



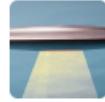
관측 데이터셋을 통한 UFO 출몰 경향 분석 및 예측

소프트웨어학부
20170730 김성권

목 차

1. 분석
2. 가설 검증
3. 회귀 예측
4. 결론

데이터 가공



Dataset

UFO Sightings (Approx. 90,000)

by Rishi Damarla

2 years ago • 40 MB • ^ 34

12 columns x 88125 records

city: 관측 도시

state: 관측 주

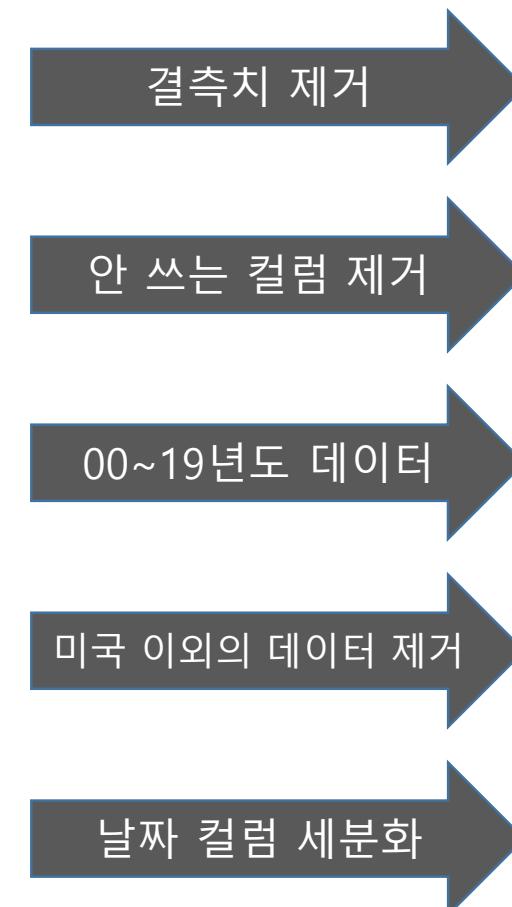
date_time: 관측 날짜, 시각

shape: 관측된 UFO의 형태

...

