Up and Coming Video Games

Team: Zubair Aslam, Eva Fuentes-Lopez, Ammad Rashid, Karyssa Tillis

Topic Selection

The video game industry is constantly growing, with content being widely available across social media, companies, and gaming platforms.

Of these sources, our data will be derived from **Steam**, a video game digital distribution service by Valve.

Users of Steam have access to a library of nearly 30,000 games of various genres, categories, and from third-party publishers.

Data Sources

Kaggle.com

- Steam Store Games (Clean dataset)
  - Combined data of 27,000 games scraped from Steam and SteamSpy APIs (May 2019)
  - https://www.kaggle.com/nikdavis/steam-store-games
  - Creator: Nik Davis
- Steam Games Complete Dataset
  - 40k Steam Games Dataset from Steam shop with detailed data. (June 2019)
  - https://www.kaggle.com/trolukovich/steam-games-complete-dataset
  - Creator: Alexander Antonov

Questions to Answer

What features determine a video game's popularity or success?

Is there a relation between a game's rating and genre?

Is there a relation between a game's rating and tagged categories?

Data Exploration

Determined the following aspects to focus on:

- Popular Tags
- Genre
- Ratings

Reducing outliers by eliminating the following:

- Low Rating counts
- Low counts of miscellaneous Popular Tags

Created separate columns to break up Genre and Popular Tag strings for each game to map its individual categories.

Data Analysis

Utilized Pandas within Jupyter Notebook to explore the datasets and perform cleaning to produce two CSVs: Popular_Tags.csv and Genre.csv. Each contain the video game name, positive review percentage, and split subsequent tags or genre categories.

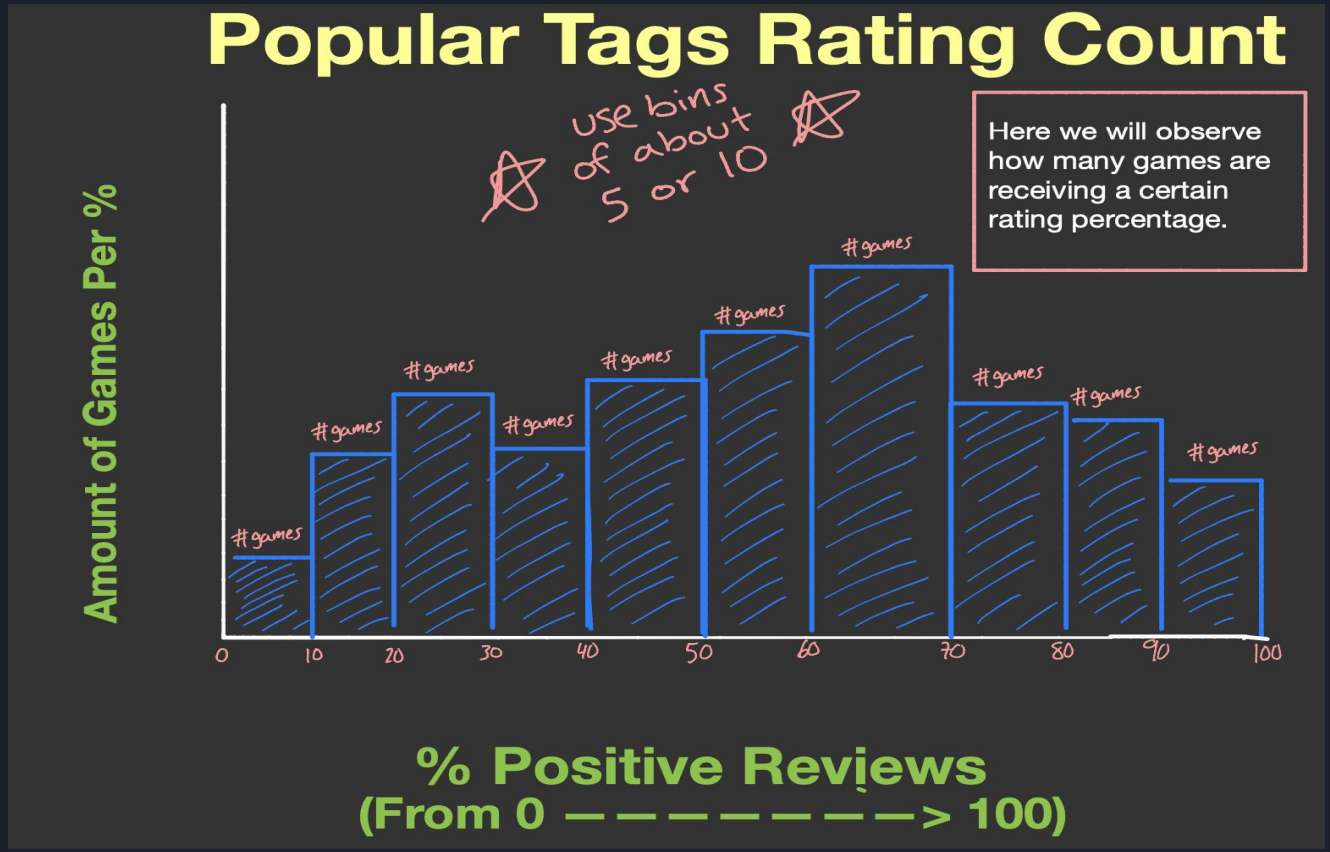
Further cleaning and analysis was performed using Excel to catch minute errors within the rows of data to allow for clean imports into Postgres, Tableau, and for the machine learning portion.

