

빅데이터 기술전문가 8기 2조

(주거환경 및 교육환경 분석을 통한 자녀 교육특구 모델 찾기)

멘토 : 이상훈강사
조원 : 김은희(조장)
김성표
노병희
노태상
전성종
최혁근

목 차

01 주제선정

1-1. 목적

1-2. 과제선정

02 추진일정

2-1. WBS

2-2. 모임 및 활동

2-3. 주별 주요 진행 사항

03 사용기술

3-1. 주요기술

3-2. 환경구축

04 과제수행

4-1. 데이터 수집

4-2. 데이터 정제, 통합

4-3. 데이터 분석

05 수행결과

5-1. 분석결과

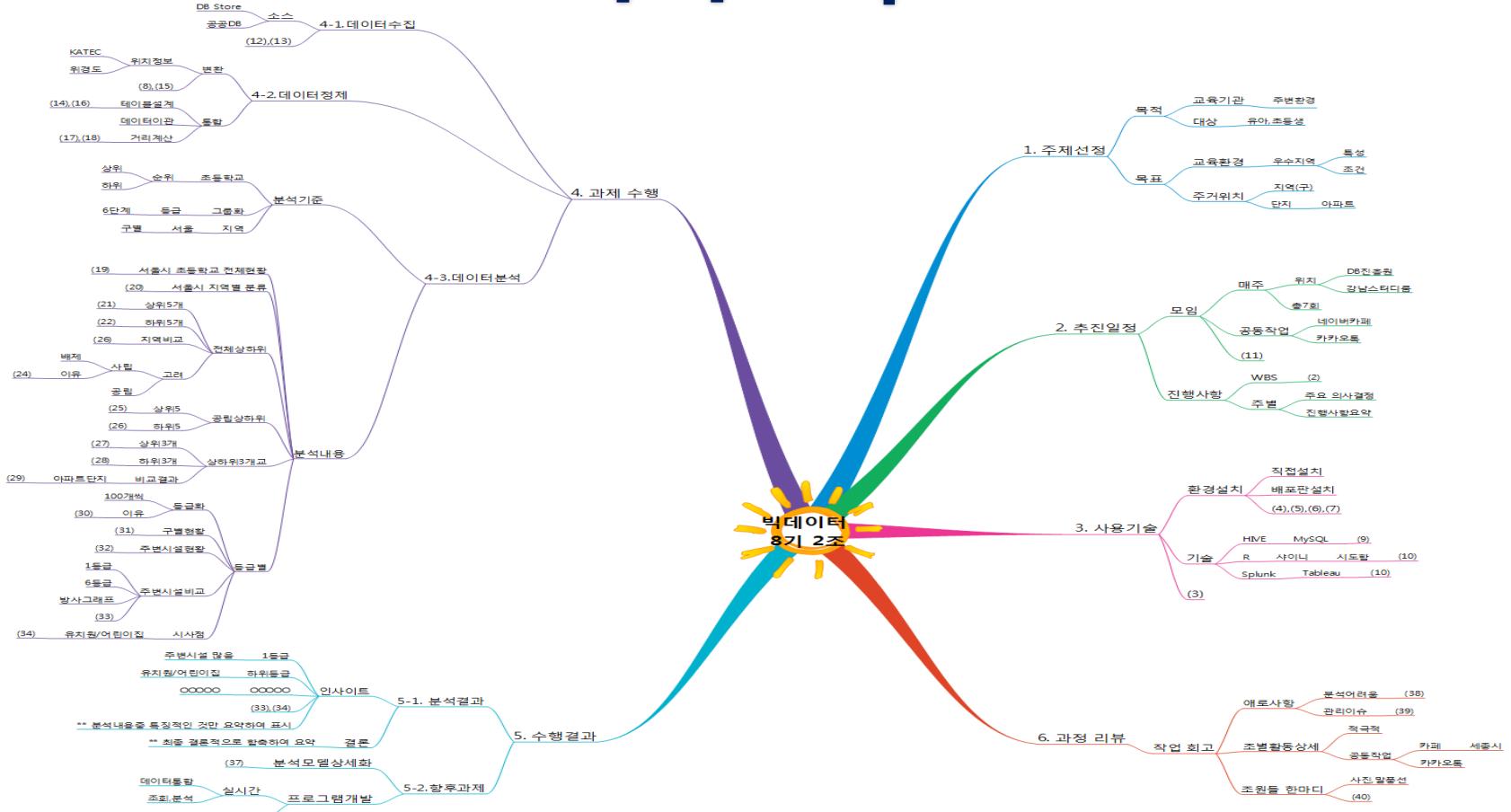
5-2. 향후 과제

06 과정리뷰

6-1. 분석애로사항

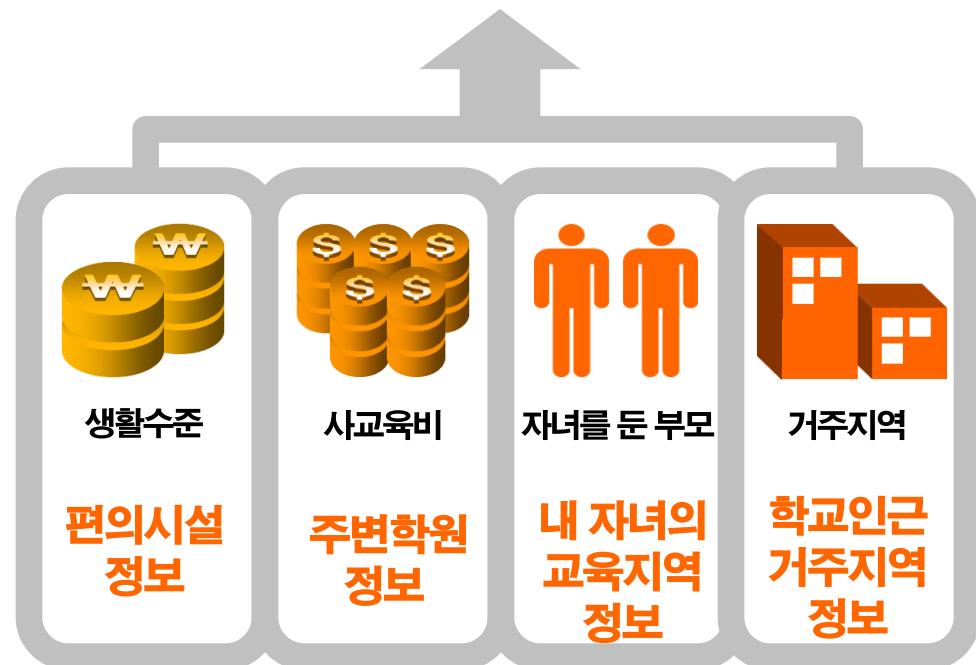
6-1. 조원별 활동 상세

목차 Map



유아 및 초등학생을 둔 부모의 교육환경

선택에 대한 잠재적 욕구 해소



“이번에 이사를 가야 하는데 우리 아이들 교육 환경이 좋은 곳으로 가고 싶은데 어디로 가야 할까?”

2조 조장님

1. 지리정보를 이용한 상권 분석
2. 학부모를 위한 자녀의 교육 특구 찾기
3. Flume을 이용한 대국민 인식관련 System 구축



2-1 WBS

2. 추진일정

KODB
한국데이터베이스진흥원

- 프로젝트 일정 : 9/1 ~ 10/24 (8주)

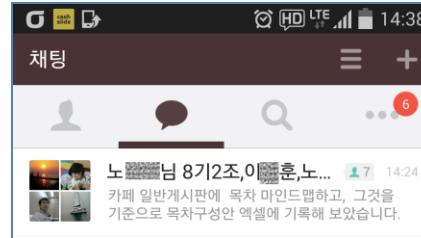
단계	세부업무	9월												10월																																
		W1			W2			W3			W4			W5			W6			W7			W8																							
1	2	3	4	5	8	9	10	11	12	15	16	17	18	19	22	23	24	25	26	29	30	1	2	3	6	7	8	9	10	13	14	15	16	17	20	21	22	23	24							
기획/분석	분석/정의/기획																																													
	필수요소 관련 기술조사																																													
	데이터 수집 검토																																													
설계	데이터 설계																																													
	기술 아키텍처 설계																																													
	분석 패턴 설계																																													
구현	데이터 수집																																													
	인프라 구축																																													
	데이터 저장																																													
	분석																																													
테스트	단위/통합 테스트																																													
	안정화																																													
적용	프로젝트 발표																																													

2-2 모임 및 활동

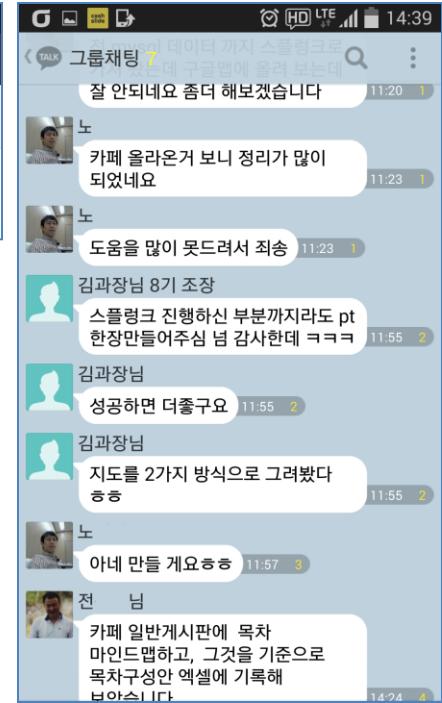
2. 추진일정 KODB 한국 데이터 베이스 진흥원

- 정기 모임 : 매주 화요일 저녁 7시 이후 DB진흥원
- 온라인 : Nave Café 빅데이터 기술전문가 8기 2조
- 기타 : 메신저(카카오톡), 유선 전화 등

네이버 카페



강남 스타디룸



카카오톡 대화방

2-3 주별 주요 행사

2. 추진일정

KODB
한국 데이터베이스 진흥원

- 데이터 수집
- 각 조원들 업무 분장
- 시스템 구성
- 테이블 구성 및 필요한 데이터 수집
- 분석에 관한 의견 수렴
- 시각화 작업에 관한 의견 수렴 등

0. 하동, Hive 설치 - 탐색

1. ~.db, 흐름/로그 재구성 : Data.

V - Table 정의서, 퀼록, + d 추가 Data 수집 결과.

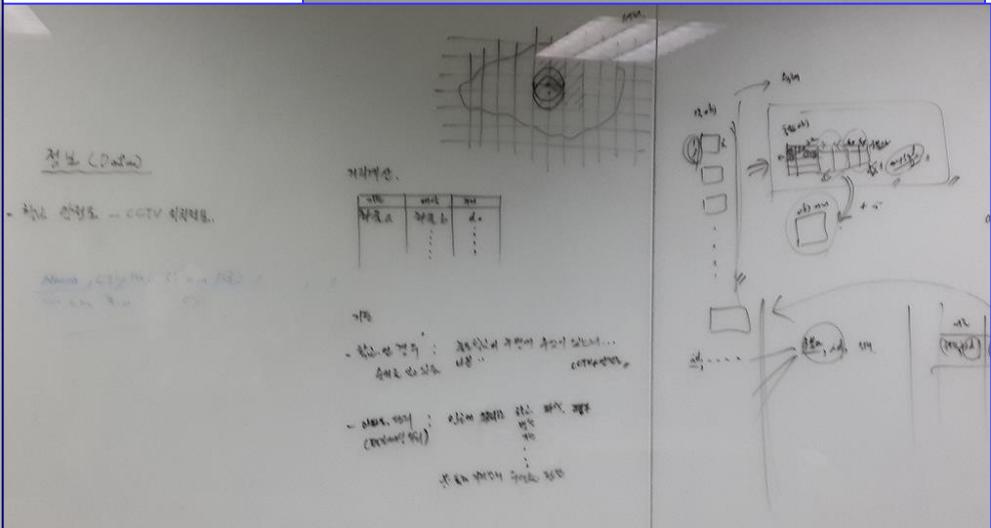
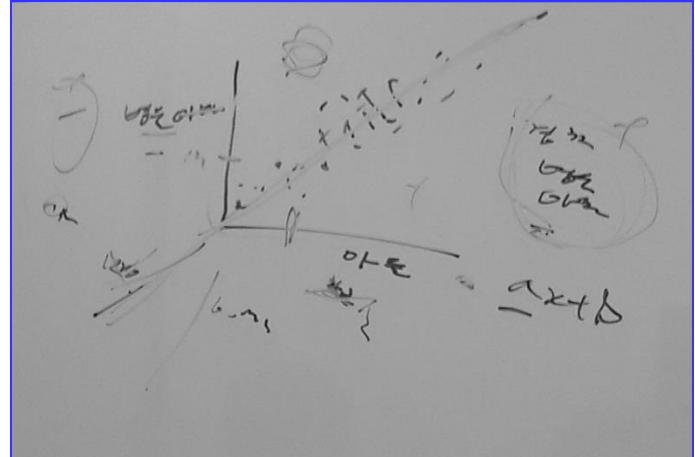
- Table Schema, 스크립트.

