

집계구별 생활인구 데이터의 격자 할당

▶ 활용 데이터 - 좌표계는 분석 과정에서 모두 5179로 통일

데이터	출처	형태
서울시 생활인구 데이터(내국인) - 집계구별	서울 열린데이터광장	.csv
서울시 집계구 경계 데이터	통계청 통계지리정보서비스	.shp
건물정보 데이터	도로명주소 안내시스템	.shp

▶ 생활인구 데이터 전처리 - R

- 2020년 11월 1일부터 2021년 10월 31일까지 일별로 집계된 데이터를 R에서 읽어들이고
필요한 컬럼을 추출하여 월별 데이터프레임 생성 및 행동동 코드를 기준으로 구로구 데이터 추출
(기초데이터 csv파일)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	20201101	0	11110515	1.1E+12	508.1686	25.5182	16.2603	21.5532	18.8104	21.1558	18.364	9.8416	16.1391	22.2531	19.4574	21.9157	11.6084	7.0801	17.2896	14.7292	8.8253
2	20201101	0	11110515	1.1E+12	799.9027	28.2966	18.0306	21.0072	20.2751	34.3375	29.0928	36.3357	18.9607	30.8776	29.3098	19.2116	16.9571	20.882	37.3312	22.9702	13.7631
3	20201101	0	11110515	1.1E+12	503.3923	33.9685	21.6448	29.1582	17.2546	12.102	5.6619	18.8367	17.9713	23.8982	12.703	23.7326	9.9973	5.3101	24.3573	28.5284	17.0934
4	20201101	0	11110515	1.1E+12	240.7788	12.0973	7.7084	8.2987	5.5018	9.5871	9.6585	9.4649	4.7948	9.177	5.4118	8.6896	8.5086 *		10.1886	14.6318	8.7669
5	20201101	0	11110515	1.1E+12	557.8979	35.6212	22.6979	30.5451	19.6565	15.7057	9.4683	18.192	19.5826	26.114	15.8758	26.0206	11.2555	6.3511	24.9263	27.7174	16.6075
6	20201101	0	11110515	1.1E+12	207.6278 *			5.525 *		5.8918	5.7873	9.8302	7.6614	12.2195	6.4854	4.6282	4.0786 *		13.5665	11.6896	7.0041
7	20201101	0	11110515	1.1E+12	252.5944	7.5445	4.8074	8.076	4.0837	8.9229	8.8028	10.3283	7.0218	11.8816	6.8946	7.8595	7.0537 *		12.8746	14.2415	8.5331
8	20201101	0	11110515	1.1E+12	680.1248	34.9353	22.2609	29.9569	26.4808	26.8181	24.6965	11.1041	22.5011	30.3667	26.8617	30.4966	14.5894	9.7732	22.1287	17.606	10.549
9	20201101	0	11110515	1.1E+12	836.0847	16.2715	10.3682	18.897	19.762	41.3378	29.7904	44.6607	18.9858	29.4864	28.6957	19.1047	19.1339	19.0595	42.4916	24.3412	14.5846
10	20201101	0	11110515	1.1E+12	725.0998	40.7412	25.9604	34.9354	27.2084	26.6349	20.9419	16.3855	24.5599	32.9882	25.5675	33.0266	15.2033	9.5899	26.9881	25.4163	15.2287
11	20201101	0	11110515	1.1E+12	263.1219 *			7.5543 *		8.197	7.9221	11.6516	8.7026	13.966	7.8642	6.7246	6.2187	4.0314	15.1033	14.3451	8.5952
12	20201101	0	11110515	1.1E+12	264.2431 *			7.0299 *		7.4966	7.3636	12.5076	9.7481	15.5477	8.2519	5.8888	5.1895	4.5869	17.2616	14.8735	8.9118
13	20201101	0	11110515	1.1E+12	386.5285	19.5298	12.4444	15.2262	12.0151	15.8177	14.714	10.6013	10.3789	16.0911	12.1227	15.6735	10.8365	4.2994	14.447	16.4795	9.874

