AI Boot Camp Project 1

Video Game Trends

Team Members:

Chris Covill

Keegan Nohavec

Ashwini Kumar

Project Overview

Goals/Questions to be addressed

Analyze leading video game sales & ratings
- **What games are used/rated the best?**
- **What games by genre are rated the best?**
- **What platform (X360, PSx, etc.) is used more for gaming?**
- **What games have the highest sales?**
- **Does the sales data follow the rating information?**
- **How are game sales distributed by region/country?**

Project Overview

Overview of data collection, cleanup and exploration process

- We gathered data from Twitch, & IGDB API's
- Kaggle was added for the Sales Data
- Removed null or NaN values
- Merged the Ratings and Sales Data into 1 dataframe (just for the sales vs ratings)
- Removed duplicate or unneeded columns

Problems Encountered

- Sales data only included up to 2016
- Sales data only includes AAA game companies (Nintendo, Sony, etc)
- Ratings data was very unclean and took a lot of cleaning
- Ratings are based on only those who actually give ratings
- Ratings are subjective, different people have different opinions
- Free API's have limited sales data available

Approach to gathering ratings data from IGDB

```python
#Function to make API request to IGDB
def make_api_request(base_url, df_name):
    offset = 0
    all_results = []
    #loop through all the results, 500 at a time since that is the max limit
    while True:

        #make request to IGDB API
        IGDB_request = requests.post(base_url, headers={'Client-ID': client_id, 'Authorization': f'Bearer {bearer_token }'}, data=f'fields *; limit 500; offset {offset};')
        IGDB_data = IGDB_request.json()

        all_results.extend(IGDB_data)

        #if it pulls less than 500 results, then it has reached the end of the data, so break the loop
        if len(IGDB_data) < 500:
            break

        offset += len(IGDB_data)
    #Put the results into a pandas normalized dataframe
    df_name = pd.json_normalize(all_results)
    df_name.head()
    return df_name
```

```python
#Make API request to IGDB to get the data for the video games
VG_DF = make_api_request(IGDB_URL, 'VG_DF')
```

```python
#make API request to IGDB to get the data for the release dates
RD_DF = make_api_request(release_dates_url, 'RD_DF')
RD_DF.head()
```
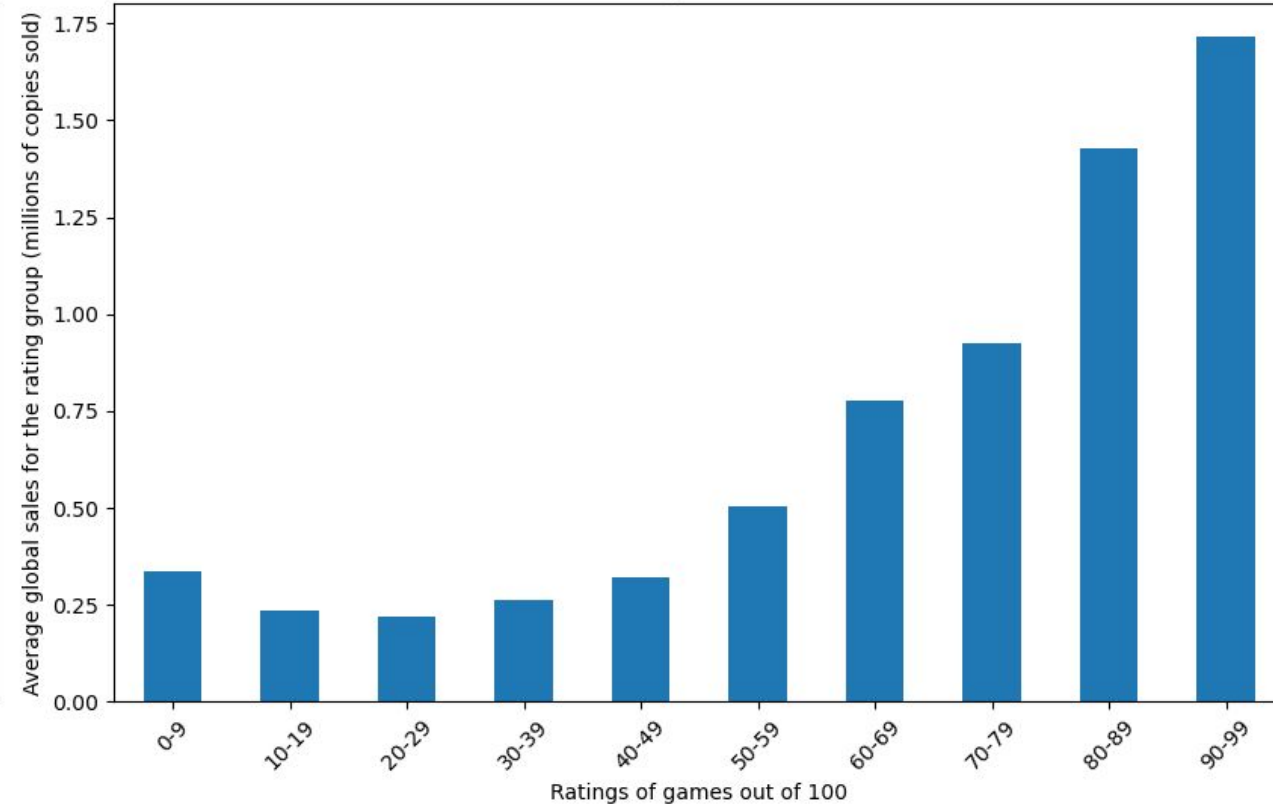
```python
genre_url = 'https://api.igdb.com/v4/genres'
genre_df = make_api_request(genre_url, 'genre_df')
```

Do Ratings Impact Sales?

