

미세먼지 발생량에 따른 세그먼트 별 소비 패턴 예측 AI모델 개발

미세조정 팀

목차

01 우리는 누구인가?



미세조정 팀 소개

02 문제는 무엇인가?



선행 연구 사례
선행 연구 사례 검증

03 어떻게 풀 것인가?



클로즈업을 통한 세분화 필요성
EDA 및 데이터 전처리
유통 인구 예측을 위한 앙상블 모델링
판매 지수 예측을 위한 다변량 LSTM 모델링

04 어떻게 활용할 것인가?



유통 부문 비즈니스 모델
카드매출 부문 비즈니스 모델



정우일

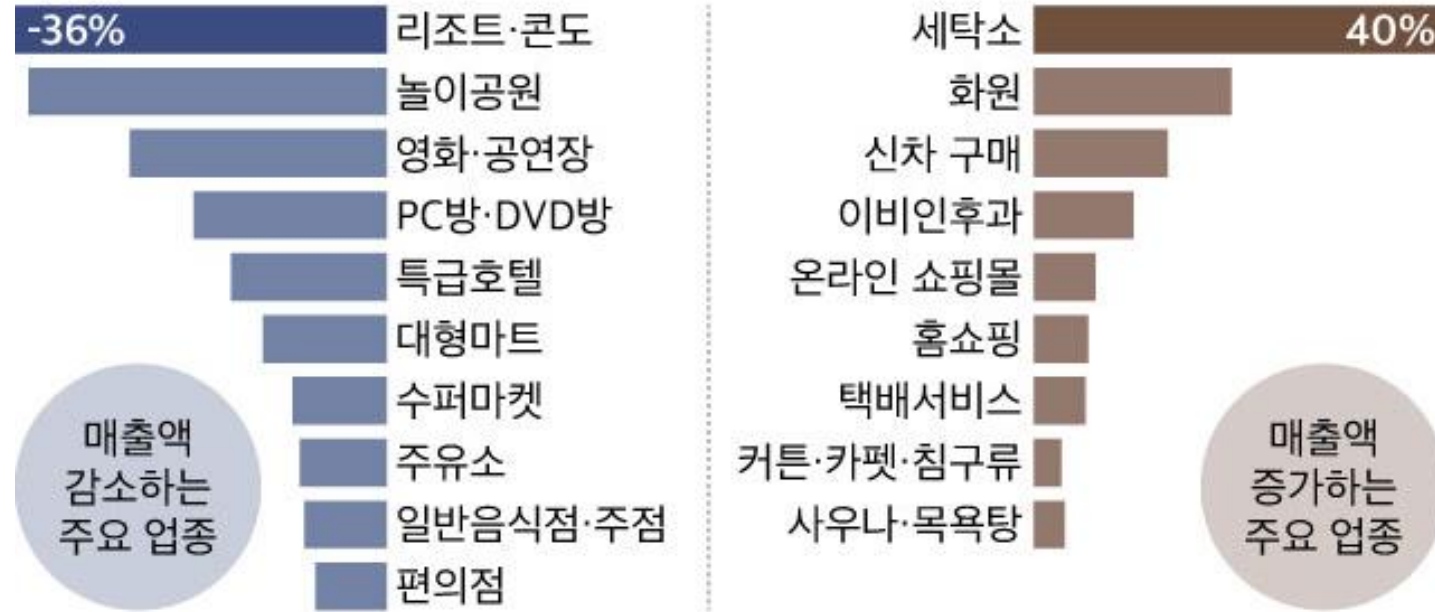
조경아

권순철

김보선

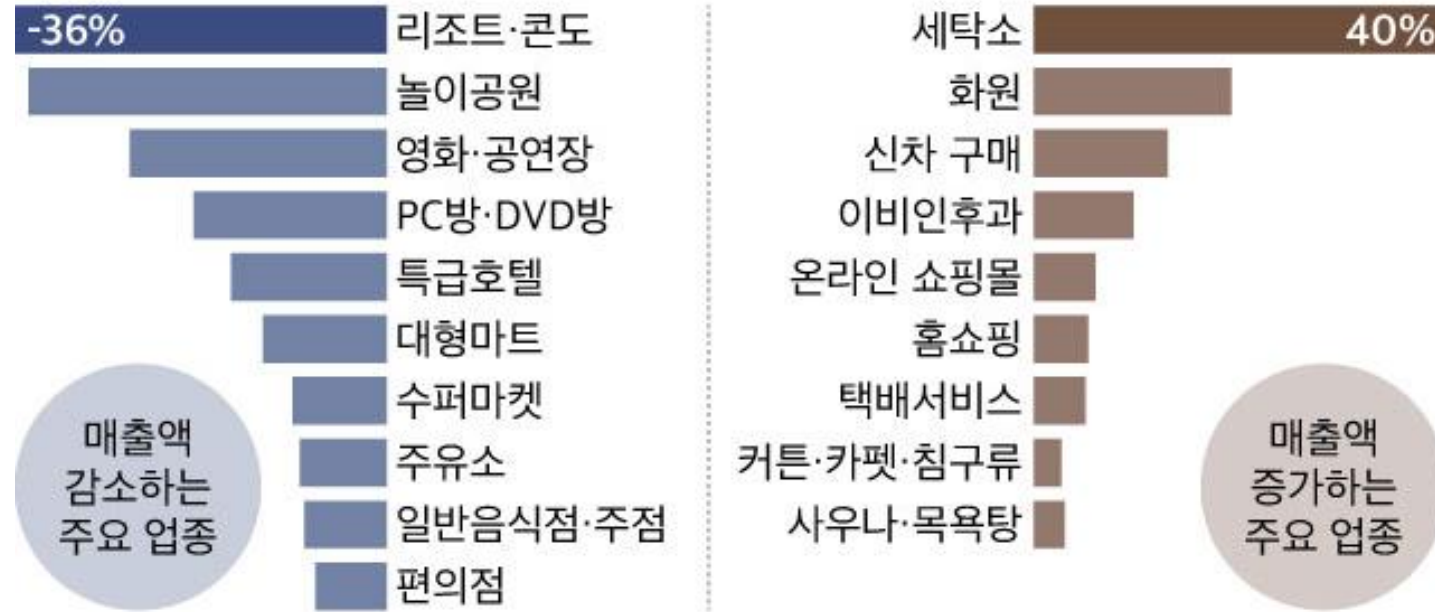
만나서 반갑습니다!
저희는 미세조정 팀 입니다!

미세먼지 뉴스 건수 증가에 따른 업종별 매출 증감률 단위: %



출처: "미세먼지 뉴스 많은 날, 목욕탕·세탁소 매출액 오른다" - Chosun Biz (2019.04.17)

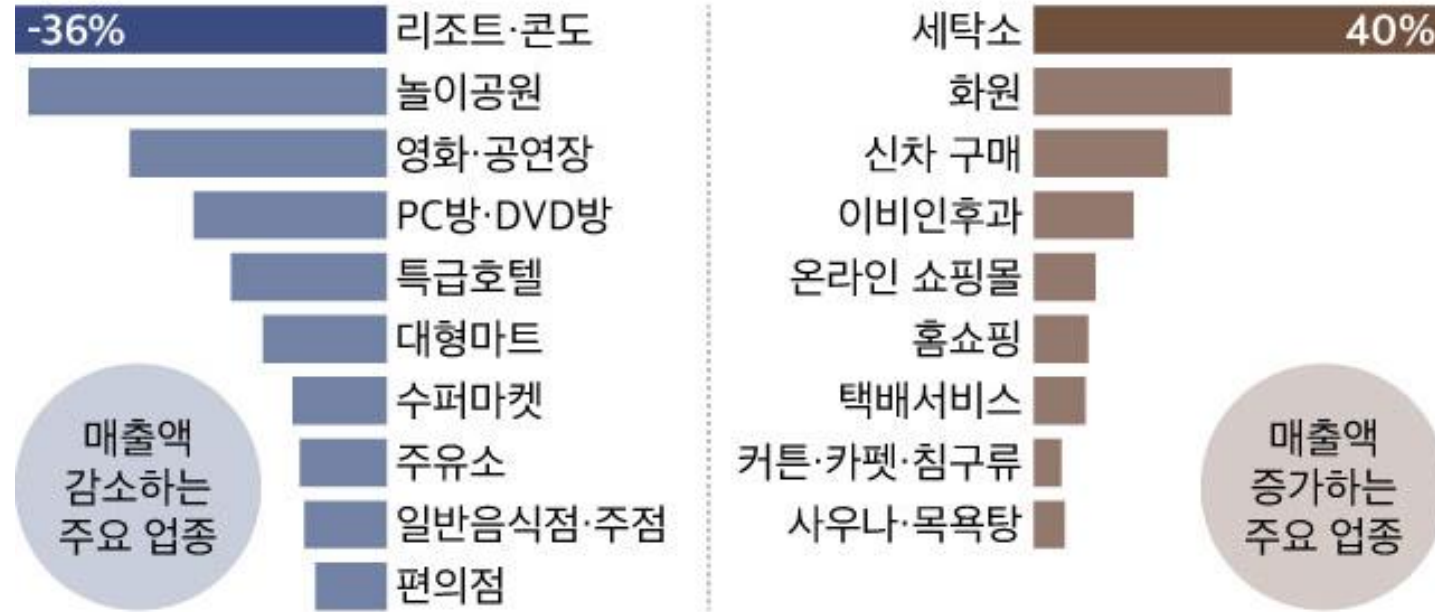
미세먼지 뉴스 건수 증가에 따른 업종별 매출 증감률 단위: %



출처: "미세먼지 뉴스 많은 날, 목욕탕·세탁소 매출액 오른다" - Chosun Biz (2019.04.17)

실제로 미세먼지 발생량이 사람들의 소비 생활에 영향을 주는 걸까요?

미세먼지 뉴스 건수 증가에 따른 업종별 매출 증감률 단위: %



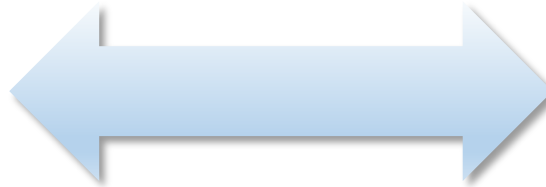
출처: "미세먼지 뉴스 많은 날, 목욕탕·세탁소 매출액 오른다" - Chosun Biz (2019.04.17)

그래서 직접 검증해보기로 했습니다!



SNS 데이터

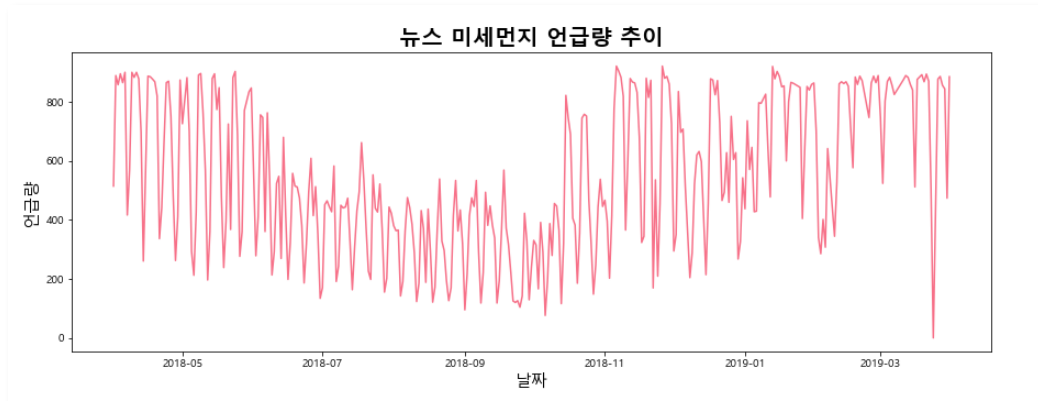
- 미세먼지 관련 뉴스 데이터
- 미세먼지 관련 블로그 데이터
- 미세먼지 관련 카페 데이터



