

산학협력캡스톤설계1

SNS 정보를 활용한 소상공인 발전 가능성 평가 서비스

1조 중간 발표



2018204035 한정수



2018204036 윤서안



2018204046 조민경

INDEX

1

연구 주제 주제 변경 이유 및 소개

2

팀 소개 팀원 별 역할

3

진행 상황 현재까지의 상황 및 중간 결과물

4

진행 계획 예상 일정 및 예상 결과물

1

연구 주제

주제 변경 이유 및 소개

원 주제

SNS 정보를 활용한
개인 신용 평가 서비스



주제 변경

SNS 정보를 활용한
소상공인 발전 가능성 평가 서비스

1. SNS만을 통한 개인 정보 수집의 한계

몇 개의 게시물만으로 정확한 성별, 나이, 직업 등의 기본 정보를 얻기가 어려움

개인 SNS 특성상 크게 부정적인 내용의 게시물은 잘 업로드 되지 않으며, 삭제와 수정이 빈번하기 때문에 꾸준한 분석이 불가능

2. 신용 평가 관련 데이터를 구할 수 없음

신용 평가 데이터 없이 SNS 데이터만을 통해 평가 기준을 선정하기 어려움

원래 다른 외부 데이터를 참고하려 했지만 크롤링한 SNS 데이터와 성격이 달라 활용이 어려움

따라서 주제를 변경하기로 결정하고 SNS 사용자들을 다음과 같이 **개인**과 **소상공인**으로 나누었음

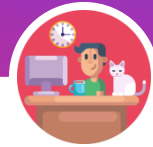
개인의 경우 여전히 앞의 문제들을 해결하기 어렵고 규모가 큰 회사들의 경우 인터넷 검색을 통해 충분히 파악이 가능하다고 판단

→ 소상공인의 데이터를 수집하고 활용하는 데에 집중하기로 결정



개인

상품 판매가 아닌 자신의 일상 공유 등을
목적으로 SNS 활동을 하는 사람 |



소상공인

상품 판매를 위해 SNS를 운영하는 개인 사업자
또는 작은 규모의 사업자 |

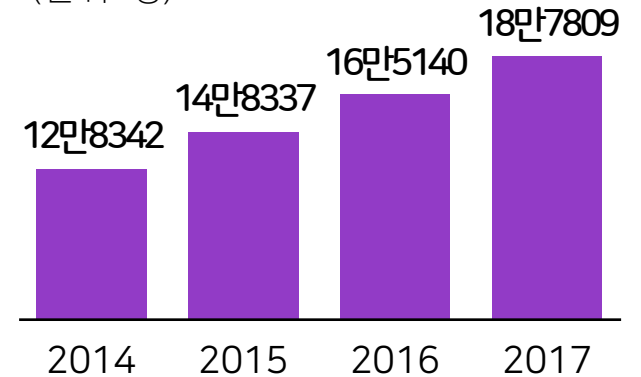
최근 SNS를 통한 광고가 인기를 끌면서 SNS에서 상품을 팔거나 홍보하는 소상공인들이 증가하고 있는 추세

하지만 여전히 자산 규모, 경제적 가치 등 신용도 관련 데이터가 부족

신용 평가를 하는 방향으로 진행하기보다는
SNS를 통해 얻을 수 있는 인지도, 신뢰도 등의 데이터를 활용하여
소상공인들의 현재 활동과 추후 발전 가능성을 평가하려고 함

통신판매업 사업자수 증가 추이

(단위: 명)