- Data 입력: CSV → Hadoop (Hive).

2. 분석.

- 아파트 단지 Point ← 주변 거리 산정

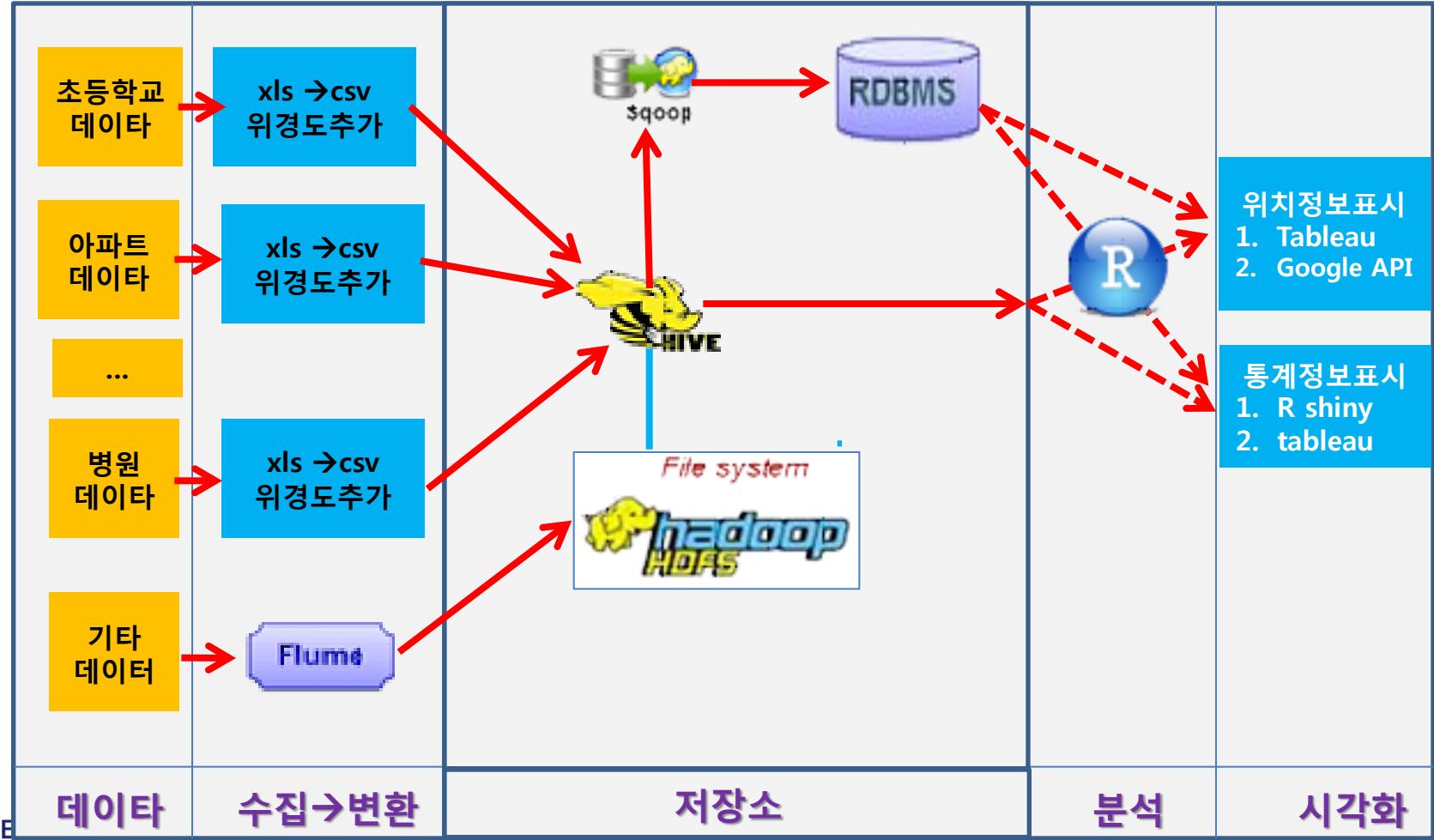
특정 단지별 대현 → 跑道분석
"도입 찾기"



3-1 주요 기술

2. 사용기술

KODB
한국데이터베이스진흥원



3-1 주요 기술

2. 사용기술

분석 툴	비고	시각화 툴	비고
	기본 빈도 및 통계 분석함		Shiny Package를 사용하여 시각화 결과물 추출 시도 함.
	Hive에서 쿼리 사용하여 결과 추출		Splunk를 이용하여 Map을 그리려 시도함
	Mysql 거리계산 함수를 이용하여 결과물을 추출		Tableau를 사용하여 기본 그래프와 Map 결과물을 추출함



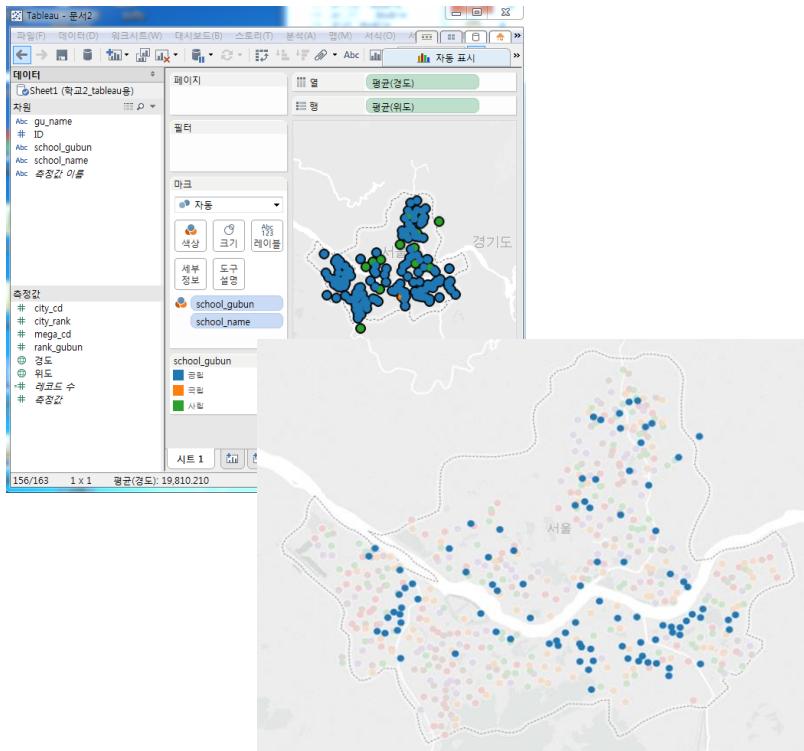
Map 결과물 추출

3-1 주요 기술

2. 사용기술

KO**DB**
한국 데이터 베이스 진흥원

- MAP 표시 비교



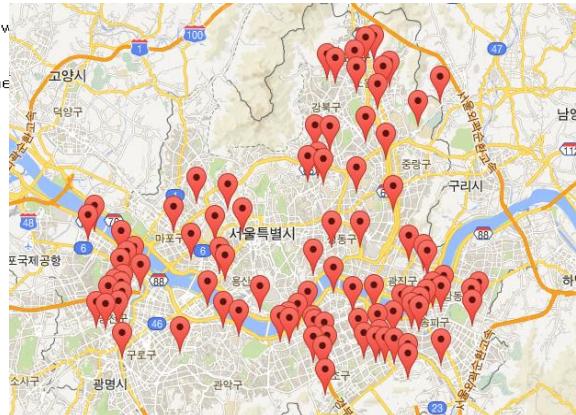
Google

```
var latlng91 = new google.maps.LatLng(37.522880387369, 126.93645238188);
var latlng92 = new google.maps.LatLng(37.5199041548971,
126.97589589122);
var latlng93 = new google.maps.LatLng(37.5426502600073,
126.97011924731);
var myOptions = {
  zoom: 11,
  center : latlng,
  mapTypeControl : google.maps.MapTypeControl.ROADMAP
}
var map = new google.maps.Map(document.getElementById
("map_canvas"), myOptions);

for(count=0;count<93;count++) {

  // eval("var marker = new google.maps.Marker({ position :
latlng"+count+", map : map, icon:'https://encrypted-tbn0.gstatic.com/images?
q=tbn:ANd9GcQBO90BkIULLnkoKurEdSCerCWp5Ymi_B66pFwFn1w9oTmFthM9i
5CabVU' }');");
  eval("var marker = new
map : map});");

  // var marker = ne
}));
```



3-2 환경 구축

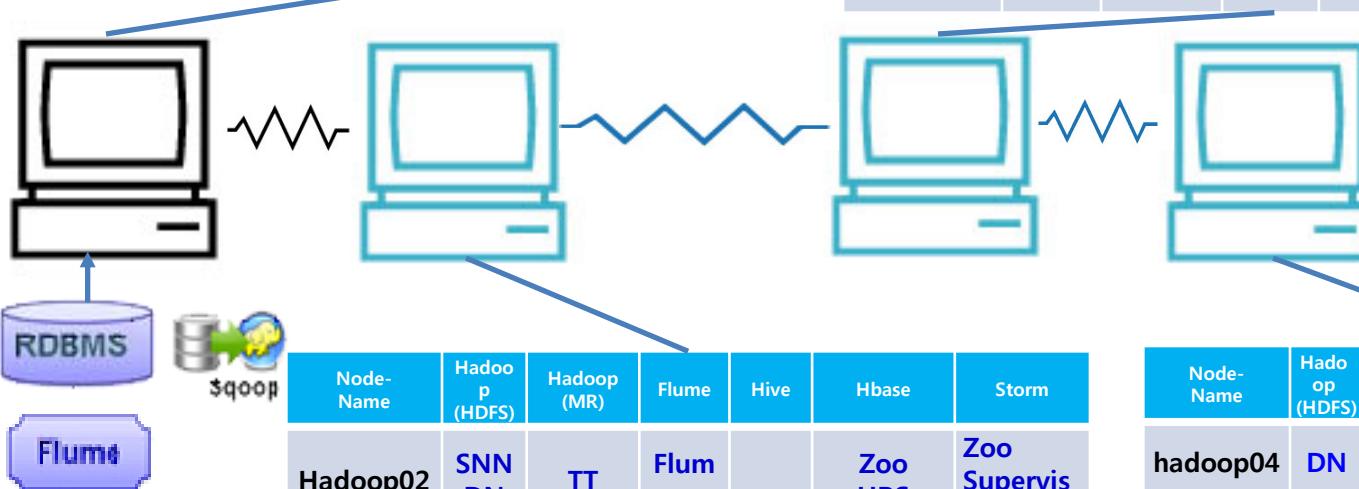
2. 사용기술

KODB
한국 데이터베이스 진흥원

- 하둡 v1.2.1 에코시스템 구성 및 분석 툴 설치
- Hive, Sqoop, Flume, Hbase, R, Mysql, JVM

Node-Name	Hadoop (HDFS)	Hadoop (MR)	Flume	Hive	Hbase	Storm
Hadoop01	NN	JT	Flume	HMS	Zoo HMS HRS	Zoo Nimbus Supervisor

Node-Name	Hadoop (HDFS)	Hadoop (MR)	Flume	Hive	Hbase	Storm
Hadoop03	DN	TT	Flume		Zoo HRS	Zoo Supervisor



Node-Name	Hadoop (HDFS)	Hadoop (MR)	Flume	Hive	Hbase	Storm
Hadoop02	SNN DN	TT	Flume		Zoo HRS	Zoo Supervisor

