

파이썬 데이터분석

- Pandas

강사 : KAIST 김동훈

I. 시계열 데이터 분석

- Open Power Systems Data
- Time Series Data Structures
- 시계열 데이터 불러오기
- 시각화 및 변동성 확인
- Resampling
- 이동평균

I. 시계열 데이터 분석

- Open Power Systems Data

- 독일 전력 회사인 OPSD 의 데이터
- 전력 소비량, 풍력 발전량, 태양열 발전량 Data, 2006-2007

데이터 항목

- Date — The date (yyyy-mm-dd format)
- Consumption — Electricity consumption in GWh
- Wind — Wind power production in GWh
- Solar — Solar power production in GWh
- Wind+Solar — Sum of wind and solar power production in GWh

I. 시계열 데이터 분석

- Open Power Systems Data

다음 질문에 답해 보자.

- 언제 전력 소비량이 최대이고 언제 최소인가?
- 풍력과 태양열 생산은 계절별로 어떻게 바뀌는가?
- 전력 소비량, 풍력 생산량, 태양열 생산량의 장기 트렌드는 어떤가?
- 전력 소비량 대비 풍력 생산량과 태양열 생산량의 비율에는 변화가 있는가?

I. 시계열 데이터 분석

- Time Series Data Structures

1. **Timestamp** : 특정 시점의 시간을 표현, to_datetime() 함수로 문자열로부터 생성 가능

```
>>> pd.to_datetime('2018-01-15 3:45pm')
Timestamp('2018-01-15 15:45:00')

>>> pd.to_datetime('7/8/1952')
Timestamp('1952-07-08 00:00:00')
```

I. 시계열 데이터 분석

- Time Series Data Structures

2. **DatetimeIndex** : 날짜 문자열 배열(array)를 to_datetime() 에 입력하면 DatetimeIndex 를 반환

```
>>> pd.to_datetime(['2018-01-05', '7/8/1952', 'Oct 10, 1995'])  
DatetimeIndex(['2018-01-05', '1952-07-08', '1995-10-10'], dtype='datetime64[ns]', freq=None)
```

※ **datetime64[ns]** : 64bit 정수형, 날짜 시간을 표현하는 데이터 타입, 나노초 단위

I. 시계열 데이터 분석

- Time Series Data Structures

3. **format** : 날짜 문자열을 datetime64 로 변환하기 위한 형식 지정 방법

```
>>> pd.to_datetime(['2/25/10', '8/6/17', '12/15/12'], format='%m/%d/%y')
DatetimeIndex(['2010-02-25', '2017-08-06', '2012-12-15'], dtype='datetime64[ns]', freq=None)
```

※ '%m/%d/%y' : %m = '월', %d='일', %y='년' 을 의미

I. 시계열 데이터 분석

- 시계열 data 불러오기

- 'Date' 컬럼을 to_datetime() 을 사용해 datetime64 로 변경
- 'Date' 컬럼을 인덱스 컬럼으로 설정
- 날짜 요소들을 하나씩 속성으로 뽑아 올 수 있다.

	Consumption	Wind	Solar	Wind+Solar	Year	Month	Day Name
Date							
2008-08-23	1152.011	NaN	NaN	NaN	2008	8	Saturday
2013-08-08	1291.984	79.666	93.371	173.037	2013	8	Thursday
2009-08-27	1281.057	NaN	NaN	NaN	2009	8	Thursday
2015-10-02	1391.050	81.229	160.641	241.870	2015	10	Friday
2009-06-02	1201.522	NaN	NaN	NaN	2009	6	Tuesday

I. 시계열 데이터 분석

- 시계열 data 불러오기

- 'Date' 컬럼을 datetime64 인덱스로 지정하면 다양한 날짜 기준으로 인덱싱을 할 수 있다.

