



GameOn

- ❑ Jesus Jimenez
- ❑ Shridhar Kamat
- ❑ Marina Mukhranskaia
- ❑ Shih-Hsien (Sherry) Ma
- ❑ Portia Pray
- ❑ Gagandeep Singh



Problem Overview

Gaming Sales:

- Was there a increase in sales during covid in comparison to 2019.

Gaming Genre:

- Which is the most published game genre, and the trend of genre over time?

Data Challenges:

- Game, Sales, Genre Data
 - Majority of accurate data located behind paywalls or is published for media purposes thus, unable to extract.
 - Not enough time in overall project to merge sales data from various company in order to create a sample size large enough to properly represent all genres.
 - Each dataset we located had missing titles, genres, and regions thus, no 1 or 2 datasets were available that encompassed all the information we needed.
- Mental Data:
 - Scientific journal data requiring subscription or purchase.
 - Recent data was harder to find as psychological studies can last for many years before data is published. As a result the data itself can become outdated for the current period of time.





What data sources made the merging and sorting?

- Online Gaming Anxiety Data
 - Data collected as part of survey among gamers worldwide.
 - Questionnaire asked questions that psychologists generally ask people who are prone to anxiety, social phobia, and less to no life satisfaction.
- Video Game Sales
 - Dataset contains list of video games with sales greater than 100k copies sold.
 - Provides breakdown by rank, name, platform, year, genre, publisher, and sales divided by region.
- Video Games Sales Dataset
 - Provides merged sales data from 2 different data sets.
 - Features columns with information such as, the Video Game Sales dataset and add new columns that highlight both critic and user scores and counts, develop name, and ESRB rating.
- Video Game Sales Dataset
 - Provides sales performance and popularity of various video games.
- IMDB Video Games
 - Purpose of the dataset is to gain insights into the trends of game genre popularity.
 - Provides game trends, plot, genre, and popularity.

What was our data process?

```
# !pip install matplotlib
# !pip install --upgrade numpy
# !conda install numpy
```

```
import matplotlib.pyplot as plt
import pandas as pd
import scipy.stats as st
import numpy as np
from scipy.stats import linregress
```

```
file_path = "../InputData/vgchartz-6_23_2020.csv"
data2 = "../InputData/vgsales-12-4-2019.csv"
data3 = "../InputData/Video_Games.csv"
```

```
game_data_2020 = pd.read_csv(file_path)
game_data2019 = pd.read_csv(data2)
game_data3 = pd.read_csv(data3)
```

Step 1: Down the CSV files from our datasource websites and import each separate CSV file.

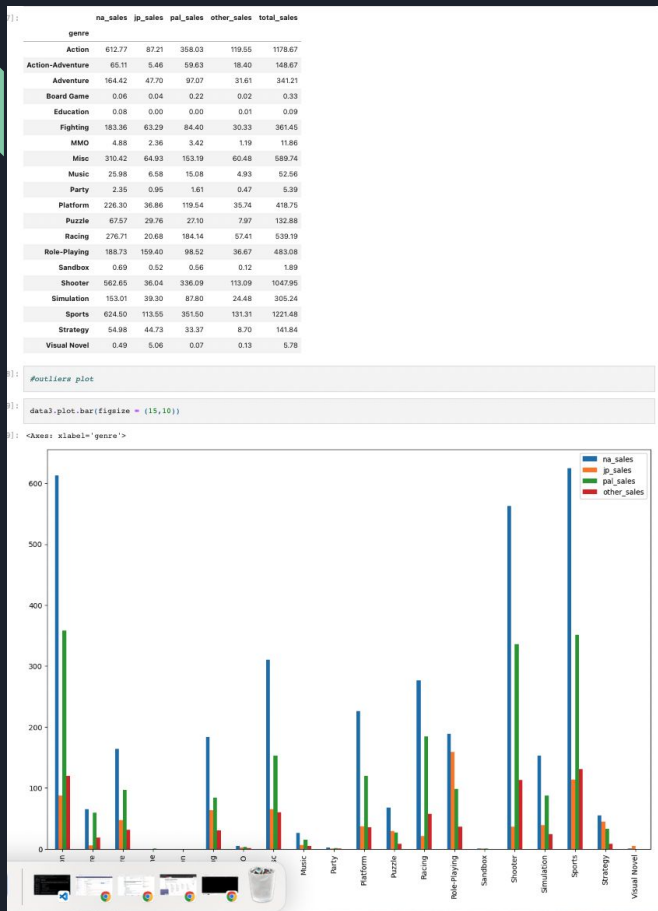
Step 2: Determine which columns we will keep for research, and create a dataframe for year 2019 and 2020.

```
list(game_data_2020)
```

```
['Unnamed: 0',
 'img',
 'title',
 'console',
 'genre',
 'publisher',
 'developer',
 'vg_score',
 'critic_score',
 'user_score',
 'total_shipped',
 'total_sales',
 'na_sales',
 'jp_sales',
 'pal_sales',
 'other_sales',
 'release_date',
 'last_update']
```

```
clean_data_20 = game_data_2020.drop(['Unnamed: 0', 'img', 'last_update', 'release_date', 'console', 'publisher', 'developer'], axis=1)
clean_data_20
```

