발표시작하겠습니다. 안녕하세요. 에이콘아카데미 빅데이터과정의 다타 팀입니다. 저는 발표를 맡은 전대한이라고 합니다. 오늘 저희는 ~~라는 주제로 데이터 분석을 하겠습니다.

팀원 소개입니다. 각자 주로 맡은 역할은 있었지만, 하나의 주제, 하나의 분석에 서로 아이디어와 분석을 "공유"하여 결과를 만들었습니다. 특히 데이터 전처리에 권나은님, 남수민님, 이소민님이 힘써주셨고 나머지 분석에 저와 선우님 태현님이 힘을 실었어주셨습니다.

20년 이후 저희는 외출과 실내이용에 대한 불편함을 겪었습니다. 그로인해 영화관은 큰 타격을 입은 것 같은데요, 왼쪽 그래프를 보면 영화 매출액이 19년 2조 5천억원에서 20년 약 1조원으로 급감한 것을 확인할 수 있습니다. 그와는 상반되게 기사를 통해 OTT서비스에 대한 관심도는 같은 코로나 시기에 증가한 것을 볼 수 있습니다. OTT서비스가 성공하고 나아가는데 어떤 요인이 있었는지, 그리고 주 고객층과 앞으로의 서비스이용에 힘을 실어줄 요인은 무엇인지 알아보겠습니다.

저희는 크게 4가지 챕터로 분류를 해봤는데요, 먼저 전처리 과정을 보여드리고 저희가 분석한 내용을 설명해드리도록 하겠습니다. ott서비스와 영화관의 이용율, 주 이용층을 알아보겠습니다. 또한 코로나 시기에 코로나가 영화관 이용에 어떤 영향을 줬는지 파악하고, 영화관 이용율과 ott영화 이용률의 관계를 알아보겠습니다. 마지막으로 향후 ott서비스 증진을 위한 요인은 무엇인지 알아보겠습니다.

먼저 전처리 부분입니다. 저희는 미디어 통계포털 사이트와 통계청에서 ott관련 데이터를, 영화관 통합 전산망과 한국데이터 거래소에서 코로나 확진자수와 영화관 데이터를 가져올 수 있었습니다.

저희가 사용한 코로나데이터는 20년에서 21년까지 약 2년간의 데이터를 사용했으며, 기간동안의 확진자 수 및 사람들의 인식 변화를 알아보고자 했습니다. ott서비스 데이터는 하나의 final데이터셋을 만들기 보단, 분석에 필요한 각각의 데이터셋을 생성했는데요, 이를 통해 ott서비스의 전반적인 양상과 코로나 기간 중 ott서비스에 대한 정보를 얻을 수 있었습니다.

ott서비스 데이터는 정보통신정책연구원에서 실시한 한국 미디어 패널조사의 결과가 정제된 데이터였습니다. 예를들어 가장많이 이용하는 ott서비스의 경우 성별, 연령별, 소득별, 가구원수별, 지역별 등 구분이 이미 지어져있었고 각 컬럼은 연도별 및 사례수에 대한 비율로 이루어져 있습니다. 이를 통해 저희는 연령별 가장 많이 이용하는 ott서비스는 무엇인지 등의 정보를 확인할 수 있었습니다.

거의 모든 ott서비스 데이터셋이 이런식으로 성별, 연령별, 소득별 등등 구분이 지어져 있기 때문에 각각 필요한 데이터를 뽑아와서 적재적소에 필요한 분석만 할 수 있게 가져올 수 있었습니다.

보시는 것은 각 ott서비스 데이터들에서 뽑아온 컬럼들입니다. (덕분에 결측치가 없고, 따로 처리할만한 이상치도 없었기 때문에 큰 전처리가 필요하지 않았지만,) 추가로 "이용경험 여부나 유료결제 여부가 있다/없다"등으로 나뉘는 범주형 컬럼을 제외하고 각 데이터 셋마다 일정기준 비율로 너무 작은 컬럼일 경우 "기타"로 병합하여 진행했습니다. 또한 연도별 증감율에 따른 파생변수를 생성하여 다른 데이터들과 비교할 수 있게 진행하였습니다.

