데이터를 통한 전력 수요 예측

문제 정의 및 변수도출

의사 결정자: 정부

기업이나 가계가 아닌 정부의 입장에서
전력수요 예측

목표: 정확한 전력 수요 예측

미리 특정 시점의 전력 수요를 정확히 예측함으로써 전력부족사태(블랙아웃)을 막는 동시에 지나치게 많은 전력을 생산하는 일도 막음

대안선택: 효용 vs. 비용

예측의 정확도만큼 예측의 신속성 중요
데이터 획득이 용이해야 함

알고리듬 I : 다중회귀분석

에측변수에 범주형 변수가 없고
선형관계가 기대

알고리듬 II: 군집분석

변수에 대한 민감도 차이를 보이는 지역 존재:
개별 그룹 타겟팅이 수요예측에 용이

예측변수: 시간, 공간 데이터

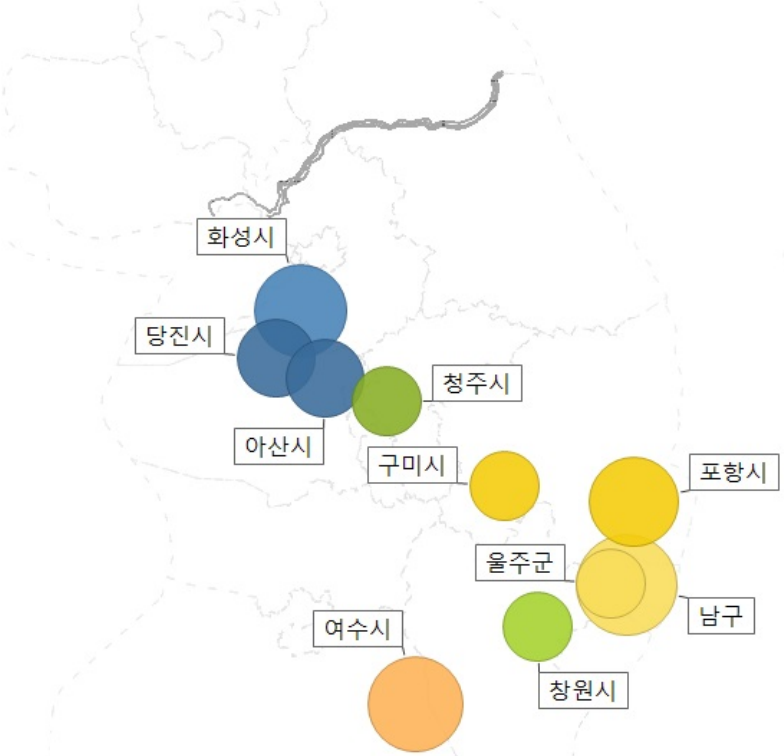
동일 지역이 시간에 따라 다른 수요를 보임
동시간에 지역마다 편차 존재

변수도출 및 가설

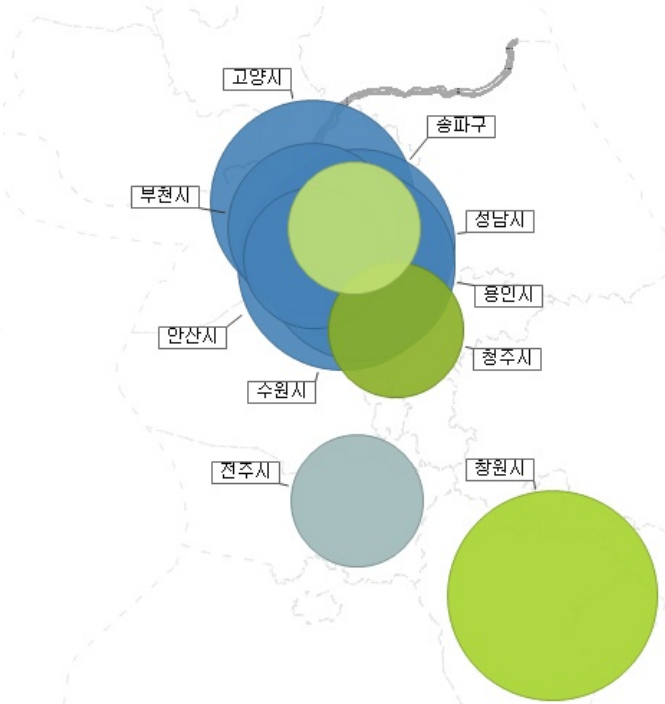
지역 인구
기후 (기온, 강우량)
계절별
전력수요 차이 존재
전력량 분석
시간이 흐를수록
더 많은 양의 전력 소비
인구가 많을수록
전력수요가 높을 것
시간
지역별 특징
전기판매단가
판매단가가 비싸다면
전력을 적게 소비할 것
공업지대는 주거지역에 비해
기후별 편차 적을 것

