
빼앗긴 산에도 봄은 오는가:

과거 산불 데이터를 기반으로 한 보험료 책정 제안

마.파.두.부 (마운틴 파이어 두목과 부하들)

2025.03.07 FRI

산불 분석 필요성

산불 연중화 현상

과거 주로 3월 중순 - 4월 중순에 집중

현재 12월 - 1월 산불 발생 급증

- 겨울 기온 상승

5월 - 6월 산불 발생 급증

- 기후변화/이상기후

피해 보장 방법 부재

국가 풍수해 피해 보험 운영

- 산불 피해 미포함

임업인 및 임산물 재해 보험 운영

- 산지 피해 보장 X

민간 산림 및 산불 피해 보험 부재

- 발생 시기 및 피해 규모 예측 어려움

sanbul.csv 전처리 과정

1_컬럼제거

20%

높은 유사성을 가진 컬럼 ↑
nan값의 정도, 내용의 명확성을 기준으로 불용 컬럼 판단

2_컬럼제거 2

40%

‘발생원인_세부원인’, ‘발생원인_구분’, ‘발생장소_동리’, ‘발생장소_읍면’,
‘발생장소_시군구’, ‘발생장소_관서’, ‘발생일시_요일’ 제거

3_데이터 병합

75%

datetime을 이용해 기온, 습도, 풍속 데이터의 년, 월, 일 분리
지역명 추가하여 조건 생성 후 일치하는 행에 **merge**

4_소모시간 추가

100%

datetime을 분단위로 변환 (화재 진압시간 - 화재 발생시간)
결과 오류 수정

한눈에 보는 sanbul.csv

화재 발생 : 자연 VS 인재

인재

전체 화재 데이터 5501건 중 20건의 낙뢰를 제외하고는
실화, 사고가 대부분을 차지

자연적 요소로 화재 예측
가능한가?

가능

BUT 화재가 발생하기 쉬운 조건의 기상학적 예측일뿐
발생 원인의 대부분이 인재이므로 완벽한 분석 어려움
(현재 산림청에서도 자연 요소를 기준으로 경보 시스템 진행 중)

계절은 4개로 분할 되나?

X

온도, 습도 등을 기준으로 클러스터링 후 2가지 혹은 3가지로 분류
특정 조건이나 목적에 따라 계절을 다르게 묶을 필요 있음

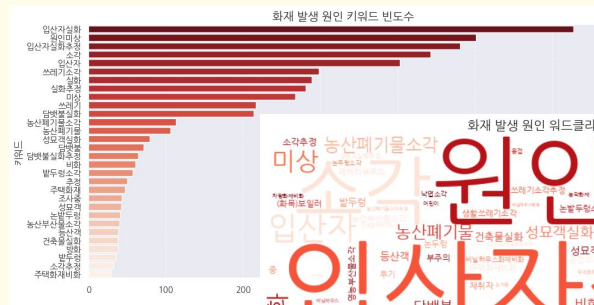
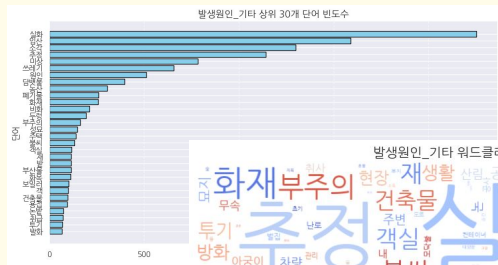
지역에 따른 화재
위험성이 있나?