신한카드

카드매출 데이터

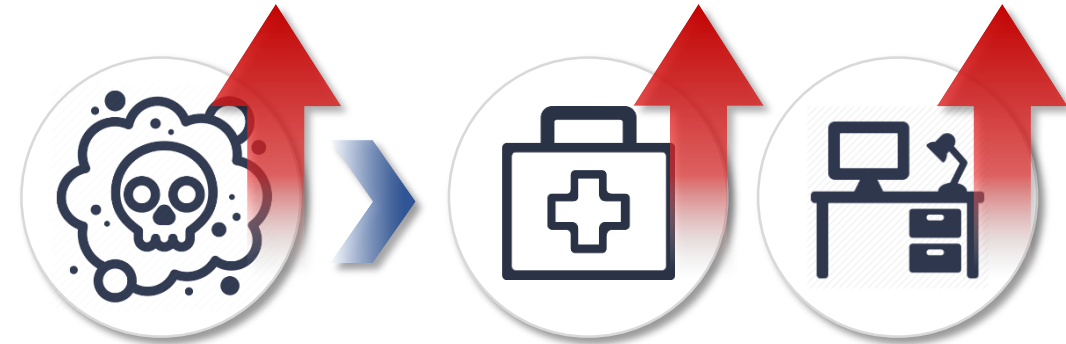
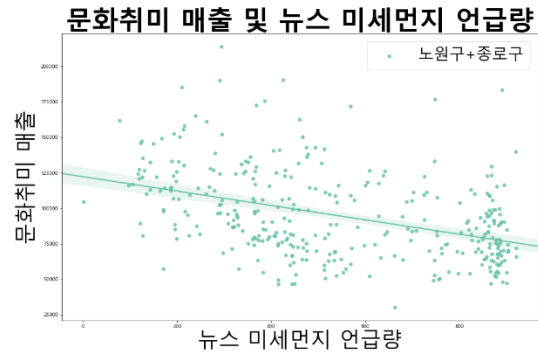
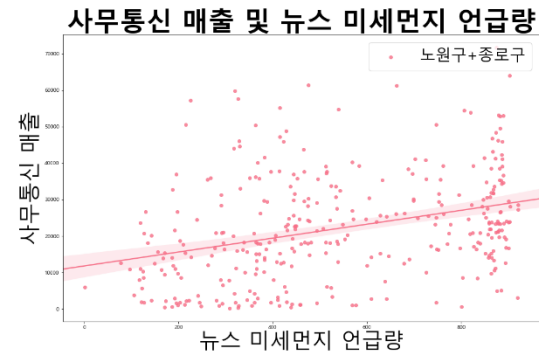
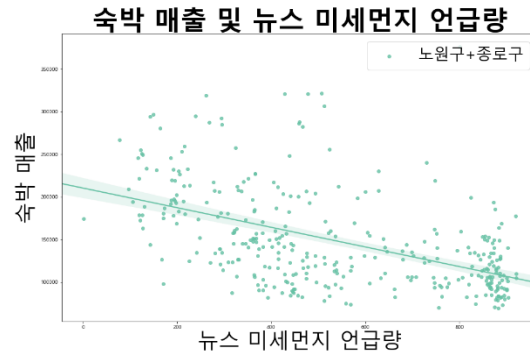
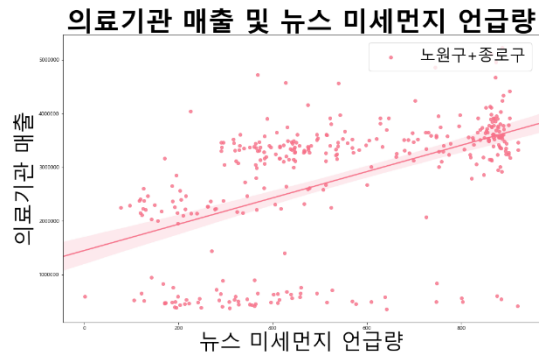
- 노원구 카드매출 데이터
- 종로구 카드매출 데이터



업종	세부 업종
의료기관	· 약국, 종합병원, 개인병원 등
사무·통신	· 사무기기, 정보통신기기, 컴퓨터 등
숙박	· 호텔, 콘도미니엄, 모텔 등
문화·취미	· 공연장, 극장, 운동경기관람 등
-	-

SNS 데이터와 카드매출 데이터를 분석하여,
미세먼지 뉴스 언급량과 카드매출 사이의 관계를 살펴보겠습니다.

뉴스 미세먼지 언급량과 업종 별 카드매출 사이의 관계



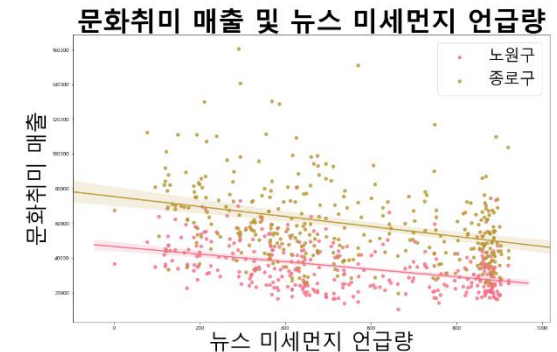
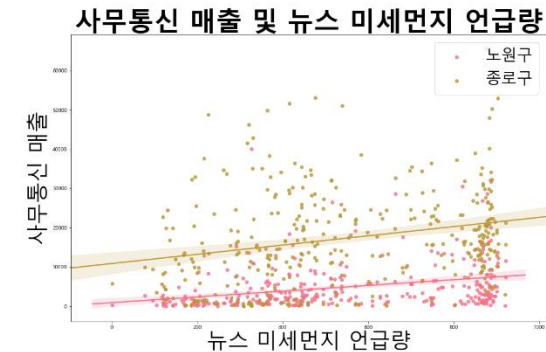
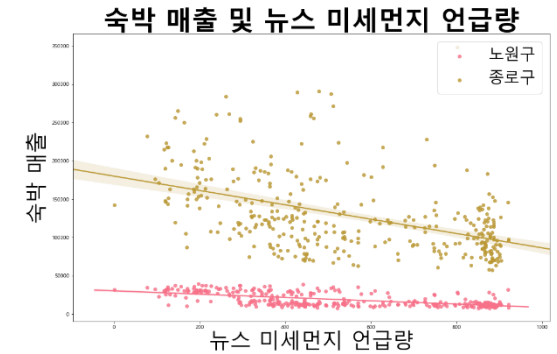
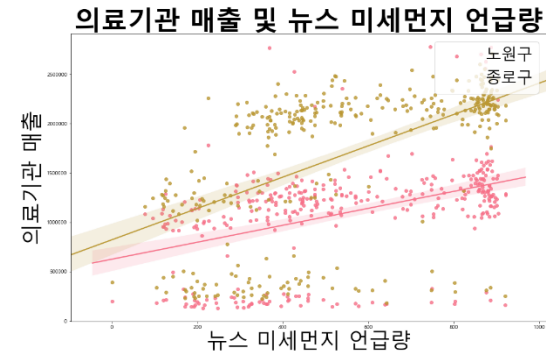
미세먼지 언급량이 높을 수록
의료기관, 사무통신 카드매출도 높은 경향



미세먼지 언급량이 높을 수록
숙박, 문화취미 카드매출은 낮은 경향

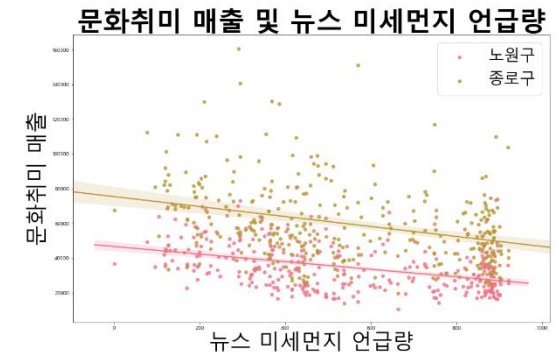
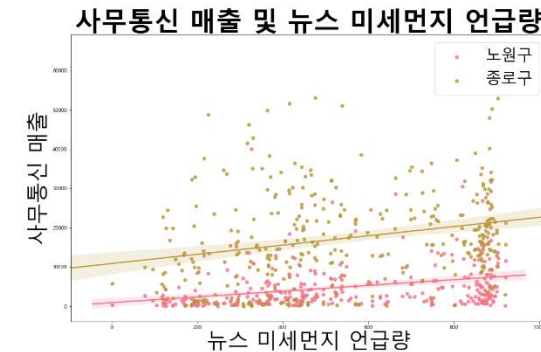
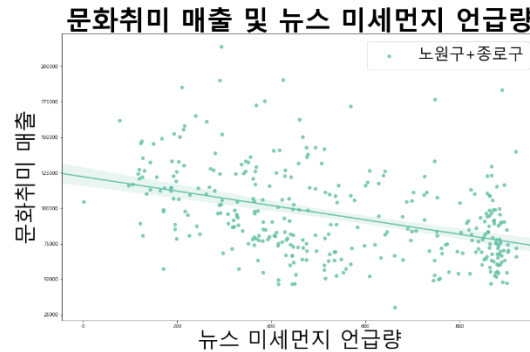
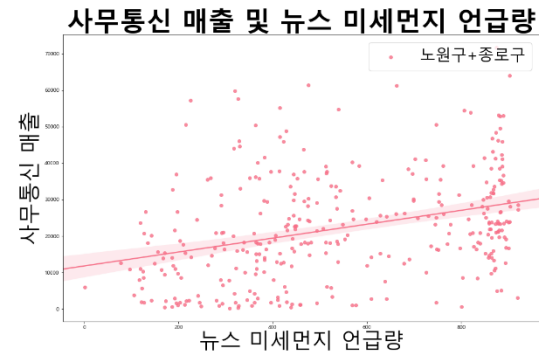
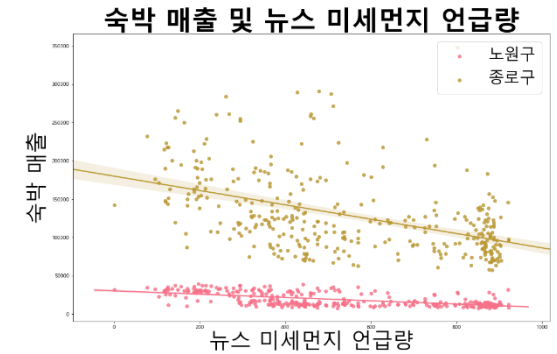
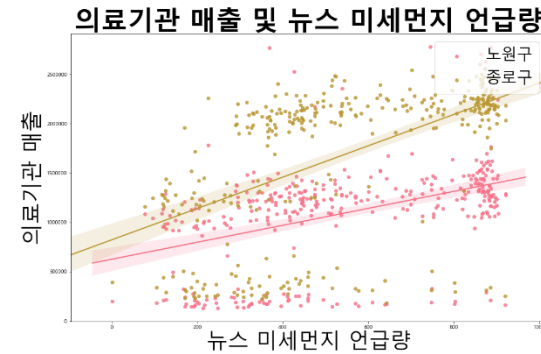
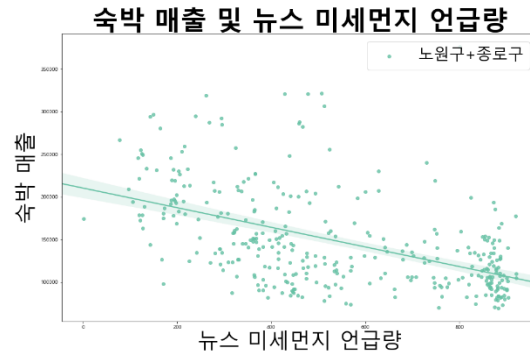
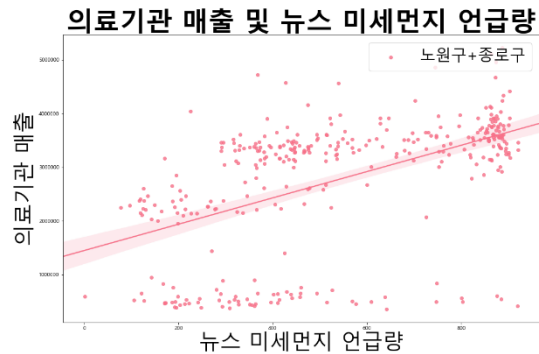
그 결과, 미세먼지 언급량과 업종 별 카드매출 사이의
선형 관계를 확인했습니다.

뉴스 미세먼지 언급량과 업종 별 카드매출 사이의 관계

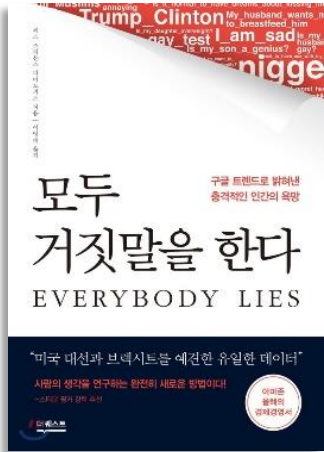


종로구와 노원구를 구분하여 살펴보면 종로가 더 높은 관련성을 보이고 있습니다.

뉴스 미세먼지 언급량과 업종 별 카드매출 사이의 관계

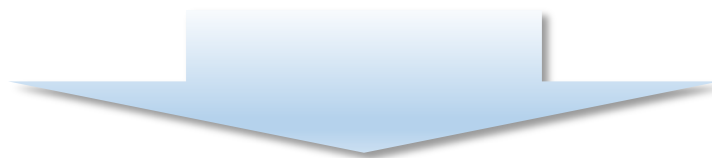


즉, 더 세밀하게 분석할수록,
미세먼지에 더 민감한 집단을 찾아낼 수 있다는 것을 확인했습니다!



“ 빅데이터는 데이터세트의 작은 일부를 유효하게 클로즈업해서 그들이 어떤 사람인지에 대한 새로운 식견을 제공한다.

연령 이외의 다른 부분을 클로즈업해볼 수도 있다. 데이터가 충분하면 특정 도시의 사람들이 어떻게 행동하는지도 알 수 있다. 사람들이 매시간, 심지어는 매분 어떻게 움직이는지도 알 수 있다. “ 세스 스티븐스 다비도위츠의 '모두 거짓말을 한다' 중에서



세그먼트

전체



구 단위



동 단위



동/성별/연령 단위

따라서 지역/성별/연령 단위까지 세분화하여,
세그먼트 별 소비 생활을 분석하기로 했습니다!

