

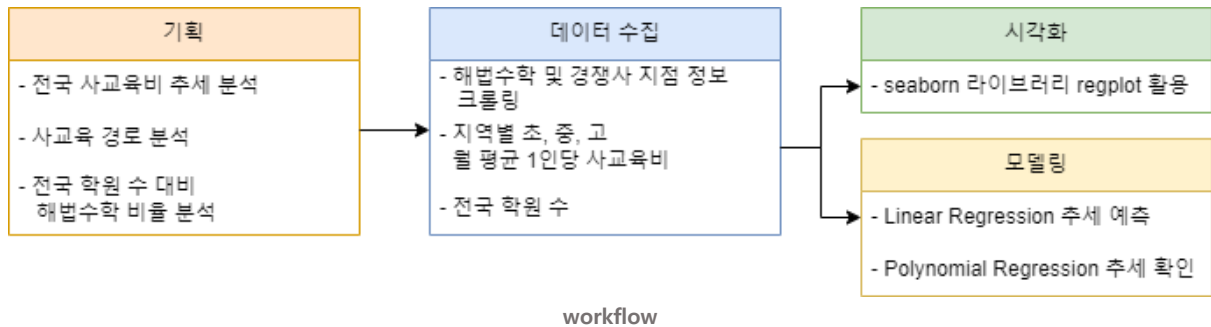
해법수학 입지 정보 분석

프로젝트 기반 빅데이터 서비스 개발자 양성 과정 (4 기)
이형석

서론

해법수학의 입지 정보를 분석하여 비즈니스적 관점에서 활용할 수 있는 인사이트를 도출하는 것에 이번 프로젝트의 목적이 있다. 대한민국 사교육 시장을 지역별로 나누어 분석했고 초, 중, 고 사교육비 정보와 과목별, 전체 학원 수 추세에 대해서도 분석했다. 추가로 모델링을 통해 인구 추세 예측도 진행했다.

방법



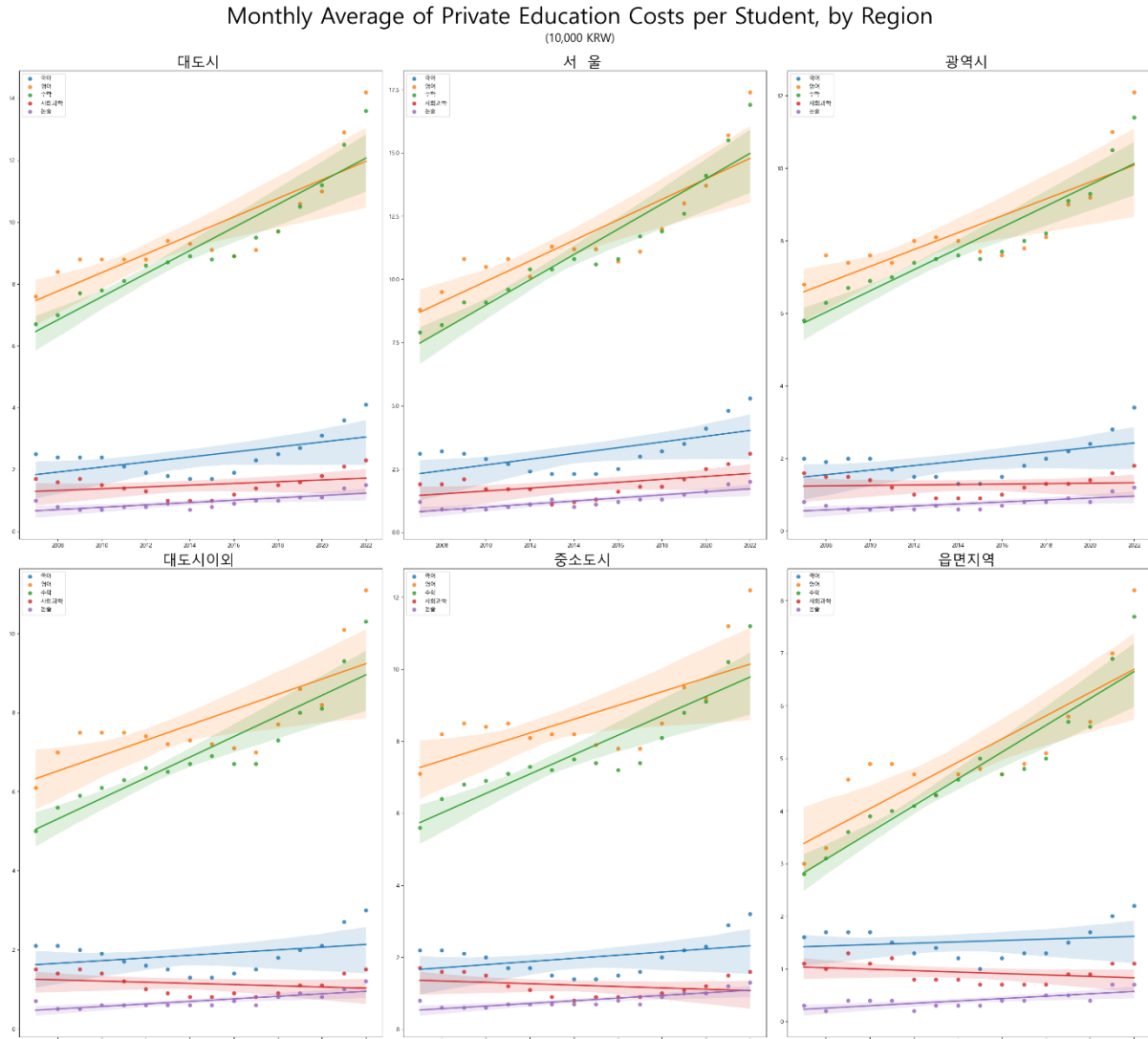
국가통계포털 데이터를 재가공하는 공공데이터포털 웹사이트를 통해 지점 정보 및 전국 사교육비, 과목별 사교육비, 학원 수에 대한 파일 데이터를 수집했다. 추가적으로 학원 위치 별 데이터를 비교하기 위해 경쟁사 중 웅진과 푸르넷 위치정보도 크롤링했다.