O

지역마다의 습도, 기온, 풍속이 다르며
인적 요소(농장, 공장, 등산객 등)에 따른 지역적 특성 존재

1_보험료 산정 요인_화재 원인 워드클라우드

발생원인 상위 30개 단어, 키워드 분포 확인



단어 빈도 분석 결과 입산, 실화 등의 분포 확인

하나의 값에서 중복 추출된 건이 많아 키워드로 재분류

1_보험료 산정 요인_화재 원인 Torch

```
import torch
from transformers import AutoTokenizer,
AutoModelForSequenceClassification

# 사전 학습된 모델 및 토크나이저 불러오기
model_name = "beomi/KcELECTRA-base-v2022"
tokenizer = AutoTokenizer.from_pretrained(model_name)
model = AutoModelForSequenceClassification.from_pretrained(model_name, num_labels=6)

# 분류 기준 정의
labels = ["부주의에 의한 발생", "고의로 인한 발생", "자연적 요인",
         "인공적 요인", "소각등의 요인", "교통 관련"]
```

데이터	모델 출력값
화목보일러재처리부주의	교통 관련
조사중	교통 관련
농산폐기물소각	인공적 요인
원인미상	소각등의 요인
담뱃불실화	소각등의 요인
입산자실화(취사행위)	소각등의 요인
쓰레기소각	소각등의 요인

Torch 기반 모델을 활용하여 분류를 시도하였지만 유의미한 분류 불가

1_보험료 산정 요인_화재 원인 K-means

```
from sklearn.feature_extraction.text import TfidfVectorizer  
from sklearn.cluster import KMeans
```

TF-IDF 변환

```
vectorizer = TfidfVectorizer()  
X = vectorizer.fit_transform(df1["발생원인_기타"])
```

KMeans 클러스터링 (6개 카테고리 자동 분류)

```
num_clusters = 6  
kmeans = KMeans(n_clusters=num_clusters, random_state=42,  
n_init=10)  
df1["발생원인_분류"] = kmeans.fit_predict(X)
```

데이터	모델 출력값
조사중	0
농산폐기물소각	0
원인미상	6
입산자 실화추정	3
입산자 실화	3
입산자실화	1

K-Means 기반 모델을 활용하여 분류를 시도하였지만 유의미한 분류 불가

1_보험료 산정 요인_화재 원인 SBERT 모델

```
from sklearn.metrics.pairwise import cosine_similarity
from sentence_transformers import SentenceTransformer
```

사전 학습된 SBERT 모델 로드

```
model = SentenceTransformer('all-MiniLM-L6-v2')
```

발생 원인 리스트 가져오기 (중복 제거)

```
unique_causes = df1["발생원인_기타"].dropna().unique()
```

발생 원인을 벡터로 변환

```
cause_vectors = model.encode(unique_causes)
```

코사인 유사도 행렬 계산

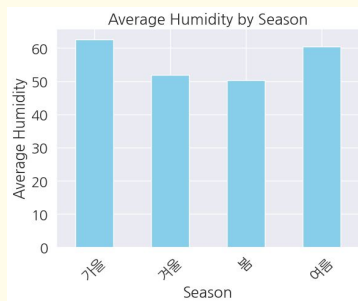
```
similarity_matrix = cosine_similarity(cause_vectors)
```

모델 출력값	
조사중 그룹	'조사중', '아궁이재처리부주의', '입산자실화', '재처리부주의'
농산폐기물소각 그룹	'농산폐기물소각', '담뱃불실화', '밭두렁소각', '차량화재비화'
원인미상 그룹	'원인미상', '농산부산물소각', '농산부산물소각 (깻대)', '미상'
주택화재비화 그룹	'입산자실화추정', '주택화재비화', '화목보일러재처리부주의', '입산자실화'
입산자실화 그룹	'조사중', '입산자실화(취사행위)', '입산자실화추정', '주택화재비화'