city_latitude: 관측 위도

city_longitude: 관측 경도



9 columns x 65420 records

state

shape

city_latitude

city_longitude

year: 년

month: 월

day: 일

Date: 년-월-일

ord_date: 그레고리력 서수

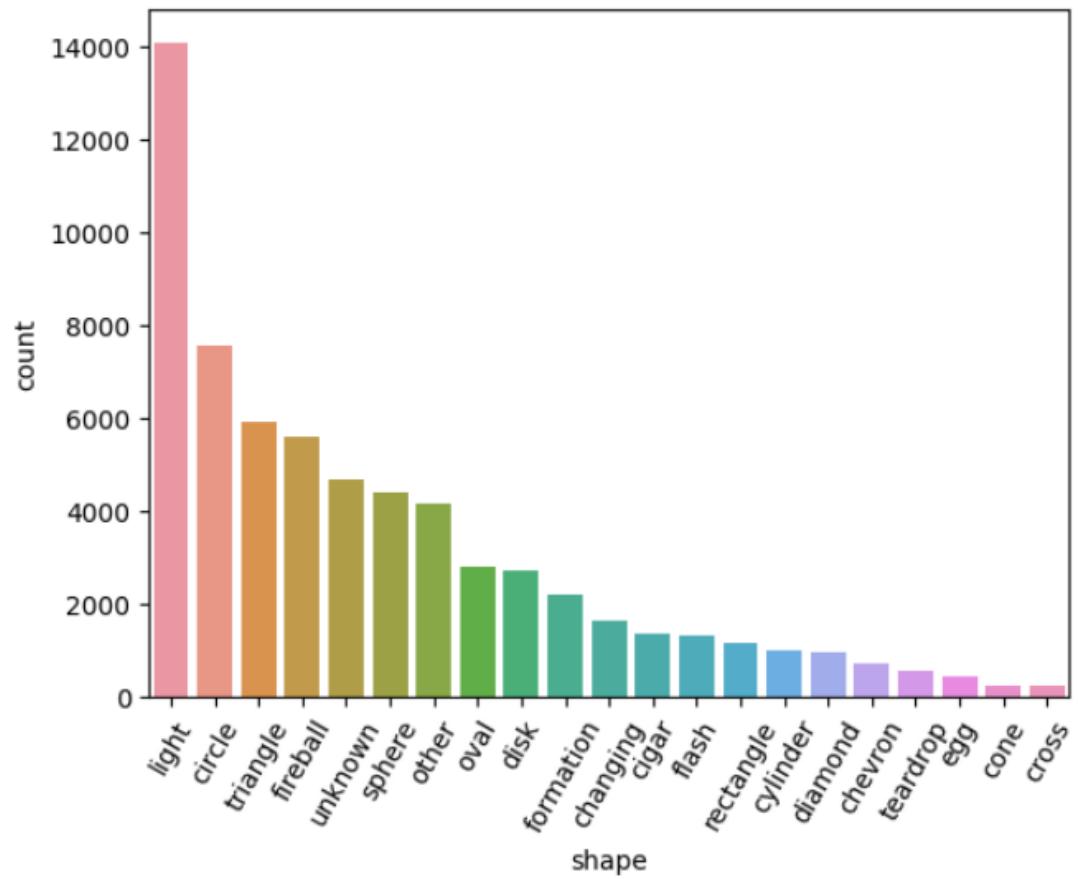
데이터

	state	shape	city_latitude	city_longitude	year	month	day	Date	ord_date
0	VA	light	37.343152	-77.408582	2019	12	12	2019-12-12	737405
1	CT	circle	41.664800	-72.639300	2019	3	22	2019-03-22	737140
5	TX	disk	30.677200	-94.005200	2019	4	2	2019-04-02	737151
6	AZ	unknown	32.259941	-110.927542	2019	5	1	2019-05-01	737180
7	AZ	circle	33.371500	-111.436900	2019	4	10	2019-04-10	737159
...
88119	LA	light	29.699692	-91.069123	2019	10	2	2019-10-02	737334
88120	AZ	formation	33.414036	-111.920920	2019	10	2	2019-10-02	737334
88121	MO	light	37.642200	-93.399600	2019	10	2	2019-10-02	737334
88122	FL	formation	27.076210	-82.223280	2019	10	2	2019-10-02	737334
88123	NC	fireball	35.605000	-82.313200	2019	10	2	2019-10-02	737334

