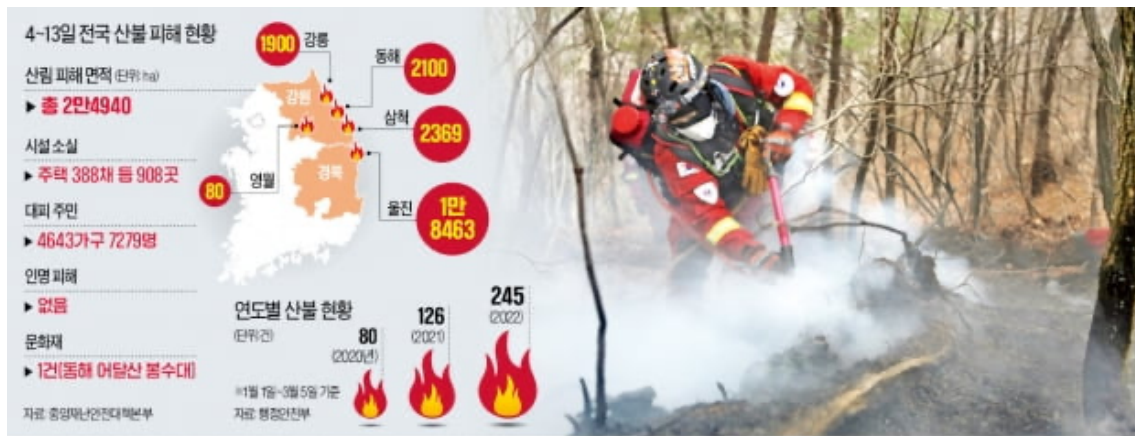


# 산불 발생 동향과 영향 요소 분석

## 1. 주제 선정 및 배경

지난 3월 6일 울산에서 산불에서 발생했다. 이번 산불로 인해 소실된 면적은 서울 면적의 41.2%에 1689억원의 피해가 발생했으며 피해 복구비용은 3400억원에 이른다. 또한 전국 송이버섯 채취량에 10%에 달하는 지역 송이버섯 산지가 초토화 되기도 했다.



행정안전부에 따르면 산불 발생 건수는 최근 10 년새 277 건에서 620 건으로 늘어났다. 또한 최근 10 년간 산불로 인한 주택 소실, 등 직접적 경제 피해 규모는 6758 억원으로 최근 심각한 사회문제로 대두되고 있다.

따라서 이러한 산불 피해를 줄이기 위해 통계청과 기상청, 통계청의 공공데이터를 활용하여 최근 15년간 산불의 발생 동향과 영향 요소를 분석하고 산불의 발생을 예측하는 모델을 만들어 보았다.

