데이터로 시작하는 작심삼일
(FEAT. 구글 데이터 스튜디오)

2017 데이터야놀자

발표자: 맹윤호

발표에 앞서…

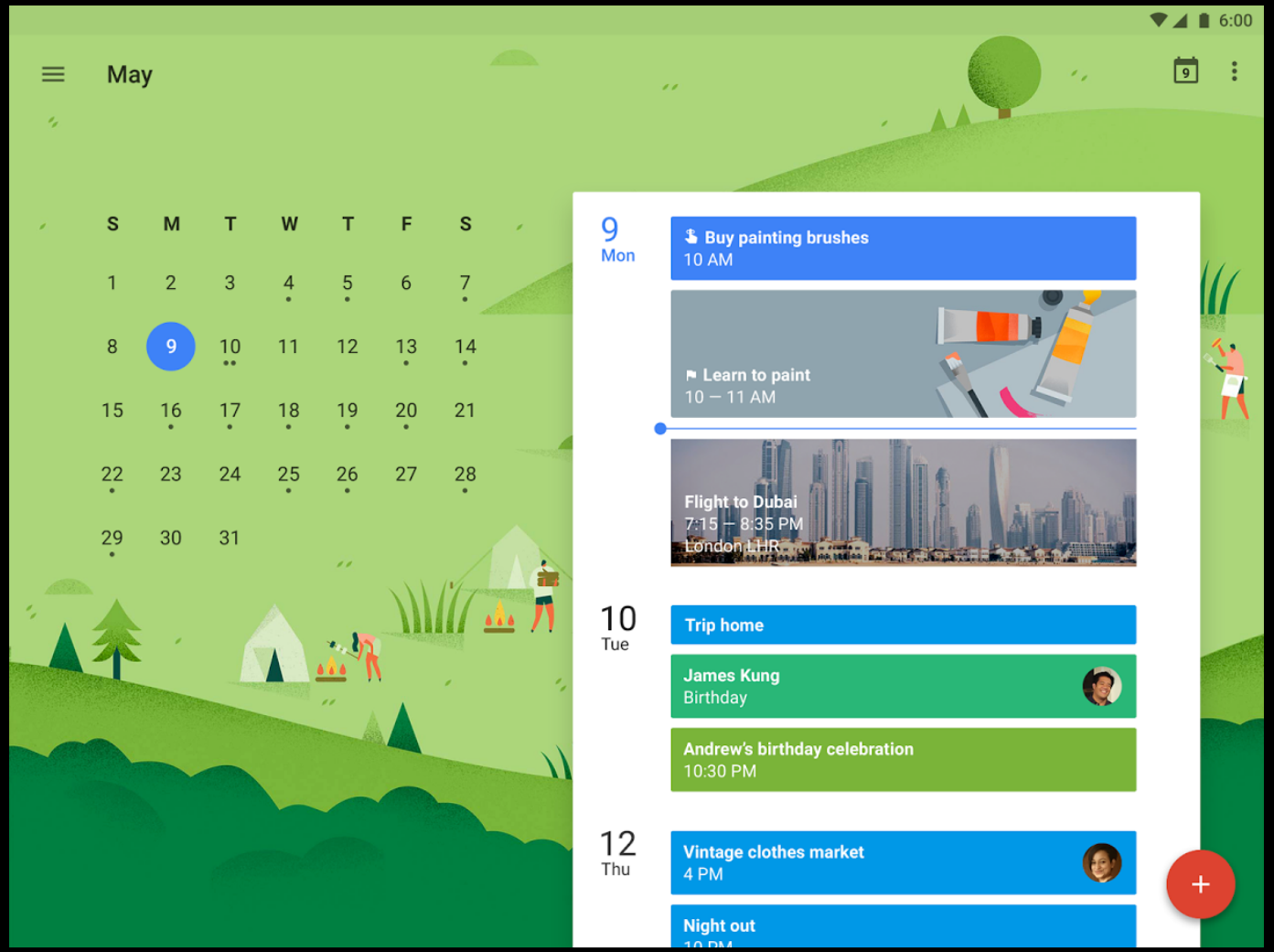
- 본 발표는 개인적인 데이터 분석 사례를 기반으로 비영리적 목적으로 진행됨을 밝힙니다.
- 본 행사로 인한 모금된 금액(티켓, 후원사 등)은 발표자와 전혀 관계가 없음을 밝힙니다.
- 본 발표는 IBM Watson과 아무런 연관이 없음을 사전에 밝힙니다.

