하다고 말할 수 있다.

## 6) Exponential Decay Transformation



exponential decay transformation 을 통해 Seasonality 효과를 다소 상쇄한 결과이다.



```

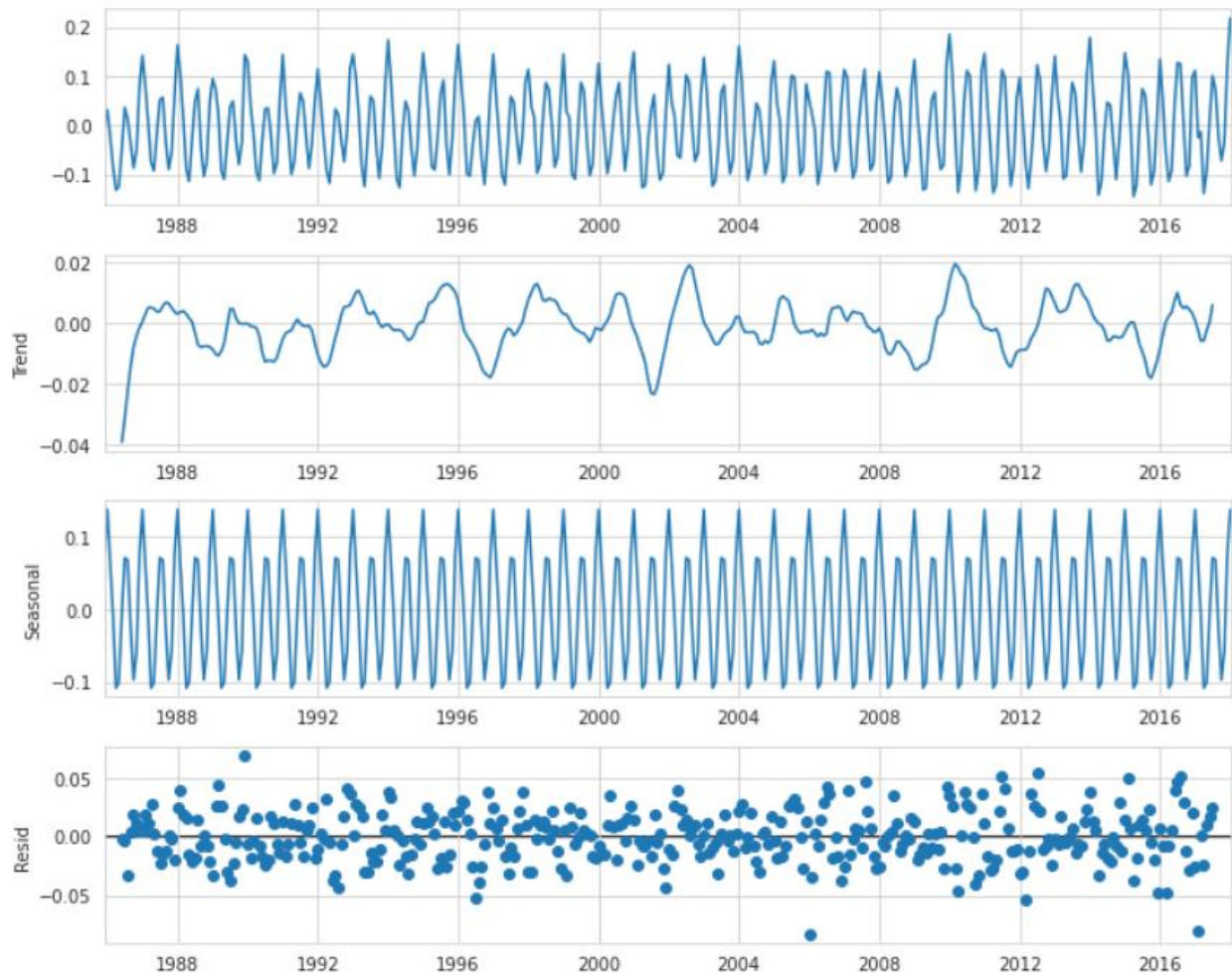
ADF Statistic: -7.213
p-value: 0.0
#####
Critical Values:
1%: -3.448
5%: -2.869
10%: -2.571
#####
ADF < Critical Values
Reject null hypothesis, time series is stationary.