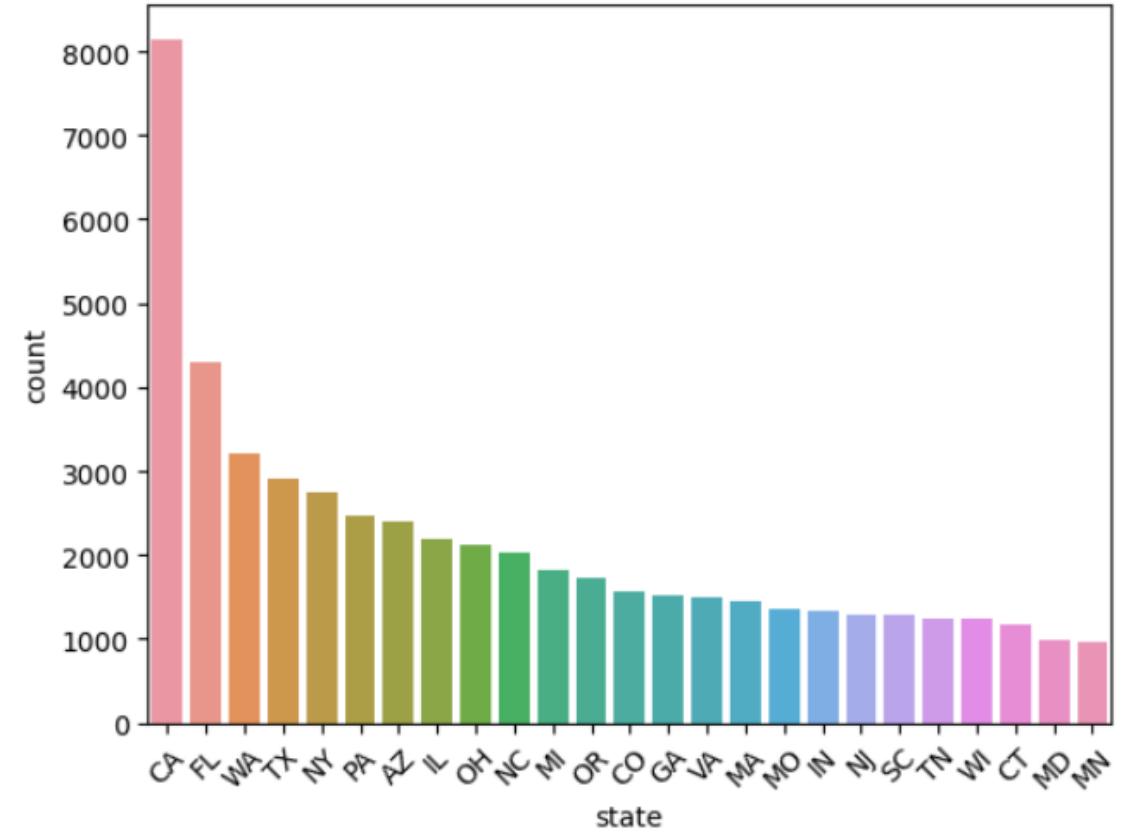
65420 rows × 9 columns

통계 분석

UFO의 모양 분포

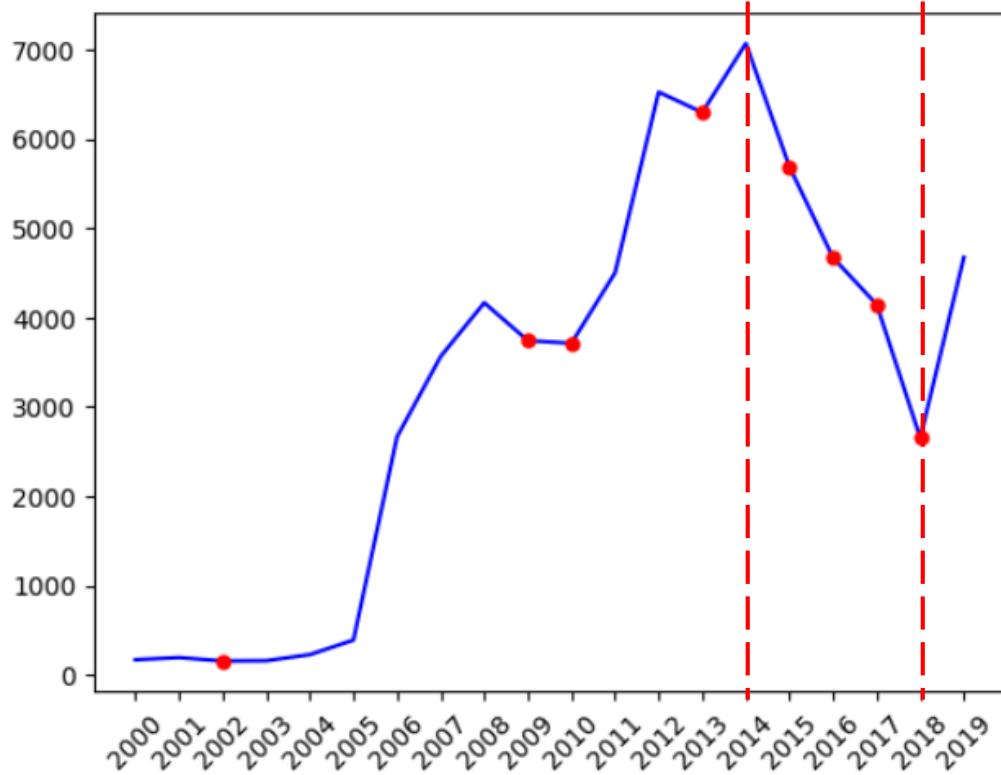


UFO가 자주 관측된 상위 25개 주 (state)



통계 분석

연도 별 UFO 관측 수



관측 수가 눈에 띄게 줄어드는 구간: 2014~2018

가설 검증

가설 1. UFO의 관측 수는 인구 수에 비례한다

추가 사용 데이터



US states Population 2000-2019
[yasser shrief](#) · Updated 2 years ago
Usability 2.9 · 1 File (CSV) · 5 kB

source: 미국 과학기술정보국(osti)

