

# 목 차

|                             |               |
|-----------------------------|---------------|
| <b>I. 개요 .....</b>          | <b>1</b>      |
| 1. 아이디어 제안 배경 .....         | 2             |
| 2. 현황 .....                 | 3             |
| 3. 세부 내용 .....              | 4             |
| <br><b>II. 데이터 관리 .....</b> | <br><b>4</b>  |
| 1. 데이터 수집 및 정제 .....        | 4             |
| 2. 데이터 분석 및 시각화 .....       | 5             |
| <br><b>III. 기대효과 .....</b>  | <br><b>9</b>  |
| 1. 아이디어의 실현에 따른 기대 효과 ..... | 9             |
| 2. 한계점 및 제언 .....           | 9             |
| <br><b>IV. 참고자료 .....</b>   | <br><b>10</b> |

## I. 개요

### 1. 아이디어 제안 배경

대한민국 사회는 2009년부터 지금까지 꾸준히 출산율이 감소하고 있어, 지금 2019년에는 출산율은 0.81명, 출생아 수는 26만명 수준으로 예상하고 있다. 출산율이 낮아지는 이유는 복합적인 요소가 있지만, 결혼 후 맞벌이 부부가 증가함에 따라 보육시설의 영향이 막대하다. 실제로 2016년 통계청 자료에 의하면 출산율 1위인 세종시가 출산율이 높은 이유는 보육시설 때문이라는 통계조사가 있다. 유치원 10곳 가운데 9곳 이상이 국공립이며 유치원 생 100명 가운데 95명 정도가 국공립 유치원을 다닌다. (자료: 통계청, 한국교육개발원)



〈자료: 통계청, 한국교육개발원〉

정부는 이러한 방안으로 유아를 돌보는 탁아시설과 유치원 수를 증원하는데 많은 자금을 쏟아붓고 있지만 유치원 아동학대 사건, 사립 유치원 비리, 급식 비리 등 좋지 않는 사건이 터짐에 따라 많은 예비부모들의 출산 후 보육에 대한 인식이 부정적으로 변하고 있다. 좋지 않은 사건을 긍정적 인식으로 전환하기 위해 국가에서 많은 예산을 투자하고 설문조사를 통해 실태를 파악하고 있긴 하지만, 표본이 낮은 설문조사는 보편적인 실태파악이 힘들고 특정지역에만 국한된 분석이기 때문에 한계점이 있다. 이러한 문제점을 해결하고자 공공데이터 포털에서 제공하는 '전국어린이집표준데이터'와 통계청의 '만3~7세 이하의 인구데이터'를 통해 전국 유치원 시설의 현황을 파악하고 보완점을 설계하여 기존에 있던 유치원에 대한 부정적 인식을 개선하여 예비 학부모들이 자신의 아이를 안심하고 맡길 수 있는 인식을 고착시켜 출산율 증가를 격려할 수 있다고 사려된다.

### 2. 현황

공공데이터에 앞서서 소셜 데이터를 먼저 살펴보았다. 대한민국에서 90%이상이 사용하고있



치를 확인한다. 또한 ‘어린이집유형구분’이 전국적으로 어떻게 되어있는지 파악한다.

위에서 가져온 칼럼을 이용하여 새로운 파생변수를 만들어 어린이집 지표를 만든다. 1)취원율, 2)부모안심지수, 3)유아 환경지수, 4)놀이터 비율, 5)스쿨버스 운영비율을 정의하고 시각화하여 지역마다 특징을 분석하고자한다.

## II. 데이터 분석 방법 및 시각화

### 1.데이터 수집 및 정제

RAW 데이터로는 공공데이터와 통계청의 데이터를 이용하였다. 공공데이터 포털에서 가져온 ‘전국어린이집표준데이터’에서 ‘시도명’, ‘시군구명’, ‘어린이집유형구분’, ‘정원수’, ‘보육교직원 수’, ‘보육실수’, ‘놀이터수’, ‘CCTV설치수’, ‘통학차량운영여부’ 데이터를 수집하였다.(표1)