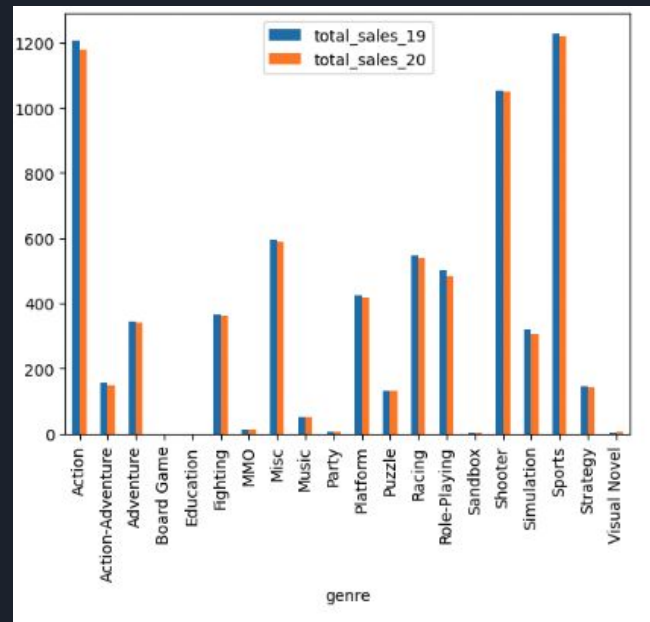
Data process continued...



Step 3: Group the data by region for sales for each year.

Step 4: Plot the data using pandas for each year.

