

불법주정차 민원다발지역 요인 분석

2018.8.28

한광종
이준범
정재민
김영우
정진호

1. 주제 선정 및 분석 배경
2. 데이터 수집
3. 데이터 전 처리
4. 데이터 탐색
5. 데이터 분석
6. 데이터 시각화
7. 기대 효과
8. 시스템 구성도

주제 선정 배경

❶ 불법 주정차 차량으로 인해 발생한 각종 교통사고 현황 자료: 부산지방경찰청

연도	발생건수
2012	1천360건
2013	977건
2014	960건

❷ 부산 소방차 진입곤란지역 현황

상습 주차	장애물 설치	경사지 등 도로구조 장애	주거지 주차장 설치
133곳			

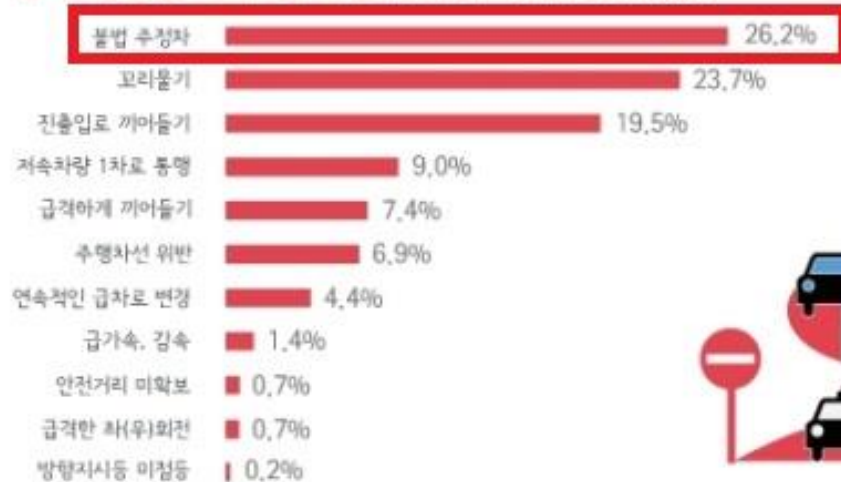
❸ 부산 비상: 방화문 자동 닫힘장치 미비

245건

※자료:부산소방안전

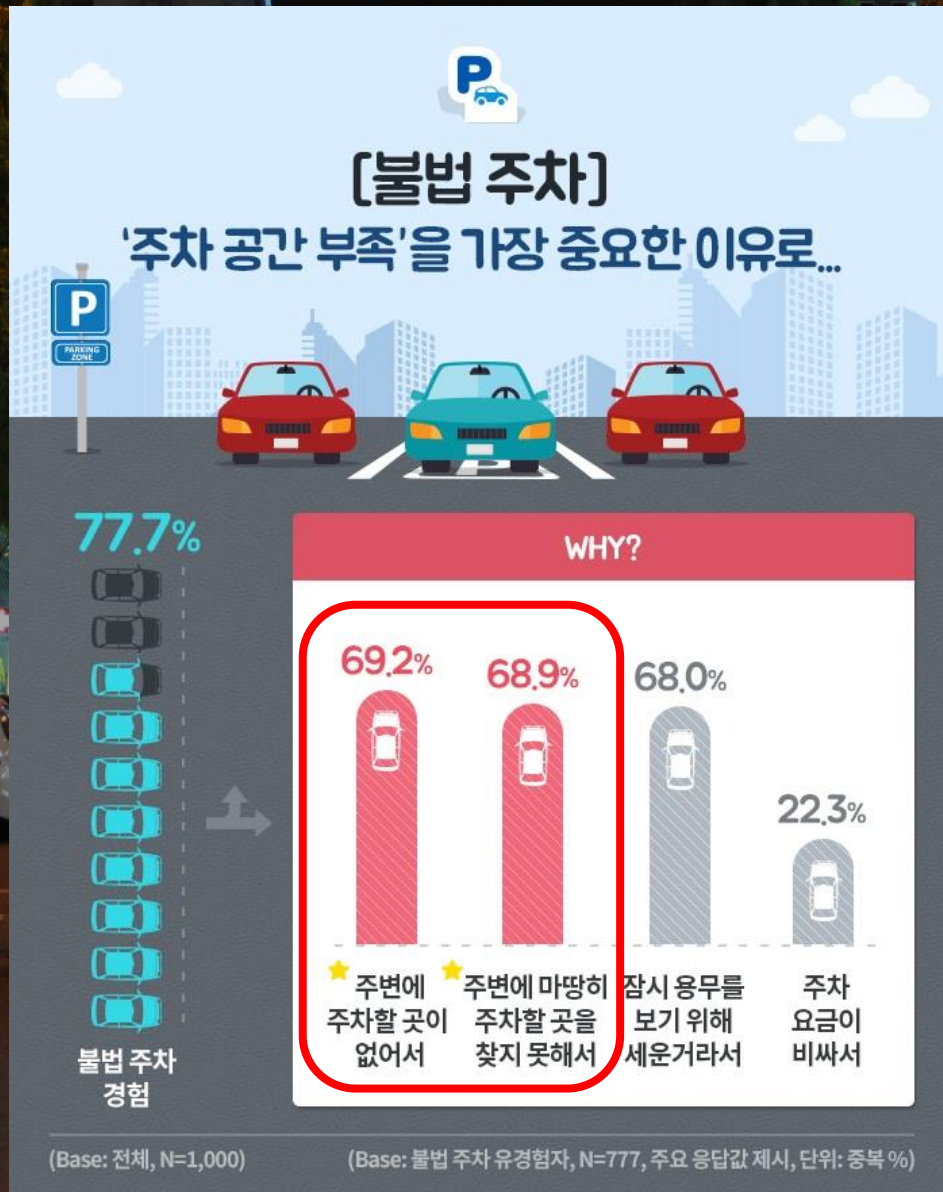
❹ 교통정체를 일으키는 나쁜 운전행태는 무엇일까?

교통정체의 가장 큰 원인으로 인식되는 운전행태로는 불법주정차, 꼬리물기, 진출입로 끼어들기 순 (운전자 인식, 복수응답)



(자료 : 서울연구원, "불법운전의 사회적 비용", 2011년)

주제 선정 배경



분석 배경

부산시 동구

속보 오늘이슈 정치 경제 사회 국제 문화 LIFE IT/과학

"주민 믿습니다" 불법주차 손 댄 부산 동구

자발적인 교통질서 유도한다며 좁은 산복도로도 단속 대신 계도

국제신문 민경진 기자 jnmin@kookje.co.kr | 입력 : 2016-08-26 20:37:26 | 본지 5면

트위터 페이스북 G+

- 무분별한 주차 탓에 교통난 가중
- "이상에 치우친 행정" 민원 빔발

부산 동구가 주민 스스로 질서를 지키도록 유도한다며 산복도로의 불법 주정차에 소극적으로 대응하면서 주차 문제가 가중되는 등 혼선을 빚고 있다. 주차공간 부족이라는 근본문제를 해결하지 않은 채 주민 자율을 강조하는 등 이상에 치우쳐 현실을 제대로 반영하지 않은 행정이라는 지적도 나온다.

부산시 해운대구



▲ 지난달 30일 광안리해수욕장 해변가, 불법 주정차 차량들이 1개 차로에 줄지어 주차돼 있다.

야경 명소로 손꼽히는 부산의 주요 해변이 불법 주정차로 몸살을 앓고 있다. 구청은 고정식 카메라, 이동식 차량으로 단속에 나서고 있지만 인근 상인, 주민들의 반발로 확실한 대책을 내놓지 못하고 있다는 지적이 나온다.



