

Twit

OpenAPI

해시태그

분석





Twit

팀원

소개

심다훈

박휘찬

이성진

홍성희



1조

TWIT

TWIT

01

02

03

04

주제 선정 배경

프로젝트 방향

프로젝트 목표

트위터 선정 이유

데이터 설명

활용 데이터 설명

파이프라인

프로젝트 활용 툴

데이터 수집

데이터 전처리

시각화

결과

시각화

시연

느낀점

01

주제 선정 배경

Twit Twit

그래서 목적이 무엇인가?

배경

django

Pandas



데이터



python

matplotlib

파이프
라인



SQL



결과

Twit Twit

데이터 엔지니어링

배경

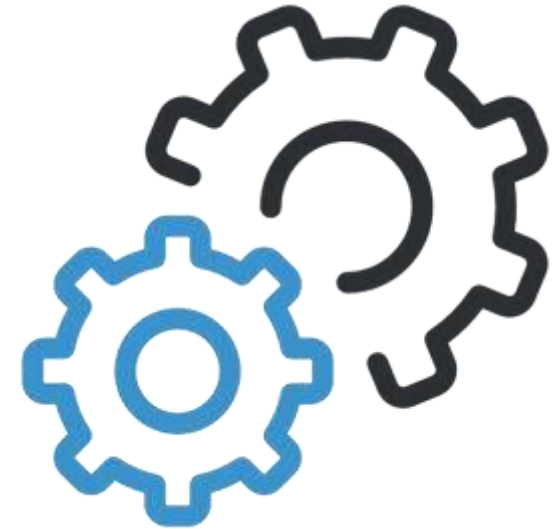
데이터

파이프
라인

결과



인사이트 도출



여러가지 툴의
사용
다양한 시각화

배경

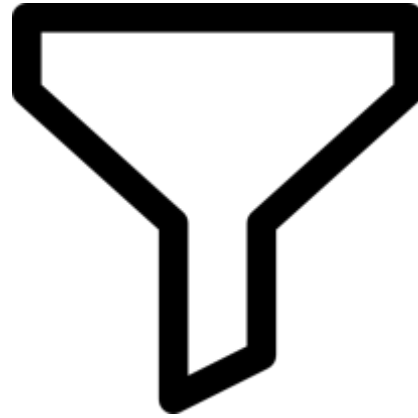
데이터

파이프
라인

결과



데이터 수집 및 저장



데이터 전처리



시각화

Twit Twit

배경

데이터

파이프
라인

결과

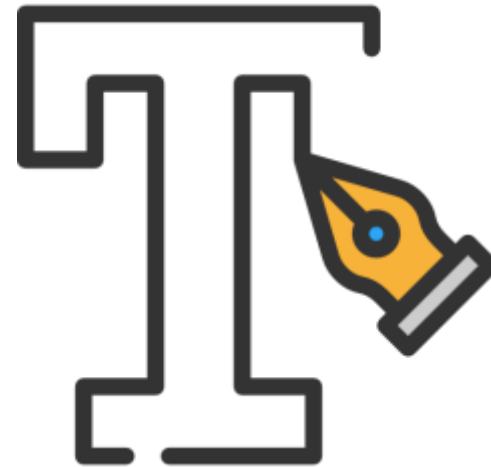
Why



?



API
제공



Text 위주의
Data(가공이 쉬움)



실시간
Trending
제공

02

테이터 설명

Twit Twit

배경

데이터

파이프
라인

결과

재밋고 흥미로운 주제로 해시태그들을 탐색

사용해시태그

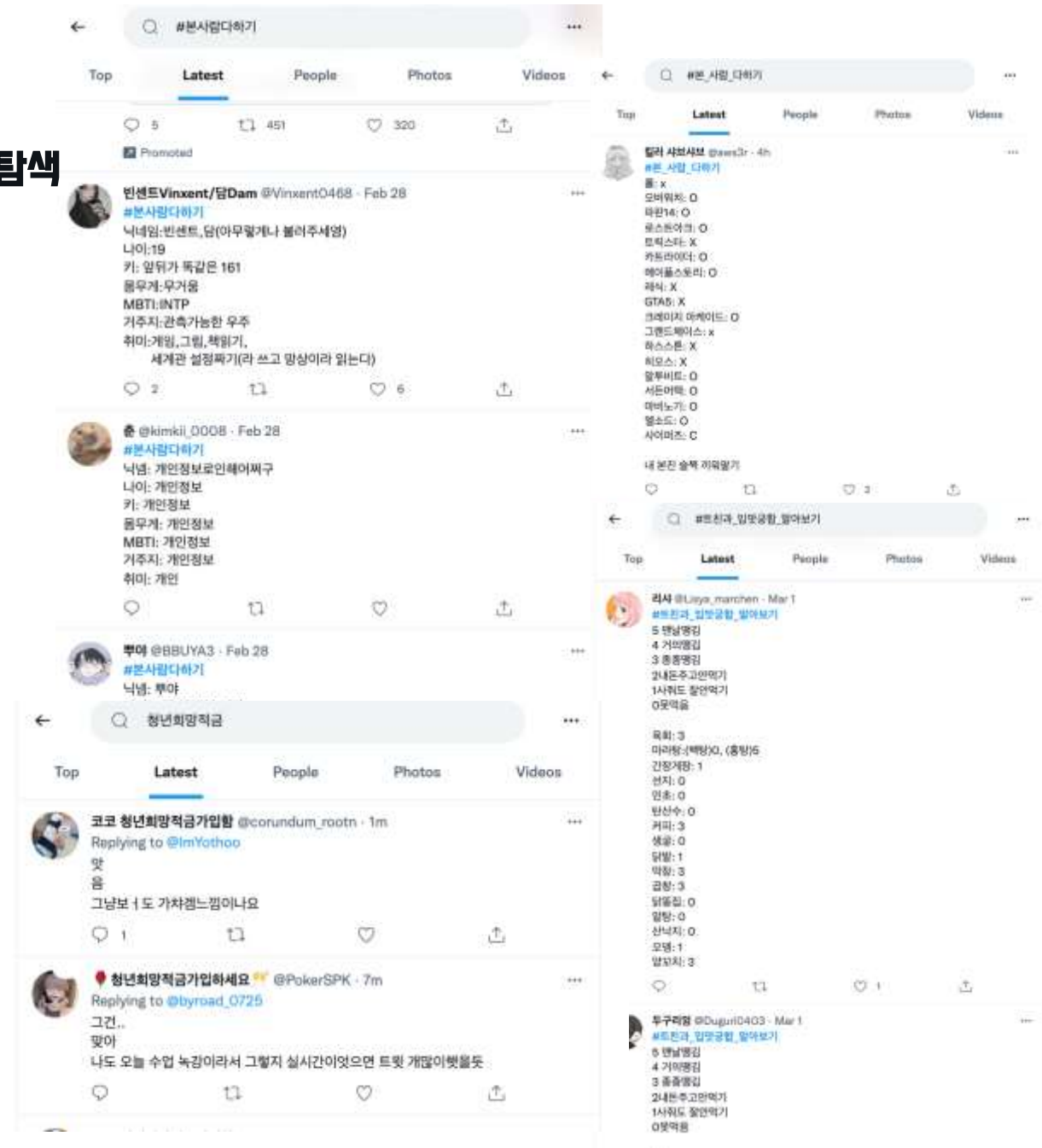
1) #본사람다하기

- 자기 소개하는 트윗
- 나이, MBTI, 거주지, 키 사용
- 해봤던 게임 올리는 트윗

1) #트친과_입맛궁합_알아보기

- 호불호 갈리는 음식으로
0-5까지 점수 매겨진 트윗

1) 청년희망적금



Twit Twit

배경

데이터

파이프
라인

결과



트윗 데이터
(JSON)

```
{
  "data":
  [{ "id": "1263150595717730305",
    "author_id": "2244994945",
    "text": "Do you 🐼our new
    Tweet settings?\n\nWe want to know
    how and why you'd use a feature like
    this in the API. Get the details and let
    us know what you
    think👉\nhttps://t.co/RtMhhfAcIB
    https://t.co/8wxeZ9fJER",
    "public_metrics": {
      "retweet_count": 12,
      "reply_count": 13,
      "like_count": 51,
      "quote_count": 7
    }
  }
],
  "includes": {
    "users":
    [{ "id": "2244994945",
      "username": "TwitterDev",
      "description": "The voice of the
      #TwitterDev team and your official source
      for updates, news, and events, related to
      the #TwitterAPI.",
      "name": "Twitter Dev"
    }
  ]
}
```