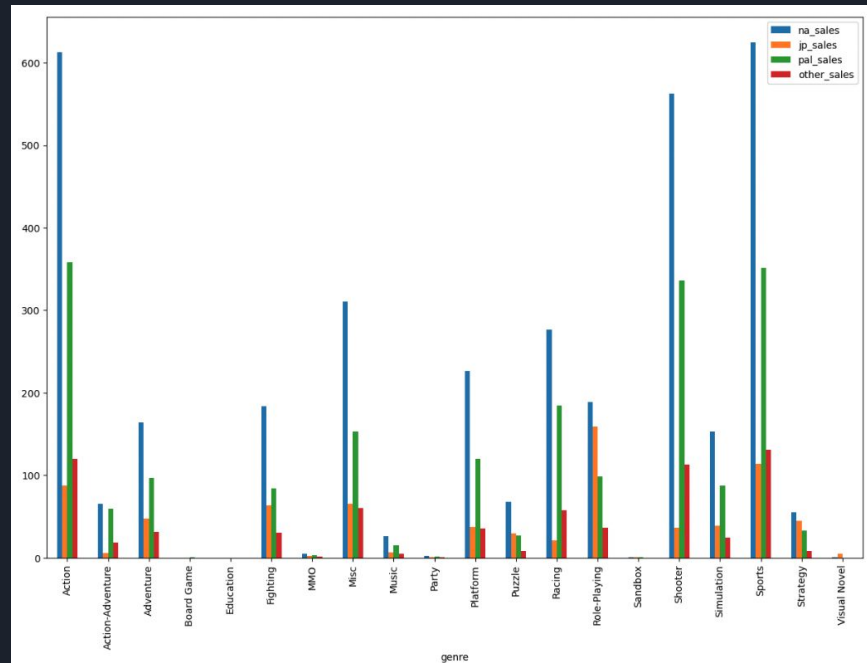
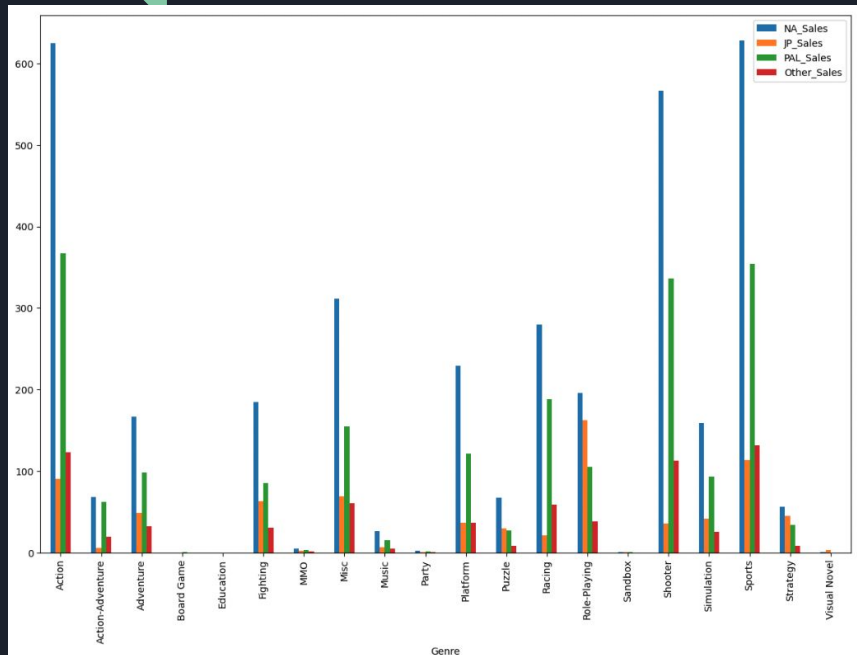
Step 5: Merge both 2019 and 2020 datasets to compare years side by side.



