한빛 -> 데이터 분석 부분: 연도별, 월별 어린이 보호구역 내 교통사고 현황

Ⅱ. 데이터 분석

3. 어린이 보호구역 내 어린이 교통사고 발생 현황

어린이 보호구역 내 어린이 교통사고 발생 현황에 대한 데이터 분석 부분입니다.

텍스트이(가) 표시된 사진

자동 생성된 설명**- 어린이 보호구역 내 어린이(12세이하) 교통사고 (=**연도별 어린이 보호구역 내 어린이 교통사고)

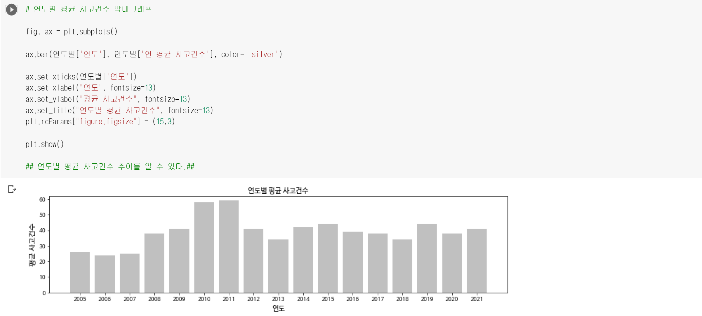
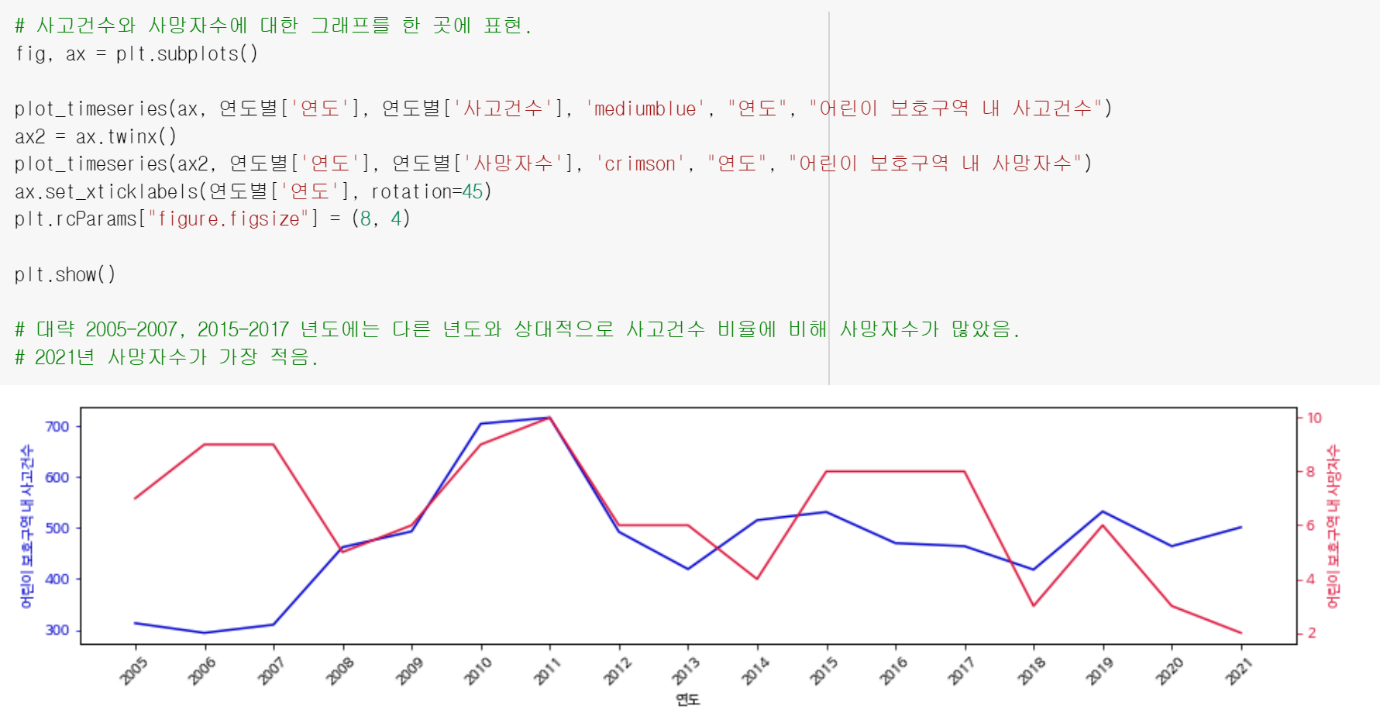
(◀여기서부터 제가 분석한 데이터 부분입니다.)