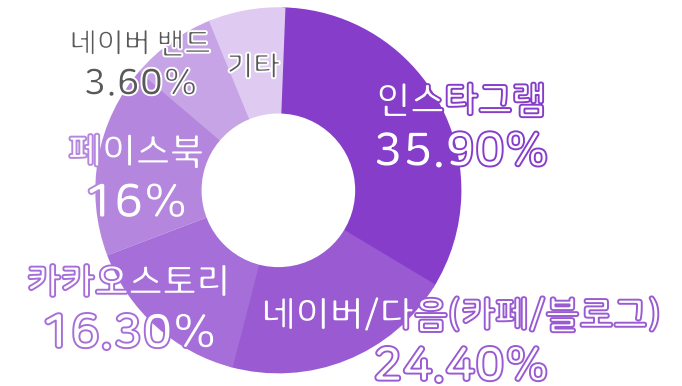


자료: 국세청 국세통계연보

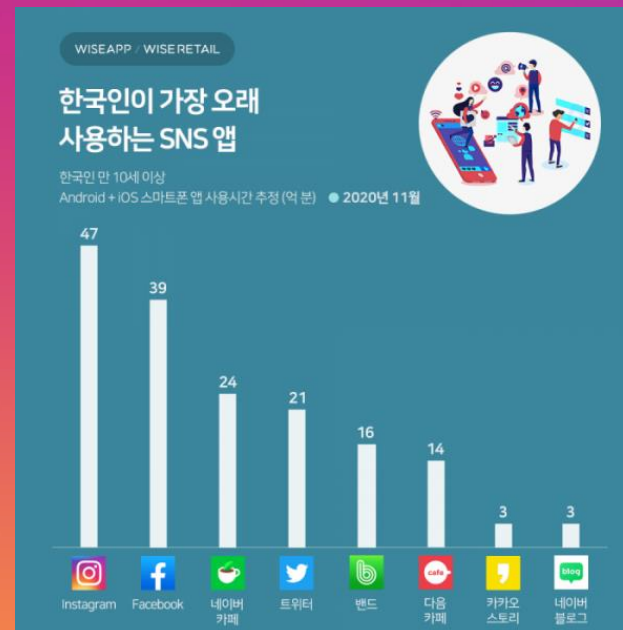
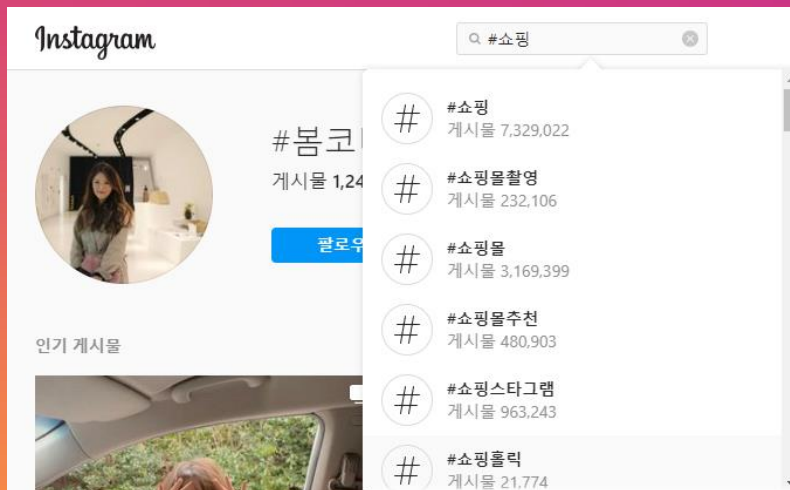
왜 인스타그램을 선택했는가?

1. 해시태그를 이용한 검색이 용이
2. 한국인이 가장 오래 사용하는 SNS 앱 1위를 기록 (지난 11월 기준)
3. 미디어 이용자 수에 관하여 인스타그램이 다양한 연령층의 상위권을 차지
4. SNS 매체 중 홍보 효과가 탁월

SNS 매체 별 쇼핑 경험 추이



자료: 서울시 전자상거래 센터



국내 소셜 미디어 연령별 월평균 이용자 수 단위: 명

	10대	20대	30대	40대	50대
1위	221만	493만	440만	502만	544만명
2위	191만	386만	319만	298만	297만명
3위	86만	178만	268만	266만	177만명

※월 평균 이용자 수는 2020년 1분기(1~3월) 내 월별로 발생한 이용자 수의 산술평균값

자료=DMC미디어

2

팀 소개

팀원 별 역할

팀원 별 역할



한정수 (팀장)

크롤링을 위한 Python 코드 작성
게시물 데이터 관련 코드 유지보수
좋아요 분석용 R 코드 작성
데이터 수집 및 데이터베이스 관리
UI 제작



좋아요 8,888개
1일 전

댓글 달기...

...



윤서안

크롤링을 위한 Python 코드 작성
해시태그 데이터 관련 코드 유지보수
해시태그 분석용 R 코드 작성
데이터 수집 및 데이터베이스 관리
UI 디자인



좋아요 8,888개
1일 전

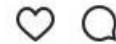
댓글 달기...

...



조민경

크롤링을 위한 Python 코드 작성
프로필 데이터 관련 코드 유지보수
게시글 분류용 R 코드 작성
데이터 시각화
UI 제작



좋아요 8,888개
1일 전

댓글 달기...

...



