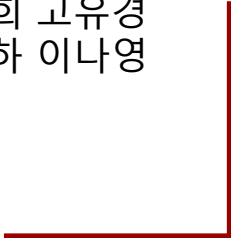




KU-BIG Fine Dust TEAM

이동빈 박수희 고유경
김재훈 김은하 이나영



INDEX

1. Project Review
2. Data Simplification
3. Revenue Transition
4. Future Plans

Project Review

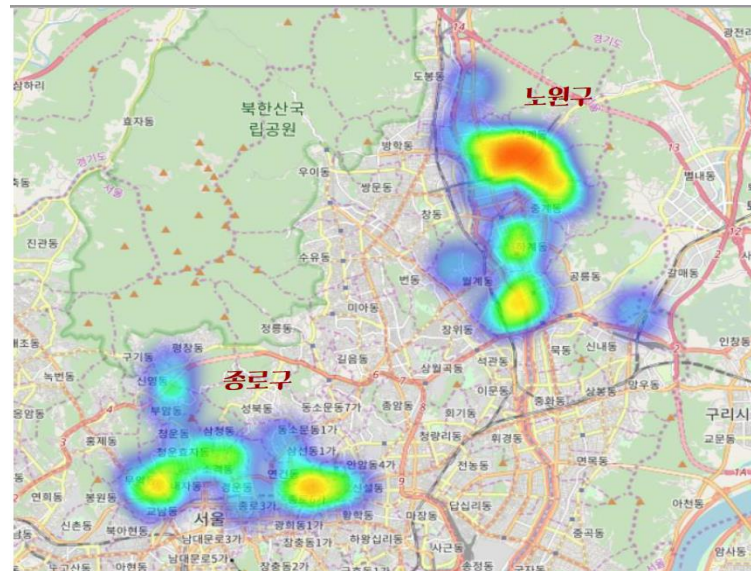
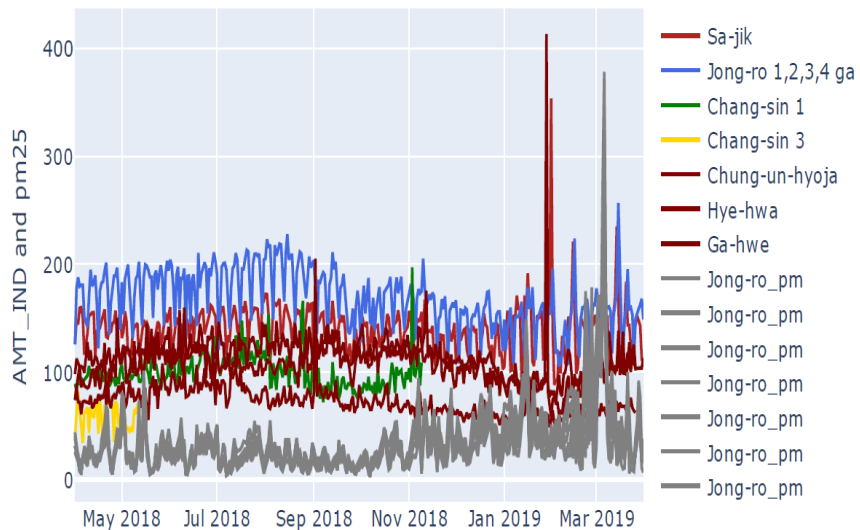
Data Preprocessing

