5 viewed per hour

Stock photo

Brand new: lowest price ⓘ

$499.99

Free Shipping

Get it by **Friday, Apr 26** from Branson, Missouri

- **Brand New** condition
- 30 day returns - Buyer pays return shipping

"Pokemon Red Spanish Version (Edicion Rojo). It would make a great addition to any collection!"

See details

See all 2 brand new listings

1 viewed per hour

Stock photo

Brand new: lowest price ⓘ

$199.99

+ $9.78 Shipping

or Best Offer

Get it by **Wednesday, May 1** from Pittsburgh, Pennsylvania

- **Brand New** condition
- No returns, but backed by eBay Money back guarantee

"It is our goal to provide an issue free buying experience. Collection Sale. INSIDE HIGH END COLLECTOR'S CASE. Look at the photos closely and get a exact details."

See details

See all 4 brand new listings

Top-Selling Video Games and Their Retention Value

Kenton Hoffmann | SI 330 | April 22, 2019

Motivation/Introduction

Initially for this project, I wanted to compare top video game sales of all time against top video game consoles sales of all time and see if the number of top video games on a console affected its total sales. However, the website for the API I planned on using (IGDB.com) went through an overhaul and dramatically changed how API requests work. There was both a new and a legacy version but the issue I encountered is that they stopped supporting their python wrapper in the new version, and I faced several errors using the legacy version because they were phasing it out.

Therefore, I started looking for other sites to get data from and I found an interesting site that tracked current & historic prices for every video game (https://www.pricecharting.com). I remember reading stories online of vintage video games selling for several times their original price on eBay, so I thought it would be interesting to see how the top video games of all time retain their value if they're preserved in mint condition. I wanted to know what top selling video games were worth the most and if the value changed for games released on several different platforms. For example, is Tetris worth more on the Gameboy or the Famicom, and is the difference substantial?

Data Sources

For the first source, I used the Wikipedia page of top-selling videogames to get the most up-to-date version of best sellers. The reason I used this page is that it had one of the longer lists of top selling video games (https://en.wikipedia.org/wiki/List_of_best-selling_video_games) and it accumulated total sales of a game across several consoles, not just one. This is critical for my report because I wanted to see if a game's resale value varied based on its console and getting the total sales for only one version of the game would ruin the comparison. I used the read_html function in Pandas to get a list of tables from the page, and through trial-and-error I managed to find the table that contained the data I needed. The important variables in this table were Title (of the video game), Sales, Platform(s), and Initial release date. This table also included the developer(s) and publisher(s) of the game, but I decided to refrain from using these variables as I already had several visualizations planned at this point and I did not want to stray too far from my original objective. I used 50 records from this site, and they cover video games released from 1981 to 2018.

For the second source, I searched the database of Price Charting (https://www.pricecharting.com/) for the games I found from the first source and filtered their suggested price for mint condition copies of the game. Originally, I was going to request the data using an API key, but then I realized that I would have had to individually request each game. I wasn't for sure if I would reach a request limit and writing each

individual request would result in a large block of code and it would take a considerable amount of time searching each json for the right data and adding them all together. Instead, I decided to use the "collection" feature on the site to add all the games to one list (https://www.pricecharting.com/offers?seller=du2htjhykngh6faeqllxbvkiya&status=collection) . Then, I could export the collection as a csv file and place that file in the project folder to avoid having to send several requests to the website. Next, I loaded the data from the csv file into a data frame and kept the variables "product-name", "console-name", and "price-in-pennies". The csv file was 12 kb and the resulting data frame contained 161 rows and 3 columns. This source did not contain a time element, but it will cover the same time periods as the first source (video games released since 1981 to 2018).

Data Manipulation

For the first source, I needed to drop several columns as well as the first row in the dataframe (game_df) so that I would only be using the necessary data to perform my analysis. Furthermore, I had to change the index labels of the dataframe so that the readers could clearly understand what each row represented. I also renamed the "Title" column to "product-name" so that when I combined the two dataframes, I would have a column to merge on. I used an inner merge to preserve the order of the left keys which meant that this dataframe's keys would be first. When I combined the dataframes, I changed the values in the "Sales column back to "int64" type so the values would order correctly in the graph.

