REPORT

방이(가) 표시된 사진

자동 생성된 설명

**[ Project Proposal ]**

**제 출 일 : 2021.09.14**

**강 의 명 : 빅데이터 입문**

**담당교수 : 최대진 교수님**

**학 부 : 자연과학대학**

**전 공 : 수학과**

**학 번 : 201600282**

**이 름 : 엄기산**

**1. Target Question**

- 당일 날씨와 영화 장르별 예매율의 변화 파악

**2. 수집할 데이터의 예상 규모**

- 영화 장르별 예매율 데이터 = 2008년 ~ 2017년도 일별 영화 예매인원 데이터

- 일별 날씨 데이터 = 2008년 ~ 2017년도 일별 날씨 ( 흐림, 맑음, 비 또는 눈 )

**3. 예상되는 어려움 및 해결방안**

|  |  |
| --- | --- |
| 예상되는 어려움 | 해결방안 |
| 코로나19로 인한 영화 예매 인원 하락으로 현재 인기 영화와 비인기 영화의 예매 율 격차가 심함 | 코로나19 이전의 데이터를 가지고 프로젝트 진행 |
| 장르별 데이터가 아닌 영화 별 예매인원 데이터만 존재 | 데이터 전처리를 통해 장르별 예매인원 데이터로 변환 |
| 현재 일별 기온 데이터와 일별 강수량 데이터 뿐임  흐림과 맑음의 구분 방법 | 흐린 날에는 맑은 날보다 기온차가 적게 발생함  강수일을 제외한 나머지 날의 월별 기온차를 파악하여 기준으로 잡고 구름양을 대략적으로 예상 |
| 마블과 해리포터 등 인기 영화가 개봉 할 때는 날씨와 상관 없이 비 인기 영화의 예매인원 감소됨 | 기간을 길게 잡아 많은 양의 데이터를 분석해서 해당 변수가 결과에 영향을 적게 주도록 함 |
| 우리나라 전체 지역을 분석하면 지역별로 날씨가 다를 수 있어 분석에 어려움이 있다.  예를 들어 서울은 맑지만 제주도는 비 오는 경우 | 우리나라 전체의 날씨가 아닌 유동인구가 가장 많은 서울만의 데이터를 분석한다. |

날씨 데이터 수집 URL = [기상자료개방포털 (kma.go.kr)](https://data.kma.go.kr/cmmn/main.do)

영화별 예매인원 데이터 수집 URL = [KOFIC 영화관 입장권 통합전산망 :: 일별 박스오피스 (kobis.or.kr)](https://www.kobis.or.kr/kobis/business/stat/boxs/findDailyBoxOfficeList.do)