

<대사파 대본>

안녕하세요, 연구 결과 부분부터 발표를 진행할 생명과학과 조명재입니다. 거두절미하고 시작하겠습니다. 먼저, 연구문제 1번에 대한 연구 결과입니다. 앞서 말했듯, 저희는 총 209편의 영화 데이터로부터 추출한 총 122명의 영화 배우를 대상으로 영화 배우와, 영화의 흥행 지표간의 관계를 설명할 수 있는 “흥행 지수”라는 변수를 제작하였습니다. 지금부터 이 흥행 지수라는 변수와 영화의 흥행 지표로 대부분의 연구에서 활용하고 있는 “연도별, 배우별 전국 총 매출 평균”, “배우별, 연도별 관객 수 비율 평균”, “배우 출연 영화 전국 총 매출 총합”, “배우 출연 영화 관객 수 총합” 간의 관계성을 분석한 결과를 말씀드리겠습니다.

먼저, Pearson 상관관계 분석 결과입니다. 파이썬의 scipy라이브러리를 활용해 진행했구요, 분석 결과 흥행 지수와 관객 수 비율 평균, 관객 수 총합, 전국 총 매출 평균, 전국 총 매출 총합 4가지 변수에 대해 강한 양적 상관관계를 나타냈습니다. 배우의 흥행 지수라는 지표가 높을 수록 영화의 흥행 척도들이 높게 나왔다, 라는 사실을 확인할 수 있었습니다. 아래 그래프는 x축을 흥행 지수로, y축을 각 변수로 설정하여 그린 산점도 및 추세선 시각화 결과입니다. 네, 그렇구요,

그 다음은 종속변수를 흥행 지수로 설정하고, 독립변수를 각각 관객 수 비율 평균, 관객 수 총합, 전국 총 매출 평균, 전국 총 매출 총합으로 설정하여 단순회귀분석을 진행한 결과입니다. 분석 결과 모든 독립 변수에 대하여 p value가 모두 0.05보다 낮게 나와, 흥행 지수의 변화와 각 독립변수의 변화 사이에 연관성이 있음을 확인할 수 있었구요, Prop(F-statistics) 값이 0.05보다 낮아 연구 집단이 아닌 모집단에 대하여 모형이 유의미하다는 사실을 확인할 수 있었습니다. 단순회귀분석이라 Adjust R squared값을 굳이 쓸 필요는 없지만, 각각 39%, 90%, 44%, 88%의 설명력을 나타내는 것으로 확인 되었습니다. 이렇게 매우 높은 설명력을 나타내는 모델을 만든 것 같지만 아쉽게도 Durbin-Watson 값을 확인했을때, 관객 수 비율 평균과 전국 총 매출 평균 변수에선 Durbin Watson 값이 유효구간보다 한참 0에 가까운 값이 나타나, 변수들간의 독립성을 인정받지 못했습니다. Durbin Watson 검정은 잔차들간의 독립성을 확인하는 검정입니다. 관측치가 122, 설명 변수가 1일때 유효구간이 1.721과 2.279 사이로 설정되었고, 이 밖을 벗어나면 독립성을 인정받지 못합니다. 두 변수의 경우엔 저희가 고려하지 못했던 변수간의 상관관계가 존재했고, 따라서 두 변수에 대한 모델은 유효하지 않다고 판단하여 이어질 잔차 검정에서 제외했습니다.

방금전 DW값이 유효구간 내에 존재했던 두 모델에 대한 추가적인 잔차검정을 진행한 결과입니다. 두 모델 모두 비슷한 경향을 보였는데, 잔차 선형성 검정 결과에선 어느정도 선형성을 띄고 있으나 잔차가 매우 산발적으로 분포되어 있음을 확인할 수 있었고, 잔차 정규성 검정 결과에선 Shapiro test결과 모두 p-value가 0.05보다 낮게 나와 두 모형 모두 유의수준 5%에서 잔차의 정규성을 위반한다는 결과를 얻었습니다. 그리고 등분산성 검정으로 잔차들간의 분산을 확인해 본 결과 흥행 지수가 낮은 구간에서 잔차의 분산이 비례하는 경향을 보였습니다. 원인은 크게 2개로 볼 수 있는데, 흥행지수가 높은 데이터보다 흥행지수가 낮은 배우 데이터가 압도적으로 적은 분포를 나타내고 있고, 저희가 고려하지 못했던 변수들간의 분포 패턴이나 관계성이 일부 존재했기 때문이라고 생각합니다.

