

고려대학교 빅데이터 연구회

# KU-BIG 빅콘테스트 TEAM 1차 발표



# 팀원소개



**팀장**

**이동빈(26)**

산업경영공학부 14학번

(곧) 삼성맨

"짱구 파자마 입기 딱 좋은 나이지~"



**박수희(23)**

통계학과 16학번

(곧) 시각화 마스터

"쿠디티 직캠 원하면 인스타 팔로우~"



**고유정(22)**

미디어학부 17학번

(곧) 출국하는 휴학생

"목디스크 아빰뻘여~"

제7회  
20

서로 다른 종류의 데이터들을  
전처리/시각화/분석하는 과정을 통해  
인사이트를 얻고, 종합적으로 이해한다.

**환경기상 데이터, 유동인구 데이터,  
유통 데이터, 카드매출 데이터,  
SNS 데이터**

# Python

# 1. 환경기상 데이터(케이웨더)

기간: 180401-190331

종로구	스테이션	측정기 등록일	위치	행정동
31지점	V1001610252	2017-12-26	교남동 46	교남동
	V1001610546	2018-03-09	세종로 100	종로 1,2,3,4가동
	V1001610540	2018-03-14	세종로 100	종로 1,2,3,4가동
	V1001610542	2018-03-14	세종로 100	종로 1,2,3,4가동
	V1001610543	2018-03-09	세종로 100	종로 1,2,3,4가동
	V1001610544	2018-03-09	세종로 100	종로 1,2,3,4가동

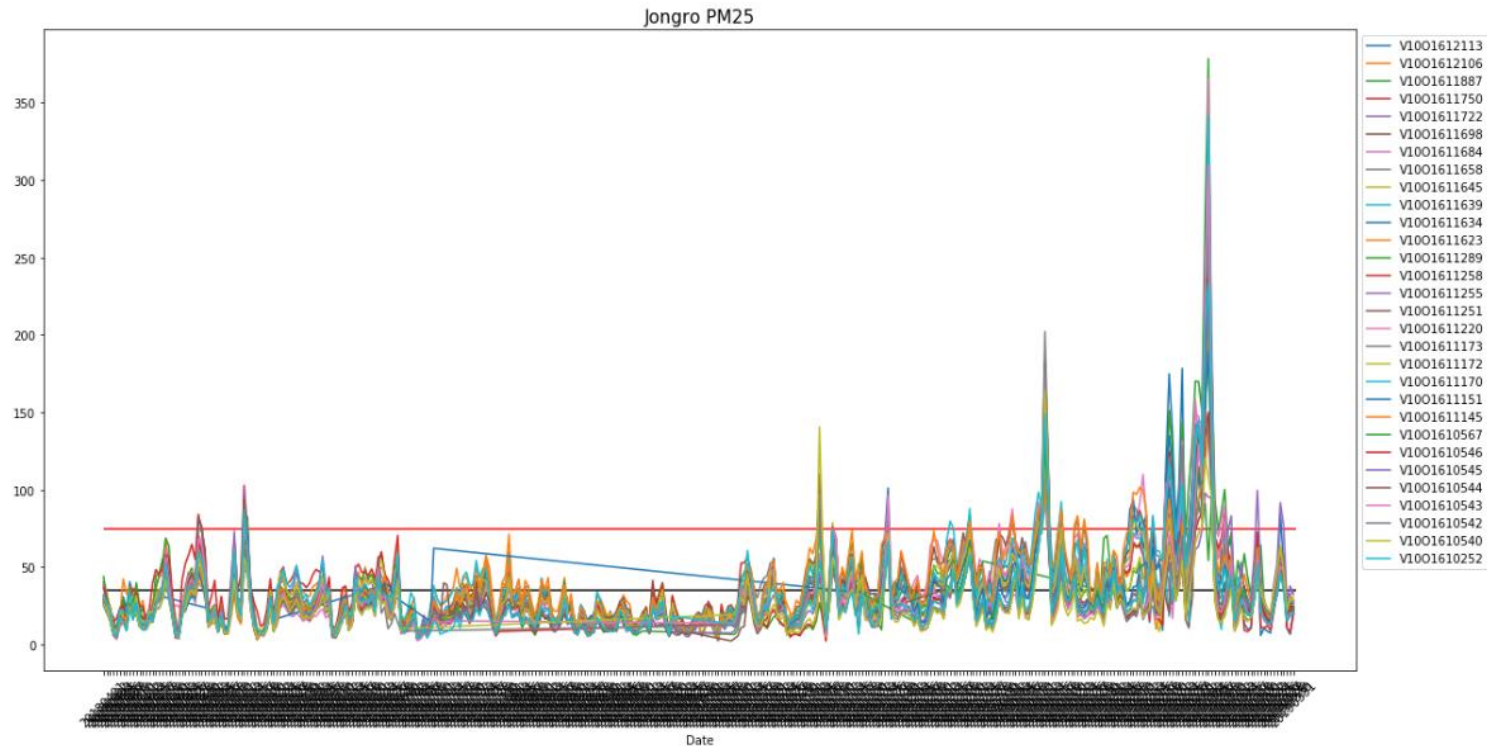
노원구	스테이션	측정기 등록일	위치	행정동
22지점	V1001610610	#	하계동 280	하계1동
	V1001610376	2017-12-27	상계동 734	상계6,7동
	V1001610293	2017-12-27	상계동 692-2	상계10동
	V1001610356	2017-12-27	상계동 1102-17	상계1동
	V1001610616	2017-12-23	상계2동 407-39	상계2동

