

빅콘테스트 Innovation분야 데이터터 소개

2019. 07. 17



빅데이터의 정의를 어떻게 내리고 있는가에 따라 다른 접근이 될 것임. 빅데이터에 관한 사이즈, 복잡성, 그 기대 가치 등...다양한 관점의 정의가 있음.

빅데이터란?

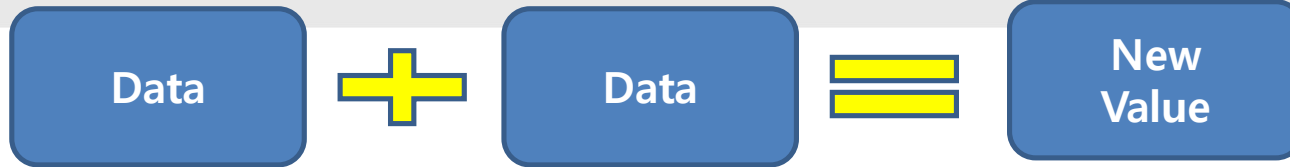
Volume, Velocity, Variety, Value
Data 기반의 문제해결 능력
Data Literacy

-
-
-
-
-
-

빅데이터는 이종의 데이터를 연결하여
새로운 가치를 만들어내는 일련의 프로세스와 분석 방법론이다.

Big data is a series of analysis processes and
methodologies to create a new value by connecting
the heterogeneous dataset. -김정선-

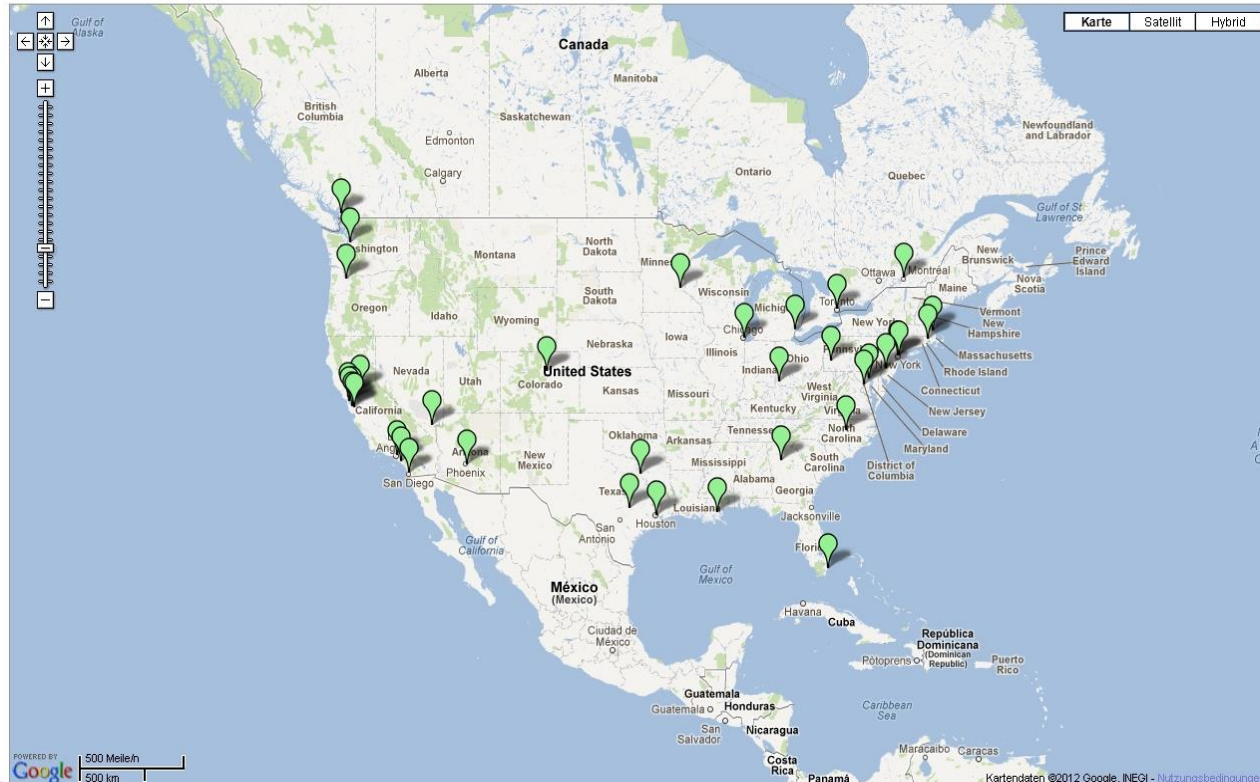
미국 mash-up 서비스 하우스징맵 이후 많은 데이터기반 서비스들이 보편화 중



For Rent For Sale Rooms Sublets

City: Price: [Show Filters](#) [New](#) [Refresh](#)

Powered by [craigslist](#) and [Google Maps](#)
(this site is in no way affiliated with craigslist or Google)
[About / Feedback](#)



Drag the map with your mouse, or double-click to center.

Click on an  icon to select a city.

In each city, click on the  and  icons to see listings.
You can also click the  and  icons next to each listing.

Yellow icons have pictures.

Set the price range using the drop-down menu above.

데이터를 통해 사회 경제적 문제를 풀고자 하는 접근도 있음

- CDR 데이터를 활용한 지역 경제편차 변이

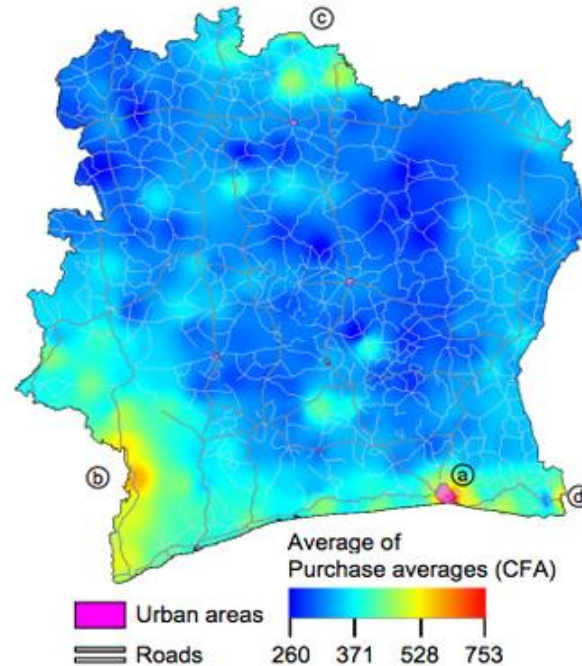


Fig. 2: Average of each user's purchase average. (a) Abidjan, (b) Liberian border (c) Roads to Mali and Burkina Faso (d) Road to Ghana

