문제점

1. 정확도를 어떻게 측정할 것인지
2. 정답지 조차 임의로 정해진거

okt 만 사용했지만,

해시태그가 대표성을 나타내지 못하므로

분류는 구석구석으로 하고 숫자 데이터는 네이버 리뷰

네이버 리뷰로 분류를 나타낼 수는 없다.

네이버 리뷰를 분류한다고 해서

네이버 방문자 수로는 분류를 해도

로그 변환을 한 상태로

로그 변환을 한 상태에서 정규화를 한 사례가 있는지 확인

-> 로그변환 -> 정규화 \_> 정규성 검정하고, 확인

1. 로그변환
2. 비모수적 방법
3. box-cox

군집화 -> 속성(변수)

추출한 데이터

근데 댓글 수가 많다고 인기 관광지인걸 알겠는데 어떤 방식으로 증명하지?

진행하는 사람과 크롤링하는 사람 분류

방법론

1. 통계적 방법론적으로 접근
2. 사분위 수 기준으로 3개로 나누는 거

음식 사용자의 선택을 받고

1. 관광이랑 음식으로 나누고,
2. 2차 리뷰 참여자 수를 기준으로 나눔(군집, 사분위수)

관광 나누는게 힘들지는 않을까아아아아아

* 리뷰 부분도 해결

자신만의 로직 만들어오기

4분위수 군집분석

정확도검사

구석구석 해시태그로 이용했을 때, ->

미드 클래스를 다시 분류 - 수린님이 분류하신거랑 상이

100%

정답

정확도 부분에서는 분류를 다시 한

450

80% train