# 1. 환경기상 데이터(케이웨더)

측정기 고유번호		미세먼지 PM-10		휘발성유기화합물 농도		온도		미세먼지 PM-2.5	
측정날짜시간	실외 측정기 구분	이산화탄소 농도		소음		습도			
tm	serial	flag	pm10	co2	vocs	noise	temp	humi	pm25
2.02E+11	V10O1610	1	38	-9999	-9999	49	13	65	20
2.02E+11	V10O1610	1	38	-9999	-9999	49	13	65	20
2.02E+11	V10O1610	1	45	-9999	-9999	50	13	65	23
2.02E+11	V10O1610	1	40	-9999	-9999	49	13	65	21
2.02E+11	V10O1610	1	44	-9999	-9999	49	12.9	66	23
2.02E+11	V10O1610	1	42	-9999	-9999	49	12.9	66	22
2.02E+11	V10O1610	1	46	-9999	-9999	48	12.9	66	24
2.02E+11	V10O1610	1	45	-9999	-9999	49	12.9	66	23

-99/-999 결측치로 간주 -> 제거

# 1. 환경기상 데이터(케이웨더)



## 2. 유동인구 데이터(SKT) \_성연령 유동

기간: 180401-190331

	A	B	C	D	E	F	G	H	I	J
1	STD_YM STD_YMD HDONG_CD HDONG_NM MAN_FLOW_POP_CNT_0004 MAN_FLOW_POP_CNT_0509 MAN_FLOW_POP_CNT_2529 MAN_FLOW_POP_CNT_70SEI									
2	201804 20180401 1111051500 청운효자동 0.05 78.93 403.26 642.66 984.24 1792.3 1936.1 2060.54 1787.98 1787.98									
3	201804 20180401 1111053000 사직동 1.25 262.54 1207.22 2177.88 4076.89 6887.72 6786.94 6873.7 5920.9									
4	201804 20180401 1111054000 삼청동 0 78.07 255.24 331.09 654.32 1264.86 1375.14 1445.4 1237.07 1658.3									
5	201804 20180401 1111055000 부암동 0 228.5 753.31 981.54 1615.77 2732 3390.58 4288.35 3857.72 4912.6									
6	201804 20180401 1111056000 평창동 0 354.38 1500.63 1963.08 2593.07 4927.24 5718.38 6997.92 6618.99 6618.99									



년월일    법정동코드/명칭

남성\_0004세/0509세 ~ 2529세 ~ 70세이상 (여성 동일)

1	STD_YM	STD_YMD	HDONG_CD	HDONG_NM	MAN_FLO	MAN_FLO	MAN_FLO	MAN_FLO	MAN_FLO	MAN_FLO
2	201804	20180401	1.11E+09	청운효자동	0.05	78.93	403.26	642.66	984.24	1792.3
3	201804	20180401	1.11E+09	사직동	1.25	262.54	1207.22	2177.88	4076.89	6887.72
4	201804	20180401	1.11E+09	삼청동	0	78.07	255.24	331.09	654.32	1264.86
5	201804	20180401	1.11E+09	부암동	0	228.5	753.31	981.54	1615.77	2732
6	201804	20180401	1.11E+09	평창동	0	354.38	1500.63	1963.08	2593.07	4927.24

## 2. 유동인구 데이터(SKT) \_시간대 유동(1시간 단위)

기간: 180401-190331

년월일

법정동코드/명칭

시간대별 유동인구(00시-23시)

STD_YM	STD_YMD	HDONG_C	HDONG_N	TMST_00	TMST_01	TMST_02	TMST_03	TMST_04	TMST_05
201804	20180401	1.11E+09	청운효자동	609.11	447.82	301.75	217.03	237.07	318.75
201804	20180401	1.11E+09	사직동	1904.32	1319.04	951.8	779.08	779.14	1355.75
201804	20180401	1.11E+09	삼청동	216.3	158.05	91.13	85.3	86.76	146.87
201804	20180401	1.11E+09	부암동	1300.61	904.38	670.74	433.02	571.34	771.25
201804	20180401	1.11E+09	평창동	2523.88	2168.69	1672.82	1300.87	1400.14	1686.09