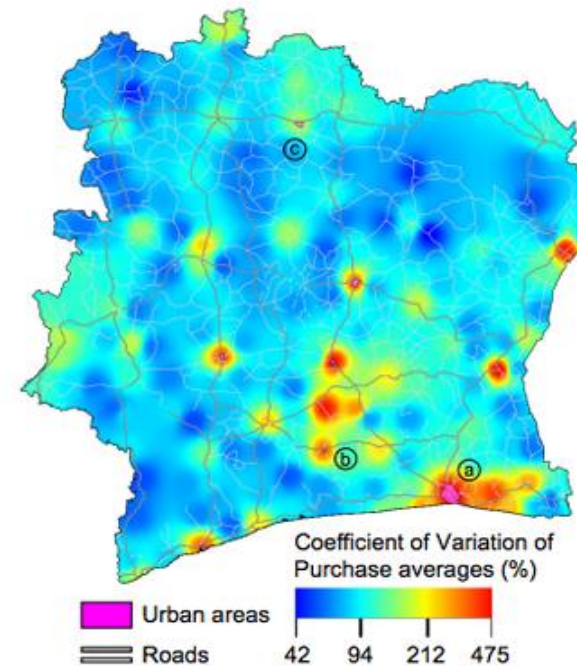


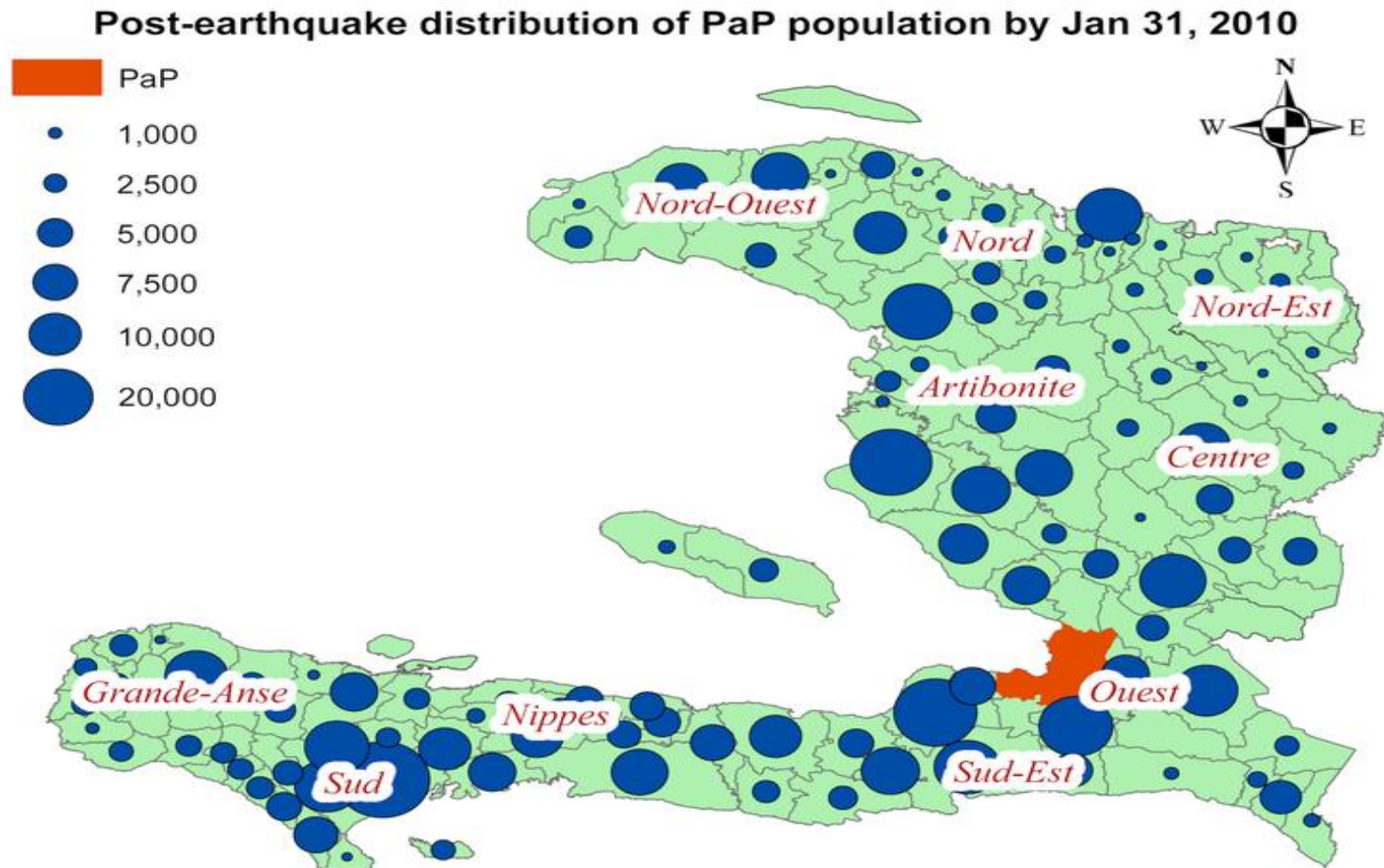
Fig. 3: Coefficient of Variation of purchase averages, measures the diversity. (a) Abidjan, (b) Divo (c) Korhogo

▲서아프리카의 빈곤국가 아이보리 코스트의 낮(왼쪽)과 밤(오른쪽) 시간에 핸드폰 사용량의 변화를 보여주는 지도입니다. 핸드폰 커뮤니케이션 및 통화가능시간(airtime credit)에 대한 구입 기록 데이터를 비교하면 지역별 소득(local-level income)과 사회경제적 차이(socioeconomic segregation)를 확인할 수 있습니다. 이러한 데이터를 활용하면 식량안보(food security)를 파악할 수 있습니다[2]

[출처] 빅데이터는 지구촌 빈곤의 감소를 위해 어떻게 활용되는가?작성자 GP3 KOREA

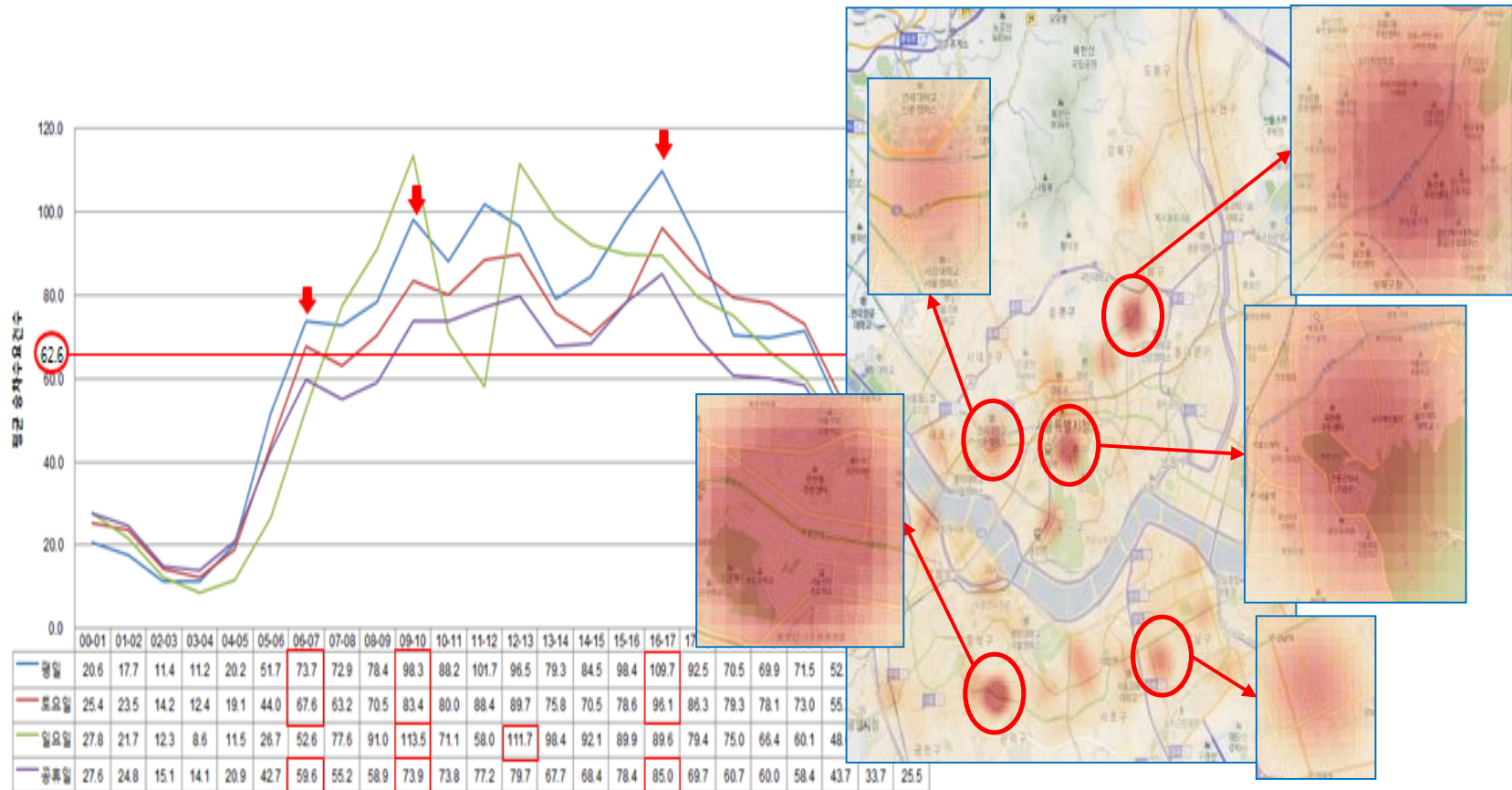
또한 재난구제관련 접근도 데이터를 통해서 가능

- 2010년 아이티 지역 피난민 분포



복지문제 해결을 위한 데이터 접근도 가능

• 시각장애인콜택시 운영 효율화를 위한 빅데이터 분석



[평일 밤시간대 수요집중지역]

경기도의 빅데이터를 활용한 CCTV 입지분석



<경기도, 빅데이터로 CCTV 최적 위치 정한다>

데이터 분석 통해 다양한 민원서비스 제공

(수원=연합뉴스) 김인유 기자 = 수원시 한 동네에 지난 해 1년간 여성을 상대로 음란행위를 하는 숙청 '바바리맨'이 나타났다'거나 '청소년 범죄가 우려된다'는 민원신고 6건이 접수됐다.

