STEAMCHARTS WHAT'S POPULAR?

Searching for trends and hidden gems
amongst the most popular games on Steam.

Ryan McDermott | DSCI510 Final Project

DATA SOURCES
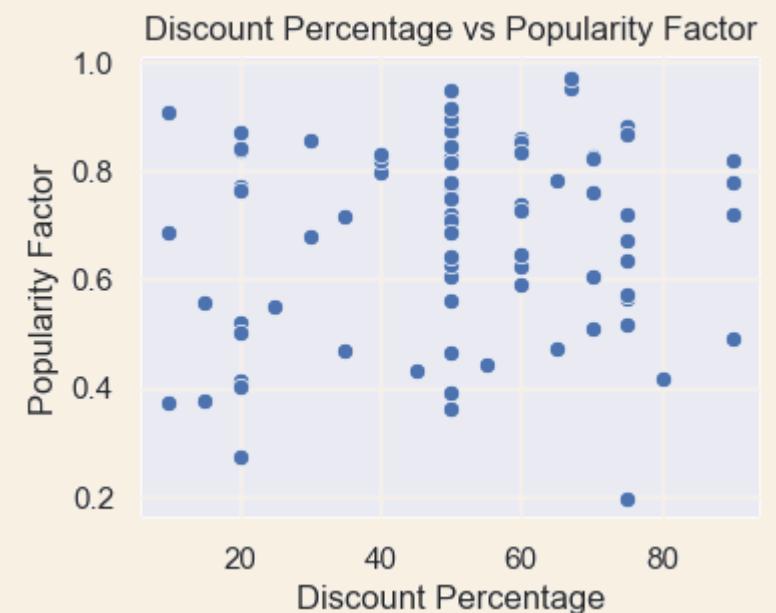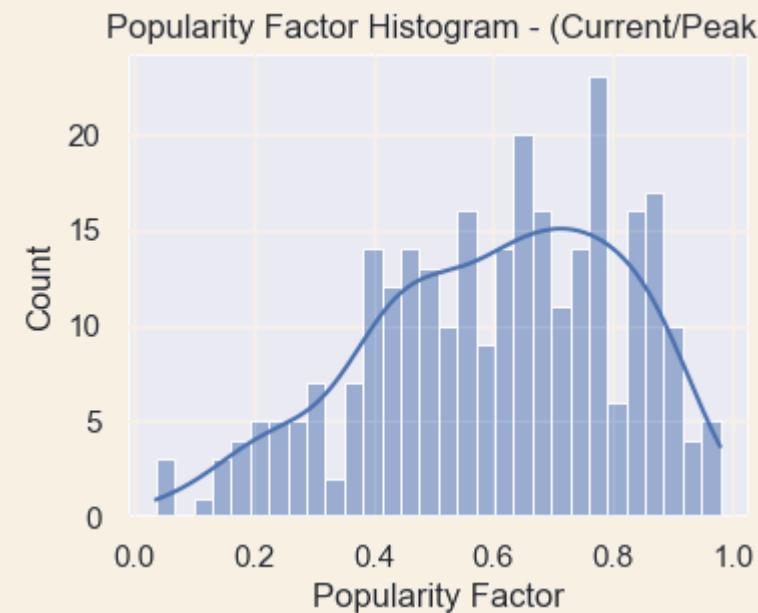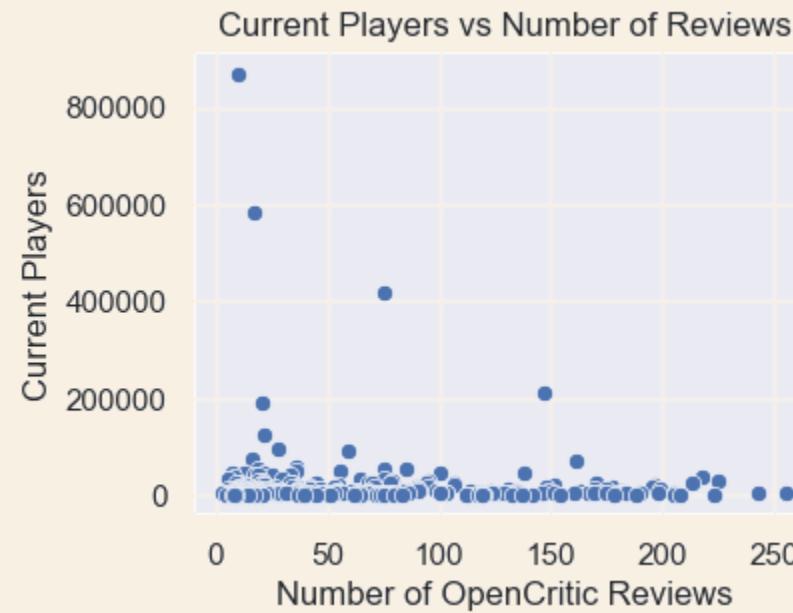

