2024년 제2회 청소년 IT경시대회 기출문제

데이터분석 부문

시행 | 2024년 3월 16일 (토)

주최 | 한국정보기술진흥원　　후원 | 서울특별시의회

< 수험자 유의사항 >

1. 본 진흥원이 주최하는 경시대회의 시험문제는 저작권법에 따라 보호되는 저작물이며, 시험문제의 일부 또는 전부를 무단 복제, 배포, (전자)출판 하는 등 저작권을 침해하는 경우 저작권법에 의하여 민·형사상 불이익을 받을 수 있습니다.
2. 신분증을 지참하지 않은 자는 시험에 응시할 수 없습니다.
 - 인정 신분증의 범위: 주민등록증, 여권, 청소년증
3. 시험 중에는 어떠한 통신기기 및 전자기기(휴대전화, 스마트폰, 태블릿PC, 스마트워치, 이어폰, 전자사전 등)도 소지 및 사용할 수 없으며, 적발 시에는 부정행위로 처리됩니다.
4. 시험 중에는 퇴실 할 수 없으며, 응시자 이외에는 시험장에 출입하실 수 없습니다.
5. 부정한 방법으로 시험에 응시하거나 시험에서 부정행위를 한 자에 대해서는 해당 회차의 시험을 정지시키거나 합격을 무효로 하며, 이후 3년간 본 진흥원에서 주최하는 시험에 응시하실 수 없습니다.

1. 주어진 numpy 배열 arr이 있을 때, 배열의 모든 원소를 제곱한 새로운 배열을 생성하는 코드는?

① new_arr = arr.square()
② new_arr = square(arr)
③ new_arr = arr ** 2
④ new_arr = numpy.square(arr)
⑤ new_arr = arr.power(2)

2. 다음 코드의 실행 결과를 고르시오.

```
import numpy as np

arr = np.arange(1, 6)
result = arr * 2 + 3
print(result)
```

① [5 7 9 11 13]
② [2 4 6 8 10]
③ [6 8 10 12 14]
④ [3 5 7 9 11]
⑤ [1 3 5 7 9]

3. 다음 코드의 실행 결과를 고르시오.

```
import numpy as np

arr = np.array([[1, 2], [3, 4]])
result = arr.flatten()
print(result)
```

① [[1 2] [3 4]]
② [1 2 3 4]
③ [[1] [2] [3] [4]]
④ [1 3 2 4]
⑤ [1 1 2 2 3 3 4 4]

4. pandas 데이터프레임 df에서 열 'A'의 평균을 계산하는 올바른 코드를 고르시오.

① mean = df['A'].avg()
② mean = df.mean('A')
③ mean = mean(df['A'])
④ mean = df['A'].mean()
⑤ mean = df.mean()[A]

5. 다음 코드의 실행 결과에 대한 설명으로 가장 올바른 것을 고르시오.

```
import pandas as pd

data = {'A': [1, 2, 3],
        'B': ['a', 'b', 'c']}
df = pd.DataFrame(data)
result = df.describe()
print(result)
```

① 데이터프레임의 크기
② 통계적 요약 정보가 포함된 데이터프레임
③ 열 'A'와 'B'의 데이터 유형
④ 열 'A'와 'B'의 최솟값, 최댓값, 평균 등의 통계적 정보
⑤ 오류가 발생한다.

6. 다음 코드의 실행 결과를 고르시오. (선택지에 있는 \n은 줄바꿈을 의미합니다.)

```
import pandas as pd

data = {'A': [1, 2, 3, 4, 5],
        'B': ['a', 'b', 'c', 'd', 'e']}
df = pd.DataFrame(data)
result = df.loc[df['A'] > 2, 'B']
print(result)
```

① 2 c\n3 d\n4 e
② c 2\nd 3\ne 4
③ c 3\nd 4n\e 5
④ a 3\nb 4\nc 5
⑤ 3 c\n4 d\n5 e

7. 다음 코드를 실행시켰을 때, 그래프의 모양에 대한 설명으로 옳은 것을 고르시오.

```
import matplotlib.pyplot as plt
import numpy as np

x = np.linspace(0, 10, 100)
y = np.sin(x)
plt.plot(x, y, linestyle='--', color='r', marker='o')
plt.xlabel('X-axis')
plt.ylabel('Y-axis')
plt.title('Sine Wave')
plt.grid(True)
plt.show()
```

① 선 그래프이며, 빨간색 점선으로 그려진다
② 선 그래프이며, 파란색 실선으로 그려진다
③ 산점도 그래프이며, 빨간색 실선으로 그려진다
④ 산점도 그래프이며, 파란색 점선으로 그려진다
⑤ 그래프가 그려지지 않는다

