



## What Makes a Video Game Great?

### Genre and Rating Disparities in User Reviews

*By: Eden Nuhamovici*

*Community Tech - Data Sciences, Toronto, Canada*

*Email: [enuhamovici@gmail.com](mailto:enuhamovici@gmail.com)*

# Background

- Video games belonging to more niche genres often have polarized views even if they are critically acclaimed.
- IMDb collects video game ratings from the public in a collaborative manner.
- It can be argued that IMDb ratings can influence the purchase of a given video game (among other factors).



# Objectives

- An IMDb rating for a video game ranges from 0.0 to 10.0.
- I am interested in understanding the deeper context behind how a game is given it's rating.
- For example:
  - A game that is rated 10.0 may be niche with only a few ratings. This can lead to inflated scores that do not reflect broad opinion.
  - A lower rated game, for example, 8.5 with many ratings may be more reliable.



TOP 10 HIGHEST RATED VIDEO GAMES RELEASED AFTER 2010		
	GOD OF WAR (2018)	9.9
	THE LAST OF US	9.8
	THE WITCHER 3: WILD HUNT	9.8
	THE WITCHER 3: BLOOD & WINE	9.8
	GRAND THEFT AUTO V	9.7
	UNCHARTED 4: A THIEF'S END	9.6
	MASS EFFECT 2	9.6
	THE WITCHER 3: HEARTS OF STONE	9.6
	THE ELDER SCROLLS V: SKYRIM	9.5
	RED DEAD REDEMPTION	9.5

TOP10WISE.COM

# Problem Statement

- What is the relationship between game rating and genre?
- Ratings alone are often insufficient as they often lack context for why they are given a specific score.
- Users can give very biased scores from recent release days excitement or personal opinion.
- Games can be purposely **review chain-bombed** by malicious users therefore, giving it a bad score and subsequently a bad view to the unknowing public.



r/cyberpunkgame • 3 yr. ago  
Aman690

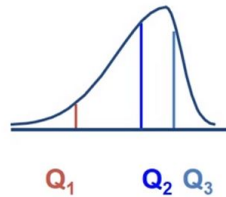
People are now Review Bombing Cyberpunk cause it won Labor of Love 🤡



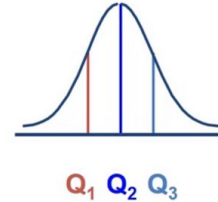
# Methodology

- I will look at game ratings using quartiles to understand how games are scored across genres.
- Segment data by genre to be able to spot differences in how games are rated.
- This will reduce the impact of outliers and skewed data by focusing on quartile distribution.

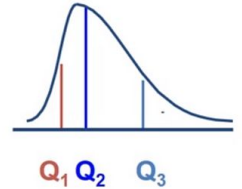
Left-Skewed



Symmetric

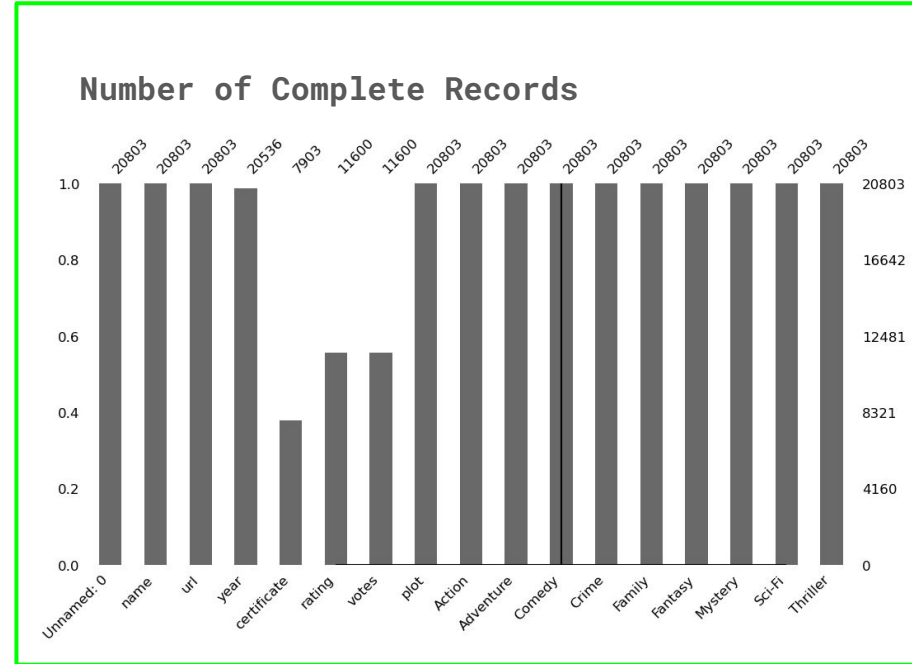


Right-Skewed



# Data Preparation

- The dataset was sourced from Kaggle.
- The raw data contained 20,803 records.
- After removing games which had missing ratings (as imputation wasn't possible), the final dataset used consists of 11600 records.





# Dashboard