Data Visualization:
전력소비량 & 인구

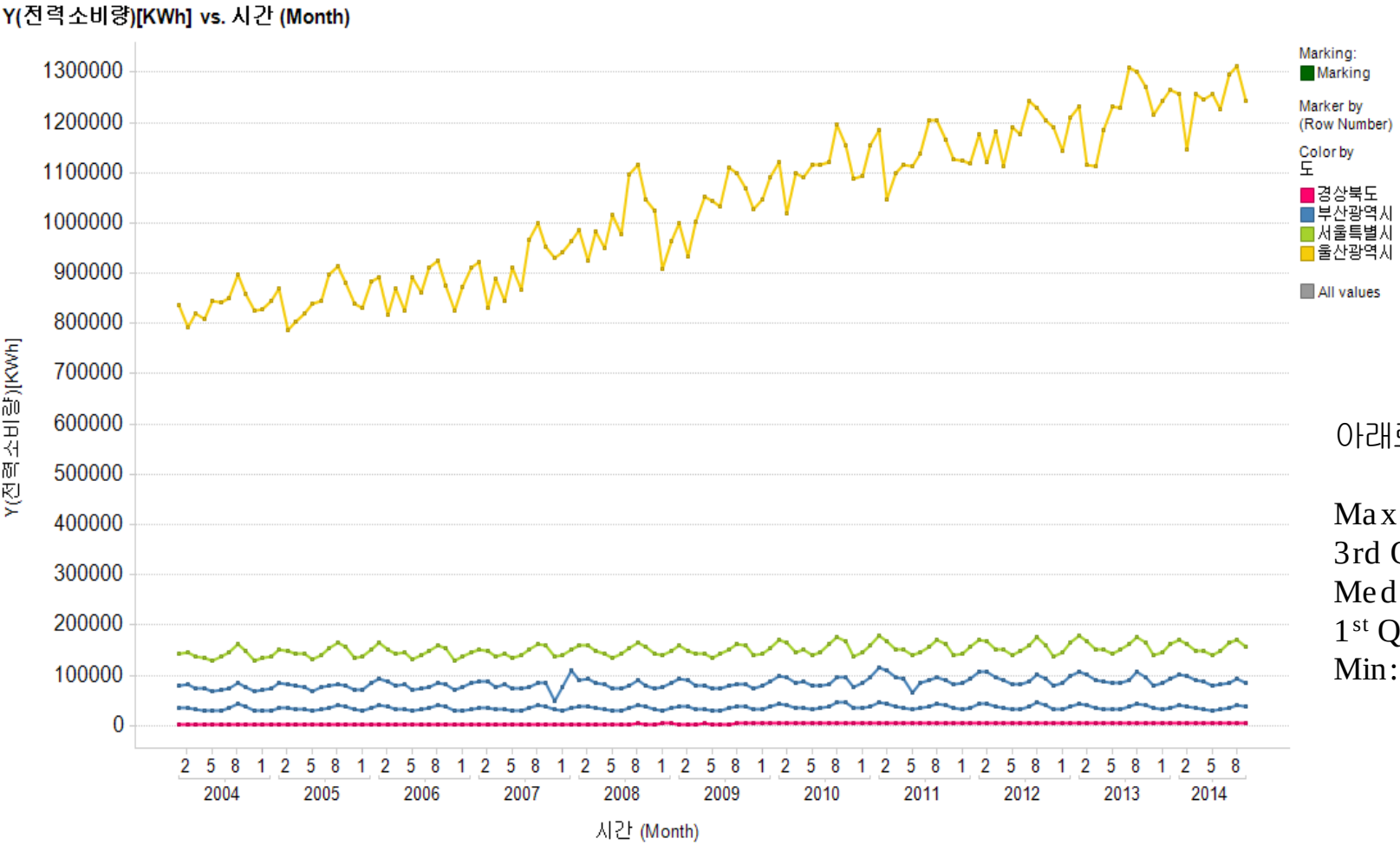
전력소비량



인구



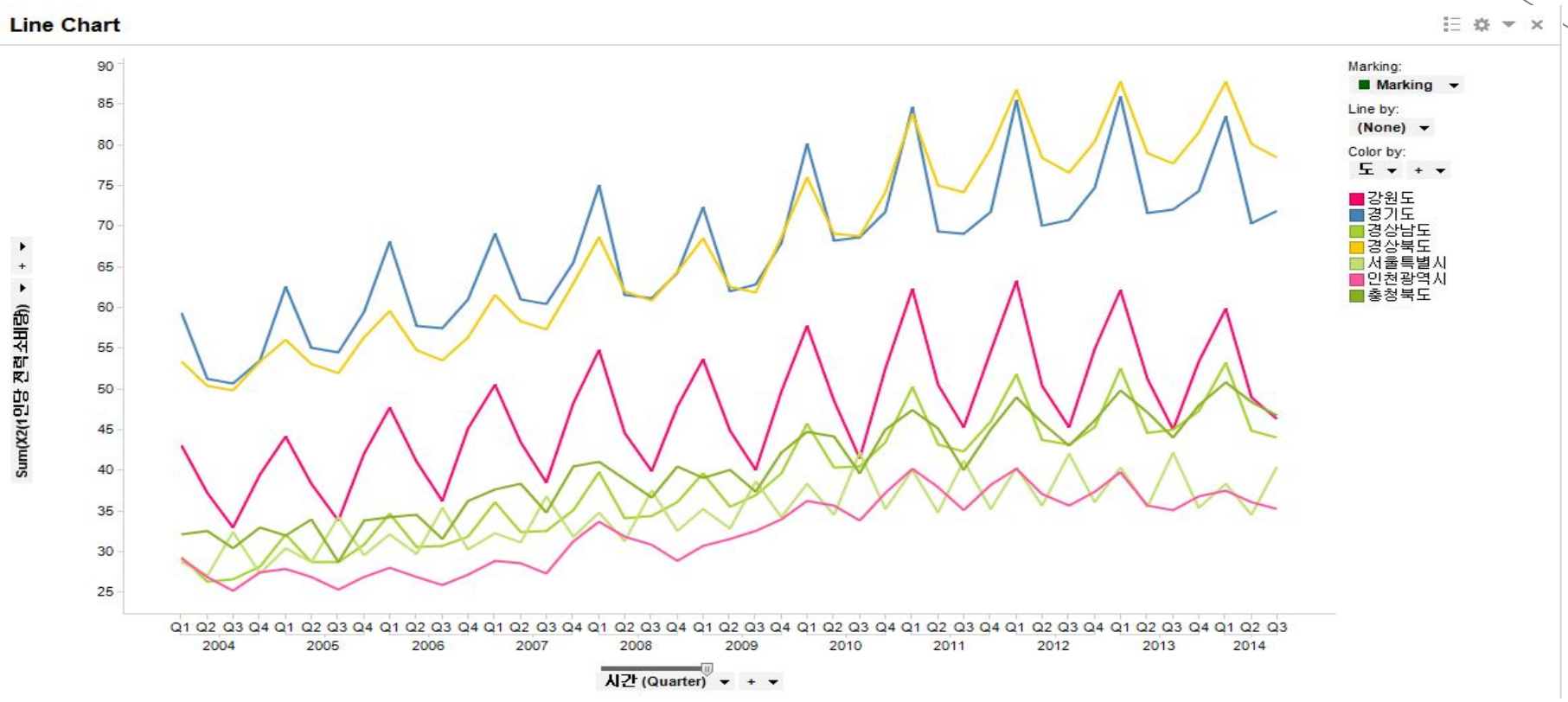
Data Visualization:
전력소비량 vs. 시간



아래로 □□□ □□

Max: □□ □□
3rd Q: □□ □□□
Med: □□ □□□