GAMING SALES



2019 vs 2020 Sales per Genre

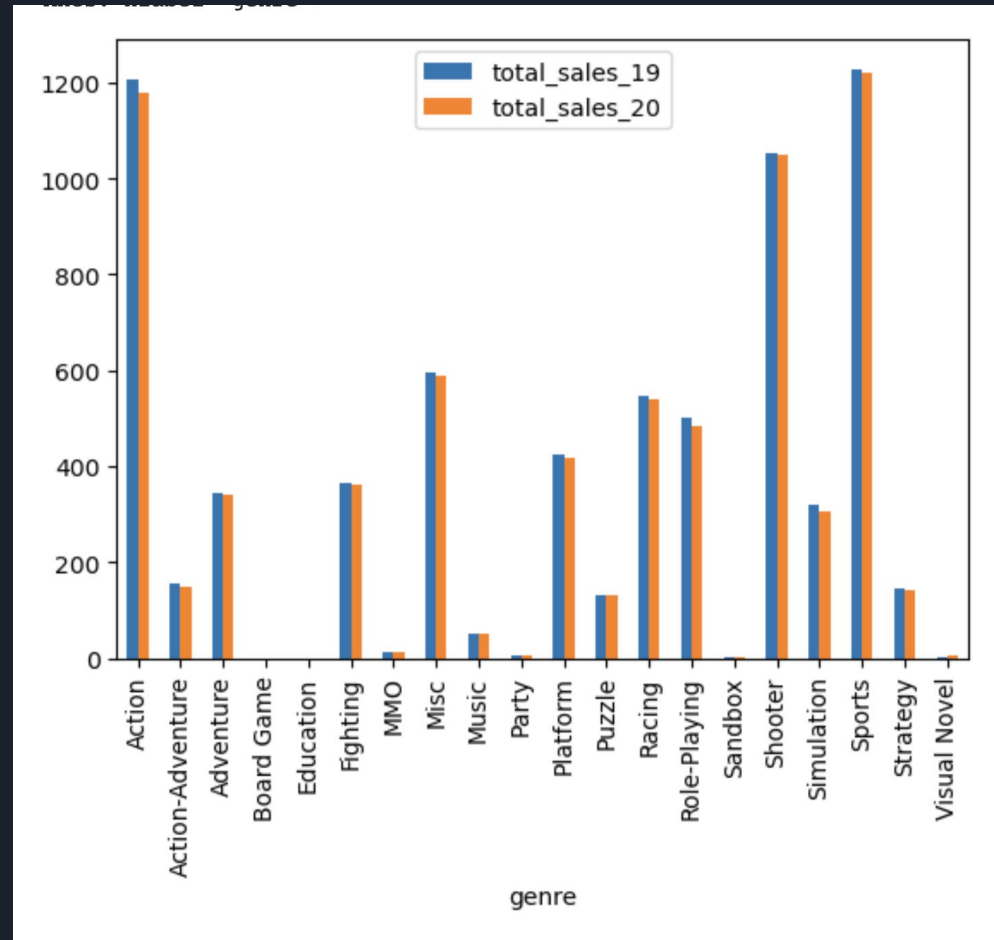


Dominant Genres over 2 year period.

- Here we took the 2 years of sales data to compare sales amount grouped by genre.
- Top 5 performing genres:
 - Sports
 - Action
 - Shooter
 - Misc
 - Racing

Conclusion:

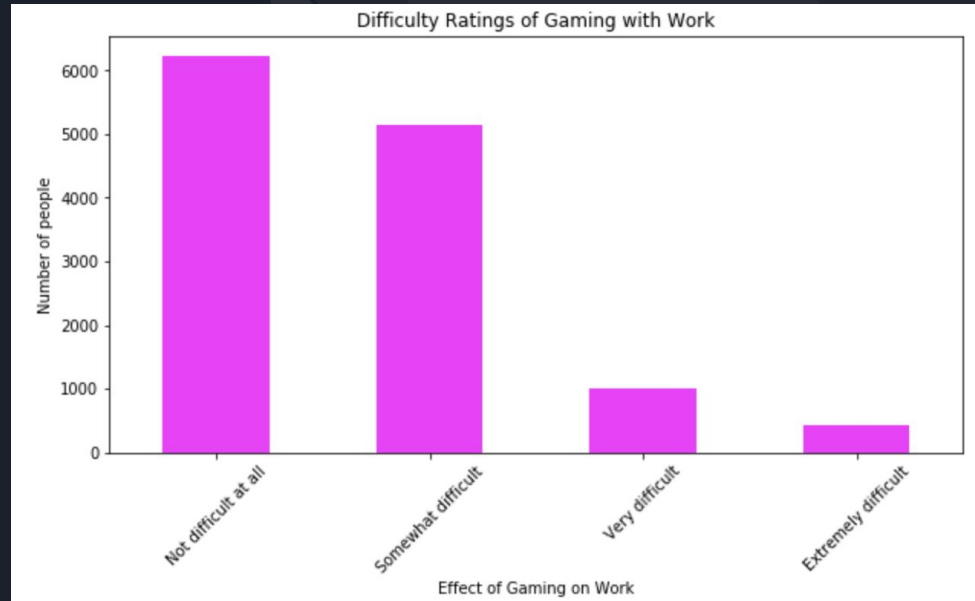
- Sales did not increase during Covid.