상시적으로 민원에 대처하는 지역을 대상으로

불법 주정차 민원다발지역의 요인을 분석

연구 가설

- 주거/상가의 종류에 따라 민원에 의해 단속된 차량 수가 차이가 있다.
- 단속 지점과 주차장의 거리가 멀수록 민원에 의해 단속된 차량 수가 차이가 있다.
- 단속 지점의 주위 시설물에 따라 민원에 의해 단속된 차량의 수가 차이가 있다.
- 단속 지점과 고정형 불법주정차단속시스템(CCTV)의 거리가 멀수록 민원에 의해 단속된 차량의 수가 많을 것이다.

데이터 수집

공공 데이터 포털



고정형 주 정차
단속 시스템(CCTV)
(4180건)



도로 시설물 현황
(327409건)



GIS 건물집합정보
(2463건)



부산시 거주자
우선 주차 지역
(2649건)

소상공인 포털



부산시
소상공인 데이터
(145084건)

웹 크롤링



주차장 데이터
(2992건)

정보 공개 포털



.....



기장군 / 남구 / 동래구 / 사상구 / 서구 / 연제구 / 영도구 / 중구 / 해운대구
주행형 주 정차 단속 데이터
(366052건)

데이터 수집 및 설명

주행형 주 정차 단속 데이터

- 부산광역시 총 16개 구/군에 정보 공개 포털에 데이터 신청.
 - ✓ 16개 구/군 중 9개 구/군이 기간 내 데이터 제공
 - ✓ 단속 일 / 단속 시간 / 단속 지점 3개의 컬럼으로 구성
 - ✓ 9개 구/군의 36052건의 단속 데이터 확보
 - ✓ 구별 단속 지점의 표기법이 상이하며 모호한 지점이 많아 정제가 필요함
 - ✓ 정기적 단속 지점의 값을 제외하고 데이터 전처리 진행

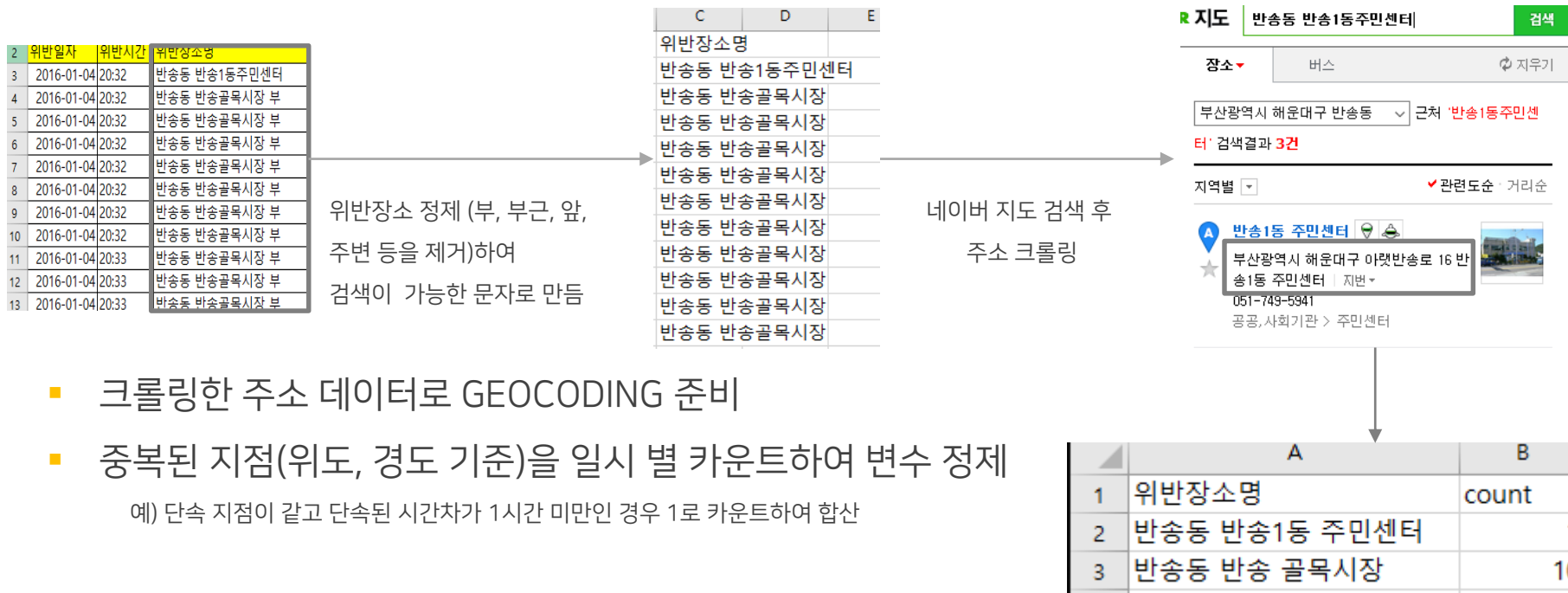
민영/공영 주차장 데이터

- 파이썬 Selenium 모듈을 이용하여 웹 크롤링
 - ✓ 네이버 지도 검색 표시 수는 300건 이상 표시 되지 않음
 - 검색 쿼리의 세분화를 통하여 해결 예) 장전1동 공영 주차장 → 세부 동 이름 + 공영/민영 + 주차장
 - ✓ 주소 / 공영 민영구분 2개 컬럼으로 구성
 - ✓ 위도, 경도가 표시 되지 않아 GEOCODING 작업 시행 필요

데이터 전처리

주행형 주 정차 단속 데이터

- 단속 지점이 모호한 지점 정제



- 크롤링한 주소 데이터로 GEOCODING 준비
- 중복된 지점(위도, 경도 기준)을 일시 별 카운트하여 변수 정제

예) 단속 지점이 같고 단속된 시간차가 1시간 미만인 경우 1로 카운트하여 합산

행정상 주소를 위도, 경도 데이터로 변경

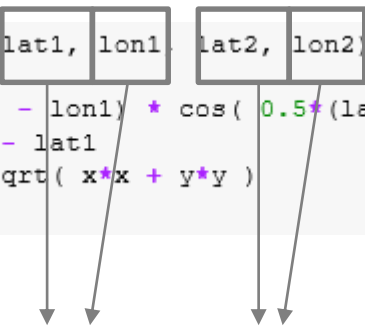
- 도로 시설물 데이터(327409건), 크롤링한 주차장 데이터(2992건), 주행형 주 정차 단속 데이터(366052건)을 NAVER API를 이용하여 위도, 경도 반환

데이터 전처리

거리 기반 COUNT 변수 생성

- 중복 값이 제거와 정기적 단속 구역을 제외한 주 정차 단속 지점(2826건)과 모든 건물 데이터와의 거리를 계산
- 2KM, 5KM 반경 내 모든 속성별 건물 수 카운트하여 변수 생성

```
def distance(lat1, lon1, lat2, lon2):
    R = 6371
    x = (lon2 - lon1) * cos( 0.5*(lat2+lat1) )
    y = lat2 - lat1
    d = R * sqrt( x*x + y*y )
    return(d)
```



① 단속지점 위도 경도 ② 건물/주차장 데이터의 위도 경도 → ①과 ②의 거리가 2KM 미만일 때, 해당 건물 속성 이름_1의 변수 1로 카운트하여 더함

- 구/군 별 단속 일 수(민원 대응 수)가 다르기 때문에 단속 지점 별 Count 값은 변환이 필요
→ (단속 지점의 단속 건수) / (해당 구의 총 단속 수)