추세 시각화 부분은 seaborn 라이브러리 내 regplot 을 사용했고, subplot 으로 정리했다.

모델링은 데이터의 수를 고려해 인구 예측을 시도했고 Linear Regression 과 다항을 고려할 수 있는 Polynomial Regression 을 사용했다.

결과

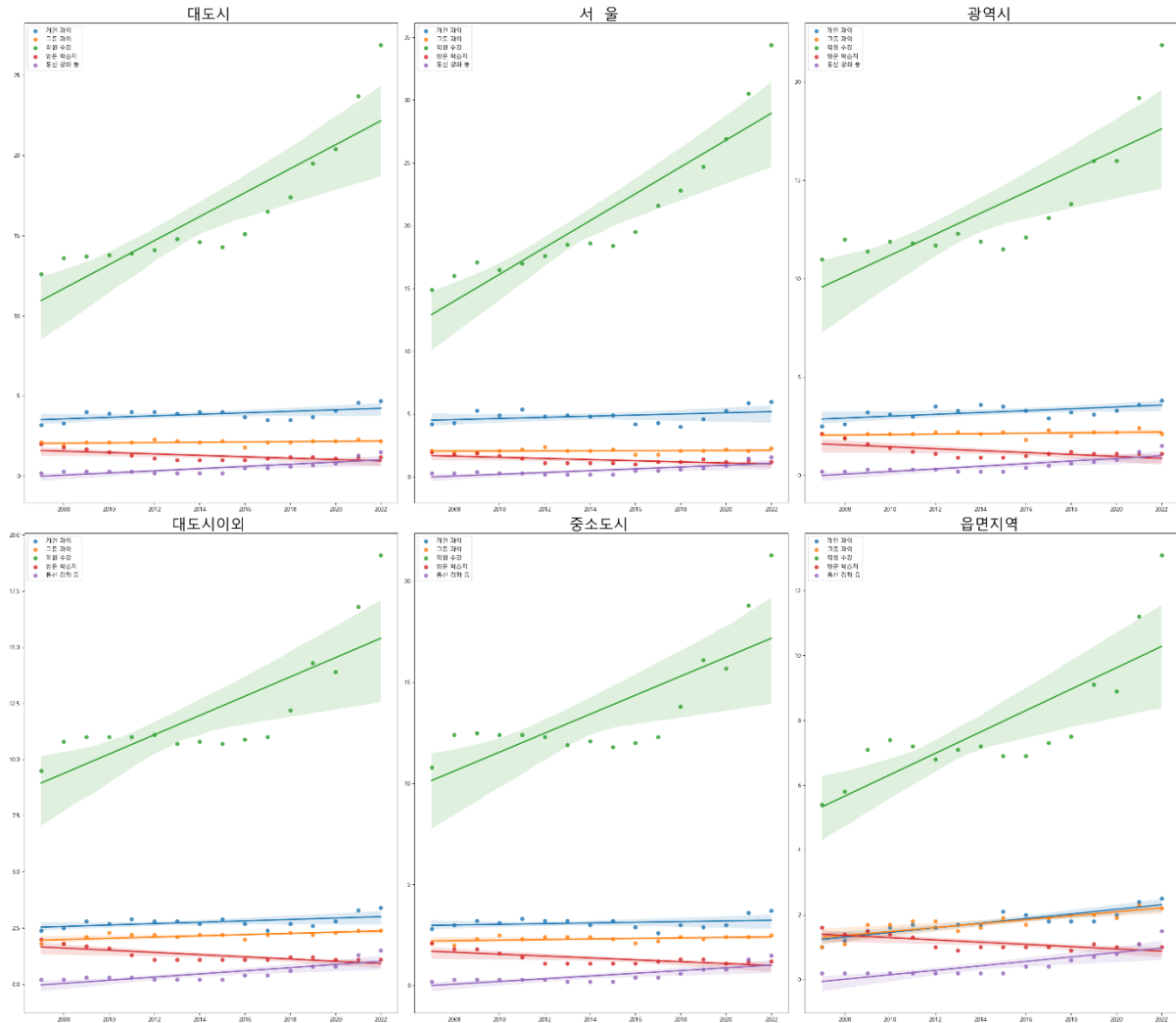
추세 시각화



