

# 서울시 빅데이터 활용

## Application of Big Data Analysis by Seoul Metropolitan Government



강진동

### 서론

‘허니버터칩’ 열풍에 대해 알고 있는가? ‘허니버터칩’은 출시가 되자마자 돈 주고도 못사는 과자가 되어버렸다. 국내 식품시장에 월 매출 10억원만 넘겨도 대박 히트상품으로 꼽히는데 ‘허니버터칩’의 경우 출시 100일만에 매출 50억원을 돌파하는 초대박 제품이 되었다. 과연 어떻게 이런 성과를 낼 수 있었을까? 바로 정답은 “빅데이터” 때문이다. 빅데이터를 기반으로 감자칩의 주요 구매층인 10-20대 여성들의 취향을 분석해 제품 개발을 하였기 때문이다.

이처럼 세대별 음식 취향이라는 일상생활의 사소한 부분들까지도 우리 주변에는 규모를 가늠할 수 없을 정도로 많은 정보와 데이터가 생산되는 빅데이터(Big Data) 환경이 도래하고 있다.

### 빅데이터의 특징과 중요성

#### 1. 빅데이터의 특징 및 중요성

빅데이터란, 기존의 정보수집 및 분석 방법으로 정보를 처리할 수 없는 대규모의 데이터량을 의미하고, 그 데이터를 처리하는 기술을 상징한다.

이러한 빅데이터는 데이터의 양(Volume), 데이터 생성 속도(Velocity), 형태의 다양성(Variety) 3V의 특징을 갖고 있다. 이처럼 다양하고 방대한 규모의 데이터는 미래 경쟁력의 우위를 좌우하는 중요한 자원으로 활용될 수 있다는 점에서 주목받고 있다. 금융, 제조, 통신, 유통, 의료 등 다양한 산업군을 넘어 교통분야 에서도 중요한 역할을 하고 있다. 대표적으로, 빅데이터 분석을 통해 보행자 및 운전자의 통행패턴을 시공간적으로 분석하여 교통사고에 대한 원인 분석과 해결 방

강진동 : 서울특별시 교통운영과, vbration@seoul.go.kr, Phone: 02-2133-2450, Fax: 02-2133-1053

안을 모색하는 근거로 활용할 수 있다. 서울시에서는 빅데이터를 활용하여 교통안전대책과 택시운행 정보를 분석하여 현재 위치한 택시승강장의 효율성을 검토하였다.

## 서울시 빅데이터 활용사례

### 1. 교통약자 안전대책

#### 1) 분석개요

어린이 보행자의 교통사고, 어르신 보행자의 교통사고, 중앙버스전용차로 정류소 교통사고, 음주운전 사고, 위험운전 행동과 사고와의 관계 등 5개 분야에 대해 빅데이터 활용하여 분석을 하였으며, 이를 바탕으로 예방 대책을 세웠다.

#### 2) 분석방법

교통안전 시설물('12-'13년, 서울시), 도로별 차량속도('12년, 서울시), 교통사고 내역('11-'13년, 도로교통공단), 위험운전 행동 데이터('12-'13년, 교통안전공단), 유동인구('14년, SKT), 기상정보('13-'14년, 서울시·기상청)등 빅데이터를 융·복합하여 분석하였다. 빅데이터 분석은 교통사고 내역에서 사고가 일어난 장소와 시간 등을 우선 확인한 뒤 그 지역의 유동인구, 그 시간대의 날씨, 주변 교통안전 시설물 등을 복합적으로 분석하는 방식으로 이뤄졌다.

이를 위해 서울시는 지난 1년여 간, 빅데이터 수집 및 분석을 위한 TF팀을 구성하여, 각각 산재됐던 데이터간 연간관계를 분석하였다. 또한, 교통사고 패턴을 밝혀낸 뒤, 총 5개 주제별 분석 결과를 토대로 교통안전 대책을 수립하였다.