1st Q: □□ □□
Min: □□□□ □□□

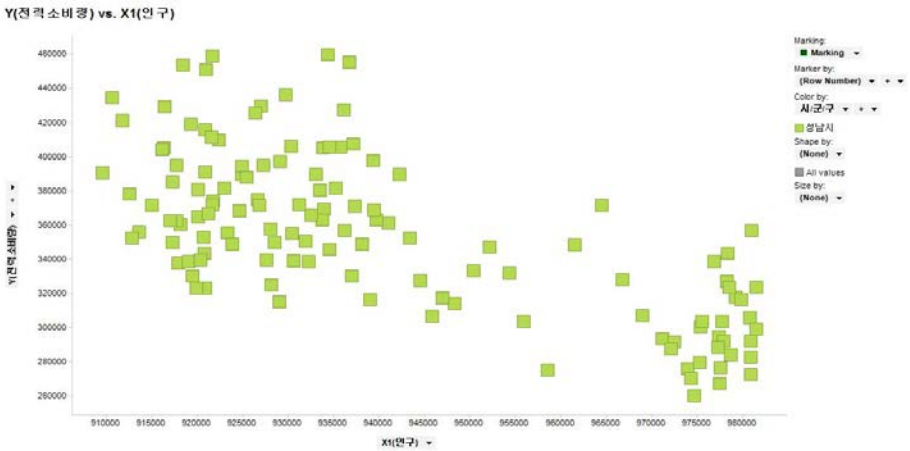
Data Visualization:
1인당 전력소비량 vs. 시간



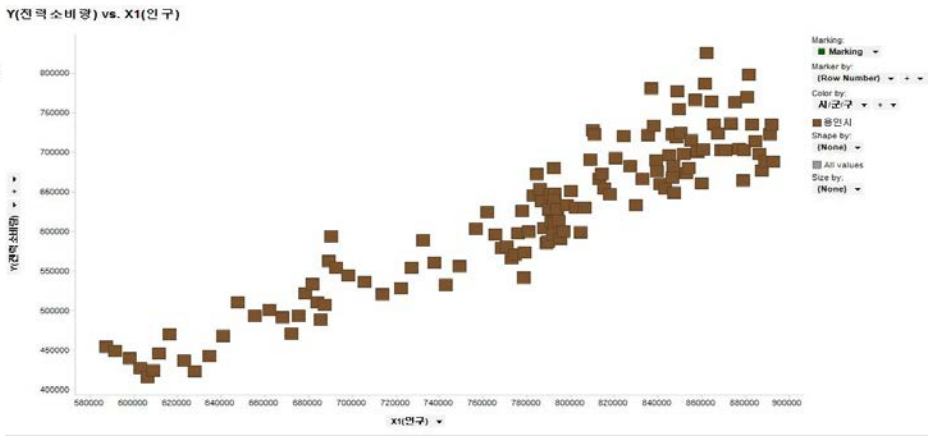
✓ 시간이 흐름에 따라, 1인당 전력소비량은 계속해서 증가하고 있음

Data Visualization: 전력소비량 vs. 인구

