


데이터분석 프로젝트

 by neul ha

팀프로젝트의 목표

- 배운 기술을 적극 활용할 수 있는가?
 - 시각화
 - 크롤링
 - 머신러닝
 - log data
- 가성비가 있는가?
- 포트폴리오로 썼을 때 기업에서 솔깃할 만 한 주제인가?
 - 기업이 알고는 있는데 안해본거. 데이터는 있지만 결과물이 없는 경우



주제 선정

논점이 무엇인가?

구조화 이전에 논점을 정해보자. 관찰한 사실이 아닌 논점

예를 들어 회사에 도둑이 들었다. ← 현상

논점 1. 방범체계가 제대로 맞춰지지 않음

논점 2. 보고체계가 미비하다.

논점 3. 회사 이미지에 손상이 간다.

문제 해결에 대한 논점은?

예를 들어 식당이 맛이 없다.

논점 1. 가격에 비해 맛이 없다 -> 다른 곳의 가격은 ~한데, ~~

논점 2. 찾아가기가 불편하다. → 자동차를 이용안하면 가기 힘들고 불편한 장소에 있으면서 주차장은 없다.

매출을 늘리기 위해 가격을 낮춘다. 광고를 한다. 판매촉진 대책 마련을 한다. 영업에 신경을 쓴다. 이런 대증요법은 일시적임.

진짜 원인, 즉 논점은 다른 곳에 있다.

상품 그자체 혹은 채널 선택, 프로모션 전략

데이터 분석 방법론

1. 흑백을 가릴 수 있는 부분부터

- 일시적인가 장기적인가?
- 부분인가 전체인가
- 분해해서 처음부터 바라본다. 우리만 그러는건가? 과연

2. 스무고개를 해보자 조원들끼리


2.논점이라고 여겨지는 것을 평면에 쪽 늘어놓는 형식으로 적는다.

대논점 - 중논점 - 소논점

논점 A - 논점a,b,c - 논점 x(잠재고객 발견),y(고객의 욕구 파악),z(고객에의 접근)

분석 경진대회를 겸해서 프로젝트를 구성해보자.

이전 프로젝트 결과물로 공모전 우수상 획득



서울특별시

2023 서울특별시 빅데이터캠퍼스 공모전

총 10개팀 900만원 상당의 상금 및 서울특별시장상

- 서울시 관련 정책이 무엇인가?
- 서울시에서 미는 사업이 무엇인가?
- 있다면 그것에서 어떤 논점을 만들 수 있을까?

»»

제가 개인적으로 생각한건...다문화라는 키워드입니다.

전세계적으로 노동인구부족, 저출산등 인구절벽에 대응방안으로 '이민자'라는 대안이 떠오르고 있습니다.

물론 논쟁의 여지도 많고 우리나라 정서에 맞지 않지만...지금과 같은 상황이 지속되면 대안으로 생각해볼 수 있는 지점이라고 생각합니다.

이민자 증대에 따른 다문화가정의 증대. 그에 따른 시설이나 복지가 고려해야 하고 그에 맞는 사업들도 진행되어야 한다고 생각합니다.(고민해볼 부분)

관련데이터는 많이 존재한다. 다문화에 관한 그에 따른 솔루션도 해외에서는 많은 것으로 안다. 국내에 도입할 시 점이고 데이터로 해결이 가능하지 않을까? 생각을 해봅니다.

통계



인구/가구

활용사례(갤러리) 등록

URL 복사

목록 이동

서울시 다문화가구 현황 통계

○ 통계개요

- * 통계명 : 다문화가구 현황
- * 통계종류 : 서울시 다문화가구 현황을 제공하는 일반 · 보고통계
- * 작성목적 : 국내 거주 외국인주민에 관한 기본정책 수립, 시행 및 제도개선을 위한

[전체 설명보기](#)

핵심은!

6주나 쓰는데.. 시간갈아서 ㅠ

누구에게나 깊은 인상을 줄 수 있는 데이터 분석 프로젝트가 될 수 있는지?

면접관은 데이터 과학 프로젝트에서 정확히 무엇을 찾고 왜 그런 프로젝트를 찾는가?를 이해해보면 좋을 것 같다.

고민해본 여러 방법론입니다.

1. 실제 데이터로 해볼 수 있을까?