Data Analysis

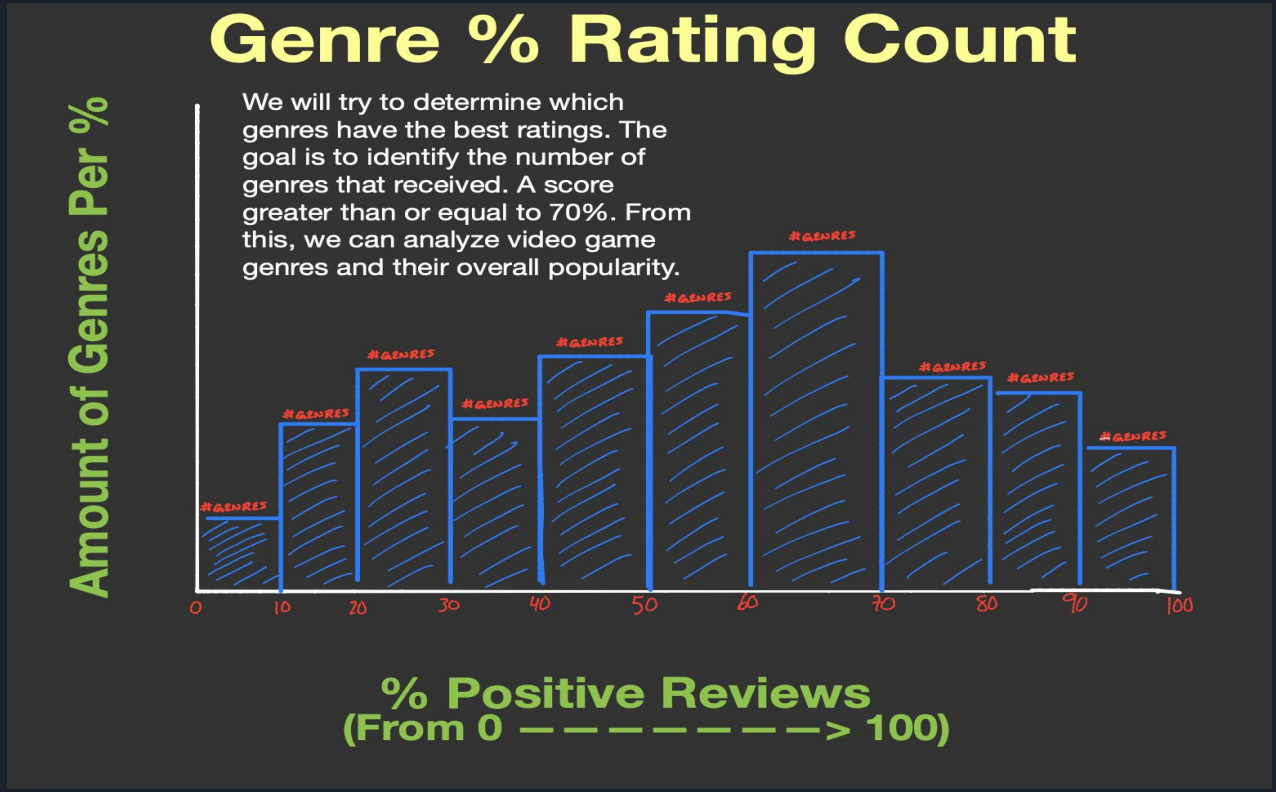
Machine Learning Model

Used the ML library SciKitLearn to create a classifier. Utilized Linear Regression for the training and testing setup to make a prediction.