불쾌지수

Discomfort_index(불쾌지수) 추가하기

- https://www.kma.go.kr/HELP/basic/help_01_05.jsp

```
find_dust['dis_index']=(9/5)*find_dust['temp']-0.55*(1-0.01*(find_dust['humi']))*((9/5)*find_dust['temp']-26)+32
```

GS인덱스

5.1.4 'gsindex1' : 동별매출지수 / 동별 유동인구

- 동별매출지수와 GS점포수와 마찬가지로 성별과 연령을 무시한 유동인구로 만들어 GS index2을 생성한다.
- 다만 지수를 생성할 때 동별 매출지수(%)와 동별 유동인구(명)의 단위가 다르므로 변수들을 정규화한 후 지수를 생성한다.

```
a1 = (tp_cir_gs['values'] - tp_cir_gs['values'].mean()) / tp_cir_gs['values'].std()
a2 = (tp_cir_gs['tp_method1'] - tp_cir_gs['tp_method1'].mean()) / tp_cir_gs['tp_method1'].std()
tp_cir_gs['gsindex1'] = a1 / a2
```

환경탄력성

카드소비의 환경탄력성 생성

idea : 미세먼지에 따른 세그먼트 별 카드소비의 민감성을 알아보기 위함

- 기준 : 동 및 월, 연령,성별,요일 및 카테고리별 카드소비값 median 값 생성
 - 연령, 성별에 따라 있는 카테고리의 수가 다르므로 데이터 수 파악 어려움

```
dfff2_1 = dfff1.groupby(['district', 'month', 'weekday', 'AGE_CD', 'SEX_CD', 'MCT_CAT_CD'])['USE_CNT'].median().reset_index()
dfff2_1 = dfff2_1.rename(columns = {'USE_CNT': 'USE_CNT_median'})
```

- 카드소비 환경탄력성
- 유동인구 환경탄력성
- 유통지수 환경탄력성

불쾌지수

Discomfort_index(불쾌지수) 추가하기

- https://www.kma.go.kr/HELP/basic/help_01_05.jsp

```
find_dust['dis_index']=(9/5)*find_dust['temp']-0.55*(1-0.01*(find_dust['humi']))*((9/5)*find_dust['temp']-26)+32
```

GS인덱스

5.1.4 'gsindex1' : 동별매출지수 / 동별 유동인구

- 동별매출지수와 GS점포수와 마찬가지로 성별과 연령을 무시한 유동인구로 만들어 GS index2을 생성한다.
- 다만 지수를 생성할 때 동별 매출지수(%)와 동별 유동인구(명)의 단위가 다르므로 변수들을 정규화한 후 지수를 생성한다.

```
a1 = (tp_cir_gs['values'] - tp_cir_gs['values'].mean()) / tp_cir_gs['values'].std()
a2 = (tp_cir_gs['tp_method1'] - tp_cir_gs['tp_method1'].mean()) / tp_cir_gs['tp_method1'].std()
tp_cir_gs['gsindex1'] = a1 / a2
```

환경탄력성

카드소비의 환경탄력성 생성

idea : 미세먼지에 따른 세그먼트 별 카드소비의 민감성을 알아보기 위함

- 기준 : 동 및 월, 연령,성별,요일 및 카테고리별 카드소비값 median 값 생성
- 연령, 성별에 따라 있는 카테고리의 수가 다르므로 데이터 수 파악 어려움

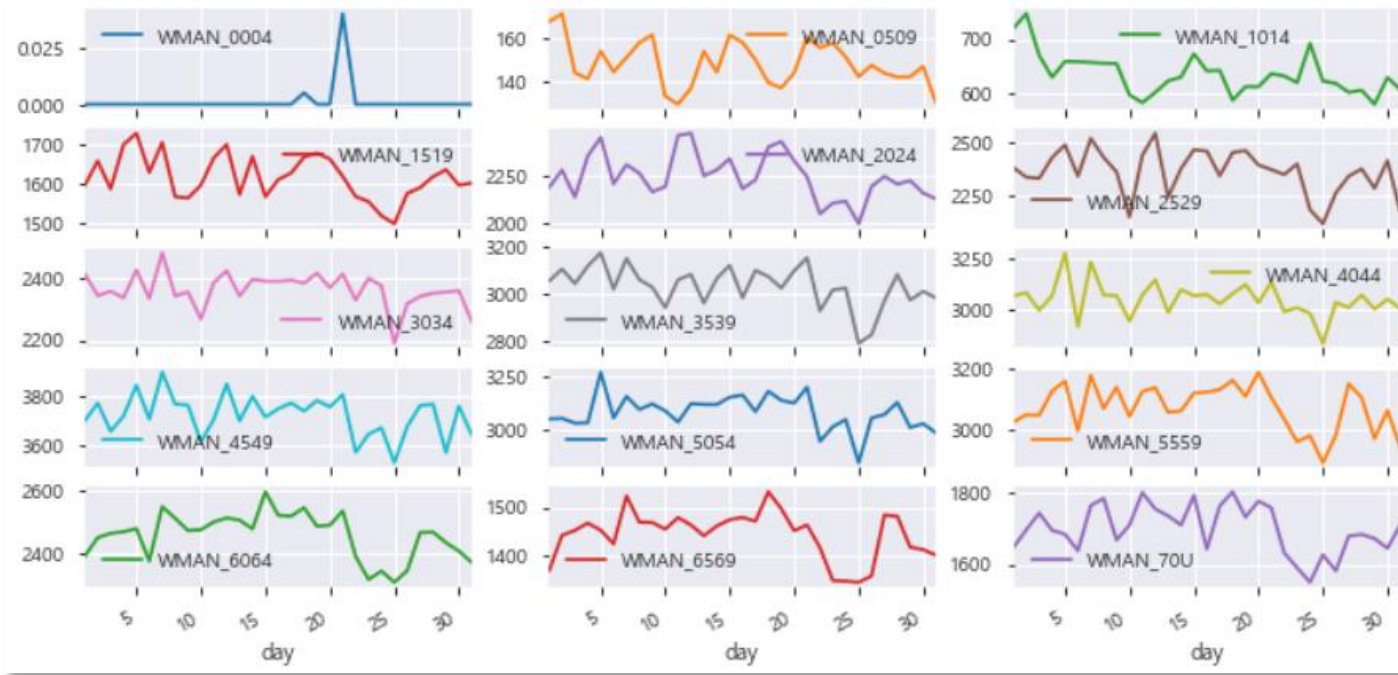
또한, 미세먼지 뿐만 아니라

- 카드소비 환경탄력성
- 유동인구 환경탄력성
- 유복지수 환경탄력성

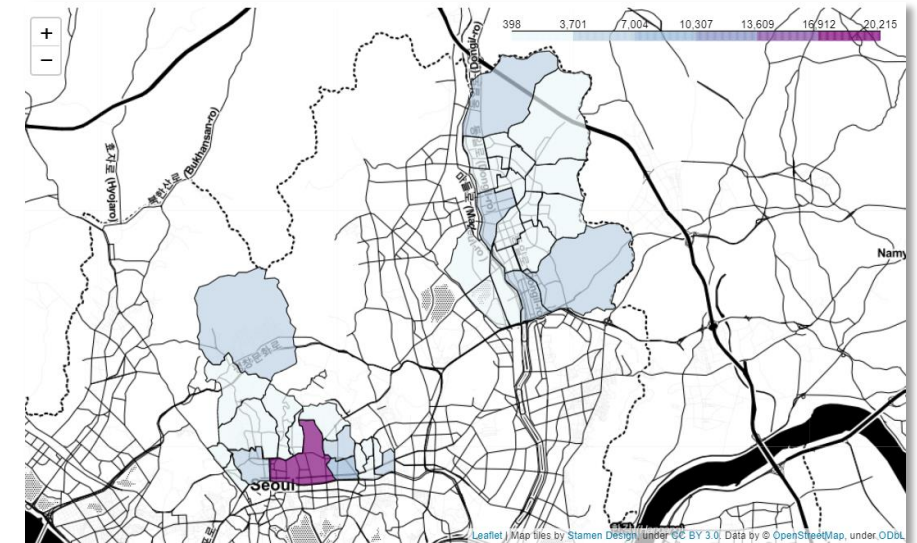
소비 생활에 영향을 주는 다양한 변수들도 고려하기로 했습니다.

```
df['2014-01-01'].ffrf1.gdcuruv['d_str', 'co', 'w', 'day', 'h3f', 'EX_CD', 'MT_CD', 'US_CNT', 'med', 'Asst', 'nd', 'x']
df['2014-01-01'].ffrf2.rename(columns = {'US_CNT': 'USE_CN', 'me', 'nd', 'x'})
```