```

p-value 가 0.05 보다 작은 수치임에 따라, 귀무가설을 기각하고 위의 데이터가 stationary 하다고 할 수 있다.



## 7) Decomposition



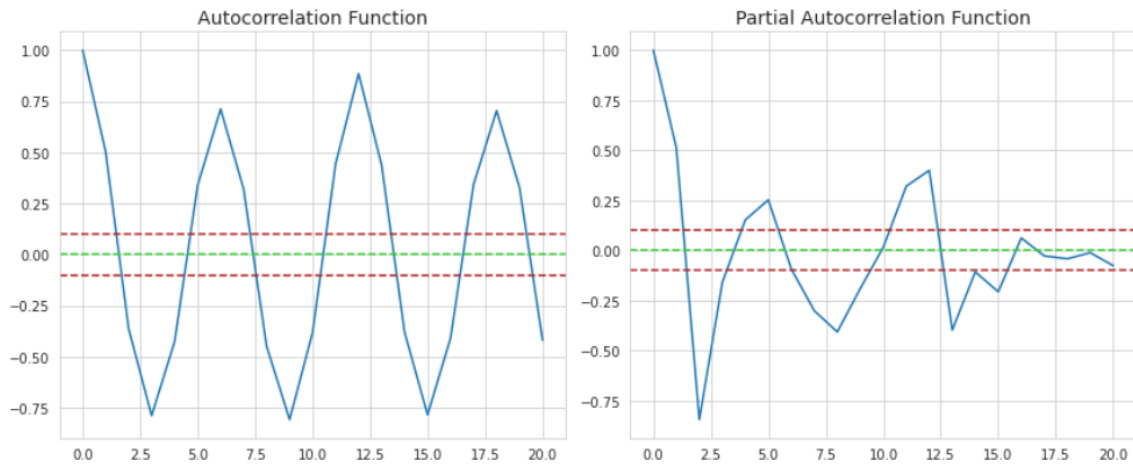
seasonal\_decompose 를 이용해서 추세, 계절성, 잔차를 파악할 수 있었다. 각각의 의미는 다음과 같다.

- 트렌드(Trend) : 연도별 크기의 윈도우를 움직여가며, 관측치들의 평균을 측정한 것
- 계절성(Seasonal) : 각 월별로 데이터를 모아 평균을 측정한 것. 각 주기의 계절성은 trend 에 대한 상대적인 증가 혹은 감소를 나타내므로 합이 0 이 되도록 조정된다.
- 잔차(residual) : 원 데이터에서 앞서 도출한 trend 와 계절성의 값을 제거한 나머지 값이다.

본 데이터는 1988 년부터 뚜렷한 상승추세는 없으나 지속적으로 미세하게 상승하는 추세를 보인다. 또한 trend 를 제거한 이후 계절성이 뚜렷하게 보이는 것을 확인할 수 있다. 최근으로 올 수록 residual 이 커지는 경향을 보인다. 즉, 전력생산량의 절대적인 양이 늘어남에 따라 수치적으로 변동성도 커졌다.



## 8) ACF(AutoCorrelation Function) & PACF(Partial AutoCorrelation Function)



ACF(자기상관함수, AutoCorrelation Function)란 시차에 따른 일련의 자기상관을 의미한다. 우리가 흔히 알고있는 Correlation 은 두 변수 간의 관계를 -1~1 사이로 정규화한 값으로 표현하는 척도인데, Autocorrelation 은 time shifted 된 자기 자신의 데이터와의 상관성을 의미한다. 시차가 커질 수록 ACF 는 0 에 가까워진다. 이는 정상성을 판단하는데 유용하다. 정상 시계열은 상대적으로 빠르게 0 에 수렴하며, 비정상 시계열은 천천히 감소하고, 종종 큰 양의 값을 가진다. 또한 PACF(편자기상관함수, Partial AutoCorrelation Function)는 시차에 따른 일련의 편자기상관이며, 시차가 다른 두 시계열 데이터 간의 순수한 상호 연관성을 나타낸다.

