수요분석과 예측으로 돌봄시설 현황파악 우선배치 지역선정 및 웹페이지 구현

서비스명 : 온종일 돌봄24
팀명 : 워킹맘

온종일 돌봄24 made by 워킹맘

목차

1 돌봄수요에 대한 필요성과 문제인식

2 데이터 분석을 통한 돌봄수요 모델링

3 모델링을 활용한 비즈니스 아이디어 구현

I. 문제인식 및 필요성과 서비스 구현 목적

문제인식 및 필요성

- 2009-2019년 '초등돌봄교실' 참여 학생 수 약 5배 증가
- 정부 온종일 돌봄정책 발표 ('22년까지 총 3,500실 확대)
- 코로나 19에 따른 사회적 거리두기 강화로 돌봄 공간 확대 필요성 대두
- 방역소독과 돌봄 병행 부담 가중에 따른 돌봄 전문 인력 부족

▶지역별 혹은 시군구별 학생수요를 고려한 선별적 우선설치가 필요

- 정부 핵심 과제로 지역사회와 함께하는 돌봄 서비스 강화
- 교육부 돌봄 서비스 사업 추진 계획 발표(2020.08.11)
- 학교에서 돌봄 서비스가 가능한 교실을 제공하면 지자체가 돌봄 운영

▶ 초등돌봄과 마을돌봄에 대한 통합적인 관리 필요

서비스 구현 목적

- 돌봄시설* 현황에 대한 데이터분석을 통해
  지역별, 시군구별 수요를 정확히 반영하여 설치가 되었는지 확인
  *초등돌봄교실/지역아동센터/다함께돌봄센터/청소년방과후아카데미

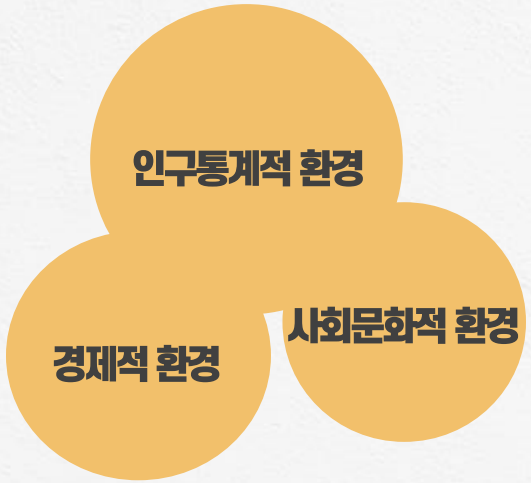
- 돌봄시설이 필요한 지역구에 예측모델링을 적용하여
  수요 파악과 시각화 진행

- 현재 학교돌봄과 마을돌봄 관련 공시정보를 같이 제시하는
  서비스가 없는 상황이므로 지자체별
  통합적인 현황을 보여줄 수 있는 서비스를 구현

Ⅱ. 다중선형 회귀분석(Multiple linear Regression)

1단계 데이터 수집 및 준비

돌봄시설에 영향을 미치는 요소는?



분석방향

1. 데이터 전처리 및 EDA

⬇

2. 회귀분석을 이용한 모델 구축

⬇

3. 예측값과 비교하여 돌봄현황 파악

⬇

4. 모델링을 통한 웹 서비스 구현

구분	의미	출처
학생수	전국의 초등학생 수	교육통계서비스 (https://kess.kedu.re.kr/index)
비만지수(BMI)	몸무게를 키의 제곱으로 나눈 값	공공데이터포털 (https://www.data.go.kr/data/15037481/fileData.do)
음주율	최근 1년 동안 한 달에 1회 이상 술을 마신 적이 있는 사람의 분율(%)	국가통계포털 (https://kosis.kr/)
교육경비보조금	지자체의 교육투자 예산	지방재정365 (http://lofin.mois.go.kr/portal/main.do)
한부모가구수	한부모가족지원관련법에 의한 급여 또는 서비스를 받는 수급가구	복지로(http://www.bokjiro.go.kr/)
도서관수	학교도서관과 공공도서관수의 합	국가도서관통계시스템 (https://www.libsta.go.kr/)