다음은 다중회귀 분석을 진행한 결과입니다. 독립변수를 다양한 변수들로 조합하면서 진행했는데요, 관객 수 비율 평균 독립변수와 전국 총 매출 독립변수가 추가되면 DW값이 너무 낮아져서 모두 유효하지 않은 결과가 나왔습니다. 그래서 독립변수를 관객 수 총합, 전국 매출 총합으로 설정한 다중회귀분석 모델만 유효하다는 결과가 나왔구요, 앞에서 확인했던 것 처럼 유효하다는 결과를 나타냈습니다. 설명력은 91%로 지금까지 확인했던 모델중에 가장 높은 설명력이 나왔구요, 저희가 만든 흥행 지수라는 지표는 관객 수 총합과 전국 총 매출 총합이라는 독립 변수와 가장 높은 연관성이 있는 것 같습니다. DW값도 유효구간 내에 있어 독립성 인정 되었구요, 다음으로 넘어 가서,

앞에서 실행한 방식 그대로 잔차 검정을 진행했습니다. 단순회귀분석 결과보다 잔차가 고르게 분포되어 있는 경향을 보였으나 앞에서 확인했던 회귀분석 모델이 가지고 있는 문제점들을 그대로 가지고 있었습니다. Shapiro 검정 결과도, 아쉽게도 크게 개선되지 못했구요. 원인도 공유하고 있다고 생각합니다. 이렇게 잔차검정 결과에서 유효한 결과를 얻지 못하면, 해결 방안으로 극단값들을 제거하고, 변수에 대한 수학적 계산을 추가적으로 실시하고, 예를들어 log를 씌운다던가 제곱을 취한다던가, 같은 방식으로 말이죠. 표본 수를 조절하는 방법이 있습니다. 아쉽게도 본 연구에선 끝까지 진행하지 못했습니다.

이런 아쉬움이 있지만 흥행 지수라는 지표는 충분히 흥행 지표와 높은 연관성이 있어보입니다. 그래서, 저희는 이 흥행 지수를 활용하여 장르별로 높은 흥행력을 가진 배우들 set을 만들어보았습니다. 액션 장르에서는 유해진, 김윤석, 하정우.. 저희가 익히 잘 알고있는 배우들이 등장하였구요, 범죄 장르의 경우엔 강동원, 하정우, 멜로/로맨스의 경우 장영남, 김향기, 유연석 등의 배우가, 사극에서는 최민식, 류승룡, 조진웅 등의 배우가, 코미디에선 오달수, 류승룡, 이하늬 등의 배우가, 드라마에선 오달수, 황정민, 송강호 등의 배우들을 확인할 수 있었습니다. 다양한 장르에서 모두 높은 흥행 지수를 나타내고 있는 배우들도 공통적으로 존재한다는 것을 확인할 수 있죠. 오달수, 황정민, 송강호, 하정우 배우들이 대표적입니다.

다음은 장르별 다양한 지수들을 분석한 결과입니다. 관객 수 비율, 스크린 점유율, 전국 총 매출 값, 출연지수들을 분석하였구요, 보시다시피 다른 장르에 비해 사극 장르는 출연 지수 평균이 가장 높고, 관객 수 비율도 높은 편이죠, 평균적으로 출연지수가 높다는 건데, 이는 기존에 흥행했던 영화에 많이 출연한 배우들을 주로 캐스팅 한다고 해석할 수 있습니다. 즉 검증된 배우들을 많이 차용하는 경향이 있다는 것이죠. 코미디의 경우 출연 지수 평균과 스크린 점유율은 상대적으로 낮는데 관객 수 비율이 높습니다. 출연 지수가 낮다는 것은 검증된 배우 보다는 새로운 배우들이 많이 출연하는 경향이 있다는 뜻으로 해석할 수 있습니다. 한마디로, 다른 장르에 비해 적은 투자로 큰 흥행 성적을 낸 케이스가 많다는 것이죠.