Node-Name	Hadoop (HDFS)	Hadoop (MR)	Flume	Hive	Hbase	Storm
hadoop04	DN	TT	Flume			

3-2 환경 구축

2. 사용기술

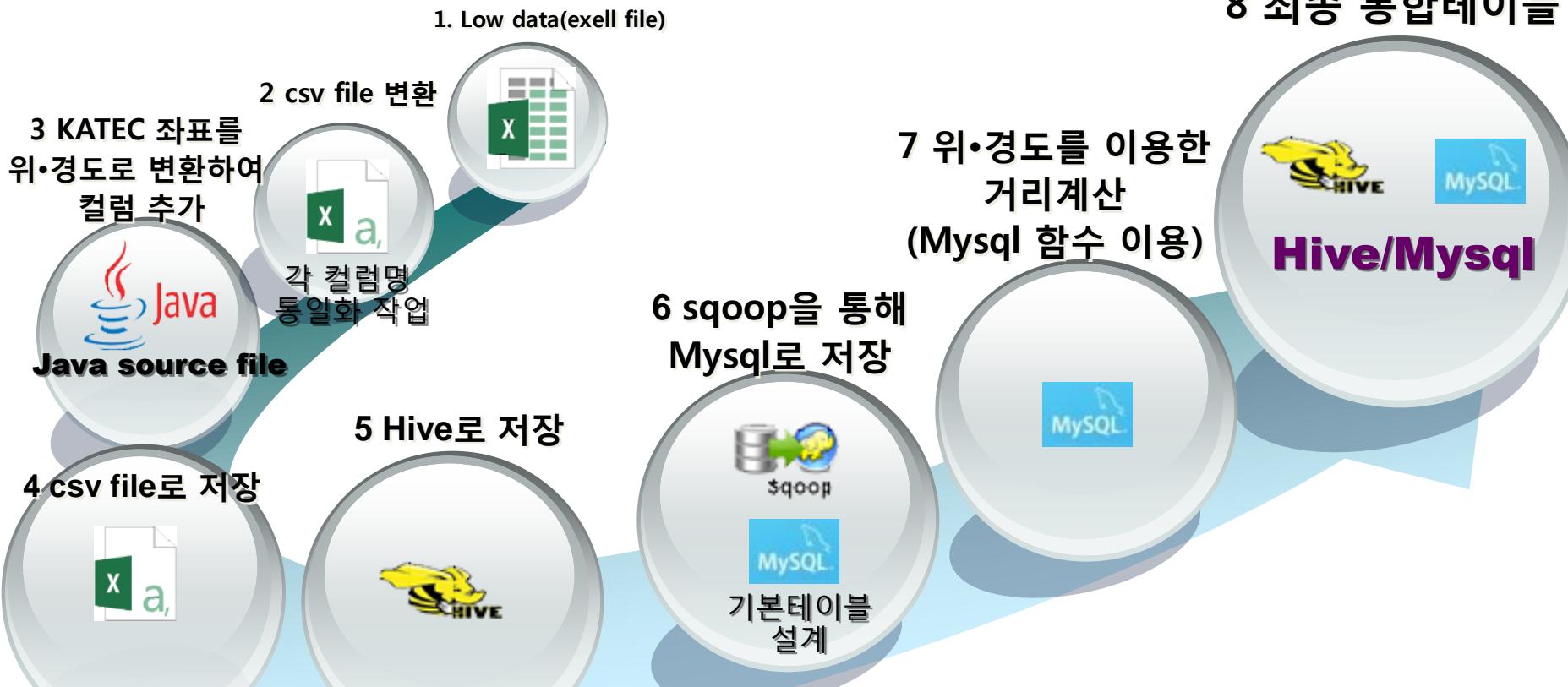


- [hadoop@hadoop01 ~]\$ jps
 - 11478 supervisor
 - 1685 NameNode
 - 26933 Jps
 - 16098 worker
 - 1804 DataNode
 - 1923 JobTracker
 - 10713 HMaster
 - 2744 RunJar
 - 11446 core
 - 10348 QuorumPeerMain
 - 11361 nimbus
 - 2046 TaskTracker
 - 10863 HRegionServer
 - [hadoop@hadoop02 ~]\$ jps
 - 5679 SecondaryNameNode
 - 5755 TaskTracker
 - 5264 Application
 - 16174 Jps
 - 5114 worker
 - 3709 QuorumPeerMain
 - 5619 DataNode
 - 4070 supervisor
 - 3843 HRegionServer
 - [hadoop@hadoop03 ~]\$ jps
 - 1146 TaskTracker
 - 3892 worker
 - 3981 Application
 - 2262 QuorumPeerMain
 - 2665 supervisor
 - 1063 DataNode
 - 2379 HRegionServer
 - 4518 Jps
 - [hadoop@hadoop04 ~]\$ jps
 - 8512 DataNode
 - 8596 TaskTracker
 - 2061 Jps
 - 3012 Application
-
- Name Node
 - Job Tracker
 - Hive
 - Hbase(Hbase Master Server)
 - Storm
 - Secondary Name Node
 - Data Node
 - Task Tracker
 - Hbase(Hbase Resource Server)
 - Flume

- DB Store 및 공공데이터, 구글링 등 포탈 사이트에서 검색함

DB Source 사이트명	사이트 주소	비고
	http://www.dbstore.or.kr/	
 API STORE	http://www.apistore.co.kr/	
 DATA GO KR 공공데이터포털	https://www.data.go.kr/	

4-2 데이터 정제, 통합



4-2 데이터 정제, 통합

4. 과제수행

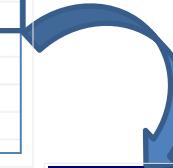
① DB Store에서 수집한 데이터

파일(F) 실행(E) 설정(Q) 도움말(H)

새로업로드 업로드열기 파일추가 업로드풀기 설치 보기

	X	미름	압축 크기	크기	압축률	종류
.../XsDB_HIGH_PC	XsDB_고등학교.POI_KA.shx	5,208	18,380	71%	SHX 파일	
	XsDB_고등학교.POI_KA.dbf	37,296	845,740	95%	DBF 파일	
	XsDB_고등학교.POI_KA.sbn	11,260	23,508	52%	SBN 파일	
	XsDB_고등학교.POI_KA.sbx	659	1,396	52%	SBX 파일	
	XsDB_고등학교.POI_KA.shp	20,364	64,080	68%	SHP 파일	

② Excel로 변환



A	B	C	D	E	F	G
MEGA_NM	MEGA_CD	CTY_NM	CTY_C_X_AXIS	Y_AXIS	SCHOOL_NM	
강원도	42	고성군	42820	437943	653550	대진고등학교
강원도	42	고성군	42820	439547	648632	거진중학교등학교
강원도	42	고성군	42820	440379	641711	고성고등학교
강원도	42	고성군	42820	447643	632589	동광산업과학교등학교
강원도	42	철원군	42780	350609	626843	김화고등학교
강원도	42	철원군	42780	350708	626796	김화공업고등학교
강원도	42	속초시	42210	451296	624839	속초고등학교
강원도	42	속초시	42210	449976	621982	속초여자고등학교
강원도	42	속초시	42210	450428	620388	설악고등학교
강원도	42	철원군	42780	339250	616232	신철원고등학교
강원도	42	인제군	42810	419111	613974	원통고등학교
강원도	42	양구군	42800	398504	610903	양구여자고등학교
강원도	42	양구군	42800	399454	610851	양구고등학교
강원도	42	양양군	42830	454027	607950	양양여자고등학교
강원도	42	양양군	42830	454300	607900	양양고등학교

③ 각 파일들의 컬럼명을 통일화 작업

④ Hive에 데이터 파일 생성

```
hive> show tables;
```

```
OK
apartment
bank
bus_stop
convenience_store
daycare_center
elementaryschool
estimation_income
highschool
hospital
kindergarten
land
living_population
middle_school
retail
worker_population
Time taken: 0.129 seconds, Fetched: 15 row(s)
```



⑤ Sqoop을 통해 MySQL 테이블 생성

```
[hadoop@hadoop01 ~]$ mysql -u root -p
```

```
Enter password: mysql
```

```
mysql> use project;
```

```
mysql> show tables;
```

+-----+
Tables_in_project
+-----+
apartment
bank
bus_stop
convenience_store
daycare_center
elementaryschool
estimation_income
highschool
hospital
kindergarten
land
living_population
middle_school
retail
worker_population
+-----+
15 rows in set (0.00 sec)

- 경도/위도 정보를 가진 두 지점 사이의 거리 계산 : Mysql 이용

(예) 학교와 은행 사이의 거리 계산

```
mysql> insert into elementary_distance
```

```
select a.city_name,a.school_name,"bank",b.city_name,b.bank_name
      ,b.x_axis_latitude,b.y_axis_longitude,1
,round(6371 *acos(cos(radians(a.x_axis_latitude))*cos(radians(b.x_axis_latitude))
*cos(radians(b.y_axis_longitude)-radians(a.y_axis_longitude))+sin(radians(a.x_axis_latitude)
*sin(radians(b.x_axis_latitude))),2) as dist
```

```
from elementaryschool a, bank b
here a.mega_name='서울특별시'
and b.mega_name='서울특별시';
```

```
Query OK, 1521825 rows affected (12.14 sec)
Records: 1521825  Duplicates: 0  Warnings: 0
```

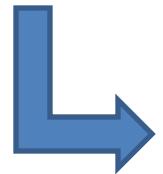
4-2 데이터 정제, 통합

4. 과제수행

학원현황 데이터 변환 – 주소 -> 위경도

- CSV 형태(주소)의 학원현황 데이터: 99,277건

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	교육자원점	학원명	설립자·성	등록일	등록상태	전화번호	위치	강사수	학원종류	분야구분	교습계열	교과과정	교습과목(정원)	교습기간	총교습시간	교습비	교재비
2	동부	도우리이별영재	1967.02.09	개원	02-969-50	서울특별시	1	평생직업고직업기술	산업용융기기미용	이용기술	20개월	01월	3,600	250,000	0		
3	동부	대성회계술오병기	1968.12.27	개원	02-966-84	서울특별시	4	평생직업고직업기술	산업서비스전산회계	전산세무2	25개월	01월	5,400	513,000	0		
4	동부	대성회계술오병기	1968.12.27	개원	02-966-84	서울특별시	4	평생직업고직업기술	산업서비스전산회계	기초회계	17개월	01월	5,400	513,000	0		
5	동부	대성회계술오병기	1968.12.27	개원	02-966-84	서울특별시	4	평생직업고직업기술	산업서비스전산회계	전산세무2	25개월	01월	9,600	850,000	0		
6	동부	대성회계술오병기	1968.12.27	개원	02-966-84	서울특별시	4	평생직업고직업기술	산업서비스전산회계	회계원리2	17개월	01월	8,400	730,000	0		
7	동부	대성회계술오병기	1968.12.27	개원	02-966-84	서울특별시	4	평생직업고직업기술	산업서비스전산회계	법인세이론	25개월	01월	7,200	810,000	0		
8	동부	대성회계술오병기	1968.12.27	개원	02-966-84	서울특별시	4	평생직업고직업기술	산업서비스전산회계	회계관리1	25개월	01월	9,000	855,000	0		
9	동부	대성회계술오병기	1968.12.27	개원	02-966-84	서울특별시	4	평생직업고직업기술	산업서비스전산회계	전산회계1	17개월	01월	3,600	400,000	0		
10	동부	대성회계술오병기	1968.12.27	개원	02-966-84	서울특별시	4	평생직업고직업기술	산업서비스전산회계	전산세무2	25개월	01월	3,600	400,000	0		
11	동부	대성회계술오병기	1968.12.27	개원	02-966-84	서울특별시	4	평생직업고직업기술	산업서비스전산회계	법인기업k	25개월	01월	3,600	420,000	0		
12	동부	대성회계술오병기	1968.12.27	개원	02-966-84	서울특별시	4	평생직업고직업기술	산업서비스전산회계	회계및전산	25개월	01월	3,600	420,000	0		
13	동부	대성회계술오병기	1968.12.27	개원	02-966-84	서울특별시	4	평생직업고직업기술	산업서비스전산회계	전산회계1	17개월	01월	8,400	790,000	0		
14	동부	대성회계술오병기	1968.12.27	개원	02-966-84	서울특별시	4	평생직업고직업기술	산업서비스전산회계	회계정보학	17개월	01월	1,800	290,000	0		



