**사망사고 데이터 분석**

다음은 **2016년 일년**동안 전국에서 사망사고가 발생한 교통사고 데이터이다. 데이터를 불러와서 아래 문제를 풀어보세요.

**※** 반드시 **적절한 함수를 활용**하여 답을 찾으세요.

가) 작업경로 확인 및 불러오기 (2점)

a)　다운받은 csv형식의 파일을 불러 오세요. (**불러올 때 변수명을 data\_raw로 설정)**

나) 데이터 확인 (8점)

a) 불러온 데이터의 **상위 6개, 하위 6개** 관측치를 미리보기 하세요. (2점)

b) 데이터구조, 변수개수, 변수 명, 관찰치 개수를 한 번에 확인 할 수 있는 함수를 활용하여 확인하세요. **위도, 경도**는 어떤 형식의 데이터인가요? (숫자형, 문자형, 논리형 중 선택) (3점)

c) **데이터 객체의 차원(행,열)**을 확인하세요. (3점)

다) 데이터 전처리 (10점)

a) “발생년월일시”열을 활용하여 “발생년월일” 새로운 열을 추가하세요. (5점)

ex) **20161223**20 → 20161223 로 변경

b) “발생년월일시”열과 “발생분”열을 활용하여 “시분” 새로운 열을 추가하세요. (5점)

ex) 20161223**20** **35**→ 2035 로 변경

라) 데이터 분석 (50점)

a) 2016년에 **“사망자수”, “사상자수”, “중상자수”, “경상자수”, “부상신고자수” 각 열의 총합**을 계산하세요. (5점)

b) 2016년에 **사망사고가 가장 많이 발생한 날짜**는 몇월 몇일인가요? (5점)

c) 2016년에 사망사고는 **주간**에 많이 발생하나요?　**야간**에 많이 발생하나요? (5점)

d) 2016년 **대전 지역의 사망사고 건수**는 몇건인가요? (5점)

e) 요일별로 사망사고가 발생한 건수를 계산하세요. **어떤 요일에 사망사고**가 가장 많이 발생했나요? (5점)

ex)　월요일 10건, 화요일 11건...일요일 20건

f) **“사상자수”가 가장 많이 발생한 요일**은 무슨 요일인가요? (5점)

g) **“사상자수”가 가장 많이 발생한 사고(행)**만 추출하세요. (5점)

h) “사고유형대분류” 별로 건수를 계산하세요. (5점)

i) **이륜차와 연관된 사고** 건수는 몇건인가요? (5점)

j) **“도로형태대분류”의 터널안**에서 발생한 사고는 몇건인가요? (5점)

마) 정부는 보행자 사고의 심각성을 깨닫고 보행자 사망사고 예방 캠페인을 펼칠 예정이다. 이에 앞서, 보행자 사망사고의 심각성을 데이터에 기반하여 살펴보고 대책을 마련하고자 한다. (20점)

a) 2016년에 보행자 사망사고는 전체 사망사고 중 몇%를 차지하는가요? (5점)

b) 보행자 사망사고 중, “법규위반”별로 건수를 계산하세요. 과속은 몇건인가요? (5점)

c)　보행자 사망사고 중, “사고유형중분류”중 **횡단중**에 발생한 사고의 비율을 계산하세요. (5점)

d)　보행자 사망사고가 많이 발생하는 시간대는 언제인가요? (5점)

바) 1인가구 증가와 배달문화의 확대로 이륜차 사고가 증가하고 있다. 정부에서는 이륜차 사고의 심각성을 파악하고자 한다. 아래의 물음에 답하세요. (10점)

a) 이륜차와 연관된 사망사고만을 필터하세요 ? (5점)

b) **야간**에 발생한 **지역별(발생지시도) 이륜차 사망사고 건수**를 Count하세요. (5점)