## 2. 분석 내용 및 결과

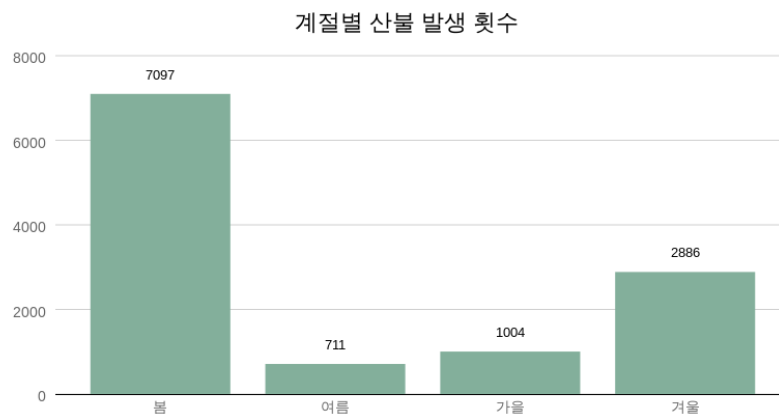
### - 데이터 전처리

	연도	건수	풍속(m/s)	습도(%)	면적(ha)	피해금액(백만원)
1	2007	418	2.0	68.6	230	377
2	2008	389	2.0	67.0	227	436
3	2009	570	2.1	65.7	1381	3883
4	2010	282	2.1	67.7	297	4451
5	2011	277	2.1	66.2	1090	29063
6	2012	197	2.1	66.1	72	2542
7	2013	296	2.1	67.7	552	25020
8	2014	492	2.0	67.8	137	9285
9	2015	623	2.0	68.6	418	30480
10	2016	391	1.9	68.9	378	15721
11	2017	692	1.9	65.5	1480	80150
12	2018	496	1.9	67.0	894	48583
13	2019	653	1.8	67.1	3255	268910
14	2020	620	1.9	69.4	2920	159141

각 연도별로 요인마다 총합을 데이터프레임으로 전처리하였고 데이터프레임에 없는 나머지 요인들은 시각화를 진행하여 분석에 활용하였다.

### - 데이터 시각화

아래와같이 그래프를 통해 시각화를 진행하였다.

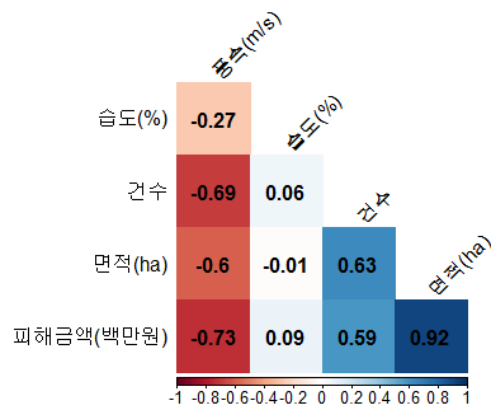


봄에 울산에서 가장 큰 화재가 발생하여 계절별로 화재발생 동향을 간단하게 확인해보았다.

우선 계절별 산불 발생횟수를 보았을때 봄에 가장 많이 발생하고 여름에 가장 적게 발생하는 것을 알 수 있었다.

계절 뿐만 아니라 다른 요소들이 화재 발생에 어떤 영향을 미치는지 확인해보기 위해 나머지 더 자세한 화재 요소 시각화 자료는 밑에서 진행할 분석에서 사용하였다.

## - 데이터 분석



데이터의 각 변수들이 산불 발생 횟수에 미치는 영향을 알아보기 위해서 각 요인들마다 상관분석을 실시하고 시각화해보았다.

기준점을 건수로 잡았을 때 풍속이 -0.69, 면적이 0.63, 피해금액 0.59로 각각 산불 발생 횟수에 영향을 미치는 것을 알 수 있었다. 즉 풍속, 면적, 피해금액 순으로 산불 발생 횟수에 영향을 미치는 요소라고 볼 수 있다.

```
Residuals:
    Min       1Q   Median       3Q      Max
-141.67  -56.04  -23.42   74.35  223.30

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    2.676e+03  9.748e+02   2.745   0.0206 *
stat$`풍속(m/s)` -1.144e+03  4.818e+02  -2.374   0.0390 *
stat$`면적(ha)`   1.475e-01  8.231e-02   1.792   0.1034
stat$`피해금액(백만원)` -1.660e-03  1.284e-03  -1.293   0.2250
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

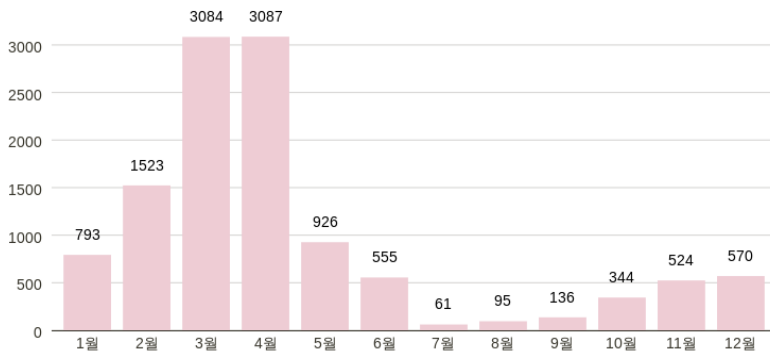
Residual standard error: 112.8 on 10 degrees of freedom
Multiple R-squared:  0.6148,    Adjusted R-squared:  0.4993
F-statistic: 5.321 on 3 and 10 DF,  p-value: 0.01889
```