또한 법적으로 유치원에 취원 할 수 있는 나이가 만3~7세이기 때문에 통계청에서 만3~7세 데이터를 가져와 정제하였다.(표2)

|   | 시도명  | 시군구명 | 어린이집유형구분 | 정원수 | 보육교직원수 | 보육실수 | 놀이터수 | CCTV설치수 | 통학차량운영여부 |
|---|------|------|----------|-----|--------|------|------|---------|----------|
| 0 | 경상남도 | 남해군  | 법인·단체    | 130 | 22.0   | 9.0  | 1.0  | 12.0    | Y        |
| 1 | 경상남도 | 남해군  | 민간       | 39  | 9.0    | 4.0  | 0.0  | 6.0     | Y        |
| 2 | 경상남도 | 남해군  | 민간       | 27  | 2.0    | 3.0  | 0.0  | 4.0     | Y        |
| 3 | 경상남도 | 남해군  | 국공립      | 30  | 10.0   | 4.0  | 0.0  | 5.0     | Y        |
| 4 | 경상남도 | 남해군  | 가정       | 20  | 7.0    | 3.0  | 0.0  | 4.0     | Y        |
| 5 | 경상남도 | 남해군  | 법인·단체    | 39  | 5.0    | 4.0  | 0.0  | 6.0     | Y        |

(표1)

| 행정구역(시군구) | 행정구역(시군구) | 만3~7세 인구수 |
|-----------|-----------|-----------|
| 서울특별시     | 소계        | 541,492   |
| 서울특별시     | 종로구       | 6,599     |
| 서울특별시     | 중구        | 6,215     |
| 서울특별시     | 용산구       | 11,665    |
| 서울특별시     | 성동구       | 18,943    |

(표2)

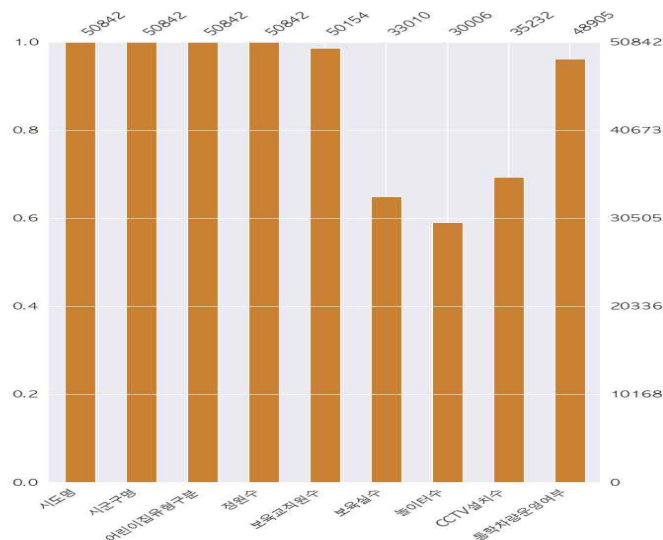
- 1) 교육학용어사전에 의하면 (해당연령층의 취원자 수 / 유치원 적령 아동수) \* 100를 취원율 이라 부른다. 또한 유치원은 법적으로 만3~7세 유아가 가도록 명시되어있다. 그래서 모든 만 3~7세가 어린이집을 간다고 가정하고 시군구별 (만 3~7세 인구수 / 정원수)을 ‘취원율’이라 정의하였다.
- 2) 어린이집 내 CCTV수가 많고, 보육실 수와 정원 수가 적으면 CCTV내 유아들이 화면 내에 기록 될 확률이 높다고 생각하여 CCTV설치수 / 보육실수 \* 정원수 = ‘부모안심지수’ 라는 지수를 만들었다.
- 3) 어린이집에 있는 유아들이 보육직원수와 보육실수가 많으면 유아들이 어린이집에서 쾌적한 환경으로 보호받을 수 있다는 ‘유아체육의 실태에 관한 조사연구: 경기도 소재 유치원 중심으로’라는 논문을 참조하여, 육교직원수 \* (보육실수 / 정원수) = ‘유아 환경지수’ 라는 지수를 만들었다.
- 4) 유치원이 놀이터가 존재하면 1, 아니면 0 이라 놓고 시군구별로 놀이터 수를 더하여 유치원 정원수로 나눈 값이다.
- 5) 유치원이 통학버스를 운영하고 있으면 1, 아니면 0이라 놓고 시군구별로 더해서 유치원 정원수로 나눈 값이다.

