

회사 밖의 데이터 - 데이터로 사이드 프로젝트 만들기

나의 브런치 구독자 데이터 분석기

파이썬을 활용한 데이터 분석기

강원양



회사 안의 데이터

01 하는 일



SOSCON 2019

SAMSUNG OPEN SOURCE CONFERENCE 2019

하는 일

데이터 시각화를 기획하고 만들고 이야기 합니다.



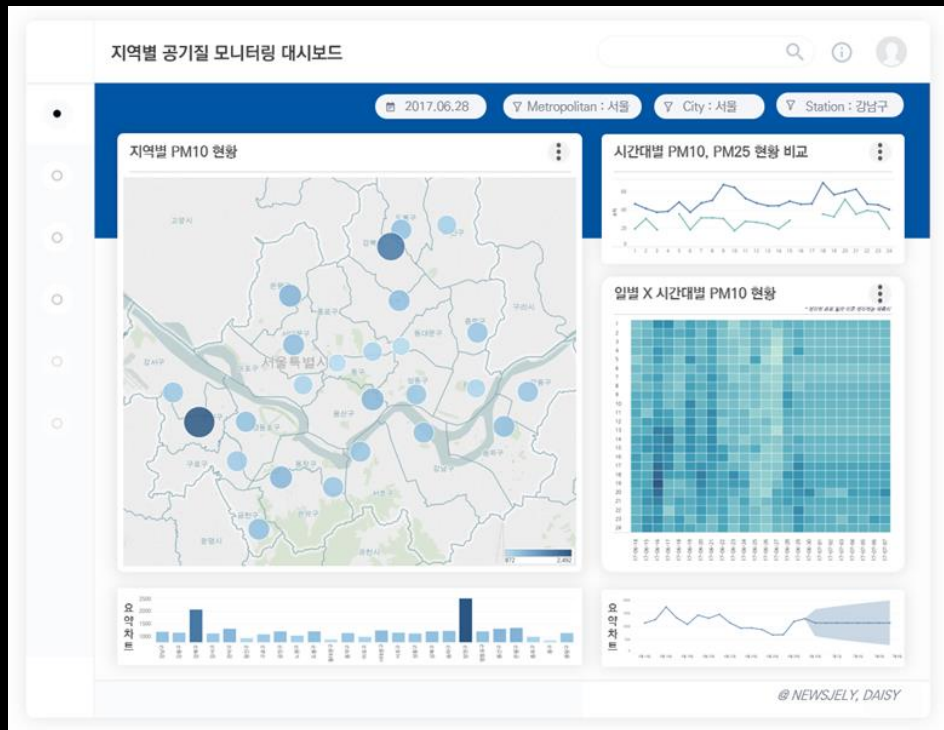
하는 일

데이터 시각화를 기획하고 만들고 이야기 합니다.



하는 일

동일한 데이터도
활용 목적에 따라
더 효과적인 방법으로 표현



하는 일

데이터 시각화를 기획하고 만들고 이야기 합니다.

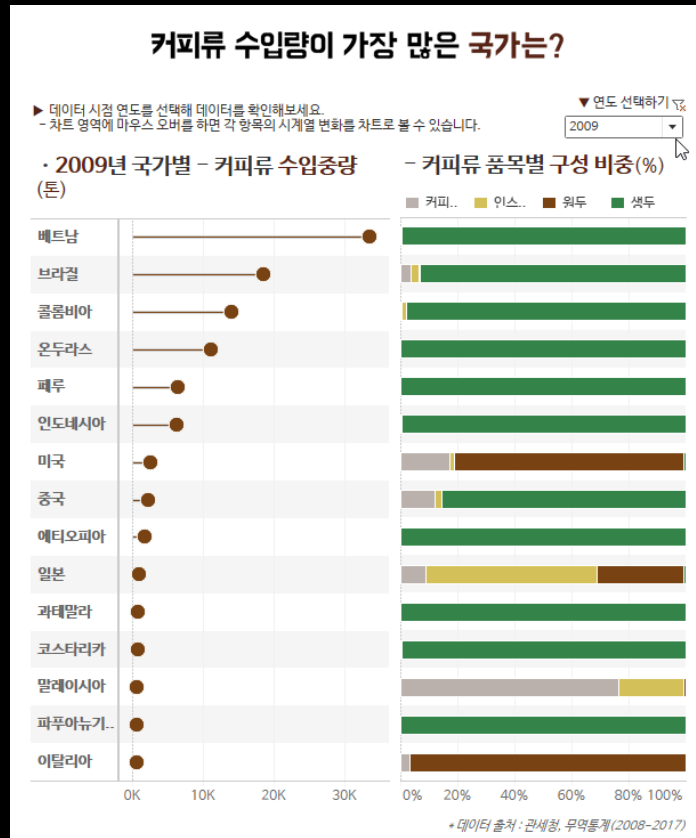


하는 일

시각화를 통한 데이터 분석

시각적 분석 기반 인사이트 도출

데이터 스토리텔링 콘텐츠



회사 밖의 데이터

02 프로젝트 배경



SOSCON 2019

SAMSUNG OPEN SOURCE CONFERENCE 2019

왜 프로젝트를 하게 되었나?

- 데이터를 활용해 궁금증을 해결하기
- 현실적인 한계를 극복하기 위한 시도하기
- 데이터 분석 과정에서 시각화의 역할 이야기하기



어떻게 프로젝트를 하게 되었나?

- 파이썬 6주 수업
- 선생님과 함께하는 미니 프로젝트



회사 밖의 데이터

03 프로젝트 개요



SOSCON 2019

SAMSUNG OPEN SOURCE CONFERENCE 2019

프로젝트 개요

← → ↺ 🔒 <https://brunch.co.kr/@dimension-value#articles> 🔍 ☆ 시크릿 모드 🍷 ☰

≡ *brunch* beta



강원양

데이터시각화 에디터

데이터로 나의 브런치 구독자 분석하기

프로젝트 개요

← → ↺ 🔒 <https://brunch.co.kr/@dimension-value#articles> 🔍 ☆ 시크릿 모드 🍷 ☰

≡ *brunch* beta



강원양

데이터시각화 에디터

나를 구독하는 사람
내가 좋아하는 작가

어떤 사람?
어떤 관계?