상관분석으로만은 인과관계를 밝힐 수 없기 때문에 산불 발생횟수에서 어떤 요소가 영향을 미치는지에 대해 인과관계와 영향을 조금 더 자세하기 알아보기 위해 풍속, 면적, 피해금액을 독립변수로 설정하고 산불 발생 횟수를 종속변수로 설정하여 회귀분석을 실시해보았다.

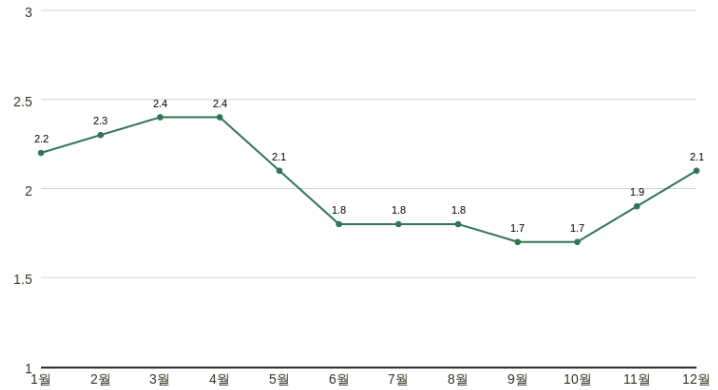
이때 p-value가 0.01889로 유의수준 0.05보다 낮으므로 대립가설이 성립하여 이 회귀모델이 유의하다고 볼 수 있다. 또한 결정계수는 0.6148로 약61%의 설명력을 갖는 모델이라고 할 수 있다. 각 독립변수의 p-value를 보았을 때 풍속이 0.039로 0.05보다 낮기 때문에 풍속이 가장 큰 인과 관계를 갖는다.

상관분석과 회귀분석을 통해 풍속이 산불 발생횟수에 가장 큰 영향을 미친다는 것을 알 수 있었다.

월별 산불 발생 횟수



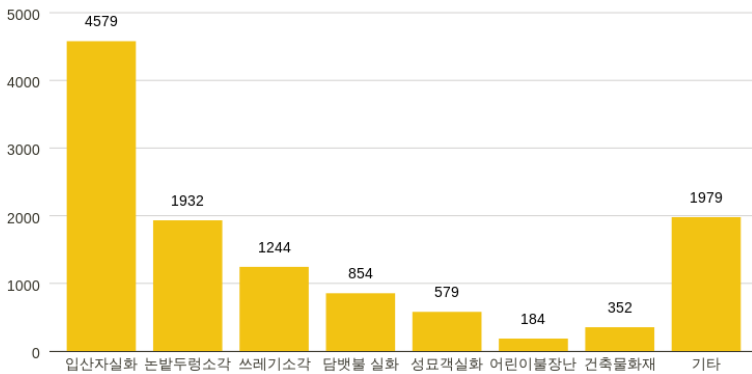
월별 평균 풍속



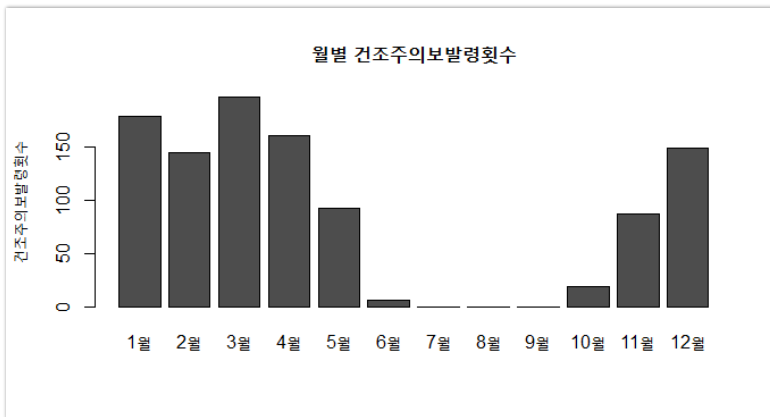
월별로 보았을 때 평균 풍속이 가장 높은 달이 3,4월에 산불 발생횟수가 가장 많은 것을 알 수 있다.