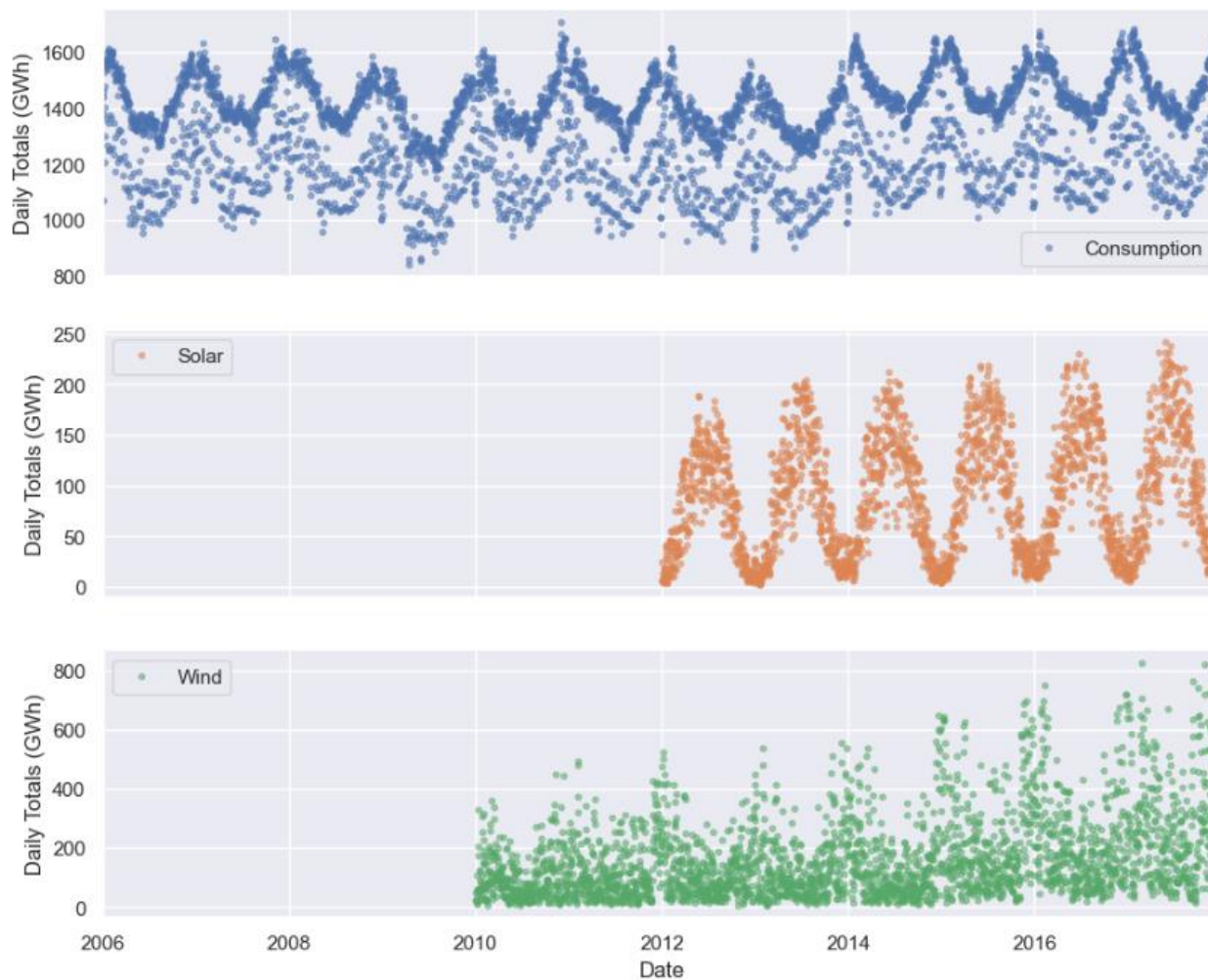
```
opds_daily.loc['2012-02']
```

	Consumption	Wind	Solar	Wind+Solar	Year	Month	Day Name
Date							
2012-02-01	1511.866	199.607	43.502	243.109	2012	2	Wednesday
2012-02-02	1563.407	73.469	44.675	118.144	2012	2	Thursday
2012-02-03	1563.631	36.352	46.510	82.862	2012	2	Friday
2012-02-04	1372.614	20.551	45.225	65.776	2012	2	Saturday
2012-02-05	1279.432	55.522	54.572	110.094	2012	2	Sunday
2012-02-06	1574.766	34.896	55.389	90.285	2012	2	Monday

I. 시계열 데이터 분석

- 시각화 및 변동성 확인

- pandas 의 plot 함수를 이용해 전력 소비량, 풍력 생산량, 태양열 생산량을 시각해 볼 수 있다.