| Dataset	Description/Type	Processing Approach	Data Size
SteamCharts	Website that presents the most popular games on Steam currently. Includes game names, appids, current players and peak players	The site is scraped via beautifulsoup to collect information one the requested number of games.	375 rows x 4 columns
Steam Store Web API	API that returns specific information such as price, discounts, if it's free, release dates	The Steam appids scraped from SteamCharts are used to request information from the Steam Store Web API.	375 rows x 9 columns
OpenCritic API	API that returns review information such as top critic review, median critic scores, % recommended, total reviews and a ranking tier.	The OpenCritic API is used to and the game name is searched and the initial results returned.	375 rows x 7 columns

RATINGS, DISCOUNTS, & POPULARITY

$$Popularity\ Factor = \frac{Current\ Players}{Peak\ Players}$$
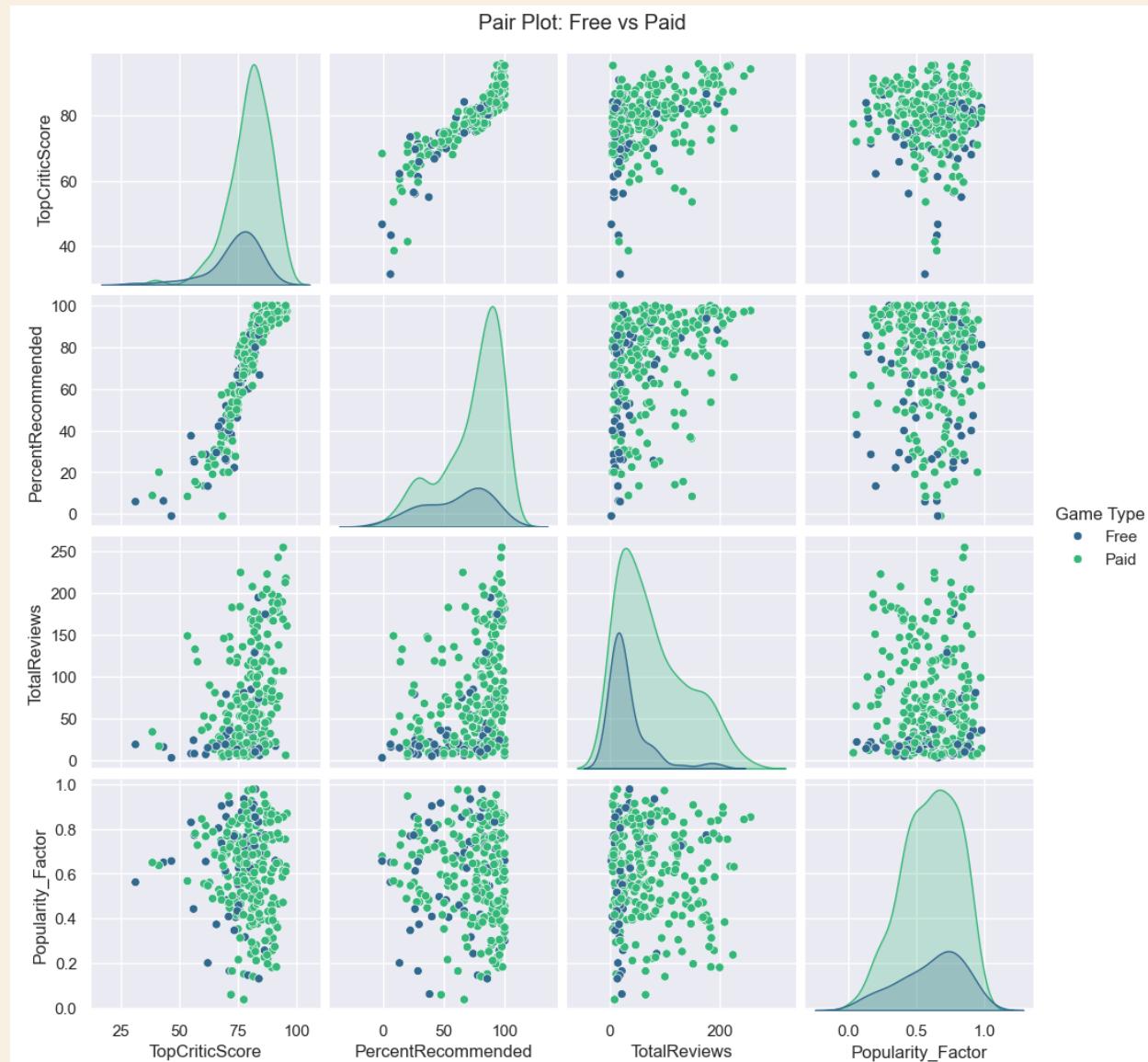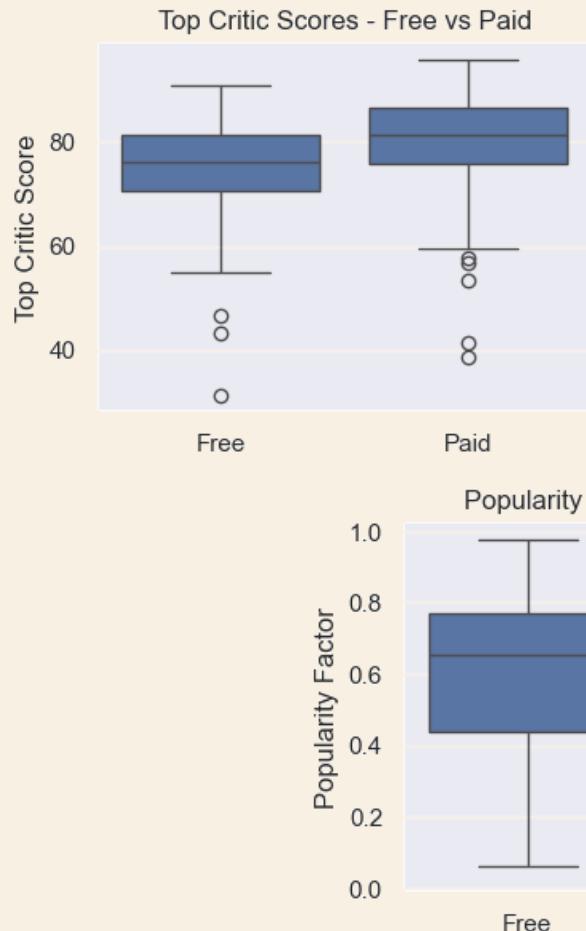
Data was cleaned to remove games that didn't have any OpenCritic reviews and were marked as not free but had a price of 0$.
- This data set went from 375 games to 286 games.
and then split between Free and Paid games
- 60 Free Games
- 226 Paid Games
    - 86 of which were discounted

