

• 봉석이와 세정이가 내주는 •

파이썬 스테디 14장

복습문제!

1번 문제



COVID-19 data

case_id		province	city	group	infection_case	confirmed	latitude	longit
0	1	Seoul	Guro-gu	True	Guro-gu Call Center	70	37.508163	126.884
1	2	Seoul	Eunpyeong-gu	True	Eunpyeong St. Mary's Hospital	14	37.63369	126.9
2	3	Seoul	Seongdong-gu	True	Seongdong-gu APT	13	37.55713	127.0
3	4	Seoul	Jongno-gu	True	Jongno Community Center	10	37.57681	127.
4	5	Seoul	Dongdaemun-gu	True	Dongan Church	9	37.592888	127.056
5	6	Seoul	Jung-gu	True	Jung-gu Fashion Company	7	37.562405	126.984

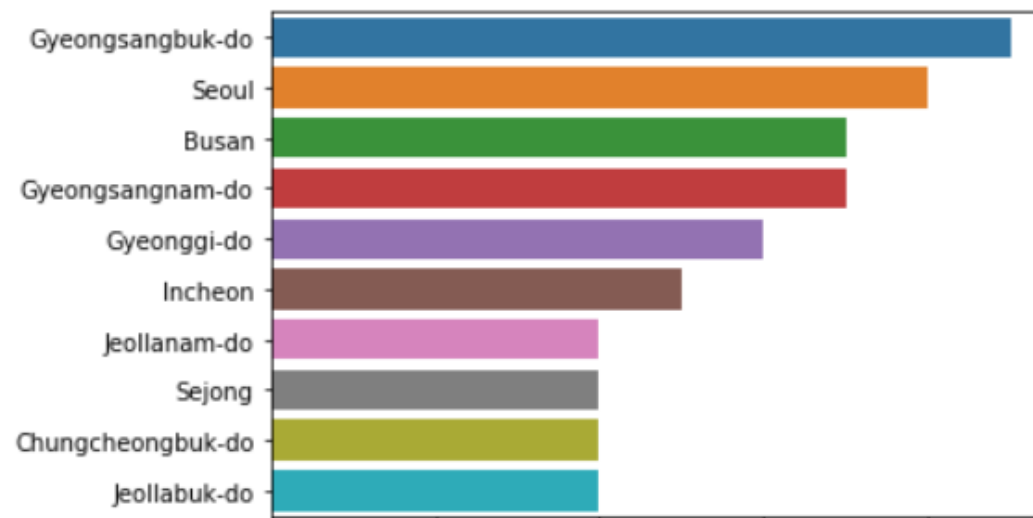
(문제 출제 흐름, 책과 완벽히 같음)

1-1

환자들의 province별 분포를 시각화 해보세요!

결과물

<matplotlib.axes._subplots.AxesSubplot at 0x1e65a3b1f08>



1-2

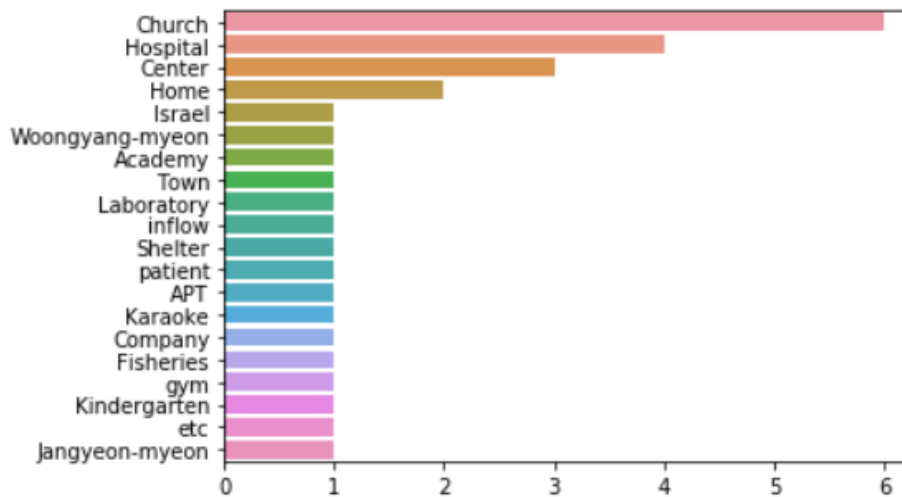
infection_case 의 데이터는 조금 길어서 보기 불편합니다.