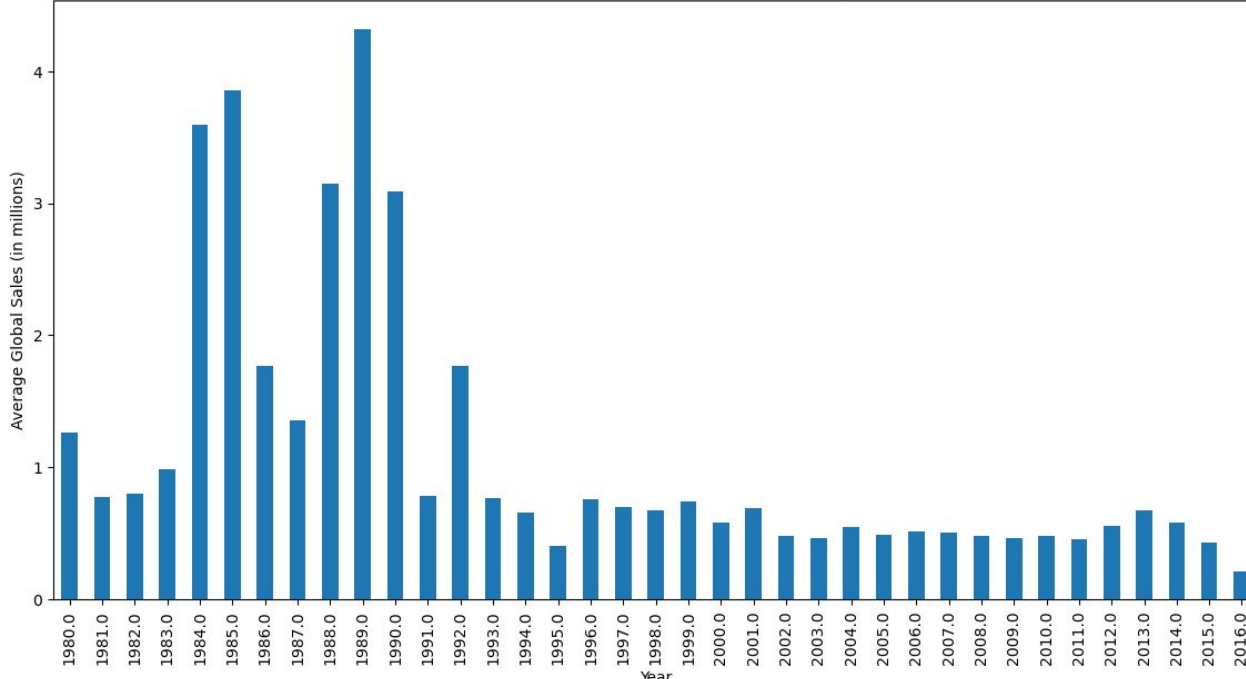Ratings of games vs number of games with that rating
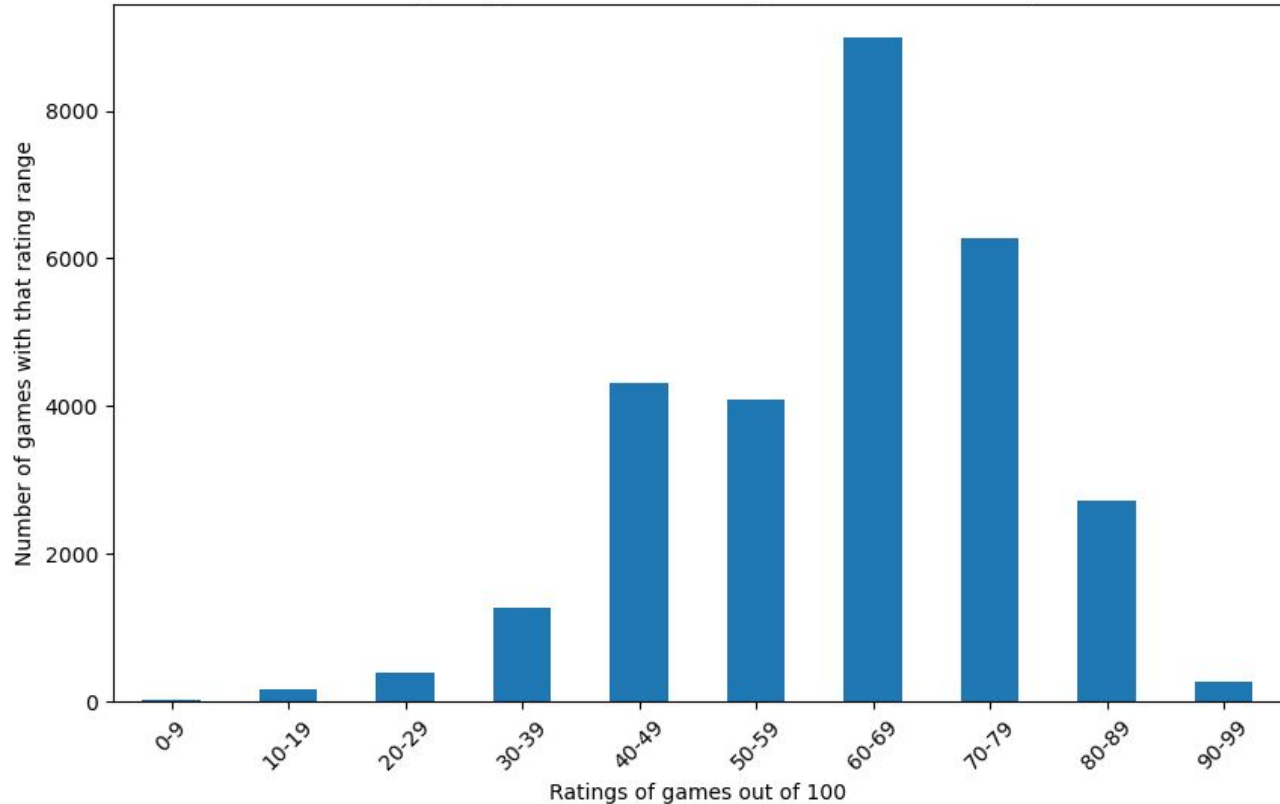
Ratings vs Sales

```
correlation = ratings_and_sales['rating'].corr(ratings_and_sales['global_sales'])
