**Members**
Michael Bradberry
Tanner Lievois
Tyler Ward

# Team 4: ETL Project Proposal

*"Wisdom is the offspring of suffering and time."*

*-Emperor Izaro, Path of Exile*

Due to the sheer volume and variety of data available, we have chosen to advance with two Kaggle datasets that present video game sales and review data.

## Datasets Chosen

1) https://www.kaggle.com/gregorut/videogamesales
2) https://www.kaggle.com/skateddu/metacritic-critic-games-reviews-20112019

## Database Structure

Moving forward, the team has decided to utilize a non-relational database, specifically MongoDB. Due to the variability in available data and the lack of consistency , the team feels that this structure best fits the data presented.

## Objectives

- Create independent collections within the database to contain game data for different platforms (Xbox, PlayStation, PC, etc.)
- Each game reviewed will be an separate document within those collections
- For each game, the individual reviews will be contained within a list of dictionaries with the key for each dictionary being the review site and the value being the review itself.

## Additional Objectives – Time Permitting

- Web scrape Metacritic for game reviews in order to fill in any blanks between the two presented datasets.
- The additional data will come from the below link with the URL being concatenated and composed of the platform and game name (i.e. URL/platform/game_name)
  - https://www.metacritic.com/game
- Feed video game reviews into a model trained to determine sentiment. The outputs of the API will be one of three sentiments: positive, negative or neutral.
- The sentiment will be analyzed using the below API.
  - https://sentim-api.herokuapp.com

RICE