그 외 휴학생 1명

크롤링을 위한 Python 코드 작성
위치 데이터 관련 코드 유지보수
날짜 분석용 R 코드 작성
데이터 시각화
UI 디자인



좋아요 8,888개
1일 전

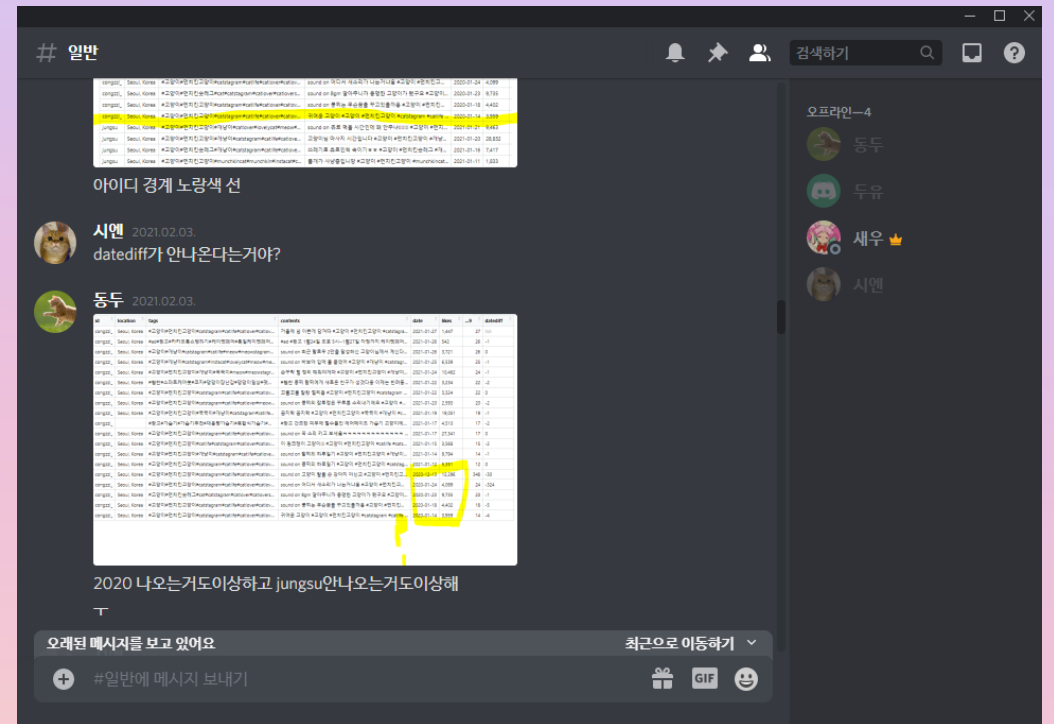
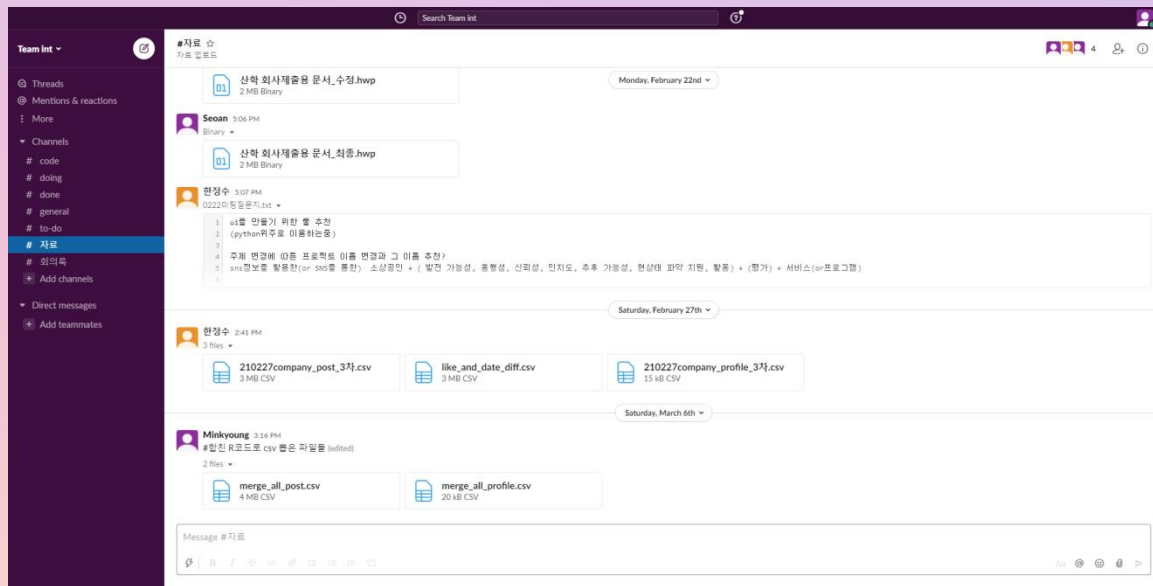
댓글 달기...

...

회의 방식

클라우드 기반 팀 협업 도구인 Slack를 이용하여 자료와 코드 공유

디스코드와 카카오톡을 사용하여 비대면 회의 진행



3

진행 상황

현재까지의 상황 및 중간 결과물

2020년 7월 ~ 8월

산학 프로젝트 수행 계획서 작성
8퍼센트(회사)와의 대면 미팅
자료 공유를 위해 Slack 페이지 생성
사전 조사
방학 동안 2주에 한번씩 회의 진행

2020년 9월 ~ 12월

학기 중(진행 느낌)
인스타그램 크롤링 코드 작성
데이터베이스 생성 후 연결
산학 프로젝트 중간 보고서 작성

현재까지의 진행 일정

2021년 1월

담당 교수님과의 비대면 미팅
주제 변경
주 3회 회의 방식으로 변경
작성했던 코드 대규모 수정
크롤링 데이터 수집

2021년 2월

8퍼센트, 교수님과의 대면 미팅
R을 이용하여 분석을 위한 여러 컬럼들 생성



인스타그램에서 활동 중인 소상공인들의 계정 약 3500개를 랜덤으로 수집

1. 마켓, 쇼핑몰, 패션, 주문제작 등 인스타 마켓과 관련도가 높은 23개의 해시태그로 게시물 검색
2. 해시태그 1개당 약 500개의 최근 게시물 크롤링
3. 게시물을 업로드한 계정이 소상공인 계정이 맞는지 아래와 같이 확인
4. 중복 제거 후 메모장에 저장