print(f"Correlation between ratings and sales: {correlation}")

✓ 0.0s

Correlation between ratings and sales: 0.15298133746818485
```
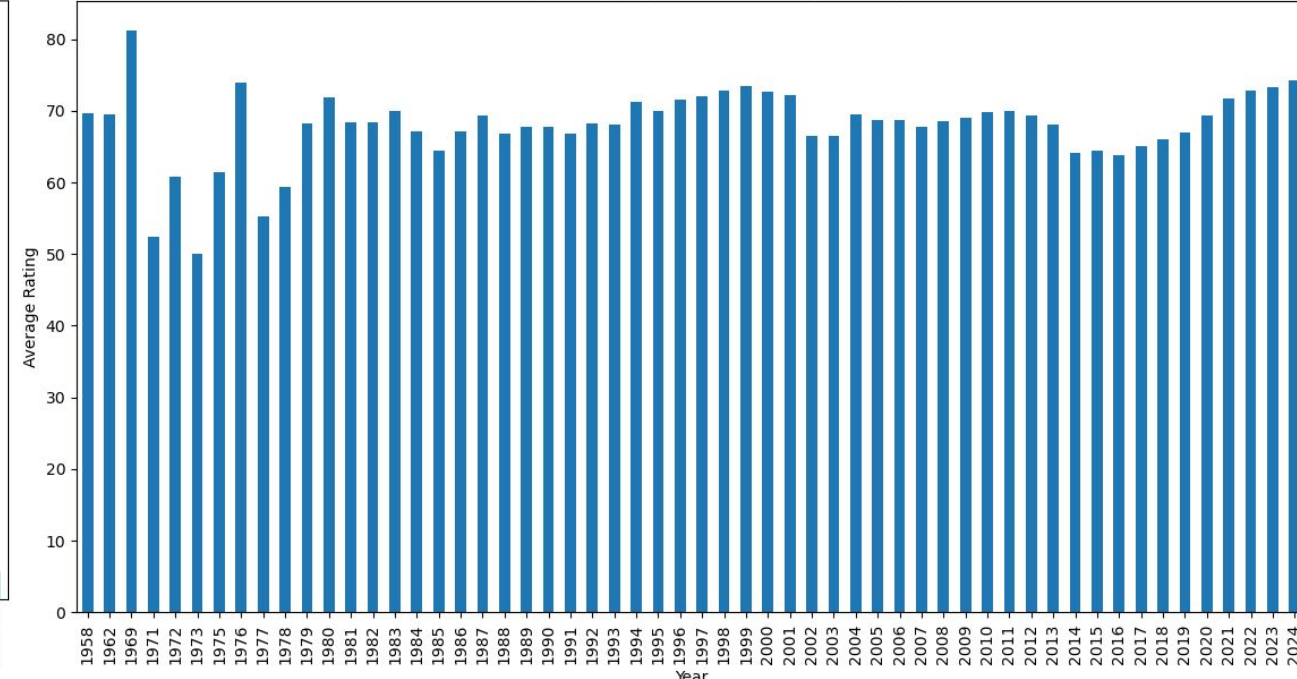
Does the Release Year Impact Sales or Ratings?