[그림 1] 지역별 1인당 월 평균 사교육비 추세

대도시와 서울의 월평균 사교육비 추세는 우상향하고 있으며, 특히 영어 수학 과목의 상승세가 가장 가파르다. 영어 수학의 경우 대도시 뿐 아니라 중소도시, 읍면지역에서도 가파른 상승세를 보여준다.

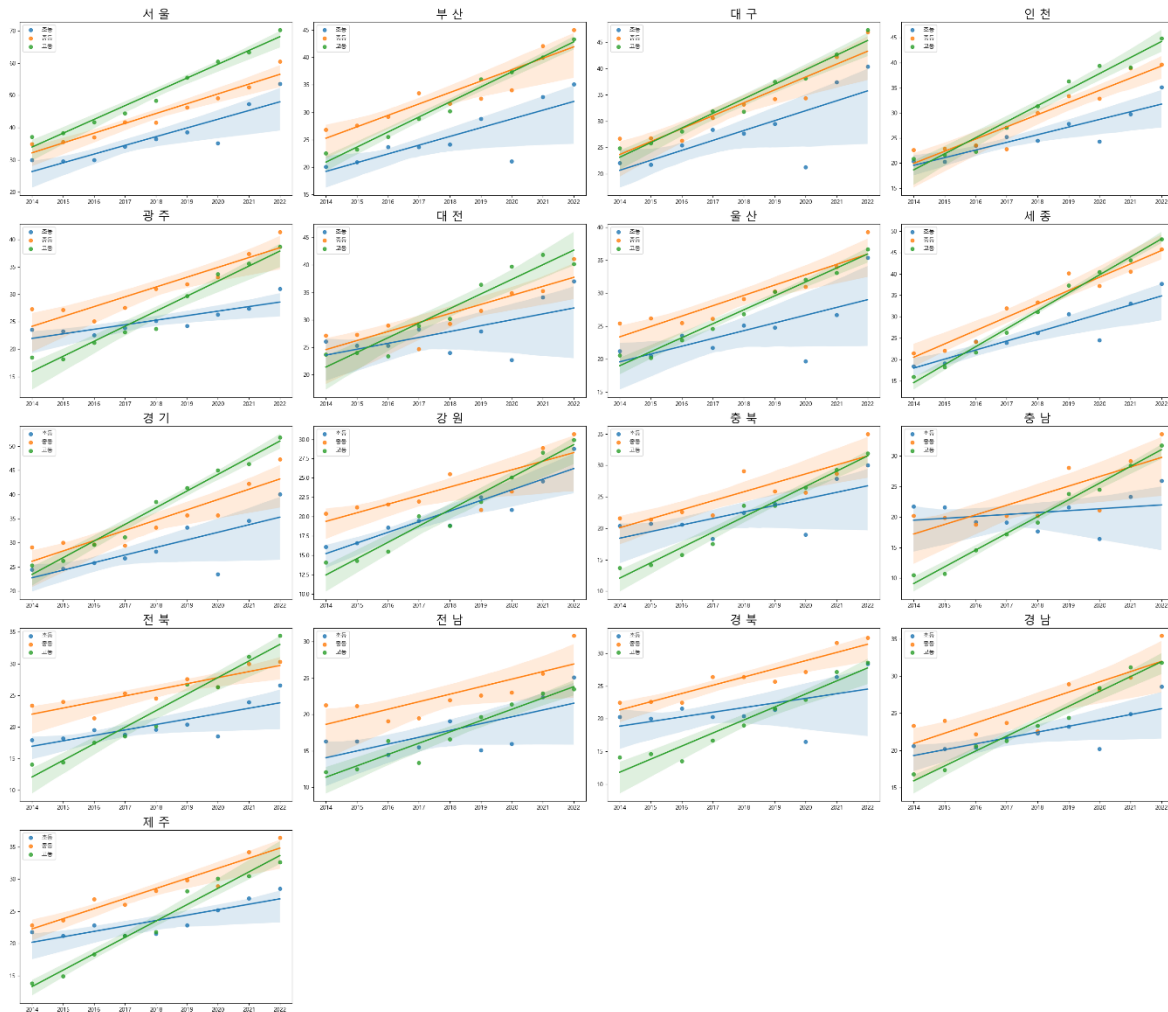
Types of Private Education, by Region (10,000 KRW)