데이터 전처리

최종 분석 테이블

	A	B	C	D	E	F	G	H	I
1	rate_all	gungu	latitude	longitude	building_01_1	building_02_1	building_03_1	building_04_1	building_05_1
2	8.85E-05	영도구	35.11372	129.04	0	1	0	0	6
3	0.002810963	중구	35.10855	129.0363	0	1	0	0	4
4	0.000353826	영도구	35.11288	129.0402	0	1	0	0	4
5	0.000265369	영도구	35.10115	129.0348	0	0	0	0	4
6	0.001405481	중구	35.10419	129.036	0	1	0	0	4
7	8.85E-05	영도구	35.10616	129.0358	0	0	0	0	4
8	0.00101626	중구	35.10195	129.0354	0	0	0	0	4
9	0.003891051	중구	35.10514	129.0357	0	0	0	0	4
10	0.001288417	동래구	35.20248	129.08	0	0	0	0	3
11	0.00607159	서구	35.10502	129.0208	0	0	0	0	3
12	0.000366703	동래구	35.19656	129.1021	0	0	0	0	3
13	0.000884564	영도구	35.09136	129.0351	0	0	0	0	3
14	0.004919185	중구	35.1045	129.0348	0	0	0	0	3
15	0.000530739	영도구	35.09352	129.0443	0	0	0	0	3
16	0.004310345	중구	35.10476	129.0345	0	0	0	0	3
17	1.77E-04	영도구	35.09228	129.0338	0	0	0	0	3

.....

BB	BC	BD	BE	BF	BG	BH	BI	BJ	BK	BL
sosang_O_2	sosang_S_2	sosang_P_2	cctv_count_1	cctv_count_2	prior_park	prior_park	private_parking_1	private_parking_2	public_parking_1	public_parking_2
8	3	0	0	2	0	0	0	3	1	0
3	0	0	0	2	0	0	0	2	1	0
11	0	0	1	0	0	0	0	2	1	0
5	0	1	1	1	0	0	0	0	4	0
0	6	0	0	0	0	0	0	1	2	0
5	0	0	0	1	0	0	0	0	0	1
2	2	0	0	1	0	0	0	0	1	0
2	3	0	0	1	0	0	0	0	0	0
0	0	0	0	0	0	0	0	1	1	0
0	6	0	0	2	0	0	0	0	0	1
0	1	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
3	3	0	0	3	0	0	0	0	1	0
0	4	0	0	3	0	0	0	0	0	0
3	2	0	1	1	0	0	0	0	0	0

- y : rate_all의 값이 상위 25%인 값은 1, 그 외의 값은 0임. 1은 민원 다발 지역, 0은 민원 지역으로 명명.
- rate_all : 각 단속 지점의 단속 수 / 해당 구의 전체 단속 수
- gungu : 총 9개 구/군 지역
- Building_건물코드_1(2): 2(5)KM 반경 내의 해당 건물 수
 - ✓ 건물 코드는 [환경정화시설, '문화/관광/레저시설', '종교시설', '근린생활시설', '보안/위험시설', '공장/창고시설', '자동차관련시설', '업무시설', '교육및복지시설', '여객(화물)운송시설', '농축수산물시설', '유통시설', '유흥/위락시설', '장묘시설', '숙박시설', '주택', '의료시설', '공공용시설'] 총 18개로 구성됨.
- Sosang_코드_1(2): 2(5)KM 반경 내의 소상공인포털 대 분류된 소상공인 영업소 수
 - ✓ 소상공인 대분류 코드는 D: 소매, F: 생활서비스, N: 관광/여가/오락, L: 부동산, O: 숙박, P: 스포츠, Q: 소매, R: 학문/교육, S: 의료

데이터 전처리

최종 분석 테이블

	A	B	C	D	E	F	G	H	I
1	rate_all	gungu	latitude	longitude	building_01_1	building_02_1	building_03_1	building_04_1	building_05_1
2	8.85E-05	영도구	35.11372	129.04	0	1	0	0	6
3	0.002810963	중구	35.10855	129.0363	0	1	0	0	4
4	0.000353826	영도구	35.11288	129.0402	0	1	0	0	4
5	0.000265369	영도구	35.10115	129.0348	0	0	0	0	4
6	0.001405481	중구	35.10419	129.036	0	1	0	0	4
7	8.85E-05	영도구	35.10616	129.0358	0	0	0	0	4
8	0.00101626	중구	35.10195	129.0354	0	0	0	0	4
9	0.003891051	중구	35.10514	129.0357	0	0	0	0	4
10	0.001288417	동래구	35.20248	129.08	0	0	0	0	3
11	0.00607159	서구	35.10502	129.0208	0	0	0	0	3
12	0.000366703	동래구	35.19656	129.1021	0	0	0	0	3
13	0.000884564	영도구	35.09136	129.0351	0	0	0	0	3
14	0.004919185	중구	35.1045	129.0348	0	0	0	0	3
15	0.000530739	영도구	35.09352	129.0443	0	0	0	0	3
16	0.004310345	중구	35.10476	129.0345	0	0	0	0	3
17	1.77E-04	영도구	35.09228	129.0338	0	0	0	0	3

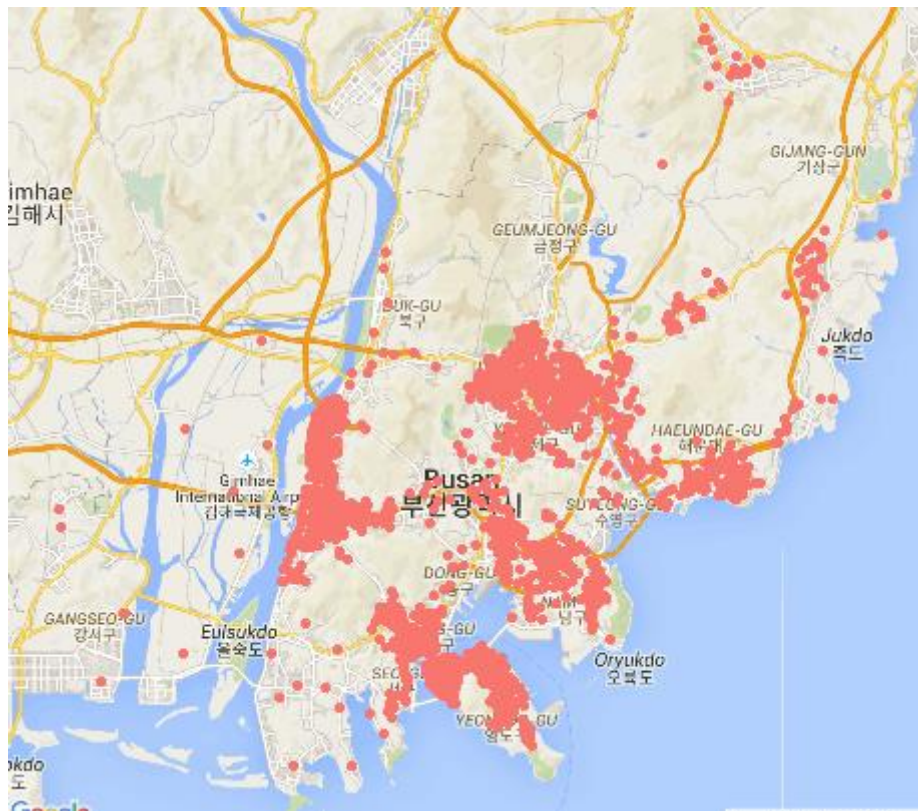
.....