Current Players vs Number of Reviews

Correlation between the total OpenCritic reviews and the current number of players: 0.0206

Popularity Factor Histogram - (Current/Peak)

Discount Percentage vs Popularity Factor

Correlation between the discount percentage and the popularity factor: 0.0734

FREE VS PAID

Interestingly, paid games tend to carry higher critic scores – both top and median, but free games edge out on popularity.
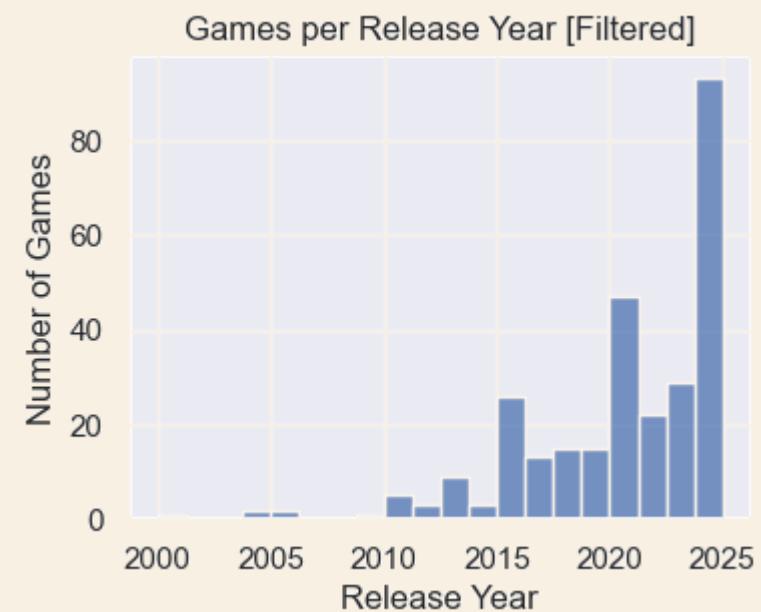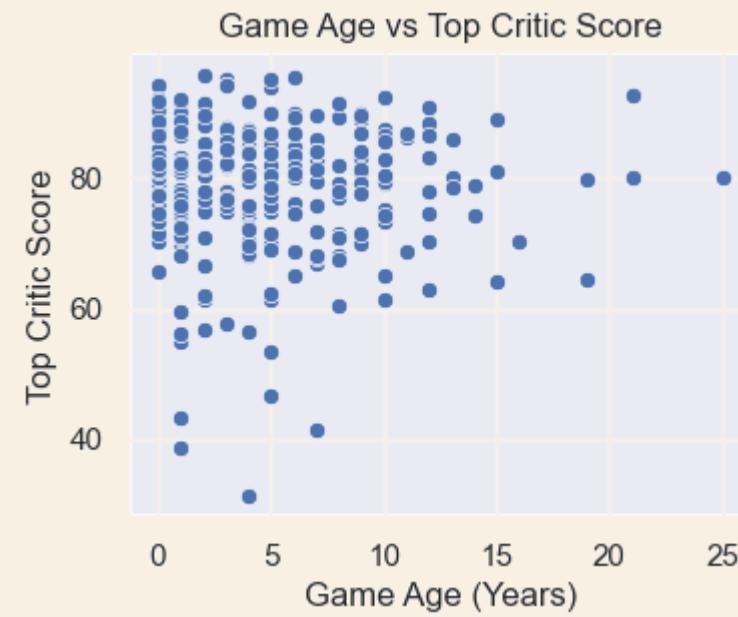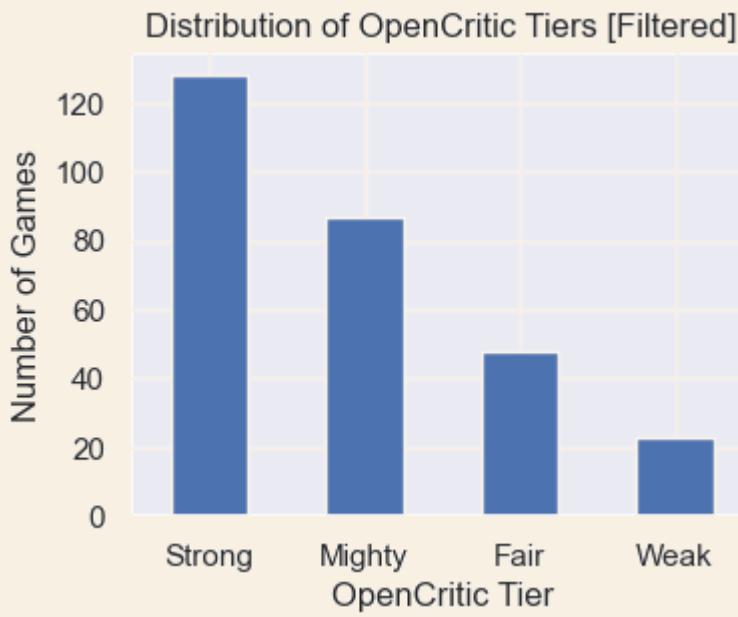
RELEASE DATES AND AGING