다음은 배우의 장르별 흥행 지수 표준편차를 분석한 결과입니다. 흥행 지수의 표준 편차가 높다는 뜻은 장르별로 흥행 지수의 평균 값이 많이 다르다는 뜻인데요, 그만큼 장르를 많이 탄다는 뜻으로 해석할 수 있습니다. 즉 흥행하는 장르가 명확하다는 것이죠, 대표적으로 최민식, 오달수, 류승룡과 같은 배우들이 이에 해당했으며 반대로 표준편차가 작은 배우들은 신하균, 수애, 유승호, 김우빈 등의 배우들로 다양한 장르에서 고른 흥행 지수를 나타내고 있다는 것을 확인할 수 있었습니다. 모든 장르에 대해 흥행 파워를 지닌다고 적혀있는데, 정확히 말해 장르를 크게 타지 않는 배우라는 뜻입니다.

그리고 표준 편차가 큰 배우들 5명에 대해 추가적인 분석을 진행한 결과 표준 편차가 큰 배우들이 장르별 흥행지수도 높다는 사실을 확인할 수 있었습니다. 배우들이 각 장르에 대해 거의 1,2,4위 등 상위권을 차지하고 있죠. 즉 각 장르에서 흥행지수가 큰 배우를 캐스팅 하는 것이 더 효과적이라는 결과를 도출해 낼 수 있습니다.

이제 연구문제 3, 특정 장르의 흥행 여부와 상영 시간간의 연관성 분석입니다. 흥행한 영화 데이터 셋이 있으니, 분기별로 빈도 분석을 진행한 것입니다. 드라마 같은 장르의 경우엔 분기에 상관없이 항상 흥행하는 특징을 보였고,

코미디, 범죄의 경우 1분기에 흥행하는 경향을 보였으며 이는 설 명절로 인한 가족 영화가 보통 이에 해당하기 때문이라 추측합니다.

2분기에는 장르별 빈도수 차이가 크게 두드러지지 않았고, 뒤에 언급하겠지만 흥행하는 영화 수 자체가 적습니다. 관람객 규모 자체가 적구요. 2분기는, 특정한 이유는 모르겠지만, 흥행한 영화 수가 상대적으로 적습니다.

3분기에는 액션, 사극 영화가 굉장히 높은 빈도로 나타났구요

4분기에는 모든 분기에서 죽을 쑤고 있었던, 멜로/로맨스 장르의 빈도수가 상대적으로 높게 나왔음을 확인할 수 있었습니다. 아마 크리스마스도 있고, 연말 시기도 있다 보니 연인끼리 로맨스 영화를 보러 오는 빈도가 높다보니 이때 멜로/로맨스 영화가 꽤 선전한게 아닌가 싶습니다. 결과적으로, 분명히 분기별로 흥행하는 장르가 다르고 이는 영화 개봉일자를 선정하는데 장르에 따른 고려가 필요하다는 것을 나타냅니다.

자, 지금까지 연구 결과 발표였습니다. 다음은 연구 결론입니다.