프로젝트 개요

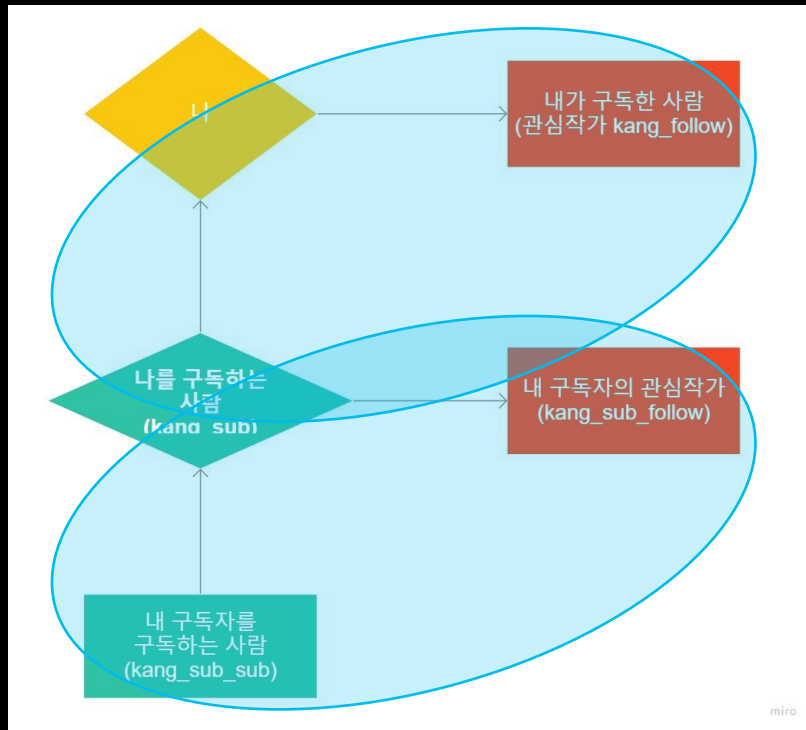


회사 밖의 데이터

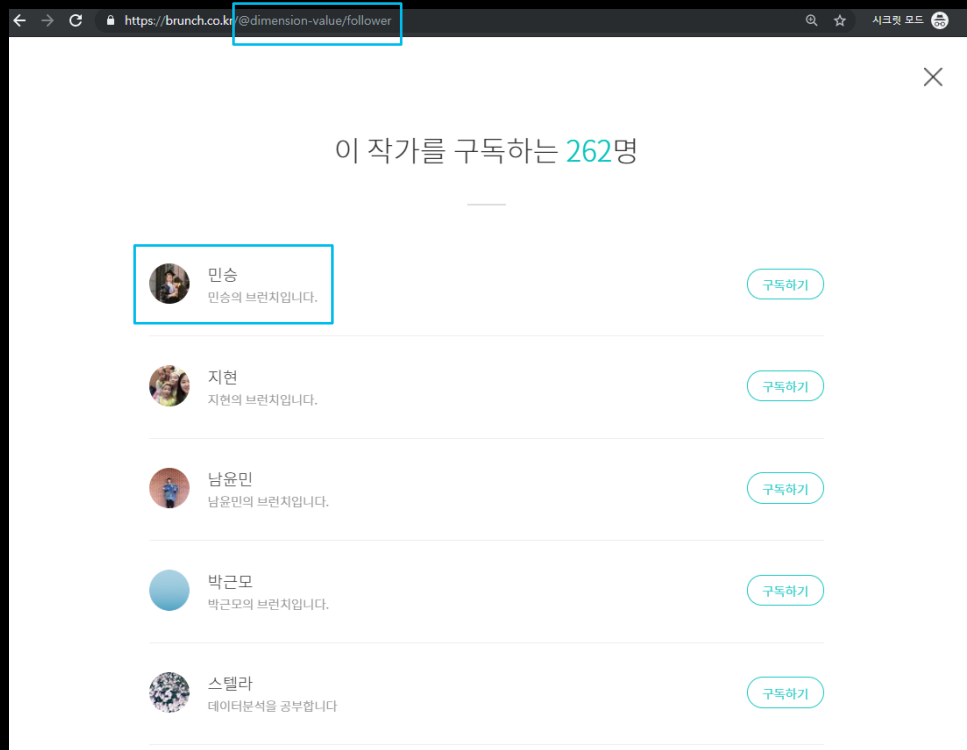
03 프로젝트 내용



파이썬을 활용한 데이터 수집 및 정제



웹 크롤링 : 구독자/ 관심작가 리스트



웹 크롤링 : 구독자/ 관심작가 리스트

```
In [1]: 1 from selenium import webdriver
2 from selenium.webdriver.common.keys import Keys
3 from bs4 import BeautifulSoup
4 import time

In [2]: 1 #구독자는 follower, 관심작가는 following
2
3 sub_url = "https://brunch.co.kr/@dimension-value/follower"

In [3]: 1 driver = webdriver.Chrome('./drivers/chromedriver.exe')
2 driver.get(sub_url)

In [4]: 1 main = driver.find_element_by_tag_name("body")

In [5]: 1 #무한 스크롤링
2
3 num_of_pagedowns = 50
4
5 while num_of_pagedowns:
6     main.send_keys(Keys.PAGE_DOWN)
7     time.sleep(0.3)
8     num_of_pagedowns -= 1

In [6]: 1 html = driver.page_source

In [7]: 1 soup = BeautifulSoup(html, 'html.parser')

In [8]: 1 profiles = soup.select('ul.list_follow > li')
2 print(type(profiles), len(profiles))

<class 'list'> 259

In [10]: 1 index = 0
2 sub_data=[]
3
4 for sub_profile in profiles:
5     sub_name = soup.select('.tit_subject')[index].text
6     sub_desc = soup.select('.desc_follow')[index].text
7     sub_link = soup.select('a.link_follow')[index].text
8     print("/@dimension-value", index, sub_name, sub_desc.strip(), sub_link, sep = ' ')
9     sub_data.append(["/@dimension-value", index, sub_name, sub_desc.strip(), sub_link,])
10    index = index + 1

/@dimension-value 0 | 스티라 | 데이터 분석을 공부합니다 | /@ivoryoh0609
/@dimension-value 1 | castellan | castellan의 브런치입니다. | /@hancastellan
/@dimension-value 2 | KyeongMi | KyeongMi의 브런치입니다. | /@kyeongmifu8
/@dimension-value 3 | 알인환 | 알인환의 브런치입니다. | /@reav1212
/@dimension-value 4 | 박구라 | 배우는 법을 배워가는 중입니다. | /@gurapark
```