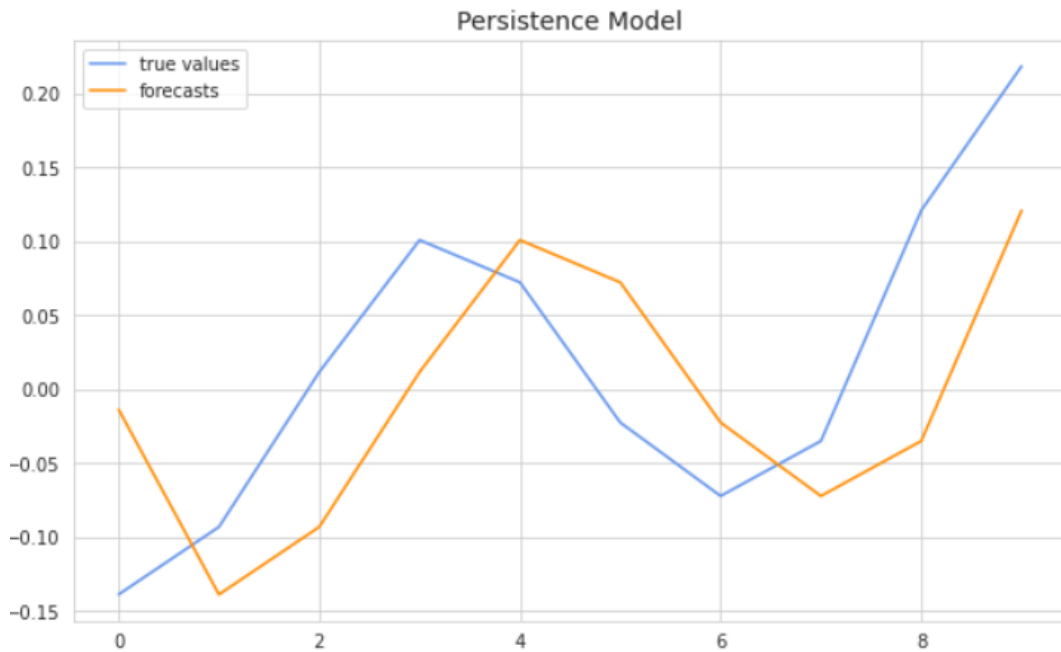
위 그래프는 autocorrelation 과 partial autocorrelation 을 시행한 결과이며, 이전 관측 값과의 관계 강도를 요약한 것이다. 이 ACF 와 PACF 는 ARIMA 모델의 파라미터가 되는  $p$ ,  $d$ ,  $q$  의 최적 차수를 탐색할 때 유용하게 사용된다.  $p$  는 AR,  $d$  는 차분 횟수,  $q$  는 MA 와 관련이 있는 파라미터이다.

## IV. 분석내용

### 4.1 예측 모델

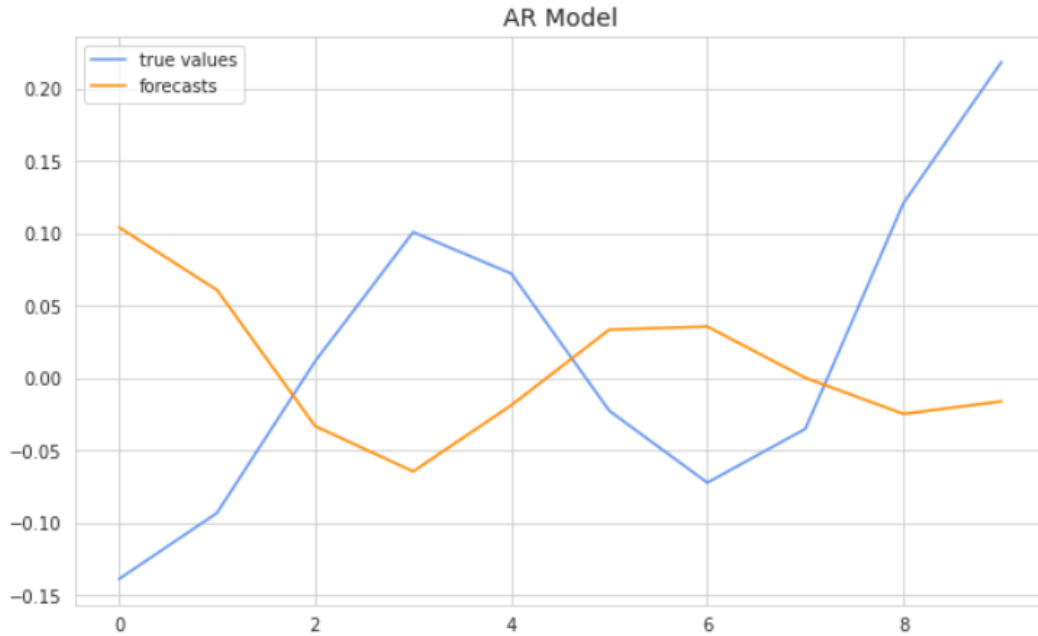
Persistence Model, Autoregression Model, Moving Average Model, ARIMA Model 들을 통해 데이터 예측을 진행해 보았다.