연도별로 발생한 어린이 보호구역 내 교통사고에 대한 데이터를 살펴보겠습니다. 데이터는 2005년부터 2021년인 최근까지 17개년도에 대한 데이터를 사용하였습니다. 먼저 2005년부터 2021년까지의 엑셀 파일을 불러왔습니다. 이 부분의 데이터 전처리는 연도와 관련 사항을 그래프로 시각화 하는 것이므로 기존의 데이터프레임에 있는 ‘합계’에 대한 행만을 남기고, 각 시도의 명칭에 대한 열을 삭제하여 필요한 부분만 출력되게 하였습니다. 그 다음으로 2005부터 2011년과 2012부터 2021년으로 분리되어 있던 데이터프레임을 concat을 사용하여 결합시키고, 행은 연도에 대한 것, 열은 사고건수, 사망자수, 부상자수로 만들어 주기 위하여 transpose를 사용하여 행과 열을 바꾸어 주었습니다. 이 결과 데이터프레임에서 각 연도에 대한 것이 index로 사고건수, 사망자수, 부상자수가 0번째 행의 data로 되는데, iloc를 사용하여 0번째 행의 데이터를 컬럼명으로 지정하고, 불필요한 0번째 행을 삭제하였습니다. 그 후 연도에 대한 부분을 컬럼의 데이터로 만들어 주기 위하여 reset\_index로 해당 인텍스를 컬럼으로 만들고, 컬럼명을 ‘연도’로 설정하여 데이터 프레임을 정리하였습니다.

그 후테이블이(가) 표시된 사진

자동 생성된 설명, 연도별 월 평균 사고건수를 막대 그래프로 나타내기 위하여 불필요한 사망자수와 부상자수에 대한 열을 삭제하고, 데이터프레임에 연 평균 사고건수에 대한 데이터 열을 추가하기 위해 연도별 총 사고건수의 각 데이터를 1년이 12개월이므로 12개월로 나누면 각 연도별 월에 대한 평균 사고건수에 대한 데이터 계산을 하였습니다.

결과 데이터 프레임 값입니다. (옆의 데이터프레임 이미지를 보여주면서)

‘연도별’이라는 데이터프레임을 막대그래프로 시각화한 결과 데이터프레임 값에서 볼 수 있었듯이 연도별 월 평균 사고건수는 2011년 59건으로 가장 높았고, 그 다음으로는 2010년 58건, 2015년과 2019년이 동일하게 44건, 2014년이 42건으로 상위 5개년도로 결과가 나왔습니다. 또한 가장 적게 나온 연도는 2006년 24건이며 민식이 법이 지정된 2020년에는 38건으로 2전년대비 44건보다 감소하였지만, 그 다음 연도인 2021년도에 41건의 증가하였습니다. 2021년의 사고건수 증가는 민식이 법 재정 의미가 무의미한 것을 볼 수 있었습니다. 이처럼 막대 그래프를 통하여 데이터프레임의 출력 값보다 어느 년도의 연도별 월 평균 사고건수가 높고 낮은 지 그리고 사고건수의 추이가 어떻게 되는지에 대한 것을 쉽게 눈으로 확인할 수 있습니다.

▲ 다음으로는 어린이 보호구역 내에서 연도별 사고가 일어난 수에 비하여 사망자가 얼만큼 발생하였는 지 비교해보기 위해 2005년부터 2021년까지 총 사고건수와 사망자수에 대한 그래프를 한 그래프 상에 표현해보았습니다. 서로 값을 비교하기 위하여 각각 해당하는 값을 다른 색으로 표현하였고, def를 사용하여 사고건수와 사망자수 그래프를 그리기 위한 함수를 선언하였습니다. 대략 2005년에서 2007년, 2015년에서 2017년 기간을 보시면 사고건수에 비하여 사망자수가 많음을 볼 수 있고, 2018년부터 2021년까지 사고건수에 비하여 사망자수가 적고, 특히 2021년에는 17개년도를 비교해 보았을 때 사망자 수가 가장 적다는 결과를 보실 수 있습니다. 이 점으로 보아 민식이 법 재정이후 사고건수 발생은 효과가 없었지만, 사망자수는 줄어든 것을 보실 수 있습니다.