- Geocoder 를 이용하여 위경도 변환

Geocoder-Xr with 다음, 네이버, 구글

No	교육지...	학원명	설립자...	등록일	등록상태	전화번호	위치	강사수	학원종류	분야구분	교습
1	동부	도우리...	박영배	1967.0...	개원	02-96...	서울특...	1	평생직...	직업기술	산...
2	동부	대성회...	오병기	1968.1...	개원	02-96...	서울특...	4	평생직...	직업기술	산...
3	동부	대성회...	오병기	1968.1...	개원	02-96...	서울특...	4	평생직...	직업기술	산...
4	동부	대성회...	오병기	1968.1...	개원	02-96...	서울특...	4	평생직...	직업기술	산...
5	동부	대성회...	오병기	1968.1...	개원	02-96...	서울특...	4	평생직...	직업기술	산...
6	동부	대성회...	오병기	1968.1...	개원	02-96...	서울특...	4	평생직...	직업기술	산...
7	동부	대성회...	오병기	1968.1...	개원	02-96...	서울특...	4	평생직...	직업기술	산...
8	동부	대성회...	오병기	1968.1...	개원	02-96...	서울특...	4	평생직...	직업기술	산...
9	동부	대성회...	오병기	1968.1...	개원	02-96...	서울특...	4	평생직...	직업기술	산...
10	동부	대성회...	오병기	1968.1...	개원	02-96...	서울특...	4	평생직...	직업기술	산...
11	동부	대성회...	오병기	1968.1...	개원	02-96...	서울특...	4	평생직...	직업기술	산...
12	동부	대성회...	오병기	1968.1...	개원	02-96...	서울특...	4	평생직...	직업기술	산...
13	동부	대성회...	오병기	1968.1...	개원	02-96...	서울특...	4	평생직...	직업기술	산...
14	동부	대성회...	오병기	1968.1...	개원	02-96...	서울특...	4	평생직...	직업기술	산...
15	동부	대성회...	오병기	1968.1...	개원	02-96...	서울특...	4	평생직...	직업기술	산...
16	동부	대성회...	오병기	1968.1...	개원	02-96...	서울특...	4	평생직...	직업기술	산...
17	동부	대성회...	오병기	1968.1...	개원	02-96...	서울특...	4	평생직...	직업기술	산...
18	동부	대성회...	오병기	1968.1...	개원	02-96...	서울특...	4	평생직...	직업기술	산...
19	동부	국제전...	이종운	1968.0...	개원	02-96...	서울특...	1	평생직...	직업기술	산...
20	동부	국제전...	이종운	1968.0...	개원	02-96...	서울특...	1	평생직...	직업기술	산...
21	동부	국제전...	이종운	1968.0...	개원	02-96...	서울특...	1	평생직...	직업기술	산...
22	동부	송산서...	신정균	1972.0...	개원	02-22...	서울특...	1	학교교...	기타(대)	기타
23	동부	송산서...	신정균	1972.0...	개원	02-22...	서울특...	1	학교교...	기타(대)	기타
24	동부	영란피...	김증란	1972.0...	개원	02-96...	서울특...	1	학교교...	예능(대)	예능
25	동부	한독자...	(주)메...	1972.0...	개원	02-92...	서울특...	3	평생직...	직업기술	산...
26	동부	한독자...	(주)메...	1972.0...	개원	02-92...	서울특...	3	평생직...	직업기술	산...
27	도내	한도인...	(주)인...	1977.0...	개원	02-02...	서울특...	2	폐쇄자...	지어기수	산...

결과 SHP 파일 C:\Users\Seunghyun\OneDrive\Desktop\제이터\result1.shp

OpenAPI Key 관리 ... Progress 시작 닫기

- [1단계] 기초테이블 생성** : 학교, 아파트 등의 이름, KATEC좌표, 지역구명 등 수집

No	테이블명	한글 테이블명	레코드 갯수
1	highschool_rank	고등학교 순위	237
2	elementaryschool_rank	초등학교 순위	591
3	highschool	고등학교 위치 정보	2,286
4	retail	리테일	3,143
5	bus_stop	버스정류장 위치 정보	45,895
6	hospital	병원 위치 정보	23,995
7	apartment	아파트 위치 정보	61,029
8	daycare_center	어린이집 위치 정보	39,309
9	kindergarten	유치원 위치 정보	6,983
10	bank	은행 위치 정보	7,273
11	living_population	시, 군, 구별 인구 정보	959,660
12	middle_school	중학교 위치 정보	3,205
13	worker_population	직장인 정보	421,515
14	elementaryschool	초등학교 위치 정보	6,296
15	estimation_income	추정 소득분위	959,661
16	land	토지 정보	400,997
17	convenience_store	편의점 위치 정보	10,823

- [2단계] 통합테이블 구성 : 학교 기준 주변 시설의 거리를 계산하여 데이터화

학교지역명, 학교명/주변시설구분자, 주변시설명, 주변시설지역명, 주변시설경도, 주변시설위도/거리

(1) 테이블 레이아웃

```
mysql> show columns from elementary_distance;
```

Field	Type	Null	Key	Default	Extra
school_city_name	varchar(64)	YES		NULL	
school_name	varchar(64)	YES		NULL	
surd_fact_gubun	varchar(64)	YES		NULL	
surd_fact_city_name	varchar(64)	YES		NULL	
surd_fact_name	varchar(64)	YES		NULL	
x_axis_latitude	double	YES		NULL	
y_axis_longitude	double	YES		NULL	
surd_fact_cnt	double	YES		NULL	
dist	double	YES		NULL	

9 rows in set (0.00 sec)

(2) 주변시설별 건수

```
mysql> select target_gubun, count(*) from `
```

target_gubun	count(*)
bank	1521825
bus_stop	5625138
convenience_store	1510596
daycare_center	3580278
highschool	185574
kindergarten	499395
middle_school	223398
retail	205668

8 rows in set (23.04 sec)

mysql>

- [2단계] 통합테이블별 건수

No	테이블명	한글 테이블명	레코드 갯수
1	elementary_distance	학교 ↔ 은행, 정류장, 편의점, 유치원, 어린이집, 고등학교, 상점 과의 거리	13,351,872
2	elementary_distance_apt	학교 ↔ 아파트 의 거리	5,671,827
3	e_school_dist_income	학교 ↔ 주변지역 인구의 평균소득 별 거리	17,527,878
4	e_school_dist_land	학교 ↔ 주변지역 토지 금액별 거리	10,421,694
5	e_school_dist_living_pop	학교 ↔ 주변지역 주거 인구별 거리	25,637,843
6	e_school_dist_worker_pop	학교 ↔ 주변지역 직장 인구별 거리	17,527,483
			90,138,597

4-3 데이터 분석

4. 과제수행

- 분석을 위하여 포탈 검색하여 초등학교 (평가) 기준으로 등급이 기제된 데이터 확보함
- 서울 초등학교 591개를 1~100위 학교 101~200위 학교 등으로 등급으로 나누어 그룹화함
- 생활권과 연계하여 중앙지역과 동, 서, 남, 북으로 나눔
- Hive와 Mysql에 동일한 환경으로 만들어 데이터 분석함

지역	구1	구2	구3	구4	구5	구6	구7	구8
중앙지역	종로구	용산구	중구					
강동지역	광진구	동대문구	중랑구	성동구	강동구	송파구		
강서지역	은평구	서대문구	마포구	금천구	강서구	양천구	영등포구	구로구
강남지역	관악구	동작구	강남구	서초구				
강북지역	성북구	노원구	강북구	도봉구				



4-3 데이터 분석

4. 과제수행

- 서울 초등학교 개수는 총 591개이며 노원구가 42개로 가장 많음
- 서울시내 구별 초등학교 평균 개수는 23.6개로 나타남
- 초등학교가 가장 많은 노원구와 가장 적은 중구의 개수 차이는 30개로 나타남

서울시 구별 초등학교 수



4-3 데이터 분석

서울시 전체 초등학교 상위 5 KODB
한국 데이터베이스 진흥원

- 서울시내 초등학교 상위 5위 학교가 있는 구는 노원구, 동작구, 서초구, 동대문구, 강북구에 각각 1개씩 위치함
- 강북지역 2개 학교, 강남지역 2개 학교, 중앙지역 1개 학교가 있음

구명	구순위	시순위	학교명
노원구	1	1	1 청원초등학교
동작구	1	2	2 중앙대학교사범대학부속초등학교
서초구	1	3	3 계성초등학교
동대문구	1	4	4 서울삼육초등학교
강북구	1	5	5 영훈초등학교



4-3 데이터 분석

서울시 전체 초등학교 하위 5

KODB
한국데이터베이스진흥원

- 서울시내 초등학교 하위 5위 학교가 있는 구는 마포, 관악구, 중랑구, 구로구, 은평구에 각각 1개씩 위치함
- 강서지역 3개 학교, 강남지역 1개 학교, 강동지역에 1개 학교가 있음

구명	구순위	시순위	학교명
마포구	22	587	서울한서초등학교
관악구	22	588	서울원당초등학교
중랑구	22	589	서울면북초등학교
구로구	23	590	서울영일초등학교
은평구	31	591	알로이시오초등학교

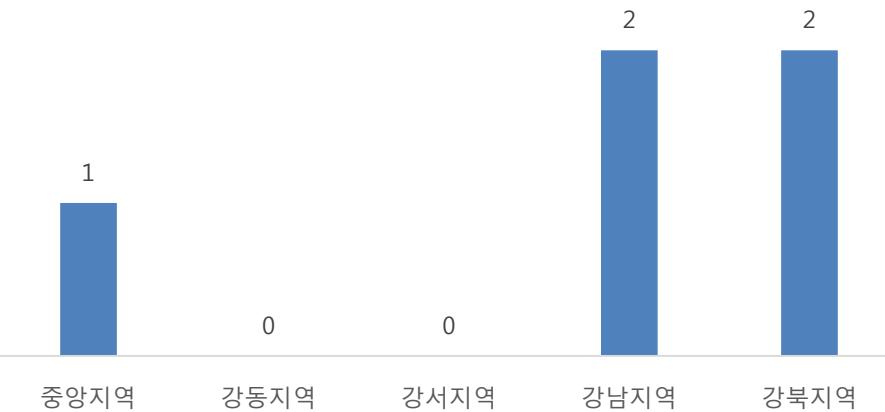


4-3 데이터 분석

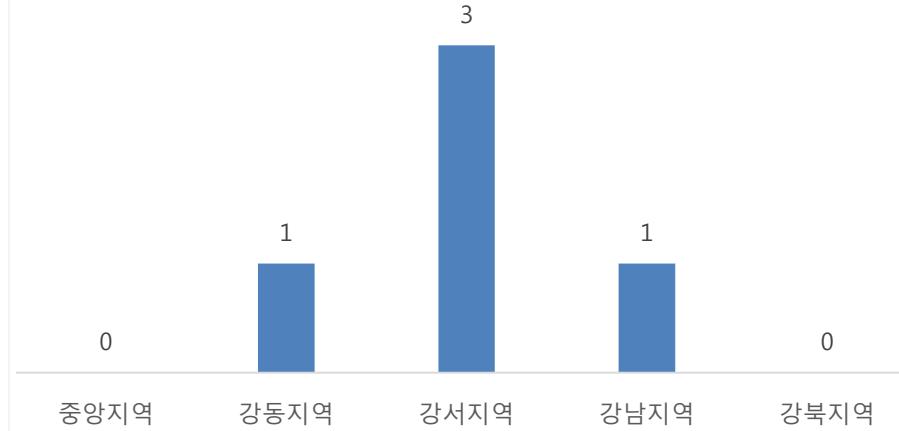
초등학교 상위 5/하위 5 지역비교

- 서울 초등학교 상위 5개 학교와 하위 5개 학교를 비교함
- 상위 5순위 내에 있는 초등학교가 강북지역 2개, 강남지역 2개, 중앙지역에 1개학교가 있음
- 중앙지역과 강북지역은 상위 학교는 포함되어 있지만 하위 학교는 포함 안됨
- 강서지역은 상위 학교가 없는 반면 하위 학교가 지역 내에 3개 학교가 있음

서울 초등학교 상위 5개 학교 지역위치



서울 초등학교 하위 5개 학교 지역위치



4-3 데이터 분석

서울시 공립 초등학교 상위 5 KODB
한국데이터베이스진흥원

- 공립 초등학교 상위 5위 학교가 있는 구는 강남구 2개, 송파구 2개, 서초구 2개가 위치함
- 공립 초등학교 상위 5개 학교는 모두 강남에 위치함

구명	구순위	시순위	학교명
서초구	2	9	서울원명초등학교
송파구	1	15	서울잠실초등학교
송파구	2	16	서울오륜초등학교
강남구	1	18	서울대곡초등학교
강북구	2	20	서울대모초등학교



4-3 데이터 분석

서울시 공립 초등학교 하위 5 KODB
한국 데이터 베이스 진흥원

- 서울시내 초등학교 하위 5위 학교가 있는 구는 마포, 관악구, 중랑구, 구로구, 은평구에 각각 1개씩 위치함
- 강서에 2개 학교, 강북에 3개 학교가 있음

구명	구순위	시순위	학교명
마포구	22	587	서울한서초등학교
관악구	22	588	서울원당초등학교
중랑구	22	589	서울면북초등학교
구로구	23	590	서울영일초등학교
은평구	31	591	알로이시오초등학교