경기도 성남



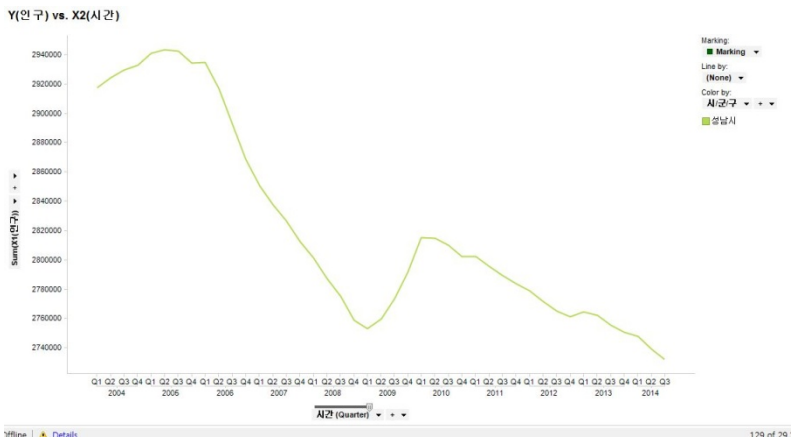
경기도 용인



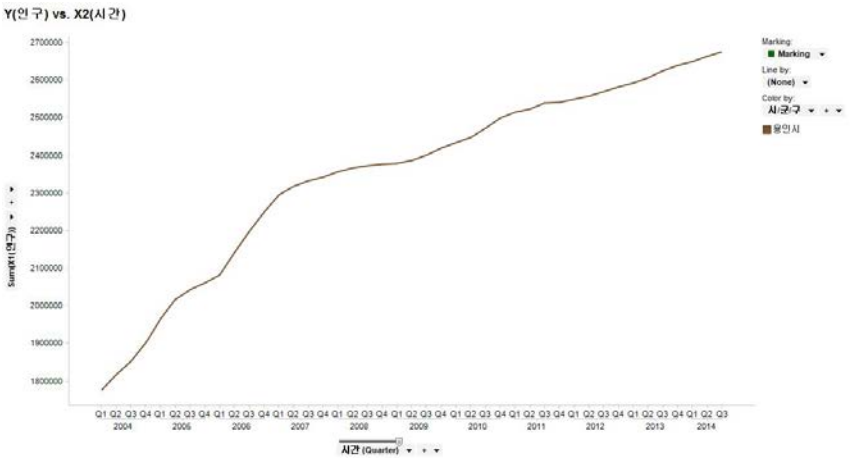
- ✓ 인구를 X, 전력소비량을 Y로 했을 때 다음과 같이 상이한 패턴을 보이는 집단 존재
- ✓ 1인당 전력소비량을 고려하면 두 집단의 차이는 시간에 따른 인구변화