Expected Visualizations

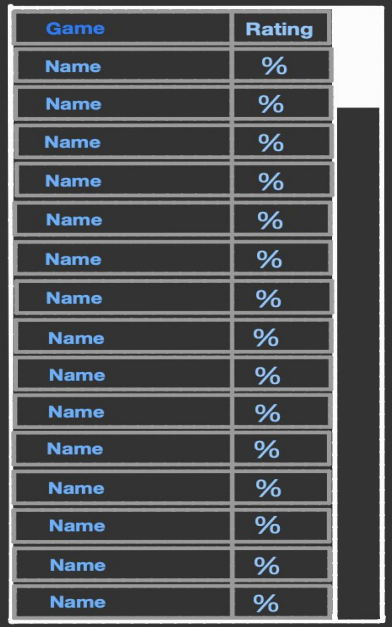


Expected Visualizations Cont...



Expected Visualizations Cont...

Individual Game %



Game	Rating
Name	%
Name	%
Name	%
Name	%
Name	%
Name	%
Name	%
Name	%
Name	%
Name	%
Name	%
Name	%
Name	%
Name	%
Name	%

Sorting = Descending

After identifying the most popular genre (i.e. the genre with the highest rating percentage), we want to find the specific game(s) with the best ratings.

Expected Visualizations Cont...

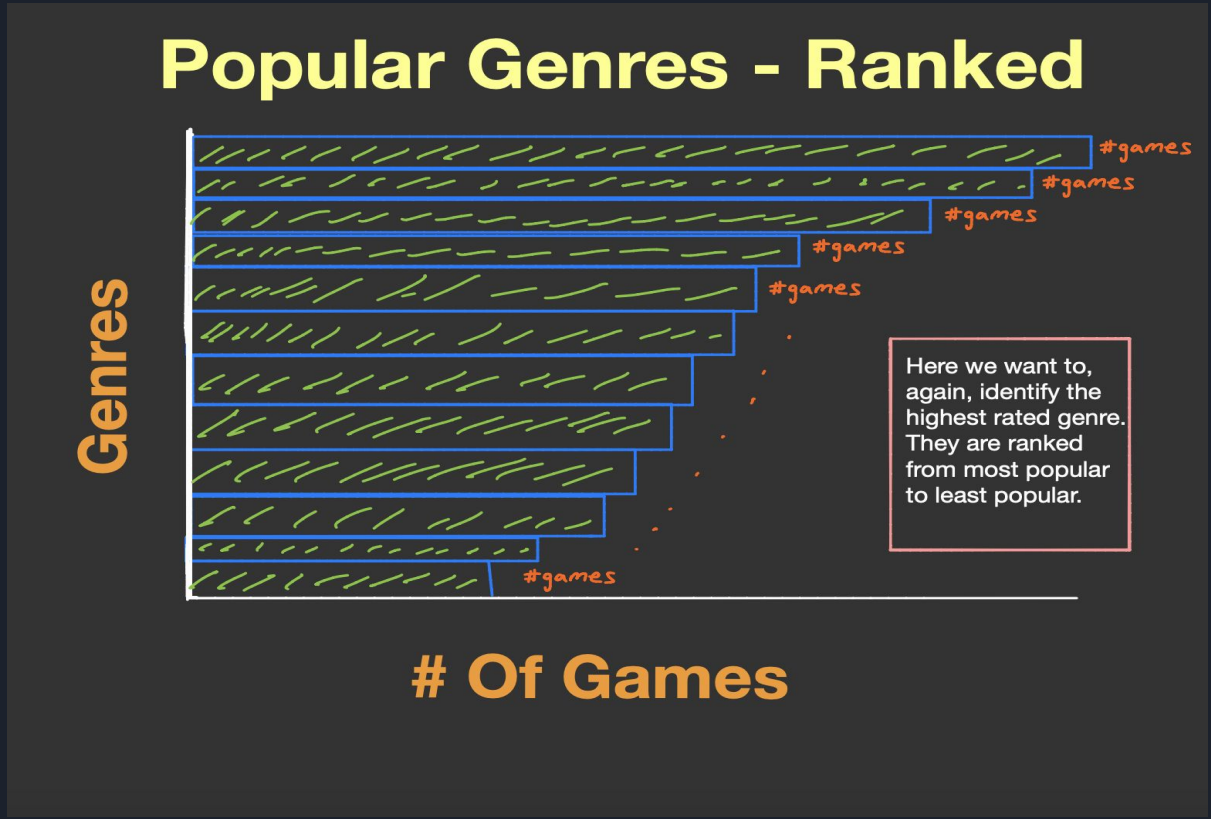


Tableau Storyboard

https://public.tableau.com/profile/eva.fuentes.lopez#!/vizhome/UpComingVideoGames/Story1?publish=yes