다음과 같이 적당히 잘라서 아래 결과를 만들어 보세요. (책 참고)

results	
Church	6
Hospital	4
Center	3
Home	2
Israel	1
Woongyang-myeon	1
Academy	1
Town	1
Laboratory	1
inflow	1
Shelter	1
patient	1
APT	1
Karaoke	1
Company	1
Fisheries	1
gym	1
Kindergarten	1
etc	1
Jangyeon-myeon	1
dtype: int64	

결과물

<matplotlib.axes._subplots.AxesSubplot at 0x1e6580d7f08>



1-3

위의 province에서 경상북도 / 비 경상북도 감염자를 다음과 같이 나누고 is_Gyeongsangbuk-do에 저장하세요.

결과물

```
case[ 'is_Gyeongsangbuk-do' ][60:]
```

```
60    No
61    Yes
62    Yes
63    Yes
64    Yes
65    Yes
66    Yes
67    Yes
68    Yes
69    Yes
70    No
71    No
72    No
73    No
74    No
75    No
76    No
77    No
```

```
Name: is_Gyeongsangbuk-do, dtype: object
```

1-4

infection_case, 와 *is_Gyeongsangbuk-do* 를 그룹별 합계를
다음과 같이 만드세요.

결과물

by_[:10]

is_Gyeongsangbuk-do	No	Yes
infection_case		
Bonghwa Pureun Nursing Home	0.0	1.0
Bundang Jesaeng Hospital	1.0	0.0
Changnyeong Coin Karaoke	1.0	0.0
Cheongdo Daenam Hospital	2.0	1.0
Dongan Church	1.0	0.0
Eunpyeong St. Mary's Hospital	1.0	0.0
Geochang Church	1.0	0.0
Geochang-gun Woongyang-myeon	1.0	0.0
Goesan-gun Jangyeon-myeon	1.0	0.0
Guro-gu Call Center	3.0	0.0

1-5

다음과 같이 `indexer`를 만들고, `take`를 사용해서 행을 정렬된 순서 그대로
선택하고 마지막 10개 행(가장 큰 값)만 잘라내세요. (책 참고)

결과물

```
indexer[:10]
```

infection_case	
Bonghwa Pureun Nursing Home	0
Bundang Jesaeng Hospital	26
Changnyeong Coin Karaoke	25
Cheongdo Daenam Hospital	23
Dongan Church	22
Eunpyeong St. Mary's Hospital	21
Geochang Church	19
Geochang-gun Woongyang-myeon	18
Goesan-gun Jangyeon-myeon	17
Guro-gu Call Center	16
dtype: int64	

```
count_subset
```

	is_Gyeongsangbuk-do	No	Yes
infection_case			
Gyeongsan Cham Joeun Community Center		0.0	1.0
Geochang-gun Woongyang-myeon		1.0	0.0
Onchun Church		2.0	0.0
gym		2.0	0.0
Guro-gu Call Center		3.0	0.0
Cheongdo Daenam Hospital		2.0	1.0
overseas inflow		6.0	0.0
contact with patient		8.0	0.0
Shincheonji Church		13.0	1.0
etc		16.0	1.0