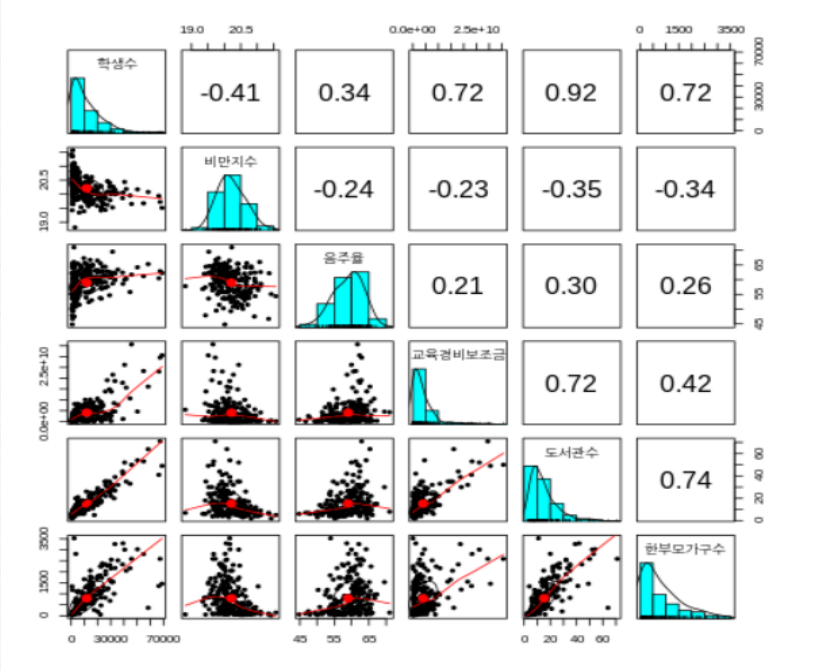
돌봄시설에 영향을 미치는 요인이 단 하나일 가능성은 희박 따라서 돌봄시설에 미치는 다양한 독립변수를 분석 이에 따른 '다중선형회귀분석' 모델을 사용

돌봄시설에 미치는 요인을 크게 3가지 환경으로 분류

인구통계적환경으로 학생수와 한부모가구수 , 경제적환경으로 지역별 교육경비보조금과 도서관수, 사회문화적환경으로 비만지수와 음주율 요인 분석시도

Ⅱ. 다중선형 회귀분석(Multiple linear Regression)

2단계 데이터 탐색(EDA)



	시군구별(1)	시군구별(2)	학생수대비초등돌봄교실수
214	경상남도	창원시	0.000497
58	인천광역시	중구	0.000710
218	경상남도	양산시	0.000767
53	인천광역시	동구	0.001197
220	경상남도	진주시	0.001242
60	광주광역시	남구	0.001773
47	대구광역시	수성구	0.002637
25	부산광역시	강서구	0.002638
77	경기도	과천시	0.002776
48	대구광역시	중구	0.002821

	시군구별(1)	시군구별(2)	학생수대비지역아동센터수
0	서울특별시	강남구	0.000154
25	부산광역시	강서구	0.000293
74	세종특별자치시	세종시	0.000441
14	서울특별시	서초구	0.000441
58	인천광역시	중구	0.000473
102	경기도	용인시	0.000508
51	인천광역시	미추홀구	0.000512
203	경상북도	울진군	0.000538
105	경기도	화성시	0.000556
209	경상남도	거제시	0.000591

- 산포도 행렬을 통한 각 변수간 상관관계 확인
- 인구통계적 환경 요인과 경제적 환경요인 간에 강한 양(+)의 상관관계를 가지고 있음.
- 비만지수는 다른 모든 독립변수와 음(-)의 상관관계를 나타냄

- 학생수 대비 초등 돌봄교실수가 가장 적은 행정구는 경상남도 창원시
- 학생수 대비 지역 아동센터수가 가장 적은 행정구는 서울특별시 강남구
- 강남구의 지역아동센터수가 적은 이유로 지역내 소득과 연관이 있을 가능성이 있고 이에 지역내 총생산 데이터를 활용해 추가적인 검증이 필요

Ⅱ. 다중선형 회귀분석(Multiple linear Regression)