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

Project Review

Visualization



Data Simplification

프로젝트 진행에 있어서 현재 데이터셋의 문제점

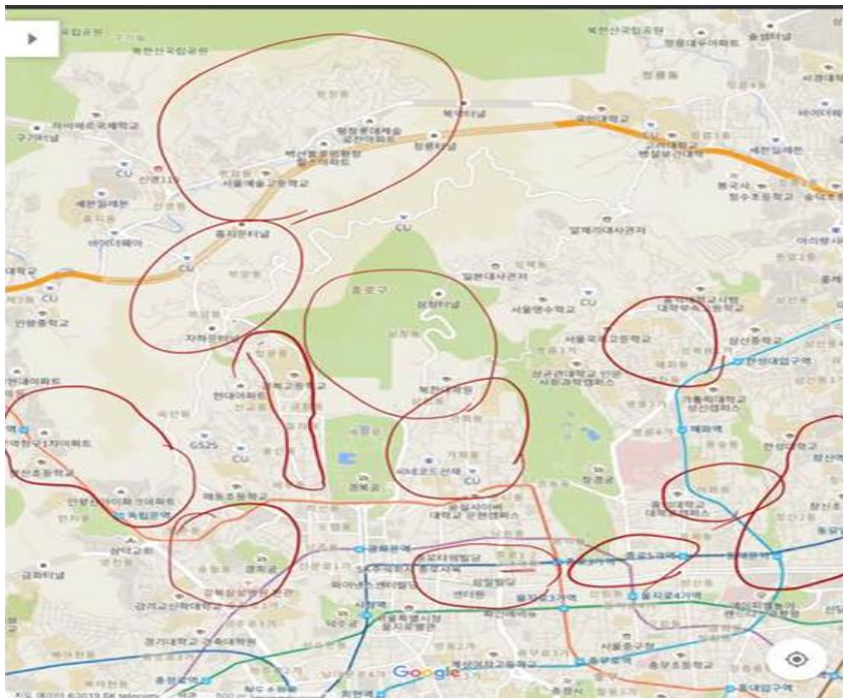
1. 방대한 양의 환경기상데이터.
2. 행정동 별 기상관측일수가 상이함.
3. 각 데이터셋 별로 지역을 대표하는 key값이 상이함.

Data Simplification

프로젝트 진행에 있어서 현재 데이터셋의 문제점

1. 방대한 양의 환경기상데이터. => 하루 동안의 미세먼지 지수의 평균을 계산
2. 행정동 별 기상관측일수가 상이함. => 각 구의 지역마다 지리적 대표성을 띄는 행정동 중 기상관측일수가 가장 많은 행정동을 뽑음
3. 각 데이터셋 별로 지역을 대표하는 key값이 상이함. => 행정동명으로 key값을 통일

Data Simplification



선정한 행정동

1. 종로구 : 가회동, 부암동, 종로1234가동, 청운효자동, 평창동, 혜화동, 창신3동
2. 노원구 : 상계1동, 상계3,4동, 상계6,7동, 월계1동, 중계본동, 하계1동

