0\_김도영 dataset의 원본dataset이 2011~2021강원도 산불 사건 통계이므로

<https://www.data.go.kr/data/3070842/openapi.do>에서

2011~2021 강원도 산불피해면적 을 뽑아 왔습니다

첨부파일(피해면적, 피해면적(강원))참고,

작업한 ipynb 파일은 추가로 사용할 일이 없을 것 같아서 첨부하지는 않습니다만 요청해주시면 바로 카톡방에 올리겠습니다

0\_김도영 instance 1056개

피해면적(강원) instance1519개

일일히 비교하면서 피해면적값을 입력했습니다…..

데이터셋을 만든 기관이 달라서 그런지는 몰라도 서로 겹치지 않는 산불사건들도 꽤 있었고 피해면적 파일에서도 중복되게 입력된 산불사건들이 꽤 있었습니다

피해면적을 입력할 수 없는 instance는 NULL이라고 입력해둠

Null 값이 예상 외로 많이 나옴

별도로 이상치와 결측치는 아직 제거를 못했는데 일단 다음 단계인 시각화를 진행하는데는 크게 무리가 없을 것 같아서 추후에 추가로 작업을 해보겠습니다.

피해면적 csv 는 위 링크에서 전국 산불사건 통계들

피해면적(강원)은 제가 편의를 위해 그중에서 강원도 사건들만 뽑아온 파일입니다

결측치, 이상치 처리

2\_김시형 csv에서

피해면적 = NULL 인 값인 인스턴스 제거

온도, 습도에서 결측치 또는 이상치가 값인 인스턴스 제거

풍향 feature 삭제 ( 분석에 필요한 instance 개수 부족이 우려되어 삭제함)

풍속 feaure 에서 NULL 값인 인스턴스 제거 하였습니다

총 1052개의 instance 에서 결측치 및 이상치 처리후

총 538개의 instance 남은 상태입니다!