앞에서 실시한 상관분석과 회귀분석을 통하여 풍속이 낮을수록 산불이 많이 발생한다는 것을 알 수 있었지만 위의 그래프에서는 평균 풍속이 가장높은 3,4월에 산불 발생 횟수가 가장 높았다. 이는 분석 내용과 상반된 결과이기 때문에 다른 요소들을 조금 더 확인해보았다.

원인별 산불 발생 횟수



월별 건조주의보발령횟수



예를들어 건조한 날씨, 산불이 발생하는 원인과 같은 앞의 분석에서 독립변수로 사용한 요인들 외에 다른 요인들이 산불 발생의 원인에 영향을 미친다는 것을 알 수 있었다.



```

Residuals:
    1      2      3      4      5
-0.04007  0.05423  0.15869 -0.11645 -0.05638

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.584e+02  3.089e+00  180.751 3.06e-05 ***
dat2$면적    -1.252e-04  1.366e-04  -0.916   0.4562
dat2$피해금액 3.909e-04  1.615e-05  24.205   0.0017 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1524 on 2 degrees of freedom
Multiple R-squared:  0.9969,    Adjusted R-squared:  0.9937
F-statistic: 317.5 on 2 and 2 DF,  p-value: 0.00314

```

전진선택법을 통해서 회귀계수값이 낮은 요소들을 제외한 나머지 요소들을 독립 변수로 설정해 회귀분석을 실시하였다. 회귀모델의 결정계수는 0.9969로 약 99%의 설명력을 갖는 모델이라고 볼 수 있고 p-value가 0.00314로 유의수준 0.05보다 낮으므로 대립가설이 성립하여 이 회귀모델은 유의하다고 볼 수 있다. 각 독립변수의 p-value를 보았을 때 피해금액이 0.0017로 0.05보다 낮기 때문에 피해금액이 가장 큰 인과관계를 갖는다.

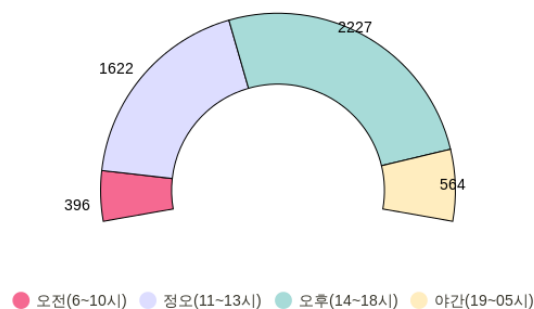
따라서 피해금액이 산불 발생횟수와 가장 큰 관계를 갖고 있다는 것을 상관분석과 회귀분석을 통해 통계적으로 증명하였다. 즉 과거의 데이터와는 다르게 미래에는 대규모 산불이 발생할 것이라는 예측을 할 수 있다.

### 3. 분석 결과 및 활용 방향

지금까지 데이터로 보았을 때, 등산객들의 방문수와 건조한 날씨, 풍속이 산불 발생 횟수에 영향을 미쳤었다. 미래 예측을 통해 얻은 가장 큰 영향을 미치는 요소는 풍속이지만 피해금액이 상관분석을 통해 가장 큰 수치인 것을 발견했고 이는 과거에 비해서 미래에 규모가 더 큰 대형산불이 발생할 가능성이 높다는 것을 예상할 수 있었다.