## 2. 데이터 분석 및 시각화



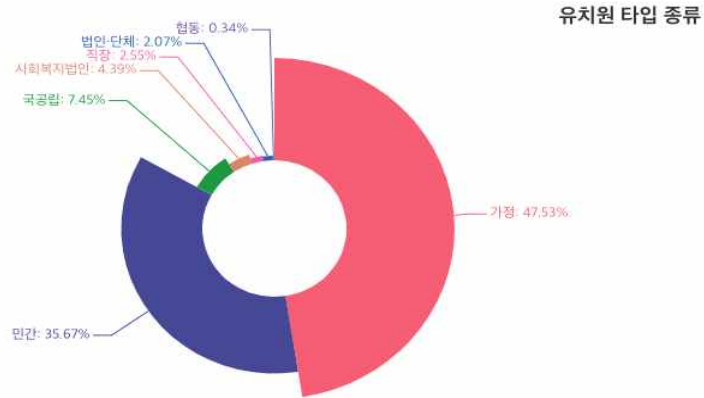
아이디어에 제시한 취원율, 부모님안심지수, 유아 환경지수, 놀이터 비율, 스쿨버스 운영비율을 시군구 별로 표시하기위해 파이썬의 pandas와 numpy를 이용하여 정제하고, 시각화는 파이썬의 matplotlib, seaborn, pyechart를 R-studio의 ggplot2를 사용하였다.

50842개의 데이터에서 ‘보육교직원수’는 결측치가 미미하여 1인 가정 유치원집이라 가정하고 1명으로 대체하였다. ‘보육실수’ 결측치는 전국 평균 보육실수로 대체하였다. 마지막으로 ‘놀이터수’, ‘CCTV설치수’, ‘통학차량운영여부’는 운영하지 않는 유치원이라고 판단하였다.(그래프 1)



(그래프 1)

결측치가 없는 어린이집유형구분을 보면 가정 유치원이 47.53%로 가장 많았고 그 다음은 35.67%로 민간이 차지했다. 서론에서 언급한 학부모들이 가장 선호하는 국공립 유치원의 비율은 7.45%로 매우 낮아 국가에서 더 증원해야 된다고 생각한다.(그래프 2)



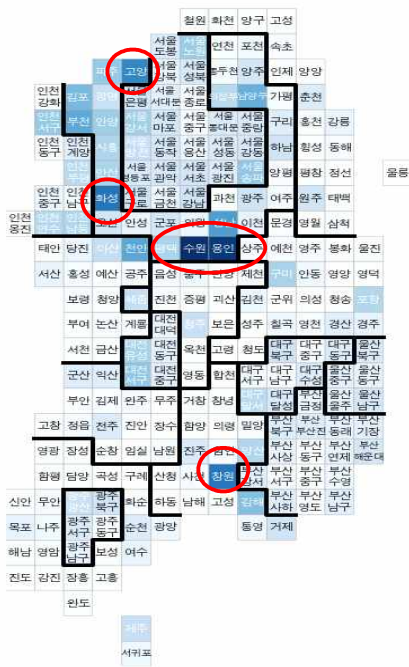
(그래프 2)

### 1) 취원율

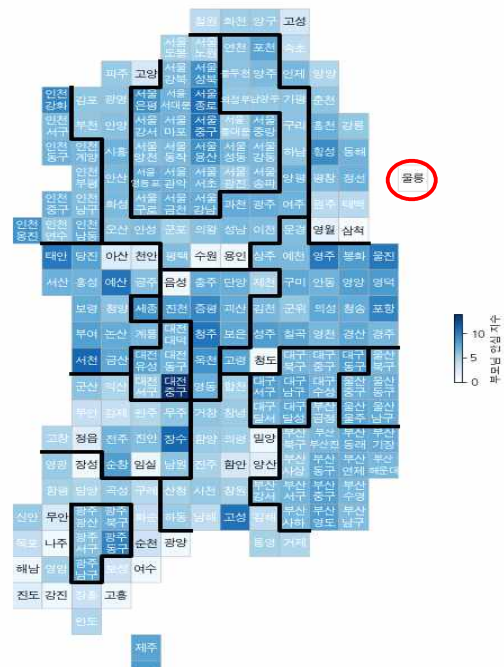
취원율은 교학용어사전에 (해당연령층의 취원자 수 / 유치원 적령 아동수) \* 100 에 이라고 명시된 것을 착안해 통계청에서 제공하는 만3~7세미만 인구수가 모두 어린이집에 간다고 가정하고, ‘전국어린이집표준데이터’의 시군구별 평균 정원수를 만3~7세미만 인구수 나눈 값을 취원율 이라고 정하였다. 위에서 구한 취원율을 전국 지도로 시각화 해보면 전체적으로 옅은 파란색으로 전국적으로 취원율이 높지 않다는 것을 확인 할 수 있다. 다른 도시보다는 서울, 경기 지역이 취원율이 높고 충청도나 전라도쪽은 취원율이 낮은 것을 확인 할 수 있다. 취원율 상위 3도시는 수원, 용인, 화성, 창원, 고양으로 파악하였다.[그림 1]