(월별 데이터프레임 생성 및 구로구 데이터 추출 코드)

```

1 setwd("C:/users/user/Desktop/공직/구로구청/업무 파일/ccvtv 분석/기초데이터/10. 생활인구/구로구 생활인구(집계구)")
2 d_11 <- data.frame(); d_12 <- data.frame(); d_01 <- data.frame(); d_02 <- data.frame()
3 d_03 <- data.frame(); d_04 <- data.frame(); d_05 <- data.frame(); d_06 <- data.frame()
4 d_07 <- data.frame(); d_08 <- data.frame(); d_09 <- data.frame(); d_10 <- data.frame()
5
6
7 for(i in 1101:1130){
8   df <- read.csv(paste0("LOCAL_PEOPLE_202011/LOCAL_PEOPLE_2020",i,".csv"), head=TRUE)
9   colnames(df) <- c("date", "hour", "d_code", "j_code", "total", "m_0", "m_10", "m_15", "m_20", "m_25", "m_30", "m_35", "m_40", "m_45",
10     "m_50", "m_55", "m_60", "m_65", "m_70", "w_0", "w_10", "w_15", "w_20", "w_25", "w_30", "w_35", "w_40", "w_45",
11     "w_50", "w_55", "w_60", "w_65", "w_70")
12   d_11 <- rbind(d_11, cbind(df[,1:5], df[,21:26]))
13 }
14 guro_11 <- subset(d_11, d_code >= 11530510 & d_code <= 11530790)
15
16
17 for(i in 1201:1231){
18   df <- read.csv(paste0("LOCAL_PEOPLE_202012/LOCAL_PEOPLE_2020",i,".csv"), head=TRUE)
19   colnames(df) <- c("date", "hour", "d_code", "j_code", "total", "m_0", "m_10", "m_15", "m_20", "m_25", "m_30", "m_35", "m_40", "m_45",
20     "m_50", "m_55", "m_60", "m_65", "m_70", "w_0", "w_10", "w_15", "w_20", "w_25", "w_30", "w_35", "w_40", "w_45",
21     "w_50", "w_55", "w_60", "w_65", "w_70")
22   d_12 <- rbind(d_12, cbind(df[,1:5], df[,21:26]))
23 }
24 guro_12 <- subset(d_12, d_code >= 11530510 & d_code <= 11530790)
25
26
27 for(i in 101:131){
28   df <- read.csv(paste0("LOCAL_PEOPLE_202101/LOCAL_PEOPLE_20210",i,".csv"), head=TRUE)
29   colnames(df) <- c("date", "hour", "d_code", "j_code", "total", "m_0", "m_10", "m_15", "m_20", "m_25", "m_30", "m_35", "m_40", "m_45",
30     "m_50", "m_55", "m_60", "m_65", "m_70", "w_0", "w_10", "w_15", "w_20", "w_25", "w_30", "w_35", "w_40", "w_45",
31     "w_50", "w_55", "w_60", "w_65", "w_70")
32   d_01 <- rbind(d_01, cbind(df[,1:5], df[,21:26]))
33 }
34 guro_01 <- subset(d_01, d_code >= 11530510 & d_code <= 11530790)
35
36
37 for(i in 201:228){
38   df <- read.csv(paste0("LOCAL_PEOPLE_202102/LOCAL_PEOPLE_20210",i,".csv"), head=TRUE)
39   colnames(df) <- c("date", "hour", "d_code", "j_code", "total", "m_0", "m_10", "m_15", "m_20", "m_25", "m_30", "m_35", "m_40", "m_45",
40     "m_50", "m_55", "m_60", "m_65", "m_70", "w_0", "w_10", "w_15", "w_20", "w_25", "w_30", "w_35", "w_40", "w_45",
41     "w_50", "w_55", "w_60", "w_65", "w_70")
42   d_02 <- rbind(d_02, cbind(df[,1:5], df[,21:26]))
43 }
44 guro_02 <- subset(d_02, d_code >= 11530510 & d_code <= 11530790)

```

2. 분석 목적에 맞게 데이터 정제 후 월별 csv파일 작성 (데이터 정제 후 월별로 생성된 csv파일)

	A	B	C	D	E	F	G	H	I	J
1	date	hour	d_code	j_code	total	women				
2	20201101	0	11530510	1.12E+12	1396.251	257.7982				
3	20201101	0	11530510	1.12E+12	363.828	60.1682				
4	20201101	0	11530510	1.12E+12	582.1388	109.5718				
5	20201101	0	11530510	1.12E+12	265.2055	37.105				
6	20201101	0	11530510	1.12E+12	224.1438	34.6167				
7	20201101	0	11530510	1.12E+12	812.688	154.1242				
8	20201101	0	11530510	1.12E+12	486.3786	96.3195				
9	20201101	0	11530510	1.12E+12	677.8325	138.4845				
10	20201101	0	11530510	1.12E+12	605.4752	111.7783				

	A	B	C	D	E	F	G	H	I	J
1	date	hour	d_code	j_code	total	women				
2	20210301	21	11530550	1.12E+12	71.624	13.9096				
3	20210301	2	11530560	1.12E+12	72.6787	18.1902				
4	20210301	20	11530540	1.12E+12	49.3446	12.7965				
5	20210301	0	11530510	1.12E+12	1386.856	276.6568				
6	20210301	5	11530510	1.12E+12	368.5566	67.0353				
7	20210301	23	11530510	1.12E+12	360.1373	68.2841				
8	20210301	4	11530510	1.12E+12	542.3448	102.2137				
9	20210301	19	11530510	1.12E+12	1224.276	224.0635				
10	20210301	22	11530510	1.12E+12	822.9723	148.1589				