웹 크롤링 : 구독자/ 관심작가 리스트

```
In [1]: 1 from selenium import webdriver
2 from selenium.webdriver.common.keys import Keys
3 from bs4 import BeautifulSoup
4 import time

In [2]: 1 #구독자는 follower, 관심작가는 following
2
3 sub_url = "https://brunch.co.kr/@dimension-value/follower"

In [3]: 1 driver = webdriver.Chrome('./drivers/chromedriver.exe')
2 driver.get(sub_url)

In [4]: 1 main = driver.find_element_by_tag_name("body")

In [5]: 1 #무한 스크롤링
2
3 num_of_pagedowns = 50
4
5 while num_of_pagedowns:
6     main.send_keys(Keys.PAGE_DOWN)
7     time.sleep(0.3)
8     num_of_pagedowns -= 1

In [6]: 1 html = driver.page_source

In [7]: 1 soup = BeautifulSoup(html, 'html.parser')

In [8]: 1 profiles = soup.select('ul.list_follow > li')
2 print(type(profiles), len(profiles))

<class 'list'> 259

In [10]: 1 index = 0
2 sub_data=[]
3
4 for sub_profile in profiles:
5     sub_name = soup.select('.tit_subject')[index].text
6     sub_desc = soup.select('.desc_follow')[index].text
7     sub_link=soup.select('.link_follow')[index]['href']
8     print("/@dimension-value", index, sub_name, sub_desc.strip(), sub_
9     sub_data.append(["/@dimension-value", index, sub_name, sub_desc.st
10    index = index + 1

/@dimension-value | 0 | 스티라 | 데이터분석을 공부합니다 | /@ivoryh0609
/@dimension-value | 1 | castellan | castellan의 브런치입니다. | /@hancastel
/@dimension-value | 2 | KyeongMi | KyeongMi의 브런치입니다. | /@kyeongaimi
/@dimension-value | 3 | 알인환 | 알인환의 브런치입니다. | /@reay1212
/@dimension-value | 4 | 박구라 | 배우는 법을 배워가는 중입니다. | /@gurapar
```

```
In [36]: 1 driver = webdriver.Chrome('./drivers/chromedriver.exe')
2
3 total_sub_data=[]
4
5 for i in sub_url:
6     driver.get(i)
7     time.sleep(3)
8
9 SCROLL_PAUSE_TIME = 1.5
10
11 # Get scroll height
12 last_height = driver.execute_script("return document.body.scrollHeight")
13
14 while True:
15     # Scroll down to bottom
16     driver.execute_script("window.scrollTo(0, document.body.scrollHeight);")
17
18     # Wait to load page
19     time.sleep(SCROLL_PAUSE_TIME)
20
21     # Calculate new scroll height and compare with last scroll height
22     new_height = driver.execute_script("return document.body.scrollHeight")
23     if new_height == last_height:
24         break
25     last_height = new_height
26
27 html = driver.page_source
28 soup = BeautifulSoup(html, 'html.parser')
29
30 profiles = soup.select('ul.list_follow > li')
31
32 index = 0
33 for sub_profile in profiles:
34     sub_name = soup.select('.tit_subject')[index].text
35     sub_desc = soup.select('.desc_follow')[index].text
36     sub_link=soup.select('.link_follow')[index]['href']
37     print('/'+i.split('/')[3], index, sub_name, sub_desc.strip(), sub_link, sep = ' | ')
38     total_sub_data.append(['/'+i.split('/')[3], index, sub_name, sub_desc.strip(), sub_link])
39     index = index + 1

/@kkw119 | 0 | 마키세라믹 | 마키세라믹의 브런치입니다. | /@hkp0402
/@kkw119 | 1 | Dan Lee | 커피 좀 좋아해서 내려 마시고 있고, 지금까지 엔지니어도 하고 개발도 하고 지금은 기획하면서 개발될 망고
```

웹 크롤링 : 구독자/ 관심작가 리스트 데이터 수집 결과

A	B	C	D	E	F
to_link	index	from_name	from_desc	from_link	
/@dimension-value	0 스텔라	데이터분석을 공부합니다	/@ivoryoh0609		
/@dimension-value	1 castellan	castellan의 브런치입니다.	/@hancastellan		
/@dimension-value	2 KyeongMi	KyeongMi의 브런치입니다.	/@kyeongmimfu8		
/@dimension-value	3 양인환	양인환의 브런치입니다.	/@reay1212		
/@dimension-value	4 박구라	배우는 법을 배워가는 중입니다.	/@gurapark		
/@dimension-value	5 김은아	김은아의 브런치입니다.	/@happynaa		
/@dimension-value	6 joo	HE.StoryFolio	/@storyfolio		
/@dimension-value	7 허성오	데이터 사이언스 연구소	/@castlefive		
/@dimension-value	8 수연	수연의 브런치입니다.	/@xy3on		
/@dimension-value	9 이민규	대기업부터 스타트업까지, 무한경쟁에서 ROI를	/@milaresoltime		
/@dimension-value	10 정동우	우물 밖 개구리, 세상을 빗대어 사람을 봅니다.	/@jucie500		
/@dimension-value	11 마치	거북이와 요구르트를 좋아합니다.	/@thj0228		
/@dimension-value	12 매티카	스타트업, 여행, 커피, 재즈를 좋아하는 20대 작	/@minwookang		
/@dimension-value	13 이규봉	이규봉의 브런치입니다.	/@lee9bong		
/@dimension-value	14 손정웅	손정웅의 브런치입니다.	/@eviljiyo		
/@dimension-value	15 유민서	유민서의 브런치입니다.	/@canadi1029		
/@dimension-value	16 heren	에디터 큐레이터 읽고 쓰고 맛보고 즐기고	/@brunchkyif		
/@dimension-value	17 이광준	이광준의 브런치입니다.	/@hdec99		
/@dimension-value	18 한재원	IT 스타트업 회사 '당근마켓'의 마케터 한재원입니다	/@onea		
/@dimension-value	19 Sunyoung Melissa Kim	Sunyoung Melissa Kim의 브런치입니다.	/@sunyoungmelissa		
/@dimension-value	20 Data Influencer		/@datainfluencer		
/@dimension-value	21 mdoong	mdoong의 브런치입니다.	/@lampyk		
/@dimension-value	22 박현지	박현지의 브런치입니다.	/@e2julie		
/@dimension-value	23 리노		/@lsdphjy		
/@dimension-value	24 천영훈	천영훈의 브런치입니다.	/@youngddangi		
/@dimension-value	25 JH	JH의 브런치입니다.	/@jihyunhaneqpw		
/@dimension-value	26 여니맘	여니맘의 브런치입니다.	/@einie98		
/@dimension-value	27 yoomyoom	yoomyoom의 브런치입니다.	/@winty33		
/@dimension-value	28 박결	박결의 브런치입니다.	/@gyul611		
/@dimension-value	29 버건디	버건디 공간	/@classicburgundy		