=> 종로구 (7동), 노원구(6동)으로 총 13개 동 선정

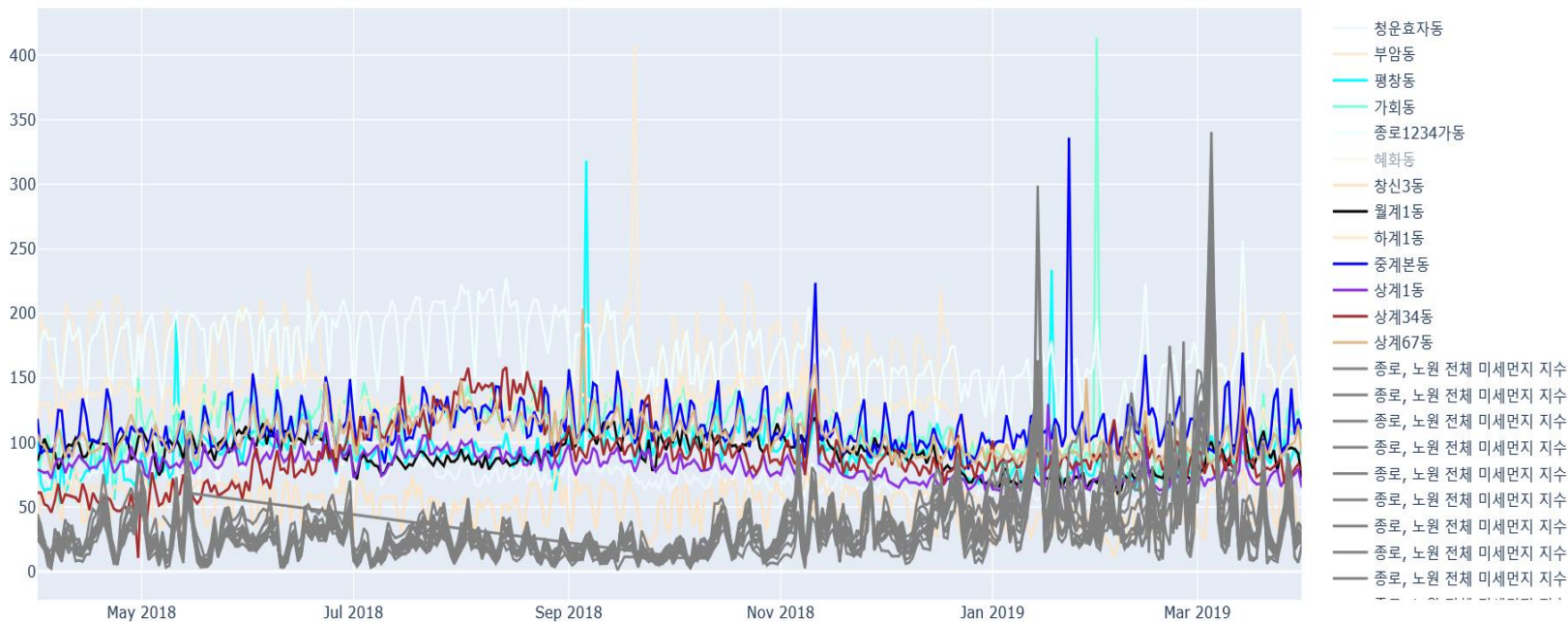
Data Simplification

STD_DD	GU_CD	DONG_CD	HDONG_NM	MCT_CAT	SEX_CD	AGE_CD	USE_CNT	USE_AMT
20180401	110	515	청운효자동	21	F	30	4	180
20180401	110	515	청운효자동	21	F	55	4	22
20180401	110	515	청운효자동	21	M	20	35	184
20180401	110	515	청운효자동	21	M	25	70	425
20180401	110	515	청운효자동	21	M	30	18	82
20180401	110	515	청운효자동	21	M	35	4	44

ADMD_CD	HDONG_NM	OPER_DT	AMT_IND	LCLS_10_P	LCLS_20_P
11110515	청운효자동	20180401	0.736932914	0.279017883	0.299568923
11110515	청운효자동	20180402	0.864277963	0.2696392	0.246127572
11110515	청운효자동	20180403	0.72225074	0.317965784	0.246600959
11110515	청운효자동	20180404	0.717211275	0.339812596	0.22223409
11110515	청운효자동	20180405	0.606730476	0.381157379	0.242948835
11110515	청운효자동	20180406	0.720851322	0.31256568	0.241087719
11110515	청운효자동	20180407	0.727591753	0.314954491	0.251578318