3. 월별 데이터를 모두 합한 후 분석 변수 생성 및 집계구별 변수값 총합계 산출 (집계구별 데이터 합계 산출 및 csv파일 작성 코드)

```
1 # 집계구별 데이터 합치기
2 setwd("C:/Users/user/Desktop/공백/구로구청/업무 파일/CCTV 분석/기초데이터/10. 생활인구/구로구 생활인구(집계구)")
3 flow <- data.frame()
4 for(i in 11:12){
5   df <- read.csv(paste0("pop_guro_2020",i,".csv"), header = TRUE)
6   flow <- rbind(flow, df)
7 }
8 for(i in 101:110){
9   df <- read.csv(paste0("pop_guro_2021",i,".csv"), header = TRUE)
10  flow <- rbind(flow, df)
11 }
12
13 # 구로구 데이터 날짜 변수 변환
14 flow$date <- as.Date(as.character(flow$date), format = '%Y%m%d')
15 flow$month <- paste(strftime(flow$date, "%Y"), strftime(flow$date, "%m"))
16 flow$month <- as.numeric(gsub("-", "", flow$month))
17 flow$day <- as.numeric(strftime(flow$date, "%d"))
18
19 # flow1_night, flow2_home 변수 생성
20 flow1 <- subset(flow, hour >= 18 & hour < 24)
21 flow1$flow1_night <- flow1$women
22 flow1$flow2_home <- 0
23
24 flow2 <- subset(flow, hour >= 0 & hour < 6)
25 flow2$flow1_night <- 0
26 flow2$flow2_home <- flow2$women
27
28 flow <- rbind(flow1, flow2)
29 flow <- cbind(flow[,3:4], flow[,7:8], flow[,2], flow[,9:10])
30 colnames(flow) <- c("d_code", "j_code", "month", "day", "hour", "flow1_night", "flow2_home")
31
32 # 집계구별 변수값 월합계 및 총합계 산출
33 flow_month <- as.data.frame(aggregate(list(flow$flow1_night, flow$flow2_home), by=list(flow$d_code, flow$j_code, flow$month), FUN=sum))
34 colnames(flow_month) <- c("d_code", "j_code", "month", "flow1_tot", "flow2_tot")
35 flow_total <- as.data.frame(aggregate(list(flow$flow1_night, flow$flow2_home), by=list(flow$d_code, flow$j_code), FUN=sum))
36 colnames(flow_total) <- c("d_code", "j_code", "flow1_tot", "flow2_tot")
37
38 setwd("C:/Users/user/Desktop/공백/구로구청/업무 파일/CCTV 분석/분석데이터/05. 범죄취약지수_유동인구/집계구별")
39 write.csv(flow_month, file="flow_month.csv", row.names=FALSE)
40 write.csv(flow_total, file="flow_total.csv", row.names=FALSE)
```

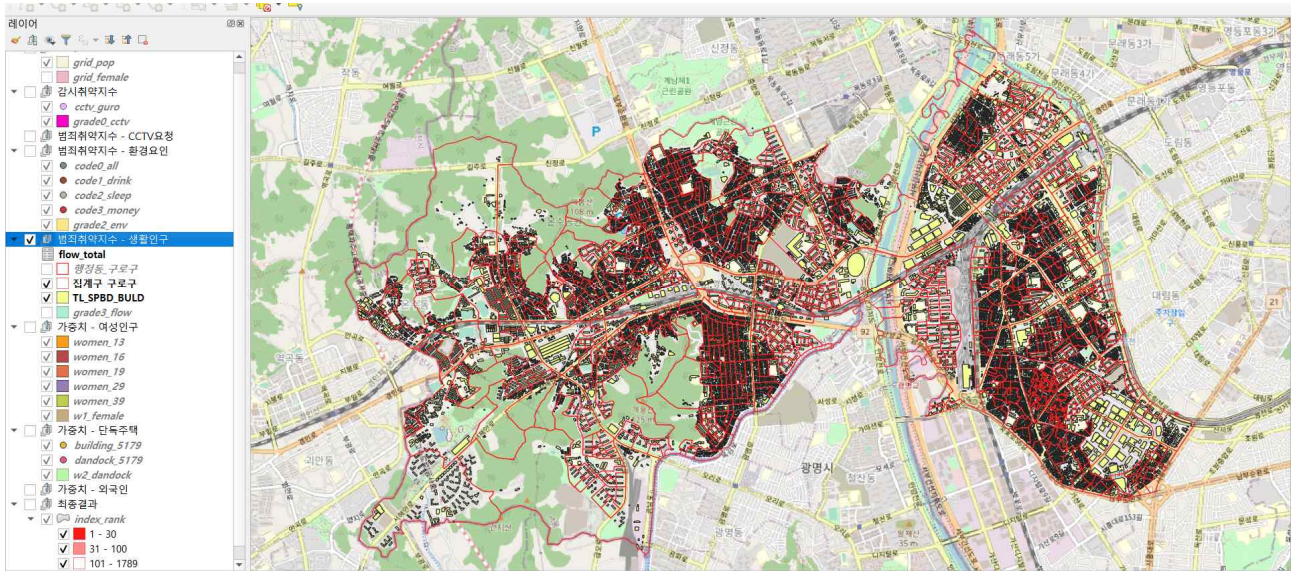
(전처리가 완료된 최종 csv파일)