웹 크롤링 : 개인별 프로필

https://brunch.co.kr/@dimension-value

강원양
데이터시각화 에디터

구독자 262 관심작가 63

제안하기 **구독하기** ⋮

작가소개 글 61 작품 3

소개

데이터를 다루고 데이터 시각화를 이야기합니다.

콘텐츠 브랜딩 IT **데이터** 기획자 크리에이터 데이터시각화

기타 이력 및 포트폴리오

데이터 수집, 정제, 분석, 시각화에 이르는 '데이터 시각화'의 과정, 그 가운데 배움과 경험, 제작을 하고 있습니다.

현재 데이터 시각화 전문 기업 뉴스젤리 브랜딩팀 소속으로 일하고 있으며, 페이스북 그룹 <데이터 시각화의 모든 것>에서 활동하고 있습니다.



웹 크롤링 : 개인별 프로필 데이터 수집 결과

link	name	company	job	sub_count	follow_count	post	megazine	desc	tag					
/@ivoryoh0609	스텔라			0	9	0	0							
/@hancastellan	castellan			0	10	0	0							
/@kyeongmimfu8	KyeongMi			0	27	0	0							
/@reay1212	양인환			0	32	0	0							
/@gurapark	박구라			1	120	0	0							
/@happynaa	김은아			1	91	0	0							
/@storyfolio	joo			1	100	0	0							
/@castlefive	허성오		개발자	263	25	3	5	데이터 사이언스 연구소	IT,자기계발,일,개발자,연구자,강연자					
/@xy3on	수연			0	18	0	0							
/@millaresoltimi	이민규	디지털콘텐츠기획자		194	35	20	7	대기업부터 스타트업까지, 무한경쟁 마케팅,콘텐츠,브랜딩,기획자,마케터,강사,디지털콘텐츠마케팅						
/@jucie500	정동우			1	257	0	0							
/@thj0228	마지			0	6	0	0							
/@minwookang	매튜강			0	24	0	0							
/@lee9bong	이규봉			0	6	0	0							
/@eviljiyo	손정웅			0	8	0	0							
/@canadi1029	유민서			0	23	0	0							
/@brunchyif	heren			0	26	0	0							
/@hdecalf99	이광준			28	2851	0	0							
/@onea	한재원	당근마켓	마케터	31	12	14	3	IT 스타트업 회사 '당근마켓'의 마케팅,콘텐츠,글쓰기,마케터,기획자,일러스트레이터,당근마켓						
/@sunyoungmelissa	Sunyoung			0	54	0	0							
/@datainfluencer	Data Infl			0	24	0	0							
/@lampyk	mdoong			0	50	0	0							
/@e2julie	박현지			0	3	0	0							
/@lsdphjy	리노			0	38	0	0							
/@youngddangi	천영훈			0	21	0	0							

데이터 시각화 기반의 인사이트 도출 with Tableau

나를 구독하는 259명은 어떤 사람일까?

A decorative graphic on the right side of the slide, consisting of several concentric circles of varying line weights. A small, bright starburst or spark is positioned near the top of the circles.