**- 월별 어린이 보호구역 내 어린이(12세이하) 교통사고** (=월별 어린이 보호구역 내 어린이 교통사고)

다음은 12개월 ‘월’에 해당하는 어린이 보호구역 내 어린이 교통사고 값을 비교해보면서 12개월 중 어느 달에 어린이 보호구역 내에서 교통사고건수가 많이 발생하는 지 분석해보았습니다. 먼저 2005년부터 2021년까지 월별에 대한 데이터를 불러온 후에 데이터프레임을 필요한 것만 남기고 정리하였습니다. 정리 과정은 2005년, 2006년에 대한 데이터프레임을 예로 들면, 시도 중에서 합계에 대한 행만을 남겨주고, 나머지 시도, 시군구, 2005년에 해당하는 합계에 대한 열을 drop으로 삭제해주었습니다. 행과 열을 transpose로 전환시키고, 사고건수, 사망자수, 부상자수와 연도.월을 칼럼명으로 바꾸어 주는 과정은 앞서 설명한 연도별 데이터프레임 정리 데이터 전처과정과 동일합니다. ‘연도.월’ 이라는 값을 연도와 월로 각각 나누어 주기 위하여 split을 사용하여 ‘ . ‘ 기준으로 칼럼을 나누어 주었습니다. get을 사용하여 기존 데이터프레임 열에 이어서 0번째 열 연도, 1번째 열 월로 설정 후, ‘연도. 월’에 대한 컬럼을 삭제하고, 컬럼 순서를 연도, 월, 사고건수, 사망자수 순으로 지정하였습니다. 그리고 월2005는 2005년에 대한 월만 선택한 데이터프레임입니다. 나머지 년도도 앞의 내용과 동일한 방식으로 데이터 전처리가 이루어집니다.

최근 5년간의 데이터를 시각화를 해보겠습니다. 각각 5년간의 월별 사고건수를 비교해 보기 위해 ‘어린이 보호구역 내 사고건수’의 y축 범위를 ylim으로 0에서 100건으로 동일하게 설정하였습니다. 앞서 데이터 전처리한 데이터프레임 중에서 각 년도에 해당하는 월2017에서 칼럼명의 ‘월’ 데이터를 x축으로 ‘사고건수’ 데이터를 y축으로 설정하였습니다. 나머지 년도도 동일합니다. 각 5개년도에 해당하는 별도의 그래프를 보여주었고, 비교를 위하여 한 곳에 그려주어 최근 5년간 월별 어린이 보호구역 내 사고건수를 비교해보았습니다. xlabel은 월로 ylabel은 사고건수로 설정하였고, 타이틀은 최근5년간 월별 어린이 보호구역 내 사고건수로 설정하였습니다. 대체적으로 다른 달에 비하여 1,2월에는 사고건수가 적지만, 3월부터 어린이 보호구역이 있는 초등학교가 개학과 동시에 어린이들이 보호구역 주변을 많이 지나가게 되는 시기부터 사고건수가 많아지기 시작하여 대게 5월, 6월이 가장 많은 것을 볼 수 있습니다. 그리고 초등학교의 여름방학인 7~8월이 겨울방학인 1~2월에 비해 사고건수가 많음을 볼 수 있습니다. 또한 2020년 3월에 민식이 법이 시행이 되었는데 그 결과로서 3월에 사고건수가 5개년도를 합한 전체의 월 중에서 가장 적었지만, 점차 사고건수가 증가하는 양상을 보이면서 해당년도 6월에는 5개년도를 합한 전체의 월 중에서 가장 많은 사고건수를 그래프상 볼 수 있습니다. 7월은 6월에 비해 줄어들었지만, 다른 년도의 7월에 비하면 높은 건수입니다. 8월부터 12월은 다른 년도와 비슷한 양상을 보여주고 있습니다. 이 그래프의 결과로서 시사하는 바는 민식이법 시행직후는 사고건수가 현저히 줄었다고 할 수 있지만 시행된지 2개월이 채 지나지도 않아 5월 그리고 6,7월을 제외한 월은 다른 년도와 사고건수가 비슷하였고, 6월, 7월에는 오히려 다른 년도보다 더 많은 사고건수를 기록하고 있다는 점에서 민식이 법 인식의 유효기간이 2개월도 안되는 안타까운 상황과 더불어 민식이 법이 법으로서 효과가 있는지 의문이 되는 점을 보여주고 있습니다.