

<section1 project 발표 스크립트>

안녕하세요 AI부트캠프 17기 송지연입니다. AI부트캠프 과정 중 Section1 project발표를 시작하겠습니다.

데이터 전처리 후에 다음 네가지 분석을 진행하겠습니다.

먼저 데이터셋의 컬럼들을 살펴보겠습니다. 데이터셋에는 게임의 이름, 게임이 지원되는 플랫폼의 이름, 게임이 출시된 연도, 게임의 장르, 게임을 배급한 회사, 각 지역에서의 출고량에 대한 정보가 있습니다.

이제 데이터 전처리를 시작해보겠습니다. 먼저 결측치가 존재하면 분석을 할 때 오류가 생기거나 부정확한 결과가 나올 수 있기 때문에 결측치가 있는 데이터는 제거했습니다.

전처리 하기 전의 Year column은 소수 첫째 자리까지 나와 있는 float 형태인데, 연도는 2022.0년이라는 것이 존재할 수 없고 2022년 형태여야 하므로 데이터 타입 변환을 했습니다. 그리고 값들을 살펴보면 1900보다 작은 값들이 있는데 이 값들은 이상치로 보여서 이상치를 삭제했습니다.

출고량 column에 대한 전처리를 하겠습니다. 지역별 출고량에서 기본단위는 1M인데 단위가 K로 표현된 데이터들이 있습니다. 이런 값들의 단위를 M로 통일해주고, 출고량 값은 숫자기 때문에 데이터 타입을 문자열 계산이 가능한 숫자로 변환했습니다.

이제 전처리 완료된 데이터셋을 살펴보면 각 column정보에 걸맞는 데이터 타입으로 변환이 완료되었고, 결측치도 없는 것을 확인했습니다.

먼저 지역에 따라 선호하는 게임 장르가 다를지에 대한 분석을 해보겠습니다. 데이터 셋을 살펴보면 각 게임별로 북미, 유럽, 일본, 기타 지역의 출고량이 존재하는데 네 가지 출고량 중 한 개만 있는 게임도 있지만, 출고량이 북미지역, 유럽지역처럼 두 개 이상인 게임이 있습니다. 이렇게 출고량이 다양한 게임의 지역을 설정하는 것이 필요해 보여서 각 게임의 네 가지 출고량 중 출고량이 가장 높은 지역을 해당 게임의 지역으로 설정했습니다. 그 후 각 지역별로 선호하는 장르에 대한 시각화를 진행했더니 북미지역과 유럽지역, 기타 지역에서 가장 선호하는 장르는 Action으로 나타났고, 일본 지역에서 가장 선호하는 장르는 Role-Playing으로 나타났습니다. 따라서 지역별로 선호하는 게임의 장르가 다르다는 사실을 확인했습니다. 특히 점으로, 북미, 유럽, 일본 지역에서 나타난 장르의 종류는 12가지였는데, 기타지역에서는 2가지 게임 장르 Puzzle, Strategy가 제외되어 있었습니다.

다음으로 연도별 게임의 트렌드가 있을지에 대한 분석을 진행하겠습니다. 여기서 트렌드는 연도별로 인기있는 게임의 장르를 이용해 파악했습니다. 1980년부터 2020년까지의 연도별 데이터를 살펴보면 2002년부터 2011년까지 기간동안 전체적으로 출시된 게임의 수가 급격히 상승했습니다. 연도별 출시된 게임 장르의 수를 시각화해서 관찰해보면, 1990년대 중후반~2000년대 초반에는 Sports장르의 게임이 가장 많이 출시되었고, 2000년대 초반부터 Action 장르의 출시가 급증하다가 2003년 이후로는 Sports 장르를 넘어서서 가장 많이 출시된 것을 확인할 수 있습니다. 그리고 Misc 장르는 2000년대 중반~2010년대 초반 기간동안 급격히 많이 출시되었습니다. 이런 특징들을 종합해보면 연도별 게임의 트렌드가 존재한다는 것을 확인할 수 있습니다.

다음으로 인기가 많은 게임에 대한 분석 및 시각화 프로세스를 진행하겠습니다. 인기가 많다

는 것은 모든 지역에서의 출고량 합계를 기준으로 정의했고, 합계 출고량이 높은 게임의 인기가 많다고 생각했습니다. 그래프를 통해 합계 출고량의 데이터 분포를 살펴보면, 12 이후로 데이터의 빈도가 급격히 낮아지므로 합계 출고량이 12이상인 게임을 인기가 많은 게임이라고 설정하겠습니다. 인기가 많은 게임들의 장르를 살펴보면 Shooter 장르가 가장 높은 비율을 차지하고 있는 것을 확인할 수 있습니다. 그리고 인기가 많은 게임 중 61.9%가 'Nintendo'회사에서 출시한 게임이고, Nintendo 회사에서 출시된 게임만 살펴보면 출고량이 가장 높은 지역으로 북미지역이 84.62%를 차지하고 있는 것을 확인 할 수 있습니다.

마지막으로 다음 분기에 어떤 게임을 설계해야 하는지에 대한 분석을 진행하겠습니다. 다음 분기는 2016년 이후 10년으로 설정했습니다. 첫 번째 그래프를 통해 2006년~2015년 데이터의 게임 플랫폼의 종류는 PS시리즈가 많은 비율을 차지하고 있는 것을 확인했고, 2006년~2015년 데이터 중 플랫폼이 PS인 게임은 일본지역과 북미지역에서의 출고량이 다른 지역보다 많았습니다. 그리고 마지막 그래프를 통해 2016년~2020년 데이터의 게임 플랫폼에서 PS시리즈가 가장 높은 비율을 차지하고 있는 것을 볼 수 있습니다. 2006~2015년의 결과를 기반으로 다음 분기에 대한 분석을 해보자면, PS의 마케팅을 일본지역과 북미지역 위주로 진행하면 좋은 영향이 있을 것입니다.

그리고 2020년 이후 코로나19의 여파와 레트로의 열풍으로 인해 nintendo의 재유행이 일어날 것으로 보이고, 닌텐도 게임은 앞선 분석의 결과를 보면 북미지역에서의 인기가 좋았기 때문에 북미지역에 맞는 마케팅을 진행한다면 판매량에 좋은 영향을 줄 것으로 보입니다.

이상으로 발표를 마치겠습니다. 감사합니다.