For the second source, I had to drop additional irrelevant columns when I created the game_prices dataframe, but I did not have to rename the saved columns liked I did for the first dataframe. When I merged the dataframes, I created another column to convert the price of the games from pennies to dollars so that when I compared it against the total sales, they would have the same units. One challenge I encountered is that I have some games listed several times because they appear on multiple consoles. For example, even though Tetris has four rows in the merged dataframe (combined_df), the only differences between each row is the console and the corresponding price. I initially tried to create a pivot table to illustrate this instead, with "product-name" as the index, "console-name" as columns, and values being "price-in-dollars", but this failed to make it easier to see. I have 50 games and over 30 unique consoles, so the pivot table had a lot of empty values that ultimately looked worse than the dataframe.

Analysis and Visualization

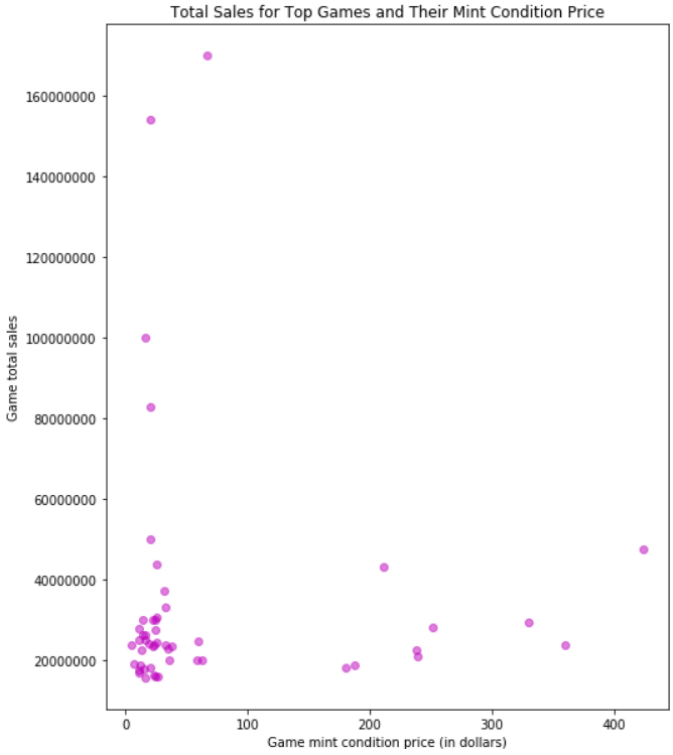
For this project, I created four different visualizations to determine what top selling video games have retained the most value and whether the platform plays a role in a game's resale value. The first visualization is a scatter plot comparing a game's total sales to the

mint condition price of the game; based on the distribution of points, there doesn't seem to be a correlation between the total sales of a game and its value, as games with the highest number of sales have a low price similar to the games with lower number of sales.

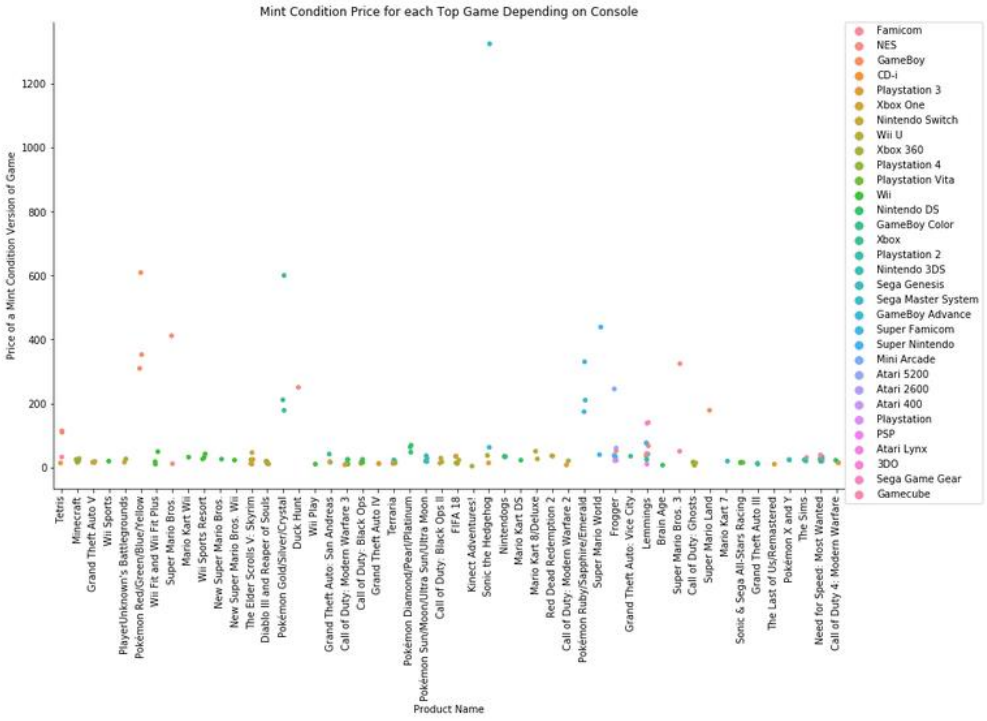
Subsequently, the next visualization utilizes Seaborn to show the mint condition price of each video game based on the platform. Based on the resulting graph, the price of a game is relatively the same across consoles, with a few exceptions. The most notable outlier is *Sonic The Hedgehog*, as a mint condition copy of this game is worth over $1,200 on the Sega Master System but is less than $100 on other platforms. This explosion in price is likely due to the Sega Master System being a rare console, and thus games compatible with this platform are even harder to find. Another inaccurate representation in this visualization is the various Pokémon games included. Upon closer inspection, it looks like there are multiple prices for each Pokémon game for the same console; for example, there are three different values for Pokémon Red/Green/Blue/Yellow that all represent GameBoy. When a new Pokémon game releases, there are often several different variations of the game and each version has minor differences. These differences are so miniscule that sales are calculating collectively rather than individually for each game. However, some versions of the game are worth more, so even though they are all roughly the same game, they fetch different prices due to each version's rarity.