Revenue Transition

- 추려진 데이터셋을 활용하여 총 매출지수와 미세먼지의 변화 추이를 관찰



Revenue Transition

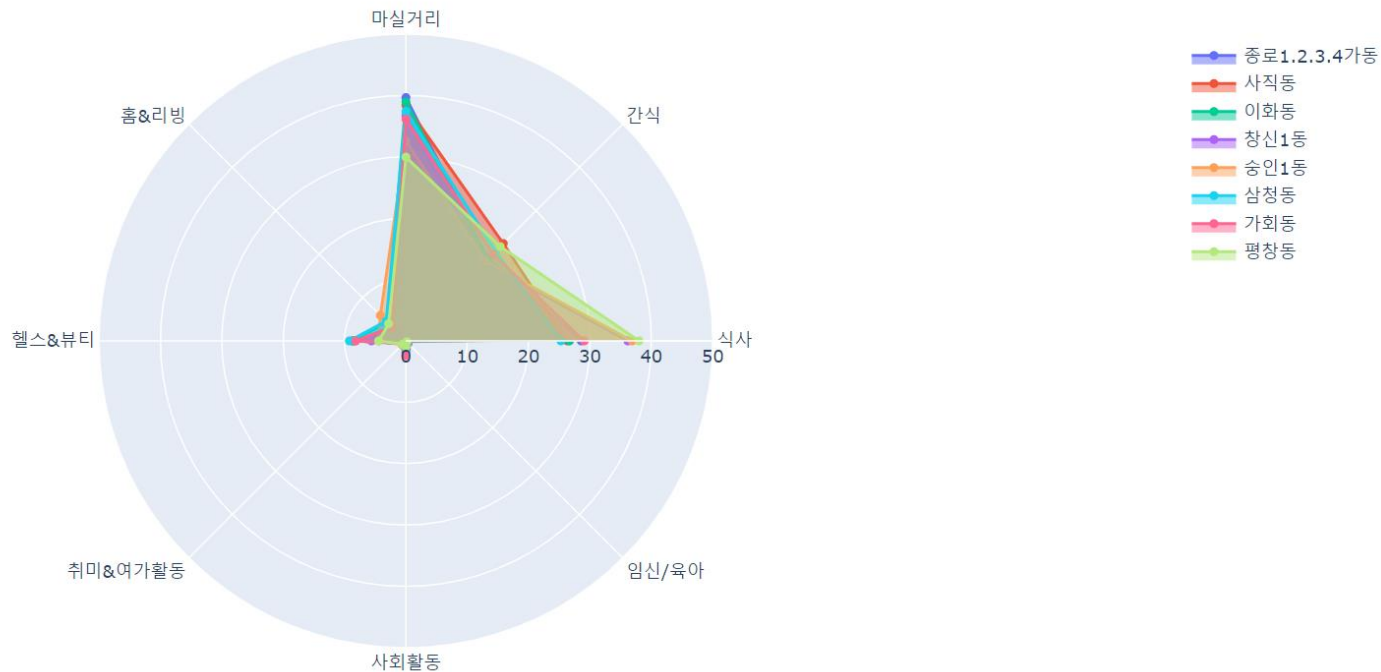
- 추후 소비 항목별 매출 지수를 활용하여 각 소비 항목들의 매출 지수 변화 추이를 관찰해볼 예정 => (ex. 미세먼지가 심한 날에는 마실거리의 소비 비중이 증가하더라.)

HDONG_NM	OPER_DT	AMT_IND	LCLS_10_P	LCLS_20_P	LCLS_30_P	LCLS_40_P	LCLS_50_P	LCLS_60_P	LCLS_70_P	LCLS_80_P
청운효자동	20180401	0.736932914	0.279017883	0.299568923	0.3348155	0.015635754	0.05378244	0.00026485	0.01309187	0.00382278
청운효자동	20180402	0.864277963	0.2696392	0.246127572	0.40501024	0.037167065	0.030789164	0.001088201	0.002976674	0.007201885

AMT_IND	매출지수
LCLS__10_P	식사_비중
LCLS__20_P	간식_비중
LCLS__30_P	마실거리_비중
LCLS__40_P	홈&리빙_비중
LCLS__50_P	헬스&뷰티_비중
LCLS__60_P	취미&여가활동_비중
LCLS__70_P	사회활동_비중
LCLS__80_P	임신/육아_비중
LCLS__90_P	기호품_비중