BB	BC	BD	BE	BF	BG	BH	BI	BJ	BK	BL
sosang_O_2	sosang_S_2	sosang_P_2	cctv_count_1	cctv_count_2	prior_park	prior_park	private_parking_1	private_parking_2	public_parking_1	public_parking_2
8	3	0	0	2	0	0	0	3	1	0
3	0	0	0	2	0	0	0	2	1	0
11	0	0	1	0	0	0	0	2	1	0
5	0	1	1	1	0	0	0	0	4	0
0	6	0	0	0	0	0	0	1	2	0
5	0	0	0	1	0	0	0	0	0	1
2	2	0	0	1	0	0	0	0	1	0
2	3	0	0	1	0	0	0	0	0	0
0	0	0	0	0	0	0	0	1	1	0
0	6	0	0	2	0	0	0	0	0	1
0	1	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
3	3	0	0	3	0	0	0	0	1	0
0	4	0	0	3	0	0	0	0	0	0
3	2	0	1	1	0	0	0	0	0	0

- cctv_count_1(2): 2(5)KM 반경 내의 CCTV 수
- prior_parking_1(2): 2(5)KM 반경 내의 거주지 우선 주차 구역 수
- private_parking_1(2): 2(5)KM 반경 내의 민영 주차장 수
- public_parking_1(2): 2(5)KM 반경 내의 공영 주차장 수

데이터 탐색

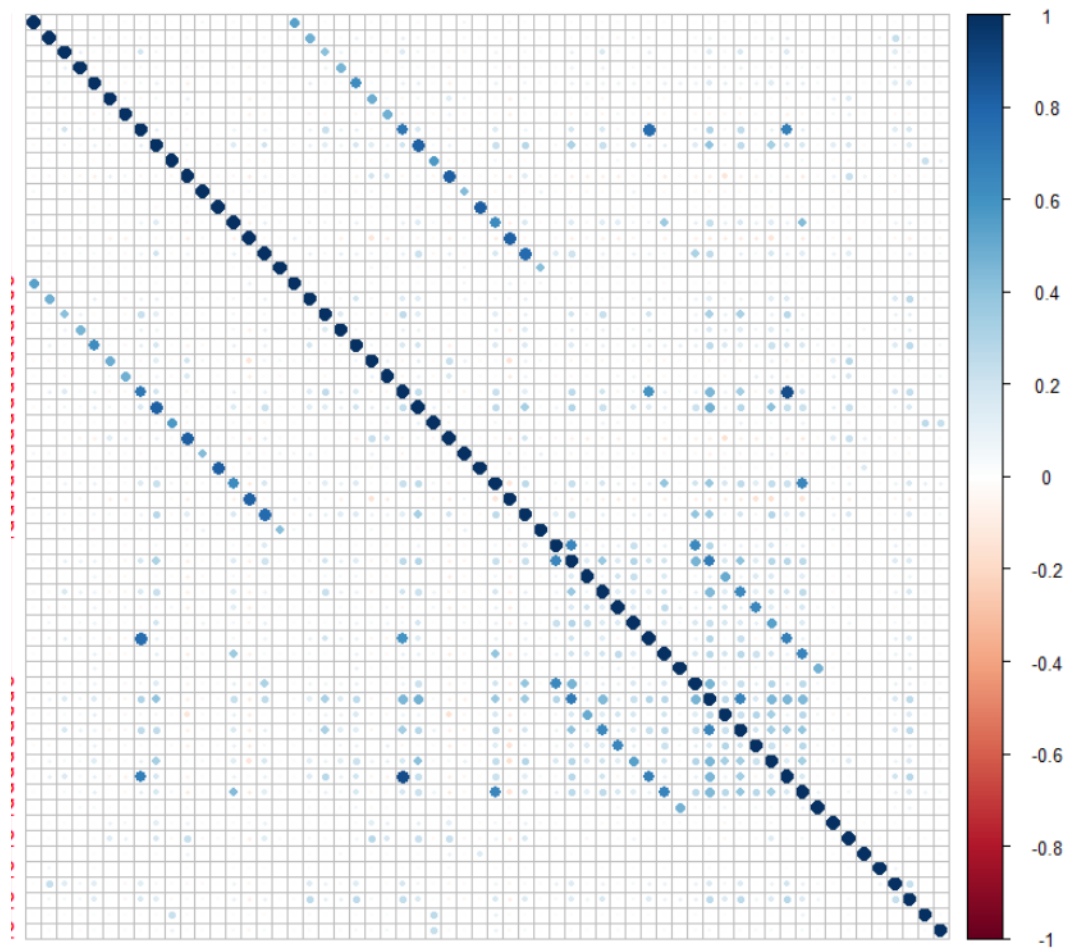
부산광역시 불법 주 정차 단속 현황 지도



- 정기적 단속 지점을 제외하지 않은 단속 현황
→ 제공 받은 구의 모든 지점에 불법 주정차 단속된 건수가 분포되어 있음.

데이터 탐색

독립 변수 간 상관관계



변수간 강한 상관 관계가 존재함. 회귀 분석 시 독립 변수 간 다중 공선성 예상

데이터 분석 - 선형 회귀 분석

- 하이브로 [0-1] 변환 후 데이터를 받아 선형 회귀 분석 시행
- 로지스틱 회귀 분석 사용을 위한 변수 선정
- 독립 변수 간 상관계수가 0.7 이상 나온 변수들이 존재하여 Stepwise를 사용하여 변수 선택을 진행

```
> summary(steped_df)
```

```
Call:
lm(formula = y ~ gungu + building_13_1 + building_10_1 + building_16_1 +
    building_04_2 + building_10_2 + building_06_2 + building_12_2 +
    sosang_Q_1 + sosang_D_2 + sosang_F_2 + sosang_S_2 + cctv_count_1 +
    public_parking_2, data = df_var)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-0.59706 -0.24411 -0.14492 -0.00456  0.99483
```

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.56421	0.05012	11.256	< 2e-16 ***
gungu	-0.26016	0.05457	-4.768	1.96e-06 ***
gungu동래구	-0.25519	0.05453	-4.680	3.01e-06 ***
gungu사상구	-0.29058	0.05425	-5.357	9.17e-08 ***
gungu서구	-0.32908	0.05956	-5.525	3.59e-08 ***
gungu연제구	-0.53900	0.06639	-8.119	6.99e-16 ***
gungu영도구	-0.40464	0.05165	-7.835	6.61e-15 ***
gungu중구	0.45393	0.06133	7.402	1.77e-13 ***
gungu해운대구	-0.27658	0.05599	-4.940	8.28e-07 ***
building_13_1	0.37860	0.14613	2.591	0.009624 **
building_10_1	0.49960	0.14156	3.529	0.000423 ***
building_16_1	-0.28924	0.07357	-3.931	8.65e-05 ***
building_04_2	-0.26510	0.11828	-2.241	0.025081 *
building_10_2	-0.17697	0.08511	-2.079	0.037682 *
building_06_2	-0.15211	0.05311	-2.864	0.004215 **
building_12_2	0.24534	0.11689	2.099	0.035922 *
sosang_Q_1	0.62147	0.19170	3.242	0.001201 **
sosang_D_2	-0.28461	0.18790	-1.515	0.129974
sosang_F_2	0.22998	0.10412	2.209	0.027271 *
sosang_S_2	-0.11103	0.07686	-1.445	0.148670
cctv_count_1	0.08174	0.03787	2.158	0.030984 *
public_parking_2	-0.33216	0.16322	-2.035	0.041944 *

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.3814 on 2803 degrees of freedom
Multiple R-squared:  0.2302,    Adjusted R-squared:  0.2244
F-statistic: 39.91 on 21 and 2803 DF,  p-value: < 2.2e-16
```

