

산더미 같은 데이터를 다루는 것은 그다지 우아한 일이 아니다. 특히 신용카드 기록처럼 잘 정리된 데이터가 아닌, 사람이 직접 입력한 데이터일 때 상황은 더 악화된다. 본격적인 분석에 들어가기에 앞서 데이터에 포함된 무수한 결점들을 바로잡는 일은, 마치 금이 있는지도 없는지도 모르는 채로 끝도 모를 지하 속 깊숙히 파고들어가는 것(mining)과도 같다. 간단히 말하자면, 그것은 매우 힘들고 오래 걸린다.

이번 전시에 사용된 건축물 대장(building account) 데이터가 그러했다. 서울에 존재하는 약 60만 동의 건물 데이터가 공개되어 있다는 것은 매력적인 사실임에 틀림없었으나, 그것은 온전히 사용할 수 있는 상태가 아니었다. 우선 전체의 50%가 용적률 항목을 지니고 있지 않았으며, 대지면적, 건폐율, 연면적 등이 마치 별레가 제멋대로 뜯어먹고 지나간 잎처럼 처참하게 비어 있었다. 이미 철거되어 사라졌음이 분명한 - 그 자리에는 아파트 단지가 지어져 있다 - 작은 건물들 수십여동이 제거되지 않고 남아있는 경우도 있었으며, 아파트 단지 전체의 연면적이 단지의 각 동에 반복적으로 기입되어 있는 경우 - 결국 전체 연면적은 열 배 이상 늘어나기도 한다 - 도 있었다. 이러한 오류들은 때로는 걸러내는 코드를 만들어서, 때로는 수동적으로 하나하나 수정해가며 바로잡았지만 사실 오류는 여전히 남아 있다. 그러나 괜찮다. 사실 서울의 건물은 무척 많기 때문에, 몇몇 개별적인 오류들은 전체 관점에서 아주 미미한 영향을 미치지 때문이다.

그런데 그렇지 않은 오류도 있었다. 매우 흔하면서 엄청난 결과의 차이를 가져다 주는 실수가 있었는데, 그것은 바로 콤마와 소수점이 혼동된 경우였다. 소수점을 찍을 자리에 콤마를 찍어서 1000배나 10000배가 되는 경우도 꽤 있었다. 한 예로, 서울 어느 곳의 작은 아파트 단지 지하에는 매우 넓은 지하주차장이 있다. 얼마나 넓은가 하면, 무려 면적이 9천만제곱미터에 달한다. 참고로 말하자면, 서울 전체 건물의 연면적 총 합계의 1/6 정도나 된다. 서울시에서 운영하는 공식 통계 웹사이트에도 이 면적이 포함된 서울 전체 건물 연면적이 게시되어 있다. 사실 모를 일이다. 실제로 그 건물 지하에 서울시가 유사시를 대비하여 비밀리에 건설한 지하도시 입구가 있을 수도 있으니까.

소수점을 이동시켜 지하도시들 마저 함락시키고 나면, 비로소 양질의 건축물 대장 데이터가 준비된다. 그렇다고 그게 끝은 물론 아니다. 분석과 매핑 작업을 진행하기 위해, 각기 다른 기관에서 관리하는 세 종류의 데이터 꾸러미를 연결시켰는데, 건축물 대장 데이터와 더불어 대지경계선 정보가 있는 지적도, 그리고 건물 polygon이 있는 GIS 데이터들이 그것들이다. 세 종류의 데이터를 연결시킬 수 있는 유일한 단서는 바로 주소인데, 한 주소에 여러 건물이 있는 경우는 그 연결이 무척 까다로워진다. 우여곡절 끝에 드디어 본격적 분석을 위한 준비 작업이 모두 완료되었다. 그래도 바로 그 지점은 금맥을 발견한 순간과도 같다. 이제 그곳부터는 시간이 허락하는 만큼 의미 있는 결과를 마이닝할 수 있다. 어렵게 산을 오른 후 정상에 도달하여 원하는 만큼 먼 곳을 한눈에 바라볼 수 있는 순간과도 같다.

그래서 그 곳에서 무엇이 보이는가? 한 줄 한 줄의 데이터, 즉 하나하나의 건물들은 개별적 사람들의 작은 욕망들이다. 욕망들이 아주 많이 모이면 사회의 거시적 흐름이 된다. 혹은, 건축을 제한하는 규제들이 점들의 불균질한 분포를 통해 명확하게 드러나기도 한다. 그리고 그런 거시적 흐름 속에서도 흐름을 벗어나려하는 특별한 individual의 미시적 자유의지들이 관찰되기도 한다.

용적률과 같은 강력한 제도들은 분석이 만들어낸 다이어그램 곳곳에 뚜렷한 흔적들을 남긴다. 특히 정부의 정책에 의해 시행되는 강력한 regulation은 그 발효시점 이전에도 특별한 흐름을 야기시키는데, 재미있는 부분은 이러한 regulation 강화가 아이러니컬하게 건설 부양책 되기도 했다는 점이다. 서울에 현존하는 단독주택(detached house)의 13%는 1990년 한 해에 지어졌다. 1991년부터 연면적 200제곱미터 이상의 단독주택은 주차장 설치가 의무화 될 예정이었기 때문에, 규제의 발효 이전에 규제를 피해서 잠재적 수요자 대부분이 집을 지은 것이다. 이것은 데이터 더미에서 의미를 마이닝한 아주 작은 사례에 불과하다. 데이터 더미들을 헤쳐나가면서 개별적인 자유의지가 모여서 만들어내는 크고 작은 의미들을 함께 찾아나서보자.