03

파이프 라인

배경

데이터

파이프
라인

결과

데이터 수집



데이터 전처리

Pandas



데이터 저장



시각화



협업



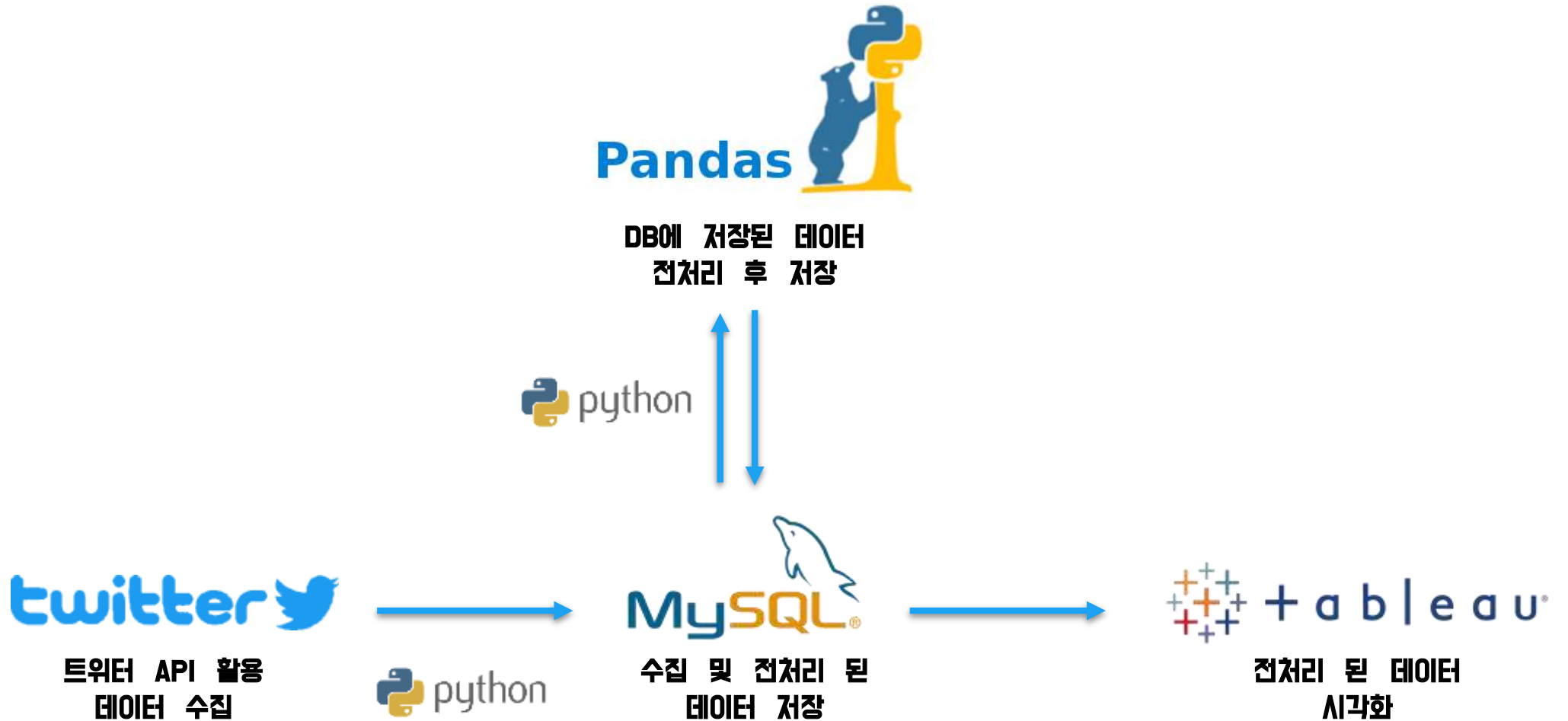
프로젝트 파이프라인

배경

데이터

파이프
라인

결과

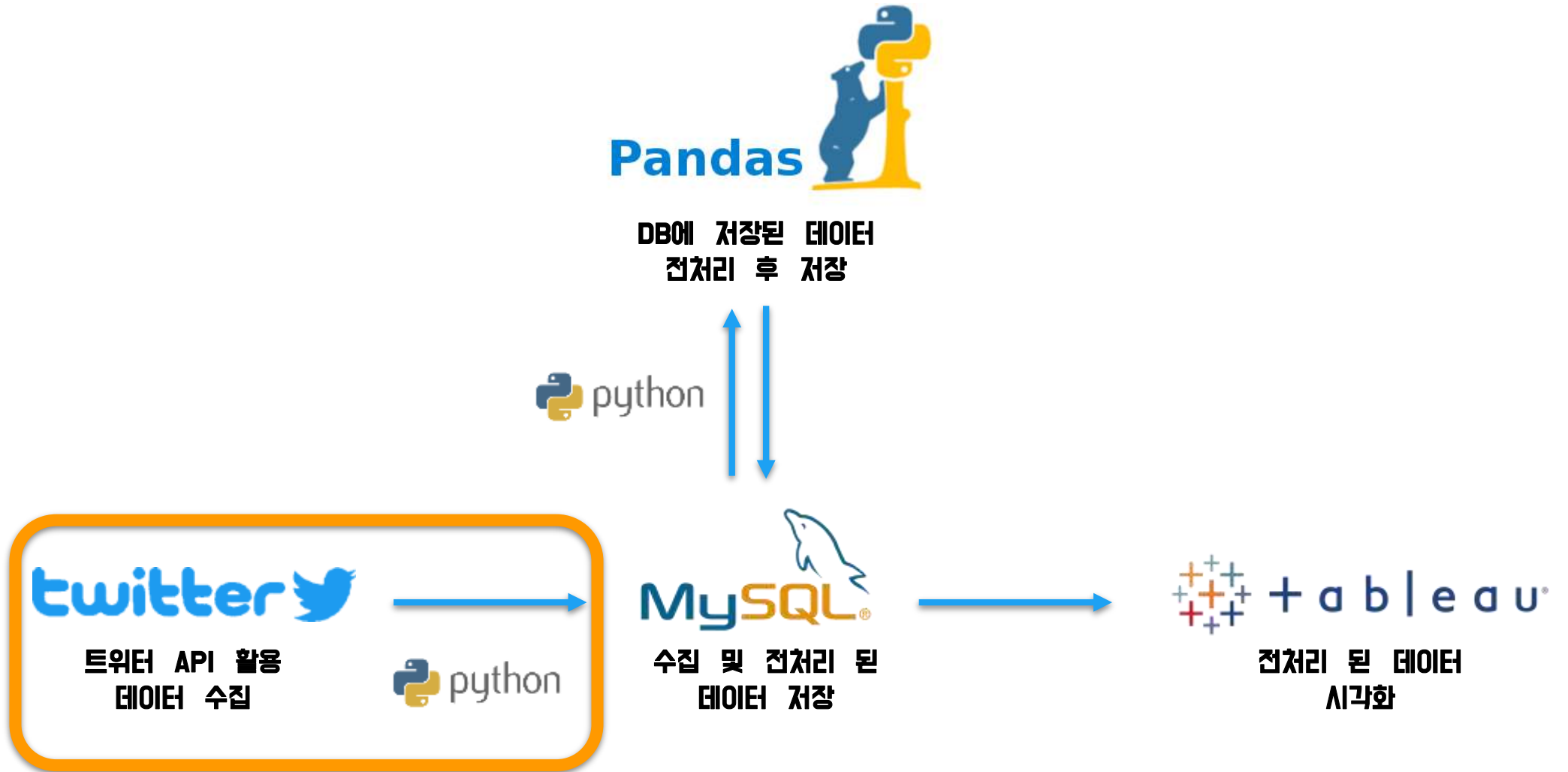


배경

데이터

파이프 라인

결과



Twit Twit

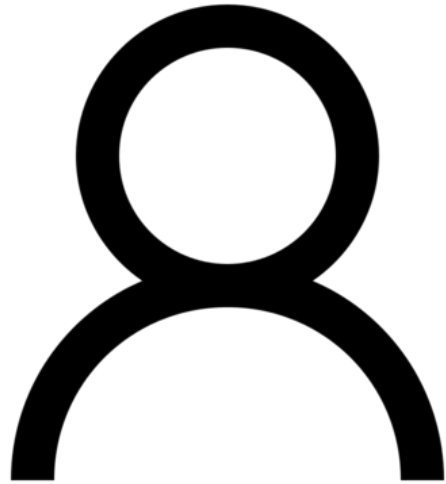
트위터 API

배경

데이터

파이프
라인

결과

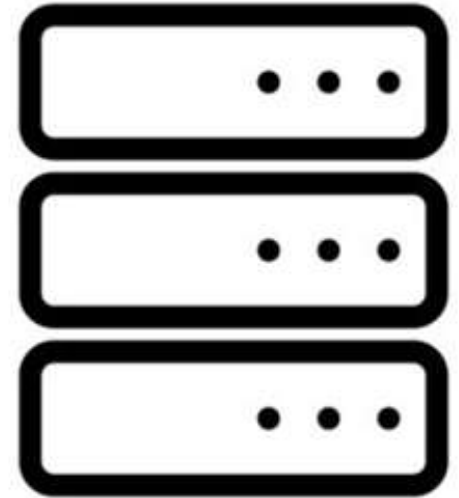


Twit Twit
Collector

HTTP REQUEST
(Query, Time, API Key, etc)



HTTP RESPONSE
(JSON)



Twitter API
Server

배경

트위터 API로 우리가 원하는 조건을 요청

api_query "검색을 원하는 키워드". "검색을 원하는 #(해시태그)"

start 검색할 구간의 시작 시간

end 검색할 구간의 끝 시간

period 검색 시간 단위 (한번의 request에 최대 100개의 tweet을 받을 수 있기때문에 적절한 시가다이를 설정)

```
api_query = Query("청년희망적금", "", False)
start = datetime.datetime(2022, 3, 2, hour=18, minute=0, second=0, microsecond=0)
end = datetime.datetime.now() - datetime.timedelta(seconds=11)
period = datetime.timedelta(hours=1)

crawler = Crawler(api_query, start, end, period)
crawler.run("tweet_tweet")
```

2022-03-02 18:00:00 와 2022-03-03 10:09:53.502114 사이에 생성된 트윗들을
1:00:00 간격으로 크롤링합니다

Progress: |#####| 100.0% Complete

총 107 개의 트윗이 dataframe에 저장되었습니다

DB에 저장?(y/n) :n

데이터

파이프
라인

결과

배경

데이터

파이프
라인

결과



트위터 API로 부터 받은 데이터

- Name(닉네임)
- Text(내용)
- Created_at(트윗발생시간)
- Author_id(작성자고유번호ID)
- Username(로그인계정)
- id(ID고유번호)



name	text	created_at	author_id	username	id
...