실제 데이터로 작업한다는 것은 특히 실시간으로 업데이트되는 데이터를 의미

사용자가 생성하는 실제 데이터로 작업하고 실시간으로 생성되는 데이터로 작업하면 면접관에게 관련성 있고

시기 적절한 데이터로 작업하는 방법을 알고 있다는 것을 증명하는 데 도움이 됨

(개인적으로 엄청 중요해봄 내가 모은 로그 데이터에 대한 갈망이 있음...설문조사라도)

```
{ "user_id": "5d468b47-6ec6-4eb3-8704-e3a96c8a504b"
, "property": { "gender": "male",
               "age": 42,
               "city": "seoul" }
, "session_id": "HZq-lYjARqWEXvRudFrKpg"
, "date": 2022-02-22 13:25:00
, "duration": 3000
, "events": [ { "name": "buy_store",
               "date": 2022-02-22 13:25:06,
               "param": { "membership": "off",
                          "use_point": 3000,
                          "use_cart": "on"
                        }
             }
            ]
}
```

- 로그 데이터를 수집해야한다. api에서 수집하여 가공 전처리... 이능력이 크게 보여질 듯. 깨끗한 정형데이터 다루는 것보다.
- 분석의 경우도 NLP도 개인적으로 해보면 좋을 것 같음. 한명이 책임지고? 그에 따른 방법은...?


1. API를 활용한다...

- API에서 데이터를 수집하고 데이터를 정리한 후에 데이터베이스에 저장한다.

→ API가 헛갈리면 이거 <https://ablearn.kr/newsletter/?idx=14351078&bmode=view>

E.g . 예시들

1.



coffee4m

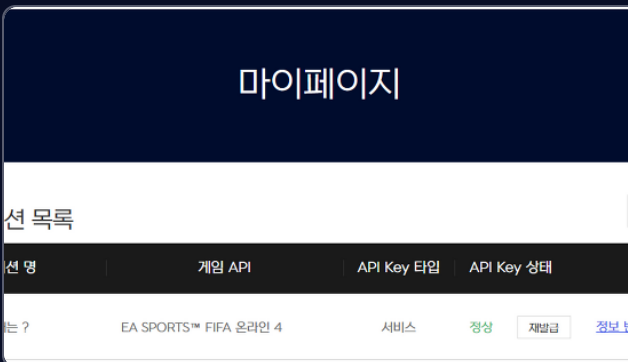
API와 파이썬 데이터 분석 - coffee4m

지난 포스팅에서는 5일치 비트코인 가격을 리스트로 만들어서 시계열 데이터 그래프를 그렸습니다. 암호화폐 거래소에서 가격을 보고 리스트를 직...

(암호화폐 데이터 분석)

2. 공공 API를 이용한 데이터 수집

3.




TISTORY

피파온라인4 API 활용 데이터 분석 프로젝트

피파온라인 시즌 1 때부터 시즌 4를 모두 플레이한 유저로서의 경험과 데이터 분석가의 꿈을 가진 현재, 피파온라인4 API를 활용한 데이터 분석...

(피파온라인4) - 시간대에 따른 공식경기 매칭 수와 승률 분석

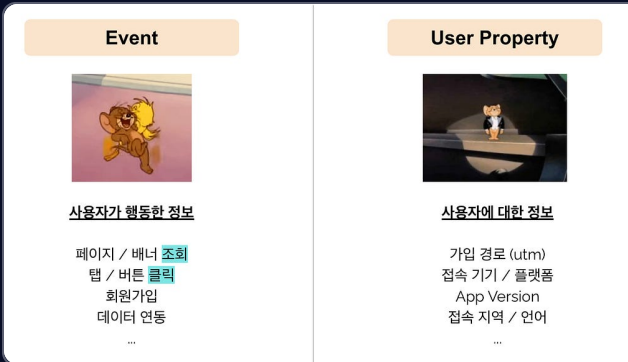


TISTORY

[디스이즈] 앱 로그 데이터 분석 프로젝트 [시작]

1. 디스이즈란? 디스이즈는 동아대학교 컴퓨터공학과 개발팀이 만든 '동아대학교 스마트 캠퍼스' 애플리케이션입니다. 학사일정, 학사공지, 교내식...

4.



Medium

사용자 행동 로그, 파종부터 수확까지 | 잘 쌓기—1편

'사용자 중심'이 당연한 세상입니다. 그런데 우리는 사용자에게 대해 잘 알고 있나요? 사용자를 낱알이 알기 위해 가장 객관적이고 대표성이 높은 방법...

