

<0909 빅캠 회의>

I. 파일에 대한 내용

1. 파일명 이름은 난수로 생성 (이용자, 경로와 관련없음)
ex. 한 스마트폰 이용자가 여러 경로를 탔다해도 파일명 정보로 알 수 없음.
2. xml 내의 변수의 개수가 다른 건 데이터가 들어오다 끊겼을 수도 있기 때문.
3. 오류 난 데이터 있을 수 있음
4. 2011년.11월~2015년.12월 데이터 받을 것
-11년, 12년 데이터는 한 폴더에 들어있음
-BOAZ측은 1년치 데이터만 받아 분석할 예정 & 나머지 4년치는 빅데이터 캠퍼스에서 VDI로
5. 테이블 변수의 정의가 명확하지 않아 개발자 측에 “테이블 정의서” 요청할 예정
6. 사용자 ID, 파일이름 매칭 가능할 수도!→ 개발업자쪽과 미팅 필요 (9월 말~10월 초 쯤)
& 테이블 정의서 필요
6. 저파일/ 고파일의 차이: 용량 차이

II. 분석방법에 대한 내용

1. 출발도착을 “월”단위로 잘라서 분석
2. 오전/ 오후시간 패턴(출근, 퇴근시간 패턴)& 전체 하루 패턴 & 평일/ 주말 패턴
- 주말은 시간을 Full로 하면 되고
- 평일 오전/ 평일 오후/평일 하루의 평균 거리는 어느 정도인지, 이동 패턴은 어떠한지
3. 출발지- 도착지는 행렬로 만들어주기 (OD matrix)
4. 경기도외곽에서 서울로 넘어오거나 그 반대의 경우 조인할 방법 구상 필요
주무관님> 보더를 넘는 것만 따로 잘라내서 비중이 어느 정도 되는지 계산하고, 비중이 높으면 따로 떼어 분석해야 함
(Categorizing)
5. 장거리 riding은 대략 20km 기준 (출/퇴근용)
- 전체 내에서 20km가 넘는 경로가 몇 프로인지 정하기
- 이동 패턴에 대한 분석 필요 (EDA): 단거리/ 중거리/ 장거리로 나오는지 or 단거리/ 장거리로 나오는지
- total_distance 변수의 단위를 알게 되면 가능할 듯

III. 녹색교통측 요구사항

1. 출발, 도착 빈도표 (1순위)

2. 시설 정보(자전거 보관소), 도로망 정보랑 매칭 → 신수현 주무관님과 나중에 논의 예정

ex. 자전거 도로망이 있는데 진짜 사람들이 많이 이용하는지

ex. 자전거 도로망이 없는데 사람들이 많이 이용한다면, 거기에 도로망을 설치

IV. 주무관님 조언

1. 래스터화 후 밀도지도로 만들 수 있음

OR 행정구역이 아닌 작은 블록으로 만들어서 그 구역을 얼마나 지나는지

OR 기존 경계를 사용하지 않고 hexagon을 만들어서

2. 경로 중 튀는 것 cleansing 필요:

ex. “route가 A, B 간의 timestamp를 기준으로 순간속도가 45이상”이라면 말도 안 되는 것!

3. 파일이 똑같이 들어가는 경우 → 유사도 판별 알고리즘 이용해야함

ex. A라인과 B라인의 유사도 98%라면 하나를 날리는 식 !!

4. 파이썬으로 여러 kml을 csv로 바꾼 후, 일괄적으로 shp로 바꾸는 작업해야 함

5. Gis로 좌표계를 뽑아내서 구 안에 들어가는지/ 보더를 넘는지 체크

6. 추가분석

- 전체 경로 중 강 주변으로 다니는 riding은 몇프로이고/ 산은 몇프로이고

7. 전체데이터를 받고 나면 Meta EDA필요 (데이터 자체의 속성들)

ex. 파일 자체의 용량은 얼마나 되는지

8. kml을 올려놓고 서울시의 Extent를 잡아서 kml 파일의 출발, 도착 좌표 or 전체 좌표 x, y 자체에 Extent에 들어가는 것만 잡아서 먼저 rendering! *뒤에 자세한 설명 있음