큰 회사나 유명인을 거르기 위해  '인증됨' 마크가 있는 계정 제외

이미 성공하여 자리를 잡아 대출 가능성이 별로 없는 계정을 거르기 위 팔로워 수가 100,000명이 넘는 계정 제외

소개글에 문의, 이벤트, 주문, 판매 등 상품 판매 계정이 자주 사용하는 35개의 단어 중 하나를 사용하면 계정 ID 수집

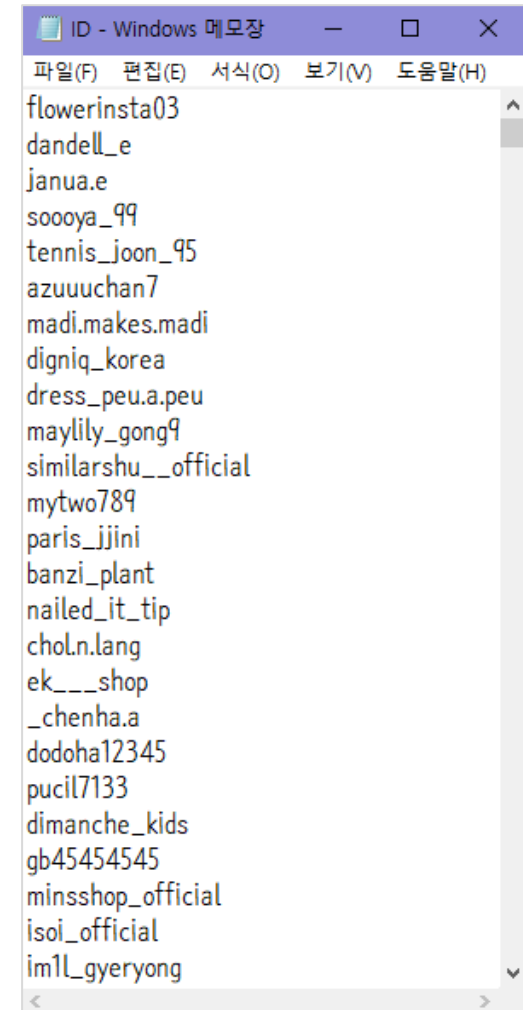
```
def insta_search(driver, search):
    url_tmp = driver.current_url
    url = "https://www.instagram.com/explore/tags/{search}?hl=ko".format(search=search) # 검색어로 찾기
    driver.get(url)
    time.sleep(2)
    insta_postList(driver)
###
def insta_postList(driver):
    url = driver.current_url
    insta_post_urls_list = []
    driver.get(url)
    time.sleep(2)
    # 일정 시간동안 스크롤
    start = datetime.datetime.now()
    end = start + datetime.timedelta(seconds = 100)
    time.sleep(2)
    while True:
        insta_post_urls = driver.find_elements_by_css_selector(".Nnq7C.weEfm.v1Nh3.kIKUG._bz0w a")
        for urlList in insta_post_urls:
            insta_post_urls_list.append(urlList.get_attribute("href"))
        driver.execute_script('window.scrollTo(0, document.body.scrollHeight);')
        time.sleep(2)
        if datetime.datetime.now() > end:
            break
    insta_post_urls_list = list(set(insta_post_urls_list))
    print(len(insta_post_urls_list))
    insta_accounts(driver, insta_post_urls_list)
```

크롤링 코드 일부

```
# 인증 여부
checks = ""
checkList = driver.find_elements_by_css_selector(".lgw0E.lwRSH.eGOV._4EzTm.soMvI")
for check in checkList:
    checks = checks + check.text
#팔로워수 제한
userFollower = userFollower.replace(',', '')
if checks.find('인증됨') == -1 and int(userFollower) <= 100000:
    for word in ['문의', '이벤트', '주문', '공구', '협찬', '판매', '마켓', '블로그', '링크', '세일', '할인', '카톡',
                 '카카오톡', '구매', '구입', '다이렉트', '택배', '배송', '제품', '상품', '스토어', '상점', '입금',
                 '플랫폼', 'DM', '디엠', '가격', '오픈', 'kakaotalk', 'shop', '쇼핑몰', '샵', 'open']:
        if intro.find(word) != -1:
            company.append(userID)
            break
```

소상공인 계정 판단

수집한 instagram 계정



flowerinsta03
dandell_e
janua.e
soooya_99
tennis_oon_95
azuuuchan7
madi.makes.madi
digniq_korea
dress_peu.a.peu
maylily_gong9
similarshu__official
mytwo789
paris_jjini
banzi_plant
nailed_it_tip
choLn.lang
ek__shop
_chenha.a
dodoha12345
pucil7133
dimanche_kids
gb45454545
minsshop_official
isoi_official
im1l_gyeryong

수집한 소상공인 계정을 크롤링

1. 소상공인 계정을 수집한 메모장 파일을 이용
2. 계정의 정보 - 계정 이름, 게시물 수, 팔로워 수, 팔로잉 수, 소개문 등의 데이터를 수집
3. 게시물 정보 - 위치정보, 해시태그, 좋아요 수, 본문 내용 등의 데이터를 수집
4. 수집한 데이터를 DB에 각각 따로 저장



코드작성



데이터베이스와 연결

Profile 계정 기본 정보 크롤링

```

userIDList = driver.find_elements_by_css_selector("_7UhW9.fKfbl.yUEEX.KV-D4.fDxYI")
userID = userIDList[0].text # 아이디

userInfoList = driver.find_elements_by_css_selector(".g47SY")

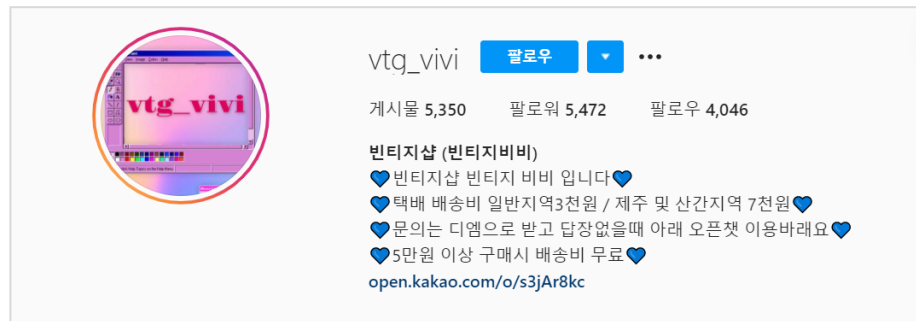
userPost = userInfoList[0] # 게시물 수
userFollower = userInfoList[1] # 팔로워 수
userFollowing = userInfoList[2] # 팔로잉 수

userPostReal = userPost.get_attribute('title')
userFollowerReal = userFollower.get_attribute('title')
userFollowingReal = userFollowing.get_attribute('title')

if userPost.text.find('백') == -1 and userPost.text.find('천') == -1 and userPost.text.find('만') == -1:
    userPost = userPost.text
else:
    userPost = userPostReal
if userFollower.text.find('백') == -1 and userFollower.text.find('천') == -1 and userFollower.text.find('만') == -1:
    userFollower = userFollower.text
else:
    userFollower = userFollowerReal
if userFollowing.text.find('백') == -1 and userFollowing.text.find('천') == -1 and userFollowing.text.find('만') == -1:
    userFollowing = userFollowing.text
else:
    userFollowing = userFollowingReal
# 소개글
intro = ""
try:
    userInfoList = driver.find_elements_by_css_selector("-vDlG span")
    intro = userInfoList[0].text
    intro = demoji.replace(intro, " ")
    print(intro)
except:
    continue

dbData = [[url, userID, userPost, userFollower, userFollowing, intro]]
connectDB1(dbData)

