

2020 NIPA 믹스앤매치 인공지능/데이터분석 아카데미

입문 오프라인 교육과정 실습 문제지

분반:

팀 번호:

이름:

모든 답안은 제시된 답안 형태에 맞게 작성되어야 합니다. 문제 자체에 대한 궁금한 사항이 있을 경우 강사/조교에게 자유롭게 질문해주세요.

1. world_corona 데이터프레임에서 male_smokers 칼럼을 삭제하는 코드를 작성하시오(1점)
2. world_corona의 수치 칼럼에 대해 기술통계량을 한 번에 볼 수 있는 코드를 1줄로 작성하시오(1점)
3. world_corona의 모든 칼럼명을 출력하여 볼 수 있는 코드를 작성하시오(1점)
4. world_corona 데이터프레임에 'is_valid'라는 칼럼을 추가하되, 'is_valid' 칼럼의 기본값은 1으로 설정하시오(2점)
5. world_corona 데이터에서 location이 'France'인 데이터만 추출하는 코드를 작성하시오(2점)
6. seoul_df와 seoul_info를 seoul_df를 기준으로 병합(left join)하는 코드를 작성하시오(3점)
7. world_corona 데이터를 인구 순으로 내림차순 정렬하는 코드를 작성하시오(1점)
8. drop_duplicates 메서드를 활용하여, location별로 가장 마지막 날짜의 데이터만 남기는 코드를 작성하시오(2점)
(코드 실행 결과 총 row의 개수는 213개가 되어야 함)
9. world_corona 데이터를 지역('location')별로 그룹화하되, 그룹화 방식은 합계(sum)으로 진행하시오. 이 때, 해당 그룹화된 데이터에서 'new_cases'만을 출력하시오(2점)
10. world_corona에는 총 몇 개의 국가('location')가 있는가?(1점)
11. world_corona에서, 2020년 10월 30일에 하루동안 발생한 전세계 확진자는 몇 명인가?(2점)
12. world_corona에서 'Continent'가 'Europe'일 때, 총 몇 개의 location이 있는가? (unique한 location을 의미)(2점)
13. world_corona에서, 'Japan'의 총 확진자는 몇 명인가?(2점)
14. world_corona에서, 2020년 7월 'France'의 확진자(new_cases)는 'South Korea'의 몇 배인가?(소수점 첫째 자리까지)(3점)
(힌트: pd.to_datetime()을 활용하면 pandas의 날짜 자료형 확인 가능)
15. world_corona에서, 1인당 gdp('gdp_per_capita')가 가장 높은 location은 어디인가?(1점)
16. seoul_df에서, 가장 많은 확진 요인('접촉력')을 차지하는 것은 무엇인가?(1점)
17. seoul_df에서, 가장 많은 확진자가 발생한 '지역'은 전체 서울 확진자의 몇 %를 차지하는가? (3점)
(정답은 '1.23%'와 같은 포맷으로 입력)
18. seoul_df에서, 가장 많은 월간 확진자가 발생한 때는 몇 월인가?(2점)
(정답은 '10월', '1월', '3월'과 같은 포맷으로 입력)
19. world_corona에서, 총 사망자가 두 번째로 적은 continent는 어디인가?(2점)
(힌트: groupby 메서드 사용)
20. world_corona에서, 전세계에서 그 간 발생한 총 사망자는 몇 명인가?(1점)
21. world_corona에서, location이 'French Polynesia'인 지역의 iso_code는 무엇인가?(1점)
22. world_corona에서, 총 확진자('total_cases')와 총 사망자('total_deaths') 간의 상관관계는 어떻게 되는가?(1점)
(소숫점 이하 둘째자리까지 출력)
23. world_corona에서, location이 'Germany'인 데이터 중 확진자가 가장 많이 발생한 월의 확진자는 직전 월의 확진자 수의 몇 배인가?(3점)
(소숫점 첫째 자리까지, "3.4배"와 같은 형태로 작성)
24. world_corona에서, continent가 'Europe'인 곳들 중 확진자 수가 두 번째로 많은 지역(location)은 어디인가?(3점)
25. world_corona에서, 확진자 대비 사망자 비율이 가장 높은 지역의 비율은 얼마인가?(3점)
(소숫점 첫째 자리까지, 입력 예시 33.5%)
26. world_corona에서, 행 번호가 1114인 데이터의 'total_deaths'의 값은 무엇인가?(1점)

27. world_corona에서, 확진자가 가장 많이 발생한 상위 10개 지역('location')의 확진자를 막대 그래프로 시각화하시오(3점)
- 힌트1: groupby를 사용
- 힌트2: sort_values()를 사용
- 힌트3: 슬라이싱[:n] 또는 .head()를 사용하여 상위 n개를 출력
28. world_corona에서, 한국('South Korea')의 일별 확진자 추이를 선 그래프로 출력하시오(1점)
29. world_corona에서, 전세계 월별 확진자 추이를 선 그래프로 출력하시오(2점)
(단, 11월과 12월 데이터는 제외하시오)
- 힌트: world_corona['month'] = pd.to_datetime(world_corona['date']).dt.month를 사용하면 특정 월 칼럼을 추가할 수 있다
30. world_corona에서, 일본('Japan')의 일별 확진자와 누적 확진자 추이를 선 그래프로 동시에 표현하시오(2점)
- 힌트: plt.plot을 한 셀에 같이 사용하면 그래프를 겹쳐 그릴 수 있다

답안 제출 링크: <https://forms.gle/VhpNG8SwnboPw9Ri7>