	A	B	C	D	E	F	G	H	I	J
1	d_code	j_code	flow1_tot	flow2_tot						
2	11530510	1.12E+12	2102369	1893037						
3	11530510	1.12E+12	671687	470573.2						
4	11530510	1.12E+12	149023.4	111640.8						
5	11530510	1.12E+12	393200.1	235068						
6	11530510	1.12E+12	69494.59	71271.54						
7	11530510	1.12E+12	58041.8	64373.38						
8	11530510	1.12E+12	250969.1	268283.5						
9	11530510	1.12E+12	245139.5	170269.5						
10	11530510	1.12E+12	380665.9	335809.3						
11	11530510	1.12E+12	119781.1	183659.8						
12	11530510	1.12E+12	8769.465	23041.47						
13	11530510	1.12E+12	45891.83	101940.1						

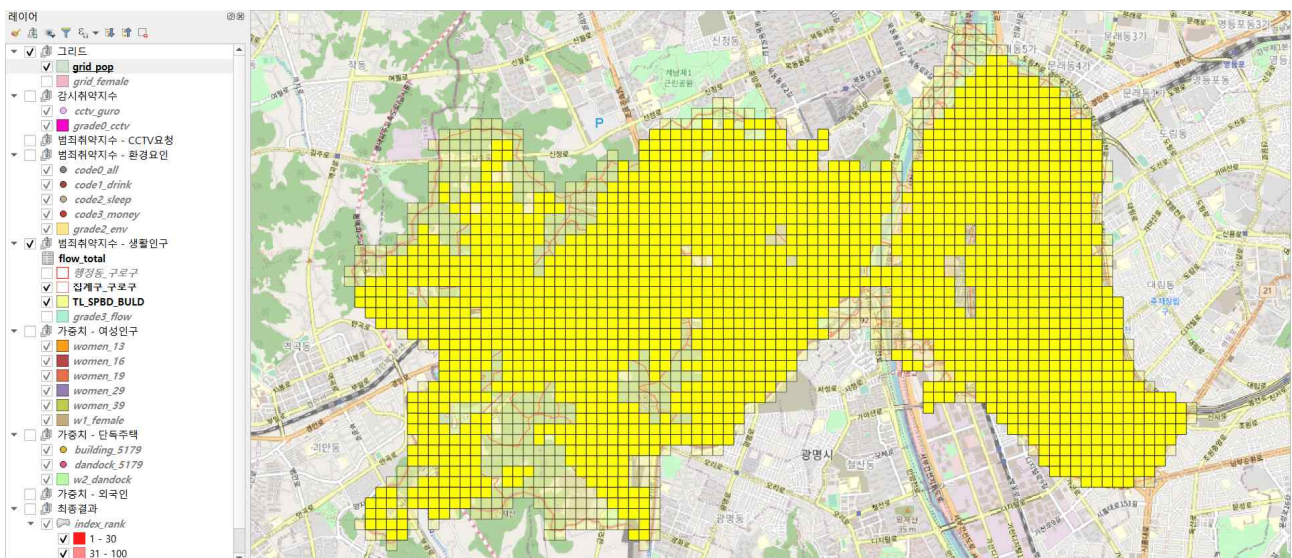
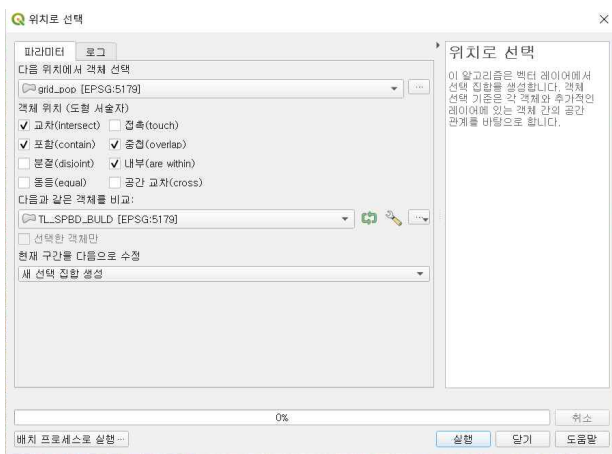
	A	B	C	D	E	F	G	H	I	J
798	11530740	1.12E+12	127377.3	189425.4						
799	11530740	1.12E+12	90923.96	143373.9						
800	11530740	1.12E+12	126156.1	216543.2						
801	11530740	1.12E+12	222899.1	150568.8						
802	11530740	1.12E+12	35885.1	91587.4						
803	11530740	1.12E+12	578032.2	270007.1						
804	11530740	1.12E+12	806581.2	247208.3						
805	11530740	1.12E+12	95459.36	187635.8						
806	11530740	1.12E+12	137581.4	194465.4						
807	11530740	1.12E+12	104808.7	163485.5						
808	11530740	1.12E+12	31399.36	39381.78						
809	11530740	1.12E+12	59765.15	76925.97						
810	11530740	1.12E+12	11698.56	24580.07						

