



10. 데이터분석-실습



이번시간에 학습할 내용은..

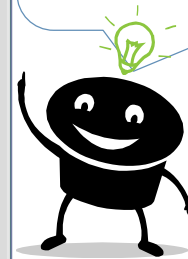
데이터를 읽어들이고 분석포커스에 맞춰 필터링
한 후 그 결과를 시각화하는 프로그램 작성 .

I. 우리가 사용할 데이터에 대해

- accidentdata.csv – 2012~2014년도 교통사건사고파일(제공:공공데이터 포털사이트)
- 분석포커스 – 요일별 중대교통사고 사상자 합계 (사상자 3명이상인 데이터에 대해)
- 경기도에서 교통사망사고가 많은 5지역을 분석하여 도식화
- 요일별/발생지시도별 교통사고 빈도분석

accidentdata.csv - Excel																		
로그인																		
어떤 작업을 원하시나요?																		
파일 홈 삽입 페이지 레이아웃 수식 데이터 검토 보기 개발 도구 ACROBAT																		
문서 작업																		
클립보드 글꼴 맞춤 표시 형식 스타일 편집																		
A1 발생년																		
1	발생년	발생년월일시	발생분	주야	요일	사망자수	사상자수	중상자수	경상자수	부상신고	발생지시도	발생지시도	사고유형	사고유형	사고유형	법규위반	법규위	
2	2012	2012050901	10 야간	수	0	1	1	0	0	0	경남	거창군	차대차	추돌	진행중 추	운전자법규안전운		
3	2012	2012033013	32 주간	금	1	1	1	0	0	0	경기	용인시	차대차	추돌	주정차중 추	운전자법규안전운		
4	2012	2012102620	55 야간	금	1	2	1	0	0	0	경북	김천시	차대차	추돌	주정차중 추	운전자법규안전운		
5	2012	2012062814	10 주간	목	1	1	1	0	0	0	경기	이천시	차대차	추돌	주정차중 추	운전자법규안전운		
6	2012	2012061804	15 야간	월	1	1	0	0	0	0	전남	광양시	차대차	추돌	진행중 추	운전자법규안전운		
7	2012	2012071508	4 주간	일	1	5	0	0	4	0	전남	광양시	차대차	추돌	진행중 추	운전자법규안전운		
8	2012	2012053104	17 야간	목	1	2	1	0	0	0	경기	하남시	차대차	추돌	진행중 추	운전자법규안전운		
9	2012	2012091714	44 주간	월	1	1	0	0	0	0	충남	논산시	차대차	추돌	진행중 추	운전자법규안전운		
10	2012	2012100102	43 야간	월	1	5	1	3	0	0	경남	밀양시	차대차	추돌	진행중 추	운전자법규안전운		
11	2012	2012081612	55 주간	목	1	1	0	0	0	0	전북	진안군	차대차	추돌	진행중 추	운전자법규안전운		
12	2012	2012083008	15 주간	목	1	5	1	3	0	0	대구	북구	차대차	추돌	진행중 추	운전자법규안전운		
13	2012	2012091604	20 야간	일	1	1	0	0	0	0	대구	달성군	차대차	추돌	진행중 추	운전자법규안전운		
14	2012	2012092410	57 주간	월	1	2	0	0	0	0	전북	무주군	차대차	추돌	진행중 추	운전자법규안전운		
15	2012	2012101602	45 야간	화	1	1	0	0	0	0	경기	화성시	차대차	추돌	진행중 추	운전자법규안전운		
16	2012	2012121404	50 야간	금	1	2	0	1	0	0	경기	광명시	차대차	추돌	진행중 추	운전자법규안전운		
17	2012	2012111110	35 주간	일	1	1	0	0	0	0	경남	진주시	차대차	추돌	진행중 추	운전자법규안전운		
18	2012	2012110601	38 야간	화	1	1	0	0	0	0	경남	진주시	차대차	추돌	진행중 추	운전자법규안전운		
19	2012	2012081100	30 야간	토	1	1	0	0	0	0	경북	구미시	차대차	추돌	진행중 추	운전자법규안전운		
20	2012	2012092002	10 야간	목	1	3	0	2	0	0	경북	구미시	차대차	추돌	진행중 추	운전자법규안전운		
21	2012	2012101011	5 야간	일	1	6	5	0	0	0	전북	정읍시	차대차	추돌	진행중 추	운전자법규안전운		
22	2012	2012101014	35 야간	일	1	1	0	0	0	0	경기	화성시	차대차	추돌	진행중 추	운전자법규안전운		
23	2012	2012101014	0 야간	일	1	2	1	0	0	0	충남	서산시	차대차	추돌	진행중 추	운전자법규안전운		
24	2012	2012101013	25 야간	일	1	1	0	0	0	0	광주	서구	차대차	추돌	진행중 추	운전자법규안전운		
25	2012	20121010416	10 주간	수	1	1	0	0	0	0	경기	김해시	차대차	추돌	진행중 추	운전자법규안전운		
26	2012	20121010420	14 야간	수	1	1	0	0	0	0	경기	안성시	차대차	추돌	진행중 추	운전자법규안전운		
27	2012	20121010910	45 주간	월	1	1	0	0	0	0	강원	춘천시	차대차	추돌	진행중 추	운전자법규안전운		
28	2012	20121010623	0 야간	금	1	1	0	0	0	0	대전	유성구	차대차	추돌	진행중 추	운전자법규안전운		
29	2012	20121010510	15 주간	목	1	1	0	0	0	0	충남	태안군	차대차	추돌	진행중 추	운전자법규안전운		
30	2012	20121010518	10 야간	목	1	1	0	0	0	0	경기	포천시	차대차	추돌	진행중 추	운전자법규안전운		
31	2012	20121011320	10 야간	금	1	1	0	0	0	0	인천	계양구	차대차	추돌	진행중 추	운전자법규안전운		
32	2012	20121011316	50 주간	화	1	1	0	0	0	0	서울	강서구	차대차	추돌	진행중 추	운전자법규안전운		
33	2012	20121011222	38 야간	목	1	1	0	0	0	0	전남	순천시	차대차	추돌	진행중 추	운전자법규안전운		
34	2012	20121011300	1 야간	금	1	1	0	0	0	0	경기	양평군	차대차	추돌	진행중 추	운전자법규안전운		
35	2012	20121011316	25 야간	화	1	1	0	0	0	0	인천	강화군	차대차	추돌	진행중 추	운전자법규안전운		