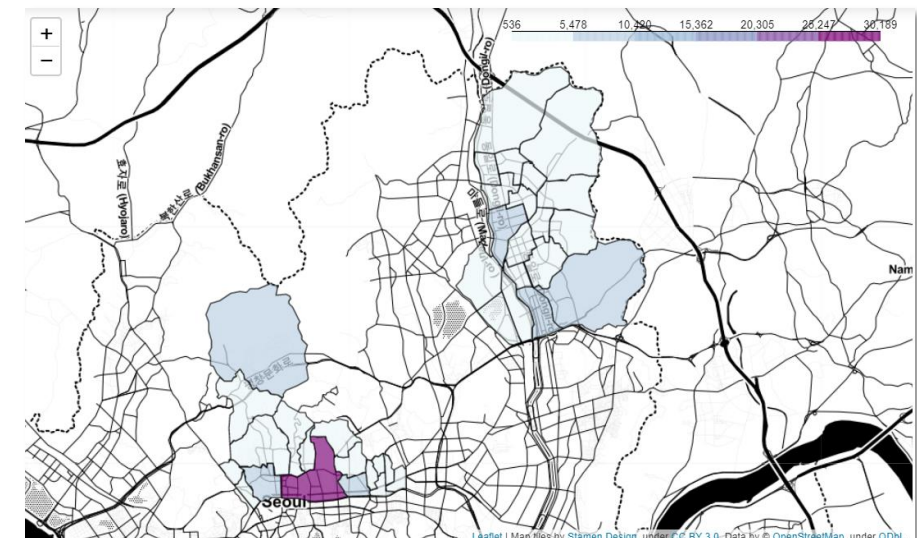

유동인구 데이터 EDA



<여성 연령 별 유동인구 추이>



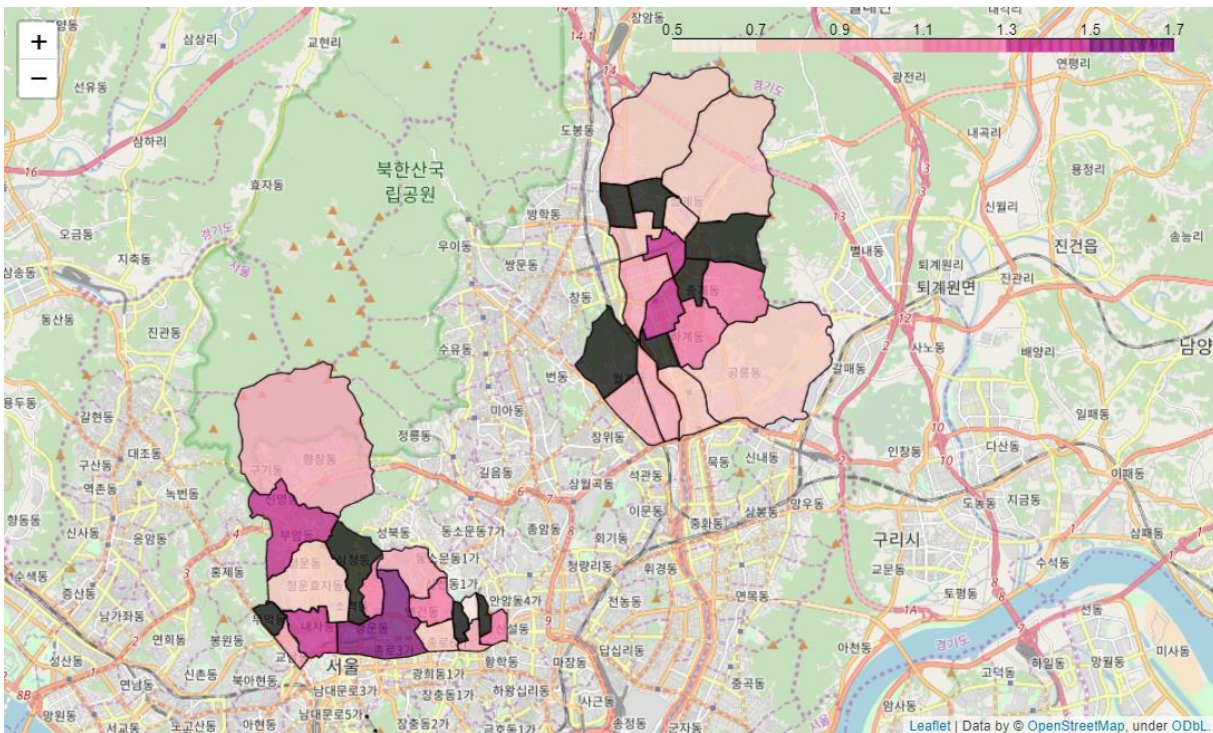
<여성 행정동 별 유동인구 비율>



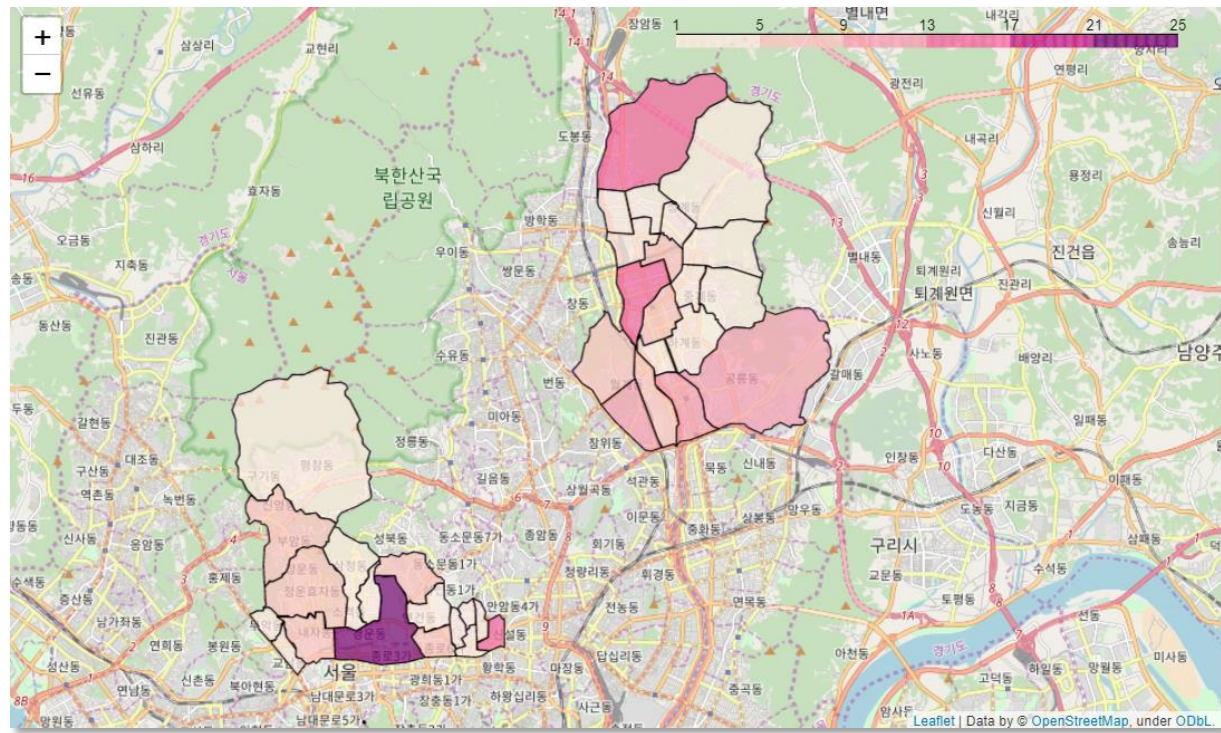
<남성 행정동 별 유동인구 비율>

» 집단 별 유동인구 패턴 및 비율 차이 확인

유통 및 상권 데이터 EDA



<행정동 별 유통지수 비교>

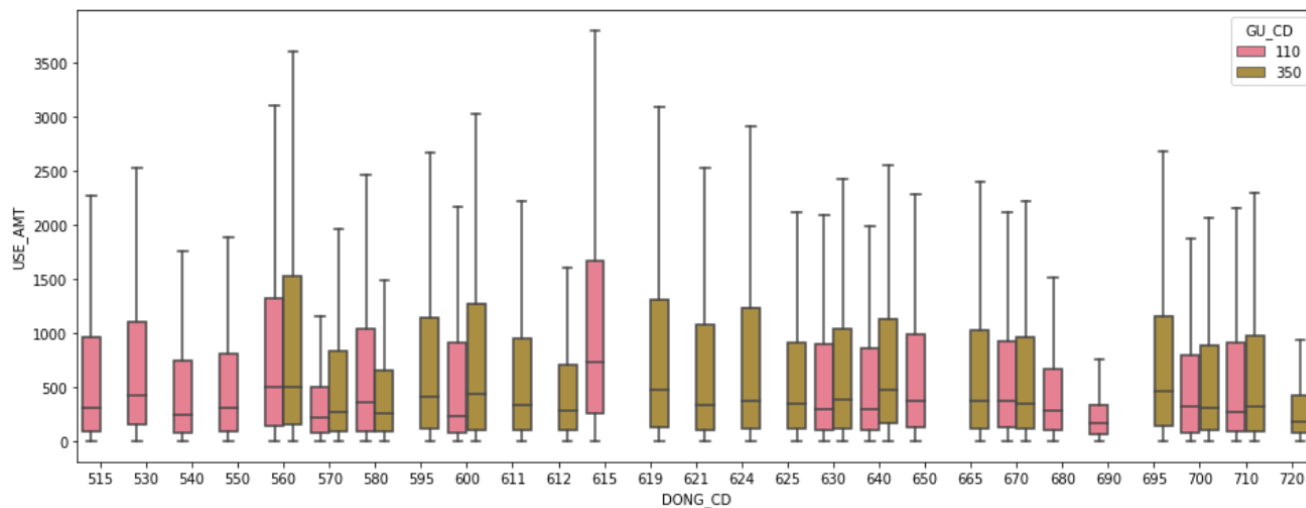


<행정동 별 GS편의점 수 비교>

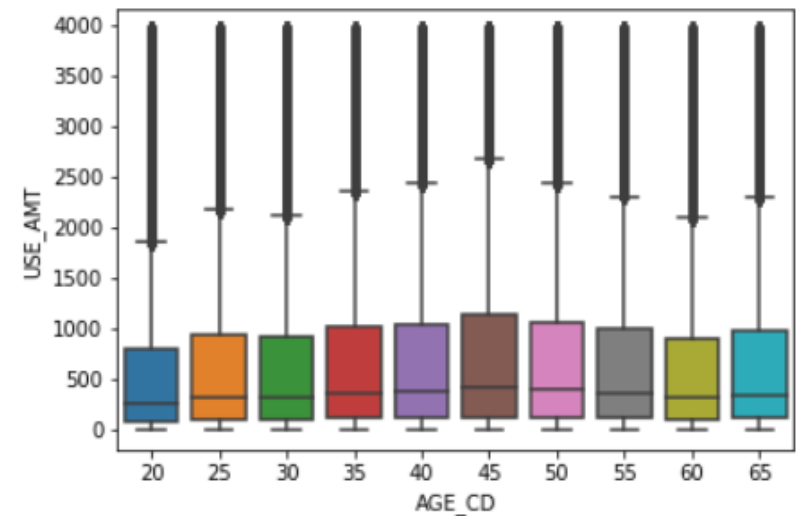


행정동 별 GS편의점 수 데이터를 활용하여 추가 파생 변수 고려

카드매출 데이터 EDA



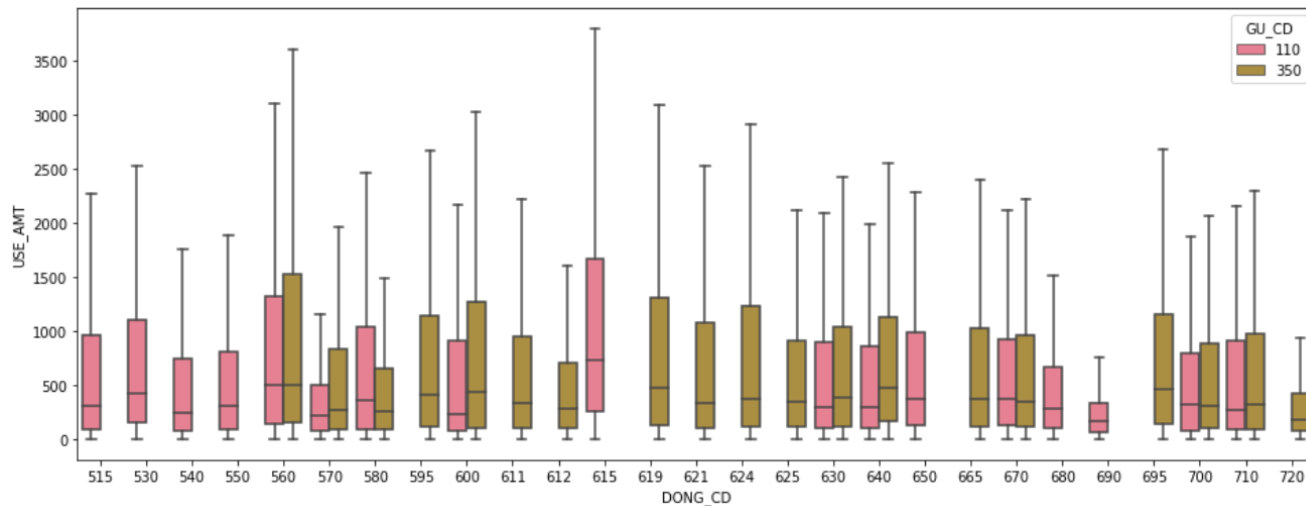
<행정동 별 카드매출 Box-Plots>



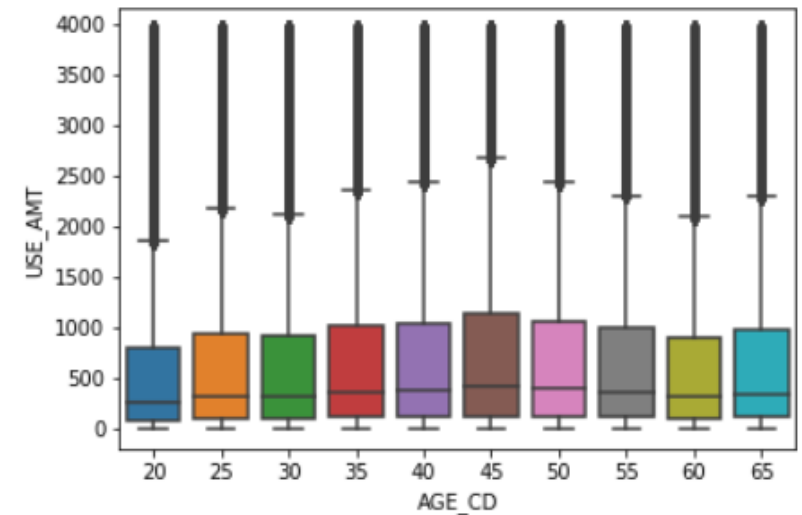
<연령 별 카드매출 Box-Plots>

» 행정동/연령 별 카드매출 분포가 다름

카드매출 데이터 EDA



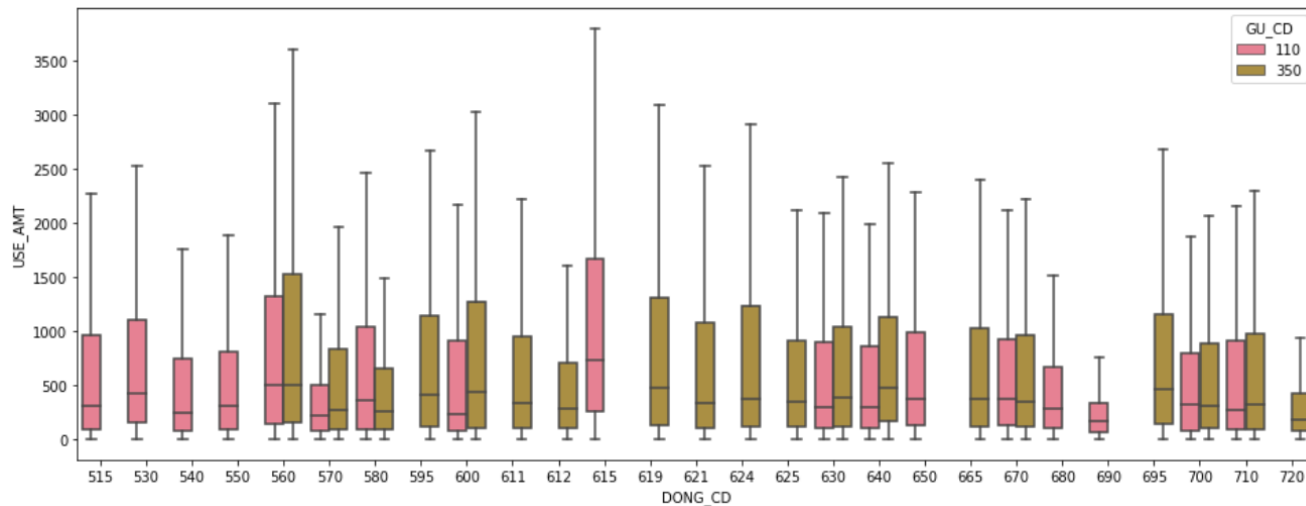
<행정동 별 카드매출 Box-Plots>



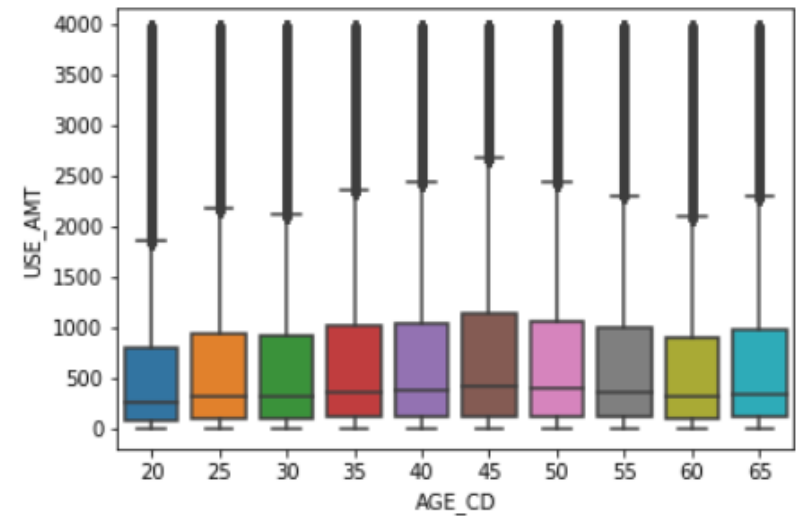
<연령 별 카드매출 Box-Plots>

» 데이터 별 EDA를 수행해본 결과
세분화된 집단들 사이의 차이가 존재한다는 것을 발견했습니다.

카드매출 데이터 EDA



<행정동 별 카드매출 Box-Plots>



<연령 별 카드매출 Box-Plots>

EDA 결과를 토대로 세분화된 집단 별 소비생활을 예측하기 위해 행정동/연령 별 카드매출 분포가 다른 본격적으로 전처리 및 모델링을 수행했습니다!

