Team:Manipogo

Group Members:

Richard Lee	rlee106@gmail.com
Deannia Lucas	deannialucas80@gmail.com
Caglar Cinar	cag.cin@gmail.com
Hershel Don	hdon@mail.depaul.edu
Hima Spandana Barla	spandanabarla26@gmail.com

Topic: Video Games

Dataset : Video Games Sales

"If we have data, let's look at data. If all we have are opinions let's look at mine" - Jim Barksdale. Fortnite, League of Legends, Tetris, and Pong, what are these you might ask? You guessed it, video games. But what makes these video games so popular and attractive to audiences of all ages? There are a large variety of video games with many different variables such as genre, a rating score, publisher, what platform the game might be on, and so on that define what kind of video game it may be. In this study we will explore the many different factors in which video game sales are affected and what makes some video games more successful than others.

The response variable we will use in this project will be the **sales (Global sales, and potentially the sales by regions: NA, EU, etc.)** We think that this variable is important to predict because it directly affects the revenue that the company gets from the game and it indicates how popular that game is.

The explanatory variables we think that the response variable is dependent of are:

- Platform (Platform of the games release - Categorical)
  - 31 unique values.
  - Some platforms are more popular than others.
- Year of Release (Year of the game's release - Numerical)
  - Between 1976 and 2017 - 8 missing values.
  - We believe that the newer games will sell more
- Genre (Genre of the game - Categorical)
  - 12 unique values.
  - Some genres simply appeal to a bigger player base
- Critic Score (Aggregate score compiled by Metacritic staff - Numerical)
  - Between 13 and 98 - 9080 missing values
  - What the critics think of this game
- Critic Count (The number of critics used in coming up with the critic score - Numerical)
  - Between 3 and 113 - 9080 missing values
  - How many critics rated this game
- User Score (Score by Metacritic's subscribers - Numerical)
  - Between 0 and 9.7 - 9618 missing values
  - What the users think of this game
- User Count (Number of users who gave the user score - Numerical)
  - Between 4 and 10766 - 9618 missing values

    - How many users rated this game
- Rating (The ESRB ratings - Categorical)
    - 8 unique values - 7164 missing values

The Data was acquired from these three sources; video games sales, and ratings was gathered from https://www.kaggle.com/kendallgillies/video-game-sales-and-ratings, Here we found a list of different video games https://www.giantbomb.com/games/ and https://thegamesdb.net/ had different developers and various platforms.

There are entries for 17,416 video games released between 1976 and 2017 in the dataset we selected. There are 14 explanatory variables, of which 4 are categorical and 10 are numerical. The categorical variables for game entries are game platform, genre, and publisher, and ESRB rating. Genre and publisher are nominal, while platform and ESRB rating are ordinal, since the game platforms can be sorted by release order and ESRB ratings can be sorted by intended audience / maturity. Out of 10 numerical variables, 5 are continuous and 5 are discrete.The continuous numerical variables are (in millions) sales in North America, Europe, Japan, all other regions, and total global sales. The discrete numerical variables in our dataset are year of release, user and critic scores, and the counts of critic and average user scores. The user and critic scores are scaled differently, yet they can both be scaled to integer values 0 through 100, so we considered them discrete numerical variables.

The major drawback of the dataset is incomplete entries for some games. Incomplete entries result from games that were released without or before the introduction of ESRB ratings, before ratings from users were collected online, or even before there were video game critic scores. Another possible drawback is that only video games that sold over 100,000 copies were included in the data, but this likely narrows the games in consideration substantially and only analyzes games that had some degree of commercial success, which is what we are interested in.

To conclude, we collaboratively worked and chose a dataset and will analyse the response variable- sales to predict the revenue of the game of a particular company based on the explanatory variables mentioned above. Complete analysis will be performed using R to analyse the output variables sales, build regression models to build using quantitative and qualitative data and visualize them.