[그림 2] 사교육 경로 추세 (2007-2022)

해법수학이 오프라인 매장임을 고려하여 어떤 경로로 사교육을 주로 소비하는지 분석했다. 학원 수강이 압도적으로 높은 지출을 보여줬고, 방문 학습지의 경우 모든 지역에서 하락하는 추세를 보인다. 통신 학습이 학원 수강과 경쟁구도를 보이지 않을까 예상했지만, 의외로 낮은 소비율과 추세를 보였다.

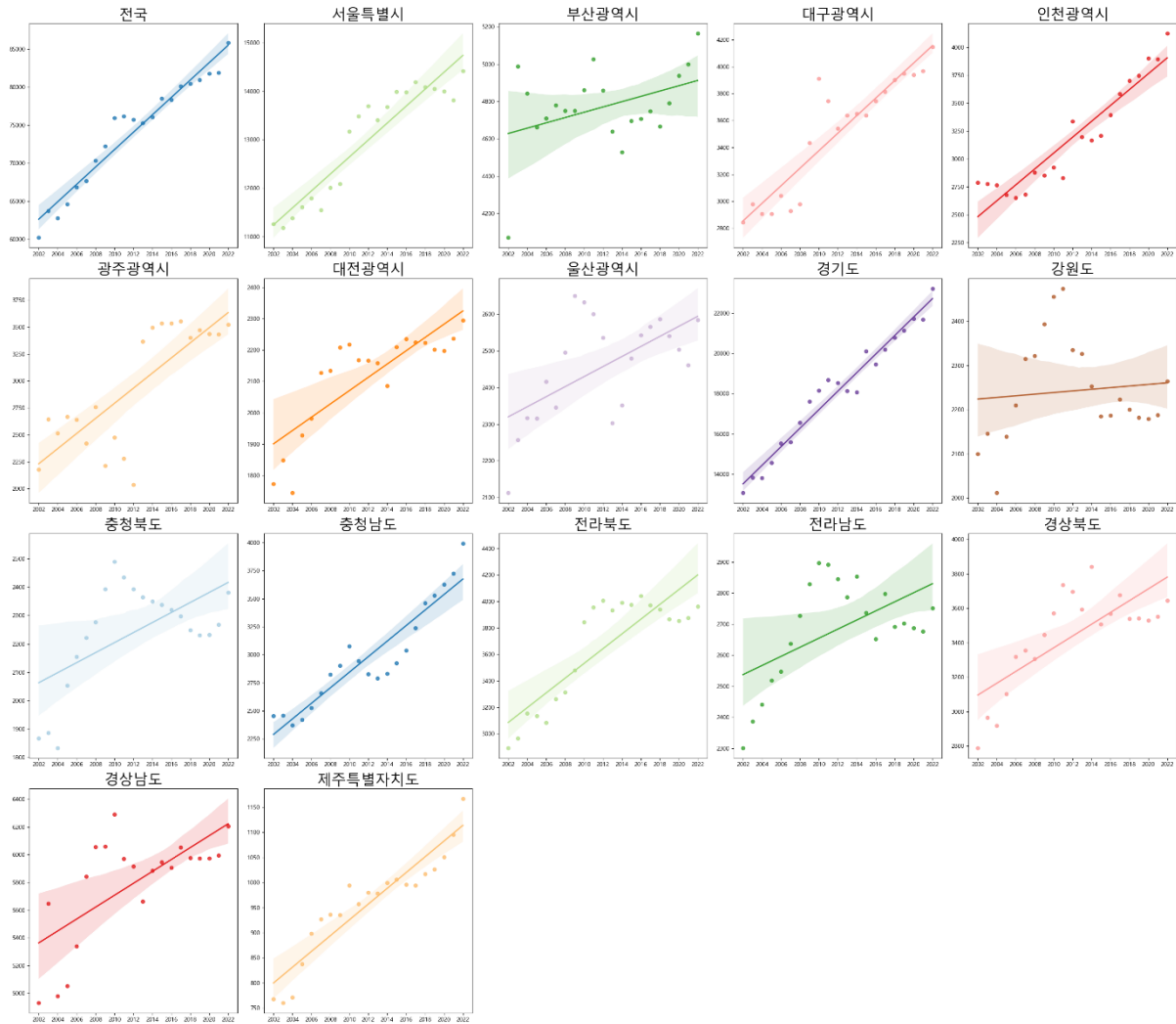