```



Post 해당 계정의 게시물들 크롤링

```

#계정아름
userID=""
userID_temp = driver.find_elements_by_css_selector(".sqd0P.yWX7d._8A5w5.ZlAJV")
userID = userID_temp[0].text

#위치정보 내용
locations = ""
locationsList = driver.find_elements_by_css_selector(".JF9hh")
for location in locationsList:
    locations = locations + location.text
#해시태그 내용
tags = ""
tagList = driver.find_elements_by_css_selector(".xi13i")
for tag in tagList:
    tags = tags + tag.text

#게시글 내용
Contents = ""
Content_List = driver.find_elements_by_css_selector(".eo2As .EtaWk .P8YgZ .04YMK span")
Content_temp = Content_List[1]
Contents = Content_temp.text

for c_tag in tagList:
    Contents = Contents.replace(c_tag.text, "")

#이모티콘 지우는 코드. 위에 import demoji 이용.
Contents = demoji.replace(Contents, " ")

# 게시물 날짜
dates = ""
dateList = driver.find_elements_by_css_selector("_lo9PC.Nzb55")
for date in dateList:
    date = date.get_attribute('title')
    dates = dates + date

```



no	url	id	postnum	followernum	followingnum	intro
1	https://www.instagram.com/vtg_vivi/	vtg_vivi	4,655	5,327	4,090	빈티지샵 빈티지 비비입니다 택배 배송비 일반지역3천원 / 제주 및 산간지역 7천원 ...
2	https://www.instagram.com/_e_room_/	_e_room_	863	571	285	안동 옷가게 〇〇 안동시 경동로 515 / 안동중 건너편
3	https://www.instagram.com/so_eun/	so_eun	222	1,025	592	이가자해어비스 서면삼정타워점 수석 디자이너 매주 일,월 휴무 DM은 답이 느립니다! 빠른 ...
4	https://www.instagram.com/jeongmiyeong580/	jeongmiyeong580	126	74	99	부산 사상구 모라로 46번길 60 (모라시장 57호) 카톡/DM 문의 및 주문가능합니다^^
5	https://www.instagram.com/euncha_official/	euncha_official	61	237	1	온차 소식 받아보세요 스토어점과 특가친구하시고 혜택받아가세요♡
6	https://www.instagram.com/you_and_sara/	you_and_sara	5,032	4,034	3	택배해용 카드*휴대폰결제가능 1호점 나와사라 @you_and_sara 2호점 함께사라 @w...
7	https://www.instagram.com/yemm_ing/	yemm_ing	177	482	2	하단 링크 타고 오시면 구매 가능 하십니다 핑퐁 스토어점 꾸욱 눌러주세요 ye...

company_profile DB

계정의 정보 크롤링

인스타그램 주소, ID, 게시물 수, 팔로워 수, 팔로잉 수, 소개문

no	url	id	location	tags	contents	date	likes
1	https://www.instagram.com/p/CLEWenhgx/	vtg_vivi		#빈티지샵#구제#구제샵#빈티지#구제소핑물#구제옷#울드스쿨#미니멀룩#오늘의코디#남녀공용#만원...		2021년 2월 9일	5
2	https://www.instagram.com/p/CK_Z1T7HhSM/	vtg_vivi		#빈티지비비구매가능#빈티지샵#구제#구제샵#빈티지#구제소핑물#구제옷#울드스쿨#미니멀룩#오늘의...	아디다스 후드티 #빈티지비비구매가능 - 사이즈 : L 가슴 : 60 중...	2021년 2월 7일	2
3	https://www.instagram.com/p/CLEZPtoHb5G/	vtg_vivi		#빈티지비비구매가능#빈티지샵#구제#구제샵#빈티지#구제소핑물#구제옷#울드스쿨#미니멀룩#오늘의...	폴로 체크 셔츠 #빈티지비비구매가능 - 사이즈 : L 가슴 : 64 중...	2021년 2월 9일	3
4	https://www.instagram.com/p/CK_aybUHm5p/	vtg_vivi		#빈티지비비구매가능#빈티지샵#구제#구제샵#빈티지#구제소핑물#구제옷#울드스쿨#미니멀룩#오늘의...	칼글 후드티 #빈티지비비구매가능 - 사이즈 : S 가슴 : 54 중자...	2021년 2월 7일	3
5	https://www.instagram.com/p/CK6CnW4nNP/	vtg_vivi		#빈티지비비구매가능#빈티지샵#구제#구제샵#빈티지#구제소핑물#구제옷#울드스쿨#미니멀룩#오늘의...	챔피언 후드티 #빈티지비비구매가능 - 사이즈 : L 가슴 : 63 중< ...	2021년 2월 5일	4
6	https://www.instagram.com/p/CK68gQ-Hv_L/	vtg_vivi		#빈티지샵#구제#구제샵#빈티지#구제소핑물#구제옷#울드스쿨#미니멀룩#오늘의코디#남녀공용#만원...		2021년 2월 5일	4
7	https://www.instagram.com/p/CK6Ci02nNmF/	vtg_vivi		#빈티지비비구매가능#빈티지샵#구제#구제샵#빈티지#구제소핑물#구제옷#울드스쿨#미니멀룩#오늘의...	아디다스 맨투맨 #빈티지비비구매가능 - 사이즈 : L 가슴 :58 중< ...	2021년 2월 5일	10
8	https://www.instagram.com/p/CK_aGeEnw-1/	vtg_vivi		#빈티지비비구매가능#빈티지샵#구제#구제샵#빈티지#구제소핑물#구제옷#울드스쿨#미니멀룩#오늘의...	타미힐피거 스물로고 맨투맨 #빈티지비비구매가능 - 사이즈 : S 가슴 : ...	2021년 2월 7일	2
9	https://www.instagram.com/p/CLEXv-EHDY7/	vtg_vivi		#빈티지비비구매가능#빈티지샵#구제#구제샵#빈티지#구제소핑물#구제옷#울드스쿨#미니멀룩#오늘의...	아디다스 빅 트레포일 크롭 트랙탑 자켓 #빈티지비비구매가능 - 사이즈 : S < ...	2021년 2월 9일	4