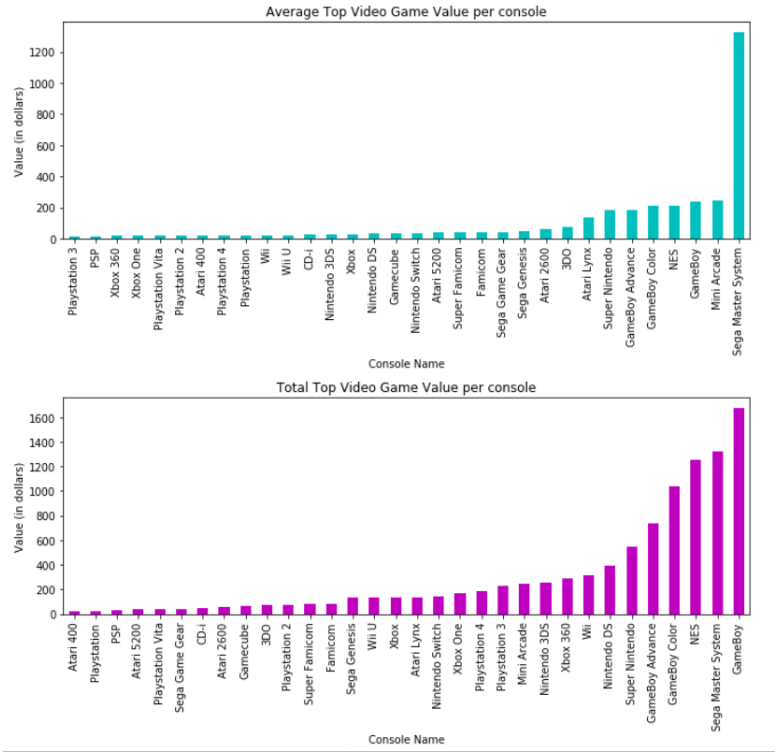
For the following visualization, I created two subplots to compare Top Video Game Values across consoles. For the first subplot, I took the average mint condition value of all video games per console, while for the second subplot, I took the of every video games' mint condition value per console. The Sega Master System is clearly the console with the highest average since it only has the 1 game (*Sonic The Hedgehog*) but compared to the total mint condition value of top video games, the GameBoy is higher because it has several valuable games. Other Nintendo products like the NES, GameBoy Color and GameBoy Advance have a higher total value as well.

Finally, I created a visualization showing how many top selling games are available on each console and I discovered a surprising result. The Xbox 360 and PlayStation 3 had the most top selling games on their platform, even doubling the amount of top selling games on the GameBoy. One explanation as to why the top selling games on Xbox 360 and PlayStation 3 sell for lower prices is that since the platforms are relatively new, it's still easy to find copies of a top selling game than it is for a console released more than 20 years ago, like the Sega Master System. Another explanation is that many of the top selling games are multi-platform and users can play them on several different consoles. However, many of the games that aren't multiplatform are only available on Nintendo products, so the only way to play games like the Pokémon series is if you have a Nintendo console such as the GameBoy. Overall, the top-selling games with the highest retention value are typically Nintendo exclusives for older consoles that are at least 15 years old.



Visualization 1 above and Visualization 2 below





Visualization 3 (two subplots) above and Visualization 4 below