Global Sales by Release Year (up to 2016)

Average Ratings by Release Year (up to 2024)

```
correlation_sales_release_year = filtered_sales_df['global_sales'].corr(filtered_sales_df['year'])
print(f"Correlation between global sales and release year: {correlation_sales_release_year}")
✓ 0.0s
Correlation between global sales and release year: -0.07463618577694422
```
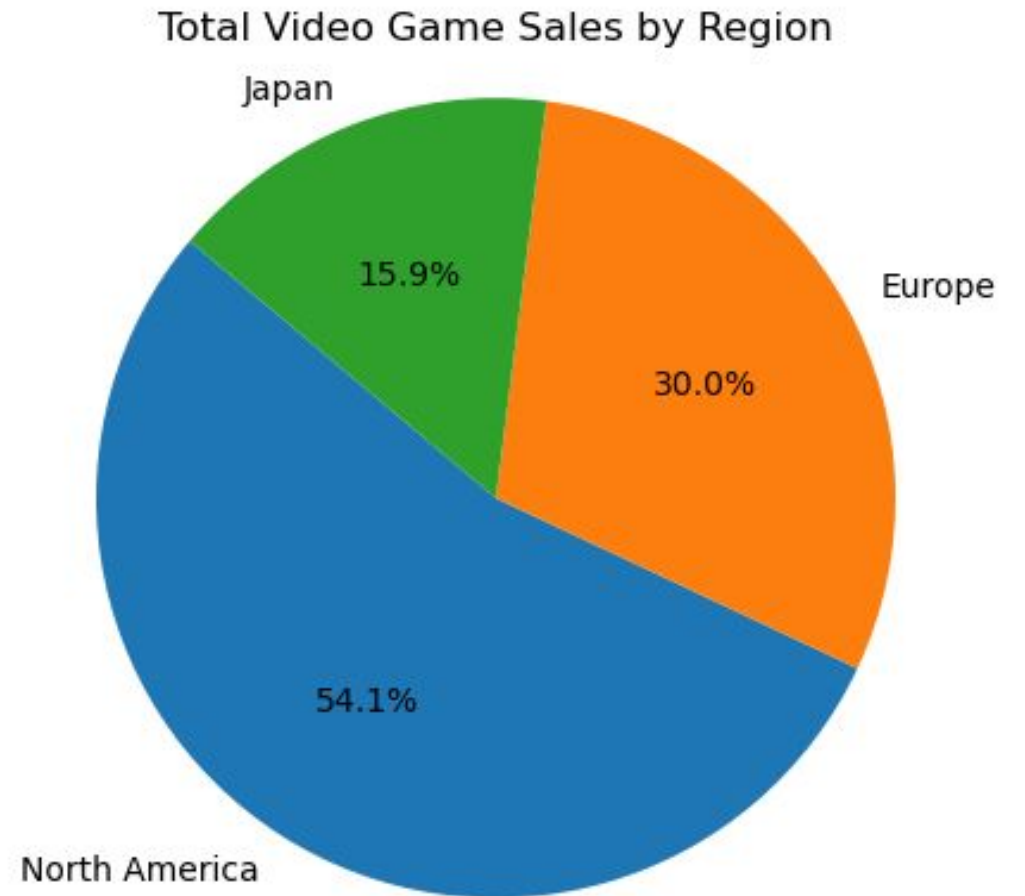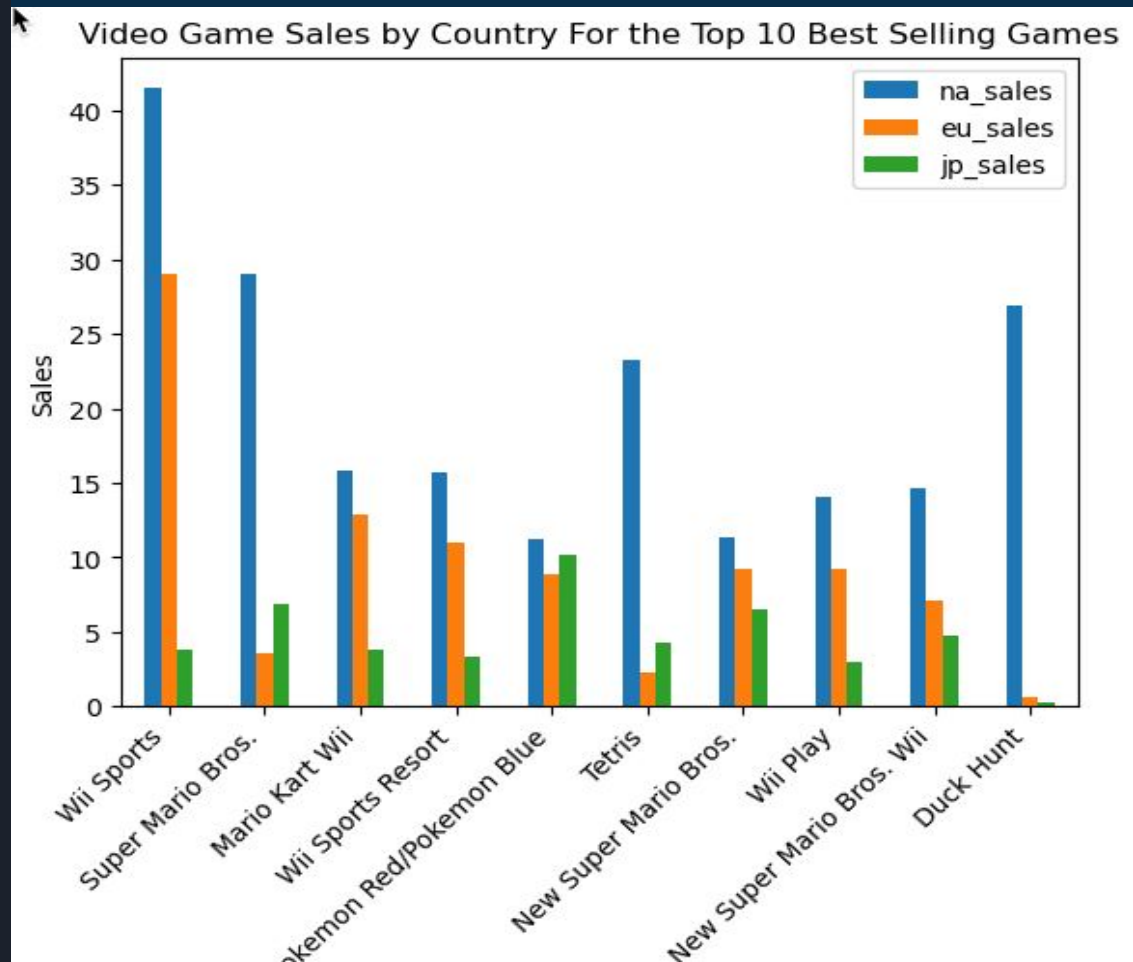
```
correlation_ratings_release_year = filtered_ratings_df['rating'].corr(filtered_ratings_df['release_year'])