### 2) 부모안심지수

부모안심지수는 최근 유치원 내에 보육지도사들의 폭행으로 아이를 믿고 맡기는 학모들의 불안감을 해소시키자는 의미에서 정의했다. 감시한다는 개념에서 부모안심지수는 CCTV 개수는 비례하고 보육실 수의 개수와 정원 수는 반비례한다고 생각하여 (CCTV설치수 / 보육실수 \* 정원수) = ‘부모안심지수’라고 정의 하였다. 위에서 구한 부모안심지수를 시각화 해보면 전반적으로 짙은 파란색으로 사회적이슈 이후로 많이 개선되었음을 확인 할 수 있다.(‘전국어린이집표준데이터’는 2018년 이후로 업데이트 되었다.) 하지만 울릉도 지역은 다소 옅은 색깔로 부모안심지수가 다소 낮은걸 파악되어진다.[그림 2]



[그림 1]



[그림 2]

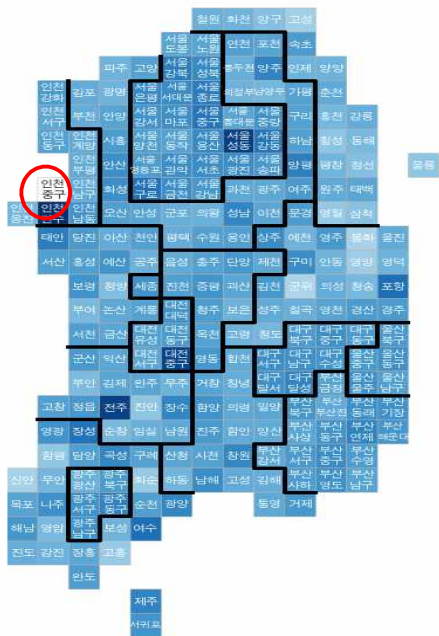
### 3) 유아환경지수

유아환경지수는 만3~7세 유아가 유치원에서 생활할 때 얼마나 편하게 생활 할 수 있는지 나타내는 지표이다. 상대적으로 보육교사들이 많고 유치원 내 방이 여러 개 존재하면 유아들이 유치원내에서 학습을 받거나 놀이 활동을 할 때, 보다 원할 한 생활을 할 수 있다. 또한 ‘한국 유아교육학회’에 의하면 유치원내에 보육교사의 놀이에 대한 교사효능감은 유아교육과정의 주요 요인인 놀이에 대한 교육적 신념과 현장실제를 위한 실천적 지식을 내포하는 교양행동의 기초가 된다고 서술 되어 있다. 또한 유치원내 보육교사의 효과는 유아의 놀이 시 교사와 유아의 상호작용의 질과 유아의 놀이 발달 수준에 영향을 미친다 한다. 위에서 구한 유아환경지수를 시각화 해보면 인천중구를 제외하고는 대체로 높은 수치를 나타낸다.[그림 3]