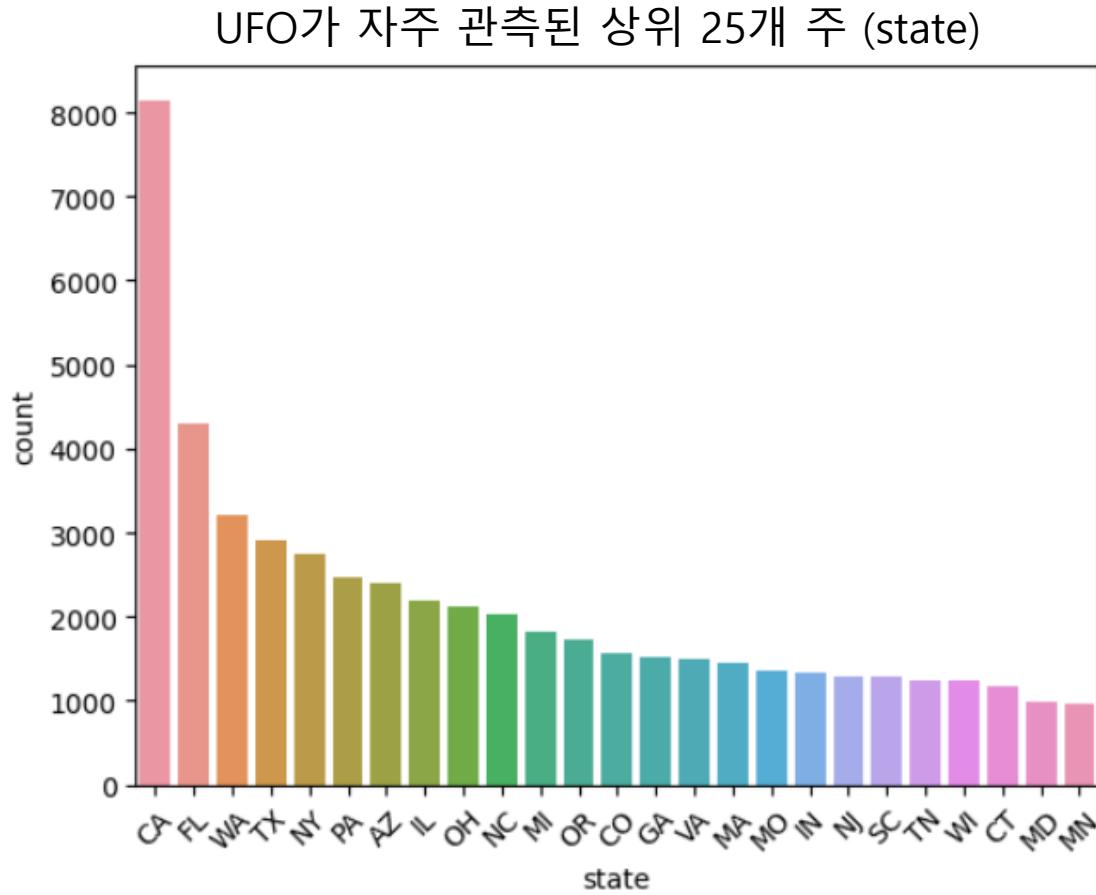
단위 수 설정

평균 컬럼 추가

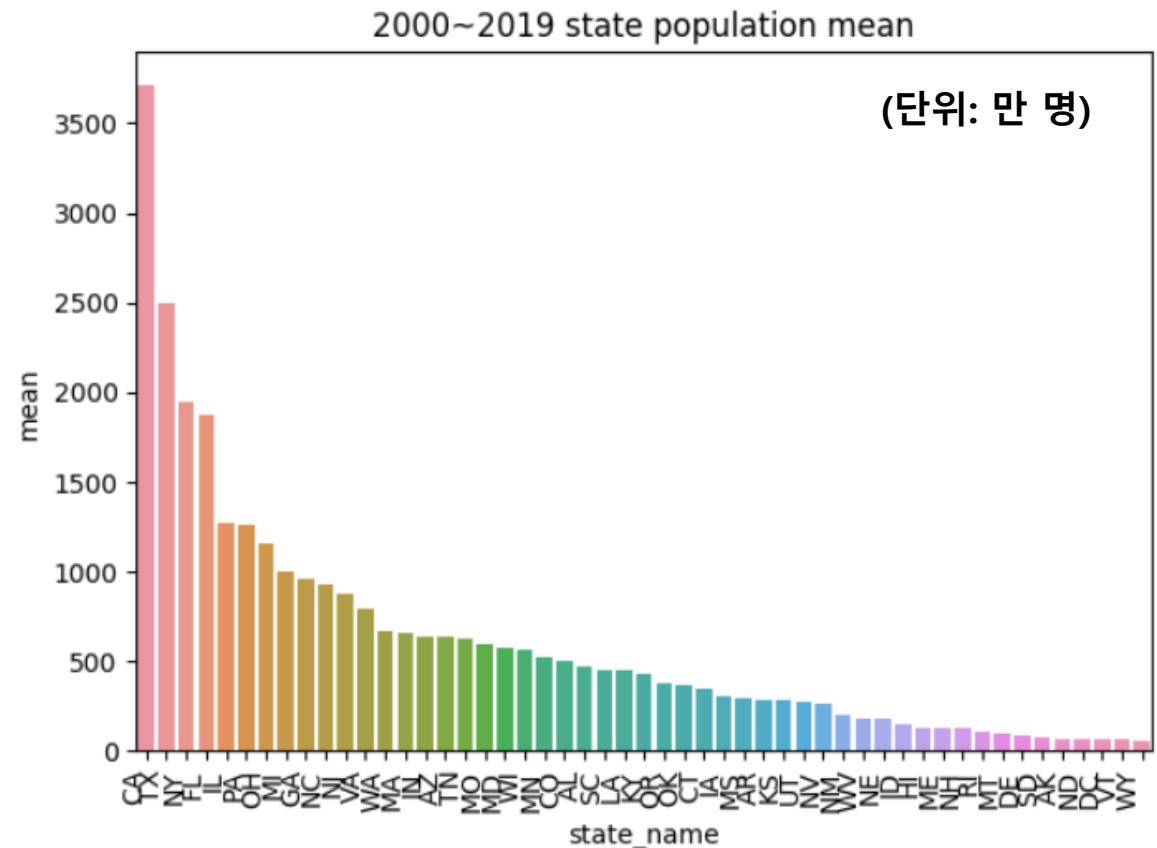
50+1 개의 주에 대한
2000~2019년의 인구수와 평균 인구수

가설 검증

가설 1. UFO의 관측 수는 인구 수에 비례한다



상위 3: CA, FL, WA

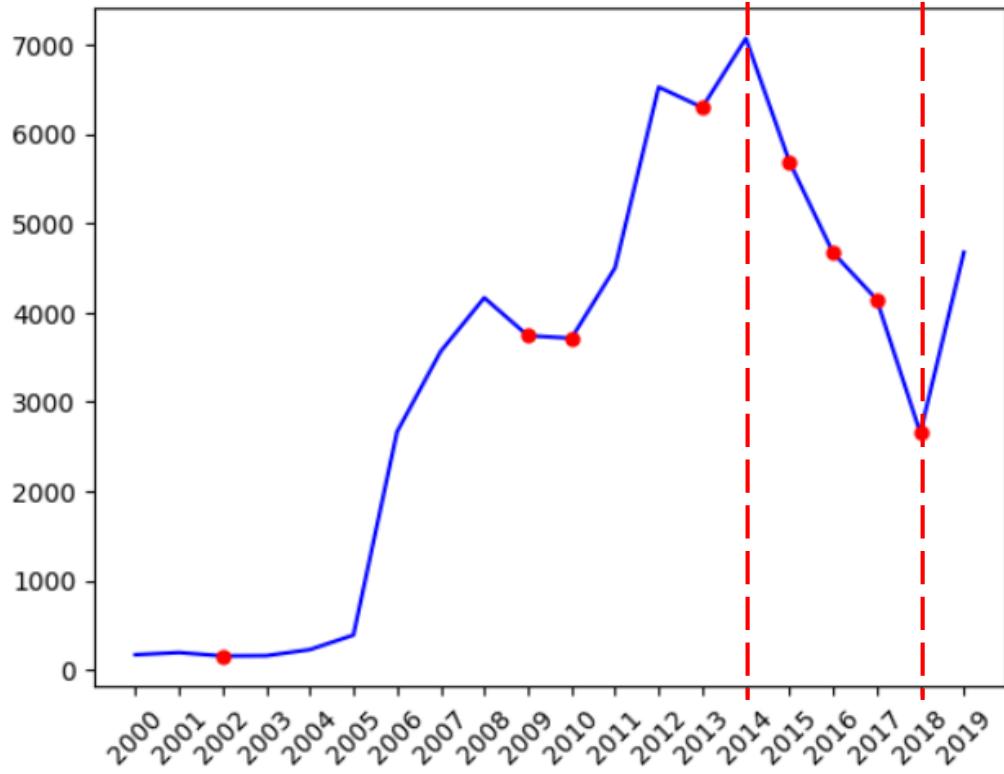


상위 3: CA, TX, NY

가설 검증

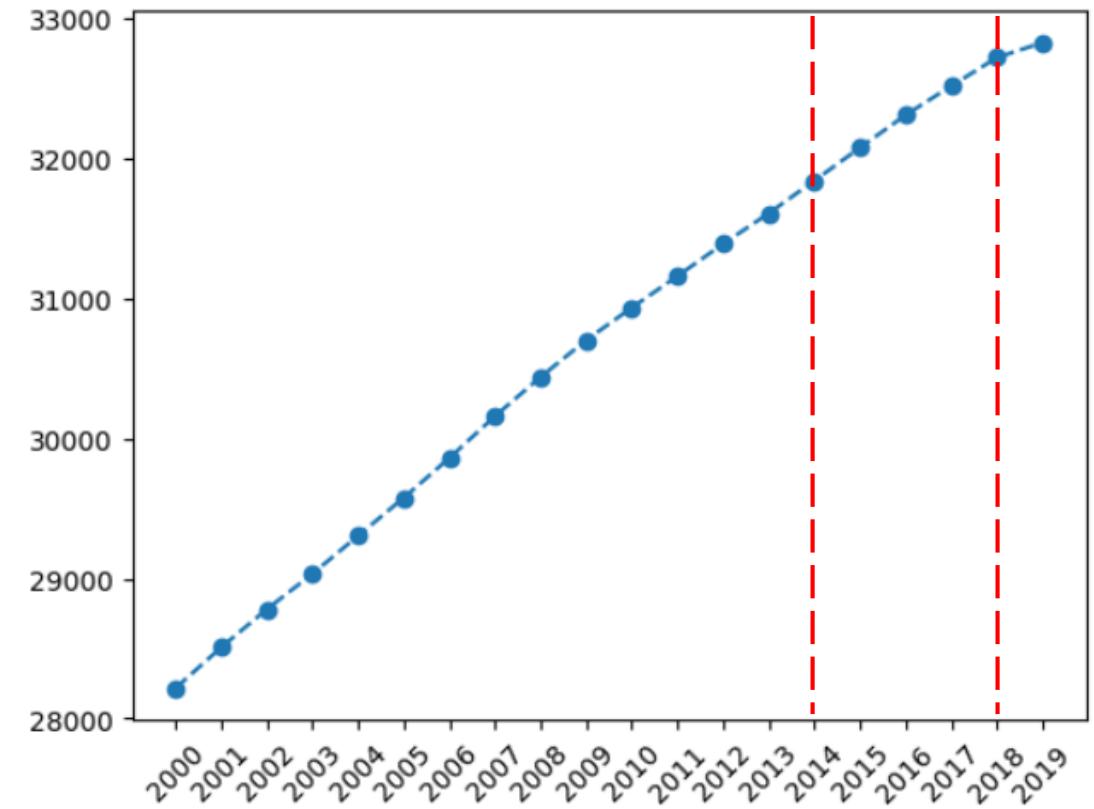
가설 1. UFO의 관측 수는 인구 수에 비례한다

연도 별 UFO 관측 수



미국의 전체 인구 수 변화 추이

(단위: 만 명)



가설 검증

가설 2. UFO는 하늘이 맑을 때 관측되기 쉽다.

추가 사용 데이터



Dataset

US Air Quality 1980-Present

by Caleb Reigada

source: 미국 환경보호청 (epa)

00~19년 데이터

여러가지 가공..

merge

UFO 데이터셋 + 대기오염 데이터(pollution, AQI)