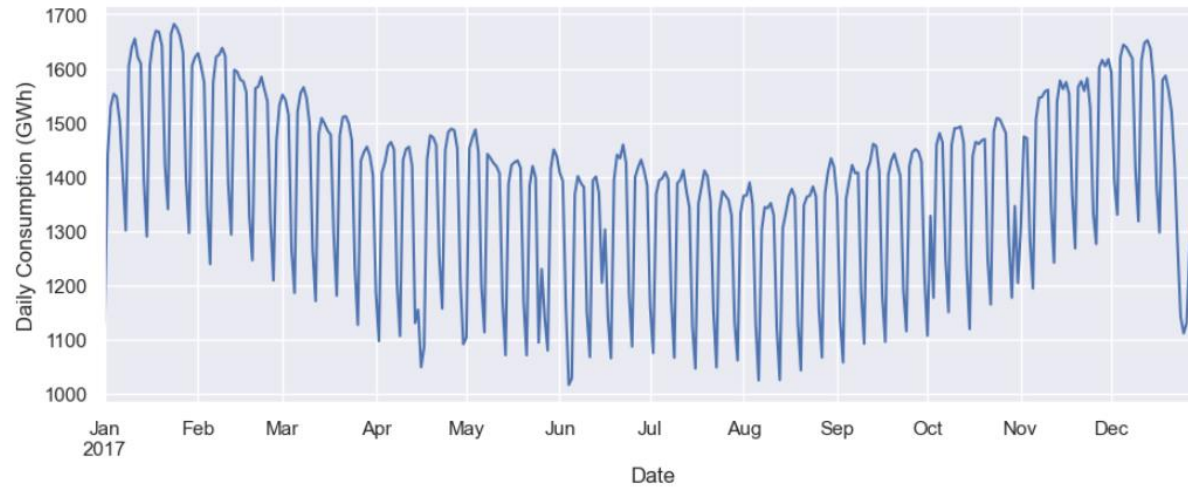


I. 시계열 데이터 분석

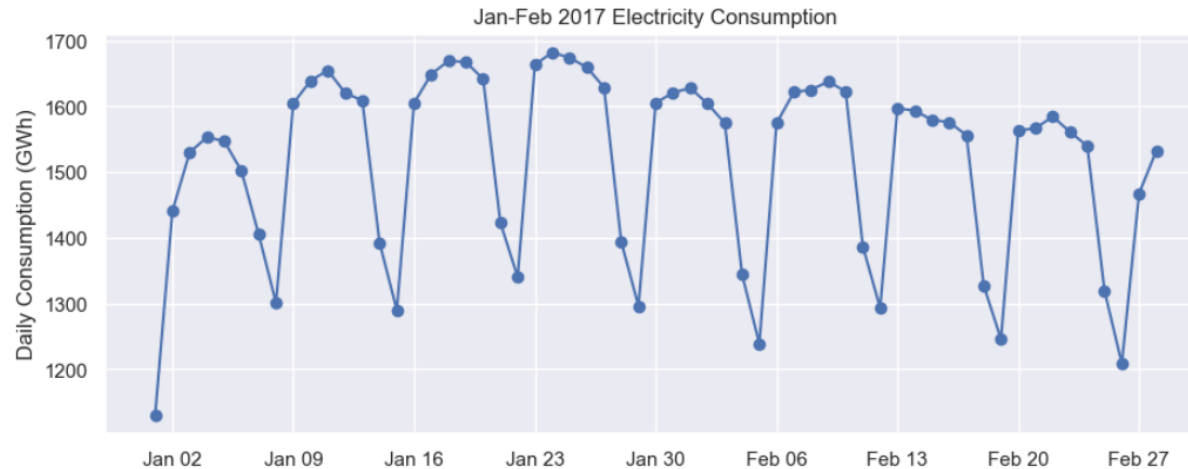
- 시각화 및 변동성 확인

- 'Date' 컬럼으로 기간을 좁혀 가면서(ZoomIn) 변동성을 더 잘 확인 할 수 있다.

