

빅데이터분석프로그래밍(30)

일자 : 2024.4.24 수요일 학번 _____ 이름 _____

1. 다음 서술내용이 맞으면 O, 틀리면 X 하세요.

- ① 수업에서 활용하는 개발환경에서 matplotlib의 한글 지원을 위해 패키지 koreanize_matplotlib를 설치해 사용한다. ()
- ② 패키지 seaborn은 matplotlib를 기반으로 활용되는 시각화 도구이다. ()

2. 다음에서 빈 부분의 영어를 적절히 채우세요.

- ① 수업에서 결측값의 시각화에 사용한 함수 matrix()가 있는 패키지 이름은 ()이다.
- ② 패키지 seaborn에서 내장 데이터를 반환 받을 수 있는 함수는 ()이다.
- ③ 구글에서 주피터 노트북을 지원하는 클라우드 서비스는 구글 ()이다.

■ 다음 코드를 가정하고 문제를 해결하세요.

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

3. 다음 각각의 문제에서 물음에 알맞은 것을 고르세요.

- ① Pandas의 특징으로 잘못된 것은? ()
- 가) 표 형식의 데이터나 다양한 형태의 테이블을 처리하기 위한 라이브러리
- 나) 주 자료 구조는 배열과 데이터프레임
- 다) 외부 모듈로 설치가 필요
- 라) 데이터프레임은 테이블 형태의 자료

- ② 판다스의 DataFrame의 메소드가 아닌 것은? ()

가) read()
나) to_csv()
다) describe()
라) info()

- ③ 다음에서 잘못된 것은? ()

가) 데이터프레임의 열은 변수라고도 부른다.
나) 결측값은 np.na로 코딩할 수 있다.
다) 데이터프레임의 isna()로 결측값인지의 논리값을 알 수 있다.
라) 데이터프레임의 isnull()은 isna()와 기능이 같다.

4. 다음 코드의 결과에서 빈 부분을 완성하세요.

```
df1 = pd.DataFrame({'var1' : [1, 2, 1],
                    'var2' : [2, 3, 2]})
df1.rename(columns = {'var2' : 'v2'}, inplace = True)
df1.rename(index = {0 : 'x', 2 : 'y'}, inplace = True)
df1
```

	var1	
	1	2
1	2	3
	1	2

* 참고: 수업에서 활용한 데이터 mpg

```
mpg.columns
Index(['manufacturer', 'model', 'displ', 'year', 'cyl', 'trans', 'drv', 'cty',
      'hwy', 'fl', 'category'],
      dtype='object')
```

5. 다음 코드와 동일한 기능을 메소드 isin()을 사용해 구현하기 위한 빈 부분을 완성하세요.(4)

```
mpg = pd.read_csv('mpg.csv') # mpg 데이터 불러오기
mpg['size'] = np.where((mpg['category'] == 'compact') |
                     (mpg['category'] == '2seater'),
                     'small', 'large')
```

```
mpg['size'] = np.where(
    _____,
    'small', 'large')
```

6. 다음 코드의 빈 부분을 완성하세요.

① 다음 결과가 나오도록 빈 부분을 완성

```
mpg.  ('drv').  (
    mean_cty = ('cty', 'mean'),
    mean_hwy = ('hwy', 'mean'))
```

	mean_cty	mean_hwy
drv		
4	14.33	19.17
f	19.97	28.16
r	14.08	21.00

② 고속도로 연비(hwy)가 40.5 초과하는 행 추출

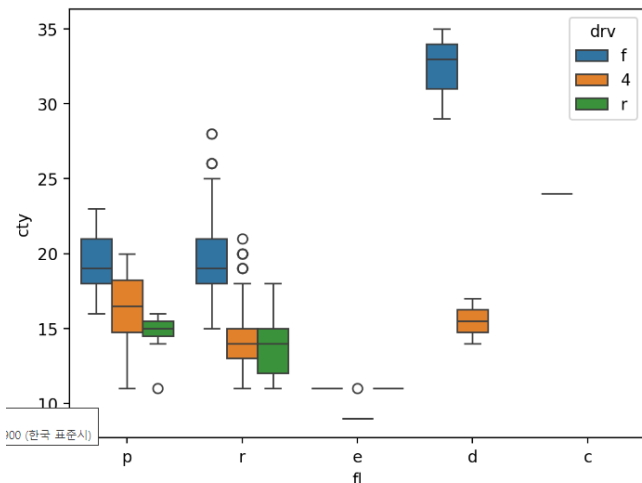
```
mpg.  (  )
```

③ 고속도로 연비(hwy)와 도심 연비의 평균을 열 mpg_mean에 저장해 데이터프레임 전체를 반환하는 코드

```
mpg.  (  =
    (mpg.hwy + mpg.cty) / 2)
```

④ 다음 그래프가 그려지도록

```
sns.boxplot(mpg, x =  , y = 'cty',
              = 'drv');
```



7. 다음 두 데이터프레임 df1, f2의 merge() 코드에서 출력 값을 쓰세요.

df1:	sub	score	df2:	sub	book
0	AI	80	0	AI	30000
1	Python	77	1	Python	28000
2	Java	99	2	C	25000

① df1.merge(df2).shape[1]

② df1.merge(df2, how='left')['book'][1]

③ df1.merge(df2, how='outer').iloc[2, 0]

④ len(df1.merge(df2, how='cross'))

8. 다음 데이터프레임 d에서 코드의 출력 값을 쓰세요.

	carat	cut	color	clarity
0	0.90	Premium	E	VS2
1	NaN	Ideal	H	SI2
2	0.58	NaN	G	NaN
3	1.51	Premium	G	VS2
4	NaN	NaN	NaN	VVS2

① d.dropna().shape[0]

② d.isna().any().sum()

③ d.isna().any(axis=1).sum()

<수고 하셨습니다.>