```
> |
```

```
> vif(steped_df)
```

	GVIF	Df	GVIF^(1/(2*Df))
gungu	2.063316	8	1.046310
building_13_1	1.018989	1	1.009450
building_10_1	1.509720	1	1.228707
building_16_1	1.158828	1	1.076489
building_04_2	1.784587	1	1.335884
building_10_2	1.624331	1	1.274492
building_06_2	1.199778	1	1.095344
building_12_2	1.516315	1	1.231387
sosang_Q_1	1.431627	1	1.196506
sosang_D_2	1.519071	1	1.232506
sosang_F_2	1.398651	1	1.182646
sosang_S_2	1.247540	1	1.116933
cctv_count_1	1.055907	1	1.027573
public_parking_2	1.084128	1	1.041215

- Stepwise를 사용 후 다중공선성이 제거 된 것을 확인

→ 해당 변수로 로지스틱 회귀 분석 진행

데이터 분석 - 로지스틱 회귀 분석

```
Call:
glm(formula = y ~ gungu + building_13_1 + building_10_1 + building_16_1 +
    building_04_2 + building_10_2 + building_06_2 + building_12_2 +
    sosang_Q_1 + sosang_D_2 + sosang_F_2 + sosang_S_2 + cctv_count_1 +
    public_parking_2, family = "binomial", data = train)
```

Deviance Residuals:

	Min	1Q	Median	3Q	Max
	-1.42815	-0.69862	-0.52309	-0.00017	2.60393

