Proposal

Yuxuan Du

2023-10-31

## group members

Yuxuan Du: yd2739

Tara Zhan: fz2377

Yaduo Wang: yw4199

Chenshuo Pan ：cp3384

Wenxin Tian: wt2369

## Tentative project title

Does trending game have similar characteristics?

## Project motivation

Two of our group members frequently use STEAM. As gamers, we often come across opinions suggesting the existence of a formula that can guarantee the success of games, such as genres like shooting, ball-games, and car-racing games being more likely to succeed. We are curious to know if this statement holds true. Furthermore, even if there are no deterministic factors that directly influence sales, are there any common key characteristics that contribute to the success of games?

## The intended final products

We will introduce a model designed to predict the popularity of games based on various features, including their genre. By utilizing these models, we aim to ascertain whether there are significant features that contribute to a game’s popularity. However, we acknowledge that there may be various definitions of popularity, and in our study, we hypothesis it as a combination of both sales and score rankings.

## anticipated data scources

We are attempting to locate a Steam dataset that contains sales data and various game features. This information should be accessible through the Steam API.

## The planned analyses / visualizations / coding challenges

Firstly, we will import our dataset and perform data cleaning, filtering out redundant information and unwanted variables.

Next, we will conduct Exploratory Data Analysis (EDA) to examine the distributions of various variables.

Also, we need to establish clear classification boundaries for popular and unpopular games, taking into account both sales figures and player satisfaction rates.

Following that, we will explore suitable models for classifying games based on their distinct features. If the selected model offers interpretability, we will also analyze the importance of different features.

A challange is that since the amount of popular games and unpopular games might be unbalance. We might need to find ways to reduce such unbalance. (maybe we need down sampling unpopular games)

## The planned timeline

We already found dataset from Kaggle. The dataset use Steam API to get datas. They provide the code for using API so it is possible if we want to find latest dataset.

We plan to do data cleaning and EDA from Nov 13-17.

Then, we plan to find appropriate models to fit our dataset from Nov 17 - Nov 24.

Then we will make our webpage from Nov 27 - Dec1.

Then, we will make the video and review our project from Dec 4 -Dec8.