이 지역에는 폐쇄회로TV(CCTV)가 3대 설치돼 있다. 그러나 6건 민원신고 가운데 2건만 CCTV감시 범위에 포함됐다.

나머지 4건은 CCTV가 잡아낼 수 없는 '사각지대'에 놓여 있었다.

주문하신

경기도 정보서비스담당관이 민원 발생지역과 CCTV설치 위치 등의 데이터를 종합적으로 분석해 내놓은 결과다.

도의 분석결과에 따라 이 동네 CCTV는 설치위치가 조정되거나 추가로 설치될 예정이다.

경기도가 올해 수원시 CCTV위치 분석처럼 빅데이터를 도민 행정서비스에 접목하는 5가지 사업을 시작했다.

우선 CCTV설치 위치 최적화 사업은 유동인구, 경찰청의 범죄정보, 지리정보시스템을 결합해 가장 효과적인 설치 지점을 찾는 작업이다.

수원시가 도가 제공한 빅데이터 자료를 검증한 뒤 올해 최적의 CCTV설치 장소를 정할 예정이다.

도 축제관할 분석사업은 축제 개최 지역에서 발생한 매출정보와 교통정보, 유동인구 등을 분석해 축제 전후 매출변화를 비교, 지역경제활성화 효과를 측정한다.

이밖에 아동 청소년의 비안 예방관리 구축, 환승센터 위치 선정, 민원지도 제작 사업도 진행한다.

도는 빅데이터 사업과 별도로 체계적인 정보공개 시스템도 마련한다.

도는 올 하반기 경기도와 산하 공공기관에서 생산하는 다양한 정보를 담은 '데이터 개방 포털 사이트' 구축사업을 완료할 계획이다.

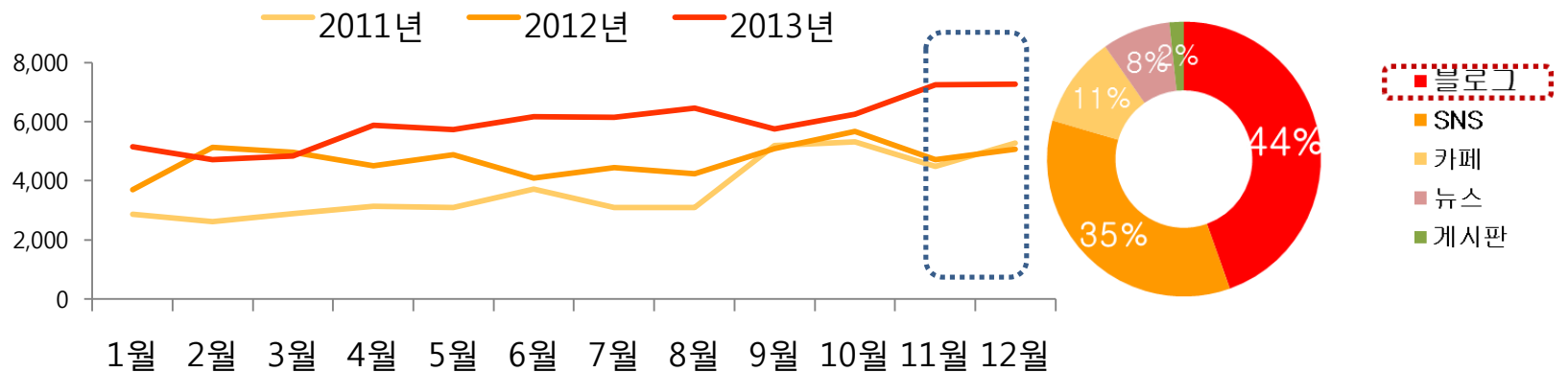
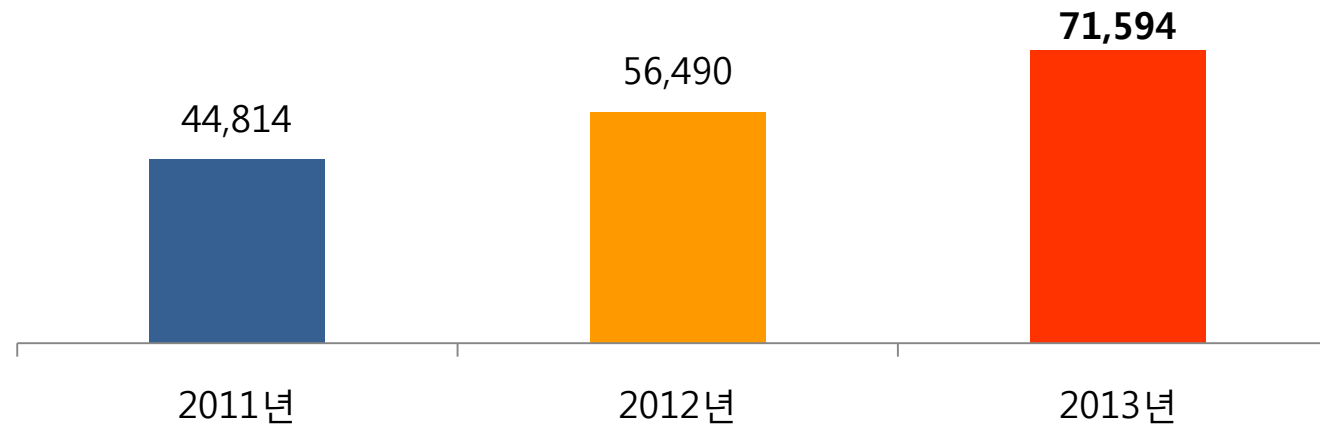
hedghog@yna.co.kr