데이터 전처리

1

MERGE

2

VISUALIZATION

3

FEATURE
ENGINEERING

4

FILL NA_VALUES

5

RESAMPLING

6

DATA
EXPORTING

(예시1) 환경기상 데이터

Merge

노원구 + 종로구 + 데이터 정의서(행정동 관련 파일)

Feature Engineering 1

1. Flag 기준 탐색: Flag = -999 (실내 공기질 측정기 삭제)
2. Date 기준: year, month, day, hour, min, weekday 변수 생성
3. Day, hour 기준: groupby 그래프 → 변수들의 계절성 확인

Fill NA_Value 1

1. Na값 처리
 - 18.04-19.03 기준 대비 비어있는 시간
 - 값: -999, -9999
 - 주최측에서 언급한 이상치 5개
2. 복수 serial: 동 단위로 통합(복수일 경우, serial 별 mean값)

Resampling 1(DownSampling)

1. Min → hour 조정 (groupby로 mean값 대치)

Fill NA_Value 2

3. Na값 채우기

- 결측치 사이 시간 길이 12h이하 → 보간법(linear) 사용
- 결측치 사이의 시간 길이 12h 초과 → 동일 시간의 과거의 최신 데이터 사용

Feature Engineering 2

4. 불쾌지수 변수 생성

5. 미세먼지, 불쾌지수 범주화

- 미세먼지: 좋음, 보통, 나쁨, 매우나쁨
- 불쾌지수: 낮음, 보통, 높음, 매우높음

Resampling 2(DownSampling)

2. Hour → day 조정

- 미세먼지, 불쾌지수 범주: hour 기준, 범주의 다빈도 범주 선택
- 모든 변수: Q1, Q2, Q3, mean 값 활용

Data Exporting

- 가공 데이터/ 환경_edit1.csv

데이터 전처리

1

MERGE

2

VISUALIZATION

3

FEATURE
ENGINEERING

4

FILL NA_VALUES

5

RESAMPLING

6

DATA
EXPORTING

(예시1) 환경기상 데이터

Merge

노원구 + 종로구 + 데이터 정의서(행정동 관련 파일)

Feature Engineering 1

1. Flag 기준 탐색: Flag = -999 (실내 공기질 측정기 삭제)
2. Date 기준: year, month, day, hour, min, weekday 변수 생성
3. Day, hour 기준: groupby 그래프 → 변수들의 계절성 확인

Fill NA_Value 1

1. Na값 처리
 - 18.04-19.03 기준 대비 비어있는 시간
 - 값: -999, -9999
 - 주최측에서 언급한 이상치 5개
2. 복수 serial: 동 단위로 통합(복수일 경우 serial 별 mean값)

Resampling 1(DownSampling)

1. Min → hour 조정 (groupby로 mean값 대입)

Fill NA_Value 2

3. Na값 채우기

- 결측치 사이 시간 길이 12h이하 → 보간법(linear) 사용
- 결측치 사이의 시간 길이 12h 초과 → 동일 시간의 과거의 최신 데이터 사용

Feature Engineering 2

4. 불쾌지수 변수 생성

5. 미세먼지, 불쾌지수 범주화

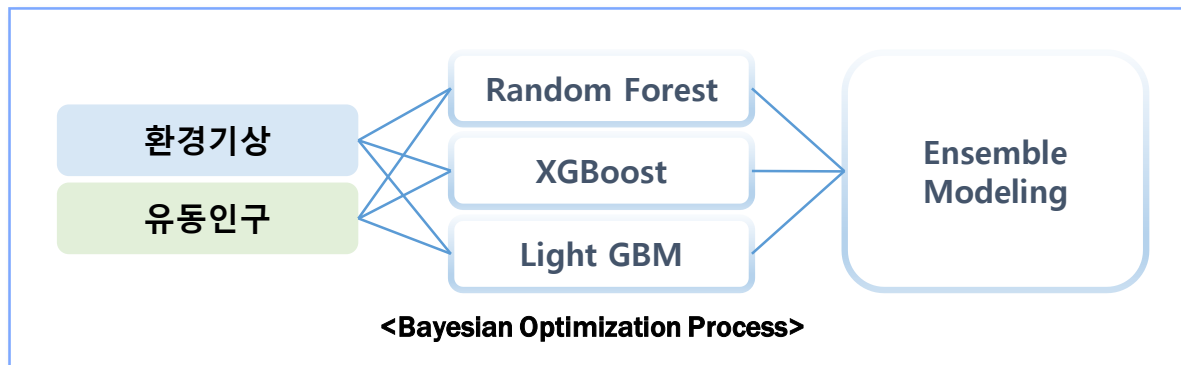
- 미세먼지: 좋음, 보통, 나쁨, 매우나쁨
- 불쾌지수: 낮음, 보통, 높음, 매우높음

Resampling 2(DownSampling)

2. Hour → day 조정

- 미세먼지, 불쾌지수 범주: hour 기준, 범주의 다빈도 범주 선택
- Day 기준: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 376, 377, 378, 379, 380, 381, 382, 383, 384, 385, 386, 387, 388, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399, 400, 401, 402, 403, 404, 405, 406, 407, 408, 409, 410, 411, 412, 413, 414, 415, 416, 417, 418, 419, 420, 421, 422, 423, 424, 425, 426, 427, 428, 429, 430, 431, 432, 433, 434, 435, 436, 437, 438, 439, 440, 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462, 463, 464, 465, 466, 467, 468, 469, 470, 471, 472, 473, 474, 475, 476, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504, 505, 506, 507, 508, 509, 510, 511, 512, 513, 514, 515, 516, 517, 518, 519, 520, 521, 522, 523, 524, 525, 526, 527, 528, 529, 530, 531, 532, 533, 534, 535, 536, 537, 538, 539, 540, 541, 542, 543, 544, 545, 546, 547, 548, 549, 550, 551, 552, 553, 554, 555, 556, 557, 558, 559, 560, 561, 562, 563, 564, 565, 566, 567, 568, 569, 570, 571, 572, 573, 574, 575, 576, 577, 578, 579, 580, 581, 582, 583, 584, 585, 586, 587, 588, 589, 590, 591, 592, 593, 594, 595, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 606, 607, 608, 609, 610, 611, 612, 613, 614, 615, 616, 617, 618, 619, 620, 621, 622, 623, 624, 625, 626, 627, 628, 629, 630, 631, 632, 633, 634, 635, 636, 637, 638, 639, 640, 641, 642, 643, 644, 645, 646, 647, 648, 649, 650, 651, 652, 653, 654, 655, 656, 657, 658, 659, 660, 661, 662, 663, 664, 665, 666, 667, 668, 669, 670, 671, 672, 673, 674, 675, 676, 677, 678, 679, 680, 681, 682, 683, 684, 685, 686, 687, 688, 689, 690, 691, 692, 693, 694, 695, 696, 697, 698, 699, 700, 701, 702, 703, 704, 705, 706, 707, 708, 709, 710, 711, 712, 713, 714, 715, 716, 717, 718, 719, 720, 721, 722, 723, 724, 725, 726, 727, 728, 729, 730, 731, 732, 733, 734, 735, 736, 737, 738, 739, 740, 741, 742, 743, 744, 745, 746, 747, 748, 749, 750, 751, 752, 753, 754, 755, 756, 757, 758, 759, 760, 761, 762, 763, 764, 765, 766, 767, 768, 769, 770, 771, 772, 773, 774, 775, 776, 777, 778, 779, 780, 781, 782, 783, 784, 785, 786, 787, 788, 789, 790, 791, 792, 793, 794, 795, 796, 797, 798, 799, 800, 801, 802, 803, 804, 805, 806, 807, 808, 809, 810, 811, 812, 813, 814, 815, 816, 817, 818, 819, 820, 821, 822, 823, 824, 825, 826, 827, 828, 829, 830, 831, 832, 833, 834, 835, 836, 837, 838, 839, 840, 841, 842, 843, 844, 845, 846, 847, 848, 849, 850, 851, 852, 853, 854, 855, 856, 857, 858, 859, 860, 861, 862, 863, 864, 865, 866, 867, 868, 869, 870, 871, 872, 873, 874, 875, 876, 877, 878, 879, 880, 881, 882, 883, 884, 885, 886, 887, 888, 889, 890, 891, 892, 893, 894, 895, 896, 897, 898, 899, 900, 901, 902, 903, 904, 905, 906, 907, 908, 909, 910, 911, 912, 913, 914, 915, 916, 917, 918, 919, 920, 921, 922, 923, 924, 925, 926, 927, 928, 929, 930, 931, 932, 933, 934, 935, 936, 937, 938, 939, 940, 941, 942, 943, 944, 945, 946, 947, 948, 949, 950, 951, 952, 953, 954, 955, 956, 957, 958, 959, 960, 961, 962, 963, 964, 965, 966, 967, 968, 969, 970, 971, 972, 973, 974, 975, 976, 977, 978, 979, 980, 981, 982, 983, 984, 985, 986, 987, 988, 989, 990, 991, 992, 993, 994, 995, 996, 997, 998, 999, 1000, 1001, 1002, 1003, 1004, 1005, 1006, 1007, 1008, 1009, 1010, 1011, 1012, 1013, 1014, 1015, 1016, 1017, 1018, 1019, 1020, 1021, 1022, 1023, 1024, 1025, 1026, 1027, 1028, 1029, 1030, 1031, 1032, 1033, 1034, 1035, 1036, 1037, 1038, 1039, 1040, 1041, 1042, 1043, 1044, 1045, 1046, 1047, 1048, 1049, 1050, 1051, 1052, 1053, 1054, 1055, 1056, 1057, 1058, 1059, 1060, 1061, 1062, 1063, 1064, 1065, 1066, 1067, 1068, 1069, 1070, 1071, 1072, 1073, 1074, 1075, 1076, 1077, 1078, 1079, 1080, 1081, 1082, 1083, 1084, 1085, 1086, 1087, 1088, 1089, 1090, 1091, 1092, 1093, 1094, 1095, 1096, 1097, 1098, 1099, 1100, 1101, 1102, 1103, 1104, 1105, 1106, 1107, 1108, 1109, 1110, 1111, 1112, 1113, 1114, 1115, 1116, 1117, 1118, 1119, 1120, 1121, 1122, 1123, 1124, 1125, 1126, 1127, 1128, 1129, 1130, 1131, 1132, 1133, 1134, 1135, 1136, 1137, 1138, 1139, 1140, 1141, 1142, 1143, 1144, 1145, 1146, 1147, 1148, 1149, 1150, 1151, 1152, 1153, 1154, 1155, 1156, 1157, 1158, 1159, 1160, 1161, 1162, 1163, 1164, 1165, 1166, 1167, 1168, 1169, 1170, 1171, 1172, 1173, 1174, 1175, 1176, 1177, 1178, 1179, 1180, 1181, 1182, 1183, 1184, 1185, 1186, 1187, 1188, 1189, 1190, 1191, 1192, 1193, 1194, 1195, 1196, 1197, 1198, 1199, 1200, 1201, 1202, 1203, 1204, 1205, 1206, 1207, 1208, 1209, 1210, 1211, 1212, 1213, 1214, 1215, 1216, 1217, 1218, 1219, 1220, 1221, 1222, 1223, 1224, 1225, 1226, 1227, 1228, 1229, 1230, 1231, 1232, 1233, 1234, 1235, 1236, 1237, 1238, 1239, 1240, 1241, 1242, 1243, 1244, 1245, 1246, 1247, 1248, 1249, 1250, 1251, 1252, 1253, 1254, 1255, 1256, 1257, 1258, 1259, 1260, 1261, 1262, 1263, 1264, 1265, 1266, 1267, 1268, 1269, 1270, 1271, 1272, 1273, 1274, 1275, 1276, 1277, 1278, 1279, 1280, 1281, 1282, 1283, 1284, 1285, 1286, 1287, 1288, 1289, 1290, 1291, 1292, 1293, 1294, 1295, 1296, 1297, 1298, 1299, 1300, 1301, 1302, 1303, 1304, 1305, 1306, 1307, 1308, 1309, 1310, 1311, 1312, 1313, 1314, 1315, 1316, 1317, 1318, 1319, 1320, 1321, 1322, 1323, 1324, 1325, 1326, 1327, 1328, 1329, 1330, 1331, 1332, 1333, 1334, 1335, 1336, 1337, 1338, 1339, 1340, 1341, 1342, 1343, 1344, 1345, 1346, 1347, 1348, 1349, 1350, 1351, 1352, 1353, 1354, 1355, 1356, 1357, 1358, 1359, 1360, 1361, 1362, 1363, 1364, 1365, 1366, 1367, 1368, 1369, 1370, 1371, 1372, 1373, 1374, 1375, 1376, 1377, 1378, 1379, 1380, 1381, 1382, 1383, 1384, 1385, 1386, 1387, 1388, 1389, 1390, 1391, 1392, 1393, 1394, 1395, 1396, 1397, 1398, 1399, 1400, 1401, 1402, 1403, 1404, 1405, 1406, 1407, 1408, 1409, 1410, 1411, 1412, 1413, 1414, 1415, 1416, 1417, 1418, 1419, 1420, 1421, 1422, 1423, 1424, 1425, 1426, 1427, 1428, 1429, 1430, 1431, 1432, 1433, 1434, 1435, 1436, 1437, 1438, 1439, 1440, 1441, 1442, 1443, 1444, 1445, 1446, 1447, 1448, 1449, 1450, 1451, 1452, 1453, 1454, 1455, 1456, 1457, 1458, 1459, 1460, 1461, 1462, 1463, 1464, 1465, 1466, 1467, 1468, 1469, 1470, 1471, 1472, 1473, 1474, 1475, 1476, 1477, 1478, 1479, 1480, 1481, 1482, 1483, 1484, 1485, 1486, 1487, 1488, 1489, 1490, 1491, 1492, 1493, 1494, 1495, 1496, 1497, 1498, 1499, 1500, 1501, 1502, 1503, 1504, 1505, 1506, 1507, 1508, 1509, 1510, 1511, 1512, 1513, 1514, 1515, 1516, 1517, 1518, 1519, 1520, 1521, 1522, 1523, 1524, 1525, 1526, 1527, 1528, 1529, 1530, 1531, 1532, 1533, 1534, 1535, 1536, 1537, 1538, 1539, 1540, 1541, 1542, 1543, 1544, 1545, 1546, 1547, 1548, 1549, 1550, 1551, 1552, 1553, 1554, 1555, 1556, 1557, 1558, 1559, 1560, 1561, 1562, 1563, 1564, 1565, 1566, 1567, 1568, 1569, 1570, 1571, 1572, 1573, 1574, 1575, 1576, 1577, 1578, 1579, 1580, 1581, 1582, 1583, 1584, 1585, 1586, 1587, 1588, 1589, 1590, 1591, 1592, 1593, 1594, 1595, 1596, 1597, 1598, 1599, 1600, 1601, 1602, 1603, 1604, 1605, 1606, 1607, 1608, 1609, 1610, 1611, 1612, 1613, 1614, 1615, 1616, 1617, 1618, 1619, 1620, 1621, 1622, 1623, 1624, 1625, 1626, 1627, 1628, 1629, 1630, 1631, 1632, 1633, 1634, 1635, 1636, 1637, 1638, 1639, 1640, 1641, 1642, 1643, 1644, 1645, 1646, 1647, 1648, 1649, 1650, 1651, 1652, 1653, 1654, 1655, 1656, 1657, 1658, 1659, 1660, 1661, 1662, 1663, 1664, 1665, 1666, 1667, 1668, 1669, 1670, 1671, 1672, 1673, 1674, 1675, 1676, 1677, 1678, 1679, 1680, 1681, 1682, 1683, 1684, 1685, 1686, 1687, 1688, 1689, 1690, 1691, 1692, 1693, 1694, 1695, 1696, 1697, 1698, 1699, 1700, 1701, 1702, 1703, 1704, 1705, 1706, 1707, 1708, 1709, 1710, 1711, 1712, 1713, 1714, 1715, 1716, 1717, 1718, 1719, 1720, 1721, 1722, 1723, 1724, 1725, 1726, 1727, 1728, 1729, 1730, 1731, 1732, 1733, 1734, 1735, 1736, 1737, 1738, 1739, 1740, 1741, 1742, 1743, 1744, 1745, 1746, 1747, 1748, 1749, 1750, 1751, 1752, 1753, 1754, 1755, 1756, 1757, 1758, 1759, 1760, 1761, 1762, 1763, 1764, 1765, 1766, 1767, 1768, 1769, 1770, 1771, 1772, 1773, 1774, 1775, 1776, 1777, 1778, 1779, 1780, 1781, 1782, 1783, 1784, 1785, 1786, 1787, 1788, 1789, 1790, 1791, 1792, 1793, 1794, 1795, 1796, 1797, 1798, 1799, 1800, 1801, 1802, 1803, 1804, 1805, 1806, 1807, 1808, 1809, 1810, 1811, 1812, 1813, 1814, 1815, 1816, 1817, 1818, 1819, 1820, 1821, 1822, 1823, 1824, 1825, 1826, 1827, 1828, 1829, 1830, 1831, 1832, 1833, 1834, 1835, 1836, 1837, 1838, 1839, 1840, 1841, 1842, 1843, 1844, 1845, 1846, 1847, 1848, 1849, 1850, 1851, 1852, 1853, 1854, 1855, 1856, 1857, 1858, 1859, 1860, 1861, 1862, 1863, 1864, 1865, 1866, 1867, 1868, 1869, 1870, 1871, 1872, 1873, 1874, 1875, 1876, 1877, 1878, 1879, 1880, 1881, 1882, 1883, 1884, 1885, 1886, 1887, 1888, 1889, 1890, 1891, 1892, 1893, 1894, 1895, 1896, 1897, 1898, 1899, 1900, 1901, 1902, 1903, 1904, 1905, 1906, 1907, 1908, 1909, 1910, 1911, 1912, 1913, 1914, 1915, 1916, 1917, 1918, 1919, 1920, 1921, 1922, 1923, 1924, 1925, 1926, 1927, 1928, 1929, 1930, 1931, 1932, 1933, 1934, 1935, 1936, 1937, 1938, 1939, 1940, 1941, 1942, 1943, 1944, 1945, 1946, 1947, 1948, 1949, 1950, 1951, 1952, 1953, 1954, 1955, 1956, 1957, 1958, 1959, 1960, 1961, 1962, 1963, 1964, 1965, 1966, 1967, 1968, 1969, 1970, 1971, 1972, 1973, 1974, 1975, 1976, 1977, 1978, 1979