Twit Twit

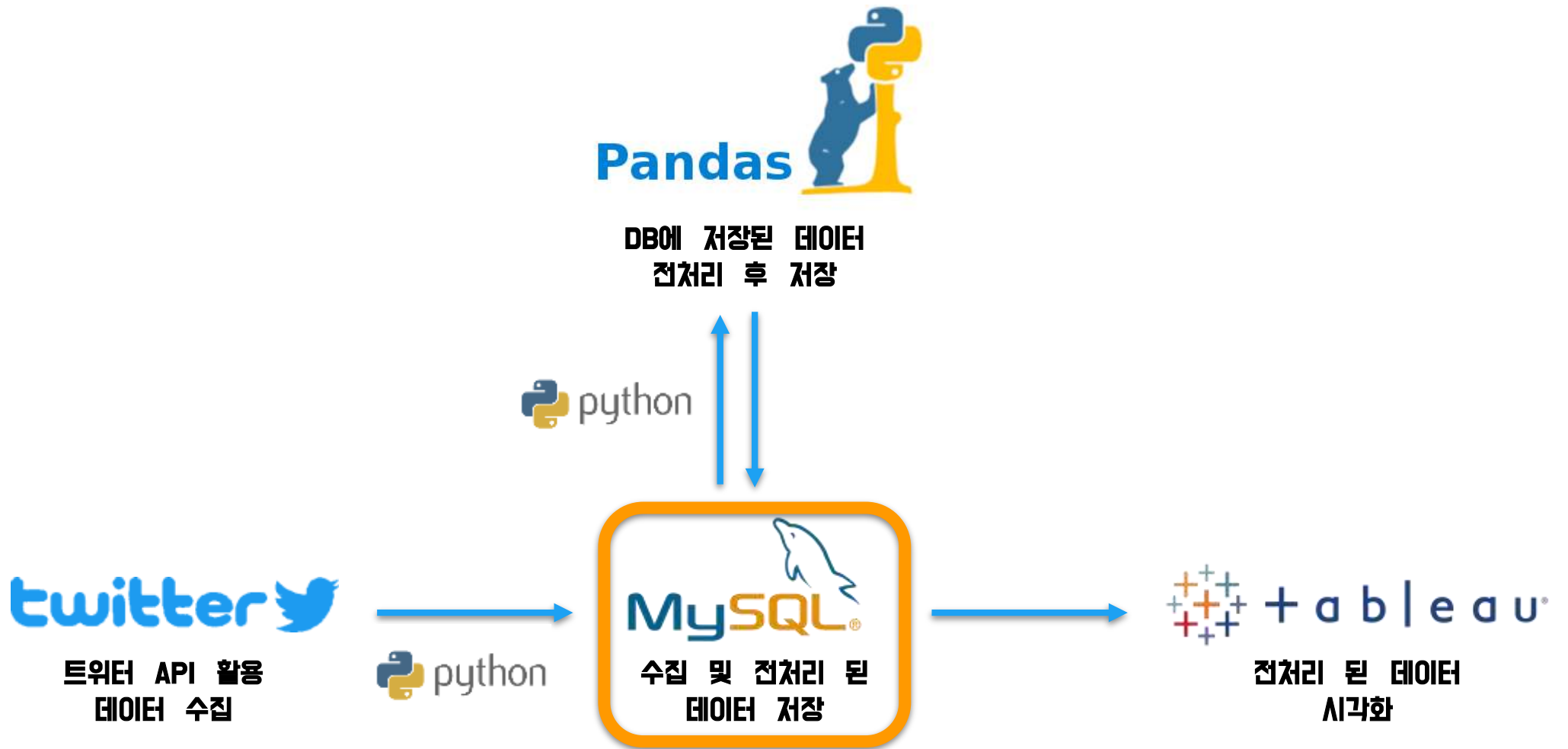
MySQL

배경

데이터

파이프
라인

결과



배경

데이터



트위터 API에서 수집한 데이터를 해시태그 별로 각각 테이블에 트위터 데이터 저장

DB

tweets

Tables

tweet_food

tweet_game

tweet_person

tweet_save

name	text	created_at	author_id
Lana	#트친과 입맛궁합 알아보...	2022-02-19T08:54:18.000Z	103631155409275289
クマ姫/윘화	#트친과 입맛궁합 알아보...	2022-02-19T11:23:05.000Z	2646683876
낭만 🍒 (A.K.A. 김아마)	#트친과 입맛궁합 알아보...	2022-02-20T11:23:09.000Z	924250587364859904
다툼 없는 사실	#트친과 입맛궁합 알아보...	2022-02-20T11:21:37.000Z	120434849041693900
GyeoL	#트친과 입맛궁합 알아보...	2022-02-20T11:05:35.000Z	122298348329689497
로지마 🌈	#트친과 입맛궁합 알아보...	2022-02-20T11:05:27.000Z	864074366056673280
수과 호포웨함!	#트친과 입맛궁합 알아보...	2022-02-20T12:25:25.000Z	711812399603408896
Vlu	#트친과 입맛궁합 알아보...	2022-02-20T12:20:57.000Z	120616584863893094
◆아그네스 복자◆	#트친과 입맛궁합 알아보...	2022-02-20T12:19:52.000Z	3034755816
피데페	#트친과 입맛궁합 알아보...	2022-02-20T12:18:19.000Z	745636519663198208
노병	#트친과 입맛궁합 알아보...	2022-02-20T12:16:35.000Z	147071862126166425
정말기여울 🌈	#트친과 입맛궁합 알아보...	2022-02-20T12:08:45.000Z	744096133908070400
마돌이	#트친과 입맛궁합 알아보...	2022-02-20T12:01:39.000Z	145839909002349772
김마오	#트친과 입맛궁합 알아보...	2022-02-20T12:00:52.000Z	135663202601150873
흔로생	#트친과 입맛궁합 알아보...	2022-02-20T11:59:15.000Z	121744653386962944
뭉이 (밍마)	#트친과 입맛궁합 알아보...	2022-02-20T11:42:46.000Z	113733583525725798
o.O.O.O.O.O	#트친과 입맛궁합 알아보...	2022-02-20T13:23:20.000Z	137067061878222438
양행숭문구매하기	#트친과 입맛궁합 알아보...	2022-02-20T13:12:38.000Z	147853219395785934
지하	#트친과 입맛궁합 알아보...	2022-02-20T12:57:35.000Z	149045543516537651
푸	#트친과 입맛궁합 알아보...	2022-02-20T12:54:49.000Z	3332423031