년 단위



월 단위

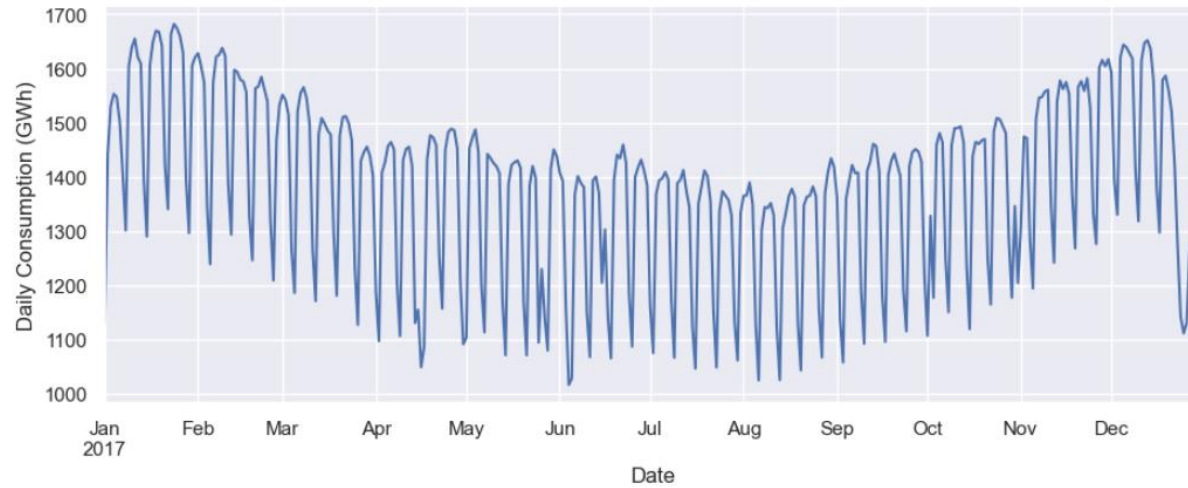


I. 시계열 데이터 분석

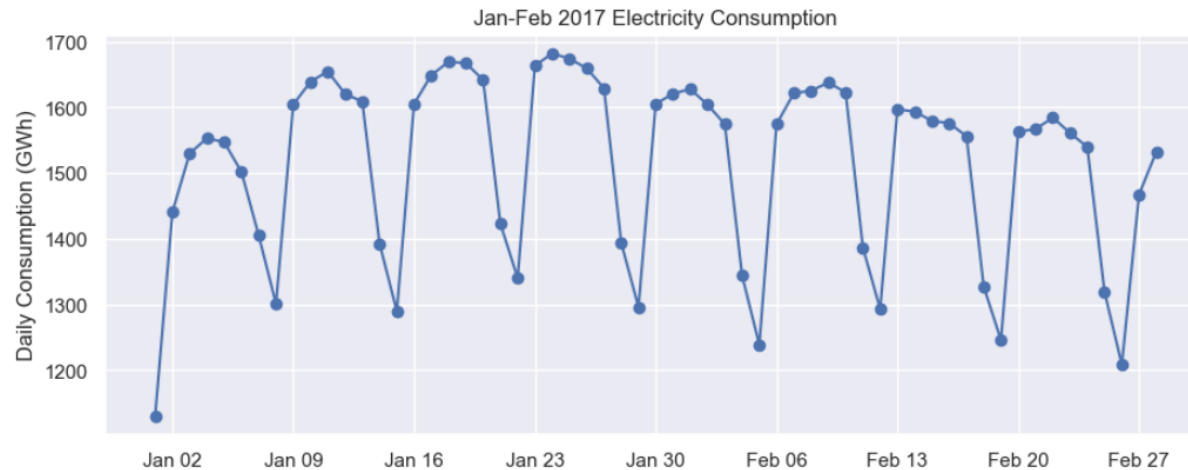
- 시각화 및 변동성 확인

- 'Date' 컬럼으로 기간을 좁혀 가면서(ZoomIn) 변동성을 더 잘 확인 할 수 있다.

