Preprocess Time Series

Data

2017 - 2021 Active Population Data

Goal: forecast 2month data

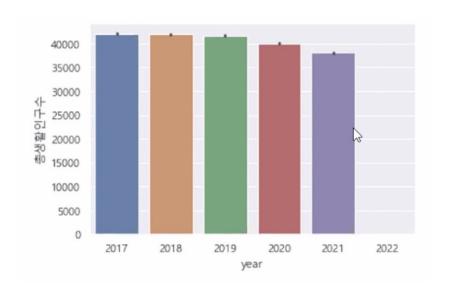
	기준일ID	시 간 대 구 분	총생활인구수	남자0세부 터9세생활 인구수	남자10세 부터14세 생활인구 수	남자15세 부터19세 생활인구 수	남자20세 부터24세 생활인구수	남자25세 부터29세 생활인구수	남자30세 부터34세 생활인구수	남자35세 부터39세 생활인구수	 여자25세 부터29세 생활인구 수	여자30세 부터34세 생활인구수 성
0	20170101	0	31535.2200	1945.4402	566.4588	631.5581	991.9184	1297.1287	1223.1529	1552.9448	1157.6661	1107.6781 13
1	20170101		31188.9174	1930.4766	562.1018	625.2151	1030.6281	1307.7381	1183.0956	1491.5827	1120.0051	1102.5198 1
2	20170101	2	31240.4974	1921.1864	559.3965	658.7463	1058.7082	1311.9591	1186.2600	1537.3732	1103.1547	1112.4579 12
3	20170101	3	31442.4314	1946.4476	566.7525	649.0804	1057.9119	1324.2638	1202,1250	1521.2732	1116.2165	1102.0668 12
4	20170101	4	31922.7751	1953.7713	568.8845	623.2087	1063.2557	1342.9146	1209.2680	1556.0675	1187.5457	1120.2232 1
8755	20211231	19	33060.4555	1858.2611	541.0746	605.9080	781.6602	1342.6763	1435.1080	1678.8904	1415.9734	1513,1706 1
8756	20211231	20	32360,1898	2093.8808	609.6807	612.5326	768.8351	1323.7124	1479.1382	1654.5005	1311.7764	1431.5813 1
8757	20211231	21	31417.3991	1983.6728	577.5909	639.2174	779.6693	1187.6783	1375.6785	1600.9946	1186.6502	1348.7581 1
8758	20211231	22	30763.0563	2028.3314	590.5941	654.0295	790.1638	1133.6311	1362.0266	1548.7366	1108.3915	1260.6435
8759	20211231	23	30404.2740	1903.8825	554.3582	615.0753	772.8499	1136.6157	1295.0434	1549.1984	993.5635	1166.8245 1
43512 r	ows × 31 col	lumns	5									

Domain knowledge

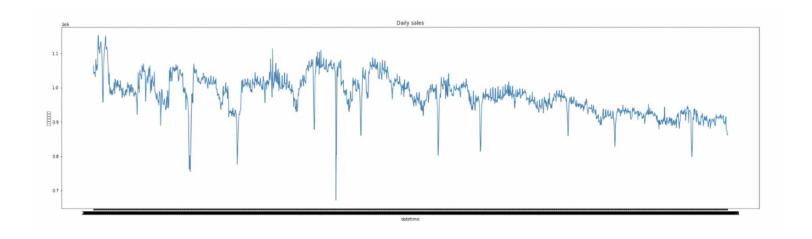
2019-2020 appearance of Covid-19

2021 infection rates peak

2022 ???



Outliers



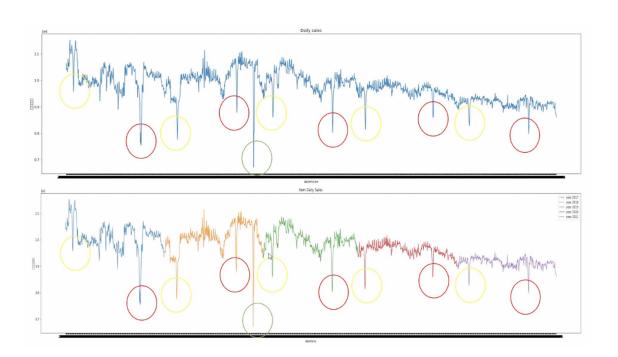
Getting Trend from Outliers

red: 추석

yellow: 설날

green: lost

(fire on data center)



Shift and Rolling

```
for i in range(1416,1416+(24*7)):

df_total[f'총생활인구수_{i}lag'] = df_total['총생활인구수'].shift(i)

for i in range(1416,1416+(24*7)):

df_total[f'총생활인구수_{i}이동평균'] = df_total[f'총생활인구수_{i}lag'].rolling(window=1416).mean()
```

	시 간 대 구 분	총생활인구 수	Ħ	앭	열	weekday_0	weekday_1	weekday_2	weekday_3	weekday_4	-	총생활인구수 _1574이동평 군	총생활인구수 _1575이동평 교	충생달인구수 _1576이동평 귣			총생활연구수 _1579이동평 균		총생활인구수 _1581이듬평 교	충생활연구수 _1582이동평 교	충생활인구수 _1583여동평 귣
2998		38583.1827	2017				0.0	0.0	0.0			44198.816904	44200.220788	44201.163320	44202.154401	44202.451685	44202.479342	44202.743557	44202,961268	44203.568294	44204.393558
2999		38857.9020	2017					0.0				44197.611944	44198.816904	44200.220788	44201.163320	44202.154401	44202.451685	44202.479342	44202.743557	44202.961268	44203.568294
3000		39289.3581	2017				0.0	0.0	0.0	0.0		44196.473656	44197.511944	44198.816904	44200.220788	44201.163320	44202.154401	44202.451685	44202.479342	44202.743557	44202.961268
3001		39498.6305	2017					0.0				44195.265283	44196.473656	44197,611944	44198.816904	44200.220788	44201.163320	44202.154401	44202,451685	44202.479342	44202.743557
3002		39392.0822	2017				0.0	0.0	0.0	0.0		44194.117890	44195.265283	44196.473656	44197.611944	44198.816904	44200.220788	44201.163320	44202.154401	44202.451685	44202.479342
-																					-
43531	19	33304.5400	2021			0.0	0.0	0.0	0.0			38004.516596	38004.197890	38003.628703	38003.203196	38002.837111	38002.590300	38002.462235	38001.357871	38000.904904	38000.933682
43532		33368.7201	2021				0.0					38004.932074	38004.516596	38004.197890	38003.628703	38003.203196	38002.837111	38002.590300	38002.462235	38001.357871	38000.904904
43533		34063.1721	2021			0.0	0.0	0.0	0.0			38006.025245	38004.932074	38004.516596	38004.197890	38003.628703	38003.203196	38002.837111	38002.590300	38002.462235	38001,357871
43534		34380.0806	2021					0.0				38006.084813	38006.025245	38004.932074	38004.516596	38004.197890	38003.628703	38003.203196	38002.837111	38002.590300	38002.462235
43535		37074.2785	2021			0.0	0.0	0.0	0.0			38005.395013	38006.064813	38006.025245	38004.932074	38004.516596	38004.197890	38003.628703	38003.203196	38002.837111	38002.590300

result

RMSE 982