코로나 확진자와 영화관 관객수 데이터에서는 각 월별과 일별로 데이터셋을 만들었고, 전월대비 증감량으로 각각의 파생변수를 생성하여 상관관계(통계분석)을 위한 데이터 셋을 만들었습니다.

특히 wide로 된 코로나와 영화 관객수 데이터를 long형으로 reshape하여 일별 확진자수와 일별 관람객수의 상관분석을 위한 데이터 셋을 만들었습니다. 그리고 20년과 21년으로 데이터를 나누어 년도별 분석도 진행할 수 있었습니다.

또한 여가시간 활용 데이터에서, 영화관의 연도별 증감율을 알기위한 파생변수를 생성하였고, ott서비스중 영화시청증감율을 알기위한 파생변수를 생성하여, 상관분석을 위한 새로운 데이터셋을 구성하였습니다.

마지막으로 beautifulsoup, urlopen등의 툴을 이용하여 뉴스기사를 크롤링하였고 이를 wordcloud로 시각화 할 수 있었습니다.

----------------------------------------------------------------------------------------------------------------

다음은 데이터 분석부분입니다. ott이용율과 주 고객층 및 영화관 매출과 관객수에 대한 시각화 자료입니다.

최근 3개월 OTT서비스 이용경험을 토대로 만든 이용 증가율입니다. 연령별로 20년과 21년의 증가율이 모두 상승했는데요, 그 중 20년의 비율이 더 큰폭으로 오른것을 확인할 수 있습니다. 저희는 이것이 코로나가 시작된 20년에 사람들이 OTT를 찾게된것이 아닐까 하는 추측을 하였습니다.

그 추측은 OTT 앱 이용 횟수로도 알 수 있는데요, 17년부터 21년까지 OTT앱 이용횟수는 쭉 증가하였는데 그 중 20년도 부터 더 큰 폭으로 증가한 것을 확인할 수 있습니다.

OTT서비스의 주 사용자를 알기위해 성별, 소득별, 나이별로 분석을 해보겠습니다. 먼저 성별 그래프입니다. 각 남성과 여성이 약간의 차이를 보이고 있지만 성별이 직접적인 이용여부의 원인이라고 판단하기는 어려웠습니다. 비슷한 양상을 보이는 두 그래프를 통해 남녀모두 OTT이용율이 증가한것을 확인할 수 있었습니다.

연령별입니다. 전체적으로 20년의 유료결제비율이 19년의 유료결제 비율보다 높았는데요, 특히 20~39세의 유료결제비율이 전체 50%로 가장 높았습니다. 여기서도 코로나가 시작된 시점인 2020년에 유료결제율이 높아진 것을 확인할 수 있었습니다.

소득별입니다. 월 평균 200만원 이상의 소득을 내는 사람들의 85%는 이용경험이 있는것으로 보이는데요, 눈에 띄는 점은 소득없음의 사람들의 80%또한 이용경험이 있다는것입니다. 이는 미성년자, 대학생 등의 사람들 또한 OTT서비스를 많이 이용한다는 의미로 해석됩니다.

다음은 영화관매출과 관객수입니다. 막대그래프와 꺾은선 그래프는, 각각 관객수와 매출액을, 시기별로 기록한 그래프입니다. 둘은 같은 양상을 보이는데요, 20년도 초부터 매출과 관객수가 급격하게 하락한 것을 확인할 수 있습니다. 이는 코로나로 인해 외출과 야외활동을 자제하는 경향이 영화관에 영향을 준 것으로 해석됩니다.

그래서 빨간색 화살표로 되어있는 실제 여가시간 영화관 이용률을 보면, 19년도에서 21년도 여가시간에 영화관을 이용하는 비율이 9.5%에서 3%로 감소한것을 확인할 수 있고 21년의 모든 연령대에서 그 비율이 감소한 것을 확인할 수 있습니다.

이는 21년도는 19년도의 약 1/5 관객수만 기록한 것으로 더욱 뒷받침 됩니다.