▶ 생활인구 데이터 100m 격자 내 할당 - QGIS

1. QGIS 프로젝트에 집계구 데이터, 건물정보 데이터 추가



2. 100m*100m 격자 중 건물이 포함된 격자 선택 후 선택한 객체 다른 이름으로 저장('그리드_건물') (벡터 - 조사도구 - 위치로 선택)



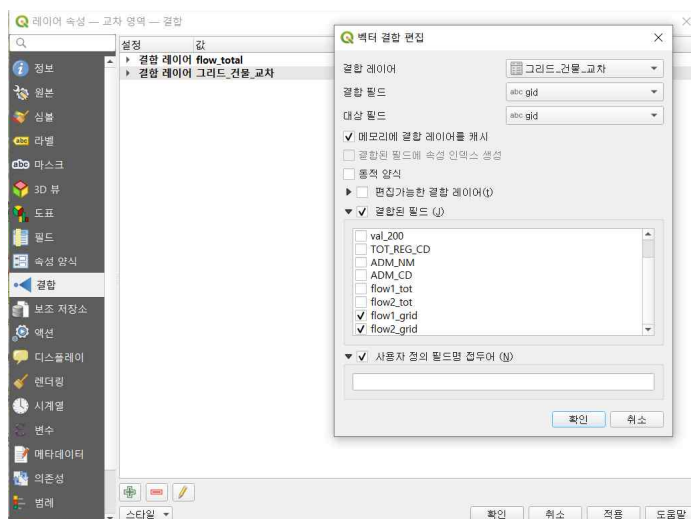
교차 영역 — 총 객체 수: 5401, 필터링된 객체 수: 5401, 선택한 객체 수: 0

	gid	lbl	val_200	TOT_REG_CD	ADM_NM	ADM_CD	flow1_tot	flow2_tot
1	다사462437	50.00	50	1117055010002	구로4동	1117055	185789.5819	197050.3479
2	다사462437	50.00	50	1117055010005	구로4동	1117055	174096.8055	95083.2143
3	다사462437	50.00	50	1117055020006	구로4동	1117055	301278.0883	145853.2629
4	다사462437	50.00	50	1117055020005	구로4동	1117055	77083.0754	120433.9053
5	다사399421	559.00	559	1117068030004	오류2동	1117068	1935162.2811	2166108.7201
6	다사412444	228.00	228	1117067020004	오류1동	1117067	128909.9944	169861.9006
7	다사412444	228.00	228	1117067020010	오류1동	1117067	45763.8436	130264.1753
8	다사412444	228.00	228	1117067020009	오류1동	1117067	81718.3492	162861.9292
9	다사412444	228.00	228	1117069020013	수궁동	1117069	105216.1843	207906.5657
10	다사412444	228.00	228	1117069020005	수궁동	1117069	210132.5781	194645.1553
11	다사428439	554.00	554	1117064010001	개봉2동	1117064	70798.9925	126716.198
12	다사428439	554.00	554	1117064010004	개봉2동	1117064	113975.7008	130862.2454
13	다사428439	554.00	554	1117064010006	개봉2동	1117064	160724.6457	169861.4645
14	다사428439	554.00	554	1117064010013	개봉2동	1117064	100218.9882	139932.6398
15	다사428439	554.00	554	1117064010022	개봉2동	1117064	170908.7532	199791.7986
16	다사428439	554.00	554	1117064010025	개봉2동	1117064	92624.1617	127263.7873
17	다사424440	591.00	591	1117064010019	개봉2동	1117064	207884.4728	213474.7636
18	다사424440	591.00	591	1117064010010	개봉2동	1117064	91103.3707	146440.1646
19	다사424440	591.00	591	1117064010009	개봉2동	1117064	102770.3437	186958.7659
20	다사424440	591.00	591	1117068020012	오류2동	1117068	327093.4981	478453.5184
21	다사439449	400.00	400	1117061010006	고척1동	1117061	511778.8839	232962.5075
22	다사439449	400.00	400	1117061010002	고척1동	1117061	150727.5172	188810.9872
23	다사439449	400.00	400	1117061010007	고척1동	1117061	107167.2091	187765.5947
24	다사460428	310.00	310	1117070010011	가리봉동	1117070	66387.6443	162211.0011