#### 1) Persistence Model



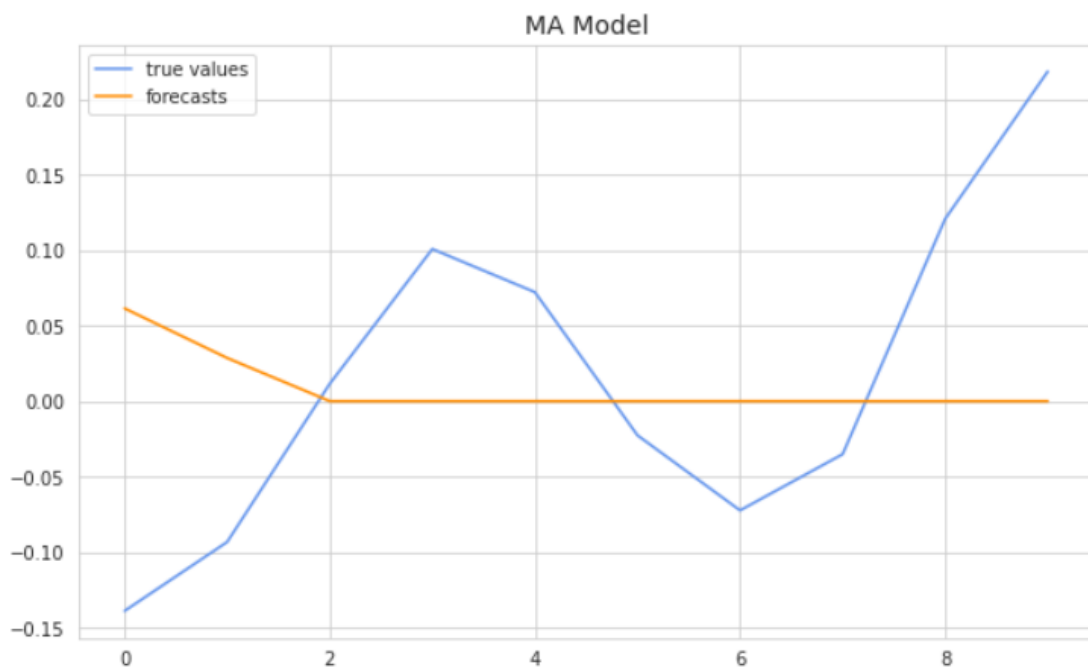
지난 10 일 간의 전기 생산을 예측하는 모델을 생성한다고 가정하고 모델을 생성하였다. 예측에 사용할 수 있는 가장 간단한 모델은 마지막 관측을 유지하는 것이다. 이를 Persistence Model 이라고 하며, Autoregression Model(자동 회귀 모델)과 비교하는 데 사용할 수 있는 성능 기준을 제공한다. 이 모델의 MSE 는 0.0084 로 측정되었다.

## 2) Autoregression Model(AR Model, 자동 회귀 모델)



Autoregression Model(AR 모델, 자동 회귀 모델)은 지연된 변수를 입력 변수로 사용하는 선형 회귀 모델이다. 과거 시점의 자기 자신의 데이터가 현 시점의 자기 자신에게 영향을 미치는 모델이라는 뜻이다. scikit-learn 의 Linear Regression 클래스를 사용하여 모델링하였다. 그래프를 확인해 보면 예측 값이 실제 값과 많은 차이를 보이는 것으로 나타났다. MSE 는 0.0212 로 Persistence Model 모델에 비해 낮은 성능을 보였다.