그에 반해 OTT서비스중 영화시청을 주된 이용으로 한 사람들의 비율은 늘어났는데요, 같은 시기 OTT를 통한 예능/오락시청률은 감소하는것에 반해, OTT영화시청율은 상승하는 것을 확인할 수 있습니다.

방금까지의 시각화 자료들을 통해 저희는 OTT이용률에 20~39세의 결제율이 가장 높았다는 점과 성별, 소득별 구분은 큰 의미가 없다는 것을 확인했습니다. 또한 코로나시대에 OTT서비스가 증가하였고 그 중 OTT로 영화를 보는 사람들이 늘어났다는 점과, 영화관 이용률은 코로나 시기 감소한다는 것을 확인했습니다. 저희는 이러한 분석을 토대로 코로나와 영화관, 그리고 영화관과 OTT이용에 대한 상관관계를 분석해보았습니다.

----------------------------------------------------------------------------------------------------------------

이제 2번째 데이터 분석, 코로나와 영화관, 그리고 영화관과 OTT서비스의 관계에 대해 분석해 보겠습니다.

저희는 앞선 시각화 자료들에서 영화관이용에 가장 큰 영향을 주는것이 코로나다 라는 가정을 해봤는데요 그 연관성을 알아보겠습니다. 저희는 먼저 실제 코로나의 확진세가 커지면, 영화관 관객수는 감소할 것이라는 추측으로 진행하였습니다.

이것은 월별로 구분한 확진자"증감율"과 관객수"증감률"입니다.

확진자가 증가하면 관객수는 감소하는 양상을 보이고 있는데요, 표시된 점선과 같이 같은 양상을 보이는 월도 있었습니다. 저희는 혹시 확진자수의 증감이 사람들의 영화관 이용에 영향을 준다면, 지난달의 확진자수가 이번달의 영화관이용에 영향을 주지 않을까 생각해봤습니다.

이 그래프는 코로나 확진자 수 증감이 "다음달" 영화관 관객수에 영향을 미치는지 확인한 그래프입니다. 몇개의 월을 제외하고 대부분의 선에서 코로나확진자와 영화관객수가 상반되게 움직이는 것을 확인할 수 있습니다. 조금 더 자세하게 상관분석을 통해 알아보겠습니다.

20년부터 21년말까지 월별로 확진자수 증감률과, 관람객수 증감률의 상관관계를 확인한 결과입니다. 확인한 결과 약 6%의 유의수준에서 -0.4라는 상관관계가 보였습니다. 확진자수가 증가하면 관람객수는 감소한다고 결론을 지었는데요, 파워검정으로 필요한 표본수를 계산해본 결과 약 198개의 표본이 필요하다는 것을 알 수 있었습니다. 그래서 저희는 월별이 아닌 일별로 자료를 가져와봤습니다.

코로나가 시작된 20년 3월부터 21년 12월 30일까지의 자료들을 월별이 아닌 일별로 가져왔습니다. 여기서도 오늘의 확진자수가 내일의 영화관 관객수의 영향을 줄 것이다는 가정에서 분석을 진행하였습니다. 보시면 저희의 예상과는 반대로 약 0.35의 상관계수로 코로나 확진자수와 관객수는 양의 상관을 보였습니다. 아마도 코로나가 심했던 20년과 비교했을때 21년은 조금 회복기에 들어섰기 때문에, "년도"가 이러한 결과에 원인이지 않을까 하고 년도별로 분석을 나눠봤습니다.

본 그래프는 2020년 코로나 확진자와 영화관 관객수의 상관분석을 표현한 그래프입니다. 코로나 확진자가 늘어나면서 영화관 관객수는 줄어들었다, 즉 이 둘은 음의 상관관계가 있다 라는 결론을 도출해 낼 수 있었습니다. 당시는 코로나가 시작된 첫해였기 때문에 사람들이 코로나에 많이 민감했을 거라는 추측입니다.