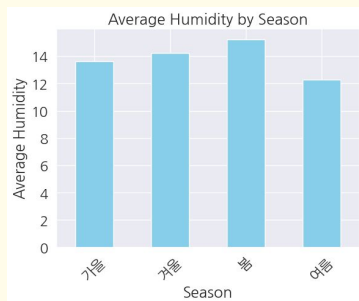
SBERT 기반 모델을 활용하여 분류를 시도하였지만 유의미한 분류 불가

1_보험료 산정 요인_환경 요소

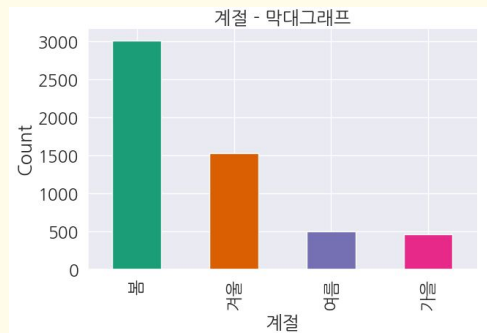
습도



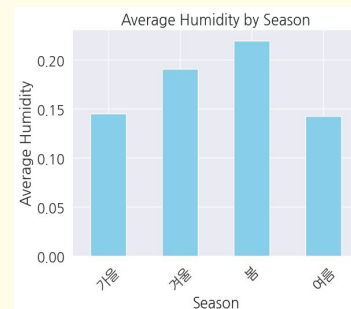
풍속



건수



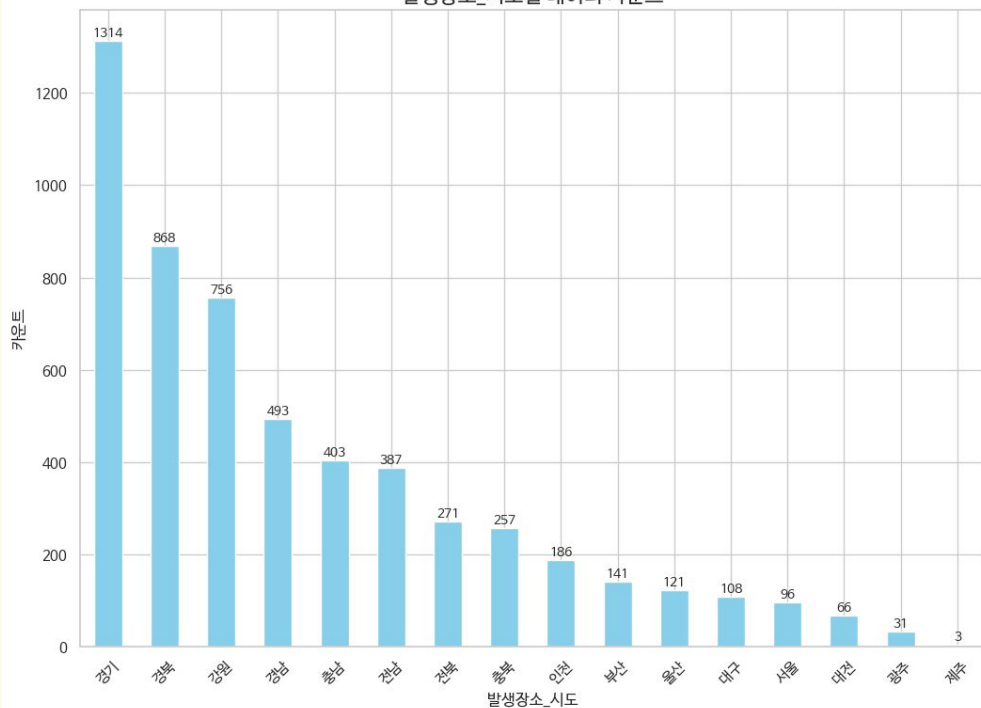
피해면적



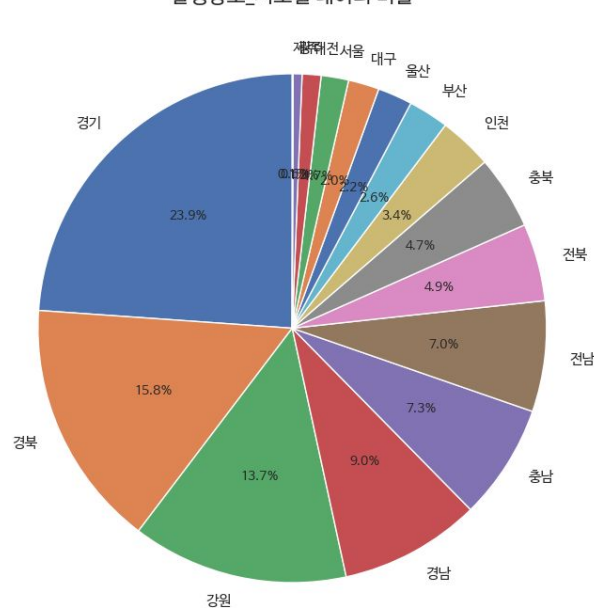
화재 요인인 습도, 풍속에 따라 건수, 면적이 비례하는 것을 확인

1_보험료 산정 요인_환경 요소

발생장소_시도별 데이터 카운트



발생장소_시도별 데이터 비율



2_보험료 산정 과정_

ML: 지도학습

지도학습(Supervised Learning)

① 분류(Classification) 알고리즘

산불이 발생할 확률 예측

위험 산불 여부 판별

② 회귀(Regression) 모델

피해 면적과 진압 시간 예측

산불 피해 규모의 정량적 평가

○ 타깃 1 : 산불 발생 가능성

사용된 분류 알고리즘 :

Random Forest Classifier

○ 타깃 2 : 진압시간

분류 진행 기준:

0 : 110분 이하

