Steam Games

Maximilian M. Krier

# Guiding questions

This report will attempt to answer the following questions.

1. Does the price of a game influence the review score of a game?
2. Does playtime influence whether the game is recommended by the reviewer?
3. Do players find reviewers with more playtime more helpful?
4. What are the highest reviewed paid games, and their prices?
5. What are the highest reviewed free games?
6. What is the optimal price point for paid games?
7. How does early access effect reviews?

# Introduction

Steam is the industry leading videogame markeplace for PC, Mac, and Linux. This service is used by nearly everyone who plays games on a computer. As a result of this huge traffic, studios all across the world use Steam. Steam has a very sophisticated community review system, this system helps to inform the buyer about the game they are purchasing. This isn’t the only factor in deciding whether or not to buy a game, price is a big motivator.

## Primary data set

* Description: The dataset is a compilation of games on the game marketplace Steam.  It contains information such as game name, price, review percentage, number of review, release date, and tags.
* Source: r/gamedev subreddit, posted by user Sadaris.
* Cases: Each case represents a game on the Steam marketplace.

## Secondary data set

* Description: This data set is a compilation of reviews of games on Steam. It includes game name, text of the review, review score (positive or negative), and review votes(was this review helpful or not)
* Source: Kaggle by user ILJUSSA
* Cases: Each case represents an individual review

# Data wrangling

These data sets will be used both separately and merged to create unique data sets to answer the questions.

## Data cleaning

The primary data set was cleaned using the select function to only include relevant data.

The secondary data set was renamed to more fit the names given to the primary, then cleaned using select to highlight relevant data.

library(devtools)  
library(tidyr)  
library(dplyr)  
library(ggplot2)  
  
df\_steam <- read.csv("C:/Users/maxkr/Downloads/Steam Trends 2023 by @evlko and @Sadari - Games Data.csv", header=TRUE)  
 View(`df\_steam`)  
   
df\_steam\_clean <- df\_steam %>% select(  
 Title,   
 Reviews.Total,  
 Reviews.Score.Fancy,   
 Release.Date,   
 Launch.Price,  
 Revenue.Estimated  
)  
  
df\_steam\_2 <- read.delim("C:/Users/maxkr/OneDrive/Desktop/archive (1)/dataset\_top20\_cleaned.csv")  
 View(`df\_steam\_2`)  
  
df\_steam\_2\_renamed <- df\_steam\_2 %>% rename(  
 Title = app\_name,  
 Reviews.Text = review\_text,  
 Review.Score = review\_score,  
 Review.Votes = review\_votes  
)  
   
df\_steam\_2\_clean <- df\_steam\_2\_renamed %>% select(  
 Title,  
 Reviews.Text,  
 Review.Score,  
 Review.Votes  
)

## Visualization