Video Games Sales Proposal

Brenda Zhu

**The Problem**:

The videogame industry has shown a great need for accurate sales data. This project aims to explore how a study of a variety of variables can help predict the videogame sales. Our hypothesis is that the sales are related to publisher, platform, genre, critic score or all of the above. This dataset has more than 10,000 games giving meaningful information can be used to create a model that is able to predict how well a game is selling. If this model can accurately predict game sales in 2016, such a model could prove useful to forecast sales in future years.

**The Purpose of the Project:**

As the videogame industry has grown, data is important to the videogame industry, its investors, publishers, and the online community of videogame fans. For example, a game company will want to keep track of the sales of its competitors’ products and investors will want to keep track of sales when planning investments. A game publisher’s marketing department may use sales data to track the effects of an advertising campaign, or to study which demographics a game is popular with so that it can use its budget effectively. Additionally, if a game is a surprise success in its first weeks, its publisher may use this information to decide whether or not to order more copies from the manufacturer in order to meet consumer demand. Fans often become loyal to certain manufacturers/developers.

**The Data**:

Original dataset was from Gregory Smith's web scrape of VGChartz Video Games Sales, and this dataset was build based on the number of variables with another web scrape from Metacritic. Unfortunately, there are missing observations as Metacritic only covers a subset of the platforms. Also, a game may not have all the observations of the additional variables discussed below. (Complete cases are 6,900)

We have the following variables in the dataset:

* **Name**
* **Platform**
* **Year\_of\_Release**
* **Genre**
* **Publisher**
* **NA\_Sales**
* **EU\_Sales**
* **JP\_Sales**
* **Other\_Sales**
* **Global\_Sales**

Other variables:

* **Critic\_score -** Aggregate score compiled by Metacritic staff
* **Critic\_count -** The number of critics used in coming up with the Critic\_score
* **User\_score -** Score by Metacritic's subscribers
* **User\_count -** Number of users who gave the user\_score
* **Developer -** Party responsible for creating the game
* **Rating -** The [ESRB](https://www.esrb.org/) ratings (ESRB: Entertainment Software Rating Board)

**Hypothesis**:

1. Does Critic score and video-game sales have any correlation, how significant is the impact?
2. Based on the variables most correlated to hits, what 2019 games can still become hits?

**Risks or limitations:**

There’re some missing values in the datasets for critic score, which directly related to our hypothesis. How to wisely handle the missing values could be a challenge for this study.

The data is available on Kaggle(<https://www.kaggle.com/rush4ratio/video-game-sales-with-ratings>),

we are making the assumption that the available sales data is accurate.