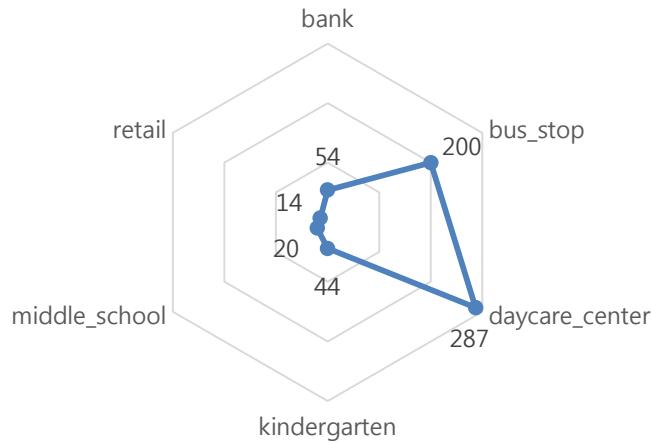


4-3 데이터 분석

전체 초등학교 상위 1

- 서울시내 전체 초등학교 1번째 청원 초등학교,
2Km 이내 주변에 대규모 아파트 단지들이 다수 있음

서울 청원 초등학교



4-3 데이터 분석

서울 청원 초등학교 주변 지도

KODB
한국데이터베이스진흥원

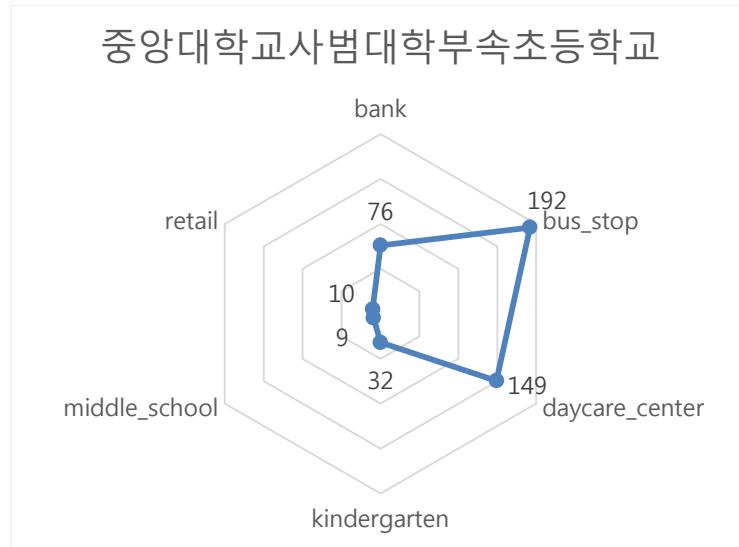


4-3 데이터 분석

서울시 전체 초등학교 상위 2

KODB
한국데이터베이스진흥원

- 서울시내 전체 초등학교 상위 2번째 중앙대학교 사범대학 부속 초등학교, 2Km 이내 주변에 대규모 아파트 단지들이 다수 있음



4-3 데이터 분석 서울 중대사범 부속 초등학교 주변 지도

KODB
한국 데이터 베이스 진흥원



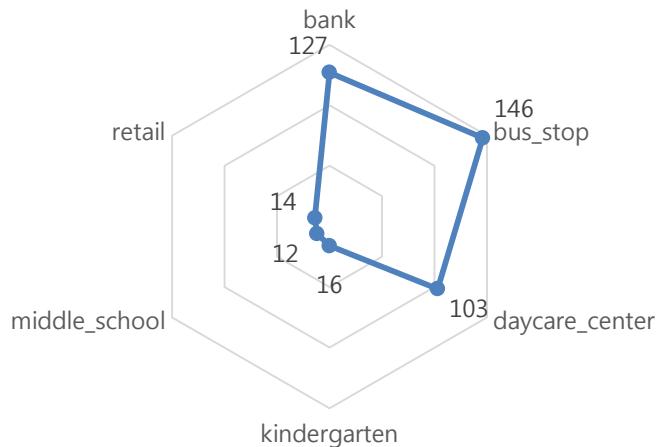
4-3 데이터 분석

서울시 전체 초등학교 상위 3

KODB
한국데이터베이스진흥원

- 서울시내 전체 초등학교 상위 3번째 계성 초등학교.
2Km 이내 주변에 대규모 아파트 단지들이 다수 있음

서울계성초등학교



4-3 데이터 분석

서울 계성 초등학교 주변 지도

KODB
한국 데이터 베이스 진흥원

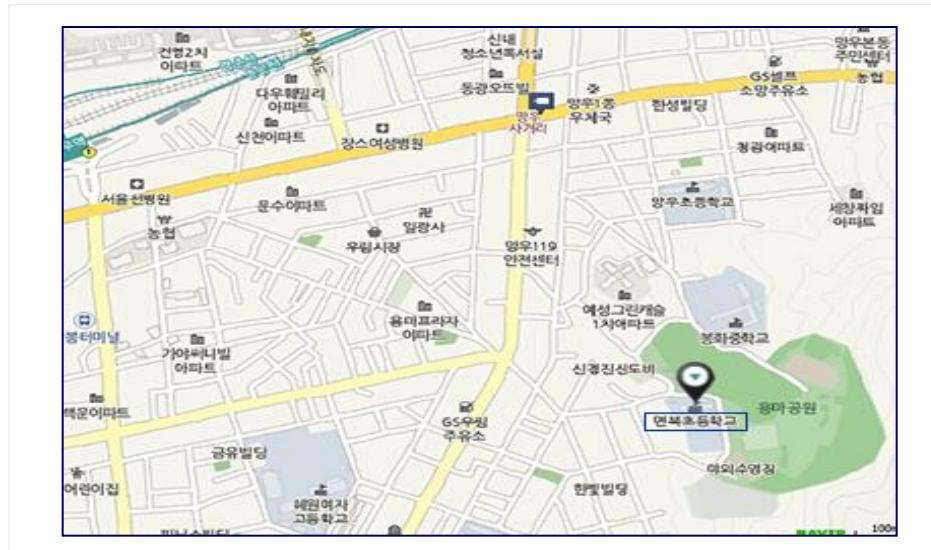
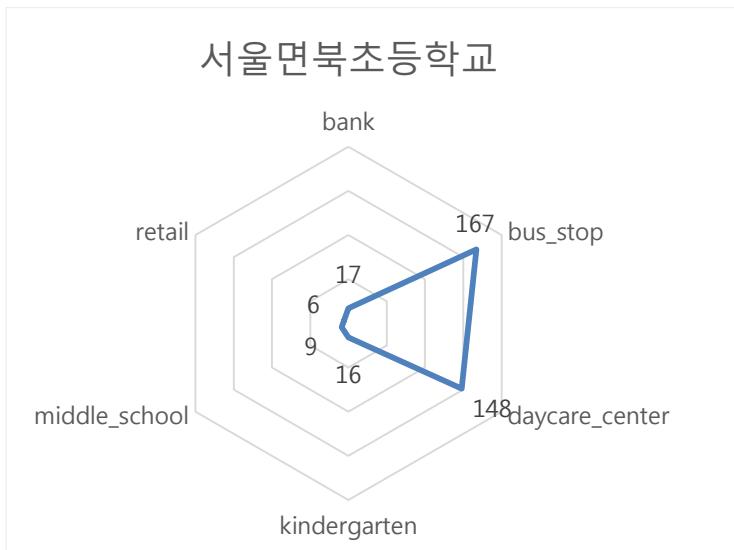


4-3 데이터 분석

서울시 전체 초등학교 하위 1

KODB
한국데이터베이스진흥원

- 서울시내 전체 초등학교 하위 면북 초등학교,
2Km 이내 주변에 대규모 아파트 단지들이 없음



4-3 데이터 분석

서울 면북 초등학교 주변 지도

KODB
한국 데이터 베이스 진흥원

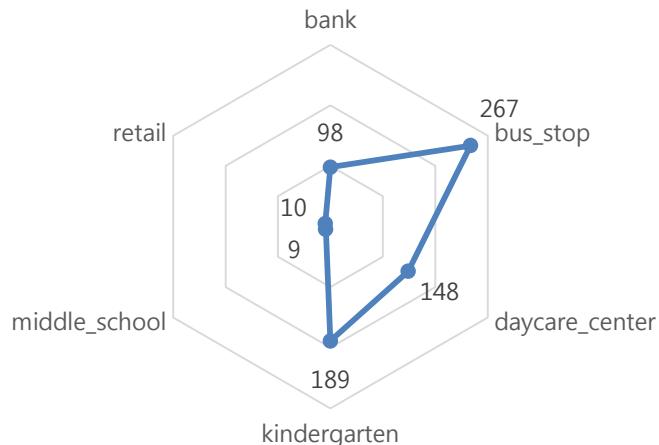


4-3 데이터 분석

서울시 전체 초등학교 하위 2

- 서울시내 전체 초등학교 하위 2번째 면북 초등학교,
2Km 이내 주변에 대규모 아파트 단지들이 없음

서울영일초등학교



4-3 데이터 분석

서울 영일 초등학교 주변 지도

KODB
한국 데이터 베이스 진흥원

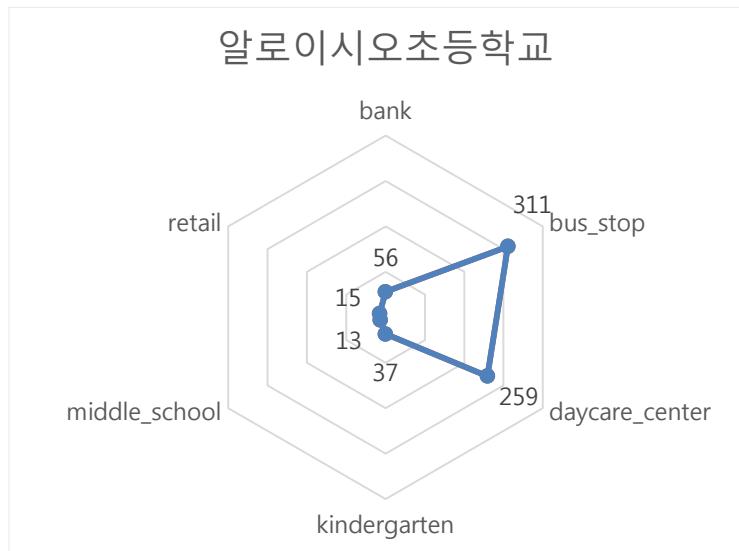


4-3 데이터 분석

서울시 전체 초등학교 하위 3

KODB
한국 데이터베이스 진흥원

- 서울시내 전체 초등학교 하위 3번째 알로이시오 초등학교,
2Km 이내 주변에 대규모 아파트 단지들이 없음



4-3 데이터 분석

서울 알로이시오 초등학교 주변 지도

KODB
한국데이터베이스진흥원



4-3 데이터 분석

서울시 전체 초등학교 상위 3/하위 3 비교

KODB
한국데이터베이스진흥원

- 서울 초등학교 상위 3개 학교와 하위 3개 학교를 비교함
- 주변 시설물등 환경적인 부분에는 큰 영향이 없음.
- 다만, 상위 학교들 주변에는 대규모 아파트 단지들이 위치함.
- 반면, 하위 학교들 주변에는 대규모 아파트 단지가 멀리 있음.
- 학교 주변 시설물 보다는 주변 아파트 단지 즉, 주거 환경이 학교수준에 영향을 줌.