4. csv파일의 각 인구수를 해당 집계구 내 격자의 개수에 따라 배분한 새로운 변수 생성 후
 QGIS 교차 영역 레이어에 결합 및 다른 이름으로 저장('그리드_생활인구')
 (엑셀 함수를 활용한 새로운 변수 생성)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	gid	lbl	val_200	TOT_REG	ADM_NM	ADM_CD	flow1_tot	flow2_tot	flow1_grid	flow2_grid										
2	다사46243	50	50	1.12E+12	구로4동	1117055	185789.6	197050.3	TIF(D:D,I2)	28150.0497										
3	다사46243	50	50	1.12E+12	구로4동	1117055	174096.8	95083.21	19344.0895	10564.80159										
4	다사46243	50	50	1.12E+12	구로4동	1117055	301278.1	145853.3	23175.23756	11219.48176										
5	다사46243	50	50	1.12E+12	구로4동	1117055	77083.08	120433.9	12847.17923	20072.31755										
6	다사39942	559	559	1.12E+12	오류2동	1117068	1935162	2166109	32799.3607	36713.70712										
7	다사41244	228	228	1.12E+12	오류1동	1117067	128910	169861.9	21484.99907	28310.31677										
8	다사41244	228	228	1.12E+12	오류1동	1117067	45763.84	130264.2	6537.691943	18609.1679										
9	다사41244	228	228	1.12E+12	오류1동	1117067	81718.35	162861.9	8171.83492	16286.19292										
10	다사41244	228	228	1.12E+12	수궁동	1117069	105216.2	207906.6	9565.107664	18900.59688										
11	다사41244	228	228	1.12E+12	수궁동	1117069	210132.6	194645.2	26266.57226	24330.64441										
12	다사42843	554	554	1.12E+12	개봉2동	1117064	70798.99	126716.2	17699.74813	31679.0495										

(QGIS 교차 영역 레이어에 격자 번호를 기준으로 새로 생성한 변수 결합)



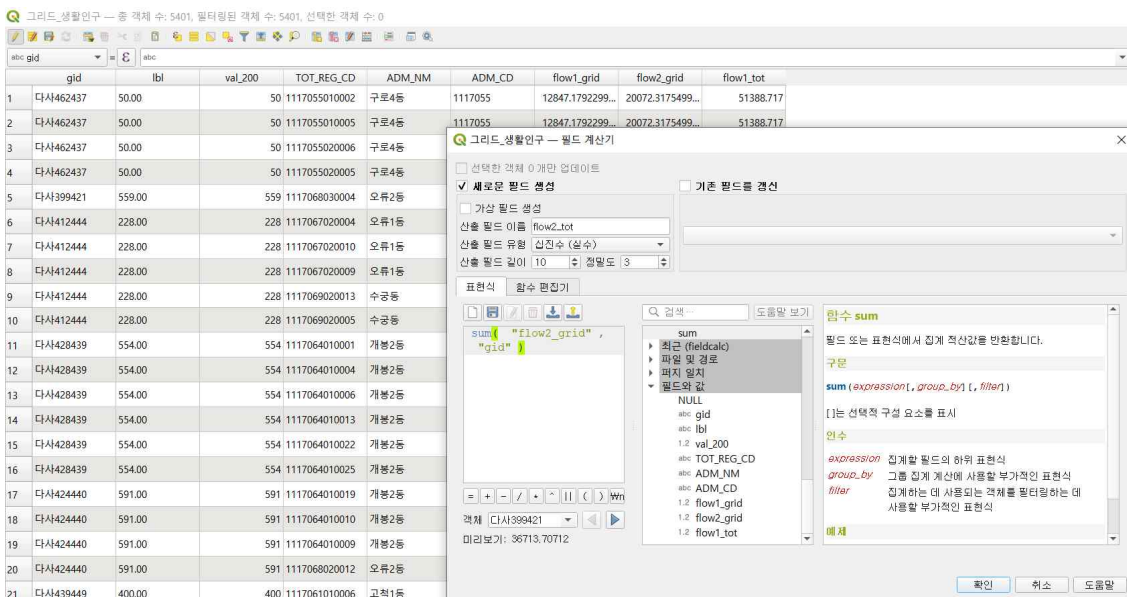
(다사462437 격자에 해당하는 4개의 영역과 각 영역에 해당하는 인구수 데이터 배분)

