

2023-12-09 회의

이전에 이어서 현재까지 완료 한 작업

1. 데이터 이상치 추출 완료
2. 네이버 크롤러 수정 완료

남은 작업

1. 데이터 시각화 및 EDA
2. 모든 csv를 한꺼번에 전처리 및 이상치 추출 할 수 있게 자동화
3. 추출된 이상치를 크롤러에 넣어 확인
4. PPT제작

변경사항

1. 기존에 네이버 뉴스기사로만 이상치의 원인을 분석했다, 하지만 데이터를 분석하는 과정에서
2022년 5월 15일 학여울역이 평소보다 2~3배 이상 인원이 많았지만 뉴스기사가 없어서 원인을 알 수 없었다.

조사 결과 학여울 역 근처에있는 SETEC전시장에서 서울 디저트페어라는 행사가 진행됐지만 일반적인 행사라 뉴스기사가 존재하지 않았다

→ 만약 뉴스 기사가 없다면 여러 사람들이 찾고 후기를 남길 수 있는 View탭의 글들을 크롤링 하도록 크롤러에 새로운 흐름을 추가