이 그래프는 2021년 그래프입니다. 20년과는 반대로 21년은 오히려 확진자 수는 늘어나더라도, 영화관 관객수또한 늘어난 양의 관계를 확인할 수 있습니다. 저희는 21년에 특히 양의상관관계를 보인이유를 사람들의 코로나 인식에 대한 부재라고 생각을 했는데요, 그 이유는 다음과 같습니다.

20년 12월과 21년 4월은 보시는 바와 같이 코로나 확진자 수가 비슷했던 월입니다. 그럼에도 영화관 관객수는 약 2배정도 차이가 있었는데요, 이 두월을 비교군으로 하여 실제 코로나 인식을 알기 위해 코로나라는 단어로 뉴스를 검색을 했을때 나오는 단어들의 빈출을 보고 어떤 키워드가 있었는지 확인해봤습니다.

두 월 모두 코로나, 시대, 치료, 등의 공통된 키워드 말고 차이가 보인것이 있었는데요, 20년은 위기, 극복, 치료, 백신 등의 키워드가 있었다면 21년은 축제, 상생, 회복, 활동 등의 키워드가 다수 검출된 것을 확인할 수 있었습니다.

이것은 한국리서치에서 실시한 "코로나 방역과 경제활성화에 대한 사람들의 인식"을 조사한 설문조사 입니다. 같은 월인 20년 12월은 약 12%의 응답자가 경제회복과 활성화는 우선시해야 한다고 했지만, 21년 4월에는 24%로 약 2배의 응답자가 경제회복과 활성화를 우선시 해야한다고 했습니다.

이를 통해 저희는 앞으로도 코로나 확진세가 영화관 영화관람에 큰 영향을 주지 않을것. 이라는 결론을 내릴 수 있었습니다.

다음은 영화관 이용증감률이 OTT영화시청증감률과 관련이 있는지를 알아봤습니다. 저희는 영화관 이용률이 감소할수록, OTT영화 시청률은 증가할 것이라는 추측으로 진행하였습니다.

먼저 상관분석을 해보니 저희의 예상과는 다르게 약 0.6의 양의 관계가 나왔습니다. 그래프로 확인하면 빨간색 원이 혼자 떨어져 있는 것이 보입니다. 이 점이 영향이 큰 것으로 보이는데요, 저 점은 70대의 OTT영화 이용율과 영화관 이용율입니다. 70대를 제외하고 확인해보겠습니다.

70대를 제외해본 데이터 셋에서는 저희 예상과 동일하게 -0.67의 음의 관계가 나왔습니다. 저희가 고민했던것은, 이 데이터들중 전체, 성별, 연령별에서 중복되어 카운팅된 데이터가 존재한다는 것이었습니다.

그래서 연령별데이터만 가져와 분석을 시도하려 했지만 아까 제외한 만 70세 이상 뿐만 아니라10대미만의 OTT이용증가율도 다른 연령에 비해 현저히 낮았기 때문에 두 나이대를 제외해 보기로 했습니다. 또한 중복될 수 있는 성별, 전체 데이터 또한 제외해보고 분석을 진행했습니다.

저희가 처음 예상한 영화관 이용증감률과 OTT영화시청률은 음의 상관을 보였으며 파워검정을 통해 2개의 데이터로 확인가능하다는 것을 얻었지만 0.45의 높은 유의 수준으로 신뢰성이 부족한 분석이라는 것을 확인했습니다.

이번 상관분석을 통해서 저희는 2가지의 결론을 얻었습니다. 첫번째, 코로나와 영화관 이용은 약한 음의 상관관계를 가지지만, 향후 그 관계가 약해질 것이라는 것과 영화관 이용증감률과 OTT영화시청증감률은 음의 상관관계가 있었지만, 낮은 설명력과 높은 유의사항으로 신뢰성이 부족하다는 것입니다.

이 두가지 사실로 저희는 향후 코로나의 확진세와 무관하게 영화관 이용률은 증가할 것이며, OTT이용률 영화관이용과는 별개로 독자적인 마케팅 전략을 찾아야 한다는 것입니다.