	gid	lbl	val_200	TOT_REG_CD	ADM_NM	ADM_CD	flow1_tot	flow2_tot	flow1_grid	flow2_grid
1	다사462437	50.00	50	1117055010002	구로4동	1117055	185789.5819	197050.3479	12847.17923	20072.31755
2	다사462437	50.00	50	1117055010005	구로4동	1117055	174096.8055	95083.2143	12847.17923	20072.31755
3	다사462437	50.00	50	1117055020006	구로4동	1117055	301278.0883	145853.2629	12847.17923	20072.31755
4	다사462437	50.00	50	1117055020005	구로4동	1117055	77083.0754	120433.9053	12847.17923	20072.31755
5	다사399421	559.00	559	1117068030004	오류2동	1117068	1935162.2811	2166108.7201	32799.3607	36713.70712
6	다사412444	228.00	228	1117067020004	오류1동	1117067	128909.9944	169861.9006	26266.57226	24330.64441
7	다사412444	228.00	228	1117067020010	오류1동	1117067	45763.8436	130264.1753	26266.57226	24330.64441
8	다사412444	228.00	228	1117067020009	오류1동	1117067	81718.3492	162861.9292	26266.57226	24330.64441
9	다사412444	228.00	228	1117069020013	수궁동	1117069	105216.1843	207906.5657	26266.57226	24330.64441
10	다사412444	228.00	228	1117069020005	수궁동	1117069	210132.5781	194645.1553	26266.57226	24330.64441
11	다사428439	554.00	554	1117064010001	개봉2동	1117064	70798.9925	126716.198	23156.04043	31815.94683
12	다사428439	554.00	554	1117064010004	개봉2동	1117064	113975.7008	130862.2454	23156.04043	31815.94683
13	다사428439	554.00	554	1117064010006	개봉2동	1117064	160724.6457	169861.4645	23156.04043	31815.94683
14	다사428439	554.00	554	1117064010013	개봉2동	1117064	100218.9882	139932.6398	23156.04043	31815.94683
15	다사428439	554.00	554	1117064010022	개봉2동	1117064	170908.7532	199791.7986	23156.04043	31815.94683
16	다사428439	554.00	554	1117064010025	개봉2동	1117064	92624.1617	127263.7873	23156.04043	31815.94683
17	다사424440	591.00	591	1117064010019	개봉2동	1117064	207884.4728	213474.7636	46727.64259	68350.50263
18	다사424440	591.00	591	1117064010010	개봉2동	1117064	91103.3707	146440.1646	46727.64259	68350.50263

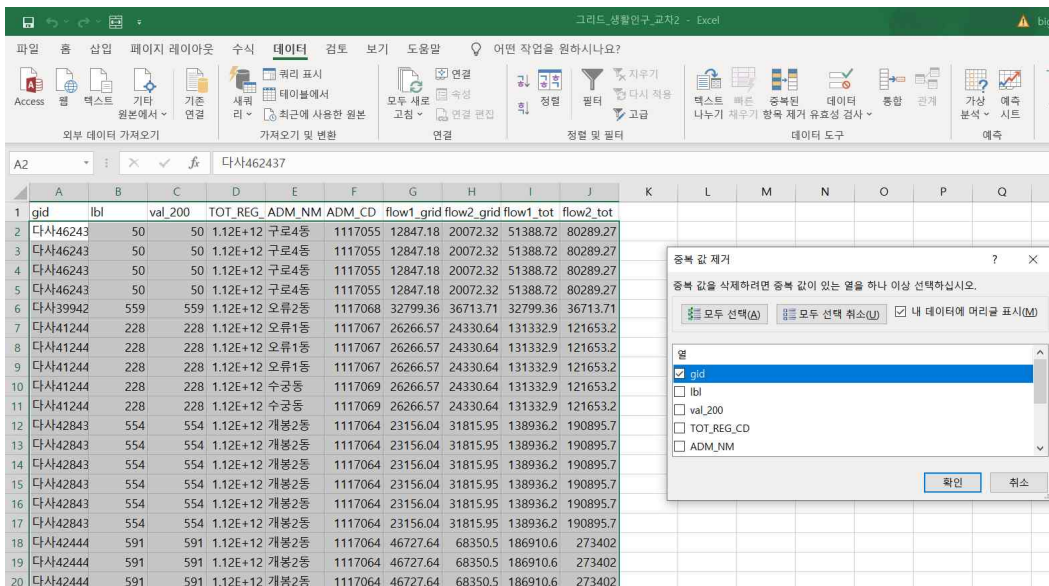
5. 불필요한 필드 제거 후 각 격자별로 인구수 데이터를 합산하여 csv파일로 저장 및 중복 데이터 제거 (‘그리드_생활인구’ 레이어의 불필요한 필드 제거)