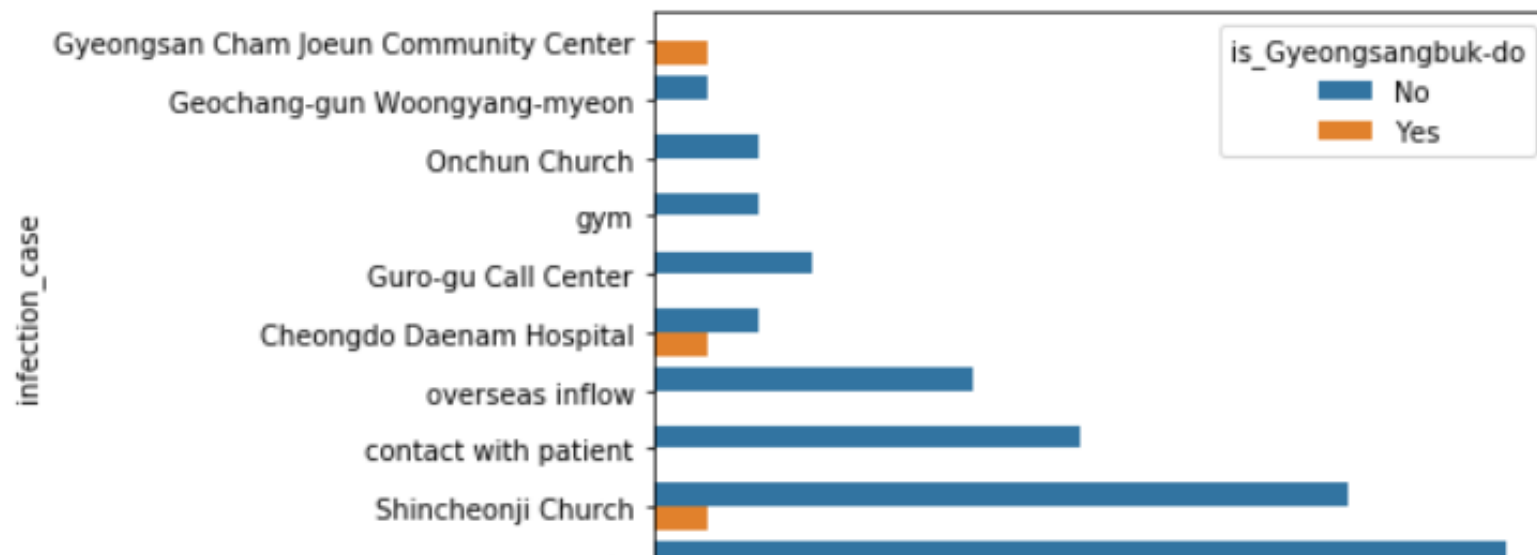
1-6

다음과 같이 시각화 해보세용

`x='total',y='infection_case',hue='is_Gyeongsangbuk-do'`

결과물

<matplotlib.axes._subplots.AxesSubplot at 0x1e65b529508>





중고차 색깔 별 평균 가격 및 평균 주행거리를 구하세요.
(pivot_table 함수 이용)

결과물

Color	other	silver	white
Odometer	37481.000000	31975.833333	37967.486486
Price	14618.969697	15220.200000	14682.432432



중고차를 색깔 별로 그룹화하고, 그룹별 총 주행거리를
계산해보세용 (groupby 함수 이용)

결과물

```
Color
other    1236873
silver    959275
white    1404797
Name: Odometer, dtype: int64
```



색깔 변수를 더미변수화하여 가격을 y변수로 놓고,
색깔 변수와 주행거리를 X변수로 하여 lm 모델의
Coefficient를 구해보세요!

결과물

	Price	Odometer	Color_other	Color_silver	Color_white
0	14636	37388	0	0	1
1	14122	44758	0	0	1
2	14470	45854	0	0	1
3	15072	40149	0	0	1
4	14802	40237	0	0	1

```
lm.coef_
```

```
array([-5.55395141e-02, -1.28652660e+02,  1.66823362e+02, -3.81707015e+01])
```

• 봉석이와 세정이가 내주는 •

파이션 스터디 14장

복습문제 끝!