**2023/05/17**

## 지금 해야 할 것  
  
1. API 열어서 안에 어떤 데이터가 있는지 세세하게 알자. → 데이터의 경향성을 파악하자.  
2. 결측치, 이상치 등 전처리!! 데이터 깔끔하게 다듬기. → 넣으면 데이터 뽑을 수 있도록. 완전 중요  
3. 데이터들 간에 어떤 상관관계가 있는지 분석하자 → 지난주에 했던 거 공유하기.  
4. 데이터를 어떻게 엮을 수 있을지 분석하자 → 지난주에 했던 거 공유하기.  
5. 머신러닝 공부 → 클러스터링 확정, 회귀와 분류 정확히  
6. 인사이트를 도출해야 하는데 이건 원래 전 오래 걸린다고 함. 이거에 초점을 맞춰서 해보기.  
7. 중요한 건 꺾이지 않는 마음……… ㅇㅈㅇㅈ

다음주까지 진행해볼 것 : 매출 데이터 분석하여 분석결과 정리하기.

[추정매출 API]

- 어떤 데이터인가  
- 분기별 데이터가 적재됨(2018~)  
- 칼럼 내용  
- 상권 코드와 상권 코드 명, 업종 코드와 업종 코드 명  
- 우리가 사용할 업종 10가지 - 한식 음식점, 중식 음식점, 분식전문점, 양식 음식점, 일식 음식점, 패스트푸드점, 커피-음료, 치킨전문점, 호프-간이주점, 제과점  
- 분기당 매출 금액, 주 중/주말 매출 금액, 월~일 매출 금액, 시간대(00~24) 매출 금액, 남성/여성 매출 금액, 연령대(10~60 이상) 매출 금액  
- 분기당 매출 건수, 주 중/주말 매출 건수, 월~일 매출 건수, 시간대(00~24) 매출 건수, 남성/여성 매출 건수, 연령대(10~60 이상) 매출 건수  
- 주 중/주말 매출 비율, 월~일 매출 비율, 시간대(00~24) 매출 비율, 남성/여성 매출 비율, 연령대(10~60 이상) 매출 비율  
- 추정 매출에 사용되는 데이터는 개인 / 법인 매출 모두 포함

요식업종 상권 분류하여 데이터 프레임 만들기  
  
- 요식업 전체 상권 - 1개. 기본적인 데이터 프레임  
  
- 업종별로 분류 - 10개  
- 업종별 총매출(제일 높은 업종, 제일 낮은 업종 - 순위 매기기)  
- 요일별 총매출(가장 높은 요일, 가장 낮은 요일)  
- 연령별 총매출(가장 높은 연령대, 가장 낮은 연령대)  
- 성별 총매출  
  
- 지역별로 분류 - 426개(행정동 코드)  
- 지역별로 상권이 각각 몇 개 있는지 알아보기(가장 많은 상권의 개수와 지역, 가장 적은 상권의 개수와 지역)  
- 지역별 총매출(상위 10개 지역, 하위 10개 지역)  
- 지역별 평균 매출(A동의 총매출을 그 지역에 있는 상권의 개수로 나누기)  
- 지역별 매출 중간값(A동의 상권들의 총매출 중 중간값)  
- 요일별 총매출(가장 높은 요일, 가장 낮은 요일)  
- 연령별 총매출(가장 높은 연령대, 가장 낮은 연령대)

- 성별 총매출  
- ex 사직동의 한식 음식점 고객은 여자가 더 많은지, 남자가 더 많은지

→ 상권을 동 단위로 묶은 데이터 생성. 상권의 지역별 분류  
  
→ ex. 사직동의 상권 매출을 묶었을 때 동 내의 상권들이 같은 경향성을 띠는지 보기.  
  
-매출 상위 다섯 개 지역, 하위 다섯 개 지역 도출(분기별)