<data1>

2011년 10월에 Frank는 지붕에 태양 전지판을 설치했고, 이 모듈의 총 전력은 5kWp입니다.  
이 분은 전기 사용량을 매일 기록하는 습관이 있어, 태양 전지판의 전력 생산량도 같이 기록하셨고, 이로 인해 이 데이터가 수집되었습니다.

‘PV Elec Gas2.csv’는 날짜, 누적 태양광 에너지, 사용된 kWh 전기, 사용된 가스 m²  
총 4가지 열을 가지고 있습니다.

<https://www.kaggle.com/datasets/fvcoppen/solarpanelspower>

- 태양 전지판 데이터 출처

data - PV\_Elec\_Gas2.csv (owner – frank)

<data2>

Frank씨가 측정한 태양광 패널이 있는 벨기에 앤트워프의 날씨 데이터입니다.

이 dataset은 2012년부터 2019년까지의 날씨 상태, 풍속, 습도, 기압, 온도의 열을 가지고 있고, Frank씨의 데이터와 같은 기간입니다.

<https://www.kaggle.com/datasets/ramima/weather-dataset-in-antwerp-belgium>

- 해당 지역 날씨 데이터 출처

data - weather\_in\_Antwerp.txt (원본 - 무료 날씨 사이트를 제공하는 Timeanddate,

전처리 - **RAMI MASHKOUK**)

<병합한 data>

두 데이터를 병합 - 병합코드 solar\_energy.ipynb

날씨와 태양광 에너지를 모두 수집하는 데이터 세트가 없다는 것을 알고 된 후 날씨 상태에 따라 태양광 생산을 분석하고 예측할 수 있는 데이터를 제공하기 위해 **RAMI MASHKOUK**는 이 병합코드를 만들게 되었습니다.

(Frank의 누적 태양광 data와 태양광 패널이 있는 벨기에 앤트워프의 날씨 data를 병합)

Frank씨의 데이터에서 누적 태양광 에너지를 일일 태양광 에너지량으로 변경하고 분단위의 날씨 데이터를 Frank씨의 데이터와 형식을 맞추기 위해 평균을 이용하여 단위를 일(day)로 변경하였습니다. 그 후 두 데이터를 병합했습니다.

<data 상태>

date: 시기

temp: 기온

cloud cover: 구름(안개)의 양(string형태로 설명)

wind: 풍속

humidity: 습도

barometer: 대기압

day\_power: 일일 태양광 발전량

<사용한 data>

2016, 2017, 2018, 2019년 data만 추출 - 코드 1619\_analysis.ipnb

병합한 dataset은 2012~2019년의 데이터이고, 2016, 2017, 2018, 2019년(11/19일까지)의 4개년 데이터를 사용하기로 결정했습니다.

최종 사용 데이터 - solar\_energy\_detail.csv

\* cloud cover형태

테이블이(가) 표시된 사진

자동 생성된 설명

‘.’을 기준으로 비/눈의 양과, 구름의 상태를 나타냅니다

태양광은 비, 눈 양보다는 구름의 상태가 중요하다고 생각하여 구름(안개)의 양을 나타내는 문자열만 추출하여 사용했습니다.

\*2018, 2019년 결측치

텍스트이(가) 표시된 사진

자동 생성된 설명

보간법 중 가장 많이 사용하는 3차 스플라인 방법을 이용하여 결측치를 채웠습니다.

\*cloud cover 점수화

'Clear': 1, 'Sunny' : 2, 'Partly sunny' : 3, 'Passing clouds' : 4, 'Scattered clouds':5, 'Broken clouds':6, 'Fog':7, 'Ice fog':7,'Haze':7, 'Overcast':8

점수가 클수록 구름/안개(안개일 때가 구름보다 비교적 습도의 양이 높기 때문에 안개에 점수를 더 크게 부여)의 양이 많이지도록 점수를 mapping했고, 1~8의 척도를 이용했습니다.

\*정규화

각 column의 범위를 일정한 범위로 조정하기 위해 MinMaxScaler를 이용하여 0~1의 범위로 변환시켰습니다.

<process model 분석>

인과관계가 성립하는 여러 모델을 돌렸을 때 유의한 결과가 있는 모델 11로 선택하였습니다.

(인과관계가 있다고 생각한 변수들의 조합으로 process를 돌렸을 때 다중조절효과가 유의하지 않은 것들이 대부분이었음)

1)

교수님께서 말씀하신 X: cloud cover, Y:태양광 발전량, M:습도, W: 기온, Z:풍속으로 모델링을 했을 때, 효과는 유의미 했지만, 계수들의 부호가 실제 인과관계와 조금 다르게 나와 부적합하다고 생각하여 다르게 모델링을 하였습니다.

2)

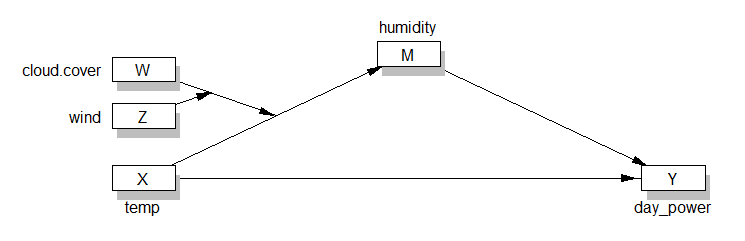
X:기온

M:습도

Y:태양광 발전량

W:구름/안개의 양

Z:풍속



<각 변수들의 관계>

X(기온)->M(습도) 관계 (조절된 조절효과)

기온증가 습도감소, 구름양이 증가하면 이러한 감소폭이 완화(완충효과)

풍속이 증가하면 구름양의 완충효과 약화 (실제 풍속 증가하면 습도를 낮춤)

X(기온)->M(습도) 관계(인과관계)

기온이 올라가면 포화 수증기량이 커져 상대습도가 낮아지고, 기온이 떨어지면 포화 수증기량이 작아져 상대습도가 높아짐

M(습도)->Y(발전량) 관계

습도가 높으면 태양전지 셀 표면에 물방울이 맺히는 현상이 일어날 수 있어 태양광에너지 발전성능을 낮춤

X(기온)->Y(발전량) 관계

기온이 높으면 태양광 에너지가 커져 발전량 증가