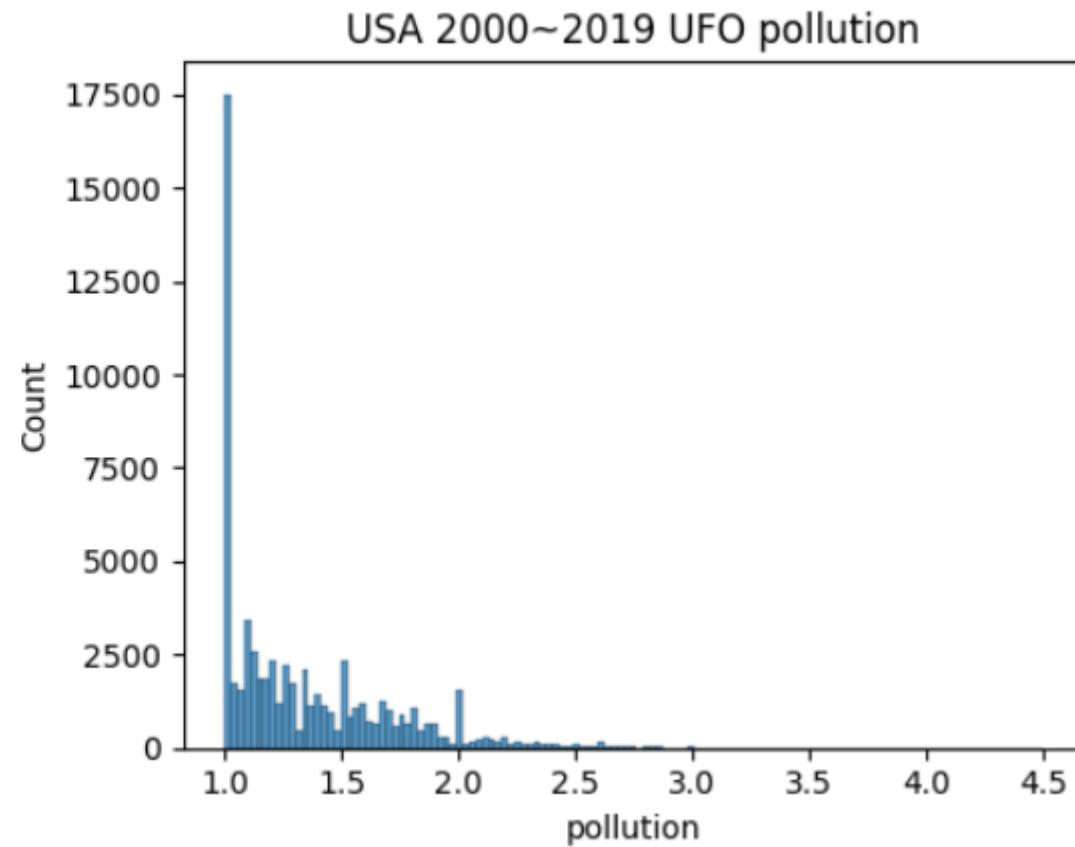
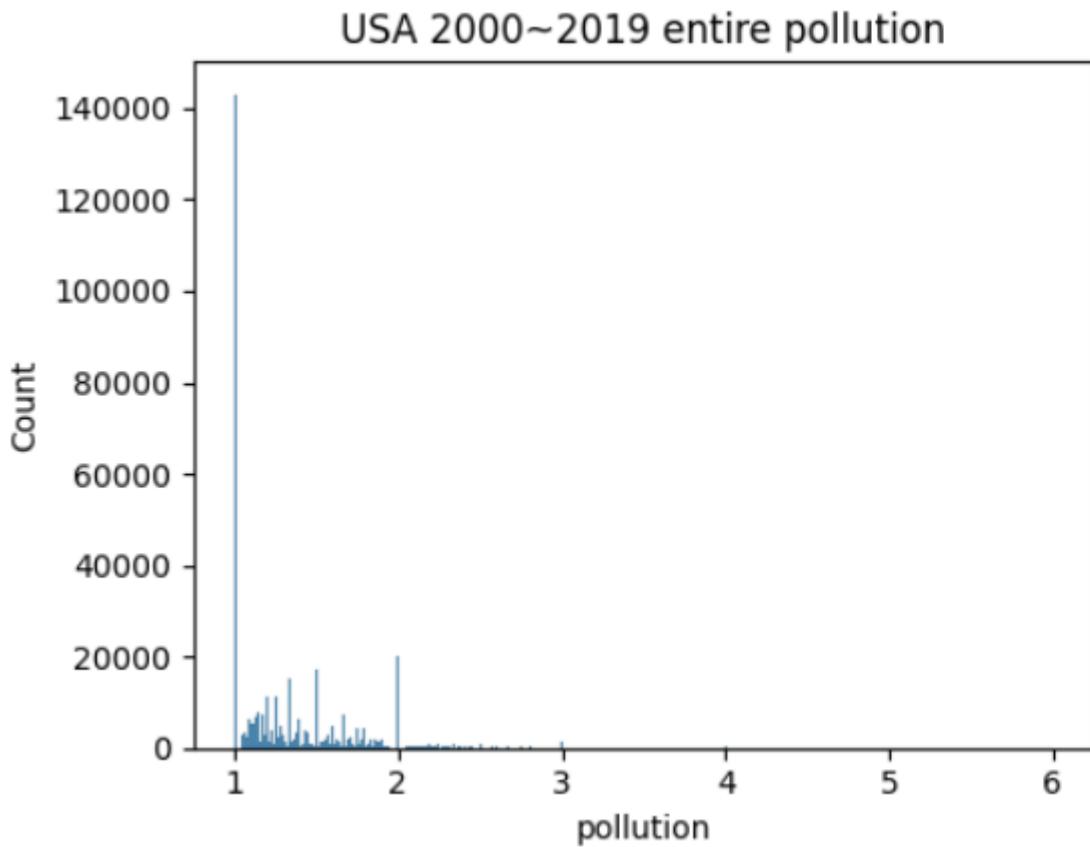
	state	shape	city_latitude	city_longitude	year	month	day	Date_x	pollution	AQI
0	VA	light	37.343152	-77.408582	2019	12	12	2019-12-12	1.000000	29.000000
1	CT	circle	41.664800	-72.639300	2019	3	22	2019-03-22	1.000000	41.400000
2	TX	disk	30.677200	-94.005200	2019	4	2	2019-04-02	1.363636	46.090909
3	AZ	unknown	32.259941	-110.927542	2019	5	1	2019-05-01	1.400000	44.100000
4	AZ	circle	33.371500	-111.436900	2019	4	10	2019-04-10	2.000000	76.200000
...
65415	LA	light	29.699692	-91.069123	2019	10	2	2019-10-02	1.571429	49.714286
65416	AZ	formation	33.414036	-111.920920	2019	10	2	2019-10-02	1.200000	43.100000
65417	MO	light	37.642200	-93.399600	2019	10	2	2019-10-02	1.142857	36.285714
65418	FL	formation	27.076210	-82.223280	2019	10	2	2019-10-02	1.000000	35.095238
65419	NC	fireball	35.605000	-82.313200	2019	10	2	2019-10-02	1.785714	58.428571