company_post DB

계정의 게시물 크롤링

게시물 주소, 작성 ID, 위치, 해시태그, 내용, 작성 날짜, 좋아요 수

R을 이용하여 분석을 위한 새로운 컬럼 추가

1. DB의 데이터를 csv 파일로 저장하여 R 코드 작성

2. Profile.csv 를 이용하여 새로운 컬럼 추가

판매 게시물 수, 일상 게시물 수, 일주일 동안의 좋아요 평균, 전체 좋아요 평균 등 다수의 컬럼 생성

3. Post.csv 를 이용하여 새로운 컬럼 추가

게시물의 성격(판매 글 or 일상 글), 게시물 작성 빈도, 좋아요 차이, 해시태그 개수 컬럼 생성

새로운 컬럼 추가 - profile

판매 게시물 수 일상 게시물 수 일주일 동안의 좋아요 평균 전체 좋아요 평균 위치 정보 기재 횟수
 게시물 업로드 주기 한 게시물당 평균 해시태그 사용 횟수 사용한 전체 해시태그 가장 많이 사용한 해시태그
 가장 많이 사용한 해시태그를 사용한 횟수

sellnum	dailynum	like_mean_7	like_mean_all	location_num	hashtag_average	post_upload	all_tags	max_tags	max_tags_num
106	2	24.4666666666667	29.787037037037	1	15.99074074	1.20833333333333	#안동옷가게#안동이름#디지니#니트#미키마우스#맨투맨#...	#안동이름	83
73	0	91.8333333333333	102.219178082192	0	27.38356164	2.14705882352941	#####	#####	245
36	0	271.666666666667	145.333333333333	36	10.83333333	10.5	#잠옷#잠옷세트#공주#귀여워#홈웨어#파자마파티#파자마...	#잠옷#파자마	27
52	0	235	267.846153846154	52	0.55769231	6.16326530612245	#선팔#맞팔#좋아요#좋아요반사#중반#데일리#팔로우#like...	#선팔#맞팔#좋아요#좋아요반사#중반#데일리#팔로우#like...	1
26	0	146	340.884615384615	24	0.65384615	6.48	#studionicholson#moonstar#내돈내산#스테디에브리웨어#...	#제품협찬	5
39	1	259.857142857143	233.35	0	2.07500000	26.5185185185185	#청주#청주옷가게#청주지하상가#대전지하상가#대전맛집...	#청주#청주지하상가#대전지하상가#데이트룩#데이트룩코...	3
107	1	20.75	19.5925925925926	0	9.71296296	3.42452830188679	#이코토이#2COTOY#빈티지토이#블로그마켓#화요마켓#케...	#2COTOY	94
72	36	15.5	24.9166666666667	0	9.06481481	7.1010101010101	#카페아마도#cafeamado#등촌동카페#등촌카페#등촌역카...	#카페아마도	104
21	0	4.6	6.61904761904762	0	21.52380952	2.08333333333333	#나이키#아웃스타그램#여자옷코디#할공점퍼#패딩#맨투맨#...	#####	65
107	0	9	13.1588785046729	0	1.90654206	2.325	#로토케이프CT#OT#Free#밍크#을90#캐시10#큐빅레이스...	#####	11
59	0	14.2	26.2372881355932	0	22.25423729	9.17857142857143	#####	#####	264
68	40	8.1666666666667	11.4259259259259	0	18.51851852	1.46875	#배색가디건#이중지니트#도톰한소재#울카디건#겨울신상...	#스토#좋아요	98
103	5	74.7142857142857	79.4444444444444	0	30.01851852	1.15463917525773	#플라워케이크#부엌떡케이크#디저트#베이킹#홈베이킹#h...	#디저트	120
108	0	16.8333333333333	13.5555555555556	101	0.70370370	1.15151515151515	#유니크#유니크스타일#보블리에#보블리에######	#보블리에	20
53	0	45.1666666666667	39.7924528301887	2	1.64150943	13.5294117647059	#soundcloud#goodnightjazz#inspiration#byaube#inspiratio...	#byaube	14
107	1	15	12.8518518518519	0	25.04629630	7.30555555555556	#삐로로#수원카페#곡반정동카페#수원아이싱쿠키#수원수...	#sweet	107
107	0	11.9411764705882	13.9345794392523	0	12.78504673	1.47727272727273	#####	#광주옷가게#풍암동#루시아#시아스토리	76
86	0	63.3333333333333	66.4418604651163	85	25.87209302	4.28395061728395	#チャノキカナザワ#チャノキ#シュクリ?ム#タピオカ#石...	#金?カフェ#クレ?フ?	85

