Problem Statement

Using information gathered from Reddit comments, is it possible to figure out what customers' overall sentiment for previous editions of Playstation (PS3 and PS4) and Xbox (360, One), as well as what features play into this, to figure out the optimal features for the newest versions (PS5 and Xbox Series S/Series X)?

Context

With the highly competitive and rapidly evolving gaming market, it is important to figure out what customers like and dislike about different consoles to figure out what needs to be added or removed on the next generation. Using a popular website like Reddit could give us an insight as to how customers feel about these products and what kind of issues they have with it.

Criteria For Success

The goal is to figure out which console has had the best ratings in previous years, as well as to pin down what features are most influential in making or breaking a customer's opinion.

Scope of Solution Space

The focus of this project is going to be on Reddit comments that mention either Playstations, Xboxes, or both in an attempt to figure out which features people generally have a positive or negative opinion on.

Constraints

As Reddit isn't designed to be a website that provides rankings for different products, I'll have to put some machine learning to use to score each comment as positive or negative, and by how much. In addition, the database I'm using only goes up to March of 2017, as it was a publicly available personal project rather than something the creators were hired to do. One of the downsides to the year range is there is little useful data on another competing console, the Nintendo Switch, which came out during the last month the database was updated, although there should be information on the Nintendo Wii.

Stakeholders

This information would be useful for any company trying to market a gaming console, despite the data being focused on Playstations and Xboxes.

Data Sources

Reddit users Dewarim and Stuck_in_the_Matrix worked together to create a database containing all public Reddit comments from 2005 through part of 2017. Kaggle has a database containing the number of mentions for various products from 2015-2017.

https://files.pushshift.io/reddit/comments/

https://www.kaggle.com/residentmario/things-on-reddit