“THOSE WHO DO NOT LEARN FROM HISTORY ARE DOOMED TO REPEAT IT.” BY GEORGE SANTAYANA

“과거를 기억하지 못하면 과거를 반복하는 운명에 처한다”

하루 24시간중 놓치는 시간을 찾아봅시다.

- 구글캘린더에 앞으로의 계획 뿐만 아니라 이미 내가 했던 일을 빠짐 없이 적어서 내가 하루에 시간을 어디에 얼마나 썼는지를 분석해봤습니다.

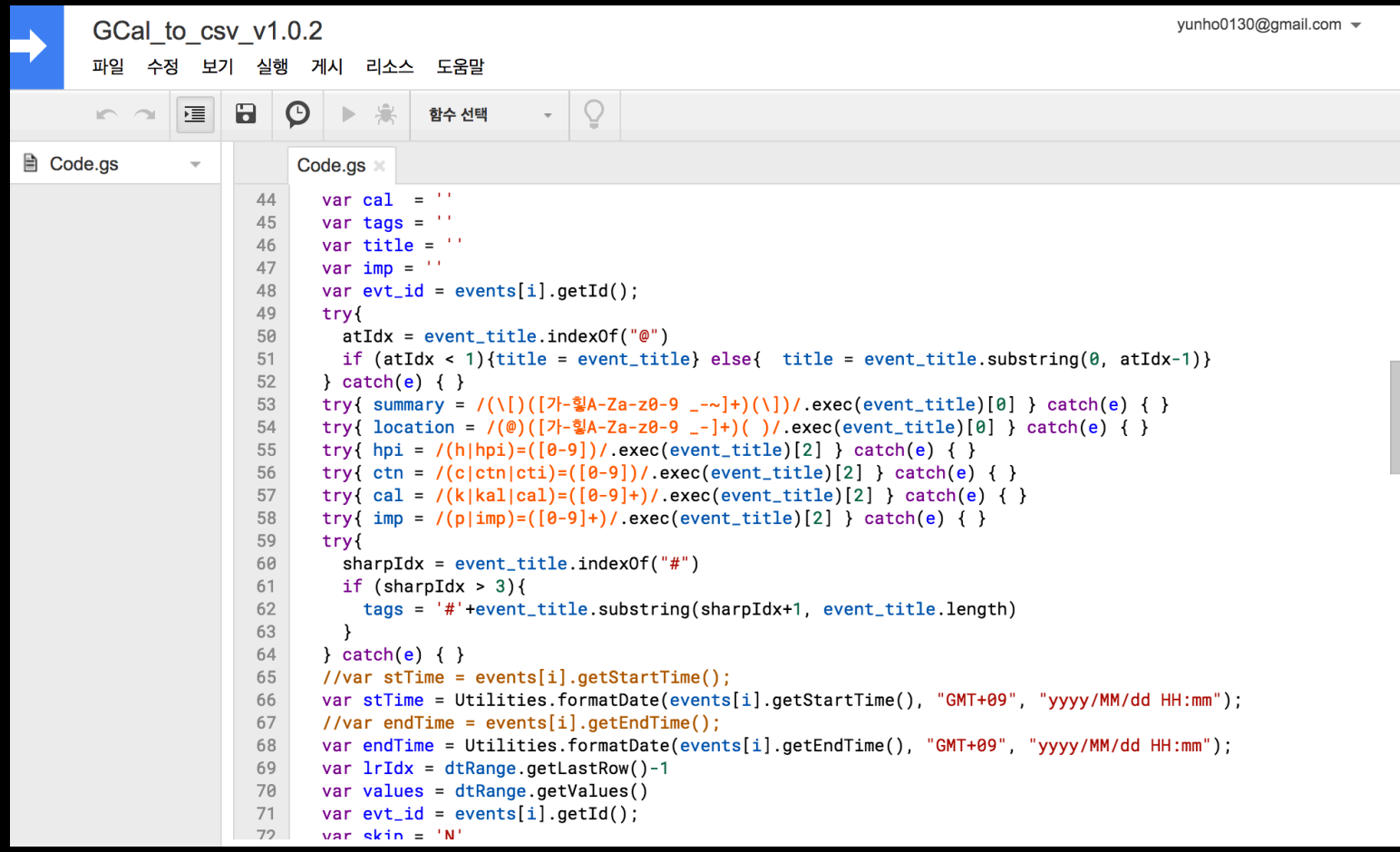


하루 24시간중 놓치는 시간을 찾아봅시다.

- 구글 캘린더 -> Google App Script -> Google Data Studio로 하루 결산을 해봤습니다.