<경제특파원(사) 연합뉴스, 무단 전재-재배포 금지>

2014/03/02 07:11 송고

국민 건강관리 관점의 소셜 모니터링 사례

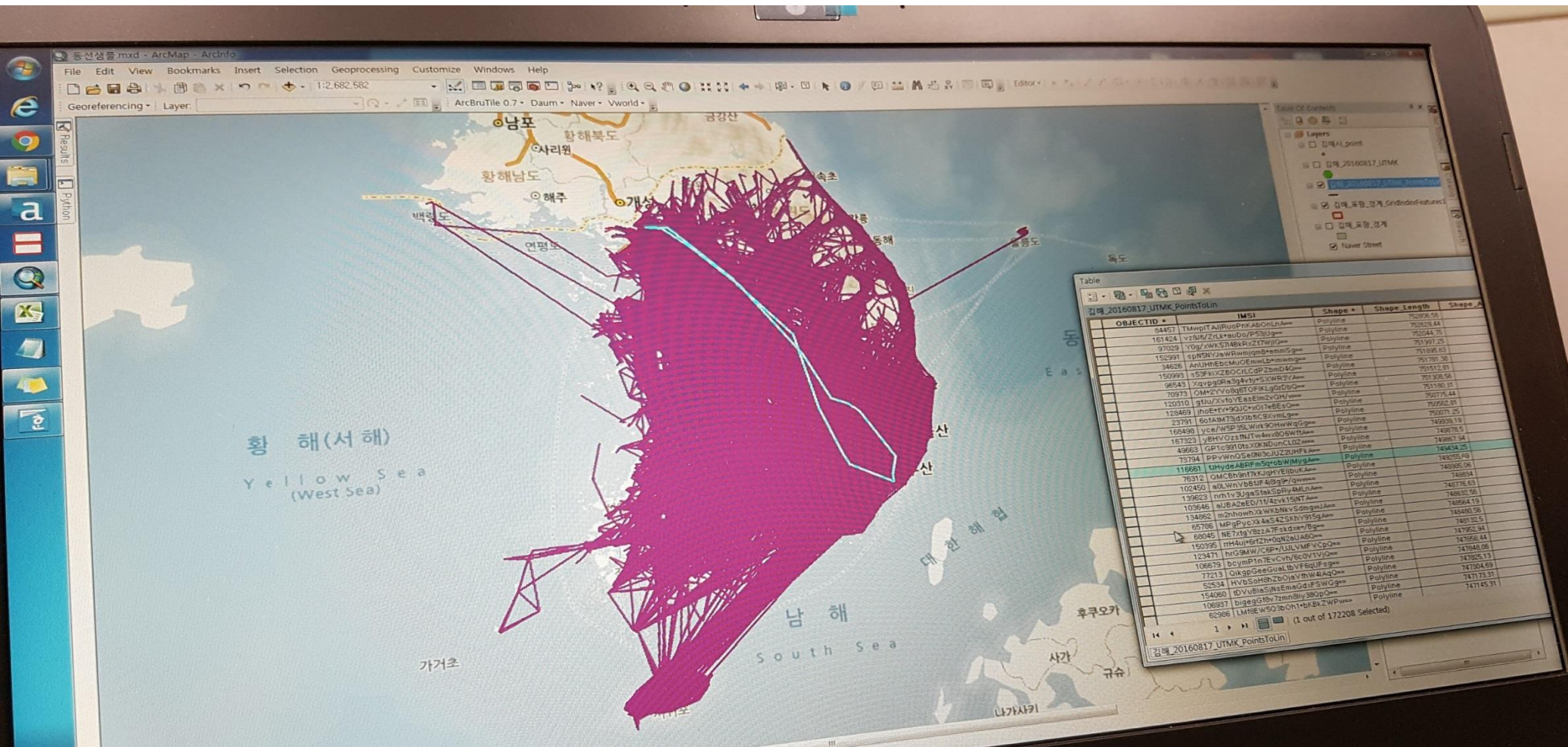
Social Buzz분석을 통한 국민건강 모니터링(Topic : 아토피/SKT 스마트인사이트)



각 사별 데이터 소개

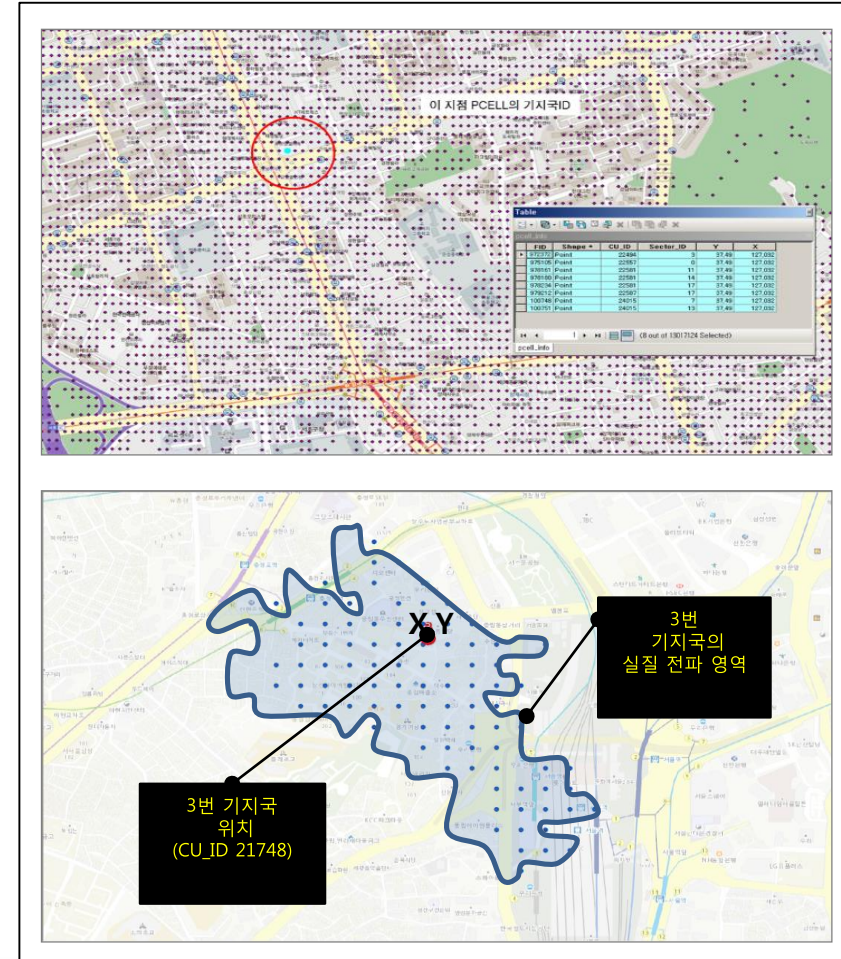
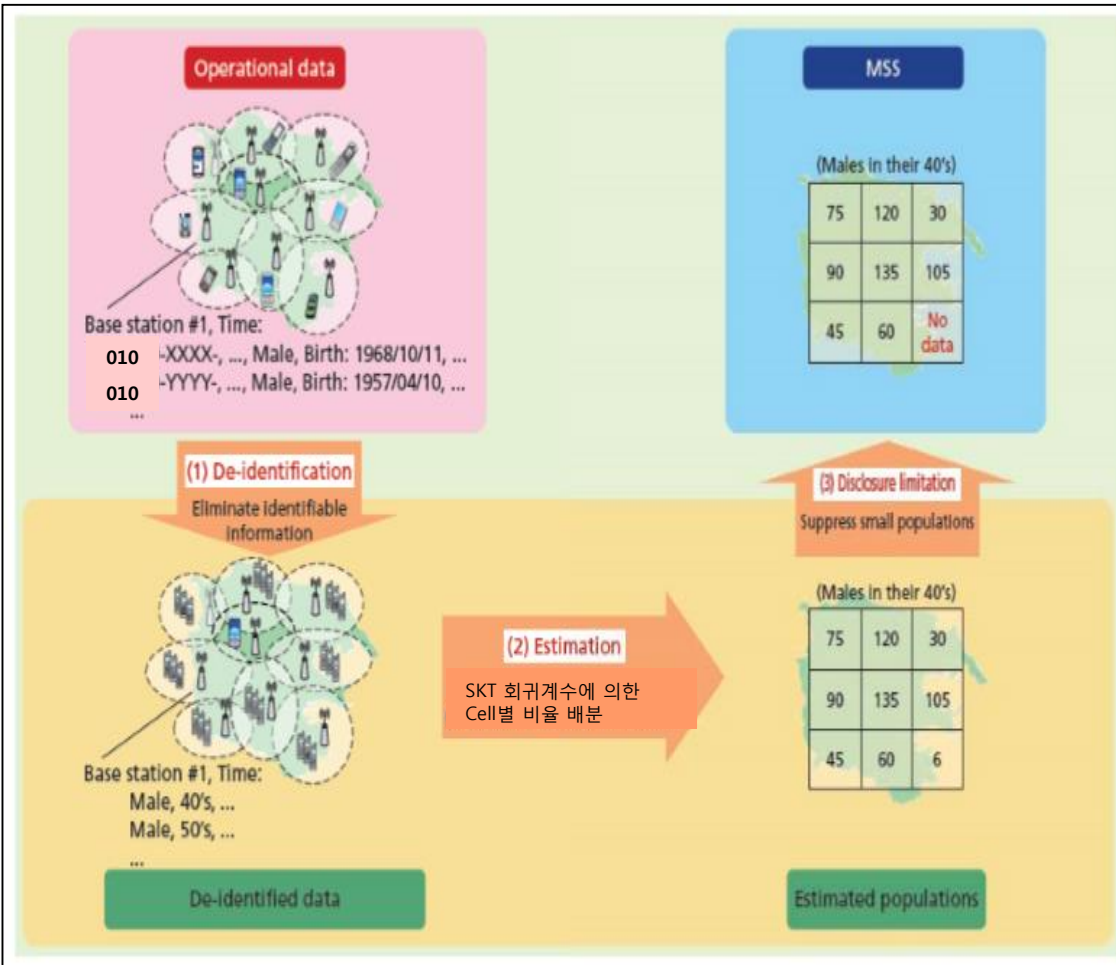
SKT 빅데이터 소개