(로그분석 방법론)

5.<https://brunch.co.kr/@leoyang99/15> (로그분석 방법론)

6.<https://sowhatmylifeismine.tistory.com/182> (본인 블로그 로그 분석)

7.<https://github.com/daje0601/Behavioral-pattern-analysis>

(핸드폰 로그 데이터를 활용하여 사용자 행동 패턴 분석 & 앱 사용자 패턴 분석)

- 1.사용자가 해당 앱에 몇분 머무는지를 예측하는 모델 생성
- 2. 다음 사용할 장르가 무엇인지 예측하는 모델 생성

2. 클라우드(구글, 아마존)

3. 다음으로. 배포 및 피드백

흥미로운 프로젝트 소개(브레인스토밍)

식습관 스몰데이터 분석을 통한 장트러블 극복기


배변 만족도, 배변 습관에 영향을 주는 음식이 개인마다 다를 것이라든 생각에서 시작
배변에 영향을 주는 음식을 알아보고 장트러블을 극복하기 위한 채변 프로젝트 시작



SlideShare

식습관 스몰데이터 분석을 통한 장트러블 극복기

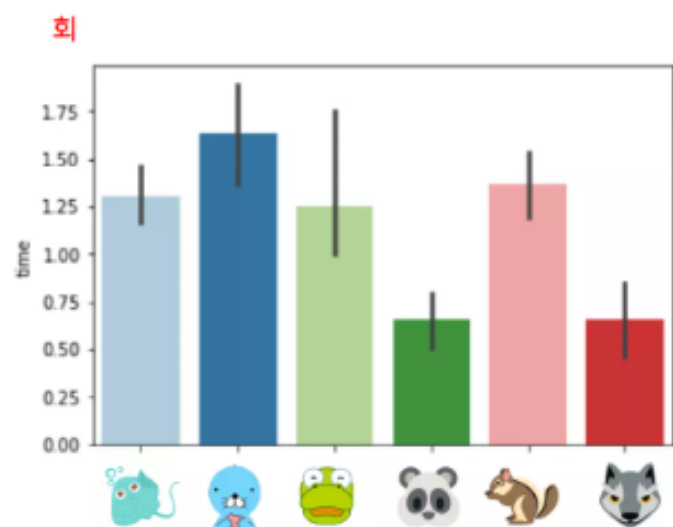
식습관 스몰 데이터 분석을 통한 장트러블 극복기 팀 푸푸 고은솔
<https://github.com/koeunsol> [김민주 https://github.com/kim4pb](https://github.com/kim4pb) ...



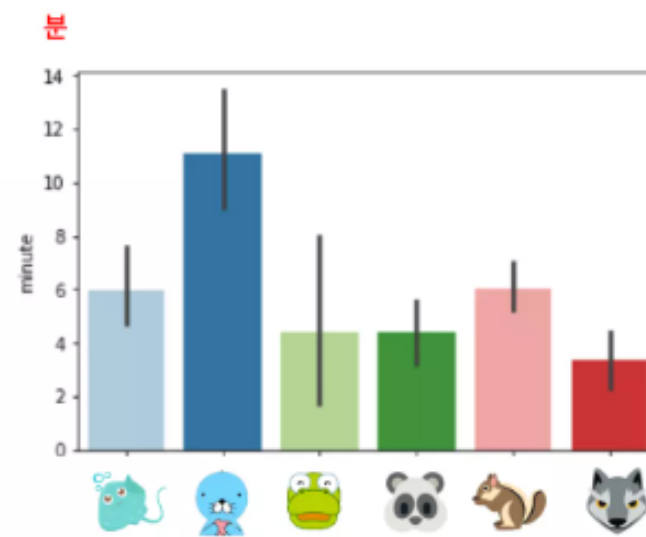
식습관 스몰 데이터 분석을 통한
장트러블 극복기

팀 푸푸

고은솔 <https://github.com/koeunsol>
김민주 <https://github.com/kim4pb>
배예솔 <https://github.com/yeseul0806>
송혜정 <https://github.com/hyeyeon11>
이신영 <https://github.com/godet1>
정보경 <https://github.com/aiceb221>



일 평균 화장실 간 횟수



일 평균 화장실에서 보낸 시간