

[화성시] 과제 - 데이터 항목(필드)설명

■ Data

주어지는 데이터 파일은 다음과 같습니다.

• PJT001_TripChain.csv	2018년 7월 승차일기준 1~4일의 버스카드 태깅정보를 담고있습니다.
• PJT001_stations_table.csv	17~18년기준, 경기도버스 정류장에 대한정보를 담고있습니다.
• PJT001_sk_emd_od.csv	2018년7월1~4일의 행정동별 이동 인구수 정보를 담고있습니다.
• PJT001_tl_scco_emd.geojson	화성시의 읍면동 정보를 담고 있습니다.
• PJT001_h_100m_cell_flow.geojson	2018년 특정일 기준.100mx100m 그리드에 시간별 유동인구정보를 담고 있습니다.
• PJT001_h_100m_cell_pop.geojson	2019년 기준.100mx100m그리드에 거주 인구정보를 담고있습니다.
• PJT001_moc_link_2018.geojson	2018년기준 화성시 도로네트워크 링크정보를 담고있습니다.
• PJT001_routestationinfo.csv	버스노선-정류장 매핑 테이블
• PJT001_routestationmapping.csv	버스노선 매핑 정보

■ Data Field

각 데이터 파일이 담고있는 열(Field, Feature)는 다음과 같습니다.

<< PJT001_TripChain.csv >>

직접 EDA 하시면서 파악할 수 있는 열들이라, 구체적인 설명은 생략합니다.

다음과 같은 열들이 있습니다.

'암호화카드번호',
'트랜잭션ID',
'환승횟수',
'교통카드발행사ID',

'총이용객수',
'사용자구분',
....
'하차역ID5',
'최초승차역ID',
'최종하차역ID',
'총이용금액',
'수집건수',
'트립체인완료코드'

카드 태깅 안하고 내리기, 전산 미스 등등의 이유로 결측치가 존재합니다.

<< PJT001_stations_table.csv >>

각 행은 정류장 1개의 정보를 담고 있습니다.
누락된 정류장 자체가 누락되어 있을 수도 있고, 특정 열의 값이 결측 되어있을 수도 있습니다.

다음과 같은 열들이 있습니다.

'표준정류장ID',
'시군명',
'정류소명',
'정류소영문명',
'정류소번호',
'중양차로여부',
'관할관청',
'위치',
'WGS84위도',
'WGS84경도',
'모바일정류장ID',
'이비카드정류장ID'

일반적으로, 표준정류장 ID 로 각각의 정류장을 구분합니다.
한편, PJT001_TripChain.csv 에는 각 정류소 ID 값으로 이비카드정류장ID가
주로 등장합니다.
PJT001_TripChain.csv 을 정류장 단위로 세세하게 살펴보고 싶을 때 사용하시면 됩니다.

<< PJT001_moc_link_2018.geojson >>

각 행은 노드와 노드를 잇는 하나의 링크 정보를 담고있습니다.

다음과 같은 열들로 이루어져있습니다.

```
'id',  
'link_id',  
'f_node',  
't_node',  
'road_use',  
'lanes',  
'road_rank',  
'road_name',  
'multi_link',  
'connect',  
'max_spd',  
'rest_veh',  
'rest_w',  
'rest_h',  
'remark',  
'user_id',  
'workstate',  
'dept_code',  
'stnl_reg',  
'road_type',  
'road_no',  
'tmpid',  
'upload_id',  
'sosfnodeid',  
'sostnodeid',  
'shape_stle',  
'geometry'
```

geojson 형식의 파일이기 때문에, 파이썬의 경우, geopandas.read_file (...) 을 사용해서 데이터프레임 형태로 불러올 수 있습니다.

예를 들어 다음과 같이 불러올 수 있습니다.

```
import geopandas as gpd
```

```
df = gpd.read_file('../Hwaseong_bus/data/raw/PJT001_moc_link_2018.geojson')
```

```
<< PJT001_100m_cell_flow.geojson >>
```

각 행은 그리드 1개의 정보를 담고 있습니다.

열은 다음과 같이 이루어져있습니다.

'id' : 그리드 각각의 id 입니다.

'ymd' : 측정된 날짜 입니다. 20171003 식으로 되어있습니다.