▶ 유동인구 예측을 위한 앙상블 모델링



유동인구 예측 예측 모델 구조

$$\widehat{Y}_{ens} = \alpha \widehat{Y}_{RF} + \beta \widehat{Y}_{XGB} + \gamma \widehat{Y}_{LGB}$$

<Ensemble Model Algorithm>

$$\alpha = \frac{ACC_{RF}}{ACC_{RF} + ACC_{XGB} + ACC_{LGB}}$$

$$\beta = \frac{ACC_{XGB}}{ACC_{RF} + ACC_{XGB} + ACC_{LGB}}$$

$$\gamma = \frac{ACC_{LGB}}{ACC_{RF} + ACC_{XGB} + ACC_{LGB}}$$

<Ensemble Coefficient>

$$ACC_{RF} = 1 - MAPE_{RF}$$

$$ACC_{XGB} = 1 - MAPE_{XGB}$$

$$ACC_{LGB} = 1 - MAPE_{LGB}$$

RF : Random Forest

XGB : XGBoost

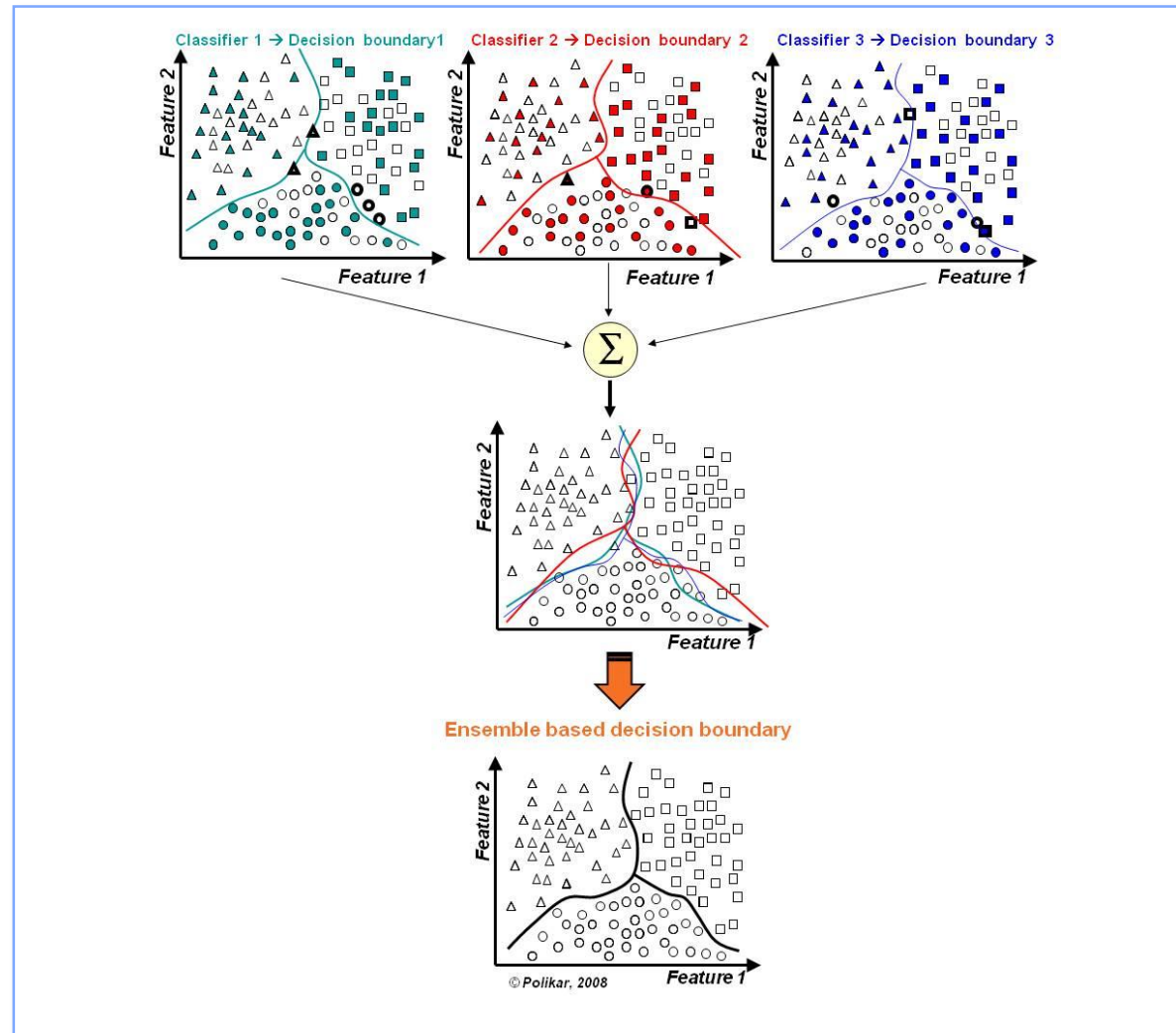
LGB : LightGBM

MAPE : Mean Absolute Percentage Error

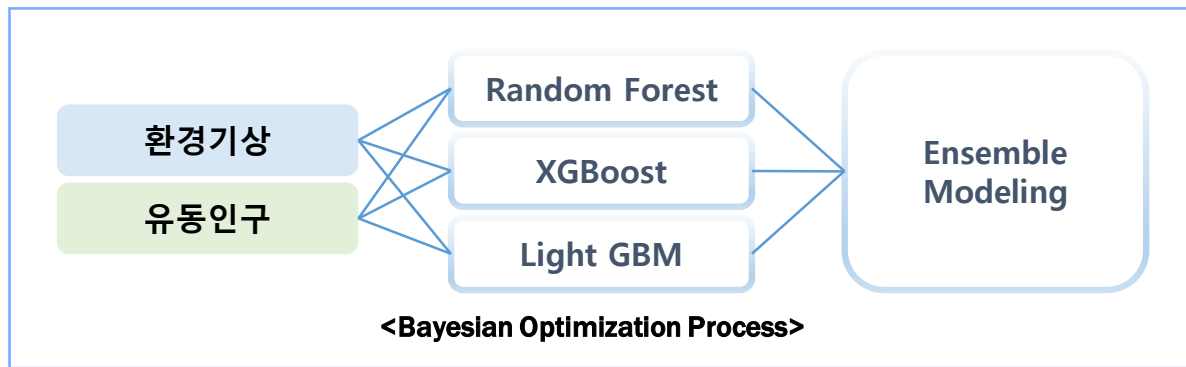
< Definition of Accuracy>

앙상블 모델 알고리즘

▶ 앙상블 모델(Ensemble Model) 선정 이유



▶ 유동인구 예측을 위한 앙상블 모델링



유동인구 예측 예측 모델 구조

$$\widehat{Y}_{ens} = \alpha \widehat{Y}_{RF} + \beta \widehat{Y}_{XGB} + \gamma \widehat{Y}_{LGB}$$