파이프 라인

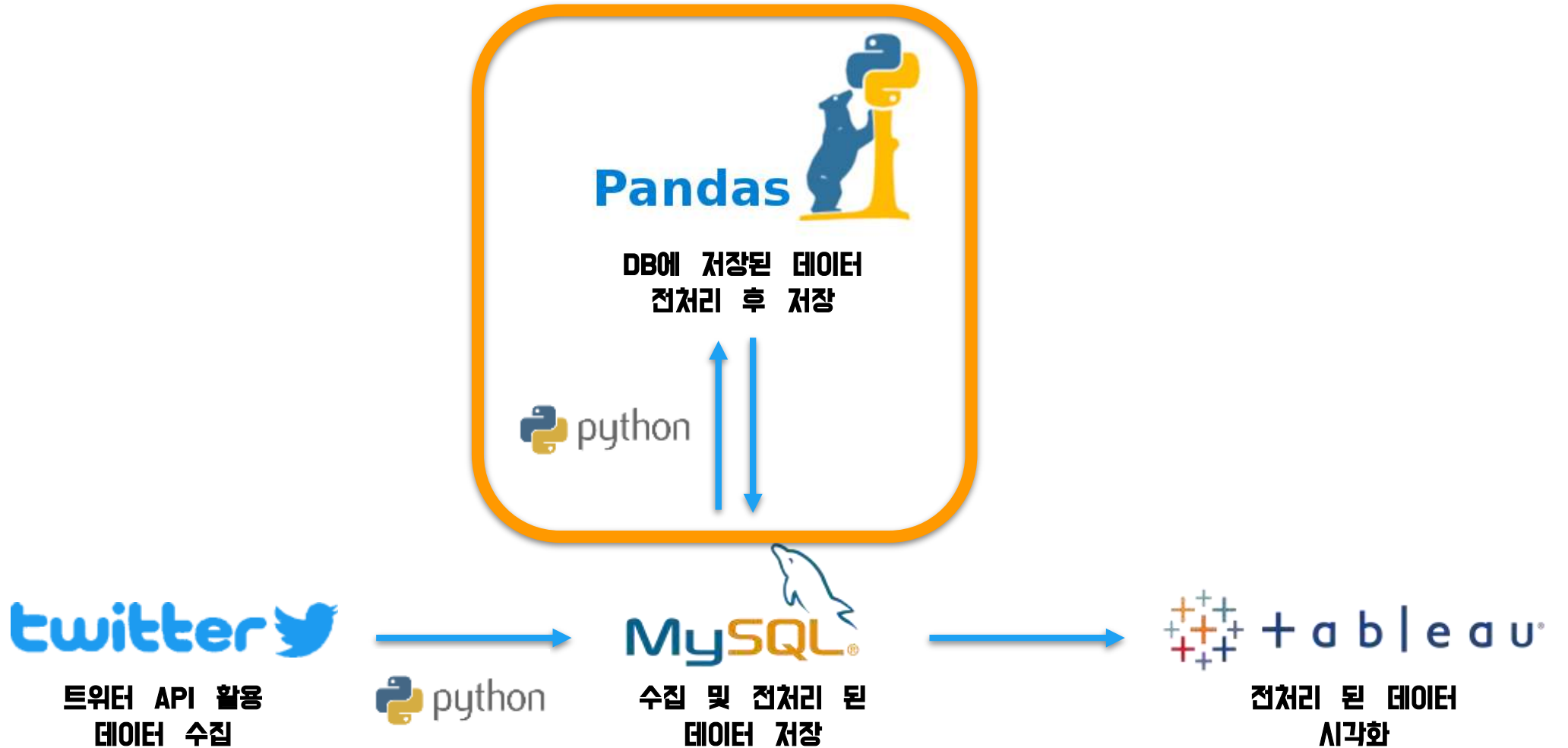
결과

배경

데이터

파이프
라인

결과



전처리_나이,MBTI,키

배경

데이터

파이프라인

결과

text 중에 나이, MBTI, 키 단어만 추출해서 저장

Input
String

#본사람다하기

본명 : 알려주겠냐

나이 : 만 14세

키 : 183

몸무게 : 7n

MBTI :ENFP

지갑 잔고 : 2만 5천원

통장 잔고 : 모름

거주지 : 외국

폰기종: 갤럭시 S21+

취미: 노래 듣기,

와일드리프트, 사진 찍기,

화분 키우기

Parser



Pandas



NumPy

Output
Data
Frame

id	mbti	age	height
08H_erb	NaN	14	NaN
0harin	ISTJ	NaN	167
mintang2728	ESFJ	19	161
Stella91114207	ISTP	16	NaN
zzangboy98	ISFJ	25	181
...
an__onnn	ISFP	29	158
mateu0830_0314	ISFP	18	163
muddthejn	NaN	NaN	160
HeheKoean	ISTP	NaN	171
37myRPECoiFdW0	NaN	NaN	159

전처리_거주지

배경

데이터

파이프라인

결과

text 중에 **거주지** 단어만 추출해서 저장

Input
String

...
...
통장 잔고 : 모름
거주지 : 경기
폰기종: 갤럭시 S21+
...
...

Input
File



대한민국행정구역.SHP

Parser

python

Pandas



NumPy



GeoPandas

* 지도로 시각화 하기위해 geopandas 사용

Output
Data
Frame

경기	MULTIPOLYGON (((931607.470 18944
경남	MULTIPOLYGON (((1053643.218 1612
경북	MULTIPOLYGON (((1171290.353 1782
광주	POLYGON ((932712.687 1696168.69
대구	POLYGON ((1107777.593 1780522.26
대전	POLYGON ((990946.725 1832389.26
산	MULTIPOLYGON (((1137080.579 1677
서울	POLYGON ((966987.226 1941110.94
세종	POLYGON ((971235.936 1844386.98
울산	MULTIPOLYGON (((1167949.952 1710
인천	MULTIPOLYGON (((897904.035 19610
전남	MULTIPOLYGON (((946598.396 15552
전북	MULTIPOLYGON (((902676.539 17174
제주	MULTIPOLYGON (((885004.560 14587

전처리_게임

배경

데이터

파이프라인

결과

text 중에 **게임** 별로 1과 0으로 값을 저장

Input
String

#본_사람_다하기
롤: 짹먹만: 0
오버워치: L
파판14: 0
로스트아크: 0
트릭스터: L
카트라이더: 0
메이플스토리: 0
...

Parser

python

Pandas



NumPy

Output
Data
Frame

롤	오버워치	파판14	로스트아크	트릭스터	카트라이더
NaN	0	1	1	0	0
0	1	0	0	0	0
1	1	1	1	1	1
0	0	0	0	0	0
1	0	0	0	0	0
...
0	NaN	0	NaN	0	0
NaN	NaN	NaN	0	0	0
1	1	1	1	1	0
1	1	1	0	0	0
1	0	0	NaN	0	0

전처리_청년희망적금

배경

데이터

파이프라인

결과

text 중에 단어별로 카운트하여 저장

Input
String

청년희망적금 하는 방법
알려드립니다.

조건 :

- 나이: 가입일당시
만19세~만34세 이하
- 2020년 국세청에서
소득금액이 잡혀야 하며
- 2020년 총급여 3.600만원
이하 또는 종합소득금액
2.600만원 이하

소득 금액 확인 하는 방법은
국세청 홈택스 들어가서
소득금액증명->2020년꺼
발급해서 확인

Parser



Pandas



KoNLPy



WordCloud Library

Output
Dictionary

338, '가입': 994, '신청': 73
' : 118, '대상자': 113, '부제'
, '가능': 87, '오류': 86, '농협'
확인': 71, '조회': 71, '농협'
62, '연봉': 62, '대기': 62,
통령': 55, '네이버': 54, '그
급': 46, '이제': 45, '요건'
'폭주': 39, '시작': 38, '미리
3, '이하': 35, '나이': 35,
달': 32, '어제': 32, '고민':
29, '혹시': 29, '비대': 29,
트': 28, '선착순': 27, '급여
세금': 26, '정보': 26, '회사
24, '부터': 24, '상품': 24,
인구': 23, '문자': 22, '시발':
21, '금요일': 21, '슈크림': 2
'광주': 19, '어케': 19, '거기
' : 18, '동안': 18, '충족':
'대그': 17, '어휘': 17, '이

* 문자열의 형태소 분리를 위해 KoNLPy
사용

전처리_음식

배경

데이터

파이프라인

결과

text 중에 음식별로 점수를 추출해서 저장

Input
String

#트친과_입맛궁합_알아보
기
5맨날팽김
4거의팽김
3종종팽김
2내돈주고안먹기
1사침도 잘안먹기

육화: 3
마라탕: 2
간장게장: 1
선지: 2
민초: 3

Parser

python

Pandas



NumPy

Output
Data
Frame

육화	곰창	닭똥집	알탕	산낙지	오뎅	양파
5	5	4	2	4	4	
3	2	3	3	3	4	N
4	4	4	3	3	3	
4	5	5	5	5	5	
5	1	2	0	3	4	N
...	
4	4	4	3	4	4	
5	5	5	NaN	4	3	N
1	0	2	0	NaN	3	
3	3	4	4	4	2	
0	0	0	0	3	3	

배경

데이터



전처리가 완료된 데이터를 해시태그 별로 각각 테이블에 저장

DB

tweets

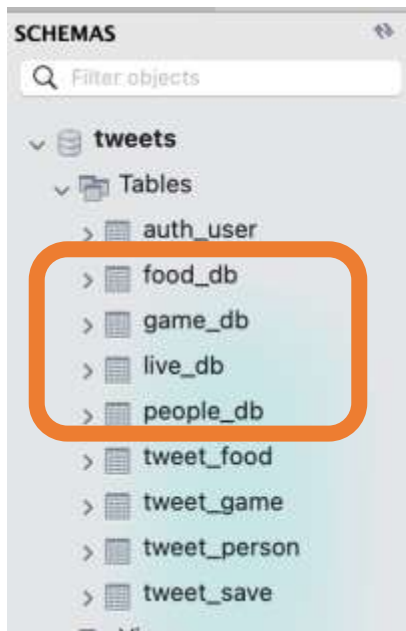
Tables

food_db

game_db

person_db

live_db



id	육회	마라탕	간장게장	선지	민초	탄산수	커피	생굴	닭발	막창	급
Adorable_Lana	3	2	1	2	3	3	5	1	3	5	5
kumahimeinlove	3	NULL	2	3	5	4	3	2	3	3	2
NangmanLaws	5	4	4	2	2	3	2	3	2	4	4
iamnotlogical_	NULL	3	5	2	4	1	3	5	4	4	5
OuO76151129	NULL	3	2	0	0	3	0	3	4	5	1
nolawschoolstay	5	NULL	1	5	4	5	0	NULL	NULL	4	1
Sugwa56	2	3	NULL	0	0	1	1	0	3	5	5
Blue_veri	5	5	5	NULL	0	1	2	0	5	5	5
xoxolts_	4	3	2	3	3	3	4	2	3	4	4
rksk9944	2	3	3	2	2	4	5	3	3	2	4
nosickk	0	NULL	0	5	5	5	0	0	0	0	0
HBKEDYEOUL	4	4	2	3	3	3	5	1	3	1	1
ymmhot	4	5	4	3	3	2	5	3	1	5	5
Nenechicken99	3	4	4	2	4	3	5	4	4	4	3
gongburobot	5	0	4	2	2	3	3	1	0	1	4
rrrararr	3	3	3	2	2	2	5	4	3	3	3
UuUu_0o0_uU...	4	5	1	2	3	3	4	0	4	5	5
EnBbang	3	3	3	1	3	3	3	1	3	3	3
z_ha	5	3	3	0	2	1	3	2	2	2	2
bAv0Ji	5	5	1	3	0	4	1	1	5	5	3
s2uuu1	3	3	0	0	3	0	5	2	0	2	3