저희는 2008년부터 2021년까지 14년간 흥행한 한국영화 209편과, 영화에 출연한 배우들을 바탕으로 영화배우 122명을 선정하여 분석에 활용했고, 기존에 존재하는 영화 데이터를 활용하여 흥행 지수라는 배우의 흥행력을 추측할 수 있는 지표를 설정했습니다. 이 흥행 지수를 활용해서, 장르와 배우의 관계성을 분석했고, 장르별로 흥행한 배우들과, 흔히 말하는 장르를 크게 타는 배우와 크게 타지 않는 배우들을 확인해 보았고, 장르를 크게 타는 배우들이 보편적으로 각 장르에서 매우 높은 흥행지수를 가지고 있다는 사실을 확인해, 각 장르에서 흥행지수가 높은 배우를 캐스팅 하는 것이 효과적이라는 결론을 내렸습니다. 거기에 분기별로 흥행하는 장르가 분명 존재하고, 영화의 개봉일자를 선정할때 장르에 따른 고려가 필요하다는 결론을 내렸습니다.

기존 연구에서 잘 다루지 않았던 배우와 영화의 흥행지표간 관계성에 대해 다뤘고, 흥행지수라는 유의미한 변수를 제작했으며, 이를 활용해 흥행력을 보장하는 배우 풀을 제작하여 제시했다는 것이 저희 연구의 가장 큰 의의라고 볼 수 있을 것 같습니다. 저희가 분석한 내용 모두가 실제 영화 제작 과정에 객관적 근거로 활용될 여지가 있으며, 영화의 흥행 또한 예측할 수 있는 자료로 활용될 수 있음을 기대할 수 있습니다. 영화는, 사실 굉장히 큰 리스크가 있는 산업이고, 다른 사업들과는 다르게 오랜 시간과 큰 비용을 들여 개봉했을때 충분히 조명받지 못한다면

흥행에 쓴맛을 보는 경우가 많았습니다. 하이 리스크라는 거죠, 이런 하이리스크 사업에 실패 변인을 줄일 수 있는 객관적 연구 지표를 만든다는 것은 큰 의의가 있다고 생각합니다.

한계도 물론 많았습니다. 첫번째로, 분석 대상 배우의 수가 부족하고, 상당히 기형적인 분포를 보여서 흥행 지수를 모델링 할때 잔차 검증 단계에서 문제가 생겼었죠, 다양한 방법을 활용해서 문제를 해결할 수 있을 것으로 보였지만, 이 연구에선 시간상의 문제로 인해 끝까지 진행해 보지 못했습니다. 많이 아쉬웠던 부분이구요, 둘째로는, 영화 흥행에 영향을 미치는 수많은 요소들에 대한 다각적인 고려가 이뤄지지 못했다는 것입니다. 오로지 배우와 장르에만 집중한 것이 저희의 연구구요, 사실 영화의 흥행 요인은 매우 다양합니다. 영화의 각본, 연출, 스토리의 사회성, 정치성, 그리고 이를 모두 총괄하는 감독과 제작진의 역량, 그리고 배우들과 배급사의 파워까지 다양한 부분을 고려하고 실제로 이 사이에 있는 네트워크를 조금씩 분석하는 연구도 존재합니다. 하지만 본 연구에선 다이나믹하게 이뤄지지 못했습니다. 그리고 이 분석은 오로지 영화관에서 개봉했던 영화에 대해 진행한 결과이기 때문에, 최근 영화 소비 트렌드가 영화관에서 OTT플랫폼으로 넘어가고 있는 추세다 보니, 객관적인 흥행 지표에 OTT플랫폼을 통한 수익 또한 포함되어야 하는게 맞습니다. 하지만 OTT플랫폼에선 조회수라던지 매출이라던지 흥행 지표에 대한 데이터를 공개하지 않고 있기 때문에, 데이터 분석에 활용할 수 없었고, 조금 현재의 트렌드와는 괴리가 생길 가능성이 있다는게 저희 연구의 한계점이었습니다.

개인적으로는, 언젠가 많은 공부를 마치고 더 신뢰성 있는 결과를 낼 수 있는 많은 분석방법들을 제대로 활용할 수 있게 된다면 더 깊이 분석을 진행해 보고 싶다는 생각이 드는,, 흥미로운 주제였다고 생각합니다. 지금까지 긴 발표 들어주셔서 감사합니다. 3조였습니다.