Profile R 코드 일부

```
###3. 최근 7일간 좋아요 평균, 전체글(100개 이하) 좋아요 평균###
#최근 년도와 그 전년도 12월만 다루기
posting_recent <- subset(post, year(post$date) == max(year(post$date)) |
#view(posting_recent)

#id 별 제일 최신글 = id별 day값이 제일 큰 것들
recently <- posting_recent %>% arrange(id, desc(date)) %>% group_by(id)
recently <- select(recently, id, day)
#view(recently)

#제일 최근글 기준 7일 간의 게시글(숫자 조절하면됨.)
join <- inner_join(posting_recent,recently, by = 'id')
#view(join)

#358인 이유 : 21년 1/1일에 작성한 글의 경우 20년 12월 25일까지이기때문(1
posting_recent <- join[(((join$day.y - join$day.x >= 0) & (join$day.y - j
#view(posting_recent)

#평균구하기
like_mean_7 <- tapply(posting_recent$likes, posting_recent$id, mean)

#전체 줄평균
posting_all <- post[order(post$id, post$date),]
like_mean_all <- tapply(posting_all$likes, posting_all$id, mean)

#평균낸거 다 하나로 붙이기
#id
id_unique <- unique(post$id)
id <-sort(id_unique)
#view(id)

like_mean <- cbind(id,like_mean_7,like_mean_all)
#view(like_mean)
```

###5. 위치정보 갯수 구하기###

```
post2 <-post[post$location != "",]
#view(post2)
```

```
location_num = ddply(post2, c("id"), summarise, location_num = length(location))
#view(location_num)
```

```
profile<- merge(x = profile, y = location_num, by='id', all.x = TRUE)
#view(profile)
```

###6. 계정별 해시태그 사용 평균 갯수 구하기(중복 허용)###

```
temp <- aggregate(hashtag_num~id, post, mean)
names(temp)[2] <- c("hashtag_average")
profile <- merge (x = profile, y = temp, by='id', all.x = TRUE)
#view(profile)
```

###7. 계정별 게시글 올리는 주기 평균 구하기###

```
temp <- post[post$datediff != 0,]
temp <- aggregate(datediff ~ id, temp, mean)
#view(temp)
names(temp)[2] <- c("post_upload")

profile <- merge (x = profile, y = temp, by='id', all.x = TRUE)
view(profile)
```

date	likes	sellingPost	day	datediff	likediff	hashtag_num
2020-12-14	28	판매글	349	0	0	18
2020-12-14	29	판매글	349	0	1	16
2020-12-15	29	판매글	350	1	0	17
2020-12-15	28	판매글	350	0	-1	16
2020-12-16	24	판매글	351	1	-4	15
2020-12-16	30	판매글	351	0	6	17
2020-12-16	23	판매글	351	0	-7	16
2020-12-16	14	판매글	351	0	-9	16
2020-12-16	21	판매글	351	0	7	23
2020-12-17	54	판매글	352	1	33	27
2020-12-17	43	판매글	352	0	-11	23
2020-12-18	65	판매글	353	1	22	9
2020-12-18	48	판매글	353	0	-17	30
2020-12-18	38	판매글	353	0	-10	28
2020-12-19	34	판매글	354	1	-4	27
2020-12-19	68	판매글	354	0	34	5
2020-12-19	49	판매글	354	0	-19	27
2020-12-20	27	판매글	355	1	-22	27

새로운 컬럼 추가 - post

게시물의 성격(판매 글 or 일상 글)

게시글 작성 빈도

좋아요 차이

해시태그 개수

결측치 처리

sellnum	dailynum	like_mean_7	like_mean_all	location_num	hashtag_average	post_upload
106	2	24.4666666666667	29.787037037037	1	15.99074074	1.20833333333333
73	NA	91.8333333333333	102.219178082192	NA	27.38356164	2.14705882352941
36	NA	271.666666666667	145.333333333333	36	10.83333333	10.5
52	NA	235	267.846153846154	52	0.55769231	6.16326530612245
26	NA	146	340.884615384615	24	0.65384615	6.48
39	1	259.857142857143	233.35	NA	2.07500000	26.5185185185185



#결측치 모두 0으로 처리
 post[is.na(post)] <- 0
 profile[is.na(profile)] <- 0

sellnum	dailynum	like_mean_7	like_mean_all	location_num	hashtag_average	post_upload
106	2	24.4666666666667	29.787037037037	1	15.99074074	1.20833333333333
73	0	91.8333333333333	102.219178082192	0	27.38356164	2.14705882352941
36	0	271.666666666667	145.333333333333	36	10.83333333	10.5
52	0	235	267.846153846154	52	0.55769231	6.16326530612245
26	0	146	340.884615384615	24	0.65384615	6.48
39	1	259.857142857143	233.35	0	2.07500000	26.5185185185185

결측치 없음을 확인

```
#결측값 측정
df.isnull().sum()
```

```
id          0
postnum     0
followernum 0
followingnum 0
sellnum     0
dailynum    0
like_mean_7 0
like_mean_all 0
location_num 0
hashtag_average 0
post_upload 0
dtype: int64
```

2021-03-01(월)

1	2021-03-01	https://www.instagram.com/vtg_vivi/	vtg_vivi	5,006	5,330	4,060
2	2021-03-01	https://www.instagram.com/_e_room_/	_e_room_	889	562	320
3	2021-03-01	https://www.instagram.com/so_eun/	so_eun	227	1,018	599
4	2021-03-01	https://www.instagram.com/jeongmiyeong580/	jeongmiyeong580	133	73	101
5	2021-03-01	https://www.instagram.com/euncha_official/	euncha_official	90	342	1

2021-03-05(금)