### 3) 분석 결과 및 대책

#### ① 어린이 보행자 교통사고 분석 결과 및 대책

어린이 보행자 교통사고의 58%는 초등학교 반경 330m 이내에서 발생하는 것으로 나타났으며,

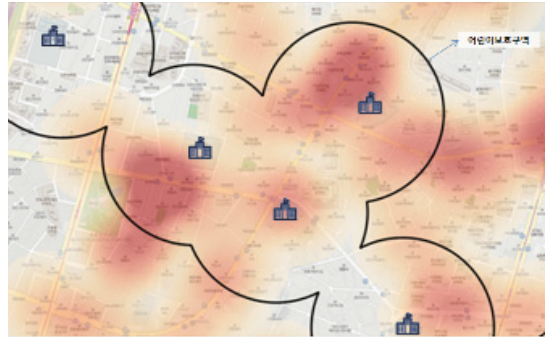


그림 1. 어린이보호구역 어린이교통사고 온도지수

초등학교에 갓 입학하는 만 7세 어린이의 사고 비율이 13.2%로 가장 높았다. 등교시간에 비해 방과 후인 오후 2시~저녁 7시에 1.5배 많이 발생했다. 남아의 교통사고 비율은 62%로 성인 남자의 교통사고 비율 49%에 비해 상대적으로 높았다.

방학 기간보다는 학기중에 교통사고가 상대적으로 많았으며 가을철보다는 신학기인 봄철, 특히 5월에 사고발생비율이 12.1%로 가장 많이 발생하는 것으로 나타났다.

'12년 한 해 서울 시내 신규 설치된 과속방지턱 62개소 반경 30m 이내에서 '11년과 '13년에 각각 발생한 교통사고를 비교한 결과, 전체 교통사고는 60%, 보행자 사고는 80%가 감소했다. 과속방지턱 10개당 연간 교통사고 1.6건이 감소한 효과가 나타났다.

높은 통행속도가 교통사고 발생의 주요 원인 중 하나라는 것은 주요근거이며 이면도로로 제한속도 하향 추진이 필요한 것을 알 수 있다.

또한, 서울시에서는 교통사고에 가장 많이 노출되어 있는 초등학교 1-2학년을 대상으로 교통안전 교육 시행을 위하여 교통안전 교육교재 및 교구를

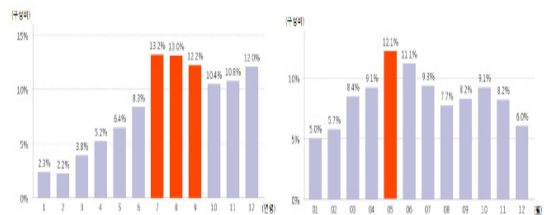


그림 2. 어린이 보행자 교통사고 연령 및 월별 분포

제작하여 시교육청의 협조를 받아 일선 학교에 배포하였다.

② **어르신 보행자 교통사고**

만 65세 이상의 어르신 보행자 교통사고는 중상 이상의 상해를 입는 경우가 65.4%로 일반인의 42.3%에 비해 1.5배 높아 어르신이 교통사고에 상대적으로 취약한 것을 알 수 있으며 전통시장이나 공원 인근 등 어르신이 주로 활동하는 지역에서의 사고 발생이 타 연령대에 비해 많은 것으로 조사되었다.

어르신보행자 교통사고 중 64.2%가 만65-74세에서 발생(65-69세 33.3%, 70-74세 30.9%) 하였고 여성 사고 비율이 남성에 비해 1.8배 높아 여성 어르신이 교통사고에 더 취약한 것으로 나타났다.