V. Q&A

1. 좌표계 UTM-K(GRS 80 타원체)를 WGS84로 통일하는 방법

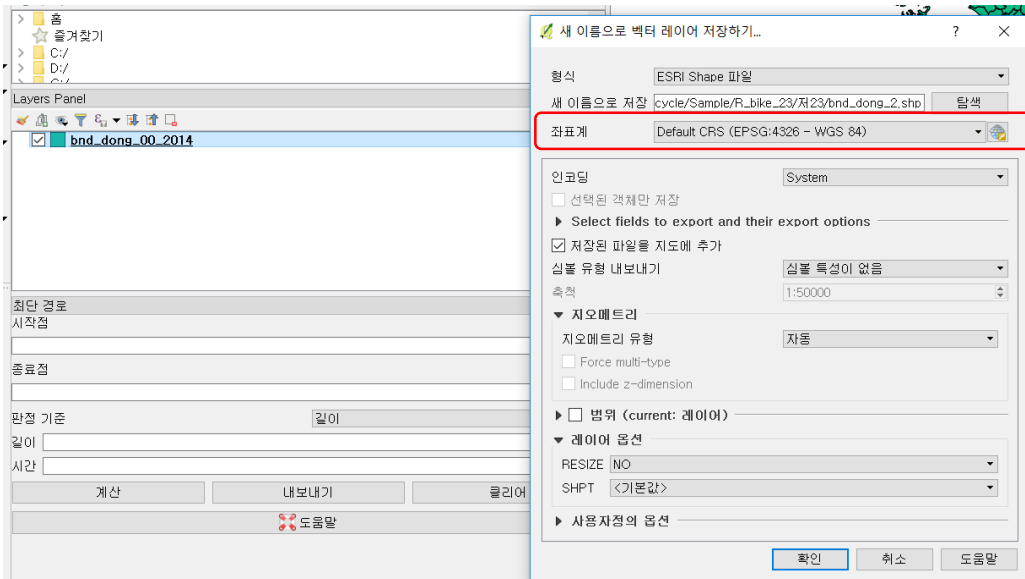
1st. bnd_dong 좌표계 utmk(grs80).prj 다운로드

➔ SGIS 통계청 사이트에서 다운함! (카톡의SOP_prj_utm.zip 파일)

2nd. .prj파일의 이름을 지도 파일 이름 (xxx.shp) 과 같게 저장

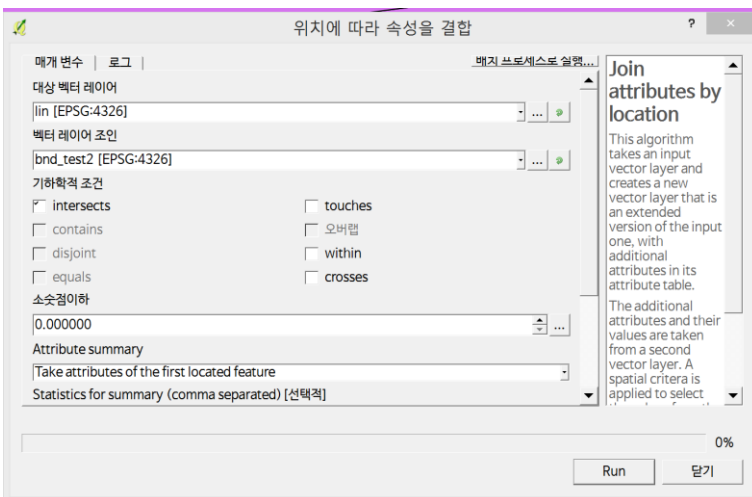
3rd. Qgis에서 불러온 후 지도 shp 파일을 불러온 후, 우클릭해 “ 다른 이름으로 저장”

➔ 좌표계를 WGS84로 바꿔서 저장



4th. 좌표계가 WGS84로 바뀐 지도파일과 경로 csv파일 불러오기

5th. 공간결합: 대상 벡터레이어에 경로, 벡터 레이어에 지도 파일 넣고, 기하학적조건 intersects



6th. 추가된 Joined Layer를 shp파일로 다른 이름으로 저장

2. 데이터 용량이 너무 커서 다룰 수 있는지?

- 전체 Route 정보 대신 시작점, 출발점정보만 분석할 수도 있음

➔ 시작점에 대한 정보 csv, 종점에 대한 정보 csv를 따로 만들어서 지도 shp파일과 공간 join을 각각 수행

3. 서울의 Extent 잡는 법?

1st. 서울만 있는 지도 shp를 가져와서 우클릭

2nd. 속성 → 일반정보 → 공간적 지표생성 → 메타데이터 → xmin, xmax, ymin, ymax를 따 (좌표값의 min, max) 서울시를 포함하는 큰 네모박스를 만들어줌!! 그게 바로 Extent!

OR 각 attribute의 xmin, xmax, ymin, ymax를 계산할 수도 있음!

속성테이블 -> 필드계산기 -> 출력필드 이름 test로 적고 ->표현식 “지오메트리” (➡ 이부분 막힘)

OR : wiki에서 다수의 폴리곤의 min, max를 찾는 법 찾기!