787	2021-03-05	https://www.instagram.com/vtg_vivi/	vtg_vivi	5,137	5,381	4,057
788	2021-03-05	https://www.instagram.com/_e_room_/	_e_room_	890	568	323
789	2021-03-05	https://www.instagram.com/so_eun/	so_eun	229	1,024	603
790	2021-03-05	https://www.instagram.com/jeongmiyeong580/	jeongmiyeong580	136	76	101
791	2021-03-05	https://www.instagram.com/euncha_official/	euncha_official	95	365	1

2021-03-08(월)

1502	2021-03-08	https://www.instagram.com/vtg_vivi/	vtg_vivi	5,220	5,409	4,053
1503	2021-03-08	https://www.instagram.com/_e_room_/	_e_room_	894	571	284
1504	2021-03-08	https://www.instagram.com/so_eun/	so_eun	230	1,023	601
1505	2021-03-08	https://www.instagram.com/jeongmiyeong580/	jeongmiyeong580	138	78	103
1506	2021-03-08	https://www.instagram.com/euncha_official/	euncha_official	97	402	1

4

진행 계획

예상 일정 및 예상 결과물

산학 연계 프로젝트



데이터



분석 보고서



검색용 코드



그 외 코드

캡스톤 프로젝트

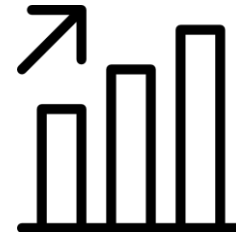
유의미한 factor들을
포함한 정제된 데이터
profile, post csv파일



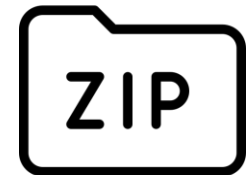
분석을 통해 발견한
인사이트 정리,
데이터 분포 시각화한
분석 보고서



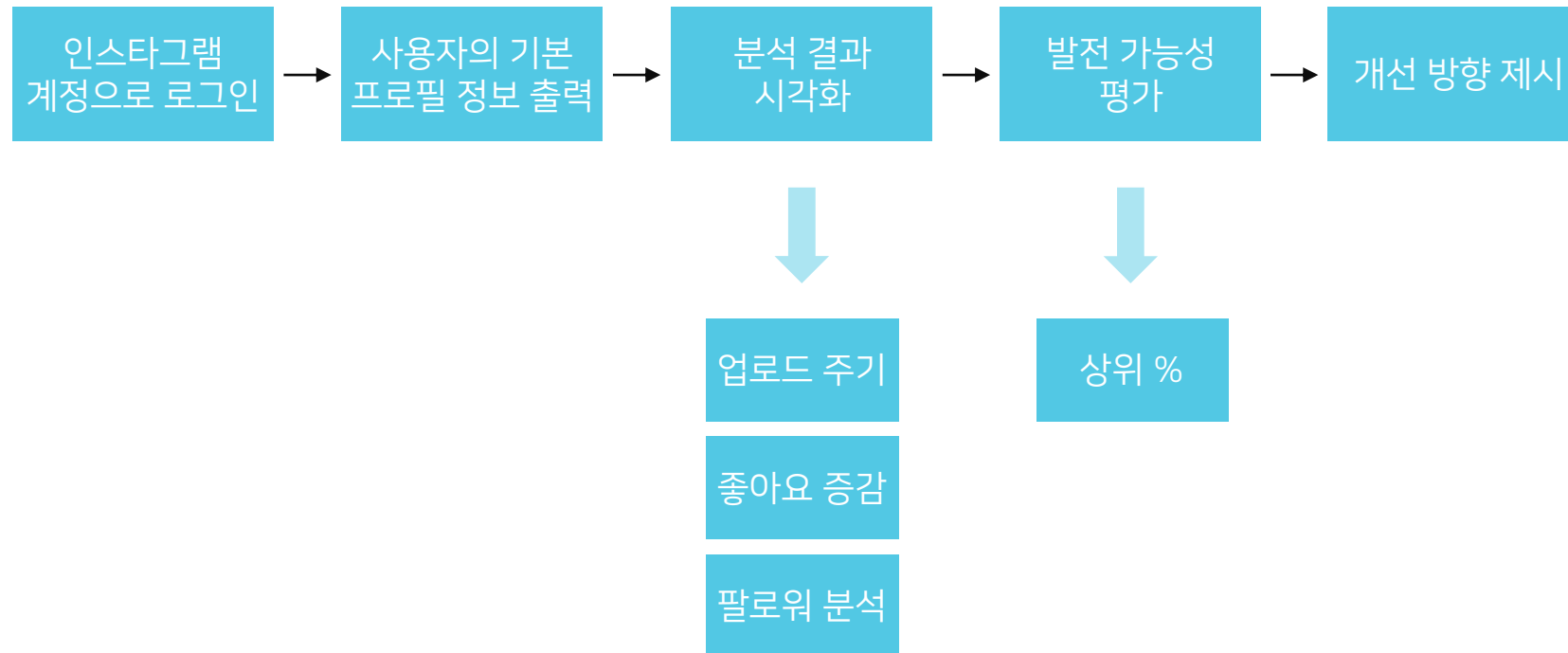
사용자(소상공인)가
자신의 인스타그램
계정으로 로그인하면
분석 결과를 시각적으로
이해하기 쉽게 보여주는
프로그램(UI 제작)



그 외 코드들과
로우 데이터



프로그램 구상도



2021년 3월

최종 데이터 수집 후 전처리
Factors 분석

2021년 5월

코드 정리 및 최종 수정
산학 프로젝트 마무리

2021년 4월

간단한 시각화
특정 데이터를 볼 수 있는 코드 제작(산학 제출용)
분석 보고서 작성(산학 제출용)

2021년 6월 ~ 10월

코드 수정
추가적인 분석과 시각화
UI 제작

앞으로의 진행 일정

Thank You