# About

This project focuses on building a database for gaming sales across major platforms for 2019. This allows the building of a prediction model using historical data in the data base.

# Datasets

|  |  |
| --- | --- |
| **Data Source** | **URL** |
| Kaggle (sales\_2019\_vg.csv) | https://www.kaggle.com/ashaheedq/video-games-sales-2019?select=vgsales-12-4-2019.csv4 |
| Pricecharting.com | https://www.pricecharting.com/ |

# Focus

Our focus in creating this database will be based on the two major platforms: PS4 and Xbox One. We will combine the data of each title based on these two platforms and create a comprehensive database. This allows the users to search and use the database and create analytical charts and reports based on game title.

# Methods

* Excel data separation
  + Dataset: sales\_2019\_vg.csv
  + What: Separate out data by PS4 and Xbox One platforms
  + How: Excel filtering
  + Output: ps4\_sales.csv, xbo\_sales.csv
* Copy-and-paste
  + Dataset: Pricincharting.com
  + What: Filter by platform using PS4 and Xbox One as parameters
  + How: Copy and pasting data into separate .csv files
  + Output: ps4\_prices.csv, xbo\_prices.csv
* Cleaning (using Excel)
  + Dataset: ps4\_sales.csv, xbo\_sales.csv, ps4\_prices.csv, xbo\_prices.csv
  + What: Dropping NA\_sales, PAL\_sales, JP\_sales, Loose Price, and CIB Price columns
  + How: Delete columns during Excel cleaning process
* Cleaning (using Python/PANDAS)
  + Dataset: ps4\_sales.csv, xbo\_sales.csv, ps4\_prices.csv, xbo\_prices.csv
  + What: Dropping any rows with N/A or empty values
  + How: Delete using df.dropna() method
* Joining (using Python/PANDAS)