1: 좋음 2: 보통
4: 안 좋음 5: 매우 안 좋음

3: 민감한 사람에게 안 좋음
6: 치명적

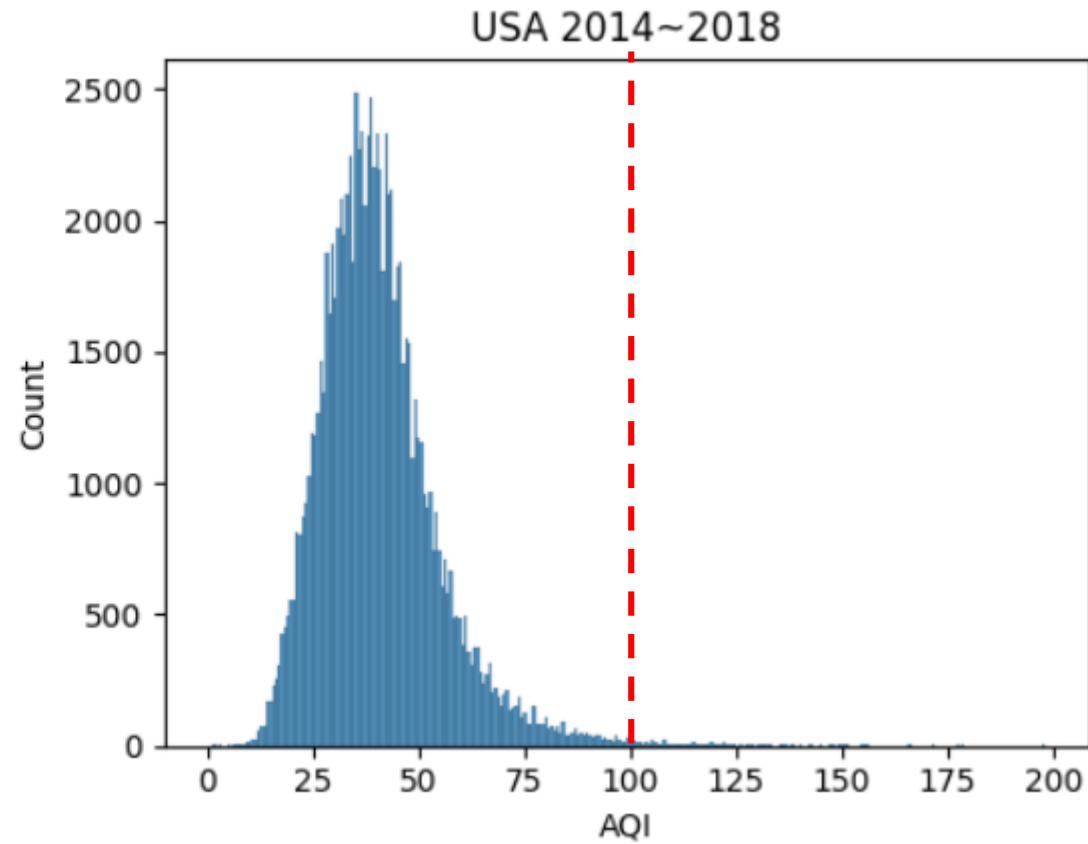
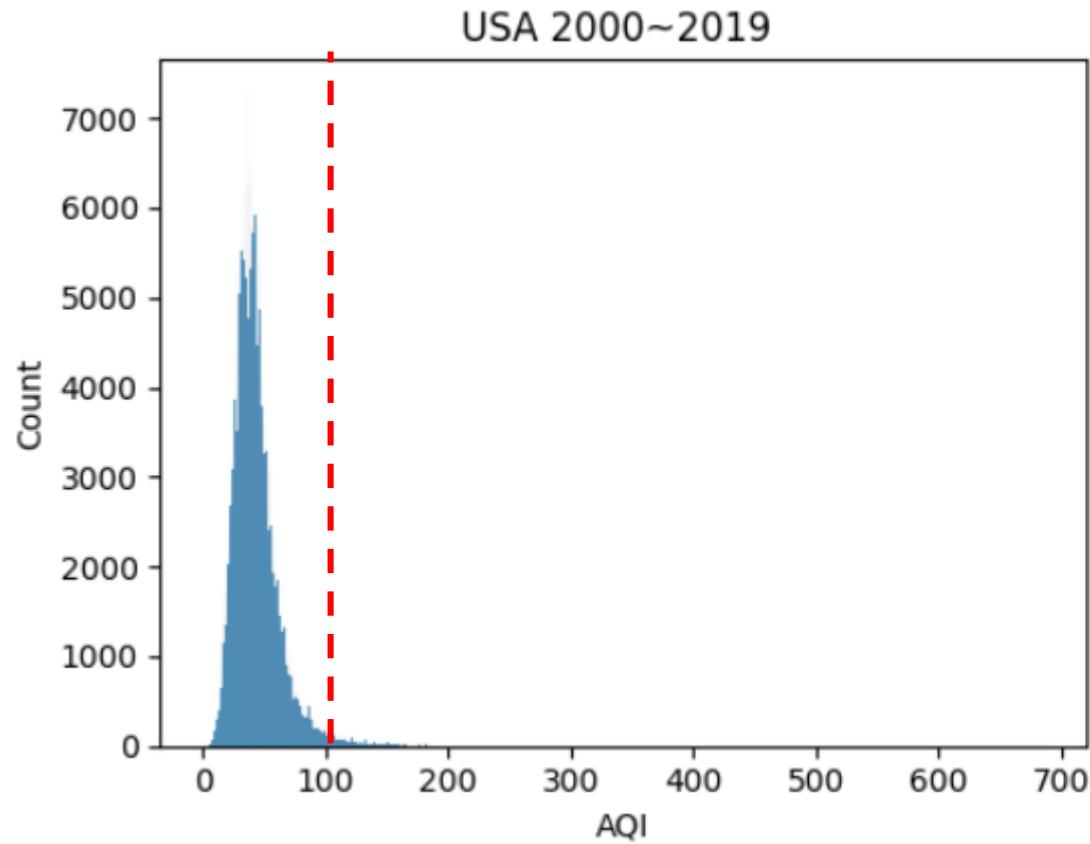
가설 검증

가설 2. UFO는 하늘이 맑을 때 관측되기 쉽다.



