분석의 완성은 스토리텔링. 통계자료 그럴듯 하다. 납득이 되느냐는 역량.

* 군산사랑상품권 데이터
  + 판매, 사용
  + 사용 품목, 액수
  + 시기(기간)
* 지역, 연령대 사용 촉발
* 비지도학습
* 목표
  + **머신러닝** 기법을 써보기 (하이퍼파라미터 튜닝)
    - 지도, 비지도 학습
  + 실질적인 데이터 다뤄보기 -> 공모전
  + 비즈니스적인 것 > 공공적인 것
* 데이터
  + 군산 / 목포 / 통영
  + 경기도
* 경기도 지역 화폐 데이터로 다른 걸 해보는 것(서로 다른 시군 비교. 차이점)
* 야구 예측 데이터, 수산물 데이터
* 군산 / 목포 / 통영 -> 다른 두 도시의 것을 군산에 적용시키기
  + **수혜군 선정 (소상공인 범위 내 우선 순위 선정) -> 지역, 인구 특성 제안**
  + 데이터? -> 세 도시를 비교
    - 온라인, 택배 : 물품의 수요 *-> 소비? 상권 부족(온라인 주문)*
    - 부동산 : 지역, 신시장, 인구 이동(내외부), 투자
    - 대출 : *기업 대출-투자*,   
       개인 대출(*투자목적?*) -> 인구통계, *지역선별* 취약 계층 선정  
      + *취약 계층 종사 분야 - 정책 제안*
    - 편의점 : 품목, 지역별 매출 -> 소비 진작
    - 음식점 : 업종 카테고리 -> 검색 유입 도시
    - 기타 : 키워드
  + 데이터프레임\_1안 : 예측 모델 - 소상공인
    - Target (Label) : 유동인구(유입).
    - Column (feature) : 날씨\_기존연구,
    - Row :
  + 데이터프레임\_2안 : 비지도 학습 EDA => 군집별 특징 참고\_*군집분석*
    - Target (Label) :
    - Column (feature) :
    - Row :
  + 데이터프레임\_3안 : 분류 모델\_feature importance
    - 목적 : F에서 차이가 나니 F에 대해 시별 비교 -> 강화, 대책 마련
    - Target (Label) : 군산 / 목포 / 통영
    - Column (feature) : 카테고리별매출, 부동산, 신용등급분포, 음식점분포
    - Row **:**
  + 분류 모델 :
  + 예측 모델 : \*소비데이터 -> 유동인구, 매출. 품목 선정
  + *시계열* X (현 시점 - 겨울)
    - 추가 데이터 조사
* 경기도 ->
  + 예측 모델 : 지역화폐 사용
  + 비지도 : 지역의 구분 -> ?

경기도 시, 군(서울, 거대도시)

군산 28만 - 여주, 양평?

연령비

클러스터링

유동인구 : 교통량 - 톨게이트. 국도.

유사. 근거?

* 데이터 join : 우창님
* 기상 데이터, 코로나 데이터
* 1) 음식물 쓰레기
* 2) 유동인구 & 3) 거주인구
* 4) 음식 소비



2시 50분

3시 30분

**전처리 EDA**

데이터 축소(기간, 행정구역) - ex. 2년치, 제주시

파생변수

**모델별 차이점 (특징, 장단점, 성능 비교)**

Regression Model 적용

**랜덤 포레스트**

GBM

**lightGBM** - GPU모델 (코랩, AWS)

+ DNN

**비즈니스 인사이트** 도출 : 나무 계열 feature\_importance 기반 제시

(특징별 회귀분석)

- 쓰레기: 지불금액 제외 배출건수, 배출량(g)

- 유동인구: 성별, 연령별, 내외국인, 방문객, 외국인-장단기

- 거주인구:

- 음식 소비 데이터: 음식점 제외 마트/슈퍼마켓, 식품, 농축수산물, 배달 업종으로 한정

PMO

**어려움 해결**

스스로 고민. 조원들과 상의.



목요일 오후 리허설 사전 요청