

# *Jame’l Brown*

# *Usha Saravanakumar*

# *Crystal Dalsania*

*Video Games*

*Sales Analysis*

ETL Project

Video Games Sales Analysis

This project extracts, transforms, and loads video game sales data from <https://www.vgchartz.com>. This dataset could be later used to analyze and answer the following questions:

* Most popular video game console in various regions
* Most popular video game across various consoles/regions/genre

# Extract

Wrote a [Jupyter notebook](https://github.com/ushaakumaar/Video_Games_Sales_Analysis_ETL/blob/main/Extract.ipynb) python script to scrape the video games data from www.vgchartz.com website using BeautifulSoup and Splinter to do the following:

* Obtained the list of genres
* Dynamically built the URL to scrape the pages of games data for each genre
* Wrote the extracted raw video games data into [CSV file](https://github.com/ushaakumaar/Video_Games_Sales_Analysis_ETL/blob/main/Output/Extracted_video_games_data.csv) to be used for Transform process

# Transform

Wrote a [Jupyter notebook](https://github.com/ushaakumaar/Video_Games_Sales_Analysis_ETL/blob/main/Transform.ipynb) python script to do the following clean up and transformation steps to the raw video games data obtained from the Extraction process:

* Deleted games without sales data
* Removed unwanted columns and renamed the columns
* Added Release year column
* Converted Release date values to date format and Sales values to float format
* Wrote the transformed video games data into [CSV file](https://github.com/ushaakumaar/Video_Games_Sales_Analysis_ETL/blob/main/Output/Transformed_video_games_data.csv) to be used for the Load process

# Load

Performed the following steps to Load the transformed video games data into PostgreSQL Database:

* Designed the database tables in normalized form and generated [ERD](https://github.com/ushaakumaar/Video_Games_Sales_Analysis_ETL/blob/main/Data_Modeling/ERD.png)
* Created a [SQL script](https://github.com/ushaakumaar/Video_Games_Sales_Analysis_ETL/blob/main/Data_Engineering/schema.sql) for table creation in PostgreSQL
* Wrote a [Jupyter notebook](https://github.com/ushaakumaar/Video_Games_Sales_Analysis_ETL/blob/main/Load.ipynb) python script to split the transformed video games data into multiple dataframes and loaded them into PostgreSQL Database tables

# Limitations

* Overall sales information for each game at the region/genre level only was available. Lack of sales history information for each game limits the analysis of sales fluctuation for each game over the years.