print(f"Correlation between ratings and release year: {correlation_ratings_release_year}")
✓ 0.0s
Correlation between ratings and release year: -0.069538058447777
```
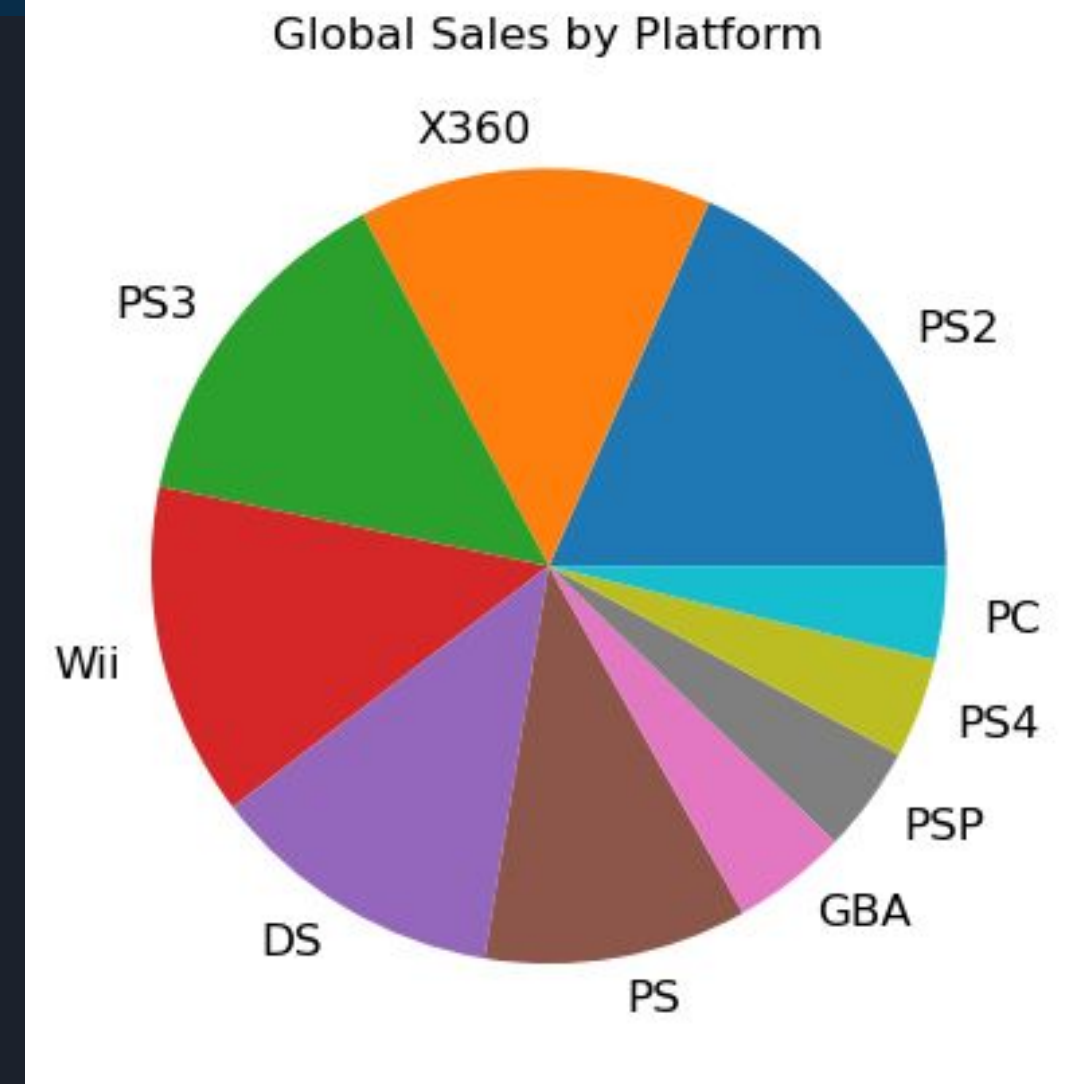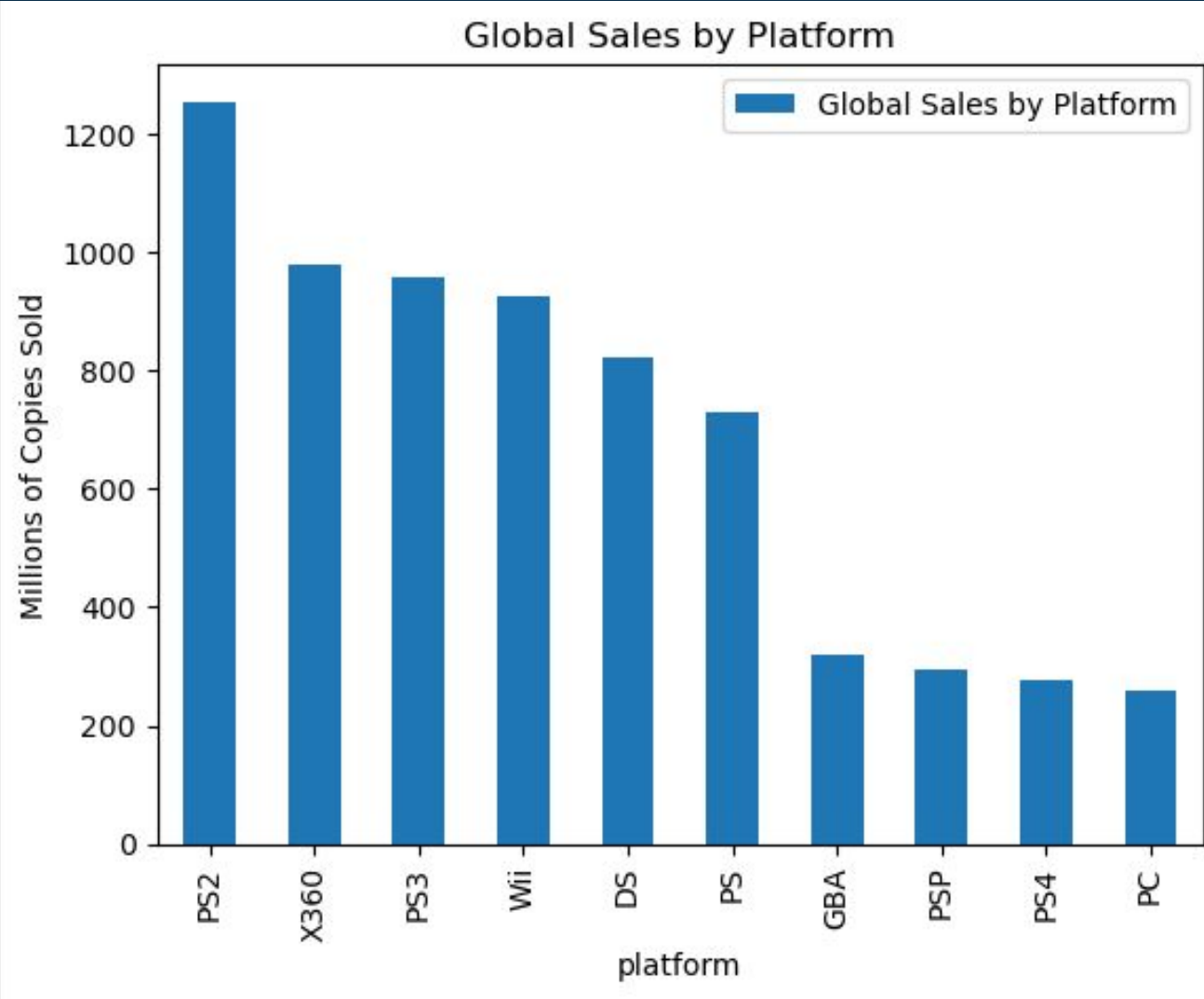
Which Countries Sell the Most Video Games and Which Games?

