

< dplyr 패키지 >

slice() 함수 # 행 슬라이스 예) | slice(delivery, c(1, 3, 5:10))

slice(데이터명, 잘라서 가져올(행)).

filter() 함수 # 조건에 맞는 행만 가져오기 예) | filter(delivery, 사군구 == "성북구")

filter(데이터명, 조건) 예2 filter(delivery, delivery\$사군구 == "성북구")

#복수조건

예) 3. filter(delivery, 사군구 == "성북구", 요일 %in% c("토", "일"), 업종 == "패션/뷰티")

arrange() 함수 # 정렬하기 (오름차순) 예1) arrange(delivery, 사군구, 요일, 업종)

예2) arrange(delivery, desc(사군구), 요일, 업종)

select() 함수 # 변수 선택하기 OR 제외하기.

내장함수

예1) select(delivery, 통화전수) # 변수 선택

예2) select(delivery, 사군구:통화전수) # 범위변수 선택

예3) select(delivery, -요일) # 변수 삭제

dplyr 패키지

unique() 함수

distinct() 함수 # 중복 내용 제거하기

예1) distinct(delivery, 업종)

mutate() 함수 # 기존 변수를 활용한 임시 변수 만들기.

예1) mutate(delivery, 새요일 = paste0(요일, "요일"))

변수 추가 방법2 예2) delivery\$새요일 = paste0(delivery\$요일, "요일")

count() 함수 # 그룹별 개수 세기 예) count(delivery, 사군구)

→ * tibble 형태

group-by() 함수 # 그룹 지정해주기 예) delivery_grp = group-by(delivery, 사군구)

summarise() 함수 # 원본 데이터 = 전체 요약 / 그룹 지정 데이터 = 그룹별 요약

예1) summarise(delivery, mean(통화전수), m = min(통화전수), M = max(통화전수))

예2) summarise(delivery_grp, mean(통화전수), m = min(통화전수), M = max(통화전수))

→ 그룹 기준 (tibble 형태)

*

전체 기준

예3) summarise(delivery_grp, length(통화전수))

→ 그룹별 → 통화전수의 크기 (tibble 형태)

Date.

No.

top-n() 함수 # 상위 n개치 뽑기

top-n(delivery, 5, 통화전)

top-n(delivery-grp, 5, 통화전)

파이프라인(%>%)을 활용한 연속작업 예1)

new_data = delivery %>%

* 결과 CSV 파일로 저장

데이터 저장

filter(업종 == "음식음료") %>%

write_csv(new_data, "result.csv", row.names = FALSE) group-by(시장구) %>%

summarise(mean_call = mean(통화전)) %>%

행 이름을 포함하지 않는다.

arrange(desc(통화전))

ungroup() 함수

delivery %>%

filter(업종 == "음식음료") %>%

group-by(시장구, 시장구) %>%

summarise(mean_call = mean(통화전)) %>%

ungroup() %>%

group-by(시장구) %>%

top-n(3, mean_call) %>%

arrange^N(desc(mean_call))_{7,}