Unsurprisingly, the best rated games were the most popular per the OpenCritic tiers.

The older a game is in the list, the more likely it was reviewed highly.

The most popular games are currently dominated by newer games.

HIDDEN GEMS

To locate the "best" and "worst" of the most popular games, a Value equation is calculated below:

$$Value = \frac{TopCriticScore * Popularity\_Factor}{Discounted\_Price + 200}$$

This value calculation attempts to prioritize popularity and the highest ratings over the cost – otherwise free games tend to dominate the value rankings.

The worst value tend to be the most expensive games that have become abandoned while also being rated lower than the majority of the games queried.

Top 5 best value games:

	game name	Current Price (USD)	TopCriticScore	Popularity_Factor	Value
119	Magic: The Gathering Arena	0.00	82.448276	0.978110	0.403218
31	Risk of Rain 2	8.24	86.052632	0.952936	0.393789
167	Fallout: New Vegas	9.99	81.166667	0.977123	0.377684
97	Lost Ark	0.00	81.691176	0.909796	0.371612
361	Disco Elysium - The Final Cut	9.99	89.235294	0.865516	0.367801

Bottom 5 value games:

	game name	Current Price (USD)	TopCriticScore	Popularity_Factor	Value
122	Escape From Duckov	17.99	77.500000	0.036804	0.024174
314	New World: Aeternum	59.99	72.000000	0.058605	0.026374
152	NARAKA: BLADEPOINT	0.00	72.166667	0.061585	0.044444
280	Black Myth: Wukong	59.99	81.530303	0.139661	0.071171
271	The Outlast Trials	39.99	72.487179	0.164542	0.085200

Top 5 best paid value games:

	game name	Current Price (USD)	TopCriticScore	Popularity_Factor	Value
31	Risk of Rain 2	8.24	86.052632	0.952936	0.393789
167	Fallout: New Vegas	9.99	81.166667	0.977123	0.377684
361	Disco Elysium - The Final Cut	9.99	89.235294	0.865516	0.367801
128	Balatro	14.99	91.638889	0.860538	0.366802
191	God of War	19.99	94.241379	0.853585	0.365667

CHALLENGES / ISSUES

- Initial SteamCharts scraping limited to 25 games due to the charts only displaying 25 at a time. Had to scrape through multiple pages.

- Steam Store Web API request limitations. API calls continued to fail after some point. Discovered I was limited to 200 requests every 5 minutes and found I could use time.sleep to slow my requests to >1.5 seconds and prevent the request limit from being hit.

- SteamCharts presents dynamic tables that include player counts per day, unsuccessfully tried to scrape data for the previous 7 days.

Dynamic Selection Example