GAMING AND MENTAL ILLNESS

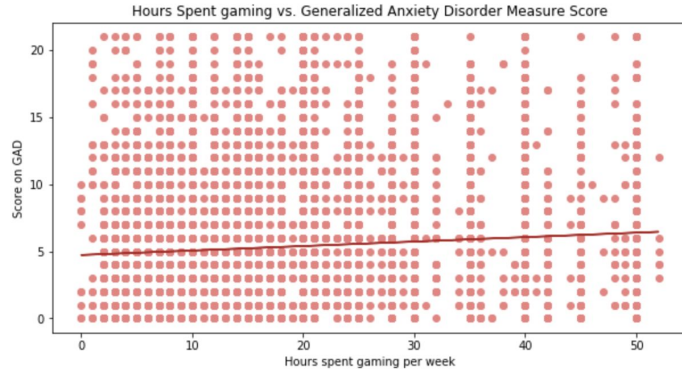


Our data

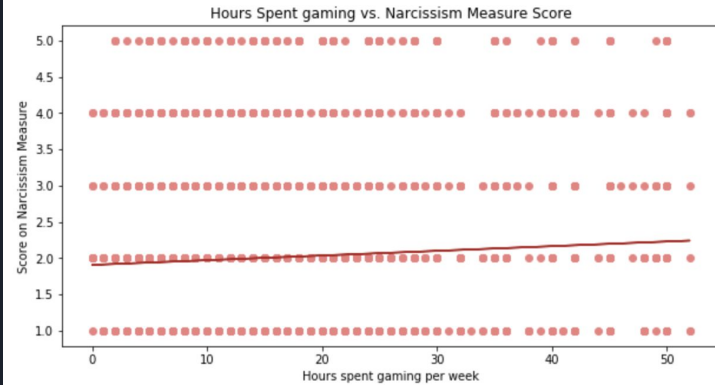


Null Hypothesis was Supported

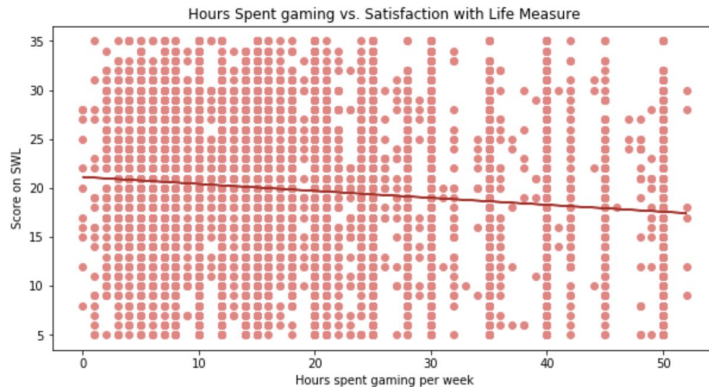
The r-value is: 0.08



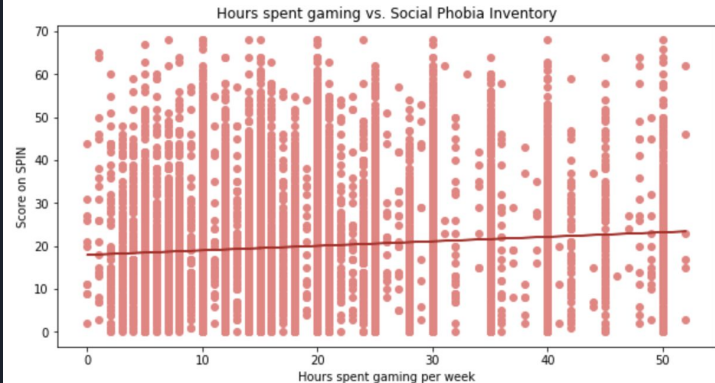
The r-value is: 0.07



The r-value is: -0.11



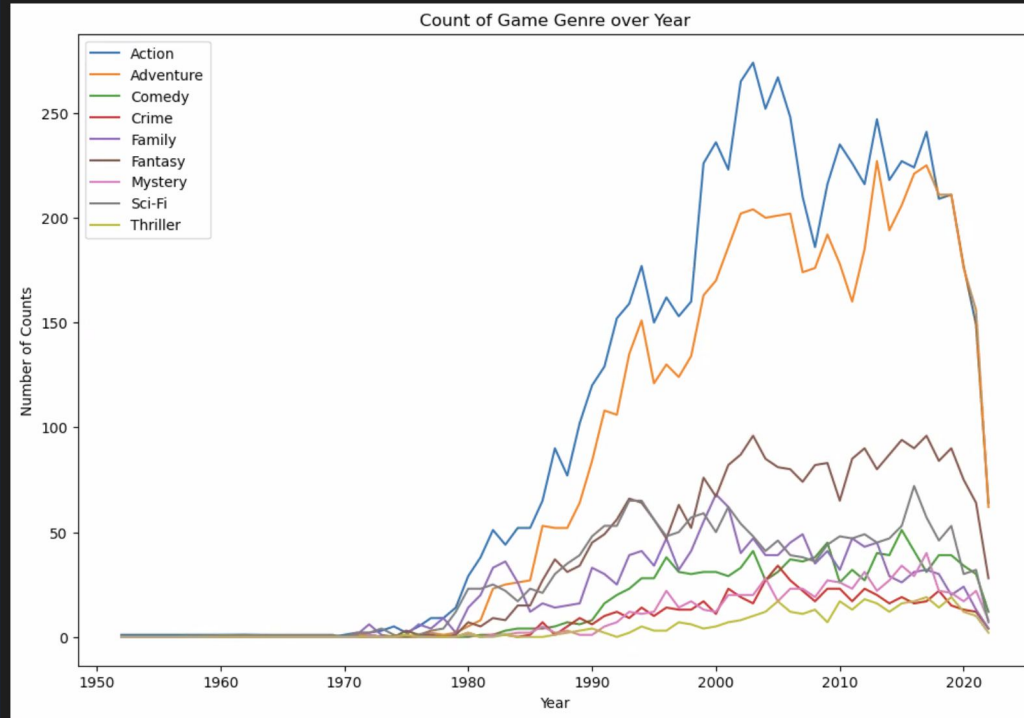
The r-value is: 0.08



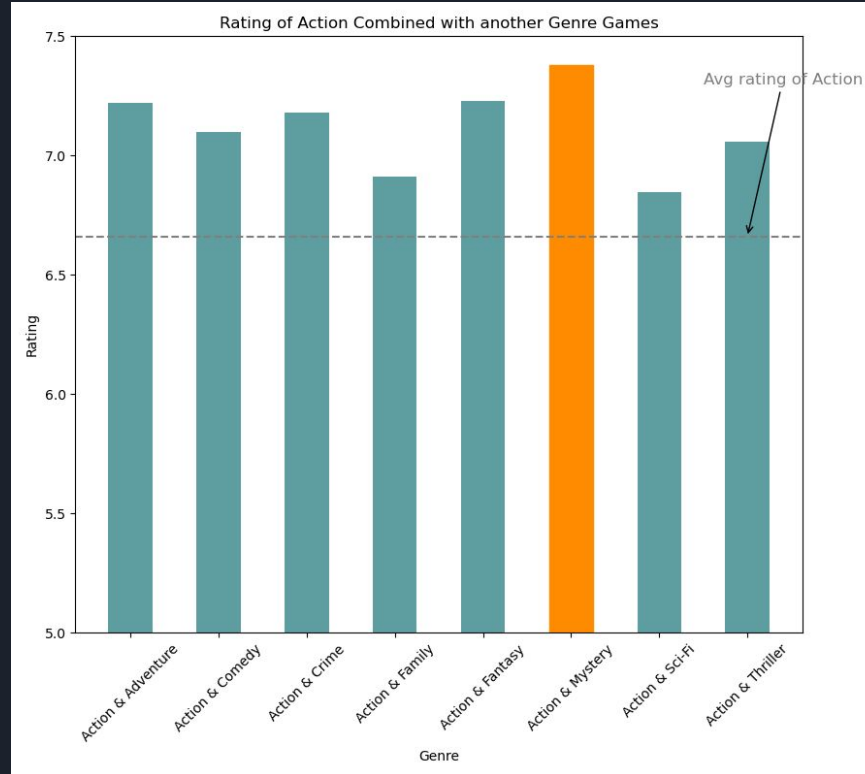
GAMING GENRE



Game Genre over Time

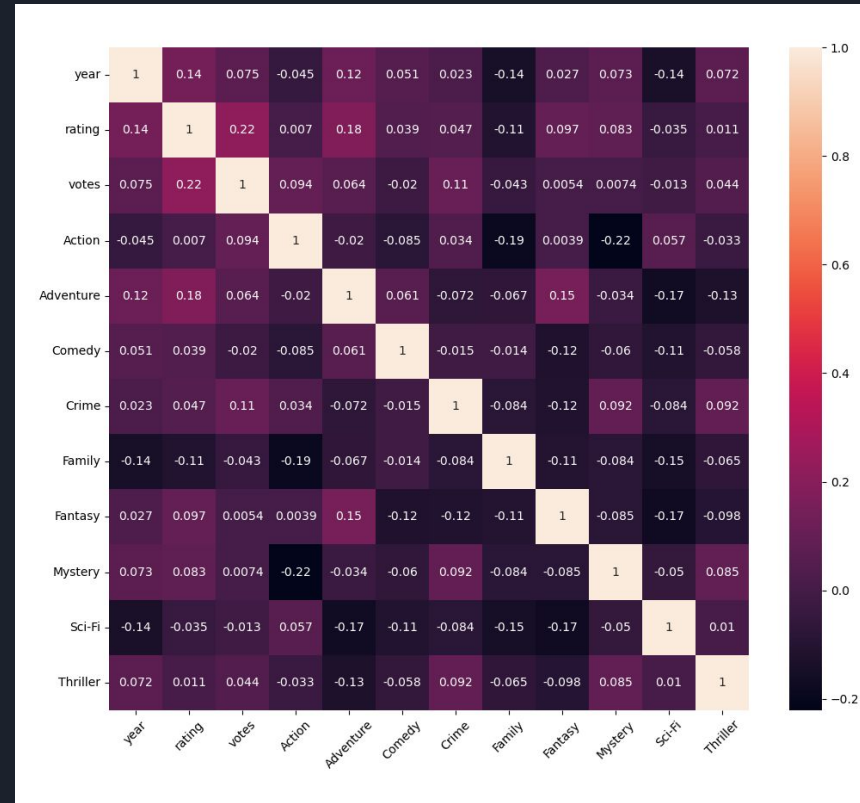