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.3746	0.3152	1.188	0.234671
gungu남구	-1.2652	0.3525	-3.590	0.000331 ***
gungu동래구	-1.0510	0.3512	-2.993	0.002764 **
gungu사상구	-1.2518	0.3512	-3.564	0.000365 ***
gungu서구	-1.4823	0.4037	-3.672	0.000241 ***
gungu연제구	-17.8456	481.1041	-0.037	0.970411
gungu영도구	-2.0522	0.3352	-6.123	9.18e-10 ***
gungu중구	17.6433	379.3027	0.047	0.962900
gungu해운대구	-1.4451	0.3701	-3.905	9.44e-05 ***
building_13_1	4.0336	1.3856	2.911	0.003602 **
building_10_1	19.9860	939.5195	0.021	0.983028
building_16_1	-2.6996	0.7341	-3.678	0.000235 ***
building_04_2	-2.6395	1.3233	-1.995	0.046091 *
building_10_2	-15.9691	939.5181	-0.017	0.986439
building_06_2	-1.0359	0.5276	-1.963	0.049596 *
building_12_2	1.5609	1.0365	1.506	0.132100
sosang_Q_1	3.3979	1.7072	1.990	0.046558 *
sosang_D_2	-2.9338	2.3341	-1.257	0.208781
sosang_F_2	2.8980	0.7761	3.734	0.000188 ***
sosang_S_2	-0.6030	0.6551	-0.921	0.357290
cctv_count_1	0.3066	0.3429	0.894	0.371244
