

CS210 Final Project Summary

Group Members (and netids):

Sihua Zhou (sz583), Hench Wu(hhw14), Deep Patel (dp1062), Joshua Clayton (jrc428)

About the Dataset:

- **Dataset name:** Video Game Sales (from Kaggle)
- **Dataset description:** The dataset contains a list of video games that sold for more than 100,000 copies. The data from Kaggle was generated by a scrape vgchartz.com. The dataset itself consist with many information about the game itself and its copies been sold at different region.
- **Data source:** <https://www.kaggle.com/datasets/gregorut/videogamesales>

Why we picked this Dataset:

- Analyze video games sale can help people who plays or planning to start gaming as entertainment making informed decision. The dataset only have games that are sold with more than 100,000 copies meaning these games are having certain popularity amount players and they are popular probably because they are good. Therefore, the analyzed dataset can serve players and new-entered players well on finding new game, finding similar games, most important, find a good game that fit the player well. In addition, the industry can use it to make informed prediction on what genre, what type, and what platform they should focus on when developing a new game. The goal and reason of

picking this dataset is to see the potential of the game industry and predict what kind of new games are likely to come out later on.

Dataset Statistic:

- 16600 rows:
 - 11493 distinct games
- 11 columns:
 - (Rank, Name, Platform, Year, Genre, Publisher, NA_Sales, EU_Sales, JP_Sales, Other_Sales, Global_Sales)
- No missing values.
- Dataset: Total 2.07M views and 5.55k downloads