### 3. 유통 테이터(GS리테일)

기간: 180401-190331

#### • 동별 매출지수

ADMD_NM	20180401	20180402	20180403	20180404	20180405	20180406
서울특별시	101.2%	109.5%	109.7%	108.9%	103.9%	109.7%
가회동	74.1%	85.6%	98.2%	106.0%	85.6%	92.4%
공릉1동	67.2%	73.1%	70.3%	73.6%	73.0%	73.3%
공릉2동	85.4%	75.8%	79.9%	81.5%	78.4%	79.9%
교남동	88.0%	78.0%	82.8%	73.0%	73.2%	67.6%
무악동	194.4%	199.8%	177.1%	208.9%	210.3%	215.9%

#### • 카테고리별 매출비중

ADMD_NM	ITEM_LCI	ANTC_ITEM_LCLS_NM	20180401	20180402	ADMD_NM	ITEM_LCI	ANTC_ITEM_LCLS_NM	20180401	20180402
가회동	10	식사	23.8%	24.8%	공릉1동	10	식사	31.9%	32.6%
가회동	20	간식	21.0%	21.5%	공릉1동	20	간식	20.7%	18.9%
가회동	30	마실거리	41.3%	41.0%	공릉1동	30	마실거리	36.2%	37.5%
가회동	40	홈&리빙	1.7%	2.5%	공릉1동	40	홈&리빙	3.3%	4.3%
가회동	50	헬스&뷰티	9.0%	8.6%	공릉1동	50	헬스&뷰티	5.0%	4.3%
가회동	60	취미&여가활동	1.0%	0.2%	공릉1동	60	취미&여가활동	1.0%	0.4%
가회동	70	사회활동	2.1%	0.2%	공릉1동	70	사회활동	1.4%	1.2%
가회동	80	임신/육아	0.1%	1.1%	공릉1동	80	임신/육아	0.5%	0.8%

## 4. 카드매출 데이터(신한카드)

기간: 180401-190331

기준일자	구코드	행정동코드	업종코드	성별코드	나이코드	이용건수	이용금액(단위: 천원)
			레저업소(21)/문화취미(22) /음료식품(81) 등				
20180401	110	515	21	F	30	4	180
20180401	110	515	21	F	55	4	22
20180401	110	515	21	M	20	35	184
20180401	110	515	21	M	25	70	425
20180401	110	515	21	M	30	18	82
20180401	110	515	21	M	35	4	44
20180401	110	515	21	M	40	4	53
	종로구	청운효자동			40-44세		

## 5. SNS 데이터(와이즈넷)

기간: 180401-190331

E2											
	A	B	C	D	E	F	G	H	I	J	K
1	DOCID	SECTION	DATE	TITLE	CONTENT						
2	204172453	블로그	201811022359	신동탄 이안파밀리에	분 양 정 보 신동탄 이안파밀리에 눈에 보이는 프리미엄 진짜네요 유쾌한 분양김						
3	204172474	블로그	201811022357	느린 점심	생활 속의 단상 느린 점심點心 묘광거사 2018 11 2 2357 URL 복사 이웃추						
4	204172506	블로그	201811022354	한화 태양	한화프렌즈 한화 태양의 숲 조성으로 미세먼지도 막고 환경도 지켜요 아르						
5	204172546	블로그	201811022354	섬유유연제	쇼핑 섬유유연제추천 미세먼지까지 잡아 주는 착한 샤프란 섬유유연제 꽃						

블로그/카페/뉴스

# (아들아 넌) **계획**(이 다 있구나)

초반  
(지금 하는중)

## — 1. 환경기상 데이터(케이웨더)

지역별/시간별 미세먼지 현황 확인,  
기간별 데이터 세분화 후 정밀 분석

중반  
(중간고사 이후)

## — 2. 유동인구 데이터(SKT)

미세먼지 농도 낮은/높은 날  
성연령별 또는 시간대별 유동인구 변화 추이 분석

## — 3. 유통 데이터(GS리테일)

노원/종로 상권 분석  
미세먼지의 영향이 특정 상권에 영향을 미치는지 파악

## — 4. 카드매출 데이터(신한카드)

상권분석을 바탕으로 성연령/지역별 소비패턴 분석

후반

## — 5. SNS 데이터(와이즈넷)

...텍스트 마이닝.....?

## — 종합적 이해, 결론 도출

# 화목한 우리 팀의 장점은?

1. 화목하다

2. 그 어디서도 쉽게 얻을 수 없는,  
실생활에 가까운, 다양한 분야의  
데이터를 다뤄볼 수 있다!



# 이런 팀원이라면 환영해요!

## 1. 전처리/시각화에 관심있는 분

정제된듯 정제되지 않은 빅콘 데이터..  
전처리를 함께할 팀원이 필요하답니다~  
또한, 환경기상이나 유동인구 데이터 분석 시  
시각화에 집중할 예정이에요. 관심있다면 들어와!

## 2. 아무말이나 던질 수 있는 분

5개 분야의 데이터를 총망라하는 만큼  
데이터 간의 연결고리를 찾는게 매우 중요하답니다~  
분석방향 설정이나 결과 해석 시 아무말이나 툭툭  
던지면서 팀의 방향키를 짚어줄 사람이 필요해요!



기다릴게요~♡~☆~♡

From. 동빈 수희 유경