1 : 110분 이상

사용된 분류 알고리즘:

Logistic Regression

○ 타깃 3 : 위험 산불 발생 가능성

사용된 분류 알고리즘 :

Logistic Regression

Random Forest Classifier

○ 타깃 4 : 피해면적

사용된 회귀 알고리즘:

Linear Regression

Polynomial Regression

XGBoost Regressor

2_보험료 산정 과정_

ML: 비지도학습

- 지도학습 분류 모델의 과적합 보완
- 낮은 위험 산불 분석 필요성
- 실용적 군집화 목적

○ 군집화 목적

기상, 산불 발생 데이터, 계절에 따른 유사한
기상·환경 조건을 클러스터링하여 분석

○ 군집화 알고리즘 채택 : K-Means

PCA를 활용하여 차원 축소
3차원 그래프로 클러스터링

○ 도출 인사이트 : 계절 & 산불 발생 현상

낮은 산불 환경이 기후 변화로 인해
4계절이 아닌 3개의 계절로 분리가 되는 것으로 확인

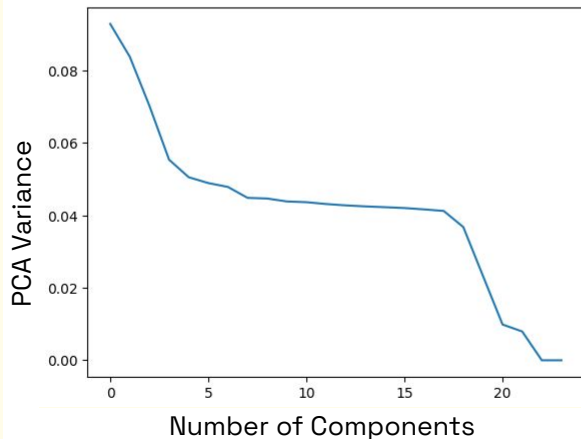
○ 도출 인사이트 활용 : 보험 상품

LA 산불 이전 계절적 피해를 크게 보아 산불 피해
보상을 중단한 보험사들을 위해 계절적 리스크를
감안한 보험 가입비 계산 알고리즘에 클러스터를 활용

