

안녕하세요. 섹션1 프로젝트 발표를 시작하도록 하겠습니다.

바로 본론으로 들어가서

우리가 달성하고자 하는 목표는 "다음 분기에 어떤 게임을 설계해야 할까" 입니다.

이 목표를 바탕으로 데이터터를 살펴보고자 합니다.

데이터터는 다음과 같이 구성되어 있구요.

~

또한 데이터터의 정보는 다음과 같습니다.

~

데이터터를 확인해본 결과 결측값이 존재하며, 그 수가 미미하다 생각하여 대체를 하는 것이 아닌 삭제를 진행구요

우리의 목표를 달성하기 위해 '데이터터'에서 파악해야 할 포인트는 '다음분기', '어떤 게임' 할 수 있습니다.

'다음분기'라는 포인트의 정보를 담고 있는 영역이 year라는 영역이기에, 어떤 정보를 담고 있는지 파악했고, 잘못 입력된 정보가 있기에 올바르게 고쳐주는 작업을 진행했습니다.

다음으로 '어떤 게임'을 설계함에 있어서 그 근거가 되는 정보는 장르가 될 수도 있고, 플랫폼이 될 수도 있지만 결국, 그 게임이 얼마나 흥행을 했는지 알 수 있는 지표 즉 '출고량'이라 할 수 있습니다. 하지만 지금 데이터터에서 출고량이 따로 수로 표현되어 있지 않고, 문자형식으로 표현되어 있기에 수로 변환해 주는 작업을 걸었습니다.

따라서 우리가 사용하게 될 자료는 다음과 같고, 필요없는 정보를 지운 데이터터는 다음과 같습니다. 앞서 살펴 봤던 자료에서 분석할 수 있도록 변환만 된 자료라고 보면 될거 같습니다.

그렇다면 이 데이터터가 우리에게 어떤 정보를 보여주는지 한눈에 알아볼 수 있도록 시각화된 자료를 보여드리도록 하겠습니다.

가장 먼저 장르별에 따른 출고량 그래프입니다.

그래프는 다음과 같습니다. 액션류의 게임이 가장 인기가 좋으며 그 뒤로 슈팅게임, 스포츠게임, 롤플레이팅 게임, 레이싱 게임 순서로 볼 수 있고, 퍼즐게임과 어드벤처 게임이 출고량이 제일 적음을 확인할 수 있습니다.

다음으로 플랫폼별-출고량 그래프입니다.

xbox360에서 가장 출고량이 높았고 다음으로 플레이스테이션, 닌텐도64며 이외에도 꽤나 많은 플랫폼에서 출고량을 보이고 있음을 확인이 가능합니다.

그렇다면 각 플랫폼별 장르의 출고량은 어떨지 확인해보고자 합니다.

가장 많은 출고량을 보여줬던 xbox360에선 액션과 슈팅게임이 많은 출고량을 보이고 있으며, 닌텐도 64의 경우 레이싱 게임이 많은 출고량을 보여줍니다. 또한 플레이스테이션의 경우 각 시리즈 별로 출

고량이 다르게 보여지는데, 3같은 경우 액션게임이 2같은 경우 스포츠 게임이, 1같은 경우 롤플레이팅과 레이싱 게임이 많은 출고량을 보입니다.

또한 닌텐도 위유나 엑스박스원 같이 극단적으로 한두개의 장르에만 출고량이 보이는 것도 확인이 가능합니다.

그렇다면 지역별-출고량은 어떤 형태를 보이는지 확인해보고자 합니다.

북미 지역의 경우 대부분 게임에서 많은 두각을 보이며, 액션, 슈팅, 레이싱, 스포츠 게임을 선호합니다.

일본 지역의 경우 롤플레이팅 게임에서 확실한 두각을 드러내고 있으며

유럽 지역의 경우 플랫폼 게임에서 두각을 드러내고 있습니다.

다음으로 연도별-출고량은 어떤 형태를 보이는지 확인해보고자 합니다.

10년을 주기로 묶어서 그래프로 나타낸 결과는 다음과 같습니다.

액션, 슈팅게임의 경우 시간이 흐를수록 출고량이 늘어나는 반면, 레이싱 게임의 경우는 시간이 지날수록 출고량이 떨어지는 것을 확인이 가능하고, 2000년대 초반에는 스포츠, 롤플레이팅, 플랫폼 게임등이 두각을 드러냈다는 점 또한 확인이 가능합니다.

덧붙여서 액션게임이 두각을 드러내는데에 가장 큰 영향을 끼친 것은 아무래도 xbox360이라는 플랫폼의 등장이 아닐까 라는 추측또한 가능합니다.

마지막으로 간단하게 퍼블리셔-출고량 그래프를 확인해보고자 합니다.

모든 퍼블리셔가 아닌 출고량 기준으로 상위 약 25%에서 / 50번 이상 존재하는 퍼블리셔 그래프는 다음과 같습니다.

그래프를 통해 확인한 점들이 과연 '통계학적'으로도 맞는지 가설을 세운뒤 확인해보고자 합니다.

가설검정은 분산분석을 사용할 예정이며 분산분석을 간단하게 설명하자면 분석하고자 하는 대상 혹은 대상들이 갖는 평균값이 같은지 비교하는 방법입니다.

우리가 앞서 봤던 그래프의 차이가 실제로 유의미한 차이인지에 대해 확인하고자 하는 것입니다.

그 전에 자료의 간격이 너무 크다는 것을 그래프를 통해 확인이 가능합니다. 이렇게 되면 유의미한 의미 도출을 컴퓨터가 하기 어렵기에, 따라서 자료를 오밀조밀하게 모으는 과정이 필요하고, 이를 스케일링, 혹은 자료변환이라 할 수 있습니다. 그 중 정규화 스케일러를 사용하는 방법을 진행했고, 이제 본격적으로 가설검정을 진행하려 합니다.

가설검정은 그래프를 바탕으로 진행했습니다.

간단하게 결과창에 대해 말하자면 여기 값이 0.05보다 작으면 우리가 도출하고자 했던 바와같이 차이가 있다고 결론을 지을 수가 있습니다. 차이가 있다는 것은 순서가 있다는 것을 의미하며 그래프에서 살펴본 것이 짐작이 아닌 확식이 되는것이죠

0.05보다 작기에 차이가 있다 라고 결론이 이어졌고, 이 또한 역시 차이가 존재하며

두 요소가 동시에 영향을 끼치며 차이가 난다고 결론 지을수 있습니다.

결과적으로 우리는 이런 결론을 지을 수 있을겁니다.