3단계 데이터 모델링 및 성능평가

```
df.null <- lm(돌봄시설합계 ~ 1,  data = df)
df.full <- lm(돌봄시설합계 ~ 학생수+비만지수+음주율+교육경비보조금+도서관수+한부모가구수 ,  data = df)

vif(df_model)

step(df.null, scope = list(upper = df.full) ,  data = df, direction = "both")

df_model <- lm(돌봄시설합계 ~학생수+비만지수+음주율+교육경비보조금+도서관수+한부모가구수, data = df)

summary(df_model)
```

```
Step:  AIC=1410.01
돌봄시설합계 ~ 학생수 + 한부모가구수 + 음주율 +
    비만지수 + 도서관수 + 교육경비보조금

                  Df Sum of Sq    RSS    AIC
<none>                         104005 1410.0
- 교육경비보조금   1      1000 105005 1410.2
- 도서관수         1      1282 105287 1410.8
- 비만지수         1      2212 106217 1412.8
- 음주율           1      3353 107359 1415.2
- 학생수           1     18988 122993 1446.2
- 한부모가구수     1     84792 188798 1544.0
```

```
Call:
lm(formula = 돌봄시설합계 ~ 학생수 + 비만지수 +
    음주율 + 교육경비보조금 + 도서관수 + 한부모가구수,
    data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-71.997 -10.228  -1.138   8.986  83.947

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)    -9.622e+01  8.051e+01  -1.195  0.23333
학생수          1.973e-03  3.107e-04   6.352  1.2e-09 ***
비만지수        8.068e+00  3.721e+00   2.168  0.03122 *
음주율         -9.383e-01  3.515e-01  -2.669  0.00816 **
교육경비보조금  5.910e-10  4.054e-10   1.458  0.14634
도서관수        5.951e-01  3.605e-01   1.651  0.10024
한부모가구수    3.927e-02  2.926e-03  13.423  < 2e-16 ***
---
Signif. codes:  0  '***'  0.001  '**'  0.01  '*'  0.05  '.'  0.1  ' '  1

Residual standard error: 21.69 on 221 degrees of freedom
Multiple R-squared:  0.8807,    Adjusted R-squared:  0.8775
F-statistic:   272 on 6 and 221 DF,  p-value: < 2.2e-16
```

P-value
: 추정된계수가 실제 0일 확률 추정치

Signif. Codes
: 독립변수가 기각될 가능성

결정계수
: 모델이 실제 관측치의 분산을 얼마나 잘 반영하고 있는지를 판단

F-통계량
: 모집단 a가 0일 확률로 모델의 통계적 유의성 판단

- VIF 분석을 진행하여 각 독립변수들 간의 multicollinearity(다중공선성) 확인 및 제거
- 종속변수와 독립변수로 선형회귀식을 세우고, R을 이용하여 독립변수 선택 Stepwise 기법 위주로 사용 타당한 종속변수들을 선택하며 최종 모델 구축

Ⅱ. 다중선형 회귀분석(Multiple linear Regression)

4단계 다중선형회귀모델 예측

인천광역시 미추홀구
MICHUHOL-GU INCHEON METROPOLITAN CITY

학생수 : 29286
초등돌봄교실 : 87
지역아동센터 : 15
다함께돌봄센터 : 0
방과후아카데미 : 0

다중선형회귀모델 예측 돌봄시설

약 174 곳

실제 돌봄시설

102 곳

실제 돌봄시설과 약 72개 차이

※ 실제사례

- 분석결과 인천 미추홀구는 돌봄시설이 가장 부족한 지역
- 특히 최근 일명 "라면화재사건"*이 인천 미추홀구에서 발생
- 위 사건을 계기로 인천광역시는
  돌봄 서비스에 대한 대대적인 전수조사를 진행 중

