Data pipeline



Before start

- 1. 데이터 (1520 rows × 4 columns)
 - 5개의 연료원
 - 기간: 2021년 1월~ 10월 (300일)
 - 정보의 한정성
- 2. 데이터 structure 변화가 필요했음
 - 제공된 데이터 상엔 '지역' 정보가 포함되어 있었으나 동일한 데이터를 갖고 있었음

Data structure

	지역	연료원	거래일	전력거래량(MWh)
0	서울시	LNG	2021-01-01	15907.170310
1	서울시	바이오	2021-01-01	50.683200
2	서울시	연료전지	2021-01-01	1241.210151

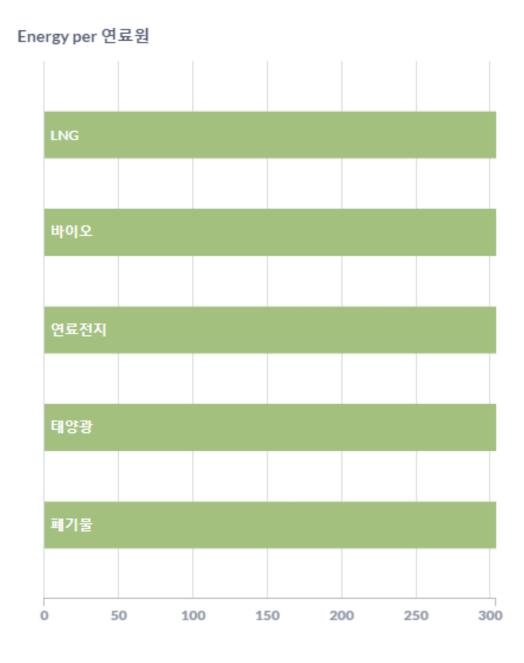
	연료원	전력거래량(MWh)	Year	Month	Day
0	LNG	15907	2021	1	1
1	바이오	50	2021	1	1
2	연료전지	1241	2021	1	1
3	태양광	39	2021	1	1
4	폐기물	492	2021	1	1

1. 지역 데이터 제거

- 2. 거래일 데이터
 - 연도, 달, 일 별 데이터로 분리
 - 연도의 통일성 --> 제거
 - column명에 괄호 --> DB상 인식 오류 --> 변경
- 3. (1520,4) --> (1520,5)
- 4. csv --> db 변환

Model

한국전력거래소가 지난 12월 6일 공개한 서울시 전력 거래량 데이터에 기반하여 전력 거래량 예측을 수행본 데이터는 2021년 1월부터 10월까지의 연료원 별 전력거래량 정보를 포함함



- 300일간의 거래정보를 포함
- 다중선형회귀모델을 이용하여 예측을 수행



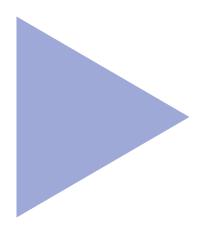




- + 전력거래소 수요자원 전력거래시스템 연동
- + 한국 에너지 공단의 신재생에너지에 대한 설명 연동하여 제공



Dashboard



https://seoul-energy.herokuapp.com/