2,700만 SKT 고객의 가입자 정보 및 위치정보를 기반으로 데이터를 생성합니다. (암호화 처리)



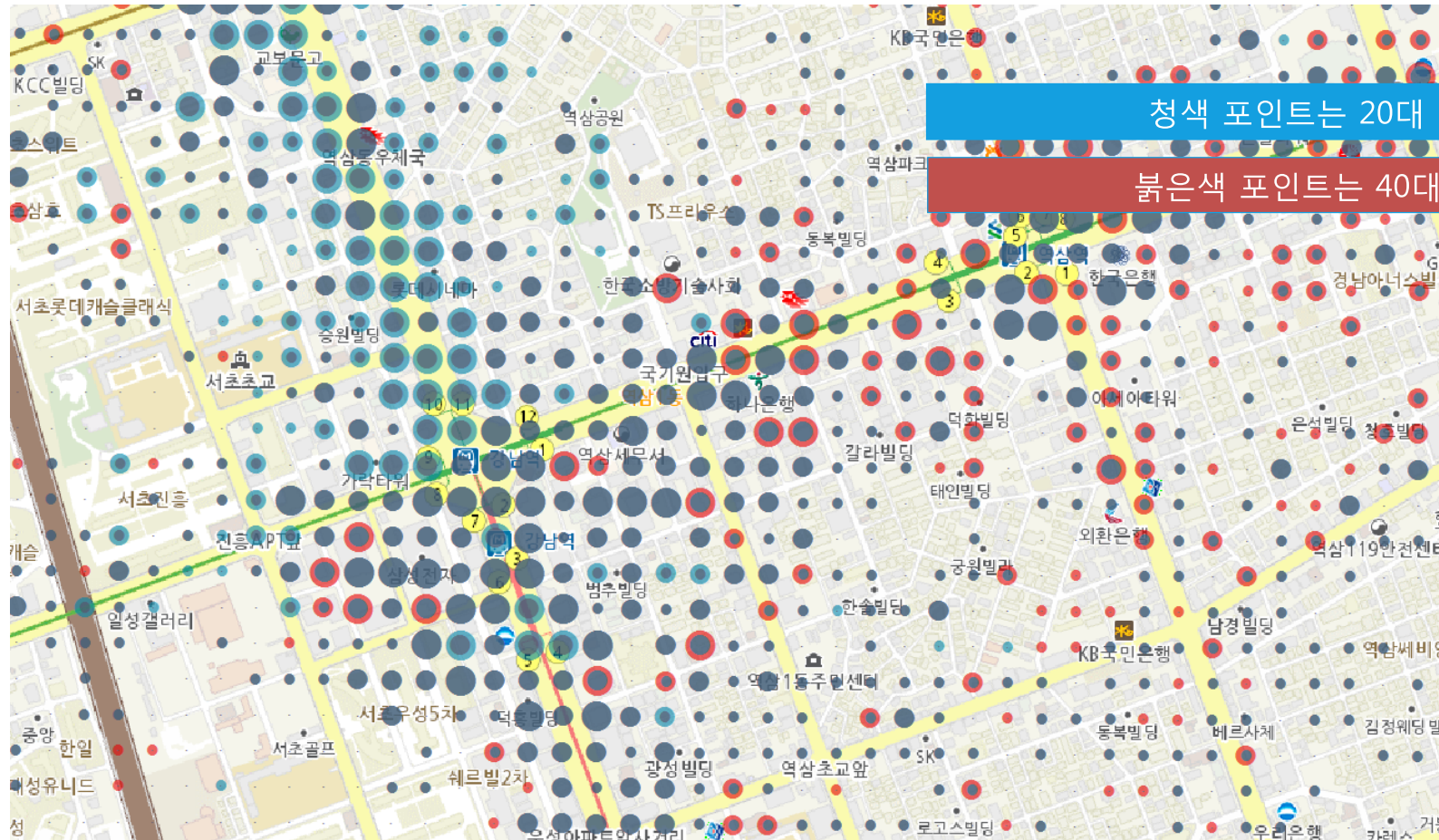
SKT 빅데이터 소개

SKT의 위치 데이터는 기지국에서 생성되는 신호를 바탕으로 해당 지역의 전파 범위를 SKT만의 pCell 기술을 활용하여 GIS에 비식별 조치를 통해서 적용하게 됩니다.



50m Cell 단위 유동인구 데이터

정제된 데이터는 50m × 50m 단위의 그리드로 해당 지역에 현재 존재하는 존재인구, 해당 지역 내 활동중인 활동인구, 통과/방문인구 데이터를 구분 가능합니다.




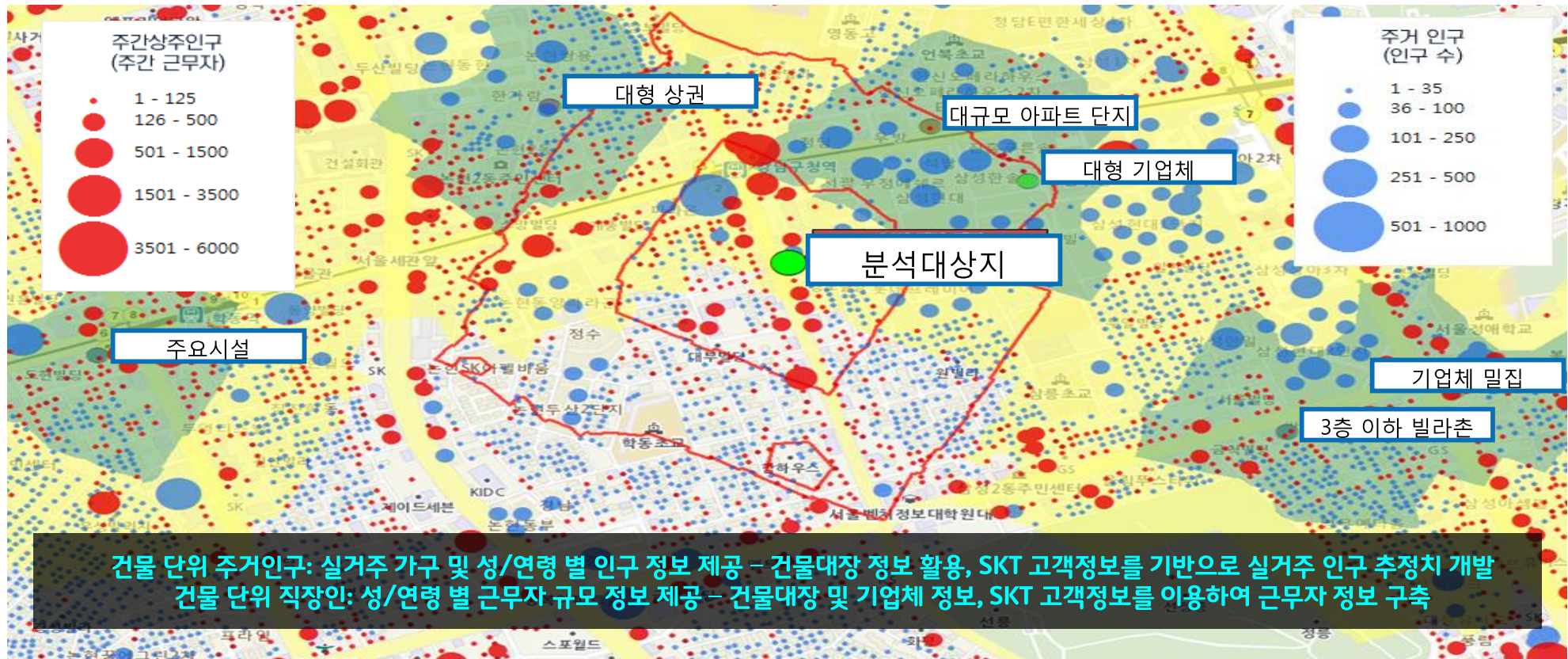
청색 포인트는 20대 여성의 수

붉은색 포인트는 40대 남성의 수

잠재고객 데이터: 주거인구, 직장인구 DB

SKT의 고객 정보를 이용하여 다양한 데이터와 연계 시 건물 단위 주거인구/직장인구 정보 구축이 가능하며 성/연령 별 실거주 인구 및 실 근무인구 수를 추정가능 합니다.

