

### Milestone 3

#### 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	<a href="mailto:spandanabarla26@gmail.com">spandanabarla26@gmail.com</a>

Historically, the genre of sports and action video games have reigned its dominance over the past few decades. With titles such as Call of Duty, FIFA, Grand Theft Auto having overwhelming results in sales, we will dive deeper by analyzing a dataset on global sales within the topic of video games sales. Between the years of 1976 and 2017, the dataset of video games released has been recorded to be 17,416 titles. This data shows the relationship between video game global sales based on the platform (Xbox 360, Wii, etc), year of release, genre, critic score, critic count, user score, user count, and the rating of the video games. For information on how we identified the data, see the appendix below.

The explanatory variables (column section) will consist of the platform, year of release, genre, critic score, critic count, user score, user count, publisher, and the rating. The global sales will be the response variable. The game itself will be rows which are the observations.

Variables in which we think that need to be deleted are NA sales, EU sales, Other sales. The reason for this is that the NA sales, EU sales, and other sales all sum up to being the total of the global sales which is in short split data which we see as unnecessary. We are also dropping rows when there are missing values in any of the variables. For example, if one game is missing a Critic Score (or Platform information, or genre, etc.), we dropped that row.

In conclusion, after processing our dataset we were left with 7191 video games (rows). We dropped 4 variables because they were considered as split data. Our next step in the process is using the information to focus on statistical modeling.

#### Appendix:

Dataset Name: Video Games

The **outcome variable** is Sales All other variables are **independent variables**.

#### Table 1: Enriched Variables

<b>Variables</b>	<b>Description</b>	<b>Type of variable</b>
Name	Name of the games	Identifier (This will not be included in the model)
Platform	Platform of the release, Ex: Wii, NES,G	Categorical
Year_of_Release	Release date of the game, between 1976 to 2017	Numerical
Genre	Genre of the game, Ex: Sports, Puzzle, Shooter	Categorical
Publisher	Publisher of the game, Ex: Sony, Microsoft, Activision, EA	Categorical
Global_Sales	Total worldwide sales in millions	Numerical
Critic_Score	Aggregate score compiled by Metacritic staff	Numerical
Critic_Count	The number of scores used in coming up with the critic scores	Numerical
User_score	Score by metacentric subscribers	Numerical
User_Count	Individual user score	Numerical
Rating	Ratings on a game confirms that it contains content suitable for a certain age group. Example: E- Everyone, M- Mature, T- Teens	Categorical