도보 2Km 이내



4-3 데이터 분석

등급분류의 필요성



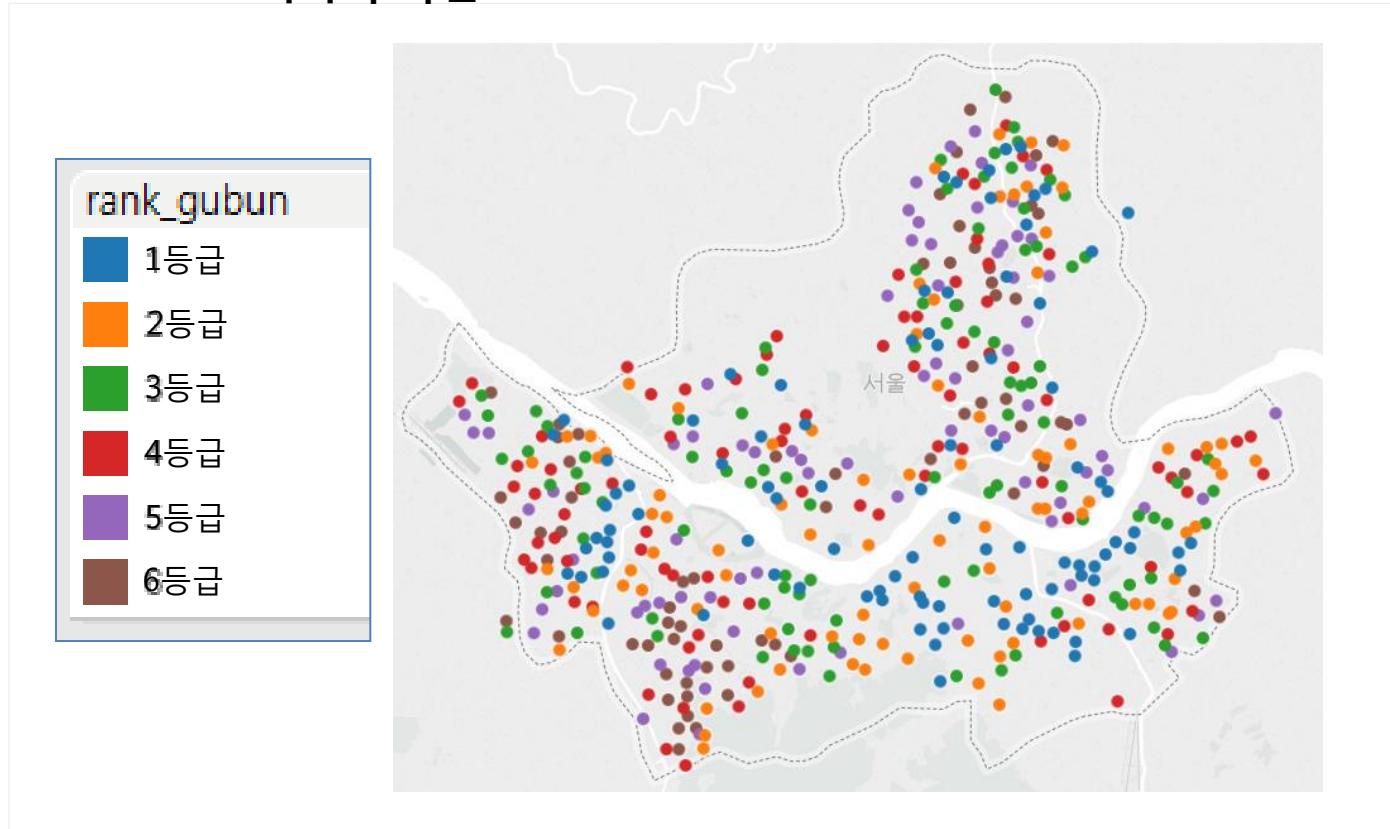
- 앞서 학교 순위별 상위 3개/하위 3개만 분석하여 특이사항 및 의미 있는 결과를 도출할 수 없었음
- 분석을 다른 각도로 재 분석이 필요했음.....
- 전체 591개 학교를 순위로 1~100, 101~200, 201~300,...으로 분류하여 총 6개 등급으로 나눔
- 주변 현황은 학교를 기준으로 2Km내에 시설물들을 검색하여 동일한 주제별로 SUM한 결과물로 분석함

4-3 데이터 분석

학교등급 구별 분포현황

KODB
한국데이터베이스진흥원

- Tableau 시각화 화면



4-3 데이터 분석

학교등급 구별 분포현황

KODB
한국데이터베이스진흥원

- 강남구, 서초구, 송파구, 양천구 상위 등급의 학교들이 많음
- 노원구는 다른 지역과 달리 등급이 골고루 분포되어 있음

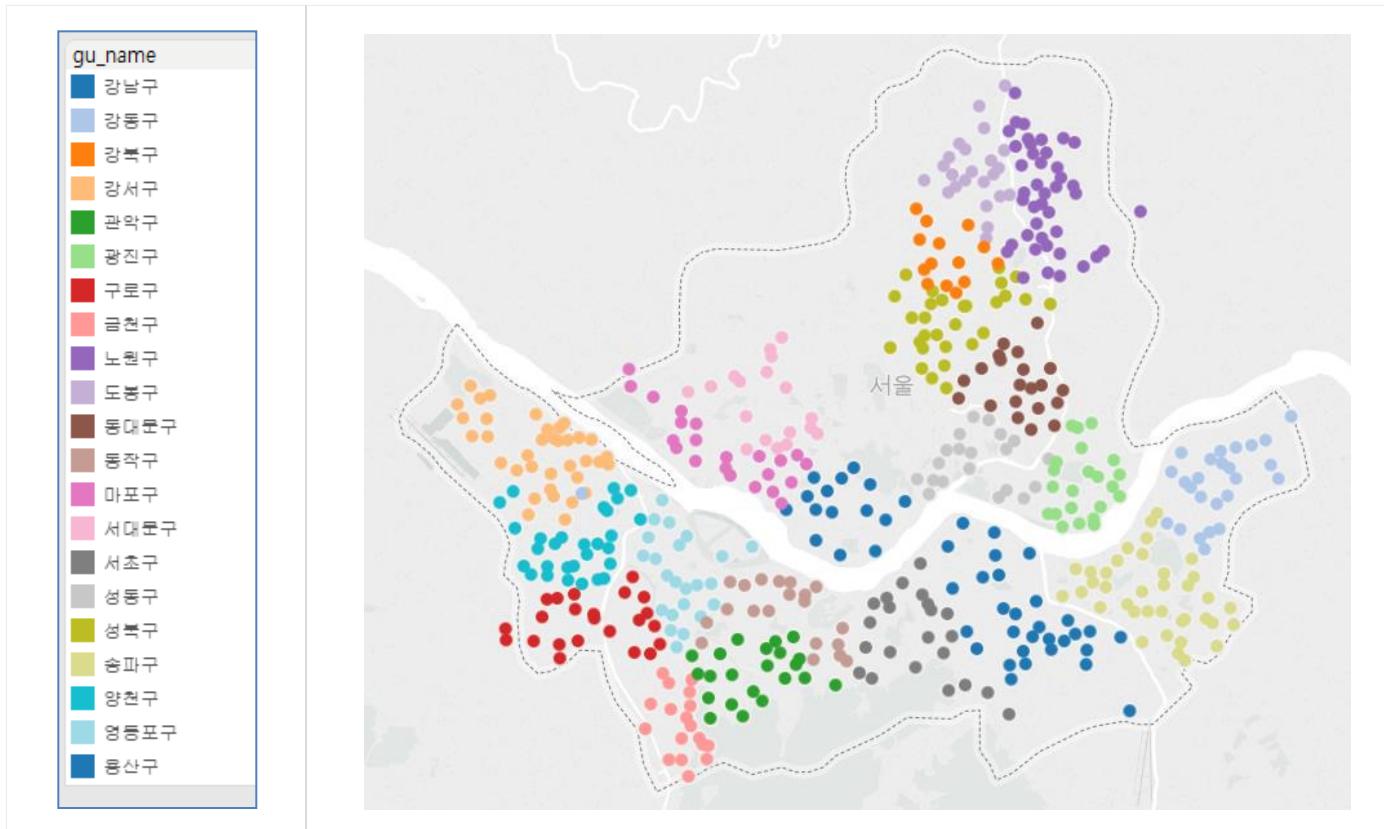
등급	강남구	강동구	강북구	강서구	관악구	광진구	구로구	금천구	노원구	도봉구	동대문구	동작구	마포구	서대문구	서초구	성동구	성북구	송파구	양천구	영등포구	용산구	은평구	종로구	중구	중랑구	총합
1등급	12	1	1	3	0	3	1	0	7	4	2	3	4	4	12	3	6	13	12	1	2	4	1	2	1	102
2등급	8	8	1	6	3	6	7	1	8	2	0	2	2	2	9	4	3	5	1	6	4	5	1	2	2	98
3등급	5	6	1	10	4	2	5	0	10	6	7	8	6	3	1	3	5	8	3	1	1	1	3	0	1	100
4등급	4	8	2	7	3	2	3	3	4	3	4	4	4	7	0	3	7	4	7	5	3	7	3	2	1	100
5등급	1	2	5	5	4	6	5	4	8	5	4	3	5	2	0	4	3	5	1	4	3	6	3	3	9	100
6등급	0	1	4	6	7	3	4	8	5	3	4	0	1	0	0	3	5	2	5	6	2	8	3	3	8	91
합계	30	26	14	37	21	22	25	16	42	23	21	20	22	18	22	20	29	37	29	23	15	31	14	12	22	591