데이터 시각화 기반의 인사이트 도출 with Tableau

연결 추가

kang_sub_data
Microsoft Excel

kang_sub_profile
Microsoft Excel

시트

Sheet1

새 유니온

kang_구독자 데이터

필터 0 추가

Sheet1

Sheet11

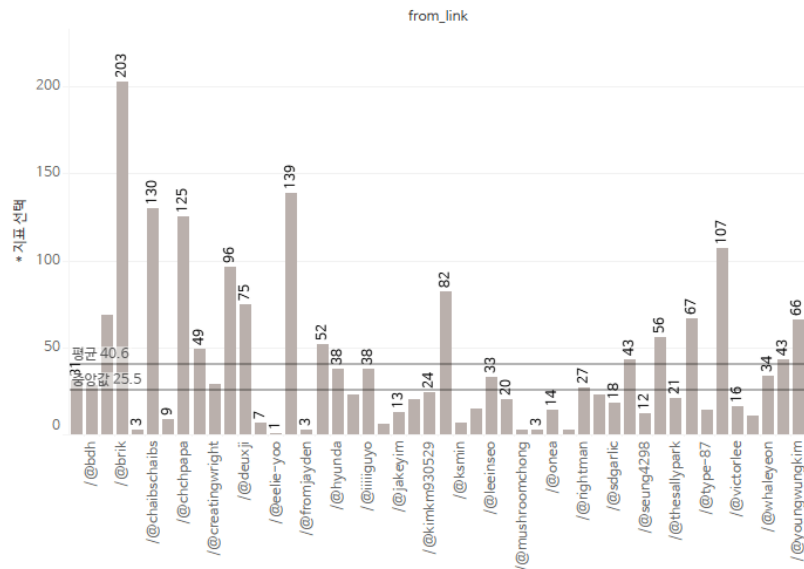
필드 정렬 데이터 원본 순서

☒ 별칭 표시 ☐ 숨겨진 필드 표시 259 행

to_link	index	from_name	from_desc	from_link	link	name	company	job	sub_count	follow_count	post	megazine	desc
/@dimension-value	0	스텔라	데이터분석을 공부합니...	/@lvoryoh0609	/@lvoryoh0609	스텔라	null	null	0	9	0	0	null
/@dimension-value	1	castellan	castellan의 브런치입니...	/@hancastellan	/@hancastellan	castellan	null	null	0	10	0	0	null
/@dimension-value	2	KyeongMi	KyeongMi의 브런치입...	/@kyeongmimfu8	/@kyeongmimfu8	KyeongMi	null	null	0	27	0	0	null
/@dimension-value	3	양인환	양인환의 브런치입니...	/@reay1212	/@reay1212	양인환	null	null	0	32	0	0	null
/@dimension-value	4	박구라	배우는 법을 배워가는 ...	/@gurapark	/@gurapark	박구라	null	null	1	120	0	0	null
/@dimension-value	5	김은아	김은아의 브런치입니...	/@happynaa	/@happynaa	김은아	null	null	1	91	0	0	null
/@dimension-value	6	joo	HE.StoryFolio	/@storyfolio	/@storyfolio	joo	null	null	1	100	0	0	null
/@dimension-value	7	허성오	데이터 사이언스 연구소	/@castlefive	/@castlefive	허성오	null	개발자	263	25	3	5	데이터 사이언스 연구소
/@dimension-value	8	수연	수연의 브런치입니...	/@xy3on	/@xy3on	수연	null	null	0	18	0	0	null
/@dimension-value	9	이만규	대기업부터 스타트업까...	/@milaresoltimi	/@milaresoltimi	이만규	디지털콘텐츠마케팅	기획자	194	35	20	7	대기업부터 스타트업까...

데이터 시각화 기반의 인사이트 도출 with Tableau

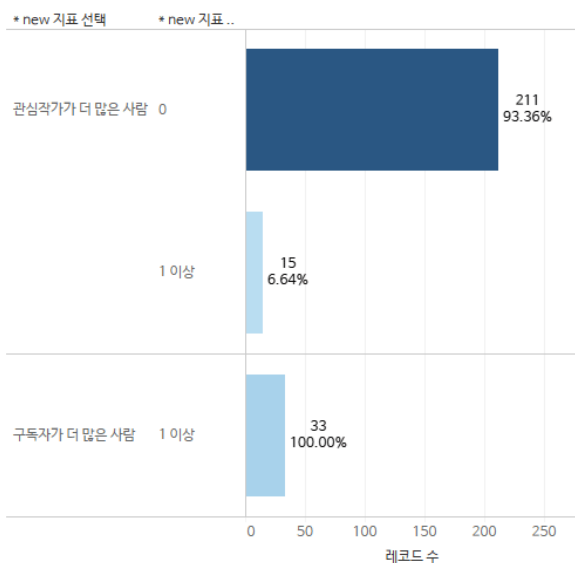
1. 구독자 특징 파악하기 - 단순 지표



지표 선택

글 수
구독자 수
관심객가 수
글 수
매거진 수

2. 구독자 특징 파악하기 - 파생변수 비교



new 지표 선택

구독자 수 vs 관심객가 수

new 지표 선택-2

글 0 vs else

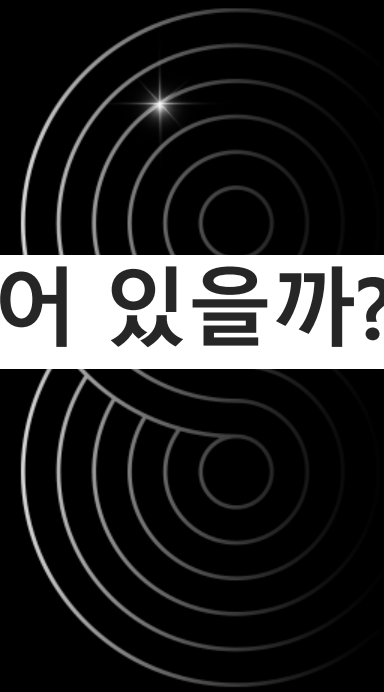
데이터 시각화 기반의 인사이트 도출 with Tableau

- **평균 구독자 수 90명**
 - 구독자가 0명인 사람이 129명, 이를 제외하면 평균 179명의 구독자 보유
- **평균 관심작가 수 135명**
 - 관심작가가 1000명 이상인 사람을 제외하면 평균 88명의 관심작가를 표시
- **평균 글 수 8개**
 - 글을 1개 이상 발행한 사람은 48명 뿐, 이들의 평균 글 수는 41개
- **평균 매거진 수 0개**
 - 매거진이 0개인 사람을 제외하면 평균 4개의 매거진 보유



데이터 시각화 기반의 인사이트 도출 with Python

나와 구독자 259명은 어떻게 연결되어 있을까?

A decorative graphic on the right side of the slide, consisting of several concentric circles of varying line weights. A small, bright starburst or spark is positioned near the top of the circles.