가설 검증

가설 2. UFO는 하늘이 맑을 때 관측되기 쉽다.



회귀 예측

```
X = ufo['ord_date'].values.reshape(-1, 1) # values.reshape(-1, 1) 메서드를 이용하여 1차원 배열을 2차원 배열로 변환
Y = ufo[['city_latitude', 'city_longitude']]

# LinearRegression() 메서드를 사용하여 선형 회귀 모델을 생성하고, fit() 메서드를 사용하여 X와 Y 데이터를 이용하여 모델을 학습
lr = LinearRegression()
lr.fit(X, Y)

Y_predict = lr.predict(X)

new_date = input("예상 날짜를 입력하세요(ex: 20AA-BB-CC) : ")
convert_date = dt.datetime.toordinal(pd.to_datetime(new_date))

new_X = pd.DataFrame({'ord_date': [convert_date]})

prediction = lr.predict(new_X)

print(prediction)

lat = prediction[0][0]
lng = prediction[0][1]
```

X = 그레고리력 서수
Y = 위도, 경도

사용자 입력값: (YYYY-MM-DD)



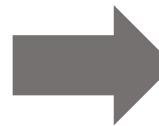
그레고리력 서수로 변환 (X value)



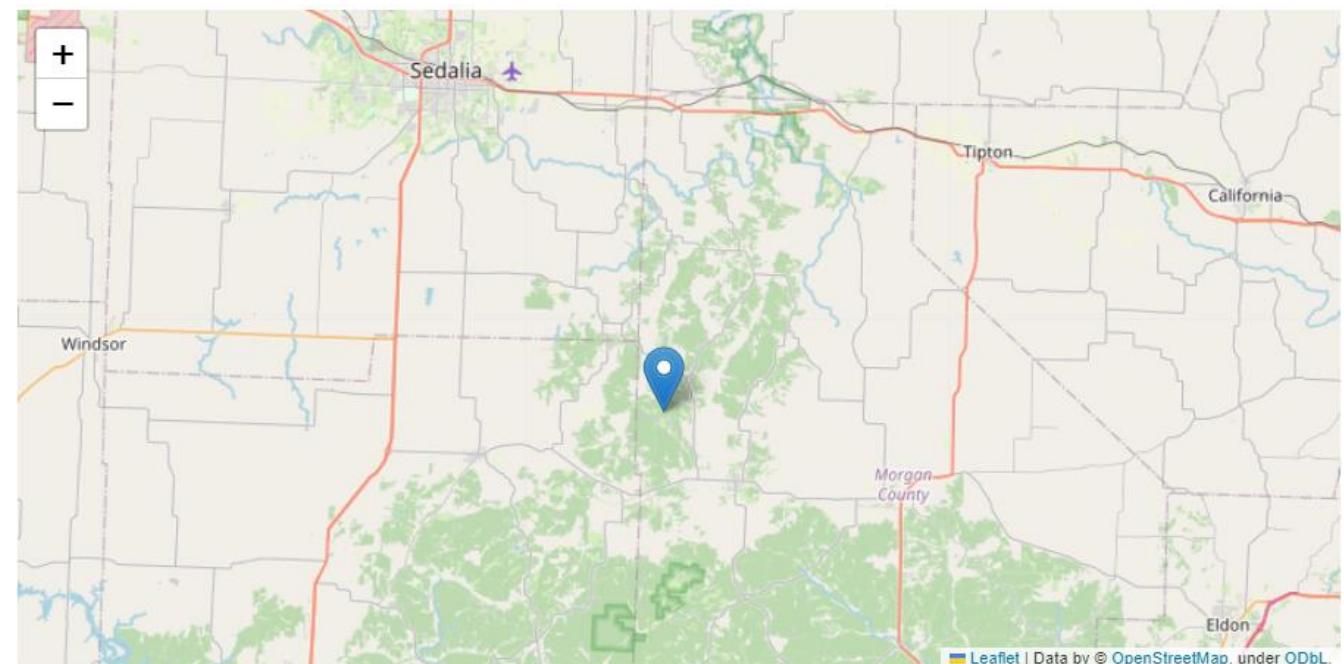
위도, 경도값 예상 (Y value)

회귀 예측

예상 날짜를 입력하세요(ex: 20AA-BB-CC) :



예상 날짜를 입력하세요(ex: 20AA-BB-CC) : 2023-07-14
[[38.48739629 -93.05045867]]



결 론

1. 가장 많이 관측된 UFO의 모양은 빛, 원형, 삼각형이며 인구수 대비 워싱턴에서의 관측률이 매우 높다.
2. 2014 ~ 2018년에 UFO 관측 수가 급락하는 이유는 밝혀지지 않았다.
3. 가설 1. UFO 관측 수는 인구 수에 비례한다 --> X
4. 가설 2. UFO는 하늘이 맑을 때 관측되기 쉽다. --> O