네이버 본사에서 UFO 보기

1. 양태양 (19930120)
2. [tyami@naver.com](mailto:tyami@naver.com)
3. 양태양 (19930120) / 김재훈 ( ) / 김종수 ( )



UFO를 보려면 어떻게 해야할까 블라블라 글을 쓰자

분석에는 Kaggle에 공개된 오픈데이터 중, “미국의 UFO 관측날의 대기 질 데이터”와 “서울시 대기 질 데이터”를 사용하였다. 각 데이터셋의 변수는 아래와 같다.

데이터셋 설명

* **미국의 UFO 관측날의 대기 질 데이터 (UFO Sightings + Air Quality)**

<https://www.kaggle.com/infof422henni/ufo-air-quality>

UFO sightings 데이터 (<https://www.kaggle.com/NUFORC/ufo-sightings>) 중 미국 지역 데이터와 U.S. Pollution Data (<https://www.kaggle.com/sogun3/uspollution>) 를 결합한 데이터셋

63,173개의 데이터 샘플, 24개의 변수

1. State.Code
2. City
3. State
4. Day
5. Month
6. Year
7. Hour
8. NO2.Mean: The arithmetic mean of concentration of NO2 within a given day
9. NO2.1st.Max.Value: The maximum value obtained for NO2 concentration in a given day
10. NO2.1st.Max.Hour: The hour when the maximum NO2 concentration was recorded in a given day
11. NO2.AQI: The calculated air quality index of NO2 within a given day
12. O3.Mean
13. O3.1st.Max.Value
14. O3.1st.Max.Hour
15. O3.AQI
16. SO2.Mean
17. SO2.1st.Max.Value
18. SO2.1st.Max.Hour
19. SO2.AQI
20. CO.Mean
21. CO.1st.Max.Value
22. CO.1st.Max.Hour
23. CO.AQI
24. ET: equal to 1 if a sighting has occured and 0 otherwise

* **서울시 대기 질 데이터 (Air pollutants measured in Seoul)**

<https://www.kaggle.com/jihyeseo/seoulairreport>

2017년 11월 17일부터 11월 24일까지 1시간 간격으로 측정된 서울시 지역구별 공기질 데이터셋

4,225개의 데이터 샘플, 8개의 변수

1. 측정일시
2. 측정소명
3. 이산화질소 (NO2) 농도 (ppm)
4. 오존 (O3) 농도 (ppm)
5. 일산화탄소 (CO) 농도 (ppm)
6. 아황산가스 (SO2) 농도 (ppm)
7. 미세먼지 (㎍/㎥)
8. 초미세먼지 (㎍/㎥)

# 데이터 처리

두 데이터셋의 대기 질 데이터를 동일하게 사용하기 위해 UFO sightings + Air Quality 데이터 중 \*.1st.Max, \*.1st.Max, \*.AQI 데이터셋은 제거하였다.

**UFO를 보기 위한 조건은 무엇이 있을까?**

데이터셋 중 크게 3가지로 나눔

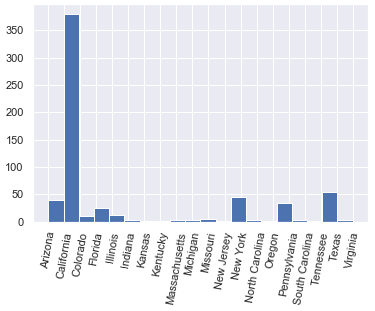
* **지역**

어느 지역에서 UFO가 많이 보였을까?

1. 미국 지도 위에 heat map 그려보자

**Insight:**

1. 주 별로 count hist plot

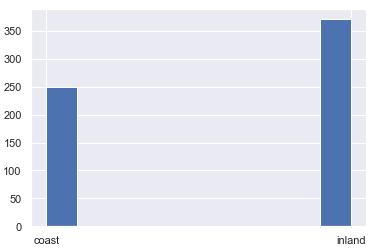


**Insight: California 주가 압도적으로 많은 편.**

1. 해안/내륙으로 나누어보자

주 별로 내륙주/해안주로 나눌 수 있지만, UFO 운전수에게는 인간의 행정구역은 중요치 않다  
**🡺 해안으로부터 20km 내에 있을 경우를 “coast”, 그 외를 “inland”로 나누었다**.

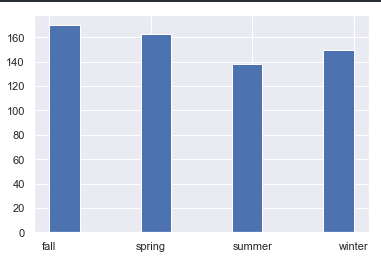
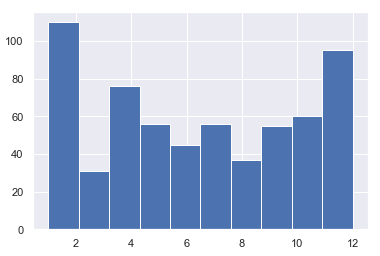
1. 해안/내륙 별 비교 hist plot



**Insight: 내륙 지역에서 발견되는 경우가 좀 더 많은 편이다.**

* **시간**

1. 월 🡪 계절 별 count hist plot



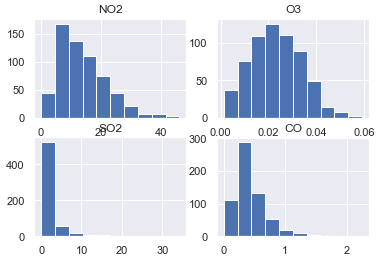
**Insight: 월별로는 차이가 나기는 하지만, 계절별로는 큰 차이가 없다. 각 계절간 n수가 3개씩뿐 안되기 때문에 별도의 통계분석은 하지 않았다.**

1. 시간 🡪 시간대 별 count hist plot

**Insight: 밤 > 새벽 > 오후 > 아침**

* **대기조건**

1. 각 대기조건별 count hist plot



#