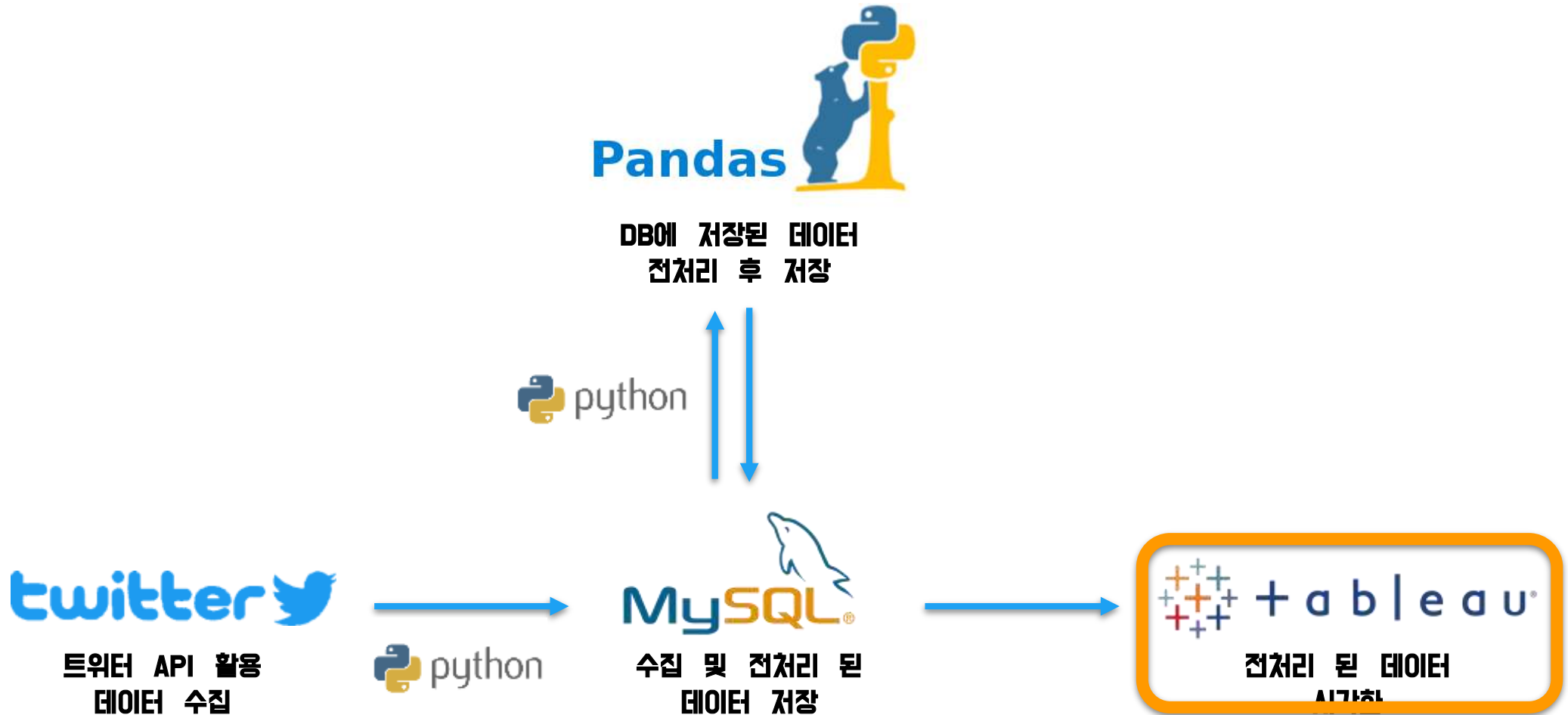
결과

배경

데이터

파이프
라인

결과



시각화 - Tableau

배경

데이터

파이프 라인

결과

MySQL과 연동하여 시각화 진행

The screenshot shows the Tableau interface with a MySQL connection established. The 'tweets_db' data source is selected, and the 'food_db' table is chosen. The interface displays a list of tables and a detailed view of the 'food_db' table structure and data.

연결 (Connection): 13.112.232.65 MySQL

데이터베이스 (Database): tweets

테이블 (Table): 전체 9 권장

테이블 이름 입력 (Table Name Input):

- auth_user
- food_db
- game_db
- live_db
- people_db
- tweet_food
- tweet_game
- tweet_person
- tweet_save
- 새 사용자 지정 SQL
- 새 유니온

food_db (17개 필드 9492개 행)

이름 (Name): food_db

필드 (Fields):

유형	필드명	물리적 테이블	원격 필드명
ABC	Id	food_db	id
#	육회	food_db	육회
#	마라탕	food_db	마라탕
#	간장게장	food_db	간장게장

데이터 테이블 (Data Table):

Id	육회	마라탕	간장게장	선지	민초	탄산수	커피
Adorable_Lana	3	null	1	2	3	3	5
kumahmeinlove	3	null	2	3	5	4	3
NangmanLaws	5	4	4	2	2	3	2
iamnotlogical_	null	3	5	2	4	1	3
OuO76151129	null	3	2	0	0	3	0
nolawschoolstay	5	null	1	5	4	5	0
Sugwa56	2	3	null	0	0	1	1