Video Game Sales by Country For the Top 10 Best Selling Games

Legend: na_sales, eu_sales, jp_sales

Games (x-axis): Wii Sports, Super Mario Bros., Mario Kart Wii, Wii Sports Resort, Pokemon Red/Pokemon Blue, Tetris, New Super Mario Bros., Wii Play, New Super Mario Bros. Wii, Duck Hunt

Total Video Game Sales by Region

Japan 15.9%
Europe 30.0%
North America 54.1%

Which Platforms are the most Popular in the Major Regions?

Global Sales by Platform

What are the Highest Rated Game by Genre?

Average Rating By Genre

Global Sales by Genre (M)
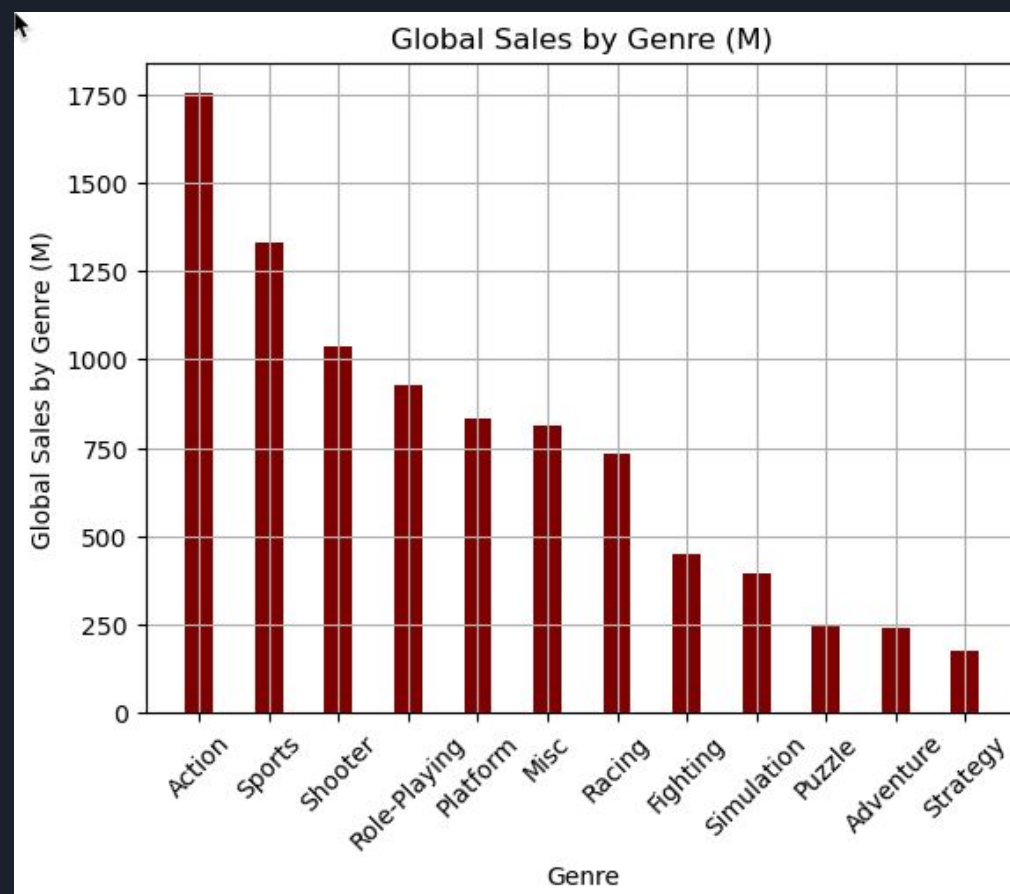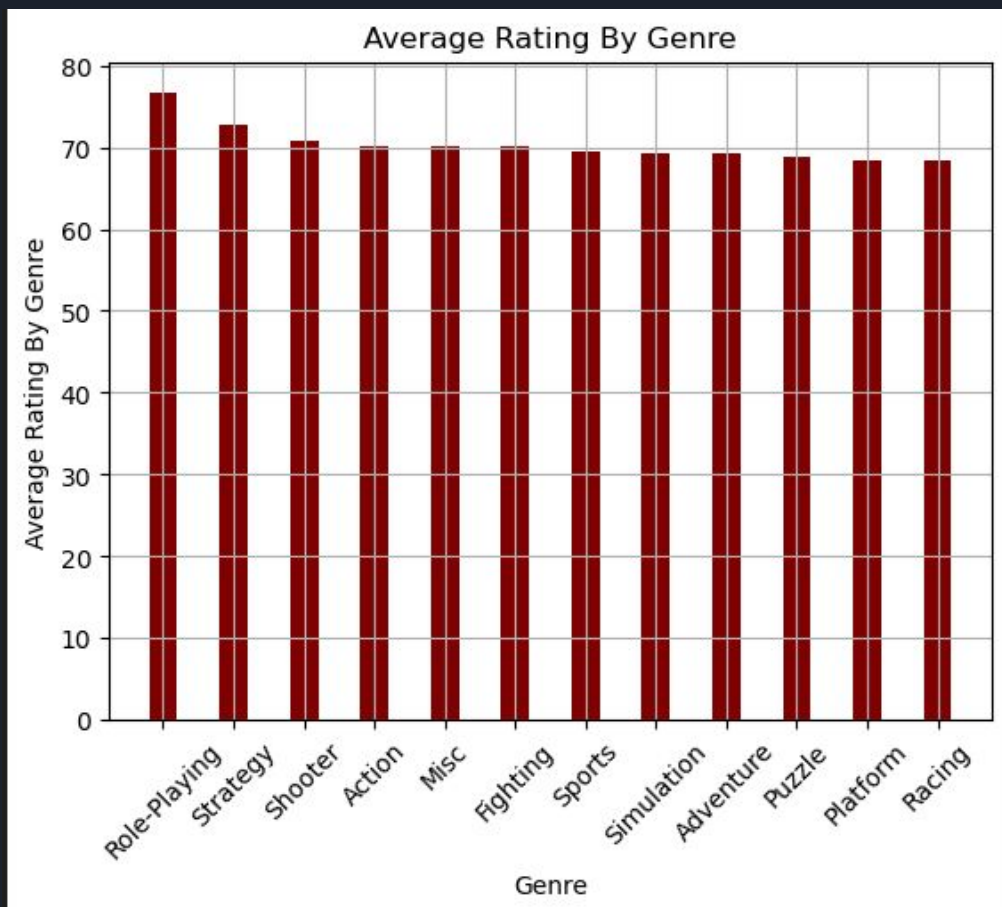
Focusing on Game Publisher Sales - Nintendo top in aggregate sales by Year