데이터 시각화 기반의 인사이트 도출 with Python

```
In [1]: 1 import pandas as pd
```

```
In [7]: 1 data = pd.read_excel('./data/kang_
```

```
Out [7]:
```

	from_link	index	to_name	
0	/@ivoryoh0609	0	취준생LAB	E
1	/@ivoryoh0609	1	강원영	
2	/@ivoryoh0609	2	Kim Keala Eddy	E
3	/@ivoryoh0609	3	벚그림	
4	/@ivoryoh0609	4	이영민	

```
In [8]: 1 from_list = data['from_link']
```

```
2 to_list = data['to_link']
```

```
In [9]: 1 for word1, word2 in zip(from_list,
```

```
2 print(word1, word2)
```

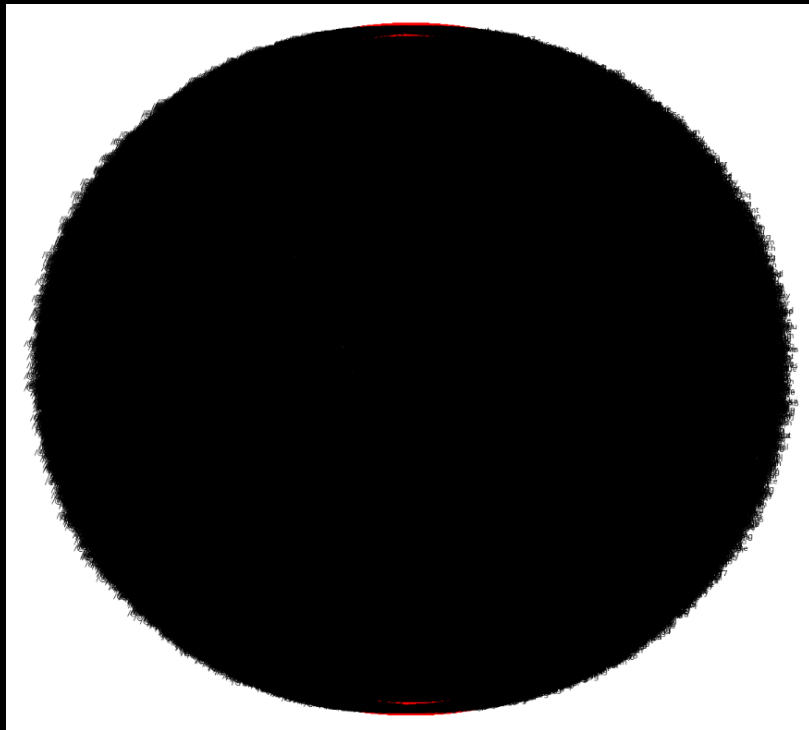
```
/@ivoryoh0609 /@jbsseekerlab  
/@ivoryoh0609 /@dimension-value  
/@ivoryoh0609 /@hero4earth  
/@ivoryoh0609 /@bdh  
/@ivoryoh0609 /@mapthecity  
/@ivoryoh0609 /@gimesilver  
/@ivoryoh0609 /@cloud09  
/@ivoryoh0609 /@bonfire  
/@ivoryoh0609 /@brunch  
/@hancastellan /@llicecoding  
/@hancastellan /@dimension-value  
/@hancastellan /@dryshin  
/@hancastellan /@bekdum  
/@hancastellan /@leeeesh  
/@hancastellan /@reinitiate  
/@hancastellan /@holidaymemories  
/@hancastellan /@ryhyrk  
/@hancastellan /@c4u  
/@hancastellan /@brunch  
/@kyeongimifus /@dimension-value
```

```
In [10]: 1 import networkx as nx  
2 import matplotlib.pyplot as plt  
3 import sys
```

```
In [11]: 1 E = nx.Graph()  
2  
3 for word1, word2 in zip(from_list, to_list):  
4     E.add_edge(word1, word2)
```

```
In [12]: 1 # 한글 표시되도록 폰트 조정  
2 if sys.platform in ["win32", "win64"]:  
3     font_name = "malgun gothic"  
4 elif sys.platform == "darwin":  
5     font_name = "AppleGothic"  
6  
7 plt.figure(figsize=(20,20))  
8 plt.axis("off")  
9  
10 G = nx.Graph(E) # 일반적인 그래프. 모든 선에 대해 선을 이어줌  
11 # G = nx.minimum_spanning_tree(E) # Minimum spanning tree 선 굵기는 것 최소화 해서 연결을 이어줌  
12  
13 nx.draw_circular(G,  
14     font_family=font_name,  
15     node_color="red",  
16     label_pos=0, #0=head, 0.5=center, 1=tail  
17     with_labels=True,  
18     font_size = 12)  
19  
20 # plt.show()  
21  
22 # 이미지 파일 저장한다  
23 plt.savefig("샘플.png")  
24
```

데이터 시각화 기반의 인사이트 도출 with Python



데이터 시각화 기반의 인사이트 도출 with Python

나와 구독자 중 48명(글 1개 이상 발행)은
어떻게 연결되어 있을까?



데이터 시각화 기반의 인사이트 도출 with Python

```
57895      /@soks90      /@brunchkxco      /@soks90/@brunchkxco
57896      /@imagineer    /@brunchkxco      /@imagineer/@brunchkxco
57897      /@prorange     /@brunchkxco      /@prorange/@brunchkxco
57898      /@eastgo        /@brunchkxco      /@eastgo/@brunchkxco
57899      /@steven        /@brunchkxco      /@steven/@brunchkxco
57900      /@kangsunseng    /@brunchkxco      /@kangsunseng/@brunchkxco
57901      /@hoonyqqq      /@brunchkxco      /@hoonyqqq/@brunchkxco
57902      /@brunch        /@brunchkxco      /@brunch/@brunchkxco
```

57903 rows × 3 columns

```
In [307]: 1 cd = (total_data['join'] == total_data['join'][10])
          2 total_data[cd]
```

Out [307]:

	to_link	from_link	join
10	/@dimension-value	/@jucie500	/@dimension-value/@jucie500

```
In [295]: 1 total_data = total_data.reset_index(drop=True)
          2 total_data = total_data.drop_duplicates('join')
```

```
In [311]: 1 df = pd.DataFrame(total_data)
          2 df.columns = ['to_link', 'from_link', 'join']
          3 df.to_excel("total_network_data.xlsx", index=True)
```

```
In [316]: 1 import itertools
          2
          3 case = list(map(''.join, itertools.permutations(id_list, 2)))
          4 case
          5
          6 /@sangmileev7qm/@t1poont004',
          7 '/@sangmileev7qm/@hyunda',
          8 '/@sangmileev7qm/@sentally',
          9 '/@sangmileev7qm/@mushroomchong',
         10 '/@sangmileev7qm/@victorlee',
         11 '/@sangmileev7qm/@thesallypark',
         12 '/@sangmileev7qm/@latebeaver',
         13 '/@sangmileev7qm/@whaleyeon',
         14 '/@sangmileev7qm/@visualoft',
         15 '/@sangmileev7qm/@dimension-value',
         16 '/@deuxji/@castlefive',
         17 '/@deuxji/@milaresoltimi',
         18 '/@deuxji/@onea',
         19 '/@deuxji/@youngwungkim',
         20 '/@deuxji/@bdh',
         21 '/@deuxji/@parkean',
         22 '/@deuxji/@yongjinjinipn',
         23 '/@deuxji/@rightman',
         24 '/@deuxji/@binjino',
         25 '/@deuxji/@gimmesilver',
         26 '/@deuxji/@dew2002'
```