Data Visualization: 전력소비량 vs. 인구

경기도 성남

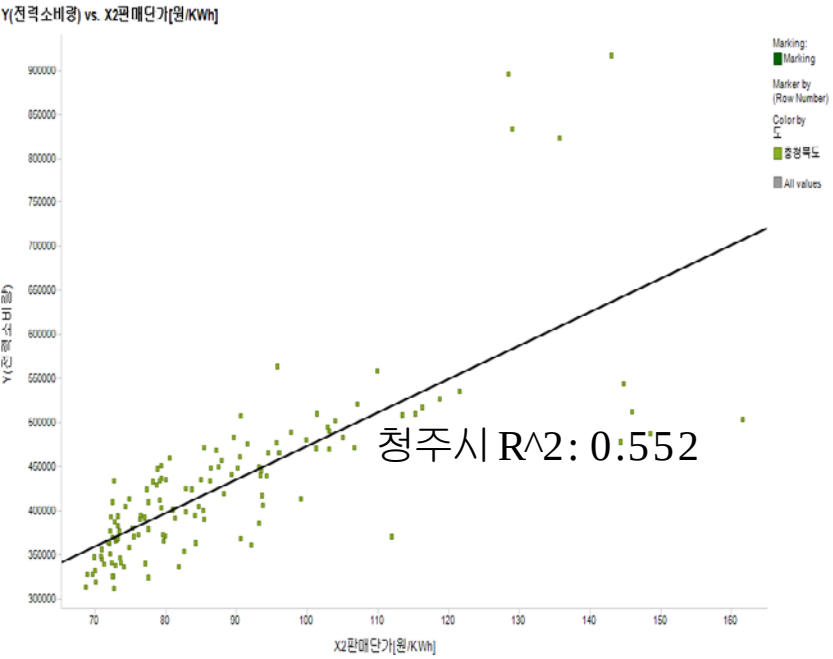


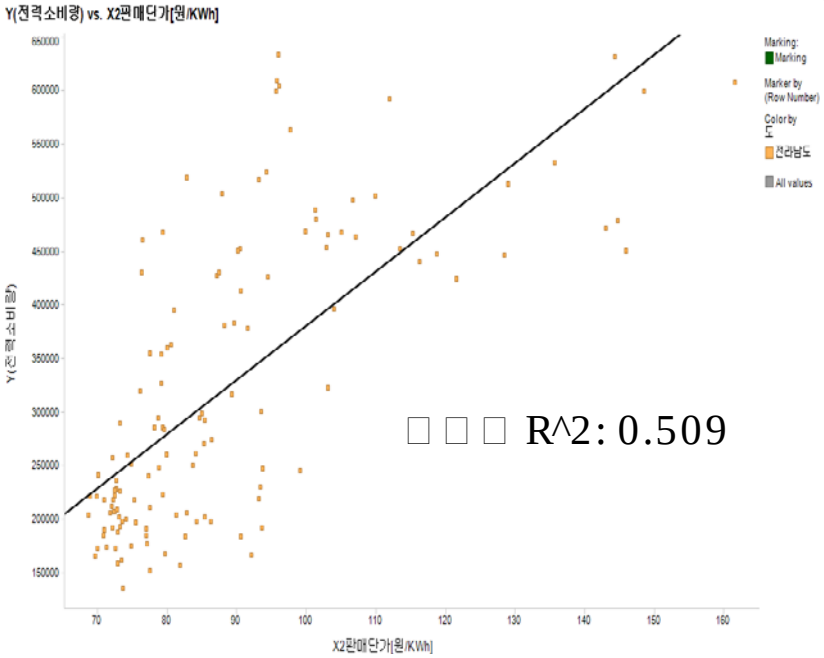
경기도 용인



- ✓ 위처럼 시간에 따른 인구변화를 갖고 설명 가능. 즉, 인구-전력 그래프는 시간-인구를 기준으로 집단으로 나뉠 필요성 존재

Data Visualization:
전력소비량 vs. 판매단가

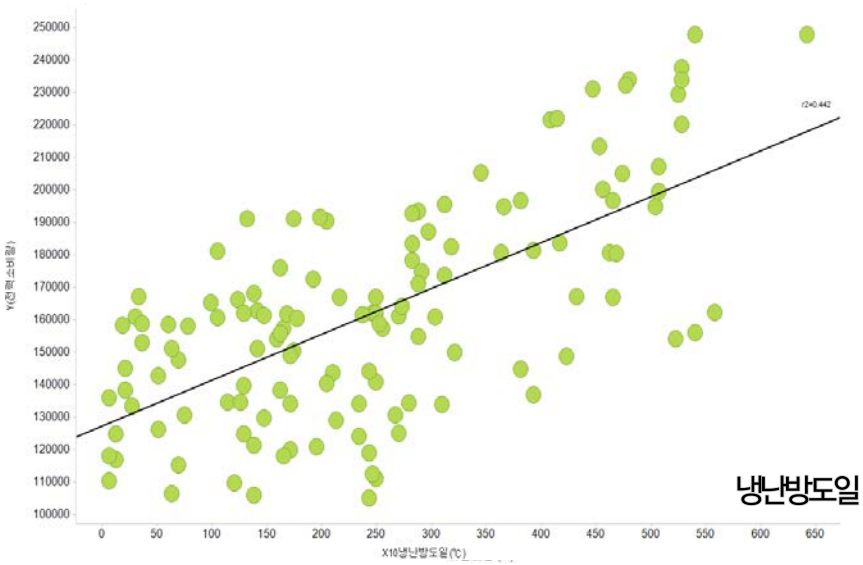