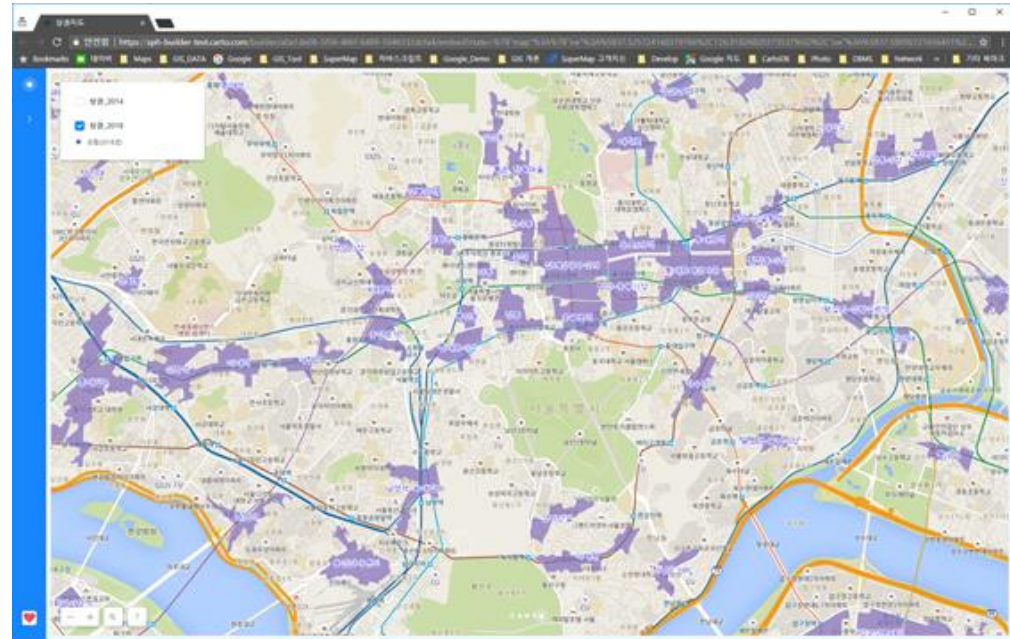
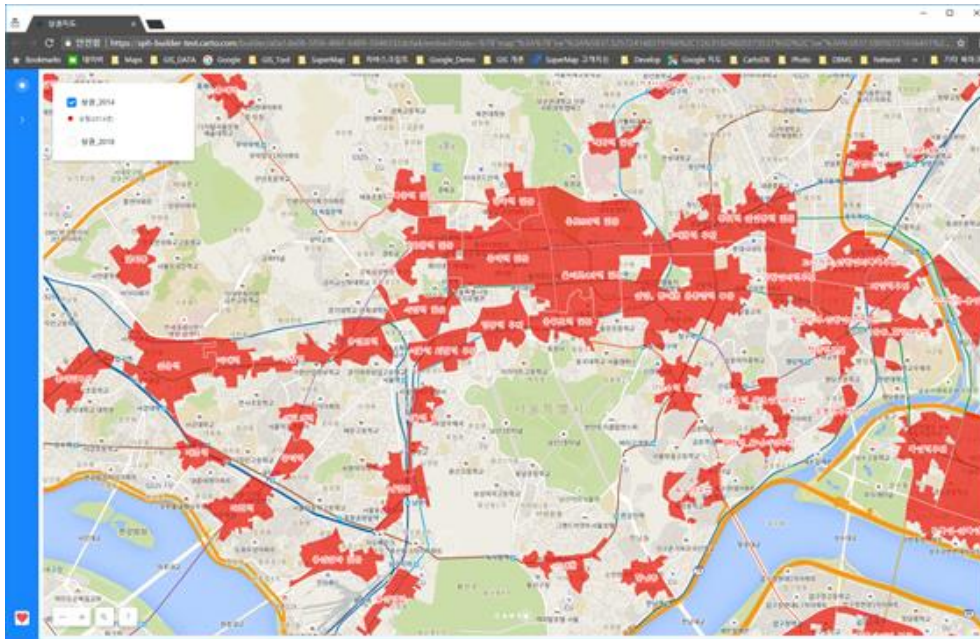
 국토부의 건물 대장 기준, 건물 단위 실거주 인구 및 근무자 수



상권분석: 특정 연령 DB

관련 데이터를 이용하여 상권분석 등이 가능합니다.

예를 들어 2014년도에 20대 연령 여성 활성화 상권과 18년도 상권의 차이를 파악 할 수 있습니다.



<https://sph-builder-test.carto.com/builder/a0a1de08-5f56-486f-8489-5046332dcfa4/embed>

SK텔레콤 제공 DB

SK텔레콤에서 생성한 주거지, 오피스지역 대표구인 종로구, 노원구의 행정동 기준 일별, 시간대별 유동인구 데이터를 제공합니다

예시

테이블정의서								
주제영역명		유동인구	작성일			작성자		
테이블ID		FLOW_AGE_(STD_YM)	테이블명		성연령 유동인구(5세단위)			
테이블설명		5세단위 노원구/종로구 유동인구						
No.	컬럼ID	컬럼명	타입	길이	NULL	KEY	DEFAULT	비고
1	STD_YM	년월	VARCHAR2	6				
2	STD_YMD	년월일	VARCHAR2	8				
3	BDONG_CD	법정동코드	VARCHAR2	10				
4	BDONG_NM	법정동명칭	VARCHAR2	10				
5	MAN_FLOW_POP_CNT_0004	남성_0004세 유동인구	NUMBER	(18, 2)				
6	MAN_FLOW_POP_CNT_0509	남성_0509세 유동인구	NUMBER	(18, 2)				
7	MAN_FLOW_POP_CNT_1014	남성_1014세 유동인구	NUMBER	(18, 2)				
8	MAN_FLOW_POP_CNT_1519	남성_1519세 유동인구	NUMBER	(18, 2)				
9	MAN_FLOW_POP_CNT_2024	남성_2024세 유동인구	NUMBER	(18, 2)				
10	MAN_FLOW_POP_CNT_2529	남성_2529세 유동인구	NUMBER	(18, 2)				
11	MAN_FLOW_POP_CNT_3034	남성_3034세 유동인구	NUMBER	(18, 2)				
12	MAN_FLOW_POP_CNT_3539	남성_3539세 유동인구	NUMBER	(18, 2)				
13	MAN_FLOW_POP_CNT_4044	남성_4044세 유동인구	NUMBER	(18, 2)				
14	MAN_FLOW_POP_CNT_4549	남성_4549세 유동인구	NUMBER	(18, 2)				
15	MAN_FLOW_POP_CNT_5054	남성_5054세 유동인구	NUMBER	(18, 2)				
16	MAN_FLOW_POP_CNT_5559	남성_5559세 유동인구	NUMBER	(18, 2)				
17	MAN_FLOW_POP_CNT_6064	남성_6064세 유동인구	NUMBER	(18, 2)				
18	MAN_FLOW_POP_CNT_6569	남성_6569세 유동인구	NUMBER	(18, 2)				
19	MAN_FLOW_POP_CNT_70U	남성_70세이상 유동인구	NUMBER	(18, 2)				
20	WMAN_FLOW_POP_CNT_0004	여성_0004세 유동인구	NUMBER	(18, 2)				

1

5세단위 연령 유동인구
(일평균)

2

5세단위 성별 유동인구
(일평균)