public_parking_2	-6.2487	2.1766	-2.871	0.004094 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 2182.5 on 1976 degrees of freedom
Residual deviance: 1693.3 on 1955 degrees of freedom
AIC: 1737.3

Number of Fisher Scoring iterations: 16

- 2KM 반경 내 유흥/위락시설 수(building_13_1), 5KM 반경 내 유통시설(building_12_2), 2KM 반경 내 소매 소상공인 업소 수(sosang_Q_1), 5KM 반경 내 생활 서비스 소상공인 업소 수(sosang_F_2)가 한 단위 증가할수록 민원 다발 지역이 될 확률이 증가함.
- 2KM 반경 내 주택 수(building_16_1), 5KM 반경 내 근린생활시설 수(building_04_2), 5KM 반경 내 공장창고시설 수(building_06_2), 5KM 반경 내 공영주차장 수(public_parking_2)가 한 단위 증가할수록 민원 다발 지역이 될 확률이 감소함.
- 각 회귀 계수는 오즈비로 한 단위 증가할 때 $\exp(\text{회귀계수})$ 만큼 확률이 증가함

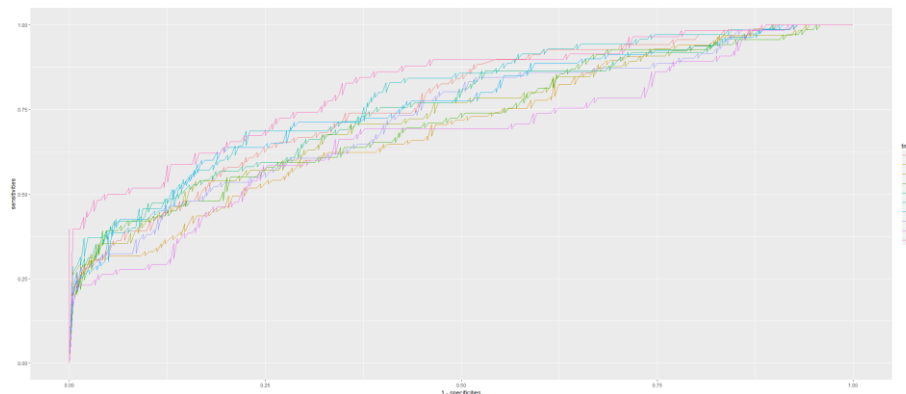
데이터 분석 - 로지스틱 회귀 분석

로지스틱 회귀 분석 모델 평가

- Train / Test 셋을 7:3 비율로 INDEX용 난수를 생성하여 모델용, 테스트용 셋으로 나눔.
- 반복문을 이용하여 시드를 바꾸어 가며 1000개의 list 형태로 데이터 프레임을 저장 후 모델을 검정
- 1000개의 모델 정확도 / 민감도 / 특이도 / 정밀도 평균, 분산, 최대값, 최소값을 구하여 과적합 여부 파악.

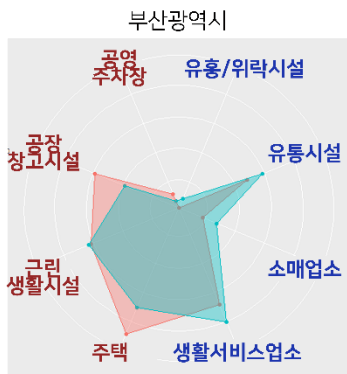
	평균	분산	최대값	최소값
ACCURACY	0.7212	0.0012	0.82429	0.5849