't00_01' : 00-01 사이의 해당 그리드의 유동인구입니다. 아래도 모두 동일합니다.

't01_02',

't02_03',

't03_04',

't04_05',

't05_06',

't06_07',

't07_08',

't08_09',

't09_10',

't10_11',

't11_12',

't12_13',

't13_14',

't14_15',

't15_16',

't16_17',

't17_18',

't18_19',

't19_20',

't20_21',

't21_22',

't22_23',

't23_24',

'emd_cd',

'geometry' : 그리드의 지리적 정보, 즉 Polygon 모양을 담고있습니다.

.geojson 파일이기 때문에, 파이썬의 경우 마찬가지로 geopandas.read_file() 로 불러올 수 있습니다.

예를 들어 다음과 같이 불러올 수 있습니다.

```
import geopandas as gpd
```

```
df = gpd.read_file('../Hwaseong_bus/data/raw/pop/PJT001_h_100m_cell_flow.geojson')
```

<< PJT001_h_100m_cell_pop.geojson >>

각 행은 그리드 1개의 정보를 담고있습니다.
열은 다음과 같이 이루어져있습니다.

'ID' : 그리드 각각의 id 입니다.
'EMD_CD',
'POP_MN_0_9' : 해당 그리드의 MN(남자), 0_9(0-9세)의 거주 인구입니다.
'POP_MN_10_' : 위의 방식과 동일합니다.
'POP_MN_20_',
'POP_MN_30_',
'POP_MN_40_',
'POP_MN_50_',
'POP_MN_60_',
'POP_MN_70_',
'POP_MN_80_',
'POP_MN_90_',
'POP_MN_TTL' : 해당 그리드의 MN(남자), TTL(Total) 총 거주 인구 입니다.
'POP_WMN_0_' : 해당 그리드의 WMN(남자), 0_9(0-9세)의 거주 인구입니다.
'POP_WMN_10',
'POP_WMN_20',
'POP_WMN_30',
'POP_WMN_40',
'POP_WMN_50',
'POP_WMN_60',
'POP_WMN_70',
'POP_WMN_80',
'POP_WMN_90',
'POP_WMN_TT',
'POP_TTL',
'geometry' : 그리드의 지리적 정보, 즉 Polygon 모양을 담고 있습니다.

<< PJT001_sk_emd_od.csv >>

각 행은 6시간별 행정동간 인구 이동 정보를 담고 있습니다.
열은 다음과 같이 이루어져있습니다.

'std_ymd' : 기준일자입니다.
'tm_trm' : 시간정보입니다. 00_05는 0시에서 6시 바로 이전을 의미합니다.
'org' : 출발 행정동을 의미합니다.

'dst' : 종료 행정동을 의미합니다.
'pop' : 인구 이동 수를 의미합니다.

<< tl_scco_emd.geojson >>

각 행은 화성시의 읍면동 정보를 담고있습니다.
열은 다음과 같이 이루어져있습니다.

'id' : 각 읍면동의 id입니다.
'emd_cd' : 각 읍면동의 8자리 행정구역 코드입니다.
'emd_eng_nm' : 각 읍면동의 영문 이름입니다.
'emd_kor_nm' : 각 읍면동의 한글 이름입니다.
'geometry' : 읍면동의 지리적 정보, 즉 Polygon 모양을 담고 있습니다.

<< PJT001_routestationinfo.csv >>

버스노선-정류장 매핑 정보이며
열은 다음과 같이 이루어져있습니다.

'seq' : 순번.
'pr_station_id' : 노선 ID.
'bus_line_no' : 버스 노선 번호.
'bus_line_no_seq' : 버스 라인 정류장 순서.
'station_nm' : 정류장 명칭.
'station_id' : 표준정류장ID.
'mobile_no' : 모바일정류장ID.

<< PJT001_routestationmapping.csv >>

버스노선 매핑 정보이며
열은 다음과 같이 이루어져있습니다.

구분 : 구분
운수사명 : 운수사명
운수사ID : 운수사 ID
이비노선ID : 이비노선ID(9자리)
표준노선ID : 표준노선ID(8자리)
노선명 : 버스 노선명

- 끝 -