<Ensemble Model Algorithm>

$$\alpha = \frac{ACC_{RF}}{ACC_{RF} + ACC_{XGB} + ACC_{LGB}}$$

$$\beta = \frac{ACC_{XGB}}{ACC_{RF} + ACC_{XGB} + ACC_{LGB}}$$

$$\gamma = \frac{ACC_{LGB}}{ACC_{RF} + ACC_{XGB} + ACC_{LGB}}$$

<Ensemble Coefficient>

$$ACC_{RF} = 1 - MAPE_{RF}$$

$$ACC_{XGB} = 1 - MAPE_{XGB}$$

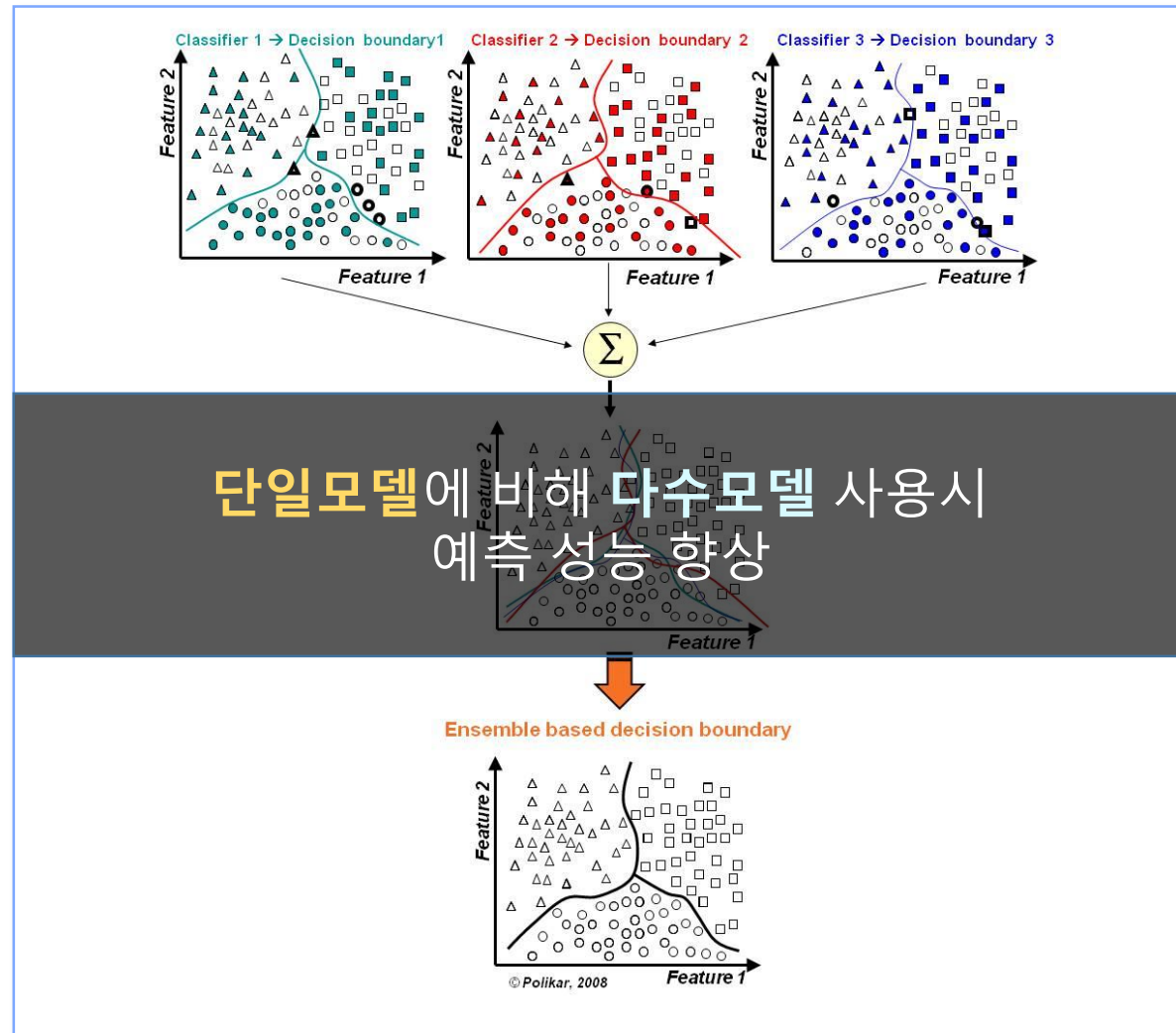
$$ACC_{LGB} = 1 - MAPE_{LGB}$$

RF : Random Forest
XGB : XGBoost
LGB : LightGBM
MAPE : Mean Absolute Percentage Error

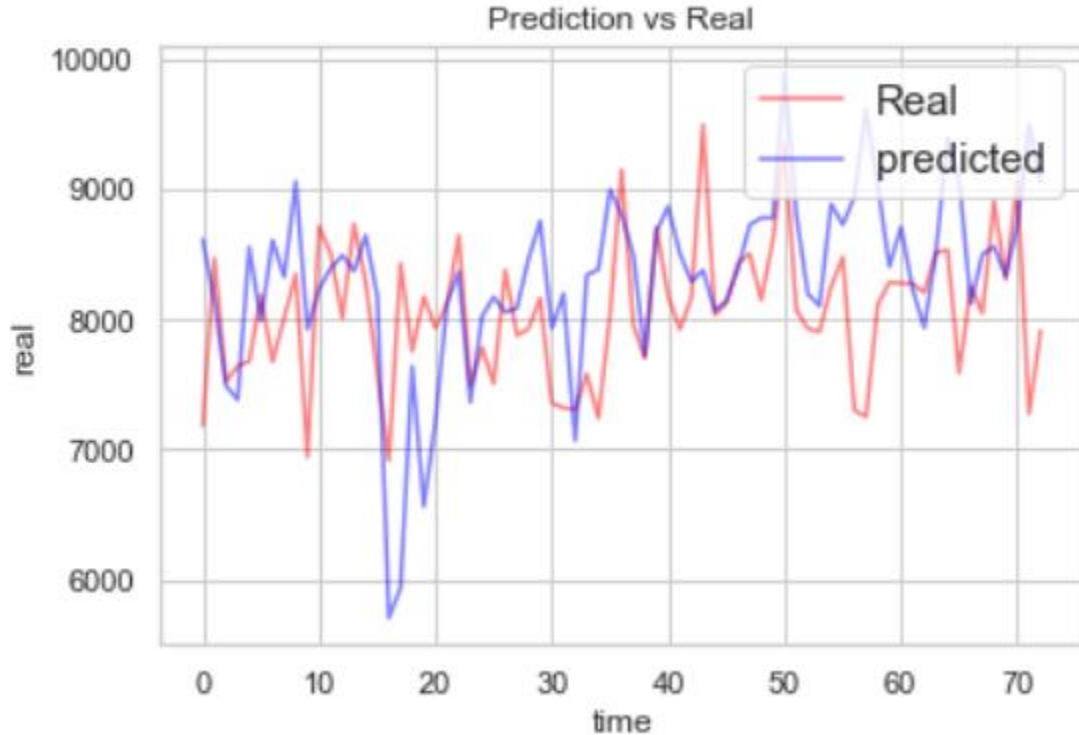
< Definition of Accuracy>

앙상블 모델 알고리즘

▶ 앙상블 모델(Ensemble Model) 선정 이유



▶ 모델 예측 결과 및 인사이트 도출



» ‘상계1동 / 60-64세 / 남성’ 집단
(19년도 1분기 유동인구)

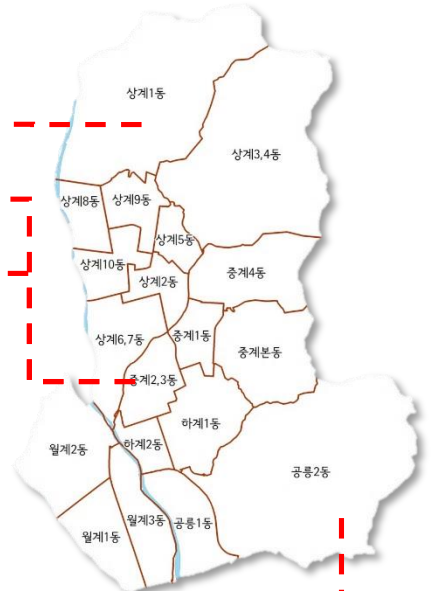
모델 예측 결과

유동인구 예측력이 높은 집단(정확도 높음)

상계1동 / 60-64세 / 남성 (91.6%)

중계2,3동 / 45-49세 / 남성 (91.4%)

중계2,3동 / 50-54세 / 남성 (91.3%)



<노원구 행정동>

유동인구 예측력이 낮은 집단

공릉2동 / 20-24세 / 여성 (40.8%)

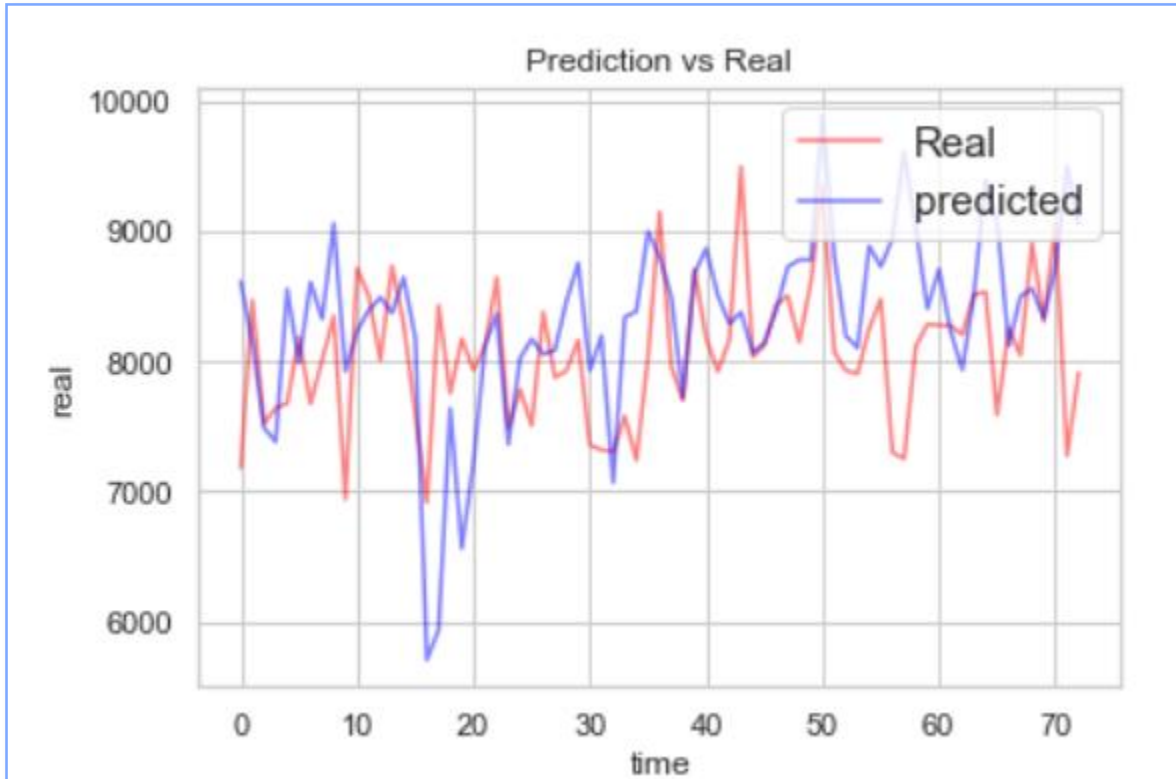
가회동 / 40-44세 / 남성 (43.7%)

가회동 / 35-39세 / 남성 (43.8%)

<종로구 행정동>

인사이트 도출

▶ 모델 예측 결과 및 인사이트 도출

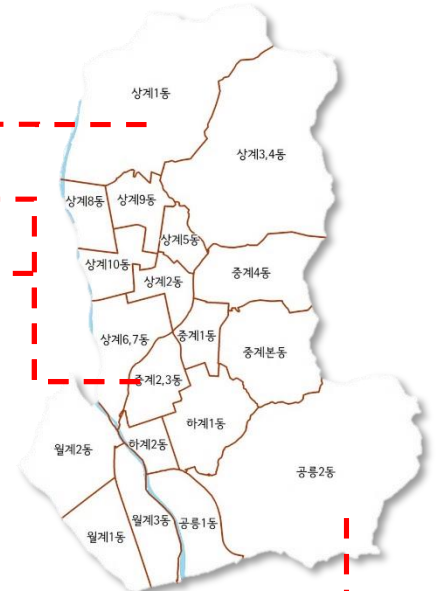


유동인구 예측력이 높은 집단(정확도 %)

상계1동 / 60-64세 / 남성 (91.6%)

중계2,3동 / 45-49세 / 남성 (91.4%)

중계2,3동 / 50-54세 / 남성 (91.3%)



<노원구 행정동>

유동인구 예측력이 낮은 집단

공릉2동 / 20-24세 / 여성 (40.8%)

가회동 / 40-44세 / 남성 (43.7%)

가회동 / 35-39세 / 남성 (43.8%)



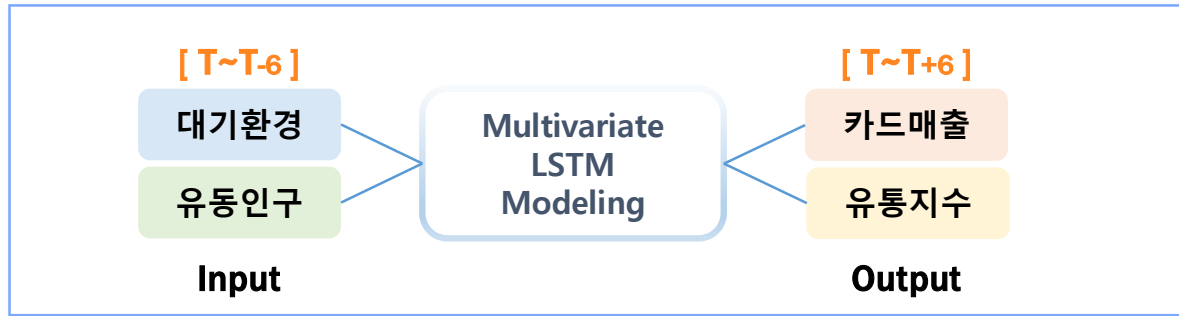
<종로구 행정동>

‘상계1동 / 60-64세 / 남성’ 집단

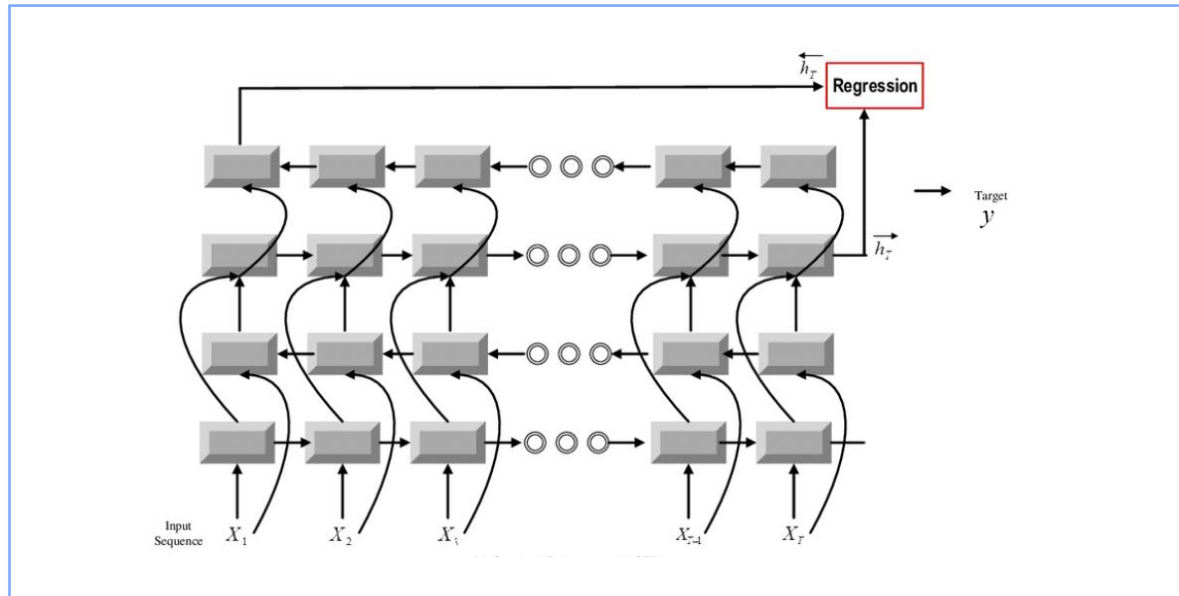
(19년도 1분기 유동인구)

앙상블 모델링 결과 대기환경 변화에 따른 유동인구 예측력이 높은 집단과 낮은 집단이 발견되기도 했습니다.