시각화 - Tableau

배경

데이터

파이프라인

결과

MySQL과 연동하여 시각화 진행



04

결과

Twit Twit

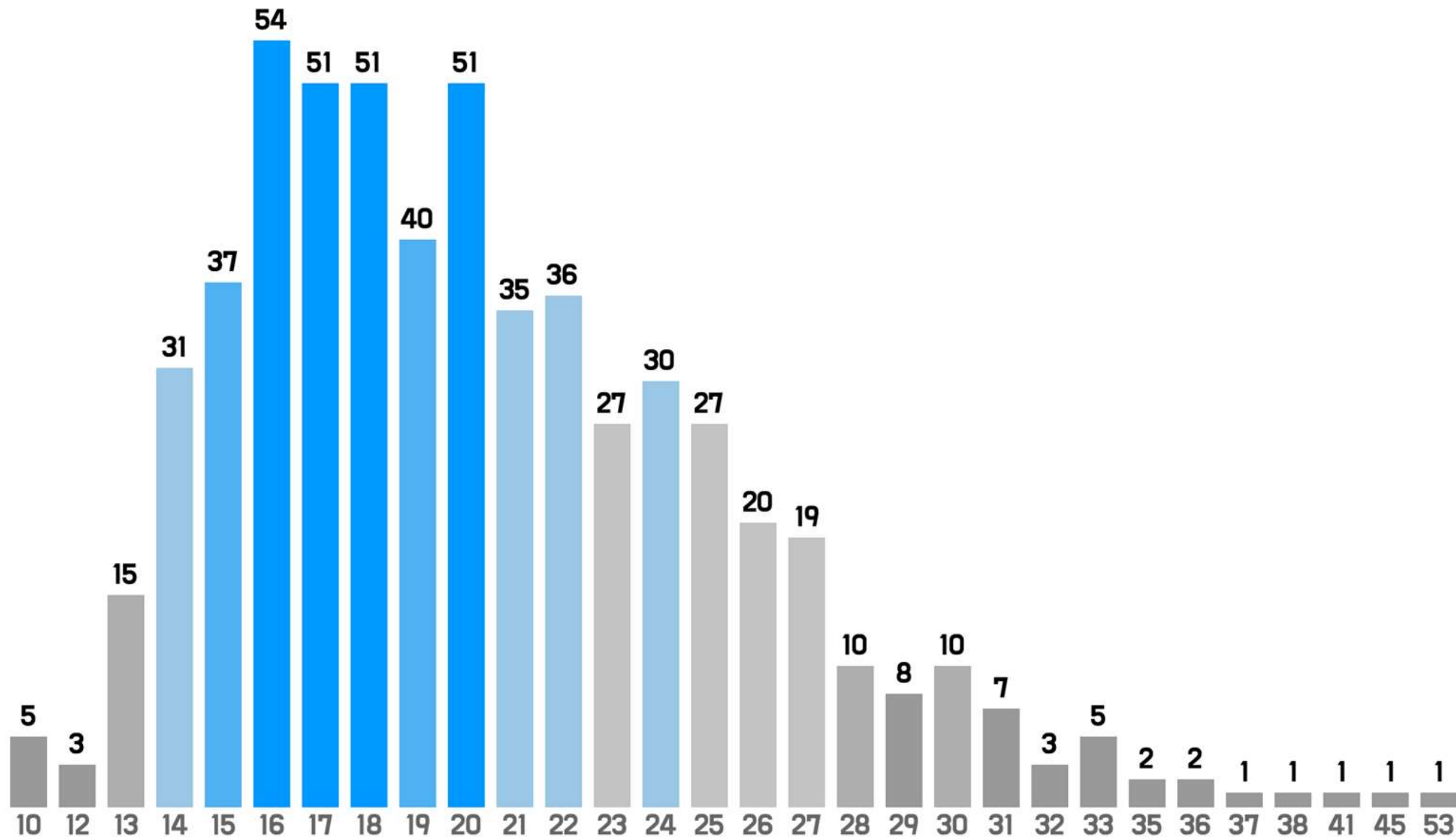
#트위터 이용자 나이 분포

배경

데이터

파이프 라인

결과



Twit Twit

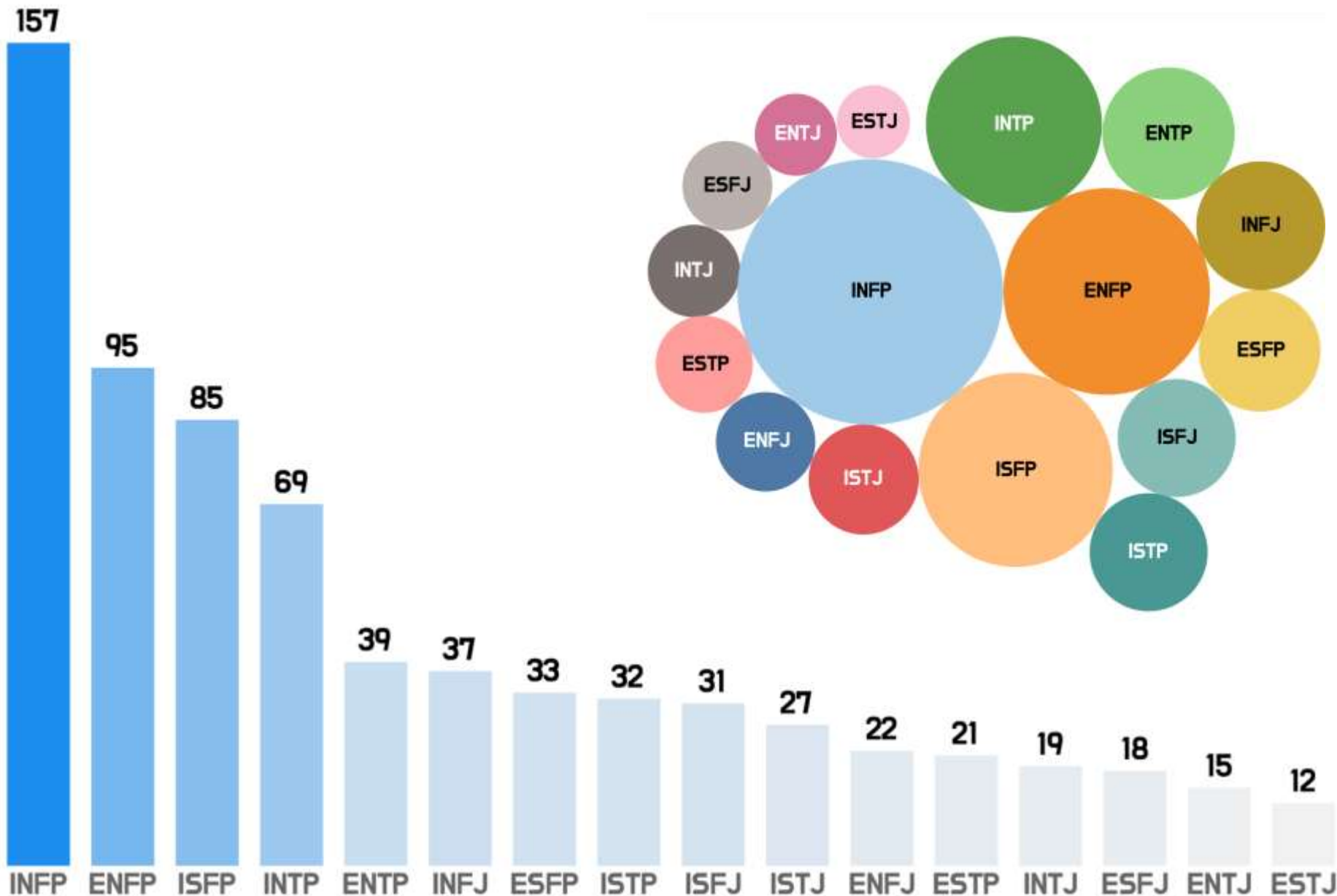
#트위터 이용자 MBTI 분석

배경

데이터

파이프 라인

결과



Twit Twit