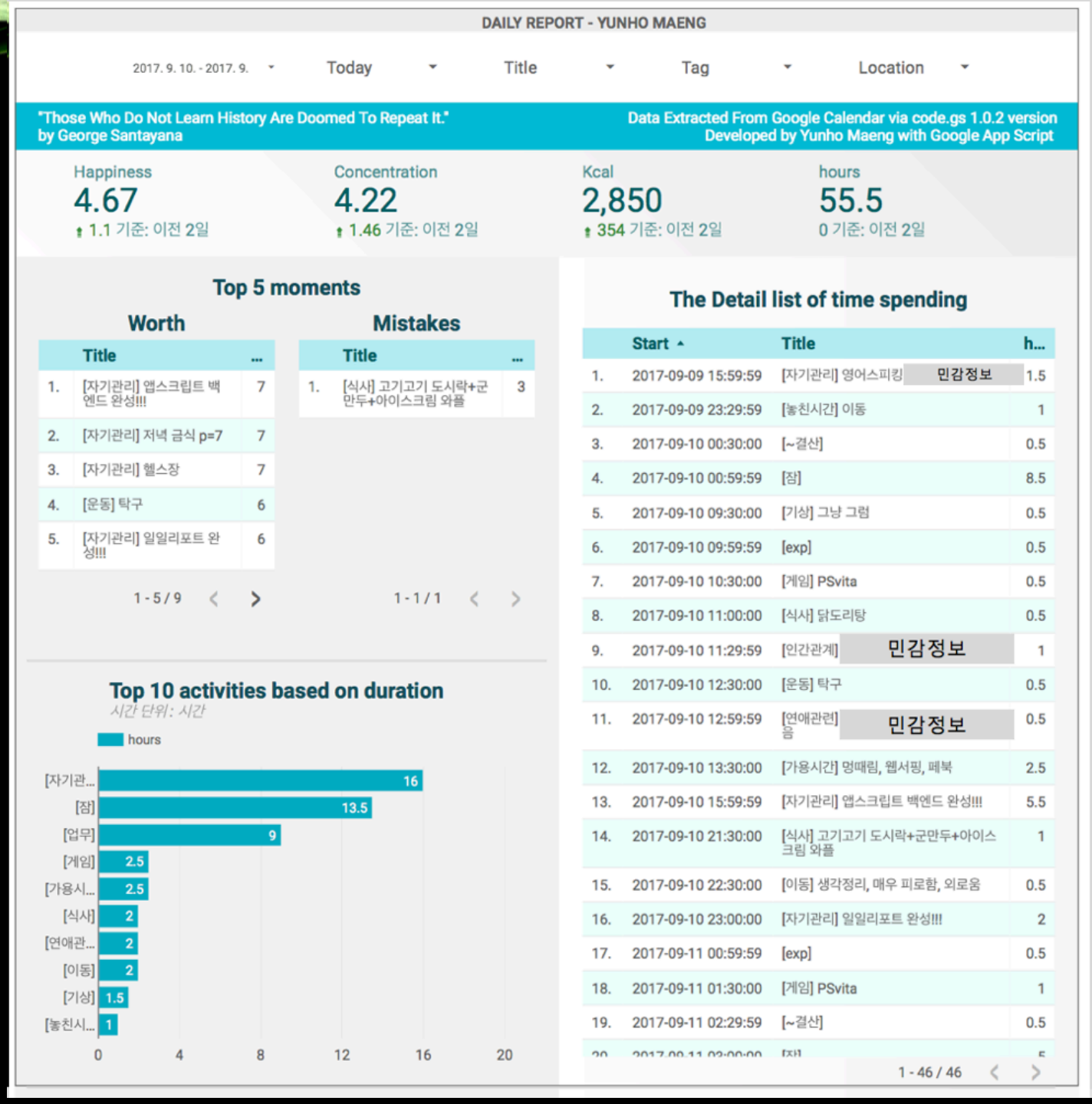
GOOGLE APP SCRIPT

- Script
- https://script.google.com/macros/d/McyjMBGa2laFDneFFH-A1eTJHHgK9ED41/edit?uiv=2&mid=ACjPJvG9v5mI1luI5nTTcQeUwnclDyHjzOUkp_1aJoUCbx-4sD7UW5P_fKsDGHAwtdoJzacL5vxJKcs87ExQsUtugru9MqdYh-uBmThzSj8DIxx6OnjiKsmTUvEN6sBwmcpF9NbLIkeMgA
- API Reference
- https://developers.google.com/apps-script/reference/calendar/



결과물

- Google Data Studio
- https://datastudio.google.com/u/0/reporting/0B0Q-zebPSeDpVDh5R2Q2QXpBUmc/page/1M
- 자동으로 구글 캘린더에 내가 '잘' 입력한 일정들을 가져옴 (1시간 마다)
- 가져온 일정들을 오른쪽의 보고서로 만들어 줌
- 모든 솔루션은 개인이 무료로 사용가능한 것으로 구성함.