년 단위



월 단위



I. 시계열 데이터 분석

- Resampling

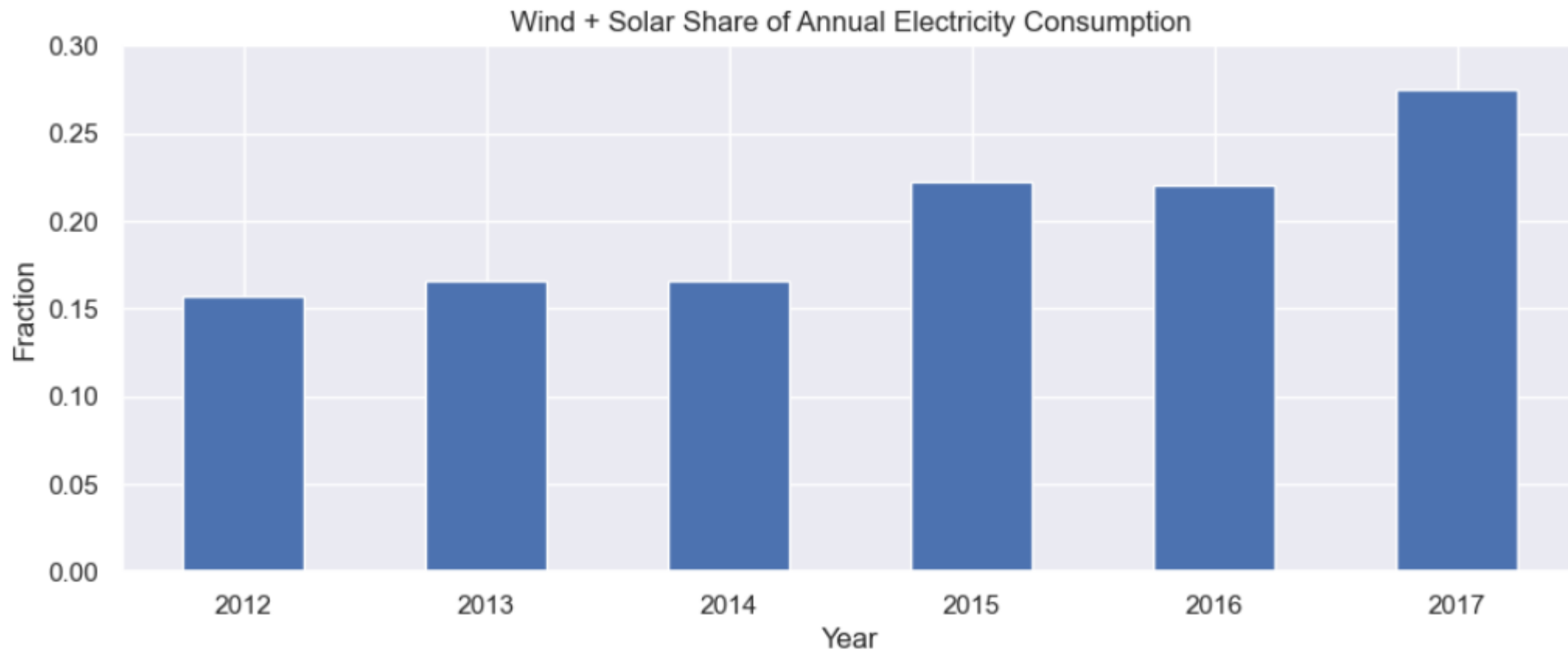
- resampling 을 통해 데이터 수를 줄이고 대표값으로 smoothing 할 수 있다.



I. 시계열 데이터 분석

- Resampling

- (풍력 생산량 + 태양열 생산량) / 전력 소비량 이 꾸준히 증가한 것을 확인 할 수 있다. (년 단위)



II. 서울시 구별 **CCTV** 현황 분석

- CCTV 데이터
- 서울시 인구 현황
- 데이터 합치기와 인덱스 변경
- CCTV와 인구현황 그래프로 분석하기

Ⅱ. 서울시 구별 CCTV 현황 분석

- 데이터 분석 목표

- 서울시 각 구별 CCTV수를 파악하고, 인구대비 CCTV 비율을 파악해서 순위 비교
- 인구대비 CCTV의 평균치를 확인하고 그로부터 CCTV가 과하게 부족한 구를 확인
- 단순한 그래프 표현에서 한 단계 더 나아가 경향을 확인하고 시각화하는 기초 확인