▶ 카드매출 및 유통지수 예측 다변량 LSTM 모델링

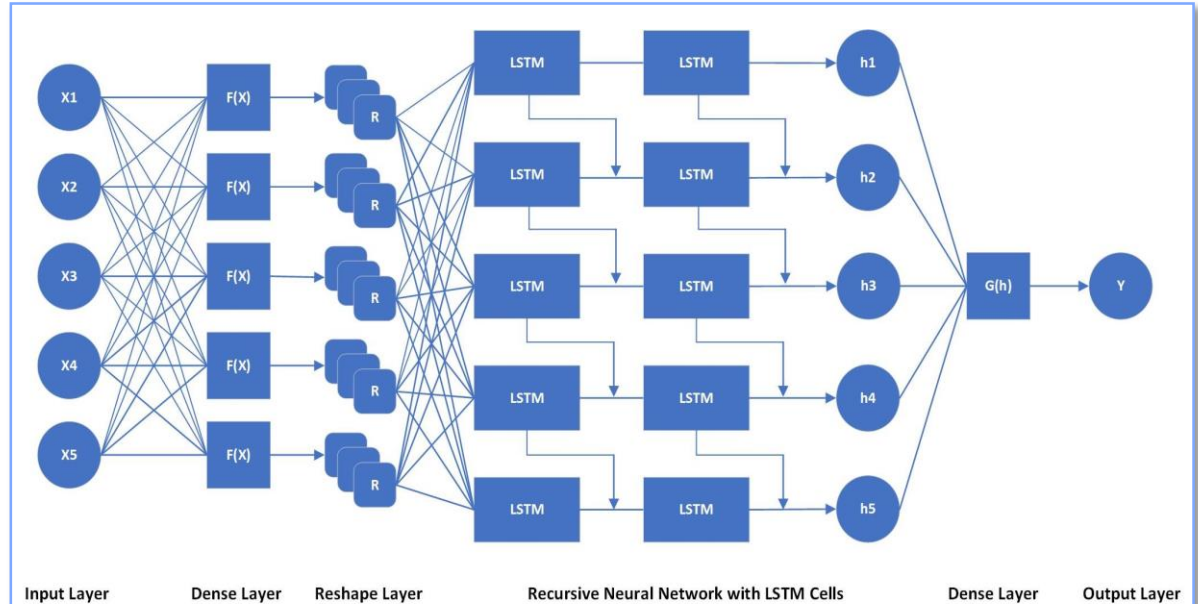


카드매출 및 유통지수 예측 모델링



다변량 LSTM 모델 알고리즘

▶ 다변량 LSTM 모델 선정 이유



향후 비즈니스 활용을 고려하여,
다변량 입력변수들의 **시계열성(환경 변화 추이, 유통인구 변화 추이 등)**을 반영할 수 있는 모델 필요



SNS 데이터

- 미세먼지 관련 뉴스 데이터
- 미세먼지 관련 블로그 데이터
- 미세먼지 관련 카페 데이터



유동인구 데이터

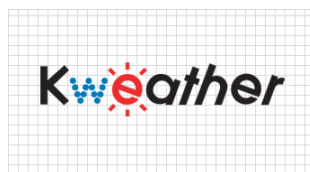
- 노원구 유동인구 데이터
- 종로구 유동인구 데이터



신한카드

카드매출 데이터

- 노원구 카드매출 데이터
- 종로구 카드매출 데이터



미세먼지 데이터

- 노원구 미세먼지 측정 데이터
- 종로구 미세먼지 측정 데이터



유통 데이터

- 노원구 GS25 유통 데이터
- 종로구 GS25 유통 데이터



소상공인시장진흥공단

상권 데이터

- 노원구 상가 업소 데이터
- 종로구 상가 업소 데이터

비즈니스 모델 활용



SNS 데이터

- 미세먼지 관련 뉴스 데이터
- 미세먼지 관련 블로그 데이터
- 미세먼지 관련 카페 데이터



유동인구 데이터

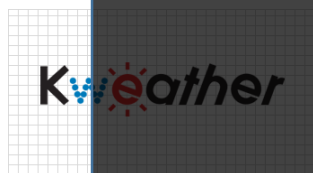
- 노원구 유동인구 데이터
- 종로구 유동인구 데이터



신한카드

카드매출 데이터

- 노원구 카드매출 데이터
- 종로구 카드매출 데이터



미세먼지 데이터

- 노원구 미세먼지 측정 데이터
- 종로구 미세먼지 측정 데이터

보유한 데이터의
활용 및 확장의 어려움

유통 데이터

- 노원구 GS25 유통 데이터
- 종로구 GS25 유통 데이터



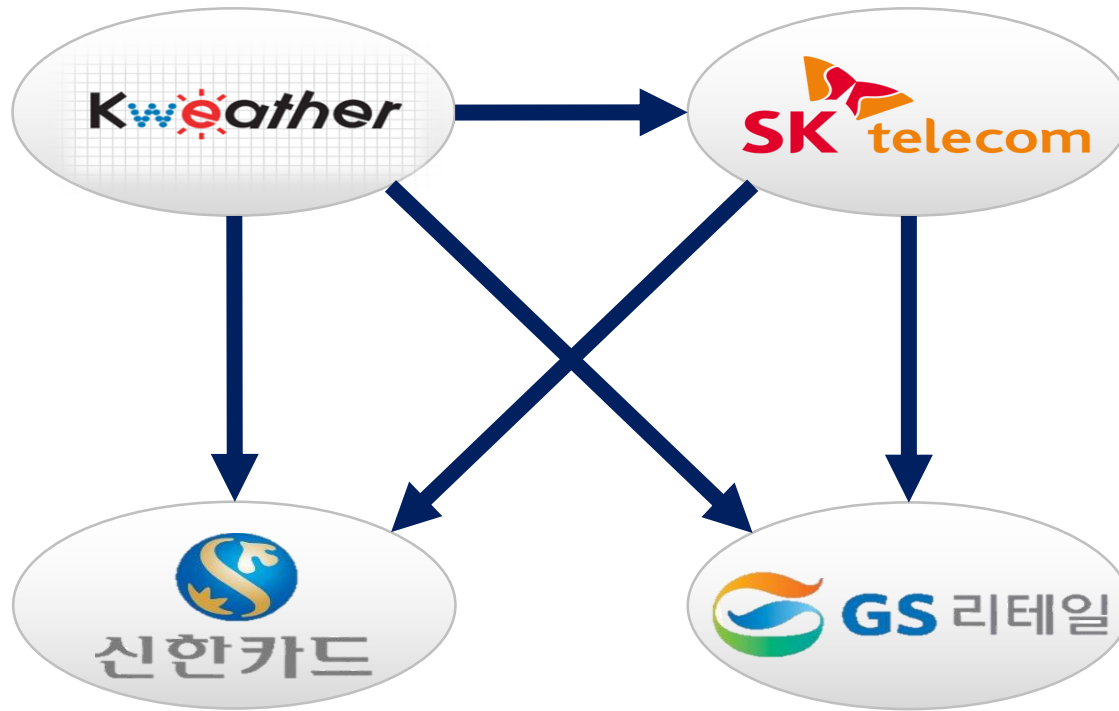
소상공인시장진흥공단

상권 데이터

- 노원구 상가 업소 데이터
- 종로구 상가 업소 데이터

비즈니스 모델 활용

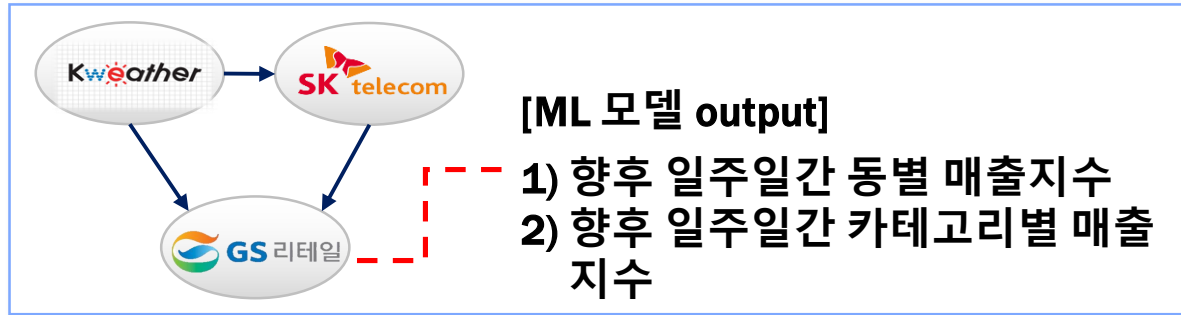
▶ WIN WIN BM



» 데이터 판매 및 비즈니스에 연계

비즈니스 모델 활용

▶ ML 모델 사용한 데이터



ML모델 Meaning

- 1) 전반적인 동별 재고 수요량 예측
- 2) 카테고리별 수요량이 많은 품목(SKU-stock keeping unit)을 사전에 알 수 있음

ML모델 Utilizing

- 1) 미세먼지 발생 시나리오에 따른 동 및 카테고리별 리포팅 서비스

▶ Development Direction

1)



국지적 적용(종로구, 노원구)

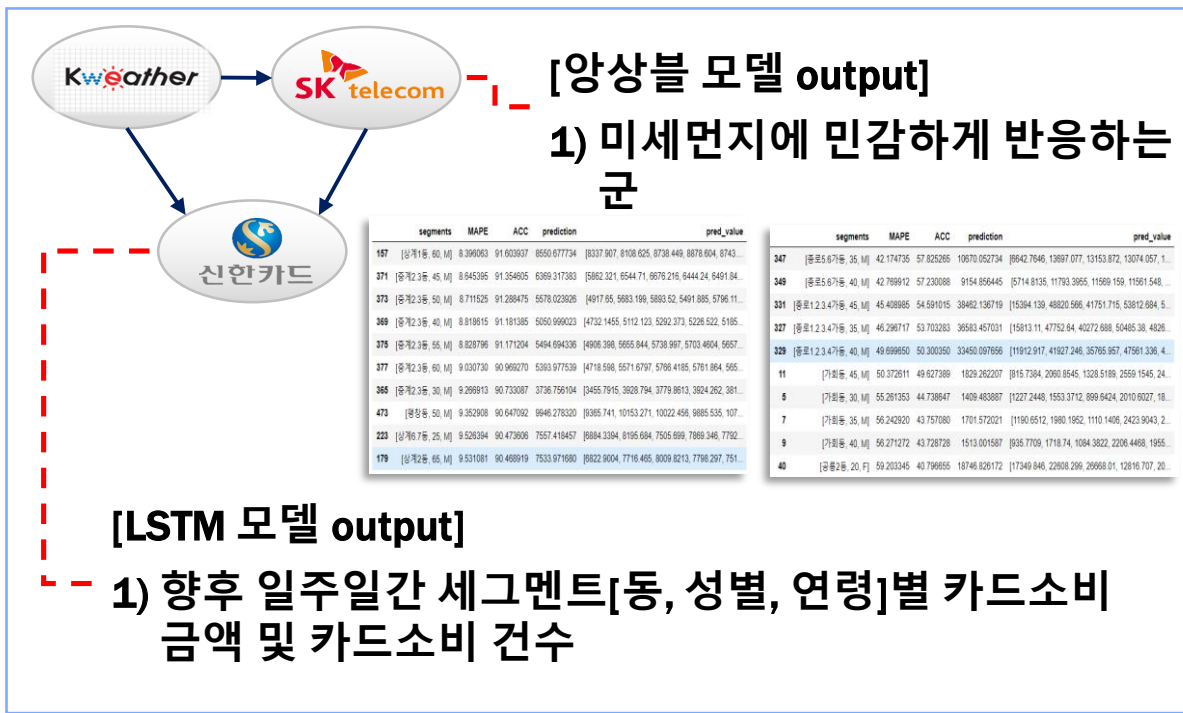


일반적 적용(전국)

2)

미세먼지 시나리오와
다른 시나리오에 맞는 BM 생성

▶ ML 모델 사용한 데이터



LSTM 모델 Meaning

- 1) 세그먼트별 카테고리에 대한 카드소비금액 변동성 파악
- 2) 세그먼트별 카테고리에 대한 소비 선호도 파악

LSTM모델+앙상블 모델(연계) Utilizing

- 1) 개별화된 마케팅: 앙상블 모델에서 도출된 민감군에 속한 개인에게 특정 카테고리 추천
- 2) 위치정보 및 소상공인 데이터를 사용하여 세그먼트별 소비선호도에 맞는 상점 추천

▶ Development Direction

1)



국지적 적용(종로구, 노원구)



일반적 적용(전국)

감사합니다!