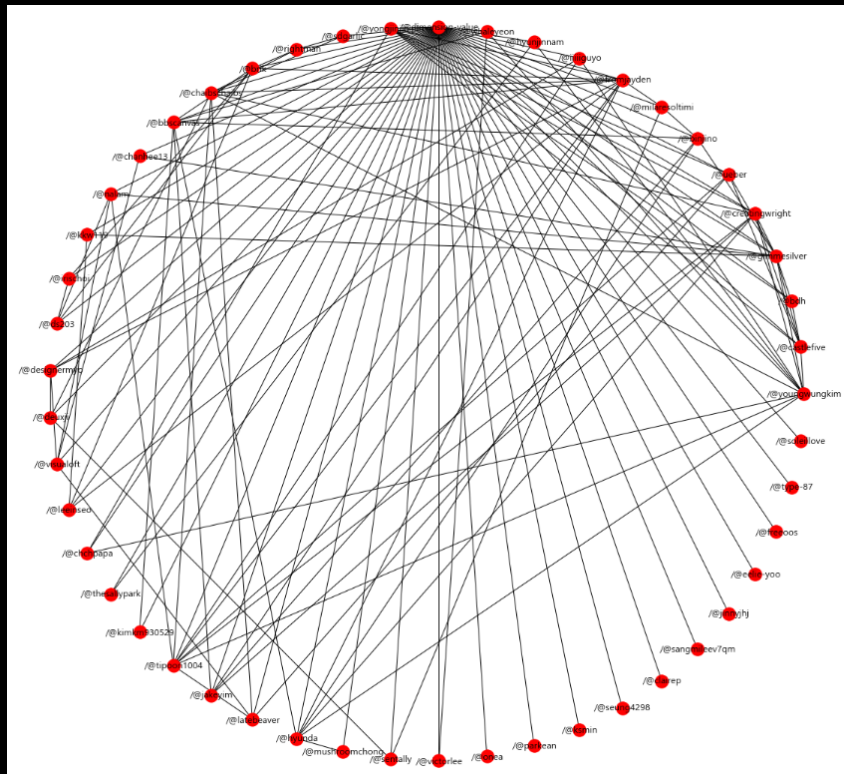
```
In [320]: 1 cd1 = (total_data['join'] == case[2351])
          2 total_data[cd1]
```

Out [320]:

	to_link	from_link	join
258	/@dimension-value	/@visualoft	/@dimension-value/@visualoft

데이터 시각화 기반의 인사이트 도출 with Python

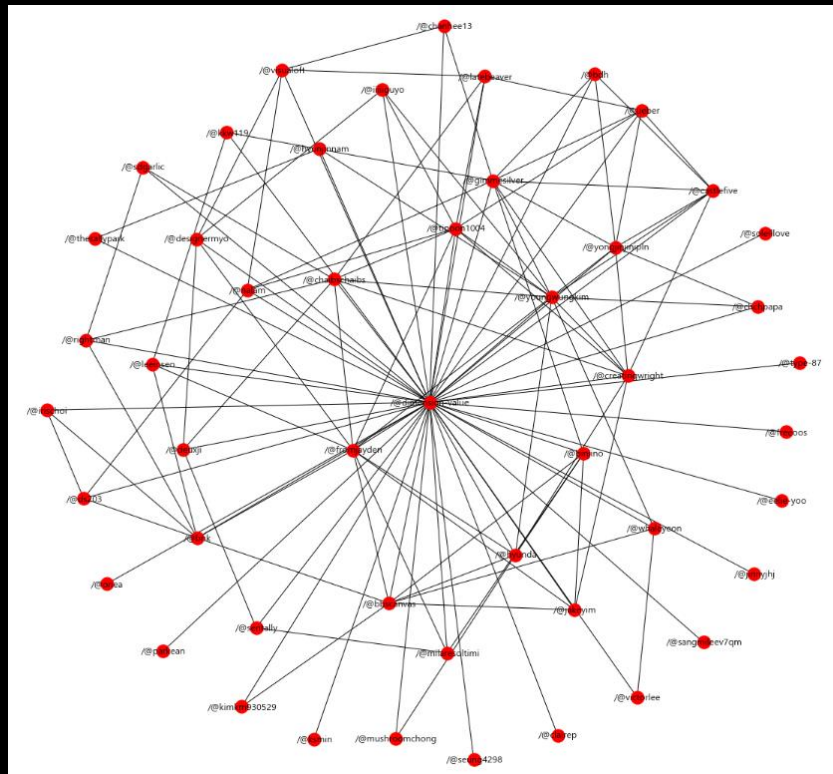
```
In [375]: 1 # 한글 표시되도록 폰트 조정
2 if sys.platform in ["win32", "win64"]:
3     font_name = "malgun gothic"
4 elif sys.platform == "darwin":
5     font_name = "AppleGothic"
6
7 plt.figure(figsize=(20,20))
8 plt.axis("off")
9
10 G = nx.Graph(E) # 일반적인 그래프. 모든 선에 대해 선을 이어줌
11 #G = nx.minimum_spanning_tree(E) # Minimum_spanning_tree 선 긋는 것 최소화 해서 전체를 이어줌
12
13 nx.draw_circular(G,
14                 font_family=font_name,
15                 node_color="red",
16                 label_pos=0, #0=head, 0.5=center, 1=tail
17                 with_labels=True,
18                 font_size = 12)
19
20 # plt.show()
```