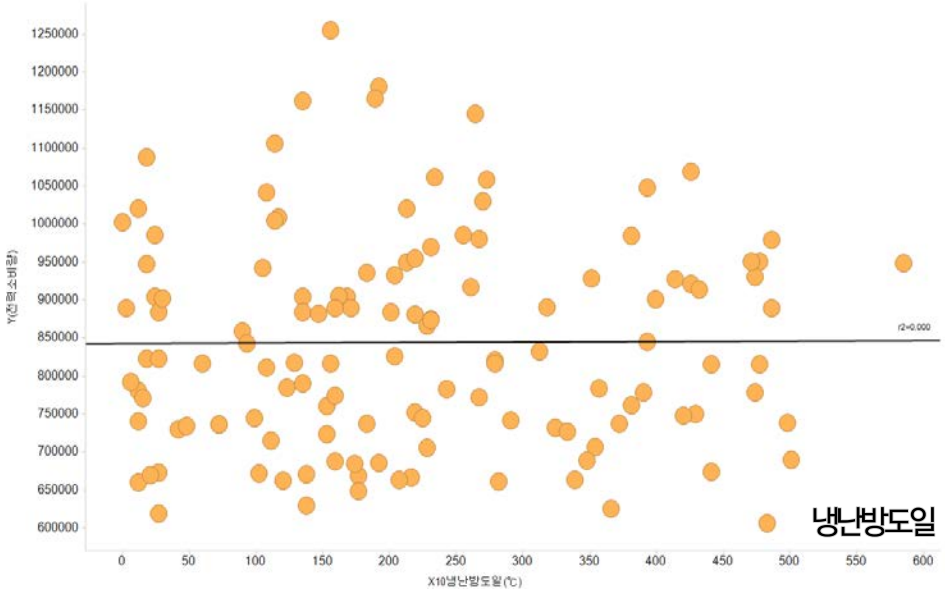
✓ 판매단가를 X로 하여 판매단가에 선형관계를 보이는 지역 탐색

Data Visualization:
전력소비량 vs. 기후요소

경기도 광주

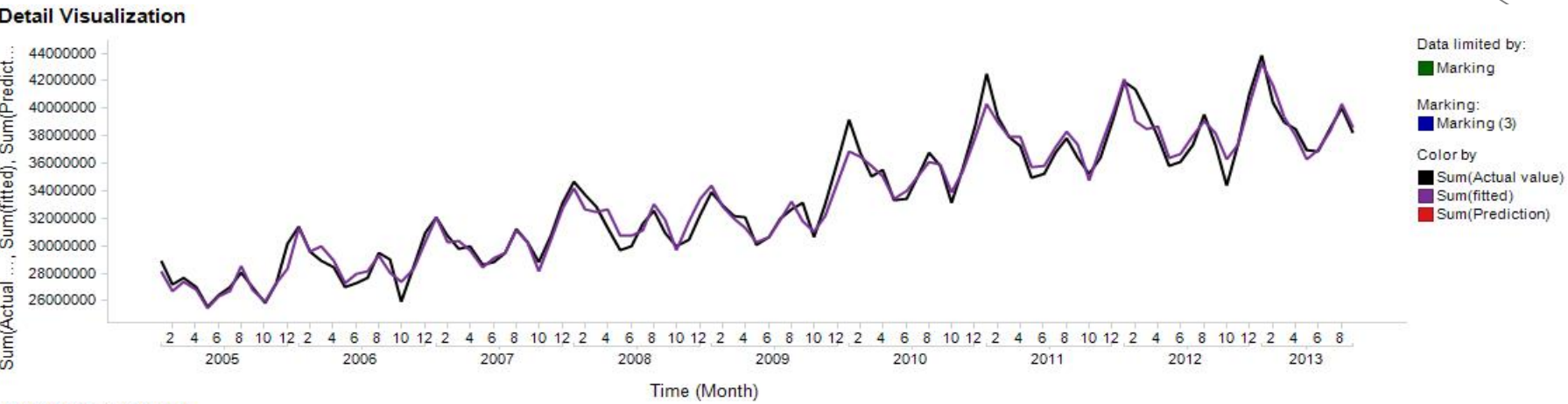


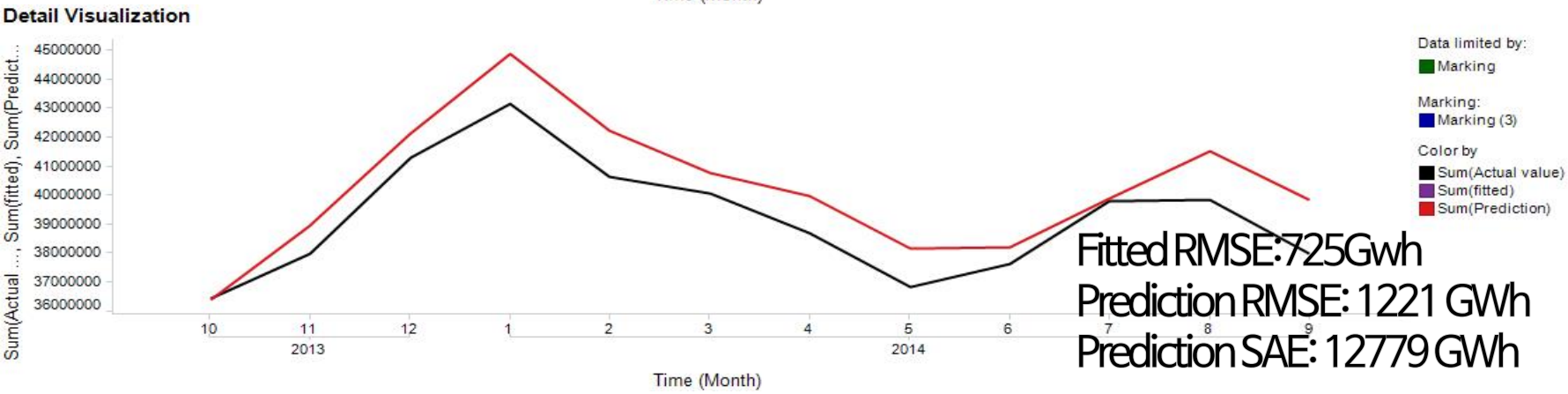
전라남도 여수



✓ 대표적 주거도시인 경기도 광주는 기후요소의 영향을 상대적으로 많이 받는 반면, 대표적 공업도시인 전라도 여수는 기후요소의 영향을 거의 받지 않는다.