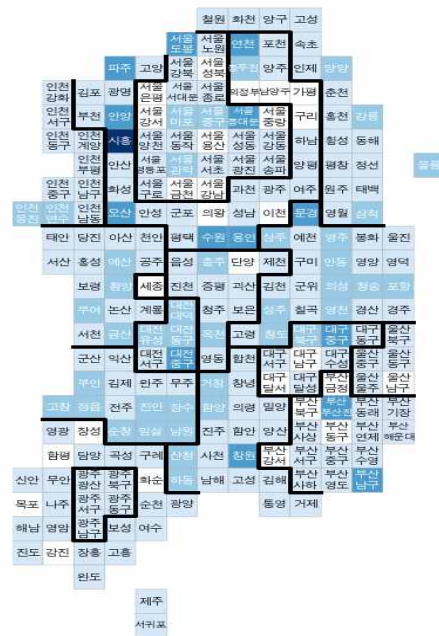
### 4) 놀이터 비율

놀이터 데이터는 50,842데이터에서 20,836개가 결측치였다. 결측치는 값은 놀이터가 존재하지 않는 유치원이라 생각하여 0값이라 넣고 시군구별 정원수 개수를 파악한 다음 놀이터 개수의 평균을 계산하였다. 놀이터비율은 시군구 별로 많은 최대 많은 곳은 0.4 밖에 안되는 수치로 유아 행동발달에 중요한 역할을 하는 놀이터 추가 설치가 필요한 상황으로 보인다.[그림 4]





[그림 3]



[그림 4]

##### 5) 스쿨버스운영 비율

스쿨버스운영 비율은 유치원이 통학버스를 운영하고 있으면 1, 아니면 0이라 놓고 시군구별로 더해서 유치원 정원수로 나눈값 이다. 스쿨버스 운영의 유무는 스쿨버스를 운영하지않는 유치원이라면 맞벌이 부부가 직접 아이를 등원시켜야 되기 때문에 맞벌이 부부가 증가하는 요즘과 같은 시대에는 중요한 요소라고 생각된다. 지도를 보면 지방권을 거의 1에 가까운 비율로 운영 하는것에 비해 서울, 경기, 경상북도는 0.5이하인 시군구가 많다.[그림 5]



[그림 5]



### Ⅲ. 기대효과

#### 1. 아이디어 실현에 따른 기대효과

공공데이터 기반의 분석을 통해서 기존에 표본이 작은 설문조사에 비해 신뢰성이 높기 때문에, 시군구 별 유치원 특성을 정확성 있게 파악 할 수 있다. 그 결과로 위에서 조사한 시군구 별로 유치원정원 수 증원, 보육교사 수 재배치, 스쿨버스 운영비율 증가 등 유치원의 환경을 개선 할 수 있어 전국적으로 예산을 투여하는 것보다 실용적일 것이라고 고려된다. 이러한 개선을 통해 유치원 비리, 유치원 폭행 등 사회적으로 유치원에 대한 예비학부모들의 나쁜 인식 전환시켜 출산율 증가에 이바지 할 수 있다고 생각한다. 또한 현 정부의 2022년까지 취원율 40%달성의 목표에 도움을 줄 지표가 될 수 있다고 생각한다.

#### 2. 한계점 및 제언

분석을 진행하면서 느꼈던 것은 주어진 데이터의 결측치 처리에 신중을 기하여야 한다는 것이었다. ‘전국어린이집표준데이터’에는 ‘과천시’, ‘용인시’, ‘화성시’, ‘부산 기장군’, ‘ 부산 수영구’, ‘서울 동대문구’, ‘서울 마포구’의 데이터는 모든 항목이 존재 하지 않았다. 이 지역의 결측치를 0이라고 대체하여 분석하였을 때 데이터 수집 프로세스에서 추출 오차 (Selection bias)가 발생하였다. 그리하여 위의 시군구는 결측치를 평균값으로 제시하였기 때문에 위에서 분석한 결과값이 나왔다. 또한 취원율에서 만3~7세 유아가 모두 유치원을 취원한다고 가정하였기 때문에 오차가 생길 수 있고, ‘전국어린이집표준데이터’가 2019년의 최신의 데이터가 아니기 때문에 포털 사이트의 유치원 수와 다르다는 점에서 오차가 생길 수 있다.

이번 연구를 통해 출산율에 막대한 영향을 가지고 있는 유치원 환경에 대해서 살펴보고 환경을 개선시켜 출산율을 장려하는 계기가 되길 바란다.

### Ⅳ. 참고자료

서연태

‘유아체육의 실태에 관한 조사연구: 경기도 소재 유치원 중심으로’  
(한국체육과학회지)

Hyeshik on

“버거 지수는”는 진짜 도시의 발전 수준을 반영할까?  
GitHub

한국 유아교육협회 학술지

OECD통계정책위원회 William Kelly

“2018년 한국의 통계청이 발표한 결혼 관련 통계”