## 3) Moving Average Model(이동평균모형)



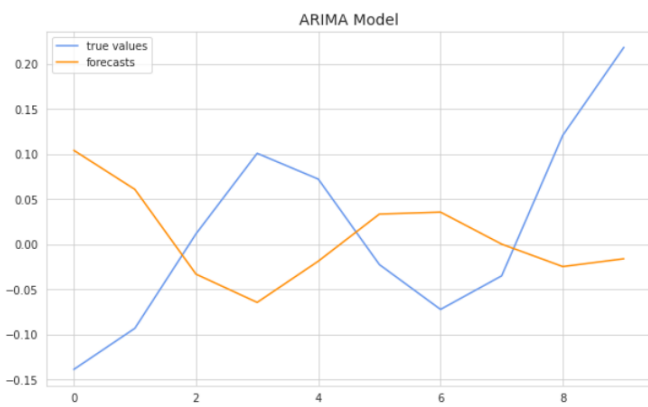
AR 모형이 과거의 값을 활용하여 미래를 예측하는데 반해, MA 모형은 과거의 예측 오차를 활용하여 미래를

예측하는데 활용한다. 윈도우라는 개념을 사용하여 시계열을 따라 윈도우 크기만큼 슬라이딩(moving)된다고 하여 이동 평균 모델이라고 한다. MSE 측정결과 0.0139 로 AR-model 에 비해 높지만 여전히 낮은 성능을 보이고 있다.

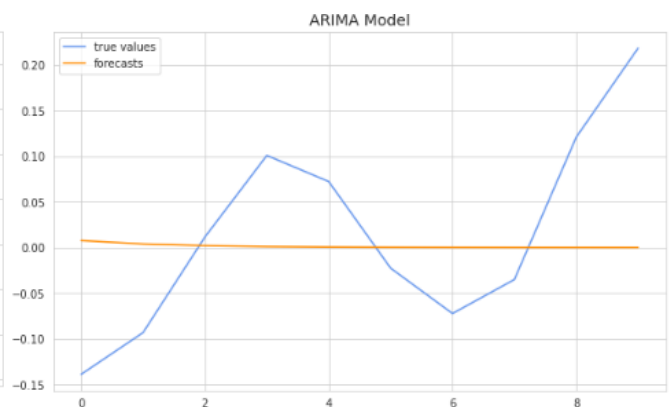
#### 4) ARIMA(AutoRegressive Integrated Moving Average, 자기 회귀 누적 이동 평균 모델)

ARIMA 모델은 자기 회귀와 이동 평균을 둘 다 고려하는 모형이다. AR, MA model 과 달리 과거 데이터의 선형 관계뿐만 아니라 추세까지 고려한 모델이다. 시계열 분석 기능을 제공하는 파이썬 패키지인 statsmodels 라이브러리를 사용해 모델링하였다. statsmodels 라이브러리의 ARIMA(p,d,q) 함수에서 모델링을 위해 파라미터를 수동으로 입력해 주어야 한다. 각 파라미터는 다음과 같다.