사고발생 시간대는 어르신들의 활동시간인 9-20시에 주로 발생하여 심야시간대에 주로 사고가 발생하는 일반적인 패턴과는 확연한 차이를 보

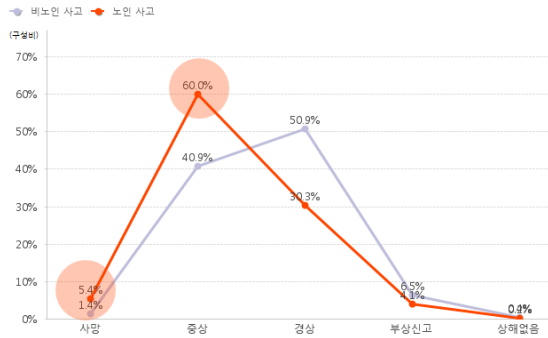


그림 3. 어르신 보행자 사고 상해정도 비교

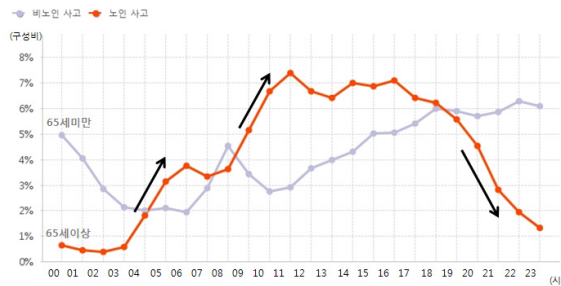


그림 4. 어르신 보행자 교통사고 시간대별 분포

였다.

어르신들의 사고발생의 가장 큰 원인은 무단횡단으로 어르신 보행자 사고가 많이 발생하는 지점에 무단횡단금지시설을 확대 설치하고 노인복지회관 등 찾아가는 교통안전 교육을 시행하여 어르신 교통사고 예방대책을 강화하였다.

장기적으로는 노인 및 장애인보호구역 지정요건에 전통시장, 공원 주변 등 어르신이 주로 활동하는 구역을 포함하여 실효성 있는 사업시행을 검토할 계획이다.

③ **중앙버스전용차로 주변 무단횡단 교통사고**

중앙버스정류소의 주변의 교통사고를 분석하기 위하여 50m×50m의 셀단위 분석, 셀에 포함된 사고내용을 텍스트 마이닝하였다.

분석셀내의 교통사고를 추출하고 텍스트마이닝을 통해 무단횡단 여부를 판단하였다.

최근 3년간 교통사고 자료 분석결과 중앙버스정류소 1개소당 무단횡단 사고 발생건수는 0.81건으로 일반 가로변 정류소에 비해 5.4배가 더 발생하는 것으로 분석되었다.

중앙정류소의 무단횡단 교통사고를 예방하기 위하여 무단횡단금지시설을 적극적으로 설치하고 있다.



그림 5. 중앙정류소 교통사고 분석방법

표 1. 정류소 유형별 무단횡단사고 건수

정류소 유형	정류소 수	무단횡단 사고 건수	1개소당 무단횡단 사고 건수
중앙 정류소	340	275	0.81
가로변정류소	10,196	1,491	0.15



그림 6. 중앙정류소 무단횡단 사고 온도지도

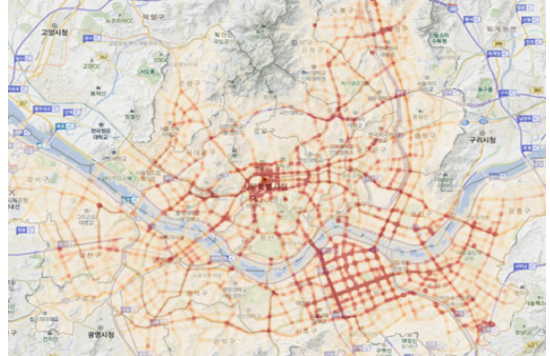


그림 8. 위험운전행동 다발 지점

'14년 개통된 충청로 중앙버스전용차로의 경우 전구간에 설치하였고 교통사고가 빈번하게 발생하는 중앙버스전용차로 구간에 무단횡단금지시설을 설치하였다. 장기적으로는 전체 중앙차로구간에 무단횡단금지시설을 설치할 예정이다. 그리고 중앙버스전용차로 구간을 운행하는 운수종사자를 대상으로 설문조사 등을 시행하여 개선대안 마련을 추진 중이다.