Benchmark Analysis;
전국 전력소비량 시계열 분석



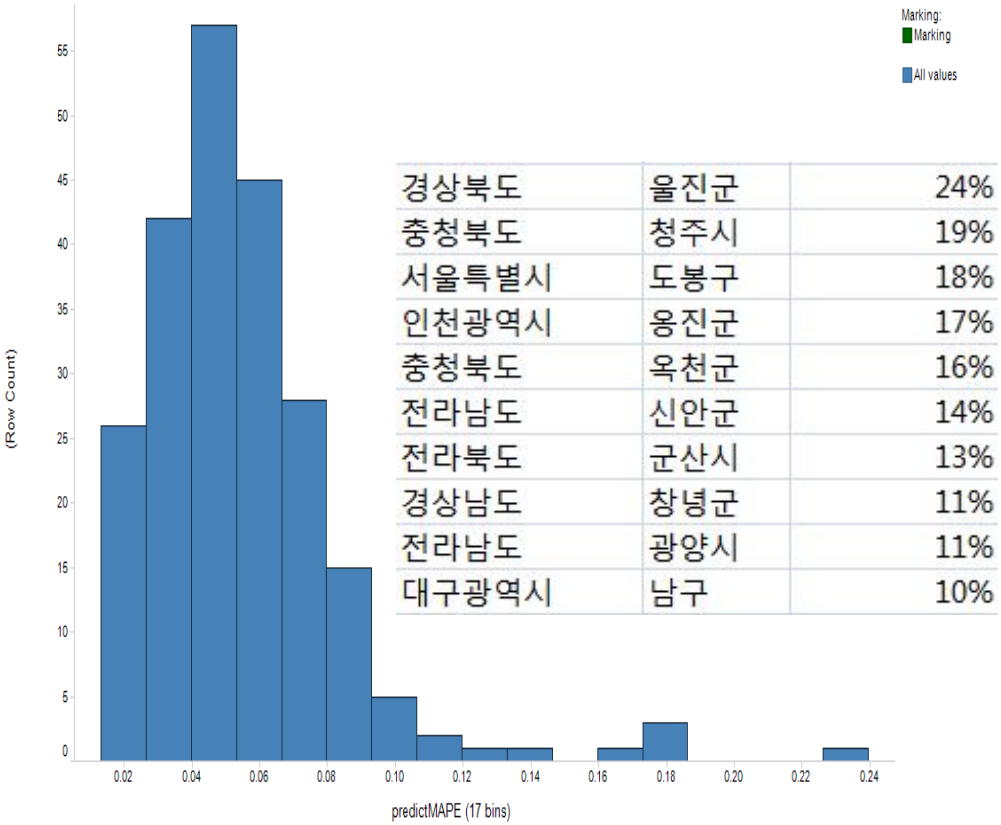


Fitted RMSE: 725Gwh
Prediction RMSE: 1221 GWh
Prediction SAE: 12779 GWh

- ✓ Holt-Winters Method를 사용해 12를 주기로 seasonality 반영
- ✓ 현 시점을 2013년 9월로 놓고, 2013년 9월까지의 data를 training set로 하여 2014년 9월까지의 기간을 예측

Error Analysis:
지역별 전력소비량 시계열 분석

시계열분석 MAPE 분포



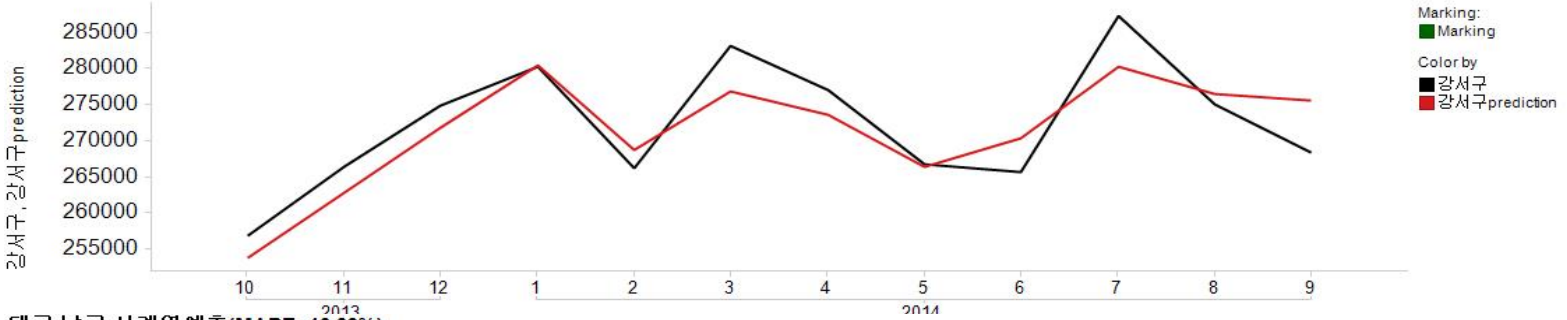
경상북도	울진군	24%
충청북도	청주시	19%
서울특별시	도봉구	18%
인천광역시	옹진군	17%
충청북도	옥천군	16%
전라남도	신안군	14%
전라북도	군산시	13%
경상남도	창녕군	11%
전라남도	광양시	11%
대구광역시	남구	10%