Trend of Private Education Costs, by School Class (10,000 KRW)



[그림 3] 초, 중, 고 지역별 사교육비 추세(2014-2022)

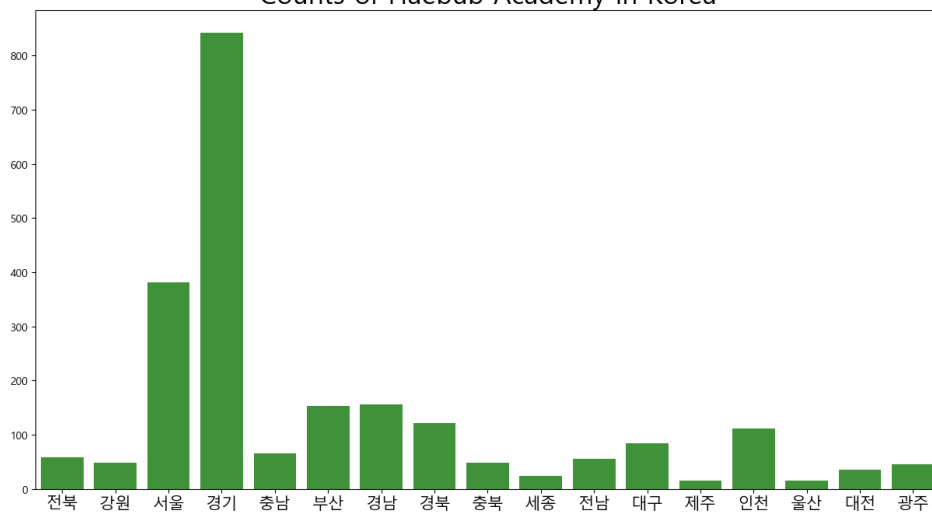
대학 진학을 위한 고등학교에서 가장 높은 사교육비가 나가지 않을까 예상했지만, 의외로 중학교 시기 사교육비 지출이 높은 경우가 많았다. 한 가지 특이한 점은 2014년 대비 고등학교 사교육비 추세가 모든 지역에서 가장 가파른 상승세를 보이고 있다는 것이다.