야크 털 깎기(YAK SHAVING)*

- 데이터야놀자의 발표 세션은 단순 기술보단 '놀이'이에 초점
- 세스 고딘(Seth Godin)의 "이제는 작은 것이 큰 것이다"에서 유래
- 아래와 같은 상황
  - 봄이 왔으니 세차를 해야 겠네.이런 호스가 겨우내 터졌네. 홈 디포우에서 새거를 하나 사야겠군.홈 디포우는 Tappan Zee 다리를 건너야 하고 그러려면 EzPass(하이 패스 같은 건가 봅니다.) 가 있어야 하는데.음 옆 집 Bob 한테 EZPass 를 빌려야 겠군.그런데 아들 넘이 옆 집 Bob한테서 보이스카웃에 가려고 베개를 빌려갔었는데.....Bob은 빌려간 베개를 돌려주기 전까지는 그의 EZPass 를 빌려주지 않을 거야...음 그런데 베개에 있는 야크 털이 많이 빠져서 지금 당장 돌려줄수가 없겠네..베개에 넣을 야크 털을 구해서 베개를 고쳐야 겠네..그래서 세차를 하기 위해 동물원에 가서 야크 털을 깍는중..

*이광섭(2016) 발췌, https://www.lesstif.com/pages/viewpage.action?pageId=29590364

야크 털 깎기(YAK SHAVING)*

- 의미있는 야크 털 깎기도 존재.
- 과연 이 미니 프로젝트가 개인적인 의미를 넘어서서 일반 사람들에게도 다가갈 수 있을지...?

> ⚠ 모든 "야크 털깍기"가 의미가 없는 것은 아니며 가장 유명한 성공 사례로는 도널드 크누스(Donald E. Knuth) 교수의 TeX 을 들수 있습니다.
>
> *"The Art of Computer Programming"* 을 저술중에 조판 시스템과 폰트가 맘에 안 들어서 메타 폰트와 조판 시스템을 만드는데 몇 년을 사용했고 그 결과물로 TeX 이라는 역사상 버그가 가장 적다는 문서 작성 도구를 만들어 냈고 이를 사용하고 저술을 마무리했습니다.
>
> 하지만 "야크 털깍기"를 한 사람이 크누스 교수라서 가능했겠죠....

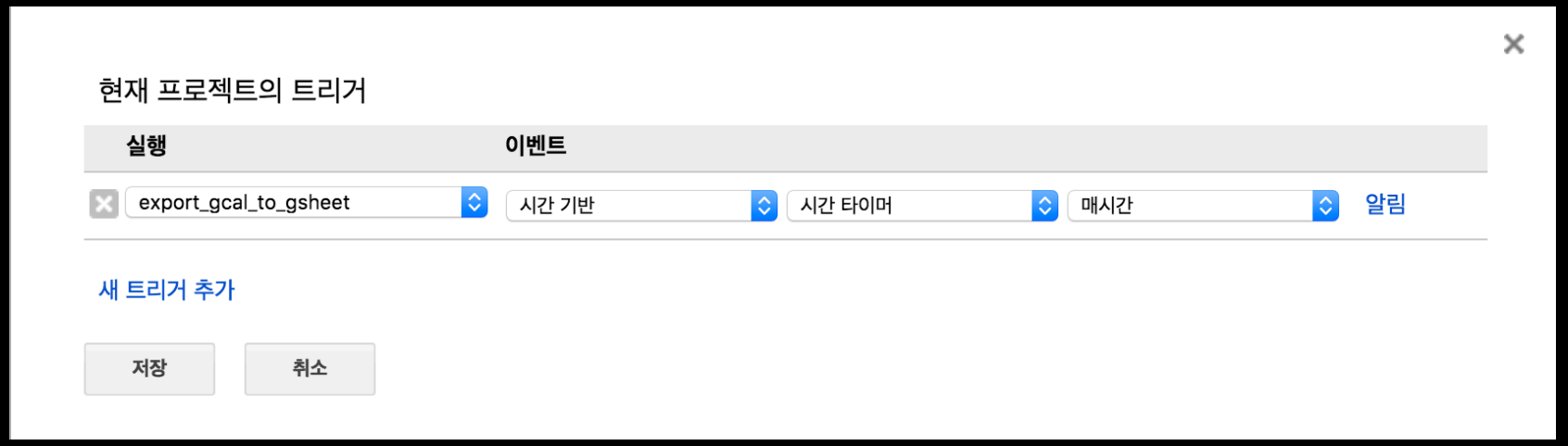
*이광섭(2016) 발췌, https://www.lesstif.com/pages/viewpage.action?pageId=29590364

