**Analysis of Console Video Game Sales**

Robert Johnson

1. INTRODUCTION

Video gaming has come a long way since the early days of Pong and Pac-man. The aggressive competition between gaming companies has driven the technological leaps to superior features and video gaming quality we see today. The more than 70 gaming consoles has produced 100’s of memorable characters that are embedded into international culture.

In 2019, video games sales were more than $116 billion globally, and nearly two-thirds of American homes have household members who play video games regularly.

The game industry is also often at the forefront of computer technology and drives popular culture with console and PC “gamers” make up the bulk of sales.

Using data collected of will focus on game console and use that data to predict future game sales.

1.1 Data Sets

Gregory Smith used a web scraper on VGChartz Video Games Sales to make the foundation of this information, the data set is extended with another web scrape from Metacritic. This data set has about 11,000 entries and about 6900 are not missing data.

The data includes 17 metrics items : Name, Platform, Year\_of\_Release, Genre, Publisher, NA\_Sales, EU\_Sales, JP\_Sales, Other\_Sales, Global\_Sales, Critic\_score, Criticcount, Criticscore, User\_score, User\_count, Developer, Rating.

<https://www.kaggle.com/rush4ratio/video-game-sales-with-ratings>

1.2 Data importation and cleaning

I imported the game sales dataset as a .csv file, which needed to be converted in to a panda data frame which I followed up with data wrangling.

Then, I cleaned the data by removing all NA’s. I also removed six columns: Critic\_score, Criticcount, Criticscore, User\_score, User\_count, Developer, Rating. Next, I convert the Year\_of\_Release to an integer.

2. DATA EXPLORATION

I initially explored preliminary relationships in the game sales data by global sales and platform from 1980 to 2019. However, the industry grew every year, but the leader changed every three to five years.

This required me to change my observation period and then eventually for platform to genre.

3. PREDICTIVE MODELS

We’ll next explore how this data can be modeled to predict future sales