**2020.07.30 회의록**

**4일(화): 데이터분석**

**목표: 계획 양식 채우기**

**12시~4시: 자료 분석 + 역할 분담**

**변수 정리 + 그래프 몇 개 (목요일까지 업로드)**

**크롤링 코드 공부 + 시계열 공부해오기(자료 만들기: 손지우)**

**Github 만들기**

**6일(목): 딥러닝 PyTorch + 데이터분석**

**예습발표: 백채빈**

**과제발표: 김민회**

**8일(토): 발표**

**발표자: 미정**

**## 데이터분석**

**<크롤링할 데이터>**

날씨(습도, 온도, 체감온도…)

검색트렌드(연령별)

경쟁채널 시청률(vs 타 홈쇼핑)

경쟁채널 시청률(vs 예능)

출연진 성별, 인원 수

\*브랜드

\*시연 여부

유명 연예인 출연 여부

\*‘매진 임박’

쇼호스트 인지도(영향력), 빈출도

\*지역데이터: 지역별 시청률, 지역별 백화점까지의 거리

**\*조사할 내용(도메인 놀리지)**

- 홈쇼핑 이용연령(통계청, 20대연구소, 엠브레인…)

- 브랜드 인지도

**\*엑셀 데이터 전처리방식**

- 시청률을 어떻게 처리할지!!!

- 정규표현식으로…

- 남성 vs. 여성

- 분당 -> 적당히 뭉치기

- 시간 분류가 가장 중요!!!

- 요일, 공휴일 분류

- 반복 등장 횟수

- 충동적 구매 vs. 합리적 구매

- 금액대별, 상품군별