데이터 IMPORT

- 체중 감량을 위해 칼로리를 적고 싶다. / 해당 활동의 소모 시간 뿐만 아니라 집중력도 추적하자! / 기왕 이렇게 된거 행복도도 넣자! / 스스로 기념하고 싶은 일과 실수 했던 일을 구분하고 싶다.
  - String으로 부터 key:value의 쌍 얻기.
  - kcal = 700와 함께 k=700도 하고 싶다. (단축 키워드)
  - 정규표현식 사용하여 String Parsing
- 태그를 넣어서 나중에 검색어로 검색을 하고 싶다.
- Demo: key:value 쌍 입력과 태그, 그리고 데이터 import

BATCH SCRIPT

- 매번 스트립트를 실행 시키는 것이 귀찮다!
 - 배치를 돌려보자. 1시간에 한 번씩!
 -- 배치를 돌려보니, 중복으로 데이터를 가져온다! 중복 이벤트 제거
 -- 이벤트 id를 통한 중복제거가 성공했으나, 수정된 일정은 추적을 못한다 왜?
 --- 이벤트 modified date 정보를 이용해서 추적을 해보자!
 - 한국 시간대로 값이 안나온다. GMT 09로 시간대 통일하기



결산 시점 설정

- '하루' 결산인데 새벽을 넘어가면 어떻게 해야 하나?
- 그 전날에 결산을 안하면 어떻게 하지? (결산 기간 설정)
  -> 특정 키워드를 통해 결산 시점 생성하기

스크립트 실행 최적화

- 스크립트의 실행이 느리다.

- 실행이 느려지는 부분을 로그를 통해 파악하기

- 해당 부분 스크립트 개선

- 데이터를 특정 시점까지만 가져오기

- 배치를 거의 실시간으로 돌리기 위해서는 선결과제.

- [실행 트랜스크립트]에서 병목현상이 어디에서 일어나는지 확인하기

리포트 만들기

- 구글 데이터 스튜디오로 대시보드 리포트 만들기
- 구글 스프레트 시트와 데이터 스튜디오의 연결
- 필터를 추가해서 특정 기간의 데이터에 대한 정보를 불러오기
- 그룹과 시간 조절
- 기본 필터와 연동 필터

어려웠던 점

- 솔직하게 적어야 하는 [요약]
  - 놀았을 때 정말 놀았다고 적을 수 있는가
  - 직접적으로 적기 어려운 요약

- [요약] 카테고리를 정하기 어려운 활동들이 있음.
  - 저녁을 먹으면서 친구를 만났으면, [식사]로 구분해야 할까 [인간관계]로 구분해야 할까?
  - 소개팅은 [연애]로 구분할 수도 없고 [썸]으로 구분할 수도 없다. 어떻게 해야 하나?
  - 이동 중에 책을 읽었으면, 어떻게 구분해야 할까?

향후 업데이트 예정

- 피로도, 지출금액을 추가 key:value 쌍으로 추가할 것을 고려 중
- Key:value 쌍을 입력하지 않더라도 기본값으로 처리할 것을 고려 중
- 장기 프로젝트로 꾸준히 이어나가서 다중회귀분석 등 통계나 머신러닝으로 스스로의 행동 변화에 영향을 미치는 요인에 대해 알아보고자 함
- key:value를 App Script가 아니라 캘린더에서 입력 받아 추가하기
  - 현재는 App Script에 직접 추가해야 하지만, 캘린더에서 key:value 쌍을 입력하면 자동으로 행을 늘려서 확장하는 방식으로의 고려
- API 혹은 서비스로 공개

후 기

- 실질적으로 가장 효과가 있었던 건, 기념하고 싶은 일의 기록이 가장 큰 일상에의 impact가 있었음
  - 하루가 끝나기 전에 하나도 기념하고 싶은 일이 없었다면, 독서를 단 10분이라도 하려고 하는 경향이 발견되었음.
- 파일로만 결산을 끝내는 것보다, 결산한 내용을 파일로 출력하고 그 이후에 후기를 적는 편이 훨씬 더 결산 완료율이 높았음.
- 칼로리 가계부를 적자 기존에 음식들에 대한 고정관념이 보다 실질적으로 바뀌었음.

Q & A

맹윤호 yunho0130@gmail.com