2_보험료 산정 과정_클러스터링

PCA

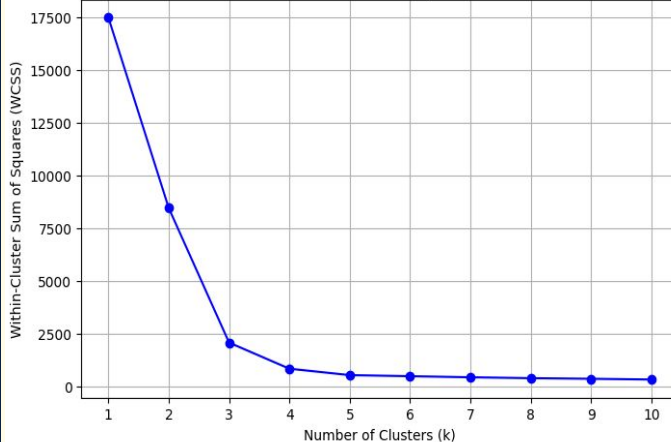
이상적 PCA 컴포넌트 수 지정



N_component에 있어 3이 이상적이지 않지만 클러스터링 시각화를 위해 **최대 시각화 가능한 3**으로 지정

Elbow Method

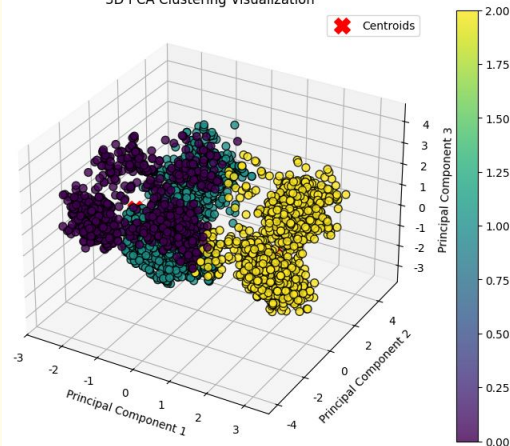
Elbow Method for Optimal k



Elbow Method 그래프로 가장 WCSS가 줄어드는 가파른 지점 **k = 3** 선정

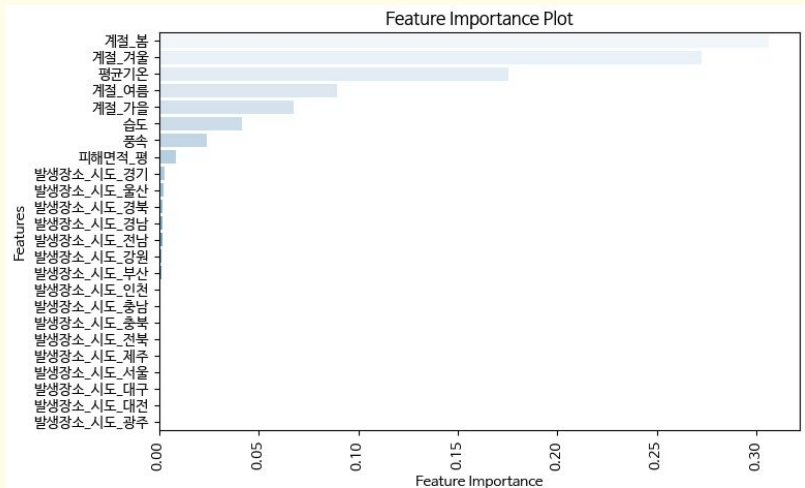
클러스터링 시각화

3D PCA Clustering Visualization

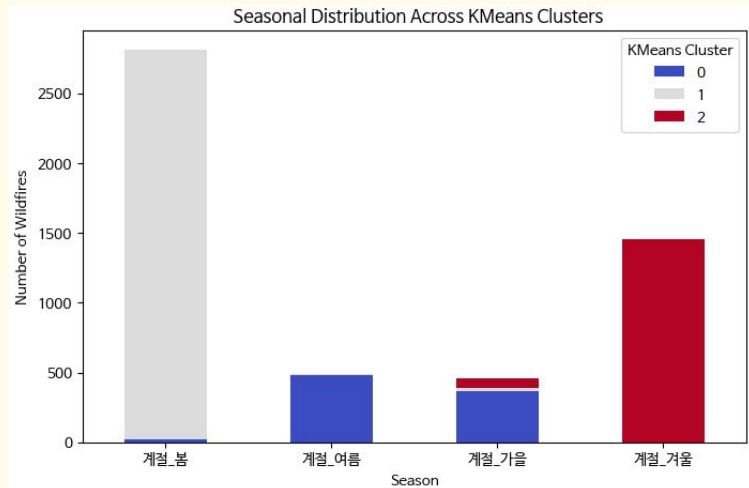


군집화가 3차원으로 시각화하고 군집의 분리도를 Davies-Bouldin Index로 **1.98**로 확인

2_보험료 산정 과정_클러스터링 계절



클러스터에 대한 Feature Importance를 시각화한 결과,