#### ④ 위험운전행동 분석

서울시 노선버스, 택시 등에 부착되어 있는 디지털 운행기록장치(DTG)를 통해 운전자의 급정거, 급출발, 급차로변경 등 위험운전행동과 교통사고와의 상관관계를 분석하였다.

분석결과 교통사고는 위험운전행동이 많은 지역에서 그리고 위험운전행동이 많은 시간대에 주로 발생하는 것으로 나타났다.

위험운전행동으로 인한 교통사고를 예방하기 위하여 택시, 버스 업체별, 운전자별 위험운전행동을 분석하고 맞춤형 교육을 시행하고 있다. 또한 위험운전행동 및 교통사고가 다발하는 지점을 선정하

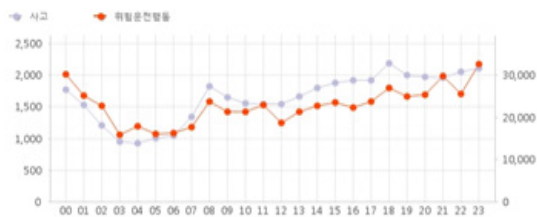


그림 7. 위험운전행동과 교통사고의 상관관계

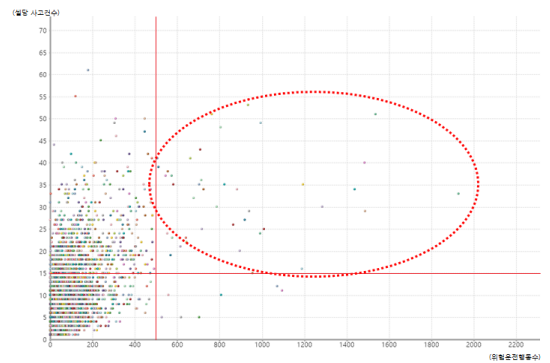


그림 9. 교통사고, 위험운전행동 다발지점 분석

여 교통안전시설 설치 등 사고예방활동을 강화할 예정이다.

## 2. 택시운행 패턴 분석

### 1) 분석개요

서울시는 주요 민원사항 중 하나인 택시서비스를 개선하기 위하여 택시승객과 택시운전자간의 탑승정보를 공유하는 택시매치메이킹(taxi matchmaking)을 활용하여 택시승강장 정비의 기초자료를 마련하기 위한 분석을 시행하였다.

### 2) 분석방법

“택시매치메이킹” 데이터셋을 구축하는 과정에서 택시의 위치 데이터를 표준노드링크에 맵핑하여 방향성을 부여하였고, 실차·공차 정보로부터



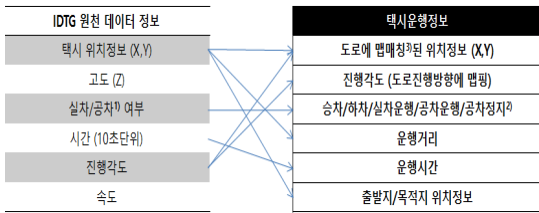


그림 10. 원천정보 및 가공정보 처리과정

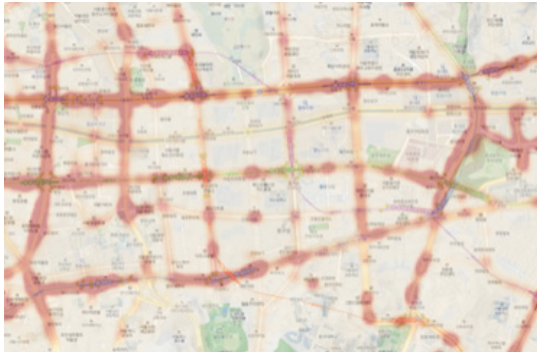


그림 11. 택시승차 온도지도(종로, 을지로)

승차·하차, 승차·공차운행, 공차정지의 상태 데이터와 한 승차행위 당 운행거리, 운행시간, 하차위치 등의 부가적인 정보를 추출하였다.