Number & Trend of Academy



[그림 4] 학원 수 및 추세 (2002-2022)

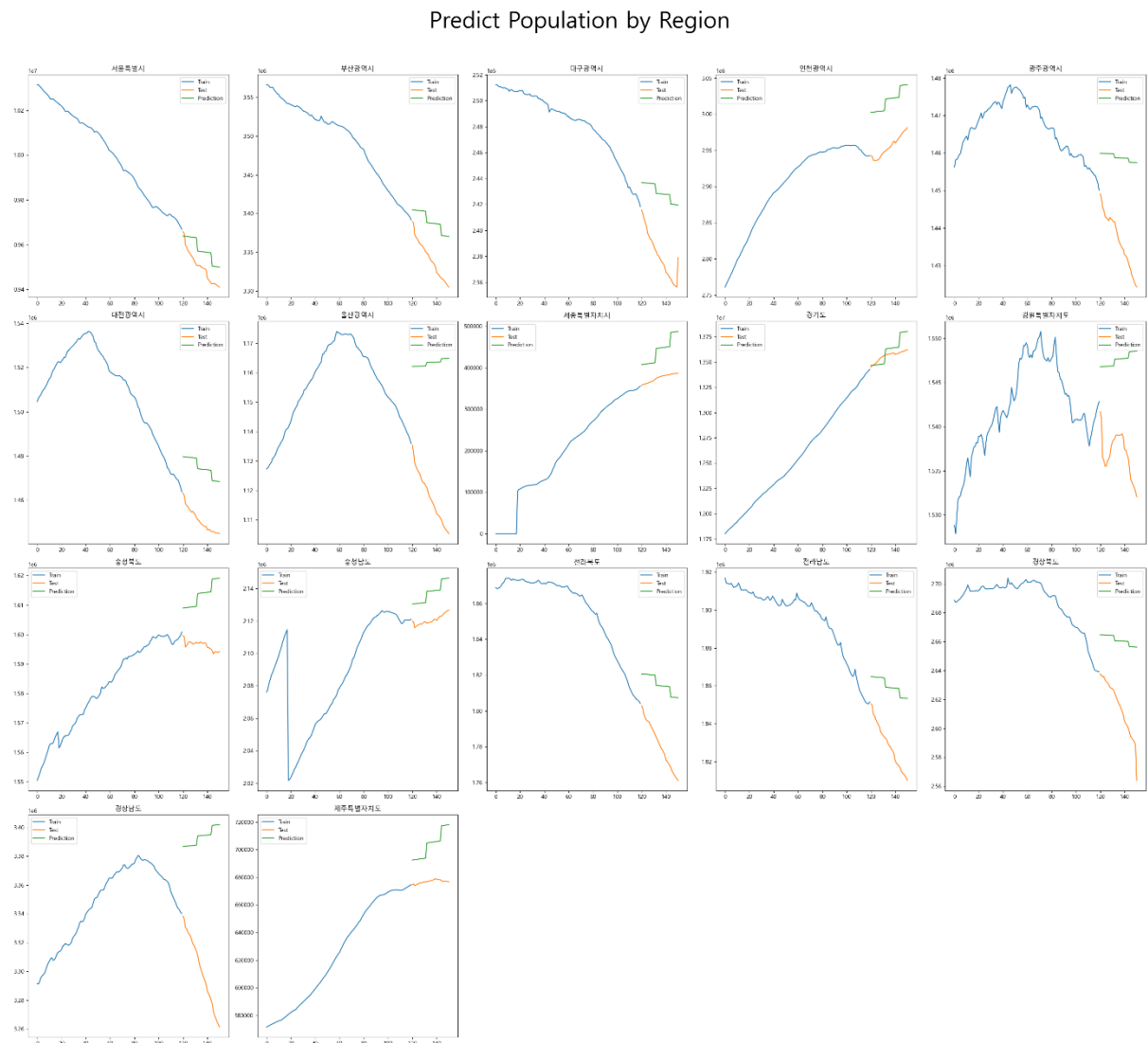
Counts of Haebub Academy in Korea



[그림 5] 지역별 해법수학 개수

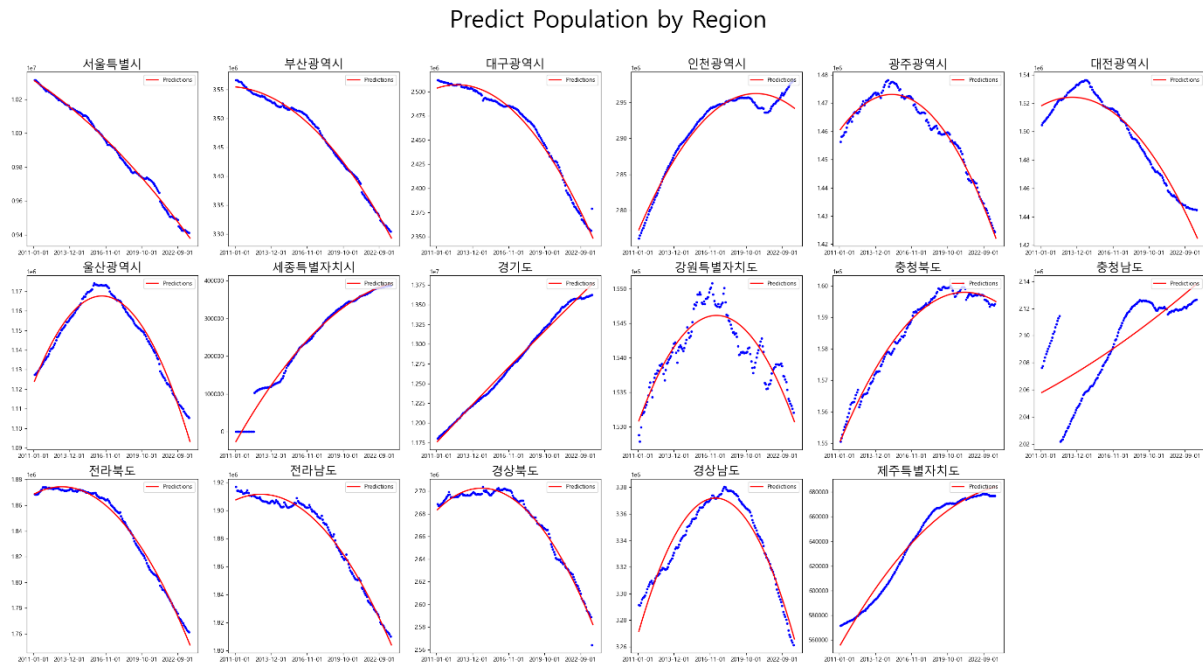
[그림 4]에서 정의한 학원은 일반 교과 뿐 아니라 다른 분야의 학원도 포함돼 있으며, 한국교육개발원에서 조사한 데이터를 기준으로 하고 있다. 대한민국의 높은 학구열과 일반 교과 학원 중 업계 상위권을 차지하고 있는 해법수학이 전체 학원 중에서도 높은 비율을 차지하지 않을까 예상했지만, 전체 학원수에 비하면 아주 낮은 비율의 수준이라 그래프에 표시하지 못했다. 편차가 큰 지역도 있지만 대부분의 지역에서 우상향하는 추세를 보인다.

모델링



[그림 6] 지역별 인구 예측 결과, Linear Regression

학원은 인구와 밀접한 관련이 있다고 분석되어 기본적인 회귀 모델인 Linear Regression을 활용해 지역별 인구 예측 모델링을 진행했다. 선형 회귀 모델이다보니 상승, 하락하는 추세에 대해 그대로 예측하는 결과를 나타냈다.



[그림 7] 지역별 인구 예측 결과, Polynomial Regression

선형회귀로 추세를 확인했을 때, 한 추세밖에 예측을 못했기 때문에 다항회귀를 활용한 추세도 확인했다. 세종, 경기, 충남 제주에서 우상향하는 추세를 확인했다.

결론 및 제언

전국의 학원 수 뿐만 아니라 사교육을 접하는 경로 중 학원의 비율이 지속적으로 증가하는 추세이기에 학원을 운영하는 시스템은 계속 유지하는 것이 좋겠다고 판단된다. 다만, 추가적으로 목표 수강생 및 지역별 경쟁 환경 등 추가적으로 고려해볼 필요가 있다.

코드

Git : https://github.com/lhshs/Genia/tree/main/mini_pjt/230809-11_hp_cometitor