배경

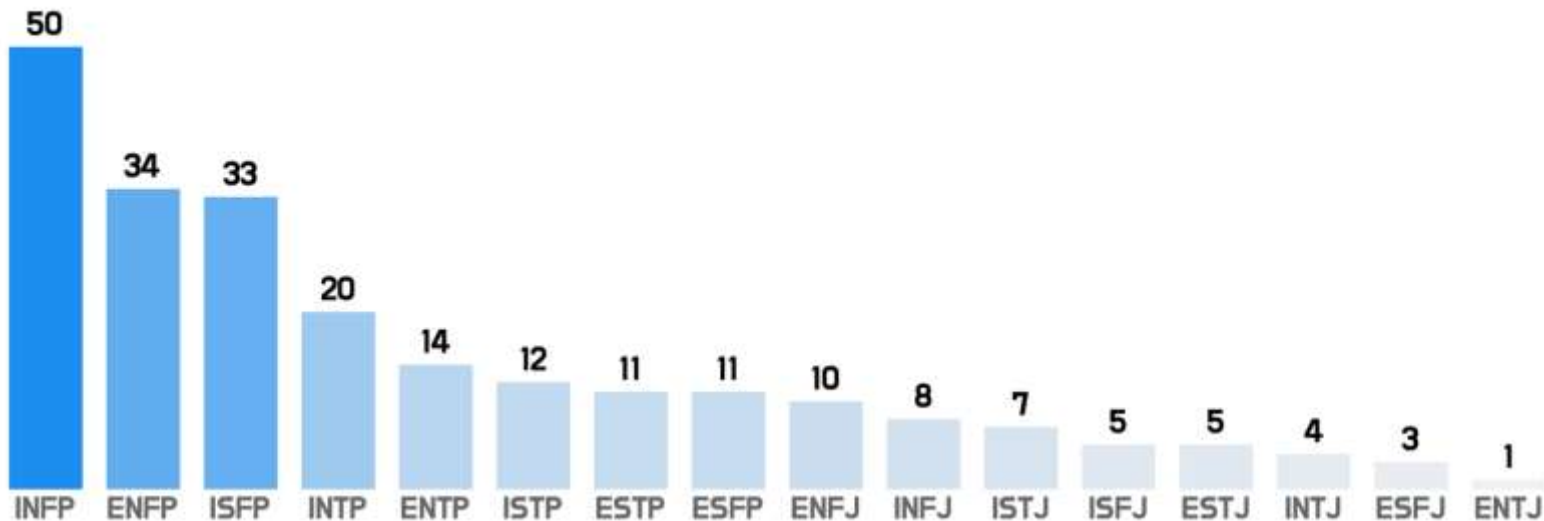
데이터

파이프
라인

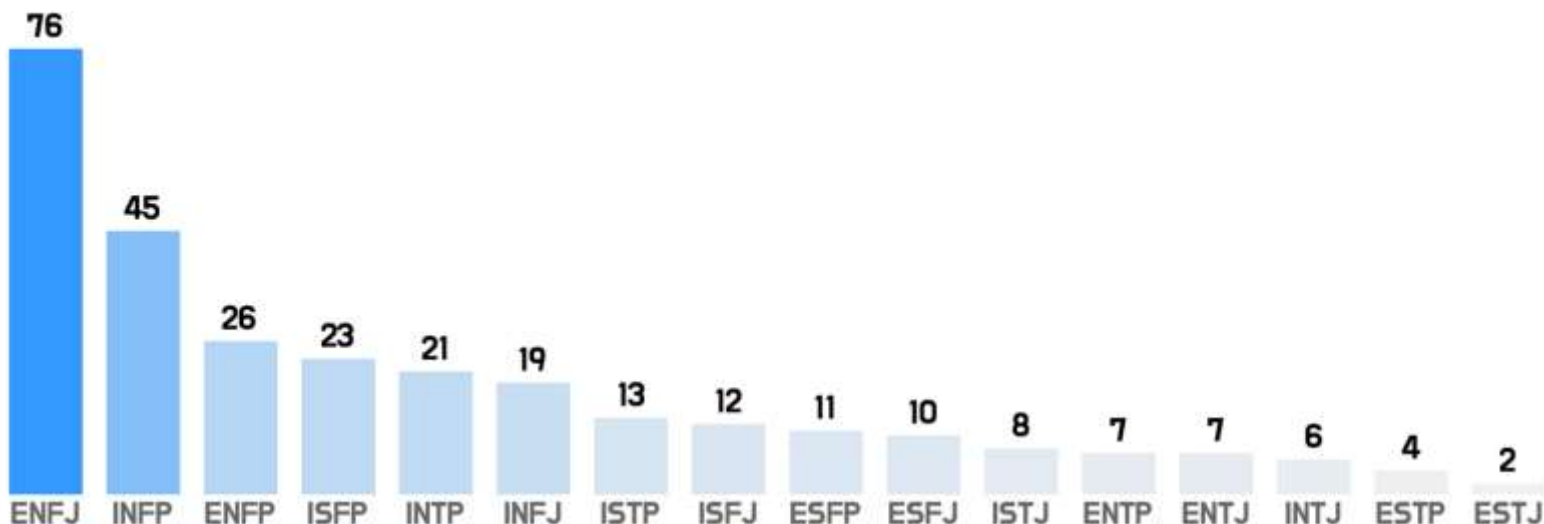
결과

#MBTI_성인,청소년

MBTI 청소년



MBTI 성인



Twit Twit

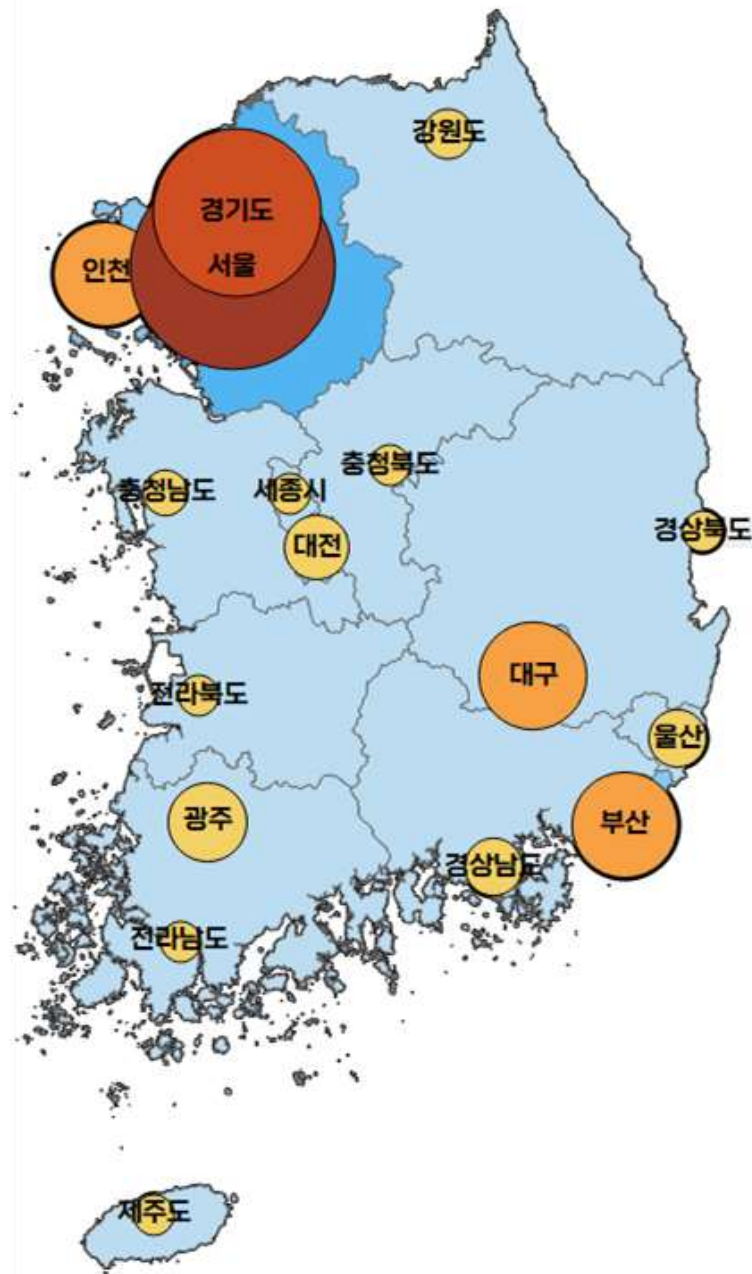
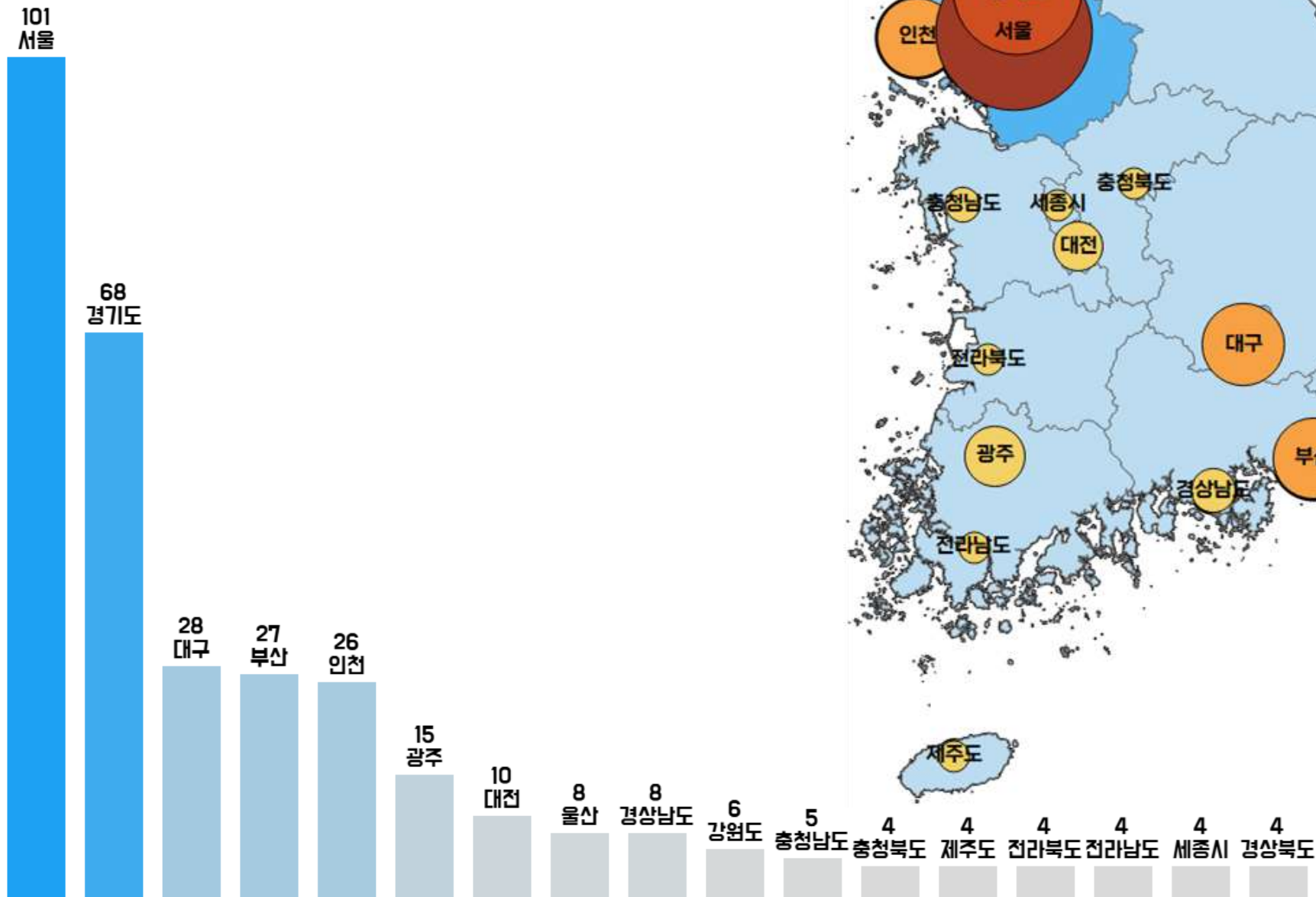
#트위터 이용자 거주지 분포