4-3 데이터 분석

등급학교 구별 분포현황

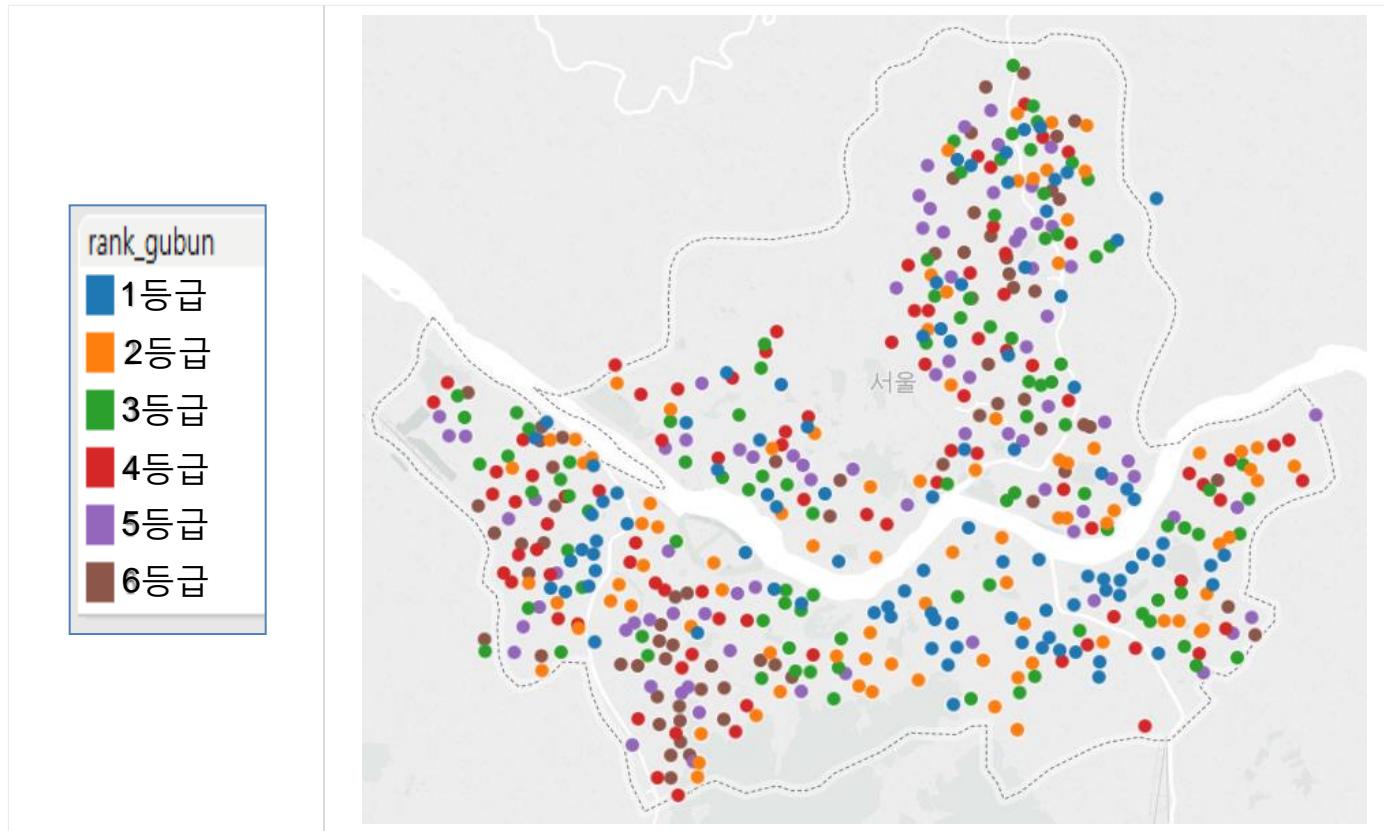
- 서울 591개 학교 구별 분포 현황



4-3 데이터 분석

등급별 분포현황

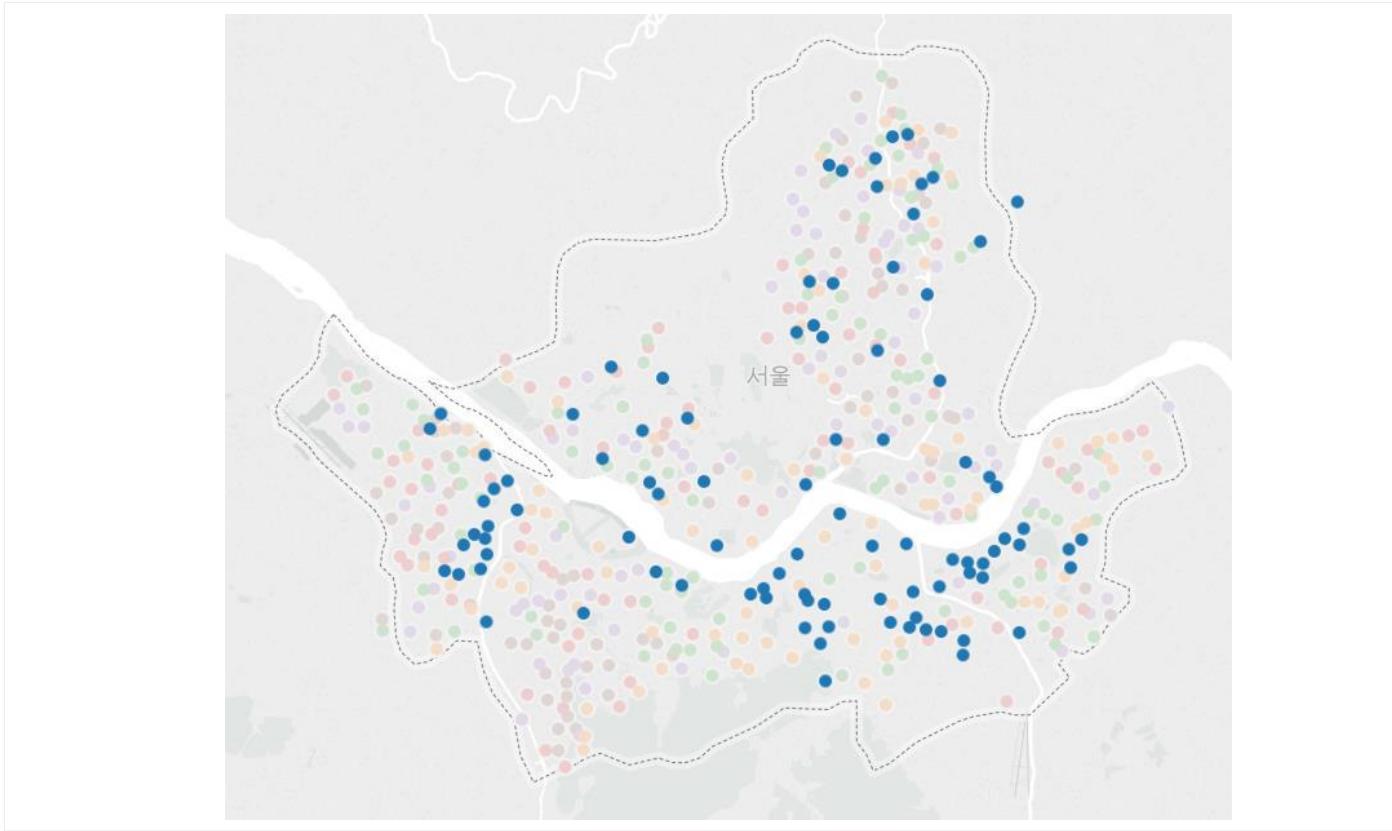
- 서울 591개 학교 등급별 분포 현황



4-3 데이터 분석

1등급 학교 분포현황 **KODB**
한국 데이터베이스 진흥원

- 서울 1등급 학교 분포 현황



4-3 데이터 분석

Splunk 작업화면

KODB
한국 데이터 베이스 진흥원

Views:
지도 | Geo Results | 이벤트

« 이전 1 2 3 4 5 6 7 8 9 10 다음 »

결과 줄바꿈

14. 10. 23. 오전 12:52:06.000 45894.0 , 인천 , 2 , biz-gis.com , 127.038885479 , 37.4712451613
geo_position=127.038885479,37.4712451613

14. 10. 23. 오전 12:52:06.000 45893.0 , 인천 , 2 , biz-gis.com , 127.038691202 , 37.4775870034
geo_position=127.038691202,37.4775870034

14. 10. 23. 오전 12:52:06.000 45892.0 , 인천 , 2 , biz-gis.com , 127.038498664 , 37.478495501
geo_position=127.038498664,37.478495501

14. 10. 23. 오전 12:52:06.000 45891.0 , 인천 , 7 , biz-gis.com , 127.038463083 , 37.4706830708
geo_position=127.038463083,37.4706830708

14. 10. 23. 오전 12:52:06.000 45890.0 , 인천 , 2 , biz-gis.com , 127.036634475 , 37.481913329
geo_position=127.036634475,37.481913329

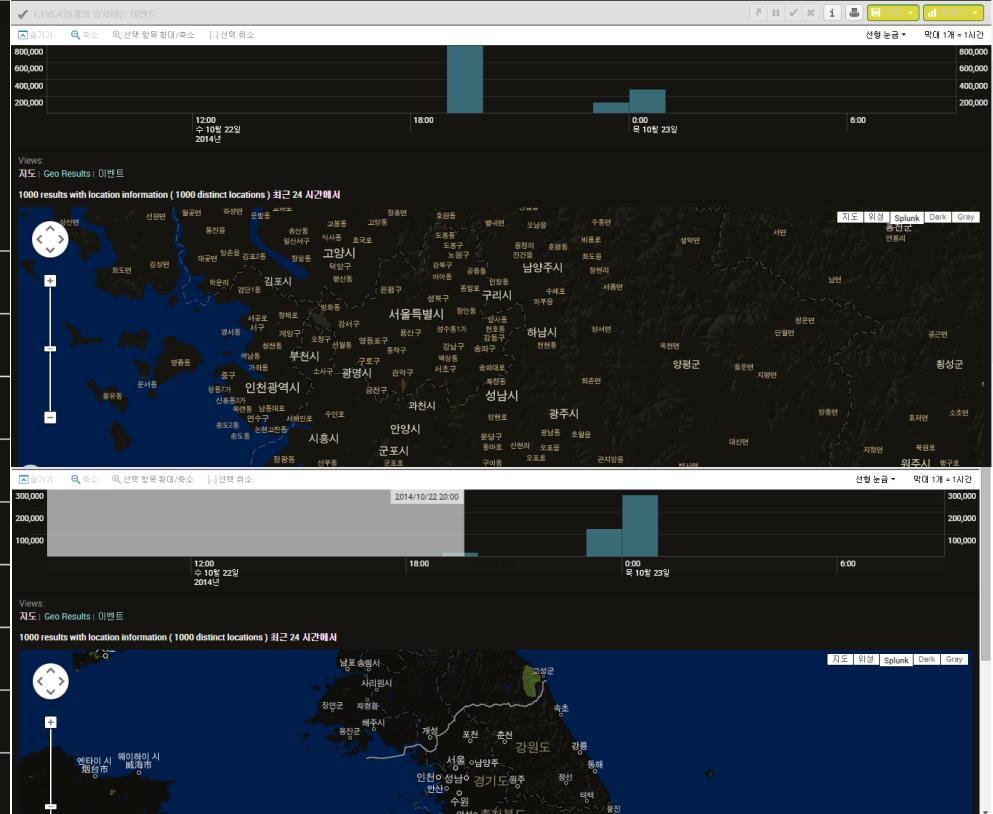
14. 10. 23. 오전 12:52:06.000 45889.0 , 인천 , 2 , biz-gis.com , 127.035743606 , 37.482608883
geo_position=127.035743606,37.482608883

14. 10. 23. 오전 12:52:06.000 45888.0 , 인천 , 5 , biz-gis.com , 127.035243143 , 37.4828390712
geo_position=127.035243143,37.4828390712

14. 10. 23. 오전 12:52:06.000 45887.0 , 인천 , 2 , biz-gis.com , 127.033162715 , 37.4863451797
geo_position=127.033162715,37.4863451797

14. 10. 23. 오전 12:52:06.000 45886.0 , 인천 , 2 , biz-gis.com , 127.032354891 , 37.4876090494
geo_position=127.032354891,37.4876090494

14. 10. 23. 오전 12:52:06.000 45885.0 , 인천 , 2 , biz-gis.com , 127.030792799 , 37.491236493
geo_position=127.030792799,37.491236493



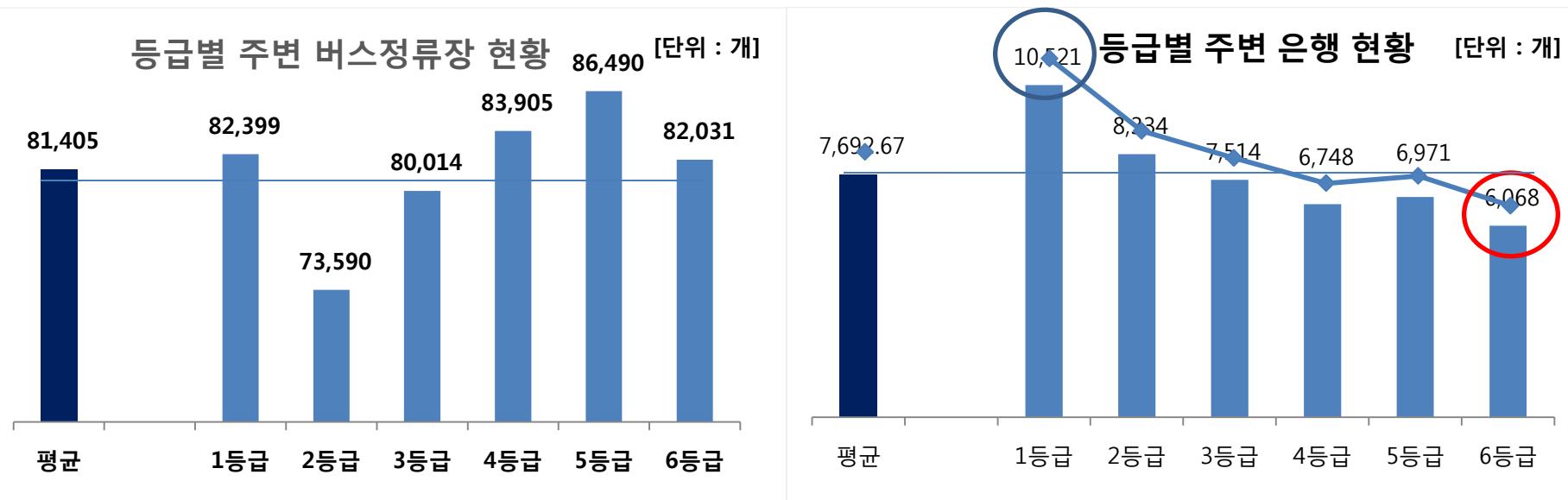
- mysql 연동 위치 정보 분석 데이터

- Splunk Google Maps 를 이용한 위치 정보 표시

4-3 데이터 분석

등급별 주변 시설현황

- 등급별 주변 버스 정류장 현황을 보면,
높은 등급과 낮은 등급간의 차이를 안보임
- 단, 등급별 은행 현황에서는 높은 등급과 낮은 등급간의 은행현황 개수가 눈에
띄게 차이가 나는데 대단위 아파트 단지와 주변시설에 영향을 받았을 것으로
추정함

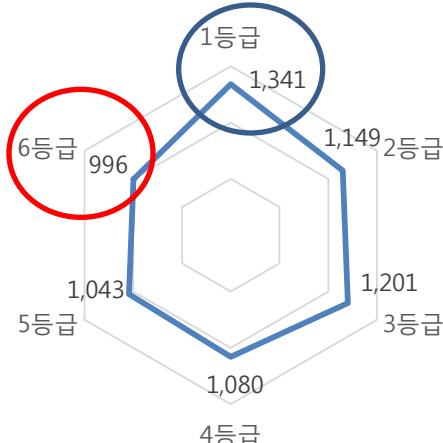


4-3 데이터 분석

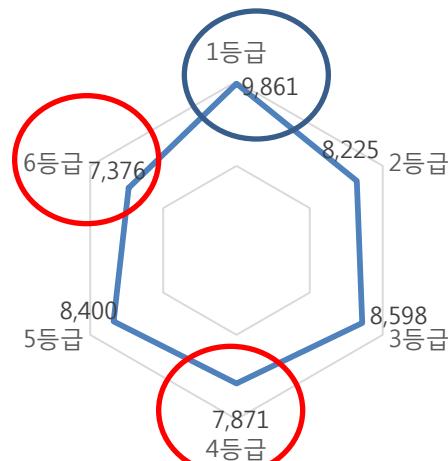
등급별 주변 시설현황

- 주변 소매점 현황에서는 앞선 은행 현황과 같은 결과가 나타남
- 주변 편의시설 현황은 1등급과 4등급, 6등급간의 차이가 크지만 2등급, 3등급, 5등급 간의 개수는 크게 차이가 나지 않으며 높은 등급학교 주변 환경에 영향을 준다고 보기 어렵음

등급별 주변 소매점 현황 [단위 : 개]



등급별 주변 편의시설 현황 [단위 : 개]