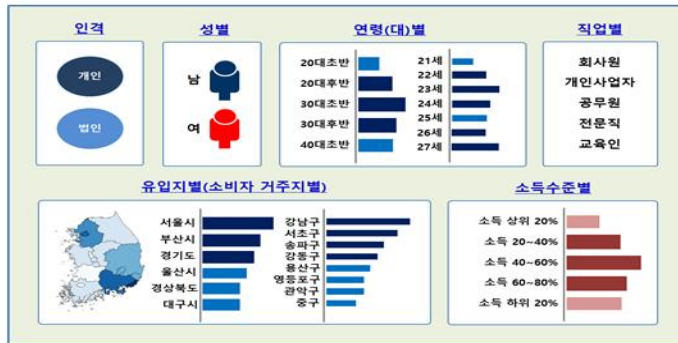
2

1시간단위 유동인구
(일평균)

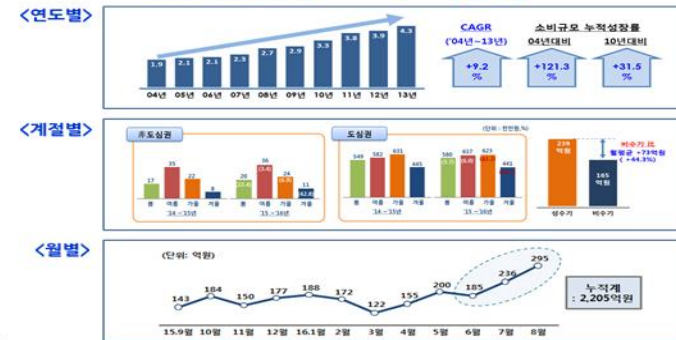
신한카드 빅데이터

2,100만 신한카드고객의 DEMO 및 실시간 카드거래 내역을 기반으로 데이터를 생성합니다. (개인정보 비식별화)
카드소비 빅데이터의 소비 주체/시점/장소/업종 정보를 활용하면 다양한 소비패턴 분석이 가능합니다.

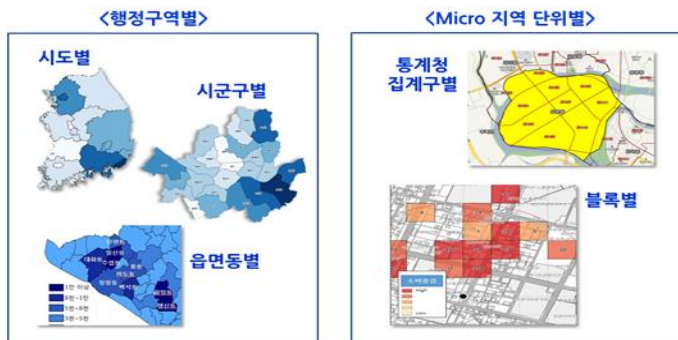
Who(소비주체)



When(소비시점)



Where(소비장소)

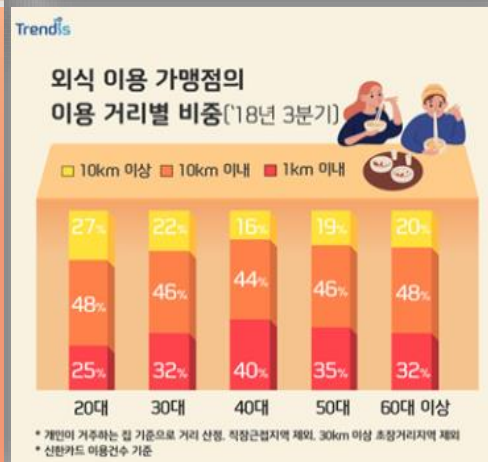
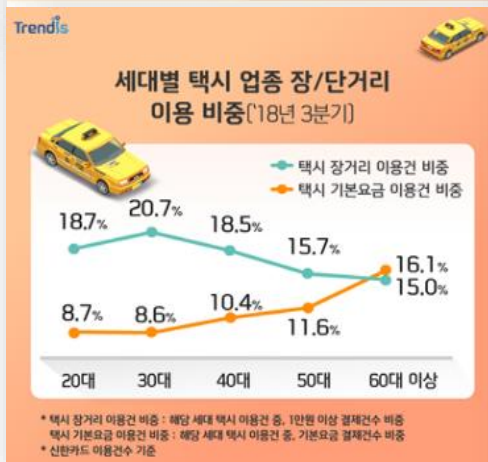


What(가맹점/업종)



카드빅데이터로 본 소비 트렌드

카드 소비데이터를 통해 고객의 다양한 Life Style을 파악하고 외부 데이터와 결합을 통해 다양한 인사이트 확보가 가능합니다.



신한카드 Trendis

신한카드 대고객서비스 Trendis

트렌드클리프 트렌드뉴스 인포그래픽스

빅데이터로 보는 소비자 라이프 스타일

신한카드가 선정한 2019년 소비 트렌드 키워드 'DETAIL'

온라인 쇼핑, 11월을 주목하라!

20대 소비스타일 작은 사치와 일상의 공유

제3의 공간, 리딩(Reading) 문화 공간'

신한카드 대회용 제공 DB

신한카드에서 생성한 서울시 종로구, 노원구의 2018.4월~2019.3월까지의 1년간 일별 DEMO*업종*행정동 별 카드 소비 데이터를 제공합니다.

1

행정동 별 카드소비데이터
(건수,금액)

2

행정동 & 업종코드

예시

No.	컬럼ID	컬럼명	타입
1	STD_DD	기준일자	VARCHAR(8)
2	GU_CD	구코드	VARCHAR(2)
3	DONG_CD	행정동코드	VARCHAR(3)
4	MCT_CAT_CD	업종코드	VARCHAR(2)
5	SEX_CD	성별코드	VARCHAR(1)
6	AGE_CD	나이코드	VARCHAR(2)
7	USE_CNT	이용건수	NUMBER
8	USE_AMT	이용금액	NUMBER

업종코드	업종명	세부업종
01	숙박	특급호텔, 1급 호텔, 2급 호텔, 콘도미니엄, 모텔,여관,기타숙박
02	여행사 및 렌터카	관광여행사, 렌터카
03	골프장	실내골프장, 실외골프장 골동품,예술품, 화랑,표구사, 관광민예,선물용품, 화방, 화원, 수족관, 음반 테이프, 악기, 피아노 대리점, 운동경기,레저용품, 레저스포츠, 테니스장, 볼링장, 스키장, 수영장, 당구장, 종합레저타운/놀이동산, 비디오방/전화방, 스포츠센터/레포츠클럽, 온천장, 공연장, 극장, 운동경기관람, 수련원, 체험장, 전자오락실, PC게임방 등
04	레저시설 및 판매	노래방
05	노래방	애완동물, 동물농장, 동물병원
06	가축병원	목재가구, 철재가구
07	가구업체	주유소, LPG가스
08	주유소	가정용연료
09	유류판매	백화점
10	백화점	쇼핑몰, 대량양판점, 생활잡화, 식품잡화, 연쇄점, 사원전용매장, 보훈매장, 복지매점, 농협, 주류판매 등
11	슈퍼마켓	연금매장, 농수축협직판장, 수입상품점, 회원제, 방문판매
12	유통업체	할인점/슈퍼마켓, 농협마트
13	할인점	면세점
14	면세점	편의점
15	편의점	수예,지물,커피,카페, 남.여기성복, 양복, 양품점, 내의류판매업, 침구(이불), 아동복, 옷감,직물류, 캐주얼의류, 혼수전문점
16	의류 및 직물업체	

GS리테일 대회용 제공 DB

GS리테일에서 생성한 서울시 종로구, 노원구의 2018.4월~2019.3월까지의 1년간 품목*행정동별 편의점 매출지수 데이터를 제공합니다.