택시 승하차정보의 시각화를 위하여 온도지도를 활용하였으며 표준노드링크상에 승하차 정보를 맵핑하여 분석에 활용하였다.

### 3) 분석결과

택시의 시간실차율은 평균 50%로 나타났으며 시간대별로는 출근시간대, 자정 무렵 60% 이상으로 가장 높은 실차율을 보이는 것으로 나타났다.

시간대별 승차건수를 분석한 결과 크게 3가지 패턴을 보이고 있었다. 심야시간에 승차가 많은 일봉형, 출퇴근시간과 심야시간에 승차가 많은 쌍봉형, 새벽시간대를 제외하고 고른분포를 보이는 지점으로 나눌 수 있다.

택시승차대가 시민들에게 이용되는 지표로서 승차대 전후 50m이내 승차건수와 승차대 전후 200m이내 승차건수를 나눈 백분율인 승차대별 활용률을 활용하였다. 승차대별 평균활용률은

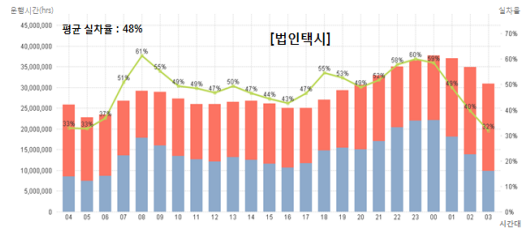
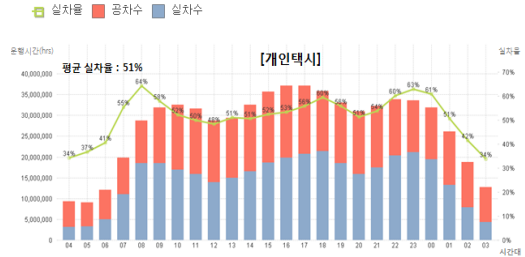


그림 12. 시간대별 실차율

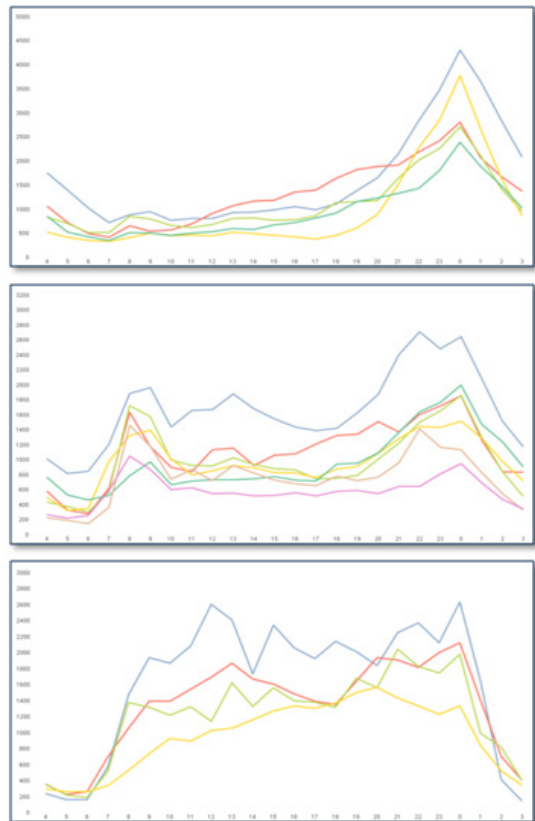


그림 13. 승차건수 기준 승차 패턴

32.5%이고 50% 이상의 활용률인 승차대는 전체 승차대의 19%에 불과한 것으로 나타났다.