각 계절 칼럼 및 기상 데이터가 판별에 높은 영향을 끼친다

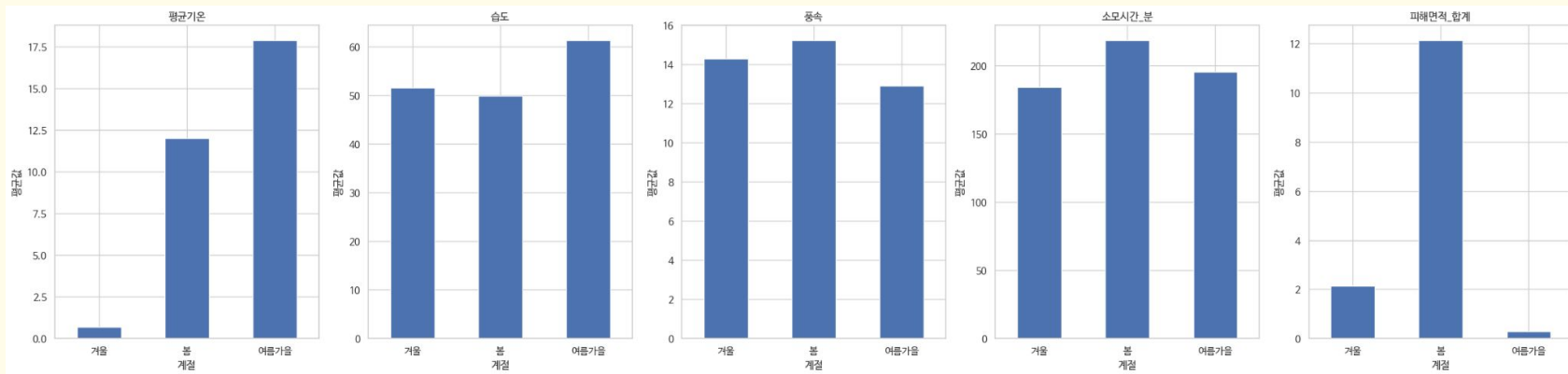


- 계절별 클러스터 분포를 분석 했을 때, 봄, 여름/가을, 겨울로 분리
- 기후 변화로 인해 4 계절이 아닌 3 계절 산불 기상 환경으로 분리되는 것으로 추측

계절과 클러스터 분리에 대한 상관성 Chi-Square Test p-value: 0.0

결론 : p-value가 0.05 이하로 상관성 ↑

2_보험료 산정 과정_클러스터링 계절



[봄] 3-5월

[여름, 가을] 6-10월

[겨울] 12-2월

2_보험료 산정 과정_알고리즘



산림청

Logistic Sigmoid Function 벤치마킹

(1) 봄철 산불위험지수 산출 알고리즘

$$[1 + \exp\{-(2.706 + (0.088 \times Tmean) - (0.055 \times Rh) - (0.023 \times Eh) - (0.014 \times Wmean))\} - 1]^{-1}$$