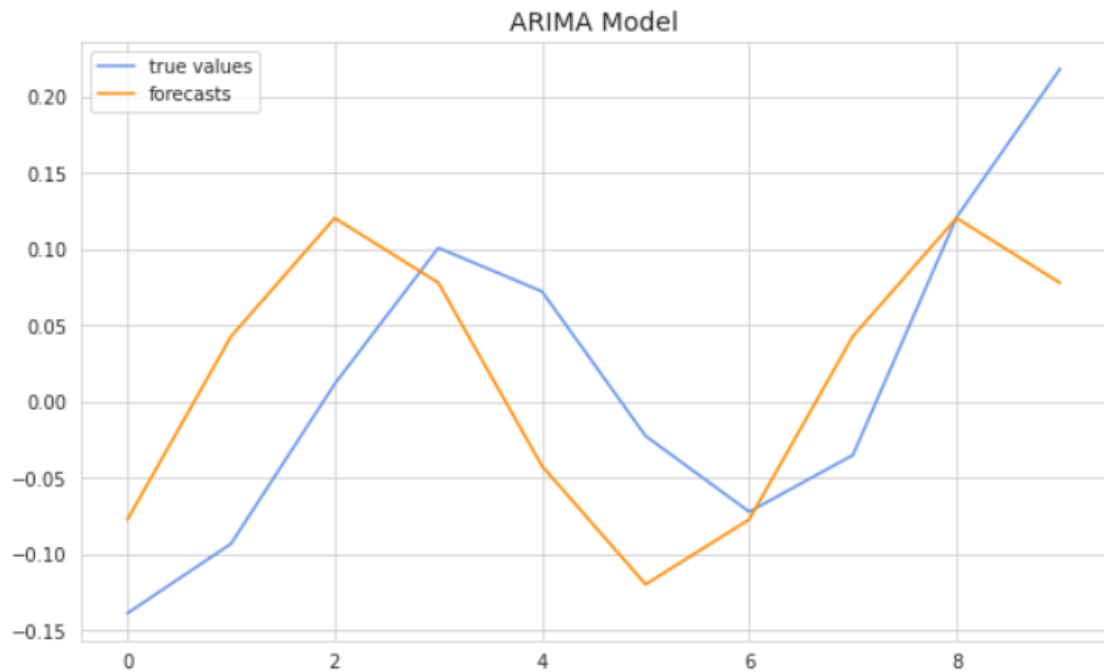
- p: 자기 회귀 차수
- d: 차분 차수
- q: 이동 평균 차수



<graph1> p = 2, d = 1, q = 0



<graph2> p = 1, d = 1, q = 1



<graph3>  $p = 2, d = 1, q = 2$

파라미터  $p$ 의 경우, 앞서 확인해 본 PACF 결과에서 2 이후로 값이 0에 가까워지고 있어 2로 설정하였다. 파라미터  $d$ 는 1차 차분 이후  $p$ -value를 만족하였기 때문에 1로 정하였으며 파라미터  $q$ 는 0보다 2일 때를 모두 시도해본 결과, 2일 때 실제 값과 더욱 비슷한 경향을 나타내고 있어 최종적으로  $p=2, d=1, q=2$ 의 값으로 파라미터를 설정하였다. ARIMA 모델을 통해 미래의 추세, 동향에 대해 예측하였다. 실제 데이터(파란색 선)와 모형 실행 결과(주황색 선)를 표시한 <graph3>을 보면 데이터가 우상향 추세를 나타내고 있으므로, 전력생산량이 향후에도 계속 증가할 것임을 예측할 수 있다.

## 4.2 모델 비교

위에서 다룬 Persistence Model, Autoregression Model, Moving Average Model, ARIMA Model에 대해 Mean Squared Errors를 통해 예측 성능평가를 진행하였다.

| MSE            |          |
|----------------|----------|
| Model          |          |
| ARIMA          | 0.008290 |
| Persistence    | 0.008401 |
| Moving Average | 0.013946 |
| Autoregression | 0.021221 |

결과는 위 표와 같이 ARIMA 모델의 MSE가 0.008290으로 가장 낮게 나와 성능이 가장 우수한 모델로 나타났다. 단순히 하루 전 데이터를 가져오는 Persistence Model과 예측 성능 결과 1위인 ARIMA 모델 간의

차이는 적은 편이고, Moving Average Model 과 Autoregression Model 의 경우 MSE 가 상위 2 개 모델보다 월등히 커, 성능이 낮은 모델인 점이 확인되었다.

## V. 결론

### 5.1 분석결과

본 리포트는 전력생산에서의 시계열 예측을 위해 지난 1985 년부터 2018 년까지 미국의 전기생산량의 월별 데이터를 살펴보았다. 데이터를 정상성을 갖춘 데이터로 변환후 Persistence Model, Autoregression Model, Moving Average Model, ARIMA Model 을 통해 10 일 간의 전력 생산량을 예측해보고 그 결과는 MSE 로 판단하였다. 결과는 ARIMA Model 이 성능이 가장 우수한 모델로 나타났다. Moving Average Model 과 Autoregression Model 은 본 데이터에 대해 Persistence Model, ARIMA Model 에 비해 성능이 낮은 것으로 확인되었다. 전력생산량이 향후에도 계속 증가할 것임을 예측할 수 있었다. 향후 우리나라 데이터 수집 및 추가 연구를 통해 단기, 장기적인 예측을 해 나간다면 전력 관련 정책 수립 및 국가 재난 상황 등 갑작스러운 전기수요를 방어할 수 있는 대비책을 마련할 수 있을 것이다.