이렇게 많은 데이터 항목을 읽어 온 다음 분석에 필요한 필드에 대해서만 필터링작업이 이뤄지겠죠?



1. 요일별 중대교통사고 사상자 합계분석

```
In [1]: %matplotlib notebook

In [2]: import pandas as pd
import matplotlib.pyplot as plt

In [3]: from matplotlib import font_manager,rc
font_info="c:/windows/fonts/malgun.ttf"
font_name=font_manager.FontProperties(fname=font_info).get_name()
rc('font',family=font_name)

In [8]: data=pd.read_csv('accidentdata.csv')
data.head(5)
```

0. pandas,matplotlib import 및 한글깨짐 문제 해결
파일읽어오는 동작까지 작업내용이에요. 코딩하여 실행해요^^
(공통사항)

	발생 년	발생년월일 시	발생 분	주 야	요 일	사 망 자 수	사 상 자 수	중 상 자 수	경 상 자 수	부 상 자 수	...	도로 형태_ 대분류	도로 형태	당사자 중별_1 당_대 분류	당사자 중별_1 당_대 분류	당사자 중별_2 당_대 분류	당사자 중별_2 당_대 분류	발생위치 X_UTMK	발생위치 Y_UTMK	경도	위도
0	2012	2012050901	10	야 간	수	1	1	0	0	0	...	단 일 로	기 타 단 일 로	승 용 차	대 형	화 물 차	대 형	1032000	1731688	127.853191	35.580719
1	2012	2012033013	32	주 간	금	1	1	0	0	0	...	단 일 로	기 타 단 일 로	화 물 차	소 형	특 수 차	경 형	960392	1924378	127.052937	37.317615
2	2012	2012102620	55	야 간	금	1	2	1	0	0	...	단 일 로	기 타 단 일 로	승 용 차	중 형	승 용 차	소 형	1045245	1799427	128.003224	36.190921
3	2012	2012062814	10	주 간	목	1	1	0	0	0	...	단 일 로	기 타 단 일 로	화 물 차	대 형	승 용 차	소 형	992942	1915569	127.420426	37.238956
4	2012	2012061804	15	야 간	월	1	1	0	0	0	...	기 타/ 불 명	기 타/ 불 명	화 물 차	대 형	화 물 차	대 형	1013547	1663287	127.648393	34.964371