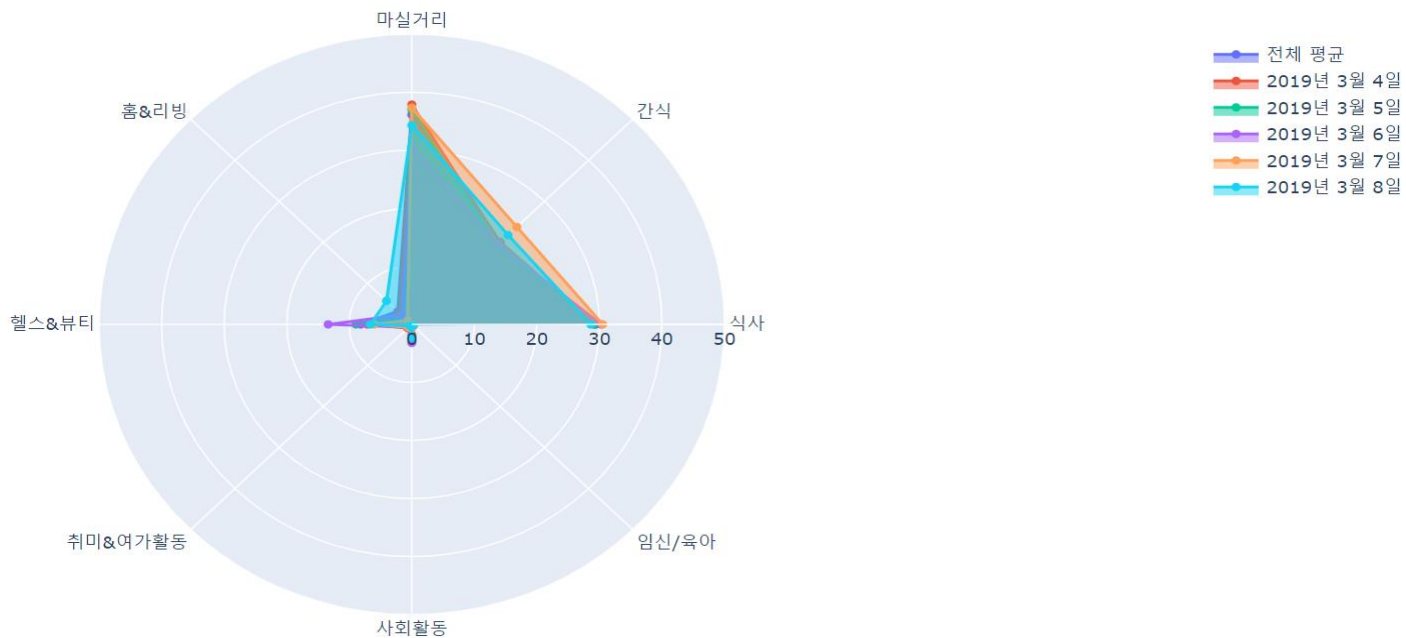
Revenue Transition

종로구 8개 동 상권 비교



Revenue Transition

종로구 가회동 미세먼지 극심한 기간



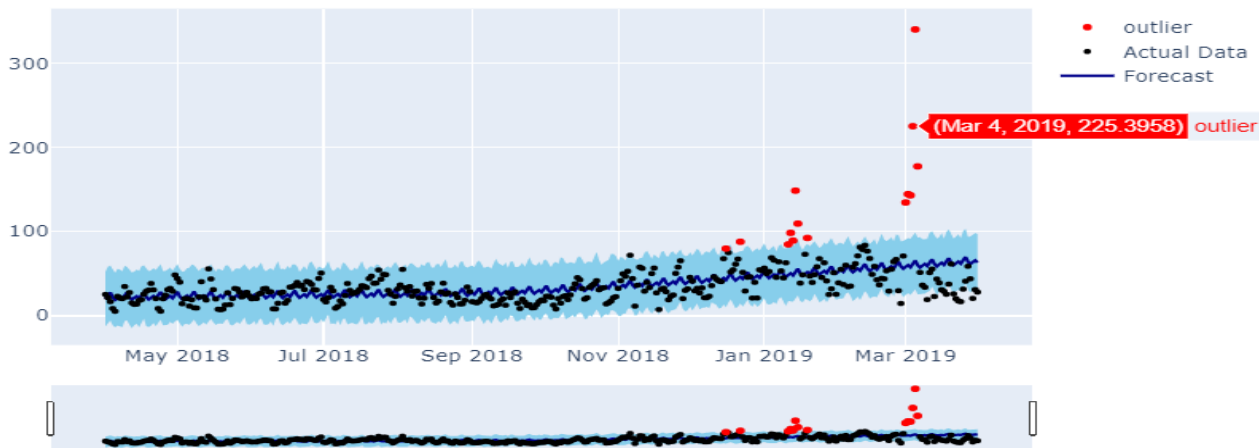
Future Plans

1. 추후 소비 항목별 매출 지수를 활용하여 각 소비 항목들의 매출 지수 변화 추이를 관찰해볼 예정 (현재는 앞서 살펴본 바와 같이 '변화 추이'를 관찰하지는 않았음)
2. 카드매출 데이터와 유동인구 데이터의 활용 방안 검토
3. 새로운 시각화 패키지를 활용하여 데이터의 변화 추이에 대하여 좀 더 세밀한 분석

Future Plans (Prophet)

Prophet : facebook에서 개발한 시계열 예측 패키지. 경험적 규칙을 사용하는 단순 회귀 model이지만 단기적 예측에서는 큰 문제없이 사용 가능

Plot : Data & Outlier using Prophet



Q & A