- ✓ Scale Independent 하게 MAPE 적용
- ✓ 히스토그램 관찰 결과 대체로 오차율이 높지 않음
- ✓ 몇몇 오차율이 큰 지역 때문에 전체 오차가 커진 것으로 추정 가능
- ✓ MAPE Top 10 지역 도출

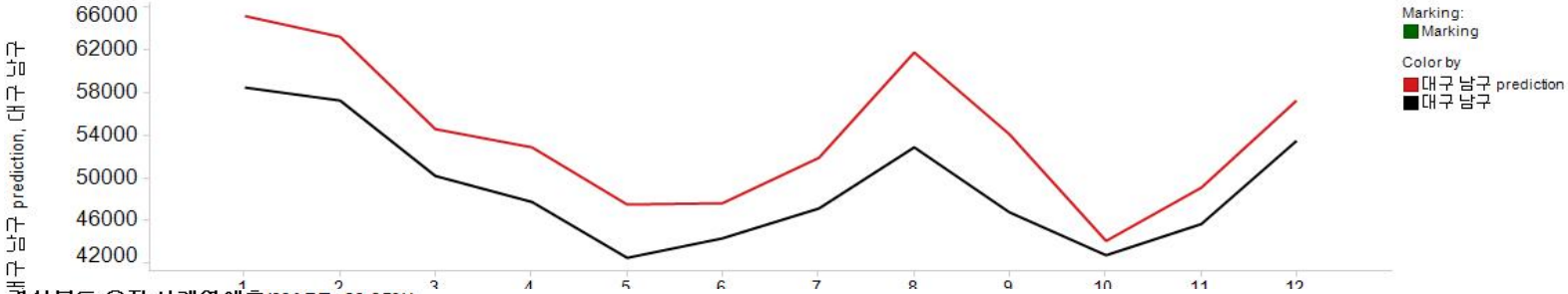
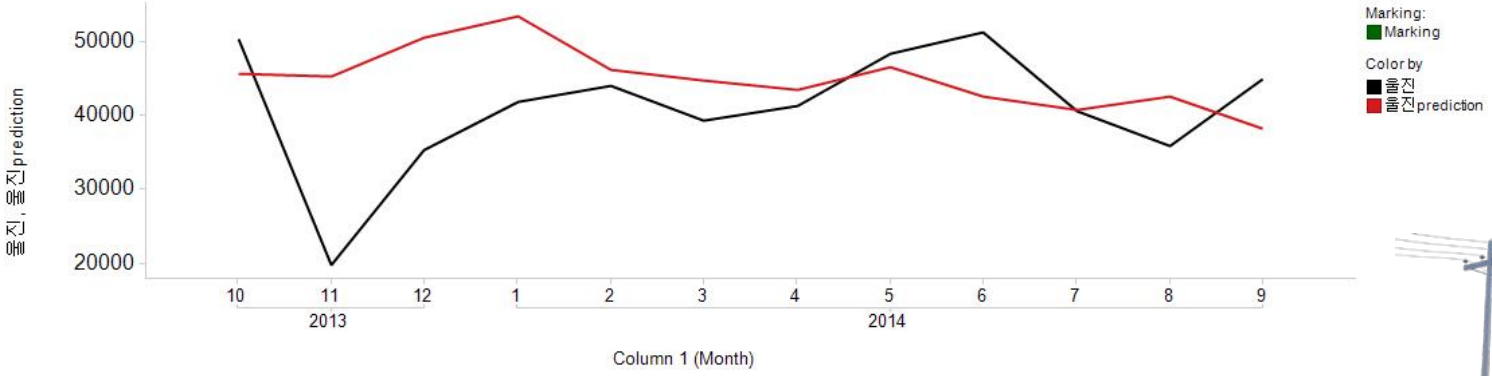
Time Series Prediction:

Best & Worst Case

부산 강서구 시계열예측(MAPE=1.32%)



대구 남구 시계열예측(MAPE=10.09%)



경상북도 울진 시계열예측(MAPE=23.95%)