### 5.2 시사점 및 활용

본 리포트는 시계열분석을 통해 전력 생산량의 추세를 확인하고 전력 생산량을 예측하여 살펴보았다. 현재 전력에 대한 수요는 5G 활성화, 전기차 수요, 이상기온, 4 차 산업혁명 등 여러 요인으로 인해 미래 예측이 매우 불확실하다. 그러나 발전설비는 단기간에 건설이 불가능한 설비이고 짧게는 5 년, 길게는 20 년 정도의 기간이 필요하므로 장기 전력 수요와 그에 따른 생산량 예측을 통한 발전설비 건설계획이 매우 중요하다. 본 리포트와 같은 시계열적 접근이 많아진다면 우리나라 정책 결정자들이 적절한 전력 공급 계획을 세우는 데 합리적인 기초 자료로서 사용될 수 있을 것이다. 예측결과가 우상향 추세를 나타나는 것으로 나타나 향후 전력생산량도 지속적으로 높을 것으로 예상된다. 본 예측결과를 활용해 장기적으로는 우리나라의 향후 전력수급 안정에 이바지할 수 있을 것이다.

### 5.3 한계 및 향후과제

본 리포트의 데이터는 FRED(연방 준비 은행 경제 데이터)의 데이터이며 실제 전력 생산량만을 가지고 분석한 것이다. 전력 생산량은 정책 및 전염병 등 다양한 요인에 의하여 영향을 받을 수 있다. 추가적인 입력변수를 고려하지 않고 분석하였기에 정확한 분석 또는 예측에 한계가 있을 수 있다. 참고 문헌 'ARIMA-Intervention 모형을 통한 최대전력 수요예측 연구'에서도 코로나 19 라는 입력변수를 넣은 ARIMA-Intervention 모델이 ARIMA 모델보다 우수한 결과를 제시할 수 있었다고 한다. 따라서, 주요 정책이나 전염병과 같은 여러 입력 변수를 넣어 분석한다면

더욱 성능이 우수한 모델을 제시할 수 있을 것이다.

## 5.4 느낀점

시계열 데이터 분석을 진행하면서 우리가 분석하고자 하는 데이터가 stationary 한 데이터인지 확인하고 검증하는 과정을 주요하게 학습할 수 있었다. 이 과정이 중요한 이유는 우리의 데이터(표본)가 일정한 평균과 분산을 가진 확률 과정(stochastic process)을 따른다고 가정해야만 시계열 예측을 할 수 있기 때문이다. 만약 시간의 흐름에 따라서 이 확률 분포가 크게 변동한다면, 그 실현 값들의 평균이나 분산 등의 의미가 없어져버린다. 그래서 적어도 이 평균과 분산이 우리가 다루고자 하는 확률 과정을 설명하기에 문제가 없도록 하기 위해 정상성(Stationarity)이라는 조건이 필요한 것이다. 본 리포트를 통해 정상성 조건의 당위성을 인지하게 되었고, 앞으로 해나갈 모든 시계열 분석에의 기초가 되는 내용으로 새기게 되었다. 또한 전력생산이라는 평소에는 관심이 덜하던 주제에 대해 다뤄보면서 정책분야의 의사결정권자의 시선을 체험해보는 소중한 기회가 되었다. 정책 결정에도 시계열 분석이 유의미한 insight 를 충분히 얻을 수 있는 도구로 작용하는 것을 알게되었다. 향후 다양한 분야의 데이터 활용에 대한 시야를 넓힐 수 있는 계기가 된 것 같다.

## 참고문헌

정현우, 김시연, 송경빈 (2013) ARIMA 모델을 이용한 주간 전력수요예측

김시연 (2014) 계절 ARIMA 모델을 이용한 104 주 주간 최대 전력수요예측

유정아, 노병준, 장성주 (2017) ARIMA 모델을 이용한 일일 건물 난방 전력 수요 예측

정현철, 정재성, 강병오 (2018) ARIMA 모델 기반 생활 기상지수를 이용한 동,하계 최대 전력 수요 예측 알고리즘 개발

손흥구 (2020) ARIMA-Intervention 모델을 통한 최대전력 수요예측 연구

김영수, 박호정 (2021) DNN 과 LSTM 활용한 일일 전력수요모델 개발 및 예측