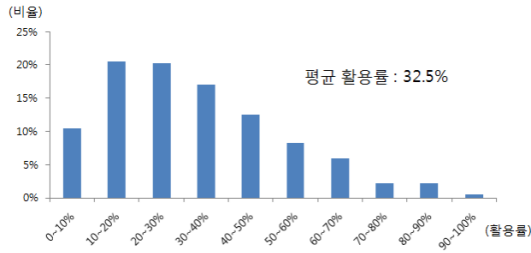


그림 14. 택시승차대 활용률

택시승차대의 적정위치 선정하기 위하여 표준노드링크 상 50m단위로 구분하여 분석을 시행하였다. 택시승차장정보를 살펴본 결과 일평균 승차건수가 1,000회 이상이 곳이 508개소이었고 그중 200m이내에 택시승차대가 없는 곳은 424개소에 달했다.

분석결과를 토대로 택시승차대 설치필요지점에 대해서 매핑을 시행하였다.

일례로 여의도역 인근의 택시정류소의 경우 기존승차장의 이용률이 낮아 인근지역으로 택시정류소의 이동이 필요한 것으로 분석되었다.

택시 승차장 기준의 빅데이터 분석결과는 향후 지역별 시간대별 택시공급 방안, 택시승차대 적정위치 선정에 활용이 가능하다. 다만, 도로별 교통량 등 교통여건, 토지이용 등 지역여건에 대한 충분한 고려가 필요하며 또한 택시부제 변경 등 제도적, 행정적인 부분에서의 충분한 검토가 선행되어야 한다.

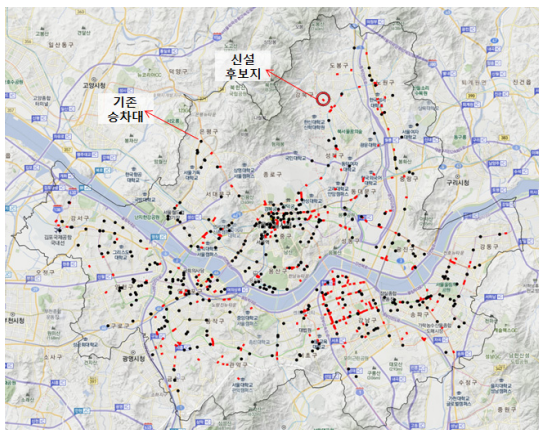


그림 15. 기존 승차대 위치 및 신설 필요지점



그림 16. 여의도역 인근 택시승차대 입지분석

## 향후 전망

교통 분야는 오랜기간 관련 데이터 축적이 이루어져 빅데이터 분석이 다른 분야에 비해 상대적으로 용이한 편이다. 70년대 신호운행을 시작으로 90년대말부터 ITS사업이 본격적으로 시행되었다. 2000년대 이후 교통안전시설물, 주차장 정보, 버스운행정보 등이 GIS맵과 연계되어 관리되고 있다. 이러한 정보를 유의미한 결과로 해석하고 정책에 반영하기 위해서는 빅데이터 분석 방법론 정립이 필요하다. 현행 데이터는 고유한 목적에 따라 구축되어 있어 교통상황과 지역여건 등을 고려할 수 없다. 그리고 데이터의 정밀도도 구축시스템별로 상이하다. 빅데이터 분석을 지속적으로 시행하고 노하우를 축적해야 하는 이유이기도 하다.

서울시는 '13년 심야전용 '울빼미버스' 9개 노선 선정에 30억건의 통화량 빅데이터 분석을 활용하는데 이어 교통뿐만 아니라 안전, 복지 분야 정책에도 빅데이터 분석을 본격 도입하기 시작했다. 이를 위한 전문부서도 신설하여 본격 가동중에 있다. 교통약자를 위한 교통사고 분석, 택시운행분석, 인생이모작 지원센터 입지선정, 노인여가복지시설 입지선정, 중국관광객 유동인구 분석 등에 빅데이터 분석 결과를 적극적으로 활용하고 있다.

또한, 향후에도 보행 및 자전거 도로설치, 공공시설물 입지 선정, 대기오염 지역별 특성, 시설물별 안전사고 유형, 지역 문화기반 일자리 분석 등 다양한 분야에 빅데이터 분석이 시행되고 이를 통해 의사결정 및 정책에 반영될 예정이다.