Ⅱ. 서울시 구별 CCTV 현황 분석

■ CCTV 데이터

	기관명	소계	2013년도 이전	2014년	2015년	2016년
0	강남구	2780	1292	430	584	932
1	강동구	773	379	99	155	377
2	강북구	748	369	120	138	204
3	강서구	884	388	258	184	81
4	관악구	1496	846	260	390	613
5	광진구	707	573	78	53	174
6	구로구	1561	1142	173	246	323
7	금천구	1015	674	51	269	354
8	노원구	1265	542	57	451	516
9	도봉구	485	238	159	42	386



	구별	소계	2013년도 이전	2014년	2015년	2016년	최근증가율
22	종로구	1002	464	314	211	630	248.922414
9	도봉구	485	238	159	42	386	246.638655
12	마포구	574	314	118	169	379	212.101911
8	노원구	1265	542	57	451	516	188.929889
1	강동구	773	379	99	155	377	166.490765
19	영등포구	904	495	214	195	373	157.979798
0	강남구	2780	1292	430	584	932	150.619195
4	관악구	1496	846	260	390	613	149.290780
23	중구	671	413	190	72	348	147.699758
11	동작구	1091	544	341	103	314	139.338235

II. 서울시 구별 CCTV 현황 분석

- 서울시 인구 현황

	구별	인구수	한국인	외국인	고령자
1	종로구	162820.0	153589.0	9231.0	25425.0
2	중구	133240.0	124312.0	8928.0	20764.0
3	용산구	244203.0	229456.0	14747.0	36231.0
4	성동구	311244.0	303380.0	7864.0	39997.0
5	광진구	372164.0	357211.0	14953.0	42214.0
6	동대문구	369496.0	354079.0	15417.0	54173.0
7	중랑구	414503.0	409882.0	4621.0	56774.0
8	성북구	461260.0	449773.0	11487.0	64692.0
9	강북구	330192.0	326686.0	3506.0	54813.0
10	도봉구	348646.0	346629.0	2017.0	51312.0



	구별	인구수	한국인	외국인	고령자	외국인비율	고령자비율
1	종로구	162820.0	153589.0	9231.0	25425.0	5.669451	15.615404
2	중구	133240.0	124312.0	8928.0	20764.0	6.700690	15.583909
3	용산구	244203.0	229456.0	14747.0	36231.0	6.038828	14.836427
4	성동구	311244.0	303380.0	7864.0	39997.0	2.526635	12.850689
5	광진구	372164.0	357211.0	14953.0	42214.0	4.017852	11.342849
6	동대문구	369496.0	354079.0	15417.0	54173.0	4.172440	14.661322
7	중랑구	414503.0	409882.0	4621.0	56774.0	1.114829	13.696885
8	성북구	461260.0	449773.0	11487.0	64692.0	2.490353	14.025062
9	강북구	330192.0	326686.0	3506.0	54813.0	1.061806	16.600342
10	도봉구	348646.0	346629.0	2017.0	51312.0	0.578524	14.717507

II. 서울시 구별 CCTV 현황 분석

- 데이터 합치기와 인덱스 변경
 - CCTV 데이터와 서울시 인구현황 데이터를 '구' 기준으로 JOIN

	소계	최근증가율	인구수	한국인	외국인	고령자	외국인비율	고령자비율
구별								
강남구	2780	150.619195	570500.0	565550.0	4950.0	63167.0	0.867660	11.072217
강동구	773	166.490765	453233.0	449019.0	4214.0	54622.0	0.929765	12.051638
강북구	748	125.203252	330192.0	326686.0	3506.0	54813.0	1.061806	16.600342
강서구	884	134.793814	603772.0	597248.0	6524.0	72548.0	1.080540	12.015794
관악구	1496	149.290780	525515.0	507203.0	18312.0	68082.0	3.484582	12.955291
광진구	707	53.228621	372164.0	357211.0	14953.0	42214.0	4.017852	11.342849
구로구	1561	64.973730	447874.0	416487.0	31387.0	56833.0	7.007998	12.689506
금천구	1015	100.000000	255082.0	236353.0	18729.0	32970.0	7.342345	12.925255
노원구	1265	188.929889	569384.0	565565.0	3819.0	71941.0	0.670725	12.634883
도봉구	485	246.638655	348646.0	346629.0	2017.0	51312.0	0.578524	14.717507

Ⅱ. 서울시 구별 CCTV 현황 분석

- CCTV와 인구현황 그래프로 분석하기

