Project-proposal

Due to Covid-19, video game market is becoming more popular than before. Now the gaming industry is booming and expected to cross $300 billion by 2025. However, there are some questions for game makers and investors such as what games are more popular in Europe or which platform should they release their games that can maximize their profits. As a gamer, I’d like to do data analysis about video game sales and solve these problems. The dataset I will use is from Kaggle.

Link: <https://www.kaggle.com/gregorut/videogamesales>

Dataset:

This dataset contains a list of video games with sales greater than 100,000 copies. It was generated by a scrape of vgchartz.com. It includes 16,598 records with 11 attributes.



Fields include:

* Rank - Ranking of overall sales
* Name - The games name
* Platform - Platform of the games release (i.e. PC,PS4, etc.)
* Year - Year of the game's release
* Genre - Genre of the game
* Publisher - Publisher of the game
* NA\_Sales - Sales in North America (in millions)
* EU\_Sales - Sales in Europe (in millions)
* JP\_Sales - Sales in Japan (in millions)
* Other\_Sales - Sales in the rest of the world (in millions)
* Global\_Sales - Total worldwide sales.

Data mining tasks:

1. Video game market analysis: popular games, genres, platforms and etc.

Descriptive statistics.

1. Use this dataset to forecast the future sales of video games.
2. Visualization of games sales

Data mining:

Association rule mining: use apriori to find which combination of platform, genre and publisher can help games be well-sold.

Decision tree: use a DT model to help game makers to decide whether to make the games or not.

Clustering: use k-means to find the games that NA, EU or JP tend to buy and summarize their characteristics (e.g., sports video games released on PS were popular in North America). I think we can get some interesting results. And it is beneficial for game publishers to avoid stereotypes.