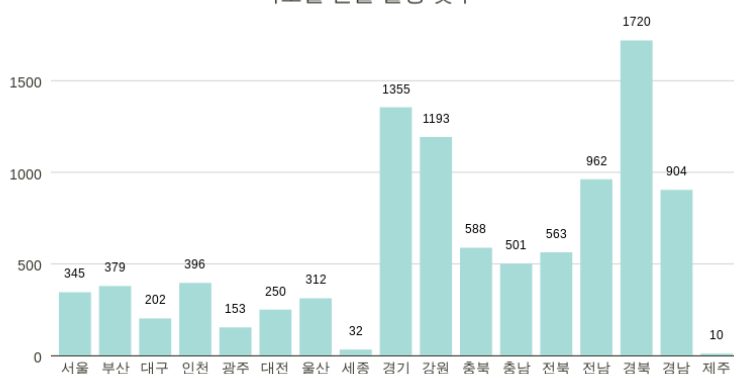
과거의 데이터와 미래의 예측값을 통합해서 미래의 규모가 더 큰 산불을 예방하기 위해서는 건조한 날씨에는 등산객들에게 산불을 주의하라는 경고가 필요할 것이라 생각한다.

시간대별 산불 발생 횟수

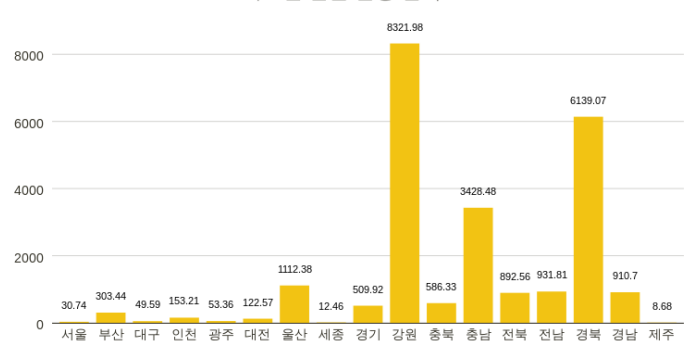


시간대별 산불 발생 횟수를 보면 정오와 오후에 가장 산불이 많이 발생한다는 것도 알 수 있으므로 이 시간대에 산불 감시 인력을 보강해야한다.

시도별 산불 발생 횟수



시도별 산불 발생 면적



시도별 산불 발생횟수와 면적 데이터를 보았을 때 강원과 경북이 산불 발생횟수와 면적이 가장 크고 실제로 통계적인 분석을 해보았을 때에도 면적과 산불 발생 횟수의 상관계수가 약0.63으로 서로 연관성이 있다.

따라서 강원, 경북 지역의 산불 전문 소방 인력을 증가시키고 수치화된 현수막을 통한 입산자들의 인식주의, 주민들에게 산불 예방 교육을 하는 노력, 등이 필요하다.

대형산불 발생했던 산들의 입산자 수, 화제원인 등의 데이터를 취합하여 예측을 통해 좀더 자세한 각각의 원인을 파악할 수 있다고 생각한다.

또한 산불 초기 발견자에게 포상금을 만드는 정부의 노력과 건조한 날씨를 예방하기 위해 인공비를 내리게 할 수 있는 기술 개발에 인력과 예산을 더 투입해 이러한 기술이 더 빨리 개발될 수 있도록 하는 노력이 필요하다.

한국농림기상학회지의 '모형실험에 의한 풍속변화에 따른 산불의 확산속도와 강도 분석' 논문에 따르면 산불의 강도는 풍속 0.5 m/sec 에서는 0.183 kW/m, 1 m/sec 에서는 0.259 kW/m 로 0.5m/sec 에 비하여 약 1.4 배정도 차이가 났으며, 속도가 빠르면 산불의 강도가 커지고 피해 면적이 넓어진다고 한다. 따라서 특정 이상의 풍속일 경우에는 등산객들의 입산을 자제시키는 제도가 마련되어야 한다.