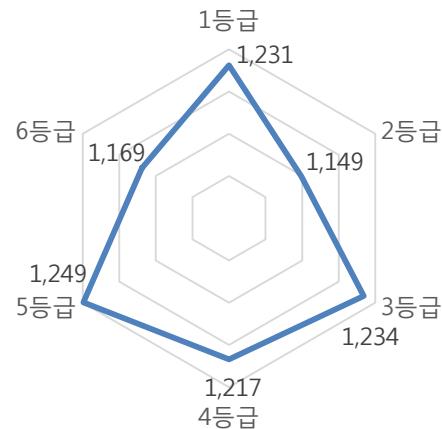
4-3 데이터 분석

등급별 주변 시설현황

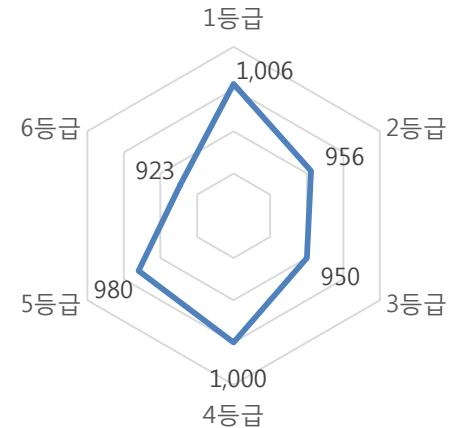
- 등급별 주변 중학교와 고등학교를 분석해본 결과,
등급별로 차이가 나타나는 결과를 보이지 않음

등급별 주변 중학교 현황

[단위 : 개]



등급별 주변 고등학교 현황 [단위 : 개]

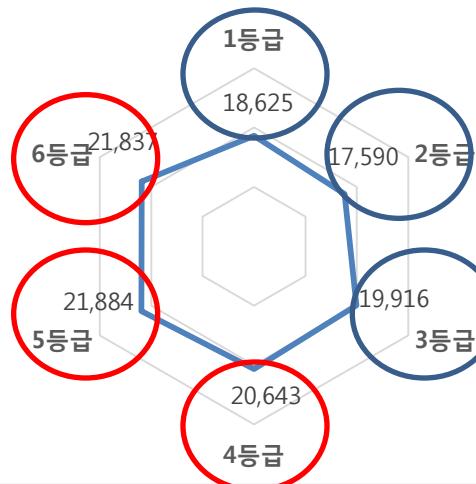


4-3 데이터 분석

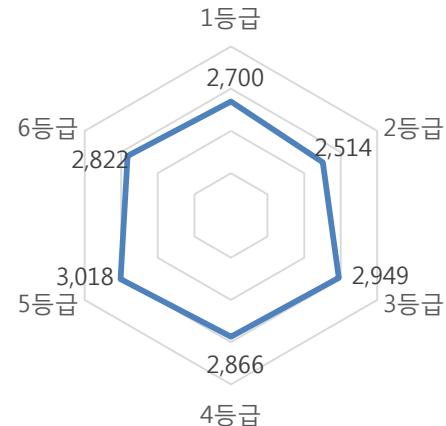
등급별 주변 시설현황

- 등급별 주변 유아원 현황을 보면,
1~3등급 주변 유아원 현황 보다 **4~6등급 유아원 현황 결과가 다소 높게 나타남**
해석 : 1~3등급의 지역보다 4~6등급 지역에 어린 자녀들을 둔 부모들이
맞벌이나 어린 자녀를 돌볼 여력이 없는 것으로 보임

등급별 주변 유아원 현황 [단위 : 개]



등급별 주변 유치원 현황 [단위 : 개]



4-3 데이터 분석

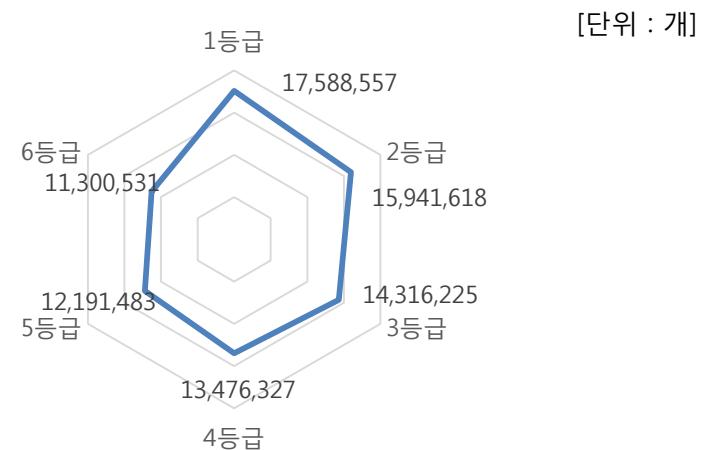
등급별 주변 시설현황

- 등급별 토지 공시지가 현황과 아파트 평단가 현황을 보면,
1~3등급이 주변 현황 보다 4~6등급 가격 현황이 높게 나타남

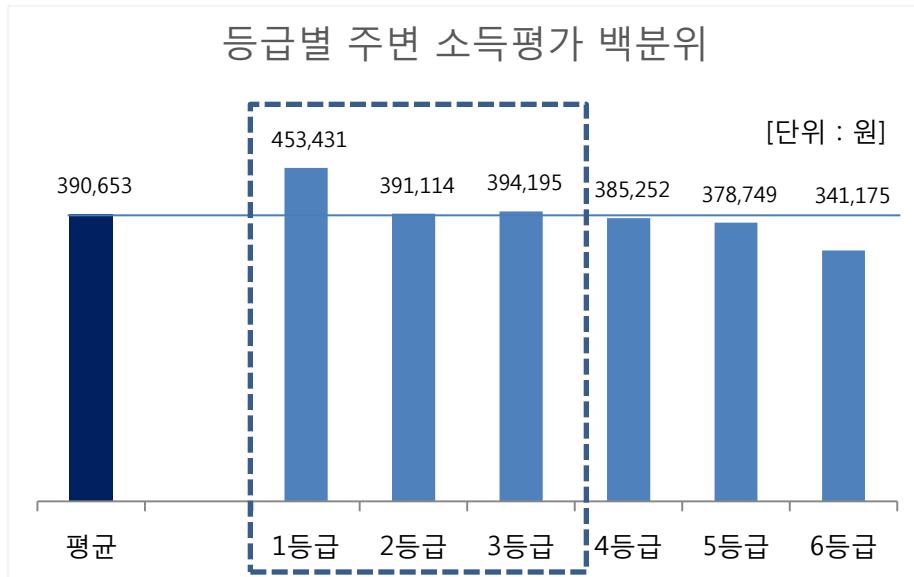
등급별 주변 토지 공시지가 현황



등급별 주변 아파트 평단가 현황



- 등급별 소득평가 백분위를 보면,
1~3등급이 4~6등급 보다 다소 높은 수준을 보임



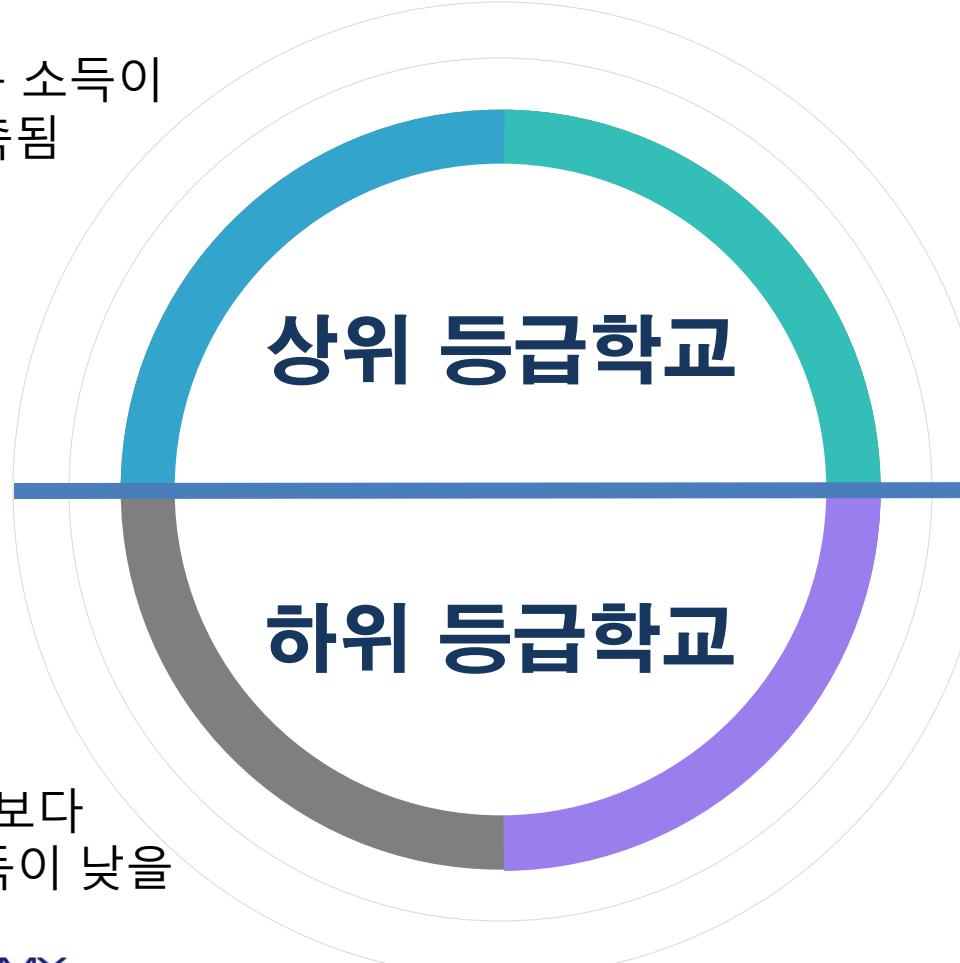
- 등급별 분석 결과 상위 등급학교와 하위 등급학교의 주변 편의 시설 중
은행 수와 공시지가, 아파트 평당단가가 상위 등급일 수로 높은 결과를 보임
- 어린이집과 유치원은 등급이 낮을 수록 많게 나타남

5-1 분석결과

인사이트

KODB
한국데이터베이스진흥원

- 상위 등급 학교는 소득이 많을 것으로 예측됨



토지 공시지가

은행

아파트 평단가

유치원

어린이집

- 상위 등급 학교 보다 상대적으로 소득이 낮을 것으로 예측됨

- 현재 수집 가공된 데이터를 통해 각 학교별, 등급별로 주변환경 요소와 상관관계 분석을 통해 모델을 개발하여 향후 서비스까지 가능할 것으로 예측됨

- 하둡 System을 이용하여 MySql에 있는 거리계산 함수 같은 지원되지 않은 것들을 사용할 기술력이 필요
- 시각화 툴의 사용법 숙지 (다양하게 시도 하였으나 완벽하게 하지는 못함)

- 구성된 데이터와 시스템을 이용하여 Web UI 개발
- 알고 싶은 지역 혹은 Map에 위치를 클릭하면 주변 학교 등급별 세부 내용 출력
- 주변 현황 (아파트 시세, 거주 인구, 편의 시설 등.) 정보 제공

- 데이터 찾기 난항
- 2주간 집체 교육을 받고 실제 프로젝트와 공부의 병행 한계
- 경험 없이 기반구축 진행방향 제시 힘듬
- 데이터 분석 시 다양한 각도에서 볼 수 있는 훈련 필요
- 분석 전 과정에서 빅 데이터 플랫폼을 다 쓰는 것이 아님
- 기본 분석 툴로 많은 부분 구현 가능

- 김은희 : 조장으로 전체 프로젝트 진두 지휘 (데이터베이스 쿼리, PT 등)
- 김성표 : 놀고 먹으면서 막판 PT작업 투입(성공적인 프로젝트를 위한 숨은 공로자)
- 노병희 : Data Search 및 스플렁크 작업 투입
- 전성종 : 회의 진행 및 Java 좌표변환 소스 개발
- 최혁근 : Hadoop 설치 및 System 관련 작업 투입
- 노태상 : 회의 때마다 늘 조원들 마음속에 투입

- 김성표 : 빅데이터 기술은 머리가 지끈 하지만 조원들을 생각하면 가슴이 뜨끈~
- 최혁근 : 데이터 전달에서 데이터에 눈을 뜨다~~^^~~
- 전성종 : 나는 개근상 줘야해~! 꼬박꼬박 회의 참석 ^^
- 노병희 : 나는 세종시~ 그래도 마음은 서울시~!!!
- 김은희 : 인프라는 본능으로 구축하고, 데이터는 마음으로 이해하자.

빅데이터 기술

8기 2조

발표 끝!!

감사합니다.