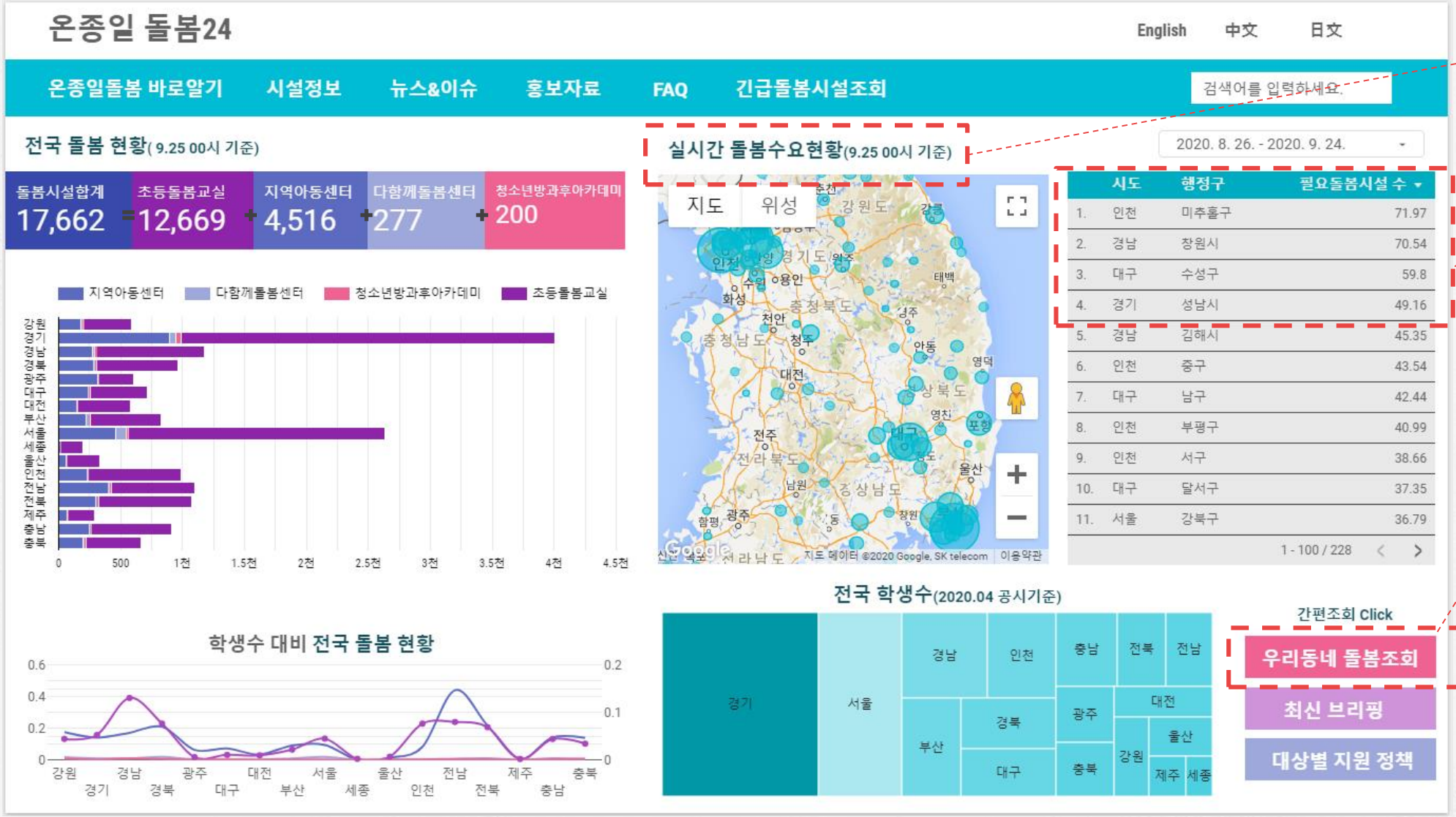
* 평소 학교에서 급식을 통해 끼니를 해결했던 기초수급대상 자녀인 두 형제가
신종 코로나바이러스 감염증(코로나19) 재확산으로
비대면 수업으로 전환하면서 급식을 먹을 수 없게 되자,
스스로 라면을 끓여 식사를 준비하려다
온몸에 화상을 입고 현재까지 의식을 잃은 상태

※ 결론

- 현재 시점 학교돌봄과 마을돌봄 공시정보를 통합적으로 확인 할 수
  있는 서비스가 부재한 상황
- 따라서

1) 지역별/행정구별 마을 돌봄과 학교돌봄의 통합하여 제시하고,
2) 각 지역에 필요한 돌봄수요를 시각화 한 서비스를 구현하고자 함

Ⅲ. 모델링을 활용하여 비즈니스 아이디어 웹페이지로 구현(서비스명 : 온종일 돌봄24)



- 실시간 돌봄시설이 부족한 지역 을 시각화
- 기간별 돌봄시설 확충이 시급한 지역을 표시
- 이용자의 거주지역 에 대한 현황 조회

온종일 돌봄24 made by 워킹맘

Ⅲ. 모델링을 활용한 비즈니스 아이디어 웹페이지로 구현(서비스명 : 온종일 돌봄24)

Ex) 우리동네 돌봄조회 -> '대구시 중구' 검색

대구시 중구의 돌봄 현황

돌봄시설합계	초등돌봄교실	지역아동센터	다함께돌봄센터	청소년방과후아카데미
18	12	5	0	1

대구시 전체 돌봄 현황

돌봄시설합계	초등돌봄교실	지역아동센터	다함께돌봄센터	청소년방과후아카데미
717	464	238	1	14

대구시 돌봄시설 우선공급지역

	행정구
1.	수성구
2.	남구
3.	달서구
4.	동구
5.	북구
6.	중구
7.	달성군
8.	서구

1 - 8 / 8

'온종일돌봄24' 서비스 기대효과

1. 정부 및 지자체 뿐만 아니라 주민도 시도별 행정구별 돌봄 유형(학교돌봄/마을돌봄)에 따른 현황파악이 용이
2. 그에 따라 수요가 시급한 지역을 파악하며 서비스 보급 우선순위 선정용이
3. 예비 초등학생 자녀를 둔 학부모에게는 '돌봄시설 우선공급지역' 정보제공을 통해 차후 주거지 결정시 도움이 되도록 함
4. 주거지 변경을 하는 가정의 경우 온종일돌봄24 서비스를 통해 자녀에 맞는 돌봄시설 수용여부 미리 확인이 가능 돌봄 공백을 최소화 하도록 함
5. 후에 돌봄 유형별 수요 파악 기대

감사합니다