

# Final Project Proposal

## Modeling and Representation of Data

### Overview

The video game industry has been growing consistently during the last two decades, and in 2017 it was worth more than 78 billion USD. Video game software sales account for around 80% of total revenue. There are several factors that influence whether a console video game is successful or not, such as the developer studio, the critics rating, the user rating, among others. I will build a model to analyze what makes a video game successful in terms of global sales.

### Research questions

The goal of the project is to find what are the most significant predictors for a video games success, measured as the number of sales around the world. I will try to use a hierarchical linear regression model to explain the number of sales a particular videogame has. Considering that the developing studio plays a major role on a customer's decision to buy a new game, and due to the fact that I cannot account for every single studio out there, I will group the data by a sample of the developers and try to build an appropriate hierarchical model.

### Data

The data was obtained from Kaggle (<https://www.kaggle.com/rush4ratio/video-game-sales-with-ratings>). It contains data from video games with more than 100,000 sales globally from 1976 to 2016. The data contains 16719 rows with the following 15 columns:

- Name
- Platform
- Year\_of\_Release
- Genre
- Publisher
- NA\_Sales
- North America Sales in Millions
- EU\_Sales
- JP\_Sales
- Other\_Sales
- Global\_Sales
- Critic\_Score
- Critic\_Count
- User\_Score
- User\_Count
- Developer
- Rating

### Project plan

I plan to build different hierarchical models to predict the number of sales a video game will have, according to the possible predictors (platform, year of release, genre, publisher, critic score, critic count, user score, user count, developer, rating). For the first week I plan to finish the data cleaning and preparation to make sure I have a useful dataset for analysis. For the second week, I plan to finish the EDA and come up with the first models. For the third week I plan to decide on the final models, write down my report and work on my presentation.