Rating of Action & another Genre



- Action: 6.66 / 10

Correlation Matrix

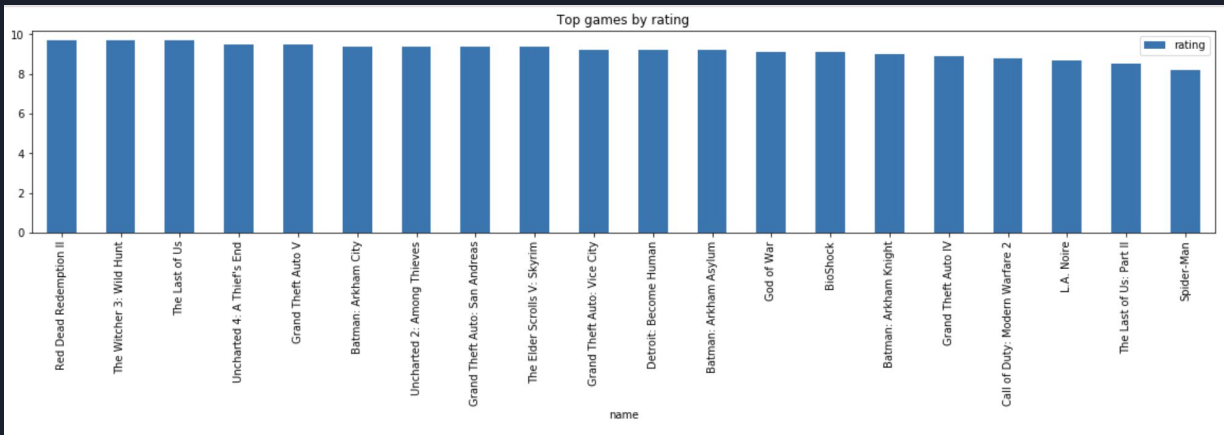
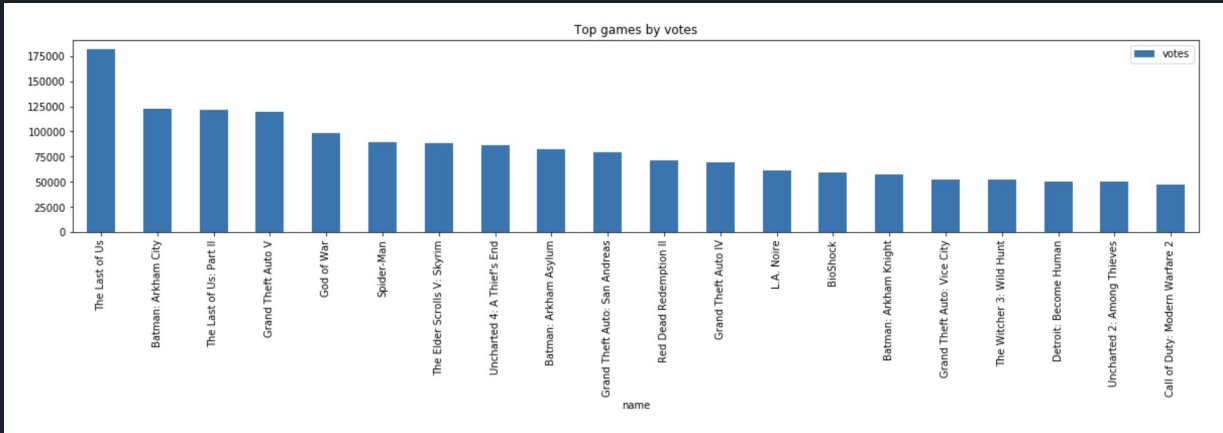


Time



Top 20 games. Ratings vs Votes







Future Plans:

Ways to expand and enhance our project?

- Having a budget
 - With funds we could access research data from scientific outlets.
- Locating API's for data
 - Having an API to pull character or gaming information from would give us more up to date information.
- Utilizing accurate and present data.
 - Having current data from sales companies within the gaming industry.
- Having expanded data over larger years and gaming platforms
 - More generalizable data to draw larger conclusions
- More time!
 - Research, finding, requesting, merging, and sorting data takes time. Having a further end date to project would result in better outcomes for our hypotheses.

Questions & Answers





Data Sources with links.

- Online Gaming Anxiety Data
 - <https://www.kaggle.com/datasets/divyansh22/online-gaming-anxiety-data?resource=download>
- Video Games Sales Dataset
 - https://www.kaggle.com/datasets/sidtwr/videogames-sales-dataset?select=Video_Games_Sales_as_at_22_Dec_2016.csv
- Video Game Sales
 - <https://www.kaggle.com/datasets/gregorut/videogamesales>
- IMDB Video Games
 - <https://www.kaggle.com/datasets/muhammadadiltalay/imdb-video-games?resource=download>
- Video Game Sales Dataset
 - <https://www.kaggle.com/datasets/ibriiee/video-games-sales-dataset-2022-updated-extra-feat>