8. 다음 코드의 실행 결과를 고르시오.

```
import pandas as pd

data = {'Name': ['Alice', 'Bob', 'Charlie',
'David'],
        'Age': [25, 30, 35, 40],
        'Score': [85, 90, 88, 75]}

df = pd.DataFrame(data)

df['Score'] += 5
df['Age'] *= 2

result = df[df['Age'] > 50]['Score'].mean()

print('%.1f' % result)
```

① 80.0 ② 83.3
③ 88.0 ④ 89.3
⑤ 90.0

9. 다음 코드의 실행 결과를 고르시오.

```
import pandas as pd

data1 = {'Name': ['Alice', 'Bob', 'Charlie'],
         'Age': [25, 30, 35]}
data2 = {'Name': ['David', 'Emily', 'Frank'],
         'Age': [40, 45, 50]}

df1 = pd.DataFrame(data1)
df2 = pd.DataFrame(data2)

df = pd.concat([df1, df2], ignore_index=True)
df['Age'] *= 2

result = df[df['Age'] > 50]['Age'].sum()

print(result)
```

① 70 ② 100
③ 180 ④ 225
⑤ 400

10. 데이터 분석에서 이상치에 대한 설명으로 가장 올바른 것을 고르시오.

① 데이터 집합의 중앙값을 의미한다.
② 데이터 집합 내에서 가장 일반적인 값을 의미한다.
③ 데이터 집합에서 평균과 중앙값의 차이를 의미한다.
④ 데이터 집합에서 예상한 값과 가장 근접한 값을 의미한다.
⑤ 데이터 집합에서 일반적인 패턴에서 벗어난 극단적인 값을 의미한다.

11. 다음 코드의 실행 결과를 작성하시오. (힌트: 정수)

```
import numpy as np
import pandas as pd

data = {'A': [1, 2, 3, 4, 5],
        'B': [6, 7, 8, 9, 10]}

df = pd.DataFrame(data)

df['C'] = np.where(df['A'] < 3, 'Low', 'High')
df['D'] = df['A'] * df['B']

result = df[(df['C'] == 'Low') & (df['D'] > 20)]['A'].count()

print(result)
```

12. 다음 코드의 실행 결과는 <아래>와 같습니다. 실행 결과의 빈 칸에 들어가는 값을 작성하시오. (힌트: "YYYY-MM-DD" 형식의 날짜)

```
import pandas as pd

data = {'A': pd.date_range(start='2023-01-01', periods=100),
        'B': pd.Series(range(100)),
        'C': pd.Series(range(100, 0, -1))}

df = pd.DataFrame(data)
df['Year'] = df['A'].dt.year
grouped_max = df.groupby('Year').max()
print(grouped_max)
```

```
< 실행 결과 >
          A   B    C
Year
2023 빈칸  99  100
```

13. 데이터 활용 관리란 데이터의 활용 여부를 점검하거나 활용도를 높이기 위해 측정 대상 데이터와 품질 지표를 선정하여 품질을 측정하고 분석하여 품질을 충족시키지 못하는 경우, 원인을 분석하여 담당자로 하여금 조치하도록 하는 작업을 말합니다. 관리 대상 중 핵심 데이터는 회사의 고객, 프로세스, 시장 환경, 재무 정보 등에 직접적으로 영향을 미치는 중요성이 높은 데이터를 말하며, 특정 기준에 따라 관리되어야 합니다. 이 관리 기준의 설명을 보고 어떤 특성에 대한 설명인지 <보기>에서 골라 작성하세요.

```
< 보기 >
완전성 일관성 최신성 유효성 유일성 명확성
```

1) 데이터의 모든 값은 의미 있게 채워져 있어야 한다
2) 데이터의 값은 업무 규칙을 준수해야 한다
3) 데이터의 값은 실제 세계의 객체들이 가지고 있는 값과 같아야 한다
4) 데이터의 값은 동일하게 관리되어야 한다
5) 데이터의 값은 동일 테이블에서 중복 관리 되어서는 안된다
6) 데이터의 의미가 혼동되지 않도록 분명하게 관리되어야 한다

2024년 제2회 청소년 IT경시대회 기출문제

데이터분석 부문 정답

1	2	3	4	5
3	**1**	**2**	**4**	**2**
6	7	8	9	10
1	**1**	**4**	**5**	**5**
11	12			
0	**2023-04-10**			
13				
완전성, 유효성, 최신성, 일관성, 유일성, 명확성				