

## 데이터 사이언스 개론 2019 중간

### 12문제 서술형 단답식

1. 데이터 사이언스의 정의와 과정 설명하기.
  2. 데이터 사이언티스트의 자질(profile)
  3. 평균, 중간값, 최빈값, 중간값 구하기 (7개 숫자중에서)
  4. 3번의 7개 숫자로 outlier구하기
  5. 데이터 탐색적 분석 정의나오고 이거 쓰는 것이 답이었어요.
  6. 차원의 저주 뜻이 뭔지 적기
  7. 모델링이 무엇인지 적기
  8. 웹페이지에서 원하는 텍스트 또는 각종 미디어자료를 추출하여 다운받는 것은 무엇일까.(답 : crawling)
  9.  $X_1(1.2)$   $x_2(3.5)$  manhattan, Euclidean, supremum distance 구하기
  10.  $X_1(1.2)$   $x_2(3.5)$  은 동일한 채로 cosine similarity 구하기
- (8, 9가 한 문제에 나왔는지 아닌지가 기억이 제대로 안나 일단 분리 해렸습니다.)
11. Support랑 confidence구하고 (Beer, Diaper 있는 프린트 내용과 동일했어요.) attribution 판단하기, apriori 사용해야 해요!
  12. 매트릭스(dissimilarity였던 것 같아요.) 채우고 가장 비슷한 object구하기

배점상 11, 12번 문제가 제일 중요하고 나머지는 단답 인 것 같아요. 교수님이 설명하시는 부분에서 주로 어렵지 않게 나오기 때문에 수업 열심히 들으시면 잘 보실 수 있으실 거예요.