1

행정동별 편의점 매출 지수 데이터

2

행정동별 편의점 품목 카테고리
지수 데이터
(카테고리 총 100% 합이 되도록
지수화)

순서	컬럼ID	컬럼명	물리데이터타입	비고
1	OPER_DT	영업일자	VARCHAR(8)	
2	BOR_CD	구코드	VARCHAR(5)	
3	ADMCD_CD	행정동코드	VARCHAR(8)	
4	AMT_IND	매출지수	NUMBER(5,2)	
5	LCLS__10_P	식사_비중	NUMBER(5,2)	비중_%
6	LCLS__20_P	간식_비중	NUMBER(5,2)	비중_%
7	LCLS__30_P	마실거리_비중	NUMBER(5,2)	비중_%
8	LCLS__40_P	홈&리빙_비중	NUMBER(5,2)	비중_%
9	LCLS__50_P	헬스&뷰티_비중	NUMBER(5,2)	비중_%
10	LCLS__60_P	취미&여가활동_비중	NUMBER(5,2)	비중_%
11	LCLS__70_P	사회활동_비중	NUMBER(5,2)	비중_%
12	LCLS__80_P	임신/육아_비중	NUMBER(5,2)	비중_%
13	LCLS__90_P	기호품_비중	NUMBER(5,2)	비중_%

케이워드 대회용 제공 DB

신한카드에서 생성한 서울시 종로구, 노원구의 측정기를 통해 수집된 2018.4월~2019.3월까지의 1년간 일별, 시간대 미세먼지 및 위험물질, 온/습도 데이터 제공합니다.

1

행정동별 미세먼지 측정 데이터

2

행정동별 유기화학물질, 이산화탄소
농도 측정 데이터(실내만 사용)

3

행정동별 온/습도 데이터

데이터	설명	비고
tm	데이터 측정 날짜 시간(YYYYMMddhhmm)	
serial	측정기 고유번호	
flag	실외 측정기 구분	
pm10	미세먼지 PM-10 데이터	
co2	이산화탄소 농도 데이터 (ppm)	실외 공기질 측정기에는 사용 하지 않음
vocs	휘발성유기화합물 농도 데이터	
noise	소음 데이터(db)	
temp	온도(°C) 데이터	
humi	습도(%) 데이터	
pm25	미세먼지 PM-2.5 데이터	

와이즈넷 대회용 제공 DB

“미세먼지”관련 수집된 블로그, 카페, 뉴스 본문을 2018.4월~2019.3월까지의 1년간 제공합니다.

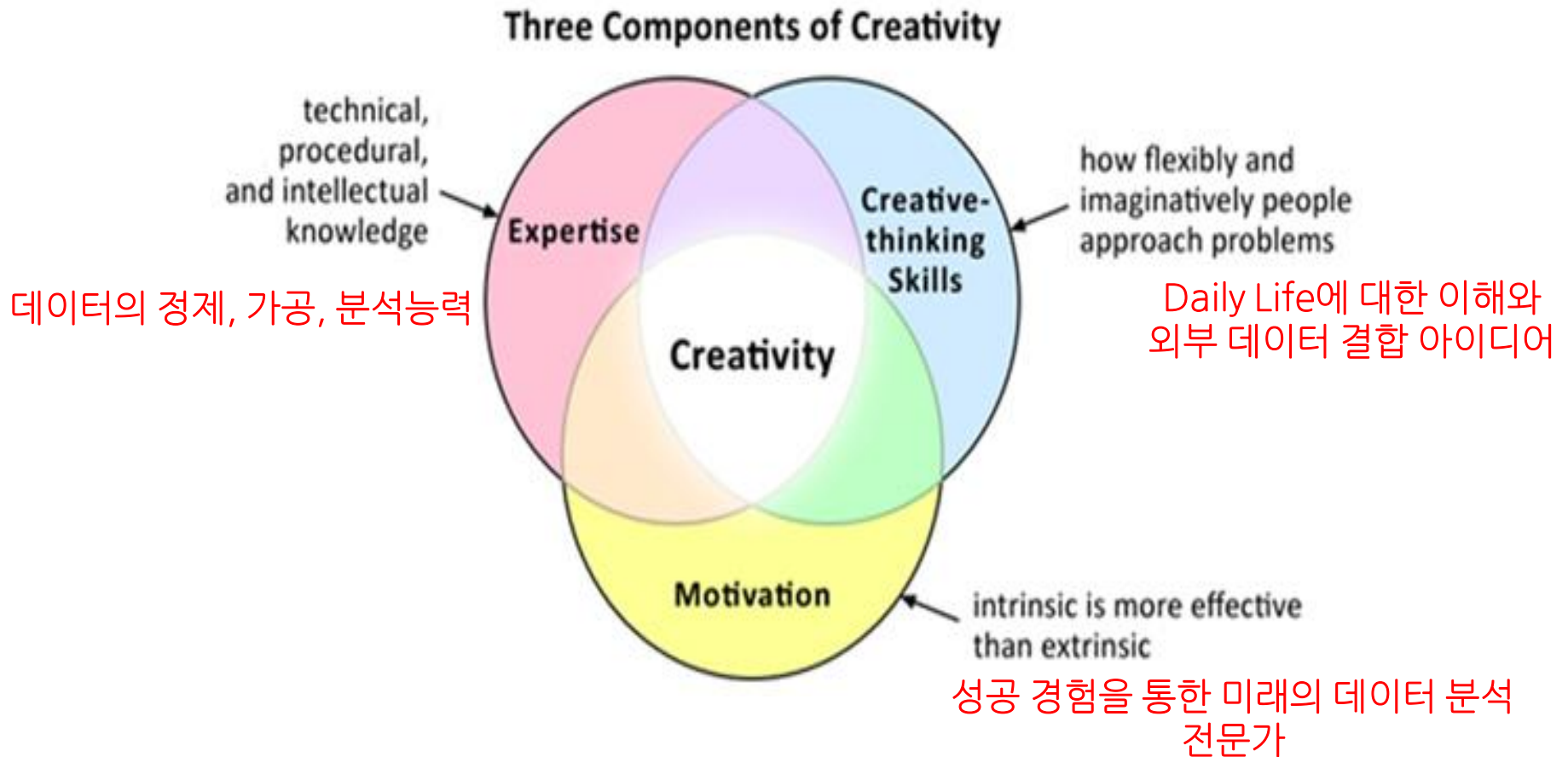
1

주어진 기간 내 블로그, 카페,
뉴스에서 발생한 본문 제공

컬럼ID	컬럼명	타입	NULL	KEY	DEFAULT	비고
DOCID	문서 KEY값	VARCHAR(8)	N	Y		ex) 12345678
REG_DT	문서 등록일	VARCHAR(14)	N	Y		ex) 20190401235959
SECTION	블로그 카페 뉴스	VARCHAR(8)	N	Y		블로그 : blog 카페 : café 뉴스 : news
TITLE	문서 제목	VARCHAR(256)	N	Y		
CONTENT	문서 본문	VARCHAR(4096)	N	Y		

출제의 의도

개방된 공공 데이터와 팀의 능력으로 구할 수 있는 모든 데이터를 상상하십시오
그리고 미세먼지가 제공되는 지역 내 사회 경제적으로 발생시키는 생활밀착형 인덱스를 개발하여 시민의 일상에 도움이 되는 numeric한 접근을 시도하길 바랍니다.



출제문제

“미세먼지의 사회적 영향 분석 및 비즈니스 아이디어 제시”



빅데이터를 활용한 “미세먼지의 사회적 영향 분석 및 비즈니스 아이디어 제시”

- 유동인구데이터(SK텔레콤), 카드매출데이터(신한카드), SNS데이터(와이즈넷), 환경기상데이터(케이웨더), 유통데이터(GS리테일), 공공데이터 등 다양한 데이터를 활용하여 미세먼지로 인한 소비/경제/행동변화에 따른 사회적 영향 분석 및 예측 모델링을 통한 비즈니스 아이디어 제시



외부 오픈 DB

다양한 데이터를 구할 수 있는 사이트를 소개 드립니다.

- 공공데이터포털 <https://www.data.go.kr/>
- 서울 열린 데이터광장 <http://data.seoul.go.kr/>
- 네이버 데이터 랩 <http://datalab.naver.com/>
- 오픈데이터포럼 <http://odf.or.kr/>
- 경기데이터드림 <https://data.gg.go.kr/>
- DATA STORE <https://www.datastore.or.kr/>
- ODPia <https://www.odpia.org/>
- 서울시설공단 https://www.sisul.or.kr/open_content/parking/guidance/info.jsp
- 오피넷 <http://www.opinet.co.kr/>
- SK텔레콤 빅데이터 허브 <https://www.bigdatahub.co.kr/>