데이터 시각화 기반의 인사이트 도출 with Python

```
In [378]: 1 # 한글 표시되도록 폰트 조정
2 if sys.platform in ["win32", "win64"]:
3     font_name = "malgun gothic"
4 elif sys.platform == "darwin":
5     font_name = "AppleGothic"
6
7 plt.figure(figsize=(20,20))
8 plt.axis("off")
9
10 G = nx.Graph(E) # 일반적인 그래프. 모든 선에 대해 선을 이어줌
11 #G = nx.minimum_spanning_tree(E) # Minimum_spanning_tree 선 긋는 것 최소화 해서 전체를 이어줌
12
13 nx.draw_kamada_kawai(G,
14     font_family=font_name,
15     node_color="red",
16     label_pos=0, #0=head, 0.5=center, 1=tail
17     with_labels=True,
18     font_size = 12)
19
20 # plt.show()
```

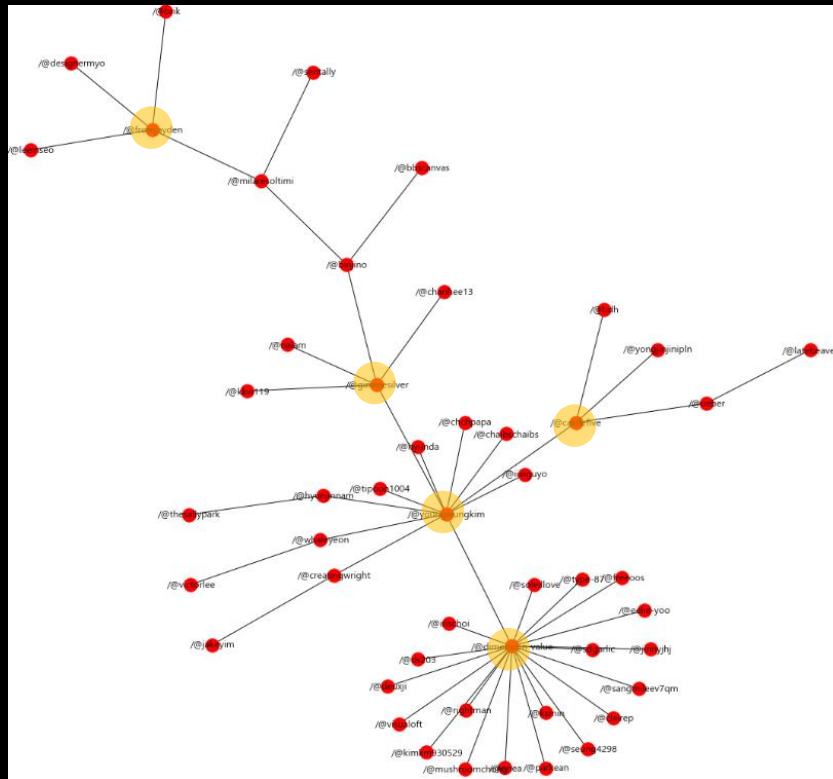
데이터 시각화 기반의 인사이트 도출 with Python



데이터 시각화 기반의 인사이트 도출 with Python

```
In [377]: 1 # 한글 표시되도록 폰트 조정
2 if sys.platform in ["win32", "win64"]:
3     font_name = "malgun gothic"
4 elif sys.platform == "darwin":
5     font_name = "AppleGothic"
6
7 plt.figure(figsize=(20,20))
8 plt.axis("off")
9
10 #G = nx.Graph(E) # 일반적인 그래프. 모든 선에 대해 선을 이어줌
11 G = nx.minimum_spanning_tree(E) # Minimum_spanning_tree 선 긋는 것 최소화 해서 전체를 이어줌
12
13 nx.draw_kamada_kawai(G,
14     font_family=font_name,
15     node_color="red",
16     label_pos=0, #0=head, 0.5=center, 1=tail
17     with_labels=True,
18     font_size = 14)
19
20 # plt.show()
```

데이터 시각화 기반의 인사이트 도출 with Python



데이터 시각화 기반의 인사이트 도출 with Python

The image displays four user profiles from the Brunch platform, arranged in a 2x2 grid. Each profile includes a profile picture, name, role, subscriber count, and post count. The profiles are as follows:

- 김영웅** (Kim Young-ung): 기획자 (Planner), 1,143 subscribers, 298 posts. Bio: 데이터와 기획 그릇을 좋아하는 경영학 박사과정생입니다. 기술교육에서 피드백과 인터랙션이 효과적으로 이루어지는 방법을 고민합니다. 관심사는 다양성 기반의 커뮤니케이션 전략입니다. Tags: IT, 스타트업, 경영, 기획자, 언택트.
- Jayden** (Jayden): 디자이너 (Designer), 113 subscribers, 705 posts. Bio: 도쿄에 살고 있으며, 스타트업에 디자이너, 프론트엔드 엔지니어입니다. 일본에서 재팬에서 느끼는 부분을 공유합니다. Tags: 재학생, 스타트업, 디자이너, 웹-디.
- 허성오** (Heung-seo): 개발자 (Developer), 271 subscribers, 29 posts. Bio: 데이터 사이언스 연구소. Tags: IT, 자기개발, 웹, 게임자, 연구자, 언택트.
- gimmesilver** (gimmesilver): 831 subscribers, 18 posts. Bio: 게임 데이터 분석을 업무로 하고 있습니다.



데이터 시각화 기반의 인사이트 도출 with Python

우리는 비슷한 사람일까?



데이터 시각화 기반의 인사이트 도출 with Python



▲ 나를 구독한 사람 = 구독자



▲ 나 + 구독자(259명)이 구독한 관심작가

◀ 내가 구독한 사람 = 관심작가

회사 밖의 데이터

04 프로젝트 후기



SOSCON 2019

SAMSUNG OPEN SOURCE CONFERENCE 2019

회사 밖의 데이터 프로젝트를 통해 얻은 결과

- 데이터를 활용해 궁금증을 직접 해결하기
- 현실적인 한계를 극복하기 위한 시도 -> 커뮤니티를 통한 학습과 성장
- 데이터 관련 컨퍼런스 등에서 콘텐츠로
데이터 분석 과정에서 시각화의 역할 이야기하기



THANK YOU

SOSCON 2019

SAMSUNG OPEN SOURCE CONFERENCE 2019