Nintendo Global Sales (M)
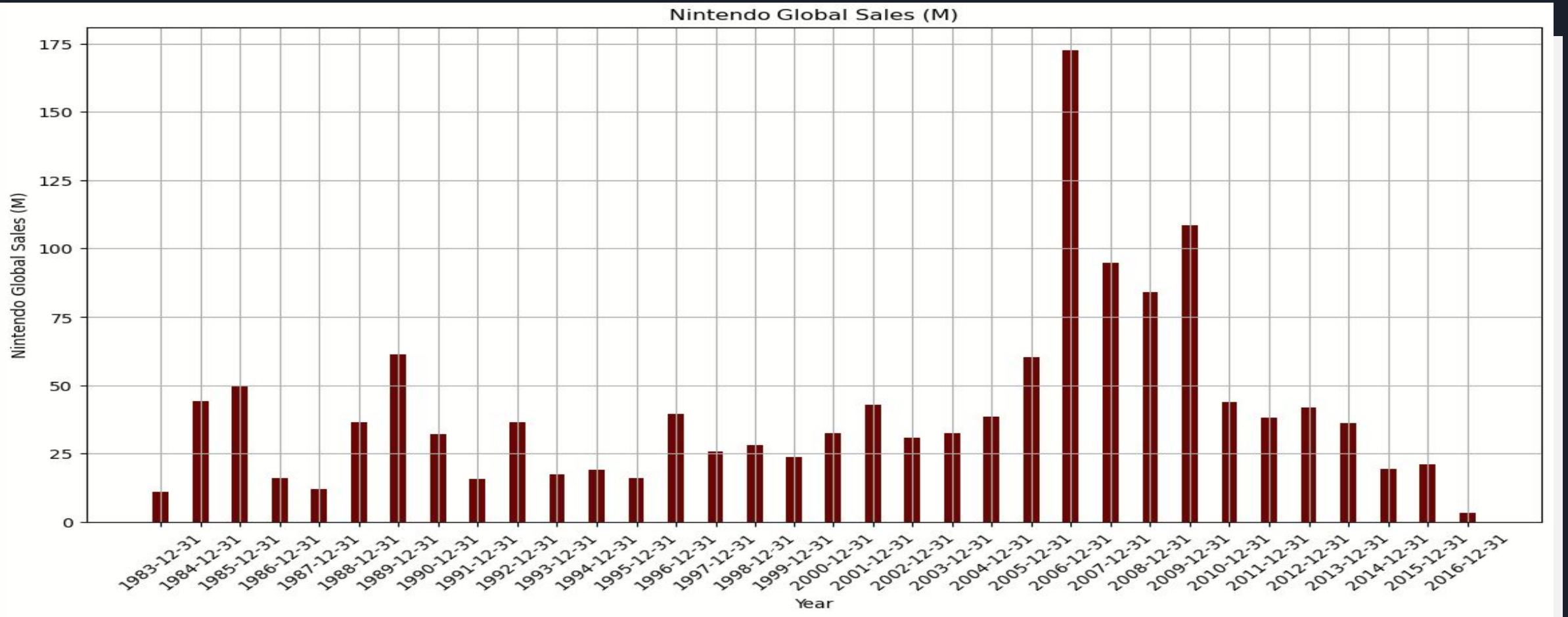
Nintendo Prophet sales forecast - 36-month period (2016 - 2019) using EOY data only
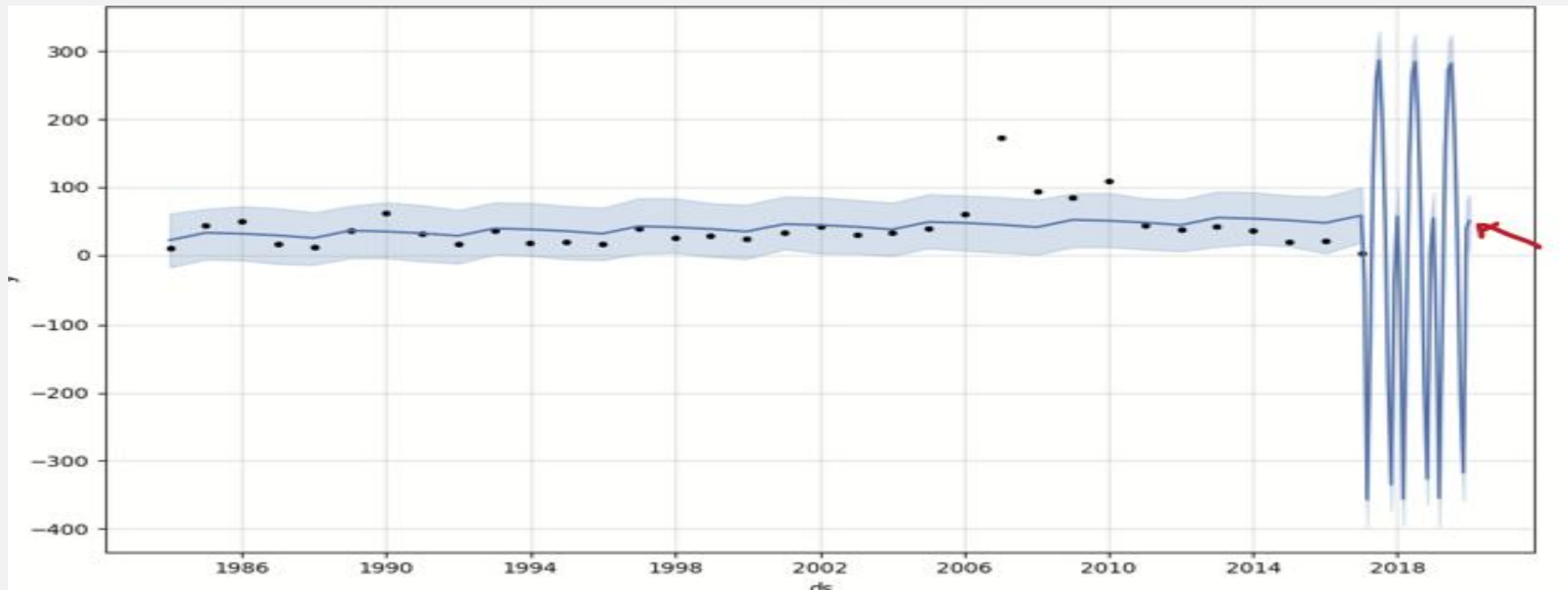
Future Considerations 💡

- **Availability to Key Sales Data from 2017 - 2024 would provide the following insights:**
  - 1 - Video Games Sales during the pandemic
  - 2 - Better insights into the fast moving nature of the game industry
  - 3 - VR Sales which became mainstream during the pandemic
- **Sales data was limited, looking at sales data from other sources, as well as Indie game companies could provide more insights**
- **Looking at ratings from other sources (steam, epic games, etc) and comparing to IGDB could provide more insights as well**

Questions?