(2) 가을철 산불위험지수 산출 알고리즘

$$[1 + \exp\{-(1.099 + (0.117 \times Tmean) - (0.069 \times Rh) - (0.182 \times Wmean))\} - 1]^{-1}$$

- Tmean : 평균온도
- Rh : 상대습도
- Eh : 실효습도
- Wmean : 평균풍속

보험 가입비 계산을 위한 위험도 점수 예측 모델 제작

• 보험 가입비 계산 로직

보험 가입비 = ((재산_평) × 평당 가격_원) × (1 + 100 × 산불 위험도 × Risk Multiplier)

* 평당 가격과 Risk Multiplier는 고정 지수며 변경 가능하다

• 위험도 점수를 계산하는 로직

산림청과 같은 Logistic Sigmoid Function을 사용하기 위해 타깃 값을 피해면적 3000평 이하는 0, 3000 < x < 6000평은 1로 선정함. 클러스터별 데이터를 트레이닝한 결과, 클러스터별 알고리즘은 다음 예시처럼 적용됐다:

* (1) 클러스터 2 산불위험지수

$$-[1 + \exp\{-(-277.8672 + (0.1147 \times Tmean) + (0.09195 \times Area) - (0.02595 \times Rh) - (0.2069 \times Wmean) + \sum \beta_{\text{시도}} + \sum \beta_{\text{계절}})\} - 1]^{-1} - \left[1 + \exp\{ \left(- \left(-277.8672 + (0.1147 \times Tmean) + (0.09195 \times Area) - (0.02595 \times Rh) - (0.2069 \times Wmean) + \sum \beta_{\text{시도}} + \sum \beta_{\text{계절}} \right) \right) \} - 1 \right]^{-1} - [1 + \exp\{-(-277.8672 + (0.1147 \times Tmean) + (0.09195 \times Area) - (0.02595 \times Rh) - (0.2069 \times Wmean) + \sum \beta_{\text{시도}} + \sum \beta_{\text{계절}})\} - 1]^{-1}$$

소 잃고 외양간 고친다?



21세기 화전민

미리 대비하지 못한 재난은 재앙입니다.
여러분이 가진 사유림은 여러분이 복구
가능한 재산인가요?

익숙함에 속아 잃을 수 있는 재산,
이제는 대비할 시기입니다.

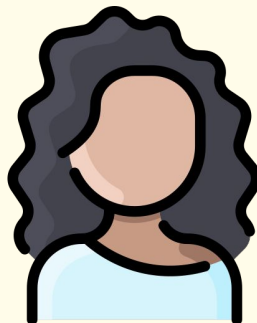
3_가입 희망자 가입비 책정 예시



홍 만 재 (57)

- 강원특별자치도 삼척시 사유림 소유
- 2,000평 (약 0.67 헥타르)
- 2월 가입 희망 (2~4월)

- Predicted Cluster: 0 (3,000평 이하)
- 🔥 화재위험 리스크 점수: 27.63
- 💰 총 가입비: 3,105,200 원
월 1,035,066 원



박 길 자 (86)

- 경기도 양주시 사유림 소유
- 3,025평 (약 1 헥타르)
- 9월 가입 희망 (9~11월)

- Predicted Cluster: 1 (3,000-6,000평)
- 🔥 화재위험 리스크 점수: 27.82
- 💰 총 가입비: 4,708,110원
월 1,569,370원

4_적용 가능 서비스 배너

삼점삼.

🔥 [전국민] 세금 환급 신청기간

1인 평균 환급액 **175,200원**
지금 무료로 확인해 보세요!

* 환급액 조회자 3명 중 1명은 환급대상

내 숨은 돈 확인하기



내 보험

BOHUM 조회몰

₩ ??? 원



Q & A

A scenic landscape featuring a calm lake reflecting the surrounding forest and distant mountains. In the foreground, a wooden bench sits on a grassy bank. The scene is framed by large green shapes on the left and top right, which contain the words 'You' and 'Thank' respectively.

Thank

You