----------------------------------------------------------------------------------------------------------------

그렇다면 향후 OTT서비스 이용증진을 위해선 어떤 전략을 수립해야 할까요? 이번 CONCLUSION파트에서 이야기 해보겠습니다.

저희는 방금의 분석1과 2를통해서, OTT서비스가 코로나라는 시대에 영향을 받지 않고 영화관이용률과는 무관하다고 결론을 지었습니다, 그렇다면 이제 코로나와 영화관 이용과는 연관없는 OTT만의 장점을 파악해야 했는데요, 저희는 아마 가격의 메리트가 있지 않을까 생각했습니다.

[클릭]이것은 현재 22년12월기준 가장많이 소비되는 OTT3개와 영화관 세곳의 가장 기본적인 가격들 입니다. 영화관은 OTT서비스보다 금액이 비싸면서 1회성이지만, OTT서비스는 한 번 결제후 한달동안 원하는 만큼 사용할 수 있으며 영화관보다 저렴하다는 장점이 있었습니다.

OTT서비스 이용빈도를 보시면 19년도에는 1주일에 1~2회의 사용빈도가 제일 많이 보이고있지만,

21년도에는 하루에도 여러 번의 비중이 가장 많은 것을 확인할 수 있습니다. 즉 소비자 대부분은 더 저렴한 가격으로 1회성이 아닌 반복해서, 자주사용 할 수 있는 OTT서비스의 장점을 적극 활용할 수 있습니다.

보시는 기사들은 영화관 가격인상에 대한 기사들의 헤드라인입니다. 영화관 가격인상에 대한 소비자들의 부정적인 시각이 기사에 고스란히 담겨있는데요, "그냥 넷플릭스를 보겠다.", "티켓값 안내리냐?등의 반응을 엿볼수 있습니다. 소비자들은 가격인상에 큰 반응을 보이는 거 같습니다.

이 그래프를 보시면 영화관 티켓값이 인상했던 5개의 시기가운데 2018년 5월의 시기를 제외한 나머지 시기에 모두 관객수가 하락한 것을 확인할 수 있습니다. 이는 소비자들이 영화관을 꺼려한 요인으로 요금이 될 수 있다는 것을 의미합니다.

이것은 해외의 자료이며 OTT서비스를 가입하지 않는 이유에 대한 설문조사입니다. 5개국 대부분의 응답자가 OTT서비스의 이용요금이 비싸다는 이유로 가입을 망설여 하는것을 볼 수 있었습니다.

넷플릭스와 유튜브 프리미엄등, OTT서비스들도 최근에 약간의 가격인상을 실시했는데요, 앞서 영화관객수의 감소와 OTT미가입이유가 모두 "가격"이 요인이 되었던 만큼 향후 OTT서비스는 가격선정에 신중해야 할 것으로 보입니다.

보시는 자료는 21년 OTT서비스 소비자들의 평균 지출액입니다. 약 80% 이상의 소비자들이 5천원~2만원미만의 금액을 지불하고있습니다. 2만원 이상의 그액대로 올라간 소비자의 비율이 한자리수로 훅 줄어든 것을 확인할 수 있는데요, 따라서 OTT서비스는 최대 2만원 미만의 이용요금을 제시해야한다고 결론을 지었습니다.

저희의 최종 결론입니다. 향후 OTT서비스 이용률을 증진하기 위해서는 코로나와, 영화관 이용률과는 별개로 20~39세가 좋아할만한, 하루에도 여러번 시청할 수 있는 콘텐츠의, 최대 2만원 미만의 금액대를 제시할 수 있어야 합니다.

저희 발표에서 아쉬웠던 점들인데요, OTT데이터는 정제되어 있어서 이상치, 결측치 처리가 필요가 없었다는 점, 표본수가 적었다는점, 분석이 심플했다는점, 그리고 상관분석의 설명력이 낮았다는 점입니다. 결국 RAW 데이터가 아니라서 더 넓고 추가적인 분석을 하지 못했던 것이 아쉬웠습니다.

함꼐 고생한 팀원들 너무 감사드리고 발표들어주셔서 감사합니다. 발표를 마칩니다. 편하게 질문해주세요.