SENSITIVITY	0.6104	0.0053	0.8164	0.3418
SPECIFICITY	0.7581	0.0047	0.9693	0.5178
PRECISION	0.4660	0.0028	0.7788	0.3403
AUC	0.7437	0.0002	0.7993	0.683

- 낮은 분산을 근거로 과적합이 없는 안정성 있는 모델로 파악이 가능함

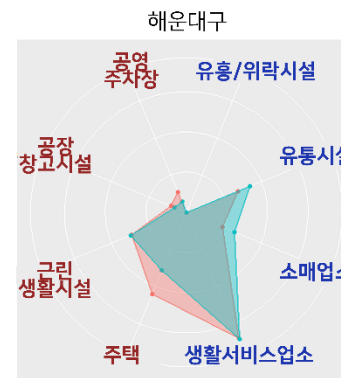


데이터 시각화

- 로지스틱 회귀 분석의 유의한 계수와 변수의 중요도를 근거로 민원다발 지역을 나누는 변수 추출
→ 유의한 변수, 유의한 구/군의 요인을 시각화 자료로 파악



민원발생지역

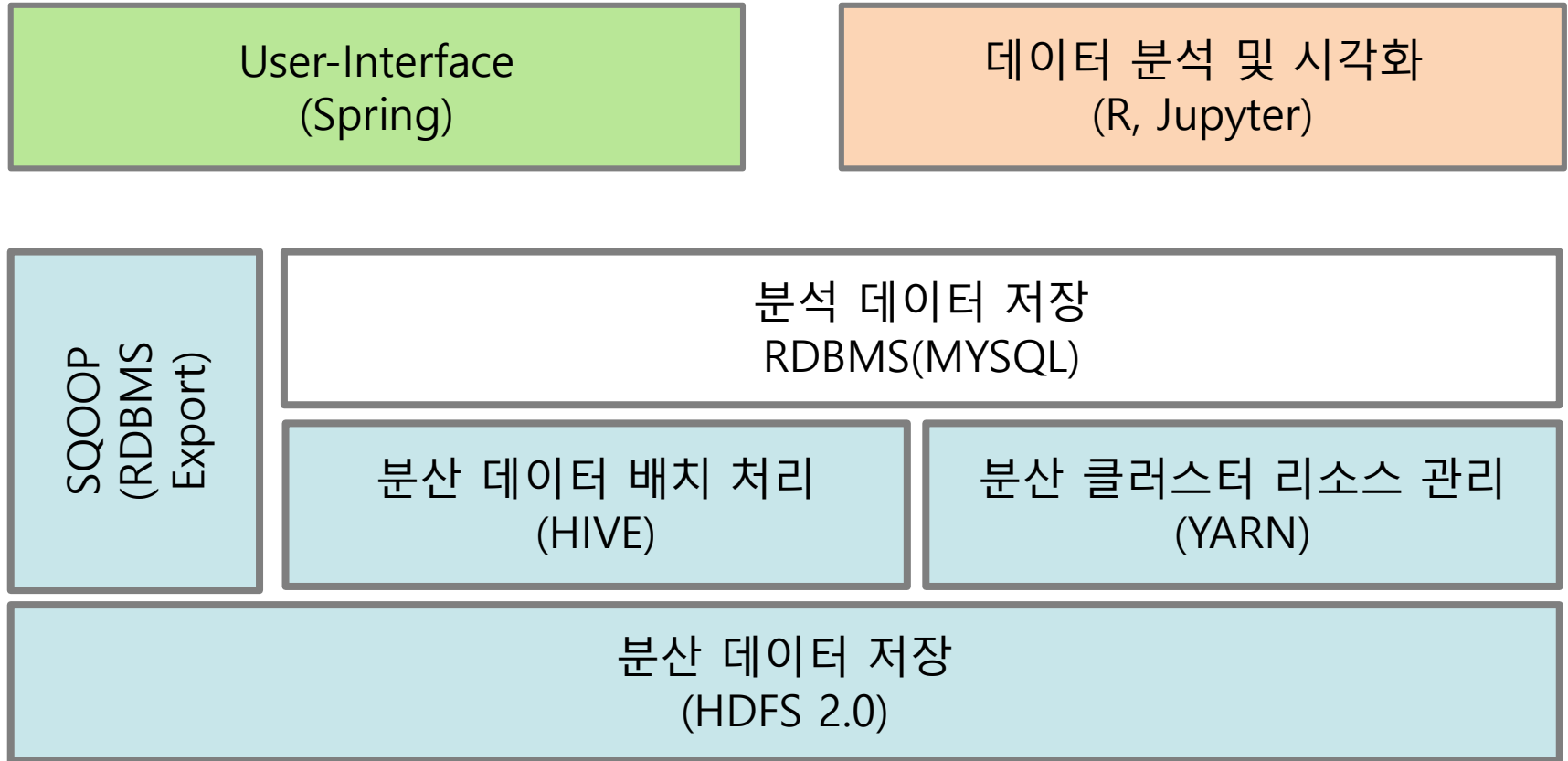


민원다발지역

기대효과

- 분석된 결과를 바탕으로 불법주정차민원에 대한 주차장, 시설물 관련 요인 제시
- 새로운 지역이 생겼을 때, 건물 용도만으로 어느 지역이 불법 주정차 민원이 많이 발생할 것인지 대략적인 예측을 할 수 있음
- 불법 주 정차 해결 정책을 주요 요인 분석 근거로 참고 될 수 있음

시스템 아키텍처



시스템 아키텍처(HDFS)



Name Node

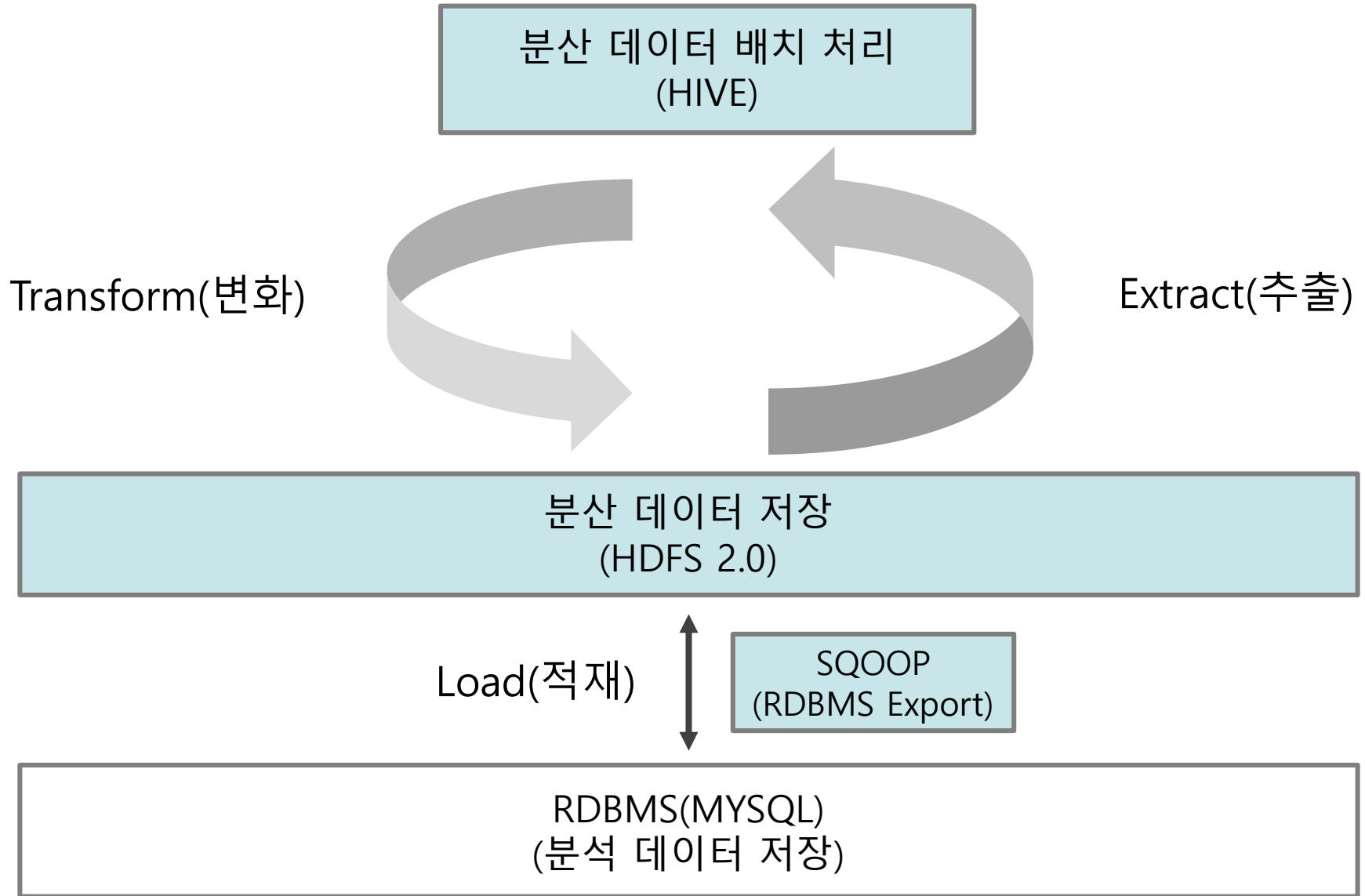
메타 데이터 정보 관리

분산 데이터 저장
(HDFS 2.0)

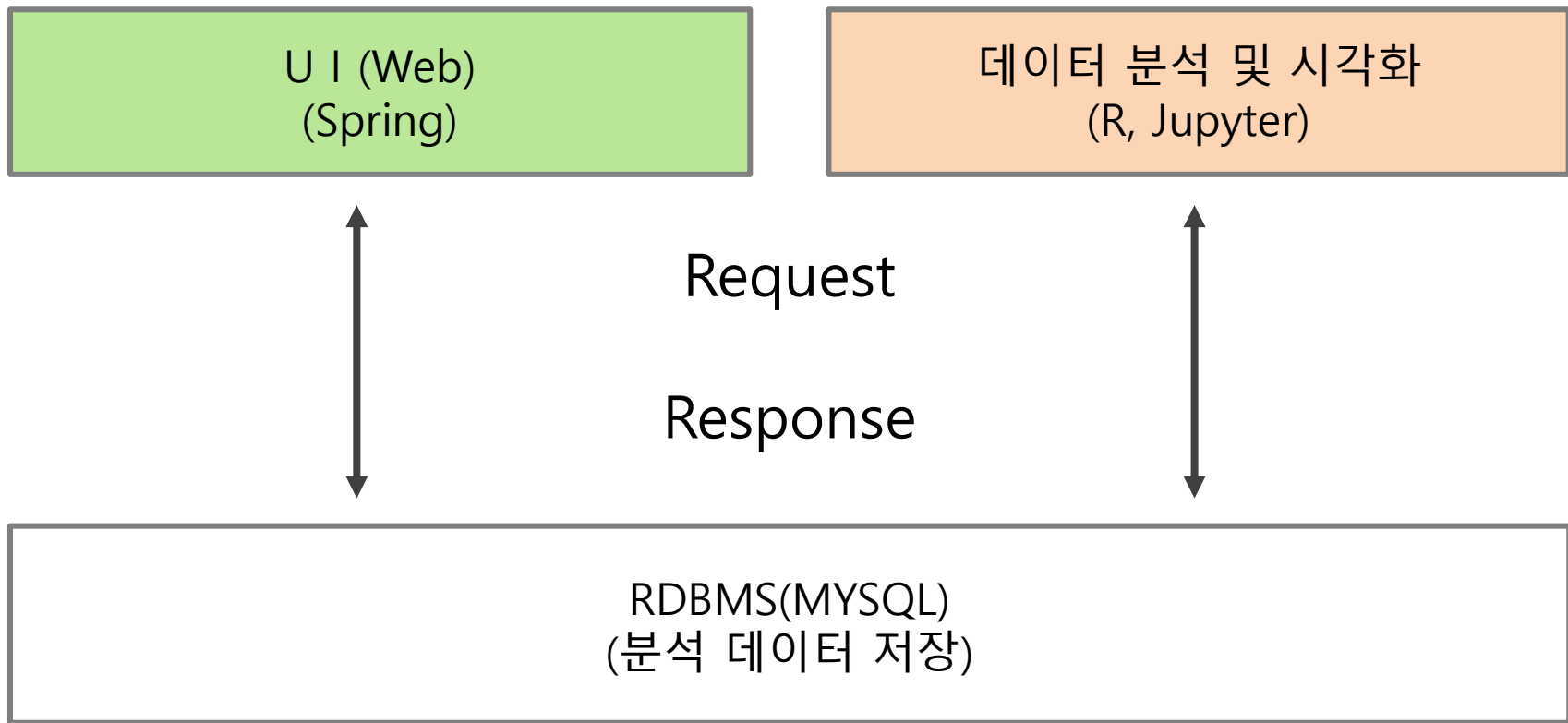


Data Node

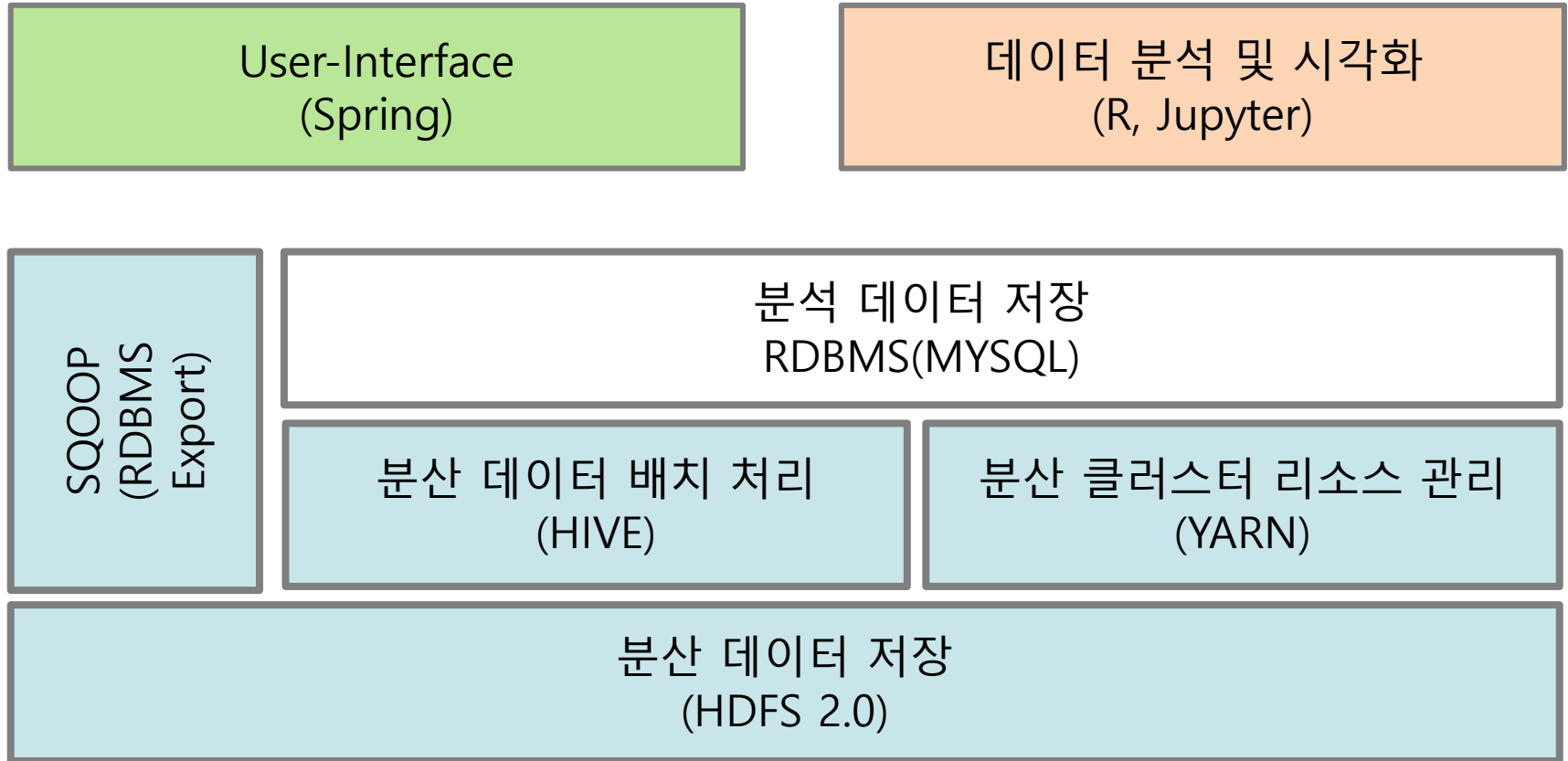
시스템 아키텍처(ETL)



시스템 아키텍처



시스템 아키텍처



시스템 아키텍처(HDFS)



Name Node

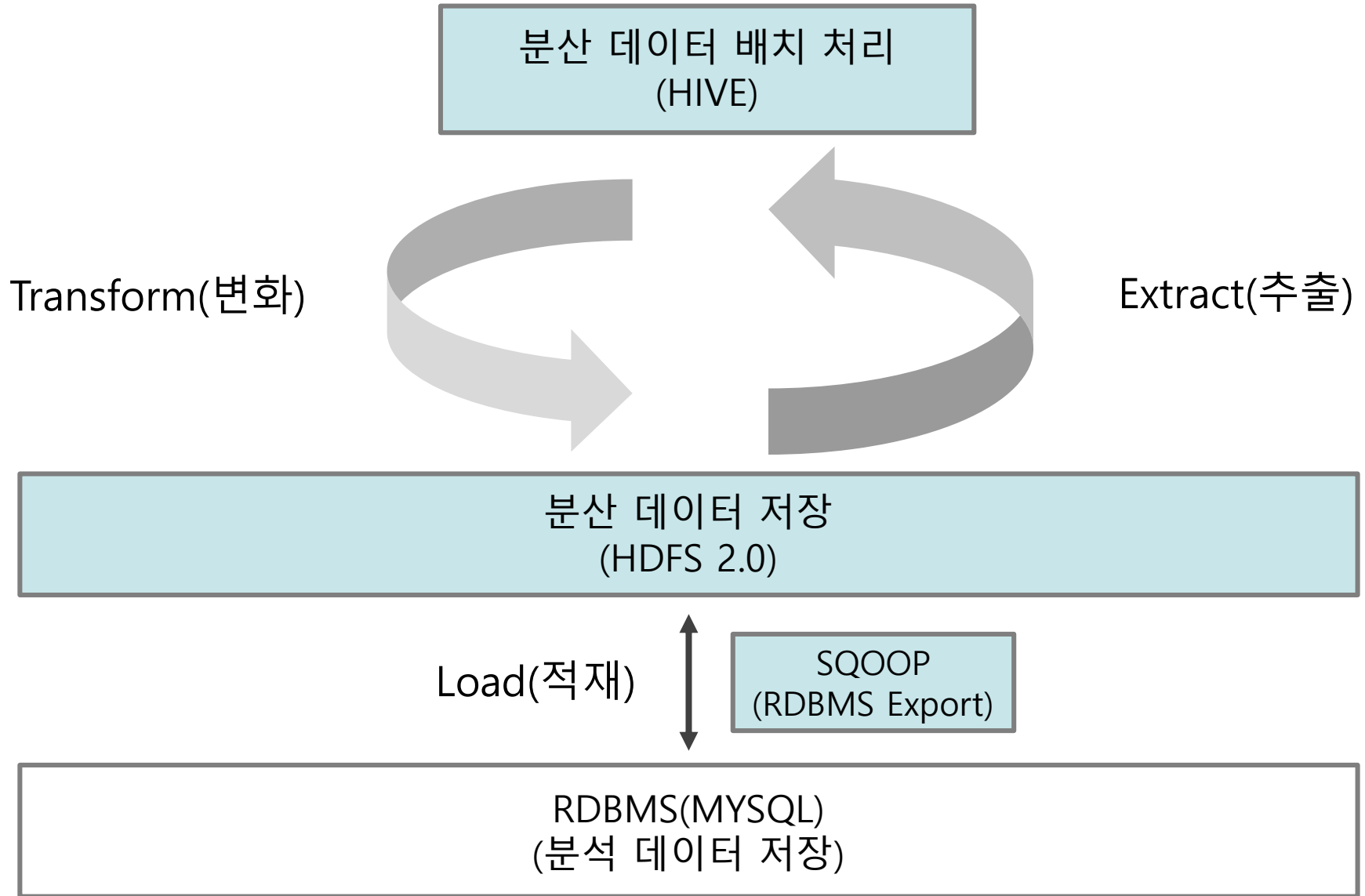
메타 데이터 정보 관리

분산 데이터 저장
(HDFS 2.0)

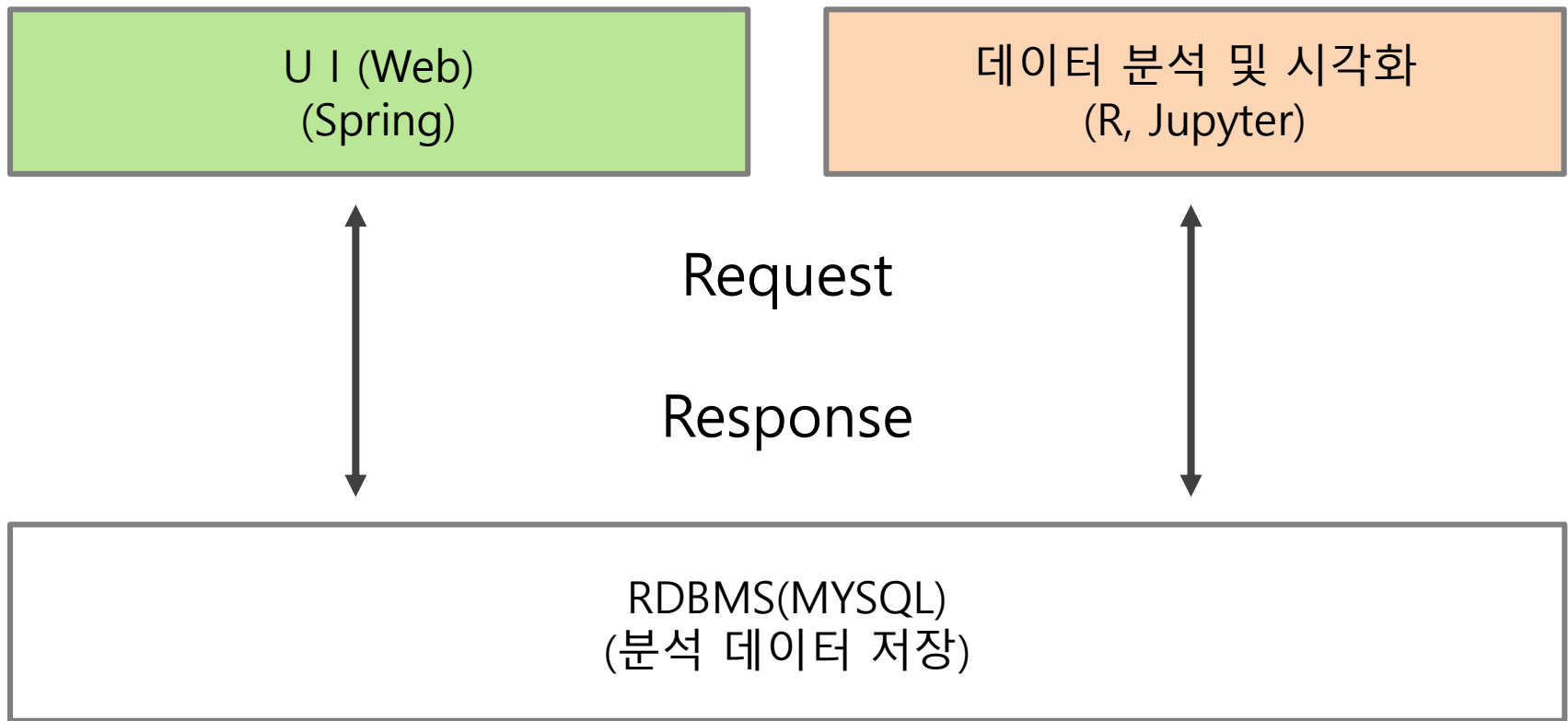


Data Node

시스템 아키텍처(ETL)



시스템 아키텍처





감사합니다.