- **Title and authors**: The Mystery Behind Video Games (Aaron He, Vivian Feng, Daisy Lee)

- **Research questions**.
    - Daisy: What are the top 5 video games with the most sales and in which region did they sell the most?
    - Vivian: What genre of video game is currently trending the most and what potential factors contribute to its appeal? Has this genre been consistently popular over time?
    - Aaron: Develop a model to predict global sales of a video game based on platform, genre, publisher, and regional sales

- **Motivation**.

Video games, as one of the primary methods for people to reduce stress, hold a substantial share in the global market. On one hand, the video game industry is a significant contributor to the global economy, exerting significant cultural influence. Within gaming, users can gain insights into diverse cultural realms such as fashion, languages, etc. Based on the current trajectory, it's evident that the video game industry will become even more popular in the future. Through exploring our research question, we can identify trends in the video game market, including changes in consumer tastes, popular genres, and favored publishers. The finding will provide valuable direction for decision-making among game designers and sales professionals, enabling them to make informed decisions based on user preferences which forms our research focus.

- **Data setting**.

The dataset our group has chosen for analyzing is 'Video Game Sales Analysis.'(https://www.kaggle.com/code/suchitravelusamy/video-game-sales-analysis). The dataset is exploring different aspects of video game sales in order to understand the video game market. The dataset contains 11,493 unique video games across diverse genres, and is a comprehensive resource for analyzing our research question aligned to video games. The dataset includes information on each game's title, platform, release year, genre, publisher, and sales performance in North America, Europe, Japan, and other countries.

By exploring 3 example data entries, we've identified that the dataset starts from the year 1984, with "Duck Hunt " being the earliest included game, hence for an analysis of user taste trends, we will commence our investigation from 1984 onwards. Additionally, the data include a classic game "Super Mario Bros."

released by Nintendo, the entry includes its year of release, genre, and sales performance across different regions. Last but not least, we noticed "Wii Sports" on the "Wii" platform. We found the titles of video games and their corresponding platforms may share identical words. Therefore, we will be particularly careful when handling such data to avoid potential confusion during analysis.

- **Challenge goals:**

Challenge Goal 1: Machine Learning

Our first challenge we plan to meet is machine learning. As a group, we are confident that we will meet this goal because there have been helpful explanations in class to teach us this concept. There are also several resources both online and in person to correctly go about this machine learning challenge. Overall, we are most confident we will meet this goal because it is material that we have already gone over, so it'll be more straightforward to us.

Challenge Goal 2: Implementing New Libraries

Our second challenge goal we plan to meet is using new libraries that we need to explore on our own. We are confident that we will meet this goal because there are plenty of helpful resources online to fully understand how to implement a new library you are using. Additionally, these libraries may have a similar format to other libraries that we have already used which will allow us to easily implement them into our code.

- **Method:**
  1. We want to make sure that the dataset is fully cleaned of NaN or null values, so we will first clean the dataset to get rid of these values.
  2. After this is done, we will make some visualizations to display the research questions. We will explore which ones best represent the question and play around with the different types of plots.
  3. Then, we will explore libraries like plotly to implement them into some of our visualizations.
  4. We will choose the visualizations that best answer our research questions.
  5. We will then perform machine learning to answer questions that require us to make predictions and display the results.
  6. Finally, we will draw conclusions from our findings and hopefully generate an insightful report.

- **Plan**:

For this project, we divided the project into two parts: visualization and machine learning. For the visualization, firstly we try to explore the top 5 video game sales and the region that sells the most; secondly, we try to visualize the trend of each genre of video game and find out the reason behind the popularity. In addition, we try to predict the global sales under each variable. All of the work above will cost us 5 hours because we should clean the data and analyze the data to do our research question. Therefore, we should clean the data like missing values or repetitive columns, and analyze the data by showing some parts of the code, then we can use the data to do our project on our own computer and finally combine it into one final project.