

1. 다음 중 누락 데이터의 표현이 아닌 것은?

- ① None ② null ③ NaN ④ na ⑤ SyntaxError

2. pandas의 자료형으로 맞는 것을 고르시오.

- ① 리스트 ② DataFrame ③ 튜플 ④ 딕셔너리 ⑤ set

3. pandas에서 누락 데이터를 제거하는데 사용되는 함수를 고르시오.

- ① mean ② median ③ groupby ④ dropna ⑤ concat

4. pandas에서 컬럼의 항목별 개수를 알려주는 함수를 고르시오.

- ① groupby ② drop ③ value\_counts  
④ concat ⑤ replace

5. 산점도와 회귀 직선을 동시에 표현하는 코드로 옳은 것을 고르시오.

- ① sns.heatmap(data)  
② sns.scatterplot(data, x, y)  
③ sns.lmplot(data, x, y)  
④ sns.histplot(data, x, y)  
⑤ sns.boxplot(data, x, y)

6. 다음과 같이 코드를 작성했을 때, 실행 결과로 알맞은 것은?(주관식)

```
arr = np.arange(10)  
print(arr)
```

7. Hypertext Markup Language의 약자로 웹 페이지를 만드는 데 사용하는 언어는 무엇인가?(주관식)

8. HTML 문서를 탐색해서 원하는 부분만 쉽게 뽑아낼 수 있는 기능을 제공하는 라이브러리를 고르시오.

- ① random ② selenium ③ datetime  
④ BeautifulSoup ⑤ math

9. pandas DataFrame 자료형은 숫자 인덱스와 문자형 인덱스를 모두 다 사용할 수 있다. (O, X 문제)

10. 이상치(outlier)란 수집된 데이터 중에서 값이 존재하지 않는 누락된 데이터이다. (O, X 문제)

11. `pd.DataFrame.describe()`가 제공하지 않는 요약 통계량을 고르시오.

- ① ver ② std ③ mean ④ min ⑤ max

12. 범주가 있는 데이터에 대해서, 범주 별 빈도수의 크기를 직사각형의 막대로 표현하는 그래프를 시각화하는 코드로 옳은 것을 고르시오.

- ① `plt.pie(ratio, labels, colors)`
- ② `sns.boxplot(data, x, y)`
- ③ `sns.countplot(data, x, hue)`
- ④ `plt.hist(x)`
- ⑤ `sns.boxplot(data, x, y)`

13. 데이터프레임에서 가장 많이 등장하는 최빈값을 구하는 코드로 옳은 것을 고르시오.

- ① `pd.dataframe.std()`
- ② `pd.dataframe.var()`
- ③ `pd.datafame.quantile()`
- ④ `pd.datafame.max()`
- ⑤ `pd.dataframe.mode()`