**팀 프로젝트 후속 모임 보고서**

팀 번호: 6

팀원 목록: 심우석 이도원 우윤형 정유경 정의진

프로젝트 주제: 데이터로 알아보는 영화의 흥행 공식

후속 모임 일시: 25.04.26.

참석자: 심우석 이도원 우윤형 정유경 정의진

**발표내용 피드백**

1. 현장 질의응답 내용 (및 이에 대한 토의)
2. 영화 제작사의 관점에서는 수익률이 가장 중요한 지표로 작용할 수 있으므로, 평균 평점이나 영화 수가 흥행에 영향을 미친다는 점에 대한 설득력 있는 근거 제시가 필요하다.
3. ROI 그래프에서 비정상적으로 높은 값이 관측되었는데, 계산 과정에서 오류는 없었는지 우선적으로 검토할 필요가 있다. 만약 계산상 문제가 없다면, 해당 이상치는 분석에 왜곡을 줄 수 있으므로 제거 후 재시각화하는 것이 더 적절할 것이다.
4. 평균 평점 상위 25%의 영화들의 점수가 일반적으로 기대되는 점수 분포보다 다소 고르게 (4-10점대) 분포되어 있는 것은 모순적으로 보일 수 있다.
5. 수익률 상위 25% 장르별 작품 수를 히트맵으로 시각화할 경우, 장르별로 표본 수에 큰 차이가 존재할 수 있다.
6. 온라인 피드백 내용 (및 이에 대한 토의)
7. 장르별 영화 분포에서 드라마는 코미디, 스릴러, 로맨스 등 다른 장르들과 달리, 내부적으로 로맨틱 코미디나 스릴러 등 다양한 하위 장르로 세분화될 수 있다. 이러한 장르 분류 방식이 고려되어야 한다.
8. 같은 평점이어도 평가자 수가 적은 경우 해당 평점의 신뢰도가 낮다고 볼 수 있다. 따라서 평점과 관련한 분석을 수행할 때, 사용한 데이터의 평가자 수 분포가 비슷한지 확인할 필요가 있다.
9. 역사적으로 오래된 장르와 그렇지 않은 장르 사이에는 평균 평점의 차이가 존재할 가능성이 있다. 이에 따라 장르별 평균 평점 비교 시, 특정 연도를 기준으로 데이터를 제한했는지 설명이 필요하다.
10. 각 장르별로 몇 월에 흥행이 많이 되는지 분석해보면 좋을 것 같다.
11. 전체에서 평점 상위 25%를 선별해서 구했을 때, 장르의 특성상 드라마 장르의 영화 수가 많아서 드라마 장르가 압도적으로 높다는 결론 및 주의점을 드러냈다. 그렇다면 전체 장르별 개수에 대한 상위 25% 개수를 비율의 형태로 구하는 것이 더 적합한 비교였을 것 같다.
12. 기타 팀 내부 토의 및 피드백

발표 자료 구성에 일부 보완이 필요하다. 장르별 상위 25% 평점의 영화 분포 시각화 그래프가 전체 데이터를 기준으로 구성된 점은 수정이 요구된다. 보다 정확한 분석을 위해서는 전체 영화 시장의 규모를 함께 검토하는 것이 유의미한 접근이 될 수 있다. 변수 간의 적절한 균형을 바탕으로 타협점을 찾아내어 하나의 결론을 도출해낸다면, 분석 결과를 더욱 명료하게 정리할 수 있다.

**팀 프로젝트 개선 방향**

1. 수익률, 평균 평점, 영화 수의 관계 분석

- 평점과 수익률, 영화 수와 수익률 간의 관계 시각화 (정유경)

- 평균 평점과 영화 수가 영화 시장에서 가지는 의미 분석 (정의진)

> 평균 평점이 가지는 의미

장르별 평균 평점은 영화의 흥행 잠재력을 가늠하는 신뢰도 높은 척도로 작용하며, 사전 제작 및 투자 전략 수립에 유의미한 근거를 제공함.

사례: PeerJ Computer Science에 실린 연구에서는 장르별 인기와 평점 데이터를 분석하여 영화의 성공 가능성을 예측하는 방법을 제시함.​ 이 연구는 장르별 평균 평점과 수익률의 상관관계를 분석하여, 제작 전에 어떤 장르가 성공 가능성이 높은지를 평가하였고​ 이를 통해 제작사들은 제작 전에 장르별 평균 평점을 고려하여 투자 결정을 내릴 수 있었음.

출처: PeerJ Computer Science에 게재된 “Investigation of time series-based genre popularity features for box office success prediction” 논문 발췌

> 장르별 영화 수의 의미

장르별 영화 수는 시청자 선호도를 분석해 영화 제작 방향과 산업 전략에 결정적 영향을 미치는 지표임

사례: 넷플릭스(Netflix)는 시청자 데이터를 분석해 특정 장르의 선호도를 파악하고, 범죄·스릴러 장르가 높은 인기를 보이자 이를 기반으로 <나르코스>, <마인드헌터> 등의 오리지널 콘텐츠를 제작했고, 이 방식은 성공 확률을 높이는 데이터 기반 제작 전략으로 주목받음.

출처: Wired (2016), *How Netflix Reverse Engineered Hollywood*

- 평점과 영화 수와 수익률 사이의 균형점 제안

1. ROI 그래프 재시각화 (우윤형)

- ROI 분포 확인 결과, 이상치가 데이터 왜곡에 영향을 주는 것으로 확인

- 이상치 제거 기준은 0.1 ≤ ROI ≤ 20000으로 설정

- 이상치 제거 후 재시각화를 통해 ROI의 왜곡된 분포를 정제하여

1. 평점의 신뢰성과 연도별, 장르별 평균 평점 그래프 (이도원)

- 평점 개수와 평균 평점 사이의 관계 분석

- 연도별 평균 평점과 평점 개수

- 장르별 평균 평점

1. 표본 수를 고려한 평점 상위 25% 영화 장르 분석

- 전체 장르별 영화 수에 대한 평점 상위 25% 영화 수 비율 시각화 (이도원)

1. 영화 시장의 규모 분석

- 장르별 작품 수를 분석하여 시장의 크기 추정 (이도원)

- 월별 장르별 관객 수 분포 (심우석, 우윤형)