배경

데이터

파이프 라인

결과



Twit Twit

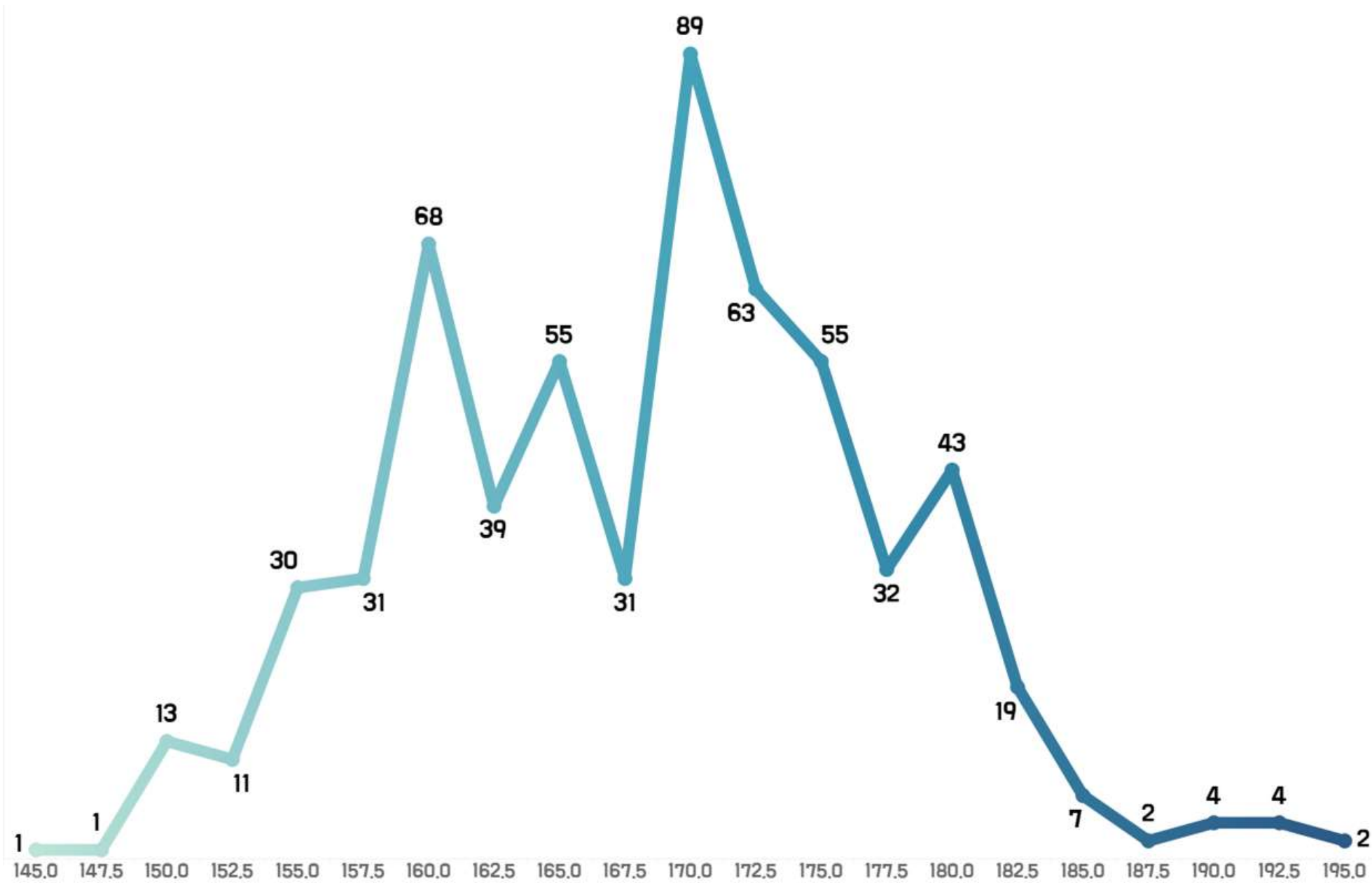
#트위터 이용자 키 분포

배경

데이터

파이프 라인

결과



Twit Twit

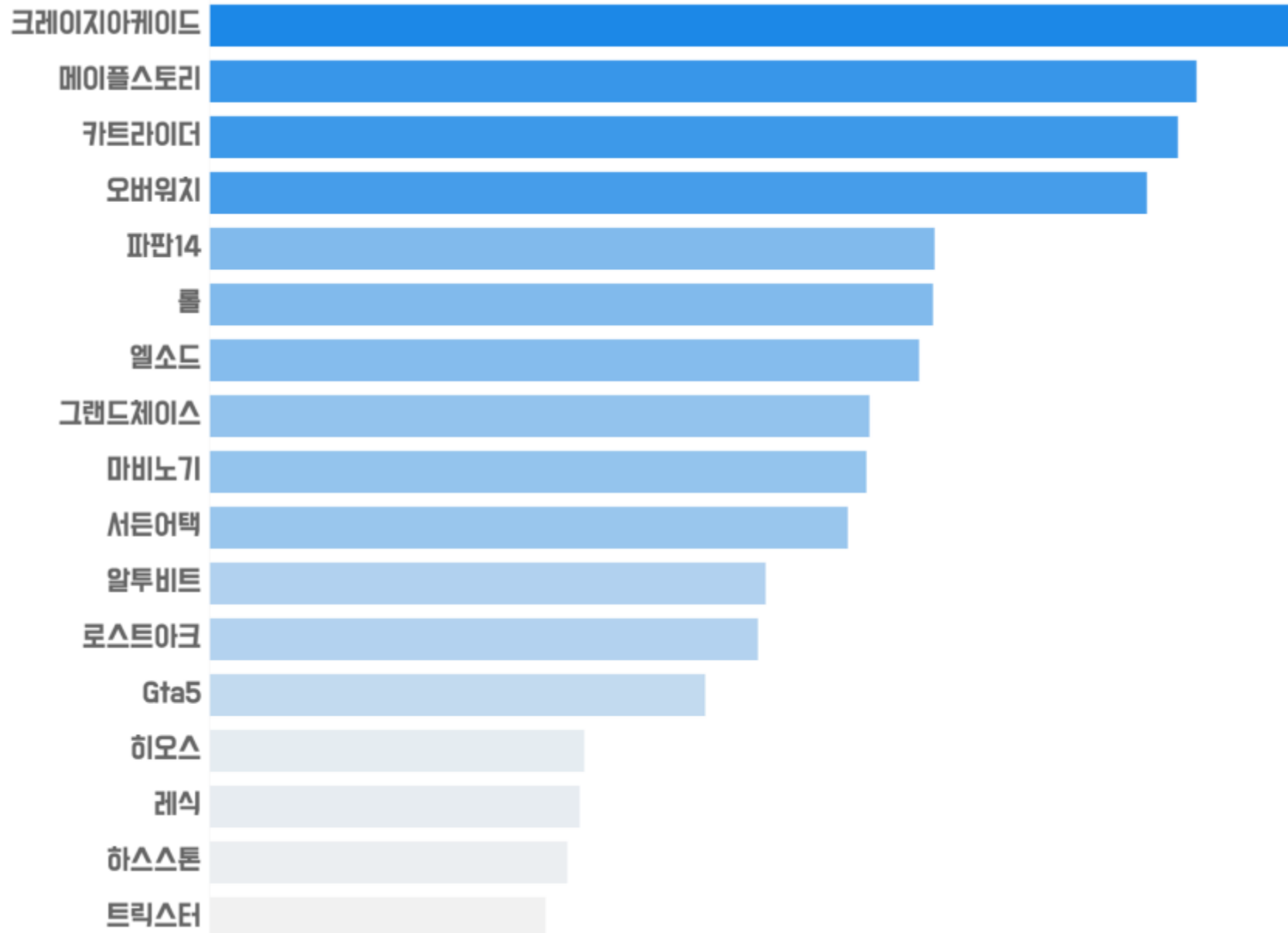
배경

데이터

파이프
라인

결과

#해본 게임 말하기



Twit Twit

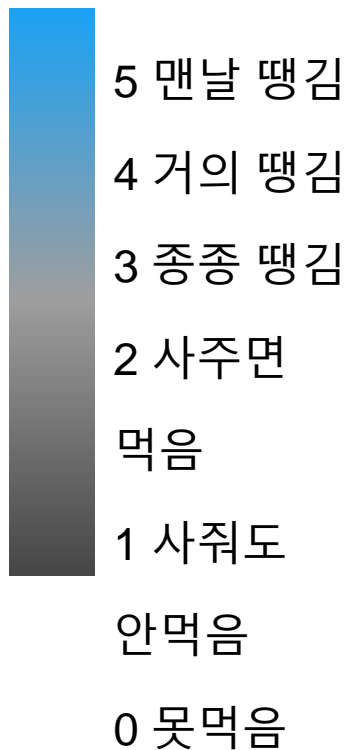
#음식 평점 매기기(0~5점)

배경

데이터

파이프 라인

결과



마라탕 3.4	막창 3.1	곱창 2.9	산낙지 2.8	커피 2.7
오뎅 3.3	닭발 2.6	닭통집 2.3	간장게장 2.2	알탕 2.0
양꼬치 3.3	민초 2.6			
육회 3.2	탄산수 2.5	선지 1.8	생굴 1.5	

Twit Twit

배경

테이터

파이프 라인

결과

#청년희망적금(워드클라우드)



Twit Twit

시연

배경

데이터

파이프
라인

결과



Twit Twit

느낌점

배경

데이터

파이프
라인

결과

심다훈

다들 프로젝트 열심히
하는 모습이 좋았고 서로
안되는 부분이나 어려운
부분에 있어서 서로
도우며 진행 되었던 점이
더 좋았던 것 같습니다.
다음에 또 만나고 싶네요
ㅎㅎㅎ

박휘찬

이번 프로젝트로 여러가지
툴을 사용해보고 다양한
시각화를 해볼 수 있어서
좋았습니다. 최종
프로젝트에서는 더 많은
것을 해보면 좋을 것
같습니다

이성진

잘하시는 분들과 함께해서
너무 즐거웠습니다. 배운
점도 정말 많고 즐겁게
웃으면서 한 점이 정말
좋았습니다. 저에게 이
프로젝트는 저를 한단계
더 높게 만들어
주었습니다. 정말
재미있었고 마지막
프로젝트 때도 함께했으면
정말 좋다고 느꼈습니다.

홍성희

프로젝트를 하면서 제가
부족함이 많았는데 다들
열정있고 인욕이 넘치는
모습을 보니 저도 자극을
많이 받아 더 열심히
공부해야겠다는 생각을
많이 하게 되었던
프로젝트였습니다.
프로젝트가 끝나는 것이
너무나 아쉽습니다. 함께해서
너무 감사하고
즐거웠습니다.

Q & A

(질문 하지 마세요)



Twit Twit

감사합니다 :D