1. k-means clustering 을 사용하는 이유

- 목적 : 배달 서비스를 제공하는 배달 기사의 휴게소 제공 => 인접한 가게가 최대가 되는 장소가 적합

- 거리기반 군집화 알고리즘 k-means clustering이 적합.

- station을 설치 하기 위한 예산이 정해져 있음을 가정 => k 개수를 미리 설정할 수 있는 k-means가 유리.

2. 다른 군집화 기술과의 차별성

- 평균이동 (Mean Shift)

거리 기준으로 clustering을 하던 k-means와는 다르게 데이터의 밀도를 기준으로 clustering을 진행.

장점 :

데이터 세트의 형태를 특정 형태로 가정하지 않기 때문에 좀 더 유연한 군집화 가능

미리 군집의 개수 정할 필요 없음

단점 :

알고리즘의 수행시간이 오래 걸림

3. k-means clustering 매개변수 선정 이유 – k, replicates

k: 현재 1가게 1배달의 서비스 제공이 활발히 이루어 고 있음.

한 station에서 수용가능한 배달 기사들의 수를 고려

주차공간, 휴식공간 고려, 20명 수용으로 가정

1가게 1배달의 원칙을 적용 한다면 하나의 station을 이용 가능한 가게는 총 20가게.

표본 데이터 => 약 4000개 내외. 따라서 200개의 station필요

k-means에서 k를 200으로 선정.

replicates : 반복 횟수는 해당 값이 수렴하게 되는 값으로 설정.

5번의 반복으로 수렴점 도달.

4. 한계점 : 가게가 밀집 되어있지 않은 지역은 k-means clustering을 통해

인접한 곳에 배달 station을 배정받지 못하게 됨.

5. 극복 - 후처리 추가 : clustering 된 점들의 반경에 포함되지 않는 배달 상점들에 대한

군집화를 다시 진행. 이때는 거리 기반이 아닌 밀도 기반으로 진행하는 평균이동

군집화를 진행. 미리 군집화 할 k를 정할 필요가 없기 때문에 적합.

k-means clustering을 진행한 뒤 남은 데이터들을 clustering하는 것이기 때문에

속도에 대한 단점 상쇄