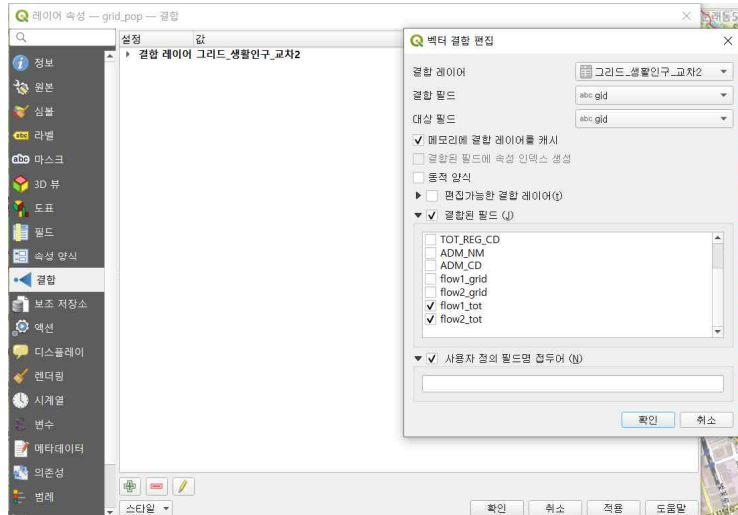
(각 격자별 인구수 데이터를 합산)



(격자 번호를 기준으로 중복된 데이터 제거)



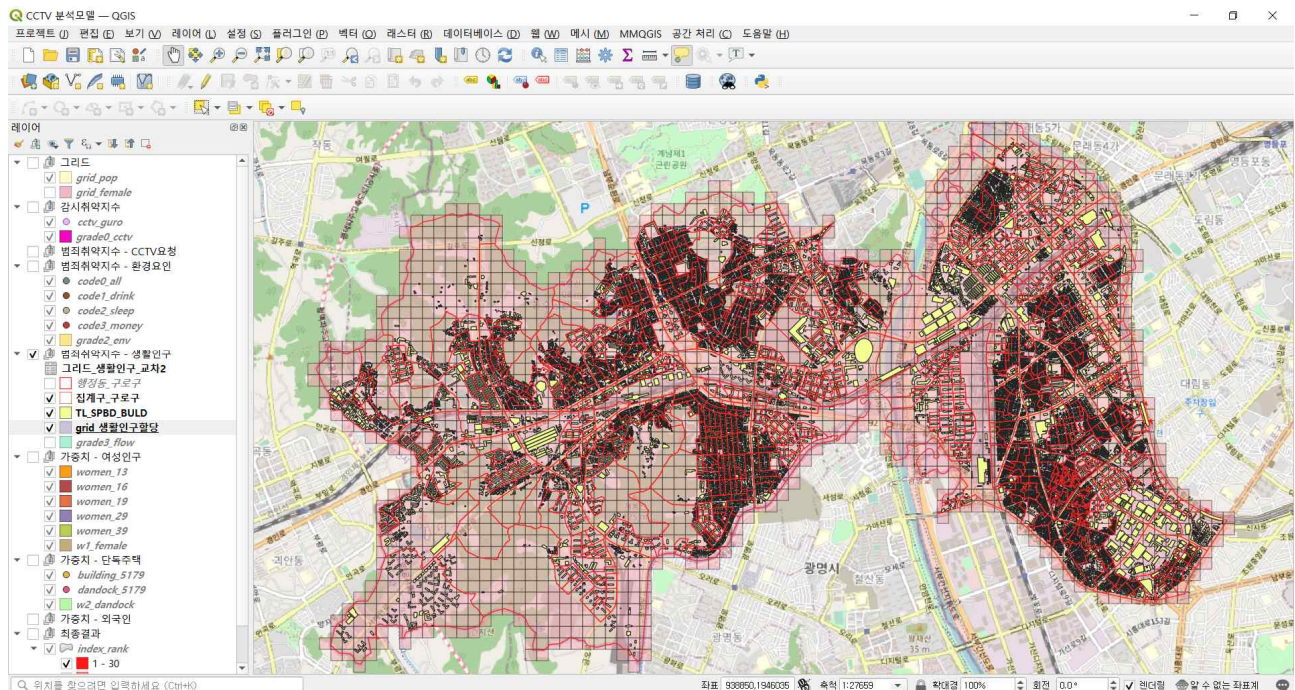
6. 최초 100m*100m 격자 레이어에 csv파일 결합 및 다른 이름으로 저장('grid_생활인구할당')
(격자 번호를 기준으로 할당된 생활인구 데이터 결합)



('grid_생활인구할당' 레이어 속성 테이블)

gid	lbl	val_200	flow1_tot	flow2_tot
다사462437	50.00	50	51388.7169999...	80289.2700000...
다사399421	559.00	559	32799.3609999...	36713.7070000...
다사412444	228.00	228	131332.861000...	121653.221999...
다사428439	554.00	554	138936.242999...	190895.681000...
다사424440	591.00	591	186910.570000...	273402.010999...
다사439449	400.00	400	80375.4070000...	140824.195999...
다사460428	310.00	310	73222.3659999...	58775.6019999...
다사462434	782.00	782	76294.4119999...	170380.225999...
다사417440	251.00	251	108550.960999...	93331.4360000...

('grid_생활인구할당' 레이어 시각화)



※ 참고자료 : 통계청 집계구 인구를 격자로 재할당하기