5 rows × 27 columns

1. 사상자수가 3명 이상인 데이터에 대해서만 분석작업을 해야하므로, 필터링작업을 그림처럼 코딩해요.

```
In [12]: d2=data[(data.사상자수>=3)]
d2.head(3)
```

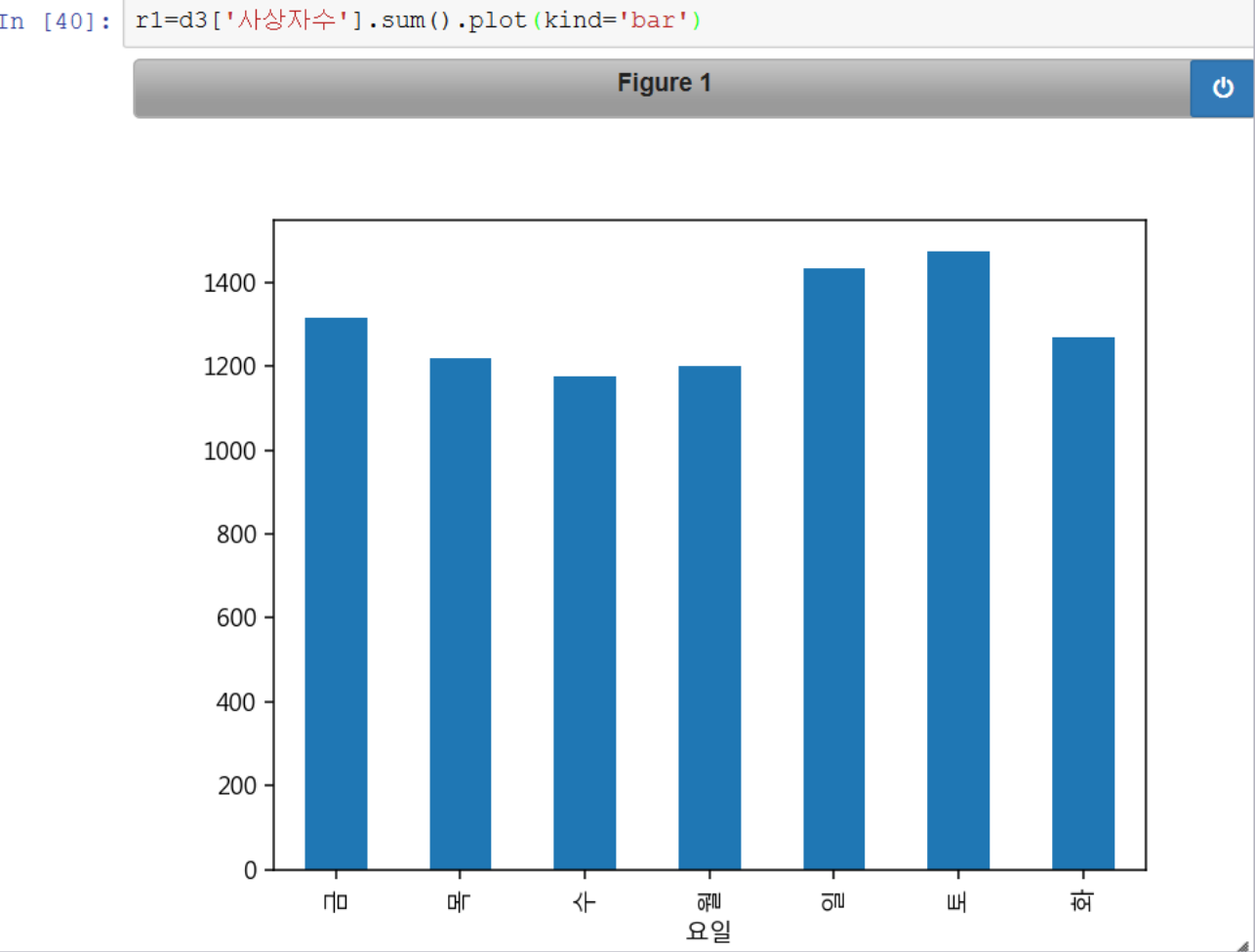
Out[12]:

	발생 년	발생년월일 시	발 생 분	주 야	요 일	사 망 자 수	사 상 자 수	중 상 자 수	경 상 자 수	부 상 고 자 수	...	도로 형태_ 대분류	도로 형태	당사자 종별_1 당_대 분류	당사 자종 별_1 당	당사자 종별_2 당_대 분류	당사 자종 별_2 당	발생위치 X_UTMK	발생위치 Y_UTMK	경도	위도
5	2012	2012071508	4	주간	일	1	5	0	4	0	...	단일로	교량 위	승합차	중형	0	0	1021668	1665087	127.727500	34.987279
8	2012	2012100102	43	야간	월	1	5	1	3	0	...	단일로	기타 단일로	승용차	중형	승용차	중형	1115618	1717852	128.773999	35.449758
10	2012	2012083008	15	주간	목	1	5	1	3	0	...	단일로	기타 단일로	화물차	대형	화물차	소형	1093633	1766952	128.537498	35.894706

2. 요일별로 그룹화하는 작업을 그림처럼 코딩해요.

```
In [22]: d3=d2.groupby('요일')
```

3. 요일별 사상자수 합계를 계산하여 결과를 막대차트로 표시하는 작업을 그림처럼 코딩해요.



2. 경기도내 교통사망사고가 높은 5지역 분석하여 도식화

In [5]:

d1=data[data.발생지시도=='경기']
d1

1. 경기지역을 대상으로 분석하기 위해 필터링하여 d1 객체생성해요.

44	2012	2012010702	27	야간	토	1	1	0	0	0	...	단일로	기타단일로	화물차	소형	0	0	982027	1930644	127.297001	37.374694
53	2012	2012032706	5	주간	화	1	3	2	0	0	...	단일로	기타단일로	화물차	중형	화물차	소형	943286	1927815	126.859543	37.347859
57	2012	2012011005	16	야간	화	1	1	0	0	0	...	교차로	교차로내	승합차	중형	이륜차	소형	972715	1918067	127.192661	37.258442
59	2012	2012011622	40	야간	월	1	4	0	2	1	...	단일로	기타단일로	승용차	중형	승용차	경형	964858	1921721	127.103510	37.293773

In [6]:

da1=d1.groupby(['발생지시군구'])
da1

2. 경기지역의 시.군별 그룹화 작업 실행하는 da1 객체 생성

Out [6]:

<pandas.core.groupby.DataFrameGroupBy object at 0x000000000A26E5F8>

In [7]:

da2=da1['사망자수'].sum()
da2

3. 사망자 수 합계를 요약하여 da2 객체 생성

Out [7]:

발생지시군구
가평군 54
고양시 172
과천시 10
광명시 41
광주시 87
구리시 28
군포시 27
김포시 91
남양주시 150
동두천시 37
부천시 88
성남시 146
수원시 173
시흥시 141

```
In [8]: result=da2.sort_values(ascending=False).head(5)  
result
```

```
Out[8]: 발생지시군구  
화성시      219  
평택시      194  
용인시      176  
수원시      173  
고양시      172  
Name: 사망자수, dtype: int64
```

4. 합계를 내림차순 정렬하여 top5까지만 result 객체로 생성해요.

```
In [10]: result.plot(kind='pie',title='2012-2014년 경기도 교통사망사고 다발지역')
```

5. 원형차트 작성

Figure 1



2012-2014년 경기도 교통사망사고 다발지역



3. 요일별/발생지시도별 교통사고 분석

```
In [5]: df_pv=pd.crosstab(data['요일'],data['발생지시도'])
df_pv
```

Out [5]:

발생지시도	강원	경기	경남	경북	광주	대구	대전	부산	서울	세종	울산	인천	전남	전북	제주	충남	충북
요일																	
금	101	379	205	239	48	76	50	75	174	6	41	70	181	162	41	169	104
목	108	404	177	216	41	78	42	89	174	10	43	79	170	134	42	173	115
수	106	390	185	210	41	71	41	81	171	8	41	68	192	122	49	162	104
월	102	400	211	225	40	69	47	85	174	5	60	66	195	165	28	189	120
일	93	377	150	193	36	90	31	88	137	4	41	47	128	129	40	156	106
토	85	417	189	198	59	87	54	95	172	7	55	76	210	137	38	181	118
화	108	410	185	230	46	74	45	83	171	13	49	80	198	158	37	184	118

1. 교차분석 실행하여 요일별 발생지시도별 데이터 건수를 계산

```
In [6]: result=df_pv.div(df_pv.sum(1).astype(float),axis=0)
result.plot(kind='bar',stacked=True)
```

Figure 1

2. 각 행의 합이 1이 되도록 비율로 설정한 후 그 내용을 누적막대형 차트로 시각화

